Integrating neuroimaging with systems biology

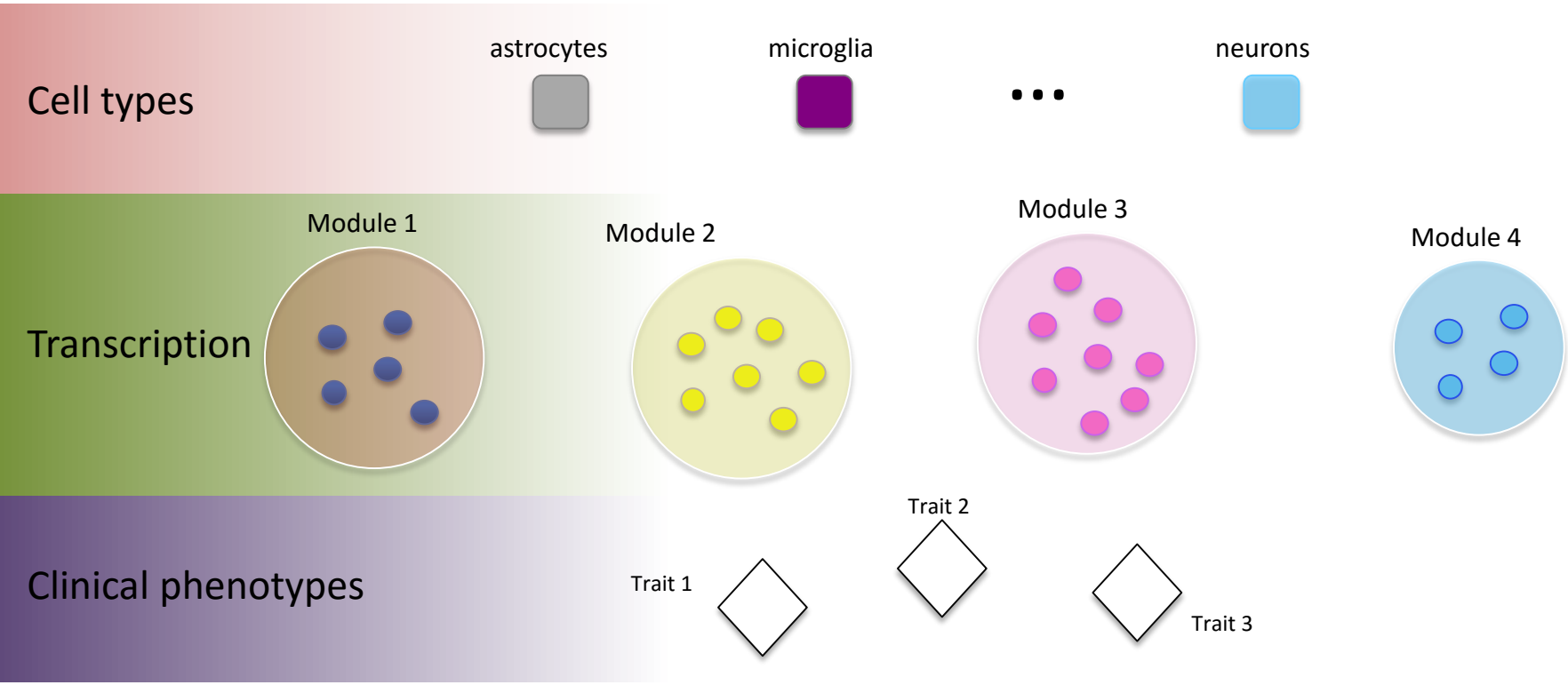
Chris Gaiteri, PhD – Rush University

Sara Mostafavi, PhD – University of British Columbia

Robert Dawe, PhD – Rush University

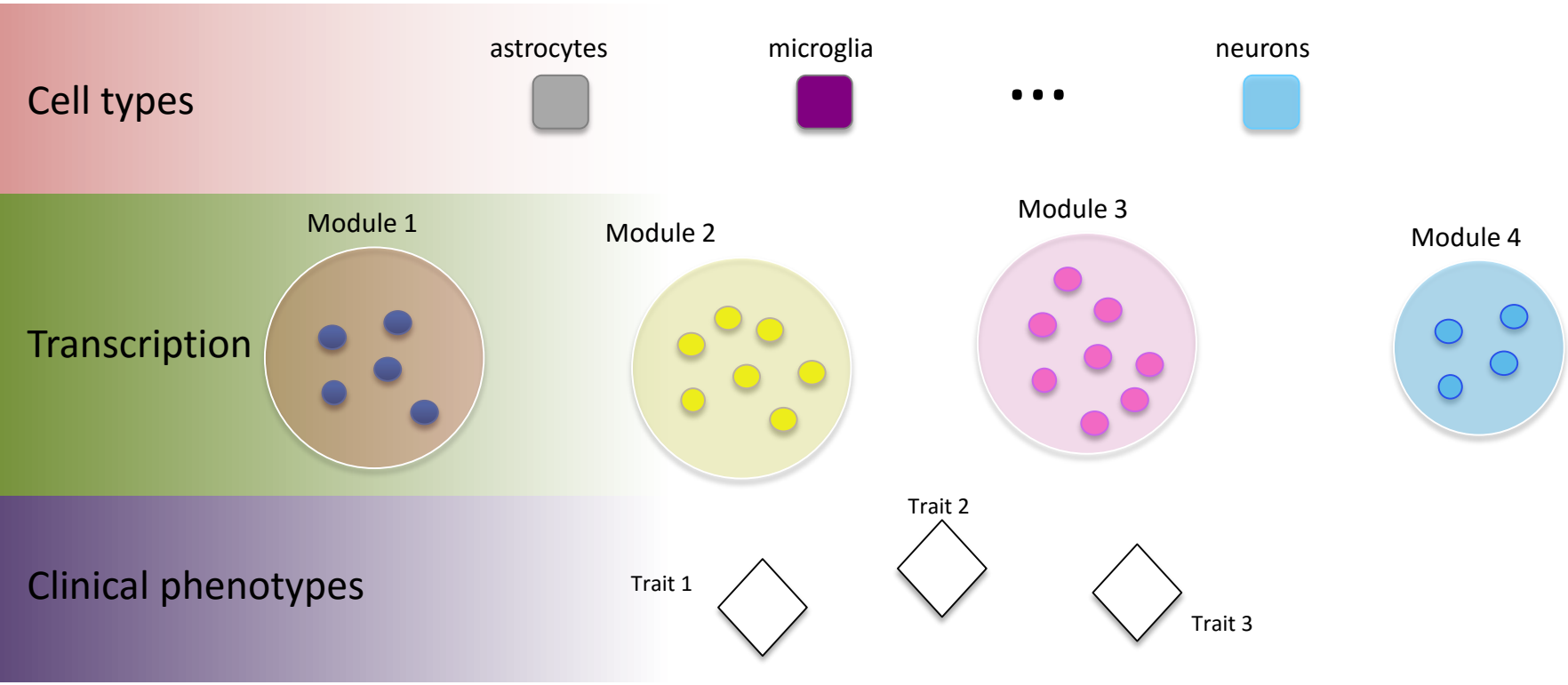
Konstantinos Arfanakis, PhD – Rush University

Combining perspectives on AD pathology



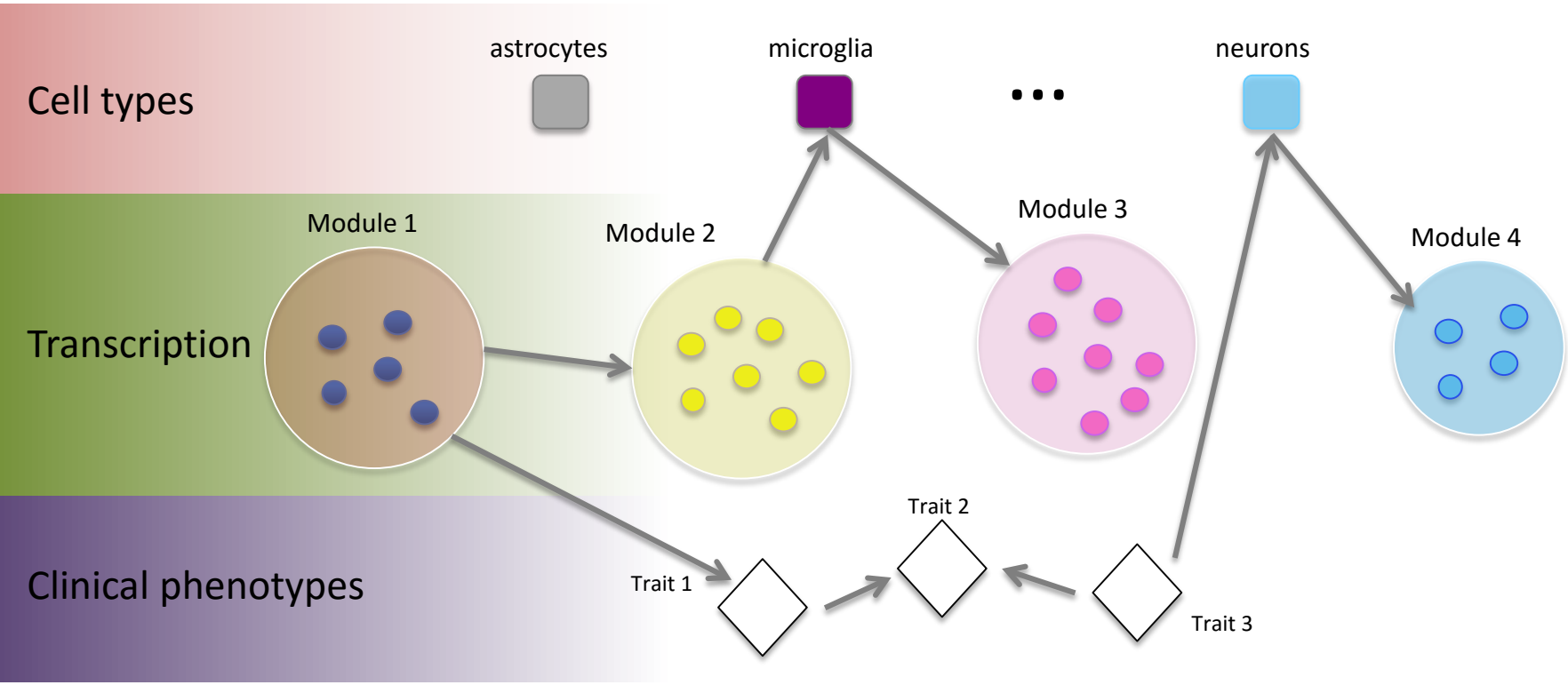
There are many parallel approaches to tracking AD severity

Combining perspectives on AD pathology



There are many parallel approaches to tracking AD severity

Combining perspectives on AD pathology



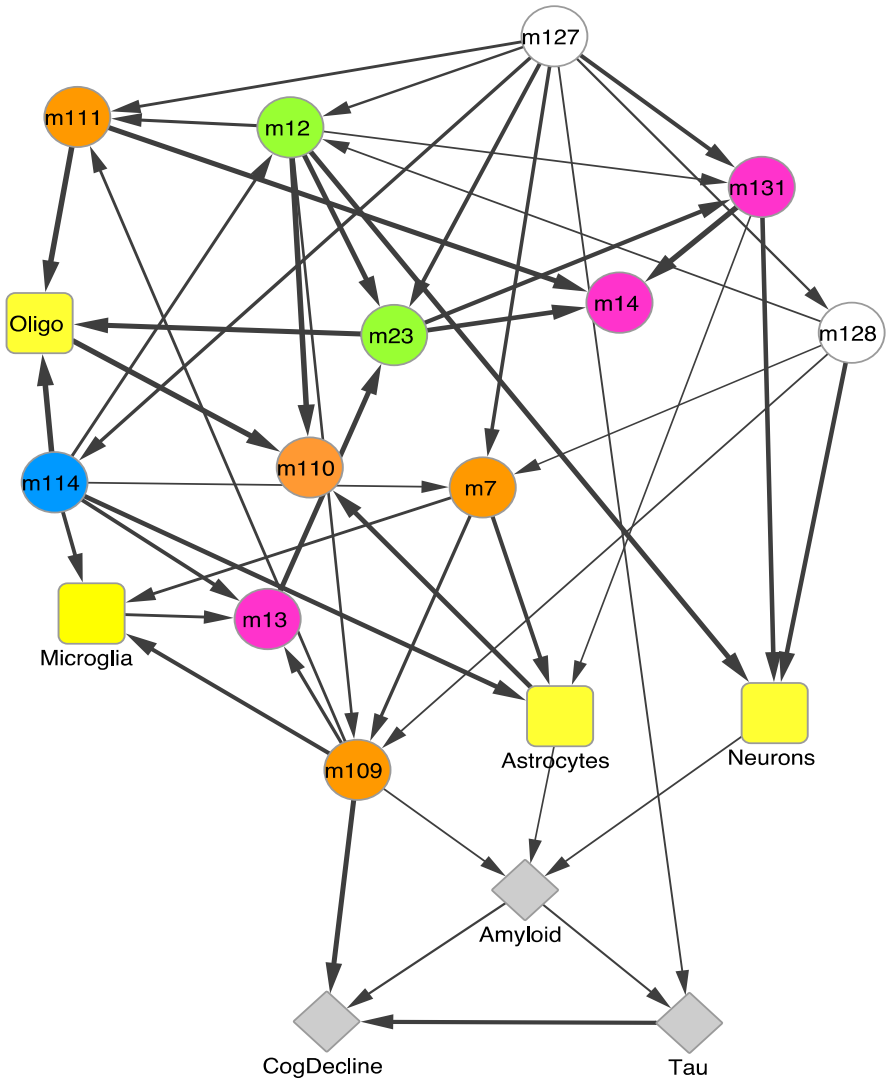
Linking these entities can help to find the molecular basis of pathology

Bringing together the ingredients of AD:

This directed graph describes the conditional independence relationships among expression modules, traits, and cell composition

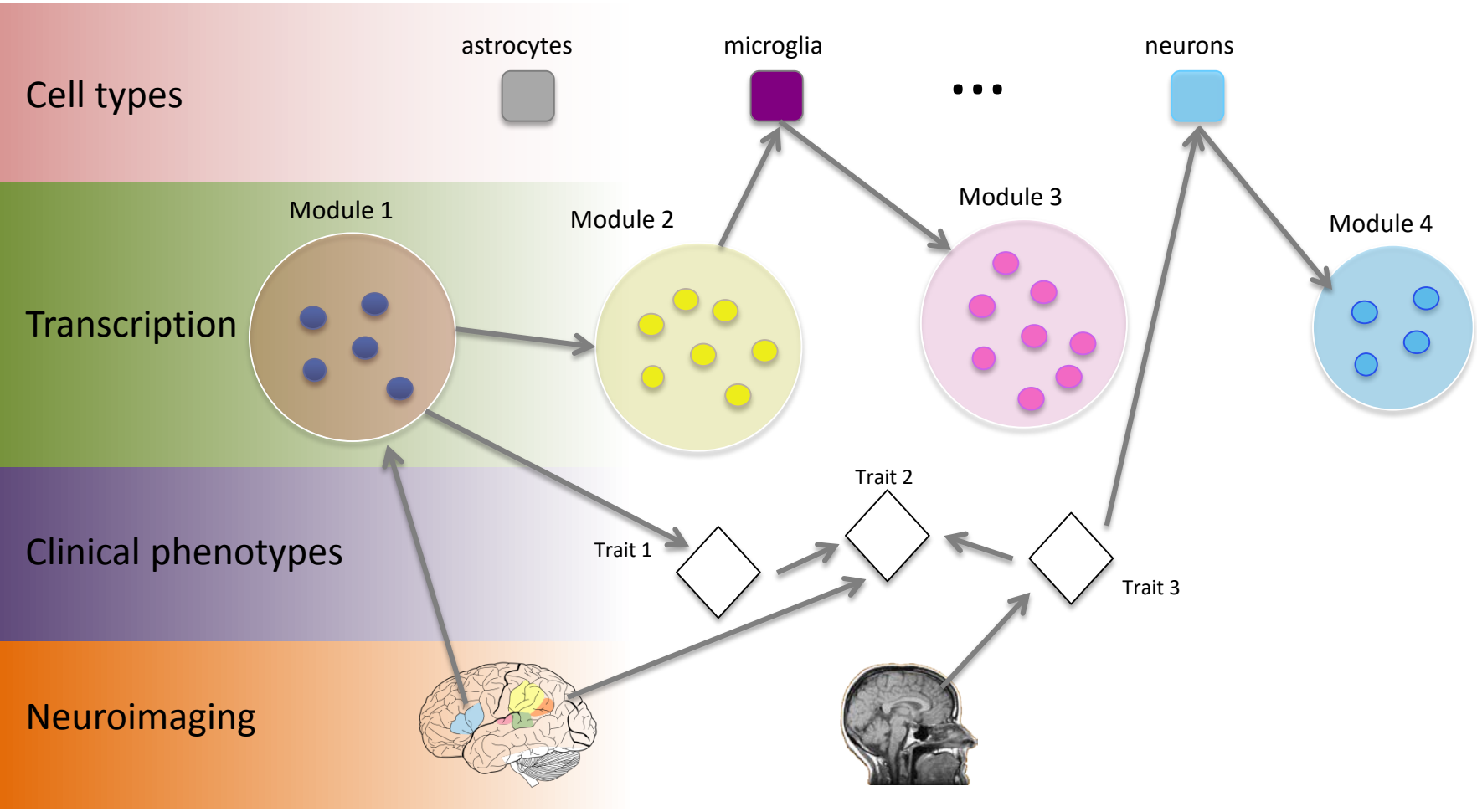
- AD traits
- Cell type markers
- Cell cycle
- Immune response
- Mitochondria funct.
- Neuronal funct.
- NA

} Molecular systems



To use method yourself, search for “Cinderella” on *Synapse* (synapase.sagebase.org)

Combining perspectives on AD pathology



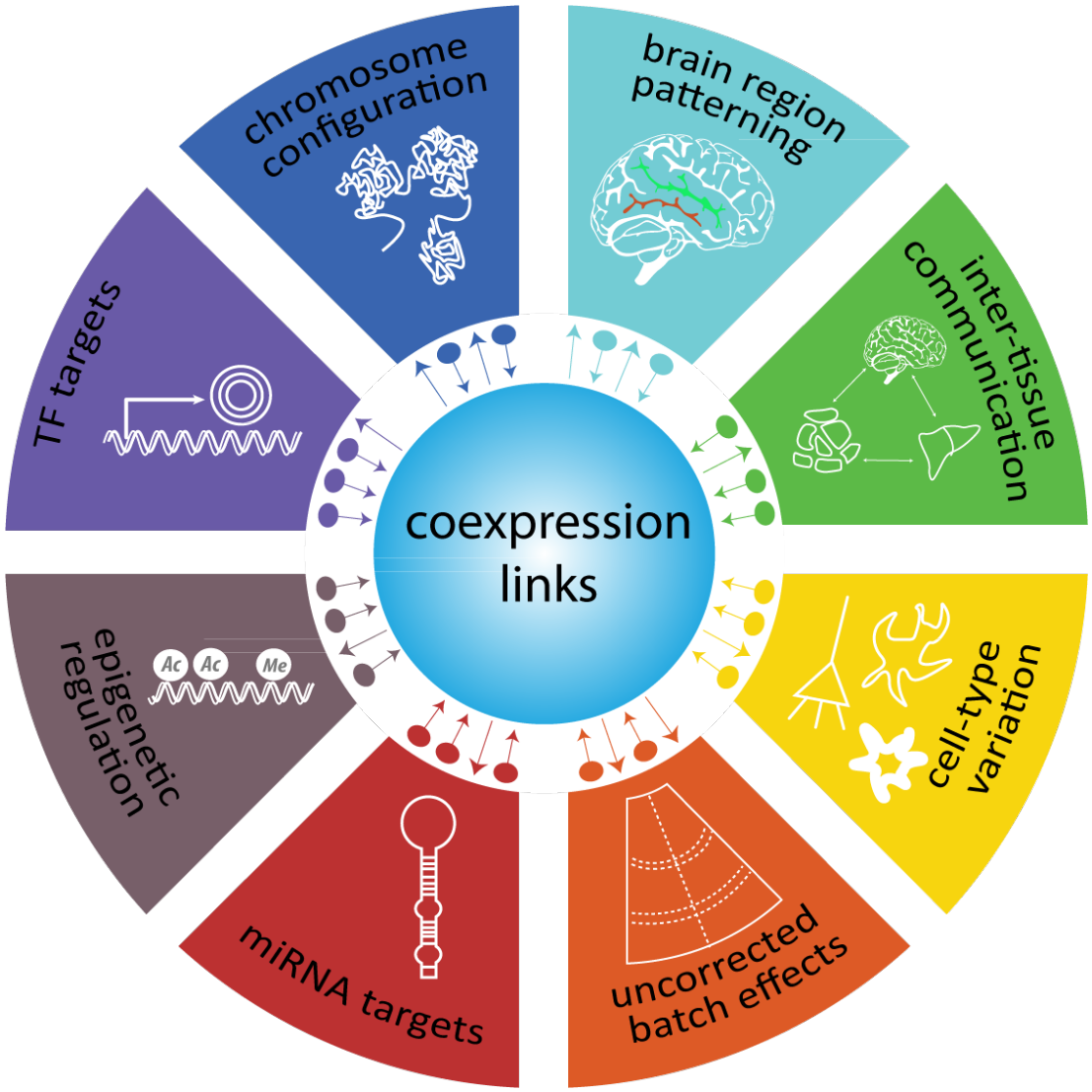
Summary of approach to “imaging-expression”

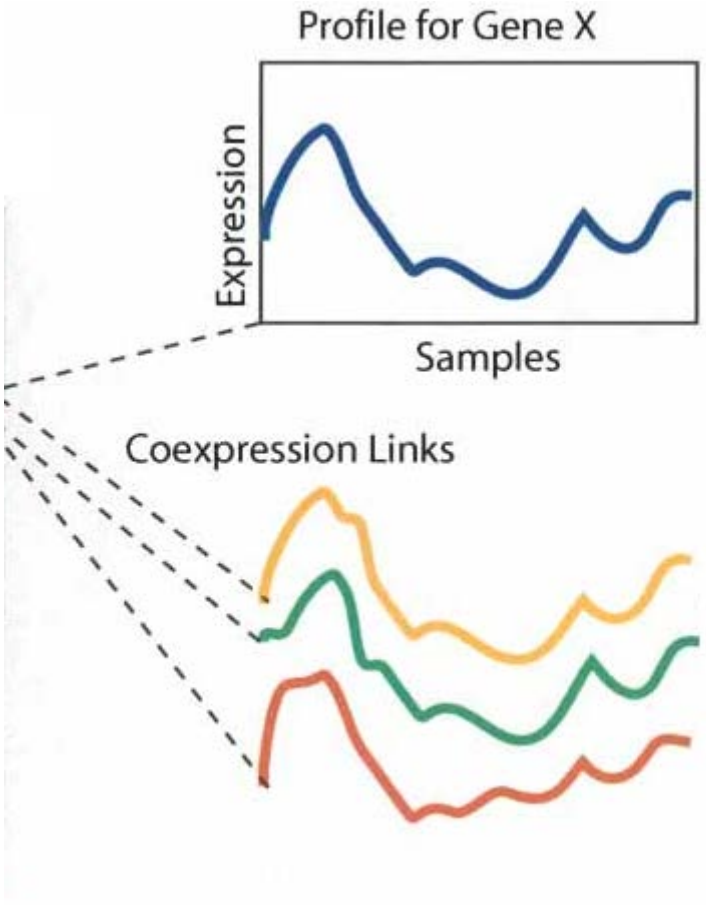
Using the ROSMAP cohort…

1. We summarize gene expression into molecular systems
2. We relate the activity of these molecular system to brain regions
3. We do this for each molecular system

Foundation of expression-imaging: sources of coexpression

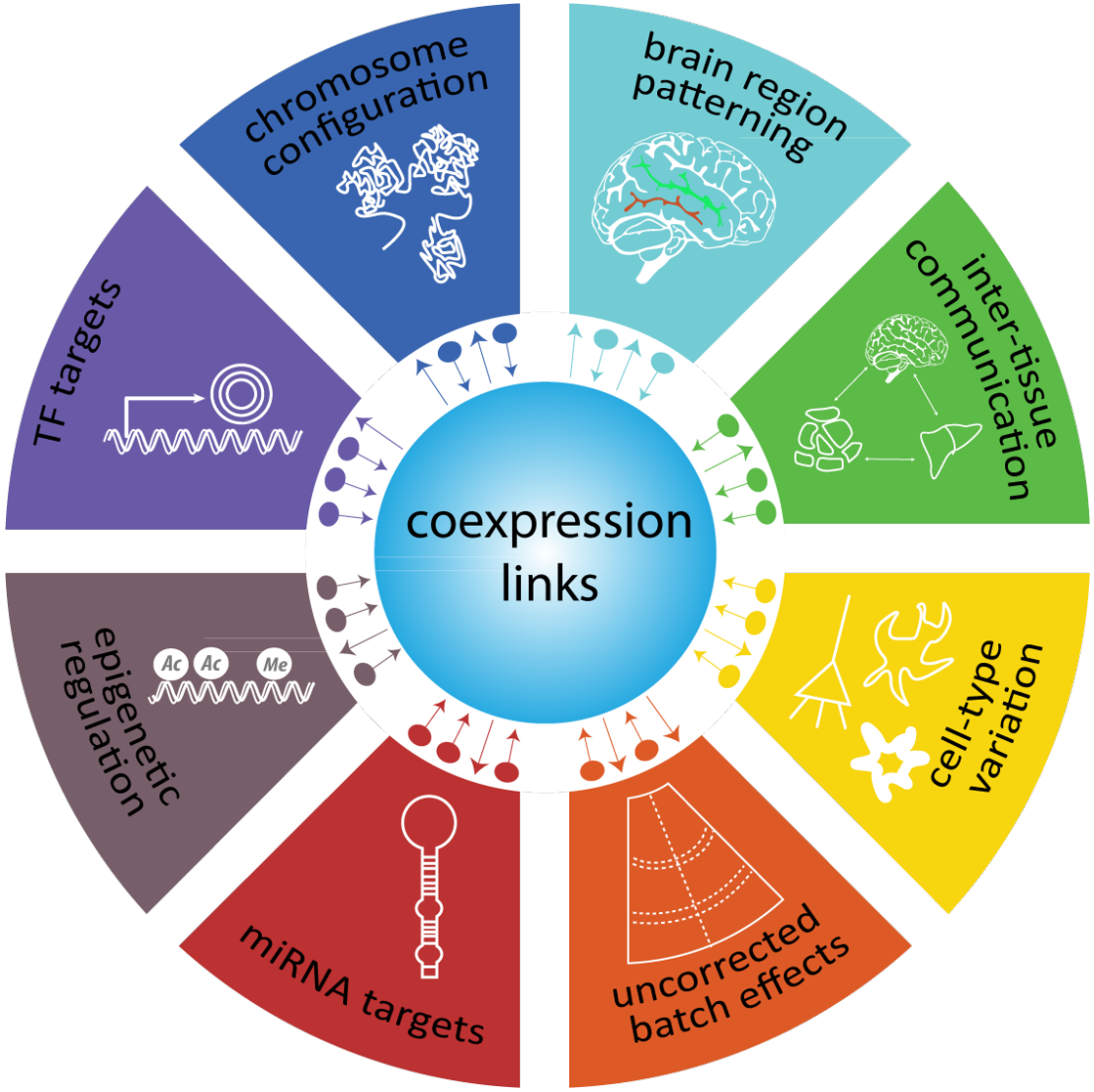
Several mechanisms generate coexpression networks



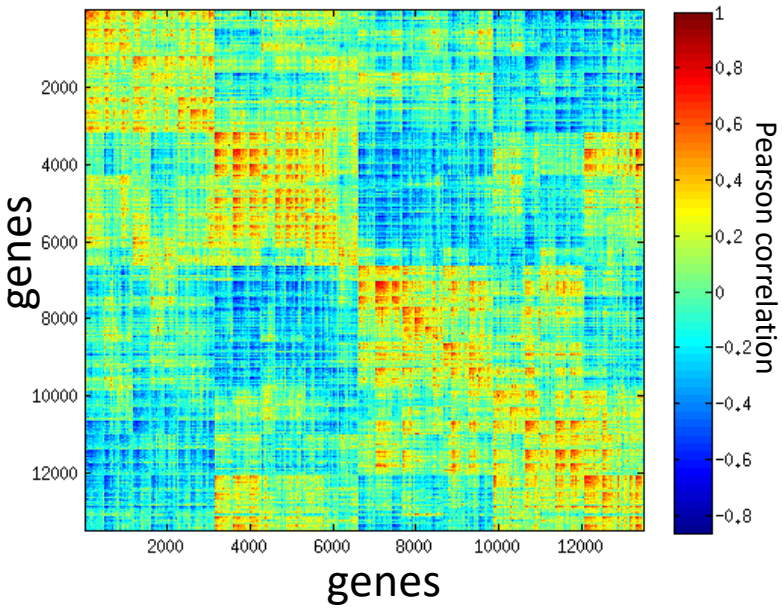


Foundation of expression-imaging: sources of coexpression

Several mechanisms generate coexpression networks



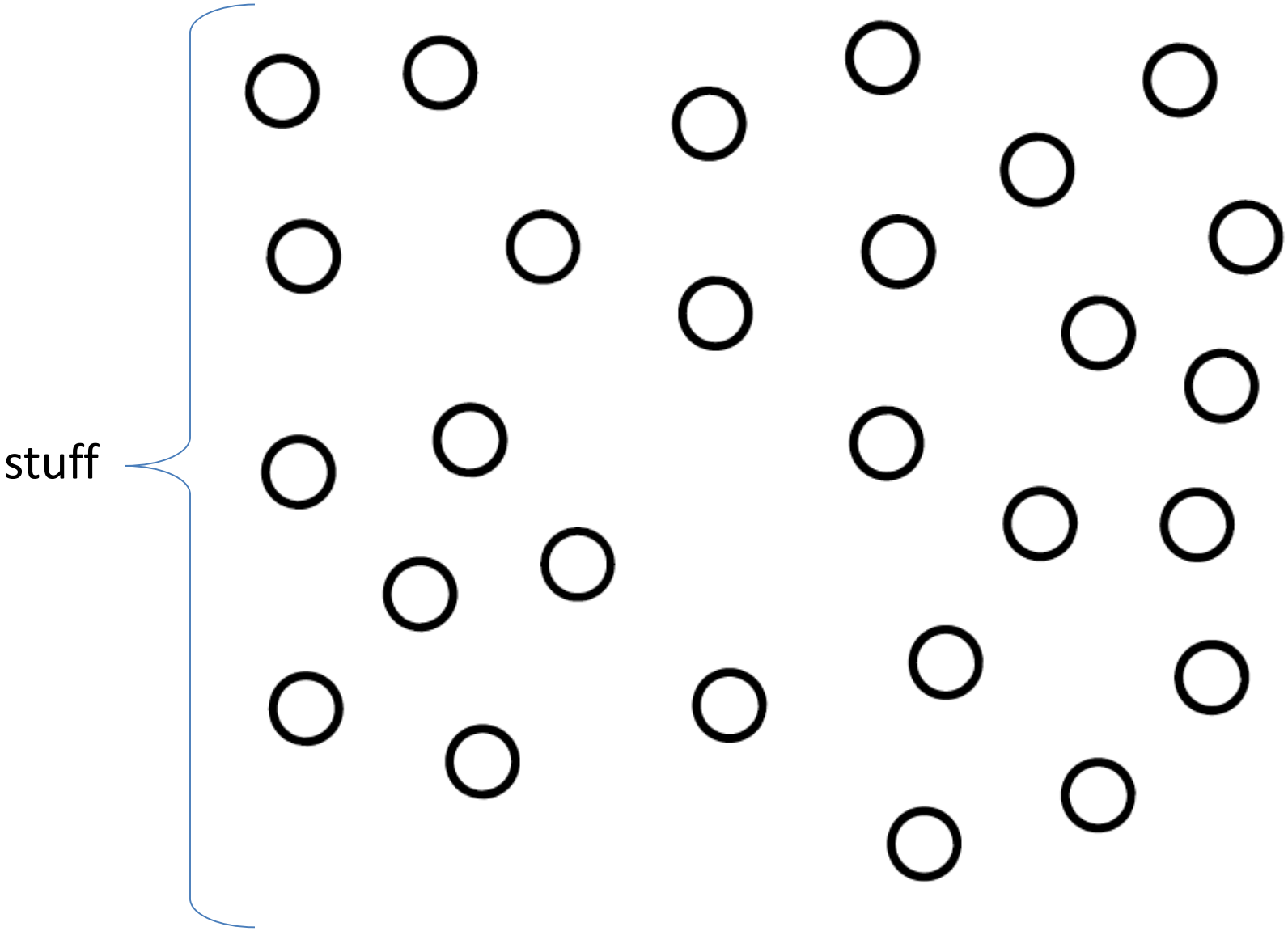
Gene-gene correlation matrix



What are the true clusters of correlated genes?

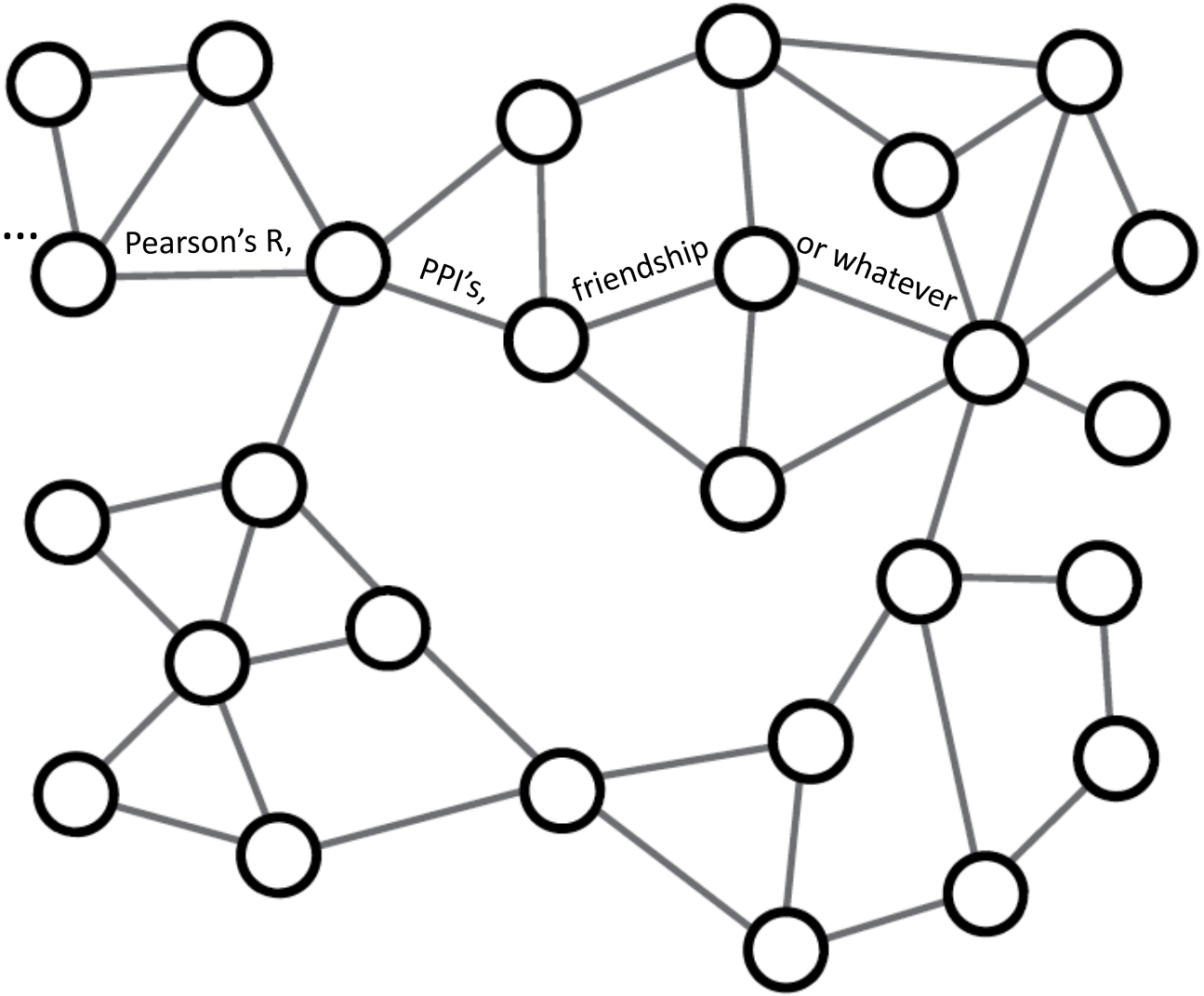
- Essential for robust analysis

SpeakEasy is an objectively better way to identify clusters



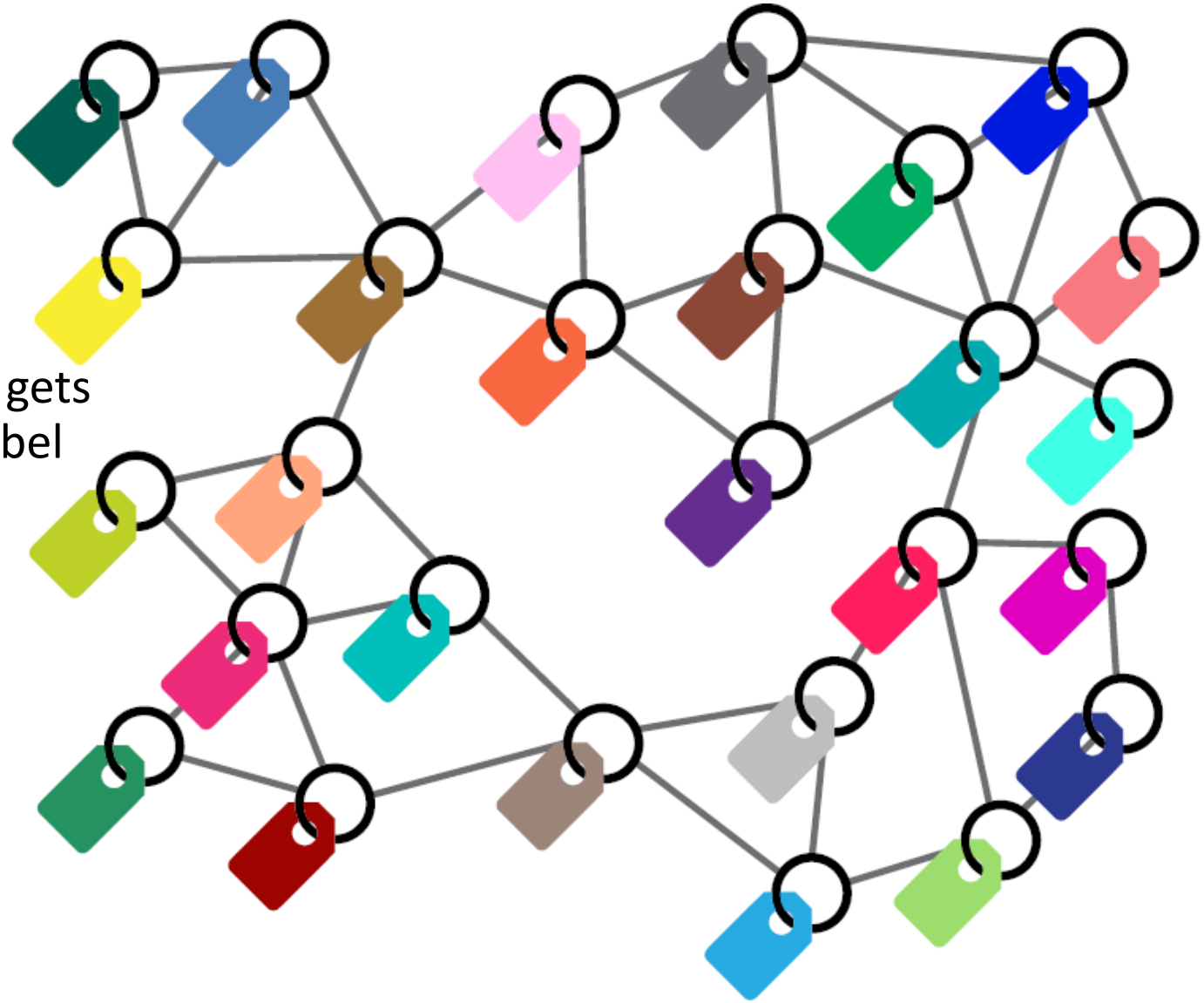
Example network - links represent object closeness

Links could be…

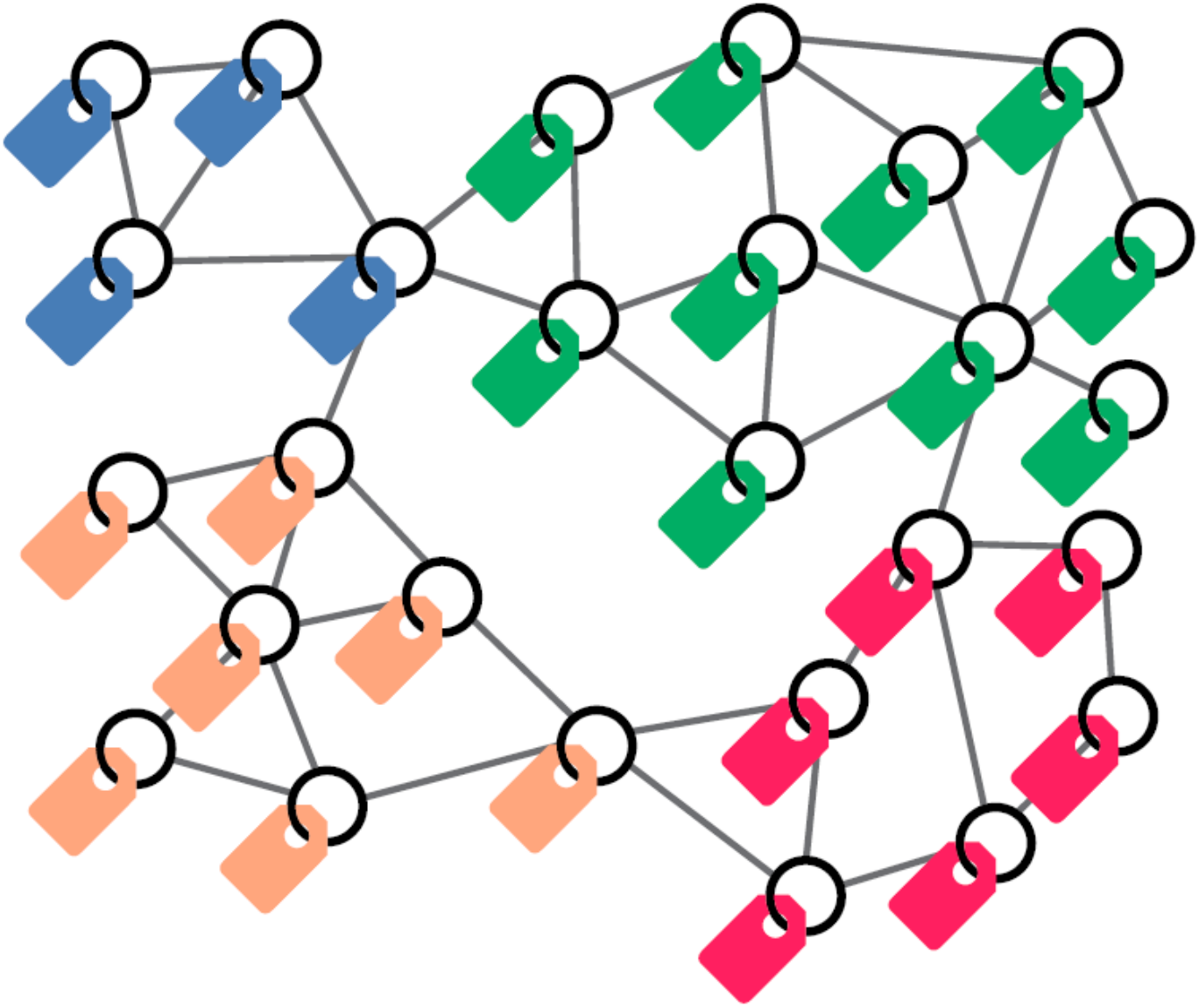


Initially we label objects randomly

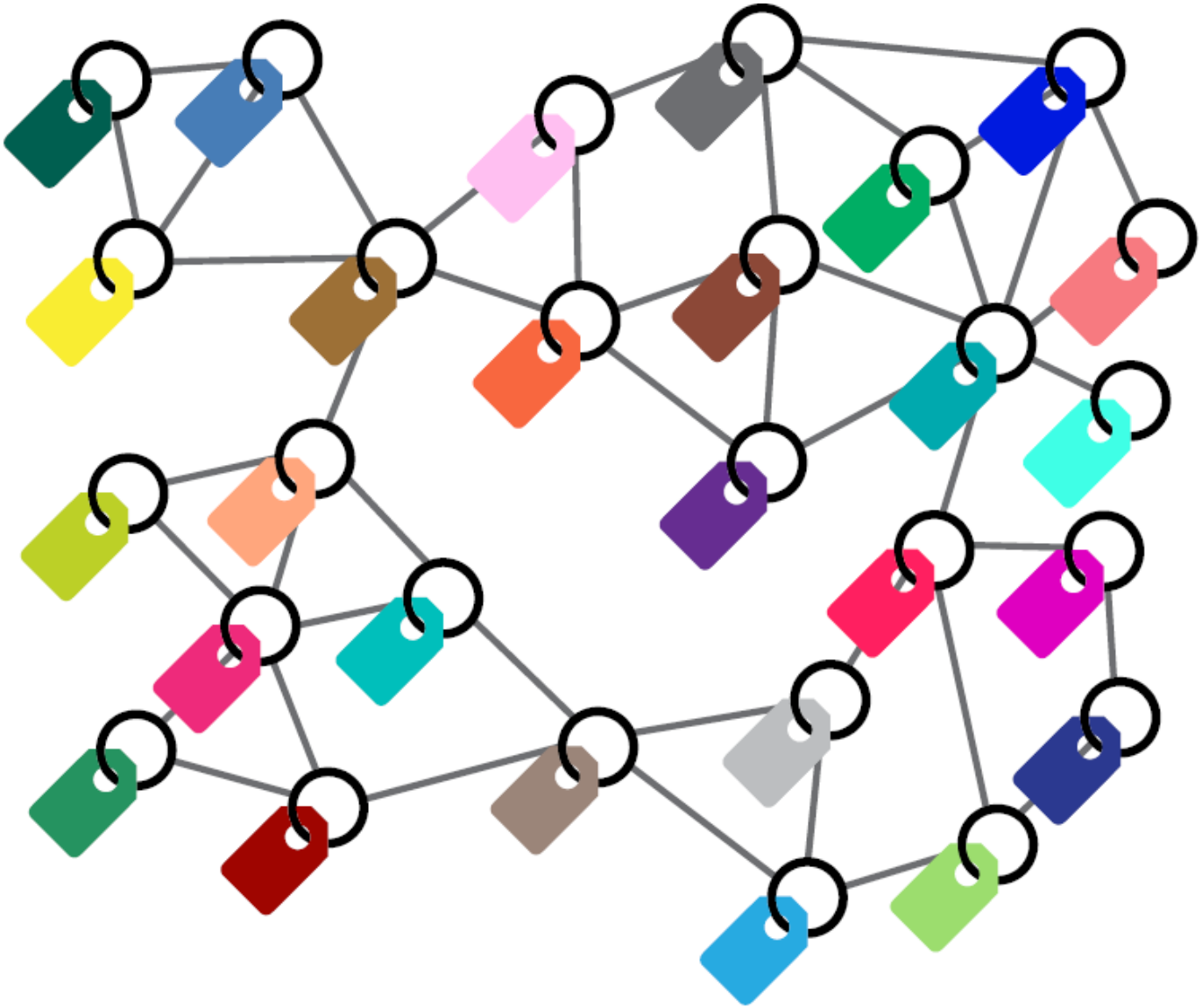
Each node gets unique label



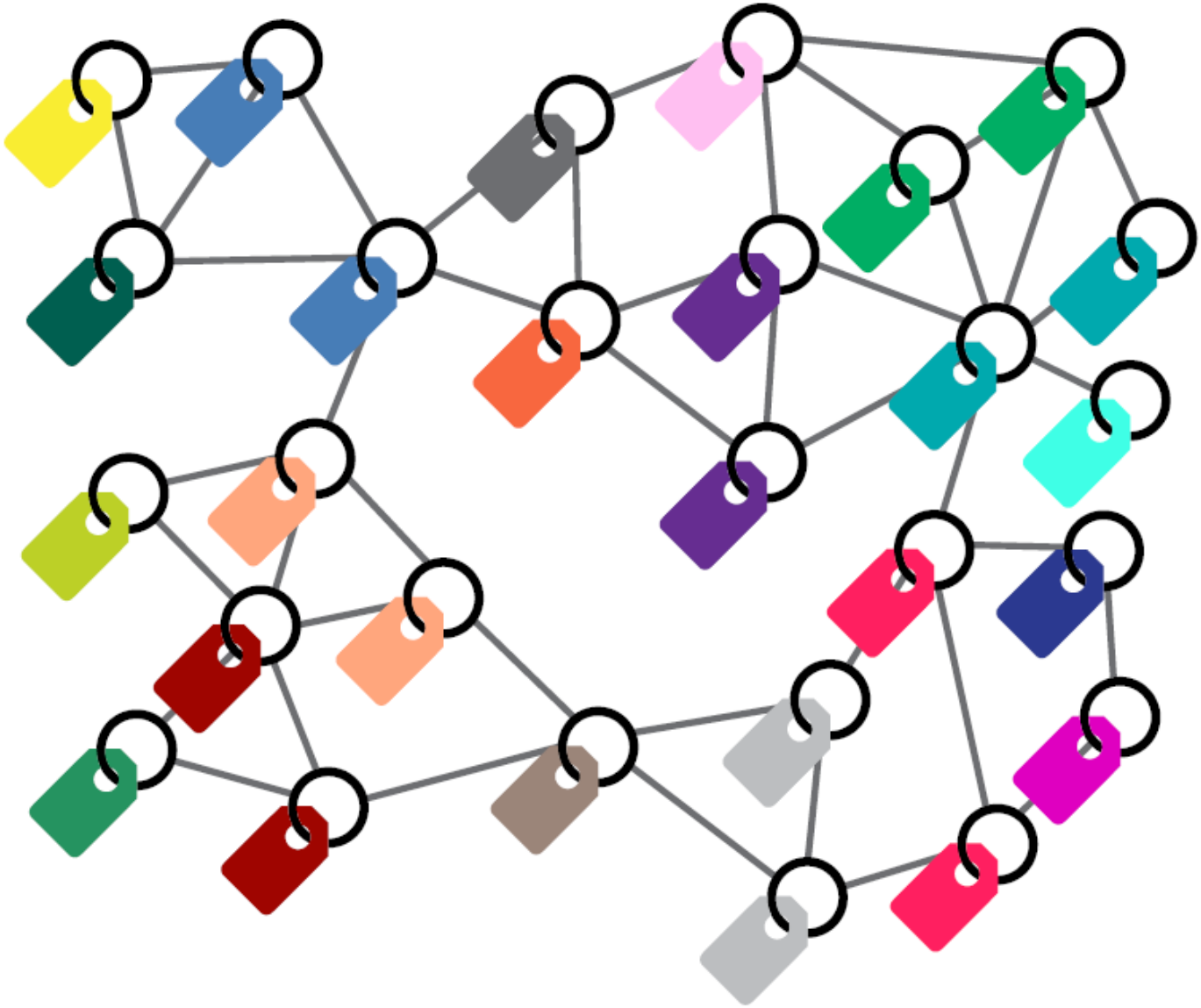
Ideally, labels will spread along links
to indicate communities



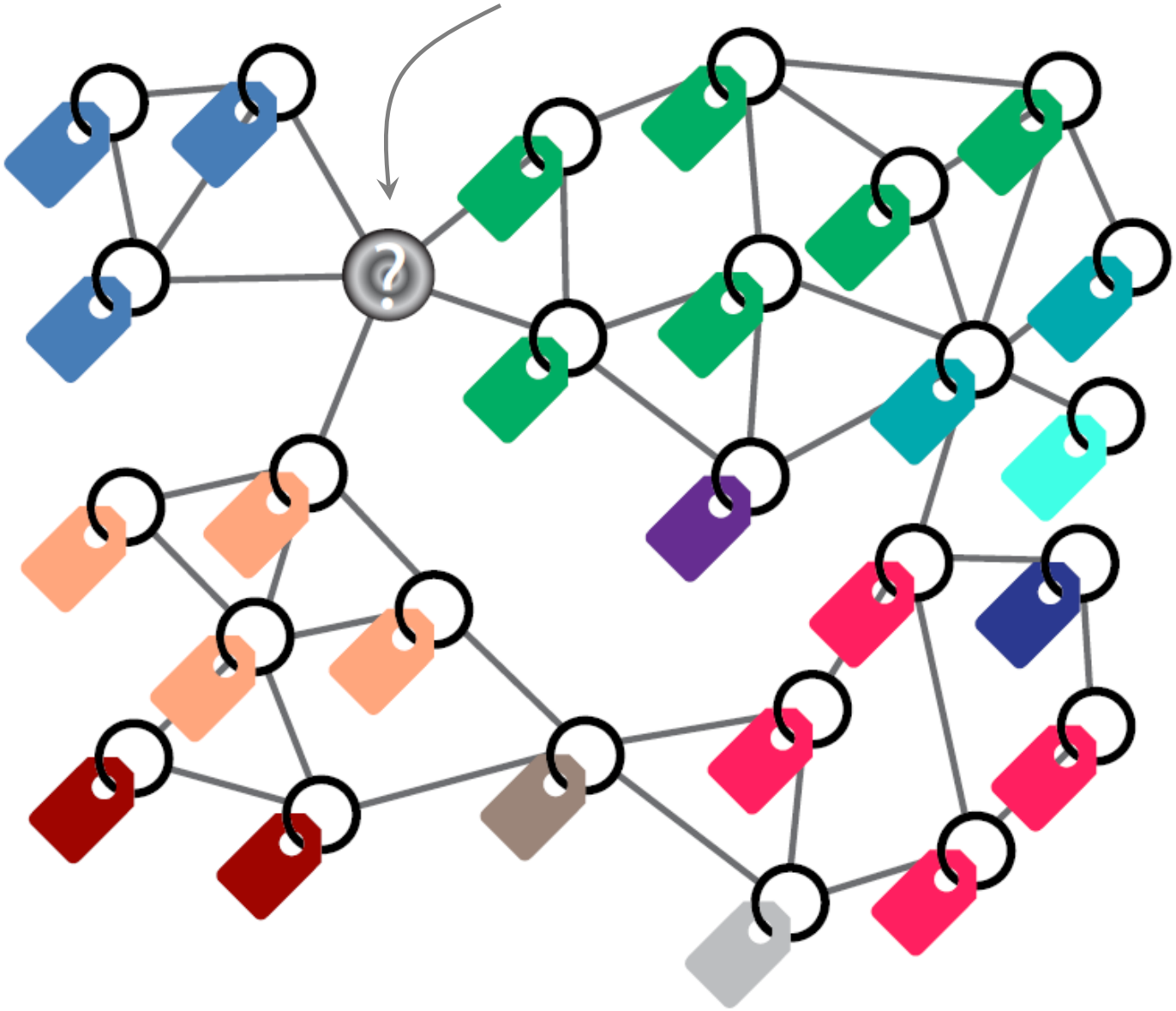
Therefore, starting from random initial labels….



We allow nodes to adopt labels
they hear frequently from their neighbors (peer pressure)

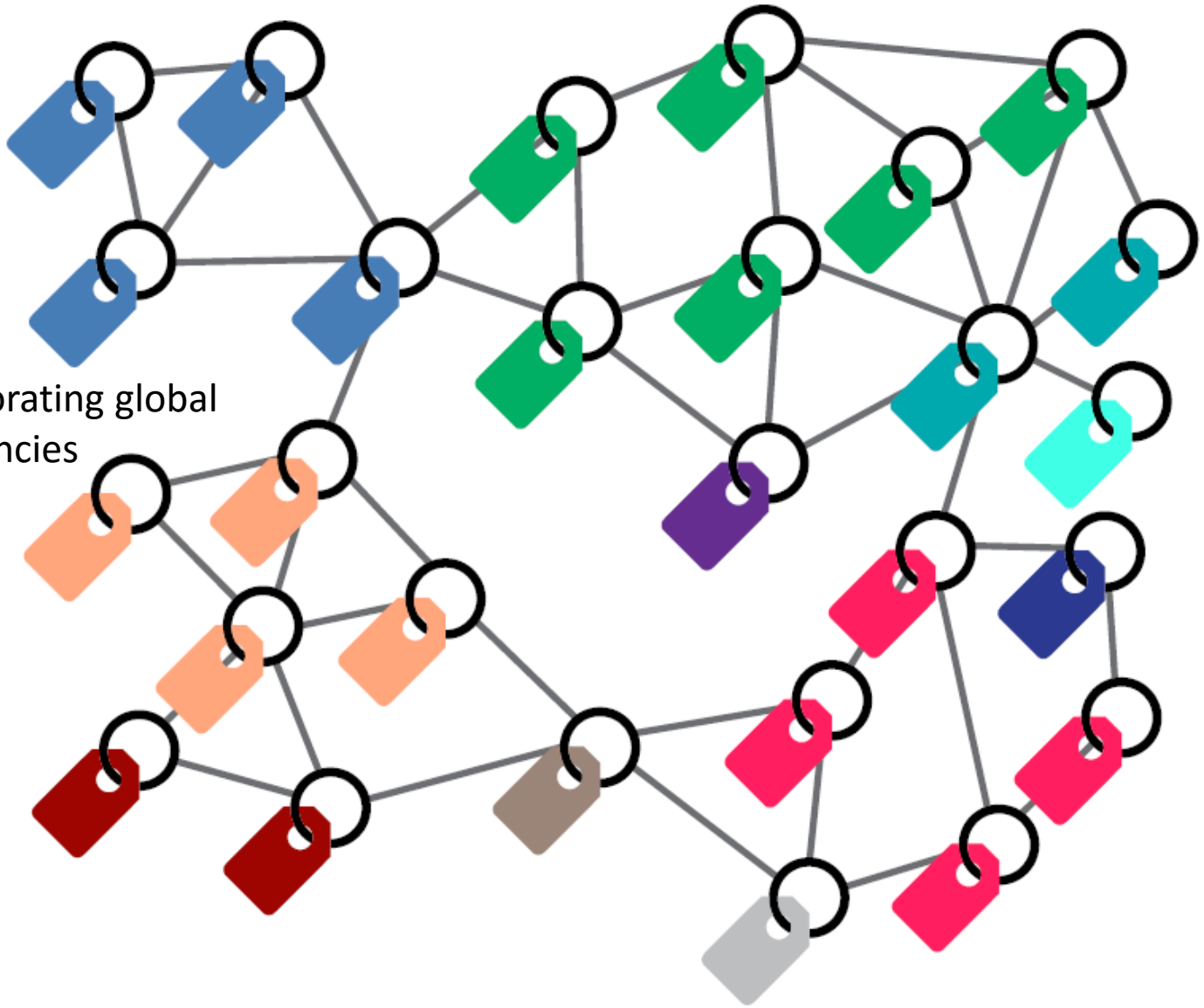


Mid-way through the process…
What will this node choose for a label?

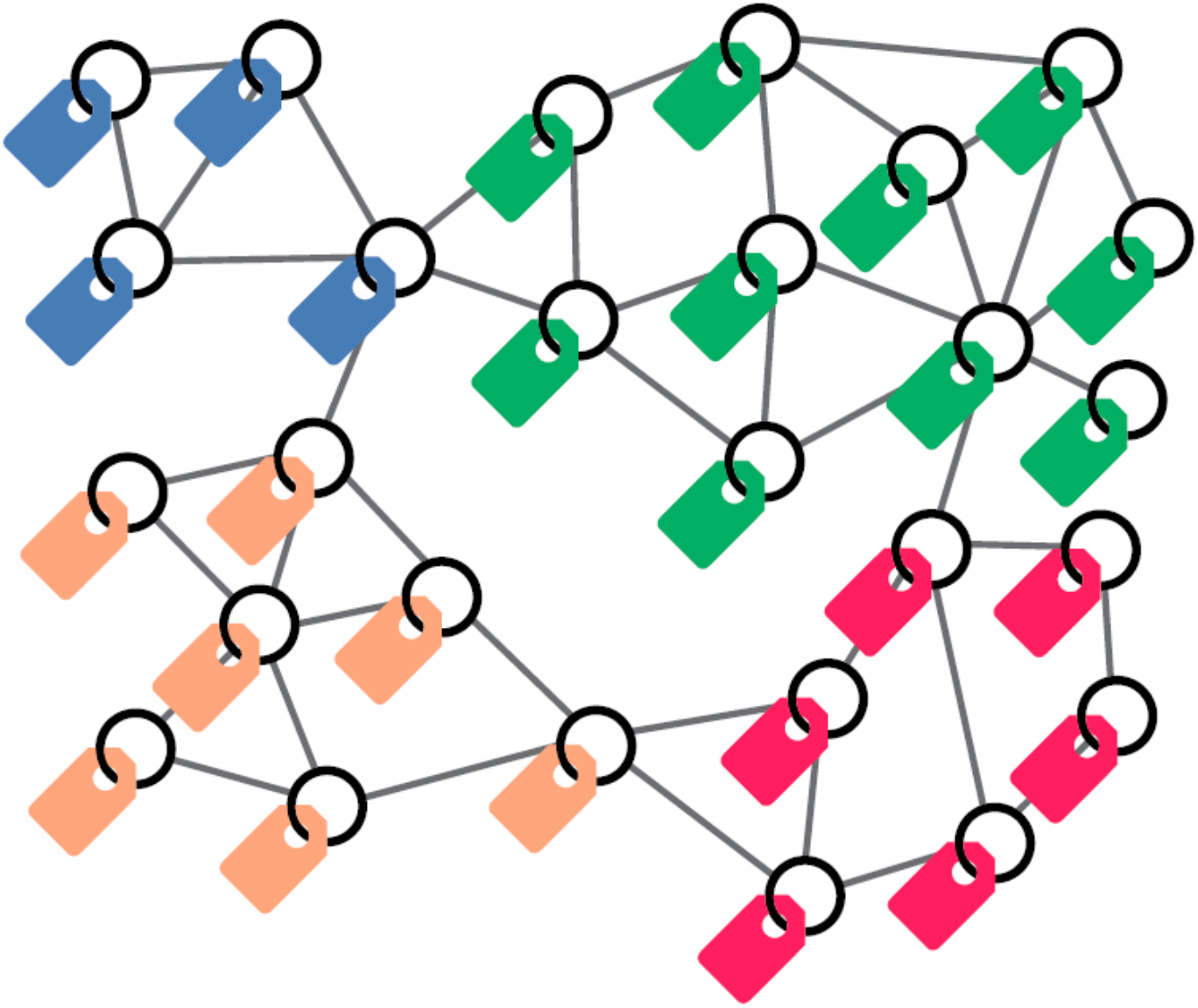


Selects the label most specific to its neighbors

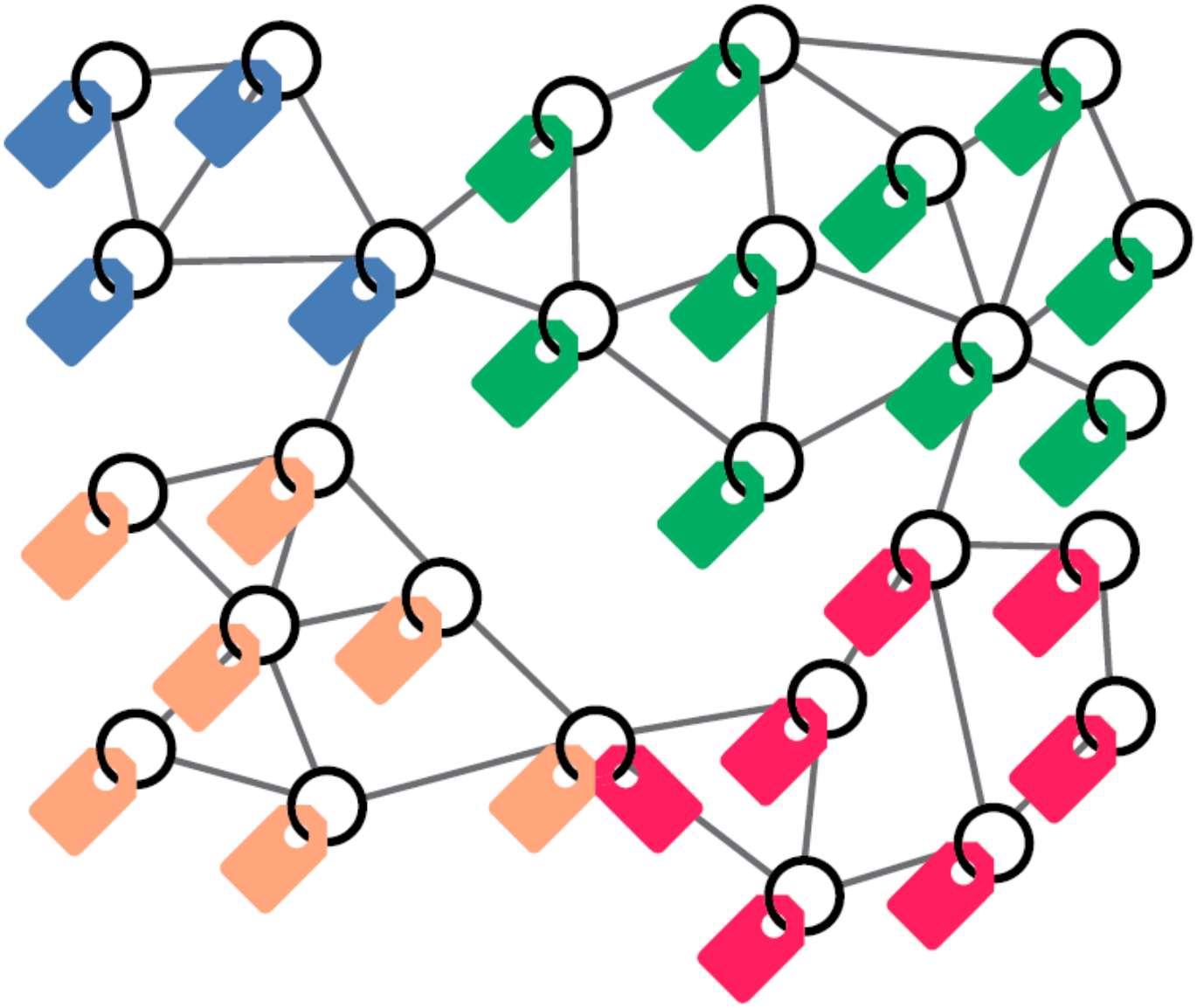
… by incorporating global label frequencies



Ultimately…
communities are identified as nodes bearing the same label



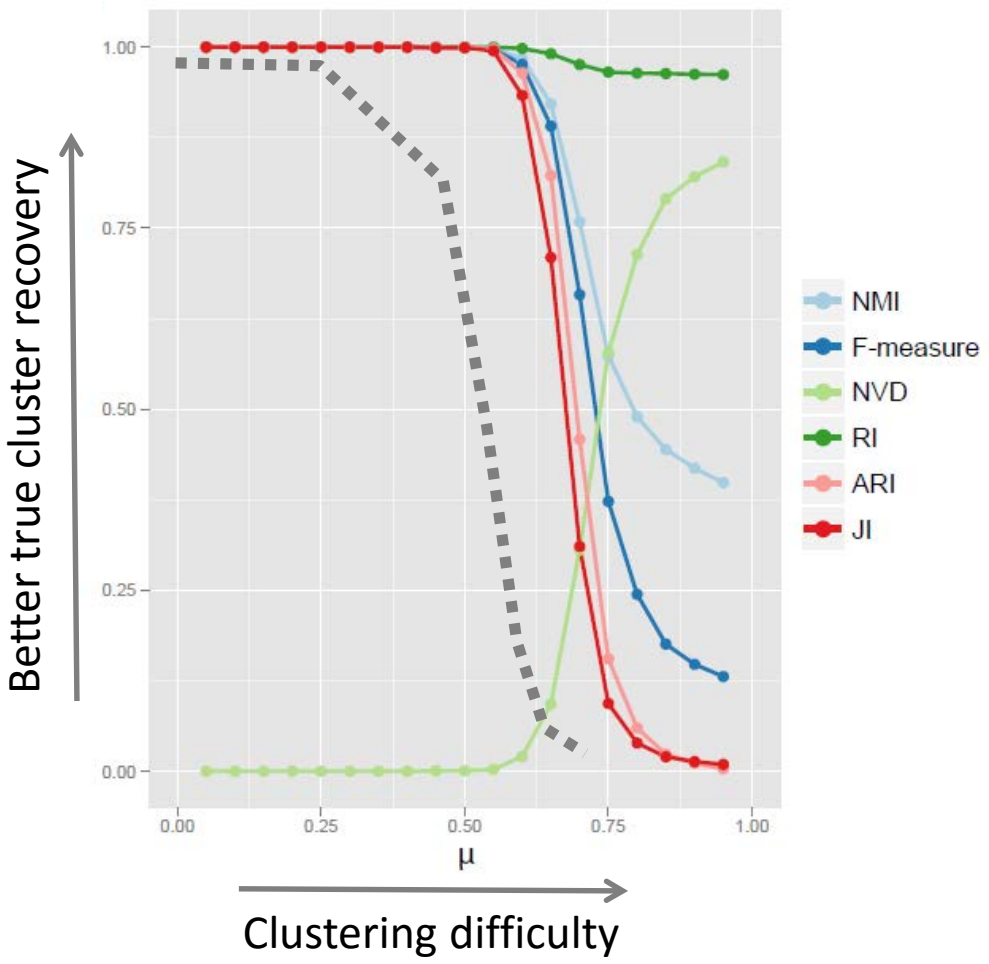
Nodes that often labeled by different communities are defined as multi-community nodes



Performance on LFR clustering benchmarks

We test the quality of clusters by trying to detect clusters in networks where the true solution is known (LFR benchmarks).

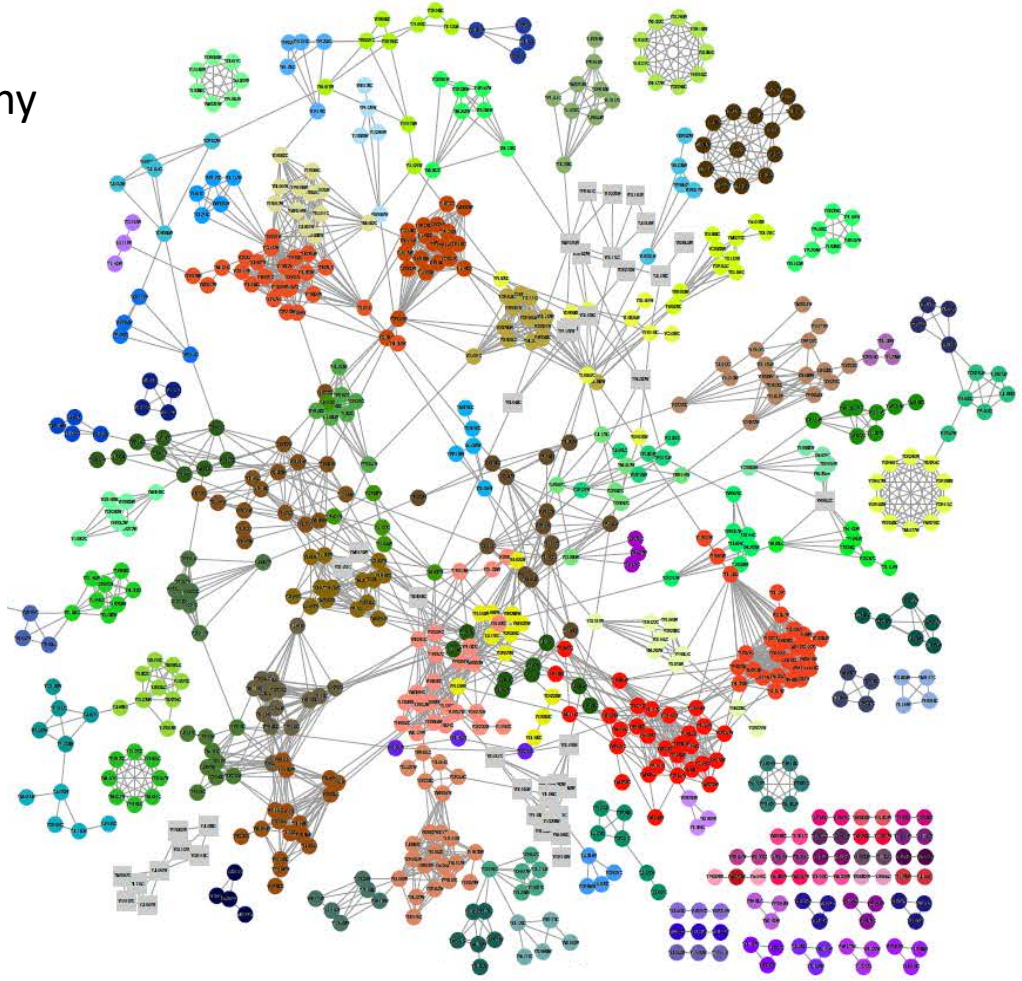
Dashed line shows typical NMI response of most other clustering algorithm



Performance on LFR clustering benchmarks

We apply SpeakEasy to clustering many types of networks, and show that it separates clusters more cleanly (high modularity) than any other method.

We applied SpeakEasy to networks of proteins (right), Amazon.com purchases, political blogs, dolphin interactions, *c.elegans* neurons, co-authorship, email, and musical collaborations. In every case it produced the highest modularity scores reported for these networks.



Summarizing RNASeq into molecular systems

24 of 47 (~50%) modules are enriched for biological function (0.05 FDR)

module id	# genes	# enriched functions	Representative function
9	243	112	Regulation of transcription
10	138	10	RNA processing
14	347	93	Mitochondrial part/function
16	352	137	Neuronal/Synapse part
23	251	18	Neuronal/Synapse part
106	489	19	Mitochondrial part/function
107	416	72	Membrane proteins/Neuronal System
109	390	71	Cell cycle damage response/Insulin signaling pathway/Proteasome
110	348	6	Cytoskeleton/protein motor (astrocytes)
111	244	70	Transcription
112	64	59	Cell membrane/Signaling peptide
113	313	307	Metabolism of protein
114	276	17	Immune response (NFKB pathway)
115	232	503	Immune response (IFN response)
116	224	432	Immune response (microglia)
117	409	134	Protein folding/unfolded protein response
118	405	140	Transcription/Protein metabolism/Immune
119	317	58	Transcription
121	403	42	Acetylation/Nucleic binding
123	317	153	Mitochondrial function
126	356	243	Mitochondrial function
187	30	10	Synaptic transmission

ROSMAP brain imaging

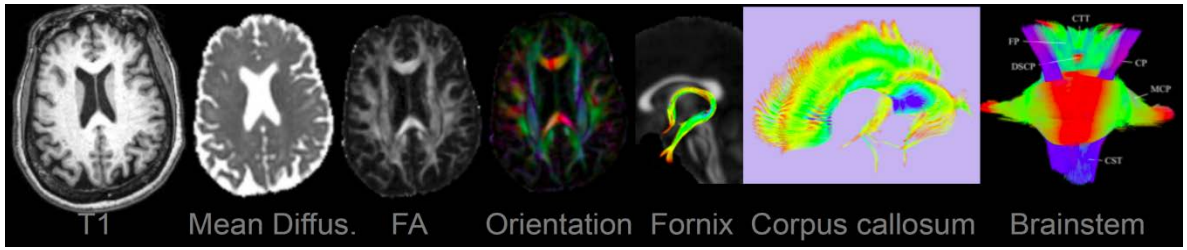
Diversity of traits recorded in ROSMAP allows us to find molecular systems associated with components of pathology that are generally not measured simultaneously with omics.

Ante-mortem imaging

1) T1-weighted MPRAGE: Macro-structure
2) SE-EPI-DTI: Micro-structure
3) Dual-echo GRE: Distortion correction
4) T2-weighted FLAIR: Lesions
5) BOLD EPI: Resting-state functional connectivity
6) Multi-echo FSE: T2-mapping, Macro-structure
7) SWI: Microbleeds

Post-mortem imaging

1) T1-weighted MPRAGE: Macro-structure
2) DTI / HARDI: Micro-structure
3) Multi-echo FSE: T2-mapping, Macrostructure
4) T2-weighted FLAIR: Lesions
5) SWI: Microbleeds



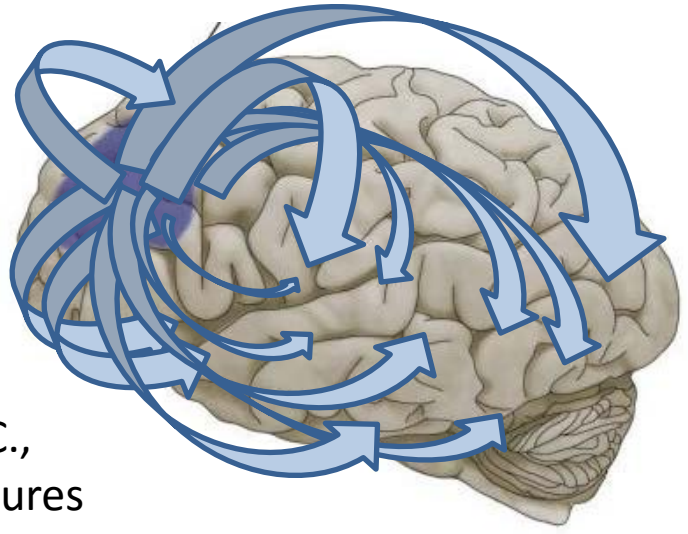
1200+ antemortem scans, 50+ counting postmortem scans. Morphological and functional features from these scans (which have already been extracted) can be linked to molecular systems and key genes predicted to drive them.

Summary of approach to “imaging-expression”

Using the ROSMAP cohort…

1. We summarize gene expression into molecular systems
2. Then we relate the activity of molecular systems to brain regions

Molecular system *A*



Molecular systems are measured in DLPFC.,
But we map them onto global brain structures

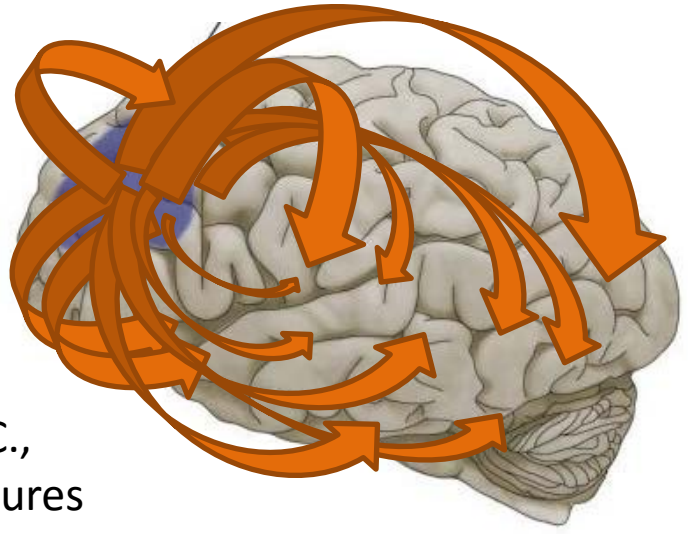
Summary of approach to “imaging-expression”

Using the ROSMAP cohort…

1. We summarize gene expression into molecular systems
2. Then we relate the activity of molecular systems to brain regions
3. Repeat for each molecular system

Molecular system *A*

Molecular system *B*



Molecular systems are measured in DLPFC.,
But we map them onto global brain structures

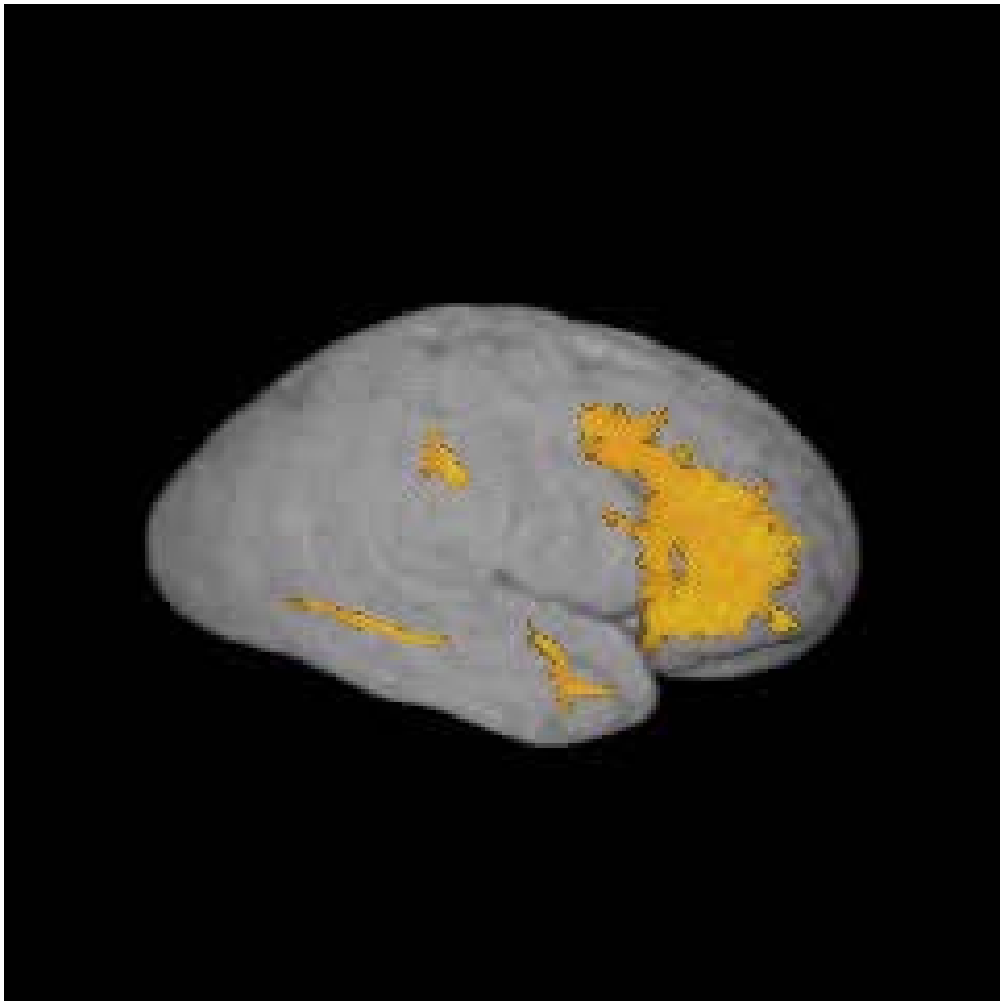
Example of expression-imaging in relation to AD

A coexpressed geneset, code named module 109, has the strongest correlations with AD path and cognitive decline.

It contains gene sets related to cell cycle and transcriptional regulation, but the overall function is not as clear as other some other coexpressed sets.

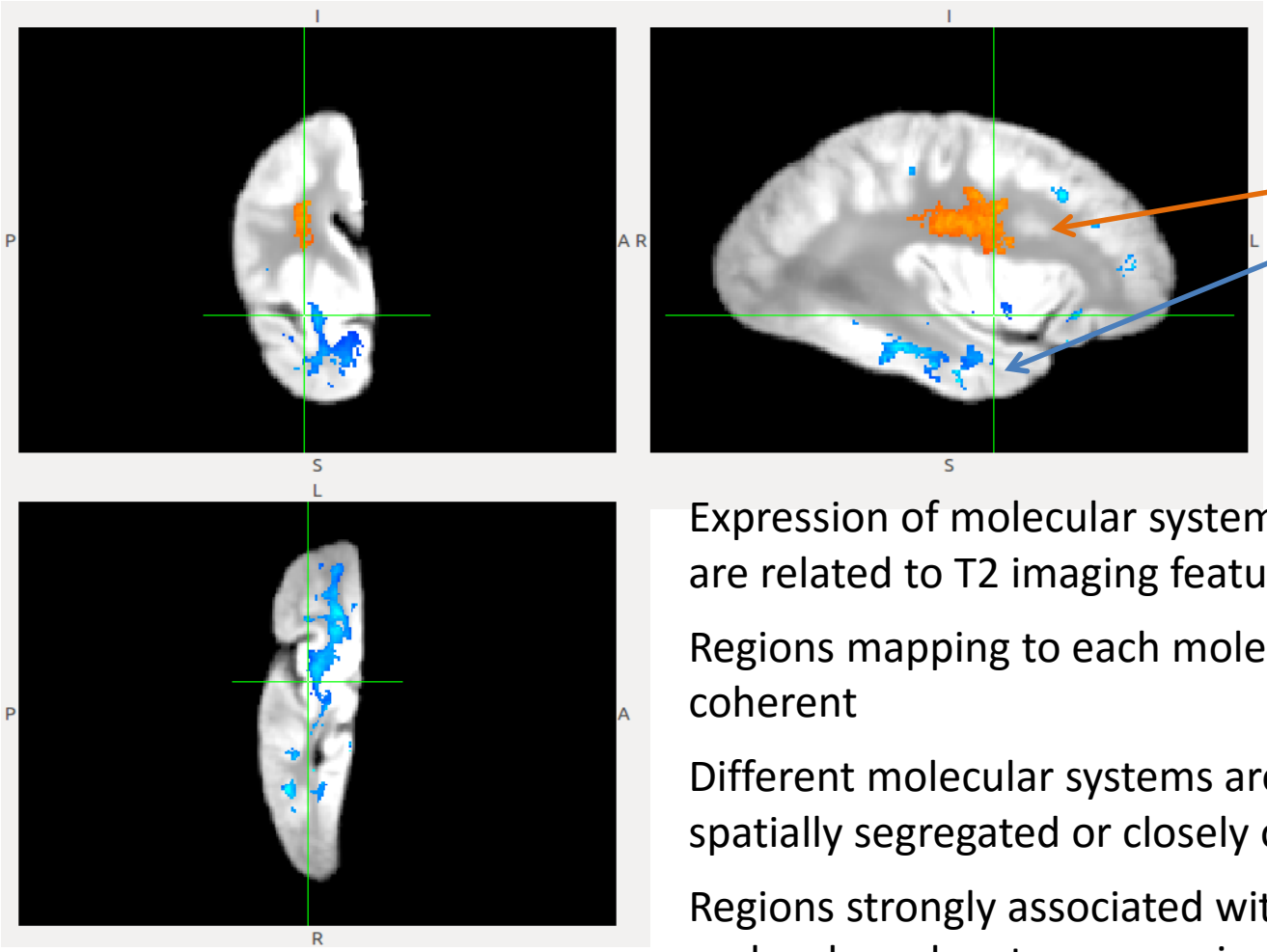
109 has correlations with TBM region-sets, and also the most extensive (spatial) correlations with T2 imaging of any molecular system.

Currently testing key genes in the set, identified by novel causal inference methods.



Regions correlated (p<.01) with m109 expression

Mapping coexpressed communities to T2 images



Ion channels
Inflammation
Lysosomes
Neurogenesis

Expression of molecular systems (measured in DLPFC) are related to T2 imaging features across the brain

Regions mapping to each molecular system are spatially coherent

Different molecular systems are linked to either highly spatially segregated or closely overlapping regions

Regions strongly associated with one or more molecular subsystems are: cingulum, temporal WM, hippocampus, internal and external capsule, insula, various portions of frontal WM

Thank you

Disease Neurobiology
- David Bennett, MD
- Sara Mostafavi, PhD – University of British Columbia

Neuroimaging
- Robert Dawe, PhD – Rush University
- Konstantinos Arfanakis, PhD – Rush University

Methods Development
- Shinya Tasaki, PhD – Takeda
- Boleslaw Szymanski, PhD – Rensselaer Polytechnic
- Miles Chen – Rensselaer Polytechnic

Grant and Data Support
- The participants of the Religious Orders Studies and the Memory and Aging Project
- Grant support: RF1AG015819, R01AG017917, R01AG036042, U01AG046152, R01AG036836

Extra slides

Current Questions

Does expression from different brain regions have more extensive or different imaging correlations?

To what extent are these correlations disease-based, cell-type based or reveal some other form of physiological “state”?

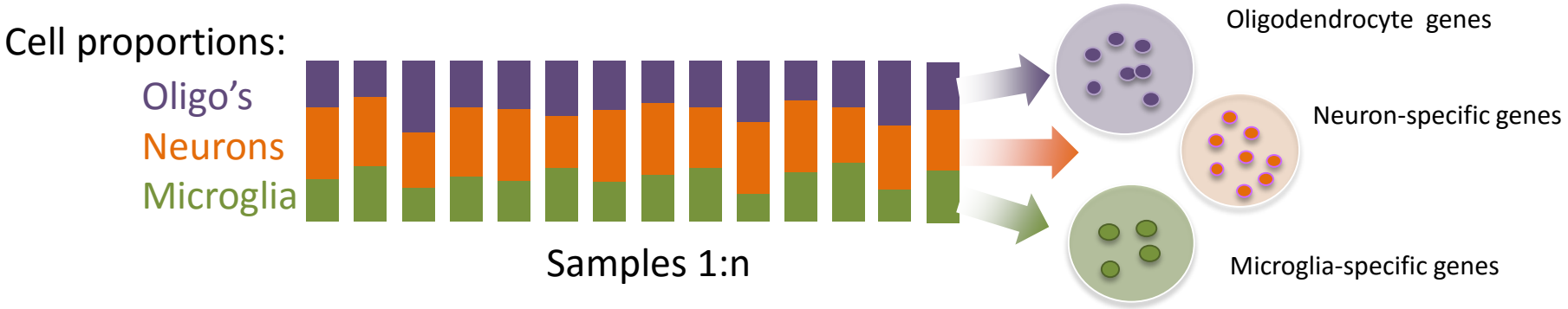
How do gene expression patterns relate to connectivity?

How are expression-correlated regions affected in AD?

What brain regions are the molecular systems affected in AD up- and down-regulated?

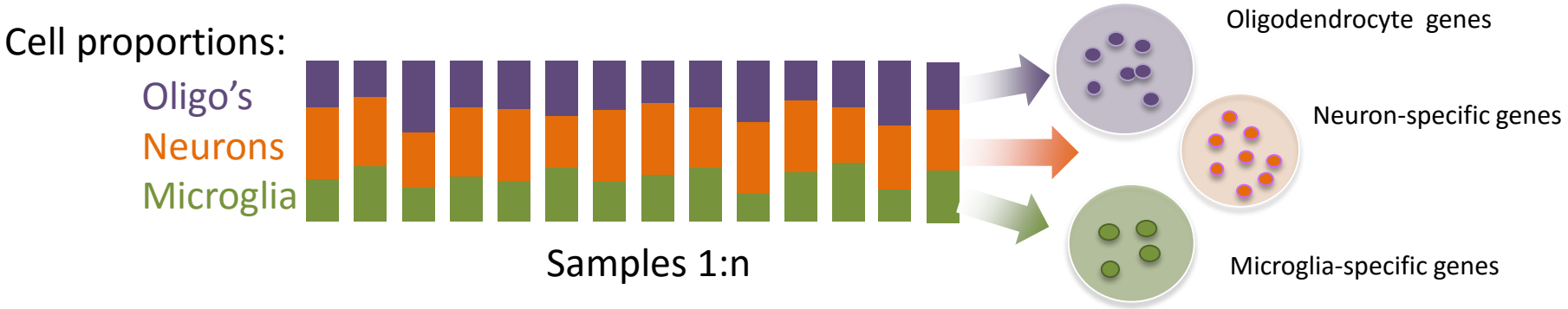
What are the mechanisms that support unified and cooperative regulation among multi-scale systems, across space and time?

What could account for expression – imaging correlations?



- Random variations in cell type proportions can create coexpressed gene sets
- Cell types have specific spatial distribution patterns

What could account for expression – imaging correlations?

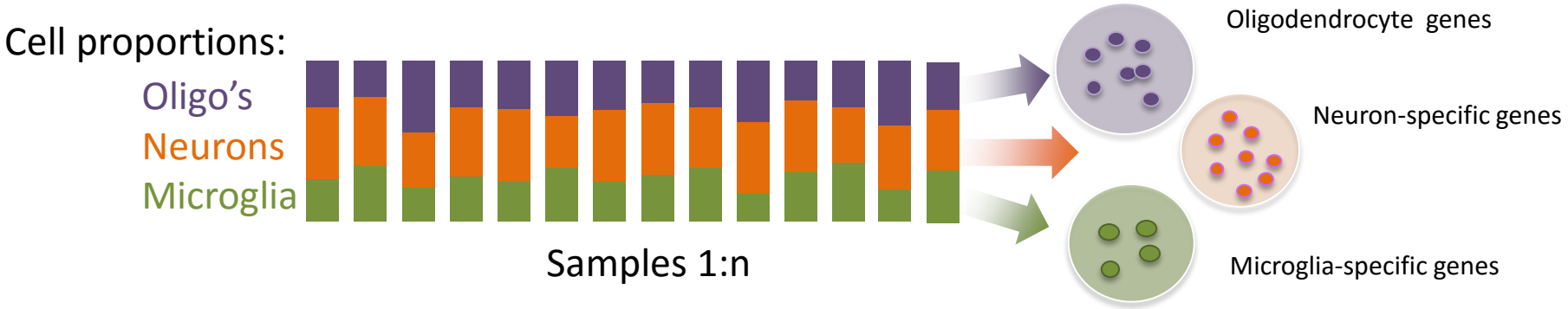


- Random variations in cell type proportions can create coexpressed gene sets
- Cell types have specific spatial distribution patterns

However

- It not obvious than multiple coexpressed sets would relate to MRI-based brain structure
- Unlike imaging-genetics, we observe strong correlations between coherent molecular systems and extensive brain regions
  - The gene sets provide an actionable basis for processes active in specific regions
  - Molecular systems correlated with AD-affected regions may be a way to explore that pathology

What could account for expression – imaging correlations?

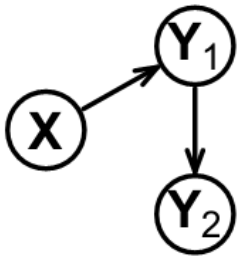


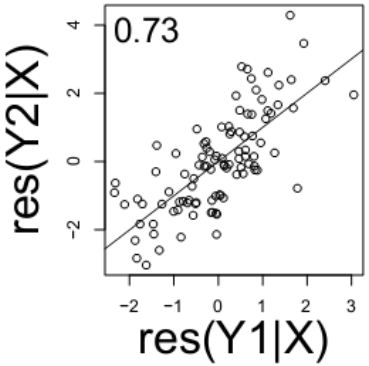
- Random variations in cell type proportions can create coexpressed gene sets
- Cell types have specific spatial distribution patterns

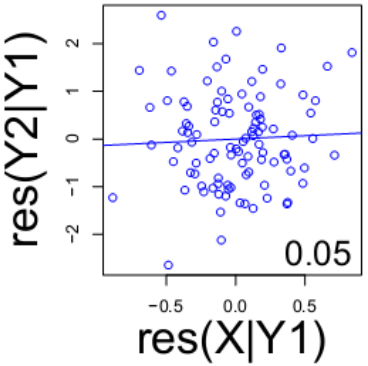
However

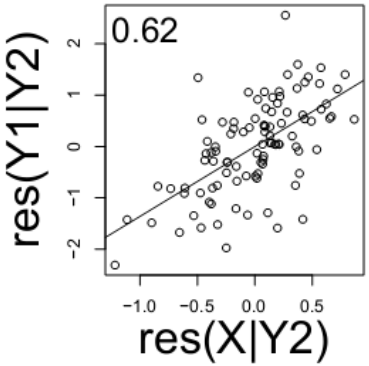
- It not obvious than multiple coexpressed sets would relate to MRI-based brain structure
- Unlike imaging-genetics, we observe strong correlations between coherent molecular systems and extensive brain regions
  - The gene sets provide an actionable basis for processes active in specific regions
  - Molecular systems correlated with AD-affected regions may be a way to explore that pathology
- Some image-correlated molecular systems appear to be subcellular
- Also possible that activity-dependent expression links genes to regions
- Also possible that cell type-specific molecular systems link genes to regions

Conditional independence as the key to causal ordering

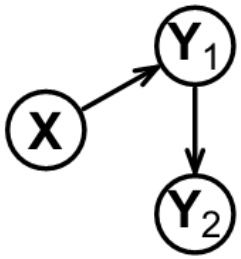


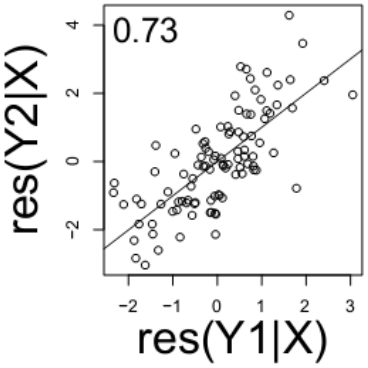


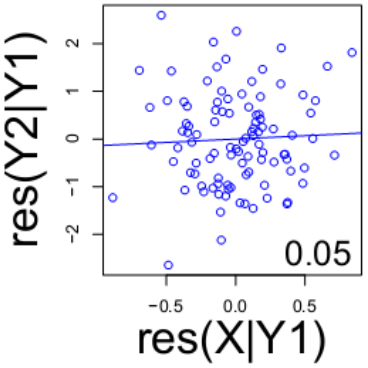


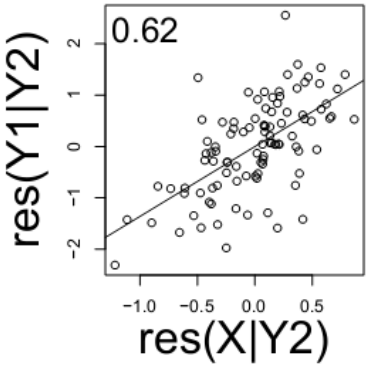


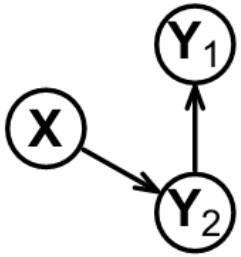
Conditional independence as the key to causal ordering

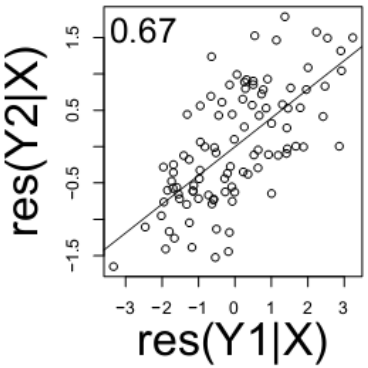


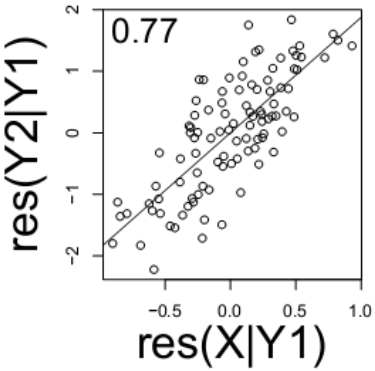


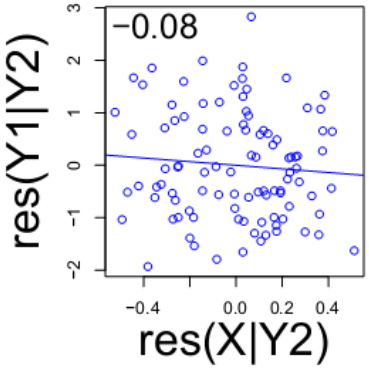




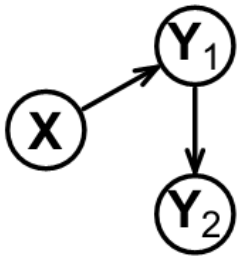


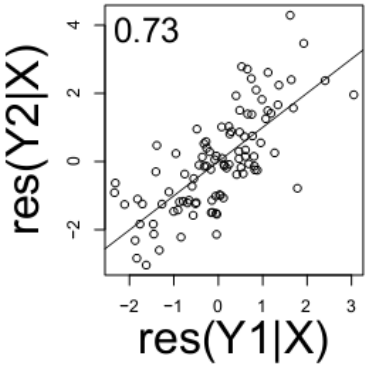


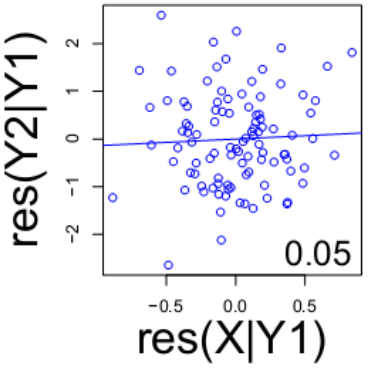


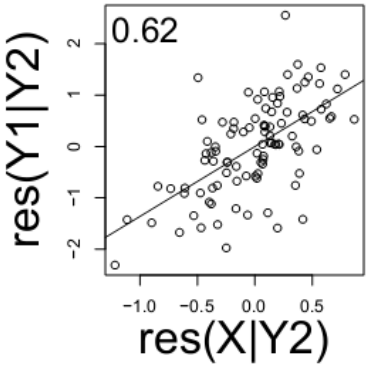


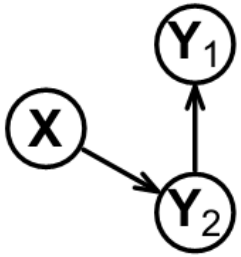
Conditional independence as the key to causal ordering

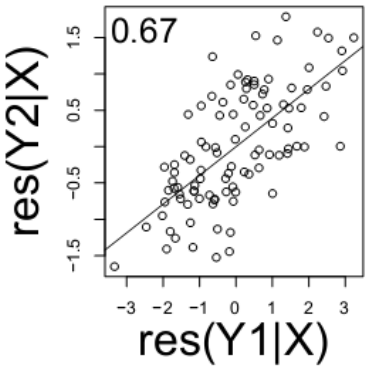


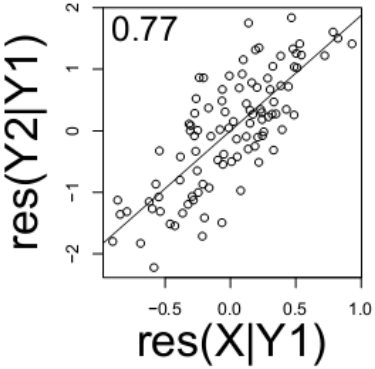


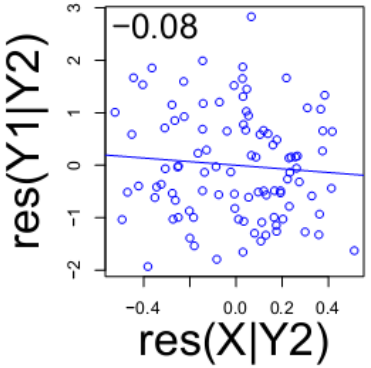


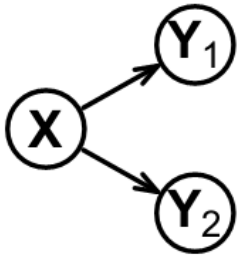


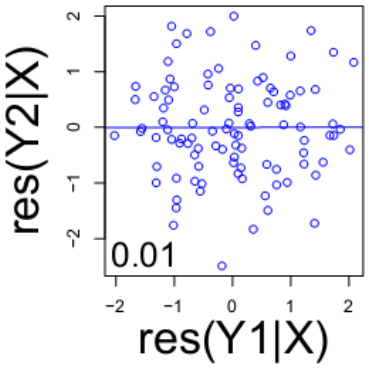


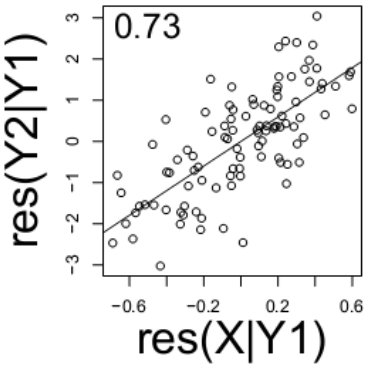


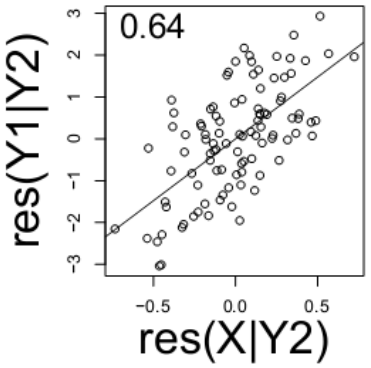












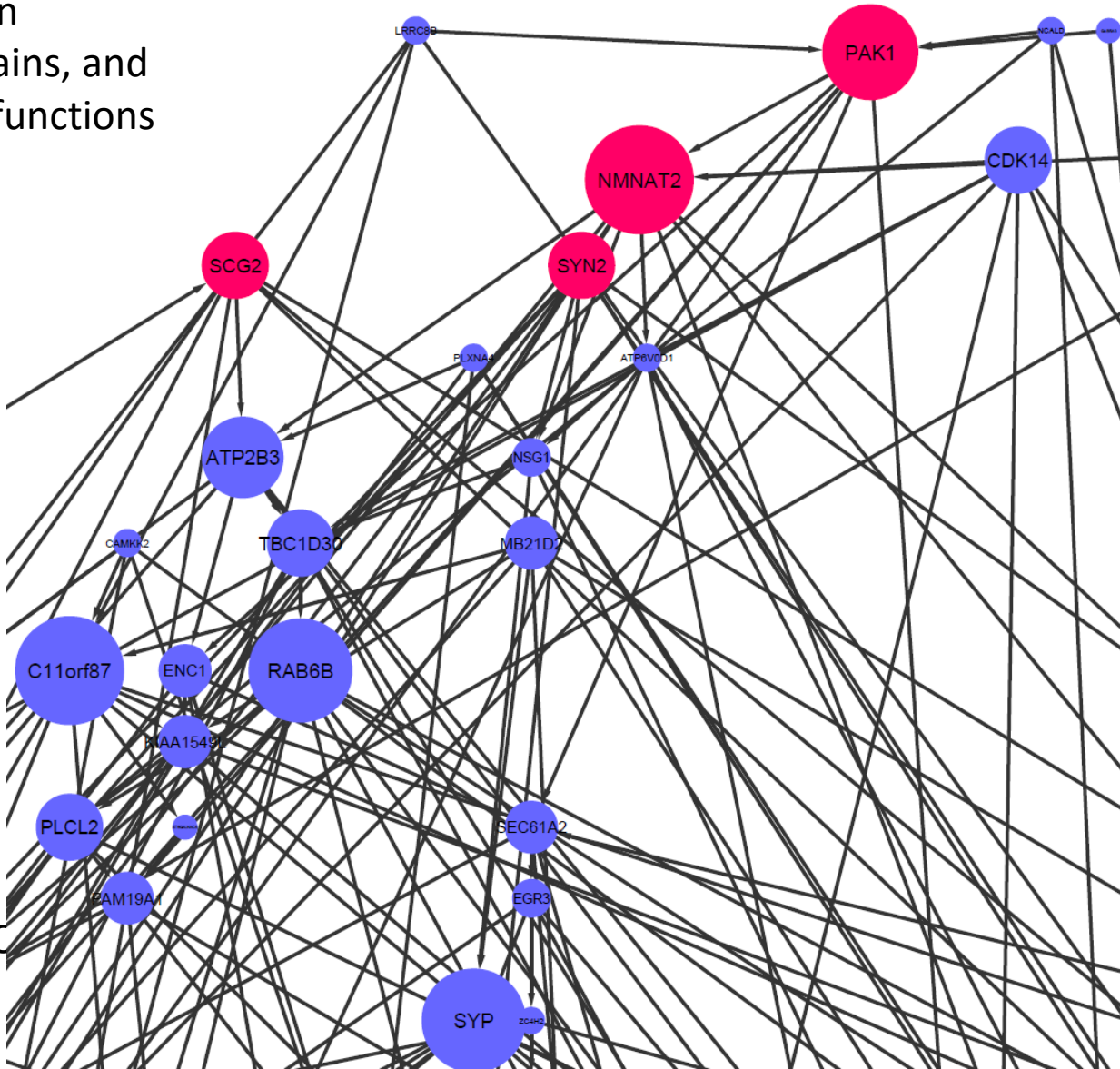
Synaptic network (zoom to key nodes)

PAK1 and the PAK family are activated by Rho GTPases, their expression in hippocampus changes in AD brains, and have synaptic and cytoskeletal functions related to memory in mice

NMNAT2 is essential for axon growth and survival and some types of repair

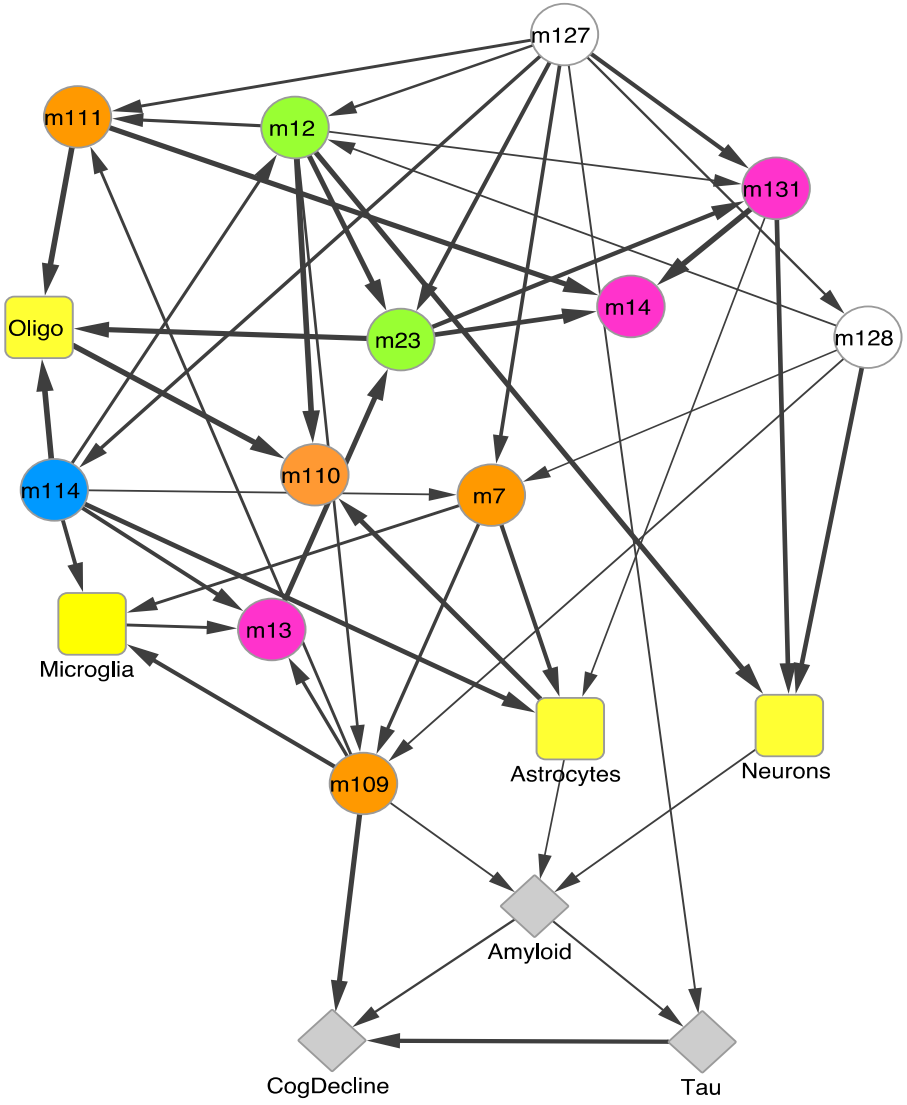
Expression and sequence changes to presynaptic gene SYN2 have been seen in SCZ and cortical expression changes are seen in AD

ERG1 induces PSEN2 expression, tau phosphorylation and activates CDK5



Bringing together the ingredients of AD:

This directed graph describes the conditional independence relationships among expression modules, traits, and cell composition



- AD traits
- Cell type markers
- Molecular systems
  - Cell cycle
  - Immune response
  - Mitochondria funct.
  - Neuronal funct.
  - NA

This network identifies molecular systems that are upstream of AD traits, and the relationships among the systems.

For instance, a specific molecular system “module 109” and astrocyte activation is predicted to affect amyloid beta levels