By forward search, $\pi(s,v) = \pi^f(s,v) + \sum_{u\in V} r^f(s,u)\pi(u,v)$, it follows that $\sum_{u\in V} r^f(s,u)\pi(u,v) \le \pi(s,v)$, and thus

$$E[X_i^2] \le 2r^f_{sum}\left(r^b_{max}\right)^2 \sum_{v\in V}\left(\frac{d_{in}(v)}{\pi(s,v)}\cdot \pi(s,v)\right) = 2r^f_{sum}\left(r^b_{max}\right)^2 \sum_{v\in V} d_{in}(v) = 2r^f_{sum}\left(r^b_{max}\right)^2\cdot m.$$

Therefore, we have $E[Y] \le 2mr^f_{sum}\left(r^b_{max}\right)^2$. Let $z = \frac{8\left(r^f_{sum}\right)^2\log n}{n_r}$, since each $X_i^2 \le \left(r^f_{sum}\right)^2$, by Chernoff inequality, the probability that $Y \le 2E[Y] + z \le 4mr^f_{sum}\left(r^b_{max}\right)^2 + \frac{8\left(r^f_{sum}\right)^2\log n}{n_r}$ is at most

$$exp\left(-n_r\left(E[Y]+z\right)^2\Big/\left(r^f_{sum}\right)^2\cdot\left(\frac{8}{3}E[Y]+\frac{2}{3}z\right)\right) \le exp\left(-\frac{3}{8}n_r\cdot\left(E[Y]+z\right)\Big/\left(r^f_{sum}\right)^2\right) \le exp\left(-\frac{3n_r\log n}{n_r}\right) \le \frac{1}{n^3},$$

and the lemma follows. □

Proof of Lemma 4.8. By Lemma 4.7, we need to show that if at some iteration, the number of random walks $n_r = \frac{cmr^f_{sum}\left(r^b_{max}\right)^2}{gap_\rho^2}\log n$ for constant c, we have $\beta(s,t) \le gap_\rho/4$ for any $t \in C$. Recall that

$$\beta(s,t) = \sqrt{\frac{2\bar{\sigma}^2(s,t)\ln\left(3n^3\log^2 n_r\right)}{n_r}} + \frac{3r^f_{sum}\ln\left(3n^3\log^2 n_r\right)}{n_r}. \quad (3)$$

By the assumption that $r^b_{max} \ge \sqrt{\frac{gap_\rho}{m}}$, we have

$$n_r \ge \frac{cmr^f_{sum}\left(r^b_{max}\right)^2}{gap_\rho^2}\log n \ge \frac{cmr^f_{sum}\cdot\frac{gap_\rho}{m}}{gap_\rho^2}\log n \ge \frac{cr^f_{sum}\log n}{gap_\rho},$$

and thus we can bound the the second term of equation (3) as

$$\frac{6r^f_{sum}\ln\left(3n^3\log^2 n_r\right)}{n_r} \le gap_\rho\frac{6\log\left(3n^3\log^2 n_r\right)}{c\log n} \le gap_\rho/8. \quad (4)$$

For c sufficiently large. The last inequality is due to $n \ge \log\frac{m}{gap_\rho}$.

To bound the first term of equation (3), we observe that by Lemma 4.1, the empirical variance $\bar{\sigma}^2(s,t) \le 2mr^f_{sum}\left(r^b_{max}\right)^2 + \frac{8\left(r^f_{sum}\right)^2\log n}{n_r}$ with probability at least $1-1/n^3$. If $2mr^f_{sum}\left(r^b_{max}\right)^2 \le \frac{8\left(r^f_{sum}\right)^2\log n}{n_r}$, we have $\bar{\sigma}^2(s,t) \le \frac{16\left(r^f_{sum}\right)^2\log n}{n_r}$ and the first term of equation (3) can be bounded by

$$\sqrt{\frac{2\bar{\sigma}^2(s,t)\ln 3n^3\log^2 n_r}{n_r}} \le \sqrt{\frac{\frac{16\left(r^f_{sum}\right)^2\log n}{n_r}\ln 3n^3\log^2 n_r}{n_r}} \le \frac{1}{8}gap_\rho.$$

The last inequality is due to inequality (4). On the other hand, if $2mr^f_{sum}\left(r^b_{max}\right)^2 \ge \frac{8\left(r^f_{sum}\right)^2\log n}{n_r}$, we have $\bar{\sigma}^2(s,t) \le 4mr^f_{sum}\left(r^b_{max}\right)^2$, and thus

$$\sqrt{\frac{2\bar{\sigma}^2(s,t)\ln\left(3n^3\log^2 n_r\right)}{n_r}} \le \sqrt{\frac{8mr^f_{sum}\left(r^b_{max}\right)^2\ln\left(3n^3\log^2 n_r\right)}{n_r}}$$

$$= \sqrt{\frac{8mr^f_{sum}\left(r^b_{max}\right)^2\ln 3n^3\log^2 n_r}{\frac{cmr^f_{sum}\left(r^b_{max}\right)^2}{gap_\rho^2}\log n}} \le gap_\rho\sqrt{\frac{24\log n\log n_r}{c\log n}} \le \frac{gap_\rho}{8}$$

for c sufficiently large. Thus we have $\beta(s,t) \le gap_\rho/4$ for c sufficiently large, and the Lemma follows. □

C.8 Proof of Theorem 4.9 and Theorem 4.10

Proof. Recall that at the i-th iteration, we have $r^f_{max} = \frac{1}{2^i m}$ and $r^b_{max} = 1/2^i\sqrt{m}$ and walk number $n_r = c2^i n\log n/|C|$. The total query cost is $cost(forward) + cost(backward) + cost(walk)$, which can be bounded by

$$O\left(\frac{1}{r^f_{max}}\right) + O\left(\frac{\sqrt{m}}{r^b_{max}}\right) + n_r|C| = O\left(2^i m + c2^i n\log n\right). \quad (5)$$

For worst-case graph, if we can prove that TopPPR stops before iteration when $2^i = 1/\sqrt{gap_\rho}$, then the total query cost is bounded by $O\left(\frac{m+n\log n}{\sqrt{gap_\rho}}\right)$ and Theorem 4.9 follows. More precisely, we have $r^f_{max} = \sqrt{gap_\rho}/m$, $r^b_{max} = \sqrt{gap_\rho/m}$. By Lemma 4.8 and $r^b_{max} = \sqrt{gap_\rho/m}$, we have $n_r \ge \frac{c2^i n\log n}{|C|} \ge \frac{cn\log n}{\sqrt{gap_\rho}\cdot|C|} \ge \frac{c\log n}{\sqrt{gap_\rho}}$. The last inequality uses the fact that $|C| \le n$. Therefore, we have $r^f_{max}\left(r^b_{max}\right)^2 = \frac{\left(\sqrt{gap_\rho}\right)^3}{m^2}$ and $n_r \ge \frac{c\log n}{\sqrt{gap_\rho}} \ge \frac{cm^2\frac{\left(\sqrt{gap_\rho}\right)^3}{m^2}}{gap_\rho^2}\log n \ge \frac{cmr^f_{sum}\left(r^b_{max}\right)^2}{gap_\rho^2}\log n$. This proves that the number of walks $n_r = \Omega\left(\frac{cmr^f_{sum}\left(r^b_{max}\right)^2}{gap_\rho^2}\log n\right)$, and by Lemma 4.7, the algorithm stops at this iteration.

Similarly, for power law graph, we can prove that TopPPR stops before iteration when $2^i = \frac{1}{\sqrt{gap_\rho}}\cdot\frac{k^{\frac{1}{4}}}{n^{\frac{1}{4}}}$, under Assumption 1. Thus Theorem 4.10 follows from equation 5. □

C.9 Proof of Lemma 5.1

Proof. We define two types of random walks. A non-stop random walk from s is a traversal of G that starts from s and, at each step, proceeds to a randomly selected out-neighbor of the current node. An $(1-\alpha)$-walk is a random walk in which at each step the random walk proceeds to the next node with probability $1-\alpha$ and terminates with probability α.

Let $p_i(s,t)$ denote the probability that a non-stop random walk from s visits t in the i step. We have $\pi(s,t) = \sum_{i=0}^{\infty}\alpha(1-\alpha)^i p_i(s,t)$. To see this, note that $\alpha(1-\alpha)^i p_i(s,t)$ is the probability that a $(1-\alpha)$-walk starts at s and terminates at t using i steps. Summing i over 0 to ∞ and the equation follows. For $(\sqrt{1-\alpha})$-walk, recall that $(\sqrt{1-\alpha})^i p_i(s,t)$ is the probability that the $\sqrt{1-\alpha}$-walk visits t at the i-th step, and $(\sqrt{1-\alpha})^i$ is the value added to $\hat{\pi}(s,t)$, we have

$$E[\hat{\pi}(s,t)] = \sum_{i=0}^{\infty}\alpha(\sqrt{1-\alpha})^i\cdot(\sqrt{1-\alpha})^i p_i(s,t) = \pi(s,t),$$

and the Lemma follows. □