

# Supplementary Information for Connectivity Measures for Signaling Pathway Topologies

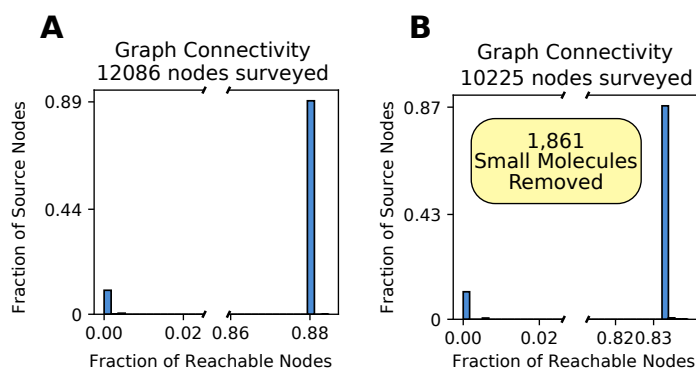
Nicholas Franzese<sup>\*1,2,3</sup>, Adam Groce<sup>2</sup>, T. M. Murali<sup>3,4</sup>, and Anna Ritz<sup>†1</sup>

<sup>1</sup>Department of Biology, Reed College, Portland, OR, US

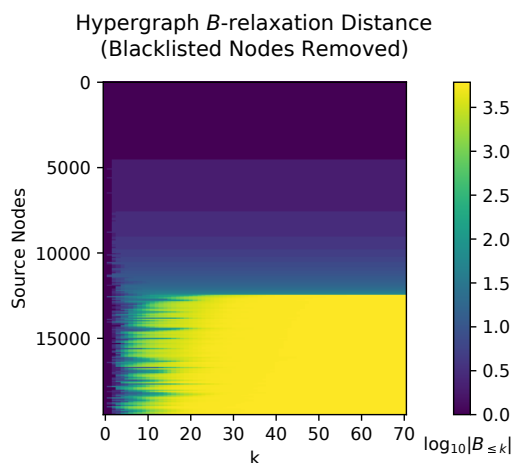
<sup>2</sup>Department of Computer Science, Reed College, Portland, OR, US

<sup>3</sup>Department of Computer Science, Virginia Tech, Blacksburg, VA, US

<sup>4</sup>ICTAS Center for Systems Biology of Engineered Tissues, Virginia Tech, Blacksburg, VA, US



**Fig S1.** Histograms of the fraction of nodes reached by each source node in the (A) full directed graph and the (B) directed graph after removing small molecules. We removed small molecules from the graph in Panel (B) by removing nodes that did not have UniProtKB protein identifier.



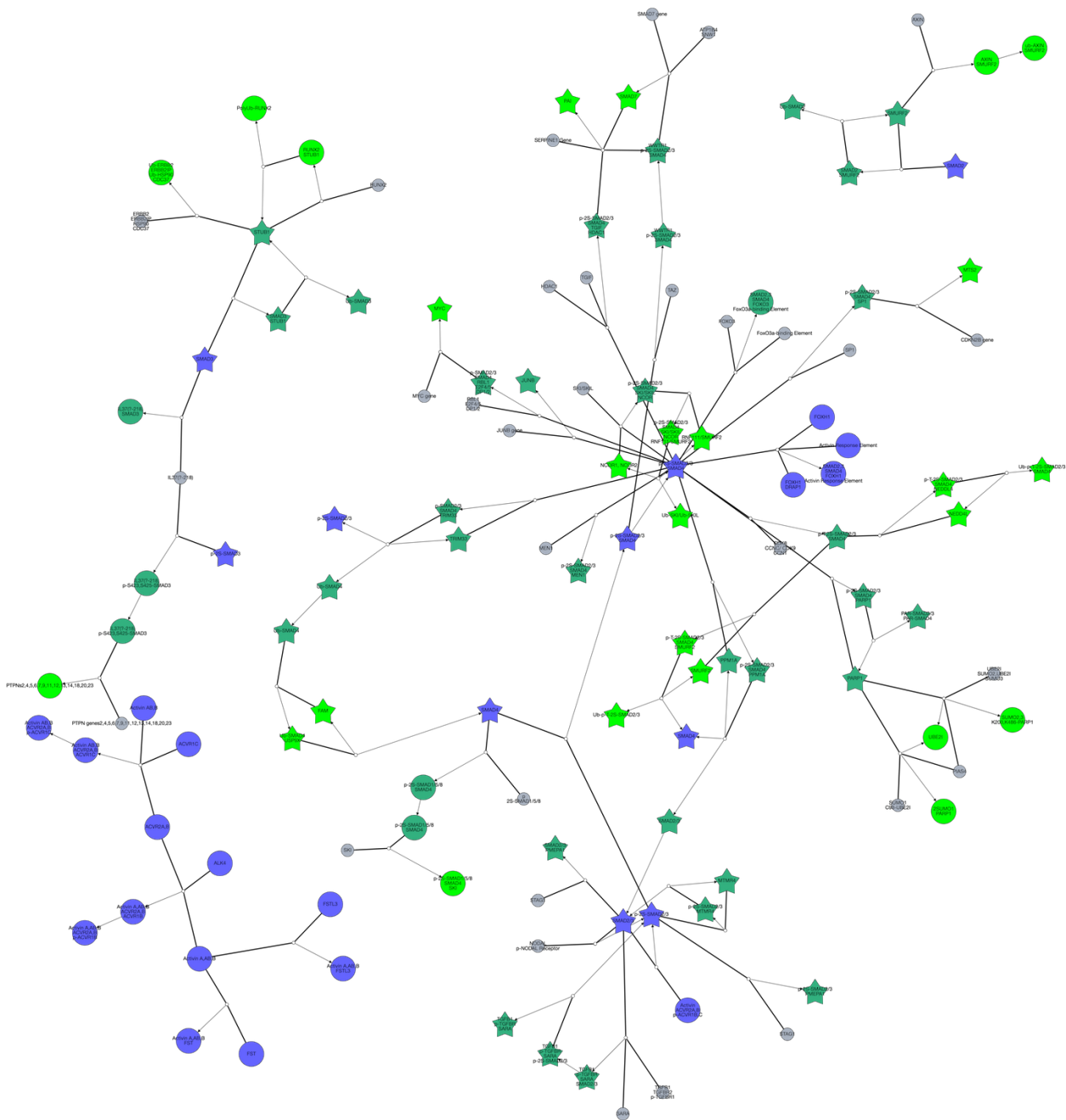
**Fig S2.**  $B$ -relaxation distance survey from each node in the hypergraph using PathwayCommons' blacklisted molecules. The heatmap shows the number of nodes  $|B_{\leq k}|$  in the  $B_k$ -connected set from each source node (rows) for values of  $k$  (columns) in the hypergraph after removing blacklisted small molecules from PathwayCommons v10 [?].

<sup>\*</sup>Current Affiliation: Department of Computer Science, University of Maryland College Park, College Park, MD, US

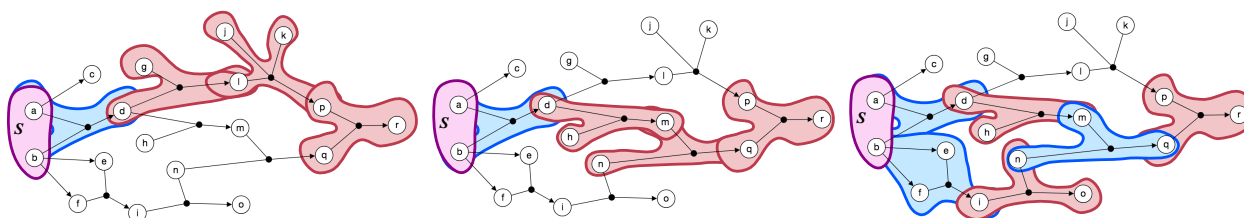
<sup>†</sup>Corresponding Author: aritz@reed.edu

Signaling Pathway	Reactome ID	# in Pathway	# in Hypergraph	# in Filtered Hypergraph
EGFR	R-HSA-177929	164	110	101
ERBB2	R-HSA-1227986	220	118	105
ERBB4	R-HSA-1236394	181	101	92
PI3K/AKT	R-HSA-1257604	753	357	333
MET	R-HSA-6806834	240	130	121
FGFR	R-HSA-190236	373	232	221
ERK1/ERK2	R-HSA-5684996	656	310	291
IGF1R	R-HSA-2404192	160	75	64
Insulin	R-HSA-74752	207	90	76
Integrin	R-HSA-9006921	88	66	53
GPCR	R-HSA-372790	2456	1006	796
DAG/IP3	R-HSA-1489509	107	50	32
PDGF	R-HSA-186797	244	88	80
VEGF	R-HSA-194138	351	215	183
NTRKs	R-HSA-166520	362	205	184
Wnt	R-HSA-195721	921	358	326
TNF	R-HSA-75893	127	97	93
PTK6	R-HSA-8848021	187	122	114
TGFB	R-HSA-170834	247	179	166
TRAIL	R-HSA-75158	17	13	13
FasL/CD95L	R-HSA-75157	10	10	10
Notch	R-HSA-157118	505	286	271
BMP	R-HSA-201451	74	38	36
Activin	R-HSA-1502540	44	30	28
MAPK4/MAPK6	R-HSA-5687128	200	113	105
NTR	R-HSA-193704	245	141	129
SCF-KIT	R-HSA-1433557	122	86	79
Hedgehog	R-HSA-5358351	417	190	172
Nuclear	R-HSA-9006931	431	261	208
Leptin	R-HSA-2586552	54	33	31
Hippo	R-HSA-2028269	79	47	45
Rho GTPases	R-HSA-194315	808	330	291
MST1	R-HSA-8852405	21	12	9
mTOR	R-HSA-165159	111	67	57

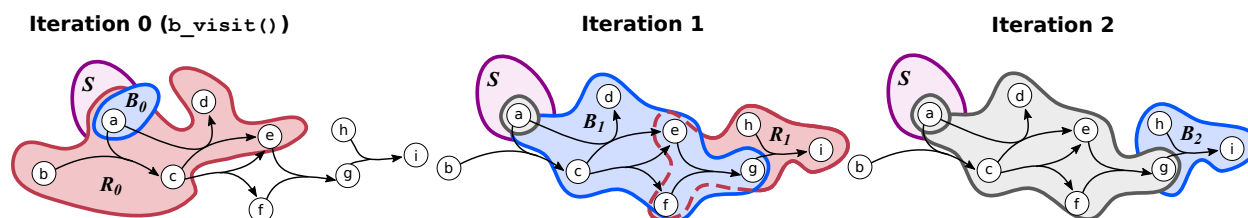
**Table S1. Thirty-four Reactome signaling pathways considered for the pathway influence analysis.** Members that are not part of any hyperedge are ignored from the hypergraph. The filtered hypergraph has removed all small molecules, two forms of Ubiquitinase, and the Nuclear Pore Complex from the hyperedges.



**Fig S3. Hyperedges traversed to compute  $B_0, B_1, \dots, B_4$  from source pathway Activin.** Node colors represent  $B$ -relaxation distance from  $k = 0$  ( $B$ -connected set, blue) to  $k = 3$  (bright green). Gray nodes are entities that are not in the  $B_k$ -connected set but are involved in traversed hyperedges. Star-shaped nodes are members of the TGF $\beta$  pathway. This network is available on GraphSpace at [http://graphspace.org/graphs/26756?user\\_layout=6713](http://graphspace.org/graphs/26756?user_layout=6713).



**Fig S4.** Examples of connectivity from  $S = \{a, b\}$  to  $r$  with a  $B$ -relaxation distance of three. Blue hyperedges denote traversals that are consistent with  $B$ -connectivity; red hyperedges denote traversals where one, but not all, nodes in the tail are connected; only hyperedges that are involved in the connectivity from  $S$  to  $r$  are highlighted for simplicity. Note that while  $B$ -relaxation distance is three, there are different sets of hyperedges that achieve this  $B$ -relaxation distance.



**Fig S5.** The restrictive set  $R_k$  may include hyperedges that have been traversed in a previous iteration's `b_visit()` call. In iteration 1, the restrictive set  $R_1$  is established by considering the  $B$ -connectivity from the heads of the two hyperedges in  $R_0$ . The hyperedge  $\{\{e, f\}, \{g\}\}$  is restrictive with respect to the heads of one hyperedge in  $R_0$  but traversable with respect to the heads of the other hyperedge. Thus,  $\{\{e, f\}, \{g\}\}$  is included in  $R_1$  but also added to the seen dictionary, saving redundant computation in Algorithm 2.

Binary Relation	# of Relations in Reactome	Conversion Rule
catalysis-precedes	49,435	Directed Edge
chemical-affects	12,219	Directed Edge
consumption-controlled-by	4,914	Directed Edge
controls-expression-of	3,730	Directed Edge
controls-phosphorylation-of	3,184	Directed Edge
controls-production-of	5,266	Directed Edge
controls-state-change-of	106,958	Directed Edge
controls-transport-of	5,218	Directed Edge
controls-transport-of-chemical	1,882	Directed Edge
in-complex-with	140,097	Undirected Edge
reacts-with	568	Undirected Edge
used-to-produce	3,271	Directed Edge

**Table S2.** Rules for converting SIF binary relations to directed edges. We ignore the “neighbor-of” binary relation.