Journal of Machine Learning Research 22 (2021) 1-62 Submitted 9/19; Revised 10/20; Published 9/21

On the Estimation of Network Complexity: Dimension of Graphons

Yann Issartel YANN.ISSARTEL@ENSAE.FR
Université Paris-Saclay, Laboratoire de Mathématiques d'Orsay,
91405 Orsay, France.
CREST, Ensae, Institut Polytechnique de Paris, 91120 Palaiseau, France.

Editor: Edo Airoldi

Abstract

Network complexity has been studied for over half a century and has found a wide range of applications. Many methods have been developed to characterize and estimate the complexity of networks. However, there has been little research with statistical guarantees. In this paper, we develop a statistical theory of graph complexity in a general model of random graphs, the so-called graphon model.

Given a graphon, we endow the latent space of the nodes with the neighborhood distance. Our complexity index is then based on the covering number and the Minkowksi dimension of this metric space. Although the latent space is not identifiable, these indices turn out to be identifiable. This notion of complexity has simple interpretations on popular examples: it matches the number of communities in stochastic block models; the dimension of the Euclidean space in random geometric graphs; the regularity of the link function in Hölder graphons.

From a single observation of the graph, we construct an estimator of the neighborhood-distance and show universal non-asymptotic bounds for its risk, matching minimax lower bounds. Based on this estimated distance, we compute the corresponding covering number and Minkowski dimension and we provide optimal non-asymptotic error bounds for these two plug-in estimators.

Keywords: random graph model, graphon, neighborhood distance, covering number, Minkowski dimension

1. Introduction

Networks appear in many areas where data is a collection of objects interacting with each other. Examples include numerous phenomena in the fields of physics, biology, neuroscience and social sciences. A major issue is to extract information from these data repositories. This exciting challenge has led researchers to seek characterizations of networks, among which their complexity has received a lot of attention for more than half a century. See (Dehmer and Mowshowitz, 2011; Zenil et al., 2018) for two recent reviews. Indeed, network complexity is a key feature used in various applications, for example, to quantify the complexity of chemical structures (Bonchev and Buck, 2005), to describe business processes (Latva-Koivisto), to characterize software libraries (Veldhuizen, 2005), and to study general graphs (Constantine, 1990).

©2021 Yann Issartel.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v22/19-747.html.

The definition and estimation of network complexity is an active line of research (Morzy et al., 2017; Zufiria and Barriales-Valbuena, 2017; Claussen, 2007). However, there appear to be little (or no) mathematical results on the statistical side of the problem. In this paper, we develop a statistical theory of graph complexity in a universal model of random graphs. To the best of our knowledge, it is the first contribution on complexity estimation with statistical guarantees.

1.1 Modeling Assumption

Statistical inference on random graphs is a fast-growing area of research (Matias and Robin, 2014; Rácz et al., 2017; Abbe, 2017) and has found a wide range of applications (Goldenberg et al., 2010; Sarkar et al., 2011). Usually, it assumes there exists an unknown feature in the underlying model and the goal is to recover this feature from a single realization of the random graph.

Here, we follow this direction with the *W-random graph model* (also known as graphon model). This general model falls into the category of non-parametric descriptions of networks (Bickel and Chen, 2009) and satisfies some forms of universality (Diaconis and Janson, 2007). See section 1.3 for details. In this paper, we define a notion of complexity for this model and then consider the problem of inferring this complexity from a single graph observation.

W-random graphs allow to model many real-world networks, such as social networks where nodes represent different people and edges people's friendships. In this example, one may expect that the friendship probability p_{ij} between individual i and j depends on their personal attributes (like jobs, ages, leisure). To model such mechanism, one may assume the observed graph is generated according to the W-random graph model, i.e. (1) for each node i of the network, an attribute ω_i is drawn from a distribution μ on a space Ω (where Ω can be seen as the social space of all possible individual features: jobs, ages,...); (2) two people are friends, independently of the others, with probability $p_{ij} = W(\omega_i, \omega_j)$, where $W : \Omega \times \Omega \to [0, 1]$ is a symmetric function. Thus, a W-random graph is specified by the triplet of parameters (Ω, μ, W), often called *graphon* in the literature (Lovász, 2012).

Such modeling falls into the popular "latent space approach" (Hoff et al., 2002). Indeed, the personal attributes may not be observed in practice and accordingly, the W-random graph model assumes that the ω_i and Ω are latent (unobserved). In fact, all parameters of the graphon (Ω, μ, W) are unknown, and the only observation is the edges of the graph, i.e. the adjacency matrix A where $A_{ij} = 1$ stands for the presence of an edge between the ith and j^{th} nodes, and $A_{ij} = 0$ otherwise. See Section 2 for a formal presentation of this model.

1.2 Contribution

1.2.1 Complexity Index

Our first objective is the definition of a complexity index in the W-random graph model. As a natural candidate, one might think of the dimension of the latent space, like d if $\Omega = [0, 1]^d$. However, this index is inadequate because of a major identifiability issue. Indeed, it is known that (see Lovász, 2012) the attribute space Ω is not identifiable from the observed adjacency matrix A. Even worse, it has been shown that all W-random graph distributions can be

represented on the specific space $\Omega = [0,1]$ (Lovász, 2012). It is therefore pointless to think about the graph complexity purely in terms of the latent space. Likewise, the regularity of the link function (like α if W is α-Hölder) is not suited due to the non-identifiability of W.

These issues motivate the introduction of a more abstract index. Given a graphon (Ω, μ, W), we endow the latent space Ω with the so-called *neighborhood distance*

$$r_W(\omega, \omega') = \left(\int_\Omega \left| W(\omega, \omega'') - W(\omega', \omega'') \right|^2 \, \mu(d\omega'') \right)^{1/2}. \tag{1}$$

From the above description of a W-random graph, we can see that the quantity $r_W(\omega_i, \omega_j)$ measures the propensity of the nodes i and j to be connected with similar nodes. Our complexity index is then defined as the covering number and the Minkowski dimension of a purified version of the (pseudo-) metric space (Ω, r_W). The purification process is detailed in section 3.1. Recall the definitions of these two standard measures for metric spaces: the ϵ-covering number $N_\Omega^{(c)}(\epsilon)$ is the minimal number of balls of radius ϵ required to entirely cover the (pseudo-) metric space (Ω, r_W). And the Minkowski dimension is the following limit on the covering number

$$dim\, \Omega := \lim_{\epsilon \to 0} \frac{\log N_\Omega^{(c)}(\epsilon)}{-\log \epsilon} \tag{2}$$

when the limit exists. In particular, the Minkowski dimension does not have to be an integer.

Although none of the three parameters Ω, μ and W are identifiable in the W-ranom graph model, we prove that the covering number and the Minkowski dimension of a purified version of (Ω, r_W) are identifiable.

We also illustrate that this notion of complexity is sound on classic examples of random graphs. Specifically, we show that $N_\Omega^{(c)}(\epsilon)$ is equal to the number of well-spaced communities in the stochastic block model; that $dim\, \Omega$ matches the dimension of the Euclidean space in some random geometric graphs; and that $dim\, \Omega$ is equal to the regularity of the link function in some Hölder graphon models. See Section 3.2 for details.

In addition to all applications listed in the introduction, these complexity indices may also be useful to adjust analytical methods to particular networks, for example, when estimating the link function W (see section 1.3 and 3.2 for related comments) or in learning representation where the goal is to find an informative metric space to place/represent the nodes of the network (Hoff et al., 2002; Perozzi et al., 2014; Grover and Leskovec, 2016).

1.2.2 Statistical Estimation

From the observed adjacency matrix A of a W-random graph, we estimate the neighborhood distance (1) on the sampled points $\omega_1, \ldots, \omega_n$. The corresponding distance estimator $\widehat{r}$ is defined in Section 4.1. We show universal non-asymptotic bounds for its risk (Theorem 1). Let $\omega_{m(i)} \in \{\omega_1, \ldots, \omega_n\} \setminus \{\omega_i\}$ denote a nearest neighbor of ω_i with respect to the distance r_W.

Theorem 1 *Consider the distance estimator $\widehat{r}$, defined in Section 4.1. Then, for any graphon (Ω, μ, W), we have*

$\forall i, j \in [n]$,

$$\left| r_W^2(\omega_i, \omega_j) - \widehat{r}^2(i,j) \right| \lesssim r_W(\omega_j, \omega_{m(j)}) + r_W(\omega_i, \omega_{m(i)}) + \sqrt{\log(n)/n}$$

with probability at least $1 - 2/n$.

In the upper bound, there is a bias term $r_W(\omega_j, \omega_{m(j)})$ which is the distance between the sampled point ω_j and its nearest neighbor $\omega_{m(i)}$ (w.r.t. the neighborhood distance). This bias depends on the form of the underlying graphon (Ω, μ, W), for example, it is equal to zero w.h.p. in the stochastic block model (i.e., when the link function W is piecewise constant on $\Omega = [0,1]$). We also derive a minimax lower bound that matches the upper bound of Theorem 1. See Section 4.2 for details on the distance estimation.

Based on the estimated distances $\widehat{r}(i,j)$, we estimate the covering number $N_\Omega^{(c)}(\epsilon)$ by plug-in and provide universal non-asymptotic error bounds for this estimator. See Section 4.3 for details. Our results on the distance and covering number are therefore valid for all graphons, unlike most results in the graphon literature.

Combining the above covering number estimator $\widehat{N}_\Omega^{(c)}$ with formula (2), we derive an estimator of the Minkowski dimension

$$\widehat{dim}_D := \frac{\log \widehat{N}_\Omega^{(c)}(\epsilon_D)}{-\log \epsilon_D}$$

which satisfies a high probability convergence rate (Theorem 2). For this result, we assume the Mikowski dimension is upper bounded by some constant D and use a particular radius ϵ_D defined in Section 4.4. We also make some mild assumptions on the graphon geometry, which are inspired by the problem of estimation of manifold dimension (see section 1.3 for this related literature). Besides, we show that this set of assumptions is minimal, in the sense that, if any of these assumptions is removed, all dimension estimators make an estimation error of the order 1.

Theorem 2 *Under some mild assumptions, defined in Section 4.4, the following holds. If* $dim\,\Omega$ *is any real in* $[0, D]$, *then*

$$\left| \widehat{dim}_D - dim\,\Omega \right| \lesssim \frac{1}{\log n}$$

with probability at least $1 - C'/n$ *for some constant* C' *independent of* n.

Finally, we prove that the upper bound $\log^{-1} n$ is optimal, which means that no estimator can improve on this error. For detailed results, see Section 4.4.

As extensions, we show that the above results also cover the important setting of sparse networks, which has been considered several times in the literature (see Bickel et al., 2011; Wolfe and Olhede, 2013; Klopp et al., 2017; Xu et al., 2014). In addition, we describe a polynomial-time algorithm to approximate the covering number estimator; we do so by using a classic greedy algorithm that is known to satisfy some theoretical guarantees. See Section 6 for these two extensions.

Finally, we test if the packing number (of a purified version of (Ω, r_W)) is smaller than K, with a specific care for controlling the type I error probability uniformly over all graphons.

We prove this error is smaller than $2/n$ for any graphon. For technical reasons detailed in Section 5, we use here the packing number instead of the covering number, which are essentially the same measures (see Appendix A for a reminder about these usual measures for metric spaces).

1.3 Connection with the Literature

1.3.1 W-random Graph Model

The most simple random graph is the Erdös-Rényi model where each edge has the same probability p of being present, independently of the other edges. The study of this generative model has been impressively fruitful in mathematics (Bollobás, 1998) but does not replicate even the simplest properties of real-world networks. Hence, the assumption of a constant connection probability p has been relaxed in the celebrated stochastic block model (Holland et al., 1983) where the connection probabilities may vary with the community membership of each node. Although this model has attracted a lot of attention (Abbe, 2017), it fails to catch some subtle aspects of very large graphs. Such modeling issues have led to a non-parametric view of network analysis (Bickel and Chen, 2009), in particular the introduction of the W-random graph model (Diaconis and Janson, 2007).

The universality of the W-random graphs has two parts. On the one hand, the graphon (Ω, μ, W) plays a key role in network analysis as a powerful representation of many graph properties. Indeed, it has been shown that many sequences of growing graphs can be represented by graphons. For details, see the theory of graph limits introduced by Lovász and Szegedy (2006) or the comprehensive monograph by Lovász (2012). On the other hand, the W-random graph model is connected with the theory of exchangeable random graphs. In fact, every distribution on random graphs that is invariant by permutation of nodes can be expressed with W-random graphs (Diaconis and Janson, 2007; Aldous, 1981; Kallenberg, 1989). Thus, the W-random graphs encompass many random graph models, including stochastic block models, random geometric graphs (Penrose et al., 2003) and random dot product graphs (Tang et al., 2013; Athreya et al., 2017).

1.3.2 Graphon Estimation

There has been much interest in the recovery of the function W (or the matrix of probabilities $[W(\omega_i, \omega_j)]_{i,j \leq n}$) on the specific space $\Omega = [0, 1]$. Usually, authors assume the graphon has some regularity (e.g. W is Hölder continuous on $[0, 1]$) and then use an approximation by SBM, which can be seen as an approximation by constant piecewise functions of W (Borgs et al., 2015; Wolfe and Olhede, 2013; Gao et al., 2015; Klopp et al., 2017; Latouche and Robin, 2016). We also mention an alternative approach based on neighborhood-smoothing (Zhang et al., 2015; Xu et al., 2014). In comparison with this literature, our objective is less ambitious since we only estimate a feature of the graph (its complexity). In return, we carry out a general analysis and do not assume any smoothness condition on $\Omega = [0, 1]$. Indeed, our results on the neighborhood distance and covering number estimations are valid for all graphons. For the dimension, we make mild assumptions which are similar to those in the "intrinsic dimension estimation" literature (see subsection 1.3.3 for a brief description of this related problem).

In the estimation problem of W, the latent space $[0,1]$ is sometimes considered instead of Ω. This choice is not restrictive (if no assumption is made on the function W on $[0,1]$) because both settings generate the same W-random graph distributions (Lovász, 2012). However, the restricted setting $[0,1]$ is not always convenient to work with, whereas the general setting Ω leads to simpler and cleaner situations (Lovász, 2012). Indeed, many random graph distributions are naturally represented on Ω so that their properties are easy to interpret. See Section 3.2 for illustrative examples.

The l_2-neighborhood distance (1) is a variant of the l_1-neighborhood distance introduced by Lovász (2012). This variant has been leveraged several times for the estimation of $[W(\omega_i, \omega_j)]_{i,j \leq n}$ (Zhang et al., 2015; Xu et al., 2014) where the authors use it as a criterion to select neighborhoods of nodes. Here, our estimator of the l_2-neighborhood distance is inspired by the work of Zhang et al. (2015), as will be discussed later.

In sparse settings as well, the matrix $[W(\omega_i, \omega_j)]_{i,j \leq n}$ can be estimated by averaging over "similar" observed entries, where "similar" is defined by some neighbor estimator. In (Li et al., 2019) for example, the neighbor estimator is simply the l_2-distance between two rows in the observed matrix, and the random fraction of observed entries is larger than $n^{-1/2+\delta}$ for any $\delta > 0$. We also consider this sparse regime for the estimation of the l_2-neighborhood distance, but we do not use the same estimator as Li et al. (2019) since it suffers from a constant bias $O(1)$. For details, see the equations page 7 in (Li et al., 2019).

However, the distance estimators in (Li et al., 2019; Zhang et al., 2015) and the current paper are only based on immediate neighborhoods, which are not sufficiently informative in very sparse regimes, where the fraction of observed entries is $o(n^{-1/2})$. Although this is not the focus of our paper, we mention that that there exist successful methods in such very sparse regimes. For example, the similarity between two nodes can be estimated by comparing the two sets of paths starting from these nodes (Borgs et al., 2017). By considering paths of length (strictly) larger than 1, this approach allows to gather information within larger neighborhoods than immediate neighbors. However, the theoretical guarantees proved for this method (Borgs et al., 2017) are restricted to functions W of finite spectrum and defined on $[0,1]$, whereas we study any function on any space Ω.

1.3.3 Intrinsic Dimension Estimation

There is a considerable body of literature on the estimation of intrinsic dimension of a manifold (Kim et al., 2016; Kégl, 2003; Koltchinskii, 2000; Levina and Bickel, 2005). In the simplest setting, points are sampled on a manifold of $\mathbb{R}^m$ whose dimension is an integer, and the objective is to recover this dimension from the sample. In contrast, here we do not assume the dimension is an integer, we do not observe the n sampled points $\omega_1, \ldots, \omega_n$, and we are not in the Euclidean metric space $\mathbb{R}^m$. Indeed, the neighborhood distance r_W is unknown, and our only observation is the connections of the graph.

Outline of the Paper. Section 2 gives a formal presentation of the problem. Section 3 presents the complexity index and some illustrations. In Section 4, we focus on statistical estimation (distance, covering number, dimension). In Section 5, we test the graph complexity. In Section 6, we provide two extensions (estimation on sparse graphs, and a polynomial-time algorithm). Proofs are deferred to the appendix.

Notation. we write $a \lesssim b$, if there exists a constant C such that $a \leq Cb$; and note $a \asymp b$, if there exist two constants c, c' such that $ca \leq b \leq c'a$. We denote by $a \vee b$ (respectively $a \wedge b$) the maximum (resp. minimum) between a and b; by $[a]_+$ the maximum between 0 and a; by $[n]$ the set $\{1, \ldots, n\}$; by $B(x, \epsilon)$ a ball of radius ϵ and center x. We note $1_{\mathcal{E}}$ the indicator function corresponding to any event $\mathcal{E}$. We write "a.e." for "almost everywhere"; and "w.r.t." for "with respect to"; and "w.h.p." for "with high probability", which means that the probability converges to 1 as the number of graph nodes tends to infinity.

2. Model

2.1 Setting

For a set of vertices $V = \{1, \ldots, n\}$, a *W-random graph* $G = (V, E)$ is generated as follows. Let (Ω, μ, W) be an unknown triplet of parameters, which is composed of a measurable set Ω, a probability measure μ on Ω, and a symmetric (measurable) function $W : \Omega \times \Omega \to [0, 1]$. Such a triplet is called a graphon and we write $\mathcal{W}$ the collection of all graphons. For each node $i \in V$, an unknown attribute $\omega_i \in \Omega$ is drawn in an i.i.d. manner from the distribution μ. Conditionally to the attributes $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$, an edge connects two vertices i and j, independently of the other edges, with probability $W(\omega_i, \omega_j)$.

$$\mathbb{P}\big((i, j) \in E \,\big|\, \boldsymbol{\omega}\big) = W(\omega_i, \omega_j) \tag{3}$$

Our data are a single observation of the W-random graph. Formally, it is an adjacency matrix $A = [A_{ij}]_{i,j \leq n}$ defined by $A_{ij} = 1$ if $(i, j) \in E$, and 0 otherwise. This symmetric binary matrix with zero-entries on the diagonal represents an undirected, unweighted graph with no self edges. The distribution of A is called the data distribution and is denoted by $\mathbb{P}^{(n)}_{(\Omega,\mu,W)}$. The goal will be to define an index characterizing the complexity of the *limiting* distribution $\mathbb{P}^{(\infty)}_{(\Omega,\mu,W)}$. In particular, the index should be identifiable from all distributions $\mathbb{P}^{(n)}_{(\Omega,\mu,W)}$, $1 \leq n \leq \infty$, in order to be estimated from the data A. The dependence in n is often dropped out in the rest of the paper.

2.2 Non-identifiability and Equivalence Class of Graphons

From the observation A, the function W is not identifiable. Indeed, for any measure-preserving bijection $\phi : \Omega \to \Omega$, we can observe that the map $W^{\phi}(x, y) = W(\phi(x), \phi(y))$ leaves the data distribution unchanged, i.e.:

$$\mathbb{P}_{(\Omega,\mu,W)} = \mathbb{P}_{(\Omega,\mu,W^{\phi})}.$$

In fact, even the latent space Ω is not identifiable. The full picture is described by Lovász (2012, chap.10):

> *Two graphons (Ω, μ, W) and (Ω', μ', W') parametrize the same data distributions for all n, if and only if, there exist some measure-preserving maps $\phi : [0, 1] \to \Omega$ and $\psi : [0, 1] \to \Omega'$ such that $W^{\phi}(x, y) = W'^{\psi}(x, y)$ a.e.*

where $[0,1]$ is the probability space endowed with the uniform measure. This characterization will be useful to prove the identifiability of our complexity index. For clarity of this future discussion, we consider the corresponding quotient space $\mathcal{W}/\sim$, which is the set of equivalence classes of graphons leading to the same data distributions.

3. Complexity Index

Given a graphon (Ω, μ, W), we endow the latent space Ω with the *neighborhood distance*

$$r_W(\omega, \omega') = \left(\int_\Omega \left| W(\omega, \omega'') - W(\omega', \omega'') \right|^2 \, \mu(d\omega'') \right)^{1/2} \tag{4}$$

which is the l_2-norm $||W(\omega, .) - W(\omega', .)||_{2,\mu}$ between the slices of the function W in ω and ω'. Then, we measure the complexity of the pseudo-metric space (Ω, r_W) in a classic way, using its covering number $N_\Omega^{(c)}(\epsilon)$ and its Minkowski dimension:

$$dim\, \Omega := \lim_{\epsilon \to 0} \frac{\log \, N_\Omega^{(c)}(\epsilon)}{-\log \epsilon} \tag{5}$$

when the limit exists. See appendix A for additional information about these two standard measures of metric spaces.

Unfortunately, the covering number and the Minkowski dimension of a graphon are not identifiable from the data distribution $\mathbb{P}_{(\Omega,\mu,W)}$. Indeed, they are not robust to changes of the graphon on null-sets, whereas such changes leave the data distribution unaltered (a null-set is a set of zero measure in the probability space (Ω, μ)). This fact is illustrated in the following example where two equivalent graphons (i.e. leading to the same data distributions) have two different Minkowski dimensions. As we can see, this problem is due to the presence of a "big" null-set in Ω.

Example. Let $\Omega := \{2\}$ and $\Omega' := \{2\} \sqcup [0,1]$ be two latent spaces endowed with a common probability distribution μ such that $\mu[\{2\}] = 1$. Let W' be a function defined on $\Omega' \times \Omega'$ such that $W'(x', y') = (x' + y')/3$ for $x', y' \in [0,1]$. Let W be any measurable function on $\Omega \times \Omega$ such that $W(2,2) = W'(2,2)$. Then, the two graphons (Ω, μ, W), (Ω', μ, W') are equivalent, and yet they have two different Minkowski dimensions: $dim\, \Omega = 0$ since $r_W = 0$ on Ω, while $dim\, \Omega' = 1$ since $r_{W'}(x', z') = |x' - z'|/3$ for $x', z' \in [0,1]$. $\square$

3.1 Purification Process for Identifiability

To define an identifiable index of complexity, we need to take care of "big" null-sets (seen in the above example). Usually, these pathological sets are not present in standard representations (Ω, μ, W) and even useless in terms of modeling. Thus, we get rid of them; we do so by using a general remedy, called pure graphon.

Definition (Lovász, 2012, chap.13) *A graphon (Ω, μ, W) is called pure if (Ω, r_W) is a complete separable metric space and the probability measure has full support (that is, every ball of non-zero radius has positive measure). Besides, there is a pure graphon in each equivalence class of graphons.*

For illustrative examples of pure graphons, see Section 3.2. There is no "big" null-set in pure graphons (since their measure μ has full support by definition) and the complexity index takes the same value on the pure graphons of a same equivalence class of $\mathcal{W}/\sim$ (Lemma 3).

Lemma 3 *If two pure graphons are equivalent, then their covering numbers are equal.*

The proof of Lemma 3 is written in Appendix C.2. Lemma 3 directly implies that the Minkowski dimension takes the same value for two equivalent pure graphons. Thus, for any limiting W-random graph distribution $\mathbb{P}^{(\infty)}_{(\Omega,\mu,W)}$, we define its complexity as the covering number and the Minkowski dimension of any pure graphon from the corresponding equivalence class. According to the above lemma, these indices are identifiable from all data distributions $\mathbb{P}^{(n)}_{(\Omega,\mu,W)}$, $1 \leq n \leq \infty$. From now on, we can work exclusively with pure graphons without the loss of generality, since there are pure graphons in each equivalence class of $\mathcal{W}/\sim$. In the remaining of the subsection, we describe two consequences of working with pure graphons.

The metric properties are preserved between equivalent pure graphons (Lemma 4).

Lemma 4 *Let (Ω, μ, W) and (Ω', μ', W') be two pure graphons, endowed with their respective neighborhood distances r_W and $r_{W'}$. If the two graphons are in a same equivalence class of $\mathcal{W}/\sim$, then for some bijective measure-preserving map $\phi : \Omega' \to \Omega$, we have*

$$r_{W'}(x, y) = r_W(\phi(x), \phi(y)) \quad \textit{almost surely on } \Omega' \times \Omega'.$$

Lemma 4 states that the metric spaces (Ω, r_W) and $(\Omega', r_{W'})$ are isometric up to a null-set, it is therefore not surprising that they share the same covering number (Lemma 3). The proof of lemma 4 is written in Appendix C.1. Note that Lemma 4 ensures that the future distance estimation is a well-posed problem.

Another consequence of working with pure graphon is that the sample $\omega_1, \ldots, \omega_n$ is asymptotically dense in Ω. Lemma 5 is proved in Appendix C.3.

Lemma 5 *For a pure graphon (Ω, μ, W) such that $N^{(c)}_{\Omega}(\epsilon) < \infty$ for all $\epsilon > 0$, the sample $\omega_1, \ldots, \omega_n$ is asymptotically dense in the metric space (Ω, r_W). That is, for all radii $\epsilon > 0$, the event*

$$\mathcal{E}(\epsilon) = \{\text{each ball of radius } \epsilon \text{ in } (\Omega, r_W) \text{ contains at least a sampled point } \omega_i\}$$

holds with a probability tending to one as $n \to \infty$.

3.2 Illustrative Examples

We exemplify the complexity index with instances of W-random graphs that are often considered in the literature: a stochastic block model (Holland et al., 1983; Abbe, 2017), a random Hölder graph (Gao et al., 2015; Zhang et al., 2015) and a random geometric graph (Penrose et al., 2003; Arias-Castro et al., 2018; De Castro et al., 2017; Bubeck et al., 2016).

Stochastic Block Model. It produces a structure of community dividing the node set into K subsets of nodes which share a same pattern of connection. More precisely, the

edges are independently sampled from each others, and the probability of an edge between two nodes only depends on their community membership. The SBM with K communities can be written in the framework of the W-random graph model, by setting $\Omega = \{c_1, \dots, c_K\}$, so that each node belongs to one of the K communities c_i, and connects to each other with probability $W(c_i, c_j)$. A natural notion of complexity for SBM is the number K of communities, which coincides with the ϵ-covering number of $\{c_1, \dots, c_K\}$ for small radii ϵ.

Approximation by SBM. In the estimation of W based on the classic approximation by SBM (Gao et al., 2015; Klopp et al., 2017), the right number of communities can be selected using the covering number. Indeed, Proposition 6 states that, for any graphon (Ω, μ, W), the function W can be "$O(\epsilon)$-approximated" in l_2-norm by an SBM with at most $N_\Omega^{(c)}(\epsilon)$ communities. The proof is written in Appendix B.1.

Proposition 6 *Consider any graphon (Ω, μ, W) and its ϵ-covering number $N_\Omega^{(c)}(\epsilon)$, defined in Section 3. There exists a graphon $(\Omega, \mu, \overline{W})$ equivalent to an SBM with $N_\Omega^{(c)}(\epsilon)$ communities, such that,*

$$\int_{\Omega^2} (W(\omega, \omega') - \overline{W}(\omega, \omega'))^2 \mu(d\omega)\mu(d\omega') \leq (4\epsilon)^2.$$

Random Hölder graph. Let $\Omega = [0, 1]^d$ be endowed with the uniform measure, and W fulfill a double Hölder condition:

$$m\big|\big|\omega' - \omega\big|\big|_2^\alpha \leq \big|W(\omega', \omega'') - W(\omega, \omega'')\big| \leq M\big|\big|\omega' - \omega\big|\big|_2^\alpha \tag{6}$$

for some Hölder exponent $\alpha > 0$ (and some constants $m, M > 0$). This means that each node has its specific attribute of d variables, and connects to another node with a probability that smoothly depends on the node attributes. A natural notion of complexity for this graph distribution should increase with the number d of variables, and decrease with the level α of smoothness. This intuitive notion is matched by the Minkowski dimension, which is equal to d/α. See Appendix A for details.

Random geometric graph. It generates simple spatial networks placing nodes in a Euclidean metric space and connecting two nodes if their Euclidean distance is small. Let $\Omega = [0, 1]^d$ be endowed with the uniform measure and the indicator function $W(\omega, \omega') = \mathbb{I}_{||\omega - \omega'||_2 \leq \delta}$ for some constant $\delta > 0$. Appendix A shows that $dim\, \Omega = 2d$. Thus, the Minkowski dimension matches the Euclidean dimension of the latent space, up to a factor 2.

4. Estimation of the Complexity Index

Given a pure graphon (Ω, μ, W), assume a W-random graph is generated from the probability distribution $\mathbb{P}_{(\Omega,\mu,W)}$ defined in Section 2.1. From a single observation of the adjacency matrix A of this graph, we want to estimate the complexity index (introduced in Section 3.1). In particular, the underlying graphon (Ω, μ, W) is unknown, and the sampled points $\omega_1, \dots, \omega_n$ are not observed.

This section is organized in the following manner. We first estimate the neighborhood distance (4) on the sampled points $\omega_1, \dots, \omega_n$. Based on these estimated distances, we then estimate the ϵ-covering number of $(\{\omega_1, \dots, \omega_n\}, r_W)$ by plug-in. Denote by $\widehat{N}_\Omega^{(c)}(\epsilon)$ this

estimator. We finally estimate the Minkowski dimension using $-\log \widehat{N}_{\Omega}^{(c)}(\epsilon)/\log \epsilon$ at a well chosen radius ϵ.

4.1 Distance-estimator

Let us explain the construction of the distance estimator. The l_2-neighborhood distance is naturally associated with a structure of inner product. Given some square-integrable functions f and g on Ω, we write their inner product $\langle f, g\rangle := \int_\Omega f(z)g(z)\mu(dz)$. Let $W(\omega_i, .)$ denote the function $x \mapsto W(\omega_i, x)$, then the neighborhood distance admits the following decomposition

$$r_W^2(\omega_i, \omega_j) \quad = \quad \langle W(\omega_i, .), W(\omega_i, .)\rangle + \langle W(\omega_j, .), W(\omega_j, .)\rangle - 2\langle W(\omega_i, .), W(\omega_j, .)\rangle. \quad (7)$$

We estimate separately the crossed term and the two quadratic terms of (7).

Note A_i the ith row vector of the adjacency matrix A, and $\langle A_i, A_j\rangle_n = \sum_{i=1}^n A_{ik}A_{jk}/n$ the inner product between two such rows. Given ω_i, ω_j, we observe that $\langle A_i, A_j\rangle_n$ is (almost) a sum of i.i.d. random variables (up to a duplicated entry because of the symmetry of the adjacency matrix A). Indeed, the $n-2$ random variables $\{A_{ik}A_{jk} : k \in [n] \text{ and } k \neq i, j\}$ are independent with the same mean conditionally to ω_i, ω_j :

$$\mathbb{E}\left[A_{ik}A_{jk}|\omega_i, \omega_j\right] = \langle W(\omega_i, .), W(\omega_j, .)\rangle$$

where the mean $\mathbb{E}$ is taken over the data distribution $\mathbb{P}_{(\Omega,\mu,W)}$. It is therefore possible to use Hoeffding's inequality to prove that $|\langle A_i, A_j\rangle_n - \langle W(\omega_i, .), W(\omega_j, .)\rangle| \lesssim \sqrt{\log n/n}$ w.h.p. (see Proposition 26 in Appendix D.1). Thus, the inner product between two *different* rows is a consistent estimator of the crossed term $\langle W(\omega_i, .), W(\omega_j, .)\rangle$ in (7).

To estimate the remaining quadratic term $\langle W(\omega_i, .), W(\omega_i, .)\rangle$ in (7), we cannot proceed in the same way since $\frac{1}{n}\langle A_i, A_i\rangle$ is an inconsistent estimator of $\langle W(\omega_i, .), W(\omega_i, .)\rangle$; indeed, we have

$$\mathbb{E}\big[A_{ik}A_{ik}|\omega_i\big] = \mathbb{E}\big[A_{ik}|\omega_i\big] = \langle W(\omega_i, .), 1\rangle \neq \langle W(\omega_i, .), W(\omega_i, .)\rangle.$$

To work around this issue, we simply approximate the quadratic term by a crossed term to be back to the previous case. Specifically, the approximation consists in replacing a sampled point by its nearest neighbor as follows: let $\omega_{m(i)} \in \{\omega_1, \ldots, \omega_n\}$ denote a nearest neighbor of ω_i according to the distance r_W, that is $m(i) \in \operatorname{argmin}_{t:\, t\neq i}\ r_W(\omega_i, \omega_t)$, then we have the following approximation:

$$\begin{aligned} |\big\langle W(\omega_i, .), W(\omega_i, .)\big\rangle - \big\langle W(\omega_i, .), W(\omega_{m(i)}, .)\big\rangle| &= |\big\langle W(\omega_i, .), W(\omega_i, .) - W(\omega_{m(i)}, .)\big\rangle| \\ &\leq r_W(\omega_i, \omega_{m(i)}) \end{aligned} \quad (8)$$

using Cauchy-Schwarz inequality. Thus, the nearest neighbor approximation (8) entails a bias in our estimation procedure, which is equal to the distance between ω_i and its nearest neighbor $\omega_{m(i)}$.

Since the index $m(i)$ is unknown, we define an index estimator $\widehat{m}(i)$ such that $\omega_{\widehat{m}(i)}$ is hopefully close to ω_i according to r_W, and then we use $\langle A_i, A_{\widehat{m}(i)}\rangle_n$ to estimate the quadratic term. Formally, $\widehat{m}(i)$ is a minimizer of the distance function $j \mapsto \widehat{f}(i, j)$ defined by

$$\widehat{f}(i, j) = \max_{k:\, k\neq i,j} |\langle A_k, A_i - A_j\rangle_n| \quad (9)$$

where $\widehat{f}(i,j)$ represents a proxy for the distance between the ith and j^{th} rows of the adjacency matrix, which is enough to define the index estimator

$$\widehat{m}(i) = \operatorname*{argmin}_{j:\, j\neq i} \widehat{f}(i,j). \tag{10}$$

Note that $\widehat{f}(i,j)$ is small in expectation if ω_i and ω_j are close according to the neighborhood distance; indeed, $\mathbb{E}\big[\,\widehat{f}(i,j)|\omega_i,\omega_j,\omega_k\big] = \max_{k\neq i,j} \left|\big\langle W(\omega_i,.) - W(\omega_j,.), W(\omega_k,.)\big\rangle\right| \leq r_W(\omega_i,\omega_j)$ using Cauchy-Schwarz inequality.

Putting together the estimators of the crossed term and the two quadratic terms, we get the following estimator of the square distance $r_W^2(\omega_i,\omega_j)$:

$$\widehat{r}^2(i,j) = \langle A_i, A_{\widehat{m}(i)}\rangle_n + \langle A_j, A_{\widehat{m}(j)}\rangle_n - 2\,\langle A_i, A_j\rangle_n \tag{11}$$

for all $i,j\in[n]$, where $\widehat{m}(i)$ is given by (10).

Remark: The distance-estimator (11) is inspired by the work of Zhang et al. (2015), in which the authors want to recover the expectation of the adjacency matrix A, based on neighborhood smoothing. They rely on the proxy (9) to select neighborhood of points with respect to the neighborhood distance. Restricting themselves on graphons of the form $([0,1],\lambda,W)$ with λ the uniform measure and W a piecewise Lipschitz function, they derive risk bounds for the estimation of W. In contrast, here we do not make any assumption on the graphon, and our objective is to provide an estimator of the neighborhood distance per se.

4.2 Consistency of the Distance-estimator

The statistical recovery of the set of distances $\{r_W(\omega_i,\omega_j) : i,j\in[n]\}$ is a well-posed problem, since the neighborhood distance is invariant on each equivalence class of graphons (Lemma 4). Theorem 7 gives non-asymptotic error bounds for the distance-estimator (11). The proof is written in Appendix D.1.

Theorem 7 *Given any (pure) graphon (Ω,μ,W), consider the data distribution $\mathbb{P}_{(\Omega,\mu,W)}$ defined in model (3). For all $1\leq i\leq n$, let $\omega_{m(i)}\in\{\omega_1,\dots,\omega_n\}\setminus\{\omega_i\}$ denote a nearest neighbor of ω_i according to the distance r_W. Then, for the distance-estimator (11), the event*

$$\mathcal{E}_{dist} = \Big\{\forall i,j\in[n] : \big|r_W^2(\omega_i,\omega_j) - \widehat{r}^2(i,j)\big| \leq 3r_W(\omega_j,\omega_{m(j)}) + 3r_W(\omega_i,\omega_{m(i)}) + 36\sqrt{\log(n)/n}\Big\}$$

holds with probability $\mathbb{P}_{(\Omega,\mu,W)}[\mathcal{E}_{dist}]\geq 1-\frac{2}{n}$.

Theorem 7 implies that the distance-estimator (11) is a consistent estimator of the neighborhood distance (4), provided that the ϵ-covering number is finite for all radii $\epsilon > 0$. Indeed, for a finite covering number, Lemma 5 ensures that the sample $\omega_1,\dots,\omega_n$ is asymptotically dense in (Ω, r_W), which implies that the bias $r_W(\omega_i,\omega_{m(i)})$ is convergent in probability to zero as n grows to infinity.

Let us describe the upper bound of Theorem 7. On the one hand, there is a fluctuation term $\sqrt{\log(n)/n}$ that corresponds to the convergence property of the inner products

between rows of A, i.e.: $|\langle A_i, A_j\rangle_n - \langle W(\omega_i,.), W(\omega_j,.)\rangle| \lesssim \sqrt{\log n/n}$ w.h.p. for $i \neq j$. On the other hand, there is a bias term $r_W(\omega_i, \omega_{m(i)})$ that results from the nearest neighbor approximation (8). Its value depends on the graphon regularity. For instance, in the SBM example of Section 3.2, the bias term $r_W(\omega_i, \omega_{m(i)})$ is equal to zero w.h.p. (indeed, ω_i and its nearest neighbor $\omega_{m(i)}$ are in the same community w.h.p., and thus separated by a distance zero w.r.t. r_W). In the random Hölder graph example, the bias term is of the order of $(\log(n)/n)^{\alpha/d}$ w.h.p..

The next result gives a lower bound matching the upper bound of Theorem 7, up to a numerical constant. Specifically, there exists a graphon (Ω, μ, W_n) for each sample size $n \geq 1$, such that the lower bound holds for (at least) some of the $O(n^2)$ distances.

Theorem 8 *There exists a sequence of graphons $(\Omega, \mu, W_n)_{n\in\mathbb{N}}$ and some numerical constants $p, c, c' > 0$, such that the following holds for any estimator $\widehat{d}$ and any permutation σ of the n indices. With a probability larger than p, the lower bound*

$$\left| r^2_{W_n}(\omega_i, \omega_j) - \widehat{d}^2(\sigma(i), \sigma(j)) \right| \geq c' \left(r_{W_n}(\omega_j, \omega_{m(j)}) + r_{W_n}(\omega_i, \omega_{m(i)}) + \sqrt{\frac{\log n}{n}} \right)$$

is satisfied for (at least) $c\,n$ different pairs (i, j).

The lower bound in Theorem 8 holds regardless of the nodes labels $i \in \{1, \ldots, n\}$ since it is satisfied for any permutation σ of the n indices $\{1, \ldots, n\}$. This is relevant in the model (3) where the data distribution is invariant by relabeling of the nodes.

In particular, Theorem 8 can be written for a graphon $(\Omega, \mu, W_n)_n$ whose bias $r_{W_n}(\omega_i, \omega_{m(i)})$ is equal to a numerical constant, thus giving the next result.

Theorem 9 *There exists a sequence of graphons $(\Omega, \mu, W_n)_{n\in\mathbb{N}}$ and some numerical constants $p, c > 0$, such that the following holds for any estimator $\widehat{d}$ and any permutation σ of the n indices. With a probability larger than p, the lower bound*

$$\left| r^2_{W_n}(\omega_i, \omega_j) - \widehat{d}^2(\sigma(i), \sigma(j)) \right| \geq \frac{1}{800}$$

is satisfied for (at least) $c\,n$ different pairs (i, j).

The constant error in this lower bound does not violate the vanishing error bound in Theorem 7. Indeed, the graphon changes with n in the above lower bounds whereas it remains fixed in the upper bound of Theorem 7.

Although Theorem 8 is a standard way of stating lower bounds in the literature, it is not sufficient to show that the right dependence on the bias $r_W(\omega_i, \omega_{m(i)})$ is linear. For example, if $r_W(\omega_i, \omega_{m(i)})$ is equal to a numerical constant, then one can replace $r_W(\omega_i, \omega_{m(i)})$ with $r^2_W(\omega_i, \omega_{m(i)})$ in Theorem 8 (by adapting the numerical constant c' which was not tight anyway). One can observe that such a change from $r_W(\omega_i, \omega_{m(i)})$ to $r^2_W(\omega_i, \omega_{m(i)})$ is also possible in the case where $r_W(\omega_i, \omega_{m(i)})$ is smaller than the fluctuation term $\sqrt{\log(n)/n}$.

Therefore, neither the lower bound of Theorem 8 nor the lower bound of Theorem 9 is enough informative to decipher whether the optimal rate is of the order of $r_W^2(\omega_i, \omega_{m(i)}) + \sqrt{\log(n)/n}$ or $r_W(\omega_i, \omega_{m(i)}) + \sqrt{\log(n)/n}$.

A quadratic dependency $r_W^2(\omega_i, \omega_{m(i)})$ would improve on the linear dependency since the neighborhood distance r_W is always smaller than 1 by definition. Hence, one may wonder whether such an improvement is possible, especially that Theorem 7 is an error bound on *square* distances, $\left|r_W^2(\omega_i, \omega_j) - \widehat{d}^2(i,j)\right|$. It turns out that even replacing the bias $r_W(\omega_i, \omega_{m(i)})$ with $r_W^{1+\gamma}(\omega_i, \omega_{m(i)})$ for some $\gamma > 0$ is impossible. In other words, no estimator $\widehat{d}$ simultaneously satisfies the following inequalities, with high probability, over all graphons (Ω, μ, W):
$\forall i, j \in [n]$:

$$\left|r_W^2(\omega_i, \omega_j) - \widehat{d}^2(i,j)\right| \leq C\left(r_W^{1+\gamma}(\omega_j, \omega_{m(j)}) + r_W^{1+\gamma}(\omega_i, \omega_{m(i)}) + \sqrt{\log(n)/n}\right) \tag{12}$$

where C is a numerical constant. Indeed, Theorem 10 ensures that the uniform bound (12) cannot be achieved by any estimator $\widehat{d}$ as soon as $\gamma > 0$, thus proving that the bias-dependence is linear ($\gamma = 0$) at best.

Theorem 10 *There exists a sequence of graphons $(\Omega, \mu, W_n)_{n\in\mathbb{N}}$ and some numerical constants $p > 0$, $\gamma \geq 0$ and $c > 0$, such that the following holds for any estimator $\widehat{d}$ and any permutation σ of the n indices. With a probability larger than p, the following lower bound is satisfied for (at least) cn different pairs (i,j),*

$$\left|r_{W_n}^2(\omega_i, \omega_j) - \widehat{d}^2(\sigma(i), \sigma(j))\right| \gtrsim \kappa_n(\gamma)\left(r_{W_n}^{1+\gamma}(\omega_j, \omega_{m(j)}) + r_{W_n}^{1+\gamma}(\omega_i, \omega_{m(i)}) + \sqrt{\frac{\log n}{n}}\right)$$

where $\kappa_n(\gamma) \underset{n}{\longrightarrow} \infty$ as soon as $\gamma > 0$, and $\kappa_n(0) = O(1)$.

Theorem 8 is a direct consequence of Theorem 10 for $\gamma = 0$. The proof of Theorem 10 can be found in Appendix D.2.

In conclusion, the upper bounds in Theorem 7 have the right dependence on the bias.

4.2.1 Discussion about the Distance Estimation and Open Questions

The lower bound in Theorem 8 only holds for $O(n)$ distances among the $O(n^2)$ ones. As to whether this could be extended to $O(n^2)$ distances, the question is left open. This generalization would confirm that the dependence on the bias should be linear and thus that the upper bounds of Theorem 7 have the optimal bias-dependence.

This gap may come from technical shortcomings of our proof, rather than an inherent feature of the estimation problem. Indeed, the instance of W-random graph constructed for the lower bounds, has only one outlier node whose $O(n)$ distances are incorrectly estimated. A possible difficulty for generalizing the proof to $O(n^2)$ distances is the randomness of the sample $\omega_1, \ldots, \omega_n \overset{iid}{\sim} \mu$ and the symmetry of the matrix A, as they bring regularization effects to the problem.

Another element suggesting that the lower bounds in Theorem 8 may not be specific to one outlier, is that the upper bounds in Theorem 7 are not sensitive to $O(1)$ nodes. Indeed, by reading the proof of Theorem 7, one can observe that the upper bounds will not improve after removing $O(1)$ nodes.

Note that our lower bound has probably no direct consequences on graphon estimation since the bound only hinges on $O(1)$ outlier nodes, which would be negligible in the estimation of the probabilities matrix $[W(\omega_i, \omega_j)]_{1\leq i,j\leq n}$ with respect to the Frobenius norm.

The fluctuation term $f_n := \sqrt{\log(n)/n}$ in the upper bounds of Theorem 7 is natural, as each inner product is computed with $\Theta(n)$ entries, leading to an error $1/\sqrt{n}$, and the $\sqrt{\log(n)}$ allows us to show high probability bounds for each of the $\Theta(n^2)$ distances. However, this explanation is specific to our estimator, and the optimality of f_n with respect to all estimators remains to be proved. The fact that f_n appears in the lower bound of Theorem 8 does not prove that f_n is inevitable since f_n is actually negligible compared to the bias term $r_{W_n}(\omega_i, \omega_{m(i)})$ in this lower bound. Indeed, Theorem 8 is based on an instance of graphon W_n satisfying $r_{W_n}(\omega_i, \omega_{m(i)}) \gtrsim f_n$.

4.3 Consistency of the Covering Number Estimator

We have defined the ϵ-covering number estimator $\widehat{N}_{\Omega}^{(c)}(\epsilon)$ as the covering number of the set $\{1, \ldots, n\}$ w.r.t. the distance-estimator $\widehat{r}$. Consider e_{sup} the supremum of the errors of $\widehat{r}$:

$$e_{sup} := \sup_{i,j\in[n]} \left| r_W(\omega_i, \omega_j) - \widehat{r}(i,j) \right| .$$

Then, the covering number estimator is linked with the true covering number of $\{\omega_1, \ldots, \omega_n\}$ by the following inequalities

$$\forall \epsilon > e_{sup}, \qquad N^{(c)}_{\omega_1,\ldots,\omega_n}(\epsilon + e_{sup}) \leq \widehat{N}^{(c)}_{\Omega}(\epsilon) \leq N^{(c)}_{\omega_1,\ldots,\omega_n}(\epsilon - e_{sup}) .$$

To compare the covering numbers of $\{\omega_1, \ldots, \omega_n\}$ and Ω, we need to measure the difference between the sample $\omega_1, \ldots, \omega_n$ and the space Ω. We do so by introducing the sampling error s_ω defined as

$$s_\omega = \sup_{\omega\in\Omega} \inf_{i\in\{1,\ldots,n\}} r_W(\omega, \omega_i) \tag{13}$$

which is the greatest distance that separates a point of Ω from the set $\{\omega_1, \ldots, \omega_n\}$. Thus, the covering numbers (w.r.t. the true distance r_W) of $\omega_1, \ldots, \omega_n$ and Ω are linked by the following inequalities

$$\forall \epsilon > s_\omega, \qquad N^{(c)}_{\Omega}(\epsilon + s_\omega) \leq N^{(c)}_{\omega_1,\ldots,\omega_n}(\epsilon) \leq N^{(c)}_{\Omega}(\epsilon - s_\omega) .$$

Finally, for

$$b^2_{sup} := 6 \sup_{i\in[n]} r_W(\omega_i, \omega_{m(i)}) + 36\sqrt{\log(n)/n}, \tag{14}$$

Theorem 7 ensures that $e_{sup} \leq b_{sup}$ with probability at least $1 - 2/n$. From the above displays, we obtain the following proposition.

Proposition 11 *Given any (pure) graphon (Ω, μ, W), consider the data distribution $\mathbb{P}_{(\Omega,\mu,W)}$ defined in model (3). Let b_{sup} and s_ω be the distance error bound* (14) *and the sampling error (13). Then, the estimator $\widehat{N}_\Omega^{(c)}$ satisfies the following non-asymptotic bounds*

$$\forall \epsilon > b_{sup} + s_\omega, \qquad N_\Omega^{(c)}\left(\epsilon + b_{sup} + s_\omega\right) \leq \widehat{N}_\Omega^{(c)}\left(\epsilon\right) \leq N_\Omega^{(c)}\left(\epsilon - b_{sup} - s_\omega\right) \tag{15}$$

with probability at least $1 - \frac{2}{n}$ according to the distribution $\mathbb{P}_{(\Omega,\mu,W)}$.

As a result, we have a consistent estimation of the ϵ-covering number for almost every ϵ, provided that the covering number is finite for all radii. Indeed, if $N_\Omega^{(c)}(\epsilon) < \infty$ for all $\epsilon > 0$, then the sample $\omega_1, \dots \omega_n$ is asymptotically dense in (Ω, r_W) by Lemma 5, which implies that b_{sup} and s_ω converge in probability to zero; Then, taking the limit $n \to \infty$ in (15), one has the convergence in probability of $\widehat{N}_\Omega^{(c)}$ towards $N_\Omega^{(c)}(\epsilon)$, for all ϵ where the step function $\epsilon \mapsto N_\Omega^{(c)}(\epsilon)$ is continuous (i.e., for almost every ϵ).

4.4 Consistency of the Dimension Estimator

We estimate the Minkowski dimension of (Ω, r_W) using the data-function $-\log \widehat{N}_\Omega^{(c)}(\epsilon) / \log \epsilon$ at a well chosen radius ϵ. The following observation makes it clear that each graphon requires a specific choice of radius, and thus no (universal) radius is suited for all graphons.

> Observation. (1) at very small scale (i.e. very small ϵ), the covering number may just count the points of the sample $\omega_1, \dots, \omega_n$ and the data look zero-dimensional; (2) if the scale is comparable to the noise due to the distance estimation, the covering number estimator $\widehat{N}_\Omega^{(c)}(\epsilon)$ is not reliable; (3) for an intermediate scale, it is possible to have a good estimation of the dimension, as we shall see in Theorem 12; (4) at very big scale, the apparent geometry may not reflect the Minkowski dimension (which is, by definition, a measure of the complexity at infinitesimal scale).

Hence, we consider a subset of graphons for which there exists a radius that is well-suited for dimension estimation. We sometimes denote $dim\,\Omega$ by d for brevity, and write $B(\omega, \epsilon)$ the ball of center $\omega \in \Omega$ with radius ϵ (w.r.t. the neighborhood distance). Given constants $D, v, \alpha > 0$ and $M \geq 1 \geq m > 0$, we define the set $\mathcal{W}(D, \alpha, m, M, v)$ of all (pure) graphons (Ω, μ, W) satisfying

1. $dim\,\Omega \leq D$.
2. For $dim\,\Omega := d$ and all $\epsilon \in (0, v]$,

$$\alpha\,\epsilon^d \;\leq\; \mu\Big[B(\omega, \epsilon)\Big] \tag{$H_1^{\alpha,v}$}$$

$$m\epsilon^{-d} \;\leq\; N_\Omega^{(c)}(\epsilon) \;\leq\; M\epsilon^{-d}. \tag{$H_2^{m,M,v}$}$$

The assumption $H_2^{m,M,v}$ links the covering number with the Minkowski dimension of the graphon. The condition $H_1^{\alpha,v}$ enforces a minimal measure for each ball of (Ω, r_W); in

particular, it strengthens the non-zero measure of balls of pure graphons, seen in Section 3.1. Mention can be made of the problem of recovery of the dimension of a manifold, where similar hypotheses are often considered (see Koltchinskii, 2000, for example). Besides, $H_1^{\alpha,v}$ may be seen as a small-ball condition used in learning problems (Mendelson, 2014; Lecué et al., 2018).

With the radius

$$\epsilon_D \asymp \left(\frac{\log n}{n}\right)^{1/(4\vee 2D)} \tag{16}$$

we consistently estimate the Minkowski dimension (Theorem 12) using the following estimator

$$\widehat{dim}_D := \frac{\log \widehat{N}_\Omega^{(c)}(\epsilon_D)}{-\log \epsilon_D}. \tag{17}$$

Theorem 12 *For all graphons (Ω, μ, W) in $\mathcal{W}(D, \alpha, m, M, v)$ and all large enough n, we have*

$$\left| \widehat{dim}_D - dim\, \Omega \right| \leq \frac{C(D, \alpha, m, M)}{\log n}$$

with probability at least $1 - C'(\alpha, M)/n$ w.r.t. the distribution $\mathbb{P}_{(\Omega,\mu,W)}$, and for some constants $C'(\alpha, M)$ and $C(D, \alpha, m, M)$ that are independent of n.

Theorem 12 is a corollary of Theorem 35 in Appendix E.1, which gives a non-asymptotic high probability bound for $-\log \widehat{N}_\Omega^{(c)}(\epsilon)/\log \epsilon$ at any radius ϵ.

One can observe that the convergence rate $\log^{-1} n$ of Theorem 12 is optimal, in the sense that faster convergence rates cannot be achieved by any estimator of the form $\log \widehat{N}_\Omega^{(c)}(\hat{\epsilon})/-\log \hat{\epsilon}$ [1]. To see it, take a graphon of dimension $d > 1$ with covering number $N_\Omega^{(c)}(\epsilon) = m\epsilon^{-d}$ for some constant $m > 1$. Even if there exists a covering number estimator that gives a perfect estimation, i.e. $\widehat{N}_\Omega^{(c)} = N_\Omega^{(c)}$, this still entails an error for the dimension estimation. Indeed, in such a case we have:

$$\left| \frac{\log \widehat{N}_\Omega^{(c)}(\epsilon)}{-\log \epsilon} - d \right| = \frac{\log m}{-\log \epsilon}$$

which is (at least) of the order $\log^{-1} n$ since the radius ϵ cannot be taken smaller than n^{-1} in general (otherwise, the estimator of the covering number may just count the n sampled points). Thus, the convergence rate $\log^{-1} n$ is optimal for the classical method of estimation of the Minkowski dimension, which is based on the the plug-in of a covering number estimate into formula (5).

Next we show that no estimator [2] can improve on the error bound $\log^{-1} n$, over the following sequence of sets. Given $n > 0$, let $\mathcal{W}_n(D, \alpha, m, M, v)$ be the class of all (pure)

1. $\widehat{N}_\Omega^{(c)}$ is any consistent estimator of the covering number, and $\hat{\epsilon}$ is any estimator of a "well chosen radius".
2. An estimator is defined as a function of the adjacency matrix $A \in \{0,1\}^{n\times n}$.

graphons fulfilling $H_1^{\alpha,v}$ and $H_2^{m,M,v}$ for all $\epsilon \in (1/n, v]$ instead of $\epsilon \in (0, v]$, so $\mathcal{W}_n(D, \alpha, m, M, v)$ is a subset of $\mathcal{W}(D, \alpha, m, M, v)$. On this sequence of sets, one can readily extend Theorem 12 and retrieve the same error bound, using the same estimator (17). This means that there exist some constants $C(D, \alpha, m, M)$ and $C'(\alpha, M)$ that are independent of n, such that for all graphons in $\mathcal{W}_n(D, \alpha, m, M, v)$ and all large enough n, the following error bound holds

$$\left| \widehat{dim}_D - dim\,\Omega \right| \leq \frac{C(D, \alpha, m, M)}{\log n} \tag{18}$$

with probability at least $1 - C'(\alpha, M)/n$. Then, Theorem 13 shows that no estimator can improve on the (order of the) bound (18). The proof is written in Appendix E.2.2.

Theorem 13 *For any $D > 2$, some numerical constants $\alpha, m, M, v > 0$ and all large enough n, we have*

$$\inf_{\widehat{d}} \sup_{\mathcal{W}_n(D,\alpha,m,M,v)} \mathbb{P}_{(\Omega,\mu,W)} \left[|\hat{d} - dim\,\Omega| \geq \frac{1}{2\log(n)} \right] \geq \frac{1}{4}$$

where $\inf_{\widehat{d}}$ *is the infimum over all estimators.*

Let us discuss the minimal aspect of the conditions defining $\mathcal{W}_n(D, \alpha, m,\ M, v)$. First, the assumption that the dimension is upper bounded seems natural, as our available data $A \in \{0,1\}^{n\times n}$ is a finite set. Indeed, for metric spaces (Ω_n, r_{W_n}) with arbitrary large dimensions (like $dim\,\Omega_n/n \to \infty$ for instance), a finite sample $\omega_1, \ldots, \omega_n$ may look like a set of distant and isolated points, which does not reflect the true geometry of (Ω_n, r_{W_n}). Since this situation is not conducive to accurate estimates of the complexity of Ω_n, we avoid it by assuming the dimension is upper bounded. Second, we show that the assumptions $H_1^{\alpha,v}$ and $H_2^{m,M,v}$ are minimal, in the sense that, removing any one of them entails a large loss for any estimator. Specifically, let $\mathcal{W}_n^{min(j)}(D, \alpha, m, M, v)$ be the collection of all (pure) graphons satisfying all conditions of the set $\mathcal{W}_n(D, \alpha, m, M, v)$ except the condition H_j (where H_j denotes $H_1^{\alpha,v}$ or $H_2^{m,M,v}$ according to the value of $j \in \{1, 2\}$). Then, Theorem 14 shows that any estimator suffers from an error of the order D, over the class $\mathcal{W}_n^{min(j)}(D, \alpha, m, M, v)$. The proof is written in Appendix E.2.2.

Theorem 14 *For any $D > 2$, some numerical constants $\alpha, m, M, v > 0$, all $j \in \{1, 2\}$ and all large enough n, we have*

$$\inf_{\widehat{d}} \sup_{\mathcal{W}_n^{min(j)}(D,\alpha,m,M,v)} \mathbb{P}_{(\Omega,\mu,W)} \left[|\hat{d} - dim\,\Omega| \geq \frac{D}{2} \right] \geq \frac{1}{4}$$

where $\inf_{\widehat{d}}$ *is the infimum over all estimators.*

Remark: our optimal rate of estimation may seem at odds with the faster rates of convergence in the literature about intrinsic dimension estimation, see (Kim et al., 2016)

for instance. This is due to the important differences in the modeling assumptions. In the work of Kim et al. (2016), for example, the observed data are n i.i.d. sampled points from a well-behaved manifold in $\mathbb{R}^m$ whose dimension is an integer. In contrast, here we do not assume the dimension is an integer, we do not observe the n sampled points $\omega_1, \ldots, \omega_n$, and do not know the metric r_W.

Comments on $H_1^{\alpha,v}$, $H_2^{m,M,v}$: we only make the assumptions $H_1^{\alpha,v}$, $H_2^{m,M,v}$ at a small scale, that is for $\epsilon \in (0, v]$. Besides, the right hand side of $H_2^{m,M,v}$ is almost free since it is already implied by $H_1^{\alpha,v}$ for $M = 2^d/\alpha$. Let us briefly explain how these assumptions imply the error bound of Theorem 12. The assumption $H_1^{\alpha,v}$ ensures that the difference between the sampled points $\omega_1, \ldots, \omega_n$ and the latent space Ω is not too large. By definition, this implies that the sampling error (13) and the distance error (14) are small. Accordingly, we can choose a radius ϵ_D that is larger than these two errors, and reliably estimate the ϵ_D-covering number $N_\Omega^{(c)}(\epsilon_D)$ by Proposition 11. Then, we use a plug-in to estimate the quantity $-\log N_\Omega^{(c)}(\epsilon_D)/\log \epsilon_D$, which is a good approximation of the dimension by assumption $H_2^{m,M,v}$. To sum up, the radius ϵ_D must be larger than the sampling and distance errors, but still small enough to well approximate the Minkowski dimension with $-\log N_\Omega^{(c)}(\epsilon_D)/\log \epsilon_D$.

5. Testing the Complexity

Given the adjacency matrix of a W-random graph, we want to known if the graph is simple or complex. In other words, we would like to test the null-hypothesis $N_\Omega^{(c)}(\epsilon) \leq K$ for a given $K > 0$, with a specific care for minimizing the assumptions on the graphon. However, instead of using the covering number we use the packing number $N_\Omega^{(p)}(\epsilon)$ for some reasons to be specified in Section 5.1. For now, note that it is essentially the same measure as the covering number, and all previous results of the paper can be adapted to the packing number (without any significant difference). See Appendix A for a reminder of this usual measure for metric spaces.

In hypothesis testing, it is common to be conservative and focus on the minimization of the type I error, which is the probability of rejecting the null-hypothesis incorrectly. Accordingly, our objective is to control the type I error without any assumption on the graphon, while keeping a control of the type II error under reasonable assumptions. (the type II error is the probability of accepting the null-hypothesis incorrectly)

5.1 Testing the Null-hypothesis without Assumption on the Graphon, via Under-estimation of the Packing Number

To test the null-hypothesis without assumption on the graphon, we want to define a complexity estimator that does not overestimate the true complexity w.h.p.. Unfortunately, the inequality on the covering number estimator from Proposition 11

$$\widehat{N}_\Omega^{(c)}\left(\epsilon + b_{sup} + s_\omega\right) \leq N_\Omega^{(c)}\left(\epsilon\right)$$

is difficult to leverage for an under-estimation since the errors b_{sup} and s_ω are unknown and take specific values for each graphon. However, we show below that the sampling error

s_ω can be removed, by working with the packing number instead of the covering number. Then, we show that the distance error bound b_{sup} can be handled with a slight modification of the distance-estimator $\widehat{r}$, defined earlier by (11).

Based on the distance estimator $\widehat{r}$, we can define a plug-in estimator $\widehat{N}_\Omega^{(p)}(\epsilon)$ of the packing number, as we did for the covering number estimator. This estimator satisfies almost the same non-asymptotic bounds as the covering number estimator, see the following proposition, which is a slight variant of Proposition 11. The proof is omitted.

Proposition 15 *Given any graphon (Ω, μ, W), consider the data distribution $\mathbb{P}_{(\Omega,\mu,W)}$ defined in model (3). Let b_{sup} and s_ω be the distance error bound (14) and the sampling error (13). Then, the packing number estimator $\widehat{N}_\Omega^{(p)}$ satisfies the following inequalities*

$$\forall \epsilon > b_{sup}, \qquad N_\Omega^{(p)}\left(\epsilon + b_{sup} + 2s_\omega\right) \leq \widehat{N}_\Omega^{(p)}\left(\epsilon\right) \leq N_\Omega^{(p)}\left(\epsilon - b_{sup}\right)$$

with probability at least $1 - \frac{2}{n}$ with respect to the distribution $\mathbb{P}_{(\Omega,\mu,W)}$.

Hence, we have

$$\widehat{N}_\Omega^{(p)}\left(\epsilon + b_{sup}\right) \leq N_\Omega^{(p)}\left(\epsilon\right)$$

without the sampling error s_ω anymore.

The next step is to control the remaining error term b_{sup} coming from the estimator $\widehat{r}$. Obviously, when b_{sup} is positive and relatively large, the estimator $\widehat{r}$ over-estimates some distances, and thus the plug-in estimator $\widehat{N}_\Omega^{(p)}$ based on $\widehat{r}$ may over-estimate the packing number. In order to avoid such an over-estimation, we modify the previous estimator $\widehat{r}$, by taking a maximum in front of the negative term involved in (the expression of) $\widehat{r}^2$:

$$\widehat{r}^2_{new}(i,j) := \left[\langle A_i, A_{\widehat{m}(i)}\rangle_n + \langle A_j, A_{\widehat{m}(j)}\rangle_n - 2 \max_{k\in\{i,\widehat{m}(i)\}, l\in\{j,\widehat{m}(j)\}} \langle A_k, A_l\rangle_n\right]_+ \tag{19}$$

We show in Appendix G.1 that the new distance estimator $\widehat{r}_{new}$ satisfies the same upper bound as $\widehat{r}$ in Theorem 7 (up to a numerical constant 5/3) but also successfully under-estimates the neighborhood distance r_W in some sense (see Lemma 42 for details). Then, defining a new packing number estimator based on $\widehat{r}_{new}$, which is denoted by $\widehat{N}_\Omega^{(p.new)}$ in the following, we finally get the wanted under-estimation of the packing number, see Theorem 16 below. The proof is written in Appendix G.1.

Theorem 16 *Given any graphon (Ω, μ, W), consider the data distribution $\mathbb{P}_{(\Omega,\mu,W)}$ defined in model (3). Then, for the radius $\widehat{\epsilon} = \sqrt{\epsilon^2 + t_n}$ with $t_n = 12\sqrt{\frac{\log n}{n}}$, the estimator $\widehat{N}_\Omega^{(p.new)}$ satisfies the following inequalities*

$$\forall \epsilon > 0, \qquad N_\Omega^{(p)}\left(\widehat{\epsilon} + \frac{5}{3}b_{sup} + 2s_\omega\right) \leq \widehat{N}_\Omega^{(p.new)}(\widehat{\epsilon}) \leq N_\Omega^{(p)}(\epsilon) \tag{20}$$

with probability at least $1 - \frac{2}{n}$ with respect to the distribution $\mathbb{P}_{(\Omega,\mu,W)}$.

Thus, without any assumption on the graphon, the estimator $\widehat{N}_\Omega^{(p.new)}(\widehat{\epsilon})$ does not overestimate the ϵ-packing number with high probability. Besides, the left hand side of (20) shows that it does not under-estimate (significantly) more than the previous estimator $\widehat{N}_\Omega^{(p)}$ of the packing number (seen in Proposition 15).

5.2 Results on the Packing Number Test

We accept the null hypothesis if and only if $\widehat{N}_{\Omega}^{(p.new)}(\widehat{\epsilon}) \leq K$. The upper bound (20) ensures that the type I error is controlled for all graphons, which gives the following result.

Corollary 17 *For any graphon, the type I error is lower than $\frac{2}{n}$ with respect to the distribution $\mathbb{P}_{(\Omega,\mu,W)}$.*

By definition of the packing number, the type II error is small as soon as $K+1$ sampled points are separated by at least a distance $\widehat{\epsilon} + err$, where err upper bounds all errors of distance estimation between the $K+1$ points. This condition on the sampled points is satisfied w.h.p. by each of the following graphons.

Given two parameters $\eta > 0$ and $\beta > 1/n$, let $\mathcal{W}(\eta, \beta)$ denote a collection of graphons for which there exist $K+1$ balls $B(x_1, \eta_1), \ldots, B(x_{K+1}, \eta_{K+1})$ in (Ω, r_W) such that

1. the $K+1$ balls are weighted enough: $\mu\left[B(x_i, \eta_i)\right] \geq \beta$ for all $i \in [K+1]$,
2. the radii are small enough: $\eta_i \leq \eta/2$ for all $i \in [K+1]$,
3. the centers are spaced enough: $r_W(x_i, x_j) \geq \sqrt{\epsilon^2 + 10\eta + 6t_n} + \eta$.

The small-ball condition 1. is similar to the assumption $H_1^{\alpha,v}$ for the dimension estimation; it ensures that some of the sampled points $\omega_1, \ldots, \omega_n$ belong to the $K+1$ balls w.h.p.. The third condition 3. ensures that these balls are enough distant from each other, so that the sampled points in these balls are separated enough, in order to have $\widehat{N}_{\Omega}^{(p.new)}(\widehat{\epsilon}) \geq K+1$ and confirm the alternative hypothesis correctly.

Theorem 18 *Assume the graphon (Ω, μ, W) belongs to $\mathcal{W}(\eta, \beta)$ for some $\beta > 1/n$. Then, the type II error is smaller than*

$$\frac{2}{n} + 2\,\beta n(K+1)\exp[-\beta(n-1)]$$

with respect to the distribution $\mathbb{P}_{(\Omega,\mu,W)}$.

The proof of Theorem 18 is written in Appendix G.2.2. This result implies that, for any graphon in $\mathcal{W}(\eta, \beta)$, the type II error is convergent to zero as soon as the measure of each ball $B(x_i, \eta_i)$ is large enough to satisfy $\beta \gtrsim n^{-1}$. For example, if each of the $K+1$ balls has a measure that is larger than $\log[Kn]/n$, then the type II error is smaller than $\log(n)/n$ up to some numerical constant. In Appendix A.3, Theorem 18 is improved by using the graphon regularity at a finer level (see Theorem 21).

6. Further Considerations

6.1 Estimation of the Complexity with Sparse Observations

In the W-random graph model (3), each node has an average degree that is linear with n the total number of nodes. However, real-world networks are often sparse with node degrees varying from zero to n. This motivates to consider a model of sparse graph where the node degree can be an order of magnitude smaller than n.

Given a sequence ρ_n such that $\rho_n \to 0$, the definition of model (3) can be modified to have average node degrees of the order of $\rho_n n$. Consider the adjacency matrix A, defined by model (3), whose edges are independently retained with probability ρ_n and erased with probability $1-\rho_n$. We refer to this set-up as "the sparse setting" and denote by $\mathbb{P}_{(\Omega,\mu,W),\rho_n}$ the corresponding data distribution. This model has been considered several times in the literature (see Bickel et al., 2011; Wolfe and Olhede, 2013; Klopp et al., 2017; Xu et al., 2014).

We now extend the results of Section 4 to this sparse setting. Corollary 19 gives non-asymtotic error bounds for the distance estimation. It is a slight variant of Theorem 7. For completeness, the proof is written in Appendix F.1.

Corollary 19 *Assume the scaling parameter ρ_n is lower bounded by*

$$\rho_n \geq 2\sqrt{\log(n)/(n-2)}. \tag{21}$$

Then, the following event

$$\mathcal{E}^{sp}_{dist} = \Big\{ \forall i,j \in [n] : \big|\rho_n^2 r_W^2(\omega_i,\omega_j) - \widehat{r}^2(i,j)\big| \\ \leq 3\rho_n \Big(\rho_n r_W(\omega_j,\omega_{m(j)}) + \rho_n r_W(\omega_i,\omega_{m(i)}) + 20\sqrt{\log(n)/n} \Big) \Big\}$$

holds with probability $\mathbb{P}_{(\Omega,\mu,W),\rho_n}\left(\mathcal{E}^{sp}_{dist}\right) \geq 1 - \frac{2}{n}$.

We estimate the Minkowski dimension using the following radius

$$\epsilon_{D,\rho_n} \asymp \left(\frac{\log n}{n}\right)^{1/(2D)} \vee \rho_n^{-1/2}\left(\frac{\log n}{n}\right)^{1/4} \tag{22}$$

Corollary 20 is an adaptation of Theorem 12 for the sparse setting. The proof is written in Appendix F.2.

Corollary 20 *For all graphons (Ω,μ,W) in $\mathcal{W}(D,\alpha,m,M,v)$, all scaling parameters ρ_n fulfilling* (21)*, and all radii satisfying* (22)*, the following rate of estimation of the dimension holds with probability tending to* 1 *as $n \to \infty$ (w.r.t. the distribution $\mathbb{P}_{(\Omega,\mu,W),\rho_n}$).*

$$\left| \frac{\log \widehat{N}^{(c)}_{\Omega}(\epsilon_{D,\rho_n})}{-\log \epsilon_{D,\rho_n}} - d \right| \leq \; C(D,\alpha,m,M,t) \begin{cases} 1 & \text{if } \rho_n \asymp \sqrt{\log(n)/n}, \\ (\log n)^{-1} & \text{if } \rho_n \asymp (\log(n)/n)^{(1/2)-t}, \end{cases}$$

where $t \in (0,1/2)$ and $C(D,\alpha,m,M,t)$ is some constant independent of n.

6.1.1 Discussion about the Distance Estimation in Sparse Regimes and Open Question

Corollary 19 extends the upper bounds to the sparse setting where each entry of the adjacency matrix $[A_{ij}]_{1\leq i,j\leq n}$ is observed with probability ρ_n independently. The scaling

parameter is assumed to satisfy $\rho_n \gtrsim \sqrt{\log(n)/n}$ where the $\sqrt{\log(n)}$ in the numerator allows us to show high probability concentration, and the $1/\sqrt{n}$ ensures that there are enough data for the distance estimator to work. Indeed, the distance estimator is a linear combination of scalar products, and most scalar products are equal to zero in the very sparse regime $\rho_n = o(\sqrt{1/n})$ where pairs of nodes tend to have no common neighbors (i.e. common non-zero entries in the adjacency matrix).

Therefore, one needs to find new strategies to estimate the distances in very sparse regimes. A direction of research could be, for example, the nearest neighbor method in (Borgs et al., 2017) which is successfully applied to matrix estimation when $\rho_n \gtrsim n^{-1}\log(n)^2$. Overall, it would be interesting to complete the picture of the distance estimation problem according to different regimes of sparsity.

6.2 Polynomial-time Algorithm (with Theoretical Guarantee)

In contrast with the previous sections, here we take into account the computational aspect of the problem. Computing the covering number of a finite set is NP-hard, hence we approximate it with a greedy algorithm (Chvatal, 1979).

For completeness, the polynomial-time procedure for estimating $N_\Omega^{(c)}(\epsilon)$ is described below. The algorithm proceeds in two steps: Step 1 computes all distances $\widehat{r}(i,j)$ using the distance-estimator (11); in particular, this step requires the computation of all index estimators $\widehat{m}(j)$ defined by (10). Step 2 approximates the ϵ-covering number of $\{1,\dots,n\}$ w.r.t. the distance estimator $\widehat{r}$, by sequentially selecting balls (of radius ϵ) according to one rule: at each stage, select the ball that contains the largest number of uncovered elements. At the end of the process, the number of selected balls is returned. This output is denoted by $\widehat{N}_\Omega^{(ap.c)}(\epsilon)$.

Covering Number Algorithm

Input: $A = [A_{ij}]$ adjacency matrix of size $n \times n$, a radius ϵ.

Step 1 : constructing the distance-estimator $\widehat{r}$

1. Compute the nearest neighbor's index of each sampled point ω_i:
$\forall i \in \{1, \dots, n\}, \qquad \widehat{m}(i) = \underset{j:\ j \neq i}{\operatorname{argmin}} \ \max_{k:\ k \neq i,j} \left| \langle A_k, A_i - A_j \rangle_n \right|.$

2. Compute all the distances:
$\forall i, j \in \{1, \dots, n\}, \quad \widehat{r}(i,j) = \langle A_i, A_{\widehat{m}(i)} \rangle_n + \langle A_j, A_{\widehat{m}(j)} \rangle_n - 2 \langle A_i, A_j \rangle_n.$

Step 2 : computing an approximation of the ϵ-covering number

3. In the space $\mathcal{S}_0 = \{1, \dots, n\}$ endowed with the distance function $\widehat{r}$, consider $\mathcal{B}_0 = \{B_j\}_{j \leq n}$ the set of all the balls of radius ϵ.

4. Obtain a cover of $\{1, \dots, n\}$ as follows:
Set $i = 0$. While $\mathcal{S}_i \neq \emptyset$, do:

   (a) Select a ball B in $\mathcal{B}_i$ that contains the largest number of elements of $\mathcal{S}_i$.

   (b) Set $\mathcal{S}_{i+1} = \mathcal{S}_i \setminus B$ to remove the elements covered by B,

   (c) Set $\mathcal{B}_{i+1} = \mathcal{B}_i \setminus \{B\}$ to update the set of available balls,

   (d) Set $i = i + 1$ to continue the algorithm.

Output: the number i of selected balls, denoted by $\widehat{N}_{\Omega}^{(ap.c)}(\epsilon)$.

We also suggest an heuristic for tuning ϵ in the estimation of the Minkowski dimension. First, run several times Covering Number Algorithm for a range of different radii $\epsilon_1, \dots, \epsilon_t$, and then plot $\log \widehat{N}_{\Omega}^{(ap.c)}(\epsilon_j) / \log \epsilon_j$ for $j = 1, \dots, t$. As in Figure 1, we look for a graph function that (roughly) admits the three following parts: (1) for big radii, the shape of the curve is irregular and seems sawtooth; (2) for medium radii, there is almost a plateau whose value is the dimension estimate; (3) for small radii, there is an abrupt drop towards zero.

According to the theoretical guarantee of the greedy algorithm (Chvatal, 1979), one has

$$\widehat{N}_{\Omega}^{(c)}(\epsilon) \leq \widehat{N}_{\Omega}^{(ap.c)}(\epsilon) \leq 2 \log(n) \widehat{N}_{\Omega}^{(c)}(\epsilon)$$

where $\widehat{N}_{\Omega}^{(c)}(\epsilon)$ is the consistent estimator introduced in Section 4. Then, for graphons fulfilling the assumptions of Theorem 12, there exist some radii ϵ such that $-\log \widehat{N}_{\Omega}^{(ap.c)}(\epsilon) / \log \epsilon$ is close to the Minkowski dimension up to a small error term $-\log \big(2 \log(n)\big) / \log \epsilon$.

We shortly illustrate the empirical performance of our algorithm on the random geometric graph, introduced in Section 3.2. Consider the latent space $[0, 1]$, endowed with the uniform measure and the function $W(x, y) = \mathbb{I}_{||x-y||_2 \leq 0.1}$, which has a Minkowski dimension 2 and satisfies the assumptions of Theorem 12. We sample $n = 1000$ points uniformly on $[0, 1]$ and plot the outputs $-\log \widehat{N}_{\Omega}^{(ap.c)}(\epsilon) / \log \epsilon$ over the range of radii $\epsilon \in$

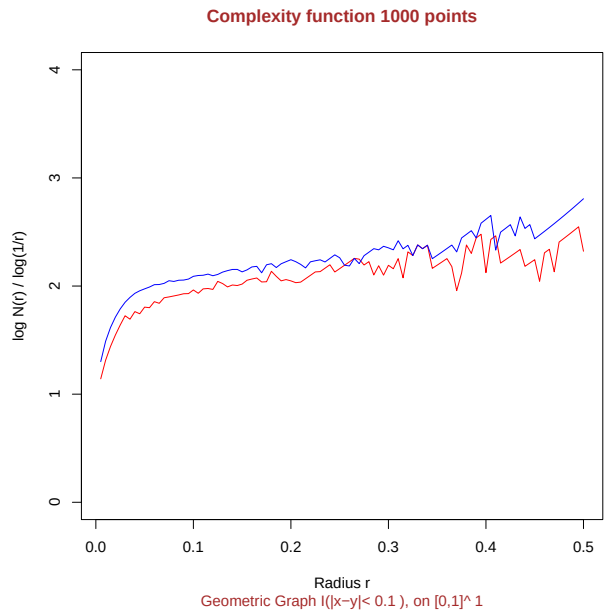

Figure 1: W-random graph with Minkowski dimension 2

$\Big\{0.005 + k * 0.005\,;\, k \in \{0, \ldots, 100\}\Big\}$. This is represented by the red curve in Figure 1. As we can see, it is close to the true dimension at some intermediate radii, which coincides with our theoretical results. Specifically, we observe the three typical parts in the graph function: (1) on the right of the figure, the sawtooth-shaped curve means that the radius is too big for approaching the Minkowski dimension (which is by definition a limit in $\epsilon \to 0$); (2) on the middle, there is a plateau whose value is close to the dimension; (3) on the left, there is an abrupt drop because the covering number estimator eventually just counts the sampled points $\omega_1, \ldots, \omega_n$. As a reference, we also plot $-\log N^{(ap.c)}_{\omega_1,\ldots,\omega_n}(\epsilon)/\log \epsilon$ in blue, where $N^{(ap.c)}_{\omega_1,\ldots,\omega_n}(\epsilon)$ is the *approximated* covering number of the sample $\{\omega_1, \ldots, \omega_n\}$ w.r.t. to the true distance r_W.

7. Discussion

7.1 On the Definition of our Complexity Index

We have introduced an index of complexity for the limiting distributions $\mathbb{P}^{(\infty)}_{(\Omega,\mu,W)}$ of W-random graphs. It has some geometric flavor as the index is based on the Minkowski dimension of the metric space (Ω, r_W), where Ω is the latent space of the W-random graph and r_W is the neighborhood distance defined on Ω. Accordingly, the index inherits from the general features of the Minkowski dimension which hold for any metric space. One of these features is the following maximum property: for any metric space (X, d) which admits a partition $X = X_1 \sqcup \ldots \sqcup X_k$, it is known that the Minkowski dimension of X is equal to the maximum of the Minkowski dimensions of the X_i, $i \in [k]$. Thus, our index captures a maximum complexity of (Ω, r_W), which is an interesting information about the

distributions $\mathbb{P}^{(\infty)}_{(\Omega,\mu,W)}$ of W-random graphs. In order to complete our index, however, it would be worthwhile to investigate other notions of complexity as well. For example, one could think of some sort of average complexity instead of the maximum complexity measured by our index, maybe replacing the Minkowski dimension with a dimension that satisfies the following kind of property: for any metric space $X = X_1 \sqcup \ldots \sqcup X_k$ endowed with a probability distribution μ, the dimension of X would be equal to an average of the dimensions of the X_i weighted according to μ. However, for such an index based on a dimension of (Ω, r_W), we remind that a major difficulty is to verify the identifiability of the index from the distribution $\mathbb{P}^{(\infty)}_{(\Omega,\mu,W)}$.

We have approximated the Minkowski dimension using a greedy algorithm in section 6.2, because the computational cost of this dimension is prohibitive. As an alternative, the correlation dimension is widely-used in manifold learning for its computational simplicity. Given sampled points $x_1, \ldots, x_n \overset{iid}{\sim} \mu$ in a metric space (X, d) endowed with a probability distribution μ, the correlation integral is usually defined as

$$C(\epsilon) = \lim_{n\to\infty} \quad \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbb{I}_{\{d(x_i,x_j)<\epsilon\}}$$

where $\mathbb{I}_{\mathcal{A}}$ denotes the indicator function of any event $\mathcal{A}$. If the limit exists, the correlation dimension of (X, d) is

$$D_{corr} = \lim_{\epsilon\to 0} \frac{\log\, C(\epsilon)}{\log\, \epsilon}.$$

It is known that D_{corr} approximates well the Minkowski dimension when the distribution μ is nearly uniform on X, whereas it may be smaller for non-uniform distributions (Kégl, 2003). In fact, one can observe that D_{corr} satisfies a minimum property, that is, the correlation dimension of $X = X_1 \sqcup \ldots \sqcup X_k$ is equal to the minimum of the correlation dimensions of the X_i, $i \in [k]$. A direction of research could be to extend the correlation dimension to W-random graph distributions, so that one get a complexity index that is simple to compute. To this end, one could consider the following form of the correlation integral

$$C(\epsilon) = \mathbb{E}_{\substack{\omega_1\sim\mu \\ \omega_2\sim\mu}} \left[\mathbb{I}_{\{r_W(\omega_1,\omega_2)<\epsilon\}}\right]$$

with respect to the metric space (Ω, r_W) associated with a graphon (Ω, μ, W).

7.2 On the Rates of Estimation

The focus of the current paper is on the whole class of graphons, which is a very general setting. In particular, our error bounds for the neighborhood distance and the covering number hold for any graphon. Therefore, a natural question is whether faster rates could be derived on sub-classes of graphons.

To estimate the Minkowski dimension of (Ω, r_W), we consider traditional assumptions in manifold learning, namely $(H_1^{\alpha,v})$ and $(H_2^{m,M,v})$, which essentially say that the metric space (Ω, r_W) behaves like a Euclidean space of dimension d endowed with a uniform distribution μ. Under these Euclidean-type assumptions, we prove that the optimal rate of estimation is $\log^{-1} n$. This rate matches known results for the Minkowski dimension in manifold learning,

see for instance (Koltchinskii, 2000). A future direction of research for W-random graph complexity could be the definition of new indices that enjoy faster rates of estimation.

Acknowledgments

We would like to thank Nicolas Verzelen for sharing his ideas, and Christophe Giraud for valuable advice. This research was initially supported by an AMX PhD funding from École Polytechnnique, and then by a grant of the French National Research Agency (ANR), "Investissements d'Avenir" (LabEx Ecodec/ANR-11-LABX-0047).

Appendix A. Additional Information

A.1 Basic Information on the Covering and Packing Numbers and the Minkowski Dimension

Given any set S, its covering number $N^{(c)}(\epsilon)$ is the minimal number of balls of radius ϵ required to entirely cover S, with the constraint that the ball centers are in S. This measure is widely used for general metric spaces. Likewise, the packing number $N^{(p)}(\epsilon)$ is the maximum number of points in a given space (strictly) separated by at least a given distance ϵ. Both measures are similar and linked by the following inequalities $N^{(c)}(\epsilon) \leq N^{(p)}(\epsilon) \leq N^{(c)}(\epsilon/2)$. In all the paper (except the last subsection 5), our results are mostly stated with the covering number, but each of them can be adapted to the packing number.

The covering number requires to choose the scale ϵ at which we look at the data. To get rid of this parameter, it is common to consider the Minkowski dimension which is defined by $\lim_{\epsilon\to 0} \; -\log N^{(c)}(\epsilon)/\log\epsilon$. Note that the same formula holds with the packing number instead. The Minkowski dimension is used for infinite (separable) spaces, when the covering number diverges to infinity as ϵ goes to zero. This dimension is therefore complementary to the covering number. It is known to match with some other classical notions of dimension in simple cases, for example the Minkowski dimension of the hypercube $[0,1]^d$ is equal to its Euclidean dimension d. The Minkowski dimension has the advantage to be applicable on a wide range of spaces (whose dimension is not necessarily an integer) and to be easy to compute (in comparison with the Hausdorff dimension for example).

A.2 Details on the Illustrative Examples

Random Hölder graph. Recall that the graphon (Ω, μ, W) is $([0,1]^d, \lambda, W)$ where λ is the uniform measure on $[0,1]^d$ and W satisfies the following condition: there exist three constants $m, M, \alpha > 0$ such that for all $\omega, \omega', \omega'' \in [0,1]^d$,

$$m\big|\big|\omega' - \omega\big|\big|_2^\alpha \leq \big|W(\omega', \omega'') - W(\omega, \omega'')\big| \leq M\big|\big|\omega' - \omega\big|\big|_2^\alpha$$

where α is the level of regularity of the function W and $||\omega' - \omega||_2$ is the Euclidean distance between ω' and ω in $[0,1]^d$. From the above display, we directly deduce some bounds on the neighborhood distance (4) :
$\forall \omega, \omega' \in [0,1]^d$,

$$m||\omega' - \omega||_2^\alpha \leq r_W(\omega', \omega) \leq M||\omega' - \omega||_2^\alpha.$$

Thus, the distance r_W behaves (up to some constants) like the Euclidean distance on $[0,1]^d$ raised to the power of α. As the covering number of the Euclidean hypercube $([0,1]^d, ||.||_2)$ is approximately equal to ϵ^{-d} for small radii, we have

$$(\epsilon/m)^{-d/\alpha} \lesssim N_\Omega^{(c)}(\epsilon) \lesssim (\epsilon/M)^{-d/\alpha}.$$

Hence $dim\,\Omega = d/\alpha$, which means that the Minkowski dimension of (Ω, r_W) is equal to the ratio between the Euclidean dimension of the latent space $[0,1]^d$ and the regularity of the function W.

Random geometric graph example. Recall that the graphon is $([0,1]^d, \lambda, W)$ where λ is the uniform measure, and W is defined as $W(\omega,\omega') = \mathbb{I}_{||\omega-\omega'||_2 \leq \delta}$ for some parameter $\delta \in]0,1[$, and $||\omega-\omega'||_2$ is the Euclidean distance between $\omega, \omega' \in [0,1]^d$. Here, the bounds on the neighborhood distance are rather involved and deferred to the Appendix B.2. The main message is that

$$r_W(\omega,\omega') \asymp \sqrt{||\omega-\omega'||_2}$$

if $||\omega-\omega'||_2$ is small enough, which means that the distance r_W behaves like the squared root of the Euclidean norm in $[0,1]^d$. Following the line of the Random Hölder graph example, we can see that $N_\Omega^{(c)}(\epsilon)$ behaves like ϵ^{-2d} for ϵ small enough. By definition of the Minkowski dimension, it follows that $dim\,\Omega = 2d$.

A.3 Test: Improvement of the Type II Error

The control of the type II error can be refined using the graphon regularity at a finer level. Instead of considering the set $\mathcal{W}(\eta,\beta)$ of graphons with $K+1$ well separated balls (Theorem 18), here we consider the new set $\mathcal{W}(\eta,\beta,M,K')$ of graphons with M disjoint collections of $K+1+K'$ separated balls. That is, for a collection of $K+1+K'$ balls, we assume the same conditions of separation, size and measure as in a collection of $K+1$ balls defined by $\mathcal{W}(\eta,\beta)$ (in Theorem 18). In addition, we assume that the M formations of $K+1+K'$ balls do not intersect each other (i.e. no ball from a collection overlaps a ball from another collection). Thus, the new set $\mathcal{W}(\eta,\beta,M,K')$ of graphons is linked with the previous one by the following equality $\mathcal{W}(\eta,\beta,1,0) = \mathcal{W}(\eta,\beta)$.

Theorem 21 *If the underlying graphon belongs to $\mathcal{W}(\eta,\beta,M,K')$ with $\beta \geq 1/n$, then the type II error is smaller than $\frac{2}{n} + \widetilde{p}_n^M$, where $\widetilde{p}_n$ admits the following upper bound*

$$\binom{K+K'+1}{K'+1}\Big(2\beta n\,\exp[-\beta(n-1)]\Big)^{(K'+1)}.$$

The proof of Theorem 18 is written in Appendix G.2.2.

Appendix B. Proofs for Illustrative Examples

B.1 Proof of Proposition 6: Approximation by SBM

Given a graphon (Ω,μ,W) and a radius $\epsilon > 0$, we consider a cover of (Ω, r_W) whose the cardinality is $N_\Omega^{(c)}(\epsilon)$ (written N for brevity), and the ball centers are $x_1,\ldots,x_N$. The

Voronoi cell V_j of x_j is the set of all elements in Ω that are closer to x_j than to any other x_k, $k \neq j$, according to the metric r_W. In the case of equality, where a point ω is at equal distance of several ball centers x_i, it belongs to the Veronoi cell of smallest index i.

$$V_j := \Big\{\omega \in \Omega : \, r_W(\omega, x_j) < r_W(\omega, x_k) \text{ if } k < j, \text{ and } r_W(\omega, x_j) \leq r_W(\omega, x_k) \text{ otherwise}\Big\}$$

Define the SBM approximation of W as follows:

$$\overline{W}(x,y) = \sum_{i,j=1}^{N} 1_{x \in V_i} 1_{y \in V_j} \frac{1}{\mu(V_i)\mu(V_j)} \int_{V_i} \int_{V_j} W(z_1, z_2)\mu(dz_1)\mu(dz_2)$$

By triangular inequality and Jensen inequality:

$$\int_{\Omega^2} (W(x,y) - \overline{W}(x,y))^2 \mu(dx)\mu(dy) \leq 2 \int_{\Omega^2} \sum_{j=1}^{N} 1_{y \in V_j} \Big[\frac{1}{\mu(V_j)} \int_{V_j} [W(x,y) - W(x,z_2)]^2 \mu(dz_2)\Big] \mu(dx)\mu(dy) +$$

$$2 \int_{\Omega^2} \Big[\sum_{i,j=1}^{N} 1_{x \in V_i} 1_{y \in V_j} \frac{1}{\mu(V_i)\mu(V_j)} \int_{V_i} \int_{V_j} [W(x,z_2) - W(z_1,z_2)]^2 \mu(dz_1)\mu(dz_2)\Big] \mu(dx)\mu(dy)$$

Note that the first term is smaller than $8\epsilon^2$ by integrating with respect to x and using the fact that y and z_2 belong to the same Voronoi cell. The second term simplifies

$$2 \int_{\Omega^2} \Big[\sum_{i,=1}^{N} 1_{x \in V_i} \frac{1}{\mu(V_i)} \int_{V_i} [W(x,y) - W(z_1,y)]^2 \mu(dz_1)\Big] \mu(dx)\mu(dy)$$

which is again smaller than $8\epsilon^2$. The approximation error of W by $\overline{W}$ is therefore lower than 4ϵ in l_2-norm. The proposition is proved. ■

B.2 The Neighborhood Distance for the Random Geometric Graph Example

Lemma 22 gives bounds on the neighborhood distance for the random geometric graph of Section 3.2. For simplicity, we neglect the side effects associated with a point too close of the side of $\Omega = [0,1]^d$. That is, we assume the parameter δ is small compared to 1 (where 1 is the length of a side of $[0,1]^d$). Write V_d the volume of the unit ball in $[0,1]^d$ endowed with the Euclidean norm $||.||_2$, and write $I_x(.,.)$ the (regularized) incomplete beta function (see DLMF, Eq.8.17.2 for a definition).

Lemma 22 *If $||x-y||_2 > 2\delta$, then $r_W^2(x,y) = 2V_d\delta^d$; otherwise $r_W^2(x,y) = 2V_d\delta^d I_x(\frac{1}{2}, \frac{d+1}{2})$ for $x = \left(\frac{||x-y||_2}{2\delta}\right)^2$. As a consequence, $\sqrt{||x-y||_2} \lesssim r_W(x,y) \lesssim \sqrt{||x-y||_2}$ as soon as $||x-y||_2$ is small enough (compared to δ).*

According to the above lemma, the neighborhood distance r_W behaves like the squared root of the Euclidean norm of $[0,1]^d$ if $||x-y||_2$ is small enough. For lower dimensions, for instance $d = 3$, we can also use the paper of Li (2011) to get the simpler formula:
if $||x-y||_2 < 2\delta$, then

$$r_W^2(x,y) = 2\pi \left(\delta^2 - \frac{||x-y||_2^2}{12}\right) ||x-y||_2.$$

Proof of Lemma 22. For the random geometric graph, observe that the computation of the neighborhood distance is equivalent to the computation of the volumes of hypersherical caps. Using the formula (3) in the paper of Li (2011) (and neglecting the side effects due to the boundary of the latent space), we have:
if $||x-y||_2 < 2\delta$, then

$$r_W^2(x,y) = 2V_d\delta^d\left[1 - I_x(\frac{d+1}{2},\frac{1}{2})\right]$$

where $x = 1 - \left(\frac{||x-y||_2}{2\delta}\right)^2$. Basic properties of the (regularized) incomplete beta function (see DLMF, Eq.8.17.4) allows to rewrite the last formula:
if $||x-y||_2 < 2\delta$, then

$$r_W^2(x,y) = 2V_d\delta^d I_x(\frac{1}{2},\frac{d+1}{2}) \tag{23}$$

where $x = \left(\frac{||x-y||_2}{2\delta}\right)^2$. Let $B(a,b)$ denote the beta function (DLMF, Eq.5.12.1), then the above formula (23) can be developed using the recurrence formula $I_x(a,b+1) = I_x(a,b) + \frac{x^a(1-x)^b}{bB(a,b)}$ (DLMF, Eq.8.17.21). It follows that r_W satisfies the following bounds: $\sqrt{||x-y||_2} \lesssim r_W(x,y) \lesssim \sqrt{||x-y||_2}$ as soon as $||x-y||_2$ is small enough. ■

Appendix C. Identifiability and Pure Graphons

C.1 Proof of Lemma 4 : Invariance of the Neighborhood Distance

Given two equivalent pure graphons (Ω,μ,W) and (Ω',μ',W'), let us show that their respective neighborhood distances r_W and $r_{W'}$ are linked by the following $\mu'\otimes\mu'$-almost surely equality

$$r_W\left(\phi(x),\phi(y)\right) = r_{W'}\left(x,y\right)$$

for some measure-preserving bijection $\phi : \Omega' \to \Omega$.

It follows from Lemma 23, which links any two equivalent pure graphons. Denote by W^ϕ the function $(x,y) \mapsto W(\phi(x),\phi(y))$.

Lemma 23 (Lovász, 2012, Section 13.3) *If two pure graphons (Ω,μ,W) and (Ω',μ',W') are equivalent, then there exists a bijective measure-preserving map $\phi : \Omega' \to \Omega$ such that $W^\phi(x,y) = W'(x,y)$ $\mu'\otimes\mu'$-almost surely.*

Indeed, by definition of the neighborhood distance,

$$r_{W'}\left(x,y\right) = \left(\int_{\Omega'}\left|W'(x,z') - W(y,z')\right|^2\mu'(dz')\right)^{1/2}$$

which gives the following $\mu'\otimes\mu'$-almost surely equality by Lemma 23,

$$r_{W'}\left(x,y\right) = \left(\int_{\Omega'}\left|W(\phi(x),\phi(z')) - W(\phi(y),\phi(z'))\right|^2\mu'(dz')\right)^{1/2}$$

for some measure-preserving bijection $\phi : \Omega' \to \Omega$. Then, using a pushforward measure (or image measure),

$$r_{W'}(x,y) = \left(\int_{\Omega} |W(\phi(x),z) - W(\phi(y),z)|^2 \, \mu(dz) \right)^{1/2}$$

$\mu' \otimes \mu'$-almost surely, so that, by definition of the neighborhood distance,

$$r_{W'}(x,y) = r_W(\phi(x), \phi(y))$$

$\mu' \otimes \mu'$-almost surely. Lemma 4 is proved. ■

C.2 Proof of Lemma 3 : Identifiability of the Covering Number

Given two equivalent pure graphons (Ω, μ, W) and (Ω', μ', W'), let us prove that their respective covering numbers are equal: $N_{\Omega}^{(c)}(\epsilon) = N_{\Omega'}^{(c)}(\epsilon)$ for all $\epsilon > 0$.

According to Lemma 4, there exists a measure-preserving bijection ϕ, such that the two metric spaces (Ω, r_W) and $(\Omega', r_{W'})$ are linked by the equality $r_{W'}(x,y) = r_W(\phi(x), \phi(y))$ on a subset of measure 1, say $\Sigma \subseteq \Omega'$ with $\mu'(\Sigma) = 1$. This means that both subpaces $(\phi(\Sigma), r_W)$ and $(\Sigma, r_{W'})$ are linked by a bijection that preserves the distances, which directly implies equality between their covering numbers: $N_{\phi(\Sigma)}^{(c)}(\epsilon) = N_{\Sigma}^{(c)}(\epsilon)$ for all $\epsilon > 0$.

Then, for proving Lemma 3, it is enough to show the two following inequalities

$$N_{\Omega}^{(c)}(\epsilon) \geq N_{\phi(\Sigma)}^{(c)}(\epsilon + \delta) \tag{24}$$

$$N_{\Sigma}^{(c)}(\epsilon + \delta) \geq N_{\Omega'}^{(c)}(\epsilon + \delta) \tag{25}$$

for any $\delta > 0$. Indeed, combining these two inequalities with the covering number equality from the above paragraph, one has $N_{\Omega}^{(c)}(\epsilon) \geq N_{\Omega'}^{(c)}(\epsilon + \delta)$. Taking the limit $\delta \to 0$ and using the right-continuity of the covering number (Lemma 24), this gives $N_{\Omega}^{(c)}(\epsilon) \geq N_{\Omega'}^{(c)}(\epsilon)$. As the reverse inequality holds by symmetry of the proof, one obtain the equality $N_{\Omega}^{(c)}(\epsilon) = N_{\Omega'}^{(c)}(\epsilon)$ of Lemma 3.

Lemma 24 *Given a pure graphon (Ω, μ, W), the function $\epsilon \mapsto N_{\Omega}^{(c)}(\epsilon)$ is piecewise constant and right-continuous (note that we use closed balls in the definition).*
Likewise, $\epsilon \mapsto N_{\Omega}^{(p)}(\epsilon)$ is a right continuous piecewise function.

Assume Σ is dense in $(\Omega', r_{W'})$. Each cover of Σ is closed as a finite union of closed balls. Hence it is also a cover of Ω' by density of Σ in Ω'. This proves (25). Likewise, assume $\phi(\Sigma)$ is dense in (Ω, r_W). An ϵ-cover of Ω can be transformed into an $(\epsilon + \delta)$-cover of $\phi(\Sigma)$ by moving the ball centers from Ω to Σ and increasing the ball radius of δ (for arbitrary small δ). This proves (24) for any $\delta > 0$.

Let us show the density of $\phi(\Sigma)$ in (Ω, r_W). One has $\mu(\phi(\Sigma)) = \mu'(\Sigma) = 1$ by definition of a (bijective) measure-preserving map, which implies that $\phi(\Sigma)$ intersects each ball of non-zero measure in (Ω, r_W). As the measure of a pure graphon has full-support by definition, then each ball of non-zero radius has a non-zero measure. Thus, $\phi(\Sigma)$ intersects each ball of non-zero radius in (Ω, r_W), which means that $\phi(\Sigma)$ is dense in (Ω, r_W). Similarly, we can show the density of Σ in $(\Omega', r_{W'})$.

Lemma 3 is proved for the covering number. The proof for the packing number is similar and omitted. ■

C.2.1 Proof of Lemma 24.

The function $\epsilon \mapsto N_\Omega^{(c)}(\epsilon)$ is non-increasing from $[0, \infty[$ to the set of all non-negative integers, it is therefore a piecewise constant function. Thus, for any radius $\epsilon_0 > 0$, there exists a (strictly) larger radius ϵ_1 such that the covering number $N_\Omega^{(c)}(\epsilon)$ is equal to a constant, say N, over the interval (ϵ_0, ϵ_1). To prove the right continuity in ϵ_0, let us show the inequality $N_\Omega^{(c)}(\epsilon_0) \leq N$ (since we already know the reverse inequality by monotonicity of the covering number function), or equivalently that there exists a cover of Ω that is composed of N balls of radius ϵ_0.

Given a radius ϵ and K points $c = (c_1, \ldots, c_K) \in \Omega^K$, denote by $C_\Omega(c, \epsilon)$ the union of K balls of centers $c_1, \ldots, c_K$. In the following, we prove: (1) the existence of some $c_0 \in \Omega^N$ such that $C_\Omega(c_0, \epsilon)$ covers Ω for all $\epsilon \in (\epsilon_0, (\epsilon_1 + \epsilon_0)/2]$; (2) for such a c_0, $C_\Omega(c_0, \epsilon_0)$ covers Ω. Thus, Lemma 24 will be proved.

(1) Define the set $E_\Omega(\epsilon) := \{c \in \Omega^N : \Omega \subseteq C_\Omega(c, \epsilon)\}$ for any given radius $\epsilon > 0$. Then, consider the following sequence of nested sets $\widetilde{E}_k := E_\Omega(\epsilon_0 + (\epsilon_1 - \epsilon_0)/k)$ where $k \geq 2$ is an integer. The Cantor's intersection theorem (recalled in Lemma 25 below) ensures that $\cap_{k \geq 2} \widetilde{E}_k \neq \emptyset$, provided that the assumptions of the theorem hold. For clarity, this verification is deferred to the end of the proof. As the set $\cap_{\epsilon_0 < \epsilon < \epsilon_1} E_\Omega(\epsilon)$ is equal to $\cap_{k \geq 2} \widetilde{E}_k$, one has $\cap_{\epsilon_0 < \epsilon < \epsilon_1} E_\Omega(\epsilon) \neq \emptyset$, which means that there exists some $c_0 \in \Omega^N$ such that $C_\Omega(c_0, \epsilon)$ covers Ω for all $\epsilon \in (\epsilon_0, (\epsilon_1 + \epsilon_0)/2]$.

(2) By contradiction, let us prove that $C_\Omega(c_0, \epsilon_0)$ covers Ω. If $C_\Omega(c_0, \epsilon_0)$ does not cover Ω, then there exists some y in the open set $\Omega \setminus C_\Omega(c_0, \epsilon_0)$, which implies that there exists an open ball $B(y, \eta)$ in $\Omega \setminus C_\Omega(c_0, \epsilon_0)$ for some radius $\eta > 0$. Hence, $r_W(y, c_{0,j}) \geq \eta + \epsilon_0$ for all $j \in \{1, \ldots, N\}$, which means that $C_\Omega(c_0, \epsilon)$ does not cover Ω for the radius $\epsilon = \epsilon_0 + \eta/2$ for instance. This is a contradiction with point (1) above.

Lemma 25 (Cantor's intersection theorem) *Suppose that (X, d) is a complete metric space, and C_n is a sequence of non-empty closed nested subsets of X whose diameters tend to zero. Then the intersection of the C_n contains exactly one point, that is $\cap_{k=1}^\infty C_k = \{x\}$ for some x in X.*

Verification of the assumptions of Lemma 25. Since (Ω, r_W) is a complete metric space by definition of a pure graphon, the product space (Ω^N, r_W^{sup}) is also complete for the sup-distance $r_W^{sup}(x, y) := \sup_{1 \leq j \leq N} r_W(x_j, y_j)$ with $x = (x_1, \ldots, x_N), y = (y_1, \ldots, y_N) \in \Omega^N$. By definition of $\widetilde{E}_k$, the sequence $(\widetilde{E}_k)_k$ is composed of nested sets, which are also non-empty since $N_\Omega^{(c)}(\epsilon) = N$ over (ϵ_0, ϵ_1). To prove that each $\widetilde{E}_k$ is a closed subset of Ω^N, it is enough to show that $E_\Omega(\epsilon)$ is closed for any $\epsilon \in (\epsilon_0, \epsilon_1)$. Let $(x^k)_{k \geq 0}$ be a sequence in $E_\Omega(\epsilon)$ such that $x^k \to x \in \Omega^N$ as $k \to \infty$. Then, for any $\eta > 0$, there exists some k_0 such that the sup-distance between $x^{k_0} = (x_1^{k_0}, \ldots, x_N^{k_0})$ and $x = (x_1, \ldots, x_N)$ is at most η. As $x^{k_0} \in E_\Omega(\epsilon)$, one know that, for any $y \in \Omega$, there exists some j_0 such that $r_W(y, x_{j_0}^{k_0}) \leq \epsilon$. Thus, using the triangle inequality, one has for any $\eta > 0$,

$$r_W(y, x_{j_0}) \leq r_W(y, x_{j_0}^{k_0}) + r_W(x_{j_0}^{k_0}, x_{j_0}) \leq \epsilon + \eta$$

which implies that $r_W(y, x_{j_0}) \leq \epsilon$. Hence, $y \in C_\Omega(x, \epsilon)$ for any $y \in \Omega$, which means that $x \in E_\Omega(\epsilon)$. $E_\Omega(\epsilon)$ is therefore a closed subset of Ω^N. All the conditions of Lemma 25 are checked.

The part of Lemma 24 on the covering number is proved. For the packing number, the proof is similar and omitted.■

C.3 Proof of Lemma 5: Asymptotic Density of the Sample

Given $\epsilon > 0$, consider a cover of (Ω, r_W) whose cardinality is the integer $N_\Omega^{(c)}(\epsilon/4)$ (written N for brevity) and whose balls are written $B_1, \ldots, B_N$. Let us upper bound the probability that (at least) one of these balls contains zero sampled point ω_i. Using the union bound, this probability is smaller than

$$\sum_{j=1}^{N} \mathbb{P}_{(\Omega,\mu,W)} \left\{ B_j \text{ contains zero sampled point among } \omega_1, \ldots, \omega_n \right\}$$

which is upper bounded by $N(1-\mu(B_j))^n \leq N(1-\beta)^n$ where $\beta := \min_{j\in[N]} \mu(B_j)$. One has $\beta > 0$ since each ball of a pure graphon has non-zero measure. And as N is not equal to infinity by assumption, this probability tends to zero with n. Thus, with high probability, all balls B_j from the cover contains at least a sampled point. Finally, the asymptotic density of the sample follows from the fact that each ball of radius ϵ of (Ω, r_W) contains a ball B_j from the cover. Lemma 5 is proved. ■

Appendix D. Estimation of the Neighborhood Distance

D.1 Proof of Theorem 7 : the Upper Bound

Theorem 7 is a direct consequence of the two following propositions. Proposition 26 shows the consistency of the inner products between the rows of the adjacency matrix A. That is, $\langle A_i, A_j \rangle_n$ is convergent in probability towards $\langle W(\omega_i, .), W(\omega_j, .) \rangle$ if $i \neq j$. Actually, Proposition 26 gives a uniform convergence over all $i, j \in [n]$, $i \neq j$.

Proposition 26 *The following event on inner products*

$$\mathcal{E}_{in} := \left\{ \forall i, j \in [n] : \ \ |\langle A_i, A_j \rangle_n - \langle W(\omega_i, .), W(\omega_j, .) \rangle| \leq 3\sqrt{\frac{\log n}{n}} \right\}$$

holds with probability $\mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}_{in}) \geq 1 - \frac{2}{n}$ *as soon as* $n \geq 6$.

We have seen that the neighborhood distance r_W can be decomposed into one crossed term and two quadratic terms as follows

$$r_W^2(\omega_i, \omega_j) = \langle W(\omega_i, .), W(\omega_i, .) \rangle + \langle W(\omega_j, .), W(\omega_j, .) \rangle - 2\langle W(\omega_i, .), W(\omega_j, .) \rangle. \tag{26}$$

Proposition 26 ensures that the crossed term is consistently estimated. Proposition 27 deals with the quadratic terms $\langle W(\omega_i, .), W(\omega_i, .) \rangle$.

Proposition 27 *Conditionally to the event* $\mathcal{E}_{in}$ *(defined above), the following inequalities*

$$\forall i \in [n] : \ \ \left| \langle A_i, A_{\widehat{m}(i)} \rangle_n - \langle W(\omega_i, .), W(\omega_i, .) \rangle \right| \leq 3\, r_W(\omega_i, \omega_{m(i)}) \, + \, 15\sqrt{\log(n)/n}$$

hold simultaneously as soon as $n \geq 6$.

The estimation error of (26) by our distance estimator

$$\widehat{r}^2(i,j) = \langle A_i, A_{\widehat{m}(i)}\rangle_n + \langle A_j, A_{\widehat{m}(j)}\rangle_n - 2\langle A_i, A_j\rangle_n$$

follows directly from Propositions 26 and 27. Theorem 7 is proved. ■

D.1.1 Proof of Proposition 26.

By triangle inequality, the expression

$$\Big|\sum_k \frac{A_{ik}A_{kj}}{n} - \int_\Omega W(\omega_i,z)W(\omega_j,z)\mu(dz)\Big|$$

is smaller than

$$\leq \frac{1}{n}\Big|\sum_{k\neq i,j} A_{ik}A_{kj} - (n-2)\int_\Omega W(\omega_i,z)W(\omega_j,z)\mu(dz)\Big|$$
$$+\frac{1}{n}\Big[(A_{ii}+A_{jj})A_{ij} + 2\int_\Omega W(\omega_i,z)W(\omega_j,z)\mu(dz)\Big]$$

which is upper bounded by

$$\leq \frac{1}{n-2}\Big|\sum_{k\neq i,j} A_{ik}A_{kj} - (n-2)\int_\Omega W(\omega_i,z)W(\omega_j,z)\mu(dz)\Big| + \frac{4}{n}.$$

Conditionally to ω_i, ω_j (with $i \neq j$), the $n-2$ random variables $\{A_{ik}A_{kj} : k \in [n], k \neq i,j\}$ are independent with a mean $\mathbb{E}\left[A_{ik}A_{kj}|\omega_i,\omega_j\right] = \int_\Omega W(\omega_i,z)W(\omega_j,z)\mu(dz)$ for all $k \neq i,j$ (where $\mathbb{E}$ is the expectation with respect to the distribution $\mathbb{P}_{(\Omega,\mu,W)}$). It follows from Hoeffding's inequality that

$$\mathbb{P}_{(\Omega,\mu,W)}\left(\frac{1}{n-2}\Big|\sum_{k\neq i,j} A_{ik}A_{kj} - (n-2)\int_\Omega W(\omega_i,z)W(\omega_j,z)\mu(dz)\Big| \geq \epsilon \;\Big|\omega_i,\omega_j\right)$$

is lower than

$$\leq 2\exp\left(-2(n-2)\epsilon^2\right) \leq 2\exp\left(-n\epsilon^2\right)$$

for $\epsilon > 0$ and $n \geq 4$. Since the above inequality is satisfied for almost every $\omega_i, \omega_j \in \Omega$, one has the same upper bound with probability 1 without conditioning. Hence, taking a union bound over all $i \neq j$ one obtain

$$\mathbb{P}_{(\Omega,\mu,W)}\left(\bigcup_{i,j:i\neq j}\left\{\frac{1}{n-2}\Big|\sum_{k\neq i,j} A_{ik}A_{kj} - (n-2)\int_\Omega W(\omega_i,z)W(\omega_j,z)\mu(dz)\Big| \geq \epsilon\right\}\right) \leq 2n^2\exp\left(-n\epsilon^2\right).$$

Then, setting $\epsilon = \sqrt{\frac{3\log n}{n}}$ one derive

$$\mathbb{P}_{(\Omega,\mu,W)}\left(\bigcup_{i,j:i\neq j}\left\{\frac{1}{n-2}\Big|\sum_{k\neq i,j} A_{ik}A_{kj} - (n-2)\int_\Omega W(\omega_i,z)W(\omega_j,z)\mu(dz)\Big| \geq \sqrt{\frac{3\log n}{n}}\right\}\right) \leq \frac{2}{n}.$$

Combining the above expressions, we get the following inequality

$$\max_{i,j:i\neq j}\Big|\sum_k \frac{A_{ik}.A_{kj}}{n} - \int_\Omega W(\omega_i,z)W(\omega_j,z)\mu(dz)\Big| \leq \sqrt{\frac{3\log n}{n}} + \frac{4}{n} \leq 3\sqrt{\frac{\log n}{n}}$$

with probability at least $1-\frac{2}{n}$ as soon as $n \geq 6$. The proof of Proposition 26 is complete. ■

D.1.2 Proof of Proposition 27.

$$\begin{aligned}\big|\langle A_i, A_{\widehat{m}(i)}\rangle_n - \langle W(\omega_i,.), W(\omega_i,.)\rangle\big| &\leq \big|\langle A_i, A_{\widehat{m}(i)} - A_{m(i)}\rangle_n\big| \\ &\quad + \big|\langle A_i, A_{m(i)}\rangle_n - \langle W(\omega_i,.), W(\omega_i,.)\rangle\big| \qquad (27)\end{aligned}$$

For the second term of the upper bound (27),

$$\begin{aligned}\big|\langle A_i, A_{m(i)}\rangle_n - \langle W(\omega_i,.), W(\omega_i,.)\rangle\big| &\leq \big|\langle A_i, A_{m(i)}\rangle_n - \langle W(\omega_i,.), W(\omega_{m(i)},.)\rangle\big| \\ &\quad + \big|\langle W(\omega_i,.), W(\omega_{m(i)},.) - W(\omega_i,.)\rangle\big| \\ &\leq 3\sqrt{\log(n)/n} + r_W(\omega_i,\omega_{m(i)})\end{aligned}$$

by Proposition 26 and Cauchy-Schwarz inequality. For the first term of the upper bound (27), if $\widehat{m}(i) \neq m(i)$,

$$\begin{aligned}\big|\langle A_i, A_{\widehat{m}(i)} - A_{m(i)}\rangle_n\big| &\leq \big|\langle A_i - A_{m(i)}, A_{\widehat{m}(i)}\rangle_n\big| + \big|\langle A_i - A_{\widehat{m}(i)}, A_{m(i)}\rangle_n\big| \\ &\leq \widehat{f}(i,m(i)) + \widehat{f}(i,\widehat{m}(i)) \\ &\leq 2\widehat{f}(i,m(i))\end{aligned}$$

by definition of $\widehat{m}(i)$ and $\widehat{f}$ in (9). We upper bound $\widehat{f}(i,m(i))$ as follows.

$$\begin{aligned}\widehat{f}(i,m(i)) := \max_{k\neq i,m(i)} \big|\langle A_k, A_i - A_{m(i)}\rangle_n\big| &\leq \max_{k\neq i,m(i)} \big|\langle W(\omega_k,.), W(\omega_i,.) - W(\omega_{m(i)},.)\rangle\big| \\ &\quad + 2\max_{l,t:\, l\neq t} |\langle A_l, A_t\rangle_n - \langle W(\omega_l,.), W(\omega_t,.)\rangle| \\ &\leq r_W(\omega_i,\omega_{m(i)}) + 6\sqrt{\log(n)/n}\end{aligned}$$

by Proposition 26 and Cauchy-Schwarz. Combining the upper bounds on (27), Proposition 27 is proved. ■

D.2 Proof of Theorem 10 : the Lower Bound

Theorem 10 is a corollary of Theorem 28 (written below). Let $\mathbf{r}_\omega$ denote the $n\times n$ symmetric matrix with entries $r_W(\omega_i,\omega_j)$, $1 \leq i \leq j \leq n$. Given a real $\delta > 0$, a graphon (Ω,μ,W), a permutation σ of $\{1,\ldots,n\}$ and an estimator $\widehat{d}$, we define

$$\begin{aligned}\mathcal{S}_{(\Omega,\mu,W)}(\widehat{d},\sigma,\mathbf{r}_\omega) = \Big\{(i,j)\ :\ 32\Big|\widehat{d}^2(\sigma(i),\sigma(j)) - r_W^2(\omega_i,\omega_j)\Big| \ \geq\ 2\,\delta \\ \text{and} \quad 2\,\delta \geq\ r_W(\omega_i,\omega_{m(i)}) + r_W(\omega_j,\omega_{m(j)})\Big\}\end{aligned}$$

and

$$\Phi_{(\Omega,\mu,W)}(\widehat{d},\mathbf{r}_\omega) = \inf_\sigma \text{ Card } \quad \mathcal{S}_{(\Omega,\mu,W)}(\widehat{d},\sigma,\mathbf{r}_\omega) \tag{28}$$

where $\Phi_{(\Omega,\mu,W)}(\widehat{d},\mathbf{r}_\omega)$ is the number of pairs (i,j) where the estimator $\widehat{d}$ is no better than our estimator $\widehat{r}$, roughly speaking. That is, $\Phi_{(\Omega,\mu,W)}(\widehat{d},\mathbf{r}_\omega)$ counts the pairs (i,j) for which the error of $\widehat{d}$ is larger than the bias of our distance estimator $\widehat{r}$, which is $r_W(\omega_i,\omega_{m(i)}) + r_W(\omega_j,\omega_{m(j)})$ up to some numerical constants. We put an infimum over all permutations σ of the n indices because we consider the problem of recovery of the set of distances $r_W(\omega_i,\omega_j)$, $1 \le i \le j \le n$, regardless of their labeling. According to Theorem 28, there exists a sequence of graphons (Ω,μ,W_n) such that for any estimator $\widehat{d}$, the quantity $\Phi_{(\Omega,\mu,W_n)}(\widehat{d},\mathbf{r}_\omega)$ grows linearly with n (on an event of positive probability).

Theorem 28 *There exists a sequence $(\Omega,\mu,W_n)_{n\ge 0}$ of SBM such that for all $n \ge 10$, all $\delta \in (\sqrt{\frac{8}{n-2}}, 1/40)$ and some numerical constants $c > 0$ and $p > 0$, the following lower bound holds*

$$\inf_{\widehat{d}} \ \mathbb{P}_{(\Omega,\mu,W_n)}\Big[\Phi_{(\Omega,\mu,W_n)}(\widehat{d},\boldsymbol{r}_\omega) > cn\Big] \ge p \tag{29}$$

where $\inf_{\widehat{d}}$ is the infimum over all estimators.

Theorem 10 follows from Theorem 28, choosing $\delta = \left(\sqrt{\frac{\log n}{n}}\right)^{1/(1+\gamma)}$. ■

Note that Theorem 9 follows from Theorem 28 for $\delta = 1/50$.

Proof of Theorem 28. The proof follows the general scheme of reduction for testing two hypotheses (see Yu, 1997; Tsybakov, 2009). We start with the definition of some SBM with five communities where the latent space Ω is $\{\mathcal{C}_1,\dots,\mathcal{C}_5\}$. We then show that for these SBM, any distance estimator suffers from a large loss.

Let $n \ge 10$ and $\delta \in (\sqrt{8/n-2}, 1/40)$. Consider the symmetric functions $W_n : \{\mathcal{C}_1,\dots,\mathcal{C}_5\}^2 \to \{\mathcal{C}_1,\dots,\mathcal{C}_5\}$ as described in Table 1 below. That is, for the two diagonal blocks $\{\mathcal{C}_1,\mathcal{C}_2\}^2$ and $\{\mathcal{C}_3,\mathcal{C}_4,\mathcal{C}_5\}^2$, it is a constant function:

$$W_n(x,y) = \begin{cases} 1/2 & \text{if } (x,y) \in \{\mathcal{C}_1,\mathcal{C}_2\}^2, \\ 1/2 & \text{if } (x,y) \ \in \{\mathcal{C}_3,\mathcal{C}_4,\mathcal{C}_5\}^2, \end{cases}$$

and for the upper right corner block $\{\mathcal{C}_1,\mathcal{C}_2\} \times \{\mathcal{C}_3,\mathcal{C}_4,\mathcal{C}_5\}$:

$$W_n(x,y) = \begin{cases} 1/2 + u_x\sqrt{\delta/2} & \text{if } y \in \mathcal{C}_3, \\ 1/2 + u_x\delta & \text{if } y \in \mathcal{C}_4, \\ 1/2 + u_x/2 & \text{if } y \in \mathcal{C}_5, \end{cases} \qquad u_x = \begin{cases} +1 & \text{if } x \in \mathcal{C}_1, \\ -1 & \text{if } x \in \mathcal{C}_2. \end{cases}$$

The latent space $\{\mathcal{C}_1,\dots,\mathcal{C}_5\}$ is endowed with the probability measure μ defined as follows:

$$\mu(\mathcal{C}_1) = \mu(\mathcal{C}_2) = \frac{1-2\eta}{2}$$

$$\frac{1}{2}\mu(\mathcal{C}_3) = \mu(\mathcal{C}_4) = \mu(\mathcal{C}_5) = \frac{\eta}{2}$$

where $\eta = 2/(n-2)$.

<table>
<tr><th></th><th>$\mathcal{C}_1$</th><th>$\mathcal{C}_2$</th><th>$\mathcal{C}_3$</th><th>$\mathcal{C}_4$</th><th>$\mathcal{C}_5$</th></tr>
<tr><td>$\mathcal{C}_1$</td><td colspan="2" rowspan="2">$1/2$</td><td>$1/2+\sqrt{\delta/2}$</td><td>$1/2+\delta$</td><td>1</td></tr>
<tr><td>$\mathcal{C}_2$</td><td>$1/2-\sqrt{\delta/2}$</td><td>$1/2-\delta$</td><td>0</td></tr>
<tr><td>$\mathcal{C}_3$</td><td></td><td></td><td colspan="3" rowspan="3">$1/2$</td></tr>
<tr><td>$\mathcal{C}_4$</td><td></td><td></td></tr>
<tr><td>$\mathcal{C}_5$</td><td></td><td></td></tr>
</table>

Table 1: values of $W_n(\mathcal{C}_i, \mathcal{C}_j)$

<table>
<tr><th></th><th>$\mathcal{C}_1$</th><th>$\mathcal{C}_2$</th><th>$\mathcal{C}_3$</th><th>$\mathcal{C}_4$</th><th>$\mathcal{C}_5$</th></tr>
<tr><td>$\mathcal{C}_1$
$\mathcal{C}_2$</td><td colspan="2">$\leq 2\eta$</td><td>$\geq \delta/4$</td><td>$\leq 5\,\delta^2$</td><td>$\geq 1/4$</td></tr>
</table>

Table 2: bounds on $r_W^2(\mathcal{C}_i, \mathcal{C}_j)$

We compute some bounds on the neighborhood distance associated with the above SBM, see Table 2 for a summary. These bounds follow easily from the definition (4) of the distance. For example,

$$r_W^2(\mathcal{C}_1, \mathcal{C}_3) \geq \int_{\{\mathcal{C}_1,\mathcal{C}_2\}} |W(\mathcal{C}_1, z) - W(\mathcal{C}_3, z)|^2 \ \mu(dz) \geq \big(\mu(\mathcal{C}_1) + \mu(\mathcal{C}_2)\big)\delta/2 \geq (1-2\eta)\delta/2$$

which is larger than $\delta/4$ since $\eta = 2/(n-2)$ and $n \geq 10$.

We now introduce two events $\mathcal{R}_1$ and $\mathcal{R}_2$ on the sampled points $\omega_1, \ldots, \omega_n$, which lead to different sets of distances (for r_W), and yet are difficult to decipher for any estimator based on the adjacency matrix A. In addition, we want these two events to happen with a positive probability p that is independent of n. Observe that the union of the two communities $\mathcal{C}_1, \mathcal{C}_2$ have a total weight $1 - 2\eta = 1 - 4/(n-2)$ and thus concentrate most of the probability measure, whereas each of the remaining communities $\mathcal{C}_3, \mathcal{C}_4$, $\mathcal{C}_5$ has a weight of the order of n^{-1}. It follows that most of the sampled points $\omega_1, \ldots, \omega_n$ belong to the communities $\mathcal{C}_1, \mathcal{C}_2$ with large probability. In particular, the two following events

$$\begin{aligned}
\mathcal{R}_1 &= \Big\{\mathcal{C}_1 \cup \mathcal{C}_2,\ \mathcal{C}_4,\ \mathcal{C}_5 \text{ respectively contain n-2, 1, 1 sampled points}\Big\} \\
\mathcal{R}_2 &= \Big\{\mathcal{C}_1 \cup \mathcal{C}_2,\ \mathcal{C}_3 \text{ respectively contain n-2, 2 sampled points}\Big\}
\end{aligned}$$

happen with a positive probability that is independent of n (Lemma 29).

Lemma 29 *The probability of each event $\mathcal{R}_1$ and $\mathcal{R}_2$ is lower bounded by some numerical constant $p > 0$:*

$$\mathbb{P}(\mathcal{R}_2) \geq \mathbb{P}(\mathcal{R}_1) \geq p$$

where $\mathbb{P}(\mathcal{R}_k) := \mathbb{P}_{(\Omega,\mu,W_n)}(\mathcal{R}_k) = \int_{(\omega_1,\ldots,\omega_n)\in\{\mathcal{C}_1,\ldots,\mathcal{C}_5\}^n} 1_{\mathcal{R}_k}(\omega_1, \ldots, \omega_n)\, \mathrm{d}\mu(\omega_1) \ldots \mu(\omega_n)$.

One of the interests of the two events $\mathcal{R}_1, \mathcal{R}_2$ is to lead to different sets of distances. Specifically, if $\mathcal{R}_1$ (resp. $\mathcal{R}_2$) holds, the random matrix $\mathbf{r}_\omega = [r_W(\omega_i, \omega_j)]_{i,j\in[n]}$ of distances is denoted by $\mathbf{r}_1 = [r_1(i,j)]_{i,j\in[n]}$ (resp. $\mathbf{r}_2 = [r_2(i,j)]_{i,j\in[n]}$). We measure the difference between both matrices $\mathbf{r}_1$, $\mathbf{r}_2$ of distances as follows:

$$\widetilde{\Phi}(\mathbf{r}_1, \mathbf{r}_2) = \inf_{\sigma} \text{ Card} \left\{ (i,j) \ : \ \begin{array}{l} 16\Big|r_2^2(i,j) - r_1^2(\sigma(i)\,\sigma(j))\Big| \geq 2\,\delta \\ r_2(i, m(i)) + r_2(j, m(j)) \leq 2\,\delta \\ r_1(\sigma(i), m(\sigma(i)) + r_1(\sigma(j), m(\sigma(j)) \leq 2\,\delta \end{array} \right\} \tag{30}$$

where $\widetilde{\Phi}$ is the number of pairs (i,j) on which $\mathbf{r}_1$ and $\mathbf{r}_2$ are separated by at least the bias of our distance estimator $\widehat{r}$ (up to some numerical constants). Note that this measure is independent of the labeling $i \in \{1\ldots, n\}$ since an infimum is taken over all permutations σ of the n indices. Lemma 30 ensures that $\mathbf{r}_1$ and $\mathbf{r}_2$ are different enough for a number of pairs (i,j) that is linear with n, regardless of their labeling.

Lemma 30 *There exists a numerical constant c such that $\widetilde{\Phi}(\,\boldsymbol{r}_1, \boldsymbol{r}_2) \geq 2c\,n$.*

So far, we have two events of positive probability which lead to two different sets of distances. It remains to see that they are difficult to decipher from the observed adjacency matrix A (Lemma 31). For simplicity, write $\mathbb{P}$ for $\mathbb{P}_{(\Omega,\mu,W)}$ in the following, and $\boldsymbol{\omega}$ the n-tuple $(\omega_1, \ldots, \omega_n)$, and $\{0,1\}^{n\times n}_{sym}$ the set of binary symmetric matrices of size $n \times n$.

Lemma 31 *For any $M \in \{0,1\}^{n\times n}_{sym}$, one has*

$$\mathbb{P}[A = M | \boldsymbol{\omega} \in \mathcal{R}_1] = \mathbb{P}[A = M | \boldsymbol{\omega} \in \mathcal{R}_2].$$

We now have all the ingredients to lower bound $\mathbb{P}\Big[\Phi_{(\Omega,\mu,W_n)}(\,\widehat{d}, \mathbf{r}_\omega) > cn\Big]$ and prove Theorem 28. For clarity, $\Phi_{(\Omega,\mu,W_n)}(\,\widehat{d}\,, \mathbf{r}_\omega)$ is denoted by $\Phi(\,\widehat{d}, \mathbf{r}_\omega)$ in the following. Then, one has

$$\begin{aligned} \mathbb{P}\Big[\Phi(\,\widehat{d}, \mathbf{r}_\omega) > cn\Big] \geq \mathbb{P}\Big[\Phi(\,\widehat{d}, \mathbf{r}_\omega) > cn | \mathcal{R}_1\Big]\mathbb{P}(\mathcal{R}_1) \\ + \mathbb{P}\Big[\Phi(\,\widehat{d}, \mathbf{r}_\omega) > cn | \mathcal{R}_2\Big]\mathbb{P}(\mathcal{R}_2) \end{aligned}$$

By definition of the SBM, the matrix $\mathbf{r}_1$ remains the same for any $\boldsymbol{\omega} \in \mathcal{R}_1$, up to a permutation of the labeling. Combining with the fact that Φ is independent of the labeling, one obtain that $\Phi(\,\widehat{d}, \mathbf{r}_1)$ takes a same value for all $\boldsymbol{\omega} \in \mathcal{R}_1$. Similarly, $\Phi(\,\widehat{d}, \mathbf{r}_2)$ takes the same value for all $\boldsymbol{\omega} \in \mathcal{R}_2$. Hence, the above display says that $\mathbb{P}\Big[\Phi(\,\widehat{d}, \mathbf{r}_\omega) > cn\Big]$ is larger than

$$\Big(\mathbb{P}\big[\Phi(\,\widehat{d}, \mathbf{r}_1) > cn | \mathcal{R}_1\big] + \mathbb{P}\big[\Phi(\,\widehat{d}, \mathbf{r}_2) > cn | \mathcal{R}_2\big]\Big)\Big(\mathbb{P}(\mathcal{R}_1) \wedge \mathbb{P}(\mathcal{R}_2)\Big)$$

and since $\mathbb{P}(\mathcal{R}_1) \wedge \mathbb{P}(\mathcal{R}_2) \geq p$ by Lemma 29, one has

$$\mathbb{P}\Big[\Phi(\,\widehat{d}, \mathbf{r}_\omega) > cn\Big] \geq \Big(\mathbb{P}\big[\Phi(\,\widehat{d}, \mathbf{r}_1) > cn | \mathcal{R}_1\big] + \mathbb{P}\big[\Phi(\,\widehat{d}, \mathbf{r}_2) > cn | \mathcal{R}_2\big]\Big)p$$

Now assume that

$$\mathbb{P}\big[\Phi(\widehat{d},\mathbf{r}_2) > cn|\mathcal{R}_2\big] \geq \mathbb{P}\big[cn > \Phi(\widehat{d},\mathbf{r}_1)|\mathcal{R}_1\big]. \tag{31}$$

Then, combining the two last inequalities gives

$$\mathbb{P}\Big[\Phi(\widehat{d},\mathbf{r}_\omega) > cn\Big] \geq p$$

which gives the lower bound of Theorem 28.

Let us show that (31) holds. Lemma 31 gives

$$\mathbb{P}\big[\Phi(\widehat{d},\mathbf{r}_2) > cn|\mathcal{R}_2\big] = \mathbb{P}\big[\Phi(\widehat{d},\mathbf{r}_2) > cn|\mathcal{R}_1\big].$$

Then, we use the generalized triangle inequality of Lemma 32 with $B = \widehat{d}$.

Lemma 32 *For any $B \in \{0,1\}^{n\times n}_{sym}$, we have $\Phi(B,\boldsymbol{r}_1) + \Phi(B,\boldsymbol{r}_2) \geq \widetilde{\Phi}(\boldsymbol{r}_1,\boldsymbol{r}_2)$.*

That is,

$$\Phi(\widehat{d},\mathbf{r}_2) \geq \widetilde{\Phi}(\mathbf{r}_1,\mathbf{r}_2) - \Phi(\widehat{d},\mathbf{r}_1)$$

which is larger than

$$2cn - \Phi(\widehat{d},\mathbf{r}_1)$$

by Lemma 30. Combing the above displays, one has

$$\mathbb{P}\big[\Phi(\widehat{d},\mathbf{r}_2) > cn|\mathcal{R}_2\big] \geq \mathbb{P}\big[cn > \Phi(\widehat{d},\mathbf{r}_1)|\mathcal{R}_1\big].$$

The line (31) is therefore proved and Theorem 28 follows. ■

We now show the technical lemmas, used in the proof of Theorem 28.

D.2.1 Proof of Lemma 29.

Let $n \geq 10$. We show that each of the two events $\mathcal{R}_1, \mathcal{R}_2$ occurs with a positive probability that is independent of n. By definition of the events, one has

$$\mathbb{P}(\mathcal{R}_2) \geq \mathbb{P}(\mathcal{R}_1) = \frac{n(n-1)}{2}\left(\frac{\eta}{2}\right)^2(1-2\eta)^{n-2}$$

which is equal to the following expression for $\eta = 2/(n-2)$,

$$\frac{n(n-2)}{2}\left(\frac{1}{n-2}\right)^2 \exp\left[(n-2)\log\left(1-\frac{4}{n-2}\right)\right]$$

Using $\log(1-x) \geq -x/(1-x)$ for all x in $]0,1[$,

$$\mathbb{P}(\mathcal{R}_1) \geq \exp\left[-\frac{4}{1-\frac{4}{n-2}}\right]$$

which is larger than some positive numerical constant. Lemma 29 is proved. ■

D.2.2 Proof of Lemma 30.

The proof consists in finding a lower bound of $\widetilde{\Phi}(\mathbf{r}_1, \mathbf{r}_2)$ that is linear with n. As $\widetilde{\Phi}$ is independent of the labeling of the set of distances $\mathbf{r}_1$ and $\mathbf{r}_2$, one can assume the two following labelings without the loss of generality. For the matrix $\mathbf{r}_1$ (defined on the event $\mathcal{R}_1$), assume the $(n-1)^{\text{th}}$ and n^{th} columns correspond to the two sampled points in $\{\mathcal{C}_4, \mathcal{C}_5\}$. For $\mathbf{r}_2$ (defined on $\mathcal{R}_2$), assume the $(n-1)^{\text{th}}$ and n^{th} columns correspond to the two sampled points in $\mathcal{C}_3$. Accordingly, the $n-2$ first columns of $\mathbf{r}_1$ and $\mathbf{r}_2$ are associated with the sampled points in $\{\mathcal{C}_1, \mathcal{C}_2\}$.

We focus on the $(n-1)^{\text{th}}$ and n^{th} columns of $\mathbf{r}_2$ corresponding to the points in $\mathcal{C}_3$. For the measure $\widetilde{\Phi}$, at least one these two columns will be necessarily compared to one of the $n-1$ first columns of $\mathbf{r}_1$. In other words, the distances associated with a point in $\mathcal{C}_3$ will be compared to the distances associated with a point in $\mathcal{C}_1, \mathcal{C}_2$ or $\mathcal{C}_4$. As we can see in Table 1 and 2, such comparisons will lead to the lower bound $\widetilde{\Phi}(\mathbf{r}_1, \mathbf{r}_2) \geq n-3$. The corresponding computation are done below, focusing on the two vectors of distances $[r_2(k, n-1)]_{k \leq n-2}$ and $[r_2(k, n)]_{k \leq n-2}$.

By definition, $\widetilde{\Phi}(\mathbf{r}_1, \mathbf{r}_2)$ is based on the infimum over all permutations. Let σ be any permutation of $\{1, \ldots, n\}$ and prove the lower bound for σ, distinguishing three cases.

Case 1: if $\sigma(n) = n$, then $\sigma(j) \in \{1, \ldots, n-1\}$ for all $j \leq n-1$. For convenience, note $\mathcal{C}_{i,j}$ for a point in $\mathcal{C}_i \cup \mathcal{C}_j$. For all $j \leq n-2$, one has

$$\left| r_2^2(j, n-1) - r_1^2(\sigma(j), \sigma(n-1)) \right| = \left| r_W^2(\mathcal{C}_{1,2}, \mathcal{C}_3) - r_W^2(\mathcal{C}_{1,2,4}, \mathcal{C}_{1,2,4}) \right|$$

according to the chosen labelings (described above). It follows from Table 2 that:

$$\left| r_W^2(\mathcal{C}_{1,2}, \mathcal{C}_3) - r_W^2(\mathcal{C}_{1,2,4}, \mathcal{C}_{1,2,4}) \right| \geq \delta/4 - \max(2\eta, 5\delta^2)$$

which is equal to $\delta(1/4 - 5\delta)$ since $\eta = 2/(n-2)$ and $\delta^2 > 8/(n-2)$ by assumption. Hence, using the condition $\delta \leq 1/40$, it is larger than $\delta/8$, so that,

$$16 \left| r_2^2(j, n-1) - r_1^2(\sigma(j), \sigma(n-1)) \right| \geq 2\delta$$

for all $j \leq n-2$.

It remains to upper bound the bias terms by 2δ. The ones related to $\mathbf{r}_2$ are easily obtained: for all $j \leq n$,

$$r_2(j, m(j)) \leq r_W(\mathcal{C}_1, \mathcal{C}_2) \leq 2\eta \leq \delta$$

since on the event $\mathcal{R}_2$, a point ω_j is either in $\mathcal{C}_1 \cup \mathcal{C}_2$ and hence $r_2(j, m(j)) \leq r_W(\mathcal{C}_1, \mathcal{C}_2)$, or in $\mathcal{C}_3$ and thus $r_2(j, m(j)) = 0$ (because its nearest neighbor is in $\mathcal{C}_3$ too). This gives the bounds on the bias terms

$$r_2(i, m(i)) + r_2(j, m(j)) \leq 2\,\delta$$

for all i, j. The corresponding bounds for $\mathbf{r}_1$ are similarly obtained from Table 1, but with more calculations. It is therefore encapsulated in the following lemma.

Lemma 33 *If $\sigma(n) = n$, we have* $r_1(\sigma(i), m(\sigma(i)) + r_1(\sigma(j), m(\sigma(j)) \leq 2\delta$ *for all $j, i \leq n-1$ such that $i \neq j$.*

Combining the above displays, we obtain the lower bound

$$\mathrm{Card}\left\{(i,j)\ :\ \begin{array}{l} 16\Big|r_2^2(i,j)-r_1^2(\sigma(i)\,\sigma(j))\Big| \geq 2\,\delta \\ r_2(i,m(i))+r_2(j,m(j)) \leq 2\,\delta \\ r_1(\sigma(i),m(\sigma(i))+r_1(\sigma(j),m(\sigma(j)) \leq 2\,\delta \end{array}\right\} \geq n-3 \tag{32}$$

for all permutations fulfilling $\sigma(n)=n$.

Case 2: if $\sigma(n-1)=n$, then $\sigma(n),\sigma(j)\in\{1,\ldots,n-1\}$ for all $j\leq n-2$. Following the same proof as above, we can show that $\Big|r_2^2(n,j)-r_1^2(\sigma(n),\sigma(j))\Big| \geq 2\delta$ for all $j\leq n-2$. Likewise, the bounds on the bias terms are obtained as before. The inequality (32) is therefore proved for all permutations fulfilling $\sigma(n-1)=n$.

Case 3: if $\sigma(n)\neq n$ and $\sigma(n-1)\neq n$. Following the same proof as above, we can show that $\Big|r_2^2(n,j)-r_1^2(\sigma(n),\sigma(j))\Big| \geq 2\delta$ for all $j\leq n-2$ such that $j\neq\sigma^{-1}(n)$. The inequality (32) is therefore proved for all permutations $\sigma(n)\neq n$ and $\sigma(n-1)\neq n$.

Finally, the lower bound (32) is true for all permutations σ, in particular for the infimum over all of them. Lemma 30 is proved. ■

D.2.3 Proof of Lemma 33.

Let us upper bound the bias terms for $\mathbf{r}_1$, in the case of an arbitrary permutation σ fulfilling $\sigma(n)=n$. On the event $\mathcal{R}_1$, one has

$$r_1(\sigma(i),m(\sigma(i))+r_1(\sigma(j),m(\sigma(j)) \leq r_W(\mathcal{C}_1,\mathcal{C}_4)+r_W(\mathcal{C}_1,\mathcal{C}_2).$$

for all $j,i\leq n-1$ such that $j\neq i$. In Table 1 and Table 2, one observes that

$$\begin{aligned} &r_W(\mathcal{C}_1,\mathcal{C}_2) \leq \sqrt{2\eta} \\ &r_W(\mathcal{C}_1,\mathcal{C}_4) \leq \sqrt{\delta^2(1-2\eta)+(\delta/2)\eta+\delta^2(\eta/2)+(1/4)(\eta/2)}. \end{aligned}$$

The second bound is smaller than $\sqrt{\delta^2(1-(3\eta/2))+\eta/4}$ since $(\delta/2)\eta \leq (1/4)(\eta/2)$ (using the assumption $\delta\leq 1/40$). Hence,

$$r_1(\sigma(i),m(\sigma(i))+r_1(\sigma(j),m(\sigma(j)) \leq \sqrt{\delta^2+\eta/4}+\sqrt{2\eta}$$

which is lower than $\delta+2\sqrt{\eta}$, and again, lower than 2δ (since $\sqrt{\eta}=\sqrt{2/(n-2)}$ is smaller than $\delta/2$ by assumption). Lemma 33 is proved. ■

D.2.4 Proof of Lemma 32.

Given any matrix $B\in\{0,1\}^{n\times n}_{sym}$, let us show the following inequality $\Phi(B,\mathbf{r}_1)+\Phi(B,\mathbf{r}_2) \geq \widetilde{\Phi}(\mathbf{r}_1,\mathbf{r}_2)$ where $\widetilde{\Phi}$ and Φ are respectively defined by (30) and (28).

For all permutations σ of $\{1,\ldots,n\}$, the triangle inequality gives

$$2\Big|B_{ij}-r_2^2(i,j)\Big|\vee 2\Big|B_{ij}-r_1^2(\sigma(i)\,\sigma(j))\Big| \geq \Big|B_{ij}-r_2^2(i,j)\Big|+\Big|B_{ij}-r_1^2(\sigma(i)\,\sigma(j))\Big| \geq \Big|r_2^2(i,j)-r_1^2(\sigma(i)\,\sigma(j))\Big|$$

so that $\mathrm{Card}\left\{(i,j)\ :\ \begin{array}{l} 16\left|r_2^2(i,j)-r_1^2(\sigma(i)\,\sigma(j))\right| \geq 2\,\delta \\ r_2(i,m(i))+r_2(j,m(j)) \leq 2\,\delta \\ r_1(\sigma(i),m(\sigma(i)))+r_1(\sigma(j),m(\sigma(j))) \leq 2\,\delta \end{array}\right\}$ lower bounds the sum of the two cardinal numbers

$$\mathrm{Card}\left\{(i,j)\quad :\quad 32\left|B_{ij}-r_2^2(i,j)\right| \ \geq\ 2\ \delta \geq\ r_2(i,m(i))+r_2(j,m(j))\right\} \quad \text{and}$$
$$\mathrm{Card}\left\{(i,j)\quad :\quad 32\left|B_{ij}-r_1^2(\sigma(i)\,\sigma(j))\right| \ \geq\ 2\ \delta \geq\ r_1(\sigma(i),m(\sigma(i)))+r_1(\sigma(j),m(\sigma(j)))\right\}.$$

Taking a permutation that minimizes the latter cardinal, one has

$$\mathrm{Card}\left\{(i,j)\ :\ 32\left|B_{ij}-r_2^2(i,j)\right| \ \geq\ 2\ \delta \geq\ r_2(i,m(i))+r_2(j,m(j))\right\}+\Phi(B,\mathbf{r}_1) \geq \widetilde{\Phi}(\mathbf{r}_1,\mathbf{r}_2)$$

by definition of Φ and $\widetilde{\Phi}$. The above inequality holds for any matrix in $\{0,1\}^{n\times n}_{sym}$, in particular for B^σ defined by $B^\sigma_{ij}=B_{\sigma(i),\sigma(j)}$ (where $B\in\{0,1\}^{n\times n}_{sym}$ and any permutation σ). Using $\Phi(B^\sigma,\mathbf{r}_1)=\Phi(B,\mathbf{r}_1)$, the above display becomes

$$\mathrm{Card}\left\{(i,j)\ :\ 32\left|B^\sigma_{ij}-r_2^2(i,j)\right| \ \geq\ 2\ \delta \geq\ r_2(i,m(i))+r_2(j,m(j))\right\}+\Phi(B,\mathbf{r}_1) \geq \widetilde{\Phi}(\mathbf{r}_1,\mathbf{r}_2)$$

and thus, choosing the permutation that minimize the left term,

$$\Phi(B,\mathbf{r}_1)+\Phi(B,\mathbf{r}_2) \geq \widetilde{\Phi}(\mathbf{r}_1,\mathbf{r}_2).$$

This generalized triangle inequality holds for all $B\in\{0,1\}^{n\times n}_{sym}$. Lemma 32 is proved. ■

D.2.5 Proof of Lemma 31.

In the following, we write $\mathbb{P}$ for $\mathbb{P}_{(\Omega,\mu,W)}$, and $\mu^{\otimes n}$ for the product measure, and $\boldsymbol{\omega}$ for the n-tuple $(\omega_1,\ldots,\omega_n)$. Lemma 31 states that for all $M\in\{0,1\}^{n\times n}_{sym}$,

$$\mathbb{P}[A=M|\boldsymbol{\omega}\in\mathcal{R}_1]=\mathbb{P}[A=M|\boldsymbol{\omega}\in\mathcal{R}_2]$$

which is equivalent to

$$p_{\mathcal{R}_1}(M)/\mathbb{P}(\mathcal{R}_1)=p_{\mathcal{R}_2}(M)/\mathbb{P}(\mathcal{R}_2) \tag{33}$$

where $p_{\mathcal{R}_1}(M)$ denotes

$$p_{\mathcal{R}_1}(M):=\mathbb{P}\left(\{A=M\}\cap\mathcal{R}_1\right)=\int_{\boldsymbol{\omega}\in\mathcal{R}_k}\mathbb{P}(A=M|\boldsymbol{\omega})\,\mathrm{d}\mu^{\otimes n}(\boldsymbol{\omega}).$$

Hence, we want to prove that

$$2p_{\mathcal{R}_1}(M)=p_{\mathcal{R}_2}(M)$$

since $2\mathbb{P}(\mathcal{R}_1)=\mathbb{P}(\mathcal{R}_2)$ by definition of the events $\mathcal{R}_1$ and $\mathcal{R}_2$.

Let $\mathcal{R}_1(k,l)$ be the the event defined by $\mathcal{R}_1 \cap \{(\omega_k, \omega_l) \in \mathcal{C}_4 \times \mathcal{C}_5\}$. Thus, the event $\mathcal{R}_1$ is the union $\cup_{1\le k\neq l\le n}\mathcal{R}_1(k,l)$. For any matrix $M = [M_{ij}]_{i,j\le n}$ in $\{0,1\}^{n\times n}_{sym}$,

$$p_{\mathcal{R}_1}(M) = \int_{\boldsymbol{\omega}\in\mathcal{R}_1} \mathbb{P}(A = M|\boldsymbol{\omega})\,\mathrm{d}\mu^{\otimes n}(\boldsymbol{\omega}) = \sum_{1\le k\neq l\le n} \int_{\boldsymbol{\omega}\in\mathcal{R}_1(k,l)} \mathbb{P}(A = M|\boldsymbol{\omega})\,\mathrm{d}\mu^{\otimes n}(\boldsymbol{\omega}).$$

Given a permutation σ of $\{1,\dots,n\}$, denote by M^σ the matrix $M^\sigma_{ij} = M_{\sigma(i),\sigma(j)}$ with $i,j \in \{1,\dots,n\}$. Write σ_{kl} for a permutation fulfilling $\sigma(n-1) = k$ and $\sigma(n) = l$. Then, the probability $p_{\mathcal{R}_1}(M)$ is equal to

$$\sum_{1\le k\neq l\le n} \int_{\boldsymbol{\omega}\in\mathcal{R}_1(k,l)} \mathbb{P}(A^{\sigma_{kl}} = M^{\sigma_{kl}}|\boldsymbol{\omega})\mathrm{d}\mu^{\otimes n}(\boldsymbol{\omega}) = \sum_{1\le k\neq l\le n} \int_{\boldsymbol{\omega}\in\mathcal{R}_1(n-1,n)} \mathbb{P}(A = M^{\sigma_{kl}}|\boldsymbol{\omega})\mathrm{d}\mu^{\otimes n}(\boldsymbol{\omega}).$$

Conditionally to $\boldsymbol{\omega}$, the entries of A for $i<j$ are independent Bernoulli variables, so that

$$p_{\mathcal{R}_1}(M) = \sum_{1\le k\neq l\le n} \int_{\boldsymbol{\omega}\in\mathcal{R}_1(n-1,n)} \prod_{1\le i<j\le n} \mathbb{P}(A_{ij} = M^{\sigma_{kl}}_{ij}|\omega_i,\omega_j)\mathrm{d}\mu^{\otimes n}(\boldsymbol{\omega}).$$

On the event $\mathcal{R}_1(n-1,n)$, the $\omega_1,\dots,\omega_{n-2}$ are in $\mathcal{C}_1\cup\mathcal{C}_1$, and (ω_{n-1},ω_n) are in $\mathcal{C}_4\times\mathcal{C}_5$. As the function W_n of the SBM is equal to $1/2$ on the diagonal blocks $\{\mathcal{C}_1,\mathcal{C}_2\}^2$ and $\{\mathcal{C}_3,\mathcal{C}_4,\mathcal{C}_5\}^2$, one has $\mathbb{P}(A_{ij} = M^{\sigma_{kl}}_{ij}|\omega_i,\omega_j) = \frac{1}{2}$ for all (i,j) in the set $\{(i,j) : i<j\le n-2\}\cup\{(n-1,n)\}$ of cardinality $g_n = n(n-1)/2 - 2(n-2)$. Hence, the probability $p_{\mathcal{R}_1}(M)$ is equal to

$$\sum_{1\le k\neq l\le n} \left(\frac{1}{2}\right)^{g_n} \int_{\boldsymbol{\omega}\in\mathcal{R}_1(n-1,n)} \prod_{1\le i\le n-2} \mathbb{P}(A_{i,n-1} = M^{\sigma_{kl}}_{i,n-1}|\omega_i,\omega_{n-1})\mathbb{P}(A_{i,n} = M^{\sigma_{kl}}_{i,n}|\omega_i,\omega_n)\mathrm{d}\mu^{\otimes n}(\boldsymbol{\omega})$$

or equivalently to

$$\sum_{1\le k\neq l\le n} \left(\frac{1}{2}\right)^{g_n} \int\limits_{(\omega_{n-1},\omega_n)\in\mathcal{C}_4\times\mathcal{C}_5} X_{M^{\sigma_{kl}}}(\omega_{n-1},\omega_n)\mathrm{d}\mu^{\otimes 2}(\omega_{n-1},\omega_n)$$

with

$$X_{M^{\sigma_{kl}}}(\omega_{n-1},\omega_n) := \prod_{1\le i\le n-2} \int_{\omega_i\in\mathcal{C}_1\cup\mathcal{C}_2} \mathbb{P}(A_{i,n-1} = M^{\sigma_{kl}}_{i,n-1}|\omega_i,\omega_{n-1})\mathbb{P}(A_{i,n} = M^{\sigma_{kl}}_{i,n}|\omega_i,\omega_n)\,\mathrm{d}\mu(\omega_i).$$

Likewise, $\mathcal{R}_2$ is the union $\cup_{1\le k<l\le n}\mathcal{R}_2(k,l)$ where each $\mathcal{R}_2(k,l)$ is the event $\mathcal{R}_2\cap\{\omega_k,\omega_l \in \mathcal{C}_3\}$. Following the same proof as for $\mathcal{R}_1$, one can show that

$$p_{\mathcal{R}_2}(M) = \sum_{1\le k<l\le n} \left(\frac{1}{2}\right)^{g_n} \int\limits_{(\omega_{n-1},\omega_n)\in\mathcal{C}_3\times\mathcal{C}_3} X_{M^{\sigma_{kl}}}(\omega_{n-1},\omega_n)\mathrm{d}\mu^{\otimes 2}(\omega_{n-1},\omega_n).$$

Lemma 34 *There exists a constant $X_{M^{\sigma_{kl}}}$ such that $X_{M^{\sigma_{kl}}}(\omega_{n-1},\omega_n) = X_{M^{\sigma_{kl}}}$ whether $\mathcal{R}_1(n-1,n)$ or $\mathcal{R}_2(n-1,n)$ holds.*

Using Lemma 34, one has

$$p_{\mathcal{R}_1}(M) = \left(\frac{1}{2}\right)^{g_n} \sum_{1\leq k\neq l\leq n} X_{M^{\sigma_{kl}}}\, \mu(\mathcal{C}_4)\, \mu(\mathcal{C}_5)$$

and

$$p_{\mathcal{R}_2}(M) = \left(\frac{1}{2}\right)^{g_n} \sum_{1\leq k< l\leq n} X_{M^{\sigma_{kl}}}\, \mu(\mathcal{C}_3)^2$$

so that $p_{\mathcal{R}_1}(M) = p_{\mathcal{R}_2}(M)/2$, since $\mu(\mathcal{C}_4) = \mu(\mathcal{C}_5) = \mu(\mathcal{C}_3)/2$ (by construction of the SBM). Lemma 31 is proved. ■

D.2.6 Proof of Lemma 34.

For brevity, write $\mathbb{P}$ for $\mathbb{P}_{(\Omega,\mu,W)}$ in the following. By definition, $X_{M^{\sigma_{kl}}}(\omega_{n-1},\omega_n)$ is the product of the $n-2$ following terms

$$\int_{\omega_i\in\mathcal{C}_1\cup\mathcal{C}_2} \mathbb{P}(A_{i,n-1} = M^{\sigma_{kl}}_{i,n-1}|\omega_i,\omega_{n-1})\, \mathbb{P}(A_{i,n} = M^{\sigma_{kl}}_{i,n}|\omega_i,\omega_n)\, \mathrm{d}\mu(\omega_i)$$

$i = 1,\ldots,n-2$. The above display is equal to

$$\begin{aligned}&\int_{\omega_i\in\mathcal{C}_1} \mathbb{P}(A_{i,n-1} = M^{\sigma_{kl}}_{i,n-1}|\omega_i,\omega_{n-1})\mathbb{P}(A_{i,n} = M^{\sigma_{kl}}_{i,n}|\omega_i,\omega_n)\mathrm{d}\mu(\omega_i)\\ &+\int_{\omega_i\in\mathcal{C}_2} \mathbb{P}(A_{i,n-1} = M^{\sigma_{kl}}_{i,n-1}|\omega_i,\omega_{n-1})\mathbb{P}(A_{i,n} = M^{\sigma_{kl}}_{i,n}|\omega_i,\omega_n)\mathrm{d}\mu(\omega_i).\end{aligned}$$

If $(\omega_{n-1},\omega_n)\in\mathcal{C}_4\times\mathcal{C}_5$, then

$$\begin{aligned}&= \int_{\omega_i\in\mathcal{C}_1} [1/2 + (2M^{\sigma_{kl}}_{i,n-1} - 1)\,\delta]\,[1/2 + (2M^{\sigma_{kl}}_{i,n} - 1)\,(1/2)]\,\mathrm{d}\mu(\omega_i)\\ &+ \int_{\omega_i\in\mathcal{C}_2} [1/2 - (2M^{\sigma_{kl}}_{i,n-1} - 1)\,\delta]\,[1/2 - (2M^{\sigma_{kl}}_{i,n} - 1)\,(1/2)]\,\mathrm{d}\mu(\omega_i)\end{aligned}$$

which is equal to $\left[1/2 + (2M^{\sigma_{kl}}_{i,n-1} - 1)(2M^{\sigma_{kl}}_{i,n} - 1)\delta\right]\mu(\mathcal{C}_1)$, since $\mu(\mathcal{C}_1) = \mu(\mathcal{C}_2)$.
If $(\omega_{n-1},\omega_n)\in\mathcal{C}_3\times\mathcal{C}_3$, then

$$\begin{aligned}&= \int_{\omega_i\in\mathcal{C}_1} [1/2 + (2M^{\sigma_{kl}}_{i,n-1} - 1)\,\sqrt{\delta/2}][1/2 + (2M^{\sigma_{kl}}_{i,n} - 1)\,\sqrt{\delta/2}]\,\mathrm{d}\mu(\omega_i)\\ &+ \int_{\omega_i\in\mathcal{C}_2} [1/2 - (2M^{\sigma_{kl}}_{i,n-1} - 1)\,\sqrt{\delta/2}][1/2 - (2M^{\sigma_{kl}}_{i,n} - 1)\,\sqrt{\delta/2}]\,\mathrm{d}\mu(\omega_i)\end{aligned}$$

which is equal to $\left[1/2 + (2M^{kl}_{i,n-1} - 1)(2M^{\sigma_{kl}}_{i,n} - 1)\delta\right]\mu(\mathcal{C}_1)$.

Hence $X_{M^{\sigma_{kl}}}(\omega_{n-1},\omega_n)$ is equal to the same constant whether (ω_{n-1},ω_n) belongs to $\mathcal{C}_3\times\mathcal{C}_3$ or $\mathcal{C}_4\times\mathcal{C}_5$. Lemma 34 is proved. ■

Appendix E. Estimation of the Minkowski Dimension

E.1 Proof of Theorem 12: the Upper Bound

Theorem 12 is a corollary of Theorem 35, which gives non-asymptotic high-probability bounds for the risk of the data-function $-\log \widehat{N}_{\Omega}^{(c)}(\epsilon)/\log \epsilon$.

Theorem 35 *Assume the graphon (Ω, μ, W) satisfies $H_1^{\alpha,v}$ and $H_2^{m,M,v}$ and has a Minkowski dimension $d \in (0, \infty)$. If n is large enough to satisfy the below inequality*

$$2\log n/n \leq \alpha\,(v/14)^{2d} \wedge (v/14)^4\,,$$

then the following holds with probability at least $1-(2+4\alpha M)/n$ with respect to the distribution $\mathbb{P}_{(\Omega,\mu,W)}$. The sum of the distance error bound (14) and the sampling error (13) is upper bounded as follows

$$\frac{b_{sup}+s_\omega}{6} \leq err_{n,d} := \left(\frac{\log n}{n}\right)^{1/4} + \left(\frac{2\log n}{\alpha\, n}\right)^{1/2d}. \tag{34}$$

For all $\epsilon \in (2(b_{sup}+s_\omega), v/7]$, the covering number estimator $\widehat{N}_{\Omega}^{(c)}(\epsilon)$ satisfies the following upper bound

$$\left|\frac{\log \widehat{N}_{\Omega}^{(c)}(\epsilon)}{-\log \epsilon} - d\right| \leq \frac{1}{-\log \epsilon}\left[\log\left(M \vee \frac{1}{m}\right) + 6d\frac{err_{n,d}}{\epsilon}\left(1+\frac{err_{n,d}}{\epsilon}\right)\right]$$

Theorem 12 follows from Theorem 35 by choosing any radius radius that minimizes the above upper bound, that is, any radius ϵ_D of the order of $\sup_{\{d:\, d\leq D\}} err_{n,d} = err_{n,D}$. ■

Comments on Theorem 35 : We first remark that the above theorem based on the covering number can also be adapted to the packing number (without difficulties). We now comment on the two additive error terms in the upper bound. The term $-\log\left(M \vee (1/m)\right)/\log\epsilon$ stands for the gap between the Minkowski dimension and the quantity that we actually estimate, i.e. $-\log N_{\Omega}^{(c)}(\epsilon)/\log\epsilon$. This gap depends on the parameters of the assumption $H_2^{m,M,v}$. The second error term $-d\, err_{n,d}/(\epsilon\log\epsilon)$ represents the gap between the latter estimated quantity and the estimator $-\log \widehat{N}_{\Omega}^{(c)}(\epsilon)/\log\epsilon$. To control this gap, we need to estimate the covering number correctly, and thus to control the error sum $b_{sup}+s_\omega$ involved in Proposition 11. Actually, the theorem ensures that this error sum is smaller than $err_{n,d}$. This comes from the fact that the difference between the sample $\omega_1,\ldots,\omega_n$ and the latent space Ω is not too large, thanks to the assumption $H_1^{\alpha,v}$. See the proof below for details.

Finally, the upper bound holds with probability at least $1-2/n-4\,\alpha\, M/n$. The first quantity $2/n$ corresponds to the event $\mathcal{E}_{dist}^c$ defined in Theorem 7, i.e. that the the distance estimator does not satisfy the distance error bound b_{sup}. The second quantity $4\,\alpha\, M/n$ corresponds to the probability of the event where the sampled points do not cover well the latent space, leading to a large sampling error s_ω and a large distance error bound b_{sup}. This event, denoted by $\mathcal{E}_{bad}$, is rigorously defined in the following proof.

Proof of Theorem 35. Assume the event $\mathcal{E}_{dist}$ of Theorem 7 holds, that is the errors of distance-estimator are uniformly bounded by b_{sup}. On this event, Proposition 11 gives

$$N_{\Omega}^{(c)}\left(\epsilon + b_{sup} + s_{\omega}\right) \leq \widehat{N}_{\Omega}^{(c)}\left(\epsilon\right) \leq N_{\Omega}^{(c)}\left(\epsilon - b_{sup} - s_{\omega}\right)$$

for all $\epsilon \in (b_{sup} + s_{\omega}, 1)$, so that

$$\frac{\log N_{\Omega}^{(c)}(\epsilon + s_{\omega} + b_{sup})}{-\log\epsilon} - d \leq \frac{\log \widehat{N}_{\Omega}^{(c)}(\epsilon)}{-\log\epsilon} - d \leq \frac{\log N_{\Omega}^{(c)}(\epsilon - s_{\omega} - b_{sup})}{-\log\epsilon} - d.$$

As the assumption $H_2^{m,M,v}$ is valid in the neighborhood $(0, v]$, we need to check that $\epsilon + s_{\omega} + b_{sup} \in (0, v]$ to use this assumption. For clarity, we do this verification at the end of the proof. Hence, using $H_2^{m,M,v}$, one has

$$\frac{\log m}{-\log\epsilon} - d\left[\frac{\log(\epsilon + s_{\omega} + b_{sup})}{-\log\epsilon} + 1\right] \leq \frac{\log \widehat{N}_{\Omega}^{(c)}(\epsilon)}{-\log\epsilon} - d \leq \frac{\log M}{-\log\epsilon} - d\left[\frac{\log(\epsilon - s_{\omega} - b_{sup})}{-\log\epsilon} + 1\right] \tag{35}$$

In the right hand side of (35), the right term is upper bounded by

$$-d\left[\frac{\log(\epsilon - s_{\omega} - b_{sup})}{-\log\epsilon} + 1\right] \leq -d\,\frac{\log\big(1 - (s_{\omega} + b_{sup})/\epsilon\big)}{-\log\epsilon}$$

which is again upper bounded by

$$d\,\frac{(s_{\omega} + b_{sup})/\epsilon + \left((s_{\omega} + b_{sup})/\epsilon\right)^2}{-\log\epsilon}$$

if $(s_{\omega} + b_{sup}) \leq \epsilon/2$. Similarly in the left hand side of (35), the right term is lower bounded by

$$-d\left[\frac{\log(\epsilon + s_{\omega} + b_{sup})}{-\log\epsilon} + 1\right] \geq -d\frac{(s_{\omega} + b_{sup})/\epsilon}{-\log\epsilon}.$$

Combining the above displays, one derive

$$\frac{\log m}{-\log\epsilon} - d\frac{(s_{\omega} + b_{sup})/\epsilon}{-\log\epsilon} \leq \frac{\log \widehat{N}_{\Omega}^{(c)}(\epsilon)}{-\log\epsilon} - d \leq \frac{\log M}{-\log\epsilon} + d\,\frac{(s_{\omega} + b_{sup})/\epsilon + \left((s_{\omega} + b_{sup})/\epsilon\right)^2}{-\log\epsilon}. \tag{36}$$

It remains to upper bound the error sum $s_{\omega} + b_{sup}$ in (36). Given a cover of Ω, composed of $N_{\Omega}^{(c)}(\eta)$ balls B_j of radius η, one define the following event

$$\mathcal{E}_{bad}(\eta) := \Big\{\exists j : B_j \text{ contains exactly 0 or 1 sampled point among } \omega_1, \ldots, \omega_n\Big\}. \tag{37}$$

Assume the complementary event $\mathcal{E}_{bad}^c(\eta)$ holds. This means that each ball of the cover of Ω contains at least two sampled points. Hence, one has

$$\begin{aligned} &s_{\omega} \leq 2\eta, \\ &\sup_{i\in\{1,..,n\}} r_W(\omega_i, \omega_{m(i)}) \leq 2\eta. \end{aligned}$$

which directly implies the following upper bound

$$b_{sup} + s_\omega \leq 6 \left(\frac{\log n}{n}\right)^{1/4} + 4\sqrt{\eta} + 2\eta$$

by definition of b_{sup} in (14). Thus, for the particular radius $\eta_n := [2\log(n)/(\alpha\, n)]^{1/d}$,

$$b_{sup} + s_\omega \leq 6 \left(\frac{\log n}{n}\right)^{1/4} + 6 \left(\frac{2\log n}{\alpha\, n}\right)^{1/2d}.$$

It follows from the definition (34) of $err_{n,d}$ that

$$s_\omega + b_{sup} \leq 6 err_{n,d}.$$

Combining the above upper bound with (36), one deduce the inequalities of the theorem.

The above displays hold conditionally to the event $\mathcal{E}_{bad}^c(\eta_n) \cap \mathcal{E}_{dist}$, which happens with probability at least $1 - (2 + 4\alpha M)/n$ (Lemma 36).

Lemma 36 *The probability $\mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}_{bad}(\eta_n) \cup \mathcal{E}_{dist}^c)$ is smaller than $(2 + 4\alpha M)/n$.*

The condition $\epsilon + s_\omega + b_{sup} \in (0, v]$ (used at the beginning of the proof) is satisfied (Lemma 37).

Lemma 37 *On the event $\mathcal{E}_{bad}^c(\eta_n) \cap \mathcal{E}_{dist}$, one has $\epsilon + s_\omega + b_{sup} \in (0, v]$.*

Theorem 35 is proved. ■

E.1.1 Proof of Lemma 37.

We want to prove that $\epsilon + s_\omega + b_{sup} \in (0, v]$ on the event $\mathcal{E}_{bad}^c(\eta_n) \cap \mathcal{E}_{dist}$. We have already seen that $s_\omega + b_{sup} \leq 6 err_{n,d}$ on this event, so it is enough to prove that $\epsilon + 6 err_{n,d} \leq v$. By assumption in Theorem 35, one has

$$2\frac{\log n}{n} \leq \alpha \left(\frac{v}{14}\right)^{2d} \wedge \left(\frac{v}{14}\right)^{4},$$

which implies

$$\left(\frac{2\log n}{\alpha\, n}\right)^{1/2d} \leq v/14 \quad \text{and} \quad \left(\frac{\log n}{n}\right)^{1/4} \leq v/14,$$

and thus

$$err_{n,d} \leq v/7.$$

Finally, one has $\epsilon + 6 err_{n,d} \leq v$, since $\epsilon \leq v/7$ by assumption. Hence, $\epsilon + s_\omega + b_{sup} \in (0, v]$ on the event $\mathcal{E}_{bad}^c(\eta_n) \cap \mathcal{E}_{dist}$. The lemma is proved. ■

E.1.2 Proof of Lemma 36.

Let us upper bound the probability $\mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}_{bad}(\eta_n) \cup \mathcal{E}^c_{dist})$. The union bound gives

$$\mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}_{bad}(\eta_n) \cup \mathcal{E}^c_{dist}) \leq \mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}_{bad}(\eta_n)) + \mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}^c_{dist}) \leq \mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}_{bad}(\eta_n)) + \frac{2}{n}$$

where the last inequality comes from Theorem 7. If the cover defined by $\mathcal{E}_{bad}(\eta_n)$ satisfies the condition (39), then Lemma 38 ensures that

$$\mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}_{bad}(\eta_n)) \leq 2N^{(c)}_{\Omega}(\eta_n) n\beta \exp[-\beta(n-1)]. \tag{38}$$

Lemma 38 *Let $B_1, \ldots, B_N$ be N balls in (Ω, r_W) of measure (strictly) larger than $1/n$, that is,*

$$\min_{j \leq N} \mu(B_j) \geq \beta > 1/n \tag{39}$$

for some real β. Then the probability that (at least) one ball contains exactly zero or one sampled point is smaller than

$$2Nn\beta \exp[-\beta(n-1)].$$

Assume that $\eta_n \in (0, v]$ to use the assumption $H_1^{\alpha,v}$. Then, one obtain the following lower bound for the cover defined by $\mathcal{E}_{bad}(\eta_n)$,

$$\mu(B_j) \geq \alpha\ \eta_n^d = 2\log(n)/n$$

so that assumption (39) is satisfied. Applying Lemma 38 for $\beta = \alpha\ \eta_n^d$, one has

$$\mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}_{bad}(\eta_n)) \leq 2N^{(c)}_{\Omega}(\eta) n\alpha\ \eta^d \exp\left[-\alpha\ \eta^d(n-1)\right].$$

Combining with the inequality $N^{(c)}_{\Omega}(\eta) \leq M\eta^{-d}$ from assumption $H_2^{m,M,v}$, one derive

$$\mathbb{P}_{(\Omega,\mu,W)}(\mathcal{E}_{bad}(\eta_n)) \leq 2Mn\alpha\ \exp\left[-\alpha\ \eta^d(n-1)\right]$$

and since $\alpha\, \eta_n^d = 2\log(n)/n$, one obtain the upper bound

$$2Mn\alpha\ \exp\left[-\frac{2\log n}{n}(n-1)\right].$$

The above display is finally smaller than

$$4Mn\alpha \exp\left[-2\log n\right] \leq (4M\alpha)/n.$$

To conclude the proof, it remains to check the condition $\eta_n \in (0, v]$ that we assume earlier. The following assumption of Theorem 35

$$2\frac{\log n}{n} \leq \alpha \left(\frac{v}{14}\right)^{2d} \wedge \left(\frac{v}{14}\right)^4$$

ensures that the radius $\eta_n = [2\log(n)/(\alpha\, n)]^{1/d}$ satisfies the condition $\eta_n \in (0, v]$. Lemma 36 is proved. ■

E.1.3 Proof of Lemma 38.

Given N balls $B_1, \ldots, B_N$, let us upper bound the probability that (at least) one of the balls contains exactly zero or one sampled point ω_i. With the union bound, this probability is lower than

$$\sum_{j=1}^{N} \mathbb{P}_{(\Omega,\mu,W)} \left\{ B_j \text{ contains exactly 0 or 1 sampled point among } \omega_1, \ldots, \omega_n \right\}$$

which is again upper bounded with the union bound by

$$\sum_{j=1}^{N} \mathbb{P}_{(\Omega,\mu,W)} \left\{ B_j \text{ contains exactly 0 point} \right\} + \sum_{j=1}^{N} \mathbb{P}_{(\Omega,\mu,W)} \left\{ B_j \text{ contains exactly 1 point} \right\}.$$

Since the probability of the event $\{B_j$ contains exactly 0 point$\}$ is equal to $(1-\mu(B_j))^n$, and since the probability of $\{B_j$ contains exactly 1 point$\}$ is $n\mu(B_j)(1-\mu(B_j))^{n-1}$, the above sum is upper bounded by

$$\sum_{j=1}^{N} (1-\mu(B_j))^n + \sum_{j=1}^{N} n\mu(B_j)(1-\mu(B_j))^{n-1}.$$

Combining the assumption $\mu(B_j) \geq \beta > 1/n$ with the monotonicity of the functions $x \mapsto (1-x)^n$ and $x \mapsto nx(1-x)^{n-1}$ on $(1/n, 1)$, one has the following upper bound

$$N\left[(1-\beta)^n + n\beta(1-\beta)^{n-1}\right]$$

which is lower than $2Nn\beta(1-\beta)^{n-1} \leq 2Nn\beta \exp[-\beta(n-1)]$. Lemma 38 is proved. ■

E.2 Lower Bound and Minimal Conditions

E.2.1 Proof of Theorem 13.

From (Falconer, chap.2), we deduce directly the following lemma.

Lemma 39 *Given $L > 1$ and $n \geq 2$, there exists a set $\Omega_0 \subset (0, 1/(Ln)) \times (0, 1/(Ln))$ with Minkowski dimension $d_2 = 1 + \log^{-1}(n)$ w.r.t the Euclidean distance of $[0,1]^2$, and a probability measure μ_0 on Ω_0.*

Based on (Ω_0, μ_0) described in Lemma 39, we construct two graphons that are difficult to distinguish for any estimator.

- $\Omega_1 = (0,1) \times \{0\} \subset [0,1]^2$ endowed with the uniform measure λ on $(0,1)$. In particular, $\lambda((0,1) \times \{0\}) = 1$.
- $\Omega_2 = \Omega_1 \cup \Omega_0 \subset [0,1]^2$ endowed with the probability measure:

$$\mu_2 = (1-n^{-1})\lambda + n^{-1}\mu_0.$$

Consider a symmetric function $W : [0,1]^2 \times [0,1]^2 \to [0,1]$ satisfying a double Hölder condition (6) with Hölder exponent $\alpha = 1$. Then, Appendix A.2 shows that the neighborhood distance (associated with such a W) behaves like the euclidean distance on $[0,1]^2$, i.e.:

$$r_W(\omega, \omega') \asymp ||\omega - \omega||_2$$

for all $\omega, \omega' \in [0,1]^2$. Hence, (Ω_1, λ, W) and (Ω_2, μ, W) satisfy $dim\, \Omega_2 = 1 + \log^{-1}(n)$ and $dim\, \Omega_1 = 1$, respectively. For brevity, we denote these dimensions by d_2 and d_1 in the following.

Let us check that all conditions of $\mathcal{W}_n(D, \alpha, m, M, v)$ are satisfied by both graphons (Ω_1, λ, W) and (Ω_2, μ_2, W). It is clear that (Ω_1, λ, W) belongs to the set $\mathcal{W}_n(D, \alpha, m, M, v)$ for large enough M and small enough α, m. For the graphon (Ω_2, μ_2, W), one has:

- Assumption $H_1^{\alpha,v}$: for any point $\omega \in \Omega_0$, note $\omega_{proj} \in \Omega_1$ its closest point in Ω_1. As $\Omega_0 \subset (0, 1/(Ln))^2$, we have $r_W(\omega, \omega_{proj}) \leq 1/(2n)$ for large enough L. Then, for all $\epsilon > 1/n$ and all $\omega \in \Omega_0$, one has

$$\begin{aligned}\mu_2\left[B(\omega, \epsilon)\right] &\geq (1 - n^{-1})\lambda\left[B(\omega, \epsilon)\right] \\ &\geq (1 - n^{-1})\lambda\left[B(\omega_{proj}, \epsilon - 1/(2n))\right] \\ &\geq \frac{1}{2}\lambda\left[B(\omega_{proj}, \epsilon/2)\right].\end{aligned}$$

  which is larger than ϵ (up to a numerical constant) since (Ω_1, λ, W) satisfies the condition $H_1^{\alpha,v}$ for all $\epsilon > 0$.

- Assumption $H_2^{m,M,v}$ lower bound: $N^{(c)}_{\Omega_2}(\epsilon) \gtrsim N^{(c)}_{\Omega_1}(\epsilon) \gtrsim \epsilon^{-d_1}$ which is larger than $\epsilon^{-d_2 + \log^{-1}(n)} \gtrsim \epsilon^{-d_2}$ because $\epsilon^{\log^{-1}(n)} \asymp 1$ for all $\epsilon \in (1/n, 1)$.

- Assumption $H_2^{m,M,v}$ upper bound: $N^{(c)}_{\Omega_2}(\epsilon) \lesssim N^{(c)}_{\Omega_1}(\epsilon) + N^{(c)}_{\Omega_0}(\epsilon) \lesssim N^{(c)}_{\Omega_1}(\epsilon)$ since $N^{(c)}_{\Omega_0}(\epsilon) \lesssim N^{(c)}_{\Omega_0}(1/n) = 1$ for $\epsilon > 1/n$ and large enough L. Combining with the fact that (Ω_1, λ, W) satisfies $H_2^{m,M,v}$, one obtain $N^{(c)}_{\Omega_2}(\epsilon) \lesssim \epsilon^{-d_1}$.

Thus, both graphons (Ω_1, λ, W) and (Ω_2, μ_2, W) fulfill all conditions of $\mathcal{W}_n(D, \alpha, m, M, v)$ for large enough constants L, M and small enough constants α, m.

We define the event $\mathcal{E}_{\Omega_1}$ where the i.i.d. sample $\omega_1, \ldots, \omega_n$ is such that all points $\omega_1, \ldots, \omega_n$ belong to Ω_1. In particular, for the graphon (Ω_2, μ_2, W), the probability of this event is larger than

$$\mu_2[\mathcal{E}_{\Omega_1}] \geq (1 - n^{-1})^n \geq \frac{1}{3}.$$

Then, for any estimator $\hat{d}$ based on the adjacency matrix A, one has

$$\begin{aligned}&\mathbb{P}_{(\Omega_2, \mu_2, W)}\left[|\hat{d} - dim\, \Omega_2| \geq \frac{1}{2}\log^{-1}(n)\right] \\ &\qquad \geq \mathbb{P}_{(\Omega_2, \mu_2, W)}\left[|\hat{d} - dim\, \Omega_2| \geq \frac{1}{2}\log^{-1}(n) \middle| \mathcal{E}_{\Omega_1}\right]\mu_2(\mathcal{E}_{\Omega_1})\end{aligned}$$

which is larger than

$$\frac{1}{3}\mathbb{P}_{(\Omega_1,\lambda,W)}\Big[|\hat{d}-dim\,\Omega_1|\leq\frac{1}{2}\log^{-1}(n)\Big]$$

since $|\hat{d}-dim\,\Omega_1|\leq\frac{1}{2}\log^{-1}(n)$ implies $|\hat{d}-dim\,\Omega_2|\geq\frac{1}{2}\log^{-1}(n)$. Thus, by writting

$$p:=\mathbb{P}_{(\Omega_1,\lambda,W)}\Big[|\hat{d}-dim\,\Omega_1|>\frac{1}{2}\log^{-1}(n)\Big],$$

the above displays entail

$$\mathbb{P}_{(\Omega_2,\mu_2,W)}\Big[|\hat{d}-dim\,\Omega_2|\geq\frac{1}{2}\log^{-1}(n)\Big]\geq\frac{1}{3}(1-p)$$

which imply that

$$\begin{aligned}\sup_{\mathcal{W}_n(D,\alpha,m,M,v)}&\mathbb{P}_{(\Omega,\mu,W)}\Big[|\hat{d}-dim\,\Omega|\geq\frac{1}{2}\log^{-1}(n)\Big]\\&\geq\max_{(\Omega_1,\lambda,W),(\Omega_2,\mu_2,W)}\mathbb{P}_{(\Omega,\mu,W)}\Big[|\hat{d}-dim\,\Omega|\geq\frac{1}{2}\log^{-1}(n)\Big]\\&\geq p\vee\frac{1-p}{3}\end{aligned}$$

which is larger than 1/4. Theorem 13 is proved. ■

E.2.2 Proof of Theorem 14.

There are two cases.

For the class $\mathcal{W}_n^{min(1)}(D,\alpha,m,M,v)$, the condition $H_1^{\alpha,v}$ is not imposed. Consider the two following graphons.

- (Ω_1,λ,W) where $\Omega_1=[0,1]\times\{0\}^{D-1}$ is endowed with the uniform measure λ on $[0,1]$, with $\lambda(\Omega_1)=1$, and where $W:[0,1]^D\times[0,1]^D\to[0,1]$ is a symmetric function that satisfies a double Hölder condition (6) with Hölder exponent $\alpha=1$.
- (Ω_2,μ_2,W) where $\Omega_2=[0,1]^D$ and $\mu_2=(1-n^{-1})\lambda+n^{-1}\nu$, with ν the uniform measure on $[0,1]^D$.

Following the proof of Theorem 13, we can show that these two graphons belong to $\mathcal{W}_n^{min(1)}(D,\alpha,m,M,v)$, and that

$$\sup_{\mathcal{W}_n^{min(1)}(D,\alpha,m,M,v)}\mathbb{P}_{(\Omega,\mu,W)}\Big[|\hat{d}-dim\,\Omega|\geq\frac{D}{2}\Big]\geq\frac{1}{4}\tag{40}$$

which gives the error bound of Theorem 14.

For the class $\mathcal{W}_n^{min(2)}(D,\alpha,m,M,v)$, the assumption $H_2^{m,M,v}$ is not assumed. As in the proof of Theorem 13, we can see that the two following graphons belong to $\mathcal{W}_n^{min(2)}(D,\alpha,m,M,v)$.

- (Ω_1,λ,W) as defined in the above case.

- (Ω_2, μ_2, W) where $\Omega_2 = [0, 1/(Ln)]^D$ for some large enough (numerical) constant L, and $\mu_2 = (1 - n^{-1})\lambda + n^{-1}\nu$, with ν the uniform measure on $[0, 1/(Ln)]^D$.

As before, one can prove the error bound (40) over the class $\mathcal{W}_n^{min(2)}(D, \alpha, m, M, v)$.

Thus, (40) is proved for $\mathcal{W}_n^{min(j)}(D, \alpha, m, M, v)$, with $j \in \{1, 2\}$, and Theorem 14 follows. ■

Appendix F. Estimation with Sparse Observations

F.1 Proof of Corollary 19 : Estimation of the Distances

Corollary 19 is a reformulation of Theorem 7 in the sparse setting and their proofs are almost identical. In this appendix, denote by W_n the function $\rho_n W$. Accordingly, r_{W_n} denotes the neighborhood distance (4) where W has been replaced with W_n. Hence, $r_{W_n} = \rho_n r_W$.

Corollary 19 is a direct consequence of the two following Lemmas.

Lemma 40 *For $\rho_n \geq 2\sqrt{\frac{\log n}{n-2}}$ and $n \geq 5$, the following event*

$$\mathcal{E}_{in}^{sp} := \left\{ \forall i, j \in [n] : \ \ |\langle A_i, A_j \rangle_n - \langle W_n(\omega_i, .), W_n(\omega_j, .) \rangle| \leq 5\, \rho_n \sqrt{\frac{\log n}{n}} \right\}$$

holds with probability at least $1 - \frac{2}{n}$ with respect to the distribution $\mathbb{P}_{(\Omega,\mu,W),\rho_n}$.

Following the proof of Proposition 26, we show Lemma 40 below, by replacing Hoeffding inequality with Bernstein inequality, in order to benefit from the small variance of A_{ij} (which is now of the order of ρ_n).

Lemma 41 *Conditionally to the event $\mathcal{E}_{in}^{sp}$, the following inequalities*

$$\forall i \in [n] : \ \ \left| \langle A_i, A_{\widehat{m}(i)} \rangle_n - \langle W_n(\omega_i, .), W_n(\omega_i, .) \rangle \right| \leq 3\rho_n\, r_{W_n}(\omega_i, \omega_{m(i)}) \, + \, 25\, \rho_n \sqrt{\log(n)/n}$$

hold simultaneously.

The proof of lemma 41 is almost the same as for Proposition 27. It is omitted. ■

F.1.1 Proof of Lemma 40

Conditionally to ω_i, ω_j, $i \neq j$, the $n-2$ random variables $\{A_{ik}A_{kj} : k \in [n], k \neq i, j\}$ are independent with expectation $\mathbb{E}[A_{ik}A_{kj}] = \int_\Omega W_n(\omega_i, z) W_n(\omega_j, z) \mu(dz)$ for all $k \neq i, j$ (where $\mathbb{E}$ is taken w.r.t. the distribution $\mathbb{P}_{(\Omega,\mu,W),\rho_n}$). Using Bernstein inequality (see Sridharan, 2002, for instance), one has

$$\mathbb{P}_{(\Omega,\mu,W),\rho_n}\left(\frac{1}{n-2} \Big| \sum_{k \neq i,j} A_{ik}A_{kj} - (n-2) \int_\Omega W_n(\omega_i, z) W_n(\omega_j, z) \mu(dz) \Big| \geq \epsilon \ \Big| \omega_i, \omega_j \right)$$

smaller than

$$\leq 2\exp\left(\frac{-(n-2)\epsilon^2}{2\rho_n^2 + 2\epsilon/3} \right)$$

for $\epsilon > 0$. Since the above inequality is satisfied for almost every $\omega_i, \omega_j \in \Omega$, we have the same upper bound for the non-conditional probability. Then, setting $\epsilon = 3\rho_n\sqrt{\frac{\log n}{n-2}}$ gives

$$2\exp\left(\frac{-(n-2)\epsilon^2}{2\rho_n^2 + 2\epsilon/3}\right) \leq 2\exp\left(\frac{-9\log n}{2 + \frac{2}{\rho_n}\sqrt{\frac{\log n}{n-2}}}\right) \leq \frac{2}{n^3}$$

since $\rho_n \geq 2\sqrt{\frac{\log n}{n-2}}$ by assumption. Thus, by using the union bound over all $i \neq j$, one obtain

$$\mathbb{P}_{(\Omega,\mu,W),\rho_n}\left(\bigcup_{i,j:i\neq j}\left\{\frac{1}{n-2}\left|\sum_{k\neq i,j} A_{ik}A_{kj} - (n-2)\int_\Omega W_n(\omega_i,z)W_n(\omega_j,z)\mu(dz)\right| \geq \epsilon\right\}\right) \leq \frac{2}{n}.$$

And finally, following the proof of Proposition 26 leads to

$$\max_{i,j:i\neq j}\left|\sum_k \frac{A_{ik}A_{kj}}{n} - \int_\Omega W_n(\omega_i,z)W_n(\omega_j,z)\mu(dz)\right| \leq 3\rho_n\sqrt{\frac{\log n}{n-2}} + \frac{4}{n}$$

with probability at least $1 - \frac{2}{n}$. To conclude the proof of Lemma 40, observe that above display is upper bounded by

$$\leq 5\rho_n\sqrt{\frac{\log n}{n}}$$

as soon as $n \geq 5$. ■

F.2 Proof of of Corollary 20 : Estimation of the Dimension

In the proof of Theorem 35, one has seen

$$\frac{\log m}{-\log\epsilon} - d\frac{(s_\omega + b_{sup})/\epsilon}{-\log\epsilon} \leq \frac{\log \widehat{N}_\Omega^{(c)}(\epsilon)}{-\log\epsilon} - d \leq \frac{\log M}{-\log\epsilon} + d\,\frac{(s_\omega + b_{sup})/\epsilon + ((s_\omega + b_{sup})/\epsilon)^2}{-\log\epsilon}.$$

The sampling error s_ω is not affected by the sparsification of the data through ρ_n, and thus takes the same value as in Theorem 35. On the other hand, the distance error bound b_{sup} changes, and is now defined as

$$b_{sup}^2 := 6\max_{1\leq i\leq n} r_W(\omega_i, \omega_{m(i)}) + \frac{60}{\rho_n}\sqrt{\log(n)/n}$$

according to Corollary 19. Following the proof of Theorem 35, one has

$$b_{sup} + s_\omega \leq 6\left(\frac{2\log n}{\alpha\, n}\right)^{1/2d} + \frac{8}{\sqrt{\rho_n}}\left(\frac{\log n}{n}\right)^{1/4}.$$

Define

$$err_{n,d,\rho_n} := \left(\frac{2\log n}{\alpha\, n}\right)^{1/2d} + \frac{1}{\sqrt{\rho_n}}\left(\frac{\log n}{n}\right)^{1/4} \tag{41}$$

so that

$$b_{sup} + s_\omega \leq 8\, err_{n,d,\rho_n}.$$

Following the proof of Theorem 35, one obtain the same error bound for the dimension estimation, after replacing $6\, err_{n,d}$ with $8\, err_{n,d,\rho_n}$. Indeed, one has

$$\left| \frac{\log \widehat{N}_\Omega^{(c)}(\epsilon)}{-\log \epsilon} - d \right| \leq \frac{1}{-\log \epsilon} \left[\log \left(M \vee \frac{1}{m} \right) + 8d \frac{err_{n,d}}{\epsilon} \left(1 + \frac{err_{n,d}}{\epsilon} \right) \right] \tag{42}$$

for all $\epsilon \in]0, v/9]$ and all n such that

$$2\log n/n \leq \alpha\, (v/18)^{2d} \wedge \rho_n^2\, (v/18)^4\,.$$

As in the proof of Theorem 12, one minimizes the error bound (42) by choosing a particular radius of the order of $\sup_{\{d:\, d \leq D\}} err_{n,d,\rho_n} = err_{n,D,\rho_n}$. This gives a radius that satisfies the following relation

$$\epsilon_{D,\rho_n} \asymp \left(\frac{\log n}{n} \right)^{1/(2D)} \vee \frac{1}{\sqrt{\rho_n}} \left(\frac{\log n}{n} \right)^{1/4}.$$

Corollary 20 follows from the plug-in of ϵ_{D,ρ_n} in (42). ■

Appendix G. Testing the Complexity via Under-estimation of the Packing Number

The current appendix is organized as follows. We first analyse the performance of the new distance estimator (19) and then deduce a control on the type I and II errors of the test.

G.1 Performance of the New Distance Estimator

Lemma 42 shows that the new distance-estimator $\widehat{r}_{new}$ does not over-estimate r_W in the sense of (44), without underestimating too much (45). Let U be the function defined by

$$U(i) = \text{argmax}_{t \in \{i, \widehat{m}(i)\}} \langle W(\omega_t, .), W(\omega_t, .) \rangle \tag{43}$$

for all $i \in [n]$. This means that $U(i)$ indicates which of the two functions $W(\omega_i, .)$ or $W(\omega_{\widehat{m}(i)}, .)$ has the largest l_2-norm $||.||_{2,\mu}$ (see Section 4.1 for the definitions of the inner product and the norm).

Lemma 42 *Consider $t_n = 12\sqrt{\frac{\log n}{n}}$ a fluctuation term and the function U introduced in* (43)*. One has the following bounds on the new distance estimator* (19)

$$\widehat{r}_{new}^2(i,j) \leq r_W^2(\omega_{U(i)}, \omega_{U(j)}) + t_n \tag{44}$$

$$\widehat{r}_{new}^2(i,j) \geq r_W^2(\omega_i, \omega_j) - 5\, r_W(\omega_i, \omega_{m(i)}) - 5\, r_W(\omega_j, \omega_{m(j)}) - 5t_n \tag{45}$$

holding simultaneously for all $i, j \in [n]$ with probability at least $1 - \frac{2}{n}$ with respect to the distribution $\mathbb{P}_{(\Omega,\mu,W)}$.

Let us prove Lemma 42. Recall the useful Proposition 26 on the convergence of the inner products: the event $\mathcal{E}_{in}$ where the following inequalities hold simultaneously for all $i \neq j$

$$|\langle A_i, A_j\rangle_n - \langle W(\omega_i, .), W(\omega_j, .)\rangle\,| \leq 3\sqrt{\log n/n} \tag{46}$$

happens with probability at least $1 - 2/n$.

Proof of (44)**.** Assume the above event $\mathcal{E}_{in}$ holds. For all $i, j \in [n]$ such that $\{i, \widehat{m}(i)\} \cap \{j, \widehat{m}(j)\} = \emptyset$, the line (46) gives

$$\widehat{r}^2_{new}(i,j) \leq \langle W(\omega_i, .), W(\omega_{\widehat{m}(i)}, .)\rangle + \langle W(\omega_j, .), W(\omega_{\widehat{m}(j)}, .)\rangle - 2 \max_{v\in\{i,\widehat{m}(i)\}, w\in\{j,\widehat{m}(j)\}} \langle W(\omega_v, .), W(\omega_w, .)\rangle + t_n$$

with $t_n = 12\sqrt{\frac{\log n}{n}}$. Then, using the function U defined by (43), one has

$$\widehat{r}^2_{new}(i,j) \leq \langle W(\omega_{U(i)}, .), W(\omega_{U(i)}, .)\rangle + \langle W(\omega_{U(j)}, .), W(\omega_{U(j)}, .)\rangle - 2\,\langle W(\omega_{U(i)}, .), W(\omega_{U(j)}, .)\rangle + t_n$$

which is upper bounded by

$$r^2_W(\omega_{U(i)}, \omega_{U(j)}) + t_n$$

with Cauchy-Schwarz inequality. The line (44) is proved in the case $\{i, \widehat{m}(i)\} \cap \{j, \widehat{m}(j)\} = \emptyset$.

If $\{i, \widehat{m}(i)\} \cap \{j, \widehat{m}(j)\} \neq \emptyset$, we can see that $\widehat{r}^2_{new}(i,j) \leq 0$. Thus (44) trivially holds in this case too. The inequalities (44) are proved. ■

Proof of (45)**.** Assume the event $\mathcal{E}_{in}$ of Proposition 26 holds.

<u>If $i, j \in [n]$ such that $\{i, \widehat{m}(i)\} \cap \{j, \widehat{m}(j)\} = \emptyset$,</u>

$$\left|r^2_W(\omega_i, \omega_j) - \widehat{r}^2_{new}(i,j)\right| \leq \left|r^2_W(\omega_i, \omega_j) - \widehat{r}^2(i,j)\right| + \left|\widehat{r}^2(i,j) - \widehat{r}^2_{new}(i,j)\right|$$

by triangle inequality. The left term is upper bounded by

$$3\, r_W(\omega_j, \omega_{m(j)}) + 3\, r_W(\omega_i, \omega_{m(i)}) + 36\sqrt{\log(n)/n}$$

thanks to Theorem 7. The right term is equal to

$$2\left|\langle A_i, A_j\rangle - \max_{k\in\{i,\widehat{m}(i)\}, l\in\{j,\widehat{m}(j)\}} \langle A_k, A_l\rangle\right|$$

which is upper bounded by

$$r_W(\omega_j, \omega_{m(j)}) + r_W(\omega_i, \omega_{m(i)}) + 12\sqrt{\log(n)/n}$$

using the same technique as in the proof of Theorem 7. Combining the above displays, one has

$$\left|r^2_W(\omega_i, \omega_j) - \widehat{r}^2_{new}(i,j)\right| \leq 5\, r_W(\omega_j, \omega_{m(j)}) + 5\, r_W(\omega_i, \omega_{m(i)}) + 60\sqrt{\log(n)/n},$$

which implies

$$\widehat{r}^2_{new}(i,j) \geq r^2_W(\omega_i, \omega_j) - 5\, r_W(\omega_j, \omega_{m(j)}) - 5\, r_W(\omega_i, \omega_{m(i)}) - 60\sqrt{\log(n)/n}. \tag{47}$$

The line (45) is therefore proved in the case $\{i, \widehat{m}(i)\} \cap \{j, \widehat{m}(j)\} = \emptyset$.

<u>If $i, j \in [n]$ such that $\{i, \widehat{m}(i)\} \cap \{j, \widehat{m}(j)\} \neq \emptyset$,</u>

$$\widehat{r}_{new}(i,j) = 0.$$

Hence, it is enough to show that the right hand side of (47) is non-positive. Consider the particular case where $\widehat{m}(i) = j$ and $i \neq \widehat{m}(j)$ for example. Then, one has

$$|\widehat{r}^2(i,j)| = \left|\langle A_{\widehat{m}(j)}, A_j\rangle - \langle A_i, A_j\rangle\right| \leq |\langle A_i, A_j - A_{\widehat{m}(j)}\rangle| + |\langle A_i - A_j, A_{\widehat{m}(j)}\rangle|$$

which is upper bounded by

$$\widehat{f}(j, m(j)) + \widehat{f}(i, m(i))$$

where $\widehat{f}$ has been introduced in (9). As in the proof of Theorem 7, one can show that the above display is upper bounded by

$$r_W(\omega_j, \omega_{m(j)}) + r_W(\omega_i, \omega_{m(i)}) + 12\sqrt{\log(n)/n}$$

on the event $\mathcal{E}_{in}$. Combining this upper bound of $\widehat{r}$ with the following lower bound from Theorem 7

$$\widehat{r}^2(i,j) \geq r_W^2(\omega_i, \omega_j) - 3\, r_W(\omega_i, \omega_{m(i)}) - 3\, r_W(\omega_j, \omega_{m(j)}) - 36\sqrt{\log(n)/n}, \tag{48}$$

one derive

$$r_W^2(\omega_i, \omega_j) \leq 4\, r_W(\omega_i, \omega_{m(i)}) + 4\, r_W(\omega_j, \omega_{m(j)}) + 48\sqrt{\log(n)/n}.$$

This implies that the right hand side of (47) is non positive. Hence (45) is proved in the particular case $\widehat{m}(i) = j$ and $i \neq \widehat{m}(j)$. By symmetry, it remains only the case $\widehat{m}(i) = \widehat{m}(j)$ to do. Following the above proof, we can show taht (45) holds for this case too. The inequality (45) is therefore proved in the case $\{i, \widehat{m}(i)\} \cap \{j, \widehat{m}(j)\} \neq \emptyset$.

The line (45) is proved, and Lemma 42 follows. ■

G.2 Control on the Type I and II Errors

In Theorem 16 on the new packing number estimator, the left hand side of (20) is similar to Section 4.3 on the covering number estimator, and thus straightforward. The right hand side of (20) and Corollary 17 are proved together below.

G.2.1 Proof for the Type I Error

Assume the null-hypothesis $N_\Omega^{(p)}(\epsilon) \leq K$ holds. We want to show that the same inequality is satisfied by the statistic $\widehat{N}_\Omega^{(p.new)}(\widehat{\epsilon})$. Proof by contradiction: assume the inequality $\widehat{N}_\Omega^{(p.new)}(\widehat{\epsilon}) \geq K+1$ holds. This means that there are $K+1$ indices $i_1, \ldots, i_{K+1} \in [n]$ such that the following inequalities hold

$$\forall s, t \in \{1, \ldots, K+1\}: \quad \widehat{\epsilon}^2 < \widehat{r}_{new}^2(i_s, i_t).$$

Combining the above inequalities with the under-estimation property (44), one has

$$\forall s,t \in \{1,\dots,K+1\}: \quad \widehat{\epsilon}^2 \,<\, r_W^2(\omega_{U(i_s)},\omega_{U(i_t)}) + t_n$$

with probability at least $1-2/n$. Replacing the radius $\widehat{\epsilon}^2$ by its value ϵ^2+t_n, it comes

$$\forall s,t \in \{1,\dots,K+1\}: \quad \epsilon^2 \,<\, r_W^2(\omega_{U(i_s)},\omega_{U(i_t)}).$$

Thus, $K+1$ sampled points are separated by at least a distance ϵ, which implies $N_\Omega^{(p)}(\epsilon) \geq K+1$. This contradicts the null-hypothesis. ■

Corollary 17 and Theorem 16 are therefore proved.

G.2.2 Proof for the Type II Error (Theorem 18).

Consider a graphon (Ω,μ,W) in the set $\mathcal{W}(\eta,\beta)$. By definition of $\mathcal{W}(\eta,\beta)$, there are $K+1$ balls in (Ω,r_W) whose centers are separated by at least a distance $\sqrt{\epsilon^2+10\eta+6t_n}+\eta$. Label these balls by $s\in\{1,\dots,K+1\}$. As in the proof for the dimension estimation, assume the complementary of the event $\mathcal{E}_{bad}$, i.e. assume that each of the $K+1$ balls contains at least two sampled points. Accordingly, denote by $i_1,j_1,\dots,i_{K+1},j_{K+1}$ the indices of the corresponding sampled points such that $\omega_{i_s},\omega_{j_s}$ belong to the sth ball with $s\in\{1,\dots,K+1\}$. Since the radius of these ball is smaller than $\eta/2$, one has

$$r_W(\omega_{i_s},\omega_{m(i_s)}) \leq r_W(\omega_{i_s},\omega_{j_s}) \leq \eta \tag{49}$$

for all $s\in\{1,\dots,K+1\}$.

On the event $\mathcal{E}_{in}$ of Proposition 26, Lemma 42 gives

$$\widehat{r}_{new}^2(i_s,i_t) \geq r_W^2(\omega_{i_s},\omega_{i_t}) - 5\,r_W(\omega_{i_s},\omega_{m(i_s)}) - 5\,r_W(\omega_{j_s},\omega_{m(j_s)}) - 5t_n.$$

for all $s\neq t\in\{1,\dots,K+1\}$. Using (49), one derive

$$\widehat{r}_{new}^2(i_s,i_t) \geq r_W^2(\omega_{i_s},\omega_{i_t}) - 10\eta - 5t_n. \tag{50}$$

The ball centers are separated by at least a distance $\sqrt{\epsilon^2+10\eta+6t_n}+\eta$ by assumption, which implies that the points in these balls are separated by

$$r_W(\omega_{i_s},\omega_{i_t}) > \sqrt{\epsilon^2+10\eta+6t_n}$$

for all $s\neq t\in\{1,\dots,K+1\}$, since the ball radii are all smaller than $\eta/2$. Combining this inequality with the line (50), one obtain

$$\widehat{r}_{new}^2(i_s,i_t) > \epsilon^2+t_n$$

for all $s\neq t\in\{1,\dots,K+1\}$. Since $\widehat{\epsilon}=\sqrt{\epsilon^2+t_n}$, this gives $\widehat{N}_\Omega^{(p.new)}(\widehat{\epsilon})\geq K+1$. Thus, the alternative hypothesis is confirmed correctly.

The above displays hold on the event $\mathcal{E}_{in}\cap\mathcal{E}_{bad}^c$. Let us upper bound the probability of the complementary event. The union bound gives

$$\mathbb{P}(\mathcal{E}_{in}^c\cup\mathcal{E}_{bad}) \leq \frac{2}{n} + (K+1)2n\beta\,\exp[-\beta(n-1)]$$

thanks to Proposition 26 and Lemma 38. Theorem 18 is then proved. ■

G.2.3 Proof for the Improvement of the Type II Error (Theorem 21).

We have seen that the type II error is upper bounded by the probability of the event $\mathcal{E}_{in}^c \cap \mathcal{E}_{bad}$. Here the only difference is that $\mathcal{E}_{bad}$ refers to the new event where, for each of the M collections of $K+1+K'$ balls, at least $K'+1$ balls contain strictly less than two sampled points. For clarity, label these collections by $\{1,\ldots,M\}$, and denote by $\mathcal{C}_j$ the event where at least $K'+1$ balls of the jth collection contain strictly less than two sampled points. Then, we have

$$\mathbb{P}[\mathcal{E}_{bad}] = \mathbb{P}[\mathcal{C}_1 \cap \ldots \cap \mathcal{C}_M]$$

where $\mathbb{P}$ denote the probability distribution $\mathbb{P}_{(\Omega,\mu,W)}$ of the W-random graph. The above display is equal to

$$\mathbb{P}[\,\mathcal{C}_1] \times \mathbb{P}[\mathcal{C}_2 \big| \mathcal{C}_1] \times \ldots \times \mathbb{P}[\mathcal{C}_M \big| \mathcal{C}_1, \ldots, \mathcal{C}_{M-1}]$$

which is upper bounded by

$$\mathbb{P}[\,\mathcal{C}_1] \times \mathbb{P}[\mathcal{C}_2] \times \ldots \times \mathbb{P}[\mathcal{C}_M]$$

since the events $\mathcal{C}_1, \ldots \mathcal{C}_M$ are negatively associated (it is shown at the end of the proof). Finally, we have

$$\mathbb{P}[\mathcal{E}_{bad}] \leq \mathbb{P}[\mathcal{C}_1]^M. \tag{51}$$

Given the first collection of $K+1+K'$ balls, denote by $\mathcal{E}_j$ the event where the jth ball of the collection contains strictly less than two sampled points. By definition of the event $\mathcal{C}_1$, we have

$$\mathbb{P}[\,\mathcal{C}_1] = \mathbb{P}[\,\exists\, i_1, \ldots, i_{K'+1} \in \{1, \ldots, K+1+K'\} : \mathcal{E}_{i_1} \cap \ldots \cap \mathcal{E}_{i_{K'+1}}].$$

The union bound gives

$$\mathbb{P}[\,\mathcal{C}_1] \leq \sum_{i_1,\ldots,i_{K'+1}} \mathbb{P}[\,\mathcal{E}_{i_1} \cap \ldots \cap \mathcal{E}_{i_{K'+1}}]$$

where the sum is taken over all possible $K'+1$ different indices. The above upper bound is equal to

$$\sum_{i_1,\ldots,i_{K'+1}} \mathbb{P}[\,\mathcal{E}_{i_1}] \times \mathbb{P}[\mathcal{E}_{i_2} \big| \mathcal{E}_{i_1}] \times \ldots \times \mathbb{P}[\mathcal{E}_{i_{K'+1}} \big| \mathcal{E}_{i_1}, \ldots, \mathcal{E}_{i_{K'}}].$$

which is smaller than

$$\sum_{i_1,\ldots,i_{K'+1}} \mathbb{P}[\,\mathcal{E}_{i_1}] \times \ldots \times P[\mathcal{E}_{i_{K'+1}}] \tag{52}$$

by negative association of the events $\mathcal{E}_k$ (this fact is proved at the end). Finally, Lemma 38 ensures that

$$\mathbb{P}[\,\mathcal{E}_k] \leq 2\beta n \, \exp[-\beta(n-1)]$$

for all k, which allows to upper bound (52) and have

$$\mathbb{P}[\,\mathcal{C}_1] \leq \binom{K+K'+1}{K'+1} \Big(2\beta n \, \exp[-\beta(n-1)]\Big)^{(K'+1)}. \tag{53}$$

Thus, setting $\widetilde{p}_n = \mathbb{P}[\mathcal{C}_1]$, we deduce from (51) that

$$\mathbb{P}(\mathcal{E}_{in}^c \cup \mathcal{E}_{bad}) \leq \mathbb{P}(\mathcal{E}_{in}^c) + \mathbb{P}(\mathcal{E}_{bad}) \leq \frac{2}{n} + \widetilde{p}_n^M,$$

where $\widetilde{p}_n$ is upper bounded by (53).

It remains to show the negative association that we use in the above proof. Given the first collection of $K+1+K'$ balls, let us show that the corresponding events $\mathcal{E}_1, \ldots, \mathcal{E}_{K+1+K'}$ are negatively associated. For the n sampled points $\omega_1, \ldots, \omega_n$, define n_j the number of points in the jth ball of the collection. Theorem 13 of Dubhashi and Ranjan (1998) ensures that the variables $n_1, \ldots, n_{K+1+K'}$ are negatively associated. Define the non-increasing function $h(n_j) = \mathbb{I}_{\mathcal{E}_j}$ where $\mathbb{I}_{\mathcal{E}_j}$ is the indicator function of $\mathcal{E}_j$. The second point of Proposition 7 of Dubhashi and Ranjan (1998) shows that $h(n_1), \ldots, h(n_{K+1+K'})$ are negatively associated. This means that the events $\mathcal{E}_1, \ldots, \mathcal{E}_{K+1+K'}$ are negatively associated.

Similarly, we show the negative association of the events $\mathcal{C}_1, \ldots, \mathcal{C}_M$. Consider n_j^t the number of sampled points in the jth ball of the tth collection. These variables are negatively associated according to Theorem 13 of Dubhashi and Ranjan (1998). Define the non-increasing functions $h_t(n_1^t, \ldots, n_{K+1+K'}^t) = \mathbb{I}_{\mathcal{C}_j}$ for all $t \leq M$. Then, Proposition 7 of Dubhashi and Ranjan (1998) shows that $\mathbb{I}_{\mathcal{C}_1}, \ldots, \mathbb{I}_{\mathcal{C}_M}$ are negatively associated.

Theorem 21 is proved. ■

References

Emmanuel Abbe. Community detection and stochastic block models: recent developments. arXiv preprint arXiv:1703.10146, 2017.

David J Aldous. Representations for partially exchangeable arrays of random variables. Journal of Multivariate Analysis, 11(4):581–598, 1981.

Ery Arias-Castro, Antoine Channarond, Bruno Pelletier, and Nicolas Verzelen. On the estimation of latent distances using graph distances. arXiv preprint arXiv:1804.10611, 2018.

Avanti Athreya, Donniell E Fishkind, Minh Tang, Carey E Priebe, Youngser Park, Joshua T Vogelstein, Keith Levin, Vince Lyzinski, and Yichen Qin. Statistical inference on random dot product graphs: a survey. The Journal of Machine Learning Research, 18(1):8393–8484, 2017.

Peter J. Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. Proceedings of the National Academy of Sciences, 106(50):21068–21073, 2009.

Peter J Bickel, Aiyou Chen, Elizaveta Levina, et al. The method of moments and degree distributions for network models. The Annals of Statistics, 39(5):2280–2301, 2011.

Béla Bollobás. Random graphs. In Modern graph theory, pages 215–252. Springer, 1998.

Danail Bonchev and Gregory A Buck. Quantitative measures of network complexity. In Complexity in chemistry, biology, and ecology, pages 191–235. Springer, 2005.

Christian Borgs, Jennifer Chayes, and Adam Smith. Private graphon estimation for sparse graphs. In Advances in Neural Information Processing Systems, pages 1369–1377, 2015.

Christian Borgs, Jennifer Chayes, Devavrat Shah, and Christina Lee Yu. Iterative collaborative filtering for sparse matrix estimation. arXiv preprint arXiv:1712.00710, 2017.

Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. Random Structures & Algorithms, 49(3):503–532, 2016.

V. Chvatal. A greedy heuristic for the set-covering problem. Mathematics of Operations Research, 4(3):233–235, 1979.

Jens Christian Claussen. Offdiagonal complexity: A computationally quick complexity measure for graphs and networks. Physica A: Statistical Mechanics and its Applications, 375(1):365–373, 2007.

Gregory M Constantine. Graph complexity and the laplacian matrix in blocked experiments. Linear and Multilinear Algebra, 28(1-2):49–56, 1990.

Yohann De Castro, Claire Lacour, and Thanh Mai Pham Ngoc. Minimax adaptive estimation of nonparametric geometric graphs. arXiv preprint arXiv:1708.02107, 2017.

Matthias Dehmer and Abbe Mowshowitz. A history of graph entropy measures. Information Sciences, 181(1):57–78, 2011.

Persi Diaconis and Svante Janson. Graph limits and exchangeable random graphs. arXiv preprint arXiv:0712.2749, 2007.

DLMF. *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.23 of 2019-06-15. URL `http://dlmf.nist.gov/`. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

Devdatt Dubhashi and Desh Ranjan. Balls and bins: A study in negative dependence. Random Structures & Algorithms, 13(2):99–124, 1998.

Kenneth J Falconer. Techniques in fractal geometry, volume 3.

Chao Gao, Yu Lu, Harrison H Zhou, et al. Rate-optimal graphon estimation. The Annals of Statistics, 43(6):2624–2652, 2015.

Anna Goldenberg, Alice X Zheng, Stephen E Fienberg, Edoardo M Airoldi, et al. A survey of statistical network models. Foundations and Trends® in Machine Learning, 2(2): 129–233, 2010.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pages 855–864. ACM, 2016.

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. Journal of the american Statistical association, 97(460):1090–1098, 2002.

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. Social networks, 5(2):109–137, 1983.

Olav Kallenberg. On the representation theorem for exchangeable arrays. Journal of Multivariate Analysis, 30(1):137–154, 1989.

Balázs Kégl. Intrinsic dimension estimation using packing numbers. In Advances in neural information processing systems, pages 697–704, 2003.

Jisu Kim, Alessandro Rinaldo, and Larry Wasserman. Minimax rates for estimating the dimension of a manifold. arXiv preprint arXiv:1605.01011, 2016.

Olga Klopp, Alexandre B Tsybakov, Nicolas Verzelen, et al. Oracle inequalities for network models and sparse graphon estimation. The Annals of Statistics, 45(1):316–354, 2017.

Vladimir I Koltchinskii. Empirical geometry of multivariate data: a deconvolution approach. Annals of statistics, pages 591–629, 2000.

Pierre Latouche and Stéphane Robin. Variational bayes model averaging for graphon functions and motif frequencies inference in w-graph models. Statistics and Computing, 26 (6):1173–1185, 2016.

Antti M Latva-Koivisto. Finding a complexity measure for business process models.

Guillaume Lecué, Shahar Mendelson, et al. Regularization and the small-ball method i: sparse recovery. The Annals of Statistics, 46(2):611–641, 2018.

Elizaveta Levina and Peter J Bickel. Maximum likelihood estimation of intrinsic dimension. In Advances in neural information processing systems, pages 777–784, 2005.

Shengqiao Li. Concise formulas for the area and volume of a hyperspherical cap. Asian Journal of Mathematics and Statistics, 4(1):66–70, 2011.

Yihua Li, Devavrat Shah, Dogyoon Song, and Christina Lee Yu. Nearest neighbors for matrix estimation interpreted as blind regression for latent variable model. IEEE Transactions on Information Theory, 66(3):1760–1784, 2019.

László Lovász. Large networks and graph limits, volume 60. American Mathematical Soc., 2012.

László Lovász and Balázs Szegedy. Limits of dense graph sequences. Journal of Combinatorial Theory, Series B, 96(6):933–957, 2006.

Catherine Matias and Stéphane Robin. Modeling heterogeneity in random graphs through latent space models: a selective review. ESAIM: Proceedings and Surveys, 47:55–74, 2014.

Shahar Mendelson. Learning without concentration. In Conference on Learning Theory, pages 25–39, 2014.

Mikołaj Morzy, Tomasz Kajdanowicz, and Przemysław Kazienko. On measuring the complexity of networks: Kolmogorov complexity versus entropy. Complexity, 2017, 2017.

Mathew Penrose et al. Random geometric graphs. Number 5. Oxford university press, 2003.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 701–710. ACM, 2014.

Miklós Z Rácz, Sébastien Bubeck, et al. Basic models and questions in statistical network analysis. Statistics Surveys, 11:1–47, 2017.

Purnamrita Sarkar, Deepayan Chakrabarti, and Andrew W Moore. Theoretical justification of popular link prediction heuristics. In Twenty-Second International Joint Conference on Artificial Intelligence, 2011.

Karthik Sridharan. A gentle introduction to concentration inequalities. Dept Comput Sci, 2002.

Minh Tang, Daniel L Sussman, Carey E Priebe, et al. Universally consistent vertex classification for latent positions graphs. The Annals of Statistics, 41(3):1406–1430, 2013.

Alexandre B Tsybakov. Introduction to nonparametric estimation. Springer, New York, 2009.

Todd L Veldhuizen. Software libraries and their reuse: Entropy, kolmogorov complexity, and zipf's law. arXiv preprint cs/0508023, 2005.

Patrick J Wolfe and Sofia C Olhede. Nonparametric graphon estimation. arXiv preprint arXiv:1309.5936, 2013.

Jiaming Xu, Laurent Massoulié, and Marc Lelarge. Edge label inference in generalized stochastic block models: from spectral theory to impossibility results. In Conference on Learning Theory, pages 903–920, 2014.

Bin Yu. Assouad, fano, and le cam. In Festschrift for Lucien Le Cam, pages 423–435. Springer, 1997.

Hector Zenil, Narsis Kiani, and Jesper Tegnér. A review of graph and network complexity from an algorithmic information perspective. Entropy, 20(8):551, 2018.

Yuan Zhang, Elizaveta Levina, and Ji Zhu. Estimating network edge probabilities by neighborhood smoothing. arXiv preprint arXiv:1509.08588, 2015.

Pedro Zufiria and Iker Barriales-Valbuena. Entropy characterization of random network models. Entropy, 19(7):321, 2017.