



**HAL**  
open science

# A degree-based goodness-of-fit test for heterogeneous random graph models

Sarah Ouadah, Stéphane Robin, Pierre Latouche

► **To cite this version:**

Sarah Ouadah, Stéphane Robin, Pierre Latouche. A degree-based goodness-of-fit test for heterogeneous random graph models. 2015. hal-01187889

**HAL Id: hal-01187889**

**<https://hal.science/hal-01187889>**

Preprint submitted on 28 Aug 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A degree-based goodness-of-fit test for heterogeneous random graph models

Sarah Ouadah<sup>\*†</sup>, Stéphane Robin<sup>\*†</sup>, Pierre Latouche<sup>‡</sup>

July 30, 2015

## Abstract

The degree variance has been proposed for many years to study the topology of a network. It can be used to assess the goodness-of-fit of the Erdős-Renyi model. In this paper, we prove the asymptotic normality of the degree variance under this model which enables us to derive a formal test. We generalize this result to the heterogeneous Erdős-Renyi model in which the edges have different respective probabilities to exist. For both models we study the power of the proposed goodness-of-fit test. We also prove the asymptotic normality under specific sparsity regimes. Both tests are illustrated on real networks from social sciences and ecology. Their performances are assessed via a simulation study.

## 1 Introduction

Interaction networks are used in many fields such as biology, sociology, ecology, economics or energy to describe the interactions existing between a set of individuals or entities. Formally, an interaction network can be viewed as a graph, the nodes of which being the individuals, and an edge between two nodes being present if these two individuals interact. Characterizing the general organization of such a network, namely its topology, can help in understanding the behavior of the system as a whole.

In the last decades, the distribution of the degrees (i.e. the number of connections of each node) has appeared as a simple and relevant way to study the topology of a network (Snijders (1981), Barabási and Albert (1999)). The degree distribution can also be used to infer complex graph model (Bickel *et al.* (2011)). From a more descriptive view-point, a very imbalanced distribution may reveal a network whose edges highly concentrate around few nodes, whereas a multi-modal distribution may reveal the existence of clusters

---

<sup>\*</sup>AgroParisTech, UMR 518, MIA, Paris, France

<sup>†</sup>INRA, UMR 518, MIA, Paris, France

<sup>‡</sup>Laboratoire SAMM, EA 4543, Université Paris 1 Panthéon-Sorbonne, France

of nodes (Channarond *et al.* (2012)). However, in practice, assessing the significance of such patterns remains an open problem.

The variance of the degrees has been considered since the earliest statistical studies of networks (Snijders (1981)). The first idea was simply to compare its empirical value to the expected one under a null random graph model, typically the Erdős-Renyi (*ER*) model, where each degree has a binomial distribution. Hagberg (2003) derives the exact moments of the degree variance and suggests to use a Gamma distribution (Hagberg (2000)). Snijders (1981) also gives the first two moments of the degree variance, but conditionally to the total number of edges. To our knowledge the first and only proof of the asymptotic normality of the degree variance under the *ER* model is given in a technical report from Bloznelis (2005).

In this paper, we derive the asymptotic distribution of the degree variance under the *ER* model and we generalize this result to the heterogeneous Erdős-Renyi model (*HER*), in which the edges are still independent but have different respective probabilities to exist. This generalization enables us to study the power of the degree variance test under a range of alternative random graph models.

Because the *ER* model is rarely a reasonable model to be tested, we define a generalized version of the degree variance statistic, which we name the degree sum of squares. This statistic generalizes the degree variance in the sense that it measures the discrepancy between the observed degrees and their expected values under an *HER* model. We also prove the asymptotic normality of this statistic and study its power under the alternative of an ill-specified *HER* model. In addition, because large networks are often sparse, we study under which sparsity regime the asymptotic distributions derived before still hold. The goodness-of-fit of the *HER* model has received little attention until now. The only reference we found is the procedure proposed by Cerqueira *et al.* (2015), who do not rely on the degrees and consider a set of independent and identically distributed random graphs.

In the rest of the paper, we consider an undirected graph  $\mathcal{G} = (\{1, \dots, n\}, \mathcal{E})$  with no self loop, that is the connection of a node to itself, and denote  $Y$  the corresponding adjacency  $n \times n$  matrix. Thus, the entry  $Y_{ij}$  of  $Y$  is 1 if  $(i, j) \in \mathcal{E}$ , and 0 otherwise. Because  $\mathcal{G}$  is undirected with no self loop, we have  $Y_{ij} = Y_{ji}, \forall i \neq j$  and  $Y_{ii} = 0$ , for all  $i$ 's. We further denote  $D_i$  the degree of node  $i$ :  $D_i = \sum_{j \neq i} Y_{ij}$ .

In terms of random graph models,  $ER(p)$  refers to the Erdős-Renyi model, according to which all edges ( $Y_{ij}$ ) are independent Bernoulli variables with same probability  $p$  to exist.  $HER(\mathbf{p})$  stands for the heterogeneous Erdős-Renyi model where edges are independent with respective probability  $p_{ij}$  to exist. The  $n \times n$  matrix  $\mathbf{p}$  has entries  $p_{ij}$ , it is symmetric with null diagonal.

The asymptotic framework for the *HER* model in the non-sparse setting is the following. We consider an infinite matrix  $\mathbf{P}$ , the elements of which are all bounded away from both 0 and 1. We then build a sequence of matrices  $\mathbf{p}^n$  made of the first  $n$  rows and columns of  $\mathbf{P}$ . Finally, we consider a sequence of independent graphs  $\mathcal{G}^n = (\{1, \dots, n\}, \mathcal{E}^n)$ , with increasing size  $n$  and respective probability matrices  $\mathbf{p}^n$ . The sequence of matrices

$\mathbf{p}^{*,n} = [p_{ij}^{*,n}]$  used in the sparse setting is constructed in a related way, based on an infinite matrix  $\mathbf{P}^*$  with all terms bounded away from 0 and 1. All quantities computed on  $\mathcal{G}^n$  should therefore be indexed by  $n$  as well. For the sake of clarity, we will drop the index  $n$  in the rest of the paper. The asymptotic framework for  $ER$  derives from this of  $HER$  described above.

In Section 2, we derive the asymptotic distribution of the degree variance under models  $ER(p)$  and  $HER(\mathbf{p})$ . We also derive its asymptotic normality under some specific sparsity regimes. We deduce a test for the null hypothesis stating that  $\mathcal{G}$  arises from  $ER(p)$  and study its power. In Section 3, we obtain a series of similar results for the mean square degree statistic under the  $HER(\bar{p})$  model with  $\bar{p} = [n(n-1)]^{-1} \sum_{i \neq j} p_{ij}$ . Both tests are illustrated on some examples. In Section 4, we study the performances of the proposed tests in a simulation study.

## 2 Degree variance test

For a given random graph, we consider the following statistic which is the empirical degree variance.

$$V = \frac{1}{n} \sum_i (D_i - \bar{D})^2 = \frac{1}{2n^2} \sum_{i \neq j} (D_i - D_j)^2,$$

where  $D_i = \sum_{i \neq j} Y_{ij}$  and  $\bar{D} = (1/n) \sum_j D_j$ .

### 2.1 Asymptotic normality

We establish the asymptotic normality of  $V$  under model  $HER(\mathbf{p})$ , then under model  $ER(p)$  as a direct consequence. The proof relies on the Hoeffding decomposition of  $V$ , in the same manner as in Bloznelis (2005) whose work is under model  $ER(p)$ . A similar strategy has been used by Nowicki and Wierman (1988) to prove the asymptotic normality of subgraph counts in random graphs. We derive all projections involved in the Hoeffding decomposition to which we eventually apply the Lindeberg-Lévy Theorem which is stated below (see e.g. Billingsley (1968), Theorem 7.2, p.42).

**Lindeberg-Lévy Theorem** *Let  $(X_{nu})_{1 \leq u \leq k_n}$  be a triangular array of independent random variables with means 0 and finite variances  $(\sigma_{nu}^2)_{1 \leq u \leq k_n}$ . Let  $B_n^2 = \sum_{u=1}^{k_n} \sigma_{nu}^2$ . If the Lindeberg condition*

$$A_n^2(\epsilon)/B_n^2 \rightarrow 0, \quad \text{as } n \rightarrow \infty, \quad \text{for each } \epsilon > 0, \quad \text{where } A_n^2(\epsilon) = \sum_{u=1}^{k_n} \int_{\{|x_{nu}| > \epsilon B_n\}} x_{nu}^2 dP \quad (1)$$

*is satisfied then*

$$\frac{1}{B_n} \sum_{u=1}^{k_n} X_{nu} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

**Remark 1** Let consider the case of binary random variables  $X_{nu}$  with mean 0. More specifically, set  $X_{nu} = a_{nu}Z_{nu}$ ,  $a_{nu} \in \mathbb{R}$ , where  $Z_{nu}$  are centered Bernoulli variables. More precisely,  $Z_{nu}$  takes value  $1 - p_{nu}$  with probability  $p_{nu}$  and value  $-p_{nu}$  with probability  $1 - p_{nu}$ . The event  $|X_{nu}| \geq \epsilon B_n$  in the definition of  $A_n^2(\epsilon)$  in (1) implies  $|a_{nu}| \geq \epsilon B_n$ . Therefore, all  $X_{nu}$  for which  $|a_{nu}| < \epsilon B_n$  do not contribute to  $A_n^2(\epsilon)$ . If this holds for all  $X_{nu}$ , then the Lindeberg condition is directly satisfied. If not, only the  $X_{nu}$  for which it does not hold have to be considered in the calculation of  $A_n^2(\epsilon)$  and, because  $|Z_{nu}| \leq 1$ , their contribution is upper-bounded by their variance  $\sigma_{nu}^2 = a_{nu}^2 p_{nu}(1 - p_{nu})$ . In the forthcoming theorems proofs, we will verify the Lindeberg condition using this observation.

**Theorem 1** Under model  $HER(\mathbf{p})$ , the degree variance is asymptotically normal:

$$(V - \mathbb{E}_{HER(\mathbf{p})}V) / \mathbb{S}_{HER(\mathbf{p})}V \xrightarrow{D} \mathcal{N}(0, 1),$$

where  $\mathbb{S}$  denotes the standard deviation and

$$\begin{aligned} \mathbb{E}_{HER(\mathbf{p})}V &= \frac{2(n-2)}{n^2} \sum_{1 \leq i < j \leq n} p_{ij} + \frac{2(n-4)}{n^2} \sum_{1 \leq i < j < k \leq n} \{p_{ij}p_{ik} + p_{ij}p_{jk} + p_{ik}p_{jk}\} \\ &\quad - \frac{8}{n^2} \sum_{1 \leq i < j < k < l \leq n} \{p_{ij}p_{kl} + p_{ik}p_{jl} + p_{il}p_{jk}\}. \end{aligned}$$

Moreover, denoting  $\sigma_{ij}^2 = p_{ij}(1 - p_{ij})$ ,

$$\begin{aligned} \mathbb{V}_{HER(\mathbf{p})}V &= \frac{1}{4n^4} \sum_{1 \leq i < j \leq n} \sigma_{ij}^2 \left( 4(n-2) + 4(n-4) \sum_{k \notin \{i,j\}} (p_{i,k} + p_{j,k}) - 16 \sum_{k < l \notin \{i,j\}} p_{kl} \right)^2 \\ &\quad + \frac{1}{n^4} \sum_{1 \leq i < j < k \leq n} 4(n-4)^2 \{ \sigma_{ij}^2 \sigma_{ik}^2 + \sigma_{ij}^2 \sigma_{jk}^2 + \sigma_{ik}^2 \sigma_{jk}^2 \} \\ &\quad + \frac{1}{n^4} \sum_{1 \leq i < j < k < l \leq n} 64 \{ \sigma_{ij}^2 \sigma_{kl}^2 + \sigma_{ik}^2 \sigma_{jl}^2 + \sigma_{il}^2 \sigma_{jk}^2 \}. \end{aligned}$$

*Proof.* Let express  $V$  as follows.

$$\begin{aligned} n^2V &= 2(n-2) \sum_{1 \leq i < j \leq n} Y_{ij} \\ &\quad + 2(n-4) \sum_{1 \leq i < j < k \leq n} \{Y_{ij}Y_{ik} + Y_{ij}Y_{jk} + Y_{ik}Y_{jk}\} \\ &\quad - 8 \sum_{1 \leq i < j < k < l \leq n} \{Y_{ij}Y_{kl} + Y_{ik}Y_{jl} + Y_{il}Y_{jk}\}. \end{aligned} \tag{2}$$

Then, we write the Hoeffding decomposition of  $V$  (see, e.g., Chapter 11 in van der Vaart (1998)) :

$$\begin{aligned} V &= P_\emptyset V + \sum_{1 \leq i < j \leq n} P_{\{ij\}} V + \sum_{1 \leq i < j < k \leq n} \{P_{\{ij,ik\}} V + P_{\{ij,jk\}} V + P_{\{ik,kj\}} V\} \\ &+ \sum_{1 \leq i < j < k < l \leq n} \{P_{\{ij,kl\}} V + P_{\{ik,jl\}} V + P_{\{il,jk\}} V\}, \end{aligned} \quad (3)$$

where

$$\begin{aligned} P_\emptyset V &= \mathbb{E}V, \\ P_{\{ij\}} V &= \mathbb{E}(V|Y_{ij}) - \mathbb{E}V, \\ P_{\{ij,ik\}} V &= \mathbb{E}(V|Y_{ij}, Y_{ik}) - \mathbb{E}(V|Y_{ij}) - \mathbb{E}(V|Y_{ik}) + \mathbb{E}V, \\ P_{\{ij,kl\}} V &= \mathbb{E}(V|Y_{ij}, Y_{kl}) - \mathbb{E}(V|Y_{ij}) - \mathbb{E}(V|Y_{kl}) + \mathbb{E}V. \end{aligned}$$

Combining the definitions above with the expression (2) of  $V$ , we obtain that,

$$\begin{aligned} P_\emptyset V &= \frac{1}{2n^2} \left( 4(n-2) \sum_{1 \leq i < j \leq n} p_{ij} + 4(n-4) \sum_{1 \leq i < j < k \leq n} \{p_{ij}p_{ik} + p_{ij}p_{jk} + p_{ik}p_{jk}\} \right) \\ &- \frac{8}{n^2} \sum_{1 \leq i < j < k < l \leq n} \{p_{ij}p_{kl} + p_{ik}p_{jl} + p_{il}p_{jk}\}, \end{aligned}$$

$$P_{\{ij\}} V = \frac{1}{2n^2} \tilde{Y}_{ij} \left( 4(n-2) + 4(n-4) \sum_{k \notin (i,j)} (p_{i,k} + p_{j,k}) - 16 \sum_{k < l \notin (i,j)} p_{kl} \right), \quad (4)$$

$$P_{\{ij,ik\}} V = \frac{2(n-4)}{n^2} \tilde{Y}_{ij} \tilde{Y}_{ik}, \quad (5)$$

$$P_{\{ij,kl\}} V = -\frac{8}{n^2} \tilde{Y}_{ij} \tilde{Y}_{kl}, \quad (6)$$

where  $\tilde{Y}_{ij} = Y_{ij} - p_{ij}$ . Observe now that,

$$\begin{aligned} n^2 \mathbb{E}V &= 2(n-2) \sum_{1 \leq i < j \leq n} p_{ij} + 2(n-4) \sum_{1 \leq i < j < k \leq n} \{p_{ij}p_{ik} + p_{ij}p_{jk} + p_{ik}p_{jk}\} \\ &- 8 \sum_{1 \leq i < j < k < l \leq n} \{p_{ij}p_{kl} + p_{ik}p_{jl} + p_{il}p_{jk}\}. \end{aligned}$$

Then, we use the fact that the  $\tilde{Y}_{ij}$ 's being independent and zero mean variables implies that each projection is orthogonal to the other ones. This leads to

$$\begin{aligned} n^4 \mathbb{V}V &= \sum_{1 \leq i < j \leq n} \sigma_{ij}^2 \left( 2(n-2) + 2(n-4) \sum_{k \notin \{i,j\}} (p_{i,k} + p_{j,k}) - 8 \sum_{k < l \notin \{i,j\}} p_{kl} \right)^2 \\ &+ \sum_{1 \leq i < j < k \leq n} 4(n-4)^2 \{ \sigma_{ij}^2 \sigma_{ik}^2 + \sigma_{ij}^2 \sigma_{jk}^2 + \sigma_{ik}^2 \sigma_{jk}^2 \} \\ &+ \sum_{1 \leq i < j < k < l \leq n} 64 \{ \sigma_{ij}^2 \sigma_{kl}^2 + \sigma_{ik}^2 \sigma_{jl}^2 + \sigma_{il}^2 \sigma_{jk}^2 \}. \end{aligned}$$

We now turn to the asymptotic normality of  $V$ . We apply the Lindeberg-Levy theorem to the projections  $P_{\{ij\}}V$ ,  $P_{\{ij,ik\}}V$  and  $P_{\{ij,kl\}}V$ , which stand for the  $X_{nu}$ . We first observe that these projections are each proportional to one of the  $\tilde{Y}_{ij}$ ,  $\tilde{Y}_{ij}\tilde{Y}_{ik}$ ,  $\tilde{Y}_{ij}\tilde{Y}_{kl}$ , ... which are all independent centered Bernoulli variables (because orthogonal and binary). We may now use Remark 1. We denote the respective  $a_{nu}$  by  $a_{n\{ij\}}$ ,  $a_{n\{ij,ik\}}$  and  $a_{n\{ij,kl\}}$ , the explicit expressions of which are given in (4), (5) and (6). Observe that  $a_{n\{ij\}} = O(1)$ ,  $a_{n\{ij,ik\}} = O(n^{-1})$  and  $a_{n\{ij,kl\}} = O(n^{-2})$ . Since  $\mathbb{V}P_{\{ij\}}V = O(1)$ ,  $\mathbb{V}P_{\{ij,ik\}}V = O(n^{-2})$  and  $\mathbb{V}P_{\{ij,kl\}}V = O(n^{-4})$ , we have  $B_n^2 = O(n^2)$ . We conclude therefore that the Lindeberg condition is fulfilled because, for any  $\epsilon$ , each  $a_{nu}$  becomes smaller than  $\epsilon B_n$  when  $n$  goes to infinity. ■

**Corollary 1** *Under model  $ER(p)$ , the degree variance is asymptotically normal:*

$$(V - \mathbb{E}_{ER(p)}V) / \mathbb{S}_{ER(p)}V \xrightarrow{D} \mathcal{N}(0, 1),$$

where

$$\begin{aligned} \mathbb{E}_{ER(p)}V &= \frac{(n-1)(n-2)pq}{n}, \\ \mathbb{V}_{ER(p)}V &= \frac{2(n-1)(n-2)^2}{n^3}pq(1 + (n-6)pq). \end{aligned}$$

This corollary is a straightforward application of Theorem 1 to the case where all  $p_{ij}$  are equal to  $p$ . Another proof of the asymptotic normality of  $V$  under  $ER(p)$  can be found in Bloznelis (2005). The moments have also been given by Hagberg (2003).

**Case of sparse graphs.** We now discuss the validity of Theorem 1 when considering sparse graphs. Sparsity can be defined in two ways. Either each of the connection probabilities vanishes as  $n$  grows, or the fraction of non-zero connection probabilities decreases as  $n$  grows. The following Proposition deals with a combination of both scenarios.

**Proposition 1** *Consider the  $HER(\mathbf{p})$  model, where  $\mathbf{p}$  is constructed as follows. First set all  $p_{ij} = p_{ij}^* n^{-a}$  where  $a \geq 0$  and  $p_{ij}^* \in [0, 1]$ . Then, a fraction  $1 - n^{-b}$ ,  $b \geq 0$ , of  $p_{ij}$ 's is set to zero. Then, provided that  $a + b < 2$ , the  $V$  statistic is asymptotically normal.*

*Proof.* The proof is a generalization of the one of Theorem 1 and follows Remark 1. The projections involved in (3) still stand for  $X_{nu}$ , and  $a_{n\{ij\}}$ ,  $a_{n\{ij,ik\}}$  as well as  $a_{n\{ij,kl\}}$  are expressed in (4), (5) and (6).

First observe that both sums  $A_n^2(\epsilon)$  and  $B_n^2$  in the Lindeberg condition can be split into three sub-sums over  $i < j$ ,  $i < j < k$  and  $i < j < k < l$ , respectively. Then, observe that the edges for which  $p_{ij}$  is zero do not contribute to  $A_n^2(\epsilon)$ . So, the number of non-zero terms in each of these sums is  $O(n^{2-b})$ ,  $O(n^{3-2b})$  and  $O(n^{4-2b})$ , respectively.

We now calculate the variance  $\sigma_{nu}^2$  of each projections. Since  $\sum_{k \notin \{i,j\}} p_{ik} = O(n^{1-a-b})$  and  $\sum_{k < l \notin \{i,j\}} p_{kl} = O(n^{2-a-b})$ , we see that  $a_{n\{ij\}} = O(n^{-(a+b)})$  if  $a+b < 1$  and  $O(n^{-1})$  if  $a+b > 1$ ,  $a_{n\{ij,ik\}} = O(n^{-1})$  and  $a_{n\{ij,kl\}} = O(n^{-2})$ . Therefore, we have  $\mathbb{V}P_{\{ij\}}V = O(n^{-3a-2b})$  if  $a+b < 1$  and  $O(n^{-a-2})$  if  $a+b > 1$ ,  $\mathbb{V}P_{\{ij,ik\}}V = O(n^{-2a-2})$  and  $\mathbb{V}P_{\{ij,kl\}}V = O(n^{-2a-4})$ . Combining this with the number of non-zero terms in the sums of the numerator of the Lindeberg condition, we get that  $B_n^2 = O(n^{2-3(a+b)})$  if  $a+b < 1$  and  $O(n^{-(a+b)})$  if  $a+b > 1$ .

Comparing  $A_n^2(\epsilon)$  with  $B_n^2$ , we see that the Lindeberg condition is fulfilled for  $a+b < 2$ . ■

## 2.2 Test and power

We now consider the use of the statistic  $V$  for the test of  $H_0 = ER$  versus  $H_1 = HER(\mathbf{p})$ . Because the probability is unknown in practice, we consider the following test statistic using a plug-in version of the moments, namely

$$(V - \mathbb{E}_{ER(\hat{p})}V) / \mathbb{S}_{ER(\hat{p})}V,$$

where  $\hat{p} = [n(n-1)]^{-1} \sum_{i \neq j} Y_{ij}$ .

**Lemma 1** *Under model ER, the degree variance is asymptotically normal:*

$$(V - \mathbb{E}_{ER(\hat{p})}V) / \mathbb{S}_{ER(\hat{p})}V \xrightarrow{D} \mathcal{N}(0, 1).$$

*Proof.* The proof relies on the concentration of  $\hat{p}$  around  $p$  and on Slutsky's lemma. First, write the statistic based on  $V$  as

$$\frac{V - \mathbb{E}_{\hat{p}}V}{\mathbb{S}_{\hat{p}}V} = \frac{\mathbb{S}_pV}{\mathbb{S}_{\hat{p}}V} \left( \frac{V - \mathbb{E}_pV}{\mathbb{S}_pV} + \frac{\mathbb{E}_pV - \mathbb{E}_{\hat{p}}V}{\mathbb{S}_pV} \right).$$

Then note that, under  $ER(p)$ ,  $(\hat{p} - p) = O_P(n^{-1})$ , so  $(\hat{p}\hat{q} - pq) = O_{\mathbb{P}}(n^{-2})$ , where  $\hat{q}$  stands for  $1 - \hat{p}$ . According to the moments given in Corollary 1, we have that  $\mathbb{E}_pV = O(n)pq$  and  $\mathbb{V}_pV = O(1)pq + O(n)p^2q^2$ . This entails that  $\mathbb{E}_pV - \mathbb{E}_{\hat{p}}V = O_{\mathbb{P}}(n^{-1})$  and  $\mathbb{V}_{\hat{p}}V - \mathbb{V}_pV = O_{\mathbb{P}}(n^{-2})$ , so  $\mathbb{S}_pV / \mathbb{S}_{\hat{p}}V$  converges in probability to 1 and  $(\mathbb{E}_pV - \mathbb{E}_{\hat{p}}V) / \mathbb{S}_pV$  converges in probability to 0. The result then follows from Slutsky's lemma, used twice. ■

The asymptotic power of this test depends on the asymptotic distribution of the statistic under the  $HER(\mathbf{p})$ . The following corollary shows that the asymptotic distribution of the test based on  $(V - \mathbb{E}_{ER(\hat{p})}V) / \mathbb{S}_{ER(\hat{p})}V$  is the same as the one of the test based on the statistic  $(V - \mathbb{E}_{ER(\bar{p})}V) / \mathbb{S}_{ER(\bar{p})}V$ .



**Lemma 2** *We have*

$$(V - \mathbb{E}_{ER(\hat{p})}V) / \mathbb{S}_{ER(\hat{p})}V - (V - \mathbb{E}_{ER(\bar{p})}V) / \mathbb{S}_{ER(\bar{p})}V \xrightarrow{\mathbb{P}} 0,$$

where  $\bar{p} = [n(n-1)]^{-1} \sum_{i \neq j} p_{ij}$ .

The proof of this Lemma is similar to this of Lemma 1 and results from the concentration of  $\hat{p}$  around  $\bar{p}$ . We now use the asymptotic normality of  $(V - \mathbb{E}_{ER(\bar{p})}V) / \mathbb{S}_{ER(\bar{p})}V$  under the  $HER(\mathbf{p})$  model to get the asymptotic power of the proposed test.

**Corollary 2** *The asymptotic power  $\pi(\mathbf{p}) = \mathbb{P}_{\mathbf{p}}\{V > t_\alpha\}$  of the test of  $H_0 = ER$  versus  $H_1 = HER(\mathbf{p})$ , with nominal level  $\alpha > 0$ , is*

$$\pi(\mathbf{p}) = 1 - \Phi\left(\frac{(\mathbb{E}_{ER(\bar{p})}V + t_\alpha \mathbb{S}_{ER(\bar{p})}V - \mathbb{E}_{HER(\mathbf{p})}V) / \mathbb{S}_{HER(\mathbf{p})}V}{1}\right),$$

where  $\Phi$  stands for the cumulative distribution function [cdf] of the standard normal distribution and  $t_\alpha = \Phi^{-1}(1 - \alpha)$ .

## 2.3 Illustration

As an illustration of the proposed test, we consider the following networks. For each of these networks, additional covariates are available, which will be considered in the next section.

**Ecological networks:** this consists in two ecological networks first introduced by Vacher *et al.* (2008) and further studied in Mariadassou *et al.* (2010). Each of these networks describe the interaction between a series of  $n = 51$  trees and  $n = 154$  fungi, respectively. In the tree network, two trees interact if they share at least one common fungal parasite. As for the fungal network, two fungi are linked if they are hosted by at least one common tree species.

**Political blogs network:** this consists in a set of  $n = 196$  French political blogs already studied by Latouche *et al.* (2011). Two blogs are connected if one contains an hyperlink to the other.

We first apply the degree variance test to each of these networks to check if the topology of these networks is similar to the one of an  $ER$  network. The results are given in Table 1. As expected, their topology are far too heterogeneous to fit an  $ER(p)$  model, and the null hypothesis is rejected for each one of them.

## 3 Degree mean square test

Because the  $ER$  model rarely fits real networks, we now consider a goodness-of-fit test for the  $HER$  model. More specifically, for a given matrix  $\mathbf{p}^0$  of connection probabilities, we consider the test statistic

$$W_{\mathbf{p}^0} = \frac{1}{n} \sum_i (D_i - \mu_i^0)^2,$$

where  $\mu_i^0$  stand for the expected degree of node  $i$  under  $HER(\mathbf{p}^0)$ , namely  $\mu_i^0 = \sum_{j \neq i} p_{ij}^0$ .

Network	$\hat{p}$	$V$	$\mathbb{E}_{ER(\hat{p})}$	$\mathbb{S}_{ER(\hat{p})}$	$(V - \mathbb{E}_{ER(\hat{p})})/\mathbb{S}_{ER(\hat{p})}$
Trees	0.540	163.1	11.93	2.34	64.6
Fungis	0.227	593.7	26.48	3.02	187.5
Blogs	0.075	104.3	13.38	1.38	65.7

Table 1: Degree variance test for the tree and fungal networks. The last column should be compared to a standard Gaussian distribution so all corresponding  $p$ -values are below the numerical precision.

### 3.1 Asymptotic normality

**Theorem 2** *Under model  $HER(\mathbf{p})$ , the statistic  $W_{\mathbf{p}^0}$  is asymptotically normal:*

$$(W_{\mathbf{p}^0} - \mathbb{E}_{HER(\mathbf{p})}W_{\mathbf{p}^0})/\mathbb{S}_{HER(\mathbf{p})}W_{\mathbf{p}^0} \xrightarrow{D} \mathcal{N}(0, 1),$$

with

$$\begin{aligned} \mathbb{E}_{HER(\mathbf{p})}W_{\mathbf{p}^0} &= \frac{2}{n} \left( \sum_{1 \leq i < j \leq n} (\sigma_{ij}^2 + \delta_{ij}^2) + \sum_{1 \leq i < j < k \leq n} (\delta_{ij}\delta_{ik} + \delta_{ij}\delta_{jk} + \delta_{ik}\delta_{jk}) \right), \\ \mathbb{V}_{HER(\mathbf{p})}W_{\mathbf{p}^0} &= \frac{4}{n^2} \left( \sum_{1 \leq i < j \leq n} \sigma_{ij}^2 (\Delta_i + \Delta_j + 1 - 2p_{ij})^2 + \sum_{1 \leq i < j < k \leq n} (\sigma_{ij}^2 \sigma_{ik}^2 + \sigma_{ij}^2 \sigma_{jk}^2 + \sigma_{ik}^2 \sigma_{jk}^2) \right), \end{aligned}$$

where  $\delta_{ij} = p_{ij} - p_{ij}^0$  and  $\Delta_i = \sum_{j \neq i} \delta_{ij}$ .

*Proof.* The proof follows the line of this of Theorem 1 and relies on Hoeffding decomposition and Lindeberg condition. First observe that,

$$\begin{aligned} nW_{\mathbf{p}^0} &= \sum_i (D_i - \mu_i + \mu_i - \mu_i^0)^2 = \sum_i \left( \sum_{j \neq i} \tilde{Y}_{ij} + \delta_{ij} \right)^2 \\ &= 2 \sum_{1 \leq i < j \leq n} (\tilde{Y}_{ij} + \delta_{ij})^2 \\ &\quad + 2 \sum_{1 \leq i < j < k \leq n} (\tilde{Y}_{ij} + \delta_{ij})(\tilde{Y}_{ik} + \delta_{ik}) + (\tilde{Y}_{ij} + \delta_{ij})(\tilde{Y}_{jk} + \delta_{jk}) + (\tilde{Y}_{ik} + \delta_{ik})(\tilde{Y}_{jk} + \delta_{jk}). \end{aligned}$$

Then, we write the Hoeffding decomposition of  $W_{\mathbf{p}^0}$ :

$$W_{\mathbf{p}^0} = P_{\emptyset}W_{\mathbf{p}^0} + \sum_{1 \leq i < j \leq n} P_{\{ij\}}W_{\mathbf{p}^0} + \sum_{1 \leq i < j < k \leq n} (P_{\{ij,ik\}}W_{\mathbf{p}^0} + P_{\{ij,jk\}}W_{\mathbf{p}^0} + P_{\{ik,kj\}}W_{\mathbf{p}^0}). \quad (7)$$

Note that the projections on disjoint pairs of edges  $P_{\{ij,kl\}}$  do not appear here. Taking all projections with respect to  $HER(\mathbf{p})$ , we have

$$nP_{\emptyset}W_{\mathbf{p}^0} = 2 \sum_{1 \leq i < j \leq n} (\sigma_{ij}^2 + \delta_{ij}^2) + 2 \sum_{1 \leq i < j < k \leq n} \delta_{ij}\delta_{ik} + \delta_{ij}\delta_{jk} + \delta_{ik}\delta_{jk},$$

which gives the expectation. The other projections provide the variance. We have

$$P_{\{ij\}}W_{\mathbf{p}^0} = \frac{2}{n}\tilde{Y}_{ij}(1 - 2p_{ij} + (\Delta_i + \Delta_j)), \quad P_{\{ij,ik\}}W_{\mathbf{p}^0} = \frac{1}{n}\tilde{Y}_{ij}\tilde{Y}_{ik}. \quad (8)$$

So,

$$\begin{aligned} n^2\mathbb{V}(P_{\{ij\}}W_{\mathbf{p}^0}) &= 4 \left( \mathbb{V}(\tilde{Y}_{ij}^2) + 2(\Delta_i + \Delta_j)\text{Cov}(\tilde{Y}_{ij}^2, \tilde{Y}_{ij}) + (\Delta_i + \Delta_j)^2\mathbb{V}\tilde{Y}_{ij} \right) \\ &= 4\sigma_{ij}^2(1 - 2p_{ij} + \Delta_i + \Delta_j)^2, \\ n^2\mathbb{V}(P_{\{ij,ik\}}W_{\mathbf{p}^0}) &= 4\sigma_{ij}^2\sigma_{ik}^2, \end{aligned}$$

and the variance of  $W_{\mathbf{p}^0}$  follows.

As for the asymptotic normality, we apply the Lindeberg-Levy theorem using Remark 1. The projections involved in (7) stand for  $X_{nu}$ . The  $a_{n\{ij\}} = O(1)$  and  $a_{n\{ij,ik\}} = O(n^{-1})$  expressed in (8) stand for  $a_{nu}$ . Since  $\mathbb{V}(P_{\{ij\}}W_{\mathbf{p}^0}) = O(1)$  and  $\mathbb{V}(P_{\{ij,ik\}}W_{\mathbf{p}^0}) = O(n^{-2})$ , we have  $B_n^2 = O(n^2)$ . We thus conclude that the Lindeberg condition is fulfilled because, for any  $\epsilon$ , each  $a_{nu}$  becomes smaller than  $\epsilon B_n$  when  $n$  goes to infinity. ■

**Case of sparse graphs.** We now extend Theorem 2 to sparse graphs, considering a setting similar to this of Proposition 1.

**Proposition 2** *Consider the  $HER(\mathbf{p})$  model, when  $p_{ij} = p_{ij}^*n^{-a}$ ,  $a > 0$ ,  $p_{ij}^* \in [0, 1]$  and a fraction  $1 - n^{-b}$ ,  $b \geq 0$ , of  $p_{ij}$ 's is set to zero. The  $p_{ij}^0$ 's follow exactly the same hypotheses. Then, provided that  $a + b < 2$ , the statistic  $W_{\mathbf{p}^0}$  is asymptotically normal.*

*Proof.* The proof is a generalization of this of Theorem 2 and follows the line of this of Proposition 1. First observe that both sums  $A_n^2(\epsilon)$  and  $B_n^2$  in the Lindeberg condition can be split into two sub-sums over  $i < j$  and  $i < j < k$ . The number of non-zero  $p_{ij}$  in each of these sums is  $O(n^{2-b})$  and  $O(n^{3-2b})$ . We now calculate the variance  $\sigma_{nu}^2$  of each projections in (8). Since  $\Delta_i = O(n^{1-a-b})$ , we see that  $a_{n\{ij\}} = O(n^{-(a+b)})$  if  $a + b < 1$  and  $O(n^{-1})$  if  $a + b > 1$ , and  $a_{n\{ij,ik\}} = O(n^{-1})$ . Therefore, we have  $\mathbb{V}P_{\{ij\}}V = O(n^{-3a-2b})$  if  $a + b < 1$  and  $O(n^{-a-2})$  if  $a + b > 1$ , and  $\mathbb{V}P_{\{ij,ik\}}V = O(n^{-2a-2})$ . This entails that  $B_n^2 = O(n^{2-3(a+b)})$  if  $a + b < 1$  and  $O(n^{-(a+b)})$  if  $a + b > 1$ . Comparing  $A_n^2(\epsilon)$  with  $B_n^2$ , we see that the Lindeberg condition is fulfilled for  $a + b < 2$ . ■

### 3.2 Test and power

We now study the test of  $H_0 = HER(\mathbf{p}^0)$  versus  $H_1 = HER(\mathbf{p})$ . The next Corollaries provide the null distribution of the test statistic  $W_{\mathbf{p}^0}$  and the power of the associate test.

**Corollary 3** Under model  $HER(\mathbf{p}^0)$  the statistic  $W_{\mathbf{p}^0}$  is asymptotically normal with moments:

$$\begin{aligned}\mathbb{E}_{HER(\mathbf{p}^0)}W_{\mathbf{p}^0} &= \frac{2}{n} \sum_{1 \leq i < j \leq n} \sigma_{ij}^2, \\ \mathbb{V}_{HER(\mathbf{p}^0)}W_{\mathbf{p}^0} &= \frac{4}{n^2} \left( \sum_{1 \leq i < j \leq n} \sigma_{ij}^2 (1 - 2p_{ij})^2 + \sum_{1 \leq i < j < k \leq n} (\sigma_{ij}^2 \sigma_{ik}^2 + \sigma_{ij}^2 \sigma_{jk}^2 + \sigma_{ik}^2 \sigma_{jk}^2) \right).\end{aligned}$$

This is a direct consequence of Theorem 2 in the special case of the  $HER(\mathbf{p}^0)$  model for which all  $\delta_{ij}$ 's are zero.

A formal test with asymptotic level  $\alpha$  can be constructed based on Corollary 3, which reject  $H_0$  as soon as  $W_{\mathbf{p}^0}$  exceeds  $\mathbb{E}_{HER(\mathbf{p}^0)}W_{\mathbf{p}^0} + t_\alpha \mathbb{S}_{HER(\mathbf{p}^0)}W_{\mathbf{p}^0}$ , where  $t_\alpha$  stands for the  $1 - \alpha$  quantile of the standard Gaussian distribution. The power of this test is given by the following Corollary.

**Corollary 4** The asymptotic power of the test for  $H_0 = HER(\mathbf{p}^0)$  versus  $H_1 = HER(\mathbf{p})$  is

$$\pi(\mathbf{p}) = 1 - \Phi \left( \left( \mathbb{E}_{HER(\mathbf{p}^0)}W_{\mathbf{p}^0} + t_\alpha \mathbb{S}_{HER(\mathbf{p}^0)}W_{\mathbf{p}^0} - \mathbb{E}_{HER(\mathbf{p})}W_{\mathbf{p}^0} \right) / \mathbb{S}_{HER(\mathbf{p})}W_{\mathbf{p}^0} \right).$$

**Special case of  $ER(p)$ .** The  $ER(p)$  model corresponds to  $HER(\mathbf{p}^0)$  where the matrix  $\mathbf{p}^0$  has all entries equal to  $p$ . In this case, the test statistic  $W_{\mathbf{p}^0}$  can be viewed as the theoretical version of the empirical variance statistic  $V$  studied in Section 2 as

$$W_{\mathbf{p}^0} = \frac{1}{n} \sum_i (D_i - (n-1)p)^2.$$

Because as  $\hat{p}$  is an average over  $O(n^2)$  edges, we have that  $(\hat{p} - p)^2 = O_P(n^{-2})$  so  $W_{\mathbf{p}^0} - V = (n-1)^2(\hat{p} - p)^2 = O_P(1)$ . Combined with arguments similar to these of Corollary 1 and Lemma 2, this implies that, under the  $ER$  model, the tests based on  $V$  and  $W_{\mathbf{p}^0}$  are asymptotically equivalent.

### 3.3 Illustration

To illustrate the use of the proposed test, we consider the same networks as in Section 2.2. Several covariates are available for each network. The genetic, taxonomic and geographical distances between tree species are available, as well as the nutritional similarities and the taxonomic distances between fungal species (see description in Mariadassou *et al.* (2010)). As for the blog network, the political party of each blog and the status of the writer (journalist or not) are also available.

The question is then to know if these covariates are sufficient to explain the heterogeneity of the network, at least in terms of degrees. To address this question, for each network

separately, we fitted a logistic regression model stating that  $\text{logit}(p_{ij}^0) = x_{ij}^\top \beta$ , where  $\text{logit}(u) = \log(u)/\log(1-u)$ ,  $u \in \mathbb{R}$ ,  $x_{ij} \in \mathbb{R}^d$  stands for the vector of covariates (intercept, distances, similarity) for the  $(i, j)$  and  $\beta$  for the vector of regression coefficients. A log-scale was used for the genetic distance.

This regression model provided us with an estimate of the connection probability matrix  $\mathbf{p}^0$ . We then applied the degree mean square test to check if the considered covariates are sufficient to explain the heterogeneity of the network. The results are given in Table 2 and again the null hypothesis  $H_0 : Y \sim \text{HER}(\mathbf{p}^0)$  is rejected for both networks. As for the ecological networks, these results are consistent with these from Mariadassou *et al.* (2010), who detected a residual heterogeneity in the valued versions of these networks after correction for these covariates.

Network	mean( $\mathbf{p}^0$ )	st-dev( $\mathbf{p}^0$ )	$W_{\mathbf{p}^0}$	$\mathbb{E}_{\text{HER}(\mathbf{p}^0)}$	$\mathbb{S}_{\text{HER}(\mathbf{p}^0)}$	$(W_{\mathbf{p}^0} - \mathbb{E}_{\text{HER}(\mathbf{p}^0)})/\mathbb{S}_{\text{HER}(\mathbf{p}^0)}$
Trees	0.540	0.192	136.8	10.57	2.09	60.4
Fungis	0.227	0.006	593.8	26.83	3.06	185.0
Blogs	0.075	0.119	78.7	10.7	1.17	57.8

Table 2: Degree mean square test for the tree and fungal networks.

## 4 Simulations study

We designed a simulation study to assess the performances of both the degree variance and the degree mean square tests. More specifically, the purpose of this study is to evaluate the power of the degree variance test for various graph sizes, densities (mean connectivities) and imbalances (in terms of degree). We also aim at illustrating for which graph size the asymptotic normality is reach; we especially focus on this point in the sparse regime. The purpose is the same for degree mean square test but, in addition, we want to study the performances of a plug-in version of this test where the probabilities  $p_{ij}^0$  would be replaced by some estimations  $\hat{p}_{ij}^0$ , obtained e.g. with a logistic regression on some observed covariates, resulting in a plug-in version of  $W_{\mathbf{p}^0}$  denoted as  $W_{\hat{\mathbf{p}}^0}$ .

The rationale behind this study is the following. Denoting  $\hat{\mu}_i^0 = \sum_{j \neq i} \hat{p}_{ij}^0$  and  $\Delta_i = \hat{\mu}_i^0 - \mu_i^0$ ,  $W_{\hat{\mathbf{p}}^0}$  writes

$$W_{\hat{\mathbf{p}}^0} := \frac{1}{n} \sum_i (D_i - \hat{\mu}_i^0)^2 = W_{\mathbf{p}^0} - \frac{2}{n} \sum_i (D_i - \mu_i^0) \Delta_i + \frac{1}{n} \sum_i \Delta_i^2. \quad (9)$$

If the  $\hat{p}_{ij}^0$  result from a typical parametric estimation based on the  $O(n^2)$  edges, we expect the estimation error  $|p_{ij}^0 - \hat{p}_{ij}^0|$  to be  $O_P(n^{-1})$ . Indeed this errors are not independent and are cumulated in each  $\Delta_i$  so the question is to know under which regime the last two terms of (9) can be neglected with respect to  $W_{\mathbf{p}^0}$ .

## 4.1 Simulation design

**Degree variance test.** We used a design similar to the one proposed in Latouche and Robin (2013). More precisely node  $i$  was associated with value  $u_i = i/(n + 1)$  and the probability  $p_{ij}$  was set to  $p_{ij} = \rho\lambda^2(u_i u_j)^{\lambda-1}$ . In this setting,  $\rho > 0$  controls the density of the graph ( $\rho = \bar{p}$ ) and  $\lambda > 0$  controls the imbalance of the degrees.  $\lambda = 1$  corresponds to the  $ER(\rho)$  model. Note that  $\lambda < \rho^{-1/2}$  must hold to keep all  $p_{ij}$  smaller than 1. For each combination of parameters, 1 000 simulations were ran and the test was carried out using the statistic  $V$ .

**Degree mean square test.** For the degree mean square, we designed a simulation that mimics the situation where an heterogeneous model is already considered, but still misses some heterogeneity. More precisely, each node  $i$  was associated with a vector of covariates  $x_i \in \mathbb{R}^d$  (all values were drawn iid with standard Gaussian distribution and  $d$  was set to 3). Each edge  $(i, j)$  was then associated with the covariate vector  $x_{ij} = x_i - x_j$ , when  $i < j$ . The edges were then drawn according a logistic model:  $\text{logit}(p_{ij}) = a + x_{ij}^\top \beta$ . The constant  $a$  was set to preserve the mean connectivity, denoted  $\rho$  in the degree variance simulation.

The probabilities  $p_{ij}^0$  of the null distribution were defined according to the same logistic model, removing the last covariate, namely  $\text{logit}(p_{ij}^0) = a + x_{ij}^{0\top} \beta^0$ , where  $x_{ij}^0$  and  $\beta^0$  are the same as  $x_{ij}$  and  $\beta$ , respectively deprived from the last coordinate. Hence, the discrepancy between the null hypothesis and the true model is measured by the coefficient of the last covariate, denoted  $\beta_1$  in the sequel. All  $\beta$ 's were set to 1 except  $\beta_1$  which ranged from 0 to 1. For each combination of parameters  $(n, \rho, \beta_1)$ , 1 000 simulations were ran and the test was carried out using the statistic  $W_{\mathbf{p}^0}$ .

We also investigated the case where the first covariates are observed but not directly the probabilities  $p_{ij}^0$ . For each simulated data set, we fitted a logistic regression model  $\text{logit}(p_{ij}^0) = x_{ij}^{0\top} \beta^0$ , ending up with an estimated  $\hat{\mathbf{p}}^0$  connection matrix. The tests were then performed using the plug-in statistic  $W_{\hat{\mathbf{p}}^0}$ .

**Sparse graphs.** Finally, we considered sparse graphs in the two settings described in Section 3. We focused on the asymptotic normality of the degree mean square statistic under the null hypothesis. To this aim, we simulated  $p_{ij}^{0*}$  probabilities as described above. We then considered the two sparsity scenarios:

- vanishing connection probabilities:  $p_{ij}^0 = p_{ij}^{0*} n^{-a}$ ;
- sparse connection probabilities:  $p_{ij}^0 = p_{ij}^{0*}$  with probability  $n^{-b}$  and 0 otherwise.

In both case the mean density of the non-sparse graph  $\rho = \overline{p_{ij}^{0*}}$  was set to 0.1. The density of the graphs therefore decrease as  $\rho n^{-a}$  and  $\rho n^{-b}$ , respectively.

**Criteria.** For each parameter configuration, we computed the moments of the respective statistics and derived the theoretical power. Based on the replicates, we estimated the empirical power. For the sparse setting, the proximity with the normal distribution was investigated plotting the empirical quantiles versus the theoretical Gaussian quantiles (QQ-plots).

## 4.2 Results

**Degree variance test.** The power curves of the degree variance test are given Figure 1. As expected, the power increases with both the graph size  $n$ , the network density  $\rho$  and the degree imbalance  $\lambda$ . The binomial confidence interval around the theoretical power informs us about the convergence to the asymptotic normality. We observe that the empirical power (dots) falls within this interval for all  $\rho$  and  $\lambda$  for  $n \geq 300$ , for  $n = 100$  as soon as  $\rho \geq 0.01$  and even for  $n \simeq 30$  for denser graphs ( $\rho \geq 0.1$ ).

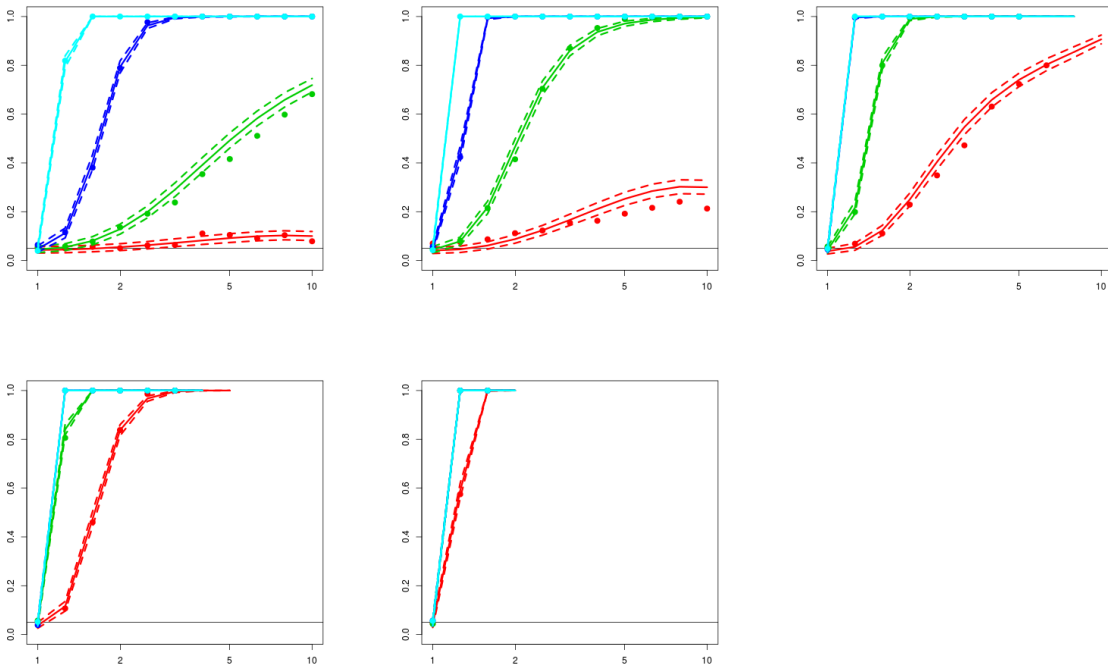


Figure 1: Power of the degree variance test as a function of the imbalance of the degree  $\lambda$  (in log scale). From top left to bottom center:  $\log_{10} \rho = -5/2$  to  $-1/2$  by step of  $1/2$ . Color refers to the graph size:  $n = 32$  (red), 100 (green), 316 (blue), 1 000 (cyan) (green, blue and cyan curves and points overlap in the last panels). Points = empirical power (average on 1 000 simulations); solid line = theoretical power; dotted = binomial confidence interval for 1 000 simulations.

**Degree mean square test.** The power curves of the degree mean square test are given Figure 2. We observe a behavior similar to this of the  $V$  based test: the power increase with both the graph size  $n$ , the density  $\rho$  and the unexpected heterogeneity  $\beta_1$ . Note that the covariates  $x_{ij}$  (and hence the probabilities  $p_{ij}$ ) were kept fixed across the simulation for a given configuration of the parameters  $n$ ,  $\rho$  and  $\beta_1$ . As the power of the test depends on both these parameters and the covariates  $x_{ij}$ , depending on their values, the power may not increase with  $\beta_1$  (see e.g. the top right panel of 2 for  $n = 32$ ). Here again, the empirical power (dots) of  $W_{\mathbf{p}^0}$  based test falls within the binomial confidence band, showing that the asymptotic normality is reach for reasonable small and dense graphs.

**Plug-in version of the degree mean square test.** The results are much more disappointing for the plug-in version  $W_{\hat{\mathbf{p}}^0}$ . Its empirical power (triangles) reaches the theoretical ones only for large and dense graphs ( $n \geq 300$ ,  $\rho \geq 0.1$ ). This suggests that the cumulative effect of all the estimation errors  $|\hat{p}_{ij}^0 - p_{ij}^0|$  on  $W_{\hat{\mathbf{p}}^0}$  vanishes much later than the convergence of  $W_{\mathbf{p}^0}$  to normality.

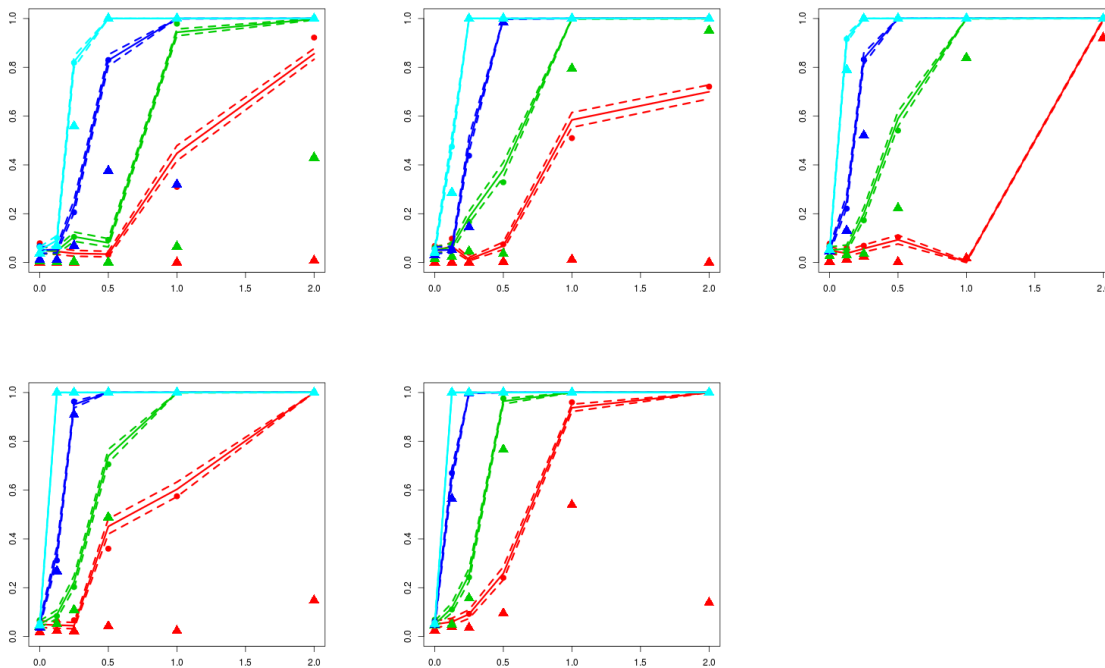


Figure 2: Power of the degree mean square test as a function of  $\beta_1$ , the effect of the last covariate. Same legend as Figure 1 for colors, points and lines. Dotted points:  $W_{\mathbf{p}^0}$  test; triangular points:  $W_{\hat{\mathbf{p}}^0}$  test.



**Sparse graphs.** Figure 3 displays the QQ-plots of the standardized  $W_{\mathbf{p}^0}$  statistic under the vanishing probabilities scenario for graphs with size  $n = 100, 1\,000, 10\,000$ . Remember that the larger the power  $a$ , the sparser the graph. We observe again that normality holds for the non sparse graphs ( $a = 0$ ) even for  $n = 100$ , but the departure is visible for  $n = 100$  as soon as  $a \geq 0.4$ . The same is observed for  $n = 1\,000$ , although a bit later ( $a \geq 0.8$ ). For the largest graph ( $n = 10\,000$ ), normality holds until  $a \simeq 1.6$  but does not seem to be reached for higher sparsity regimes. Similar conclusions can be drawn from Figure 4 for the sparse probabilities scenario, each distribution being slightly closer to normal. Based on this simulation study, we would not advise to rely on asymptotic normality to perform a test for graphs of size  $n = 100$  and density smaller that  $\rho n^{-0.2} \simeq 5\%$  or for graphs of size  $n = 1\,000$  and density smaller that  $\rho n^{-0.4} \simeq 5\%$ . From a practical point-of-view, in the very sparse regime, normality can only be relied on for very large graphs.

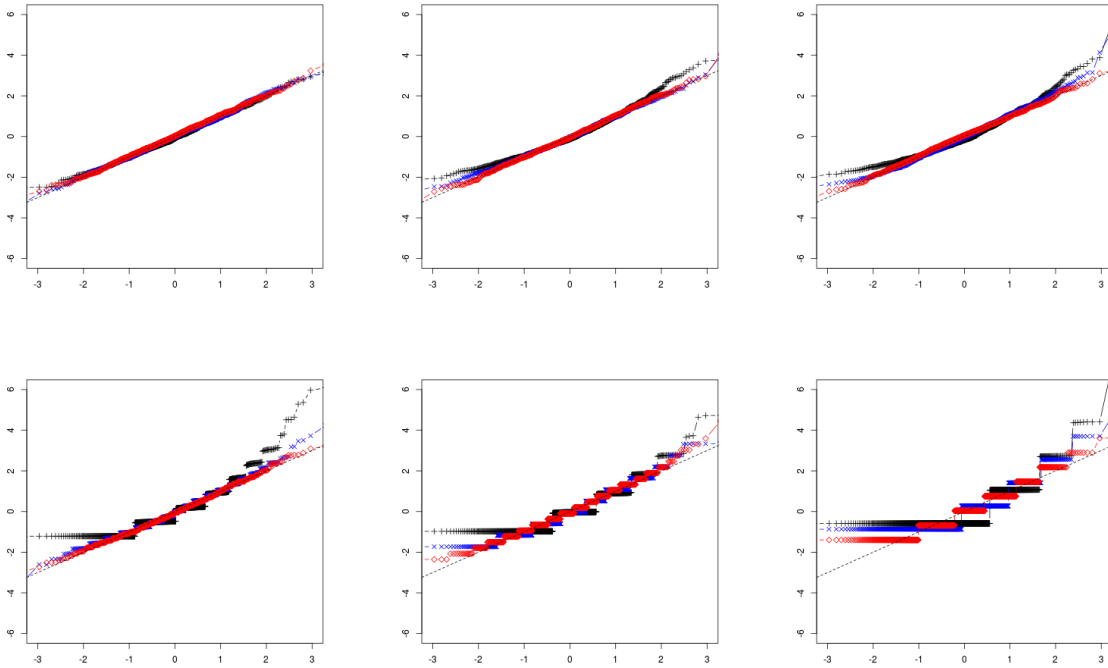


Figure 3: QQ-plots of the degree mean square statistic for vanishing connection probabilities:  $p_{ij} = p_{ij}^* n^{-a}$  and initial mean density  $\rho = 0.1$ . From top left to bottom right:  $a = 0, 0.4, 0.8, 1.2, 1.4, 1.6$ . Graph size  $n = 100$  (+),  $1\,000$  ( $\times$ ) and  $10\,000$  ( $\diamond$ ).

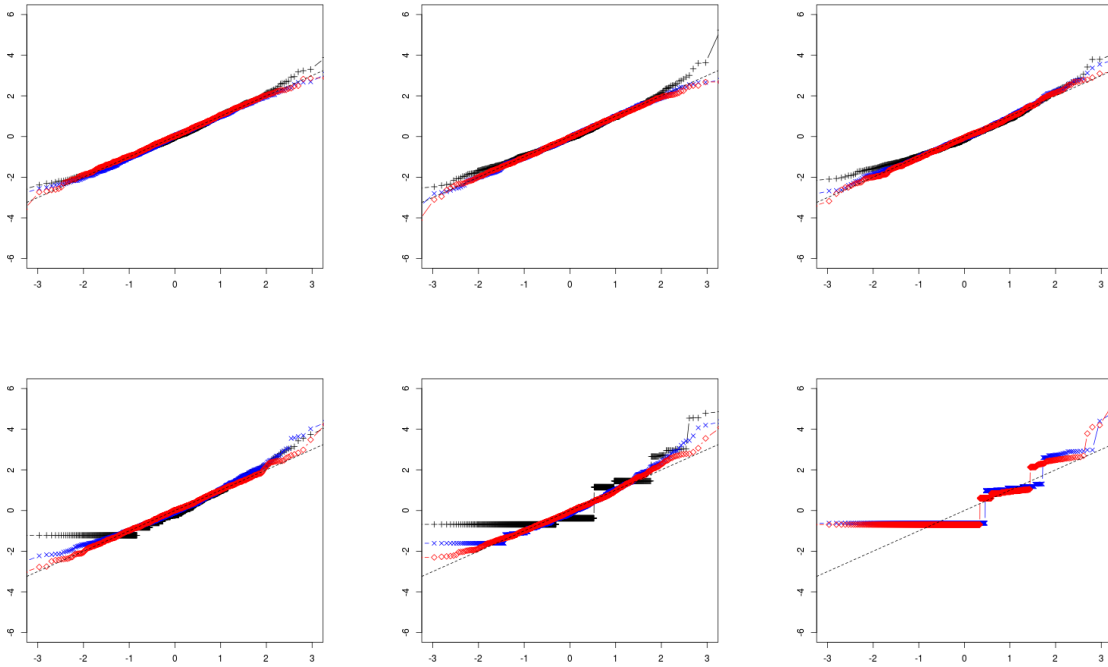


Figure 4: QQ-plots of the degree mean square statistic for a fraction  $n^{-b}$  of non-zero connection probabilities  $p_{ij}$ , with initial mean density  $\rho = 0.1$ . From top left to bottom right:  $b = 0, 0.4, 0.8, 1.2, 1.4, 1.6$ . Graph size  $n = 100$  (+), 1000 ( $\times$ ) and 10000 ( $\diamond$ ).

**Acknowledgements.** We thank Pr Bloznelis for providing us with his report referred to as Bloznelis (2005).

## References

- [Barabási and Albert (1999)] BARABÁSI, A. L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science*. **286** 509–512.
- [Bickel *et al.* (2011)] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *The Annals of Statistics*. **39** (5) 2280–2301.
- [Billingsley (1968)] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley: New-York.
- [Bloznelis (2005)] BLOZNELIS, M. (2005), Degree variance is asymptotically normal. Technical report, Vilnius university, Faculty of Mathematics and Informatics.

- [Cerqueira *et al.* (2015)] CERQUEIRA, A., FRAIMAN, D., VARGAS, C. D. and LEONARDI, F. (Apr., 2015), A test of hypotheses for random graph distributions built from EEG data. Technical report, ArXiv:1504.06478.
- [Channarond *et al.* (2012)] CHANNAROND, A., DAUDIN, J.-J. and ROBIN, S. (2012). Classification and estimation in the stochastic block model based on the empirical degrees. *Elec. J. Stat.* **6** 2574–601.
- [Hagberg (2000)] HAGBERG, J. (2000), Centrality testing and the distribution of the degree variance in bernoulli graphs. Technical report, Department of Statistics, Stockholm University.
- [Hagberg (2003)] HAGBERG, J. (2003), General moments of degrees in random graphs. Technical report, Stockholm University, Department of Statistics.
- [Latouche and Robin (2013)] LATOUCHE, P. and ROBIN, S. (Oct., 2013), Bayesian Model Averaging of Stochastic Block Models to Estimate the Graphon Function and Motif Frequencies in a W-graph Model. Technical report, ArXiv:1310.6150.
- [Latouche *et al.* (2011)] LATOUCHE, P., BIRMELÉ, E. and AMBROISE, C. (2011). Overlapping stochastic block models with application to the French political blogosphere. *Ann. Appl. Stat.* **5** (1) 309–336.
- [Mariadassou *et al.* (2010)] MARIADASSOU, M., ROBIN, S. and VACHER, C. (2010). Uncovering structure in valued graphs: a variational approach. *Ann. Appl. Statist.* **4** (2) 715–42.
- [Nowicki and Wierman (1988)] NOWICKI, K. and WIERMAN, J. C. (1988). Subgraph counts in random graphs using incomplete u-statistics methods. *Discrete Mathematics.* **72** (1) 299–310.
- [Snijders (1981)] SNIJDERS, T. A. (1981). The degree variance: An index of graph heterogeneity. *Social Networks.* **3** (3) 163–174.
- [van der Vaart (1998)] VAN DER VAART, A. W. (1998). *Asymptotic statistics.* volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge.
- [Vacher *et al.* (2008)] VACHER, C., PIOUS, D. and DESPREZ-LOUSTAU, M.-L. (2008). Architecture of an antagonistic tree/fungus network: The asymmetric influence of past evolutionary history. *PLoS ONE.* **3** (3) 1740. doi:10.1371/journal.pone.0001740.