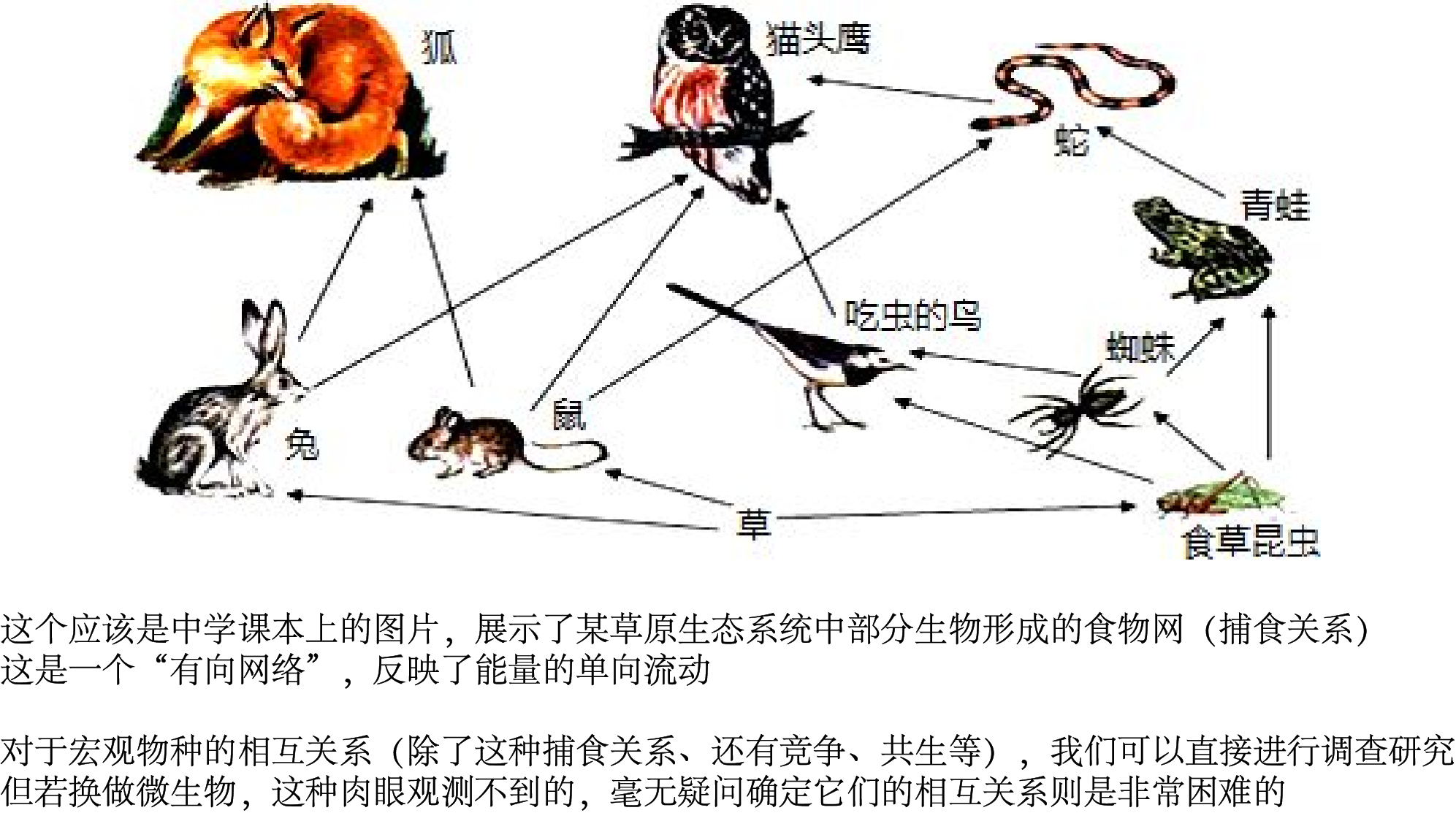
微生物关联网络推断及一个简单的相关网络示例

生物系统通常表示为复杂的二元相互作用集或不同实体之间的关系。系统生物学的目的是在系统层面理解生物实体，除了在单组件（individual components）层面，还包含了互动系统（interacting systems）。与此相关的是网络生物学，其允许使用源自图论的工具来表示和分析生物系统。基本上，每个生物实体都与其他生物实体具有相互作用，从分子层面到生态系统层面，可通过网络反映/模拟生物学过程，如生态、神经、代谢或分子相互作用网络。[网络分析基础可参考前文](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484701&idx=1&sn=c744d7dc780988c537353dc71b795663&chksm=97f5b505a0823c13e65b487699d80c34bd2e5ebef37db7ccf3443e2e75b1b9baf6390d48370e&token=645846601&lang=zh_CN#rd)。

白鱼同学在读研期间是土壤微生物学专业，相对而言对微生物生态学的了解比代谢组、转录组的了解更多（好吧其实对微生物生态学也是皮毛水平），所以本篇整体以微生物网络做个很基础的简介。

# 微生物网络简介

在自然环境中，微生物并不是以分离的个体形式存在，而是通过直接或间接的相互作用形成复杂的共发生网络。微生物之间的相互作用包括互利共生、共栖、寄生、捕食、偏害共栖和竞争等类型，相互作用会对参与者产生正向、负向和中性三种影响。同样地，环境影响群落的组成，因此微生物-环境之间也存在着密切的交互。微生物体积小、数量大、代谢旺盛、繁殖迅速，造成微生物群落在组成和功能上极高的复杂度，然而，在生态系统中直接探究这些不同的相互作用类型非常困难。



## 技术允许

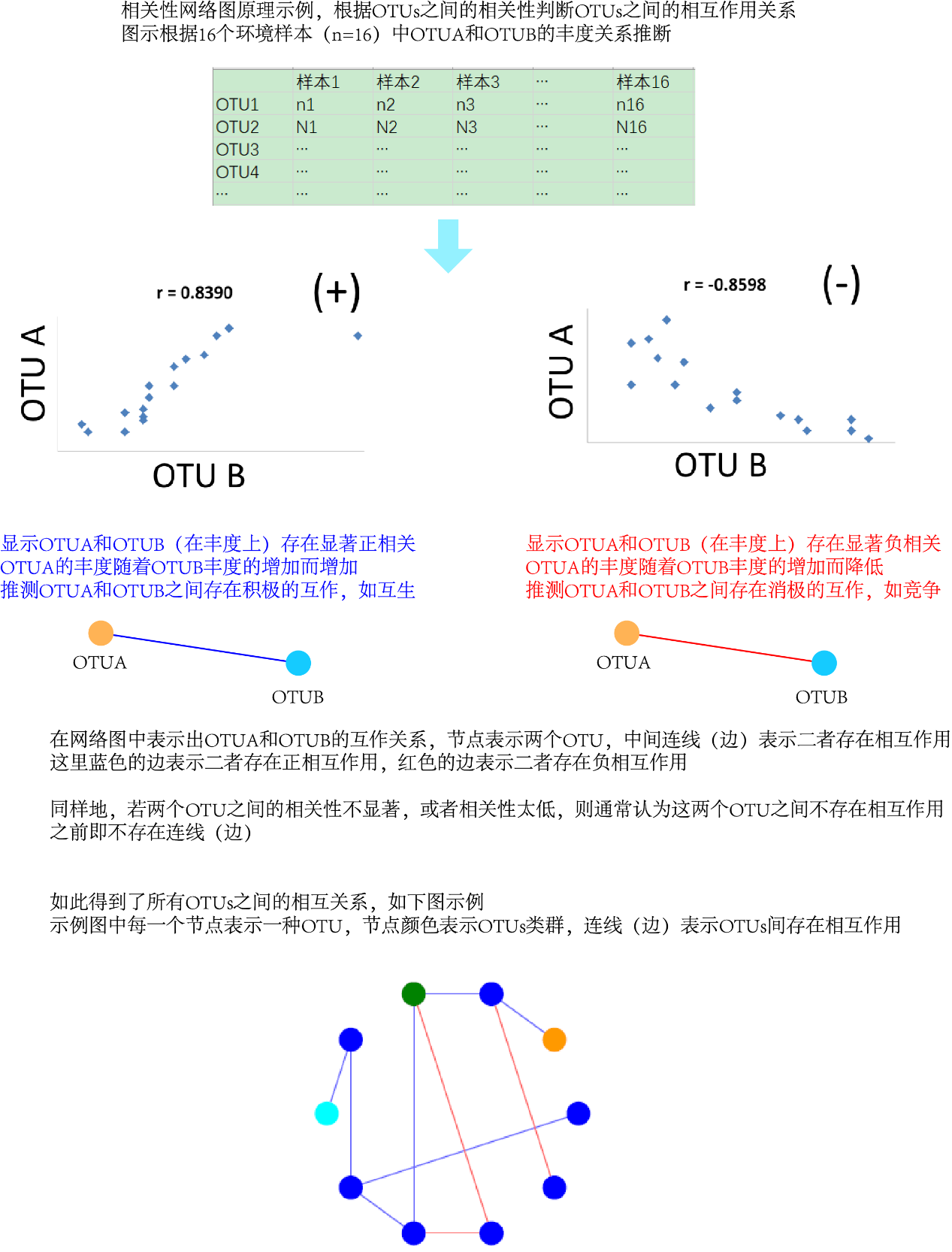
高通量测序数据包含大量种群信息，具有模拟微生物共关系的极大优势。将网络分析与高通量测序相结合，对于发现微生物群落的构建过程和群落的稳定维持中所必需的微生物关系，推断各种相互对宿主健康影响等提供了丰富的研究手段，而这些结果很难通过对生态群落的Alpha多样性或Beta多样性分析来得到（Layeghifard et al. 2017）。

将某个数据集表示为网络图时，定义边的规则通常是两个相邻节点的某些属性具有足够程度的“关联”，例如通过“相关性”建立联系。这种“关联网络”（association network）可以再很多领域中见到，在微生物领域，关联网络分析为微生物生态学的发展提供了强有力的研究手段，并为探索环境中的复杂的微生物-微生物、微生物-环境等相互作用提供了有力的参考依据。近年来，网络分析被广泛应用在多种生态系统中，例如海洋、河流、湖泊、森林、农田、草地、活性污泥和人体，来研究微生物的共发生模式。网络分析能够揭示菌群中非随机的物种共发生模式，使物种间的直接互作或生态位共享特征得到较好的重现，对于理解微生物群落的组建机制、生态系统的功能以及识别群落中的关键物种至关重要。

## 微生物共发生关系推断

如上所述，由于很难直接观测微生物之间的相互作用类型，因此，这里的微生物互作网络实质上是基于数学方法推断的。从最简单的相似性（相关性或距离）计算、到更复杂的多元回归和贝叶斯网络分析，多种方法可以被用来构建微生物共发生网络。这些方法的效率、准确性、速度和计算要求各不相同。其中基于两两物种之间相似性计算的方法具有构建速度快和易于操作的特点，因而最为流行。理论上，任何计算两两之间关系的方法都可以使用。例如：Bray–Curtis距离计算物种丰度之间的距离，Pearson相关系数评估线性关系，Spearman相关系数计算非线性关系等，多元回归R2作为环境对微生物结构的解释程度等。

其中，相关性网络可能是其中最简单、最容易理解的一种方法了，主要依据物种丰度之间的 [Pearson或Spearman相关系数](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484692&idx=1&sn=15e67b1b2f9621a161d1e865c3c9563d&chksm=97f5b50ca0823c1ac57f52dc3ee27bac82df99eda8a87f95a9d737a961a97c1f5152fdb9760c&token=645846601&lang=zh_CN#rd)来构建共发生网络。如下展示了基于OTUs丰度间相关性的关系构建网络图的原理示例。（在16S/18S/ITS扩增子测序中，微生物类群的基本单位称为OTU，即Operational Taxonomic Unit，可以大致理解为一种OTU代表了一种微生物，通常OTU很难定位到种水平，介于种到属之间；最后可获得每个环境样本中都包含了哪些种类的OTU，以及它们的丰度，并在此基础上，继续统计得到各门纲目科属等水平的微生物类群种类和丰度）

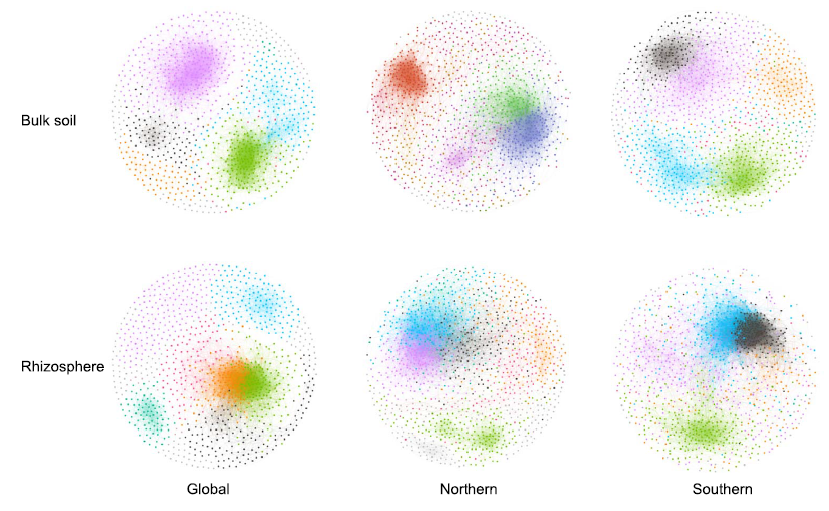


上述大致讲述了相关性网络的构建原理，结果中包含了谁和谁具有相关性、正相关还是负相关、相关性的大小等信息。

更多的网络推断方法，本文不再细说，大家有兴趣可自行查阅相关文献了解。

看到这里构建的微生物互作网络是“无向网络”。毕竟该网络仅为由统计推断得到的，无法像上文中展示的那种“草原生物食物网”一样明确两两物种间的实际相互作用关系，因此通常不指明边的方向。

对于高通量测序所得的细菌群落数据，除了在OTUs水平（如下图所示，大豆根际微生物共发生网络）外，也可以根据实际所要关注的问题，选择在其它水平构建网络，如属水平、门水平等。无论使用哪种水平，能够合理解释生物学现象才是最实际的。



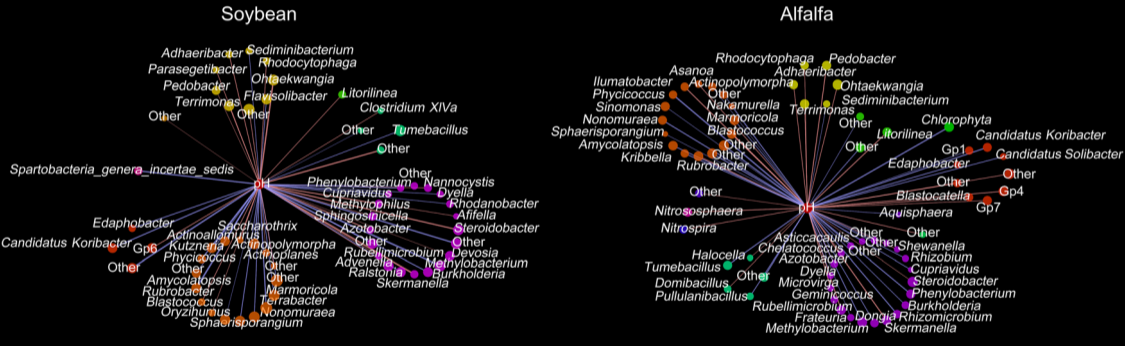
（Zhang et al, 2018）

## 微生物-环境-功能关系的推断

上述示例仅根据微生物的丰度数据来计算，不包含其它的非物种类型的数据，那么这种类型的网络图实际上反映了“物种”间的直接或间接的相互作用。

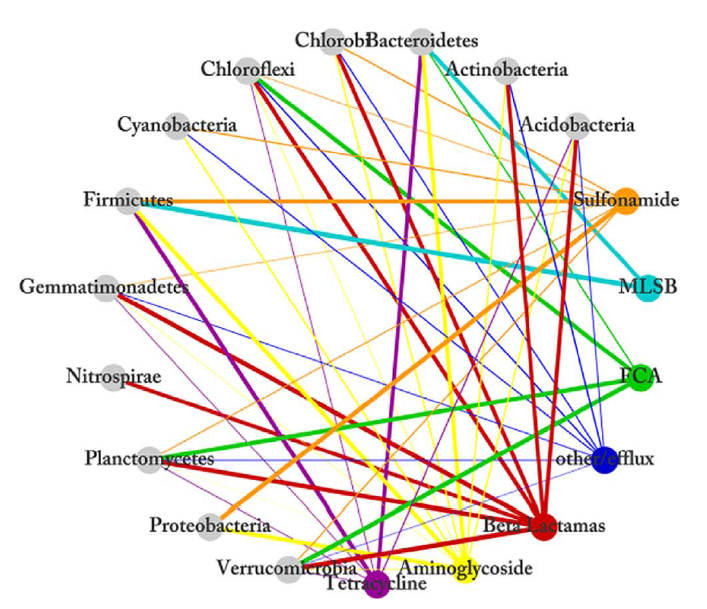
除了识别微生物-微生物之间的互作，同样可将环境因子、功能基因等添加在内，用来推测微生物与环境之间、种群与功能之间的相互关系等。在计算方法上和纯物种数据相比没有本质区别，均是通过某种“关联程度”作为判断的依据。不过相较于上述纯物种的互作网络，在这些网络中可能更侧重于体现物种-环境、物种-功能的关系。

如下示例，使用网络分析了两种植物根际环境中主要的细菌OTUs和土壤理化属性的关系，反映环境对细菌群落组建的影响（图中仅展示了微生物-环境因子的相互作用，不关注微生物-微生物或环境因子-环境因子之间的相互作用）。节点不同的颜色代表对应的细菌OTUs归属于不同的细菌门（同时这些OTUs属于哪些细菌属，直接在节点中标识出属名），边的粗细对应相关性的强弱，红色的边代表呈正相关，蓝色的代表负相关。



（Xiao et al, 2017）

再如下示例，构建相关性网络反映环境中主要的细菌类群和抗生素抗性基因（Antibiotic resistance genes，ARGs）之间的关系，图中各节点反映了不同的细菌门（灰色节点）或抗生素抗性基因大类（彩色节点），连线（边）表示微生物和抗性基因存在相关（图中仅展示了微生物-抗性基因的关联度，不关注微生物-微生物或抗性基因-抗性基因之间的关联度）。边的粗细反映了相关性的大小，即相关性越大，表明该抗性基因更倾向由这些细菌类群所产生。



（Zheng et al, 2018）

## 微生物（生态）网络实际应用举例

网络分析具备多种优点：通用性（数据可以是基因或者物种），可以整合不同的数据类型（物种的相对丰度数据和环境因子），可以适用于分层数据（可以在多个水平如 OTUs、属水平上构建），可以预测群落属性如生态位、核心物种。

网络分析是帮助理解微生物群落组建规则的重要手段之一，能够反映出菌群中可能存在的生态过程，如合作、竞争和生态位分化等，并为更好地认识生态系统中复杂的物种相互作用以及探索群落关键成员提供了新的思路。

以下列举一些应用场景。

**反映生态位共享/重叠程度**

当两个物种同相同的环境因素存在较强的相关性，意味着这两个物种具有一定程度的生态位重叠。同时，微生物共发生网络一般可以被划分成多个模块，模块是网络中高度连接的区域。模块可能反应了栖息地的异质性、系统发育上亲缘关系较近物种的聚集、生态位的重叠和物种的共进化，被认为是系统发育、进化或功能上独立的单元（Olesen et al. 2007）。

除了分析群落物种的共发生模式，网络分析还可以提供微生物群落构建方面的依据。微生物共发生网络的度符合幂律分布，即大部分物种具有少量的连接数，极个别的物种具有非常多的连接数，是典型的无标度网络（Steele et al. 2011），说明微生物群落构建方式是非随机的（Barberan et al. 2012）。共发生（正关联）可能反应出生态位的重叠，而共排斥（负关联）可能由生态位分化所引起。根据这一观点，可以推断非生物的环境选择作用或者生物之间的竞争作用在微生物群落构建中的重要性。当系统发育上亲缘关系相近的物种在观测网络中倾向于共发生，亲缘关系较远的物种倾向于共排斥时，意味着确定性过程在微生物群落构建中占主要地位。反之，当亲缘关系较近的物种倾向于共排斥时，而亲缘关系较远的物种倾向于共发生时，意味着竞争作用占主要地位（Perez-Valera et al. 2017）。Hu等（2017）在研究河流微生物共发生模式时发现，来源于同一纲和科的物种，以及Firmicutes中不同纲的物种倾向于共发生，而Proteobacteria和其他门的物种倾向于共排斥，说明环境选择和竞争作用在微生物群落构建中同时发挥作用。微生物网络结构会随实验条件或者环境改变而发生变化。例如，相比长期施用化肥，施用有机肥的农田其土壤微生物共发生网络结构更加复杂，功能微生物之间联系更加紧密（Ling et al. 2016）；空气中CO2 浓度的增加使得土壤微生物功能基因网络结构更加复杂（Zhou et al. 2011），等等。

**识别关键物种**

在微生物网络中，有一些节点处于枢纽位置，这些核心节点的缺失可能会引起模块和网络的分解，因此在维持网络结构的稳定性中发挥重要作用。这些核心节点常被解读为关键种，在维持微生物群落结构稳定性上可能起重要作用。有三种方法常被用来鉴定这些核心节点。第一种是选择网络中连接数最高的一些节点， Eiler等（2012）据此发现放线菌门acTH1-A1分化枝上的物种是淡水湖泊浮游细菌网络中的核心节点；第二种是选择那些处于网络中心地位的，也就是中介中心性最高的一些节点，Gokul等（2016）据此发现在北极冰盖的细菌网络中，放线菌门的一些OTUs是核心节点；第三种更为复杂，根据节点的模块内连接度（Zi）和模块间连接度（Pi）将网络中所有节点归为四类：外围节点（Peripherals）、连接节点（Connectors）、模块中心点（Module hubs）和网路中心点（Network hubs），通常将连接节点、模块中心点和网络中心点这三类节点归为假定的关键种，Shi等（2016）据此在野燕麦根际网络中寻找核心OTUs。

**网络分析中需要注意的问题**

尽管微生物网络分析据有很多优点，但毕竟只是由数学方法推断得到的，因此真实情况中所有的网络构建方法都不能区分出真正的生物相互作用和非随机的生态过程（例如互养和生态位重叠）。在构建好网络后，可以计算一系列的网络拓扑属性，虽然这些拓扑属性可以简化复杂的网络系统，但解读时如果不加注意，可能会得出错误的结论。例如中心节点常被解读成关键种，去除其之后被认为会引起微生物群落结构发生极大改变。但是中心节点的连接数多可能是由于生态位的重叠引起，在这种情况下，去除并不会引起微生物群落发生较大改变。相反，如果连接数少的节点是食物链顶端的捕食者，其去除会对微生物群落结构产生重大影响。

**网络分析与其它技术的结合**

使用网络分析来研究微生物相互作用带有强烈的推测性，因此越来越多的研究者将其和其他技术相结合应用。共培养是研究微生物相互作用最传统有效的实验手段。共培养实验在受控环境中人为构建群落，因此为验证有关群落稳定性和动态变化的生态理论提供了理想的条件。Harcombe（2010）将*Salmonella Typhimurium*和不能合成蛋氨酸的*Escherichia coli*突变菌株共培养，发现这两株菌不能在醋酸培养基和葡萄糖培养基中生长，而在乳糖培养基上，*S. Typhimurium*可以利用*E. coli*分解乳糖的副产物同时产蛋氨酸供后者利用，说明互利共生可以通过实验进化来获得。传统的共培养技术一次只能研究少量的物种，要同时研究大量物种的关系，需借助其他技术。Lima-Mendez等（2015）利用激光共聚焦显微技术和三维重构技术证实了基于网络分析推测到的微生物间寄生关系。结合网络分析和荧光原位杂交技术，Cardinale等（2015）证实莴苣的根内生细菌间存在很大程度的生态位重叠。

## 微生物网络构建的常用工具以及注意事项

上文提到，无论微生物-微生物互作关系，还是微生物-环境因子、微生物-功能基因等的相互关系，均是根据数学的方法去推断的。在这方面，研究者们开发出了多种工具来构建共发生网络。

**微生物网络构建的常用工具**

常见的如分子生态网络分析（Molecular Ecological Network Analysis，MENA）、局部相似性分析（Local Similarity Analysis，LSA）、成分数据的稀疏关联分析（Sparse Correlations for Compositional data，SparCC）、CoNet分析等；在常用的开源语言R、python中，也有很多被开发用于构建网络的包或模块，如igraph等。

不同的工具构建网络图的算法各有区别，适用情形也不一样。MENA分析将物理学中的随机矩阵理论（Random Matrix Theory，RMT）应用到共发生网络构建中，来自动确定相关系数的阈值（Deng et al. 2012）。在计算时间序列上的关系时，面临着一个问题：一个物种丰度对另一个物种丰度变化的响应需要一定的时间。LSA分析被用来处理这种迟滞效应带来的问题（Ruan et al. 2006）。采用高通量测序技术并不能获得微生物物种的绝对丰度，一般使用均一化后的OTU表，然后根据相对丰度计算任意OTU对之间的相关系数和显著性。OTU相对丰度表中任何一个样品的物种相对丰度之和为1，是非独立的成分数据。使用成分数据计算相关性可能会产生假阳性，也就是成分偏差。SparCC 分析通过计算对数转化后的物种相对丰度之间的Pearson相关系数，来针对性地处理成分偏差问题（riedman et al. 2012）。CoNet采用多种相关性和距离算法计算物种之间的关系，利用置换分析计算每个算法的p并将其合并，在最终的网络中仅保留那些在多个算法中同时显著的关系，同时解决了相似性分析带来的成分偏差问题和距离分析带来的异常值问题（Faust et al. 2012b）。开源语言R、python中提供了非常多的网络构建方法供选择，简单的如基于相关性阈值和显著性p值筛选的网络构建方法，这也是常见的通用方法；更多的网络构建算法可能更为复杂，通常针对于特定的研究（一般可见相关文献中的描述）。话虽如此，但这里的互作网络毕竟仅为根据数学推断，所得结果并非“绝对真实”的相互作用，因此各种方法的适用条件、结果优劣等只是相对而言的。

**微生物网络构建时需要考虑的问题**

大多数分析方法需要人为设定相关系数阈值、显著性阈值等来构建共发生网络，因此研究结果存在一定的主观性。例如在最常见的相关性网络中，对于Pearson或Spearman相关系数阈值的选取，选择0.6或0.7的结果肯定不同，显著性p值阈值选择0.05或0.01的结果肯定也不一样，但是很难说明哪种阈值的选择是更合理的。再例如在CoNet分析中，如果采用自动阈值法，相关性、相异性度量方法的选择，初始连线数的选择等也将显著影响最终结果。总之参数的选择没有确切的衡量标准，主观性是不可避免的。

无论使用哪种网络构建方法，选择了怎样的参数，这些都没有明确的衡量标准，尽管去尝试。毕竟网络分析只是工具手段性质的，只要能通过网络分析合理地解释科学问题/达到预期的效果，那么选择的方法通常就是合适的。

网络构建还要求足够多的样本量，这是毋庸置疑的，这样才能充分体现物种丰度变化的协同作用。就拿相关性网络来讲，参考上文的“原理图”，图中对于OTUA和OTUB互作类型（丰度变化趋势）的探究，根据16个环境样本的群落数据计算二者的相关性，结果清晰明了。假设此处仅有3个样本，那么OTUA和OTUB之间相关性的计算是不是将会大打折扣？类似地，如果预期关注微生物和环境的关系，那么样点还最好覆盖一定的环境变化梯度，短梯度下（环境差异不大，甚至无明显差异）很容易出现偶然误差。当然，与上述网络构建方法、参数选择等相比不同，这里的内容属于试验设计性质的问题了，在进行试验前就一定要慎重决定的。

# 一个简单的相关网络推断示例（R语言计算）

接下来在R中，展示一种最简单的方法，就是上文提到的基于相关系数构建的微生物网络。其它方法，如MENA、CoNet、SparCC等，以及其它领域的工具，例如计算基因共表达网络的方法如WGCNA，将在后面的推文中一一讲解。

示例数据和R代码链接（提取码：xto3）：

<https://pan.baidu.com/s/1OjBRx5mbs_ybNi5LjoU9aw>

## 微生物-微生物相关网络

首先展示使用丰度数据，推断微生物-微生物互作的相关性网络。

**相关系数矩阵计算**

首先计算微生物丰度间的相关系数。

##计算微生物丰度间的相关系数

library(Hmisc)

#以属水平丰度为例，“genus\_table.txt” 是一个属水平的微生物丰度表

genus <- read.delim('genus\_table.txt', row.name = 1, check.names = FALSE)

#可选事先过滤一些低丰度或低频的类群

genus <- genus[which(rowSums(genus) >= 0.005), ] #例如只保留相对丰度总和高于 0.005 的属

genus1 <- genus

genus1[genus1>0] <- 1

genus <- genus[which(rowSums(genus1) >= 5), ] #例如只保留在 5 个及以上样本中出现的属

#计算两属之间是否存在丰度变化的相关性，以 spearman 相关系数为例

genus\_corr <- rcorr(t(genus), type = 'spearman')

#阈值筛选

#将 spearman 相关系数低于 0.7 的关系剔除，即 r>=0.7

r <- genus\_corr$r

r[abs(r) < 0.7] <- 0

#选取显著性 p 值小于 0.05 的相关系数，即 p<0.05

p <- genus\_corr$P

p <- p.adjust(p, method = 'BH') #可选 p 值校正，这里使用 BH 法校正 p 值

p[p>=0.05] <- -1

p[p<0.05 & p>=0] <- 1

p[p==-1] <- 0

#根据上述筛选的 r 值和 p 值保留数据

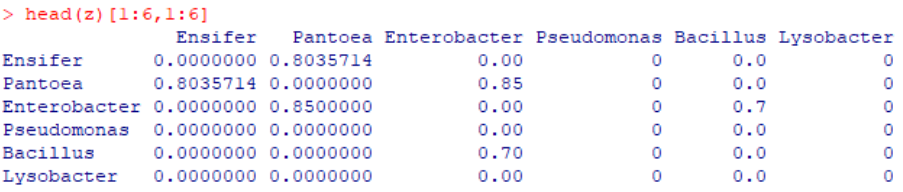
z <- r \* p

diag(z) <- 0 #将相关矩阵中对角线中的值（代表了自相关）转为 0

head(z)[1:6,1:6]

#如此便得到了邻接矩阵格式的网络文件（微生物属的相关系数矩阵）

write.table(data.frame(z, check.names = FALSE), 'genus\_corr.matrix.txt', col.names = NA, sep = '\t', quote = FALSE)



获得了微生物属丰度间的相关系数矩阵，已剔除了相关系数过低且不显著的值。

**igraph构建网络**

该相关系数矩阵可作为[邻接矩阵](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484701&idx=1&sn=c744d7dc780988c537353dc71b795663&chksm=97f5b505a0823c13e65b487699d80c34bd2e5ebef37db7ccf3443e2e75b1b9baf6390d48370e&token=645846601&lang=zh_CN#rd)类型的网络输入文件，通过它转化为igraph的网络格式。

##获得网络

library(igraph)

#将邻接矩阵转化为 igraph 网络的邻接列表

#构建含权的无向网络，权重代表了微生物属间丰度的 spearman 相关系数

g <- graph.adjacency(z, weighted = TRUE, mode = 'undirected')

g

#自相关也可以通过该式去除

g <- simplify(g)

#孤立节点的删除（删除度为 0 的节点）

g <- delete.vertices(g, names(degree(g)[degree(g) == 0]))

#该模式下，边权重代表了相关系数

#由于权重通常为正值，因此最好取个绝对值，相关系数重新复制一列

E(g)$correlation <- E(g)$weight

E(g)$weight <- abs(E(g)$weight)

#为节点（微生物属）添加属性信息（界门纲目科属水平注释）

#“genus\_taxonomy.txt” 记录了微生物的属性，读入该表后根据已知网络节点匹配对应的行

tax <- read.delim('genus\_taxonomy.txt', row.name = 1, check.names = FALSE, stringsAsFactors = FALSE)

tax <- tax[as.character(V(g)$name), ]

V(g)$kingdom <- tax$kingdom

V(g)$phylum <- tax$phylum

V(g)$class <- tax$class

V(g)$order <- tax$order

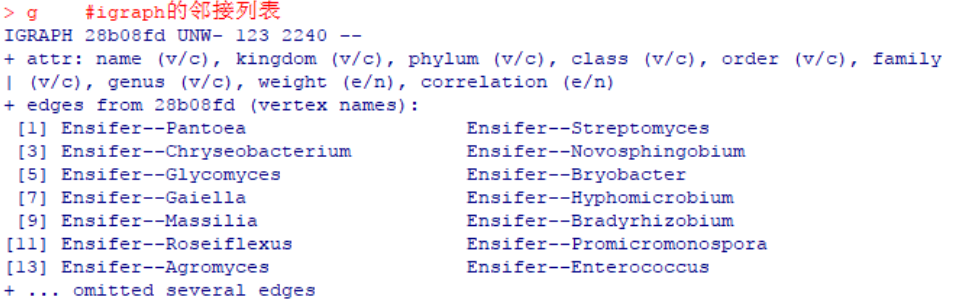
V(g)$family <- tax$family

V(g)$genus <- tax$genus

#查看网络图

g

plot(g)



**网络拓扑特征计算**

略，可参考前文等描述：

[微生物共发生网络中节点度的幂律分布](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484722&idx=1&sn=cc04698aa90a477c917576421f584e89&chksm=97f5b52aa0823c3cf22fd50712bfcfad7fdc3e0ccecb8bb86ff662e71bd62d693ca4c4ff9557&token=645846601&lang=zh_CN#rd)

网络拓扑结构-节点和边特征

网络拓扑结构-网络图的凝聚性特征

[网络模块内连通度（Zi）和模块间连通度（Pi）](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484722&idx=2&sn=0adb81212e38015236be940d1423377d&chksm=97f5b52aa0823c3ce89202bd152e25ae866488a87f51ebe1f8d754d6c24d420aa4fea1da414e&token=645846601&lang=zh_CN#rd)

**文件输出**

考虑将构建好的网络保存，如常见的[邻接矩阵](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484701&idx=1&sn=c744d7dc780988c537353dc71b795663&chksm=97f5b505a0823c13e65b487699d80c34bd2e5ebef37db7ccf3443e2e75b1b9baf6390d48370e&token=645846601&lang=zh_CN#rd)或[边列表](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484701&idx=1&sn=c744d7dc780988c537353dc71b795663&chksm=97f5b505a0823c13e65b487699d80c34bd2e5ebef37db7ccf3443e2e75b1b9baf6390d48370e&token=645846601&lang=zh_CN#rd)、[节点属性列表](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484701&idx=1&sn=c744d7dc780988c537353dc71b795663&chksm=97f5b505a0823c13e65b487699d80c34bd2e5ebef37db7ccf3443e2e75b1b9baf6390d48370e&token=645846601&lang=zh_CN#rd)格式等。

##网络文件输出，输出特定的网络文件类型，便于后续数据分析需求

#邻接矩阵，出了上述提到的在计算相关系数后，输出筛选后的相关系数矩阵外

#还可以由 igraph 的邻接列表转换

adj\_matrix <- as.matrix(get.adjacency(g, attr = 'correlation'))

write.table(data.frame(adj\_matrix, check.names = FALSE), 'network.adj\_matrix.txt', col.names = NA, sep = '\t', quote = FALSE)

#边列表

edge <- data.frame(as\_edgelist(g)) #igraph 的邻接列表转为边列表

edge\_list <- data.frame(

source = edge[[1]],

target = edge[[2]],

weight = E(g)$weight,

correlation = E(g)$correlation

)

head(edge\_list)

write.table(edge\_list, 'network.edge\_list.txt', sep = '\t', row.names = FALSE, quote = FALSE)

#节点属性列表

node\_list <- data.frame(

label = names(V(g)),

kingdom = V(g)$kingdom,

phylum = V(g)$phylum,

class = V(g)$class,

order = V(g)$order,

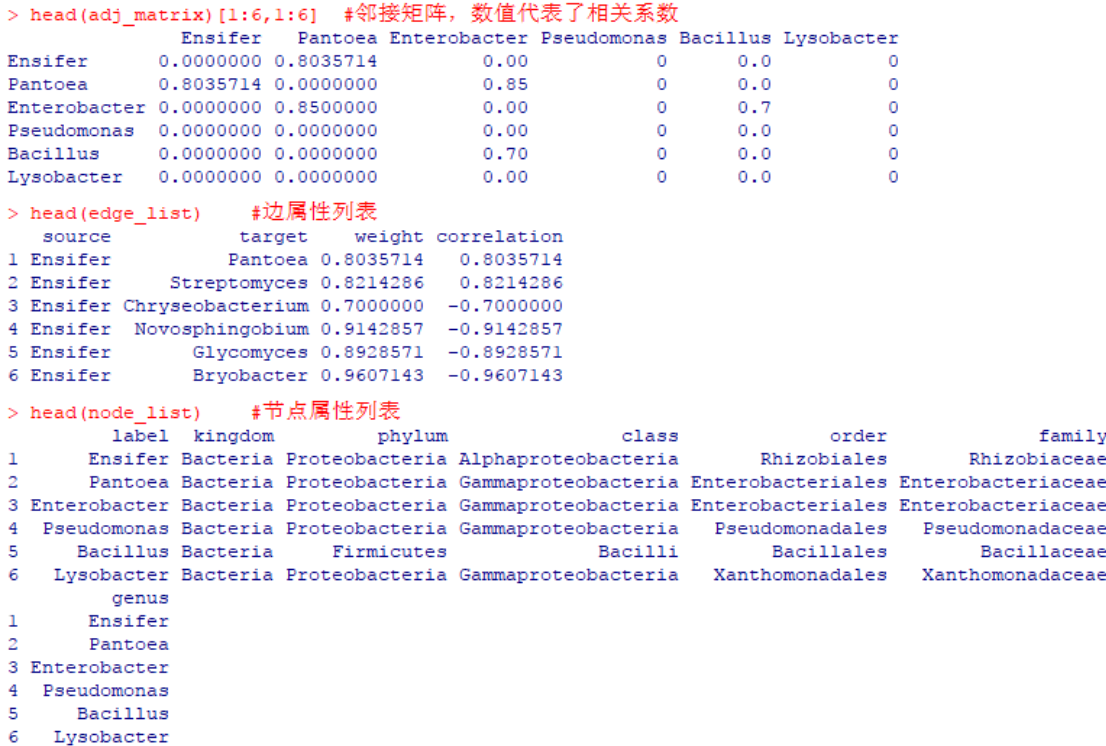
family = V(g)$family,

genus = V(g)$genus

)

head(node\_list)

write.table(node\_list, 'network.node\_list.txt', sep = '\t', row.names = FALSE, quote = FALSE)



以及一些特殊的文件格式，可以输入至其它软件（如cytoscape）中使用。

#边列表节点属性列表可以导入至 gephi 或 cytoscape 等网络可视化软件中进行编辑

#此外 igraph 也提供了可以被 gephi 或 cytoscape 等直接识别的格式

#graphml 格式，可使用 gephi 软件打开并进行可视化编辑

write.graph(g, 'network.graphml', format = 'graphml')

#gml 格式，可使用 cytoscape 软件打开并进行可视化编辑

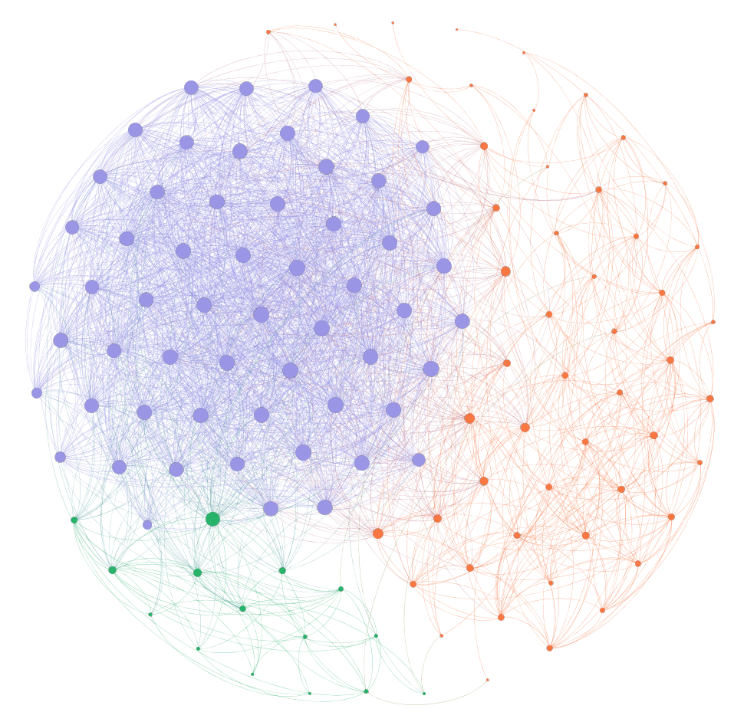
write.graph(g, 'network.gml', format = 'gml')

**网络可视化**

对于网络可视化，我个人不推荐使用R来进行，尽管也有很多优秀的网络可视化R包。

归因于网络要素过多，节点间的结构通过代码调整真的很麻烦，相比之下图形化的操作界面，鼠标可拖动的样式就很容易操作。

例如Gephi的可视化，过程本篇略，Gephi的可视化方法将来我会录视频……



## 微生物-功能、微生物-环境相关网络

再展示一个探索微生物和功能相关的情况。

微生物和环境的关系推断，和该方法类似，仿照来即可，就不再展示了。

##计算微生物类群丰度和功能基因丰度的相关系数

library(Hmisc)

#以门水平丰度为例，“phylum\_table.txt” 是一个门水平的微生物丰度表

phylum <- read.delim('phylum\_table.txt', row.name = 1, check.names = FALSE)

#“ARGs\_table.txt”是一个抗生素抗性基因丰度表

ARGs <- read.delim('ARGs\_table.txt', row.name = 1, check.names = FALSE)

#计算群落组成与功能的相关性，以 spearman 相关系数为例

phylum\_ARGs\_corr <- rcorr(as.matrix(phylum), as.matrix(ARGs), type = 'spearman')

#相关系数 r 值和显著性 p 值矩阵

r <- phylum\_ARGs\_corr$r

p <- phylum\_ARGs\_corr$P

#只保留微生物丰度-功能基因丰度的相关系数

#去除微生物-微生物、功能基因-功能基因之间的相关系数

r <- r[colnames(phylum),colnames(ARGs)]

p <- p[colnames(phylum),colnames(ARGs)]

#阈值筛选

#将 spearman 相关系数低于 0.7 的关系剔除，即 r>=0.7

#该模式下，一定要注意负值的选择是否是合适的，因为很多情况下可能负相关无意义

r[abs(r) < 0.7] <- 0

#选取显著性 p 值小于 0.05 的相关系数，即 p<0.05

p <- p.adjust(p, method = 'BH') #可选 p 值校正，这里使用 BH 法校正 p 值

p[p>=0.05] <- -1

p[p<0.05 & p>=0] <- 1

p[p==-1] <- 0

#根据上述筛选的 r 值和 p 值保留数据

z <- r \* p

#再转换为对称矩阵，igraph 只能识别这种样式的邻接矩阵类型

z1 <- phylum\_ARGs\_corr$r

z1[z1 != 0] <- 0

z1[rownames(z),colnames(z)] <- z

z1[colnames(z),rownames(z)] <- z

#write.table(data.frame(z1, check.names = FALSE), 'phylum\_ARGs\_corr.matrix.txt', col.names = NA, sep = '\t', quote = FALSE)

##获得网络

library(igraph)

#将邻接矩阵转化为 igraph 网络的邻接列表

#构建含权的无向网络，权重代表了微生物丰度和功能基因丰度间的 spearman 相关系数

g <- graph.adjacency(z1, weighted = TRUE, mode = 'undirected')

g

#孤立节点的删除（删除度为 0 的节点）

g <- delete.vertices(g, names(degree(g)[degree(g) == 0]))

#该模式下，边权重代表了相关系数

#由于权重通常为正值，因此最好取个绝对值，相关系数重新复制一列

E(g)$correlation <- E(g)$weight

E(g)$weight <- abs(E(g)$weight)

#查看网络图

plot(g)

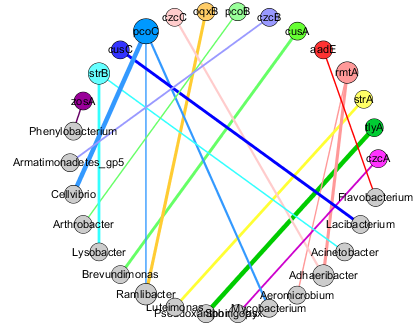
##网络文件输出，略，邻接矩阵、边列表等获取方式参考上文即可

#例如 gml 格式，可使用 cytoscape 软件打开并进行可视化编辑

write.graph(g, 'network.gml', format = 'gml')

更多细节部分这里不再展示了，参考上文的微生物-微生物相关网络即可。

最后展示Cytoscape的可视化示例，过程本篇略，Cytoscape的可视化方法将来我也会录视频……



# 参考资料

师兄的博士论文（师兄总结得太好了，很多段话直接复制粘贴的，感谢师兄允许，就是可惜师兄不让提到他名字和论文标题……）

Barberán, Albert, Bates S T, Casamayor E O , et al. Using network analysis to explore co-occurrence patterns in soil microbial communities. The ISME Journal, 2012, 6(2):343-351.

Cardinale M, Grube M, Erlacher A, et al. Bacterial networks and co-occurrence relationships in the lettuce root microbiota. Environmental Microbiology, 2015, 17(1):239-252.

Eiler A, Heinrich F, Bertilsson S. Coherent dynamics and association networks among lake bacterioplankton taxa. The ISME Journal, 2012, 6(2):330-342.

Gokul J K, Hodson A J, Saetnan E R, et al. Taxon interactions control the distributions of cryoconite bacteria colonizing a High Arctic ice cap. Molecular Ecology, 2016.

Hu A, Ju F, Hou L, et al. Strong impact of anthropogenic contamination on the co-occurrence patterns of a riverine microbial community. Environmental Microbiology, 2017.

Layeghifard M, Hwang D M, Guttman D S . Disentangling Interactions in the Microbiome: A Network Perspective. Trends in Microbiology, 2017, 25(3):217-228.

Lima-Mendez G, Faust K, Henry N, et al. Ocean plankton. Determinants of community structure in the global plankton interactome. Science, 2015, 348(6237):1262073.

Ling N, Zhu C, Xue C, et al. Insight into how organic amendments can shape the soil microbiome in long-term field experiments as revealed by network analysis. Soil Biology and Biochemistry, 2016, 99:137-149.

Olesen J M, Bascompte J, Dupont Y L, et al. The modularity of pollination networks. Proc Natl Acad Sci U S A, 2007, 104(50):19891-19896.

Pérez-Valera, Eduardo, Goberna M, et al. Fire modifies the phylogenetic structure of soil bacterial co-occurrence networks. Environmental Microbiology, 2016.

Shi S, Nuccio E E, Shi Z J, et al. The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. Ecology Letters, 2016, 19(8):926-936.

Steele J A, Countway P D, Xia L, et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. The ISME Journal, 2011, 5(9):1414-1425.

Zhou J, Deng Y, Luo F, et al. Phylogenetic Molecular Ecological Network of Soil Microbial Communities in Response to Elevated CO2. mBio, 2011, 2(4):e00122-11-e00122-11.