Sharp Local Minimax Rates for Goodness-of-Fit Testing in Large Random Graphs, multivariate Poisson families and multinomials

Julien Chhor\textsuperscript{1} and Alexandra Carpentier\textsuperscript{2}

\textsuperscript{1}CREST/ENSAE
\textsuperscript{2}OvGU, Magdeburg

Abstract

We consider the identity testing problem in large inhomogeneous random graphs (also called goodness-of-fit testing problem), multivariate Poisson families and multinomials. Given a known probability distribution $P$ and $n$ iid samples drawn from an unknown probability distribution $Q$, we investigate how large $\rho$ should be to distinguish, with high probability, the case $P = Q$ from the case $d(P, Q) \geq \rho$. We answer this question in the case of a family of distances: $d(P, Q) = \|P - Q\|_t$ for $t \in [1, 2]$. Besides being locally minimax-optimal (i.e. characterizing the detection threshold associated to the known matrix $P$), our tests have simple expressions and are easily implementable. Our results are closely related to important and popular results in the multinomial setting, and complete them by providing the missing matching upper and lower bounds.

Keywords: Minimax Identity Testing, Goodness-of-fit Testing, Large Random Graphs, Multinomials, Multivariate Poisson Families, Locality

1 Introduction

In recent machine learning and statistical applications, the increasing use of networks has made large random graphs a decisive field of interest. To name a few topics, let us mention community detection, especially in the stochastic block model ([28], [17], [21], [23], [12]), in social networks ([24], [22]), as well as network modelling ([8], [44]), or network dynamics ([9]).

On the other hand, the existing literature about hypothesis testing [1] is profuse: the goodness-of-fit problem has been thoroughly studied, especially in the case of signal detection in the Gaussian setting, for instance by Ingster - see [43] - and has given rise to a vast literature.
In parallel to the study of hypothesis testing, there exists a broad literature on the related problem of property testing with seminal papers such as [5], [6].

In the present paper, we consider the problem of identity testing or goodness-of-fit testing in random graphs, multivariate Poisson families and multinomials - that is to say, testing whether the distribution of the data matches a given distribution. Throughout the paper, we will state the problems in the graph setting, and we will establish later on the link with multivariate Poisson families and multinomials. The problem is the following one: given \( n \) i.i.d. realizations of an unknown inhomogeneous Erdős-Rényi graph - see Section 2 - denoted as \( Q \), and given a known Erdős-Rényi graph \( P \), we want to test

\[
H_0 : P = Q \quad \text{vs} \quad H_1 : d(P, Q) \geq \rho,
\]

for a given distance \( d \) and separation distance \( \rho \). In other words, we want to test whether \( Q \) is equal to \( P \), or whether it is separated away from \( P \) with a distance \( \rho \), in terms of the metric \( d \).

The difficulty of this testing problem is characterized by the minimal separation radius \( \rho \) needed to ensure the existence of a test uniformly consistent under both the null and the alternative hypothesis - i.e. a test whose worst-case error is smaller than a given \( \eta > 0 \). The aim is then to devise the minimal separation distance \( \rho \) so that a uniformly consistent test exists for the worst-case error level \( \delta > 0 \). See Section 2 for a precise definition of the setting.

In this paper, we will mostly focus on the following goals:

- We focus on the case where the distance \( d \) is the \( l_t \) distance, namely \( d(P, Q) = \left( \sum_{i,j} |Q_{i,j} - P_{i,j}|^t \right)^{1/t} \) for any \( t \in [1, 2] \). Typically, the case \( t = 2 \) (Frobenius norm) and \( t = 1 \) (total variation distance) are considered, and we interpolate between these two extreme cases.

- Our main objective will be to develop tests - as well as matching lower bounds - for this identity testing problem that are locally optimal in that the minimax separation distance \( \rho \) should depend tightly on \( P \). Indeed, it is clear that some \( P \) will be more 'easy' to test than others. Consider e.g. the two following extreme cases: (i) the very 'easy' case where all entries of \( P \) are 0, i.e. all vertices of the random graphs are disconnected (which implies a very low noise) and (ii) the very 'difficult' case where all entries of \( P \) are equal to \( 1/2 \), i.e. all edges of the random graphs connect with probability \( 1/2 \) (and the noise is maximised). It is clear that the minimax local separation distance should differ between these two cases and be much smaller in case (i) than in case (ii). We aim at studying the minimax local separation distance for any \( P \), and characterise tightly its shape depending on \( P \).

Only little literature exists about identity testing in random graphs - and to the best of our knowledge, no litterature exists about local identity testing in the sense that we described above. [41] and [35] propose an analysis of the two sample case, under sparsity: Given two
populations of mutually independent random graphs, each population being drawn respectively from the distributions \( P \) and \( Q \), they perform the minimax hypothesis testing \( H_0 : P = Q \) vs \( H_1 : d(P, Q) \geq \rho \) for a variety of distances \( d \), and identify optimal tests over the classes of sparse graphs that they consider. \([19]\) identifies a computationally efficient algorithm for testing the separability of two hypotheses. Testing between a stochastic block model versus an Erdös-Rényi model has been studied in \([30]\) and \([27]\). Phase transitions are also known for detecting strongly connected groups or high dimensional geometry in large random graphs \([25]\). \([33]\) tests random dot-product graphs in the two sample setting with low-rank adjacency matrices. \([31]\) examines a more general case in which the graphs are not necessarily defined on the same set of vertices. To summarize, only few papers address the construction of efficient tests in random graphs - although this would be valuable in various areas such as social networks \([16]\), brain or ‘omics’ networks \([32]\ \[15]\), testing chemicals \([13]\) or ecology and evolution \([11]\). Moreover, and to the best of our knowledge, no paper considers the local version of the testing problem - i.e. and focuses on obtaining separation distances that depend on the null hypothesis.

On the other hand, the related identity testing problem in multinomials - i.e. probability distributions over a finite set - is a property testing problem that has been studied in the literature in many forms. The aim is to test equality of an unknown distribution represented by an unknown probability vector \( q \) - observed through \( n \) multinomial samples - to a given, known, multinomial distribution \( p \) (null hypothesis), versus the fact that \( p \) is separated from \( q \) according to a given distance. More precisely, given a sample of multinomial independent random variables generated through \( q \), the aim is to test whether \( H_0 : p = q \) vs \( H_1 : d(p, q) \geq \rho \) for some distance \( d \). As in the graph setting, the aim is then to devise the minimal separation distance \( \rho \) so that a uniformly consistent test exists the some worst-case error level \( \delta > 0 \). Note that this problem is also often considered in the dual setting of sample complexity that is to say, finding the minimal number of samples \( n \) so that a consistent test exists for a given separation distance. In other words, given a separation distance \( \rho \) and a target worst-case error probability \( \delta > 0 \), they want to characterise the minimal \( n \) that enables the existence of a uniformly consistent test with worst-case error level \( \delta > 0 \)

One distinguishes between global results which are obtained for the worst case of the distribution \( p \), and local results, where the minimax separation distance is required to depend precisely on any given \( p \). For global results, see e.g. \([2]\) (in Russian), \([3]\ \[4]\ \[12]\ \[10]\), and also in the related two-sample testing problem - where both \( p, q \) are unknown and observed through samples - see e.g. \([7]\ \[18]\). In the present paper, we focus on local results as mentioned, and this problem has been tackled in some cases in the case of multinomial distributions. Important contributions, in the case where the separation distance is the total variation distance - i.e. the distance \( \| \cdot \|_1 \) in our case - are found in e.g. \([34]\ \[26]\). Note that these papers provide results in terms of sample complexity, and more recently, the paper \([38]\) has re-considered this problem in terms of minimax separation distance - focusing mainly on the case of smooth densities. Another quite related work is \([40]\), investigating the rate of goodness-of-fit testing in the multinomial case, in the \( l^1 \) and \( l^2 \) distances, under privacy constraints. Regarding the related - yet quite different - two sample testing problem, see \([14]\ \[20]\ \[26]\ \[36]\). This multinomial framework proves very
useful for a wide range of applications, which include Ising models [39], bayesian networks [29] or even quantum mechanics [37]. The papers [34, 38] are most related to our present results concerning testing in random graphs, since the multivariate Poisson distribution setting and the multinomial setting can be seen as a special case of the random graph setting, after a Poissonization trick - see section 4 for more details on why our setting encompasses those settings. We postpone a precise discussion between our result and this stream of literature to the core of the paper[1], since it is technical. As high-level comments, we restrict to remarking here that (i) there are some regimes where the bounds in [34, 38] are not tight when it comes to the separation distance - we give examples in Section 5 and (ii) this stream of literature restricts to the case of total variation distance, namely $\|\cdot\|_1$ norm.

The paper is organized as follows: In Section 2 we describe the setting by defining the random graph model and the minimax framework. In Section 3 we describe our tests and state theoretical results guaranteeing their optimality. In Section 4 we establish the equivalence between the binomial, the Poisson and the multinomial settings. In Section 5 we discuss our results, by comparing them with the state of the art, especially with the multinomial setting.

2 PROBLEM STATEMENT

For the sake of readability, we only introduce here the random graph setting. In Section 4 we will introduce three other very related settings (the Multinomial setting corresponding to the discrete case, the Poisson setting, and the Binomial setting) and prove that the associated minimax rates can be deduced from the results in the graph setting.

2.1 Setting

Description of the data. We consider directed or undirected graphs defined on a finite set of vertices $\{1, \cdots, m\}$ for $m \geq 2$. In addition, a graph is defined by specifying its set of edges, which is a subset $E \subset \{(i, j) ; i, j \in \{1, \cdots, m\}\}$. In words, a vertex $i$ can point or not to a vector $j$ through an edge $(i, j)$. Given a graph, we can define its adjacency matrix $G = (g_{i,j})_{1 \leq i,j \leq m}$ as
g_{i,j} = \begin{cases} 1 & \text{if } (i, j) \in E \\ 0 & \text{otherwise}. \end{cases}

The set of adjacency matrices is therefore $\{0, 1\}^{m \times m}$ and we identify any directed (resp. undirected) graph with its adjacency matrix (resp. its upper triangular adjacency sub-matrix). Indeed, a graph is entirely characterized by its adjacency matrix.

\footnote{We compare with this stream of literature under our upper and lower bounds in Sections 3 and also in the discussion in Section 5}
We call matrix of connection probabilities any matrix
\[ P = (p_{i,j})_{1 \leq i,j \leq m} \in [0,1]^{m \times m}. \]

We say that \( G \) is (the adjacency matrix of) an inhomogeneous Erdös-Rényi (IER) graph with matrix of connection probabilities \( P \), and we write
\[ G \sim \text{IER}(P), \]
if \( G \) is a random matrix such that for all \( 1 \leq i,j \leq m \): \( g_{i,j} \) is distributed as a Bernoulli random variable of parameter \( p_{i,j} \) and if the \( (G_{i,j})_{1 \leq i,j \leq m} \) are mutually independent:
\[ G_{i,j} \sim \mathcal{B}(p_{i,j}), \quad (G_{i,j})_{1 \leq i,j \leq m} \text{ independent}. \]

In practice, we can choose to take into account the auto-connection probabilities or not. For instance, if we are interested in testing in social networks and care about the probability of connection between individuals, we can ignore the auto-connection - in which case all the information is contained in resp. \( (P_{i,j})_{i \neq j} \) and \( (G_{i,j})_{i \neq j} \). We can also consider undirected graphs, in which case the whole information is contained in resp. \( (P_{i,j})_{i \leq j} \) and \( (G_{i,j})_{i \leq j} \).

Let \( n \geq 1 \). The data is defined as follows: we suppose that there exists an unknown matrix of connection probabilities \( Q \in [0,1]^{m \times m} \), and that we are given \( n = 2k \) (\( k \in \mathbb{N}^* \)) independent observations \( G = (G^{(1)}, \ldots, G^{(n)}) \) of inhomogeneous Erdös-Rényi random graphs drawn from \( \text{IER}(Q) \), i.e. for any \( i \in \{1,\ldots,n\} \)
\[ G^{(i)} \sim \text{IER}(P) \quad \text{i.i.d.}. \]

We write \( P_P \) for the probability associated to \( G \).

**Additional notations.** Let \( \eta > 0 \). We denote by \( \lesssim_\eta \) and \( \gtrsim_\eta \) the inequalities up to constants depending on \( \eta \), i.e., for \( f \) and \( g \) two real-valued functions defined, we say that \( f \lesssim_\eta g \) (resp. \( f \gtrsim_\eta g \)) if there exists a constant \( c_\eta > 0 \) (resp. \( C_\eta > 0 \)) depending only on \( \eta \), such that \( c_\eta g \leq f \) (resp. \( f \geq C_\eta g \)). We write \( f \asymp g \) if \( g \lesssim f \) and \( f \gtrsim g \). Whenever the constants are absolute, we drop the index \( \eta \) and just write \( \lesssim, \gtrsim, \asymp \). We denote by \( x \vee y \) and \( x \wedge y \) the respectively maximum and minimum of the two real values \( x \) and \( y \). We denote the total variation distance between two probability measures by \( d_{TV} \) and for any matrix \( A \in \mathbb{R}^{m \times m} \) and for \( t > 0 \), we define
\[ \|A\|_t = \left( \sum_{1 \leq i,j \leq n} |A_{i,j}^t| \right)^{\frac{1}{t}}. \]
2.2 Minimax Testing Problem

We now define the testing problem considered in the paper. Let \( \eta \in (0, 1) \) be a fixed constant and let \( t \in [1, 2] \). We are given a known matrix of connection probabilities \( P \in [0, 1]^{m \times m} \) and we suppose that the data is generated from an unknown matrix \( Q \): \( G = (G^{(1)}, \ldots, G^{(n)}) \sim IER(Q) \). We are interested in the following testing problem:

\[
\mathcal{H}_0^P : Q = P \quad \text{vs} \quad \mathcal{H}_1^{\rho, P, t} : \|Q - P\|_t \geq \rho.
\]

This problem is called ’goodness-of-fit problem’. When no ambiguity arises, we write \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \) to denote the null and alternative hypotheses.

A test \( \psi \) is a measurable function of the observations \( G \), taking only the values 0 or 1. If \( \psi = 0 \) (resp. \( \psi = 1 \)), we say that the test decides in favor of \( \mathcal{H}_0 \) (resp. \( \mathcal{H}_1 \)). We would like to control the probability of deciding in favor of the wrong hypothesis by finding an optimal test for the testing problem (1). To do so, we measure the quality of any test \( \psi \) by its maximum risk, defined as:

\[
R(\psi) := R_{\rho, P, t, n}(\psi) = \mathbb{P}_P(\psi = 1) + \sup_{Q \text{ s.t. } \|P - Q\|_t \geq \rho} \mathbb{P}_Q(\psi = 0).
\]

\( R(\psi) \) is the sum of the type-I and the type-II errors:

- \( \mathbb{P}_P(\psi = 1) \) is the type-I error, i.e. the probability of deciding \( \mathcal{H}_1 \) when \( \mathcal{H}_0 \) holds.

- \( \sup_{\|P - Q\|_t \geq \rho} \mathbb{P}_Q(\psi = 0) \) is the worst-case type-II error. Indeed, for \( Q \) such that \( \|P - Q\|_t \geq \rho \), \( \mathbb{P}_Q(\psi = 0) \) corresponds to the probability of deciding in favor of \( \mathcal{H}_0 \) when \( \mathcal{H}_1 \) holds. Therefore, \( \sup_{\|P - Q\|_t \geq \rho} \mathbb{P}_Q(\psi = 0) \) is the type-II error of \( \psi \) if the data were drawn from the most unfavorable distribution \( Q \in \mathcal{H}_1 \) for \( \psi \).

The minimax risk is the risk of the best possible test, if any:

\[
R^* := R^*_{\rho, P, t, n} = \inf_{\psi \text{ test}} R(\psi) = \inf_{\psi \text{ test}} \left[ \mathbb{P}_P(\psi = 1) + \sup_{Q : \|P - Q\|_t \geq \rho} \mathbb{P}_Q(\psi = 0) \right].
\]

Note that \( R^* := R^*_{\rho, P, t, n} \) depends on the choice of the norm indexed by \( t \), the matrix \( P \), the separation radius \( \rho \), and the sample size \( n \). Since all quantities depend on \( P \), we say that the testing problem is local around \( P \) - as opposed to classical approaches in the minimax testing literature, where one generally only considers a family of matrices \( P \) and focuses only on the worst case results over this family - see e.g. [31].
In the following, we fix an absolute constant $\eta \in (0, 1)$ and we are interested in finding the smallest $\rho^*_{P,t,n}$ such that $R^*_{\rho^*_{P,t,n},P,t,n} \leq \eta$:

$$
\rho^*_{P,t,n}(\eta) = \inf \left\{ \rho > 0 : R^*_{\rho,P,t,n} \leq \eta \right\}.
$$

We call $\rho^*_{P,t,n}(\eta)$ the $\eta$-minimax separation radius. Whenever no ambiguity arises, we drop the indexation in $n, P, t, \eta$ and write simply $\rho^*, R^*_{\rho}$, but these variables remain important, as will appear later on.

The question addressed in the paper is the following: given $0 < \eta < 1$, a graph adjacency matrix $P$ and $n$ i.i.d. observations $G$ from IER$(Q)$ where $Q$ is some unknown graph adjacency matrix, what is the minimal separation radius $\rho^*_{P,t,n}$ ensuring $R^*_{\rho^*_{P,t,n},P,t,n} \leq \eta$, and can we construct minimax optimal tests? We study this question for any given graph adjacency matrix $P$, and any norm indexed by $t \in [1, 2]$.

## 3 RESULTS

Without loss of generality, we can assume that $\max_{i,j} P_{i,j} \leq \frac{1}{2}$. Otherwise, if for some $i < j$ we have $P_{i,j} > \frac{1}{2}$, we replace $P_{ij}$ by $1 - P_{ij}$ and replace accordingly $G_{i,j}^{(l)}$ by $1 - G_{i,j}^{(l)}$ for all $l = 1, \cdots, n = 2k$.

We now define the vectorization of matrix $P$, which we will use throughout the paper. It is defined as the vector $p$ containing all entries of $P$ sorted in decreasing order:

$$
p = (p_1 \geq p_2 \geq \cdots \geq p_N)
$$

where $N$ denotes the number of independent edges in the graph - if the graph is directed and with auto-connection, then $N = m^2$, if the graph is directed and without auto-connection, then $N = m(m-1)$, if the graph is undirected and without autoconnection then $N = m(m-1)/2$ etc.

Define now a bijection $\sigma : \{1, \cdots, N\} \rightarrow E$ for this transformation, where $E$ denotes the set of edges. It is defined such that, $\forall l \leq N : p_l = P_{\sigma(l)}$. We now define the vectorization of all adjacency matrices according to $\sigma$. We transform matrix $Q$ and the adjacency matrices $G^{(1)}, \cdots, G^{(n)}$ into vectors by defining

$$
\forall j = 1, \cdots, n : \begin{cases}
q = (q_1, \cdots, q_N) = (Q_{\sigma(1)}, \cdots, Q_{\sigma(N)}) \\
g^{(j)} = (g_1^{(j)}, \cdots, g_N^{(j)}) = (G_{\sigma(1)}^{(j)}, \cdots, G_{\sigma(N)}^{(j)})
\end{cases}
$$

(4)
We can notice here that we do not fully use the structure of graph, and that we could write all results for very general multivariate binomial distributions.

For an index $1 \leq U \leq N$, we define the vector $p_{\leq U} = (p_1, \cdots, p_U, 0, \cdots, 0)$ and $p_{>U} = (0, \cdots, 0, p_{U+1}, \cdots, p_N)$.

Let $\eta > 0$. In what follows, we write

$$r = \frac{2t}{4-t}.$$ 

for $p$ we also write $I$ for the smallest index such that

$$\sum_{i>I} p_i^2 \leq \frac{C_{\text{tail}}(\eta)}{n^2},$$

where $C_{\text{tail}}(\eta) = \eta^{10} \wedge \frac{1}{2}(1 - \eta)^2$. Index $A$ denotes the largest index less than $I$ such that

$$p_A \geq \frac{\sqrt{l_{\eta}}}{2^{\frac{1}{2}} \sqrt{n} \left(\sum_{i\leq I} r_p\right)^{\frac{1}{4}}} p_A^{2-t}$$

where $l_{\eta} = \log(1 + 4(1 - \eta)^2) \wedge \frac{\eta^2}{100}$, i.e.

$$A = A_{P,t,n}(\eta) := \max \left\{ a \leq I : p_a \geq \frac{\sqrt{l_{\eta}}}{2^{\frac{1}{2}} \sqrt{n} \left(\sum_{i\leq I} r_p\right)^{\frac{1}{4}}} p_A^{2-t} \right\}.$$

We will prove the following theorem.

**Theorem 1.** For all $t \in [1, 2]$, the following bound holds, up to a constant depending only on $\eta$ and $t$:

$$\rho^* \leq \eta \sqrt{\frac{\|p_{\leq I}\|_r}{n}} + \frac{\|p_A\|_{1\frac{2-t}{r}}}{n^{\frac{2-t}{r}}} + \frac{1}{n}.$$ 

### 3.1 Lower bounds

We start by presenting the lower bound part of Theorem 1. We divide the analysis into two parts: a lower bound for the large coefficients of $P$ (bulk) and a lower bound for the small coefficients of $P$ (tail). The bulk will be defined as the set $p_{\leq A}$ and the tail as $p_{>A}$. 

(Note that we do not necessarily have $q_1 \geq \cdots \geq q_N$.)
3.1.1 Lower bound for the bulk

We first provide the following lower bound for the bulk regime: In words, we identify a radius \( \rho \) such that, if the \( L^t \) distance between \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \) is less than \( \rho \), then any test has risk at least \( \eta \). Therefore, to get a decision procedure \( \psi^* \) with risk less than \( \eta \), we need the separation distance to be greater than \( \rho \), so that \( \rho^* \geq \rho \), yielding the lower bound.

**Proposition 1.** Let \( t \in [1,2] \) and consider a test \( \psi \). There exists a constant \( c' \eta > 0 \) that depends only on \( \eta \) and a distribution \( Q \) such that we have

\[
\| (q - p)_{\leq A} \|_t \geq c' \eta \left( \frac{\| P_{\leq A} \|_r^t}{\sqrt{n} \| P_{\leq I} \|_r^t} + \frac{1}{n} \right),
\]

and

\[
P_P(\psi = 1) + P_Q(\psi = 0) \geq \eta.
\]

This implies that \( \rho = \frac{\| P_{\leq A} \|_r^t}{\sqrt{n} \| P_{\leq I} \|_r^t} + \frac{1}{n} \) is a lower bound on the minimax separation radius \( \rho^* \).

Note first that the lower bound in \( 1/n \) is trivial since changing any entry of \( P \) by \( 1/n \) is not detectable with high probability. Now let us examine the first part of the rate. To prove this lower bound, we use Le Cam’s two points method by defining a prior distribution over a discrete subset of the matrices of connection probability satisfying \( \mathcal{H}_1 \). More precisely, for all \((\delta_1, \cdots, \delta_A) \in \{\pm 1\}^A\) we define the distribution \( q_\delta \) such that:

\[
(q_\delta)_i = \begin{cases} p_i + \delta_i \gamma_i & \text{if } i \leq A \\ p_i & \text{otherwise}, \end{cases}
\]

where

\[
\gamma_i = \frac{l_\eta}{2^{1/4}} \frac{p_i^{2/7}}{\sqrt{n} \left( \sum_{i \leq I} p_i^r \right)^{1/7}}.
\]

Consider the data distribution - i.e. the distribution of our \( n \) random graphs - corresponding to the case where the connection probability matrix \( Q_\delta \) has vectorization \( q_\delta \) where \( \delta \) is taken uniformly on \( \{\pm 1\}^A \) - and write it \( P_{\text{bulk}} \).

The core of the proof is to prove that observations \((G^{(1)}, \cdots, G^{(n)})\) drawn from this mixture distribution \( P_{\text{bulk}} \) are so difficult to distinguish from observations \((G'^{(1)}, \cdots, G'^{(n)})\) drawn from \( P_P \), that the risk of any decision procedure (i.e. test) is necessarily larger than \( \eta \). This brings us to the conclusion of our proposition since any distribution \( q_\delta \) is separated away from \( p \) by an \( L^t \) distance equal to \( \left( \sum_{i=1}^A \gamma_i^t \right)^{1/t} \leq \frac{\| P_{\leq A} \|_r^t}{\sqrt{n} \| P_{\leq I} \|_r^t} \). Therefore the separation radius \( \rho^* \) is necessarily at least \( \frac{\| P_{\leq A} \|_r^t}{\sqrt{n} \| P_{\leq I} \|_r^t} \).
3.1.2 Lower bound for the tail

We now derive a lower bound for the tail, i.e, the set of coefficients $p_i$ such that $i > A$. This corresponds to the case of the smallest $p_i$. We will see that the phenomena involved are quite different from those occurring the bulk of the distribution. The reason is that the definition of $A$ implies that on the tail, whp, no same coordinate (here, edge) is observed twice or more among the $n$ data.

**Proposition 2.** Let $t \in [1, 2]$, and consider any test $\psi$. There exists a constant $\zeta_\eta > 0$ depending only on $\eta$ and a distribution $Q$ such that

$$\| (q - p)_{> A} \|_t \geq \frac{\zeta_\eta \| p_{> I} \|_1^{2-t}}{n^{2-t}} ,$$

and

$$P_P(\psi = 1) + P_Q(\psi = 0) \geq \eta .$$

To prove this lower bound, we once more use Le Cam’s two points method with a sparse prior distribution. We here specify this prior. Define the smallest index $U > I$ such that $n^2 p_U \| P_{\geq U} \|_1 \leq c_u < 1$ where $c_u > 0$ is a small constant defined in the appendix. We define

$$\bar{\pi} = \frac{c_u}{n^2 \| p_{\geq U} \|_1} \text{ and } \pi_i = \frac{p_i}{\bar{\pi}} .$$

The index $U$ has no further meaning than to guarantee for all $i \geq U : \pi_i \in [0, c_u]$. In particular, $\pi_i$ is a Bernoulli parameter. Now, we define the following prior on $q$. For any $i < U$ we set $q_i = p_i$. Otherwise for $i \geq U$, we set

$$q_i = \begin{cases} \bar{\pi} \text{ with probability } \pi_i \\ 0 \text{ with probability } 1 - \pi_i , \end{cases}$$

imposing moreover that all $q_i$ are independent. Consider now the data distribution (of our $n$ random graphs) associated to a connection probability matrix re-indexed as $q$, where $q$ is distributed as described above. We denote it by $\tilde{P}_{\text{tail}}$.

Like above, we prove that data $(G^{(1)}, \ldots, G^{(n)})$ drawn from this mixture $\tilde{P}_{\text{tail}}$ is difficult to distinguish from the data $(G'^{(1)}, \ldots, G'^{(n)})$ drawn from $P_P$. Moreover, we show that with high probability, the $L^t$ distance between adjacency matrices that correspond to the mixture $\tilde{P}_{\text{tail}}$, and $P$, is larger, up to an absolute constant than

$$\frac{\| P_{> U} \|_1^{2-t}}{n^{2(t-1)}} .$$
Finally, to conclude the proof, we show in Lemma 8 that
\[
\frac{\|p > U\|}{n^{2(t-1)/t}} + \frac{1}{n} \ll \eta \frac{\|p > I\|}{n^{2(t-1)/t}} + \frac{1}{n}
\]
in words, that we can replace \(U\) by \(I\).

### 3.1.3 Combination of both lower bounds

By combining Propositions 1 and 2, we obtain the following theorem.

**Theorem 2.** Let \(t \in [1, 2]\), and consider any test \(\psi\). There exists a constant \(c_\eta' > 0\) depending only on \(\eta\) and a distribution \(Q\) such that
\[
\|Q - P\| \geq c_\eta' \left(\sqrt{\frac{\|p \leq I\|}{n}} + \frac{\|p > A\|}{n^{2t-2}} + \frac{1}{n}\right),
\]
and
\[
P_P(\psi = 1) + P_Q(\psi = 0) \geq \eta.
\]

This theorem implies that
\[
\rho^* \gtrsim \eta \sqrt{\frac{\|P \leq A\|}{n}} + \frac{\|p > A\|}{n^{2t-2}} + \frac{1}{n},
\]
which is a lower bound on the separation radius \(\rho^*\), up to a positive constant depending only on \(\eta\).

Note that when combining Propositions 1 and 2 we do not get exactly the expression in Theorem 2. We actually obtain:
\[
\rho^* \gtrsim \eta \sqrt{\frac{\|P \leq A\|}{n}} + \frac{\|p > I\|_{r/4}}{n^{2t-2}} + \frac{1}{n}.
\]

We therefore need to show that this expression is equivalent to that in Theorem 2. This is done by using first Lemma 9 which states that we can replace \(\frac{\|P \leq A\|}{\sqrt{n}\|P \leq I\|_{r/4}}\) by \(\frac{\|p \leq I\|}{n}\) without changing the rate, i.e.
\[
\frac{\|P \leq A\|_{r/4}}{\sqrt{n}\|P \leq I\|_{r}} + \frac{\|p > A\|_{1/4}}{n^{2t-2}} + \frac{1}{n} \ll \sqrt{\frac{\|p \leq I\|}{n}} + \frac{\|p > I\|}{n^{2t-2}} + \frac{1}{n}
\]
We then use Lemma 10 to show that we can replace $\|p_{>I}\|_1$ by $\|p_{>A}\|_1$ without changing the rate, which yields the expression in Theorem 2.

**Remark on the index $A$:** As explained in (5), the optimal prior is of the form $p_i \pm \gamma_i$ where $\gamma_i$ is proportional to $p_i^{2-t}$, according to (6). Since $\frac{2}{4-t} \leq 1$, we can have $\gamma_i > p_i$ if $p_i$ is too small, so that it is impossible to set the optimal prior $p_i \pm \gamma_i$, since $p_i - \gamma_i$ has to be a Bernoulli parameter. The index $A$ is just the last index ensuring $p_A \geq \gamma_A$ so that we can do the lower bound construction.

**Remark on the index $I$:** The index $I$ defines the largest set of coefficients $p_{>I}$ such that, whp, when observing $n$ graphs, no edge $j > I$ is observed twice or more. This is exactly the interpretation of the relation $\sum_{j>I} n^2 p_j^2 \leq c_{\text{tail}}$ for a small constant $c_{\text{tail}}$. As shown in Lemma 14, the miracle is that the definition of $A$ also implies that $\sum_{j>A} n^2 p_j^2 \leq c_{\text{tail}} + l_n^2$, which leads us to tune the constants $c_{\text{tail}}$ and $l_n$ such that this sum is small. Therefore, on the actual tail ($p_{>A}$), no same coordinate will be observed twice whp. This is the reason why the phenomena involved are different on the bulk and on the tail. On the bulk, many coordinates are observed at least twice, which allows us to build an estimator based on the dispersion of the data around its mean, namely the renormalized $\chi^2$ estimator which is a modified estimator of the variance. Like in the classical signal detection setting, the optimal procedure for detecting whether or not the data is drawn from $P$ is to estimate the dispersion of the data.

On the tail, however, each coordinate is observed at most once, so we cannot access the dispersion of the data. On this set, we rather design a prior distribution which mimics the behavior of the null, while being as separated away from it as possible. More precisely, we impose that whp, no coordinate is observed twice, and such that coordinate-wise, the expected number of observations is equal to that under the null hypothesis $P$. In short, this prior is designed such that its first order moment is equal to that under the null and its second order moment is unobserved whp. Under both of these constraints, we maximise the $L^t$ distance between the null hypothesis $P$ and the possible distributions composing the prior.

### 3.2 Upper bounds

We write first let

$$S = \sum_{j=1}^{k} G(j), \quad \text{and} \quad S' = \sum_{j=k+1}^{n} G'(j),$$

where $S$ and $S'$ are vectors containing the number of times each edge occurs in the data, after sample splitting. We also write

$$b = \frac{4 - 2t}{4 - t}.$$
3.2.1 Test for the bulk coefficients

We now introduce the following test statistic on the bulk coefficients, i.e. the coefficients with index smaller than $A$:

$$T_{\text{bulk}} = \sum_{i \leq A} \frac{1}{p_i} \left( \frac{S_i}{k} - p_i \right) \left( \frac{S_i'}{k} - p_i \right),$$

(8)

which is a weighted $\chi^2$ statistic. We now define the test

$$\psi_{\text{bulk}} = 1\{T_{\text{bulk}} > \frac{\zeta_\eta}{\|p\|_{r,n}} \},$$

where $\zeta_\eta = 4/\sqrt{\eta}$.

We prove the following proposition regarding this statistic and the bulk of the vector $p$.

**Proposition 3.** There exists $\xi_\eta > 0$, such that the following holds.

- **Type I error is bounded:**
  $$\mathbb{P}_P(\psi_{\text{bulk}} = 1) \leq \eta/2.$$

- **Type II error is bounded:** for any $Q$ such that
  $$\|q\|_{t} \geq \xi_\eta \left( \sqrt{\frac{\|p\|_{r}^2}{n}} + \frac{1}{n} \right),$$

  it holds that
  $$\mathbb{P}_Q(\psi_{\text{bulk}} = 0) \leq \eta/2.$$

For $t = 1$, we get $r = \frac{2}{3}$, which is the norm identified in [34] by Valiant and Valiant. However, the setting is slightly different since we consider probability distributions on graphs rather than on multinomials, since we consider separation distance for a fixed $n$ instead of sample complexity - and since we also consider all $t \in [1, 2]$. However, we prove later on that the two settings - graphs and multinomials - are related and that the rates can be transferred from our setting to the multinomial case.

Note however that we define the bulk in a different way than in [34]. Indeed, in [34], the index $I'$ defining the tail is the smallest index such that, for a fixed $\epsilon$, the total mass of small coefficients is less than $\epsilon$: $\sum_{i > I'} p_i \leq \epsilon$. This definition therefore only involves the first order moment of the null distribution. In our setting, conversely, we define the index $I$ using the second order moment of the null distribution, as the smallest index such that $\sum_{i > I} p_i^2 \leq \frac{\zeta_{\text{tail}}}{n}$. The above result also generalizes the bound identified in [34], by characterizing the rate of testing for all $t \in [1, 2]$ and sheds light on a duality between the norms $\| \cdot \|_t$ and $\| \cdot \|_r$ when $r = \frac{2t}{4-t}$. 

13
3.2.2 Test for the tail coefficients

The tail test will be a combination of two tests. We define the histogram of the data which is a sufficient statistic:

\[ \forall j > A, N_j := \sum_{i=1}^{n} \mathbb{1}\{X_i = j\} \]

We first define the test \( \psi_2 \) which rejects \( H_0 \) whenever one tail coordinate is observed twice.

\[ \psi_2 = \mathbb{1}\{\exists j > A: N_j \geq 2\} \]  

(9)

We also define a statistic counting the number of observations on the tail, and the associated test, recalling that \( \xi_n = 4 / \sqrt{n} \):

\[ T_1 = \sum_{i > A} \frac{N_i}{n} - p_i, \quad \psi_1 = \mathbb{1}\{|T_1| > \xi_n \sqrt{\frac{\sum_{i > A} p_i}{n}}\} \]

We prove the following proposition regarding this statistic.

**Proposition 4.** There exists \( \xi'_n > 0 \), such that the following holds.

- **Type I error is bounded:**
  \[ \mathbb{P}_P(\psi_1 \lor \psi_2 = 1) \leq \eta / 2. \]

- **Type II error is bounded:** for any \( Q \) such that
  \[ \| q_{> A} \|_t \geq \xi'_n \left( \frac{\| p_{> A} \|_1 2^{-t}}{n^{2t-2}} + \frac{1}{n} \right), \]
  it holds that
  \[ \mathbb{P}_Q(\psi_1 \lor \psi_2 = 0) \leq \eta / 2. \]

Recall that the tail of the distribution is defined such that, \( \text{whp} \), no same coordinate is observed at least twice. On this set, the best that we can do is to count how many coordinates are observed - each being only observed once - and to compare to what should be observed under the null. We therefore combine two tests: one which rejects \( H_0 \) if one of the coordinates is observed at least twice- test \( \psi_2 \) - and an other one which rejects \( H_0 \) if the total mass of observed coordinates is too different from its expectation under the null - test \( \psi_1 \). Proposition 4 proves that this combination of tests reaches the optimal rate.

In [34], the tail test only involves the first order moment, which is sufficient in the case of the \( \| \cdot \|_1 \) norm. Moreover, in the proof of Proposition 4, it becomes clear that for \( t = 1 \) we only need the test \( \psi_1 \) and for \( t = 2 \) we only need the test \( \psi_2 \). However in the case of the \( \| \cdot \|_t \) for \( t \in (1, 2) \), the combination of both \( \psi_1 \) and \( \psi_2 \) is necessary.
3.2.3 Aggregated test

We now combine the above results to define the aggregated test. We define our test as

\[ \psi = \psi_{\text{bulk}} \lor \psi_1 \lor \psi_2. \]

This is the test rejecting the null whenever one of the three tests rejects. Denote by

\[ \bar{\rho} = \sqrt{\frac{\|p_{\leq I}\|_r}{n} + \frac{\|p_{>A}\|_1^{2-t}}{n^{\frac{2-t}{2}}} + \frac{1}{n}}. \]

The following theorem states that this test reaches the rate \( \bar{\rho} \), which is the minimax rate \( \rho^* \) given in Theorem 1. In words, it guarantees that, whenever the two hypotheses are separated away with an \( L^t \) distance of the order of \( \bar{\rho} \), this test has type-I and type-II errors upper bounded by \( \eta/2 \), ensuring that its risk is less than \( \eta \). Since the minimax separation radius \( \rho^* \) is the smallest radius ensuring the existence of a test satisfying this condition, we can conclude that \( \rho^* \lesssim \bar{\rho} \). Hence the upper bound.

**Theorem 3.** There exists \( c' \eta > 0 \), such that the following holds.

- Type I error is bounded:
  \[ P_{\psi}(\psi = 1) \leq \eta/2. \]

- Type II error is bounded: for any \( Q \) such that
  \[ \|p - q\|_t \geq c' \eta \left( \sqrt{\frac{\|p_{\leq I}\|_r}{n} + \frac{\|p_{>A}\|_1^{2-t}}{n^{\frac{2-t}{2}}} + \frac{1}{n}} \right), \]
  it holds that
  \[ P_{\Psi}(\psi = 0) \leq \eta/2. \]

4 Equivalence between the Binomial, the multinomial and the Poisson setting

We now move to the multinomial and Poisson settings and show that we can transfer the above results, using a Poissonization trick. In what follows, we consider the parameter vector as being an element of \( \mathbb{R}^d \) rather than \( \mathbb{R}^N \), unlike before.

**Prop 1** (Poissonization trick for multinomials). Let \( n \geq 2 \) and \( \bar{n} \sim Poi(n) \). Conditional on \( \bar{n} \), let \( Z_1, \ldots, Z_{\bar{n}} \overset{\text{iid}}{\sim} M(q) \). For all \( j = 1, \ldots, d \), let \( N_j = \sum_{i=1}^{\bar{n}} 1 \{ Z_i = j \} \). Then for all \( j \), \( N_j \sim Poi(nq_j) \) and \( N_1, \ldots, N_d \) are mutually independent.
Prop 2 (Poissonization trick for binomial families). Let \( n \geq 2 \) and \( \tilde{n} \sim \text{Poi}(n) \). Conditional on \( \tilde{n} \), let \( X_1, \cdots, X_n \overset{iid}{\sim} \bigotimes_{j=1}^d \text{Ber}(p_j) \). Then \( \sum_{i=1}^{\tilde{n}} X_i \sim \bigotimes_{j=1}^d \text{Poi}(np_j) \).

In all the sequel we suppose \textit{wlog} \( p_1 \geq \cdots \geq p_d \). We consider the following settings:

1. **Binomial case:** This is the set considered in the graph example, after vectorisation of the adjacency matrix. We define \( \mathcal{P}^{\text{(Bin)}} = \{ \text{Ber}(p); \ p \in \mathbb{R}_+^d \} \) where by convention, \( \text{Ber}(p) := \bigotimes_{j=1}^d \text{Ber}(p_j) \). We fix \( p \in \mathcal{P}^{\text{(Bin)}} \) and suppose we are given observe \( X_1, \cdots, X_n \overset{iid}{\sim} \text{Ber}(q) \) for \( q \in \mathcal{P}^{\text{(Bin)}} \) unknown. We consider the \textbf{binomial} testing problem:

\[
H_0^{\text{(Bin)}}(q = p) \quad \text{vs} \quad H_1^{\text{(Bin)}}(\|q - p\|_t \geq \rho).
\]

2. **Poisson case:** \( \mathcal{P}^{\text{(Poi)}} = \{ \text{Poi}(p); \ p \in \mathbb{R}_+^d \} \) where by convention, \( \text{Poi}(p) := \bigotimes_{j=1}^d \text{Poi}(p_j) \). We fix \( p \in \mathcal{P}^{\text{(Poi)}} \) and suppose we are given observe \( Y_1, \cdots, Y_n \overset{iid}{\sim} \text{Poi}(q) \) for \( q \in \mathcal{P}^{\text{(Poi)}} \) unknown. We consider the \textbf{Poisson} testing problem:

\[
H_0^{\text{(Poi)}}(q = p) \quad \text{vs} \quad H_1^{\text{(Poi)}}(\|q - p\|_t \geq \rho).
\]

3. **Multinomial case** \( \mathcal{P}^{\text{Mult}} = \{ \mathcal{M}(p); \ p \in \mathbb{R}_+^d \text{ and } \sum_{j=1}^d p_j = 1 \} \) where \( \mathcal{M}(p) \) denotes the multinomial distribution over \( \{1, \ldots, d\} \). We fix \( p \in \mathcal{P}^{\text{Mult}} \) and suppose we are given observe \( Z_1, \cdots, Z_n \overset{iid}{\sim} \text{Mult}(q) \) for \( q \in \mathcal{P}^{\text{Mult}} \) unknown. We consider the \textbf{Poisson} testing problem:

\[
H_0^{\text{Mult}}(q = p) \quad \text{vs} \quad H_1^{\text{Mult}}(\|q - p\|_{\mathcal{M}, t} \geq \rho).
\]

where for \( x = (x_1, \cdots, x_d) \): \( \|x\|_{\mathcal{M}, t} = \left[ \sum_{j=2}^d |x_j|^t \right]^{1/t} \) is the multinomial norm, defined without taking the first coordinate into account. Indeed, because of the shape constraint \( \sum p_j = 1 \), the first coordinate does not bring any information and can be deduced from the \( d - 1 \) coordinates.

For these three testing problems, we define respectively \( \rho^{\text{Bin}}_*(n, p, \eta) \), \( \rho^{\text{Poi}}_*(n, p, \eta) \), \( \rho^{\text{Mult}}_*(n, p, \eta) \) for the minimax separation distances in the sense of Equation \( (3) \), for each one of the testing problems.

We state the following statement regarding the equivalence between all models.
Lemma 1. (Equivalence between Binomial and Poisson settings) There exist two absolute constants $c_{BP}, C_{BP} > 0$ such that $\forall p \in [0,1]^d$, $\forall n \geq 2 \eta > 0$, :

$$c_{BP} \rho^*_{\text{Bin}}(n,p,\eta) \leq \rho^*_{\text{Poi}}(n,p) \leq C_{BP} \rho^*_{\text{Bin}}(n,p,\text{eta}).$$

Lemma 2. (Equivalence between Multinomial and Poisson settings) It holds that $\forall p \in [0,1]^d$, $\forall n \geq 2 \eta > 0$, if $\sum_{i=1}^{d} p_i = 1$:

$$\rho^*_{\text{Mult}}(n,p,\eta) \preceq \eta \rho^*_{\text{Poi}}(n,p^{-\max}) \preceq \eta \rho^*_{\text{Mult}}(n,p,\eta)$$

where $p^{-\max} := (p_2, \cdots, p_d)$.

This entails the following corollary regarding the minimax rates of testing in the multinomial model:

**Corollary 1.** The minimax separation radii in the Poisson and multinomial cases are respectively given by:

$$\rho^*_{\text{Poi}}(n,p,\eta) \asymp \eta \sqrt{\frac{\|p_{\leq I}\|_r}{n} + \frac{\|p_{>A}\|_{1_{1-r}}^2}{n^{2-2}} + \frac{1}{n}} \quad \text{for } p \in \mathcal{P}(\text{Poi})$$

$$\rho^*_{\text{Mult}}(n,p,\eta) \asymp \eta \sqrt{\frac{\|p_{\leq I}^{-\max}\|_r}{n} + \frac{\|p_{>A}\|_{1_{1-r}}^2}{n^{2-2}} + \frac{1}{n}} \quad \text{for } p \in \mathcal{P}(\text{Mult})$$

where $p_{\leq I}^{-\max} = (p_2, \ldots, p_I)$.

Note that the upper bounds in the Poisson model are obtained using our tests on the Poisson vector, and the upper bounds in the Multinomial model are obtained using our tests on the $d-1$ last coordinates of the estimates of probabilities of each categories.

## 5 DISCUSSION

In this entire section, we mostly discuss the Multinomial setting and results - whose rates are given, as a corollary of our graph results, in Corollary [1] - since it is the setting that is mostly studied in the literature. Since, in what follows, we do not focus on the influence of the target error probability $\eta$, we will write to alleviate notations $\rho^*(n,p)$ for the minimax separation distance in the Multinomial model, dropping the dependency on $\eta$. In this section and as elsewhere in the paper, $\rho^*$ (without indexing by $n,p,\eta$ to alleviate notations) refers to the minimax rate in the graph setting.
5.1 Locality of the results

In the present paper, we derive sharp local minimax rates of testing in the binomial, Poisson and multinomial settings. The locality property is a major aspect of the results: For each fixed $p$ we identify the detection threshold associated to $p$, where $p$ is allowed to be any distribution in the class. This approach is less standard than the usual global approach, which consists in finding the largest detection threshold in the class, that is, the worst case of $p$. Yet, local results can substantially improve global results: For instance, in the multinomial case, the global separation radius for an $N$-dimensional multinomial in terms of the $\| \cdot \|_2$ norm is classically $N^{1/4}/\sqrt{n}$, and is reached in the case where $p$ is uniform distribution. However, if $p = (1, 0, \ldots, 0)$ is a Dirac multinomial, then from our results the rate of testing is $1/n$, hence much faster than the global rate. Even for fixed $N$, one can actually find a sequence of null distributions $p^{(n)}$ whose associated separation distance $\rho^{*}(p^{(n)}, n)$ reaches any rate $1/n^\alpha$ for any $1/2 \leq \alpha \leq 1$. This consequently improves the global rate even for less extreme discrete distributions than Dirac multinomials. To give an example, consider an exponentially decreasing multinomial distribution $p^{(n)} = (\frac{c n^{(2\alpha-1)/2}}{\sum_{j=1}^{N} j^{2\alpha-1}})^{N}$ for the renormalizing constant $c = n^{2\alpha-1} \frac{1 - \frac{1}{n^{\alpha-1}}}{1 - \frac{1}{n^{(2\alpha-1)N}}} \approx n^{2\alpha-1}$. Then, evaluating the local rate in $\| \cdot \|_2$ (allowing us to consider the whole set of coefficients as the bulk, see Section 5.4 below), we get:

$$\rho^{*}(n, p^{(n)}) \approx \sqrt{\frac{\|p - \max\|_2}{n}} + \frac{1}{n} \approx \frac{1}{n^\alpha}.$$

5.2 Links with previous work

Our results are quite related to those of [34], which examines the multinomial testing problem for the $L^1$ distance and in terms of sample complexity. More precisely, for a fixed $N$-dimensional multinomial distribution $p$, and for a fixed separation $\rho$, this work investigates the smallest number $n^{*}(p, \rho)$ of samples $X_1, \ldots, X_n \sim M(p)$ needed to ensure that the Multinomial testing problem introduced in Section 4 has a minimax risk - i.e. sum of worst-case type I and type II testing error - less than $2/3$:

$$H_0 : q = p \quad \text{vs} \quad H_1 : q; \|q - p\|_1 \geq \rho$$

However, the local upper- and lower bounds identified in [34], though matching in most usual cases - also when translated in terms of separation distance - actually do not match in some specific cases. [33] proves the following bounds to characterize the optimal sample complexity $n^{*}(p, \rho)$ when given a fixed $\rho > 0$:

$$\frac{1}{\rho} + \frac{\|p - \max\|_2/3}{\rho} \lesssim n^{*}(p, \rho) \lesssim \frac{1}{\rho} + \frac{\|p - \max\|_2/3}{\rho}$$

18
In the above bound, \( p = (p_1, \cdots, p_N) \) where \( p_1 \geq \cdots \geq p_N \geq 0 \) and \( \sum_{i=1}^{n} p_i = 1 \). For \( \rho > 0 \), let \( J \) be the smallest index such that \( \sum_{i>J} p_i \leq \rho \). The notation: \( \|p_{-\rho}^{\text{max}}\|_1 \) denotes \( \sum_{i=2}^{J} p_i \).

We show here that these bounds do not match in general. Let \( 1 > \rho > 0 \), let \( N > 0 \) an integer and

\[
p = (1 - \rho, \frac{\rho}{N-1}, \cdots, \frac{\rho}{N-1})
\]

so that \( \|p\|_1 = 1 \).

Then, \( \|p_{-\rho}^{\text{max}}\|_2/3 = 0 \) and \( \|p_{-\rho/16}^{\text{max}}\|_2/3 \simeq \left[ \frac{15(N-1)}{16} \left( \frac{\rho}{16(N-1)} \right)^{\frac{2}{3}} \right]^{\frac{3}{2}} \simeq \rho \sqrt{N} \) so that:

\[
\frac{1}{\rho} + \frac{\|p_{-\rho}^{\text{max}}\|_2/3}{\rho} = \frac{1}{\rho}
\]

and

\[
\frac{1}{\rho} + \frac{\|p_{-\rho/16}^{\text{max}}\|_2/3}{\rho} = \frac{1 + \sqrt{N}}{\rho} \gg \frac{1}{\rho}
\]

for \( N \gg 1 \), proving that the bounds do not match.

In the present work, we fix this by providing the matching upper- and lower bounds in terms of separation distance, which were unknown to the best of our knowledge.

We also generalize the result in several respects:

- We consider the whole range of \( L^t \) distances for \( t \) in the segment \([1, 2]\) and characterize the \textit{local} rates of testing in each case,
- We generalize the multinomial case to the graph case (binomial case) and to the Poisson setting, through the Poissonization trick.

All these cases involve comparable phenomena. The distribution can be split into bulk (set of large coefficients, with a subgaussian phenomenon) and tail (set of small coefficients, with a subpoissonian phenomenon). To the best of our knowledge, the way we define the tail is new. It allows us to establish a clear cut-off between these two optimal sets, fundamentally differing through the behavior of the second order moment of \( p \).

5.3 Remarks on the tests

In the bulk tests, we propose test statistics based on sample splitting, whose variance is easier to express. However, those tests could be define slightly differently without sample splitting,
allowing also for the analysis of the case $n = 1$. Denoting by $N = \sum_{i=1}^{n} G^{(i)}$ the component-wise sum of the data, define

$$\tilde{T}_{\text{Bulk}} = \sum_{j \leq A} \frac{1}{p_j} \left[ \left( \frac{N_j}{n} - p_j \right)^2 - N_j \right]$$

and the associated test:

$$\tilde{\psi}_{\text{bulk}} = 1\{\tilde{T}_{\text{bulk}} > \frac{c}{n} \|p_{\leq A}\|_2^2\}.$$

Then this test is equivalent to the bulk test we define (8), and is therefore also optimal in the bulk regime.

To understand the interpolation between the extreme cases $t = 1$ and $t = 2$, one need to understand what the tail tests $\psi_1$ and $\psi_2$ capture. Under $H_1$, the test $\psi_1$ checks that the total mass of the tail coefficients $\|q_{> A}\|_1$ is not to far away from $\|p_{> A}\|_1$. As to test $\psi_2$, on the tail, that is, on a set for which $\sum_{j > A} n^2 p_j^2 \ll 1$, it is actually equivalent to using a test for the second order moment. In other words, test $\psi_2$ is equivalent to $\tilde{\psi}_2 = 1\{|T_2| > \frac{c}{n} \|p_{> A}\|_2\}$ for a small constant $c$, where

$$T_2 = \sum_{i > A} \left( \frac{S_i}{k} - p_i \right) \left( \frac{S_i'}{k} - p_i \right).$$

Therefore, the test $\tilde{\psi}_2$ checks that the second order moment of the tail of distribution $q_{> A}$ is not too different from that of $p_{> A}$, in other words, that it does not contain much greater coefficients than the corresponding values of $p_{> A}$.

### 5.4 Influence of the norm $\| \cdot \|_t$

In this paper, we consider the separation distance in all $\| \cdot \|_t$ norms for $t \in [1, 2]$. The choice of $t$ influences the minimax separation distance.

In the extreme case $t = 2$, the minimax separation distance reduces to: $\rho^* \asymp \sqrt{\frac{\|p_{< I}\|_2}{n}} + \frac{1}{n}$, which can be further simplified as:

$$\rho^* \asymp \sqrt{\frac{\|p\|_2}{n}} + \frac{1}{n}.$$  

Indeed, by definition of $I$: $\|p_{> I}\|_2 \lesssim \frac{1}{n}$. In this case, as discussed earlier, a single test would suffice for reaching this separation distance. We could simply apply the bulk test - in the case of $t = 2$, a simple $\chi^2$ test - on both the bulk and tail coefficients.

Define $\Delta = p - q$. In the extreme case $t = 1$, all perturbations $|\Delta_j|$ contribute as much as each other to the discrepancy $\|\Delta\|_1$. Therefore, in terms of the $\| \cdot \|_1$ norm, it makes no difference whether all perturbations $|\Delta_j|$ are equally small, or if all the total mass of the discrepancy
is concentrated in one unique perturbation $|\Delta_j|$. Only the total mass of the perturbations $\|\Delta\|_1$ is important. We therefore do not need to use test $\psi_2$, which is tailored to detect extreme values of the perturbations.

Between the two extreme cases, that is, for $t \in (1, 2)$ all $|\Delta_j|$ contribute differently the total discrepancy $\|\Delta\|_t$. The larger perturbations increase the discrepancy much more than the smaller ones, and this phenomenon becomes more pronounced as $t$ increases. We therefore need to combine both the first order moment test $\psi_1$ and the second order moment test $\psi_2$.

For $t > 2$, the underlying phenomenon is fundamentally different. The $\| \cdot \|_t$ norm for $t > 2$ emphasises so much the large deviations that the second order moment seems to be no more optimal for testing. We conjecture that, like in the gaussian signal detection setting in $\| \cdot \|_t$ norm, the optimal procedure would rather be to check for the extreme values in the upper bound and set a sparse prior in the lower bound, at least on the bulk (defined possibly differently from what we define). Most likely, the analysis of this case can not be deduced from our results.

An interesting phenomenon is the duality between the norms $\| \cdot \|_t$ and $\| \cdot \|_r$ when $r = \frac{4t}{4-t}$. This phenomenon comes from a combination of Hölder’s inequality and information theory. Define $\gamma = (\gamma_1, \ldots, \gamma_A) \in [0, 1]^A$, and define the random vector $q = (p_1 + \delta_1 \gamma_1, \ldots, p_A + \delta_A \gamma_A)$ for $\delta_i \sim \text{Rad}(\frac{1}{2})$ like in [4], except that this time, we do not impose that $(\gamma_i)_i$ is defined as in [6]. Introduce, for $L^2_{\eta} = \log(1 + 4(1 - \eta)^2)$:

$$\Gamma := \left\{ (\gamma_1, \ldots, \gamma_A) \in [0, 1]^A : \sum_{i=1}^A \frac{\gamma_i^4}{p_i^2} \leq \frac{L^2_{\eta}}{n^2}; \ p_i - \gamma_i \in [0, 1], \ p_i + \gamma_i \in [0, 1] \right\}.$$ 

Then by Lemma 4 found in the Appendix, whenever $\gamma \in \Gamma$, the $n$ graphs generated from the random vector $q$ have a probability distribution indistinguishable from the null hypothesis $p$. The largest $\gamma \in \Gamma$, when measured in $\| \cdot \|_t$, therefore provides a lower bound on the minimax separation radius. It is found by solving: $\max_{\gamma \in \Gamma} \sum_{i=1}^A \gamma_i^t$, which can be done using Hölder’s inequality:

$$\sum_{i=1}^A \gamma_i^t = \sum_{i=1}^A \left( \frac{\gamma_i^4}{p_i^2} \right)^{t/4} p_i^{t/2} \leq \text{Hölder} \left( \sum_{i=1}^A \frac{\gamma_i^4}{p_i^2} \right)^{t/4} \left( \sum_{i=1}^A p_i^2 \right)^{(4-t)/4} \leq \left( \frac{L^2_{\eta}}{n^2} \right)^{t/4} \|p\|_{r}^{1/2t}$$

where we have used Hölder’s inequality with $a = \frac{t}{4}$ and $b = \frac{4-t}{4-t}$. Setting $\gamma^*$ the vector on the frontier of $\Gamma$ reaching the equality case in Hölder’s inequality, we obtain for fixed $n$: $\|\gamma^*\|_t \propto \|p\|_{r}^{1/2t}$. 

21
5.5 Asymptotics

In the case of a fixed multinomial distribution $p$ different from a Dirac, it is interesting to notice that for $n$ large enough, we have $I = A = N$. In words, we eventually no longer need to split the distribution between bulk and tail and we can define the bulk as the whole set of coefficients. The asymptotic local minimax rate therefore rewrites:

$$\rho^*(p, n) \nonumber \quad \begin{aligned}
\sqrt{\frac{\|p-\max\|}{n}} + \frac{1}{n} & \leq \frac{N^{1/4}}{\sqrt{n}} \\
\sqrt{\frac{\|p\|}{n}} + \frac{1}{n} & \leq \frac{N^{1/4}}{\sqrt{n}}
\end{aligned}$$

in the multinomial case

$$\sqrt{\frac{\|p\|}{n}} + \frac{1}{n} \leq \frac{N^{1/4}}{\sqrt{n}}$$

in the binomial or Poisson case.

The fast rate $\frac{1}{n}$ asymptotically dominates if, and only if, $p$ is a Dirac multinomial distribution or $p$ is the null distribution in the binomial and Poisson setting.

Acknowledgements. Both authors acknowledge fruitful discussions with Alexandre Tsybakov, Cristina Butucea and Rajarshi Mukherjee. The work of A. Carpentier is partially supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether grant MuSyAD (CA 1488/1-1), by the DFG - 314838170, GRK 2297 MathCoRe, by the FG DFG , by the DFG CRC 1294 'Data Assimilation', Project A03, and by the UFA-DFH through the French-German Doktorandenkolleg CDFA 01-18 and by the SFI Sachsen-Anhalt for the project RE-BCI.

References

[1] Jerzy Neyman and Egon S Pearson. IX. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. Lond. A*, 231(694-706):289–337, 1933.

[2] Yu I Ingster. Asymptotically minimax testing of nonparametric hypotheses on the density of the distribution of an independent sample. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.(LOMI)*, 136:74, 1984.

[3] Yuri Izmailovich Ingster. The minimax test of nonparametric hypothesis on a distribution density in metrics l_p. *Teoriya Veroyatnostei i ee Primeneniya*, 31(2):384–389, 1986.

[4] Michael Sergeevich Ermakov. Minimax nonparametric testing of hypotheses on the distribution density. *Theory of Probability & Its Applications*, 39(3):396–416, 1995.

[5] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[6] Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM (JACM)*, 45(4):653–750, 1998.
[7] Tugkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D Smith, and Patrick White. Testing that distributions are close. In Proceedings 41st Annual Symposium on Foundations of Computer Science, pages 259–269. IEEE, 2000.

[8] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. Reviews of modern physics, 74(1):47, 2002.

[9] Noam Berger, Christian Borgs, Jennifer T Chayes, and Amin Saberi. On the spread of viruses on the internet. In Soda, volume 5, pages 301–310, 2005.

[10] Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. IEEE Transactions on Information Theory, 54(10):4750–4755, 2008.

[11] Darren P Croft, Joah R Madden, Daniel W Franks, and Richard James. Hypothesis testing in animal social networks. Trends in ecology & evolution, 26(10):502–507, 2011.

[12] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. Physical Review E, 84(6):066106, 2011.

[13] Nino Shervashidze, Pascal Schweitzer, Erik Jan Van Leeuwen, Kurt Mehlhorn, and Karsten M Borgwardt. Weisfeiler-lehman graph kernels. Journal of Machine Learning Research, 12(9), 2011.

[14] Jayadev Acharya, Hirakendu Das, Ashkan Jafarpour, Alon Orlitsky, Shengjun Pan, and Ananda Suresh. Competitive classification and closeness testing. In Conference on Learning Theory, pages 22–1, 2012.

[15] Daniel R Hyduke, Nathan E Lewis, and Bernhard Ø Palsson. Analysis of omics data with genome-scale models of metabolism. Molecular BioSystems, 9(2):167–174, 2013.

[16] Sebastian Moreno and Jennifer Neville. Network hypothesis testing using mixed kronecker product graph models. In 2013 IEEE 13th International Conference on Data Mining, pages 1163–1168. IEEE, 2013.

[17] Ery Arias-Castro, Nicolas Verzelen, et al. Community detection in dense random networks. The Annals of Statistics, 42(3):940–969, 2014.

[18] Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms, pages 1193–1203. SIAM, 2014.

[19] Jing Qian and Venkatesh Saligrama. Efficient minimax signal detection on graphs. Advances in Neural Information Processing Systems, 27:2708–2716, 2014.
[20] Bhaswar Bhattacharyya and Gregory Valiant. Testing closeness with unequal sized samples. In Advances in Neural Information Processing Systems, pages 2611–2619, 2015.

[21] Nicolas Verzelen, Ery Arias-Castro, et al. Community detection in sparse random networks. The Annals of Applied Probability, 25(6):3465–3510, 2015.

[22] Meng Wang, Chaokun Wang, Jeffrey Xu Yu, and Jun Zhang. Community detection in social networks: an in-depth benchmarking study with a procedure-oriented framework. Proceedings of the VLDB Endowment, 8(10):998–1009, 2015.

[23] Emmanuel Abbe and Colin Sandon. Achieving the ks threshold in the general stochastic block model with linearized acyclic belief propagation. In Proceedings of the 30th International Conference on Neural Information Processing Systems, pages 1342–1350. Citeseer, 2016.

[24] Punam Bedi and Chhavi Sharma. Community detection in social networks. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 6(3):115–135, 2016.

[25] 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.

[26] Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS), pages 685–694. IEEE, 2016.

[27] Jing Lei et al. A goodness-of-fit test for stochastic block models. The Annals of Statistics, 44(1):401–424, 2016.

[28] Emmanuel Abbe. Community detection and stochastic block models: recent developments. The Journal of Machine Learning Research, 18(1):6446–6531, 2017.

[29] Clément L Canonne, Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Testing bayesian networks. In Conference on Learning Theory, pages 370–448. PMLR, 2017.

[30] Chao Gao and John Lafferty. Testing network structure using relations between small subgraph probabilities. arXiv preprint arXiv:1704.06742, 2017.

[31] Debarghya Ghoshdastidar, Maurilio Gutzeit, Alexandra Carpentier, and Ulrike von Luxburg. Two-sample tests for large random graphs using network statistics. arXiv preprint arXiv:1705.06168, 2017.

[32] Cedric E Ginestet, Jun Li, Prakash Balachandran, Steven Rosenberg, Eric D Kolaczyk, et al. Hypothesis testing for network data in functional neuroimaging. The Annals of Applied Statistics, 11(2):725–750, 2017.
[33] Minh Tang, Avanti Athreya, Daniel L Sussman, Vince Lyzinski, Youngser Park, and Carey E Priebe. A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, 26(2):344–354, 2017.

[34] Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017.

[35] Debarghya Ghoshdastidar and Ulrike von Luxburg. Practical methods for graph two-sample testing. *Advances in Neural Information Processing Systems*, 31:3019–3028, 2018.

[36] Iman Kim, Sivaraman Balakrishnan, and Larry Wasserman. Robust multivariate nonparametric tests via projection-pursuit. arXiv preprint arXiv:1803.00715, 2018.

[37] Costin Bădescu, Ryan O’Donnell, and John Wright. Quantum state certification. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 503–514, 2019.

[38] Sivaraman Balakrishnan, Larry Wasserman, et al. Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *Annals of Statistics*, 47(4):1893–1927, 2019.

[39] Constantinos Daskalakis, Nishanth Dikkala, and Gautam Kamath. Testing ising models. *IEEE Transactions on Information Theory*, 65(11):6829–6852, 2019.

[40] Thomas B Berrett and Cristina Butucea. Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. arXiv preprint arXiv:2005.12601, 2020.

[41] Debarghya Ghoshdastidar, Maurilio Gutzeit, Alexandra Carpentier, Ulrike Von Luxburg, et al. Two-sample hypothesis testing for inhomogeneous random graphs. *Annals of Statistics*, 48(4):2208–2229, 2020.

[42] Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*, volume 40. Cambridge University Press, 2016.

[43] Yuri Ingster and Irina A Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Springer Science & Business Media, 2012.

[44] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[45] Alexandre B Tsybakov. *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008.
A Lower bound

Let $P$ be a connection probability matrix. For $\mathcal{P}_1 := \mathcal{P}_1(\rho)$ a particular collection of connection probability matrices satisfying $\mathcal{H}_{1,\rho}$ we denote by $U(\mathcal{P}_1)$ the uniform distribution on $\mathcal{P}_1$.

Let $G$ be the set of $n$-sequences of adjacency matrices on all vertices $\{1, \ldots, m\}$ - e.g. the set of $n$-sequences of matrices in $\{0, 1\}^{m \times m}$. Let us also write $N$ for the number of independent edges in our graph - which is $m^2$ for directed graphs with auto-connection, $m(m - 1)$ for directed graphs without auto-connection, $m(m - 1)/2$ for undirected graphs without auto-connection.

Then the following lemma gives a way to derive a lower bound on $\rho^*$ by giving a sufficient condition, for a fixed $\rho$, that $R^*(\rho) \geq \eta$:

**Lemma 3.** If

\[
\frac{1}{|G|} \sum_{G \in G} \left( \frac{\mathbb{E}_{Q \sim U(\mathcal{P}_1)} \mathbb{P}_Q(G)}{\mathbb{P}_P(G)} \right)^2 \leq 1 + 4(1 - \eta)^2,
\]

Then $R^*(\rho) \geq \eta$.

**Proof of Lemma** We have that:

\[
R^*(\epsilon_t) \geq \inf_{\psi} \mathbb{P}_P(\psi = 1) + \sup_{Q \in \mathcal{P}_1} \mathbb{P}_Q(\psi = 0) \quad \text{(all elements of $\mathcal{P}_1$ satisfy $\mathcal{H}_1$)}
\]

\[
\geq \inf_{\psi} \mathbb{P}_P(\psi = 1) + \mathbb{E}_{Q \sim U(\mathcal{P}_1)} \mathbb{P}_Q(\psi = 0) \quad \text{(the supremum is greater than the integral)}
\]

\[
= 1 + \inf_{\psi} \mathbb{P}_P(\psi = 1) - \mathbb{E}_{Q \sim U(\mathcal{P}_1)} \mathbb{P}_Q(\psi = 1)
\]

\[
= 1 - \sup_{\psi} \left| \mathbb{P}_P(\psi = 1) - \mathbb{E}_{Q \sim U(\mathcal{P}_1)} \mathbb{P}_Q(\psi = 1) \right|
\]

\[
= 1 - d_{TV}(\mathbb{P}_P, \mathbb{E}_{Q \sim U(\mathcal{P}_1)} \mathbb{P}_Q)
\]

\[
\geq 1 - \frac{1}{2} \sqrt{\chi^2(\mathbb{P}_P \mid \mathbb{E}_{Q \sim U(\mathcal{P}_1)} \mathbb{P}_Q)},
\]

where the definition of the $\chi^2$ divergence can be found in [45], as well as the proof for the inequality $d_{TV} \leq \frac{1}{2} \sqrt{\chi^2}$. And so:

\[
R^*(\rho) \geq 1 - \frac{1}{2} \sqrt{\chi^2(\mathbb{P}_P \mid \mathbb{E}_{Q \sim U(\mathcal{P}_1)} \mathbb{P}_Q)}
\]

\[
= 1 - \frac{1}{2} \sqrt{\frac{1}{|G|} \sum_{G \in G} \left( \frac{\mathbb{E}_{Q \sim U(\mathcal{P}_1)} \mathbb{P}_Q(G)}{\mathbb{P}_P(G)} \right)^2} - 1
\]

Therefore, to have $R^*(\epsilon_t) \geq \eta$ it suffices that

\[
\frac{1}{|G|} \sum_{G \in G} \left( \frac{\mathbb{E}_{Q \sim U(\mathcal{P}_1)} \mathbb{P}_Q(G)}{\mathbb{P}_P(G)} \right)^2 \leq 1 + 4(1 - \eta)^2
\]

26
From now on, we identify any connection probability matrix \( Q \) with its vectorization \( q \) as defined in (4) - where the reordering is taken with respect to \( P \). For all \( i = 1, \cdots, N \), let \( \gamma_i \in [0, p_i] \) and let \( \gamma = (\gamma_i)_i \). We now apply the previous lemma with \( \mathcal{P}_1 = \{p + (\delta_i\gamma_i)_{i \leq N} \mid \delta \in \{\pm 1\}^{|G|}\} \).

We recall that \( n \) is the total number of observations. Let \( L_\eta = \sqrt{\log 1 + 4(1 - \eta)^2} \). This yields the following lemma:

**Lemma 4.** We suppose that

\[
\sum_{i=1}^{|G|} \frac{\gamma_i^4}{p_i^3} \leq \frac{L_\eta^2}{2n^2}.
\]

Then, for all \( \rho \leq \|\gamma\|_1 : R^*(\rho) \geq \eta \).

**Proof.** We will use Lemma [3](#) with \( p \) and \( \mathcal{P}_1 \) defined as above.

- We first compute \( \mathbb{P}_Q(G) \) for some realization \( G = (G^{(1)}, \ldots, G^{(n)}) \). Let \( S = \sum_{i=1}^n G^{(i)} \) and let \( s \) be the vectorisation of \( S \) as in (4). We have that

\[
\mathbb{P}_P(G) = \prod_{i=1}^{|G|} p_i^{s_i}(1 - p_i)^{n-s_i}
\]

- We now compute \( \mathbb{E}_{Q \sim \mathcal{U}(\mathcal{P}_1)}\mathbb{P}_Q(G) \): for any \( (\delta_i)_i \in \{\pm 1\}^N \), we define \( q_\delta = p + (\delta_i\gamma_i)_{1 \leq i \leq N} \), and \( Q_\delta \) the matrix matrix vectorized as \( q_\delta \). Then we have:

\[
\mathbb{P}_{Q_\delta}(G) = \prod_{i=1}^N (p_i + \delta_i\gamma_i)^{s_i}(1 - p_i - \delta_i\gamma_i)^{n-s_i}
\]

Therefore we have:
Lemma 3.

Note that

\[
\frac{1}{|G|} \sum_{G \in \mathcal{G}} \left( \mathbb{E}_{Q \sim \mathcal{U}(G)} P_Q(G) \right)^2 = \frac{1}{|G|} \sum_{G \in \mathcal{G}} \sum_{\delta, \delta'} \frac{N \prod_{i=1}^N (p_i + \delta_i \gamma_i)^{s_i} (1 - p_i - \delta_i \gamma_i)^{n-s_i}}{p_i^{s_i} (1 - p_i)^{n-s_i}}
\]

\[
\times (p_i + \delta_i \gamma_i)^{s_i} (1 - p_i - \delta_i \gamma_i)^{n-s_i}
\]

\[
= \frac{1}{|G|} \sum_{\delta, \delta'} \prod_{i=1}^N \left( 1 + \frac{\delta_i \delta'_i \gamma_i^2}{p_i (1 - p_i)} \right)^n = \prod_{i=1}^N \left[ \frac{1}{4} \sum_{\delta_i, \delta'_i \in \{\pm 1\}} \left( 1 + \frac{\delta_i \delta'_i \gamma_i^2}{p_i (1 - p_i)} \right)^n \right]
\]

\[
\leq \prod_{i=1}^N \left[ \frac{1}{2} \exp \left( \frac{n \gamma_i^2}{p_i (1 - p_i)} \right) + \frac{1}{2} \exp \left( \frac{-n \gamma_i^2}{p_i (1 - p_i)} \right) \right]
\]

\[
= \prod_{i=1}^N \cosh \left( \frac{n \gamma_i^2}{p_i (1 - p_i)} \right) \leq \exp \left( \sum_{i=1}^N \frac{n^2 \gamma_i^4}{2 p_i^2 (1 - p_i)^2} \right)
\]

Note that

\[
\exp \left( \sum_{i=1}^N \frac{n^2 \gamma_i^4}{2 p_i^2 (1 - p_i)^2} \right) \leq 1 + 4(1 - \eta)^2
\]

\[
\iff \sum_{i=1}^N \frac{\gamma_i^4}{p_i^2 (1 - p_i)^2} \leq \frac{2 L_\eta^2}{n^2}
\]

\[
\iff \sum_{i=1}^N \frac{\gamma_i^4}{p_i^2 (1 - p_i)^2} \leq \frac{L_\eta^2}{2n^2}
\]

where \( L_\eta^2 := \log \left( 1 + 4(1 - \eta)^2 \right) \) and since \( \forall 1 \leq i \leq |G|, \ p_i \leq \frac{1}{2} \). The result follows by Lemma 3.

\[\square\]
This means the following: let $\gamma := (\gamma_i)_i$ satisfying (11) and let $\rho = \|\gamma\|_t$. Then all points $p + (\delta_i \gamma_i)_{1 \leq i \leq |G|}$ are located at a distance $\rho$ from $p$ in terms of $L_t$ norm - so that the corresponding adjacency matrices are at a distance $\rho$ from each other in $L_t$ norm. Moreover we proved that for the uniform prior on this set of points $\mathcal{P}_1$, we have $R^*(\rho) \geq \eta$, which yields $\rho^* \geq \rho$.

We now prove the lower bound by combining the following four lemmas.

**Lemma 5.** It holds that

$$\rho_1^* \gtrsim \eta \rho_1 := \frac{\|p_{\leq A}\|_t^r}{\sqrt{n} \|p_{\leq I}\|_t^r}.$$

**Lemma 6.** Assume that $\|p_{> I}\|_1 \geq \frac{1}{n}$. Then it holds that

$$\rho_2^* \gtrsim \eta \rho_2 := \frac{\|p_{> I}\|_1^{\frac{2-t}{t}}}{n^{\frac{t-2}{t}}}.$$  

**Lemma 7.** Assume that $\|p_{\geq I}\|_1 \leq \frac{1}{n}$. Then it holds that

$$\rho_3^* \gtrsim \rho_3 := \frac{1}{n}.$$  

**Lemma 8.** It holds : $\|P_{\geq U}\|_1 + \frac{1}{n} \asymp \|P_{> I}\|_1 + \frac{1}{n}$. Moreover, we have either $\|P_{\geq U}\|_1 \geq \frac{1}{3} \|P_{> I}\|_1$ or $\|P_{> I}\|_1 \leq \|P_{\geq U}\|_1 + \frac{\sqrt{C_{\text{tail}}}}{n}$.

**Lemma 9.** $\rho_1 + \rho_2 \asymp \sqrt{\frac{\|p_{\leq I}\|_r}{n}} + \rho_2$

**Lemma 10.** $\|P_{> I}\|_1 + \frac{1}{n} \asymp \|P_{> A}\|_1 + \frac{1}{n}$

**Proof of Lemma 5.** We recall that $l_\eta = L_\eta \wedge \frac{n^2}{100}$. We then define the constant

$$a = \frac{\sqrt{l_\eta}}{2^{\frac{1}{4}} \sqrt{n} \left(\sum_{i \leq I} p_i^r\right)^{\frac{1}{4}}}.$$  

(10)

For all $\delta \in \{\pm 1\}^N$ let $q_\delta = ((q_\delta)_i)_{i=1,\ldots,N}$ such that
• \( \forall i \leq A, \ (q_\delta)_i = p_i + a \delta_i p_i^{\frac{2}{\rho}} \) where \( a = \) is defined in (10).

• \( \forall i > A, \ (q_\delta)_i = p_i. \)

Let \( \mathcal{P}_1 = \{ q_\delta | \delta \in \{ \pm 1 \}^A \} \) - we recall that we identify connection probability matrices with their. We set a uniform prior on \( \mathcal{P}_1 \). With the notation of Lemma 4 we just set \( \gamma_i = a q_i^{\frac{2}{\rho}} \) if \( i \leq A \) and 0 otherwise. In terms of \( \| \cdot \|_t \), any probability matrix where this prior puts mass is separated from \( P \) with a distance \( \rho \) such that:

\[
\rho = a \left\| \left( p_i^{\frac{2}{\rho}} \right)_{i=1,\ldots,A} \right\|_t = \frac{\sqrt{\| \gamma \|_t}}{\sqrt{\mathbb{E} (\sum_i p_i^t)}} \left( \sum_i p_i^t \right)^{\frac{1}{2}} \gtrsim \eta \frac{\| p_{\leq A} \|_t}{\sqrt{\mathbb{E} (\| p_{\leq I} \|_t^2)}} = \rho_1.
\]

According to Lemma 4 this prior gives a minimax risk greater than \( \eta \) since

\[
\sum_{i \leq A} \gamma_i^A = a^4 \sum_{i \leq A} p_i^{8\rho - 2} = \frac{L^2}{2n^2} \leq \frac{L^2}{2n^2}.
\]

Proof of Lemma 6. By assumption we have \( \| p_{> I} \|_1 \gtrsim \frac{1}{n} \).

Let \( U \) be the smallest index greater than or equal to \( A \) such that \( n^2 p_U \| p_{\geq U} \|_1 \leq c_u \) where \( c_u = \frac{\eta}{10} \wedge \frac{1}{2} (1 - \eta)^2 \).

Let

\[
\bar{\pi} = \frac{c_u}{n^2 \| P_{\geq U} \|_1} \quad \text{and} \quad \pi_i = \frac{p_i}{\bar{\pi}}.
\]

We set the following sparse prior on the matrices of connection probability: for all \( i < U \) we set \( q_i = p_i \) and for all \( i \geq U \) we draw \( b_i \sim \mathcal{B}(\pi_i) \) mutually independent, and we define \( q_i = b_i \bar{\pi}. \)

We write \( q = (q_i)_i \) and \( Q \) for the corresponding random connection probability matrix - for which we write \( Q \) for the distribution.

Before showing that the data distribution coming from this prior - namely \( \mathbb{E}_{Q \sim Q} \mathbb{P}_Q - \) is close enough to \( \mathbb{P}_\pi \) in total variation, we first prove that \( Q \sim Q \) is such that \( \| Q - P \|_t \) is with high probability larger - up to a positive multiplicative constant that depends only on \( u \) - than \( \rho_2 \) from the null. We have

\[
\mathbb{E}_{Q \sim Q} \left[ \| P - Q \|_t \right] = \mathbb{E}_{(b_i)_i \sim \mathcal{B}(\pi_i)} \left[ \sum_{i \geq U} |b_i \bar{\pi} - p_i|^t \right] = \pi^t \mathbb{E}_{(b_i)_i \sim \mathcal{B}(\pi_i)} \left[ \sum_{i \geq U} |b_i - \pi_i|^t \right] = \pi^t \sum_{i \geq U} \pi_i (1 - \pi_i)^t + (1 - \pi_i) \pi_i^t \geq 4^{-1} \pi^t \sum_{i \geq U} \pi_i + \pi_i^t \geq 4^{-1} \pi^t \sum_{i \geq U} \pi_i,
\]

30
since \( \forall i \geq U, \pi_i \leq c_u \leq \frac{1}{2} \), and

\[
\mathbb{V}_{Q \sim Q} \left[ \| P - Q \|_f^2 \right] = \pi^2 \sum_{i \geq U} b_i \sim \mathcal{B}(\pi_i) \| b_i - \pi_i \|^2 = \pi^2 \sum_{i \geq U} \pi_i (1 - \pi_i) \left[ (1 - \pi_i)^t - \pi_i^t \right]^2 \leq \pi^2 \sum_{i \geq U} \pi_i.
\]

We now show that \( \left( \mathbb{E}_{Q \sim Q} \left[ \| P - Q \|_f^2 \right] \right)^2 \gg \mathbb{V}_{Q \sim Q} \left[ \| P - Q \|_f^2 \right] \). This is equivalent to proving \( \sum_{i \geq U} \pi_i \gg 1 \), or equivalently: \( n^2 \| P_{\geq U} \|^2 \gg c_u \).

By Lemma \( \mathbb{8} \) we are necessarily in the case \( \| P_{\geq U} \|_1 \geq \frac{1}{3} \| P_{> I} \|_1 \). Indeed, suppose that \( \| P_{\geq U} \|_1 < \frac{1}{3} \| P_{> I} \|_1 \), then by Lemma \( \mathbb{8} \) we would have

\[
\| P_{> I} \|_1 \leq \| P_{\geq U} \|_1 + \frac{\sqrt{C_{\text{tail}}}}{n} \leq \frac{1}{3} \| P_{> I} \|_1 + \frac{\sqrt{C_{\text{tail}}}}{n},
\]

hence \( \| P_{> I} \|_1 \leq \frac{3 \sqrt{C_{\text{tail}}}}{n} \), which is excluded because we assume \( \| P_{> I} \|_1 \geq \frac{1}{n} \).

Therefore, \( \| P_{\geq U} \|_2 n^2 \geq \frac{1}{9} \gg c_u \). We conclude using Chebyshev’s inequality. Therefore, this prior is indeed separated away from the null distribution by a distance greater than \( \pi \sum_{i \geq U} \pi_i \) up to a constant, or equivalently, greater than \( \frac{P_{\geq U}}{n^2} \).

We now show that this prior is such that \( R^* \gg \eta \). We write \( \bar{P}_{\text{tail}} = \mathbb{E}_{Q \sim Q} [P_Q] \), the prior distribution used to lower bound the minimax risk. We always have:

\[
R^* \geq 1 - d_{TV} \left( P, \bar{P}_{\text{tail}} \right).
\]

Moreover, we remind that for any realization \( G = (G^{(1)}, \ldots, G^{(n)}) \) we write \( S = \sum_{i=1}^n G^{(i)} \) and \( s \) for the vectorization of \( S \) as in \( \mathbb{1} \). We have

\[
d_{TV} \left( P, \bar{P}_{\text{tail}} \right) = \frac{1}{2} \sum_{G \in \mathcal{G}} \left| P(G) - \bar{P}_{\text{tail}}(G) \right| = \frac{1}{2} \sum_{G \in \mathcal{G} : \forall i \geq U, s_i \leq 2} \left| P(G) - \bar{P}_{\text{tail}}(G) \right| + \frac{1}{2} \sum_{G \in \mathcal{G} : \exists i \geq U, s_i > 2} \left| P(G) - \bar{P}_{\text{tail}}(G) \right|.
\]

This implies that - since \( s \) is a sufficient statistic

\[
\sum_{G \in \mathcal{G} : \exists i \geq U, s_i > 2} \left| P(G) - \bar{P}_{\text{tail}}(G) \right| \leq \left| P \left( \exists i \geq U ; s_i \geq 2 \right) + \bar{P}_{\text{tail}} \left( \exists i \geq U ; s_i \geq 2 \right) \right|
\]

\[
\leq \sum_{i=U} \left[ 1 - P(s_i = 0) - P(s_i = 1) + 1 - \bar{P}_{\text{tail}}(s_i = 0) - \bar{P}_{\text{tail}}(s_i = 1) \right].
\]
Let’s fix \( i \in \{ U, \ldots, N \} \). We will use the following inequalities true for all \( n \in \mathbb{N}, x \in [0, 1] \):

\[
(1 - x)^n \geq 1 - nx; \quad (1 - x)^n \geq 1 - nx + \frac{n^2}{4}x^2; \quad (1 - x)^n \leq 1 - nx + \frac{n^2}{2}x^2.
\]

**First term in the sum:** \( \sum_{i=U}^{N} [1 - \mathbb{P}(s_i = 0) - \mathbb{P}(s_i = 1)] \). We recall that by the definition of \( U \) we have \( \forall i \geq U \; np_i \leq c_u \) so that for any \( i \geq U \)

\[
1 - \mathbb{P}(s_i = 0) - \mathbb{P}(s_i = 1) = 1 - (1 - p_i)^n - np_i(1 - p_i)^{n-1}
\]

\[
\leq 1 - \left[ 1 - np_i + \frac{n}{4}p_i^2 \right] - np_i \left[ 1 - (n - 1)p_i \right] \leq n^2p_i^2.
\]

Summing over all \( i = U, \ldots, N \) yields that

\[
\sum_{i=U}^{N} [1 - \mathbb{P}(s_i = 0) - \mathbb{P}(s_i = 1)] \leq C_{tail}.
\]

**Second term in the sum:** \( \sum_{i=U}^{N} [1 - \tilde{\mathbb{P}}_{tail}(s_i = 0) - \tilde{\mathbb{P}}_{tail}(s_i = 1)] \). We recall that by the definition of \( U \) we have \( \forall i \geq U \; np_i \leq c_u \) so that for any \( i \geq U \)

\[
1 - \tilde{\mathbb{P}}_{tail}(s_i = 0) - \tilde{\mathbb{P}}_{tail}(s_i = 1) = 1 - \left[ 1 - \pi_i + \pi_i(1 - \tilde{\pi})^n \right] - \pi_i n\tilde{\pi}(1 - p)^{n-1}
\]

Summing over all \( i = U, \ldots, N \) yields that

\[
\sum_{i=U}^{N} [1 - \tilde{\mathbb{P}}_{tail}(s_i = 0) - \tilde{\mathbb{P}}_{tail}(s_i = 1)] \leq c_u\frac{\|P_{\geq U}\|_1}{\|P_{\geq U}\|_1} = c_u.
\]

Therefore

\[
d_{TV} \left( \mathbb{P}_{\cdot}, \mathbb{P}_{\text{tail}} \right) = \frac{1}{2} \sum_{G \in \mathcal{G}; \forall i \geq U, s_i \leq 2} \left| \mathbb{P}_{\cdot}(G) - \tilde{\mathbb{P}}_{\text{tail}}(G) \right| + C_{\text{tail}} + c_u \tag{11}
\]

Now, we can upper bound the total variation by the \( \chi^2 \) divergence on the high probability event that we only observe 0 or 1 for each coordinate \( i \geq U \). We have - since \( s \) is a sufficient
We can now sum the two terms:

\[
\sum_{k=0}^{1} \frac{\Pr_{\text{tail}}(s_i = k)^2}{\Pr(s_i = k)} = 1 + 2\bar{c}_u = \frac{N}{\prod_{i=U} \left( \sum_{k=0}^{1} \frac{\Pr_{\text{tail}}(s_i = k)^2}{\Pr(s_i = k)} \right)} - 1 + 2\bar{c}_u. \tag{12}
\]

**Computation of \( \sum_{k=0}^{1} \frac{\Pr_{\text{tail}}(s_i = k)^2}{\Pr(s_i = k)} \).** Now, we compute \( \sum_{k=0}^{1} \frac{\Pr_{\text{tail}}(s_i = k)^2}{\Pr(s_i = k)} \):

\[
\frac{\Pr_{\text{tail}}(s_i = k)^2}{\Pr(s_i = k)} = \frac{\left[ 1 - \pi_i + \pi_i (1 - \bar{\pi})^n \right]^2}{(1 - p_i)^n} + \frac{\pi_i n \bar{\pi} (1 - \bar{\pi}^{-n})^2}{np_i (1 - p_i)^n-1}
\]

The first term writes:

\[
\frac{\left[ 1 - \pi_i + \pi_i (1 - \bar{\pi})^n \right]^2}{(1 - p_i)^n} \leq \frac{\left[ 1 - \pi_i + \pi_i (1 - \bar{\pi} + \frac{n^2}{2} \bar{\pi}^2) \right]^2}{1 - np_i} = 1 - np_i + n^2 p_i \bar{\pi} + \frac{n^2 p_i \bar{\pi}}{1 - np_i}
\]

\[
\leq 1 - np_i + n^2 p_i \bar{\pi} + \frac{n^4 p_i^2 \bar{\pi}^2}{4(1 - C_{\text{tail}})} \leq 1 - np_i + n^2 p_i \bar{\pi} + \frac{n^2 p_i \bar{\pi}}{4(1 - C_{\text{tail}})}.
\]

The second term writes:

\[
\frac{\pi_i n \bar{\pi} (1 - \bar{\pi}^{-n})^2}{np_i (1 - p_i)^n-1} = np_i (1 - \bar{\pi})^{2n} \leq np_i \text{ since } \bar{\pi} \geq p_i
\]

We can now sum the two terms:

\[
\sum_{k=0}^{1} \frac{\Pr_{\text{tail}}(s_i = k)^2}{\Pr(s_i = k)} = 1 + n^2 p_i \bar{\pi} + \frac{n^2 p_i \bar{\pi}}{4(1 - C_{\text{tail}})}
\]

So that

\[
\prod_{i=U} \left( \sum_{k=0}^{1} \frac{\Pr_{\text{tail}}(s_i = k)^2}{\Pr(s_i = k)} \right) = \prod_{k=U} \left( 1 + n^2 p_i \bar{\pi} + \frac{n^2 p_i \bar{\pi}}{4(1 - C_{\text{tail}})} \right) \leq \exp \left( \sum_{k=U}^{N} n^2 p_i \bar{\pi} + \frac{n^2 p_i \bar{\pi}}{4(1 - C_{\text{tail}})} \right)
\]

\[
= \exp \left( cu + \frac{cu}{1 - C_{\text{tail}}} \right) \leq \exp \frac{3}{2} cu \leq 1 + 3cu \text{ since } \frac{3}{2} cu \leq 1.
\]
Now, using (11) and (12), we have: \( d_{TV}(P_P, P_{\text{tail}}) \leq \frac{1}{2} \sqrt{5c_u} + C_{\text{tail}} + c_u \leq 1 - \eta \) by the definition of \( c_u, C_{\text{tail}} \). This concludes the proof. 

\[ \]

**Proof of Lemma 7** We introduce \( Q \) such that \( q_1 = p_1 + \frac{1-n}{n} \) and \( q_j = p_j \) for all \( j \geq 2 \).

\[
R^* \geq \inf_{\psi_{\text{test}}} P_P(\psi = 1) + P_Q(\psi = 0) = 1 - d_{TV}(P_P, P_Q)
\]
\[
= 1 - n \ d_{TV} \left( \bigotimes_{i<j} B(p_i), \bigotimes_{i<j} B(q_i) \right)
\]
\[
= 1 - n \ d_{TV} \left( B(p_1), B(q_1) \right) = 1 - n \ |p_1 - q_1| = 1 - n \ \frac{1-\eta}{n}
\]
\[
= \eta.
\]
This concludes the proof.

**Proof of lemma 8** If \( \|P_{\geq U}\|_1 \geq \frac{1}{7} \|P_{\geq I}\|_1 \) then the result is clear. Now, suppose \( \|P_{\geq U}\|_1 < \frac{1}{3} \|P_{\geq I}\|_1 \). We have \( \|P_{\geq U}\|_1 < \frac{1}{2} \|P_{I \rightarrow U}\| \) where \( P_{I \rightarrow U} = (p_{I+1}, \cdots, p_{U-1}) \). We have:

\[
p_{U-1}^2 + \frac{C_{\text{tail}}}{2n^2} \geq p_{U-1}^2 + \frac{1}{2} \sum_{i=I+1}^{U-1} p_i^2 \geq p_{U-1} \left( p_{U-1} + \frac{1}{2} \sum_{i=I+1}^{U-1} p_i \right) > p_{U-1} \left( p_{U-1} + \sum_{i \geq U} p_i \right)
\]
\[
\geq p_{U-1} \sum_{i \geq U-1} p_i = p_{U-1} \|P_{\geq U-1}\|_1 > \frac{c_u}{n^2}
\]
by the definition of \( U \).

Therefore,

\[
p_{U-1}^2 > \frac{2c_u - C_{\text{tail}}}{2n^2} \implies \forall I < i < U, \ p_i^2 > \frac{C_{\text{tail}}}{2n^2} \) since \( c_u \geq C_{\text{tail}} \).

Moreover,

\[
\frac{C_{\text{tail}}}{n^2} \geq \sum_{U < i < U} p_i^2 > (I - U - 1)p_{U-1}^2 > (I - U - 1) \frac{C_{\text{tail}}}{2n^2}
\]

34
So that
\[ I - U - 1 < 2 \quad \text{i.e.} \quad I - U - 1 \leq 1 \]
Thus:
\[
\|P_{> I}\|_1 \leq \|P_{1\rightarrow U}\|_1 + \|P_{\geq U}\|_1 \leq (I - U - 1)p_{I+1} + \|P_{\geq U}\|_1 \\
\leq \frac{\sqrt{C_{\text{tail}}}}{n} + \|P_{\geq U}\|_1 \lesssim \|P_{\geq U}\|_1 + \frac{1}{n}.
\]
Hence the result. □

Proof of Lemma 9 Clearly, \( \rho_1 + \rho_2 \leq \sqrt{\frac{\|P_{\leq I}\|_r}{n}} + \rho_2 \). To prove \( \rho_1 + \rho_2 \gtrsim \eta \sqrt{\frac{\|P_{< I}\|_r}{n}} + \rho_2 \), there are two cases.

- If \( A = I \) then the result is clear.
- Otherwise, \( I > A \). Note that by setting \( p'_i := np_i \) for all \( i = 1, \cdots, N \), the result to show can be rewritten as:

\[
\frac{\|p'_{\leq A}\|_r^2}{\|p'_{\leq I}\|_r^2} + \|p'_{\geq I}\|_1^{2-t} \gtrsim \sqrt{\|p'_{\leq I}\|_r} + \|p'_{\geq I}\|_1^{2-t}. \tag{13}
\]

We have by definition of \( A \) and \( I \):

\[
p'_I^{2-r} \left( \sum_{i \geq I} p'_i \right)^{2-r} = \left( \sum_{i \geq I} p'_I p'_i \right)^{2-r} \geq \left( \sum_{i \geq I} p_i^2 \right)^{2-r} \gtrsim \eta 1 \text{ and}
\]

\[
p'_I^{2b} \sum_{i \leq I} p_i^{tr} \leq p'_A^{2b} \sum_{i \leq I} p_i^{tr} \leq \eta^2 < 1 \text{ by definition of } A.
\]

Hence, by noticing that \( 2b = 2 - r \) we have \( \left( \sum_{i \geq I} p'_i \right)^{2-r} > \sum_{i \leq I} p_i^{tr} \), which yields
\[
\|p'_{\geq I}\|_1^{2-t} \geq \sqrt{\|p'_{\leq I}\|_r} \geq \frac{\|p'_{< A}\|_r^2}{\|p'_{\leq I}\|_r^2} \text{ by raising to the power } \frac{1}{2r}. \]
This condition yields the result of the lemma, by replacing \( p' \) by \( np \). □
Proof of lemma 10. If \( A = I \) then the result is clear. Now, suppose that \( A < I \). We have, by the definition of \( A \):
\[
\frac{l_\eta}{n^2} > p_{A+1}^{2b} \sum_{i \leq I} p_i^2 \geq \sum_{i = A+1}^I p_i^2 \geq p_I \sum_{i = A+1}^I p_i \implies \frac{l_\eta}{n^2 \sum_{i = A+1}^I p_i} \geq p_I
\]
Moreover if \( I < N \),
\[
\frac{C_{\text{tail}}}{n^2} \leq \sum_{i > I} p_i^2 \leq p_{I+1} \sum_{i > I} p_i \implies p_{I+1} \geq \frac{C_{\text{tail}}}{n^2 \sum_{i > I} p_i}
\]
So that
\[
\sum_{i > I} p_i \geq \frac{C_{\text{tail}}}{l_\eta} \sum_{i = A+1}^I p_i
\]
and consequently \( \|p_{> I}\|_1 \gtrsim \|p_{> A}\|_1 \)
Now if \( I = N \), we have \( \|p_{> I}\|_1 = 0 \) and \( p_N > \sqrt{\frac{C_{\text{tail}}}{n}} \) and
\[
p_{A+1}^{2b} < \frac{l_\eta^2}{n^2 \sum_{i=1}^N p_i^2} \implies \sum_{j = A+1}^N p_{j}^{2b} p_{A+1} p_j \leq \frac{l_\eta^2}{n^2}
\]
\[
\implies \sum_{j = A+1}^N p_j^2 \leq \frac{l_\eta^2}{n^2}
\]
\[
\implies p_N \|p_{> A}\|_1 \leq \frac{l_\eta^2}{n^2}
\]
\[
\implies \sqrt{\frac{C_{\text{tail}}}{n}} \|p_{> A}\|_1 \leq \frac{l_\eta^2}{n^2}
\]
Hence \( \|p_{> A}\|_1 \lesssim \frac{1}{n} \) so that \( \|p_{> A}\|_1 + \frac{1}{n} \asymp \|p_{> I}\|_1 + \frac{1}{n} \asymp \frac{1}{n} \)

\[\square\]

B Upper bound

We write \( p, q \) for the vectorisation of \( P, Q \) as in (4), and define \( \Delta = q - p \). In the following, \( c > 0 \) denotes an absolute constant, depending only on \( \eta \). We call
\[
\rho = \sqrt{\frac{\|p_{\leq I}\|_r}{n}} + \frac{\|p_{> I}\|_r^{2-t}}{n^{2-2t}} + \frac{1}{n},
\]
and we prove: \( \rho^* \gtrsim \eta \rho \).

We start with the three following lemmas which control the expectation and variance of the statistics \( T_{\text{bulk}}, T_1, T_2 \). We remind that \( \tilde{n} = \left\lfloor \frac{n}{2} \right\rfloor \).
Lemma 11 (Bounds on expectation and variance of $T_{\text{bulk}}$). The expectation and variance of $T_{\text{bulk}}$ satisfy:

$$
E[T_{\text{bulk}}] = \sum_{i \leq A} \frac{\Delta_i^2}{p_i},
$$

$$
V[T_{\text{bulk}}] \leq \sum_{i \leq A} \frac{1}{p_i^{2\theta}} \left( \frac{q_i^2}{\bar{n}^2} + \frac{2}{\bar{n}} q_i \Delta_i^2 \right).
$$

Lemma 12 (Bounds on expectation and variance of $T_1$). The expectation and variance of $T_1$ satisfy:

$$
E[T_1] = \sum_{i > A} q_i - p_i,
$$

$$
V[T_1] \leq \sum_{i > A} \frac{q_i}{n}.
$$

Lemma 13 (Bounds on expectation and variance of $T_2$). The expectation and variance of $T_2$ satisfy:

$$
E[T_2] = \|(p - q)_A\|^2_2,
$$

$$
V[T_2] \leq \sum_{i > A} \frac{q_i^2}{\bar{n}^2} + \frac{2}{\bar{n}} q_i \Delta_i^2.
$$

We then study the null and alternative hypotheses in the following subsection, bounding the probability of error of the test $\psi$.

B.1 Under the null hypothesis $H_0$.

We start by assuming that $P = Q$. We recall that $\zeta_\eta = \frac{4}{\sqrt{\eta}}$.

Test $\psi_{\text{bulk}}$. Moreover, for the bulk, since $P = Q$, we have by lemma 11: $E[T_{\text{bulk}}] = 0$ and $V[T_{\text{bulk}}] = \sum_{i \leq A} \frac{q_i^2}{\bar{n}^2}$. Therefore by Chebyshev’s inequality:

$$
P \left( T_{\text{bulk}} > \zeta_\eta \sqrt{\sum_{i \leq A} \frac{q_i^2}{\bar{n}^2}} \right) \leq \frac{\eta}{16}
$$

so that:

$$
P (\psi_{\text{bulk}} = 1) \leq \frac{\eta}{16}, \quad (14)
$$
Test $\psi_1$. Since $P = Q$, we have by Lemma 12 that $\mathbf{E}(T_1) = 0$ and $\mathbf{V}(T_1) \leq \sqrt{\frac{\sum_{i > A} p_i}{n}}$. By the same argument $\psi_1$’s type-I error is upper bounded as:

$$P(\psi_1 = 1) = P \left( T_1 > \xi \sqrt{\frac{\sum_{i > A} p_i}{n}} \right) \leq \frac{1}{\xi^2} = \frac{\eta}{16},$$

so that by definition of $\psi_1$

$$P(\psi_1 = 1) \leq \frac{\eta}{16}. \quad (15)$$

Test $\psi_2$. Finally, under the null and since $P = Q$, we have $\mathbf{E}(T_2) = 0$ and $\mathbf{V}(T_2) = \frac{1}{n^2} \sum_{i > A} p_i^2$ by Lemma 13 so that

$$P \left( T_2 > \xi \sqrt{\frac{\sum_{i > A} p_i^2}{n^2}} \right) \leq \frac{\eta}{16},$$

which rewrites:

$$P(\psi_2 = 1) \leq \frac{\eta}{16}. \quad (16)$$

Conclusion: Putting together equations (15), (14) and (16) we get that the type I error of $\psi = \psi_{\text{bulk}} \lor \psi_1 \lor \psi_2$ is upper bounded as

$$P(\psi = 1) \leq \sum_{i \in \{\text{bulk,1,2}\}} P(\psi_i = 1) \leq \frac{3\eta}{16} < \frac{\eta}{2}.$$

B.2 Under the alternative hypothesis $\mathcal{H}_1(\rho)$

Suppose that for some constant $\bar{c}_\eta > 0$, we have $\|\Delta\|_t \geq 2\bar{c}_\eta \rho$. By the triangle inequality, there are two cases:

- **First case:** Either $\|\Delta_{\leq A}\|_t \geq \bar{c}_\eta \rho$
- **Second case:** Or $\|\Delta_{> A}\|_t \geq \bar{c}_\eta \rho$

**Proposition 5** (Study in the **First case**). There exists a large enough constant $\bar{c}_{\eta}^{(\text{bulk})} > 0$ such that if $\|\Delta_{\leq A}\|_t \geq \bar{c}_{\eta}^{(\text{bulk})} \rho$, then

$$P(\psi_{\text{bulk}} = 1) \geq 1 - \eta/6.$$

**Proposition 6** (Study in the **Second case**). If $\|\Delta_{> A}\|_t \geq c\rho$, then

$$P(\psi_1 \lor \psi_2 = 1) \geq 1 - \frac{2\eta}{3}.$$
Proof of Proposition 5. Suppose $\|\Delta_{\leq A}\|_t \geq c\rho$ for some constant $c$. We show that if $c$ is large enough, then the test $\psi_{Bulk}$ will detect it. To do so, we compute a constant $c'$ depending on $c$ such that if $\|\Delta_{\leq A}\|_t \geq c\rho$, then $V(T_{Bulk}) \leq c' E(T_{Bulk})^2$ and such that $\lim_{c \to +\infty} c' = 0$.

By definition of $\rho$, we have in particular: $\|\Delta_{\leq A}\|_t \geq \sqrt{\frac{\|p_{\leq I}\|_r}{n}} \wedge \frac{c}{n}$, hence

$$\frac{1}{n^2} \leq \frac{1}{c^4} \|p_{\leq I}\|_r^2 \wedge \frac{\|\Delta_{\leq A}\|_i^2}{c^2}$$

Using Lemma 11 we split $V[T_{bulk}]$ into four terms

$$V[T_{bulk}] \leq \sum_{i \leq A} \frac{1}{p_i^{2b}} \left( \frac{(p_i + \Delta_i)^2}{n^2} + \frac{2}{n} (p_i + \Delta_i) \Delta_i^2 \right)$$

$$\leq \frac{2}{n^2} \sum_{i \leq A} p_i^r + \frac{2}{n^2} \sum_{i \leq A} \frac{\Delta_i^2}{p_i^{2b}} + \frac{2}{n} \sum_{i \leq A} p_i^{1-2b} \Delta_i^2 + \frac{2}{n} \sum_{i \leq A} \frac{\Delta_i^3}{p_i^{2b}}.$$

Now, we show that each of the four terms is less than $E[T_{bulk}]^2$, up to a constant

**Term 1:** We have by Hölder’s inequality:

$$\sum_{i \leq A} \Delta_i^t \leq \left[ \sum_{i \leq A} \left( \frac{\Delta_i^t}{p_i^t} \right) \right]^{\frac{t}{2}} \left[ \sum_{i \leq A} \left( \frac{p_i^{2b}}{1-t} \right) \right]^{\frac{1}{2}} = \left( \sum_{i \leq A} \frac{\Delta_i^t}{p_i^t} \right)^{\frac{t}{2}} \left( \sum_{i \leq I} p_i^r \right)^{1-\frac{t}{2}}.$$

Hence

$$\|\Delta_{\leq A}\|_t \leq \left( \sum_{i \leq A} \frac{\Delta_i^t}{p_i^t} \right)^{\frac{1}{2}} \left( \sum_{i \leq I} p_i^r \right)^{\frac{2-t}{4}}.$$

Moreover, we have $\frac{1}{n^2} \leq \frac{\|\Delta_{\leq A}\|_t^4}{c^4 \|p_{\leq I}\|_r^2}$ so that the term 1 writes:
\[
\frac{2}{n^2} \sum_{i \leq A} p_i^R \leq 2 \sum_{i \leq A} p_i^R \left( \sum_{i \leq A} \Delta_i^2 \right)^{\frac{4}{n}} \frac{1}{c^A \left( \sum_{i \leq I} p_i^R \right)^{\frac{2}{n}}}
\leq \frac{2}{c^A} \left( \sum_{i \leq A} p_i^R \right)^{1-\frac{2}{r}} \left( \sum_{i \leq A} p_i^b \right) \left( \sum_{i \leq A} p_i^r \right)^{4-2r/4} \text{ by (18)}
= \frac{2}{c^4} \left( \sum_{i \leq A} \Delta_i^2 \right)^{2} = \frac{2}{c^4} \mathbb{E}[T_{\text{bulk}}]^2.
\]

**Term 2:*** The condition \(a \leq p_A^b\) ensures that:
\[
p_A^b \geq a^2 = \frac{l_\eta}{\sqrt{2(\sum_{j \leq I} p_j^r)^{1/2}}} =: \tilde{c} \frac{1}{(\sum_{j \leq I} p_j^r)^{1/2}}.
\]
Using this condition, the term \(2\) writes:
\[
\sum_{i \leq A} \frac{1}{n^2} \frac{\Delta_i^2}{p_i^b} \leq \frac{1}{n^2} \frac{1}{p_A^b} \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \leq \frac{\tilde{c}}{n} \left( \sum_{j \leq I} p_j^r \right)^{\frac{1}{r}} \left( \sum_{i \leq A} p_i^b \right)^{\frac{1}{b}}.
\]
Moreover, since \(\sqrt{\frac{\|P \leq I\|_r}{n}} \lesssim \rho \lesssim \|\Delta \leq A\|_l\) we have, using (18):
\[
\frac{1}{n} \left( \sum_{j \leq I} p_j^r \right)^{\frac{1}{r}} = \frac{1}{n^b} \left( \sqrt{\frac{\|P \leq A\|_r}{n}} \right)^{r} \leq \frac{1}{n^b} \left( \sum_{i \leq A} \Delta_i^2 \right)^{\frac{2}{r}} \left( \sum_{i \leq A} p_i^b \right)^{\frac{1}{b}} \lesssim \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b}.
\]
In the last inequality, we use the fact proved in case number \(1\) that \(\frac{1}{n^b} \left( \sum_{i \leq A} p_i^r \right)^{b} \lesssim \mathbb{E}[T_{\text{bulk}}]^b\) and the relation \(\frac{r}{b} + b = 1\)

Plugging in (20) yields that the second term \(2\) is bounded by \(\mathbb{E}[T_{\text{bulk}}]^2\)
**Term 3:** This term writes:

\[
\frac{1}{n} \sum_{i \leq A} p_i^{1-2b} \Delta_i^2 \leq \frac{\|\Delta_{\leq A}\|_2^2}{c^2 \left( \sum_{i \leq I} p_i^r \right)^\frac{1}{r}} \sum_{i \leq A} p_i^{1-2b} \Delta_i^2 \\
\leq \frac{1}{c^2} \left( \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \right) \left( \sum_{i \leq A} p_i^r \right)^{-\frac{1}{r}} \sum_{i \leq A} p_i^{1-2b} \Delta_i^2 \quad \text{using (18)} \\
\leq \frac{1}{c^2} \left( \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \right) \left( \sum_{i \leq A} p_i^r \right)^{-\frac{1}{r}} \left( \sum_{i \leq A} p_i^{\frac{2}{3}(1-2b)} \Delta_i^\frac{4}{3} \right)^\frac{3}{2} \text{ since } \|\cdot\|_1 \leq \|\cdot\|_\frac{4}{3}.
\]

Moreover, by Hölder’s inequality with \(\frac{1}{\frac{4}{3}} + \frac{1}{3} = 1\):

\[
\sum_{i \leq A} p_i^{\frac{2}{3}(1-2b)} \Delta_i^{\frac{4}{3}} \leq \left( \sum_{i \leq A} \left( \frac{p_i^{\frac{2}{3}(1-2b)} \Delta_i^{\frac{4}{3}}}{p_i^b} \right) \right)^{\frac{3}{2}} \left( \sum_{i \leq A} \left( \frac{p_i^r}{p_i^b} \right) \right)^{\frac{1}{2}} \leq \left( \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \right)^{\frac{2}{3}} \left( \sum_{i \leq A} p_i^r \right)^{\frac{1}{3}}.
\]

So that

\[
\left( \sum_{i \leq A} p_i^{\frac{2}{3}(1-2b)} \Delta_i^{\frac{4}{3}} \right)^{\frac{3}{2}} \leq \left( \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \right) \left( \sum_{i \leq A} p_i^r \right)^{\frac{1}{2}} \text{ i.e. } \left( \sum_{i \leq I} p_i^r \right)^{-\frac{1}{2}} \left( \sum_{i \leq A} p_i^{\frac{2}{3}(1-2b)} \Delta_i^{\frac{4}{3}} \right)^{\frac{3}{2}} \leq \left( \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \right)^{\frac{2}{3}} \left( \sum_{i \leq A} p_i^r \right)^{\frac{1}{3}}.
\]

This yields that the third term satisfies:

\[
\frac{1}{n} \sum_{i \leq A} p_i^{1-2b} \Delta_i^2 \leq \frac{1}{c^2} \left( \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \right)^2 = \frac{1}{c^2} E[T_{\text{bulk}}^2].
\]

**Term 4:** The fourth term writes:

\[
\frac{1}{n} \|\left( \frac{|\Delta_i|_{2b}}{p_i^{2b}} \right)_{i \leq A} \|_3^3 \leq \frac{1}{n} \|\left( \frac{|\Delta_i|_{2b}}{p_i^{2b}} \right)_{i \leq A} \|_3^3 = \frac{1}{n} \left( \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \right)^{\frac{1}{2}} \leq \frac{1}{n^\frac{1}{2}} \left( \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \right)^{\frac{1}{2}} \left( \sum_{i \leq I} p_i^r \right)^{\frac{1}{4}},
\]

\[41\]
where in the last step we have used the fact that

\[ p_i^\frac{1}{3} \geq \frac{1}{(\sum_{i \leq I} p_i^r)^{\frac{1}{3}}} n^{\frac{1}{3}}. \]

Then using (20):

\[ \frac{1}{\sqrt{n}} \left( \sum_{i \leq I} p_i^r \right)^{\frac{1}{4}} \lesssim \left( \sum_{i \leq A} \frac{\Delta_i^2}{p_i^b} \right)^{\frac{1}{4}}. \]

So the term \( 4 \) is upper-bounded by \( \frac{1}{c'} \mathbb{E}[T_{\text{bulk}}]^2 \).

**Conclusion** By Chebyshev’s inequality, the type-II error of \( \psi_{\text{Bulk}} \) is bounded as

\[ \mathbb{P}(\psi_{\text{Bulk}} = 0) = \mathbb{P}\left( T_{\text{bulk}} \leq \frac{\xi}{n} \| P_{\leq A} \| \frac{\tilde{r}}{r} \right) = \mathbb{P}\left( \mathbb{E}(T_{\text{bulk}}) - T_{\text{bulk}} \geq \mathbb{E}(T_{\text{bulk}}) - \frac{\xi}{n} \| P_{\leq A} \| \frac{\tilde{r}}{r} \right) \leq \mathbb{P}\left( \mathbb{V}(T_{\text{bulk}}) \left( \mathbb{E}(T_{\text{bulk}}) - \frac{\xi}{n} \| P_{\leq A} \| \frac{\tilde{r}}{r} \right)^2 \right) \leq \frac{c' \mathbb{E}(T_{\text{bulk}})^2}{\left( \mathbb{E}(T_{\text{bulk}}) - \frac{\xi}{n} \| P_{\leq A} \| \frac{\tilde{r}}{r} \right)^2}. \]

Moreover, using (19), we have that for \( c \) large enough, \( \mathbb{E}(T_{\text{bulk}}) \geq \frac{\xi}{n} \| P_{\leq A} \| \frac{\tilde{r}}{r} \geq 2 \frac{\xi}{n} \| P_{\leq A} \| \frac{\tilde{r}}{r} \) so that the denominator is well defined. Finally, since \( \lim_{c \to +\infty} c' = 0 \), the type-II error of this test goes to 0 as \( c \) goes to infinity, so for \( c \) large enough, the type-II error is upper-bounded by \( \eta / 6 \)

\[ \Box \]

We now move to the proof of Proposition \( \Box \)

**Proof of Proposition \( \Box \)** We will need the two following lemmas:

**Lemma 14.** It holds by definition of \( A \) that: \( \| P_{>A} \|_2^2 \leq \frac{C_A}{n^2} \) for \( C_A = l_{\eta} + c_{\text{tail}} \).

**Proof of lemma \( \Box \)** If \( A = I \) then the result is clear, by definition of \( I \). Otherwise, by definition of \( A \):

\[ p_{A+1}^{2b} \sum_{i \leq I} p_i^r < \frac{l_{\eta}^2}{n^2} \Rightarrow p_{A+1}^{2b} \sum_{i = A+1}^I p_i^r < \frac{l_{\eta}^2}{n^2} \Rightarrow \sum_{i = A+1}^I p_i^2 < \frac{l_{\eta}^2}{n^2} \Rightarrow \sum_{i > A} p_i^2 < \frac{l_{\eta}^2 + c_{\text{tail}}}{n^2} \]

\[ \Box \]
Lemma 15. For fixed \( j > A \), the probability that coordinate \( j \) is observed at least twice is upper-bounded by \( n^2 p_j^2 \).

Proof of Lemma 15. The probability that coordinate \( j \) is observed at least twice is

\[
1 - (1 - p_j)^n - np_j(1 - p_j)^{n-1} \leq 1 - (1 - np_j) - np_j[1 - (n-1)p_j] \leq n^2 p_j^2
\]

\( \square \)

Under \( H_0 \): We upper bound the type-I error of tests \( \psi_1 \) and \( \psi_2 \). For \( \psi_2 \): by Lemma 14

\[
p(\psi_2 = 1) \leq \sum_{j > A} n^2 p_j^2 \leq C A \leq \frac{n^2}{4}.
\]

As to test \( \psi_1 \): \( p(\psi_1 = 1) = p(|T_1| > \xi_n \sqrt{\frac{\sum_{j > A} p_j}{n}}) \leq \frac{n}{4} \) by Chebyshev’s inequality. By union bound, the type-I error of \( \psi_1 \cup \psi_2 \) is less than \( \eta/2 \).

Under \( H_1 \): If \( \|\Delta > A\|_t \geq c \rho \), we now show that either \( \psi_1 \) or \( \psi_2 \) will detect it. Until the end of the proof, we drop from now on the indexation \( > A \) and write only e.g. \( \|p\|_2, \|\Delta\|_2 \) instead of \( \|p_{>A}\|_2, \|\Delta_{>A}\|_2 \).

We have by Hölder’s inequality:

\[
\|\Delta\|_2^{2(t-1)} \|\Delta\|_1^{2-t} \geq \|\Delta\|_t^{t} \geq C \left( \frac{\|p\|_1^{2-t} + \frac{1}{n^t}}{n^{2t-2}} \right) = C \frac{1}{n^{2t-2}} \left( \|p\|_1^{2-t} + \frac{1}{n^{2-t}} \right)
\]

for \( C = C_1 C_2 = \text{where } C_1 = \left( \frac{20}{\eta} (\xi_n + 1) + 1 \right)^{2-t}, C_2 = \left( \frac{1}{4} \left( \log(4/\eta)^2 \lor 9/100 \right) + c_{tail} \right)^{(t-1)/2} \) so that one of the two relations must hold:

\[
\|\Delta\|_2^{2(t-1)} \geq C_2 \frac{1}{n^{2t-2}} \quad \text{or} \quad \|\Delta\|_1^{2-t} \geq C_1 \left( \|p\|_1^{2-t} + \frac{1}{n^{2-t}} \right)
\]

- First case: \( \|\Delta\|_2^{2(t-1)} \geq C_2 \frac{1}{n^{2t-2}} \). Then \( \|\Delta\|_2 \geq C_2^{1/(t-1)} / n \) so that \( \|q\|_2 \geq C_2^{1/(t-1)} / n - \|p\|_2 \geq \frac{1}{n} \left( C_2^{1/(t-1)} - c_{tail} \right) \).

\( \psi_2 \) accepts if, and only if, all coordinates are observed at most once. This probability corresponds to:

\[
q(\forall j > A, N_j = 0 \text{ or } N_j = 1) = \prod_{j > A} \left[ (1 - q_j)^n + nq_j(1 - q_j)^{n-1} \right] = \prod_{j > A} (1 - q_j)^{n-1}(1 + (n-1)q_j) = \prod_{j > A} (1 - q_j)^{n'}(1 + n' q_j), \text{ writing } n' = n - 1
\]
Let \( I_- = \{ j > A : nq_j \leq \frac{1}{2} \} \) and \( I_+ = \{ j > A : nq_j > \frac{1}{2} \} \). Recall that for \( x \in (0, 1/2] \), \( \log(1 + x) \leq x - x^2/3 \). Then, for \( j \in I_- \):

\[
(1 - q_j)n'(1 + n'q_j) = \exp \left\{ n' \log(1 - q_j) + \log(1 + n'q_j) \right\}
\leq \exp \left\{ -n'q_j + n'q_j - \frac{n'^2q_j^2}{3} \right\}
= \exp \left( -\frac{n'^2q_j^2}{3} \right)
\]

Now, for \( j \in I_+ \), we have:

\[
\leq -n'q_j + \log(1 + n'q_j) \leq -\frac{1}{10} n'q_j
\]

using the inequality \(-0.9x + \log(1 + x) \leq 0\) true for all \( x \geq \frac{1}{2} \). Therefore, we have upper bounded the type-II error of \( \psi_2 \) by:

\[
q(\psi = 0) \leq \exp \left( -\frac{1}{3} \sum_{j \in I_-} n'^2q_j^2 - \frac{1}{10} \sum_{j \in I_+} n'q_j \right)
\leq \exp \left( -\frac{1}{3} \sum_{j \in I_-} n'^2q_j^2 - \frac{1}{10} \left( \sum_{j \in I_+} n'^2q_j^2 \right)^{1/2} \right)
= \exp \left( -\frac{1}{3}(S - S_+) - \frac{1}{10} (S_+)^{1/2} \right)
\]

for \( S = \sum_{j > A} n'^2q_j^2 \) and \( S_+ = \sum_{j \in I_+} n'^2q_j^2 \).

Now, \( S_+ \mapsto -\frac{S}{3} + \frac{1}{3} S_+ - \frac{\sqrt{S}}{10} \) is convex over \([0, S]\) so its maximum is reached on the boundaries of the domain and is therefore equal to \((-\frac{\sqrt{S}}{10}) \vee -\frac{S}{3} = -\frac{\sqrt{S}}{3}\) for \( S \geq 9/100 \).

Now, since \( ||q||_2^2 \geq \frac{C_2^{2/(t-1)}}{n^2} \geq \frac{4C_2^{2/(t-1)}}{n^2} \), we have \( S = n'^2||q||_2^2 \geq \log(4/\eta)^2 \vee 9/100 \) which ensures \( q(\psi_2 = 0) \leq \eta/4 \).

- Second case: \( ||\Delta||_1^{2-t} \geq C_1 \left( ||p||_1^{2-t} + \frac{1}{n^{2-t}} \right) \). Then

\[
||\Delta||_1 \geq C_1^{1/(2-t)} \left( ||p||_1 \vee \frac{1}{n} \right) \geq \left( \frac{C_1^{1/(2-t)} - \left( ||p||_1 + \frac{1}{n} \right) \right. \). We will need the following lemma:

**Lemma 16.** If \( \sum_{j > A} \Delta_j \geq 3 \sum_{j > A} p_j \) then \( |\sum_{j > A} \Delta_j| \geq \frac{1}{2} ||\Delta||_1 \)

**Proof.** Define \( J_+ = \{ j > A : q_j \geq p_j \} \) and \( J_- = \{ q_j < p_j \} \). Define also:

\[
s = \frac{\sum_{j > A} \Delta_j}{\sum_{j > A} p_j}, \quad s_+ = \frac{\sum_{j \in J_+} \Delta_j}{\sum_{j \in J_+} p_j}, \quad s_- = \frac{\sum_{j \in J_-} \Delta_j}{\sum_{j \in J_-} p_j}
\]

44
Then by assumption: \( s_+ - s_- = s \geq 3 \). Moreover, \( s_- = \frac{\sum_{j \in J} (p_j - q_0)}{\sum_{j \in J} p_j} \leq 1 \). Thus, \( s_+ \geq 3 \geq 3s_- \) so that \( 2(s_+ - s_-) \geq s_+ + s_- \), which yields the result.

We can now upper bound the type-II error of \( \psi_1 \):

\[
q(\psi_1 = 0) = q \left( \left| \sum \frac{N_j}{n} - p_j \right| \leq \omega \sqrt{\frac{\|p\|_1}{n}} \right) \\
\leq q \left( \left| \sum_{j > A} q_j - p_j \right| - \left| \sum \frac{N_j}{n} - q_j \right| \leq \omega \sqrt{\frac{\|p\|_1}{n}} \right) \text{ by triangular inequality} \\
\leq q \left( \frac{1}{2}q - p\|1 - \omega \sqrt{\frac{\|p\|_1}{n}} \leq \left| \sum \frac{N_j}{n} - q_j \right| \right) \text{ by Lemma } [16] \\
\leq \frac{1}{n} \sum_{j > A} q_j \\
\left( \frac{1}{2}q - p\|1 - \omega \sqrt{\frac{\|p\|_1}{n}} \right)^2 \text{ by Chebyshev's inequality} \\
\leq \frac{\|q\|_1/n}{\left( \frac{1}{2}q\|1 - \frac{1}{2}p\|1 - \omega \sqrt{\frac{\|p\|_1}{n}} \right)^2} \text{ by triangular inequality} \\
\leq \frac{\|q\|_1/n}{\left( \frac{1}{2}q\|1 - (\omega + 1)(\|p\|_1 + 1/n) \right)^2} \text{ using } \sqrt{xy} \leq x + y \\
\leq \frac{\|q\|_1/n}{\left( \frac{1}{2}q\|1 - (\omega + 1)(\|p\|_1 + 1/n) \right)^2} \text{ using } \sqrt{xy} \leq x + y
\]

Now set \( z = (\omega + 1)(\|p\|_1 + 1/n) \). The function \( f : x \mapsto \frac{x}{n(x/2 - z)^2} \) is decreasing. Moreover, for \( x \geq 20z/\omega \), we have:

\[
f(x) \leq \frac{20z/\omega}{n(10z/\omega - z)^2} = \frac{20\omega}{nz(10 - \omega)^2} \leq \frac{20\omega}{81} \leq \omega/4
\]

which proves that, whenever \( \|q\|_1 \geq \frac{20\omega}{\eta}(\omega + 1)(\|p\|_1 + 1/n) \), we have \( q(\psi_1 = 0) \leq \omega/4 \). This condition is guaranteed when \( \|\Delta\|_1 \geq \left( \frac{20\omega}{\eta}(\omega + 1) \right)(\|p\|_1 + 1/n) \) = \( C_1^{1/(2-t)}(\|p\|_1 + 1/n) \)

**Proof of lemma** [12]: Since the graphs incidence matrices \( G^{(1)}, \ldots, G^{(n)} \) are matrices filled with independent Bernoulli random variables (with associated probability matrix \( Q \)), we know that \( S_i, S'_i \) are mutually independent and distributed as \( S_i, S'_i \sim \mathcal{B}(k, q_i) \).
• **Expectation:** We start with the expectation of $T_{\text{bulk}}$ and since the $(S_i, S'_i)_i$ with $S_i, S'_i \sim \mathcal{B}(\bar{n}, q_i)$

\[
E[T_{\text{bulk}}] = \sum_{i \leq A} \frac{1}{p_i^{2b}} \left( E \left[ \frac{S_i}{\bar{n}} - p_i \right] E \left[ \frac{S'_i}{\bar{n}} - p_i \right] \right) \\
= \sum_{i \leq A} \frac{1}{p_i^{2b}} (p_i - q_i)^2.
\]

• **Variance:** We continue with the upper bound on the variance of $T_{\text{bulk}}$

\[
\mathbb{V}(T_{\text{bulk}}) = \sum_{i \leq A} \frac{1}{p_i^{2b}} \left( E \left[ \left( \frac{S_i}{\bar{n}} - p_i \right)^2 \left( \frac{S'_i}{\bar{n}} - p_i \right)^2 \right] - E \left[ \left( \frac{S_i}{\bar{n}} - p_i \right) \left( \frac{S'_i}{\bar{n}} - p_i \right) \right]^2 \right) \\
= \sum_{i \leq A} \frac{1}{p_i^{2b}} \left( E \left[ \left( \frac{S_i}{\bar{n}} - p_i \right)^2 \right]^2 - (p_i - q_i)^4 \right),
\]

Since the $(S_i, S'_i)_i$ are independent. And so by a bias-variance decomposition, and since $S_i, S'_i \sim \mathcal{B}(\bar{n}, q_i)$

\[
\mathbb{V}(T_{\text{bulk}}) = \sum_{i \leq A} \frac{1}{p_i^{2b}} \left( \mathbb{V} \left( \frac{S_i}{\bar{n}} \right) + E \left[ \left( \frac{S_i}{\bar{n}} - p_i \right)^2 \right]^2 - (p_i - q_i)^4 \right) \\
= \sum_{i \leq A} \frac{1}{p_i^{2b}} \left( \frac{q_i (1 - q_i)}{\bar{n}} + (p_i - q_i)^2 \right)^2 - (p_i - q_i)^4 \\
= \sum_{i \leq A} \frac{1}{p_i^{2b}} \left( \frac{q_i^2 (1 - q_i)^2}{\bar{n}^2} + \frac{2}{\bar{n}} q_i (1 - q_i) (p_i - q_i)^2 \right) \\
\leq \sum_{i \leq A} \frac{1}{p_i^{2b}} \left( \frac{q_i^2}{\bar{n}^2} + \frac{2}{\bar{n}} q_i (p_i - q_i)^2 \right).
\]

\[\square\]

*Proof of lemma 12.* Since the graphs incidence matrices $(G_1, \cdots, G_n)$ are matrices filled with independent Bernoulli distributions with associated probability matrix $Q$, we know that the random variables $S_i, S'_i$ are such that $S_i, S'_i \sim \mathcal{B}(k, q_i)$ and are all independent.

We therefore have

\[
E[T_i] = E \left[ \frac{\sum_{i > A} S_i + S'_i}{n} - p_i \right] = \sum_{i > A} p_i - q_i.
\]
and
\[
V[T_1] = V \left[ \sum_{i > A} \frac{S_i + S_i'}{n} \right] = \sum_{i > A} \frac{V[S_i] + V[S_i']}{n^2}
\]
by independence of the \((S_i, S_i')_i\)
\[
= \sum_{i > A} \frac{q_i(1 - q_i)}{n} \leq \sum_{i > A} \frac{q_i}{n}
\]

\[\square\]

Proof of lemma 13. The proof is similar to the proof of lemma 11 by replacing \(b\) with 0 and summing over \(i > A\) instead of \(i \leq A\).
\[\square\]

C  Equivalence between the Binomial, Poisson and Multinomial settings

We now prove that the rates for goodness of fit testing in the Binomial, Poisson and Multinomial case are equivalent.

Proof of Lemma 1. We first prove \(\rho_{Poi}(n, p) \leq C_{BP} \rho_{Bin}(n, p)\). Let \(n \geq 2\), and let \(Y_1, \ldots, Y_n \sim \text{Poi}(q)\). We consider a random function \(\phi\) such that for any Poisson family \(Y_1, \ldots, Y_n \sim \text{Poi}(q)\),
\[
\left\{ \begin{array}{l}
\phi(Y_1, \ldots, Y_n) = (X_1, \ldots, X_{\tilde{n}}) \sim \text{Ber}(q) \quad \text{where} \quad \tilde{n} \sim \text{Poi}(n) \text{ is independent of the } (Y_i)_i \\
\sum_{i=1}^{\tilde{n}} X_i = \sum_{i=1}^{n} Y_i 
\end{array} \right.
\]
In words, \(\phi\) is a function which takes \(n\) Poisson random variables (or equivalently one Poisson random variable \(\text{Poi}(nq)\)) and decomposes them into \(\tilde{n} \sim \text{Poi}(n)\) Bernoulli iid random variables whose sum is \(\sum_{i=1}^{n} Y_i\).

Let \(\tilde{n} \sim \text{Poi}(n)\) be the random length of \(\phi(Y_1, \ldots, Y_n)\). We can choose a small constant \(c = c(\eta)\) such that the event:
\[\mathcal{A}_1 := \{\tilde{n} \geq cn\}\]
has probability larger than \(1 - \eta/4\). Moreover, for \(m \geq cn\) we can define the function
\[
\pi(x_1, \ldots, x_m) = (x_1, \ldots, x_{\lfloor cn\rfloor})
\]
Let $\psi_{\text{Bin}}$ be the test associated to the binomial testing problem:

$$H_0 : q = p \quad \text{v.s.} \quad H_1 : \|p - q\|_t \geq \rho_{\text{Bin}}(cn, p, \frac{\eta}{2})$$

In particular, $R(\psi_{\text{Bin}}) \leq \eta/2$. Now, we define the test

$$\psi = \begin{cases} 
\psi_{\text{Bin}} \circ \pi \circ \phi & \text{if } A_1 \\
0 & \text{otherwise}
\end{cases}$$

and we show that, when associated to the Poissonian testing problem

$$H_0 : q = p \quad \text{v.s.} \quad H_1 : \|p - q\|_t \geq \rho$$

with $\rho = \rho_{\text{Bin}}(cn, p, \frac{\eta}{2})$, it has a risk less than $\eta$. We first analyse its type-I error.

$$\mathbb{P}_{H_0} \left( \psi(Y_1^n) = 1 \right) \leq \mathbb{P}_{H_0} \left( A_1 \cap \psi(Y_1^n) = 1 \right) + \mathbb{P}_{H_0}(\bar{A}_1)$$

$$\leq \mathbb{P}_{H_0} \left( \psi(Y_1^n) = 1 \middle| A_1 \right) + \frac{\eta}{4}$$

$$\leq \mathbb{P}_{H_0} \left( \psi_{\text{Bin}}(X_1, \cdots, X_{\lceil cn \rceil}) = 1 \middle| A_1 \right) + \frac{\eta}{4}$$

$$= \mathbb{P}_{X_1^{\lceil cn \rceil} \sim \text{Ber}(p) \otimes \lceil cn \rceil} \left( \psi_{\text{Bin}}(X_1, \cdots, X_{\lceil cn \rceil}) = 1 \right) + \frac{\eta}{4}$$

For the Type-II error, the same steps show that for any vector $q$:

$$\mathbb{P}_{q} \left( \psi(Y_1^n) = 0 \right) \leq \mathbb{P}_{X_1^{\lceil cn \rceil} \sim \text{Ber}(q) \otimes \lceil cn \rceil} \left( \psi_{\text{Bin}}(X_1, \cdots, X_{\lceil cn \rceil}) = 0 \right) + \frac{\eta}{4}$$

We can now compute the risk of $\psi$ when $\rho = \rho_{\text{Bin}}(cn, p, \frac{\eta}{2})$:

$$R(\psi) = \mathbb{P}_{H_0} \left( \psi(Y_1^n) = 1 \right) + \sup_{\|p - q\|_t \geq \rho} \mathbb{P}_{q} \left( \psi(Y_1^n) = 0 \right)$$

$$\leq \frac{\eta}{2} + \mathbb{P}_{X_1^{\lceil cn \rceil} \sim \text{Ber}(p) \otimes \lceil cn \rceil} \left( \psi_{\text{Bin}}(X_1, \cdots, X_{\lceil cn \rceil}) = 1 \right)$$

$$+ \sup_{\|p - q\|_t \geq \rho} \mathbb{P}_{X_1^{\lceil cn \rceil} \sim \text{Ber}(q) \otimes \lceil cn \rceil} \left( \psi_{\text{Bin}}(X_1, \cdots, X_{\lceil cn \rceil}) = 0 \right)$$

$$= \frac{\eta}{2} + R(\psi_{\text{Bin}})$$

$$= \frac{\eta}{2} + \frac{\eta}{2} = \eta$$

This proves $\rho_{\text{Poi}}(n, p) \leq \rho_{\text{Bin}}(cn, p, \frac{\eta}{2}) \preceq \rho_{\text{Bin}}(n, p, \eta)$.
We now show \( \rho_{\text{Poi}}(n, p) \geq c_{\text{BP}} \rho_{\text{Bin}}(n, p) \). Let \( X_1, \cdots, X_n \sim \text{Ber}(q) \) iid. For some small constant \( \tau > 0 \) let \( \tilde{n} \sim \text{Poi}(\lfloor \tau n \rfloor) \). We choose \( \tau > 0 \) such that
\[
\mathcal{A}_2 = \{ \tilde{n} \leq n \}
\] (22)
has probability larger than \( 1 - \frac{\eta}{4} \). Consider the extended sequence of multivariate Bernoulli random variables \((\tilde{X}_i)_i\) such that
\[
\begin{cases}
\tilde{X}_i = X_i & \text{if } i \leq n \\
\tilde{X}_i \sim \text{Ber}(q) & \text{otherwise}
\end{cases}
\]
and such that \((\tilde{X}_i)_i\) are mutually independent. Let \( Y = \sum_{i=1}^{\tilde{n}} X_i \sim \text{Poi}(\lfloor \tau n \rfloor q) \). The sum is a sufficient statistic of the parameter \( q \) for Poisson random variables so we can define a function
\[
\tilde{\phi}(Y) = (Y_1, \cdots, Y_{\lfloor \tau n \rfloor})
\]
such that \( Y_i \sim \text{Poi}(q) \) iid. Moreover, we set for \( m \leq n \):
\[
\tilde{\pi}(y_1, \cdots, y_n, m) = (y_1, \cdots, y_m)
\]
On \( \mathcal{A}_2 \), we do not even need to extend the sequence of observations. We call \( \psi_{\text{Poi}} \) the test associated to the **Poisson** testing problem:
\[
H_0 : q = p \quad \text{v.s.} \quad H_1 : \|p - q\|_t \geq \rho_{\text{Poi}}(\lfloor \tau n \rfloor, p, \frac{\eta}{2})
\]
We define the randomized test
\[
\tilde{\psi} = \begin{cases} 
\psi_{\text{Poi}} \circ \tilde{\pi} \circ \tilde{\phi}(Y) & \text{if } \mathcal{A}_2 \\
0 & \text{otherwise}
\end{cases}
\] (23)
We show that this test has a risk less than \( \eta \). For the type-I error:
\[
\mathbb{P}_{H_0} \left( \tilde{\psi}(Y) = 1 \right) \leq \mathbb{P}_{H_0} \left( \mathcal{A}_2 \cap \tilde{\psi}(Y) = 1 \right) + \mathbb{P}_{H_0}(\mathcal{A}_2)
\]
\[
\leq \mathbb{P}_{H_0} \left( \tilde{\psi}(Y) = 1 \mid \mathcal{A}_2 \right) + \frac{\eta}{4}
\]
\[
\leq \mathbb{P}_{H_0} \left( \psi_{\text{Poi}}(Y_1, \cdots, Y_{\lfloor \tau n \rfloor}) = 1 \mid \mathcal{A}_2 \right) + \frac{\eta}{4}
\]
\[
= \mathbb{P}_{Y_1, \cdots, Y_{\lfloor \tau n \rfloor} \sim \text{Poi}(p) \otimes \lfloor \tau n \rfloor} \left( \psi_{\text{Poi}}(Y_1, \cdots, Y_{\lfloor \tau n \rfloor}) = 1 \right) + \frac{\eta}{4}
\]
For the Type-II error, the same steps show that for any vector \( q \):
\[
\mathbb{P}_q \left( \tilde{\psi}(Y) = 0 \right) \leq \mathbb{P}_{Y_1, \cdots, Y_{\lfloor \tau n \rfloor} \sim \text{Poi}(q) \otimes \lfloor \tau n \rfloor} \left( \psi_{\text{Poi}}(Y_1, \cdots, Y_{\lfloor \tau n \rfloor}) = 0 \right) + \frac{\eta}{4}
\]
49
We can now compute the risk of $\tilde{\psi}$ when $\rho = \rho_{Poi}(\overline{c}n, p, \frac{\eta}{2})$:

$$
R(\psi) = P_{H_0}(\tilde{\psi}(Y) = 1) + \sup_{\|p-q\| \geq \rho} \mathbb{P}_q(\tilde{\psi}(Y) = 0)
$$

$$
= \frac{\eta}{2} + \mathbb{P}_{Y_1, \ldots, Y_{\overline{c}n} \sim Poi(p)}(\psi_{Poi}(Y_1, \ldots, Y_{\overline{c}n}) = 1)
$$

$$
+ \sup_{\|p-q\| \geq \rho} \mathbb{P}_{Y_1, \ldots, Y_{\overline{c}n} \sim Poi(q)}(\psi_{Poi}(Y_1, \ldots, Y_{\overline{c}n}) = 0)
$$

$$
= \frac{\eta}{2} + R(\psi_{Poi})
$$

$$
= \frac{\eta}{2} + \frac{\eta}{2} = \eta
$$

This proves $\rho_{Bin}(n, p) \leq \rho_{Poi}(\overline{c}n, p, \frac{\eta}{2}) \asymp \rho_{Poi}(n, p, \eta)$.

Proof of Lemma 2: We first prove that $\rho_{Mult}(n, p) \lesssim \rho_{Poi}(n, p^{\max})$ when $\sum p_i = 1$ by following the same steps as for proving $\rho_{Bin} \lesssim \rho_{Poi}$: we draw $\tilde{n} \sim Poi(\overline{c}n)$ and $Z_1, \ldots, Z_{\tilde{n}} \sim \mathcal{M}(q)$. Then the histogram (or fingerprints) is a sufficient statistic of $Z_1, \ldots, Z_{\tilde{n}}$ for $q$. It is defined as

$$
\left( \begin{array}{c}
N_1 \\
\vdots \\
N_d
\end{array} \right) := \left( \begin{array}{c}
\sum_{i=1}^{\tilde{n}} 1\{Z_i = 1\} \\
\vdots \\
\sum_{i=1}^{\tilde{n}} 1\{Z_i = d\}
\end{array} \right) \sim Poi(nq).
$$

On $A_2$, defined in (22), we have

$$
\left( \begin{array}{c}
N_2 \\
\vdots \\
N_d
\end{array} \right) \sim Poi(n(q_2, \ldots, q_d))
$$

so we can just apply the exact same steps to prove that, if $q = p$ then the test $\tilde{\psi}$ from (23) its type-I error is less than $\frac{\eta}{2}$ and if $\|q - p\|_{\mathcal{M}, t} \geq \rho_{Poi}(\overline{c}n, p, \frac{\eta}{2})$, its type-II error is less than $\frac{\eta}{2}$.

We now prove the converse lower bound: $\rho_{Poi}^*(n, p, \eta) \leq \rho_{Mult}^*(n, p, \eta)$. For this, we come back to the prior distributions defined in (5) and (7) except that we do not set any perturbation on $p_1$. This defines a probability distribution $\tilde{p}$ such that $\forall 1 < j \leq A, \tilde{p}_j = p_j = \delta_j \gamma_j$ where $\delta_j \sim \text{Rad}(\frac{1}{2})$ iid and $\forall j \geq U, \tilde{p}_j = \bar{\pi} b_j$ where $b_j \sim \text{Ber}(\frac{2}{\bar{\pi}})$ iid and $\bar{\pi} = \frac{\overline{c}n}{\|p\|_{1}}$. We will

50
project \( \tilde{p} \) onto the simplex so that it is a probability vector. Define \( p' = \frac{\tilde{p}}{\|\tilde{p}\|_1} \). \( p' \) therefore follows a prior distribution on the set of \( d \)-dimensional probability vectors. We now show that this prior concentrates on a zone located at \( \rho^*_{Poi}(n,p) \) from \( p \) (up to a constant), and that it is undetectable when observing \( n \) iid data drawn from \( p' \) where \( p' \) follows this prior.

Now, consider the high probability event

\[
A_3 = \left\{ \left| \sum_{j=2}^{A} \delta_j \gamma_j + \pi \|p_U\|_1 - \sum_{j=2}^{d} p_j \right| \leq \frac{c_1}{\sqrt{n}} \right\}
\]

for a small constant \( c_1 \), and

\[
A_4 = \left\{ \left| \sum_{j=2}^{A} \delta_j \right| \leq c_2 \sqrt{A} \right\}
\]

For the bulk, define

\[
J_+ = \{2 \leq j \leq A : \delta_j = 1\} \quad \text{and} \quad J_- = \{2 \leq j \leq A : \delta_j = -1\}
\]

On \( A_4 \), we have \( \|\gamma_{J_+}\|_t \gtrsim \|\gamma\|_t \) and \( \|\gamma_{J_-}\|_t \gtrsim \|\gamma\|_t \). There are two cases:

- First case: \( \|\tilde{p}\|_1 \geq 1 \). Then we shrink \( \tilde{p} \). This means that
  \[
  \left\| \left( \frac{\tilde{p}}{\|\tilde{p}\|_1} \right)_{J_-} - p_{J_-} \right\|_t \geq \left\| \tilde{p}_{J_-} - p_{J_-} \right\|_t = \|\gamma_{J_-}\|_t \gtrsim \|\gamma\|_t
  \]

- Second case: \( \|\tilde{p}\|_1 < 1 \). Then similarly:
  \[
  \left\| \left( \frac{\tilde{p}}{\|\tilde{p}\|_1} \right)_{J_+} - p_{J_+} \right\|_t \gtrsim \|\gamma\|_t
  \]

In both cases, the rescaled vector \( p' \) is still separated away from the null distribution by a distance at least \( \rho^*_{Poi,Bulk}(n,p) \).

For the tail: On \( A_3 \), we have \( \|\tilde{p}\|_1 \in \left[ \frac{1}{2}, \frac{3}{2} \right] \) so that \( \forall j \geq U : \frac{\tilde{p}_j}{\|\tilde{p}\|_1} \approx \tilde{p}_j \). The exact same calculation as in the proof of lemma \( \square \) shows that, with high probability, \( \|p'_{\geq U} - p_{\geq U}\|_t \gtrsim \frac{\|p_{\geq U}\|_t^{2-t}}{n^{2(t-1)}} \). Combining the above results, we get that \( \|p' - p\|_t \gtrsim \rho^*_{Poi}(n,p) \) and that this prior is indistinguishable from the null distribution, ensuring \( \rho^*_{Poi}(n,p) \lesssim \rho^*_{Mult}(n,p) \). Indeed, on \( A_3 \), the rescaling factor is between 1/2 and 3/2 so that on the bulk, we still have \( \forall j = 2, \ldots, A : |p'_j - p| \leq 2\gamma_j \) and the perturbation \( (\pm \gamma_j)_j \) is already undetectable. \( \square \)