Hypergraph Matching

Jiaxin Hu

February 5, 2022

1 Problem Setup

Consider two correlated Erdős-Rényi m-hypergraphs $\mathcal{A}, \mathcal{B} \sim \mathcal{G}(n, q)$ on the same vertex sets $[n]$. Let $\pi : [n] \mapsto [n]$ denote the latent permutation. Assume that conditional on $\mathcal{A}$, for all $\omega = (\omega_1, \ldots, \omega_m)$ where $1 \leq \omega_1 < \cdots < \omega_m \leq n$, $\mathcal{B}_{\pi(\omega)}$ are independent and distributed as

$$B_{\pi(\omega)} \sim \begin{cases} Ber(s) & \text{if} \quad \mathcal{A}_\omega = 1 \\ Ber\left(\frac{q(1-s)}{1-q}\right) & \text{if} \quad \mathcal{A}_\omega = 0. \end{cases}$$

Note that $\mathcal{A}, \mathcal{B}$ are symmetric tensors, and $\mathcal{A}(\omega) = \mathcal{B}(\omega) = 0$ for all the entries ω in which $\omega_i = \omega_j$ for some $i \neq j \in [n]$.

2 Derivation of distance statistics

Define the open neighbourhood of i as the collection of other nodes that are connected to node i

$$\mathcal{N}_\mathcal{A}(i) = \{(i_2, \ldots, i_m) \in [n]^{m-1} : \mathcal{A}_{i, i_2, \ldots, i_m} = 1\}.$$

Define the plain neighbourhood of i as

$$\mathcal{C}_\mathcal{A}(i) = \{i \in [n] : i \in \omega, \omega \in \mathcal{N}_\mathcal{A}(i)\} \cup \{i\},$$

where $i \in \omega$ is equivalent to $i = \omega_j$ for some $j \in [n]$. Then $\mathcal{C}_\mathcal{A}(i)$ collects all the nodes in $[n]$ connected with i via a higher-order connection $\mathcal{A}_\omega$ for some $\omega \in \mathcal{N}_\mathcal{A}(i)$. Let $c_a(i) = |\mathcal{C}_\mathcal{A}(i)|$.

Define the "innerneighbourhood" of vertex j of i as

$$\begin{aligned} \mathcal{D}_\mathcal{A}(j)^{(i)} = &\{(i_1, \ldots, i_{m-1}) : i_k \neq i_l, \text{ for all } k \neq l \in [m-1], i_k \in \mathcal{C}_\mathcal{A}(i)/\{j\} \text{ for all } k \in [m-1]\} \\ &\cup \{(i_1, \ldots, i_{m-1}) : \text{at least one of } i_k = j\}. \end{aligned}$$

Note that $\mathcal{D}_\mathcal{A}(j)^{(i)}$ collects all the sub-edges that (1) are generated from the plain neighbourhood of i and (2) leads to non-defined hyperedges involving at least two elements equal to j, e.g., $(j, 1, \ldots, j)$. Also note that

$$|\mathcal{D}_\mathcal{A}(j)^{(i)}| = P^{m-1}_{c_a(i)} + \sum_{k=1}^{m-1} \binom{m-1}{k} n^{m-1-k},$$

where $P_{c_a(i)}^{m-1}$ is the number of permutation of $m-1$ elements out of $c_a(i)$ elements.

Then, the "outdegree" of vertex j is

$$
\begin{aligned}
a_j^{(i)} &= \frac{1}{\sqrt{n^{m-1} - |\mathcal{D}_{\mathcal{A}}(j)^{(i)}|}} \sum_{\omega \notin \mathcal{D}_{\mathcal{A}}(j)^{(i)}} (\mathcal{A}_{\omega,j} - q) \\
&= \frac{1}{\sqrt{n^{m-1} - (P_{c_a(i)}^{m-1} + \sum_{k=1}^{m-1} \binom{m-1}{k} n^{m-1-k})}} \sum_{\omega \notin \mathcal{D}_{\mathcal{A}}(j)^{(i)}} (\mathcal{A}_{\omega,j} - q).
\end{aligned}
$$

Next, note that we need to consider all the $j \in \mathcal{C}_{\mathcal{A}}(i)$ except the node i itself. So, we define the empirical distribution

$$
\bar{\mu}_i = \frac{1}{c_a(i) - 1} \sum_{\mathcal{C}_{\mathcal{A}}(i)/\{i\}} \delta_{a_j^{(i)}} - \overline{\mathrm{Binom}}(n^{m-1} - |\mathcal{D}_{\mathcal{A}}(j)^{(i)}|, q).
$$

Similar for network $\mathcal{B}$.

References