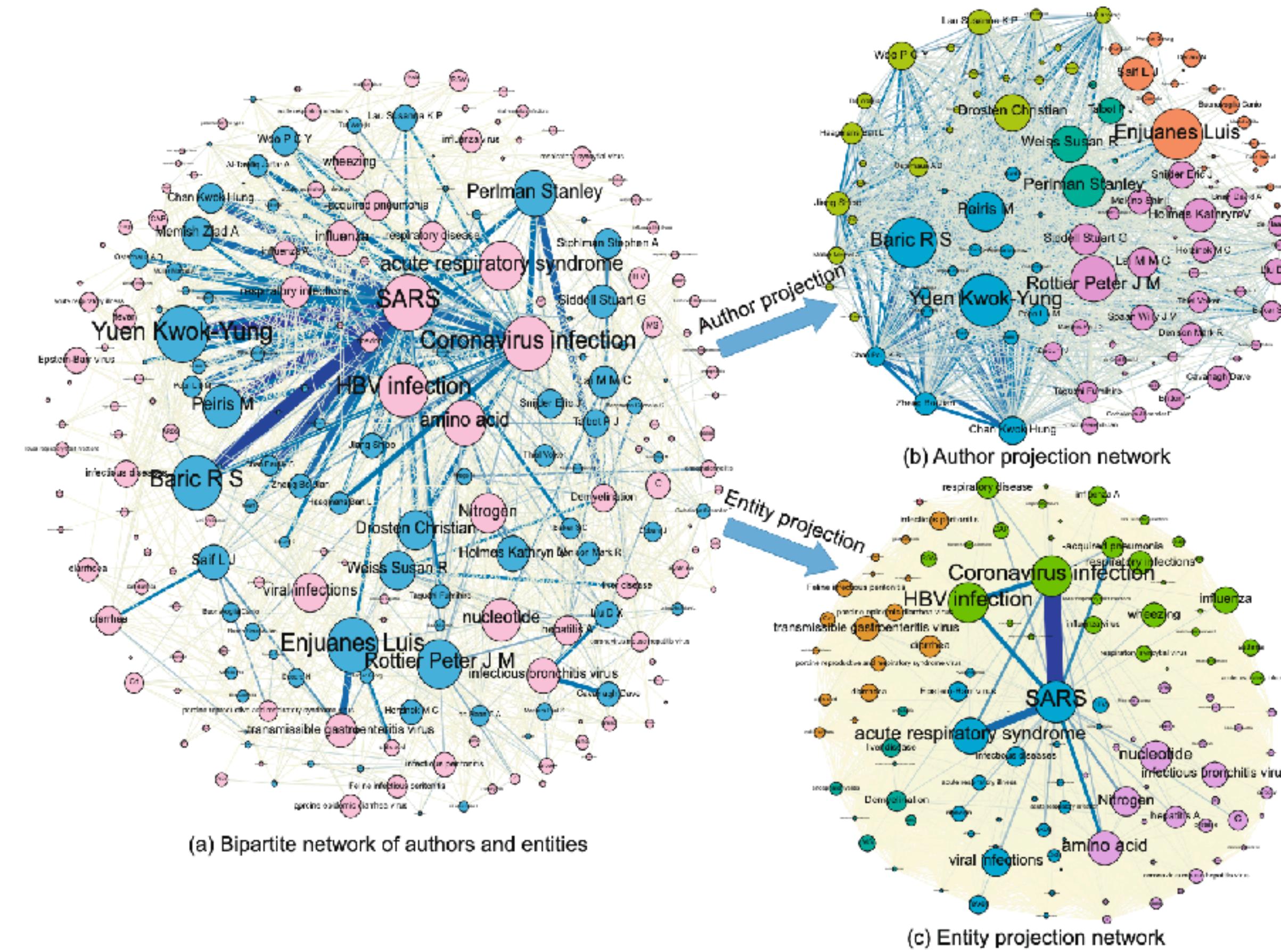
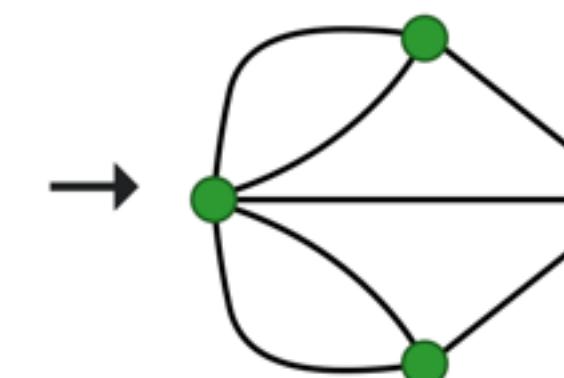
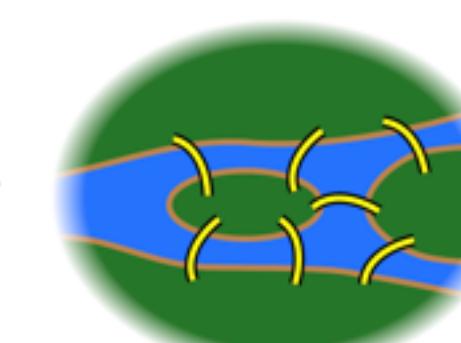
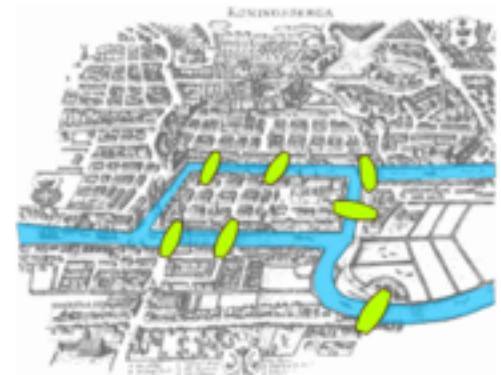


Biological Network Analysis



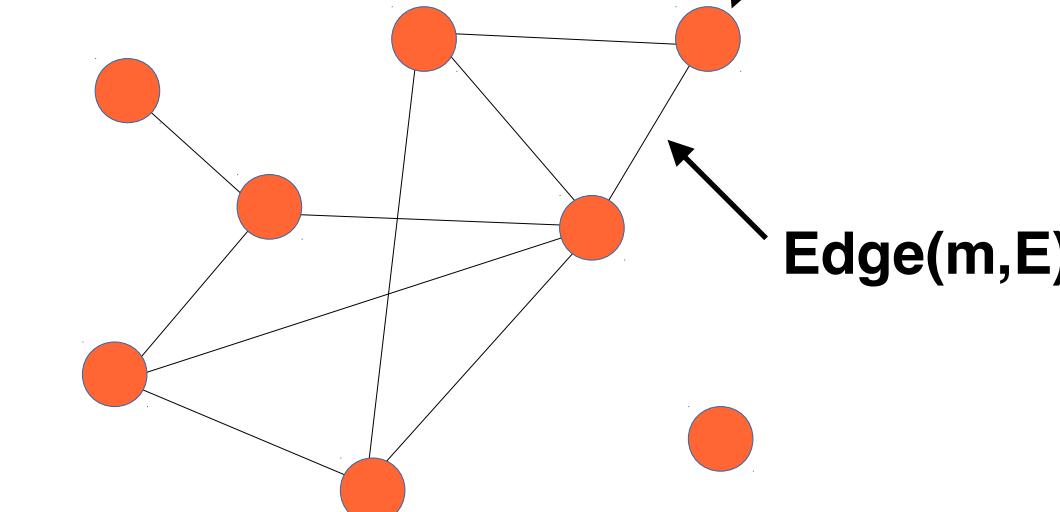
The Anatomy of a Network



The study of networks is concerned with understanding and modelling the behaviour of real-world systems.

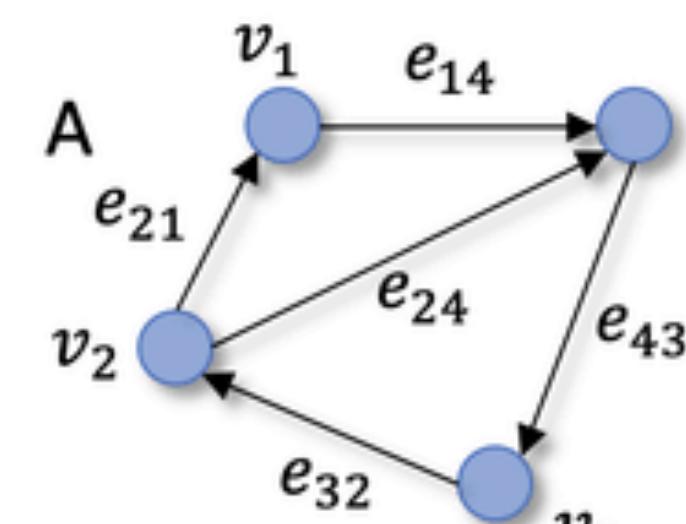
- In mathematics called Graph theory.
- First network Euler's Königsberg bridge problem (1736).

Node(n) / Vertex(v)

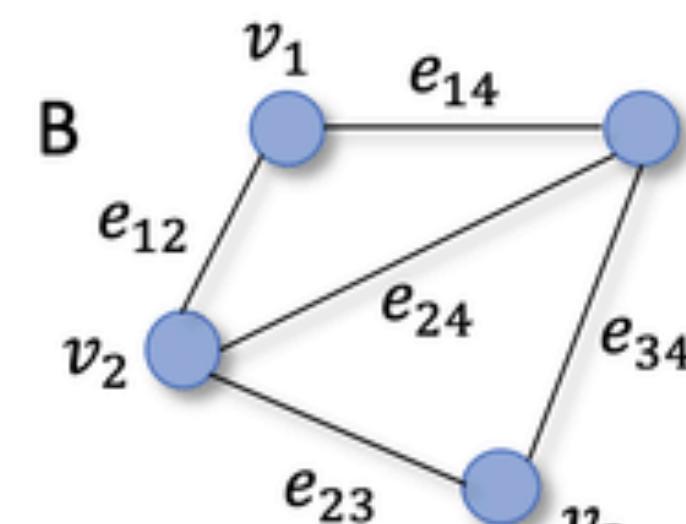


Edge(m,E)

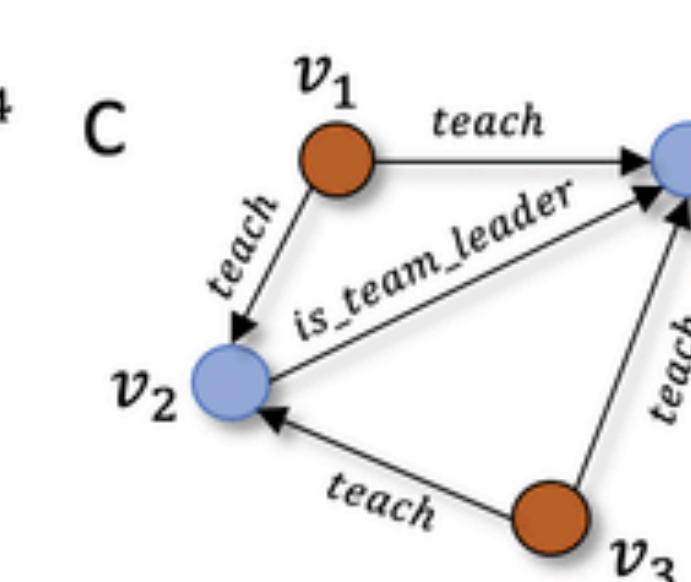
Directed graph $G(V,E)$



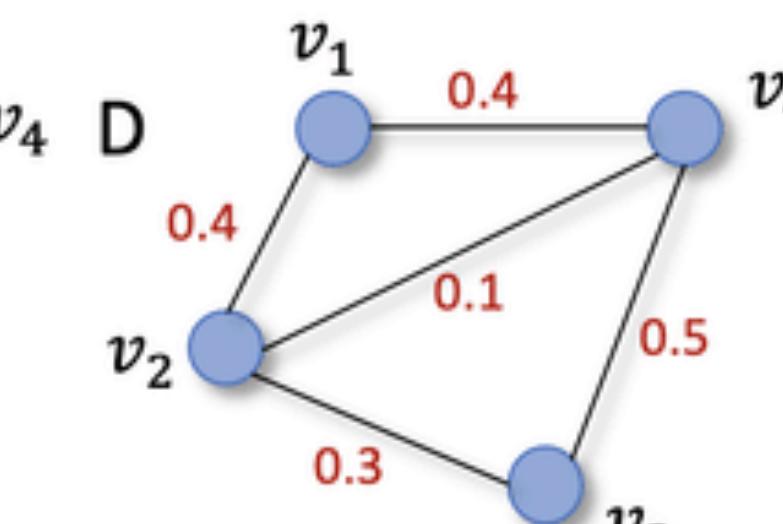
Undirected graph $G(V,E)$



Knowledge graph $G(V,E)$



Weighted graph $G(V,E)$



E

	v_1	v_2	v_3	v_4
v_1	0	0	0	1
v_2	1	0	0	1
v_3	0	1	0	0
v_4	0	0	1	0

F

	v_1	v_2	v_3	v_4
v_1	0	1	0	1
v_2	1	0	1	1
v_3	0	1	0	1
v_4	1	1	1	0

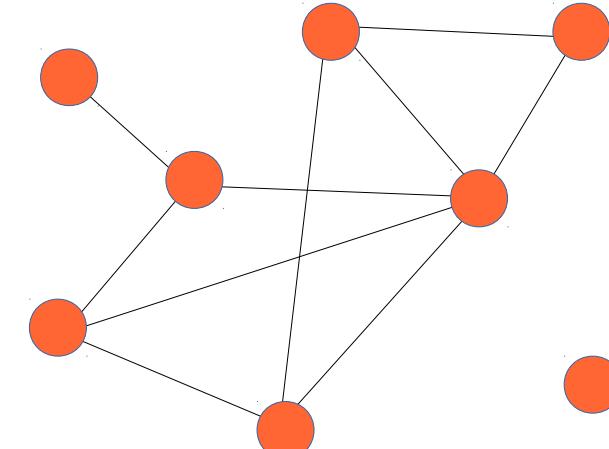
G

	v_1	v_2	v_3	v_4
v_1	0	1	0	1
v_2	0	0	0	1
v_3	0	1	0	1
v_4	0	0	0	0

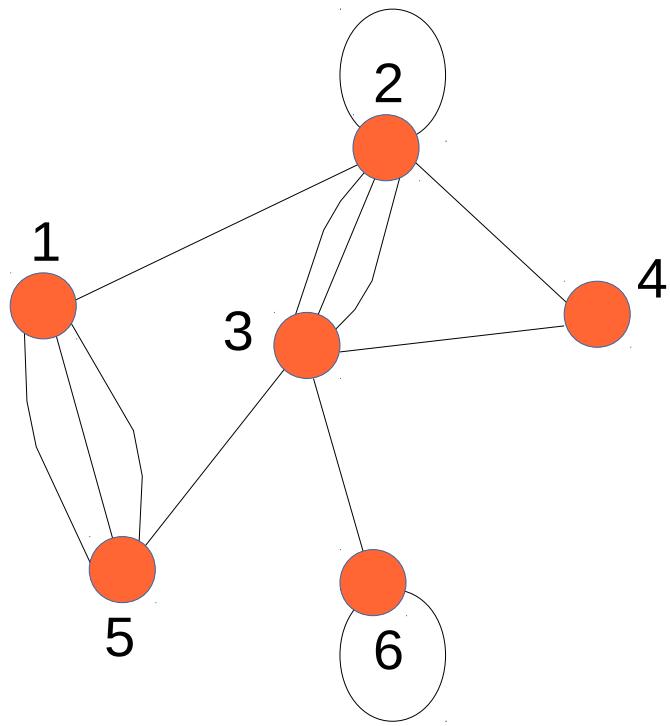
H

	v_1	v_2	v_3	v_4
v_1	0	0.4	0	0.4
v_2	0.4	0	0.3	0.1
v_3	0	0.3	0	0.5
v_4	0.4	0.1	0.5	0

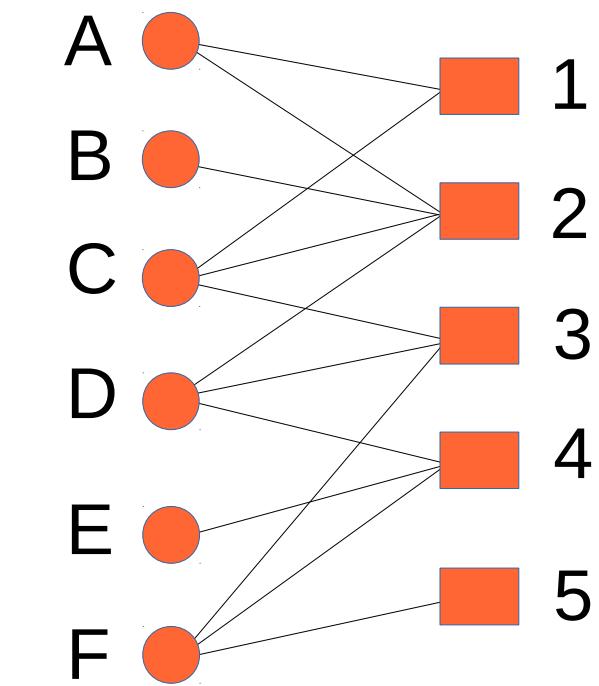
Types of Network



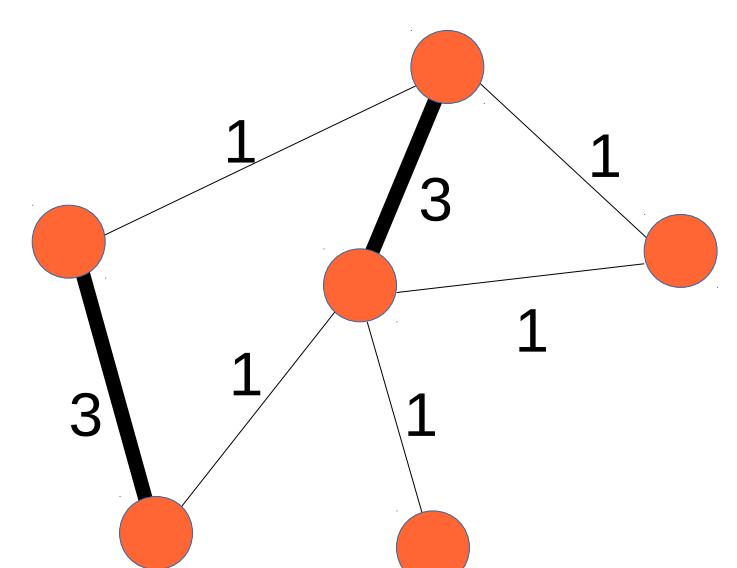
Simple



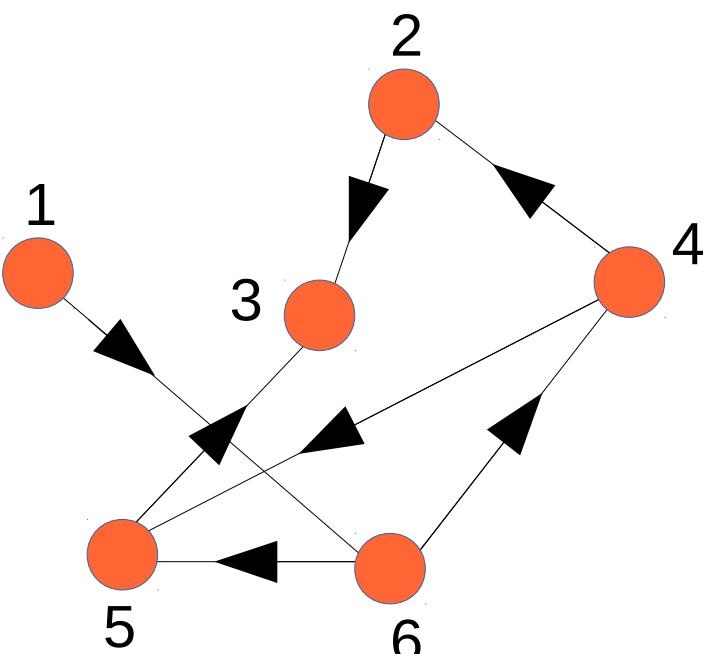
Multigraph



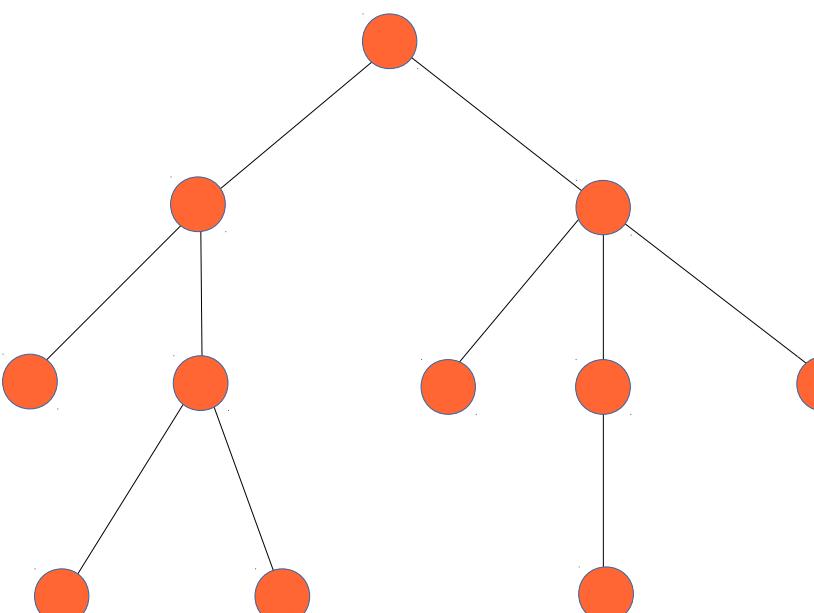
Bipartite



Simple (weighted)

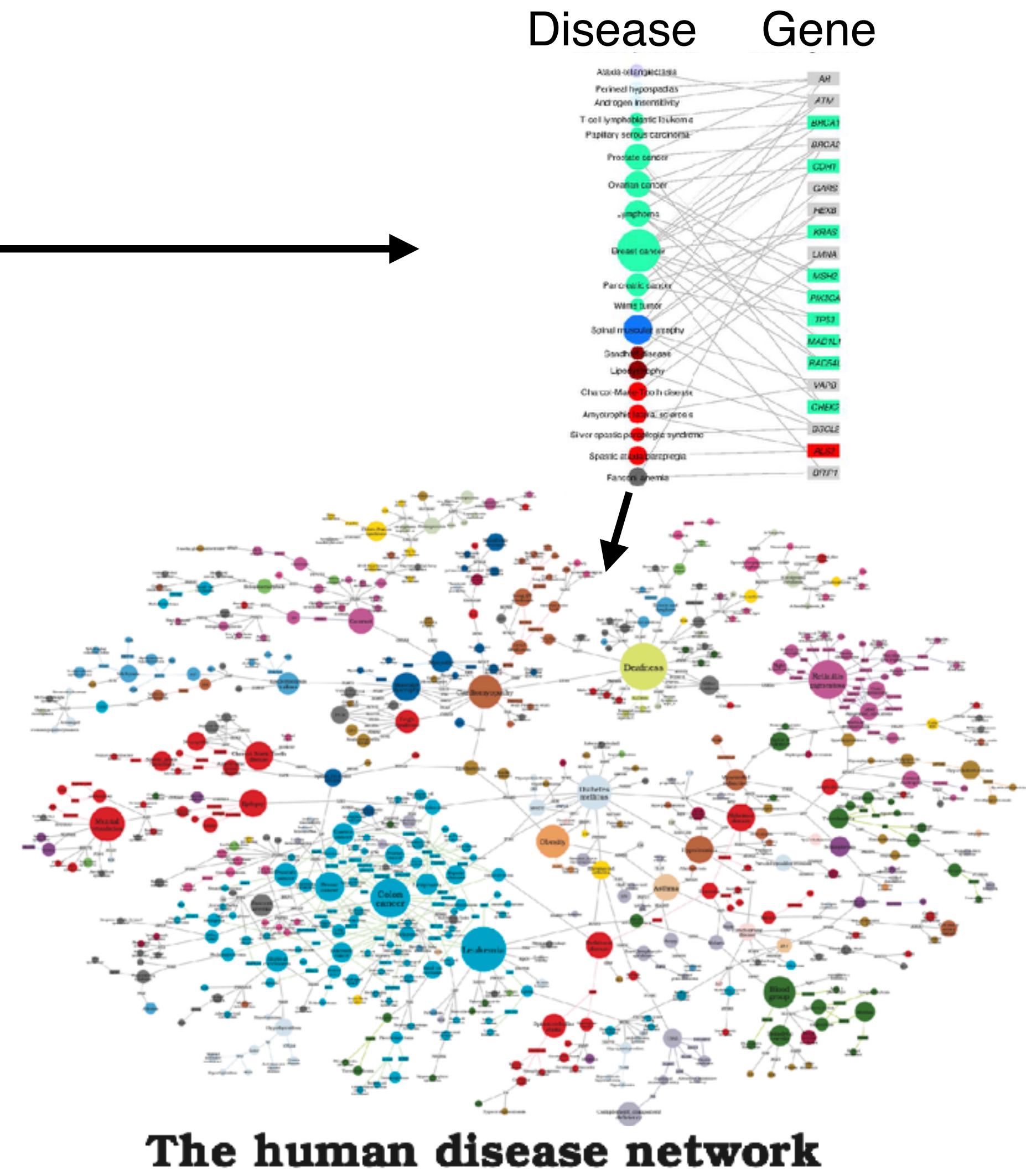
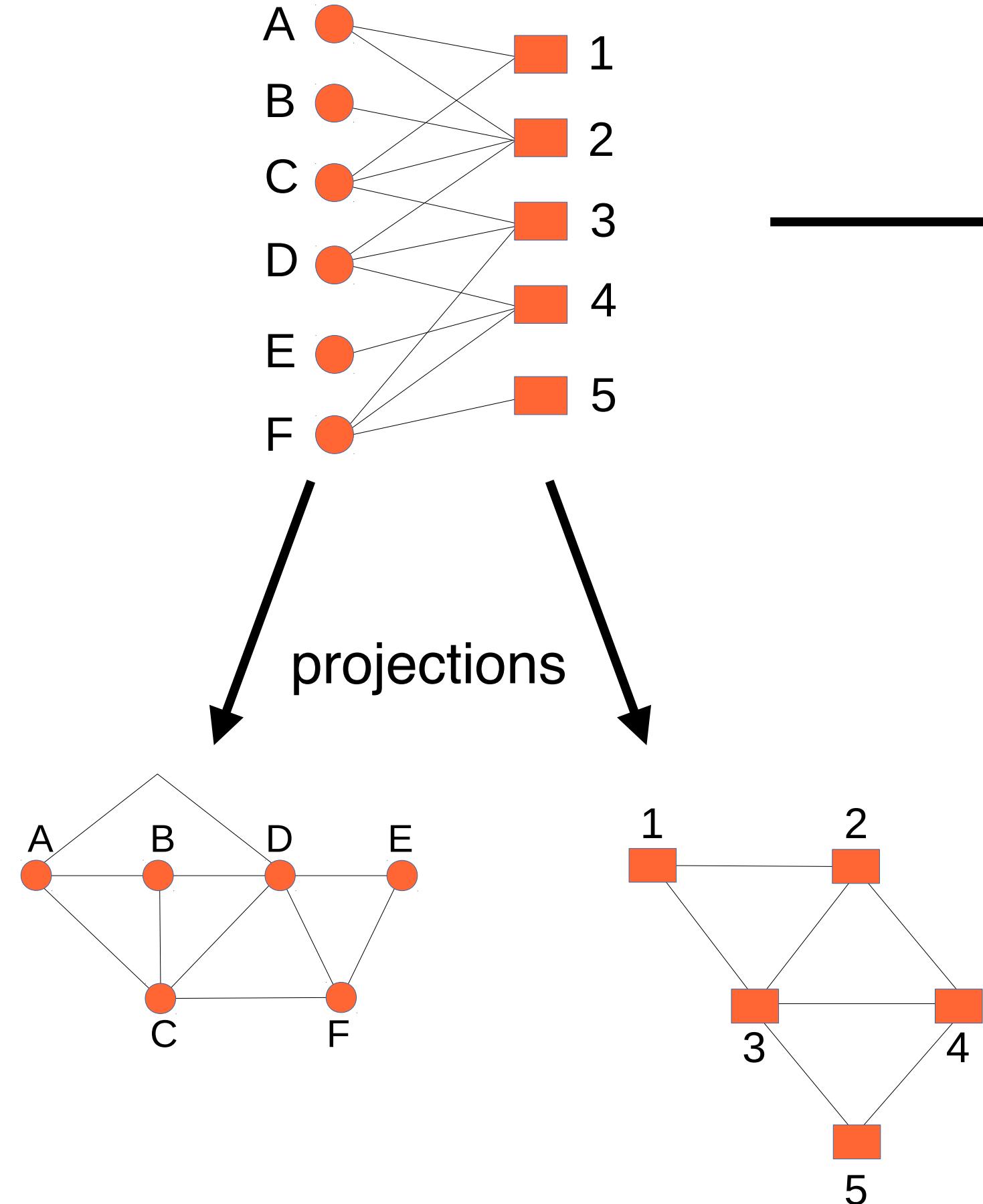


Directed

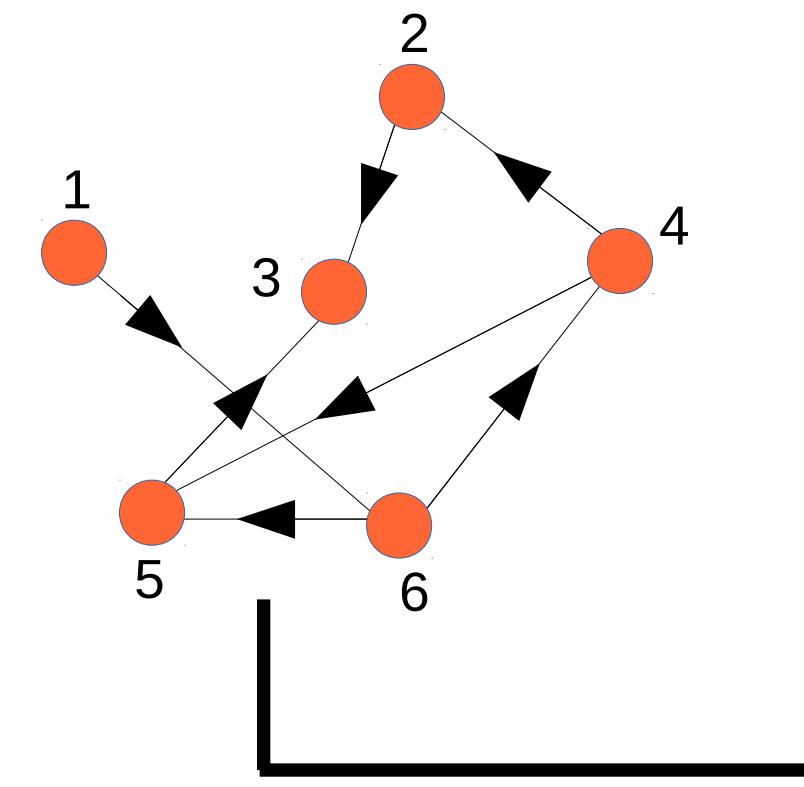


Tree

Network - Bipartite



Network - Directed



Persistent activity...

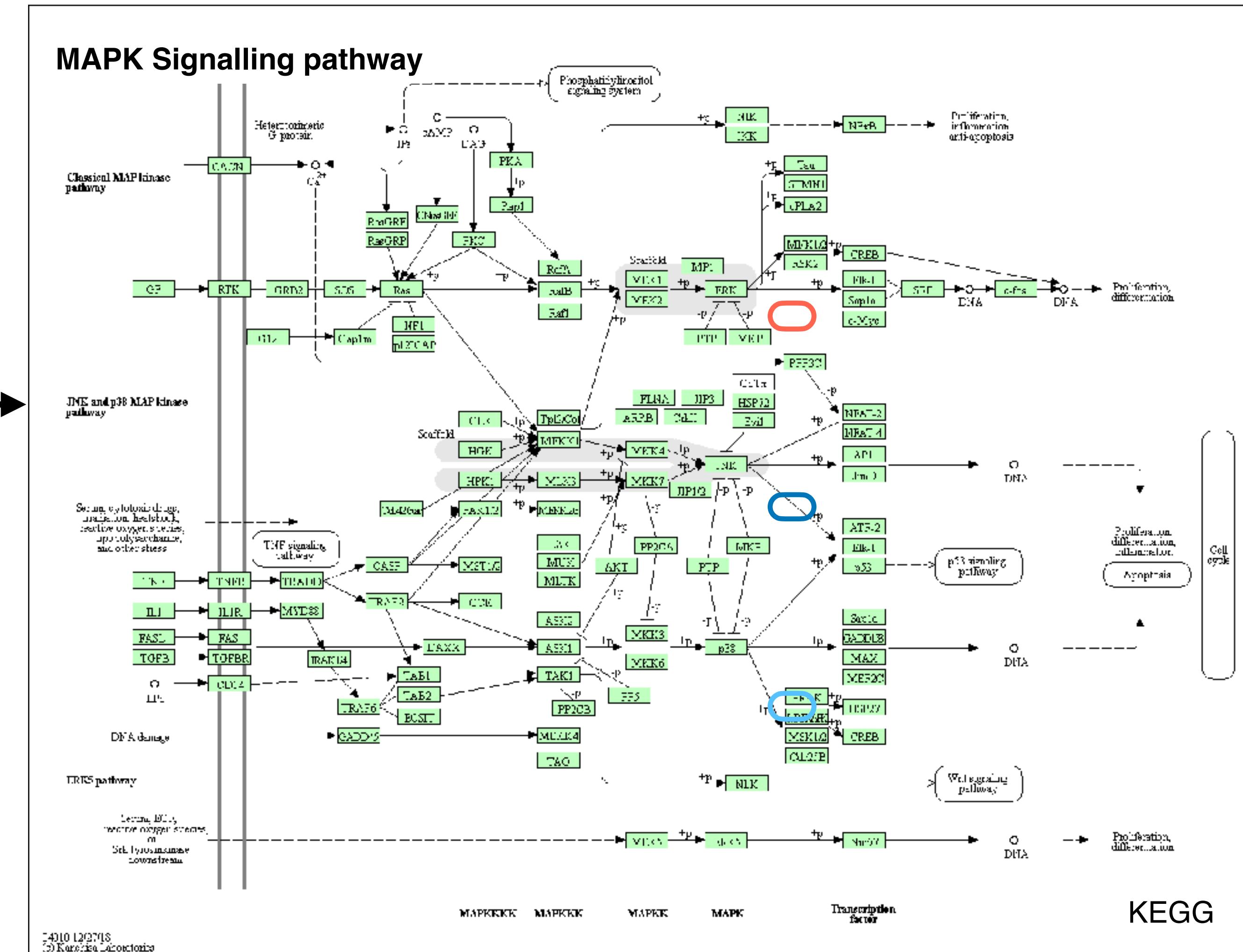
→ ERK →

Key role in tumorigenesis

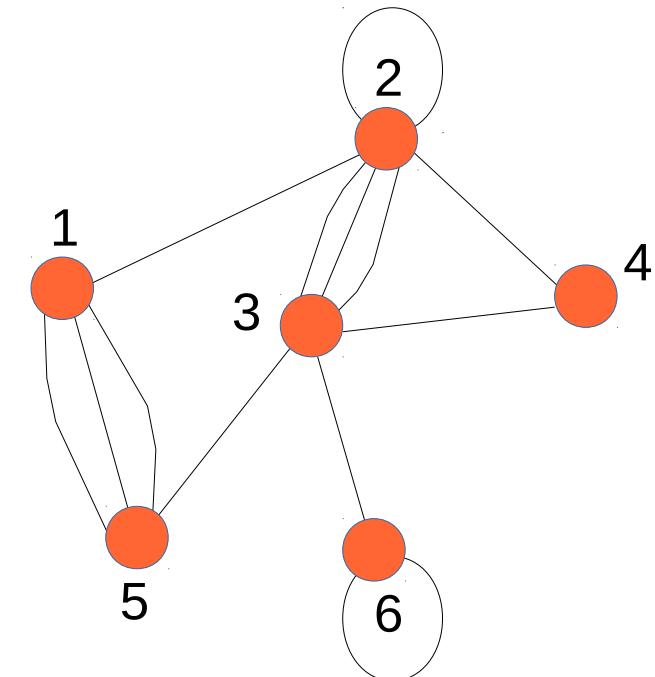
→ JNK →

mediating neural apo

→ p38 →

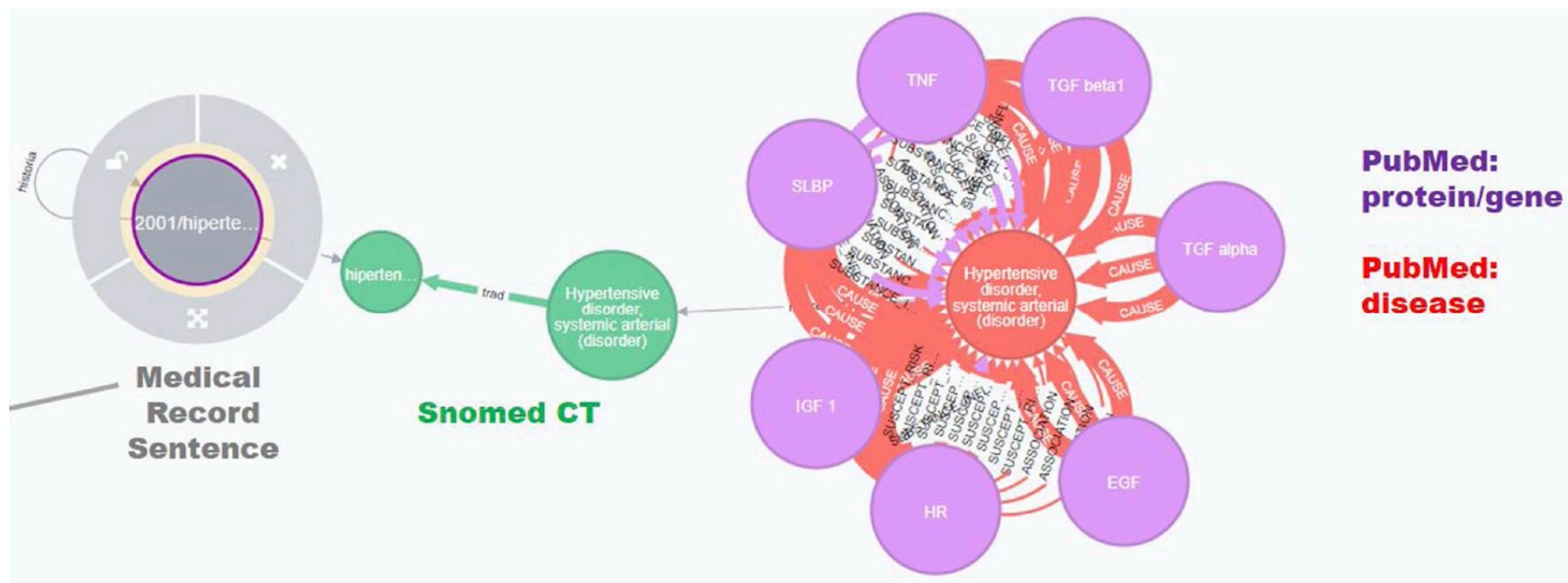


Network - Multigraph



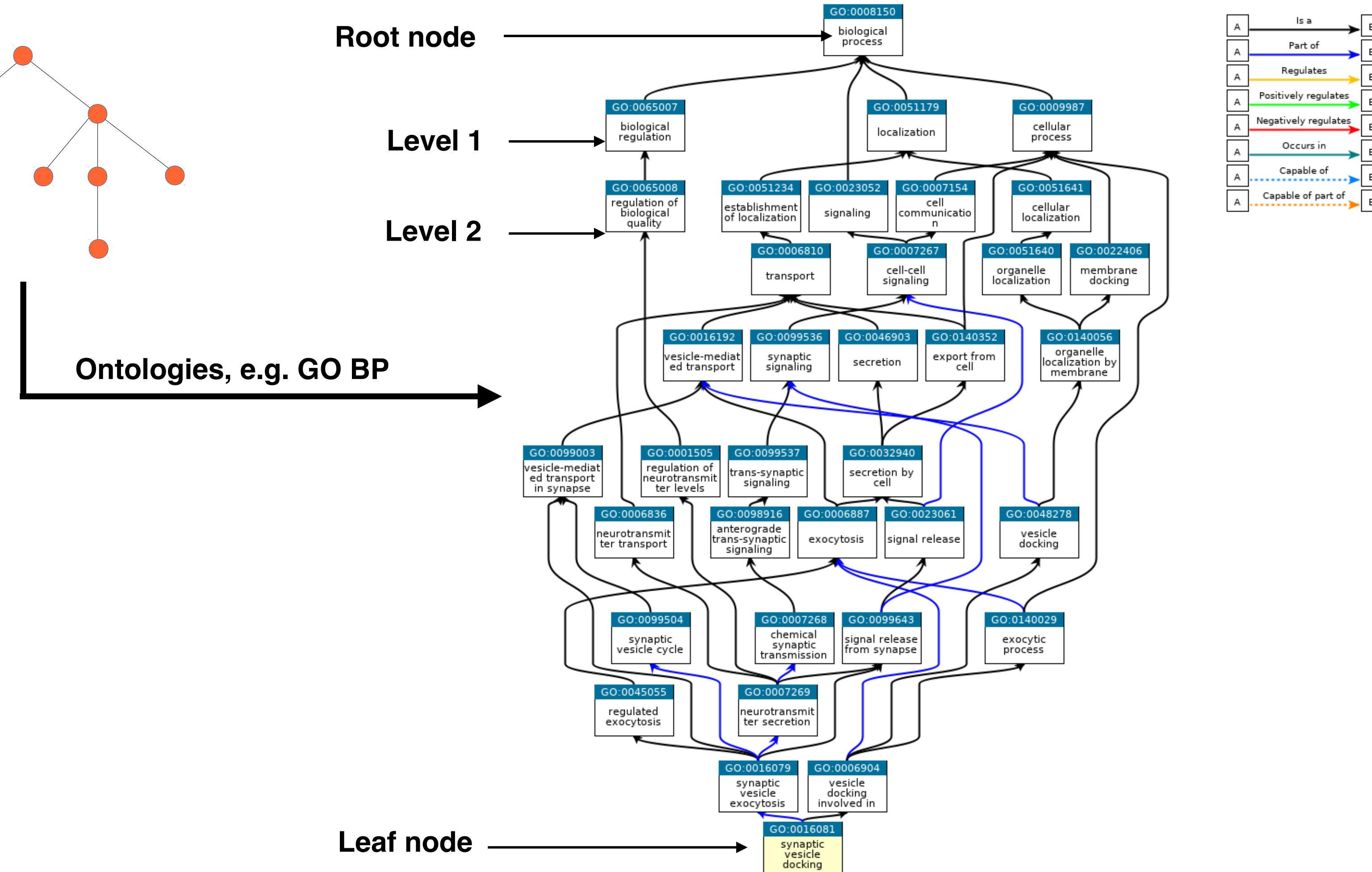
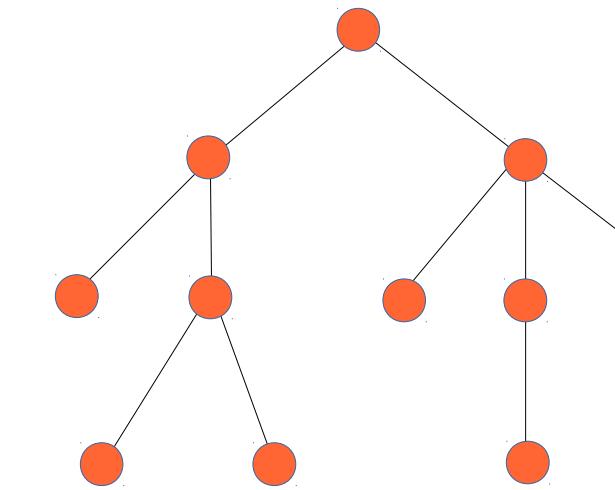
Graphical Database - Neo4j:

- knowledge path between patient medical records, to corpus of curated disease and gene data from pubmed articles.
- multiple node types
- multiple edges types between nodes.



(Bello, F.L. et al. 2019)

Network - Tree



QuickGO - <https://ebi.ac.uk>

Network - Simple

Our Example Network

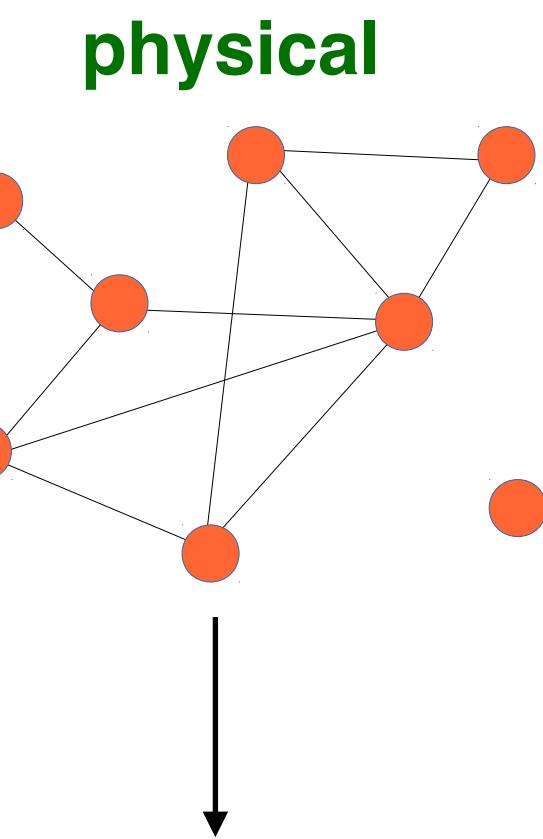
- We'll focus on a simple network type for our analysis examples.
 - Proteomic network with edges known to exist from measurement.

Sparsity:

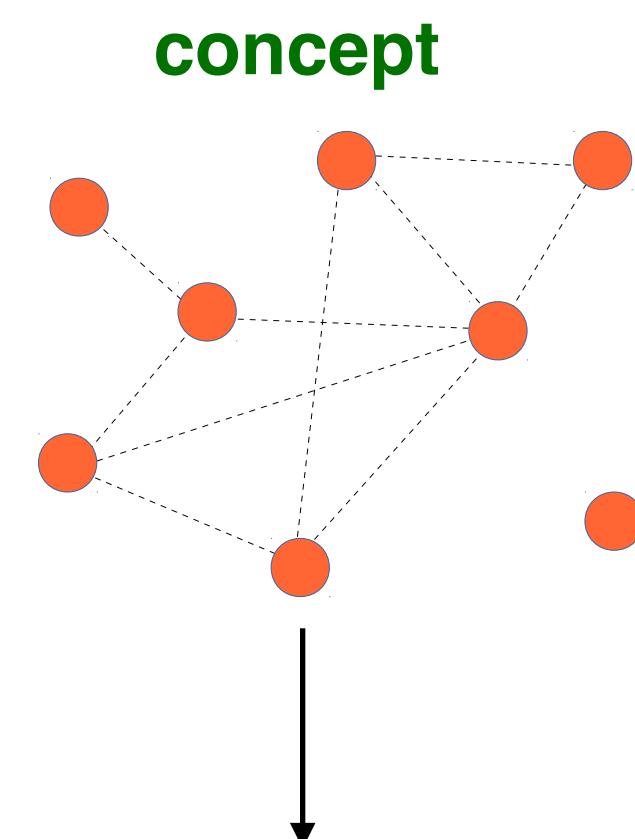
- each pair of nodes could have an edge, but most don't

Concept Network

- Another simple network where edges are not known, but inferred from data associated with each node.
 - Common as includes node data, e.g. fMRI, microarray, medical records.

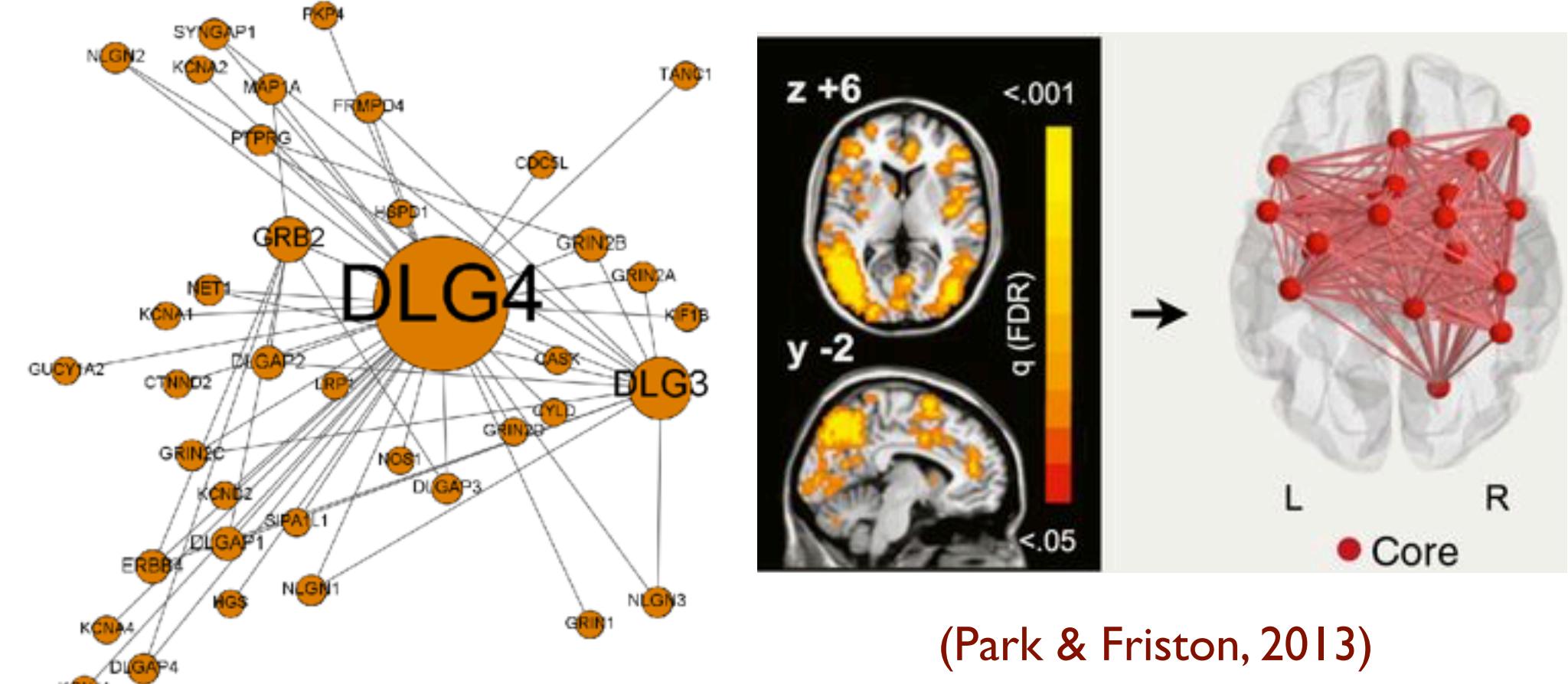


Protein Interactions - for DLG4



fMRI image data

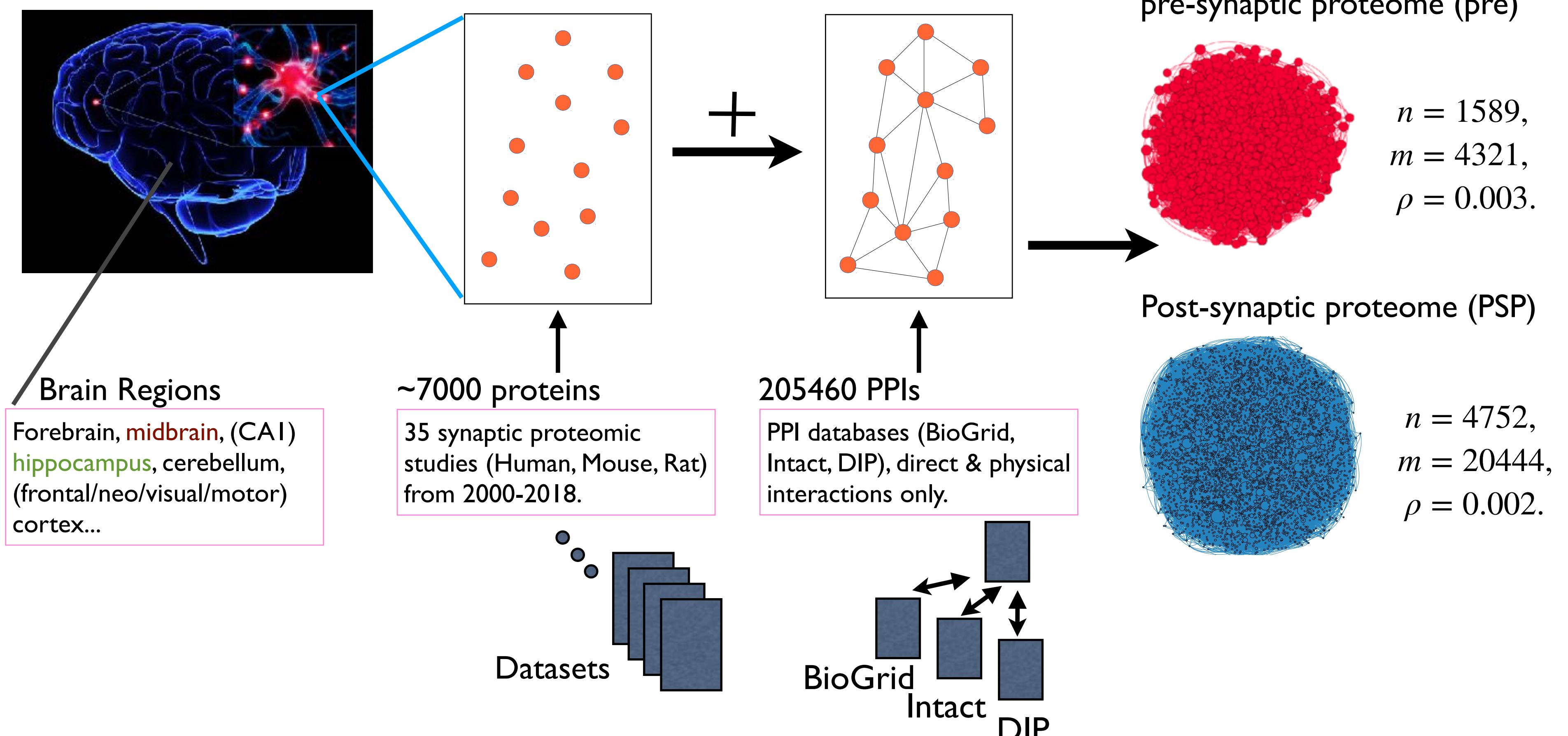
- core regions in task



Simple Network - Protein-Protein Interactions in the Synapse

What we can do with the data:

- Enrichment of data for diseases.
- Construct interaction network from data
- Underlying structure in network (clustering)
- Enrichment of clusters for diseases & synaptic function.



Network - Constraints

physical:

- useful thinking about our networks as being either physical or concept.

unweighted:

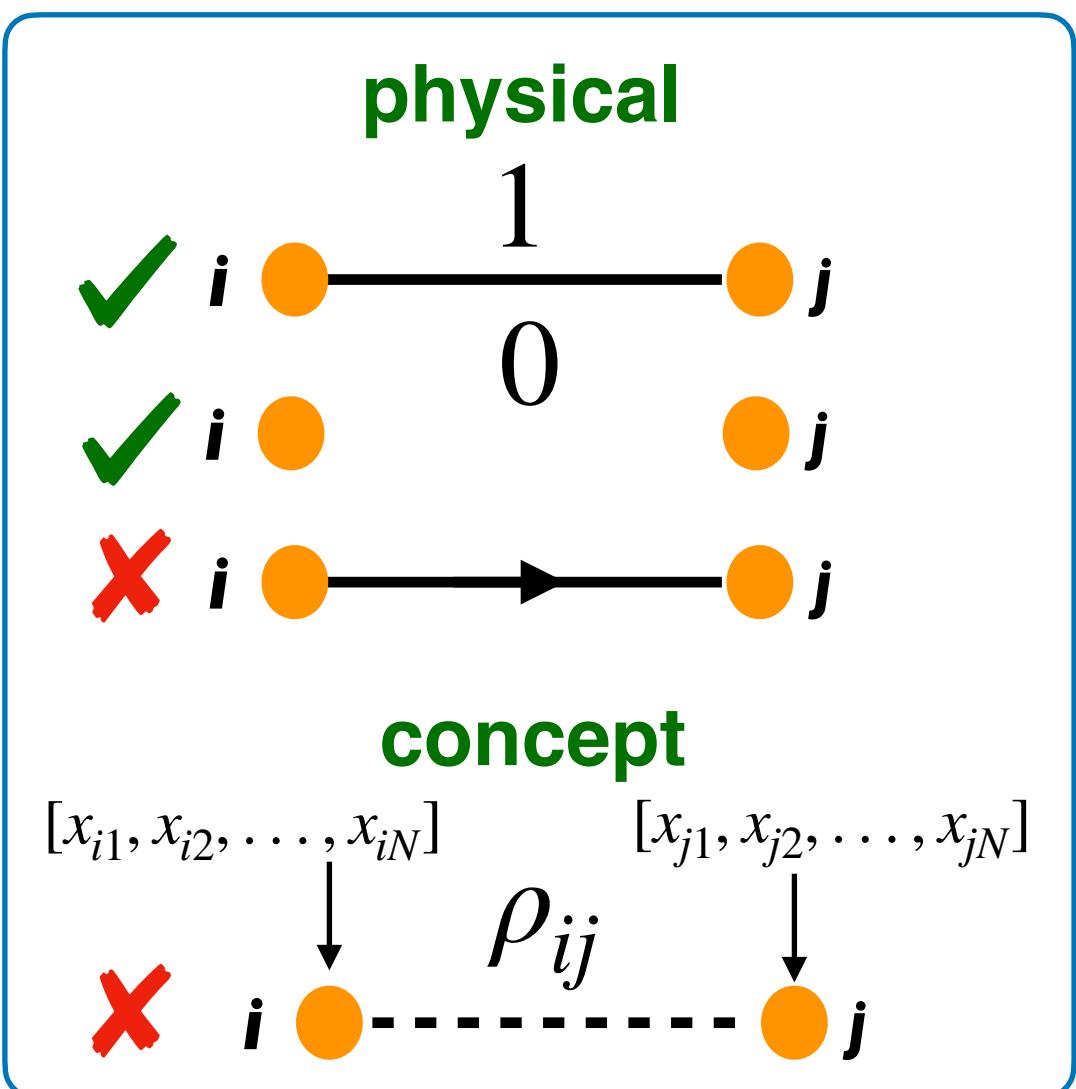
- every edge in network assigned either 1 "measured" or 0 "not measured".

undirected:

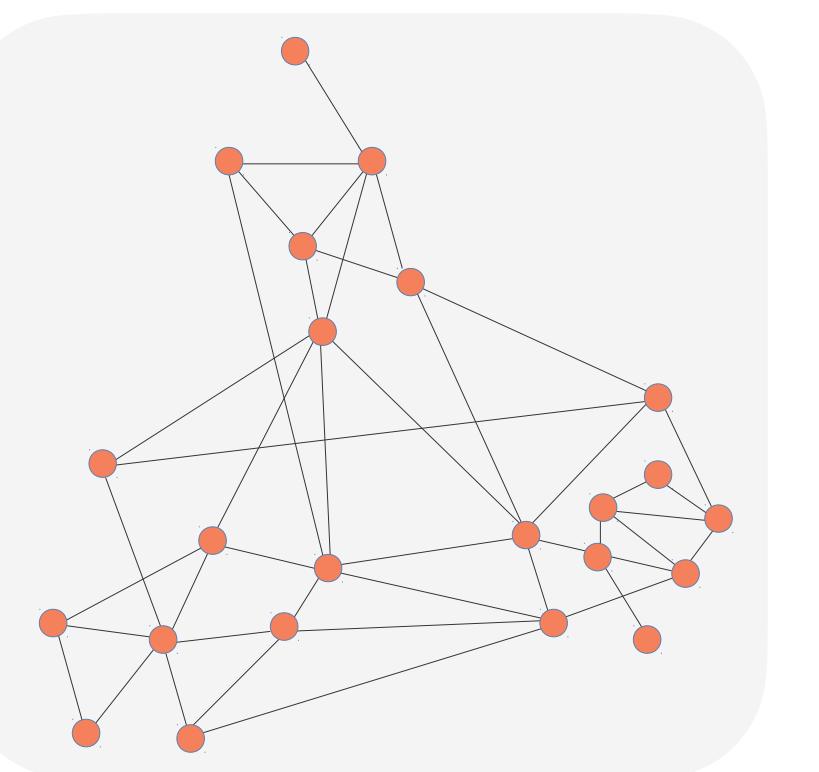
- every measured edge in network can go from $i \rightarrow j$ or $j \rightarrow i$ with equal magnitude.

single component:

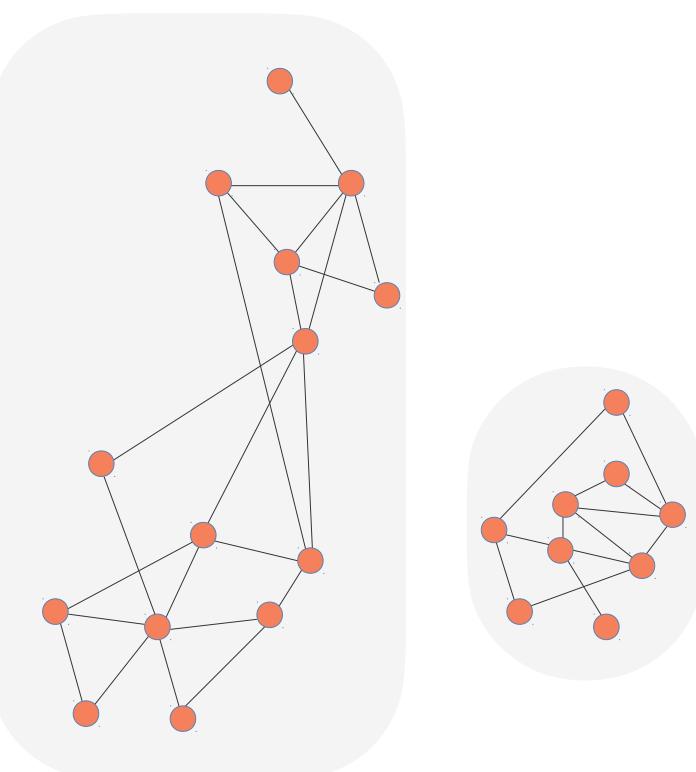
- fraction of nodes (S) in the largest component is always 1.



$S = 1$



$S = 0.6$



Network - Representations

Adjacency Matrix						
	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	1	0
3	0	1	0	1	0	0
4	0	0	1	0	1	0
5	0	1	0	1	0	1
6	0	0	0	0	1	0

$$k_i = \sum_{j=1}^n A_{ij} \text{ (degree of node } i\text{)}$$

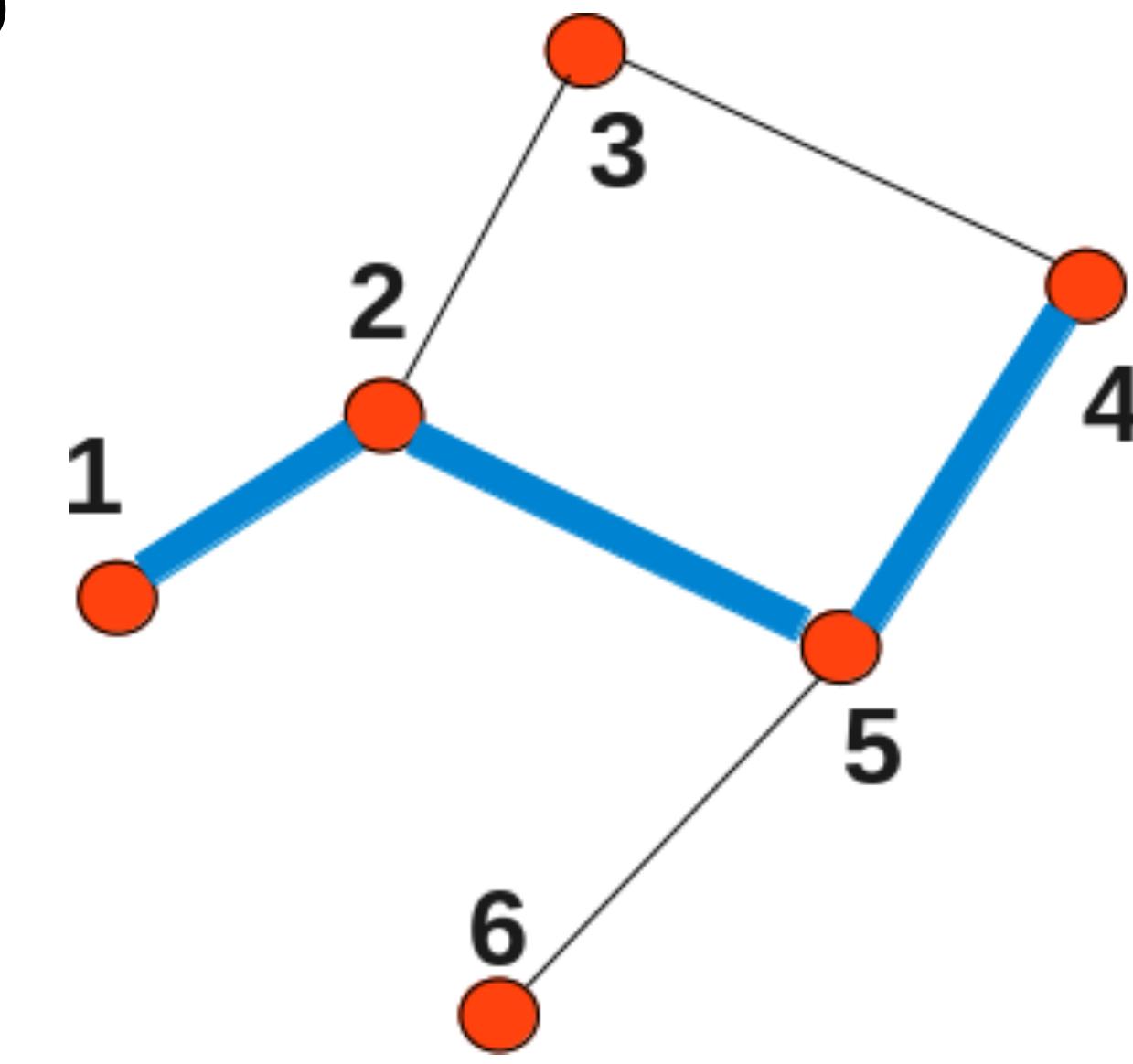
Degree Matrix						
	1	2	3	4	5	6
1	1	0	0	0	0	0
2	0	3	0	0	0	0
3	0	0	2	0	0	0
4	0	0	0	2	0	0
5	0	0	0	0	3	0
6	0	0	0	0	0	1

$A_{ij} \in \mathbb{R}$ (unweighted = 0,1)

$A_{ij} = A_{ji}$ (undirected)

$A_{ii} = 2$ (if self-edge)

(edges in network)



$$\frac{1}{2} \sum_{i=1}^n k_i = 6$$

Laplacian Matrix

	1	2	3	4	5	6
1	1	-1	0	0	0	0
2	-1	3	-1	0	-1	0
3	0	-1	2	-1	0	0
4	0	0	-1	2	-1	0
5	0	-1	0	-1	3	-1
6	0	0	0	0	0	1

$$L = D - A =$$

Network - Measures

Measures:

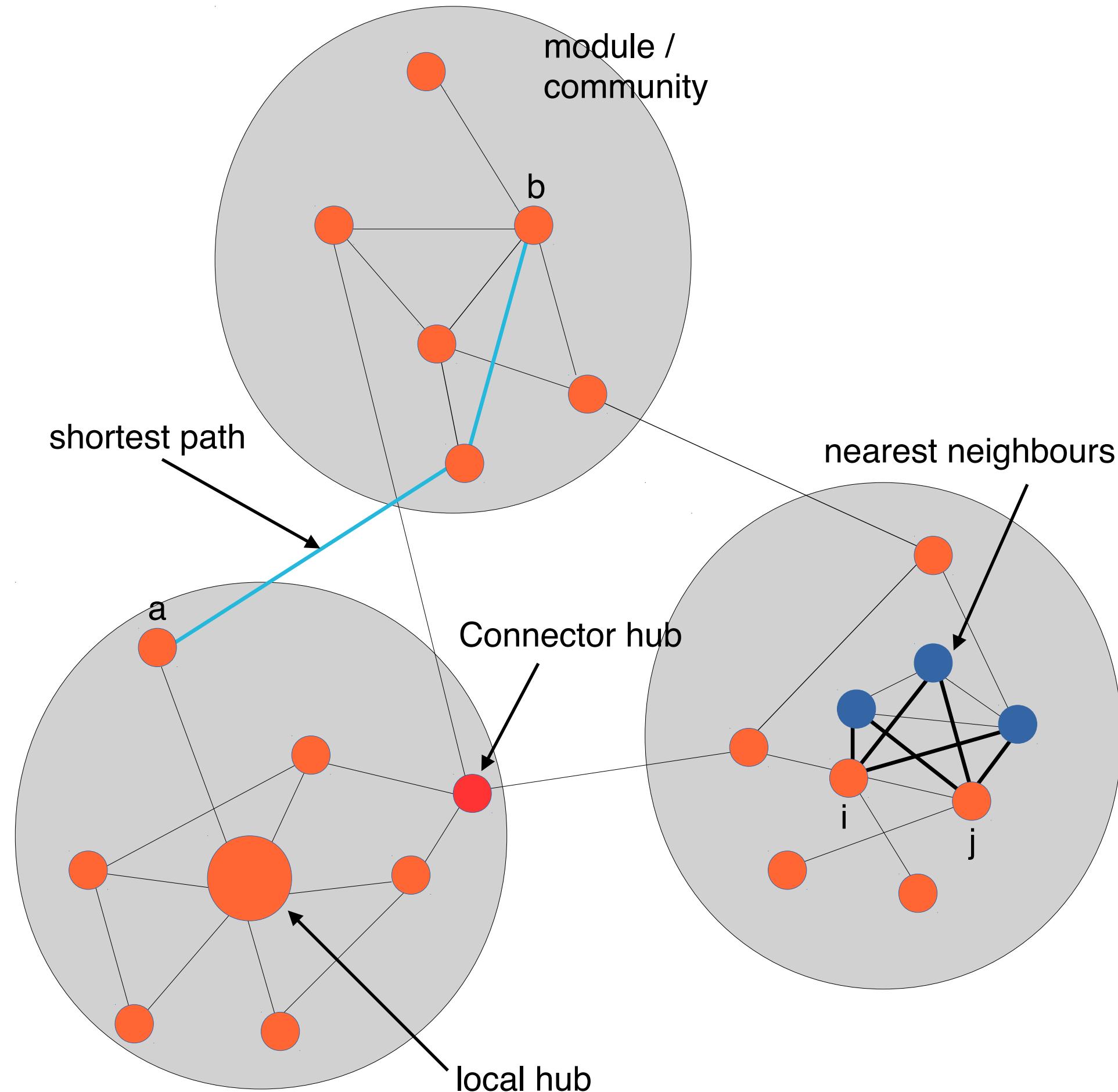
- From our mathematical representation, we can define many concepts which reveal a networks underlying character.

Centrality:

- "What are the most important or central nodes (i.e. influencer, disease hub) in our network."
- Degree, Betweenness Google's PageRank, Semi-local, Closeness...

Similarity:

- "How similar two nodes are to each other."
- how many shared neighbours.
- Pearson correlation, Modularity.



Network - Degree

One of the most useful concepts, encoding a network's local structure:

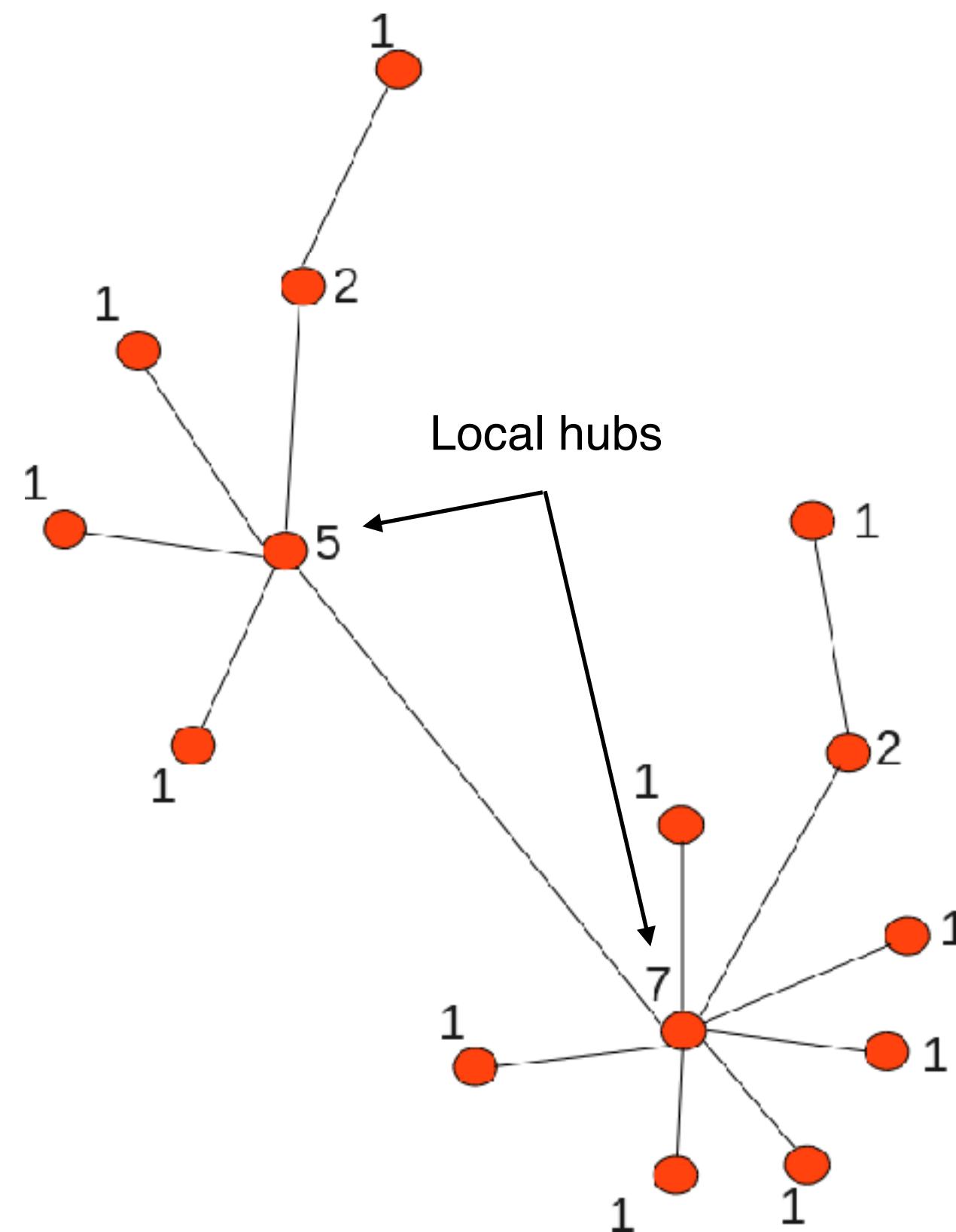
- can be used as a centrality score.
- generate random networks...
- (a test) degree distribution of many real-world networks follows a power-law: $\propto k^{-\alpha}$

Node Degree Centrality:

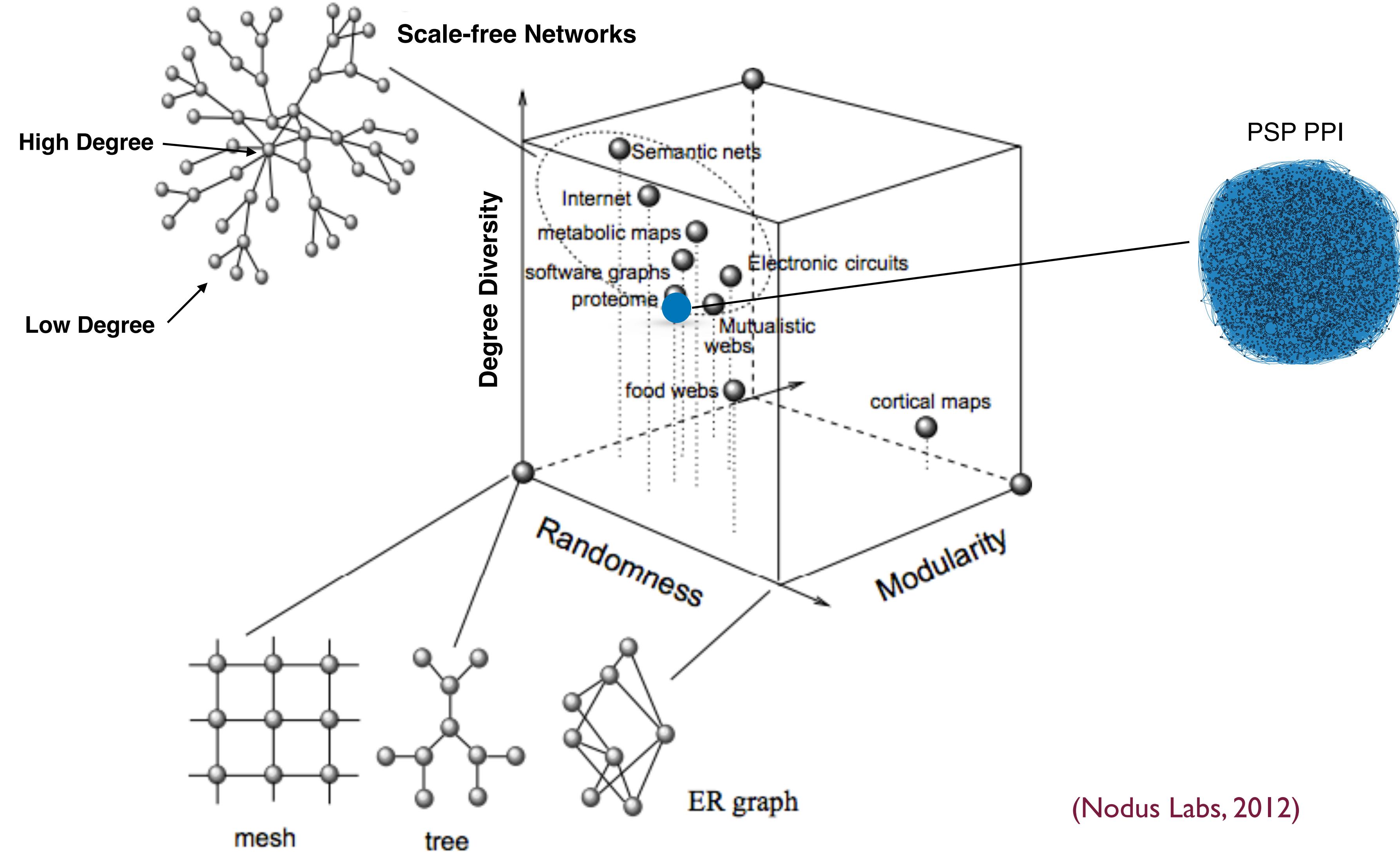
- Number of neighbours, denoted by k
- local hubs $\Rightarrow \begin{cases} \text{highest degree nodes} \\ \text{disease related proteins} \end{cases}$

5 highest degree nodes in PPI PSP

Gene	Degree	Diseases
EWSR1	232	MND
GRB2	205	SCH
SPRK2	183	SCH
LMNA	180	AD
EGFR	174	AD;Epi;SCH;HD;BD



Network - Degree

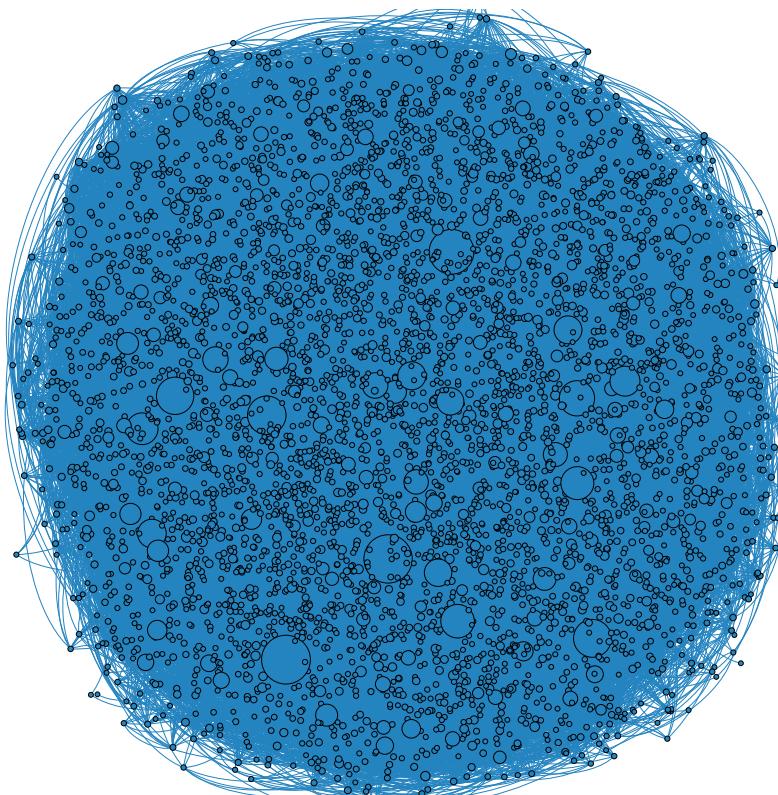


Network - Degree

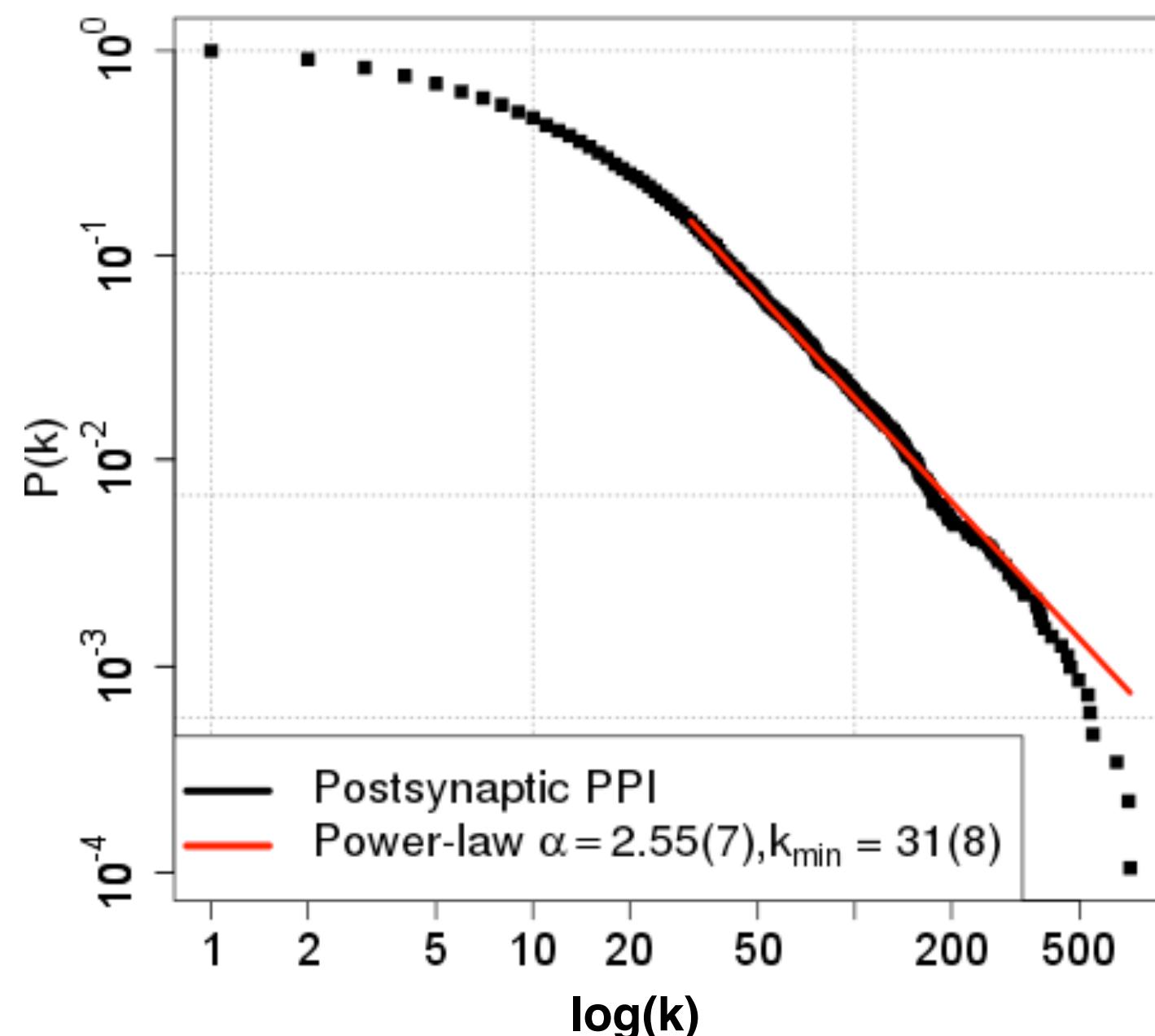
One of the most useful concepts, encoding a network's local structure:

- can be used as a centrality score.
- generate random networks...
- (a test) degree distribution of many real-world networks follows a power-law: $\propto k^{-\alpha}$

PSP PPI network

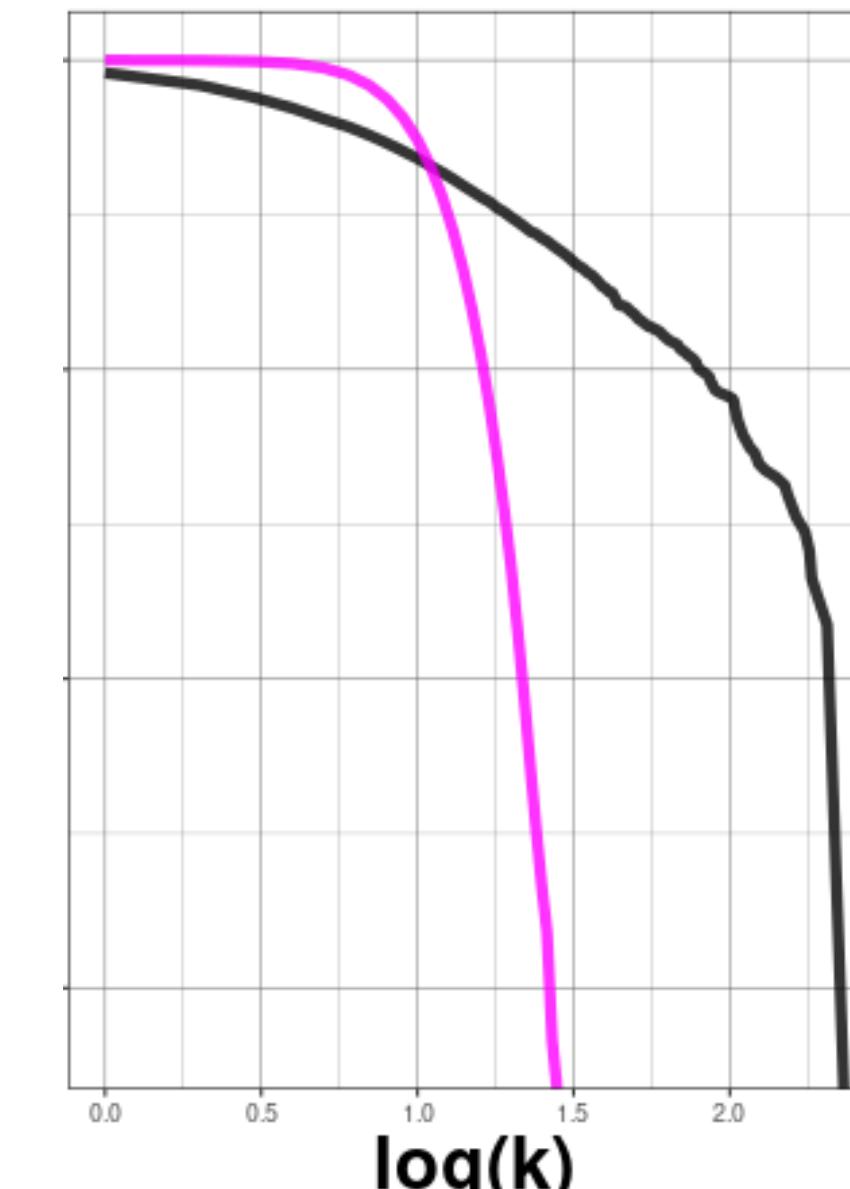


Degree Distribution



(Clauset, A. et al, 2009)
(Gillespie, C., 2013)

Random Distribution



— PPI_{obs}
— E-R

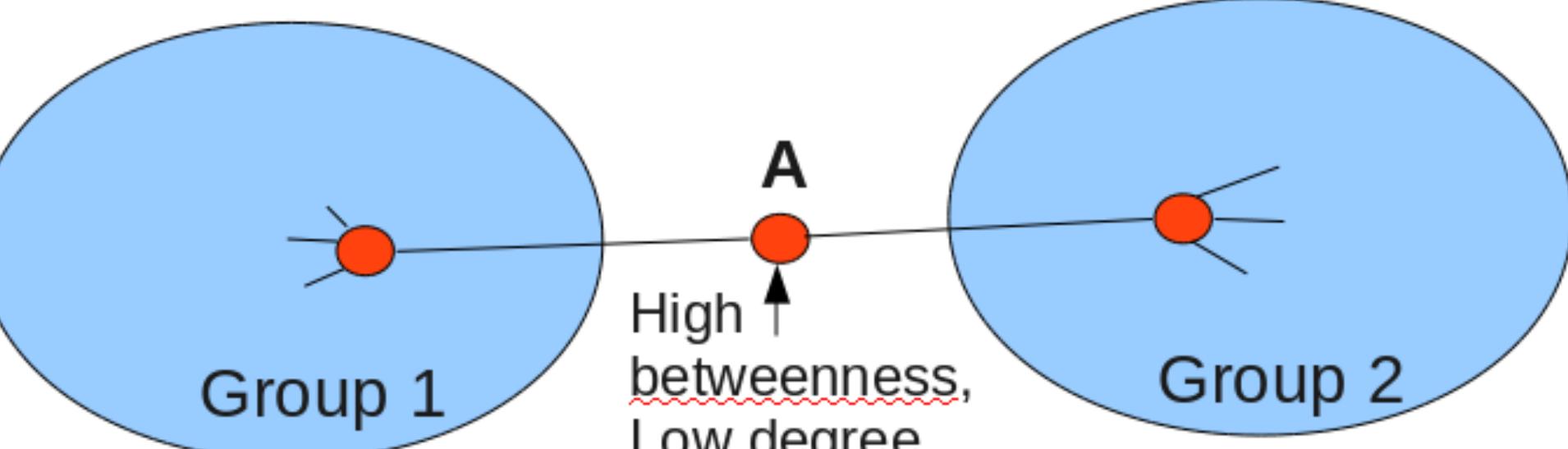


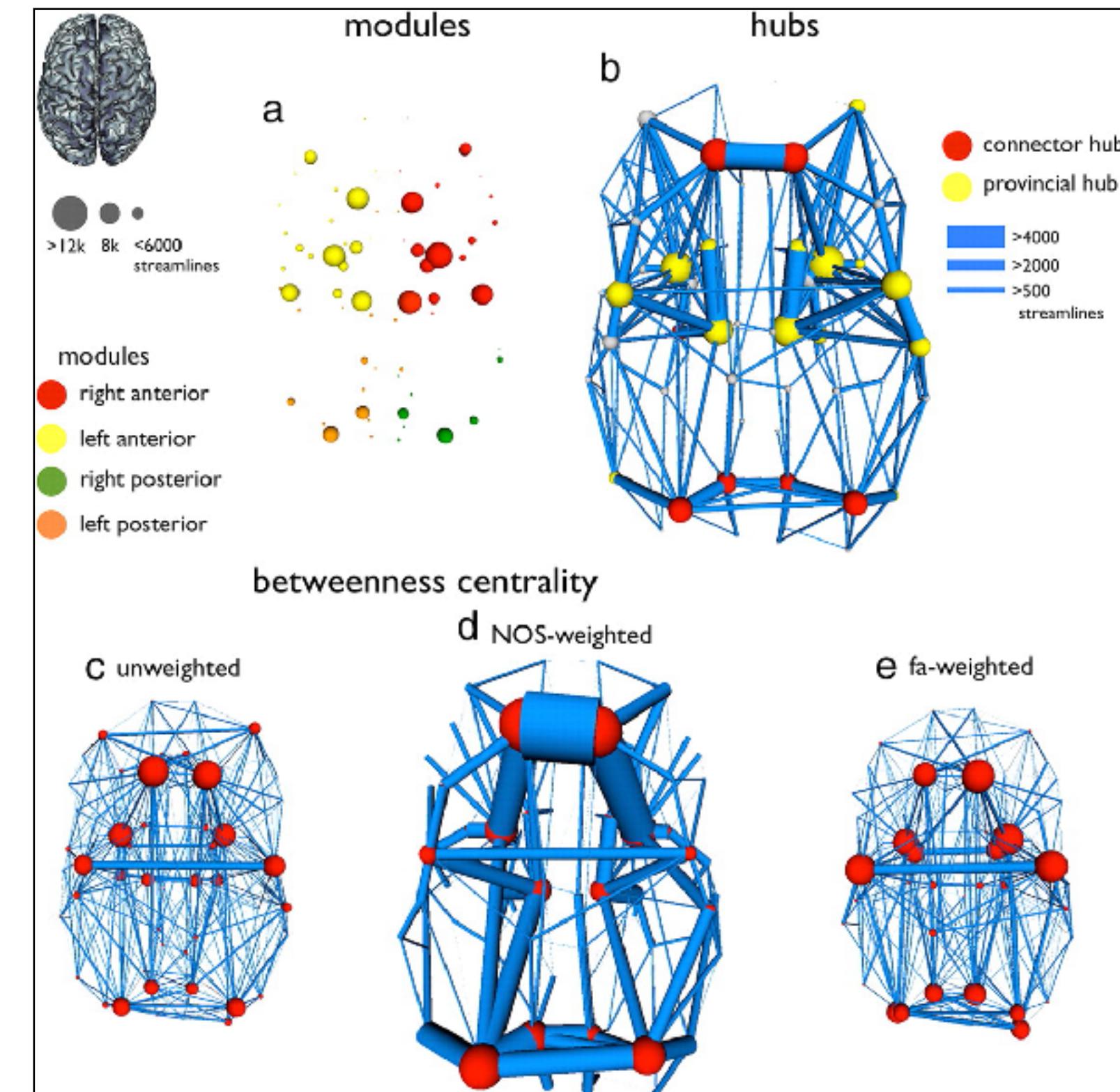
Erdős & Rényi (E-R)

Network - Betweenness

Betweenness:

- Fraction of shortest paths passing through a node.
 - Nodes can have low degree, be a long distance on average from others, and still have high betweenness. For instance 'A' below.

- connector hubs \Rightarrow 
 { high betweenness nodes
 controlling information flow



(Van der Heuvel & Sporns, 2011)

Network - Similarity

common neighbours (n_{ij}):

- nearest neighbours of nodes i and j is:

$$n_{ij} = \sum_k A_{ik} A_{kj} = \vec{A}_{i\cdot} \cdot \vec{A}_{j\cdot} = 3$$

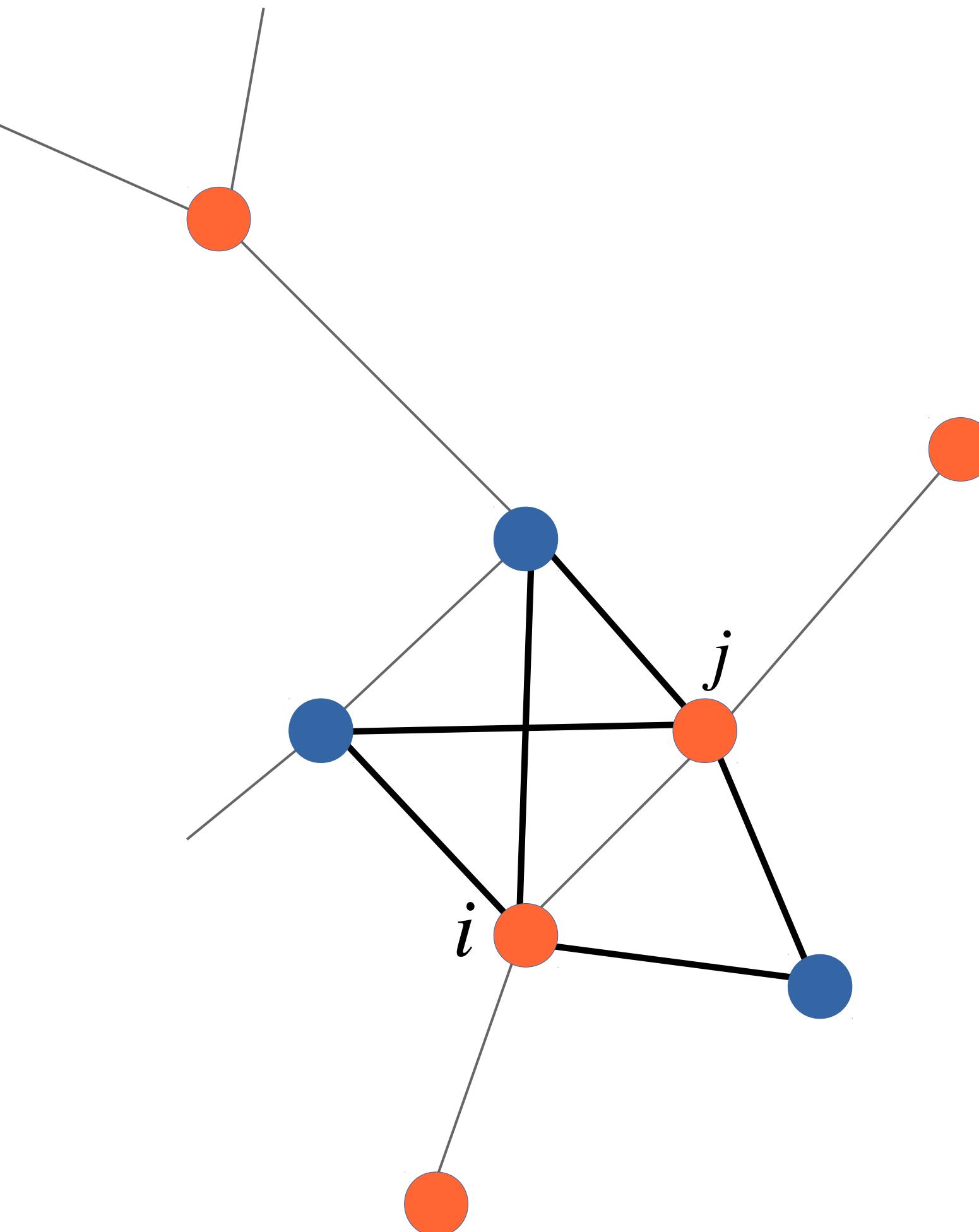
↓ ↓
each row(column) of \mathbf{A} is a vector

cosine:

$$- \sigma_{ij} = \frac{n_{ij}}{\sqrt{k_i} \sqrt{k_j}} = \frac{3}{\sqrt{5} \sqrt{5}} = 0.6$$

Pearson correlation:

$$- r_{ij} = \frac{\sum_k (A_{ik} - \langle A_i \rangle)(A_{jk} - \langle A_j \rangle)}{\sqrt{\sum_k (A_{ik} - \langle A_i \rangle)^2} \sqrt{\sum_k (A_{jk} - \langle A_j \rangle)^2}}$$



Network - Paths

Geodesic distance (d):

- Shortest path (excluding loops) between two nodes (d_{ik}) through network.

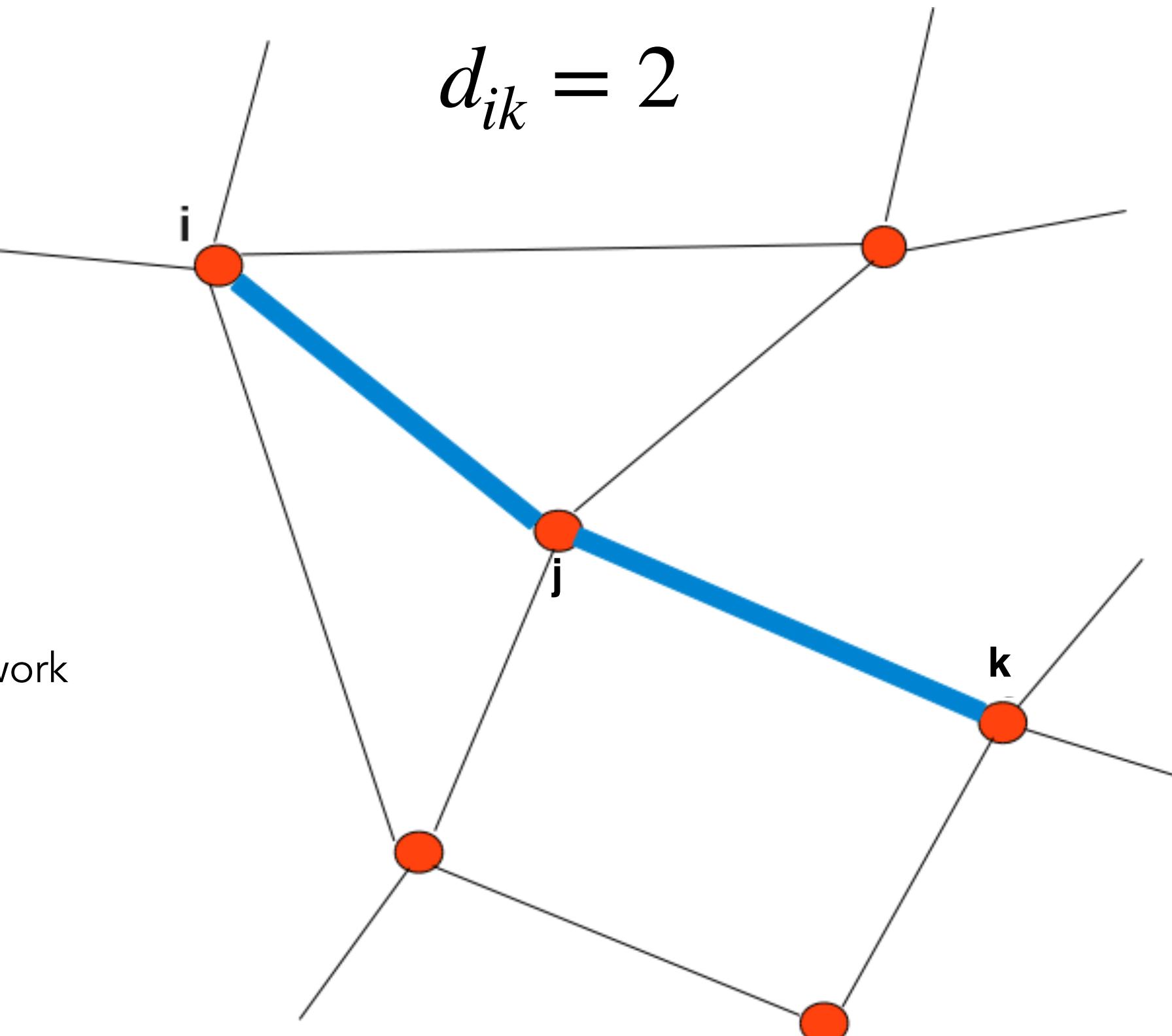
Random walks:

- Transition matrix \mathbf{P} , where probability going from i to j is: $P_{ij} = \frac{1}{k_i}$
- the steady state, i.e. after making many random walks:

$$\begin{array}{c} \xrightarrow{\quad} \text{Initial prob. dist. of nodes in network} \\ \downarrow \\ \mathbf{P}\vec{\pi} = \vec{\pi} \\ \mathbf{L}[\mathbf{D}^{-1}\vec{\pi}] = 0 \\ \uparrow \\ \mathbf{D}^{-1}\vec{\pi} = a\vec{1} \\ \therefore \pi_i = \frac{k_i}{2m} \end{array}$$

- probability $p(i \rightarrow j)$ of walk along edge from i to j at any time of random walk is:

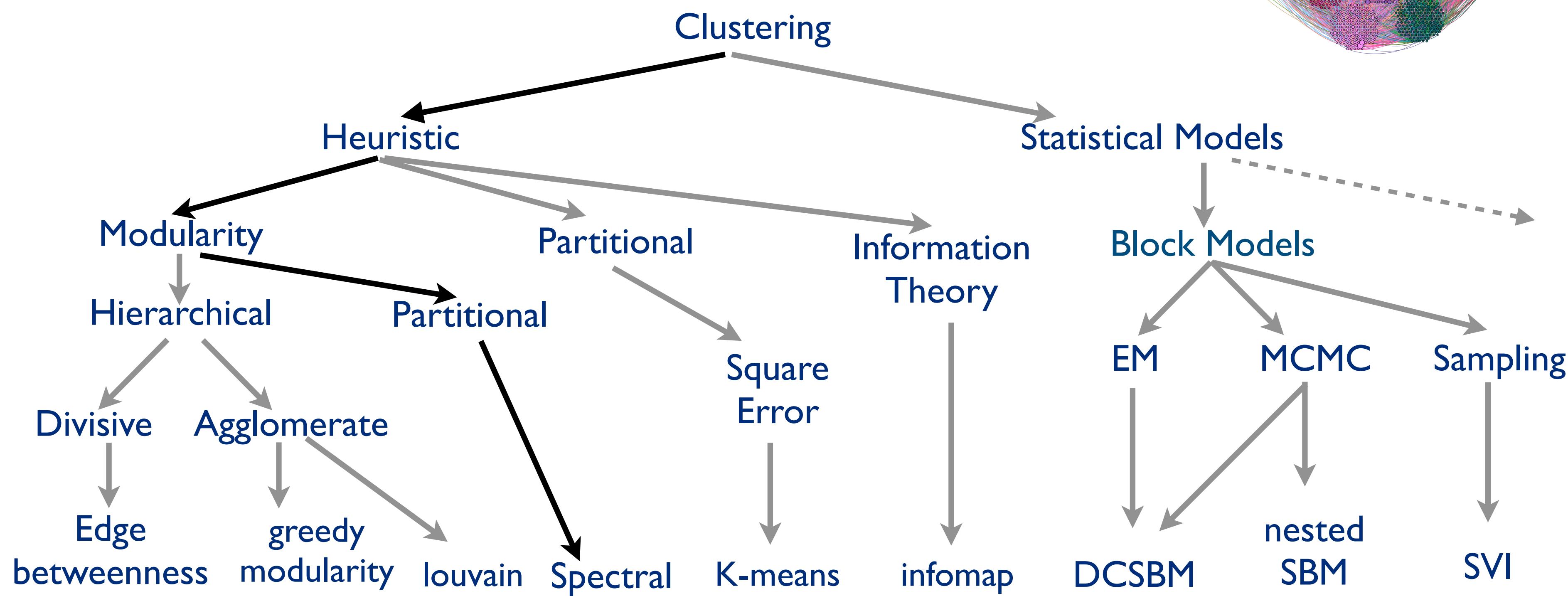
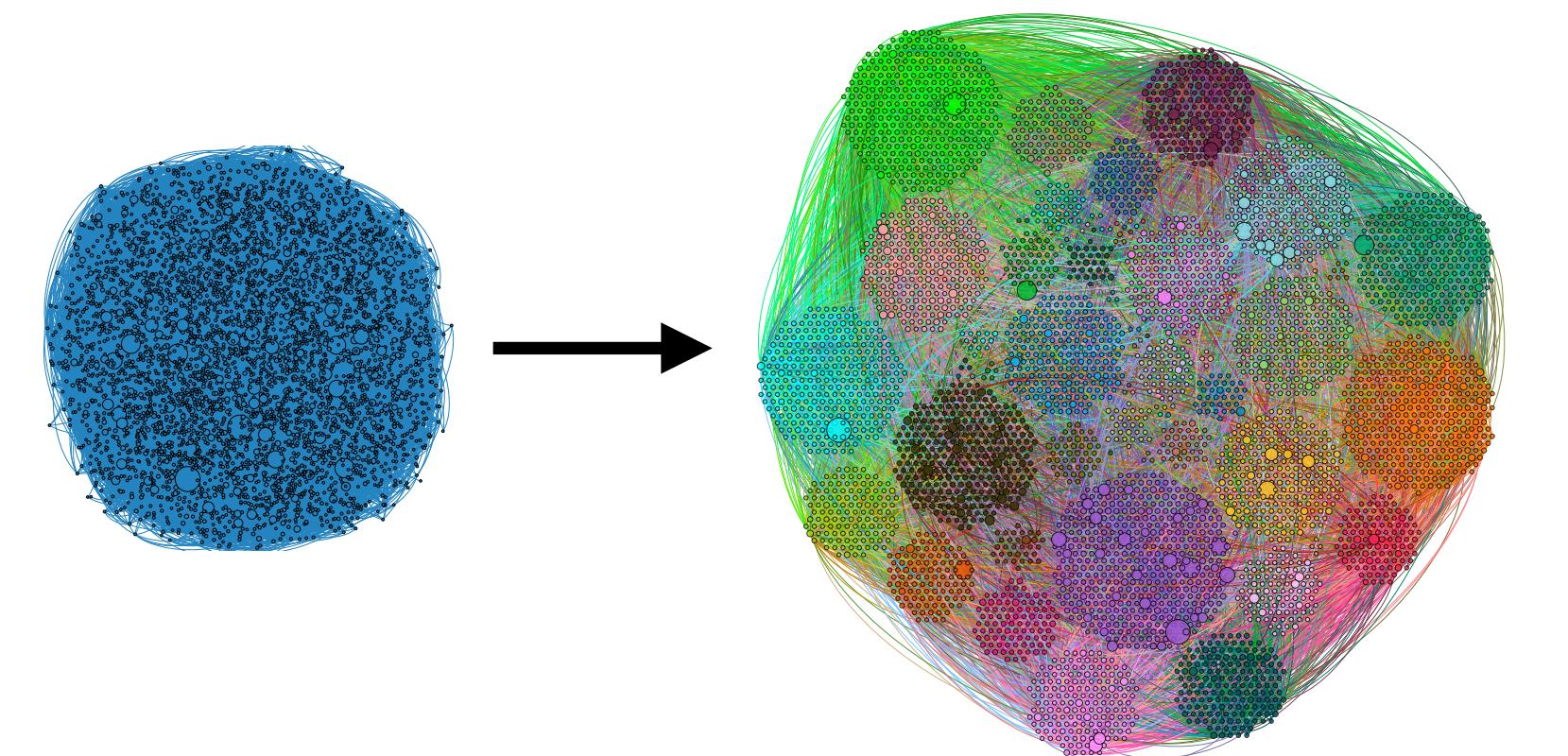
$$p(i \rightarrow j) = \frac{1}{k_i} \frac{k_i}{2m}$$



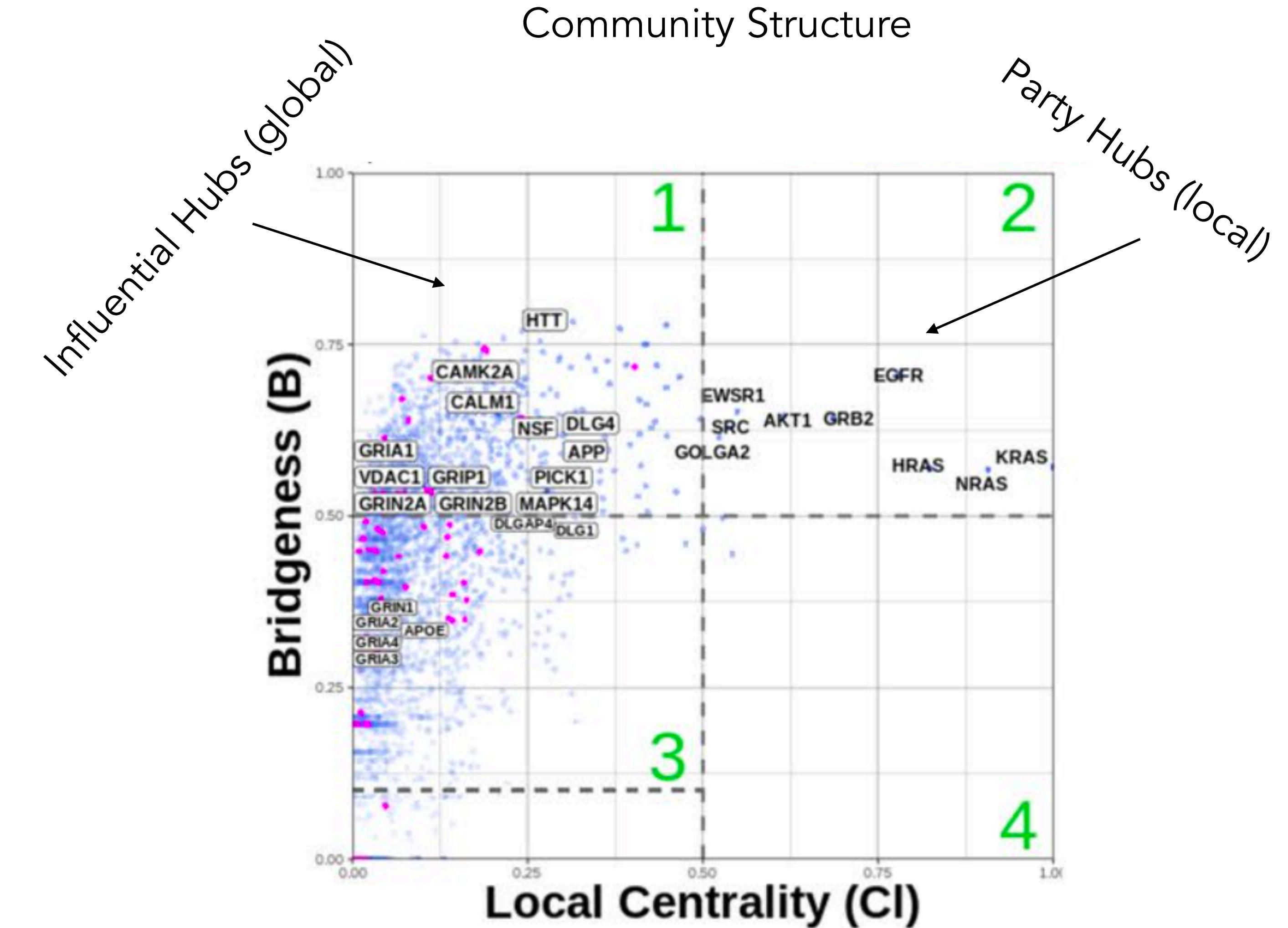
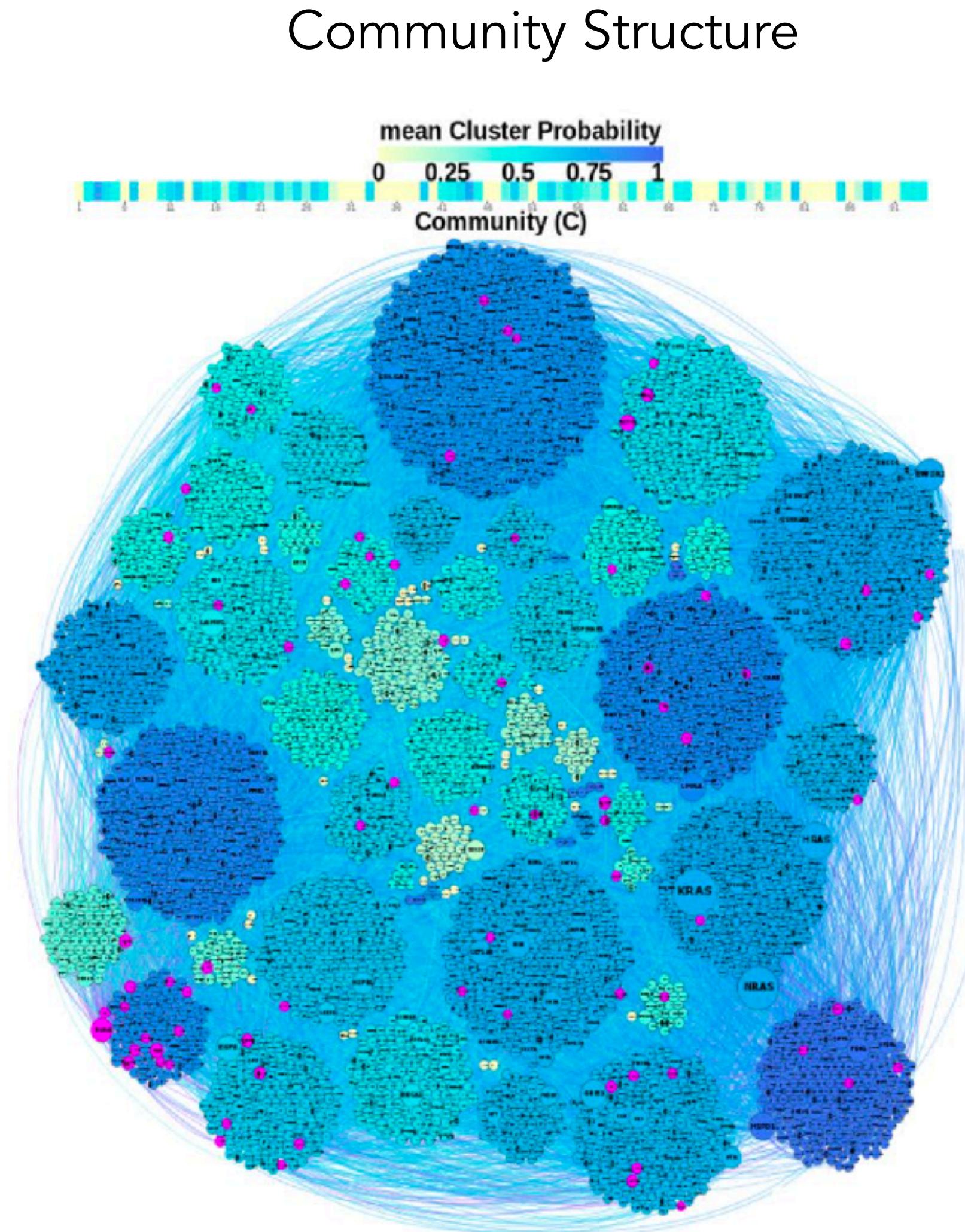
Network - Clustering

What is a Community?

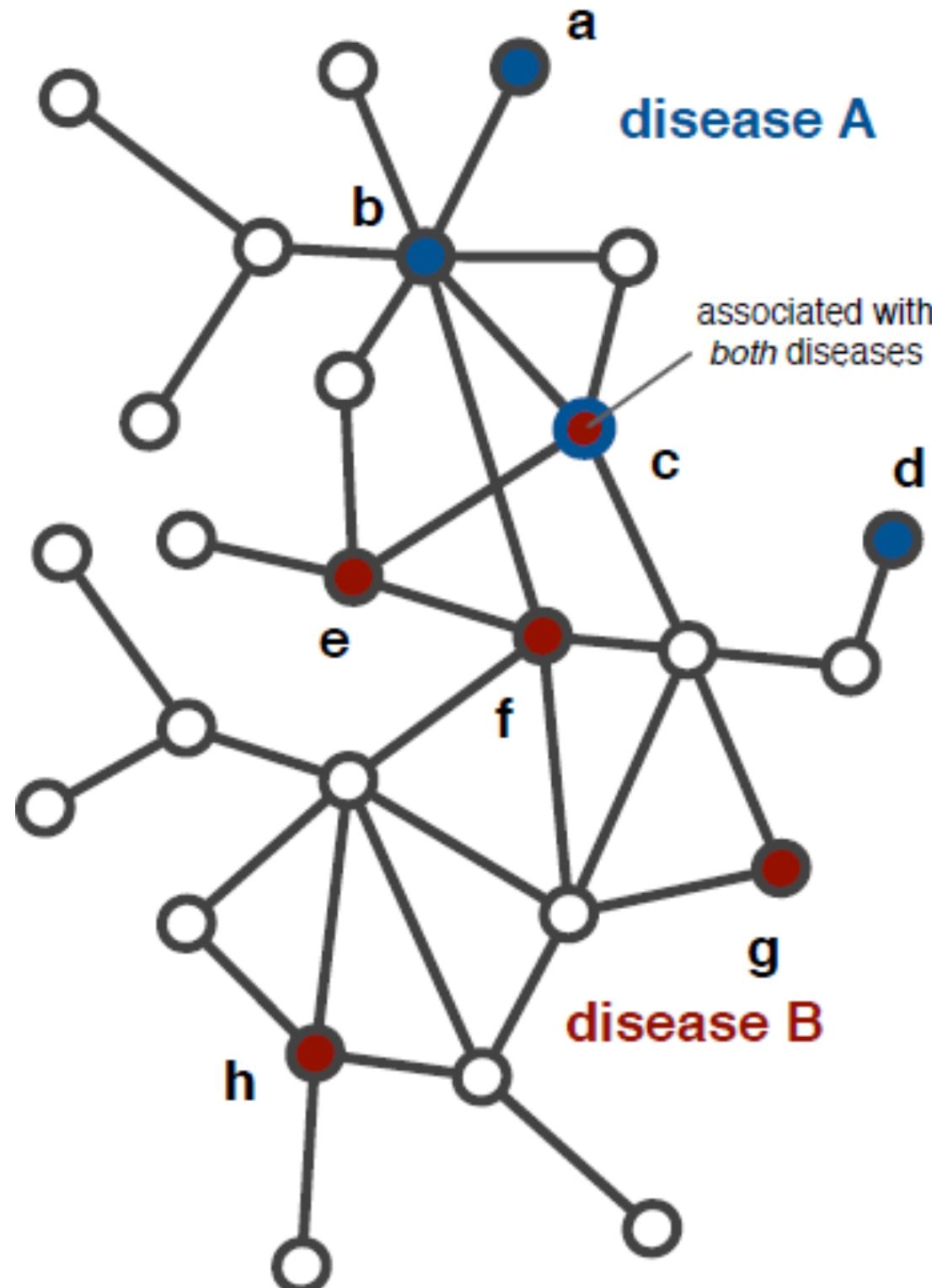
- locally dense regions of nodes in our network which are similar.
- modules, complexes, groups, blocks.



Community Detection in the Synaptic Interactome



Network Example - Disease Comparison



Geodesic distance (d):

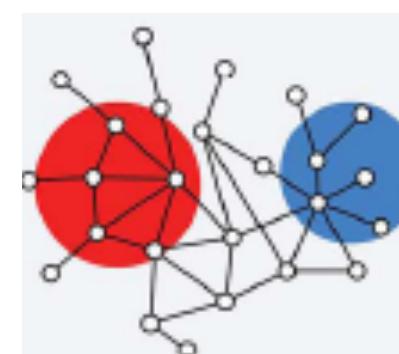
- average shortest path between node pairs.
- $\langle d_{AA} \rangle$ with-disease average.
- $\langle d_{AB} \rangle$ between-disease average.

Disease co-occurrence:

$$S_{AB} = \langle d_{AB} \rangle - \frac{\langle d_{AA} + d_{BB} \rangle}{2}$$

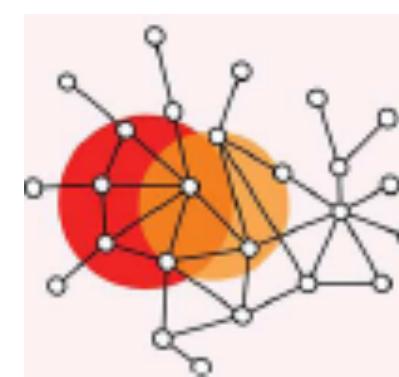
$$S_{AB} \geq 0$$

Separated disease pair



$$S_{AB} < 0$$

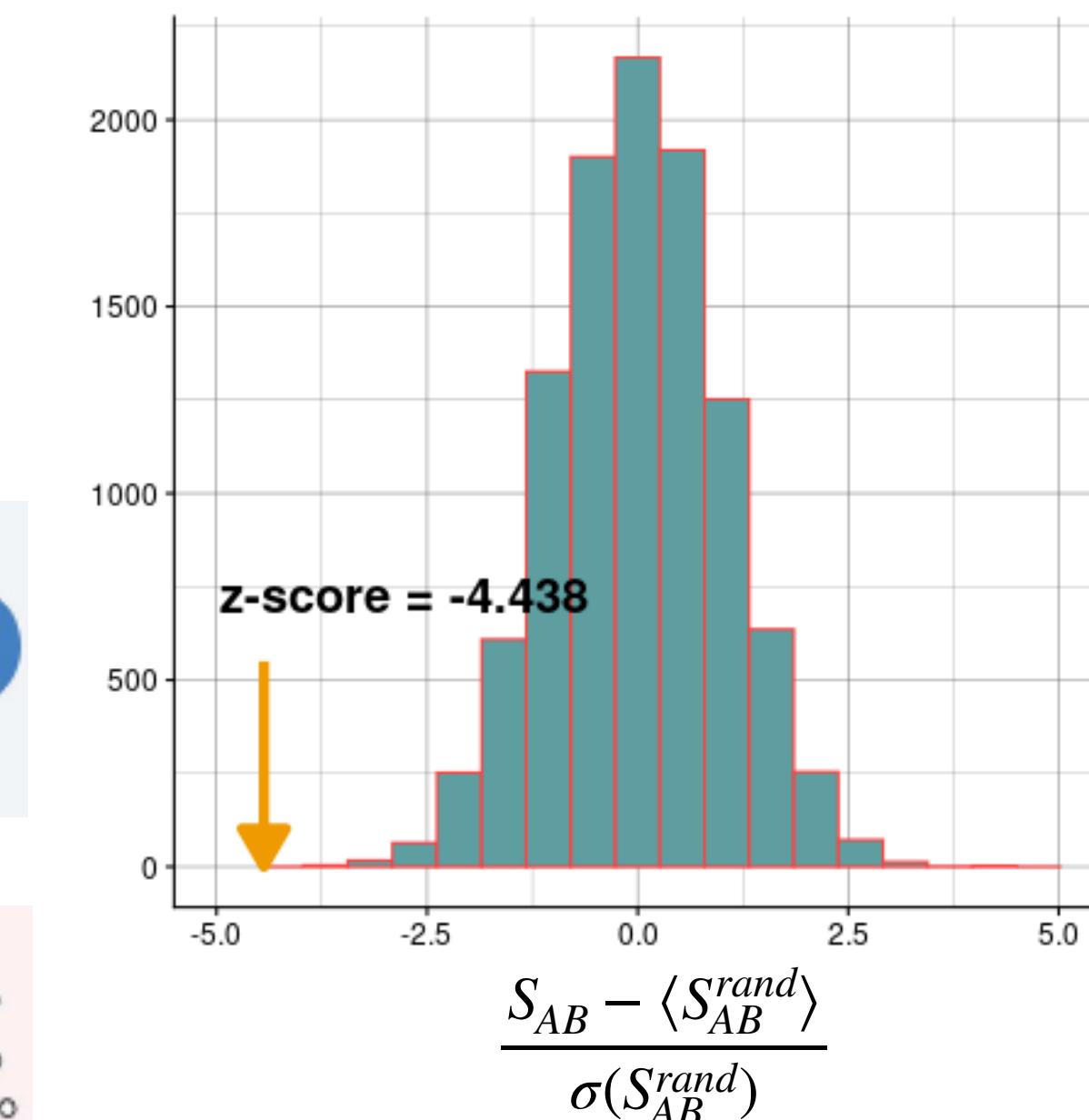
Overlap disease pair



Randomised model:

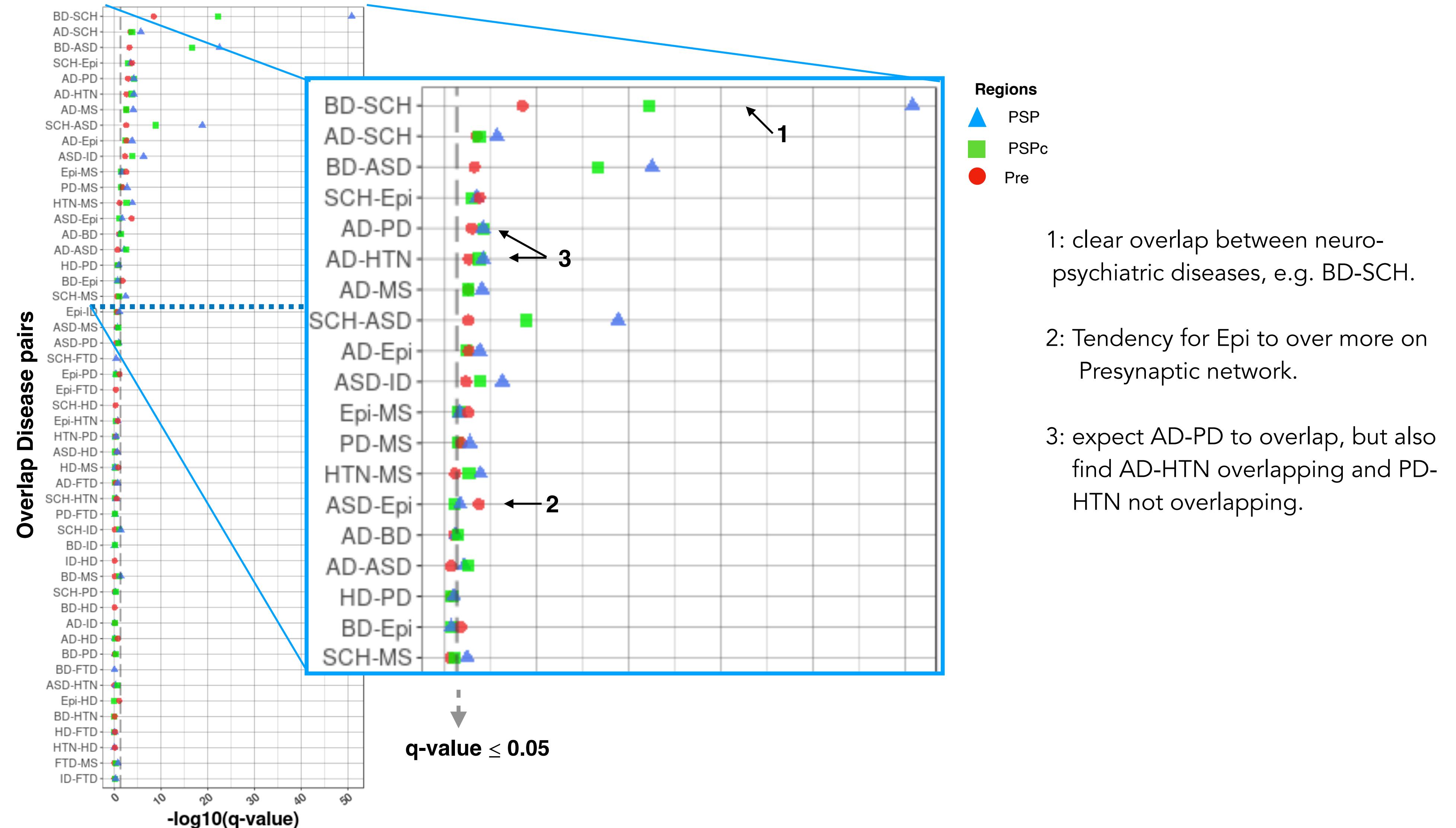
- place each diseases GDA's at random compute S_{AB}^{rand} .
- repeat 10,000.

Alzheimer's disease Vs Parkinson's disease



(Menche et al., 2015)

Network Example - Disease Comparison



Predicting Autism Genes Using Differential RNA-seq Expression Data

SFARI Gene is a database for the autism research community that is centred on genes implicated in autism susceptibility.



RNA-seq Datasets Comparing ASD affected and unaffected individuals

Science

Current Issue First release papers Archive About Submit

HOME > SCIENCE > VOL. 362, NO. 6420 > TRANSCRIPTOME-WIDE ISOFORM-LEVEL DYSREGULATION IN ASD, SCHIZOPHRENIA, AND BIPOLAR DISORDER

SPECIAL ISSUE RESEARCH ARTICLE

f t in g m e

Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder

MICHAEL J. GANDAL, PAN ZHANG, EVI HADJIMICHAEL, REBECCA L. WALKER, CHAO CHEN, SHUANG LIU, HYEJUNG WON, HARM VAN BAKEL, MERINA VARGHESE, DANIEL H. GESCHWIND, +27 authors Authors Info & Affiliations

SCIENCE • 14 Dec 2018 • Vol 362, Issue 6420 • DOI: 10.1126/science.aat8122

nature communications

Explore content About the journal Publish with us

nature > nature communications > articles > article

Open Access | Published: 10 December 2014

Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism

Simone Gupta, Shannon E. Ellis, Foram N. Ashar, Anna Moes, Joel S. Bader, Jianan Zhan, Andrew B. West & Dan

E. Arking

Nature Communications 5, Article number: 5748 (2014) | Cite this article

26k Accesses | 294 Citations | 118 Altmetric | Metrics

nature > translational psychiatry > original article > article

Open Access | Published: 09 May 2017

Altered expression of histamine signaling genes in autism spectrum disorder

C Wright, J H Shin, A Rajpurohit, A Deep-Soboslay, L Collado-Torres, N J Brandon, T M Hyde, J E Kleinman, A E Jaffe, A J Cross & D R Weinberger

Translational Psychiatry 7, e1126 (2017) | Cite this article

4512 Accesses | 56 Citations | 19 Altmetric | Metrics

Wright, C. et al. Altered expression of histamine signaling genes in autism spectrum disorder. *Transl Psychiatr* 7, e1126–e1126 (2017).

Gupta, S. et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun* 5, 5748 (2014)

Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* 362, (2018).

Gandal - 53 ASD, 24 unaffected

Gupta - 32 ASD, 42 unaffected

Wright - 13 ASD, 39 unaffected

Weighted Gene Correlation Network Analysis (WGCNA)

BMC Bioinformatics



Software

Open Access

WGCNA: an R package for weighted correlation network analysis

Peter Langfelder¹ and Steve Horvath^{*2}

Address: ¹Department of Human Genetics, University of California, Los Angeles, CA 90095, USA and ²Department of Human Genetics and Department of Biostatistics, University of California, Los Angeles, CA 90095, USA

Email: Peter Langfelder - Peter.Langfelder@gmail.com; Steve Horvath^{*} - shorvath@mednet.ucla.edu

* Corresponding author

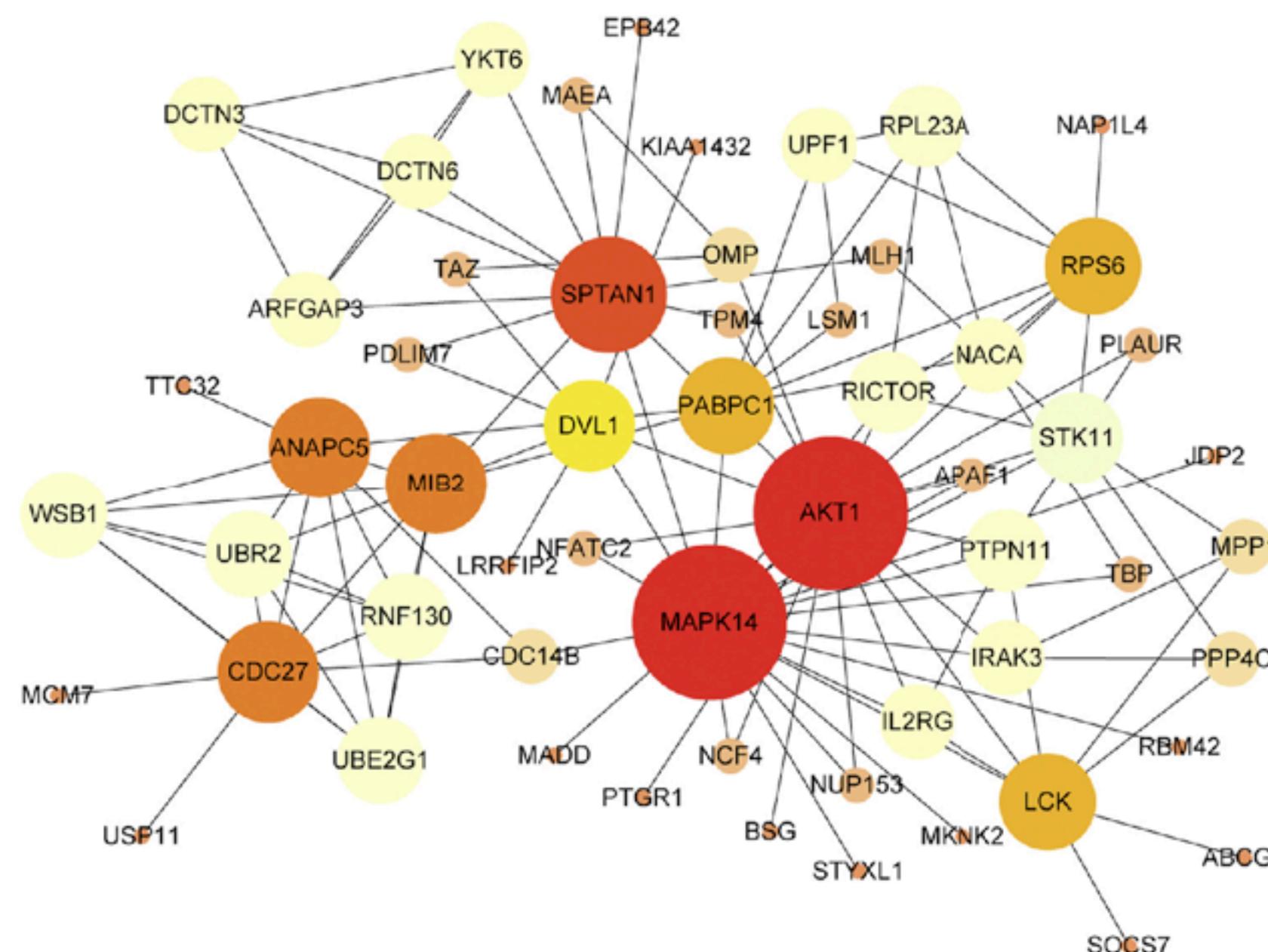
Published: 29 December 2008

Received: 24 July 2008

Accepted: 29 December 2008

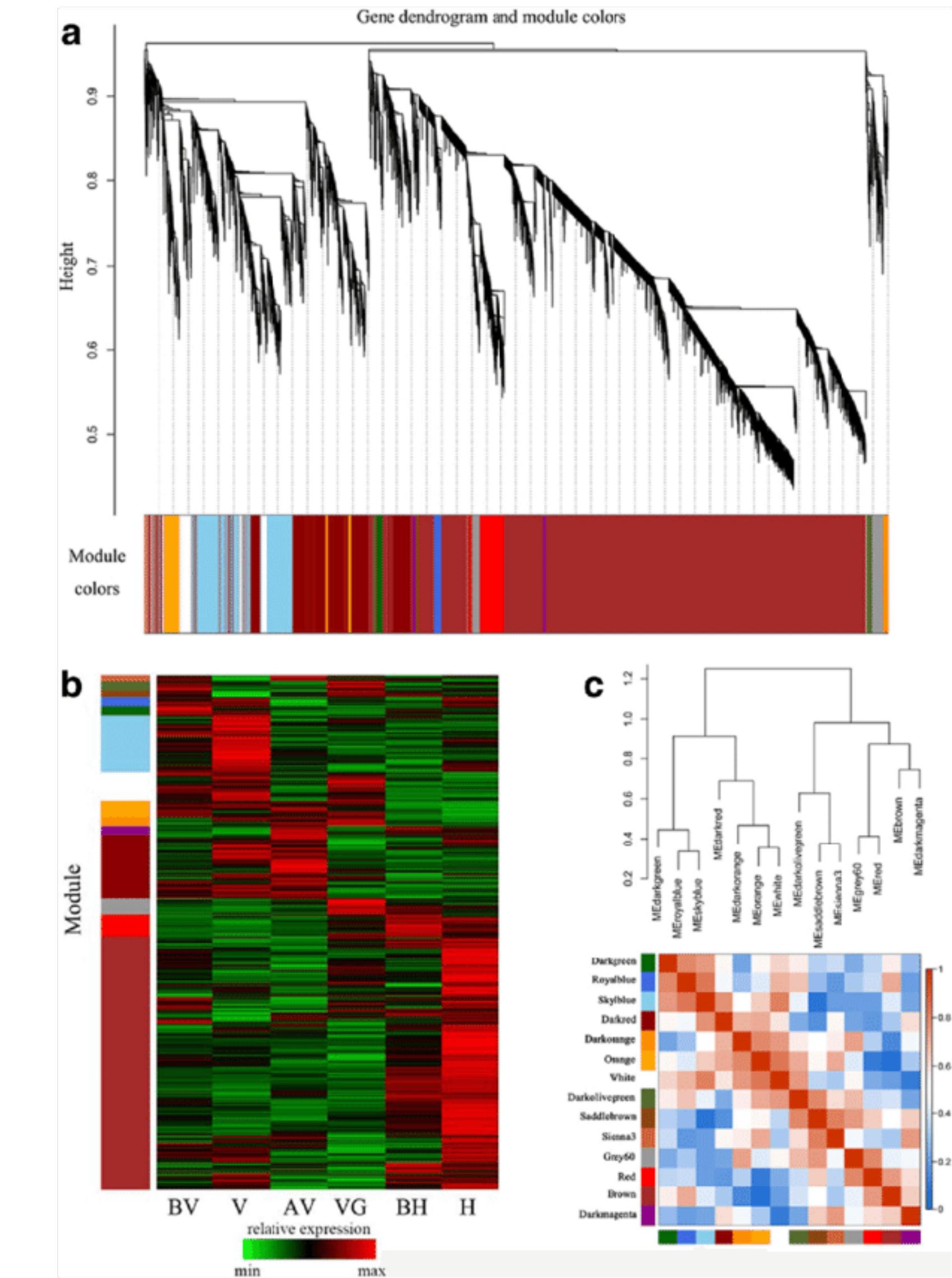
BMC Bioinformatics 2008, 9:559 doi:10.1186/1471-2105-9-559

This article is available from: <http://www.biomedcentral.com/1471-2105/9/559>



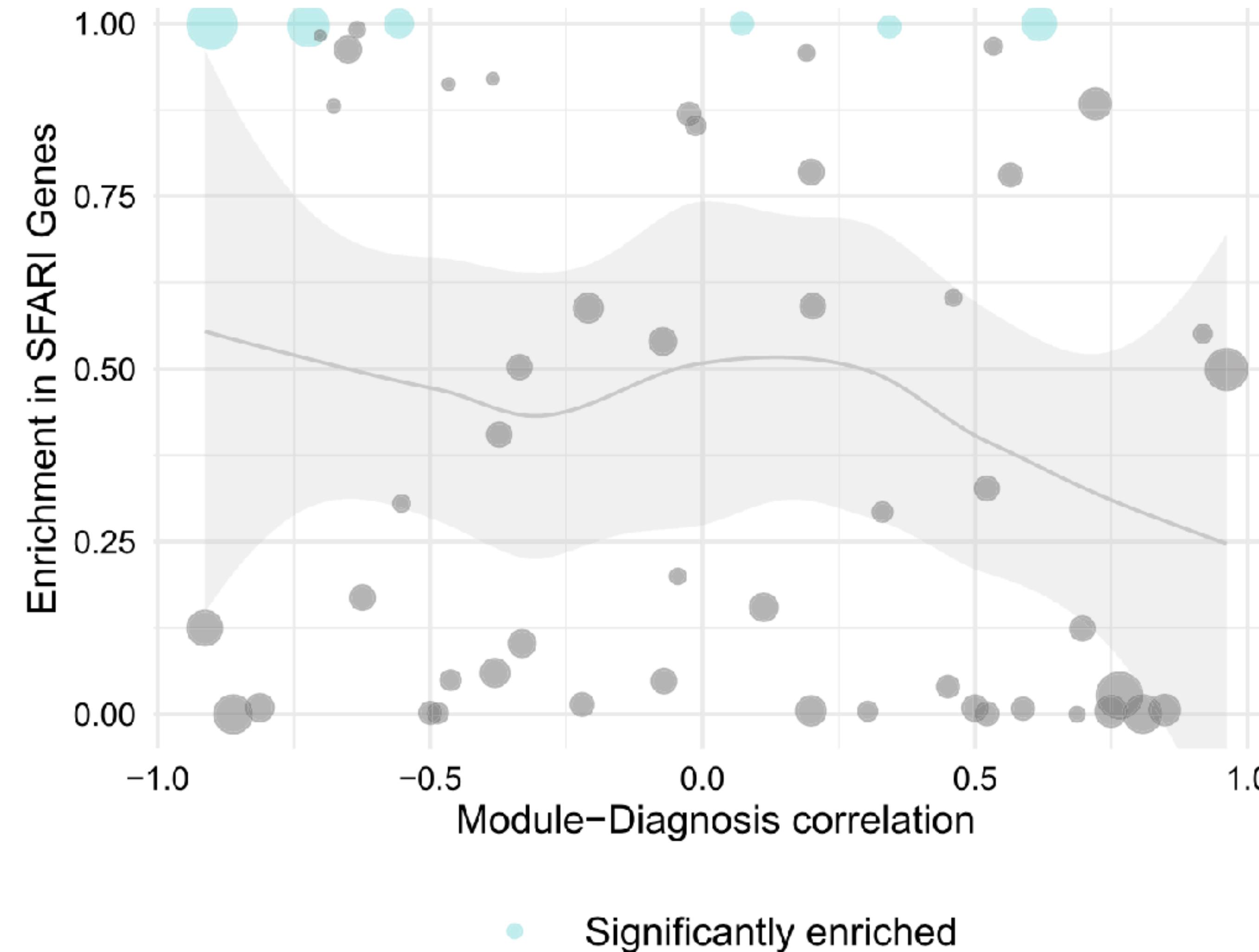
<https://homepages.inf.ed.ac.uk/tsimpson/>

ian.simpson@ed.ac.uk



<https://github.com/biomedicalinformaticsgroup>

Network Modules Correlated with ASD Diagnosis are not Enriched in ASD Genes



Arpi, M. N. T. & Simpson, T. I. SFARI genes and where to find them; modelling Autism Spectrum Disorder specific gene expression dysregulation with RNA-seq data. *Scientific Reports* 12, 10158 (2022)

Considering the Whole Network Topology Reveals Novel ASD Candidate Genes

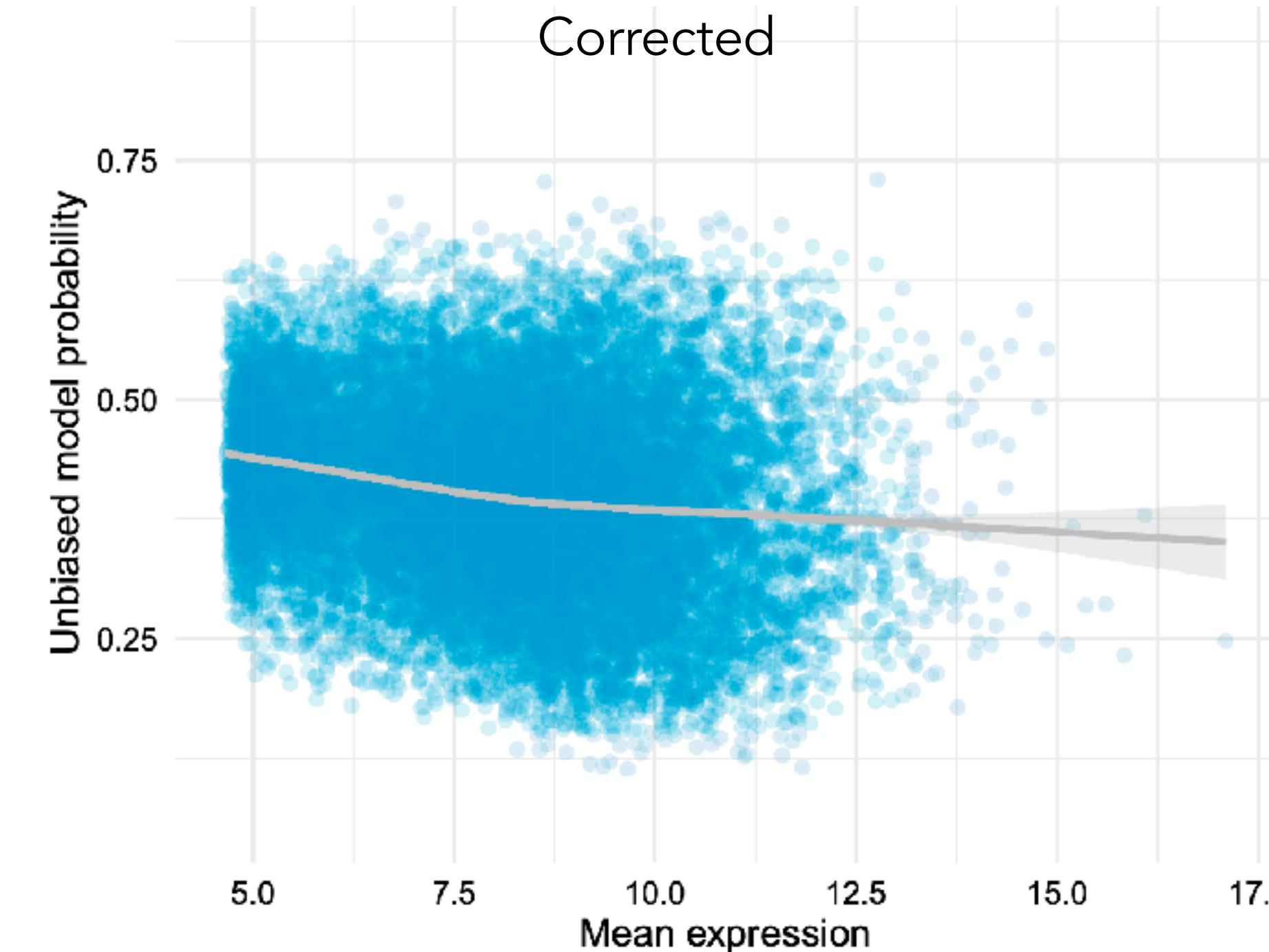
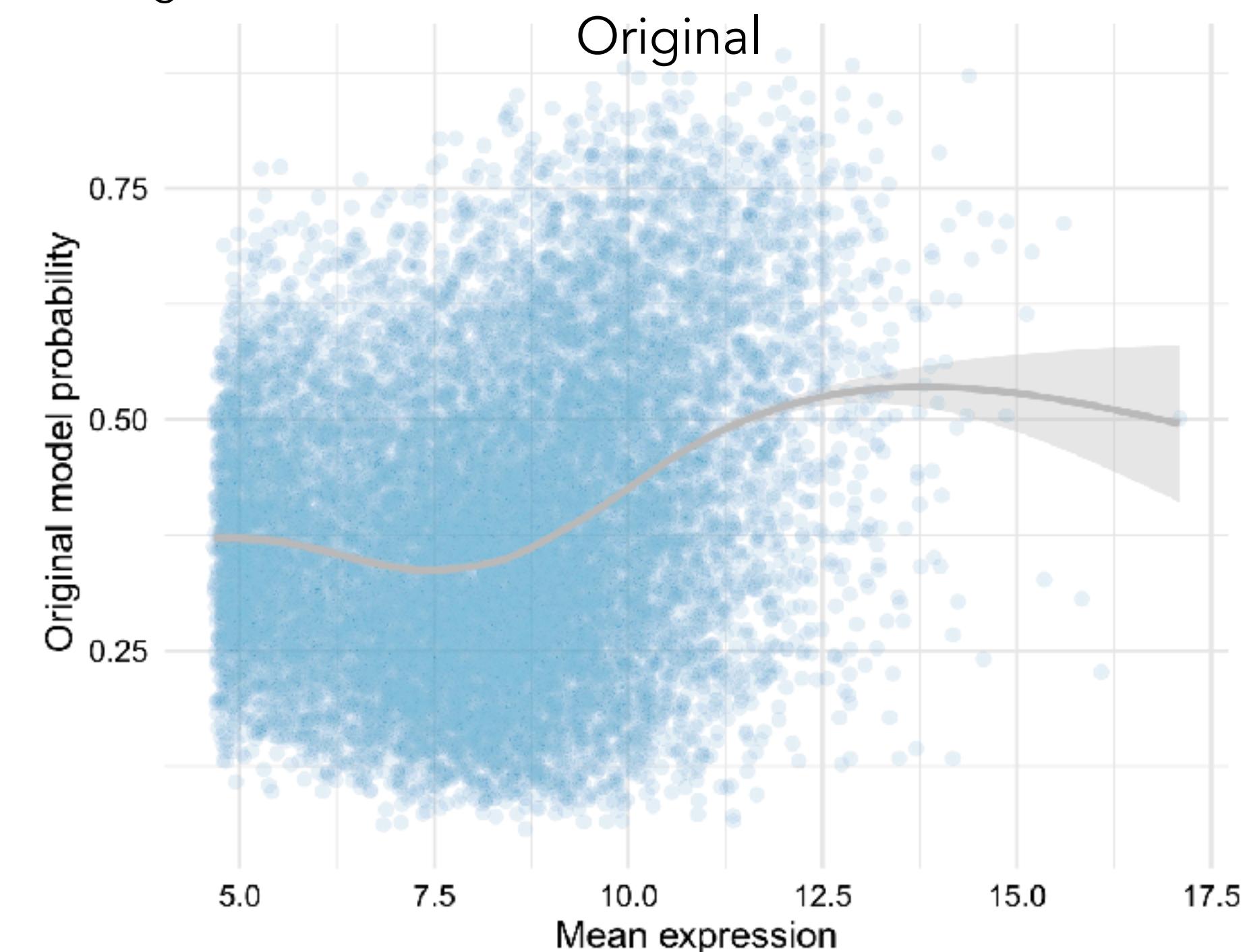
Using only “local” network information (module level) is not sufficient, we need to use features of the whole co-expression network to capture the signal

Model Features

The correlation between a gene’s expression pattern and i) diagnosis, ii) each module’s eigen-gene, iii) between a gene’s assigned network module and diagnosis, and iv) a binary objective variable indicating whether the gene is a SFARI list gene.

Model Performance

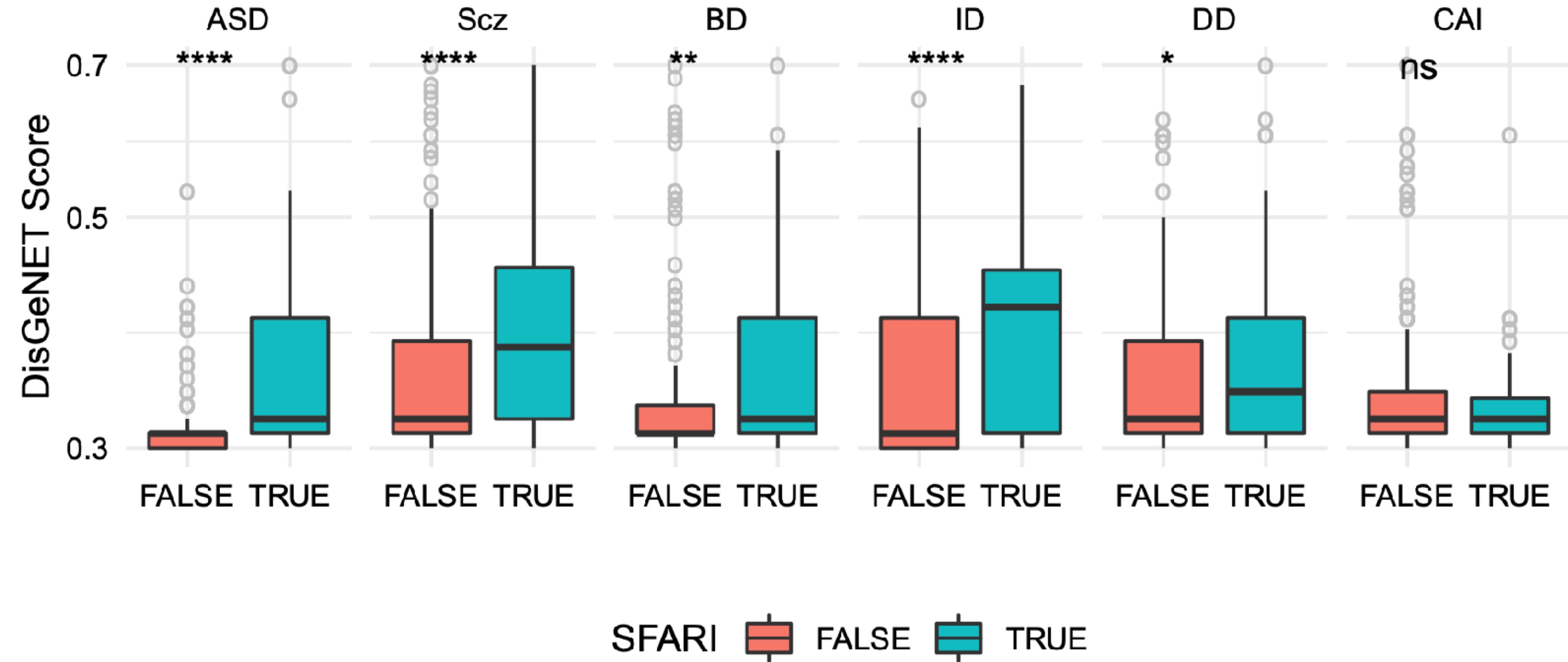
Model	AUC	MLP	Balanced accuracy
Original	$0.69 \pm 4 \times 10^{-4}$	20.43 ± 0	0.64 ± 0.0018
Unbiased	0.58 ± 0.03	13.61 ± 6.84	0.56 ± 0.01
Shuffled labels	0.50 ± 0.02	2.83 ± 3.4	0.50 ± 0.01



Top10 Predicted ASD (SFARI-like) Genes with the Corrected Whole Network Model

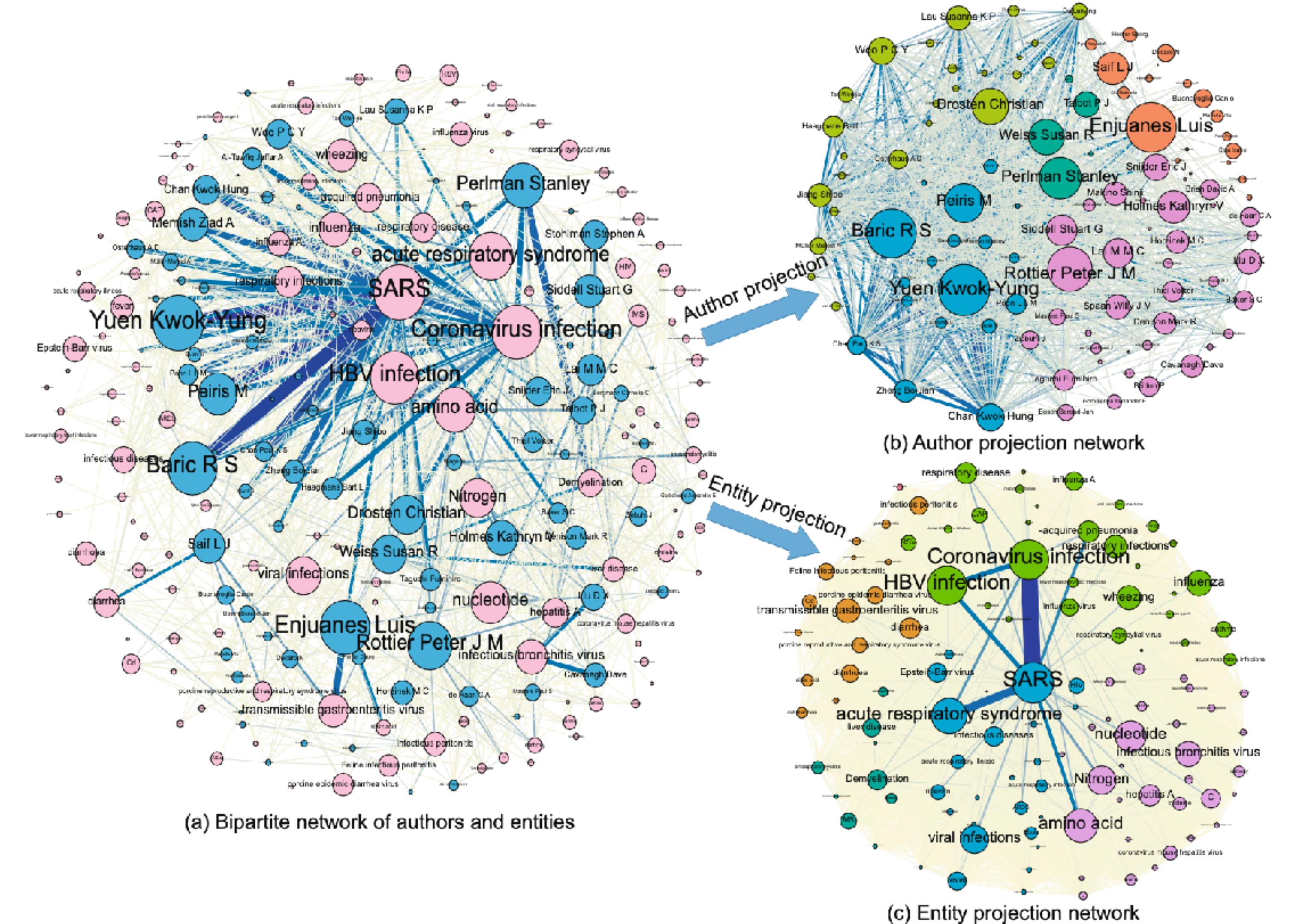
	Gene	Probability	Literature review
1	SNX25	0.73	CNV associated both to ASD and ADHD ⁴⁴
2	CLMP	0.71	QTN associated to play skills in twins with ASD ⁴⁵
3	EGR1	0.70	Role in the aberrant regulation of synaptic maturation in ASD ⁴⁶
4	HECTD2	0.69	Phylogenetically similar to UBE3A (SFARI Gene Score 1) ⁴⁷
5	PLXNC1	0.69	Part of the Axonal Guidance signaling pathway, one of the canonical pathways significantly associated with dysregulated genes with LINE-1 insertion ⁴⁸
6	AHI1	0.69	Mutations associated to ASD ⁴⁹
7	CORO1A	0.69	Now a SFARI Gene with Score 1 in the latest version of the dataset ⁵
8	ARC	0.68	Target protein of gene UBE3A (SFARI Gene Score 1) ⁵⁰
9	ARPP21	0.68	Gene associated to candidate intergenic risk loci in ASD ⁵¹
10	ARHGAP20	0.68	Differential expression related to ASD ⁵²

Is there a Shared Aetiology Between ASD and Other Neurological Disorders?



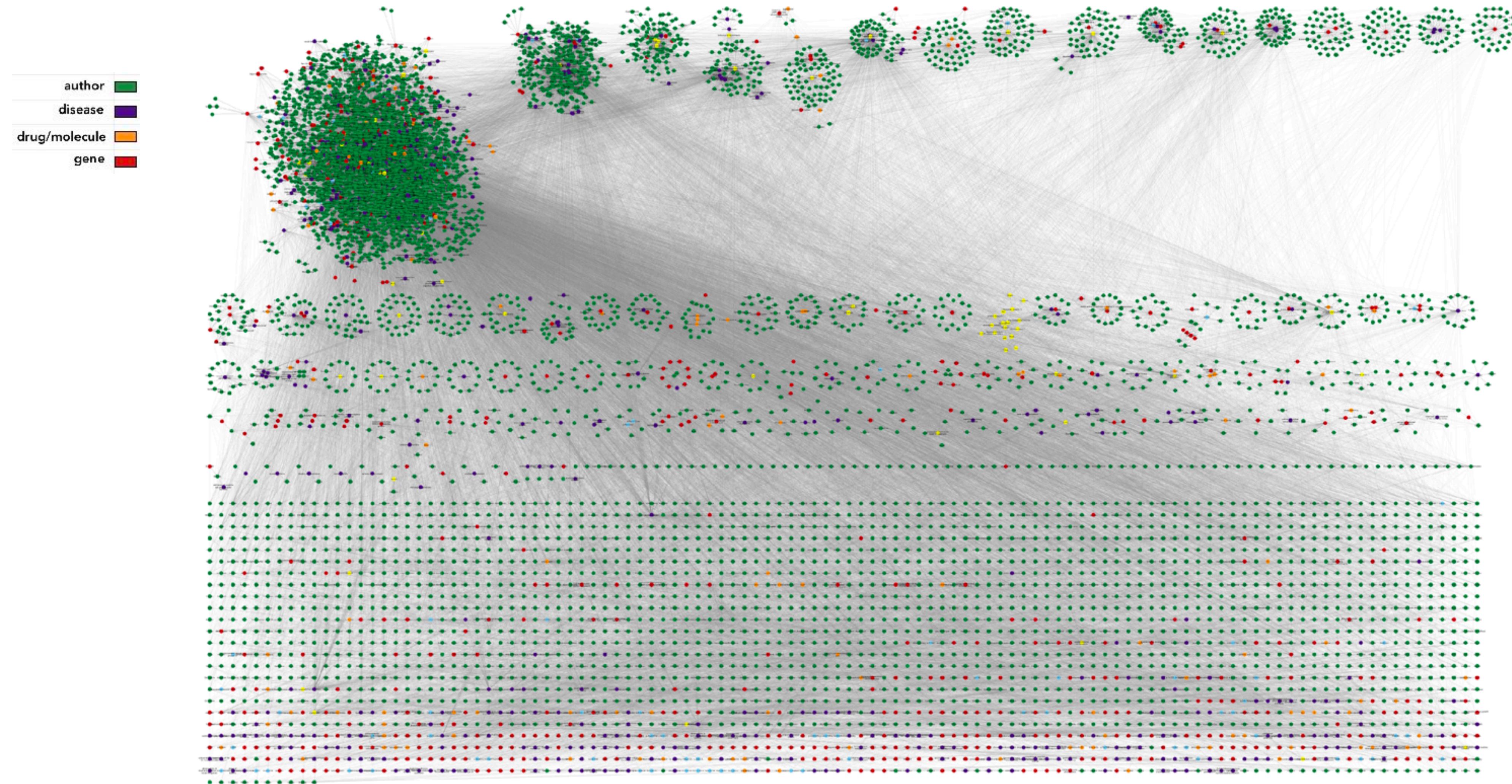
Arpi, M. N. T. & Simpson, T. I. SFARI genes and where to find them; modelling Autism Spectrum Disorder specific gene expression dysregulation with RNA-seq data. Scientific Reports 12, 10158 (2022)

Building a Literature Driven Knowledgegraph for Autism Spectrum Disorders



after Xu, J. et al. Building a PubMed knowledge graph. Sci Data 7, 205 (2020).

ASD BioEntity Network



BioBERT - 1353 unique concepts

UoE-UMLSModel - 50,000 unique concepts

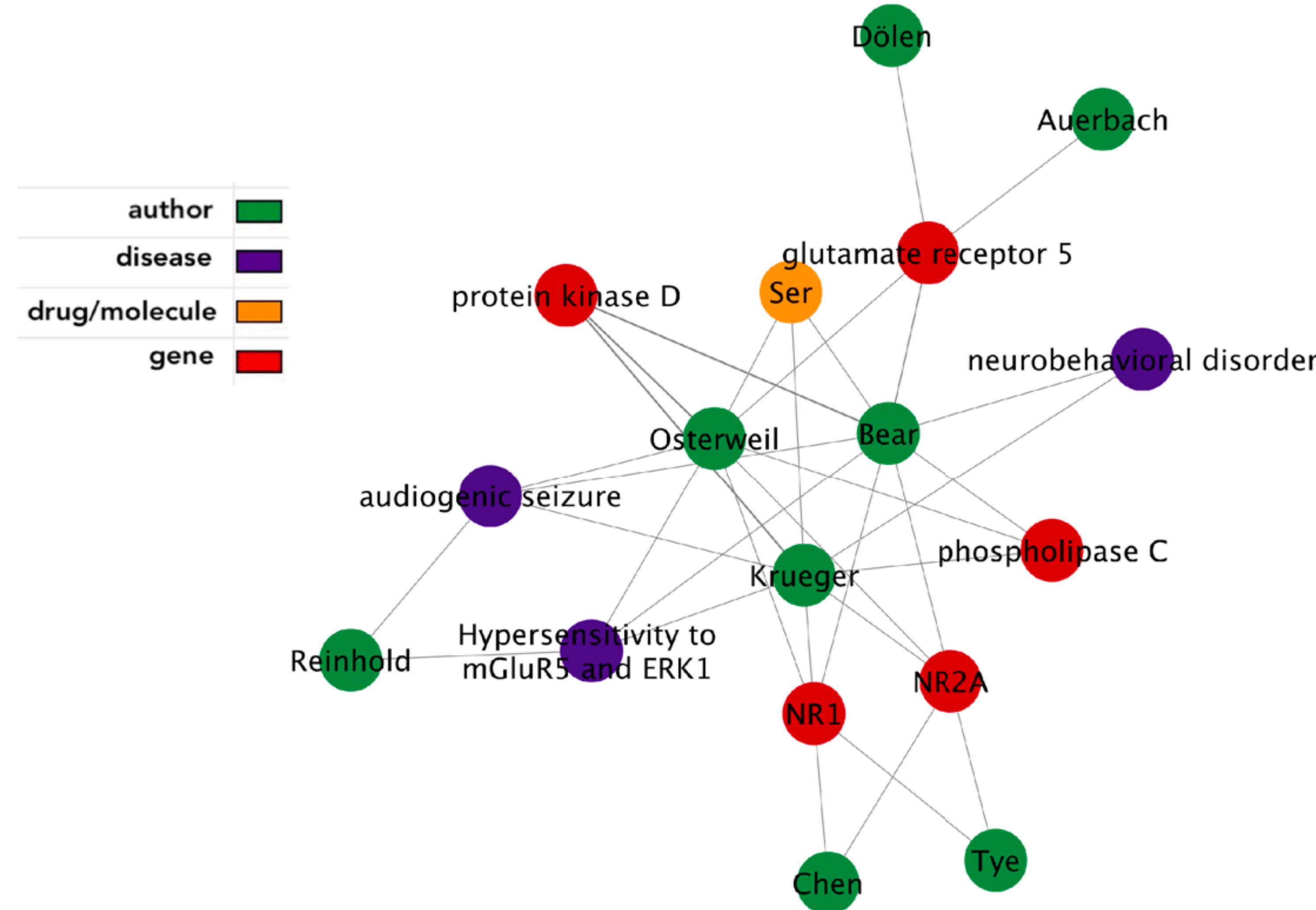
(1673 HPO terms alone)

<https://homepages.inf.ed.ac.uk/tsimpson/>

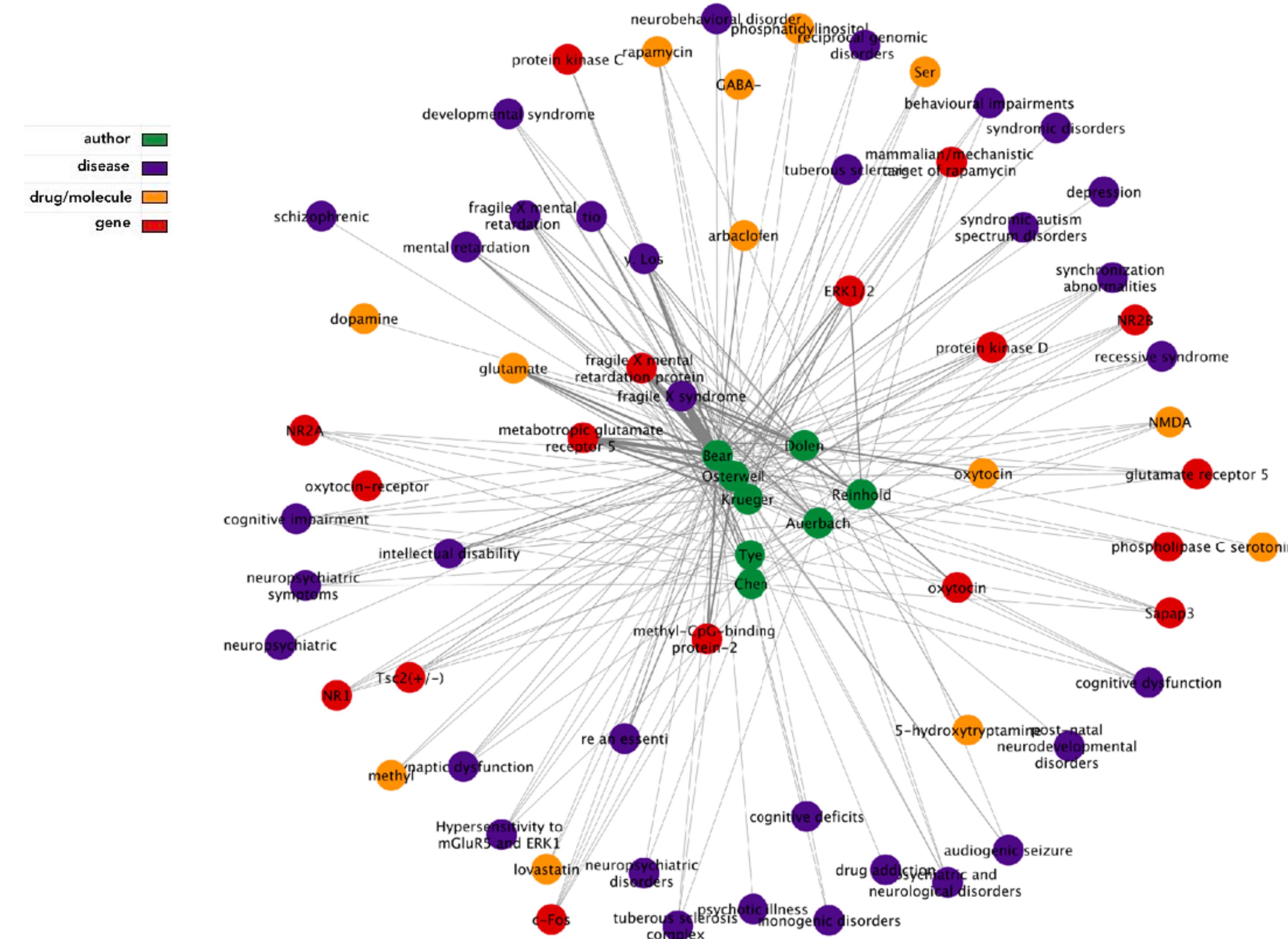
ian.simpson@ed.ac.uk

<https://github.com/biomedicalinformaticsgroup>

Author Entity Cluster



Author Entity Cluster Neighbourhood



Multi-modal Data for Autism Spectrum Disorders

SSC Collection Overview

Families: 2644

Individuals: 10474

Female probands: 352

Male probands: 2292

Simons Searchlight

155 genes, 23 CNVs

>1500 individuals



SFARI GeneBase

Scored Genes: 1028

ASD linked genes: 969

Animal Models: 276 genes

CNVs: 17



Phenotypic data from **283,520** individuals (including more than **100,000** individuals with ASD)

94,116 children (<18y) with ASD

17,604 adults with ASD

83,897 male pro-bands

27,973 female pro-bands

43,158 unaffected siblings

7,074 enrolled twins, triplets and quadruplets

16,951 multiplex families

Types of data include basic medical screening, Developmental Coordination, Repetitive Behaviour, Social Communication, Adaptive Behaviour.

11/02/22



Open Access Patient Records: **40,544**

Phenotype Observations: 173,565

CNVs 41,679

DDG2P

Unique Genes: **2132**

Diseases: 2161

With Phenotype: 1550



Literature corpora for Autism Spectrum Disorder

OA and UoE licensed for Data Science

Full-Text Retrieval HTML, XML, & PDF

68,329 target papers 04/02/22 (**54,591** - 86% retrieved)

SPARK Recruitment Data Gathering

DCDQ - 17 questions with 5-point Likert scale

[scores are 0 for "not at all like this child to 5 for "extremely like this child"]

RBS-R - 43 questions 4-point Likert scale

[scores are from 0 "behaviour doesn't occur" to 3 "behaviour occurs and is a severe problem"]

SCQ - 40 yes/no questions with 0 = 0 absence of abnormal behaviour and 1 presence.

[SCQ can be asked as current or lifetime, the latter being better in very young children]

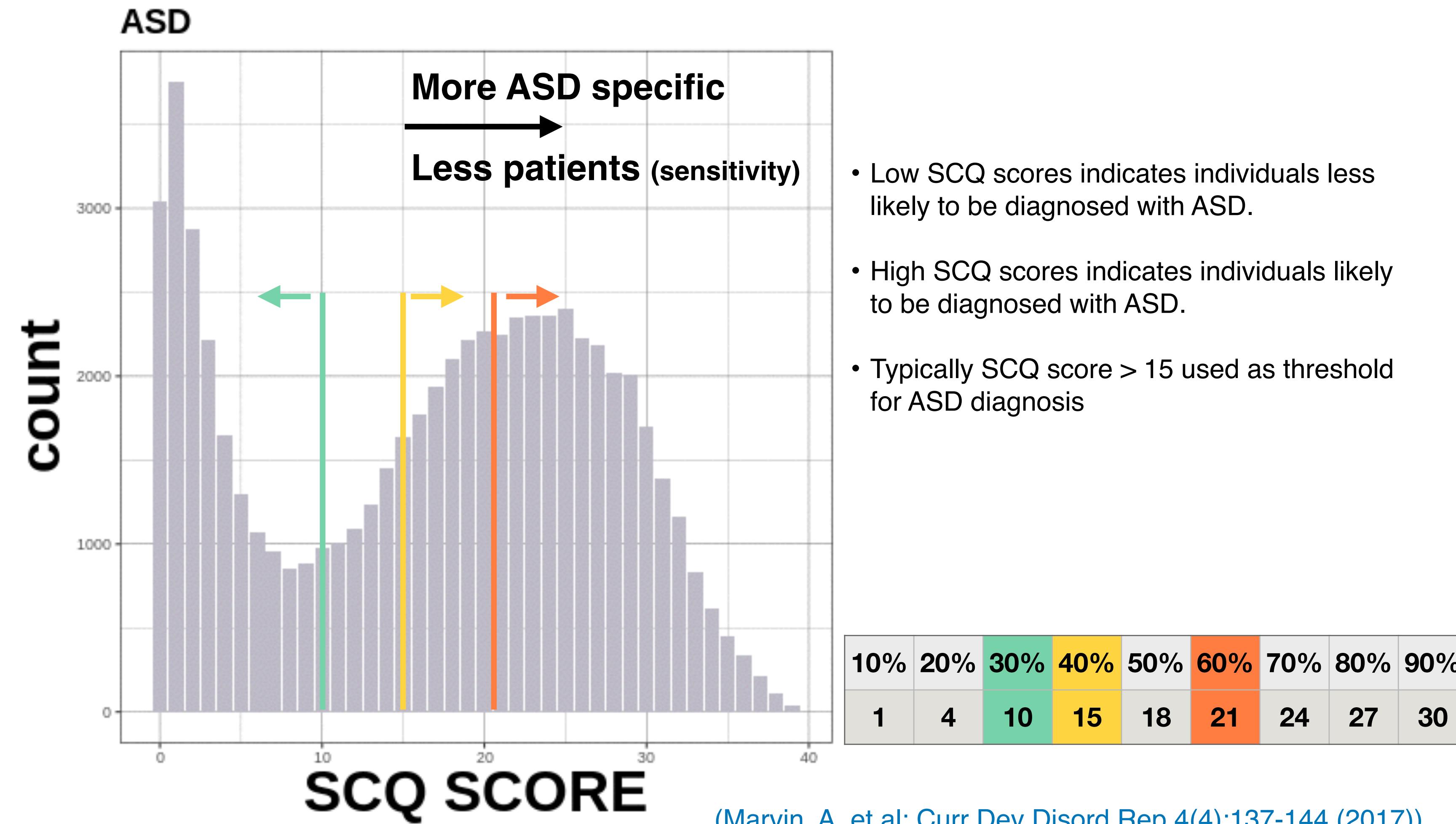
Wilson, B. N., Kaplan, B. J., Crawford, S. G., Campbell, A. & Dewey, D. Reliability and Validity of a Parent Questionnaire on Childhood Motor Skills. *Am J Occup Ther* 54, 484–493 (2000).

Lam, K. S. L. & Aman, M. G. The Repetitive Behavior Scale-Revised: Independent Validation in Individuals with Autism Spectrum Disorders. *J Autism Dev Disord* 37, 855–866 (2007).

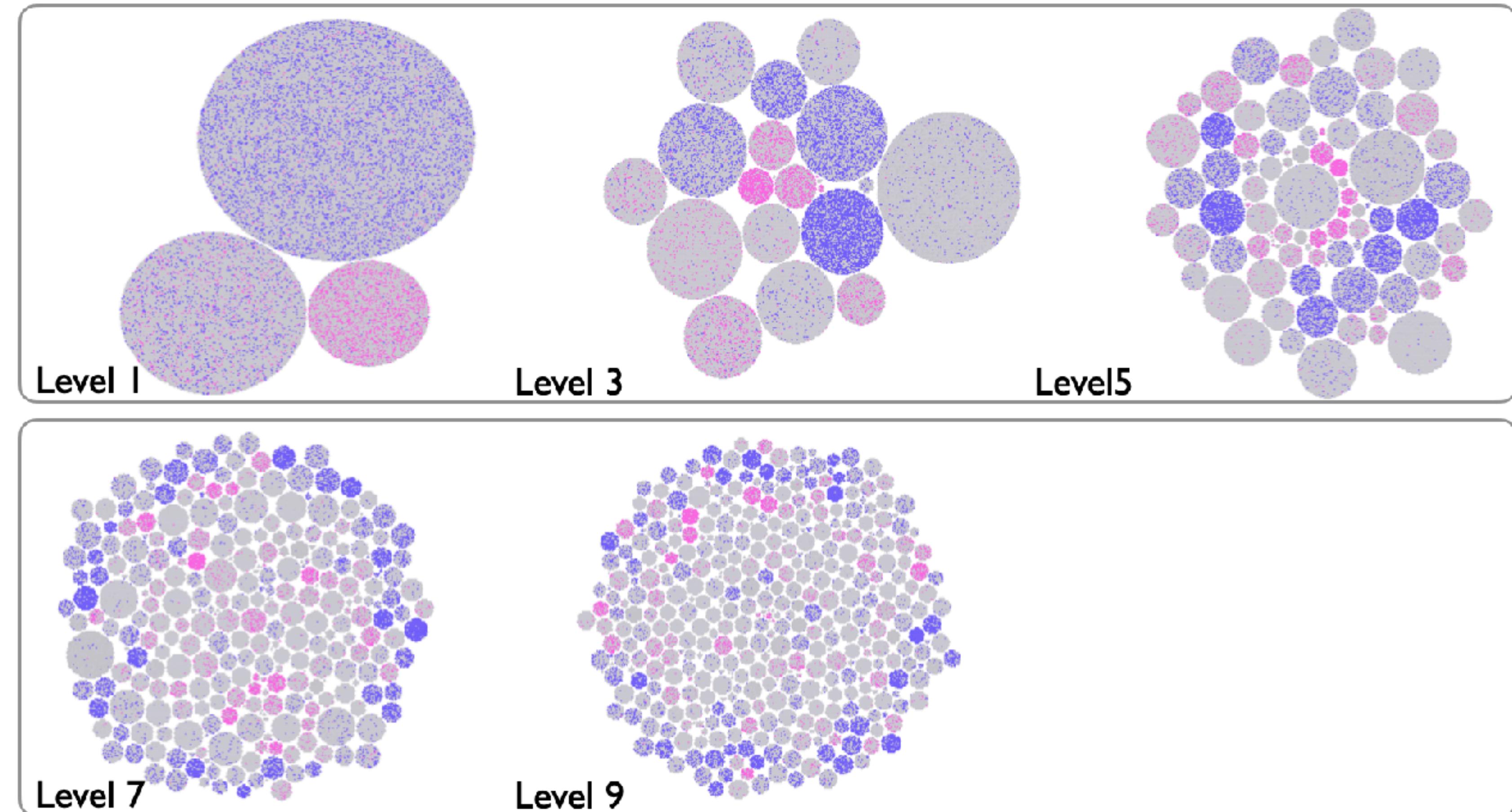
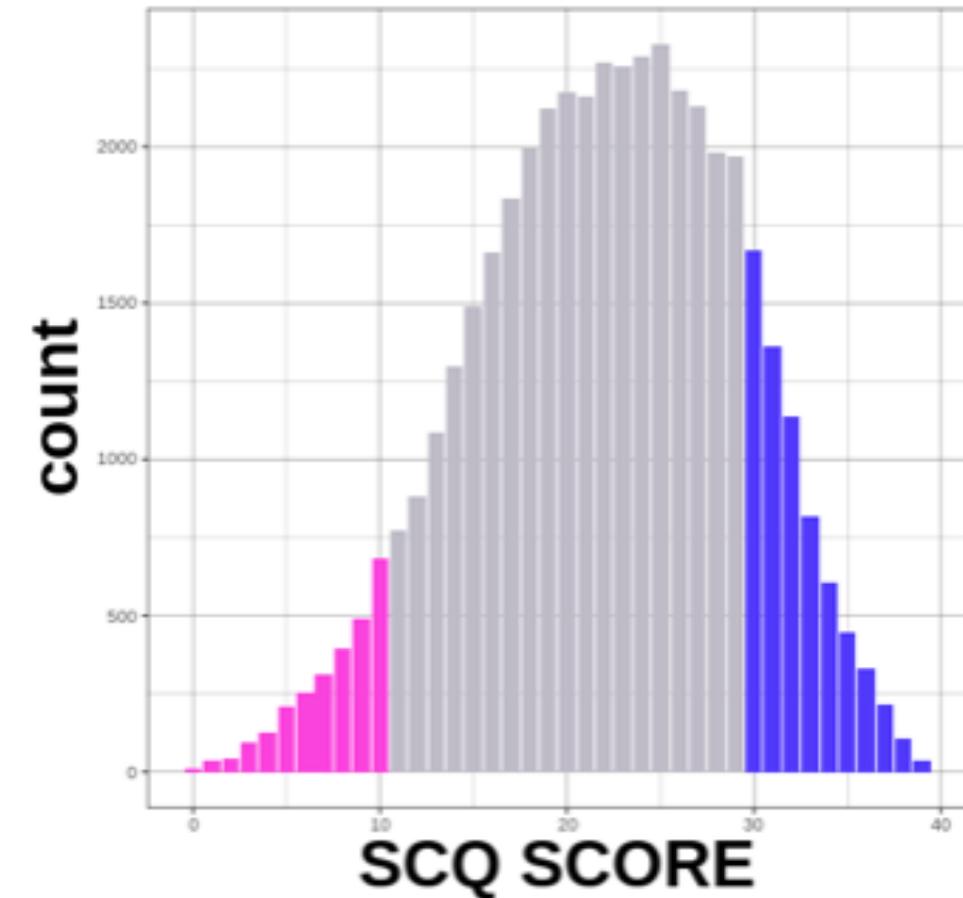
Marvin, A. R., Marvin, D. J., Lipkin, P. H. & Law, J. K. Analysis of Social Communication Questionnaire (SCQ) Screening for Children Less Than Age 4. *Curr Dev Disord Reports* 4, 137–144 (2017).

Instrument	Parents	Individuals with ASD	Unaffected Siblings
Background History Questionnaire-Adult Measure of: Demographics and social history Format: Online Questionnaire Completed by: Independent adults with ASD Regarding: Independent ASD adults over 18 years Author: SFARI		x	
Background History Questionnaire-Child/Dependent Measure of: Demographics and developmental history Format: Online Questionnaire Completed by: Parent/guardian Regarding: Dependent ASD individuals Author: SFARI		x	
Background History Questionnaire-Sibling Measure of: Demographics and developmental history Format: Online Questionnaire Completed by: Parent/guardian Regarding: Non-ASD siblings under 18 years Author: SFARI			x
Basic Medical Screening Questionnaire Measure of: Medical history Format: Online Questionnaire Completed by: Parent/guardian or self Regarding: All ASD individuals, their parents and siblings Author: SFARI	x	x	x
Developmental Coordination Disorder Questionnaire Measure of: Motor delays Format: Online Questionnaire Completed by: Parent/guardian Regarding: Dependent ASD individuals age 5–15 years Author: B.N. Wilson		x	
Social Communication Questionnaire (SCQ) — Lifetime Measure of: Screen of ASD markers Format: Online Questionnaire Completed by: Parent/guardian Regarding: ASD individuals and non-ASD siblings ages 2 to 17 years 11 months Publisher: WPS		x	x
Repetitive Behaviors Scale-Revised Measure of: Repetitive behaviors Format: Online Questionnaire Completed by: Parent/guardian Regarding: Dependent ASD individuals age 3 years and up Author: Bodfish		x	

Diagnostic Decisions by SCQ Score



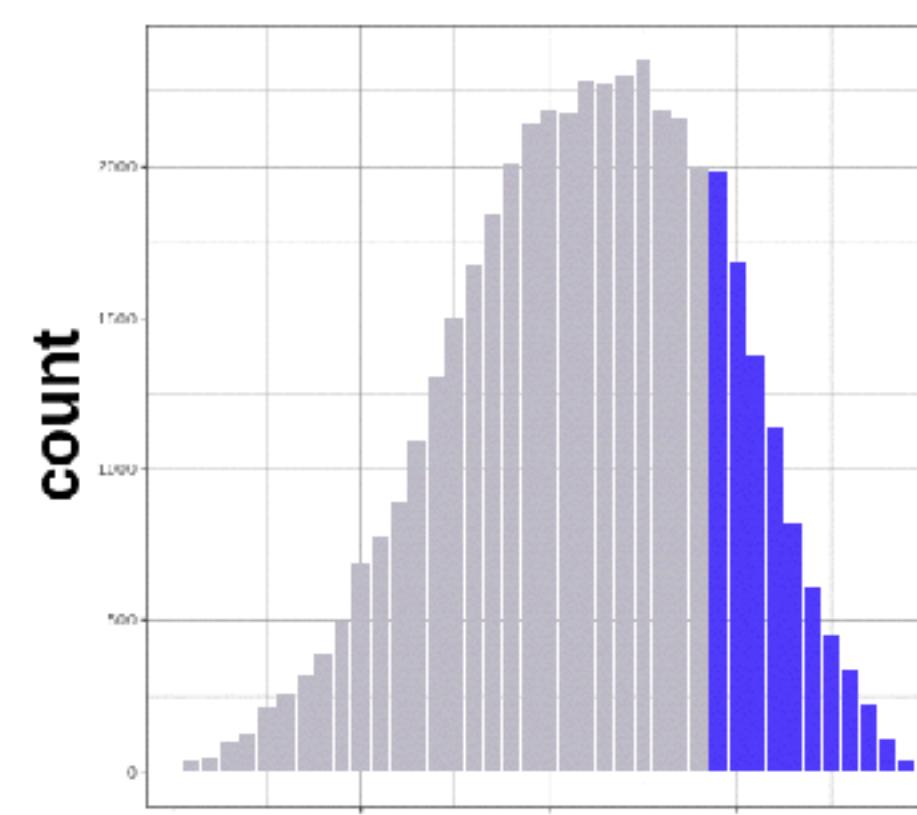
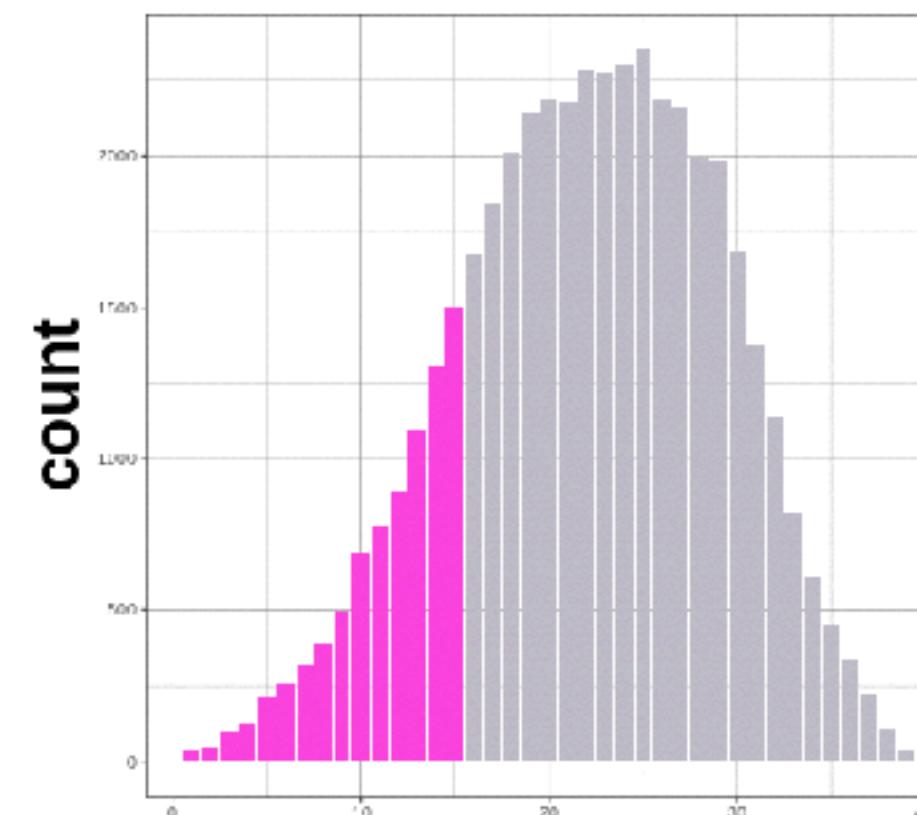
Social Communication Questionnaire Scores



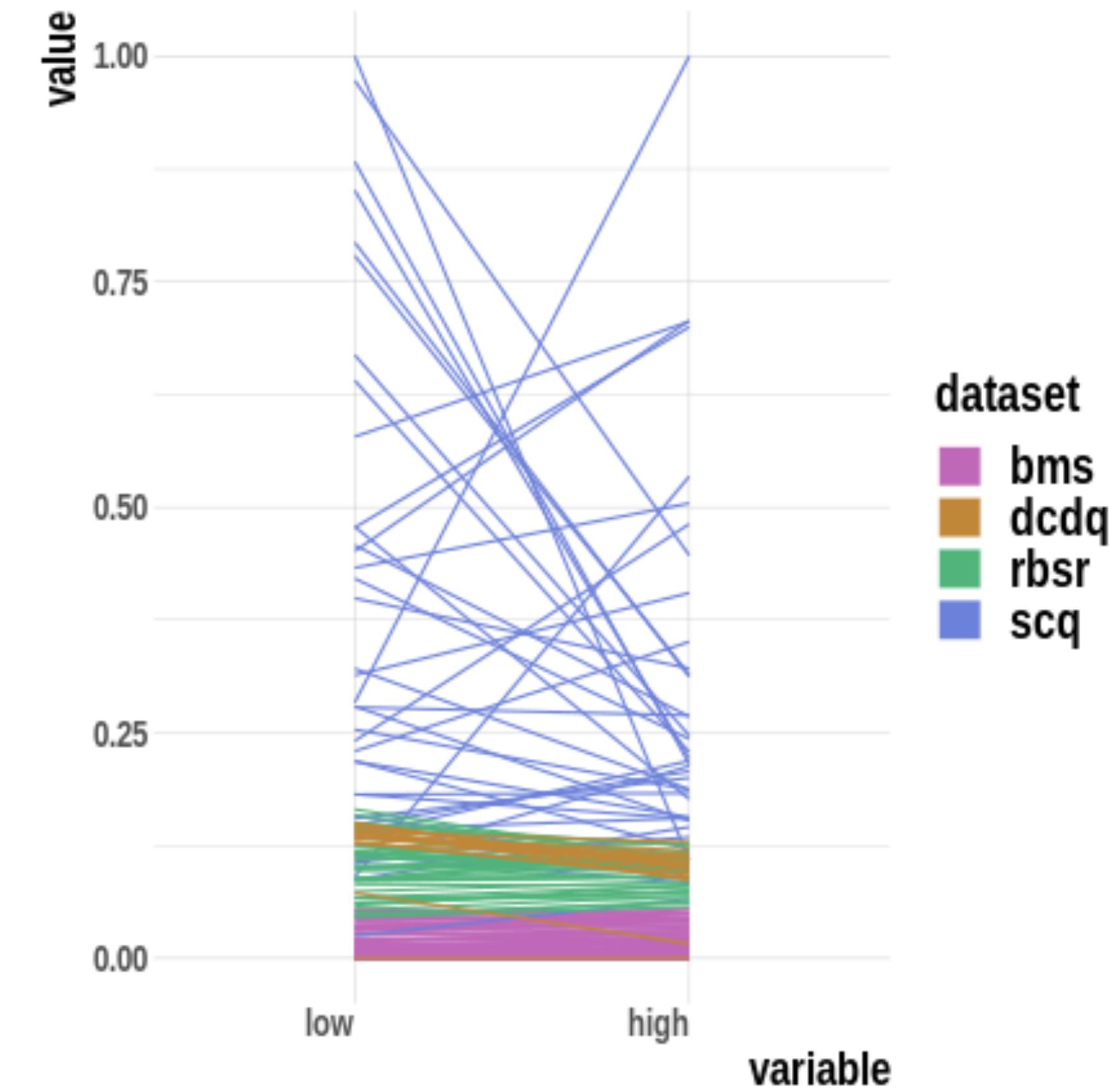
Colin McLean

Social Coordination is the Strongest Predictor of ASD Status in SPARK

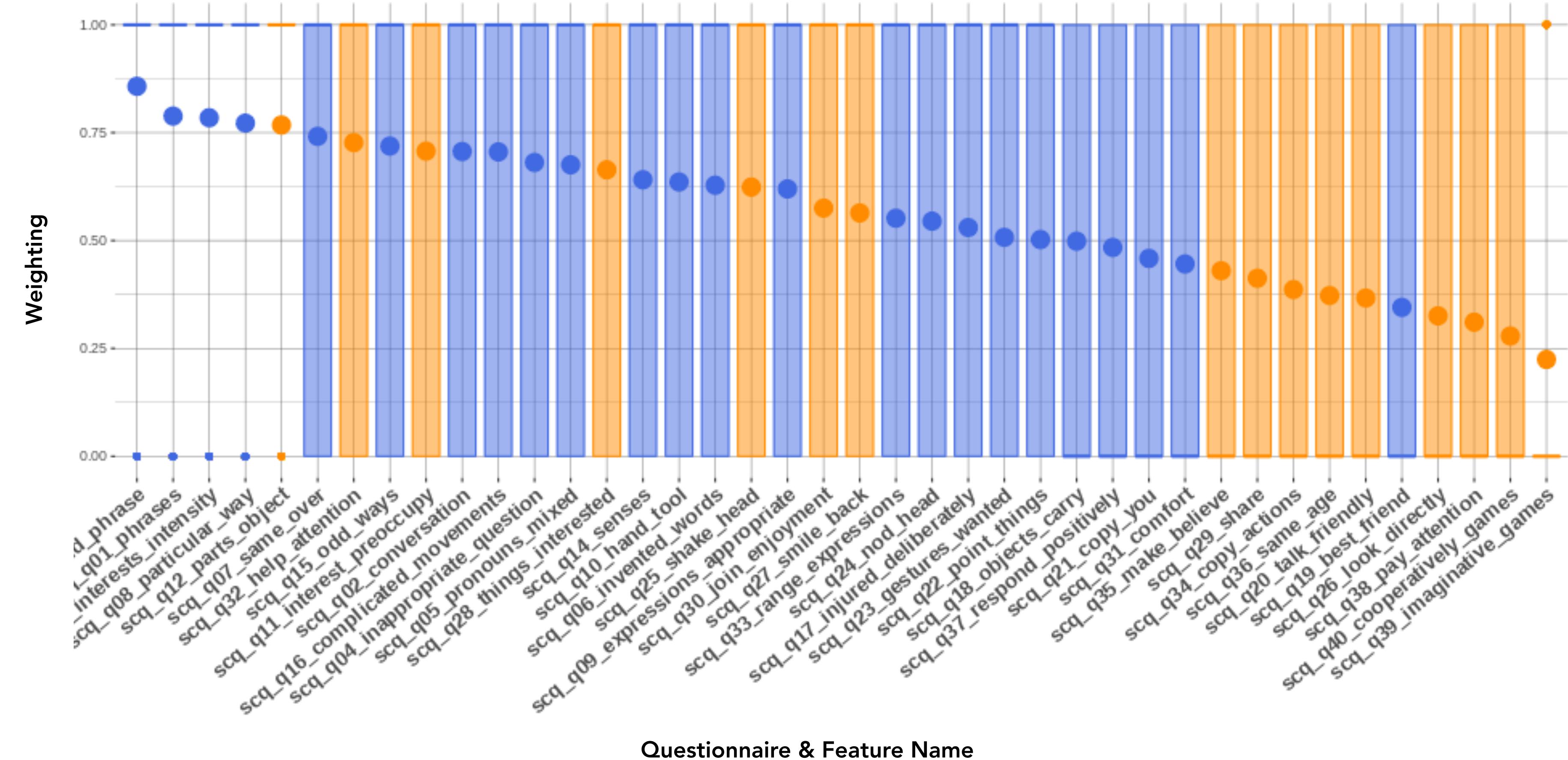
SCQ Scores



Feature Importance



Diagnostic SPARK Features

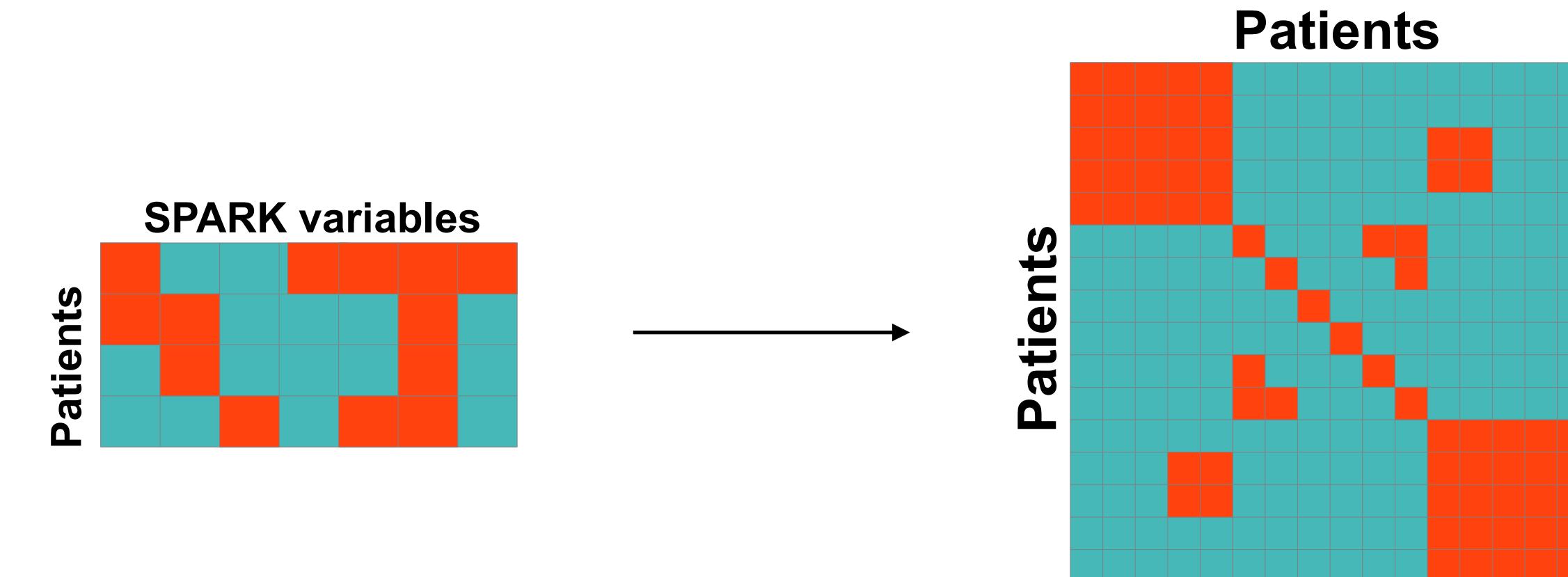


SPARK Feature Guided RF Predicts ASD Status

- Prediction of ASD status using as few as 9 features (out of c.200) used in the parental questionnaires given to parents upon joining the scheme.
- Raises possibility of new low impact approach for ASD triage
- Further work can stratify performance by other factors such as age
- Network clusters can be evaluated using this schema to identify sub-groups of clinical features.
- Establishes joint modelling framework for additional data fusion using genomics (exome and genome/ GWAS - Clustering of genes and variants with genomic features)

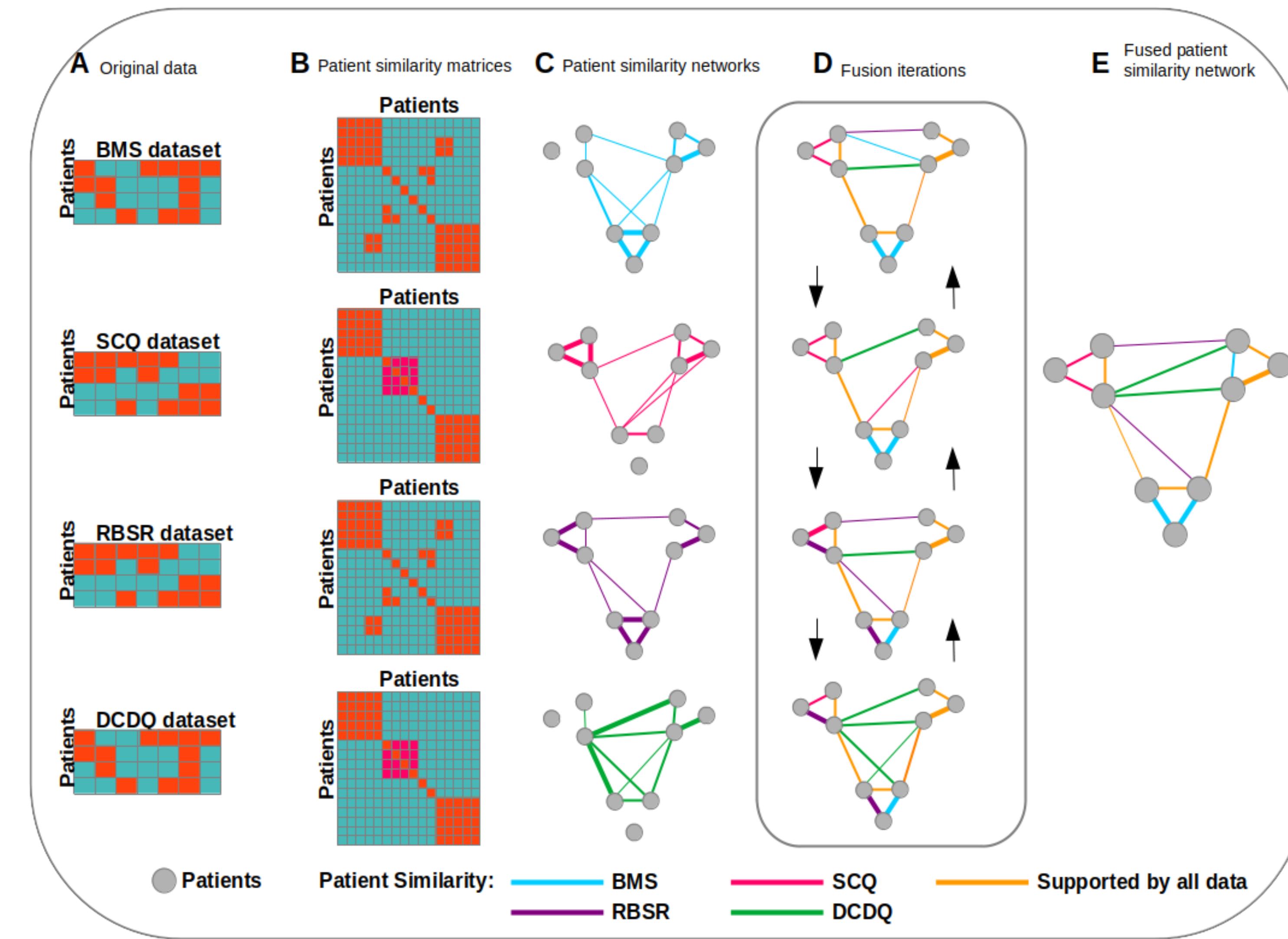
low extreme gamma = 1			low extreme gamma = 0.6			high extreme gamma = 1			high extreme gamma = 0.6																																						
<table border="1"><thead><tr><th></th><th>Joint</th><th>non-joint</th></tr></thead><tbody><tr><td>N Patients</td><td>23990</td><td>31915</td></tr><tr><td>N features</td><td>9</td><td>9</td></tr></tbody></table>				Joint	non-joint	N Patients	23990	31915	N features	9	9	<table border="1"><thead><tr><th></th><th>Joint</th><th>non-joint</th></tr></thead><tbody><tr><td>N Patients</td><td>23990</td><td>31915</td></tr><tr><td>N features</td><td>35</td><td>35</td></tr></tbody></table>				Joint	non-joint	N Patients	23990	31915	N features	35	35	<table border="1"><thead><tr><th></th><th>Joint</th><th>non-joint</th></tr></thead><tbody><tr><td>N Patients</td><td>23990</td><td>31915</td></tr><tr><td>N features</td><td>6</td><td>6</td></tr></tbody></table>				Joint	non-joint	N Patients	23990	31915	N features	6	6	<table border="1"><thead><tr><th></th><th>Joint</th><th>non-joint</th></tr></thead><tbody><tr><td>N Patients</td><td>23990</td><td>31915</td></tr><tr><td>N features</td><td>32</td><td>32</td></tr></tbody></table>				Joint	non-joint	N Patients	23990	31915	N features	32	32
	Joint	non-joint																																													
N Patients	23990	31915																																													
N features	9	9																																													
	Joint	non-joint																																													
N Patients	23990	31915																																													
N features	35	35																																													
	Joint	non-joint																																													
N Patients	23990	31915																																													
N features	6	6																																													
	Joint	non-joint																																													
N Patients	23990	31915																																													
N features	32	32																																													
Joint	ASD	nonASDext	Joint	ASD	nonASDext	Joint	ASD	nonASDext	Joint	ASD	nonASDext																																				
precision	0.95	0.62	precision	0.98	0.83	precision	0.93	0.57	precision	0.98	0.83																																				
recall	0.90	0.79	recall	0.96	0.93	recall	0.90	0.69	recall	0.96	0.92																																				
fscore	0.92	0.7	fscore	0.97	0.88	fscore	0.92	0.63	fscore	0.97	0.87																																				
<table border="1"><thead><tr><th></th><th>ASD</th><th>nonASDext</th></tr></thead><tbody><tr><td>ASD</td><td>6562</td><td>239</td></tr><tr><td>nonASDext</td><td>438</td><td>740</td></tr></tbody></table>				ASD	nonASDext	ASD	6562	239	nonASDext	438	740	<table border="1"><thead><tr><th></th><th>ASD</th><th>nonASDext</th></tr></thead><tbody><tr><td>ASD</td><td>6738</td><td>78</td></tr><tr><td>nonASDext</td><td>239</td><td>924</td></tr></tbody></table>				ASD	nonASDext	ASD	6738	78	nonASDext	239	924	<table border="1"><thead><tr><th></th><th>ASD</th><th>nonASDext</th></tr></thead><tbody><tr><td>ASD</td><td>6876</td><td>191</td></tr><tr><td>nonASDext</td><td>477</td><td>435</td></tr></tbody></table>				ASD	nonASDext	ASD	6876	191	nonASDext	477	435	<table border="1"><thead><tr><th></th><th>ASD</th><th>nonASDext</th></tr></thead><tbody><tr><td>ASD</td><td>6917</td><td>94</td></tr><tr><td>nonASDext</td><td>174</td><td>794</td></tr></tbody></table>				ASD	nonASDext	ASD	6917	94	nonASDext	174	794
	ASD	nonASDext																																													
ASD	6562	239																																													
nonASDext	438	740																																													
	ASD	nonASDext																																													
ASD	6738	78																																													
nonASDext	239	924																																													
	ASD	nonASDext																																													
ASD	6876	191																																													
nonASDext	477	435																																													
	ASD	nonASDext																																													
ASD	6917	94																																													
nonASDext	174	794																																													
non-joint	ASD	nonASDext	non-joint	ASD	nonASDext	non-joint	ASD	nonASDext	non-joint	ASD	nonASDext																																				
precision	0.96	0.63	precision	0.98	0.80	precision	0.97	0.48	precision	0.99	0.82																																				
recall	0.94	0.76	recall	0.97	0.92	recall	0.94	0.70	recall	0.98	0.89																																				
fscore	0.95	0.67	fscore	0.98	0.85	fscore	0.95	0.57	fscore	0.98	0.86																																				

Building Patient Similarity Matrices - “Scalability”



Dataset	ASD	Elements (10^9)	Matrix Size (GB)
bms	55066	3.03223	24.3
scq	44631	1.99192	15.9
rbsr	34033	1.58245	9.3
dcdq	24855	0.61777	4.9
unique	55905	3.12537	25

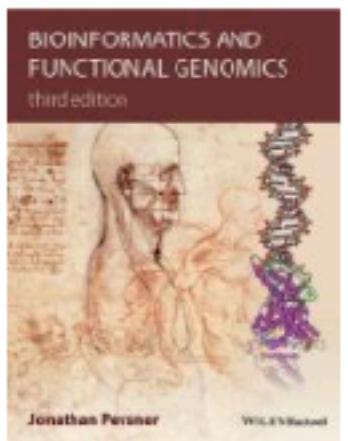
Network Fusion Allows Joint Modelling of ASD SPARK Clinical Data



after Wang, B. et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* **11**, 333–337 (2014).

Next Week

If you would like to read ahead then please look at the following chapter, if not we will cover this in next week's lecture
This is available from the Bio1 course "Resource List"



BOOK Bioinformatics and functional genomics ✓

Pevsner, Jonathan, 1961-, Third edition., Hoboken, John Wiley & Sons, Incorporated, 2015

Note: Please read Chapter 9 - Analysis of Next-Generation Sequencing Data.

[Add tags to item](#)

[Complete](#) [Check availability >](#)