Modeling Cell Signaling Networks with Prize-Collecting Subhypernetworks

{ Barney Potter, James Fix, and Anna Ritz }

REED COLLEGE

Introduction

Cell signaling pathways are important tools used by biologists to model ho signals are transduced through cells. Though signaling pathways can be (and often are) modeled using graphs, we instead use a generalization known as directed hypergraphs. We find directed hypergraphs to be a useful alternative to standard graphs, because hypergraphs give us a means to represent complex biological reactions that may have more than one reactant or product.

Network & Weighting

We built directed hypergraphs from Reactome Pathways. Each protein was given a prize based on false discovery rate [1] between healthy and basal cell carcinoma (BCC) gene expression:

- Node n was assigned a prize according to $p(n) = -log(FDR_n)$
- Multi-protein nodes were given the *minimum* prize of their component proteins
- Nodes with missing data were manually assigned a defalult value

What is a directed hypergraph?

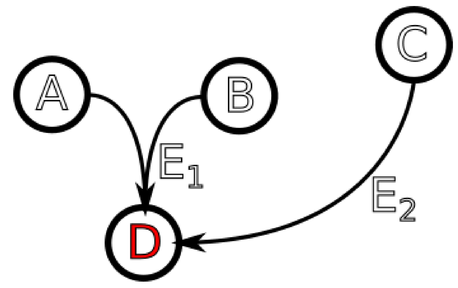

Figure 1: A directed hypergraph containing four nodes and two hyperedges.

We define a directed hypergraph, $\mathcal{H}$, as a pair $(V, \mathcal{E})$, where V is a finite set of vertices and $\mathcal{E} \subseteq 2^V \times 2^V$ is a finite set of *directed hyperedges* connecting members of V. A directed hyperedge $e \in \mathcal{E}$ relates a tail set T(e) to a head set H(e). We require that H(e) and T(e) are disjoint and are each nonempty.

Using hypergraphs, we model proteins, protein complexes, and small molecules as nodes and we use hyperedges to denote their involvement in known reactions [2, 3].

Hypershrubs

We develop a formulation that seeks *hypershrubs*, meaningful subnetworks that, roughly speaking, connect interesting sources to interesting targets. Consider the following for an edge subset $A \subseteq \mathcal{E}$. Define $T(A) = \cup_{e \in A} T(e)$, $H(A) = \cup_{e \in A} H(e)$, and $V(A) = T(A) \cup H(A)$.

- The *induced root set of* A is $D(A) = V(A) \setminus H(A)$ and the *induced leaf set of* A is $C(A) = V(A) \setminus T(A)$.
- Given $S, T \subseteq V$ where $S \cap T = \emptyset$, an S,T*-hypershrub* is an $A \subseteq \mathcal{E}$ where $S, T \subseteq V(A)$, $D(A) \subseteq S$ and $C(A) \subseteq T$.

When sources S and targets T are determined ahead of time, an S,T-hypershrub A is a subnetwork that includes all the source and target nodes, directed only from sources as roots to targets as leaves.

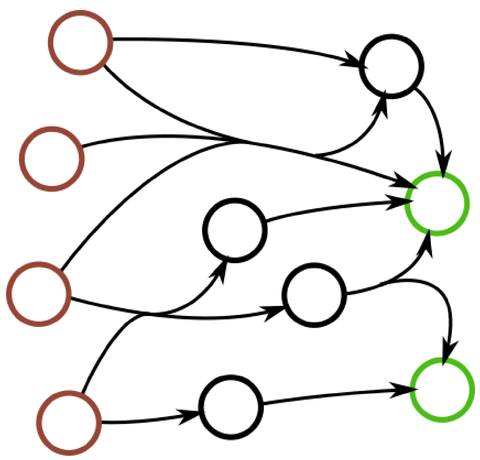

Figure 2: An example hypershrub $\mathcal{S}$ comprised of eight hyperedges A. The induced roots, $D(A)$, are shown in brown on the left and the induced leaves, $C(A)$, are shown in green on the right.

PCSHN

To find the nearest approximation of a hypershrub given only a weighted hypergraph $\mathcal{H} = (V, \mathcal{E})$, we seek the edge set $A \subseteq \mathcal{E}$ which maximizes:

$$o(A) = \sum_{v \in V(A)} g(v) - \sum_{e \in A} c(v) - \sum_{v \in D(A)} h(v) - \sum_{v \in C(A)} \ell(v)$$

We call the subhypergraph formed by A and its induced vertices the Prize-Collecting Subhypernetwork (PCSHN). This formulation seeks to (1) Maximize node prizes, (2) Minimize edge costs, and (3) Minimize nodes without an incoming or outgoing hyperedge. We implement this optimization through the use of an integer linear program (ILP).

Results

After restricting analysis to the Hedgehog (Hh) pathway, a subset of the optimal solution is the is the interaction of Kif7, SUFU and Gli2 molecules and the EVC:EVC2 complex to yield Gli2 and SUFU.

	# Nodes	# Hyperedges
Interactome	35,914	8,615
Hh Pathway	413	82
Solution	99	44

1. SUFU and Gli2 bind together
2. Kif7 assists the complex to the ciliary membrane
3. EVC:EVC2 assists in the dismantling of the SUFU:Gli2 complex, diverting Gli2 from the degradation pathway
4. Gli2 can then localize to the nucleus and cause transcriptional changes

This particular part of the Hedgehog pathway has been implicated in the upregulation of Hedgehog signaling. Since upregulation of the pathway has been connected to increased risk for BCC, this may be one predictor for BCC.

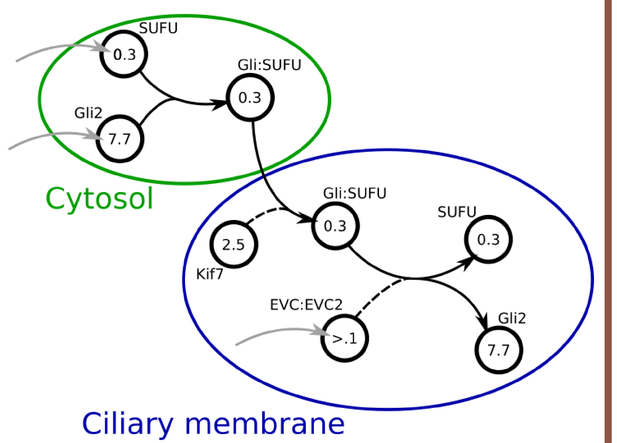

Figure 3: One sub-pathway found in the PCSHN when applied to BCC data in the Hedgehog signaling pathway.

Discussion

- Hypershrubs identify minimal network of source-sink pairs
- PCSHNs provide high-confidence networks which approximate hypershrubs

References

[1] Byul A J et al. *Molecular carcinogenesis*, 2015.

[2] A. Ritz et al. *Trends in Biotechnology.*, 2014.

[3] A. Ritz et al. *IEEE Transactions on Computational Biology and Bioinformatics*, 2015.

Contact Information

Web www.barneyipotter.com;
http://www.reed.edu/biology/ritz/

Email barney.potter.24@gmail.com;
aritz@reed.edu