

DETECTING A LATE CHANGEPOINT IN THE PREFERENTIAL ATTACHMENT MODEL

BY GIANMARCO BET^{1,a} , KAY BOGERD^{2,b} RUI M. CASTRO^{2,c} 
REMCO VAN DER HOFSTAD^{2,d} 

¹Department of Mathematics and Computer Science, University of Florence, ^agianmarco.bet@unifi.it

²Department of Mathematics and Computer Science, Eindhoven University of Technology, ^bkaybogerd@hotmail.com;
^crmcastro@tue.nl; ^dr.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 δ_0 , and the alternative hypothesis is a preferential attachment model where the affine attachment parameter changes from δ_0 to δ_1 at an unknown changepoint time τ_n . For our analysis we focus on the regime where δ_0 and δ_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 the 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 there is no powerful test based on the final network snapshot when $\gamma < \frac{1}{2}$. The first test we propose requires knowledge of δ_0 . The second test is significantly more involved, and does not require the knowledge of δ_0 while still achieving the same 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.

CONTENTS

1	Introduction	2
1.1	Related work	3
2	Model	3
2.1	Assumptions and notation	4
3	Minimal degree tests	5
3.1	Powerful test for known δ_0	5
3.2	Powerful test for unknown δ_0	7
3.3	Asymptotically calibrated tests	8
4	Asymptotic normality of test statistics	9
5	Numerical experiments	10
6	Discussion and open problems	12
6.1	Early changepoint	12
6.2	Dynamical graph observations	12
6.3	Lower bounds	13

MSC2020 subject classifications: Primary 05C80; secondary 60F05, 62M02, 60C05.

Keywords and phrases: Preferential attachment model, Changepoint detection.

6.4	Boundary case $\gamma = 1$	14
6.5	Other test statistics	14
6.6	Boundary case $\gamma = \frac{1}{2}$	14
7	Powerful test for known δ_0 : Proof of Theorem 3.1	14
7.1	Type-I error	15
7.2	Type-II error	15
8	Powerful test for unknown δ_0 : Proof of Theorem 3.4	17
8.1	Type-I error	21
8.2	Type-II error	22
9	Asymptotic normality proofs	22
9.1	Asymptotic normality under the null hypothesis	22
9.2	Asymptotic normality under the alternative hypothesis	26
A	Proof of auxiliary results	31
A.1	Proof of Proposition 8.2	31
A.2	Proof of Lemma A.1	36
A.3	Proof of Proposition 8.3	37
A.4	Sketch proof of Lemma A.3	38
	Funding	38
	References	39

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, 11] or internet [20], biological networks [21, 27, 29], collaboration networks of movie actors [2, 22], and citation networks [6, 30, 33, 37]. Furthermore, the typical distance between vertices in the preferential attachment model is small [19] (see also [26, Chapter 8] and references therein). This is called the *small-world* phenomenon [38, 39].

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 [9, 10], see also [25, 26] 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 have 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 δ_0 to δ_1 at an unknown moment τ_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 relevant when one wants to detect the change as quickly as possible.

1.1. Related work. Our work nicely complements earlier results [4, 7] 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 also some 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, 7] 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. The authors of [15] propose a likelihood-ratio testing procedure to detect a changepoint in a preferential attachment tree and the associated changepoint estimator. Crucially the methods in [15] 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 [15] 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 [12–14, 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 [8, 31, 35, 36, 40]. 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 set of edges. 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 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

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

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 [25, 26] 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. This is our only observation, i.e., we do not have access to the sequence $\{G_1, \dots, G_{n-1}\}$. In particular, the order of the vertices is unknown to us. Clearly the distribution of this random graph is parameterized by m and δ . 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 δ .

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 δ 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 δ_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 γ 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 δ is a step-function, the attachment rule in (1) can be further simplified to get, for $v \in \{v_0, \dots, v_{t-1}\}$,

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

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, δ_0, δ_1, c , and γ 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 δ_0 .* We begin with the assumption that δ_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 δ_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, 25]. 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 α (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. The proof of Theorem 3.1 is given in Section 7.

THEOREM 3.1 (Asymptotically powerful test: known δ_0). *Consider the test that rejects the null hypothesis for large values of $T(G_n)$. Namely, define the test*

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

where $\alpha \in (0, 1)$. The type-I error of this test is asymptotically bounded by α , 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 $\gamma < \frac{1}{2}$.

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

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

where

$$(3) \quad \eta(\delta_0, \delta_1) := c(1 - p_m(\delta_0)/p_m(\delta_1)).$$

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 $\gamma > \frac{1}{2}$.

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. Although a formal answer to this question is still open, we conjecture that no test can be powerful in that scenario:

CONJECTURE 3.2 (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 should 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 multi-variate normal distribution. As shown in Lemma 9.2 these degree counts are also asymptotically normal, with the same covariance matrix, but a rather small mean-shift. We conjecture this shift is small enough so to imply 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, but proving such a statement requires a careful formalization of this insight.

3.2. *Powerful test for unknown δ_0 .* The knowledge of δ_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 δ_0 . For such an approach to be fruitful a candidate estimator must be “close enough” to δ_0 both under the null and alternative models. Gao and van der Vaart in [23] consider the problem of estimating δ_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 δ_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 [23], to avoid the usual issues at the boundary of the parameter space we make an extra assumption that the range of possible values for δ_0 and δ_1 is known:

ASSUMPTION 3.3 (Containment of δ_0, δ_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 [23], 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

$$(4) \quad \hat{\delta}_n := \underset{\delta \in [\delta_{\min}, \delta_{\max}]}{\operatorname{argmax}} \iota_n(\delta).$$

Equivalently (for large n) we can define $\hat{\delta}_n$ as the solution in $\delta \in [\delta_{\min}, \delta_{\max}]$ of $\frac{\partial}{\partial \delta} \iota_n(\delta) := \iota'_n(\delta) = 0$. Although not obvious, this definition coincides with (4) for large n , since it is shown in [23] that the solution of $\iota'_n(\delta) = 0$ exists and is unique for large enough n with high probability. Note that the score function is given by

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

Motivated by this estimator we consider the test statistic

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

This is analogous to the previously considered statistic, with the exception that δ_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 δ_0 , under both the null and alternative models. When $\frac{1}{2} < \gamma < 1$ this is indeed the case, and we show that $\hat{\delta}_n$ is a consistent estimator of δ_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 3.4 (Asymptotically powerful test, unknown δ_0). *Consider Assumption 3.3 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 Theorem 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 $\gamma = \frac{1}{2}$, one can expect Gaussian fluctuations of $N_m(n)$ under the null and alternative hypotheses to compete with the resulting change in expectations of $N_m(n)$, so the Type-II error can not be expected to 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 α . Theorem 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

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

Furthermore, let z_α denote the right-quantile function of the standard normal distribution¹. In particular, $z_{\alpha/2} > 0$ when $\alpha \in (0, 1)$. We are now ready to present the asymptotically calibrated test when δ_0 is assumed known.

THEOREM 3.5 (Asymptotically calibrated test for known δ_0). *Let $\alpha \in (0, 1)$ and define the test*

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

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

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

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

$$\mathbb{P}_1(\psi_{\text{cal}}(G_n) = 0) \rightarrow 0.$$

¹For $\alpha \in (0, 1)$, let z_α be the unique solution of $\alpha = \int_{z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$.

A statement regarding partial power when $\gamma = \frac{1}{2}$ is also possible, but not particularly insightful. To define the next test we require some additional notation. Let

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

and

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

The next theorem presents the asymptotically calibrated test for the unknown δ_0 case:

THEOREM 3.6 (Asymptotically calibrated test, unknown δ_0). *Consider Assumption 3.3, let $\alpha \in (0, 1)$, and define the test*

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

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

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

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

$$\mathbb{P}_1(\phi_{\text{cal}}(G_n) = 0) \rightarrow 0.$$

The proofs of Theorems 3.5 and 3.6 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 δ_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 δ_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 δ_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 [23] 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 μ and variance σ^2 , and let \xrightarrow{D} denote convergence in distribution. The following theorem, proved in Section 9, 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. As $n \rightarrow \infty$,*

$$(10) \quad \frac{T(G_n)}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, w(\delta_0, m)).$$

Moreover, under Assumption 3.3 and the null model, as $n \rightarrow \infty$,

$$(11) \quad \frac{Q(G_n)}{\sqrt{n}} = \frac{N_m(n) - np_m(\hat{\delta}_n)}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, w(\delta_0, m) + u(\delta_0, m)) .$$

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

$$(12) \quad \frac{T(G_n) - \mathbb{E}_1[T(G_n)]}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, w(\delta_0, m)) .$$

Moreover, under the alternative model with $\gamma \in (\frac{1}{2}, 1)$ and Assumption 3.3, as $n \rightarrow \infty$,

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

Note that (10) has been proved in [3]. In the second statement (11), $u(\delta_0, m)$ captures the adjustment in the variability in the test statistic by using an estimate of δ_0 , instead of the actual value. The convergence results (10) and (11) provide an avenue for asymptotically exact calibration of the proposed test as given in Theorems 3.5 and 3.6 (equivalently, for computation of asymptotically exact p -values).

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 Theorems 3.1 and 3.4). It is important to remark that, for the convergence result (13), the assumption $\gamma > \frac{1}{2}$ appears to be merely technical, and it should be possible to drop it. Finally, 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 Section 9.

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 δ_0 has slightly less power than the test making use of that knowledge. Finally, there is an asymmetry of 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 δ_1

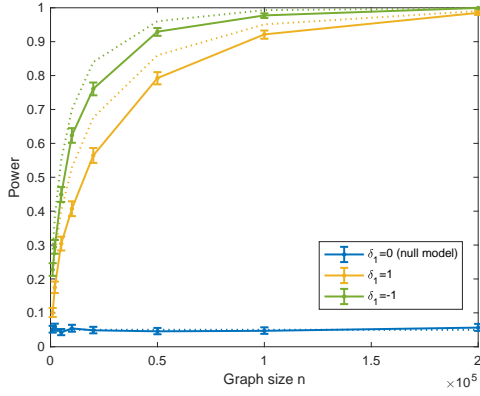
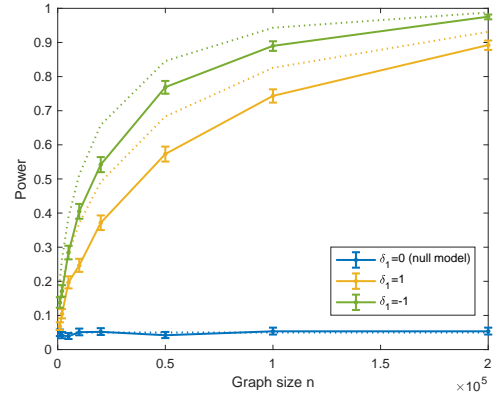
(a) Power of the tests (known δ_0 and $\alpha = 0.05$).(b) Power of the tests (unknown δ_0 and $\alpha = 0.05$).

FIG 1. Power of the tests calibrated using asymptotic normality for $m = 5$. Panel (a) corresponds to the case of known δ_0 and panel (b) corresponds to the unknown δ_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.

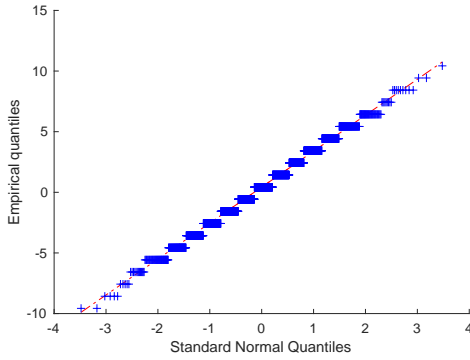
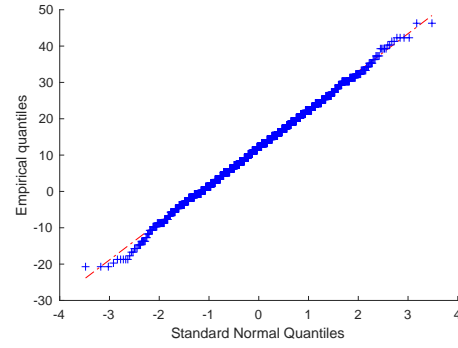
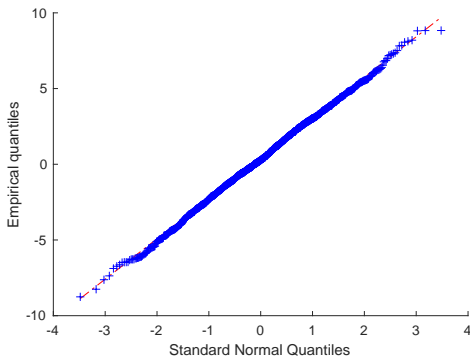
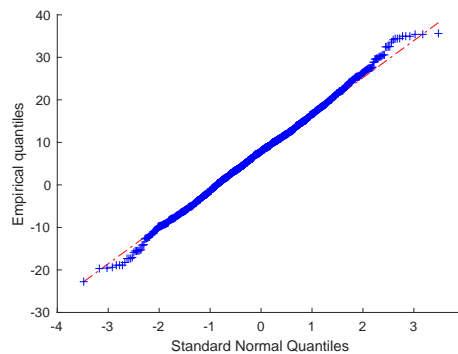
(a) $T(G_n)$ with $n = 100$ and $\delta_0 = \delta_1 = 0$.(b) $T(G_n)$ with $n = 1000$ and $\delta_0 = \delta_1 = 0$.(c) $Q(G_n)$ with $n = 100$ and $\delta_0 = \delta_1 = 0$.(d) $Q(G_n)$ with $n = 1000$ and $\delta_0 = \delta_1 = 0$.

FIG 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 δ_0 statistic) and panels (c) and (d) correspond to $Q(G_n)$ (unknown δ_0 statistics). Even for very small values of n distribution of the statistics is very well approximated by a normal distribution.

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

$$\hat{\mu}_n^{(T)} := \frac{1}{B} \sum_{b=1}^B T(g_n^{(b)})$$

and

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

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^γ 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 $\eta(\delta_0, \delta_1)$ from (3) and $\alpha(\delta_0, \delta_1)$ from (24) 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 δ_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^γ , 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, 7, 15], 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^γ . 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.

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
	$\eta(\delta_0, \delta_0)$ = 0	$\alpha(\delta_0, \delta_0)$ = 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
	$\eta(\delta_0, \delta_1)$ = -0.0649	$\alpha(\delta_0, \delta_1)$ = -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
	$\eta(\delta_0, \delta_1)$ = 0.0794	$\alpha(\delta_0, \delta_1)$ = 0.0567	$w(\delta_0, m)$ = 0.1020	$w(\delta_0, m) + u(\delta_0, m)$ = 0.0811

6.3. *Lower bounds.* Conjecture 3.2 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 Lemma 9.2, 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 Theorems 3.1 and 3.4 the case $\gamma = 1$ is excluded. This is in contrast with the results in [4, 7, 15]. 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 3.4 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 will likely not be a good practical choice (particularly when c is large), and the statistics used in [4, 7] will likely lead to more powerful tests.

6.5. *Other test statistics.* Note that information about the presence of a changepoint is present not only in $N_m(n)$, but also on other counts of low-degree vertices. With this in mind, a test based on $(N_m(n), N_{m+1}(n))$ can be potentially more powerful (in a finite sample sense) than the test we proposed. Although we expect such tests to have exactly the same asymptotic performance, they can perform much better for finite 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 .

7. Powerful test for known δ_0 : Proof of Theorem 3.1. The main idea of the proof is to decompose the test statistic $T(G_n)$ in two terms:

$$(14) \quad T(G_n) = N_m(n) - np_m(\delta_0) = \underbrace{\mathbb{E}_\ell[N_m(n)] - np_m(\delta_0)}_{:=A} + \underbrace{N_m(n) - \mathbb{E}_\ell[N_m(n)]}_{:=B},$$

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 [25, Lemma 8.5], and such that $\mathbb{E}_\ell[N_m(n)] = M_1$ and $N_m(n) = M_n$. Furthermore, by [25, 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 $x > 0$,

$$(15) \quad \forall \ell \in \{0, 1\} \quad \mathbb{P}_\ell(|N_m(n) - \mathbb{E}_\ell[N_m(n)]| \geq x) \leq 2e^{-\frac{x^2}{8m^2n}}.$$

This completes the characterization of the term B in (14). We now proceed by considering the type-I and type-II errors separately.

7.1. *Type-I error.* To control the term A in (14) under the null model we use [18, Proposition 2.2] (see also [25, Proposition 8.7]), which states that there exists a constant $C_0 = C_0(\delta_0, m)$ such that, for all $n \geq 1$,

$$(16) \quad |\mathbb{E}_0[N_m(n)] - np_m(\delta_0)| \leq C_0.$$

Combining this with (15) 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 α , 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

$$(17) \quad \begin{aligned} \mathbb{E}_1[N_m(n)] - np_m(\delta_0) &= \mathbb{E}_0[N_m(n)] - np_m(\delta_0) \\ &\quad + \mathbb{E}_1[N_m(n)] - \mathbb{E}_1[N_m(\tau_n)] - (\mathbb{E}_0[N_m(n)] - \mathbb{E}_0[N_m(\tau_n)]). \end{aligned}$$

The equality holds as the law of G_{τ_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 ℓ , 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 3.4:

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

$$\begin{aligned} & \mathbb{P}_\ell(D_v(n) - D_v(\tau_n) > 0 \mid D_v(\tau_n) = k) \\ &= 1 - \mathbb{P}_\ell(D_v(n) - D_v(\tau_n) = 0 \mid D_v(\tau_n) = k) \\ &= 1 - \prod_{j \in [n] \setminus [\tau_n]} \prod_{i=1}^m \left(1 - \frac{k + \delta_\ell}{j(2m + \delta_\ell) - 2m + i - 1}\right) \\ &= 1 - \exp \left(\sum_{j \in [n] \setminus [\tau_n]} \sum_{i=1}^m \log \left(1 - \frac{k + \delta_\ell}{j(2m + \delta_\ell) - 2m + i - 1}\right) \right) \\ (18) \quad &= 1 - \exp \left(\sum_{j \in [n] \setminus [\tau_n]} \sum_{i=1}^m -(1 + o(1)) \frac{k + \delta_\ell}{j(2m + \delta_\ell) - 2m + i - 1} \right) \\ &= 1 - \exp \left(-(1 + o(1))cmn^\gamma \frac{k + \delta_\ell}{n(2m + \delta_\ell)} \right) \\ (19) \quad &= 1 - \left(1 - (1 + o(1))cmn^{\gamma-1} \frac{k + \delta_\ell}{2m + \delta_\ell} \right) \\ &= (1 + o(1))cmn^{\gamma-1} \frac{k + \delta_\ell}{2m + \delta_\ell}, \end{aligned}$$

where in (18) we relied on the fact that $k = o(n)$ and for (19) it is crucial that $k = o(n^{1-\gamma})$. \square

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 (17) and the fact that $\mathbb{E}_0[N_m(\tau_n)] = \mathbb{E}_1[N_m(\tau_n)]$, yields

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

where $0 < \gamma < 1$, and we have again used (16) to relate $\mathbb{E}_0[N_m(n)]$ to $np_m(\delta_0)$. \square

Similarly as for the type-I error, we can use Proposition 7.1 together with the Azuma-Hoeffding inequality (15) 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

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

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 δ_0 : Proof of Theorem 3.4. To prove Theorem 3.4 we partially leverage on the results and analysis in Theorem 3.1. However, we must take into account that δ_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

$$(20) \quad Q(G_n) = N_m(n) - np_m(\hat{\delta}_n)$$

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

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 Theorem 3.1 is in the characterization of B , the second term. For this we need to understand how fast $\tilde{\delta}_n$ converges to δ_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 $\mathbb{E}_\ell[\iota'_n(\delta)] = 0$ 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 $\frac{1}{2} < \gamma < 1$. Under the alternative model $\tilde{\delta}_n$ converges to δ_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 $\iota'_n(\delta)$ is a score function, therefore we have immediately that $\mathbb{E}_0[\iota'_n(\delta_0)] = 0$. Under the alternative model $\tilde{\delta}_n$ will not be equal to δ_0 and quantifying this deviation is crucial to our analysis. The following technical result, examining the difference between $\mathbb{E}_1[\iota'_n(\delta)]$ and $\mathbb{E}_0[\iota'_n(\delta)]$, is instrumental:

PROPOSITION 8.2 (Analysis of differences of means). *Let $\frac{1}{2} < \gamma < 1$. As $n \rightarrow \infty$, and uniformly for every $\delta \in [\delta_{\min}, \delta_{\max}]$,*

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

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 the appendix. 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

$$\iota'(\delta) := \sum_{k \geq m} \frac{p_{>k}(\delta_0)}{k + \delta} - \frac{1}{2 + \delta/m},$$

where $p_{>k}(\delta_0) = \sum_{j>k} p_j(\delta_0)$. Intuitively, this should be the limit of $\iota'_n(\delta)$ as $n \rightarrow \infty$, as we see next. Note also that we can rewrite this expression in terms of p_k , by noticing (see [23, Lemma 2]) that, for $k \geq m$

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

Therefore,

$$\iota'(\delta) = \frac{1}{2 + \delta_0/m} \sum_{k \geq m} \frac{k + \delta_0}{k + \delta} p_k(\delta_0) - \frac{1}{2 + \delta/m} .$$

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

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

[23, 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 $|\iota'_n(\delta)|$ and $|\iota'(\delta)|$ are bounded, uniformly in δ 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 ι' 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 $n \rightarrow \infty$,

$$(21) \quad \forall \ell \in \{0, 1\} \quad \mathbb{E}_\ell \left[\sup_{\delta \in [\delta_{\min}, \delta_{\max}]} |\iota'_n(\delta) - \iota'(\delta)| \right] \rightarrow 0 .$$

By definition $\mathbb{E}_1[\iota'_n(\tilde{\delta}_n)] = 0$, therefore we conclude that $\iota'(\tilde{\delta}_n) \rightarrow 0$. [23, Lemma 4] shows that ι' has a unique zero at δ_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 ι'_n is a score function. Furthermore, by definition $\mathbb{E}_1[\iota'_n(\tilde{\delta}_n)] = 0$, therefore Proposition 8.2 implies that

$$\begin{aligned} 0 &= \mathbb{E}_1[\iota'_n(\tilde{\delta}_n)] - \mathbb{E}_0[\iota'_n(\tilde{\delta}_n)] + \mathbb{E}_0[\iota'_n(\tilde{\delta}_n)] \\ &= \kappa(\delta_1, \delta_0, \tilde{\delta}_n)n^{\gamma-1}(1 + o(1)) + \mathbb{E}_0[\iota'_n(\delta_0) + \iota''_n(\tilde{\delta}_n)(\tilde{\delta}_n - \delta_0)] \\ (22) \quad &= \kappa(\delta_1, \delta_0, \tilde{\delta}_n)n^{\gamma-1}(1 + o(1)) + \mathbb{E}_0[\iota''_n(\tilde{\delta}_n)](\tilde{\delta}_n - \delta_0), \end{aligned}$$

where $|\tilde{\delta}_n - \delta_0| \leq |\tilde{\delta}_n - \delta_0|$. Clearly $\tilde{\delta}_n \rightarrow \delta_0$. As shown in [23], $\iota''_n(\delta)$ 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 ι'_n is uniformly bounded). In addition, it is also shown in [23] 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 κ is a continuous function. Putting all this together and re-writing the expression (22) 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

$$(23) \quad p'_m(\delta) = -\frac{1}{(m + \delta + 2 + \delta/m)^2}.$$

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

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

□

In conclusion, the sum of the terms A and B in (20) equals $\alpha(\delta_0, \delta_1)(1 + o(1))n^\gamma$ with

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

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

To complete the proof of Theorem 3.1 we need to characterize the term C in (20). This has two components, the first one already studied in the proof of Theorem 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 the appendix:

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 3.4. We proceed separately for the type I and type II error:

8.1. Type-I error. Recall the definition of a_n in Theorem 3.4, 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 (20). Under the null model the term B is equal to zero, and for the term A we know (from (16)) that

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

Therefore

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

where the final result follows from (15) 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 (20) 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 $\alpha(\delta_0, \delta_1)n^\gamma + o(n^\gamma)$, where

$$\begin{aligned} \alpha(\delta_0, \delta_1) &= c \left[1 - \frac{p_m(\delta_0)}{p_m(\delta_1)} + \frac{1}{(m + \delta_0 + 2 + \delta_0/m)^2} (\delta_1 - \delta_0) \frac{2m + \delta_0}{2m + \delta_1} \right] \\ (24) \quad &= c(\delta_0 - \delta_1) \frac{m + \delta_0}{(2 + \delta_1/m)(m + \delta_0 + 2 + \delta_0/m)^2} . \end{aligned}$$

Clearly, this takes the value 0 when $\delta_0 = \delta_1$, and it is non-zero otherwise. The characterization of the term C remains exactly the same as under the null model. Therefore,

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

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 3.4. \square

9. Asymptotic normality proofs. This section is dedicated to the proof of Theorem 4.1. The proof for the null model relies on a somewhat involved application of a martingale central limit theorem to an appropriately constructed martingale. The proof for the alternative model hinges on the null model result, as well as further estimates of the effect of the last $n - \tau_n$ vertices on the distribution of $N_k(n)$ for $k \geq m$. Since one proof hinges on the other, and to avoid any confusion between the null and alternative models, we separate these two cases in two sections. Specifically, in Section 9.1 we prove the asymptotic normality of test statistics under the null model, i.e., (11). In Section 9.2 we prove the asymptotic normality of test statistics under the alternative model, i.e., (12) and (13).

9.1. *Asymptotic normality under the null hypothesis.* Let $\mathcal{N}(\mu, \Sigma)$ denote the normal distribution with mean μ and covariance matrix Σ . The following result establishes the joint asymptotic normality of $N_m(n)$ and $\hat{\delta}_n$, which can be then used to deduce (11) by an application of the delta method:

PROPOSITION 9.1 (Joint normality of count of degree m vertices and estimator for δ_0). *Under the null model, as $n \rightarrow \infty$,*

$$(25) \quad \sqrt{n} \begin{pmatrix} \frac{N_m(n) - np_m(\delta_0)}{n} \\ \hat{\delta}_n - \delta_0 \end{pmatrix} \xrightarrow{D} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma(\delta_0, m) \right) ,$$

where the covariance matrix is

$$(26) \quad \Sigma(\delta_0, m) := \begin{pmatrix} w(\delta_0, m) - \frac{b(\delta_0, m)}{v(\delta_0, m)} & -\frac{b(\delta_0, m)}{v(\delta_0, m)} \\ -\frac{b(\delta_0, m)}{v(\delta_0, m)} & \frac{1}{v(\delta_0, m)} \end{pmatrix} ,$$

where w and v are defined in (7) and (8), and $b(\delta_0, m) := \frac{m^2}{(\delta_0 + m(2 + m + \delta_0))^2}$.

PROOF OF PROPOSITION 9.1. The convergence of the marginals in (25) follows from results in the literature, which we briefly recall here. From the general result in [3] it follows that, under the null model,

$$\frac{N_m(n) - np_m(\delta_0)}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, w(\delta_0, m)) ,$$

as $n \rightarrow \infty$. Moreover, from [23] follows that, under the null model,

$$\sqrt{n}(\hat{\delta}_n - \delta_0) \xrightarrow{D} \mathcal{N}(0, 1/v(\delta_0, m)) ,$$

as $n \rightarrow \infty$. To show joint convergence, we apply the multivariate martingale central limit theorem (MMCLT) to (a linear transformation of) the vector on the left-hand side of (25). For more details on the MMCLT, see, e.g., [16] and references therein. In fact, the asymptotic normality of the marginals has been proven respectively in [3] and [23] by applying a martingale central limit theorem to appropriately constructed martingale difference sequences $X_{t,i}$ and $Y_{t,i}$, which we define precisely later on. To obtain our results, we apply the MMCLT to the (square-integrable) martingale difference array $(X_{t,i}, Y_{t,i})$. The MMCLT requires two ingredients. First, the components of the array must satisfy the Lindeberg condition. Since this is also an ingredient for the univariate martingale central limit theorem, it has already been verified for $X_{t,i}$ in [3, Section 3.2] and for $Y_{t,i}$ in [23, Section 4]. Second, one should compute the asymptotic expression for the covariance matrix. Note that the asymptotic variance of $X_{t,i}$ (resp. $Y_{t,i}$) has already been computed explicitly in [3] (resp. [23]).

We begin by introducing the notation required to define the martingale difference array $(X_{t,i}, Y_{t,i})$, which, roughly speaking, is a linear transformation of $(N_m(t), \hat{\delta}_t)$. We then compute the asymptotic covariance matrix of $(X_{t,i}, Y_{t,i})$. Finally, we perform a linear transformation to deduce a joint central limit theorem for $(N_m(n), \hat{\delta}_n)$.

To construct the martingale difference sequences, it is important to consider the “intermediate” m steps in the preferential attachment process, at each time t . For $i = 1, \dots, m-1$, we denote by $N_k(s, i)$ the number of vertices of degree k after the i -th edge has been attached at time s , excluding the vertex s . We set $N_k(s+1, 0) := N_k(s, m) = N_k(s)$. We further define $D_{s,i} := D_{v_{s,i}}(G_{s,i-1})$. In other words, $D_{s,i}$ is the degree of the vertex which will be attached to when constructing $G_{s,i}$ from $G_{s,i-1}$. For compactness, we define the constants

$$W_{t,i} := \prod_{s=1}^{t-1} \prod_{j=1}^m a_{s,j} \prod_{j=1}^i a_{t,j} \quad \text{where} \quad a_{t,i} := 1 - \frac{m + \delta_0}{S_{t,i-1}} ,$$

and where

$$S_{t,i} := t\delta_0 + 2(t-1)m + i .$$

Define also

$$\xi := \frac{m + \delta_0}{2m + \delta_0} ,$$

$$\Lambda := \prod_{j=0}^{m-1} \frac{\Gamma(1 - \frac{2m-j}{2m+\delta_0})}{\Gamma(1 - \frac{3m-j+\delta_0}{2m+\delta_0})} .$$

In the constants defined above, we omitted the dependence on δ_0 and m to facilitate readability of the computations that follow. The martingale difference sequences (from [3, 24]) are defined as

$$X_{t,i} := \left(\frac{N_m(t, i) - a_{t,i-1}N_m(t, i-1) + \mathbb{1}\{i=m\}}{W_{t,i-1}} \right) \frac{1}{n^{\frac{1}{2}+m\xi}} ,$$

$$Y_{t,i} := \left(\frac{1}{D_{t,i} + \delta_0} - \frac{t}{S_{t,i-1}} \right) \frac{1}{n^{\frac{1}{2}}} .$$

Here $X_{t,i}$ is the first component of the martingale difference array in [3], and $Y_{t,i}$ is the martingale difference defined in [23]. Next, we compute the asymptotic covariance matrix. By [3, Section 3.1] and [23, Section 4], as $n \rightarrow \infty$,

$$(27) \quad \sum_{t=2}^n \sum_{i=1}^m \mathbb{E}[X_{t,i}^2 | G_{t,i-1}] \xrightarrow{\mathbb{P}_0} \frac{1}{\Lambda^2} \frac{m}{1+2m\xi} p_m \xi (1-p_m \xi) \\ = \frac{1}{\Lambda^2} \frac{m^2(m+\delta_0)(1+m+\delta_0)(2m+\delta_0)}{(\delta_0+2m(1+m+\delta_0))(\delta_0+m(2+m+\delta_0))^2},$$

$$(28) \quad \sum_{t=2}^n \sum_{i=1}^m \mathbb{E}[Y_{t,i}^2 | G_{t,i-1}] \xrightarrow{\mathbb{P}_0} \sum_{k=1}^{\infty} \frac{mp_k}{(k+\delta_0)(2m+\delta_0)} - \frac{m}{(2m+\delta_0)^2}.$$

We are left to explicitly compute $\mathbb{E}[X_{t,i}Y_{t,i} | G_{t,i-1}]$. First, since $\mathbb{E}[Y_{t,i} | G_{t,i-1}] = 0$,

$$\begin{aligned} \mathbb{E}[X_{t,i}Y_{t,i} | G_{t,i-1}] &= \mathbb{E}\left[\frac{1}{n^{\frac{1}{2}+m\xi}} \frac{N_m(t,i)}{W_{t,i-1}} Y_{t,i} \middle| G_{t,i-1}\right] \\ &= \frac{1}{n^{1+m\xi}} \frac{1}{W_{t,i-1}} \mathbb{E}\left[\frac{N_m(t,i)}{D_{t,i}+\delta_0} \middle| G_{t,i-1}\right] \\ &\quad - \frac{1}{n^{1+m\xi}} \frac{1}{W_{t,i-1}} \frac{t}{S_{t,i-1}} \mathbb{E}[N_m(t,i) | G_{t,i-1}]. \end{aligned}$$

To continue, we determine the above two (conditional) expectations. We start by computing

$$\begin{aligned} \mathbb{E}\left[\frac{N_m(t,i)}{D_{t,i}+\delta_0} \middle| G_{t,i-1}\right] &= \frac{N_m(t,i-1) - 1 + \mathbb{1}\{i=m\}}{m+\delta_0} \mathbb{P}(D_{t,i}=m | G_{t,i-1}) \\ &\quad + \sum_{k=m+1}^{\infty} \frac{N_m(t,i-1) + \mathbb{1}\{i=m\}}{k+\delta_0} \mathbb{P}(D_{t,i}=k | G_{t,i-1}) \\ &= \sum_{k=m}^{\infty} \frac{N_m(t,i-1) + \mathbb{1}\{i=m\}}{k+\delta_0} \mathbb{P}(D_{t,i}=k | G_{t,i-1}) \\ &\quad - \frac{1}{m+\delta_0} \mathbb{P}(D_{t,i}=m | G_{t,i-1}) \\ &= \frac{t}{S_{t,i-1}} (N_m(t,i-1) + \mathbb{1}\{i=m\}) - \frac{N_m(t,i-1)}{S_{t,i-1}}. \end{aligned}$$

Also, from [3],

$$\mathbb{E}[N_m(t,i) | G_{t,i-1}] = \left(1 - \frac{m+\delta_0}{S_{t,i-1}}\right) N_m(t,i-1) + \mathbb{1}\{i=m\}.$$

Combining the above, we get

$$\begin{aligned} \mathbb{E}[X_{t,i}Y_{t,i} | G_{t,i-1}] &= \frac{1}{n^{1+m\xi}} \frac{1}{W_{t,i-1}} \mathbb{E}\left[\frac{N_m(t,i)}{D_{t,i}+\delta_0} \middle| G_{t,i-1}\right] \\ &\quad - \frac{1}{n^{1+m\xi}} \frac{1}{W_{t,i-1}} \frac{t}{S_{t,i-1}} \mathbb{E}[N_m(t,i) | G_{t,i-1}] \\ &= \frac{1}{n^{1+m\xi}} \frac{1}{W_{t,i-1}} \frac{t}{S_{t,i-1}} \left[N_m(t,i-1) + \mathbb{1}\{i=m\} - \frac{N_m(t,i-1)}{t} \right] \end{aligned}$$

$$\begin{aligned}
 & - \left(1 - \frac{m + \delta_0}{S_{t,i-1}} \right) N_m(t, i - 1) - \mathbb{1}\{i = m\} \Big] \\
 & = \frac{1}{n^{1+m\xi}} \frac{N_m(t, i - 1)}{W_{t,i-1} S_{t,i-1}} \left[\frac{t}{S_{t,i-1}} (m + \delta_0) - 1 \right].
 \end{aligned}$$

To compute the limiting correlation we use that, uniformly over $i = 1, \dots, m$, as $t \rightarrow \infty$,

$$\frac{N_m(t, i)}{t} \xrightarrow{\mathbb{P}_0} p_m, \quad \frac{S_{t,i}}{t} \rightarrow 2 + \delta_0.$$

Furthermore, tedious but straightforward computations show that, uniformly over $i = 1, \dots, m$, as $t \rightarrow \infty$,

$$W_{t,i} \asymp \frac{\Lambda}{t^{m\xi}},$$

where, crucially, Λ is not a function of t . Hence, as $t \rightarrow \infty$,

$$\mathbb{E}[X_{t,i} Y_{t,i} | G_{t,i-1}] = \frac{1}{n^{1+m\xi}} \frac{t^{m\xi}}{\Lambda} \frac{p_m}{(2m + \delta_0)} (\xi - 1)(1 + o_{\mathbb{P}}(1)).$$

Summing these terms and using $\sum_{t=1}^n t^{x-1} \asymp n^x/x$ as $n \rightarrow \infty$, gives

$$\begin{aligned}
 (29) \quad & \sum_{t=2}^n \sum_{i=1}^m \mathbb{E}[X_{t,i} Y_{t,i} | G_{t,i-1}] \xrightarrow{\mathbb{P}_0} \frac{1}{\Lambda} \frac{m p_m}{(2m + \delta_0)} \frac{\xi - 1}{1 + m\xi} \\
 & = - \frac{m^2}{(\delta_0 + m(2 + m + \delta_0))^2} \frac{1}{\Lambda}.
 \end{aligned}$$

In conclusion, putting together (27), (28), and (29) and applying the MMCLT (see, e.g., [16]) gives, as $n \rightarrow \infty$,

$$(30) \quad \sqrt{n} \left(\left(\frac{N_m(n)/n}{\iota'_n(\delta_0)} \right) - \left(\frac{p_m}{\iota'(\delta_0)} \right) \right) \xrightarrow{D} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tilde{\Sigma} \right),$$

where

$$\tilde{\Sigma} := \begin{pmatrix} \frac{1}{\Lambda^2} \frac{m^2(m + \delta_0)(1 + m + \delta_0)(2m + \delta_0)}{(\delta_0 + 2m(1 + m + \delta_0))(\delta_0 + m(2 + m + \delta_0))^2} - \frac{1}{\Lambda} \frac{m^2}{(\delta_0 + m(2 + m + \delta_0))^2} \\ - \frac{1}{\Lambda} \frac{m^2}{(\delta_0 + m(2 + m + \delta_0))^2} \quad v(\delta_0, m) \end{pmatrix}.$$

Recall that $v(\delta_0, m)$ is given in (8). Note that the above central limit theorem result is not for $(N_m(n)/n, \hat{\delta}_n)$. This is because the martingale difference sequences are defined as a linear transformation of $(N_m(n)/n, \hat{\delta}_n)$. Therefore, next we apply a linear transformation to (30) to obtain a central limit theorem for $(N_m(n)/n, \hat{\delta}_n)$. To this end, observe that (see [23, proof of Theorem 2]),

$$\sqrt{n}(\hat{\delta}_n - \delta_0) = -\sqrt{n} \frac{(\iota'_n(\delta_0) - \iota'(\delta_0))}{\iota''_n(\bar{\delta}_n)},$$

where $\bar{\delta}_n$ lies between δ_0 and $\hat{\delta}_n$, and so $\iota''_n(\bar{\delta}_n) \xrightarrow{\mathbb{P}_0} \iota''(\delta_0) = -v(\delta_0, m)$. Using the above we obtain

$$\begin{aligned}
 \sqrt{n}(N_m(n)/n - p_m) & = \sqrt{n} \Lambda \left(\frac{N_m(n)/n}{\Lambda} - \frac{p_m}{\Lambda} \right) \\
 \sqrt{n}(\hat{\delta}_n - \delta_0) & = \frac{\sqrt{n}}{v(\delta_0, m)} (\iota'_n(\delta_0) - \iota'(\delta_0)).
 \end{aligned}$$

In conclusion, using Slutsky's lemma, as $n \rightarrow \infty$,

$$\sqrt{n} \left(\begin{pmatrix} N_m(n)/n \\ \hat{\delta}_n \end{pmatrix} - \begin{pmatrix} p_m \\ \delta_0 \end{pmatrix} \right) \xrightarrow{D} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma(\delta_0, m) \right),$$

where

$$\Sigma(\delta_0, m) := \begin{pmatrix} \frac{m^2(m+\delta)(1+m+\delta)(2m+\delta)}{(\delta+2m(1+m+\delta))(\delta+m(2+m+\delta))^2} - \frac{m^2}{(\delta+m(2+m+\delta))^2} \frac{1}{v(\delta_0, m)} & \\ -\frac{m^2}{(\delta+m(2+m+\delta))^2} \frac{1}{v(\delta_0, m)} & \frac{1}{v(\delta_0, m)} \end{pmatrix}.$$

□

The proof of (11) is now a simple consequence of Proposition 9.1 together with the delta method. Define the function $h(x, y) := x - p_m(y)$. The gradient of $h(\cdot, \cdot)$ is given by

$$\nabla h(x, y) = (1, -p'_m(y)) = \left(1, \frac{m^2}{(y + m(2 + m + y))^2} \right),$$

where $p'_m(y)$ is given in (23). Using $h(\cdot, \cdot)$, we can rewrite

$$(31) \quad \frac{Q(G_n)}{\sqrt{n}} = \frac{N_m(n) - np_m(\hat{\delta}_n)}{\sqrt{n}} = \sqrt{n} h(N_m(G_n)/n, \hat{\delta}_n).$$

Hence, by applying the delta method we get that (31) converges in distribution as $n \rightarrow \infty$ to a normal distribution with variance

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Var} \left(\frac{N_m(G_n) - np_m(\hat{\delta}_n)}{\sqrt{n}} \right) &= (\nabla h(p_m(\delta_0), \delta_0))^T \Sigma(\delta_0, m) (\nabla h(p_m(\delta_0), \delta_0)) \\ &= \frac{m^2(m+\delta_0)(1+m+\delta_0)(2m+\delta_0)}{(\delta_0+2m(1+m+\delta_0))(\delta_0+m(2+m+\delta_0))^2} - \frac{m^4}{v(\delta_0, m)(\delta_0+m(2+m+\delta_0))^4}. \end{aligned}$$

where $\Sigma(\delta_0, m)$ and $v(\delta_0, m)$ are given by (26) and (8), respectively. □

9.2. Asymptotic normality under the alternative hypothesis. The main insight is that most of the contribution for the asymptotic distribution of the degree counts is due to the attachment process up to the changepoint. In fact, the asymptotic distribution of $(N_m(n), N_{m+1}(n), \dots)$ is normal, both under the null and alternative models, with exactly the same covariance structure but with different means. Specifically, in [3] it was shown that under the null model

$$\left(\frac{N_k(n) - \mathbb{E}_0[N_k(n)]}{\sqrt{n}} \right)_{k \geq m} \xrightarrow{D} (Z_k)_{k \geq m},$$

as $n \rightarrow \infty$, where the right-hand side is a zero-mean Gaussian process with covariance given in [3, Theorem 2.5], and the notation $(X_k(n))_{k \geq m} \xrightarrow{D} (Z_k)_{k \geq m}$ means that for any $k \geq m$ we have $(X_m(n), \dots, X_k(n)) \xrightarrow{D} (Z_m, \dots, Z_k)$ as $n \rightarrow \infty$ (i.e., the infinite vector converges in the product topology). The following lemma generalizes this result to our alternative model, immediately implying (12). Furthermore, it provides a stepping stone towards the proof of (13):

LEMMA 9.2. *Under the alternative model with $\gamma \in (0, 1)$*

$$\left(\frac{N_k(n) - \mathbb{E}_1[N_k(n)]}{\sqrt{n}} \right)_{k \geq m} \xrightarrow{D} (Z_k)_{k \geq m},$$

as $n \rightarrow \infty$ where the right-hand side is a zero-mean Gaussian process with covariance given in [3, Theorem 2.5].

PROOF. Let $k \geq m$ be arbitrary. We must show that the standardized version of $(N_m(n), \dots, N_k(n))$ is asymptotically normal with the correct covariance structure. Let us first inspect the asymptotic marginal distributions of $N_k(n)$, as this illustrates the main premise of the argument, that can be extended easily by using an application of the Cramér-Wold device. Proceed by using the following decomposition

$$(32) \quad \begin{aligned} \frac{1}{\sqrt{n}} (N_k(n) - \mathbb{E}_1[N_k(n)]) &= \frac{1}{\sqrt{n}} (\mathbb{E}_1[N_k(n) | G_{\tau_n}] - \mathbb{E}_1[N_k(n)]) \\ &\quad + \frac{1}{\sqrt{n}} (N_k(n) - \mathbb{E}_1[N_k(n) | G_{\tau_n}]) . \end{aligned}$$

The second term in (32) can be dealt with conveniently using the Azuma-Hoeffding inequality. Define the Doob martingale, for $t \in [n] \setminus [\tau_n]$,

$$M_k(t) = \mathbb{E}_1[N_k(n) | G_t] .$$

As used before we know that almost surely $|M_k(t) - M_k(t-1)| \leq 2m$. Therefore, by the Azuma-Hoeffding inequality,

$$\mathbb{P}_1(|N_k(n) - \mathbb{E}_1[N_k(n) | G_{\tau_n}]| \geq x) \leq 2e^{-\frac{x^2}{8m^2(n-\tau_n)}} .$$

Thus, the second term in (32) is $O_{\mathbb{P}_1}(\sqrt{n-\tau_n}/\sqrt{n}) = O_{\mathbb{P}_1}(n^{(\gamma-1)/2}) = o(1)$ when $\gamma < 1$.

We now shift the focus to the first term in (32). Define

$$N_k(\tau_n, n) = \sum_{v \in [\tau_n]} \mathbb{1}\{D_v(n) = k\} .$$

In words, this is the number of vertices of degree k that were added to the graph G_n up to the changepoint. Note that

$$N_k(n) = N_k(\tau_n, n) + \sum_{v \in [n] \setminus [\tau_n]} \mathbb{1}\{D_v(n) = k\} .$$

Given the affine nature of the preferential attachment function (after time τ_n) we know that $\sum_{v \in [n] \setminus [\tau_n]} \mathbb{1}\{D_v(n) = k\}$ is independent of G_{τ_n} . Therefore, we can simplify the first term in (32) as

$$\frac{1}{\sqrt{n}} (\mathbb{E}_1[N_k(n) | G_{\tau_n}] - \mathbb{E}_1[N_k(n)]) = \frac{1}{\sqrt{n}} (\mathbb{E}_1[N_k(\tau_n, n) | G_{\tau_n}] - \mathbb{E}_1[N_k(\tau_n, n)]) .$$

At this point, it is useful to write $\mathbb{E}_1[N_k(\tau_n, n) | G_{\tau_n}]$ in a slightly more explicit way. Note that

$$\begin{aligned} \mathbb{E}_1[N_k(\tau_n, n) | G_{\tau_n}] &= \mathbb{E}_1 \left[\sum_{v \in [\tau_n]} \mathbb{1}\{D_v(n) = k\} \middle| G_{\tau_n} \right] \\ &= \sum_{v \in [\tau_n]} \sum_{j=m}^k \mathbb{1}\{D_v(\tau_n) = j\} \underbrace{\mathbb{P}_1(D_v(n) = k | D_v(\tau_n) = j)}_{:= p_{j,k}(\tau_n, n)} \\ &= \sum_{j=m}^k N_j(\tau_n) p_{j,k}(\tau_n, n) . \end{aligned}$$

Using this we immediately see that

$$\mathbb{E}_1[N_k(\tau_n, n)] = \sum_{j=m}^k \mathbb{E}_1[N_j(\tau_n)] p_{j,k}(\tau_n, n) = \sum_{j=m}^k \mathbb{E}_0[N_j(\tau_n)] p_{j,k}(\tau_n, n) .$$

In conclusion

$$\begin{aligned}
& \frac{1}{\sqrt{n}} (\mathbb{E}_1 [N_k(n) | G_{\tau_n}] - \mathbb{E}_1 [N_k(n)]) \\
&= \frac{1}{\sqrt{n}} \sum_{j=m}^k (N_j(\tau_n) - \mathbb{E}_0[N_j(\tau_n)]) p_{j,k}(\tau_n, n) \\
(33) \quad &= \sum_{j=m}^k \frac{1}{\sqrt{\tau_n}} (N_j(\tau_n) - \mathbb{E}_0[N_j(\tau_n)]) \sqrt{\frac{\tau_n}{n}} p_{j,k}(\tau_n, n) .
\end{aligned}$$

To proceed all that is needed is to characterize $p_{j,k}(\tau_n, n)$. Lemma 7.2 immediately provides the necessary result, specifically

$$p_{k,k}(\tau_n, n) = 1 - (1 + o(1)) c n^{\gamma-1} m \frac{k + \delta_1}{2m + \delta_1} = 1 + o(1)$$

and for $j < k$

$$p_{j,k}(\tau_n, n) \leq \mathbb{P}_1 (D_v(n) > k | D_v(\tau_n) = j) = o(1) .$$

In conclusion, since $\sqrt{\tau_n/n} \rightarrow 1$ and $p_{k,k}(\tau_n, n) \rightarrow \mathbb{1}\{j = k\}$ as $n \rightarrow \infty$ we conclude that (33) converges to the same normal distribution as $\frac{1}{\sqrt{n}} (N_k(n) - \mathbb{E}_0[N_k(n)])$ under the null model.

Owing to the linearity of (33), the same argument also shows that any finite linear combination of the (centered and rescaled) elements of $(N_m(n), \dots, N_k(n))$ is asymptotically normal with the appropriate variance. An application of the Cramér-Wold device then shows that, for any $k \geq m$, this vector converges in distribution to the desired finite-dimensional multi-variate Gaussian distribution. This concludes the proof. \square

With this lemma at hand (12) is immediate. The second result (13) is, however, not a trivial consequence of the lemma, as the convergence in the product topology in Lemma 9.2 is unfortunately not sufficient to obtain the final result.

To formally show (13), begin by recalling that $\iota'_n(\hat{\delta}_n) = 0$. Expanding $\delta \mapsto \iota'_n(\delta)$ around δ_0 we get

$$\hat{\delta}_n - \delta_0 = -\frac{\iota'_n(\delta_0)}{\iota''_n(\bar{\delta}_n)} ,$$

where $\bar{\delta}_n = \delta_0 + \bar{\zeta}_n(\hat{\delta}_n - \delta_0)$ for some $\bar{\zeta}_n \in [0, 1]$. Therefore, and considering also a Taylor expansion of $p_m(\hat{\delta}_n)$, we conclude that

$$\begin{aligned}
Q(g_n) &= N_m(n) - n p_m(\hat{\delta}_n) \\
&= N_m(n) - n p_m(\delta_0) + n \frac{p'_m(\check{\delta}_n)}{\iota''_n(\bar{\delta}_n)} \iota'_n(\delta_0) ,
\end{aligned}$$

where $\check{\delta}_n = \delta_0 + \check{\zeta}_n(\hat{\delta}_n - \delta_0)$ for some $\check{\zeta}_n \in [0, 1]$. Since $\hat{\delta}_n \xrightarrow{\mathbb{P}_1} \delta_0$ as $n \rightarrow \infty$, and using the uniform convergence results of ι'' from [23] (as already used in the proof of Proposition 8.1) we conclude that

$$\frac{p'_m(\check{\delta}_n)}{\iota''_n(\bar{\delta}_n)} \xrightarrow{\mathbb{P}_1} \frac{p'_m(\delta_0)}{\iota''(\delta_0)} \neq 0 .$$

To aid the presentation we rewrite $Q(G_n)$ as

$$(34) \quad \begin{aligned} Q(G_n) = & N_m(n) - np_m(\delta_0) + n \frac{p'_m(\delta_0)}{\iota''(\delta_0)} \iota'_n(\delta_0) \\ & + n \left(\frac{p'_m(\check{\delta}_0)}{\iota''_n(\delta_0)} - \frac{p'_m(\delta_0)}{\iota''(\delta_0)} \right) \iota'_n(\delta_0) . \end{aligned}$$

We argue that the last term is negligible, and therefore it suffices to characterize the asymptotic normality of the first two terms. Note that $N_m(n)$ and $\iota'_n(\delta_0)$ are not independent. To avoid unnecessarily cluttering the presentation we focus first on the asymptotic normality of $\iota'_n(\delta_0)$ and then argue that extending the analysis to the joint normality is straightforward.

The following result generalizes [23, Lemma 7] to the alternative model:

LEMMA 9.3. *Recall the definition of v in (8). Under the alternative model,*

$$\sqrt{n}(\iota'_n(\delta_0) - \mathbb{E}_1[\iota'_n(\delta_0)]) \xrightarrow{D} \mathcal{N}(0, v(\delta_0, m)) .$$

PROOF. We proceed similarly as in the proof of Lemma 9.2 by isolating the contributions of the vertices that join after the changepoint as follows

$$(35) \quad \begin{aligned} \sqrt{n}(\iota'_n(\delta_0) - \mathbb{E}_1[\iota'_n(\delta_0)]) = & \frac{\sqrt{n}}{n+1} \sum_{k \geq m} \frac{1}{k + \delta_0} (N_{>k}(n) - \mathbb{E}_1[N_{>k}(n)|G_{\tau_n}]) \\ & + \frac{\sqrt{n}}{n+1} \sum_{k \geq m} \frac{1}{k + \delta_0} (\mathbb{E}_1[N_{>k}(n)|G_{\tau_n}] - \mathbb{E}_1[N_{>k}(n)]) . \end{aligned}$$

Following the exact same argument as in the proof of Lemma A.2 the first term on the right-hand side is sufficiently small since

$$\mathbb{P}_1(n|\iota'_n(\delta_0) - \mathbb{E}_1[\iota'_n(\delta_0)|G_{\tau_n}]| \geq x) \leq 2 \exp \left(\frac{x^2}{2(n - \tau_n)c_{n,m}^2} \right) ,$$

where $c_{n,m} = 2m \log(n)(1 + o(1))$, and thus

$$\sqrt{n}(\iota'_n(\delta_0) - \mathbb{E}_1[\iota'_n(\delta_0)|G_{\tau_n}]) = O_{\mathbb{P}}(n^{\gamma-1} \log(n)) = o_{\mathbb{P}_1}(1) .$$

For the following term, it is convenient to introduce a convenient notation, similarly as done in the proof of Lemma 9.2:

$$N_{>k}(\tau_n, n) := \sum_{v \in [\tau_n]} \mathbb{1}\{D_v(n) > k\} .$$

Note that

$$\mathbb{E}_1[N_{>k}(n)|G_{\tau_n}] - \mathbb{E}_1[N_{>k}(n)] = \mathbb{E}_1[N_{>k}(\tau_n, n)|G_{\tau_n}] - \mathbb{E}_1[N_{>k}(\tau_n, n)] ,$$

and also

$$\mathbb{E}_1[N_{>k}(n)|G_{\tau_n}] = N_{>k}(\tau_n) + \sum_{v \in [\tau_n]} \mathbb{1}\{D_v(\tau_n) \leq k\} \underbrace{\mathbb{P}_1(D_v(n) > k | D_v(\tau_n) \leq k)}_{:= q_k(\tau_n, n)} .$$

With this in hand, the second term on the right-hand side of (35) is

$$\begin{aligned}
& \frac{\sqrt{n}}{n+1} \sum_{k \geq m} \frac{1}{k + \delta_0} (\mathbb{E}_1[N_{>k}(\tau_n, n) | G_{\tau_n}] - \mathbb{E}_1[N_{>k}(\tau_n, n)]) \\
&= \frac{\sqrt{n\tau_n}}{n+1} \frac{1}{\sqrt{\tau_n}} \sum_{k \geq m} \frac{1}{k + \delta_0} (N_{>k}(\tau_n) - \mathbb{E}_1[N_{>k}(\tau_n)]) \\
&\quad + \frac{\sqrt{n}}{n+1} \sum_{k \geq m} \frac{q_k(\tau_n, n)}{k + \delta_0} \sum_{v \in [\tau_n]} (\mathbb{1}\{D_v(\tau_n) \leq k\} - \mathbb{P}_1(D_v(\tau_n) \leq k)) \\
(36) \quad &= \frac{\sqrt{n\tau_n}}{n+1} \frac{1}{\sqrt{\tau_n}} \sum_{k \geq m} \frac{1}{k + \delta_0} (N_{>k}(\tau_n) - \mathbb{E}_0[N_{>k}(\tau_n)])
\end{aligned}$$

$$(37) \quad - \frac{\sqrt{n}}{n+1} \sum_{k \geq m} \frac{q_k(\tau_n, n)}{k + \delta_0} \sum_{v \in [\tau_n]} (\mathbb{1}\{D_v(\tau_n) > k\} - \mathbb{P}_0(D_v(\tau_n) > k)) .$$

The term (36) converges in law to $\mathcal{N}(0, v(\delta_0, m))$ by [23, Lemma 7]. To finalize our argument, we are left to prove that the term (37) is negligible. We do this by rewriting it as a Doob martingale difference, similarly as was done earlier in the proof. Let

$$M_t := \frac{\sqrt{n}}{n+1} \mathbb{E}_1 \left[\sum_{k \geq m} \frac{q_k(\tau_n, n)}{k + \delta_0} N_{>k}(\tau_n) \middle| G_t \right] .$$

With this notation (37) is precisely $M_{\tau_n} - M_1$. Furthermore, the martingale differences are bounded by

$$\begin{aligned}
|M_t - M_{t-1}| &\leq \frac{\sqrt{n}}{n+1} \sum_{k \geq m} \frac{2m}{k + \delta_0} q_k(\tau_n, n) \\
&= \frac{2m\sqrt{n}}{(n+1)\tau_n} \sum_{k \geq m} \sum_{v \in [\tau_n]} \frac{1}{k + \delta_0} \frac{\mathbb{P}_1(D_v(n) > k, D_v(\tau_n) \leq k)}{\mathbb{P}_1(D_v(\tau_n) \leq k)} \\
&\leq \frac{2m\sqrt{n}}{(n+1)\tau_n} \sum_{k \geq m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_1(D_v(n) > k, D_v(\tau_n) \leq k)}{k + \delta_0} \frac{1}{\mathbb{P}_0(D_v(\tau_n) \leq m)} \\
&\leq \frac{2mn^{-3/2}}{p_m(\delta_0)} (1 + o(1)) \sum_{k \geq m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_1(D_v(n) > k, D_v(\tau_n) \leq k)}{k + \delta_0} .
\end{aligned}$$

The double-summation is controlled in the proof of Proposition 8.2, and it is the sum of the terms in expressions (41) and (42). In conclusion

$$|M_t - M_{t-1}| \leq \underbrace{\frac{2m^2c}{(2m + \delta_0)p_m(\delta_0)}}_{:= \text{const}(m, \delta_0, c)} n^{\gamma-3/2} (1 + o(1)) .$$

Based on this and using the Azuma-Hoeffding inequality we conclude that, for any $x \geq 0$

$$\begin{aligned}
\mathbb{P}(|M_{\tau_n} - M_1| \geq x) &\leq 2 \exp \left\{ - \frac{x^2}{(1 + o(1))\tau_n (\text{const}(m, \delta_0, c)n^{\gamma-3/2})^2} \right\} \\
&\leq 2 \exp \left\{ - \frac{x^2}{(1 + o(1))\text{const}^2(m, \delta_0, c)n^{2(\gamma-1)}} \right\} .
\end{aligned}$$

This implies that $M_{\tau_n} - M_1 = O_{\mathbb{P}_1}(n^{\gamma-1}) = o_{\mathbb{P}_1}(1)$, showing that (37) is negligible, and concluding the proof of the lemma. \square

Lemma 9.3 suffices to show that the last term in (34) gives a negligible contribution to the limit distribution of $(Q(G_n) - \mathbb{E}_1[Q(G_n)])/\sqrt{n}$, since

$$\begin{aligned} \sqrt{n} \left(\frac{p'_m(\check{\delta}_0)}{\iota''_n(\check{\delta}_0)} - \frac{p'_m(\delta_0)}{\iota''(\delta_0)} \right) \iota'_n(\delta_0) - \mathbb{E}_1 \left[\sqrt{n} \left(\frac{p'_m(\check{\delta}_0)}{\iota''_n(\check{\delta}_0)} - \frac{p'_m(\delta_0)}{\iota''(\delta_0)} \right) \iota'_n(\delta_0) \right] \\ = o_{\mathbb{P}_1}(1) O_{\mathbb{P}_1}(1) = o_{\mathbb{P}_1}(1). \end{aligned}$$

In conclusion

$$\frac{Q(G_n) - \mathbb{E}_1[G_n]}{\sqrt{n}} = \frac{N_m(n) - \mathbb{E}_1[N_m(n)]}{\sqrt{n}} + \sqrt{n} \frac{p'_m(\delta_0)}{\iota''(\delta_0)} (\iota'_n(\delta_0) - \mathbb{E}_1[\iota'_n(\delta_0)]) + o_{\mathbb{P}_1}(1).$$

Since $\iota'_n(\delta_0)$ and $N_m(n)$ are not independent we cannot directly rely on Lemma 9.3 and 9.2 to obtain the final result. However, using exactly the same type of argument leads to the following sequence of statements:

$$\begin{aligned} & \frac{Q(G_n) - \mathbb{E}_1[Q(G_n)]}{\sqrt{n}} \\ &= o_{\mathbb{P}_1}(1) + \frac{\mathbb{E}_1[N_m(n) | G_{\tau_n}] - \mathbb{E}_1[N_m(n)]}{\sqrt{n}} \\ & \quad + \frac{\sqrt{n}}{n+1} \sum_{k \geq m} \frac{1}{k + \delta_0} (\mathbb{E}_1[N_{>k}(n) | G_{\tau_n}] - \mathbb{E}_1[N_{>k}(n)]) \\ &= o_{\mathbb{P}_1}(1) + \frac{N_m(\tau_n) - \mathbb{E}_0[N_m(\tau_n)]}{\sqrt{n}} \\ & \quad + \frac{\sqrt{n}}{n+1} \sum_{k \geq m} \frac{1}{k + \delta_0} (N_{>k}(\tau_n) - \mathbb{E}_0[N_{>k}(\tau_n)]) \\ & \xrightarrow{D} \mathcal{N}(0, w(\delta_0, m) + u(\delta_0, m)), \end{aligned}$$

where the last statement follows from the joint convergence, after appropriate rescaling, of $N_m(\tau_n)$ and ι'_{τ_n} , which is guaranteed by (30) (cf. the definition of ι'_n in (5)).

APPENDIX A: PROOF OF AUXILIARY RESULTS

A.1. Proof of Proposition 8.2. Recall that $\delta \in [\delta_{\min}, \delta_{\max}]$. Begin by noting that

$$(n+1) (\mathbb{E}_1[\iota'_n(\delta)] - \mathbb{E}_0[\iota'_n(\delta)]) = \sum_{k \geq m} \frac{\mathbb{E}_1[N_{>k}(n)] - \mathbb{E}_0[N_{>k}(n)]}{k + \delta}.$$

The numerator in the above summand can be decomposed in a similar manner as used for the proof of Proposition 7.1:

$$\begin{aligned} & \mathbb{E}_1[N_{>k}(n)] - \mathbb{E}_0[N_{>k}(n)] \\ &= \mathbb{E}_1[N_{>k}(n)] - \mathbb{E}_1[N_{>k}(\tau_n)] + \mathbb{E}_0[N_{>k}(\tau_n)] - \mathbb{E}_0[N_{>k}(n)] \\ &= (\mathbb{E}_1[N_{>k}(n)] - \mathbb{E}_1[N_{>k}(\tau_n)]) - (\mathbb{E}_0[N_{>k}(n)] - \mathbb{E}_0[N_{>k}(\tau_n)]). \end{aligned}$$

The first equality holds as the law of G_{τ_n} is the same under the null and alternative models. Therefore,

$$(38) \quad (n+1) (\mathbb{E}_1[\iota'_n(\delta)] - \mathbb{E}_0[\iota'_n(\delta)]) = \sum_{k \geq m} \frac{\mathbb{E}_1[N_{>k}(n)] - \mathbb{E}_1[N_{>k}(\tau_n)]}{k + \delta} - \sum_{k \geq m} \frac{\mathbb{E}_0[N_{>k}(n)] - \mathbb{E}_0[N_{>k}(\tau_n)]}{k + \delta}.$$

The treatment of the two terms is entirely analogous and it is done simultaneously.

For the rest of the proof, let $\ell \in \{0, 1\}$. Clearly

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

Like in the proof of Proposition 7.1 we distinguish the behavior of “old” vertices (that arrived before the change-point) from the remaining vertices (the “new” vertices). The contribution of the latter plays an insignificant role, as we see next. Note that, since $k \geq m$ and for $v \in [n] \setminus [\tau_n]$, the event $D_v(n) > k$ is only possible when there is a vertex $v' > v$ that attached to v . Referring to (2) we see that the probability of this happening is at most $(m + \delta_\ell) / ((2m + \delta_\ell)\tau_n - 2m)$. Since there are at most $m(n - \tau_n)$ possible edges that could attach we get the simple bound

$$\sum_{v \in [n] \setminus [\tau_n]} \mathbb{P}_\ell(D_v(n) > k) \leq m(n - \tau_n)^2 \frac{m + \delta_\ell}{(2m + \delta_\ell)\tau_n - 2m} = O(n^{2\gamma-1}).$$

Using that result and the fact that the largest degree in G_n is at most nm , this implies that

$$\sum_{k=m}^{nm} \sum_{v \in [n] \setminus [\tau_n]} \frac{\mathbb{P}_\ell(D_v(n) > k)}{k + \delta} \leq O(n^{2\gamma-1}) \sum_{k=m}^{nm} \frac{1}{k + \delta} \leq O(n^{2\gamma-1} \log n).$$

Therefore,

$$\begin{aligned} &\sum_{k \geq m} \frac{\mathbb{E}_\ell[N_{>k}(n)] - \mathbb{E}_\ell[N_{>k}(\tau_n)]}{k + \delta} \\ &= \sum_{k \geq m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(n) > k) - \mathbb{P}_\ell(D_v(\tau_n) > k)}{k + \delta} + O(n^{2\gamma-1} \log n) \\ &= \sum_{k \geq m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(\tau_n) \leq k, D_v(n) > k)}{k + \delta} + o(n^\gamma) \\ (39) \quad &= \sum_{k \geq m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(\tau_n) = k, D_v(n) > k)}{k + \delta} \end{aligned}$$

$$(40) \quad + \sum_{k > m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(\tau_n) < k, D_v(n) > k)}{k + \delta} + o(n^\gamma).$$

The bulk of the analysis is therefore the characterization of the double summations above. This is somewhat delicate, and requires a good understanding of the behavior of $\mathbb{P}_\ell(D_v(\tau_n) \leq$

$k, D_v(n) > k$) for $v \in [\tau_n]$ and $k \geq m$. For an arbitrary vertex $v \in [\tau_n]$ and “small” k we know that, most likely, the degree of the vertex will not change in G_n . Most of the contribution in the above expression will therefore be due to the attachment of a single edge to v . This reasoning does not apply when k is “large”, but in that case the denominator $k + \delta$ is large enough to make the contribution to the above summations negligible.

Note that up to time τ_n both null model and alternative models coincide. Therefore $\frac{1}{\tau_n} \sum_{v \in [\tau_n]} \mathbb{P}_\ell(D_v(\tau) = k) = p_k(\delta_0)(1 + o(1))$, regardless of the value of ℓ . To characterize the term (39) we proceed by truncating that series and using Lemma 7.2. Define an auxiliary series $b_n = \lceil n^{1-\gamma}/\log n \rceil$. This is a divergent sequence of integers such that $b_n = o(n^{1-\gamma})$. Then

$$\begin{aligned}
 & \sum_{k \geq m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(\tau_n) = k, D_v(n) > k)}{k + \delta} \\
 &= \sum_{k \geq m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(n) - D_v(\tau_n) > 0 \mid D_n(\tau_n) = k) \mathbb{P}_\ell(D_v(\tau_n) = k)}{k + \delta} \\
 &= \sum_{k=m}^{b_n} \sum_{v \in [\tau_n]} (1 + o(1)) c n^{\gamma-1} m \frac{k + \delta_\ell}{2m + \delta_\ell} \frac{\mathbb{P}_\ell(D_v(\tau_n) = k)}{k + \delta} \\
 &\quad + \sum_{k=b_n+1}^{\infty} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(n) - D_v(\tau_n) > 0 \mid D_n(\tau_n) = k) \mathbb{P}_\ell(D_v(\tau_n) = k)}{k + \delta} \\
 (41) \quad &= (1 + o(1)) c n^{\gamma-1} \frac{m}{2m + \delta_\ell} \tau_n \sum_{k=m}^{b_n} \frac{k + \delta_\ell}{k + \delta} p_k(\delta_0)
 \end{aligned}$$

$$(42) \quad + \sum_{k=b_n+1}^{\infty} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(n) - D_v(\tau_n) > 0 \mid D_n(\tau_n) = k) \mathbb{P}_\ell(D_v(\tau_n) = k)}{k + \delta}.$$

The series in (41) is convergent. For the series in (42), note that

$$\begin{aligned}
 & \sum_{k=b_n+1}^{\infty} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(n) - D_v(\tau_n) > 0 \mid D_n(\tau_n) = k) \mathbb{P}_\ell(D_v(\tau_n) = k)}{k + \delta} \\
 &\leq \sum_{k=b_n+1}^{\infty} \sum_{v \in [\tau_n]} \frac{k^2}{b_n^2} \frac{\mathbb{P}_\ell(D_v(\tau_n) = k)}{k + \delta} \\
 &= \frac{1}{b_n^2} \sum_{k=b_n+1}^{\infty} \sum_{v \in [\tau_n]} \mathbb{E}_\ell \left(\frac{D_v^2(\tau_n)}{D_v(\tau_n) + \delta} \mathbb{1}\{D_v(\tau_n) = k\} \right) \\
 &\leq \frac{1}{b_n^2} \sum_{k=1}^{\infty} \sum_{v \in [\tau_n]} \mathbb{E}_\ell \left(\underbrace{\frac{D_v(\tau_n)}{D_v(\tau_n) + \delta}}_{\leq C_m} D_v(\tau_n) \mathbb{1}\{D_v(\tau_n) = k\} \right) \\
 &= \frac{C_m}{b_n^2} \mathbb{E}_\ell \left(\sum_{v \in [\tau_n]} D_v(\tau_n) \right)
 \end{aligned}$$

$$= \frac{C_m}{b_n^2} 2m\tau_n = O(n^{2\gamma-1} \log n) = o(n^\gamma),$$

where $C_m := \frac{m}{m+\delta_{\min}}$. In conclusion,

$$(43) \quad \sum_{k \geq m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(\tau_n) = k, D_v(n) > k)}{k + \delta} = (1 + o(1)) c n^{\gamma-1} \frac{m}{2m + \delta_\ell} \tau_n \sum_{k \geq m} \frac{k + \delta_\ell}{k + \delta} p_k(\delta_0).$$

The characterization of (40) is significantly more delicate. Note that

$$(44) \quad \begin{aligned} & \sum_{k > m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(\tau_n) < k, D_v(n) > k)}{k + \delta} \\ &= \sum_{v \in [\tau_n]} \mathbb{E}_\ell \left[\sum_{k > m} \frac{1}{k + \delta} \mathbb{1}\{D_v(\tau_n) < k, D_v(n) > k\} \right] \\ &\leq \sum_{v \in [\tau_n]} \mathbb{E}_\ell \left[\sum_{k > m} \frac{1}{D_v(\tau_n) + 1 + \delta} \mathbb{1}\{D_v(\tau_n) < k, D_v(n) > k\} \right] \\ &= \sum_{v \in [\tau_n]} \mathbb{E}_\ell \left[\frac{1}{D_v(\tau_n) + 1 + \delta} \mathbb{1}\{D_v(n) - D_v(\tau_n) \geq 2\} \sum_{k=D_v(\tau_n)+1}^{D_v(n)-1} 1 \right] \\ &= \sum_{v \in [\tau_n]} \mathbb{E}_\ell \left[\mathbb{1}\{D_v(n) - D_v(\tau_n) \geq 2\} \frac{D_v(n) - D_v(\tau_n) - 1}{D_v(\tau_n) + 1 + \delta} \right] \\ &= \sum_{v \in [\tau_n]} \mathbb{E}_\ell \left[\mathbb{1}\{D_v(n) - D_v(\tau_n) \geq 1\} \frac{D_v(n) - D_v(\tau_n) - 1}{D_v(\tau_n) + 1 + \delta} \right]. \end{aligned}$$

The key quantity to control is the expectation in (44). Let σ_v denote the first time after τ_n when an edge is attached to vertex v . With this in hand we can bound (44) as

$$(45) \quad \begin{aligned} & \sum_{k > m} \sum_{v \in [\tau_n]} \frac{\mathbb{P}_\ell(D_v(\tau_n) < k, D_v(n) > k)}{k + \delta} \\ &\leq \sum_{v \in [\tau_n]} \sum_{s \in [n] \setminus [\tau_n]} \mathbb{E}_\ell \left[\mathbb{1}\{\sigma_v = s\} \frac{D_v(n) - D_v(s)}{D_v(\tau_n) + 1 + \delta} \right]. \end{aligned}$$

The following lemma allows us to bound (45):

LEMMA A.1. For $\ell \in \{0, 1\}$ and $t \geq \tau > \tau_n$

$$\begin{aligned} & \mathbb{E}_\ell [D_v(t) + \delta_\ell \mid D_v(\tau)] \\ &= (D_v(\tau) + \delta_\ell) \prod_{j=1}^{t-\tau} \prod_{i=1}^m \left(1 + \frac{1}{2(\tau + j - 1)m + \delta_\ell(\tau + j) + (i - 1)} \right). \end{aligned}$$

We postpone the proof of Lemma A.1 to later on. With this in hand, for any $s \in [n] \setminus [\tau_n]$,

$$\mathbb{E}_\ell [D_v(n) - D_v(s) \mid D_v(s)]$$

$$\begin{aligned}
&= \mathbb{E}_\ell [D_v(n) + \delta_\ell \mid D_v(s)] - (D_v(s) + \delta_\ell) \\
&= (D_v(s) + \delta_\ell) \left(\prod_{j \in [n] \setminus [\tau_n]} \prod_{i=1}^m \left(1 + \frac{1}{j(2m + \delta_\ell) - 2m + i - 1} \right) - 1 \right) \\
&= (1 + o(1))(D_v(s) + \delta_\ell)(n - \tau_n) \frac{m}{(2m + \delta_\ell)\tau_n} \\
&\leq O(1)(D_v(s) + \delta_\ell)n^{\gamma-1}.
\end{aligned}$$

When $s = \sigma_v$ we know that $D_v(s) \leq D_n(\tau_n) + m$ (at time s the first edge was attached to v , and therefore at most m edges were attached to v after all the intermediate steps). That means that

$$\begin{aligned}
\frac{D_v(s) + \delta_\ell}{D_v(\tau_n) + 1 + \delta} &\leq \frac{D_v(\tau_n) + m + \delta_\ell}{D_v(\tau_n) + 1 + \delta} \\
&\leq \frac{D_v(\tau_n) + m + \delta_{\max}}{D_v(\tau_n) + 1 + \delta_{\min}} \\
&\leq \frac{2m + \delta_{\max}}{m + 1 + \delta_{\min}} := \text{const},
\end{aligned}$$

where $\text{const} > 0$ is simply a constant. Therefore,

$$\begin{aligned}
&\sum_{v \in [\tau_n]} \sum_{s \in [n] \setminus [\tau_n]} \mathbb{E}_\ell \left[\mathbb{1}\{\sigma_v = s\} \frac{D_v(n) - D_v(s)}{D_v(\tau_n) + 1 + \delta} \right] \\
&= \sum_{v \in [\tau_n]} \sum_{s \in [n] \setminus [\tau_n]} \mathbb{E}_\ell \left[\frac{1}{D_v(\tau_n) + 1 + \delta} \mathbb{E}_\ell [\mathbb{1}\{\sigma_v = s\} (D_v(n) - D_v(s)) \mid D_v(s)] \right] \\
&= O(1) \sum_{v \in [\tau_n]} \sum_{s \in [n] \setminus [\tau_n]} \mathbb{E}_\ell \left[\frac{1}{D_v(\tau_n) + 1 + \delta} \mathbb{E}_\ell [\mathbb{1}\{\sigma_v = s\} (D_v(s) + \delta_\ell)n^{\gamma-1} \mid D_v(s)] \right] \\
&\leq O(n^{\gamma-1}) \sum_{v \in [\tau_n]} \sum_{s \in [n] \setminus [\tau_n]} \mathbb{P}_\ell(\sigma_v = s) \\
&= O(n^{\gamma-1}) \sum_{v \in [\tau_n]} \mathbb{P}_\ell(D_v(n) - D_v(\tau_n) \geq 1) \\
&\leq O(n^{\gamma-1}) \sum_{v \in [\tau_n]} (n - \tau_n)m \mathbb{E}_\ell \left[\frac{D_v(\tau_n) + \delta_\ell}{(2m + \delta_\ell)\tau_n + \delta_\ell} \right] \\
&= O(n^{2\gamma-2}) \mathbb{E}_\ell \left[\sum_{v \in [\tau_n]} D_v(\tau_n) + \delta_\ell \right] = O(n^{2\gamma-1}) = o(n^\gamma),
\end{aligned}$$

where the last inequality follows from the same reasoning used to obtain (43), and the last step follows since $\gamma > \frac{1}{2}$. This means that the term in (40) is of smaller order than the term in (39). In conclusion,

$$\sum_{k \geq m} \frac{\mathbb{E}_\ell[N_{>k}(n)] - \mathbb{E}_\ell[N_{>k}(\tau_n)]}{k + \delta} = (1 + o(1))cn^\gamma \frac{m}{2m + \delta_\ell} \sum_{k=m}^{\infty} \frac{k + \delta_\ell}{k + \delta} p_k(\delta_0).$$

We are now ready to go back to (38) to get

$$\begin{aligned}
& (n+1) (\mathbb{E}_1[\iota'_n(\delta)] - \mathbb{E}_0[\iota'_n(\delta)]) \\
&= \sum_{k \geq m} \frac{\mathbb{E}_1[N_{>k}(n)] - \mathbb{E}_1[N_{>k}(\tau_n)]}{k + \delta} - \sum_{k \geq m} \frac{\mathbb{E}_0[N_{>k}(n)] - \mathbb{E}_0[N_{>k}(\tau_n)]}{k + \delta} \\
&= (1 + o(1)) \frac{cmn^\gamma}{(2m + \delta_0)(2m + \delta_1)} (\delta_1 - \delta_0) \sum_{k \geq m} \frac{2m - k}{k + \delta} p_k(\delta_0) \\
&= (1 + o(1)) \frac{cmn^\gamma}{(2m + \delta_0)(2m + \delta_1)} (\delta_1 - \delta_0) \left(-1 + \sum_{k \geq m} \frac{2m + \delta}{k + \delta} p_k(\delta_0) \right),
\end{aligned}$$

as required and where in the last step we used the fact that $\sum_{k \geq m} p_k(\delta_0) = 1$. \square

A.2. Proof of Lemma A.1. Note that between t and τ the attachment function is affine with parameter δ_ℓ . Note also that the graph G_t has precisely $t + 1$ vertices and mt edges. Let us describe what happens at each one of the intermediate steps. Let v be a vertex in G_τ and let $D_v(\tau + 1, i)$ denote its degree in the graph $G_{\tau+1, i}$, where $i \in \{1, \dots, m\}$. Then,

$$\begin{aligned}
& \mathbb{E}_\ell[D_v(\tau + 1, i) + \delta_0 \mid D_v(\tau + 1, i - 1)] \\
&= D_v(\tau + 1, i - 1) + \delta_0 + \mathbb{E}_\ell[D_v(\tau + 1, i) - D_v(\tau + 1, i - 1) \mid D_v(\tau + 1, i - 1)] \\
&= D_v(\tau + 1, i - 1) + \delta_0 + \frac{D_v(\tau + 1, i - 1) + \delta_0}{2\tau m + \delta_0(\tau + 1) + (i - 1)} \\
&= (D_v(\tau + 1, i - 1) + \delta_0) \left(1 + \frac{1}{2\tau m + \delta_0(\tau + 1) + (i - 1)} \right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}_\ell[D_v(\tau + 1) + \delta_0 \mid D_v(\tau)] &= \mathbb{E}_\ell[D_v(\tau + 1, m) + \delta_0 \mid D_v(\tau + 1, 0)] \\
&= (D_v(\tau) + \delta_0) \prod_{i=1}^m \left(1 + \frac{1}{2\tau m + \delta_0(\tau + 1) + (i - 1)} \right).
\end{aligned}$$

Thus, in general, for $t > \tau$,

$$\begin{aligned}
& \mathbb{E}_\ell[D_v(t) + \delta_0 \mid D_v(\tau)] \\
&= \mathbb{E}_\ell[\mathbb{E}_\ell[D_v(t) + \delta_0 \mid D_v(t - 1)] \mid D_v(\tau)] \\
&= \mathbb{E}_\ell[\mathbb{E}_\ell[D_v(t) + \delta_0 \mid D_v(t - 1)] \mid D_v(\tau)] \\
&= \prod_{i=1}^m \left(1 + \frac{1}{2(t - 1)m + \delta_0 t + (i - 1)} \right) \mathbb{E}_\ell[D_v(t - 1) + \delta_0 \mid D_v(\tau)] \\
&\vdots \\
&= (D_v(\tau) + \delta_0) \prod_{j=1}^{t-\tau} \prod_{i=1}^m \left(1 + \frac{1}{2(\tau + j - 1)m + \delta_0(\tau + j) + (i - 1)} \right).
\end{aligned}$$

\square

A.3. Proof of Proposition 8.3. To show consistency of $\hat{\delta}_n$ note that $\iota'_n(\hat{\delta}_n) = 0$ by definition. Recalling (21) we conclude immediately that

$$\mathbb{E}[\iota'(\hat{\delta}_n)] \rightarrow 0 .$$

[23, Lemma 4] shows that ι' has a unique zero at δ_0 , and $\iota'(\delta) > 0$ for $\delta < \delta_0$ and $\iota'(\delta) < 0$ for $\delta > \delta_0$. This immediately implies that $\mathbb{E}_1[|\hat{\delta}_n - \delta_0|] \rightarrow 0$ as $n \rightarrow \infty$, proving the first assertion in the proposition.

Note also that in Proposition 8.1 we have shown that $\tilde{\delta}_n \rightarrow \delta_0$, therefore we also have $\mathbb{E}_\ell[|\hat{\delta}_n - \tilde{\delta}_n|] \rightarrow 0$. To characterize the rate of convergence of $\hat{\delta}_n$ to $\tilde{\delta}_n$ we need the following lemma:

LEMMA A.2. For $\ell \in \{0, 1\}$ and $x > 0$

$$\mathbb{P}_\ell \left(\sup_{\delta \in [\delta_{\min}, \delta_{\max}]} (n+1) |\iota'_n(\delta) - \mathbb{E}_\ell[\iota'_n(\delta)]| \geq x \right) \leq 2e^{-\frac{x^2}{2nc_{n,m}^2}} ,$$

where $c_{n,m} = \sum_{k=m}^{nm} \frac{2m}{k+\delta_{\min}}$.

PROOF. To prove this lemma we use a similar argument used in the proof of Theorem 3.1, resorting to the Azuma-Hoeffding's inequality. Note that in the expression of $\iota'_n(\delta)$ in (5) only the first term is not deterministic. Begin by constructing the Doob martingale

$$M_t(\delta) = \sum_{k \geq m} \frac{\mathbb{E}_\ell[N_{>k}(n) | G_t]}{k + \delta} ,$$

where $t \in [n]$. Clearly $M_1(\delta) = \sum_{k \geq m} \frac{\mathbb{E}_\ell[N_{>k}(n)]}{k + \delta}$ and $M_n(\delta) = \sum_{k \geq m} \frac{N_{>k}(n)}{k + \delta}$, therefore

$$M_n(\delta) - M_1(\delta) = (n+1) (\iota'_n(\delta) - \mathbb{E}_\ell[\iota'_n(\delta)]) .$$

Furthermore, at each timestep in the construction of G_n we add $2m$ edges. Therefore

$$|\mathbb{E}_\ell[N_{>k}(n) | G_t] - \mathbb{E}_\ell[N_{>k}(n) | G_{t-1}] \leq 2m ,$$

where $t \in \{2, \dots, n\}$ and $m \leq k \leq nm$. As a result,

$$|M_t(\delta) - M_{t-1}(\delta)| \leq \sum_{k=m}^{nm} \frac{2m}{k + \delta} \leq \sum_{k=m}^{nm} \frac{2m}{k + \delta_{\min}} .$$

Note that the bound on the martingale differences holds uniformly in δ . With this in hand we can simply apply the Azuma-Hoeffding's inequality to get the desired result. \square

Note that $c_{n,m} = 2m(1 + o(1) \log n)$ as $n \rightarrow \infty$. Let a_n be an arbitrary sequence satisfying $a_n = \omega(\sqrt{n} \log n)$. The above lemma tells us that

$$\mathbb{P}_\ell \left(\sup_{\delta \in [\delta_{\min}, \delta_{\max}]} |\iota'_n(\delta) - \mathbb{E}_\ell[\iota'_n(\delta)]| \geq a_n/n \right) = o(1) .$$

Now define $h_\ell : \delta \mapsto \mathbb{R}$ as $h_\ell(\delta) := \mathbb{E}_\ell[\iota'_n(\delta)]$.

We have in particular that

$$\iota'_n(\hat{\delta}_n) - h_\ell(\hat{\delta}_n) = o_{\mathbb{P}}(a_n/n) .$$

Using a Taylor expansion of h_ℓ allows us to characterize the difference between $\hat{\delta}_n$ and $\tilde{\delta}_n$. Let $h'_\ell(\delta) = \frac{\partial}{\partial \delta} h_\ell(\delta) = \mathbb{E}_\ell[\iota''_n(\delta)]$ and recall that $\iota_n(\hat{\delta}_n) = h_\ell(\tilde{\delta}_n) = 0$ by definition. Then

$$\begin{aligned}
 0 &= \iota_n(\hat{\delta}_n) - h_\ell(\tilde{\delta}_n) \\
 &= \iota_n(\hat{\delta}_n) - h_\ell(\hat{\delta}_n) - h'_\ell(\tilde{\delta}_n)(\tilde{\delta}_n - \hat{\delta}_n) \\
 (46) \quad &= o_\mathbb{P}(a_n/n) - h'_\ell(\tilde{\delta}_n)(\tilde{\delta}_n - \hat{\delta}_n) ,
 \end{aligned}$$

where $|\tilde{\delta}_n - \hat{\delta}_n| \leq |\tilde{\delta}_n - \hat{\delta}_n|$. To proceed we must understand the behavior of $h'_\ell(\delta) = \mathbb{E}_\ell[\iota''_n(\delta)]$. As argued in the proof of Proposition 8.1, thanks to the results in [23] and the fact that $\iota''_n(\delta)$ and $\iota''(\delta)$ are uniformly bounded,

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

Actually, this result also holds under the alternative hypothesis by using the following fact:

LEMMA A.3. *Let $\frac{1}{2} < \gamma < 1$. Then $\mathbb{E}_1[\iota''_n(\delta)] - \mathbb{E}_0[\iota''_n(\delta)] \rightarrow 0$ as $n \rightarrow \infty$ uniformly in $\delta \in [\delta_{\min}, \delta_{\max}]$.*

The proof of this result follows almost immediately from the arguments used to prove Proposition 8.2. In addition, it is also shown in [23] that $\iota''(\delta_0) < 0$. Since $\tilde{\delta}_n$ converges in probability to δ_0 we therefore conclude that

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

This, together with (46) implies that

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

proving the second statement in the proposition. The third statement is a rather trivial consequence of the second statement by using a Taylor expansion of $p_m(\delta)$ around δ_0 (recalling Equation (23)) to obtain

$$n(p_m(\hat{\delta}_n) - p_m(\tilde{\delta}_n)) = np'_m(\delta_0)(1 + o_\mathbb{P}(1))(\hat{\delta}_n - \tilde{\delta}_n) = o_\mathbb{P}_\ell(a_n) .$$

□

A.4. Sketch proof of Lemma A.3. Similarly to the proof of Proposition 8.2 note that

$$\mathbb{E}_1[\iota''_n(\delta)] - \mathbb{E}_0[\iota''_n(\delta)] = -\frac{1}{n+1} \sum_{k \geq m} \frac{\mathbb{E}_1[N_{>k}(n)] - \mathbb{E}_0[N_{>k}(n)]}{(k+\delta)^2} .$$

This is quite similar to

$$\mathbb{E}_1[\iota'_n(\delta)] - \mathbb{E}_0[\iota'_n(\delta)] = \frac{1}{n+1} \sum_{k \geq m} \frac{\mathbb{E}_1[N_{>k}(n)] - \mathbb{E}_0[N_{>k}(n)]}{k+\delta} .$$

Note that $\sum_{k \geq m} \frac{1}{(k+\delta)^2}$ is a convergent series, unlike $\sum_{k \geq m} \frac{1}{k+\delta}$. By the same (and in fact somewhat simpler) arguments as in the proof of Proposition 8.2 we conclude therefore that $\mathbb{E}_1[\iota''_n(\delta)] - \mathbb{E}_0[\iota''_n(\delta)] = o(n^{\gamma-1})$, as we wanted to show. □

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

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* **287.5461** 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. *Physical Review Letters* **85.24** 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 Processes and Related Fields* **28.4** 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. *Annals of Applied Probability* **33.4** 2919–2980. <https://doi.org/10.1214/22-AAP1881>
- [5] BARABÁSI, A. L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286.5439** 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. *Physica A: Statistical Mechanics and its Applications* **311.3-4** 590–614. [https://doi.org/10.1016/S0378-4371\(02\)00736-7](https://doi.org/10.1016/S0378-4371(02)00736-7)
- [7] BHAMIDI, S., JIN, J. and NOBEL, A. (2018). Change point detection in network models: Preferential attachment and long range dependence. *Annals of Applied Probability* **28.1** 35–78. <https://doi.org/10.1214/17-AAP1297>
- [8] BHATTACHARJEE, M., BANERJEE, M. and MICHAELIDIS, G. (2020). Change point estimation in a dynamic stochastic block model. *The Journal of Machine Learning Research* **21.1** 4330–4388.
- [9] BOLLOBÁS, B. and RIORDAN, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24.1** 5–34. <https://doi.org/10.1007/s00493-004-0002-2>
- [10] BOLLOBÁS, B., RIORDAN, O., SPENCER, J. and TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms* **18.3** 279–290. <https://doi.org/10.1002/rsa.1009>
- [11] BRODER, A., KUMAR, R., MAGHOUL, F., RAGHAVAN, P., RAJAGOPALAN, S., STATA, R., TOMKINS, A. and WIENER, J. (2000). Graph structure in the web. *Computer Networks* **33.1** 309–320. [https://doi.org/10.1016/S1389-1286\(00\)00083-9](https://doi.org/10.1016/S1389-1286(00)00083-9)
- [12] BUBECK, S., DEVROYE, L. and LUGOSI, G. (2017). Finding Adam in random growing trees. *Random Structures & Algorithms* **50.2** 158–172. <https://doi.org/10.1002/rsa.20649>
- [13] BUBECK, S., ELKAN, R., MOSSEL, E. and RÁČZ, M. Z. (2017). From trees to seeds: On the inference of the seed from large trees in the uniform attachment model. *Bernoulli* **23.4A** 2887–2916. <https://doi.org/10.3150/16-BEJ831>
- [14] BUBECK, S., MOSSEL, E. and RÁČZ, M. Z. (2015). On the influence of the seed graph in the preferential attachment model. *IEEE Transactions on Network Science and Engineering* **2.1** 30–39. <https://doi.org/10.1109/TNSE.2015.2397592>
- [15] CIRKOVIC, D., WANG, T. and ZHANG, X. (2022). Likelihood-based Changepoint Detection in Preferential Attachment Networks. Available at arXiv: [2206.01076](https://arxiv.org/abs/2206.01076)
- [16] CRIMALDI, I. and PRATELLI, L. (2005). Convergence results for multivariate martingales. *Stochastic Processes and their Applications* **115.4** 571–577. <https://doi.org/10.1016/j.spa.2004.10.004>
- [17] CURIEN, N., DUQUESNE, T., KORTCHEMSKI, I. and MANOLESCU, I. (2015). Scaling limits and influence of the seed graph in preferential attachment trees. *Journal de l'École polytechnique — Mathématiques* **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. *Arkiv for Matematik* **47.1** 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. *Journ. Stat. Phys.* **139** 72–107. <https://doi.org/10.1007/s10955-010-9921-z>
- [20] FALOUTSOS, M., FALOUTSOS, P. and FALOUTSOS, C. (1999). On power-law relationships of the internet topology. *Computer Communication Review* **29.4** 251–261. <https://doi.org/10.1145/316194.316229>
- [21] 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*. *Physica A: Statistical Mechanics and its Applications* **318.3-4** 601–612. [https://doi.org/10.1016/S0378-4371\(02\)01731-4](https://doi.org/10.1016/S0378-4371(02)01731-4)
- [22] GAO, F. (2011). Modeling and interference of the internet movie database, Master Thesis, Eindhoven University of Technology.
- [23] GAO, F. and VAN DER VAART, A. (2017). On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees. *Stochastic Processes and their Applications* **127.11** 3754–3775. <https://doi.org/10.1016/J.SPA.2017.03.008>

- [24] GAO, F., VAN DER VAART, A., CASTRO, R. M. and VAN DER HOFSTAD, R. (2017). Consistent estimation in general sublinear preferential attachment trees. *Electronic Journal of Statistics* **11.2** 3979–3999. <https://doi.org/10.1214/17-EJS1356>
- [25] VAN DER HOFSTAD, R. (2017). *Random graphs and complex networks - Volume one*. Cambridge University Press. <https://doi.org/10.1017/9781316779422>
- [26] VAN DER HOFSTAD, R. (2024). *Random graphs and complex networks - Volume two*. (in preparation).
- [27] JEONG, H., TOMBOR, B., ALBERT, R., OLTVAL, Z. N. and BARABÁSI, A. L. (2000). The large-scale organization of metabolic networks. *Nature* **407.6804** 651–654. <https://doi.org/10.1038/35036627>
- [28] MARCHAND, D. C. and MANOLESCU, I. (2020). Influence of the seed in affine preferential attachment trees. *Bernoulli* **26.3** 1665–1705. <https://doi.org/10.3150/19-BEJ1152>
- [29] MIDDENDORF, M., ZIV, E. and WIGGINS, C. H. (2005). Inferring network mechanisms: The Drosophila melanogaster protein interaction network. *Proceedings of the National Academy of Sciences of the United States of America* **102.9** 3192–3197. <https://doi.org/10.1073/pnas.0409515102>
- [30] NEWMAN, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences* **98.2** 404–409. <https://doi.org/10.1073/pnas.021544898>
- [31] PENSKY, M. and ZHANG, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electronic Journal of Statistics* **13.1** 678–709. <https://doi.org/10.1214/19-EJS1533>
- [32] PERC, M. (2014). The Matthew effect in empirical data. *J. R. Soc. Interface* **11.98**. <https://doi.org/10.1098/rsif.2014.0378>
- [33] REDNER, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B* **4.2** 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. *Advances in Applied Probability* **48.A** 283–299. <https://doi.org/10.1017/apr.2016.56>
- [35] WANG, D., YU, Y. and RINALDO, A. (2021). Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics* **49.1** 203 – 232. <https://doi.org/10.1214/20-AOS1953>
- [36] WANG, H., TANG, M., PARK, Y. and PRIEBE, C. E. (2014). Locality statistics for anomaly detection in time series of graphs. *IEEE Transactions on Signal Processing* **62.3** 703–717. <https://doi.org/10.1109/TSP.2013.2294594>
- [37] WANG, M., YU, G. and YU, D. (2008). Measuring the preferential attachment mechanism in citation networks. *Physica A: Statistical Mechanics and its Applications* **387.18** 4692–4698. <https://doi.org/10.1016/j.physa.2008.03.017>
- [38] WATTS, D. J. (1999). *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press.
- [39] WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature* **393** 440–442. <https://doi.org/10.1038/30918>
- [40] ZHAO, Z., CHEN, L. and LIN, L. (2019). Change-point detection in dynamic networks via graphon estimation. Available at arXiv: [1908.01823](https://arxiv.org/abs/1908.01823)