Appendix

A   Proof of Lemma 4.3

We show that $D(uvw) \leq 2.5L(uvw)$ for all $c \in L$ and $u, v, w \in S_c^t$ by enumerating all possible combinations of $\big(\phi(uv), \phi(vw), \phi(wu)\big)$. Note that due to symmetry, we do not consider the order of $\big(\phi(uv), \phi(vw), \phi(wu)\big)$.

Without loss of generality, we first consider the special case of $u = v$. Observing that $D(uuu) = L(uuu) = 0$, so we consider $w \neq u$. By definition, it is easy to obtain that

$$D(uuw) = 2d(uw \mid u) = \begin{cases} 2x_{uw}^c, & \text{if } \phi(uw) = c, \\ 2, & \text{if } \phi(uv) \in L \setminus \{c\}, \\ 2(1 - x_{uw}^c), & \text{if } \phi(uw) = \gamma, \end{cases}$$

$$L(uuw) = 2lp(uw \mid u) = \begin{cases} 2x_{uv}^{\phi(uw)}, & \text{if } \phi(uw) \in L, \\ 2\sum_{i \in L}(1 - x_{uw}^i), & \text{if } \phi(uw) = \gamma. \end{cases}$$

Thus, it is trivial to see that $D(uuw) \leq L(uuw)$ if $\phi(uw) = c$ or $\phi(uw) = \gamma$. If $\phi(uw) \in L \setminus \{c\}$, then by Observation 4.1 we know that $x_{uw}^{\phi(uw)} \geq \frac{1}{2}$, which implies $D(uuw) \leq 2L(uuw)$.

Now, it remains to consider the general case that u, v, w are distinct vertices. In the following, we give a useful property of the fractional solution to **[CCC-LP]**.

Observation A.1. *Given three distinct vertices $u, v, w \in S_c$ and a color $\bar{c} \in L \setminus \{c\}$, then $x_{uv}^{\bar{c}} \geq \max\{\frac{1}{2}, 1 - x_{uv}^c, 1 - x_{vw}^c, 1 - x_{wu}^c\}$.*

Proof. By the definition of S_c, Observation 4.1 and constraint (2), we have

$$x_{uv}^{\bar{c}} \geq x_u^{\bar{c}} \geq 1 - x_u^c > \frac{1}{2},$$
$$x_{uv}^{\bar{c}} + x_{uv}^c \geq x_u^{\bar{c}} + x_u^c \geq 1,$$
$$x_{uv}^{\bar{c}} + x_{vw}^c \geq x_v^{\bar{c}} + x_v^c \geq 1,$$
$$x_{uv}^{\bar{c}} + x_{wu}^c \geq x_u^{\bar{c}} + x_u^c \geq 1.$$

Rearranging them completes the proof. $\qquad\qquad\square$

For brevity, we use $x_1 = x_{uv}^c, x_2 = x_{vw}^c, x_3 = x_{wu}^c$ and $m = \max\{\frac{1}{2}, 1 - x_1, 1 - x_2, 1 - x_3\}$, and refer to the triangle $\big(\phi(uv), \phi(vw), \phi(wu)\big)$ as (c_1, c_2, c_3) where c_i is one of the colors in $\{c, \bar{c}, \gamma\}$ and $\bar{c}$ is any color in $L \setminus \{c\}$. In what follows, we investigate all combinations of (c_1, c_2, c_3).

A.1 (γ, γ, γ) Triangles

By definition,

$$D(uvw) = (1 - x_1)(1 - x_2) + (1 - x_2)(1 - x_3) + (1 - x_3)(1 - x_1).$$

Meanwhile,

$$\begin{aligned} L(uvw) &= (1 - x_1 x_2) \sum_{i \in L}(1 - x_{wu}^i) + (1 - x_2 x_3) \sum_{i \in L}(1 - x_{uv}^i) + (1 - x_1 x_3) \sum_{i \in L}(1 - x_{vw}^i) \\ &\geq (1 - x_1 x_2)(1 - x_3) + (1 - x_2 x_3)(1 - x_1) + (1 - x_1 x_3)(1 - x_2) \\ &\geq (1 - x_1)(1 - x_3) + (1 - x_2)(1 - x_1) + (1 - x_3)(1 - x_2) \\ &= D(uvw). \end{aligned}$$

14

A.2 (γ, γ, c) Triangles

We have

$$D(uvw) = (1 - x_2)(1 - x_3) + (1 - x_1)(1 - x_3) + (1 - x_1)x_2 + (1 - x_2)x_1,$$

$$\text{and} \quad L(uvw) = (1 - x_2 x_3) \sum_{i \in L} (1 - x_{wu}^i) + (1 - x_1 x_3) \sum_{i \in L} (1 - x_{uv}^i) + (1 - x_1 x_2)x_3$$

$$\geq (1 - x_2 x_3)(1 - x_1) + (1 - x_1 x_3)(1 - x_2) + (1 - x_1 x_2)x_3,$$

$$\geq \tfrac{1}{2}\big((1 - x_3)(1 - x_1) + (1 - x_3)(1 - x_2) + (1 - x_1 + 1 - x_2)(1 + x_3)\big).$$

Thus,

$$2L(uvw) - D(uvw) \geq (1 - x_1 + 1 - x_2) - (1 - x_1)x_2 - (1 - x_2)x_1 = 2(1 - x_1)(1 - x_2) \geq 0.$$

A.3 $(\gamma, \gamma, \bar{c})$ Triangles

We have

$$D(uvw) = (1 - x_2)(1 - x_3) + (1 - x_1)(1 - x_3) + 1 - x_1 x_2,$$

$$\text{and} \quad L(uvw) = (1 - x_2 x_3) \sum_{i \in L} (1 - x_{wu}^i) + (1 - x_1 x_3) \sum_{i \in L} (1 - x_{uv}^i) + (1 - x_1 x_2)x_{wu}^{\bar{c}}$$

$$\geq (1 - x_2 x_3)(1 - x_1) + (1 - x_1 x_3)(1 - x_2) + (1 - x_1 x_2)m$$

$$\geq (1 - x_3)(1 - x_1) + (1 - x_3)(1 - x_2) + \tfrac{1}{2}(1 - x_1 x_2).$$

Thus, $2L(uvw) - D(uvw) \geq 0$.

A.4 (γ, c, c) Triangles

We have

$$D(uvw) = (1 - x_2)(1 - x_3) + (x_1 + x_3 - 2x_1 x_3) + (x_1 + x_2 - 2x_1 x_2)$$

$$= 1 + x_2 x_3 + 2x_1(1 - x_2 - x_3),$$

$$\text{and} \quad L(uvw) = (1 - x_2 x_3) \sum_{i \in L} (1 - x_{wu}^i) + (1 - x_1 x_3)x_2 + (1 - x_1 x_2)x_3$$

$$\geq (1 - x_2 x_3)(1 - x_1) + (1 - x_1 x_3)x_2 + (1 - x_1 x_2)x_3$$

$$= 1 - x_1 + x_2 + x_3 - x_2 x_3 - x_1 x_2 x_3.$$

Thus,

$$\frac{5}{2} \cdot L(uvw) - D(uvw) \geq \frac{3}{2} + \frac{5(x_2 + x_3)}{2} - \frac{7x_2 x_3}{2} - x_1\left(\frac{9}{2} - 2x_2 - 2x_3 + \frac{5x_2 x_3}{2}\right) \triangleq F.$$

Obviously,

$$\frac{\partial F}{\partial x_1} = -\frac{9}{2} + 2x_2 + 2x_3 - \frac{5x_2 x_3}{2} < 0.$$

This implies that F decreases along with x_1. Recall that by constraint (3), we have $x_1 \leq x_2 + x_3$. Thus, when $x_2 + x_3 \geq 1$, F achieves its minimum at $x_1 = 1$ such that

$$F \geq \frac{3}{2} + \frac{5(x_2 + x_3)}{2} - \frac{7x_2 x_3}{2} - \left(\frac{9}{2} - 2x_2 - 2x_3 + \frac{5x_2 x_3}{2}\right)$$

$$= -3 + \frac{9(x_2 + x_3)}{2} - 6x_2 x_3$$

$$= -\frac{3}{2}(x_2 + x_3 - 1)(x_2 + x_3 - 2) + \frac{3}{2}(x_2 - x_3)^2$$

$$\geq 0,$$

where the last inequality is because $1 \leq x_2 + x_3 \leq 2$.

15

On the other hand, when $x_2 + x_3 \leq 1$, F achieves its minimum at $x_1 = x_2 + x_3$ such that

$$F \geq \frac{3}{2} + \frac{5(x_2 + x_3)}{2} - \frac{7x_2x_3}{2} - (x_2 + x_3)\left(\frac{9}{2} - 2x_2 - 2x_3 + \frac{5x_2x_3}{2}\right)$$

$$= \frac{3}{2} - 2(x_2 + x_3) - \frac{7x_2x_3}{2} + 2(x_2 + x_3)^2 - \frac{5x_2x_3(x_2 + x_3)}{2} \triangleq F^*.$$

Due to symmetry, without loss of generality, let $x_2 = y + a$ and $x_3 = y - a$ with $0 \leq a \leq y \leq \frac{1}{2}$. Then, we can rewrite F^* as

$$F^* = \frac{3}{2} - 4y - \frac{7(y^2 - a^2)}{2} + 8y^2 - 5y(y^2 - a^2)$$

$$\geq -5y^3 + \frac{9y^2}{2} - 4y + \frac{3}{2}$$

$$= -5y^2\left(y - \frac{1}{2}\right) + (2y - 3)\left(y - \frac{1}{2}\right) \geq 0.$$

Putting it together yields $\frac{5}{2} \cdot L(uvw) - D(uvw) \geq 0$.

A.5 $(\gamma, c, \bar{c})$ Triangles

We have

$$D(uvw) = (1 - x_2)(1 - x_3) + (x_1 + x_3 - 2x_1x_3) + (1 - x_1x_2)$$
$$= 2 - x_2 + x_2x_3 + x_1(1 - x_2 - 2x_3),$$

and $\quad L(uvw) = (1 - x_2x_3)\sum_{i \in L}(1 - x_{wu}^i) + (1 - x_1x_3)x_2 + (1 - x_1x_2)x_{wu}^{\bar{c}}$

$$\geq (1 - x_2x_3)(1 - x_1) + (1 - x_1x_3)x_2 + (1 - x_1x_2)m$$
$$= 1 - x_1 + x_2 + m - x_2x_3 - x_1x_2m.$$

Thus,

$$\frac{5}{2} \cdot L(uvw) - D(uvw) \geq \frac{1 + 5m}{2} + \frac{7x_2}{2} - \frac{7x_2x_3}{2} - x_1\left(\frac{7}{2} + \frac{(5m - 2)x_2}{2} - 2x_3\right) \triangleq F.$$

Again, when m is independent of x_1 and $m \geq 1/2$, we have

$$\frac{\partial F}{\partial x_1} = -\frac{7}{2} - \frac{(5m - 2)x_2}{2} + 2x_3 < 0.$$

This implies that F decreases along with x_1. When $x_2 + x_3 \geq 1$, F achieves its minimum at $x_1 = 1$ such that

$$F \geq \frac{1 + 5m}{2} + \frac{7x_2}{2} - \frac{7x_2x_3}{2} - \left(\frac{7}{2} + \frac{(5m - 2)x_2}{2} - 2x_3\right)$$

$$= \frac{5m(1 - x_2)}{2} - 3 + \frac{9x_2}{2} + 2x_3 - \frac{7x_2x_3}{2} \triangleq F^*.$$

If $x_2 \geq 1/2$, observing that $5m = m + 4m \geq 1 - x_3 + 4/2 = 3 - x_3$, we have

$$F^* \geq \frac{(3 - x_3)(1 - x_2)}{2} - 3 + \frac{9x_2}{2} + 2x_3 - \frac{7x_2x_3}{2} = \frac{3(2x_2 - 1)(1 - x_3)}{2} \geq 0.$$

If $x_2 < 1/2$, observing that $5m \geq 5(1 - x_2)$, we have

$$F^* \geq \frac{5(1 - x_2)^2}{2} - 3 + \frac{9x_2}{2} + \frac{4 - 7x_2}{2} \cdot (1 - x_2) = \frac{3(2x_2 - 1)^2}{2} \geq 0,$$

where the first inequality is because $4 - 7x_2 \geq 0$ and $x_3 \geq 1 - x_2$.

On the other hand, when $x_2 + x_3 \leq 1$, F achieves its minimum at $x_1 = x_2 + x_3$ such that

$$F \geq \frac{1 + 5m}{2} + \frac{7x_2}{2} - \frac{7x_2x_3}{2} - (x_2 + x_3)\left(\frac{7}{2} + \frac{(5m - 2)x_2}{2} - 2x_3\right)$$

$$= \frac{5m(1 - x_2(x_2 + x_3))}{2} + \frac{1}{2} - \frac{7x_3}{2} - \frac{x_2x_3}{2} + x_2^2 + 2x_3^2 \triangleq F^*.$$

16

If $x_2 \geq 1/2$, observing that $5m \geq 5/2$, we have

$$
\begin{aligned}
F^* &\geq \frac{5(1 - x_2(x_2 + x_3))}{4} + \frac{1}{2} - \frac{7x_3}{2} - \frac{x_2 x_3}{2} + x_2^2 + 2x_3^2 \\
&= \frac{7(1 - x_3)^2 - x_2^2 - 7x_2 x_3 + x_3^2}{4} \\
&\geq \frac{7x_2^2 - x_2^2 - 7x_2 x_3 + x_3^2}{4} \\
&= \frac{(6x_2 - x_3)(x_2 - x_3)}{4} \geq 0,
\end{aligned}
$$

where the second inequality is by $1 - x_3 \geq x_2$, and the last inequality is by $x_3 \leq 1 - x_2 \leq 1/2 \leq x_2$.

If $1/8 \leq x_2 < 1/2$, observing that $5m \geq 5(1 - x_2)$, we have

$$
\begin{aligned}
F^* &\geq \frac{5(1 - x_2)(1 - x_2(x_2 + x_3))}{2} + \frac{1}{2} - \frac{7x_3}{2} - \frac{x_2 x_3}{2} + x_2^2 + 2x_3^2 \\
&\geq \frac{5(1 - x_2)^2}{2} + x_2^2 + 2\left(x_3 - \frac{7}{8}\right)^2 - \frac{x_2 x_3}{2} - \frac{33}{32},
\end{aligned}
$$

which is increasing in x_3 since $x_3 \leq 1 - x_2 \leq 7/8$. Thus,

$$
\begin{aligned}
F^* &\geq \frac{5(1 - x_2)^2}{2} + x_2^2 + 2\left(1 - x_2 - \frac{7}{8}\right)^2 - \frac{x_2(1 - x_2)}{2} - \frac{33}{32} \\
&= \frac{3(2x_2 - 1)^2}{2} \\
&\geq 0.
\end{aligned}
$$

If $x_2 < 1/8$, similarly, we have

$$
\begin{aligned}
F^* &\geq \frac{5(1 - x_2)^2}{2} + x_2^2 + 2\left(x_3 - \frac{7}{8}\right)^2 - \frac{x_2 x_3}{2} - \frac{33}{32} \\
&\geq \frac{5(1 - 1/8)^2}{2} - \frac{1}{16} - \frac{33}{32} \\
&\geq 0.
\end{aligned}
$$

Putting it together yields $\frac{5}{2} \cdot L(uvw) - D(uvw) \geq 0$.

A.6 $(\gamma, \bar{c}, \bar{c})$ Triangles

We have

$$
\begin{aligned}
D(uvw) &= (1 - x_2)(1 - x_3) + (1 - x_1 x_3) + (1 - x_1 x_2), \\
\text{and} \quad L(uvw) &= (1 - x_2 x_3) \sum_{i \in L} (1 - x_{wu}^i) + (1 - x_1 x_3) x_{vw}^{\bar{c}} + (1 - x_1 x_2) x_{wu}^{\bar{c}} \\
&\geq (1 - x_2 x_3)(1 - x_1) + (1 - x_1 x_3)m + (1 - x_1 x_2)m.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\frac{5}{2} \cdot L(uvw) &\geq 2m \cdot \left((1 - x_1 x_3) + (1 - x_1 x_2)\right) + \frac{(1 - x_1 x_3)m}{2} + \frac{(1 - x_1 x_2)m}{2} \\
&\geq \left((1 - x_1 x_3) + (1 - x_1 x_2)\right) + \frac{(1 - x_1 x_3)(1 - x_2)}{2} + \frac{(1 - x_1 x_2)(1 - x_3)}{2} \\
&\geq \left((1 - x_1 x_3) + (1 - x_1 x_2)\right) + \frac{(1 - x_3)(1 - x_2)}{2} + \frac{(1 - x_2)(1 - x_3)}{2} \\
&= D(uvw).
\end{aligned}
$$

17

A.7 (c, c, c) Triangles

We have

$$D(uvw) = (x_2 + x_3 - 2x_2x_3) + (x_1 + x_3 - 2x_1x_3) + (x_1 + x_2 - 2x_1x_2)$$
$$= 2(x_1 + x_2 + x_3 - x_1x_2 - x_2x_3 - x_1x_3),$$
$$\text{and} \quad L(uvw) = (1 - x_2x_3)x_1 + (1 - x_1x_3)x_2 + (1 - x_1x_2)x_3$$
$$= x_1 + x_2 + x_3 - 3x_1x_2x_3.$$

Thus, $2L(uvw) - D(uvw) = 2x_1x_2(1 - x_3) + 2x_2x_3(1 - x_1) + 2x_1x_3(1 - x_2) \geq 0$.

A.8 $(c, c, \bar{c})$ Triangles

We have

$$D(uvw) = (x_2 + x_3 - 2x_2x_3) + (x_1 + x_3 - 2x_1x_3) + (1 - x_1x_2),$$
$$\text{and} \quad L(uvw) = (1 - x_2x_3)x_1 + (1 - x_1x_3)x_2 + (1 - x_1x_2)x_{wu}^{\bar{c}}$$
$$\geq (1 - x_2x_3)x_1 + (1 - x_1x_3)x_2 + \frac{1}{2} \cdot (1 - x_1x_2).$$

Thus,

$$\frac{5}{2} \cdot L(uvw) - D(uvw) \geq \frac{(3 + 4x_3)(x_1 + x_2)}{2} - 2x_3 + \frac{1 - x_1x_2}{4} - 5x_1x_2x_3 \triangleq F.$$

Due to symmetry, without loss of generality, let $x_1 = y + a$ and $x_2 = y - a$ with $0 \leq a \leq y \leq 1$. Then, we have

$$F = -\left(5x_3 + \frac{1}{4}\right)(y^2 - a^2) + (3 + 4x_3)y - 2x_3 + \frac{1}{4} \geq -\left(5x_3 + \frac{1}{4}\right)y^2 + (3 + 4x_3)y - 2x_3 + \frac{1}{4}.$$

It is easy to verify that $-5y^2 + 4y - 2 \leq 0$, which indicates $(-5y^2 + 4y - 2)x_3$ decreases along with x_3. Thus, if $y \geq 1/2$, we can get that

$$F \geq -5x_3y^2 + (3 + 4x_3)y - 2x_3 \geq -5y^2 + 7y - 2 \geq \min\{-5 + 7 - 2, -5/4 + 7/2 - 2\} = 0.$$

Meanwhile, if $y \leq 1/2$, using $x_3 \leq x_1 + x_2 = 2y$ (by constraint 3), we can get that

$$F \geq -10y^3 + \frac{31}{4}y^2 - y + \frac{1}{4} = y^2(-10y + 5) + \left(y - \frac{1}{2}\right)^2 + \frac{7}{4}y^2 \geq 0.$$

Putting it together yields $\frac{5}{2} \cdot L(uvw) - D(uvw) \geq 0$.

A.9 $(c, \bar{c}, \bar{c})$ Triangles

We have

$$D(uvw) = (x_2 + x_3 - 2x_2x_3) + (1 - x_1x_3) + (1 - x_1x_2),$$
$$\text{and} \quad L(uvw) = (1 - x_2x_3)x_1 + (1 - x_1x_3)x_{vw}^{\bar{c}} + (1 - x_1x_2)x_{wu}^{\bar{c}}$$
$$\geq (1 - x_2x_3)x_1 + \frac{1}{2} \cdot (1 - x_1x_3) + \frac{1}{2} \cdot (1 - x_1x_2).$$

Thus,

$$\frac{5}{2} \cdot L(uvw) - D(uvw) \geq \frac{5(1 - x_2x_3)x_1}{2} - (x_2 + x_3 - 2x_2x_3) + \frac{2 - x_1x_2 - x_1x_3}{4} \triangleq F.$$

Observing that F is linear in x_1, we first assume that F decreases along with x_1. Then, as $x_1 \leq 1$, we have

$$F \geq \frac{5(1 - x_2x_3)}{2} - (x_2 + x_3 - 2x_2x_3) + \frac{2 - x_3 - x_2}{4} = \frac{1 - x_2x_3}{2} + \frac{5(2 - x_2 - x_3)}{4} \geq 0.$$

Next, we consider that F increases along with x_1. Due to symmetry, without loss of generality, suppose that $x_2 \geq x_3$. Using $x_1 \geq x_2 - x_3$ (by constraint 3), we have

$$F \geq \frac{5(1 - x_2x_3)(x_2 - x_3)}{2} - (x_2 + x_3 - 2x_2x_3) + \frac{2 - (x_2 + x_3)(x_2 - x_3)}{4} \triangleq F^*.$$

18

Then, we can get that

$$\frac{\partial F^*}{x_2} = \frac{5(1 - 2x_2x_3 + x_3^2)}{2} - (1 - 2x_3) - \frac{x_2}{2} \geq \frac{5x_3^2}{2} - 3x_3 + 1 = \frac{5}{2}\left(x_3 - \frac{3}{5}\right)^2 + \frac{1}{10} \geq 0,$$

where the second inequality is because $x_2 \leq 1$. Thus, F^* increases along with x_2, which gives rise to

$$F^* \geq \frac{5(1 - x_3x_3)(x_3 - x_3)}{2} - (x_3 + x_3 - 2x_3x_3) + \frac{2 - (x_3 + x_3)(x_3 - x_3)}{4} = 2\left(x_3 - \frac{1}{2}\right)^2 \geq 0.$$

Putting it together yields $\frac{5}{2} \cdot L(uvw) - D(uvw) \geq 0$.

A.10 $(\bar{c}, \bar{c}, \bar{c})$ Triangles

We have

$$D(uvw) = (1 - x_2x_3) + (1 - x_1x_3) + (1 - x_1x_2),$$
$$\text{and} \quad L(uvw) = (1 - x_2x_3)x_{uv}^{\bar{c}} + (1 - x_1x_3)x_{vw}^{\bar{c}} + (1 - x_1x_2)x_{wu}^{\bar{c}}$$
$$\geq \frac{1}{2}\big((1 - x_2x_3) + (1 - x_1x_3) + (1 - x_1x_2)\big).$$

Thus, $2L(uvw) - D(uvw) \geq 0$.

This completes the case-by-case analysis for proving $D(uvw) \leq 2.5L(uvw)$.

B Time Complexity Analysis of the **GreedyVote** Algorithm

Note that computing $\frac{|N_{\phi(uv)}(u) \cap N_{\phi(uv)}(v)|}{|N(u) \cup N(v)|}$ for any $uv \in E$ can be implemented in $O(\Delta)$ time given the adjacent vertex lists of u, v and the adjacent matrix of G. Besides, the while-loop in Lines 2–12 iterates for at most $\mathcal{O}(|V|)$ times due to the reason that at least two vertices are removed from V in each iteration. Therefore, Lines 3–4 cost $\mathcal{O}(\Delta|V|)$ running time in total, considering that m is a predefined constant.

In Lines 6–10 of GreedyVote, when a new vertex w is added into S_k, only the vote counts of the vertices in $C = \{v \mid wv \in E \wedge v \notin S_k \cup \{w\}\}$ would be changed. Therefore, we can check each edge in $\{wv \mid v \in C\}$ to determine the next vertex w' to be added into S_k, while determining the new best color of $S_k \cup \{w\}$ (i.e., the most frequent color of the edges with both endpoints in $S_k \cup \{w\}$) at the same time. This implies that the operations in Lines 2–12 except Lines 3–4 can be implemented in $\mathcal{O}(|E|)$ time in total.

Synthesizing the above analysis, we know that the time complexity of the GreedyVote algorithm is $\mathcal{O}(|E| + \Delta|V|)$ for any constant m.

C Derandomization

In this section, we show that our approximation algorithm 1 can be de-randomized using some tricks inspired by Chawla et al. [10]. The resultant deterministic algorithm also guarantees a 2.5-approximation.

It is noted that phase 1 (Lines 2–4) is a deterministic procedure according to Observation 4.1. Lines 5–6 can also be derandomized by a fixed coloring method easily. Therefore, to get a deterministic algorithm, it is sufficient to derandomize the clustering process on subgraph S_c.

According to the definitions of $d(uv \mid w)$ and $lp(uv \mid w)$, we can rewrite them as follows:

$$d(uv \mid w) = \begin{cases} (1 - p_{vw}^c)p_{uw}^c + (1 - p_{uw}^c)p_{vw}^c, & \text{if } \phi(uv) = c, \\ 1 - p_{uw}^c p_{vw}^c, & \text{if } \phi(uv) \in L \setminus \{c\}, \\ (1 - p_{uw}^c)(1 - p_{vw}^c), & \text{if } \phi(uv) = \gamma; \end{cases}$$

and

$$lp(uv \mid w) = \begin{cases} (1 - p_{uw}^c p_{vw}^c)x_{uv}^{\phi(uv)}, & \text{if } \phi(uv) \in L, \\ (1 - p_{uw}^c p_{vw}^c)\sum_{i \in L}(1 - x_{uv}^i), & \text{if } \phi(uv) = \gamma. \end{cases}$$

19
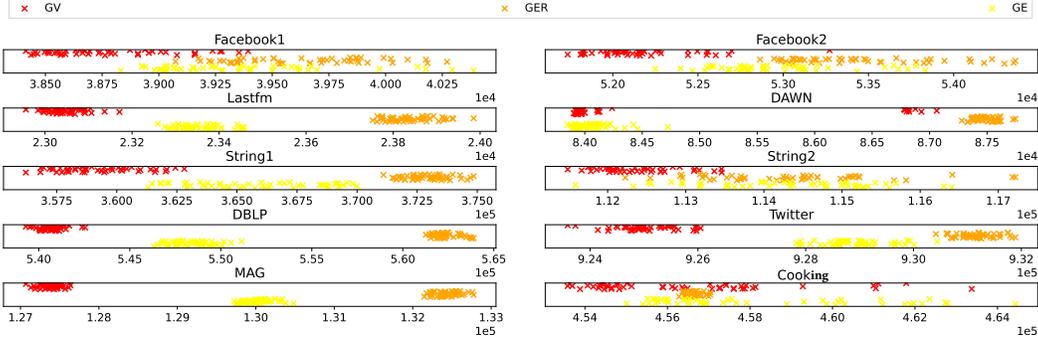
Figure A: The distribution of NOD for GE (yellow), GER (orange), and GV (red)

Here $1 - p_{uv}^c$ denotes the probability of the event that u and v are in the same cluster colored by c, used in line 11. Now we attempt to change the value of each p_{uv}^c to either 0 or 1 to maximize the function:

$$2.5\mathbb{E}[L_t \mid S_c^t] - \mathbb{E}[D_t \mid S_c^t]. \tag{9}$$

This can be done as follows: for each edge $uv \in E_c^t$, we pick a value from $\{0, 1\}$ to assign to p_{uv}^c that maximizes the function (9). After all p_{uv}^c for $uv \in E_c^t$ are integers (e.g. 0 or 1), the algorithm picks the vertex w to maximize $2.5\mathbb{E}[L_t \mid S_c^t, p_t = w] - \mathbb{E}[D_t \mid S_c^t, p_t = w]$ where $p_t = w$ denotes the event that the center of the new cluster at iteration t is vertex w. Algorithm gets the deterministic cluster in iteration t because the center of the cluster is deterministic and the probability of the event that each vertex is add to the cluster is either 0 or 1.

Note that all terms of $L(uvw)$ and $D(uvw)$ are linear functions of p_{uv}^c, p_{vw}^c, and p_{wu}^c according to the definitions. Then, $2.5\mathbb{E}[L_t \mid S_c^t] - \mathbb{E}[D_t \mid S_c^t] = \frac{1}{6|S_c^t|} \sum_{u,v,w \in S_c^t} (2.5L(uvw) - D(uvw))$ are linear in each p_{uv}^c for $uv \in E_c^t$. Therefore, the value of (9) may only increase when we greedily change p_{uv}^c to 0 or 1. Combined with our previous analysis, function (9), which is equal to $\frac{1}{6|S_c^t|} \sum_{w \in S_c^t} \sum_{uv \in E_c^t} (2.5lp(uv \mid w) - d(uv \mid w))$, is nonnegative after each p_{uv}^c is replaced with 0 or 1. Hence, we can find some vertex $w^\star \in S_t^c$ to make $\sum_{uv \in E_t^c} (2.5lp(uv \mid w^\star) - d(uv \mid w^\star)) \geq 0$. It means that the number of disagreements is upper bounded by 2.5 times the LP cost among the edges removed at iteration t, and we get $D_t \leq 2.5L_t$. Therefore, this deterministic algorithm achieves an approximation ratio of 2.5.

D   Addtional Experiments

Table 4: Compare all algorithms on the variance of NOD

Datasets	Pivot [2]	RC [3]	DC [3]	CB [6]	GE [25]	GER [25]	GV (ours)
Facebook1	2.17e7	3.47e6	4.78e6	3.21e6	1.34e5	9.11e4	**7.36e4**
Facebook2	1.48e7	8.68e6	6.92e6	3.75e6	1.34e5	1.76e5	**8.19e4**
Lastfm	7.45e6	9.42e5	1.13e6	6.91e3	2.62e3	2.82e3	**1.69e3**
Twitter	7.83e7	8.07e6	1.69e7	8.55e4	3.19e4	2.21e4	**1.56e4**
DAWN	1.68e9	2.21e8	1.44e8	1.12e6	2.75e4	1.02e4	1.23e6
Cooking	7.05e9	2.06e9	3.63e8	1.81e7	6.50e6	3.16e4	5.38e6
String1	4.54e9	2.49e8	1.03e9	1.23e7	6.81e6	7.79e5	2.61e6
String2	3.17e7	4.27e7	4.87e7	3.90e6	1.12e6	1.11e6	**2.30e5**
DBLP	2.16e10	1.50e8	2.15e8	4.26e6	3.42e5	1.26e5	2.84e5
MAG	3.24e8	1.08e7	9.11e6	2.67e6	1.23e6	4.52e5	**4.31e5**

In this section, we present more experimental results under the same experimental settings in Sec. 6. Each implemented algorithm is run for 50 times for each dataset.

First, we compare the implemented algorithms on their variances, with the results shown by Table 4. As GV, GE, GER are the three algorithms with the best performance on the average Number of Disagreements (NOD), we also plot the distribution of their NOD in a visualized way in Figure A.

20

From Table 4 and Figure A, it can be seen that the variances of GV are smaller than those of GE on all 10 datasets except DAWN. Compared to GER, the variances of GV are smaller on 6 out of 10 datasets, but GV outperforms GER in terms of average NOD on all 10 datasets (see Sec. 2) . Moreover, Figure A shows that GV achieves smaller NOD than GE and GER in *every* of the 50 runs on the four datasets including Lastfm, DBLP, Twitter and MAG. Finally, among all the seven implemented algorithms, GV achieves the smallest variance on 6 out of 10 datasets.

Next, we evaluate the impact of the parameter m of GV on the number of disagreements (NOD) and running time (RT), shown in Figure B. As expected, the running time of GV increases with m, while the NOD of GV decreases when m increases. This reveals the tradeoff on efficiency and effectiveness for selecting the value of m. However, the overall time complexity of GV (i.e., $\mathcal{O}(|E| + \Delta|V|)$) is independent on m.

E   Code

The code of this paper can be found at: https://github.com/xiuq04/heuristics-for-ccc

21
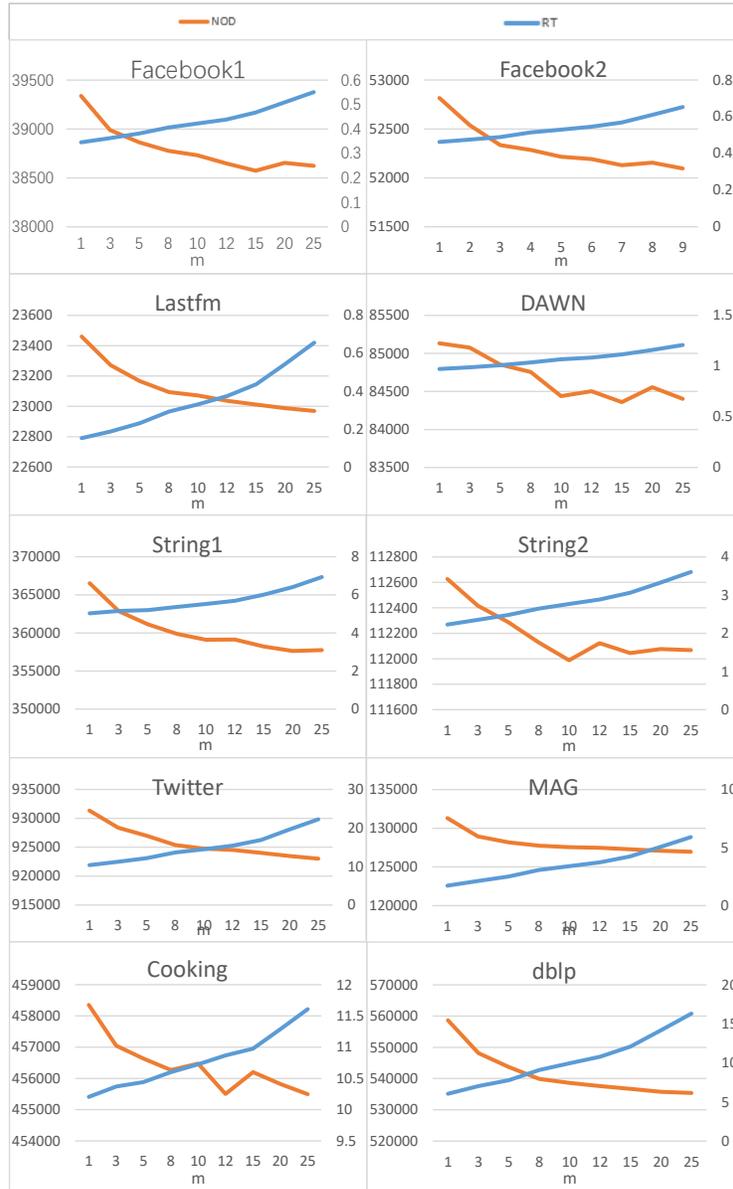
Figure B: Impact of GV's parameter m on NOD (the number of disagreements, plotted by orange lines) and RT (running time, plotted by blue lines) on all 10 datasets. For each dataset, the left Y-axis denotes the values of NOD, and the right Y-axis denotes the values of RT (seconds).

22