

Augmenting Signaling Pathway Reconstructions

Tobias Rubel

Biology Department, Reed College
Philosophy Department, Reed College
rubelato@reed.edu

Anna Ritz

Biology Department, Reed College
aritz@reed.edu

ABSTRACT

Signaling pathways drive cellular response, and understanding such pathways is fundamental to molecular systems biology. A mounting volume of experimental protein interaction data has motivated the development of algorithms to computationally reconstruct signaling pathways. However, existing methods suffer from low recall in recovering protein interactions in ground truth pathways, limiting our confidence in any new predictions for experimental validation. We present the Pathway Reconstruction AUGmenter (PRAUG), a higher-order function for producing high-quality pathway reconstruction algorithms. PRAUG modifies any existing pathway reconstruction method, resulting in augmented algorithms that outperform their un-augmented counterparts for six different algorithms across twenty-nine diverse signaling pathways. The algorithms produced by PRAUG collectively reveal potential new proteins and interactions involved in the Wnt and Notch signaling pathways. PRAUG offers a valuable framework for signaling pathway prediction and discovery.

CCS CONCEPTS

• **Applied computing** → **Systems biology**; **Computational biology**; • **Theory of computation** → **Graph algorithms analysis**.

KEYWORDS

Systems biology, signaling pathways, graph algorithms, pathway reconstruction

ACM Reference Format:

Tobias Rubel and Anna Ritz. 2020. Augmenting Signaling Pathway Reconstructions. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '20)*, September 21–24, 2020, Virtual Event, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3388440.3412411>

1 INTRODUCTION

Signaling pathways describe the series of molecular interactions that occur in response to a certain stimulus, which mediates the expression of relevant genes. Understanding the specific reactions that occur within pathways is a fundamental question in molecular systems biology and through decades of research we now have some

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

BCB '20, September 21–24, 2020, Virtual Event, USA

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-7964-9/20/09...\$15.00
<https://doi.org/10.1145/3388440.3412411>

understanding of how cells grow, proliferate, and die. Many of these reactions are stored in databases that characterize pathways such as NetPath [13], KEGG [14], Reactome [11], and dozens of others. Biologists have extensively used these databases to help determine which pathways are activated or inhibited in high-throughput gene expression experiments. Despite the promise of pathway database resources, we know that they are incomplete. The same pathway can be represented very differently across databases, and they often have little overlap among the proteins and interactions. There are also new discoveries about protein membership and interactions involved in canonical signaling pathways that are not yet in these databases. Further, pathways in non-model systems or under-studied pathways are often missing from these resources.

In the past decade, there have been two emerging solutions for improving pathway databases. First, pathways have been integrated into resources that attempt to capture the current knowledge base of signaling. Repositories such as WikiPathways [17] and PathwayCommons [23] now contain thousands of pathways comprised of millions of interactions by aggregating information from other databases. WikiPathways also offers a community-driven platform, allowing researchers to submit and curate pathways [17].

While these resources are useful for exploration, they are not intended to make predictions about *new* proteins and interactions that may be associated with a pathway of interest. Another area of signaling pathway research develops new methods that can predict new players in signaling pathways of interest. Many of these approaches generate new predictions by integrating protein-protein interaction data with gene [7, 12, 21, 21, 29, 30, 33] or protein [4, 15, 19, 24] expression. Other approaches work to remove biologically implausible predictions [20, 34]. Here, we will focus on methods that use protein-protein interaction data to predict new proteins and molecular interactions involved in canonical signaling pathways.

1.1 Network-based Pathway Reconstruction

Protein-protein interactions can be modeled as graphs where the nodes stand in for proteins and the edges represent a (possibly directed) interaction between two proteins. These graphs, called *interactomes*, can be built from experimental dataset repositories and can be weighted by confidence in the interaction, functional relationship, or tissue [1, 2, 5, 6, 22, 26, 32, 34]. Interactomes provide a background set of plausible interactions for consideration in a pathway of interest, and network-based methods generally have been successful with amplifying the signal of interacting proteins [9].

We are interested in predicting new proteins and interactions involved with a particular pathway of interest. The **Pathway Reconstruction Problem** is summarized as the following [22]: Given an interactome, a set of receptors in a specific pathway, and a set of transcriptional regulators in the pathway, recover the intermediate proteins and interactions responsible for transmitting the signal

from the receptors to the transcriptional regulators. The pathway-specific receptors and transcriptional regulators may come from proteins in existing pathway databases [22, 25] or from experimental data [21, 21, 29, 30, 33].

1.2 Contributions

We present the Pathway Reconstruction AUGmenter (PRAUG), a higher-order function which maps any pathway reconstruction method to an augmented method that improves pathway reconstruction performance. PRAUG is designed based on the observation that pathway reconstruction methods typically perform well when predicting the proteins in a pathway. PRAUG takes as input a pathway reconstruction method and provides a method which uses a traversal on the input method's predicted nodes to explore protein interactions. Despite PRAUG's simplicity, this augmentation improves the protein interaction accuracy of almost any pathway reconstruction method that is used to seed the algorithm. Thus, PRAUG can serve as a framework to boost the performance of any existing method for the Pathway Reconstruction Problem. We highlight PRAUG's generalizability by running it on an interactome of 612516 weighted edges to recover 29 diverse human signaling pathways from the NetPath pathway database [13], using 6 state-of-the-art pathway reconstruction methods as seeds. We highlight the improved pathway reconstructions using a case study of the Wnt and Notch pathways.

2 METHODS

We first describe PRAUG and the pathway reconstruction methods used as input to PRAUG. We then describe the interactome and pathway datasets, and details about assessment.

2.1 PRAUG

We are given a potentially directed interactome $G = (V, E)$, which may have weights w_{uv} for every edge $(u, v) \in E$. A pathway $P = (V_P, E_P)$ can be described as a subgraph of G (e.g. $V_P \subseteq V$ and $E_P \subseteq E$). Pathway P includes a set $S \subseteq V_P$ of receptors and a set $T \subset V_P$ of transcriptional regulators (TRs). A solution to the Pathway Reconstruction Problem connects the *sources* S to the *targets* T through the interactome G .

Many existing pathway reconstruction methods offer a solution to the Pathway Reconstruction Problem. Here, we focus on the space of methods $\mathbf{M}$ that take an interactome G , a set of sources S and targets T , and potentially other user-defined parameters which we denote $\{*\}$. These methods return a subgraph of G . Examples of such methods are outlined in Section 2.2. An instance of a method $\mathcal{M} \in \mathbf{M}$ will return a subgraph H of G :

$$\mathcal{M}(G, S, T, \{*\}) \rightarrow H = (V_{\mathcal{M}}, E_{\mathcal{M}}), \quad (1)$$

where $V_{\mathcal{M}} \subseteq V$ and $E_{\mathcal{M}} \subseteq E$. The edges of H are often ranked by the order in which they were found by $\mathcal{M}$, but may also be unranked.

PRAUG transforms a method $\mathcal{M} \in \mathbf{M}$ into another method $\widetilde{\mathcal{M}} \in \mathbf{M}$ (Figure 1). An instance of $\widetilde{\mathcal{M}}$ will return a subgraph $\widetilde{H}$ of G :

$$\widetilde{\mathcal{M}}(G, S, T, \{*\}) \rightarrow \widetilde{H} = (V_{\widetilde{\mathcal{M}}}, E_{\widetilde{\mathcal{M}}}), \quad (2)$$

where $V_{\widetilde{\mathcal{M}}} \subseteq S \cup T \cup V_{\mathcal{M}}$ and $E_{\widetilde{\mathcal{M}}} \subseteq E$. The PRAUG-augmented counterpart of $\mathcal{M}$ works as follows:

- (1) Run $\mathcal{M}(G, S, T, \{*\})$ and return a subgraph $H = (V_{\mathcal{M}}, E_{\mathcal{M}})$.
- (2) Define a node set $X = S \cup T \cup V_{\mathcal{M}}$.
- (3) Introduce a super source node σ to G and add edges from σ to every node $v \in X$.
- (4) Starting from σ , perform an unweighted depth-first traversal on the X -induced subgraph of G . Terminate when no new edge can be traversed.
- (5) Return the traversed edges as a subgraph $\widetilde{H} = (V_{\widetilde{\mathcal{M}}}, E_{\widetilde{\mathcal{M}}})$, with edges sorted by their traversal order.

$\widetilde{\mathcal{M}}$ uses the nodes predicted by $\mathcal{M}$ as a scaffold to find edges that connect predicted nodes (Figure 1). Some of these edges may not have been discovered by the original $\mathcal{M}$, which may potentially include known interactions for a particular pathway. Since the sources S and targets T depend on the specific pathway P that we wish to reconstruct, we parameterize the methods by the pathway P instead of S and T . We also drop the interactome G from the parameterization for simplicity, leaving the methods parameterized as $\mathcal{M}(P, \{*\})$ and $\widetilde{\mathcal{M}}(P, \{*\})$.

2.2 Pathway Reconstruction Methods

We use six different pathway reconstruction methods as inputs for PRAUG (Table 1).

2.2.1 PathLinker (PL) [22]. This method computes the k shortest paths from any source S to any target T in a weighted, directed interactome. PathLinker uses an A^* speedup to the classic Yen's k -shortest loopless paths algorithm [22]. PL requires a number of shortest paths $k \in \mathbb{Z}^+$ (by default $k = 500$), and edges are ranked in increasing order by the first path in which they appear.

2.2.2 ResponseNet (RN) [7, 33]. This method formulates the pathway reconstruction problem as a network flow algorithm, and presents a linear program to find a subgraph that balances outgoing flow from the sources and incoming flow to the targets. RN requires a sparsity parameter $\gamma \in \mathbb{R}^+$ that penalizes flow through multiple sources (by default $\gamma = 20$). The unranked set of edges with positive flow are considered the predicted pathway reconstruction.

2.2.3 BowTieBuilder (BTB) [25]. This method iteratively connects sources to targets using short paths that aim to determine an hourglass, or "bowtie" structure in the pathway reconstruction. The method terminates when as many sources and targets as possible have been added to the graph. The unranked set of edges are considered the predicted pathway reconstruction.

2.2.4 Prize Collecting Steiner Forest (PCSF) [29, 30]. This method places node prizes on the sources and targets in a weighted, undirected interactome and identifies a forest (set of trees) that connect prizes while simultaneously minimizing the edge costs [29]. We convert the directed, weighted interactome G into an undirected graph by converting any bidirected edge between nodes u and v to an undirected edge with the minimum cost of (u, v) and (v, u) ; the remaining directed edges simply become undirected. In addition to a prize $p \in \mathbb{R}^+$, the method also has a parameter $b \in \mathbb{R}^+$ that controls the tradeoff between including more prizes and taking higher-cost edges, a weight $\omega \in \mathbb{R}^+$ for dummy edges, and a degree

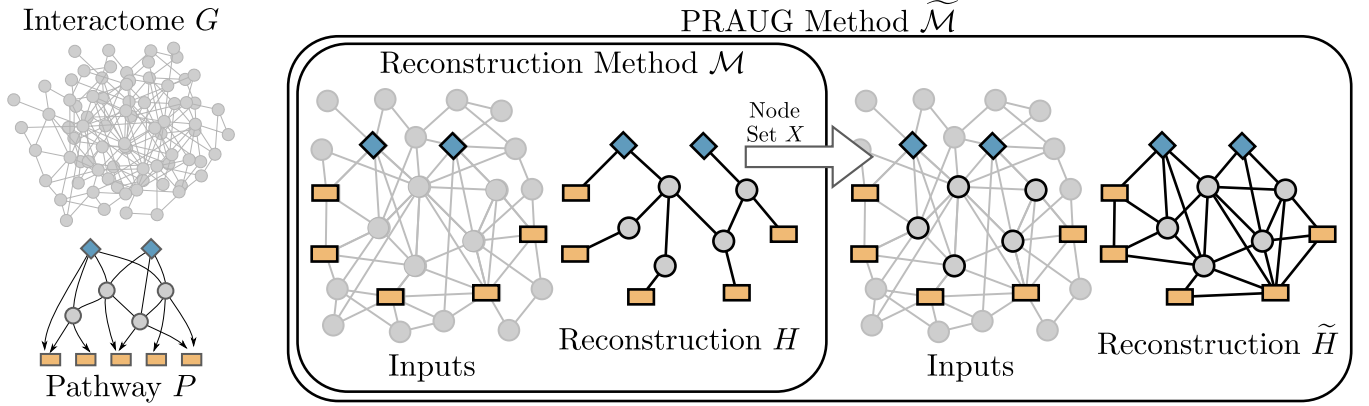


Figure 1: Overview of PRAUG. Pathway reconstruction methods take as input an interactome G and a pathway comprised of sources (blue diamonds) and targets (orange rectangles). Given a pathway reconstruction method $\mathcal{M}$, PRAUG defines a new method $\tilde{\mathcal{M}}$ that calls $\mathcal{M}$ and performs a traversal on the resultant node set.

penalty $g \in \mathbb{R}^+$. Default parameters are $b = 1, \omega = 5$, and $g = 3$. We use the optimal forest returned by Omics Integrator 2 [30], and take the unranked set of edges in the forest as the predicted pathway reconstruction.

2.2.5 Random Walk with Restarts (RWR). Inspired by TiedIE [21], we implemented a random walk with restarts using the following procedure. First, we run a random walk with restarts from the sources S , teleporting to sources uniformly at random. We then reverse the edges of the graph and run a random walk with restarts from the targets T , teleporting to targets uniformly at random. For a node $v \in V$, let the forward visitation probability be denoted as $p_F(v)$ and let the backward visitation probability be denoted as $p_B(v)$. We calculate the combined flux $f(u, v)$ for an edge (u, v) as

$$f(u, v) = \text{fwd}(u, v) \times \text{bkwd}(u, v), \text{ where} \quad (3)$$

$$\text{fwd}(u, v) = \frac{p_F(u)w_{uv}}{d^{\text{out}}(v)} \text{ and } \text{bkwd}(u, v) = \frac{p_B(v)w_{uv}}{d^{\text{in}}(u)}, \quad (4)$$

where $d^{\text{in}}(v)$ and $d^{\text{out}}(v)$ are the in-degree and out-degree of v , respectively. We return the number of edges that capture $\tau\%$ of the total $f(u, v)$ in the entire graph. RWR requires a damping factor $\alpha \in (0, 1]$ which is one minus the teleportation probability and a threshold $\tau \in (0, 1]$. Default parameters are $\alpha = 0.85$ and $\tau = 0.3$. We take the negative $\log(f(u, v))$ such that the predictions are ranked in increasing order.

2.2.6 All Pairs Shortest Paths. Finally, we also computed the shortest path from every source to every target and took the union of such paths as the reconstruction. If there are many tied paths between a source and a target, one path is arbitrarily chosen. The unranked set of edges in the shortest paths are considered the predicted pathway reconstruction.

2.3 Data

2.3.1 Interactome. We use PLNET₂, a weighted, directed interactome constructed from both molecular interaction data and signaling pathway databases (including the database used for ground truth pathways) [34]. PLNET₂ is weighted using an evidence-based

Method Name	Abbrev.	Parameters
PathLinker [22]	PL	number of shortest paths k
ResponseNet [7, 33]	RN	sparsity parameter γ
BowTieBuilder [25]	BTB	
Prize Collecting Steiner Forest [29, 30]	PCSF	terminal prize p edge reliability b dummy edge weight ω degree penalty g
Random Walk with Restarts (insp. by [21])	RWR	teleportation probability α flux threshold τ
Shortest Paths	SP	

Table 1: Pathway Reconstruction Methods.

Bayesian method introduced by RN [33] that assigns a high confidence to edges supported by experimental methods that successfully predict signaling interactions [34]. PLNET₂ contains 17168 nodes (UniProtKB identifiers [8]) and 612516 directed edges, including 286520 physical interactions that are converted to bidirected edges.

2.3.2 Signaling Pathways. We consider a set of 29 signaling pathways from the NetPath database, a repository of cancer and immune related pathways [13]. Sources and targets are automatically detected for each of the pathways from curated lists of human receptors and transcriptional regulators (see [22] for more details). The Advanced Glycation End Products (AGE/RAGE) pathway, Inhibitor of Differentiation (ID) pathway, and Interleukin 11 (IL-11) pathway did not have an automatically-detected receptor and/or a transcriptional regulator, and were removed from consideration. Twenty-nine remaining pathways are listed in Table 2.

2.4 Assessment

2.4.1 Precision/Recall. We evaluate reconstruction methods based on their precision as well as their recall. Recall that Section 2.1 defines an interactome $G = (V, E)$, a pathway $P = (V_P, E_P)$, and a method $\mathcal{M}$'s prediction $H = (V_{\mathcal{M}}, E_{\mathcal{M}})$. Let N_P be a set of negative edges from E with respect to pathway P ; that is, N_P contains interactions that are unlikely to be related to signaling and are disjoint from E_P . The negatives N_P need to be selected carefully, as

Pathway ($n = 29$)	# Nodes			# Edges	
	Total	S	T	Dir.	Undir.
$\alpha 6 \beta 4$ Integrin	66	7	3	223	116
Androgen Receptor	165	2	75	491	251
BCR	137	1	18	456	261
BDNF	72	5	4	139	76
CRH	24	2	8	54	27
EGFR1	231	6	33	1456	756
FSH	19	2	1	30	18
Hedgehog	36	6	13	124	64
IL1	43	3	5	178	93
IL2	67	3	12	242	139
IL3	70	2	9	176	97
IL4	57	5	12	173	91
IL5	30	2	3	71	36
IL6	53	4	14	162	83
IL7	18	2	3	52	28
IL9	13	2	3	29	15
Kit	76	6	8	207	109
Leptin	55	3	15	135	74
Notch	74	4	27	255	154
OSM	37	5	12	82	42
Prolactin	68	4	10	199	103
RANKL	57	2	12	142	76
TCR	154	7	20	504	271
TWEAK	17	1	4	30	15
TSLP	7	1	2	14	7
TSH	48	2	6	90	47
TGF β	209	5	77	863	452
TNF α	239	4	44	913	473
Wnt	106	14	14	428	220

Table 2: NetPath pathways used in this study. Edges are taken as undirected for the assessment.

is described in more detail in Section 2.4.2. If $\mathcal{M}$ produces unranked predictions (as is the case for RN, BTB, PCSF, and SP), then the edge (or interaction) precision p_{edge} and recall r_{edge} of method $\mathcal{M}$ for pathway P are

$$p_{\text{edge}} = \frac{|E_{\mathcal{M}} \cap E_P|}{|E_{\mathcal{M}} \cap (E_P \cup N_P)|} \quad \text{and} \quad r_{\text{edge}} = \frac{|E_{\mathcal{M}} \cap E_P|}{|E_P|}.$$

For ranked predictions we index precision and recall to a rank i such that, for instance, $p_{\text{edge}}(i)$ is the precision for the top i predictions. Thus the precision and recall for ranked methods (such as PL and RWR) comprise many points. Node (protein) precision and recall equations are the same *mutatis mutandis*.

We assessed the performance of reconstruction methods by way of their maximum F_1 score. Given a precision p and a recall r ,

$$F_1 = 2 \cdot \frac{p \cdot r}{p + r}$$

The F_{max} is defined as the maximum F_1 for a method on a pathway. For unranked methods this is just the F_1 score of their single precision-recall point.

2.4.2 Negative Set. Because our task is not only to reconstruct known pathway interactions, but also to discover unknown interactions, we construct negatives N_P by sub-sampling interactions from

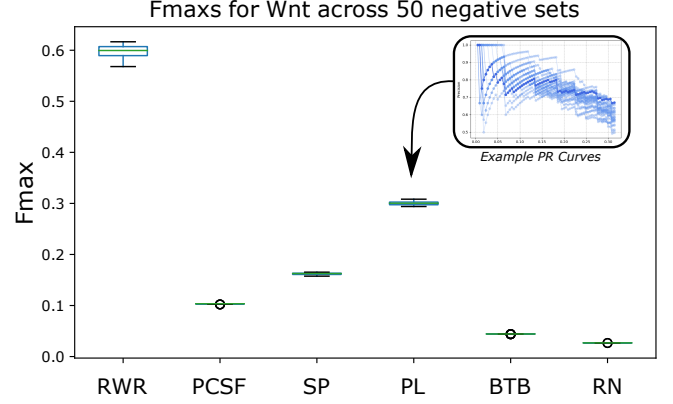


Figure 2: F_{max} scores for each method across 50 distinct sub-sampled negative sets for the Wnt pathway. Inset shows example precision-recall curves for PathLinker that are used to compute the F_{max} distribution.

E that are not in pathway P . Following [22] we generate negative sets for each pathway by randomly selecting interactions which are not in the ground truth pathway at a rate of fifty negatives for every known interaction. However, the choice of sub-sampled negatives can affect a method's performance in terms of precision. We designed a procedure to select a set of negatives for each pathway that is used for all precision-recall calculations for all reconstruction methods. For each pathway P , we compute the F_{max} for each of the six original methods ($\mathcal{M}$) using 50 distinct randomly sampled negative sets and place them in a 6×50 matrix M . Each column of M has the F_{max} scores for a given method across all 50 negative sets. The F_{max} variance for the 50 sets is relatively small for each method (Figure 2). We then create a new 6×50 matrix N such that

$$N_{ij} = |M_{ij} - \text{med}(M_i)|,$$

where $\text{med}(x)$ is the median of a vector. Intuitively, the rows of N represent the differences from the median F_{max} for a given negative set for all methods. We sum the rows of N to get the aggregated difference from the median F_{max} for each negative set, and select the negative set N_P that has the minimum of these values (reflecting the negative set that is the closest to the median across all methods). We do this for each pathway P in order to get reasonable negative sets across all pathways that do not advantage a particular method $\mathcal{M}$.

2.4.3 Composites. In order to evaluate the overall performance of each prediction method across all pathways, we construct composite predictions and negative sets. Given a collection of negative sets specific to each pathway P , we create a composite negative set N_{comp} by taking a union over these edge sets, keeping track of the pathway for each negative edge:

$$N_{\text{comp}} = \bigcup_P \{(e, P) \mid e \in N_P\}.$$

Likewise for a method $\mathcal{M}$ with predictions $H = (V_{P, \mathcal{M}}, E_{P, \mathcal{M}})$ for a pathway P , let H_{comp} be the union over these edges:

$$H_{\text{comp}} = \bigcup_P \{(e, P) \mid e \in E_{P, \mathcal{M}}\}.$$

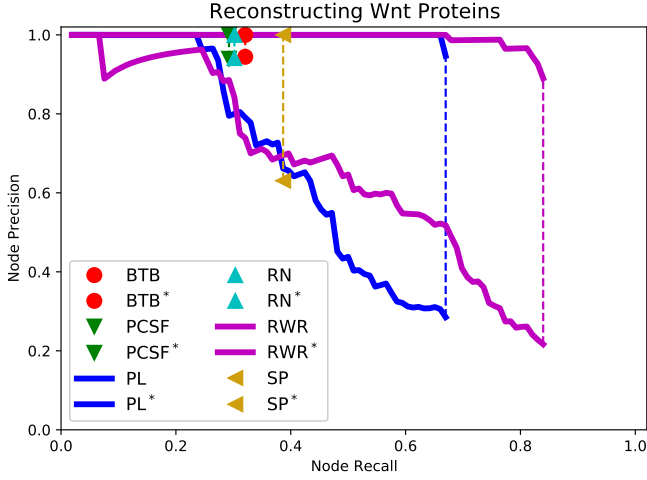


Figure 3: Existing methods recover proteins with high precision. Asterisks denote that negative nodes connected to a positive are ignored. (PL was run with $k = 5000$ and RWR was run with $\tau = 0.5$ for illustration purposes.)

In the case of ranked predictions the edges of H_{comp} are then sorted such that a partial ordering is restored. The set of positives P_{comp} are constructed in a similar way from the ground truth pathways $P = (V_P, E_P)$:

$$P_{\text{comp}} = \bigcup_P \{(e, P) \mid e \in E_P\}.$$

Precision, recall, and F_{max} can be calculated as described in Section 2.4.1, providing a general overall view of the performance of each method, though it privileges the performance of methods on larger pathways and obscures the heterogeneity of the performances on individual pathways.

3 RESULTS

3.1 Reconstruction Methods Successfully Recover Proteins in Pathways

The Pathway Reconstruction Problem is solved by recovering both proteins and interactions that connect receptors to transcriptional regulators, and previous work has shown that recovering interactions is a challenging task [22]. However, current pathway reconstruction methods are relatively successful at recovering pathway involved proteins. We evaluated the six original pathway reconstruction methods $\mathcal{M}$ on their ability to recover the nodes in the Wnt pathway, subsampling negative nodes at a rate of 50 negatives to every positive (Figure 3). As has been previously shown [22], the methods maintain relatively high precision at larger values of recall — especially PL and RWR which offer ranked lists of candidate nodes (rather than a node set). Strikingly, if we ignore negative nodes that share an edge with a positive node, all methods jump to nearly perfect precision (asterisked methods in Figure 3). For instance when reconstructing the Wnt pathway, almost all predicted nodes for all methods are either in Wnt or one edge away from Wnt in G . These results indicate that predicted nodes from existing pathway reconstruction methods may be able to provide reliable information for predicting protein interactions.

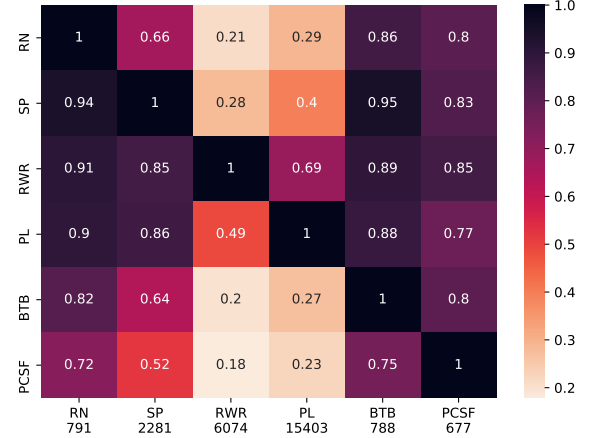


Figure 4: Jaccard overlap of nodes predicted by pathway reconstruction methods, normalized by the columns. Number of predicted interactions given below x-axis labels.

In order to assess the similarity of different reconstruction methods we computed the Jaccard overlap of the proteins they predicted across all pathways using default parameters. Given the interactome $G = (V, E)$, let $A \subseteq V$, $B \subseteq V$, be two sets of nodes. The asymmetric Jaccard overlap of A with B is the percentage of elements of A which are also in B . We computed the Jaccard overlap of composite predictions H_{comp} for the six methods across all 29 pathways (Figure 4). The algorithms, which are quite heterogeneous, have a considerable amount of agreement in their reconstructed proteins, ranging from Jaccard indices of 0.18 (PCSF with RWR) to 0.95 (BTB with SP).

3.2 PRAUG Improves Upon Reconstruction Methods

We have shown that reconstruction methods can recover the proteins in a pathway; however recovering the interactions is more challenging. We compared each reconstruction method to PRAUG in terms of composite precision and recall across all 29 signaling pathways (Figure 5). Each panel compares the performance of a pathway reconstruction method $\mathcal{M}$ from Section 2.2 to $\widetilde{\mathcal{M}}$, using default values and taking care to keep the negative set consistent for each pathway (Section 2.4.2). While we note that there are differences among reconstruction methods, we did not optimize the parameters for each method with respect to the Pathway Reconstruction Problem, so we do not compare un-augmented methods against one another.

In every case, the PRAUG method $\widetilde{\mathcal{M}}$'s F_{max} value is larger than the original method $\mathcal{M}$'s F_{max} in terms of recovering the interactions, outperforming $\mathcal{M}$ by an average of 89% (Figure 5). PRAUG gains this large improvement in F_{max} by improving recall while sacrificing sparingly with respect to precision. While precision at maximum recall is lower across the board for augmented methods as compared to their counterparts, precision for $\widetilde{\mathcal{M}}$ is equal to or greater than the precision of $\mathcal{M}$ at the top recall of $\mathcal{M}$.

The methods $\mathcal{M}$ that produce unranked predictions (BTB, PCSF, RN, and SP) are single points in Figure 5, and the PRAUG-augmented counterparts of these methods maintain the precision of $\mathcal{M}$ for

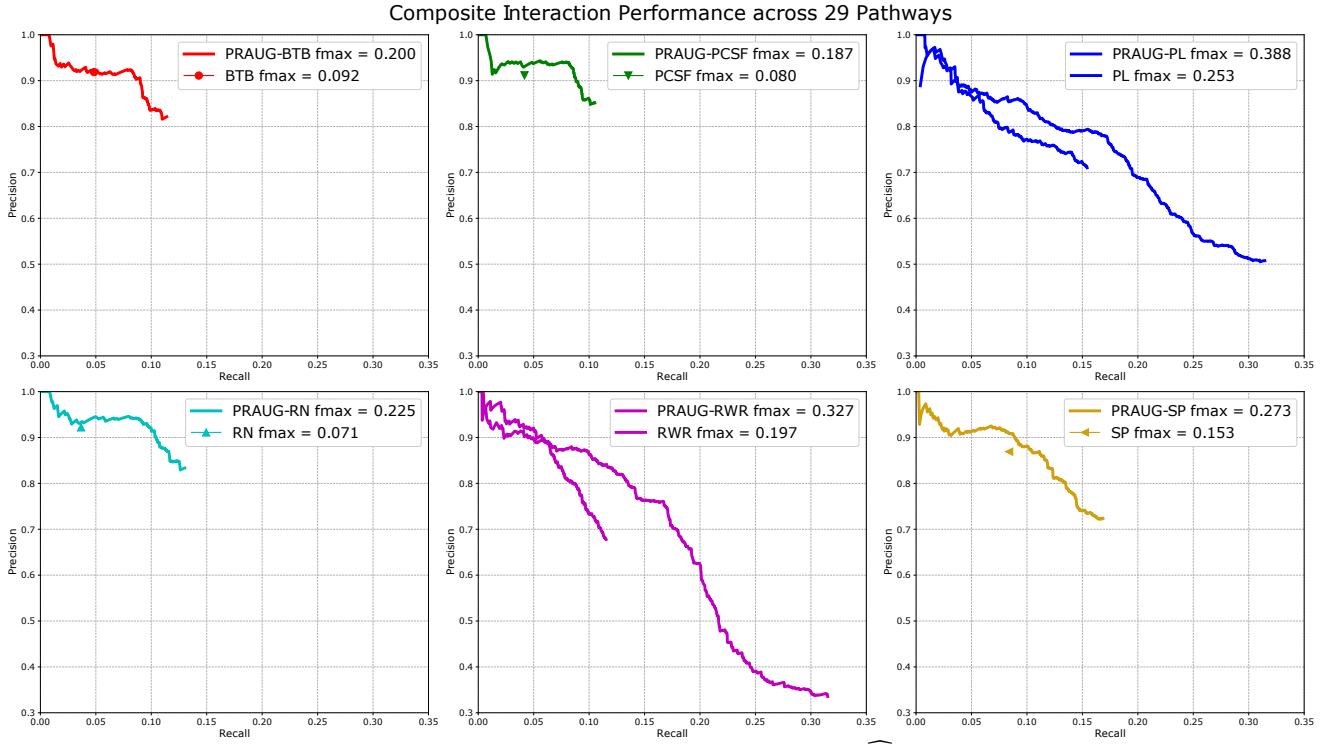


Figure 5: Performance of method M and PRAUG-transformed $\widehat{M}$ across all 29 pathways.

larger values of recall. However, many exhibit a plateau of precision after which precision drops off rapidly with increased recall (for example, precision remains nearly constant for PRAUG-BTB from recall values of 0.0125 to 0.08). The PRAUG counterparts of the ranked methods (PL and RWR) do not exhibit this behavior. Despite the differences between the unranked and ranked methods, the key result is that the PRAUG counterparts $\widehat{M}$ of each pathway reconstruction method outperforms M in $F_{\max}$, recall, and often precision.

3.3 Robustness Analysis and Benchmarking

PRAUG is a straightforward traversal-based algorithm, and it may not be immediately clear why this approach can be so successful. In this section we justify PRAUG’s depth-first traversal, provide empirical upper bounds on the Pathway Reconstruction Problem given methods which assume a traversal paradigm, and illustrate the effect of parameter selection on PRAUG methods compared to their original counterparts.

3.3.1 Choice of Traversal. Step 4 of a PRAUG-augmented pathway reconstruction method is to perform a depth first traversal from the sources of a pathway ignoring edge weights (Section 2.1). We considered three alternatives to PRAUG’s unweighted depth-first traversal:

PRAUG-BFS: Perform an unweighted breadth-first traversal on G starting from σ , taking only nodes in X .

PRAUG-WEIGHTED: Perform a weighted depth-first traversal on G starting from σ , taking only nodes in X .

PRAUG-BFS-WEIGHTED: Perform a weighted breadth-first traversal on G starting from σ , taking only nodes in X .

In all cases, the method terminates when no new edge can be traversed from σ to nodes in X . Thus, all of these methods recover the exact same interactions but traverse them in a different order. The choice of traversal or edge weighting has little effect on the precision and recall, as is illustrated for PL and RWR on the Wnt pathway (Figure 6A,B). When there is a difference, using breadth first traversal performs worse than using depth first traversal, and weighting has a negligible effect on the performance of the augmented methods.

3.3.2 Ground Truth as Upper Bounds. We also investigated how well PRAUG could recover the pathways when given the ground truth nodes and interactions, which serve as an upper bound on a traversal-like performance. We first define an oracle pathway reconstruction method GT which, given an interactome G , a source set S , and a target set T , produces the unique ground truth signaling pathway $P = (V_P, E_P)$ such that $S \subseteq V_P$ and $T \subseteq V_P$.¹ Of course, just because we can define something doesn’t mean we can build it. If we knew how to realize GT — an algorithm which produces the ground truth pathway by definition — then we would doubtlessly be publishing a paper on that instead. In lieu of such good fortune we provide our implementation of GT with the actual pathway P to be produced. Thus, our implementation of GT does not meaningfully solve the pathway reconstruction problem. However, GT provides the complete set of ground truth nodes to its PRAUG-augmented counterpart:

PRAUG-GT-NODES: Perform an unweighted depth-first traversal on G starting from σ , taking only nodes in V_P .

¹The oracle GT is an element of the set of all pathway reconstruction methods $\mathcal{M}$.

PRAUG-GT-NODES informs us how well a PRAUG-augmented method could perform if a pathway reconstruction method $\mathcal{M}$ perfectly reconstructed the proteins involved in a given pathway. In this first upper-bound, PRAUG-GT-NODES provides the precision and recall of interactions assuming (a) perfect node precision and recall and (b) we begin the traversal from the sources.

We can ask a similar question about PRAUG's performance when we have the ground truth interactions available for traversal. This requires augmenting the definition of Step 4 of PRAUG as follows:

PRAUG-GT-EDGES: Perform an unweighted depth-first traversal on G starting from σ , taking only edges from E_p .

PRAUG-GT-EDGES differs from PRAUG because it traverses edges rather than nodes. PRAUG-GT-EDGES is guaranteed to maintain perfect precision and marks the upper bound on recall for a traversal based method starting from the sources which only makes predictions from the annotated ground truth.

In Figure 6A and 6B, PRAUG-GT-NODES are shown in black and PRAUG-GT-EDGES are shown in brown. The ground truth Wnt pathway is the same in both panels A and B, and the only difference between panels is the subsampled negative set involved in the calculation. PRAUG-GT-NODES achieves nearly total recall while maintaining nearly perfect precision for Wnt, which is representative of PRAUG-GT-NODES generally. As expected, PRAUG-GT-EDGES maintains perfect precision up to much larger values of recall than any other pathway reconstruction method. Interestingly, PRAUG-GT-NODES achieves higher recall than PRAUG-GT-EDGES, because there are some incorrect interactions that allow the traversal to reach more positive interactions when only considering nodes.

3.3.3 Parameter Sweeps. Several of the reconstruction methods considered in this paper take additional inputs beyond an interactome, source nodes, and target nodes (Table 1). These additional parameters substantially affect the behavior of the methods. In order to assess how additional parameters affected the performance of both original methods $\mathcal{M}$ as well as their corresponding augmented methods $\tilde{\mathcal{M}}$, we performed parameter sweeps for both PRAUG-PL and PRAUG-RWR (Figure 6C and 6D). PL and RWR both have parameters which determine the number of edges in the reconstructed pathway. In the case of PL, k denotes the number of paths returned.² In the case of RWR, τ marks the percentage of total edge flux in the interactome to return. Thus for PL (resp. RWR) lower k (resp. τ) values produce truncated precision recall curves of higher k (resp. τ) values.

The same does not hold for the PRAUG-augmented counterparts to PL and RWR. Because the ranking of edges in a PRAUG-augmented method is given by their order in the traversal, the introduction of new nodes (by way of parameter change) to the traversal leads to markedly different precision recall curves (Figure 6C and 6D). Further, PRAUG-augmented methods are highly sensitive to parameter changes. In general large parameters lead to degraded precision at low values of recall, but recover a large portion of positive interactions. This can be seen for instance in PRAUG-RWR $\tau = 0.75$ in Figure 6D. Thus, optimal parameters

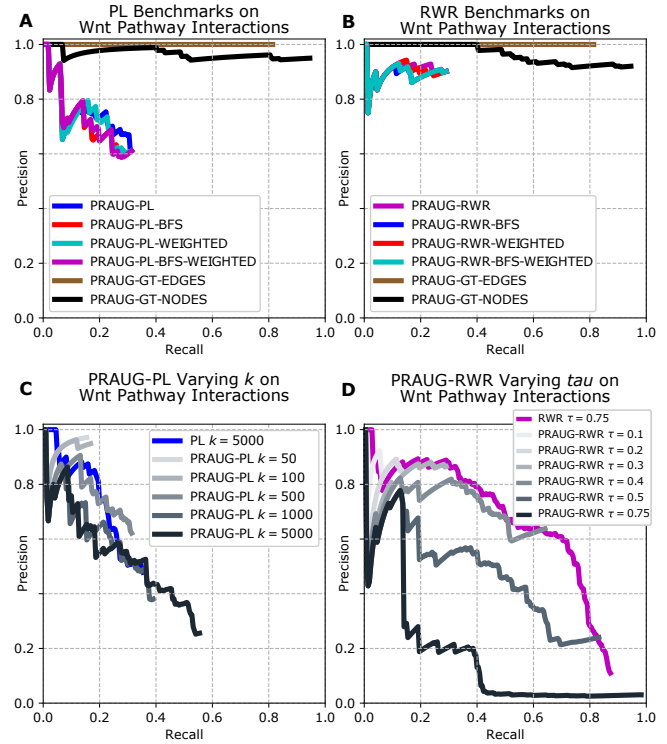


Figure 6: (A,B) PRAUG compared to breadth-first search, using a weighted interactome, and using ground truth (GT) nodes and edges as inputs for PL and RWR. (C,D) PRAUG performance when varying k for PL and varying τ for RWR. PL (blue) and RWR (magenta) are shown for the largest parameter, since all smaller parameters simply truncate this curve.

for an unaugmented method may deviate significantly from the optimal parameters for its PRAUG-augmented counterpart.

3.4 Case Studies

To illustrate the benefits of pathway reconstructions augmented by PRAUG, we focus on two well-studied signaling pathways: Wnt and Notch. For each pathway, we show the single-pathway precision and recall curves for all methods $\mathcal{M}$ and $\tilde{\mathcal{M}}$, as well as PRAUG-GT-NODES and PRAUG-GT-EDGES. We also consider the first 500 edges predicted by each of the six PRAUG methods, and we visualize the interactions that are supported by all six methods. It is important to note that PRAUG ranks interactions according to the order in which edges are traversed, which does not necessarily imply confidence in the interaction. While the choice of using the first 500 interactions is arbitrary, they provide a representative set of interactions from the traversal and produce a network that connects most of the receptors to transcriptional regulators.

3.4.1 The Wnt Signaling Pathway. The canonical Wnt signaling pathway regulates a variety of developmental events during embryogenesis and helps maintain tissue homeostasis [18]. The non-canonical Wnt signaling pathway, which is not mediated through the canonical Frizzled receptors or β -catenin, also regulates cell movement and tissue polarity [13]. The combination of canonical

²Though note that new paths may use no new edges. Thus k neither fully nor uniquely determines the number of edges in the reconstructed pathway.

and non-canonical Wnt signaling makes this a good case study for PRAUG reconstructions. Pathway reconstruction methods that return a single subgraph tend to have high precision but quite low recall (Figure 7A). PL and RWR extend this recall while maintaining relatively high precision, which has been previously shown [22]). The PRAUG methods are able to improve both precision and recall for all methods except for PL, in which it only improves recall (Figure 7A). PRAUG-GT-EDGES and PRAUG-GT-NODES are described in Section 3.3.2 and are provided here for benchmark purposes.

The interactions that appear in the first 500 interactions of all PRAUG methods (Figure 7B) are largely consistent with the pathway reconstruction reported by the first 200 paths from Path-Linker [22]. The Wnt network includes CFTR, a chloride ion transporter that interacts with non-canonical RYK receptor, which has been shown to be involved in Wnt signaling via a subset of the Wnt ligands [22]. Many of the 22 intermediate gray nodes have previously been implicated in Wnt pathway crosstalk, including the MAP Kinases, SRC, EGFR, PII3KR1, and SMAD9 [22, 27]. There are also known ubiquitination proteins (UBQLN1, UBQLN4, and TRIM21) which are general proteins and may not be specific to Wnt signaling [8]. Other non-positives have documented roles in Wnt signaling despite not being present in NetPath: for example, DACT1 interacts through Disheveled (DVL) family proteins and MDPI regulates TCF family transcription factors [8]. Further, proteins such as DLG4 and DAZAP2 are not directly associated with Wnt signaling, and may offer new hypotheses for follow-up investigation. DLG4, for example, is involved in cell-cell adhesion (a classic role for Wnt) but is known to interact with the NMDA receptor in the brain [8].

In addition to the nodes within the Wnt network, the interactions themselves also help put the reconstructions in context. Only a handful of the interactions in the Wnt network appear in the NetPath Wnt pathway (black edges in Figure 7B). The remaining edges are either negatives or ignored due to subsampling in the precision-recall plots. However, 66% of these edges appear as interactions in some KEGG signaling pathway (shown in pink), indicating that they are signaling interactions. Further, some of the edges that do not appear in NetPath or KEGG (red edges) connect proteins in NetPath's Wnt signaling (e.g. DVL1-DVL2 and LRP6-GSK3B), which may be known interactions that have been missed by signaling pathways. In this way, pathway reconstruction methods that focus on recovering interactions have the potential for generating more mechanistic-driven hypotheses.

3.4.2 The Notch Signaling Pathway. The Notch signaling pathway regulates fundamental processes such as morphogenesis, cell differentiation and cell-fate determination, proliferation, and cell death, and is highly conserved across metazoans [13, 16]. Notch is a unique case study for the Pathway Reconstruction Problem because its primary receptors function both as cell surface receptors and nuclear transcription factors, seemingly contradicting our assumptions that a signaling pathway's receptors and TRs are mutually exclusive. However, most pathway reconstruction methods recover Notch signaling with high precision at low recall, and PRAUG methods typically improve recall while remaining close to the original method's precision (Figure 8A). Unlike Wnt, PRAUG-GT-EDGES for Notch only recovers about half of the positive interactions, but PRAUG-GT-NODES boasts a recall greater than 0.8.

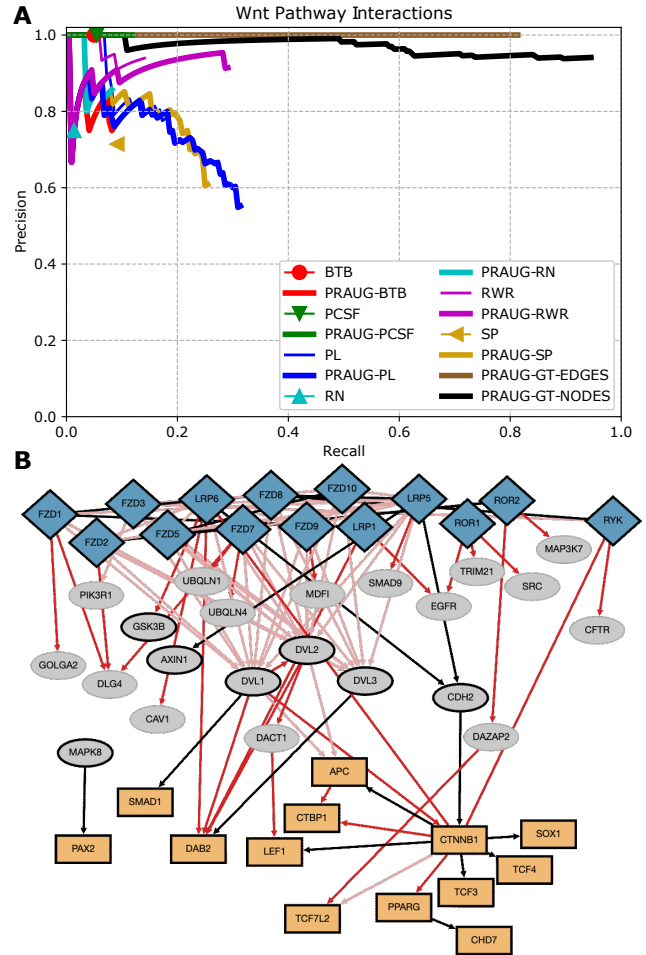


Figure 7: Precision recall curve for the Wnt signaling pathway (A), and the network comprised of the top interactions predicted by the PRAUG reconstructions (B). Blue diamonds are receptors and orange rectangles are transcriptional regulators. Black nodes and edges appear in the Wnt NetPath pathway; pink edges appear in some KEGG pathway, and red edges do not appear in NetPath or KEGG databases.

When we examine the interactions recovered in the first 500 interactions of all six PRAUG methods, we see that the majority of the interactions connect receptors to TRs, highlighting the nuclear role of Notch family proteins (Figure 8B). One of the six intermediate proteins is a member of the MAML transcriptional coactivator, and two are SMAD proteins, which along with NF κ B are known to associate with the Notch intracellular domain [16]. Five of the 53 non-positive interactions appear in NetPath's TGF- β Receptor pathway [13], and another twelve interactions appear in KEGG as members of Notch signaling (e.g. edges from MAML proteins to EP300) or Th1 and Th2 cell differentiation (e.g. edges from Notch3 to RELA and NF κ B1, which co-complex to form NF κ B), of which Notch is known to initiate [28]. While two thirds of non-positive interactions do not appear in NetPath or KEGG (shown in red), they connect proteins that are in the NetPath Notch pathway, suggesting

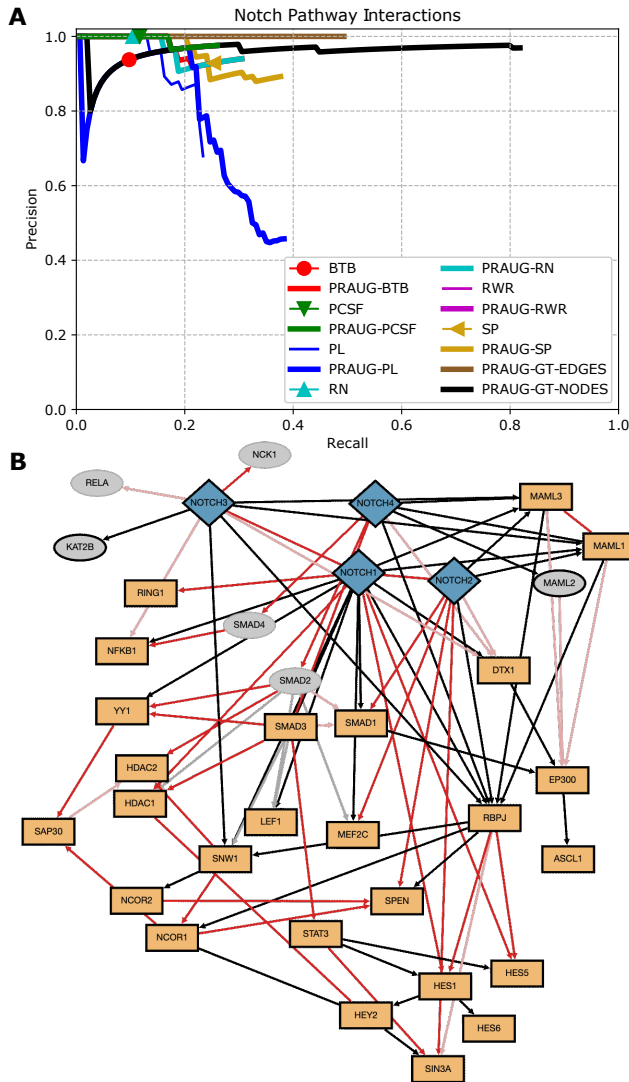


Figure 8: Precision recall curve for the Notch signaling pathway (A), and the network comprised of the top interactions predicted by the PRAUG reconstructions (B). Edges and nodes are colored as described in Figure 7; and gray edges appear in some NetPath pathway other than Wnt.

that these may be missed interactions in NetPath. Further, the connection from Notch3 to NCK1, an adaptor protein which associates with certain growth factor receptors, has recently been shown to maintain Notch signaling by preventing processes induced by Notch inhibition, and provides a promising testable hypothesis [10].

4 DISCUSSION

We present PRAUG, a higher-order algorithm for augmenting pathway reconstruction methods which consistently recovers interactions from ground truth pathways better than the original methods. We have shown that this framework is highly generalizable, improving reconstructions for six pathway reconstruction algorithms that include shortest-paths (PL, BTB, SP), network flow (RN), random walk (RWR), and Steiner tree (PCSF) approaches. We demonstrate

an improved $F_{\max}$ when PRAUG is applied to all of these methods across 29 signaling pathways (Figure 5), which cover diverse pathways with respect to both size (e.g. number of nodes and interactions in Table 2) as well as mechanism (immune signaling pathways, cancer hallmark pathways, and hormone signaling pathways). We illustrate PRAUG’s power in generating plausible hypotheses for follow-up validation by analyzing interactions predicted by all PRAUG method predictions for Wnt and Notch (Figures 7 and 8).

PRAUG does not present a new pathway reconstruction algorithm, but rather a paradigm for improving other algorithms designed to reconstruct pathways. The foundational observation of our method is that reconstructing the nodes in a pathway is a relatively easy task at which many existing algorithms excel (Figure 3). In particular, The performance of PRAUG-GT-NODES demonstrates that PRAUG can produce substantially better pathway reconstruction methods if it is provided with a method which reliably produces correct node sets. Even when limiting the inputs to ground truth edges, PRAUG-GT-EDGES achieves recall much higher than current algorithms. Every algorithm that we used for this paper attempts to explicitly reconstruct the *interactions* of a pathway; further work to improve protein reconstructions may lead us closer to these empirical upper bounds.

There are variants of the Pathway Reconstruction Problem, such as a “curation-aware” pathway reconstruction that uses all proteins and interactions from pathway databases as input [31], and pathway reconstructions that involve changing the underlying interactome [3]. Considering a traversal-based approach for these formulations may prove similarly useful. Further, PRAUG offers a multitude of opportunities for defining a “pathway reconstruction method” $\mathcal{M}$ – for example, we could focus on gene ontology node sets or experimentally-derived genes as the node set X for PRAUG traversal. In this way, PRAUG may be able to recover interactions from methods which explicitly produce node reconstructions.

Of the two pathway reconstruction methods with ranked interactions, we were surprised by the dramatic performance of RWR. For large values of τ , RWR achieves precision of 0.8 at a recall of 0.4, blowing away all other non-PRAUG methods (magenta curve in Figure 6D). This method, which was inspired by TieDIE [21], performs a random walk from sources, then another random walk on a reversed graph from targets, and combines the edge flux scores for every edge. This random walk was different than the one presented in the PathLinker paper [22], which performed a random walk with restarts only from the sources. Here, RWR shows significant promise for improving pathway reconstructions, and is an improvement over prior algorithms in its own right.

One feature of PRAUG-augmented methods is that they tend to produce more edges than their unaugmented counterparts. While the $F_{\max}$ values are higher for PRAUG-augmented methods, the topology of the produced networks may differ significantly from the ground truth. Combining the results of several PRAUG-augmented methods through some sort of voting system as done in Section 3.4 can highlight biologically meaningful interactions. However, we acknowledge that the returned subnetworks are optimized for our ground truth measures and may not accurately reflect the real underlying pathways, the topology of which is unknown.

Despite its simplicity, we described some subtleties to PRAUG. For example, an optimal choice of parameters for a method $\mathcal{M}$ may

not be optimal for the PRAUG version $\widetilde{\mathcal{M}}$ (Figure 6C,D). In general, we found that the larger the reconstruction returned by $\mathcal{M}$, the worse the PRAUG augmented method performed. In essence, if the pathway reconstruction returned a large subset of the nodes in G , then PRAUG would return an induced subgraph of the nodes, which is bound to have many negative interactions. We do not explicitly compare pathway reconstruction methods to each other, as more careful selection of parameters for each method $\mathcal{M}$ and PRAUG method $\widetilde{\mathcal{M}}$ would be necessary.

PRAUG shows great potential for improving upon existing pathway reconstruction methods. We have shown six applications of PRAUG using extant reconstruction methods, and demonstrated promise for improving untested algorithms. PRAUG brings us one step closer to practical pathway predictions for emerging and understudied signaling pathways.

Data Availability. All code and datasets are available on <https://github.com/TobiasRubel/Pathway-Reconstruction-Tools>. Users can use GraphSpace to explore Wnt (<http://graphspace.org/graphs/28928>) and Notch (<http://graphspace.org/graphs/28927>) networks.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (BIO-DBI #1750981 to AR).

REFERENCES

- [1] Gregorio Alanis-Lobato. 2015. Mining protein interactomes to improve their reliability and support the advancement of network medicine. *Frontiers in genetics* 6 (2015), 296.
- [2] Gregorio Alanis-Lobato, Miguel A Andrade-Navarro, and Martin H Schaefer. 2016. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic acids research* (2016), gkw985.
- [3] Gregorio Alanis-Lobato, Pablo Mier, and Miguel Andrade-Navarro. 2018. The latent geometry of the human protein interaction network. *Bioinformatics* 34, 16 (2018), 2826–2834.
- [4] Özgün Babur, Augustin Luna, Anil Korkut, Funda Durupinar, Metin Can Siper, Ugur Dogrusoz, Joseph E Aslan, Chris Sander, and Emek Demir. 2018. Causal interactions from proteomic profiles: molecular data meets pathway knowledge. *BioRxiv* (2018), 258855.
- [5] Omer Basha, Ruth Barshir, Moran Sharon, Eugene Lerman, Binyamin F Kirson, Idan Hekselman, and Esti Yeger-Lotem. 2017. The TissueNet v. 2 database: a quantitative view of protein-protein interactions across human tissues. *Nucleic acids research* 45, D1 (2017), D427–D431.
- [6] Omer Basha, Dvir Flom, Ruth Barshir, Ilan Smoly, Shoval Tirman, and Esti Yeger-Lotem. 2015. MyProteinNet: build up-to-date protein interaction networks for organisms, tissues and user-defined contexts. *Nucleic acids research* 43, W1 (2015), W258–W263.
- [7] Omer Basha, Omry Mauer, Eyal Simonovsky, Rotem Shpringer, and Esti Yeger-Lotem. 2019. ResponseNet v. 3: revealing signaling and regulatory pathways connecting your proteins and genes across human tissues. *Nucleic acids research* 47, W1 (2019), W242–W247.
- [8] UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic acids research* 47, D1 (2019), D506–D515.
- [9] Lenore Cowen, Trey Ideker, Benjamin J Raphael, and Roded Sharan. 2017. Network propagation: a universal amplifier of genetic associations. *Nature Reviews Genetics* 18, 9 (2017), 551.
- [10] Alexandre Dubrac, Gael Genet, Roxana Ola, Feng Zhang, Laurence Pibouin-Fragner, Jinah Han, Jiasheng Zhang, Jean-Léon Thomas, Alain Chedotal, Martin A Schwartz, et al. 2016. Targeting NCK-mediated endothelial cell front-rear polarity inhibits neovascularization. *Circulation* 133, 4 (2016), 409–421.
- [11] Antonio Fabregat, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, Bijay Jassal, Florian Korfinger, Bruce May, et al. 2018. The reactome pathway knowledgebase. *Nucleic acids research* 46, D1 (2018), D649–D655.
- [12] Yuxu Jiang, Yanchun Liang, Duolin Wang, Dong Xu, and Trupti Joshi. 2020. A dynamic programming approach to integrate gene expression data and network information for pathway model generation. *Bioinformatics* 36, 1 (2020), 169–176.
- [13] Kumaran Kandasamy, S Sujatha Mohan, Rajesh Raju, Shivakumar Keerthikumar, Ghantasala S Sameer Kumar, Abhilash K Venugopal, Deepthi Telikicherla, J Daniel Navarro, Suresh Mathivanan, Christian Pecquet, et al. 2010. NetPath: a public resource of curated signal transduction pathways. *Genome biology* 11, 1 (2010), R3.
- [14] Minoru Kanehisa and Susumu Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28, 1 (2000), 27–30.
- [15] Ali Sinan Köksal, Kirsten Beck, Dylan R Cronin, Aaron McKenna, Nathan D Camp, Saurabh Srivastava, Matthew E MacGilvray, Rastislav Bodik, Alejandro Wolf-Yadlin, Ernest Fraenkel, et al. 2018. Synthesizing signaling pathways from temporal phosphoproteomic data. *Cell reports* 24, 13 (2018), 3607–3618.
- [16] Raphael Kopan and Ma Xenia G Ilagan. 2009. The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* 137, 2 (2009), 216–233.
- [17] Martina Kutmon, Anders Riutta, Nuno Nunes, Kristina Hanspers, Egon L Wilhelmsen, Anwesha Bohler, Jonathan Mélius, Andra Waagmeester, Sravanthi R Sinha, Ryan Miller, et al. 2016. WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic acids research* 44, D1 (2016), D488–D494.
- [18] Catriona Y Logan and Roel Nusse. 2004. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* 20 (2004), 781–810.
- [19] Matthew E MacGilvray, Evgenia Shishkova, Deborah Chasman, Michael Place, Anthony Gitter, Joshua J Coon, and Audrey P Gasch. 2018. Network inference reveals novel connections in pathways regulating growth and defense in the yeast salt response. *PLoS computational biology* 13, 5 (2018).
- [20] Chris S Magnano and Anthony Gitter. 2019. Automating parameter selection to avoid implausible biological pathway models. *BioRxiv* (2019), 845834.
- [21] Evan O Paull, Daniel E Carlin, Mario Niepel, Peter K Sorger, David Haussler, and Joshua M Stuart. 2013. Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE). *Bioinformatics* 29, 21 (2013), 2757–2764.
- [22] Anna Ritz, Christopher L Poirel, Allison N Tegge, Nicholas Sharp, Kelsey Simmons, Allison Powell, Shiv D Kale, and TM Murali. 2016. Pathways on demand: automated reconstruction of human signaling networks. *NPJ systems biology and applications* 2, 1 (2016), 1–9.
- [23] Igor Rodchenkov, Ozgun Babur, Augustin Luna, Bulent Arman Aksoy, Jeffrey V Wong, Dylan Fong, Max Franz, Metin Can Siper, Manfred Cheung, Michael Wrana, et al. 2020. Pathway Commons 2019 Update: integration, analysis and exploration of pathway data. *Nucleic acids research* 48, D1 (2020), D489–D497.
- [24] Jan Daniel Rudolph, Marjo de Graauw, Bob van de Water, Tamar Geiger, and Roded Sharan. 2016. Elucidation of signaling pathways from large-scale phosphoproteomic data using protein interaction networks. *Cell systems* 3, 6 (2016), 585–593.
- [25] Jochen Supper, Lucia Spangenberg, Hannes Planatscher, Andreas Dräger, Adrian Schröder, and Andreas Zell. 2009. BowTieBuilder: modeling signal transduction pathways. *BMC systems biology* 3, 1 (2009), 67.
- [26] Damian Szklarczyk, John H Morris, Helen Cook, Michael Kuhn, Stefan Wyder, Milan Simonovic, Alberto Santos, Nadezhda T Doncheva, Alexander Roth, Peer Bork, et al. 2016. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* (2016), gkw937.
- [27] Allison N Tegge, Nicholas Sharp, and TM Murali. 2016. Xtalk: a path-based approach for identifying crosstalk between signaling pathways. *Bioinformatics* 32, 2 (2016), 242–251.
- [28] Irma Tindemans, Marlies JW Peeters, and Rudi W Hendriks. 2017. Notch signaling in T helper cell subsets: instructor or unbiased amplifier? *Frontiers in immunology* 8 (2017), 419.
- [29] Nurcan Tuncbag, Alfredo Braunstein, Andrea Pagnani, Shao-Shan Carol Huang, Jennifer Chayes, Christian Borgs, Riccardo Zecchina, and Ernest Fraenkel. 2013. Simultaneous reconstruction of multiple signaling pathways via the prize-collecting steiner forest problem. *Journal of computational biology* 20, 2 (2013), 124–136.
- [30] Nurcan Tuncbag, Sara JC Gosline, Amanda Kedaigle, Anthony R Soltis, Anthony Gitter, and Ernest Fraenkel. 2016. Network-based interpretation of diverse high-throughput datasets through the omics integrator software package. *PLoS computational biology* 12, 4 (2016).
- [31] Mitchell J Wagner, Aditya Pratapa, and TM Murali. 2019. Reconstructing signaling pathways using regular language constrained paths. *Bioinformatics* 35, 14 (2019), i624–i633.
- [32] Aaron K Wong, Arjun Krishnan, and Olga G Troyanskaya. 2018. GIANT 2.0: genome-scale integrated analysis of gene networks in tissues. *Nucleic acids research* 46, W1 (2018), W65–W70.
- [33] Esti Yeger-Lotem, Laura Riva, Linhui Julie Su, Aaron D Gitler, Anil G Cashikar, Oliver D King, Pavan K Auluck, Melissa L Geddie, Julie S Valastyan, David R Karger, et al. 2009. Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nature genetics* 41, 3 (2009), 316.
- [34] Ibrahim Youssef, Jeffrey Law, and Anna Ritz. 2019. Integrating protein localization with automated signaling pathway reconstruction. *BMC bioinformatics* 20, 16 (2019), 1–15.