Connectivity Measures for Signaling Pathway Topologies
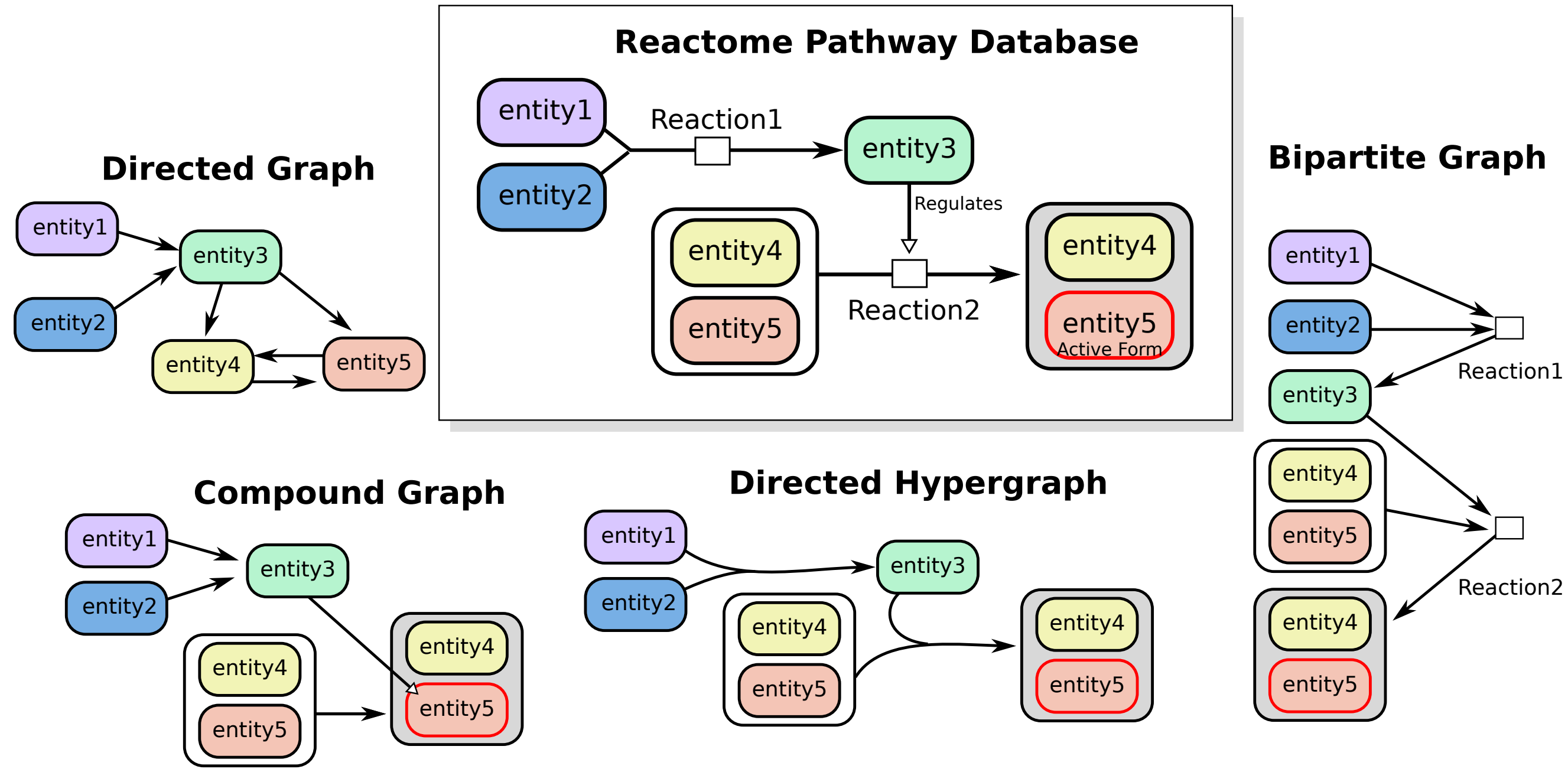
Nicholas Franzese[1,2,3†], Adam Groce[2], T. M. Murali[3], and Anna Ritz[1]

[1]Biology and [2]Mathematics Departments, **Reed College**, Portland, OR, USA
[3]Computer Science Department, **Virginia Tech**, Blacksburg, VA, USA
[†]Current Affiliation: CS Department, University of Maryland College Park, College Park, MD, USA

Background: Signaling Pathway Representations

When are two molecules connected in a network? We computed distances between all pairs of nodes in Reactome representations.

1. **Directed graphs** and **bipartite graphs** use breadth first search.
2. **Compound graphs** use a traversal with biologically-inspired rules [1].
3. **Directed Hypergraphs** use the notion of B-connectivity [2]: a hyperedge is traversable w.r.t. a node s if all nodes in the tail (inputs) are reachable from s [2,3].
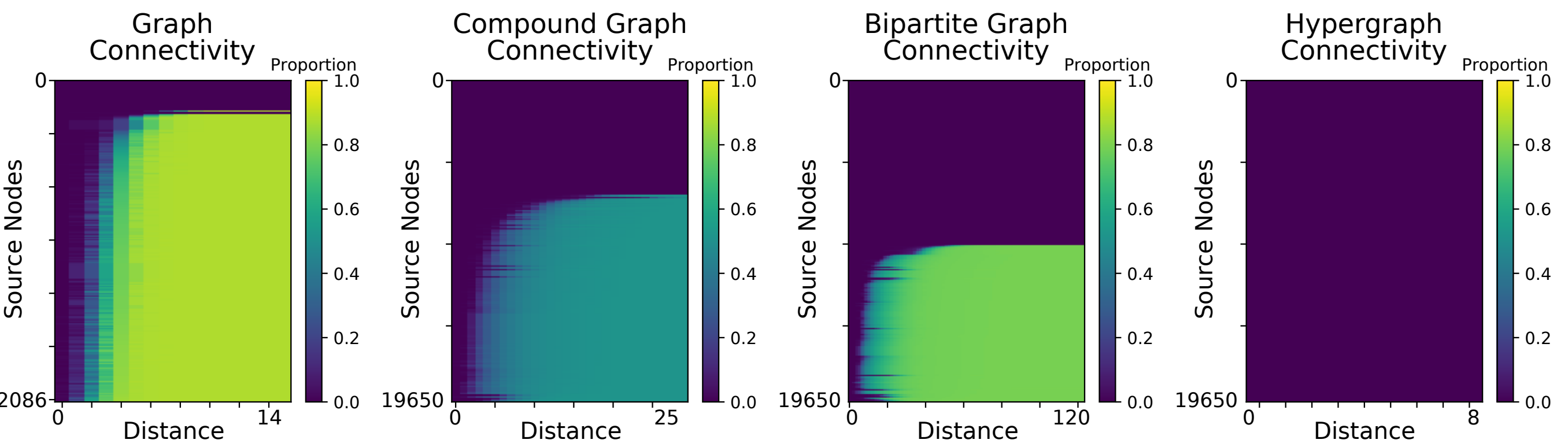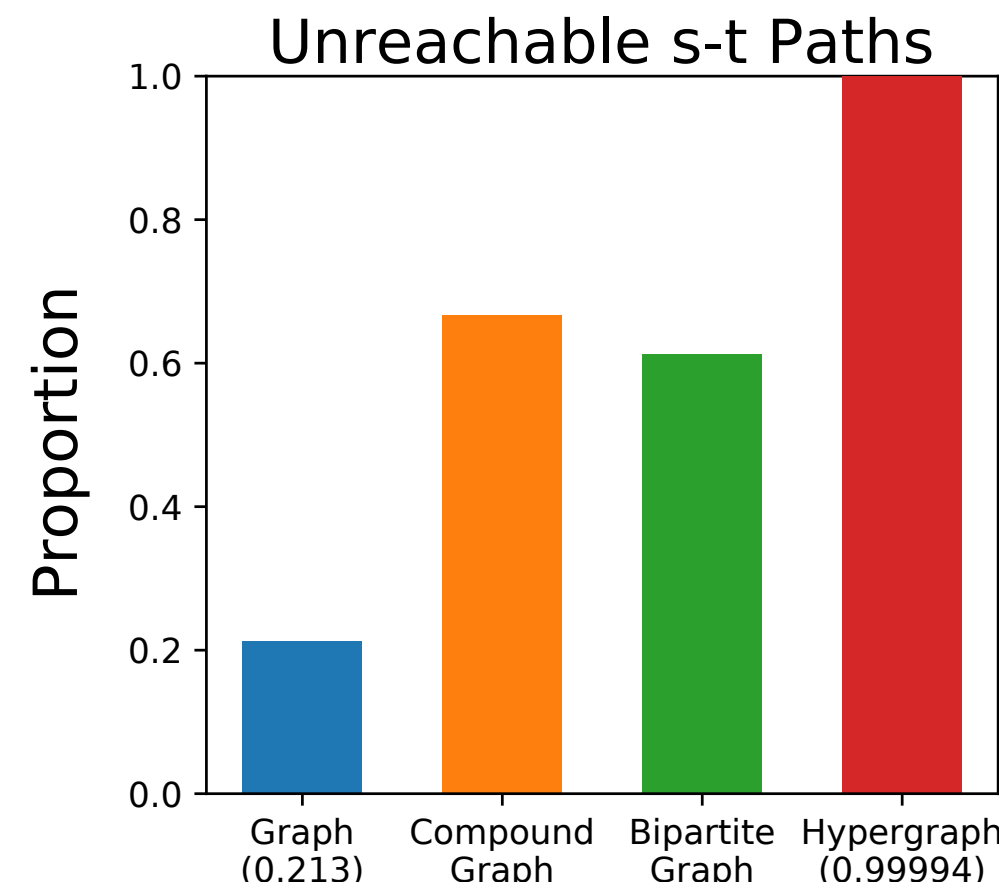

Hypergraphs are more informative than other representations, but B-connectivity is overly strict for Reactome's topology.

Methods: B-Relaxation Distance [4]

B-relaxation distance relaxes B-connectivity in hypergraphs while preserving hypergraph topology by iteratively lifting the B-connectivity restriction on hyperedges. Given a set S of source nodes, the algorithm computes two sets:

1. B_0: nodes B-connected to S.
2. R_0: *restrictive* hyperedges where at least one tail node is in B_0 but it was not traversed.
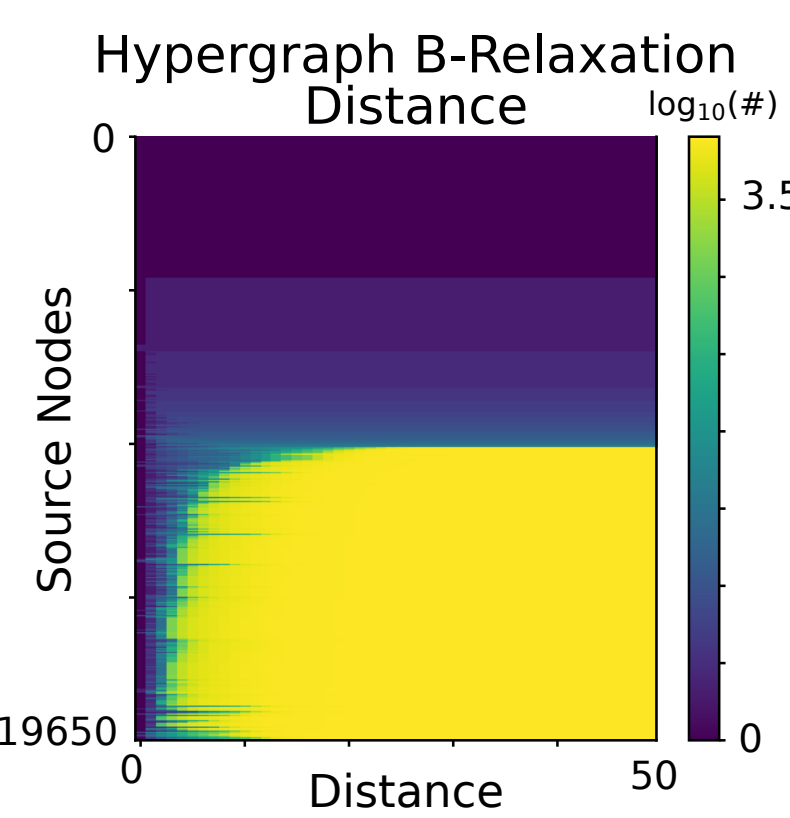
The algorithm then computes B_1 and R_1 from the heads of each restrictive hyperedge in the previous iteration (e.g., R_0). This process continues until no new nodes are reachable; this set is exactly the set of reachable nodes in the bipartite graph.

The B-Relaxation Distance of a node v from a source set S is the smallest value of k where v is in B_k.
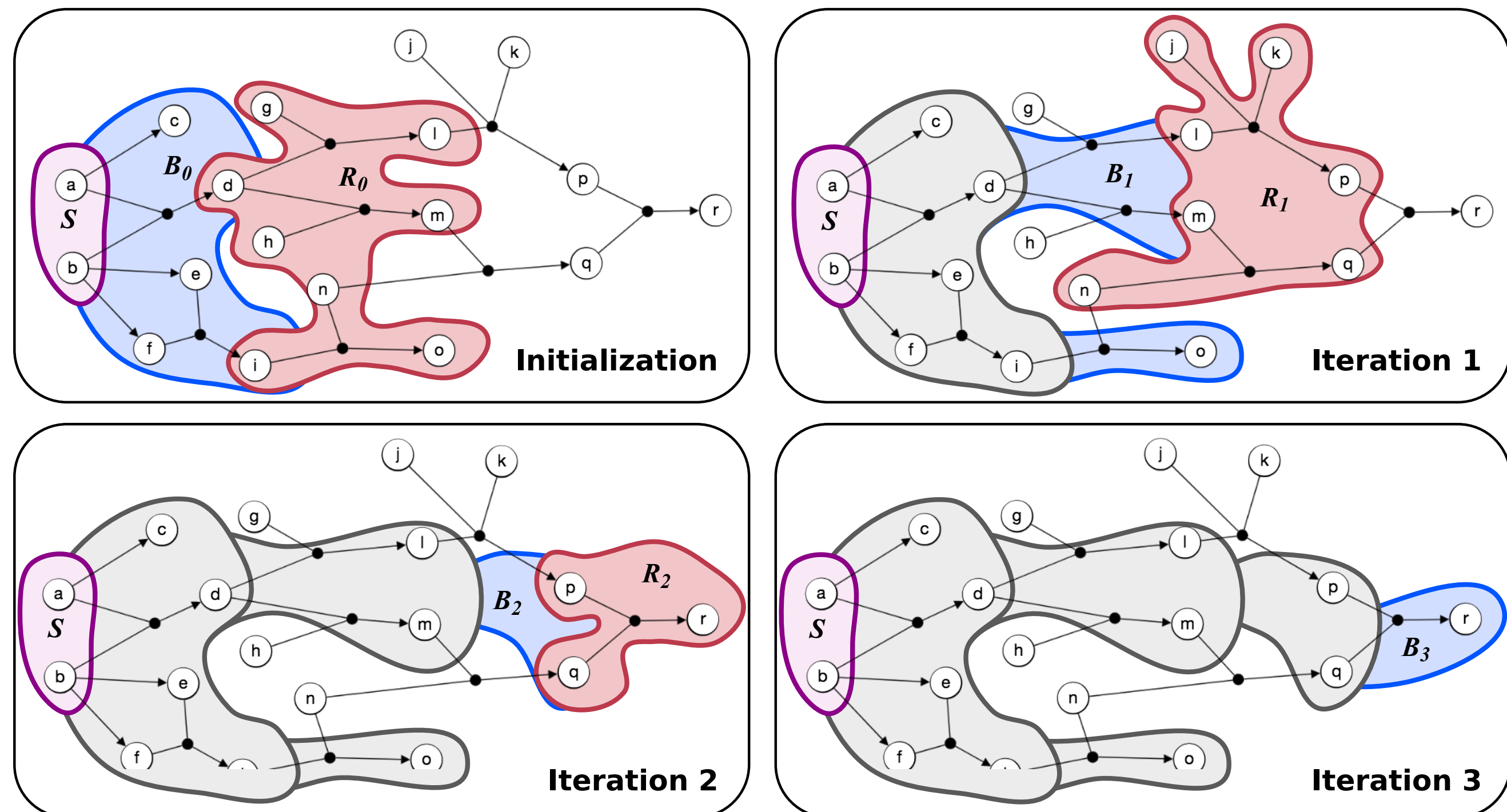
## References & Acknowledgements

Refs: [1] Dogrusoz et al, *BMC Bioinf.* 2009; [2] Gallo et al., *Discrete Appl. Math.* 1993; [3] Ritz et al., *IEEE TCBB* 2017; [4] Ritz et al., *bioRxiv* 2019.

Results: Pathway influence across Reactome

We compute the *influence score* s_k of a source pathway S (w/ nodes P_S) on target pathway T (w/ nodes P_T) for B-relaxation distance up to k (denoted $B_{\leq k}$):

$$s_k = \frac{|\overbrace{(B_{\leq k}(P_S) \cap P_T)}^{\text{"reachable" overlap}} \setminus \overbrace{(P_S \cap P_T)}^{\text{initial overlap}}|}{|B_{\leq k}(P_S) \setminus (P_S \cap P_T)|}.$$

We compute the significance of s_k with a permutation test that preserves the node overlap of all 34 Reactome pathways.
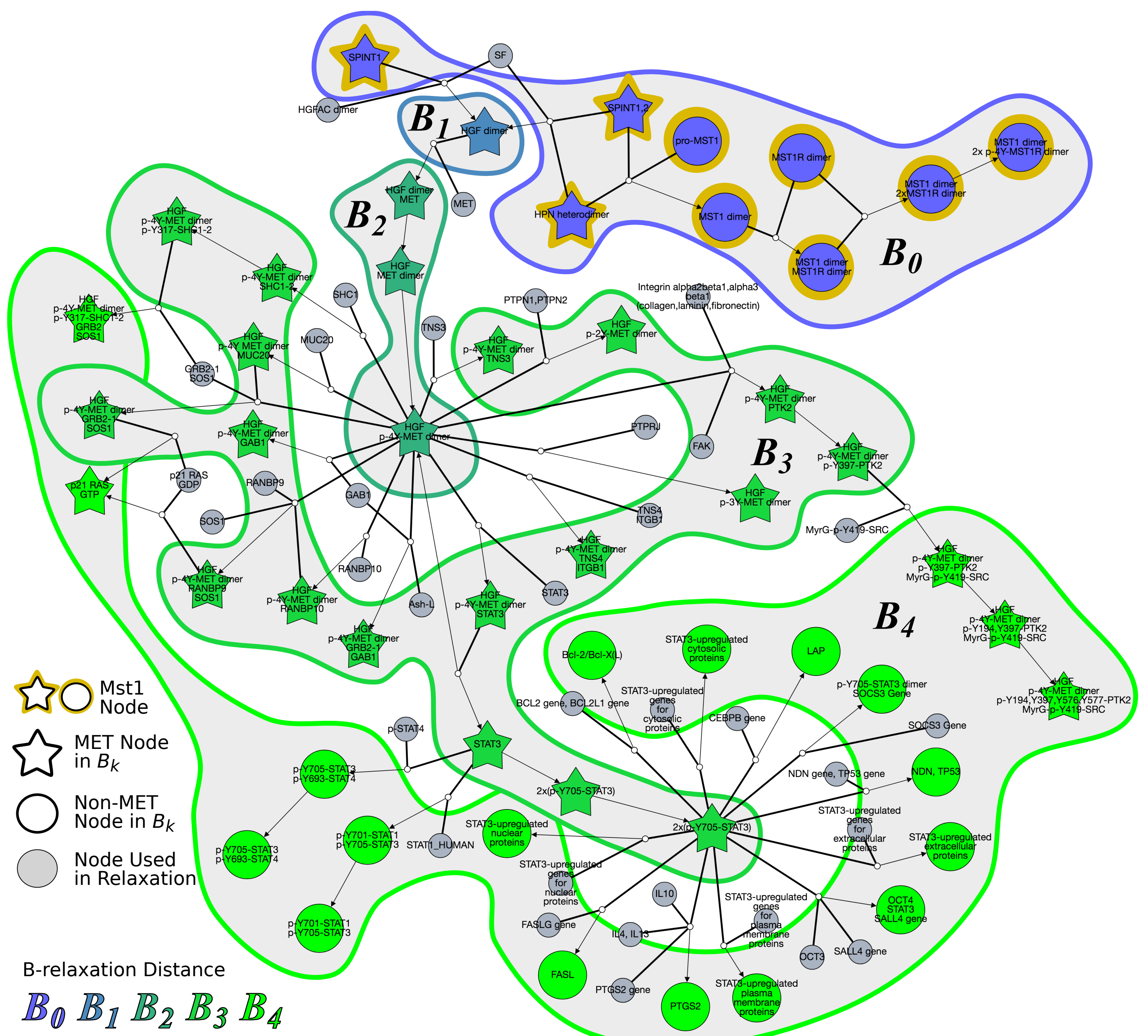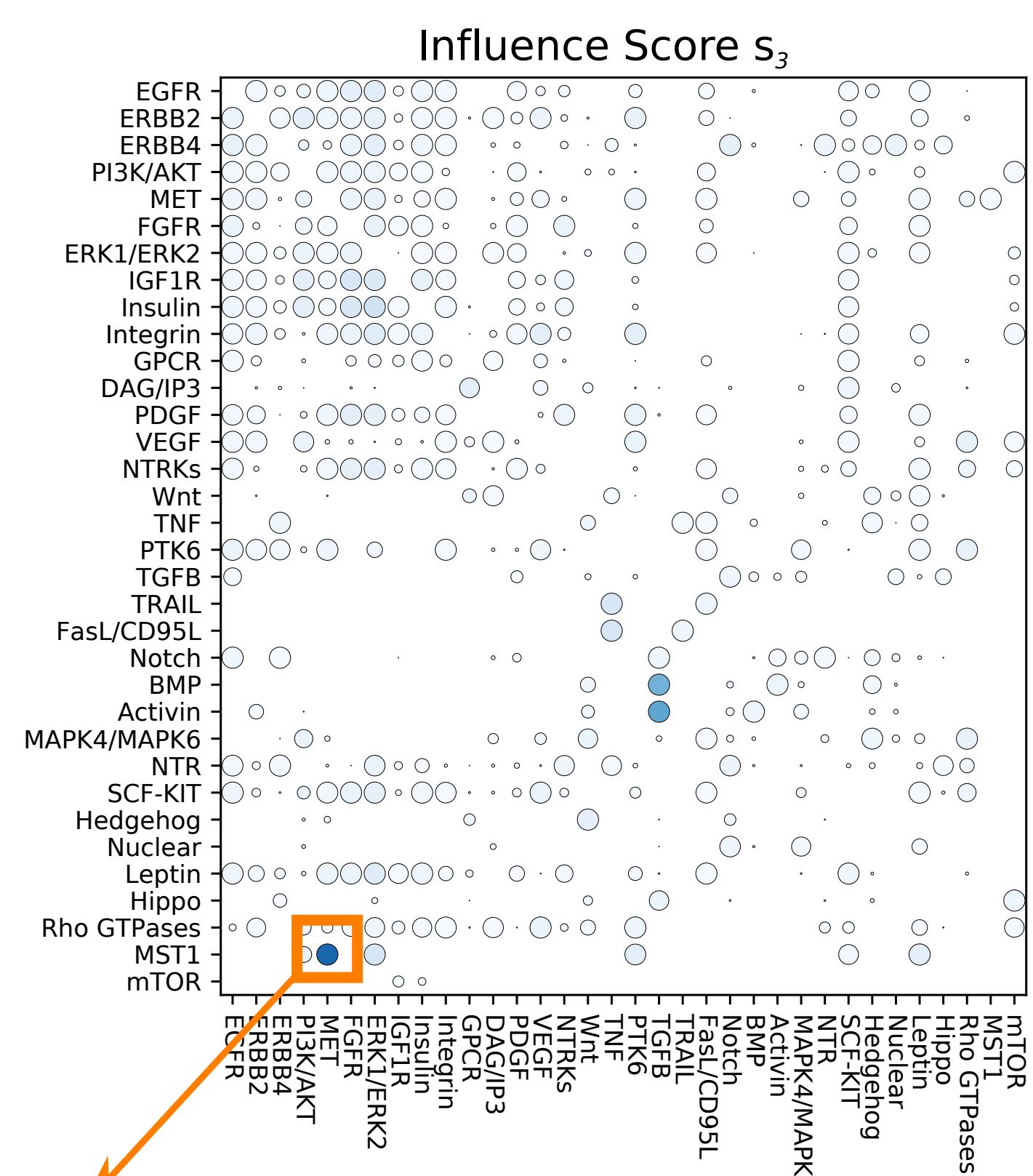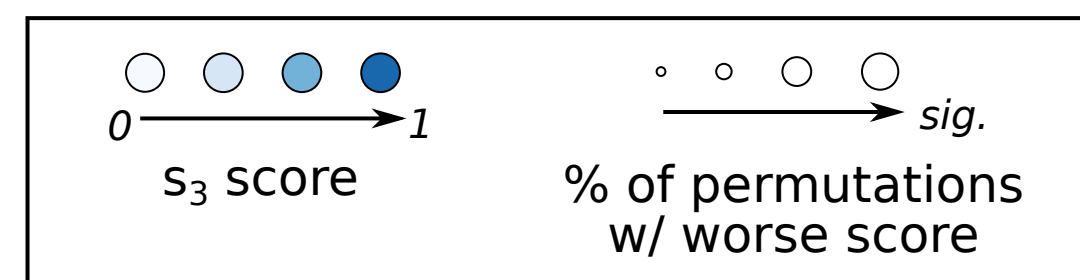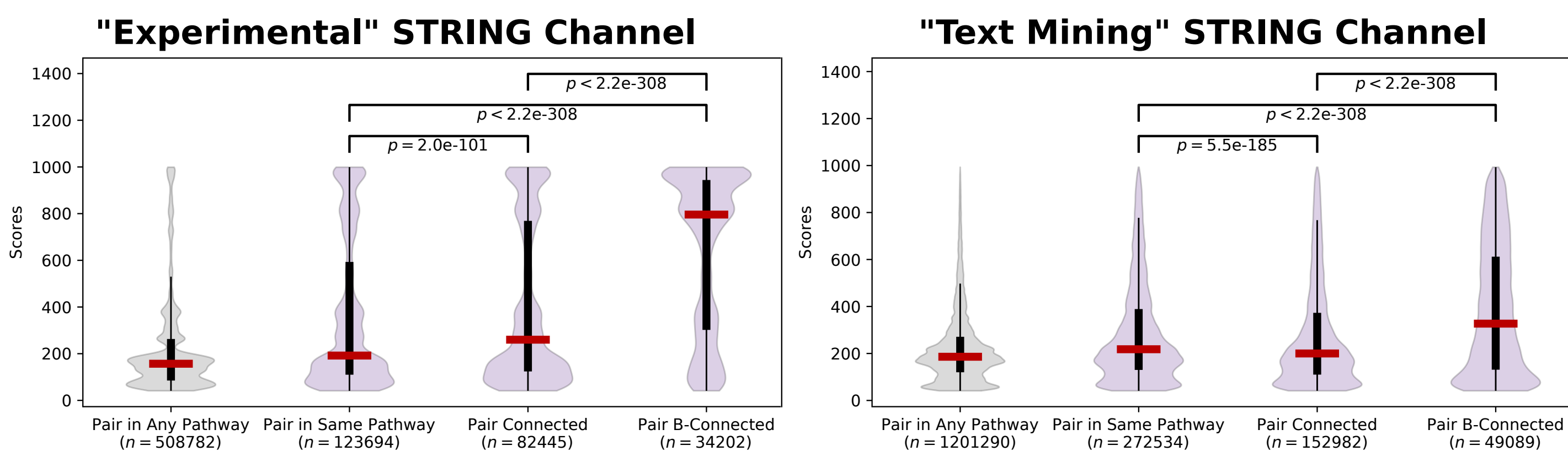

Case Study: Mst1 Pathway Influence on MET Signaling

Results: Benchmarking STRING functional relationships

We examined the enrichment of STRING interaction scores for node pairs with different definitions of connectivity: (1) node pairs in any Reactome pathway; (1) node pairs in the *same* Reactome pathway; (3) node pairs connected in the Reactome bipartite graph; and (4) node pairs connected in the Reactome hypergraph using B-connectivity.

Node pairs that are B-connected in Reactome are enriched for higher-scoring STRING interactions across different channels.

Availability