

ggraph 优雅的绘制网络图

阎俊安 2022-01-07

目录

1 安装并加载 R 包	1
2 导入数据	2
3 数据拆分	2
4 数据清洗	2
5 计算相似度	3
6 整合边文件与点文件	3
7 数据格式转换	3
8 绘制网络图	4

本文档主要来介绍如何使用 **ggraph** 包来绘制网络图，示例数据为微生物 16S 测序所生成的 OTU 表，通过对 OTU 表格的拆分整合使用门水平的信息来绘制网络图，下面来看具体代码

1 安装并加载 R 包

```
package.list=c("tidyverse","ggraph","tidygraph","magrittr","tidytext","widyr")

for (package in package.list) {
  if (!require(package,character.only=T, quietly=T)) {
    install.packages(package)
    library(package, character.only=T)
  }
}
```

2 导入数据

导入了 OTU 表，并针对 **taxonomy** 列内容以; 为分割符号将其拆分为 7 列

```
otu <- read_tsv("otu_taxa_table.xls") %>%
  separate(taxonomy,
           into=c("domain","phylum","class","order","family",
                  "genus","species"),sep=";") %>%
  mutate_at(vars(c(`domain`,`species`)),~str_split(.,"_",simplify=TRUE)[,2]) %>%
  column_to_rownames("OTU")

otu %>% select(1:13) %>% head() %>% knitr::kable()
```

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3
OTU1	1005	1023	1137	4522	2672	35	18	2	30	40	0	1	60
OTU2	5518	5454	4204	4734	4087	429	725	237	1786	2244	164	362	119
OTU3	12	45	122	53	22	66	142	87	83	75	2594	535	1882
OTU4	107	10	18	20	9	5823	2621	12299	7560	1279	372	1220	572
OTU5	2080	1782	2245	1830	1645	443	359	218	579	603	479	3	398
OTU6	1	59	0	8	15	48	78	73	33	11	2750	540	1727

3 数据拆分

将 OTU 表丰度与注释信息进行拆分

```
table <- otu %>% select_if(~is.numeric(.)) %>% rownames_to_column("ID")
tax <- otu %>% select_if(!is.numeric(.)) %>% rownames_to_column("ID")
```

4 数据清洗

```
titles_node <- table %>% left_join(.,tax %>% select(1,phylum),by="ID") %>%
  select(-ID) %>%
  group_by(phylum) %>%
  summarise(across(where(is.numeric), ~ sum(.x, na.rm=TRUE))) %>% select(1) %>%
  distinct(phylum) %>%
  filter(phylum!="") %>%
  rowid_to_column("id")
```

```
title_description_tf_idf <- table %>% left_join(.,tax %>% select(1,phylum),by="ID") %>%
  select(-ID) %>%
  group_by(phylum) %>%
  summarise(across(where(is.numeric), ~ sum(.x, na.rm=TRUE))) %>%
  pivot_longer(-phylum) %>%
  filter(phylum!="",value!=0) %>%
  set_colnames(c("title","word","n")) %>%
  bind_tf_idf(word, title , n)
```

5 计算相似度

```
title_similarity <- title_description_tf_idf %>%
  pairwise_similarity(title, word, tf_idf, sort = TRUE)
```

6 整合边文件与点文件

```
titles_edges <- title_similarity %>%
  left_join(titles_node, by = c("item1" = "phylum")) %>%
  rename(from = id)

titles_edges %<>%
  left_join(titles_node, by = c("item2" = "phylum")) %>%
  rename(to = id)

titles_edges %<>% select(from, to, similarity)
```

7 数据格式转换

```
titles_graph <- tbl_graph(nodes = titles_node,
  edges = titles_edges,directed = TRUE)
```

```
titles_graph
```

```
## # A tbl_graph: 26 nodes and 466 edges
```

```
## #
```

```
## # A directed simple graph with 1 component
```

```
## #
## # Node Data: 26 x 2 (active)
##      id phylum
##      <int> <chr>
## 1      1 Acidobacteria
## 2      2 Actinobacteria
## 3      3 Armatimonadetes
## 4      4 Ascomycota
## 5      5 Bacteroidetes
## 6      6 Calditrichaeota
## # ... with 20 more rows
## #
## # Edge Data: 466 x 3
##      from      to similarity
##      <int> <int>      <dbl>
## 1      24      12      0.996
## 2      12      24      0.996
## 3      20      12      0.935
## # ... with 463 more rows
```

8 绘制网络图

```
ggraph(titles_graph, layout = "fr") +
  geom_edge_fan0(aes(color = similarity)) +
  geom_node_point(shape=21, fill = "green", color="white", size=3) +
  geom_node_text(aes(label = phylum), vjust = -1, hjust=0.5, size=3.5, color="black") +
  scale_edge_color_gradientn(colours=colorRampPalette(c("#5eaf5", "#f4d963", "red"))(10),
                             na.value="grey80") +
  theme_graph() +
  theme(legend.title = element_blank())
```

