

The interdependent network of gene regulation and metabolism is robust where it needs to be

姓名：姜震波

学院：生命科学学院

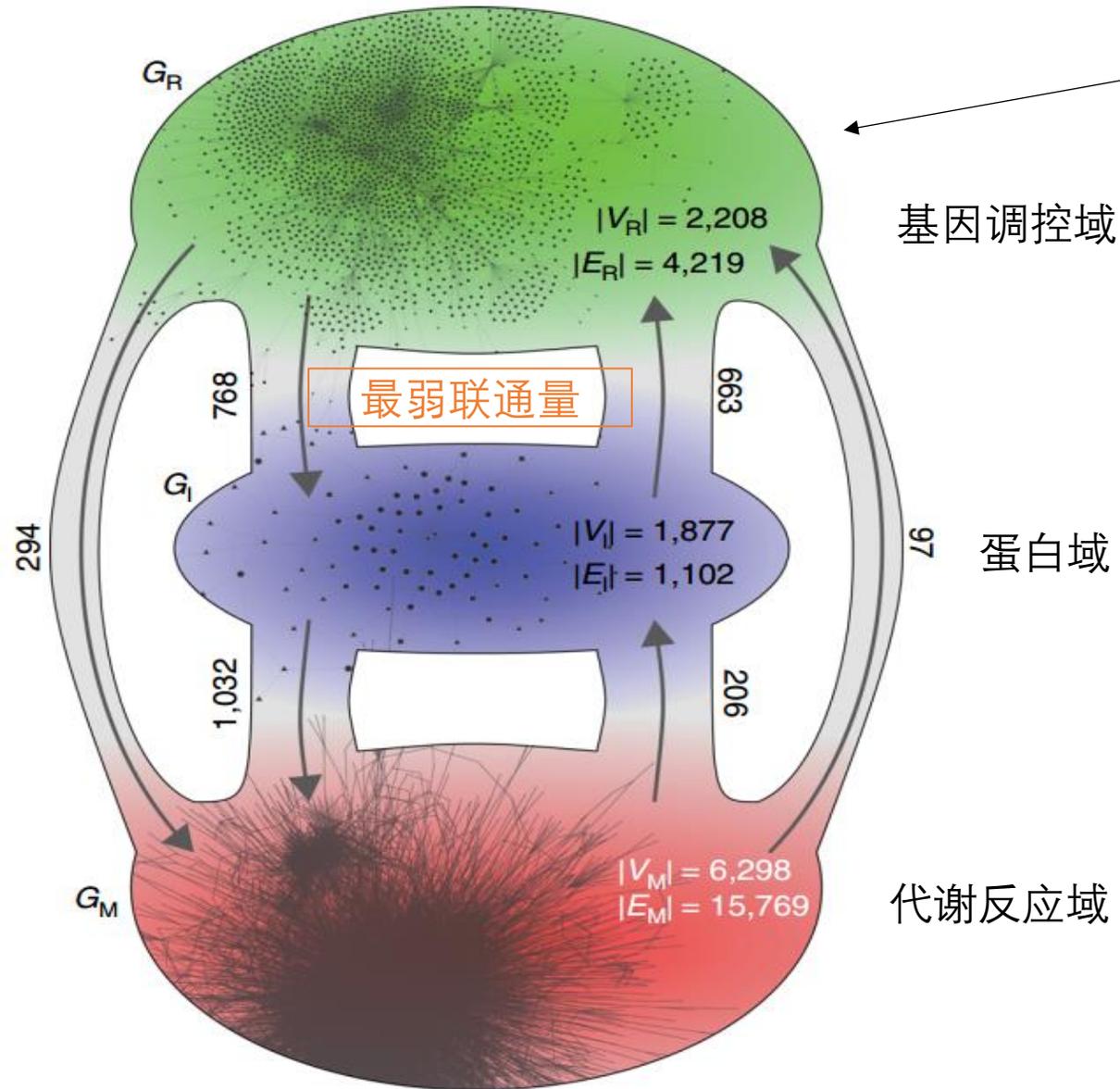
Background :

- 基因调控和代谢反应网络存在着高度相互依赖性，但是经常被独立研究，其实它们是可以通过各种形式的蛋白质联系到一起（举例），这篇文章就是想通过系统生物学的手段对这一组和体系施加扰动来分析相互依赖网络的稳健性

网络：

- 大肠杆菌的整合网络已经使用EcoCyc数据库进行组装，其中提供了大量的有关代谢过程和基因调控的数据。广泛的元数据可以将顶点分配给三个功能域。相应的图形表示由10,383个顶点24,150个有向边组成。

大肠杆菌域组织网络图



为了方便阐述扰动的影响，引入了公式：

$$p=1-q$$

P为占有概率（不受影响区域所占百分比）

q为互补量（扰动影响区域百分比）

同理：

$$p_c=1-q_c$$

P_c为渗透阈值（大小决定稳健性）

Q_c为扰动的临界尺寸

网络中边的三个类别：

C型 (conjunct)：这类边的目标顶点依赖于源点（化合物），一旦源点失效，那么这个边就失效了。例如：要发生某个反应，所有需要的化合物都必须满足。

D型 (disjunct)：这一类边被认为是多余的，即如果把所有和d点连接的点都失效，他才会变的失效。

R型 (regulation)：这一类的边覆盖了14种不同类型的反应事件，我们把这些边和C型边一样对待。

关于渗流模型：

1.在描述网络中初始扰动传播的动力学规则中，我们根据LCE来区分给定的边在更新目标顶点中的不同角色。在我们的渗流模型中，每个顶点都被分配一个布尔状态变量 $\sigma \in \{0,1\}$ 。

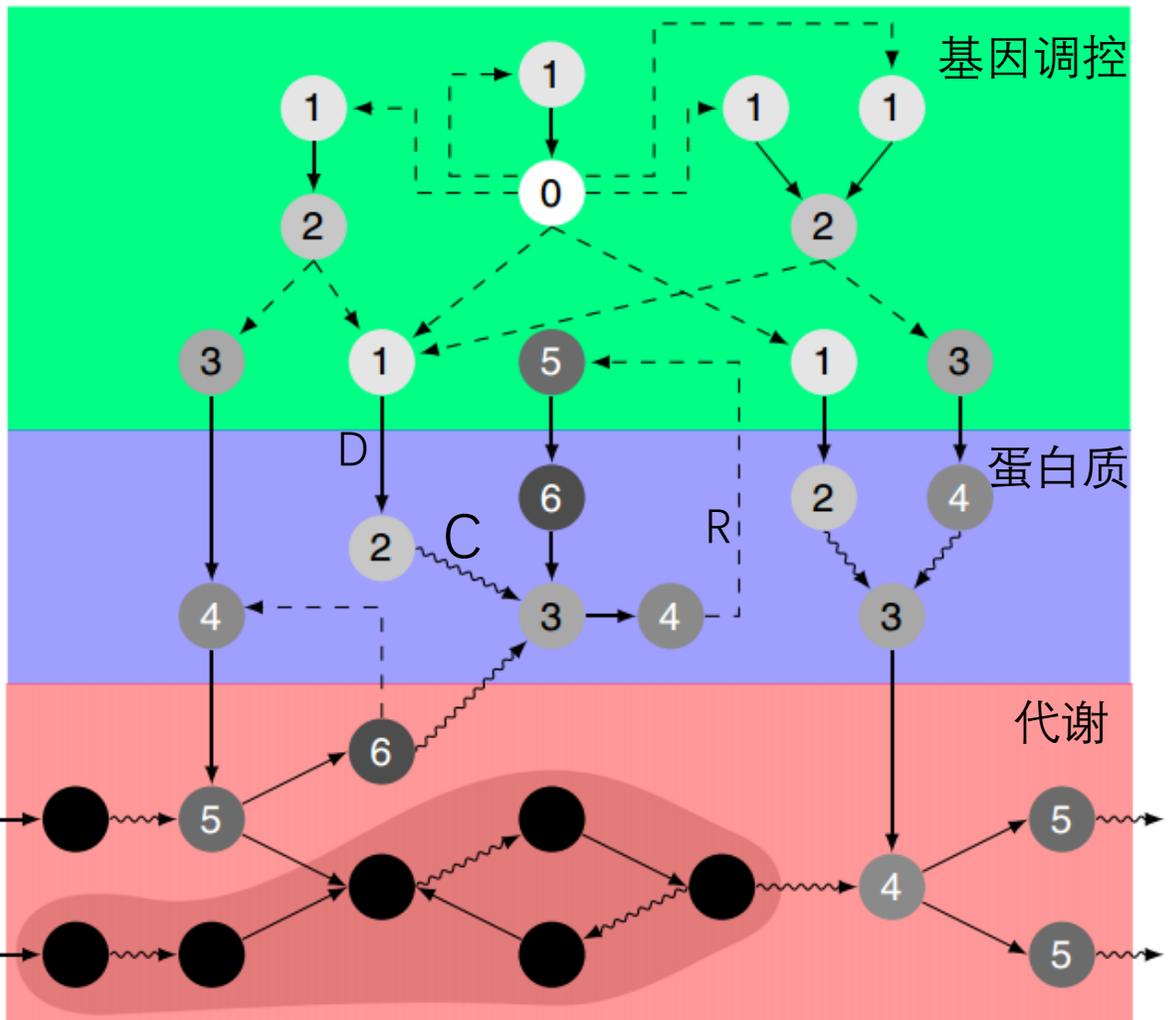
2.因为我们打算模仿扰动的传播，我们将状态1标识为“尚未受扰动影响”，而状态0对应于“受扰动影响”。

3.如果一个顶点没有通过C或R的边连接，那么一个顶点就会被认为不会受扰动的影响。

(之前被用于疾病蔓延的研究)

扰动扩散图解：

单点扰动传播动力学的简图 (C:卷曲, D:固体, R:虚线)



通过大肠杆菌基因调控和代谢整合系统的图表表示三个领域来研究扰动的扩散。我们将产生的**连锁故障**量化作为初始扰动的大小和定位。

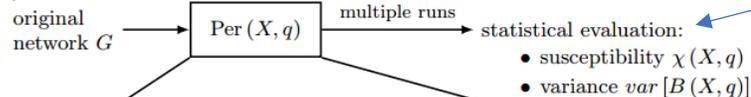
(frozen network state 整个阶段都不受影响的区域)

对局部扰动分析的示意图：

输入扰动域 X 和扰动 q 的大小的摄动算法 $\text{Per}(X; q)$

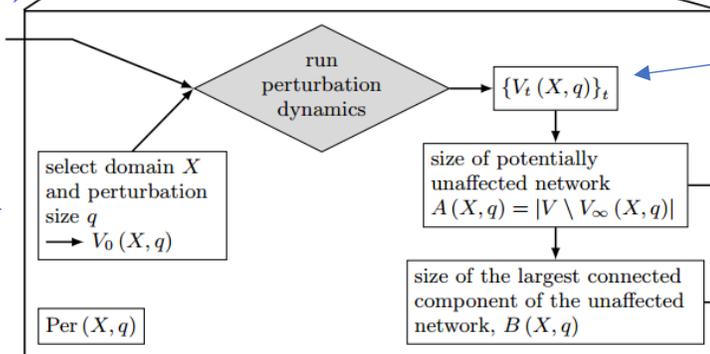
然后根据敏感度和方差评估多次运行统计数据;

(a) 数据处理



(b) 扰动细节运算

随机选择初始扰动点的集合 V_0

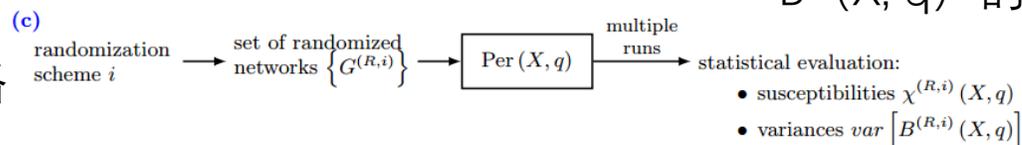


排列扰动导致受影响的点的时间序列 V_t

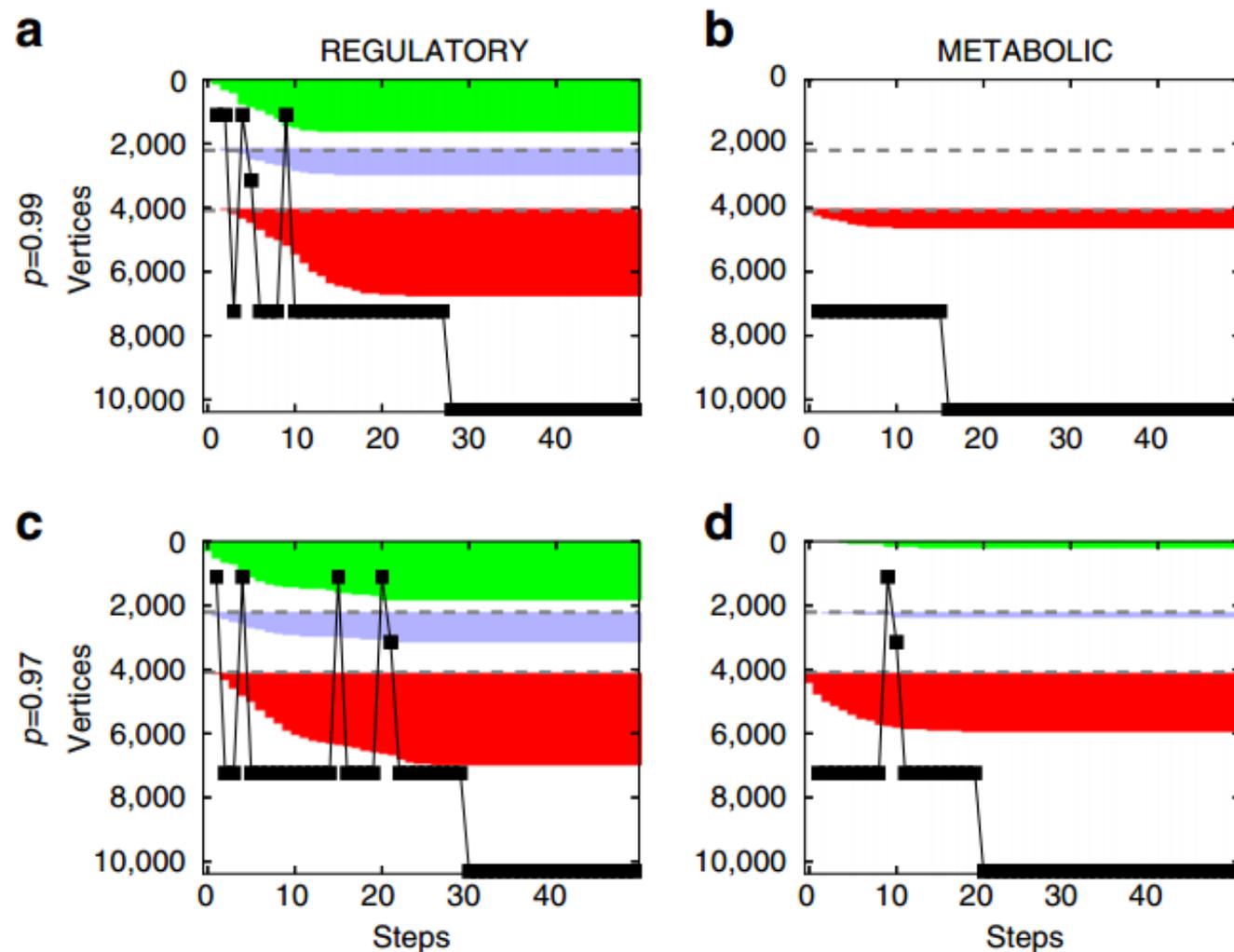
评估为产生未受影响的网络的大小

未受影响网络的最大（弱）连通量 $B(X; q)$ 的相对大小

(c) 与 (a) 相同, 但是针对于随机网络



大肠杆菌集成网络的样本轨迹图：

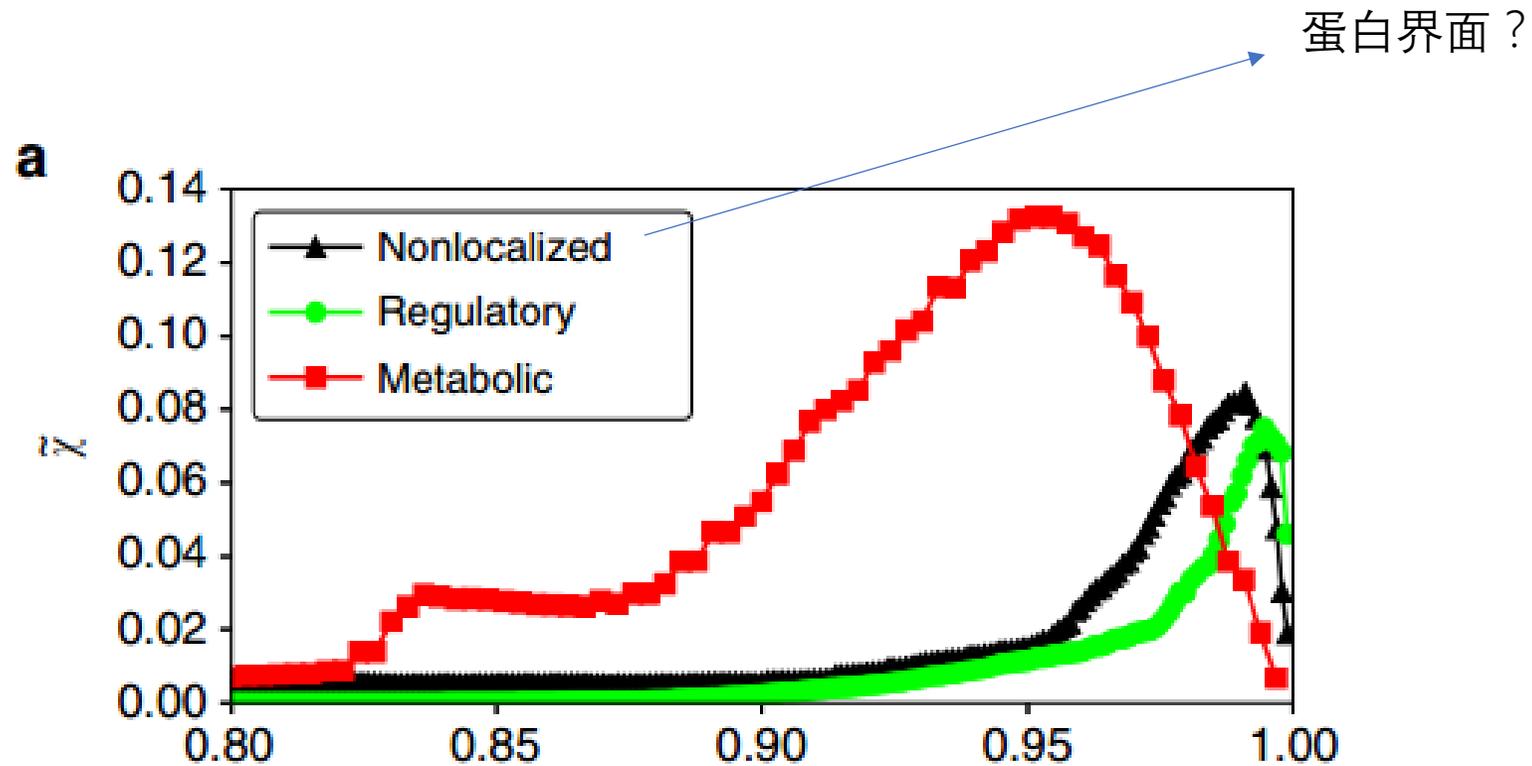


点是按照更新的步骤排列在每个域中，因此不同的点在不同的生物学运动中有不同的排序，点的属性有变化也是同理。

关于动态网络分析：

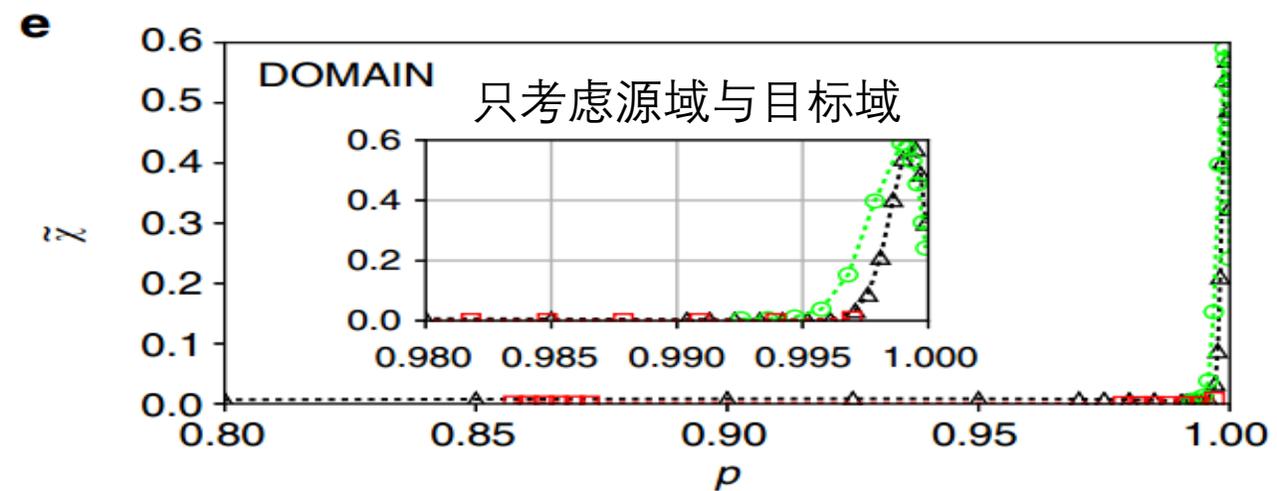
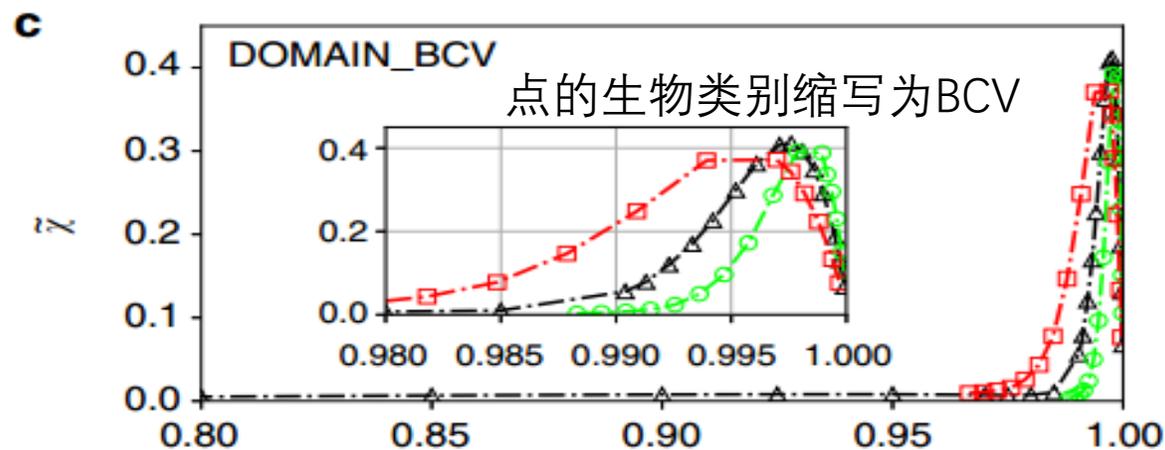
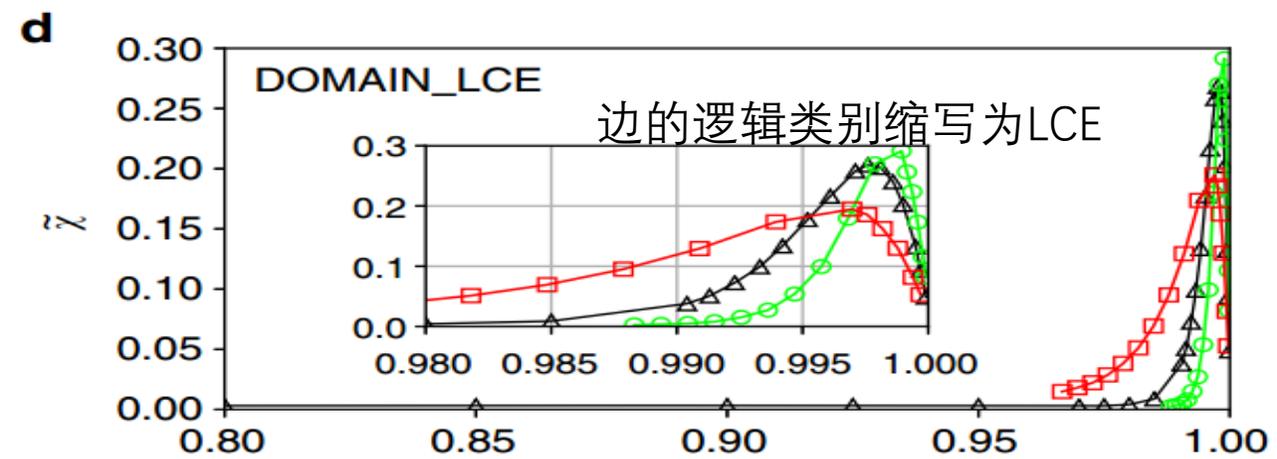
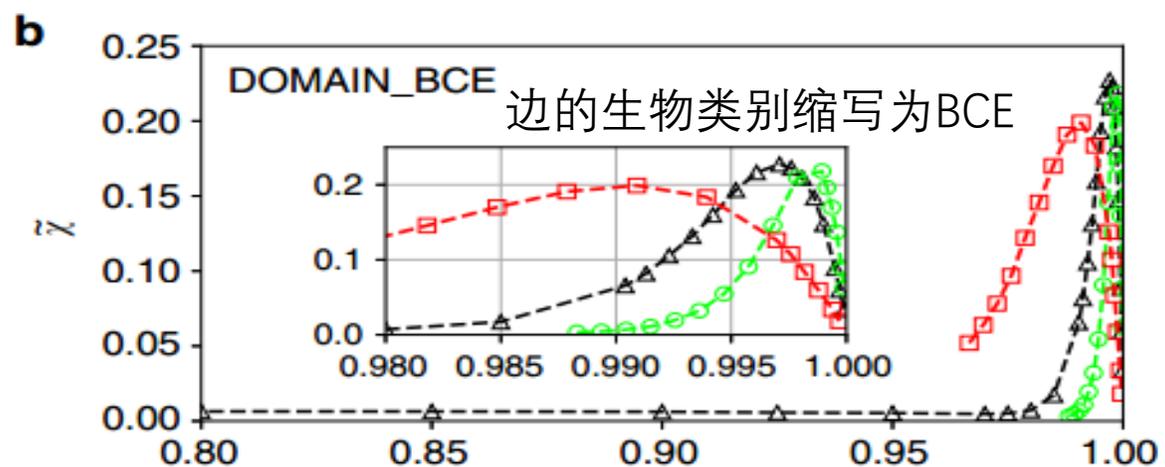
- 把边的源点和目标点的域称为源域SD,和目标域TD。EcoCyc数据库中的元数据可以用来给图的点和边分配属性，用于渗流模型和随机化方案的编译。
- 区分边的**生物类别**（边的不同生物功能）和**逻辑类别**（确定渗流过程的方向）。在网络的动态变化中每个点时刻在更新。
- 将点的生物类别缩写为**BCV**，将边的生物类别缩写为**BCE**。八个BCE中的每一个都可以被唯一映射到三个**LCE**（逻辑生物类别）中的一个。

大肠杆菌三域网络对不同局部扰动的敏感性：



每一个随机图都是通过五百次扰动产生的重复图的权重集合。非局部干扰呈黑色，在基因调控域和代谢域中引起的扰动分别以绿色和红色给出。

大肠杆菌三域网络随机版本对不同局部扰动的敏感性：



为了评估上述行为是否是由于网络的特定属性造成的，我们使用一组随机图。

分析：

- 我们看到整个原始网络似乎比随机化网络强大得多，一个非常小的扰动就足以形成鲜明的对比。因此，原始图的稳健性不能仅仅归因于随机化方案中保留的边和顶点属性。

结论：

- 1.相互依赖的系统对基因调控和蛋白质水平的扰动敏感，但对代谢变化的稳健性却很强。
- 2.真实网络的未受影响顶点的数量远大于所有四个随机化方案的数量，这表明生物系统具有较强的整体稳健性。（对于不太大的初始扰动， $p > 0.94$ ）。

猜想：

代谢网络具有稳健性是为了抵御时刻变化的环境因素，而基因调控区的敏感性使为了快速对其他部分的干扰作出反应

依据：相对于基因调控区域，代谢区的点更多而且也有更多的D型边。

不足之处：

在具体的应用中C型边与R型边的区别并不是很清楚。如果系统被限制在C、R和D三种边来连接点，相当多的信息将会丢失。

在无穷系统中，可以直接研究 $B(X, q)$ 作为 q 的函数，以产生 q_c 。研究的有限系统中，不得不估计 q_c ，寻找可能性的峰值位置。

亮点和启发：

- 处理不完整的生物数据时，在简化的系统和有适当细节的水平之间找到平衡，并利用渗流理论来评估。动力学、功能水平。
- 在这样大型的网络分析中，从复杂网络的科学里汲取了大量的工具。将图论的数学学科与统计物理学的方法相结合，大大有助于理解

THANKS!