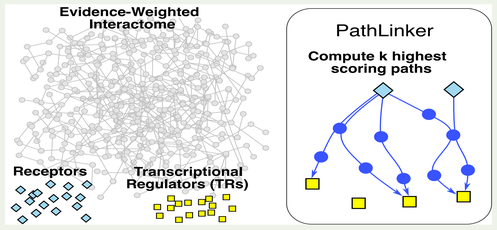
Integrating Protein Localization with Automated Signaling Pathway Reconstruction

Ibrahim Youssef[1,2] and Anna Ritz[1]

[1]Biology Dept., Reed College, OR, USA; [2]Biomedical Engineering Dept., Cairo University, Egypt

Introduction

- Cellular signaling pathways are a series of protein-protein interactions (PPIs) that eventually regulate target genes so that a cell responds to its environment.
- Graph algorithms for automatic reconstruction of signaling pathways consider proteins as nodes and interactions as edges, and work on interactomes built from PPIs [1].



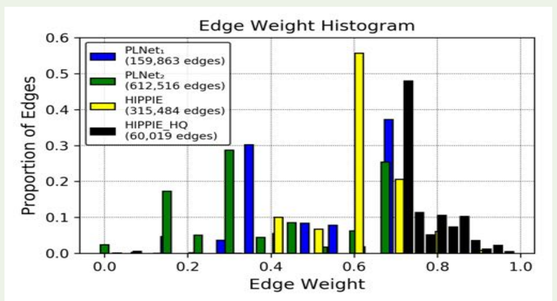
- Interactomes often include weights on the edges to denote confidence/relevance of an interaction.

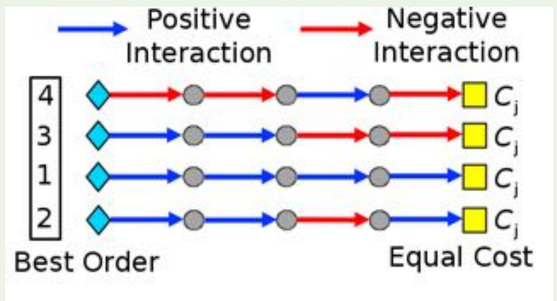
PathLinker:

- PathLinker [1] is a technique to automatically reconstruct signaling pathways.
- It takes as inputs: 1) a weighted interactome, 2) a set of receptors for a pathway of interest, and 3) a set of transcriptional regulators (TRs) for that pathway; and outputs the k-shortest signaling paths, $\mathcal{P} = [P_1, P_2, \ldots, P_k]$, where each has a cost c_i and each connects a receptor to a TR.

Problem

- Many interactomes such as those of PathLinker ($PLNet_1$ and $PLNet_2$) and HIPPIE [2] have a general problem of coarsely-weighted edges.
- This causes PathLinker to produce "equally good" multiple reconstructed paths that share the same cost.





Can we (i) break tied paths and (ii) promote paths having higher portions of true positive signaling edges?

Contribution

We propose *LocPL*, a method to improve the automatic reconstruction of signaling pathways from PPIs by incorporating information about protein localization to ensure that (i) the proteins in a reconstruction are localized in cellular compartments involved in signaling transduction and that (ii) the interactions are consistent with signaling from the membrane to the nucleus.

Case study: We applied *LocPL* on signaling pathways from the NetPath database [3], as the ground truth for signaling interactions of specific pathways, utilizing localization information from ComPPI [4]—a database that predicts cellular compartments for proteins and PPIs.

Conclusion and Future Directions

LocPL reconstructions are more enriched with signaling interactions and the paths constituting interactions, from a receptor to a TR, are spatially coherent with the signaling cellular compartments, leading to more accurate and biologically meaningful reconstructions. *LocPL* can be widely applied to other interactome-based analyses since it is (i) a post-processing method after constructing the signaling pathways and (ii) based on the localization score of the edge nodes, 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 extend *LocPL* to compare paths under different conditions, such as dysregulated signaling due to diseases.

Contact Information

Web: http://www.reed.edu/biology/ritz Email: aritz@reed.edu youssefi@reed.edu

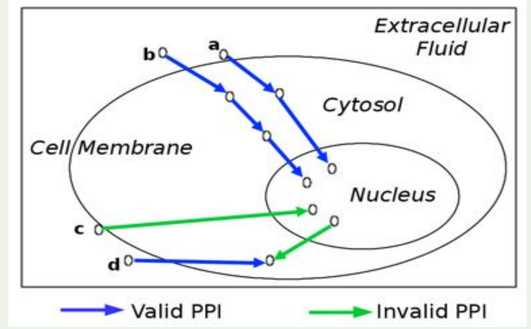
Methods

Localized PPIs from ComPPI:

- ComPPI computes: (i) a protein *localization score* describing the likelihood of a protein to be found in a major sub-cellular compartment, and (ii) an *interaction score* representing the likelihood that an interaction takes place inside the cell.
- We filter the interactome by keeping only PPIs with non-zero ComPPI *interaction scores*.

Signaling Assumptions:

- For every PathLinker path $P_i \in \mathcal{P}$ connecting m proteins, $P = (v_1, v_2, \ldots, v_m)$, we compute a signaling score $s(P_i)$ using a tailored signaling-based dynamic program.
- We use three cellular compartments: extracellular fluid or cell membrane (*ExtMem*), *Cytosol*, and *Nucleus*.
- We only consider intracellular signaling that begins with activation of a membrane-bound protein receptor and is transmitted to a DNA-binding transcription factor through PPIs within the cytosol.
- We assume a fixed unidirectional signaling flow from *ExtMem* through *Cytosol* to *Nucleus*.
- Multiple interactions per compartment are allowed.



From Compartment	To Compartment
ExtMem	*ExtMem*
ExtMem or *Cytosol*	*Cytosol*
Cytosol or *Nucleus*	*Nucleus*

Dynamic Program for Path-Based Signaling Scores:

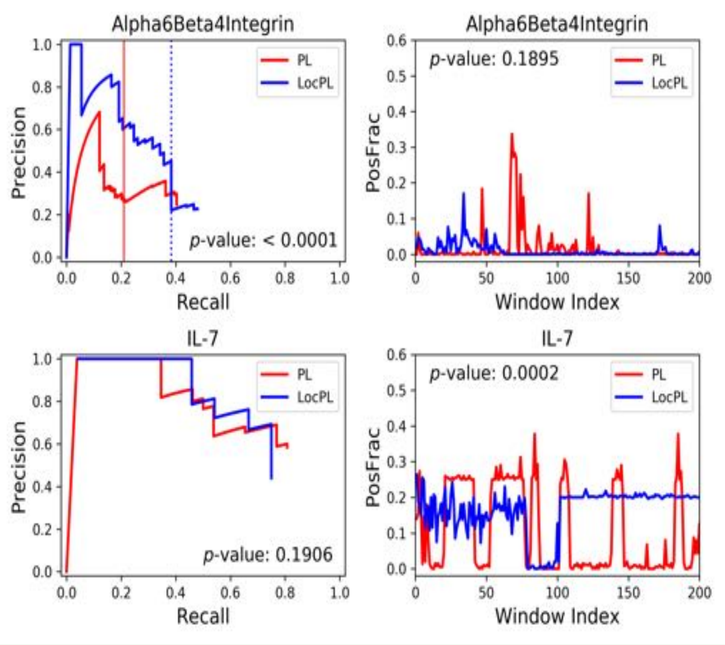
- Let ℓ_v^t denote the ComPPI localization score of protein v for the cellular compartment $t = \{ExtMem, Cytosol, Nucleus\}$.
- $\mathcal{P}$ is partitioned into ties: $\mathcal{P} = [P_{c_1}, \ldots, P_{c_n}]$, where $P_{c_j} \subseteq \mathcal{P}$ is the group of paths sharing the same path cost c_j, 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 $s(P_i)$, the signaling score.

- Log-transform localization scores: $\ell_v^t = -\log \ell_v^t$.
- At node v_j, $j = 2, 3, \ldots, (m-1)$,
  $s(v_j, ext) = s(v_{j-1}, ext) + \ell_{v_j}^{ext}$
  $s(v_j, cyt) = \min[s(v_{j-1}, ext), s(v_{j-1}, cyt)] + \ell_{v_j}^{cyt}$
  $s(v_j, nuc) = \min[s(v_{j-1}, cyt), s(v_{j-1}, nuc)] + \ell_{v_j}^{nuc}$
- The final path signaling score:
  $s(v_m, nuc) = \min[s(v_{m-1}, cyt), s(v_{m-1}, nuc)] + \ell_{v_m}^{nuc}$

Results

We evaluate performance in two ways:

1. **Precision-Recall (PR) curve:** this provides a global assessment across all the k-shortest paths. It shows how quickly (in terms of k) the technique can discover new positive edges and what percentage of the positives are discovered by reaching the k^{th} path.
2. **Fraction of positives per path (*PosFrac*):** this provides a path-based assessment by computing the within-path percentage of positives. The average of this metric across non-overlapping windows of 100 paths each is plotted on the right.



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

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

[3] K. Kandasamy et al. NetPath: a public resource of curated signal transduction pathways. *Genome Biol.*, 11(1):R3, Jan 2010.

[4] D. Veres et al. ComPPI: a cellular compartment-specific database for protein–protein interaction network analysis. *Nucleic Acids Research*, 43(D1):D485–D493, 2015.