

# Tail Flexibility in the Degrees of Preferential Attachment Networks

Thomas William Boughen<sup>1</sup>, Clement Lee<sup>1</sup>,  
Vianey Palacios Ramirez<sup>1</sup>

<sup>1</sup>School of Mathematics, Statistics and Physics, Newcastle University.

## Abstract

Devising the underlying generating mechanism of a real-life network is difficult as, more often than not, only its snapshots are available, but not its full evolution. One candidate for the generating mechanism is preferential attachment which, in its simplest form, results in a degree distribution that follows the power law. Consequently, the growth of real-life networks that roughly display such power-law behaviour is commonly modelled by preferential attachment. However, the validity of the power law has been challenged by the presence of alternatives with comparable performance, as well as the recent findings that the right tail of the degree distribution is often lighter than implied by the body, whilst still being heavy. In this paper, we study a modified version of the model with a flexible preference function that allows super/sub-linear behaviour whilst also guaranteeing that the limiting degree distribution has a heavy tail. We relate the distributions tail index directly to the model parameters, allowing direct inference of the parameters from the degree distribution alone.

**Keywords:** networks, discrete extremes, power law, preferential attachment

# 1 Introduction

Networks appear in many fields such as sociology, politics, epidemiology, and economics. Statistical methods have been used to study networks that appear in these fields and beyond, from the use of stochastic block models to detect communities within the French political blogosphere (Latouche et al, 2011), to using Exponential Random Graph Models (ERGMs) to analyse the global trade network (Setayesh et al, 2022), and the use of mechanistic models in neuroscience to investigate the patterns of connections in neural systems (Betz et al, 2017).

One example of mechanistic models is the preferential attachment (PA) model, where as a network grows and new vertices join, they connect to those already present in the network with probabilities proportional to some preference function of their in-degrees. The well-known Barabási-Albert (BA) model is a special case of this when  $b(k) = k + 1$  (Barabási and Albert, 1999).

These mechanistic models are generally useful when simulating networks, but not for fitting to real data as usually only snapshots of a network are available and not the full evolution - without this conclusions about the growth are difficult to make. However, one statistic that is usually available for snapshots of networks is the degree distribution, making it valuable to study the degrees in a network. Often, it is appealing to model the degree distributions of real networks using the power law<sup>1</sup>, as it is seemingly observed in much of the data. Furthermore, as the BA model results in a power law degree distribution (Barabási and Albert, 1999), that the power law fits well offers the interpretation that the network may have grown according to PA or similar simple rules. However, attempting to explain the networks growth using preferential attachment when the degrees seemingly follow a power law may not be the best course of action as Krapivsky and Redner (2001) show that a general preference function does not necessarily lead to power law degrees. Specifically, showing that a preference function of form  $b(k) = k^\alpha$  leads to light tails when  $\alpha < 1$ , a degenerate network (a finite number of vertices end up with all of the added edges) when  $\alpha > 1$ , and only yields heavy tails when  $\alpha = 1$ . Rudas et al (2007) derives the limiting degree distribution in terms of the preference function  $b(\cdot)$  and its parameters, we build upon this by finding a class of preference functions that result in realistic tails i.e. heavy but not as heavy as implied by the power law. On the other hand, it is still being argued whether or not many real data actually follow the power law. Broido and Clauset (2019) provides some evidence that real world networks rarely have degrees that follow a power law, through the use of statistical analysis of almost 1000 networks across a wide variety of fields; Voitalov et al (2019), on the other hand, disagrees and claims that these kinds of networks are not nearly as rare and only appear so as a result of an unrealistic expectation of power law without deviations or noise.

Extreme value methods have been scarcely used when it comes to analysis of degree distributions and whether they follow the power law. The absence of these methods essentially downweights the deviation of vertices with the largest degrees, which are

---

<sup>1</sup>Power law:  $\bar{F}(k) = L(k) \times k^{-1/\xi}$ , where  $\xi > 0$  and  $L(k)$  is slowly varying in  $k$ .

usually the most influential, from the power law. Voitalov et al (2019) proposes estimating the power law exponent of a degree distribution using three estimators based in extreme value theory, in contrast to the estimators generally used in network science, the degree distributions are then split into one of three categories (not power-law, hardly power-law or power-law) depending on the values of the estimates obtained from each of the estimators. Lee et al (2024) proposes a framework for modelling entire degree distributions using a combination of two generalisations of the power law with a model selection step that is capable of suggesting whether or not the data is adequately modelled by a power law, with one conclusion being that most data seems to have a tail lighter than what is implied by the power law in body whilst still being heavy tailed. One of the models used by Lee et al (2024) is the Zipf-IGP model where up until a certain threshold the data is modelled using the Zipf distribution (a discrete power law) and beyond it a discrete version of the Generalised Pareto distribution is used, we will later compare our model to this as they share some similarities. Wang and Resnick (2022) considers a preferential attachment model with an affine preference function and reciprocal edges, finding that the joint limiting distribution of the in-degrees and out-degrees is multivariate regularly varying and has the property of hidden regular variation.

Evidently, the study of degree distributions is a very active but the vast majority of research in the field (whether using extremes or not) are purely descriptive in the sense that no information about the preference function is revealed even if preferential attachment was the underlying mechanism. This paper aims to address this gap, asking if the degree distribution of a network is assumed to come from the preferential attachment model, can we directly infer the model parameters?

The remainder of this paper is as follows: Section 2 gives a detailed description of the PA model alongside the theoretical results for the limiting degree distribution with a focus on its tail behaviour in terms of the preference function  $b(\cdot)$ . It is shown that a heavy tailed degree distribution can only be the result of an asymptotically linear preference function ending with a suggestion for a preference function that can guarantee heavy tails while remaining flexible. Section 3 utilises the preference function proposed at the end of Section 2, and illustrates how the tail index of the degree distribution varies with the model parameters. Section 4.1 consists of a simulation study where networks are simulated from the proposed model for various parameter combinations, demonstrating that the parameters can be recovered from fitting the model to only the degree distribution. Section 4.2 fits the model to some real data and provides posterior estimates for the preference function. Section 5 concludes the article.

## 2 Tail Behaviour of Preferential Attachment Model

The network generative model that we will focus on in this paper is the General Preferential Attachment model in [Rudas et al \(2007\)](#) and is defined as follows:

Starting at time  $t = 0$  with an initial network of  $m$  vertices that each have no edges, at times  $t = 1, 2, \dots$  a new vertex is added to the network bringing with it  $m$  directed edges (with the new vertex as the source); the target for each of these edges are selected from the vertices already in the network with weights proportional to some preference function  $b(\cdot)$  of their degree, where  $b : \mathbb{N} \mapsto \mathbb{R}^+ \setminus \{0\}$  is such that:

$$\sum_{k=0}^{\infty} \frac{1}{b(k)} = \infty. \quad (1)$$

Special cases of this model include the Barabási-Albert (BA) model when  $b(k) = k + \varepsilon$ , which leads to a power-law degree distribution with  $\xi = 1/2$ , and the Uniform Attachment (UA) model where  $b(k) = c$  leading to a degree distribution that is light tailed.

Given condition 1 the survival function of the limiting degree distribution, called the limiting survival hereafter, can be analytically derived in the case where  $m = 1$ , obtained by considering a branching process that is equivalent to the growth of the network, as in [Rudas et al \(2007\)](#). Theorem 1 from [Rudas et al \(2007\)](#) states that for the tree  $\Upsilon(t)$  at time  $t$ :

$$\lim_{t \rightarrow \infty} \frac{1}{|\Upsilon(t)|} \sum_{x \in \Upsilon(t)} \varphi(\Upsilon(t)_{\downarrow x}) = \lambda^* \int_0^{\infty} e^{-\lambda^* t} \mathbb{E}[\varphi(\Upsilon(t))] dt \quad (2)$$

where  $\lambda^*$  satisfies  $\hat{\rho}(\lambda^*) = 1$  and  $\hat{\rho}$  is the Laplace transform of the density of the point process associated with the pure birth process that corresponds to a nodes individual growth, that is:

$$\hat{\rho}(\lambda) := \int_0^{\infty} e^{-\lambda t} \rho(t) dt.$$

The limiting survival can be viewed as the limit of the empirical proportion of vertices with degree over a threshold  $k \in \mathbb{N}$ , that is:

$$\bar{F}(k) = \lim_{t \rightarrow \infty} \frac{\sum_{x \in \Upsilon(t)} \mathbb{I}\{\deg(x, \Upsilon(t)_{\downarrow x}) > k\}}{\sum_{x \in \Upsilon(t)} 1}$$

which can also be written using Equation 2 as:

$$\bar{F}(k) = \frac{\int_0^{\infty} e^{-\lambda^* t} \mathbb{E}[\mathbb{I}\{\deg(x, \Upsilon(t)) > k\}] dt}{\int_0^{\infty} e^{-\lambda^* t} dt} = \prod_{i=0}^k \frac{b(i)}{\lambda^* + b(i)}.$$

Therefore, the corresponding probability mass function of the degree distribution  $f(k) = \bar{F}(k-1) - \bar{F}(k)$  is

$$f(k) = \frac{\lambda^*}{\lambda^* + b(k)} \prod_{i=0}^{k-1} \frac{b(i)}{\lambda^* + b(i)}.$$

We are interested in how the tail behaviour of the discrete limiting degree distribution is affected by the preference function  $b$ , which can be determined by the results of [Shimura \(2012\)](#).

For a distribution  $F$  with survival function  $\bar{F}$  and some  $k \in \mathbb{Z}^+$  let

$$\Omega(F, k) = \left( \log \frac{\bar{F}(k+1)}{\bar{F}(k+2)} \right)^{-1} - \left( \log \frac{\bar{F}(k)}{\bar{F}(k+1)} \right)^{-1}.$$

[Shimura \(2012\)](#) states that if  $\lim_{k \rightarrow \infty} \Omega(F, k) = 1/\alpha$  ( $\alpha > 0$ ), then  $F$  is heavy-tailed with  $\bar{F}(k) \sim k^{-\alpha}$ . On the other hand, if  $\lim_{k \rightarrow \infty} \Omega(F, k) = 0$  then the distribution is light-tailed. This allows us to show the following:

**Proposition 2.1.** *If  $\bar{F}(k) = \prod_{i=0}^k \frac{b(i)}{\lambda^* + b(i)}$  and  $b(k) \rightarrow \infty$  as  $k \rightarrow \infty$ , then*

$$\lim_{k \rightarrow \infty} \Omega(F, k) = \lim_{k \rightarrow \infty} \frac{b(k+1) - b(k)}{\lambda^*}.$$

See Appendix B for the details of the proof.

Proposition 2.1 aligns with the result from [Krapivsky and Redner \(2001\)](#) demonstrating that a sub-linear preference function will lead to a light-tailed distribution, as  $\lim_{k \rightarrow \infty} b(k+1) - b(k) = 0$  if  $b(k) = k^\alpha$  where  $\alpha < 1$ . Proposition 2.1 also aligns with the fact that BA model produces a heavy-tailed degree distribution with index  $\xi = 0.5$  by considering the preference function  $b(k) = k + \varepsilon$ , as  $\lim_{k \rightarrow \infty} b(k+1) - b(k) = 1$  leaving the tail index to be  $1/\lambda^*$  which using  $\hat{\rho}$  can be found to be  $1/2$ . So, in order for the degree distribution to be heavy-tailed we need that the limit  $\lim_{k \rightarrow \infty} b(k+1) - b(k)$  exists and is positive. To determine the class of functions that will result in heavy-tailed limiting degree distributions, we use the following result:

**Theorem 2.1** (Stolz-Cesàro Theorem ([Cesàro, 1888](#))). *Let  $(a_k)_{k \geq 1}$  and  $(b_k)_{k \geq 1}$  be two sequence of real numbers. Assume  $(a_k)_{k \geq 1}$  is a strictly monotone and divergent sequence and the following limit exists:*

$$\lim_{k \rightarrow \infty} \frac{b_{k+1} - b_k}{a_{k+1} - a_k} = l.$$

*Then,*

$$\lim_{k \rightarrow \infty} \frac{b_k}{a_k} = l.$$

**Proposition 2.2.** *Consider a GPA model with preference function  $b(\cdot)$  satisfying Proposition 2.1. Then the limiting degree distribution has a heavy tail with index  $c/\lambda^*$  if and only if  $\lim_{k \rightarrow \infty} b(k)/k = c > 0$ .*

*Proof*

From Proposition 2.1, we have that:

$$\lim_{k \rightarrow \infty} [b(k+1) - b(k)] = c > 0.$$

Now, setting  $b_k = b(k)$  and  $a_k = k$ :

$$\lim_{k \rightarrow \infty} [b(k+1) - b(k)] = \lim_{k \rightarrow \infty} \frac{b_{k+1} - b_k}{a_{k+1} - a_k} = c > 0,$$

meaning by Theorem 2.1:

$$\lim_{k \rightarrow \infty} \frac{b_k}{a_k} = \lim_{k \rightarrow \infty} \frac{b(k)}{k} = c > 0 \quad \square.$$

Considering Proposition 2.2 we propose the following preference function:

$$b(k) = \begin{cases} k^\alpha + \varepsilon, & k < k_0, \\ k_0^\alpha + \varepsilon + \beta(k - k_0), & k \geq k_0 \end{cases} \quad (3)$$

for  $\alpha, \beta, \varepsilon > 0$  and  $k_0 \in \mathbb{N}$ .

This preference function, as per Proposition 2.1, will produce a degree distribution with tail index  $\beta/\lambda^*$  guaranteeing a heavy tail. We study this preference function further in the next section.

### 3 Preferential Attachment with Flexible Heavy Tail

In the previous section, we found that using a preference function with linearity in the limit allows for the inclusion of sub/super-linear behaviour below the threshold, while simultaneously guaranteeing a heavy tail. In this section, we study how the limiting degree distribution varies with the preference function according to Equation 3. Its survival is

$$\bar{F}(k) = \begin{cases} \prod_{i=0}^k \frac{i^\alpha + \varepsilon}{\lambda^* + i^\alpha + \varepsilon}, & k < k_0, \\ \left( \prod_{i=0}^{k_0-1} \frac{i^\alpha + \varepsilon}{\lambda^* + i^\alpha + \varepsilon} \right) \frac{\Gamma(\lambda^* + k_0^\alpha + \varepsilon)/\beta}{\Gamma((k_0^\alpha + \varepsilon)/\beta)} \frac{\Gamma(k - k_0 + 1 + \frac{k_0^\alpha + \varepsilon}{\beta})}{\Gamma(k - k_0 + 1 + \frac{\lambda^* + k_0^\alpha + \varepsilon}{\beta})}, & k \geq k_0, \end{cases} \quad (4)$$

with  $\lambda^*$  satisfying  $\hat{\rho}(\lambda^*) = 1$  where

$$\hat{\rho}(\lambda) = \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{i^\alpha + \varepsilon}{\lambda + i^\alpha + \varepsilon} + \left( \frac{k_0^\alpha + \varepsilon}{\lambda - \beta} \right) \prod_{i=0}^{k_0-1} \frac{i^\alpha + \varepsilon}{\lambda + i^\alpha + \varepsilon} \quad (5)$$

which must be solved numerically for most parameter choices. Also, note that  $\lambda > \beta$ .

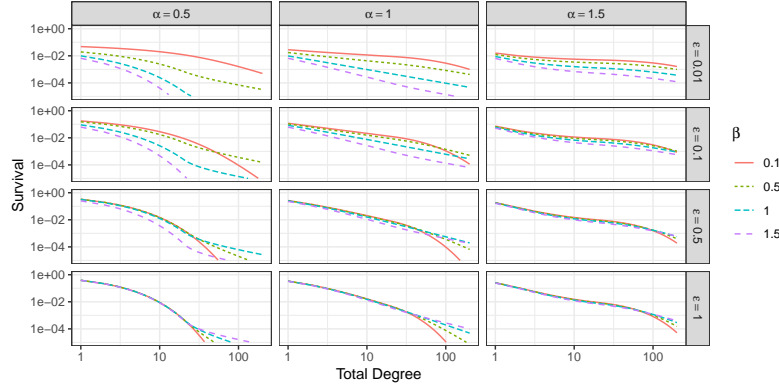
Some examples of the limiting degree distribution for various parameter combinations are shown below on log-log scale in Figure 1:

```

a      b      eps k0 x      lambda      surv
10 0.5 1.5 0.01 20 0 1.500094 0.006622105

[1] 0.9999374

```



**Figure 1:** Theoretical survival distributions of the limiting degree distributions, according to various combinations of  $(\alpha, \beta, \varepsilon)$  and  $k_0 = 20$  of the proposed preferential attachment model.

Figure 1 demonstrates that this model can capture a wide variety of shapes for the survival function, including a large range of possible tail indices ranging from 0.035 ( $\alpha = 1.5, \beta = 0.1, \varepsilon = 1$ ) to 0.999 ( $\alpha = 0.5, \beta = 1.5, \varepsilon = 0.01$ ).

The survival function (4) can be connected to the discrete version of the generalised Pareto distribution (GP), called the Integer GP (IGP) (Rohrbeck et al, 2018) with conditional survival:

$$\Pr(X > x | X > v) = \left( \frac{\xi(x-v)}{\sigma} + 1 \right)^{-1/\xi}, \quad x = v+1, v+2, \dots$$

for  $v \in \mathbb{Z}^+, \sigma > 0, \xi \in \mathbb{R}$ , denoted as  $X | X > u \sim \text{IGP}(\xi, \sigma, u)$  where  $\xi$  is the shape parameter controlling the tail index

By Equation 4 and using Stirling's approximation:

$$\begin{aligned} \bar{F}(k | k \geq k_0) &= \frac{\Gamma\left(\frac{\lambda^* + k_0^\alpha + \varepsilon}{\beta}\right)}{\Gamma\left(\frac{k_0^\alpha + \varepsilon}{\beta}\right)} \times \frac{\Gamma\left(k - k_0 + 1 + \frac{k_0^\alpha + \varepsilon}{\beta}\right)}{\Gamma\left(k - k_0 + 1 + \frac{\lambda^* + k_0^\alpha + \varepsilon}{\beta}\right)} \\ &\approx \left(\frac{k_0^\alpha + \varepsilon}{\beta}\right)^{\lambda^*/\beta} \left(k - k_0 + 1 + \frac{k_0^\alpha + \varepsilon}{\beta}\right)^{-\lambda^*/\beta} \\ &= \left(\frac{k_0^\alpha + \varepsilon}{k_0^\alpha + \varepsilon + \beta}\right)^{\lambda^*/\beta} \left(\frac{\beta(k - k_0)}{\beta + k_0^\alpha + \varepsilon} + 1\right)^{-\lambda^*/\beta} \\ &= \left(\frac{\beta(k + 1 - k_0)}{k_0^\alpha + \varepsilon} + 1\right)^{-\lambda^*/\beta} \end{aligned}$$

Therefore,

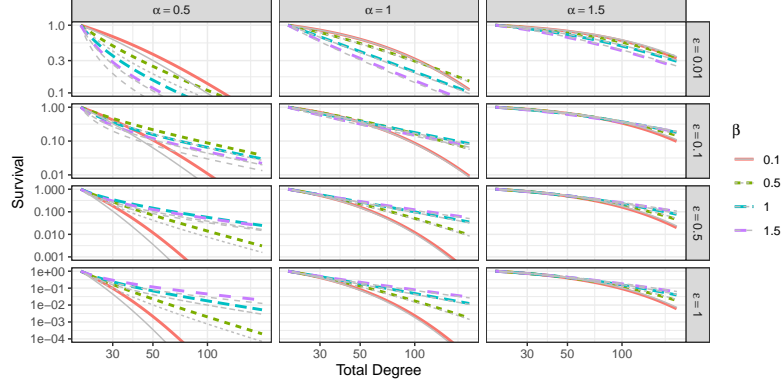
$$\bar{F}(k) \begin{cases} = \prod_{i=0}^k \frac{i^\alpha + \varepsilon}{\lambda^* + i^\alpha + \varepsilon}, & k < k_0, \\ \approx \left(\prod_{i=0}^{k_0-1} \frac{i^\alpha + \varepsilon}{\lambda^* + i^\alpha + \varepsilon}\right) \left(\frac{\beta(k+1-k_0)}{k_0^\alpha + \varepsilon} + 1\right)^{-\lambda^*/\beta}, & k \geq k_0, \end{cases}$$

meaning that for  $k \geq k_0$  the limiting degree distribution (for large  $k_0^\alpha$ ) is approximated by  $\text{IGP}\left(\frac{\beta}{\lambda^*}, \frac{k_0^\alpha + \varepsilon}{\lambda^*}, k_0 - 1\right)$ .

To assess how close of an approximation this is, the theoretical conditional survivals are shown in Figure 2 in colour and their IGP approximations are shown in grey. The approximation seems to hold up fairly well even for large degrees and more so when  $\alpha$  is larger.

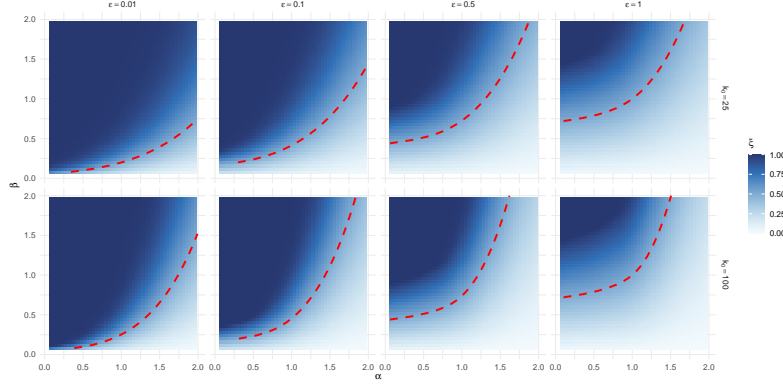
In agreement with Proposition 2.2,  $\beta > 0$  ensures that the shape parameter of the IGP is positive and thus the distribution is heavy-tailed. Additionally the value of the shape parameter  $\xi$  is shown in Figure 3 for various parameter choices. The darker and lighter regions on the heat maps correspond to a heavier and a lighter tail, respectively,





**Figure 2:** Theoretical conditional survivals (grey) alongside their IGP approximations (coloured).

the red dashed line shows combinations of  $\alpha$  and  $\beta$  that produce a limiting degree distribution with the same tail index as the Barabási-Albert model, that is,  $\xi = 0.5$ .



**Figure 3:** Heat maps of  $\xi$  for various combinations of the parameters of the proposed model.

To perform inference of the model parameters, we consider a network with degree count vector  $\mathbf{n} = (n_0, n_1, \dots, n_M)$ , where  $M$  is the maximum degree. We can then write the likelihood as:

$$\begin{aligned} L(\mathbf{x}, \mathbf{n} | \boldsymbol{\theta}) &= \left( \frac{\lambda^*}{\lambda^* + \varepsilon} \right)^{n_0} \left( \prod_{j=l}^{k_0-1} \frac{j^\alpha + \varepsilon}{\lambda^* + j^\alpha + \varepsilon} \right)^{\left( \sum_{i \geq k_0} n_i \right)} \\ &\times \prod_{l \leq i < k_0} \left( \frac{\lambda^*}{\lambda^* + i^\alpha + \varepsilon} \prod_{j=l}^{k_0-1} \frac{j^\alpha + \varepsilon}{\lambda^* + j^\alpha + \varepsilon} \right)^{n_i} \\ &\times \prod_{i \geq k_0} \left( \frac{B(i - k_0 + (k_0^\alpha + \varepsilon)/\beta, 1 + \lambda^*/\beta)}{B((k_0^\alpha + \varepsilon)/\beta, \lambda^*/\beta)} \right)^{n_i} \end{aligned}$$

where  $B(y, z)$  is the beta function, and  $l \geq 0$  is a variable that allows truncating the data such that the minimum degree is  $l$ . This will allow the model to be fitted whilst ignoring the influence of the lower degrees (those less than  $l$ ) as the model does not capture the behaviour well at the lower degrees since [Rudas et al \(2007\)](#) only provides results for the case of a preferential attachment tree.

## 4 Applications

### 4.1 Simulated Data

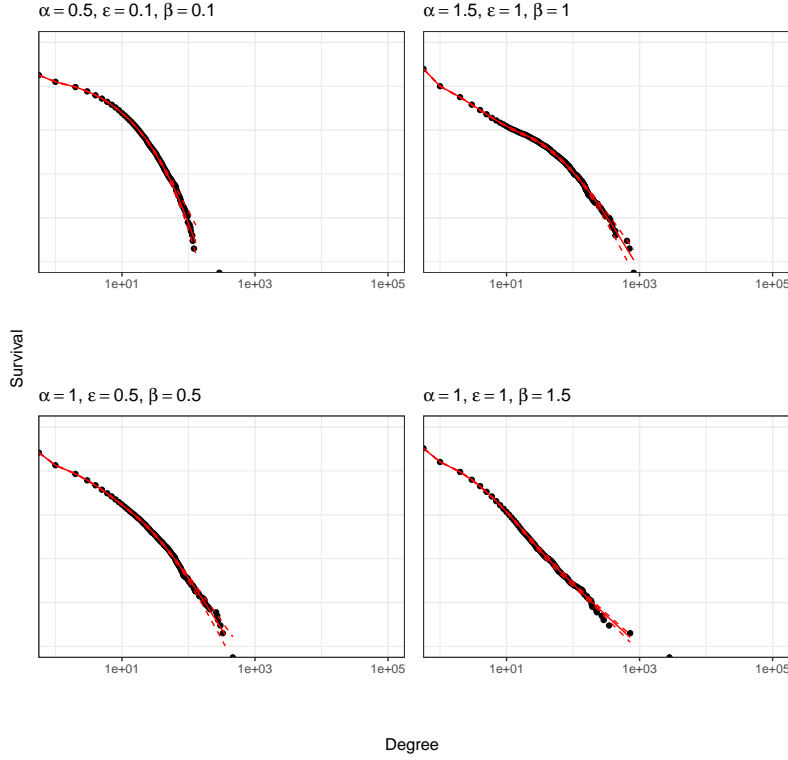
This first subsection aims to show that the parameters of the model in Section 3 can be recovered from simulating a network from the model, and fitting it to the observed degree distribution, using the likelihood in (3).

The procedure for recovering the parameters begins with simulating a network from the model with  $N = 100,000$  vertices and  $m = 1$  given some set of parameters  $\boldsymbol{\theta} = (\alpha, \beta, \varepsilon, k_0)$ , obtaining the degree counts and using the likelihood from the previous section alongside the priors:

$$\begin{aligned} \alpha &\sim \text{Ga}(1, 0.01), \\ \beta &\sim \text{Ga}(1, 0.01), \\ k_0 &\sim \text{U}(1, 10,000), \\ \varepsilon &\sim \text{Ga}(1, 0.01), \end{aligned}$$

where  $\text{Ga}(a, b)$  is the Gamma distribution with shape  $a$  and rate  $b$ , and  $\text{U}(a, b)$  is uniform distribution with lower and upper bounds  $a$  and  $b$ , to obtain a posterior distribution, up to the proportionality constant. Posterior samples can then be obtained

by an adaptive Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm. For these simulated networks  $l = 0$ .

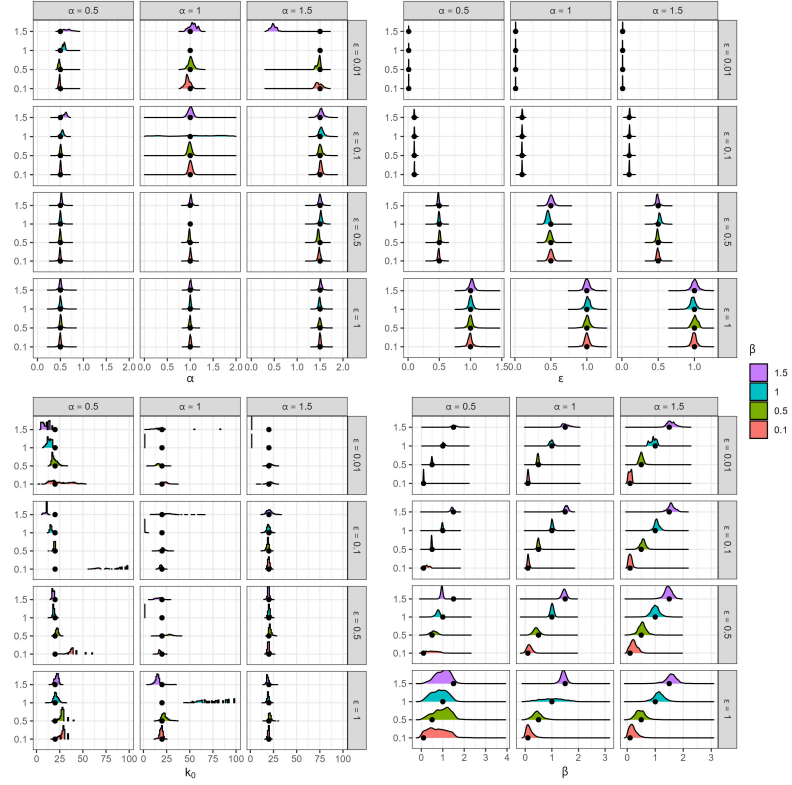


**Figure 4:** Posterior estimates of survival function for data simulated from the proposed model with various combinations of  $(\alpha, \beta, \epsilon)$  and  $k_0 = 20$ .

Figure 4 and Figure 5 shows the estimations demonstrates the usefulness of the model, as we can recover the model parameters fairly well from only the final degree distribution of the simulated network. This indicates that the method may also be applied to the degree distributions of real networks, estimating the model parameters assuming they evolved according to the GPA scheme.

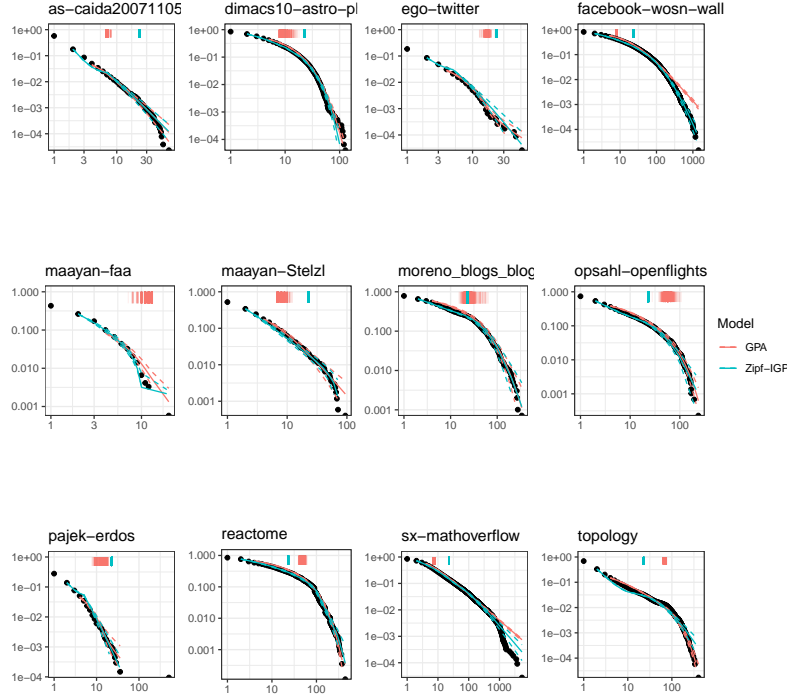
## 4.2 Real Data

In this section, we fit the proposed model to the degree distributions of various real networks and learn about the mechanics of their growth. While we also compare the fit to that of the mixture distribution by Lee et al (2024) we note that the proposed method has the additional benefit of learning directly about the growth of a network from the inference results. The data consists of 12 networks sourced from KONECT and the Network Data Repository (Rossi and Ahmed, 2015):



**Figure 5:** Posterior estimates of paramters for data simulated from the proposed model with various combinations of  $(\alpha, \beta, \epsilon)$  and  $k_0 = 20$ .

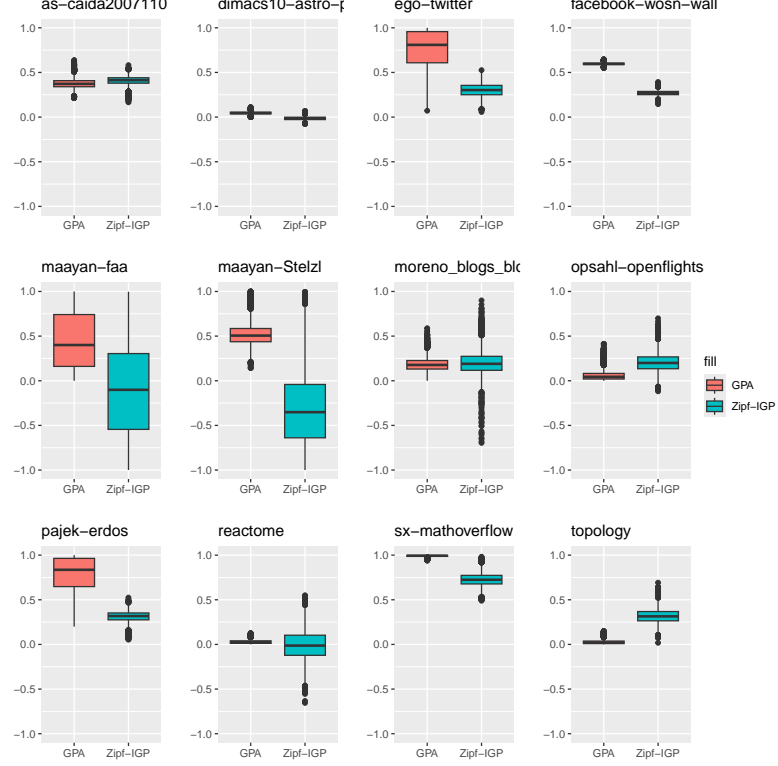
- **as-caida20071105:** network of autonomous systems of the Internet connected with each other from the CAIDA project
- **dimacs10-astro-ph :** co-authorship network from the “astrophysics” section (astro-ph) of arXiv
- **ego-twitter:** network of twitter followers
- **facebook-wosn-wall:** subset of network of Facebook wall posts
- **maayan-faa:** USA FAA (Federal Aviation Administration) preferred routes as recommended by the NFDC (National Flight Data Center)
- **maayan-Stelzl:** network representing interacting pairs of proteins in Humans
- **moreno-blogs-blogs:** network of URLs found on the first pages of individual blogs
- **opsahl-openflights:** network containing flights between airports of the world.
- **pajek-erdos:** co-authorship network around Paul Erdős
- **reactome:** network of protein–protein interactions in Humans
- **sx-mathoverflow:** interactions from the StackExchange site [MathOverflow](#)
- **topology:** network of connections between autonomous systems of the Internet



**Figure 6:** Posterior estimates (solid red) of survival for several real data sets and their 95% credible intervals (dotted red).

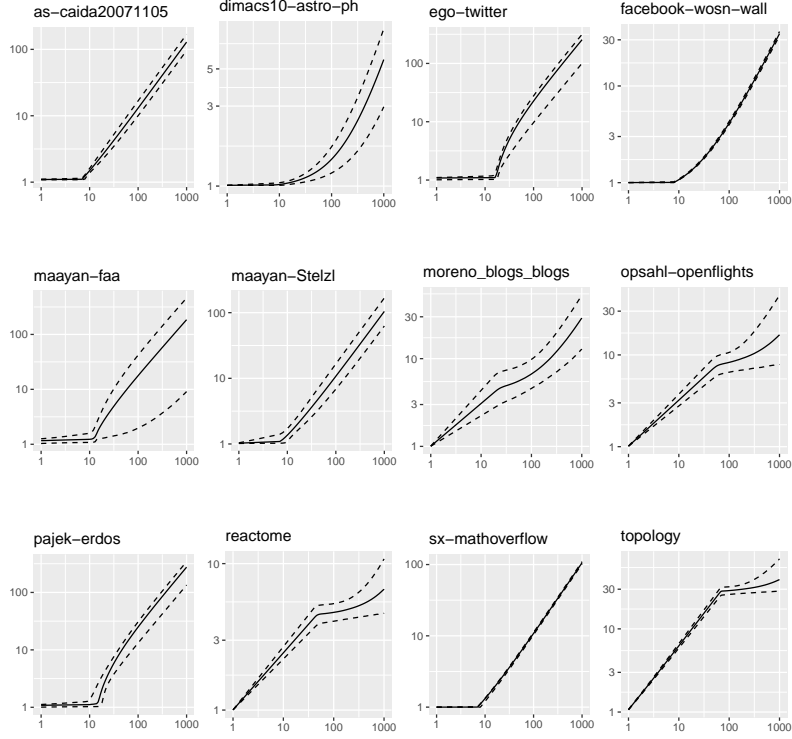
Figure 6 displays the posterior estimates of the survival function for various data sets, obtained from fitting the GPA model and the Zipf-IGP mixture model from [Lee et al \(2024\)](#). In most cases, the GPA model does not necessarily provide an improvement in fit when compared to the Zipf-IGP model but where the GPA model fits well we gain additional information about the preference function assuming that the network evolved according to the GPA scheme.

Figure 7 shows the posterior of the shape parameter  $\xi$  obtained from the Zipf-IGP model alongside the posterior of the equivalent shape parameter  $\beta/\lambda^*$  obtained from fitting the GPA model. Generally, the GPA model performs similarly to the Zipf-IGP when estimating the tail behaviour of the degree distribution. In the cases of substantial discrepancies, it is either because the GPA model fits the tail better than the Zipf-IGP model does, or because of the threshold being estimated as too low forcing almost all of the data to be modelled by the linear part of the GPA. This again shows the effects that small degrees have on this model, which is somewhat expected as the theory used for this model is for trees and none of these real networks (nor many real networks) are.



**Figure 7:** Posterior estimates (solid red) of survival for several real data sets and their 95% credible intervals (dotted red).

Figure 8 shows the estimated preference function  $b(k)$  alongside the 95% credible interval on a log-log plot. Although the credible interval becomes very large for the largest degrees, this is expected as not all of these networks had data in that region, and for those that do the credible interval is much narrower, as is the case for **sx-mathoverflow**. Looking at the shape of the preference function, there appears to be two distinct shapes of preference function. The first appears mostly flat (similar to uniform attachment) for the smallest degrees and then after a threshold preferential attachment kicks in, some with this shape are **pajek-erdos** and **sx-mathoverflow**. The second distinct shape appears to provide some clear preferential attachment behaviour that then slows down after a certain point, examples of this are seen in the two infrastructure networks **opsahl-openflights** and **topology**. This slowing down could be viewed as a kind of diminishing returns on the degree of a vertex i.e. as a vertex gets larger gaining more connections has less of an effect than it did before some threshold  $k_0$ .



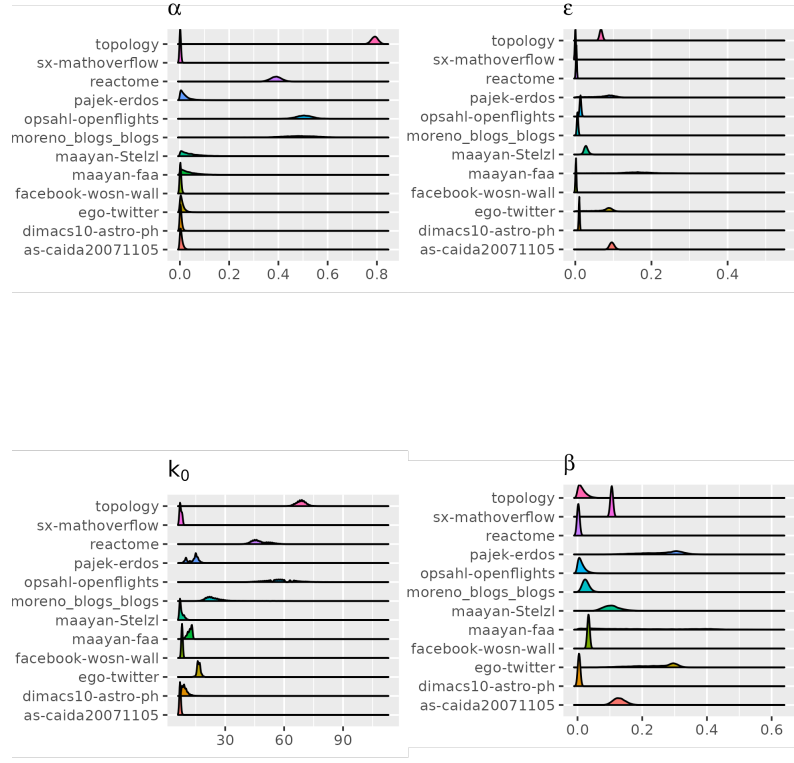
**Figure 8:** Posterior estimate for preference function (solid) with 95% credible interval (dashed) on log-log scale.

## 5 Conclusion and Discussion

In this paper we introduced a class of preference functions that, under the GPA scheme, generate a network with a flexible yet heavy-tailed degree distribution. From the simulation study we showed that the parameters can be recovered from fitting the model to the degrees alone. We also applied this method to the degree distributions of real networks, estimating their model parameters assuming they evolved in the same way. Not only did this yield fairly good fits for the degree distribution, similar to that of the Zipf-IGP, it also came with the added benefit of giving a posterior estimate for a preference function.

One limitation of this method is that the lowest degrees needed to be truncated as they had a very large effect on the fit of the model as a result of using theory developed for trees and applying it to general networks. Future work could apply theory developed for general networks using a similar method to this, allowing us to compare the results here something that is more accurate. This could include fixing the out-degree of new nodes at a constant greater than one, or allowing the out-degree of new nodes to vary.

## A Additional Plots



**Figure A1:** : Posterior estimates of parameters for real data.

## B Proofs and Derivations

### B.1 Proof of Proposition 2.1

Taking the form of the GPA degree survival:

$$\bar{F}(k) = \prod_{i=0}^k \frac{b(i)}{\lambda + b(i)}$$

and substituting into the formula for  $\Omega(F, n)$ :



$$\begin{aligned}
\Omega(F, k) &= \left( \log \frac{\prod_{i=0}^{k+1} \frac{b(i)}{\lambda+b(i)}}{\prod_{i=0}^{k+2} \frac{b(i)}{\lambda+b(i)}} \right)^{-1} - \left( \log \frac{\prod_{i=0}^k \frac{b(i)}{\lambda+b(i)}}{\prod_{i=0}^{k+1} \frac{b(i)}{\lambda+b(i)}} \right)^{-1} \\
&= \left( \log \frac{\lambda+b(k+2)}{b(k+2)} \right)^{-1} - \left( \log \frac{\lambda+b(k+1)}{b(k+1)} \right)^{-1} \\
&= \left( \log \left[ 1 + \frac{\lambda}{b(k+2)} \right] \right)^{-1} - \left( \log \left[ 1 + \frac{\lambda}{b(k+1)} \right] \right)^{-1}.
\end{aligned}$$

Clearly if  $b(k) = c$  or  $\lim_{k \rightarrow \infty} b(k) = c$  for some  $c > 0$  then  $\Omega(F, k) = 0$ . Now consider a non-constant  $b(k)$  and re-write  $\Omega(F, k)$  as:

$$\begin{aligned}
\Omega(F, k) &= \left( \log \left[ 1 + \frac{\lambda}{b(k+2)} \right] \right)^{-1} - \frac{b(k+2)}{\lambda} + \frac{b(k+2)}{\lambda} - \left( \log \left[ 1 + \frac{\lambda}{b(k+1)} \right] \right)^{-1} \\
&\quad + \frac{b(k+1)}{\lambda} - \frac{b(k+1)}{\lambda} \\
&= \left\{ \left( \log \left[ 1 + \frac{\lambda}{b(k+2)} \right] \right)^{-1} - \frac{b(k+2)}{\lambda} \right\} - \left\{ \left( \log \left[ 1 + \frac{\lambda}{b(k+1)} \right] \right)^{-1} - \frac{b(k+1)}{\lambda} \right\} \\
&\quad + \frac{b(k+2)}{\lambda} - \frac{b(k+1)}{\lambda}.
\end{aligned}$$

Then if  $\lim_{k \rightarrow \infty} b(k) = \infty$  it follows that:

$$\begin{aligned}
\lim_{k \rightarrow \infty} \Omega(F, k) &= \lim_{k \rightarrow \infty} \left\{ \left( \log \left[ 1 + \frac{\lambda}{b(k+2)} \right] \right)^{-1} - \frac{b(k+2)}{\lambda} \right\} \\
&\quad - \lim_{k \rightarrow \infty} \left\{ \left( \log \left[ 1 + \frac{\lambda}{b(k+1)} \right] \right)^{-1} - \frac{b(k+1)}{\lambda} \right\} \\
&\quad + \lim_{k \rightarrow \infty} \left( \frac{b(k+2)}{\lambda} - \frac{b(k+1)}{\lambda} \right) \\
&= \frac{1}{2} - \frac{1}{2} + \lim_{k \rightarrow \infty} \left( \frac{b(k+2)}{\lambda} - \frac{b(k+1)}{\lambda} \right) \\
&= \frac{1}{\lambda} \lim_{k \rightarrow \infty} [b(k+2) - b(k+1)]. \quad \square
\end{aligned}$$

## B.2 Derivation of Equation 5

For a preference function of the form:

$$b(k) = \begin{cases} g(k), & k < k_0, \\ g(k_0) + \beta(k - k_0), & k \geq k_0, \end{cases}$$

for  $\beta > 0, k_0 \in \mathbb{N}$  we have that

$$\begin{aligned} \hat{\rho}(\lambda) &= \sum_{n=0}^{\infty} \prod_{i=0}^{n-1} \frac{b(i)}{\lambda + b(i)} \\ &= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \sum_{n=k_0+1}^{\infty} \left( \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)} \prod_{i=k_0}^{n-1} \frac{g(k_0) + \beta(i - k_0)}{\lambda + g(k_0) + \beta(i - k_0)} \right) \\ &= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \left( \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)} \right) \sum_{n=k_0+1}^{\infty} \prod_{i=k_0}^{n-1} \frac{g(k_0) + \beta(i - k_0)}{\lambda + g(k_0) + \beta(i - k_0)}. \end{aligned}$$

Now using the fact that:

$$\prod_{i=0}^n (x + yi) = x^{n+1} \frac{\Gamma(\frac{x}{y} + n + 1)}{\Gamma(\frac{x}{y})}$$

and reindexing the product in the second sum,

$$\begin{aligned} \hat{\rho}(\lambda) &= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \left( \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)} \right) \sum_{n=k_0+1}^{\infty} \frac{\Gamma(\frac{g(k_0)}{\beta} + n - k_0)}{\Gamma(\frac{\lambda + g(k_0)}{\beta} + n - k_0)} \frac{\Gamma(\frac{\lambda + g(k_0)}{\beta})}{\Gamma(\frac{g(k_0)}{\beta})} \\ &= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \frac{\Gamma(\frac{\lambda + g(k_0)}{\beta})}{\Gamma(\frac{g(k_0)}{\beta})} \left( \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)} \right) \sum_{n=k_0+1}^{\infty} \frac{\Gamma(\frac{g(k_0)}{\beta} + n - k_0)}{\Gamma(\frac{\lambda + g(k_0)}{\beta} + n - k_0)} \\ &= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \frac{\Gamma(\frac{\lambda + g(k_0)}{\beta})}{\Gamma(\frac{g(k_0)}{\beta})} \left( \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)} \right) \sum_{n=1}^{\infty} \frac{\Gamma(\frac{g(k_0)}{\beta} + n)}{\Gamma(\frac{\lambda + g(k_0)}{\beta} + n)}. \end{aligned}$$

In order to simplify the infinite sum, consider:

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{\Gamma(n + x)}{\Gamma(n + x + y)} &= \frac{1}{\Gamma(y)} \sum_{n=0}^{\infty} B(n + x, y) \\ &= \frac{1}{\Gamma(y)} \sum_{n=0}^{\infty} \int_0^1 t^{n+x-1} (1-t)^{y-1} dt \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\Gamma(y)} \int_0^1 t^{x-1} (1-t)^{y-1} \sum_{n=0}^{\infty} t^n a t \\
&= \frac{1}{\Gamma(y)} \int_0^1 t^{x-1} (1-t)^{y-1} \frac{1}{1-t} a t \\
&= \frac{1}{\Gamma(y)} \int_0^1 t^{x-1} (1-t)^{y-2} a t \\
&= \frac{1}{\Gamma(y)} y(x, y-1) \\
&= \frac{\Gamma(x)}{(y-1)\Gamma(x+y-1)}.
\end{aligned}$$

This infinite sum does not converge when  $x \leq 1$  as each term is  $O(n^{-x})$ . We can now use this in  $\hat{\rho}(\lambda)$ :

$$\begin{aligned}
\hat{\rho}(\lambda) &= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \frac{\Gamma\left(\frac{\lambda + g(k_0)}{\beta}\right)}{\Gamma\left(\frac{g(k_0)}{\beta}\right)} \left( \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)} \right) \left( \frac{\Gamma\left(\frac{g(k_0)}{\beta}\right)}{\left(\frac{\lambda}{\beta} - 1\right) \Gamma\left(\frac{g(k_0) + \lambda}{\beta} - 1\right)} - \frac{\Gamma\left(\frac{g(k_0)}{\beta}\right)}{\Gamma\left(\frac{g(k_0) + \lambda}{\beta}\right)} \right) \\
&= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \left( \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)} \right) \left( \frac{\Gamma\left(\frac{g(k_0) + \lambda}{\beta}\right)}{\left(\frac{\lambda}{\beta} - 1\right) \Gamma\left(\frac{g(k_0) + \lambda}{\beta} - 1\right)} - 1 \right) \\
&= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \left( \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)} \right) \left( \frac{\frac{g(k_0) + \lambda}{\beta} - 1}{\frac{\lambda}{\beta} - 1} - 1 \right) \\
&= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \left( \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)} \right) \left( \frac{g(k_0) + \lambda - \beta}{\lambda - \beta} - 1 \right) \\
&= \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{g(i)}{\lambda + g(i)} + \frac{g(k_0)}{\lambda - \beta} \prod_{i=0}^{k_0-1} \frac{g(i)}{\lambda + g(i)}. \quad \square
\end{aligned}$$

## References

- Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512. <https://doi.org/10.1126/science.286.5439.509>, URL <https://www.science.org/doi/abs/10.1126/science.286.5439.509>, <https://www.science.org/doi/pdf/10.1126/science.286.5439.509>
- Betzal RF, Bassett DS (2017) Generative models for network neuroscience: prospects and promise. *Journal of The Royal Society Interface* 14(136):20170623. <https://doi.org/10.1098/rsif.2017.0623>
- Broido AD, Clauset A (2019) Scale-free networks are rare. *Nature Communications* 10(1). <https://doi.org/10.1038/s41467-019-08746-5>, URL <http://dx.doi.org/10.1038/s41467-019-08746-5>
- Cesàro E (1888) Sur la convergence des séries. *Nouvelles annales de mathématiques : journal des candidats aux écoles polytechnique et normale* 3e série, 7:49–59. URL [https://www.numdam.org/item/NAM\\_1888\\_\\_3\\_\\_7\\_\\_49\\_1/](https://www.numdam.org/item/NAM_1888__3__7__49_1/)
- Krapivsky PL, Redner S (2001) Organization of growing random networks. *Phys Rev E* 63:066123. <https://doi.org/10.1103/PhysRevE.63.066123>, URL <https://link.aps.org/doi/10.1103/PhysRevE.63.066123>
- Latouche P, Birmelé E, Ambroise C (2011) Overlapping stochastic block models with application to the French political blogosphere. *The Annals of Applied Statistics* 5(1):309 – 336. <https://doi.org/10.1214/10-AOAS382>, URL <https://doi.org/10.1214/10-AOAS382>
- Lee C, Eastoe EF, Farrell A (2024) Degree distributions in networks: Beyond the power law. *Statistica Neerlandica* 78(4):702–718. <https://doi.org/https://doi.org/10.1111/stan.12355>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/stan.12355>, <https://onlinelibrary.wiley.com/doi/pdf/10.1111/stan.12355>
- Rohrbeck C, Eastoe EF, Frigessi A, et al (2018) Extreme value modelling of water-related insurance claims. *The Annals of Applied Statistics* 12(1):246 – 282. <https://doi.org/10.1214/17-AOAS1081>, URL <https://doi.org/10.1214/17-AOAS1081>
- Rossi RA, Ahmed NK (2015) The network data repository with interactive graph analytics and visualization. In: *AAAI*, URL <https://networkrepository.com>
- Rudas A, Tóth B, Valkó B (2007) Random trees and general branching processes. *Random Structures & Algorithms* 31(2):186–202. <https://doi.org/https://doi.org/10.1002/rsa.20137>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20137>, <https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20137>
- Setayesh A, Sourati Hassan Zadeh Z, Bahrak B (2022) Analysis of the global trade network using exponential random graph models. *Applied Network Science* 7. <https://doi.org/10.1007/s41467-022-00000-0>

[//doi.org/10.1007/s41109-022-00479-7](https://doi.org/10.1007/s41109-022-00479-7)

Shimura T (2012) Discretization of distributions in the maximum domain of attraction. *Extremes* 15:1–19. <https://doi.org/10.1007/s10687-011-0137-7>

Voitalov I, van der Hoorn P, Hofstad R, et al (2019) Scale-free networks well done. *Physical Review Research* 1. <https://doi.org/10.1103/PhysRevResearch.1.033034>

Wang T, Resnick S (2022) Random networks with heterogeneous reciprocity. **2208.00348**