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

**Student:** Yiyang Chen

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

**1. Introduction:**

The plant microbiome system plays a vital role in nitrogen cycling, influencing plant access to its vital nutrient. Some nitrogen fixer bacteria (e.g., Rhizobia) convert atmospheric nitrogen gas into ammonia, utilized by plants for growth, while some microbes (e.g., Nitrobacter) oxidize nitrite to nitrate that can be absorbed by plants. The composition and balance of these microbiome community significantly impact nitrogen availability for plants. The composition and balance of these microbiome community significantly impact nitrogen availability for plants. This research project aims to unravel the microbial interactions within the plant microbiome under varying nitrogen conditions using graphical network analysis, a powerful tool for understanding complex biological systems.

**2. Microbial Association Network Inference:**

Traditional methods for analysing microbiome data often focus on explicitly identifying differentially abundant taxa, neglecting the complex web of interactions within microbial communities. Network inference methods, such as correlation-based approaches (e.g., Pearson, Spearman), Bayesian networks, and probability based graphical modelling (e.g., Graphical Lasso), on the other hand, offer a more holistic view of microbial community dynamics. However, inherent challenges of microbiome data, including zero inflation, undermine such network construction techniques.

**3. SPRING Framework and Extensions:**

This project leverages the SPRING framework, a robust approach for inferring sparse microbial association networks from quantitative abundance data. Unlike many methods built for relative abundances, SPRING specifically handles absolute microbial counts, leveraging the richness of quantitative data. Modified CLR Transformation, utilized by SPRING for preprocessing, accommodates relative abundance data by using modified centred log-ratio transformation, effectively handling zeros and non-zeros without adding pseudo-counts. The semi-parametric rank-based correlation estimator also effectively handles non-normality and zero inflation. Moreover, by integrating with Fused Graphical Lasso (FGL) or Group Graphical Lasso (GGL), it is competent to construct group-wise networks, enabling comparative analyses across nitrogen conditions to identify subtle network variations (via. FGL) or a core set of conserved interactions (via. GGL).

**4. Graphical Modeling Approaches:**

Graphical modelling approaches, such as the Fused Graphical Lasso (FGL) and Group Graphical Lasso (GGL), offer powerful tools for estimating sparse inverse covariance matrices (i.e., Θ), representing conditional dependencies between microbial taxa. FGL is particularly adept at capturing subtle variations in network structure and edge strengths across different conditions, making it ideal for investigating how nitrogen availability shapes microbial interactions and pinpointing key interactions involved in plant adaptation. GGL, on the other hand, focuses on identifying a core set of interactions consistent across conditions, revealing the fundamental relationships within the plant microbiome, and allowing for comparative analysis of edge strengths to highlight interactions specifically affected by nitrogen stress. Both methods contribute to a more robust and nuanced understanding of microbial association networks.

**5. Proposed Deep Learning Approach:**

Recent advancements in deep learning for biological network analysis suggest that integrating Graph Convolutional Networks (GCNs) could significantly enhance network inference. GCNs excel at capturing complex, non-linear relationships in graph-structured data, potentially leading to more accurate and informative microbial association networks. GCNs can leverage both the topological structure of the microbial co-occurrence network and node features (e.g., phylogenetic information, functional annotations) to learn robust representations of microbial interactions.

**6. Validation:**

The reliability of the extended SPRING method, with and without GCN integration, will be assessed by comparing identified key microbial taxa and interactions to known nitrogen stress biomarkers in plants and by evaluating network stability through subsampling.