E-Companion

The following sections are part of the e-companion.

EC.1. Proof for Theorem 1

THEOREM 1 *For a general graph $G = (V, E)$, unit populations $\{p_i, p_i^A, p_i^B\}_{i \in V}$ and number of districts $K \geq 2$, the Aggregate Competitive Districting Problem (ACDP) with or without the population balance constraints is NP-complete if $K = 2$ and is strongly NP-complete if $K \geq 3$.*

Proof of Theorem 1. The decision version of ACDP is defined as follows. The inputs are a graph $G = (V, E)$, unit populations $\{p_i, p_i^A, p_i^B\}_{i \in V}$, a balance threshold $\tau \geq 0$, number of districts $K \geq 2$, and a target $W \geq 0$. Then, the decision version of ACDP without population balance constraints asks if there exists a feasible district plan $z : V \rightarrow [K]$ such that

$$\max_{k \in [K]} |P_k^A(z) - P_k^B(z)| \leq W, \tag{EC.1}$$

where for each district $k \in [K]$ and party $l \in \{A, B\}$, $P_k^l(z) = \sum\limits_{i \in V : z(i) = k} p_i^l$. Here, a feasible district plan is one that satisfies the MIP constraints (3)-(9). For ACDP with population balance constraints, a feasible district plan also satisfies constraints (2) for a given $\tau \geq 0$. Given a district plan, since feasibility for ACDP with or without population balance constraints can be verified in polynomial time using the constraints (2)-(9) and (EC.1), it is clear that ACDP with or without population balance constraints is in NP.

It is now shown that ACDP is NP-complete for $K = 2$ districts, and is strongly NP-complete for $K \geq 3$ districts. A polynomial-time reduction is shown from the NP-complete 2-*partitioning problem* for $K = 2$ and from the strongly NP-complete 3-*partitioning problem* for $K \geq 3$ (Garey and Johnson 1978). The two reductions are first shown for ACDP without the population balance constraints. Then, it is shown how a τ value can be chosen so that the reductions hold when the population balance constraints are added.

For $K = 2$ without population balance:

The 2$-$partitioning problem (or 2-PART) is defined as follows. For a given set of integers $V' = \{a_1, a_2 \ldots, a_n\}$ and a target $W' \geq 0$, 2-PART divides V' into two partitions $\{S_1, S_2\}$ such that $\max_{k=1,2} |\sum_{i \in S(k)} a_i| \leq W'$. For a given instance of 2-PART (V', W'), an instance of ACDP can be constructed such that solving the ACDP is equivalent to solving 2-PART. The set of units V is chosen to be the indices of the set of integers $[n]$, the number of districts K is equal to 2, and the adjacency graph G is constructed to be a complete graph. For every unit $i \in V$, the unit populations (p_i, p_i^A, p_i^B) are chosen to be $(a_i, a_i, 0)$. The target W is chosen to be W'. To show that the two problems are equivalent, it shown that the constructed instance of ACDP is a "yes" instance if and only if the given 2-PART instance is a "yes" instance.

First, let the instance of 2-PART be a "yes" instance, i.e., there exists a partition $\{S_1, S_2\}$ of V' such that $\max_{k=1,2} |\sum_{i \in S_k} a_i| \leq W'$. Consider a district plan z constructed by assigning units in V to the two districts according to the corresponding assignment of integers in the two sets S_1 and S_2. Then, for every district $k = 1, 2$,

$$|P_k^A(z) - P_k^B(z)| = |\sum_{i \in V: z(i) = k} (p_i^A - p_i^B)| = |\sum_{i \in V'_k} (a_i - 0)| \leq W'.$$

Since $W = W'$, the district plan z satisfies constraint (EC.1). Note that since G is a complete graph, z satisfies the contiguity constraints. Hence, the constructed instance to ACDP is a "yes" instance.

Second, let the constructed instance of ACDP be a "yes" instance, i.e., there exists a district plan z that satisfies constraints (3)-(9) and (EC.1). Consider a partition of V' into 2 partitions $\{S_1, S_2\}$ constructed according to the assignment of the units in $V = V'$ to the $K = 2$ districts. For every $k = 1, 2$, as determined by z, $|\sum_{i \in S(k)} a_i| = |\sum_{i \in V: z(i) = k} (p_i^A - p_i^B)| \leq W$. Since $W = W'$, the partition $\{S_1, S_2\}$ satisfies $\max_{k=1,2} |\sum_{i \in S(k)} a_i| \leq W'$. Hence, the instance of 2-PART is a "yes" instance.

For $K \geq 3$ without population balance:

The 3−partitioning problem (or 3-PART) is defined as follows. For a given set of integers $V' = \{a_1, a_2 \ldots, a_n\}$ and a target $W' \geq 0$, 3-PART divides V' into three partitions $\{S_1, S_2, S_3\}$ such that $\max_{k=1,2,3} |\sum_{i \in S(k)} a_i| \leq W'$. For a given instance of 3-PART (V', W'), an instance of ACDP can be constructed such that solving the ACDP is equivalent to solving 3-PART. The number of districts is chosen to be $K \geq 3$. The set of units V is chosen to be the indices of the set of integers in V' padded with $K - 3$ additional units, i.e., $V = [n] \cup \{n+1, n+2, \ldots, n+K-3\}$. The adjacency graph G is constructed to be a complete graph. For every unit $i \in V$, the unit populations (p_i, p_i^A, p_i^B) are chosen to be $(a_i, a_i, 0)$ if $i \in [n]$, or $(W', W', 0)$ if $n + 1 \leq i \leq n + K - 3$. The target W is chosen to be W'. To show that the two problems are equivalent, it is to be shown that an instance of ACDP is a "yes" instance if and only if the constructed 3-PART instance is a "yes" instance.

First, let the instance of 3-PART be a "yes" instance, i.e., there exists a partition $\{S_1, S_2, S_3\}$ of V' such that $\max_{k=1,2,3} |\sum_{i \in S(k)} a_i| \leq W'$. Consider a district plan z constructed by assigning units in $[n]$ to districts 1, 2 and 3 according to the corresponding assignment of integers in the three sets S_1, S_2, S_3. Each unit in $\{n+1, n+2, \ldots, n+K-3\}$ is assigned to occupy a single district among the remaining districts in $[K] \backslash [3]$. Then, for every district $k = 1, 2, 3$,

$$|P_k^A(z) - P_k^B(z)| = |\sum_{i \in V: z(i) = k} (p_i^A - p_i^B)| = |\sum_{i \in V'_k} (a_i - 0)| \leq W'.$$

For every district $k \in K \backslash [3]$,

$$|P_k^A(z) - P_k^B(z)| = |\sum_{i \in V: z(i)=k} (p_i^A - p_i^B)| = |(W' - 0)| = W'.$$

Since $W = W'$, the district plan z satisfies constraint (EC.1). Note that since G is a complete graph, z satisfies the contiguity constraints. Hence, the constructed instance to ACDP is a "yes" instance.

Second, let the instance of ACDP be a "yes" instance, i.e., there exists a district plan z that satisfies constraints (3)-(9) and (EC.1). Note that each of the $K-3$ padded units will have to be assigned to its own district in order to satisfy constraint (EC.1). These units will occupy $K-3$ of the K districts. A "yes" to 3-PART can be constructed using the units in $[n]$ to the remaining three districts; suppose, without loss of generality, that the three remaining districts correspond to districts $k = 1,2,3$. Consider a partition of V' into 3 partitions, $\{S_1, S_2, S_3\}$, constructed according to the assignment of the units in $[n]$ to the three districts. For every $k = 1,2,3$, $|\sum_{i \in S(k)} a_i| = |\sum_{i \in V: z(i)=k} (p_i^A - p_i^B)| \leq W$. Since $W = W'$, the partition $\{S_1, S_2, S_3\}$ satisfies $\max_{k=1,2,3} |\sum_{i \in S(k)} a_i| \leq W'$. Hence, the instance of 3-PART is a "yes" instance.

Inclusion of population balance constraints:

The population balance constraints (2) require that the population in every district is in the range $[\overline{P}(1-\tau), \overline{P}(1+\tau)]$, where $\overline{P} := (\sum_{i \in V} p_i)/K$. When these constraints are included in ACDP, it is needed to show that an instance of ACDP can be constructed for some $\tau \geq 0$ such that the two reductions in this proof hold true. Let $\tau = K-1$. Then, the range of allowable district populations is $[\overline{P}(2-K), \overline{P}K]$ which is equivalent to $[0, \sum_{i \in V} p_i]$ for $K \geq 2$ since unit populations are non-negative. Then, every district plan will satisfy the population balance constraints, and the two reductions in this proof will continue to hold true. □

EC.2. Proof for Proposition 1

PROPOSITION 1 *Symmetric-DP can be formulated as a Mixed Integer Linear Program (MILP).*

Proof of Proposition 1. To construct an MILP formulation for Symmetric-DP, each of the three constraints (23) - (25) are linearized one at a time in the following three parts.

(a) First, constraints (23) are linearized for given values of $\{x_{ij}\}_{i,j \in V}$, the decision variables that assign units to district centers. For each unit $i \in V$, the following set of constraints define ω_i to be the vote-share for party A in a district with center i; 0 if i is not a district center. Constraints (23) are given by,

$$\omega_i = \begin{cases} \dfrac{\sum_{j \in V} p_j^A x_{ij}}{\sum_{j \in V} (p_j^A + p_j^B) x_{ij}}, & \text{if } x_{ii} = 1 \\ 0, & \text{otherwise} \end{cases} \qquad \forall\, i \in V,$$

The following additional variables are defined.

$$\beta_{ij} = \begin{cases} \omega_i, \text{ if } i \in V \text{ is a district center, and } j \in V \text{ is assigned to that district.} \\ 0, \text{ otherwise }. \end{cases}$$

The first claim is that the following set of constraints linearize constraints (23).

$$\omega_i \leq x_{ii} \quad \forall\, i \in V, \tag{EC.2}$$

$$\beta_{ij} \leq \omega_i \quad \forall\, i,j \in V, \tag{EC.3}$$

$$\omega_i - 1 + x_{ij} \leq \beta_{ij} \quad \forall\, i,j \in V, \tag{EC.4}$$

$$\beta_{ij} \leq x_{ij} \quad \forall\, i,j \in V, \tag{EC.5}$$

$$\sum_{j \in V} (p_j^A + p_j^B)\beta_{ij} = \sum_{j \in V} p_j^A x_{ij} \quad \forall\, i \in V, \tag{EC.6}$$

$$\beta_{ij}, \omega_i \geq 0 \quad \forall\, i,j \in V. \tag{EC.7}$$

For every $i \in V$, constraint (EC.2) ensures that if i is not a district center, then $\omega_i = 0$. For every $i, j \in V$, constraints (EC.3)- (EC.5) together establish the relationship $\beta_{ij} = \omega_i \cdot x_{ij}$. When $x_{ij} = 0$, constraint (EC.5) ensures that $\beta_{ij} = 0$, and when $x_{ij} = 1$, constraints (EC.3) and (EC.4) ensure that $\beta_{ij} = \omega_i$. Given that $\beta_{ij} = \omega_i \cdot x_{ij}$, it can be seen that constraint (EC.6) is equivalent to the non-linear constraint (23). Constraints (EC.7) establish the non-negativity of the variables.

(b) Second, constraints (24) are linearized. Here, the values for $\{\omega_i\}_{i \in V}$ defined by constraints (EC.2)-(EC.7) are used to define α_k, the $k-$th largest district vote-share for $k \in [K]$ among the set of all district vote-shares. Constraints (24) are given by,

$$\alpha_k = k\max_{i \in V}\{\omega_i\} \quad \forall\, k \in [K],$$

where the $k\max$ function returns the $k-$th largest value of a given set. To linearize this constraint, for each $i \in V$ and $k \in [K]$, additional variables $\delta_{i,k}$ and $\gamma_{i,k}$ are now defined. Some additional notation is first introduced for each $k \in [K]$. Let $\omega_{(k)}$ denote the value of the $k-$th largest element in the set $\{\omega_i\}_{i \in V}$. For each $k \in [K]$, let $S^+(k), S(k), S(k)^- \subseteq V$ be subsets of units that partition V which are defined as $S^+(k) \coloneqq \{i \in V : \omega_i > \omega_{(k)}\}$, $S(k) \coloneqq \{i \in V : \omega_i = \omega_{(k)}\}$, and $S(k)^- \coloneqq \{i \in V : \omega_i < \omega_{(k)}\}$. It can be observed that $|S^+(k)| < k$ and $|S^-(k)| < |V| - k + 1$ since $|S(k)| \geq 1$ by definition. Among the units in $S(k) \cup S^+(k)$, choose an arbitrary subset of units $S'(k) \subseteq S(k) \cup S^+(k)$ such that $|S'(k)| = k$. Such an $S'(k)$ is possible since $|S(k) \cup S^+(k)| = |V| - |S^-(k)| > |V| - (|V| - k + 1) = k - 1$, and hence $|S(k) \cup S^+(k)| \geq k$ since the size of the set is an integer. Further, among the units in $S(k) \cup S^-(k)$, choose an arbitrary subset of units $S''(k) \subseteq S(k) \cup S^-(k)$ such that $|S''(k)| = |V| - k + 1$. Such an $S''(k)$

is possible since $|S(k) \cup S^-(k)| = |V| - |S^+(k)| > |V| - k$, and hence $|S(k) \cup S^-(k)| \geq |V| - k + 1$ since the size of the set is an integer. Based on the constructed sets $S'(k)$ and $S''(k)$, the $\delta_{i,k}$ and $\gamma_{i,k}$ values are defined as follows.

$$\delta_{i,k} = \begin{cases} 1, \text{ if } i \in S'(k), \\ 0, \text{ otherwise.} \end{cases} \tag{EC.8}$$

$$\gamma_{i,k} = \begin{cases} 1, \text{ if } i \in S''(k), \\ 0, \text{ otherwise.} \end{cases} \tag{EC.9}$$

For each $k \in [K]$, ensuring that α_k is the $k-$th largest element in the set $\{\omega_i\}_{i \in V}$ is equivalent to ensuring that α_k is less than or equal to at least k elements in the set, and that α_k is greater than or equal to at least $|V| - k + 1$ elements in the set. The second claim in this proof is that the following set of constraints linearize constraints (24).

$$\alpha_k \leq \omega_i + 1 - \delta_{i,k} \quad \forall\ i \in V,\ k \in [K] \tag{EC.10}$$

$$\sum_{i \in V} \delta_{i,k} = k \quad \forall\ k \in [K] \tag{EC.11}$$

$$\omega_i - 1 + \gamma_{i,k} \leq \alpha_k \quad \forall\ i \in V,\ k \in [K] \tag{EC.12}$$

$$\sum_{i \in V} \gamma_{i,k} = |V| - k + 1 \quad \forall\ k \in [K] \tag{EC.13}$$

$$\gamma_{i,k}, \delta_{i,k} \in \{0,1\} \quad \forall\ i \in V,\ k \in [K]. \tag{EC.14}$$

For each $k \in [K]$, constraints (EC.10) ensure that for every $i \in V$, $\delta_{i,k} = 1 \Rightarrow \alpha_k \leq \omega_i$. Constraints (EC.11) ensure that $\delta_{i,k} = 1$ for exactly k elements in $\{\omega_i\}_{i \in V}$. Constraints (EC.12) ensure that for $i \in V$, $\gamma_{i,k} = 1 \Rightarrow \alpha_k \geq \omega_i$. Constraints (EC.13) ensure that $\gamma_{i,k} = 1$ for exactly $|V| - k + 1$ elements in $\{\omega_i\}_{i \in V}$. Constraints (EC.14) define the binary nature of the variables. This linearization is similar to the constraints provided by Rubin (2015), where the goal is to minimize the sum of the k largest elements in a vector.

The correctness of constraints (EC.10)-(EC.14) is now shown. It is required to show that a solution satisfies these constraints if and only if it defines α_k for all $k \in [K]$ to be the $k-$th largest element in the set $\{\omega_i\}_{i \in V}$.

For the sufficiency condition, consider a solution $(\alpha_k, \delta_{i,k}, \gamma_{i,k})_{i \in V, k \in [K]}$ that satisfies constraints (EC.10)-(EC.14). It is required to show that $\alpha_k = \omega_{(k)}$ for all $k \in [K]$. Suppose for the purpose of contradiction that $\alpha_{k'} \neq \omega_{(k')}$ for some $k' \in [K]$. Then, one of the following two cases holds for that k':

- *Case 1:* $\alpha_{k'} < \omega_{(k')}$. Any $i \in V$ with $\gamma_{i,k'} = 1$ must have $i \in S^-_{k'}$, since it must have $\omega_i \leq \alpha_{k'} < \omega_{(k')}$. Hence, $\{i \in V \mid \gamma_{i,k'} = 1\} \subseteq S^-_{k'}$ and it can be shown that $\sum_{i \in V} \gamma_{i,k'} \leq \sum_{i \in S^-_{k'}} 1 = |S^-_{k'}| < |V| - k' + 1$, which provides a contradiction with constraint (EC.13).

- *Case 2:* $\alpha_{k'} > \omega_{(k')}$. Any $i \in V$ with $\delta_{i,k'} = 1$ must have $i \in S^+_{k'}$, since it must have $\omega_i \geq \alpha_{k'} > \omega_{(k')}$. Hence, $\{i \in V | \ \delta_{i,k'} = 1\} \subseteq S^+_{k'}$ and it can be shown that $\sum_{i \in V} \delta_{i,k'} \leq \sum_{i \in S^+_{k'}} 1 = |S^+_{k'}| < k'$, which provides a contradiction with constraint (EC.11).

Hence, either case contradicts the assumption that $\alpha_{k'} \neq \omega_{(k')}$ for some $k' \in [K]$. Therefore, it can be concluded that $\alpha_k = \omega_{(k)}$ for all $k \in [K]$.

For the necessary condition, for each $k \in [K]$, it is required to show that a solution $(\alpha_k, \delta_{i,k}, \gamma_{i,k})_{i \in V}$ that defines α_k to be the $k-$th largest element in $\{\omega_i\}_{i \in V}$ will satisfy satisfies constraints (EC.10)-(EC.14). In such a solution, $\alpha_k = \omega_{(k)}$ by definition. The $\delta_{i,k}$ and $\gamma_{i,k}$ values for each $i \in V$ are derived from (EC.8) and (EC.9). First, consider constraint (EC.10). For all $i \in S'(k)$, since $\delta_{i,k} = 1$ and $\alpha_k \leq \omega_i$ by the definitions of $S(k)$ and $S^+(k)$ (whose union is a superset of $S'(k)$), this constraint is satisfied. For all $i \in V \backslash S'(k)$, since $\delta_{i,k} = 0$ and $\alpha_k \leq 1$ by constraint (27), this constraint is satisfied. Hence, this solution satisfies constraint (EC.10). Further, since

$$\sum_{i \in V} \delta_{i,k} = \sum_{i \in S'(k)} \delta_{i,k} + \sum_{i \in V \backslash S'(k)} \delta_{i,k} \tag{EC.15}$$

$$= k + 0 = k, \tag{EC.16}$$

this solution satisfies constraint (EC.11). Similarly, using the $\{\gamma_{i,k}\}_{i \in V}$ values from (EC.9), it can be shown that the solution satisfies constraints (EC.12) and (EC.13). The proof proceeds similar to the proof for the solution satisfying constraints (EC.10) and (EC.11), and is hence omitted. Finally, the $\{\delta_{i,k}, \gamma_{i,k}\}_{i \in V}$ values satisfy constraint (EC.14) by definition.

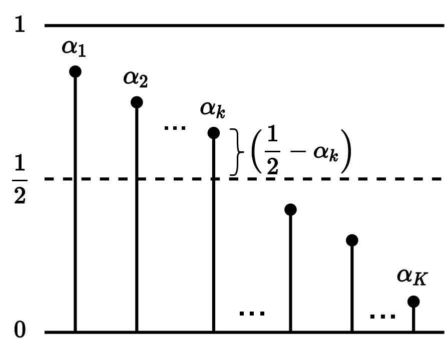

Figure EC.1 District votes-shares in K districts sorted in a non-increasing order, $\{\alpha_k\}_{k \in [K]}$.

(c) Third, constraints (25) are linearized. Here, for each $k \in [K]$, the value of α_k defined by constraints (EC.10)-(EC.14) is used to define μ_{km} to be the minimum average vote-share for A to win exactly k districts. To achieve this, the vote-share α_m in each district $m \in [K]$ is decreased by a value of $\alpha_k - \frac{1}{2}$, which is the fraction of voters needed to just flip district k;

note that this will increase the vote-share if $\alpha_k < 1/2$. The intuition behind this is illustrated in Figure EC.1. Constraints (25) are given by,

$$\mu_{km} = \begin{cases} 0, \text{ if } \alpha_m + \frac{1}{2} - \alpha_k \leq 0 \\ \alpha_m + \frac{1}{2} - \alpha_k, \text{ if } \alpha_m + \frac{1}{2} - \alpha_k \in (0,1) \\ 1, \text{ if } \alpha_m + \frac{1}{2} - \alpha_k \geq 1 \end{cases} \quad \forall\ k, m \in [K].$$

It can be seen that for each $k, m \in [K]$, μ_{km} is a non-convex piecewise linear function of $(\alpha_m + 1/2 - \alpha_k)$ with 2 breakpoints to ensure that μ_{km} does not leave the $[0,1]$ interval. Linearizing a piecewise linear function in an MIP has been well studied (Grimstad and Knudsen 2020), and this paper presents a concise version of the standard big-M constraints. The concision comes from using only 2 binary variables for each $k, m \in [K]$ instead of the standard 3 variables needed when the piecewise linear function has 2 breakpoints. See Vielma et al. (2010) for convex combination models to linearize non-convex piecewise linear functions.

The linearization is now described. For each $k, m \in [K]$, additional binary variables Ω_{km} and Ω'_{km} are defined as follows

$$\Omega_{km} = \begin{cases} 1, \text{ if } \alpha_m + \frac{1}{2} - \alpha_k > 0. \\ 0, \text{ if } \alpha_m + \frac{1}{2} - \alpha_k < 0. \end{cases}$$

$$\Omega'_{km} = \begin{cases} 1, \text{ if } \alpha_m + \frac{1}{2} - \alpha_k < 1. \\ 0, \text{ if } \alpha_m + \frac{1}{2} - \alpha_k > 1. \end{cases}$$

Here, Ω_{km} (Ω'_{km}) arbitrarily takes the value of either 0 or 1 if $\alpha_m + \frac{1}{2} - \alpha_k = 0$ (1). The first idea in the linearization is to use big-M constraints to ensure that Ω_{km} and Ω'_{km} are defined with respect to the value of $\left(\alpha_m + \frac{1}{2} - \alpha_k\right)$. The second idea is to define μ_{km} using Ω_{km}, Ω'_{km} and $\left(\alpha_m + \frac{1}{2} - \alpha_k\right)$. Note that for every $k, m \in [K]$, the expression $\left(\alpha_m + \frac{1}{2} - \alpha_k\right)$ ranges between $-\frac{1}{2}$ and $\frac{3}{2}$. The third claim is that the following set of constraints linearize constraints (25). The linearized constraints are given by,

$$\alpha_m + \frac{1}{2} - \alpha_k \leq \frac{3}{2}\Omega_{km} \quad \forall\ k, m \in [K], \tag{EC.17}$$

$$\alpha_m + \frac{1}{2} - \alpha_k \geq \mu_{km} + \frac{3}{2}(\Omega_{km} - 1) \quad \forall\ k, m \in [K], \tag{EC.18}$$

$$\alpha_m + \frac{1}{2} - \alpha_k \geq -\frac{3}{2}\Omega'_{km} + 1 \quad \forall\ k, m \in [K], \tag{EC.19}$$

$$\alpha_m + \frac{1}{2} - \alpha_k \leq \frac{3}{2}(1 - \Omega'_{km}) + \mu_{km} \quad \forall\ k, m \in [K], \tag{EC.20}$$

$$-\Omega'_{km} + 1 \leq \mu_{km} \leq \Omega_{km} \quad \forall\ k, m \in [K], \tag{EC.21}$$

$$\Omega_{km}, \Omega'_{km} \in \{0, 1\} \quad k, m \in [K]. \tag{EC.22}$$

For each $k, m \in [K]$, constraint (EC.22) establishes the binary domain of both Ω_{km} and Ω'_{km}. To prove that constraints (EC.17) - (EC.22) linearize constraint (25), there are two parts to the proof. First, it is shown that for a given value of $(\alpha_m + \frac{1}{2} - \alpha_k)$, $\mu_{k,m}$ is correctly defined through the variables $\Omega_{k,m}$ and $\Omega'_{k,m}$. Consider the following five cases for $(\alpha_m + \frac{1}{2} - \alpha_k)$:

- *Case 1*: when $\alpha_m + \frac{1}{2} - \alpha_k < 0$, then $\Omega_{km} = 0$ by combining constraints (EC.18) and (27), and $\Omega'_{km} = 1$ by constraint (EC.19). Substituting these values of Ω_{km} and Ω'_{km} into constraint (EC.21) gives $\mu_{km} = 0$. These values also satisfy constraints (EC.17) and (EC.20).
- *Case 2*: when $\alpha_m + \frac{1}{2} - \alpha_k = 0$, then $\Omega'_{km} = 1$ by constraint (EC.19). There are two subcases based on the value of Ω_{km}:
  (a) if $\Omega_{km} = 1$, then $\mu_{km} = 0$ by combining constraints (EC.18) and (27). Constraints (EC.17), (EC.20), and (EC.21) are also satisfied.
  (b) if $\Omega_{km} = 0$, then $\mu_{km} = 0$ by constraint (EC.21). Constraints (EC.17), (EC.18), and (EC.20) are also satisfied.
- *Case 3*: when $\alpha_m + \frac{1}{2} - \alpha_k \in (0, 1)$, then $\Omega_{km} = 1$ and $\Omega'_{km} = 1$ by constraints (EC.17) and (EC.19), respectively. Substituting these values of Ω_{km} and Ω'_{km} into constraints (EC.18) and (EC.20) gives $\mu_{km} = \alpha_m + \frac{1}{2} - \alpha_k$. These values also satisfy constraint (EC.21).
- *Case 4*: when $\alpha_m + \frac{1}{2} - \alpha_k = 1$, then $\Omega_{km} = 1$ by constraint (EC.17). There are two subcases based on the value of Ω'_{km}:
  (a) if $\Omega'_{km} = 1$, then $\mu_{km} = 1$ by combining constraints (EC.20) and (27). Constraints (EC.18), (EC.19), and (EC.21) are also satisfied.
  (b) if $\Omega'_{km} = 0$, then $\mu_{km} = 1$ by constraint (EC.21). Constraints (EC.18), (EC.19), and (EC.20) are also satisfied.
- *Case 5*: when $\alpha_m + \frac{1}{2} - \alpha_k$ ¿ 1, then $\Omega_{km} = 1$ by constraint (EC.17) and $\Omega'_{km} = 0$ by combining constraints (EC.20) and (27). Substituting these values of Ω_{km} and Ω'_{km} into constraint (EC.21) gives $\mu_{km} = 1$. These values also satisfy constraints (EC.18) and (EC.19).

Therefore, regardless of the value of $\alpha_m + \frac{1}{2} - \alpha_k$, the value of μ_{km} is determined as defined in constraint (25).

Second, it is shown that for given values of Ω_{km} and Ω'_{km}, the values of μ_{km} and $(\alpha_m + \frac{1}{2} - \alpha_k)$ are correctly determined. There are 4 cases:

- *Case 1:* $\Omega_{km} = \Omega'_{km} = 0$ is not possible by constraint (EC.21).
- *Case 2:* when $\Omega_{km} = 0$ and $\Omega'_{km} = 1$, then $\mu_{km} = 0$ by (EC.21), and $(\alpha_m + \frac{1}{2} - \alpha_k) \leq 0$ by constraint (EC.17) (Other constraints are also satisfied by noting that $\alpha_m + \frac{1}{2} - \alpha_k \geq -1/2$).
- *Case 3:* when $\Omega_{km} = 1$ and $\Omega'_{km} = 0$, then $\mu_{km} = 1$ by (EC.21), and $(\alpha_m + \frac{1}{2} - \alpha_k) \geq 1$ by constraint (EC.19) (Other constraints are also satisfied by noting that $\alpha_m + \frac{1}{2} - \alpha_k \leq 3/2$).
- *Case 4:* when $\Omega_{km} = \Omega'_{km} = 1$, then $\mu_{km} = \alpha_m + \frac{1}{2} - \alpha_k$ by combining constraints (EC.18) and (EC.20). Other constraints are also satisfied. $\square$

EC.3. Approximating the Minmax Maximal Matching Problem (MLP)

This section presents an approximation result for the *minmax maximal matching problem* (MLP). Given a connected undirected graph $G = (V, E)$ and edge weights u_e for all $e \in E$, MLP finds a maximal matching M in G that minimizes the maximum edge weight $\max_{e\in M} u_e$. MLP is NP-complete via a reduction from the dominating set problem (Lavrov 2019). This paper uses a greedy algorithm to find a feasible solution to MLP; this matching can then be used to merge in the coarsening procedure in Section 4.1.2. For this greedy algorithm, the edge weights for the population-based coarsening is given by $u_{(i,j)} = p_i + p_j$ for every edge $(i, j) \in E$, where $p_i \geq 0$ is the population in every unit $i \in V$.

This section shows that any arbitrary maximal matching (and by extension, the solution found by the greedy algorithm in Section 4.1.2) has an approximation factor that depends on the relative populations of neighboring units in G. To formalize this, let $\rho := \max_{(i,j)\in E} \max\{p_i/p_j, p_j/p_i\}$ be the largest ratio of populations among neighboring units in G. Clearly, $\rho \geq 1$. In defining ρ, a key assumption is that unit populations are strictly positive. Theorem 2 shows that the maximum edge weight (u') in any arbitrary maximal matching is at the most ρ times the maximum edge weight (u^*) in an optimal maximal matching to MLP. In the special case when all the unit populations are equal, this result is trivially true since all the edge weights are equal, where $\rho = 1$. In this case, every maximal matching is optimal to MLP. Theorem 2 derives an approximation ratio for the general case for any $\rho \geq 1$.

Theorem 2 *Consider a connected graph $G = (V, E)$, positive unit populations $p_i > 0$ for every unit $i \in V$, and edge weights $u_{ij} = p_i + p_j$ for every edge $(i, j) \in E$. Let M^* be an optimal solution to the minmax maximal matching problem with inputs $(G, \{u_{ij}\}_{(i,j)\in E})$, and let $u^* := \max_{e\in M^*} u_e$ be its maximum edge weight. Let M be an arbitrary maximal matching with maximum edge weight $u' := \max_{e\in M} u_e$. Then, $u'/\rho \leq u^* \leq u'$, where $\rho := \max_{(i,j)\in E} \max\{p_i/p_j, p_j/p_i\}$.*

Proof of Theorem 2. Since M is a maximal matching (i.e., M is a feasible solution to the minmax maximal matching problem), $u^* \leq u'$. The proof for $u^* \geq u'/\rho$ proceeds by contradiction. Assume the contrary, that $u^* < u'/\rho$. Let $e' = (i', j')$ be an edge with maximum weight in M, i.e., $e' := \arg\max_{e\in M} u_e$ and $u_{e'} = u'$. There are two cases:

1. *Case 1:* edge e' is a part of M^*. In this case, $u' \leq u^*$ since u^* is the maximum edge weight in M^*. This is contrary to the assumption that $u^* < u'/\rho$ since $\rho \geq 1$.
2. *Case 2:* edge e' is not a part of M^*. Since M^* is a maximal matching, at least one of i' or j' is a part of a matched edge in M^* that is not e'. Without loss of generality, select i' as a unit

that is part of a matched edge in M^*. Let that edge be (i', j'') for some $j'' \in V$, and its weight is given by,

$$u_{(i',j'')} = p_{i'} + p_{j''} \geq p_{i'} + \frac{p_{i'}}{\rho} = p_{i'} \frac{(1+\rho)}{\rho}. \tag{EC.23}$$

Since (i', j'') is a part of M^*, and u^* is the maximum edge weight in M^*,

$$u_{(i',j'')} \leq u^*. \tag{EC.24}$$

From (EC.23) and (EC.24),

$$p_{i'} \frac{(1+\rho)}{\rho} \leq u^*. \tag{EC.25}$$

On the other hand, the weight of edge (i', j') is given by,

$$u' = p_{i'} + p_{j'} \leq p_{i'} + \rho p_{i'} = (1+\rho) p_{i'}. \tag{EC.26}$$

From (EC.25) and (EC.26), it is evident that $u' \leq \rho u^*$. This is contrary to the assumption that $u^* < u'/\rho$. $\square$

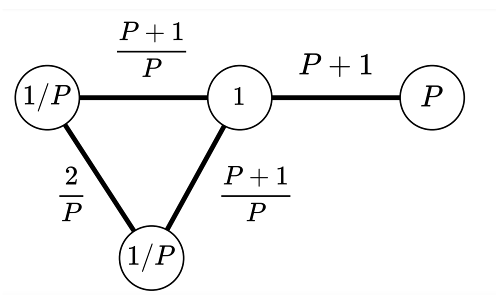

Figure EC.2 An example instance of MLP illustrating that Theorem 2 provides a tight approximation; a graph with four units with respective unit populations $1/P$, $1/P$, 1 and P for some $P > 1$. The unit populations and edge weights are depicted at the units and edges, respectively.

It can be shown that the greedy algorithm in Section 4.1.2 provides a tight approximation, i.e, there exists an instance of MLP such that $u'/u^* = \rho$, where u' and u^* are the greedy and optimal solutions, respectively.

Example EC.1. Consider an instance of MLP with four units and four edges as depicted in Figure EC.2. The respective unit populations are $1/P$, $1/P$, 1 and P for some $P > 1$. The maximum population ratio among neighboring units is $\rho = P$.

For the instance in Example 1, the optimal maximal matching to MLP selects exactly one edge whose edge weight is $u^* = (P+1)/P$. On the other hand, the greedy algorithm in Section 4.1.2 first sorts the edges in the non-decreasing order: $\{2/P, (P+1)/P, (P+1)/P, P+1\}$. The maximal matching returned by the greedy algorithm selects the first and the fourth edges with respective weights $2/P$ and $P+1$. The maximum edge weight for the greedy solution is $u' = P+1$. Hence, the approximation ratio is $u'/u^* = (P+1)/((P+1)/P) = P$, which is equal to ρ. Note that if the unit populations are restricted to integers, setting P to be an integer and multiplying all the unit populations in this example by P would yield an instance with integral unit populations, and the tightness result still holds.

EC.4. Computational Results for Pareto-Frontier

EC.4.1. Efficiency Gap and Compactness

Coarsening	ϕ_{EG}	ϕ_{comp} ($\times 10^9$)	Clock time (s)	CPU time (ticks)	# B&C nodes	Opt. gap ($\frac{UB-LB}{UB}$)
ML	0.1561	2.683	416	180,309	112	-
	0.1511	2.683	784	290,876	199	-
	0.0277	2.686	1,161	467,881	987	-
	0.0272	2.688	28,927	9,402,030	7,091	-
	0.0257	2.688	8,480	3,010,334	3,187	-
	0.0234	2.689	1,695	638,511	1,756	-
	0.0224	2.690	1,625	599,558	2,347	-
	0.0220	2.691	12,849	4,160,289	4,860	-
	0.0217	2.695	81,616	24,191,404	16,204	-
	0.0215	2.695	52,968	15,764,469	12,876	-
	0.0198	2.697	13,370	3,801,803	10,131	-
MM	0.0307	3.923	295	134,708	123	-
	0.0290	3.932	7,584	2,452,971	4,200	-
	0.0286	3.937	10,582	3,517,502	5,491	-
	0.0283	3.943	9,689	2,610,137	3,907	-
	0.0281	3.952	23,973	7,019,182	14,997	-
	0.0274	3.954	15,944	5,100,907	7,839	-
	0.0268	3.961	14,622	4,349,295	8,056	-
	0.0255	3.966	3,289	1,093,726	6,296	-
	0.0252	3.973	45,793	14,142,057	23,586	-
	0.0251	3.979	32,342	9,499,810	16,455	-
	0.0250	3.982	5,678	1,828,732	11,261	-
	0.0243	3.983	48,467	15,287,890	32,957	-
	0.0235	3.989	34,041	10,235,375	22,377	-
	0.0190	4.003	39,687	12,684,359	44,313	-

Table EC.1 Computational results from the $\epsilon-$constraint method to obtain the Pareto-frontier between the efficiency gap and compactness, when coarsening using maximum matchings (MM) and maximal matchings (ML). An optimality gap of "-" indicates that the solution is optimal, i.e., the upper bound (UB) and lower bound (LB) on the compactness objective found in the branch-and-cut algorithm are equal.

EC.4.2. Partisan Asymmetry and Compactness

Coarsening	ϕ_{PA}	$\phi_{comp}(\times 10^9)$	Clock time (s)	CPU time (ticks)	# B&C nodes	Opt. gap ($\frac{UB-LB}{UB}$)
ML	0.0633	3.166	21,601*	5,457,399	4,285	0.154
	0.0127	3.277	21,600	6,187,473	1,554	0.184
	0.0020	3.535	21,600	6,141,634	6,796	0.243
	0.0016	3.929	21,600	5,934,803	3,484	0.319
MM	0.0034	4.605	21,608*	6,102,360	14,573	0.148
	0.0024	5.516	21,600	6,034,105	14,503	0.289
	0.0018	5.800	21,600	6,045,607	8,643	0.324
	0.0002	6.729	21,600	5,861,623	4,098	0.420

Table EC.2 Computational results from the $\epsilon-$constraint method to obtain the Pareto-frontier between the partisan asymmetry and compactness objectives, when coarsening using maximum matchings (MM) and maximal matchings (ML). *The solver takes extra time beyond the 6 hour MIP time limit to completely terminate.

EC.4.3. Competitiveness and Compactness

Coarsening	ϕ_{cmpptv}	$\phi_{comp}(\times 10^9)$	Clock time (s)	CPU time (ticks)	# B&C nodes	Opt. gap ($\frac{UB-LB}{UB}$)
ML	0.333	2.691	897	443,614	642	-
	0.313	2.699	590	251,040	1,208	-
	0.303	2.709	651	296,992	1,211	-
	0.290	2.714	475	220,096	426	-
	0.276	2.724	574	278,495	1,608	-
	0.266	2.738	529	243,615	807	-
	0.258	2.746	520	239,000	604	-
	0.248	2.761	693	317,177	2,105	-
	0.239	2.768	755	349,210	2,952	-
	0.231	2.781	825	383,338	2,746	-
	0.223	2.788	972	496,707	548	-
	0.213	2.802	1,639	869,083	1,916	-
	0.163	2.875	1,755	809,647	4,702	-
	0.153	2.895	11,345	3,633,036	7,867	-
	0.138	2.905	1,825	710,101	1,506	-
	0.123	2.923	11,290	3,416,250	2,539	-
	0.111	2.939	16,952	5,092,444	3,451	-
	0.099	2.959	33,652	11,747,083	5,430	-
MM	0.286	3.923	1,094	503,219	918	-
	0.265	3.935	321	156,208	182	-
	0.247	3.965	2,866	1,029,225	3,429	-
	0.246	3.972	11,033	3,425,895	5,827	-
	0.238	3.988	995	417,346	1,199	-
	0.234	3.994	444	203,912	414	-
	0.224	4.018	1,025	463,780	2,001	-
	0.188	4.081	4,212	1,368,974	2,730	-
	0.186	4.088	6,886	2,225,135	4,722	-
	0.171	4.089	3,115	1,095,628	3,175	-
	0.159	4.103	556	253,614	616	-
	0.135	4.131	16,198	5,681,976	9,785	-
	0.131	4.146	2,928	1,161,383	3,175	-
	0.129	4.154	5,835	2,082,055	6,933	-
	0.125	4.155	3,121	1,083,149	5,181	-
	0.121	4.157	2,255	780,248	4,666	-
	0.119	4.164	2,952	1,000,392	4,791	-

Table EC.3 Computational results from the $\epsilon-$constraint method to obtain the Pareto-frontier between competitiveness and compactness objectives, when coarsening using maximum matchings (MM) and maximal matchings (ML). From the 72 solutions obtained when coarsening using ML, a subset of 18 solutions are showcased here for simplicity. This subset is selected by picking every fourth solution in the sorting of the 72 solutions according to their respective competitiveness (ϕ_{cmpttv}) values. An optimality gap of "-" indicates that the solution is optimal, i.e., the upper bound (UB) and lower bound (LB) on the compactness objective found in the branch-and-cut algorithm are equal.