A note on sampling scale-free graphs

K. M. Briggs keith.briggs@bt.com
BT Research, Adastral Park
Martlesham UK IP5 3RE

2007 November 19 16:47

In Proc. Nat. Acad. Sci. **102**, 4221-4224 (2005), Stumpf et al. have shown that when a fraction p of the nodes of a scale-free graph are sampled, the observed degree distribution is not that of a scale-free graph. This was achieved by computing the exact probability generating function (pgf), and then expanding it in powers of the parameter p, thus obtaining approximate formula. The purpose of this note is to show that exact formulas involving finite sums of polylogarithms are possible.

The setting is this: we fix a parameter $\gamma > 1$, and suppose that we have an infinite graph in which the fraction of nodes with degree k is $k^{-\gamma}/\zeta(\gamma)$. We now fix a parameter p, $0 \leqslant p \leqslant 1$, choose uniformly from the graph a fraction p of the nodes and keep the edges between these; that is, we form the induced subgraph. We delete nodes of degree 0. I claim that the degree distribution P^{**} of the subgraph is given by

$$P^{**}(k) = \left(\frac{p}{1-p}\right)^k \frac{\sum_{i=1}^{k} S_1(k, k-i+1) \log(\gamma+i-k-1, 1-p)}{k!\,(\zeta(\gamma)-\log(\gamma, 1-p))}. \qquad (1)$$

Here $\log(\alpha, x) \equiv \sum_{i=1}^{\infty} x^i/i^\alpha$ is the polylog function, ζ is the Riemann zeta function, and S_1 is the signed Stirling number of the first kind.

The proof starts with the degree pgf of the original graph:

$$G(s) = \frac{1}{\zeta(\gamma)} \sum_{k=0}^{\infty} k^{-\gamma} s^k$$

From this Stumpf et al. (equation [5]) show that the pgf of the sampled graph is

$$G^{**}(s) = \frac{G(1+ps-p) - G(1-p)}{1-G(1-p)}.$$

We see immediately from the definition of the polylog that this is

$$G^{**}(s) = \frac{\log(\gamma, 1+ps-p) - \log(\gamma, 1-p)}{\zeta(\gamma) - \log(\gamma, 1-p))}.$$

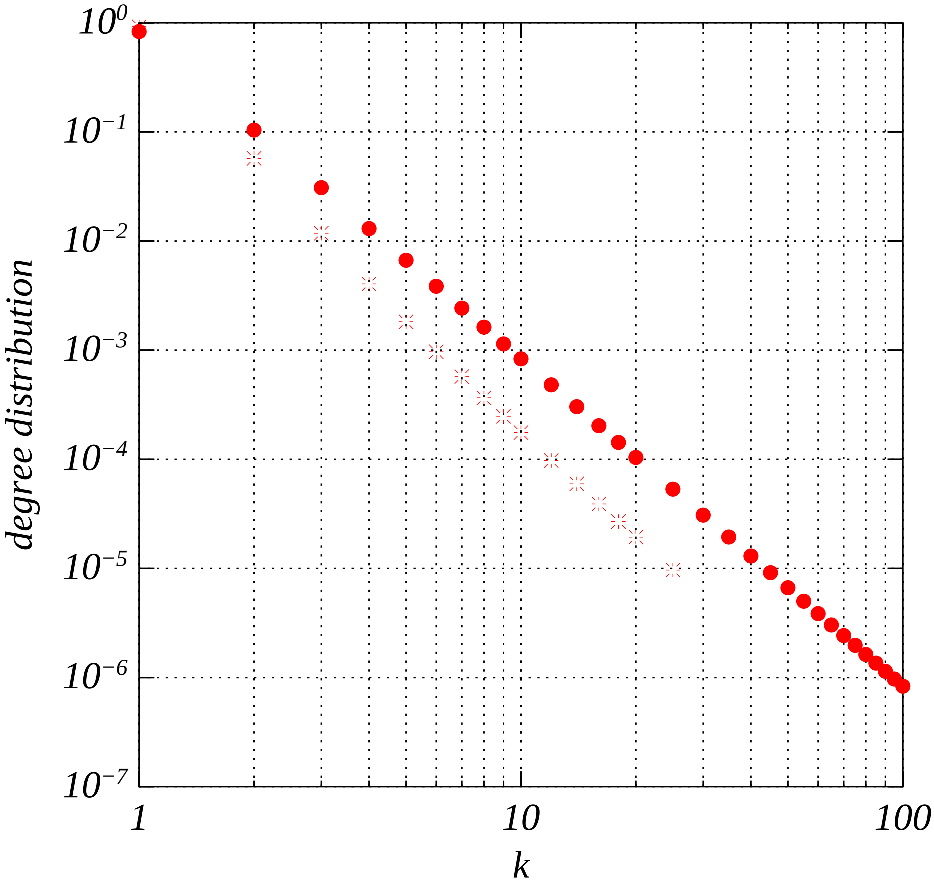

Figure 1: $\gamma = 3$: circles: degree distribution of original graph; stars: degree distribution of sampled graph for $p = 0.2$.

The proof of my equation 1 is now a simple induction on k, using $P^{**}(k) = (\mathrm{d}/\mathrm{d}s)^k\, G^{**}(s)/k!\big|_{s=0}$, and the recurrence for Stirling numbers of the first kind: $S_1(n+1, m) = S_1(n, m-1) - nS_1(n, m)$.

Note that this formula is bad for numerical evaluation, since the Stirling numbers become large and have alternating signs. It is better to derive series expansions useful for small p; for example, for $k = 3$ we have

$$P(1) = 1 + \frac{6\log(p) + \pi^2 - 3}{2\pi^2} p + \mathcal{O}(p^2).$$

The plot verifies that my exact formula agrees with Stumpf's approximations. It should be compared with their Figure 2, the $\gamma = 3$ case.