Gene networks

Why networks

Data lends itself to network analysis (time series, developmental, large numbers of samples)

Complementary to DE

Centrality metrics can provide other insights

Gene clusters with associations

How are networks inferred

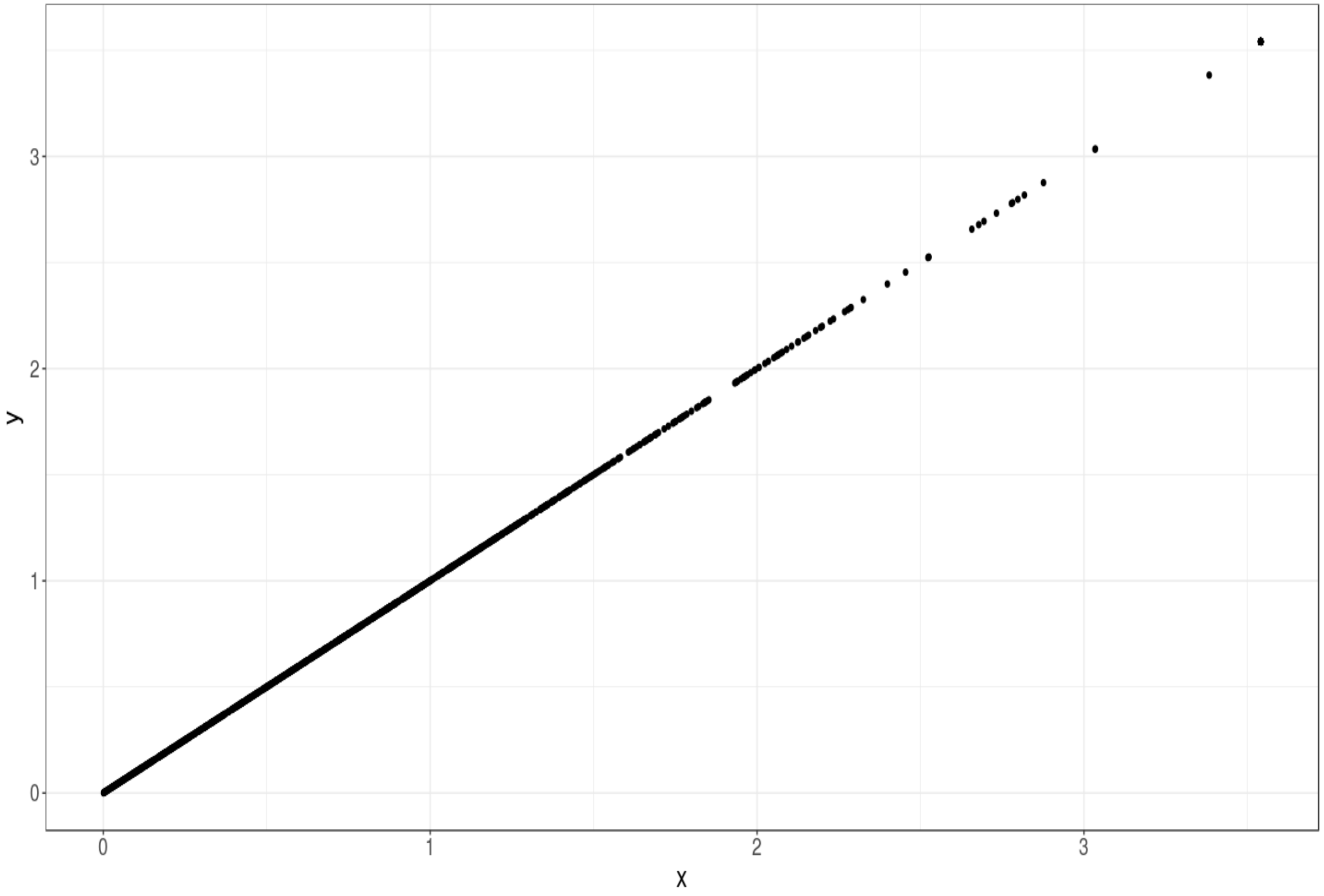
Vast number of options

- Correlation (Pearson, Spearman)
- Mutual Information and extensions (CLR, ARACNE)
- Regression (TIGRESS, GENIE3)
- Other (PLSNET, ANOVERENCE)

Pearson correlation

Linear relationship (parametric)

x and *y* behave the same in a linear manner

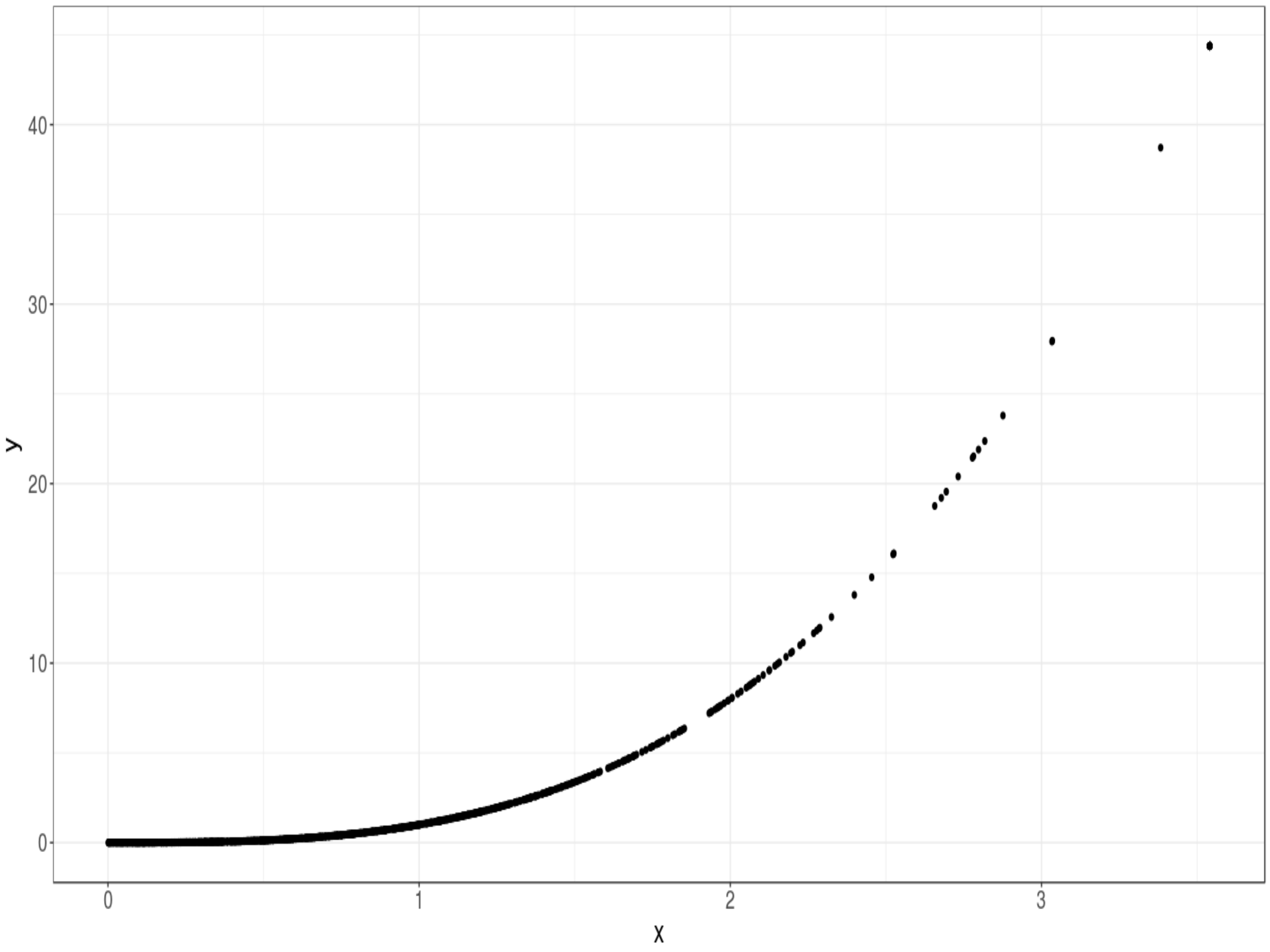


Spearman correlation

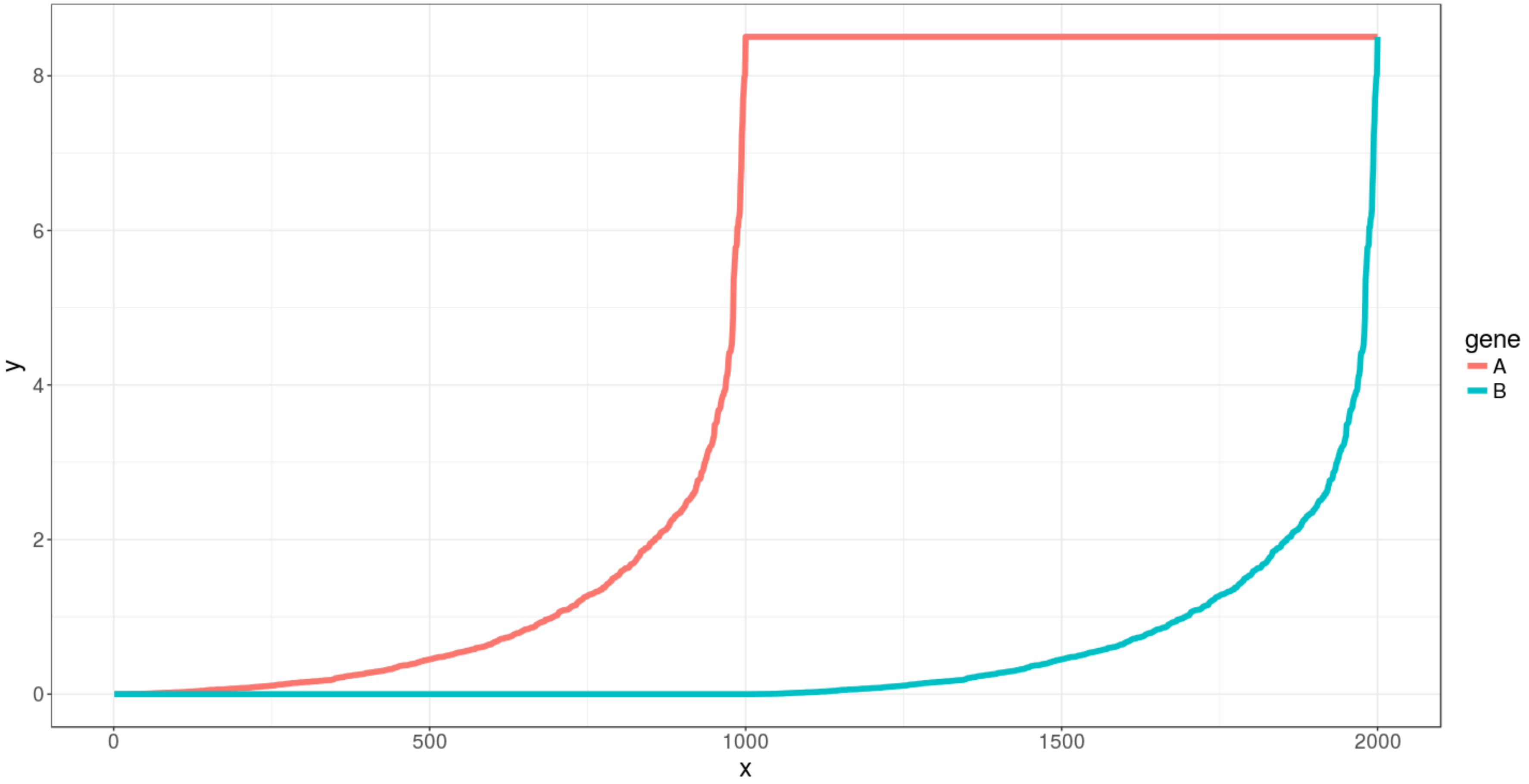
As Pearson, but computed on ranks

Also detects nonlinear relationships

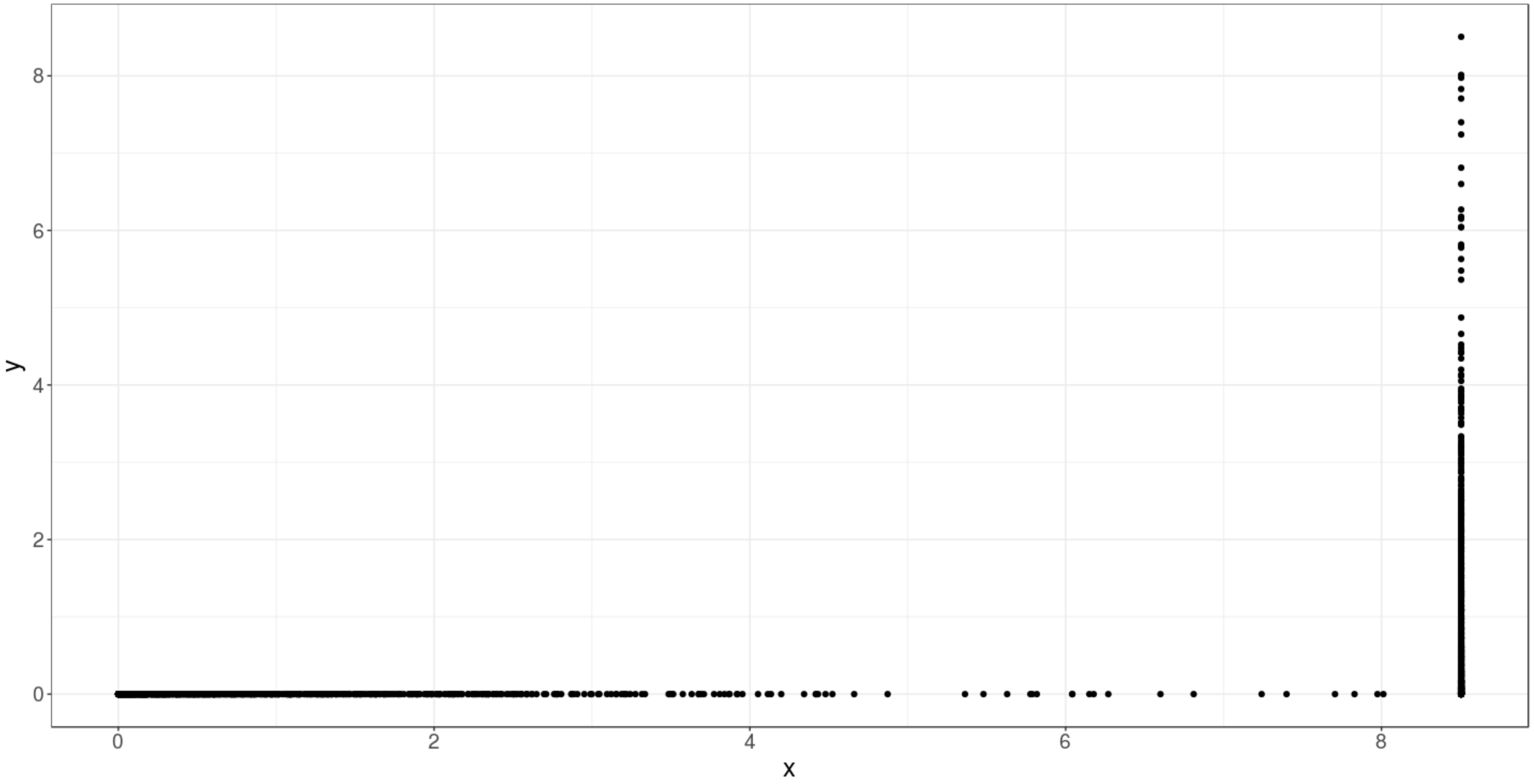
x and *y* behave the same in a linear or nonlinear manner



What about these two genes



Complex interactions are harder to detect

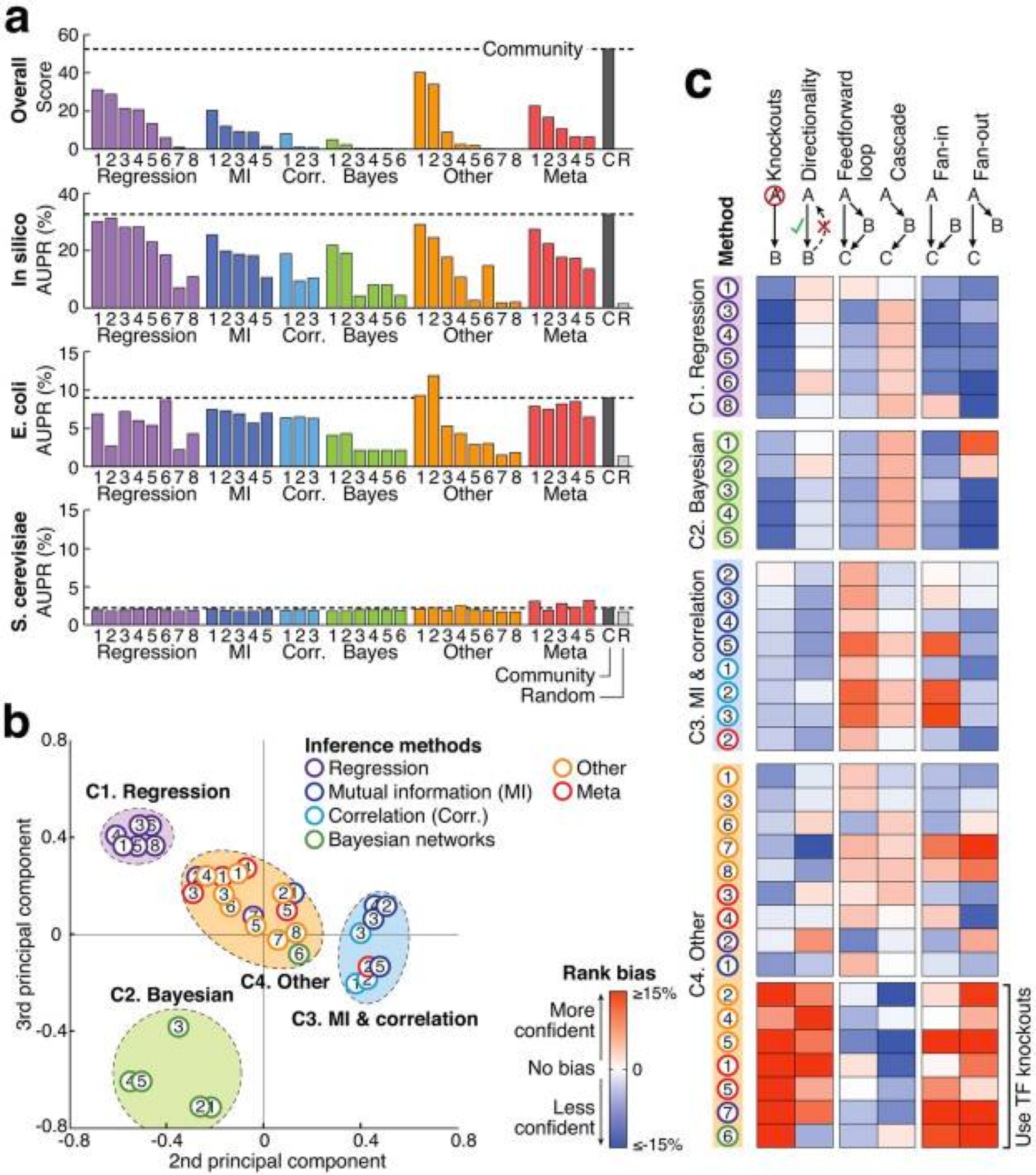


DREAM5 consortium

Benchmark of GRN predictions

Different groups of methods tend to detect some interactions better than others

Combining all networks into a meta-network was better than any single method

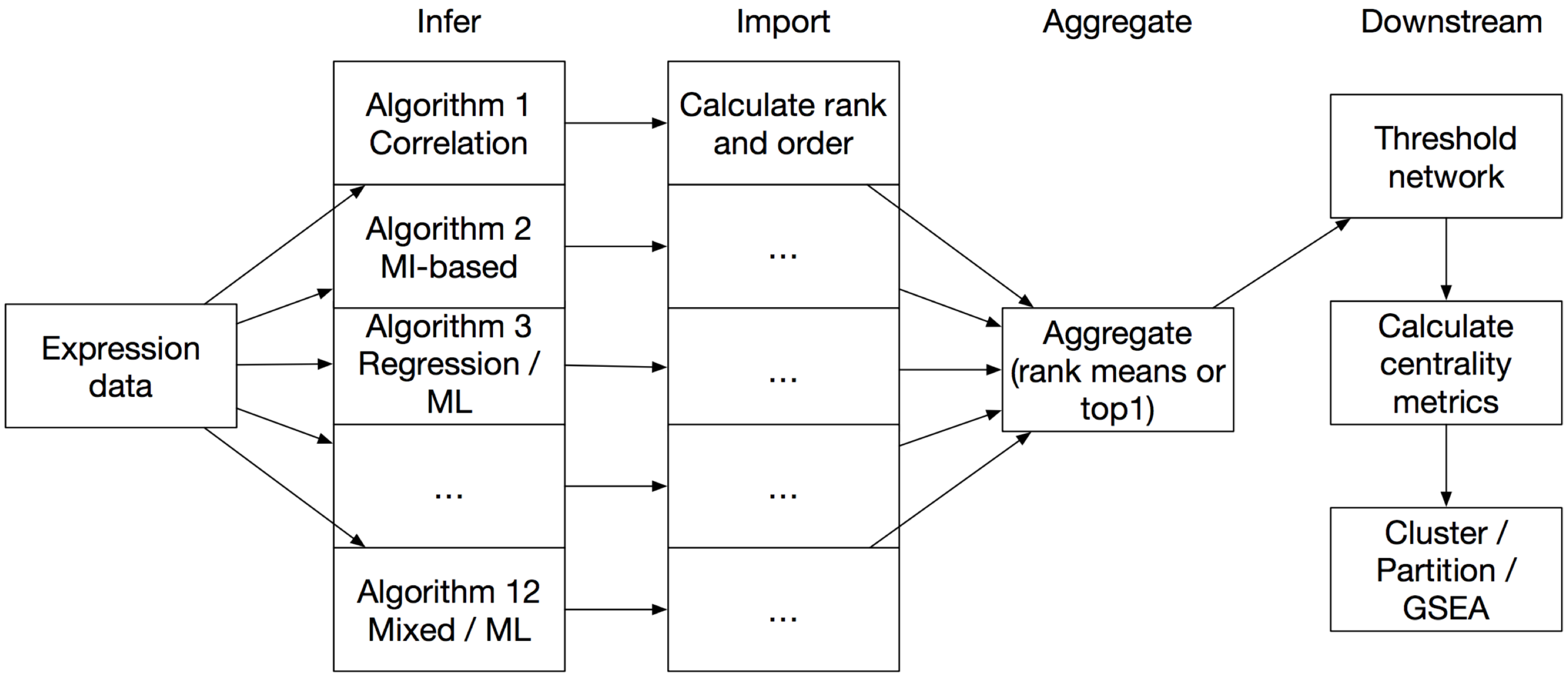


Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., … Zimmer, R. (2012). Wisdom of crowds for robust gene network inference. *Nature Methods*, *9*(8), 796–804. https://doi.org/10.1038/nmeth.2016

Seidr makes life easier-ish

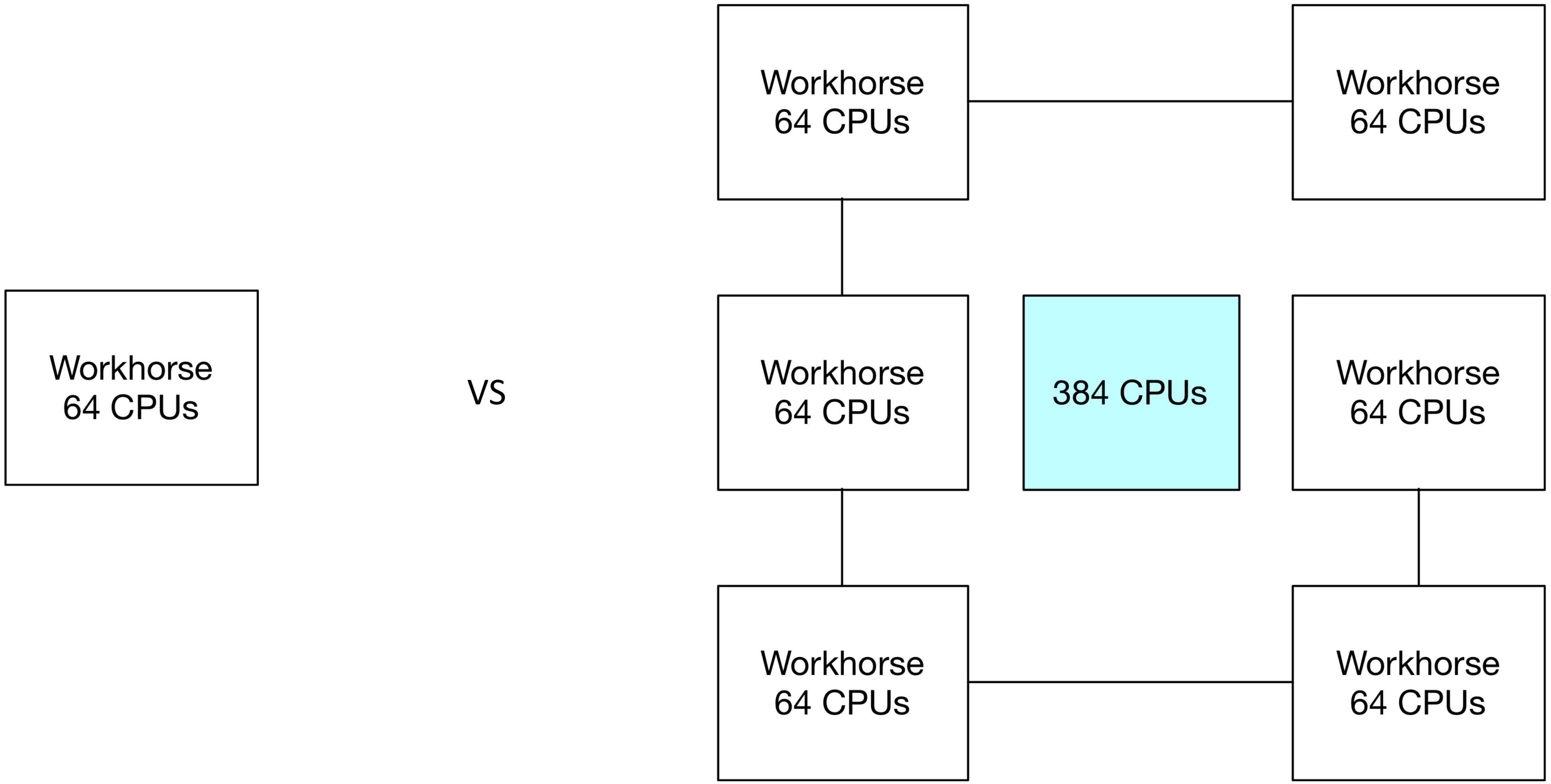
ANOVERENCE	Küffner et al., 2012	ANOVA	C++	C++	No	No
ARACNE	Margolin et al., 2006	MI + DPI	C++	C++	Yes	Yes
CLR	Faith et al., 2007; Daub et al., 2004	MI + CLR	MATLAB / C / C++	C++	No	Yes
Elastic Net ensemble	Ruyssinck et al., 2014	Elastic Net Regression	R (glmnet)	C++ (glmnet)	No	Yes
GENIE3	Huynh-Thu et al., 2010	Random Forest Regression	R (randomForest)	C++ (ranger)	No	Yes
NARROMI	Zhang et al., 2013	MI + Linear Programming	MATLAB	C++ (glpk)	No	Yes
Partial Correlation	Schäfer and Strimmer, 2005	Correlation	R	C++	No	No
Pearson Correlation	NA	Correlation	NA	C++	No	No
PLSNET	Guo et al., 2016	PLS	MATLAB	C++	No	Yes
Spearman Correlation	NA	Correlation	NA	C++	No	No
SVM ensemble	Ruyssinck et al., 2014	SVM regression	R (libsvm) / C	C++ (libsvm, liblinear)	No	Yes
TIGRESS	Haury et al., 2012	LASSO Regression	MATLAB / R	C++ (glmnet)	No	Yes

Seidr workflow



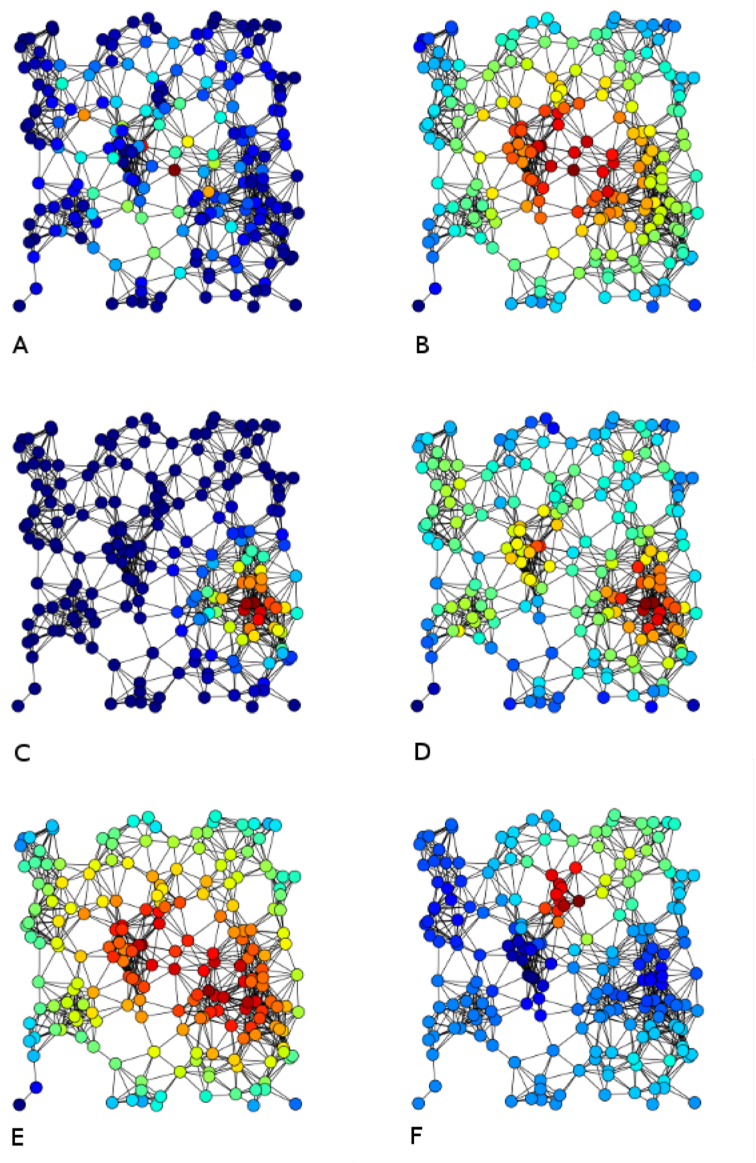
Schiffthaler, B., Serrano, A., Delhomme, N., & Street, N. R. (2018). Seidr: A toolkit for calculation of crowd networks. *bioRxiv*, 250696. https://doi.org/10.1101/250696

OpenMPI enables cluster computing



Network centrality

A) **Betweenness** – Nodes that are placed in between others. Control flow
B) **Closeness** – Nodes that are placed most central. Can act quickly on other nodes
C) **Eigenvector** – Nodes that are connected to other important nodes and can influence the entire network
D) **Degree** – Nodes that have the most direct connections
E) **Harmonic** – Variant of closeness defined for disconnected graphs
F) **Katz** – Variant of eigenvector centrality as a measure of influence



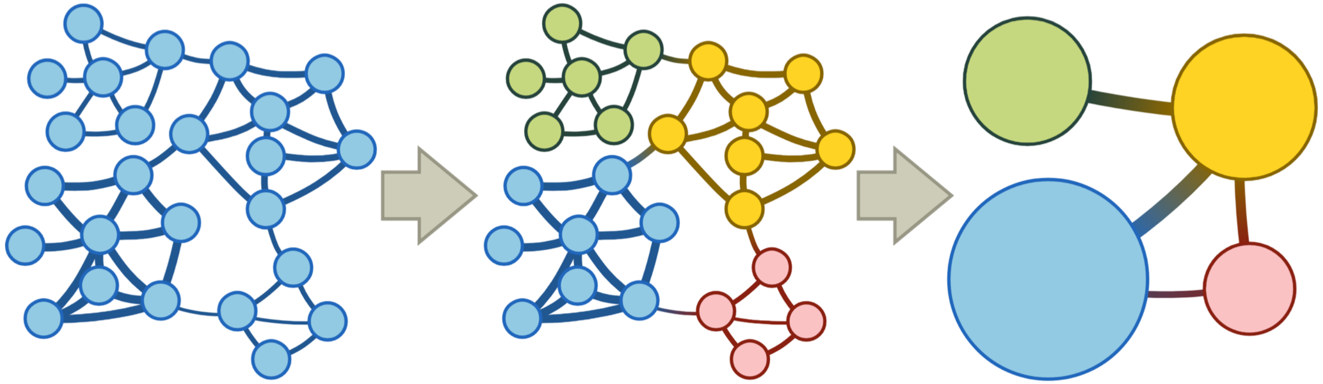
https://en.wikipedia.org/wiki/Katz_centrality

Graph partitioning

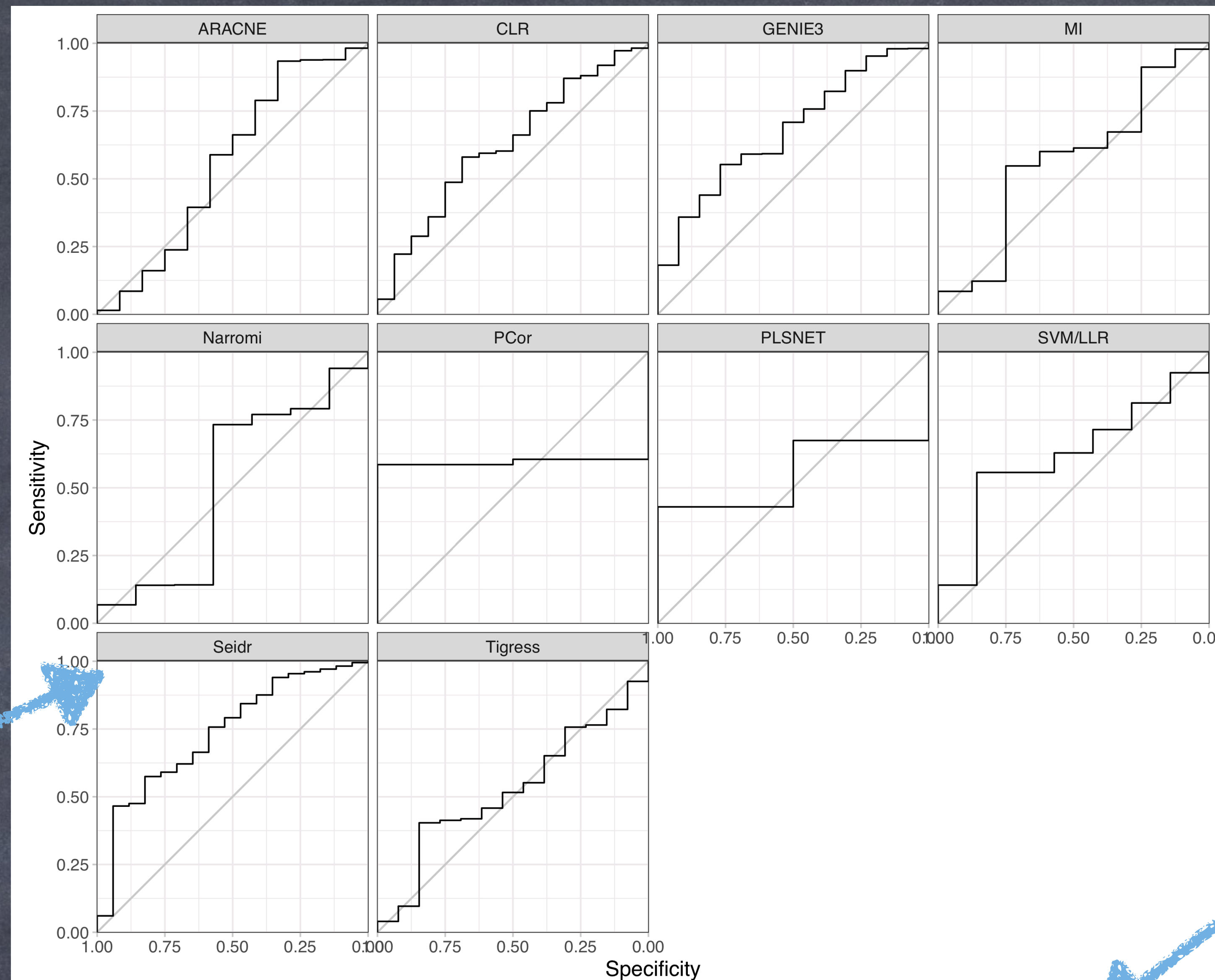
Objective is to divide the graph into meaningful clusters

Only topology as input, not underlying data

We use InfoMap, which partitions via random walks

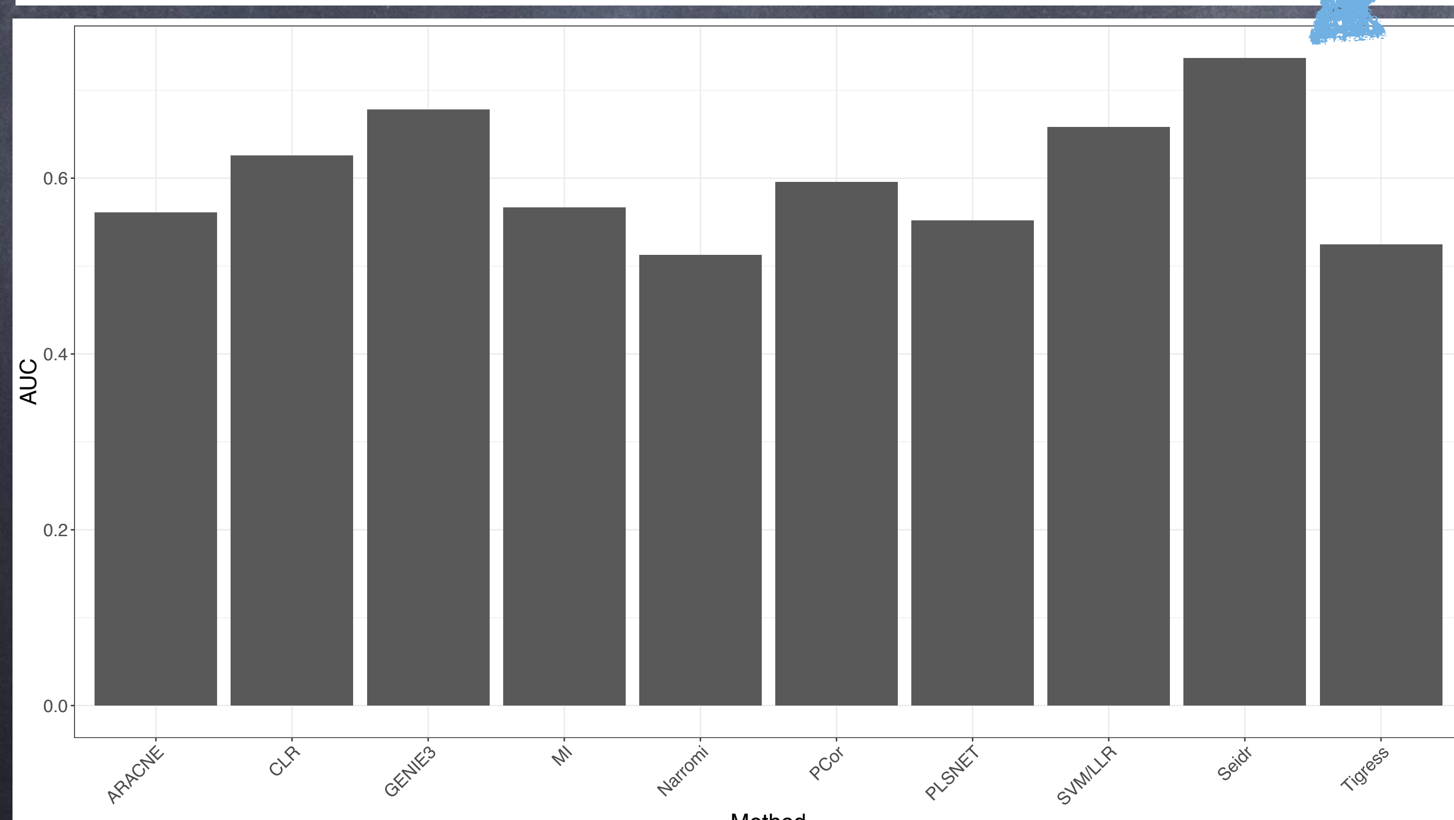


Rovall, M., & Bergstrom, C. . (2008). Maps of Random Walks on Complex Network Reveal Community Structure. In *Proceedings of the National Academy of Sciences,* (p. 105(4), 1118-1123.).



Best spot

You would want it to be 1



0.6

0.3

0