Network Legos: Building Blocks of Cellular Wiring Diagrams

T. M. Murali and Corban G. Rivera

Department of Computer Science
Virginia Polytechnic Institute and State University

Eleventh Annual International Conference on Research in Computational Molecular Biology
May 22, 2007

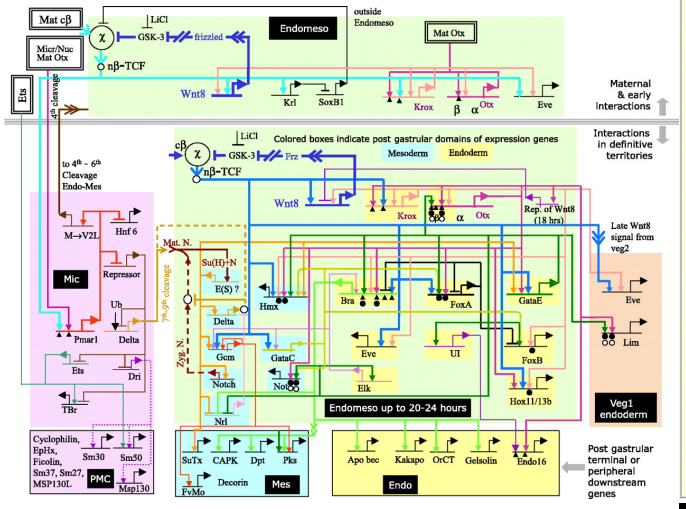
Goals of Systems Biology

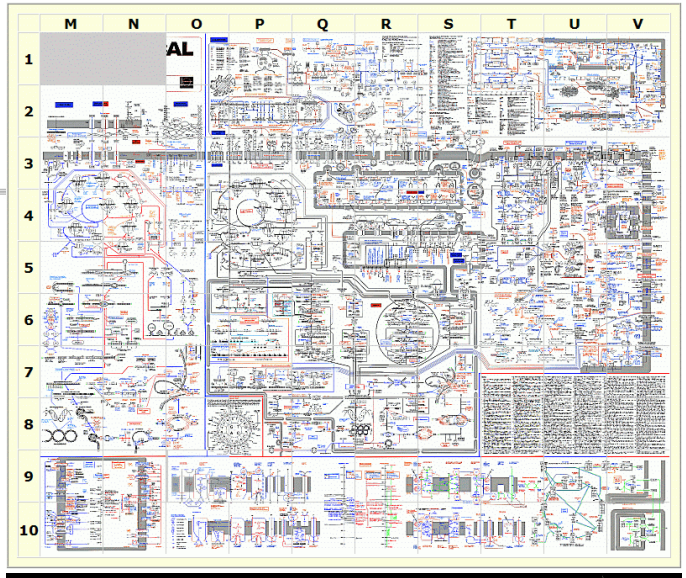
- Identify the building blocks of molecular interaction networks
- Interconnect the building blocks to build high level models of the cell
- Understand the interaction of the building blocks over time and under different conditions

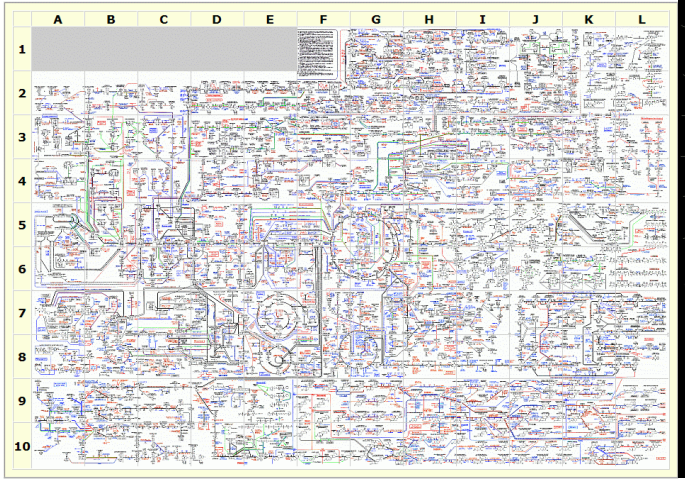
Goals of Systems Biology

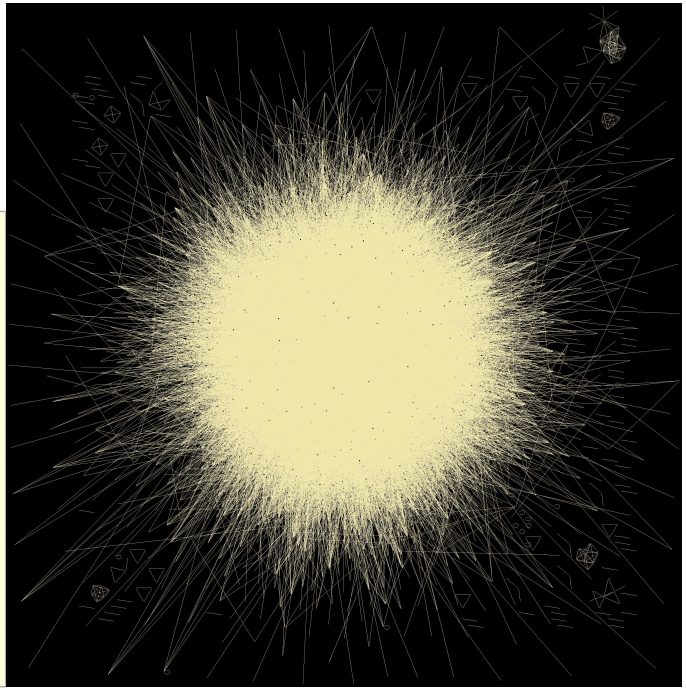
- Identify the building blocks of molecular interaction networks
- Interconnect the building blocks to build high level models of the cell
- Understand the interaction of the building blocks over time and under different conditions
- How do we automatically construct these building blocks?

Cellular Wiring Diagrams

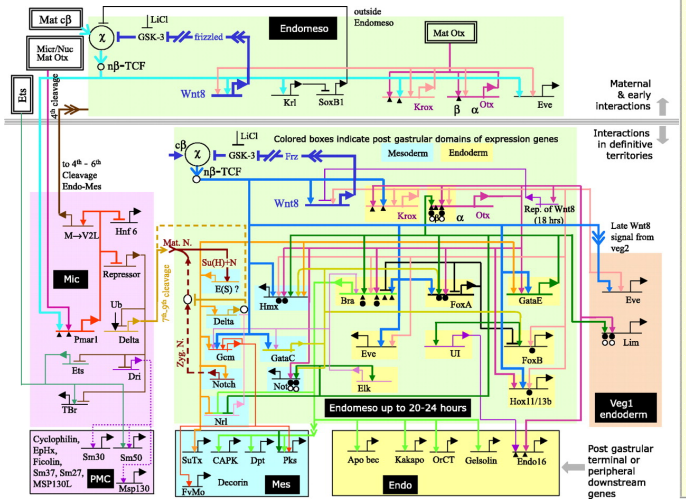


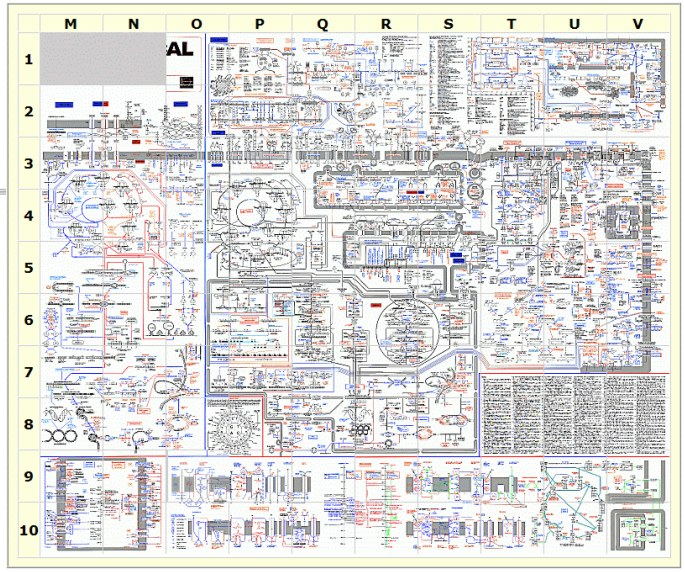


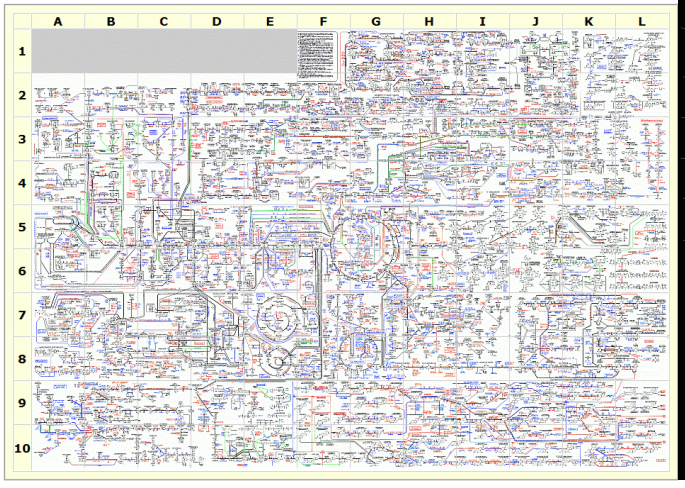


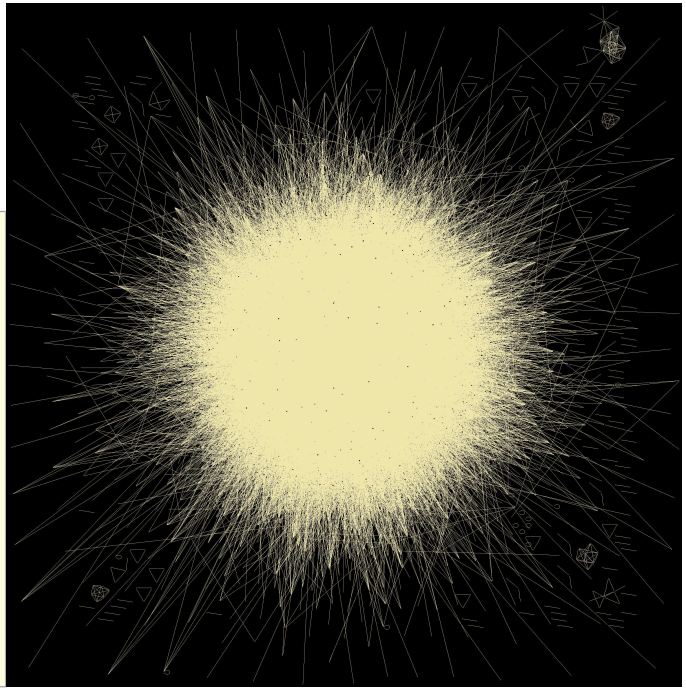


Cellular Wiring Diagrams









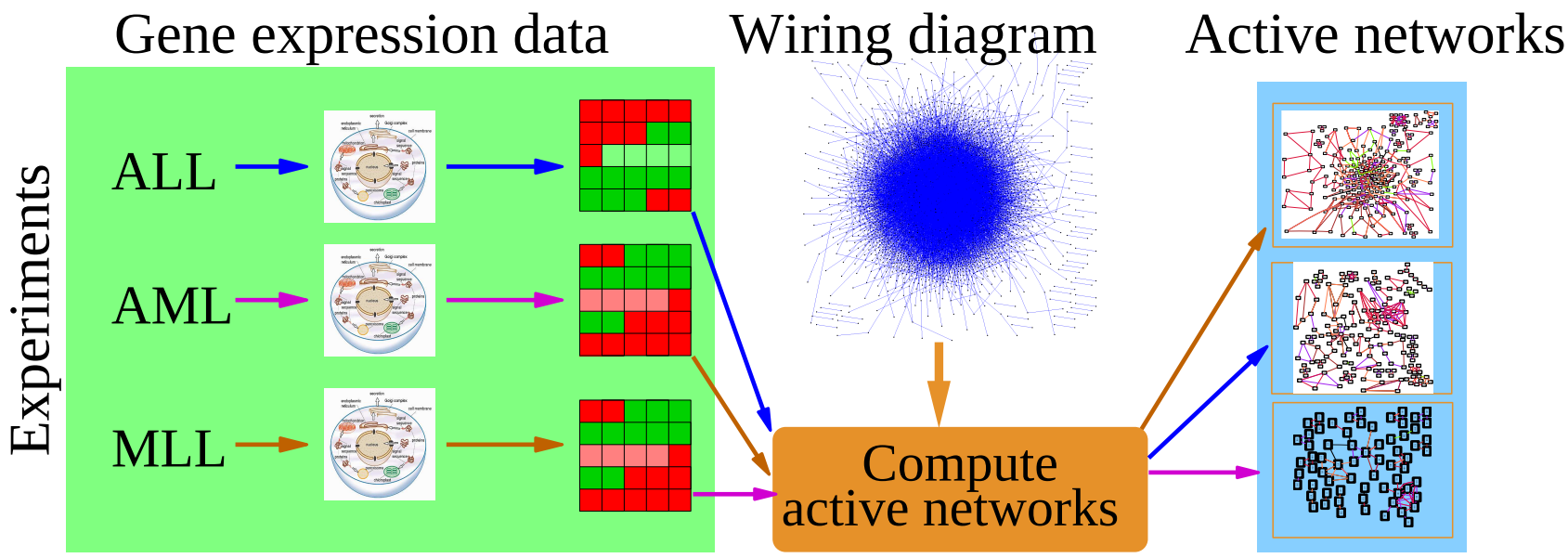
- These wiring diagrams are static.

Cell State is Dynamic

- Active molecular interactions change with time, external signals, and perturbations.
- We need to integrate wiring diagram with other types of data to compute the cell's response to different conditions.

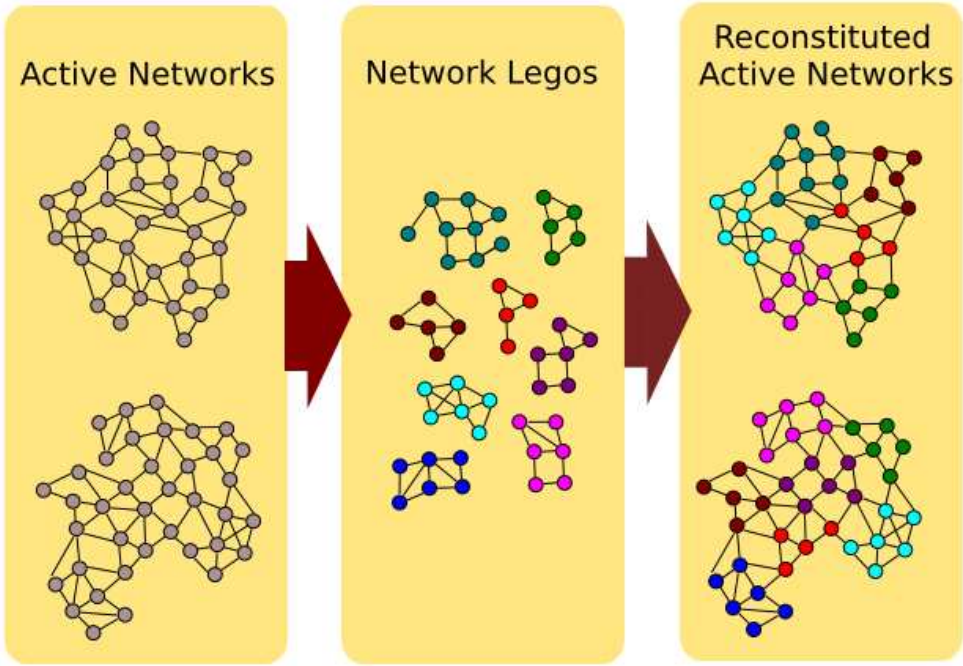
Active Networks

- Transcriptional regulatory measurements provide dynamic snapshots of cellular activity.
- *Active networks*: Molecular interactions activated by the cell in response to a stimulus.
- Methods to integrate wiring diagram with transcriptional measurements to compute response networks:
  - Ideker et al., Bioinformatics 2002
  - Luscombe et al., Nature 2004
  - Han et al., Nature 2004
  - Ulitsky and Shamir, BMC Sys Bio 2007

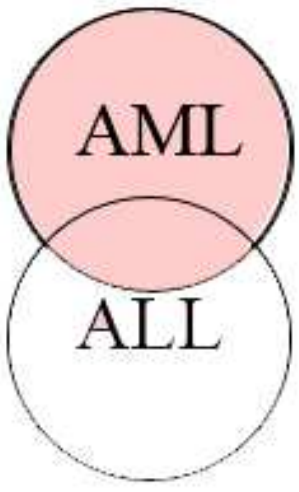


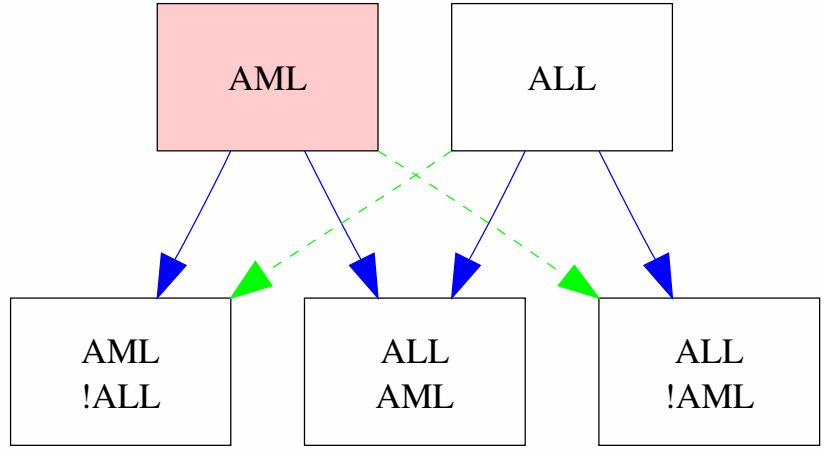
Goals of the Network Lego Approach

- What are the similarities and differences between active networks?
- Can we identify building blocks or network legos that constitute each of the active networks?



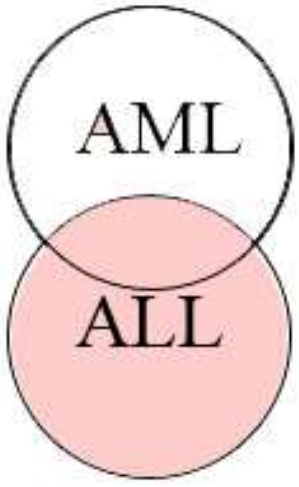
Comparing Two Active Networks

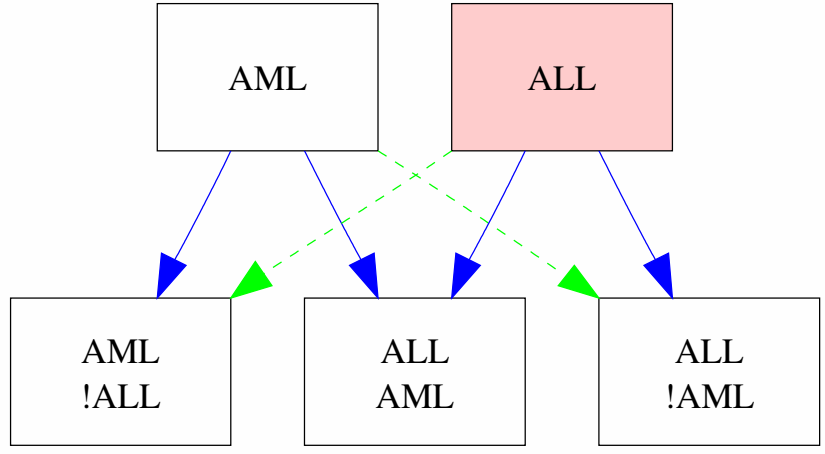




AML

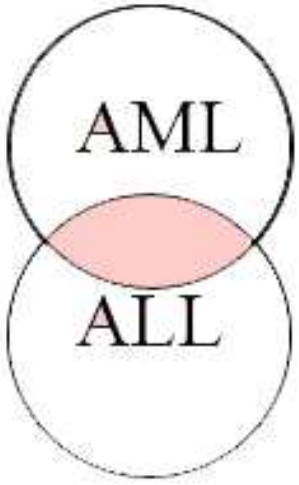
Comparing Two Active Networks

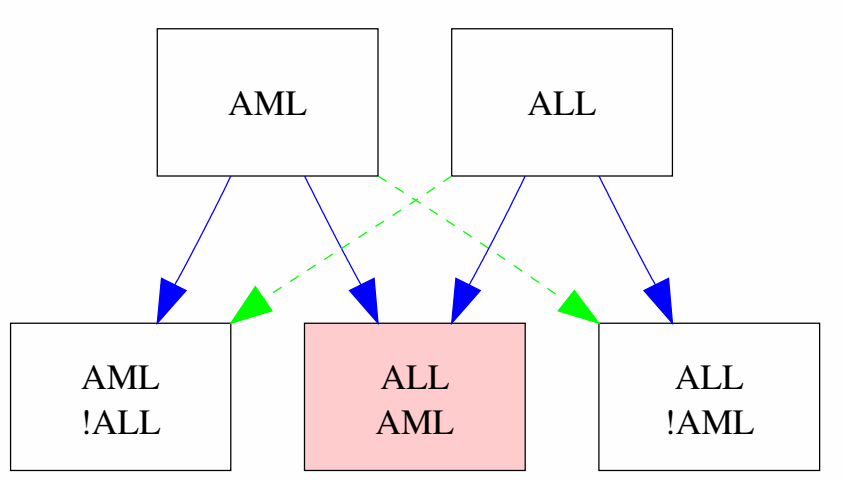




ALL

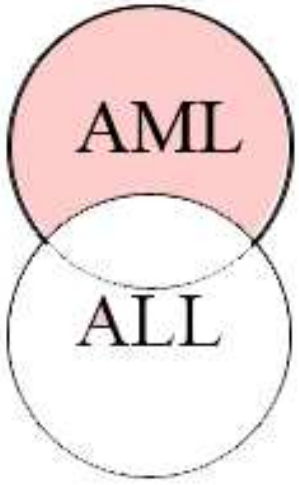
Comparing Two Active Networks

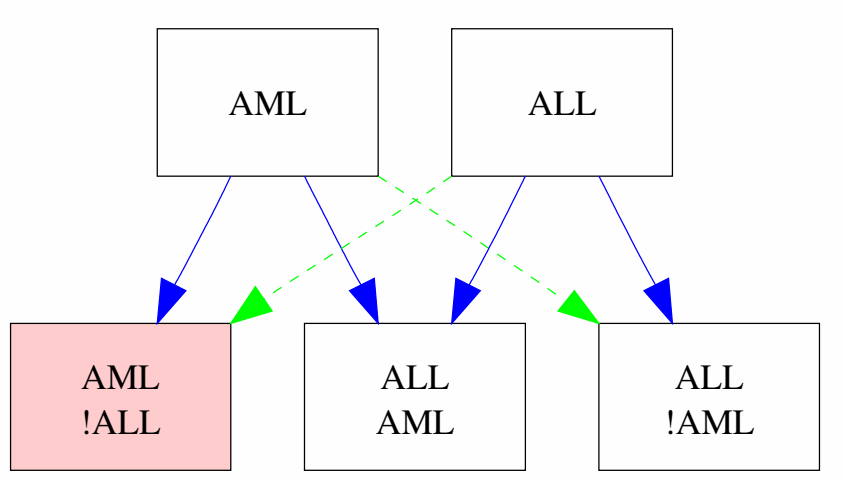




AML ∩ ALL

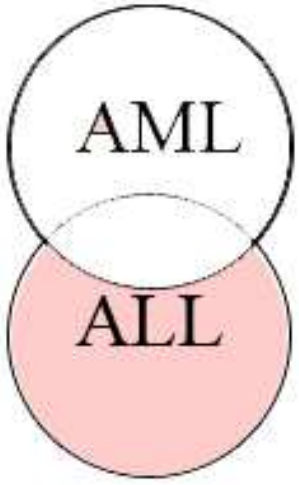
Comparing Two Active Networks

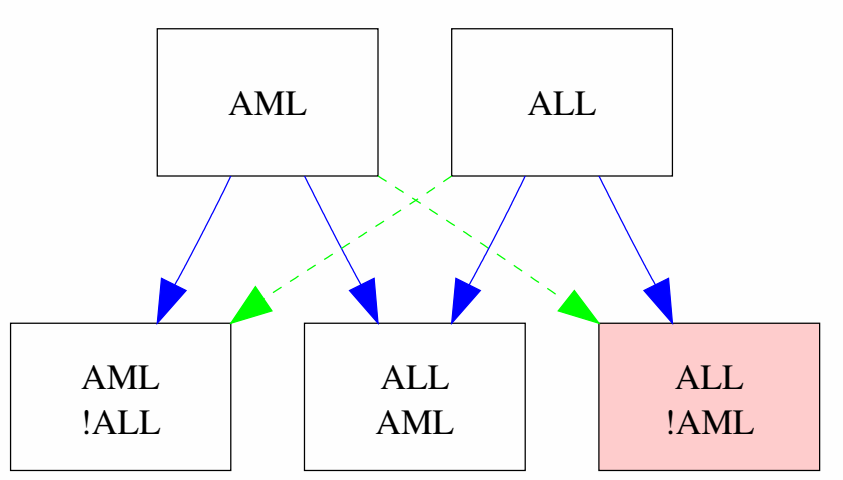




$AML \cap !ALL$

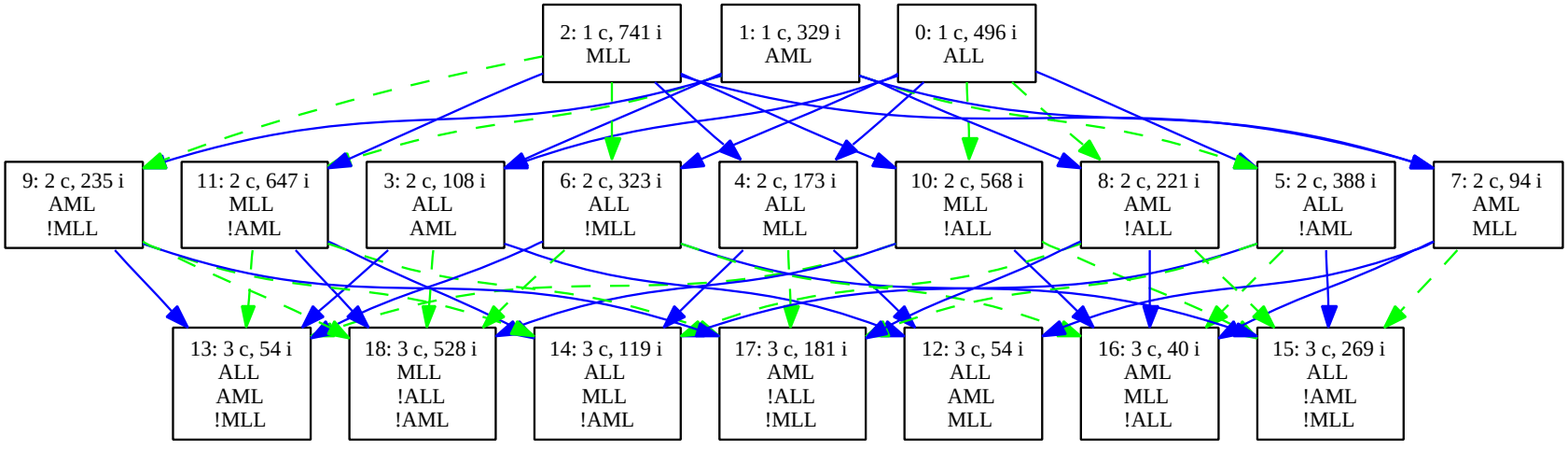
Comparing Two Active Networks



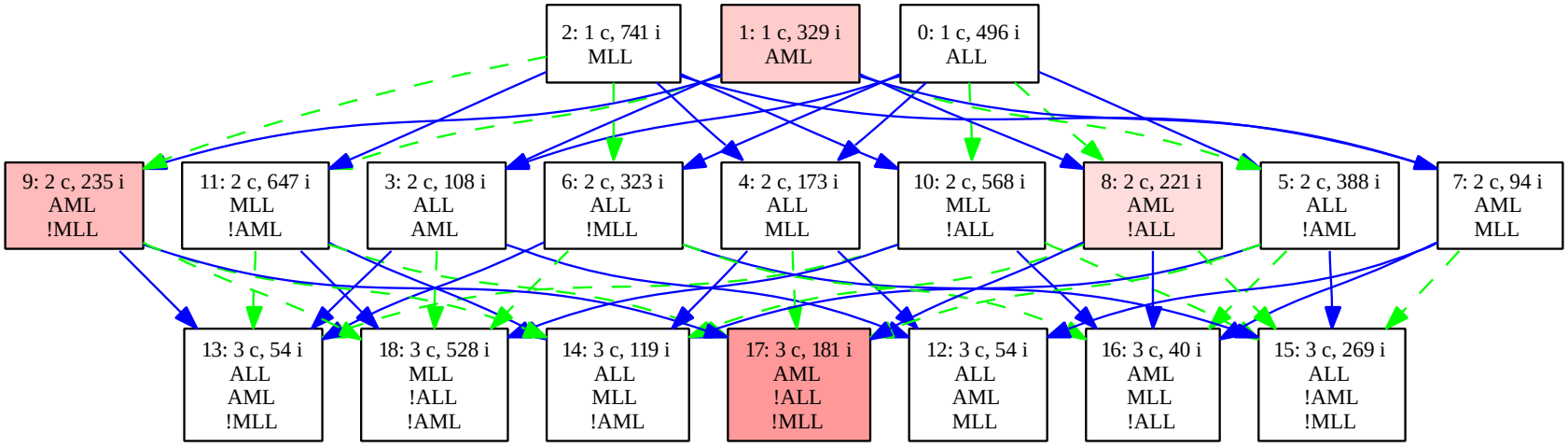


$ALL \cap !AML$

Comparing Three Active Networks

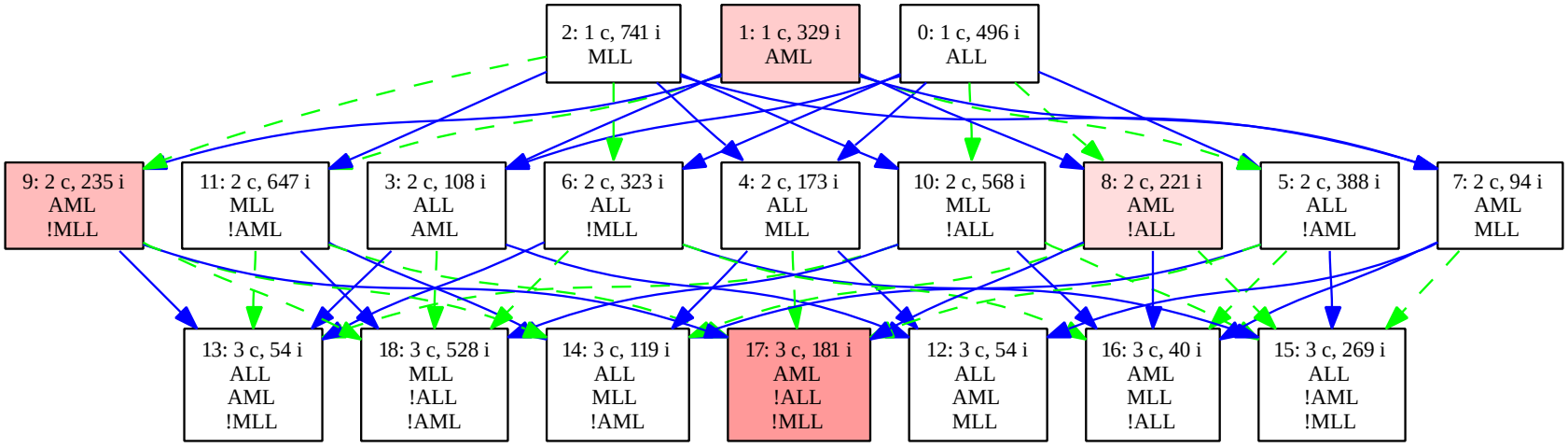


Differential Activation of the Kit Receptor Pathway in AML



- AML: p-value 2×10^{-4}
- AML $\cap$!ALL: p-value 1×10^{-3}
- AML $\cap$!MLL: p-value 6.7×10^{-5}
- AML $\cap$!ALL $\cap$!MLL: p-value 3.5×10^{-7}

Differential Activation of the Kit Receptor Pathway in AML



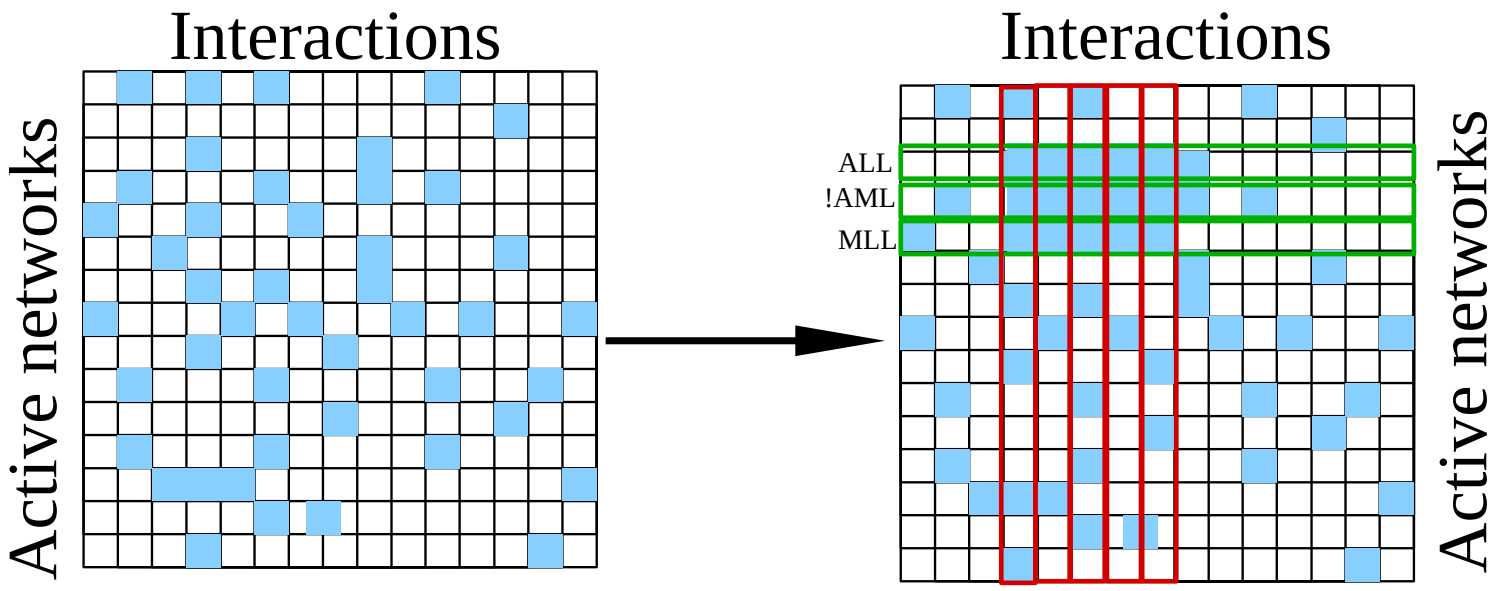
- AML: p-value 2×10^{-4}
- AML $\cap$!ALL: p-value 1×10^{-3}
- AML $\cap$!MLL: p-value 6.7×10^{-5}
- AML $\cap$!ALL $\cap$!MLL: p-value 3.5×10^{-7}

- c-KIT receptor is activated in almost all subtypes of AML but not in ALL (Reuss-Borst et al., *Leukemia*, 1994, Bene et al., *Blood*, 1998, Schwartz et al., *Leuk Lymphoma.*, 1999).

Challenges in Comparing Arbitrary Numbers of Active Networks

- How to do we efficiently compute all combinations?
- Which combinations are the network legos?
- How do we demonstrate that we have found network legos?

How to do we efficiently compute all combinations?

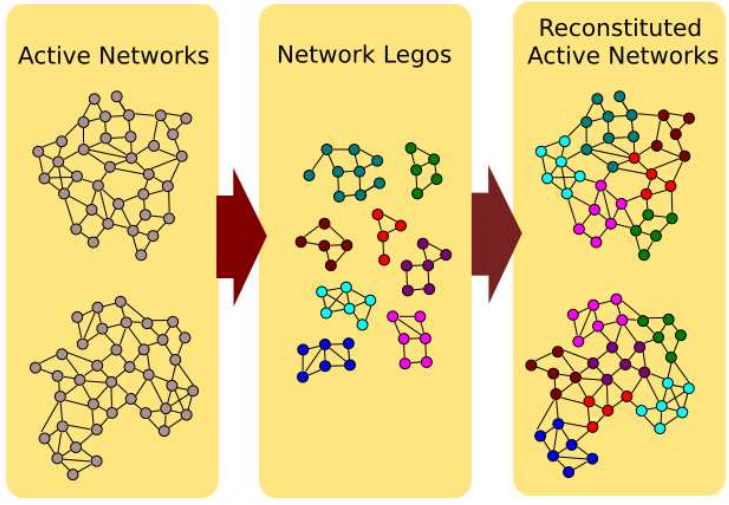


- Construct a binary matrix M whose columns are interactions.
- Represent each active network and its complement in M's rows.
- Compute all closed biclusters in M.
- Connect biclusters in a DAG.

Which combinations are the network legos?

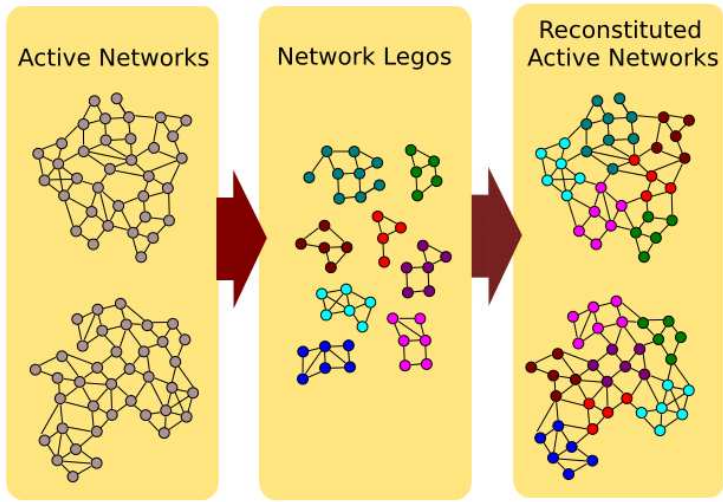
- For each bicluster B with n non-complemented and c complemented active networks
  1. Pick n non-complemented and c complemented active networks repeatedly at random, compute the number of interactions induced by this combination, and build a distribution of the number of interactions.
  2. Set the p-value of B to be the fraction of random biclusters with more interactions than B.
- B is a *network lego* if it is more significant than any of its ancestors or descendants in the DAG.

How do we demonstrate that we have found network legos?



- Stability
  - Sequentially remove each active network from the input and recompute network legos.
  - For each original network lego, compute the fraction of leave one out datasets for which the network lego occurs with at least 95% fidelity.

How do we demonstrate that we have found network legos?



- Stability
  - Sequentially remove each active network from the input and recompute network legos.
  - For each original network lego, compute the fraction of leave one out datasets for which the network lego occurs with at least 95% fidelity.
- Recoverability
  - Compute the union of network legos.
  - Measure the size of the intersection of each active network with this union.

Analysis of Human Stress Data

- 13 distinct stresses applied to human cells (Murray et al., Mol. Bio. Cell, 2004).
- Stress conditions include heat shock, oxidative stress, cell cycle arrest, and crowding.
- Two cell types: WI38 Fibroblasts and Hela.
- Murray et al. note that each stress elucidated a unique response.

Human Stress Results

- 13 stresses and their active networks yielded 444201 closed biclusters.
- 143 biclusters are network legos.
- The network legos contained between 165 and 1148 proteins.
- The network legos have 95% stability.
- The network legos provide better than 86% recoverability for all active networks.
- We recovered 11 active networks at 100%.

#conditions	5	6	7	8	9	10	11	12
#legos	1	6	10	36	34	20	28	8

Human Stress Results without Cell Cycle Arrest Treatment

- The response networks for cell cycle arrest treatments contain interactions that are distinct compared to the interaction from other treatments.
- 11 stresses yielded only 15 network legos.
- The network legos provide better than 71% recoverability for all active networks.
- We recovered five active networks at 100%.
- Each formula contained at least 7 active networks

WI38 Menadione and WI38 DTT Network Lego

- One network lego contained endoplasmic reticulum stress and oxidative stress to fibroblasts in non-complemented form.
- All other stresses appeared in complemented form.
- This network lego is the only one enriched in functions related to the cell cycle and targets of the E2F1 transcription factor.
- Fibroblasts respond differently from HeLa cells to these two stresses.

Our Contributions

- Combined representation of biological processes using formulae and network legos.
- A formula relates different cellular states or perturbations by explicitly denoting their participation via intersections and complements.
- Each network lego corresponds to a functional module of coherently interacting genes in the universal network.
- Network legos serve as building blocks of active networks.

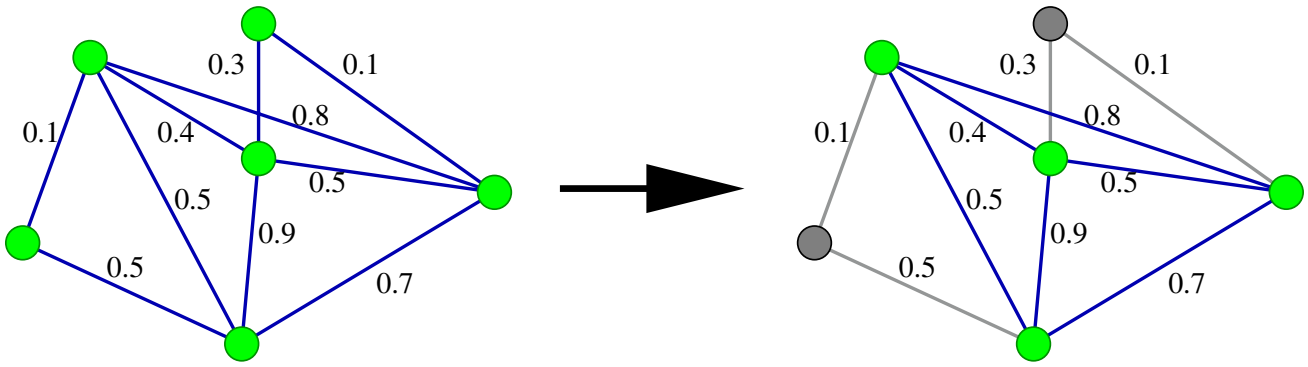
Future Work

- Explore network legos in the context of a larger compendium of cellular stresses.
- Develop an algorithm to directly compute network legos without searching the space of all active network combinations.
- Determine rules and grammar for combining network legos into active networks.

Algorithmic Ingredients: Active Networks

(i) Assign Pearson's correlation
as the interaction weight

(ii) Compute dense subgraphs



- Compute the Pearson's correlation coefficient of the expression profiles of the interacting genes.
- Search for pockets of concerted activity using an algorithm for finding dense subgraphs.