Preferential Attachment

Doug Jungreis

Power-Law Degree Distribution

- Graph has “power-law degree distribution” if number of nodes of degree k is proportional to $k^{-\beta}$ for some constant β.
- Approximate power-law distributions have been observed in many complex networks:

network	nodes	edges
Web	web-pages	hyperlinks
Internet	routers	physical connections
collaboration	authors	co-authored papers
telephone	phone nums	phone calls
biological	proteins	bindings

Why Are They Common?

- One explanation: Preferential Attachment
- Graph grows over time.
- Add one half-edge at a time.
- Two assumptions:
  1. With fixed probability p, the new half-edge introduces a new node.
  2. Otherwise, new half-edge goes into existing node chosen with probability proportional to degree.
- Resulting graph has power-law distribution.

Do Assumptions Make Sense?

1. With fixed probability p, the new half-edge introduces a new node.
   - Average degree always $\approx 1/p$.
   - Doesn't change over time.
   - Plausible, for some networks.
2. Otherwise, new half-edge goes into existing node chosen with probability proportional to degree.
   - If a has made twice as many calls as b, he's twice as likely to make next one.
   - But that assumes they've had phones for same amount of time.
   - Seems unlikely.

Poor Motivation

- Power-law distributions are supposed to motivate preferential attachment.
- In fact, preferential attachment is **not** a good explanation for power-laws.
- Homework: Find the good explanation.
- Talk should end here.
- Instead, I'll talk about preferential attachment and generalizations.

Balls and Bins

Consider only nodes and half-edges. (Pairing of half-edges not relevant to degree distrib.) Replace nodes by bins, half-edges by balls.

Process: Fix p and γ. Given bins each with one ball, introduce balls one at a time. With probability p, ball gets a new bin. Otherwise, probability ball goes into a bin is proportional to k^γ, where $k =$ number of balls in that bin.

Question: What happens to the bin-sizes?

Six cases:

- $p = 0$ or $p > 0$ (Are new bins created?)
- $\gamma < 1$, $\gamma = 1$, or $\gamma > 1$.

Polya's Urn Problem: $p = 0, \gamma = 1$

Problem: Start with two bins, each with one ball. Never introduce bins. Probability ball goes into bin is proportional to number of balls in bin. After a long time, what fraction of balls is in each bin?

Answer: Fraction of balls in first bin converges to some real number $x \in [0, 1]$. Furthermore, distribution of x is uniform.

Generalizes to > 2 bins.

$p = 0$, $\gamma \neq 1$

Easy to couple this with Polya's Urn Problem.

If $\gamma > 1$, number of balls in k largest bins is always at least number of balls in k largest bins for $\gamma = 1$ case. Fraction of balls going into largest bin must then converge to 1.

If $\gamma < 1$, number of balls in k largest bins is always at most number of balls in k largest bins for $\gamma = 1$ case. All bins must then converge to same size.

$p \neq 0$

Let n_i and f_i denote number and fraction of bins with i balls. Suppose f_i converges for each i, and suppose $\sum f_i i^\gamma$ converges. Then they're easy to compute.

Expected change in n_i is proportional to $n_{i-1}(i-1)^\gamma - n_i i^\gamma$. This must be proportional to n_i. Thus, for some constant K:

$$
\begin{aligned}
K n_i &= n_{i-1}(i-1)^\gamma - n_i i^\gamma, \\
n_i &= \frac{n_{i-1}(i-1)^\gamma}{K + i^\gamma}, \\
f_i &= \frac{f_{i-1}(i-1)^\gamma}{K + i^\gamma},
\end{aligned}
$$

Resulting Distributions

Solving the recurrence, we find the following types of distributions, depending on γ.

$$f_i \;\propto\; \begin{cases} i^{-(1+1/(1-p))} & \text{if } \gamma = 1, \\ i^{-\gamma} e^{-K i^{1-\gamma}/(1-\gamma)} & \text{if } 0 < \gamma < 1, \\ (K+1)^{-i} & \text{if } \gamma = 0, \\ O\left(\frac{((i-1)\,!)^{\gamma}}{K^i}\right) & \text{if } \gamma < 0. \end{cases}$$

We get power-law distribution only if $\gamma = 1$.

Do f_i and $\sum f_i i^\gamma$ converge?

- If $\gamma > 1$, the latter does not.
- Homework: prove convergence for $\gamma < 1$.
- Easy to prove convergence for $\gamma = 1$.

Convergence when $\gamma = 1$

- Let $x =$ fraction of balls in size 1 bins.
- Net rate of change of number of balls in size 1 bins is $p - (1-p)x$:
  - Bins of size 1 appear at rate p.
  - Disappear at rate $(1-p)x$.
- In limit, $x = p - (1-p)x$, i.e. $x = p/(2-p)$
  - If $x < p - (1-p)x$, then x grows.
  - If $x > p - (1-p)x$, then x shrinks.
- Then can repeat the argument for size 2 bins, then size 3, etc.

$p > 0$ and $\gamma > 1$

Theorem: If $p > 0$ and $\gamma > 1$, then one bin dominates, i.e., probability goes to 1 that a new ball either goes in a new bin or goes into that one bin.

Proved by Chung-Graham/Handjani/me; similar result proved by Oliveira/Spencer.

Step 1 of Proof

Size of largest bin grows arbitrarily large:

- Probability i^{th} ball goes into largest bin is at least $(1-p)/i$.
- $\sum_{i=1}^{\infty}(1-p)/i$ is infinite.
- Ball is placed in largest bin infinitely often.
- Size of largest bin grows infinitely often.

Step 2 of Proof

Always nearly p of balls are in "small" bins.

- When we have i balls, we have $\approx ip$ bins.
- Average bin-size is $1/p$.
- At most $1/N$ bins bigger than N/p.
- E.g., 99.9% of bins smaller than $1000/p$.
- Then at least $\approx (1 - 1/N)p$ of balls are in bins smaller than N/p.

Step 3 of Proof

Largest bin grows at almost linear rate:

- Let f be fraction of balls in largest bin.
- Prob new ball goes in existing bin is $1-p$.
- Since largest bin grows arbitrarily large, and $\gamma > 1$, balls in small bins become negligible for computing prob that new ball goes into largest bin.
- About p of balls in small bins, so prob ball goes into largest bin, given it goes into existing bin, is at least close to $f/(1-p)$.
- Prob ball goes into largest bin is almost f.

Step 4 of Proof

If $\frac{k}{k-1} < \gamma$, then only finitely many bins ever reach size k. (E.g., if $\gamma > 2$, then only finitely many bins ever get a second ball.)

- Measure time t by number of balls.
- Size of largest bin is almost order t.
- Probability small bin gets t^{th} ball is at most approximately order $t^{-\gamma}$.
- Probability bin created at time t_1 gets balls at times $t_2, \ldots, t_k$ is approximately of order $(t_2 \cdots t_k)^{-\gamma}$.
- Integrate.

End of Proof

- At time t, number of bins is $\leq t$; largest bin has size almost of order t.
- Prob that ball goes in bin smaller than k is at most $\approx \frac{tk^\gamma}{t^\gamma}$, which goes to 0.
- Prob goes to 1 that ball goes into the finitely many bins with $\geq k$ balls.
- Reduced to finite case ($p = 0$).
- One bin dominates.