

# 信息路径的多样性与现实世界网络的稀疏性的关联性研究

## 摘要

本文研究了现实世界网络中信息路径的多样性如何驱动稀疏性。现实网络既需要响应外部扰动，又需在内部协调各个组成部分的信息传递。本文表明，网络在保持信息流通的同时，会牺牲一定的响应多样性，这种权衡类似于热力学中热与功的效率关系。通过将网络状态编码为密度矩阵，本文提出了一个变分原理，能够从宏观上解释现实世界网络的稀疏性和规模规律，并通过数学分析和数值模拟验证了该原理在多个领域的普适性。研究结果表明，网络中的模块化、小世界特性和异质性等拓扑特征可以通过最大化信息流与响应多样性之间的平衡来解释，尤其在中到大时间尺度上，体现了这些特征的热力学来源。这些发现为理解复杂系统中的网络结构和功能提供了新的理论框架。

**关键词：**复杂网络；信息流；响应多样性；热力学效率

## 1 引言

现实世界中的网络通常具有稀疏连接的特性 [5]——即实际存在的链接数量远小于潜在可用的链接数量——并展示出一些独特的拓扑性质，如异质性连接性 [25]、小世界性 [24]、模块化 [17] 和层次结构 [21]，以及在分离与整合之间的平衡 [22]。为了复制这些特征，许多研究者提出了多种网络增长机制 [2, 7, 10, 13, 16, 19, 20] 和不直接基于网络生长的方法 [6, 9]。然而，目前仍缺乏一个理论框架，能够解释为何某种从断开状态到相对稳定的连接配置的过渡会自然发生。

这种过渡在许多复杂生物系统中得到了观察：系统内各单元之间的通信，通常可以通过化学、电或电化学信号的交换来理解，亦或是通过二进制数据包或语言的传递，促使系统开始正常运作。例如，人类口腔中的细菌通过信号分子（如粘附素和受体）在多物种社区内进行信息传递，从而促进了附着与社区的形成 [12]。当真菌菌落建立后，细胞网络通过信号调控菌落的生长与发展 [8, 14]。广泛来说，信息的搜索与交换在生命体、合成体或工程复杂系统的形成、适应与进化中起着重要作用，尽管这一过程仍未完全揭示 [23]。为了应对内外部扰动，复杂系统需要平衡动态功能的快速响应与所需能量成本 [18]。例如，生命体中从单细胞到生态系统的规律无处不在，这些规律可能与各个组织层次上最优网络的出现密切相关 [15]。一般来说，涉及资源（如能量和信息）获取、存储、分配或转化的结构与动态过程，被认为是塑造复杂网络功能属性的关键因素，这些网络广泛存在于生物和非生物运输及供需系统中。

在复杂网络领域 [1]，识别这些组织原理仍是一个未解的挑战，尽管显然在多种系统中存在某种成本效益机制 [24]。一个典型的例子是人脑，它被认为通过最小化连接成本及枢纽介

导的模块间信息交换来优化网络结构 [4]。因此，为了优化网络性能，研究者提出了多种复杂网络的进化原理，从抗扰动能力到可导航性与运输效率等。某些方法已成功恢复了特定的拓扑特征，如枢纽结构、无标度性和层次化组织结构，无论是否有空间约束 [3]。然而，准确通过从第一原理出发的模拟，揭示不同规模和领域的复杂网络行为高效性背后的普遍过程，仍然是一个难题。

在平衡态和接近平衡态的统计物理学中，以上问题通常可以通过基本原理得到自然解答，如吉布斯熵最大化或自由能最小化。然而，对于那些开放且远离平衡的复杂系统，我们缺乏足够的理论框架来描述、解释和预测其状态转变。因此，如何在不同的现实网络中出现这些拓扑特征，以及它们在多大程度上是由信息交流过程中的基本机制和相关成本引起的因果副产品，仍然是一个未解之谜。

本文将首先引入一个理论框架，来分析网络形成过程。随后，我们将展示如何利用这一框架，预测复杂网络的特定拓扑特征，如连接性、密度、分离与整合的共存以及小世界性中有序与无序的共存。最后，我们将对数百个经验网络的拓扑稀疏性进行比较，发现理论预测与实际数据之间有着极好的符合。

## 2 本文方法

### 2.1 本文方法概述

本工作采用物理中的热力学的概念（扩散方程、冯诺依曼熵和自由能）和方法（密度矩阵和变分法）将网络的形成描述为一个物理过程，采用热力学中的熵最大化和自由能最小化的平衡，量化了响应多样性（response diversity）的损失、信息流（information flow）的增益及两者之间的权衡（定义为  $\eta$ ），解释了网络稀疏拓扑结构的出现 [11]。图 1 展现了本工作的理论和实验完整思路。

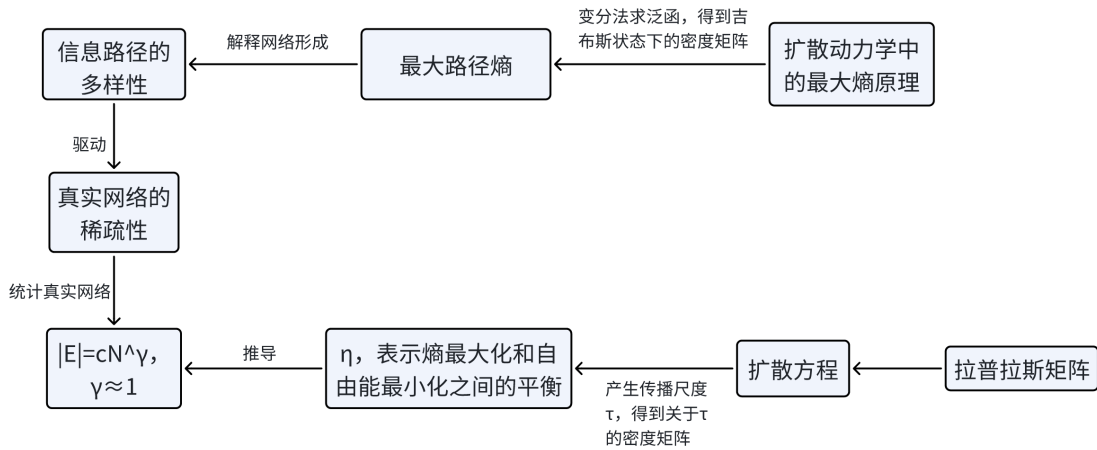


图 1. 理论框架图

### 2.2 网络密度矩阵

设定一个复杂网络  $G$ ，该网络有  $N$  个节点， $E$  条边。通过邻接矩阵  $A_{ij}$  表示网络单元之间的链路，1 为通路，0 为断开。 $D$  是对角矩阵， $D_{ii} = k_i$ ， $k_i$  是节点  $i$  的度。结合拉普拉斯

矩阵  $L = D - A$ ，扩散方程及其解可以写成态演化的具体形式（含时演化）：

$$\partial_\tau \psi_\tau^T = -L \psi_\tau^T \psi_\tau^T = e^{-\tau L} \psi_0^T \quad (1)$$

其中， $\psi_0$  是初态，演化  $\tau$  时间后，变成  $\psi_\tau$ ， $e^{-\tau L}$  是时间演化算子（the time-evolution operator），例如， $v_i(e^{-\tau L})v_i^T$  表示从节点  $i$  到  $j$  的流， $v_i, v_j$  是单位向量。特别地，有  $\tau$  表示传播尺度，决定信号在网络中的传播距离。如果  $\tau$  很小，信号主要在相邻节点之间传播，而  $\tau$  很大时，可以实现长距离通信。为了更好地了解  $\tau$  在确定网络中不同长度路径的贡献的方式，时间演化算子可拓展为

$$e^{-\tau L} = \sum_{n=0}^{\infty} \frac{(-\tau)^n}{n!} L^n \quad (2)$$

当  $\tau$  增大，第二邻居（距离为 2 的节点）开始通过  $\frac{\tau^2}{2} L^2$  项进行通信，当  $\tau$  取更大的值时，更长路径的激活可以通过  $L$  的更高次幂实现。接着，讨论另一个重要数学对象是统计传播算子（the statistical propagator）。假设每个结点  $N$  可以被扰动的概率是  $\frac{1}{N}$ ，如果结点  $i$  被扰动，信号流由  $e^{-\tau L} v_i^T$  给出，他与自身的外积  $e^{-\tau L} v_i v_i^T e^{-\tau L^T}$  就是他的局部传播算子，编码了从结点  $i$  接收信号的任意两个节点之间的关联，作为非对角元素，以及信号能量在节点之间的分布，作为对角元素。统计传播算子是局部传播算子（the local propagators）的叠加，这反映了整个系统对扰动的整体响应：

$$U_{2\tau} = \frac{1}{N} \sum_{i=1}^N e^{-\tau L} v_i^T v_i e^{-\tau L^T} = \frac{e^{-2\tau L}}{N} \quad (3)$$

尽管统计传播算子的迹是系统中的总信号能量，它也提供了一个合适的配分函数  $Z = \text{Tr}(U_\tau)$ ，对信号能量单元进行计数，并能够定义网络密度矩阵：

$$\rho_\tau = \frac{U_\tau}{Z} \quad (4)$$

接着，我们需要明确网络冯诺依曼熵和网络自由能的定义。网络冯诺依曼熵  $S$  是用来衡量系统对扰动响应的多样性。它定义为  $S = -\text{Tr}(\rho_\tau \log \rho_\tau)$ ，这里的  $\rho$  是网络的密度矩阵。冯诺依曼熵在量子力学中用于衡量量子系统的混合度，即系统状态的不确定性或纯度。在网络理论中，较高的熵表明网络对扰动有更广泛的响应，而不是集中在某几个特定的状态。网络自由能  $F$  是用来衡量信号在节点之间传输速度的指标。它定义为  $F = -\log Z/\tau$ ，这可能与系统的稳定性和效率有关。在物理中，自由能通常与系统在某状态下做工的能力相关联。这两个概念都表明，使用密度矩阵可以对网络的动态行为进行更深入的分析，包括系统如何对随机扰动做出反应以及信号如何在网络中传播。

### 2.3 网络形成过程

将网络形成过程描述为初始状态  $G_0$  转变为  $G_1$ ，从只有点没有边，转变为具有拓扑特征的网络。该过程只能降低响应多样性（response diversity），增强传输特性（transport properties）。为了比较这两个因素，我们定义了一个增益函数  $W = \delta F$  和一个损失函数  $Q = \delta S/\tau$ ，有  $W \geq |Q|$ ，以及它们之间的相对权衡可以表示为

$$\eta = 1 - \frac{|Q|}{W} \quad (5)$$

从数学角度来看，它们类似于热量  $Q$  和功  $W$  的热力学函数，尽管这些概念不能直接推广到网络中。在图 2 中，显示了一个象征性的例子，其中在具有不同连通性的四个图之间比较了响应多样性和信号流。 $\eta$  越高的网络更好地平衡了响应分集的损失和信息流的增益。所有由微观机制涵盖的拓扑特征都可以自然地熵和自由能 ( $\eta$ ) 的平衡中出现，而不需要显式地强加这些特征。也就是网络的宏观结构和行为可以通过这些原理来理解，而不一定需要从微观层面的详细规则出发。

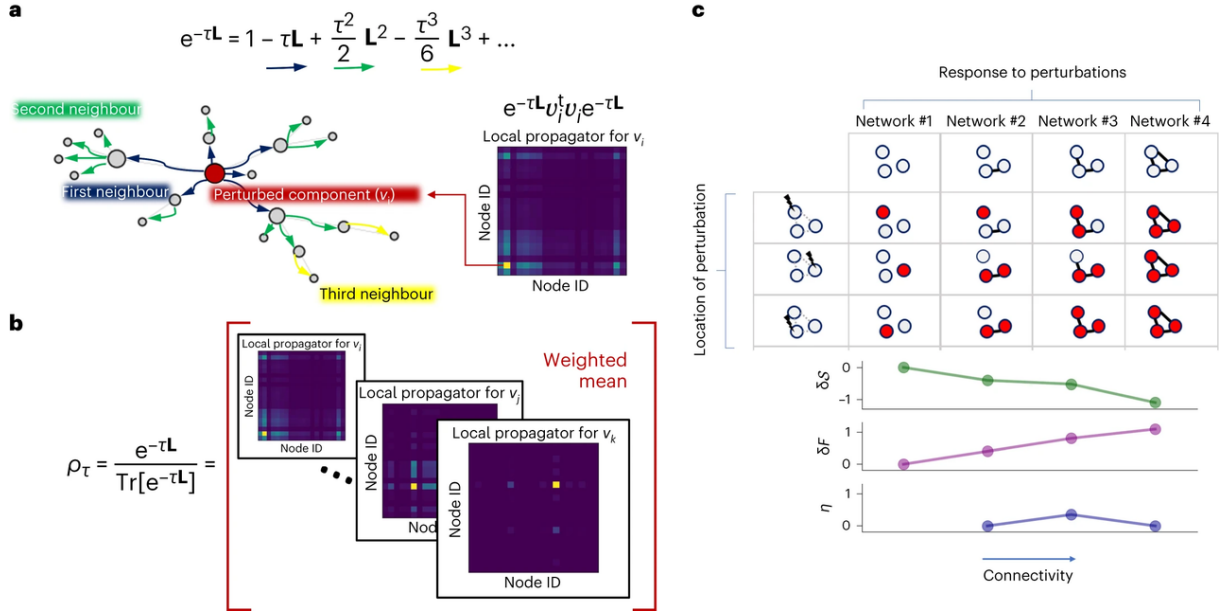


图 2. 响应多样性与信息传播

## 2.4 $\eta$ 和信号传播尺度

首先，我们对广义热功函数进行泰勒展开，只保留前导项。我们证明了在非常小的尺度上，其中信号是局部包含的 ( $\tau \approx 0$ )，所有的网络形成过程都是不可区分的，其特征为  $\eta = 1$ 。值得注意的是，在这种情况下，网络拓扑结构与系统的功能无关。其次，我们发现具有较少链路的网络中以公式 2 中的第零项 ( $n = 0$ ) 和第一项 ( $n = 1$ ) 展开。最后，二阶校正揭示了连通性的重要性：假设网络中的链路数量与节点数量成正比，即

$$|E| = cN^\gamma \quad (6)$$

其中  $c$  是一个常数，我们发现使  $\eta$  最大化的最优指数为  $\gamma = 1$ 。因此，通过要求  $\eta$  在网络形成过程中实现最大化，我们自然地得出相应的连通性分布必须是稀疏的，并且具有特定的指数。

## 2.5 网络中的自由能

在网络中，自由能的概念对应于扩散动力学中的网络分割函数  $Z$ ，这个函数衡量了网络动力学陷阱的程度，也就是网络在局部阻碍信息流动的倾向。拉普拉斯算子的特征值被记为

$\lambda_\ell$  , 其中  $\ell = 1, 2, \dots, N$  , 对于有  $N$  个节点的网络: 网络分割函数遵循  $z = \sum_{\ell=1}^N e^{-\tau \lambda_\ell}$ 。因为所有的特征值都是非负的,  $\lambda_\ell \geq 0$ , 并且至少有一个特征值  $\lambda_1 = 0$ , 所以分割函数被限定在  $1 \leq Z \leq N$  之间。下限  $Z = 1$  对应于在非常大的传播尺度 ( $\tau \gg 1$ ) 的连通网络中, 而上限  $Z = N$  出现在最小的尺度上, 即  $\tau = 0$ , 或者当网络由孤立节点组成时, 因为在这种情况下拉普拉斯矩阵的所有特征值都是零。网络的自由能与分割函数成反比, 定义为  $F = -\log Z_\tau / \tau$ 。因此, 孤立节点网络有最低的自由能  $F_{min} = F_{iso} = -\log N / \tau$ 。同时, 最小的动力学陷阱 ( $Z_\tau = 1$ ) 给出最大的自由能  $F_{max} = 0$ 。因此, 网络自由能的界限被定义在  $0 \geq F \geq -\log N / \tau$ 。

## 2.6 网络中的冯诺依曼熵

冯·诺依曼熵是量子力学中用于描述量子系统混乱程度或信息含量的一个量度。在网络科学中, 这个概念被借用来描述网络状态的不确定性或者复杂性。网络冯·诺依曼熵由式子  $S = -\text{Tr}(\rho_\tau \log \rho_\tau)$  给出。这里的密度矩阵的特征值对应着玻尔兹曼分布的概率 ( $e^{-\tau \lambda_\ell} / Z$ ), 而微正则系综在所有特征值相等时给出最大熵, 即  $\lambda_\ell = \lambda_\ell$ 。在  $\tau > 0$  的条件下唯一满足这个条件的网络是孤立节点的网络, 它的特征值全为 0, 并且它的冯·诺依曼熵为  $S_{max} = S_{iso} = \log N$ 。相比之下, 在连通网络中有一个特征值为 0, 其他的都为正。因此, 在大尺度上, 当我们观察大尺度上的网络扩散行为时, 几乎所有状态的概率都会随时间指数级衰减 ( $e^{-\tau \lambda_\ell}$ ), 这意味着系统趋向于达到一种稳定状态。因此, 除了因为  $\lambda_1 = 0$  而不衰减的基态外, 系统的熵会趋向于最小值  $S_{min} = 0$ 。因此, 网络冯·诺依曼熵的界限可以表示为  $0 < S < \log N$ 。这个范围给出了网络在最不确定 (孤立节点) 和最确定 (完全稳定状态) 之间的熵变化范围。

## 2.7 网络在可扩散动力学中的内能

内能  $U$  通过与统计物理中的吉布斯态相似的网络密度矩阵来表征。内能可以通过下面的配分函数  $Z$  表达式计算得到:

$$U = -\partial_\tau \log Z = \frac{1}{Z} \sum_{\ell=1}^N \lambda_\ell e^{-\tau \lambda_\ell} \quad (7)$$

文中指出, 拉普拉斯矩阵特征值的总和是网络中链接数目的两倍, 也就是

$$\text{Tr}(L) = \sum_{\ell=1}^N \lambda_\ell = 2|E| \quad (8)$$

其中,  $|E|$  是网络中边的数量。所以, 当  $\tau = 0$  时, 网络的内能为

$$U = \frac{1}{Z_\tau} \sum_{\ell=1}^N \lambda_\ell = \frac{2|E|}{N} \quad (9)$$

最后, 网络的内能还可以通过自由能  $F$  和熵  $S$  来表示, 公式为

$$U = F + \frac{S}{\tau} \quad (10)$$

式子10对应的是热力学的一个标准问题: 自由能  $F$  和内能  $U$ , 熵  $S$  之间的关系是

$$F = U - k_B T S = U - \frac{1}{\beta} S \quad (11)$$

按照前面的类比有

$$F = U - \frac{1}{\tau} S \implies U = F + \frac{1}{\tau} S \quad (12)$$

## 2.8 网络形成作为物理过程在热力学中的类比

首先定义了一个量  $Q$ ，类似于热力学中的热量，表示网络形成的熵代价：

$$Q = \frac{S - \log N}{\tau} \leq 0 \quad (13)$$

这里  $S$  是网络的熵， $S_{iso} = \log N$  是孤立节点网络的熵， $\tau$  是传播尺度。 $Q$  被解释为类似于热量的东西，在热力学过程中散失掉，因此它是负的或为零。接着，信息流的增益  $W$  是通过网络自由能的变化得到的：

$$W = F + \frac{\log N}{\tau} \geq 0 \quad (14)$$

其中  $F$  是网络自由能， $F_{iso} = -\frac{\log N}{\tau}$  是孤立节点网络的最小可能自由能。因为  $W + Q = F + \frac{S}{\tau} = U$  且  $U \geq 0$ ，所以在网络形成的过程中，增益与损失的总和总是非负的。而且  $Q \leq 0$ ，所以  $W \geq |Q|$ 。再有  $0 < W + Q < W$ ，则  $0 < \frac{W+Q}{W} < 1$ 。最后，定义了一个效率参数  $\eta$ ，用来衡量网络形成中增益与损失之间的相对权衡

$$\eta = \frac{W + Q}{W} = 1 - \frac{|Q|}{W} = \frac{U}{W} \quad (15)$$

这个参数  $\eta$  的值在 0 和 1 之间，它量化了网络形成过程中的效率。

## 2.9 求 $\eta$ 的近似值

首先，定义拉普拉斯矩阵的特征值为  $\lambda_\ell$ ， $\bar{\lambda}$  作为平均特征值的近似，用于简化计算。配分函数  $Z$  的泰勒展开到第三项

$$Z \approx N(1 - \frac{\tau \bar{\lambda}}{N} + \frac{\tau^2 \bar{\lambda}^2}{2N} - \frac{\tau^3 \bar{\lambda}^3}{6N}) \quad (16)$$

接着，近似计算配分函数的对数

$$-\log \frac{Z}{N} \approx \frac{\tau \bar{\lambda}}{N} - \frac{\tau^2 \bar{\lambda}^2}{2N} + \frac{\tau^3 \bar{\lambda}^3}{6N} \quad (17)$$

这个对数是内能  $U$  和信息流增益  $W$  计算的基础。对应的内能  $U$  和信息流增益  $W$  的近似表达式为

$$U = -\partial_\tau \log Z = -\partial_\tau \log \frac{Z}{N} \approx \frac{\bar{\lambda}}{N} - \tau \frac{\bar{\lambda}^2}{N} + \frac{\tau^2 \bar{\lambda}^3}{2N} \quad (18)$$

$$W = -\frac{1}{\tau} \log \frac{Z}{N} \approx \frac{\bar{\lambda}}{N} - \frac{\tau \bar{\lambda}^2}{2N} + \frac{\tau^2 \bar{\lambda}^3}{6N} \quad (19)$$

为了方便计算，接着采用平均场近似，得到

$$\lambda_\ell^2 = (\frac{\bar{\lambda}}{N} - \delta \lambda_\ell)^2 \implies \bar{\lambda}^2 \approx \sum_{\ell=1}^N (\frac{\bar{\lambda}}{N})^2 + \delta \lambda_\ell^2 - 2\delta \lambda_\ell \frac{\bar{\lambda}}{N} \approx \frac{(\bar{\lambda})^2}{N} \quad (20)$$

$$\lambda_\ell^3 = (\frac{\bar{\lambda}}{N} - \delta \lambda_\ell)^3 \implies \bar{\lambda}^3 \approx \frac{(\bar{\lambda})^3}{N^2} \quad (21)$$

最后，我们假设连接数与节点数成正比，这是许多自组织网络（如无标度网络）的常见特性。这一假设带出的是  $\bar{\lambda}$  与节点数  $N$  的关系表达式

$$\bar{\lambda} = 2|E| = cN^{\delta+1} \quad (22)$$



其中,  $|E|$  是边数,  $c$  和  $\delta$  是与网络拓扑结构相关的常数。接着, 重写  $U$  和  $W$  的表达式

$$U \approx cN^\delta(1 - \tau cN^\delta + \frac{\tau^2}{2}c^2N^{2\delta}) \quad (23)$$

$$W \approx cN^\delta(1 - \frac{2}{\tau}cN^\delta + \frac{\tau^2}{6}c^2N^{2\delta}) \quad (24)$$

然后, 将  $\frac{1}{W}$  泰勒级数展开和一些近似计算

$$\begin{aligned} \frac{1}{W} &= \frac{1}{cN^\delta} \frac{1}{1 - \frac{\tau}{2}cN^\delta + \frac{\tau^2}{6}c^2N^{2\delta}} \\ &\approx \frac{1}{cN^\delta} (1 + \frac{\tau}{2}cN^\delta - \frac{\tau^2}{6}c^2N^{2\delta} + (\frac{\tau}{2}cN^\delta - \frac{\tau^2}{6}c^2N^{2\delta})^2) \\ &\approx \frac{1}{cN^\delta} (1 + \frac{\tau}{2}cN^\delta - \frac{\tau^2}{6}c^2N^{2\delta} + \frac{\tau^2}{4}c^2N^{2\delta}) \\ &= \frac{1}{cN^\delta} (1 + \frac{\tau}{2}cN^\delta + \frac{\tau^2}{12}c^2N^{2\delta}) \end{aligned} \quad (25)$$

最后, 将  $U$  和  $W$  的近似表达式23和25代入到效率参数的  $\eta$  式子中, 然后进行二阶近似并展开后, 只保留其指数为 0,1,2 的项:

$$\begin{aligned} \eta = \frac{U}{W} &\approx (1 - \tau cN^\delta + \frac{\tau^2}{2}c^2N^{2\delta})(1 + \frac{\tau}{2}cN^\delta + \frac{\tau^2}{12}c^2N^{2\delta}) \\ &\approx 1 - \frac{\tau}{2}cN^\delta + \frac{\tau^2}{12}c^2N^{2\delta} \end{aligned} \quad (26)$$

这个结果表示网络效率参数  $\eta$ , 它考虑了网络的大小 (通过  $N^\delta$  项) 和扩散尺度 (通过  $\tau$  项)。观察前两项 (一阶近似), 要使得  $\eta$  最大, 则第二项最小,  $\delta < 0$  或者  $N$  趋近于无穷时,  $N^\delta$  趋近于 0。即, 当网络越稀疏,  $\eta$  越大。

接着, 对表达式26的  $\delta$  求导, 得到27, 27的解即为两个最优点 ( $\eta$  的极值), 得到:  $N^\delta = 0$ ,  $\delta$  不随  $N$  的增长而缩放, 即  $N^\delta$  为常数, 则  $\delta = 0$ 。

$$\partial_\delta \eta = -\frac{\tau c}{2}N^\delta [1 - \frac{\tau c}{3}N^\delta] \log N = 0 \quad (27)$$

然后, 为了确定泰勒展开的二阶近似在小尺度、中尺度或大尺度传播条件下是否适用, 对于第  $i$  个特征值, 必须满足  $\frac{\tau^2}{2}\lambda_i^2 > \epsilon$ ,  $\epsilon$  是一个很小的正数, 用于确定何时可以忽略某个量。

基于最小的非零特征值  $\lambda$ , 提出了关于扩散时间  $\tau_{diff}$  的不等式:

1. 对于任意非零特征值  $\lambda_i$ , 确保  $\frac{\tau^2}{2}\lambda^2$  是小于  $\epsilon$  的, 从而可以确保泰勒展开到二阶是有效的。

$$\frac{\tau}{\tau_{diff}} > \sqrt{2\epsilon}$$

2. 同时, 要确保三阶项  $\frac{\tau^3}{6}\lambda^3$  也是可以忽略的。

$$\sqrt{2\epsilon} < \frac{\tau}{\tau_{diff}} < \sqrt{3\epsilon}$$

这表明, 传播尺度  $\tau$  必须选择在一个使得二阶和三阶近似都有效的范围内。另外, 选择  $\epsilon$  的大小会影响  $\tau$  的合理取值范围。

## 3 复现过程

### 3.1 实验环境搭建

操作系统：Microsoft Windows 10

编译环境：python 3.12

扩展包：networkx, matplotlib, numpy 等

数据库：ICON

### 3.2 与已有开源代码对比

本文复现过程的工作是基于开源的关于计算  $\eta$  的近似值的代码框架，模拟分析网络从不连通到连通、从分立到整合、从有序到无序以及从异构到同构等四个拓扑特征形成过程中的传播尺度和  $\eta$  的联系，另外，还利用了 ICON 上的 543 个现实世界的经验网络数据集验证理论结果。

### 3.3 复现细节

#### 3.3.1 关于 $\eta$ 的计算

代码的计算思路如下：

1. 输入拉普拉斯矩阵  $L_s$  和传播尺度  $\beta$ 。
2. 从拉普拉斯矩阵中删除接近于 0 的元素，以及相应的概率  $p$ ，防止数值计算中的下溢问题。
3. 计算配分函数  $Z$ ，这是所有  $e^{-\tau L}$  的和。
4. 计算概率分布  $p$ ，即每个状态的概率。
5. 计算自由能的变化  $dF$ ，使用的是网络可能状态总数  $N$  与配分函数  $Z$  的对数差。
6. 计算熵的变化  $dS$ ，及概率分布  $p$  的期望负对数减去  $\log N$ 。
7. 计算网络形成的效率参数  $\eta$ ，定义为  $\frac{dF+dS}{dF}$ 。

#### 3.3.2 改进 BA 模型

函数的输入参数包括：

1.  $n$ ：网络中的节点总数。
2.  $m$ ：每次添加新节点时，需要连接到现有节点的数量。这个参数也代表了模型中的优先连接（preferential attachment）特性，即新节点倾向于与已经拥有更多连接的节点建立连接。
3.  $\alpha$ ：逆转“富者更富”规律的概率，这是对标准 BA 模型的一个调整，以允许较不受欢迎的节点有时也能获得新的连接。



方法的原理如下：

1. 首先，创建一个包含  $m$  个节点的空图。
2. 定义初始的目标节点列表，这些是新节点将要连接的节点。
3. 通过优先连接机制添加剩余的  $n - m$  个节点。每次添加新节点时，都会选择  $m$  个目标节点进行连接。
4. 在选择目标节点时，函数会根据节点的连接度进行加权选择，但是使用了一个概率调整机制，其中概率由  $alpha$  参数影响。

## 4 实验结果分析

为了研究拓扑特征如何影响  $\eta$ ，对这些参数进行了调整，以观察在不同传播尺度下  $\eta$  的最大值。网络的大小： $N = 10^3$ ，通过平均 10 次实现，得出网络的平均扩散时间大约为  $\tau_{diff} \approx 10^2$ 。其中，大尺度传播有  $\tau_{large} \in (10^1, 10^2)$ ；中尺度传播有  $\tau_{mid} \in (1, 10^1)$ ；小尺度传播有  $\tau_{small} \in (10^{-1}, 1)$ 。图 3 展示了现实世界网络中一些最突出的拓扑特征如何影响  $\eta$ 。通过不断改变 Erdős-Rényi (ER)、Watts-Strogatz 的参数和随机块模型来找到每个传播尺度上  $\eta$  的最大值。同时，还引入了 Barabási-Albert (BA) 网络的修改，其中包含一个参数来调整异质性并使用它来评估高度连接节点对  $\eta$  的影响。

如图3(a)所示，网络的每一对节点都以相同的概率连接，这种模型通常生成随机网络。其中  $\eta$  的最大值出现在连接概率  $p$  接近  $N^{-1}$  和  $N^{-1} \log(N)$  之间。随着网络规模  $N$  趋近于无穷大（热力学极限），这两个概率之间的差异变得可以忽略不计，最优连接概率可被近似为  $p \approx N^{-1}$ 。

如图3(b)所示，在小尺度上，整合的社区显示出最高的  $\eta$ ，而在更大的尺度上，信息流与响应多样性之间的权衡更为优化，其中  $\log_{10}$  在 -2 和 -1 之间的结果是更理想的。

如图3(c)所示，当网络较为无序 ( $p_{rew} \approx 1$ ) 时， $\eta$  较高。而在大尺度上，有序结构 ( $p_{rew} \approx 0$ ) 的表现优于无序结构。最优的  $\eta$  出现在有序与无序之间的中间体制，这种体制表现出高度的小世界特性。

如图3(d)所示，引入一个可调节网络异质性的参数，修改后的 BA 模型，新加入的节点与现有节点连接的概率是原始的 BA 模型概率  $p_{BA}(i) \propto k(i)$  和一个逆概率  $q_{BA}(i) \propto \frac{1}{k_i^{10}}$  的混合。混合概率公式为  $(1 - p_{rew})p_{BA} + p_{rew}q_{BA}$ 。其中， $p_{rew}$  是一个控制参数，当  $p_{rew}$  较小或较大时，网络分别表现出异质性或均匀性。在小尺度上，网络节点均匀连接的  $\eta$  更高，而在大尺度上，异质性结果更好。

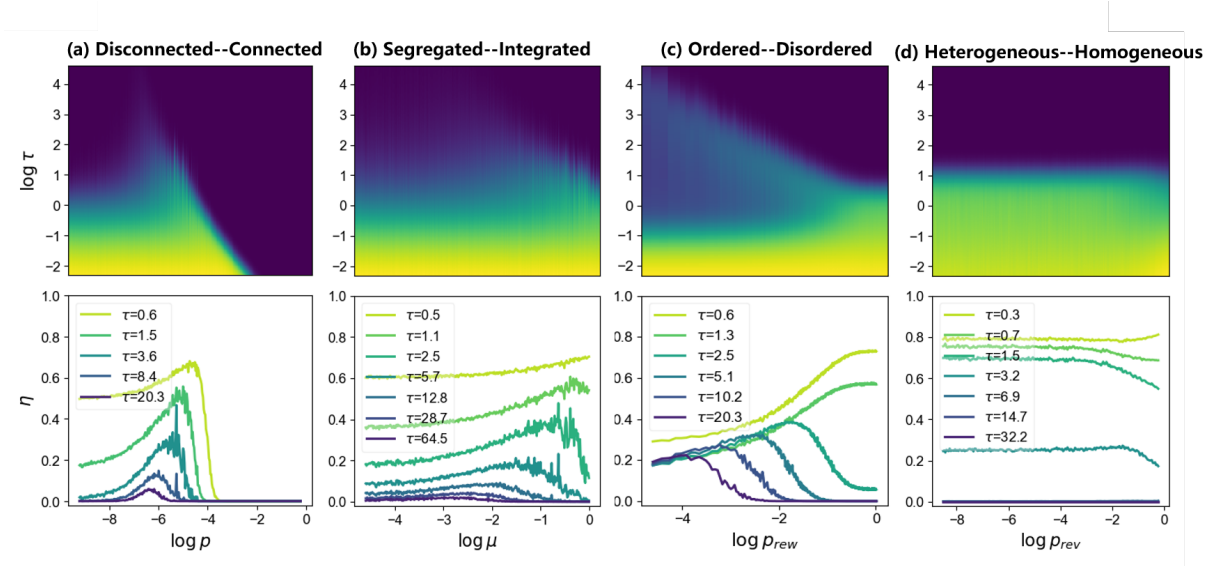


图 3. 拓扑特征对  $\eta$  的影响

通过分析来自 *ICON* 的 543 个社会、生物、信息、经济、交通和技术网络，大小范围为  $N \approx 10^2$  至  $N \approx 10^8$  节点。首先假设网络的连接数与网络大小之间存在幂律关系  $|E| \propto N^\gamma$ ，并在对数-对数空间进行线性回归以得出  $\gamma$  的值。实验发现，如图4所示，其总体最佳拟合的值为  $1.23 \pm 0.02$ ，虽然与作者的  $\gamma$  值为  $1.07 \pm 0.02$  略有出入，但还是与理论预测的  $\gamma = 1$  相当接近。

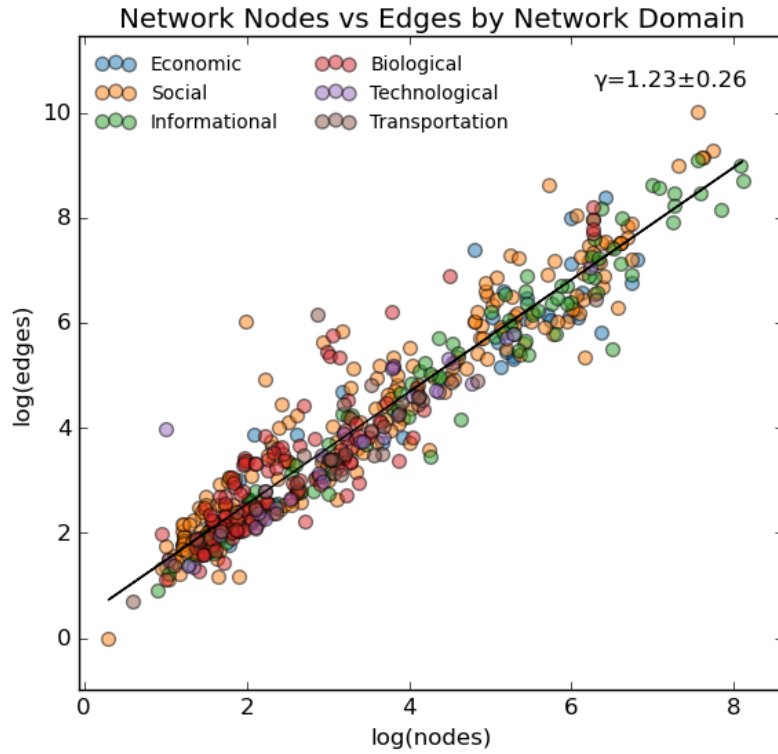


图 4. 经典网络的稀疏性

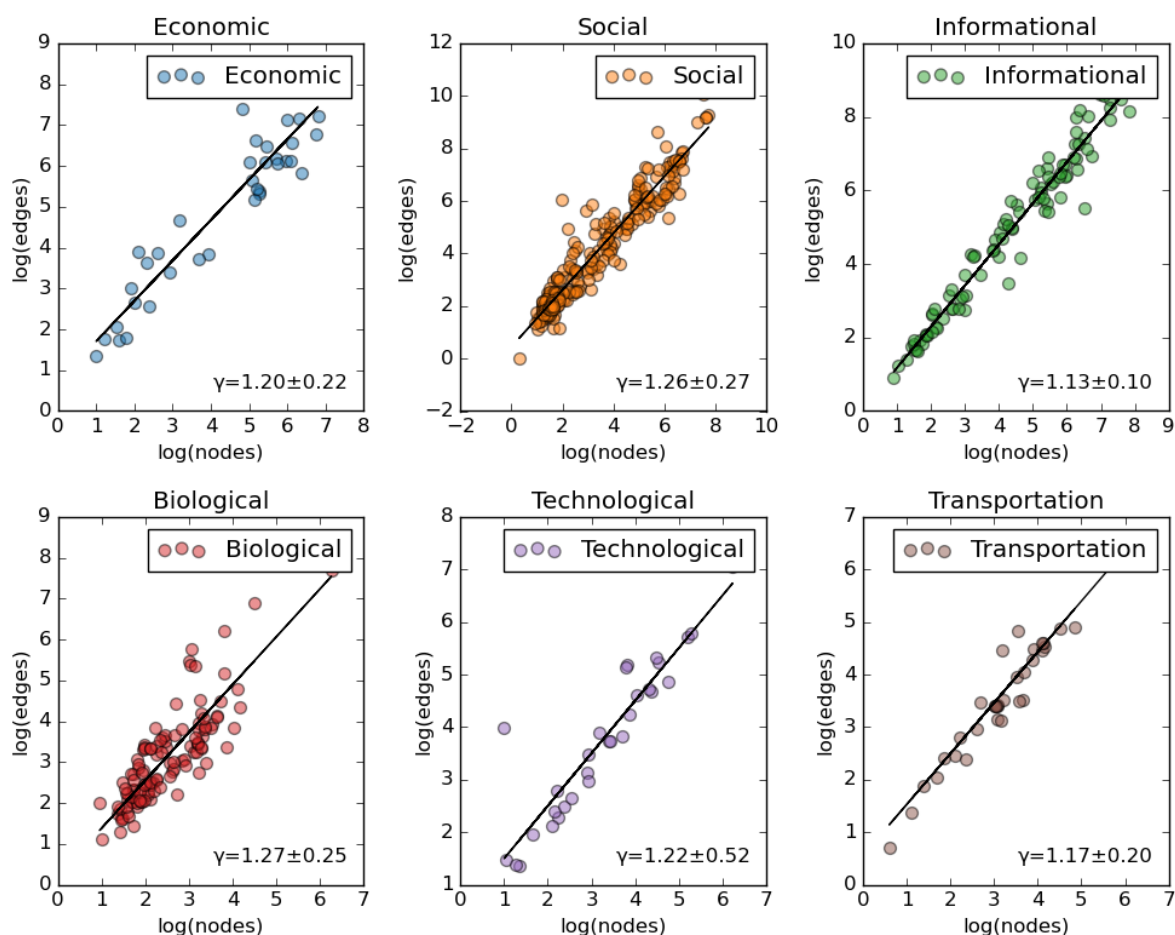


图 5. 不同分类下的经典网络

此外，如图6和图7所示，通过分析数百个经验网络，从真菌到人类、秀丽隐杆线虫，以及从城市和飞行网络到社交和食物网。这些经典网络与不同的零模型集合之间的直接比较表明，相对于其随机零模型，它们在中大型传播尺度上具有明显更高的  $\eta$ 。这一发现提供了进一步的证据，表明了复杂系统必须表现出拓扑相关性以维持其信息流，同时在宏观尺度上维持其路径多样性。

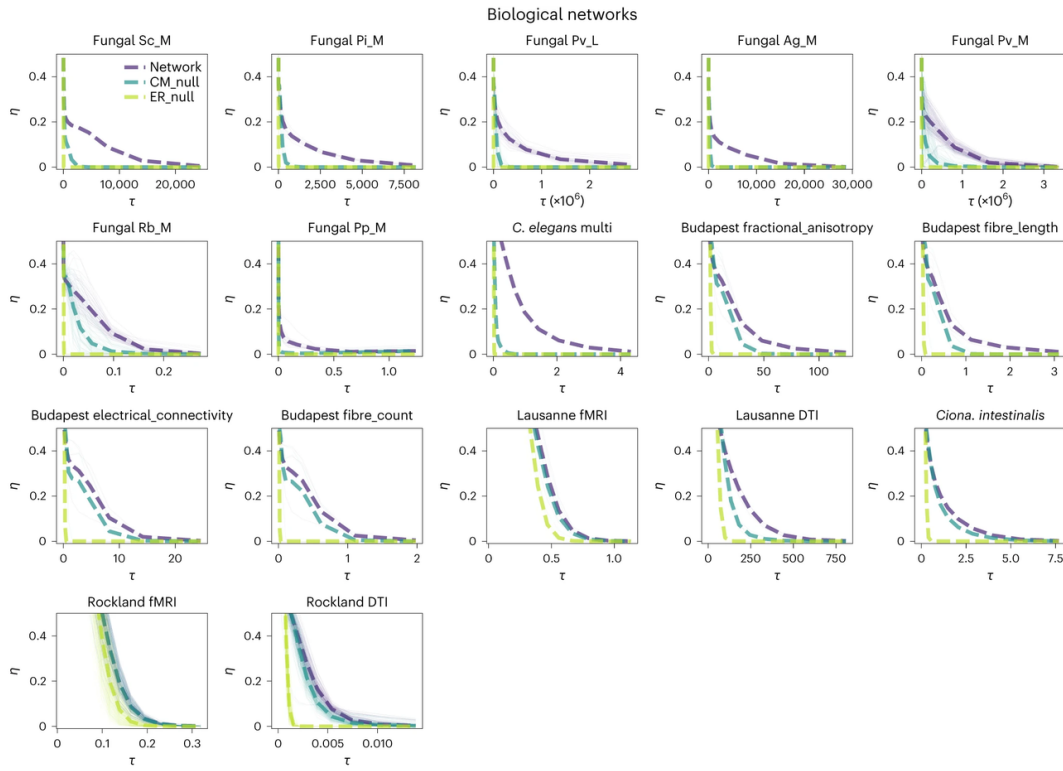


图 6. 零模型和生物模型

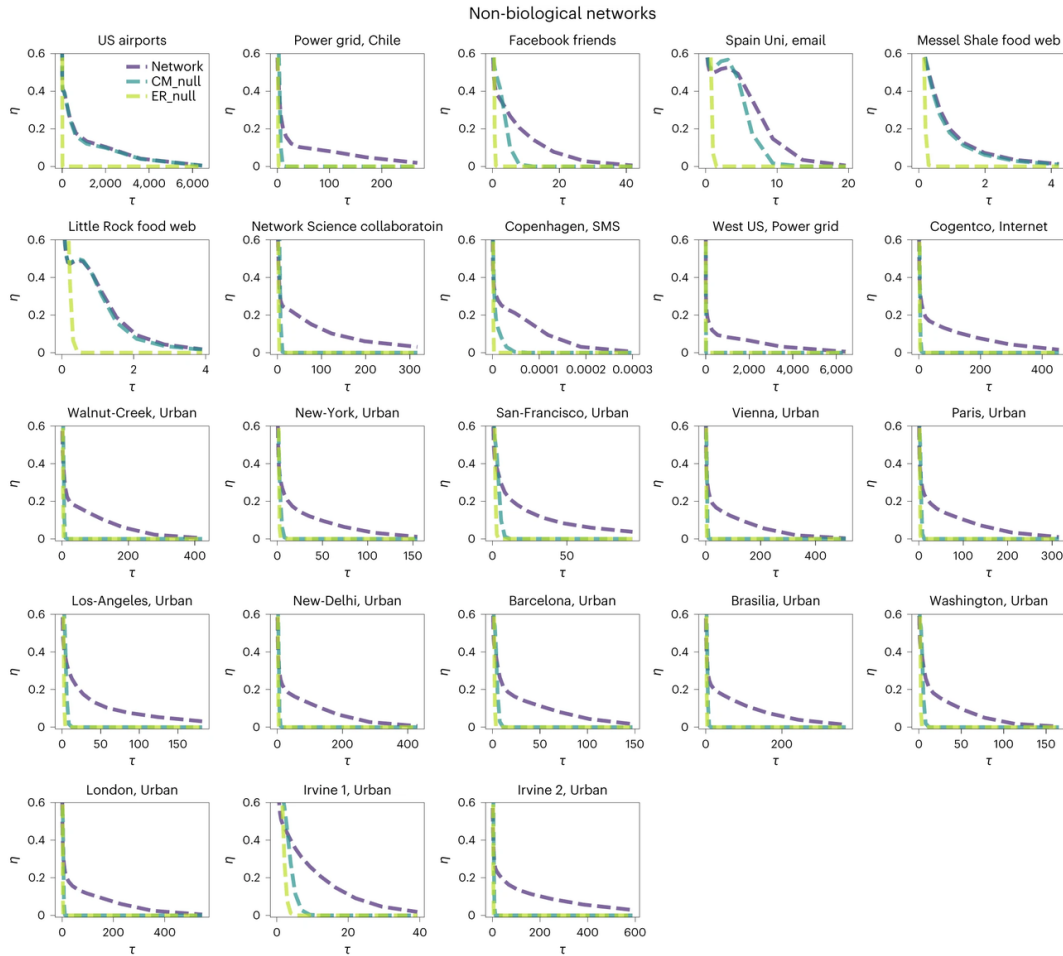


图 7. 零模型和非生物模型

## 5 总结与展望

本文研究了复杂网络中的拓扑特征对信息流和系统响应多样性之间的权衡所起的重要作用。通过引入网络密度矩阵并量化响应多样性损失、信息流增益及其权衡，我们提出了一种新的量化指标 ( $\eta$ )，类似于热力学中的效率概念。分析结果表明，优化信息流和响应多样性需要网络保持适当的稀疏性，并且其连通性遵循一种缩放定律 ( $|E| \propto N^\gamma$ )，其中  $\gamma \approx 1$ ，与理论预期高度吻合。通过对来自生物、社会、信息和运输等领域的 543 个经验网络进行分析，我们验证了这一规律，并且发现网络的模块化、小世界性等复杂拓扑特征能有效增强  $\eta$  值，促进系统单元之间的中远程信息通信。

进一步的数值实验表明，拓扑特征的引入，如模块化结构和小世界效应，有助于在多时间尺度下优化信息流和响应多样性之间的权衡。这表明，拓扑相关性在复杂网络中扮演着至关重要的角色，对于网络的功能和结构起到决定性作用。此外，通过与零模型的比较，我们提出了一种有效的方法来识别哪些拓扑特征对于信息交换和响应多样性之间的平衡至关重要。

尽管本研究基于扩散过程来近似信息流的动力学，但这一优化方法依然能够确保即便在复杂的非线性动力学下，系统也能实现高效的响应多样性和信息流。未来的研究可以进一步推广这一框架，探索更复杂的非线性过程，尤其是在反应-扩散过程中的应用。此外，随着多层网络和超图等高阶网络的兴起，这些新型网络结构的拓扑特征仍有待进一步研究。

在未来，可以进一步拓展确定合适的约束，当这些约束施加在变分原理上时，可以从分析上解释拓扑特征的出现，如模块化、小世界性和连通性异质性。这种观点相当令人兴奋，因为它可以根据信息如何流动，用新的动态视角来补充现有的结构方法。

## 参考文献

- [1] S. Boccaletti A, V. Latora B C, Y. Moreno D E, M. Chavez F, and D. U. Hwang A. Complex networks: Structure and dynamics. *Physics Reports*, 424(4–5):175–308, 2006.
- [2] Réka Albert and Albert  Laszl Barabsi. Topology of evolving networks: local events and universality. *Physical review letters*, 85 24:5234–7, 2000.
- [3] Marc Barthelemy. Spatial networks. In *Encyclopedia of Social Network Analysis and Mining*, 2010.
- [4] Edward T. Bullmore and Olaf Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13:336–349, 2012.
- [5] Daniel Maria Busiello, Samir Simon Suweis, Jorge Hidalgo, and Amos Maritan. Explorability and the origin of network sparsity in living systems. *Scientific Reports*, 7, 2016.
- [6] Guido Caldarelli, Andrea Capocci, Paolo De Los Rios, Paolo De Los Rios, and Miguel Angel Muoz. Scale-free networks from varying vertex intrinsic fitness. *Physical review letters*, 89 25:258702, 2002.
- [7] Sergey N. Dorogovtsev and Jos F. F. Mendes. Evolution of networks. *Advances in Physics*, 51:1079 – 1187, 2001.

- [8] Karoline Faust and Jeroen Raes. Microbial interactions: from networks to models. *Nature Reviews Microbiology*, 10:538–550, 2012.
- [9] Lazaros K. Gallos, Shlomo Havlin, H. Eugene Stanley, and Nina H. Fefferman. Propinquity drives the emergence of network structure and density. *Proceedings of the National Academy of Sciences*, 116:20360 – 20365, 2019.
- [10] Diego Garlaschelli, Andrea Capocci, and Guido Caldarelli. Self-organized network evolution coupled to extremal dynamics. *Topologica*, 2008.
- [11] Arsham Ghavasieh and Manlio De Domenico. Diversity of information pathways drives sparsity in real-world networks. *Nature Physics*, 20:512–519, 2024.
- [12] Paul E. Kolenbrander, Roxanna N. Andersen, David S. Blehert, Paul G. Eglund, Jamie S. Foster, and Robert J. Palmer. Communication among oral bacteria. *Microbiology and Molecular Biology Reviews*, 66:486 – 505, 2002.
- [13] P. L. Krapivsky, P. L. Krapivsky, Sidney Redner, and François Leyvraz. Connectivity of growing random networks. *Physical review letters*, 85 21:4629–32, 2000.
- [14] Abigail C. Leeder, Javier Palma-Guerrero, and N. Louise Glass. The social network: deciphering fungal language. *Nature Reviews Microbiology*, 9:440–451, 2011.
- [15] Pablo A. Marquet, Renato A. Quiñones, Sebastián R. Abades, Fabio A. Labra, Marcelo F. Tognelli, Matías Arim, and Marcelo M. Rivadeneira. Scaling and power-laws in ecological systems. *Journal of Experimental Biology*, 208:1749 – 1769, 2005.
- [16] Nora Molkenhuth and Marc Timme. Scaling laws in spatial network formation. *Physical review letters*, 117 16:168301, 2016.
- [17] Mark E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103 23:8577–82, 2006.
- [18] Schuyler B. Nicholson, Luis Pedro García-Pintos, Adolfo del Campo, and Jason R. Green. Time–information uncertainty relations in thermodynamics. *Nature Physics*, 16(12):1211–1215, December 2020.
- [19] Jorge M. Pacheco, Arne Traulsen, and Martin A. Nowak. Coevolution of strategy and structure in complex networks with dynamical linking. *Physical review letters*, 97 25:258103, 2006.
- [20] Fragkiskos Papadopoulos, Marián Boguñá, and Dmitri V. Krioukov. Popularity versus similarity in growing networks. *Nature*, 489:537–540, 2011.
- [21] Erzsébet Ravasz and Albert-András-László Barabási. Hierarchical organization in complex networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 67 2 Pt 2:026112, 2002.



- [22] Giulio Tononi, Olaf Sporns, and Gerald M. Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences of the United States of America*, 91 11:5033–7, 1994.
- [23] Massimo Vergassola, Emmanuel Villermanx, and Boris I. Shraiman. ‘infotaxis’ as a strategy for searching without gradients. *Nature*, 445:406–409, 2007.
- [24] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442, 1998.
- [25] Albert-*Á*szló Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286 5439:509–12, 1999.