Using iDIRECT (Inference of Direct and Indirect Relationship with Efficient Copula-based Transitivity) to improve network modelling of rhizosphere microbial assemblages

Naijia Xiao,[1*] (naijia.xiao@ou.edu), Daliang Ning,[1] Aifen Zhou,[1] Megan Kempher,[1] Zhou Jason Shi,[2,3] Benjamin Zhou,[4] Shengjing Shi,[1,5] Erin Nuccio,[5,6] **Mary Firestone[5,7], and Jizhong Zhou[1,7,8]**

[1]University of Oklahoma, Norman; [2]University of California, San Francisco; [3]Chan Zuckerberg Biohub, San Francisco, CA; [4]Stanford University, Stanford, CA; [5]University of California, Berkeley, [6]Lawrence Livermore National Laboratory, Livermore, CA; [7]Lawrence Berkeley National Laboratory, Berkeley, CA; and [8]Tsinghua University, Beijing, China

Project Goals: The overall goals of our project are to advance our understanding of the complex interactions controlling C flow in the rhizosphere by addressing two fundamental topics: 1) How multi-trophic interactions control soil C dynamics, and 2) How changing precipitation regimes alter these interactions, and thus impact flow and fate of soil C.

Networks are a vital tool for understanding and modeling the interactions of complex systems in science and engineering, and direct and indirect interactions are pervasive in all networks including those that occur in rhizosphere soil. However, quantitatively discriminating direct and indirect relationships in association networks remains a formidable task. Here, we present iDIRECT (**I**nference of **D**irect and **I**ndirect **R**elationship with **E**ffective **C**opula-based **T**ransitivity), a novel statistical approach for inferring direct dependences on association networks. Specifically, we first developed new algorithms using Copula-based transitivity, which overcomes several challenging mathematical problems inherent in previous network inference approaches, such as singularity or ill conditioning, self-looping and outside the natural range of [0, 1] in association data. By analyzing simulated data from microbiomes as benchmark examples, an average increase of 11% in prediction accuracy were obtained, demonstrating the effectiveness and robustness of iDIRECT. Compared to several existing approaches, iDIRECT have great advantages with 24-79% up to 2600% improvements in inferring direct relationships in terms of network topology, network size, accuracy and robustness. iDIRECT outperforms previous state-of-the-art approaches for inferring direct relationships. We applied this approach to previously reported bacterial networks that evaluate16S bacterial succession data from *Avena sp.* rhizosphere soil; this analysis shows a change in the composition of dominating modules in several complex networks. We also applied iDIRECT to real gene expression data from *Escherichia coli* and yeast to infer gene regulatory networks. Our results indicated that iDIRECT consistently improves the prediction by up to 12% as compared to the best performing approaches in DREAM5. The iDIRECT approach is clearly applicable to inferring direct relationships in plant root, bacterial, fungal, and mesofaunal association networks.

References

1. Alipanahi, B., & Frey, B. J. (2013). Network cleanup. Nature biotechnology, 31(8), 714.

2. Barzel, B., & Barabási, A. L. (2013). Network link prediction by global silencing of indirect correlations. Nature biotechnology, 31(8), 720-725.
3. Feizi, S., Marbach, D., Médard, M., & Kellis, M. (2013). Network deconvolution as a general method to distinguish direct dependencies in networks. Nature biotechnology, 31(8), 726.
4. Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. PLoS computational biology, 11(5), e1004226.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Numbers DE-SC0010570 and DE-SC0016247 to UC Berkeley and the University of Oklahoma.