Barabási-Albert random graphs, scale-free distributions and bounds for approximation through Stein's method

Elizabeth Ford

University of Oxford

A thesis submitted for the degree
of *Doctor of Philosophy.*

Trinity Term 2009

Barabási-Albert random graphs, scale-free distributions and bounds for approximation through Stein's method

Elizabeth Ford, Jesus College
A thesis submitted for the degree of *Doctor of Philosophy.*
Trinity Term 2009

Abstract

Barabási-Albert random graph models are a class of evolving random graphs that are frequently used to model social networks with scale-free degree distributions. It has been shown that Barabási-Albert random graph models have asymptotic scale-free degree distributions as the size of the graph tends to infinity. Real world networks, however, have finite size so it is important to know how close the degree distribution of a Barabási-Albert random graph of a given size is to its asymptotic distribution.

Stein's method is chosen as one main method for obtaining explicit bounds for the distance between distributions. We derive a new version of Stein's method for a class of scale-free distributions and apply the method to a Barabási-Albert random graph.

We compare the evolution of a sequence of Barabási-Albert random graphs with a continuous-time stochastic processes motivated by Yule's model for evolution. Through a coupling of the models we bound the total variation distance between their degree distributions. Using these bounds, we extend degree distribution bounds that we find for specific models within the scheme to find bounds for every member of the scheme.

We apply the Azuma-Hoeffding inequality and Chernoff bounds to find bounds between the degree sequences of the random graph models and the given scale-free distribution. These bounds prove that the degree sequences converge completely (and therefore also converge almost surely) to our scale-free distribution.

We discuss the relationship between the random graph processes and the Chinese restaurant process. Aided by the construction of an inhomogeneous Markov chain, we apply our results for the degree distribution in a Barabási-Albert random graph to a particular statistic of the Chinese restaurant process.

Finally, we explore how our methods can be adapted and extended to other evolving random graph processes. We study a Bernoulli evolving random graph process, for which we bound the distance between its degree distribution and a geometric distribution and we bound the distance between the number of triangles in the graph and a normal distribution.

Acknowledgements

I am pleased to acknowledge the many sources of assistance and support that I have received during the course of this work.

To my supervisor Prof. Gesine Reinert I express my sincere thanks for her guidance, patience and enthusiastic encouragement.

I thank the Institute for Mathematical Sciences of the National University of Singapore for enabling my participation in the *Progress in Stein's Method* programme. I thank all those with whom I have had useful discussions both on this programme and elsewhere.

My thanks go to Sophie Schbath and to the Statistics for Systems Biology research group in Paris for inviting me to talk about my work as well as financing my visit.

Finally I thank the Engineering and Physical Sciences Research Council for its financial support.

Contents

Notation

The following provides a summary of common notation and conventions used throughout this thesis.

- The natural numbers are defined not to include zero.

  $$\mathbb{N} := \{1, 2, 3, ...\}$$

  $$\mathbb{Z}_0^+ := \{0, 1, 2, ...\}$$

- The forward difference is denoted by Δ. For a function f on $\mathbb{Z}$,

  $$\Delta f(k) := f(k+1) - f(k).$$

- We use the usual order notation:

  $$\begin{aligned} f(k) = o(g(k)) \text{ as } k \to \infty &\iff \frac{f(k)}{g(k)} \to 0 \text{ as } k \to \infty; \\ f(k) = O(g(k)) \text{ as } k \to \infty &\iff \exists C, k_0 \geq 0, \quad |f(k)| \leq C|g(k)| \quad \forall k \geq k_0. \end{aligned}$$

- We write $f(k) << g(k)$ to mean $f(k) = o(g(k))$.
- The supremum norm is denoted by $\|\cdot\|_\infty$. For a function f,

  $$\|f\|_\infty := \sup_x |f(x)|.$$

- For random variables X and Y, we write $X \perp Y$ to mean that X and Y are independent.
- We use the usual notation for the gamma and beta functions (defined on $\mathbb{C}^+$ and $\mathbb{C}^+ \times \mathbb{C}^+$ respectively),

  $$\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} dt;$$

  $$B(x, y) := \int_0^1 t^{x-1}(1-t)^{y-1} dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

- A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is described by its vertex set $\mathcal{V}$ and edge set $\mathcal{E}$.

  If $\mathcal{G}$ is a directed graph then an edge from vertex $i \in \mathcal{V}$ to vertex $j \in \mathcal{V}$ is denoted by the ordered pair $(i, j) \in \mathcal{E}$. If $\mathcal{G}$ is an undirected graph then an edge between vertices $i, j \in \mathcal{V}$ is denoted by the set $\{i, j\} \in \mathcal{E}$.

- Following the definition by Hsu and Robbins [29] we say that a sequence of random variables $(X_n)_n$ converges completely to a constant c if the series

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - c| > \varepsilon)$$

converges for every $\varepsilon > 0$. We write $X_n \to^C c$.

Distributions

- $Geometric_{\mathbb{N}}(p)$ denotes the geometric distribution with parameter $p \in (0,1]$ and state space $\mathbb{N}$; $\mu(k) = p(1-p)^{k-1}$ for $k \in \mathbb{N}$.
- $Geometric_{\mathbb{Z}_0^+}(p)$ denotes the geometric distribution with parameter $p \in (0,1]$ and state space $\mathbb{Z}_0^+$; $\mu(k) = p(1-p)^k$ for $k \in \mathbb{Z}_0^+$.
- $Yule-Simon(\rho)$ denotes the Yule-Simon distribution with parameter $\rho > 0$. For a random variable V,

$$V \sim Yule-Simon(\rho) \iff \big\{\mathbb{P}(V = k) = \rho B(k, \rho + 1) \quad \forall k \in \mathbb{N}\big\}.$$

- $Negbinom(r,p)$ denotes the negative binomial distribution with parameters $r > 0$ and $p \in (0,1)$. For a random variable V,

$$V \sim Negbinom(r,p) \iff \left\{\mathbb{P}(V = k) = \frac{\Gamma(k+r)}{k!\Gamma(r)} p^r (1-p)^k \quad \forall k \in \mathbb{Z}_0^+\right\}.$$

Equivalently,

$$V \sim Negbinom(r,p) \iff \left\{\mathbb{P}(V \le k) = \frac{\int_0^p s^{r-1}(1-s)^k ds}{B(r; k+1)} \quad \forall k \in \mathbb{Z}_0^+\right\}.$$

For $r \in \mathbb{N}$, if $U_1, U_2, ..., U_r$ are independent identically distributed random variables with $\mathcal{L}(U_1) = Geometric_{\mathbb{Z}_0^+}(p)$, equivalently we can say,

$$V \sim Negbinom(r,p) \iff \mathcal{L}(V) = \mathcal{L}\left(\sum_{i=1}^{r} U_i\right).$$

- For a random variable R that takes values in $[0,1]$, we use $V \sim Negbinom(\alpha; R)$ to mean that the random variable V is defined conditionally on R, such that given $R = r$, $V \sim Negbinom(\alpha; r)$. The notation is extended to other such mixtures of distributions.
- For $\alpha > 0, \rho > 1$ we define μ_ρ^α to be the measure on $\mathbb{Z}_0^+$ given by

$$\mu_\rho^\alpha(k) := \frac{B(k+\alpha; \rho+1)}{B(\alpha; \rho)}.$$

Distances

- The total variation distance between two random variables, X and Y, defined jointly on a state space $\mathcal{S}$ is defined by,

$$\begin{aligned} d_{TV}(X,Y) &:= \sup_{A\subset\mathcal{S}} |\mathbb{P}(X\in A)-\mathbb{P}(Y\in A)| \\ &= \sup_{h\in\mathcal{H}} |\mathbb{E}(h(X))-\mathbb{E}(h(Y))|, \end{aligned}$$

  where $\mathcal{H} := \{\mathbf{1}_A : A\subset\mathcal{S}, \quad A \text{ measurable}\}$.

  If $\mathcal{S}$ is countable then, equivalently,

$$d_{TV}(X,Y) = \frac{1}{2}\sum_{k\in\mathcal{S}} |\mathbb{P}(X=k)-\mathbb{P}(Y=k)|.$$

  If $\mathcal{S}$ is Hausdorff separable then, equivalently (see [8] p. 254),

$$d_{TV}(X,Y) = \min_{X'=^D X, Y'=^D Y} \mathbb{P}(X'\neq Y').$$

  There are many other equivalent definitions, but these will be adequate for our purposes.

- The Kolmogorov distance between two random variables, X and Y, taking values in a subset of the real line is defined by,

$$\begin{aligned} d_K(X,Y) &:= \sup_{x\in\mathbb{R}} |\mathbb{P}(X\leq x)-\mathbb{P}(Y\leq x)| \\ &= \sup_{h\in\mathcal{H}} |\mathbb{E}(h(X))-\mathbb{E}(h(Y))|, \end{aligned}$$

  where $\mathcal{H} := \{\mathbf{1}_{(-\infty,x]} : x\in\mathbb{R}, \quad h \text{ measurable}\}$.

- The Wasserstein distance between random variables, X and Y, defined jointly on a state space $\mathcal{S}$ is defined by,

$$d_W(X,Y) := \sup_{h\in\mathcal{H}} |\mathbb{E}(h(X))-\mathbb{E}(h(Y))|,$$

  where $\mathcal{H} := \{h:\mathbb{R}\to\mathbb{R} : h \text{ differentiable } ||h'||_\infty \leq 1\}$.

Chapter 1

Introduction

This thesis is a rigorous study of convergence to a scale-free distribution, with bounds on the distance to the limit. One achievement of note is the derivation of Stein's method for a generalised scale-free distribution, of which the Yule-Simon distribution is a special case.

The motivation for this work is the study of scale-free random graphs and, in particular, Barabási-Albert random graph models. Our method connects the evolution of Barabási-Albert random graphs with the continuous-time stochastic process known as Yule's model for the evolution of species. We consider a sequence of discrete-time and continuous-time preferential attachment processes and use couplings to bound the total variation distance between the degree distributions of the models in our scheme.

From Yule's model for evolution we identify a candidate asymptotic degree distribution for our models. We study a range of different methods for bounding the total variation distance between the degree distributions and the asymptotic distributions of our models. Each method applies directly only to certain models in our scheme, but a coupling of the models allows us to extend our results to every model. We see that the method providing the better bounds depends on the choice of model and the chosen parameters of that model.

In Chapter 10 we indicate how our results for random graph models also apply to a statistic of the Chinese restaurant process. To aid this comparison we construct an inhomogeneous Markov chain, whose evolution represents the evolution of the degree distribution in one of the models of our scheme. Later we discuss a possible method of applying Stein's method directly to this Markov chain to obtain bounds for the distance between the chain at step

n and its asymptotic distribution.

Chapter 12 extends our methods to obtain similar results for a Bernoulli evolving random graph. Guided by the successful approach for Barabási-Albert models, we bound the distance between the degree distribution and a geometric distribution, and the distance between the number of triangles in the graph and a normal distribution.

1.1 Social network models

Barabási-Albert random graph models have been motivated by social networks. Social networks are networks that are derived from the study of relationships in social systems. An easy example of a social network is a friendship network. We can construct a graphical representation of a friendship network where vertices represent individuals and edges indicate friendships between individuals. Vertex i and vertex j are joined by an edge if and only if individuals i and j are considered to be friends.

Social networks may be directed or undirected. For example, in a friendship network we might consider friendships to be mutual, meaning that the resulting network is undirected. Alternatively, we might form a network by asking every individual in the population who they consider their friends to be. The results could be represented as a directed network with an edge from vertex i to vertex j if and only if individual i lists individual j as a friend.

The Barabási-Albert random graph model that we study is one that has been used as a model for social networks. This model was proposed in [5] with the intention of emulating the growth of networks such as citation networks and the World Wide Web. In particular, Barabási and Albert aimed to produce networks that exhibited a scale-free degree distribution. The idea was that in a social network the network grows over time, under the preferential attachment condition that high degree (or 'popular') vertices are more likely to gain new edges than low degree vertices. The Barabási-Albert random graph model evolves in discrete time steps. At each step a new vertex is added to the graph and m edges are drawn between that vertex and m old vertices. The probability that an old vertex is chosen is proportional to the degree of that vertex.

1.2 Scale-free distributions

We use the term '*scale-free distribution*' or '*power law distribution*' to describe a probability distribution with a probability mass function or probability distribution function asymptotically of the form $f(x) = Ax^{-\gamma}$ as $x \to \infty$, for some constants A and γ.

The description, '*scale-free distribution*', comes from a property that functions of this type satisfy known as '*scale invariance*'. Caldarelli [14] discusses the concept of scale invariance in detail. In the case of these functions we mean that whatever scale we use, the shape of the function is the same. An alternative definition is that a function f satisfies an exact power law property if, for all x in the domain of f and for any scaling constant, c,

$$f(cx) \propto f(x).$$

It is clear that this is the case for any function of the form $f(x) = Ax^{-\gamma}$, with the exponent determining the decay of the power law tail. In physics it has been observed that power laws with the same exponent are seen in systems that share dynamical properties. This is known as '*universality*' or '*self-organized criticality*' and systems with the same exponent are described as belonging to the same '*universality class*'. Solow [48] discusses self-organized criticality in relation to power laws, concluding that, although there is a connection between power laws and self-organized criticality, the first does not imply existence of the second.

1.2.1 Degree distributions

When discussing the '*degree distribution*' of a random graph authors use this description to refer to various statistics of the graph. This section clarifies our usage of the term and alternative meanings found in the literature.

For a vertex i in an undirected random graph, the '*degree*' of i refers to the number of edges incident to vertex i. In a directed graph we can talk about the number of edges directed into the vertex (the '*in-degree*'), the number of edges directed out of the vertex (the '*out-degree*'), or the total number of edges directed into or out of the vertex (the '*total degree*').

Let $\mathcal{G}$ be a random graph on n vertices labelled $1, 2, ..., n$ and let I_n be a random variable distributed uniformly on $\{1, 2, ..., n\}$. In this thesis we use the description '*the degree distribution of* $\mathcal{G}$' to refer to the distribution of the random variable $D(I_n)$ defined to be the degree of vertex I_n in $\mathcal{G}$. This is the meaning of degree distribution used by Drmota, Giménez and Noy [20]. Initially the definition might appear strange, but knowing the distribution of $D(I_n)$ is equivalent to knowing the values of the expected number of vertices of every possible degree in the graph $\mathcal{G}$. If $N(k)$ represents the number of vertices of degree k in $\mathcal{G}$, then

$$\mathbb{E}(N(k)) = \sum_{i=1}^{n} \mathbb{P}(D(i) = k) = n\mathbb{P}(D(I_n) = k).$$

By our definition of degree distribution, therefore, the knowing the degree distribution corresponds to knowing the sequence $(\mathbb{E}(N(k)))_k$.

Confusion could arise from the alternative usage of '*degree distribution*' in some literature to describe the distribution of the sequence $(N(k))_k$ where $N(k)$ again represents the number of vertices of degree k in the random graph $\mathcal{G}$. In this thesis we follow the example of Bollobás et al. [11] in referring to this sequence as the degree sequence of the graph rather than as the degree distribution of the graph.

1.2.2 The scale-free order (γ)

In this thesis we bound the total variation distance between Barabási-Albert random graph models on n vertices and specific scale-free probability distributions. The models that we study are random graphs with scale-free degree distributions that have power law tails of order $\gamma > 2$.

Durrett [22] lists previously studied social networks and the orders of the associated power laws. Durrett references ten different observed power laws, of which nine are of the form $\gamma > 2$ and the majority are of the form $\gamma \in (2, 3]$. In Chapter 3 we gain some insight into the parameters of the model that determine the order of the resulting power law, and in Chapter 7 the relationship for our specific class of models is proved.

For now, it is worth noting that power laws of order $\gamma \leq 2$ are precisely those with an undefined expectation. In the case of our random graphs, having such a distribution as asymptotic degree distribution translates to the situation where the number of edges added per vertex diverges as the number of vertices tends to infinity. Barabási-Albert random graphs aim to add precisely m edges with every vertex, so scale-free distributions of order $\gamma \leq 2$ would not be expected to result.

1.2.3 Log-log plots

A common method of estimating the power-law exponent in a power law distribution is to take the gradient of a log-log plot. This method is illustrated in [49], where a discussion is given of methods for testing the power law model. We now explain how the total variation distance bounds for degree distribution approximations have implications for the use of log-log plots when estimating the power law exponent in a Barabási-Albert random graph process. In Chapter 3 we shall define a scale-free distribution, μ_ρ^α, which has a power law tail. In later chapters of this thesis, we shall find total variation distance bounds between the degree distributions of Barabási-Albert random graphs and the μ_ρ^α-distribution.

Let $V \sim \mu_\rho^\alpha$ and suppose that W_n represents the degree distribution of a particular Barabási-Albert random graph on n vertices. We consider the range of vertex degrees for which the gradient of a log-log plot gives a good approximation of the power law exponent, $\rho + 1$. This assumes that we are able to sample data from a sufficiently large number of individual networks. In Section 3.2.3 we shall prove that, for $k \geq 0$, $\mu_\rho^\alpha(k) \propto k^{-(\rho+1)} + O(k^{-(\rho+2)})$. Thus for some constant, c,

$$\forall k \quad |\mathbb{P}(W_n = k) - ck^{-(\rho+1)}| \leq d_{TV}(W_n, V) + O(k^{-(\rho+2)}).$$

It follows that the log-log plot will provide good approximations for ρ for k such that $k^{-(\rho+2)} + d_{TV}(W_n, V) << k^{-(\rho+1)}$.

Now

$$k^{-(\rho+2)} << k^{-(\rho+1)} \iff 1 << k \iff 0 << \log k,$$

and,

$$d_{TV}(W_n, V) << k^{-(\rho+1)} \iff \log k << \log d_{TV}(W_n, V).$$

Thus a log-log plot gives a good approximation for ρ in the region,

$$0 << \log k << \log d_{TV}(W_n, V).$$

The total variation distance bounds that will be found in Chapter 7 and Chapter 8 are all of order $\frac{\log n}{n}$ or $n^{-\gamma}$ for some $\gamma \in (0,1)$, thus for these graphs the region is equivalently

$$0 << \log k << \log n.$$

1.3 Outline of thesis

The thesis divides naturally into three parts. Chapters 2 and 3 introduce preferential attachment models from two different areas of mathematics. Chapters 4 and 5 introduce the approximation method known as Stein's method and develop ways of applying Stein's method to the particular power law distribution that we have found in Chapter 3. The remaining chapters present bounds between sequences of random variables and the asymptotic distributions of those sequences. Firstly we look at the degree distribution of the evolving Barabási-Albert random graph model, and then at other mathematical problems that use similar methods.

We begin Chapter 2 by describing the properties of Barabási-Albert random graph models and how the precise forms that we choose to study fit into the context of previous work. Chapter 3 introduces Yule's model for evolution. We compare the continuous-time stochastic model given by Yule for the evolution of species to the network models of Barabási and Albert. We show how a specific power law degree distribution, which we call the μ_ρ^α-distribution, arises naturally in these continuous-time models. Later, in Chapter 6, Yule's model for evolution motivates our derivation of a class of continuous-time Barabási-Albert random graph models. We bound the distance between the degree distributions of these models and a scale-free distribution in Chapter 7.

Having looked at both discrete-time and continuous-time preferential attachment models with asymptotic power law distributions, Chapter 4 introduces a major tool, known as Stein's method, that we develop for the μ_ρ^α-distribution and use to analyse the models. Stein's method is a natural choice for our purposes because it provides explicit bounds for distances between distributions under a given metric. In Chapter 4 we present known results for Stein's method for the geometric distribution, the normal distribution and general discrete distributions. These results will be used and extended later in the thesis.

In Chapter 5 we derive a new method of applying Stein's method to the μ_ρ^α-distribution. We derive a Stein equation and Stein factors for Kolmogorov and total variation distances. This method is applied to a Barabási-Albert random graph model in Chapter 11.

In Chapter 6 we use coupling constructions to bound for the total variation distance between the degree distributions of each of a sequence of Barabási-Albert random graph models. These bounds provide a means of extending degree distribution bounds for any model in our scheme to find bounds for the other models. Since each bounding method explored later applies only to some of the models, these coupling constructions add wider significance to the later results.

Chapter 7 uses a theorem bounding the total variation distance between two negative binomial distributions to bound the distance between the degree distribution of one of the models of Chapter 6 and the μ_ρ^α-distribution. We illustrate how these bounds can be extended to other models within the scheme using the results of Chapter 6.

Chapter 8 explores how the method of recurrence equations gives us alternative bounds for the distance between the expected number of vertices of a given degree and a power law measure. From this we obtain different bounds for the total variation distance found in Chapter 7. Compared with the bounds obtained by other methods, the bounds from method are stronger for some models and parameters, and of the same order for others.

In the case of our original Barabási-Albert random graph, we are able to apply the Azuma-Hoeffding inequality to obtain bounds and prove complete convergence, which is stronger than almost sure convergence, for the degree sequence, as well as other degree statistics of the random graph. We do this in Chapter 9. For other models, which lack a condition necessary to apply the Azuma-Hoeffding inequality, we find similar bounds using Chernoff bounds, again proving complete convergence.

Chapter 10 compares a stochastic process known as the Chinese restaurant process to the random graphs that we have studied and proves that our results for the degree distribution apply to a particular statistic of this process. We see that the degree distribution of one model in our scheme can be studied as the outcome of a Markov chain.

Chapter 11 discusses a possible avenue for bounding the distance between the outcome of the Markov chain from Chapter 10 and its asymptotic distribution directly, using the new version of Stein's method derived in Chapter 5. Here we prove that, should the Markov chain converge, then its asymptotic distribution would be the μ_ρ^α-distribution.

In Chapter 12 we look at an evolving Bernoulli random graph. We show that many of our methods translate easily to the Bernoulli case, indicating their versatility. We bound the distance between the degree sequence for the Bernoulli random graph and a geometric distribution, and we prove complete convergence following the same method as in the Barabási-Albert case. For this model we use Stein's method to bound the number of triangles in the graph from a normal distribution.

We conclude in Chapter 13, with a discussion of the achievements of this thesis and suggestions of further avenues for future research.

Finally, the Appendices provide material for reference. Appendix A gives properties of beta and gamma functions that are referred to in calculations throughout the thesis, while Appendix B provides a list of definitions for models in our scheme.

Chapter 2

Barabási-Albert models

In order to model networks that exhibit a scale-free distribution, Barabási and Albert [5] proposed a class of evolving random graph models that follow a so-called '*preferential attachment*' condition. By preferential attachment they mean that the graphs evolve over time in such a way that high degree vertices are more likely to gain additional edges than low degree vertices. In fact, their models specify that the probability of attaching a new vertex to a particular previous vertex at a specific time point should be proportional to the degree of that vertex at that time.

The need for network models with a scale-free degree distribution is illustrated by Barabási and Albert in a number of specific real network examples. One example is a citation network of academic publications. In such a network vertices represent academic publications and a directed edge is drawn from vertex i to vertex j if and only if publication i cites publication j. The preferential attachment condition for a model of this network is justified by the idea that a new publication is more likely to cite well-known and frequently cited previous publications than less frequently cited previous publications.

Caldarelli [14] provides an overview of current thinking on the subject of scale-free networks. In particular, Caldarelli notes that the idea of preferential attachment has been independently rediscovered several times and thus has been studied under several different names including *Yule's model for evolution*, *the Matthew effect*, *rich gets richer*, *preferential attachment* and *cumulative advantage*. For us the most important of these formulations will be the preferential attachment condition of Barabási-Albert random graphs and the Yule's model for evolution. Indeed, it is the relationship between the preferential attachment evolution condition of Barabási and Albert and the formulation of Yule's model for evolution that will prove key to our method for obtaining error bounds. Bhamidi [9] makes use of this relationship and uses embedding techniques to obtain asymptotic distribution results (but not bounds) for characteristics of preferential attachment trees, including the degree distribution. Here we do not restrict our attention to trees.

Yule's model for evolution was first described in 1925 [56] as a model for the evolution of species. The model will be discussed in detail in Chapter 3. For now we concentrate on defining a precise version of the Barabási-Albert random graph.

Barabási and Albert defined their models as follows,

> "...starting with a small number (m_0) vertices, at every time step we add a new vertex with m ($\leq m_0$) edges that link the new vertex to m different vertices already present in the system. To incorporate preferential attachment, we assume that the probability Π that a new vertex will be connected to a vertex i depends on the connectivity k_i of that vertex, so that $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$." ([5] p. 511)

As Durrett ([22] p. 90) correctly points out, when $m > 1$ this description is imprecise. Do we choose our m vertices one by one or at the same time? If we choose vertices consecutively, do we update the vertex degrees in between choices? How do we ensure that we do not choose a vertex more than once in a particular step? Thus the above description indeed fits a whole class of models, a subclass of which we study.

There are more difficulties even once a precise definition has been chosen. Barabási-Albert random graph models prove difficult to analyse due to the high levels both of inhomogeneity and of dependence between edges of the graphs.

Various precise versions of this description have been studied, and we discuss our choice of formulation in the context of previous work in Section 2.2.

This thesis looks at a sequence of types of Barabási-Albert random graph models. Model BA, described in the next section, is chosen in an attempt to follow the Barabási-Albert description as closely as possible. In this vein we insist that precisely m vertices will be chosen at each time-step (disallowing loops and multiple edges) and that the marginal probability of choosing a particular vertex at time n is proportional to the vertex degree. We do not specify the probabilities of choosing each set of m vertices at each time-step, preferring instead to prove results that hold for all probability distributions with the given marginal probabilities. Example 2.1 will show that random graphs following the specification of Model BA exist.

2.1 A subclass of Barabási-Albert models

We define a class of models that satisfy the preferential attachment property, namely that the probability of adding an edge to a particular vertex given its degree is proportional to the degree of that vertex. For each sequence of parameters, (α, m_0, m), a whole family of random graph models is described. In Chapter 7 we bound the distance between a power law* distribution and the degree of a randomly chosen vertex in an arbitrary graph belonging to this class of models.

Let $\alpha > 0$ be a real parameter and let $m \leq m_0$ be integer parameters. Some of our results will only apply for $\alpha \geq 1$ but in general $\alpha > 0$ is allowed.

Model BA

- **(BA1)** Start with the empty graph on m_0 vertices.
- **(BA2)** At step n select m distinct vertices from the graph in such a way that the vertex i with $D_{n-1}^{(BA)}(i) := (\alpha +$ in-degree of i at step $n-1)$ is chosen with probability

*The distribution considered is a discrete probability distribution on $\mathbb{Z}_0^+$ with measure $\mu(k) = \frac{B(k+\alpha; \frac{2m+\alpha}{m})}{B(\alpha; \frac{m+\alpha}{m})}$. The special case of $\alpha = 1$ is known as the Yule-Simon distribution. We discuss this scale-free distribution and its properties in Chapter 3.

proportional to $D_{n-1}^{(BA)}(i)$. More precisely,

$$\mathbb{P}(\text{vertex } i \text{ is chosen}|D_{n-1}^{(BA)}(i)) = \frac{mD_{n-1}^{(BA)}(i)}{(n-1)(m+\alpha)+m_0\alpha}.$$

(Note that the conditioning here is not on the whole degree vector at step $n-1$.)

- **(BA3)** Attach the new vertex, $n+m_0$, to the m chosen vertices with m edges directed out of the new vertex.

This definition is included in Appendix B for reference.

Of course we need to check that there are random graphs that satisfy these conditions so that the subclass is non-empty. Example 2.1 provides a construction for a random graph satisfying **(BA1)**, **(BA2)** and**(BA3)**.

Example 2.1. *Let $\mathcal{G}_0$ be the empty graph on $m_0 \geq 1$ vertices so that **(BA1)** is satisfied and so that **(BA1)**, **(BA2)** and **(BA3)** have been satisfied at every previous step.*

*Now suppose that for some $n \in \mathbb{Z}_0^+$, $\mathcal{G}_n$ is a random graph that has evolved in n steps, satisfying **(BA1)**, **(BA2)** and **(BA3)** at all previous steps. Then $\mathcal{G}_n$ is a graph on $n+m_0$ vertices. Moreover the first m_0 vertices have out-degree 0 while the remaining vertices have out-degree m.*

*The random graph $\mathcal{G}_n$ satisfied the condition **(BA2)** above at every previous step in its evolution. Thus for each previous step there is an associated random set of m vertices that were chosen at that step. If $n > 0$ then for $k = 1, ..., n$ define,*

$$S_k := \{\textit{vertices selected at step } k\}.$$

Note that each set S_k has cardinality m.

We consider the following rule for selecting m vertices at step $n+1$. Define the random variables $I_n \sim Uniform\{1, 2, ..., n\}$ independent of the random graph process, and let T_{n+1}

be a set of m vertices chosen uniformly at random from $\mathcal{V}_n$, the vertex set $\{1, 2, ..., n+m_0\}$. At step 1 all vertices have in-degree 0 so we let $S_1 := T_1$. For $n \geq 1$, then with probability $\frac{mn}{n(m+\alpha)+m_0\alpha}$ let $S_{n+1} := S_{I_n}$, i.e. we select the same vertices at step $n+1$ as at step I_n. Otherwise let $S_{n+1} := T_{n+1}$.

*Our definition of S_{n+1} ensures that the set contains m elements, so that the condition **(BA3)** is satisfied. It remains to prove that the marginal probability of choosing vertex i satisfies condition **(BA2)**. For $n = 0$ this holds since all vertices have the same initial degree and S_0 is uniform on the set of all sets of m vertices.*

*Now consider a vertex i in the graph $\mathcal{G}_n$ for $n \geq 1$. The in-degree of vertex i in $\mathcal{G}_n$ is given by the number of steps at which vertex i was chosen via condition **(BA2)**. Each of these occasions corresponds to precisely one value of k for which $i \in S_k$. Thus,*

$$\begin{aligned} D_n^{(BA)}(i) &:= (\alpha + \text{ in-degree of } i \text{ at step } n) \\ &= \alpha + \#\{k : 1 \leq k \leq n, i \in S_k\}. \end{aligned}$$

We see that $D_n^{(BA)}(i) = \alpha + \#\{k : 1 \leq k \leq n, i \in S_k\}$ is a function of the sequence of *random variables* $(S_1, ..., S_n)$ *and* $D_n^{(BA)}(i) \perp T_{n+1}$. *Thus,*

$$
\begin{aligned}
\mathbb{P}(i \in S_{n+1} | D_n^{(BA)}(i)) &= \frac{mn}{n(m+\alpha)+m_0\alpha}\mathbb{P}(i \in S_{I_n}|D_n^{(BA)}(i)) \\
&\quad + \frac{\alpha(m_0+n)}{n(m+\alpha)+m_0\alpha}\mathbb{P}(i \in T_{n+1}|\#\{k : 1 \leq k \leq n, i \in S_k\}) \\
&= \frac{mn}{n(m+\alpha)+m_0\alpha}\mathbb{P}(i \in S_{I_n}|D_n^{(BA)}(i)) + \frac{\alpha(m_0+n)}{n(m+\alpha)+m_0\alpha}\mathbb{P}(i \in T_{n+1}) \\
&= \left(\frac{mn}{n(m+\alpha)+m_0\alpha}\right)\mathbb{E}\left(\frac{\#\{k : 1 \leq k \leq n, i \in S_k\}}{n}\Big| D_n^{(BA)}(i)\right) \\
&\quad + \left(\frac{\alpha(m_0+n)}{n(m+\alpha)+m_0\alpha}\right)\left(\frac{\binom{n-1+m_0}{m-1}}{\binom{n+m_0}{m}}\right) \\
&= \frac{m(D_n^{(BA)}(i)-\alpha)}{n(m+\alpha)+m_0\alpha} + \left(\frac{\alpha(m_0+n)}{n(m+\alpha)+m_0\alpha}\right)\left(\frac{m}{n+m_0}\right) \\
&= \frac{mD_n^{(BA)}(i)}{n(m+\alpha)+m_0\alpha},
\end{aligned}
$$

as required.

Defining S_{I_n} in the proof above was conditional on the evolution of the graph at time steps $1, 2, ..., n$. However, since knowing $\mathcal{G}_n$ allows us to determine the entire evolution, $(\mathcal{G}_0, \mathcal{G}_1, ..., \mathcal{G}_n)$, the graph evolution is trivially Markov. Therefore we were able to construct these probability distributions based on observing the current graph structure.

The construction at step n required only that **(BA1)**, **(BA2)** and **(BA3)** were satisfied at the previous steps. Therefore we have also proved that, given any construction of an n-step evolving random graph satisfying these conditions, it is possible to construct further steps $n+1, n+2, ...$ to create a random graph that is a member of this class.

2.2 This thesis in context: a discussion of previous results from the literature

Since Barabási and Albert [5] suggested preferential attachment as a method for modelling scale-free networks in 1999, many different precise forms of the model have been studied. We attempt to provide an overview of relevant literature concerning this class of models, concentrating in particular on mathematical results relating to them and especially those that describe the degree distributions of such models.

Introductions to the subject of Barabási-Albert random graph models can be found in, for example, [14], [18], [22] and [35]. All give an overview of the subject, and Caldarelli [14] discusses other occurrences of the scale-free property including in Yule processes, an important connection for this thesis that will be discussed in detail in Chapter 3. Newman, Barabási and Watts [37] have collected papers on network theory with introductions to each section provided by the authors; several of these papers are discussed below.

There are a number of different precise versions of the Barabási-Albert random graph for which results have been proved. One model of note is known as the '*LCD model*'. This model was first described by Bollobás et al. [11] in 2001 and is a special version of the Barabási-Albert random graph model constructed in such a way that the random graph $\mathcal{G}_n$ on n vertices can be defined non-recursively. The model allows both loops and multiple edges and when $m > 1$ edges are added at each time step, the edges are added one by one and use the updated vertex degrees that include the previous edges from that time step.

Bollobás et al. [11] prove that, when m vertices are added at each step in this model, the asymptotic power law distribution is of the form

$$\mathbb{E}(\#\text{degree k vertices}) \sim \frac{2m(m+1)n}{(k+m)(k+m+1)(k+m+2)},$$

for $0 \leq k \leq n^{\frac{1}{15}}$. We observe that this model gives a power law of the form $k^{-\gamma}$ where

$\gamma = 3$ for any value of $m \in \mathbb{N}$. This contrasts with our version of the model, where, as we shall prove in Chapter 7, $\gamma = 2 + \frac{\alpha}{m}$ and so depends on the values of m and α and lies in the interval $(2, \infty)$.

Further, Bollobás et al. [11] introduce a method using the Azuma-Hoeffding inequality and martingale theory to bound

$$\mathbb{P}|(\mathbb{E}(\#\text{degree k vertices}) - (\#\text{degree k vertices})| > \sqrt{n \log n}).$$

This method applies to our Model BA and will be used in Chapter 7 to extend our results to bound the degree sequence of Model BA from a given power law distribution.

Szymański [53] uses a recurrence equations to study the degree sequence in the LCD model. In contrast to the other authors, Szymański finds bounds for finite sized graphs. We apply the same method in Chapter 8 for a model in our scheme. We show how total variation distance bounds between the degree distribution and a power law can be deduced from the degree sequence bounds.

Bollobás and Riordan [10] and Riordan [42] further study the LCD model, looking at the giant component of the model.

Cooper and Frieze [17] give a very general preferential attachment model. In their model at each time step there is a choice of procedure, with the choice determined by the outcome of a $Bernoulli(\alpha)$ random variable. In the first procedure, when a new vertex is added to the model a random number of new edges is generated. A further $Bernoulli(\beta)$ random variable determines whether the edges attach to old vertices either uniformly at random or by preferential attachment. If the second procedure is chosen then new edges are added within the model according to a prescribed random procedure, but the number of vertices remains the same.

The Cooper and Frieze model allows multiple edges. The choice of the two procedures also means that the number of vertices in the graph at step n is random. Cooper and

Frieze prove that the degree sequence of such a model asymptotically follows a power law, but they do not find the exact form of the power law.

Katona and Móri [32] study a preferential attachment model where edges are added independently to each old vertex in the model at step n. This means that a random number of edges is added at each step. This model can be compared to our Model I of Chapter 6, where we add an edge to each old vertex independently with probability proportional to $\alpha+$ in-degree. A difference between Model I and the Katona and Móri model is that in Model I the probability of choosing a vertex i at step n is conditional only on the current in-degree of i, whereas Katona and Móri condition both on the degree of i and the sum of the degrees in the graph.

It is proved that models of type described by Katona and Móri converge to a power law distribution of order k^{-3}, the exact form of which is given. Additionally, Katona and Móri look at the maximum degree in their model and prove a strong law of large numbers result.

So far most of the models that we have discussed have exhibited asymptotic power law degree distributions of order k^{-3}. Krapivsky, Redner and Leyvraz [34] and Krapivsky and Redner [33] look at preferential attachment conditions where instead of choosing new vertices with probability proportional to their degree they use a probability proportional to a power of the vertex degree. Through this method they obtain asymptotic power laws of order $k^{-\gamma}$ with $\gamma \neq 3$.

Dorogovtsev, Mendes and Samukhin [19] suggest an alternative method for obtaining power laws of order $k^{-\gamma}$ with $\gamma \neq 3$. They use a linear preferential attachment condition, but instead of updating the vertex degrees between the addition of each new edge, m edges are added at each time step, independently and with probability distribution given by the vertex in-degrees after the previous time step. Multiple edges are allowed, but it is expected that they will occur infrequently. For a vertex, i, of in-degree $D_{n-1}(i)$ at step $n-1$, the probability that each edge is added directed into vertex i, conditional on

$D_{n-1}(i)$, is given by

$$\frac{A + D_{n-1}(i)}{(m+A)(n-1)}.$$

The value of A is a positive constant parameter for the model, and is described as the 'attractiveness' of the vertices.

Dorogovtsev, Mendes and Samukhin use a 'rate equation' method to study the asymptotic degree distribution of their model. They let $A = ma$ and show that for large n,

$$\mathbb{E}\left(\frac{\#\text{degree } k \text{ vertices at step } n}{n}\right) \approx \frac{(1+a)\Gamma((m+1)a+1)\Gamma(k+ma)}{\Gamma(ma)\Gamma(k+2+(m+1)a)}.$$

This can alternatively be written as

$$\mathbb{E}\left(\frac{\#\text{degree } k \text{ vertices at step } n}{n}\right) \approx \frac{B(k+ma; a+2)}{B(ma; a+1)},$$

and is precisely the scale-free distribution that we study in Chapter 3 with $\alpha = ma$ and $\rho = a+1$. This proof is made mathematically precise in [13].

Of the models that we discuss here, the Dorogovtsev, Mendes and Samukhin model is the closest model to our Model BA. The difference between the models is that we do not allow multiple edges to occur, a condition that seems sensible since many networks represent relationships between two vertices that are either present or absent so that multiple edges are meaningless. As we shall see, the asymptotic distribution of our Model BA turns out to be the same as that found by Dorogovtsev, Mendes and Samukhin, and this gives weight to claims made in many papers that the effect of multiple edges is small and thus may be ignored. Of course, further work is needed to fully understand the effect.

The final Barabási-Albert model that we discuss here is given by Athreya, Ghosh and Sethuraman [3]. They describe an 'embedding' of a random graph process in a continuous-time branching process. In this sense it is similar to our Model Y of Chapter 7. Through their embedding the authors obtain a formula for the asymptotic degree distribution of their model and see that this is a power law distribution. A major difference between this model and many Barabási-Albert models is that multiple edges are not only possible but are highly likely. At each time step a random number of edges are added between a new vertex and a single old vertex of the graph.

In this thesis we look at a sequence of different types of Barabási-Albert random graph models. We couple the models together, thus showing how they are related to one another and bounding the distances between the degree distributions of the models. Describing the relationships between such models helps us gain a fuller understanding of the reasons behind the power law degree distribution in Barabási-Albert random graphs.

Our initial model, Model BA, is chosen to be as close as possible to the imprecise description given by Barabási and Albert [5]. We insist that precisely m edges are added at each time step, a restriction that proves key in bounding the distance between the degree sequence of Model BA and its asymptotic distribution, following the example of Bollobás et al. [11].

Of the models that we describe here, two evolve in a very similar way to Model BA. These are the model by Bollobás et al. [11] and the one by Dorogovtsev, Mendes and Samukhin [19]. Both of these models use linear preferential attachment and, in contrast to our model, allow multiple edges to occur.

The differences between the model by Bollobás et al. and the one by Dorogovtsev, Mendes and Samukhin are firstly that Bollobás et al. update the vertex degrees between the addition of new edges, while Dorogovtsev, Mendes and Samukhin update the vertex degrees only at the times of addition of new vertices, and secondly the inclusion of the '*attractiveness*' parameter, A, in the model by Dorogovtsev, Mendes and Samukhin. The result of these differences is that the asymptotic power law distribution obtained by Bollobás et al. [11] is of order k^{-3} for any value of m, while the order depends on A in the model by Dorogovtsev, Mendes and Samukhin.

Like Dorogovtsev, Mendes and Samukhin we do not use updated vertex degrees after the addition of each new edge. In fact, we disallow loops and multiple edges, and insist that the marginal probabilities of choosing a vertex at a given time step are proportional to their degrees.

We obtain an asymptotic degree distribution whose order depends on m and α, in contrast to the asymptotic degree distribution of k^{-3} obtained by Bollobás et al. for all values of their single parameter, m. Like Krapivsky, Redner and Leyvraz [34], we obtain asymptotic power laws of order $k^{-\gamma}$, where $\gamma > 2$ is determined by the model parameters, but we are able to do this using a linear form of preferential attachment.

Our results in Chapter 3 for Yule's model for evolution provide a simple way of determining a candidate for the asymptotic power law distributions of our models. In Chapter 6 we derive a similar Model Y and couple it with discrete-time models in our scheme. The coupling of the models provides insight and an explanation of asymptotic power law distributions for Barabási-Albert random graphs. Moreover, it is hoped that knowing how these models are connected to one another will be useful when studying statistics of the models other than degree sequence. Having chosen the model for which the distribution of the statistic is easiest to study, one would need to consider whether the statistic evolves similarly in the other models and then construct suitable couplings.

Finally, in contrast to most authors (with Szymański [53] as an exception), we not only prove that the asymptotic degree distributions of our Barabási-Albert random graphs are power law distributions, but we bound the distance between a specific power law distribution (described in Chapter 3) and the degree distribution of each graph. Since the original purpose of the Barabási-Albert random graph model was to model real world networks, networks that have a finite number of vertices, it is essential to know how close the degree distributions of our models are to their asymptotic distributions when the graphs are of a given finite size. Additionally, as discussed in Section 1.2.3, degree distribution bounds indicate how good the method of log-log plots is for estimating the exponent parameter of such graphs.

2.3 The approach in this thesis

We study a sequence of models, coupling our Model BA via this sequence with a continuous-time branching process known as Model Y. The models are coupled in such a way that bounds for the distances between their degree distributions can be found. Once these bounds are obtained (in Chapter 6) any degree distribution bounds for a particular model extend to all models within the scheme.

The construction of Model Y is motivated by Yule's model for evolution, and is defined to be of a form that allows bounds for the distance between its degree distribution and the μ_ρ^α-distribution defined in Chapter 3 to be calculated. We use Theorem 3.18, which bounds the distance between two negative binomial distributions, to obtain this result in Chapter 7.

Degree distribution bounds are obtained directly for Model BA in Chapter 8 using recurrence equations and following the example of Szymański [53]. This method is particularly easy once we have a target asymptotic distribution, something that we obtain from studying the continuous-time Model Y.

Once we have obtained bounds for the total variation distance between the specific asymptotic power law distribution and the degree distribution of the graph at step n, we automatically have bounds for $|\mathbb{E}(M_n^{(BA)}(S)) - \mu_\rho^\alpha(S)|$, where $M_n^{(BA)}(S)$ is the random variable representing the number of vertices of in-degree some $k \in S \subset \mathbb{Z}_0^+$ in the graph at time n. Following the example of Bollobás et al. [11], we apply the Azuma-Hoeffding inequality to obtain bounds for the convergence of the random variables $M_n^{(BA)}(S)$ (and therefore for the degree sequence) in the random graph described by Model BA. From these bounds we deduce complete convergence (as the number of vertices tends to infinity) of the degree sequence to our power law distribution, a stronger form of convergence than almost sure convergence.

The Azuma-Hoeffding inequality method applies only to Model BA within our scheme. For other models instead we use Chernoff bounds to bound $M_n(S)$. Again, we are able to deduce complete convergence.

One possible alternative approach would be the '*rate equation*' method used by Dorogovtsev, Mendes and Samukhin for a very similar Barabási-Albert model. However, Dorogovtsev, Mendes and Samukhin only obtain the asymptotic degree distribution for the model, while we aim to bound the distance between the degree distribution and its asymptotic distribution at any given time.

Another possible strategy is explored in Chapter 10. Here we look at an inhomogeneous Markov chain. Studying a sequence of related homogeneous Markov chains (the n^{th} of whose transition probabilities are given by the transition probabilities at step n in the inhomogeneous Markov chain) we show that the sequence of stationary distributions of these homogeneous Markov chains converges to the μ_ρ^α-distribution defined in Chapter 3. Thus, the method in Chapter 10 works as an alternative way of finding a candidate limiting distribution for the degree distribution of Model BA.

Chapter 11 builds on this. We apply Stein's method for the μ_ρ^α-distribution, which is derived in Chapter 5, directly to the inhomogeneous Markov chain of Chapter 10. Although we do not prove convergence to the μ_ρ^α-distribution through this method, we do prove that if the Markov chain converges then its limiting distribution must be the μ_ρ^α-distribution.

Chapter 3

Yule's model for evolution and a general power law distribution

3.1 Introduction

Chapter 2 described a class of Barabási-Albert random graph models, which use preferential attachment mechanisms with the aim of producing random graphs that exhibit a scale-free property. This chapter returns to an earlier model published by Yule [56] and later discussed and extended to further situations by Simon [47]. A brief history of models of this type can be found in [35].

Yule's model was derived as a model for evolution of species. Nevertheless, this model is very closely related to Barabási-Albert random graph models, since species in the model increase in number with rate proportional to the number of species present. We call this a continuous-time preferential attachment condition. Moreover, the asymptotic species-size distribution has been proved to be what is now known as a Yule-Simon* distribution. This is a distribution with a power law tail, and it is a special case of the more general power law distribution from which we shall bound the degree distribution of our Barabási-Albert random graph models.

The structure of this chapter is as follows. We begin by giving definitions of the Yule-Simon probability distribution and a generalisation of this distribution. We discuss some properties of the distributions, in particular showing that the distributions have power law tails.

*The asymptotic distribution was described by Yule [56] and written down in the form that we study by Simon [47].

This is followed by a description of Yule's model for the evolution of species. We discuss how a generalisation of this model could describe the evolution of certain continuous-time Barabási-Albert random graph models. To finish, we prove some new results concerning the rate of convergence of the species-size distribution to our generalisation of the Yule-Simon distribution. The results of this chapter will motivate the construction of the continuous-time Model Y in in Chapter 7.

3.2 Power law distributions

3.2.1 The Yule-Simon distribution

Let us begin by defining the probability distributions that we are going to work with. The Yule-Simon distribution is a discrete probability distribution on the natural numbers. In general, the Yule-Simon distribution with parameter $\rho > 0$ is defined as follows.

Definition 3.1. *For $k \in \mathbb{N}$ and B and Γ representing the usual beta and gamma functions, the measure μ_ρ of the Yule-Simon distribution with parameter ρ is given by*

$$\mu_\rho(k) \quad := \quad \rho B(k; \rho + 1) = \frac{\rho\Gamma(k)\Gamma(\rho + 1)}{\Gamma(k + \rho + 1)}.$$

A Yule-Simon distribution with parameter $\rho > 1$ has power-law tail, $\mu_\rho(k) \sim \frac{\rho\Gamma(\rho+1)}{k^{\rho+1}}$, as $k \to \infty$. This is a standard result (for example see Titchmarsh [54]); in Section 3.2.3 we calculate bounds for the approximation for each $k \in \mathbb{N}$.

Remark 3.2. *Sometimes described as an exponential-geometric mixture distribution, the Yule-Simon distribution can be formed by taking a mixture over geometric distributions with parameters determined by an exponentially distributed random variable. Suppose $\rho > 0$, $U \sim Exponential(\rho)$ and $V \sim Geometric_{\mathbb{N}}(e^{-U})$, by which we mean that, conditional* *on $U = u$, V takes a geometric distribution with parameter e^{-u}. (This notation will be used throughout this thesis and is discussed on page viii.) Then V has a Yule-Simon distribution with parameter ρ.*

This is given as a result from Mathematica by [44] and can also be shown by direct calculation. To see this, let $\rho > 0$, $U \sim Exponential(\rho)$ and $V \sim Geometric_{\mathbb{N}}(e^{-U})$. Then, for $k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}(V = k) &= \int_0^\infty \rho e^{-\rho x} e^{-x}(1-e^{-x})^{k-1}dx \\ &= \rho \int_0^1 (1-t)^\rho t^{k-1} dt \quad \textit{(substituting } t = 1-e^{-x}\textit{)} \\ &= \rho B(k; \rho+1),\end{aligned}$$

as required.

3.2.2 A more general power law distribution

The definition of a more general power law distribution is motivated by the result of Remark 3.2. We note that the negative binomial distribution is a generalisation of the geometric distribution. Following the usual notation of this thesis, we consider the $Negbinom(\alpha, e^{-U})$ where $U \sim Exponential(\rho)$, $\alpha, \rho > 0$. The special case of $\alpha = 1$ provides a geometric distribution on $k = 0, 1, 2, ...$, which is a shifted Yule-Simon distribution, since the Yule-Simon distribution is usually defined on $k = 1, 2, ...$

Definition 3.3. *For $k \in \mathbb{Z}_0^+$, $\alpha, \rho > 0$ and B representing the usual beta function, define the measure μ_ρ^α with parameters ρ and α by*

$$\mu_\rho^\alpha(k) := \frac{B(k+\alpha; \rho+1)}{B(\alpha; \rho)}.$$

Proposition 3.4. *The cumulative distribution function for the μ_ρ^α power law distribution is given, for $n \in \mathbb{Z}_0^+$ by*

$$\sum_{i=0}^{n} \mu_\rho^\alpha(i) = 1 - \frac{B(n+\alpha+1;\rho)}{B(\alpha;\rho)}. \tag{3.1}$$

Proof. Let $n \in \mathbb{Z}_0^+$. We prove the result by calculation using the definition of the beta function,

$$\begin{aligned}
\sum_{i=0}^{n} \mu_\rho^\alpha(i) &= \sum_{i=0}^{n} \frac{B(i+\alpha;\rho+1)}{B(\alpha;\rho)} \\
&= \frac{1}{B(\alpha;\rho)} \sum_{i=0}^{n} \int_0^1 t^{i+\alpha-1}(1-t)^\rho dt \\
&= \frac{1}{B(\alpha;\rho)} \int_0^1 t^{\alpha-1}(1-t)^\rho \sum_{i=0}^{n} t^i dt \\
&= \frac{1}{B(\alpha;\rho)} \int_0^1 t^{\alpha-1}(1-t)^\rho \left(\frac{1-t^{n+1}}{1-t}\right) dt \\
&= \frac{1}{B(\alpha;\rho)} \int_0^1 t^{\alpha-1}(1-t)^{\rho-1} dt - \frac{1}{B(\alpha;\rho)} \int_0^1 t^{n+\alpha}(1-t)^{\rho-1} dt \\
&= 1 - \frac{B(n+\alpha+1;\rho)}{B(\alpha;\rho)},
\end{aligned}$$

as required. □

Remark 3.5. *It follows immediately from Proposition 3.1 that μ_ρ^α is a probability distribution, since trivially $\mu_\rho^\alpha(i) \geq 0$, and*

$$\sum_{i=0}^{n} \mu_\rho^\alpha(i) = 1 - \frac{B(n+\alpha+1;\rho)}{B(\alpha;\rho)} \to 1 \text{ as } n \to \infty.$$

Proposition 3.6. *Let $\alpha, \rho > 0$ and let μ_ρ^α be defined as in Definition 3.3. Let $U \sim$ Exponential(ρ) and let $V \sim Negbinom(\alpha, e^{-U})$. Then for every $k \in \mathbb{Z}_0^+$,*

$$\mathbb{P}(V = k) = \mu_\rho^\alpha(k).$$

Proof. By definition of the negative binomial distribution,

$$\begin{aligned}
\mathbb{P}(V=k) &= \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)}\mathbb{E}(e^{-\alpha U}(1-e^{-U})^k) \\
&= \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)}\int_0^\infty \rho e^{-\rho x}e^{-\alpha x}(1-e^{-x})^k dx \\
&= \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)}\int_0^1 \rho t^k(1-t)^{\rho+\alpha-1}dt \quad \text{(putting } t=1-e^{-x}\text{)} \\
&= \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)} \times \rho B(k+1;\rho+\alpha) \quad \text{(by definition)} \\
&= \frac{\rho\Gamma(k+\alpha)\Gamma(\rho+\alpha)\Gamma(k+1)}{k!\Gamma(\alpha)\Gamma(k+\rho+\alpha+1)} \\
&= \left(\frac{\Gamma(k+\alpha)\Gamma(\rho+1)}{\Gamma(k+\alpha+\rho+1)}\right)\left(\frac{\rho\Gamma(\rho+\alpha)}{\Gamma(\rho+1)\Gamma(\alpha)}\right) \\
&= \frac{B(k+\alpha;\rho+1)}{B(\alpha;\rho)} \\
&= \mu_\rho^\alpha(k),
\end{aligned}$$

as required. □

3.2.3 Distance from a power law

This section gives bounds for the approximation by an exact power law for the tail of the general distribution described above. Simon [47] notes that, for the Yule-Simon distribution, $\mu_\rho(k) \sim \frac{\rho\Gamma(\rho+1)}{k^{\rho+1}}$ as $k \to \infty$. For the general distribution (which includes a shifted Yule-Simon distribution) on $\mathbb{Z}_0^+$ we now bound $|\mu_\rho^\alpha(k) - \frac{\Gamma(\rho+1)}{(k+\alpha)^{\rho+1}B(\alpha;\rho)}|$ for finite k. The result in Simon [47] is proved by Titchmarsh [54]. Our proof closely follows the proof in Titchmarsh [54] but gives precise bounds for the difference in values. An interpretation of the result is that the μ_ρ^α-distribution exhibits power law decay in the tail of the distribution.

Theorem 3.7. *For $\rho > 1$, $\alpha > 0$, let $\mu_\rho^\alpha(k) := \frac{B(k+\alpha;\rho+1)}{B(\alpha;\rho)}$ for $k = 0, 1, 2, ...$ Then for any $k = 0, 1, 2, ...,$*

$$\left|\mu_\rho^\alpha(k) - \frac{\Gamma(\rho+1)}{B(\alpha;\rho)(k+\alpha)^{\rho+1}}\right| \leq \frac{\rho\Gamma(\rho+2)}{B(\alpha;\rho)(k+\alpha)^{\rho+2}}.$$

Proof. Let $k \in \mathbb{Z}_0^+$. By definition,

$$\begin{aligned}
\mu_\rho^\alpha(k) &= \frac{B(k+\alpha;\rho+1)}{B(\alpha;\rho)} \\
&= \frac{1}{B(\alpha;\rho)} \int_0^1 t^\rho (1-t)^{k+\alpha-1} dt \\
&= \frac{1}{B(\alpha;\rho)} \int_0^\infty (1-e^{-t})^\rho e^{-(k+\alpha)t} dt \\
&= \mathcal{I}_1 - \mathcal{I}_2,
\end{aligned}$$

where $\mathcal{I}_1 = \frac{1}{B(\alpha;\rho)} \int_0^\infty t^\rho e^{-(k+\alpha)t} dt$ and $\mathcal{I}_2 = \frac{1}{B(\alpha;\rho)} \int_0^\infty (t^\rho - (1-e^{-t})^\rho) e^{-(k+\alpha)t} dt.$

By substitution of $s = (k+\alpha)t$, $\mathcal{I}_1 = \frac{\Gamma(\rho+1)}{B(\alpha;\rho)(k+\alpha)^{\rho+1}}$, so

$$\left|\mu_\rho^\alpha(k) - \frac{\Gamma(\rho+1)}{B(\alpha;\rho)(k+\alpha)^{\rho+1}}\right| = |\mu_\rho^\alpha(k) - \mathcal{I}_1| = |\mathcal{I}_2|.$$

Now

$$\begin{aligned}
|\mathcal{I}_2| &= \frac{1}{B(\alpha;\rho)} \int_0^\infty (t^\rho - (1-e^{-t})^\rho) e^{-(k+\alpha)t} dt \\
&= \frac{1}{B(\alpha;\rho)} \int_0^1 (t^\rho - (1-e^{-t})^\rho) e^{-(k+\alpha)t} dt \\
&\quad + \frac{1}{B(\alpha;\rho)} \int_1^\infty (t^\rho - (1-e^{-t})^\rho) e^{-(k+\alpha)t} dt \\
&\leq \frac{1}{B(\alpha;\rho)} \int_0^1 \left(1 - \left(1 - \frac{1}{2}t\right)^\rho\right) t^\rho e^{-(k+\alpha)t} dt + \frac{1}{B(\alpha;\rho)} \int_1^\infty t^\rho e^{-(k+\alpha)t} dt.
\end{aligned}$$

Consider $f(t) := \frac{1-(1-\frac{1}{2}t)^{\rho}}{t}$ on $(0,1]$. Then,

$$\sup_{t\in(0,1]} f(t) = \max\{\lim_{t\to 0} f(t), f(1), \sup_{t\in(0,1]}\{f(t_0) : f'(t_0)=0\}\}.$$

Now we can see immediately that $f(1) = \frac{1-(\frac{1}{2})^{\rho}}{1} < 1$ since $\rho > 1 > 0$. Moreover, the function f is a well-defined ratio of continuously differentiable functions on $(0,1]$. Thus, by l'Hôpital's rule,

$$\begin{aligned}
\lim_{t\to 0} f(t) &= \lim_{t\to 0}\left(\frac{d}{dt}\left(1-\left(1-\frac{1}{2}t\right)^{\rho}\right)\right) \\
&= \frac{\rho}{2}\left(1-\frac{1}{2}t\right)^{\rho-1}\Bigg|_{t=0} \\
&= \frac{\rho}{2}.
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{d}{dt}\left(\frac{(1-(1-\frac{1}{2}t)^{\rho})}{t}\right) = 0 \quad &\Longleftrightarrow \quad \frac{t(\frac{1}{2}\rho(1-\frac{1}{2}t)^{\rho-1}) - (1-(1-\frac{1}{2}t)^{\rho})}{t^2} = 0 \\
&\Longleftrightarrow \quad \frac{\rho}{2}t\left(1-\frac{1}{2}t\right)^{\rho-1} = 1-\left(1-\frac{1}{2}t\right)^{\rho} \\
&\Longleftrightarrow \quad \left(1-\frac{1}{2}t\right)^{\rho-1} = \frac{1}{1+\frac{1}{2}(\rho-1)t}.
\end{aligned}$$

Thus, if t_0 satisfies $f'(t_0) = 0$, then,

$$\begin{aligned}
f(t_0) &= \frac{1-\frac{1-\frac{1}{2}t_0}{1+\frac{1}{2}(\rho-1)t_0}}{t_0} \\
&= \frac{1+\frac{1}{2}(\rho-1)t_0 - 1 + \frac{1}{2}t_0}{t_0(1+\frac{1}{2}(\rho-1)t_0)} \\
&= \frac{\rho}{2+(\rho-1)t_0} \le \frac{\rho}{2}, \qquad \text{since } \rho > 1.
\end{aligned}$$

The combination of these results gives,

$$\sup_{t\in(0,1]} f(t) \leq \max\left(1, \frac{\rho}{2}\right) \leq \rho, \text{ since } \rho > 1.$$

So,

$$\begin{aligned} |\mathcal{I}_2| &\leq \frac{\rho}{B(\alpha;\rho)}\int_0^1 t^{\rho+1}e^{-(k+\alpha)t}dt + \frac{1}{B(\alpha;\rho)}\int_1^\infty t^{\rho}e^{-(k+\alpha)t}dt \\ &\leq \frac{\rho}{B(\alpha;\rho)}\int_0^\infty t^{\rho+1}e^{-(k+\alpha)t}dt \\ &= \frac{\rho\Gamma(\rho+2)}{B(\alpha;\rho)(k+\alpha)^{\rho+2}}. \end{aligned}$$

□

Remark 3.8. *It follows that the relative error of $\mu_\rho^\alpha(k)$ from $\frac{\Gamma(\rho+1)}{B(\alpha;\rho)(k+\alpha)^{\rho+1}}$ is bounded above by $\frac{\rho(\rho+1)}{k+\alpha}$.*

3.2.4 The value of ρ

Section 1.2.2 discussed the values of γ seen in scale-free degree distributions of real-world networks. We now consider which values of ρ these correspond to in the distribution with measure μ_ρ^α described above. As we have seen, $\mu_\rho^\alpha(k) \sim \frac{\Gamma(\rho+1)}{B(\alpha;\rho)(k+\alpha)^{\rho+1}}$, as $k \to \infty$.

If our random graph model were to have in-degree distribution with measure μ_ρ^α, the expected number of edges directed into a randomly chosen vertex would be $\mathbb{E}(V)$ for the random variable V satisfying $\mathcal{L}(V) = \mu_\rho^\alpha$.

Proposition 3.9. *For $\rho > 1$, $\alpha > 0$, let $\mu_\rho^\alpha(k) := \frac{B(k+\alpha;\rho+1)}{B(\alpha;\rho)}$ for $k = 0, 1, 2, \ldots$. Let V be a random variable with measure μ_ρ^α. Then*

$$\mathbb{E}(V) = \frac{\alpha}{\rho-1}.$$

Proof. By integration,

$$
\begin{aligned}
\mathbb{E}(V) &= \frac{1}{B(\alpha;\rho)} \sum_{k=0}^{\infty} kB(k+\alpha;\rho+1) \\
&= \frac{1}{B(\alpha;\rho)} \sum_{k=1}^{\infty} \int_0^1 kt^{k+\alpha-1}(1-t)^{\rho} dt \\
&= \frac{1}{B(\alpha;\rho)} \int_0^1 (1-t)^{\rho} t^{\alpha-1} \left(\sum_{k=1}^{\infty} kt^k \right) dt \\
&= \frac{1}{B(\alpha;\rho)} \int_0^1 (1-t)^{\rho} t^{\alpha-1} \left(\frac{t}{(1-t)^2} \right) dt \\
&= \frac{1}{B(\alpha;\rho)} \int_0^1 (1-t)^{\rho-2} t^{\alpha} dt \\
&= \frac{B(\alpha+1;\rho-1)}{B(\alpha;\rho)} \\
&= \frac{\alpha}{\rho-1}.
\end{aligned}
$$

□

The value of $\mathbb{E}(V)$ corresponds to the expected number of edges added per vertex. Thus values of ρ lying between 1 and $1+\alpha$ correspond to those models where the expected number of edges added per vertex lies between 1 and infinity. This explains the observation discussed in Chapter 1 that most of the studied real-world power-law tails take the form $k^{\rho+1}\mu(k) \to c$ as $k \to \infty$ for some constant c and for some relatively small $\rho > 1$. Many of the results proved in Chapter 7 require the condition that $\rho > 1$, a condition that is required for $\mathbb{E}(V)$ to be defined.

3.3 Yule's model for evolution

Yule's model for evolution is relevant to us since it is a continuous-time stochastic process that exhibits a preferential attachment condition. We consider the parallels between Yule's model for evolution and Model BA. The next section looks at a generalisation of Yule's model for evolution. The more general version will be compared with versions of Model BA.

Yule's model for evolution consists of a number of genera, each of which consists of a number of species. It is assumed that the number of genera and sizes of the genera are non-decreasing as time passes. The precise evolution of the model can be described using a type of stochastic process, now known as a *Yule process.*

Before defining the model in terms of stochastic processes we look at the requirements that Yule [56] set out for the model.

3.3.1 The requirements of Yule's model for evolution

We have a number of different genera, each genus consisting of a number of different species. Two types of mutation are permitted. The first type of mutation causes the creation of a new species within a particular genus while the second type of mutation causes the creation of new genus. The mathematical requirements are as follows.

1. Over a given length of time-interval, within each species there is a chance of a mutation causing the creation of a new species of the same genus. The probability of this happening is constant for all species and for all time.

2. Over a given length of time-interval, within each genus there is a chance of a mutation causing the creation of a new genus of one species. The probability of this happening is constant for all genera and for all time.

The first condition is the preferential attachment part of the model. The larger the number of species within a genus, the more likely that genus is to gain a new species. Mathematically, we can describe this process using a particular branching process or, with some loss of information, as a pure birth process. Precise definitions are given below.

3.3.2 The Yule process as a branching process

Defining the Yule process as a branching process provides a helpful visualisation for understanding Yule's model for evolution. We begin this section by defining suitable notation for branching processes in general, before giving the conditions for the Yule process. Introductory literature on branching processes is easy to find, for example see [2], [26], [30].

A branching process is a particular type of stochastic process. We may picture a branching process as representing the size of a population through time. At any time t we have a population consisting of z_t individuals, $(i_t(1), ..., i_t(z_t))$. Each individual i has associated random variables for the number of '*offspring*' it will produce at the time of its death, ξ_i, and the length of its life, λ_i. Individuals branch independently of one another and the number of offspring, ξ_i, for a given individual, i, is independent of the rest of the process.

We start the process at time $t = 0$ with a given number z_0 of individuals, all of whom have birth time $\beta_{i_0(j)} = 0$. Every other individual in the process is a descendant of one of these individuals. The birth time of an individual is defined to be the death time of its '*parent*'. The death time of individual i is given by $\delta_i = \beta_i + \lambda_i$.

A Yule process with rate parameter ρ is defined to be a branching process where individual's life-lengths are independent identically distributed $Exponential(\rho)$ random variables, and each individual gives birth to precisely 2 offspring at its death. In the notation above this is equivalent to the described branching process with $\lambda_i \sim Exponential(\rho)$ for all i and $\xi_i = 2$ for all i.

In the branching process definition of a Yule process, at the death of an individual one of its offspring represents the individual itself while the other represents the new individual (created by the mutation in Yule's model for evolution). For this reason an alternative definition is a specific pure birth process.

3.3.3 The Yule process as a pure birth process

Alternatively, the definition of a Yule process is sometimes given in the form of a pure birth process (for example Ross [46]). In this case we do not keep track of the relationship between '*parents*' and '*offspring*' or, using the language of Yule's model for evolution, we record when new genera and species are formed but not in which old genus or species the mutation occurred.

If we think about the branching process definition above, when the total population size is n the time until the next death of any individual in the process is $Exponential(n\rho)$, owing to the exponentially distributed inter-arrival times. At this time of the next death, the total population size increases by 1.

Thus a Yule process can be thought of as a pure birth process where the time between the birth of the n^{th} and $n + 1^{st}$ individuals is $Exponential((n + 1)\rho)$.

3.3.4 The mathematical model

To relate the Yule process to Yule's model for evolution, we consider the two required types of mutation in Yule's model for evolution in turn.

Yule's first requirement is that within each species in any time-interval there should be a chance of a mutation causing the creation of a new species of the same genus, and that the probability of this happening should be constant for all species and for all time. We can model each genus as an independent Yule process of a given rate with the number of individuals in the process representing the number of species of that genus. By definition and properties of the exponential distribution, each species branches (i.e. mutates to create a new species) at the same rate, and this rate is constant for all species and for all time. Let us define this rate to be 1, so that each species is modelled by an independent Yule process of rate 1.

We have now fulfilled the first requirement. The second requirement is that within each genus in any time-interval there should be a chance of a mutation causing the creation of a new genus, and that the probability of this happening should be constant for all species and for all time. By the same argument this condition is satisfied by modelling the genera as a Yule process. In this case let us define the rate of the Yule process to be some constant $\rho > 0$ to allow any ratio desired between the rates of the two types of mutation.

Thus Yule's assumptions correspond to the following mathematical model. We begin with one genus consisting of one species. The number of genera follows a Yule process of some rate $\rho > 0$, while the number of species within each genus follows an independent Yule process of rate 1. Yule [56] provides a proof that the asymptotic frequency distribution, as time tends to infinity, for the sizes of genera follows what we now call the $Yule-Simon(\rho)$ distribution, that is a shifted version of the μ_ρ^α-distribution with $\alpha = 1$.

3.3.5 A class of continuous-time Barabási-Albert random graph models

We now discuss how Yule's model can be understood directly to model scale-free networks. First consider what is happening within the model at a given time-point, t. A random number of genera is present and each genus is gaining new species at a rate given by the current size of that genus. For a given genus, the probability that the next new species will be within that genus is proportional to the size of that genus. This is analogous to the preferential attachment condition of the Barabási-Albert random graph model.

We now describe a class of continuous-time Barabási-Albert random graph models. We do not define the evolution of the random graphs completely, but the mathematical model described by Yule is used to determine how the degrees of vertices in these models evolve. Each genus in Yule's model represents a vertex of the evolving random graphs, and the number of species in that genus represents the degree of that vertex. The result proved by Yule tells us that in such models the asymptotic degree distribution for each of the random graphs is the $Yule - Simon(\rho)$ distribution.

Graph models described by Yule's model for evolution

The models consist of two independent sets of processes:

1. **Births of new vertices into the graphs.** The time between arrivals is exponentially distributed with rate, $\rho\times$ number of vertices in the graph.
2. **Growth of edges for current vertices.** Each vertex gains incident in-edges independently with rate proportional to 1+ in-degree.

Each vertex has 'degree' 1+ number of incident half-edges. There are various options for using these half-edges to create a directed or undirected graph from the graph. The choice of method affects the properties of the graphs created and should be chosen depending on the properties required of the model. Examples of possible options include,

- Edges can be formed joining half-edges together in pairs in the order in which they are added to the graph. This means that loops and multiple edges can occur and that the graph may not be connected.
- We can consider half-lines as edges directed in (or out) of the vertices. When a new vertex appears in the graph, create out (or in) half-lines from that vertex to match with any remaining unattached half-lines of the graph. This creates a random graph with no loops, however the graph still may have multiple edges and may not be connected. In the case of this graph it is the 1+ in-degree distribution that takes the Yule-Simon distribution.
- If we want to prevent multiple edges we could consider attaching 'degree 0' vertices to half-lines whenever the alternative would mean creating a multiple edge or a loop. Such vertices would never gain new edges, since attachment is proportional to vertex degree.

Notation

- We write S_i for the time between the births of the i^{th} and $(i+1)^{st}$ new vertices, so $S_i \sim Exponential(\rho i)$.
- We write $Y_i(t)$ for the degree of vertex i at time t. We set $Y_1(0) = 1$, and for $i \geq 2$ we set $Y_i(S_1 + ... + S_{i-1}) := 1$. For $i \geq 1$ and $t \geq S_1 + ... + S_{i-1}$, the process $Y_i(t)$ is defined to evolve as a Yule process with rate 1.

3.4 A generalisation of Yule's model for evolution

In the previous section we considered a Barabási-Albert graph model that can be represented in terms of Yule processes in Yule's model for evolution. This model can be compared to the $\alpha = 1$ case in Model BA. We now consider a generalisation that allows us to study more values of α. Recall the graph model defined in Section 3.3.5.

A more general class of models

We would like to change the rate of addition of new edges so that the following are satisfied. The models that we study consist of two independent sets of processes:

1. **Births of new vertices into the graphs.** The time between arrivals is exponentially distributed with rate, $\rho\times$ number of vertices in the graph.
2. **Growth of edges for current vertices.** Each vertex gains incident in-edges independently with rate proportional to $\alpha+$ in-degree.

The definition of this model is included in Appendix B for easy reference.

It is easy to construct such a model using modified Yule processes. We use a process that is sometimes known as a Yule process with immigration with immigration rate α and birth rate 1. The process starts with 0 individuals and is defined as a pure birth process with rate $\alpha + i$ where i is the number of individuals in the process. The 'degree' of a vertex is now α+in-degree.

Notation

- We write S_i for the time between the births of the i^{th} and $(i+1)^{st}$ new vertices, so $S_i \sim Exponential(\rho i)$.
- We write $Y_i(t)$ for the in-degree of vertex i at time t. We set $Y_1(0) = 0$, and for $i \geq 2$ we set $Y_i(S_1 + ... + S_{i-1}) := 0$. For $i \geq 1$ and $t \geq S_1 + ... + S_{i-1}$, $Y_i(t)$ is defined to evolve as a Yule process with immigration rate $\alpha > 0$ and birth rate 1.

3.5 Some new results at a sequence of stopping times

Given that one aim of this thesis is to study scale-free random graph models, the results proved concerning the generalised version of Yule's model are stated in the context of the represented continuous-time Barabási-Albert graph models. Of course, the mathematical results hold for whatever real-life interpretation of this mathematical model we choose to make.

The scale-free probability distribution described in Section 3.2.2 was seen to be a mixture distribution formed by a negative binomial distribution with random parameter. The next results will link the Yule processes in these continuous-time random graph models to our μ_ρ^α-distribution.

Lemma 3.10 (Ross [46]). *For a Yule (pure-birth) process, $Y(t)$ of rate λ satisfying $Y(0) = 1$, and for $n = 1, 2, ...,$*

$$\mathbb{P}(Y(t) = n) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}.$$

The following theorem is a generalisation of Lemma 3.10 when $\lambda = 1$, for Yule processes with immigration.

Theorem 3.11. *For a Yule process with immigration, $Y(t)$, that has immigration rate $\alpha > 0$ and birth rate* 1*,*

$$Y(t) \sim Negbinom(\alpha, e^{-t}).$$

Proof. Let $Y(t)$ be a Yule process with immigration that has immigration rate $\alpha > 0$ and birth rate 1. We want to prove that, for all $k \in \mathbb{Z}_0^+$,

$$\mathbb{P}(Y(t) = k) = \frac{\Gamma(k+\alpha)}{k!\Gamma(a)} e^{-\alpha t}(1-e^{-t})^k. \tag{3.2}$$

We proceed by induction on k. For $k = 0$,

$$\begin{aligned} \mathbb{P}(Y(t) = 0) &= 1 - \mathbb{P}(Y(t) \geq 1) \\ &= 1 - \mathbb{P}(S_1 \leq t) \\ &= e^{-\alpha t}, \end{aligned}$$

as required.

Suppose that for some $n \in \mathbb{N}$, (3.2) holds for all $k < n$. Then by the equivalence of the two alternative definitions of the negative binomial distribution on page viii we know that, for any $t > 0$, $\mathbb{P}(Y(t) \leq n-1) = \frac{\int_0^{e^{-t}} s^{\alpha-1}(1-s)^{n-1} ds}{B(\alpha;n)}$.

Since $\mathbb{P}(S_1 + ... + S_n \leq x) = \mathbb{P}(Y(x) \geq n)$, by differentiation we obtain the density function for $S_1 + ... + S_n$,

$$f_{S_1+...+S_n}(x) = \frac{d}{dx}\left(1 - \frac{\int_0^{e^{-x}} s^{\alpha-1}(1-s)^{n-1} ds}{B(\alpha;n)}\right) = \frac{e^{-\alpha x}(1-e^{-x})^{n-1}}{B(\alpha;n)}.$$

Thus, conditioning on $S_1 + ... + S_n = x$,

$$\begin{aligned}
\mathbb{P}(Y(t) = n) &= \mathbb{P}(S_1 + ... + S_n \leq t) - \mathbb{P}(S_1 + ... + S_{n+1} \leq t) \\
&= \int_0^t (1 - \mathbb{P}(S_{n+1} \leq t - x | S_1 + ... + S_n = x)) \frac{e^{-\alpha x}(1-e^{-x})^{n-1}}{B(\alpha; n)} dx \\
&= \int_0^t e^{-(\alpha+n)(t-x)} \frac{e^{-\alpha x}(1-e^{-x})^{n-1}}{B(\alpha; n)} dx \\
&= \frac{e^{-(\alpha+n)t}}{B(\alpha; n)} \int_0^t e^{nx}(1-e^{-x})^{n-1} dx \\
&= \frac{\Gamma(n+\alpha)}{n!\Gamma(a)} e^{-\alpha t}(1-e^{-t})^n,
\end{aligned}$$

as required.

$\square$

Since vertices follow independent but identically defined Yule processes which start at different time points, it is helpful to have a concept of the '*age*' of a vertex. We define this in such a way that all vertices of a given age have the same degree distribution.

Definition 3.12. *We say that vertex 1 has age t at time t, and that vertex $i > 1$ has age t at time $S_1 + ... + S_{i-1} + t$.*

Corollary 3.13. *For each vertex i of age t, the degree distribution of vertex i takes an independent identically distributed negative binomial distribution. More exactly,*

$$Y_1(t) \sim Negbinom(\alpha, e^{-t}),$$

and, for $i \geq 2$,

$$Y_i(S_1 + ... + S_{i-1} + t) \sim Negbinom(\alpha, e^{-t}).$$

Proof. This is an immediate corollary of Theorem 3.11 and the definition of $\{Y_i\}_{i \geq 1}$. $\square$

We are interested in bounding the degree distributions of our random graphs from the power law distribution of Section 3.2.2. As the discussion in Section 1.2.1 explains, by '*degree distribution*' we mean the distribution of the random variable $D(I_n)$, defined to be the degree of vertex $I_n \sim Uniform\{1, 2, ..., n\}$ in a random graph $\mathcal{G}$ on n vertices.

By Corollary 3.13 we know that the degree distribution of a particular vertex in our continuous-time random graph depends only on its age and initial degree. Since all the vertices have the same initial in-degree, 0, it is sufficient to find the distribution of the age of a vertex selected uniformly at random. The following theorem considers the time infinitesimally before each vertex enters the graph. We prove that at the sequence of stopping times, $\{S_1 + S_2 + ... + S_n\}_{n\in\mathbb{N}}$, given by the arrivals of new vertices, the age of a randomly chosen vertex takes an *Exponential*(ρ) distribution. That this is asymptotically true is a known result. For example it is indicated by Yule's [56] calculations when finding the limiting form of the frequency distribution.

Theorem 3.14. *For $n \in \mathbb{N}$, $1 \le i \le n$, $\rho > 0$, let $S_i \sim Exponential(\rho i)$ be independent random variables. Let $I_n \sim Uniform\{1, 2, ..., n\}$ and independent of $S_1, ..., S_n, ...$. Then,*

$$R(n) := S_{I_n} + S_{I_n+1} + ... + S_n \sim Exponential(\rho).$$

Proof. We prove that for $n \in \mathbb{N}$, $R(n)$ has the required moment generating function. First note that the moment generating function for an *Exponential*(ρ) random variable is,

$$M(t) = \frac{\rho}{\rho - t}, \text{ for } t < \rho.$$

We proceed by induction. Clearly $R(1) \sim Exponential(\rho)$ by definition.

Now suppose that for some $k \ge 1$ we know that $R(k) \sim Exponential(\rho)$. Conditioning on I_k gives, for $t < \rho$,

$$\frac{\rho}{\rho - t} = M_{R(k)}(t) = \frac{1}{k} \sum_{i=1}^{k} \prod_{r=i}^{k} \frac{\rho r}{\rho r - t}.$$

Then, to calculate $M_{R(k+1)}(t)$ we condition on I_{k+1}. This gives, for $t < \rho$,

$$\begin{aligned}
M_{R(k+1)}(t) &= \frac{1}{k+1}\sum_{i=1}^{k+1}\prod_{r=i}^{k+1}\frac{\rho r}{\rho r - t}, \\
&= \frac{k}{k+1}\left(\frac{\rho(k+1)}{\rho(k+1)-t}\left(\frac{1}{k}+\frac{1}{k}\sum_{i=1}^{k}\prod_{r=i}^{k}\frac{\rho r}{\rho r - t}\right)\right) \\
&= \frac{k}{k+1}\left(\frac{\rho(k+1)}{\rho(k+1)-t}\left(\frac{1}{k}+M_{R(k)}(t)\right)\right) \\
&= \frac{k}{k+1}\left(\frac{\rho(k+1)}{\rho(k+1)-t}\left(\frac{1}{k}+\frac{\rho}{\rho - t}\right)\right) \quad \text{(by inductive hypothesis)} \\
&= \frac{\rho}{\rho - t}.
\end{aligned}$$

□

Corollary 3.15. *For $n \in \mathbb{N}$, $1 \le i \le n$, and both $\alpha, \rho > 0$, let $S_i \sim Exponential(\rho i)$ be independent random variables. Let $I_n \sim Uniform\{1, 2, ..., n\}$ and independent of $S_1, ..., S_n, ...$. Then, using the standard notation of this thesis, if V is a random variable with distribution $V \sim Negbinom(\alpha, e^{-(S_{I_n}+...+S_n)})$ then for $k = 0, 1, 2, ...$,*

$$\mathbb{P}(V = k) = \frac{B(k+\alpha;\rho+1)}{B(\alpha;\rho)} = \mu_\rho^\alpha(k).$$

Proof. Let $n \in \mathbb{N}$, both $\alpha, \rho > 0$ and let S_i, $1 \le i \le n$, I_n be defined as above.

Suppose $V \sim Negbinom(\alpha, e^{-(S_{I_n}+...+S_n)})$. Then it follows from Theorem 3.14 that $V \sim Negbinom(\alpha, e^{-U})$ for $U \sim Exponential(\rho)$. Thus by Proposition 3.2

$$\mathbb{P}(V = k) = \frac{B(k+\alpha;\rho+1)}{B(\alpha;\rho)} = \mu_\rho^\alpha(k),$$

as required. □

We can now prove that at the sequence of stopping times, $\{S_1 + S_2 + ... + S_n\}_{n\in\mathbb{N}}$, which represents the times when new vertices enter the graph, the degree distribution of the graph is precisely the power law distribution with measure μ_ρ^α.

Theorem 3.16. *For $n \in \mathbb{N}$, let I_n be a uniform random variable on $\{1, 2, ..., n\}$. Then for $k = 0, 1, 2, ...$,*

$$\mathbb{P}(Y_{I_n}(S_1 + ... + S_n) = k) = \mu_\rho^\alpha(k),$$

where μ_ρ^α is the measure defined in Section 3.2.2.

Proof. By construction, for $t > S_1 + S_2 + ... + S_{i-1}$, the distributions of $(S_1, S_2, S_3, ...)$ and $\{Y_i(t - S_1 - S_2 - ... - S_{i-1})\}_{i \in \mathbb{N}}$ are independent. By Theorem 3.14, $S_{I_n} + S_{I_n+1} + ... + S_n \sim$ *Exponential*(ρ). Since, by Corollary 3.13, $Y_{I_n}(S_1 + ... + S_{I_n-1} + t)$ is identically distributed $Y_1(t)$, it follows that for $k = 0, 1, 2, ...$,

$$\begin{aligned}
&\mathbb{P}(Y_{I_n}(S_1 + ... + S_n) = k) \\
&\quad = \mathbb{P}(Y_1(S_{I_n} + ... + S_n)) = k) \\
&\quad = \mathbb{P}(Y_1(U) = k) \quad (\text{by Corollary 3.15, with } U \sim \mathit{Exponential}(\rho)) \\
&\quad = \mu_\rho^\alpha(k) \quad (\text{by Corollary 3.13 and definition of } \mu_\rho^\alpha),
\end{aligned}$$

as required. □

This result provides a class of models that generate random graphs with any given number of vertices, where the degree distribution has measure μ_ρ^α. Graphs produced by this process take precisely the required distribution at the random sequence of times infinitesimally before new vertices are added to the graph. It follows that a graph formed by taking the union of any finite number of such finite-sized random graphs (where each graph is observed at the moment infinitesimally before the addition of a vertex) will have the same distribution. This provides scope for producing scale-free random graphs with many different structural properties.

The particular case of $\alpha = 1, \rho = 2$ is the most closely related to the more traditional Barabási-Albert random graph models; in this case edges and new vertices are added at approximately the same rates. The resulting distribution of a randomly chosen vertex takes the Yule-Simon distribution with degree 3 power law tail,

$$\mu(k) = \frac{4}{k(k+1)(k+2)}.$$

3.6 Bounds for convergence in Yule's model for evolution

We now find bounds in total variation distance between the degree distribution in Yule's model for evolution and a power law for all times after the graph has at least n vertices. The method hinges on Theorem 3.18 below, where we prove bounds between two negative binomial distributions. The success of this method motivates the construction in Chapter 6 of a continuous-time evolving random graph, Model Y, which we compare to discrete-time evolving random graphs. In Chapter 7 we apply the method below to Model Y.

We begin by proving general results giving total variation distance bounds between two negative binomial distributions.

Lemma 3.17. *Let $p, q \in (0,1]$ and let $X \sim Geometric_{\mathbb{Z}_0^+}(p)$, $Y \sim Geometric_{\mathbb{Z}_0^+}(q)$ with X and Y independent. Then*

$$d_{TV}(X,Y) \le \left|\frac{p}{q} - 1\right|.$$

Proof. We need to prove two cases:

$$\begin{aligned} &(i) \quad \text{if} \quad q \le p \quad \text{then} \quad d_{TV}(X,Y) \le \frac{p}{q} - 1; \\ &(ii) \quad \text{if} \quad q > p \quad \text{then} \quad d_{TV}(X,Y) \le 1 - \frac{p}{q}. \end{aligned}$$

Now, $0 \leq 1 - \frac{q}{p} \leq \frac{p}{q} - 1$ so for (i) it is sufficient to prove that

$$\text{if } q \leq p \text{ then } d_{TV}(X,Y) \leq 1 - \frac{q}{p}.$$

Moreover, by symmetry, the same result proves (ii). Assume that $p \geq q$.

Let $(U_i)_i$, $(V_i)_i$ be sequences of independent identically distributed random variables with $U_i \sim Bernoulli(p)$ and $V_i \sim Bernoulli(q)$. Using a well-known construction for the geometric distribution we let $X = \inf\{i \in \mathbb{Z}_0^+ : U_i = 1\}$, and $Y = \inf\{i \in \mathbb{Z}_0^+ : V_i = 1\}$.

For each $i \in \mathbb{Z}_0^+$ we couple (U_i, V_i) so that $\mathbb{P}(U_i = V_i = 1) = q$ and $\mathbb{P}(U_i = V_i = 0) = 1 - p$. Thus

$$\begin{aligned} d_{TV}(X,Y) &\leq \mathbb{P}(X \neq Y) \\ &= \sum_{j=0}^{\infty} \mathbb{P}(\{\forall i < j \quad U_i = V_i = 0\} \quad \cap \quad U_j \neq V_j) \\ &= \sum_{j=0}^{\infty} (1-p)^j (p-q) \\ &= \left(1 - \frac{q}{p}\right), \end{aligned}$$

as required. □

Theorem 3.18. *Let $p, q \in (0,1]$ and let $\alpha > 0$. Let X, Y be random variables satisfying $X \sim Negbinom(\alpha, p)$ and $Y \sim Negbinom(\alpha, q)$. Then for all $\alpha \in \mathcal{A}$*

$$d_{TV}(X,Y) \leq K(\alpha)\Big|\frac{p}{q} - 1\Big|,$$

where,

$$K(\alpha) := \alpha \quad \textit{and} \quad \mathcal{A} = \mathbb{N}. \tag{3.3}$$

Proof. We extend the coupling of Lemma 3.17. Since $\alpha \in \mathbb{N}$, we can write $X = \sum_{i=1}^{\alpha} U_i$ and $Y = \sum_{i=1}^{\alpha} W_i$ where $\{U_i\}_i$ and $\{W_i\}_i$ are sets of independent identically distributed random variables, $U_i \sim Geometric_{\mathbb{Z}_0^+}(p)$ and $W_i \sim Geometric_{\mathbb{Z}_0^+}(q)$ and are coupled in the same way as X and Y in the proof of Lemma 3.17. Thus,

$$\begin{aligned}
d_{TV}(X,Y) &= d_{TV}\left(\sum_{i=1}^{\alpha} U_i, \sum_{i=1}^{\alpha} W_i\right) \\
&= \inf_{\{U_i,W_i\}_i} \mathbb{P}\left(\sum_{i=1}^{\alpha} U_i \neq \sum_{i=1}^{\alpha} W_i\right) \\
&\leq \sum_{i=0}^{\alpha} \inf_{(U_i,W_i)} \mathbb{P}\left(U_i \neq W_i\right) \\
&= \sum_{i=1}^{\alpha} d_{TV}(U_i,W_i) \leq \alpha \left|\frac{p}{q} - 1\right|,
\end{aligned}$$

as required. $\square$

Remark 3.19. *It seems likely that Theorem 3.18 can be extended to a larger set $\mathcal{A}$, perhaps for a different function $K(\alpha)$. It is for this reason that we use the notation of $\mathcal{A}$ and $K(\alpha)$, indicating that any results based on this theorem extend to all sets $\{\mathcal{A}, K(\alpha)\}$ to which Theorem 3.18 can be extended.*

The next theorem illustrates a method, using ideas from [1], that gives a bound for general $\alpha \geq 1$. However, the bound is not as strong as in Theorem 3.18 since

$$p > q \implies \frac{p}{q} - 1 > 1 - \frac{q}{p}.$$

Theorem 3.20. *Let $p > q \in (0,1]$ and let $\alpha > 1$. Let X, Y be random variables satisfying $X \sim Negbinom(\alpha, p)$ and $Y \sim Negbinom(\alpha, q)$. Then*

$$d_{TV}(X,Y) \leq \overline{K}(\alpha)\left(\frac{p}{q} - 1\right), \quad \textit{where} \quad \overline{K}(\alpha) := \alpha.$$

Proof. We proceed as follows. By definition,

$$d_{TV}(X,Y) = \sum_{r=0}^{\infty} \frac{\Gamma(\alpha+r)}{2r!\Gamma(\alpha)}\Big| ((1-p)^r p^\alpha - (1-q)^r q^\alpha)\Big|.$$

Since $p \geq q$, there exists some $k \in \mathbb{N}$ such that $\forall r < k \quad (1-q)^r q^\alpha \leq (1-p)^r p^\alpha$ and $\forall r \geq k \quad (1-q)^r q^\alpha \geq (1-p)^r p^\alpha$. Thus,

$$\begin{aligned}
d_{TV}(X,Y) &= \sum_{r=0}^{\infty} \frac{\Gamma(\alpha+r)}{2r!\Gamma(\alpha)}|(1-p)^r p^\alpha - (1-q)^r q^\alpha| \\
&= \sum_{r=0}^{k-1} \frac{\Gamma(\alpha+r)}{2r!\Gamma(\alpha)}\{(1-p)^r p^\alpha - (1-q)^r q^\alpha\} + \sum_{r=k}^{\infty} \frac{\Gamma(\alpha+r)}{2r!\Gamma(\alpha)}\{(1-q)^r q^\alpha - (1-p)^r p^\alpha\} \\
&= \frac{1}{2}\left(\mathbb{P}(X \leq k-1) - \mathbb{P}(Y \leq k-1) + \mathbb{P}(Y > k-1) - \mathbb{P}(X > k-1)\right) \\
&= \sum_{r=0}^{k-1} \frac{\Gamma(\alpha+r)}{r!\Gamma(\alpha)}\{g_r(p) - g_r(q)\},
\end{aligned}$$

where $g_r(x) = (1-x)^r x^\alpha$. Now $g_r'(x) = \left(\frac{-r}{1-x} + \frac{\alpha}{x}\right)(1-x)^r x^\alpha$ and so applying the Mean Value Theorem to the sum gives that, for some $0 < \theta < 1$,

$$\begin{aligned}
d_{TV}(X,Y) &\leq (p-q)\sum_{r=0}^{k-1} \frac{\Gamma(\alpha+r)}{r!\Gamma(\alpha)}\left(\frac{-r}{1-x} + \frac{\alpha}{x}\right)(1-x)^r x^\alpha \\
&\leq \alpha\left(\frac{p}{q} - 1\right)\sum_{r=0}^{k-1} \frac{\Gamma(\alpha+r)}{r!\Gamma(\alpha)}(1-x)^r x^\alpha
\end{aligned}$$

where $x = q + \theta(p-q)$, and so $0 < x < 1$. Summing the probabilities gives

$$d_{TV}(X,Y) \leq \alpha\left(\frac{p}{q} - 1\right),$$

as required. $\square$

Recall the generalisation of Yule's model for evolution, which can be found in Appendix B. We have already seen in Theorem 3.16 that at the sequence of stopping times, $\{S_1 + S_2 + ... + S_n\}_n$, corresponding to the random time points when new vertices enter the graph, the degree distribution has scale-free distribution μ_ρ^α, where $\mu_\rho^\alpha(k) = \frac{B(k+\alpha;\rho+1)}{B(\alpha;\rho)}$ for $k \in \mathbb{Z}_0^+$. We now bound the distance between the μ_ρ^α-distribution and the degree distribution at a time $S_1 + S_2 + ... + S_n + \theta S_{n+1}$, where $\theta \in (0,1)$. This tells us how close the degree distribution of the graph is to its asymptotic distribution when the graph has reached a size of at least $n+1$ vertices. At time 0 the graph contains one vertex, so at time $S_1 + ... + S_n + \theta S_{n+1}$ there are $n+1$ vertices.

As usual, to study this distribution we define, $I_{n+1} \sim Uniform\{1,2,3,...,n+1\}$ to be the random variable representing the vertex whose degree we observe at time $S_1 + ... + S_n + \theta S_{n+1}$. We denote by W_n^θ the random variable representing the degree of I_{n+1} at time $S_1 + ... + S_n + \theta S_{n+1}$ and we note that it has the mixture distribution, $W_n^\theta \sim Negbinom(\alpha, e^{-(S_{I_{n+1}}+...+S_{n+1})+(1-\theta)S_{n+1}})$ by Theorem 3.11. By Theorem 3.14, since the $S_i's$ and I_{n+1} are all independent of one another, equivalently, $W_n^\theta \sim Negbinom(\alpha; e^{-U+(1-\theta)S_{n+1}})$ for $U \sim Exponential(\rho)$. Note that U and S_{n+1} are dependent random variables, since $U = S_{I_{n+1}} + ... + S_{n+1}$.

Theorem 3.21. *Let $\theta \in (0,1)$, $\rho > 1$, $\alpha \geq 1$ and let S_i and W_n^θ be defined as above. Let $\mathcal{L}(V) = \mu_\rho^\alpha$, the measure defined above. Then, for $\alpha \in \mathcal{A}$ as defined in (3.3),*

$$d_{TV}(W_n^\theta, V) \leq \left(\frac{K(\alpha)}{(n+1)}\right)\left(\frac{n}{\rho(n+1)+1} + \frac{\rho}{\rho-1}\right),$$

for $K(\alpha)$ the constant determined by α as defined in (3.3).

Proof. Without loss of generality define I_n recursively by

$$I_1 := 1,$$

$$\mathbb{P}(I_{n+1} = I_n | I_n) = \frac{n}{n+1},$$

$$\mathbb{P}(I_{n+1} = n+1|I_n) = \frac{1}{n+1}.$$

By Corollary 3.15, a random variable $V' \sim Negbinom(\alpha, e^{-(S_{I_n}+...+S_n)})$ has measure μ_ρ^α, so $V' =^D V$. Now, W_n^θ and V' are both mixtures of distributions that take negative binomial distributions conditional on the outcomes of the random variables in the set $\{S_1, ..., S_{n+1}, I_1, ..., I_{n+1}\}$.

$$V' \sim Negbinom(\alpha, e^{-(S_{I_n}+...+S_n)});$$

$$W_n^\theta \sim Negbinom(\alpha, e^{-(S_{I_{n+1}}+...+S_{n+1})+(1-\theta)S_{n+1}}).$$

Thus it follows from Theorem 3.18 that

$$\begin{aligned} d_{TV}(V, W_n^\theta) &= d_{TV}(V', W_n^\theta) \\ &\leq K(\alpha)\mathbb{E}|e^{(S_{I_n}+...+S_n)-(S_{I_{n+1}}+...+S_{n+1})+(1-\theta)S_{n+1}} - 1|. \end{aligned}$$

Conditioning on the outcome of I_{n+1}, we get

$$\begin{aligned} d_{TV}(V, W_n^\theta) &\leq K(\alpha)\mathbb{E}|e^{(S_{I_n}+...+S_n)-(S_{I_{n+1}}+...+S_{n+1})+(1-\theta)S_{n+1}} - 1| \\ &= K(\alpha)\mathbb{E}\left(|e^{(S_{I_n}+...+S_n)-\theta S_{n+1}} - 1|\Big|I_{n+1} = n+1\right)\mathbb{P}(I_{n+1} = n+1) \\ &\quad +K(\alpha)\mathbb{E}\left(|e^{-\theta S_{n+1}} - 1|\Big|I_{n+1} = I_n\right)\mathbb{P}(I_{n+1} = I_n) \\ &= \frac{K(\alpha)}{(n+1)}\mathbb{E}(e^{(S_{I_n}+...+S_n)}|e^{-\theta S_{n+1}} - e^{-(S_{I_n}+...+S_n)}|\Big|I_{n+1} = n+1) \\ &\quad +\frac{K(\alpha)n}{(n+1)}\mathbb{E}\left(|e^{-\theta S_{n+1}} - 1|\Big|I_{n+1} = I_n\right) \\ &\leq \frac{K(\alpha)}{(n+1)}\mathbb{E}(e^{(S_{I_n}+...+S_n)}|I_{n+1} = n+1) + \frac{K(\alpha)n}{(n+1)}\mathbb{E}(1 - e^{-\theta S_{n+1}}|I_{n+1} = I_n). \end{aligned}$$

Now $I_{n+1} \perp (S_1, ..., S_{n+1})$ and for all $k = 1, 2, ..., n$, $\mathbb{P}(I_n = k | I_{n+1} = n+1) = \mathbb{P}(I_n = k)$. Therefore, recalling from Theorem 3.14 that $S_{I_n} + ... + S_n \sim Exponential(\rho)$,

$$\begin{aligned}
d_{TV}(V, W_n^\theta) &\leq \frac{K(\alpha)}{n+1}\mathbb{E}(e^{(S_{I_n}+...+S_n)}) + \frac{K(\alpha)n}{n+1}\mathbb{E}(1 - e^{-\theta S_{n+1}}) \\
&= \frac{K(\alpha)}{(n+1)}\int_0^\infty e^u \rho e^{-\rho u} du \\
&\quad + \frac{K(\alpha)n}{(n+1)}\left(1 - \int_0^\infty e^{-\theta t}\rho(n+1)e^{-\rho(n+1)t}dt\right) \\
&= \frac{K(\alpha)}{(n+1)} \times \frac{\rho}{\rho - 1} + \frac{K(\alpha)n}{(n+1)}\left(1 - \frac{\rho(n+1)}{\rho(n+1)+\theta}\right) \\
&\leq \left(\frac{K(\alpha)}{(n+1)}\right)\left(\frac{n}{\rho(n+1)+1} + \frac{\rho}{\rho-1}\right),
\end{aligned}$$

as required. □

The method we have used above is illustrative of the types of methods that will be employed for Model Y in Chapter 7. However it is possible to get a similar result for this particular model with less work, as we now show in the next theorem.

Theorem 3.22. *Let $\theta \in (0, 1)$, $\rho > 1$, $\alpha \geq 1$ and let S_i and W_n^θ be defined as above. Let $\mathcal{L}(V) = \mu_\rho^\alpha$, the measure defined above. Then,*

$$d_{TV}(W_n^\theta, V) \leq \frac{1}{n+1} + \frac{\alpha}{(n+1)(\rho-1)}.$$

Proof. Without loss of generality define I_n recursively by

$$I_1 := 1,$$

$$\mathbb{P}(I_{n+1} = I_n | I_n) = \frac{n}{n+1},$$

$$\mathbb{P}(I_{n+1} = n+1 | I_n) = \frac{1}{n+1}.$$

Now,

$$\begin{aligned}
d_{TV}(V, W_n^\theta) &\le \mathbb{P}(W_n^\theta \neq W_{n+1}) \\
&\le \mathbb{P}(I_{n+1} \neq I_n) + \mathbb{P}(W_n^\theta \neq W_{n+1} | I_{n+1} = I_n) \\
&\le \frac{1}{n+1} + \mathbb{P}(W_{n+1} \neq W_n | I_{n+1} = I_n).
\end{aligned}$$

The probability $\mathbb{P}(W_{n+1} \neq W_n | I_{n+1} = I_n)$ is determined by the number of vertices that change in degree in the interval $(S_1 + ... + S_n, S_1 + ... + S_{n+1}]$. An upper bound for this is given by the number of edges that are added to the graph in this interval. Writing $\mathcal{E}_n$ for the set of edges in the graph after time $S_1 + ... + S_n$, we get,

$$\begin{aligned}
d_{TV}(V, W_n^\theta) &\le \frac{1}{n+1} + \sum_{r=1}^{\infty} \frac{r}{n+1} \mathbb{P}(\#\mathcal{E}_{n+1} - \#\mathcal{E}_n = r) \\
&= \frac{1}{n+1} + \frac{1}{n+1} \mathbb{E}(\#\mathcal{E}_{n+1} - \#\mathcal{E}_n) \\
&= \frac{1}{n+1} + \frac{\alpha}{(n+1)(\rho - 1)},
\end{aligned}$$

by Proposition 3.9, as required. $\square$

Chapter 4

Stein's method

4.1 Introduction

This chapter introduces the theory of Stein's method. Stein's method can provide explicit bounds under given metrics for the distance between probability distributions. Therefore it is highly relevant to our framework.

Through Stein's method we can quantify how well the degree distributions of Barabási-Albert random graphs of particular finite sizes approximate a power law distribution. In contrast, most* previous authors have given asymptotic degree distribution results as the number of vertices in the graph tends to infinity. The real networks that random graphs model have a finite size, so it is important to know how well finite-sized models approximate the required degree distributions.

In Chapter 5 we derive a new version of Stein's method for the μ_ρ^α-distribution. We apply this version to a Barabási-Albert random graph model in Chapter 11.

Stein's method has developed from work by Stein, [50], [51], where he used a new method to obtain bounds for the accuracy of the central limit theorem in the presence of dependence. Following this, versions of Stein's method have been developed for various distributions including the Poisson distribution, [15], the geometric distribution, [38], [39], the exponential distribution, [55], as well as general methods for discrete probability measures, [23], [28]. A very brief introduction to Stein's method can be found in [45], while collections of

*The exception to this is Szymański [53] who provides explicit bounds for one particular model using a recurrence equation method. In Chapter 8 we apply the recurrence equation method to our scheme of models.

papers on the subject are found in [6] and [7].

This chapter begins with a short discussion of the idea behind Stein's method. We then summarise known Stein equations and Stein factors for those probability distributions that will be required later in the thesis. The ideas behind Stein's method that are explained here will be used in Chapter 5, where we derive a Stein equation and Stein factors for the power law distribution of Chapter 3.

4.1.1 The general idea

Stein's method relies on a suitable characterisation of the target probability distribution in terms of a functional operator. In particular, Stein [51] showed that the operator

$$\mathcal{A}(f)(x) := f'(x) - xf(x) \tag{4.1}$$

characterises the $N(0,1)$ distribution. By this we mean that for the real random variable W to have a standard normal distribution, it is necessary and sufficient that, for all continuous and piecewise continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with $||f'||_\infty < \infty$,

$$\mathbb{E}\{\mathcal{A}(f)(W)\} = 0.$$

The idea behind Stein's method is that if a random variable U is close in distribution to W, then $|\mathbb{E}\{\mathcal{A}(f)(U)\}|$ should be in some sense 'small'. This is made precise by taking suprema of $|\mathbb{E}\{\mathcal{A}(f)(U)\}|$ over certain large classes of functions and proving that these suprema correspond to bounds for the distances between the distributions under particular metrics.

Let us now consider an arbitrary probability distribution with measure μ. Suppose that V is a random variable satisfying $\mathcal{L}(V) = \mu$. Reinert [43] gives us the following sketch for applying Stein's method for μ.

1. "Find a suitable characterisation, namely an operator $\mathcal{A}$ such that $X \sim \mu$ if and only if for all smooth functions f, $\mathbb{E}(\mathcal{A}(f)(X)) = 0$ holds.
2. For each smooth functions h find a solution $f = f_h$ of the *Stein equation*,

$$h(x) - \mu(h) = \mathcal{A}(f)(x).$$

3. Then for any variable W it follows that

$$\mathbb{E}h(W) - \mu(h) = \mathcal{A}(f)(W).\text{"}$$

([43] in [6], p. 184)

The key to the method is the relationship between the solution of the Stein equation and certain metrics. We see from the definitions of probability metrics from page ix that the distance between U and V is given by $\sup_{h \in \mathcal{H}} |\mathbb{E}(h(U)) - \mu(h)|$ where $\mathcal{H}$ determines the metric. For example $\mathcal{H} := \{(-\infty, x] : x \in \mathbb{R}\}$ gives us the Kolmogorov distance.

The final step of Stein's method is to find properties of the solutions $f = f_h$ of the Stein equation for different classes of functions, $\mathcal{H}$. For example, we might show that for a particular $\mathcal{H}$, all solutions f_h satisfy $||f_h|| \leq c$ for a given constant c. This bound is known as a *Stein factor*.

4.1.2 Stein's method for discrete probability distributions

Following the above approach, we describe a quick method for writing down a Stein equation for a discrete† probability distribution.

Lemma 4.1. *Let μ be a probability measure on $\mathbb{Z}_0^+$ and suppose that there exist functions h_1 and h_2 such that, for all $k \in \mathbb{Z}_0^+$,*

$$h_1(k+1)\mu(k+1) = h_2(k+1)\mu(k).$$

†See Stein [52] for a continuous version.

Then the following is a Stein operator for μ,

$$\mathcal{A}(f)(k) = h_1(k)f(k) - h_2(k+1)f(k+1) - \mu(0)h_1(0)f(0),$$

that is to say,

$$X \sim \mu \iff \forall f \quad \mathbb{E}\mathcal{A}(f)(X) = 0.$$

Proof. First suppose that V is a random variable satisfying $\mathcal{L}(V) = \mu$. Then for any function f,

$$\begin{aligned}
\mathbb{E}\{h_1(V)f(V) - \mu(0)h_1(0)f(0)\} &= \sum_{k=1}^{\infty} \mu(k)h_1(k)f(k) \\
&= \sum_{k=0}^{\infty} \mu(k+1)h_1(k+1)f(k+1) \\
&= \sum_{k=0}^{\infty} \mu(k)h_2(k+1)f(k+1) \\
&\quad \text{(by definition of } h_1,\ h_2) \\
&= \mathbb{E}\{h_2(V+1)f(V+1)\},
\end{aligned}$$

so $\mathbb{E}\{\mathcal{A}(f)(V)\} = 0$ as required.

Now suppose that $\mathbb{E}\{\mathcal{A}(f)(V)\} = 0$ for all functions f and that $\mathcal{L}(V) = \nu$. Then in particular this is true for $f = \mathbf{1}_{\{k+1\}}$ for each $k \in \mathbb{Z}_0^+$. It follows that for $k \in \mathbb{Z}_0^+$,

$$h_1(k+1)\nu(k+1) = h_2(k+1)\nu(k).$$

Since ν is a probability distribution, we also know that $\sum_{k=0}^{\infty} \nu(k) = 1$, thus uniquely defining $\nu = \mu$, as required. $\square$

This simple method for obtaining Stein equations for discrete random variables is applied in Chapter 5 to obtain a Stein equation for the power law distribution derived in Chapter 3. Of course there are many possible choices for h_1 and h_2, showing that the Stein equation is far from unique, but the operators should be chosen in such a way that they give rise to suitable bounds.

Some papers that look in detail at Stein's method for discrete probability distributions are [12], [23] and [28]. All three papers obtain Stein equations for a discrete measure μ by looking at a birth-death chain where μ is the stationary distribution. When the chain has birth rates b_k and death rates d_k the following Stein equation is used

$$\mathcal{B}(g)(k) = b_k \Delta g(k) - d_k \Delta g(k-1).$$

Since, by detailed balance equations,

$$d_{k+1}\mu(k+1) = b_k \mu(k),$$

we see that one Stein equation resulting from the method described above, taking $h_1(k+1) = d_{k+1}$, $h_2(k+1) = b_k$, would be

$$\mathcal{A}(f)(k) = d_k f(k) - b_k f(k+1) - \mu(0) d_0 f(0).$$

We can obtain $\mathcal{B}$ by applying $\mathcal{A}$ to $f(k) := -\Delta g(k-1)$ and setting $\Delta g(-1) = 0$. Thus the birth-death process method is equivalent to our approach described above. For both approaches the choice of death rate (or function $h_2(k+1)$) is free, but a unit per capita death rate is commonly chosen. Eichelsbacher and Reinert [23] give general results for solutions to the Stein equation for discrete random variables. In the case of the power law described in Chapter 3, however, we do not obtain sufficiently strong Stein factors‡ for convergence using their results. By studying the solutions to a particular Stein equation we are able to find stronger Stein factors in Chapter 5.

‡By '*Stein factors*' we mean bounds for the norms of solutions of the Stein equation, or functions of those solutions. These are sometimes known as '*magic factors*' or '*Stein magic factors*'. Strong Stein factors are essential for obtaining good bounds from a Stein equation.

4.2 Stein's method for the geometric distribution

Having discussed Stein's method in general, we now give known results for the geometric distribution that will be used in Chapter 12. The Stein operator below is equivalent to those developed in [38] and [39], since each can be obtained from the other through the application of an invertible operator (Stein factors must then be adapted for the new class of relevant functions). We state Stein's method in the form given in [23] and we use their Stein factors.

Lemma 4.2. *Let $\mu(k) = p(1-p)^k$ for some $p \in [0,1]$ and for $k = 0, 1, 2, \ldots$. Then for any bounded function f, with $|f(k)| \leq 1, \forall k \in \mathbb{N}$ there is a function g satisfying $g(0) = 0$, $|g(j+1)| \leq \frac{1}{p(j+1)}$ for $j = 0, 1, 2, \ldots$, and $|\Delta g(j)| \leq \min\left(\frac{1}{j}, \frac{1+p}{1+j}\right)$ for $j = 0, 1, 2, \ldots$, such that,*

$$f(k) - \mu(f) = (1-p)(k+1)g(k+1) - kg(k).$$

Moreover, for any random variable V with law μ and any function g,

$$\mathbb{E}\{(1-p)(V+1)g(V+1) - Vg(V)\} = 0.$$

The following corollary illustrates how equivalent versions of the Stein equation may be obtained by the application of an invertible operator to the Stein operator. This Stein equation and the corresponding Stein factors are simpler to work with in Chapter 12.

Corollary 4.3. *Let $\mu(k) = p(1-p)^k$ for some $p \in [0,1]$ and for $k = 0, 1, 2, \ldots$. Then for any bounded function f,with $|f(k)| \leq 1, \forall k \in \mathbb{N}$, there is a function h satisfying $h(1) = 0$, $|h(j+1)| \leq \frac{1}{p}$ and $|\Delta h(j)| \leq \frac{1}{(j+1)p} + j \min\left(\frac{1}{j}, \frac{1+p}{j+1}\right)$, for $j = 0, 1, 2, \ldots$ that satisfies,*

$$f(k) - \mu(f) = (1-p)h(k+1) - h(k).$$

Moreover, for any random variable V with law μ and any function h,

$$\mathbb{E}\{(1-p)h(V+1) - h(V)\} = 0.$$

Proof. Let $\mu(k) = p(1-p)^{k-1}$ for some $p \in [0,1]$ and for $k = 0, 1, 2, \ldots$. Let f be a bounded function satisfying $|f(k)| \leq 1, \forall k \in \mathbb{N}$. By Lemma 4.2 we can find an associated function g satisfying $g(0) = 0$, $|g(j+1)| \leq \frac{1}{p(j+1)}$ for $j = 0, 1, 2, \ldots$, and $|\Delta g(j)| \leq \min\left(\frac{1}{j}, \frac{1+p}{1+j}\right)$ for $j = 0, 1, 2, \ldots$, such that,

$$f(k) - \mu(f) = (1-p)(k+1)g(k+1) - kg(k).$$

First note that the Stein equation in Lemma 4.2 can be written,

$$f(k) - \mu(k) = \mathcal{A}(g)(k),$$

where $\mathcal{A}$ is the operator defined by,

$$\mathcal{A}(g)(k) := (1-p)(k+1)g(k+1) - kg(k).$$

Define $h(k) := kg(k)$. Then

$$\begin{aligned} f(k) - \mu(k) &= \mathcal{A}(g)(k) \\ &= (1-p)(k+1)g(k+1) - kg(k) \\ &= (1-p)h(k+1) - h(k). \end{aligned}$$

Alternatively we can think of this as $\mathcal{A}(g)(k) = (\mathcal{A} \circ \mathcal{B})(h)(k)$ where $\mathcal{B}$ is the invertible operator

$$\mathcal{B}(g)(k) := \begin{cases} 0 \text{ if } k = 0; \\ \frac{g(k)}{k} \text{ otherwise.} \end{cases}$$

We have proved the existence of a function h satisfying the required equation for $k \neq 0$. Moreover by definition and our conditions on g,

$$h(0) = g(0) = 0$$

and for $j \geq 0$,

$$\begin{aligned} |h(j+1)| &= j|g(j+1)| \\ &\leq \frac{1}{p}, \end{aligned}$$

and

$$\begin{aligned} |\Delta h(j)| &= |(j+1)g(j+1) - jg(j)| \\ &= |g(j+1) + j\Delta g(j)| \\ &\leq \frac{1}{(j+1)p} + j \min\left(\frac{1}{j}, \frac{1+p}{j+1}\right). \end{aligned}$$

Finally, suppose h is any function and let $\mathcal{L}(V) = \mu$. Then, with the operators defined above,

$$\begin{aligned} &\mathbb{E}\{(1-p)h(V+1) - h(V)\} \\ &= \mathbb{E}\{(1-p)(V+1)\mathcal{B}(h)(V+1) - V\mathcal{B}(h)(V)\} \\ &= 0, \end{aligned}$$

by Lemma 4.2 applied to the function $\mathcal{B}(h)$. □

4.3 Stein's method for the normal distribution

We now outline some results for Stein's method for the standard normal distribution. As well as being the original probability distribution studied by Stein, we shall use these results in Chapter 12 to obtain bounds from $N(0,1)$ for the distribution of the number of triangles in an evolving Bernoulli random graph.

As explained above, Stein gave the following operator for the standard normal distribution,

$$\mathcal{A}(f)(x) := f'(x) - xf(x).$$

The following theorem, proved in [16], gives Stein factors for the Wasserstein distance for this Stein equation.

Theorem 4.4. *Let $Z \sim N(0,1)$. For every bounded absolutely continuous*[§] *function h there is a function $f_h : \mathbb{R} \to \mathbb{R}$, $f_h(x) := e^{\frac{x^2}{2}} \int_{-\infty}^{x} (h(t) - \mathbb{E}h(Z)) e^{\frac{-t^2}{2}} dt$, satisfying,*

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = \mathbb{E}(f_h'(W) - Wf_h(W)). \tag{4.2}$$

If h is absolutely continuous and f_h is any function $f_h : \mathbb{R} \to \mathbb{R}$ satisfying Equation (4.2) then the following conditions hold,

$$\begin{aligned} ||f_h||_\infty &\leq \min\left(\sqrt{\frac{\pi}{2}}||h(.) - \mathbb{E}h(Z)||_\infty, 2||h'||_\infty\right) & (4.3)\\ ||f_h'||_\infty &\leq \min(2||h(.) - \mathbb{E}h(Z)||_\infty, 4||h'||_\infty) & (4.4)\\ ||f_h''||_\infty &\leq 2||h'||_\infty. & (4.5) \end{aligned}$$

[§]By Rademacher's theorem, h is almost everywhere differentiable.

Chapter 5

Stein's method for the μ_ρ^α-distribution

In Chapter 4 we discussed how Stein's method can be used to bound the distances between two distributions under a given metric. In this chapter we develop a version of Stein's method for the general μ_ρ^α-distribution with Stein factors given both for the Kolmogorov distance and for the total variation distance. We apply this method to a preferential attachment random graph model in Chapter 11.

5.1 A Stein equation for the μ_ρ^α-distribution

This section presents a characterising Stein operator for the general power law distribution given in Chapter 3. We begin by using properties of the μ_ρ^α-distribution to derive an appropriate Stein equation. We then consider properties of the solutions of the Stein equation and use these properties to obtain Stein factors for this Stein equation.

Recall that the μ_ρ^α-distribution with parameters $\alpha > 0, \rho > 1$ has support $\mathbb{Z}_0^+$ and is defined on its support by the measure

$$\mu_\rho^\alpha(k) := \frac{B(k+\alpha;\rho+1)}{B(\alpha,\rho)}, \tag{5.1}$$

with

$$B(x,y) \;\; := \;\; \int_0^1 t^{x-1}(1-t)^{y-1}dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

As a consequence of (A.2) of Appendix A, for $k \geq 0$

$$\frac{\mu_\rho^\alpha(k+1)}{\mu_\rho^\alpha(k)} \;\; = \;\; \frac{B(k+\alpha+1;\rho+1)}{B(k+\alpha;\rho+1)} \tag{5.2}$$

$$\;\; = \;\; \frac{k+\alpha}{k+\alpha+\rho+1}. \tag{5.3}$$

By the results of Section 4.1.2, Chapter 4, the ratio in Equation (5.3) suggests an appropriate Stein equation to characterise the μ_ρ^α-distribution,

$$f(k) - \mu_\rho^\alpha(f) = (k+\alpha+\rho)g(k) - (k+\alpha)g(k+1) - (\alpha+\rho)g(0)\mu_\rho^\alpha(0). \tag{5.4}$$

If $g(0) = 0$, the solutions, g, of (5.4) for f satisfy

$$g(k+1) \;\; = \;\; \frac{k+\alpha+\rho}{k+\alpha}g(k) - \frac{f(k)-\mu_\rho^\alpha(f)}{k+\alpha}, \text{ for } k \geq 0, \tag{5.5}$$

and (5.5) gives an iterative construction for a solution g with $g(0) = 0$. Since every solution g with $g(0) = 0$ must satisfy (5.5), we see that the solution is unique.

Lemma 5.1. *If a random variable V has law μ_ρ^α then for all functions g such that $g(0) = 0$,*

$$\mathbb{E}((V+\alpha+\rho)g(V)) = \mathbb{E}((V+\alpha)g(V+1)).$$

Proof. Suppose that V is a random variable with $\mathcal{L}(V) = \mu_\rho^\alpha$. Then

$$\begin{aligned}
\mathbb{E}((V+\alpha)g(V+1)) &= \sum_{k=0}^{\infty} \mu_\rho^\alpha(k)(k+\alpha)g(k+1) \\
&= \sum_{k=0}^{\infty} (k+\alpha+\rho+1)\mu_\rho^\alpha(k+1)g(k+1) \text{ (by 5.3)} \\
&= \sum_{k=1}^{\infty} (k+\alpha+\rho)\mu_\rho^\alpha(k)g(k) \\
&= \sum_{k=0}^{\infty} \mu_\rho^\alpha(k)(k+\alpha+\rho)g(k) \text{ (since } g(0)=0) \\
&= \mathbb{E}((V+\alpha+\rho)g(V)).
\end{aligned}$$

□

Proposition 5.2. *Let f be a bounded function. Let g_f be the function defined iteratively as in* (5.5)*, with $g_f(0) = 0$. Then for $k \in \mathbb{Z}_0^+$,*

$$g_f(k+1) = -\sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(k+\alpha;\rho+1)} \left(\frac{f(r) - \mu_\rho^\alpha(f)}{k+\alpha} \right).$$

Proof. We proceed by induction. First consider the case $k = 0$. Then, by (5.5),

$$\begin{aligned}
g_f(0+1) &= \frac{0+\alpha+\rho}{0+\alpha} g(0) - \frac{f(0) - \mu_\rho^\alpha(f)}{0+\alpha} \\
&= -\frac{f(0) - \mu_\rho^\alpha(f)}{0+\alpha} \\
&= -\sum_{r=0}^{0} \frac{B(r+\alpha;\rho+1)}{B(0+\alpha;\rho+1)} \left(\frac{f(r) - \mu_\rho^\alpha(f)}{0+\alpha} \right),
\end{aligned}$$

as required.

Now suppose that the proposition holds for $k = n$, so that

$$g_f(n+1) = -\sum_{r=0}^{n} \frac{B(r+\alpha;\rho+1)}{B(n+\alpha;\rho+1)} \left(\frac{f(r) - \mu_\rho^\alpha(f)}{n+\alpha} \right).$$

Then, by (5.5),

$$
\begin{aligned}
g(n+2) &= \frac{n+\alpha+\rho+1}{n+\alpha+1} g(n+1) - \frac{f(n+1)-\mu_\rho^\alpha(f)}{n+\alpha+1} \\
&= -\frac{n+\alpha+\rho+1}{n+\alpha+1} \sum_{r=0}^{n} \frac{B(r+\alpha;\rho+1)}{B(n+\alpha;\rho+1)} \left(\frac{f(r)-\mu_\rho^\alpha(f)}{n+\alpha} \right) \\
&\quad -\frac{f(n+1)-\mu_\rho^\alpha(f)}{n+\alpha+1} \\
&\quad \text{(by inductive hypothesis)} \\
&= -\frac{1}{n+\alpha+1} \sum_{r=0}^{n} \frac{B(r+\alpha;\rho+1)}{B(n+\alpha+1;\rho+1)} \left(f(r)-\mu_\rho^\alpha(f) \right) \\
&\quad -\frac{B(n+\alpha+1;\rho+1)}{B(n+\alpha+1;\rho+1)} \left(\frac{f(n+1)-\mu_\rho^\alpha(f)}{n+\alpha+1} \right) \\
&\quad \text{(by Appendix A, (A.2))} \\
&= -\sum_{r=0}^{n+1} \frac{B(r+\alpha;\rho+1)}{B(n+\alpha+1;\rho+1)} \left(\frac{f(r)-\mu_\rho^\alpha(f)}{n+\alpha+1} \right),
\end{aligned}
$$

as required. □

Lemma 5.3. *Let V be a discrete random variable with support $\mathbb{Z}_0^+$. If*

$$\mathbb{E}((V+\alpha+\rho)g(V)) = \mathbb{E}((V+\alpha)g(V+1)) \tag{5.6}$$

holds for all functions g satisfying $g(0)=0$, then V takes the μ_ρ^α-distribution.

Proof. Suppose that V is a discrete random variable with measure ν, satisfying the condition that (5.6) holds whenever $g(0)=0$. Consider the sequence of indicator functions $(\mathbf{1}_{\{k+1\}})_{k\geq 0}$. The functions satisfy the condition $\mathbf{1}_{\{k+1\}}(0)=0$ for $k\in\mathbb{Z}_0^+$, so, putting $g=\mathbf{1}_{\{k+1\}}$ in (5.6),

$$\mathbb{E}((V+\alpha+\rho)\mathbf{1}_{\{k+1\}}(V)) = \mathbb{E}((V+\alpha)\mathbf{1}_{\{k+1\}}(V+1)) \quad \forall k\geq 0.$$

Equivalently,

$$(k+\alpha+\rho+1)\nu(k+1) = (k+\alpha)\nu(k) \quad \forall k \geq 0.$$

There exists at most one probability measure ν satisfying this recursion as well as the condition $\sum_{i=0}^{\infty} \nu(i) = 1$. Thus, by (5.3) we see that this uniquely defines ν to be precisely the measure μ_ρ^α defined in (5.1). $\square$

5.2 Stein factors

Lemma 5.1 and Lemma 5.3 showed that Equation (5.4) does indeed characterise the μ_ρ^α-distribution. We now consider the solutions given by (5.5) for the members of two particular classes of functions. First we consider the class of half-lines. This allows us to obtain Stein factors for the Kolmogorov distance, defined on page ix by,

$$\begin{aligned} d_K(X,Y) &:= \sup_{x\in\mathbb{R}} |\mathbb{P}(X \leq x) - \mathbb{P}(Y \leq x)| \\ &= \sup_{x\in\mathbb{R}} |\mathbb{E}(\mathbf{1}_{(-\infty,x]}(X) - \mathbf{1}_{(-\infty,x]}(Y))|. \end{aligned}$$

We then look at the class of functions $\{\mathbf{1}_A : A \subset \mathbb{Z}_0^+\}$, from which we obtain Stein factors for the total variation distance. Using the definition of page ix, we get

$$\begin{aligned} d_{TV}(X,Y) &:= \sup_{A\subset\mathbb{Z}_0^+} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \\ &= \sup_{A\subset\mathbb{Z}_0^+} |\mathbb{E}(\mathbf{1}_A(X) - \mathbf{1}_A(Y))|. \end{aligned}$$

It is an immediate consequence of their definitions that total variation distance is a stronger metric than the Kolmogorov distance.

We have already seen in Proposition 5.2 that solutions, g, of the Stein equation exist and are uniquely defined given $g(0) = 0$. Properties of the solutions provide bounds for the Stein factors for different metrics. We now consider the class of half-lines, which are an appropriate class of test functions for the Kolmogorov distance.

5.2.1 Stein factors for Kolmogorov distance approximation

Lemma 5.4. *For $n \geq 0$, let f_n be the half-line $f_n := \mathbf{1}_{(-\infty,n]}$, and let g_n be defined as in Proposition 5.2. Then, for $n \in \mathbb{Z}_0^+$,*

$$\begin{aligned}
(i) &\quad g_n(k) < 0 \quad \forall k \geq 1;\\
(ii) &\quad g_n(k+1) \leq g_n(k) \quad \forall k \geq 0;\\
(iii) &\quad |g_n(k)| \leq \tfrac{1}{\rho}, \quad \forall k \geq 0;\\
(iv) &\quad |\Delta g_n(k)| \leq \tfrac{1}{k+\alpha+\rho} \quad \forall k \geq 2,
\end{aligned}$$

where $\Delta g(k)$ represents the forward difference, $\Delta g(k) = g(k+1) - g(k)$.

Proof. We know from (3.1) in Chapter 3, Proposition 3.4, that $\mu_\rho^\alpha(f_n) = 1 - \frac{B(n+\alpha+1;\rho)}{B(\alpha;\rho)}$.

Fix $n \in \mathbb{Z}_0^+$ and suppose $k \leq n$. Since $f_n(r) = 1 \quad \forall r \leq k$,

$$g_n(k+1) = -\sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(k+\alpha;\rho+1)} \left(\frac{1-\mu_\rho^\alpha(f_n)}{k+\alpha}\right) \leq 0.$$

So $g_n(k+1) \leq 0$ for $1 \leq k \leq n$. Furthermore, by (5.5), for $1 \leq k \leq n$,

$$\begin{aligned}
g_n(k+1) - g_n(k) &= \frac{\rho}{k+\alpha} g_n(k) - \frac{1 - (1 - \frac{B(n+\alpha+1;\rho)}{B(\alpha;\rho)})}{k+\alpha}\\
&= \frac{\rho}{k+\alpha} g_n(k) - \frac{B(n+\alpha+1;\rho)}{(k+\alpha)B(\alpha;\rho)}.
\end{aligned}$$

Thus, since $g_n(0) = 0$, $g_n(k)$ is a monotonically decreasing function between $k = 0$ and $k = n + 1$.

We now calculate $g_n(n + 1)$:

$$\begin{aligned}
g_n(n+1) &= -\sum_{r=0}^{n} \frac{B(r+\alpha;\rho+1)}{B(n+\alpha;\rho+1)} \left(\frac{1-\mu_\rho^\alpha(f_n)}{n+\alpha}\right) \\
&= -\sum_{r=0}^{n} \frac{B(r+\alpha;\rho+1)B(n+\alpha+1;\rho)}{(n+\alpha)B(n+\alpha;\rho+1)B(\alpha;\rho)} \\
&\quad \text{(by (3.1) of Proposition 3.4, Chapter 3)} \\
&= -\sum_{r=0}^{n} \frac{B(r+\alpha;\rho+1)}{\rho B(\alpha;\rho)} \\
&\quad \text{(by Appendix A, (A.4))} \\
&= -\frac{\mu_\rho^\alpha(f_n)}{\rho} \geq -\frac{1}{\rho}.
\end{aligned}$$

Thus, by (5.5), as $f_n(n + 1) = 0$,

$$g_n(n+2) - g_n(n+1) = \frac{1}{n+\alpha+1}(-\mu_\rho^\alpha(f_n) + \mu_\rho^\alpha(f_n)) = 0,$$

so,

$$g_n(n+2) = g_n(n+1) = -\frac{\mu_\rho^\alpha(f_n)}{\rho}.$$

We see that for $k \geq n + 1$, $f_n(k) = 0$ and,

$$\begin{aligned}
g_n(k+1) - g_n(k) &= \frac{1}{k+\alpha}\{\rho g_n(k) - (0 - \mu_\rho^\alpha(f_n))\} \\
&= \frac{1}{k+\alpha}(\rho g_n(k) + \mu_\rho^\alpha(f_n)),
\end{aligned}$$

so $g_n(k)$ is constant for $k \geq n + 1$. Since $g_n(k) \leq 0$, $|g_n(k)| \leq \frac{1}{\rho}$ for $k \leq n + 1$ and any $n \in \mathbb{Z}_0^+$. Thus $|g_n(k)| \leq \frac{1}{\rho}$ for all k, as required.

We now obtain the required bound on the forward difference, $|\Delta g_n(k)|$. It clearly follows from the previous argument that for $k \geq n+1$, $|\Delta g_n(k)| = 0$, since g_n has been shown to be constant for such k. Suppose $1 \leq k \leq n$. By Proposition 5.2 and by (3.1) of Chapter 3, Proposition 3.4,

$$\begin{aligned}
\Delta g_n(k) &= -\sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(k+\alpha;\rho+1)} \left(\frac{B(n+\alpha+1;\rho)}{(k+\alpha)B(\alpha;\rho)} \right) \\
&\quad + \sum_{r=0}^{k-1} \frac{B(r+\alpha;\rho+1)}{B(k+\alpha-1;\rho+1)} \left(\frac{B(n+\alpha+1;\rho)}{(k+\alpha-1)B(\alpha;\rho)} \right) \\
&= -\frac{B(n+\alpha+1;\rho)}{(k+\alpha)B(k+\alpha;\rho+1)} \sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(\alpha;\rho)} \\
&\quad + \frac{B(n+\alpha+1;\rho)}{(k+\alpha-1)B(k+\alpha-1;\rho+1)} \sum_{r=0}^{k-1} \frac{B(r+\alpha;\rho+1)}{B(\alpha;\rho)}.
\end{aligned}$$

Applying (5.1) with (3.1) again gives,

$$\begin{aligned}
\Delta g_n(k) &= -\frac{B(n+\alpha+1;\rho)}{(k+\alpha)B(k+\alpha;\rho+1)} \left(1 - \frac{B(k+\alpha+1;\rho)}{B(\alpha;\rho)} \right) \\
&\quad + \frac{B(n+\alpha+1;\rho)}{(k+\alpha-1)B(k+\alpha-1;\rho+1)} \left(1 - \frac{B(k+\alpha;\rho)}{B(\alpha;\rho)} \right) \\
&= \frac{B(n+\alpha+1;\rho)}{B(k+\alpha;\rho+1)} \left(\frac{1}{k+\alpha+\rho} - \frac{1}{k+\alpha} \right) + \frac{B(n+\alpha+1;\rho)}{\rho B(\alpha;\rho)} - \frac{B(n+\alpha+1;\rho)}{\rho B(\alpha;\rho)} \\
&\quad \text{(by A.4)} \\
&= \frac{-\rho B(n+\alpha+1;\rho)}{(k+\alpha)(k+\alpha+\rho)B(k+\alpha;\rho+1)} \\
&= \frac{-B(n+\alpha+1;\rho)}{(k+\alpha+\rho)B(k+\alpha+1;\rho)} \quad \text{(by A.4)} \\
&= \frac{-1}{k+\alpha+\rho} \left(\frac{1-\mu_\rho^\alpha(f_n)}{1-\mu_\rho^\alpha(f_k)} \right) \quad \text{(by (3.1))} \\
&\in \left(\frac{-1}{k+\alpha+\rho}, 0 \right),
\end{aligned}$$

since $k \leq n$, as required. $\square$

Theorem 5.5. *Let V, W be discrete random variables taking values in the non-negative* *integers, with V satisfying $\mathcal{L}(V) = \mu_\rho^\alpha$, $\mu_\rho^\alpha(k) = \frac{B(k+\alpha;\rho+1)}{B(\alpha;\rho)}$. Then the Kolmogorov distance between V and W is bounded by,*

$$d_K(V,W) \leq \sup_{g\in\mathcal{H}_\rho^\alpha} |\mathbb{E}\left((W+\alpha+\rho)g(W) - (W+\alpha)g(W+1)\right)|,$$

where

$$\mathcal{H}_\rho^\alpha = \left\{ g : g(0) = 0, \quad \forall k \geq 0 \quad g(k) \leq 0 \quad |g(k)| \leq \frac{1}{\rho}, \quad |\Delta g(k)| \leq \frac{1}{k+\alpha+\rho} \right\}. \tag{5.7}$$

Proof. Let $\mathcal{J} := \{\mathbf{1}_{(-\infty,n]} : n \in \mathbb{Z}_0^+\}$. For each element $\mathbf{1}_{(-\infty,n]} \in \mathcal{J}$, by Proposition 5.2 we can define a solution of the Stein equation of the form in Proposition 5.2. Lemma 5.4 tells us that $g_n \in \mathcal{H}_\rho^\alpha$ as defined in (5.7). so $\mathcal{H}_\rho^\alpha$ contains all solutions of Stein equations for indicator functions of the form $\mathbf{1}_{(-\infty,n]}$. Thus,

$$\begin{aligned} d_K(V,W) &= \sup_{f\in\mathcal{J}} |\mathbb{E}(f(W) - \mu_\rho^\alpha(f))| \\ &\leq \sup_{g\in\mathcal{H}_\rho^\alpha} |\mathbb{E}\left((W+\alpha+\rho)g(W) - (W+\alpha)g(W+1)\right)|, \end{aligned}$$

as required. □

5.2.2 Stein factors for total variation distance approximation

Ideally we would prefer to use total variation distance approximation, since this is a stronger metric than the Kolmogorov distance. We now obtain Stein factors for total variation distance, defining an arbitrary test function for the total variation distance in such a way that we can use properties obtained in Lemma 5.4.

For a non-empty set $A \subset \mathbb{Z}_0^+$ define $f_A := \mathbf{1}_A$ and note that by Proposition 5.2 we can define g_A to be the unique solution to the Stein equation (5.1) satisfying the conditions,

$$\begin{aligned} g(0) &= 0, \\ g(k+1) &= \frac{k+\alpha+\rho}{k+\alpha} g(k) - \frac{f(k)-\mu_\rho^\alpha(f)}{k+\alpha} \\ &= -\sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(k+\alpha;\rho+1)} \left(\frac{f(r)-\mu_\rho^\alpha(f)}{k+\alpha} \right). \end{aligned} \tag{5.8}$$

Throughout this subsection we write g_s for the solution to the Stein equation (satisfying these conditions) for the function $f_s = \mathbf{1}_{(-\infty,s]}$, $s \in \mathbb{Z}_0^+$. Lemma 5.4 gives properties of g_s.

Lemma 5.6. *Let $A, B \subset \mathbb{Z}_0^+$ such that $A \cap B = \emptyset$. Let f_A, f_B be the associated indicator functions and let g_A, g_B be defined as in (5.8). Then the following equalities hold,*

$$\begin{aligned} \mu_\rho^\alpha(f_{A\cup B}) &= \mu_\rho^\alpha(f_A) + \mu_\rho^\alpha(f_B); \\ g_{A\cup B} &= g_A + g_B. \end{aligned}$$

Proof. The first result is a direct consequence of the linearity of measures.

We prove the second result by induction on k. First note, by definition, that $g_{A\cup B} = 0 = g_A + g_B$.

Now suppose that $g_{A\cup B}(r) = g_A(r) + g_B(r)$ for some $r \in \mathbb{Z}_0^+$. We know that

$$g_{A\cup B}(r+1) = \frac{r+\alpha+\rho}{r+\alpha} g_{A\cup B}(r) - \frac{f_{A\cup B}(r) - \mu_\rho^\alpha(f_{A\cup B})}{r+\alpha}.$$

So

$$
\begin{aligned}
g_{A\cup B}(r+1) &= \frac{r+\alpha+\rho}{r+\alpha}(g_A(r)+g_B(r)) - \frac{f_A(r)+f_B(r)-\mu_\rho^\alpha(f_{A\cup B})}{r+\alpha} \\
&\quad \text{(by inductive hypothesis and definition of } f_A) \\
&= \left(\frac{r+\alpha+\rho}{r+\alpha}(g_A(r)+g_B(r)) - \frac{f_A(r)+f_B(r)-\mu_\rho^\alpha(f_A)-\mu_\rho^\alpha(f_B)}{r+\alpha}\right) \\
&\quad \text{(by the first equality of this lemma)} \\
&= g_a(r+1)+g_B(r+1) \quad \text{(by (5.8))},
\end{aligned}
$$

thus by induction the second equality holds. ∎

Lemma 5.7. *For $s \in \mathbb{Z}_0^+$ let $f_s = \mathbf{1}_{(-\infty,s]}$ and let g_s be the associated function defined as in (5.8). Then*

- *for $s \geq k$,* $\quad g_{s+1}(k+1) - g_s(k+1) > 0$;
- *for $s < k$,* $\quad g_{s+1}(k+1) - g_s(k+1) < 0$.

Proof. To begin note that, by definition,

$$
\begin{aligned}
&g_{s+1}(k+1) - g_s(k+1) \\
&= \ -\sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(k+\alpha;\rho+1)} \left(\frac{f_{s+1}(r)-f_s(r)+\mu_\rho^\alpha(f_s)-\mu_\rho^\alpha(f_{s+1})}{k+\alpha}\right).
\end{aligned}
$$

Suppose $s \geq k$. Then, if $r \leq k$, the appropriate half-lines satisfy $f_{s+1}(r) - f_s(r) = 1 - 1 = 0$.

Thus,

$$
\begin{aligned}
& g_{s+1}(k+1) - g_s(k+1) \\
= & -\sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(k+\alpha;\rho+1)} \left(\frac{\mu_\rho^\alpha(f_s) - \mu_\rho^\alpha(f_{s+1})}{k+\alpha} \right) \\
= & -\sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(k+\alpha;\rho+1)} \left(\frac{(-\mu_\rho^\alpha(\mathbf{1}_{\{s+1\}}))}{k+\alpha} \right) > 0.
\end{aligned} \tag{5.9}
$$

Now suppose $s < k$. Whenever $r \neq s+1$, then the appropriate half-lines satisfy $f_{s+1}(r) - f_s(r) = 0$. If $r = s+1$ then $f_{s+1}(r) - f_s(r) = 1$ Thus,

$$
\begin{aligned}
& g_{s+1}(k+1) - g_s(k+1) \\
= & -\sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(k+\alpha;\rho+1)} \left(\frac{\mu_\rho^\alpha(f_s) - \mu_\rho^\alpha(f_{s+1})}{k+\alpha} \right) - \frac{B(s+\alpha+1;\rho+1)}{(k+\alpha)B(k+\alpha;\rho+1)}.
\end{aligned}
$$

As for (5.9), this becomes

$$
\begin{aligned}
& g_{s+1}(k+1) - g_s(k+1) \\
= & \frac{B(\alpha;\rho)}{B(k+\alpha;\rho+1)} \left(\frac{\mu_\rho^\alpha(\mathbf{1}_{\{s+1\}})}{k+\alpha} \right) \sum_{r=0}^{k} \frac{B(r+\alpha;\rho+1)}{B(\alpha;\rho)} - \frac{B(s+\alpha+1;\rho+1)}{(k+\alpha)B(k+\alpha;\rho+1)} \\
= & \frac{B(\alpha;\rho)}{B(k+\alpha;\rho+1)} \left(\frac{B(s+\alpha+1;\rho+1)}{(k+\alpha)B(\alpha;\rho)} \right) \left(1 - \frac{B(k+\alpha+1;\rho)}{B(\alpha;\rho)} \right) \\
& -\frac{B(s+\alpha+1;\rho+1)}{(k+\alpha)B(k+\alpha;\rho+1)} \quad \text{(by (3.1))} \\
= & -\frac{B(s+\alpha+1;\rho+1)B(k+\alpha+1;\rho)}{(k+\alpha)B(k+\alpha;\rho+1)B(k+\alpha;\rho+1)} < 0,
\end{aligned}
$$

as required. □

Lemma 5.8. *Let $A := \{a_1, a_2, ..., a_m\}$ be a finite subset of Z_0^+, where the elements are* *ordered $a_1 < a_2 < ... < a_m$, and let f_A and g_A be defined as above. Then*

$$\forall k \geq 1, \quad |g_A(k)| \leq \frac{2}{\rho} \quad \textit{and} \quad |\Delta g_A(k)| \leq \frac{3}{k}.$$

Proof. By Lemma 5.6 we can write, $g_A = \sum_{i=1}^{m}(g_{a_i} - g_{a_i-1})$. for n_i an increasing sequence of non-negative integers. Define the non-decreasing sequence $(n_i)_i$ by $n_{2i-1} := a_i - 1$, $n_{2i} := a_i$. Then $g_A = \sum_{i=1}^{2m}(-1)^i g_{n_i}$. By Lemma 5.4 we know that for each $i = 1, 2, ..., m$, $||g_{n_i}||_\infty \in [-\frac{1}{\rho}, 0]$.

Now by Lemma 5.7 we see that the function of s given by evaluating $g_s(k)$ is monotonically decreasing from $s = 0$ to $s = k$ and is monotonically increasing for $s \geq k$.

There are three possible cases. If $n_i \leq k \quad \forall i \leq 2m$ then $\{g_{n_i}(k)\}_{i \leq 2m}$ is non-increasing. Let $(a_i)_{i \leq 2N}$ be the non-increasing sequence taking values in $[-\frac{1}{\rho}, 0]$ given by $a_i = g_{n_i}(k)$. Then,

$$g_A(k) \quad = \quad \sum_{i=1}^{2m}(-1)^i a_i = \sum_{j=1}^{m}(a_{2j} - a_{2j-1}).$$

Now $a_{2j} - a_{2j-1} \leq 0$ for $j = 1, 2, ..., m$, since the sequence $(a_i)_{i \leq 2m}$ is non-increasing. Thus,

$$\begin{aligned}
0 \geq g_A(k) \quad &= \quad \sum_{j=1}^{m}(a_{2j} - a_{2j-1}) \\
&= \quad a_{2m} - a_1 + \sum_{j=1}^{m-1}(a_{2j} - a_{2j+1}) \\
&\geq \quad a_{2m} \geq -\frac{1}{\rho},
\end{aligned}$$

since $a_i \in \left[-\frac{1}{\rho}, 0\right]$. The second case is $n_1 \geq k$, so that $n_i \geq k$ for all $i = 1, 2, ...$ Then with the definitions above $\{g_{n_i}(k)\}_{i\leq 2m}$ is non-decreasing and takes values in $[-\frac{1}{\rho}, 0]$. Thus,

$$\begin{aligned}
0 \leq g_A(k) &= \sum_{j=1}^{m}(a_{2j} - a_{2j-1}) \\
&= a_{2m} - a_1 + \sum_{j=1}^{m-1}(a_{2j} - a_{2j+1}) \\
&\leq a_{2m} \leq \frac{1}{\rho},
\end{aligned}$$

as required. The third case is that there is some $1 \leq r \leq 2m$ such that $n_j \leq k$ whenever $j \leq r$ and $n_j > k$ whenever $j > r$. Then $\{g_{n_i}(k)\}_{i\leq 2m}$ is non-increasing for $i = 1, 2, ..., r$ and non-decreasing for $i = r+1, r+2, ..., 2m$. Then, by the above,

$$\begin{aligned}
|g_A(k)| &= \Big|\sum_{i=1}^{2m}(-1)^i g_{n_i}(k)\Big| \\
&= \Big|\sum_{i=1}^{r}(-1)^i g_{n_i}(k)\Big| + \Big|\sum_{i=r}^{2m}(-1)^i g_{n_i}(k)\Big| \\
&\leq \frac{1}{\rho} + \frac{1}{\rho} = \frac{2}{\rho},
\end{aligned}$$

as required.

For the first difference bounds, recall that $\Delta g_A(k) = \frac{\rho}{k} g_A(k) - \frac{f_A(k) - \mu_\rho(f_A)}{k}$. The Stein factors obtained for $|g_A|$ give us,

$$|\Delta g_A(k)| \leq \Big|\frac{2\rho}{\rho k}\Big| + \Big|\frac{1}{k}\Big| = \frac{3}{k},$$

as required. $\square$

Theorem 5.9. *Let V and W be discrete random variables taking values in the non-negative integers, with V satisfying $\mathcal{L}(V) = \mu_\rho^\alpha$, $\mu_\rho^\alpha(k) = \frac{B(k+\alpha;\rho+1)}{B(\alpha;\rho)}$. Then the total variation distance between V and W is bounded by,*

$$d_{TV}(V,W) \leq \sup_{g \in \mathcal{I}_\rho^\alpha} \left|\mathbb{E}\left((W+\alpha+\rho)g(W) - (W+\alpha)g(W+1)\right)\right|,$$

where,

$$\mathcal{I}_\rho^\alpha = \left\{ g : g(0) = 0, \quad \forall k \geq 0 \quad g(k) \leq 0 \quad |g(k)| \leq \frac{2}{\rho}, \quad |\Delta g(k)| \leq \frac{3}{k} \right\}. \tag{5.10}$$

Proof. Let $\mathcal{J} := \{\mathbf{1}_A : A \subset \mathbb{Z}_0^+\}$. For $n \leq |A|$, define $A_n \subset A$ to be the set containing the first n elements of A in order of size.

For each function $\mathbf{1}_{A_n}$, we can define a solution, g_{A_n} of the Stein equation as given by Proposition 5.2. Moreover, Lemma 5.8 tells us that $g_{A_n} \in \mathcal{I}_\rho^\alpha$ given by (5.10), so $\mathcal{I}_\rho^\alpha$ contains all solutions of Stein equations for indicator functions of the form $\mathbf{1}_{A_n}$. Thus,

$$\begin{aligned}
d_{TV}(V,W) &= \sup_{\mathbf{1}_A \in \mathcal{J}} |\mathbb{E}(\mathbf{1}_A(W) - \mu_\rho^\alpha(\mathbf{1}_A))| \\
&= \sup_{\mathbf{1}_A \in \mathcal{J}} \lim_{n\to\infty} |\mathbb{E}(\mathbf{1}_{A_n}(W) - \mu_\rho^\alpha(\mathbf{1}_{A_n}))| \\
&\quad \text{(by dominated convergence and continuity of the modulus)} \\
&\leq \sup_{\mathbf{1}_A \in \mathcal{J}} \lim_{n\to\infty} \sup_{g \in \mathcal{I}_\rho^\alpha} \left|\mathbb{E}\left((W+\alpha+\rho)g(W) - (W+\alpha)g(W+1)\right)\right| \\
&\leq \sup_{g \in \mathcal{I}_\rho^\alpha} \left|\mathbb{E}\left((W+\alpha+\rho)g(W) - (W+\alpha)g(W+1)\right)\right|,
\end{aligned}$$

as required. □

Remark 5.10. *Comparing (5.7) and (5.10) we see that the total variation distance bounds that we can obtain using Theorem 5.9 are of the same order as the Kolmogorov distance bounds from Theorem 5.5. This suggests that there are additional properties of the functions corresponding to Kolmogorov distance that are not captured in (5.7) and that would* *need to be understood in order to obtain optimal bounds for many random variables.*

5.2.3 Example

To end this chapter we apply Theorem 5.5 and Theorem 5.9 to obtain bounds for the Kolmogorov and total variation distances between elements of a sequence of random variables and the μ_ρ^α-distribution. The sequence consists of truncated versions of the μ_ρ^α-distribution.

Although we can calculate the distances for such random variables explicitly, we perform these calculations to test our Stein factors for the μ_ρ^α-distribution. We shall see that the bounds that we obtain are of the same order as the actual distances, indicating that the Stein factors found in Lemma 5.4 and Lemma 5.8 are of optimal order for this example.

As usual define the measure μ_ρ^α on $\mathbb{Z}_0^+$ by (5.1). For $r \in \mathbb{Z}_0^+$, we define the truncated version of μ_ρ^α by

$$\begin{aligned} \nu_r(s) &:= \mu_\rho^\alpha(s) \text{ for } s = 0, 1, 2, ..., r; \\ \nu_r(r+1) &:= \sum_{i=r+1}^{\infty} \mu_\rho^\alpha(i) = \frac{B(n+\alpha+1;\rho)}{B(\alpha;\rho)}; \\ \nu_r(s) &:= 0 \text{ for } s = r+2, r+3, r+4, ... \end{aligned} \tag{5.11}$$

Remark 5.11. *We can see that for $r \in \mathbb{Z}_0^+$, if W_r and V are random variables satisfying $\mathcal{L}(W_r) = \nu_r$ and $\mathcal{L}(V) = \mu_\rho^\alpha$ respectively, then*

$$\begin{aligned} d_K(W_r, V) &= \sup_{k \in \mathbb{Z}_0^+} |\mathbb{P}(W_r \le k) - \mathbb{P}(V \le k)| \\ &= |\mathbb{P}(W_r \le r+1) - \mathbb{P}(V \le r+1)| \\ &= \sum_{i=r+2}^{\infty} \mu_\rho^\alpha(i) = \frac{B(r+\alpha+2;\rho)}{B(\alpha;\rho)}, \end{aligned}$$

and

$$\begin{aligned} d_{TV}(W_r, V) &= \frac{1}{2}\sum_{i=0}^{\infty} |\mathbb{P}(W_r = i) - \mathbb{P}(V = i)| \\ &= \frac{1}{2}\sum_{i=r+2}^{\infty} \mu_\rho^\alpha(i) + \frac{1}{2}\sum_{i=r+2}^{\infty} \mu_\rho^\alpha(i) = \frac{B(r+\alpha+2;\rho)}{B(\alpha;\rho)}. \end{aligned}$$

Proposition 5.12. *Let $r \in \mathbb{Z}_0^+$ and let W_r and V be random variables satisfying $\mathcal{L}(W_r) = \nu_r$ and $\mathcal{L}(V) = \mu_\rho^\alpha$ with μ_ρ^α and ν_r defined as above. Then,*

$$\begin{aligned} d_K(W_r, V) &\le \frac{B(r+\alpha+2;\rho)}{B(\alpha;\rho)}; \\ d_{TV}(W_r, V) &\le \frac{B(r+\alpha+2;\rho)}{B(\alpha;\rho)} \left(\frac{3(r+\alpha+\rho+1)}{r+1}\right). \end{aligned}$$

Proof. Define $\mathcal{H}_\rho^\alpha$ and $\mathcal{I}_\rho^\alpha$ as in (5.7) and (5.10) respectively. It follows from Theorem 5.5 and Theorem 5.9 that

$$\begin{aligned} d_K(V, W_r) &\le \sup_{g\in\mathcal{H}_\rho^\alpha} |\mathbb{E}\{(W_r+\alpha+\rho)g(W_r) - (W_r+\alpha)g(W_r+1)\}|; \\ d_{TV}(V, W_r) &\le \sup_{g\in\mathcal{I}_\rho^\alpha} |\mathbb{E}\{(W_r+\alpha+\rho)g(W_r) - (W_r+\alpha)g(W_r+1)\}|. \end{aligned}$$

Then, for $g \in \mathcal{H}_\rho^\alpha \cup \mathcal{I}_\rho^\alpha$,

$$
\begin{aligned}
&\mathbb{E}\{(W_r+\alpha+\rho)g(W_r) - (W_r+\alpha)g(W_r+1)\} \\
&= \sum_{i=0}^{\infty}\{(i+\alpha+\rho)g(i) - (i+\alpha)g(i+1)\}\nu_r(i) \\
&= \sum_{i=0}^{r}\{(i+\alpha+\rho)g(i) - (i+\alpha)g(i+1)\}\mu_\rho^\alpha(i) \\
&\quad +\{(r+\alpha+\rho+1)g(r+1) - (r+\alpha+1)g(r+2)\}\sum_{i=r+1}^{\infty}\mu_\rho^\alpha(i).
\end{aligned}
$$

By (5.3) and since $g(0)=0$,

$$
\begin{aligned}
&\sum_{i=0}^{r}\{(i+\alpha+\rho)g(i) - (i+\alpha)g(i+1)\}\mu_\rho^\alpha(i) \\
&= \sum_{i=1}^{r}(i+\alpha-1)g(i)\mu_\rho^\alpha(i-1) - \sum_{i=0}^{r}(i+\alpha)g(i+1)\mu_\rho^\alpha(i) \\
&= \sum_{i=0}^{r-1}(i+\alpha)g(i+1)\mu_\rho^\alpha(i) - \sum_{i=0}^{r}(i+\alpha)g(i+1)\mu_\rho^\alpha(i) \\
&= -(r+\alpha)g(r+1)\mu_\rho^\alpha(r). \qquad (5.12)
\end{aligned}
$$

So

$$
\begin{aligned}
&|\mathbb{E}\{(W_r+\alpha+\rho)g(W_r)-(W_r+\alpha)g(W_r+1)\}| \\
&= \Big| -(r+\alpha)g(r+1)\mu_\rho^\alpha(r) \\
&\quad +((r+\alpha+\rho+1)g(r+1)-(r+\alpha+1)g(r+2))\sum_{i=r+1}^{\infty}\mu_\rho^\alpha(i)\Big| \\
&= \Big| -(r+\alpha)g(r+1)\frac{B(r+\alpha;\rho+1)}{B(\alpha;\rho)} \\
&\quad +((r+\alpha+\rho+1)g(r+1)-(r+\alpha+1)g(r+2))\,\frac{B(r+\alpha+1;\rho)}{B(\alpha;\rho)} \\
&= ((-\rho+r+\alpha+\rho+1)g(r+1)-(r+\alpha+1)g(r+2))\,\frac{B(r+\alpha+1;\rho)}{B(\alpha;\rho)} \\
&\quad \text{(by Appendix A, (A.4))} \\
&= (r+\alpha+1)|\Delta g(r+1)|\frac{B(r+\alpha+1;\rho)}{B(\alpha;\rho)} \\
&= (r+\alpha+\rho+1)|\Delta g(r+1)|\frac{B(r+\alpha+2;\rho)}{B(\alpha;\rho)}
\end{aligned}
$$

by Appendix A, (A.2). Thus,

$$
\begin{aligned}
d_K(V,W_r) &\le \sup_{g\in\mathcal{H}_\rho^\alpha}(r+\alpha+\rho+1)\frac{B(r+\alpha+2;\rho)}{B(\alpha;\rho)}|\Delta g(r+1)| \\
&\le \frac{B(r+\alpha+2;\rho)}{B(\alpha;\rho)},
\end{aligned}
$$

and

$$
\begin{aligned}
d_{TV}(V,W_r) &\le sup_{g\in\mathcal{I}_\rho^\alpha}(r+\alpha+\rho+1)\frac{B(r+\alpha+2;\rho)}{B(\alpha;\rho)}|\Delta g(r+1)| \\
&\le \frac{B(r+\alpha+2;\rho)}{B(\alpha;\rho)}\left(\frac{3(r+\alpha+\rho+1)}{r+1}\right),
\end{aligned}
$$

as required. □

Not only does our upper bound take the actual order of convergence in both total variation distance and in Kolmogorov distance of the sequence $(W_r)_{r\in\mathbb{N}}$ to V, but we gain precisely the actual Kolmogorov distance for the Kolmogorov distance bound. This suggests that, for this problem, our Stein factors for Δg are of optimal order as $r \to \infty$ in both metrics and that the bound for the Kolmogorov distance is optimal for this problem.

The fact that our upper bound for the Kolmogorov distance is the exact distance is not only a consequence of our optimal Stein factor for Δg but is also consequence of our special construction of ν_r, with $\nu_r(i) = \mu_\rho^\alpha(i)$ for $i = 0, 1, ..., r$ and $\nu_r(i) = 0$ for $i > r + 1$. One would expect such a definition to be particularly easy to bound using our Stein equation. In particular note that in (5.12) we were able to discard the first $r - 1$ terms of the sum

$$\sum_{i=0}^{\infty} \left((i + \alpha + \rho)g(i) - (i + \alpha)g(i + 1)\right) \nu_r(i).$$

This is unsurprising since our Stein equation was chosen based on the fact that this is a telescoping series when ν_r is replaced with μ_ρ^α.

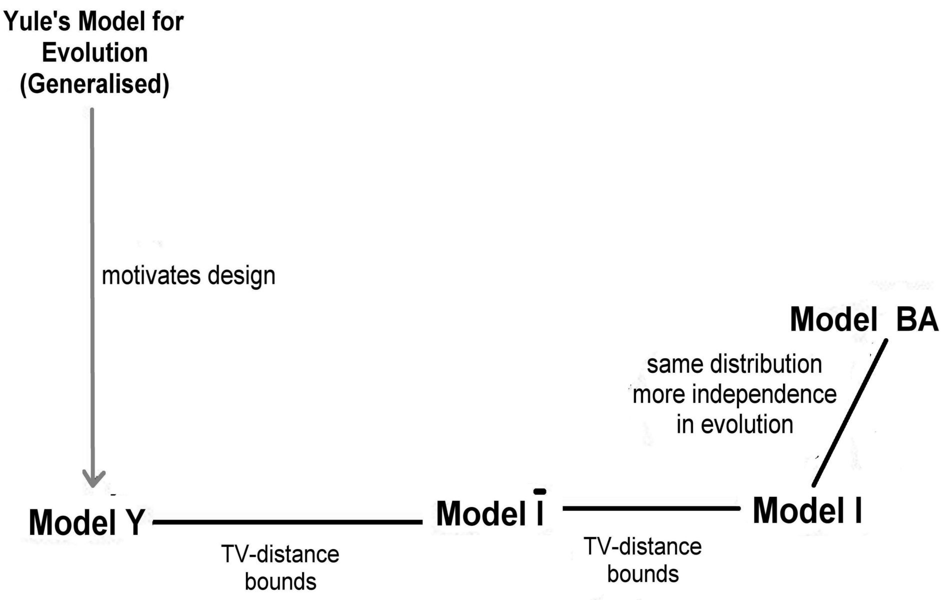
Chapter 6

Coupling the degree evolution in the discrete-time and continuous-time models

Chapters 2 and 3 gave continuous-time and discrete-time random graph models that evolve via a preferential attachment condition. Particular versions of the models have already been proved asymptotically to display power law degree distributions. For example, Yule [56] demonstrates this for the continuous-time process described in Chapter 3 and Bollobás et al. [11] prove convergence of the degree sequence for a discrete-time model.

This chapter presents a sequence of discrete-time and continuous-time preferential attachment models. We couple the models so that we can bound the total variation distance between their degree distributions. Later in the thesis we shall use a number of different methods to bound the total variation distances between degree distributions of random graphs within this scheme and the μ_ρ^α-distribution. Each method applies to specific models within our framework. The bounds of this chapter in contrast extend those results to cover all graphs within this framework.

Throughout this thesis we denote the degree distribution of a model, Model M, at step n by $W_n^{(M)}$. The diagram below shows how our models are related and indicates which degree distribution bounds are calculated.



6.1 Barabási-Albert models

6.1.1 Model BA

As explained in Chapter 2, models from the family described below are straightforward versions of the Barabási-Albert model, with a precise description for the vertex degree evolution. Let $\alpha > 0$ be a real parameter and let $m \leq m_0$ be integer parameters.

Model BA

- **(BA1)** Start with the empty graph on m_0 vertices.
- **(BA2)** At step n select m distinct vertices from the graph in such a way that the vertex i with $D_{n-1}^{(BA)}(i) := (\alpha +$ in-degree of i at step $n-1)$ is chosen with probability proportional to $D_{n-1}^{(BA)}(i)$. More precisely,

$$\mathbb{P}(\text{vertex } i \text{ is chosen}|D_{n-1}^{(BA)}(i)) = \frac{mD_{n-1}^{(BA)}(i)}{(n-1)(m+\alpha)+m_0\alpha}.$$

(Note that the conditioning here is not on the whole degree vector at step $n-1$.)

- **(BA3)** Attach the new vertex, $n+m_0$, to the m chosen vertices with m edges directed out of the new vertex.

Theorem 2.1 of Chapter 3 proved that models in this family exist. We note that for all vertices except the initial vertices the total degree is ($m+$ in-degree), so knowing the in-degree distribution is asymptotically the same as knowing the total degree distribution.

One cause of difficulty in analysing this model is the dependence between the degrees of distinct vertices. We now describe a model with independent edge growth, Model I, and prove that the distributions given by observing the degrees of one randomly chosen vertex from each model are the same.

6.1.2 Model I: more independence

Model I

- Start with the empty graph on m_0 vertices;
- At step n, independently for each vertex in the graph, select vertex i with conditional probability $\frac{mD_{n-1}^{(BA)}(i)}{(n-1)(m+\alpha)+m_0\alpha}$, where $D_{n-1}^{(I)}(i) := (\alpha+$ in-degree of i at step $n-1)$, for some $\alpha > 0$.
- Attach the new vertex $n+m_0$ to all the selected vertices with edges directed out of the new vertex.

Note that the number of vertices selected is random in this model. This model is included in Appendix B for reference.

Theorem 6.1. *Let $I_{n+m_0} \sim Uniform \in \{1, 2, ..., n+m_0\}$. If $D_n^{(BA)}(I_{n+m_0})$ and $D_n^{(I)}(I_{n+m_0})$ are defined to be the degrees of vertex I_{n+m_0} in Models BA and I respectively with the same* *parameters, then,*

$$D_n^{(BA)}(I_{n+m_0}) =^D D_n^{(I)}(I_{n+m_0}).$$

Proof. We proceed by induction on n.

Clearly $D_0^{(BA)}(I_{m_0}) =^D D_0^{(I)}(I_{m_0})$ since the degrees in both models at time 0 are identically deterministic.

For $n = k \geq 0$ suppose $D_k^{(BA)}(I_{k+m_0}) =^D D_k^{(I)}(I_{k+m_0})$. Consider $n = k+1$. Then for any $r \in \mathbb{Z}_0^+$,

$$\begin{aligned}
&\mathbb{P}(D_{k+1}^{(BA)}(I_{k+m_0+1}) = \alpha + r) \\
&= \frac{1}{k+m_0+1} \sum_{i=1}^{k+m_0+1} \mathbb{P}(D_{k+1}^{(BA)}(i) = \alpha + r) \\
&= \frac{1}{k+m_0+1} \left(\mathbf{1}_{\{r=0\}} + \sum_{i=1}^{k+m_0} \sum_{s=0}^{r} \mathbb{P}(D_{k+1}^{(BA)}(i) = \alpha + r | D_k^{(BA)}(i) = \alpha + s) \mathbb{P}(D_k^{(BA)}(i) = \alpha + s) \right) \\
&\quad \text{(since } D_{k+1}^{(BA)}(k+m_0+1) = \alpha) \\
&= \frac{1}{k+m_0+1} \left(\mathbf{1}_{\{r=0\}} + \sum_{i=1}^{k+m_0} \sum_{s=0}^{r} \mathbb{P}(D_{k+1}^{(I)}(i) = \alpha + r | D_k^{(I)}(i) = \alpha + s) \mathbb{P}(D_k^{(I)}(i) = \alpha + s) \right) \\
&\quad \text{(by inductive hypothesis and conditional probabilities in the two models)} \\
&= \mathbb{P}(D_{k+1}^{(I)}(I_{k+m_0+1}) = \alpha + r).
\end{aligned}$$

□

Thus the distributions of the random variables $W_n^{(BA)} = D_n^{(BA)}(I_{n+m_0}) - \alpha$ and $W_n^{(I)} = D_n^{(I)}(I_{n+m_0}) - \alpha$ are identical and it follows that $d_{TV}(W_n^{(BA)}, W_n^{(I)}) = 0$.

6.1.3 Model $\bar{\text{I}}$: starting with one vertex

We now present a slightly altered version of Model I, Model $\bar{\text{I}}$, whose evolution will be coupled with that of the continuous-time Model Y so that we can bound the distances between their degree distributions. This model evolves in the same way as Model I, but for ease of comparison with Yule's model for evolution we start with only one vertex.

Model $\bar{\text{I}}$ looks at a different set of parameters from Model I. Instead of parameters $\alpha > 0, m \in \mathbb{N}$ and $m_0 \in \mathbb{N}$ where $m_0 \geq m$, Model $\bar{\text{I}}$ uses parameters $\alpha > 0, \beta > 0, \rho > 1$. To compare the models we use the correspondence $\alpha = \alpha$, $\frac{m}{m+\alpha} = \frac{1}{\rho}$ and $\frac{(m_0-1)\alpha - m}{m} = \beta$.

The model evolves as follows.

Model $\bar{\text{I}}$

- Start with one vertex;
- At step n, independently for each vertex, select vertex i with conditional probability $\frac{D_{n-1}^{(\bar{I})}(i)}{\rho n+\beta}$, where $D_{n-1}^{(\bar{I})}(i) := (\alpha + \text{in-degree of } i \text{ at step } n-1)$, $\rho > 1$, $\alpha > 0, \beta \geq 0$ satisfying $\alpha \leq \beta + \rho$. The condition $\alpha \leq \beta + \rho$ ensures that all probabilities lie between 0 and 1.
- Attach a new vertex $n+1$ to all the selected vertices with edges directed out of the new vertex. (The number of vertices selected at step n is random.)

Theorem 6.2. *Let $m_0 \geq m \geq 1$, let $\alpha \geq 1$ and let $G_n^{(\bar{I})}$ be an evolving random graph satisfying Model $\bar{\text{I}}$ with $\rho := \frac{m+\alpha}{m}$, $\beta = \frac{(m_0-1)\alpha - m}{m}$ and assume that $\beta \geq 0$,*

Then, with the usual notation,

$$d_{TV}(W_n^{(\bar{I})}, W_n^{(I)}) \leq \frac{m_0 - 1}{n + m_0 - 1}.$$

If $m_0 = 1$ the two models are identical so we rightly get a distance of 0.

Proof. Let I_n and J_{n+m_0-1} be the vertices whose degrees are observed by $W_n^{(\bar{I})}$ and $W_n^{(I)}$ respectively, so $I_n \sim Uniform\{1,2,...,n\}$ and $J_{n+m_0-1} \sim Uniform\{1,2,...,n+m_0-1\}$. Couple I_n and J_{n+m_0-1} so that,

$$\begin{aligned} \mathbb{P}(J_{n+m_0-1} = I_n + m_0 - 1 | I_n) &= \frac{n}{n+m_0-1}, \\ \mathbb{P}(J_{n+m_0-1} = I_{m_0-1} | I_n) &= \frac{m_0-1}{n+m_0-1}. \end{aligned}$$

Values of J_{n+m_0-1} lying between 1 and m_0-1 represent the extra initial vertices of Model I not included in Model Ī, while values of J_{n+m_0-1} greater than m_0-1 represent the initial vertex that the models have in common and the vertices added at later time steps.

Conditional on $J_{n+m_0} \geq m_0$, $W_n^{(\bar{I})} =^D W_n^{(I)}$. This can be seen from the fact that, subject to this condition, the vertices J_{n+m} in Model I and I_n in Model Ī evolve independently of the other vertices in their graphs, they enter the graphs at the same time steps, they share the same initial conditions and they follow the same stochastic processes up to step n.

It follows that we can couple our models by forcing vertex i in Model Ī to take the same degree as vertex $i + m_0$ in Model I, for $i = 0, 1, ..., n$. In this coupling $W_n^{(\bar{I})} = W_n^{(I)}$ whenever $J_{n+m_0-1} \geq m_0$. We get,

$$\begin{aligned} d_{TV}(W_n^{(\bar{I})}, W_n^{(I)}) &\leq \mathbb{P}(W_n^{(\bar{I})} \neq W_n^{(I)}) \\ &\leq \mathbb{P}(J_{n+m_0-1} < m_0) \\ &= \frac{m_0-1}{n+m_0-1}. \end{aligned}$$

□

6.1.4 Model Y: motivated by Yule's model for evolution

Yule's original model described in the previous section consists of a *'Yule process of Yule processes'*. Vertices arrive in the graph following a Yule process, and for each vertex in the graph, the addition of edges to each vertex forms a Yule process that is independent of the rest of the system. Section 3.4 introduced a more general random graph model in which vertices gained edges following a Yule process with immigration. Later, in Chapter 3, we calculated bounds for convergence of the degree distribution of this general model.

We now attempt to simplify this model to find a model that can be compared with discrete-time Barabási-Albert random graphs. In the following model vertices still grow in degree according to independent Yule processes with immigration, but vertices now enter the graph at deterministic time points. We define the times of arrival using the $\alpha > 0$ and $\beta \geq 0$ notation of Model $\bar{\text{I}}$, so that Model Y matches up with this model as closely as possible.

Model Y

1. **Arrivals of new vertices into the graph.** The time between the arrival of vertex n and vertex $n+1$ is given by $s_n := \frac{1}{\rho n+\beta}$ for $\rho > 1$, $\beta \geq 0$.
2. **Growth of edges for current vertices.** Each vertex i gains incident in-edges independently with rate given by $D_n^{(Y)}(i) := \alpha+$ in-degree, for $\alpha > 0$.

As usual this definition can be found in Appendix B for reference. In Chapter 7 we shall obtain bounds for the distance between the degree distribution of this random graph and the μ_ρ^α-distribution.

6.2 Coupling Model Y and Model $\bar{\text{I}}$

We now define a coupling of Model Y and Model $\bar{\text{I}}$, through which we bound the distance between the degree distributions of the two models. Once we have this bound we are able to use the triangle inequality to bound the distances between the degree distributions of any pair of models in our scheme.

To help us to define the coupling we first look at an alternative description for the evolution of Model Y. This description gives the evolution of the graph as a discrete-time process where time steps correspond to the arrivals of new vertices in the graph, and the distributions of the changes in vertex degree at each time step is described.

First, recall the original definition of Model Y given above. Since the time of the arrival of the n^{th} vertex is deterministic, equivalently we can define this model by the distribution of the number of edges added to each vertex in the graph during each time interval $(s_1 + ... + s_{n-1}, s_1 + ... + s_n]$. Then conditional on the degree of vertex i at step n, $D_{n-1}^{(Y)}(i) = d_{n-1}^{(Y)}(i)$, the degree of vertex i at step $n+1$ can be viewed as the outcome at time $s_n = e^{-\frac{1}{\rho n+\beta}}$ of a Yule process with immigration, where the birth rate is 1 and the immigration rate is $d_{n-1}^{(Y)}(i)$ and the initial size of the population is 0.

It is a consequence of Theorem 3.11 that, equivalently,

$$D_n^{(Y)}(i) = D_{n-1}^{(Y)}(i) + E_n^{(Y)}(i),$$

where $E_n^{(Y)}(i) \sim Negbinom(D_{n-1}^{(Y)}(i), e^{-\frac{1}{\rho n+\beta}})$ with the negative binomial distribution defined as on page viii.

We therefore reformulate Model Y as follows,

- Start with one vertex, 1, of 'degree' α;
- At step n add one new vertex, $n+1$, of 'degree' α to the graph.

- Conditional on the degree sequence,

$$D_{n-1}^{(Y)}(1) = d_{n-1}^{(Y)}(1), D_{n-1}^{(Y)}(2) = d_{n-1}^{(Y)}(2), ..., D_{n-1}^{(Y)}(n) = d_{n-1}^{(Y)}(n)$$

  of the first n vertices at step $n-1$, generate independent random variables and call them $E_n^{(Y)}(1), ..., E_n^{(Y)}(n)$ where $E_n^{(Y)}(i) \sim Negbinom(d_{n-1}^{(Y)}(i), e^{-\frac{1}{\rho n+\beta}})$. At step n add $E_n^{(Y)}(i)$ incident in-edges to vertex i.

The random variables $E_n^{(Y)}(i)$ can take any non-negative integer values, but if n is large $E_n^{(Y)}(i)$ will be 0 with high probability unless $D_{n-1}^{(Y)}(i)$ is large. We consider the probabilities that $E_n^{(Y)}(i)$ takes the values 0 and 1 in Model Y.

Lemma 6.3. *For the random variables defined in the model above the following inequalities hold,*

$$1 - \frac{D_{n-1}^{(Y)}(i)}{\rho n + \beta} \le \mathbb{P}(E_n^{(Y)}(i) = 0 | D_{n-1}^{(Y)}(i)) \le 1 - \frac{D_{n-1}^{(Y)}(i)}{\rho n + \beta} + \frac{(D_{n-1}^{(Y)}(i))^2}{2(\rho n + \beta)^2},$$

and

$$\mathbb{P}(E_n^{(Y)}(i) = 1 | D_{n-1}^{(Y)}(i)) \ge \frac{D_{n-1}^{(Y)}(i)}{\rho n + \beta} - \frac{D_{n-1}^{(Y)}(i)}{2(\rho n + \beta)^2} - \frac{(D_{n-1}^{(Y)}(i))^2}{(\rho n + \beta)^2}.$$

Proof. It can be shown analytically that for any $x \ge 0$,

$$1 - x \le e^{-x} \le 1 - x + \frac{x^2}{2},$$

and from this it follows that for any $x \ge 0$,

$$x - \frac{x^2}{2} \le 1 - e^{-x} \le x.$$

Since $\frac{D_{n-1}^{(Y)}(i)}{\rho n+\beta} \ge 0$ the first result follows immediately from the definition, $\mathbb{P}(E_n^{(Y)}(i) = 0 | D_{n-1}^{(Y)}(i)) = e^{-\frac{D_{n-1}^{(Y)}(i)}{\rho n+\beta}}$.

For the second, first note by Theorem 3.11 that, $\mathbb{P}(E_n^{(Y)}(i) = 1|D_{n-1}^{(Y)}(i)) = e^{-\frac{D_{n-1}^{(Y)}(i)}{\rho n+\beta}}(1 - e^{-\frac{1}{\rho n+\beta}})D_{n-1}^{(Y)}(i)$. Taking the lower bounds for e^{-x} and $1 - e^{-x}$ found above,

$$e^{-\frac{D_{n-1}^{(Y)}(i)}{\rho n+\beta}} \geq 1 - \frac{D_{n-1}^{(Y)}(i)}{\rho n + \beta}$$

and

$$(1 - e^{-\frac{1}{\rho n+\beta}}) \geq \frac{1}{\rho n + \beta} - \frac{1}{2(\rho n + \beta)^2}.$$

So,

$$\begin{aligned}
&\mathbb{P}(E_n^{(Y)}(i) = 1|D_{n-1}^{(Y)}(i)) \\
&\quad \geq \left(1 - \frac{D_{n-1}^{(Y)}(i)}{\rho n + \beta}\right)\left(\frac{1}{\rho n + \beta} - \frac{1}{2(\rho n + \beta)^2}\right) D_{n-1}^{(Y)}(i) \\
&\quad \geq \frac{D_{n-1}^{(Y)}(i)}{\rho n + \beta} - \frac{D_{n-1}^{(Y)}(i)}{2(\rho n + \beta)^2} - \frac{(D_{n-1}^{(Y)}(i))^2}{(\rho n + \beta)^2},
\end{aligned}$$

as required. □

The above lemma bounds the conditional probabilities $\mathbb{P}(E_n^{(Y)} = 0|D_{n-1}^{(Y)}(i))$ from $1 - \frac{D_{n-1}^{(Y)}(i)}{\rho n+\beta}$ and $\mathbb{P}(E_n^{(Y)} = 1|D_{n-1}^{(Y)}(i))$ from $\frac{D_{n-1}^{(Y)}(i)}{\rho n+\beta}$. This is a promising start since these probabilities are similar in form to the transition probabilities in Model $\bar{\text{I}}$.

We now recall the definition of Model $\bar{\text{I}}$ that can be found in Appendix B. Our main result is the following set of bounds.

Theorem 6.4. *Let $\alpha \geq 1$, $\beta \geq 0$ and $\rho > 1$. Let $I_n \sim Uniform\{1, 2, ..., n\}$ and let $W_n^{(Y)}$ and $W_n^{(\bar{I})}$ represent the in-degrees of vertex I_n at step n in Model Y and Model $\bar{\text{I}}$*

respectively. Then,

$$
\begin{aligned}
d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) \quad \leq \quad & \frac{3\alpha(\alpha+1)}{2n}\sum_{i=2}^{n}\sum_{r=i+1}^{n}\frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho(i-1)+\beta)^{\frac{2}{\rho}}} \\
& +\frac{3\alpha(\alpha+1)}{2n}\sum_{r=2}^{n}\frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho(i-1)+\beta)^{\frac{2}{\rho}}}e^{\frac{1}{\rho+\beta}} \\
& +\frac{3\alpha^2(2\rho+\beta)}{2\rho(\rho+\beta)^2 n}.
\end{aligned}
$$

In particular, if $\rho \in (1,2)$,

$$
d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) \leq B_1(\alpha,\beta,\rho,n) = O(n^{2(\frac{1}{\rho}-1)}) \text{ as } n \to \infty,
$$

if $\rho = 2$,

$$
d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) < B_2(\alpha,\beta,\rho,n) = O\left(\frac{(\log n)^2}{n}\right) \text{ as } n \to \infty,
$$

and if $\rho > 2$,

$$
d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) \leq B_3(\alpha,\beta,\rho,n) = O\left(\frac{\log n}{n}\right) \text{ as } n \to \infty,
$$

where

$$
\begin{aligned}
B_1(\alpha,\beta,\rho,n) &= \frac{3\alpha(\alpha+1)(\rho n+\beta)^{\frac{2}{\rho}-1}}{2(2-\rho)n}\left(1+\frac{(\beta+\rho)^{1-\frac{2}{\rho}}}{2-\rho}\right) \\
&\quad +\frac{3e^{\frac{1}{\beta+\rho}}\alpha(\alpha+1)}{2n(2-\rho)(\beta+\rho)}+\frac{3\alpha^2(2\rho+\beta)}{2\rho(\beta+\rho)^2 n}, \qquad (6.1)
\end{aligned}
$$

$$
\begin{aligned}
B_2(\alpha,\beta,\rho,n) &= \frac{3\alpha(\alpha+1)}{2n}\left(3\log\frac{2n+\beta}{2+\beta}+\frac{1}{16}(\log(2n+\beta))^2\right) \\
&\quad +\frac{3\alpha^2(4+\beta)}{4(2+\beta)^2 n}, \qquad (6.2)
\end{aligned}
$$

$$
\begin{aligned}
B_3(\alpha,\beta,\rho,n) &= \frac{3\alpha(\alpha+1)}{2(\rho-2)n}\left((\beta+\rho)^{\frac{2}{\rho}-1}+\frac{1}{\rho}\log\frac{\rho n+\beta}{2\rho+\beta}\right) \\
&\quad +\frac{3\alpha(\alpha+1)}{2n}\left(\frac{e^{\frac{1}{\beta+\rho}}}{(\rho-2)(\beta+\rho)}\right). \qquad (6.3)
\end{aligned}
$$

Proof. To prove the theorem we use that the total variation distance between $W_n^{(Y)}$ and $W_n^{(\bar{I})}$ can be bounded by $\mathbb{P}(W_n^{(Y)} \neq W_n^{(\bar{I})})$ ([8], p. 254). We then couple Model Y and Model Ī so that the degree evolution of I_n coincides with high probability. We bound $\mathbb{P}(D_n^{(Y)}(i) \neq D_n^{(\bar{I})}(i))$ and condition on I_n to obtain bounds for $\mathbb{P}(W_n^{(Y)} \neq W_n^{(\bar{I})})$.

Lemma 6.5. *Using the notation above,*

$$
d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) \leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{P}(D_n^{(Y)}(i) \neq D_n^{(\bar{I})}(i)).
$$

Proof. Conditioning on I_n,

$$\begin{aligned} d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) &\leq \mathbb{P}(W_n^{(Y)} \neq W_n^{(\bar{I})}) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(W_n^{(Y)} \neq W_n^{(\bar{I})} | I_n = i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(D_n^{(Y)}(i) \neq D_n^{(\bar{I})}(i)), \end{aligned}$$

as required. □

We need to bound $\mathbb{P}(W_n^{(Y)} \neq W_n^{(\bar{I})}) = \mathbb{P}(D_n^{(Y)}(i) \neq D_n^{(\bar{I})}(i))$ to obtain the required total variation distance bound for each vertex i.

Fix $i \in \{1, 2, ..., n\}$ and condition on the evolution of the degree of i up to time n,

$$D_1^{(Y)}(i) = d_1^{(Y)}(i), D_2^{(Y)}(i) = d_2^{(Y)}(i), ..., D_{n-1}^{(Y)}(i) = d_{n-1}^{(Y)}(i).$$

We look at the following couplings for Model Y and Model $\bar{\text{I}}$. Let $E_n^{(M)}(i)$ represent the number of edges added to vertex i at step n in Model $M \in \{Y, \bar{\text{I}}\}$

Coupling for Y and $\bar{\textbf{I}}$: We define the evolution of Model $\bar{\text{I}}$ based on the evolution of Model Y as follows for each step $s \geq i$,

$$\begin{aligned} &\mathbb{P}(E_s^{(\bar{I})}(i) = 1 | E_s^{(Y)}(i) \geq 1, D_{s-1}^{(Y)}(i) = D_{s-1}^{(\bar{I})}(i) = d_{s-1}(i)) = 1, \\ &\mathbb{P}(E_s^{(\bar{I})}(i) = 1 | E_s^{(Y)}(i) = 0, D_{s-1}^{(Y)}(i) = D_{s-1}^{(\bar{I})}(i) = d_{s-1}(i))) \\ &\quad= \frac{\mathbb{P}(E_s^{(Y)}(i) = 0 | D_{s-1}^{(Y)}(i) = d_{s-1}(i)) - \left(1 - \frac{d_{s-1}(i)}{\rho s + \beta}\right)}{\mathbb{P}(E_s^{(Y)}(i) = 0 | D_{s-1}^{(Y)}(i) = d_{s-1}(i))}, \\ &\mathbb{P}(E_s^{(\bar{I})}(i) = 0 | E_s^{(Y)}(i) = 0, D_{s-1}^{(Y)}(i) = D_{s-1}^{(\bar{I})}(i) = d_{s-1}(i)) \\ &\quad= 1 - \frac{\mathbb{P}(E_s^{(Y)}(i) = 0 | D_{s-1}^{(Y)}(i) = d_{s-1}(i)) - \left(1 - \frac{d_{s-1}(i)}{\rho s + \beta}\right)}{\mathbb{P}(E_s^{(Y)}(i) = 0 | D_{s-1}^{(Y)}(i) = d_{s-1}(i))}. \end{aligned} \tag{6.4}$$

By Lemma 6.3 these are well-defined probabilities for all possible outcomes of $D^{(Y)}_{s-1}(i)$. This coupling gives the required marginal probability distribution for addition of edges in Model $\bar{\mathrm{I}}$,

$$\begin{aligned}
\mathbb{P}(E^{(\bar{I})}_s(i) = 1 | D^{(\bar{I})}_{s-1}(i) = d_{s-1}(i)) &= \frac{d_{s-1}(i)}{\rho s + \beta}, \\
\mathbb{P}(E^{(\bar{I})}_s(i) = 0 | D^{(\bar{I})}_{s-1}(i) = d_{s-1}(i)) &= 1 - \frac{d_{s-1}(i)}{\rho s + \beta}.
\end{aligned}$$

We couple the $\{E^{(\bar{I})}_s(i) : i = 1, 2, ..., s\}$ such that independence is maintained.

For any vertex i at step s, using this coupling of Models Y and Model $\bar{\mathrm{I}}$,

$$\begin{aligned}
&\mathbb{P}(E^{(Y)}_s(i) \neq E^{(\bar{I})}_s(i) | D^{(Y)}_{s-1}(i) = D^{(\bar{I})}_{s-1}(i) = d_{s-1}(i)) \\
&= \mathbb{P}(E^{(Y)}_s(i) \neq 0 \cap E^{(\bar{I})}_s(i) = 0 | D^{(Y)}_{s-1}(i) = D^{(\bar{I})}_{s-1}(i) = d_{s-1}(i)) \\
&\quad + \mathbb{P}(E^{(Y)}_s(i) \neq 1 \cap E^{(\bar{I})}_s(i) = 1 | D^{(Y)}_{s-1}(i) = D^{(\bar{I})}_{s-1}(i) = d_{s-1}(i)) \\
&= 0 + \mathbb{P}(E^{(Y)}_s(i) = 0 \cap E^{(\bar{I})}_s(i) = 1 | D^{(Y)}_{s-1}(i) = D^{(\bar{I})}_{s-1}(i) = d_{s-1}(i)) \\
&\quad + \mathbb{P}(E^{(Y)}_s(i) > 1 | D^{(Y)}_{s-1}(i) = D^{(\bar{I})}_{s-1}(i) = d_{s-1}(i)),
\end{aligned}$$

since $\mathbb{P}(E^{(Y)}_s(i) \neq 0 \cap E^{(\bar{I})}_s(i) = 0 | D^{(Y)}_{s-1}(i) = D^{(\bar{I})}_{s-1}(i) = d_{s-1}(i)) = 0$. Using the probabilities defined in the coupling (6.4),

$$
\begin{aligned}
&\mathbb{P}(E_s^{(Y)}(i) \neq E_s^{(\overline{I})}(i) | D_{s-1}^{(Y)}(i) = D_{s-1}^{(\overline{I})}(i) = d_{s-1}(i)) \\
&= \mathbb{P}(E_s^{(Y)}(i) = 0 | D_{s-1}^{(Y)}(i) = d_{s-1}(i)) \left(\frac{\mathbb{P}(E_s^{(Y)}(i) = 0 | D_{s-1}^{(Y)}(i) = d_{s-1}(i)) - \left(1 - \frac{d_{s-1}(i)}{\rho s + \beta}\right)}{\mathbb{P}(E_s^{(Y)}(i) = 0 | D_{s-1}^{(Y)}(i) = d_{s-1}(i))} \right) \\
&\quad + \mathbb{P}(E_s^{(Y)}(i) > 1 | D_{s-1}^{(Y)}(i) = d_{s-1}(i)) \\
&= \mathbb{P}(E_s^{(Y)}(i) \neq 1 | D_{s-1}^{(Y)}(i) = d_{s-1}(i)) - \left(1 - \frac{d_{s-1}^{(Y)}(i)}{\rho s + \beta}\right) \\
&\leq 1 - \left(\frac{d_{s-1}(i)}{\rho s + \beta} - \frac{d_{s-1}(i)}{2(\rho s + \beta)^2} - \frac{(d_{s-1}(i))^2}{(\rho s + \beta)^2} \right) - \left(1 - \frac{d_{s-1}(i)}{\rho s + \beta}\right) \quad \text{(by Lemma 6.3)} \\
&\leq \frac{3(d_{s-1}(i))^2}{2(\rho s + \beta)^2}, \text{ since } \alpha \geq 1. \qquad (6.5)
\end{aligned}
$$

In view of Lemma 6.5 we want to bound use the bound $\mathbb{P}(D_n^{(Y)}(i) \neq D_n^{(\overline{I})}(i))$ for each vertex i. We know that

$$
D_n^{(Y)}(i) \neq D_n^{(\overline{I})}(i) \implies \exists r \in \{i, ..., n\} \quad E_r^{(Y)}(i) \neq E_r^{(\overline{I})}(i).
$$

Thus,

$$
\begin{aligned}
&\mathbb{P}(D_n^{(Y)}(i) \neq D_n^{(\overline{I})}(i)) \\
&\leq \mathbb{P}(\bigcup_{r=i}^{n} \{\forall s = i, ..., r-1 \quad E_s^{(Y)}(i) = E_s^{(\overline{I})}(i) \text{ and } E_r^{(Y)}(i) \neq E_r^{(\overline{I})}(i)\}) \\
&\leq \sum_{r=i}^{n} \mathbb{P}\{\forall s = i, ..., r-1 \quad E_s^{(Y)}(i) = E_s^{(\overline{I})}(i) \text{ and } E_r^{(Y)}(i) \neq E_r^{(\overline{I})}(i)\} \\
&\leq \sum_{r=i}^{n} \mathbb{E}\left(\mathbf{1}_{\{E_s^{(Y)}(i) = E_s^{(\overline{I})}(i) \quad \forall s = i, ..., r-1\}} \left(\frac{3(D_{r-1}^{(Y)}(i))^2}{2(\rho r + \beta)^2} \right) \right) \quad \text{(by (6.5))} \\
&\leq \frac{3}{2} \sum_{r=i}^{n} \frac{1}{(\rho r + \beta)^2} \mathbb{E}((D_{r-1}^{(Y)}(i))^2).
\end{aligned}
$$

Therefore, by Lemma 6.5 and the conditioning argument for I_n,

$$d_{TV}(W_n^{(Y)}, W_n^{(\overline{I})}) \leq \frac{3}{2n}\sum_{i=1}^{n}\sum_{r=i}^{n}\frac{1}{(\rho r+\beta)^2}\mathbb{E}((D_{r-1}^{(Y)}(i))^2). \tag{6.6}$$

Now, $(D_{i-1}^{(Y)}(i))^2 = \alpha^2$ and for $r \geq i+1$, by Corollary 3.13 we know that the degrees satisfy the distribution, $D_{r-1}^{(Y)}(i) - \alpha \sim Negbinom(\alpha, e^{-(\frac{1}{\rho i+\beta}+\ldots+\frac{1}{\rho(r-1)+\beta})})$. So using the well-known result that a $Negbinom(\alpha, p)$ distribution has mean $\frac{\alpha(1-p)}{p}$ and variance $\frac{\alpha(1-p)}{p^2}$, we get, for $r \geq i+1$,

$$\begin{aligned}
\mathbb{E}((D_{r-1}^{(Y)}(i))^2) &= Var(D_{r-1}^{(Y)}(i)-\alpha) + [\mathbb{E}(D_{r-1}^{(Y)}(i)-\alpha)]^2 + 2\alpha\mathbb{E}(D_{r-1}^{(Y)}(i)-\alpha) + \alpha^2 \\
&= \alpha\left(e^{2(\frac{1}{\rho i+\beta}+\ldots+\frac{1}{\rho(r-1)+\beta})} - e^{(\frac{1}{\rho i+\beta}+\ldots+\frac{1}{\rho(r-1)+\beta})}\right) \\
&\quad +\alpha^2\left(e^{(\frac{1}{\rho i+\beta}+\ldots+\frac{1}{\rho(r-1)+\beta})} - 1\right)^2 \\
&\quad +2\alpha^2\left(e^{(\frac{1}{\rho i+\beta}+\ldots+\frac{1}{\rho(r-1)+\beta})} - 1\right) + \alpha^2 \\
&= \alpha(\alpha+1)e^{2(\frac{1}{\rho i+\beta}+\ldots+\frac{1}{\rho(r-1)+\beta})} - \alpha e^{(\frac{1}{\rho i+\beta}+\ldots+\frac{1}{\rho(r-1)+\beta})} \\
&\leq \alpha(\alpha+1)e^{2(\frac{1}{\rho i+\beta}+\ldots+\frac{1}{\rho(r-1)+\beta})}, \text{ since } \alpha \geq 1.
\end{aligned}$$

We know that for any integers $r > s > 1$, $\sum_{j=s}^{r}\frac{1}{\rho j+\beta} \leq \int_{s-1}^{r}\frac{1}{\rho x+\beta}dx = \frac{1}{\rho}\log(\frac{\rho r+\beta}{\rho(s-1)+\beta})$.

Thus, from (6.6),

$$
\begin{aligned}
& d_{TV}(W_n^{(Y)}, W_n^{(\overline{I})}) \\
& \leq \frac{3}{2n} \sum_{i=1}^{n} \sum_{r=i}^{n} \frac{1}{(\rho r+\beta)^2} \mathbb{E}((D_{r-1}^{(Y)}(i))^2) \\
& \leq \frac{3\alpha(\alpha+1)}{2n} \sum_{i=1}^{n} \sum_{r=i+1}^{n} \frac{1}{(\rho r+\beta)^2} \left(e^{2(\frac{1}{\rho i+\beta}+\ldots+\frac{1}{\rho(r-1)+\beta})} \right) \\
& \quad + \frac{3}{2n} \sum_{r=1}^{n} \frac{\alpha^2}{(\rho r+\beta)^2} \\
& \leq \frac{3\alpha(\alpha+1)}{2n} \sum_{i=2}^{n} \sum_{r=i+1}^{n} \frac{1}{(\rho r+\beta)^2} \left(e^{\frac{2}{\rho} \log(\frac{\rho r+\beta}{\rho(i-1)+\beta})} \right) \\
& \quad + \frac{3\alpha(\alpha+1)}{2n} \sum_{r=2}^{n} \frac{1}{(\rho r+\beta)^2} e^{\frac{2}{\rho}\log(\frac{\rho r+\beta}{\rho+\beta})+\frac{1}{\rho+\beta}} + \frac{3\alpha^2}{2n} \sum_{r=1}^{n} \frac{1}{(\rho r+\beta)^2}.
\end{aligned}
$$

From here it is easy to bound the order of the distance by comparison. To obtain exact bounds we bound each of these sums using integration, depending on the value of ρ. Firstly,

$$
\begin{aligned}
\sum_{r=1}^{n} \frac{1}{(\rho r+\beta)^2} & \leq \frac{1}{(\beta+\rho)^2} + \int_1^n \frac{1}{(\rho x+\beta)^2} dx \\
& = \frac{1}{(\beta+\rho)^2} + \frac{1}{\rho} \left(\frac{1}{\rho+\beta} - \frac{1}{\rho n+\beta} \right) \\
& \leq \frac{2\rho+\beta}{\rho(\rho+\beta)^2}.
\end{aligned}
$$

So

$$
\begin{aligned}
d_{TV}(W_n^{(Y)}, W_n^{(\overline{I})}) & \leq \frac{3\alpha(\alpha+1)}{2n} \sum_{i=2}^{n} \sum_{r=i+1}^{n} \frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho(i-1)+\beta)^{\frac{2}{\rho}}} \\
& \quad + \frac{3\alpha(\alpha+1)}{2n} \sum_{r=2}^{n} \frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho+\beta)^{\frac{2}{\rho}}} e^{\frac{1}{\rho+\beta}} + \frac{3\alpha^2}{2n} \left(\frac{2\rho+\beta}{\rho(\rho+\beta)^2} \right),
\end{aligned}
$$

as required.

For the first two terms we consider three cases, $\rho \in (1,2)$, $\rho = 2$ and $\rho > 2$. First suppose $\rho \in (1,2)$ so that $\frac{2}{\rho} - 1 > 0$. Then,

$$\begin{aligned}
\sum_{r=2}^{n} \frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho+\beta)^{\frac{2}{\rho}}} e^{\frac{1}{\rho+\beta}} &\leq \frac{e^{\frac{1}{\rho+\beta}}}{(\rho+\beta)^{\frac{2}{\rho}}} \int_{1}^{n} (\rho x+\beta)^{2(\frac{1}{\rho}-1)} dx \\
&= \frac{e^{\frac{1}{\rho+\beta}}}{\rho(\rho+\beta)^{\frac{2}{\rho}}} \int_{\rho+\beta}^{\beta+n\rho} u^{2(\frac{1}{\rho}-1)} du \\
&\leq \frac{e^{\frac{1}{\rho+\beta}}}{(2-\rho)(\rho+\beta)^{\frac{2}{\rho}}} (\rho n+\beta)^{\frac{2}{\rho}-1}.
\end{aligned}$$

Also if $\rho \in (1,2)$,

$$\begin{aligned}
\sum_{i=2}^{n} \sum_{r=i+1}^{n} \frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho(i-1)+\beta)^{\frac{2}{\rho}}} &\leq \sum_{i=2}^{n} \sum_{r=i+1}^{n} (\rho r+\beta)^{2(\frac{1}{\rho}-1)} (\rho i+\beta)^{-\frac{2}{\rho}} \\
&\leq \sum_{r=3}^{n} (\rho r+\beta)^{\frac{2}{\rho}-2} \left(1 + \int_{1}^{r-1} (\rho x+\beta)^{-\frac{2}{\rho}} dx\right) \\
&\leq \sum_{r=3}^{n} (\rho r+\beta)^{\frac{2}{\rho}-2} \left(1 + \frac{1}{\rho} \int_{\rho+\beta}^{\rho r+\beta} u^{-\frac{2}{\rho}} du\right),
\end{aligned}$$

We bound the two parts separately to get,

$$\sum_{r=3}^{n} (\rho r+\beta)^{\frac{2}{\rho}-2} \leq \int_{2}^{n} (\rho x+\beta)^{\frac{2}{\rho}-2} dx \leq \frac{(\rho n+\beta)^{\frac{2}{\rho}-1}}{2-\rho}, \tag{6.7}$$

as above. For the second part,

$$
\begin{aligned}
\frac{1}{\rho}\sum_{r=3}^{n}(\rho r+\beta)^{\frac{2}{\rho}-2}\int_{\rho+\beta}^{r\rho+\beta}x^{-\frac{2}{\rho}}dx &\leq \frac{1}{\rho}\sum_{r=3}^{n}(\rho r+\beta)^{\frac{2}{\rho}-2}\frac{(\rho+\beta)^{1-\frac{2}{\rho}}}{\frac{2}{\rho}-1} \qquad \left(\text{since } \frac{2}{\rho}-1>0\right)\\
&= \frac{(\rho+\beta)^{1-\frac{2}{\rho}}}{2-\rho}\sum_{r=3}^{n}(\rho r+\beta)^{\frac{2}{\rho}-2}\\
&\leq \frac{(\rho+\beta)^{1-\frac{2}{\rho}}}{(2-\rho)^2}(\rho n+\beta)^{\frac{2}{\rho}-1},\ \text{(by (6.7))}.
\end{aligned}
$$

Thus, for $\rho\in(1,2)$,

$$
\begin{aligned}
d_{TV}(W_n^{(Y)},W_n^{(\overline{I})}) &\leq \frac{3\alpha(\alpha+1)(\rho n+\beta)^{\frac{2}{\rho}-1}}{2n(2-\rho)}\left(1+\frac{(\beta+\rho)^{1-\frac{2}{\rho}}}{2-\rho}\right)\\
&\quad+\frac{3e^{\frac{1}{\rho+\beta}}\alpha(\alpha+1)}{2n(2-\rho)(\beta+\rho)}+\frac{3\alpha^2(2\rho+\beta)}{2n\rho(\rho+\beta)^2}\\
&= B_1(\alpha,\beta,\rho,n) \quad = O(n^{2(\frac{1}{\rho}-1)}) \text{ as } n\to\infty.
\end{aligned}
$$

If $\rho=2$,

$$
\begin{aligned}
\sum_{r=2}^{n}\frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho+\beta)^{\frac{2}{\rho}}}e^{\frac{1}{\rho+\beta}} &\leq \frac{e^{\frac{1}{2+\beta}}}{2+\beta}\int_1^n(2x+\beta)^{-1}dx\\
&= \frac{e^{\frac{1}{2+\beta}}}{2(2+\beta)}\int_{2+\beta}^{2n+\beta}u^{-1}du\\
&\leq \frac{e^{\frac{1}{2+\beta}}}{(2+\beta)}\log\frac{2n+\beta}{2+\beta}.
\end{aligned}
$$

As in the $\rho\in(1,2)$ case,

$$
\sum_{i=2}^{n}\sum_{r=i+1}^{n}\frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho(i-1)+\beta)^{\frac{2}{\rho}}} \leq \sum_{r=3}^{n}(\rho r+\beta)^{\frac{2}{\rho}-2}\left(1+\frac{1}{\rho}\int_{\rho+\beta}^{r\rho+\beta}u^{-\frac{2}{\rho}}du\right),
$$

so we bound the two parts separately as above. Firstly,

$$\begin{aligned}\sum_{r=3}^{n}(2r+\beta)^{-1} &\leq \int_2^n \frac{1}{2x+\beta}dx \\ &= \frac{1}{2}\log\frac{2n+\beta}{2+\beta},\end{aligned}$$

and secondly,

$$\begin{aligned}\frac{1}{2}\sum_{r=3}^{n}(2r+\beta)^{-1}\int_{2+\beta}^{2r+\beta}x^{-1}dx &\leq \frac{1}{4}\sum_{r=3}^{n}(2r+\beta)^{-1}\log\frac{2r+\beta}{2+\beta} \\ &\leq \frac{1}{4}\int_2^n (2x+\beta)^{-1}\log(2x+\beta)dx \\ &= \frac{1}{8}\int_{4+\beta}^{2n+\beta}\frac{\log u}{u}du \\ &\leq \frac{1}{16}(\log(2n+\beta))^2.\end{aligned}$$

Thus, for $\rho = 2$,

$$\begin{aligned}d_{TV}(W_n^{(Y)}, W_n^{(\overline{I})}) &\leq \frac{3\alpha(\alpha+1)}{2n}\left(\frac{1}{2}\log\frac{2n+\beta}{2+\beta}+\frac{1}{16}(\log(2n+\beta))^2\right) \\ &\quad +\frac{3\alpha(\alpha+1)e^{\frac{1}{2+\beta}}}{2(2+\beta)n}\log\frac{2n+\beta}{2+\beta}+\frac{3\alpha^2(4+\beta)}{4(2+\beta)^2 n} \\ &< \frac{3\alpha(\alpha+1)}{2n}\left(3\log\frac{2n+\beta}{2+\beta}+\frac{1}{16}(\log(2n+\beta))^2\right)+\frac{3\alpha^2(4+\beta)}{4(2+\beta)^2 n} \\ &= B_2(\alpha,\beta,\rho,n) \quad = O\left(\frac{(\log n)^2}{n}\right) \text{ as } n\to\infty.\end{aligned}$$

If $\rho > 2$,

$$\begin{aligned}
\sum_{r=2}^{n} \frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho+\beta)^{\frac{2}{\rho}}} e^{\frac{1}{\rho+\beta}} &\leq \frac{e^{\frac{1}{\rho+\beta}}}{(\rho+\beta)^{\frac{2}{\rho}}} \int_1^n (\rho x+\beta)^{2(\frac{1}{\rho}-1)} dx \\
&= \frac{e^{\frac{1}{\rho+\beta}}}{\rho(\rho+\beta)^{\frac{2}{\rho}}} \int_{\rho+\beta}^{\rho n+\beta} u^{2(\frac{1}{\rho}-1)} du \\
&\leq \frac{e^{\frac{1}{\rho+\beta}}}{(\rho-2)(\rho+\beta)^{\frac{2}{\rho}}} (\beta+\rho)^{\frac{2}{\rho}-1} \\
&= \frac{e^{\frac{1}{\rho+\beta}}}{(\rho-2)(\rho+\beta)}.
\end{aligned}$$

As in the $\rho \in (1,2)$ case,

$$\sum_{i=2}^{n} \sum_{r=i+1}^{n} \frac{(\rho r+\beta)^{\frac{2}{\rho}-2}}{(\rho(i-1)+\beta)^{\frac{2}{\rho}}} \leq \sum_{r=3}^{n} (\rho r+\beta)^{\frac{2}{\rho}-2} \left(1+\frac{1}{\rho}\int_{\rho+\beta}^{\rho r+\beta} u^{-\frac{2}{\rho}} du\right),$$

so we bound the two parts separately as above. Firstly,

$$\sum_{r=3}^{n} (\rho r+\beta)^{\frac{2}{\rho}-2} \leq \int_2^n (\rho x+\beta)^{\frac{2}{\rho}-2} dx \leq \frac{(\rho+\beta)^{\frac{2}{\rho}-1}}{\rho-2},$$

as above. For the second part,

$$\begin{aligned}
\frac{1}{\rho}\sum_{r=3}^{n} (\rho r+\beta)^{\frac{2}{\rho}-2} \int_{\rho+\beta}^{\rho r+\beta} x^{-\frac{2}{\rho}} dx &\leq \frac{1}{\rho}\sum_{r=3}^{n} (\rho r+\beta)^{\frac{2}{\rho}-2} \frac{(\rho r+\beta)^{1-\frac{2}{\rho}}}{1-\frac{2}{\rho}} \quad \left(\text{since } \frac{2}{\rho}-1<0\right) \\
&= \frac{1}{\rho-2}\sum_{r=3}^{n} (\rho r+\beta)^{-1} \\
&\leq \frac{1}{\rho(\rho-2)} \log \frac{\rho n+\beta}{2\rho+\beta},
\end{aligned}$$

by the previous result. Thus, for $\rho > 2$,

$$
\begin{aligned}
d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) &\leq \frac{3\alpha(\alpha+1)}{2(\rho-2)n}\left((\beta+\rho)^{\frac{2}{\rho}-1} + \frac{1}{\rho}\log\frac{\rho n+\beta}{2\rho+\beta}\right) \\
&\quad + \frac{3\alpha(\alpha+1)}{2n}\left(\frac{e^{\frac{1}{\beta+\rho}}}{(\rho-2)(\beta+\rho)}\right) \\
&\quad + \frac{3\alpha^2(2\rho+\beta)}{2\rho(\beta+\rho)^2 n} \\
&= B_3(\alpha,\beta,\rho,n) \quad = O\left(\frac{\log n}{n}\right) \text{ as } n \to \infty.
\end{aligned}
$$

$\square$

This chapter has provided the means for us to bound the total variation distance between any pair of models in our scheme using the bounds given and applying the triangle inequality. Later chapters bound the degree distributions of models in our scheme using a variety of methods, each of which applies to a subset of the models of interest. The bounds in this chapter extend the results to the remaining models.

We shall see this method illustrated in Chapter 7 where we use the negative binomial approximation method already seen in Section 3.6 to bound the total variation distance between the degree distribution of Model Y and the μ_ρ^α-distribution. Theorem 7.3 uses the results of the current chapter to extend these bounds to Model BA.

Chapter 7

Bounds for convergence using negative binomial approximation

Having defined a sequence of continuous-time and discrete-time preferential attachment models, in Chapter 6 we bounded the total variation distances between the degree distributions in our scheme. We now follow the method applied to Yule's model for evolution in Section 3.6 to bound the distance between the degree distribution of Model Y and the μ_ρ^α-distribution.

In the generalisation of Yule's model for evolution, bounds for the distance between the degree distribution and the μ_ρ^α-distribution hinged on the result of Theorem 3.14, that at the stopping-time given by the time of addition of the n^{th} vertex to the graph, the age of a uniformly chosen vertex takes precisely the *Exponential*(ρ) distribution. Similarly, adding new vertices at the prescribed deterministic sequence of times for Model Y we find that the age of a uniformly chosen vertex at time $s_1 + s_2 + ... + s_n$, (where $s_i := \frac{1}{\rho i}$, as before), is close in distribution to the *Exponential*(ρ) distribution as $n \to \infty$.

Theorem 7.1. *Let $n \in \mathbb{N}$, $\beta \geq 0$ and $\rho > 0$. Let $U \sim Exponential(\rho)$, $I_n \sim Uniform\{1, 2, ..., n\}$ and let $U_n = s_{I_n} + s_{I_n+1} + ... + s_n$ for the $s_i's$ defined above. Then,*

$$d_K(U, U_n) \leq \frac{\rho + \beta}{\rho n + \beta}.$$

Proof. We bound the difference between $\mathbb{P}(U \leq x)$ and $\mathbb{P}(U_n \leq x)$ for all values of x.

For $x \leq s_n$,

$\mathbb{P}(s_{I_n} + ... + s_n \leq x) = 0, \quad \mathbb{P}(U \leq x) = 1 - e^{-\rho x} \leq 1 - e^{-\frac{\rho}{\rho n+\beta}} \leq \frac{\rho}{\rho n+\beta}, \quad$ so

$$\sup_{x \leq s_n} |\mathbb{P}(U \leq x) - \mathbb{P}(s_{I_n} + ... + s_n \leq x)| < \frac{\rho}{\rho n + \beta}.$$

For $x > s_1 + s_2 + ... + s_n = \sum_{i=1}^{n} \frac{1}{\rho i+\beta}$**,** so, since

$$\sum_{i=1}^{n} \frac{1}{\rho i + \beta} \quad > \quad \frac{1}{\rho + \beta} + \int_1^n \frac{1}{\rho x + \beta} dx = \frac{1}{\rho + \beta} + \frac{1}{\rho} \log \frac{\rho n + \beta}{\rho + \beta},$$

it follows that

$$\begin{aligned} \mathbb{P}(U \leq x) &= 1 - e^{-\rho x} \\ &\geq 1 - e^{-(\frac{1}{\rho+\beta} + \frac{1}{\rho} \log \frac{\rho n+\beta}{\rho+\beta})\rho} \\ &= 1 - \frac{(\rho + \beta) e^{-\frac{\rho}{\rho+\beta}}}{\rho n + \beta}. \end{aligned}$$

Also,

$$\mathbb{P}(s_{I_n} + ... + s_n \leq x) = 1,$$

so

$$\begin{aligned} \sup_{x > s_1 + s_2 + ... + s_n} |\mathbb{P}(U \leq x) - \mathbb{P}(s_{I_n} + ... + s_n \leq x)| &\leq \frac{(\rho + \beta) e^{-\frac{\rho}{\rho+\beta}}}{\rho n + \beta} \\ &\leq \frac{\rho + \beta}{\rho n + \beta}. \end{aligned}$$

For $i \geq 2$**,** $x \in (s_{i+1} + ... + s_n, s_i + ... + s_n] =: \mathcal{R}_i$,

$$\begin{aligned} x &\in \left(\frac{1}{\rho(i+1)+\beta} + ... + \frac{1}{\rho n + \beta}, \frac{1}{\rho i + \beta} + ... + \frac{1}{\rho n + \beta} \right] \\ &\subset \left(\int_{i+1}^{n+1} \frac{1}{\rho x + \beta} dx, \int_{i-1}^{n} \frac{1}{\rho x + \beta} dx \right] \\ &= \left[\frac{1}{\rho} \log \frac{\rho(n+1)+\beta}{\rho(i+1)+\beta}, \frac{1}{\rho} \log \frac{\rho n + \beta}{\rho(i-1)+\beta} \right]. \end{aligned}$$

So $x = \frac{1}{\rho} \log y$ for some $y \in [\frac{\rho(n+1)+\beta}{\rho(i+1)+\beta}, \frac{\rho n+\beta}{\rho(i-1)+\beta}]$ and $\mathbb{P}(U \leq x) = 1 - e^{-\log y} = 1 - \frac{1}{y}$.

Also, $\mathbb{P}(s_{I_n} + ... + s_n \leq x) = \mathbb{P}(I_n > i) = 1 - \frac{i}{n}$, so,

$$\begin{aligned} \sup_{x \in \mathcal{R}_i} |\mathbb{P}(U \leq x) - \mathbb{P}(s_{I_n} + ... + s_n \leq x)| &= \sup_{x \in \mathcal{R}_i} \left| \left(1 - \frac{1}{y}\right) - \left(1 - \frac{i}{n}\right) \right| \\ &\leq \sup_{y \in [\frac{\rho(n+1)+\beta}{\rho(i+1)+\beta}, \frac{\rho n+\beta}{\rho(i-1)+\beta}]} \left| \frac{1}{y} - \frac{i}{n} \right| \\ &= \max \left(\frac{|\beta i + (\rho - \beta) n|}{(\rho n + \beta) n}, \frac{(\rho+\beta)(n-i)}{(\rho(n+1)+\beta)n} \right) \\ &\leq \frac{\rho + \beta}{\rho n + \beta}. \end{aligned}$$

For $x \in (s_2 + ... + s_n, s_1 + ... + s_n] =: \mathcal{R}_1$, we have

$$\begin{aligned} x &\in \left(\frac{1}{2\rho+\beta} + ... + \frac{1}{\rho n + \beta}, \frac{1}{\rho + \beta} + ... + \frac{1}{\rho n + \beta} \right] \\ &\subset \left[\int_{2}^{n+1} \frac{1}{\rho x + \beta} dx, \frac{1}{\rho+\beta} + \int_{1}^{n} \frac{1}{\rho x + \beta} dx \right] \\ &= \left[\frac{1}{\rho} \log \frac{\rho(n+1)+\beta}{2\rho+\beta}, \frac{1}{\rho+\beta} + \frac{1}{\rho} \log \frac{\rho n + \beta}{\rho+\beta} \right]. \end{aligned}$$

Now, in this case, $\mathbb{P}(U_n \le x) = \mathbb{P}(I_n > 1) = 1 - \frac{1}{n}$, and, by the above,

$$\begin{aligned}
\mathbb{P}(U < x) &\in [1 - e^{-\log\frac{\rho(n+1)+\beta}{2\rho+\beta}}, 1 - e^{\frac{\rho}{\rho+\beta}+\log\frac{\rho n+\beta}{\rho+\beta}}] \\
&= \left[1 - \frac{2\rho+\beta}{\rho(n+1)+\beta}, 1 - \frac{(\rho+\beta)e^{-\frac{\rho}{\rho+\beta}}}{\rho n+\beta}\right].
\end{aligned}$$

So,

$$\begin{aligned}
&\sup_{x \in \mathcal{R}_1} |\mathbb{P}(U \le x) - \mathbb{P}(s_{I_n} + ... + s_n \le x)| \\
&= \max\left(\left|\frac{1}{n} - \frac{2\rho+\beta}{\rho(n+1)+\beta}\right|, \left|\frac{1}{n} - \frac{(\rho+\beta)e^{-\frac{\rho}{\rho+\beta}}}{\rho n+\beta}\right|\right) \\
&< \frac{\rho+\beta}{\rho n+\beta}.
\end{aligned}$$

Combining these results gives the required Kolmogorov distance,

$$d_K(U, U_n) \le \frac{\rho+\beta}{\rho n+\beta}.$$

$\square$

This result explains why the degree distribution of Model Y has approximately a μ_ρ^α-distribution for fixed ρ as $n \to \infty$. We have already seen in Theorem 3.20 that for a function $\overline{K}$, if $\alpha \ge 1$ then the total variation distance between a $Negbinom(\alpha; p)$ distribution and a $Negbinom(\alpha; q)$ distribution can be bounded by

$$\overline{K}(\alpha) \max\left(\left|\frac{p}{q} - 1\right|, \left|\frac{q}{p} - 1\right|\right).$$

We use this result to bound the distance between the degree distribution in Model Y and the μ_ρ^α-distribution. We follow the strategy employed in Theorem 3.21 for Yule's model for evolution. We apply Theorem 3.18, noting by Proposition 3.6 that if $U \sim Exponential(\rho)$ then the mixture distribution, $Negbinom(\alpha, e^{-U})$ is the μ_ρ^α-distribution. Moreover, by Corollary 3.13 the degree distribution of Model Y takes a $Negbinom(\alpha; e^{-U_n})$ mixture distribution. Constructing U_n and U, as in Theorem 7.1, allows us to obtain bounds for $\mathbb{E}\left(\max\left(\left|e^{U_n-U}-1\right|, \left|e^{U-U_n}-1\right|\right)\right)$ that are small for large n.

Theorem 7.2. *Let $\alpha \geq 1$, $\beta \geq 0$ and $\rho > 1$. For $n \in \mathbb{N}$, as usual, let $D_n^{(Y)}(i)$ represent $\alpha+$ in-degree of vertex i in Model Y at time $s_1 + s_2 + ... + s_n$ where $s_i := \frac{1}{\rho i+\beta}$ as above. Let $I_n \sim Uniform\{1, 2, ..., n\}$ with $I_n \perp \{D_n^{(Y)}(i)\}_i$ and let $W_n^{(Y)} = D_n^{(Y)}(I_n) - \alpha$. Then, if $\mathcal{L}(V) = \mu_\alpha^\rho$,*

$$d_{TV}(W_n^{(Y)}, V) \leq A(\alpha, \beta, \rho, n),$$

where

$$A(\alpha, \beta, \rho, n) = \frac{1}{n} + \frac{\overline{K}(\alpha)}{n}\left(\frac{\rho+\beta}{\rho}\log(n-1)\right), \tag{7.1}$$

for $\overline{K}(\alpha)$ the constant determined by α, defined in Theorem 3.20.

Proof. Let $n \in \mathbb{N}$. Without loss of generality, by Proposition 3.6 we can assume that the random variable V satisfies $V \sim Negbinom(\alpha, e^{-U})$ for $U \sim Exponential(\rho)$.

Assume that V and $W_n^{(Y)}$ are coupled so that $d_{TV}(V, W_n^{(Y)}) = \mathbb{P}(V \neq W_n^{(Y)})$. Thus,

$$\begin{aligned} d_{TV}(V, W_n^{(Y)}) &= \frac{1}{n}\mathbb{P}(V \neq W_n^{(Y)}|I_n = 1) + \frac{n-1}{n}\mathbb{P}(V \neq W_n^{(Y)}|I_n \neq 1) \\ &\leq \frac{1}{n} + \frac{n-1}{n}\mathbb{P}(V \neq W_n^{(Y)}|I_n \neq 1) \\ &\leq \frac{1}{n} + \frac{n-1}{n}d_{TV}(V, \overline{W_n^{(Y)}}), \end{aligned} \tag{7.2}$$

where $\overline{W_n^{(Y)}} = D_n^{(Y)}(\overline{I_n}) - \alpha$ and $\overline{I_n} \sim Uniform\{2, 3, ..., n\}$.

Let $U_n = s_{I_n} + s_{I_n+1} + ... + s_n$. Then U_n gives the age of the randomly chosen vertex, I_n.

Thus by Corollary 3.13, for any $k \in \mathbb{N}$,

$$\mathcal{L}(W_n^{(Y)}) = Negbinom(\alpha, e^{-U_n}).$$

It follows from Theorem 3.20 that

$$d_{TV}(\overline{W_n^{(Y)}}, V) \quad \leq \quad \overline{K}(\alpha)\mathbb{E}\left(\max\left(\left|e^{U_n-U}-1\right|, \left|e^{U-U_n}-1\right|\right)|I_n \neq 1\right). \tag{7.3}$$

We use the following coupling construction to define U_n based on U in such a way that $\mathbb{E}\left(\max\left(\left|e^{U_n-U}-1\right|, \left|e^{U-U_n}-1\right|\right)|I_n \neq 1\right) \to 0$ as $n \to \infty$. Define $I_n := \lceil ne^{-\rho U} \rceil$. Then, for $k = 2, 3, ..., n$,

$$\begin{aligned}
\mathbb{P}(I_n = k) &= \mathbb{P}(ne^{-\rho U} \in (k-1, k]) \\
&= \mathbb{P}\left(U \in \left[-\frac{1}{\rho}\log\left(\frac{k}{n}\right), -\frac{1}{\rho}\log\left(\frac{k-1}{n}\right)\right)\right) \\
&= e^{\log(\frac{k}{n})} - e^{\log(\frac{k-1}{n})} = \frac{1}{n},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(I_n = 1) &= \mathbb{P}(ne^{-\rho U} \in (0, 1]) \\
&= \mathbb{P}\left(U \in \left[-\frac{1}{\rho}\log\left(\frac{k}{n}\right), \infty\right)\right) \\
&= e^{\log(\frac{1}{n})} = \frac{1}{n},
\end{aligned}$$

so $I_n \sim Uniform\{1, 2, ..., n\}$ as required.

Moreover, for $i \geq 2$, if $I_n = i$ then,

$$\begin{aligned}
U_n &= \frac{1}{\rho i+\beta} + ... + \frac{1}{\rho n+\beta} \\
&\in \left[\int_i^{n+1} \frac{1}{\rho x+\beta}dx, \int_{i-1}^{n} \frac{1}{\rho x+\beta}\right] \\
&= \left[\frac{1}{\rho}\log\left(\frac{\rho(n+1)+\beta}{\rho i+\beta}\right), \frac{1}{\rho}\log\left(\frac{\rho n+\beta}{\rho(i-1)+\beta}\right)\right],
\end{aligned}$$

and, by the coupling of U with I_n, when $I_n = i$,

$$U \in \left[\frac{1}{\rho}\log\left(\frac{n}{i}\right), \frac{1}{\rho}\log\left(\frac{n}{i-1}\right)\right].$$

Thus if $I_n = i \geq 2$, both $U, U_n \in [\frac{1}{\rho}\log\frac{\rho n}{\rho i+\beta}, \frac{1}{\rho}\log\frac{n}{(i-1)}]$, so

$$\begin{aligned}
|U - U_n| &\leq \frac{1}{\rho}\left(\log\left(\frac{n}{i-1}\right) - \log\left(\frac{\rho n}{\rho i+\beta}\right)\right) \\
&= \frac{1}{\rho}\log\left(\frac{\rho i+\beta}{\rho(i-1)}\right).
\end{aligned}$$

It follows that,

$$\begin{aligned}
\{e^{U-U_n}, e^{U_n-U}\} &\subset \left[\left(\frac{(\rho i+\beta)}{\rho(i-1)}\right)^{-\frac{1}{\rho}}, \left(\frac{(\rho i+\beta)}{\rho(i-1)}\right)^{\frac{1}{\rho}}\right] \\
&\subset \left[1 - \frac{\rho+\beta}{\rho i+\beta}, 1 + \frac{\rho+\beta}{\rho(i-1)}\right]
\end{aligned}$$

since $\rho > 1$, so $\max\left(\left|e^{U_n-U} - 1\right|, \left|e^{U-U_n} - 1\right|\right) \leq \frac{\rho+\beta}{(I_n-1)\rho}$ for $I_n \geq 2$.

It follows that,

$$\begin{aligned}
\mathbb{E}\left(\max\left(\left|e^{U_n-U}-1\right|,\left|e^{U-U_n}-1\right|\right)|I_n\neq 1\right) &\leq \frac{1}{n-1}\sum_{i=2}^{n}\frac{\rho+\beta}{\rho(i-1)} \\
&\leq \frac{\rho+\beta}{\rho(n-1)}\log(n-1).
\end{aligned}$$

Substituting into (7.2), using (7.3), gives

$$\begin{aligned}
d_{TV}(W_n^{(Y)},V) &\leq \frac{1}{n}+\frac{(n-1)\overline{K}(\alpha)}{n}\mathbb{E}\left(\max\left(\left|e^{U_n-U}-1\right|,\left|e^{U-U_n}-1\right|\right)|I_n\neq 1\right) \\
&\leq \frac{1}{n}+\frac{\overline{K}(\alpha)}{n}\left(\frac{\rho+\beta}{\rho}\log(n-1)\right) \\
&= A(\alpha,\beta,\rho,n),
\end{aligned}$$

as required. □

The coupling and bounds of Chapter 6 allow us to extend these results to other models within our scheme. To illustrate this point, and because Model BA is of particular interest to us, the next theorem gives the extension Model BA.

Theorem 7.3. *The total variation distance between a random variable V taking the power law μ_ρ^α-distribution, with $\rho = \frac{m+\alpha}{m}$, and the degree distribution of our Barabási-Albert Model BA with $m_0 \geq m \geq 1$, $\alpha \geq 1$, satisfying $(m_0-1)\alpha > m$ takes the following upper bounds depending on the value of α. If $\alpha < m$,*

$$\begin{aligned}
&d_{TV}(W_n^{(BA)},V) \\
&\leq \frac{m_0-1}{n+m_0+1}+A\left(\alpha,\frac{(m_0-1)\alpha-m}{m},\frac{m+\alpha}{m},n\right) \\
&\quad +B_1\left(\alpha,\frac{(m_0-1)\alpha-m}{m},\frac{m+\alpha}{m},n\right) \\
&= O(n^{-\frac{2\alpha}{\alpha+m}}) \text{ as } n\to\infty,
\end{aligned}$$

if $\alpha = m$,

$$
\begin{aligned}
& d_{TV}(W_n^{(BA)}, V) \\
\leq & \ \frac{m_0 - 1}{n + m_0 + 1} + A\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
& +B_2\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
= & \ O\left(\frac{(\log n)^2}{n}\right) \ \textit{as } n \to \infty,
\end{aligned}
$$

and if $\alpha > m$

$$
\begin{aligned}
& d_{TV}(W_n^{(BA)}, V) \\
\leq & \ \frac{m_0 - 1}{n + m_0 + 1} + A\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
& +B_3\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
= & \ O\left(\frac{\log n}{n}\right) \ \textit{as } n \to \infty,
\end{aligned}
$$

where $A(\alpha, \beta, \rho, n)$, $B_1(\alpha, \beta, \rho, n)$, $B_2(\alpha, \beta, \rho, n)$ and $B_3(\alpha, \beta, \rho, n)$are defined as in (7.1), (6.1), (6.2) and (6.3) respectively.

Proof. There is a correspondence between the two sets of parameters for the various models, $\beta = \frac{(m_0-1)\alpha - m}{m}$ and $\rho = \frac{m+\alpha}{m}$. Throughout this proof we use Model I with the same parameters as Model BA and use parameters $\alpha = \alpha$, $\beta = \frac{(m_0-1)\alpha - m}{m}$ and $\rho = \frac{m+\alpha}{m}$ for Model $\bar{\mathrm{I}}$ and Model Y. We apply the triangle inequality with these parameter substitutions,

$$
\begin{aligned}
d_{TV}(W_n^{(BA)}, V) \leq & \ d_{TV}(W_n^{(BA)}, W_n^{(I)}) + d_{TV}(W_n^{(I)}, W_n^{(\bar{I})}) \\
& +d_{TV}(W_n^{(\bar{I})}, W_n^{(Y)}) + d_{TV}(W_n^{(Y)}, V). \qquad (7.4)
\end{aligned}
$$

Now let us consider the distance bounds one by one. Firstly, by Theorem 6.1

$$d_{TV}(W_n^{(BA)}, W_n^{(\bar{I})}) = 0.$$

Also, by Theorem 6.2

$$d_{TV}(W_n^{(\bar{I})}, W_n^{(I)}) \leq \frac{m_0 - 1}{n + m_0 - 1}.$$

By Theorem 7.2

$$d_{TV}(W_n^{(Y)}, V) \quad \leq \quad A\left(\alpha, \frac{(m_0 - 1)\alpha - m}{m}, \frac{m + \alpha}{m}, n\right).$$

Finally we look at the bounds of Theorem 6.4. These bounds depend of the value of ρ, or equivalently the relative values of α and m. If $\rho \in (1, 2) \iff \alpha < m$,

$$d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) \quad \leq \quad B_1\left(\alpha, \frac{(m_0 - 1)\alpha - m}{m}, \frac{m + \alpha}{m}, n\right).$$

If $\rho = 2 \iff \alpha = m$,

$$d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) \quad < \quad B_2\left(\alpha, \frac{(m_0 - 1)\alpha - m}{m}, \frac{m + \alpha}{m}, n\right).$$

Finally, if $\rho > 2 \iff \alpha > m$,

$$d_{TV}(W_n^{(Y)}, W_n^{(\bar{I})}) \quad \leq \quad B_3\left(\alpha, \frac{(m_0 - 1)\alpha - m}{m}, \frac{m + \alpha}{m}, n\right).$$

Substituting in these values to (7.4) gives, if $\alpha < m$,

$$\begin{aligned}
&d_{TV}(W^{(BA)}, V) \\
&\leq \frac{m_0 - 1}{m_0 + n - 1} + A\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
&\quad + B_1\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
&= O\left(n^{\frac{-2\alpha}{\alpha+m}}\right) \text{ as } n \to \infty,
\end{aligned}$$

if $\alpha = m$,

$$\begin{aligned}
&d_{TV}(W^{(BA)}, V) \\
&\leq \frac{m_0 - 1}{m_0 + n - 1} + A\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
&\quad + B_2\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
&= O\left(\frac{(\log n)^2}{n}\right) \text{ as } n \to \infty,
\end{aligned}$$

and if $\alpha > m$

$$\begin{aligned}
&d_{TV}(W^{(BA)}, V) \\
&\leq \frac{m_0 - 1}{m_0 + n - 1} + A\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
&\quad + B_3\left(\alpha, \frac{(m_0-1)\alpha - m}{m}, \frac{m+\alpha}{m}, n\right) \\
&= O\left(\frac{\log n}{n}\right) \text{ as } n \to \infty,
\end{aligned}$$

as required. $\square$

Chapter 8

Bounds for convergence using recurrence equations

In Chapter 7 we found total variation distance bounds between the degree distributions of each class of evolving random graph models and the μ_ρ^α-distribution whenever $\alpha \geq 1$. From our total variation distance bounds we are able to deduce bounds for

$$\left|\frac{\mathbb{E}N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k)\right|,$$

where $N_n^{(BA)}(k)$ is the number of in-degree k vertices at step n in Model BA, so that $\frac{\mathbb{E}(N_n^{(BA)}(k))}{n+m_0} = \mathbb{P}(W_n^{(BA)} = k)$. As we saw in Section 1.2.3, this quantity is of special interest when estimating the exponent of a power law from log-log plots.

We now follow the example of Szymański [53] and bound this directly for Model BA using recurrence equations. The recurrence equation method gives us better bounds than those found in Chapter 7 for $\left|\frac{\mathbb{E}N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k)\right|$. The method that provides better bounds for the total variation distance is determined by the model of interest and the choice of parameters. By giving both results we allow the better bounds to be chosen for any given model and parameters.

8.1 The recurrence equations

Define the random variables $(N_n^{(BA)}(k))_{(n,k)}$ so that $N_n^{(BA)}(k)$ represents the number of vertices of in-degree k in Model BA at step n. Here we fix α, m and m_0. Recall the definition of Model BA, found in B.1. We get the following recurrence equations and initial conditions.

The random graph begins with m_0 vertices of in-degree 0, so $N_0^{(BA)}(0) = m_0$ and $N_0^{(BA)}(i) = 0$ for all $i \in \mathbb{N}$. Also, each vertex gains at most one in-edge at each step, so $N_n^{(BA)}(n+i) = 0$ for all $n \in \mathbb{Z}_0^+$, $i \in \mathbb{N}$.

Now at step n a new vertex is added with in-degree 0. For each vertex of in-degree 0 already in the graph there is a probability of $\left(1 - \frac{m\alpha}{(n-1)(m+\alpha)+m_0\alpha}\right)$ that the vertex remains of in-degree 0, the vertex not having been chosen for a new edge. Thus,

$$\mathbb{E}(N_n^{(BA)}(0)) = 1 + \left(1 - \frac{m\alpha}{(n-1)(m+\alpha)+m_0\alpha}\right)\mathbb{E}(N_{n-1}^{(BA)}(0)).$$

For $n \geq k > 0$, if a vertex has in-degree $k-1$ at step $n-1$ there is a probability of $\frac{m(k+\alpha-1)}{(n-1)(m+\alpha)+m_0\alpha}$ that it has in-degree k at step n. If a vertex has in-degree k at step $n-1$, there is a probability of $\left(1 - \frac{m(k+\alpha)}{(n-1)(m+\alpha)+m_0\alpha}\right)$ that it still has in-degree k at step n. Therefore,

$$\begin{aligned}\mathbb{E}(N_n^{(BA)}(k)) &= \frac{m(k+\alpha-1)}{(n-1)(m+\alpha)+m_0\alpha}\mathbb{E}(N_{n-1}^{(BA)}(k-1)) \\ &\quad + \left(1 - \frac{m(k+\alpha)}{(n-1)(m+\alpha)+m_0\alpha}\right)\mathbb{E}(N_{n-1}^{(BA)}(k)). \qquad (8.1)\end{aligned}$$

8.2 Bounding the error term

The following theorem proves that $\left|\frac{\mathbb{E}N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k)\right| \leq \frac{m_0(m+1)\alpha(\alpha+1)}{(m+\alpha)(n+m_0)(k+\alpha)} + \frac{m_0 m \mu_\rho^\alpha(k)}{(m+\alpha)(n+m_0)}$. We set up recurrence equations for the error and deduce the bound by an inductive argument.

Theorem 8.1. *For* $\rho = \frac{m+\alpha}{m}$,

$$\Big|\frac{\mathbb{E}N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k)\Big| \le \frac{m_0(m+1)\alpha(\alpha+1)}{(m+\alpha)(n+m_0)(k+\alpha)} + \frac{m_0 m}{(m+\alpha)(n+m_0)}\mu_\rho^\alpha(k).$$

Proof. Let $\rho = \frac{m+\alpha}{m}$. For $n, k \ge 0$, we define the sequence of errors, $(\varepsilon_{n,k})_{(n,k)}$, by

$$\mathbb{E}N_n^{(BA)}(k) = \frac{n(m+\alpha)+m_0\alpha}{m+\alpha}\mu_\rho^\alpha(k) + \varepsilon_{n,k}.$$

It follows that,

$$\frac{\mathbb{E}N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k) = \frac{n(m+\alpha)+m_0\alpha}{(n+m_0)(m+\alpha)}\mu_\rho^\alpha(k) + \frac{\varepsilon_{n,k}}{n+m_0} - \mu_\rho^\alpha(k),$$

so that

$$\begin{aligned}
\Big|\frac{\mathbb{E}N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k)\Big| &\le \Big|\frac{\varepsilon_{n,k}}{n+m_0}\Big| + \Big|\frac{n(m+\alpha)+m_0\alpha}{(m+\alpha)(n+m_0)} - 1\Big|\mu_\rho^\alpha(k)\\
&= \frac{|\varepsilon_{n,k}|}{n+m_0} + \frac{m_0 m}{(m+\alpha)(n+m_0)}\mu_\rho^\alpha(k).
\end{aligned}$$

We need only prove that $|\varepsilon_{n,k}| \le \frac{m_0(m+1)\alpha(\alpha+1)}{(m+\alpha)(k+\alpha)}$ for all $n, k \ge 0$.

From the recurrence equations (8.1) for $(\mathbb{E}(N_n^{(BA)}(k)))_{(n,k)}$ we deduce the following recurrence equations for $(\varepsilon_{n,k})_{(n,k)}$. For $k = 0$,

$$\begin{aligned}
\varepsilon_{n,0} &= \mathbb{E}N_n^{(BA)}(0) - \frac{n(m+\alpha)+m_0\alpha}{m+\alpha}\mu_\rho^\alpha(0)\\
&= 1 + \left(1 - \frac{m\alpha}{(n-1)(m+\alpha)+m_0\alpha}\right)\mathbb{E}N_{n-1}^{(BA)}(0) - \frac{n(m+\alpha)+m_0\alpha}{m+\alpha}\mu_\rho^\alpha(0)\\
&= 1 + \left(1 - \frac{m\alpha}{(n-1)(m+\alpha)+m_0\alpha}\right)\left(\frac{(n-1)(m+\alpha)+m_0\alpha}{m+\alpha}\mu_\rho^\alpha(0) + \varepsilon_{n-1,0}\right)\\
&\quad -\frac{n(m+\alpha)+m_0\alpha}{m+\alpha}\mu_\rho^\alpha(0)\\
&= \left(1 - \frac{\alpha m}{(n-1)(m+\alpha)+m_0\alpha}\right)\varepsilon_{n-1,0}, \quad \text{since } \mu_\rho^\alpha(0) = \frac{\rho}{\alpha+\rho} = \frac{m+\alpha}{\alpha m+\alpha+m}.
\end{aligned}$$

Thus, $|\varepsilon_{n,0}| \leq |\varepsilon_{n-1,0}| \leq |\varepsilon_{0,0}|$ for all $n \in \mathbb{N}$. Similarly, for $n \geq k > 0$, since $\frac{\mu_\rho^\alpha(k)}{\mu_\rho^\alpha(k-1)} = \frac{k+\alpha}{k+\alpha+\rho}$,

$$\begin{aligned}
\varepsilon_{n,k} &= \mathbb{E}N_n^{(BA)}(k) - \frac{n(m+\alpha)+m_0\alpha}{m+\alpha}\mu_\rho^\alpha(k) \\
&= \frac{m(k+\alpha-1)}{(n-1)(m+\alpha)+m_0\alpha}\mathbb{E}(N_{n-1}^{(BA)}(k-1)) \\
&\quad + \left(1 - \frac{m(k+\alpha)}{(n-1)(m+\alpha)+m_0\alpha}\right)\mathbb{E}(N_{n-1}^{(BA)}(k)) - \frac{n(m+\alpha)+m_0\alpha}{m+\alpha}\mu_\rho^\alpha(k) \\
&= \frac{m(k+\alpha-1)}{(n-1)(m+\alpha)+m_0\alpha}\varepsilon_{n-1,k-1} + \left(1 - \frac{m(k+\alpha)}{(n-1)(m+\alpha)+m_0\alpha}\right)\varepsilon_{n-1,k}.
\end{aligned}$$

Finally, the definition of Model BA gives us $N_0^{(BA)}(0) = m_0$ and $N_0^{(BA)}(k) = 0$ for $k > 0$. It follows that we have boundary conditions $\varepsilon_{0,0} = m_0\left(1 - \frac{\alpha}{m+\alpha}\mu_\rho^\alpha(0)\right) = \frac{m_0 m(\alpha+1)}{m+\alpha+m\alpha}$ and for $k > 0$, $\varepsilon_{0,k} = -\frac{m_0\alpha}{m+\alpha}\mu_\rho^\alpha(k)$.

For $k, n \in \mathbb{N}$, define

$$\sigma_{n,k} := (k+\alpha)\varepsilon_{n,k}.$$

We already know that for all $n \in \mathbb{N}$,

$$|\sigma_{n,0}| \leq \alpha|\varepsilon_{0,0}| = \frac{\alpha m_0 m(\alpha+1)}{m+\alpha+m\alpha} < \frac{m_0(m+1)\alpha(\alpha+1)}{m+\alpha}.$$

It remains to prove the same bound for $n \geq k > 0$. We obtain the following recurrence equations for $\sigma_{n,k}$ when $n \geq k > 0$,

$$\sigma_{n,k} = \frac{m(k+\alpha)}{(n-1)(m+\alpha)+m_0\alpha}\sigma_{n-1,k-1} + \left(1 - \frac{m(k+\alpha)}{(n-1)(m+\alpha)+m_0\alpha}\right)\sigma_{n-1,k}.$$

Thus $|\sigma_{n,k}| \leq \max(|\sigma_{n-1,k-1}|, |\sigma_{n-1,k}|)$.

By double induction, it is sufficient to prove that for all $n, k \in \mathbb{N}$, $|\sigma_{n,0}| \leq \frac{m_0(m+1)\alpha(\alpha+1)}{m+\alpha}$

and $|\sigma_{0,k}| \le \frac{m_0(m+1)\alpha(\alpha+1)}{m+\alpha}$. We have already proved the former. For the latter we use the boundary conditions for $\varepsilon_{0,k}$,

$$\begin{aligned}
|\sigma_{0,k}| &= (k+\alpha)\frac{m_0\alpha}{m+\alpha}\mu_\rho^\alpha(k) \\
&= \frac{m_0\alpha(k+\alpha)B(\alpha+k;\rho+1)}{(m+\alpha)B(\alpha;\rho)} \\
&= \frac{m_0\alpha\rho(\alpha+\rho-1)B(\alpha+k+1;\rho)}{(m+\alpha)(\rho-1)B(\alpha;\rho-1)} \text{ (by (A.4))} \\
&\le \frac{m_0\alpha\rho(\alpha+\rho-1)}{(m+\alpha)(\rho-1)} \text{ (since } \frac{B(\alpha+k+1;\rho)}{B(\alpha;\rho-1)} \text{ is a probability)} \\
&= \frac{m_0(m+1)\alpha}{m+\alpha} < \frac{m_0(m+1)\alpha(\alpha+1)}{m+\alpha},
\end{aligned}$$

since $\rho = \frac{m+\alpha}{m}$, as required. $\square$

We have obtained bounds for $\left|\frac{\mathbb{E}N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k)\right|$. In the next theorem we deduce total variation distance bounds between the degree distribution and a random variable $V \sim \mu_\rho^\alpha$ using Theorem 8.1.

Theorem 8.2. *Let $W_n^{(BA)}$ represent the degree distribution of Model BA with parameters α, m and m_0. Let V be a random variable satisfying $\mathcal{L}(V) = \mu_\rho^\alpha$ where $\rho = \frac{m+\alpha}{m}$. Then*

$$\begin{aligned}
d_{TV}(W_n^{(BA)}, V) &\le \frac{m_0}{2(n+m_0)(m+\alpha)}\left(\alpha(\alpha+1)(m+1)\log\frac{n+\alpha}{\alpha} + 2m+1\right) \\
&\quad + \frac{B(n+\alpha+1;\frac{m+\alpha}{m})}{2B(\alpha;\frac{m+\alpha}{m})} \\
&= O\left(\frac{\log n}{n}\right) \text{ as } n\to\infty.
\end{aligned}$$

Proof. By the definition of total variation distance,

$$\begin{aligned} d_{TV}(W_n^{(BA)}, V) &= \frac{1}{2}\sum_{k=0}^{\infty} |\mathbb{P}(W_n^{(BA)} = k) - \mathbb{P}(V = k)| \\ &= \frac{1}{2}\sum_{k=0}^{\infty} \left|\frac{\mathbb{E}(N_n^{(BA)}(k))}{n+m_0} - \mu_\rho^\alpha(k)\right|. \end{aligned}$$

Noting that $N_n^{(BA)}(n+k) = 0$ for all $n, k \in \mathbb{N}$ and applying Theorem 8.1 and (3.1) we get,

$$\begin{aligned} & d_{TV}(W_n^{(BA)}, V) \\ \leq\ & \frac{1}{2}\sum_{k=0}^{n} \frac{m_0(m+1)\alpha(\alpha+1)}{(m+\alpha)(n+m_0)(k+\alpha)} + \frac{1}{2}\sum_{k=n+1}^{\infty} \mu_\rho^\alpha(k) + \frac{1}{2}\sum_{k=0}^{n} \frac{m_0 m \mu_\rho^\alpha(k)}{(m+\alpha)(n+m_0)} \\ \leq\ & \frac{1}{2}\sum_{k=0}^{n} \frac{m_0(m+1)\alpha(\alpha+1)}{(m+\alpha)(n+m_0)(k+\alpha)} + \frac{B(n+\alpha+1; \frac{m+\alpha}{m})}{2B(\alpha; \frac{m+\alpha}{m})} + \frac{m_0 m}{2(m+\alpha)(n+m_0)} \\ \leq\ & \frac{m_0}{2(n+m_0)(m+\alpha)}\left(\alpha(\alpha+1)(m+1)\log\frac{n+\alpha}{\alpha} + 2m+1\right) + \frac{B(n+\alpha+1; \frac{m+\alpha}{m})}{2B(\alpha; \frac{m+\alpha}{m})} \\ =\ & O\left(\frac{\log n}{n}\right) \quad \text{as } n \to \infty, \text{ as required.} \end{aligned}$$

□

This method provides bounds of the same order or better (depending on α and m) as the bounds in Chapter 7. The recurrence equation method always gives better bounds for $\left|\frac{\mathbb{E}N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k)\right|$ than those that can be deduced from Chapter 7. These results prove particularly useful in Chapter 9 when we bound the convergence for the degree sequence of Model BA.

Chapter 9

Convergence of the degree sequence

In Chapter 8 we found order $\frac{1}{kn}$ bounds for $\left|\frac{\mathbb{E}N_n(k)}{n+m_0} - \mu_\rho^\alpha(k)\right|$. In this chapter we extend the results to prove complete convergence and to give exact bounds for the rates of convergence for Model BA, Model $\bar{\mathrm{I}}$ and Model Y.

We begin by applying the Azuma-Hoeffding inequality in the case of Model BA. This follows the example of Bollobás et al. [11], and is appropriate only to Model BA in our scheme, since the method requires that at most m edges are added at each time-step. For the other models we are able to use Chernoff bounds, since the vertex degrees evolve independently of one another.

The method is generalised to bound the number of vertices of in-degree lying in some set S, based on the total variation distance between the degree distribution and the μ_ρ^α-distribution. The results of the previous chapters for the total variation distance between the degree distribution and the μ_ρ^α-distribution then make these bounds exact.

9.1 The Azuma-Hoeffding inequality and Model BA

We now apply the method of Bollobás et al. [11] to bound the degree sequence of Model BA. This method relies on the requirement of this particular model that precisely m edges are added at each time step, so the method cannot be extended to the other models in our scheme.

As in the previous chapter we let $N_n^{(BA)}(k)$ be the random variable representing the number of vertices of in-degree k at step n in Model BA. By Theorem 8.1, if $\rho = \frac{m+\alpha}{m}$,

$$\left|\frac{\mathbb{E}(N_n^{(BA)}(k))}{n+m_0} - \mu_\rho^\alpha(k)\right| \leq \frac{m_0(m+1)\alpha(\alpha+1)}{(m+\alpha)(n+m_0)(k+\alpha)} + \frac{m_0 m}{(m+\alpha)(n+m_0)}\mu_\rho^\alpha(k).$$

It will be proved in Theorem 9.4 that to bound the distance between the degree sequence and the μ_ρ^α-distribution it is therefore sufficient to bound

$$\mathbb{P}(|\mathbb{E}(N_n^{(BA)}(k)) - N_n^{(BA)}(k)| \geq \varepsilon), \text{ for } \varepsilon > 0.$$

In fact we find a more general bound for the random variable $M_n^{(BA)}(S)$ defined to be the number of vertices of some in-degree $k \in S \subset \mathbb{Z}_0^+$.

We follow the example of Bollobás et al. [11] and apply the following inequality of Azuma [4] and Hoeffding [27].

Lemma 9.1 (The Azuma-Hoeffding Inequality). *Let $(X_t)_{t=0}^n$ be a martingale that satisfies $|X_{t+1} - X_t| \leq c$ for $t = 0, 1, 2, ..., n-1$. Then*

$$\mathbb{P}(|X_n - X_0| \geq x) \leq \exp\left(-\frac{x^2}{2c^2 n}\right).$$

Lemma 9.2. *Let $(G_n)_{n\geq 0}$ be an evolving random graph sequence with some initial number of vertices, m_0, of in-degree 0. Let the graph grow at time step n by the addition of one new vertex of in-degree 0 that is linked by edges to precisely $m \leq m_0$ vertices, with each linked vertex increasing in in-degree by 1. Suppose also that the probability that a particular vertex is linked to at step n depends only on the current degree of that vertex. Let $S \subset \mathbb{Z}_0^+$ and let $M_n(S)$ represent the number of vertices of in-degree some $k \in S$ at step n.*

Fix $n \in \mathbb{N}$ and for $t = 0, 1, ..., n,\ k \in \mathbb{Z}_0^+$ define

$$X_t = \mathbb{E}(M_n(S)|\sigma(G_t, G_{t-1}, ..., G_0)).$$

Then $(X_t)_{t=0,1,...,n}$ is a martingale satisfying $|X_{t+1} - X_t| \leq 2m$.

Proof. First we need to check that $(X_t)_{t=0,1,...,n}$ is a martingale. Since there are at most $n + m_0$ vertices in the graph at time n, it follows that $0 \leq M_n(S) \leq n$. Thus, for all $t = 1, 2, ..., n$, we have $\mathbb{E}|X_t| \leq n < \infty$.

Also, for $t = 0, 1, ..., n-1,\ s \leq t$,

$$\begin{aligned}
&\mathbb{E}(X_{t+1}|\sigma(G_s, G_{s-1}, ..., G_0)) \\
&\quad = \mathbb{E}(\mathbb{E}(M_n(S)|\sigma(G_{t+1}, G_t, ..., G_0))|\sigma(G_s, G_{s-1}, ..., G_0)) \\
&\quad = \mathbb{E}(M_n(S)|\sigma(G_s, G_{s-1}, ..., G_0)) \\
&\quad = X_s,
\end{aligned}$$

so $(X_t)_{t=0,1,...,n}$ is a martingale as required. It remains to prove that $|X_{t+1} - X_t| \leq 2m$. To do this we consider the possible transitions from G_t to G_{t+1}. At this transition we attach new edges to precisely m vertices. Whether we choose a vertex set $\{a_1, ..., a_m\}$ or a vertex set $\{b_1, ..., b_n\}$ does not affect the joint distribution of the degrees of the remaining vertices at time n. Thus our choice can affect the total number of vertices of degree in S at step n by at most $2m$, so the upper bound of $|X_{t+1} - X_t| \leq 2m$ follows. $\square$

Corollary 9.3. *For a random graph, $(G_n)_{n\geq 0}$, of the form described in Lemma 9.2*

$$\mathbb{P}\left(\left|\frac{\mathbb{E}(M_n(S))}{n+m_0} - \frac{M_n(S)}{n+m_0}\right| > \frac{n^{\frac{3}{4}}}{n+m_0}\right) \leq \left(\exp\left(-\frac{1}{8m^2}\right)\right)^{\sqrt{n}}.$$

Proof. As a direct result of applying Lemma 9.1 to Lemma 9.2 we can say that, for any

$\varepsilon > 0$,

$$\mathbb{P}\left(|\mathbb{E}(M_n(S)) - M_n(S)| \geq \varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2}{8m^2 n}\right).$$

Substitute in $\varepsilon = n^{\frac{3}{4}}$ and divide through by $n + m_0$ on the left hand side to obtain the result. $\square$

The next theorem obtains bounds for the rate of convergence of the degree sequence in Model BA to the μ_ρ^α-distribution with $\rho = \frac{m+\alpha}{m}$.

Theorem 9.4. *Let $N_n^{(BA)}(k)$ be the random variable representing the number of vertices of in-degree $k \in \mathbb{Z}_0^+$ at step $n \in \mathbb{Z}_0^+$ in Model BA with $m_0 \geq m \geq 1$ and $\alpha > 0$. Then the following bounds hold,*

$$\mathbb{P}\left(\left|\frac{N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k)\right| > \frac{n^{\frac{3}{4}}}{n+m_0} + \frac{m_0(m+1)\alpha(\alpha+1)}{(m+\alpha)(n+m_0)(k+\alpha)} + \frac{m_0 m \mu_\rho^\alpha(k)}{(m+\alpha)(n+m_0)}\right)$$
$$\leq \left(\exp\left(-\tfrac{1}{8m^2}\right)\right)^{\sqrt{n}}.$$

Proof. By Theorem 8.1 we know that

$$\left|\frac{\mathbb{E}(N_n^{(BA)}(k))}{n+m_0} - \mu_\rho^\alpha(k)\right| \leq \frac{m_0(m+1)\alpha(\alpha+1)}{(m+\alpha)(n+m_0)(k+\alpha)} + \frac{m_0 m \mu_\rho^\alpha(k)}{(m+\alpha)(n+m_0)},$$

thus, by the triangle inequality,

$$\begin{aligned}\left|\frac{N_n^{(BA)}(k)}{n+m_0} - \mu_\rho^\alpha(k)\right| \leq\ & \frac{m_0(m+1)\alpha(\alpha+1)}{(m+\alpha)(n+m_0)(k+\alpha)} + \frac{m_0 m \mu_\rho^\alpha(k)}{(m+\alpha)(n+m_0)} \\ & + \left|\frac{\mathbb{E}(N_n^{(BA)}(k))}{n+m_0} - \frac{N_n^{(BA)}(k)}{n+m_0}\right|.\end{aligned}$$

By Corollary 9.3

$$\mathbb{P}\left(\left|\frac{\mathbb{E}(N_n^{(BA)}(k))}{n+m_0} - \frac{N_n^{(BA)}(k)}{n+m_0}\right| > \frac{n^{\frac{3}{4}}}{n+m_0}\right) \leq \left(\exp\left(-\frac{1}{8m^2}\right)\right)^{\sqrt{n}}.$$

So substituting this into the above expression gives the required result. ☐

In Corollary 9.3 we bounded the distance between the random variable representing the number of vertices of in-degree in a set S and the expected value of that random variable. Theorem 9.4 combined this with bounds found via recurrence relations in Chapter 8 to give bounds for the convergence of the degree sequence.

The next theorem bounds the convergence of the random variable $M_n^{(BA)}(S)$ for any $S \subset \mathbb{Z}_0^+$.

Theorem 9.5. *Let $S \subset \mathbb{Z}_0^+$ and let $M_n^{(BA)}(S)$ be the random variable representing the number of vertices of in-degree some $k \in S$ at step $n \in \mathbb{Z}_0^+$ in Model BA with $m_0 \geq m \geq 1$ and $\alpha > 0$. Let V be a random variable satisfying $\mathcal{L}(V) = \mu_\rho^\alpha$. Then the following bound holds,*

$$\mathbb{P}\left(\left|\frac{M_n^{(BA)}(S)}{n} - \mu_\rho^\alpha(S)\right| > \frac{n^{\frac{3}{4}}}{n+m_0} + d_{TV}(W_n^{(BA)}, V)\right) \leq \left(\exp\left(-\frac{1}{8m^2}\right)\right)^{\sqrt{n}}.$$

Proof. First note that

$$\begin{aligned}\left|\frac{\mathbb{E}(M_n^{(BA)}(S))}{n} - \mu_\rho^\alpha(S)\right| &= |\mathbb{P}(W_n^{(BA)} - \alpha \in S) - \mu_\rho^\alpha(S)| \\ &\leq d_{TV}(W_n^{(BA)}, V).\end{aligned}$$

Now, by Corollary 9.3,

$$\mathbb{P}\left(\left|\frac{\mathbb{E}(M_n^{(BA)}(k))}{n+m_0} - \frac{M_n^{(BA)}(S)}{n+m_0}\right| > \frac{n^{\frac{3}{4}}}{n+m_0}\right) \leq \left(\exp\left(-\frac{1}{8m^2}\right)\right)^{\sqrt{n}}.$$

So applying the triangle inequality gives the required result. $\square$

We are now able to use the total variation distance bounds of previous chapters to prove complete convergence for all random variables $M_n^{(BA)}(S)$. Of course one special case is the complete convergence of the degree sequence.

Corollary 9.6. *For $S \subset \mathbb{Z}_0^+$, let $M_n^{(BA)}(S)$ be the random variable defined above for Model BA with $m_0 \geq m \geq 1$ and $\alpha > 0$. Then, using the notation for complete convergence given on page vii, and with $\rho = \frac{m+\alpha}{m}$,*

$$\frac{M_n^{(BA)}(S)}{n+m_0} \to^C \mu_\rho^\alpha(S) \text{ as } n \to \infty.$$

Proof. Fix α, m, m_0 and let $\rho = \frac{m+\alpha}{m}$. Let $\varepsilon > 0$ be given. By definition, we need to show that

$$\sum_{n=0}^{\infty} \mathbb{P}\left(\left|\frac{M_n^{(BA)}(S)}{n+m_0} - \mu_\rho^\alpha(S)\right| > \varepsilon\right) \quad < \quad \infty.$$

We know by (for example) Theorem 8.2 that if V is a random variable satisfying $\mathcal{L}(V) = \mu_\rho^\alpha$ then $d_{TV}(W_n^{(BA)}, V) \to 0$ as $n \to \infty$. Therefore, by Theorem 9.5, there exists some positive function f, determined by α, that satisfies $f(n) \to 0$ as $n \to \infty$ and that, for every $n \in \mathbb{N}$,

$$\mathbb{P}\left(\left|\frac{M_n^{(BA)}(S)}{n+m_0} - \mu_\rho^\alpha(S)\right| > f(n)\right) \leq \left(\exp\left(-\frac{1}{8m^2}\right)\right)^{\sqrt{n}}.$$

So we can find $N_\varepsilon \in \mathbb{N}$ such that whenever $n \geq N_\varepsilon$ then $f(n) < \varepsilon$. Thus,

$$\begin{aligned}
&\sum_{n=0}^{\infty} \mathbb{P}\left(\left|\frac{M_n^{(BA)}(S)}{n+m_0} - \mu_\rho^\alpha(S)\right| > \varepsilon\right) \\
&\quad \leq \quad N_\varepsilon + \sum_{n=N_\varepsilon}^{\infty} \mathbb{P}\left(\left|\frac{M_n^{(BA)}(S)}{n+m_0} - \mu_\rho^\alpha(S)\right| > f(n)\right) \\
&\quad \leq \quad N_\varepsilon + \sum_{n=N_\varepsilon}^{\infty} \left(\exp\left(-\frac{1}{8m^2}\right)\right)^{\sqrt{n}} < \infty,
\end{aligned}$$

as required. $\square$

Remark 9.7. *Complete convergence implies almost sure convergence. This can be shown easily using the Borel-Cantelli lemma, for example see Gut, [25] page 168.*

9.2 Chernoff bounds and models with independent vertices

This section is of a similar structure to the previous section, this time looking at models where the number of new edges at a given time step is not bounded by m. Instead of using the Azuma-Hoeffding inequality, we use Chernoff bounds to obtain our bounds. The method relies on the fact that, for each model considered in this section, its vertex degrees evolve independently of one another.

For $S \subset \mathbb{Z}_0^+$, let $M_n^{(Y)}(S)$ and $M_n^{(\overline{I})}(S)$ be the random variables representing the number of vertices of in-degree given by some $k \in S$ at step n in Model Y and Model Ī respectively with the same parameter values, α, β and ρ. For $M \in \{Y, \overline{I}\}$, we know that

$$\frac{1}{n}\mathbb{E}(M_n^{(M)}(S)) = \mathbb{P}(W_n^{(M)} \in S).$$

This can be bounded by the total variation distance from the μ_ρ^α distribution,

$$\begin{aligned}\left|\frac{\mathbb{E}(M_n^{(M)}(S))}{n} - \mu_\rho^\alpha(S)\right| &= |\mathbb{P}(W_n^{(M)} \in S) - \mu_\rho^\alpha(S)| \\ &\leq d_{TV}(W_n^{(M)}, V),\end{aligned}$$

where $\mathcal{L}(V) = \mu_\rho^\alpha$.

We already have obtained bounds for $d_{TV}(W_n^{(M)}, V)$ by various methods, for example from Theorem 6.4 and Theorem 7.2. As in the previous section, Theorem 9.10 will prove that to bound the distances between the degree sequences and the μ_ρ^α-distribution it is therefore sufficient to bound $\mathbb{P}(|\mathbb{E}(M_n^{(M)}(S)) - M_n^{(M)}(S)| \geq \varepsilon)$ for $\varepsilon > 0$. For this we use the following Chernoff bounds.

Lemma 9.8 (Mitzenmacher, Upfal [36], p. 64 and p. 66.)**.** *Let $X_1, X_2, ..., X_n$ be independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = p_i$ for $i = 1, 2, ..., n$. Then for any $\varepsilon \in (0, 1]$,*

$$\mathbb{P}\left(\Big|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} p_i\Big| \geq \varepsilon \sum_{i=1}^{n} p_i\right) \leq e^{-\frac{\varepsilon^2}{3}\sum_{i=1}^{n} p_i}.$$

Corollary 9.9. *Let $M \in \{Y, \overline{I}\}$ and let $S \subset \mathbb{Z}_0^+$. Then for Model M with $M_n^{(M)}(S)$ defined to be random variable representing the number of vertices of in-degree some $k \in S$ at step $n \in \mathbb{Z}_0^+$ we have the following bound,*

$$\mathbb{P}\left(\left|\frac{\mathbb{E}(M_n^{(M)}(S))}{n} - \frac{M_n^{(M)}(S)}{n}\right| > n^{-\frac{1}{4}}\left(\mu_\rho^\alpha(S) + d_{TV}(W_n^{(M)}, V)\right)\right)$$
$$\leq \exp\left(\frac{n^{-\frac{1}{2}}}{3}\left(d_{TV}(W_n^{(M)}, V) - \mu_\rho^\alpha(S)\right)\right).$$

Proof. Let $\varepsilon \in (0, 1]$. First note that $M_n^{(M)}(S)$ can be written as a sum of independent Bernoulli random variables,

$$M_n^{(M)}(S) = \sum_{i=1}^{n} \mathbf{1}_{\{D_n^{(M)}(i) \in S\}}.$$

By Lemma 9.8,

$$\mathbb{P}\left(\Big|\frac{M_n^{(M)}(S)}{n} - \frac{\mathbb{E}(M_n^{(M)}(S))}{n}\Big| > \frac{\varepsilon\mathbb{E}(M_n^{(M)}(S))}{n}\right) \quad \leq \quad \exp\left(-\frac{\varepsilon^2}{3}\mathbb{E}(M_n^{(M)}(S))\right),$$

so,

$$\mathbb{P}\left(\left|\frac{M_n^{(M)}(S)}{n}-\frac{\mathbb{E}(M_n^{(M)}(S))}{n}\right|>\varepsilon\left(\mu_\rho^\alpha(S)+d_{TV}(W_n,V)\right)\right)$$
$$\leq\ \exp\left(-\frac{n\varepsilon^2}{3}\left(\mu_\rho^\alpha(S)-d_{TV}(W_n,V)\right)\right),$$

since $|\mathbb{E}(M_n^{(M)}(S)-n\mu_\rho^\alpha(S)|\leq nd_{TV}(W_n,V)$. The result follows from the substitution of $\varepsilon=n^{-\frac{1}{4}}$. □

Theorem 9.10. *Let $M\in\{Y,\overline{I}\}$, let $S\subset\mathbb{Z}_0^+$ and let $M_n^{(M)}(S)$ be the random variable representing the number of vertices of in-degree some $k\in S$ at step $n\in\mathbb{Z}_0^+$ in Model M with $\alpha\geq 1$, $\beta\geq 0$ and $\rho>1$. Let V be a random variable satisfying $\mathcal{L}(V)=\mu_\rho^\alpha$. Then the following bound holds,*

$$\mathbb{P}\left(\left|\frac{M_n^{(M)}(S)}{n}-\mu_\rho^\alpha(S)\right|>n^{-\frac{1}{4}}+(1+n^{-\frac{1}{4}})d_{TV}(W_n^{(M)},V+\alpha)\right)$$
$$\leq\exp\left(\frac{n^{-\frac{1}{2}}}{3}\left(d_{TV}(W_n^{(M)},V)-\mu_\rho^\alpha(S)\right)\right).$$

Proof. We know that

$$\left|\frac{\mathbb{E}(M_n^{(M)}(S))}{n}-\mu_\rho^\alpha(S)\right|\leq d_{TV}(W_n^{(M)},V).$$

By Corollary 9.9, for $M\in\{Y,\overline{I}\}$,

$$\mathbb{P}\left(\left|\frac{\mathbb{E}(M_n^{(M)}(S))}{n}-\frac{M_n^{(M)}(S)}{n}\right|>n^{-\frac{1}{4}}\left(\mu_\rho^\alpha(S)+d_{TV}(W_n^{(M)},V)\right)\right)$$
$$\leq\exp\left(\frac{n^{-\frac{1}{2}}}{3}\left(d_{TV}(W_n^{(M)},V)-\mu_\rho^\alpha(S)\right)\right).$$

So by the triangle inequality we get the required result. □

In Section 9.1 we were able to deduce complete convergence from our bounds for convergence of the degree sequence. We now do the same for the models discussed here.

Corollary 9.11. *Let $M \in \{Y, \overline{I}\}$, let $S \subset \mathbb{Z}_0^+$ and let $M_n^{(M)}(S)$ be the random variable representing the number of vertices of in-degree some $k \in S$ at step $n \in \mathbb{Z}_0^+$ in Model M with $\alpha \geq 1$, $\beta \geq 0$ and $\rho > 1$. Then,*

$$M_n^{(M)}(S) \to^C \mu_\rho^\alpha(S) \text{ as } n \to \infty.$$

Proof. Let $\varepsilon > 0$ be given. By definition, we need to show that

$$\sum_{n=0}^{\infty} \mathbb{P}\left(\left|\frac{M_n^{(M)}(S)}{n} - \mu_\rho^\alpha(S)\right| > \varepsilon\right) \quad < \quad \infty.$$

We know from previous results of (for example) Chapter 6 and Chapter 8 that if V is a random variable satisfying $\mathcal{L}(V) = \mu_\rho^\alpha$ then $d_{TV}(W_n^{(M)}, V) \to 0$ as $n \to \infty$. Therefore, by Theorem 9.10, there exists some positive function f, determined by α and ρ, that satisfies $f(n) \to 0$ as $n \to \infty$ and that, for every $n \in \mathbb{N}$,

$$\mathbb{P}\left(\left|\frac{M_n^{(M)}(S)}{n} - \mu_\rho^\alpha(S)\right| > f(n)\right) \leq \exp\left(\frac{n^{-\frac{1}{2}}}{3}\left(d_{TV}(W_n^{(M)}, V) - \mu_\rho^\alpha(S)\right)\right).$$

So we can find $N_\varepsilon \in \mathbb{N}$ such that whenever $n \geq N_\varepsilon$ then $f(n) < \varepsilon$. Thus,

$$\begin{aligned}
&\sum_{n=0}^{\infty} \mathbb{P}\left(\left|\frac{M_n^{(M)}(S)}{n} - \mu_\rho^\alpha(S)\right| > \varepsilon\right) \\
&\quad \leq \quad N_\varepsilon + \sum_{n=N_\varepsilon}^{\infty} \mathbb{P}\left(\left|\frac{M_n^{(M)}(S)}{n} - \mu_\rho^\alpha(S)\right| > f(n)\right) \\
&\quad \leq \quad N_\varepsilon + \sum_{n=N_\varepsilon}^{\infty} \exp\left(\frac{n^{-\frac{1}{2}}}{3}\left(d_{TV}(W_n^{(M)}, V) - \mu_\rho^\alpha(S)\right)\right) < \infty,
\end{aligned}$$

as required. $\square$

Chapter 10

A connection with the Chinese restaurant process

This chapter connects the degree distributions of our random graphs to a statistic of the Chinese Restaurant process. We are able to deduce results for this statistic from our previous work. We discuss how an aspect of the evolution of a preferential attachment random graph can be thought of in terms of an inhomogeneous Markov chain. The Markov chain is defined so that the n^{th} element of the chain has the same distribution as the degree of a randomly chosen vertex in Model $\bar{\mathrm{I}}$ of Chapter 7.

The results of Chapter 6 and Chapter 7 have proved that the Markov chain has the μ_ρ^α-distribution as its limiting distribution. We look at related homogeneous Markov chains with transition probabilities based on those for a particular step of this inhomogeneous Markov chain. Bounds for the rate of convergence (to truncated power law distributions) for these Markov chains are given. Chapter 11 uses Stein's method for the μ_ρ^α-distribution to prove directly that the limiting degree distribution is the μ_ρ^α-distribution.

We discuss a variant of the Chinese restaurant process [40] and note that the inhomogeneous Markov chain that we define represents the precise evolution of a particular statistic in this process. The bounds that we have found therefore apply to this statistic of the Chinese restaurant process.

10.1 An inhomogeneous Markov chain

10.1.1 Model $\bar{\mathrm{I}}$

First recall the definition of Model $\bar{\mathrm{I}}$ that can be found in Appendix B. We have been studying the in-degree distribution of Model $\bar{\mathrm{I}}$, denoted by $W_n^{(\bar{I})}$. We now define a Markov chain, $(X_n)_{n\geq 0}$, such that the marginal distribution of X_n satisfies $\mathcal{L}(X_n) = \mathcal{L}(W_n^{(\bar{I})}+1)$ for all $n \in \mathbb{Z}_0^+$.

10.1.2 The Markov chain

Theorem 10.1. *For $n = 0, 1, 2, ...$ let*

$$X_n := D_n^{(\bar{I})}(I_{n+1}) - \alpha + 1,$$

with $(I_n)_{n\in\mathbb{N}}$ satisfying

$$\begin{aligned}
\mathbb{P}(I_{n+1} = I_n | I_n) &= \frac{n}{n+1} \\
\mathbb{P}(I_{n+1} = n+1 | I_n) &= \frac{1}{n+1}.
\end{aligned}$$

with $(I_n)_{n\in\mathbb{N}} \perp (\mathcal{G}_n^{(\bar{I})})_{n\in\mathbb{N}}$, the evolving random graph of Model $\bar{I}$. Then $(X_n)_{n\in\mathbb{Z}_0^+}$ is an inhomogeneous Markov chain on $\mathbb{N}$ with $X_0 = 1$ and with transition probabilities for $n \in \mathbb{N}$ given by,

$$\begin{aligned}
\mathbb{P}(X_n = 1 | X_{n-1} = 1) &= 1 - \frac{n\alpha}{(n+1)(\rho n + \beta)}; \\
\mathbb{P}(X_n = 2 | X_{n-1} = 1) &= \frac{n\alpha}{(n+1)(\rho n + \beta)};
\end{aligned}$$

and, for $2 \leq i \leq n$,

$$
\begin{aligned}
\mathbb{P}(X_n = 1|X_{n-1} = i) &= \frac{1}{n+1}; \\
\mathbb{P}(X_n = i|X_{n-1} = i) &= \frac{n}{n+1}\left(1 - \frac{i+\alpha-1}{\rho n+\beta}\right); \\
\mathbb{P}(X_n = i+1|X_{n-1} = i) &= \frac{n}{n+1}\left(\frac{i+\alpha-1}{\rho n+\beta}\right).
\end{aligned}
$$

Proof. Let us begin by proving that the Markov property is satisfied for $(X_n)_{n\in\mathbb{Z}_0^+}$. For $1 \leq i \leq n+1$,

$$
\begin{aligned}
\mathbb{P}(X_n = i|\{X_j\}_{j<n}) &= \mathbb{P}(D_n^{(\overline{I})}(I_{n+1}) = i+\alpha-1|\{D_j^{(\overline{I})}(I_{j+1})\}_{j<n}) \\
&= \frac{1}{n+1}\mathbb{P}(D_n^{(\overline{I})}(n+1) = i+\alpha-1|\{D_j^{(\overline{I})}(I_{j+1})\}_{j<n}) \\
&\quad + \frac{n}{n+1}\mathbb{P}(D_n^{(\overline{I})}(I_n) = i+\alpha-1|\{D_j^{(\overline{I})}(I_{j+1})\}_{j<n}) \\
&= \frac{1}{n+1}\mathbb{P}(D_n^{(\overline{I})}(n+1) = i+\alpha-1|D_{n-1}^{(\overline{I})}(I_n)) \\
&\quad + \frac{n}{n+1}\mathbb{P}(D_n^{(\overline{I})}(I_n) = i+\alpha-1|D_{n-1}^{(\overline{I})}(I_n)) \\
&= \mathbb{P}(X_n = i|X_{n-1}).
\end{aligned}
$$

Now let us calculate the transition probabilities. For $1 \leq i \leq n$,

$$
\begin{aligned}
\mathbb{P}(X_n = i+1|X_{n-1} = i) &= \mathbb{P}(D_n^{(\overline{I})}(I_{n+1}) = i+\alpha|D_{n-1}^{(\overline{I})}(I_n) = i+\alpha-1) \\
&= \mathbb{P}(D_n^{(\overline{I})}(I_n) = i+\alpha \cap I_{n+1} = I_n|D_{n-1}^{(\overline{I})}(I_n) = i+\alpha-1) \\
&\quad \text{(since if } I_{n+1} = n+1 \text{ then } D_n^{(\overline{I})}(I_{n+1}) = \alpha \text{ by definition)} \\
&= \mathbb{P}(I_{n+1} = I_n)\mathbb{P}(D_n^{(\overline{I})}(I_n) = i+\alpha|D_{n-1}^{(\overline{I})}(I_n) = i+\alpha-1) \\
&\quad \text{(by construction } (I_n)_{n\in\mathbb{N}} \perp (\mathcal{G}_n^{(\overline{I})})_{n\in\mathbb{N}}) \\
&= \frac{n}{n+1}\left(\frac{i+\alpha-1}{\rho n+\beta}\right),
\end{aligned}
$$

as required.

Similarly, for $2 \leq i \leq n$,

$$\begin{aligned}
\mathbb{P}(X_n = i | X_{n-1} = i) &= \mathbb{P}(D_n^{(\overline{I})}(I_{n+1}) = i + \alpha - 1 | D_{n-1}^{(\overline{I})}(I_n) = i + \alpha - 1) \\
&= \mathbb{P}(D_n^{(\overline{I})}(I_n) = i + \alpha - 1 \cap I_{n+1} = I_n | D_{n-1}^{(\overline{I})}(I_n) = i + \alpha - 1) \\
&= \mathbb{P}(I_{n+1} = I_n)\mathbb{P}(D_n^{(\overline{I})}(I_n) = i + \alpha - 1 | D_{n-1}^{(\overline{I})}(I_n) = i + \alpha - 1) \\
&\quad \text{(by construction } (I_n)_{n\in\mathbb{N}} \perp (\mathcal{G}_n^{(\overline{I})})_{n\in\mathbb{N}}) \\
&= \frac{n}{n+1}\left(1 - \frac{i + \alpha - 1}{\rho n + \beta}\right).
\end{aligned}$$

Also, for $2 \leq i \leq n$,

$$\begin{aligned}
\mathbb{P}(X_n = 1 | X_{n-1} = i) &= \mathbb{P}(D_n^{(\overline{I})}(I_{n+1}) = \alpha | D_{n-1}^{(\overline{I})}(I_n) = i + \alpha - 1) \\
&= \mathbb{P}(I_{n+1} = n + 1 | D_{n-1}^{(\overline{I})}(I_n) = i + \alpha - 1) \\
&\quad \text{(since } I_{n+1} = I_n \implies D_n^{(\overline{I})}(I_{n+1}) \geq D_{n-1}^{(\overline{I})}(I_n) \geq 2) \\
&= \mathbb{P}(I_{n+1} = n + 1) \\
&= \frac{1}{n+1}.
\end{aligned}$$

Of course either of the last two options have the same effect on the process if we are currently in state 1, so

$$\begin{aligned}
\mathbb{P}(X_n = 1 | X_{n-1} = 1) &= \mathbb{P}(D_n^{(\overline{I})}(I_{n+1}) = \alpha | D_{n-1}^{(\overline{I})}(I_n) = \alpha) \\
&= \mathbb{P}(D_n^{(\overline{I})}(I_n) = \alpha \cap I_{n+1} = I_n | D_{n-1}^{(\overline{I})}(I_n) = \alpha) \\
&\quad + \mathbb{P}(I_{n+1} = n + 1 | D_{n-1}^{(\overline{I})}(I_n) = \alpha) \\
&= \frac{n}{n+1}\left(1 - \frac{\alpha}{\rho n + \beta}\right) + \frac{1}{n+1} \\
&= 1 - \frac{\alpha n}{(n+1)(\rho n + \beta)}.
\end{aligned}$$

Finally we see that, since X_i can only increase by at most 1 at each transition, $X_{n-1} \in \{1, 2, ..., n\}$. Thus we have considered all possible transition probabilities for X_{n-1}. $\square$

Later, in Subsection 10.3.1 we study homogeneous Markov chains derived by taking the transition probabilities at a given step n of the above Markov chain. We show that the stationary distribution of such a Markov chain is a truncated power law distribution and that it takes precisely the form of the truncated μ_ρ^α-distribution given in Example 5.12 when $\beta = 0$.

10.2 The Chinese restaurant process

This section looks at how our Markov chain $(X_n)_{n\in\mathbb{N}\cup\{0\}}$ can describe a particular statistic in a variant of the Chinese restaurant process.

10.2.1 Simple random seating plan

The Chinese restaurant process [40] can be thought of as a description of customers entering a restaurant and sitting at tables.

In the simplest version of the model, customers choose a table to sit at following a preferential mechanism. Person 1 sits at Table 1. For $n \geq 1$, suppose the first n customers are already seated. Person $n + 1$ enters the restaurant and chooses uniformly at random between $n + 1$ possible spaces. These are n spaces, one next to each of the n previous customers, or one space at a new table.

After the n^{th} customer we can consider the probability distribution on the set of partitions of $\{1, 2, ..., n\}$, where elements i, j are assigned to the same set whenever customers i and j are sitting on the same table.

10.2.2 Random seating plan with departure

We now describe a slight variant on the simple random seating plan model above that can be related to our inhomogeneous Markov chain.

Person 1 sits at Table 1. For $n \geq 1$, suppose the first n customers are already seated. Person $n+1$ enters the restaurant and chooses where to sit by the following process.

1. With probability $\frac{1}{n+1}$ Person n sits on a new table.
2. Otherwise Person $n+1$ chooses uniformly at random between the first n individuals. With probability $\frac{n}{\rho n+\beta}$ Person $n+1$ sits at the table of the chosen individual.
3. If Person $n+1$ is still not seated, a table is chosen uniformly at random from all the occupied tables. Let N_n be the random variable representing the number of occupied tables at this time. Person $n+1$ sits at the chosen table with probability, conditional on N_n, given by $\frac{(\alpha-1)N_n}{((\rho-1)n+\beta)}$. Since $N_n \leq n$ this conditional probability takes a value in $[0,1]$.
4. If not seated by now, Person $n+1$ leaves the restaurant.

Note that Step 3 can be omitted if $\alpha = 1$.

To describe the statistic that our Markov chain represents we use the concept of a waiter serving at tables. The random variable X_n will describe the size of the table that the waiter is serving at when the $(n+1)^{st}$ customer has arrived.

The waiter begins by serving at Table 1. For $n \geq 1$, suppose the first n customers are already seated and that the waiter is serving Table T_n. Person $n+1$ arrives. If Person $n+1$ sits at a new table then the waiter moves to that table. Otherwise the waiter stays at Table T_n. Since the tables are ordered based on when they first become occupied, the table that the waiter serves at is the final table in the ordering of occupied tables. Equivalently, $T_n = N_n$, with N_n defined as above to be the number of occupied tables.

Let X_n be the size of the table that the waiter is serving at time n. In terms of the partition of customers represented by the tables that they sit on, X_n represents the size of the set in the partition with the greatest least element.

The following example illustrates the process.

Example 10.2. *1. Person 1 sits at Table 1. Waiter serves Table 1.*

$T_1 = 1,\ X_0 = 1.$

Corresponds to partition $\{\{1\}\}$.

2. *Person 2 starts a new table. Waiter serves Table 2.*

   $T_2 = 2,\ X_1 = 1.$

   Corresponds to partition $\{\{1\},\{2\}\}$.

3. *Person 3 sits at Table 1. Waiter serves Table 2.*

   $T_3 = 2,\ X_2 = 1.$

   Corresponds to partition $\{\{1,3\},\{2\}\}$.

4. *Person 4 sits at Table 2. Waiter serves Table 2.*

   $T_4 = 2,\ X_3 = 2.$

   Corresponds to partition $\{\{1,3\},\{2,4\}\}$.

5. *Person 5 sits at a new table. Waiter serves Table 3.*

   $T_5 = 3,\ X_4 = 1.$

   Corresponds to partition $\{\{1,3\},\{2,4\},\{5\}\}$.

6. *Person 6 sits at Table 1. Waiter serves Table 3.*

   $T_6 = 3,\ X_5 = 1.$

   Corresponds to partition $\{\{1,3,6\},\{2,4\},\{5\}\}$.

7. *Person 7 sits at Table 3. Waiter serves Table 3.*

$T_7 = 3,\ X_6 = 2.$

Corresponds to partition $\{\{1,3,6\},\{2,4\},\{5,7\}\}$.

8. *Person 8 sits at Table 1. Waiter serves Table 3.*

   $T_8 = 3,\ X_7 = 2.$

   Corresponds to partition $\{\{1,3,6,8\},\{2,4\},\{5,7\}\}$.

9. *Person 9 sits at a Table 3. Waiter serves Table 3.*

   $T_9 = 3,\ X_8 = 3.$

   Corresponds to partition $\{\{1,3,6,8\},\{2,4\},\{5,7,9\}\}$.

10. *Person 10 sits at a new table. Waiter serves Table 4.*

    $T_{10} = 4,\ X_9 = 1.$

    Corresponds to partition $\{\{1,3,6,8\},\{2,4\},\{5,7,9\},\{10\}\}$.

Theorem 10.3. *In the random seating plan with departure process described above, let X_{n-1} represent the size of Table T_n when n people have arrived. Then $(X_n)_{n\in\mathbb{Z}_0^+}$ is a Markov chain satisfying $X_0 = 1$ and with the transition probabilities given in the statement of Theorem 10.1.*

Proof. Since the first customer sits on a table alone we know that $X_0 = 1$. We want to prove that the Markov property holds for $(X_n)_{n\in\mathbb{Z}_0^+}$ and that the transition probabilities given in Theorem 10.1 apply. The random variable X_{n-1} is bounded above by n (when all individuals sit at Table 1) so we need only consider transitions from states $\{1, 2, ..., n\}$.

The possible transitions are as follows. Suppose $X_{n-1} = i$. Then we know that the waiter is serving a table of size i. There are three possible transitions at the arrival of Person $n + 1$.

- Person $n + 1$ sits at a new table. Then the waiter moves to serve Person $n + 1$ so $X_n = 1$. The probability of this occurring is independent of all previous steps in the process.

- Person $n+1$ sits at Table T_n. Then $X_n = i+1$.

  Given X_{n-1}, which is the size of Table T_n, we shall see that this is independent of the previous steps in the process.

- Person $n+1$ sits at an another occupied table or leaves. Then $X_n = i$.

  Given X_{n-1}, which is the size of Table T_n, we shall see that this is independent of the previous steps in the process since the sum of table sizes is fixed.

We now calculate the associated transition probabilities. The calculation of transition probabilities proves that if Person $n+1$ sits at an occupied table other than table T_n or leaves then which of these tables is chosen, if any, does not affect the evolution of X_n. This explains why, despite Step 3 of the process conditioning on N_n, we can prove the Markov property for $(X_n)_{n\in\mathbb{Z}_0^+}$ and not only for $((X_n, N_n))_{n\in\mathbb{N}}$. It can be shown that the Markov property holds for $(N_n)_{n\in\mathbb{N}}$ but the result is unnecessary and so the proof is omitted.

For $1 \leq i \leq n$,

$$
\begin{aligned}
&\mathbb{P}(X_n = i+1 | X_{n-1} = i, N_n = r_n, (X_s = x_s, N_{s+1} = r_{s+1})_{\{s=0,1,\ldots,n-2\}}) \\
&\quad = \mathbb{P}(\text{Person } n+1 \text{ sits at Table } T_n) \\
&\quad = \mathbb{P}(\text{Person } n+1 \text{ sits at Table } T_n \text{ at Step 2}) \\
&\qquad +\mathbb{P}(\text{Person } n+1 \text{ sits at Table } T_n \text{ at Step 3}).
\end{aligned}
$$

The number of customers at Table T_n is deterministic given $X_{n-1} = i$, and is equal to i.

Therefore,

$$
\begin{aligned}
&\mathbb{P}(X_n = i+1 | X_{n-1} = i, N_n = r_n, (X_s = x_s, N_{s+1} = r_{s+1})_{\{s=0,1,\ldots,n-2\}}) \\
&= \frac{n}{n+1}\left(\frac{\#\text{customers at } T_n}{n}\right)\left(\frac{n}{\rho n+\beta}\right) \\
&\quad + \frac{n}{n+1}\left(\frac{(\rho-1)n+\beta}{\rho n+\beta}\right)\left(\frac{1}{r_n}\right)\left(\frac{(\alpha-1)r_n}{((\rho-1)n+\beta)}\right) \\
&= \frac{n}{n+1}\left(\frac{\alpha+i-1}{\rho n+\beta}\right) \qquad (10.1) \\
&= \mathbb{P}(X_n = i+1 | X_{n-1} = i).
\end{aligned}
$$

For $2 \le i \le n$,

$$
\begin{aligned}
&\mathbb{P}(X_n = 1 | X_{n-1} = i, N_n = r_n, (X_s = x_s, N_{s+1} = r_{s+1})_{\{s=0,1,\ldots,n-2\}}) \\
&= \mathbb{P}(\text{Person } n+1 \text{ sits at new table}) \\
&= \frac{1}{n+1} \\
&= \mathbb{P}(X_n = 1 | X_{n-1} = i).
\end{aligned}
$$

Also, since there are only three possible transitions from state i, for $2 \le i \le n$,

$$
\begin{aligned}
&\mathbb{P}(X_n = i | X_{n-1} = i, N_n = r_n, (X_s = x_s, N_{s+1} = r_{s+1})_{\{s=0,1,\ldots,n-2\}}) \\
&= 1 - \mathbb{P}(X_n = i+1 | X_{n-1} = i) - \mathbb{P}(X_n = 1 | X_{n-1} = i) \\
&= 1 - \frac{1}{n+1} - \frac{n}{n+1}\left(\frac{\alpha+i-1}{\rho n+\beta}\right) \\
&= \frac{n}{n+1}\left(1 - \frac{\alpha+i-1}{\rho n+\beta}\right) \\
&= \mathbb{P}(X_n = i | X_{n-1} = i).
\end{aligned}
$$

Finally,

$$\begin{aligned}
&\mathbb{P}(X_n = 1 | X_{n-1} = 1, N_n = r_n, (X_s = x_s, N_{s+1} = r_{s+1})_{\{s=0,1,\dots,n-2\}}) \\
&= \mathbb{P}(\text{Person } n+1 \text{ does not sit at table } T_n) \\
&= 1 - \mathbb{P}(\text{Person } n+1 \text{ sits at table } T_n) \\
&= 1 - \mathbb{P}(X_n = 2 | X_{n-1} = 1) \\
&= 1 - \frac{n}{n+1}\left(\frac{\alpha}{\rho n + \beta}\right) \quad \text{(by (10.1))} \\
&= \mathbb{P}(X_n = 1 | X_{n-1} = 1),
\end{aligned}$$

as required. □

It follows that the size of the newest table at step $n+1$, X_n, is identically distributed to the degree of a randomly chosen vertex in Model $\bar{\text{I}}$ of Chapter 7. Thus it can be is approximated by a μ_ρ^α-distribution for $\rho > 1$, and the bounds from Chapter 7 apply.

Remark 10.4. *The simple random seating plan version of the Chinese restaurant problem corresponds to this model with $\alpha = 1, \beta = 0, \rho = 1$. Our bounds for convergence in Chapter 7 require $\rho > 1$, and we see that the convergence is $O(n^{2(\rho-1)})$, which would be a constant if we allowed $\rho = 1$.*

10.3 Associated homogeneous Markov chains

Now that we have an inhomogeneous Markov chain, it is instructive to look at associated homogeneous Markov chains for fixed time-steps. For each time-step n we define a Markov chain on the state space $\mathcal{S}_n = \{1, 2, ..., n+1\}$ using the transition probabilities at the n^{th} step of the inhomogeneous Markov chain.

We prove that the stationary distribution of such a Markov chain is a truncated power law distribution on $\mathcal{S}_n$ and we bound the rate of convergence for each such n. This gives us a snapshot of the convergence of the inhomogeneous Markov chain at each time-step. In particular, we note that the stationary distributions converge to the μ_ρ^α-distribution as $n \to \infty$.

For fixed n, define the Markov chain $(X_k^{(n)})_{k\in\mathbb{N}\cup\{0\}}$, using the following transition probabilities taken from X_n,

$$\begin{aligned}
\mathbb{P}(X_k^{(n)} = 1|X_{k-1}^{(n)} = 1) &= 1 - \frac{\alpha n}{(\rho n+\beta)(n+1)}; \\
\mathbb{P}(X_k^{(n)} = 2|X_{k-1}^{(n)} = 1) &= \frac{\alpha n}{(\rho n+\beta)(n+1)};
\end{aligned} \tag{10.2}$$

and, for $1 \le i \le n$,

$$\begin{aligned}
\mathbb{P}(X_k^{(n)} = 1|X_{k-1}^{(n)} = i) &= \frac{1}{n+1}; \\
\mathbb{P}(X_k^{(n)} = i|X_{k-1}^{(n)} = i) &= \frac{n}{n+1}\left(1 - \frac{i+\alpha-1}{\rho n+\beta}\right); \\
\mathbb{P}(X_k^{(n)} = i+1|X_{k-1}^{(n)} = i) &= \frac{(i+\alpha-1)n}{(\rho n+\beta)(n+1)}.
\end{aligned} \tag{10.3}$$

Our Markov chain moves from a state in $\{1, 2, ..., n\}$ to one in $\{1, 2, ..., n+1\}$, but nowhere else. To define a homogeneous Markov chain based on these conditions that is valid for more than one transition we define transition probabilities from $\mathcal{S}_n \setminus \{1, 2, ..., n\}$ to $\mathcal{S}_n$ by,

$$\begin{aligned}
\mathbb{P}(X_k^{(n)} = n+1|X_{k-1}^{(n)} = n+1) &= 1 - \frac{1}{(n+1)}; \\
\mathbb{P}(X_k^{(n)} = 1|X_{k-1}^{(n)} = n+1) &= \frac{1}{(n+1)}.
\end{aligned}$$

Thus there is no absorbing state, the chain as aperiodic and irreducible. Hence it possesses a unique stationary distribution.

10.3.1 The stationary distribution

We now derive the stationary distribution for the homogeneous Markov chain, $(X_k^{(n)})_{k\in\mathbb{N}\cup\{0\}}$. When $\beta = 0$ this distribution is precisely the truncated μ_ρ^α-distribution described in Example 5.12 and truncated at n. For every fixed value of $\beta \geq 0$, the distribution converges to the μ_ρ^α-distribution as $n \to \infty$, the known asymptotic distribution of the inhomogeneous Markov chain.

Similarly to (5.11), we define the truncated μ_ρ^α-distribution as follows,

$$\begin{aligned}
\nu_n^{\alpha,\beta,\rho}(s) &= \frac{B(s+\alpha,\rho+\frac{\beta}{n}+1)}{B(\alpha;\rho+\frac{\beta}{n})} \text{ for } s = 0,1,2,...,n-1; \\
\nu_n^{\alpha,\beta,\rho}(n) &:= 1 - \sum_{s=0}^{n-1} \nu_n^{\alpha,\beta,\rho}(s) \qquad (10.4)\\
&= \frac{B(n+\alpha;\rho+\frac{\beta}{n})}{B(\alpha;\rho+\frac{\beta}{n})} \text{ (by (3.1))}; \\
\nu_n^{\alpha,\beta,\rho}(s) &:= 0 \text{ for } s = n+1, n+2, n+3, ...
\end{aligned}$$

Theorem 10.5. *Let $(X_k^{(n)})_{k\in\mathbb{N}\cup\{0\}}$ be the homogeneous Markov chain defined above. Then $\nu_n^{\alpha,\beta,\rho}$ is the stationary distribution of $(X_k^{(n)}-1)_{k\in\mathbb{N}\cup\{0\}}$.*

Proof. Let $n \in \mathbb{N}$ be fixed and suppose that for some $r \in \mathbb{Z}_0^+$ we know that for every $s \in \mathbb{N}$, $\mathbb{P}(X_r^{(n)} - 1 = s) = \nu_n^{\alpha,\beta,\rho}(s)$. We need to show that it follows that for all $s \in \mathbb{N}$, $\mathbb{P}(X_{r+1}^{(n)} - 1 = s) = \nu_n^{\alpha,\beta,\rho}(s)$.

We calculate these probabilities using the transition probabilities for $(X_k^{(n)})_{k\in\mathbb{N}\cup\{0\}}$. Firstly,

$$\begin{aligned}
\mathbb{P}(X_{r+1}^{(n)}-1=0) &= \sum_{s=0}^{\infty}\nu_n^{\alpha,\beta,\rho}(s)\mathbb{P}(X_{r+1}^{(n)}=1|X_r^{(n)}=s)\\
&= \nu_n^{\alpha,\beta,\rho}(0)\left(1-\frac{\alpha n}{(\rho n+\beta)(n+1)}\right)+\sum_{s=1}^{\infty}\nu_n^{\alpha,\beta,\rho}(s)\frac{1}{n+1}\\
&= \frac{B(\alpha;\rho+\frac{\beta}{n}+1)}{B(\alpha;\rho+\frac{\beta}{n})}\left(\frac{n(\rho n+\beta-\alpha)}{(\rho n+\beta)(n+1)}\right)+\frac{1}{n+1}\sum_{s=0}^{\infty}\nu_n^{\alpha,\beta,\rho}(s)\\
&= \frac{B(\alpha;\rho+\frac{\beta}{n}+1)}{B(\alpha;\rho+\frac{\beta}{n})}\left(\frac{n(\rho n+\beta-\alpha)}{(\rho n+\beta)(n+1)}\right)+\frac{1}{n+1}\\
&= \frac{B(\alpha;\rho+\frac{\beta}{n}+1)}{B(\alpha;\rho+\frac{\beta}{n})}\left(\frac{n(\rho n+\beta-\alpha)}{(\rho n+\beta)(n+1)}+\frac{\rho+\frac{\beta}{n}+\alpha}{(\rho+\frac{\beta}{n})(n+1)}\right)\\
&\quad \text{(by Appendix A, (A.3))}\\
&= \frac{B(\alpha;\rho+\frac{\beta}{n}+1)}{B(\alpha;\rho+\frac{\beta}{n})}\left(\frac{(\rho+\frac{\beta}{n})(n+1)}{(\rho+\frac{\beta}{n})(n+1)}\right)\\
&= \nu_n^{\alpha,\beta,\rho}(0),
\end{aligned}$$

as required.

Similarly, for $1\le i\le n-1$,

$$\begin{aligned}
\mathbb{P}(X_{r+1}^{(n)}-1=i) &= \sum_{s=0}^{\infty}\nu_n^{\alpha,\beta,\rho}(s)\mathbb{P}(X_{r+1}^{(n)}=i+1|X_r^{(n)}=s+1)\\
&= \nu_n^{\alpha,\beta,\rho}(i)\frac{n}{n+1}\left(1-\frac{i+\alpha}{\rho n+\beta}\right)+\nu_n^{\alpha,\beta,\rho}(i-1)\frac{(i+\alpha-1)n}{(\rho n+\beta)(n+1)}\\
&= \nu_n^{\alpha,\beta,\rho}(i)\left(\frac{1}{n+1}\left(n-\frac{i+\alpha}{\rho+\frac{\beta}{n}}+\frac{B(i+\alpha-1;\rho+\frac{\beta}{n}+1)}{B(i+\alpha;\rho+\frac{\beta}{n}+1)}\left(\frac{i+\alpha-1}{\rho+\frac{\beta}{n}}\right)\right)\right)\\
&\quad \text{(by (10.4))}\\
&= \nu_n^{\alpha,\beta,\rho}(i)\left(\frac{1}{n+1}\left(n-\frac{i+\alpha}{\rho+\frac{\beta}{n}}+\frac{i+\alpha+\rho+\frac{\beta}{n}}{i+\alpha-1}\left(\frac{i+\alpha-1}{\rho+\frac{\beta}{n}}\right)\right)\right)\\
&\quad \text{(by Appendix A, (A.2))}\\
&= \nu_n^{\alpha,\beta,\rho}(i)\left(\frac{1}{n+1}\left(n-\frac{i+\alpha}{\rho+\frac{\beta}{n}}+\frac{i+\alpha+\rho+\frac{\beta}{n}}{\rho+\frac{\beta}{n}}\right)\right)\\
&= \nu_n^{\alpha,\beta,\rho}(i),
\end{aligned}$$

and, finally,

$$
\begin{aligned}
&\mathbb{P}(X_{r+1}^{(n)} - 1 = n) \\
= \;& \sum_{s=0}^{\infty} \nu_n^{\alpha,\beta,\rho}(s)\mathbb{P}(X_{r+1}^{(n)} = n+1 | X_r^{(n)} = s+1) \\
= \;& \nu_n^{\alpha,\beta,\rho}(n)\left(1 - \frac{1}{n+1}\right) + \nu_n^{\alpha,\beta,\rho}(n-1)\left(\frac{n(n+\alpha-1)}{(\rho n+\beta)(n+1)}\right) \\
& \text{(by definition of the transition probabilities)} \\
= \;& \frac{B(n+\alpha;\rho+\frac{\beta}{n})}{B(\alpha;\rho+\frac{\beta}{n})}\left(1-\frac{1}{n+1}\right) + \frac{B(n+\alpha-1;\rho+\frac{\beta}{n}+1)}{B(\alpha;\rho+\frac{\beta}{n})}\left(\frac{n+\alpha-1}{(\rho+\frac{\beta}{n})(n+1)}\right) \\
= \;& \frac{B(n+\alpha;\rho+\frac{\beta}{n})}{B(\alpha;\rho+\frac{\beta}{n})}\left(\frac{n}{n+1}+\frac{1}{n+1}\right) \quad \text{(by Appendix A, (A.4))} \\
= \;& \nu_n^{\alpha,\beta,\rho}(n),
\end{aligned}
$$

as required. □

10.3.2 Rate of convergence to stationarity

Finally, we bound the rate of convergence of each of these Markov chains to its stationary distribution. This explains further the convergence results that we have already seen in Chapter 7. We see that the bounds on the rate of convergence decrease in order with n, equivalent to decreasing with increasing time in the inhomogeneous Markov chain.

In order to bound the rate of convergence of one of these Markov chains to its stationary distribution, we use the relationship this process has with an evolving random graph similar to Model $\bar{\text{I}}$. Let $n \in \mathbb{N}$ and consider the following evolving random graph, which we call Model $X^{(n)}$.

Model $X^{(n)}$

- Start with one vertex, $\mathcal{G}_0 = K_1$;

- At step k, independently for each vertex in the graph with in-degree at most n, select vertex vertex i with probability $\frac{D_{k-1}^{(X^{(n)})}(i)}{\rho n+\beta}$, conditional on the degree of vertex i at step $k-1$, $D_{k-1}^{(X^{(n)})}(i) := (\alpha + \text{in-degree of } i \text{ at step } k-1)$, $\alpha > 0$, $\beta \geq 0$, $\rho > 1$ and $\alpha \leq \rho + \beta$.

- Attach the new vertex $k+1$ to all the selected vertices, with edges directed out of the new vertex. (The number of vertices selected at step k is random.)

Now define the sequence $(I_k^{(n)})_{k\in\mathbb{N}}$ recursively as follows. Let $I_1^{(n)} := 1$ and, independently of the evolution of the graph, let,

$$I_{k+1}^{(n)} = \begin{cases} k+1 \text{ with probability } \frac{1}{n+1}; \\ I_k^{(n)} \text{ otherwise.} \end{cases} \tag{10.5}$$

Thus $I_k^{(n)} \in \{1, 2, ..., k\}$ but $I_k^{(n)}$ is not, in general, uniformly distributed.

Lemma 10.6. *For $n = 0, 1, 2, ...,$ we can construct $(X_k^{(n)})_{k\in\mathbb{Z}_0^+}$ with transition probabilities given by (10.2) and (10.3), by putting*

$$X_k^{(n)} := D_k^{(X^{(n)})}(I_{k+1}^{(n)}),$$

with $(I_k^{(n)})_{k\in\mathbb{N}}$ satisfying (10.5), and $(I_k^{(n)})_{k\in\mathbb{N}} \perp (\mathcal{G}_n^{(\overline{I})})_{n\in\mathbb{N}}$, the evolving random graph.

Proof. The proof for a given step k is exactly as for the n^{th} step in Theorem 10.1, with the transitions probabilities now the same for every step k.

□

Lemma 10.6 is now used to obtain upper bounds for the rate of convergence of the Markov chain $(X_k^{(n)})_{k\in\mathbb{N}\cup\{0\}}$ to its stationary distribution. We use the fact that, once in stationarity, we remain in stationarity. Thus if we adapt Model $X^{(n)}$ giving our initial vertex a degree chosen from $\nu_n^{\alpha,\beta,\rho}$, then the distribution of $D_k^{(X^{(n)})}(I_{k+1}^{(n)})$ will always have this measure. We can then bound the distance between the distributions by looking at the probability that $I_k^{(n)} = I_1^{(n)}$, the only vertex that may differ in degree between the models.

Theorem 10.7. *Let $(X_k^{(n)})_{k\in\mathbb{N}\cup\{0\}}$ be the homogeneous Markov chain satisfying transition probabilities (10.2) and (10.3), and with $X_0^{(n)} = 1$. Let V_n be a random variable with truncated power law measure $\nu_n^{\alpha,\beta,\rho}$. Then, for $k \in \mathbb{N}\cup\{0\}$,*

$$d_{TV}(X_{k+1}^{(n)} - 1, V_n) \leq \left(\frac{n}{n+1}\right)^k.$$

Proof. Define a random graph, Model $Z^{(n)}$, that evolves in an identical way to Model $X^{(n)}$, but where the initial degree of the first vertex is identically distributed to the random variable $V_n + \alpha$. Consider $(D_k^{(Z^{(n)})}(I_{k+1}^{(n)}))_{k\in\mathbb{Z}_0^+}$ the evolution of the degrees of the vertices $(I_k^{(n)})_{k\in\mathbb{N}}$ for this model.

By Lemma 10.6, the evolution of vertex degrees given by the sequence, $(D_k^{(Z^{(n)})}(I_{k+1}^{(n)}) - \alpha + 1)_{k\in\mathbb{N}\cup\{0\}}$, is represented by a Markov chain, $(Z_k^{(n)})_{k\in\mathbb{N}\cup\{0\}}$, with the same transition probabilities as the Markov chain $(X_k^{(n)})_{k\in\mathbb{N}\cup\{0\}}$. Moreover, $\mathcal{L}(Z_0^{(n)} - 1) = \nu_n^{\alpha,\beta,\rho}$ so the chain begins in stationarity (by Theorem 10.5) and it follows that for all $k \in \mathbb{N}$, $\mathcal{L}(D_k^{(Z^{(n)})}(I_{k+1}^{(n)}) - \alpha) = \nu_n^{\alpha,\beta,\rho}$.

Let $(D_k^{(X^{(n)})}(I_{k+1}^{(n)}))_{k\in\mathbb{N}\cup\{0\}}$ be the degree sequence defined above in Model $X^{(n)}$. It follows from Lemma 10.6 that,

$$d_{TV}(X_{k+1}^{(n)}, V_n) = d_{TV}(D_k^{(X^{(n)})}(I_{k+1}^{(n)}), D_k^{(Z^{(n)})}(I_{k+1}^{(n)})).$$

Now the random graphs in Model $X^{(n)}$ and Model $Z^{(n)}$ consist of vertices that grow in degree independently of one another. Moreover the evolution of the degree of any vertex $i \geq 2$ follows the same stochastic process in either random graph model. We couple the evolution of vertices $2, 3, 4, \ldots$ in Model $X^{(n)}$ and Model $Z^{(n)}$ so that $D_k^{(X^{(n)})}(i) = D_k^{(Z^{(n)})}(i)$

whenever $i \geq 2$.

With this coupling $D_k^{(X^{(n)})}(I_{k+1}^{(n)}) = D_k^{(Z^{(n)})}(I_{k+1}^{(n)})$ whenever $I_{k+1} \geq 2$, so the bound corresponds to the probability that $I_{k+1}^{(n)} = 1$,

$$\begin{aligned}
d_{TV}(X_{k+1}^{(n)}, V_n) &= d_{TV}(D_k^{(X^{(n)})}(I_{k+1}^{(n)}), D_k^{(Z^{(n)})}(I_{k+1}^{(n)})) \\
&\leq \mathbb{P}(D_k^{(X^{(n)})}(I_{k+1}^{(n)}) \neq D_k^{(Z^{(n)})}(I_{k+1}^{(n)})) \\
&\leq \mathbb{P}(I_{k+1}^{(n)} = 1) \\
&= \mathbb{P}(\{I_s^{(n)} = 1\}_{s=1,2,\ldots,k+1}) \\
&= \left(\frac{n}{n+1}\right)^k,
\end{aligned}$$

as required. $\square$

Thus for fixed n (as in the homogeneous random graph) the rate of convergence is at most exponential with k. However for fixed k as $n \to \infty$ the bound converges to 1. This suggests that the convergence rate in the inhomogeneous Markov chain slows as n increases. Of course the distribution to which we are converging changes with n as well, being a truncated power law distribution on a expanding state space. In contrast to the bounds found through other methods, the bound that we see here is not very refined and does not depend on α, β or ρ.

Chapter 11

Applying Stein's method for the μ_ρ^α-distribution

So far we have not applied Stein's method for the μ_ρ^α-distribution directly to bound the degree distribution of a preferential attachment random graph model. In this chapter we discuss a possible route that uses the inhomogeneous Markov chain defined in Chapter 10.

Recall from Theorem 10.1 that the degree distribution of Model $\bar{\mathrm{I}}$ can be described as the Markov chain $(X_n)_{n\in\mathbb{Z}_0^+}$ with $X_0 = 1$ and transition probabilities for $n \in \mathbb{N}$ given by,

$$\begin{aligned}
\mathbb{P}(X_n = 1 | X_{n-1} = 1) &= 1 - \frac{n\alpha}{(n+1)(\rho n + \beta)}; \\
\mathbb{P}(X_n = 2 | X_{n-1} = 1) &= \frac{n\alpha}{(n+1)(\rho n + \beta)};
\end{aligned}$$

and, for $2 \leq i \leq n$,

$$\begin{aligned}
\mathbb{P}(X_n = 1 | X_{n-1} = i) &= \frac{1}{n+1}; \\
\mathbb{P}(X_n = i | X_{n-1} = i) &= \frac{n}{n+1}\left(1 - \frac{i+\alpha-1}{\rho n + \beta}\right); \\
\mathbb{P}(X_n = i+1 | X_{n-1} = i) &= \frac{n}{n+1}\left(\frac{i+\alpha-1}{\rho n + \beta}\right).
\end{aligned}$$

The next theorem uses Stein's method for the μ_ρ^α-distribution, as defined in Chapter 4, to bound the total variation distance between X_n and the μ_ρ^α-distribution based on how quickly $(X_n)_{n\in\mathbb{Z}_0^+}$ is converging to a limiting distribution.

Theorem 11.1. *Let $(X_n)_{n\in\mathbb{Z}_0^+}$ be the Markov chain defined above with $\alpha > 0$, $\beta \geq 0$, $\rho > 1$ and $\alpha \leq \beta + \rho$. Let V be a random variable satisfying $\mathcal{L}(V) = \mu_\rho^\alpha$. Then, if $\beta > 0$,*

$$d_{TV}(X_n - 1, V) \leq \frac{3(\alpha+1)}{n+1}\left(\sum_{i=0}^{n-1} d_{TV}(X_i, X_n) + 1 + \frac{\beta}{\rho}\log\frac{\rho n + \beta}{\beta}\right),$$

and if $\beta = 0$,

$$d_{TV}(X_n - 1, V) \leq \frac{3(\alpha+1)}{n+1}\left(\sum_{i=0}^{n-1} d_{TV}(X_i, X_n) + 1\right).$$

Proof. Let

$$\mathcal{I}_\rho^\alpha = \left\{ g : g(0) = 0, \quad \forall k \geq 0 \quad g(k) \leq 0 \quad |g(k)| \leq \frac{2}{\rho}, \quad |\Delta g(k)| \leq \frac{3}{k} \right\}.$$

Then, by Theorem 5.9,

$$\begin{aligned} d_{TV}(X_n - 1, V) &\leq \sup_{g\in\mathcal{I}_\rho^\alpha} |\mathbb{E}\left((X_n + \alpha + \rho - 1)g(X_n - 1) - (X_n + \alpha - 1)g(X_n)\right)| \\ &= \sup_{g\in\mathcal{I}_\rho^\alpha} |\mathbb{E}\left(-(X_n + \alpha - 1)\Delta g(X_n - 1) + \rho g(X_n - 1)\right)|. \end{aligned}$$

From the transition probabilities and iteration,

$$\begin{aligned} \mathbb{E}(\rho g(X_n - 1)) &= \frac{\rho n}{(n+1)(\rho n + \beta)}\mathbb{E}((X_{n-1} + \alpha - 1)\Delta g(X_{n-1} - 1)) + \frac{\rho n}{n+1}\mathbb{E}(g(X_{n-1} - 1)) \\ &= \sum_{i=0}^{n-1} \frac{\rho(i+1)}{(n+1)(\rho(i+1) + \beta)}\mathbb{E}((X_i + \alpha - 1)\Delta g(X_i - 1)). \end{aligned}$$

Thus,

$$
\begin{aligned}
&d_{TV}(X_n - 1, V) \\
&\le \sup_{g\in\mathcal{I}_\rho^\alpha} \Big| -\mathbb{E}(X_n+\alpha-1)\Delta g(X_n-1) \\
&\quad + \sum_{i=0}^{n-1} \frac{\rho(i+1)}{(n+1)(\rho(i+1)+\beta)} \mathbb{E}((X_i+\alpha-1)\Delta g(X_i-1))\Big|.
\end{aligned}
$$

Defining, for $k \ge 1$, $h(k) := (k+\alpha-1)\Delta g(k-1)$, we see that if $g \in \mathcal{I}_\rho^\alpha$ then $|h(k)| \le \frac{3(k+\alpha-1)}{k} \le 3(\alpha+1)$. So

$$
\begin{aligned}
&d_{TV}(X_n - 1, V) \\
&\le \sup_{||h||_\infty \le 3(\alpha+1)} \Big| -\mathbb{E}h(X_n) + \sum_{i=0}^{n-1} \frac{\rho(i+1)}{(n+1)(\rho(i+1)+\beta)} \mathbb{E}(h(X_i))\Big| \\
&= \sup_{||h||_\infty \le 3(\alpha+1)} \frac{1}{n+1} \Big| \sum_{i=0}^{n-1} \frac{\rho(i+1)}{(\rho(i+1)+\beta)} \mathbb{E}(h(X_i) - h(X_n)) \\
&\quad -\mathbb{E}h(X_n)\left(1 + \sum_{i=0}^{n-1} \frac{\beta}{\rho(i+1)+\beta}\right)\Big| \\
&\le \frac{3(\alpha+1)}{n+1}\left(\sum_{i=0}^{n-1} d_{TV}(X_i, X_n) + 1 + \sum_{i=0}^{n-1} \frac{\beta}{\rho(i+1)+\beta}\right).
\end{aligned}
$$

Thus, by integration, if $\beta > 0$,

$$
d_{TV}(X_n - 1, V) \le \frac{3(\alpha+1)}{n+1}\left(\sum_{i=0}^{n-1} d_{TV}(X_i, X_n) + 1 + \frac{\beta}{\rho}\log\frac{\rho n+\beta}{\beta}\right),
$$

and if $\beta = 0$,

$$
d_{TV}(X_n - 1, V) \le \frac{3(\alpha+1)}{n+1}\left(\sum_{i=0}^{n-1} d_{TV}(X_i, X_n) + 1\right).
$$

as required. $\square$

In Chapter 10 we asserted that if our inhomogeneous Markov chain had a limiting distribution then, based on the stationary distributions of a sequence of related homogeneous Markov chains, we would expect that limiting distribution to be the μ_ρ^α-distribution. The next theorem proves this formally using the results of Theorem 11.1.

Theorem 11.2. *Let $(X_n)_{n\in\mathbb{Z}_0^+}$ be the Markov chain defined above with $\alpha > 0$, $\beta \geq 0$, $\rho > 1$ and $\alpha \leq \beta + \rho$. Let V be a random variable satisfying $\mathcal{L}(V) = \mu_\rho^\alpha$ and let W be some discrete random variable.*

Suppose that $X_n \to^{d_{TV}} W$ as $n \to \infty$. Then $W =^D V$.

Proof. Since $X_n \to^{d_{TV}} W$, to show that $W =^D V$ it is equivalent to prove that $X_n \to^{d_{TV}} V$ as $n \to \infty$. By Theorem 11.1, if $\beta > 0$,

$$d_{TV}(X_n - 1, V) \leq \frac{3(\alpha+1)}{n+1}\left(\sum_{i=0}^{n-1} d_{TV}(X_i, X_n) + \max\left(1, \frac{\beta}{\rho}\log\frac{\rho n+\beta}{\beta}\right)\right),$$

and if $\beta = 0$,

$$d_{TV}(X_n - 1, V) \leq \frac{3(\alpha+1)}{n+1}\left(\sum_{i=0}^{n-1} d_{TV}(X_i, X_n) + 1\right).$$

For any $\beta > 0$, $\frac{\beta}{\rho(n+1)}\log\frac{\rho n+\beta}{\beta} \to 0$ as $n \to \infty$. Thus it remains to prove that

$$\frac{1}{n+1}\sum_{i=0}^{n-1} d_{TV}(X_i, X_n) \to 0 \text{ as } n \to \infty.$$

We know that $d_{TV}(X_i, W) \to 0$ as $n \to \infty$. Define f to be a non-negative function with $f(x) \downarrow 0$ as $x \uparrow \infty$ and $d_{TV}(X_i, W) \leq f(i)$ for all i.

Then, by the triangle inequality,

$$\begin{aligned}
\frac{1}{n+1}\sum_{i=0}^{n-1} d_{TV}(X_i, X_n) &\leq \frac{1}{n+1}\sum_{i=0}^{n-1}(f(i)+f(n)) \\
&\leq \frac{2}{n+1}\sum_{i=0}^{n-1} f(i) \\
&= \frac{2}{n+1}\sum_{i=0}^{\lceil\sqrt{n-1}\rceil} f(i) + \frac{2}{n+1}\sum_{i=\lceil\sqrt{n-1}\rceil+1}^{n-1} f(i) \\
&\leq \frac{2f(0)\lceil\sqrt{n-1}\rceil}{n+1} + \frac{2(n-\lceil\sqrt{n-1}\rceil)f(\lceil\sqrt{n-1}\rceil)}{n+1} \\
&\to 0 \text{ as } n \to \infty,
\end{aligned}$$

as required. ☐

Chapter 12

Evolving Bernoulli random graph models

So far we have considered only random graph models that follow a preferential attachment condition. This chapter presents similar results for an evolving Bernoulli random graph model. The fact that we can successfully employ the same methods within a different type of model demonstrates their versatility. Additionally, the new bounds presented for this model are of interest in their own right and follow on from previous work giving the asymptotic degree distribution of this evolving Bernoulli random graph model, for example in [11].

As in the case of Barabási-Albert random graph models, we begin this chapter by defining a discrete-time random graph model. We then look at an analogous continuous-time model. The discrete-time model that we study was mentioned by Bollobás et al. [11], who proved that the asymptotic degree distribution is geometric. We apply Stein's method for the geometric distribution directly to our model to obtain bounds for the total variation distance between the degree distribution at step n and the geometric distribution. Additionally, we apply Stein's method for the normal distribution to obtain bounds from normality for the number of triangles in the model.

Motivated by our study of the continuous-time Barabási-Albert random graph model, we derive a continuous-time version of this evolving Bernoulli random graph. Again, we look at the degree distribution at a sequence of stopping times and show that this is precisely geometric.

12.1 A discrete-time evolving Bernoulli random graph

We define a Bernoulli evolving random graph that uses uniform attachment rather than the preferential attachment of Barabási-Albert models. As such, it can be thought of as perhaps the simplest type of evolving random graph.

The model begins with a complete graph on $m+1^*$ vertices. At each time step we add one new vertex to the graph with $m \in \mathbb{N}$ new edges. The edges attach the new vertex to m distinct vertices already in the graph, with the m vertices chosen uniformly at random from all possible sets of vertices of size m. We write $\mathcal{V}_n$ for the vertex set at step n and $\mathcal{E}_n$ for the edge set at step n. The evolution is summarised as follows.

- Start with the complete graph on $m+1$ vertices, $\mathcal{G}_0 = K_{m+1}$;
- at step $n+1$ choose a set of m vertices $\{i_1, i_2, ..., i_m\}$ uniformly at random from all m-element subsets of $\mathcal{V}_n$;
- $\mathcal{V}_{n+1} := \mathcal{V}_n \cup \{n+m+2\}$, $\mathcal{E}_{n+1} := \mathcal{E}_n \cup \{\{i_1, n+m+2\}, \{i_2, n+m+2\}, ..., \{i_m, n+m+2\}\}$.

12.1.1 Bounds for the degree distribution

Bollobás et al. [11] prove that the asymptotic degree distribution of this model is $m+V$ where $V \sim Geometric_{\mathbb{Z}_0^+}(\frac{1}{m+1})$. We use Stein's method for the geometric distribution to bound this approximation for the degree distribution at step n.

Let $I_{n+m+1} \sim Uniform\{1, 2, ..., n+m+1\}$ and denote the degree of vertex i at step n by $D_n(i)$. Then $D_n(I_{n+m+1})$ denotes the degree of a vertex chosen uniformly at random from the graph at step n. We want to find the total variation distance between $D_n(I_{n+m+1})-m$

*We choose to start with the complete graph on $m+1$ vertices since this means that every vertex (including the initial vertex) enters the graph with initial degree m. Of course, we could prove similar results for a general initial size m_0 as we do in the case of the Barabási-Albert random graph, but having seen the method in the Barabási-Albert case it is not particularly instructive to include it here.

and $V \sim Geometric(\frac{1}{m+1})$. We use Stein's method for the geometric distribution, and specifically the methods of Chapter 4.

The first step is to compare the degrees of consecutively added vertices. Let $X_j^{(i)}$ take the value 1 if vertex i is chosen at step j and take the value 0 otherwise. Then for $i \in \{1, 2, ..., n+m\}$,

$$D_n(i) = m + \sum_{j=\max(i-m,1)}^{n} X_j^{(i)}.$$

Note that for a given vertex i, $X_j^{(i)} \sim Bernoulli(\frac{m}{j+m})$ independently for each $j \in \mathbb{N} \cap \{i-m, i-m+1, i-m+2, ...\}$. Also $D_j(j+m+1) = m$.

Lemma 12.1. *Let $n \in \mathbb{N}$ and suppose $i \in \{m+1, m+2, ..., m+n\}$. Then for any function h,*

$$\mathbb{E}(h(D_n(i))) = \frac{m}{i}\mathbb{E}(h(D_n(i+1)+1)) + \frac{i-m}{i}\mathbb{E}(h(D_n(i+1))).$$

Proof. Condition on whether vertex i is chosen at step $i-m$, that is the first step after i enters the graph. This gives

$$\begin{aligned}
&\mathbb{E}(h(D_n(i))) \\
&= \mathbb{E}(h(D_n(i))|X_{i-m}^{(i)} = 1)\mathbb{P}(X_{i-m}^{(i)} = 1) + \mathbb{E}(h(D_n(i))|X_{i-m}^{(i)} = 0)\mathbb{P}(X_{i-m}^{(i)} = 0) \\
&= \frac{m}{i}\mathbb{E}(h(\sum_{j=i-m+1}^{n} X_j^{(i)} + m + 1)) + \frac{i-m}{i}\mathbb{E}(h(\sum_{j=i-m+1}^{n} X_j^{(i)} + m)) \\
&= \frac{m}{i}\mathbb{E}(h(\sum_{j=i-m+1}^{n} X_j^{(i+1)} + m + 1)) + \frac{i-m}{i}\mathbb{E}(h(\sum_{j=i-m+1}^{n} X_j^{(i+1)} + m)) \\
&\quad \text{(by definition since } X_j^{(i)} \sim Bernoulli(\frac{m}{j+m}) \text{ for all } i) \\
&= \frac{m}{i}\mathbb{E}(h(D_n(i+1)+1)) + \frac{i-m}{i}\mathbb{E}(h(D_n(i+1))).
\end{aligned}$$

□

Corollary 12.2. *Let $n \in \mathbb{N}$ and suppose $i \in \{m+1, m+2, ..., m+n\}$. Then for any function h, for all $N \geq n$,*

$$\begin{aligned}
\mathbb{E}(h(D_N(i))) &= m \sum_{j=0}^{n+m-i} \frac{(i-1)!(i+j-m-1)!}{(i+j)!(i-m-1)!} \mathbb{E}(h(D_N(i+j+1)+1)) \\
&\quad + \frac{(i-1)!n!}{(n+m)!(i-m-1)!} \mathbb{E}(h(D_N(n+m+1))) \\
&= \frac{m}{m+1} \sum_{j=0}^{n+m-i} \frac{\binom{i-1}{m}}{\binom{i+j}{m+1}} \mathbb{E}(h(D_N(i+j+1)+1)) \\
&\quad + \frac{\binom{i-1}{m}}{\binom{n+m}{m}} \mathbb{E}(h(D_N(n+m+1))).
\end{aligned}$$

Proof. We proceed by induction on n. First note that for $n = 1$ this is a reformulation of the statement of Lemma 12.1 and so holds.

Suppose that Corollary 12.2 holds for $n = k \geq 1$.

Consider $n = k + 1$. Let $N \geq k + 1$. Then $n \geq k$, so by the inductive hypothesis

$$\begin{aligned}
\mathbb{E}(h(D_N(i))) &= m \sum_{j=0}^{k+m-i} \frac{(i-1)!(i+j-m-1)!}{(i+j)!(i-m-1)!} \mathbb{E}(h(D_N(i+j+1)+1)) \\
&\quad + \frac{(i-1)!k!}{(k+m)!(i-m-1)!} \mathbb{E}(h(D_N(k+m+1)))
\end{aligned}$$

Applying Lemma 12.1 gives

$$\begin{aligned}
\mathbb{E}(h(D_N(k+m+1))) &= \frac{m}{k+m+1} \mathbb{E}(h(D_N(k+m+2)+1)) \\
&\quad + \frac{k+1}{k+m+1} \mathbb{E}(h(D_N(k+m+2))).
\end{aligned}$$

Combining the two, we get

$$\begin{aligned}
\mathbb{E}(h(D_N(i))) &= m\sum_{j=0}^{k+m-i}\frac{(i-1)!(i+j-m-1)!}{(i+j)!(i-m-1)!}\mathbb{E}(h(D_N(i+j+1)+1))\\
&\quad+\frac{(i-1)!k!}{(k+m)!(i-m-1)!}\frac{m}{k+m+1}\mathbb{E}(h(D_N(k+m+2)+1))\\
&\quad+\frac{(i-1)!k!}{(k+m)!(i-m-1)!}\frac{k+1}{k+m+1}\mathbb{E}(h(D_N(k+m+2)))\\
&= m\sum_{j=0}^{k+1+m-i}\frac{(i-1)!(i+j-m-1)!}{(i+j)!(i-m-1)!}\mathbb{E}(h(D_N(i+j+1)+1))\\
&\quad+\frac{(i-1)!(k+1)!}{(k+m+1)!(i-m-1)!}\mathbb{E}(h(D_N(k+m+2)))\\
&= \frac{m}{m+1}\sum_{j=0}^{n+m-i}\frac{\binom{i-1}{m}}{\binom{i+j}{m+1}}\mathbb{E}(h(D_N(i+j+1)+1))\\
&\quad+\frac{\binom{i-1}{m}}{\binom{n+m}{m}}\mathbb{E}(h(D_N(n+m+1))),
\end{aligned}$$

as required. $\square$

Theorem 12.3. *Let $\mu_m(k) := \frac{m^k}{(m+1)^{k+1}}$ for $k = 0, 1, 2, \ldots$ be the measure of the random variable $V \sim Geometric_{\mathbb{Z}_0^+}(\frac{1}{m+1})$. Let f be a bounded real-valued function satisfying $|f(k)| \le 1$. Then, for all $n \in \mathbb{N}$,*

$$|\mathbb{E}(f(D_{I_{n+m+1}}(n)-m)) - \mu_m(f)| \le \frac{m(m+1)}{n+m+1}.$$

Moreover, the total variation distance between $D_n(I_n) - m$ and the random variable V is bounded above by,

$$d_{TV}(V, D_n(I_n)-m) \le \frac{m(m+1)}{n+m+1}.$$

Proof. Let g be an arbitrary function that satisfies $g(0) = 0$ and $|g(j+1)| \le m+1$ and $\Delta g(j) \le \frac{m+1}{j+1} + j\min\left(\frac{1}{j}, \frac{m+2}{(m+1)(j+1)}\right)$ for $j \in \mathbb{Z}_0^+$. From Corollary 4.3, taking $p = \frac{1}{m+1}$, we know that it is sufficient to show that

$$\left|\mathbb{E}\left(\frac{m}{m+1}g(D_n(I_{n+m+1})+1-m)-g(D_n(I_{n+m+1})-m)\right)\right|\leq\frac{m(m+1)}{n+m+1}.$$

Conditioning on I_{n+m+1} gives

$$\begin{aligned}
&\mathbb{E}(g(D_n(I_{n+m+1})-m))\\
&= \frac{1}{n+m+1}\sum_{i=1}^{n+m+1}\mathbb{E}(g(D_n(i)-m))\\
&= \frac{m}{n+m+1}\mathbb{E}(g(D_n(1)-m))+\frac{1}{n+m+1}\sum_{i=m+1}^{n+m}\mathbb{E}(g(D_n(i)-m)), \qquad (12.1)
\end{aligned}$$

since $g(D_n(n+m+1)-m) = g(0) = 0$ and the first $m+1$ vertices have identically distributed degrees.

Now, by Corollary 12.2, if $i \geq m+1$

$$\begin{aligned}
&\mathbb{E}(g(D_n(i)-m))\\
&= \frac{m}{m+1}\sum_{j=0}^{n+m-i}\frac{\binom{i-1}{m}}{\binom{i+j}{m+1}}\mathbb{E}(g(D_n(i+j+1)+1-m))\\
&\quad +\frac{\binom{i-1}{m}}{\binom{n+m}{m}}\mathbb{E}(g(D_n(n+m+1)-m))\\
&= \frac{m}{m+1}\sum_{j=0}^{n+m-i}\frac{\binom{i-1}{m}}{\binom{i+j}{m+1}}\mathbb{E}(g(D_n(i+j+1)+1-m)),
\end{aligned}$$

since $D_n(n+m+1)-m=0$ and $g(0)=0$.

Putting $s = i + j$, this becomes

$$\begin{aligned}
&\mathbb{E}(g(D_n(i) - m)) \\
&= \frac{m}{m+1} \sum_{s=i}^{n+m} \frac{\binom{i-1}{m}}{\binom{s}{m+1}} \mathbb{E}(g(D_n(s+1) + 1 - m)).
\end{aligned}$$

We substitute in (12.1),

$$\begin{aligned}
&\mathbb{E}(g(D_n(I_{n+m+1}) - m)) \\
&= \frac{m}{n+m+1} \mathbb{E}(g(D_n(1) - m)) \\
&\quad + \frac{m}{(m+1)(n+m+1)} \sum_{i=m+1}^{n+m} \sum_{s=i}^{n+m} \frac{\binom{i-1}{m}}{\binom{s}{m+1}} \mathbb{E}(g(D_n(s+1) + 1 - m)).
\end{aligned}$$

Changing the order of summation gives

$$\begin{aligned}
&\frac{m}{(m+1)(n+m+1)} \sum_{i=m+1}^{n+m} \sum_{s=i}^{n+m} \frac{\binom{i-1}{m}}{\binom{s}{m+1}} \mathbb{E}(g(D_n(s+1) + 1 - m)) \\
&= \frac{m}{(m+1)(n+m+1)} \sum_{s=m+1}^{n+m} \frac{1}{\binom{s}{m+1}} \mathbb{E}(g(D_n(s+1) + 1 - m)) \sum_{i=m+1}^{s} \binom{i-1}{m} \\
&= \frac{m}{m+1} \frac{1}{n+m+1} \sum_{s=m+2}^{n+m+1} \mathbb{E}(g(D_n(s) + 1 - m)) \\
&= \frac{m}{m+1} \mathbb{E}(g(D_n(I_{n+m+1}) + 1 - m)) - \frac{m}{m+1} \frac{1}{n+m+1} \sum_{s=1}^{m+1} \mathbb{E}(g(D_n(s) + 1 - m)) \\
&= \frac{m}{m+1} \mathbb{E}(g(D_n(I_{n+m+1}) + 1 - m)) - \frac{m}{(n+m+1)} \mathbb{E}(g(D_n(1) + 1 - m)),
\end{aligned}$$

since the first $m + 1$ vertices have identical degree distributions.

Thus

$$\begin{aligned}
&\mathbb{E}(g(D_n(I_{n+m+1})-m)) - \frac{m}{m+1}\mathbb{E}(g(D_n(I_{n+m+1})+1-m)) \\
&= \frac{m}{n+m+1}\mathbb{E}(g(D_n(1)-m)) - \frac{m}{n+m+1}\mathbb{E}(g(D_n(1)+1-m)).
\end{aligned}$$

This gives,

$$\begin{aligned}
&\left|\mathbb{E}(g(D_n(I_{n+m+1})-m)) - \frac{m}{m+1}\mathbb{E}(g(D_n(I_{n+m+1})+1-m))\right| \\
&= \left|\frac{m}{n+m+1}\left(\mathbb{E}(g(D_n(1)-m)) - \mathbb{E}(g(D_n(1)+1-m))\right)\right| \\
&\leq \frac{m}{n+m+1}||\Delta g||_\infty \\
&\leq \frac{m(m+1)}{n+m+1},
\end{aligned}$$

using the Stein bounds from Corollary 4.3, $||g||_\infty \leq \frac{1}{p}+1 = m+1$.

The total variation distance bound follows immediately by definition. □

12.1.2 Bounds for the degree sequence

In Section 9.1 we applied the Azuma-Hoeffding inequality to our results for Model BA to obtain bounds for the degree sequence from degree distribution bounds. As explained in Section 9.1, this method relied on the property that precisely m vertices increased in degree at each time-step in Model BA. In fact, Lemma 9.2 and Corollary 9.3 apply to a class of evolving random graph models including the Bernoulli evolving random graph of current interest.

We now apply these results to obtain degree sequence bounds and to prove complete convergence of the degree sequence in the case of the current model.

For $n \in \mathbb{Z}_0^+$, $S \subset \mathbb{Z}_0^+$, let us define $M_n(S)$ to be the number of vertices of degree $m+k$ where $k \in S$. In the language of Lemma 9.2, if a vertex has degree $m+k$ then we say that its in-degree is k.

We already know from Theorem 12.3 that $\left|\frac{\mathbb{E}(M_n(S))}{n+m+1} - \mu_m(S)\right| \leq \frac{m(m+1)}{n+m+1}$.

Theorem 12.4. *For the evolving Bernoulli random graph model above the following bounds hold for any $m, n \in \mathbb{N}$, $S \subset \mathbb{Z}_0^+$,*

$$\mathbb{P}\left(\left|\frac{M_n(S)}{n+m+1} - \mu_m(S)\right| > \frac{n^{\frac{3}{4}}}{n+m_0} + \frac{m(m+1)}{n+m+1}\right) \leq \left(\exp\left(-\frac{1}{8m^2}\right)\right)^{\sqrt{n}}.$$

Proof. We follow the proof of Theorem 9.4.

By Theorem 12.3 we know that,

$$\left|\frac{\mathbb{E}(M_n(S))}{n+m+1} - \mu_m(S)\right| \leq \frac{m(m+1)}{n+m+1}.$$

Thus, by the the triangle inequality,

$$\left|\frac{M_n(S)}{n+m+1} - \mu_m(S)\right| \leq \frac{m(m+1)}{n+m+1} + \left|\frac{\mathbb{E}(M_n(S))}{n+m+1} - \frac{M_n(S)}{n+m+1}\right|.$$

So,

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{M_n(S)}{n+m+1} - \mu_m(S)\right| > \frac{n^{\frac{3}{4}}}{n+m_0} + \frac{m(m+1)}{n+m+1}\right) \\
\leq\ & \mathbb{P}\left(\left|\frac{\mathbb{E}(M_n(S))}{n+m+1} - \frac{M_n(S)}{n+m+1}\right| > \frac{n^{\frac{3}{4}}}{n+m_0}\right) \\
\leq\ & \left(\exp\left(-\frac{1}{8m^2}\right)\right)^{\sqrt{n}},
\end{aligned}$$

by Corollary 9.3. $\square$

Remark 12.5. *Again, as in Corollary 9.6, complete convergence follows from Theorem 12.4. That is to say,* $\frac{M_n(S)}{n+m+1} \to^C \mu_m(S)$ *as* $n \to \infty$.

12.1.3 Bounds for the number of triangles

This section looks at the $m = 2$ version of the above model. We show how the distance between the distribution of the number of triangles in this model and the normal distribution can be bounded using Stein's method for the normal distribution. This illustrates the fact that Stein's method can be applied to study various statistics, not only the degree distribution of the graph. The number of triangles is a statistic of interest since it indicates the level of clustering in the graph. The $m = 2$ version of the model can be described as follows.

- Start with the complete graph on 3 vertices, $\mathcal{G}_0 = K_3$;
- at step $n+1$ choose a vertex pair $\{i, j\}$ uniformly at random from $\{\{i, j\} : i, j \in \mathcal{V}_n, i \neq j\}$;
- $\mathcal{V}_{n+1} := \mathcal{V}_n \cup \{k+n+1\}$, $\mathcal{E}_{n+1} := \mathcal{E}_n \cup \{\{i, k+n+1\}, \{j, k+n+1\}\}$.

We see that at each step the number of triangles formed is either 1, if the chosen vertices are joined by an edge, or otherwise 0.

Proposition 12.6. *For* $n = 1, 2, \ldots,$ *let*

$$X_n = \begin{cases} 1 & \text{if a triangle is formed at step } n; \\ 0 & \text{otherwise.} \end{cases}$$

Let W_n be the total number of triangles in the graph at step n,

$$W_n = 1 + \sum_{i=1}^{n} X_i.$$

Then $X_n \sim Bernoulli(p_n)$ where $p_n = \frac{2(2n+1)}{(n+2)(n+1)}$,

and $\mathbb{E}(W_n) \sim 4 \ln n$ as $n \to \infty$.

Proof. As already noted, X_n is 1 if the pair of vertices chosen at step n have an edge between them and 0 otherwise. Since the method of choosing a vertex pair at step n is uniformly random then $X_n \sim Bernoulli(p_n)$ where p_n represents the ratio of edges to pairs of vertices in the graph. After the $n-1^{st}$ step, there are $n+2$ vertices and $2n+1$ edges in the graph, so we get

$$\begin{aligned} p_n &= \frac{2n+1}{\binom{n+2}{2}} \\ &= \frac{2(2n+1)}{(n+2)(n+1)}, \end{aligned}$$

as required.

For the second result,

$$\begin{aligned} \mathbb{E}(W_n) &= 1 + \sum_{i=1}^{n} p_i \\ &= 1 + \sum_{i=1}^{n} \frac{2(2n+1)}{((n+2)(n+1)} \\ &= 1 + \sum_{i=1}^{n} \frac{4}{i+2} - \sum_{i=1}^{n} \frac{2}{(i+2)(i+1)} \\ &\sim 4 \ln n \text{ as } n \to \infty, \end{aligned}$$

as required. ∎

Let us now consider the mean number of steps at which a new triangle is added to the graph. Define $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_k (= \frac{W_n}{n})$. From Proposition 12.6 we know that $\mathbb{E}(\overline{X}_n) \to 0$ as $n \to \infty$. We apply the Chernoff bounds from Lemma 9.8 to prove that $\overline{X}_n \xrightarrow{C} 0$ as $n \to \infty$.

Theorem 12.7. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent Bernoulli(p_n) random variables satisfying the condition that $\sum_{i=1}^n \frac{p_i}{n}$ converges to some limit μ as $n \to \infty$. Then, for any fixed $\varepsilon \in (0,1)$, there exists $k > 0$ such that for all sufficiently large n,*

$$\mathbb{P}\left(\left|\overline{X}_n - \frac{1}{n}\sum_{i=1}^n p_i)\right| > \mu n^{-\frac{1}{4}}\right) \leq e^{-\frac{\mu n^{\frac{1}{2}}}{6}},$$

and

$$\overline{X}_n \left(:= \frac{1}{n}\sum_{i=1}^n X_i\right) \xrightarrow{C} \mu \quad \textit{as } n \to \infty.$$

Proof. By Lemma 9.8 we know that, for every $\varepsilon \in (0,1]$,

$$\mathbb{P}\left(\left|\overline{X}_n - \frac{1}{n}\sum_{i=1}^n p_i)\right| > \frac{\varepsilon}{n}\sum_{i=1}^n p_i\right) \leq e^{-\frac{\varepsilon^2}{3}\sum_{i=1}^n p_i}.$$

Thus, taking $\varepsilon = n^{-\frac{1}{4}}$, we get

$$\mathbb{P}\left(\left|\overline{X}_n - \frac{1}{n}\sum_{i=1}^n p_i)\right| > n^{-\frac{5}{4}}\sum_{i=1}^n p_i\right) \leq e^{-\frac{n^{-\frac{1}{2}}}{3}\sum_{i=1}^n p_i}.$$

Now, $\frac{1}{n}\sum_{i=1}^n p_i \to \mu$ as $n \to \infty$, so for all $\delta > 0$ there exists $N_\delta \in \mathbb{N}$ such that whenever $n \geq N_\delta$ then $\left|\frac{1}{n}\sum_{i=1}^n p_i - \mu\right| < \delta$. Thus, for all $n \geq N_\delta$,

$$\mathbb{P}\left(\left|\overline{X}_n - \frac{1}{n}\sum_{i=1}^n p_i)\right| > \mu n^{-\frac{1}{4}}\right) \leq e^{-\frac{n^{\frac{1}{2}}(\mu-\delta)}{3}}.$$

Choosing $\delta = \frac{\mu}{2}$ gives the first part, while the second follows from the observation that $\sum_{n=1}^\infty e^{-\frac{\mu n^{\frac{1}{2}}}{6}} < \infty$. $\square$

We now bound the distance between the distribution of the number of triangles in our evolving Bernoulli random graph and the normal distribution. We use the so-called '*local approach*', following the example of Reinert [41], to apply Stein's method and obtain bounds in Wasserstein distance. Alternatively, the Berry-Esseen Theorem (for example see [21], Chapter 11) can be applied directly to obtain Kolmogorov distance bounds of the same order. Since the Wasserstein distance that we obtain is the stronger metric on $\mathbb{Z}$ (see [24]) and since the method is not particularly instructive, we omit this proof.

Since the Stein equations for the normal distribution given in Chapter 4 are for the standardised normal distribution, we begin by defining standardised versions of our random variables X_i and W_n. For $n \in \mathbb{N}$, $i = 1, 2, ..., n$ define

$$Y_i^{(n)} := \frac{X_i - p_i}{\sqrt{\sum_{j=1}^n p_j(1-p_j)}},$$

$$U^{(n)} := \sum_{i=1}^n Y_i.$$

The random variables $(Y_i^{(n)})_i$ inherit independence from $(X_i)_i$, so that $\mathbb{E}(U^{(n)}) = 0$ and $Var(U^{(n)}) = 1$. We shall bound the distance between $\mathcal{L}(U^{(n)})$ and the $N(0,1)$ distribution.

Proposition 12.8. *For $U^{(n)} := \sum_{i=1}^n Y_i^{(n)}$ defined for our model as above, h any absolutely continuous bounded function and $Z \sim N(0,1)$,*

$$|\mathbb{E}(h(U^{(n)})) - \mathbb{E}(h(Z))| \leq \frac{5||h'||_\infty}{(\sum_{j=1}^n p_j(1-p_j))^{\frac{1}{2}}}.$$

Proof. Fix n and let $U_i^{(n)} := U^{(n)} - Y_i^{(n)} = \sum_{j \neq i, 1 \leq j \leq n} Y_j^{(n)}$. By construction, $U_i^{(n)}$ and $Y_i^{(n)}$ are independent. Let h be an absolutely continuous bounded function and let f be the solution of the Stein equation given in Theorem 4.4. Applying Taylor's expansion,

$$
\begin{aligned}
&\mathbb{E}(U^{(n)}f(U^{(n)}))\\
= &\sum_{i=1}^{n}\mathbb{E}(Y_i^{(n)}f(U^{(n)}))\\
= &\sum_{i=1}^{n}\mathbb{E}\left(Y_i^{(n)}\left(f(U_i^{(n)})+(U^{(n)}-U_i^{(n)})f'(U_i^{(n)})\right)\right)\\
&+\sum_{i=1}^{n}\mathbb{E}\left(Y_i^{(n)}\frac{(U^{(n)}-U_i^{(n)})^2}{2}f''(U_i^{(n)}\ +\theta_i(U^{(n)}-U_i^{(n)}))\right),
\end{aligned}
$$

for some $\theta_i \in (0,1)$. Thus,

$$
\begin{aligned}
&\mathbb{E}(U^{(n)}f(U^{(n)}))\\
= &\sum_{i=1}^{n}\left(\mathbb{E}(Y_i^{(n)}f(U_i^{(n)}))+\mathbb{E}((Y_i^{(n)})^2f'(U_i^{(n)}))\right)\\
&+\sum_{i=1}^{n}\mathbb{E}\left(\frac{(Y_i^{(n)})^3}{2}f''(U_i^{(n)}+\theta_i(U^{(n)}-U_i^{(n)}))\right)\\
= &\sum_{i=1}^{n}\mathbb{E}((Y_i^{(n)})^2)\mathbb{E}(f'(U_i^{(n)}))+R_1\\
&\text{where } R_1=\sum_{i=1}^{n}\mathbb{E}\left(\frac{(Y_i^{(n)})^3}{2}f''(U_i^{(n)}+\theta_i(U^{(n)}-U_i^{(n)}))\right).
\end{aligned}
$$

Applying Taylor's expansion again,

$$
\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}((Y_i^{(n)})^2)\mathbb{E}(f'(U_i^{(n)}))\\
= &\sum_{i=1}^{n}\mathbb{E}((Y_i^{(n)})^2)\mathbb{E}\left(f'(U^{(n)})+(U_i^{(n)}-U^{(n)})f''(U_i^{(n)}+\tau_i(U^{(n)}-U_i^{(n)}))\right)\\
&\text{for some } \tau_i\in(0,1)\\
= &\sum_{i=1}^{n}\mathbb{E}((Y_i^{(n)})^2)\mathbb{E}(f'(U^{(n)}))+R_2\\
&\text{where } R_2=\sum_{i=1}^{n}\mathbb{E}(Y_i^2)\mathbb{E}\left((U_i^{(n)}-U^{(n)})f''(U_i^{(n)}+\tau_i(U^{(n)}-U_i^{(n)}))\right).
\end{aligned}
$$

So we have,

$$\begin{aligned}
\mathbb{E}(U^{(n)}f(U^{(n)})) &= \sum_{i=1}^{n}\mathbb{E}((Y_i^{(n)})^2)\mathbb{E}(f'(U^{(n)})) + R_1 + R_2 \\
&= \mathbb{E}(f'(U^{(n)}))\sum_{i=1}^{n}\mathbb{E}((Y_i^{(n)})^2) + R_1 + R_2 \\
&= \mathbb{E}(f'(U^{(n)})) + R_1 + R_2,
\end{aligned}$$

since

$$1 = Var(U^{(n)}) = \sum_{i=1}^{n} Var(Y_i) = \sum_{i=1}^{n}(\mathbb{E}((Y_i^{(n)})^2) - (\mathbb{E}(Y_i^{(n)}))^2) = \sum_{i=1}^{n}\mathbb{E}((Y_i^{(n)})^2).$$

If we can bound $R_1 + R_2$ we will have bounded the right-hand side of the Stein equation. Now,

$$\begin{aligned}
|R_1 + R_2| &\leq |R_1| + |R_2| \\
|R_1| &= \left|\sum_{i=1}^{n}\mathbb{E}\left(\frac{(Y_i^{(n)})^3}{2}f''(U_i^{(n)} + \theta_i(U^{(n)} - U_i^{(n)}))\right)\right| \\
&\leq \frac{1}{2}\parallel f'' \parallel \sum_{i=1}^{n}\mathbb{E}|(Y_i^{(n)})^3|, \qquad (12.2)
\end{aligned}$$

where $Y_i^{(n)}$ either takes value

$$\frac{1-p_i}{(\sum_{j=1}^{n} p_j(1-p_j))^{\frac{1}{2}}}$$

with probability p_i, or value

$$\frac{-p_i}{(\sum_{j=1}^{n} p_j(1-p_j))^{\frac{1}{2}}}$$

with probability $1 - p_j$. This gives,

$$\sum_{i=1}^{n} \mathbb{E}|(Y_i^{(n)})^3| \leq \frac{1}{(\sum_{j=1}^{n} p_j(1-p_j))^{\frac{1}{2}}}, \qquad (12.3)$$
$$\text{since } (1-p_i)^2 + (p_i)^2 \leq 1.$$

Similarly,

$$|R_2| \leq \| f'' \| \sum_{i=1}^{n} \mathbb{E}((Y_i^{(n)})^2)\mathbb{E}|Y_i^{(n)}|, \qquad (12.4)$$

and

$$\mathbb{E}|Y_i^{(n)}| = \frac{2p_i(1-p_i)}{(\sum_{j=1}^{n} p_j(1-p_j))^{\frac{1}{2}}},$$

so

$$\begin{aligned} \sum_{i=1}^{n} \mathbb{E}((Y_i^{(n)})^2)\mathbb{E}|Y_i^{(n)}| &= 2\sum_{i=1}^{n} \frac{{p_i}^2(1-p_i)^2}{(\sum_{j=1}^{n} p_j(1-p_j))^{\frac{3}{2}}} \\ &\leq \frac{2}{(\sum_{j=1}^{n} p_j(1-p_j))^{\frac{1}{2}}}. \qquad (12.5) \end{aligned}$$

Combining (12.2), (12.3), (12.4) and (12.5) we get

$$|R_1 + R_2| \leq \frac{5 \| f'' \|}{2(\sum_{j=1}^{n} p_j(1-p_j))^{\frac{1}{2}}}$$

or

$$|R_1 + R_2| \leq \frac{5||h'||}{(\sum_{j=1}^{n} p_j(1-p_j))^{\frac{1}{2}}} \text{ from Theorem 4.4.}$$

□

Corollary 12.9. *If* $Z \sim N(0,1)$, *then*

$$d_W(U^{(n)}, Z) \leq \frac{5}{\sqrt{\log(n+3) + \frac{3}{n+2} - \frac{3}{2}}}.$$

Proof. By the definition of Wasserstein distance on page ix, for $\mathcal{H} := \{h : \mathbb{R} \to \mathbb{R} : ||h'||_\infty \leq 1\}$, Proposition 12.8 gives that

$$\sup_{h\in\mathcal{H}} |\mathbb{E}(h(U_n)) - \mathbb{E}(h(Z))| \leq \frac{5}{(\sum_{j=1}^n p_j(1-p_j))^{\frac{1}{2}}}.$$

By Proposition 12.6, we know that $p_j = \frac{2(2j+1)}{(j+2)(j+1)}$. Thus,

$$\begin{aligned}
\sum_{j=1}^n p_j(1-p_j) &= \sum_{j=1}^n \frac{2(2j+1)}{(j+2)(j+1)}\left(1 - \frac{2(2j+1)}{(j+2)(j+1)}\right) \\
&= \sum_{j=1}^n \frac{2j(2j+1)(j-1)}{(j+2)^2(j+1)^2} \\
&\geq \sum_{j=1}^n \frac{(j-1)}{(j+2)^2} \\
&= \sum_{j=1}^n \frac{1}{j+2} - 3\sum_{j=1}^n \frac{1}{(j+2)^2} \\
&\geq \int_1^{n+1} \frac{1}{x+2}dx - 3\int_0^n \frac{1}{(x+2)^2}dx \\
&= \log(n+3) + \frac{3}{n+2} - \frac{3}{2}.
\end{aligned}$$

Substituting this into the previous expression gives the required result. □

12.2 An analogous continuous-time evolving Bernoulli random graph

In a similar vein to Chapter 3 we construct a continuous-time evolving Bernoulli random graph model based on the discrete-time model. In the case of the Barabási-Albert random graph, the continuous-time version of the model helped us to bound the rate of convergence of the degree distribution in the discrete-time version. Although we already have bounds for this rate in the discrete Bernoulli case, looking at the continuous-time model provides an instructive illustration of the generality of such comparisons.

12.2.1 The model

As in the continuous-time Barabási-Albert evolving random graph, the model consists of two independent sets of processes.

1. **Arrivals of new vertices into the graph.** The time between arrivals is exponentially distributed with rate $\frac{1}{\rho}\times$ number of vertices in the graph.
2. **Growth of edges for current vertices.** Each vertex gains new half-edges independently with rate 1.

To define a model that is comparable to the discrete-time Bernoulli random graph, we require all vertices to increase in degree at the same rate regardless of the degree of the vertex. Instead of a Yule process of Yule processes as in the Barabási-Albert case, we now have a Yule process of Poisson processes.

Notation

- Write S_i for the time between the start of the i^{th} and $(i+1)^{st}$ new processes. $S_i \sim Exponential(\frac{i}{\rho})$.

- Write $X_i(t)$ for the degree of vertex i at time t, so that, conditional on $t > S_1 + ... + S_{i-1}$, $(X_i(S_1 + ... + S_{i-1}) := 1$ and $X_i(t)$ is a Poisson process with rate 1.

As in Chapter 3 for the models discussed previously we define the concept of the '*age*' of a vertex.

Definition 12.10. *We say that vertex* 1 *has age* t *at time* t*, and that vertex* $i > 1$ *has age* t *at time* $S_1 + ... + S_{i-1} + t$*.*

12.2.2 Results

We have already seen that in the discrete case the asymptotic degree distribution of an evolving Bernoulli random graph is geometric. Following the example of Chapter 3, we now prove a result for the continuous version of the model. At a sequence of stopping times corresponding the times that new vertices enter the graph, the degree distribution is precisely geometric.

Theorem 12.11. *For* $n \in \mathbb{N}$*, let* I_n *be a uniform random variable on* $\{1, 2, ..., n\}$ *with* $I_n \perp (S_i)_{i \in \mathbb{N}}$*. Then for* $k = 0, 1, 2, 3, ...,$

$$\mathbb{P}(X_{I_n}(S_1 + ... + S_n) = k) = \frac{1}{1+\rho}\left(1 - \frac{1}{1+\rho}\right)^k.$$

Proof. By construction, since the degree distribution of a vertex depends only on its age, the distributions of $(S_1, S_2, S_3, ...)$ and $\{X_i(S_1 + ... + S_{i-1} + t)\}_{i \in \mathbb{N}}$.

By Theorem 3.14, $S_{I_n} + S_{I_n+1} + ... + S_n \sim Exponential(\frac{1}{\rho})$.

By definition, for $i = 1, 2, ..., n$, the random variable $X_i(S_1 + ... + S_{i-1} + t)$ is identically distributed to the Poisson process of rate 1, $X_1(t)$. It follows that

$$\begin{aligned}\mathbb{P}(X_{I_n}(S_1 + ... + S_n) = k) &= \mathbb{P}(X_1(S_{I_n} + ... + S_n) = k) \\ &= \mathbb{E}\left(\frac{e^{-U}U^k}{k!}\right),\end{aligned}$$

where $U \sim Exponential(\frac{1}{\rho})$. Integrating to find this expectation gives the required result.

$\square$

The similar proof in Chapter 3 hinged on the result that for $U_\rho \sim Exponential(\rho)$, we have $Negbinom(\alpha; e^{-U_\rho}) = \mu_\rho^\alpha$. The corresponding result in this instance is that for $U_{\frac{1}{\rho}} \sim Exponential(\frac{1}{\rho})$, we have $Poisson(U) = Geometric(\frac{1}{\rho+1})$.

Chapter 13

Concluding remarks

This thesis has studied a sequence of different classes of Barabási-Albert random graph models. In Chapter 6 we compared discrete-time and continuous-time preferential attachment models and found total variation distance bounds between their degree distributions. There is scope for statistical applications of the bounds that we obtain. For example, as explained in Section 1.2.3, the degree distribution bounds give an indication of a suitable range of degrees for consideration when using log-log plots to estimate the exponent of a power law.

In Chapters 7, 8 and 11 we used three different methods to prove that the asymptotic degree distributions of models in our scheme are scale-free. For two of the methods we gave total variation distance bounds. Each method applied to specific models within the scheme, but the bounds of Chapter 6 allowed us to extend the results to the other models.

The comparison of different types of preferential attachment models has provided insight into the scale-free asymptotic degree distributions seen in preferential attachment models. In particular, when studying Yule's model for evolution in Chapter 3 we were easily able to write down the exact degree distribution at a sequence of stopping-times. This motivated the derivation of a continuous-time preferential attachment model in Chapter 6. We bounded the distances between the degree distribution of this model and both the scale-free distribution found in Chapter 3 and the original Barabási-Albert random graph, Model BA defined in Chapter 2.

Having found a candidate asymptotic degree distribution in Chapter 3, we used the recurrence relation method of Szymański [53] to bound the expected proportion of vertices of a given degree in Model BA. The bounds from this method were of particular interest in Chapter 9, where we applied bounds on the proportion of vertices of a given degree to bound the degree sequence of Model BA.

The method in Chapter 9 for bounding the degree sequence of Model BA used the Azuma-Hoeffding inequality, following the example of Bollobás et al. [11]. The application was restricted to certain types of models, more precisely those models where the number of edges added at every time-step is bounded above by some constant. For Model BA we proved complete convergence to the μ_ρ^α-distribution, a stronger result than almost sure convergence.

For evolving random graphs where there is no constant that bounds the number of edges added at every time-step, but instead with the condition that vertex degrees evolve independently, we obtained similar results through the use of the Chernoff bounds. Again, we were able to prove complete convergence for the degree sequence.

Stein's method has been one key method through which we have obtained approximation bounds for the graph statistics studied in this thesis. The method has proved powerful once a target distribution is known, but is difficult to apply without a target distribution. An avenue for further study is the question of whether the method can be extended to allow us to work with a target 'type' of distribution. For example, could a general method be found that compares the tail-degree distribution with a general scale-free distribution? Within the current framework, we cannot hope to apply Stein's method unless we know both the asymptotic power order parameter and the normalising constant.

In Chapter 4 we derived a Stein equation for the μ_ρ^α-distribution and found good Stein factors (as demonstrated in Example 5.2.3). In Chapter 11, we applied this to an inhomogeneous Markov chain that represents the degree evolution of Model $\bar{\mathrm{I}}$. Through this we formally proved the assertion that, should the Markov chain converge, its asymptotic distribution would be the μ_ρ^α-distribution. It is hoped that further work might develop this method to derive total variation distance bounds using either this or another Stein equation for the μ_ρ^α-distribution.

Chapter 10 proved that the degree distribution of Model $\bar{\mathrm{I}}$ is identically distributed to a particular statistic of the Chinese restaurant process. Not only does this mean that we can deduce bounds for the distribution of this statistic, but making the comparison gave us a new understanding of Model $\bar{\mathrm{I}}$ as a inhomogeneous Markov chain. Examining related homogeneous Markov chains and their stationary distributions showed us an alternative way of obtaining the μ_ρ^α-distribution (a scale free distribution derived in Chapter 3) as a candidate for the asymptotic degree distribution of Model $\bar{\mathrm{I}}$.

Finally, in Chapter 12 we used methods that we earlier applied to preferential attachment models to study a Bernoulli evolving random graph. As well as using Stein's method for the geometric distribution to bound the degree distribution and degree sequence we used Stein's method for the normal distribution to bound the distance between the number of triangles in the graph and a normal distribution. The fact that we were able to apply the same methods that we used earlier in the thesis for this new model (and to even more statistics of the model) gives weight to our hope that these methods will prove successful in studying further properties of these and other evolving random graph processes.

Appendix A

Gamma and beta functions

The following are standard results for gamma and beta functions that are used frequently throughout the thesis. On page 5 we defined the notation for the gamma and beta functions (defined on $\mathbb{C}^+$ and $\mathbb{C}^+ \times \mathbb{C}^+$ respectively),

$$\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} dt;$$

$$B(x, y) := \int_0^1 t^{x-1} (1-t)^{y-1} dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

It is well-known (and easy to calculate) that

$$\Gamma(z+1) = z\Gamma(z). \tag{A.1}$$

The following ratios follow from (A.1):

$$\frac{B(x+1; y)}{B(x; y)} = \frac{x}{x+y}; \tag{A.2}$$

$$\frac{B(x; y+1)}{B(x; y)} = \frac{y}{x+y}; \tag{A.3}$$

$$\frac{B(x+1; y)}{B(x; y+1)} = \frac{x}{y}. \tag{A.4}$$

Appendix B

A list of random graphs

For reference we list the definitions for random graphs involved in the coupling construction of Chapter 6.

B.1 Discrete-time models

Model BA

Let $\alpha > 0$ be a real parameter and let $m \le m_0$ be integer parameters.

- **(BA1)** Start with the empty graph on m_0 vertices.
- **(BA2)** At step n select m distinct vertices from the graph in such a way that the vertex i with $D_{n-1}^{(BA)}(i) := (\alpha +$ in-degree of i at step $n-1)$ is chosen with probability proportional to $D_{n-1}^{(BA)}(i)$. More precisely,

$$\mathbb{P}(\text{vertex } i \text{ is chosen}|D_{n-1}^{(BA)}(i)) = \frac{mD_{n-1}^{(BA)}(i)}{(n-1)(m+\alpha)+m_0\alpha}.$$

  (Note that the conditioning here is not on the whole degree vector at step $n-1$.)
- **(BA3)** Attach the new vertex, $n + m_0$, to the m chosen vertices with m edges directed out of the new vertex.

B.1.1 Model I

The following definition is given on page 84.

- Start with the empty graph on m_0 vertices;
- At step n, independently for each vertex in the graph, select vertex i with conditional probability $\frac{mD_{n-1}^{(BA)}(i)}{(n-1)(m+\alpha)+m_0\alpha}$, where $D_{n-1}^{(I)}(i) := (\alpha+$ in-degree of i at step $n-1)$, for some $\alpha > 0$.
- Attach the new vertex $n+m_0$ to all the selected vertices with edges directed out of the new vertex.

B.1.2 Model $\bar{\text{I}}$

The following definition is given on 86.

- Start with one vertex;
- At step n, independently for each vertex, select vertex i with conditional probability $\frac{D_{n-1}^{(\bar{I})}(i)}{\rho n+\beta}$, where $D_{n-1}^{(\bar{I})}(i) := (\alpha +$ in-degree of i at step $n-1)$, $\rho > 1$, $\alpha > 0, \beta \geq 0$ satisfying $\alpha \leq \beta + \rho$. The condition $\alpha \leq \beta + \rho$ ensures that all probabilities lie between 0 and 1.
- Attach a new vertex $n+1$ to all the selected vertices with edges directed out of the new vertex. (The number of vertices selected at step n is random.)

B.2 Continuous-time models

B.2.1 Graph based on Yule's model for evolution

The following definition is given on page 38.

1. **Births of new vertices into the graphs.** The time between arrivals is exponentially distributed with rate, $\rho\times$ number of vertices in the graph.

2. **Growth of edges for current vertices.** Each vertex gains incident in-edges independently with rate proportional to $\alpha+$ in-degree.

We use the following notation for this model.

- We write S_i for the time between the births of the i^{th} and $(i+1)^{st}$ new vertices, so $S_i \sim Exponential(\rho i)$.

- We write $Y_i(t)$ for the in-degree of vertex i at time t. We set $Y_1(0) = 0$, and for $i \geq 2$ we set $Y_i(S_1 + ... + S_{i-1}) := 0$. For $i \geq 1$ and $t \geq S_1 + ... + S_{i-1}$, $Y_i(t)$ is defined to evolve as a Yule process with immigration rate α and birth rate 1.

B.2.2 Model Y

The following definition is given on page 88.

1. **Arrivals of new vertices into the graph.** The time between the arrival of vertex n and vertex $n+1$ is given by $s_n := \frac{1}{\rho n+\beta}$ for $\rho > 1$, $\beta \geq 0$.

2. **Growth of edges for current vertices.** Each vertex i gains incident in-edges independently with rate given by $D_n^{(Y)}(i) := \alpha+$ in-degree, for $\alpha > 0$.

In Section 6.2 we introduce the following alternative formulation of Model Y.

- Start with one vertex, 1, of 'degree' α;
- At step n add one new vertex, $n+1$, of 'degree' α to the graph.
- Conditional on the degree sequence,

$$D_{n-1}^{(Y)}(1) = d_{n-1}^{(Y)}(1), D_{n-1}^{(Y)}(2) = d_{n-1}^{(Y)}(2), ..., D_{n-1}^{(Y)}(n) = d_{n-1}^{(Y)}(n)$$

  of the first n vertices at step $n-1$, generate independent random variables and call them $E_n^{(Y)}(1), ..., E_n^{(Y)}(n)$ where $E_n^{(Y)}(i) \sim Negbinom(d_{n-1}^{(Y)}(i), e^{-\frac{1}{\rho n+\beta}})$. At step n add $E_n^{(Y)}(i)$ incident in-edges to vertex i.

Bibliography

[1] Adell, J. A., Jodrá, P, (2006). Exact Kolmogorov and total variation distances between some familiar discrete distributions, *Journal of Inequalities and Applications*, **2006**, p. 1-8.

[2] Athreya, K. B., Ney, P. E., (2004). *Branching processes*, Courier Dover Publications.

[3] Athreya, K. B., Ghosh, A. P., Sethuraman, S., (2008). Growth of preferential attachment random graphs via continuous time branching processes, *Proceedings Mathematical Sciences*, **85**, pp. 473-494.

[4] Azuma, K., (1967). Weighted sums of certain dependent random variables, *Tôhoku Mathematical Journal*, **3**, pp. 375-367.

[5] Barabási, A.-L., Albert, R., (1999). Emergence of scaling in random networks science, *Science*, **286**, pp. 509-512.

[6] Barbour, A. D., Chen, L. H. Y., (2005). *An introduction to Stein's method.* World Scientific Press, Singapore.

[7] Barbour, A. D., Chen, L. H. Y., (2005). *Stein's method and applications.* World Scientific Press, Singapore.

[8] Barbour, A. D., Holst, L., Janson, S., (1992). *Poisson approximation*, Oxford University Press.

[9] Bhamidi, S., (2007). Universal techniques to analyze preferential attachment trees: global and local analysis, *Preprint, http://www.stat.berkeley.edu/users/shanky/preferent.pdf.*

[10] Bollobás, B., Riordan, O., (2004). Coupling scale-free and classical random graphs, *Internet Mathematics*, **1**, pp. 215-225.

[11] BOLLOBÁS, B., RIORDAN, O., SPENCER, J., TUSNÁDY, G., (2001). The degree sequence of a scale-free random graph process, *Random Structures & Algorithms*, **18**, pp. 279-280.

[12] BROWN, T. C., XIA, A., (2001). Stein's method and birth-death processes *Annals of Applied Probability,* **29**, pp. 1373-1403.

[13] BUCKLEY, P. G., OSTHUS, D., (2004) Popularity based random graph models leading to a scale-free distribution, *Discrete Mathematics*, **282**, pp. 53-68.

[14] CALDARELLI, G., (2007). *Scale-free networks : complex webs in nature and technology*, Oxford University Press.

[15] CHEN, L. H. Y., (1975). Poisson approximation for dependent trials, *Annals of Probability*, **3**, pp. 534-545.

[16] CHEN, L. H. Y., SHAO, Q., (2005). Stein's method for normal approximation, in [6], pp. 1-59.

[17] COOPER, C., FRIEZE, A., (2003). A general model of web graphs, *Random Structures and Algorithms,* **22**, pp. 311-335.

[18] CHUNG, F., (2008). A whirlwind tour of random graphs, *Encyclopedia on Complex Systems,* Springer.

[19] DOROGOVTSEV, S. N., MENDES, J. F. F., SAMUKHIN, A. N., (2000). Structure of growing networks with preferential linking, *Physical Review Letters,* **85**, pp. 4633-4636.

[20] DRMOTA, M., GIMÉNEZ, O., NOY, M., (2008). Degree distribution in random planar graphs, *Proceedings of the Fifth Colloquium on Mathematics and Computer Science, DTMCS.*

[21] DASGUPTA, A., (2008). *Asymptotic theory of statistics and probability*, Springer.

[22] DURRETT, R., (2006). *Random graph dynamics*, Cambridge University Press.

[23] Eichelsbacher, P., Reinert, G., (2008). Stein's method for discrete Gibbs measures, *Annals of Applied Probability,* **18**, pp. 1588-1618.

[24] Gibbs, A. L., Su, F. E., (2002). On choosing and bounding probability metrics, *International Statistical Review*, **70**, pp. 419-435.

[25] Gut, A., (1995). *An intermediate course in probability*, Springer.

[26] Harris, T. D., (2002). *The theory of branching processes*, Courier Dover Publications.

[27] Hoeffding, W., (1963). Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association*, **58**, pp. 13-30.

[28] Holmes, S., (2004). Stein's method for birth and death chains *Stein's method: expository lectures and applications,* Diaconis, P., Holmes, S., *Lecture Notes - Monograph Series, Beachwood, Ohio, USA: Institute of Mathematical Statistics*, **46**.

[29] Hsu, P. L., Robbins, H., (1947). Complete convergence and the law of large numbers, *Proceedings of the National Academy of Sciences*, **33**, pp. 25-31.

[30] Jagers, P., (1975). *Branching processes with biological applications*, John Wiley & Sons, Ltd.

[31] Kallenberg, O., (2002). *Foundations of modern probability,* Second Edition, Springer-Verlag.

[32] Katona, Z., Móri, T., (2006). A new class of scale free random graphs, *Statistics and Probability Letters,* **76**, pp. 1587-1593.

[33] Krapivsky, P. L., Redner, S., (2001). Organisation of growing random networks, *Physical Review E,* **63**, 066123.

[34] Krapivsky, P. L., Redner, S., Leyvraz, F., (2000). Connectivity of growing random networks, *Physical Review Letters,* **85**, pp. 4629-4632.

[35] Mitzenmacher, M., (2004). A brief history of generative models for power law and lognormal distributions *Internet Mathematics*, **1**, pp. 215-225.

[36] Mitzenmacher, M., Upfal, E., (2005). *Probability and computing: randomized algorithms and probabilistic analysis,* Cambridge University Press.

[37] Newman, M., Barabási, A.-L., Watts, D. J., (2006). *The structure and dynamics of networks,* Princeton University Press.

[38] Peköz, E. A., (1996). Stein's method for geometric approximation, *Journal of Applied Probability,* **33**, pp. 707-713.

[39] Phillips, M. J., Weinberg, G. V., (2000). Non-uniform bounds for geometric approximation, *Statistics & Probability Letters.* **49**, pp. 305-311.

[40] Pitman, J., (2006). *Combinatorial stochastic processes*, Springer-Verlag.

[41] Reinert, G., (1998). Couplings for normal approximations with Stein's method *Microsurveys in Discrete Probability*, Aldous, D., Propp, J., *Dimacs series. AMS*, pp. 193-207.

[42] Riordan, O., (2005). The small giant component in scale-free random graphs, *Combinatorics, Probability and Computing archive,* **14**, pp. 897-938.

[43] Reinert, G., (2005). Three general approaches, in [6] pp. 183-221.

[44] Rose, C., Smith, M. D., (2002). *Mathematical sciences with Mathematica*, Springer.

[45] Ross, Sheldon M., Peköz, Erol A., (2007). *A second course in probability* www.ProbabilityBookstore.com.

[46] Ross, S. M., (1996). *Stochastic processes,* 2nd edition, John Wiley & Sons.

[47] Simon, H. A., (1955). On a class of skew distribution functions, *Biometrika*, **42**, pp. 425-440.

[48] Solow, S. R., (2005). Power laws without complexity, *Ecology Letters*, **2**, pp. 361-363.

[49] Solow, A. R., Costello, C. J., Ward, M., (2003). Testing the power law model for discrete size data, *The American Naturalist*, **162**, pp. 685-689.

[50] Stein, C., (1972). A bound for the error in normal approximation to the distribution of a sum of dependent random variables, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, **2**, pp. 583-602.

[51] Stein, C., (1986). *Approximate computation of expectations*, IMS Lecture Notes - Monograph Series, Institute of Mathematical Statistics, Hayward, California, **7**.

[52] Stein, C., with Diaconis, P., Holmes, S., Reinert, G., *Use of exchangeable pairs in the analysis of simulations*, IMS Lecture Notes - Monograph Series, Institute of Mathematical Statistics, Hayward, California, **46**.

[53] Szymański, J., (2005). *Concentration of vertex degrees in a scale-free random graph process*, *Random Structures and Algorithms*, **26**, pp. 224-236.

[54] Titchmarsh, E. C., (1932). *Theory of functions*, Oxford University Press.

[55] Weinberg, G. V., (2005). *Stein's method and its application in radar signal Processing,* DSTO Systems Sciences Laboratory, Commonwealth of Australia.

[56] Yule, G. U., (1925). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S., *Philosophical Transactions of the Royal Society of London*, **213**, pp. 21-87.