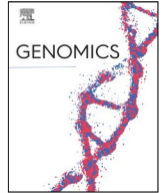




在 ScienceDirect 上提供的内容清单

基因组学

期刊主页: www.elsevier.com/locate/ygeno



无健中枢是前列腺癌网络的潜在靶基因

Irengbam Rocky ^{Mangangchaa,b,c,d,1}, Md.Zubair ^{Malikd,1}, Omer ^{Kucuke}, Shakir ^{Alib}, R.K. Brojen ^{Singhd},

^a School of Interdisciplinary Sciences and Technology, Jamia Hamdard, New Delhi 110062, 印度。

^b Bioinformatics Infrastructure Facility, BIF & Department of Biochemistry, School of Chemical and Life Sciences Jamia Hamdard, New Delhi 110062, India.

^c Department of Zoology, Deshbandhu College, University of Delhi, New Delhi 110019, India.

^d School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India.

^e Winship Cancer Institute of Emory University, 1365 Clifton Road NE, Atlanta, GA 30322, USA.



文章信息

关键词:

枢纽
系统生物学
关键疾病基因
前列腺癌
网络理论
中心性致命性规则

ABSTRACT

利用网络理论方法可以成功研究复杂的疾病网络,有助于寻找关键疾病基因和相关疾病模块。我们研究了由患者基因表达数据集构建的前列腺癌(PCa)蛋白-蛋白相互作用(PPI)网络,发现该网络表现出缺乏中心性致死规则的分层无尺度拓扑结构。从网络中敲除领先中心的集合实验导致从分层(HN)过渡到无尺度(SF)拓扑结构,影响网络的整合和组织。这种过渡, $HN \rightarrow SF$, 由于去掉了大量最高程度的枢纽,导致信息处理效率、信号传播的性价比、紧凑性、节点的聚类能量分布相对下降。在去除前 50 个枢纽后,系统性地从非同化的 PCa PPI 网络过渡到同化网络,然后在进一步去除枢纽后又恢复到非同化性质,这也表明最大的枢纽在 PCa 网络整合中的主导地位。此外,利用 PCa 网络的模块内度和参与系数对枢纽进行功能分类,领先枢纽敲除实验表明,无亲缘关系的枢纽作为构成模块和异质节点之间建立联系的基础,以维持网络稳定。然后,我们通过对枢纽进行 Fisher's 精确检验,检验了敲除实验中枢纽的本质性,结果表明,去除无亲缘关系枢纽对应于网络的最大致死率。然而,过量去除这些中枢基本上可能导致网络崩溃。

1.1.简介

在复杂疾病中,寻找关键疾病基因是理解疾病调控机制、潜在药物靶点和疾病生物标志物的最 important 任务。近年来,基于生物复杂网络的网络医学研究已成为理解疾病模块、识别疾病生物标志物和药物靶点的新兴疾病研究领域之一[1]。在复杂网络中,高度节点、枢纽在维持网络结构、调节功能动态和稳定网络方面发挥着最重要的作用[2,3]。在现实世界中,许多这样的网络呈现出无尺度拓扑结构,枢纽在其中起着核心作用,干扰这些枢纽会影响网络的稳定性和信号传播,从而导致网络的崩溃[3-5]。进一步。

这些网络与高程度中心之间丰富的俱乐部形成的同源性网络拓扑结构相辅相成,使它们更容易受到这种定向破坏[4,6-11]。最近对乳腺癌蛋白-蛋白相互作用(PPI)复合物网络工作的研究表明其构象为无层次规模的拓扑结构,这就提出了这些网络不遵循中心性规则(centrality lethality rule)的问题,即当中心被锁定时,网络的稳定性和动态性会受到干扰,但不会完全受到影响[12-14]。这可能是由于在复杂网络和其他生物逻辑网络中,模块/子模块在不同拓扑层次上的分层排列,它们与特定功能相关联[2,3,15,16]。因此,枢纽分布在这些模块之间,保持着特定的功能,并帮助建立它们之间的联系,以协调网络系统的功能[17]。在真核生物的 PPI 网络中,大多数枢纽是必不可少的。

通讯作者:

电子邮件地址: brojen@jnu.ac.in (R.K.B.Singh)。

¹ 作者有同等贡献。

<https://doi.org/10.1016/j.ygeno.2020.09.033>

2019 年 10 月 19 日收到; 2020 年 8 月 28 日收到修改稿; 2020 年 9 月 14 日接受。

可于 2020 年 9 月 23 日上线

0888-7543/© 2020 由 Elsevier Inc.

细胞增殖和发育所必需的基因,而这些基因的敲除对生物体来说是致命的[5,18].因此,对枢纽的功能分类和研究它们在模块水平上的活动是至关重要的,特别是在理解不同枢纽在分析复杂生物网络中的作用时。因此,对中心的功能分类和研究它们在模块水平上的活动是至关重要的,特别是在了解不同中心在分析复杂生物网络中的作用方面。

前列腺癌(PCa)是仅次于肺癌的第二大常见癌症,也是全球男性的第五大杀手[19],随着高转移性抗雄激素转移性前列腺癌的出现,病情加重。这是由于雄激素受体(AR)的突变和其他改变,或是信号通路导致其表达升高[20],最近有研究显示AR基因敲除可降低PCa细胞的侵袭和迁移[21]。在本研究中,我们利用网络定理了解包括AR在内的PCa相关基因的PPI网络的结构拓扑,特别是高程度节点、枢纽在调控复杂疾病网络中的作用。研究表明,PCa PPI网络表现出分层尺度的自由到pology与模块化组织。Knockout analysis of PCa PPI network was performed to reveal the indispensability of the hubs and network stability. Owing to an inevitable relevance of hubs in hier-archical PCa PPI network, this study classifies hubs on the basis of their participation coefficients (p_i) and within module degree (z score) (z_i). 枢纽敲除研究表明,无亲缘关系的模块枢纽是维持PCa系统稳定的枢纽中最重要的。随后,该研究报道了PCa前100个枢纽中必要基因的最大富集度。

2.2. 材料和方法

2.1. 方法流程图

我们通过图1给出的流程图总结了我们在这项工作中实现的方法和算法。

2.2. PPI网络的构建和节点的敲除。

从GeneMANIA交互组中提取了3871个前列腺癌相关基因的2960个节点和20372个边缘的蛋白质-蛋白质交互网络,这些基因根据BioXpressv3.0 ($FC>1$, 调整后的 p -值 <0.05) (<https://hive.biochemistry.gwu.edu/bioexpress>) 过度表达,根据COSMIC (<https://cancer.sanger.ac.uk/cosmic>) 在PCa患者中突变[22-24]。这

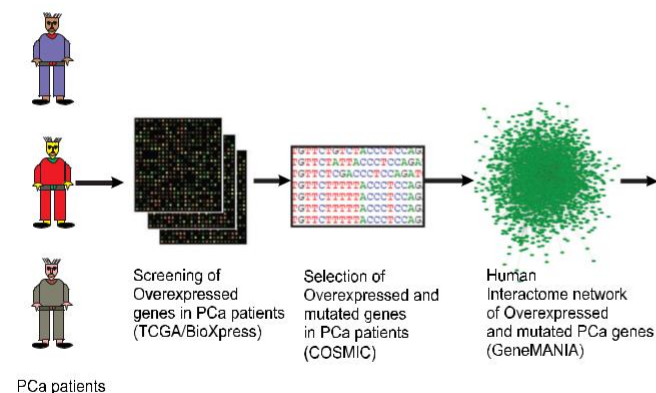


图1. PCa PPI网络研究的实施方法流程图。研究PCa PPI网络的方法实施流程图。

表示PCa PPI初级网络,表示一个图 $G(N,M)$, $N=\{n_i\}$; $i=1, 2, \dots, N$, N 是节点的集合, M 是边的集合, $M=\{m_{ij}\}$; $i, j=1, 2, 3, \dots, N$ 。

通过对PCa PPI网络中的节点及其各自的边缘进行敲除,首先随机生成PCa PPI网络中的基因列表,然后从网络中删除前10,30,50,70,100,200,300,400,500,600,700,800,900和1000个最高度节点。对删除节点后形成的子网络及其各自的边缘进一步分析其拓扑特性。

2.3. 网络的拓扑分析

利用Cytoscape3.6.0中的Network analyzer和CytoNCA[25-27]计算主网络的节点度(k)、中心度(特征向量 CE 、紧密度 CB 、紧密度 CC 、子图 CS)、节点度分布($P(k)$)、聚类系数($C(k)$)和节点邻接性($cn(k)$),并对网络/模块进行敲除实验。模块化程度(Q)采用基于Louvain群体检测算法的Igraph模块化函数计算[28]。使用R中的Igraph包"brainGraph"(<https://github.com/cwatson/brainGraph>)计算富俱乐部系数(ϕ)、参与系数(p_i)和模块内度或 z 分(z_i)。使用Ma-tlab package 计算单片图的局部社区范式相关性(LCP - cor) [29]。

度(k)。它是指一个节点在网络中的连接总数,其计算公式为:

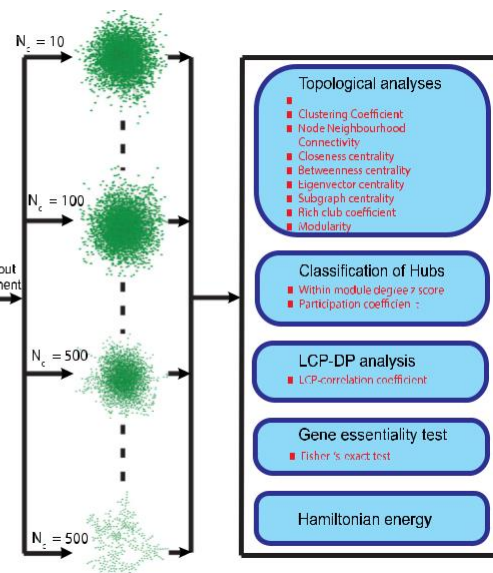
$$k_i = \sum_{j=1}^N A_{ij} \quad (1)$$

其中, k_i 为第 i 个节点的度, A_{ij} 表示图 $G=(N,M)$ 中图的邻接矩阵元素,其中 N 表示节点的集合, M 表示边的集合。

度分布概率, $P(k)$ 。它是整个网络中度数分布的概率,并根据网络中度数的分布情况,定义网络的无尺度、分层、小世界或随机网络的特性拓扑结构[2-4]。其计算公式为:

$$P(k) = \frac{nk}{n} \quad (2)$$

其中, nk 为度数为 k 的节点总数, n , 为总节点数。



在网络中。

本地聚类系数 $c_i(k)$ 。网络中一个节点与其最近的邻居之间的内部联系强度是通过其与最近的邻居形成三角形小团体的能力与网络中三角形小团体的总数来衡量的[2-4]。这样，节点间高度连接的数值较高，计算公式为：

$$c_i(k) = \frac{2m_i}{k_i(k-1)} \quad (3)$$

其中， k_i 为第 i 个节点的度数， m_i 为其最近邻居间的边总数。

平均聚类系数 $C(k)$ ：它衡量网络中节点与其最近邻居的内部连通性的强弱，对于节点间连通性密集的高聚类紧凑型网络，其数值较高，计算方法为网络中总节点数 n 的局部聚类系数的平均值[2-4]。

$$C(k) = \frac{1}{n} \sum_{i=1}^n c_i(k) \quad (4)$$

邻域连通性 $CN(k)$ 。它是网络中一个节点的最近邻居的平均连通性，它定义了网络中的随机性或非随机性混合[6,7,30]，计算公式为：

$$CN(k) = \sum_q P(q|k) \quad (5)$$

其中， $P(q|k)$ 是具有 k 个连接的节点与另一个具有 q 个连接的节点的连接的条件概率。

富俱乐部系数 $\phi(k)$ 。它是高度富节点与低度节点之间形成良好连接的度量，与网络的排序混合有关[6,7,16]。对于一个度数为 k 的节点，其富俱乐部系数用以下公式表示

$$\phi(k) = \frac{2m_{>k}}{n_{>k}(n-k+1)} \quad (6)$$

其中， $m_{>k}$ 为度数 $\geq k$ 的节点之间的边数， $n_{>k}$ 为度数 $\geq k$ 的节点数。

$$\phi_{\text{规范}}(k) = \frac{\phi(k)}{\phi_{\text{兰德}}(k)} \quad (7)$$

其中， $\phi_{\text{rand}}(k)$ 为 si-milar 大小和度数序列的随机网络的富俱乐部系数， $\phi(k)$ 为度数为 k 的富俱乐部系数， $\phi_{\text{norm}}(k)$ 为归一化富俱乐部系数。

密切性中心度 cc ：它是网络中两个节点之间的归一化最短路径距离，它给出了一个节点与网络中其他节点的密切程度[31-33]。它的计算公式为

$$CC(k) = \frac{n}{\sum_{j \neq i} d_{ij}} \quad (8)$$

其中， n 为网络中的节点总数， d_{ij} 为节点 i 和 j 之间的大地路径长度之和

介于**中心性** CB ：它衡量的是网络中节点的中心控制能力，具有较高介于中心性的节点对网络有较高的控制能力，使其在网络中更加中心[32-34]。它的计算公式为

$$CB(v) = \sum_{i \neq j} \frac{d_{ij}(v)}{d_{ij}} \quad (9)$$

其中， $CB(v)$ 为节点 v 的无中心性， $d_{ij}(v)$ 表示节点 v 之间的从节点 i 到节点 j 经过节点 v 的大地测量路径数。

那么，归一化之间的中心度 CB 由以下公式给出。

$$CB(v) = \frac{1}{M} \sum_{i \neq j} \frac{d_{ij}(v)}{d_{ij}} \quad (10)$$

其中， M 表示节点对的数量，不包括 v ，**特征向量中心性** CE ：它衡量一个节点对

网络中，与更有影响力的节点连接度较高的节点将比其他节点具有更高的**特征向量中心性**，对网络中的信号传播有更大的控制力[32,33,35]。

它的计算公式为：

$$CE(i) = \frac{1}{\lambda} \sum_{j \in nn(i)} v_j \quad (11)$$

其中， $nn(i)$ 表示网络中节点 i 的最近邻，特征值 λ ，特征值方程的特征向量 v_i ， $Av = \lambda v$ 其中， A 为网络邻接矩阵。

子图中心度 CS ：它衡量的是节点在网络中子图数量的参与度，最终越有影响力的节点参与度越大，**子图中心度**越高[36]。它的计算公式为

$$CS(i) = \sum_{j=1}^N v_j(i) 2e_j \quad (12)$$

其中 λ_j 为第 j 个特征值， $v_j(i)$ ，为相关特征向量的第 i 个元素。

2.4. 衡量网络紧凑程度的参数

以下参数和**平均聚类系数** $C(k)$ 可以用来衡量网络的聚类强度和紧凑程度。

模块化 Q ：它衡量的是网络划分为模块/社区的能力，**模块化程度**越高的网络，其模块内连接越多，形成模块/社区之间的稀疏和失去连接的节点[15]。**模块化度**的计算公式为

$$Q = \frac{1}{2m} \sum_{i,j} \left[\frac{A_{ij}}{2m} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (13)$$

其中， m 为边数， A_{ij} 邻接矩阵， k_i 为节点 i 的度数， k_j 为节点 j 的度数， c_i 和 c_j 为节点 i 和 j 的分量，如果 i 和 j 属于同一组， δ 为 1，否则为 0。

本地社区范式相关系数($LCP - corr$)：通过 **LCP-分解图**($LCP - DP$)可以评估链路预测及其相关的本地社区范式对网络拓扑的影响。这是通过计算**共同邻居**(CNs)数量与各自的**本地社区链路**($LCLs$)数量之间的**皮尔逊相关系数**来计算的，**本地社区链路**是指交互节点之间的共同第一邻居之间的链路数量，并绘制在一个双向图上[29]。**LCP-DP 相关系数**>0.80 的网络具有高度的紧凑性和动态性，自组织性强。它的计算公式为

$$LCP - corr = \frac{con(CN, LCP)}{\sigma_{CN} \sigma_{LCL}}, CN > 1 \quad (14)$$

其中， $con(CN, LCP)$ 是 CN 和 LCP 的协方差， σ_{CN} 和

σ_{LCL} 是 CN 和 LCP 的标准差。

哈密顿能量：**哈密顿能量** (H) 表征了 partitioning 参数，从而实现更好的网络分区，并在网络中进行分解。

复杂的网络根据不同的情况被分割成更小的子图。

大小和网络密度与分辨率 γ 参数

[37,38].高密度的网络对解析力的影响更大。

tion 参数，从而实现更好的网络分区，并在网络中进行分解。

可分化子图，它是由

$$H = [m^{-\gamma} n] \quad (15)$$

其中， m 为社区/模块中边的总数， n 为

总节点和 γ 为分辨率参数。

2.5. 中心的分类方法

以下参数可用于对网络中的枢纽类型进行分类。

模块内度或 z 分, (z_i)。探讨复杂网络中模块层次的节点作用, 可以通过其在模块中的功能分类来实现。一个节点的模块内 z 分 (z_i) 由其模块度得出, 相应地将节点划分为模块枢纽 ($z \geq 2.5$), 与其他非枢纽模块节点 ($z < 2.5$) 相比, 这些节点在其模块中具有更高的连接度和更大的影响力[17]。其计算公式为

$$Z_i = \frac{k_i - \bar{k}_{si}}{\sigma_{ksi}} \quad (16)$$

其中, k_i 为节点 i 在其模块 si 中的度数, ksi 表示模块 si 中所有节点的度数 k 的平均值, σ_{ksi} 表示度数 k 在 si 中的标准差。

参与协同效率 (P_i)。模块化节点的进一步局部和全局功能作用是通过它们参与建立其自身模块内或模块外节点之间的联系来确定的。一个节点的 P_i 越高 > 0.70 , 说明它参与连接不同模块的程度越高。

$$P_i = 1 - \sum_{s=1}^{NM} \left(\frac{k_{is}}{k_i} \right)^2 \quad (17)$$

其中, k_{is} 为节点 i 对 s 模块中节点的度数, k_i 为节点 i 的总度数, NM 为网络中模块的总数。

P_i 与 Z_i 一起将模块化集线器分为在所有其他模块之间建立节点连接的模块化无缘集线器 ($Z \geq 2.5$ 且 $P_i > 0.70$)、与自身模块外的节点有最大联系的模块化连接器集线器 ($Z \geq 2.5$ 且 $0.30 < P_i \leq 0.75$) 和模块化省级集线器 ($Z \geq 2.5$ 且 $P_i \leq 0.30$) [17]。

2.6. 基因本质分析

OGEE v2 (<http://ogee.medgenius.info>) 在 21 556 个基因中共鉴定了 7 168 个人类必需基因[39]。这些基因对细胞的生存、发展和增殖至关重要, 这在 7 种不同的癌细胞系中得到了证实[39]。以 PPI PCa 原生网络中的前 100, 200, 300, 400, 500, 600, 700, 800, 900 和 1000 个最高程度的基因为一组, 计算这些重要基因在每一组中的百分比, 并采用双尾 Fisher's 精确检验对整个数据集的 21,556 个基因进行富集分析[40]。计算了每个基因组最高程度基因的几率比以及 95% CI 值及其 P 值。

3. 结果

PCa PPI 网络的拓扑特性, $P(k)$ 、 $C(k)$ 及 $CN(k)$ (方法中的公式(2)、(3)、(4))遵循幂律行为, $P(k) \sim k^{-\gamma}$, $C(k) \sim k^{-\alpha}$ 及 $CN(k) \sim k^{-\beta}$, $\gamma \sim 1.92$, $\alpha \sim 0.30$ 和 $\beta \sim 0.006$ (公式(18)、图 2)。由于这种幂律行为是网络中分形或尺度自由度的 signature, 而这个网络又被组织成不同层次的模块和子模块, 因此 PCa、PPI 网络遵循分层无尺度网络拓扑结构[2,3,41]。

3.1. 由枢纽驱动的网络拓扑结构的过渡

枢纽作为最高程度的节点, 在信号传播和维护网络稳定方面承担着重要的责任, 当部分枢纽从网络中移除时, 可能会在网络系统的组织/重新组织中发挥各种作用。这可能反映出

枢纽在网络组织中可能的重要性, 例如, 在信号处理的优化, 网络属性的变化, 系统级组织的变化等。因此, 为了研究网络中枢纽的重要义务, 我们对原发性 PCa PPI 网络中的第一个领先枢纽 ($10 < NC < 1000$, 如 Methods 中给出的) 进行了基因敲除, 并分析了由此产生的网络的 topological 属性 (图 2)。从分析中发现, 这些网络和子网络遵循层次尺度的自由拓扑结构, 其中, $P(k)$ 、 $C(k)$ 和 $CN(k)$ 对度数 k 仍表现出负指数的幂律性质 (式(18))。拟合的幂律曲线对所得到的 $P(k)$ 数据的斜率随着基因敲除中枢纽数量的增加而变化并系统地增加 (如图 2 第一幅上图, $10 < NC < 700$)。由于分层网络的 γ 值为 $\sim 1 + \ln(4)$ (3) ~ 2.26 [3], 而对于无尺度网络 $2 < \gamma < 3.4$ [4], 因此, 我们可以确定三种不同的不同体制, 其中, 网络特性有显著的不同。首先, 分层机制, 其中, γ 的值大约在 1.9 和 2.3 之间, $1.9 < \gamma < 2.3$, 相应的敲除中心数在 0 和 200 之间, $0 < NC < 200$ 。在这种体制下, 网络中模块的出现主导了其中枢纽的出现[2], 系统级组织的模块和各级组织的子模块在网络中的作用更为突出。在这里, 网络缺乏中心控制机制, 因此, 违反了 centrality-lethality 规则[5], 网络一般是自组织的。第二种制度, 可以称为强无规模制度, 是以 $2.3 < \gamma < 4$, 它对应于将 $200 < NC < 500$ 的中心从网络中移除。在这种制度下, 枢纽的功能可能比模块的作用更加突出/主导[3]。在这种体制下, 去除集线器可能对网络工作更具杀伤力, 导致网络组织的严重崩溃。第三, 网络崩溃体制, 其中, γ 的值非常大, $\gamma > 4$, 去除枢纽 $NC > 500$, 导致网络随机崩溃。

$$\begin{bmatrix} P(k) \\ C(k) \\ CN(k) \end{bmatrix} \sim \begin{bmatrix} k^{-\gamma} \\ k^{-\alpha} \\ k^{-\beta} \end{bmatrix} \rightarrow \begin{bmatrix} 1.92 - 4.73 \\ (-0.52) - 0.86 \\ (-0.04) - 0.29 \end{bmatrix} \quad (18)$$

下面提到三种制度下聚类系数的行为。在层次制度下, α 的范围为 $0.10 < \alpha < 0.86$, 在强无规模制度下, α 随着 NC 的增加而减小, $0.86 > \alpha > 0.5$, 而在网络分解制度下, α 的值变得随机, 有时为负, 显示出几乎随机的稀疏网络拓扑结构 (图 2 中行面板)。另一方面, 服从幂律行为的邻域连通性 $CN(k)$ 的结果, 在枢纽活动的驱动下, 表现出两种截然不同的行为及其转变 (图 2 最低行面板)。首先, 观察网络拓扑中的同种性, 当枢纽的敲除范围为 $0 < NC < 100$ ($k \geq 85$), 其中, 指数 β 的值在 $-0.04 > \beta > -0.1$ 的范围内。在这种情况下, 集线器的显著作用在网络中表现出来, 使它们之间的联系更加密集, 这种现象被称为富俱乐部的形成, 除了模块化组织外, 少数集线器开始一起工作, 可能是为了控制网络[16], 提供网络中的功能多样性[42], 并使网络组织中的能量成本最小化[43]。因此, 移除一个或几个这样的枢纽对网络来说是致命的/有毒的[18]。在这种情况下, 网络工作带有中心或富俱乐部拓扑结构的无规模网络的特征[42]。进一步计算这些基因敲除网络的富俱乐部系数 (ϕ) 和归一化富俱乐部系数 (ϕ_{Norm}), 使用公式 (6) 和 (7) 的方法显示, 这些网络中的高位节点对应于 PCa 网络中的中间度节点 ($19 \leq k \leq 107$), 富俱乐部形成的证据 (图 3)。其次, 当敲除中心范围在 $100 < NC < 1000$ 之间时, 观察网络拓扑结构中的失调性, 其中, β 的范围在 $0.01 < \beta < 0.29$ 之间[30]。在这种情况下, 尺度

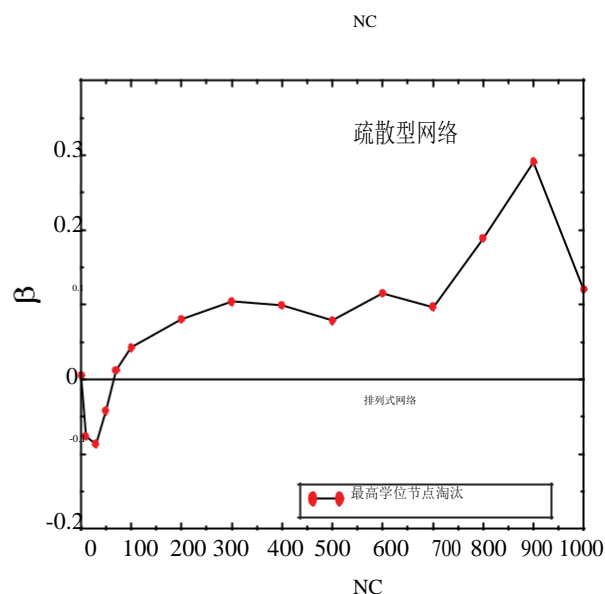
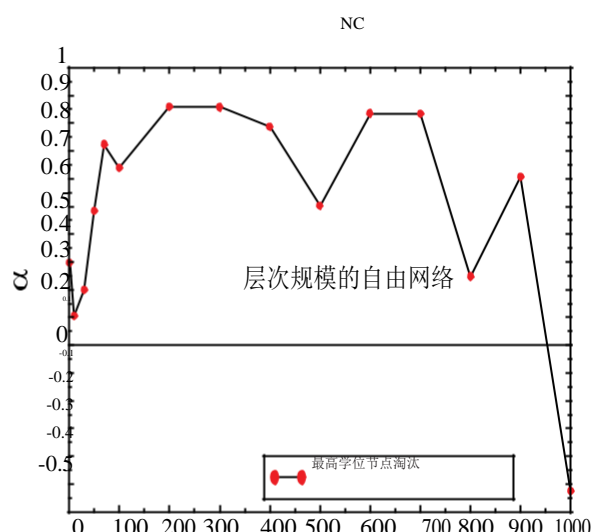
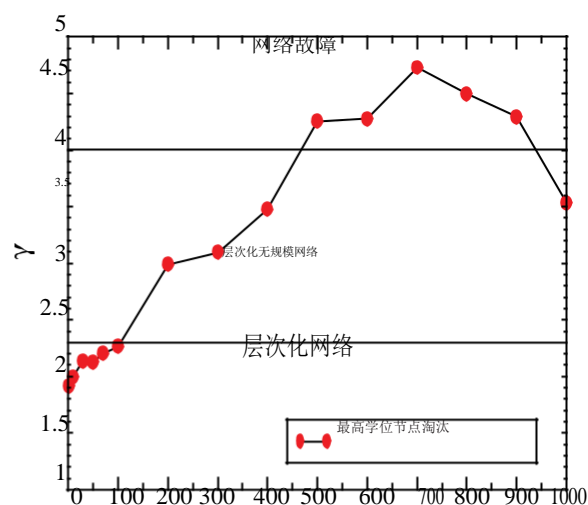


Fig.2.PCa PPI network and sub-networks followed hierarchical scale free topologies.PCa PPI 网络和子网络遵循分层尺度的自由拓扑结构。

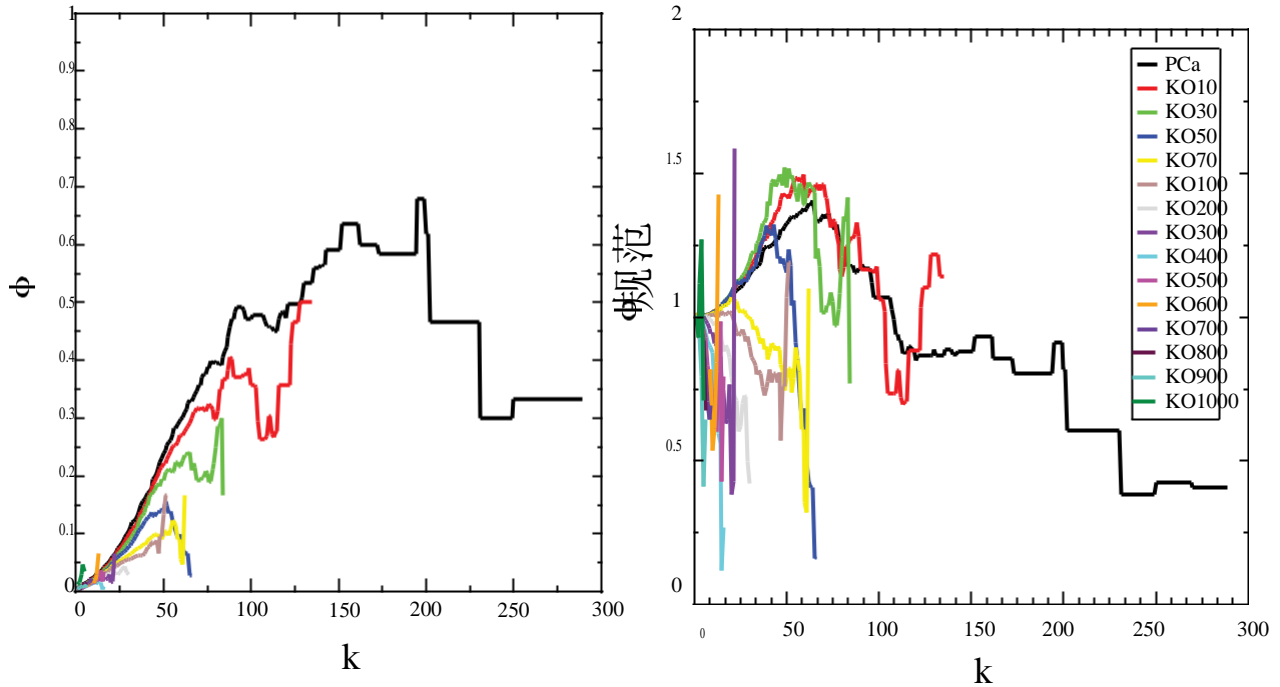


Fig.3. Rich club analysis of PCa PPI network and sub networks. PCa PPI 网络和子网络的富俱乐部分析。

自由拓扑失去了富俱乐部签名，其中，网络失去了 functional 多样性，网络组织中的能量成本增加[42,43]。

然后，我们研究了 PCa 网络中因敲除枢纽而引起的网络信息处理的变化，用中心性参数，即闭合度 CC 、 CB 之间、特征向量 CE 和子图 CS 中心性来衡量（见方法），发现这些中心性参数服从幂律，分别为 $CC \sim k^\epsilon$ 、 $CB \sim k^\eta$ 、 $CE \sim k^\delta$ 和 $CS \sim k^\zeta$ ，其中， ϵ 、 η 、 δ 和 ζ 分别为相应的外因子（图 4）。敲除实验的结果与

$0 < NC < 1000$ ，总结在公式（19）中。 ϵ 、 η 、 δ 和 ζ 的值是通过用相关对应的幂律方程拟合各自的数据点来计算的，其中拟合的良好度范围为 $r^2 = 93\% - 98\%$ 。由于 CC 的值随着 k 的增加而增加（图 4 第一行左图），这表明较大的集线器与网络中其他节点的距离更近，说明集线器与其他节点的交叉对话更密切，增强了节点间的快速信息传播。此外，我们还发现，随着敲除集线器 NC 数量的增加， ϵ 被增加允许在 CC 值增加（图 4 第一行右图）。因此，随着 NC 的增加（ $0 < NC < 500$ ）网络中的信息传播速度越快。然后对于 $NC > 500$ 时，信息内传播变得缓慢，并急剧下降（ $NC > 700$ ），这可能表明网络组织的崩溃。同样，指数 η 的 CB 也随着 k 的增加而增加，说明枢纽具有通过快速信号处理的中枢控制能力（图 4 第二行上图左图）。在基因敲除实验中，重新移动集线器，首先导致 η 随着 NC 的增加而降低（ $NC < 50$ ），然后 η 随着 NC 的增加而单调增加（ $50 < NC < 150$ ），接着 η 随着 NC 的增加而稳定降低（ $150 < NC < 500$ ）（图 4 第二行右图）。进一步增加 NC （ $NC > 500$ ）， η 几乎随机波动。然而，在 δ 相对于 NC 的情况下，情况却大不相同，其中， δ 先是随着 NC 的增加而增加（ $0 < NC < 50$ ），然后随着 NC 的增加而单调地减少（ $50 < NC < 100$ ），接着是随着 NC 的增加而保持不变（ $100 < NC < 500$ ），之后， δ 变得近乎随机。类似的行为也被观察到 CS 与 k 以及 ζ 与 NC 的关系（图 4 最低行面板）。由于这些指数的值为正，因此，领先的枢纽（大 k 节点）。

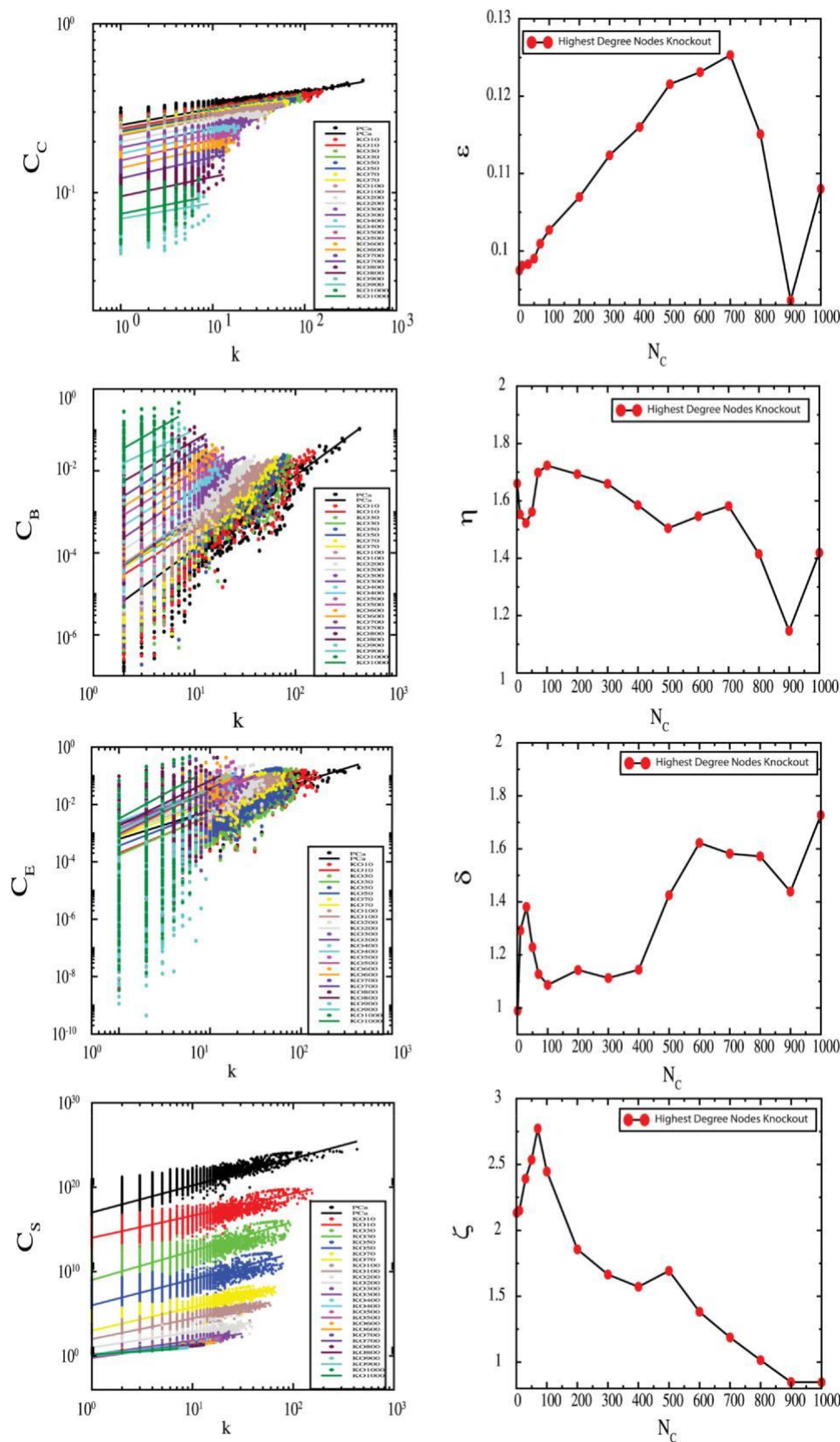
各自的中心值都很大，说明中心是网络中快速信息处理的关键节点，包括模态间、模态内、局部和全局信号传播。

$$\begin{bmatrix} CC \\ CB \\ CE \\ CS \end{bmatrix} \sim \begin{bmatrix} k^\epsilon \\ k^\eta \\ k^\delta \\ k^\zeta \end{bmatrix} \rightarrow \begin{bmatrix} \epsilon \\ \eta \\ \delta \\ \zeta \end{bmatrix} = \begin{bmatrix} 0.09 & 0.13 \\ 1.15 & 1.72 \\ 0.99 & 1.73 \\ 0.85 & 2.77 \end{bmatrix} \quad (19)$$

The knockout of hubs experiment clearly showed the significant change of network topology driven by hubs in the network. The observed clear transition from hierarchical to scale free properties and vice versa was due to the dominance of activities of modules over hubs in the network and vice versa.然而，模块和集线器都可能承担起建设性的责任，以维持网络的稳定性和组织性，而过多地将集线器从网络中移除，或者如果系统中没有领导者，网络将无法运行或可能崩溃。

3.2. 中枢扰动危及 PCa 网络的自组织性。

集线器扰动对网络相当敏感，可能对网络系统造成致命的影响，但注入系统网络的致命性大小取决于网络拓扑结构。例如，在无规模网络中，集线器扰动可能导致网络崩溃[5,18]。因此，我们通过两种方式对 PCa 网络进行比较敲除分析，首先是随机删除 PCa 网络中的节点，其次是依次删除网络中的领先枢纽。然后，我们通过计算敲除节点后 PCa 网络和子网络的平均聚类系数 $C(k)$ 、模块化 Q 、局部社区范式相关系数 $(LCP - Corr)$ 和 Hamiltonian 能量 H （ H 的计算结果为 $\gamma=0.5$ ）来分析网络 properties，利用公式（4）、（13）、（14）和（15）分别计算出 PCa 网络和敲除节点后的子网络的希尔顿能量 H 。平均聚类系数 $C(k)$ 作为 NC 的函数的行为表明，在 $NC \rightarrow [0 - 1000]$ 之间的领先枢纽的系统去除显示出显著的变化与枢纽的去除，其中它单调地随着 NC 的增加而减少（图 5 左上角面板）。然而，在 NC 节点被 randomly 从网络中移除的地方并没有显示出显著的变化。



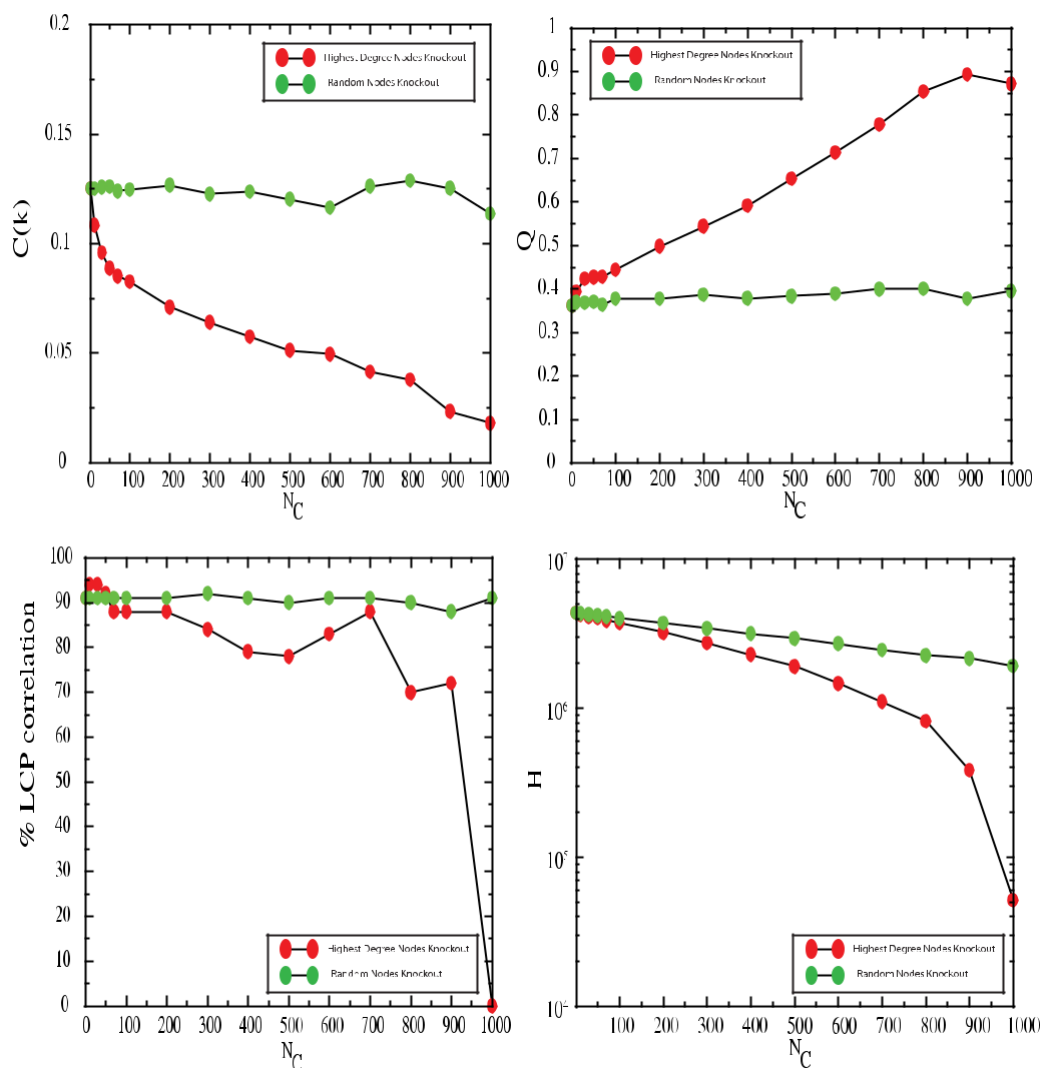


图 5.高度枢纽高度枢纽对网络的稳定性和完整性非常重要。

$C(k)$ 相对于 N_c 。这表明枢纽在维持网络聚类和密度方面的重要作用。在随机节点去除过程中，模块化 Q 与 N_c 的关系也表现出类似的行为，而在系统化的枢纽去除过程中， Q 随着 N_c 的增加而增加(图 5 右上图)，显示出模块内连接增加而模块间连接减少的现象[15]。因此，枢纽节点在 PCa 网络中承担了更多的模块内和模块间组织责任。因此，枢纽节点在 PCa 网络中承担了更多的模块内以及模块间组织的责任。接下来，我们还发现随机节点去除实验的 LCP 相关性%没有明显变化，大致在 90% 左右波动(图 5 左下图)，而在系统性去除领先枢纽的实验中，随着 N_c 的增加，LCP 相关性从 95% 急剧下降到 0% 以下。这些结果表明，枢纽在维持网络紧凑性方面又发挥了重要作用。然后，我们分析了两种情况下的基因敲除实验中，通过 Hamiltonian 能量函数(H)测得的网络中的能量分布作为 N_c 的函数(图 5 右下图)。结果发现，在系统去除领先枢纽敲除过程的情况下， H 作为 N_c 的函数下降的速度比随机去除节点的情况下快得多。因此，枢纽在网络中的能量分配中表现出维持网络组织/重组的重要作用。

从 PCa 网络的随机和系统敲除实验中我们可以推断出，在维持网络稳定、优化信号方面起重要作用的是枢纽

的传播和网络组织/重组。一般来说，在分层网络中，集线器分布在模块的层面上，被特别分为无亲属、连接器和模块化集线器[17]，对网络的扰动非常敏感，它们是模块内和模块间节点的连接者，是网络中的能量分配器，在本地和全球范围内扮演着许多其他角色。它们是模块内和模块间节点的连接者，是网络中的能量分配器，在本地和全球范围内扮演着许多其他角色。将这些枢纽从网络中移除，可能会对网络产生致命的影响，也可能导致网络拓扑结构的转变，甚至造成网络的崩溃。

3.3.无健枢纽是 PCa 网络完整性的关键

通过计算网络构成节点的模块内度(z 分)和相应的参与系数 p_i ，对 PCa PPI 网络中的节点进行功能分类，每-形成[17]。利用公式计算 PCa PPI 网络及各子网络的构成节点的 z 及其对应的 p_i 值，从领先枢纽 ($0 < N_c < 1000$) 系统性敲除。(16)和(17)如方法中所述。计算出的 PCa 的 z 和 p_i 以及由此产生的子网络的分析表明，无亲属(图 6 中上图中的 R7)和连接器(图 6 中上图中的 R6)枢纽在初级 PCa 网络以及分别敲除领先枢纽 $N_c \rightarrow [300 - 400]$ 和 $N_c \rightarrow [600, 700]$ 后的网络中是最大的，而

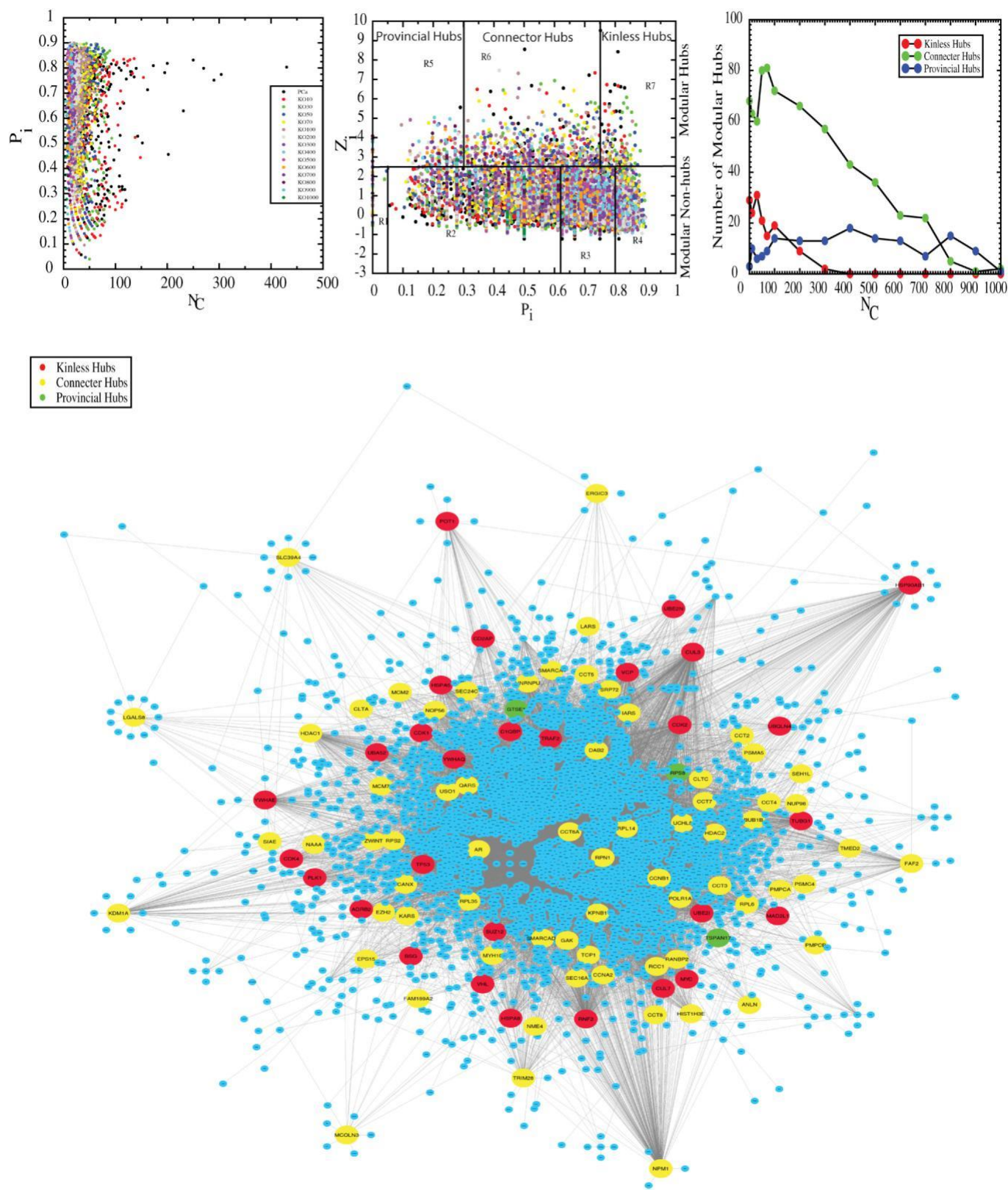


图 6.PCa PPI 网络中模块化枢纽的分类。PCa PPI 网络中模块化枢纽的分类。

省级（图 6 中上图 R5）枢纽在整个网络计算过程中几乎保持一致。这些结果表明，无缘枢纽和连接器枢纽是 PCa 网络中各模块内节点以及各构成模块间节点的关键纽带。此外，这些无缘关系的枢纽对以下方面的变化很敏感

网络拓扑结构，由于去除领先的枢纽 N_C 和执行几乎类似的责任，在所有的网络，主要 PCa PPI 网络工作，以及敲除 N_C 枢纽后的结果网络。因此，无缘关系的枢纽负责本地以及全球网络的整合，以维持网络稳定，最佳信号处理。

表 1
PCa PPI 网络的模块化枢纽。

S.No.	无缝隙集线器	程度(k)	S.No.	连接器集线器	程度(k)	S.S. No.	连接器集线器	程度(k)	S.S. No.	外围集线器	程度(k)
1.	CUL3	430	1.	NPM1	231	39.	NUP98	44	1.	RPS8	120
2.	RNF2	304	2.	HNRNPU	202	40.	PSMA5	42	2.	GTSE1	37
3.	TP53	290	3.	HDAC1	152	41.	QARS	41	3.	TSPAN17	13
4.	CUL7	270	4.	RPL35	143	42.	SEC24C	38			
5.	CDK2	250	5.	NOP56	118	43.	MYH10	38			
6.	VCP	200	6.	RPL6	113	44.	SLC39A4	37			
7.	MYC	195	7.	SRP72	113	45.	MCM2	37			
8.	HSP90AB	173	8.	RPS2	108	46.	ZWINT	37			
9.	VHL	140	9.	RPL14	107	47.	BUB1B	37			
10.	UBE2I	136	10.	EZH2	91	48.	CCNA2	37			
11.	SUZ12	130	11.	CLTC	87	49.	IARS	35			
12.	YWHAQ	129	12.	HDAC2	87	50.	EPS15	34			
13.	TUBG1	121	13.	AR	86	51.	LGALS8	31			
14.	CDK1	117	14.	CCT3	84	52.	PSMC4	31			
15.	HSPA5	111	15.	TRIM28	79	53.	SEH1L	30			
16.	HSPA8	101	16.	HIST1H3E	74	54.	GAK	29			
17.	YWHAE	96	17.	KPNB1	73	55.	KARS	29			
18.	C1QBP	92	18.	RANBP2	72	56.	ANLN	27			
19.	ADRB2	84	19.	KDM1A	72	57.	CLTA	26			
20.	PLK1	69	20.	FAF2	67	58.	LARS	26			
21.	UBA52	64	21.	RPN1	67	59.	POLR1A	24			
22.	MAD2L1	61	22.	CCT4	66	60.	DAB2	23			
23.	TRAF2	58	23.	CANX	64	61.	FAM189A2	21			
24.	CD2AP	55	24.	CCT2	63	62.	NME4	21			
25.	POT1	55	25.	SMARCA4	63	63.	ERGIC3	20			
26.	CDK4	51	26.	TMED2	62	64.	PMPCB	20			
27.	BSG	47	27.	RCC1	56	65.	MCOLN3	19			
28.	UBQLN4	40	28.	SMARCA4	56	66.	PMPCA	19			
29.	UBE2N	37	29.	CCT7	55	67.	NAAA	17			
			30.	CCT6A	54	68.	SIAE	14			
			31.	TCP1	54						
			32.	CCT8	51						
			33.	USO1	50						
			34.	MCM7	49						
			35.	SEC16A	47						
			36.	UHL5	47						
			37.	CCNB1	46						
			38.	CCT5	46						

和自组织。这个结果符合网络从分层（更紧凑的集群）到无规模拓扑（相对而言，枢纽的作用不那么紧凑的）的过渡行为。
(s) 是显著的），再到如图 5 所示的高度稀疏网络，节点的聚类能力不断降低，模块化程度不断提高，从而使模块之间的连接变得松散，多重影响网络的整体拓扑结构。
然后，我们，估计了这些模块化枢纽在领先枢纽敲除实验中的数量及其在维持网络完整性中的作用（图 6 右上角）。结果显示，随着 NC 的增加，无缘枢纽的数量（ n_{kh} ）减少最快，其中， $n_{kh} \rightarrow 0$ ，因为 $NC \rightarrow 500$ ，这是 PCa 网络崩溃的开始。因此，无亲缘关系枢纽的去除导致 PCa 网络的快速崩溃。接下来更负责任的枢纽是连接器枢纽，从结果可以看出（表 1）。因此，无仁（主要）和连接器集线器是 PCa 网络完整性的关键，对它们的攻击可能会对网络造成致命的影响。这些模块化集线器在 PCa 网络中的表现如图 6 下图所示。

3.4. 费舍尔精确检验。前 100 个中心显示出最大的杀伤力

我们应用 Fisher's exact test[40]对 PCa 网络中最高度节点集合中的必要基因进行了富集分析，以确定去除后可能对网络造成致命的必要枢纽。我们针对一组人类 21 556 个基因[39]报道的 PCa 网络中共 7 168 个必要基因（图 7），做了这个检验，找出 $0 \leq NC \leq 1000$ 范围内每个 NC 主导枢纽（基因）的%。结果显示，在最高度 100 节点的组中，90%为必要基因。

表明在其他组最高度节点中富集度最高， p 值最低，ODDs 比值最高（图 7）。
所得结果表明，在最高程度的枢纽中，前 100 个枢纽是对网络中诱导的扰动最敏感的基本基因。有趣的是，我们还发现前 100 个枢纽内的无亲缘关系枢纽和连接器枢纽的数量是最大的（图 6 上行右图）。因此，去除前 100 个枢纽将导致 PCa 网络中最大的致死率，这可能会导致网络结构和属性的转变。

4.4. 讨论情况

生物分子系统的功能特性一般是通过蛋白质之间的相互作用来实现的，这种有组织的 PPI 网络表现出不同的拓扑结构，[1-3,15]。这些 PPI 网络表现为无尺度拓扑结构，其中高程度的节点、枢纽在信号传播中起着核心作用，并维持网络的稳定性；或具有系统级模块化组织特征的分层无尺度拓扑结构，其中模块代表不同的功能/通路模块，[2,3,15]。在 PCa PPI 网络中，层次无尺度拓扑的表现形式表明，高程度的节点、枢纽协调特定的模块功能/通路，并在这些模块之间建立连接，维持系统的完整性[16,41]。在这样的网络中，枢纽和模块的相互作用是非常重要的，一般这些枢纽是多功能的。本研究对 PCa PPI 网络的一个重要发现是观察到前 50 个最高程度的枢纽（ $k \geq 85$ ）重新移动后形成的网络中存在富俱乐部形成的现象，表示为

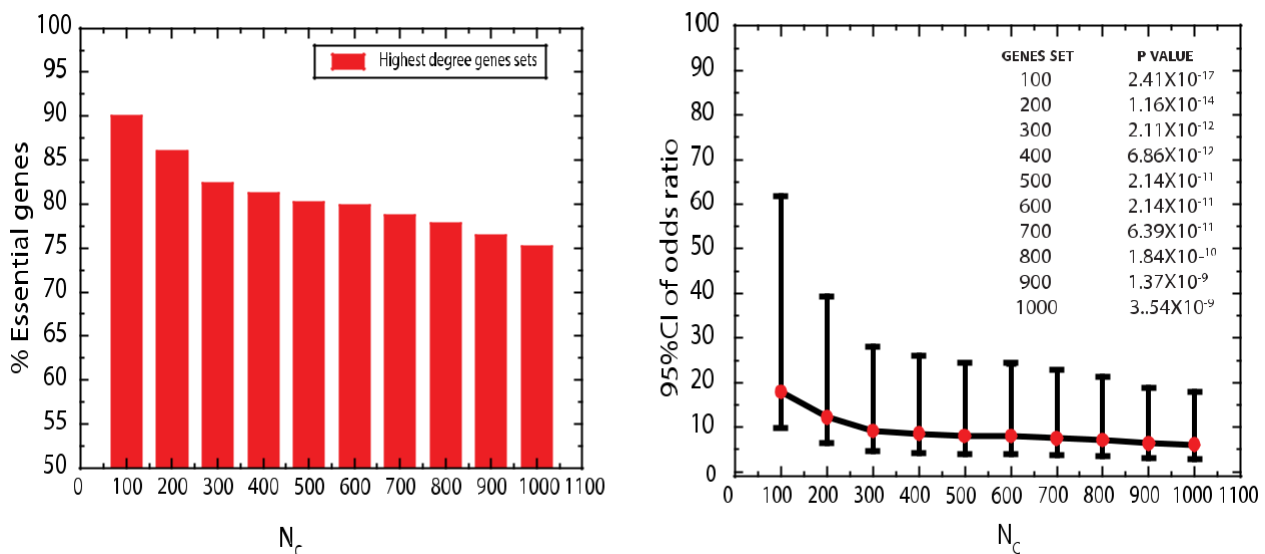


Fig. 7. Essentiality of genes are highest amongst top 100 highest degree hubs of PCa PPI network.

网络同向性的特征[6,7,30]), 然后在进一步删除枢纽时恢复到非同向性(图2)。这说明最大的枢纽在网络完整性中占据主导地位, 主要维持着 PCa PPI 网络的模块化组织, 而它们的删除会使网络转变为一个更多的同向性网络, 其中较高度度的枢纽[7,16]在它们之间建立了强大的连接, 使网络对随机攻击更具弹性, 但同时也增加了它们对枢纽的定向攻击的脆弱性[3,4,9]。这些中心站的脆弱性和网络的完整性被观察到, 系统性地从高度节点从网络中重新移动, 扰乱了系统的自组织。这导致了一个更模块化的分层网络的拓扑结构过渡, 系统中不同功能或路径之间有很强的通信, 首先是规模自由的 *topology* 到稀疏的随机和解体系统(图2)。在这一过渡过程中, 由于最高度枢纽的重新移动, 以中心性(CB、CC、CE和CS)测量为特征的网络中信号传播的信息处理效率和成本效益明显降低(图4)[44]。PCa PPI 网络中高程度节点删除后 centrality 测量值的变化表明, 网络中任何一对节点之间的平均路径长度减少, 将其还原为小世界网络拓扑结构, 并降低了节点在网络内融合中的参与度, 鼓励节点形成更小群体的隔离[45,46,47,48]。由于从原始网络中敲除了大量的最高度枢纽, 网络工作的紧凑性(由 LCP-correlation 测量)、节点的 clustering(由网络拓扑参数表示)和能量分布(由 Hamiltonian energy, H 归属)相对减少(图5)。因此, 对最高度节点尤其是前 100 个枢纽的敲除将 PCa 网络从一个紧凑的、高度集群的网络转变为具有松散约束的、模块间通信较弱的稀疏网络(图5)[47,48]。此外, 在去除最大的枢纽, ($k \geq 85$)时, 网络从无序性向有序性的对称性转变, 也说明了中间度节点 ($19 \leq k \leq 107$) (图3)[41], 特别是无模数枢纽和连接器枢纽(表1), 在连接高度节点和低度节点, 建立强大的 PCa 网络完整性和稳定性方面的重要性。然后, 过量去除高度枢纽导致网络断裂机制完全失去网络完整性。在这种情况下, 网络变得相当稀疏, 几乎失去了所有重要的网络拓扑特性。

从高度组织化的结构向稀疏随机的结构过渡。

和解体的系统与失去最多的模块化无亲缘关系和连接器枢纽有关, 这些枢纽负责连接不同的功能模块和途径(图6)[17,49,50]。PCa PPI 网络的一些模块化无亲缘关系枢纽(如 c-MYC 和 p53)和连接器枢纽(表1)(如雄激素受体, AR)在前列腺肿瘤的发育和进展中发挥着重要作用[51]。Gene amplification of the cellular proto-oncogene MYC increases cell proliferation, survival and metastasis specially in androgen-deprived PCa patients[51]。基因组广泛关联研究也报道了 MYC 基因中的几个易感位点增加了这些基因位点发生某些突变的人的 PCa 风险[52]。肿瘤抑制基因 p53 是在包括 PCa 在内的各种癌症中发现最常见的突变和失去肿瘤抑制功能的基因[51,53]。P53 通过诱导细胞周期停滞、DNA 修复、细胞凋亡等参与细胞的肿瘤抑制, 其在癌症中功能的丧失是由于基因突变或缺失所致[53]。突变的 p53 蛋白在高等级前列腺肿瘤中可再移植, 具有高度转移性特征和晚期阶段[51]。前列腺肿瘤组织中从雄激素依赖型向雄激素独立型前列腺肿瘤的转变也与该肿瘤抑制因子的突变有关[51]。雄激素受体过度表达增加其转录活性, 增强了 PCa 内诱导细胞生长和增殖的雄激素信号和依赖性[51-53]。AR 基因突变使受体即使在较低水平的雄激素中也能保持刺激状态, AR 基因扩增使其表达增强, 与雄激素剥夺疗法的抗性有关, 且多见于激素难治性前列腺癌[51-53]。

此外, 在 PCa PPI 网络的 19 个新型关键调控因子中, CUL7、HSPA5 是模块化无亲缘关系枢纽, CCT4、RANBP2 和 RPL6 是模块化连接器枢纽[41]。这些蛋白与不同癌症相关, 诱导细胞生长和增殖, 抑制细胞凋亡, 并被报道为多种癌症(in-包括 PCa)的预后标志物[41]。此外, 在 PCa 的关键调控因子中, 有 17 个调控因子(ACCT4、EIF3A、HSPA5、NOP2、RANBP2、RPL11、RPL15、RPL19、RPL23A、RPL3、RPL5、RPL6、RPLP0、RPS11、RPS8、RPSA 和 SNU13)是人类生存的必要基因[39,41]。由于在 PCa PPI 网络的前 100 个最高度节点、枢纽(90%)中观察到的必要基因富集度最高, 因此在这前 100 个枢纽中也发现了最多的无亲缘关系(共 29 个, 其中 16 个)(图7)。因此, PCa PPI 网络的高位节点对应着大部分的无亲缘关系的枢纽和重要基因, 说明它们对 PCa 细胞系统的维持和多重生存的重要性[5,18]。这些重要的无亲缘关系基因的丢失

PCa PPI 网络中与重要基因相对应的无亲缘关系和连接器枢纽可能会诱导蛋白质之间的相互作用发生变化, 最终导致 PCa 细胞的功能/通路模块解体, 使其难以生存。因此, 无亲缘关系的枢纽是 PCa PPI 网络中最重要的大脑, 可能是开发抗 PCa 药物的潜在靶点候选。这些模块化枢纽的突变及其表达的失调可能会改变蛋白质之间的相互作用, 影响多蛋白复合体 formations 和信号通路, 扰乱系统的动态, 导致 PCa 和其他相关疾病的肿瘤发展[54]。因此, 这些模块化枢纽可以作为 PPI 网络的驱动节点, 扰动它们可以引起网络动态的变化, 改变生物系统的表型特征[55]。

5.5. 结论

我们根据 PCa 患者的基因表达数据构建了 PCa PPI 网络, 并对其拓扑特性进行研究, 以了解中枢在网络调控和组织中的作用/s。在研究中发现, PCa PPI 网络遵循无层次规模的拓扑结构, 节点之间存在不同步混合, 表明其不遵循中心性致死规则, 缺乏中心控制的枢纽在网络组织中的作用。在系统性地剔除网络

中最高程度的节点或 *枢纽*时，我们观察到*网络拓扑结构*从*分层拓扑结构*向*无规模拓扑结构*的系统性转变，而后在过度剔除领先枢纽时，网络变得稀疏而随机，呈现网络崩溃的现象。这种网络拓扑结构的转变与网络的自组织性和紧凑性的丧失有关，同时减少了节点之间的聚类和信息处理。此外，将集线器分为模块式无缘集线器、连接器集线器和外围集线器，并对它们进行分析，发现*无缘集线器*的丢失，是网络中所有模块相互连接的最大责任，对应于网络组织的最大杀伤力，其次是*连接器集线器*。此外，最大数量的*无亲缘关系枢纽*被发现是负责保持网络稳定和属性的生存的必要基因。因此，我们认为，*无亲缘关系的枢纽*基因可以成为药物设计和开发的良好潜在目标。

竞争利益声明

无人申报。

鸣谢

IRM 感谢德里大学 Deshbandhu 学院提供的攻读博士学位的假期。MZM 感谢印度政府卫生和家家庭福利部卫生研究司根据青年科学家计划提供的财政援助 (制裁文件号：R.12014/01/2018-HR，FTS 号：3146887)。SA ac-knowledge the Department of Biotechnology, Ministry of Science and Technology, Government of India for the bioinformatics facility at Jamia Hamdard under BTISNet, the Biotechnology Information System Network (Sanction no.BT/BI/25/062/2012(BIF).SA 和 OK ac-knowledge Indian Council of Medical Research for International fel-lowship to SA to visit Emory Winship Cancer Institute, Atlanta.RKBS 感谢印度贾瓦哈拉尔-尼赫鲁大学和印度生物技术系卓越中心（COE）的资助。