CoNet – a Co-occurrence Network inference tool

Karoline Faust and Jeroen Raes

Research group of Bioinformatics and (Eco-)Systems Biology
1) Structural Biology, VIB 2) Applied Biological Sciences, VUB
Pleinlaan 2, 1050 Brussels
karoline.faust@vib-vub.be, jeroen.raes@vib-vub.be

Microorganisms interact in various ways, e.g. as parasites or predators, as competitors for scarce resources, in cross-feeding chains etc. Such ecological interactions as well as niche preferences and random factors shape the abundances of microorganisms. In recent times, more and more metagenomic experiments measure the abundance of microbial DNA across multiple samples, spanning diverse environments from the ocean to the human body. These abundances are the starting point for co-occurrence analysis, which aims at predicting ecological relationships (who interacts with whom) by constructing association networks (who is frequently found together with whom). Here we present CoNet, which offers a variety of association network construction routines. CoNet is available at:
**http://systemsbiology.vub.ac.be/conet**

## Typical similarity-based association network construction workflow
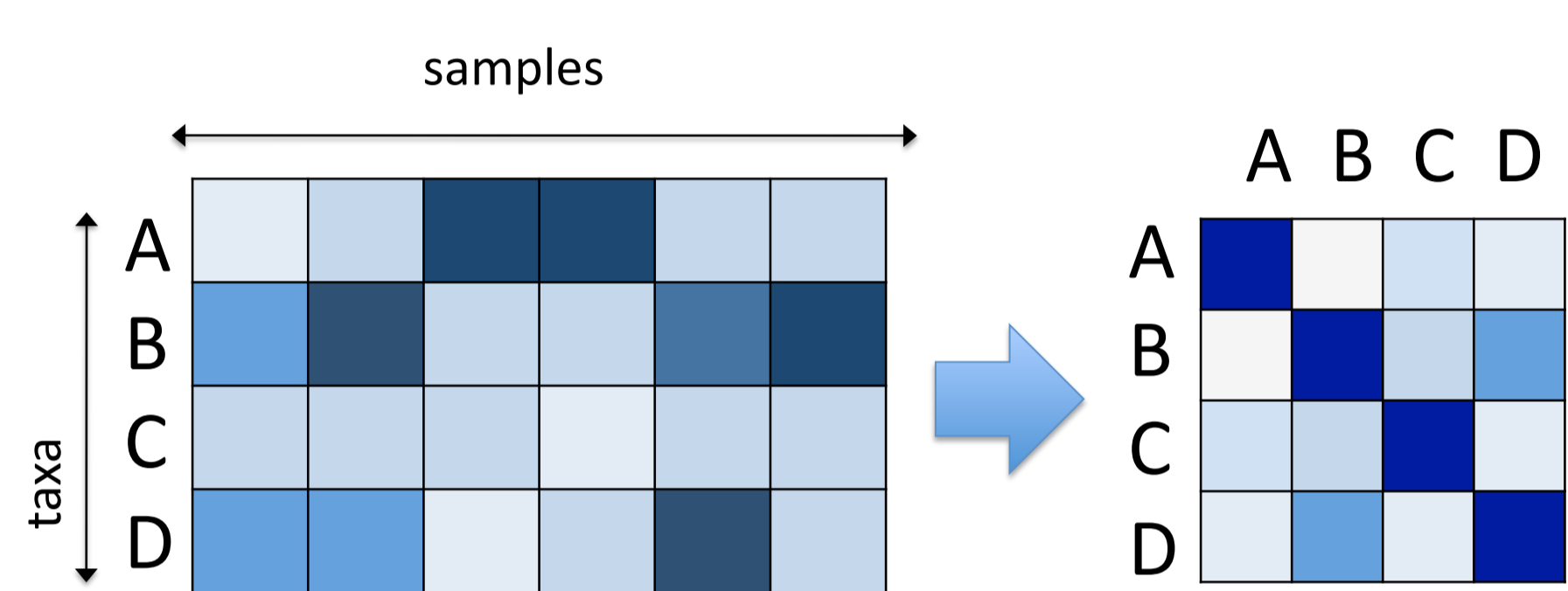
For each pair of taxa, the similarity of the two abundance profiles is assessed. Popular scoring functions include measures of correlation (e.g. Spearman, Pearson), dissimilarity (e.g. Bray-Curtis, Kullback-Leibler) and similarity (e.g. mutual information).
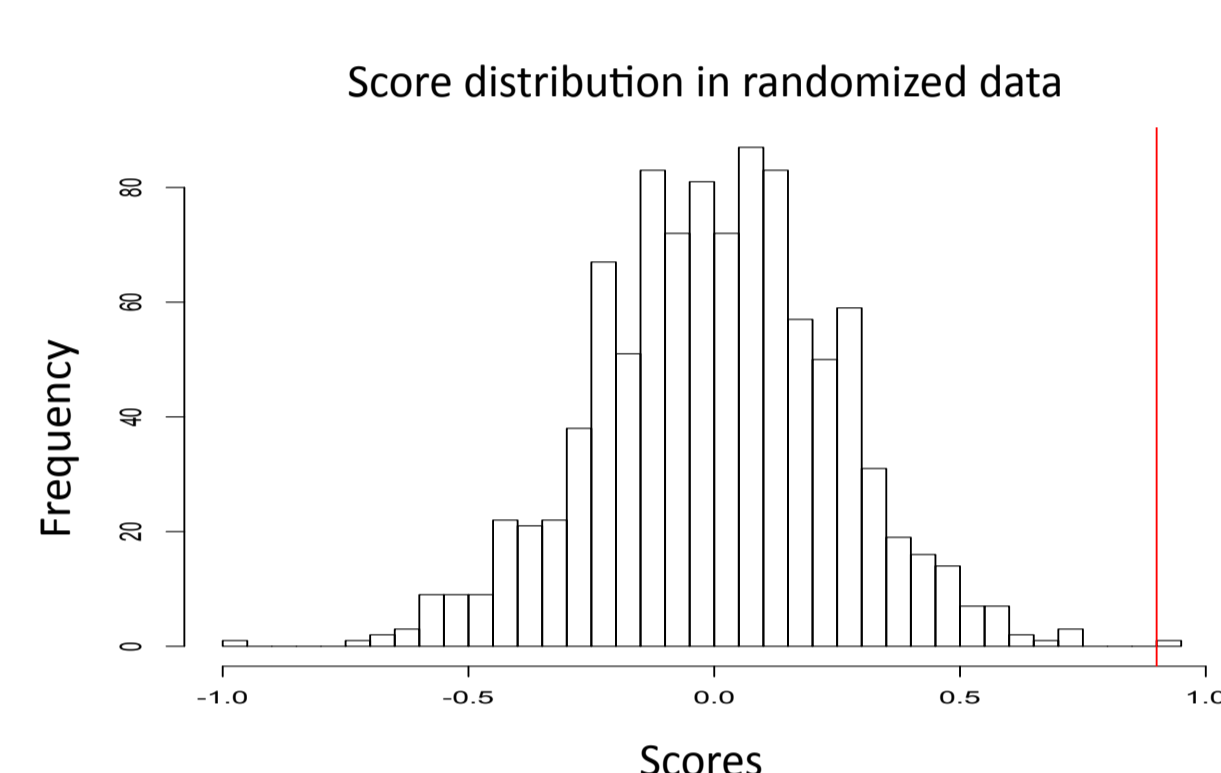
To assess the significance of an observed score, a null distribution of scores is computed from randomized data. Taxon pairs with scores above the threshold p-value (i.e. scores that do not differ enough from the bulk of the random scores) are discarded.
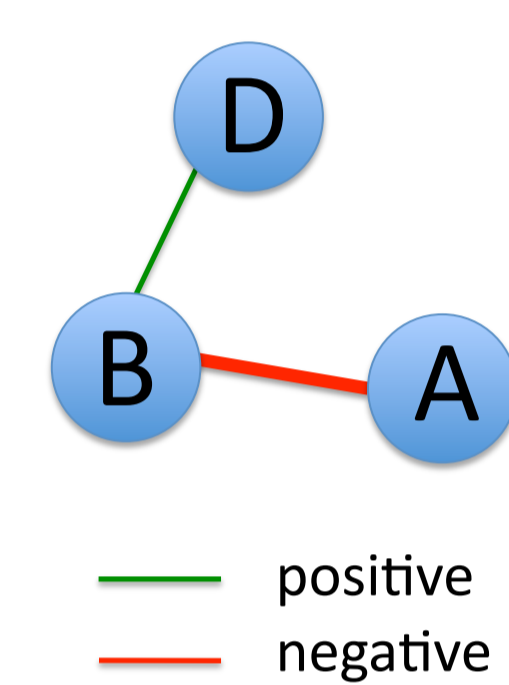
The remaining taxon pairs are visualized as a network. Edge thickness can code for the strength of the taxon association.
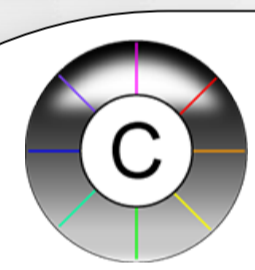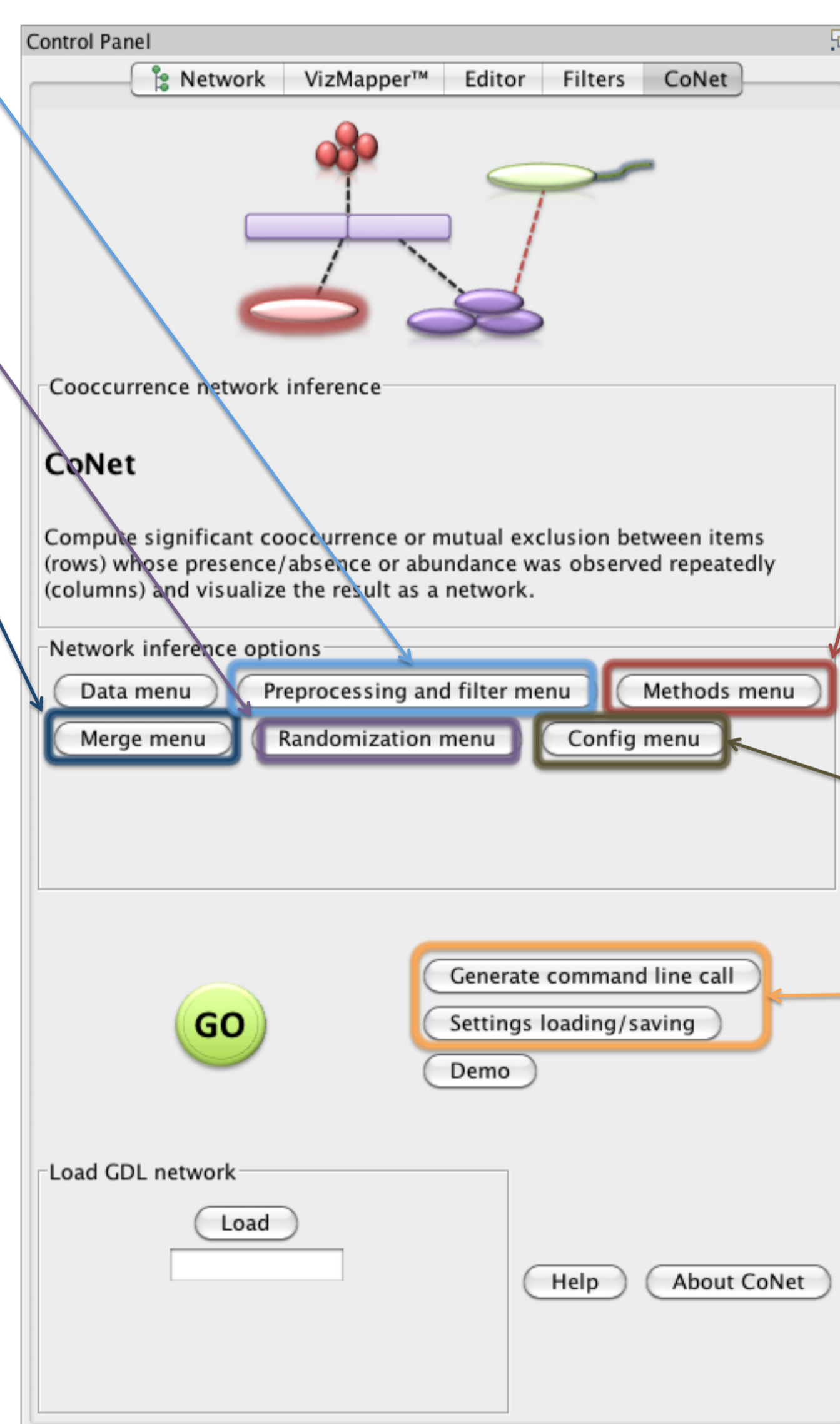
CoNet – interface and features

Matrix filtering and normalization options

randomization routines (permutation, bootstrap, p-value merging, multiple-test correction) including a routine that mitigates compositionality bias [2]

Various options to merge results from different methods (union, intersection, multigraph etc.)

similarity measures for presence/absence (hypergeometric distribution, Jaccard distance, association rule mining) and abundance matrices (Pearson, Spearman, Euclidean, Bray Curtis, mutual information etc.), manual or automated threshold setting

missing value treatment, minet [3] integration

current parameter settings can be exported as a java command to run CoNet on command line or saved to a file to re-run an experiment in CoNet

As a Cytoscape plugin, CoNet makes use of the network visualization and analysis power of Cytoscape [1]. A special feature of CoNet is its capacity to combine several measures into an ensemble approach. The idea of ensemble is that different prediction methods have different weaknesses, so that faulty predictions are only supported by a minority of the methods.
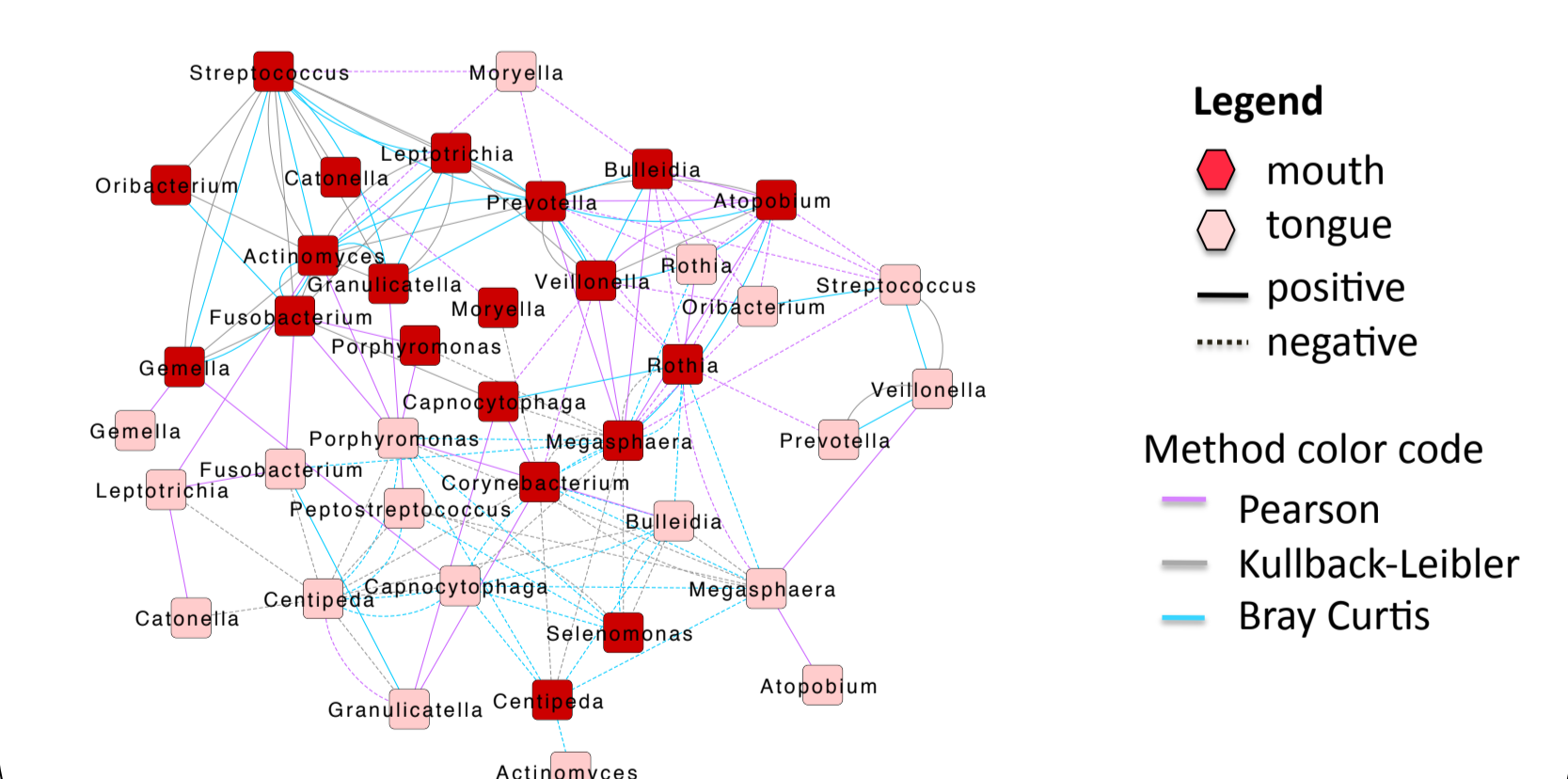

Examples*

A) co-presence network from global presence/absence data [4] obtained with the hypergeometric distribution
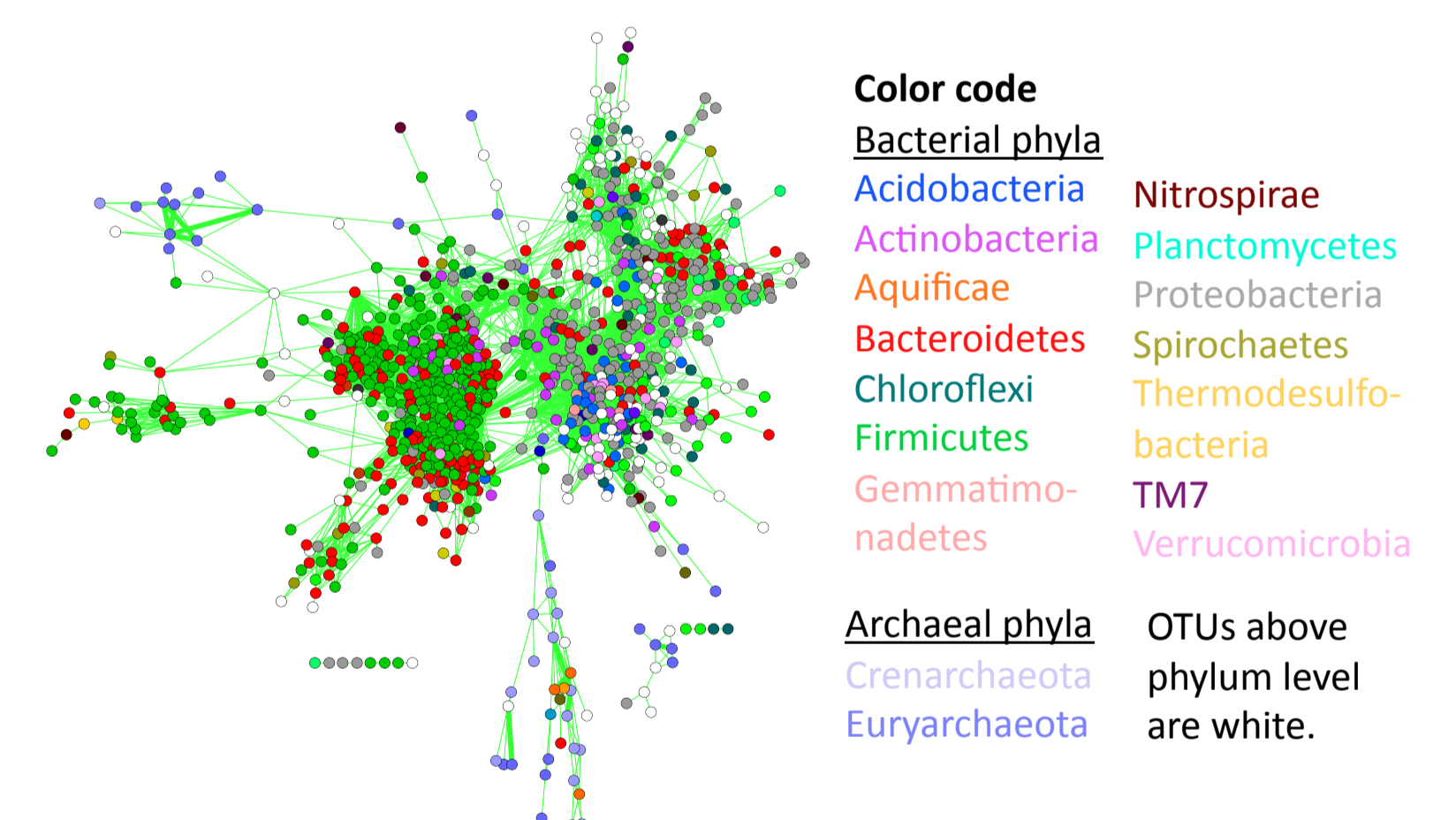
Color code
Bacterial phyla
Acidobacteria
Actinobacteria
Aquificae
Bacteroidetes
Chloroflexi
Firmicutes
Gemmatimo-nadetes

Nitrospirae
Planctomycetes
Proteobacteria
Spirochaetes
Thermodesulfo-bacteria
TM7
Verrucomicrobia

Archaeal phyla
Crenarchaeota
Euryarchaeota

OTUs above phylum level are white.

B) association network obtained for abundance data from human body sites [5]

Legend
mouth
tongue
positive
negative

Method color code
Pearson
Kullback-Leibler
Bray Curtis

*These examples are available as tutorials on the CoNet homepage.

CoNet is a general-purpose association network inference tool that supports a variety of popular similarity-based network inference approaches and offers unique features such as the combination of methods into an ensemble approach. Though CoNet was designed to analyze data from metagenomic experiments, it can infer relationships between any set of repeatedly observed objects (genes, compounds etc.) and could thus prove useful in a variety of settings.

References
[1] Smoot et al. 2011. Bioinformatics, 27: 431-432.
[2] Sathirapongsasuti, Faust et al. 2012. PLoS Computational Biology, 8: e1002606.
[3] Meyer et al. 2008. BMC Bioinformatics 9: 461.
[4] Chaffron et al. 2010. Genome Research 20: 947-959.
[5] Costello et al. 2009. Science 326: 1694-1697.