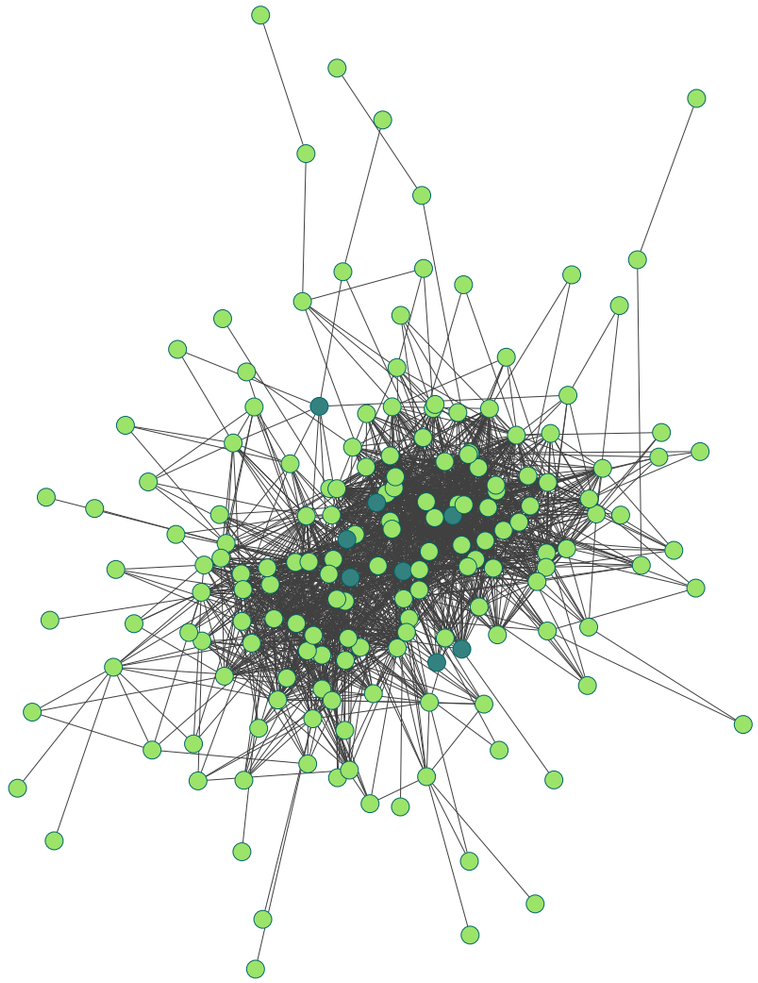
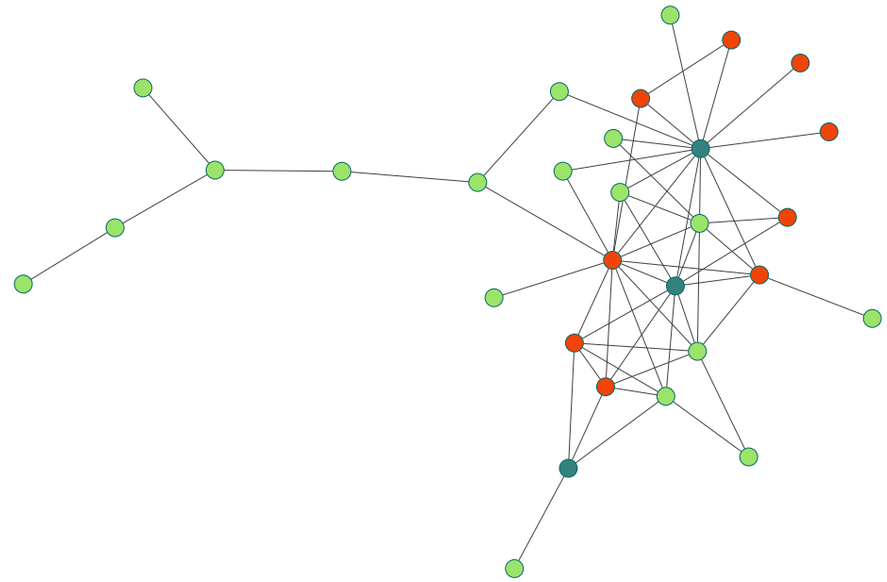




We have the OTUs, so what now?
Building and exploring OTUs co-occurrence networks

Antonio Fernandez-Guerra
Max Plank Institute for Marine Microbiology

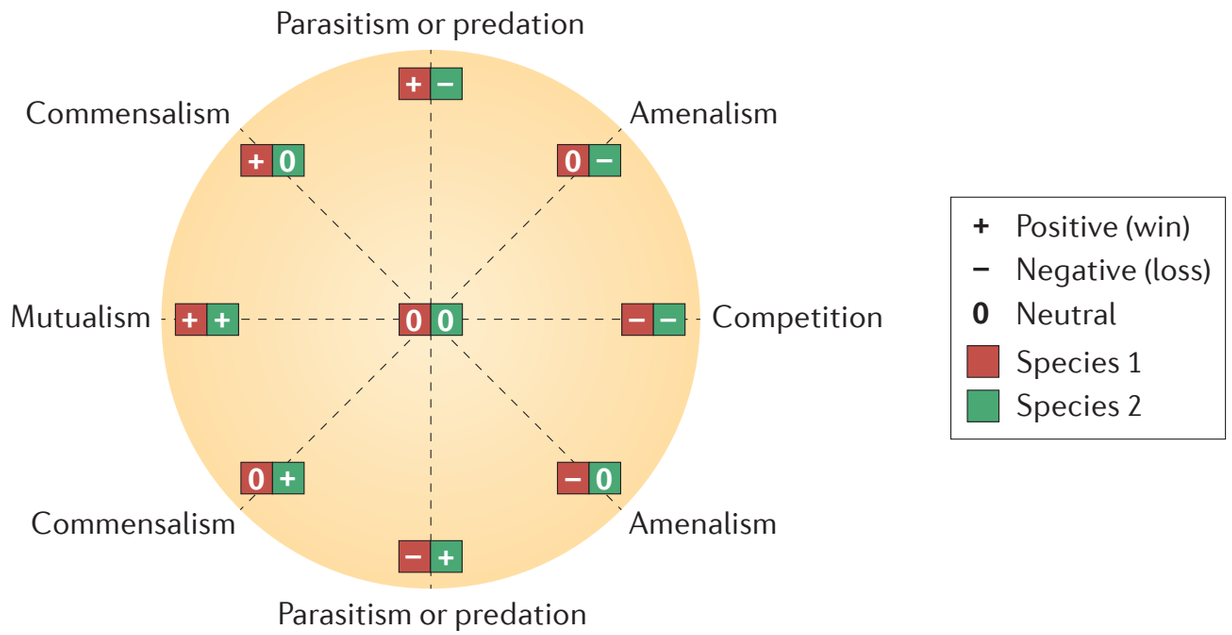


Figure 1 | **Summary of ecological interactions between members of different species.** The wheel display introduced by Lidicker[1] has been adapted to summarize all possible pairwise interactions. For each interaction partner, there are three possible outcomes: positive (+), negative (–) and neutral (0). For instance, in parasitism, the parasite benefits from the relationship (+), whereas the host is harmed (–); this relationship is thus represented by the symbol pair + –.

Commensalism: benefits one organism and the other organism is neither benefited nor harmed, i.e. biodegradation

Amensalism: an organism inflicts harm to another organism without any costs or benefits received i.e. metabolic by-products of a microbial species alter the environment to the detriment of others

Mutualism: interaction between two or more species, where species derive a mutual benefit, i.e. biofilms

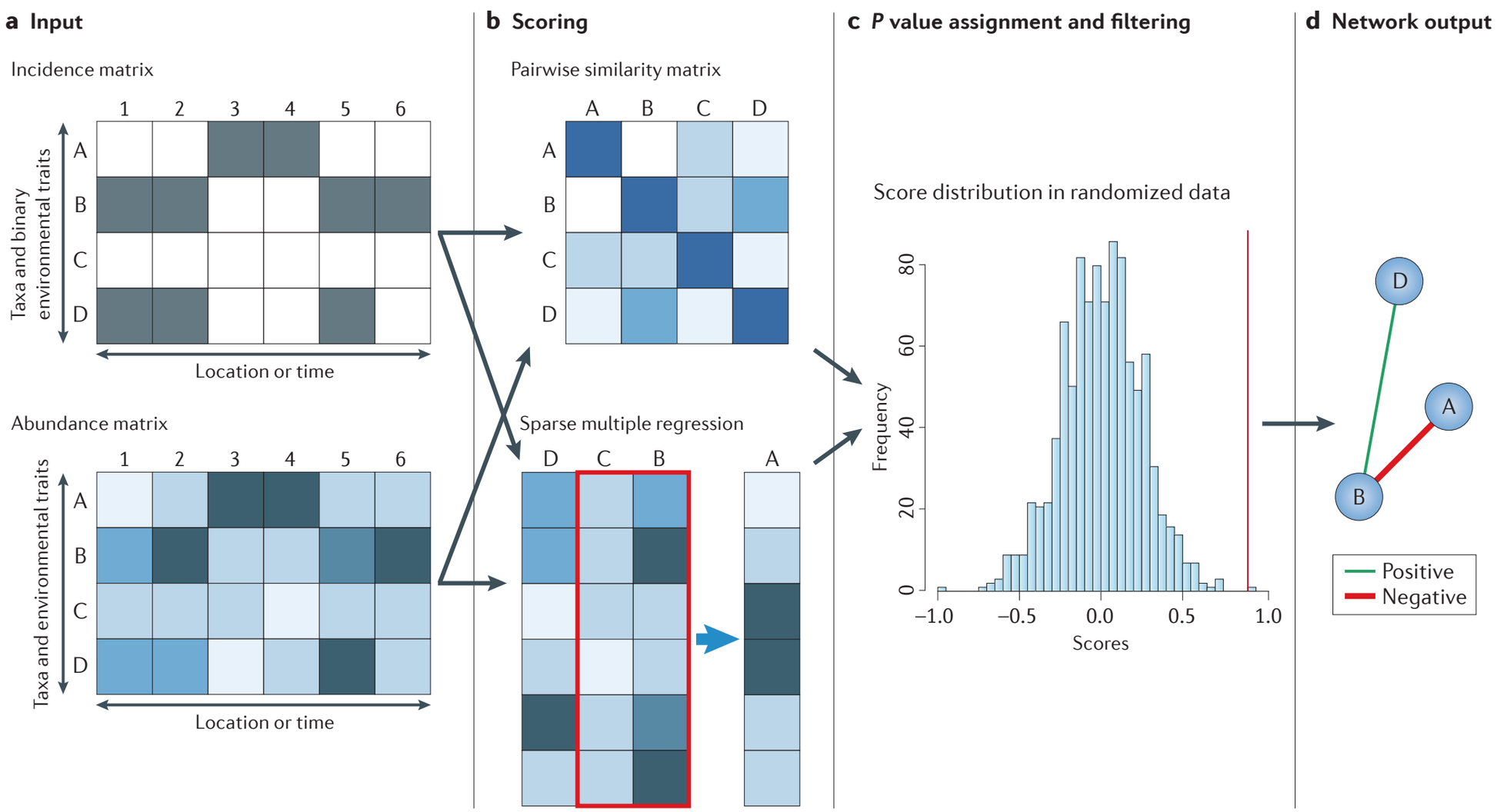
Competitive exclusion: two organisms competing for the same resources cannot coexist if other ecological factors are constant

Co-occurrence and correlation patterns can be used for the prediction of species interactions

Faust & Raes, 2012

Network Inference Methods

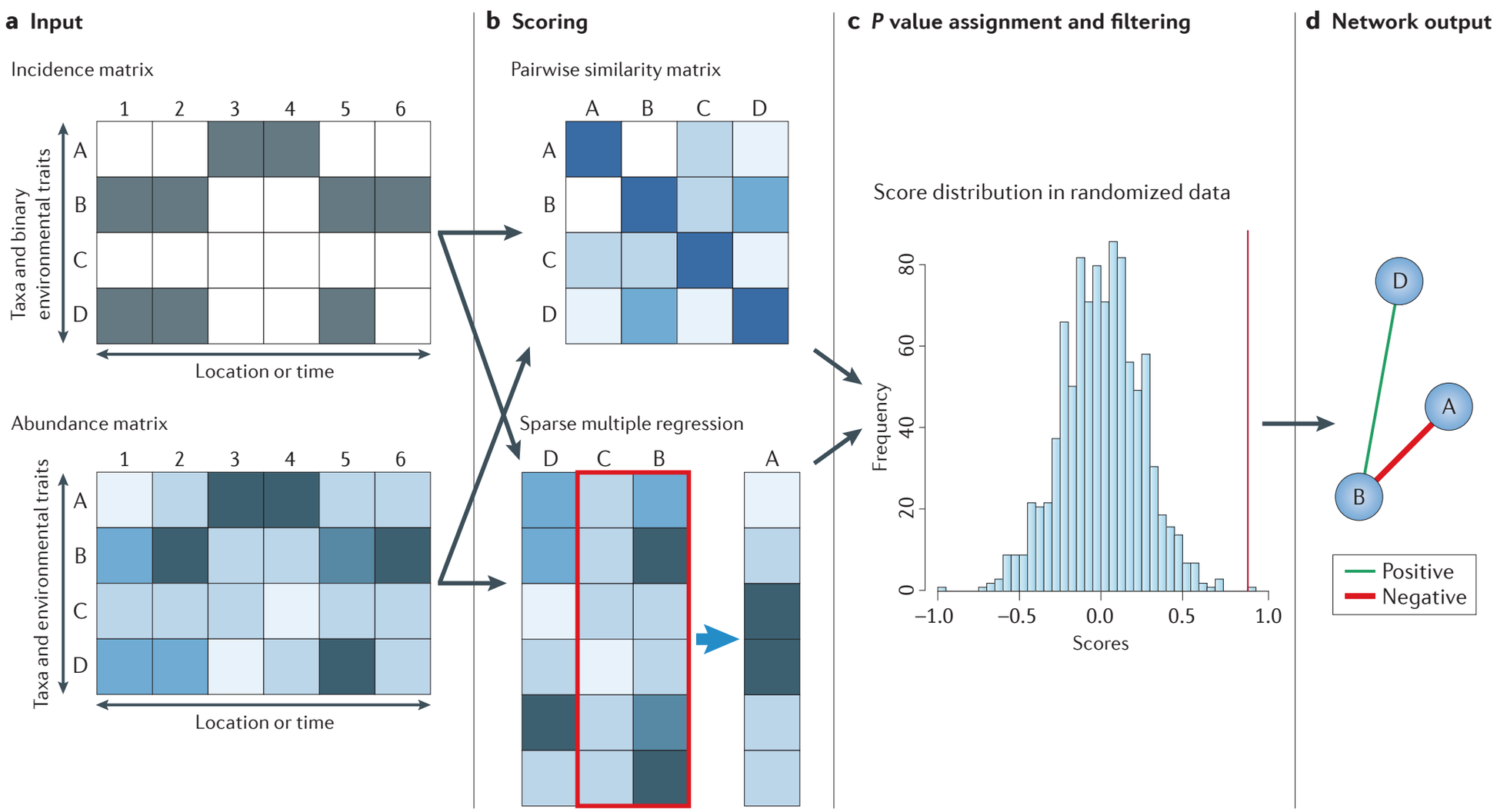
- Pairwise relationships: similarity-based network inference.
- Complex relationships: regression- and rule-based networks.



Faust & Raes, 2012

Network Inference Methods

- **Pairwise relationships: similarity-based network inference.**
- Complex relationships: regression- and rule-based networks.



Faust & Raes, 2012

Be aware of...

bias from samples to abundance data → Relative abundances by normalisation or downsampling

downsampling: all samples have the same total counts (doesn't alleviate compositional effects)

normalisation/standardisation: each OTU is normalised by the total counts in the sample (overestimates the number of zero fractions)

↓

Resulting fractions fall in what is call compositional data (particular geometrical and statistical properties)

large percentage of zeros taxon is absent from the sample or its abundance is below the detection level

Generate a good **null model** to asses the significance of predicted associations by **shuffling** the data

multiple testing correction: High number of OTU pairs being compared the more likely it is that some associations will be significant by chance alone.

p-values are adjusted: to control the expected proportion of wrongly rejected null hypotheses (number of false-positive associations).

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Inferring Correlation Networks from Genomic Survey Data

Jonathan Friedman[1], Eric J. Alm[1,2,3]*

SparCC: Sparse Correlations for Compositional data

SparCC
yonatanf

Clone | Fork | Compare

Overview | **Source** | Commits | Branches | Pull requests | Issues 1 | Wiki | Downloads

default | SparCC / | New file

example			
lib			
.project	360 B	2011-07-12	Init commint of SparCC. Currently based on the MatrixDictionary class.
.pydevproject	288 B	2011-07-12	Init commint of SparCC. Currently based on the MatrixDictionary class.
MakeBootstraps.py	2.4 KB	2011-07-12	Init commint of SparCC. Currently based on the MatrixDictionary class.
PseudoPvals.py	3.2 KB	2012-10-31	FIX: fixed error is readme, and a few typos
SampleDist.py	2.1 KB	2011-07-24	FIX: print output file name in sample_dist
SparCC.py	3.4 KB	2011-11-10	STY: iteration averaging moved from SurveyMatrix to basis_correlation (Main function).
__init__.py	0 B	2011-07-12	Init commint of SparCC. Currently based on the MatrixDictionary class.
readme.txt	5.7 KB	2012-10-31	FIX: fixed error is readme, and a few typos

https://bitbucket.org/yonatanf/sparcc/

What sparCC does...

Assumptions:

the number of different components (e.g., OTUs or genes) is large,
the true correlation network is 'sparse' (most components are not strongly correlated with each other)

SparCC estimates the linear Pearson correlations between the log-transformed components

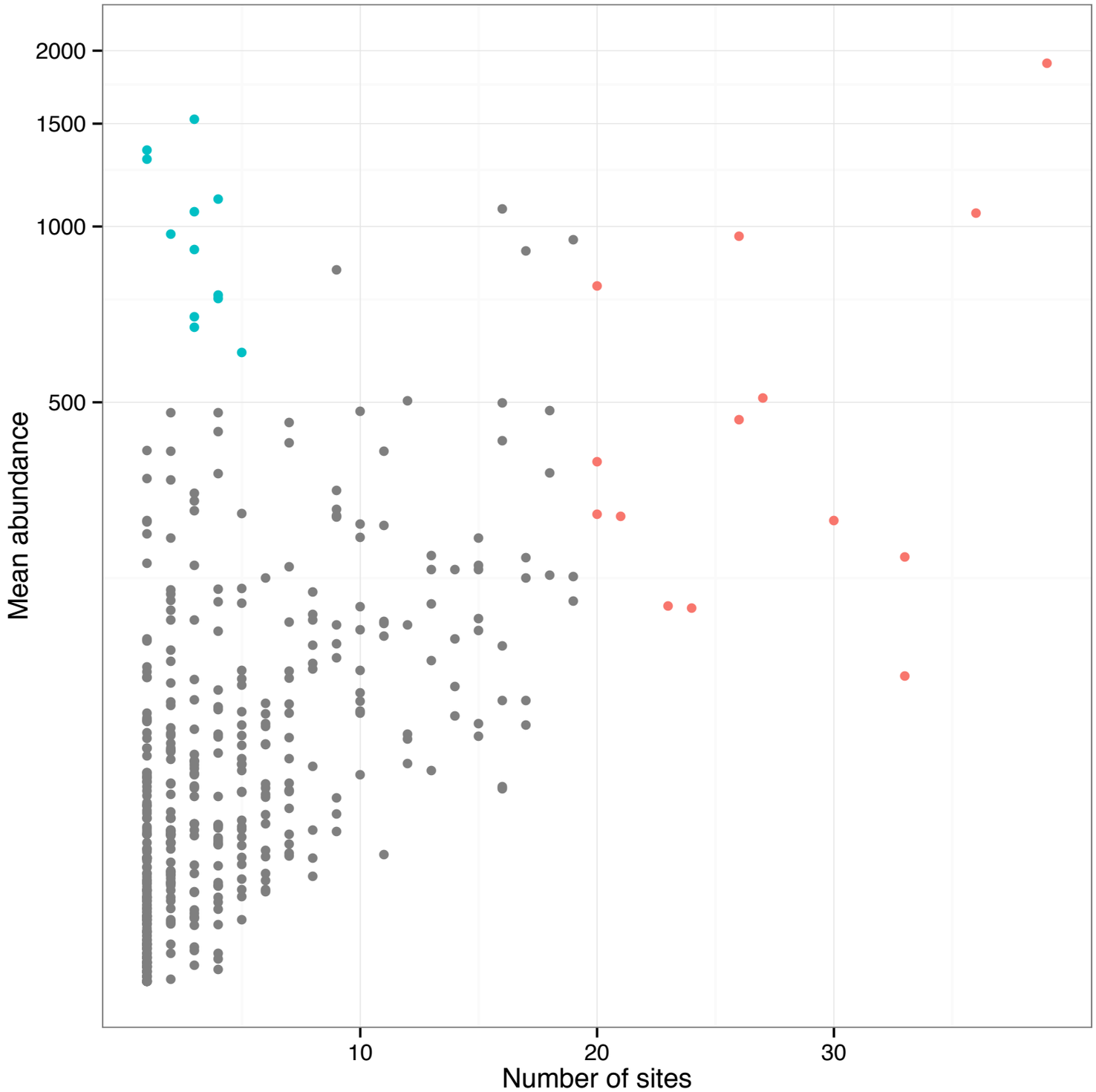
normalisation/standardisation: employs a Bayesian approach to estimate the true fractions from the observed counts. It assumes unbiased sampling in the sequencing procedure where the Dirichlet distribution is appropriate (multivariate generalisation of the ß-distribution)

large percentage of zeros: eliminate zero fractions by adding small pseudocounts. Similar to add a pseudocount of 1 to all count values.

bootstrapping: assigning each OTU in each sample a number of counts that is randomly sampled from the OTU's observed counts across all samples, with replacement.

multiple testing correction: pseudo p-values are assigned to be proportion of simulated data sets for which a correlation value at least as extreme as the one computed for the original data was obtained

Friedman & Alm, 2012

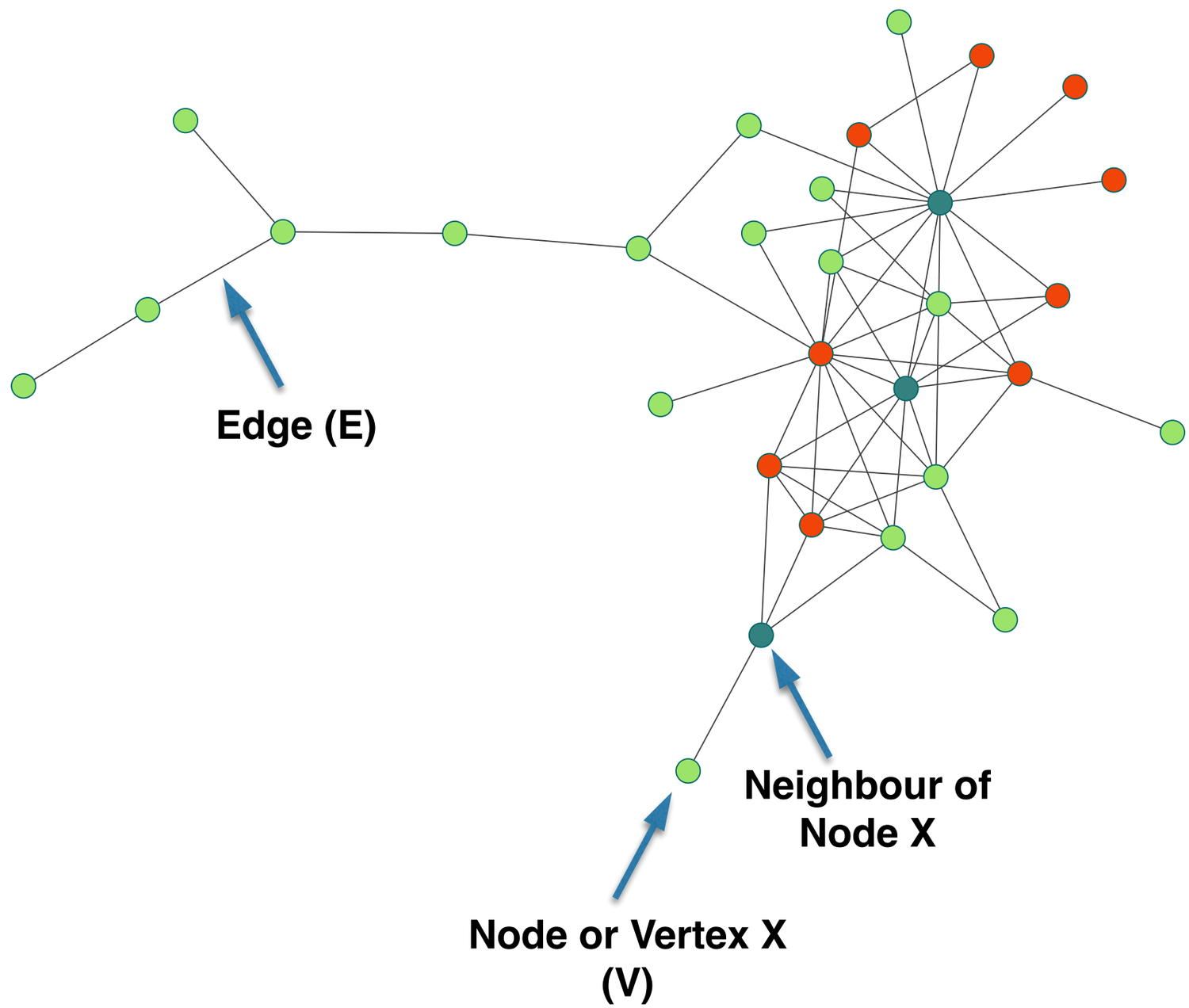


Cosmopolitan: Present in more than 20 sites

Endemic: Present in at least 8 sites
and with a mean abundance >= 500

OTU distribution

Basic network concepts



Basic network concepts

sparse networks: networks where the number of edges is much smaller than the number of possible edges

shortest path: the minimal number of edges that need to be traversed to travel from one vertex to another

distance: the length the length of the shortest path between the vertices

diameter: the maximal distance of any pair of vertices.

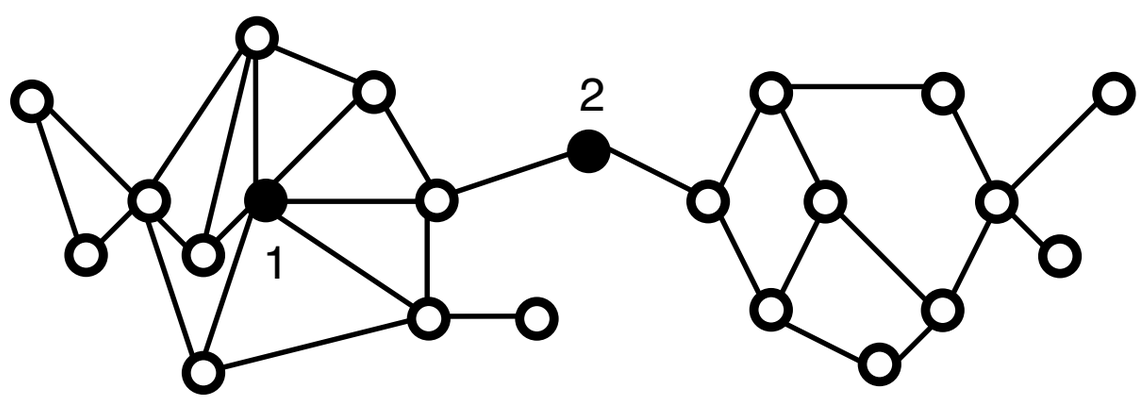
average path length: the average distance between all pairs of vertices

node degree: the number of edges adjacent to the vertex

clustering coefficient: the local cohesiveness of a network and measures the probability that two vertices with a common neighbour are connected

closeness centrality: specifies which vertices have the shortest paths to all others

betweenness centrality: measures how often a vertex or edge is present in the set of all shortest paths



Björn H. Junker, Falk Schreiber. Analysis of Biological Networks. 2008

Hands-On

1. Prepare ICOMM data - **R**
2. Calculate correlation matrix - **SparCC**
3. OTU distributions - **Occupancy–abundance** relationship
4. Build networks - **igraph** (R)
5. Network exploration - **Cytoscape**

Cytoscape/Network tutorial

http://wiki.bio.dtu.dk/teaching/index.php/Course27040Spring2013