Network Legos: Building Blocks of Cellular Wiring Diagrams

T. M. Murali

August 19, 2008

Goals of Data-Driven Molecular Systems Biology

- Identify the building blocks of wiring diagrams.
- 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 Data-Driven Molecular Systems Biology

- Identify the building blocks of wiring diagrams.
- 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.

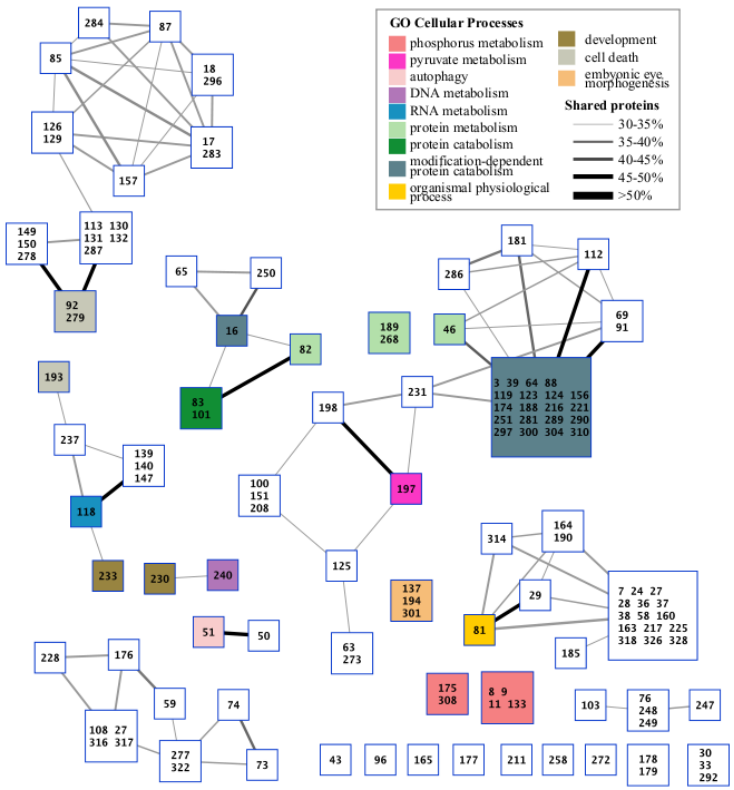
How do we automatically construct these building blocks?

Molecular Interactomes $\longrightarrow$ Modules

- Number of existing techniques decompose interactomes into modules (reviewed by Sharan and Ideker, *Nat. Biotech.*, 2006).
- Map computed modules to known protein complexes, pathways, biological processes, functions, etc.

Molecular Interactomes → Modules

- Number of existing techniques decompose interactomes into modules (reviewed by Sharan and Ideker, *Nat. Biotech.*, 2006).
- Map computed modules to known protein complexes, pathways, biological processes, functions, etc.



(From Sharan et al., *PNAS*, 2005)

Molecular Interactomes $\longrightarrow$ Modules

- Number of existing techniques decompose interactomes into modules (reviewed by Sharan and Ideker, *Nat. Biotech.*, 2006).
- Map computed modules to known protein complexes, pathways, biological processes, functions, etc.

Molecular Interactomes → Modules

- Number of existing techniques decompose interactomes into modules (reviewed by Sharan and Ideker, *Nat. Biotech.*, 2006).
- Map computed modules to known protein complexes, pathways, biological processes, functions, etc.

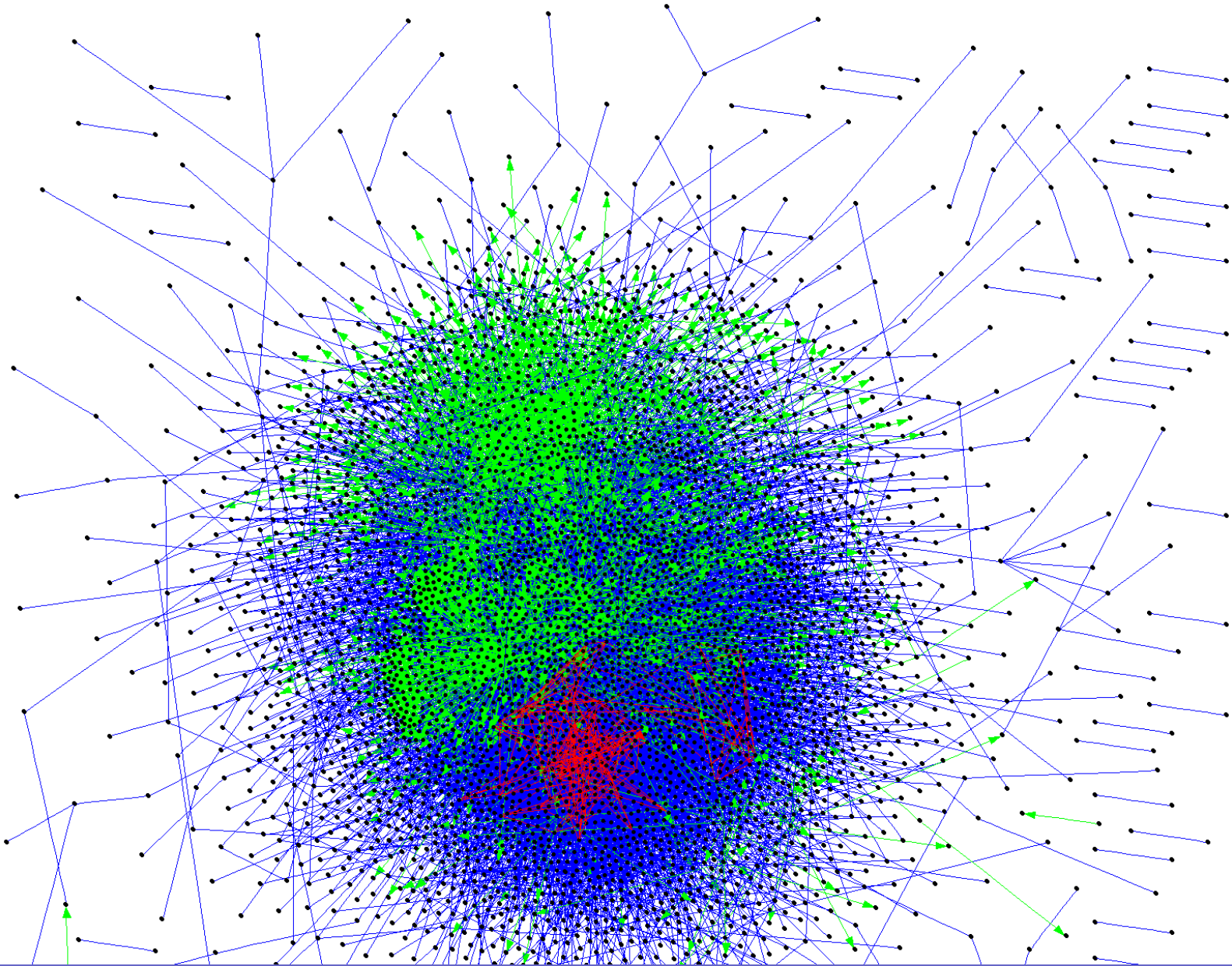
But cell state is dynamic!

- Active molecular interactions change with time, external signals, and perturbations.
- Decompositions of static and *universal* interactomes may miss many important aspects of cellular activity.
- We must integrate interactomes with dynamic measurements of cell state to compute the cell's response to different conditions.

Molecular Interactomes $\longrightarrow$ Active Networks

- Gene expression data provide dynamic snapshots of cellular activity.
- *Active network*: Molecular interactions activated by the cell in response to a stimulus.
- Methods to integrate interactomes with transcriptional measurements to compute active networks:
  - Ideker et al., *Bioinformatics* 2002, *Mol. Sys. Bio* 2007.
  - Bar-Joseph et al., *Nat. Biotech.*, 2003.
  - Luscombe et al., *Nature* 2004.
  - Ulitsky and Shamir, *BMC Sys Bio* 2007.
  - Murali and Rivera, *Journal of Computational Biology*, 2008.

Molecular Interactomes → Active Networks



Molecular Interactomes $\longrightarrow$ Active Networks

- Gene expression data provide dynamic snapshots of cellular activity.
- *Active network*: Molecular interactions activated by the cell in response to a stimulus.
- Methods to integrate interactomes with transcriptional measurements to compute active networks:
  - Ideker et al., *Bioinformatics* 2002, *Mol. Sys. Bio* 2007.
  - Bar-Joseph et al., *Nat. Biotech.*, 2003.
  - Luscombe et al., *Nature* 2004.
  - Ulitsky and Shamir, *BMC Sys Bio* 2007.
  - Murali and Rivera, *Journal of Computational Biology*, 2008.

Molecular Interactomes → Active Networks

- Gene expression data provide dynamic snapshots of cellular activity.
- *Active network*: Molecular interactions activated by the cell in response to a stimulus.
- Methods to integrate interactomes with transcriptional measurements to compute active networks:
  - Ideker et al., *Bioinformatics* 2002, *Mol. Sys. Bio* 2007.
  - Bar-Joseph et al., *Nat. Biotech.*, 2003.
  - Luscombe et al., *Nature* 2004.
  - Ulitsky and Shamir, *BMC Sys Bio* 2007.
  - Murali and Rivera, *Journal of Computational Biology*, 2008.
- These methods usually compute active networks one condition at a time or simultaneously across multiple conditions.

Hypotheses Guiding a New Approach

1. Wiring diagram can be broken up into a number of multiple overlapping modules, where each module is a network of coherently-interacting molecules.

Hypotheses Guiding a New Approach

1. Wiring diagram can be broken up into a number of multiple overlapping modules, where each module is a network of coherently-interacting molecules.
2. Cell responds to a stress by appropriately modulating the activity of a subset of modules.

Hypotheses Guiding a New Approach

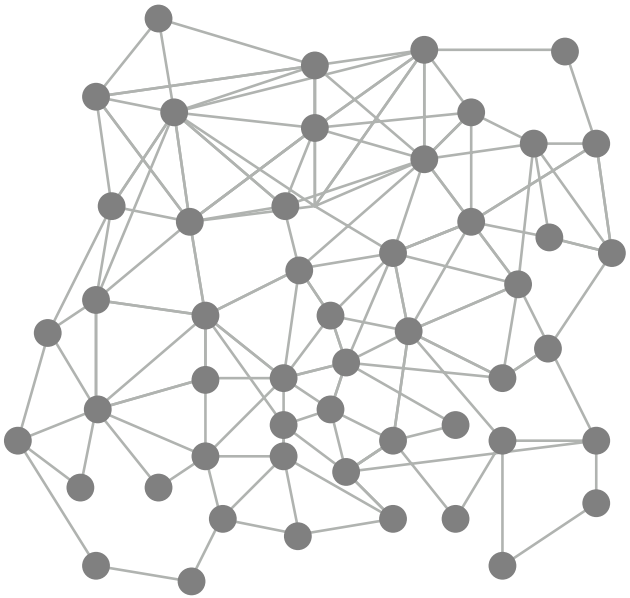
1. Wiring diagram can be broken up into a number of multiple overlapping modules, where each module is a network of coherently-interacting molecules.
2. Cell responds to a stress by appropriately modulating the activity of a subset of modules.

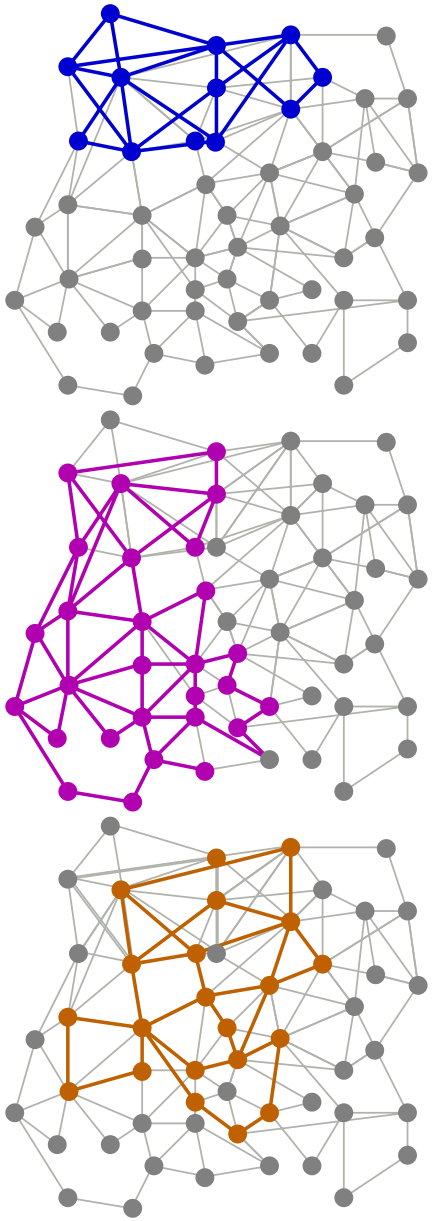
Idea: turn second hypothesis on its head to compute modules from multiple response networks.

Goals of the Network Lego Approach

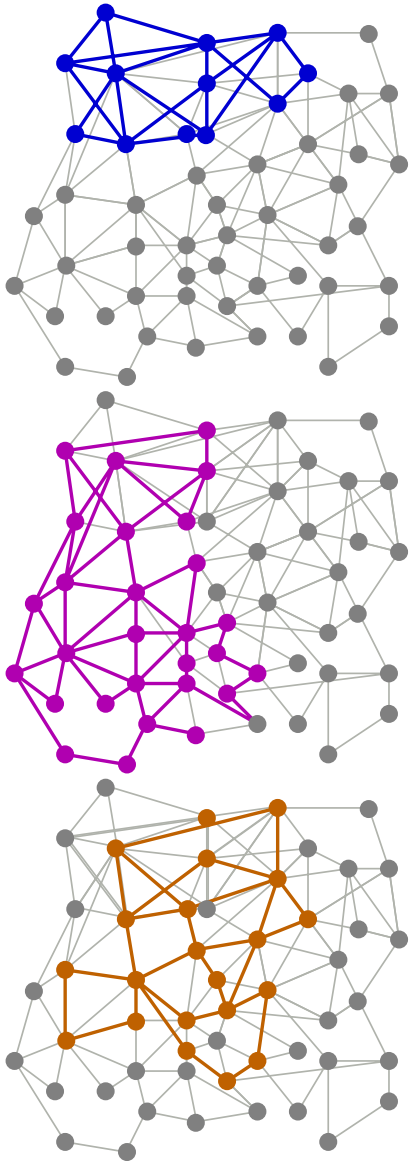
- Combine active network computation with module detection to compute *network legos*: context-sensitive building blocks of wiring diagrams.
- Potential applications:
  1. Identify pathways uniquely activated in one or more conditions.
  2. Compare and contrast responses of different cell types to the same stress.
  3. Develop a formalism for expressing any active network as a combination of network legos.

Step 1: Molecular Interactome to Active Networks

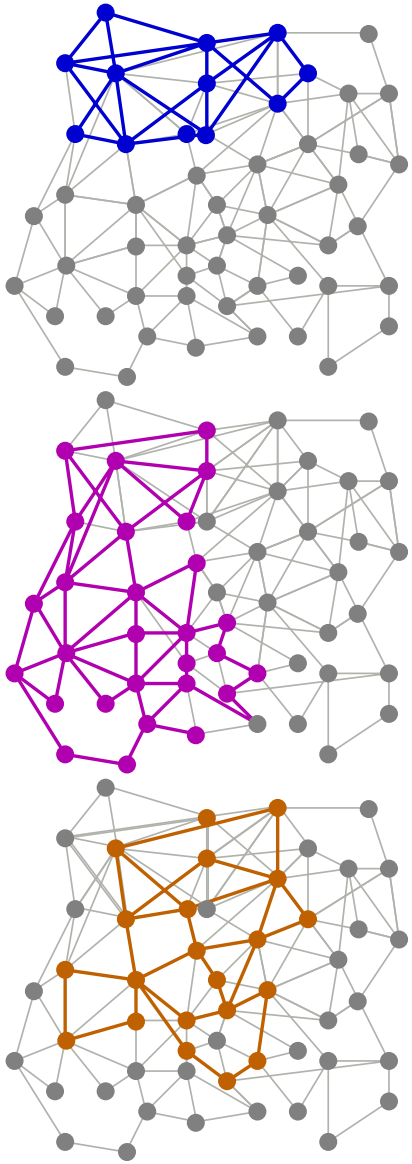


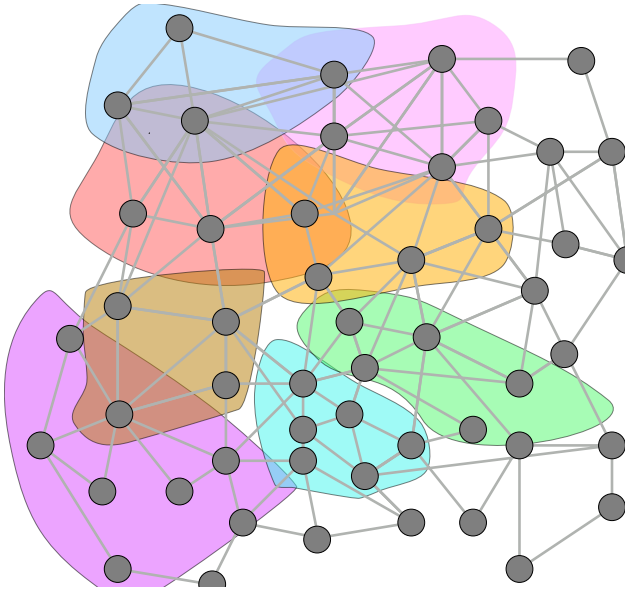


Step 2: Active Networks to Network Legos

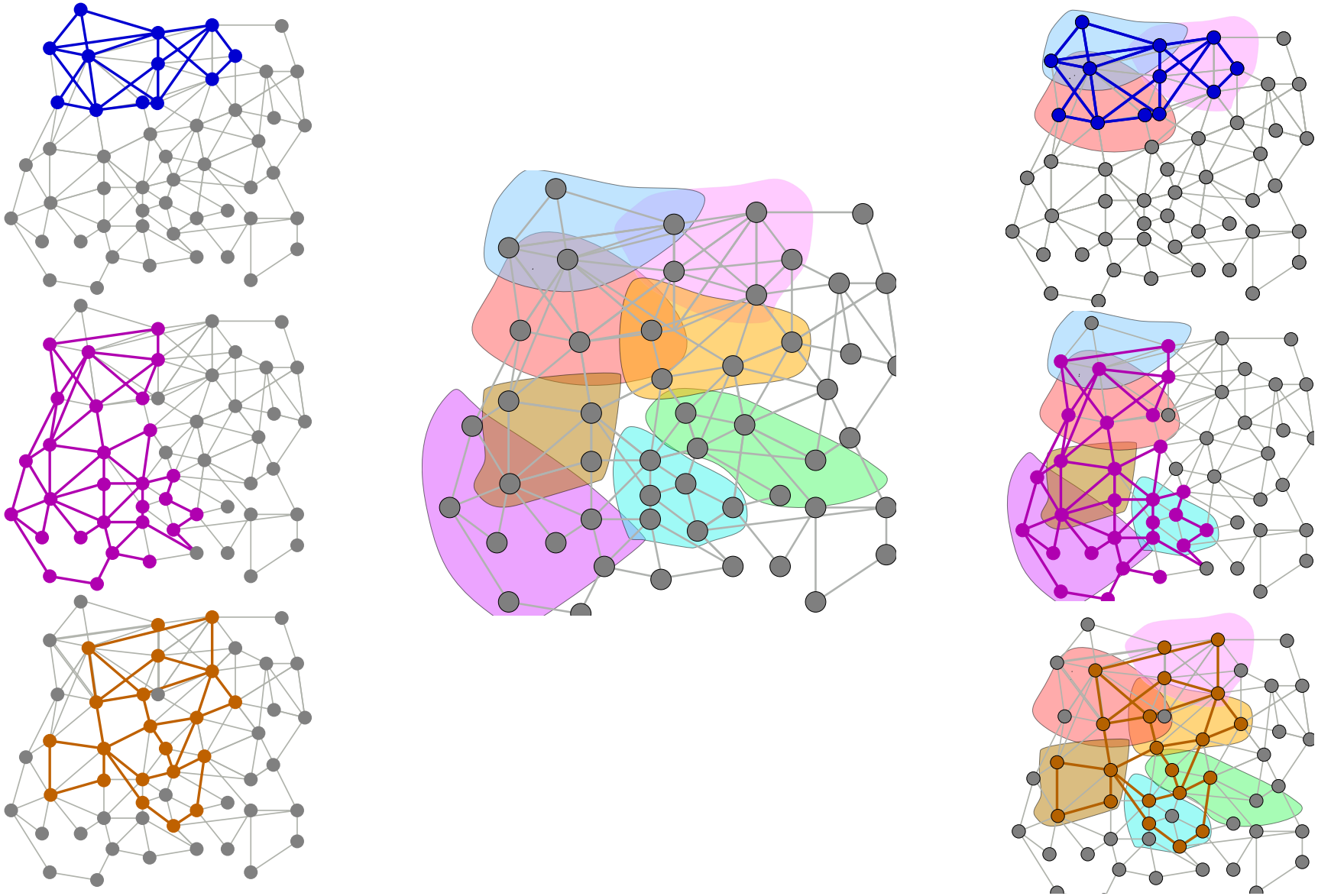


Step 2: Active Networks to Network Legos





Step 2: Active Networks to Network Legos



Caveats

- Interactomes are incomplete and noisy.
- Gene expression measurements miss many aspects of cellular state.
- We will consider only presence or absence of an interaction in an active network.
- Network legos are only a mental model of how the cell may operate.

Network Blocks

- Suppose we have gene expression datasets for a number of conditions.
- Compute the active network for each condition.
  - Consider each active network to be a set of interactions.
  - Any set operation on these active networks will yield another network of interactions.

Network Blocks

- Suppose we have gene expression datasets for a number of conditions.
- Compute the active network for each condition.
  - Consider each active network to be a set of interactions.
  - Any set operation on these active networks will yield another network of interactions.
- Let $\mathcal{A}$ be the set of all active networks.
- A *block* is a triple $(G, \mathcal{I}, \mathcal{E})$ where
  - $\mathcal{I} \subseteq \mathcal{A}$, $\mathcal{I}$ is non-empty.
  - $\mathcal{E} \subseteq \mathcal{A}$, disjoint from $\mathcal{I}$.
  - $\mathcal{I}$ and $\mathcal{E}$ are inclusion-maximal
  - G is a network where each interaction
    - is present in every active network in $\mathcal{I}$.
    - is absent in every active network in $\mathcal{E}$.

$$G = \left(\bigcap_{P \in \mathcal{I}} P \right)$$

Network Blocks

- Suppose we have gene expression datasets for a number of conditions.
- Compute the active network for each condition.
  - Consider each active network to be a set of interactions.
  - Any set operation on these active networks will yield another network of interactions.
- Let $\mathcal{A}$ be the set of all active networks.
- A *block* is a triple $(G, \mathcal{I}, \mathcal{E})$ where
  - $\mathcal{I} \subseteq \mathcal{A}$, $\mathcal{I}$ is non-empty.
  - $\mathcal{E} \subseteq \mathcal{A}$, disjoint from $\mathcal{I}$.
  - $\mathcal{I}$ and $\mathcal{E}$ are inclusion-maximal
  - G is a network where each interaction
    - is present in every active network in $\mathcal{I}$.
    - is absent in every active network in $\mathcal{E}$.

$$G = \left(\bigcap_{P \in \mathcal{I}} P \right) - \left(\bigcup_{N \in \mathcal{E}} N \right)$$

ALL, AML, and MLL

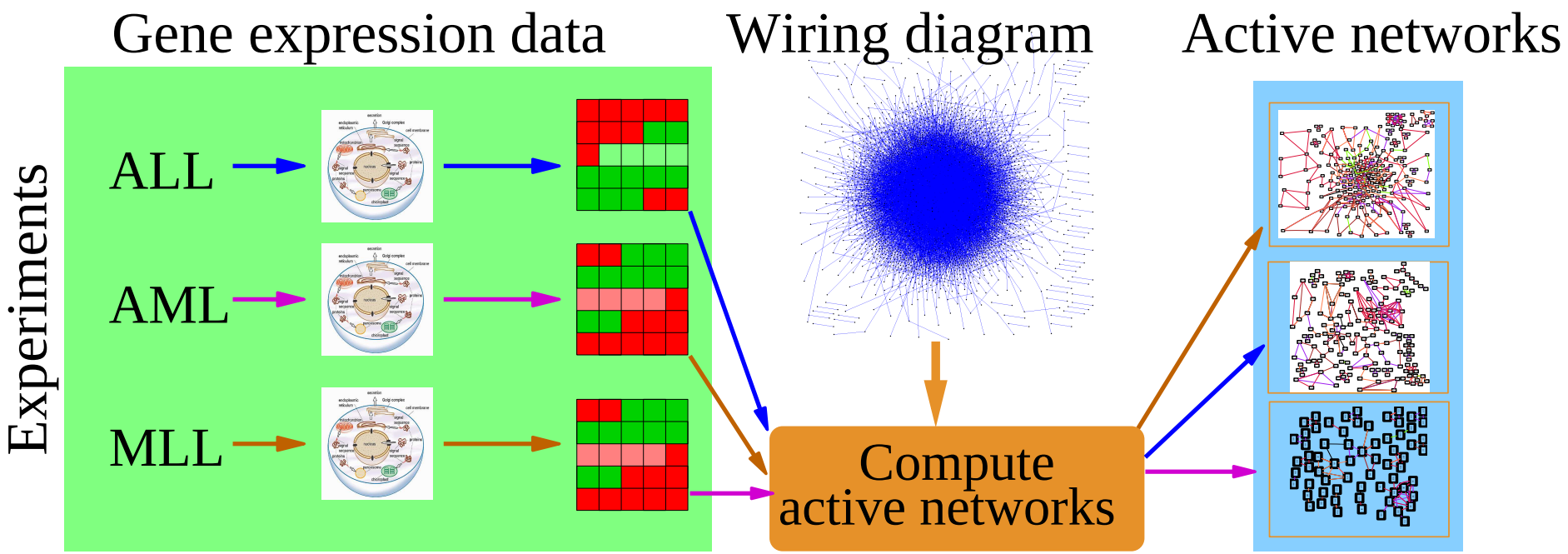
- Acute Lymphoblastic Leukaemia (ALL) and Acute Myeloid Leukaemia (AML) are two types of leukaemia.
- Armstrong et al., Nature Genetics 2003 argued that translocations in the Mixed Lineage Leukaemia (MLL) gene identify a disease distinct from ALL and AML.

ALL, AML, and MLL

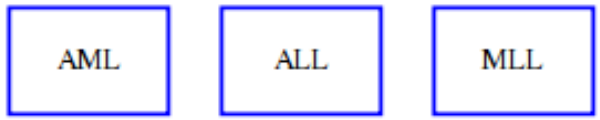
- Acute Lymphoblastic Leukaemia (ALL) and Acute Myeloid Leukaemia (AML) are two types of leukaemia.
- Armstrong et al., Nature Genetics 2003 argued that translocations in the Mixed Lineage Leukaemia (MLL) gene identify a disease distinct from ALL and AML.

Can we compare active networks to identify subsets of interactions differentially activated in each leukaemia?

Comparing ALL, AML, and MLL

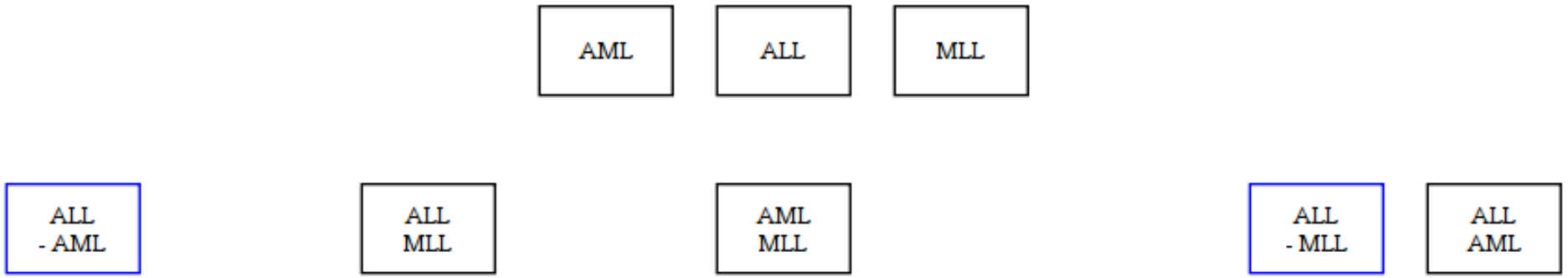


Computing 17 Comparisons

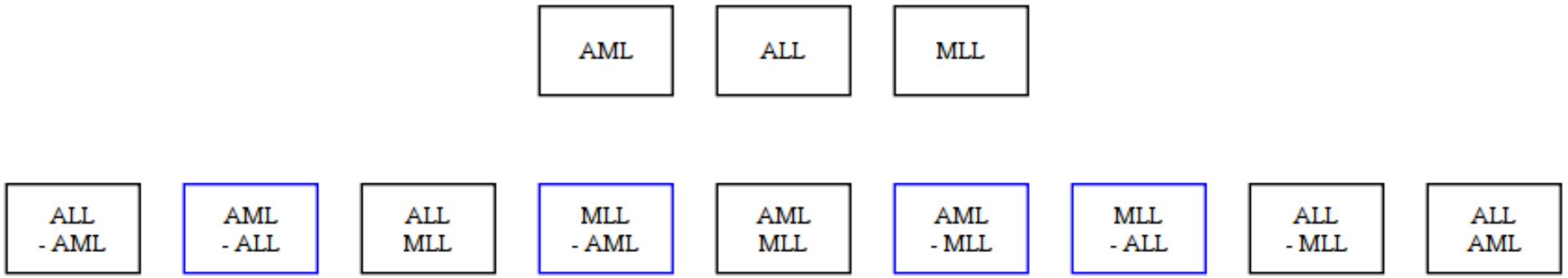


Computing 17 Comparisons

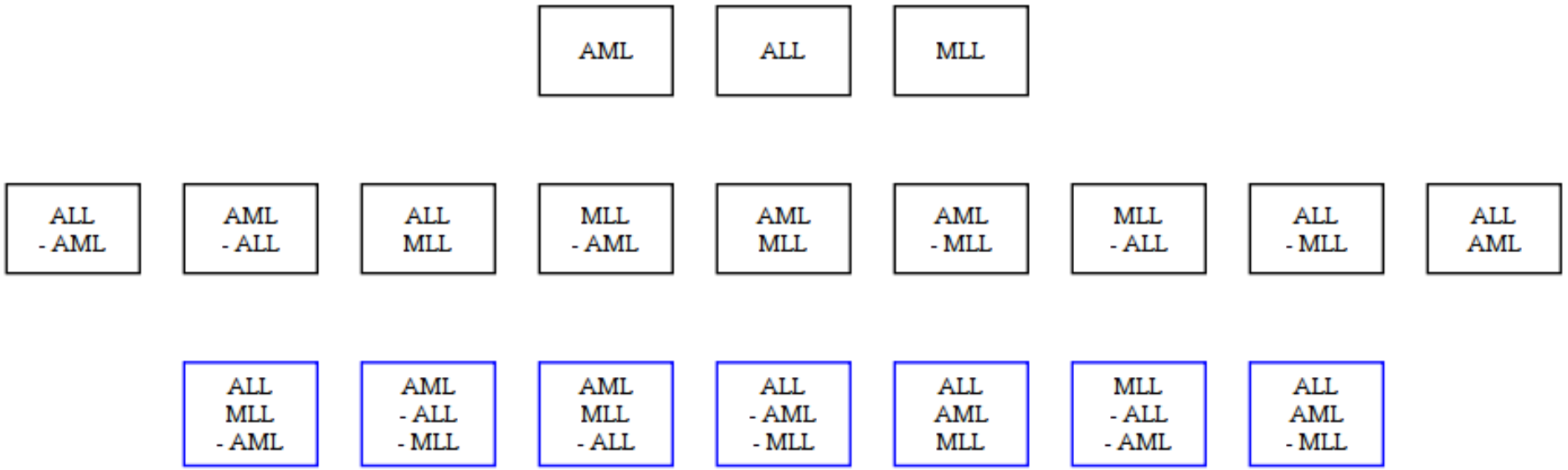
Computing 17 Comparisons



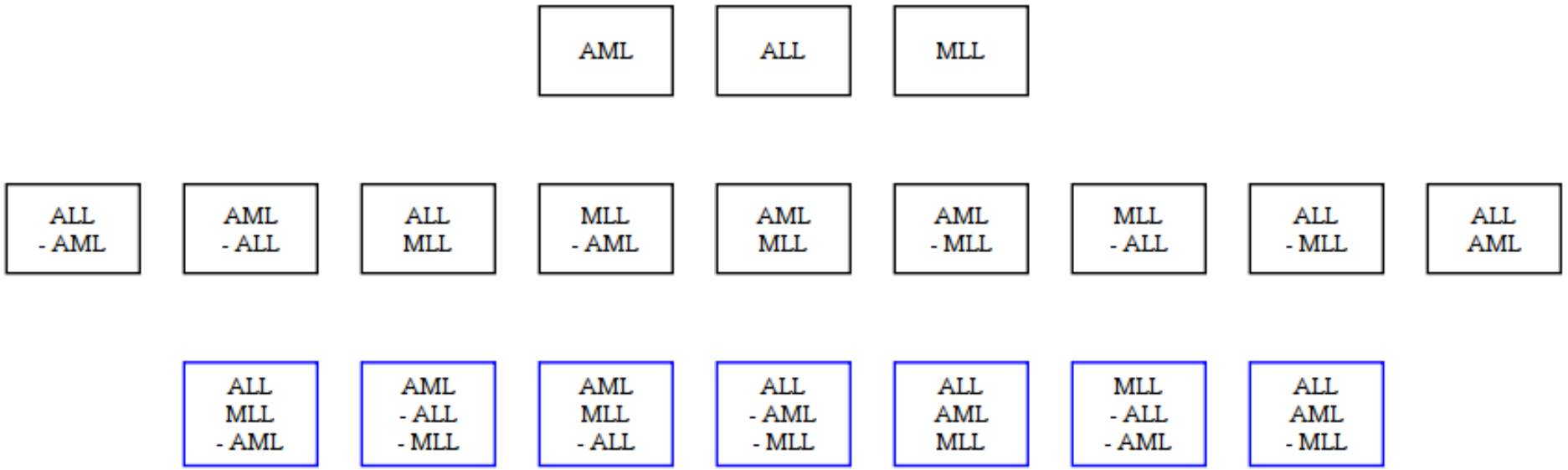
Computing 17 Comparisons



Computing 17 Comparisons

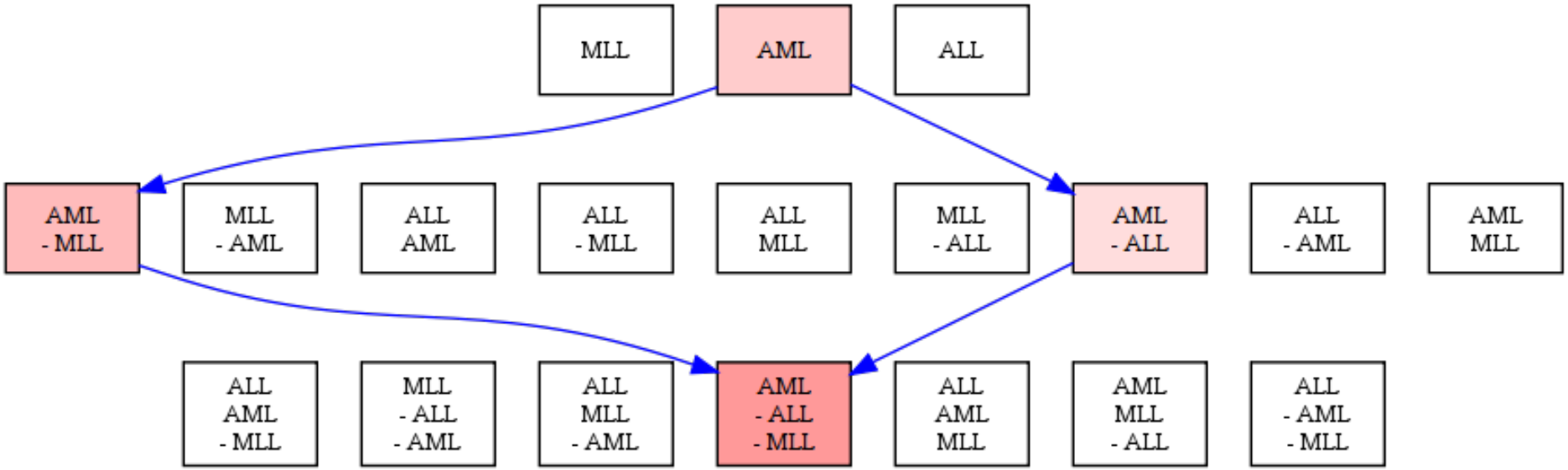


Computing 17 Comparisons



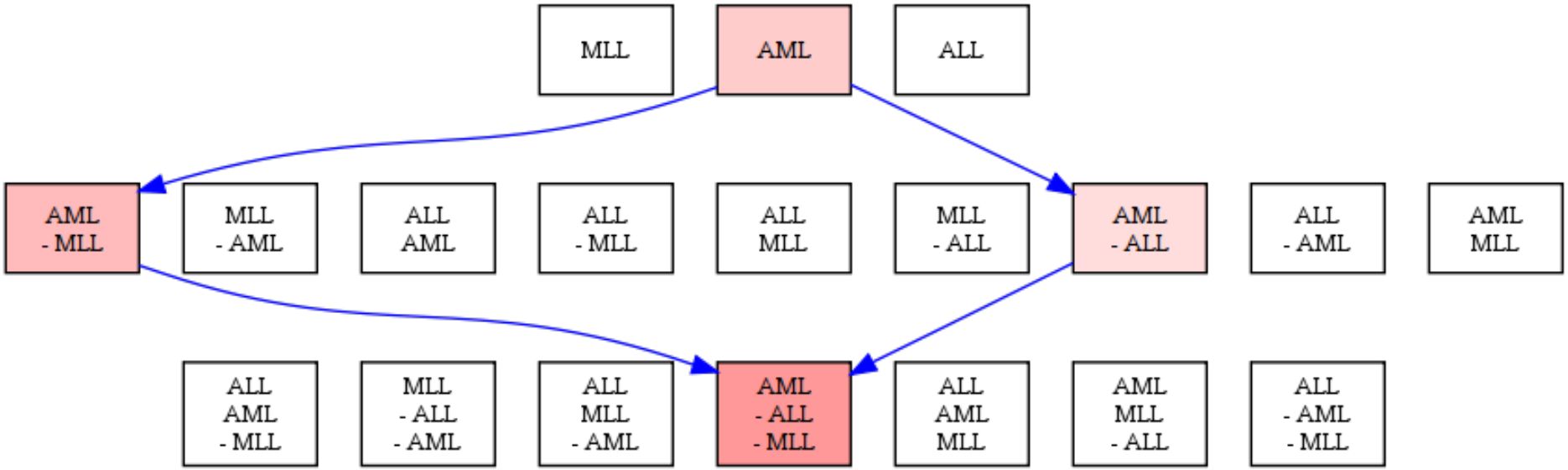
- Each node represents
  1. a boolean conjunction of (possibly negated) conditions and
  2. a network of interactions
- Can compute enrichment of known processes or pathways (Gene Ontology, Netpath, REACTOME, etc.) in each network.

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 can we efficiently compute all combinations?
- How do we identify which combinations are the network legos?
- How do we demonstrate that the network legos we have found are building blocks?

Challenges in Comparing Arbitrary Numbers of Active Networks

- How can we efficiently compute all combinations?
  - Construct binary matrix of interactions vs. active networks and use closed itemset mining algorithms.
- How do we identify which combinations are the network legos?
- How do we demonstrate that the network legos we have found are building blocks?

Challenges in Comparing Arbitrary Numbers of Active Networks

- How can we efficiently compute all combinations?
  - Construct binary matrix of interactions vs. active networks and use closed itemset mining algorithms.
- How do we identify which combinations are the network legos?
  - Compute statistical significance of each combination and exploit DAG structure.
- How do we demonstrate that the network legos we have found are building blocks?

Challenges in Comparing Arbitrary Numbers of Active Networks

- How can we efficiently compute all combinations?
  - Construct binary matrix of interactions vs. active networks and use closed itemset mining algorithms.
- How do we identify which combinations are the network legos?
  - Compute statistical significance of each combination and exploit DAG structure.
- How do we demonstrate that the network legos we have found are building blocks?
  - Define and measure stability and recoverability.

Network Blocks

- Let $\mathcal{A}$ be the set of all active networks.
- A *block* is a triple $(G, \mathcal{I}, \mathcal{E})$ where
  - $\mathcal{I} \subseteq \mathcal{A}$, $\mathcal{I}$ is non-empty.
  - $\mathcal{E} \subseteq \mathcal{A}$, disjoint from $\mathcal{I}$.
  - $\mathcal{I}$ and $\mathcal{E}$ are inclusion-maximal such that

$$G = \left(\bigcap_{P \in \mathcal{I}} P \right) \bigcap \left(\bigcap_{N \in \mathcal{E}} !N \right)$$

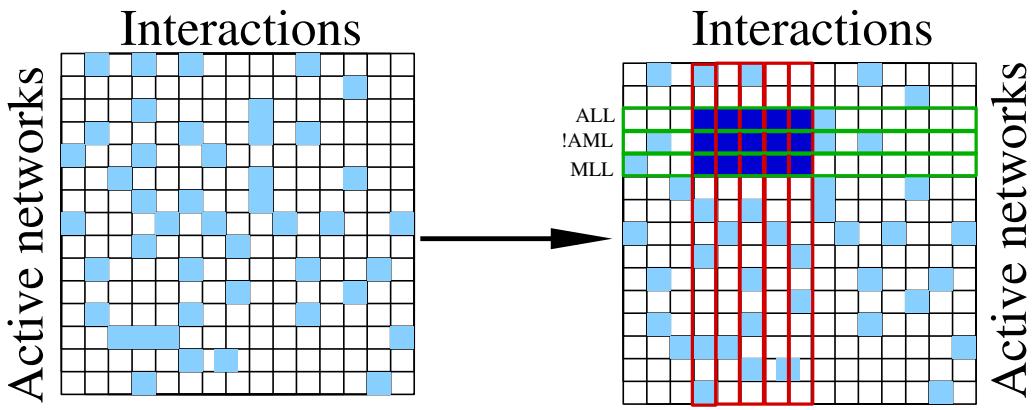
Network Blocks

- Let $\mathcal{A}$ be the set of all active networks.
- A *block* is a triple $(G, \mathcal{I}, \mathcal{E})$ where
  - $\mathcal{I} \subseteq \mathcal{A}$, $\mathcal{I}$ is non-empty.
  - $\mathcal{E} \subseteq \mathcal{A}$, disjoint from $\mathcal{I}$.
  - $\mathcal{I}$ and $\mathcal{E}$ are inclusion-maximal such that

$$G = \left(\bigcap_{P \in \mathcal{I}} P \right) \bigcap \left(\bigcap_{N \in \mathcal{E}} !N \right)$$

- Partial order exists between blocks, e.g.,
  - $\mathsf{ALL} < \mathsf{ALL} \cap \mathsf{AML}$.
  - $\mathsf{ALL} \cap \mathsf{MLL} < \mathsf{ALL} \cap \mathsf{MLL} \cap !\mathsf{AML}$.

Efficiently Compute Network Blocks

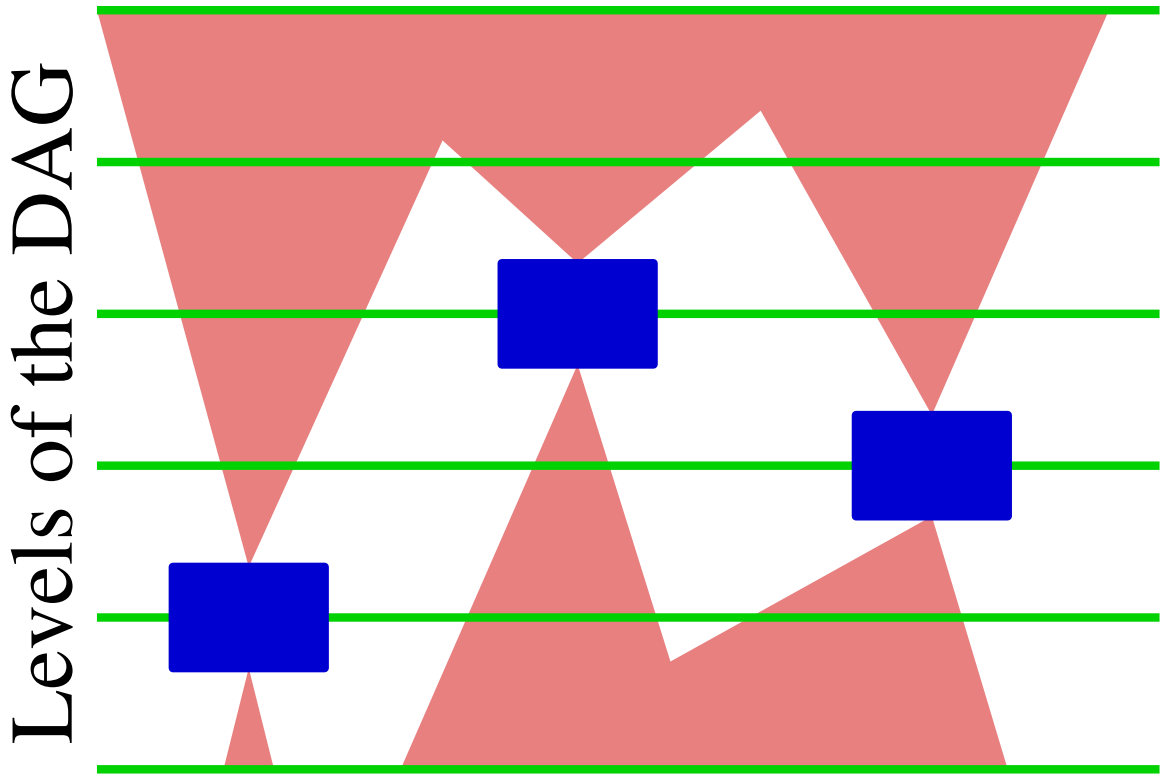


- Construct a binary matrix M whose columns are interactions.
- Represent each active network and its complement in M's rows.
- A *bicluster* is a subset of rows and subset of columns such that M only has 1s in this submatrix.
  - Rows of bicluster $\equiv$ formula.
  - Columns of bicluster $\equiv$ network.
- Compute all closed biclusters in M.
- Connect biclusters in the DAG induced by the partial order.

Assessing Statistical Significance of a Bicluster

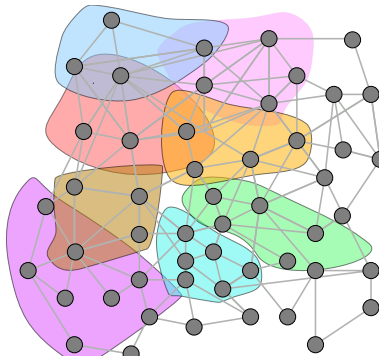
- Suppose a bicluster B has n included and c excluded active networks.
  1. Pick n active networks and the complements of c active networks repeatedly at random, compute the number of interactions induced by this combination, and build a distribution of the number of interactions.
  2. Set the p-value of B to be the fraction of random biclusters with more interactions than B.

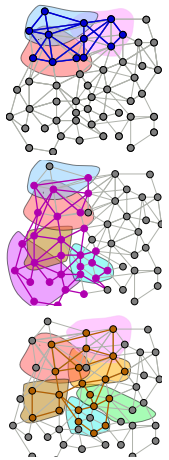
Identify Network Blocks that are Network Legos



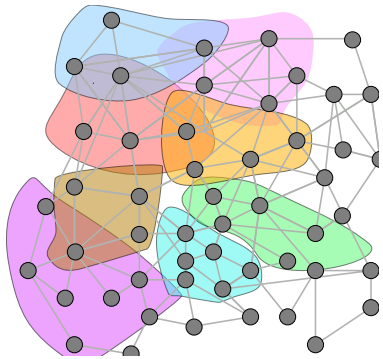
- Assess the statistical significance of each bicluster by simulation.
- B is a *network lego* if it is more significant than any of its ancestors or descendants in the DAG.

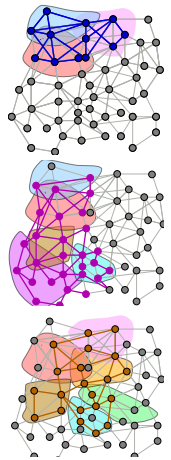
Show that Network Legos are Building Blocks





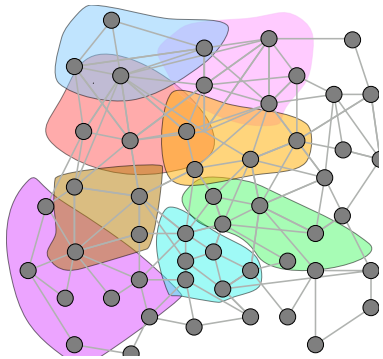
Show that Network Legos are Building Blocks

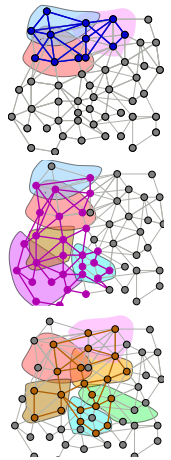




- Stability
  - Remove each active network and recompute network legos.
  - For each original network lego, compute the fraction of leave one out datasets for which the network lego occurs with at least t% fidelity.

Show that Network Legos are Building Blocks





- Stability
  - Remove each active network 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 t% fidelity.
- Recoverability
  - Compute the union of network legos.
  - Measure the size of the intersection of each active network with union.

Analysis of Human Stress Data

- Human protein-protein interaction network with 9243 proteins and 31000 interactions.
  - PPIs from (Ramani et al., Genome Biology, 2005; Rual et al., Nature, 2005; Stelzl et al., Cell, 2005).
- 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.
- Each network lego has 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 active networks for cell cycle arrest treatments contain interactions that are distinct compared to those in active networks for 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 Response to Menadione and DTT

- 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 wiring diagram.
- Network legos serve as building blocks of active networks.

Compendium Approach