Learning growth mechanisms of tail realistic preferential attachment models from network degree distributions

Thomas William Boughen[1*], Clement Lee[1] and Vianey Palacios Ramirez[1]

[1]School of Mathematics, Statistics and Physics, Newcastle University.

*Corresponding author(s). E-mail(s): t.w.boughen1@newcastle.ac.uk;

Abstract

Identifying the generating mechanism of a network is challenging as, more often than not, only snapshots are available, but not the 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 display such power-law behaviour is commonly modelled by preferential attachment. The ubiquity of the power law has been challenged by the presence of alternatives with comparable performance, as well as the recent findings that the tail of the degree distribution is often lighter than implied by the body, whilst still being regularly varying. In this paper, we propose a preferential attachment model with a flexible preference function. Using methods for discrete extremes, we characterise the tail behaviour of the limiting degree distribution directly by the form of the preference function. Directly relating the tail index to the model parameters enables them to be inferred when fitting to degree distributions alone, which is supported by simulation studies. Results of applications to real data are promising and comparable to alternatives.

1 Introduction

Networks have become powerful tools for representing and analysing complex systems, with uses in a large array of fields. In network science and statistics, they have been studied by various families of models, from stochastic block models for detecting communities online (**?**), to exponential random graph models (ERGMs) for analysing the global trade network (**?**), and mechanistic models for investigating patterns in neural systems (**?**).

Amid the recent rise of interest in networks, there has been a debate on whether most real networks are scale-free. Claiming a real network is scale-free is equivalent to saying that its degree distribution follows a power law, that is the fraction of nodes with degree k is proportional to $k^{-\alpha}$, and therefore has a regularly varying tail with tail index α. On the side against the claim is **?** who compared the fits of a power law model against that of several non-scale-free models to nearly a thousand networks, only to find strong evidence for scale-freeness in four percent and weak evidence in over half of the networks, thus claiming that scale-free networks do not make up a majority in real networks. On the other side of this debate is **?** who disagrees and claims that these networks are not nearly as rare and only appear to be so as a result of an unrealistic expectation of a power law without deviations or noise. Additional evidence of these deviations from a power law is shown by **?** who demonstrate that a lot of networks are partially scale-free, in that the body of the degree distribution is often modelled well by a power law, while the tail is lighter than what is implied by the body, albeit still regularly varying. Nevertheless, most studies into the appropriateness of a power law for the degrees of real networks, the aforementioned references included, have been largely descriptive in the sense that no information about the growth of the networks is revealed.

The popularity of using the power law for network degrees can be traced back to the preferential attachment (PA) model popularised by **?**. In the general model, as new nodes join the network, an existing node with degree k gains edges at a rate proportional to $b(k)$, where $b(\cdot)$ is a non-negative preference function. **?** showed that, when $b(k) = k + 1$, in the limit the resulting degree distribution is regularly varying with index 2. Subsequently, if a real network is shown to be scale-free, one can loosely justify PA as the underlying mechanism of its growth.

The model from **?** provided the foundations for various generalisations — **?** considered $b(k) = (k + 1)^{\alpha}$, and showed that the degree distribution is not regularly varying (and therefore not following the power law) when $0 < \alpha < 1$, and when $\alpha > 1$ a finite number of nodes end up with all edges after a certain point resulting in a degenerate degree distribution. **?** returns to a linear preference function of the form $b(k) = k + \varepsilon$ but adds the possibility for reciprocal edges to be sent, resulting in the joint distribution of in-degree and out-degree being multivariate regularly varying and having the property of hidden regular variation. **?** followed in the footsteps of **?**, by considering a PA tree and using theory from continuous branching processes, deriving a limiting degree distribution in terms of the preference function $b(\cdot)$. Nevertheless, research in this area tends to only focus on the theoretical asymptotic results of network growth models with little analysis of real networks.

This paper aims to address the gap between the applied and theoretical works, by asking if a network is assumed to come from a PA model, can we use the degree distribution alone to directly infer the parameters of the the preference function and learn about the growth mechanisms? Moreover, proper consideration is given to the tail of the degree distribution, because otherwise the effects of the largest degrees, which correspond to the most influential nodes, deviating from the power law will be discounted.

While Voitalov et al. (2019) have pioneered using methods in extreme value theory to analyse degree distributions, they overlooked the inapplicability of standard tools for continuous extremes, due to the discrete nature of degrees. A key difficulty arises because many standard discrete distributions, such as the Poisson, geometric, or negative binomial, do not satisfy the conditions required to belong to a maximum domain of attraction (MDA). Shimura (2012) showed that for a discrete distribution F must be long-tailed in order to be in an MDA, and introduced the notion of recoverability to the MDA, also referred to as the discrete MDA (Hitz et al., 2024). Using the theoretical guarantees of regular variation for a discrete distribution by Shimura (2012), we demonstrate how the tail of the degree distribution is directly influenced by $b(\cdot)$, and subsequently proposed a class of preference functions that not only imply regularly varying degree distributions but are also tail realistic for real networks. These analytical results enable the likelihood of the degree distribution to be expressed in terms of the parameters of $b(\cdot)$, which in turn allows the underlying mechanism of the network, assumed to grow according to PA, to be inferred directly.

The remainder of the paper is as follows: Section 2 gives a detailed description of the PA model alongside the theoretical results for the survival function of the limiting degree distribution, with a focus on the tail behaviour in terms of the preference function $b(\cdot)$. A class of asymptotically linear preference functions will be introduced and shown to guarantee regular variation in the degree distribution while remaining flexible up until a threshold. Section 3 utilises the proposed preference function and illustrates numerically how the tail index of the degree distribution varies with the model parameters. The simulation study in Subsection 4.1 demonstrates that the parameters can be recovered from fitting the model to only the degree distribution. Subsection 4.2 fits the model to some real data and provides posterior estimates for the preference function. Section 5 provides a discussion of this paper and possible avenues for future work.

2 Tail Behaviour of Preferential Attachment Model

The model that we will focus on in this paper is the General Preferential Attachment (GPA) model in **?** and is defined as follows:

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

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

Special cases of this model include the Barabási-Albert (BA) model when $b(k) = k+1$, which in the limit of $t \to \infty$ leads to a power-law degree distribution with tail index 2, and the Uniform Attachment (UA) model where $b(k) = c$ leading to a degree distribution that is not regularly varying.

The survival function of the limiting degree distribution, called the limiting survival hereafter, under condition 1 can be analytically derived in the case where $m = 1$, which is presented below.

Consider a continuous time branching process $\zeta(t)$ driven by a Markovian pure birth process, with $\zeta(0) = 0$ and birth rates depending on a non-negative function $b(\cdot)$:

$$\Pr(\zeta(t + \mathrm{d}t) = k + 1 | \zeta(t) = k) = b(k)\mathrm{d}t + o(\mathrm{d}t).$$

Denote the density of the point process associated with the pure birth process corresponding to the growth of an individual node by $\rho(t)$, and its Laplace transform by $\widehat{\rho}(\lambda) := \int_0^\infty e^{-\lambda t} \rho(t)\mathrm{d}t$. Next, denote the Malthusian parameter of this process by λ^*, that is λ^* satisfies $\widehat{\rho}(\lambda^*) = 1$.

Now, let $\Upsilon(t)$ be the tree determined by $\zeta(t)$ as follows: $\Upsilon(t) = \{\emptyset\}$ and $\Upsilon(t) = G$ where each existing node x in $\Upsilon(t)$ gives birth to a child with rate $b(\deg(x, \Upsilon(t)))$ independently of the other nodes where $\deg(x, \Upsilon(t))$ is the degree of node x in the tree $\Upsilon(t)$ at time t, denote by $\Upsilon(t)_{\downarrow x}$ the tree when treating node x as the root.

Theorem 1 from **?** states that for the tree $\Upsilon(t)$ at time t and a characteristic function of the tree $\varphi(\cdot)$:

$$\lim_{t \to \infty} \frac{1}{|\Upsilon(t)|} \sum_{x \in \Upsilon(t)} \varphi(\Upsilon(t)_{\downarrow x}) = \lambda \int_0^\infty e^{-\lambda t} \mathbb{E}\left[\varphi(\Upsilon(t))\right] \mathrm{d}t. \tag{2}$$

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:

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

which can also be written using Equation 2 as:

$$\bar{F}(k) = \frac{\int_0^\infty e^{-\lambda^* t}\mathbb{E}\left[\mathbb{I}\left\{\deg(x, \Upsilon(t)) > k\right\}\right]\mathrm{d}t}{\int_0^\infty e^{-\lambda^* t}\mathrm{d}t} = \prod_{i=0}^{k} \frac{b(i)}{\lambda^* + b(i)}. \tag{3}$$

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

$$f(k) = \frac{\lambda^*}{\lambda^* + b(k)} \prod_{i=0}^{k-1} \frac{b(i)}{\lambda^* + b(i)}. \tag{4}$$

We investigate how the tail behaviour of the discrete limiting degree distribution is influenced by the preference function $b(\cdot)$, using the characterisation by **?**.

A central tool is the quantity $\Omega(F, k)$. For a discrete distribution F with survival function $\bar{F}$ and $k \in \mathbb{Z}^+$ define:

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

? established that if $\lim_{k\to\infty} \Omega(F, k) = 1/\alpha$ ($\alpha > 0$), then F is regularly varying with $\bar{F}(k) \sim k^{-\alpha}$, and hence belongs to the Frechet MDA. If instead, $\lim_{k\to\infty} \Omega(F, k) = 0$ then the distribution is recoverable to the Gumbel MDA; in this cases we say that the distribution is light-tailed. Moreover, if the distribution is also long-tailed, we can conclude that F itself lies in the Gumbel MDA.

This framework offers a natural approach for analysing the tail behaviour of the limiting degree distribution. The following proposition gives the limiting behaviour of $\Omega(F, k)$ when F is a limiting degree distribution resulting from the GPA model with preference function $b(\cdot)$.

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

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

Here, λ^* is the Malthusian parameter of the corresponding branching process. See Appendix A for the details of the proof.

Remark 2.1. If $\lim_{k\to\infty} b(k) < \infty$, then $\Omega(F, k) = 0$.

Corollary 2.1. *Let $\bar{F}(k)$ be the limiting survival function of degrees in a GPA network with preference function b such that $b(k) \to \infty$. Then:*

- *$\bar{F}(k)$ is regularly varying if and only if* $\lim_{k\to\infty}[b(k+1)-b(k)] = c > 0$*, in which case the tail index is λ^*/c, where λ^* is the Malthusian parameter of the corresponding branching process.*
- *$\bar{F}(k)$ is light-tailed if and only if* $\lim_{k\to\infty}[b(k+1)-b(k)] = 0$*.*

Corollary 2.1 is a direct consequence of Proposition 2.1 and the results of **?**, and it aligns with previous findings in the literature. In particular it is consistent with **?**, who showed that a sub-linear preference function leads to a light-tailed distributions, since for $b(k) = k^\alpha$ with $\alpha < 1$, we have $\lim_{k\to\infty}[b(k+1)-b(k)] = 0$. Conversely, it also agrees with the limiting behavior of BA model, which produces a regularly degree distribution with tail index 2; for the preference function $b(k) = k+\varepsilon$, $\lim_{k\to\infty}[b(k+1)-b(k)] = 1$, giving a tail index of $\lambda^* = 2$

Using Corollary 2.1 we can directly connect the preference function to the tail behaviour of the degree distribution. By choosing an appropriate preference function, we can ensure that the tail of the degree distribution is regularly varying while allowing the body to deviate from the power-law shape, as observed in real networks. Inspired by **?**, we consider a piecewise function:

$$b(k) = \begin{cases} k^\alpha + \varepsilon, & k < k_0, \\ k_0^\alpha + \varepsilon + \beta(k - k_0), & k \geq k_0 \end{cases} \tag{5}$$

for $\alpha, \beta, \varepsilon > 0$ and $k_0 \in \mathbb{N}$. By Corollary 2.1, the resulting degree distribution is regularly varying with tail index λ^*/β, since

$$\lim_{k\to\infty} \Omega(F, k) = \frac{\beta}{\lambda^*}.$$

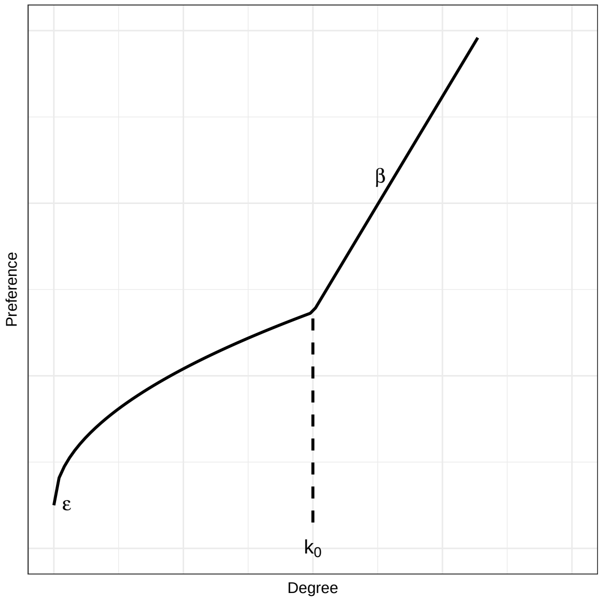

Fig. 1: Example construction of the preference function in Equation 5

3 A PA Model with Flexible Regular Variation

In the previous section, we found that using an asymptotically linear preference function allows for the inclusion of sub/super-linear behaviour below the threshold, while simultaneously guaranteeing regular variation of the degrees. In this section, we demonstrate the flexibility of the preference function in Equation 5, with regard to the tail behaviour of the limiting degree distribution. Using Equation 3, the limiting survival is

$$\bar{F}(k) = \begin{cases} \prod_{i=0}^{k} \frac{i^\alpha+\varepsilon}{\lambda^*+i^\alpha+\varepsilon}, & k < k_0, \\ \left(\prod_{i=0}^{k_0-1} \frac{i^\alpha+\varepsilon}{\lambda^*+i^\alpha+\varepsilon}\right) \frac{\Gamma(\lambda^*+k_0^\alpha+\varepsilon)/\beta)}{\Gamma((k_0^\alpha+\varepsilon)/\beta)} \frac{\Gamma\left(k-k_0+1+\frac{k_0^\alpha+\varepsilon}{\beta}\right)}{\Gamma\left(k-k_0+1+\frac{\lambda^*+k_0^\alpha+\varepsilon}{\beta}\right)}, & k \geq k_0, \end{cases} \tag{6}$$

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

$$\hat{\rho}(\lambda) = \sum_{n=0}^{k_0} \prod_{i=0}^{n-1} \frac{i^\alpha + \varepsilon}{\lambda + i^\alpha + \varepsilon} + \left(\frac{k_0^\alpha + \varepsilon}{\lambda - \beta}\right) \prod_{i=0}^{k_0-1} \frac{i^\alpha + \varepsilon}{\lambda + i^\alpha + \varepsilon} \tag{7}$$

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

For some parameter combinations, the limiting survival $\bar{F}(k)$ is shown on log-log scale in Figure 2:

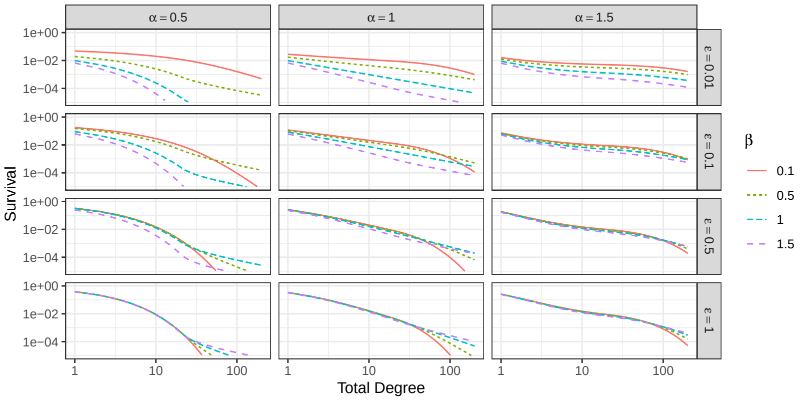

Fig. 2: The limiting survival, according to various combinations of $(\alpha, \beta, \varepsilon)$ and $k_0 = 20$ of the proposed PA model.

Figure 2 demonstrates that this model can capture a range of tail behaviour, 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 analytic form of the survival function in (6), offers a natural connection to the discrete version of the generalised Pareto (GP) distribution , providing a link to a well established component of discrete extremes in the literature. Specifically, it is connected to the Integer GP (IGP) distribution seen in **?** with conditional survival:

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

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

By Equation 6 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,

$$\bar{F}(k) \begin{cases} = \prod_{i=0}^{k} \frac{i^\alpha+\varepsilon}{\lambda^*+i^\alpha+\varepsilon}, & k<k_0,\\ \approx \left(\prod_{i=0}^{k_0-1}\frac{i^\alpha+\varepsilon}{\lambda^*+i^\alpha+\varepsilon}\right)\left(\frac{\beta(k+1-k_0)}{k_0^\alpha+\varepsilon}+1\right)^{-\lambda^*/\beta}, & k\geq k_0,\end{cases} \tag{8}$$

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

To assess how close of an approximation this is, the theoretical conditional survival Equation 6 are shown in Figure 3 in colour and their IGP approximations Equation 8 are shown in grey. The approximation holds up well even for large degrees and more so when α is larger.

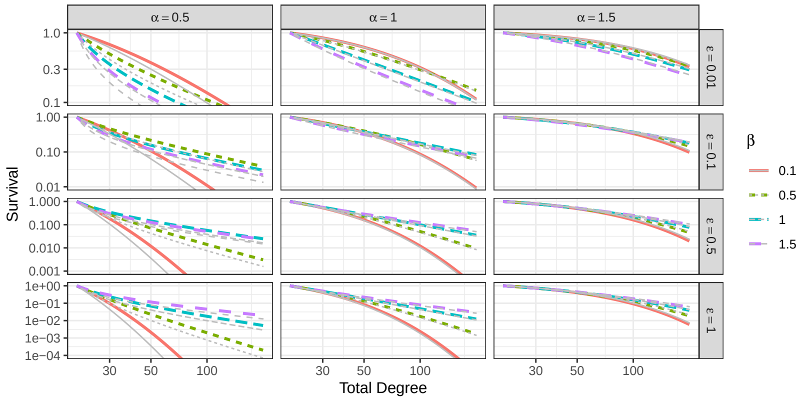

Fig. 3: Theoretical conditional survivals (grey) of the proposed model alongside their IGP approximations (coloured).

In agreement with Corollary 2.1, $\beta > 0$ ensures that the shape parameter of the IGP distribution is positive and thus the distribution is regularly varying. Additionally the shape parameter ξ is shown in Figure 4 for various parameter choices. The darker and lighter regions on the heat maps correspond to a heavier and a lighter tail, respectively, and the red dashed line shows combinations of α and β that produce a limiting degree distribution with the same tail index as the BA model.

Through the connection implied by Equation 8, fitting the proposed model is almost equivalent to fitting the IGP distribution to the degrees and estimating its parameters. However, instead of only describing the shape of the degree distribution, we would also gain estimates for the shape of the preference function, thus gaining a direct understanding the mechanisms underlying the growth of the network.

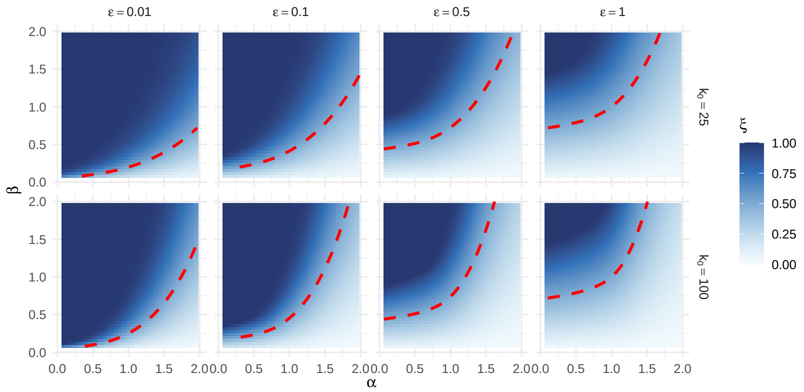

Fig. 4: Heat maps of ξ for various combinations of the parameters of the proposed model.

To perform inference of the model parameters, we consider a network with degree count vector $\boldsymbol{n} = (n_0, n_1, \dots, n_M)$, where M is the maximum degree. Using Equation 4, the likelihood is:

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

where $\mathrm{B}(\cdot, \cdot)$ is the the beta function, $\boldsymbol{\theta} = (\alpha, \varepsilon, k_0, \beta)$, and $l \geq 0$ is a quantity 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 **?** only provides results for the case of a preferential attachment tree.

4 Applications

4.1 Simulated Data

This subsection aims to show that the parameters of the model (and therefore the preference function) 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:

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

where Gamma(a, b) is the gamma distribution with shape a and rate b, and 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 $l = 0$.

Figures 5 and 6 demonstrate the usefulness of the model, as we can recover the model parameters well from only the final degree distribution of the simulated network. This indicates that the method may also be applied to real networks, with the assumption that they evolved according to the GPA scheme.

4.2 Real Data

In this subsection, 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 **?** 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(**?**):

- `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

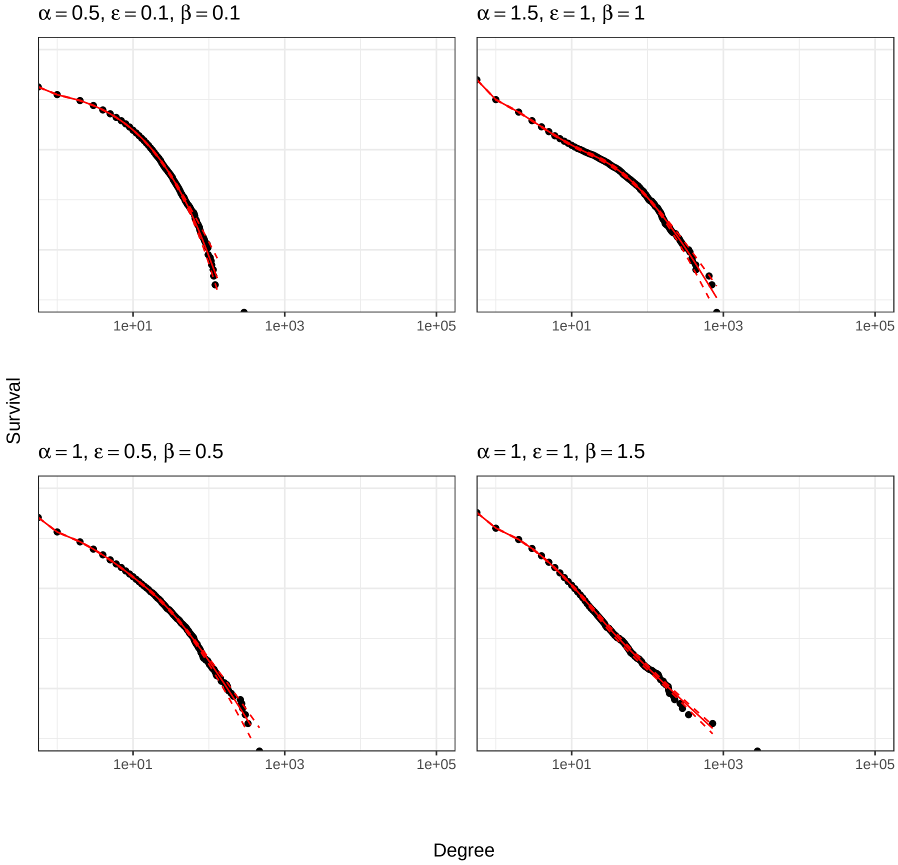

Fig. 5: Empirical (black dots) and fitted (red line) survival functions for data simulated from the proposed model with various combinations of $(\alpha, \beta, \epsilon)$ and $k_0 = 20$. The 95% credible (dashed lines) are included but too narrow to be seen clearly.

- `topology`: network of connections between autonomous systems of the Internet

Figure 7 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 **?**. In most cases, the GPA model gives a similar fit 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 the the GPA scheme.

Figure 9 shows the posterior of the shape parameter ξ obtained from the Zipf-IGP model alongside the posterior of the equivalent shape parameter β/λ^* 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 too low, forcing almost

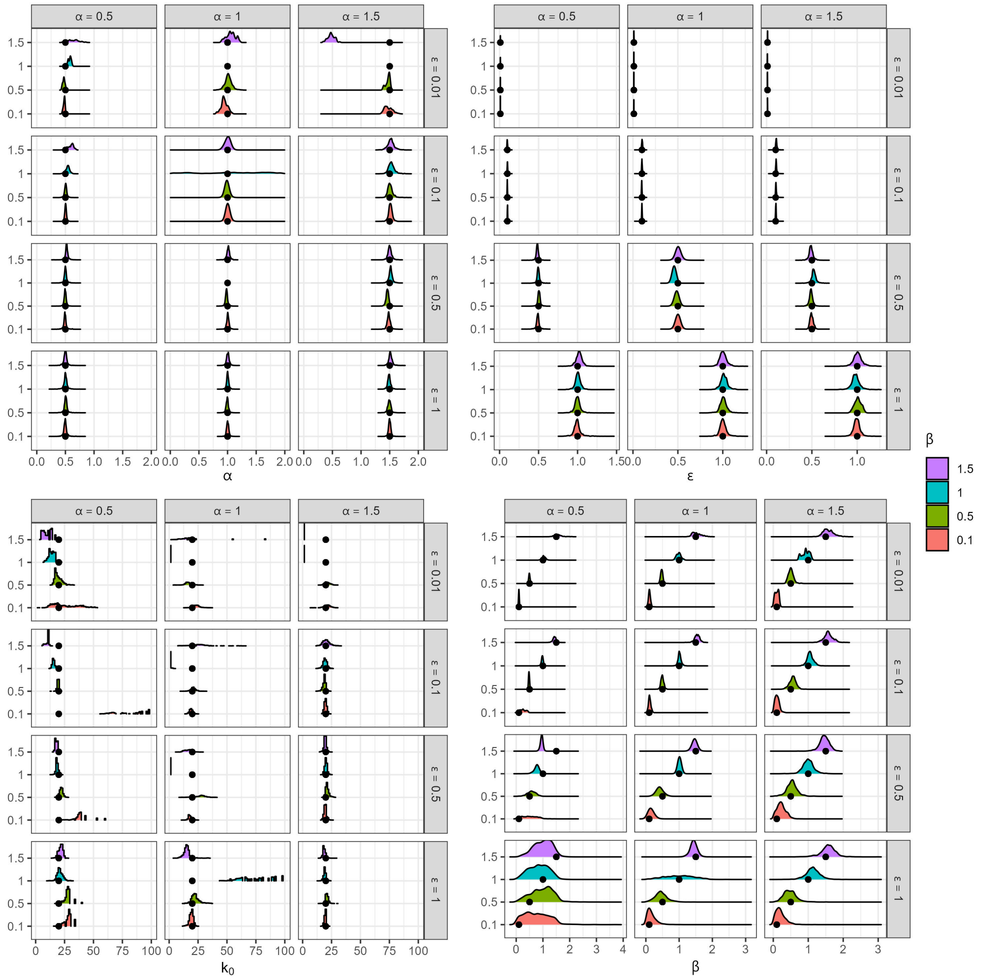

Fig. 6: : Posterior densities of parameters for data simulated from the proposed model with various combinations of (α,β,ε) and $k_0 = 20$. True parameter values shown with black dots.

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.

Figure 10 shows the estimated preference function $b(k)$ alongside the 95% credible interval on a log-log scale. 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 PA

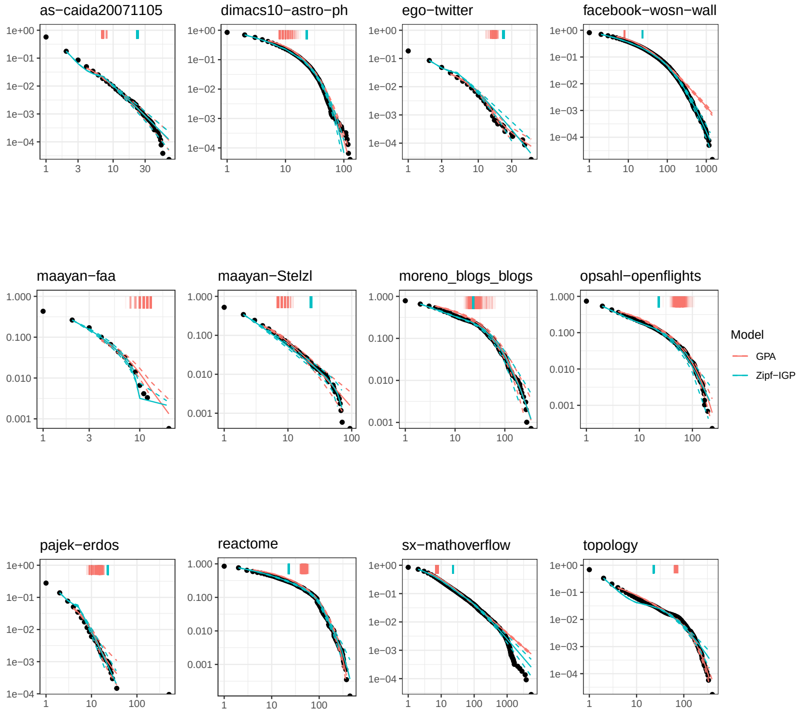

Fig. 7: Empirical (black dots) and posterior medians (solid red) of the fitted survival function for several real data sets and their 95% credible intervals (dotted red).

kicks in, some with this shape are `pajek-erdos` and `sx-mathoverflow`. The second distinct shape appears to provide some clear PA behaviour that then slows down after a certain point, and 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.

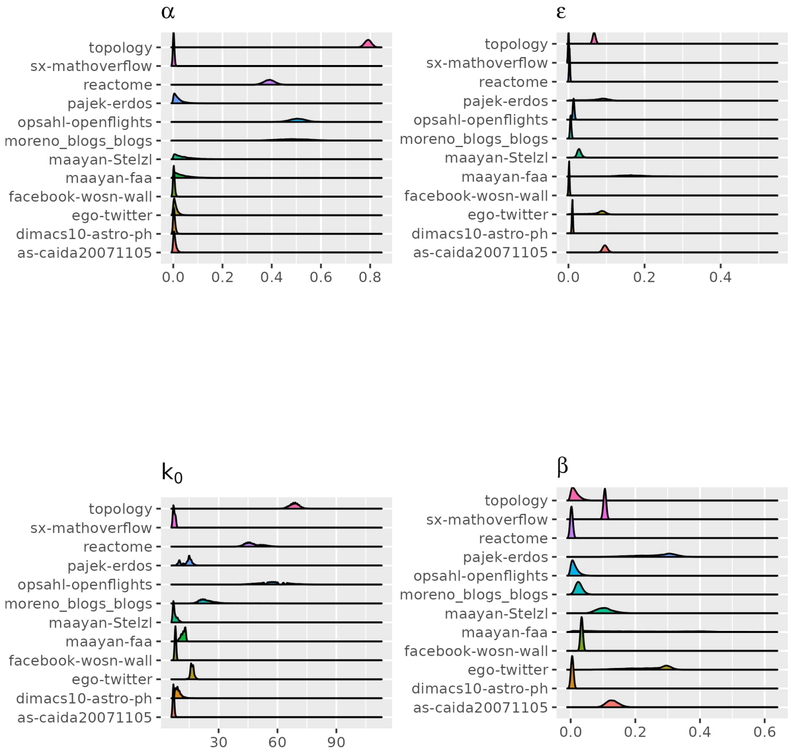

Fig. 8: Posterior densities of the parameters of the proposed model fitted to real data.

5 Conclusion and Discussion

We introduced a class of preference functions that, under the GPA scheme, generate a network with a flexible yet regularly varying 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 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.

This paper contributes to the understanding of the relationship between the mechanism underlying a networks growth and the resulting degree distribution. As well as demonstrating that under certain conditions information about this mechanism can be garnered from the degrees alone. In future, something similar could be done by

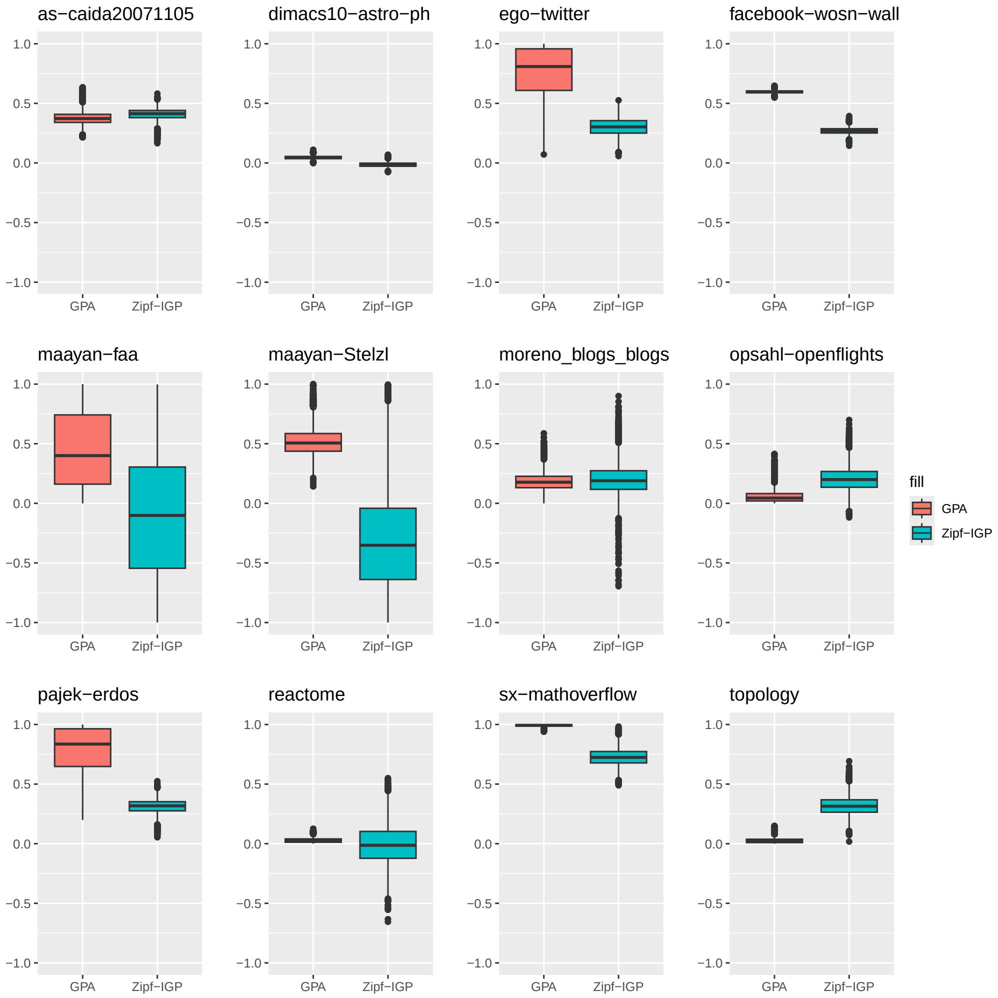

Fig. 9: Posterior estimates of shape parameter of Zipf-IGP distribution (right) and the analagous quantity obtained from fitting the proposed model (left).

looking at another statistic for the network (e.g. the triangle distribution) and using it in combination with the degrees to gain further insights into the growth mechanism when the networks full evolution is not available. The triangle distribution would be a good avenue for future work within the realm of extremes as it has also been shown by **?** that many real networks seem to have power law triangle distributions.

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

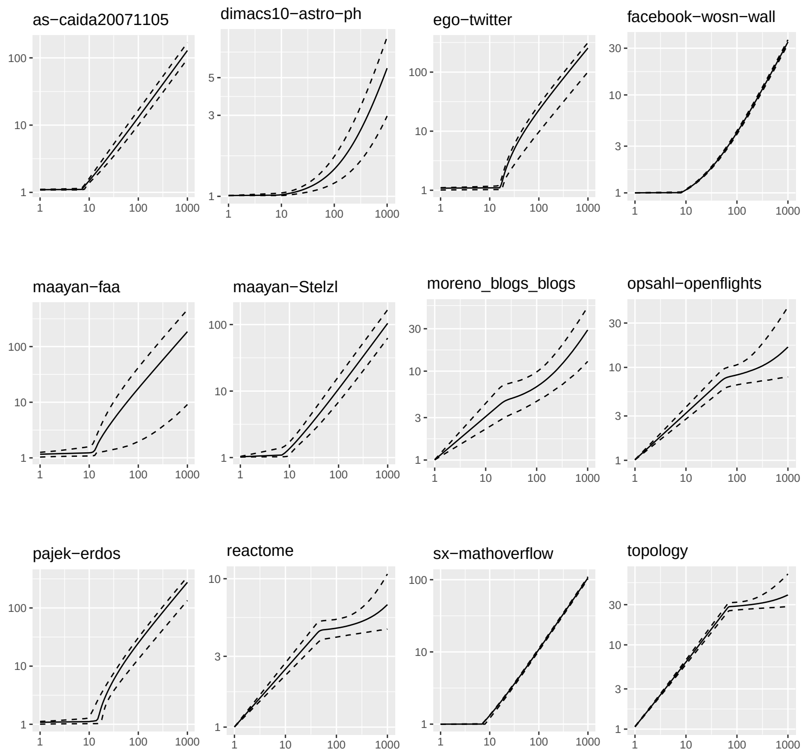

Fig. 10: Posterior median for the preference function (solid) with 95% credible interval (dashed) on log-log scale.

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.

Its worth noting the recent work by **?**, that provides results for more general networks beyond trees utilising the same underlying branching process. They consider a model that grows in the exact same way as the model that we have considered, but then they collapse the nodes of the tree resulting in a more general network much more like those in reality. However, there is no expression for the probability mass function of the degrees and therefore no likelihood that can be used for modeling in the same way that we have. In spite of this, using Remark 2 from **?** it can be shown that the survival function of the limiting degree distribution (when using a preference function of the form from Corollary 2.1) can be bounded by two regularly varying functions showing that the degree distribution is still heavy tailed, although not necessarily regularly varying.

Obtaining an expression for the degree distributions of more general cases of preferential attachment models are, for the moment, seemingly inaccessible and provide a real barrier to using more realistic models in a way similar to what we have in this paper, we leave these open problems for future analysis.

A Proofs and Derivations

A.1 Proof of Proposition 2.1

Taking the form of the GPA degree survival from Equation 3 :

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

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 $\lim_{k\to\infty} b(k) = c$ for some $c > 0$ then $\Omega(F,k) = 0 (= \lim_{k\to\infty} \frac{1}{\lambda^*}[b(k+2) - b(k+1)])$. 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 $\lim_{k\to\infty} b(k) = \infty$ it follows that:

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

$$
\begin{aligned}
&+\lim_{k\to\infty}\left(\frac{b(k+2)}{\lambda}-\frac{b(k+1)}{\lambda}\right)\\
&=\frac{1}{2}-\frac{1}{2}+\lim_{k\to\infty}\left(\frac{b(k+2)}{\lambda}-\frac{b(k+1)}{\lambda}\right)\\
&=\frac{1}{\lambda}\lim_{k\to\infty}\left[b(k+2)-b(k+1)\right]. \qquad \square
\end{aligned}
$$

A.2 Derivation of Equation 7

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:

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

and reindexing the product in the second sum,

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

In order to simplify the infinite sum, consider:

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

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

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

References