Network Science Cheatsheet

UNIVERSITÉ DE LYON

Made by
Remy Cazabet

Scale-Free Networks

Scale-Free: Definition

A network is said to be **Scale-Free** when its degree distribution follows a Power-Law distribution, or can be approximated by a Power-Law distribution.

Power-Law (PL)- Approximate Distribution

A **Power-Law distribution** is defined as follows:

$$P(k) \sim k^{-\alpha} = \frac{1}{k^{\alpha}}$$

α is called the *exponent* of the distribution.
Intuitively, the more α is large, the more large values are rare. For instance, with $\alpha = 0$, it corresponds to a uniform distribution (any degree is equivalently probable). With $\alpha = 1$, the probability of a node taken at random to have degree k is $\sim \frac{1}{k}$. Usually, a distribution is considered scale free when $2 \leq \alpha \leq 3$, as we will see.

PL - Boundaries

In most settings, the Power-Law degree distribution exists only for a certain range of degrees.
This makes sense in real networks: few people have 0 or 1 social contacts, for instance, few websites have no incoming nor outgoing hypertext links, or we wouldn't even be aware of them. Thus there is a **lower bound** k_{min} from which the distribution exists. Similarly, real networks represent entities of the real world, which are in finite numbers, therefore the number of elements itself is a limit. But in many situations, even lower thresholds exist: Social networks often impose a limit to the number of connections to avoid spammers, time and space also typically impose limits to what is possible or not in a network. An **upper bound** k_{max} can be used to limit the distribution.

Power-Law - Exact Distribution

For a distribution to be properly defined, the sum of all probability must be equal to one, we therefore add a normalization constant C to ensure this property.

$$\int P(k) = 1 = \int Ck^{-\alpha} = C \int k^{-\alpha}$$

From which we can define C:

$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\alpha} dk} = (\alpha - 1)k_{\min}^{\alpha-1}$$

And finally, the exact definition of the Power-Law degree distribution with lower bound:

$$P(k) = (\alpha - 1)k_{\min}^{\alpha-1} k^{-\alpha} P(k) = \frac{\alpha - 1}{k_{\min}} \left(\frac{k}{k_{\min}} \right)^{-\alpha}$$

Power-Law - Plotting

A famous property of the Power-Law distribution is that it looks like a line when plotted in a log-log plot, i.e., a plot in which the x-axis (degrees) and the y-axis (frequency of degrees) are represented using a logarithmic scale.
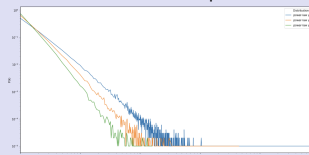
Power-Law distributions in linear scale for degrees [1-10]
(100 000 samples)
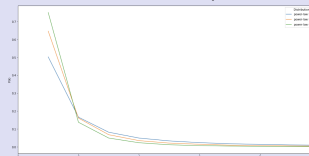
Power-Law distributions in linear scale for degrees [1-100000]
(100 000 samples). The distribution is so heterogeneous that is is not readable.

Power-Law distributions in log-log scale for degrees [1-100000]
(100 000 samples).

Power-Law - Long tail

Compared with other well-known distributions such as Poisson or exponential distribution, a key difference is what is called the **long-tail** property: very large values are rare, but possible. We can observe this long tail by comparing with other distributions on log-log plots.

Comparing power-law with Poisson distributions
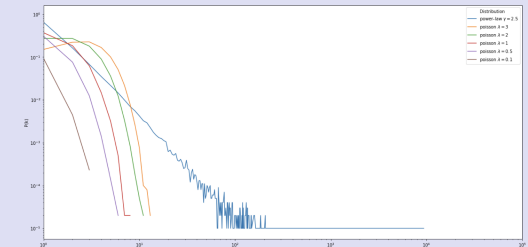
Comparing power-law with Exponential distributions
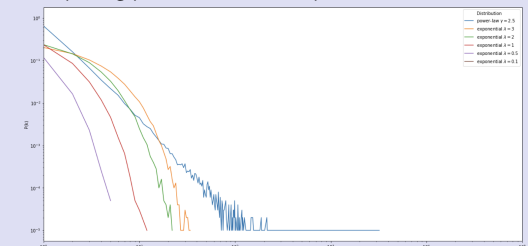
SF networks - universality

Scale-Free networks are widely studied because they are considered to be very frequent in the real world. Some important papers discovered the existence of Power-Law degree distribution in a variety of large real networks, notably:

1. The World Wide Web (webpages) (Barabási and Albert 1999)

2. The internet (physical network) (M. Faloutsos, P. Faloutsos, and C. Faloutsos 1999)

3. Airline connections (Guimera and Amaral 2004)

4. Scientific collaborations (Newman 2001)

5. Romantic interactions (Liljeros et al. 2001)

It must be noted, however, that many real world networks **are not scale-free**. A typical counter-example is a road-network, in which nodes correspond to intersection and edges to roads: for practical reasons, intersections with large degrees do not make sense.

Why is it called *scale free*

Because they have no (typical) scale!
It is defined in opposition to Poisson and other Bell-Shaped distributions, which are **centered around their average value**. Let's take a typical example: The height of humans follow a Bell-shaped distribution: the average height is 1.65m, and most humans are quite close to this value, there is a **typical scale** of human height. On the contrary, human wealth distribution follows approximately a power-law [a]: a few humans are extremely wealthy (Billions of \$), while more than half the world population posses less than 10 000\$. As a consequence, the average human wealth (70 000\$) is not at all representative of human wealth.

[a]`https://en.wikipedia.org/wiki/Distribution_of_wealth`

Central moments

The first two central moments of a distribution are the mean $\langle k^1 \rangle$ and the variance $\langle k^2 \rangle$. They are defined as

$$\langle k^m \rangle = \int_{k_{\min}}^{\infty} k^m p(k)dk = (\alpha - 1)k_{\min}^{\alpha-1} \int_{k_{\min}}^{\infty} k^{-\alpha+m}dk$$

From this, we can conclude that **central moments are defined only if $\alpha > m+1$**, otherwise they diverge towards infinity, they are not properly defined.
Thus:

1. $\langle k^2 \rangle = \frac{\alpha-1}{\alpha-2}k_{\min}$, **if and only if $\alpha \geq 2$**

2. $\langle k^3 \rangle = \frac{\alpha-1}{\alpha-3}k_{\min}^2$, **if and only if $\alpha \geq 3$**

Divergence in practice

In practice, one can always compute the mean and variance of a provided, observed degree distribution. So what does it mean that they *diverge*?
The problem arises when we are not certain to observe the whole network. Usually, a large sample of a population has the same mean and variation than the whole population, and the largest the sample, the more precise the value.
But in a power law, moments are **dominated** by the largest values in the long tail: some rare values are *so large* that they shift the moments. So the more data we observe, the higher the moments.

Divergence: consequences

The consequence of diverging moments is that **if the distribution follows a power law**, then if the exponent is below 2, you should now rely on the mean degree or the variance. If the exponent is between 2 and 3, you can (relatively) rely on the mean, but not on the variance. Be careful though, even if $\alpha > 2$, the mean converges slowly, i.e., you need a very large sample for your mean to be close to the real value.

Fitting power laws

When confronted with a power law degree distribution, we might want to **fit** the distribution, i.e., to **find the exponent** of the distribution. A naive and simple way to do it is to plot the distribution on a log-log plot and to find the slope of the line, either graphically or through least-square regression on the log-transformed values of degrees and frequencies.
This however suffers from a strong bias: values in the tail are based on a few samples, and introduce noise.
The most appropriate method is to use Maximum Likelihood Estimation (MLE[a]), taking into account min and max-boundaries, as described in[b]

[a]https://towardsdatascience.com/a-gentle-introduction-to-maximum-likelihood-estimation-9fbff27ea12f

[b]Goldstein, Morris, and Yen 2004.

Exponent limits

In real networks, we consider that we should have $\alpha \geq 2$, because a lower exponent would mean that the distribution is so skewed that we expect to find nodes with a degree larger than the size of the network.
Furthermore, if the exponent is too large, large degree nodes becomes so rare, that the network would need to be enormous to observe such a node. For instance, with $\alpha = 5$, we need to observe $N = 10^{12}$ nodes to expect to observe a single node of size 1000.

Exponent and shortest-paths

Random networks with Poisson degree distribution already have a *short average distance*. However, it is possible to define classes of networks with even smaller average distance based on the exponent α:

- $\alpha = 2$: The biggest hub degree is of order $\mathcal{O}(N)$, thus most nodes are at distance 2. The average path length can be considered a small constant, independent of N

- $2 < \alpha < 3$: **Ultra Small World**: $\langle \ell \rangle = \frac{\log \log N}{\log(\alpha-1)}$

- $\alpha = 3$: $\langle \ell \rangle = \frac{\log N}{\log \log N}$

- $\alpha > 3$: $\langle \ell \rangle = \log N$, the network behaves approximately like an ER network.

Scale-free network controversy

There is an on-going debate in the network science community over the prevalence of scale-free networks. For some authors[a], most real networks follow to some extent a power-law degree distribution, while, for some others, scale-free networks are rare[b]. The controversy has been studied[c] and can be interpreted as differences between scientific approaches: one popular among (some) physicists (*scale-freeness is the sign of a universal law*) and another one common among statisticians (*scale-freeness is an empirical characterization*).

[a]Barabási and Bonabeau 2003.
[b]Broido and Clauset 2019.
[c]Jacomy 2020.

SF networks: what to do

Is your network Scale-Free? The first question you might ask yourself is: *why do you need to know?*.

- If the goal is to characterize a network, then plotting the degree distribution might be more useful than fitting a power-law exponent to it

- If the goal is to show that the distribution is broad, significantly different from a bell-shaped distribution, then plotting the distribution might be enough

- If the goal is to show that the distribution is *approximately* a power-law, for instance because an algorithm complexity or a proof can be made for such cases, then fitting a line on a log-log plot and talking about *power-law-ish* might be enough

- If on the contrary it is scientifically important to argue that the network is, without doubts, a scale-free networks, then you need to be fully aware of the controversy and to position your work relatively to it.

References

[1] Albert-László Barabási and Réka Albert. "Emergence of scaling in random networks". In: *science* 286.5439 (1999), pp. 509–512.

[2] Albert-László Barabási and Eric Bonabeau. "Scale-free networks". In: *Scientific american* 288.5 (2003), pp. 60–69.

[3] Anna D Broido and Aaron Clauset. "Scale-free networks are rare". In: *Nature communications* 10.1 (2019), pp. 1–10.

[4] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. "On power-law relationships of the internet topology". In: *ACM SIGCOMM computer communication review* 29.4 (1999), pp. 251–262.

[5] Michel L Goldstein, Steven A Morris, and Gary G Yen. "Problems with fitting to the power-law distribution". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 41.2 (2004), pp. 255–258.

[6] Roger Guimera and Luis A Nunes Amaral. "Modeling the world-wide airport network". In: *The European Physical Journal B* 38.2 (2004), pp. 381–385.

[7] Mathieu Jacomy. "Epistemic clashes in network science: Mapping the tensions between idiographic and nomothetic subcultures". In: *Big Data & Society* 7.2 (2020), p. 2053951720949577.

[8] Fredrik Liljeros et al. "The web of human sexual contacts". In: *Nature* 411.6840 (2001), pp. 907–908.

[9] Mark EJ Newman. "The structure of scientific collaboration networks". In: *Proceedings of the national academy of sciences* 98.2 (2001), pp. 404–409.