A Directed Web Graph Model with Deletions

Aladdin Presentation

Roy Liu

ruoranl@andrew.cmu.edu

Carnegie Mellon University

Web Graphs

Why study web graphs?

- A fitting topic for the information age
- Model the web
- Gain insight into web structure by a random process

Web Graph Properties

Web graphs obey certain properties observed in real life.

- Since the growth of web graphs are random, they are nice for modeling large structures like the WWW, which have asymptotic properties.
- The standard random graph $G(n, p)$ is insufficient for explaining the power law distribution observed on $\mathsf{E}\left[D_k\right]$, the expectation of number of vertices of degree k over whatever random process that generated the graph. Namely, $\mathsf{E}\left[D_k\right] = Ck^{-\beta}$, for C, β constants.

History

Web graphs are a fairly new subject. Recent papers in the area propose models that achieve the power law distribution in the limit. Below are the two I am concerning myself with:

- *Directed Scale-Free Graphs*, Bollobás et al
- *Random Deletion in a Scale-Free Random Graph Process*, Cooper, Frieze, Vera

Proposing a Model

The model I propose will attempt to combine the ideas from the two papers, allowing for **deletions** in a **directed** graph. Let $G_t = (V_t, E_t)$ be a directed graph. To go from time step $t-1$ to t, I randomly select an action:

- With probability $\alpha - \alpha_1$, add m edges to the graph. $(*)$
- With probability α_1, add a vertex x_t and m edges incident to it. Each edge is chosen to be an in-edge or an out-edge of x_t with probabilities β and $1-\beta$, respectively. $(**)$
- With probability $1-\alpha-\alpha_0$, delete a vertex, chosen uniformly at random, and all edges incident to it.
- With probability α_0, delete m random edges, chosen independently, uniformly at random. If an edge is marked for deletion multiple times, simply delete it.

Just to Make Things Interesting

$(*)$ and $(**)$ are more complicated actions.

- In $(*)$, edge addition, we attach a directed edge (u, v) *preferentially*, meaning $\Pr[u \text{ chosen}] = \mathrm{d}_{\text{out}}(u)/|E_{t-1}|$ and $\Pr[v \text{ chosen}] = \mathrm{d}_{\text{in}}(v)/|E_{t-1}|$. u, v are chosen in this way independently for each of the m edges.
- In $(**)$, vertex addition, independently for each of the m edges, preferentially attach it to a vertex in V_{t-1} depending on whether it is an in-edge or an out-edge.
- After either action, we coalesce multiple edges into a single edge and delete self-loops.

Plan of Attack

I will follow the steps taken in [CFV], coming up with analogous proofs to theirs.

a. Write a recurrence relation for $\mathsf{E}\left[D_{i,j}(t)\right]$, the expectation of the random variable measuring the number of vertices of out-degree i and in-degree j in G_t.

b. Make simplifying assumptions about the recurrence and prove that these produce small deviations from the actual solution (i.e., show that coalescence produces small error terms).

c. Convert the recurrence into a differential equations problem using Laplace's method.

d. Solve the differential equation, obtain an answer, prove that this closely approximates the actual value.