

## Lecture 3 – Popularity-Power Law

Instructor: *Augustin Chaintreau*Scribe: *Bohyun Kim*

As we discussed in Lecture 2, power law seems to take place in many different scenarios. For this reason, it is often called "the ubiquitous Power Law". For example, one can observe the Power Law in many popularity of items such as Amazon, YouTube, and Flickr. They are also found in a broad spectrum of natural systems such as number of species in genus or protein sequences in genome.

**Skewed and power law :** Generally speaking a distribution is skewed, or heavy tailed, when very large values occur with non-negligible probability, and these large values also account for an important fraction of the total sum of values. As an example, if we consider how income are distributed (i.e. how money is divided and given to each individual per month), a skewed distribution would typically have a small percentage with very large income, that would overall receive an important fraction of the total wealth.

The power law distribution is one example of such skewed distribution. A characteristic of a power Law distribution is that its CCDF and its density function, plotted using a logarithmic scale for both x-axis and y-axis, both decrease linearly. In contrast, exponential or light-tailed distribution decreases faster than linearly when these same scales are used.

**Explanation :** Previous works have exhibited three possible sources (or explanation) for the appearance of skewed distribution such as power law: first is reinforcement, second is optimization and the third is artefacts of random stopping. We will first examine the reinforcement case in Power Law.

## 1 Reinforcement

The particular case which we will study has been originally motivated by the appearance of species inside a genus (in classification of plants and animals), which were found to obey also power law distribution. We can see that new specie or genus can appear by mutation and the key to this phenomenon is that a genus containing more species is more likely to see mutation occurring among them.

**Dynamics of a simple balls and bins reinforcement process** The following dynamics have been introduced by G. Yule in 1925 to model evolution of species and their repartition within different classes, or "genus". We model a species as a ball which is contained in a bin that model its associated class or genus. Once a species appear it is never removed and it does not change genus (*i.e.*, once a ball is created, it remains forever in the same bin). The main dynamics is the apparition of new species.

We assume that the process is initialized at time  $t = 0$  with an arbitrary number of genera containing each an arbitrary number of species. The time follows discrete time slots denoted  $t = 1, 2, \dots$ , and we introduce the following notation:

$\forall t \geq 0, \forall i \geq 0$ ,  $X_i(t)$  is the number of genera containing exactly  $i$  species.

Hence the distribution of species among genus at time  $t$  is entirely described by the sequence  $(X_i(t))_{i \geq 1}$ . In particular,

We assume that during a time slot, a mutation occurs in exactly one species, that is chosen uniformly at random among all species. The consequence of this mutation is the following:

- With a probability  $p > 0$  (chosen independently of the past), this mutation is so important that it creates a new genus by itself. As a consequence, a new bin is created that contains exactly one ball. All other bins remains unchanged.
- Otherwise (and hence with probability  $(1 - p)$ ), this mutation generates a new species which is contained in the same genus as the original one that was chosen. As a consequence, a new ball is created in the bin that contained the original ball that was chosen for the mutation.

Intuitively, in the second case, the process entails *reinforcement* for the following reason: in a genus that contains twice more species, the chance of having a mutation is twice more likely and hence this genus is also twice more likely to “grow” by having another species created and associated with it. In other words, the larger a genus becomes, the faster it will grow.

The following result proves formally that this reinforcement implies that as time grows large, the number of species contains by a genus approaches a power law.

**Theorem 1** (Analysis of Yule Process). *For the dynamics described above,*

- (i) *There exist  $C_1, C_2, \dots$  such that for any  $i \geq 1$  we have,  $X_i(t)/t \rightarrow C_i$  almost surely as  $t$  grows large.*
- (i) *We have  $C_1 = \frac{p}{2-p}$  and  $C_i = C_{i-1}(1 - a/i + O(i^{-2}))$  where  $a = \frac{2-p}{1-p}$*
- (i) *this implies  $\ln(\frac{C_i}{C_1}) \sim -\alpha \ln(i)$  which implies that  $C_i$  is roughly proportional  $i^{-a}$  and hence is well approximated by a power law.*

*Proof.* The proof follows from two ingredients

1. Analysis of the expected value evolution
2. A probabilistic concentration result and its consequence

Let  $X_1(t)$  be the number of genus with exactly 1 species, we wish to prove  $\frac{X_1(t)}{t} \rightarrow C_1$ .

Let us first translate the dynamics of the system in a given step into the evolution of the variable  $X_1$  between time slots  $t$  and  $t + 1$ .

Let us denote by  $N(t)$  the number of species in the system. Note that since during each time slot a mutation occurs and create exactly one species, we have  $N(t) = N(0) + t$ .

We have:

$$X_1(t+1) = \begin{cases} X_1(t) + 1 & \text{with probability } p \\ X_1(t) - 1 & \text{with probability } (1-p) \frac{X_1(t)}{N(t)} \\ X_1(t) & \text{otherwise} \end{cases}$$

The first case follows from the fact that with probability  $p$  a new genus is created with exactly one species. The second case represents the case when the mutation occurs in one of the  $X_1(t)$  species associated with genera containing a single species, and it is not creating a new genera. Indeed in this case, one of these

genera will contain two species in the next time slot. Finally the last case denotes any other event.

Finding the Expected Value of  $X_1(t)$

As a consequence of the previous dynamics we can very precisely characterize the evolution of the *expectation* of  $X_1(t)$  with time  $t$ :

That is to multiply the probability of each event by the each case which are defined above

$$E[X_1(t+1)] = p * E[X_1(t) + 1] + \frac{E[X_1(t)]}{N(t)} * (1-p) * E[X_1(t) - 1] + \left\{1 - p - \frac{E[X_1(t)]}{N(t)} * (1-p) * E[X_1(t)]\right\}$$

$$E[X_1(t+1)] = E[X_1(t)] + p - \frac{E[X_1(t)]}{N(t)} * (1-p)$$

How about  $X_i(t)$  and its expected value?

Let  $X_i(t)$  be a number of genus with exactly  $i$  species. We wish to show  $\frac{X_i(t)}{t} \rightarrow C_i$

$$X_i(t+1) = \begin{cases} X_i(t) + 1 & \text{with probability } (i-1) \frac{X_{i-1}(t)}{N(t)} (1-p) \\ X_i(t) - 1 & \text{with probability } \frac{iX_i(t)}{N(t)} (1-p) \\ X_i(t) & \text{otherwise} \end{cases}$$

The same principle applies for finding the expected value for  $X_i(t)$ , that is to multiply the probability of each event.

$$E[X_i(t+1)] = E[X_i(t)] + \frac{(i-1)E[X_{i-1}(t)]}{N(t)} (1-p) - \frac{iE[X_i(t)]}{N(t)} (1-p)$$

**Limit of expected value of  $X_i(t)/t$**  Let  $\Delta_i(t) = E[X_i(t)] - t \cdot C_i$ , we want  $\frac{\Delta_i(t)}{t} = 0$  as  $t$  goes to  $\infty$ .

Let us first prove it for  $i = 1$ . We want to prove that  $\Delta_1(t)$  is small According to the above evolution of  $\mathbb{E}[X_1(t)]$  we have:

$$\begin{aligned} \Delta_1(t+1) &= \Delta_1(t) - C_1 + E[X_1(t+1)] - E[X_1(t)] \\ &= \Delta_1(t) - C_1 + p - \frac{\Delta_1(t) + t \cdot C_1}{N(t)} (1-p) \end{aligned}$$

Putting all factors of  $\Delta_1(t)$  together, we obtain

$$= \Delta_1(t) \left[1 - \frac{1-p}{N(t)}\right] - C_1 + p - t \cdot \frac{C_1(1-p)}{N(t)}$$

$$\underbrace{\frac{(-C_1+p)N(t) - C_1(1-p)t}{N(t)}}_{\text{where } N(t) \text{ is nothing but } N(0) + t}$$

the underbraced expression may be rewritten as:  $\frac{(-C_1+p)N(0) + t[-C_1+p - C_1(1-p)]}{N(t)}$

As  $t$  grows large,  $N(t)$  grows large as well. In order to show that this term becomes small, we wish to have the coefficient multiplying  $t$  to be zero. That is we assume:

$$-C_1 + p - C_1(1-p) = 0 \implies C_1 = \frac{p}{2-p}.$$

Note that we can assume that since, until now, the value of the constant  $C_1$  was not fixed and can be chosen arbitrarily for all these results to hold.

Now we have that the underbraced expression reduces to a term becoming small as  $t$  grows:

$$\Delta_1(t+1) = \Delta_1(t) \left[1 - \frac{1-p}{N(t)}\right] + \frac{(-C_1+p)N(0)}{N(t)}$$

Since the term multiplying  $\Delta_1(t)$  is less than 1 in absolute value and the right term is less than  $\frac{A}{t}$  for a constant  $A > 0$  chosen sufficiently large, we can apply Lemma 2 below

$$|\Delta_1(t+1)| \leq \Delta_1(0) + A \sum_{s=1}^t \frac{1}{s}$$

Hence, using Lemma 3, we deduce that there exists  $A' > 0$  such that

$$\Delta_1(t) \leq A' \ln(t), \text{ which proves in particular that } \frac{\Delta_1(t)}{t} \text{ goes to zero as } t \text{ gets large.}$$

We wish to prove the following hypothesis for any  $i \geq 1$ :  $\forall \varepsilon > 0 \exists A$  such that  $(|\Delta_i(t)| \leq At^\varepsilon)$ ,

Indeed, we have just shown that this is true for  $i = 1$  since we found a logarithmic upper bound on the size of  $\Delta_1(t)$ . By recurrence, it is sufficient to prove that if it holds for  $i - 1$  it holds for  $i$  as well.

Note that, following similar steps as used for  $i = 1$  (rewriting evolution of expectation  $\mathbb{E}[X_i(t+1)]$  using  $\Delta_i(t)$  and  $\Delta_{i-1}(t)$ ), we have:

$$\Delta_i(t+1) = \Delta_i(t) \left( 1 - \frac{i(1-p)}{N(t)} \right) + \frac{(i-1)(1-p)}{N(t)} \Delta_{i-1}(t) + \underbrace{\left( -C_i + \frac{(i-1)(1-p)t \cdot C_{i-1}}{N(t)} - \frac{i(1-p)t \cdot C_i}{N(t)} \right)}_{= \frac{N(0)C_i + t \cdot (-C_i + (1-p)(i-1)C_{i-1} - (1-p)iC_i)}{N(t)}}.$$

Note that, again the value of  $C_i$  is not fixed so that we can choose  $C_i$  so that the coefficient of  $t$  in the underbraced term is zero. This implies:

$$(1-p)(i-1)C_{i-1} = C_i + (1-p)iC_i$$

$$\text{or, in other words } C_i = C_{i-1} \left( 1 - \frac{2-p}{1+(1-p)i} \right) = C_{i-1} \left( 1 - \frac{2-p}{(1-p)i} + \underbrace{\frac{2-p}{(1-p)i(1+(1-p)i)}}_{\leq \frac{A}{i^2} \text{ for any constant } A \text{ larger than } \frac{2-p}{(1-p)^2}} \right).$$

Once this is shown, for any  $\varepsilon > 0$ , we can use the hypothesis for  $i - 1$  to deduce that  $\Delta_{i-1} \leq At^\varepsilon$  and hence

$$\Delta_i(t+1) = \Delta_i(t)\gamma_t + S_t$$

where  $|\gamma_t| < 1$  and  $|S_t| \leq (i-1)(1-p)At^{\varepsilon-1} + A'/t \leq A''t^{\varepsilon-1}$ . We can then conclude, using lemma 4 below that the hypothesis remains true for  $i$ .

**Lemma 2.** *If  $\chi_{n+1} = \gamma_n \chi_n + S_n$ , where  $|\gamma_n| \leq 1$ , then we have  $|\chi_n| \leq |\chi_0| + \sum_{m=1}^n |S_m|$ .*

**Lemma 3.** *For any  $n \geq 1$  we have:  $\sum_{j=1}^n \frac{1}{j} \leq 1 + \ln(n)$*

**Lemma 4.** *For any  $\varepsilon > 0$  and  $n \geq 1$  we have:  $\sum_{j=1}^n j^{\varepsilon-1} \leq 1 - \frac{1}{\varepsilon} + \frac{1}{\varepsilon} j^\varepsilon$ .*

**Second ingredient probabilistic concentration result:** So far, we have been able to prove that there exists  $C_1, C_2, \dots$  such that, as  $t$  grows, the expectation of  $X_1(t), X_2(t), \dots$  grows approximately as  $C_1 \cdot t, C_2 \cdot t$  etc. (i.e., for all  $i \geq 1$ , we have  $\lim_{t \rightarrow \infty} \frac{\mathbb{E}[X_i(t)]}{t} = C_i$ ). We now wish to establish a much stronger result, comparable to a law of large number, which states that the sequence of random variables  $X_i(t)$  grows approximately as  $C_i \cdot t$  (i.e.  $\frac{X_i(t)}{t}$  converges to  $C_i$ , with probability 1).

To prove this we use the following probabilistic concentration result which states that, as  $t$  goes large, the sequence  $X_i(t)$  is not far from its average value. More precisely, we admit the following result:

$$\forall i \geq 1, \forall M > 0, \text{ we have } P[|X_i(t) - E[X_i(t)]| > M] \leq 2 * \exp(-\frac{M^2}{8t}). \quad (1)$$

If we choose  $M = \sqrt{t * \ln(t)} * 4$

$$\begin{aligned} \frac{-M^2}{8t} &= \frac{-16t \ln(t)}{8t} = -2 \ln(t) \\ \exp(\underbrace{\frac{-M^2}{8t}}_{-2 \ln(t)}) &= \exp(-2 \ln t) = \exp(\ln(t^{-2})) = \frac{1}{t^2} \end{aligned}$$

We then have  $P[|X_1(t) - E[X_i(t)]| > M] = \frac{1}{t^2}$

Then we can say,

$$\begin{aligned} \implies P[|X_i(t) - E[X_i(t)]| > M] &\leq \frac{1}{t^2} \\ \implies \sum_{t=0}^{\infty} P[|X_i(t) - E[X_i(t)]| > M] &\leq \sum \frac{1}{t^2} < \infty \end{aligned}$$

Let us now conclude. For any  $i \geq 1$  using Lemma 5 below (known as Borel Cantelli lemma), we know that there exists  $T$  which is almost surely finite such that when  $t \geq T, |X_i(t) - E[X_i(t)]| \leq M$ , which also implies that  $|\frac{X_i(t)}{t} - \frac{E[X_i(t)]}{t}| \leq \frac{M}{t} = 4\sqrt{\frac{\ln(t)}{t}}$

Since  $\lim_{t \rightarrow \infty} \frac{E[X_i(t)]}{t} = C_i$  (note that this is a deterministic convergence for a sequence of real numbers). This means that almost surely (*i.e.*, on the event  $\{T < \infty\}$ ), the values taken by the sequence of random variable  $\frac{X_i(t)}{t}$  form a sequence of real numbers which are converging to a convergence sequence. This implies that these values form a sequence that converge to the same limit, and hence that almost surely (*i.e.*, on the event  $\{T < \infty\}$ ),  $\frac{X_i(t)}{t}$  converges to  $C_i$ .

**Lemma 5.** *Let  $(A_n)_{n \geq 0}$  be a sequence of event satisfying  $\sum P(A_n) < \infty$ .*

*There exists  $N$  which is finite almost surely such that for any  $n \geq N$ ,  $A_n$  does not occur.*

**Finally, let us prove that  $C_i$  is well approximated by a power law with coefficient  $a = \frac{2-p}{1-p}$ :**

We have seen that  $C_i = C_{i-1}(1 - \frac{\alpha}{i} + O(\frac{1}{i^2}))$

Note that this implies that there exists  $A > 0$  such that  $C_{i-1}(1 - \frac{\alpha}{i} - \frac{A}{i^2}) \leq C_i \leq C_{i-1}(1 - \frac{\alpha}{i} + \frac{A}{i^2})$

Which may be rewritten  $C_1 \prod_{j=1}^i (1 - \frac{\alpha}{j} - \frac{A}{j^2}) \leq C_i \leq C_1 \prod_{j=1}^i (1 - \frac{\alpha}{j} + \frac{A}{j^2})$

This implies  $\ln(C_1) + \sum_{j=1}^i \ln(1 - \frac{\alpha}{j} - \frac{A}{j^2}) \leq \ln(C_i) \leq \ln(C_1) + \sum_{j=1}^i \ln(1 - \frac{\alpha}{j} + \frac{A}{j^2})$

Hence  $\exists A' > 0$  such that  $\ln(C_1) + \sum_{j=1}^i \frac{\alpha}{j} - \sum_{j=1}^i \frac{A'}{j^2} \leq \ln(C_i) \leq \ln(C_1) - \sum_{j=1}^i \frac{\alpha}{j} + \sum_{j=1}^i \frac{A'}{j^2}$

Note that  $\alpha \sum_{j=1}^i \frac{1}{j} \sim -\alpha \ln(i)$  as  $i$  grows. This implies the result as the two other series are convergent and hence bounded.

If we assume that this approximation is exact we obtain

$$\implies \ln(C_i) = \ln(C_1) - \alpha * \ln(i)$$

$$\implies C_i = C_1 * i^{-\alpha}$$

$$\implies C_i \propto i^{-a}$$

This equality is not true in general (since the approximation introduces a constant that may play a role after being exponentiated). But one can show that there exist two constants  $A, B$  such that

$Ai^{-\alpha} \leq C_i \leq Bi^{-\alpha}$ , so that  $C_i$  remains not too far from a power law with the corresponding exponent.  $\square$

**Example 6.** *The copying model*

We take nodes as webpages to join the graph in sequence, creating edges linking to a previous webpage. Each webpage will create  $k$  outgoing edges as follows.

1. Choose a node  $v$  uniformly.
2. With probability  $p$ , edge  $u \rightarrow v$  is created.
3. With probability  $(1-p)$ , edge  $u \rightarrow w$  for  $w$ , a child of  $v$ .

We will have the fraction of nodes with  $i$  incoming links  $= i^{-\alpha}$  and  $\alpha = (2 - p)/(1 - p)$

**Example 7.** *How to make our own Power Law?*

- We need many apples
- With fixed probability  $p$  and for every  $m$  apple, send it to  $x$  who already received one. Then ask  $x$  to send it to one who received an apple by her  $\Rightarrow$  REINFORCEMENT ACHIEVED

## 2 Optimization

We first need to show that the Power laws are optimal by using resources in the most efficient manner and creating the optimal topology. Secondly, we will need to show that the systems are lead to optimal.

**Example 8.** *Word Frequency*

The first example of Power Law is a word frequency. When we need to convey information, we tend to use short words more frequently and longer words almost rarely.

The Mandelbrot experiment was to design a language over a  $d$ -ary alphabet to optimize information per character, where the probability of  $j$ th most frequently used word is  $p_j$  and the length of  $j$ th most frequently used word is  $c_j$ .

We show that by making the optimization ratio  $A = \text{MAX}(H/C)$  where the average information per word as  $H = -\sum_j p_j \log_2 p_j$ , and the average characters per word as  $C = \sum_j p_j c_j$

For any  $j$ , derive with respect to  $p_j$ ,  $H'_j/C'_j = H/C$ , and if  $c_j = \log_d j$ , then Power Law results.

**Example 9.** *Internet Topology*

In this example, we will assume that nodes are joined sequentially, that nodes will be chosen uniformly on a plane. We then choose a neighboring node to connect to minimize its cost to destination, which in mathematical formula can be written as  $\text{Cost}(i) = \alpha d(i, j) + h_j$ , where  $d(i, j)$  is the euclidean distance and  $h_j$  is the network distance.

The theorem has simply three cases that,

- if  $\alpha < 2^{-0.5}$  then the network is a star
- if  $\alpha < c\sqrt{n}$  then degrees are exponential
- if  $4 < \alpha < o(\sqrt{n})$  then degrees are heavy tailed

**Example 10.** *P2P topology*

$-q_i$  = rate for item  $i$  is requested

$-p_i$  = fraction of cache allocated to item  $i$

We want  $\min \sum \frac{q_i}{p_i}$  where  $\frac{1}{p_i}$  is the expected search time for item  $i$

If  $q_i$  follows PL then  $p_i$  also follows PL

$$C = \sum_{i=1}^{m-1} \frac{q_i}{p_i} + \frac{q_m}{1 - \sum_{i=1}^{m-1} p_i}$$
$$\frac{\partial C}{\partial p_i} = q_i * \frac{-1}{p_i^2} + q_m (-1) \frac{1}{pm^2}$$
$$\frac{p_i^2}{p_m^2} = \frac{q_m}{q_i} \Rightarrow p_i = \frac{\sqrt{q_i}}{\sum_j \sqrt{q_j}}$$

### 3 Artifacts of Random Stopping

**Example 11.** Assume that one produces a large corpus randomly and the space bar is hit with probability,  $p$ . Each character has the same probability of hit,  $q = (1-p)/d$ . Clearly, all words of length  $j$  has the same probability,  $pq^j$ , and there are  $d^j$  such words. Hence  $P[\text{word with rank } i] = pi^{-\alpha}$  with  $\alpha = (-\ln(q)/\ln(d))$ . It should come intuitive to one that random corpus also exhibits Power Law.

**Example 12.** Internet router degrees also follow Power Law. Traceroute picks a source and computes the shortest path to many different destinations. Traceroutes do not discover all edges, however, as it needs to be on the shortest path from a source to a destination. Moreover, this reinforces the presence of edges near the source. There is a theorem, which states that for general degree distance random graphs, the degree distribution changes radically and they may exhibit heavy tail/Power Law artificially.

### 4 Summary

It is no doubt that Power Law exists everywhere. The skewed distribution can be explained by reinforcement and proportional effect. This also applies to degree distribution in graph, that also exhibits small-world and "hub" structure. One should be careful that both lognormal and Power Law may be obtained and that those graphs obtained do not exhibit clustering coefficient. To summarize, the frequency of items, popularity of objects and distributions of many variables are highly skewed. This proves that it cannot be the sum of independent effect and it can be reinforcement or the result of a bias, which typically is random stopping. It's also good to note that in some cases such as caching, it can relate to optimal efficiency.