Reverse Engineering Molecular Hypergraphs

Ahsanur Rahman, Christopher L. Poirel, David J. Badger, and T. M. Murali
Department of Computer Science
ICTAS Center for Systems Biology of Engineered Tissues
Virginia Tech
Blacksburg, VA, USA
ahsanur@vt.edu, poirel@vt.edu, dbadger@vt.edu, murali@cs.vt.edu

1. PROOFS OF BOUNDS ON HYPEREDGE DENSITY

Lemma 1 *If S is a (β, σ)-hyperedge with k nodes, then the density of $\mu_S(G)$ is at least*

$$\frac{\beta\left(\sum_{i=0}^{l-1} i\binom{\binom{k}{2}}{i} + l\left(\sigma 2^{\binom{k}{2}} - \sum_{i=0}^{l-1}\binom{\binom{k}{2}}{i}\right)[l>0]\right)}{\binom{k}{2}},$$

where l is the smallest integer such that

$$\sum_{i=0}^{l}\binom{\binom{k}{2}}{i} \geq \sigma 2^{\binom{k}{2}}.$$

In the lemma, [ ] denotes an indicator function, which is true if and only if l is positive.

Before outlining the proof of Lemma 1, we remind the reader of three concepts that are important in the definition of a (β, σ) hyperedge:

(i) $\mathcal{G}(S)$ is the multiset of the subgraphs induced by S as we vary the graphs in $\mathcal{G}$,

(ii) $\mathcal{P}(S)$ is the set of $2^{\binom{k}{2}}$ possible graphs on the nodes in S, and

(iii) if H is a graph in $\mathcal{P}(S)$, then $\psi(H)$ is the number of occurrences of H in $\mathcal{G}(S)$.

PROOF. To prove this lower bound, we consider the sparsest graphs in $\mathcal{G}(S)$ that enable S to be a hyperedge. To assist this analysis, we partition $\mathcal{P}(S)$ into $\binom{k}{2}+1$ sets where $\mathcal{P}_i(S), 0 \leq i \leq \binom{k}{2}$ is the set of graphs on the nodes in S that contain exactly i edges. By construction, $\mathcal{P}_i(S)$ contains $\binom{\binom{k}{2}}{i}$ graphs. It is easy to see that the lower bound is achieved when the following conditions are satisfied:

(i) if $H \in \mathcal{P}(S)$ occurs at least once in $\mathcal{G}(S)$, i.e., $\psi(H) > 0$, then H is one of the $\sigma 2^{\binom{k}{2}}$ sparsest graphs in $\mathcal{P}(S)$, i.e., the graphs with the smallest number of edges;

(ii) for each such graph H, $\psi(H) = \beta n$, and

(iii) each of the remaining graphs in $\mathcal{G}(S)$ is the empty graph, i.e., the only graph in $\mathcal{P}_0(S)$.

Using the definition of l in the statement of the lemma, we select the $\sigma 2^{\binom{k}{2}}$ sparsest graphs in $\mathcal{P}(S)$ as follows: (i) pick all the graphs in the sets $\mathcal{P}_0(S), \mathcal{P}_1(S), \ldots, \mathcal{P}_{l-2}(S), \mathcal{P}_{l-1}(S)$ and (ii) pick as many graphs as necessary from $\mathcal{P}_l(S)$ so as to obtain $\sigma 2^{\binom{k}{2}}$ graphs. To obtain the lower bound on the density, we simply compute the total number of edges in these graphs, note that each graph occurs βn times in $\mathcal{G}(S)$, and divide by $n\binom{k}{2}$. To compute the total number of edges in these graphs, note that each graph in $\mathcal{P}_i(S), 1 \leq i < l$ contains i edges. We use $\sigma 2^{\binom{k}{2}} - \sum_{i=0}^{l-1}\binom{\binom{k}{2}}{i}$ graphs from $\mathcal{P}_l(S)$, with each of these graphs containing l edges. Finally, we need the indicator function $[l > 0]$ to avoid double counting in the special case when $l = 0$, i.e., when $\sigma = 1/2^{\binom{k}{2}}$. $\square$

Lemma 2 *If S is a (β, σ)-hyperedge, then the density of $\mu_S(G)$ is at most*

$$\frac{\beta u\left(\sigma 2^{\binom{k}{2}} - \sum_{i=u+1}^{\binom{k}{2}}\binom{\binom{k}{2}}{i}\right)[u<\binom{k}{2}]}{\binom{k}{2}} + \frac{\beta\left(\sum_{i=u+1}^{\binom{k}{2}-1} i\binom{\binom{k}{2}}{i}\right)}{\binom{k}{2}} + \left(1+\beta-\beta\sigma 2^{\binom{k}{2}}\right),$$

where u is the largest integer such that

$$\sum_{i=u}^{\binom{k}{2}} \binom{\binom{k}{2}}{i} \geq \sigma 2^{\binom{k}{2}}.$$

We sketch the proof of Lemma 2, since it is very similar to the proof of Lemma 1.

PROOF. To obtain the upper bound on the density of $\mu_S(G)$, we pack the densest graphs in $\mathcal{P}(S)$ into the set of $\sigma 2^{\binom{k}{2}}$ graphs that occur at least βn times in $\mathcal{G}(S)$. Using the definition of u from the statement of lemma, these graphs belong to the sets $\mathcal{P}_{\binom{k}{2}}(S), \mathcal{P}_{\binom{k}{2}-1}(S), \ldots, \mathcal{P}_{u+2}(S), \mathcal{P}_{u+1}(S)$ and as many graphs as necessary from $\mathcal{P}_u(S)$. Each of these graphs occurs βn times in $\mathcal{G}(S)$ to satisfy the constraint imposed by β . We fill in the remain elements of $\mathcal{G}(S)$ using the complete graph in $\mathcal{P}_{\binom{k}{2}}(S)$; this graph gives rise to the term $(1 + \beta - \beta\sigma 2^{\binom{k}{2}})$ in the upper bound. $\square$