Fragmentation of Random Trees

Eli Ben-Naim
Los Alamos National Laboratory

with: Ziya Kalay (Kyoto University)

poster & paper available from: http://cnls.lanl.gov/~ebn

Random Graph Processes, Austin TX, March 11, 2016
Formation of a Random Tree

- Start with a single node, the root
- Nodes are added one at a time
- Each new node links to a randomly-selected existing node
- A single connected component with $N$ nodes, $N-1$ links
- Degree distribution is exponential
  $$n_k = 2^{-k}$$
- In-component degree distribution is power-law
  $$b_s = \frac{1}{s(s + 1)}$$
Fragmentation of a Random Tree

- Nodes are removed one at a time: many previous studies on removal of links [Janson, Baur, Bertoin, Kuba]
- When a node is removed, all links associated with it are removed as well
- Random Forest: a collection of trees formed by the node removal process
- Degree distribution of individual nodes is known (Moore/Ghosal/Newman PRE 2006)

What is the size distribution of trees in the forest?
Main Result:
Size Distribution of Trees in Random Forest

distribution of trees of size $s$ is controlled by one parameter: fraction $m$ of remaining nodes*

$$\phi_s = \frac{1 - m}{m^2} \frac{\Gamma(s)\Gamma\left(\frac{1}{m}\right)}{\Gamma(s + 1 + \frac{1}{m})}$$

size distribution has a power-law tail

$$\phi_s \sim s^{-1 - \frac{1}{m}} \quad \text{for} \quad s \gg 1$$

*exact result, valid in the infinite $N$ limit
Removal of a Single Node

- Remove a single, randomly-chosen, node from a random tree with $N$ nodes.
- Let $P_{s,N}$ be the average number of trees with size $s$.
- Two “conservation” laws:
  \[ \sum_s P_{s,N} = \frac{2(N-1)}{N} \] and \[ \sum_s sP_{s,N} = N - 1 \]
  - Tree with $N$ nodes has $N-1$ links.
  - Every link connects two nodes.
- Recursion equation (add node to original random tree):
  \[ P_{s,N+1} = \frac{N}{N+1} \left( \frac{s-1}{N}P_{s-1,N} + \frac{N-s}{N}P_{s,N} \right) + \frac{1}{N+1}(\delta_{s,1} + \delta_{s,N}) \]
  - Existing trees grow in size due to new node.
  - New trees attributed to new node.
Size Distribution of Trees

- Manual iteration of recursion equation gives

\[ P_{s,2} = \left( \frac{1}{1.2} + \frac{1}{1.2} \right) \delta_{s,1} \]
\[ P_{s,3} = \left( \frac{1}{1.2} + \frac{1}{2.3} \right) (\delta_{s,1} + \delta_{s,2}) \]
\[ P_{s,4} = \left( \frac{1}{1.2} + \frac{1}{3.4} \right) (\delta_{s,1} + \delta_{s,3}) + \left( \frac{1}{2.3} + \frac{1}{2.3} \right) \delta_{s,2} \]
\[ P_{s,5} = \left( \frac{1}{1.2} + \frac{1}{4.5} \right) (\delta_{s,1} + \delta_{s,4}) + \left( \frac{1}{2.3} + \frac{1}{3.4} \right) (\delta_{s,2} + \delta_{s,3}) \]

- By induction: incredibly simple distribution

\[ P_{s,N} = \frac{1}{s(s + 1)} + \frac{1}{(N - s)(N + 1 - s)} \]

- Scaling form

\[ P_{s,N} \approx \frac{1}{N^2} \Psi \left( \frac{s}{N} \right) \quad \Psi(x) = \frac{1}{x^2} + \frac{1}{(1 - x)^2} \]
The Scaling Function

![Graph of the Scaling Function](image)
Iterative Removal of Nodes

• Remove randomly-selected nodes, one at a time

• Key observation: all trees in the random forest are statistically equivalent to a random tree!

• Treat the number of removed nodes as time $t$

• Let $F_{s,N}(t)$ be the average number of trees with size $s$ at time $t$

• A single conservation law

$$\sum_s s F_{s,N}(t) = N - t$$

• Recursion equation (represents removal of one node)

$$F_s(t+1) = F_s(t) - s f_s(t) + \sum_{l>s} l f_l(t) P_{s,l} \quad \text{with} \quad f_s(t) = \frac{F_s(t)}{\sum_s s F_s(t)}$$

- loss of trees
- loss rate = tree size
- gain of trees
- by fragmentation of larger ones
- normalized tree-size distribution
Rate Equation Approach

- Take the infinite tree-size limit: $N \to \infty$
- Treat time as continuous variable
- Recursion equation becomes a differential equation
  \[
  \frac{dF_s}{dt} = -sf_s + \sum_{l > s} l f_l P_{s,l}
  \]
- Use limiting size distribution, fraction of remaining nodes
  \[
  \phi_s(m) = \lim_{N \to \infty} \lim_{t \to \infty} \frac{F_{s,N}(t)}{\sum_s s F_{s,N}(t)}
  \]
  and
  \[
  m = \frac{N - t}{N}
  \]
- Problem reduces to the differential equation
  \[
  (\alpha - 1) \frac{d\phi_s}{d\alpha} = (1 - s)\phi_s + \sum_{l > s} \left[ \frac{l \phi_l}{s(s + 1)} + \frac{l \phi_l}{(l - s)(l + 1 - s)} \right]
  \]
  \[
  \alpha = 1 + \frac{1}{m}
  \]
The Size Distribution

- Miraculously, exact solution of the rate equation feasible

\[ \phi_s = \frac{1 - m}{m^2} \frac{\Gamma(s)\Gamma(\frac{1}{m})}{\Gamma(s + 1 + \frac{1}{m})} \]

- Power-law tail

\[ \phi_s \sim s^{-1 - \frac{1}{m}} \]

- Special case

\[ m = \frac{1}{2} \]

\[ \phi_s = \frac{2}{s(s + 1)(s + 2)} \]
Addition and Removal of Nodes

- **Addition**: Nodes are added at constant rate $r$
- **Removal**: Nodes are removed at constant rate $l$
- **Outcome**: random forest with growing number of nodes
- **Straightforward generalization of rate equation**

$$\frac{dF_s}{dt} = r \left[(s-1)f_{s-1} - sf_s\right] - sf_s(t) + \sum_{l>s} l f_l(t) P_{s,l}.$$ 

- **Normalized distribution of tree size decays exponentially**

$$\phi_s \sim s^{-r} \left(1 - e^{-r}\right)^s.$$
Summary

- Studied fragmentation of a random tree into a random forest
- Nodes removed one at a time
- Distribution of tree size becomes universal in the limit of infinitely many nodes
- Distribution of tree size has a power law tail
- Exponent governing the power law depends only on the fraction of remaining nodes
- Rate equation approach is a powerful analysis tool