SPRAS: A Workflow for Streamlining Network-based Pathway Reconstruction

Adam Shedivy[1], Nina Young[2], Christopher Magnano[1,3], Pramesh Singh[2], Tobias Rubel[4], Anna Ritz[2], and Anthony Gitter[1]

[1]Morgridge Institute and University of Wisconsin-Madison, Madison, WI, USA; [2]Biology Department, Reed College, Portland, OR, USA; [3]Center for Computational Biomedicine, Harvard Medical School, Cambridge, MA, USA; [4]Computer Science Department, University of Maryland - College Park, College Park, MD, USA
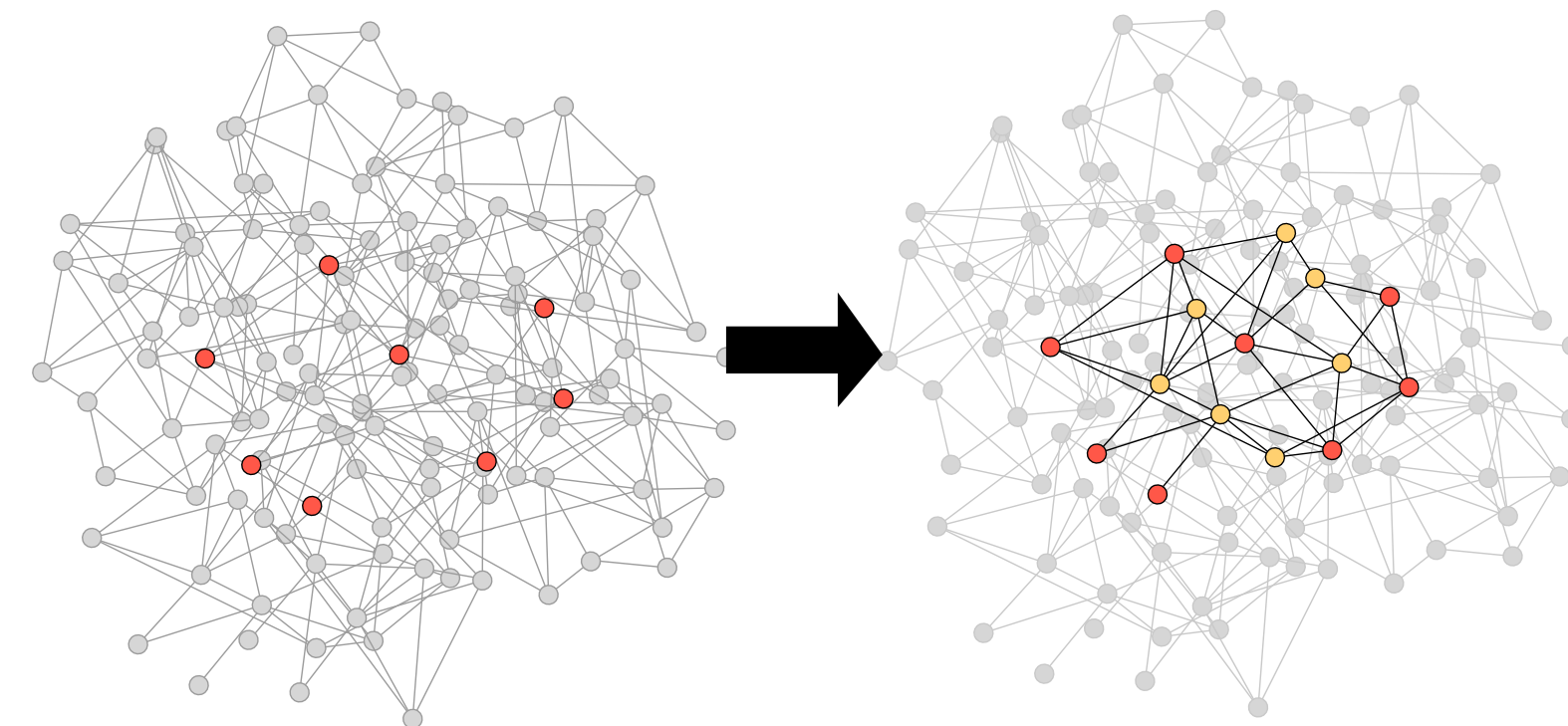
Pathway Reconstruction

Analyzing data from transcriptomic, proteomic, and other high-throughput assays in a biological network context provides a systems-level understanding of individual events. **Pathway reconstruction** aims to connect genes or proteins of interest (nodes) by selecting relevant interactions (edges) from a background network that connect these genes. Dozens of network biology methods exist that solve the pathway reconstruction problem (examples include [1, 2, 3, 4]).

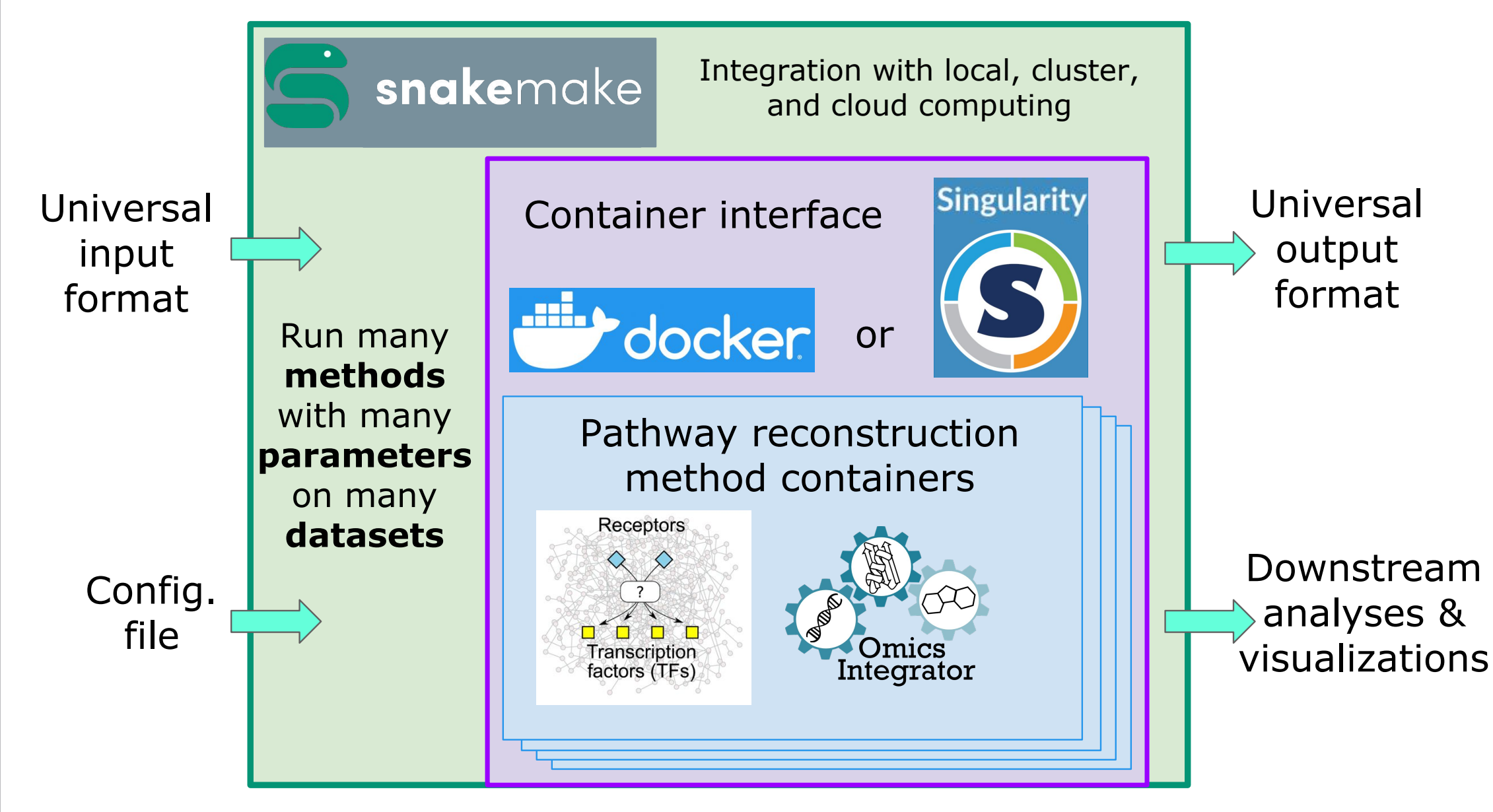**Node Inputs:**
Gene Expression,
Protein Expression,
GWAS,
…

Network Input:
Protein interactions,
Functional similarities,
…

Pathway Reconstruction Challenges

Practical challenges have limited the adoption of pathway reconstruction methods.

Each method has its own file formats, dependencies, installation process, and approach to setting parameters. These decisions are shown for two methods [1, 2].

SPRAS unifies different methods so they can all be run in a single consistent workflow.

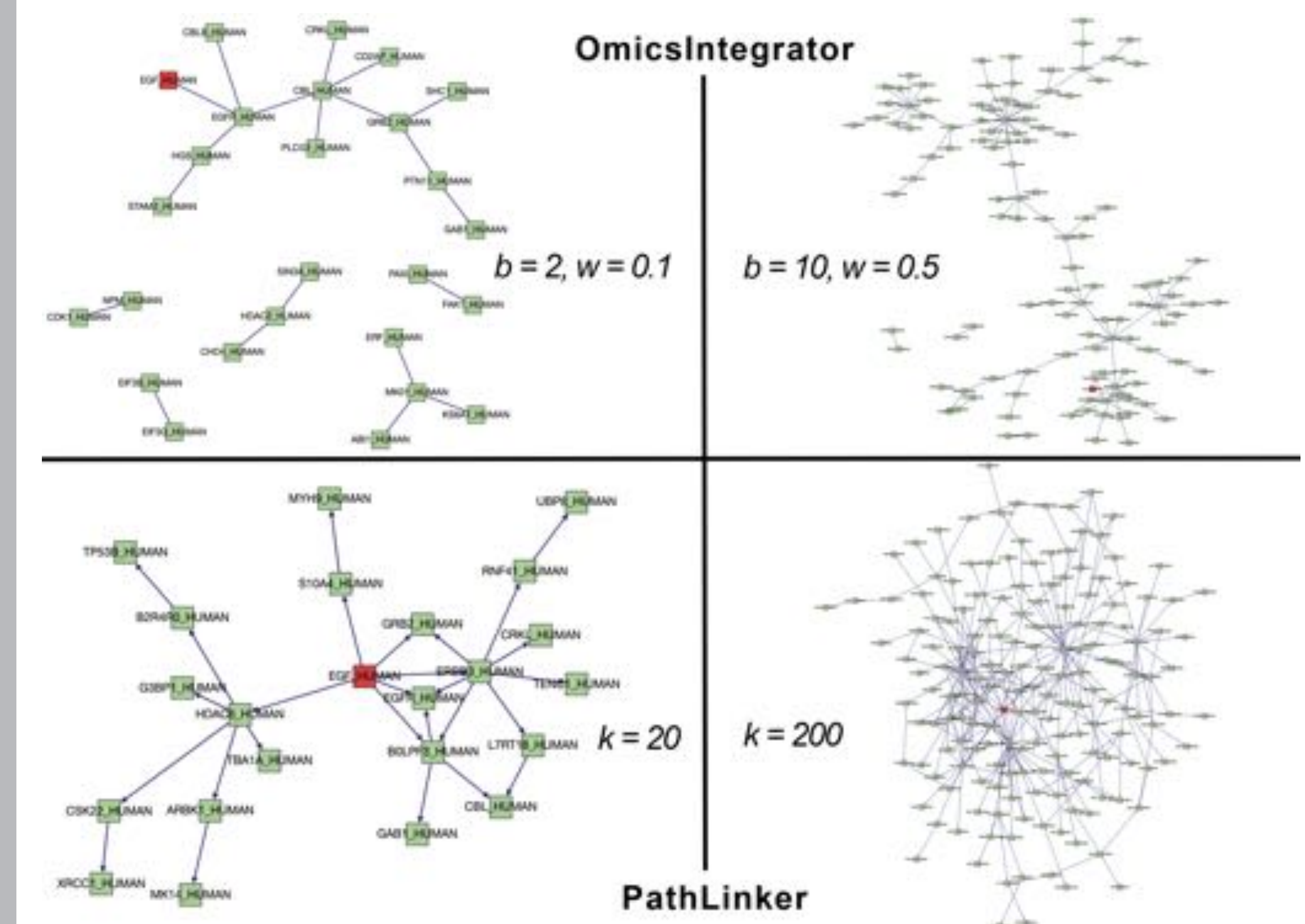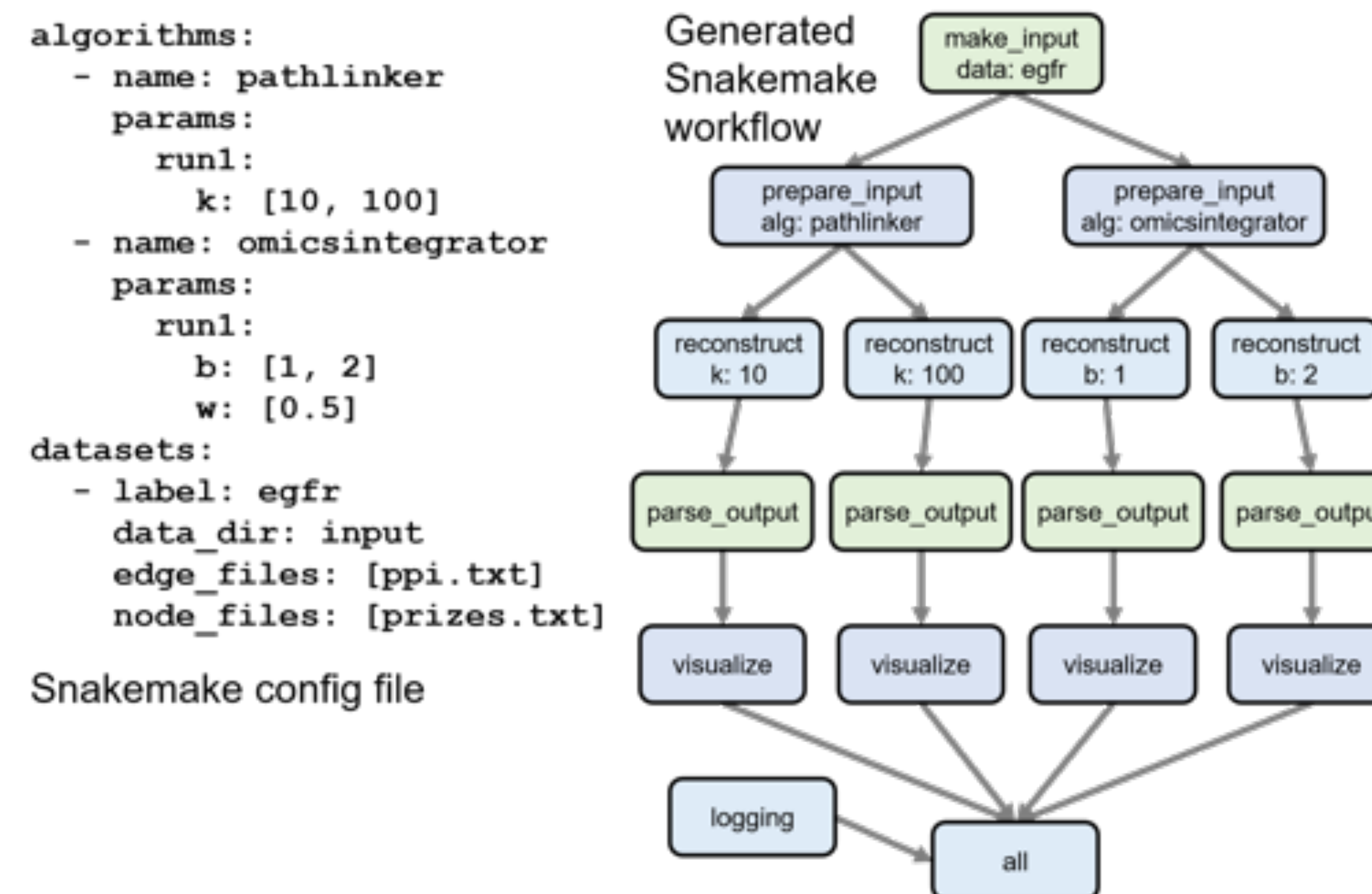|	PathLinker	Omics Integrator

Dependencies	Python packages
	`python 3.5`
	`networkx 1.11`
		Python packages
		`python 2.7`
		`networkx 1.11`
		`numpy 1.13`
		…
		Boost C++ libraries
		Compiled `msgsteiner`
		C++ code

Setting parameters	Command line
	`python run.py \`
	`-k 100`
		Configuration file
		`w = 1`
		`b = 2`
		`D = 10`
		`mu = 0.5`

Input files	Node types file
	`#Node  Node type`
	`A      source`
	`D      target`
		Node prizes file
		`name    prize`
		`A       3.5`
		`D       1.2`

	Network file
	`#N1 N2`
	`A   B   0.1`
	`A   C   0.9`
	`B   D   0.1`
		Network file
		`p1 p2 weight dir`
		`A  B   0.1    U`
		`A  C   0.9    D`
		`B  D   0.1    U`

Output file	`#tail head i cost`
	`A    B    1 0.01`
	`B    D    1 0.01`
		`A   pp   B`
		`B   pp   D`

SPRAS: Signaling Pathway Reconstruction Analysis Streamliner

SPRAS provides a unified pathway reconstruction workflow through Snakemake. It also provides containers for popular algorithms [1, 2].

SPRAS Workflow

A single configuration file sets parameters for all methods. Snakemake runs all combinations of datasets, algorithms, and parameters and visualizes network outputs.

```
algorithms:
  - name: pathlinker
    params:
      run1:
        k: [10, 100]
  - name: omicsintegrator
    params:
      run1:
        b: [1, 2]
        w: [0.5]
datasets:
  - label: egfr
    data_dir: input
    edge_files: [ppi.txt]
    node_files: [prizes.txt]
```

Snakemake config file

Generated Snakemake workflow

SPRAS Case Study

Output networks from the workflow on the left visualize phosphorylation changes in response to EGF treatment [3].

The example outputs illustrate reconstructed Steiner forests (Omics Integrator [1]) and k-shortest paths (PathLinker [2]) run with parameters that produce small and large reconstructions.

Ongoing and Future Work

Pathway visualization is a common downstream task. We now have a protoype Cytoscape implementation.

We plan to implement additional pathway reconstruction methods such as ResponseNet [4]. We also plan to develop evaluation criteria for benchmarking.

Interested? We are looking for collaborators and contributors! Email aritz@reed.edu or gitter@biostat.wisc.edu for more info.

References

[1] Tuncbag *et al.* Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. *PLoS Comp. Biol.* **12**, e1004879 (2016).
[2] Ritz *et al.* Pathways on demand: automated reconstruction of human signaling networks. *npj Systems Biology and Applications* **2**, 16002 (2016).
[3] Köksal *et al.* Synthesizing Signaling Pathways from Temporal Phosphoproteomic Data. *Cell Reports* **24**, 3607–3618 (2018).
[4] Lan *et al.* ResponseNet: revealing signaling and regulatory networks linking genetic and transcriptomic screening data. *Nucleic Acids Research* **39**, W424–W429 (2011).

Acknowledgements

https://github.com/Reed-CompBio/spras