

Annual Progress Review

Thomas William Boughen



Newcastle University

School of Mathematics, Statistics and Physics

1 Introduction

Since the aim is to gain understanding about the behaviour of the degree distribution of networks at the right tail, it seems natural to look to using methods from extreme value theory.

2 Extreme Value Theory

This section begins with a review of the theory and methodology for modelling the extreme values of continuous random variables, before moving to considerations for modelling the extreme values of discrete random variables.

2.1 Continuous Extremes

Studying the properties of the extreme values of a random variable first requires determining what exactly is considered to be an extreme value. In this section extreme values of two kinds are considered, both of which can be characterised.

The first kind of extreme value considers the distribution of block maxima. That is, for a set of independent and identically distributed (iid) random variables $\{X_1, \dots, X_n\}$ with common cumulative density function (cdf) F what is the limiting distribution of $M_n = \max\{X_1, \dots, X_n\}$? This is answered by the Fisher–Tippett–Gnedenko theorem [REF](#).

Theorem 2.1.1 (Extreme Value Theorem). *Let X_1, \dots, X_n be a sample of iid random variables with common cdf F with block maxima $M_n = \max\{X_1, \dots, X_n\}$ and suppose that there exists $a_n > 0, b_n \in \mathbb{R}$ such that*

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{1}{a_n} [M_n - b_n] \right) = G(x),$$

then F is said to be in the domain of attraction of G , denoted $F \in \mathcal{D}(G)$, and G is of one of three types:

- *Gumbel:* $\Lambda(x) = \exp\{-\exp(-x)\}$, $x \in \mathbb{R}$
- *Fréchet:* $\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}$, $x \geq 0, \alpha > 0$
- *Negative-Weibull:* $\Psi_\alpha(x) = \exp\{-x^{-\alpha}\}$, $x < 0, \alpha > 0$

These three types of distribution can be collected into one more general distribution, called the generalised extreme value (GEV) distribution which is defined below.

Definition 2.1.1 (Generalised Extreme Value Distribution). Denoted by $\text{GEV}(\mu, \sigma, \xi)$ the distribution is characterised by three parameters $\mu \in \mathbb{R}$ the location, $\sigma \in \mathbb{R}^+$ the scale, and the shape $\xi \in \mathbb{R}$. It has support on $\{x \in \mathbb{R} : 1 + \xi(x - \mu)/\sigma > 0\}$ and has cdf given by:

$$G(x) = \begin{cases} \exp \left\{ - \left(1 + \frac{\xi(x-\mu)}{\sigma} \right)_+^{-1/\xi} \right\}, & \xi \neq 0 \\ \exp \{ - \exp(-\frac{x-\mu}{\sigma}) \}, & \xi = 0. \end{cases}$$

The three types of extremal distribution are obtained from changing the shape parameter ξ , which corresponds to $1/\alpha$ in Theorem [2.1.1](#). This change is generally to be made so that increasing ξ corresponds to increasing how heavy the tails of the distribution are. Specifically, $\xi < 0$, $\xi = 0$, $\xi > 0$, correspond to the negative Weibull, Gumbel and the Fréchet domains of attraction respectively.

Identifying what domain of attraction a distribution belongs to via Theorem [2.1.1](#) may not be so trivial, however, there are other ways to determine what domain of attraction a distribution belongs to.

Definition 2.1.2 (Conditions for Domains of Attraction). The distribution F belongs to the Fréchet domain of attraction $\mathcal{D}(\Phi_\alpha)$ if and only if the survival function $\bar{F} = 1 - F$ is regularly varying with index $-\alpha$ i.e.:

$$\bar{F}(x) = x^{-\alpha} L(x), \quad \text{for } L \text{ slowly varying}$$

A similar condition applies to the Weibull domain of attraction $\mathcal{D}(\Psi_\alpha)$ in that a distribution F belongs to the Weibull domain of attraction if and only if:

$$\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x), \quad \text{for } L \text{ slowly varying}$$

where x_F is the finite right endpoint of the support of F .

The condition for the Gumbel domain of attraction is not defined through regular variation but instead, a distribution F belongs to the Gumbel domain if and only if there exists a positive function $a : \mathbb{R} \rightarrow \mathbb{R}^+$ and a $t \in \mathbb{R}$ such that:

$$\lim_{x \rightarrow x_F} \frac{\bar{F}(x + ta(x))}{\bar{F}(x)} = e^{-t}$$

Throughout this report the term “heavy tailed” distribution will be used to describe any distribution in the Fréchet domain of attraction, although some of the literature refers to “heavy tailed” distributions as being the distributions that decay slower than the exponential.

At this point it will also be useful to introduce the concept of distributions that have super-heavy tails.

Definition 2.1.3 (Super Heavy Tails). [FIND SUPER HEAVY TAILS DEFINITION]

Additionally, if the survival function \bar{F} is slowly varying itself then F has super heavy tails i.e.

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = 1, \forall t \in \mathbb{R}^+ \implies \text{super heavy tails}$$

Another kind of extreme values are the observations above a large threshold, like the limiting distribution of block maxima, the limiting distribution of these extreme values can be characterised by the generalised pareto (GP) distribution.

Definition 2.1.4 (Generalised Pareto Distribution). Consider a random variable X with the same cdf F as in Theorem 2.1.1, the Generalised Pareto (GP) distribution can be obtained by using the GEV distribution and conditional probability such that for large enough threshold the GP distribution approximately describes the conditional distribution of threshold exceedances. More precisely, for sufficiently large threshold u and the change of variable to $Y = X - u$:

$$\Pr(Y \leq y | Y > 0) = H(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, & y > 0, \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & y > 0, \xi = 0 \end{cases}$$

Since this distribution was obtained using a $\text{GEV}(\mu, \sigma^*, \xi)$ the shape parameter ξ is identical in both distributions and the shape parameter σ is defined such that $\sigma = \sigma^* + \xi(u - \mu)$.

It is also possible to derive the result without using the GEV, as shown in [REF].

2.2 Discrete Extremes

A lot of Section 2.1 is appropriate only for continuous random variables and some of the results may not hold in a discrete setting. In particular, a continuous distribution F being in certain domain of attraction may not necessarily imply that a discretisation of F remains in that domain of attraction.

Shimura (2012) provides conditions for a discrete distribution to belong to the domain of attraction. In particular the following theorem which corresponds to Theorem 1 in Shimura (2012).

Theorem 2.2.1 (Domain of attraction consistency).

- (a) Every discretisation of distribution in $\mathcal{D}(\Phi_\alpha)$ remains in $\mathcal{D}(\Phi_\alpha)$.
- (b) The discretisation of a distribution remains in $\mathcal{D}(\Lambda)$ if and only if the original is in $\mathcal{D}(\Lambda) \cap \mathcal{L}$.

Where \mathcal{L} is the set of long-tailed distributions that have the property:

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x+1)}{\bar{F}(x)} = 1$$

In addition Shimura (2012) introduces a quantity useful for determining the domain of a attraction that a discrete distribution belongs to.

Definition 2.2.1 (Omega Function). For a distribution F with survival function \bar{F} and some $n \in \mathbb{Z}^+$ let:

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

This quantity plays an important role in Section 4 when determining the domain of attraction to which the degree distribution of a network generative model belongs.

Applying ideas from Section 2.1 to modelling discrete random variables has been approached from many different directions. What follows is a overview of some of the approaches that have been taken but will see use in this report.

One of the main issues when it comes to applying results from Section 2.1 to discrete data, is the discretisation of the GP distribution and making sure that it still maintains most of the same properties.

[AHMAD] introduces several ways to discretise the GP distribution including mixing the geometric distribution with a Gamma distribution, they also introduce an extended GP distribution (and its discretisation) which is obtained from a probability integral transform (PIT) of a distribution with support on $(0, 1)$. This is a modification of how the regular GP is obtained, which is from a PIT uniform distribution on $(0, 1)$.

Also considering mixtures, [VALIQUETTE] investigated the tail properties of Poisson mixtures and found that changing the mixing distribution changes the tail heaviness of the mixed distribution. They divided the domain of attraction of the mixing distribution into various subsets that lead to different tail behaviours, through the limit of the ratio of consecutive values of the survival function. These limits do help describe the tail behaviour of the resulting mixture but do not quite correspond to different domains of attraction.

The approach used in Section 4 will be very similar to the one taken in [ROHRBECK] with a minor change to the derivation, this will be explained in detail in Section 4.

2.3 Modelling

The results from Section 2.1 allow the GEV and GP to be fitted to the block maxima and threshold exceedances respectively. Typically, when fitting the GP, a sufficiently high threshold needs to be specified beforehand [REFER TO COLES 2001] and give examples.

Another more recent approach shown in [MACDONALD 2012], uses a spliced threshold mixture to model the threshold exceedances where one distribution is assumed for the bulk of the data and the GP is used for those values above the threshold. This approach can also be applied in the discrete setting, and is what is used in Section 4.

3 Networks

Networks are the structures that will be the source of data that the results from Section 2 will be used to analyse. Networks appear across a wide range of fields when attempting to represent complex systems and the relationships between the components within, showing up in anything from micro-biology (e.g. protein interactions in cells) to sociology (e.g. the social network of Harvard graduates).

This makes networks a valuable source of data and understanding the mechanics of the network generation process can provide insights to the components themselves and into the networks future.

3.1 Mathematical Definitions

Networks on the face of it are fairly simple objects, nothing more than a collection of objects with connections between each other. Here, graphs constructed from vertices and edges will be used as an analogue for these networks.

Definition 3.1.1 (Graph/Network). A graph $G = (V, E)$ is constructed from a vertex set $V \in \mathbb{Z}^+$ and an edge set E . The edge set can take on one of two forms depending on if the graph is directed or un-directed. If the graph is directed then $E \subseteq V^2$ i.e the edge set is contained within the set of ordered pairs of vertices, whereas if the graph is **un-directed** then $E \subseteq [V]^2$ i.e. the edge set is contained within the set of un-ordered pairs of vertices. The focus from now on will be on un-directed networks and graphs.

Throughout this section and the next the concept of a vertices “degree” will come up, and in fact the main focus of Section 4 is the degree distribution of networks.

Definition 3.1.2 (Degree). For an un-directed graph a vertex’s degree denoted $d(v)$ or k_v for $v \in V$ is the number of edges that are connected to vertex v :

$$d(v) = |\{e \in E : v \in e\}|$$

Directed graphs have something analogous, called the in-degree d_{in} , out-degree d_{out} and total degree d_{tot} . The in-degree of a vertex v is the number edges with endpoint at v , whereas the out-degree is the number of edges with start point at v and the total degree is the sum of these i.e.:

$$\begin{aligned} d_{in}(v) &= |\{(w_1, w_2) \in E : w_2 = v\}| \\ d_{out}(v) &= |\{(w_1, w_2) \in E : w_1 = v\}| \\ d_{tot}(v) &= d_{in}(v) + d_{out}(v) \end{aligned}$$

There are many reasons to analyse network like data, one of which is to gain an insight into the mechanics that governed the growth of the network. The next sub-section is focused on presenting several network generative models increasing in generality.

3.2 Network Generative Models

The models in this section begin with the most general considered in this report, and then two special cases of this model are introduced.

General Preferential Attachment (GPA)

Under this model, at each time step one vertex is added to the network and brings an edge with it that connects the existing vertices with a probability proportional to some function of the vertices' degrees.

Definition 3.2.1 (General Preferential Attachment Model). Starting with a graph $G_1 = (V_1, E_1) = (\{1, \dots, m_0\}, \emptyset)$, at each following time step $t > 1$ the graph is denoted by $G_t = (V_t, E_t)$ and is generated by repeating:

1. **Growth:** Add a new vertex to the vertex set i.e.

$$V_t = V_{t-1} \cup \{t\}$$

2. **Preferential Attachment:** Add $m \leq m_0$ edges connecting the new vertex to those already in the graph selected at random proportional to a function of their degree(minus one)¹ i.e.:

$$E_t = E_{t-1} \cup \tilde{E}$$

where $\tilde{E} = \{\tilde{e}_1, \dots, \tilde{e}_m\}$ and $\tilde{e}_i = \{t, \tilde{v}_i\}$ for $\tilde{v}_i \sim \text{Cat}(V_{t-1}, P)$

$$P = \left\{ \frac{g(k_v - 1)}{\sum_{w \in V_{t-1}} g(k_w - 1)} : v \in V_{t-1} \right\}$$

for some function $g : \mathbb{Z} \mapsto \mathbb{R}^+ \setminus \{0\}$, which will be referred to as the preferential attachment function

Expected Degree Distribution

In [GPA REF] the expected degree distribution for $m = 1$ was calculated in terms of the preferential attachment function does not have a general explicit form. It is defined as follows, let λ^* be the solution, if it exists, to:

$$1 = \sum_{n=1}^{\infty} \prod_{i=1}^{n-1} \frac{g(i)}{g(i) + \lambda}$$

then the expected degree distribution resulting from the GPA model has pmf:

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

Barabási-Albert (BA)

The first special case is the Barabási-Albert model, which is equivalent to setting the preferential attachment function g to be the identity function $g(k) = k$

This model defined in Barabási and Albert (1999) and also very closely related to the Yule-Simon process from [YS REF] changes the attachment mechanism from being purely uniform on the vertices already in the network to being random with a probability proportional to the degrees of the vertices in the network.

Definition 3.2.2 (Barabási-Albert Model). Starting with a graph $G_1 = (V_1, E_1)$ where $V_1 = \{1, \dots, m_0\}$ and $E_1 = \{\{v\} : v \in V_1\}$ i.e a graph with m_0 vertices with one self-loop each. At each time step $t > 1$ the graph denoted by $G_t = (V_t, E_t)$ is generated by repeating the following:

1. **Growth:** Add a new vertex to the vertex set i.e.

$$V_t = V_{t-1} \cup \{t\}$$

2. **Preferential Attachment:** Add $m \leq m_0$ edges between the new vertex and those already in the graph with probability proportional to each vertices degree i.e:

¹The probabilities are proportional to the degree minus one to align with the results from [GPA REF]

$$E_t = E_{t-1} \cup \tilde{E}$$

where $\tilde{E} = \{\tilde{e}_1, \dots, \tilde{e}_m\}$ and $\tilde{e}_i = \{t, \tilde{v}_i\}$ for $\tilde{v}_i \sim \text{Cat}(V_{t-1}, P)$

$$P = \left\{ \frac{d(v)}{\sum_{w \in V_{t-1}} d(w)} : v \in V_{t-1} \right\}$$

Expected Degree Distriubtion

In the same paper (Barabási and Albert (1999)) it was shown that for large values of t the expected degree distribution for this model is approximately:

$$f(k) = \frac{2m^2t}{m_0 + t} k^{-3} \approx 2m^2 k^{-3}, \quad k \geq m$$

This is clearly a regularly varying function and therefore in in the Fréchet domain of attraction $\mathcal{D}(\Phi_2)$.

Uniform Attachment (UA)

The final special case presented here is obtained from setting the preferential attachment function g to be some constant value.

Definition 3.2.3 (Uniform Attachment Model). Start with a graph $G_1 = (V_1, E_1) = (\{1, \dots, m_0\}, \emptyset)$, at each time step $t > 1$ the graph is denoted by $G_t = (V_t, E_t)$ and generated by repeating the following two steps:

1. **Growth:** Add a new vertex to the vertex set i.e.

$$V_t = V_{t-1} \cup \{t\}$$

2. **Attachment:** Add $m \leq m_0$ random edges between the new vertex and those already in the graph i.e.

$$E_t = E_{t-1} \cup \tilde{E}$$

where $\tilde{E} = \{\tilde{e}_1, \dots, \tilde{e}_m\}$ and $\tilde{e}_i = \{t, \tilde{v}_i\}$ and $\tilde{v}_i \sim U(V_{t-1})$.

Expected Degree Distribution

As showing in [REF] the expected degree distribution of this model for large values of t is:

$$f(k) = \frac{e}{m} \exp\left(-\frac{k}{m}\right), \quad k \geq m$$

Since this distribution has exponential form, it is in the Gumbel domain of attraction.

4 Methods

As mentioned in Section 2.3, the method used here to model the extreme values of the data will be a spliced threshold mixture. Specifically, it will be a spliced threshold mixture of a power law and a discretisation of the generalised pareto distribution similar to what is defined in [ROHRBECK].

Definition 4.0.1 (Integral Generalised Pareto Distribution (IGP)). Consider a random variable X with cdf F , and consider the random variable $Y = \lfloor X \rfloor$. From Definition 2.1.4, $X|X > u \sim GP(\sigma, \xi)$ for some sufficiently large $u \in \mathbb{R}^+$ and it can be obtained that the distribution of $Y|Y > u$ has distribution defined below:

$$\Pr(Y = y > Y > u) = \left(1 + \frac{\xi(y + 1 - \lceil u \rceil)}{\sigma_0 + \xi \lceil u \rceil}\right)_+^{-1/\xi} - \left(1 + \frac{\xi(y - \lceil u \rceil)}{\sigma_0 + \xi \lceil u \rceil}\right)_+^{-1/\xi}$$

For $y = \lceil u \rceil, \lceil u \rceil + 1, \dots$ and $\xi \in \mathbb{R}$ and $u, \sigma_0 \in \mathbb{R}^+$.

Since the some degree distributions of real networks seen in [FIG] seem to be approximately linear for the bulk of the data and then begin to change, the spliced threshold mixture that will be used consists of a truncated discrete power law for the bulk of the data and a GP above a threshold.

Definition 4.0.2 (Power-Law IGP Distribution).

$$f(y) = \begin{cases} (1 - \phi) \frac{y^{-(\alpha+1)}}{\sum_{k=1}^v}, & y = 1, 2, \dots, v \\ \phi \left[\left(1 + \frac{\xi(y+1-v)}{\sigma_0 + \xi v}\right)_+^{-1/\xi} - \left(1 + \frac{\xi(y-v)}{\sigma_0 + \xi v}\right)_+^{-1/\xi} \right] & \end{cases}$$

4.1 Fitting model to real data

4.2 GPA analyses

4.3 Conclusion and a Conjecture

5 Next Steps

References

- Barabási, Albert-László, and Réka Albert. 1999. “Emergence of Scaling in Random Networks.” *Science* 286 (5439): 509–12. <https://doi.org/10.1126/science.286.5439.509>.
- Shimura, Takaaki. 2012. “Discretization of Distributions in the Maximum Domain of Attraction.” *Extremes* 15 (September): 1–19. <https://doi.org/10.1007/s10687-011-0137-7>.