

Mini-course Compositional Data Analysis

Estimating microbial association networks

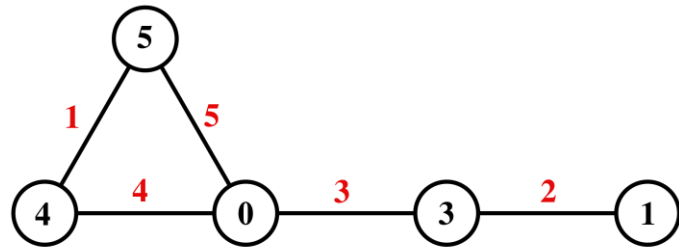
Agenda

1. Introduction to network learning
2. Association estimation
3. The SPIEC-EASI approach and R package
4. From associations to adjacencies
5. Network analysis

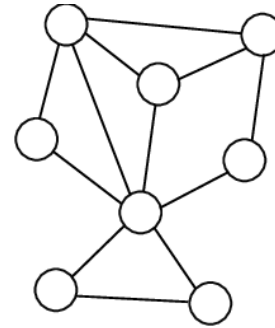
Introduction to network learning

Networks and graphs

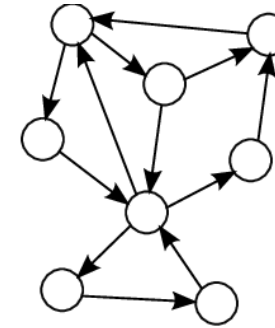
- The term **graph** is used in computer science and math



Graph with labeled **vertices** and **edges**

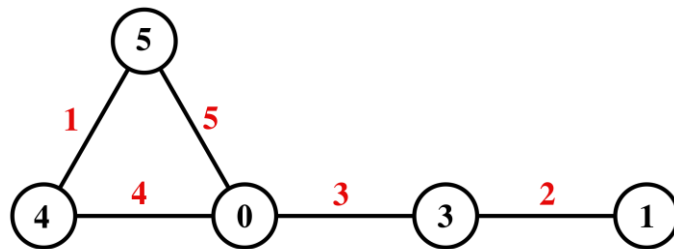


Directed graph



Undirected graph

- The term **network** is used in physics, biology, and social sciences



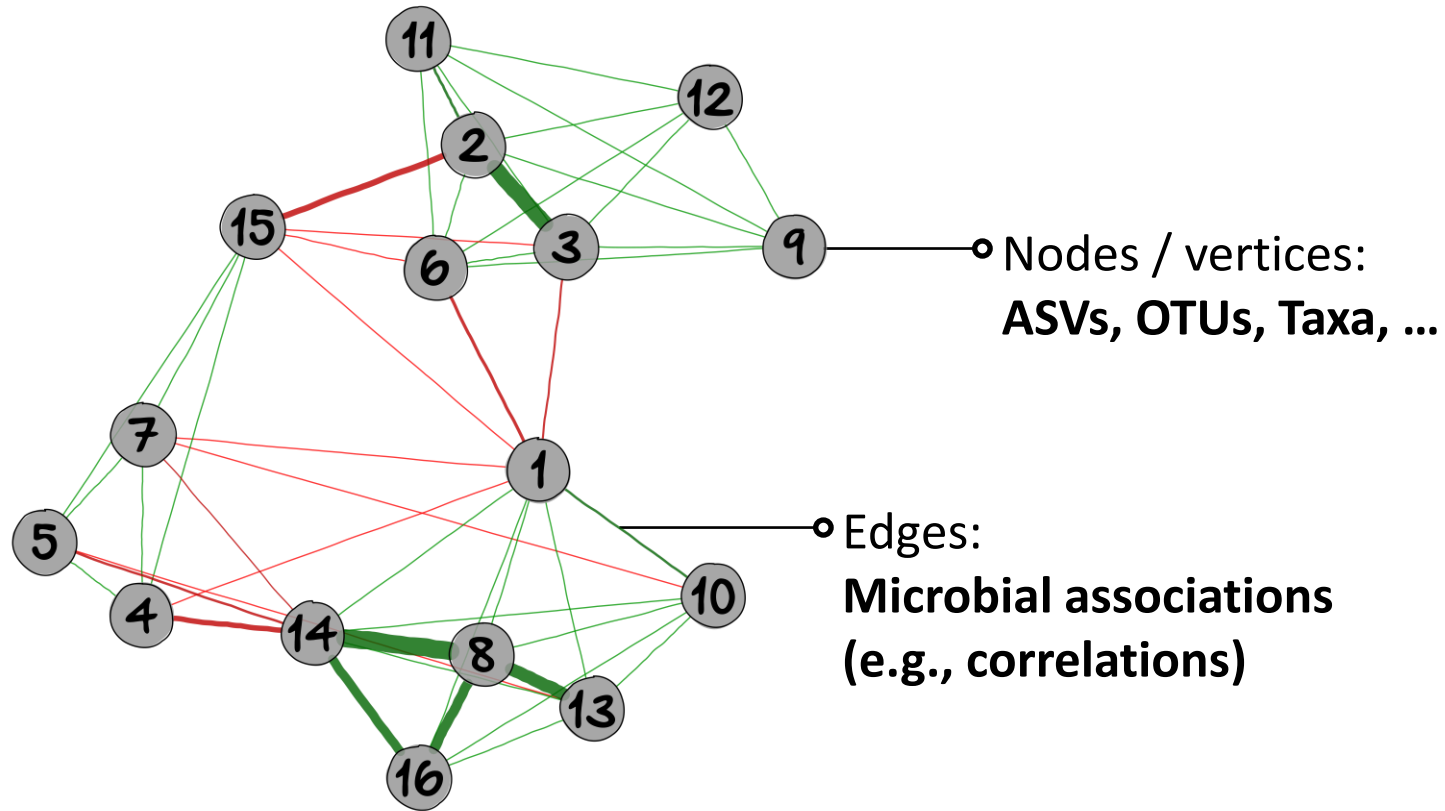
Network with labeled **nodes** and **links**

Networks in cell / molecular biology

- Protein-protein interaction (PPI) networks
- Gene regulatory (co-expression) networks
- Metabolic networks
- Signaling networks
- Neuronal networks
- Microbial (ecological) interaction networks

Microbiome networks

→ Insights into the organizational structure of a microbial community



Association strength and direction:

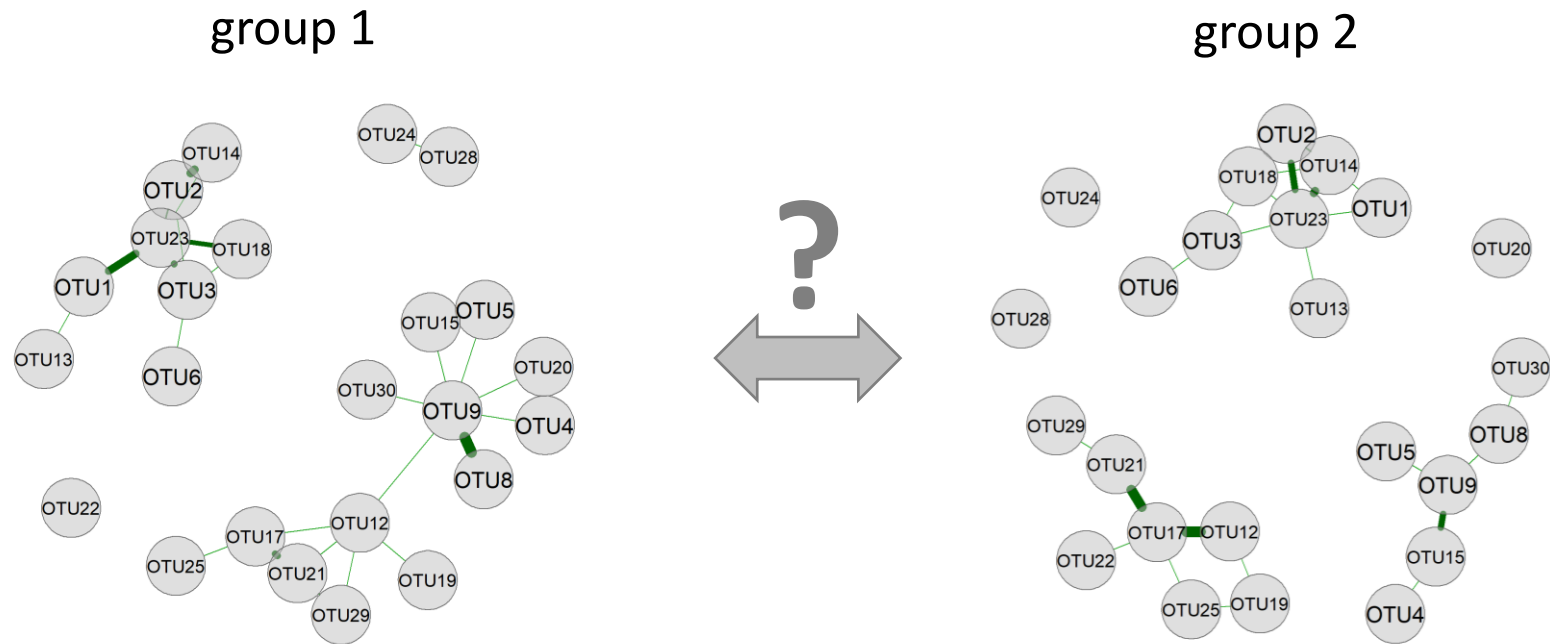
- ▬ Strongly positive
- ▬ Weakly positive
- ▬ Strongly negative
- ▬ Weakly negative

Layout:

Force-directed layout algorithm

Network comparison

→ Does the microbial composition change across different conditions?



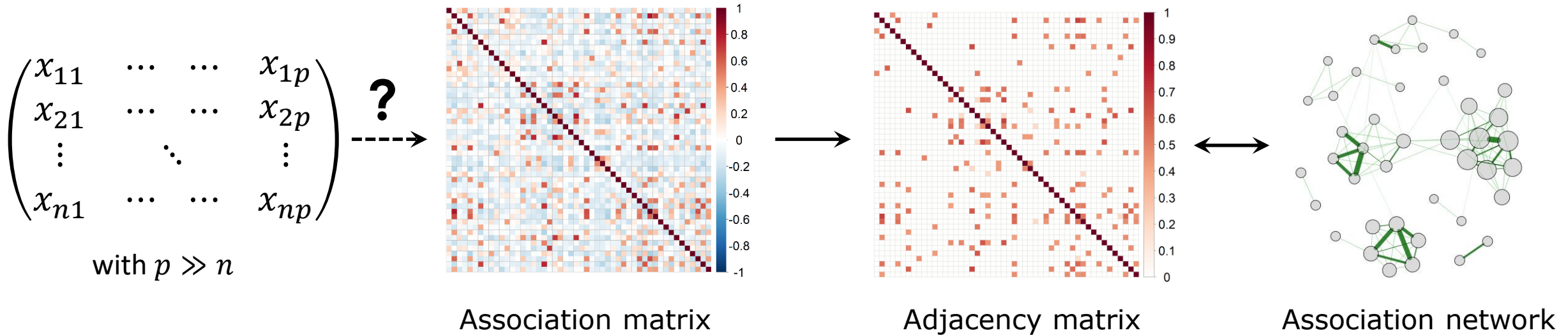
Possible groups:

- Cases and controls
- Two environmental states
- Two time points
- ...

Typical network analysis workflow



From sequencing data to networks



Data characteristics:

- We observe only a **sample** of the true microbial composition
- **Compositional** (only relative information)
- **Sequencing depth** (total number of reads) **varies** across samples
- **Zero-inflated**
- **High-dimensional** (number of taxa $p \gg$ sample size n)

Outlook

In the remainder of this lesson, we will answer the following questions:

- Why is the sample correlation matrix not a good estimator in the high-dimensional case?
- How can we measure microbial associations?
- How can LASSO regression be useful in microbial association estimation?
- How do we get from an association matrix to the final network?
- How can the constructed network be analyzed?