probability theory → stochastic processes ★★

Journal Citations

Suppose we wanted to understand the probability distribution of journal citations for something like the Physical Review. We might crudely expect that at zeroth order, the distribution of citations is scale invariant, because preferential attachment makes sense as a mechanism (popular papers are more likely to be cited again in the future). On the other hand, this doesn't perfectly match the data, so we should suspect this mechanism is not perfect. One proposal for a mechanism which appears to work is given by the following variant of preferential attachment. At each time step, we add a single new paper to the journal, with m citations. With probability $1 - q$, it cites an older paper according to linear preferential attachment: the probability it attaches to a paper with n citations is given by $n + \alpha$, for some constant α. However, with probability q, it instead attaches to a node chosen uniformly at random from the older nodes, irrespective of the number of edges it has. Assume that $mq \gtrsim 1$.

(a) Let $p_n(t)$ denote the probability that, after t articles have been added to the journal, a given article has n citations. You don't need to worry about the minor issue of what happens for t very small. Write down a recursive formula for $p_n(t+1)$ in terms of $p_m(t)$, for all m.

(b) As $t \to \infty$, show that these master equations admit an equilibrium distribution of the form

$$\mathrm{p}(n) \sim (n + n_0)^{-\gamma} .$$

and find expressions for n_0 and γ in terms of α, m and q. This is called the shifted power law distribution, and it matches the data surprisingly well!