A. Bias and Variance

Lemma A.1. *Let $X \sim$ Bernoulli(θ) and $Y = aX + b(1-X)$, where a and b are some constants. Then*

$$\mathit{Var}(Y) = \theta(1-\theta)(a-b)^2.$$

Proof of Lemma 3.1

Lemma. *The bias and variance of $\widehat{R}_{naive}(\widehat{o})$ are*

$$\begin{aligned} B(\widehat{R}_{naive}) &= \left|\mathbf{E}[\widehat{R}_{naive}] - R(\widehat{o})\right| \\ &= \frac{\Delta}{|\mathcal{U}|}\left|\sum_{(i,j)\in\mathcal{U}} y_{ij}(1-\pi_{ij})(1-2\widehat{o}_{ij})\right|, \\ \mathit{Var}(\widehat{R}_{naive}) &= \frac{\Delta^2}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}} y_{ij}\pi_{ij}(1-y_{ij}\pi_{ij}). \end{aligned}$$

Proof. We have

$$\begin{aligned} \widehat{R}_{\text{naive}}(o,\widehat{y}) &= \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} \delta(o_{ij},\widehat{o}_{ij}) \\ &= \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} [o_{ij}\delta(1,\widehat{o}_{ij}) + (1-o_{ij})\delta(0,\widehat{o}_{ij})] \\ \therefore\ \mathbf{E}_o[\widehat{R}_{\text{naive}}(o,\widehat{y})] &= \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} [y_{ij}\pi_{ij}\delta(1,\widehat{o}_{ij}) + (1-y_{ij}\pi_{ij})\delta(0,\widehat{o}_{ij})] \\ &= \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} [y_{ij}\pi_{ij}(1-\widehat{o}_{ij})\delta(1,0) + (1-y_{ij}\pi_{ij})\widehat{o}_{ij}\delta(0,1)] \\ &= \frac{\Delta}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} [y_{ij}\pi_{ij}(1-\widehat{o}_{ij}) + (1-y_{ij}\pi_{ij})\widehat{o}_{ij}]\,. \end{aligned}$$

The true risk is

$$\begin{aligned} R(\widehat{y}) &= \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} [y_{ij}\delta(1,\widehat{o}_{ij}) + (1-y_{ij})\delta(0,\widehat{o}_{ij})] \\ &= \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} [y_{ij}(1-\widehat{o}_{ij})\delta(1,0) + (1-y_{ij})\widehat{o}_{ij}\delta(0,1)] \\ &= \frac{\Delta}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} [y_{ij}(1-\widehat{o}_{ij}) + (1-y_{ij})\widehat{o}_{ij}]\,. \end{aligned}$$

Thus the bias is

$$\begin{aligned} \mathrm{B}(\widehat{R}_{\text{naive}}) &= \left|\mathbf{E}[\widehat{R}_{\text{naive}}] - R(\widehat{o})\right| \\ &= \left|\frac{\Delta}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} [y_{ij}\pi_{ij}(1-\widehat{o}_{ij}) + (1-y_{ij}\pi_{ij})\widehat{o}_{ij} - y_{ij}(1-\widehat{o}_{ij}) - (1-y_{ij})\widehat{o}_{ij}]\right| \\ &= \frac{\Delta}{|\mathcal{U}|}\left|\sum_{(i,j)\in\mathcal{U}} y_{ij}(1-\pi_{ij})(1-2\widehat{o}_{ij})\right|. \end{aligned}$$

The variance is

$$
\begin{aligned}
\operatorname{Var}(\widehat{R}_{\text{naive}}) &= \operatorname{Var}\left(\frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}}\left[o_{ij}\delta(1,\widehat{o}_{ij}) + (1-o_{ij})\delta(0,\widehat{o}_{ij})\right]\right) \\
&= \frac{1}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}}\operatorname{Var}\left(o_{ij}\delta(1,\widehat{o}_{ij}) + (1-o_{ij})\delta(0,\widehat{o}_{ij})\right) \\
&= \frac{1}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}} y_{ij}\pi_{ij}(1-y_{ij}\pi_{ij})\left(\delta(1,\widehat{o}_{ij}) - \delta(0,\widehat{o}_{ij})\right)^2 \quad \text{(using Lemma A.1)} \\
&= \frac{\Delta^2}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}} y_{ij}\pi_{ij}(1-y_{ij}\pi_{ij}).
\end{aligned}
$$

□

Lemmas 3.2, 3.3, and 3.4 can be proved similarly.

Proof of Theorem 3.1

Theorem (Comparison of Variances). *For all values of* $\widehat{\pi}, \widehat{y}$, *we have* $Var(\widehat{R}_{AP}) < Var(\widehat{R}_{naive})$, *and* $Var(\widehat{R}_{AP}) < Var(\widehat{R}_{w}) < Var(\widehat{R}_{PU})$

Proof. First we show that $\operatorname{Var}(\widehat{R}_{\text{AP}}) < \operatorname{Var}(\widehat{R}_{\text{naive}})$. We have

$$
\begin{aligned}
\operatorname{Var}(\widehat{R}_{\text{AP}}) &= \frac{\Delta^2}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}} y_{ij}\pi_{ij}(1-y_{ij}\pi_{ij})\psi_{ij}^2, \\
\text{where } \psi_{ij} &= \frac{1-\widehat{y}_{ij}}{1-\widehat{\pi}_{ij}\widehat{y}_{ij}} < 1, \\
\operatorname{Var}(\widehat{R}_{\text{naive}}) &= \frac{\Delta^2}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}} y_{ij}\pi_{ij}(1-y_{ij}\pi_{ij}).
\end{aligned}
$$

Using the fact that $\psi_{ij}^2 < 1\ \forall (i,j)\in\mathcal{U}$, we get $\operatorname{Var}(\widehat{R}_{\text{AP}}) < \operatorname{Var}(\widehat{R}_{\text{naive}})$.

Next, we show that $\operatorname{Var}(\widehat{R}_{w}) < \operatorname{Var}(\widehat{R}_{\text{PU}})$:

$$
\begin{aligned}
\operatorname{Var}(\widehat{R}_{w}) &= \frac{\Delta^2}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}} y_{ij}\pi_{ij}(1-y_{ij}\pi_{ij})\left(\frac{1-\widehat{o}_{ij}}{\widehat{\pi}_{ij}^2} + \widehat{o}_{ij}\psi_{ij}^2\right) \\
\operatorname{Var}(\widehat{R}_{\text{PU}}) &= \frac{\Delta^2}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}} \frac{y_{ij}\pi_{ij}(1-y_{ij}\pi_{ij})}{\widehat{\pi}_{ij}^2} \\
&= \frac{\Delta^2}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}} y_{ij}\pi_{ij}(1-y_{ij}\pi_{ij})\left(\frac{1-\widehat{o}_{ij}}{\widehat{\pi}_{ij}^2} + \frac{\widehat{o}_{ij}}{\widehat{\pi}_{ij}^2}\right) \\
\therefore\ \operatorname{Var}(\widehat{R}_{w}) - \operatorname{Var}(\widehat{R}_{\text{PU}}) &= \frac{\Delta^2}{|\mathcal{U}|^2}\sum_{(i,j)\in\mathcal{U}} y_{ij}\pi_{ij}(1-y_{ij}\pi_{ij})\widehat{o}_{ij}\left(\psi_{ij}^2 - \frac{1}{\widehat{\pi}_{ij}^2}\right) \\
\therefore\ \operatorname{Var}(\widehat{R}_{w}) - \operatorname{Var}(\widehat{R}_{\text{PU}}) &< 0 \quad \left(\text{because } \psi_{ij} < 1 \text{ and } \frac{1}{\widehat{\pi}_{ij}} > 1\right) \\
\therefore\ \operatorname{Var}(\widehat{R}_{w}) &< \operatorname{Var}(\widehat{R}_{\text{PU}}).
\end{aligned}
$$

Next, we show that $\mathrm{Var}(\widehat{R}_{\mathrm{AP}}) < \mathrm{Var}(\widehat{R}_w)$:

$$\mathrm{Var}(\widehat{R}_{\mathrm{AP}}) - \mathrm{Var}(\widehat{R}_w) = \frac{\Delta^2}{|\mathcal{U}|^2} \sum_{(i,j)\in\mathcal{U}} y_{ij}\pi_{ij}(1 - y_{ij}\pi_{ij})(1 - \widehat{o}_{ij}) \left(\psi_{ij}^2 - \frac{1}{\widehat{\pi}_{ij}^2}\right)$$

$$\therefore\ \mathrm{Var}(\widehat{R}_{\mathrm{AP}}) - \mathrm{Var}(\widehat{R}_w) < 0 \ \left(\text{because } \psi_{ij} < 1 \text{ and } \frac{1}{\widehat{\pi}_{ij}} > 1\right)$$

$$\therefore\ \mathrm{Var}(\widehat{R}_{\mathrm{AP}}) < \mathrm{Var}(\widehat{R}_w).$$

□

Proof of Theorem 3.2

Theorem (Comparison of Biases). *Under the bias approximations, a sufficient condition for $B(\widehat{R}_w) = B(\widehat{R}_{PU}) < B(\widehat{R}_{naive})$ is*

$$\frac{\pi_{ij}}{2 - \pi_{ij}} < \widehat{\pi}_{ij} < 1,\ \forall (i,j) \in \mathcal{U},$$

and for $B(\widehat{R}_{AP}) < B(\widehat{R}_{naive})$ is

$$\frac{\pi_{ij}}{2 - \pi_{ij}} < \widehat{\pi}_{ij} < 1 \textit{ and } 0 < \widehat{y}_{ij} < c y_{ij},\ \forall (i,j) \in \mathcal{U}$$

$$\textit{where } c = \frac{2(1 - \pi_{ij})}{1 - \widehat{\pi}_{ij} - \pi_{ij}y_{ij} + (2 - \pi_{ij})\widehat{\pi}_{ij}y_{ij}} \geq 1.$$

Proof. We first derive the sufficient condition for $\mathrm{B}(\widehat{R}_w) = \mathrm{B}(\widehat{R}_{\mathrm{PU}}) < \mathrm{B}(\widehat{R}_{\mathrm{naive}})$. We have

$$\mathrm{B}(\widehat{R}_{\mathrm{naive}}) \approx \frac{\Delta}{|\mathcal{U}|} \sum_{(i,j)\in\mathcal{U}'} y_{ij}(1 - \pi_{ij}),$$

$$\mathrm{B}(\widehat{R}_w) \approx \mathrm{B}(\widehat{R}_{\mathrm{PU}}) \approx \frac{\Delta}{|\mathcal{U}|} \left| \sum_{(i,j)\in\mathcal{U}'} y_{ij} \left(1 - \frac{\pi_{ij}}{\widehat{\pi}_{ij}}\right) \right|.$$

If $1 > \widehat{\pi}_{ij} > \pi_{ij} \forall (i,j) \in \mathcal{U}$, we have

$$\left(1 - \frac{\pi_{ij}}{\widehat{\pi}_{ij}}\right) > 0\ \ \forall (i,j) \in \mathcal{U}$$

$$\therefore \mathrm{B}(\widehat{R}_w) \approx \mathrm{B}(\widehat{R}_{\mathrm{PU}}) \approx \frac{\Delta}{|\mathcal{U}|} \sum_{(i,j)\in\mathcal{U}'} y_{ij} \left(1 - \frac{\pi_{ij}}{\widehat{\pi}_{ij}}\right).$$

$$\begin{aligned}
\therefore \mathrm{B}(\widehat{R}_w) - \mathrm{B}(\widehat{R}_{\mathrm{naive}}) &= \frac{\Delta}{|\mathcal{U}|} \sum_{(i,j)\in\mathcal{U}'} y_{ij} \left(\pi_{ij} - \frac{\pi_{ij}}{\widehat{\pi}_{ij}}\right) \\
&= \frac{\Delta}{|\mathcal{U}|} \sum_{(i,j)\in\mathcal{U}'} y_{ij}\pi_{ij} \left(1 - \frac{1}{\widehat{\pi}_{ij}}\right) \\
&< 0 \ \ (\text{because } \widehat{\pi}_{ij} < 1)
\end{aligned}$$

$$\therefore \mathrm{B}(\widehat{R}_w) < \mathrm{B}(\widehat{R}_{\mathrm{naive}}).$$

If $0 < \widehat{\pi}_{ij} \leq \pi_{ij} \forall (i,j) \in \mathcal{U}$, we have

$$\left(1 - \frac{\pi_{ij}}{\widehat{\pi}_{ij}}\right) \leq 0\ \ \forall (i,j) \in \mathcal{U}$$

$$\therefore \mathrm{B}(\widehat{R}_w) \approx \mathrm{B}(\widehat{R}_{\mathrm{PU}}) \approx \frac{\Delta}{|\mathcal{U}|} \sum_{(i,j)\in\mathcal{U}'} y_{ij} \left(\frac{\pi_{ij}}{\widehat{\pi}_{ij}} - 1\right). \quad .$$

Then, a sufficient condition for $\mathrm{B}(\widehat{R}_w) = \mathrm{B}(\widehat{R}_{\text{PU}}) < \mathrm{B}(\widehat{R}_{\text{naive}})$ is

$$y_{ij}\left(\frac{\pi_{ij}}{\widehat{\pi}_{ij}} - 1\right) < y_{ij}(1-\pi_{ij}) \;\; \forall (i,j) \in \mathcal{U}$$
$$\therefore\;\; y_{ij}\left(\frac{\pi_{ij}}{\widehat{\pi}_{ij}} - 1\right) < y_{ij}(1-\pi_{ij}) \;\; \forall (i,j) \in \mathcal{U}$$
$$\therefore\;\; \widehat{\pi}_{ij} > \frac{2}{2-\pi_{ij}} \;\; \forall (i,j) \in \mathcal{U}.$$

Next, we derive the sufficient condition for $\widehat{R}_{\text{AP}} < \widehat{R}_{\text{naive}}$. Observe that

$$\frac{\pi_{ij}}{2-\pi_{ij}} < \widehat{\pi}_{ij} < 1 \;\; \forall (i,j) \in \mathcal{U}$$
$$\therefore (1-\pi_{ij})y_{ij} - (1-\pi_{ij}y_{ij})\tau_{ij} \geq 0 \;\; \forall (i,j) \in \mathcal{U}$$
$$\therefore \widehat{R}_{\text{AP}} \approx \frac{\Delta}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}'}\left[(1-\pi_{ij})y_{ij} - (1-\pi_{ij}y_{ij})\tau_{ij}\right], \text{ where } \tau_{ij} = \left(\frac{\widehat{y}_{ij}(1-\widehat{\pi}_{ij})}{1-\widehat{\pi}_{ij}\widehat{y}_{ij}}\right).$$

Therefore, when $\frac{\pi_{ij}}{2-\pi_{ij}} < \widehat{\pi}_{ij} < 1 \;\; \forall (i,j) \in \mathcal{U}$, a sufficient condition for $\widehat{R}_{\text{AP}} < \widehat{R}_{\text{naive}}$ is

$$(1-\pi_{ij})y_{ij} - (1-\pi_{ij}y_{ij})\tau_{ij} < y_{ij}(1-\pi_{ij}) \;\; \forall (i,j) \in \mathcal{U}$$
$$\therefore\;\; 0 < \widehat{y}_{ij} < \left(\frac{2(1-\pi_{ij})}{1-\widehat{\pi}_{ij} - \pi_{ij}y_{ij} + (2-\pi_{ij})\widehat{\pi}_{ij}y_{ij}}\right) y_{ij} \;\; \forall (i,j) \in \mathcal{U}.$$

□

B. Generalization Bound

Proof of Theorem 4.1

Theorem (Generalization Bound). *Let $\mathcal{F}$ be a class of functions $(\widehat{\pi}, \widehat{y})$. Let $\delta(o_{ij}, \widehat{y}_{ij}) \leq \eta \; \forall (i,j) \in \mathcal{U}$ and $\widehat{\pi}_{ij} \geq \epsilon > 0 \; \forall (i,j) \in \mathcal{U}$. Then, for $\widehat{R} \in \left\{\widehat{R}_w, \widehat{R}_{PU}, \widehat{R}_{AP}\right\}$, with probability at least $1-\delta$, we have*

$$R(\widehat{y}) \leq \widehat{R}(\widehat{y}, \widehat{\pi}) + B(\widehat{R}) + 2\mathcal{G}(\mathcal{F}, \widehat{R}) + M \tag{6}$$
$$\leq \widehat{R}(\widehat{y}, \widehat{\pi}) + B(\widehat{R}_w) + 2\widehat{\mathcal{G}}(\mathcal{F}, \widehat{R}_w) + 3M, \tag{7}$$

where $M = \sqrt{\frac{4\eta^2}{\epsilon^2|\mathcal{U}|}\log(\frac{2}{\delta})}$ and $B(\widehat{R})$ is the bias of $\widehat{R}$ derived in Section 3.

Proof. We proceed similarly to the standard Rademacher complexity generalization bound proof (Shalev-Shwartz & Ben-David, 2014)[Ch. 26]. Observe that

$$R(\widehat{y}) = R(\widehat{y}) - \mathbf{E}_o[\widehat{R}(o, \widehat{y}, \widehat{\pi})] + \mathbf{E}_o[\widehat{R}(o, \widehat{y}, \widehat{\pi})]$$
$$\leq B(\widehat{R}) + \mathbf{E}_o[\widehat{R}(o, \widehat{y}, \widehat{\pi})]. \tag{8}$$

Let $\Phi(o) = \sup_{(\widehat{\pi}, \widehat{y})\in\mathcal{F}}\left[\mathbf{E}_o[\widehat{R}(o, \widehat{y}, \widehat{\pi})] - \widehat{R}(o, \widehat{y}, \widehat{\pi})\right]$. Then

$$\mathbf{E}_o[\widehat{R}(o, \widehat{y}, \widehat{\pi})]| \leq \widehat{R}(o, \widehat{y}, \widehat{\pi}) + \Phi(o). \tag{9}$$

Now we upper bound $\Phi(o)$. Since $\delta(o_{ij}, \widehat{y}_{ij}) \leq \eta \; \forall (i,j)$ and $\widehat{\pi}_{ij} \geq \epsilon > 0, \; \forall (i,j)$ and $\forall \, \widehat{R} \in \left\{\widehat{R}_w, \widehat{R}_{\text{PU}}, \widehat{R}_{\text{AP}}\right\}$, we have

$$|\Phi(o) - \Phi(\tilde{o})| \leq \frac{2\eta}{\epsilon},$$

if o and $\tilde{o}$ differ in only one coordinate, i.e., $o_{ij} \neq \tilde{o}_{ij}$ for some $(i,j) \in \mathcal{U}$ and $o_{lm} = \tilde{o}_{lm} \forall (l,m) \in \mathcal{U}$ s.t. $(i,j) \neq (l,m)$. Using McDiarmid's Inequality, with probability at least $1-\delta$, we have

$$\Phi(o) \leq \mathbf{E}[\Phi(o)] + C. \tag{10}$$

Next, we upper bound $\mathbf{E}[\Phi(o)]$. Let $\bar{o}$ be a ghost sample independently drawn having the same distribution as o. We have

$$\begin{aligned}
\mathbf{E}[\Phi(o)] &= \mathbf{E}_o\left[\sup_{(\widehat{\pi},\widehat{y})\in\mathcal{F}}\left[\mathbf{E}_o[\widehat{R}(o,\widehat{y},\widehat{\pi})] - \widehat{R}(o,\widehat{y},\widehat{\pi})\right]\right] \\
&= \mathbf{E}_o\left[\sup_{(\widehat{\pi},\widehat{y})\in\mathcal{F}} \mathbf{E}_{\bar{o}}\left[\widehat{R}(\bar{o},\widehat{y},\widehat{\pi}) - \widehat{R}(o,\widehat{y},\widehat{\pi}) \,\middle|\, o\right]\right] \\
&= \mathbf{E}_o\left[\sup_{(\widehat{\pi},\widehat{y})\in\mathcal{F}} \mathbf{E}_{\bar{o}}\left[\frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} r(\bar{o}_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij}) - \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} r(o_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij}) \,\middle|\, o\right]\right] \\
&\leq \mathbf{E}_{o,\bar{o}}\left[\sup_{(\widehat{\pi},\widehat{y})\in\mathcal{F}}\left[\frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} r(\bar{o}_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij}) - \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} r(o_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij})\right]\right] \quad \text{(Jensen's Inequality)} \\
&= \mathbf{E}_{o,\bar{o},\sigma}\left[\sup_{(\widehat{\pi},\widehat{y})\in\mathcal{F}}\left[\frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} \sigma_{ij} r(\bar{o}_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij}) - \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} \sigma_{ij} r(o_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij})\right]\right] \\
&= \mathbf{E}_{o,\bar{o},\sigma}\left[\sup_{(\widehat{\pi},\widehat{y})\in\mathcal{F}}\left[\frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} \sigma_{ij} r(\bar{o}_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij}) + \frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} \sigma_{ij} r(o_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij})\right]\right] \\
&\leq \mathbf{E}_{o,\bar{o},\sigma}\left[\sup_{(\widehat{\pi},\widehat{y})\in\mathcal{F}}\left[\frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} \sigma_{ij} r(\bar{o}_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij})\right] + \sup_{(\widehat{\pi},\widehat{y})\in\mathcal{F}}\left[\frac{1}{|\mathcal{U}|}\sum_{(i,j)\in\mathcal{U}} \sigma_{ij} r(o_{ij},\widehat{\pi}_{ij},\widehat{y}_{ij})\right]\right] \\
&= 2\mathcal{G}(\mathcal{F},\widehat{R}).
\end{aligned} \tag{11}$$

Combining Eqs. 8, 9, 10, and 11, we get Eq. 6. Another application of McDiarmid's Inequality allows us to obtain Eq. 7 from Eq. 6. □

C. Feedback Loops

Lemma C.1 (Binomial Tail Bound). *If the random variable $X_n \sim \frac{1}{n}$Binomial(n,θ), then for $\epsilon > 0$, we have*

$$\mathbf{P}(|X_n - \theta| > \epsilon) \leq 2\exp\left(-2n\epsilon^2\right).$$

Proof. Observe that $X_n \in [0,1]$. Applying Hoeffding's inequality gives us the desired result. □

Lemma C.2. *Let $n \in \mathbb{N}$ and κ be a fixed $C-1$ simplex such that $\kappa_v n \in \mathbb{N}$ $\forall v \in [C]$. The random variable $\tilde{q}_v \sim \frac{1}{\kappa_v n}$Binomial$(\kappa_v n, q_v)$, where $q_v \in (0,1)$. Assume that $q_v > q_w$ if $v > w$. We denote as $\widehat{e}$ the following $C-1$ simplex:*

$$\widehat{e} = \frac{1}{Z}\left[\kappa_1\tilde{q}_1, \kappa_2\tilde{q}_2, \ldots, \kappa_C\tilde{q}_C\right], \;\; \textit{where } Z = \sum_{i\in[C]} \kappa_i\tilde{q}_i.$$

Let $\widehat{e}_{vw} = \frac{\widehat{e}_v}{\widehat{e}_v + \widehat{e}_w} = \frac{\kappa_v\tilde{q}_v}{\kappa_w\tilde{q}_w + \kappa_w\tilde{q}_w}$ and $\kappa_{vw} = \frac{\kappa_v}{\kappa_v+\kappa_w}$. Then for a constant ρ_{vw} such that

$$0 < \rho_{vw} < \frac{\kappa_v\kappa_w(q_v - q_w)}{q_v\kappa_v^2 + (q_v+q_w)\kappa_v\kappa_w + q_w\kappa_w^2},$$

we have

$$\begin{aligned}
&|\tilde{q}_v - q_v| < \epsilon_{vw}, \; |\tilde{q}_w - q_w| < \epsilon_{vw} \implies \widehat{e}_{vw} - \kappa_{vw} > \rho_{vw}, \\
&\textit{for some constant } \epsilon_{vw} \textit{ s.t. } 0 < \epsilon_{vw} < \frac{\rho_{vw}q_v\kappa_v^2 - \kappa_v\kappa_w(q_w - q_v) + q_w\rho_{vw}\kappa_v(\kappa_v - \kappa_w)}{\rho_{vw}(\kappa_v^2 - \kappa_w^2) - 2\kappa_v\kappa_w}.
\end{aligned}$$

This is saying that, for (v, w) s.t. $v > w$ the simplex $\widehat{e}$ will be more skewed towards v than the simplex κ if the sampled $\tilde{q}_v$ and $\tilde{q}_w$ are close to their mean values q_v and q_w, respectively.

Proof. Observe that if $|\tilde{q}_v - q_v| < \epsilon_{vw}$ and $|\tilde{q}_w - q_w| < \epsilon_{vw}$, then the lowest value that $\widehat{e}_{vw}$ can take is

$$\widehat{e}_{vw}^{(\text{min})} = \frac{\kappa_v(q_v - \epsilon_{vw})}{\kappa_v(q_v - \epsilon_{vw}) + \kappa_w(q_w + \epsilon_{vw})}, \text{ and}$$
$$\widehat{e}_{vw}^{(\text{min})} - \kappa_{vw} > \rho_{vw} \implies \widehat{e}_{vw} - \kappa_{vw} > \rho_{vw}.$$

Therefore, we have

$$\begin{aligned} &\widehat{e}_{vw}^{(\text{min})} - \kappa_{vw} > \rho_{vw} \text{ and } \epsilon_{vw} < q_w \\ \Longleftarrow\; &\underbrace{\frac{\kappa_v(q_v - \epsilon_{vw})}{\kappa_v(q_v - \epsilon_{vw}) + \kappa_w(q_w + \epsilon_{vw})} - \frac{\kappa_v}{\kappa_v + \kappa_w} > \rho_{vw}}_{(1)} \text{ and } \rho_{vw} < \frac{\kappa_v\kappa_w(q_v - q_w)}{q_v\kappa_v^2 + (q_v + q_w)\kappa_v\kappa_w + q_w\kappa_w^2}. \end{aligned}$$

The inequality (1) above can further be simplified as

$$\begin{aligned} &\frac{\kappa_v(q_v - \epsilon_{vw})}{\kappa_v(q_v - \epsilon_{vw}) + \kappa_w(q_w + \epsilon_{vw})} - \frac{\kappa_v}{\kappa_v + \kappa_w} > \rho_{vw} \\ \Longleftarrow\; &\epsilon_{vw} < \frac{\rho_{vw}q_v\kappa_v^2 - \kappa_v\kappa_w(q_w - q_v) + q_w\rho_{vw}\kappa_v(\kappa_v - \kappa_w)}{\rho_{vw}(\kappa_v^2 - \kappa_w^2) - 2\kappa_v\kappa_w}. \end{aligned}$$

This completes the proof. □

Lemma C.3. *Let α be a fixed $C - 1$ simplex and $\widehat{e}$ be the following $G - 1$ simplex, $\widehat{e} = \frac{1}{Z}[\alpha_1\tilde{q}_1, \alpha_2\tilde{q}_2, \ldots, \alpha_C\tilde{q}_C]$, where $Z = \sum_{z\in[C]} \alpha_z\tilde{q}_z$ and the vector $\kappa \sim \frac{1}{n}$Multinomial$(n, \widehat{e})$. Let $\widehat{e}_{vw} = \frac{\widehat{e}_v}{\widehat{e}_v + \widehat{e}_w} = \frac{\tilde{q}_v}{\tilde{q}_w + \tilde{q}_w}$ and $\kappa_{vw} = \frac{\kappa_v}{\kappa_v + \kappa_w}$.*
Assume that $|\tilde{q}_z - q_z| < \epsilon\ \forall z \in [C]$ where $q_z \in (0, 1)$ are fixed. If $|\kappa_v - \widehat{e}_v < \frac{\eta_{nw}}{C}|$ and $|\kappa_w - \widehat{e}_w < \frac{\eta_{nw}}{C}|$, then for some constant ρ, we have

$$\widehat{e}_{vw} - \kappa_{vw} < \rho, \;\; \textit{when} \;\; \eta_{vw} < \rho\left(\frac{q_v + q_w}{\max_{z\in[C]} q_z + \epsilon}\right).$$

Proof. If $|\kappa_v - \widehat{e}_v < \frac{\eta_{nw}}{C}|$ and $|\kappa_w - \widehat{e}_w < \frac{\eta_{nw}}{C}|$, then the smallest value that κ_{vw} can achieve is

$$\kappa_{vw}^{(\text{min})} = \frac{\widehat{e}_v - \frac{\eta_{nw}}{C}}{\widehat{e}_v + \widehat{e}_w}.$$

This means that

$$\begin{aligned} &\widehat{e}_{vw} - \kappa_{vw} < \rho \\ \Longleftarrow\; &\widehat{e}_{vw} - \kappa_{vw}^{(\text{min})} < \rho \\ \Longleftrightarrow\; &\frac{\frac{\eta_{nw}}{C}}{\widehat{e}_v + \widehat{e}_w} < \rho \\ \Longleftrightarrow\; &\frac{\eta_{nw}}{C} < \rho(\widehat{e}_v + \widehat{e}_w). \end{aligned}$$

Since $|\tilde{q}_z - q_z| < \epsilon \ \forall z \in [C]$, we have

$$\begin{aligned}
\widehat{e}_v &= \frac{\alpha_v \tilde{q}_v}{\sum_{z\in[C]} \alpha_z \tilde{q}_z} \\
&> \frac{\alpha_v (q_v - \epsilon)}{\alpha_v (q_v - \epsilon) + \sum_{z\in[C], z\neq v} \alpha_z (q_z + \epsilon)} \\
&> \frac{\alpha_v (q_v - \epsilon)}{C \max_{z\in[C]} \alpha_z (q_z + \epsilon)}, \\
\text{and similarly } \ \widehat{e}_w &> \frac{\alpha_w (q_w - \epsilon)}{C \max_{z\in[C]} \alpha_w (q_z + \epsilon)} \\
\therefore \ \widehat{e}_v + \widehat{e}_w &> \frac{\alpha_v (q_v - \epsilon) + \alpha_w (q_w - \epsilon)}{C \max_{z\in[C]} (q_z + \epsilon)}.
\end{aligned}$$

Therefore, we can set η_{vw} such that

$$\begin{aligned}
\frac{\eta_{nw}}{C} &< \rho \left(\frac{\alpha_v (q_v - \epsilon) + \alpha_w (q_w - \epsilon)}{C \max_{z\in[C]} (q_z + \epsilon)} \right) \\
\therefore \ \eta_{nw} &< \rho \left(\frac{\alpha_v (q_v - \epsilon) + \alpha_w (q_w - \epsilon)}{\max_{z\in[C]} q_z + \epsilon} \right).
\end{aligned}$$

□

Proof of Theorem 5.1

Theorem. *WLOG, assume that $q_v > q_w$ if $v > w$. Let $\kappa_{vw}^{(t)} = \frac{\kappa_v^{(t)}}{\kappa_v^{(t)} + \kappa_w^{(t)}}$. Let $A_{vw}^{(t)}$ represent the event that relative fraction of recommendations from g_v to that from g_w increases at time t, i.e., $\kappa_{vw}^{(t+1)} > \kappa_{vw}^{(t)}$. Let $A^{(t)}$ be the event that all relative fractions get skewed towards g_v from g_w if $q_v > q_w$, i.e. $A^{(t)} = \bigcap_{(v,w)\in\mathcal{S}} A_{vw}^{(t)}$, where $\mathcal{S} = \{(v, w) : v \in [C], w \in [C], v > w\}$. Then, for constants $\epsilon, \eta > 0$ that only depend on $\kappa^{(t)}$ and q, we have*

$$\begin{aligned}
\mathbf{P}(A^{(t)}|\kappa^{(t)}) &\geq 1 - 2C \left[\exp\left(-2n\epsilon^2\right) + \exp\left(-\frac{2n\eta^2}{C^2}\right) \right] \\
&\geq 1 - 2C \exp\left(-\mathcal{O}\left(\frac{n}{C^2}\right)\right)
\end{aligned}$$

Proof. We know that the estimated probabilities $\widehat{q}_v^{(t)}$ have distribution $\widehat{q}_v^{(t)}|\kappa^{(t)} \sim \frac{1}{n}\text{Binomial}(n\kappa_v^{(t)}, q_v)$. The simplex with normalized probabilities is $\widehat{e}^{(t+1)} = \frac{1}{Z}[\widehat{q}_1^{(t)}, \widehat{q}_2^{(t)}, \ldots, \widehat{q}_C^{(t)}]$, where $Z = \sum_{z\in[C]} \widehat{q}_z^{(t)}$.

Let $\tilde{q}_v^{(t)} = \frac{\widehat{q}_v^{(t)}}{\kappa_v^{(t)}}$. Observe that $\tilde{q}_v^{(t)}|\kappa^{(t)} \sim \frac{1}{n\kappa_v^{(t)}}\text{Binomial}(n\kappa_v^{(t)}, q_v)$. We denote by $\widehat{e}_{vw}^{(t+1)}$,

$$\widehat{e}_{vw}^{(t+1)} = \frac{\widehat{e}_v^{(t+1)}}{\widehat{e}_v^{(t+1)} + \widehat{e}_w^{(t+1)}} = \frac{\kappa_v^{(t)} \tilde{q}_v^{(t)}}{\kappa_v^{(t)} \tilde{q}_v^{(t)} + \kappa_w^{(t)} \tilde{q}_w^{(t)}}.$$

There are two main parts to the proof. First, we show that, with high probability, $\widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t)} > \rho \ \forall (v, w) \in \mathcal{S}$ for some constant ρ. Then, we show that, with high probability, $\widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t+1)} < \rho \ \forall (v, w) \in \mathcal{S}$. We combine these two results to show that, with high probability, $\kappa_{vw}^{(t+1)} > \kappa_{vw}^{(t)} \ \forall (v, w) \in \mathcal{S}$.

Using Lemma C.2, for some $(v, w) \in \mathcal{S}$, we know that for some constant ρ_{vw} such that

$$0 < \rho_{vw} < \frac{\kappa_v^{(t)} \kappa_w^{(t)} (q_v - q_w)}{q_v (\kappa_v^{(t)})^2 + (q_v + q_w) \kappa_v^{(t)} \kappa_w^{(t)} + q_w (\kappa_w^{(t)})^2},$$

we have

$$
\begin{aligned}
&|\tilde{q}_v^{(t)} - q_v| \leq \epsilon_{vw} \textit{ and } |\tilde{q}_w^{(t)} - q_w| \leq \epsilon_{vw} \implies \widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t)} \geq \rho_{vw}, \\
&\text{for a constant } \epsilon_{vw} \text{ s.t. } 0 < \epsilon_{vw} < \frac{\rho_{vw} q_v (\kappa_v^{(t)})^2 - \kappa_v^{(t)} \kappa_w^{(t)} (q_w - q_v) + q_w \rho_{vw} \kappa_v^{(t)} (\kappa_v^{(t)} - \kappa_w^{(t)})}{\rho_{vw}((\kappa_v^{(t)})^2 - (\kappa_w^{(t)})^2) - 2\kappa_v^{(t)} \kappa_w^{(t)}} \\
\implies &\mathbf{P}\left(\widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t)} \geq \rho_{vw}\right) \geq \mathbf{P}\left(|\tilde{q}_v^{(t)} - q_v| \leq \epsilon_{vw}, |\tilde{q}_w^{(t)} - q_w| \leq \epsilon_{vw}\right).
\end{aligned}
$$

Intuitively, this is saying that $\widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t)} > \rho_{vw}$ if $\tilde{q}_v^{(t)}$ and $\tilde{q}_w^{(t)}$ are close to q_v and q_w, respectively. Let $\rho = \min_{(v,w)\in\mathcal{S}} \rho_{vw}$ and $\epsilon = \min_{(v,w)\in\mathcal{S}} \epsilon_{vw}$. Then we have

$$
\begin{aligned}
\mathbf{P}\left(\bigcap_{(v,w)\in\mathcal{S}} \widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t)} \geq \rho\right) &\geq \mathbf{P}\left(\bigcap_{z\in[C]} |\tilde{q}_z^{(t)} - q_z| \leq \epsilon\right) && (12) \\
&= 1 - \mathbf{P}\left(\bigcup_{z\in[C]} |\tilde{q}_z^{(t)} - q_z| \geq \epsilon\right) && (13) \\
&\geq 1 - \sum_{z=1}^{C} \mathbf{P}\left(|\tilde{q}_z^{(t)} - q_z| \geq \epsilon\right) \quad \text{(Union Bound)} && (14) \\
&\geq 1 - \sum_{z=1}^{C} 2\exp\left(-2n\epsilon^2\right) \quad \text{(using Lemma C.1)} \\
&= 1 - 2C\exp\left(-2n\epsilon^2\right). && (15)
\end{aligned}
$$

Now, we show that $\widehat{e}_{vw}^{(t+1)}$ is close to $\kappa_{vw}^{(t+1)}$. We know that $\kappa^{(t+1)} \sim \frac{1}{n}\text{Multinomial}(n, \widehat{e}^{(t+1)})$. Let the event $Q^{(t)} = \bigcap_{z\in[C]} |\tilde{q}_z^{(t)} - q_z| \leq \epsilon$. Using Lemma C.3, we know that, under $Q^{(t)}$, for some constant η_{vw}, we have

$$
\begin{aligned}
&\left|\widehat{e}_v^{(t+1)} - \kappa_v^{(t+1)}\right| < \frac{\eta_{vw}}{C} \text{ and } \left|\widehat{e}_w^{(t+1)} - \kappa_w^{(t+1)}\right| < \frac{\eta_{vw}}{C} \implies \widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t+1)} < \rho, \\
&\text{where } 0 < \eta_{vw} < \frac{\kappa_v^{(t+1)}(q_v - \epsilon) + \kappa_w^{(t+1)}(q_w - \epsilon)}{\max_{z\in[C]} \kappa_z^{(t+1)}(q_z + \epsilon)} \\
\implies &\mathbf{P}\left(\widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t+1)} < \rho \,\middle|\, Q^{(t)}\right) \geq \mathbf{P}\left(\left|\widehat{e}_v^{(t+1)} - \kappa_v^{(t+1)}\right| < \frac{\eta_{vw}}{C},\ \left|\widehat{e}_w^{(t+1)} - \kappa_w^{(t+1)}\right| < \frac{\eta_{vw}}{C}\right).
\end{aligned}
$$

Intuitively, this is saying that $\widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t+1)} < \rho$ if $\kappa_v^{(t+1)}$ and $\kappa_w^{(t+1)}$ are close to $\widehat{e}_v^{(t+1)}$ and $\widehat{e}_w^{(t+1)}$, respectively. Thus, for $\eta = \min_{(v,w)\in\mathcal{S}} \eta_{vw}$, we have

$$
\begin{aligned}
\mathbf{P}\left(\bigcap_{(v,w)\in\mathcal{S}} \widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t+1)} \leq \rho \,\middle|\, Q^{(t)}\right) &\geq \mathbf{P}\left(\bigcap_{z\in[C]} |\widehat{e}_z^{(t+1)} - \kappa_z^{(t+1)}| \leq \frac{\eta}{C}\right) \\
&= 1 - \mathbf{P}\left(\bigcup_{z\in[C]} |\widehat{e}_z^{(t+1)} - \kappa_z^{(t+1)}| \geq \frac{\eta}{C}\right) \\
&\geq 1 - \sum_{z=1}^{C} \mathbf{P}\left(|\widehat{e}_z^{(t+1)} - \kappa_z^{(t+1)}| > \frac{\eta}{C}\right) \quad \text{(Union Bound)} \\
&\geq 1 - 2C\exp\left(-\frac{2n\eta^2}{C^2}\right) \quad \text{(using Lemma C.1)}. && (16)
\end{aligned}
$$

Combining Eq. 15 and 16, we get the desired result as follows:

$$\begin{aligned}
\mathbf{P}\left(\bigcap_{(v,w)\in\mathcal{S}} A_{vw}^{(t)}\right) &= \mathbf{P}\left(\bigcap_{(v,w)\in\mathcal{S}} \kappa_{vw}^{(t+1)} > \kappa_{vw}^{(t)}\right) \\
&\geq \mathbf{P}\left(\bigcap_{(v,w)\in\mathcal{S}} \left(\widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t)} \geq \rho,\ \ \widehat{e}_{vw}^{(t+1)} - \kappa_{vw}^{(t+1)} \leq \rho\right)\right) \\
&\geq \mathbf{P}\left(\bigcap_{z\in[C]} \left(|\widehat{e}_{z}^{(t+1)} - \kappa_{z}^{(t+1)}| \leq \frac{\eta}{C},\ \ |\tilde{q}_z^{(t)} - q_z| \leq \epsilon\right)\right) \\
&= \mathbf{P}\left(\bigcap_{z\in[C]} |\widehat{e}_{z}^{(t+1)} - \kappa_{z}^{(t+1)}| \leq \frac{\eta}{C} \,\middle|\, Q^{(t)}\right) \mathbf{P}\left(\bigcap_{z\in[C]} |\tilde{q}_z^{(t)} - q_z| \leq \epsilon\right) \\
&\geq \left(1 - 2C\exp\left(-\frac{2n\eta^2}{C^2}\right)\right)\left(1 - 2C\exp\left(-2n\epsilon^2\right)\right) \\
&\geq 1 - 2C\left[\exp\left(-2n\epsilon^2\right) + \exp\left(-\frac{2n\eta^2}{C^2}\right)\right] \\
&\geq 1 - 2C\exp\left(-\mathcal{O}\left(\frac{n}{C^2}\right)\right).
\end{aligned}$$

□

Proof of Theorem 5.2

Lemma C.4 (Convergence in Probability)**.** *Let X_n, Y_n, and Z be random variables such that $X_n \xrightarrow{p} Y_n$ and $Y_n \xrightarrow{p} Z$, then $X_n \xrightarrow{p} Z$.*

Proof. For any $\epsilon > 0$, we have

$$\begin{aligned}
\mathbf{P}(|X_n - Z| \geq \epsilon) &= \mathbf{P}(|X_n - Y_n + Y_n - Z| \geq \epsilon) \\
&\leq \mathbf{P}(|X_n - Y_n| + |Y_n - Z| \geq \epsilon) \\
&\leq \mathbf{P}\left(|X_n - Y_n| \geq \frac{\epsilon}{2}\right) + \mathbf{P}\left(|X_n - Y_n| \geq \frac{\epsilon}{2}\right) \\
&= 0.
\end{aligned}$$

Therefore, $X_n \xrightarrow{p} Z$. □

Theorem. *Let $q_v > q_w$. As $n \to \infty$, $\kappa_{vw}^{(t)} \xrightarrow{p} 1 - \frac{1}{1+c^t}$, where $c = \frac{q_v}{q_w}$.*

Proof. At time step t, the fraction of recommendations from each group is κ_t. From group g_v, the user cites papers according to probability q_v. Therefore, $\widehat{q}_v^{(t)} \xrightarrow{p} \kappa_v^{(t)} q_v$. And the normalized estimate is $\widehat{e}^{(t+1)} = \frac{1}{S}[\kappa_1^{(t)} q_1, \ldots, \kappa_C^{(t)} q_C]$, where $S = \sum_{z\in[C]} \kappa_z^{(t)} q_z$. Since $\kappa^{(t+1)} \sim \frac{1}{n}\text{Multinomial}(n, \widehat{e}^{(t+1)})$, we have

$$\begin{aligned}
\kappa^{(t+1)} &\xrightarrow{p} \widehat{e}^{(t+1)} \\
\frac{\kappa_v^{(t+1)}}{\kappa_w^{(t+1)}} &\xrightarrow{p} \frac{q_v \kappa_v^{(t)}}{q_w \kappa_w^{(t)}} \\
&= c\frac{\kappa_v^{(t)}}{\kappa_w^{(t)}}.
\end{aligned} \tag{17}$$

Table 5. The distribution of the FOS in the two real-world datasets.

FOS	DATASET 1	DATASET 2
ART	0.03%	0.08%
BIOLOGY	26.48%	23.43%
BUSINESS	0.38%	0.10%
CHEMISTRY	10.11%	15.67%
COMPUTER SCIENCE	9.40%	3.42%
ECONOMICS	2.51%	0.03%
ENGINEERING	6.24%	17.98%
ENVIRONMENTAL SCIENCE	0.13%	0.03%
GEOGRAPHY	0.48%	0.40%
GEOLOGY	1.45%	0.46%
HISTORY	0.04%	0.03%
MATERIALS SCIENCE	3.06%	19.09%
MATHEMATICS	7.17%	1.03%
MEDICINE	21.28%	13.90%
PHILOSOPHY	0.03%	0.01%
PHYSICS	2.99%	3.14%
POLITICAL SCIENCE	0.18%	0.01%
PSYCHOLOGY	7.49%	1.14%
SOCIOLOGY	0.55%	0.05%

We know that $\frac{\kappa_v^{(1)}}{\kappa_w^{(1)}} \xrightarrow{p} c$. Combining this with Eq. 17 and using Lemma C.4 recursively, we get

$$
\begin{aligned}
\frac{\kappa_v^{(t)}}{\kappa_w^{(t)}} &\xrightarrow{p} c^t \\
\therefore\ 1 - \frac{1}{1 + \frac{\kappa_v^{(t)}}{\kappa_w^{(t)}}} &\xrightarrow{p} 1 - \frac{1}{1 + c^t} \quad \text{(Continuous mapping theorem)} \\
\therefore\ \frac{\kappa_v^{(t)}}{\kappa_v^{(t)} + \kappa_w^{(t)}} &\xrightarrow{p} 1 - \frac{1}{1 + c^t} \\
\therefore\ \kappa_{vw}^{(t)} &\xrightarrow{p} 1 - \frac{1}{1 + c^t}.
\end{aligned}
$$

□

D. Experiments

Table 5 provides the distribution of the various FOS in both the datasets used for the real-world dataset experiments (Section 6.2). We can see that the FOS distributions are different. For example, Dataset 2 has substantially more *Materials Science* and *Engineering* papers.