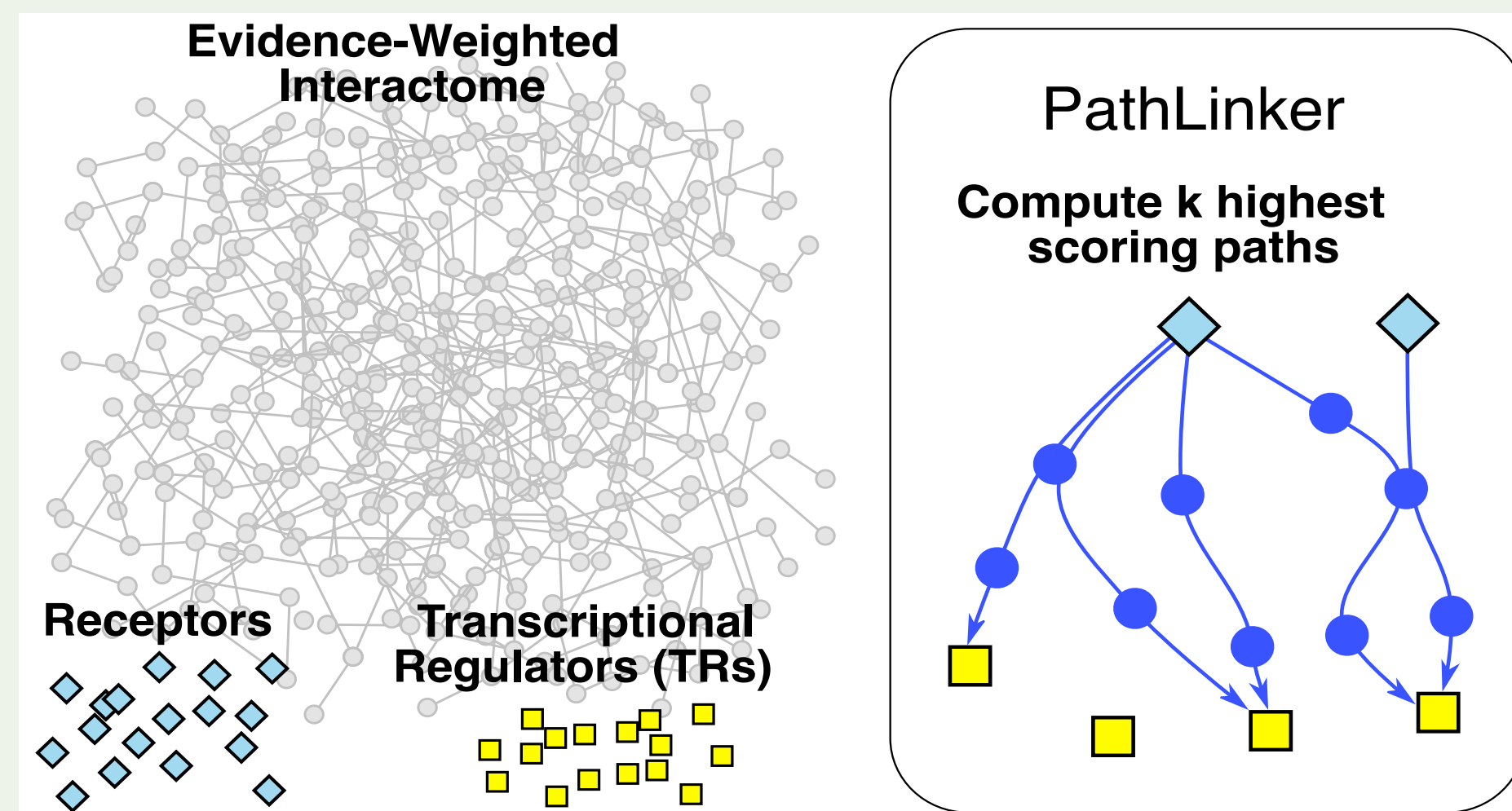


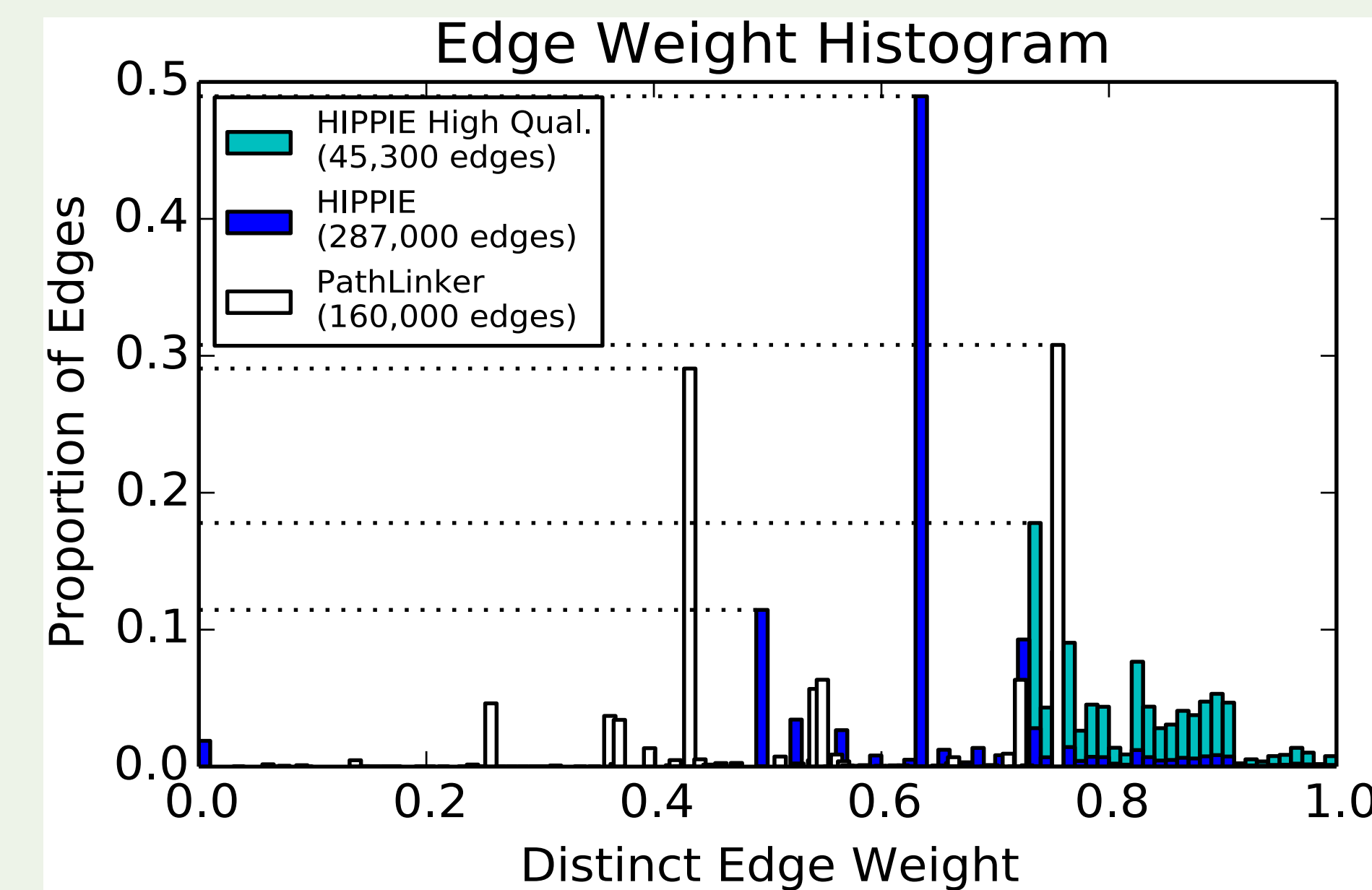
Introduction

- Cellular signaling pathways are a series of protein-protein (PP) interactions that eventually regulate target genes so that a cell responds to its environment.
- Graph algorithms for automatic reconstruction of signaling pathways, like PathLinker [1], take interactomes built from PP interactions as one of the inputs.
- These interactomes often include weights on the interactions (graph edges) to denote confidence or relevance of an interaction.
- PathLinker takes as inputs: 1) a weighted interactome, 2) a set of receptors, and 3) a set of transcriptional regulators (TRs), and outputs the k -shortest signaling paths, where each connects a receptor to a TR.



Problem

- The unit/coarsely-weighted edges in the PP interactomes lead to multiple reconstructed PathLinker paths sharing the same cost.
- This is a general problem for weighted interactomes such as PathLinker and HIPPIE [2].
- Paths containing more signaling interactions are not favored over their tied paths containing fewer or no signaling interactions.



Can we break tied paths and promote those having higher portions of true positive signaling edges?

Methods

Symbols:

G : graph, V : nodes (proteins)
 E : edges (interactions), W : edges weight
 P : Paths [$P_i \subset E$], C : path cost

PathLinker: Given a weighted graph $G = (V, E, W)$, PathLinker outputs a ranked list of the k shortest paths, $\mathcal{P} = [P_1, P_2, \dots, P_k]$, where a path P_i has a cost of c_i .

Classification Scheme: We use a support vector machine classifier with a polynomial kernel of the third degree trained with the nodes gene expression values.

- $\forall v \in V$, we assign a binary classification label:

$$y_v = \begin{cases} 1; & \text{if } v \text{ is predicted a (+ve) signaling edge;} \\ 0; & \text{otherwise.} \end{cases}$$

Path Reordering:

- $\forall e = (u, v) \in E$, we assign a score s_{uv} :

$$s_{uv} = y_u y_v$$

- $\forall P_i \in \mathcal{P}$, we assign a score $\mathcal{S}(P_i)$:

$$\mathcal{S}(P_i) = \sum_{(u,v) \in P_i} s_{uv}.$$

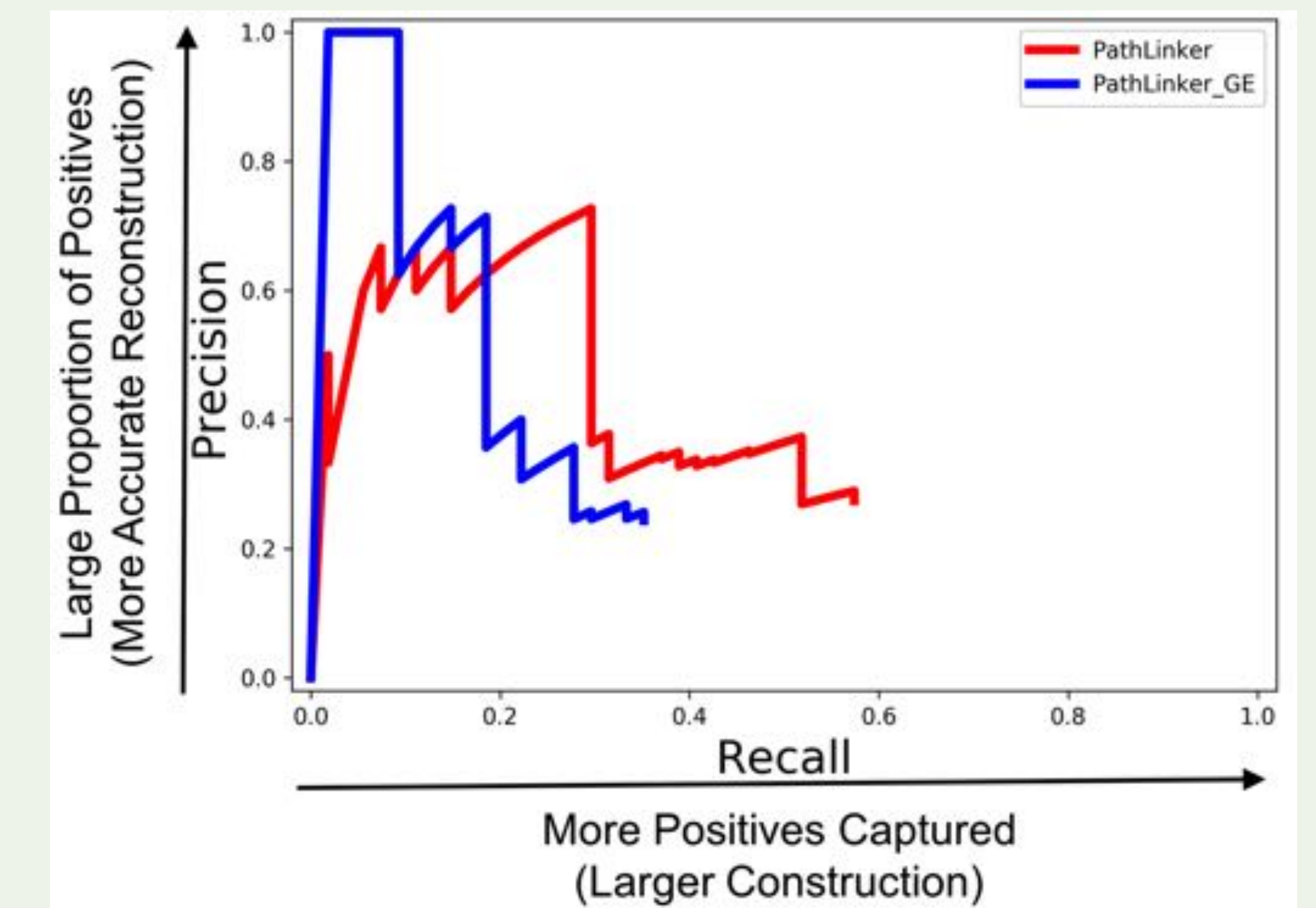
- We retain paths, $\mathcal{P}' \subseteq \mathcal{P}$, with at least one inferred positive PP interaction:

$$\mathcal{P}' = \{P_i : \mathcal{S}(P_i) > 0\}$$

- \mathcal{P}' is partitioned into ties: $\mathcal{P}' = [P_{c_1}, \dots, P_{c_n}]$, where $P_{c_j} \subseteq \mathcal{P}'$ is the group of paths sharing the same path cost c_j , and $c_j < c_{j+1}$ for $1 \leq j \leq n$ for n distinct costs.
- Paths $P_i \in P_{c_j}$ within each group are reordered according to $\mathcal{S}(P_i)$.

Results

- The proposed method has higher precision at the early recall values (blue curve below), meaning that the first shortest paths contain more true positives (signaling interactions from the NetPath Wnt pathway) and fewer negatives.
- This is because some positives were ranked higher (promoted) and some negatives were ranked lower (demoted) or removed completely.



Contribution

We propose a machine learning approach to: 1) use gene expression data to break ties in the PathLinker output paths, and 2) re-prioritize these paths based on the number of inferred true signaling edges. This approach promotes paths with more positives and demotes paths with more negatives or deletes them completely.

Case study: We use PathLinker [1] to reconstruct the Wnt signaling pathway from the NetPath database [3] and apply gene expression data from colon cancer samples, a disease that is known to dysregulate Wnt signaling [4], to the paths to break ties.

Conclusion and Future Directions

The proposed method can be widely applied to other interactome-based analyses since it is:

- a post-processing method after constructing the signaling pathways.
- built upon computing an edge score, so it can be applied to any signaling reconstruction method that outputs an ordered list of paths, trees, or other subgraphs.

The next step is to build a tissue-specific interactome to reconstruct signaling pathways specific to that tissue.

[1] A. Ritz *et al.*, "Pathways on demand: automated reconstruction of human signaling networks," *npj Systems Biology and Applications*, vol. 2, p. 16002, 2016.

[2] G. Alanis-Lobato *et al.*, "HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks.," *Nucleic Acids Research*, vol. 45, pp. D408–D414, 2017.

[3] K. Kandasamy *et al.*, "NetPath: a public resource of curated signal transduction pathways," *Genome Biol.*, vol. 11, p. R3, Jan 2010.

[4] D. M. Muzny *et al.*, "Comprehensive molecular characterization of human colon and rectal cancer," *Nature*, vol. 487, pp. 330–337, Jul 2012.

Contact Information

- Web: <http://www.reed.edu/biology/ritz>
- Email: aritz@reed.edu & youssefi@reed.edu