

实验笔记

## 1 数据概要

### 1.1 老鼠数据

表 1: Notations

PART	VALUE	DESCRIPTION
$n_1$	1376	基因 - 表型关联矩阵中的基因数目
$m$	5420	基因 - 表型关联矩阵中的表型数目
$n_2$	5985	PPI 基因 - 基因关联矩阵中的基因数目
$n_3$	7754	pathway 中所有出现的基因数目
$n_1 \cap n_2$	670	基因 - 表型关联矩阵与 PPI 网络中共有的基因数目
$n_1 \cap n_3$	790	基因 - 表型关联矩阵与 pathway 中共有的基因数目
$n_1 \cap n_2 \cap n_3$	441	基因 - 表型关联矩阵, PPI, pathway, 中共有的基因数目

获取基因的原则：基因或者出现在 PPI 网络中，或者出现在 pathway 中，这样在目标函数中便有约束可以作用在所出现的基因上面。

- 如果选择 670 个基因作为基准基因集合，则会有 229 (670-441) 个基因没有出现在任何 pathway 中，在计算目标函数与梯度过程中，这 229 个基因并不影响计算，在遍历 pathway 时，只考虑那 441 个基因便可。因此我们应将每个 pathway 下的基因重新筛选一次，每条 pathway 下，只用这 441 个基因来表示即可。
- 如果选择 790 个基因作为基准基因集合，则会有 349 (790-441) 个基因没有出现在 PPI 网络中，此时 PPI 网络要保留为  $790 \times 790$  大小，也就是说在 PPI 网络中有 349 个基因与其他基因是没有任何关联的。

## 2 具体实验流程

选择 670 个基因作为基准基因集合的实验流程：

---

**Algorithm 1 具体实验流程**

---

- 1: 获取 670 个基因的基因列表，用这 1 到 670 作为基因索引
- 2: 用新的 670 个基因索引表示基因 -表型网络和基因 PPI 网络
- 3: 在 670 个基因索引中进一步挑选出 441 个基因的索引列表，用于表示每个 pathway 下的基因

---

选择 790 个基因作为基准基因集合的实验流程：

---

**Algorithm 2 具体实验流程**

---

- 1: 获取 790 个基因的基因列表，用这 1 到 790 作为基因索引
- 2: 用新的 790 个基因索引表示基因 -表型网络和基因 pathway
- 3: 在 790 个基因索引中进一步挑选出 441 个基因的索引列表，用于表示 PPI 网络，实际上在 790 个基因组成的 PPI 网络中，有 349 个基因与其他基因是没有任何关联的

---