Bipartite graph Matching

Jiaxin Hu

February 3, 2022

This note aims to generalize the matching techniques in Ding et al. (2021, Section 3.1, 3.2) for bipartite graphs (unbalanced matrices with Bernoulli entries).

1 Problem Setup

We first define the generalized Erdos-Renyi graph and the generalized correlated Erdos-Renyi graphs.

Definition 1 (Generalized Erdos-Renyi Graph $\mathcal{G}(n,m,q)$)**.** Let $A \in \{0,1\}^{n\times m}$ denote the adjacency matrix of a generalized Erdos-Renyi graph $\mathcal{G}(n,m,q)$, where $q \in (0,1)$ and

$$P(A_{ij}=1)=q, \quad \text{for all} \quad (i,j)\in[n]\times[m].$$

Note that the generalized Erdos-Renyi graph is not necessarily symmetric.

Definition 2 (Generalized Correlated Erdos-Renyi Graph $\mathcal{G}(n,m,q;s)$)**.** Let $A, B \in \{0,1\}^{n\times m}$ denote two adjacency matrices of two Erdos-Renyi graphs $\mathcal{G}(n,m,q)$, where $q\in(0,1)$. Let $\pi_1^*: [n]\mapsto[n]$ and $\pi_2^*:[m]\mapsto[m]$ denote the latent permutations for the row and column indices. Conditional on A, we assume for all $(i,j)\in[n]\times[m]$, $B_{\pi_1^*(i),\pi_2^*(j)}$ are independent and distributed as

$$B_{\pi_1^*(i),\pi_2^*(j)} = \begin{cases} Ber(s) & \text{if} \quad A_{ij}=1 \\ Ber\left(\frac{q(1-s)}{1-q}\right) & \text{if} \quad A_{ij}=0 \end{cases}.$$

2 Generalized Matching via Degree Profile

Without loss of generality, we consider the matching problem to find π_1^*.

For each vertex $i\in[n]$, define the *connected set* $N_A(i)$ (corresponding to the "open/closed neighbourhood" in (Ding et al., 2021)) as

$$N_A(i)=\{j\in[m]: A_{ij}=1\}, \quad \text{with} \quad a_i=|N_A(i)|.$$

Define $N_B(i)$ and b_i similarly. Also, define the "degree" (corresponding to the "outdegree" in (Ding et al., 2021)) of vertex $j\in[m]$ in A,B as

$$a_j^{(i)}=\frac{1}{\sqrt{(n-1)q(1-q)}}\sum_{l\neq i}(A_{lj}-q), \quad b_j^{(i)}=\frac{1}{\sqrt{(n-1)q(1-q)}}\sum_{l\neq i}(B_{lj}-q).$$

Consider the empirical distributions of $a_j^{(i)}$ for all $j \in N_A(i)$ and $b_j^{(i)}$ for all $j \in N_B(i)$ as

$$\mu_i = \frac{1}{a_i} \sum_{j \in N_A(i)} \delta_{a_j^{(i)}}, \quad \nu_i = \frac{1}{b_i} \sum_{j \in N_B(i)} \delta_{b_j^{(i)}},$$

where δ_x refers to the point mass at point x, and the centered version

$$\bar{\mu}_i = \mu_i - \overline{Bin(n-1,q)}, \quad \bar{\nu}_i = \nu_i - \overline{Bin(n-1,q)},$$

where $\overline{Bin(k,p)}$ denotes the standardized binomial distribution, that is, the law of $\frac{X-kp}{\sqrt{kp(1-p)}}$ for $X \sim Bin(k,p)$. Then, we obtain the the distance Z_{ik} with $\bar{\mu}_i$ and $\bar{\nu}_k$ as Ding et al. (2021).

Without the symmetry, we need to repeat the above procedures for the column matching to find π_2^*.

The possible generalization of Algorithm 1 is in Algorithm 1.

Algorithm 1 Generalized graph matching via degree profile

Input: Graphs $A, B \in \{0,1\}^{n \times m}$, an integer L (tuning parameter).

1: For each $i, k \in [n]$, calculate the row distances Z_{ik}^r; for each $j, l \in [m]$, calculate the column distances Z_{jl}^c.
2: Sort $\{Z_{ik}^r : i, k \in [n]\}$ and let $\mathcal{S}_1$ be the set of indices of the smallest n elements; sort $\{Z_{jl}^c : j, l \in [m]\}$ and let $\mathcal{S}_2$ be the set of indices of the smallest m elements.
3: **if** there exists $\hat{\pi}_1$ such that $\mathcal{S}_1 = \{(i, \hat{\pi}_1(i)) : i \in [n]\}$; there exists $\hat{\pi}_2$ such that $\mathcal{S}_2 = \{(i, \hat{\pi}_2(i)) : i \in [n]\}$ **then**
4: Output $\hat{\pi}_1$ and $\hat{\pi}_2$
5: **else**
6: Output error.
7: **end if**

Output: Estimated permutations $\hat{\pi}_1, \hat{\pi}_2$ or error.

References

Ding, J., Ma, Z., Wu, Y., and Xu, J. (2021). Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, 179(1):29–115.