Let $\mathbf{p} = (p_u)_{u \in \mathbf{X}}$ denote the PPR vector computed over $G_{n,r}$ (where for ease of reading we suppress dependence on the hyperparameter α and seed node v.)

Lemma 1. *Consider running Algorithm 1 with any $r < \sigma$ and*

$$\frac{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{10} \leq \alpha \leq \frac{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{9}. \tag{A.1}$$

There exists a good set $\mathcal{C}_\sigma[\mathbf{X}]^g \subseteq \mathcal{C}_\sigma[\mathbf{X}]$ with $\mathrm{vol}(\mathcal{C}_\sigma[\mathbf{X}]^g) \geq \mathrm{vol}(\mathcal{C}_\sigma[\mathbf{X}])/2$ such that the following statements hold for all $v \in \mathcal{C}_\sigma[\mathbf{X}]^g$:

- *For all $u \in \mathcal{C}[\mathbf{X}]$,*

$$p_u \geq \frac{4}{5}\widetilde{\pi}_{n,r}(u) - \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])\widetilde{D}_{\min}}$$

- *For all $u' \in \mathcal{C}'_\sigma[\mathbf{X}]$,*

$$p_{u'} \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])\widetilde{D}_{\min}}$$

Proof. For $S \subseteq X_1^n$, let I_S be a diagonal matrix where $I_{jj} = 1$ if $X_j \in S$ and 0 otherwise. Let D_S be the corresponding degree matrix for the subgraph induced by S. D_S is a diagonal matrix where for $X_j \in S$, $(D_S)_{jj} = \sum_{i: X_i \in S} A_{ij}$, and if $X_j \notin S$ then $(D_S)_{jj} = 0$. Then, define *leakage* and *soakage* vectors

$$\begin{aligned}
\ell_t &:= e_v(WI_{\mathcal{C}_\sigma[\mathbf{X}]})^t(I - D^{-1}D_{\mathcal{C}_\sigma[\mathbf{X}]}) \\
\ell &:= \sum_{t=0}^{\infty}(1-\alpha)^t \ell_t \\
s_t &:= e_v(WI_{\mathcal{C}_\sigma[\mathbf{X}]})^t(WI_{G/\mathcal{C}_\sigma[\mathbf{X}]}) \\
s &:= \sum_{t=0}^{\infty}(1-\alpha)^t s_t
\end{aligned}$$

Roughly, the proof will unfold in four steps. The first two will result in the lower bound of (**??**), while the latter two will imply the upper bound in (**??**).

1. For $u \in \mathcal{C}'[\mathbf{X}]$, use the results of [1] to lower bound $p_v(u) \geq 4/5\widetilde{\pi}(u) - \widetilde{p}_\ell(u)$, where $\widetilde{p}$ is the PPR random walk over the subgraph induced by $\mathcal{C}_\sigma[\mathbf{X}]$, and ℓ has bounded norm $||\ell||_1 \leq 2\frac{\Phi^{btw}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha}$.
2. Since $r < \sigma$, there are no edges between u and $G/\mathcal{C}_\sigma[\mathbf{X}]$. Therefore, the page-rank vector $\widetilde{p}_\ell$ will not assign more than $||\ell||_1/d_{\min}(\mathcal{C}_\sigma[\mathbf{X}])$ probability mass to any vertex in $\mathcal{C}'[\mathbf{X}]$. This will conclude our proof of (**??**).

3. For vertices $u' \in G/\mathcal{C}_\sigma[\mathbf{X}]$, we can upper bound $p_v(u) \leq p_s(u')$. In particular, this hold for all $u' \in \mathcal{C}'[\mathbf{X}]pr$.
4. Since $r < \sigma$, there are no edges between u' and $G/\mathcal{C}'[\mathbf{X}]prthick$. Therefore, the page-rank vector p_s will assign no more than $||s||_1/d_{\min}(\mathcal{C}_\sigma[\mathbf{X}])$ probability mass to any vertex in $\mathcal{C}'[\mathbf{X}]$. Additionally, s has bounded norm $||s||_1 \leq ||\ell||_1$. This will conclude our proof of (**??**), and hence Proposition **??**.

Step 1 We will begin by restating the results of [1].

Denote by $\widetilde{p}$ the PageRank vector computed only over the subgraph induced by $\mathcal{C}_\sigma[\mathbf{X}]$.

$$\begin{aligned}\widetilde{p}_v &= \alpha e_v + (1-\alpha)\widetilde{p}_v\widetilde{W} &\text{(A.2)}\\ &= \alpha\sum_{t=0}^{\infty}(1-\alpha)^t\left(e_v\widetilde{W}^t\right) &\text{(A.3)}\end{aligned}$$

and correspondingly for the leakage vector

$$\widetilde{p}_\ell = \alpha\ell + (1-\alpha)\widetilde{p}_\ell\widetilde{W}.$$

From Lemma 3.1 of [1] we have for all $u \in \mathcal{C}_\sigma[\mathbf{X}]$

$$p_v(u) \geq \widetilde{p}_v(u) - \widetilde{p}_\ell(u).$$

with $||\ell||_1 \leq \frac{2\Phi^{btw}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha}$. Moreover if, as we have specified, $\alpha \leq \frac{1}{9T_\infty(\mathcal{C}_\sigma[\mathbf{X}])}$, Lemma 3.2 of [1] yields a lower bound on $\widetilde{p}$

$$\widetilde{p}_v(u) \geq \frac{4}{5}\widetilde{\pi}(u). \tag{A.4}$$

Step 2 We turn to upper bounding $\widetilde{p}_\ell(u)$. We have

$$
\begin{aligned}
\widetilde{p}_\ell(u) &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\ell \widetilde{W}^t\right)[u] \\
&= \|\ell\|_1 \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \widetilde{W}^t\right)[u] \\
&\overset{\text{(i)}}{=} \|\ell\|_1 \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \widetilde{W}^t\right)[u] \\
&\overset{\text{(ii)}}{\leq} \|\ell\|_1 \frac{1}{d_{\min}(\mathcal{C}_\sigma[\mathbf{X}])}
\end{aligned} \tag{A.5}
$$

where (i) follows from the fact that since $r < \sigma$, $E(\mathcal{C}'[\mathbf{X}], G/\mathcal{C}_\sigma[\mathbf{X}]) = 0$ and therefore $\ell(u) = 0$. To see (ii), let $q = \frac{\ell}{\|\ell\|_1}\widetilde{W}^{t-1}$, and then

$$
\begin{aligned}
\left(\frac{\ell}{\|\ell\|_1} \widetilde{W}^t\right)[u] &= (qW)[u] \\
&\leq \|q\|_1 \|W(\cdot, u)\|_\infty \\
&\overset{\text{(iii)}}{\leq} \frac{1}{d_{\min}(C_n^\sigma)}.
\end{aligned}
$$

where (iii) comes from the fact that $\|u - v\| \leq r$ means $v \in \mathcal{C}_\sigma[\mathbf{X}]$. Combining (A.5) with (A.4), and since $||\ell||_1 \leq 2\frac{\Phi^{btw}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha}$, we have

$$
p_v(u) \geq \frac{4}{5}\widetilde{\pi}(u) - 2\frac{9\Phi^{btw}(\mathcal{C}_\sigma[\mathbf{X}])}{d_{\min}(\mathcal{C}_\sigma[\mathbf{X}])\alpha}.
$$

Step 3 To get the corresponding upper bound on $p_v(u')$, we will use the soakage vectors s and s_t. We will first argue that s is a worse starting distribution – meaning it puts uniformly more mass outside the cluster – than simply starting at v.

Lemma 2. *For all $u' \notin \mathcal{C}_\sigma[\mathbf{X}]$,*

$$
p_v(u') \leq p_s(u'). \tag{A.6}
$$

The proof of Lemma 2 is left to the supplement. It follows largely the same steps as Lemma 3.1 of [1], except over $G/\mathcal{C}_\sigma[\mathbf{X}]$ rather than $\mathcal{C}_\sigma[\mathbf{X}]$.

Step 4 Just as we upper bounded the probability mass $\widetilde{p}_\ell$ could assign to any one vertex, we can upper bound

$$\begin{aligned}
\widetilde{p}_s(u) &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(s\widetilde{W}^t \right) [u] \\
&= \|s\|_1 \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1} \widetilde{W}^t \right) [u] \\
&= \|s\|_1 \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1} \widetilde{W}^t \right) [u] \\
&\leq \|s\|_1 \frac{1}{d_{\min}(\mathcal{C}_\sigma[\mathbf{X}])}. \qquad (A.7)
\end{aligned}$$

Finally, by the definition of $s_t = e_v(WI_{\mathcal{C}_\sigma[\mathbf{X}]})^t(WI_{G/\mathcal{C}_\sigma[\mathbf{X}]})$, letting $q_t = e_v(WI_{\mathcal{C}_\sigma[\mathbf{X}]})^t$ for ease of notation, we have

$$\begin{aligned}
\|s_t\|_1 &= \|q_t(WI_{G/\mathcal{C}_\sigma[\mathbf{X}]})\|_1 \\
&= \sum_{u' \in G} \sum_{u \in G} q_t(u)(WI_{G/\mathcal{C}_\sigma[\mathbf{X}]})[u, u'] \\
&= \sum_{u' \in G/\mathcal{C}_\sigma[\mathbf{X}]} \sum_{u \in \mathcal{C}_\sigma[\mathbf{X}]} \frac{q(u)}{d(u)} I(e_{u,u'} \in G) \\
&= \sum_{u \in \mathcal{C}_\sigma[\mathbf{X}]} \frac{q(u)(d(u) - d_{\mathcal{C}_\sigma[\mathbf{X}]}(u))}{d(u)} \\
&= \|q_t(I - D^{-1}D_{\mathcal{C}_\sigma[\mathbf{X}]})\|_1 = \|\ell_t\|_1.
\end{aligned}$$

and as a result $\|s\|_1 = \|\ell\|_1$. Combining with $\|\ell\|_1 \leq 2\frac{\Phi^{btw}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha}$ and (A.7) yields the desired upper bound.

□

Lemma 3. *Let $\mathcal{C}_\sigma$ satisfy the conditions of Theorem 4. For $r < \sigma$, the following statements hold with probability tending to one as $n \to \infty$:*

$$\begin{aligned}
D_{\min}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r}) &\geq \frac{1}{2}\nu_d r^d \lambda_\sigma \\
D_{\max}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r}) &\leq 2\nu_d r^d \Lambda_\sigma \\
\widetilde{\mathrm{vol}}_{n,r}(\widetilde{G}_{n,r}) &\leq 2\nu(\mathcal{C}_\sigma)\Lambda_\sigma
\end{aligned}$$

where $D_{\min}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r})$ is the minimum degree of any vertex $v \in \mathcal{C}_\sigma[X]$ in the subgraph $\widetilde{G}_{n,r}$, and analogously for $D_{\max}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r})$.

The statement follows immediately from Lemma **??**.

.1 Proof of Theorem 4

We note that by Theorems 1 and 2,

$$\kappa_2(\mathcal{C}) \geq \frac{\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}.$$

As a result Lemma 1 implies

$$\begin{aligned} p_u &\geq \frac{4}{5}\widetilde{\pi}_{n,r}(u) - \frac{18\kappa_2(\mathcal{C})}{\widetilde{D}_{\min}} && (u \in \mathcal{C}[\mathbf{X}]) \\ p_{u'} &\leq \frac{18\kappa_2(\mathcal{C})}{\widetilde{D}_{\min}} && (u' \in \mathcal{C}'[\mathbf{X}]) \end{aligned} \tag{A.8}$$

We then have

$$\begin{aligned} \widetilde{\pi}_{n,r}(u) &\geq \frac{D_{\min}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r})}{\widetilde{\mathrm{vol}}_{n,r}(\widetilde{G}_{n,r})} \\ &\geq \frac{D_{\min}(\mathcal{C}_\sigma[X]; \widetilde{G}_{n,r})}{\widetilde{n}\widetilde{D}_{\max}} \end{aligned}$$

and application of Lemma 3 yields

$$\widetilde{\pi}_{n,r}(u) \geq 8\frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2} \tag{A.9}$$

and

$$\frac{1}{\widetilde{D}_{\min}} \leq \frac{2}{\nu_d r^d \lambda_\sigma} \tag{A.10}$$

with probability tending to 1 as $n \to \infty$, for all $u \in \mathcal{C}[\mathbf{X}]$.

Combining (A.8), (A.9) and (A.10), along with the requirement on $\kappa_2(\mathcal{C})$ given by (17), we have

$$\begin{aligned} p_u &\geq 3/5\frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2} \\ p_{u'} &\leq 1/5\frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2} \end{aligned}$$

for any $u \in \mathcal{C}$, $u' \in \mathcal{C}'$. As a result, if $\pi_0 \in (2/5, 3/5) \cdot \frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2}$, as $n \to \infty$ with probability tending to one any sweep cut of the form of (6), including the output set $\widehat{C}$, will successfully recover $\mathcal{C}$ in the sense of (9).

References

[1] Zeyuan Allen Zhu, Silvio Lattanzi, and Vahab S Mirrokni. A local algorithm for finding well-connected clusters. In *ICML (3)*, pages 396–404, 2013.