Random walk with uniform restart

This random walk consists from simple random walk (when probability to transit from the node to one of his neighbors is equal) and uniform jumps (uniformly choosing one of the nodes).

If random walk is in the node i and d_i is the degree of the node i, then with probability $\frac{\alpha}{d_i+\alpha}$ the jump to one of n nodes will be performed or with probability $1-\frac{\alpha}{d_i+\alpha}=\frac{d_i}{d_i+\alpha}$ transition to one of i's neighbors.

So if j is the neighbor of i:

$$p_{ij}=\frac{d_i}{d_i+\alpha}\cdot\frac{1}{d_i}+\frac{\alpha}{d_i+\alpha}\cdot\frac{1}{n}=\frac{1+\frac{\alpha}{n}}{d_i+\alpha}$$

If j is not the neighbor of i:

$$p_{ij}=\frac{\alpha}{d_i+\alpha}\cdot\frac{1}{n}=\frac{\frac{\alpha}{n}}{d_i+\alpha}$$

So transition probabilities in such random walk are:

$$p_{ij}=\begin{cases}\dfrac{1+\frac{\alpha}{n}}{d_i+\alpha} & \text{if } i \text{ has a link to } j\\[2ex] \dfrac{\frac{\alpha}{n}}{d_i+\alpha} & \text{if } i \text{ does not have a link to } j\end{cases}$$

The stationary distribution is following:

$$\pi_i(\alpha)=\frac{d_i+\alpha}{\sum_j d_j+n\alpha}\qquad \forall i\in V$$

Sampling with random walk with uniform restart

If we take only samples after jump during such random walk, then all of them are distributed uniformly (due to uniform jump). As the random walk on the graph is time reversible, then if we take only samples before jump, then all of them should be distributed uniformly too. In this way, intuitively, performing one jump [uniform sampling] we get not one but too uniform samples. Let's calculate the gain that we have with such sampling comparing to the uniform sampling.

The average probability to jump [expected probability to jump]:

$$p_{jump}=\sum_i\left(\frac{d_i+\alpha}{\sum_j d_j+n\alpha}\cdot\frac{\alpha}{d_i+\alpha}\right)=\sum_i\frac{\alpha}{\sum_j d_j+n\alpha}=\frac{n\alpha}{\sum_j d_j+n\alpha}$$

Or if we denote $\bar{d}$ as average degree

$$\bar{d} = \sum_j \frac{d_j}{n}$$

Then

$$p_{jump} = \frac{\alpha}{\bar{d} + \alpha}$$

Let C be the cost of taking one uniform sample [or the cost of jump]. The cost of transition to neighbor is 1.

Then the cost of one step is

$$\frac{\alpha}{\bar{d} + \alpha} \cdot C + \left(1 - \frac{\alpha}{\bar{d} + \alpha}\right) \cdot 1$$

So as it was said one jump gives us 2 uniform samples. But due to the consecutive jumps the gain will be actually less. Let's count it exactly.

Let's divide nodes that we selected as samples (T is number of selected samples) into the nodes that end sequence of jumps (I) [it means that on the next step transition to the neighbor is performed], the nodes that start sequence of jumps (II) and the rest of them (III). [T = I + II + III = 2*I + III]

Probability to jump is

$$\frac{\alpha}{\bar{d} + \alpha}$$

Probability to be in one particular node after jump is

$$\frac{\alpha}{\bar{d} + \alpha} \cdot \frac{1}{n}$$

Probability to jump and then transit is

$$\frac{\alpha}{\bar{d} + \alpha} \cdot \frac{1}{n} \cdot \sum_j \frac{d_j}{d_j + \alpha}$$

Probability to jump and then jump again is

$$\frac{\alpha}{\bar{d} + \alpha} \cdot \frac{1}{n} \cdot \sum_j \frac{\alpha}{d_j + \alpha}$$

The expected number of samples we get for one step is:

$$2 \cdot \frac{\alpha}{\bar{d} + \alpha} \cdot \frac{1}{n} \cdot \sum_j \frac{d_j}{d_j + \alpha} + \frac{\alpha}{\bar{d} + \alpha} \cdot \frac{1}{n} \cdot \sum_j \frac{\alpha}{d_j + \alpha}$$

So the cost of one sample is:

$$\frac{\frac{\alpha}{\bar{d}+\alpha}\cdot C+\left(1-\frac{\alpha}{\bar{d}+\alpha}\right)\cdot 1}{2\cdot\frac{\alpha}{\bar{d}+\alpha}\cdot\frac{1}{n}\cdot\sum_j\frac{d_j}{d_j+\alpha}+\frac{\alpha}{\bar{d}+\alpha}\cdot\frac{1}{n}\cdot\sum_j\frac{\alpha}{d_j+\alpha}}$$

If we assume the next approximation

$$\frac{1}{n}\cdot\sum_j\frac{d_j}{d_j+\alpha}=\frac{\bar{d}}{\bar{d}+\alpha}$$

$$\frac{1}{n}\cdot\sum_j\frac{\alpha}{d_j+\alpha}=\frac{\alpha}{\bar{d}+\alpha}$$

Then the cost of one sample of such random walk is (*)

$$\frac{\frac{\alpha}{\bar{d}+\alpha}\cdot C+\left(1-\frac{\alpha}{\bar{d}+\alpha}\right)\cdot 1}{2\cdot\frac{\alpha}{\bar{d}+\alpha}\cdot\frac{1}{n}\cdot\sum_j\frac{d_j}{d_j+\alpha}+\frac{\alpha}{\bar{d}+\alpha}\cdot\frac{1}{n}\cdot\sum_j\frac{\alpha}{d_j+\alpha}}=\frac{\frac{\alpha C+\bar{d}}{\bar{d}+\alpha}}{2\cdot\frac{\alpha\bar{d}}{\bar{d}+\alpha}+\left(\frac{\alpha}{\bar{d}+\alpha}\right)^2}=\frac{(\alpha C+\bar{d})\cdot(\bar{d}+\alpha)}{2a\bar{d}+a^2}$$

Minimize α

Let’s find α that minimizes the cost of one sample

$$f(\alpha)=\frac{(\alpha C+\bar{d})\cdot(\bar{d}+\alpha)}{2a\bar{d}+a^2}=\frac{C\alpha\bar{d}+C\alpha^2+\bar{d}^2+\alpha\bar{d}}{2a\bar{d}+a^2}=\frac{C\alpha^2+\alpha\bar{d}(1+C)+\bar{d}^2}{2a\bar{d}+a^2}$$

$$f'(\alpha)=\frac{\left(2\alpha C+\bar{d}(1+C)\right)\cdot(2a\bar{d}+a^2)-(2\bar{d}+2\alpha)\cdot(C\alpha^2+\alpha\bar{d}(1+C)+\bar{d}^2)}{(2a\bar{d}+a^2)^2}$$
$$=\frac{\alpha^2(4C\bar{d}+\bar{d}+C\bar{d}-2C\bar{d}-2\bar{d}-2C\bar{d})+\alpha(2\bar{d}^2+2C\bar{d}-2\bar{d}^2-2C\bar{d}-2\bar{d}^2)-2\bar{d}^3}{(2a\bar{d}+a^2)^2}$$
$$=\frac{\alpha^2(C\bar{d}-\bar{d})-2\bar{d}^2\alpha-2\bar{d}^3}{(2a\bar{d}+a^2)^2}$$

Critical values are

$$\alpha_{12}=\frac{\bar{d}\pm\bar{d}\sqrt{2C-1}}{C-1}$$

When α is following the expression of the sample cost is minimized

$$\alpha=\frac{\bar{d}+\bar{d}\sqrt{2C-1}}{C-1}$$

If we put this α in (*) then the cost for one sample is

$$\frac{\sqrt{2C-1}+C}{2}$$

Now knowing that the cost of one sample with uniform sampling is C we can compare these two methods

$$\delta = \frac{cost\ of\ one\ sample\ with\ RW\ with\ restarts}{cost\ of\ one\ sample\ with\ uniform\ sampling} = \frac{\frac{\sqrt{2C-1}+C}{2}}{C} = \frac{\sqrt{2C-1}+C}{2C}$$
$$= \frac{1}{2} + \frac{\sqrt{2C-1}}{2C}$$

So with uniform restarts we can't have gain 2 times more and bigger. But as C grows the gain also grows.

Jensen's inequality

As

$$\frac{1}{n} \cdot \sum_j \frac{d_j}{d_j+\alpha} = 1 - \frac{1}{n} \cdot \sum_j \frac{\alpha}{d_j+\alpha}$$

We can rewrite the expression for the cost of one sample

$$\frac{\frac{\alpha}{\bar{d}+\alpha} \cdot C + \left(1 - \frac{\alpha}{\bar{d}+\alpha}\right) \cdot 1}{2 \cdot \frac{\alpha}{\bar{d}+\alpha} \cdot \frac{1}{n} \cdot \sum_j \frac{d_j}{d_j+\alpha} + \frac{\alpha}{\bar{d}+\alpha} \cdot \frac{1}{n} \cdot \sum_j \frac{\alpha}{d_j+\alpha}} = \frac{\frac{\alpha}{\bar{d}+\alpha} \cdot C + \left(1 - \frac{\alpha}{\bar{d}+\alpha}\right) \cdot 1}{\frac{\alpha}{\bar{d}+\alpha}\left(2\frac{1}{n} \cdot \sum_j \frac{d_j}{d_j+\alpha} + \frac{1}{n} \cdot \sum_j \frac{\alpha}{d_j+\alpha}\right)}$$
$$= \frac{\frac{\alpha}{\bar{d}+\alpha} \cdot C + \left(1 - \frac{\alpha}{\bar{d}+\alpha}\right) \cdot 1}{\frac{\alpha}{\bar{d}+\alpha}\left(2\left(1 - \frac{1}{n} \cdot \sum_j \frac{\alpha}{d_j+\alpha}\right) + \frac{1}{n} \cdot \sum_j \frac{\alpha}{d_j+\alpha}\right)} = \frac{\frac{\alpha}{\bar{d}+\alpha} \cdot C + \left(1 - \frac{\alpha}{\bar{d}+\alpha}\right) \cdot 1}{\frac{\alpha}{\bar{d}+\alpha}\left(2 - \frac{1}{n} \cdot \sum_j \frac{\alpha}{d_j+\alpha}\right)}$$

As $f(d_i)$ is convex

$$f(d_i) = \frac{\alpha}{d_i+\alpha}$$

And knowing Jensen's inequality

$$f\left(\sum_{i=1}^{n} q_i x_i\right) \le \sum_{i=1}^{n} q_i f(x_i)$$

$$\frac{\alpha}{\bar{d}+\alpha} = \frac{\alpha}{\sum_i \frac{1}{n} d_i + \alpha} \le \sum_i \frac{1}{n} \cdot \frac{\alpha}{d_i+\alpha}$$

We get

$$\frac{\frac{\alpha}{\bar{d}+\alpha}\cdot C + \left(1-\frac{\alpha}{\bar{d}+\alpha}\right)\cdot 1}{\frac{\alpha}{\bar{d}+\alpha}\left(2-\frac{1}{n}\cdot\sum_j\frac{\alpha}{d_j+\alpha}\right)} \geq \frac{\frac{\alpha}{\bar{d}+\alpha}\cdot C + \left(1-\frac{\alpha}{\bar{d}+\alpha}\right)\cdot 1}{\frac{\alpha}{\bar{d}+\alpha}\left(2-\frac{\alpha}{\bar{d}+\alpha}\right)}$$

So in fact having counted α with following approximation

$$\frac{1}{n}\cdot\sum_j\frac{\alpha}{d_j+\alpha} = \frac{\alpha}{\bar{d}+\alpha}$$

shows the gain less then theoretical gain.