

# Detecting a late changepoint in the preferential attachment model

GIANMARCO BET<sup>1,a</sup> , KAY BOGERD<sup>2,b</sup>, RUI M. CASTRO<sup>2,c</sup>  and  
REMCO VAN DER HOFSTAD<sup>2,d</sup> 

<sup>1</sup>*Department of Mathematics and Computer Science, University of Florence, Florence, Italy,*

<sup>a</sup>[gianmarco.bet@unifi.it](mailto:gianmarco.bet@unifi.it)

<sup>2</sup>*Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands,* <sup>b</sup>[kaybogerd@hotmail.com](mailto:kaybogerd@hotmail.com), <sup>c</sup>[rmcastro@tue.nl](mailto:rmcastro@tue.nl), <sup>d</sup>[r.w.v.d.hofstad@tue.nl](mailto:r.w.v.d.hofstad@tue.nl)

Motivated by the problem of detecting a change in the evolution of a network, we consider the preferential attachment random graph model with a *time-dependent* attachment function. Our goal is to detect whether the attachment mechanism changed over time, based on a single snapshot of the network and without directly observable information about the dynamics. We cast this question as a hypothesis testing problem, where the null hypothesis is a preferential attachment model with a constant affine attachment parameter  $\delta_0$ , and the alternative hypothesis is a preferential attachment model where the affine attachment parameter changes from  $\delta_0$  to  $\delta_1$  at an unknown changepoint time  $\tau_n$ . For our analysis we focus on the regime where  $\delta_0$  and  $\delta_1$  are fixed, and the changepoint occurs close to the observation time of the network (i.e.,  $\tau_n = n - cn^\gamma$  with  $c > 0$  and  $\gamma \in (0, 1)$ ). This corresponds to a rather relevant scenario where we aim to detect the changepoint shortly after it has happened. We present two tests based on the number of vertices with minimal degree, and show that these are asymptotically powerful when  $\frac{1}{2} < \gamma < 1$ . We conjecture that any test based on the final network snapshot will be powerless when  $\gamma < \frac{1}{2}$ . The first test we propose requires knowledge of  $\delta_0$ . The second test is significantly more involved, and does not require the knowledge of  $\delta_0$  while still achieving the same asymptotic performance guarantees. Furthermore, we prove that the test statistics for both tests are asymptotically normal, allowing for accurate calibration of the tests. This is demonstrated by numerical experiments, that also illustrate the finite sample test properties.

*Keywords:* Changepoint detection; preferential attachment model

## 1. Introduction

One of the most celebrated successes of complex network theory has been the recognition that simple *dynamical* random graph models with local connection rules are able to successfully explain important macroscopic features observed in real-world networks. The preferential attachment model and its generalizations are perhaps the most successful of such models. Barabási and Albert [5] proposed this model to explain the occurrence of power-law degree sequences, which are often observed in real-world networks such as the world wide web [1,12] or internet [22], biological networks [23,26,29], collaboration networks of movie actors [2,24], and citation networks [6,30,33,38]. Furthermore, the typical distance between vertices in the preferential attachment model is small [19] (see also [36, Chapter 8] and references therein). This is called the *small-world* phenomenon [40,41].

The preferential attachment model considers the entire evolution of a network by adding vertices one by one using a simple *preferential attachment* rule. Informally, as new vertices are added to the graph, they are more likely to attach to vertices that already have a large degree, hence further increasing the degree of these vertices. This formalism essentially creates a paradigm where “the rich get richer”, which is often invoked to explain the wide-spread inequality in socio-economic contexts [32]. Accordingly, the degree of the oldest vertices grows as new vertices attach to the graph. On the

other hand, the degree of the last few vertices to join is typically quite small. Since its introduction in [5], the preferential attachment model has received a tremendous amount of attention thanks to its early explanatory successes. The structural properties of the model are investigated formally in [10,11], see also [35,36] for a detailed overview on this model and many of its properties.

In our work we are interested in situations where the growth dynamics of the network do not remain constant over time, but undergoes a change at some point. This captures a situation where a major event could cause a change in the subsequent evolution of the network. To model this, we consider a time-inhomogeneous affine preferential attachment model, where a new vertex  $v_t$  that enters the graph at time  $1 \leq t \leq n$  connects to a pre-existing vertex with degree  $k$  with probability proportional to  $f(k) = k + \delta(t)$ . We consider the hypothesis testing problem where  $\delta(t) = \delta_0$  remains constant under the null hypothesis, whereas under the alternative the affine attachment parameter  $\delta(t)$  changes from  $\delta_0$  to  $\delta_1$  at an unknown moment  $\tau_n$ , called a *changepoint*. For our work we are particularly interested in scenarios where the change occurs very late, and affects only a very small part of the graph. Specifically, in the regime we are interested in, the changepoint has the form  $\tau_n = n - cn^\gamma$  with  $c > 0$  and  $\gamma \in (0, 1)$ , as explained in Section 2. From a practical standpoint this is the most relevant regime, particularly when one wants to detect the change as quickly as possible.

## 1.1. Related work

Our work nicely complements earlier results [4,8] that focus on the detection of a changepoint in the setting of preferential attachment trees, where every vertex that enters the graph connects to  $m = 1$  other vertices. There are, however, some important differences. First, our results consider the more general case of preferential attachment graphs, where vertices may enter the graph with  $m \geq 1$  edges. The other difference is that we focus on a *late* changepoint  $\tau_n = n - cn^\gamma$ , whereas [4,8] focus on a changepoint that happens at a linear time  $O(n)$  or even  $o(n)$ . Thus, in our setting a much smaller number of vertices enter the graph after the changepoint, making it harder to detect. Moreover, in practice one desires to detect changes as quickly as possible, further justifying the focus on late changepoints. From the moment a pre-print of our work appeared on arXiv further progress on the late changepoint detection problem has been made first in [27], and then in [21]. In the former work, the changepoint is parametrized as  $\tau_n = n - \Delta_n$  and it is proven that, when the *labeled* graph is observed, detection is possible if and only if  $\Delta_n \rightarrow \infty$ . In other words, in the labeled case, detection is impossible if and only if a bounded number of incoming vertices attach according to the new attachment function. The authors also prove that, when the *unlabeled* graph is observed, detection is impossible if  $\Delta_n = o(n^{1/3})$ . In the work [21], the authors build on the results and the techniques of [27] to prove that detection is impossible if  $\Delta_n = o(n^{1/2})$ . These results complement our main result, which implies that, when the unlabeled graph is observed, asymptotically accurate detection is possible when  $\Delta_n = cn^\gamma$  with  $\gamma > 1/2$ .

The authors of [16] propose a likelihood-ratio testing procedure to detect a changepoint in a preferential attachment tree and the associated changepoint estimator. Crucially the methods in [16] rely on the knowledge of the entire network evolution. This is not the case for our test, which only requires a snapshot of the network at the final time. The authors of [16] extend their test to detect multiple changepoints by applying two general techniques (namely, Screening and Ranking, and Binary Segmentation) to decompose the multiple-changepoints problem into a sequence of single changepoint problems. This work, however, is still in the scenario where the changepoint occurs at a linear time, in stark contrast with the regime we investigate.

Although different from this work, there has been much interest in understanding and detecting the effect of an initial seed graph on the evolution of the preferential attachment tree [13–15,17,28]. Here one starts with a given initial graph at time  $t = 1$  and then grows the remaining tree according to a preferential

(or uniform) attachment. The goal is to estimate the initial seed graph based on an observation of the fully developed graph at a much later time. Finally, changepoint detection and related inference questions have also received much attention in the setting of dynamic stochastic block models [9,31,37,39,42]. In those works the aim is primarily to understand the evolution of the network’s community structure.

## 2. Model

We formalize the problem of detecting a changepoint in a dynamical network as a hypothesis testing problem on random graphs. We first explain the model that we use in general, and then define concrete versions of this model for the null and alternative hypothesis. This model has parameters  $m \in \mathbb{N}$  and  $\delta : \mathbb{N} \rightarrow (-m, \infty)$  and produces a sequence of undirected graphs without loops. Let  $G_n = (V_n, E_n)$  be an undirected graph, where  $V_n = \{v_0, \dots, v_n\}$  denotes the vertex set and  $E_n \subseteq \{\{i, j\} : i, j \in V_n\}$  denotes a random *multiset* of edges. In particular, the graph  $G_n$  will not, in general, be a simple graph. Note that  $G_n$  has  $n + 1$  vertices. For  $v \in V_n$  let  $D_v(G_n)$  denote the degree of vertex  $v$  in the graph  $G_n$ .

There exist various versions of the preferential attachment model, each following slightly different conventions for adding new vertices. Here we consider the following model: the first graph  $G_1$ , also called the seed graph, consists of two vertices  $v_0$  and  $v_1$  connected by  $m$  edges. For  $t > 1$ , the graph  $G_t$  is constructed by taking  $G_{t-1}$  and adding one extra vertex  $v_t$ , that is then connected to the vertices in  $G_t$  by exactly  $m$  edges. In the model we consider, this process is better described by introducing a number of intermediate steps, described by graphs  $G_{t,0}, G_{t,1}, \dots, G_{t,m}$ , with vertex set  $V_t = \{v_0, \dots, v_t\}$ . Begin by defining  $G_{t,0}$  to be identical to  $G_{t-1}$  together with an isolated vertex  $v_t$ . The graph  $G_{t,1}$  is obtained by adding an edge between  $v_t$  and one of the vertices in  $V_{t-1}$  with probability proportional to  $D_v(G_{t,0}) + \delta(t)$ . In general, for  $i \in [m]$ , we proceed by sampling vertex  $v_{t,i} \in \{v_0, \dots, v_{t-1}\}$  with conditional probability

$$\mathbb{P}(v_{t,i} = v \mid G_{t,i-1}) = \frac{D_v(G_{t,i-1}) + \delta(t)}{\sum_{j=0}^{t-1} D_j(G_{t,i-1}) + \delta(t)}, \tag{1}$$

and constructing  $G_{t,i}$  by adding the edge  $\{v_{t,i}, v_t\}$  to  $G_{t,i-1}$ . Finally, define  $G_t = G_{t,m}$ . Note that the degree of  $v_t$  in  $G_t$  is exactly  $m$ .

The above model is rather general, as it allows for quite a bit of flexibility in terms of the function  $\delta(t)$ , as the only requirement is that  $\delta(t) > -m$  to ensure that (1) is indeed a valid probability. A classical choice is to take  $\delta(t)$  as a constant, and the properties of the ensuing graphs are well studied (e.g., see [35,36] and the references therein). However, we are interested in knowing when it is possible to distinguish graphs generated by a model where  $\delta(t)$  is constant versus graphs generated by a model where  $\delta(t)$  is a step function. The latter models a preferential attachment evolution, where at some point the characteristics of the attachment process change.

Consider the above model, and let  $G_n$  denote the “last” graph obtained. Informally, our only observation is  $G_n$ , without any vertex labels. Formally, our only observation is the equivalence class of  $G_n$  under vertex relabeling. In other words, we have only access to the topological structure of  $G_n$  and not to the sequence  $\{G_1, \dots, G_{n-1}\}$ . In particular, the arrival order of the vertices is unknown to us. With an abuse of notation, from now on we will denote by  $G_n$  the equivalence class of  $G_n$  under vertex relabeling. Clearly the distribution of this random graph is parameterized by  $m$  and  $\delta$ . Since  $G_n$  has exactly  $n + 1$  vertices and  $nm$  edges, we have knowledge of  $m$  (so this is not an unknown parameter). Therefore the only unknown parameter is the function  $\delta$ .

Our goal is to determine when one can find evidence in the final graph that the growth dynamics of the network has changed at some point. This can be rather naturally formulated as an hypothesis

testing problem. Namely, we would like to conduct the following binary hypothesis test: under the null hypothesis (denoted by  $H_0$ ) we assume  $\delta(t) = \delta_0 > -m$  for all  $t \in \mathbb{N}$ . Under the alternative hypothesis (denoted by  $H_1$ ) we assume  $\delta$  is a step function, namely

$$\delta(t) = \mathbb{1}\{t \leq \tau_n\}\delta_0 + \mathbb{1}\{t > \tau_n\}\delta_1,$$

for some  $\delta_0 \neq \delta_1$  with  $\delta_0, \delta_1 > -m$ , and  $\tau_n \in \mathbb{N}$  with  $\tau_n \leq n$ .

Our main research goal is to determine when it is possible to distinguish the two hypotheses, based solely on  $G_n$  (where we do not know the order in which the various vertices have arrived). For our first result we consider  $\delta_0$  to be known, but we then relax this assumption and devise a test that does not require this knowledge while retaining the same asymptotic power characterization. In both cases we use the parameterization  $\tau_n = n - cn^\gamma$ , where  $c > 0$  and  $\gamma \in (0, 1)$ , and obviously  $c$  and  $\gamma$  are unknown.

Note that the alternative model does coincide with the null model when either  $\delta_1 = \delta_0$  or  $\tau_n = n$ . Furthermore, since  $\delta$  is a step-function, the attachment rule in (1) can be further simplified to get, for  $v \in \{v_0, \dots, v_{t-1}\}$ ,

$$\mathbb{P}(v_{t,i} = v \mid G_{t,i-1}) = \begin{cases} \frac{D_v(G_{t,i-1}) + \delta_0}{2^{(t-1)m+t}\delta_0 + (i-1)} & \text{if } t \leq \tau_n, \\ \frac{D_v(G_{t,i-1}) + \delta_1}{2^{(t-1)m+t}\delta_1 + (i-1)} & \text{if } t > \tau_n. \end{cases} \tag{2}$$

### 2.1. Assumptions and notation

Throughout this paper, when limits are unspecified, they are taken as the graph size  $n \rightarrow \infty$ . We recall that we consider the parameterization  $\tau_n = n - cn^\gamma$ . All the parameters  $m, \delta_0, \delta_1, c$ , and  $\gamma$  are assumed to remain constant as a function of  $n$ . We use the subscripts 0 and 1 in the expectation and probability operators to indicate whether we are considering the null or alternative hypothesis. Finally, we also use standard asymptotic notation:  $a_n = O(b_n)$  when  $|a_n/b_n|$  is bounded,  $a_n = \Omega(b_n)$  when  $b_n = O(a_n)$ ,  $a_n = o(b_n)$  when  $a_n/b_n \rightarrow 0$ ,  $a_n = \omega(b_n)$  when  $b_n = o(a_n)$ , and  $a_n \asymp b_n$  when  $a_n = (1 + o(1))b_n$ . Also, we use the probabilistic versions of these:  $a_n = O_{\mathbb{P}}(b_n)$  when  $|a_n/b_n|$  is stochastically bounded,  $a_n = \Omega_{\mathbb{P}}(b_n)$  when  $b_n = O_{\mathbb{P}}(a_n)$ , and  $a_n = o_{\mathbb{P}}(b_n)$  when  $a_n/b_n$  converges to 0 in probability.

## 3. Minimal degree tests

All the tests we consider use the information contained in low-degree vertices, in particular the number of vertices of minimal degree. Although this is a rather simple idea, the number of minimal degree vertices is substantially affected by the presence of a changepoint, even at late stages in the growth of the graph.

### 3.1. Powerful test for known $\delta_0$

We begin with the assumption that  $\delta_0$  is known. Although this might seem unrealistic, it provides a wealth of information about the properties of the test statistic we consider, and paves the way for more general results when  $\delta_0$  is unknown.

To define our test we first introduce some notation. To reduce the notational burden we identify the set of vertices  $\{v_0, \dots, v_n\}$  with  $[n] := \{0, 1, \dots, n\}$ . Furthermore, let  $D_v(t) := D_v(G_t)$  denote the degree

of vertex  $v$  in graph  $G_t$ , and let  $N_k(t)$  be the number of vertices of degree  $k$  in the graph  $G_t$ , that is,

$$N_k(t) := \sum_{v \in [t]} \mathbb{1}\{D_v(t) = k\}.$$

Since each vertex is attached to at least  $m$  other vertices, in our model we naturally have that  $N_k(n) = 0$  for  $k < m$ , and  $N_m(n)$  denotes the number of vertices with minimal degree. The latter quantity plays a crucial role in our test.

It is well-known that in the classical preferential attachment model, which corresponds to our null model, the number of vertices of degree  $k \geq m$  is highly concentrated [18,35]. In particular,  $N_k(n)$  is well concentrated around  $np_k(\delta_0)$  where  $p_k = p_k(\delta_0)$  satisfies the recursion

$$p_k = \frac{k - 1 + \delta_0}{2 + \delta_0/m} p_{k-1} - \frac{k + \delta_0}{2 + \delta_0/m} p_k,$$

for  $k > m$  with

$$p_m(\delta_0) = \frac{2 + \delta_0/m}{m + \delta_0 + 2 + \delta_0/m}.$$

This recursion is easily solved, giving rise to the following expression for  $p_k(\delta_0)$ :

$$p_k(\delta_0) := (2 + \delta_0/m) \frac{\Gamma(k + \delta_0)\Gamma(m + 2 + \delta_0 + \delta_0/m)}{\Gamma(m + \delta_0)\Gamma(k + 3 + \delta_0 + \delta_0/m)}.$$

Thus,  $p_k(\delta_0)$  is the limiting degree distribution of the random graph  $G_n$  under the null model.

We are now able to introduce our test statistic, that simply compares the number of minimal degree vertices to its asymptotic expected value under the null model, as

$$T(G_n) := N_m(n) - np_m(\delta_0).$$

If the observed value of  $T(G_n)$  is significantly different from zero, then we have evidence to reject the null model. This brings us to the first result, characterizing when such a test is powerful. Specifically we introduce a test based on this statistic that is guaranteed to have type-I error that is at most  $\alpha$  (asymptotically), and that is asymptotically powerful provided  $\gamma > \frac{1}{2}$ . In other words, under the alternative hypothesis the type-II error converges to zero provided  $\gamma > \frac{1}{2}$ . When  $\gamma < \frac{1}{2}$  the test is powerless, and when  $\gamma = \frac{1}{2}$ , the type-II error is bounded by a constant that depends on the specific model parameters.

Although this result indicates when the test is powerful or powerless, it provides only a conservative upper bound on the type I error. It follows from [3] that this test statistic is asymptotically normal, and therefore we can calibrate this test to guarantee that the type I error is  $(1 + o(1))\alpha$  as  $n \rightarrow \infty$ . We do this in Section 3.3. However, we present Proposition 3.1 first because it allows us to introduce the essential ideas of the proof while avoiding many technical details. The proof of Proposition 3.1 is given in Section 7.

**Proposition 3.1 (Asymptotically powerful test: known  $\delta_0$ ).** *Consider the test that rejects the null hypothesis for large values of  $T(G_n)$ . Namely, define the test*

$$\psi(G_n) := \mathbb{1}\left\{|T(G_n)| \geq m\sqrt{8n \log(2/\alpha)}\right\},$$

where  $\alpha \in (0, 1)$ . The type-I error of this test is asymptotically bounded by  $\alpha$ , i.e.,

$$\mathbb{P}_0(\psi(G_n) \neq 0) \leq (1 + o(1))\alpha.$$

Furthermore, the type-II error of this test satisfies

$$\begin{aligned} &\mathbb{P}_1(\psi(G_n) = 0) \\ &\leq \begin{cases} o(1) & \text{when } \gamma > \frac{1}{2}, \\ (2 + o(1)) \exp\left(-\left(\left(\frac{c|1 - p_m(\delta_0)/p_m(\delta_1)|}{m\sqrt{8}} - \sqrt{\log(2/\alpha)}\right) \vee 0\right)^2\right) & \text{when } \gamma = \frac{1}{2}, \end{cases} \end{aligned}$$

and  $\mathbb{P}_1(\psi(G_n) = 0) \geq (1 + o(1))(1 - \alpha)$  when  $\gamma < \frac{1}{2}$ .

**Remark 3.2 (Parametrization of the changepoint).** We assume that the changepoint is parametrized as  $\tau_n = n - cn^\gamma$ , and thus a late changepoint corresponds to  $\gamma \in (0, 1)$ . However, our proof of Proposition 3.1 works under the more general late changepoint assumption  $n - \tau_n = o(n)$ , and under the assumption that  $\delta_1 = \delta_1^{(n)}$  depends on  $n$ . In that case, the proof shows that, if

$$\frac{n - \tau_n}{\sqrt{n}} |\delta_0 - \delta_1^{(n)}| \rightarrow \infty,$$

then  $\mathbb{P}_1(\psi(G_n) = 0) \rightarrow 0$  as  $n \rightarrow \infty$ . Furthermore, the proof shows that, if

$$\frac{n - \tau_n}{\sqrt{n}} |\delta_0 - \delta_1^{(n)}| \rightarrow 0,$$

then the test  $\psi$  is powerless, i.e., the sum of type-I and type-II errors converges to one as  $n \rightarrow \infty$ . However, the proofs of later results rely explicitly on the choice  $\tau_n = n - cn^\gamma$ . Because of this, we have decided to maintain the parametrization  $\tau_n = n - cn^\gamma$  throughout the manuscript for consistency of presentation.

The proof of this result is based on the following observations. Under the null model it is known that  $\mathbb{E}_0(N_m(n)) - np_m(\delta_0) = O(1)$ . Under the alternative model we can show that

$$\mathbb{E}_1[N_m(n)] - np_m(\delta_0) = (1 + o(1))\eta(\delta_0, \delta_1, m, c)n^\gamma,$$

where

$$\eta(\delta_0, \delta_1, m, c) := c(1 - p_m(\delta_0)/p_m(\delta_1)). \tag{3}$$

Therefore there is a substantial difference in the expected values of the test statistic under the null and alternative models. Note that both  $\mathbb{E}_1[N_m(n)]$  and  $np_m(\delta_0)$  have the same order of magnitude  $O(n)$ , so the above result characterizes the second-order behavior of  $\mathbb{E}_1[N_m(n)]$  and thus is somewhat delicate. Besides characterizing the mean of  $N_m(n)$ , we must also characterize the fluctuations of  $N_m(n)$  around it. These are of small order, and controlled by a rather standard application of Azuma-Hoeffding’s inequality. Specifically  $N_m - \mathbb{E}[N_m(n)] = O_{\mathbb{P}}(\sqrt{n})$ . This result holds both under the null and alternative hypothesis, showing that it is possible to construct a powerful test when  $\gamma > \frac{1}{2}$ . When  $\gamma = \frac{1}{2}$  both the shift in the expected value  $N_m(n) - np_m(\delta_0)$  and its stochastic fluctuations have the same order, leading to the result in the theorem.

As the proposed test is powerless when  $\gamma < \frac{1}{2}$ , one might wonder if there is any test that can have power in that situation. In the pre-print of our work which appeared on arXiv we set forth the following conjecture:

**Conjecture 3.3 (Powerless tests when  $\gamma < \frac{1}{2}$ ).** Consider the case  $\gamma < \frac{1}{2}$ . We conjecture the following:

- (i) All tests based on the vertex degrees  $\{D_v(n)\}_{v \in [n]}$  are powerless.
- (ii) All tests based on  $G_n$  are powerless.

Obviously the statement (ii) implies (i). The main motivation for (i) is that when  $\gamma < \frac{1}{2}$  the number of vertices with degree  $k$  will deviate from  $np_k(\delta_0)$  by at most  $O(n^\gamma)$ . These deviations become smaller when  $k$  gets larger. On the other hand, the fluctuations of  $N_k(n)$  around its mean will also become smaller, but will always be of higher order than this mean-shift. Actually, the results in [3] characterize the joint distribution of the degree counts under the null model and show this is asymptotically a multivariate normal distribution. As shown in [7, Lemma 1.2], under the alternative model these degree counts are also asymptotically normal with the same covariance matrix, but a rather small mean-shift. If this shift is small enough, then the total variation distance between the two distributions is close to zero, implying (i). The conjecture (ii) is significantly stronger, stating that higher-level information contained on the edge structure of  $G_n$  will not be helpful for this testing problem. This is expected given that the attachment dynamics are only driven by the vertex degrees. In fact, from the moment the pre-print of our work appeared on arXiv, the stronger conjecture (ii) has been proven, first partially in [27], and then as stated above in [21], by building on [27].

### 3.2. Powerful test for unknown $\delta_0$

The knowledge of  $\delta_0$  was crucial for the test above, as it gives a benchmark to compare  $N_m(n)$  against, namely  $np_m(\delta_0)$ . Without this knowledge we must essentially estimate, from  $G_n$ , the value of  $\delta_0$ . For such an approach to be fruitful a candidate estimator must be “close enough” to  $\delta_0$  both under the null and alternative models. Gao and van der Vaart in [25] consider the problem of estimating  $\delta_0$  when the preferential attachment function is *constant*, meaning that we are under the null model in our formulation. The authors proposed a maximum likelihood estimator based on  $G_n$ , and showed it consistently estimates  $\delta_0$  and is asymptotically normal. A natural idea is to start by considering this estimator as well, and understand how its properties change when  $G_n$  is generated under the alternative model.

As done in [25], to avoid the usual issues at the boundary of the parameter space we make an extra assumption that the range of possible values for  $\delta_0$  and  $\delta_1$  is known:

**Assumption 3.4 (Containment of  $\delta_0, \delta_1$ ).** Let  $-m < \delta_{\min} \leq \delta_{\max}$  be known, and assume that  $\delta_0, \delta_1 \in (\delta_{\min}, \delta_{\max})$ .

As shown in [25], under the null model the (normalized) log-likelihood function  $\iota_n : [\delta_{\min}, \delta_{\max}] \rightarrow \mathbb{R}$  is given by

$$\begin{aligned} \iota_n(\delta) &:= \frac{1}{n+1} \left( \sum_{k=1}^{\infty} \log(k+\delta) (N_{>k}(n) - (n+1)\mathbb{1}\{k < m\}) - \sum_{t=2}^n \sum_{i=1}^m \log S_{t,i-1}(\delta) \right) \\ &= \frac{1}{n+1} \sum_{k=m}^{\infty} \log(k+\delta) N_{>k}(n) - \frac{1}{n+1} \sum_{t=2}^n \sum_{i=1}^m \log S_{t,i-1}(\delta), \end{aligned}$$

where  $S_{t,i-1}(\delta) := t\delta + 2m(t-1) + (i-1)$  and  $N_{>k}(n) := \sum_{j>k} N_j(n)$ . The maximum-likelihood estimator is defined as

$$\hat{\delta}_n := \arg \max_{\delta \in [\delta_{\min}, \delta_{\max}]} \iota_n(\delta). \tag{4}$$

Equivalently (for large  $n$ ) we can define  $\hat{\delta}_n$  as the solution in  $\delta \in [\delta_{\min}, \delta_{\max}]$  of  $\frac{\partial}{\partial \delta} \ell_n(\delta) := \ell'_n(\delta) = 0$ . Although not obvious, this definition coincides with (4) for large  $n$ , since it is shown in [25] that the solution of  $\ell'_n(\delta) = 0$  exists and is unique for large enough  $n$  with high probability. Note that the score function is given by

$$\ell'_n(\delta) = \frac{1}{n+1} \sum_{k=m}^{\infty} \frac{1}{k+\delta} N_{>k}(n) - \frac{1}{n+1} \sum_{t=2}^n \sum_{i=1}^m \frac{t}{S_{t,i-1}(\delta)}. \tag{5}$$

Motivated by this estimator we consider the test statistic

$$Q(G_n) := N_m(n) - np_m(\hat{\delta}_n). \tag{6}$$

This is analogous to the previously considered statistic, with the exception that  $\delta_0$  is replaced by the above estimator.

A test based on the above statistic will only be sensible if  $\hat{\delta}_n$  is a good surrogate for  $\delta_0$ , under both the null and alternative models. When  $\gamma < 1$  this is indeed the case and we show that  $\hat{\delta}_n$  is a consistent estimator of  $\delta_0$  under both the null and alternative models. However, consistency is not enough, and it is necessary to carefully characterize the rate of convergence of this estimator under the alternative model. It turns out that the deviations of  $N_m(n)$  and  $np_m(\hat{\delta}_n)$  around their respective means have exactly the same order, but with *different* leading constants. A careful characterization of those constants is the crucial result leading to our main result:

**Theorem 3.5 (Asymptotically powerful test, unknown  $\delta_0$ ).** *Consider Assumption 3.4 and the test statistic defined in (6). Let  $a_n$  be a positive diverging sequence such that  $a_n = \omega(\sqrt{n} \log n)$  and  $a_n = n^{\frac{1}{2}+o(1)}$  and define the test*

$$\phi(G_n) := \mathbb{1}\{|Q(G_n)| \geq a_n\}.$$

The type-I error of this test converges to zero as  $n \rightarrow \infty$ , i.e.,

$$\mathbb{P}_0(\phi(G_n) \neq 0) = o(1).$$

Furthermore, when  $\delta_0 \neq \delta_1$  and  $\frac{1}{2} < \gamma < 1$  this test has vanishing type II error, i.e., as  $n \rightarrow \infty$ ,

$$\mathbb{P}_1(\phi(G_n) = 0) = o(1).$$

The proof of the theorem is deferred to Section 8 and it is rather involved. It builds upon some of the results used to prove Proposition 3.1, namely the characterization of  $N_m(n) - np_m(\delta_0)$ . However, it does require a very careful characterization of  $np_m(\hat{\delta}_n) - np_m(\delta_0)$ . It turns out that both quantities have essentially the same order of magnitude, namely  $O_{\mathbb{P}}(n^\gamma)$ . However, the leading constants are different, and this fact allows the test to be powerful in the regime  $\frac{1}{2} < \gamma < 1$ .

The cases  $\gamma = \frac{1}{2}$  and  $\gamma = 1$  are special. For  $\gamma = \frac{1}{2}$  the Gaussian fluctuations of  $N_m(n)$  under the null and alternative hypotheses will compete with the resulting change in expectations of  $N_m(n)$ , so the Type-II error does not vanish, as it does for  $\gamma \in (\frac{1}{2}, 1)$ . For  $\gamma = 1$ , on the other hand, both the maximal and minimal degree tests seem to perform well, but it is unclear which performs best, or whether there even is a better test available.

### 3.3. Asymptotically calibrated tests

The two theorems above characterize the regime when the proposed tests are powerful. However, they fall short of providing guidelines to properly calibrate the tests. Particularly, in a fixed significance testing framework, for any  $\alpha \in (0, 1)$  we would like to ensure that under the null model the type-I error is approximately  $\alpha$ . Proposition 3.1 provides only an rather conservative asymptotic upper bound on the type-I error, due to the worst-case nature of the Azuma-Hoeffding inequality. To introduce our calibrated test, we define

$$w(\delta_0, m) := \frac{m^2(m + \delta_0)(1 + m + \delta_0)(2m + \delta_0)}{(\delta_0 + 2m(1 + m + \delta_0))(\delta_0 + m(2 + m + \delta_0))^2}. \tag{7}$$

Furthermore, let  $\Phi : \mathbb{R} \rightarrow [0, 1]$  denote the cumulative distribution function of a standard normal distribution and  $z_\alpha := \Phi^{-1}(1 - \alpha)$  denote the corresponding right-quantile function for  $\alpha \in (0, 1)$ . In particular,  $z_{\alpha/2} > 0$  when  $\alpha \in (0, 1)$ . We are now ready to present the asymptotically calibrated test when  $\delta_0$  is assumed known.

**Theorem 3.6 (Asymptotically calibrated test for known  $\delta_0$ ).** *Let  $\alpha \in (0, 1)$  and define the test*

$$\psi_{\text{cal}}(G_n) := \mathbb{1} \left\{ |T(G_n)| \geq z_{\alpha/2} \sqrt{nw(\delta_0, m)} \right\}.$$

As  $n \rightarrow \infty$ , the type-I error of this test converges to  $\alpha$ :

$$\mathbb{P}_0(\psi_{\text{cal}}(G_n) \neq 0) \rightarrow \alpha.$$

Furthermore, when  $\delta_0 \neq \delta_1$  the type-II error satisfies

$$\mathbb{P}_1(\psi_{\text{cal}}(G_n) \neq 1) \rightarrow \begin{cases} 0 & \text{if } \frac{1}{2} < \gamma < 1 \\ 1 - \Phi\left(z_{\alpha/2} - \frac{\eta(\delta_0, \delta_1, m, c)}{\sqrt{w(\delta_0, m)}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\eta(\delta_0, \delta_1, m, c)}{\sqrt{w(\delta_0, m)}}\right) & \text{if } \gamma = \frac{1}{2} \end{cases}$$

as  $n \rightarrow \infty$ , where  $\eta(\delta_0, \delta_1, m, c)$  is defined in (3).

To define the next test and result we require some additional notation. Let

$$v(\delta_0, m) := \sum_{k=m}^{\infty} \frac{mp_k(\delta_0)}{(k + \delta_0)(2m + \delta_0)} - \frac{m}{(2m + \delta_0)^2}, \tag{8}$$

$$u(\delta_0, m) := -\frac{m^4}{v(\delta_0, m)(\delta_0 + m(2 + m + \delta_0))^4}, \tag{9}$$

and

$$\alpha(\delta_0, \delta_1, m, c) := c(\delta_0 - \delta_1) \frac{m + \delta_0}{(2 + \delta_1/m)(m + \delta_0 + 2 + \delta_0/m)^2}. \tag{10}$$

The next theorem presents the asymptotically calibrated test for the unknown  $\delta_0$  case:

**Theorem 3.7 (Asymptotically calibrated test, unknown  $\delta_0$ ).** Consider Assumption 3.4, let  $\alpha \in (0, 1)$ , and define the test

$$\phi_{\text{cal}}(G_n) := \mathbb{1} \left\{ |Q(G_n)| \geq z_{\alpha/2} \sqrt{n(w(\hat{\delta}_n, m) + u(\hat{\delta}_n, m))} \right\}.$$

As  $n \rightarrow \infty$ , the type-I error of this test converges to  $\alpha$ :

$$\mathbb{P}_0(\phi_{\text{cal}}(G_n) \neq 0) \rightarrow \alpha.$$

Furthermore, when  $\delta_0 \neq \delta_1$  and  $n \rightarrow \infty$  the type-II error satisfies

$$\mathbb{P}_1(\phi_{\text{cal}}(G_n) \neq 1) \rightarrow \begin{cases} 0 & \text{if } \frac{1}{2} < \gamma < 1, \\ 1 - \Phi\left(z_{\alpha/2} - \frac{\alpha(\delta_0, \delta_1, m, c)}{\sqrt{w(\delta_0, m) + u(\delta_0, m)}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\alpha(\delta_0, \delta_1, m, c)}{\sqrt{w(\delta_0, m) + u(\delta_0, m)}}\right) & \text{if } \gamma = \frac{1}{2}. \end{cases}$$

The proofs of Theorems 3.6 and 3.7 are an immediate consequence of the asymptotic normality of the test statistics as discussed in the next section, together with the consistency of  $\hat{\delta}_n$  as an estimator of  $\delta_0$ .

### 4. Asymptotic normality of test statistics

In this section we characterize the asymptotic distribution of the proposed test statistics which allowed us to calibrate the tests in Section 3.3.

When  $\delta_0$  is known, the situation is relatively simple. Under the null model and when  $m = 1$ , it is known that  $N_m(n)$  with  $k \geq m$  admits a central limit theorem [34]. In particular, this shows that  $N_m(n)$  is asymptotically normally distributed. Furthermore, [3] extends these results to the general case  $m \geq 1$ , making them applicable to our setting. For the case of unknown  $\delta_0$  the situation is a bit more complicated. Recall that our test statistic is  $Q(G_n) = N_m(n) - np_m(\hat{\delta}_n)$ . It is known from [25] that, under the null model,  $\hat{\delta}_n$  is asymptotically normal. This does not, however, immediately imply that  $Q(G_n)$  is also asymptotically normal under the null model. The results in [3] establish that  $(N_m(n), N_{m+1}(n), \dots)$  is asymptotically normal (under the null model), strongly hinting at asymptotic normality of  $Q(G_n)$ . Furthermore, even under the alternative model, one might expect asymptotic normality of the test statistics, with exactly the same asymptotic variance, as the number of vertices that enter after the (late) changepoint is too small to change the asymptotic variance.

Let  $\mathcal{N}(\mu, \sigma^2)$  denote the normal distribution with mean  $\mu$  and variance  $\sigma^2$ , and let  $\xrightarrow{D}$  denote convergence in distribution. The following theorem, proved in [7, Section 1], establishes the asymptotic properties of the test statistics.

**Theorem 4.1 (Asymptotic normality of test statistics).** Recall the definitions of  $w$  and  $u$  in (7) and (9) respectively. Under the null model, as  $n \rightarrow \infty$ ,

$$\frac{T(G_n)}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, w(\delta_0, m)). \tag{11}$$

Moreover, under Assumption 3.4 and the null model, as  $n \rightarrow \infty$ ,

$$\frac{Q(G_n)}{\sqrt{n}} = \frac{N_m(n) - np_m(\hat{\delta}_n)}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, w(\delta_0, m) + u(\delta_0, m)). \tag{12}$$

Furthermore, under the alternative model with  $\gamma \in (0, 1)$ , as  $n \rightarrow \infty$ ,

$$\frac{T(G_n) - \mathbb{E}_1[T(G_n)]}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, w(\delta_0, m)). \tag{13}$$

Moreover, under Assumption 3.4, as  $n \rightarrow \infty$ ,

$$\frac{Q(G_n) - \mathbb{E}_1[Q(G_n)]}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, w(\delta_0, m) + u(\delta_0, m)). \tag{14}$$

Note that (11) has been proved in [3]. In the second statement (12),  $u(\delta_0, m)$  captures the adjustment in the variability in the test statistic by using an estimate of  $\delta_0$ , instead of the actual value. The convergence results (11) and (12) provide an avenue for asymptotically exact calibration of the proposed test as given in Theorems 3.6 and 3.7 (equivalently, for computation of asymptotically exact  $p$ -values). On the other hand, the convergence results (13) and (14) give an asymptotically exact quantification of the type-II error of the tests.

To establish asymptotic normality under the alternative model, we exploit the corresponding statements under the null model, together with a correction term quantifying the effects due to the presence of a changepoint (partially relying on the arguments in Proposition 3.1 and Theorem 3.5). Note that the asymptotic variances remain the same, whether one is considering the null or alternative model, as expected. The proof of Theorem 4.1 is carried out in [7].

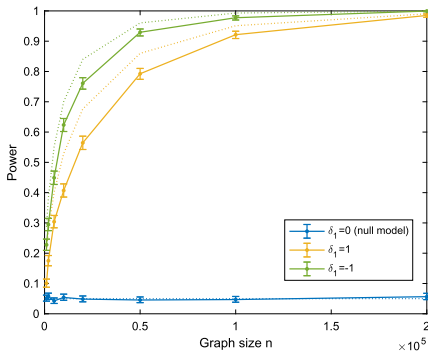
## 5. Numerical experiments

In this section we examine the properties of the proposed tests using simulation. This serves a two-fold purpose, namely to empirically assess the validity of the theoretical guarantees given, as well as to investigate the finite-sample properties of the tests.

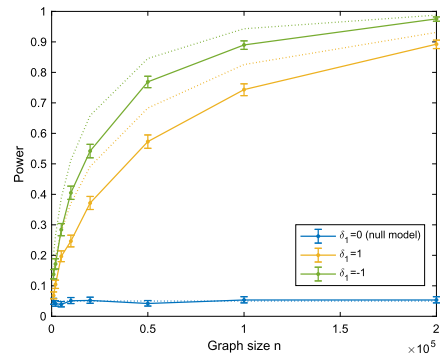
To assess the finite sample properties of the asymptotically calibrated tests we conduct several numerical experiments. For simplicity we fix  $\delta_0 = 0$  for all the experiments (this is the so-called linear preferential attachment model in [5]). We take  $m = 5$ ,  $c = 1$ ,  $\gamma = \frac{3}{4}$ , and  $\delta_1 \in \{-1, 0, 1\}$ . Note that  $\delta_1 = \delta_0 = 0$  corresponds to the null model. We consider graphs of different sizes, namely  $n \in \{1000, 2000, 5000, 10000, 20000, 50000, 100000, 200000\}$ , and for each value of  $n$  we generate  $B = 2000$  independent graphs using the preferential attachment model in (1). Specifically, for the three cases  $\delta_1 \in \{-1, 0, 1\}$  and each value  $n$  we obtain  $\{g_n^{(b)}\}_{b=1}^B$  graphs.

Note that as  $n$  increases the relative distance between the changepoint and  $n$  decreases considerably, and only a minute part of the graph is affected after the changepoint. For instance, for  $n = 5000$  there are only 594 vertices that join the graph after the changepoint, so about 12% of the vertices. However, for  $n = 200000$  only 9457 vertices join after the changepoint, a mere 4.7% of the total vertices.

In Figure 1 we depict the power of the two proposed tests. For concreteness we consider fixed significance testing at level  $\alpha = 0.05$  (qualitatively the results are similar for other significance levels). We clearly see that both tests are well calibrated, even for small values of  $n$ . Also, as expected, the power increases as a function of  $n$ . As intuitively expected, the test that does not assume knowledge of  $\delta_0$  has slightly less power than the test making use of that knowledge. Finally, there is an asymmetry of

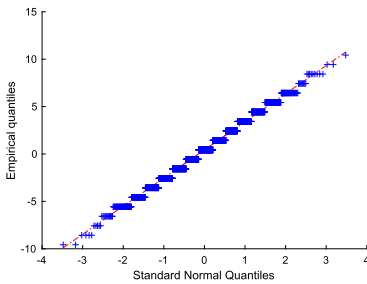


(a) Power of the tests (known  $\delta_0$  and  $\alpha = 0.05$ ).

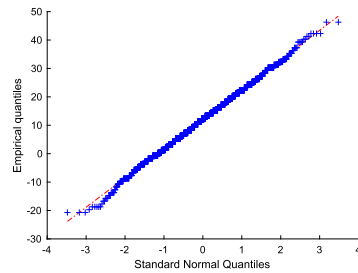


(b) Power of the tests (unknown  $\delta_0$  and  $\alpha = 0.05$ ).

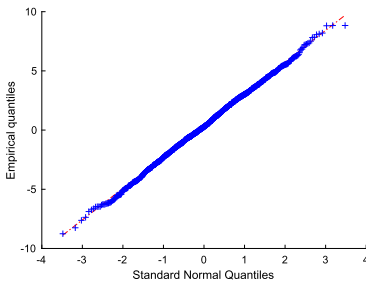
**Figure 1.** Power of the tests calibrated using asymptotic normality for  $m = 5$ . Panel (a) corresponds to the case of known  $\delta_0$  and panel (b) corresponds to the unknown  $\delta_0$  case. Both plots consider a fixed significance level testing with level  $\alpha = 0.05$ . The solid lines correspond to the power estimated using  $B = 2000$  graph samples, and (the pointwise 95% confidence bands are computed using the Clopper-Pearson exact method. The dashed lines are estimates of the power based on the asymptotic characterization of the test statistics.



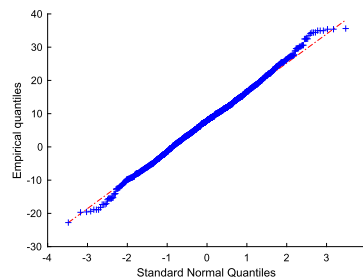
(a)  $T(G_n)$  with  $n = 100$  and  $\delta_0 = \delta_1 = 0$ .



(b)  $T(G_n)$  with  $n = 1000$  and  $\delta_0 = \delta_1 = 0$ .



(c)  $Q(G_n)$  with  $n = 100$  and  $\delta_0 = \delta_1 = 0$ .



(d)  $Q(G_n)$  with  $n = 1000$  and  $\delta_0 = \delta_1 = 0$ .

**Figure 2.** Normal QQ plots of the test statistics under the null model, for  $n \in \{100, 1000\}$ . Panels (a) and (b) correspond to  $T(G_n)$  (known  $\delta_0$  statistic) and panels (c) and (d) correspond to  $Q(G_n)$  (unknown  $\delta_0$  statistics). Even for very small values of  $n$  distribution of the statistics is very well approximated by a normal distribution.

**Table 1.** Numerical estimates of (rescaled) mean and variance of the test statistics, and comparison with corresponding asymptotic values. The estimates above are based on  $B = 2000$  independently generated graphs.

(a) Null model $\delta_1 = \delta_0 = 0$ .				
$n$	$\hat{\mu}_n^{(T)}$	$\hat{\mu}_n^{(Q)}$	$\hat{v}_n^{(T)}/n$	$\hat{v}_n^{(Q)}/n$
1000	0.0127	0.0105	0.1014	0.0801
2000	0.0017	0.0042	0.1016	0.0795
5000	0.0056	0.0041	0.1009	0.0789
10000	-0.0052	-0.0038	0.1049	0.0831
20000	0.0007	0.0012	0.1034	0.0828
50000	0.0049	-0.0009	0.0986	0.0780
100000	-0.0066	-0.0054	0.1000	0.0814
200000	0.0031	0.0046	0.1048	0.0811
	$\eta(\delta_0, \delta_0, m, c)$ = 0	$\alpha(\delta_0, \delta_0, m, c)$ = 0	$w(\delta_0, m)$ = 0.1020	$w(\delta_0, m) + u(\delta_0, m)$ = 0.0811

(b) Alternative model $\delta_0 = 0$ and $\delta_1 = 1$ .				
$n$	$\hat{\mu}_n^{(T)}/n^\gamma$	$\hat{\mu}_n^{(Q)}/n^\gamma$	$\hat{v}_n^{(T)}/n$	$\hat{v}_n^{(Q)}/n$
1000	-0.0460	-0.0276	0.0944	0.0772
2000	-0.0487	-0.0308	0.0992	0.0776
5000	-0.0554	-0.0371	0.1017	0.0826
10000	-0.0558	-0.0375	0.0994	0.0799
20000	-0.0574	-0.0395	0.1041	0.0799
50000	-0.0594	-0.0409	0.1033	0.0832
100000	-0.0610	-0.0421	0.1042	0.0809
200000	-0.0610	-0.0427	0.0974	0.0759
	$\eta(\delta_0, \delta_1, m, c)$ = -0.0649	$\alpha(\delta_0, \delta_1, m, c)$ = -0.0464	$w(\delta_0, m)$ = 0.1020	$w(\delta_0, m) + u(\delta_0, m)$ = 0.0811

(c) Alternative model $\delta_0 = 0$ and $\delta_1 = -1$ .				
$n$	$\hat{\mu}_n^{(T)}/n^\gamma$	$\hat{\mu}_n^{(Q)}/n^\gamma$	$\hat{v}_n^{(T)}/n$	$\hat{v}_n^{(Q)}/n$
1000	0.0672	0.0436	0.1087	0.0831
2000	0.0679	0.0443	0.1020	0.0760
5000	0.0700	0.0463	0.1033	0.0799
10000	0.0725	0.0491	0.0952	0.0759
20000	0.0713	0.0488	0.1049	0.0817
50000	0.0738	0.0509	0.1043	0.0833
100000	0.0743	0.0515	0.1057	0.0875
200000	0.0750	0.0522	0.0998	0.0793
	$\eta(\delta_0, \delta_1, m, c)$ = 0.0794	$\alpha(\delta_0, \delta_1, m, c)$ = 0.0567	$w(\delta_0, m)$ = 0.1020	$w(\delta_0, m) + u(\delta_0, m)$ = 0.0811

the power depending on whether  $\delta_1 > \delta_0$  or  $\delta_1 < \delta_0$ , the later scenario leading to higher power. This is expected, as for smaller  $\delta_1$  values the empirical degree distribution has heavier tails, and detecting the presence of a changepoint becomes easier.

To illustrate the results of Theorem 4.1, in Figure 2 we display normal quantile-quantile plots of the test statistics for very small values of  $n$ . We see that, even for  $n = 100$  (so, a graph with  $nm = 500$  edges) both test statistics are approximately normally distributed. In order to have a better assessment of the mean and variance of the test statistics we compute (for  $\delta_1 \in \{-1, 0, 1\}$ ) the first two empirical moments, and compare them with their asymptotic counterparts, both under the null and alternative hypothesis. Define

$$\hat{\mu}_n^{(T)} := \frac{1}{B} \sum_{b=1}^B T(g_n^{(b)}), \quad \text{and} \quad \hat{v}_n^{(T)} := \frac{1}{B} \sum_{b=1}^B (T(g_n^{(b)}) - \hat{\mu}_n^{(T)})^2,$$

and the analogous definitions of  $\hat{\mu}_n^{(Q)}$  and  $\hat{v}_n^{(Q)}$ . In Table 1 we compare these values (adequately rescaled) with the expected asymptotic values. As can be seen, the empirical variance closely matches the asymptotic variance, both under the null and alternative models, even for small values of  $n$ . This is in agreement with Theorem 4.1. For the mean of the statistics, one sees that the scaling by  $n^\gamma$  and the corresponding leading constant is also accurate, but finite sample effects are more evident under the alternative model when  $n$  is small.

Finally, making use of Theorem 4.1 we can compare the empirical power with an estimate based on the asymptotic normality of the statistic. Namely, we know the test statistics are asymptotically normal with exactly the same variance, and a small mean-shift proportional to  $n^\gamma(1 + o(1))$ , where the proportionality constant is given by  $\eta(\delta_0, \delta_1, m, c)$  from (3) and  $\alpha(\delta_0, \delta_1, m, c)$  from (10) for  $T(G_n)$  and  $Q(G_n)$  respectively. Based on this, one can get an estimate for the power of the tests for different values of  $\delta_1$ . This is shown by the dashed lines in Figure 1. As one can see, although not terribly accurate, the estimates capture the exact behavior of the empirically observed power. This lack of accuracy is not unexpected, as all we know is that the mean-shifts are of the form  $\text{const}(\delta_0, \delta_1, m)n^\gamma + o(n^\gamma)$ . However, the remainder term might still have an order only slightly smaller than  $n^\gamma$ , which will lead to rather poor power estimates for small values of  $n$ .

## 6. Discussion and open problems

In this section, we compare our results to the literature and state some open problems.

### 6.1. Early changepoint

In previous work [4,8,16], the case of an *early* changepoint was considered for preferential attachment trees, i.e., for  $m = 1$ . Thus, our work extends this setting from trees to graphs, as well as from an early changepoint to a late one. Arguably, the latter case is more relevant in practice, since one would rather detect a changepoint quickly, meaning, close to the time after which it occurs. This setting corresponds to a changepoint close to the time of observation of the final network.

### 6.2. Dynamical graph observations

It would be of interest to extend our results to a *dynamic* setting, where we detect the changepoint as the graph changes dynamically. Bear in mind though that we currently only assume that we observe the graph

at the final time, and observing the graph dynamically thus provides much more information. Thus, it is an interesting extension to devise an appropriate setting where we only observe *partial information* on the network, while still detecting the changepoint dynamically. There are several settings that could be of interest. In the first, one observes the network snapshots only at multiples of  $n^\gamma$ . In the second, we assume that we only dynamically observe information about the degree counts, and not the entire network. We defer such problems to future work.

### 6.3. Lower bounds

Conjecture 3.3 states that no test will be powerful when  $\gamma < \frac{1}{2}$ . Proving such lower bounds in the context of preferential attachment models is challenging, due to the latent nature of these models. However, part (i) of the conjecture might be approached by relying on the asymptotic normality characterization in [7, Lemma 1.2], together with bounds on  $\mathbb{E}_1[N_k(n)] - \mathbb{E}_0[N_k(n)]$  obtained using the methods developed in this paper.

### 6.4. Boundary case $\gamma = 1$

Note that in Proposition 3.1 and Theorem 3.5 the case  $\gamma = 1$  is excluded. This is in contrast with the results in [4,8,16]. The proof of the two theorems relies on Proposition 7.1, that explicitly excludes the case  $\gamma = 1$ . It should be possible to extend that result for  $\gamma = 1$ . Specifically, the current argument quantifies the contribution to  $\mathbb{E}_1[N_m(n)] - np_m(\delta_0)$  made by vertices that arrived after the changepoint. Due to the late changepoint, most of those vertices will have degree  $m$  in  $G_n$ . However, for  $\gamma = 1$ , a small, but non-vanishing fraction of those vertices will have higher degree. Therefore, to extend the result, this needs to be quantified, and a slightly more refined argument will be needed, where the role of the parameter  $c$  will become much more prevalent. Extending the result of Theorem 3.5 to this setting will likely be significantly more challenging, as it requires extending Proposition 8.2, where the assumption that  $\gamma < 1$  was crucially used to bound the terms of order  $o(n^\gamma)$ . On the other hand, when  $\gamma = 1$ , our test statistic might not be a good practical choice (particularly when  $c$  is large), and the statistics used in [4,8] will possibly lead to more powerful tests. This remains an interesting avenue of future research.

### 6.5. Other test statistics

The recent work [21] proves that, when  $n - \tau_n = o(\sqrt{n})$ , all tests are *asymptotically* powerless. This implies that our test based on  $N_m(n)$  is asymptotically powerful up to the detectability threshold. However, information about the presence of a changepoint is also present in other statistics of the graph, which might not even be measurable with respect to the degree sequence, so it cannot be excluded that other tests have better performance than ours for *finite*  $n$ . For example, the number of lowest-degree neighbours of lowest-degree vertices, which is not measurable with respect to the degree sequence, carries information on the changepoint, a fact that was verified via numerical simulation. Another, more straightforward, possibility is including higher degree counts to improve the finite- $n$  performance of the test. Following this idea, the master thesis [20] numerically investigates the finite- $n$  performance of a likelihood-ratio test based on  $(N_m(n), N_{m+1}(n))$  and proves that it is slightly better than the performance of our test based solely on  $N_m(n)$ . An interesting avenue of research is to identify, in a principled way, statistics that lead to tests that have higher power than the ones proposed.

### 6.6. Boundary case $\gamma = \frac{1}{2}$

When  $\gamma = \frac{1}{2}$ , the fluctuations of  $N_m(n)$  around its mean under  $H_0$  are of the same order as  $\mathbb{E}_1[N_m(n)] - p_m(\delta_0)$ . Moreover, since a central limit theorem holds for  $N_m(n) - \mathbb{E}_1[N_m(n)]$  under  $H_1$ , with the same limiting variance as under  $H_0$  (cf. Theorem 4.1), it follows that, when  $\gamma = \frac{1}{2}$ , the type-II error of our test is strictly bounded away from zero. In other words, when  $\gamma = \frac{1}{2}$ , a large value of  $N_m(n) - p_m$  can be explained either by a large deviation away from  $\mathbb{E}_0[N_m]$  under  $H_0$ , or by a deviation around  $\mathbb{E}_1[N_m]$  under  $H_1$ . Although this setting might be perceived as too specific to be of practical interest, it does provide an avenue to compare different testing procedures from an asymptotic lens, and can yield valuable insights regarding the finite-sample performance of different tests.

## 7. Powerful test for known $\delta_0$ : Proof of Proposition 3.1

The main idea of the proof is to decompose the test statistic  $T(G_n)$  in two terms:

$$T(G_n) = N_m(n) - np_m(\delta_0) = \underbrace{\mathbb{E}_\ell[N_m(n)] - np_m(\delta_0)}_{:=A} + \underbrace{N_m(n) - \mathbb{E}_\ell[N_m(n)]}_{:=B}, \tag{15}$$

where  $\ell \in \{0, 1\}$ . The characterization of the stochastic term  $B$  is the same under both the null and alternative models, and follows a somewhat standard argument. Let  $\ell \in \{0, 1\}$  be fixed. Define the stochastic process  $\{M_t\}_{t=1}^n$  such that

$$M_t := \mathbb{E}_\ell[N_m(n)|G_t].$$

This is a Doob martingale [35, Lemma 8.5], and such that  $\mathbb{E}_\ell[N_m(n)] = M_1$  and  $N_m(n) = M_n$ . Furthermore, by [35, Lemma 8.6], we know that, for every  $t \in \{2, \dots, n\}$ ,  $|M_n - M_{n-1}| \leq 2m$  almost surely. Although strictly speaking these two lemmas were not stated for our model, their arguments do not depend on the specific sequence  $\delta(t)$  in (1). Therefore, those results still hold and follow simply from the fact that every step in time we add precisely  $m$  edges to the graph. With this in hand, we can directly apply the Azuma-Hoeffding inequality to get that, for any  $x > 0$ ,

$$\forall \ell \in \{0, 1\} \quad \mathbb{P}_\ell (|N_m(n) - \mathbb{E}_\ell[N_m(n)]| \geq x) \leq 2e^{-\frac{x^2}{8m^2n}}. \tag{16}$$

This completes the characterization of the term  $B$  in (15). We now proceed by considering the type-I and type-II errors separately.

### 7.1. Type-I error

To control the term  $A$  in (15) under the null model we use [18, Proposition 2.2] (see also [35, Proposition 8.7]), which states that there exists a constant  $C_0 = C_0(\delta_0, m)$  such that, for all  $n \geq 1$ ,

$$|\mathbb{E}_0[N_m(n)] - np_m(\delta_0)| \leq C_0. \tag{17}$$

Combining this with (16) we see that the type-I error of the minimal degree test is bounded by

$$\begin{aligned} \mathbb{P}_0(\psi(T(G_n)) \neq 0) &= \mathbb{P}_0\left(|N_m(n) - np_m(\delta_0)| \geq m\sqrt{8n \log(2/\alpha)}\right) \\ &\leq \mathbb{P}_0\left(|N_m(n) - \mathbb{E}_0[N_m(n)]| \geq m\sqrt{8n \log(2/\alpha)} - C_0\right) \\ &\leq 2 \exp\left(-\frac{(m\sqrt{8n \log(2/\alpha)} - C_0)^2}{8m^2n}\right) = (1 + o(1))\alpha. \end{aligned}$$

This shows that the type-I error is essentially at most  $\alpha$ , completing the first part of the proof.

### 7.2. Type-II error

Again we must control the term  $A$  above, for which the following proposition is instrumental:

**Proposition 7.1.** *Let  $0 < \gamma < 1$ . Then*

$$\begin{aligned} \mathbb{E}_1[N_m(n)] - np_m(\delta_0) &= (1 + o(1)) cn^\gamma \left(1 - \frac{p_m(\delta_0)}{p_m(\delta_1)}\right) \\ &= (1 + o(1))cn^\gamma(\delta_0 - \delta_1) \frac{1}{(2 + \delta_1/m)(m + \delta_0 + 2 + \delta_0/m)}. \end{aligned}$$

**Proof.** Note that

$$\begin{aligned} \mathbb{E}_1[N_m(n)] - np_m(\delta_0) &= \mathbb{E}_0[N_m(n)] - np_m(\delta_0) \\ &\quad + \mathbb{E}_1[N_m(n)] - \mathbb{E}_1[N_m(\tau_n)] - (\mathbb{E}_0[N_m(n)] - \mathbb{E}_0[N_m(\tau_n)]). \end{aligned} \tag{18}$$

The equality holds as the law of  $G_{\tau_n}$  is the same under the null and alternative models, and therefore  $\mathbb{E}_0[N_m(\tau_n)] = \mathbb{E}_1[N_m(\tau_n)]$ . The last two terms are controlled in a similar way.

Let  $\ell \in \{0, 1\}$  and note that

$$\begin{aligned} \mathbb{E}_\ell[N_m(n)] - \mathbb{E}_\ell[N_m(\tau_n)] &= \sum_{v \in [\tau_n]} (\mathbb{P}_\ell(D_v(n) = m) - \mathbb{P}_\ell(D_v(\tau_n) = m)) \\ &\quad + \sum_{v \in [n] \setminus [\tau_n]} \mathbb{P}_\ell(D_v(n) = m). \end{aligned}$$

In the above there is a contribution from vertices that arrived before the change-point (so-called “old” vertices), and vertices that arrived afterwards (the “new” vertices). The contribution by the new vertices  $v \in [n] \setminus [\tau_n]$  is essentially the same regardless of the value of  $\ell$ , as  $\mathbb{P}_\ell(D_v(n) = m) \approx 1$  when  $v \in [n] \setminus [\tau_n]$ . To see this, note that  $\mathbb{P}_\ell(D_v(n) = m) = 1 - \mathbb{P}_\ell(D_v(n) > m)$  and that the event  $D_v(n) > m$  can only occur if there is at least one vertex  $v' > v$  attaching to  $v$ . Referring to (2) the probability of this happening is at most  $(m + \delta_\ell)/((2m + \delta_\ell)\tau_n - 2m)$ , and there are at most  $m(n - \tau_n)$  possible edges that can lead to that connection. Therefore

$$\mathbb{P}_\ell(D_v(n) = m) \geq 1 - m(n - \tau_n) \frac{m + \delta_\ell}{(2m + \delta_\ell)\tau_n - 2m} = 1 - O(n^{\gamma-1}).$$

In conclusion (since  $\mathbb{P}_\ell(D_v(n) = m)$  is bounded above by 1)

$$\sum_{v \in [n] \setminus [\tau_n]} \mathbb{P}_\ell(D_v(n) = m) = c(n - \tau_n)(1 + O(n^{\gamma-1})), \quad \ell \in \{0, 1\}.$$

For the term involving the “old” vertices we use the following lemma, which will be used to the full extent for the proof of Theorem 3.5:

**Lemma 7.2.** *Let  $v \in [\tau_n]$ ,  $\gamma < 1$  and  $m \leq k = o(n^{1-\gamma})$  and  $\ell \in \{0, 1\}$ . Then*

$$\mathbb{P}_\ell(D_v(n) - D_v(\tau_n) > 0 \mid D_v(\tau_n) = k) = (1 + o(1))cn^{\gamma-1}m \frac{k + \delta_\ell}{2m + \delta_\ell},$$

as  $n \rightarrow \infty$ .

**Proof.** Note that

$$\begin{aligned} & \mathbb{P}_\ell(D_v(n) - D_v(\tau_n) > 0 \mid D_v(\tau_n) = k) \\ &= 1 - \mathbb{P}_\ell(D_v(n) - D_v(\tau_n) = 0 \mid D_v(\tau_n) = k) \\ &= 1 - \prod_{j \in [n] \setminus [\tau_n]} \prod_{i=1}^m \left(1 - \frac{k + \delta_\ell}{j(2m + \delta_\ell) - 2m + i - 1}\right) \\ &= 1 - \exp\left(\sum_{j \in [n] \setminus [\tau_n]} \sum_{i=1}^m \log\left(1 - \frac{k + \delta_\ell}{j(2m + \delta_\ell) - 2m + i - 1}\right)\right) \\ &= 1 - \exp\left(\sum_{j \in [n] \setminus [\tau_n]} \sum_{i=1}^m -(1 + o(1)) \frac{k + \delta_\ell}{j(2m + \delta_\ell) - 2m + i - 1}\right) \tag{19} \end{aligned}$$

$$\begin{aligned} &= 1 - \exp\left(-(1 + o(1))cmn^\gamma \frac{k + \delta_\ell}{n(2m + \delta_\ell)}\right) \\ &= 1 - \left(1 - (1 + o(1))cmn^{\gamma-1} \frac{k + \delta_\ell}{2m + \delta_\ell}\right) \tag{20} \\ &= (1 + o(1))cmn^{\gamma-1} \frac{k + \delta_\ell}{2m + \delta_\ell}, \end{aligned}$$

where in (19) we relied on the fact that  $k = o(n)$  and for (20) it is crucial that  $k = o(n^{1-\gamma})$ . □

With this lemma in hand, we clearly see that

$$\begin{aligned} & \mathbb{P}_\ell(D_v(n) = m) - \mathbb{P}_\ell(D_v(\tau_n) = m) \\ &= \mathbb{P}_\ell(D_v(n) - D_v(\tau_n) = 0 \mid D_v(\tau_n) = m) \mathbb{P}_\ell(D_v(\tau_n) = m) \\ &= (1 - \mathbb{P}_\ell(D_v(n) - D_v(\tau_n) > 0 \mid D_v(\tau_n) = m)) \mathbb{P}_\ell(D_v(\tau_n) = m) \\ &= \left(1 - (1 + o(1))cn^{\gamma-1}m \frac{m + \delta_\ell}{2m + \delta_\ell}\right) \mathbb{P}_\ell(D_v(\tau_n) = m). \end{aligned}$$

Putting the two results together we get

$$\begin{aligned} & \mathbb{E}_\ell[N_m(n)] - \mathbb{E}_\ell[N_m(\tau_n)] \\ &= \left(1 - (1 + o(1))cn^{\gamma-1}m \frac{m + \delta_\ell}{2m + \delta_\ell}\right) \mathbb{E}_\ell[N_m(\tau_n)] + c(n - \tau_n)(1 + O(n^{\gamma-1})). \end{aligned}$$

Note that this results characterizes what happens both under the null and alternative models. Hence this result, together with (18) and the fact that  $\mathbb{E}_0[N_m(\tau_n)] = \mathbb{E}_1[N_m(\tau_n)]$ , yields

$$\begin{aligned} & \mathbb{E}_1[N_m(n)] - np_m(\delta_0) \\ &= \mathbb{E}_0[N_m(n)] - np_m(\delta_0) \\ & \quad + (1 + o(1))cn^{\gamma-1}m \left(\frac{m + \delta_0}{2m + \delta_0} - \frac{m + \delta_1}{2m + \delta_1}\right) \mathbb{E}_0[N_m(\tau_n)] + O(n^{2\gamma-1}) \\ &= O(1) + (1 + o(1))cn^{\gamma-1}(\delta_0 - \delta_1) \frac{1}{(2 + \delta_1/m)(2 + \delta_0/m)} \tau_n (p_m(\delta_0) + O(1/n)) \\ & \quad + O(n^{2\gamma-1}) \\ &= (1 + o(1))cn^\gamma \frac{\delta_0 - \delta_1}{(2 + \delta_1/m)(m + \delta_0 + 2 + \delta_0/m)} = (1 + o(1)) cn^\gamma \left(1 - \frac{p_m(\delta_0)}{p_m(\delta_1)}\right), \end{aligned}$$

where  $0 < \gamma < 1$ , and we have again used (17) to relate  $\mathbb{E}_0[N_m(\tau_n)]$  to  $\tau_n p_m(\delta_0)$ . □

Similarly as for the type-I error, we can use Proposition 7.1 together with the Azuma-Hoeffding inequality (16) to get

$$\begin{aligned} & \mathbb{P}_1(T(G_n) \neq 1) \\ &= \mathbb{P}_1\left(|N_m(n) - np_m(\delta_0)| < m\sqrt{8n \log(2/\alpha)}\right) \\ &\leq \mathbb{P}_1\left(|N_m(n) - \mathbb{E}_1[N_m(n)]| > \left(|\mathbb{E}_1[N_m(n)] - np_m(\delta_0)| - m\sqrt{8n \log(2/\alpha)}\right) \vee 0\right) \\ &\leq 2 \exp\left(-\frac{\left(\left(|\mathbb{E}_1[N_m(n)] - np_m(\delta_0)| - m\sqrt{8n \log(2/\alpha)}\right) \vee 0\right)^2}{8m^2n}\right). \end{aligned}$$

Considering the cases  $\gamma > \frac{1}{2}$  and  $\gamma = \frac{1}{2}$  separately, this gives

$$\mathbb{P}_1(T(G_n) \neq 1) \leq \begin{cases} o(1) & \text{when } \gamma > \frac{1}{2}, \\ (2 + o(1)) \exp\left(-\left(\frac{c|1 - p_m(\delta_0)/p_m(\delta_1)|}{m\sqrt{8}} - \sqrt{\log(2/\alpha)}\right) \vee 0\right)^2 & \text{when } \gamma = \frac{1}{2}. \end{cases}$$

The case  $\gamma < \frac{1}{2}$  does not follow immediately from the analysis above, as the characterization obtained by the Azuma-Hoeffding only provides an upper bound on the variability of the test statistic. However, in Theorem 4.1 it is shown that  $(T(G_n) - \mathbb{E}_1[T(G_n)])/\sqrt{n}$  is asymptotically normal with mean 0 and

variance  $w(\delta_0, m) > 0$ . As shown above  $\mathbb{E}_1[T(G_n)]/\sqrt{n} = O(n^{\gamma - \frac{1}{2}})$ . Therefore  $\mathbb{E}_1[T(G_n)]/\sqrt{n} = o(1)$  when  $\gamma < \frac{1}{2}$ , and therefore  $T(G_n)/\sqrt{n}$  converges to the same distribution under the null and alternative models, meaning the type II error is asymptotically just the complement of the type I error.  $\square$

### 8. Powerful test for unknown $\delta_0$ : Proof of Theorem 3.5

To prove Theorem 3.5 we partially leverage on the results and analysis in Proposition 3.1. However, we must take into account that  $\delta_0$  is not known, and rather we have only its estimate  $\hat{\delta}_n$ . The main idea is to decompose the test statistic  $Q(G_n)$  as

$$\begin{aligned}
 Q(G_n) &= N_m(n) - np_m(\hat{\delta}_n) && (21) \\
 &= \underbrace{\mathbb{E}_\ell[N_m(n)] - np_m(\delta_0)}_{:=A} + \underbrace{np_m(\delta_0) - np_m(\tilde{\delta}_n)}_{:=B} \\
 &\quad + \underbrace{N_m(n) - \mathbb{E}_\ell[N_m(n)] + np_m(\tilde{\delta}_n) - np_m(\hat{\delta}_n)}_{:=C},
 \end{aligned}$$

where  $\ell \in \{0, 1\}$ . In the above  $\tilde{\delta}_n$  is a deterministic quantity, and it is formally defined below. At this moment one might simply think of it as a ‘‘population’’ version of  $\hat{\delta}_n$ .

Clearly  $A$  is already characterized by Proposition 7.1. The bulk of the argument needed to prove Proposition 3.1 is in the characterization of  $B$ , the second term. For this we need to understand how fast  $\tilde{\delta}_n$  converges to  $\delta_0$  as  $n \rightarrow \infty$ . It turns out that under the alternative model both  $A$  and  $B$  have the same first-order asymptotic behavior, but with different leading constants. This fact is crucial to ensure that the test is powerful. Finally  $C$ , the last term, appears complicated but it is very well concentrated around zero.

Define  $\tilde{\delta}_n$  as a solution of  $\mathbb{E}_\ell[l'_n(\delta)] = 0$  in  $\delta \in [\delta_{\min}, \delta_{\max}]$  (if more than one solution exists, then we choose an arbitrary one). Most of the analysis focuses on the alternative model, but the stated results apply to the null model simply by taking  $\delta_1 = \delta_0$ . The following proposition gives a characterization of  $\tilde{\delta}_n$  and the term  $B$  above:

**Proposition 8.1.** *Let  $\gamma < 1$ . Under the alternative model  $\tilde{\delta}_n$  converges to  $\delta_0$  as  $n \rightarrow \infty$ . Furthermore,*

$$\tilde{\delta}_n - \delta_0 = (1 + o(1))n^{\gamma-1}c(\delta_1 - \delta_0)\frac{2m + \delta_0}{2m + \delta_1},$$

and

$$n(p_m(\delta_0) - p_m(\tilde{\delta}_n)) = (1 + o(1))cn^\gamma \frac{1}{(m + \delta_0 + 2 + \delta_0/m)^2}(\delta_1 - \delta_0)\frac{2m + \delta_0}{2m + \delta_1}.$$

The proof of this result is rather involved, due to the implicit nature of the definition of  $\tilde{\delta}_n$ . Note, however, that  $l'_n(\delta)$  is a score function, therefore we have immediately that  $\mathbb{E}_0[l'_n(\delta_0)] = 0$ . Under the alternative model  $\tilde{\delta}_n$  will not be equal to  $\delta_0$  and quantifying this deviation is crucial to our analysis. The following technical result, examining the difference between  $\mathbb{E}_1[l'_n(\delta)]$  and  $\mathbb{E}_0[l'_n(\delta)]$ , is instrumental:

**Proposition 8.2 (Analysis of differences of means).** *Let  $\gamma < 1$ . As  $n \rightarrow \infty$ , and uniformly for every  $\delta \in [\delta_{\min}, \delta_{\max}]$ ,*

$$(n + 1) (\mathbb{E}_1[l'_n(\delta)] - \mathbb{E}_0[l'_n(\delta)]) = \kappa(\delta_1, \delta_0, \delta)n^\gamma(1 + o(1)),$$

where  $\kappa(\delta_1, \delta_0, \delta)$  equals

$$\kappa := \kappa(\delta_1, \delta_0, \delta) = (\delta_1 - \delta_0) \frac{cm}{(2m + \delta_1)(2m + \delta_0)} \left( -1 + \sum_{k \geq m} \frac{2m + \delta}{k + \delta} p_k(\delta_0) \right).$$

The proof of this result, which is long and rather technical, is given in [7]. With this result in hand we are ready to prove Proposition 8.1:

**Proof of Proposition 8.1.** We first establish that  $\tilde{\delta}_n$  is well defined and that  $\tilde{\delta}_n \rightarrow \delta_0$ . Define

$$t'(\delta) := \sum_{k \geq m} \frac{p_{>k}(\delta_0)}{k + \delta} - \frac{1}{2 + \delta/m},$$

where  $p_{>k}(\delta_0) = \sum_{j>k} p_j(\delta_0)$ . Intuitively, this should be the limit of  $t'_n(\delta)$  as  $n \rightarrow \infty$ , as we see next. Note also that we can rewrite this expression in terms of  $p_k$ , by noticing (see [25, Lemma 2]) that, for  $k \geq m$

$$p_{>k}(\delta_0) = \frac{k + \delta_0}{2 + \delta_0/m} p_k(\delta_0).$$

Therefore,

$$t'(\delta) = \frac{1}{2 + \delta_0/m} \sum_{k \geq m} \frac{k + \delta_0}{k + \delta} p_k(\delta_0) - \frac{1}{2 + \delta/m}.$$

[25, Lemma 6] shows essentially that, as  $n \rightarrow \infty$ ,

$$\sup_{\delta \in [\delta_{\min}, \delta_{\max}]} |t'_n(\delta) - t'(\delta)| \xrightarrow{\mathbb{P}_0} 0.$$

[25, Lemma 6] is actually stated for a more general setting, where the number of edges added at each step is *random*. However, if the support of the distribution of the number of edges is bounded below by  $m$ , then following the steps in the proof of the lemma leads to the above result. Actually, in our setting we can state a slightly stronger result, namely convergence in  $L_1$ . First note that both  $|t'_n(\delta)|$  and  $|t'(\delta)|$  are bounded, uniformly in  $\delta$  and  $n$ . To see this note that, regardless of the value  $k$ ,

$$kN_{>k}(n) = k \sum_{\ell > k} N_\ell(n) \leq \sum_{\ell > k} \ell N_\ell(n) \leq \sum_{\ell \geq m} \ell N_\ell(n) = 2nm.$$

Therefore,

$$\begin{aligned} 0 &\leq \frac{1}{n+1} \sum_{k=m}^{\infty} \frac{1}{k + \delta} N_{>k}(n) \leq \frac{1}{n+1} \sum_{k=m}^{\infty} \frac{1}{k(k + \delta)} 2nm \\ &\leq \frac{2nm}{n+1} \sum_{k=m}^{\infty} \frac{1}{k(k + \delta_{\min})} \leq m \sum_{k=m}^{\infty} \frac{1}{k(k + \delta_{\min})} \\ &\leq m \left( \frac{1}{m(m + \delta_{\min})} + \sum_{k=1}^{\infty} \frac{1}{k^2} \right) = m \left( \frac{1}{m(m + \delta_{\min})} + \pi^2/6 \right). \end{aligned}$$

On the other hand,

$$\begin{aligned}
 0 &\leq \frac{1}{n+1} \sum_{t=2}^n \sum_{i=1}^m \frac{t}{t\delta + 2m(t-1) + (i-1)} \leq \frac{m}{n+1} \sum_{t=2}^n \frac{t}{t\delta_{\min} + 2m(t-1)} \\
 &\leq \frac{m(n-1)}{n+1} \frac{2}{2(m + \delta_{\min})} \leq \frac{2m}{2(m + \delta_{\min})}.
 \end{aligned}$$

These two results together imply that  $|\iota'_n(\delta)|$  is almost surely uniformly bounded for all  $\delta \in [\delta_{\min}, \delta_{\max}]$ . For  $\iota'$  we must simply note that this is a continuous function defined in the compact set  $[\delta_{\min}, \delta_{\max}]$ , and is therefore bounded. In conclusion, as  $n \rightarrow \infty$ ,

$$\mathbb{E}_0 \left[ \sup_{\delta \in [\delta_{\min}, \delta_{\max}]} |\iota'_n(\delta) - \iota'(\delta)| \right] \rightarrow 0.$$

This result, together with Proposition 8.2, shows that this is also true under the alternative model, and therefore, as  $n \rightarrow \infty$ ,

$$\forall \ell \in \{0, 1\} \quad \mathbb{E}_\ell \left[ \sup_{\delta \in [\delta_{\min}, \delta_{\max}]} |\iota'_n(\delta) - \iota'(\delta)| \right] \rightarrow 0. \tag{22}$$

By definition  $\mathbb{E}_1[\iota'_n(\tilde{\delta}_n)] = 0$ , therefore we conclude that  $\iota'(\tilde{\delta}_n) \rightarrow 0$ . [25, Lemma 4] shows that  $\iota'$  has a unique zero at  $\delta_0$ , and  $\iota'(\delta) > 0$  for  $\delta < \delta_0$  and  $\iota'(\delta) < 0$  for  $\delta > \delta_0$ . This immediately implies that  $\tilde{\delta}_n \rightarrow \delta_0$  as  $n \rightarrow \infty$ , proving the first assertion in the proposition.

To quantify the speed of convergence note first that  $\mathbb{E}_0[\iota'_n(\delta_0)] = 0$ , since  $\iota'_n$  is a score function. Furthermore, by definition  $\mathbb{E}_1[\iota'_n(\tilde{\delta}_n)] = 0$ , therefore Proposition 8.2 implies that

$$\begin{aligned}
 0 &= \mathbb{E}_1[\iota'_n(\tilde{\delta}_n)] - \mathbb{E}_0[\iota'_n(\tilde{\delta}_n)] + \mathbb{E}_0[\iota'_n(\tilde{\delta}_n)] \\
 &= \kappa(\delta_1, \delta_0, \tilde{\delta}_n)n^{\gamma-1}(1 + o(1)) + \mathbb{E}_0[\iota'_n(\delta_0) + \iota''_n(\tilde{\delta}_n)(\tilde{\delta}_n - \delta_0)] \\
 &= \kappa(\delta_1, \delta_0, \tilde{\delta}_n)n^{\gamma-1}(1 + o(1)) + \mathbb{E}_0[\iota''_n(\tilde{\delta}_n)](\tilde{\delta}_n - \delta_0),
 \end{aligned} \tag{23}$$

where  $|\tilde{\delta}_n - \delta_0| \leq |\tilde{\delta}_n - \delta_0|$ . Clearly  $\tilde{\delta}_n \rightarrow \delta_0$ . As shown in [25],  $\iota''_n(\delta)$  also converges in probability to  $\iota''(\delta)$  uniformly in  $\delta \in [\delta_{\min}, \delta_{\max}]$ . In fact, convergence holds also in  $L_1$ , since  $\iota''_n(\delta)$  is uniformly bounded (following the type of argument used before showing that  $\iota'_n$  is uniformly bounded). In addition, it is also shown in [25] that  $\iota''(\delta_0) < 0$ . Therefore, for  $n$  large enough,

$$\mathbb{E}_0[\iota''_n(\tilde{\delta}_n)] = (1 + o(1))\iota''(\delta_0) < 0.$$

We are now ready to show the second assertion in Proposition 8.1. Note that  $\kappa(\delta_1, \delta_0, \tilde{\delta}_n) \rightarrow \kappa(\delta_1, \delta_0, \delta_0)$  since  $\kappa$  is a continuous function. Putting all this together and re-writing the expression (23) we conclude that

$$\tilde{\delta}_n - \delta_0 = (1 + o(1))n^{\gamma-1} \frac{\kappa(\delta_1, \delta_0, \delta_0)}{|\iota''(\delta_0)|},$$

where

$$\begin{aligned} \iota''(\delta_0) &= \frac{m}{(2m + \delta_0)^2} - \sum_{k \geq m} \frac{p_{>k}(\delta_0)}{(k + \delta_0)^2} \\ &= \frac{m}{(2m + \delta_0)^2} - \frac{m}{2m + \delta_0} \sum_{k \geq m} \frac{p_k(\delta_0)}{k + \delta_0}. \end{aligned}$$

Therefore, after trivial algebraic manipulation, we conclude that

$$\begin{aligned} \tilde{\delta}_n - \delta_0 &= (1 + o(1))n^{\gamma-1} \frac{c(\delta_1 - \delta_0)}{2m + \delta_1} \frac{-1 + \sum_{k \geq m} \frac{2m + \delta_0}{k + \delta_0} p_k(\delta_0)}{\sum_{k \geq m} \frac{1}{k + \delta_0} p_k(\delta_0) - \frac{1}{2m + \delta_0}} \\ &= (1 + o(1))n^{\gamma-1} c(\delta_1 - \delta_0) \frac{2m + \delta_0}{2m + \delta_1}, \end{aligned}$$

proving the second assertion in the proposition. For the last assertion note that  $\delta \mapsto p_m(\delta)$  is a continuously differentiable function, so that

$$p_m(\tilde{\delta}_n) = p_m(\delta_0) + p'_m(\tilde{\delta}_n)(\tilde{\delta}_n - \delta_0),$$

where  $|\tilde{\delta}_n - \delta_0| \leq |\tilde{\delta}_n - \delta_0|$  and

$$p'_m(\delta) = -\frac{1}{(m + \delta + 2 + \delta/m)^2}. \tag{24}$$

Clearly  $\tilde{\delta}_n \rightarrow \delta_0$ , and  $p'_m(\delta_0) \neq 0$  and so

$$\begin{aligned} &n(p_m(\delta_0) - p_m(\tilde{\delta}_n)) \\ &= -(1 + o(1))n p'_m(\delta_0)(\tilde{\delta}_n - \delta_0) \\ &= (1 + o(1))n^\gamma \frac{1}{(m + \delta_0 + 2 + \delta_0/m)^2} c(\delta_1 - \delta_0) \frac{2m + \delta_0}{2m + \delta_1}. \end{aligned}$$

□

In conclusion, the sum of the terms *A* and *B* in (21) equals  $\alpha(\delta_0, \delta_1, m, c)(1 + o(1))n^\gamma$  with

$$\begin{aligned} \alpha(\delta_0, \delta_1, m, c) &= c \left[ 1 - \frac{p_m(\delta_0)}{p_m(\delta_1)} + \frac{1}{(m + \delta_0 + 2 + \delta_0/m)^2} (\delta_1 - \delta_0) \frac{2m + \delta_0}{2m + \delta_1} \right] \\ &= c(\delta_0 - \delta_1) \frac{m + \delta_0}{(2 + \delta_1/m)(m + \delta_0 + 2 + \delta_0/m)^2}. \end{aligned}$$

Clearly, this takes the value 0 when  $\delta_0 = \delta_1$ , and it is non-zero otherwise.

To complete the proof of Proposition 3.1 we need to characterize the term *C* in (21). This has two components, the first one already studied in the proof of Proposition 3.1. For the second component we must characterize the deviations of  $\hat{\delta}_n$  around  $\tilde{\delta}_n$ . Again, due to the implicit definition of the estimator this requires some care. The results are summarized in the following proposition, proven in [7]:

**Proposition 8.3.** *The estimator  $\hat{\delta}_n$  is consistent, both under the null and alternative models. Specifically,*

$$\mathbb{E}_\ell [|\hat{\delta}_n - \delta_0|] \rightarrow 0,$$

as  $n \rightarrow \infty$ , where  $\ell \in \{0, 1\}$ . In addition,

$$|\hat{\delta}_n - \tilde{\delta}_n| = o_{\mathbb{P}_\ell}(a_n/n)$$

where  $a_n = \omega(\sqrt{n} \log n)$ . Finally,

$$n(p_m(\hat{\delta}_n) - p_m(\tilde{\delta}_n)) = o_{\mathbb{P}_\ell}(a_n).$$

With these results in hand we are ready to complete the proof of Theorem 3.5. We proceed separately for the type I and type II error:

### 8.1. Type-I error

Recall the definition of  $a_n$  in Theorem 3.5, which satisfies  $a_n = \omega(\sqrt{n} \log n)$  and  $a_n = n^{\frac{1}{2}+o(1)}$ . Refer to the decomposition of the test statistic  $Q(G_n)$  in (21). Under the null model the term  $B$  is equal to zero, and for the term  $A$  we know (from (17)) that

$$|\mathbb{E}_0[N_m(n)] - np_m(\delta_0)| \leq C_0(\delta_0, m) := C_0.$$

Therefore

$$\begin{aligned} & \mathbb{P}_0(|Q(G_n)| \geq a_n) \\ &= \mathbb{P}_0(|\mathbb{E}_0[N_m(n)] - np_m(\delta_0) + N_m(n) - \mathbb{E}_0[N_m(n)] + np_m(\tilde{\delta}_n) - np_m(\hat{\delta}_n)| \geq a_n) \\ &\leq \mathbb{P}_0(|N_m(n) - \mathbb{E}_0[N_m(n)] + np_m(\tilde{\delta}_n) - np_m(\hat{\delta}_n)| \geq a_n - C_0) \\ &\leq \mathbb{P}_0\left(|N_m(n) - \mathbb{E}_0[N_m(n)]| \geq \frac{a_n - C_0}{2}\right) + \mathbb{P}_0\left(|np_m(\tilde{\delta}_n) - np_m(\hat{\delta}_n)| \geq \frac{a_n - C_0}{2}\right) \\ &= o(1), \end{aligned}$$

where the final result follows from (16) and the last statement in Proposition 8.3, since  $a_n = \omega(\sqrt{n} \log n)$ .

### 8.2. Type-II error

Referring again to the decomposition in (21) we see that the terms  $A$  and  $B$  now play a crucial role. Namely, we have shown that the sum of the two terms equals  $\alpha(\delta_0, \delta_1, m, c)n^\gamma + o(n^\gamma)$  where  $\alpha(\delta_0, \delta_1, m, c)$  is given by (10), which is non-zero provided  $\delta_0 \neq \delta_1$ , and it is non-zero otherwise. The characterization of the term  $C$  remains exactly the same as under the null model. Therefore,

$$\mathbb{E}_\ell[N_m(n)] - np_m(\delta_0) + np_m(\delta_0) - np_m(\tilde{\delta}_n) = \alpha(\delta_0, \delta_1, m, c)(1 + o(1))n^\gamma = \omega(a_n),$$

since  $a_n = n^{\frac{1}{2}+o(1)}$  and  $\gamma > \frac{1}{2}$ , and therefore

$$\mathbb{P}_1(|Q(G_n)| \geq \tau_n) = 1$$

concluding the proof of Theorem 3.5. □

## Acknowledgements

We are grateful to one anonymous referee for suggesting that our result actually holds under a weaker condition, and many other helpful suggestions.

## Funding

The work of RvdH was supported in part by the Netherlands Organisation for Scientific Research (NWO) through Gravitation-grant NETWORKS-024.002.003.

## Supplementary Material

**Asymptotic normality and technical proofs** (DOI: [10.3150/25-BEJ1935SUPP](https://doi.org/10.3150/25-BEJ1935SUPP); .pdf). In the supplement, we collect the proofs of various statements made in the paper which are too technical to be presented in the main manuscript. More precisely, we present a proof of Theorem 4.1, Proposition 8.2, and Proposition 8.3.

## References

- [1] Adamic, L.A., Huberman, B.A., Barabási, A.L., Albert, R., Jeong, H. and Bianconi, G. (2000). Power-law distribution of the world wide web. *Science* **287** 2115. <https://doi.org/10.1126/science.287.5461.2115a>
- [2] Albert, R. and Barabási, A.L. (2000). Topology of evolving networks: Local events and universality. *Phys. Rev. Lett.* **85** 5234–5237. <https://doi.org/10.1103/PhysRevLett.85.5234>
- [3] Baldassarri, S. and Bet, G. (2022). Asymptotic normality of degree counts in a general preferential attachment model. *Markov Process. Related Fields* **28** 577–603.
- [4] Banerjee, S., Bhamidi, S. and Carmichael, I. (2023). Fluctuation bounds for continuous time branching processes and evolution of growing trees with a change point. *Ann. Appl. Probab.* **33** 2919–2980. <https://doi.org/10.1214/22-AAP1881>
- [5] Barabási, A.L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. <https://doi.org/10.1126/science.286.5439.509>
- [6] Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Phys. A* **311** 590–614. [https://doi.org/10.1016/S0378-4371\(02\)00736-7](https://doi.org/10.1016/S0378-4371(02)00736-7)
- [7] Bet, G., Bogerd, K., Castro, R.M. and van der Hofstad, R. (2026). Supplement to “Detecting a late changepoint in the preferential attachment model.” <https://doi.org/10.3150/25-BEJ1935SUPP>
- [8] Bhamidi, S., Jin, J. and Nobel, A. (2018). Change point detection in network models: Preferential attachment and long range dependence. *Ann. Appl. Probab.* **28** 35–78. <https://doi.org/10.1214/17-AAP1297>
- [9] Bhattacharjee, M., Banerjee, M. and Michailidis, G. (2020). Change point estimation in a dynamic stochastic block model. *J. Mach. Learn. Res.* **21** 4330–4388.
- [10] Bollobás, B. and Riordan, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24** 5–34. <https://doi.org/10.1007/s00493-004-0002-2>
- [11] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** 279–290. <https://doi.org/10.1002/rsa.1009>
- [12] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. and Wiener, J. (2000). Graph structure in the web. *Comput. Netw.* **33** 309–320. [https://doi.org/10.1016/S1389-1286\(00\)00083-9](https://doi.org/10.1016/S1389-1286(00)00083-9)
- [13] Bubeck, S., Devroye, L. and Lugosi, G. (2017). Finding Adam in random growing trees. *Random Structures Algorithms* **50** 158–172. <https://doi.org/10.1002/rsa.20649>
- [14] Bubeck, S., Mossel, E. and Rácz, M.Z. (2015). On the influence of the seed graph in the preferential attachment model. *IEEE Trans. Netw. Sci. Eng.* **2** 30–39. <https://doi.org/10.1109/TNSE.2015.2397592>

- [15] Bubeck, S., Eldan, R., Mossel, E. and Rácz, M.Z. (2017). From trees to seeds: On the inference of the seed from large trees in the uniform attachment model. *Bernoulli* **23** 2887–2916. <https://doi.org/10.3150/16-BEJ831>
- [16] Cirkovic, D., Wang, T. and Zhang, X. (2022). Likelihood-based changepoint detection in preferential attachment networks. Preprint. Available at [arXiv:2206.01076](https://arxiv.org/abs/2206.01076).
- [17] Curien, N., Duquesne, T., Kortchemski, I. and Manolescu, I. (2015). Scaling limits and influence of the seed graph in preferential attachment trees. *J. Éc. Polytech. Math.* **2** 1–34. <https://doi.org/10.5802/jep.15>
- [18] Deijfen, M., van den Esker, H., van der Hofstad, R. and Hooghiemstra, G. (2007). A preferential attachment model with random initial degrees. *Ark. Mat.* **47** 41–72. <https://doi.org/10.1007/s11512-007-0067-4>
- [19] Dommers, S., van der Hofstad, R. and Hooghiemstra, G. (2010). Diameters in preferential attachment graphs. *J. Stat. Phys.* **139** 72–107. <https://doi.org/10.1007/s10955-010-9921-z>
- [20] Douwes, A.S. (2024). Detection of change points and anomalies in Preferential Attachment Models. Master thesis.
- [21] Du, H., Gong, S. and Xu, J. (2025). A proof of the changepoint detection threshold conjecture in preferential attachment models. Preprint. Available at [arXiv:2502.00514](https://arxiv.org/abs/2502.00514).
- [22] Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999). On power-law relationships of the internet topology. *Comput. Commun. Rev.* **29** 251–261. <https://doi.org/10.1145/316194.316229>
- [23] Farkas, I., Jeong, H., Vicsek, T., Barabási, A.L. and Oltvai, Z.N. (2003). The topology of the transcription regulatory network in the yeast, *Saccharomyces cerevisiae*. *Phys. A* **318** 601–612. [https://doi.org/10.1016/S0378-4371\(02\)01731-4](https://doi.org/10.1016/S0378-4371(02)01731-4)
- [24] Gao, F. (2011). Modeling and interference of the internet movie database. Master thesis, Eindhoven University of Technology.
- [25] Gao, F. and van der Vaart, A. (2017). On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees. *Stoch. Process. Appl.* **127** 3754–3775. <https://doi.org/10.1016/J.SPA.2017.03.008>
- [26] Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabási, A.L. (2000). The large-scale organization of metabolic networks. *Nature* **407** 651–654. <https://doi.org/10.1038/35036627>
- [27] Kaddouri, I., Naulet, Z. and Gassiat, É. (2024). On the impossibility of detecting a late change-point in the preferential attachment random graph model. Preprint. Available at [arXiv:2407.18685](https://arxiv.org/abs/2407.18685).
- [28] Marchand, D.C. and Manolescu, I. (2020). Influence of the seed in affine preferential attachment trees. *Bernoulli* **26** 1665–1705. <https://doi.org/10.3150/19-BEJ1152>
- [29] Middendorff, M., Ziv, E. and Wiggins, C.H. (2005). Inferring network mechanisms: The *Drosophila melanogaster* protein interaction network. *Proc. Natl. Acad. Sci. USA* **102** 3192–3197. <https://doi.org/10.1073/pnas.0409515102>
- [30] Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98** 404–409. <https://doi.org/10.1073/pnas.021544898>
- [31] Pensky, M. and Zhang, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electron. J. Stat.* **13** 678–709. <https://doi.org/10.1214/19-EJS1533>
- [32] Perc, M. (2014). The Matthew effect in empirical data. *J. R. Soc. Interface* **11**. <https://doi.org/10.1098/rsif.2014.0378>
- [33] Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* **4** 131–134. <https://doi.org/10.1007/s100510050359>
- [34] Resnick, S.I. and Samorodnitsky, G. (2016). Asymptotic normality of degree counts in a preferential attachment model. *Adv. Appl. Probab.* **48** 283–299. <https://doi.org/10.1017/apr.2016.56>
- [35] van der Hofstad, R. (2017). *Random Graphs and Complex Networks 1*. Cambridge University Press. <https://doi.org/10.1017/9781316779422>
- [36] van der Hofstad, R. (2024). *Random Graphs and Complex Networks 2*. Cambridge University Press. <https://doi.org/10.1017/9781316795552>
- [37] Wang, D., Yu, Y. and Rinaldo, A. (2021). Optimal change point detection and localization in sparse dynamic networks. *Ann. Statist.* **49** 203–232. <https://doi.org/10.1214/20-AOS1953>
- [38] Wang, M., Yu, G. and Yu, D. (2008). Measuring the preferential attachment mechanism in citation networks. *Phys. A* **387** 4692–4698. <https://doi.org/10.1016/j.physa.2008.03.017>

- [39] Wang, H., Tang, M., Park, Y. and Priebe, C.E. (2014). Locality statistics for anomaly detection in time series of graphs. *IEEE Trans. Signal Process.* **62** 703–717. <https://doi.org/10.1109/TSP.2013.2294594>
- [40] Watts, D.J. (1999). *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press.
- [41] Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of ‘small-world’ networks. *Nature* **393** 440–442. <https://doi.org/10.1038/30918>
- [42] Zhao, Z., Chen, L. and Lin, L. (2019). Change-point detection in dynamic networks via graphon estimation. Preprint. Available at [arXiv:1908.01823](https://arxiv.org/abs/1908.01823).

*Received March 2024 and revised September 2025*