**GO/NO-GO Report: Microbial Association Network**

**Student:** Yiyang Chen

**Project Title: *Inferring Sparse Microbial Association Networks for Plants Under Stress***

1. **Plant Microbiome and Nitrogen Stress**

**1.1 Microbiome's Role in Nitrogen Cycling**

The plant microbiome system plays a vital role in nitrogen cycling, influencing plant access to its vital nutrient. Key microbial processes include:

* **Nitrogen Fixation:** Some bacteria (e.g., *Rhizobia*) convert atmospheric nitrogen gas into ammonia, utilized by plants for growth.
* **Nitrification:** Other microbes (e.g., *Nitrobacter*) oxidize nitrite to nitrate that can be absorbed by plants.
* **Denitrification:** Some bacteria (e.g., *Pseudomonas*) convert nitrate back into nitrogen gas, potentially leading to nitrogen loss from the soil.

The composition and balance of these microbiome community significantly impact nitrogen availability for plants.

**1.2 Microbiome Shifts Under Nitrogen Stress**

Nitrogen deficiency triggers significant shifts in the plant microbiome composition and function:

* **Enrichment of Nitrogen Fixers:** Plants under nitrogen stress often recruit nitrogen-fixing bacteria to enhance nitrogen acquisition.
* **Altered Community Structure:** Overall microbiome diversity and abundance can be altered, with specific taxa better adapted to low nitrogen conditions becoming more prevalent.
* **Functional Adaptation:** Microbial communities under nitrogen stress might exhibit altered gene expression profiles, shifting their metabolic priorities towards nitrogen scavenging and utilization pathways.

Understanding these shifts is crucial for developing strategies to relieve nitrogen stress in plants.

**2. Microbial Association Network Inference**

**2.1 Traditional Methods**

* **Correlation-based Methods:** These rely on calculating pairwise correlations (e.g., Pearson, Spearman) between microbial abundance profiles. Simple to implement but susceptible to false correlations.
* **Bayesian Networks:** Employ probabilistic graphical models to infer causal relationships between microbes. Can capture complex dependencies but computationally demanding.

**2.2 Network Structure and Organization Metrics**

Common metrics to characterize networks include:

* **Node Degree Centrality:** Measures the number of connections a node (microbe) has, reflecting its importance in the network.
* **Betweenness Centrality:** Quantifies how often a node lies on the shortest path between other nodes, indicating its role in connecting different parts of the network.
* **Modularity:** Identifies clusters of nodes (i.e. modules) with dense connections within the module and sparser connections between modules.
  1. **SPRING Framework**

Unlike many methods built for relative abundances, SPRING specifically handles absolute microbial counts, leveraging the richness of quantitative data.

* **Robust to Compositionality and Zero Inflation:**
  + **Semi-parametric Rank-Based Correlation:** Employs a rank-based correlation estimator that is not affected by non-normality and is less sensitive to outliers and zero inflation compared to standard correlation methods.
  + **Modified CLR Transformation:** Accommodates relative abundance data by using a modified centered log-ratio transformation, effectively handling zeros and the unit-sum constraint without needing pseudo-counts.
* **Group-Wise Network Comparisons:**
  + Can be extended with fused lasso or group lasso regularization to construct group-wise networks, encouraging shared sparsity patterns between groups. This allows for powerful comparisons across different conditions (i.e., nitrogen availability) to identify key shifts in microbial interactions.
* **Robustness to Total Abundance Misspecification:**
  + Resile to inaccuracies in total abundance estimates, making it suitable for real-world scenarios where precise cell counts are challenging to obtain.
  1. **Graphical Modeling Approaches:**

Estimate sparse inverse covariance matrices (i.e. precision matrices) representing conditional dependencies. Examples include the Graphical Lasso and its variants. (e.g., FGL, GGL).

**2.4.1 Fused Graphical Lasso (FGL):**

* **Capturing Subtle Differences:** FGL excels at identifying subtle variations in network structures and edge strengths between conditions(K≥2). This is relevant because the project aims to understand how nitrogen availability shapes the network, implying the existence of both shared and unique interaction patterns.
* **Biological Rationale:** Microbiome responses to environmental changes often involve gradual shifts in species abundances and interaction strengths. FGL's ability to model both shared edges and differing edge values aligns well with this biological expectation.
* **Hypothesis Generation:** By highlighting edges that differ significantly between nitrogen conditions, FGL can pinpoint key microbial interactions potentially responsible for plant adaptation to nitrogen stress. This can guide further targeted experimental investigations.

**2.4.2 Group Graphical Lasso (GGL):**

* **Identifying Core Structure:** GGL is ideal for identifying a core set of interactions that remain consistent across both nitrogen conditions. This core network represents the fundamental relationships within the plant microbiome, regardless of nitrogen availability.
* **Robustness to Noise**: By focusing on shared sparsity patterns, GGL can be more robust to noise or variations in individual datasets compared to methods that estimate networks independently for each condition.
* **Comparative Analysis:** After identifying the core structure, GGL allows for comparisons of edge strengths between nitrogen conditions, highlighting interactions that are specifically amplified or weakened under nitrogen stress.