The Annals of Statistics
2014, Vol. 42, No. 3, 940–969
DOI: 10.1214/14-AOS1208

COMMUNITY DETECTION IN DENSE RANDOM NETWORKS

BY ERY ARIAS-CASTRO[1] AND NICOLAS VERZELEN[2]

University of California, San Diego and INRA

Dedicated to the memory of Yuri I. Ingster

We formalize the problem of detecting a community in a network into testing whether in a given (random) graph there is a subgraph that is unusually dense. Specifically, we observe an undirected and unweighted graph on N nodes. Under the null hypothesis, the graph is a realization of an Erdős–Rényi graph with probability p_0. Under the (composite) alternative, there is an unknown subgraph of n nodes where the probability of connection is $p_1 > p_0$. We derive a detection lower bound for detecting such a subgraph in terms of N, n, p_0, p_1 and exhibit a test that achieves that lower bound. We do this both when p_0 is known and unknown. We also consider the problem of testing in polynomial-time. As an aside, we consider the problem of detecting a clique, which is intimately related to the planted clique problem. Our focus in this paper is in the quasi-normal regime where np_0 is either bounded away from zero, or tends to zero slowly.

1. Introduction. In recent years, the problem of detecting communities in networks has received a large amount of attention, with important applications in the social and biological sciences, among others [Fortunato (2010)]. The vast majority of this expansive literature focuses on developing realistic models of (random) networks [Albert and Barabási (2002), Barabási and Albert (1999)], on designing methods for extracting communities from such networks [Girvan and Newman (2002), Newman (2006), Reichardt and Bornholdt (2006)] and on fitting models to network data [Bickel, Chen and Levina (2011)].

The underlying model is that of graph $\mathcal{G} = (\mathcal{E}, \mathcal{V})$, where $\mathcal{E}$ is the set of edges and $\mathcal{V}$ is the set of nodes. For example, in a social network, a node would represent an individual and an edge between two nodes would symbolize a friendship or kinship of some sort shared by these two individuals. In the literature just mentioned, almost all the methodology has concentrated on devising graph partitioning methods, with the end goal of clustering the nodes in $\mathcal{V}$ into groups with strong inner-

Received February 2013; revised February 2014.

[1]Supported in part by a grant from the Office of Naval Research (N00014-13-1-0257).

[2]Supported in part by the french Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

MSC2010 subject classifications. 62C20, 62H30, 94A13.

Key words and phrases. Community detection, detecting a dense subgraph, minimax hypothesis testing, Erdős–Rényi random graph, scan statistic, planted clique problem, dense k-subgraph problem, sparse eigenvalue problem.

connectivity and weak inter-connectivity [Bickel and Chen (2009), Lancichinetti and Fortunato (2009), Newman and Girvan (2004)].

In this euphoria, perhaps the most basic problem of actually detecting the *presence* of a community in an otherwise homogeneous network has been overlooked. For example, imagine a finite number of individuals making a number of, say, binary decisions. A network is then created with the nodes representing the individuals, and where two individuals are connected if they made the same decision, say, 50% of the time. In such a setting,[3] the task of community detection—as we define it here—corresponds to determining whether the individuals are acting independently, or not. Such an abstract setting could have applications in economics, where the individuals may be corporations, and in biology, where the individuals could be genes and the decisions could be their different expressions. It could arise in a dynamic setting where a network is growing over time and monitored for clustering [Heard et al. (2010), Mongiovı et al. (2013), Park, Priebe and Youssef (2013)]. Once a community is detected, the next step is often to identify it, but from a mathematical perspective, probing the limits of detection (i.e., hypothesis testing) often offers insight into what is possible in terms of identification (i.e., estimation).

Many existing community extraction methods can be turned into community detection procedures. For example, one could decide that a community is present in the network if the modularity of Newman and Girvan (2004) exceeds a given threshold. To set this threshold, one needs to define a null model. Newman and Girvan (2004) implicitly assume a random graph conditional on the node degrees. Here, we make the simplest assumption that the null model is an Erdős–Rényi random graph [Bollobás (2001)].

In this context, we also touch on another line of work, that of detecting a clique in a random graph—the so-called planted (or hidden) clique problem [Alon, Krivelevich and Sudakov (1998), Dekel, Gurel-Gurevich and Peres (2011), Feige and Ron (2010)].

1.1. *The framework.* We address a stylized community detection problem, where the task is to detect the presence of clustering in the network and is formalized as a hypothesis testing problem. We observe an *undirected* graph $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ with $N := |\mathcal{V}|$ nodes. Without loss of generality, we take $\mathcal{V} = [N] := \{1, \ldots, N\}$. The corresponding adjacency matrix is denoted $\mathbf{W} \in \{0, 1\}^{N \times N}$, where $W_{i,j} = 1$ if, and only if, $(i, j) \in \mathcal{E}$, meaning there is an edge between nodes $i, j \in \mathcal{V}$. Note that $\mathbf{W}$ is symmetric, and we assume that $W_{ii} = 0$ for all i. Under the null hypothesis, the graph $\mathcal{G}$ is a realization of $\mathbb{G}(N, p_0)$, the Erdős–Rényi random graph on N nodes with probability of connection $p_0 \in (0, 1)$; equivalently, the upper diagonal

[3]This setting is better model with a random dot-product graph, which is more complex than what we consider in this paper.

entries of $\mathbf{W}$ are independent and identically distributed with $\mathbb{P}(W_{i,j} = 1) = p_0$ for any $i \neq j$. Under the alternative, there is a subset of nodes indexed by $S \subset \mathcal{V}$ such that $\mathbb{P}(W_{i,j} = 1) = p_1$ for any $i, j \in S$ with $i \neq j$, with everything else the same. We assume that $p_1 > p_0$, implying that the connectivity is stronger between nodes in S. When $p_1 = 1$, the subgraph with node set S is a clique, meaning all pairs of nodes within S are connected. The subset S is not known, although in most of the paper we assume that its size $n := |S|$ is known.

We study detectability in this framework in asymptotic regimes where $n, N \to \infty$, and p_0, p_1 may also change; all these parameters are assumed to be functions of N. A test ϕ is a function that takes $\mathbf{W}$ as input and returns $\phi = 1$ to claim there is a community in the network, and $\phi = 0$ otherwise. The (worst-case) risk of a test ϕ is defined as

$$\gamma_N(\phi) = \mathbb{P}_0(\phi = 1) + \max_{|S|=n} \mathbb{P}_S(\phi = 0), \tag{1}$$

where $\mathbb{P}_0$ is the distribution under the null and $\mathbb{P}_S$ is the distribution under the alternative where S indexes the community. We say that a sequence of tests (ϕ_N) for a sequence of problems $(\mathbf{W}_N)$ is asymptotically powerful (resp., powerless) if $\gamma_N(\phi_N) \to 0$ (resp., $\to 1$). Practically speaking, a sequence of tests is asymptotically powerless if it does not perform substantially better than any guessing that ignores the adjacency matrix $\mathbf{W}$. We will often speak of a test being powerful or powerless when in fact referring to a sequence of tests and its asymptotic power properties. As the alternative hypothesis is completely symmetric with respect to S, the results presented here remain valid if we replace the maximum in (1) by an average over all subsets S of size n.

1.2. *Closely related work.* We take the beaten path, following the standard approach in statistics for analyzing such composite hypothesis testing problems, in particular, the work of Ingster (1997) and others [Donoho and Jin (2004), Hall and Jin (2010), Ingster and Suslina (2002)] on the detection of a sparse (normal) mean vector. Most closely related to our work is that of Butucea and Ingster (2011). Specializing their results to our setting, they derive lower bounds and upper bounds for the same detection problem when the graph is directed and the probability of connection under the null (denoted p_0) is fixed, which is a situation where the graph is very dense. Their work leaves out the interesting regime where $p_0 \to 0$, which leads to a null model that is much more sparse.

1.3. *Main contribution.* Our main contribution in this paper is to derive a sharp detection boundary for the problem of detecting a community in a network as described above. We focus here on the quasi-normal regime where np_0 is either bounded away from zero, or tends to zero slowly, specifically,

$$\log\left(1 \vee \frac{1}{np_0}\right) = o\left[\log\left(\frac{N}{n}\right)\right]. \tag{2}$$

The other regime is studied in Arias-Castro and Verzelen (2013).

On the one hand, we derive an information theoretic bound that applies to all tests, providing conditions under which all tests are powerless. On the other hand, we display a test that basically achieves the best performance possible. The test is the combination of the two natural tests that arise in Butucea and Ingster (2011) and much of the work in that field [Arias-Castro, Candès and Plan (2011), Cai, Jeng and Jin (2011), Ingster, Tsybakov and Verzelen (2010)]:

- *Total degree test.* This test rejects when the total number of edges is unusually large. This is global in nature in that it cannot be directly turned into a method for extraction.
- *Scan (or maximum modularity) test.* This test amounts to turning modularity (calibrated according to our null model) into a test statistic by rejecting when its maximum value is unusually large. It is strictly speaking the generalized likelihood ratio test under our framework.

We also consider the situation, common in practice, where p_0 is unknown. We derive the corresponding lower bound in this situation and design a test that achieves this bound. The test is again the combination of the two tests:

- *Degree variance test.* This test is based on the differences between two estimates for the degree variance. (Note that the total degree test cannot be calibrated without knowledge of p_0.)
- *Scan test.* This test can be calibrated in various ways when p_0 is unknown, for example, by estimation of p_0 based on the whole graph, or by permutation. We study the former.

Finally, we consider various polynomial-time algorithms, the main one being a convex relaxation of the scan test based on a sparse eigenvalue problem formulation. Our inspiration there comes from the recent work of Berthet and Rigollet (2012). We discuss the discrepancy between the performances of the scan test and the relaxed scan test and compare it with other polynomial-time tests.

We summarize our findings in Tables 1 and 2, where

$$R = \frac{\sqrt{n}(p_1 - p_0)}{\sqrt{p_0(1 - p_0)}}$$

is (up to $\sqrt{n/2}$ factor) the signal-to-noise ratio for deciding whether a given subgraph of size n is unusually dense or not.

In fact, we start by addressing the problem of detecting the presence of a large clique in the graph, and treat it separately, as it is an interesting case in its own right. It is simpler and allows us to focus on the regime where $n/\log N \to \infty$ in the rest of the paper. We establish a lower bound and prove that the following (natural) test achieves that bound:

TABLE 1

Detection boundary and near-optimal algorithms when p_0 is known and when p_0 is unknown. For any sequence a and b going to infinity, $a \lll b$ (resp., $a \ggg b$) means that there exists $\varepsilon > 0$ arbitrarily small such that $a \le b^{1-\varepsilon}$ (resp., $a \ge b^{1+\varepsilon}$). See Section 3 (p_0 known) and Section 4 (p_0 unknown)

p_0 **known**	$n \lll N^{2/3}$	$n \ggg N^{2/3}$
	SCAN TEST	TOT. DEG. TEST
$p_0 \gg \frac{\log(N/n)}{n}$	$R > 2\sqrt{\log(N/n)}$	$R > N/n^{3/2}$
$p_0 \ll \frac{\log(N/n)}{n}$	$R > \frac{2\log(N/n)}{\sqrt{np_0}\log(\log(N/n)/(np_0))}$	
p_0 **unknown**	$n \lll N^{3/4}$	$n \ggg N^{3/4}$
	SCAN TEST	DEG. VAR. TEST
$p_0 \gg \frac{\log(N/n)}{n}$	$R > 2\sqrt{\log(N/n)}$	$R > N^{3/4}/n$
$p_0 \ll \frac{\log(N/n)}{n}$	$R > \frac{2\log(N/n)}{\sqrt{np_0}\log(\log(N/n)/(np_0))}$	$R > N^{3/4}/n$

- *Clique number test.* This tests rejects when the size of the clique number (the size of a largest clique) of the graph is unusually large. It can be calibrated without knowledge of p_0, for example, by permutation, but we do not know of a polynomial-time algorithm that comes even close.

The rest of the paper is organized as follows. In Section 2, we consider the problem of detecting the presence of a large clique and analyze the clique number test. In Section 3, we consider the more general problem of detecting a densely connected subgraph and analyze the total degree test and the scan test. The more realistic situation of unknown p_0 is handled in Section 4. In Section 5, we investigate polynomial-time tests. We then discuss our results and the outlook in Section 6. The technical proofs are postponed to Section 7 and to the supplementary material [Arias-Castro and Verzelen (2014)].

TABLE 2

The performance of our best polynomial time algorithms. See Section 5

p_0 **known**	$n \lll \sqrt{N}$	$n \ggg \sqrt{N}$
	RELAX. SCAN TEST	TOT. DEG. TEST
	$R > \sqrt{2}(N\log N)^{1/4}$	$R > N/n^{3/2}$
p_0 **unknown**	$n \lll \sqrt{N}$	$n \ggg \sqrt{N}$
	RELAX. SCAN TEST	DEG. VAR. TEST
	$R > \sqrt{2}(N\log N)^{1/4}$	$R > N^{3/4}/n$

1.4. *General assumptions and notation.* We assume throughout the paper that $N \to \infty$ and the other parameters n, p_0, p_1 (and more) are allowed to change with N, unless specified otherwise. This dependency is left implicit. In particular, we assume that $n/N \to 0$, emphasizing the situation where the community to be detected is small compared to the size of the whole network. (When n is of the same order as N, the total degree test is basically optimal.) We assume that p_0 is bounded away from 1, which is the most interesting case by far, and that $N^2 p_0 \to \infty$, the latter implying that the number of edges in the network (under the null) is not bounded. We also hypothesize that either $p_1 = 1$, or $n \to \infty$ with $n^2 p_1 \to \infty$, for otherwise there is a nonvanishing chance that the community does not contain any edges, precluding any test to be powerful.

We say that the test that rejects for large values of a (real-valued) statistic T is asymptotically powerful if there is a critical value $t = t(N)$ such that the test $\{T \geq t\}$ has risk (1) tending to 0. The choice of t that makes this possible may depend on p_1. (In practice, t is chosen to set the probability of type I error, which does not necessitate knowledge of p_1 as long as T itself does not depend on p_1, which is the case of all the tests we consider here, except for the likelihood ratio test.) Similarly, we say that the test is asymptotically powerless if, for any sequence of reals $t = t(N)$, the risk of the test $\{T \geq t\}$ is at least 1 in the limit.

We use standard notation such as $a_n \sim b_n$ when $a_n/b_n \to 1$; $a_n = o(b_n)$ when $a_n/b_n \to 0$; $a_n = O(b_n)$ when a_n/b_n is bounded; $a_n \asymp b_n$ when $a_n = O(b_n)$ and $b_n = O(a_n)$; $a_n \prec b_n$ when there exists a positive constant C such that $a_n \leq C b_n$. For an integer n, let $n^{(2)} = n(n-1)/2$.

Because of its importance in describing the tails of the binomial distribution, the following function—which is the relative entropy or Kullback–Leibler divergence of Bern(q) to Bern(p)—will appear in our results:

$$H_p(q) = q \log\left(\frac{q}{p}\right) + (1-q)\log\left(\frac{1-q}{1-p}\right), \qquad p, q \in (0,1). \tag{3}$$

We will only consider $H_p(q)$ with $q \geq p$.

2. Detecting a large clique in a random graph. We start with specializing the setting to that of detecting a large clique, meaning we consider the special case where $p_1 = 1$. In this section, n is not necessarily increasing with N.

2.1. *Lower bound.* We establish the detection boundary, giving sufficient conditions for the problem to be too hard for any test.

THEOREM 1. *All tests are asymptotically powerless if*

$$\binom{N}{n} p_0^{n(n-1)/2} \to \infty. \tag{4}$$

The result is, in fact, very intuitive. Condition (4) implies that, with high probability under the null, the clique number is at least n, which is the size of the clique planted under the alternative. This is a classical result in random graph theory, and finer results are known; see [Bollobás (2001), Chapter 11]. This is not enough to prove the result, however, as the clique number could still be even larger under the alternative; and even if this is not the case, it would only imply that the clique number test is powerless, but would not say anything about other tests. To prove Theorem 1, we adopt the standard approach based on studying the likelihood ratio test when the clique is chosen uniformly at random; see, for example, Lehmann and Romano (2005), Chapter 8, or Tsybakov (2009), Chapter 2. In this specific setting, the second moment method—which consists in showing that the variance of the likelihood ratio tends to 0—suffices.

2.2. *The clique number test.* Computational considerations aside, the most natural test for detecting the presence of a clique is the clique number test defined in the Introduction. We obtain the following.

PROPOSITION 1. *The clique number test is asymptotically powerful if*

$$\binom{N}{n} p_0^{n(n-1)/2} \to 0. \tag{5}$$

Hence, the clique number test is able to achieve the detection boundary established in Theorem 1. The proof is entirely based on the fact that, when (5) holds, the clique number under the null is at most $n-1$ with high probability [Bollobás (2001), Theorem 11.6], while it is at least n under the alternative. (Thus, the proof is omitted.)

3. Detecting a dense subgraph in a random graph. We now consider the more general setting of detecting a dense subgraph in a random graph. We start with an information bound that applies to all tests, regardless of their computational requirements. We then study the total degree test and the scan test, showing that the test that combines them with a simple Bonferroni correction is essentially optimal.

3.1. *Lower bound.* When assuming infinite computational power, what is left is the purely statistical challenge of detecting the subgraph. For simplicity, we assume that n is not too small, specifically,

$$\frac{n}{\log N} \to \infty, \tag{6}$$

though our result below partially extends to the case where $n = O(\log N)$, particularly when p_1 is constant. As usual, a minimax lower bound is derived by choosing a prior over the composite alternative. Assuming that p_0 and p_1 are known,

because of symmetry, the uniform prior over the community S is least favorable, so that we consider testing

$$H_0 : \mathcal{G} \sim \mathbb{G}(N, p_0) \text{ versus } \bar{H}_1 : \mathcal{G} \sim \mathbb{G}(N, p_0; n, p_1), \tag{7}$$

where the latter is the model where the community S is chosen uniformly at random among subset of nodes of size n, and then for $i \neq j$, $\mathbb{P}(W_{i,j} = 1) = p_1$ if $i, j \in S$, while $\mathbb{P}(W_{i,j} = 1) = p_0$ otherwise. For this simple versus simple testing problem, the likelihood ratio test is optimal, which is what we examine to derive the following lower bound. Remember the entropy function defined in (3).

THEOREM 2. *Assuming* (6) *and* (2) *hold, all tests are asymptotically powerless if*

$$\frac{p_1 - p_0}{\sqrt{p_0}} \frac{n^2}{N} \to 0 \tag{8}$$

and

$$\limsup \frac{n H_{p_0}(p_1)}{2 \log(N/n)} < 1. \tag{9}$$

Conditions (8) and (9) have their equivalent in the work of Butucea and Ingster (2011). That said, (9) is more complex here because of the different behaviors of the entropy function according to whether p_1/p_0 is small or large—corresponding to the difference between large deviations and moderate deviations of the binomial distribution. Only in the case where $p_1/p_0 \to 1$ is the normal approximation to the binomial in effect.

To better appreciate (9), note that it is equivalent to

$$\limsup \frac{(p_1 - p_0)^2}{4 p_0 (1 - p_0)} \frac{n}{\log(N/n)} < 1 \qquad \text{when } \frac{n p_0}{\log(N/n)} \to \infty \tag{10}$$

and

$$\limsup \frac{p_1}{2(1 - p_0)} \frac{n}{\log(N/n)} \log\left(\frac{\log(N/n)}{n p_0} \right) < 1 \qquad \text{when } \frac{n p_0}{\log(N/n)} \to 0. \tag{11}$$

In (10), np_0 is larger and only the moderate deviations of the binomial distribution are involved, while in (11), np_0 is smaller and the large deviations come into play.

Theorem 2 happens to be sharp because, as we show next, the test that combines the total degree test and the scan test is asymptotically powerful when conditions (8) and (9) are—roughly speaking—reversed.

3.2. *The total degree test.* The total degree test rejects for large values of

$$W := \sum_{1 \le i < j \le N} W_{i,j}. \tag{12}$$

PROPOSITION 2. *The total degree test is asymptotically powerful if*

$$\frac{p_1 - p_0}{\sqrt{p_0}} \frac{n^2}{N} \to \infty. \tag{13}$$

It is equally straightforward to show that the total degree test has risk strictly less than one—meaning has some nonnegligible power—when the same ratio tends to a strictly positive and finite constant, while it is asymptotically powerless when that ratio tends to zero.

3.3. *The scan test.* The scan test is another name for the generalized likelihood ratio test, and corresponds to the test that is based on the maximum modularity (calibrated according to our null model). It is particularly simple when p_0 is known, as it rejects for large values of

$$W_n^* := \max_{|S|=n} W_S, \qquad W_S := \sum_{i,j \in S, i<j} W_{i,j}. \tag{14}$$

Unlike the total degree (12), the scan statistic (14) has an intricate distribution as the partial sums W_S are not independent. Nevertheless, the union bound and standard tail bounds for the binomial distribution lead to the following result.

PROPOSITION 3. *The scan test is asymptotically powerful if*

$$\liminf \frac{n H_{p_0}(p_1)}{2 \log(N/n)} > 1. \tag{15}$$

3.4. *The combined test.* Having studied these two tests individually, we are now in a position to consider them together, by which we mean a simple Bonferroni combination which rejects when either of the two tests rejects. Looking back at our lower bound and the performance bounds, we established for these tests, we come to the following conclusion. When the limit in (8) is infinite—yielding (13)—then the total degree test is asymptotically powerful by Proposition 2. When the limit inferior in (9) exceeds one—yielding (15)—then the scan test is asymptotically powerful by Proposition 3.

3.5. *Adaptation to unknown n.* The scan statistic in (14) requires knowledge of n. When this is unknown, the common procedure is to combine the scan tests

at all different sizes n using a simple Bonferroni correction, which amounts to considering a test that rejects for large values of

$$\max_{n \geq u_N} \frac{W_n^*}{w_n}, \tag{16}$$

for a carefully chosen sequence of positive reals (w_n), and a sequence $u_N \to \infty$ slowly. We call this the multiscale scan test. This is done in Butucea and Ingster (2011), with the conclusion that the resulting test is essentially as powerful as the individual tests. It is straightforward to see that here, too, the tail bound used in the proof of Proposition 3 allows for enough room to scan over all subgraphs of all sizes.

PROPOSITION 4. *Assuming* (6), *the multiscale scan test with* $u_N = \log N$ *and*

$$w_n = n^{(2)} H_{p_0}^{-1}\left[2\frac{\log(N/n)+2}{n-1}\right]$$

is asymptotically powerful when (15) *holds*.

4. When p_0 is unknown: The fixed expected total degree model. Although it leads to interesting mathematics, the setting where p_0 is known is, for the most part, impractical. In this section, we evaluate how not knowing p_0 changes the difficulty of the problem. Formalizing the situation where p_0 is unknown amounts to considering the same hypothesis testing problem, but maximize the risk over relevant subsets of p_0's and p_1's, since now even the null hypothesis is composite:

$$\gamma_N^u(\phi) = \sup_{p_0 \in \mathcal{C}_0} \mathbb{P}_0(\phi = 1) + \sup_{(p_0,p_1)\in\mathcal{C}_1} \max_{|S|=n} \mathbb{P}_S(\phi = 0), \tag{17}$$

with collections $\mathcal{C}_0 \subset (0,1)$ and $\mathcal{C}_1 \subset (0,1)^2$.

Lower bounding the optimal detection boundary for this new problem is not straightforward. As a first step, we reduce our problem to an alternate testing problem, where the graph has the same expected total degree under the null and under the alternative hypotheses. We still observe a graph $\mathcal{G}$ on N nodes. As before, under the null, $\mathcal{G}$ is Erdős–Rényi with parameter p_0. Under the alternative where S (still of size n) is the community, $\mathbb{P}(W_{ij} = 1) = p_1$ if $i, j \in S$, while $\mathbb{P}(W_{ij} = 1) = p_0'$ otherwise, where $p_0 = p_0' + (p_1 - p_0')\frac{n^{(2)}}{N^{(2)}}$; let $\mathbb{P}_S'$ and $\mathbb{E}_S'$ denote the probability distribution and expectation under that model. We check that, indeed,

$$\mathbb{E}_S'(W) = N^{(2)} p_0' + n^{(2)}(p_1 - p_0') = N^{(2)} p_0 = \mathbb{E}_0(W).$$

Note that we still assume that p_0, p_1, n are known to the statistician. The risk of a test ϕ for this problem is defined as

$$\gamma_N'(\phi) = \mathbb{P}_0(\phi = 1) + \max_{|S|=n} \mathbb{P}_S'(\phi = 0).$$

Observe that, for all tests ϕ, $\gamma'_N(\phi) \le \gamma^u_N(\phi)$ as soon as $p_0 \in \mathcal{C}_0$ and $(p'_0, p_1) \in \mathcal{C}_1$ so that the detection boundary for the fixed expected total degree model is smaller than for the unknown p_0 setting.

To obtain a lower bound on the minimax risk, we do as in (7) and reduce it to the following simple versus simple testing problem:

$$H_0 : \mathcal{G} \sim \mathbb{G}(N, p_0) \text{ versus } \bar{H}'_1 : \mathcal{G} \sim \mathbb{G}(N, p'_0; n, p_1). \tag{18}$$

As we did before, we first compute the detection boundary for the testing problem (18), and then exhibit some tests achieving this detection boundary. Interestingly, these tests do not require the knowledge of p_0 and p_1, or even n, so that they can be used in the original setting (7) when these parameters are unknown. This will also imply that the detection boundaries are the same in the expected total degree model and in the unknown p_0 setting.

4.1. *Lower bound.*

THEOREM 3. *Assuming* (6) *holds and that*

$$\log\left(1 \vee \frac{1}{np'_0}\right) = o\left[\log\left(\frac{N}{n}\right)\right], \tag{19}$$

all tests are asymptotically powerless for the problem (18) *if*

$$\frac{p_1 - p'_0}{\sqrt{p'_0}} \frac{n^{3/2}}{N^{3/4}} \to 0 \tag{20}$$

and

$$\limsup \frac{n H_{p'_0}(p_1)}{2\log(N/n)} < 1. \tag{21}$$

Comparing with Theorem 2, where p_0 is assumed to be known, condition (20) is substantially weaker than the corresponding condition (8), while we shall see in the proof that (21) is comparable to (9). That said, when $n^2 < N$, the entropy condition (9) is a stronger requirement than either (8) or (20), implying that the setting where p_0 is known and the setting where unknown are asymptotically as difficult in that case.

4.2. *Degree variance test.* By construction, the total degree W has the same expectation under the null and under the alternative in the testing problem with fixed expected total degree—and same variance also up to second order—making it difficult to fruitfully use this statistic in this context.

We design instead a test based on comparing the two estimators for the node degree variance, not unlike an analysis of variance. Let

$$W_{i\cdot} = \sum_{j \ne i} W_{i,j} \tag{22}$$

denote the degree of node i in the whole network. The first estimate is simply the maximum likelihood estimator under the null

$$V_1 = (N-1)\frac{N^{(2)}}{N^{(2)}-1}\hat{p}_0(1-\hat{p}_0), \qquad \hat{p}_0 := \frac{W}{N^{(2)}}.$$

The second estimator is some sort of sample variance, modified to account for the fact that the $W_{i\cdot}$'s are not independent

$$V_2 = \frac{1}{N-2}\sum_{i=1}^{N}\bigl(W_{i\cdot} - (N-1)\hat{p}_0\bigr)^2.$$

Both estimators are unbiased for the degree variance under the null, meaning, $\mathbb{E}_0 V_1 = \mathbb{E}_0 V_2 = (N-1)p_0(1-p_0)$. Under the alternative, V_2 tends to be larger than V_1, leading to a test that rejects for large values of

$$V^* := \frac{V}{\sqrt{N}\hat{p}_0}, \qquad V := V_2 - V_1. \tag{23}$$

PROPOSITION 5. *Assume that* $\liminf Np_0 > 2$. *The degree variance test is asymptotically powerful under fixed expected total degree if*

$$\frac{(p_1 - p_0')^2}{p_0'}\frac{n^3}{N^{3/2}} \to \infty. \tag{24}$$

The test based on V^* achieves the part (20) of the detection boundary. We note that computing V^* does not require knowledge of p_0, p_1 or n, and in fact, its calibration can be done without any knowledge of these parameters via a form of parametric bootstrap, as we do for the scan test below.

4.3. *The scan test.* When p_0 is not available a priori, we have at least three options:

- *Estimate* p_0. We replace p_0 with its maximum likelihood estimator under the null, that is, $\hat{p}_0 = W/N^{(2)}$, and then compare the magnitude of the observed scan statistic (14) with what one would get under a random graph model with probability of connection equal to $\hat{p}_0$.
- *Generalized likelihood ratio test.* We simply implement the actual generalized likelihood ratio test [Kulldorff (1997)], which rejects for large values of

$$\max_{|S|=n}\bigl[n^{(2)}h(\hat{p}_{1,S}) + \bigl(N^{(2)} - n^{(2)}\bigr)h(\hat{p}_{0,S}) - N^{(2)}h(\hat{p}_0)\bigr],$$

where $h(p) := p\log p + (1-p)\log(1-p)$, $\hat{p}_0$ as above, and

$$\hat{p}_{1,S} := \frac{W_S}{n^{(2)}}, \qquad \hat{p}_{0,S} := \frac{W - W_S}{N^{(2)} - n^{(2)}},$$

which are the maximum likelihood estimates of p_1 and p_0 for a given subset S.

- *Calibration by permutation.* We compare the observed value of the scan statistic to simulated values obtained by generating a random graph with either the same number of edges—which leads to a calibration very similar to the first option—or the same degree distribution—which is the basis for in the modularity function of Newman and Girvan (2004).

We focus on the first option.

PROPOSITION 6. *Assume that* $\liminf p_0 N^2/n > 1$. *The scan test that rejects for* $W_n^* \geq n^{(2)} H_{\hat{p}_0}^{-1}[2\frac{\log(N/n)+2}{n-1}]$ *is asymptotically powerful under fixed expected total degree if*

$$\liminf \frac{n H_{p_0'}(p_1)}{2\log(N/n)} > 1. \tag{25}$$

Hence, the scan test calibrated by estimation of p_0 achieves the entropy condition (9) without requiring the knowledge of p_0 or p_1. We mention that adaptation to unknown n may be achieved as described in Section 3.5.

4.4. *Combined test and full adaptation to unknown* p_0. A combination of the degree variance test and of the scan test calibrated by estimation of p_0 is seen to achieve the detection boundary established in Theorem 3, without requiring knowledge of p_0 or p_1.

Recall the definition of the risk $\gamma_N^u(\phi)$ in (17).

PROPOSITION 7. *Consider any fixed number* $\varepsilon > 0$ *and any sequence* $v_n = o(1)$. *Define* $\mathcal{C}_0$ *as the collection of sequences* p_0 *such that* $p_0 \leq 1/2$ *and* $Np_0 \geq 1$. *Define* $\mathcal{C}_1$ *as the collection of sequences* (p_0', p_1) *with* $p_0' \leq 1/2$, $Np_0' \geq 1$ *and either*

$$\frac{n H_{p_0'}(p_1)}{2\log(N/n)} \geq 1+\varepsilon \quad \text{or} \quad \frac{(p_1-p_0')^2}{p_0'}\frac{n^3}{N^{3/2}} v_n \geq 1. \tag{26}$$

Then the test ϕ *that rejects if* $V^* \geq v_n^{-1/2}$ *or* $W_n^* \geq n^{(2)} H_{\hat{p}_0}^{-1}[2\frac{\log(N/n)+2}{n-1}]$ *satisfies* $\gamma_N^u(\phi) = o(1)$.

In particular, this entails that the minimax detection boundary for the unknown p_0 problem is the same as for the fixed expected degree model. Adaptation to n can be handled as in the previous section.

5. Testing in polynomial-time. A question of particular importance in modern times is determining the tradeoff between statistical performance and computational complexity. At the most basic level, this boils down to answering the following question: *What can be done in polynomial-time?*

While computing the total degree (12) or the degree variance statistic (23) can be done in linear time in the size of the network, that is, in $O(N^2)$ time, computing the scan statistic (14) seems intractable: it is NP-hard by reduction to the clique problem (that of computing the clique number) [Karp (1972)], which is even hard to approximate [Zuckerman (2006)].

We consider below various polynomial-time alternatives. The main test in this section, the one for which we prove the strongest performance, is the relaxed scan test presented in Section 5.2 below.

5.1. *The dense n-subgraph problem.* Feige, Kortsarz and Peleg (2001) consider the closely related (but harder) problem of identifying $S \subset \mathcal{V}$ that maximizes W_S among subsets of nodes of size n. They call this the dense n-subgraph problem. The algorithm they develop achieves, as far as we know, the best approximation ratio for the scan statistic. More precisely, their procedure returns some quantity A_n^* satisfying

$$N^{-\delta} W_n^* \le A_n^* \le W_n^*,$$

where δ is a universal constant between $1/3$ and $5/18$, which is not made explicit in [Feige, Kortsarz and Peleg (2001)]. Arguing as in the analysis of the scan test, we derive that the approximate scan test based on A_n^* is asymptotically powerful if $p_0 \succ \log(N/n)/n$ and $p_1 \succ p_0 N^{\delta}$. Comparing this with Proposition 8, our performance guarantee for the relaxed scan test are stronger in the regime $n \ll N^{1/2}$ and $p_0 \succ \log(N/n)/n$.

5.2. *Convex relaxation scan test.* We now suggest a convex relaxation to the problem of computing the scan statistic. To do so, we follow the footsteps of Berthet and Rigollet (2012), who consider the problem of detecting a sparse principal component based on a sample from a multivariate Gaussian distribution in dimension N. Assuming the sparse component has at most n nonzero entries, they show that a near-optimal procedure is based on the largest eigenvalue of any n-by-n submatrix of the sample covariance matrix. Computing this statistic is NP-hard, so they resort to the convex relaxation of d'Aspremont et al. (2007), which they also study. We apply their procedure to $\mathbf{W}^2$.

Formally, for a positive semidefinite matrix $\mathbf{B} \in \mathbb{R}^{N \times N}$ and $1 \le n \le N$, define

$$\lambda_n^{\max}(\mathbf{B}) = \max_{|S|=n} \lambda^{\max}(\mathbf{B}_S),$$

where $\mathbf{B}_S$ denotes the principal submatrix of $\mathbf{B}$ indexed by $S \subset \{1, \dots, N\}$ and $\lambda^{\max}(\mathbf{B})$ the largest eigenvalue of $\mathbf{B}$. d'Aspremont et al. (2007) relaxed this to

$$\mathrm{SDP}_n(\mathbf{B}) = \max_{\mathbf{Z}} \mathrm{Trace}(\mathbf{B}\mathbf{Z}) \qquad \text{subject to } \mathbf{Z} \succeq 0, \mathrm{Trace}(\mathbf{Z}) = 1, |\mathbf{Z}|_1 \le n,$$

where the maximum is over positive semidefinite matrices $\mathbf{Z} = (Z_{st}) \in \mathbb{R}^{N\times N}$ and $|\mathbf{Z}|_1 = \sum_{s,t} |Z_{st}|$. We consider the relaxed scan test, which rejects for large values of

$$\mathrm{SDP}_n(\mathbf{W}^2). \tag{27}$$

When p_0 is known, we simply calibrate the procedure by Monte Carlo simulations, effectively generating $\mathbf{W}_1, \dots, \mathbf{W}_B$ i.i.d. from $\mathbb{G}(N, p_0)$ and computing $\mathrm{SDP}_n(\mathbf{W}_b^2)$ for each $b = 1, \dots, B$, and estimating the p-value by the fraction of b's such that $\mathrm{SDP}_n(\mathbf{W}_b^2) \geq \mathrm{SDP}_n(\mathbf{W}^2)$. Typically B is a large number, and below we consider the asymptote where $B = \infty$.

When p_0 is unknown, we estimate p_0 as we did for the scan test in Proposition 6, and then calibrate the statistic by Monte Carlo, effectively using a form of parametric bootstrap.

In either case, we have the following.

PROPOSITION 8. *Assume that* (2) *holds and* $n \leq N^{1/2-t}$ *for some* $t > 0$. *Then, the relaxed scan test is asymptotically powerful if*

$$\liminf \frac{n}{\sqrt{N\log(N)}} \frac{(p_1 - p_0)^2}{p_0} > 2. \tag{28}$$

To gain some insights on the relative performance of the scan test and the relaxed scan test, let us assume that $n^2 \ll N$, and $np_0 \gg \log(N/n)$. Applying Proposition 3 (or Proposition 6) in this setting, we find that the scan test is asymptotically powerful when

$$\frac{(p_1 - p_0)^2}{p_0} \succ \frac{\log(N/n)}{n}.$$

Thus, comparing with (28), we lose a factor of $\sqrt{N/\log(N)}$ when using the relaxed version. In the regime where $n^2 \gg N\log(N)$, the total degree test and degree variance test both have stronger theoretical guarantees established in Proposition 2 and Proposition 5, respectively. Below we explain why the $\sqrt{N/\log(N)}$ loss is not unexpected. Consider the specific case where $p_0 = 1/2$ and $p_1 = 1$, known as the planted clique problem [Feige and Ron (2010)]. According to Proposition 8, the power of the relaxed scan test in the planted clique problem is not guaranteed for $n = o(\sqrt{N})$, while the clique test can detect a clique of size $n \asymp \log N$, as shown in Proposition 1. In fact, there is no known polynomial-time algorithm that can detect a clique of size $n = o(\sqrt{N})$ [Dekel, Gurel-Gurevich and Peres (2011)] and the problem is provably hard in some computational models, such as monotone circuits [Feldman et al. (2012), Rossman (2010)]. We refer to [Berthet and Rigollet (2012)] for a thorough discussion of the difficulty of the planted clique problem.

5.3. *Densest subgraph test.* Another possible avenue for designing computationally tractable tests for the problem at hand lies in algorithms for finding dense subgraphs of a given size. We follow [Khuller and Saha (2009)], where the reader will find appropriate references and additional results. Define the density of a subgraph $S \subset \mathcal{V}$ as

$$h(S) = \frac{W_S}{|S|}.$$

Although maximizing $h(S)$ among subsets of given size n reduces this to the densest n-subgraph problem described earlier, the same optimization with constraint on $|S|$ may be done in polynomial-time.

PROPOSITION 9. *Assume that* $p_0 \gg \log(N)/N$.

1. *Under the null hypothesis*,

$$\max_S h(S) \sim_{\mathbb{P}_0} h(\mathcal{V}) \sim Np_0/2,$$

and this maximum is achieved at subsets S *satisfying* $|S| \sim N$.

2. *The densest subgraph test is asymptotically powerful if*

$$\liminf \frac{np_1}{Np_0} > 1.$$

3. *Assume that* $\frac{np_1}{Np_0} \to 0$. *Under the alternative hypothesis*,

$$\max_S h(S) \sim_{\mathbb{P}_S} h(\mathcal{V}) \sim_{\mathbb{P}_S} Np_0/2,$$

and this maximum is achieved at subsets S *satisfying* $|S| \sim N$.

The condition $\liminf \frac{np_1}{Np_0} > 1$ is stronger than what we obtained for the relaxed scan test in (28) in the regime where $n \le N^{1/2-t}$ for some $t > 0$, and also stronger than what we obtained for the total degree test (13) and the degree variance test (24) in the regime $n \gg \sqrt{N}$. If $np_1/Np_0 \to 0$, then the densest subgraph statistic seems to behave like the total degree statistic and we therefore expect similar performances although we have no proof of this statement.

Maximizing $h(S)$ over subsets of size $|S| \ge n$ is harder, but can be approximated within a constant factor in polynomial-time. However, the power of the resulting test is not better. Indeed, this test is asymptotically powerful only if $np_1 \ge CNp_0$ where C is positive constant that depends on this approximation factor.

5.4. *The maximum degree test.* Consider the test based on the maximum degree

$$\max_{i=1,\dots,N} W_{i\cdot}, \tag{29}$$

where $W_{i\cdot}$ is the degree of node i in the graph, defined in (22).

PROPOSITION 10. *The maximum degree test is asymptotically powerful if* $p_0 \gg \log(N)/N$ *and*

$$\liminf \frac{n^2}{N\log(N)} \frac{(p_1 - p_0)^2}{p_0(1-p_0)} > 2.$$

Under condition (2), *the maximum degree test is asymptotically powerless if* $\limsup \log(n)/\log(N) < 1$ *and*

$$\frac{n^2}{N\log(N)} \frac{(p_1 - p_0)^2}{p_0(1-p_0)} \to 0. \tag{30}$$

Comparing with Propositions 2 and 8, we observe that the maximum degree test is either less powerful than the relaxed scan test (when $n \le N^{1/2-t}$ for any $t > 0$) or less powerful than the total degree test [when $n \gg \sqrt{N/\log(N)}$]. For unknown p_0, the maximum degree test (which can be calibrated as we did for the scan test) is also less powerful than the degree variance test.

6. Discussion. With this paper, we have established the fundamental statistical (information theoretic) difficulty of detecting a community in a network, modeled as the detection of an unusually dense subgraph within an Erdős–Rényi random graph, in the quasi-normal regime where np_0 is not too small as made explicit in (2). When np_0 is smaller, the arguments are more complex and a number of other tests, not presented here, play an important role. This is detailed in our recent work [Arias-Castro and Verzelen (2013)]. For the time being, in the quasi-normal regime, we learned the following. In the setting where $n \gg N^{2/3}$ for known p_0, and $n \gg N^{3/4}$ for unknown p_0, this detection boundary is achieved by the total degree test and the degree variance test, respectively, which can be computed in polynomial-time. Otherwise, there is a large discrepancy between the information theoretic detection boundary, achieved by the scan test, and what polynomial tests are shown to achieve, which in view of the planted clique problem is not surprising. It is of great interest to study this optimal detection boundary, this time under computational constraints, a theme of contemporary importance in statistics, machine learning and computer science. This promisingly rich line of research is well beyond the scope of the present paper.

7. Proofs.

7.1. *Auxiliary results.* The following is Chernoff's bound for the binomial distribution. Remember the definition of H_p in (3).

LEMMA 1 (Chernoff's bound). *For any positive integer* n, *any* $0 < p \le q \le 1$, *we have*

$$\mathbb{P}\big(\mathrm{Bin}(n, p) \ge qn\big) \le \exp\big(-nH_p(q)\big). \tag{31}$$

A consequence of Chernoff's bound is Bernstein's inequality for the binomial distribution.

LEMMA 2 (Bernstein's inequality). *For positive integer n, any $p \in (0,1)$ and any $x \geq 0$, we have*

$$\mathbb{P}\big[\mathrm{Bin}(n,p) \geq np + x\big] \leq \exp\left[-\frac{x^2}{2[np(1-p)+x/3]}\right].$$

We will need the following basic properties of the entropy function.

LEMMA 3. *For $p \in (0,1)$, $H_p(q)$ is convex and increasing in $q \in [p,1]$. Moreover,*

$$H_p(q) = \begin{cases} \dfrac{(q-p)^2}{2p(1-p)} + O\left(\dfrac{(q-p)^3}{p^2}\right), & \dfrac{q}{p} \to 1; \\ p(r\log r - r + 1), & \dfrac{q}{p} \to r \in (1,\infty),\ p \to 0; \\ q\log\left(\dfrac{q}{p}\right) + O(q), & \dfrac{q}{p} \to \infty. \end{cases} \tag{32}$$

We will also use the following upper bound on the binomial coefficients.

LEMMA 4. *For any integers $1 \leq k \leq n$,*

$$k\log(n/k) \leq \log\binom{n}{k} \leq k\log(ne/k), \tag{33}$$

where $e = \exp(1)$.

The next result bounds the hypergeometric distribution with the corresponding binomial distribution. Let $\mathrm{Hyp}(N,m,n)$ denotes the hypergeometric distribution counting the number of red balls in n draws from an urn containing m red balls out of N.

LEMMA 5. *For any $m \leq N/2$, $\mathrm{Hyp}(N,m,n)$ is stochastically smaller than $\mathrm{Bin}(n, \frac{m}{N-m})$.*

PROOF. Suppose the balls are picked one by one without replacement. At each stage, the probability of selecting a red ball is smaller than $m/(N-m)$. More formally, sample n i.i.d. uniform variable $Z_i \in (0,1)$. Then, on the one hand, $Y := \sum_{i=1}^n \mathbb{1}_{\{Z_i \leq m/(N-m)\}}$ follows a binomial distribution with parameters $(n, m/(N-m))$, while on the other hand $X := \sum_{i=1}^n X_i$, with

$$X_i := \mathbb{1}_{\{Z_i \leq (1/(N-i+1))(m-\sum_{j=1}^{i-1} X_i)\}},$$

follows a hypergeometric distribution with parameters N,m,n. And by construction $X \leq Y$. $\square$

7.2. *Proof of Theorem 1.* Following standard lines, we start by reducing the composite alternative to a simple alternative by considering the uniform prior π on subsets $S \subset [N] := \{1, \dots, N\}$ of size $|S| = n$. Indeed, for a test ϕ, recall its worst-case risk $\gamma_N(\phi)$ defined in (1), and define its average risk

$$\bar{\gamma}_N(\phi) = \mathbb{P}_0(\phi = 1) + \frac{1}{n} \sum_{|S|=n} \mathbb{P}_S(\phi = 0).$$

Then the average risk is bounded by the worst-case risk, meaning, $\bar{\gamma}_N(\phi) \le \gamma_N(\phi)$, this being valid for all ϕ. It suffices, therefore, to lower bound the average risk, and the advantage is that we know that the likelihood ratio test minimizes the average risk, and we can even compute its risk. The likelihood ratio is

$$L = \frac{\#\{S \subset [N] : |S| = n,\, W_S = n^{(2)}\}}{\binom{N}{n} p_0^{n(n-1)/2}}, \tag{34}$$

which is the observed number of cliques of size n divided by the expected number under the null. Then the test $\{L > 1\}$ minimizes the average risk [Lehmann and Romano (2005), Problem 3.10], with risk equal to

$$\gamma_L := \mathbb{P}_0(L > 1) + \mathbb{E}_0\big(L\{L \le 1\}\big).$$

Therefore, it suffices to show that $\gamma_L \to 1$. Here, we use arguably the simplest method, a second moment argument, which is based on the fact that

$$\gamma_L = 1 - \tfrac{1}{2}\,\mathbb{E}_0\,|L - 1| \ge 1 - \tfrac{1}{2}\sqrt{\mathrm{Var}_0(L)},$$

by the Cauchy–Schwarz inequality, so it is enough to prove that $\mathrm{Var}_0(L) \to 0$. We do so by showing that $\mathbb{E}_0(L^2) \le 1 + o(1)$.

Note that

$$L = p_0^{-n^{(2)}} \pi\big[W_S = n^{(2)}\big],$$

where $\pi[\cdot]$ denotes the expectation with respect to π. Hence, by Fubini–Tonelli's theorem, we have

$$\mathbb{E}_0\, L^2 = \pi^{\otimes 2}\big[p_0^{-2n^{(2)}}\, \mathbb{P}_0\big(W_{S_1} = W_{S_2} = n^{(2)}\big)\big] = \pi^{\otimes 2}\big[p_0^{-K(K-1)/2}\big],$$

where $K := |S_1 \cap S_2|$, and $\pi^{\otimes 2}$ is the joint distribution of $S_1, S_2 \overset{\text{i.i.d.}}{\sim} \pi$. Indeed, the event $\{W_{S_1} = W_{S_2} = n^{(2)}\}$ means that all edges between pairs of nodes in S_1 exist, and similarly for S_2, and there are a total of $n(n-1) + K(K-1)/2$ such edges.

Before going further, note that (4) and (33) imply that

$$\log(N/n) - \frac{(n-1)}{2} \log(1/p_0) \to \infty. \tag{35}$$

In particular, this means that $n \le 3 \log N$, eventually and, therefore,

$$\frac{n^2}{N} = O\big((\log N)^2/N\big) \to 0. \tag{36}$$

Since $K \sim \mathrm{Hyp}(N, n, n)$, by Lemma 5, K is stochastically bounded by $\mathrm{Bin}(n, \rho)$, where $\rho := n/(N-n)$. Hence, by Lemmas 1 and 5, we have

$$\begin{aligned}\mathbb{P}(K \geq k) &\leq \mathbb{P}\big(\mathrm{Hyp}(N, n, n) \geq k\big) \\ &\leq \mathbb{P}\big(\mathrm{Bin}(n, \rho) \geq k\big) \\ &\leq \exp\big(-nH_\rho(k/n)\big).\end{aligned} \tag{37}$$

Now, using Lemma 3 and (36), for $k \geq 2$ we get

$$nH_\rho(k/n) = k\log\big(k/(n\rho)\big) + O(k) = k\log\big(kN/n^2\big) + O(k).$$

Hence,

$$\begin{aligned}&\pi^{\otimes 2}\big[p_0^{-K(K-1)/2}\big] \\ &\quad= \mathbb{P}_0(K \leq 1) + \sum_{k=2}^{n} \exp\bigg(\frac{k(k-1)}{2}\log(1/p_0) - nH_\rho(k/n)\bigg) \\ &\quad\leq 1 + \sum_{k=2}^{n} \exp\bigg(k\bigg[\frac{(k-1)}{2}\log(1/p_0) - \log\big(kN/n^2\big) + O(1)\bigg]\bigg).\end{aligned} \tag{38}$$

For $a > 0$ fixed, the function $x \mapsto ax - \log x$ is decreasing on $(0, 1/a)$ and increasing on $(1/a, \infty)$. Therefore,

$$\frac{(k-1)}{2}\log(1/p_0) - \log\big(kN/n^2\big) \leq -\omega,$$

where

$$\omega := \min\bigg(\log\big(N/n^2\big) - \frac{1}{2}\log(1/p_0), \log(N/n) - \frac{n-1}{2}\log(1/p_0)\bigg).$$

By (35), the second term in the minimum tends to ∞. This also the case of the first term, since

$$\log\big(N/n^2\big) - \frac{1}{2}\log(1/p_0) = \log(N/n) - \frac{n-1}{2}\log(1/p_0) + \frac{n}{2}\log(1/p_0) - \log n,$$

with the second difference bounded from below. Hence, $\omega \to \infty$. Hence, the sum in (38) is bounded by

$$\sum_{k=2}^{n} \exp\big(-k\big[\omega + O(1)\big]\big) \leq \sum_{k=2}^{n} e^{-k\omega/2} = \frac{e^{-\omega}}{1 - e^{-\omega/2}} \to 0,$$

eventually.

Hence, we showed that $\mathbb{E}_0(L^2) \leq 1 + o(1)$ and the proof of Theorem 1 is complete.

7.3. *Proof of Theorem 2.* We assume that (2), (8) and (9) hold. We reduce the composite alternative to a simple alternative by considering the uniform prior π on subsets $S \subset [N] := \{1, \dots, N\}$ of size $|S| = n$. The resulting likelihood ratio is

$$L = L(\mathbf{W}) = \binom{N}{n}^{-1} \sum_{|S|=n} L_S = \pi[L_S], \tag{39}$$

where $\pi[\cdot]$ is the expectation with respect to $S \sim \pi$, and

$$L_S := \exp\bigl(\theta_{p_1} W_S - \Lambda(\theta_{p_1}) n^{(2)}\bigr), \tag{40}$$

with

$$\theta_q := \log\biggl(\frac{q(1-p_0)}{p_0(1-q)}\biggr) \tag{41}$$

and

$$\Lambda(\theta) := \log\bigl(1 - p_0 + p_0 e^{\theta}\bigr),$$

which is the moment generating function of $\mathrm{Bern}(p_0)$.

It is well known that H is the Fenchel–Legendre transform of Λ; more specifically, for $q \in (p_0, 1)$,

$$H_{p_0}(q) = \sup_{\theta \geq 0} \bigl[q\theta - \Lambda(\theta)\bigr] = q\theta_q - \Lambda(\theta_q). \tag{42}$$

The second moment argument used in Section 7.2 is also applicable here, though it does not yield sharp bounds. In Case 1 below [see (44)], which is the regime where the moderate deviations of the binomial come into play, this method leads to a requirement that the limit superior in (9) be bounded by $1/2$ instead of 1. And, worse than that, in Case 3 below, which is the regime where the large deviations of the binomial are involved, it does not provide any useful bound whatsoever.

Fortunately, a finer approach was suggested by Ingster (1997). The refinement is based on bounding the first and second moments of a truncated likelihood ratio. Here, we follow Butucea and Ingster (2011). They work with the following truncated likelihood:

$$\tilde{L} = \binom{N}{n}^{-1} \sum_{|S|=n} \mathbb{1}_{\Gamma_S} L_S,$$

where the events Γ_S will be specified below (50). We note $\Gamma = \bigcap_{|S|=n} \Gamma_S$. Using the triangle inequality, the fact that $\tilde{L} \leq L$ and the Cauchy–Schwarz inequality, we have the following upper bound:

$$\begin{aligned} \mathbb{E}_0 |L - 1| &\leq \mathbb{E}_0 |\tilde{L} - 1| + \mathbb{E}_0 (L - \tilde{L}) \\ &\leq \sqrt{\mathbb{E}_0[\tilde{L}^2] - 1 + 2\bigl(1 - \mathbb{E}_0[\tilde{L}]\bigr)} + \bigl(1 - E_0[\tilde{L}]\bigr), \end{aligned}$$

so that $\gamma_L \to 1$ when $\mathbb{E}_0[\tilde{L}^2] \to 1$ and $\mathbb{E}_0[\tilde{L}] \to 1$. Note that contrary to what Butucea and Ingster (2011) do, we do not require that $\mathbb{P}_0(\Gamma) \to 1$. More precisely, we shall prove that $(1, 1)$ is an accumulation point of any subsequence of $(\mathbb{E}_0 \tilde{L}, \mathbb{E}_0[\tilde{L}^2])$. Adopting this approach allows us to assume that p_1/p_0 converges to $r \in [1, \infty]$, p_1^2/p_0 converges to $r_2 \in [0, \infty]$ and that

$$\frac{nH_{p_0}(p_1)}{2\log(N/n)} < 1 - \eta_0, \tag{43}$$

for some $\eta_0 \in (0, 1)$ fixed. Notice that (6) and (9) imply that $H_{p_0}(p_1) \to 0$, which by Lemma 3 forces either $p_1/p_0 \to 1$ or $p_1 \to 0$; in any case, p_1 is bounded away from 1 this time.

In what follows, we provide the general arguments while the proof of the technical results (Lemmas 6–8) is postponed to the supplementary material Arias-Castro and Verzelen (2014). To show these technical results, we divide the analysis depending on the behavior of p_1/p_0

$$\frac{p_1}{p_0} \to \begin{cases} r = 1, & (44) \\ r \in (1, \infty), & (45) \\ r = \infty. & (46) \end{cases}$$

In regime (44), the moderate deviations of the binomial distribution dominate and these are asymptotically equivalent to normal (Gaussian) deviations; in particular, it is in this setting (with p_0 constant) that Butucea and Ingster (2011) successfully reduce the binary setting to the normal setting. In regime (46), the large deviations of the binomial distribution dominate, which do not resemble the normal deviations and lead to a completely different regime. Regime (45) is intermediary and requires a special treatment.

Define the numbers

$$k_* = \left\lceil 1 + 2\frac{\log(N/n)}{\log(1 + (p_1 - p_0)^2/(p_0(1 - p_0)))} \right\rceil \wedge n, \tag{47}$$

$$k_{\min} = \left\lceil 1 + 2\frac{\log(Nk_*/n^2) - \log\{\log(n/\log(N/n)) \wedge \log(N/n)\}}{\log(1 + (p_1 - p_0)^2/(p_0(1 - p_0)))} \right\rceil \wedge n. \tag{48}$$

The exact expression of $k_{\min}$ will be useful for bounding the second moment of $\tilde{L}$. For the time being, we only need to have in mind the properties summarized in the following lemma.

LEMMA 6. *We have* $k_{\min} \sim k_* \to \infty$ *and* $\log(n/k_{\min}) = o[\log(N/n)]$.

We next define a sequence of number (q_k) via the following result.

LEMMA 7. *For any integer k between $k_{\min}+1$ and n, there exists a unique $q_k \in (p_0, 1)$ such that*

$$\frac{(k-1)}{2} H_{p_0}(q_k) = \log(N/k) + 2. \tag{49}$$

Moreover, q_k satisfies $\theta_{q_k} \le 2\theta_{p_1}$.

Let $w_k = q_k k^{(2)}$ and define

$$\Gamma_S := \{W_T \le w_{|T|}, \forall T \subset S \text{ such that } |T| \ge \lfloor k_{\min} \rfloor + 1\}. \tag{50}$$

7.3.1. *First truncated moment.* We first prove that $\mathbb{E}_0 \tilde{L} \to 1$. By Fubini's theorem, we have

$$\mathbb{E}_0 \tilde{L} = \pi[\mathbb{E}_0[L_S \mathbb{1}_{\Gamma_S}]] = \pi[\mathbb{P}_S(\Gamma_S)] = \mathbb{P}_S(\Gamma_S),$$

where S is any fixed subset of size n in $\{1, \dots, N\}$ and this last inequality is by the fact that $\mathbb{P}_S(\Gamma_S)$ does not depend on S by symmetry. By the union bound, Chernoff's bound (31) and (33),

$$\begin{aligned} 1 - \mathbb{P}_S(\Gamma_S) &\le \sum_{k=\lfloor k_{\min} \rfloor + 1}^{n} \sum_{T \subset S, |T|=k} \mathbb{P}_S\big(W_T > q_k k^{(2)}\big) \\ &\le \sum_{k=\lfloor k_{\min} \rfloor + 1}^{n} \binom{n}{k} \mathbb{P}\big(\mathrm{Bin}\big(k^{(2)}, p_1\big) > q_k k^{(2)}\big) \\ &\le \sum_{k=\lfloor k_{\min} \rfloor + 1}^{n} \exp\left[k\left(\log(ne/k) - \frac{(k-1)}{2} H_{p_1}(q_k)\right)\right]. \end{aligned}$$

We then conclude that $1 - \mathbb{P}_S(\Gamma_S) = o(1)$ using the following result.

LEMMA 8. *We have*

$$\min_{k=\lfloor k_{\min} \rfloor + 1, \dots, n} \left(\frac{k-1}{2} H_{p_1}(q_k) - \log\left(\frac{n}{k}\right)\right) \to \infty. \tag{51}$$

7.3.2. *Second truncated moment.* We now prove that $\mathbb{E}_0 \tilde{L}^2 \le 1 + o(1)$, which with $\mathbb{E}_0 \tilde{L} \to 1$ shows that $\mathrm{Var}_0(\tilde{L}) \to 0$. Let $S_1, S_2 \overset{\text{i.i.d.}}{\sim} \pi$ and define $K = |S_1 \cap S_2|$. By Fubini's theorem, we have

$$\begin{aligned} \mathbb{E}_0 \tilde{L}^2 &= \pi^{\otimes 2}[\mathbb{E}_0(L_{S_1} L_{S_2} \mathbb{1}_{\Gamma_{S_1}} \mathbb{1}_{\Gamma_{S_2}})] \\ &= \pi^{\otimes 2}[\mathbb{E}_0(\exp\{\theta_{p_1}(W_{S_1} + W_{S_2}) - 2\Lambda(\theta_{p_1}) n^{(2)}\} \mathbb{1}_{\Gamma_{S_1} \cap \Gamma_{S_2}})]. \end{aligned}$$

Define

$$W_{S \times T} = \frac{1}{2} \sum_{i \in S, j \in T} W_{i,j},$$

and note that $W_S = W_{S\times S}$. We use the decomposition

$$W_{S_1} + W_{S_2} = W_{S_1\times(S_1\setminus S_2)} + W_{S_2\times(S_2\setminus S_1)} + 2W_{S_1\cap S_2}, \tag{52}$$

the fact that

$$\Gamma_{S_1}\cap\Gamma_{S_2}\subset\{W_{S_1\cap S_2}\le w_K\},$$

and the independence of the random variables on the RHS of (52), to get

$$\mathbb{E}_0\big(\exp\{\theta_{p_1}(W_{S_1}+W_{S_2})-2\Lambda(\theta_{p_1})n^{(2)}\}\mathbb{1}_{\Gamma_{S_1}\cap\Gamma_{S_2}}\big)\le \mathrm{I}\cdot\mathrm{II}\cdot\mathrm{III},$$

where

$$\mathrm{I} := \mathbb{E}_0\exp\left(\theta_{p_1}W_{S_1\times(S_1\setminus S_2)} - \frac{\Lambda(\theta_{p_1})}{2}(n-K)(n+K-1)\right) = 1,$$

$$\mathrm{II} := \mathbb{E}_0\exp\left(\theta_{p_1}W_{S_2\times(S_2\setminus S_1)} - \frac{\Lambda(\theta_{p_1})}{2}(n-K)(n+K-1)\right) = 1,$$

$$\mathrm{III} := \mathbb{E}_0\big(\exp\big(2\theta_{p_1}W_{S_1\cap S_2} - 2\Lambda(\theta_{p_1})K^{(2)}\big)\mathbb{1}_{\{W_{S_1\cap S_2}\le w_K\}}\big).$$

The first two equalities are due to the fact that the likelihood integrates to one.

To bound III, we follow Butucea and Ingster (2011), with a twist. When $K\le k_{\min}$, we will use the obvious bound

$$\mathrm{III}\le \mathbb{E}_0\exp\big(2\theta_{p_1}W_{S_1\cap S_2} - 2\Lambda(\theta_{p_1})K^{(2)}\big) = \exp\big(\Delta K^{(2)}\big),$$

where

$$\Delta := \Lambda(2\theta_{p_1}) - 2\Lambda(\theta_{p_1}) = \log\left(1+\frac{(p_1-p_0)^2}{p_0(1-p_0)}\right). \tag{53}$$

When $K > k_{\min}$, we use a different bound. For any $\xi\in(0,2\theta_{p_1})$, we have

$$\begin{aligned}\mathrm{III} &\le \mathbb{E}_0\big[\exp\big(\xi W_{S_1\cap S_2} + (2\theta_{p_1}-\xi)w_K - 2\Lambda(\theta_{p_1})K^{(2)}\big)\{W_{S_1\cap S_2}\le w_K\}\big]\\ &\le \mathbb{E}_0\exp\big[\xi W_{S_1\cap S_2} + (2\theta_{p_1}-\xi)w_K - 2\Lambda(\theta_{p_1})K^{(2)}\big],\end{aligned}$$

so that

$$\mathrm{III}\le\exp\big(\Delta_K K^{(2)}\big),$$

where

$$\Delta_k := \min_{\xi\in[0,2\theta_{p_1}]}\Lambda(\xi) + (2\theta_{p_1}-\xi)q_k - 2\Lambda(\theta_{p_1}). \tag{54}$$

By the variational definition of the entropy (42), the minimum of $\Lambda(\xi)+(2\theta_{p_1}-\xi)q_k - 2\Lambda(\theta_{p_1})$ over ξ in $\mathbb{R}^+$ is achieved at $\xi=\theta_{q_k}$, and we know from Lemma 7 that $\theta_{q_k}\le 2\theta_{p_1}$. Hence, we have

$$\begin{aligned}\Delta_k &= -H_{p_0}(q_k) + 2\theta_{p_1}q_k - 2\Lambda(\theta_{p_1})\\ &= -2H_{p_1}(q_k) + H_{p_0}(q_k).\end{aligned} \tag{55}$$

Following our tracks, we have

$$\mathbb{E}_0\,\tilde{L}^2 \le \mathbb{E}\big[\mathbb{1}_{\{K\le k_{\min}\}}\exp\big(\Delta K^{(2)}\big)\big] + \mathbb{E}\big[\mathbb{1}_{\{K>k_{\min}\}}\exp\big(\Delta_K K^{(2)}\big)\big],$$

where the expectation is with respect to $\pi^{\otimes 2}$.

Let b be an integer sequence such that $b\to\infty$ so slowly that

$$\frac{(p_1-p_0)}{\sqrt{p_0}}\frac{bn^2}{N}\to 0, \tag{56}$$

which is possible because of (8). Recall that $\rho = n/(N-n)$ and define $k_0 = \lceil bn\rho\rceil$. We divide the expectation into two parts: $K\le k_0$ and $k_0+1\le K\le n$. When $k_0=1$, we simply have

$$\mathbb{E}\big[\mathbb{1}_{\{K\le k_0\}}\exp\big(\Delta K^{(2)}\big)\big] = \mathbb{P}(K\le 1)\le 1.$$

When $k_0\ge 2$, we use the expression (53) of Δ to derive

$$\begin{aligned}\mathbb{E}\big[\mathbb{1}_{\{K\le k_0\}}\exp\big(\Delta K^{(2)}\big)\big] &\le \exp\big[\Delta k_0^2\big]\\ &\le \exp\bigg[O(1)\frac{(p_1-p_0)^2}{p_0(1-p_0)}\frac{b^2n^2}{N^2}\bigg] = 1+o(1),\end{aligned}$$

because of (56).

When $k_0+1\le K\le \lfloor k_{\min}\rfloor$, we use the bound (37) and the identity $(1-x)\log(1-x)\ge -x$, to get

$$\begin{aligned}&\mathbb{E}\big[\mathbb{1}_{\{k_0+1\le K\le\lfloor k_{\min}\rfloor\}}\exp\big(\Delta K^{(2)}\big)\big]\\ &\quad\le \sum_{k=k_0+1}^{\lfloor k_{\min}\rfloor}\exp\bigg[\Delta\frac{k(k-1)}{2} - nH_\rho\Big(\frac{k}{n}\Big)\bigg]\\ &\quad\le \sum_{k=k_0+1}^{\lfloor k_{\min}\rfloor}\exp\bigg[k\Big(\Delta\frac{k-1}{2} - \log\Big(\frac{k}{n\rho}\Big)+1\Big)\bigg].\end{aligned}$$

For $a>0$ fixed, the function $f(x)=ax-\log x$ is decreasing on $(0,1/a)$ and increasing on $(1/a,\infty)$. Therefore, for $k_0+1\le k\le n$,

$$\Delta\frac{k-1}{2} - \log\Big(\frac{k}{n\rho}\Big)\le -\omega,$$

where

$$\omega := \min\bigg[\log b - \Delta\frac{k_0-1}{2}, \log\Big(\frac{k_{\min}}{n\rho}\Big) - \Delta\frac{k_{\min}-1}{2}\bigg].$$

From what we did previously, we know that $\Delta(k_0-1)=o(1)$, so that the first term in the maximum tends to ∞. Therefore, it suffices to look at the second term in

the maximum. In fact, $k_{\min}$ has been precisely defined in (48) to make this second term diverge. Indeed, by (48) and (53), we have

$$\Delta\frac{k_{\min}-1}{2} \le \log\biggl(\frac{Nk^*}{n^2}\biggr) - \log\log\biggl[\frac{n}{\log(N/n)}\biggr].$$

By Lemma 6 and since $\rho \asymp n/N = o(1)$, we get $\log(k_{\min}/(n\rho)) - \log(\frac{Nk^*}{n^2}) = o(1)$. Consequently,

$$\log\biggl(\frac{k_{\min}}{n\rho}\biggr) - \Delta\frac{k_{\min}-1}{2} \ge \log\log\biggl[\frac{n}{\log(N/n)}\biggr] + o(1) \to \infty,$$

because of (6).

When $K > k_{\min}$, we have

$$\begin{aligned}&\mathbb{E}\bigl[\mathbb{1}_{\{K>k_{\min}\}}\exp\bigl(\Delta_K K^{(2)}\bigr)\bigr]\\&\quad\le \sum_{k=\lfloor k_{\min}\rfloor+1}^{n} \exp\biggl[k\biggl(\Delta_k\frac{k-1}{2} - \log\biggl(\frac{k}{n\rho}\biggr)+1\biggr)\biggr].\end{aligned}$$

Now, using (55), we have

$$\begin{aligned}&\Delta_k\frac{k-1}{2} - \log\biggl(\frac{k}{n\rho}\biggr)\\&\quad= \frac{k-1}{2}\bigl[-2H_{p_1}(q_k) + H_{p_0}(q_k)\bigr] - \log\biggl(\frac{N}{k}\biggr) + 2\log\biggl(\frac{n}{k}\biggr) + o(1),\end{aligned}$$

which goes to $-\infty$ uniformly over all k between $\lfloor k_{\min}\rfloor + 1$ and n by the definition (49) of q_k and by the control of $H_{p_1}(q_k)$ from Lemma 8. Hence, the sum above tends to zero.

This concludes the proof that $\mathbb{E}_0\,\tilde{L}^2 \le 1 + o(1)$.

7.4. *Proof of Proposition 2.* We start with a useful result for proving that a test is asymptotically powerful based on the first two moments of the corresponding test statistic.

LEMMA 9. *Suppose that for testing H_0 versus H_1, a statistic T satisfies*

$$R_T := \frac{\mathbb{E}_1(T) - \mathbb{E}_0(T)}{\max(\sqrt{\mathrm{Var}_1(T)}, \sqrt{\mathrm{Var}_0(T)})} \to \infty. \tag{57}$$

Then there is a test based on T that is asymptotically powerful.

PROOF. Consider the test that rejects when $T \ge \mathbb{E}_0(T) + \sqrt{R_T\,\mathrm{Var}_0(T)}$. By Chebyshev's inequality, the probability of type I error tends to zero:

$$\mathbb{P}_0\bigl(T \ge \mathbb{E}_0(T) + \sqrt{R_T\,\mathrm{Var}_0(T)}\bigr) \le \frac{1}{R_T} \to 0.$$

For the probability of type II error, we have

$$\mathbb{P}_1\big(T \geq \mathbb{E}_0(T) + \sqrt{R_T \operatorname{Var}_0(T)}\big) = \mathbb{P}_1\left(\frac{T - \mathbb{E}_1(T)}{\sqrt{\operatorname{Var}_1(T)}} \geq -\xi\right) \geq 1 - \frac{1}{\xi^2},$$

where

$$\xi := \frac{R_T \max(\sqrt{\operatorname{Var}_1(T)}, \sqrt{\operatorname{Var}_0(T)}) - \sqrt{R_T \operatorname{Var}_0(T)}}{\sqrt{\operatorname{Var}_1(T)}} \to \infty. \qquad \square$$

We now apply Lemma 9 to the total degree test. Under the null,

$$\mathbb{E}_0(W) = \frac{N(N-1)}{2} p_0, \qquad \operatorname{Var}_0(W) = \frac{N(N-1)}{2} p_0(1-p_0),$$

while under the alternative,

$$\mathbb{E}_1(W) = \frac{N(N-1)}{2} p_0 + \frac{n(n-1)}{2}(p_1 - p_0)$$

and

$$\operatorname{Var}_1(W) = \frac{N(N-1)}{2} p_0(1-p_0) + \frac{n(n-1)}{2}\big[p_1(1-p_1) - p_0(1-p_0)\big].$$

In any case,

$$\max\big(\operatorname{Var}_1(W), \operatorname{Var}_0(W)\big) \leq \tfrac{1}{2} N^2 p_0 + \tfrac{1}{2} n^2 (p_1 - p_0).$$

Recalling the definition of R_W in (57), under (13) we have

$$R_W \geq \frac{n(n-1)(p_1 - p_0)}{\sqrt{N^2 p_0 + n^2(p_1 - p_0)}} \succ \frac{n^2}{N} \frac{p_1 - p_0}{\sqrt{p_0}} \to \infty.$$

Therefore, the total degree test is asymptotically powerful when (13) holds.

7.5. *Proof of Proposition 3.* We use the union bound, Chernoff's bound (31) and (33) to get

$$\begin{aligned}\mathbb{P}_0\big(W_n^* \geq a n^{(2)}\big) &\leq \binom{N}{n} \exp\big(-n^{(2)} H_{p_0}(a)\big) \\ &\leq \exp\big(n \log(Ne/n) - n^{(2)} H_{p_0}(a)\big),\end{aligned}$$

which goes to zero when

$$\log(N/n) - \frac{(n-1)}{2} H_{p_0}(a) \to -\infty. \tag{58}$$

Choose $a = \eta p_0 + (1-\eta) p_1$ with $\eta = \log^{-1}(N/n) \in (0,1)$ fixed, sufficiently small that

$$\liminf \frac{n H_{p_0}(a)}{2 \log(N/n)} > 1.$$

This is possible because of how H varies, which is described in Lemma 3.

We then consider the test that rejects when $W_n^* \geq an^{(2)}$. We just chose a so that its level tends to zero. Under the alternative, let S denote the community. By definition, $W_n^* \geq W_S$, and since $W_S \sim \mathrm{Bin}(n^{(2)}, p_1)$ and $p_1 n^{(2)} \to \infty$, $W_S = p_1 n^{(2)} + O_P(\sqrt{p_1 n^{(2)}})$. Therefore, the test is asymptotically powerful when $n^2(p_1 - a) \gg \sqrt{p_1 n^{(2)}}$. Since $p_1 - a = \eta(p_1 - p_0)$ and $\eta > 0$ is constant, this is the same as $(p_1 - p_0)n^2 \gg \sqrt{p_1 n^2}$. Now, if p_1/p_0 is bounded away from 1, this is true because $p_1 - p_0 \asymp p_1$ and $p_1 n^2 \to \infty$; while if $p_1/p_0 \to 1$, we use Lemma 3 and (15) to get that $(p_1 - p_0)^2 n/p_0 \geq \mathrm{cst} \log(N/n)$, implying that $(p_1 - p_0)n^2/\sqrt{p_1 n^2} \sim (p_1 - p_0)n/\sqrt{p_0} \to \infty$.

7.6. *Proof of Proposition 4.* The proof is a simple refinement of that of Proposition 3. Let $a_n = H_{p_0}^{-1}[\frac{2[\log(N/n)+2]}{n-1}] = w_n/n^{(2)}$. Then

$$\begin{aligned}\mathbb{P}_0\Big(\max_{n \geq u_N} \frac{W_n^*}{w_n} \geq 1\Big) &= \sum_{n=u_N}^{N} \mathbb{P}_0\big(W_n^* \geq a_n n^{(2)}\big) \\ &\leq \sum_{n=u_N}^{\infty} \exp\big(n \log(Ne/n) - n^{(2)} H_{p_0}(a_n)\big) \\ &= \sum_{n=u_N}^{\infty} e^{-n} = o(1).\end{aligned}$$

Now, assume we are under the alternative, with a community S of size n. As in the proof of Proposition 3, let $\eta \in (0, 1)$ be fixed, but large enough that $a := \eta p_0 + (1 - \eta)p_1$ satisfies $\liminf \frac{nH_{p_0}(a)}{2\log(N/n)} > 1$. Since $\frac{nH_{p_0}(a_n)}{2\log(N/n)} \to 1$, necessarily $a_n \leq a$ for n large enough, in which case $p_1 - a_n \geq p_1 - a \gg \sqrt{p_1 n^{(2)}}$, as shown earlier, and we conclude in the same way.

Acknowledgements. We would like to thank Jacques Verstraete for helpful discussions on the clique number of a random graph, and anonymous referees for constructive feedback that helped improve the presentation of the paper. We first learned about the work of Butucea and Ingster (2011) at the *New Trends in Mathematical Statistics* conference held at the Centre International de Rencontres Mathématiques (CIRM), Luminy, France, in 2011.

SUPPLEMENTARY MATERIAL

Technical appendix (DOI: 10.1214/14-AOS1208SUPP; .pdf). This supplementary material contains the remaining proofs and some technical details.

REFERENCES

ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys*. **74** 47–97. MR1895096

ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. In *Proceedings of the Eighth International Conference "Random Structures and Algorithms"* (*Poznan*, 1997). *Random Structures Algorithms* **13** 457–466. MR1662795

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist*. **39** 2533–2556. MR2906877

ARIAS-CASTRO, E. and VERZELEN, N. (2013). Community detection in sparse random networks. Available at http://arxiv.org/abs/1308.2955.

ARIAS-CASTRO, E. and VERZELEN, N. (2014). Supplement to "Community detection in dense random networks." DOI:10.1214/14-AOS1208SUPP.

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

BERTHET, Q. and RIGOLLET, P. (2012). Optimal detection of sparse principal components in high dimension. Available at http://arXiv.org/abs/1202.5070.

BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Givan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist*. **39** 2280–2301. MR2906868

BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966

BUTUCEA, C. and INGSTER, Y. I. (2011). Detection of a sparse submatrix of a high-dimensional noisy matrix. Available at http://arxiv.org/abs/1109.0898.

CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **73** 629–662. MR2867452

D'ASPREMONT, A., EL GHAOUI, L., JORDAN, M. I. and LANCKRIET, G. R. G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Rev.* **49** 434–448 (electronic). MR2353806

DEKEL, Y., GUREL-GUREVICH, O. and PERES, Y. (2011). Finding hidden cliques in linear time with high probability. In *ANALCO*11—*Workshop on Analytic Algorithmics and Combinatorics* 67–75. SIAM, Philadelphia, PA. MR2815485

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist*. **32** 962–994. MR2065195

FEIGE, U., KORTSARZ, G. and PELEG, D. (2001). The dense k-subgraph problem. *Algorithmica* **29** 410–421. MR1799268

FEIGE, U. and RON, D. (2010). Finding hidden cliques in linear time. In 21*st International Meeting on Probabilistic*, *Combinatorial*, *and Asymptotic Methods in the Analysis of Algorithms* (*AofA'*10). 189–203. Assoc. Discrete Math. Theor. Comput. Sci., Nancy. MR2735341

FELDMAN, V., GRIGORESCU, E., REYZIN, L., VEMPALA, S. and XIAO, Y. (2012). Statistical algorithms and a lower bound for planted clique. Available at http://arXiv.org/abs/1201.1214.

FORTUNATO, S. (2010). Community detection in graphs. *Phys. Rep*. **486** 75–174. MR2580414

GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826 (electronic). MR1908073

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist*. **38** 1686–1732. MR2662357

HEARD, N. A., WESTON, D. J., PLATANIOTI, K. and HAND, D. J. (2010). Bayesian anomaly detection methods for social networks. *Ann. Appl. Stat*. **4** 645–662. MR2758643

INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist*. **6** 47–69. MR1456646

INGSTER, Y. I. and SUSLINA, I. A. (2002). On the detection of a signal with a known shape in a multichannel system. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **294** 88–112, 261. MR1976749

INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131

KARP, R. M. (1972). Reducibility among combinatorial problems. In *Complexity of Computer Computations* (*Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, NY*, 1972) 85–103. Plenum, New York. MR0378476

KHULLER, S. and SAHA, B. (2009). On finding dense subgraphs. In *Automata, Languages and Programming. Part I. Lecture Notes in Computer Science* **5555** 597–608. Springer, Berlin. MR2544878

KULLDORFF, M. (1997). A spatial scan statistic. *Comm. Statist. Theory Methods* **26** 1481–1496. MR1456844

LANCICHINETTI, A. and FORTUNATO, S. (2009). Community detection algorithms: A comparative analysis. *Phys. Rev. E* (3) **80** 056117.

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York. MR2135927

MONGIOVI, M., BOGDANOV, P., RANCA, R., PAPALEXAKIS, E. E., FALOUTSOS, C. and SINGH, A. K. (2013). NetSpot: Spotting significant anomalous regions on dynamic networks. In *SIAM International Conference on Data Mining*. Austin, TX.

NEWMAN, M. E. J. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103** 8577–8582.

NEWMAN, M. E. J. and GIRVAN, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* (3) **69** 026113.

PARK, Y., PRIEBE, C. E. and YOUSSEF, A. (2013). Anomaly detection in time series of graphs using fusion of graph invariants. *IEEE Journal of Selected Topics in Signal Processing* **7** 67–75.

REICHARDT, J. and BORNHOLDT, S. (2006). Statistical mechanics of community detection. *Phys. Rev. E* (3) **74** 016110, 14. MR2276596

ROSSMAN, B. (2010). Average-case complexity of detecting cliques. Ph.D. thesis, MIT, Cambridge, MA.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359

ZUCKERMAN, D. (2006). Linear degree extractors and the inapproximability of max clique and chromatic number. In *STOC'*06: *Proceedings of the* 38*th Annual ACM Symposium on Theory of Computing* 681–690. ACM, New York. MR2277193

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093-0112
USA
E-MAIL: eariasca@ucsd.edu

INRA
UMR 729 MISTEA
2 PLACE VIALA, BÂT. 29
F-34060 MONTPELLIER
FRANCE
E-MAIL: nicolas.verzelen@supagro.inra.fr