Modeling Cell Signaling Networks with Prize-Collecting Subhypernetworks

Barney Potter
Reed College Departments of
Biology and Mathematics
3203 SE Woodstock Blvd
Portland, OR
bpotter@reed.edu

James Fix
Reed College Department of
Mathematics
3203 SE Woodstock Blvd
Portland, OR
jimfix@reed.edu

Anna Ritz
Reed College Department of
Biology
3203 SE Woodstock Blvd
Portland, OR
aritz@reed.edu

ABSTRACT

Cell signaling pathways are important tools used by biologists to model how 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 way to represent complex biological reactions that may have more than one reactant or product.

Prior work analyzing cell signaling as hypergraphs sought structures called *hyperpaths*. We develop the notion of a *hypershrub*, a multi-source, multi-target generalization of a hyperpath. We create a formulation that finds specific, prize-dense hypershrubs that may be representative of biological phenomena that went previously unnoticed in standard graph representations of the same signaling pathways.

To build our graphs, we introduce a weighting scheme that uses differential expression data to assign weights to nodes, representative of proteins and protein complexes, based on confidence that a given node is differentially regulated between conditions. This method is agnostic about whether or not a node is up-regulated or down-regulated between conditions, allowing us to find pathways that may involve both positive and negative regulation. We apply this formulation to the human Hedgehog signaling pathway, using data from Basal Cell carcinoma (BCC), to analyze the effectiveness of our algorithm on real data.

Keywords

hypergraphs, cell signaling networks, cancer, BCC

1. INTRODUCTION

One of the most promising areas where computational methods are now being applied in a biological context concerns the study of signaling pathways within cells. Cell signaling pathways are useful to understand how information propagates throughout cells, ultimately changing their be-

behavior. Understanding how signaling information spreads throughout a cell helps us to better grasp dysregulations that result in changes to cellular dynamics. These dysregulations are often manifested in heterogeneous diseases such as cancer [7].

One of the difficulties presented by applying traditional computational methodologies to analysis of signaling pathways is that modeling pathways that contain complex reactions using standard graphs leads to information loss and oversimplification. To better model these systems, we use directed hypergraphs [6] to more accurately model the biological reactions taking place within cells. 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 $\mathrm{T}(e)$ to a head set $\mathrm{H}(e)$. We require that $\mathrm{H}(e)$ and $\mathrm{T}(e)$ are disjoint and are each nonempty [2]. Using hypergraphs, we model proteins, protein complexes, and small molecules as nodes and we use hyperedges to denote their involvement in known reactions.

2. 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 $C(A) = V(A) \setminus T(A)$. An example of this is shown in Figure 1. 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.

For the pathway analysis applications we consider here, we seek hypershrub-like structures in hypergraphs by instead inferring sources and targets from additional experimental data. This leads us to a "prize-collecting" formulation that places weights on nodes and hyperedges, a series of scores and costs, to find optimal subnetworks that include valued nodes, while minimizing the edges and induced leaves and roots that connect them.

3. WEIGHTING FROM DATA

To build hypergraphs, we draw upon the wealth of publically available databases that have arisen in the last decade
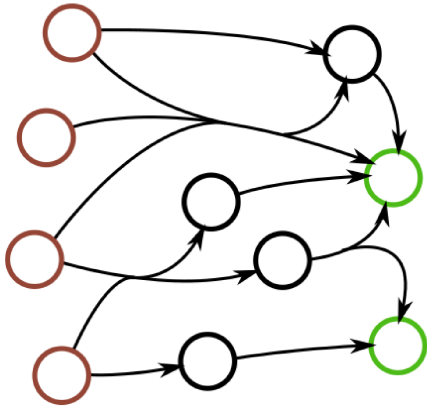
Figure 1: An example hypershrub $\mathcal{S}$ comprised of eight hyperedges A. The induced roots, $D(A)$, are shown in brown on the left and the induced leaves, $C(A)$, are shown in green on the right.

that contain information about known signaling pathways. For our analysis, we use the Reactome database [1, 5], which can be easily parsed to create hypergraphs from the BioPAX Level 3 file format. We assigned prizes to the nodes of the Hedgehog Pathway by considering differential gene expression values in a BCC study [4]. Prizes were assigned according to the prize function $p(n) = -log(FDR_n)$ where FDR_n is the false discovery rate between cancer and normal conditions for node n. If n was a multi-protein complex, it was assigned a prize equal to the minimum value given when p is applied to the proteins in n. Finally, if n was a small molecule other than a protein or complex, or if data was not available for n, then it was given a default value which was optimized for manually.

4. RESULTS

We computed a prize-rich hypershrub from the Reactome Hedgehog Signaling pathway with nodes weighted from BCC data. The solution contained 44 hyperedges and 99 nodes. Since the output hypergraph was still relatively large, we restricted our analysis to one subcomponent of the Hedgehog signaling pathway, known as the *Hedgehog 'on' state* that accounts for approximately one third of the reactions in the Hedgehog signaling pathway. Using this, we were able to search for any nodes that were roots or leaves in the solution, which were present in this pathway. Similarly, we were able to look at the biochemical reactions in the *Hedgehog 'on' state*, to determine if those reactions connected the roots and leaves that we found.

One reaction that we find in this solution is the path from Kif7, SUFU, and Gli2 molecules and the EVC:EVC2 complex to yield Gli2 and SUFU. This particular pathway involves three hyperedges, two cases of positive regulation, and eight total nodes. Kif7 is a kinesin that is known for regulating the Hedgehog signaling pathway and controling localization of SUFU:Gli2 to the ciliary membrane. SUFU is a kinesin-binding protein that is known to down-regulate Gli2-mediated activation of genes. Finally, EVC and EVC2 are proteins that exist in the downstream Hedgehog-Smoothened signaling pathway that are known to regulate signaling by Hedgehog ([8]).

Biologically, this corresponds to the movement of SUFU and Gli2 from the cytosol to the ciliary membrane, assisted by Kif7 and EVC:EVC2. First, SUFU and Gli2 bind together to form the SUFU:Gli2 complex. Then Kif7 assists the complex in to the ciliary membrane. Finally, EVC:EVC2 assists in the dismantling of the SUFU:Gli2 complex, helping to divert Gli2 from the degradation pathway, allowing for Gli2 to localize to the nucleus and cause transcriptional changes ([3]). This particular part of the Hedgehog pathway has been implicated in the up-regulation of Hedgehog signaling ([9]). Since upregulation of the Hedgehog signaling pathway has been connected to increased risk for BCC, this may be one mechanism that could be a predictor for increased risk of BCC.

5. CONCLUSIONS

We have defined a multi-source, multi- sink structure that we have called a hypershrub. From this, we have modeled dysregulations in protein-protein interaction networks. We have found a way to weight these graphs appropriately based on differential gene expression data, in a way that is agnostic about whether genes are up- or down-regulated. From our subhypernetworks, we then show that we can use our implementation to reconstruct one pathway that has been implicated in the upregulation of Hedgehog signaling, which may be an indicator for potential tumorgenesis of basal-cell carcinoma.

6. REFERENCES

[1] D. Croft et al. The Reactome pathway knowledgebase. *Nucleic Acids Research*, 42(Database):472–477, 2014.

[2] G. Gallo et al. Directed hypergraphs and applications. *Discrete Applied Mathematics*, 42:177–201, apr 1993.

[3] E. W. Humke et al. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes & development*, 24(7):670–82, apr 2010.

[4] B. A. Jee et al. Molecular classification of basal cell carcinoma of skin by gene expression profiling. *Molecular carcinogenesis*, 54(12):1605–12, dec 2015.

[5] M. Milacic et al. Annotating cancer variants and anti-cancer therapeutics in Reactome. *Cancers*, 4(4):1180–1211, 2012.

[6] A. Ritz, B. Avent, and T. M. Murali. Pathway Analysis with Signaling Hypergraphs. *IEEE Transactions on Computational Biology and Bioinformatics*, pages 249–258, 2014.

[7] I. W. Taylor et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature biotechnology*, 27(2):199–204, feb 2009.

[8] The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Research*, 43(D1):D204–212, oct 2014.

[9] H. Tukachinsky, L. V. Lopez, and A. Salic. A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu-Gli protein complexes. *The Journal of cell biology*, 191(2):415–28, oct 2010.