

Suppose  $A, B \in \mathbb{R}^{(m+n) \times (m+n)}$  are thought of as adjacency matrices (ie symmetry, simplicity, hollowness, etc etc not required), partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \quad \text{order of these matrices being } (m+n) \times (m+n)$$

such that the first  $m$  vertices of  $A$  and  $B$ 's graphs are “hard seeds” (ie known to match 1 to 1, 2 to 2, 3 to 3,...  $m$  to  $m$ ) and the other  $n$  vertices are matched in an unknown way.

The objective function of interest which we wish to maximize over the permutation matrices  $P \in \{0, 1\}^{n \times n}$ , temporarily relaxed to the doubly stochastic matrices  $P \in \mathbb{R}^{n \times n}$ , is

$$\begin{aligned} f(P) &= \text{trace} \left( \begin{bmatrix} A_{11}^T & A_{21}^T \\ A_{12}^T & A_{22}^T \end{bmatrix} \begin{bmatrix} I_{m \times m} & 0_{m \times n} \\ 0_{n \times m} & P \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} I_{m \times m} & 0_{m \times n} \\ 0_{n \times m} & P^T \end{bmatrix} \right) \\ &= \text{trace} \left( \begin{bmatrix} A_{11}^T & A_{21}^T \\ A_{12}^T & A_{22}^T \end{bmatrix} \begin{bmatrix} B_{11} & B_{12}P^T \\ PB_{21} & PB_{22}P^T \end{bmatrix} \right) \\ &= \text{trace} A_{11}^T B_{11} + \text{trace} A_{21}^T P B_{21} + \text{trace} A_{12}^T B_{12} P^T + \text{trace} A_{22}^T P B_{22} P^T \\ &= \text{trace} A_{11}^T B_{11} + \text{trace} P^T A_{21} B_{21}^T + \text{trace} P^T A_{12}^T B_{12} + \text{trace} A_{22}^T P B_{22} P^T \end{aligned}$$

which has gradient

$$G(P) := A_{21} B_{21}^T + A_{12}^T B_{12} + A_{22} P B_{22}^T + A_{22}^T P B_{22}.$$

Given any particular doubly stochastic matrix  $\hat{P} \in \mathbb{R}^{n \times n}$ , the Frank-Wolfe-step linearization involves maximizing  $\text{trace} Q^T G(\hat{P})$  over all of the doubly stochastic matrices  $Q \in \{0, 1\}^{n \times n}$ . The Hungarian Algorithm will in fact find the optimal  $Q$ , call it  $\hat{Q}$  (and it will be a permutation matrix). The next task in the Frank-Wolfe algorithm step will be maximizing the objective function over the line segment from  $\hat{P}$  to  $\hat{Q}$ ; ie maximizing  $g(\alpha) := f(\alpha \hat{P} + (1 - \alpha) \hat{Q})$  over  $\alpha \in [0, 1]$ . Denote  $c := \text{trace} A_{22}^T \hat{P} B_{22} \hat{P}^T$  and  $d := \text{trace}(A_{22}^T \hat{P} B_{22} \hat{Q}^T + A_{22}^T \hat{Q} B_{22} \hat{P}^T)$  and  $e := \text{trace} A_{22}^T \hat{Q} B_{22} \hat{Q}^T$  and  $u := \text{trace}(\hat{P}^T A_{21} B_{21}^T + \hat{P}^T A_{12}^T B_{12})$  and  $v := \text{trace}(\hat{Q}^T A_{21} B_{21}^T + \hat{Q}^T A_{12}^T B_{12})$ . Then (ignoring the additive constant  $\text{trace} A_{11}^T B_{11}$  without loss of generality, since it won't affect the maximization) we have  $g(\alpha) = c\alpha^2 + d\alpha(1 - \alpha) + e(1 - \alpha)^2 + u\alpha + v(1 - \alpha)$  which simplifies to  $g(\alpha) = (c - d + e)\alpha^2 + (d - 2e + u - v)\alpha + (e + v)$ . Setting the derivative of  $g$  to zero yields potential critical point  $\hat{\alpha} := \frac{-(d - 2e + u - v)}{2(c - d + e)}$  (if indeed  $0 \leq \hat{\alpha} \leq 1$ ); thus the next Frank-Wolfe algorithm iterate will either be  $\hat{P}$  (in which case algorithm would halt) or  $\hat{Q}$  or  $\hat{\alpha} \hat{P} + (1 - \hat{\alpha}) \hat{Q}$ . At termination of the Frank-Wolfe Algorithm, we would “unrelax” by maximizing  $\text{trace} R^T \hat{P}$  over permutation matrices  $R \in \{0, 1\}^{n \times n}$ —using the Hungarian Algorithm—and the optimal permutation matrix  $R$  encodes our proposed bijection between the last  $n$  vertices of  $A$ 's graph and the last  $n$  vertices of  $B$ 's graph.