Let $\mathbf{p} = (p_u)_{u \in \mathbf{X}}$ denote the PPR vector computed over $G_{n,r}$ (where for ease of

Recalling that $\widetilde{\pi}_{n,r}$ is the stationary distribution over $\widetilde{G}_{n,r}$, we write $\widetilde{\pi}_{n,r}(u)$ to denote the stationary distribution evaluated at $u \in \mathcal{C}_\sigma[X]$.

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

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

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

$$p_{u'} \leq \frac{2\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\Psi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])\widetilde{D}_{\min}} \tag{A.3}$$

Proof. We will write $\mathbf{W}_n = \mathbf{D}_n \mathbf{A}_n^{-1}$ for the transition probability matrix over $G_{n,r}$, and let $\widetilde{\mathbf{D}}_n$ and $\widetilde{\mathbf{W}}_n$ be the degree and random walk matrices for the subgraph $\widetilde{G}_{n,r}$.

We introduce *leakage* and *soakage* vectors, defined by

$$\begin{aligned}
\ell_t &:= e_v(\mathbf{W}_n\widetilde{\mathbf{I}}_n)^t(\mathbf{I}_n - \mathbf{D}_n^{-1}\widetilde{\mathbf{D}}_n) \\
\ell &:= \sum_{t=0}^{\infty}(1-\alpha)^t \ell_t \\
s_t &:= e_v(\mathbf{W}_n\widetilde{\mathbf{I}}_n)^t(\mathbf{W}_n\widetilde{\mathbf{I}}_n^c) \\
s &:= \sum_{t=0}^{\infty}(1-\alpha)^t s_t
\end{aligned}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\widetilde{\mathbf{I}}_n$ is a diagonal matrix with $(\widetilde{\mathbf{I}}_n)_{uu} = 1$ if $u \in \mathcal{C}_\sigma[\mathbf{X}]$ and $\widetilde{\mathbf{I}}_n^c = \mathbf{I}_n - \widetilde{\mathbf{I}}_n$.

Roughly, the proof will unfold in four steps. The first two will result in the lower bound of (A.2), while the latter two will imply the upper bound in (A.3).

1. For $u \in \mathcal{C}'[\mathbf{X}]$, use the results of [1] to produce the lower bound

   $$\mathbf{p}(u) \geq 4/5\widetilde{\pi}_{n,r}(u) - \widetilde{\mathbf{p}}_\ell(u)$$

   where

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

is the PPR random walk over $\widetilde{G}_{n,r}$, and ℓ has bounded norm $||\ell||_1 \leq 2\frac{\Phi_{n,r}(\mathcal{C}_\sigma[\mathbf{X}])}{\alpha}$.

2. Since $r < \sigma$, for any $u \in \mathcal{C}[\mathbf{X}]$ there are no edges between u and $G_{n,r}/\mathcal{C}_\sigma[\mathbf{X}]$. Therefore, the page-rank vector $\widetilde{\mathbf{p}}_\ell$ 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 observation will conclude our proof of (A.2).

3. For vertices $u' \in G_{n,r}/\mathcal{C}_\sigma[\mathbf{X}]$, we can upper bound $p_v(u) \leq p_s(u')$. In particular, this hold for all $u' \in \mathcal{C}'[\mathbf{X}]$.

4. Since $r < \sigma$, there are no edges between u' and $G/\mathcal{C}'[\mathbf{X}]$. 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 (A.3), and hence Lemma 1.

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

For seed node v, we write

$$\widetilde{\mathbf{p}}_v = \alpha e_v + (1-\alpha)\widetilde{\mathbf{p}}_v \widetilde{\mathbf{W}}_n \tag{A.4}$$

$$= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(e_v \widetilde{\mathbf{W}}_n^t\right) \tag{A.5}$$

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

$$p_u \geq \widetilde{\mathbf{p}}_v(u) - \widetilde{\mathbf{p}}_\ell(u)$$

$$||\ell||_1 \leq \frac{2\widetilde{\Phi}_{n,r}}{\alpha} \tag{A.6}$$

where $\widetilde{\mathbf{p}}_v = (\widetilde{\mathbf{p}}_v(u))$ and likewise for $\widetilde{\mathbf{p}}_\ell = (\widetilde{\mathbf{p}}_\ell(u))$.

Moreover if, as we have specified, $\alpha \leq \widetilde{\Psi}_{n,r}/9$, Lemma 3.2 of [1] yields a lower bound on $\widetilde{p}$

$$\widetilde{\mathbf{p}}_v(u) \geq \frac{4}{5}\widetilde{\pi}_{n,r}(u). \tag{A.7}$$

Step 2 We turn to upper bounding $\widetilde{\mathbf{p}}_\ell(u)$. For any $u \in \mathcal{C}[\mathbf{X}]$, we have

$$
\begin{aligned}
\widetilde{\mathbf{p}}_\ell(u) &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\ell \widetilde{\mathbf{W}}_n^t\right)(u) \\
&= \|\ell\|_1 \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \widetilde{\mathbf{W}}_n^t\right)(u) \\
&\overset{\text{(i)}}{=} \|\ell\|_1 \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{\ell}{\|\ell\|_1} \widetilde{\mathbf{W}}_n^t\right)(u) \\
&\overset{\text{(ii)}}{\leq} \|\ell\|_1 \frac{1}{\widetilde{D}_{\min}}
\end{aligned} \tag{A.8}
$$

where we use $\left(\ell \widetilde{\mathbf{W}}_n^t\right)(u)$ to denote $\ell \widetilde{\mathbf{W}}_n^t e_u$.

(i) follows from the fact that since $r < \sigma$, $\mathrm{cut}(\mathcal{C}'[\mathbf{X}], G_{n,r}/\mathcal{C}_\sigma[\mathbf{X}]; G_{n,r}) = 0$. Therefore $(\mathbf{D}_n^{-1})_{uu}(\widetilde{\mathbf{D}}_n)_{uu} = 1$, and as a result

$$
(\ell \widetilde{\mathbf{W}}_n^0)(u) = \ell(u) = 0.
$$

To see (ii), let $q = \frac{\ell}{\|\ell\|_1} \widetilde{\mathbf{W}}_n^{t-1}$. Then

$$
\begin{aligned}
\left(\frac{\ell}{\|\ell\|_1} \widetilde{\mathbf{W}}_n^t\right)(u) &= \left(q \widetilde{\mathbf{W}}_n\right)(u) \\
&\leq \|q\|_1 \|\widetilde{\mathbf{W}}_{\cdot u}\|_\infty \\
&\overset{\text{(iii)}}{\leq} \frac{1}{\widetilde{D}_{\min}}.
\end{aligned}
$$

where $\widetilde{\mathbf{W}}_{\cdot u}$ is the uth column of $\widetilde{\mathbf{W}}_n$. (iii) then follows from the fact that any vertex in $\mathcal{C}[\mathbf{X}]$ is connected only to vertices in $\mathcal{C}_\sigma[\mathbf{X}]$, and therefore every entry of $\widetilde{\mathbf{W}}_{\cdot u}$ is either 0 or at most $1/\widetilde{D}_{\min}$.

Combined, (A.8), (A.7), and (A.6) imply

$$
p_v(u) \geq \frac{4}{5} \widetilde{\pi}_{n,r}(u) - 18 \frac{\widetilde{\Phi}_{n,r}}{\widetilde{D}_{\min} \alpha}.
$$

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* $u' \notin \mathcal{C}_\sigma[\mathbf{X}]$,

$$
\mathbf{p}_v(u') \leq \mathbf{p}_s(u'). \tag{A.9}
$$

Proof. We have

$$\begin{aligned}
\mathbf{p}_v(u') &= \alpha \sum_{T=0}^{\infty} (1-\alpha)^T (e_v \mathbf{W}_n^T)(u) \\
&\overset{(i)}{=} \alpha \sum_{T=1}^{\infty} (1-\alpha)^T (e_v \mathbf{W}_n^T)(u')
\end{aligned}$$

where (i) follows from $v \in \mathcal{C}_\sigma$, $u \notin \mathcal{C}_\sigma$ and therefore $e_v(u) = 0$.

Lemma 3 allows us to make the transition to sums of soakage vectors.

Lemma 3. *Let $G = (V, E)$ be a graph, with associated random walk matrix W. For any $T \geq 1$, q vector, $S \subset V$, and $s_t = s_t(S^c, q)$*

$$qW^T = \sum_{t=0}^{T-1} s_t W^{T-t-1} + q(WI_S)^T \tag{A.10}$$

We prove Lemma 3 after completing the proof of Lemma 2.

Now, along with the fact $u \notin \mathcal{C}_\sigma$, we have

$$\left(e_v \mathbf{W}_n^T\right)(u') = \sum_{t=0}^{T-1} \left(s_t \mathbf{W}_n^{T-t-1}\right)(u')$$

and so

$$\begin{aligned}
\mathbf{p}_v(u) &= \alpha \sum_{T=1}^{\infty} (1-\alpha)^T \left(\sum_{t=0}^{T-1} s_t \mathbf{W}^{T-t-1} \right)(u') \\
&= \alpha \sum_{t=0}^{\infty} \sum_{T=t+1}^{\infty} (1-\alpha)^T \left(s_t \mathbf{W}^{T-t-1}\right)(u') \\
&= \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t+1} \left(s_t \mathbf{W}_n^{\Delta}\right)(u') \\
&\leq \alpha \sum_{t=0}^{\infty} \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta+t} \left(s_t \mathbf{W}_n^{\Delta}\right)(u') \\
&= \alpha \sum_{\Delta=0}^{\infty} (1-\alpha)^{\Delta} \left(s \mathbf{W}_n^{\Delta}\right)(u') \\
&= \mathbf{p}_s(u')
\end{aligned}$$

□

Proof of Lemma 3. Proceed by induction. When $T = 1$,

$$\begin{aligned} qW &= q(WI_S) + q(WI_{S^c}) \\ &= q(WI_S)^T + s_0 \end{aligned}$$

Assume true for T_0. For $T = T_0 + 1$,

$$\begin{aligned} qW^T &= qW^{T_0}W \\ &= \left\{ \sum_{t=0}^{T_0-1} s_t W^{T_0-1-t} + q(WI_S)^{T_0} \right\} W \\ &= \sum_{t=0}^{T_0-1} s_t W^{T-1-t} + q(WI_S)^{T_0}(WI_S + WI_{S^c}) \\ &= \sum_{t=0}^{T-1} s_t W^{T-1-t} + q(WI_S)^T \end{aligned}$$

□

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

$$\begin{aligned} \mathbf{p}_s(u') &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(s\mathbf{W}_n^t \right)(u') \\ &= \|s\|_1 \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1} \mathbf{W}_n^t \right)(u') \\ &= \|s\|_1 \alpha \sum_{t=1}^{\infty} (1-\alpha)^t \left(\frac{s}{\|s\|_1} \mathbf{W}_n^t \right)(u') \\ &\leq \|s\|_1 \frac{1}{\widetilde{D}_{\min}}. \end{aligned} \tag{A.11}$$

Finally, letting $q_t = e_v(\mathbf{W}_n\widetilde{\mathbf{I}}_n)^t$ for ease of notation, and writi we have

$$\begin{aligned}
\|s_t\|_1 &= \|q_t(\mathbf{W}_n\widetilde{\mathbf{I}}_n)\|_1 \\
&= \sum_{u'\in\mathbf{X}}\sum_{u\in\mathbf{X}} q_t(u)(\mathbf{W}_n\widetilde{\mathbf{I}}_n)(u,u') \\
&= \sum_{u'\in\mathbf{X}/\mathcal{C}_\sigma[\mathbf{X}]}\sum_{u\in\mathcal{C}_\sigma[\mathbf{X}]} \frac{q_t(u)}{(\mathbf{D}_n)_{uu}}\mathbf{1}(e_{u,u'}\in G_{n,r}) \\
&= \sum_{u\in\mathcal{C}_\sigma[\mathbf{X}]} \frac{q(u)\Big((\mathbf{D}_n)_{uu}-(\widetilde{\mathbf{D}}_n)_{uu}\Big)}{(\mathbf{D}_n)_{uu}} \\
&= \|q_t(I-\mathbf{D}_n^{-1}\widetilde{\mathbf{D}}_n)\|_1 = \|\ell_t\|_1.
\end{aligned}$$

and as a result $\|s\|_1 = \|\ell\|_1$. Combining with $\|\ell\|_1 \leq 2\frac{\widetilde{\Phi}_{n,r}}{\alpha}$ and (A.11) yields the desired upper bound.

□

Lemma 4. *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

$$\begin{aligned}
p_u &\geq \frac{4}{5}\widetilde{\pi}_{n,r}(u) - \frac{18\kappa_2(\mathcal{C})}{\widetilde{D}_{\min}} \qquad (u\in\mathcal{C}[\mathbf{X}]) \\
p_{u'} &\leq \frac{18\kappa_2(\mathcal{C})}{\widetilde{D}_{\min}} \qquad (u'\in\mathcal{C}'[\mathbf{X}]) \qquad \text{(A.12)}
\end{aligned}$$

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 4 yields

$$\widetilde{\pi}_{n,r}(u) \geq 8\frac{\lambda_\sigma}{\nu(\mathcal{C}_\sigma)\Lambda_\sigma^2} \tag{A.13}$$

and

$$\frac{1}{\widetilde{D}_{\min}} \leq \frac{2}{\nu_d r^d \lambda_\sigma} \tag{A.14}$$

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

Combining (A.12), (A.13) and (A.14), 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.