



复杂生物系统临界状态的识别与预警*

刘锐, 杨茜然

华南理工大学数学学院, 广东 广州 510640

摘要: 很多生物系统的动态发展过程中都存在状态的临界迁移现象, 例如复杂疾病, 在离临界状态较远时, 病情不明显; 而到达临界状态后, 病情有可能在很短的时间内从稳定期突然恶化而成为重病期. 如何基于高维生物医学数据, 找到可用于识别生物系统状态改变临界期的标记物, 对包括疾病的早期预警等课题具有重要意义. 围绕生物系统临界状态识别与预警这一课题, 本文主要综述了动态网络生物标志物(DNB)方法, 及针对不同的数据条件所发展的一系列适用的改进方法及其应用.

关键词: 状态迁移; 临界状态; 动态网络生物标志物(DNB); 早期预警信号

中图分类号: O29 **文献标志码:** A **文章编号:** 2097-0137(2024)06-0275-16

Identification and early warning of critical states in complex biological systems

LIU Rui, YANG Xiran

School of Mathematics, South China University of Technology, Guangzhou 510640, China

Abstract: Many biological systems exhibit critical transitions in their dynamic development processes. For instance, in complex diseases, the condition may not be apparent when far from the critical state; however, upon reaching the critical state, the condition may rapidly deteriorate from a stable phase to a severe phase. It is of great importance to identify biomarkers from high-dimensional biomedical data that can be used to recognize the critical periods of state changes in biological systems, including early warning of diseases. This paper reviews the methods of dynamic network biomarkers and a series of applicable modified methods developed for different data conditions and their applications in the identification and early warning of critical states in biological systems.

Key words: state transition; critical state; dynamic network biomarkers(DNB); early warning signal

1 复杂生物系统中普遍存在临界状态

在许多自然科学和社会科学领域, 我们经常观察到复杂动力系统经历状态转变的现象, 即系统状态的临界改变. 当系统处于特定条件下时, 微小的变化或外界扰动可以引发系统状态的定性变化, 例如物种灭绝、生态环境恶化以及市场崩溃等. 这类复杂自然系统往往存在临界点或临界状态, 在该点附近, 系统

*收稿日期: 2024-04-09

录用日期: 2024-05-02

网络首发日期: 2024-07-31

基金项目: 国家自然科学基金(62172164, 12026608)

作者简介: 刘锐(1983年生), 男; 研究方向: 非线性动力系统、计算生物学与生物信息学;

E-mail: sclirui@scut.edu.cn

通信作者: 杨茜然(2001年生), 女; 研究方向: 非线性动力系统、计算生物学与生物信息学;

E-mail: mayangxiran@mail.scut.edu.cn

全文阅读



ZR20240109

在短时间内发生从一种状态到另一种状态的突变. 例如在地球系统中, 海洋环流或气候可能会突变(Lenton et al., 2008), 冰川时期可能突然结束(Lenton, 2011). 在环境科学中, 湖泊可能在短时间内从维持数千年的贫营养状态转变为不利于动物生存的富营养状态(Wang et al., 2012), 牧场、鱼类种群或野生动物种群可能经历灾难性变化(Rietkerk et al., 2004; Drake et al., 2010). 在金融领域, 我们可能观察到系统性市场崩溃(Kambhu et al., 2013)、经济衰退(Mukherji, 2013)以及大型公司的破产(Quax et al., 2013). 由于状态迁移在短时间内完成, 使得系统从一个稳态进入另一个不可逆的稳态, 因此这一现象也被称为“灾变”(catastrophic shift)(Scheffer et al., 2009). 从数学的角度来说, 这样的系统状态变化通常具有两个特点: 首先, 系统至少具有两个稳定的平衡状态, 在稳定状态下, 外界的持续扰动只会使系统缓慢变化; 第二, 在两个稳定状态之间存在着一个短暂的临界状态, 处于该状态时, 系统的回复性很弱且对扰动敏感, 容易被外界的扰动驱使而迅速地失稳, 从而进入另一个稳定状态(Scheffer et al., 2009). 上述的复杂系统大多是由多个相互作用的变量或子系统组成, 往往拥有多个自由度. 这些系统的复杂性不仅来自于多个变量/多个子系统的动力学, 更重要的是来自于这些变量/子系统之间的相互作用(或耦合)关系的非线性动力学(Chen et al., 2010). 在这类系统中, 几个关联变量或一个局部子系统的变化就可能导致整个系统的状态突变.

在生命科学和医学领域的研究中, 同样普遍观察到生物系统的发展过程中存在系统状态的临界变化现象. 举例而言, 在细胞分化(Bargaje et al., 2017)、细胞命运决策(Sarkar et al., 2019)、组织急性损伤(Sciuto et al., 2005)、器官失能的急性发作(Olde Rikkert et al., 2016)、机体抗药性形成(Liu et al., 2019a)等复杂生物过程中, 系统随着时间的变化从一个稳定状态转变为另一个稳定状态. 这些生物系统状态的临界变化可能由外部因素引起, 如生活环境变化、药物诱导等; 或者由内部生理机制变化引起, 如衰老和基因突变等. 在许多复杂疾病中, 这种状态临界变化反映了疾病的突然恶化. 如图 1 所示, 生物系统迅速从相对正常状态转变为疾病状态(Viera, 2011; Liu et al., 2014a). 由于这种状态变化通常是不可逆的, 系统很难从疾病状态恢复到相对正常状态. 对很多患者来说, 早期诊断和早期治疗是提高疗效、挽救患者生命的关键. 然而, 复杂疾病的早期往往不具备典型的症状和体征, 不易引起重视. 现今威胁人类健康的复杂疾病, 如 2 型糖尿病(Li et al., 2014)与其他代谢功能综合征(Koizumi et al., 2020)、肿瘤细胞远端转移(Chen et al., 2006; Tanaka et al., 2015)、抑郁症等(Stapelberg et al., 2019), 一旦形成后生物系统大多处于稳定状态, 因此治愈难度较大. 这一类具有突然恶化现象的疾病都具有一个很相似的特点, 即在病情恶化过程中存在一个突变点, 在离突变点较远时, 病情不明显, 这往往使得病人忽视了病情, 耽误了治疗的最佳时机; 而到达突变点时, 病情不再是平缓地发展, 而是在很短的时间内从稳定期突然恶化而成为重病期(如图 1 所示). 正是由于这个原因, 对这类疾病的确诊常常不及时, 而一旦疾病恶化就会面临治疗难度大、疗效差等棘手的问题, 所以肿瘤等疾病恶性突变后患者生存时间一般很短. 如何及时地早期诊断这类复杂疾病, 关键在于找到疾病突然恶化的预警特征或临界信号, 从而预警突变点, 这已经成为了生物信息和数学医学研究的一个热点问题.

从复杂系统和动力学的观点来看, 上述许多疾病的发展和恶性转化过程可以看作随机动力系统的时间动态演化过程, 把影响疾病的外在因素视为动力系统参数, 把参与疾病演变的分子浓度当作系统中的状态变量, 于是疾病的突然恶化现象就对应了系统的状态突变/分岔现象(Shi et al., 2016; 2021). 这样一来, 根据疾病的动态发展规律, 很多具有临界突变现象的复杂疾病动态发展过程可以用 3 个阶段来进行描述和建模(图 1), 即正常状态、临界状态(或前疾病状态)、疾病状态(Chen et al., 2012; Lesterhuis et al., 2017). 其中正常状态描述正常阶段或病情较疾病期轻微的缓慢变化阶段, 包括疾病的潜伏阶段、癌变前的慢性炎症阶段或者病情得到有效控制而处于相对健康的阶段, 这时患者的生物系统处于一个稳定状态, 对扰动具有较强的抗性; 前疾病状态是系统处于临界点之前的一个临界时期(或正常期的极限状态), 这时患者的生物系统处于不稳定状态, 对扰动很敏感, 在该阶段积极地改变生活方式和适当的医疗措施可以使系统重新恢复到正常状态(可逆性)(Viera, 2011). 然而, 如果没有及时的医疗干预和生活方式的改变, 疾病将进一步发展——在外界或内部不良因素(不良生活习惯、过大的压力、代谢紊乱等)的持续驱动下,

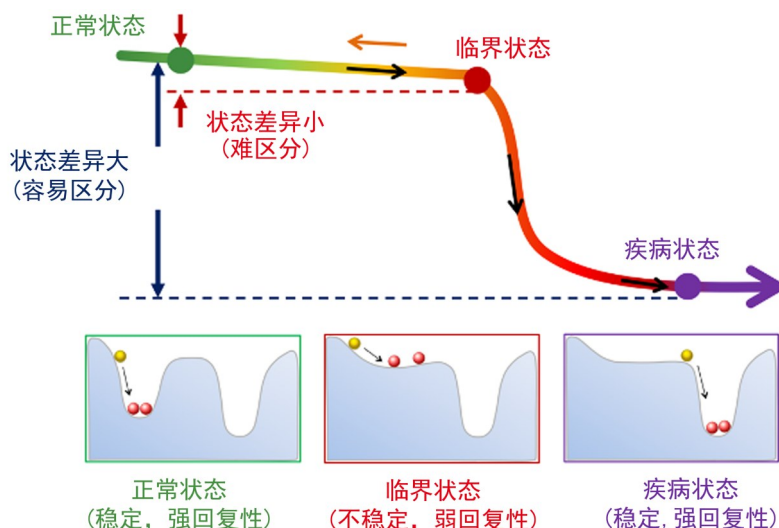


图1 复杂疾病发展的3个状态

Fig. 1 Three states of complex disease development

系统就很容易越过临界点, 进入另外一个稳定状态, 即疾病状态, 该状态代表病情已经恶化成为重病期, 或者慢性炎症已经恶性转化成为癌症. 对诸如恶性肿瘤、糖尿病等复杂疾病来说, 当系统已经到达疾病状态这一阶段时, 治疗的难度非常大, 很难再使系统回到正常状态(非可逆性). 因此, 前疾病状态是疾病发生发展过程中的关键时间节点, 因此预警疾病临界点的关键就在于及时发现系统处于前疾病状态的信号 (Liu et al., 2014a). 在生物医学领域, 尽管已经有大量基于生物分子数据的工作对复杂疾病进行研究, 但传统基于生物分子标志物的疾病诊断是为了区分正常状态和疾病状态, 即如果待检样本中分子标志物的表达值显著高于/低于正常值, 就能判断系统是处于疾病状态. 因此, 寻找分子标志物是基于具有差异表达的分子或分子群的“对照式研究”. 然而前疾病状态是正常期的极限状态(如图1所示), 分子表达的均值和在正常状态时并没有大的差异. 因此, 与诊断疾病状态不同, 对很多复杂疾病来说, 诊断前疾病状态或疾病的恶性突变预警是一个非常困难的问题.

2 临界慢化现象与动态网络生物标志物(DNB)

2.1 临界慢化现象

根据对自然现象中复杂动力系统的长时间观察, 研究者们基于动力系统分岔理论提出了状态的“临界慢化”(CSD, critical slowing down)作为研究非线性动力系统状态临界变化的指标 (Tredicce et al., 2004; van Nes et al., 2007; Rietkerk et al., 2021). 临界慢化是通过一维动力系统来进行描述的现象, 在系统的状态受到扰动的情形下, 考虑如下的一维动力系统:

$$dx = f(x, p)dt + \gamma dW_t, \quad (1)$$

其中 x 是系统的状态, p 是系统的控制参数, $f(x, p)$ 是系统的确定性部分, 且 $f(x^*, p) = 0$ 对所有参数 p 成立(即 x^* 是确定性系统的稳定平衡点), dW_t 是具有零均值的和振幅为 γ 的白噪声项, 代表了外界对系统的随机驱动 (Dakos et al., 2012).

当一个小扰动 ϵ 推动系统离开平衡点 x^* 时, 系统回到该平衡点的速率(rate)大约等于系统(1)的线性化系统

$$d\epsilon = \lambda(x^*, p)dt + \gamma dW_t, \quad (2)$$

其中 $\lambda(x^*, p) = f'_x(x^*, p) < 0$ 是系统在 x^* 处的主特征值 (Horsthemke et al., 1984). 时滞为 1 的自相关系数 $\rho_\epsilon(1)$ 和方差 σ_ϵ^2 可近似表达为 (Gardiner, 1985):

$$\rho_\epsilon(1) = e^{\lambda(x^*, p)}, \quad (3)$$

$$\sigma_{\epsilon}^2 = -\frac{\gamma^2}{2\lambda(x^*, p)}. \quad (4)$$

显然, 当主特征值 $\lambda(x^*, p)$ 从负值趋向于 0 的时候, 即系统趋近于临界点, 自相关系数 $\rho_{\epsilon}(1)$ 和方差 σ_{ϵ}^2 都会显著增加. 这就是临界慢化现象的数学原理.

近年来, 生命科学研究迅速发展, 产生了大量的生物医学数据. 基于这些高通量数据, 国内外研究者们已经开始探究各种复杂疾病和生物过程中的临界状态. 在医学领域, 人体和其他生物系统的自发性和系统性状态变化得到了观察, 例如哮喘等突发性疾病的发作 (Litt et al., 2001; Venegas et al., 2005). 进一步研究表明: 癫痫发作前存在临界状态, 并将临界慢化现象引入作为预警癫痫发作的生物标志物 (Maturana et al., 2020); 由于负面情绪的持续生成和积累导致重度抑郁症心理状态的临界变化现象 (van de Leemput et al., 2014); 在应用激素类药物治疗前列腺癌时, 肿瘤细胞可能出现获得性抗药性的临界现象, 并提出了间歇性激素治疗的药物策略 (Tanaka et al., 2008; Hirata et al., 2010); 在高维流行病动力系统中存在临界现象, 并指出在麻疹等季节性传染病传播过程中, 通过临界慢化现象可能进行传染病暴发的监测 (Brett et al., 2020). 显然, 基于一维动力系统的临界慢化现象可以较好的对低维系统进行临界点识别与预警. 但现实中, 更多的系统是拥有大量变量的高维系统, 需要开发更合适的方法来处理这类系统的临界点识别.

2.2 动态网络生物标志物 (DNB) 方法

在不断涌现的生物医学高维数据驱动下, 基于动力系统分岔理论, 对应于“状态的临界慢化”现象, 近年来研究者们建立了高维复杂生物系统“变量的协同波动”的量化准则, 即动态网络生物标志物 (DNB, dynamic network biomarker) 方法 (Chen et al., 2012). 该方法是一种数据驱动的普适方法, 即使对于每个采样期只有少量的样本, 只需要每个样本有高通量数据或高维数据, DNB 方法就可以量化疾病恶性转变前的预警信号, 并因此可以识别疾病发生前的临界状态及重要因子和网络.

与传统的分子标志物相比, 动态网络生物标志物的提出具有重要意义. 传统的分子标志物主要依赖于静态的差异信息, 如基因或蛋白质的表达量差异来区分疾病状态与正常状态. 然而, 前疾病状态与正常状态之间的差异往往较小, 因此传统标志物在预警前疾病状态或临界状态方面可能失效. 动态网络生物标志物则不同, 它们建立在分子标志物的波动和关联的动态差异信息之上. 这种动态的差异信息可以更准确地识别前疾病状态或临界状态的到来, 为及时干预和治疗提供了新的机会.

从动力系统理论的角度看, 生物系统或复杂疾病的发展过程可以被描述为一个非线性动力系统或动态网络. 复杂疾病的发展过程沿着时间轴演化, 其恶性转化是系统经过一个临界状态进入疾病状态的现象. 外部因素影响疾病的发展被视为动力系统参数, 而参与疾病演变的分子浓度等则被视为系统中的状态变量. 因此, 考虑以下表示动态演化的高维离散时间动力系统:

$$\mathbf{Z}(k+1) = f(\mathbf{Z}(k); \mathbf{P}), \quad (5)$$

其中 $k \in \mathbb{Z}$, $\mathbf{Z}(k) = (z_1(k), z_2(k), \dots, z_n(k)) \in \mathbb{R}^n$ 是一个在时刻 k 的 n 维的状态向量, 代表系统中的 n 个生物分子的表达值; $\mathbf{P} = (p_1, \dots, p_s) \in \mathbb{R}^s$ 是一个参数向量, 代表了驱动高维生物分子系统缓慢变化的因素; $f: \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}^n$ 是该高维分子系统中的非线性光滑函数.

非线性系统在从一个状态到另一个状态之间发生分岔, 并且在分岔发生之前或之后, 系统的动力学通常变得非常复杂. 然而, 当系统在参数的驱动下接近临界点时, 就会进入一个特殊的阶段, 其动力学可以由一个低维空间来表示, 即中心流形. 基于分岔理论和中心流形理论, 可以证明系统在临近分岔点时的动力学具有特殊的特征, 这使得系统状态将要发生改变的信号能够被检测到. 这也是动态网络生物标志物 (DNB) 的理论基础 (Chen et al., 2012; Liu et al., 2013a, 2014b).

对于系统 (5), 如果 $\mathbf{Z}^* = (z_1^*, \dots, z_n^*)$ 是系统 (5) 的一个不动点, 则存在一个分岔参数值或临界参数值 \mathbf{P}_c , 使得一个 (或一对) Jacobin 矩阵 $\left. \frac{\partial f(\mathbf{Z}; \mathbf{P}_c)}{\partial \mathbf{Z}} \right|_{\mathbf{Z}=\mathbf{Z}^*}$ 的特征值的模为 1; 当参数 $\mathbf{P} \neq \mathbf{P}_c$ 时, Jacobin 矩阵

$\left. \frac{\partial f(\mathbf{Z}; \mathbf{P})}{\partial \mathbf{Z}} \right|_{\mathbf{Z}=\mathbf{Z}^*}$ 的特征值的模不恒等于 1. 这 3 个假设与其他横截条件意味着当参数 \mathbf{P} 达到临界值 \mathbf{P}_c 时, 系统在 \mathbf{Z}^* 处经历相变或余维一分岔. 假设系统具有一个稳定的不动点 \mathbf{Z}^* . 当系统(5)的参数 \mathbf{P} 趋向于 \mathbf{P}_c 时, 在不动点 \mathbf{Z}^* 附近, 系统所有的特征值的模都在 $(0, 1)$ 之内.

DNB 的理论和方法是基于系统(5)的线性化以及它在 \mathbf{Z}^* 附近的小噪声扰动的基础上建立的. 具体来说, 通过引入新的变量 $\mathbf{Y}(k) = (y_1(k), \dots, y_n(k))$ 和变换矩阵 S , 即

$$\mathbf{Y}(k) = S^{-1}(\mathbf{Z}(k) - \mathbf{Z}^*). \tag{6}$$

于是有线性化系统

$$\mathbf{Y}(k+1) = \Lambda(\mathbf{P})\mathbf{Y}(k) + \boldsymbol{\zeta}(k), \tag{7}$$

其中 $E(y_i(k)) = 0$. $\Lambda(\mathbf{P})$ 是系统在不动点附近线性项系数矩阵 $\left. \frac{\partial f(\mathbf{Z}; \mathbf{P})}{\partial \mathbf{Z}} \right|_{\mathbf{Z}=\mathbf{Z}^*}$ 的对角化矩阵, 其中 $\boldsymbol{\zeta}(k) = (\zeta_1(k), \dots, \zeta_n(k))$ 是均值为 0 的高斯小噪声. 记 $k_{ii} > 0$ 为 $\zeta_i (i = 1, 2, \dots, n)$ 的方差, $k_{ij} = \text{Cov}(\zeta_i, \zeta_j)$ 为任意两个小噪声之间的协方差, 对角化矩阵 $\Lambda(\mathbf{P}) = \text{diag}(\lambda_1(\mathbf{P}), \dots, \lambda_n(\mathbf{P}))$ 且其中的 $|\lambda_i|$ 在 0 和 1 之间. 对于主特征值或最大模的特征值来说, 根据矩阵的对角化过程有两种典型的情况, 即主特征值是实数(包括多个具有相同值的实特征值)和主特征值是一对共轭的复数. 当主特征值的模接近于 1 时, 系统会出现 3 类普适的分岔现象.

在主特征根是实数的前提下, 如果主特征值接近 1, 临界点是鞍结点(saddle-node)或 fold 分岔, 而如果主特征值接近 -1, 临界点是倍周期(period-doubling)分岔. 值得注意的是, 在 Λ 的特征值中, 可能存在多个实的主特征值(特征方程有重根的情况), 其推导类似于无重根的情况. 不失一般性, 记主特征值为 λ_1 . 当系统从稳态开始演化时, λ_1 是绝对值最接近 1 的特征值. 当参数 $\mathbf{P} \rightarrow \mathbf{P}_c$ 时, 主特征值 $|\lambda_1|$ 趋近于 1. 临界前的状态对应于 $|\lambda_i| < 1$ 的时期, 而临界状态对应于 $|\lambda_i| \rightarrow 1$ 的时期. 显然, \mathbf{Y} 中的第一个变量 y_1 对应于 λ_1 , 即 $(y_1, 0, \dots, 0)$ 是特征值 λ_1 的特征向量.

由于 Λ 是一个对角化矩阵, 根据式(7), 我们有协方差(Cov, covariance)

$$\text{Cov}(y_i, y_j) = \lambda_i \lambda_j \text{Cov}(y_i, y_j) + \kappa_{ij},$$

其中 k_{ij} 是两个小噪声之间的协方差. 于是有

$$\text{Cov}(y_i, y_j) = \frac{k_{ij}}{1 - \lambda_i \lambda_j}. \tag{8}$$

当 $i = j$ 时, 协方差成为方差(Var, variance), 于是有

$$\text{Var}(y_i) = \frac{k_{ii}}{1 - \lambda_i^2}, \tag{9}$$

其中 $k_{ii} > 0$ 是噪声 ζ_i 的方差. 当参数 \mathbf{P} 远离临界值 \mathbf{P}_c 时, 系统具有稳定的不动点, 特征值满足 $0 \leq |\lambda_i| < 1$, 当 $\mathbf{P} \rightarrow \mathbf{P}_c$ 或主特征值 $|\lambda_1| \rightarrow 1$ 时, $\text{Var}(y_1) \rightarrow +\infty$. 对于其他的特征值 $\lambda_i (i = 2, 3, \dots, n)$ 来说, 当参数 $\mathbf{P} \rightarrow \mathbf{P}_c$ 时, 方差 $\text{Var}(y_i)$ 趋近于一个有界的正常数.

当 $i \neq j$ 时, 皮尔逊相关系数具有如下的形式

$$\text{PCC}(y_i, y_j) = \frac{\text{Cov}(y_i y_j)}{\sqrt{\text{Var}(y_i) \text{Var}(y_j)}} = \frac{k_{ij}}{\sqrt{k_{ii} k_{jj}}} \frac{\sqrt{(1 - \lambda_i^2)(1 - \lambda_j^2)}}{1 - \lambda_i \lambda_j}.$$

回到原空间, 有 $z_i(k) = s_{i1}y_1(k) + \dots + s_{in}y_n(k) + z_i^*$, 方差为

$$\text{Var}(z_i) = s_{i1}^2 \frac{\kappa_{11}}{1 - \lambda_1^2} + \sum_{k=2}^n s_{ik}^2 \frac{\kappa_{kk}}{1 - \lambda_k^2} + \sum_{k, m=1, k \neq m}^n s_{ik} s_{im} \frac{\kappa_{km}}{1 - \lambda_k \lambda_m}.$$

同样地, 对于皮尔逊相关系数有

$$\begin{aligned} \text{PCC}(z_i, z_j) &= \frac{\text{Cov}(z_i, z_j)}{\sqrt{\text{Var}(z_i)\text{Var}(z_j)}} \\ &= \frac{s_{i1}s_{j1}\frac{\kappa_{11}}{1-\lambda_1^2} + \sum_{k=2}^n s_{ik}s_{jk}\frac{\kappa_{kk}}{1-\lambda_k^2} + \sum_{k,m=1,k \neq m}^n s_{ik}s_{jm}\frac{\kappa_{km}}{1-\lambda_k\lambda_m}}{\sqrt{\left(\frac{s_{i1}^2\kappa_{11}}{1-\lambda_1^2} + \sum_{k=2}^n \frac{s_{ik}^2\kappa_{kk}}{1-\lambda_k^2} + \sum_{k,m=1,k \neq m}^n \frac{s_{ik}s_{im}\kappa_{km}}{1-\lambda_k\lambda_m}\right)\left(\frac{s_{j1}^2\kappa_{11}}{1-\lambda_1^2} + \sum_{k=2}^n \frac{s_{jk}^2\kappa_{kk}}{1-\lambda_k^2} + \sum_{k,m=1,k \neq m}^n \frac{s_{jk}s_{jm}\kappa_{km}}{1-\lambda_k\lambda_m}\right)}} \end{aligned}$$

因此当参数 P 接近鞍结点分岔或倍周期分岔的临界值 P_c 时(即主特征值 $|\lambda_1| \rightarrow 1$ 时), 存在一个 DNB 变量群, 具有以下性质:

(i) 如果变量 z_i 和 z_j 都属于该 DNB, 即 $s_{i1} \neq 0$ 且 $s_{j1} \neq 0$, 则 $|\text{PCC}(z_i, z_j)| \rightarrow 1$, 并且 $\text{SD}(z_i) \rightarrow +\infty$ 和 $\text{SD}(z_j) \rightarrow +\infty$;

(ii) 如果变量 z_i 属于该 DNB, 但变量 z_j 不属于 DNB, 即 $s_{i1} \neq 0$ 且 $s_{j1} = 0$, 则 $\text{PCC}(z_i, z_j) \rightarrow 0$, 并且 $\text{SD}(z_i) \rightarrow +\infty$, 而 $\text{SD}(z_j)$ 趋向于一个有界正数;

(iii) 如果变量 z_i 和 z_j 都不属于该 DNB, 即 $s_{i1} = 0$ 且 $s_{j1} = 0$, 则 $\text{PCC}(z_i, z_j)$ 趋于一个常数, 并且 $\text{SD}(z_i)$ 和 $\text{SD}(z_j)$ 都趋向于有界正常数.

也就是说, 当系统接近临界点时, 会存在一个满足以下 3 个条件或性质的 DNB:

- 1) DNB 中每一对成员之间的相关性都变得很强, 如生物分子表达的皮尔逊相关系数(PCC, Pearson correlation coefficient)的绝对值迅速增长;
- 2) DNB 的每个成员表达的波动性都急剧变强, 如标准偏差(SD, standard deviation)急剧增加;
- 3) DNB 某个成员和非 DNB 的其他分子之间的相关性变得很弱, 如皮尔逊相关系数的绝对值迅速下降.

在许多情况下, 只需要前两个条件, 即“临界协同波动”的分子群的出现就意味着临界状态的到来. 特别需要强调的是, 状态的临界慢化是描述这样的临界状态的现象, 通常很难观察到系统的状态, 而在 DNB 理论中, 变量的临界协同波动是可以基于观测变量的信息来量化这种临界状态的. 显然, 动态网络生物标志物方法是一种无模型(model-free)的方法, 通过识别“临界协同波动”的分子群来检测状态临界转变的早期信号不依赖于特定的疾病模型, 即使在样本数量有限的情况下也能预警疾病的突然恶化. 基于描述分子表达波动和分子间相关性的动态变化的条件, 可以进一步构建一个复合指数来识别疾病临界状态的预警信号(Chen et al., 2012).

3 动态网络生物标志物(DNB)的几类改进算法

3.1 基于多样本数据的动态网络生物标志物方法

多样本数据可以提供更多样本间的比较和对比, 从而更全面地了解网络的动态变化, 还可以减少个体差异的影响, 提高标志物的稳定性和可靠性. 不仅如此, 多样本数据的动态网络生物标志物方法可以支持群体水平的分析, 从而更好地了解群体的整体特征和趋势. 通过对不同群体之间的比较, 可以发现群体间的差异和共性, 为群体水平的干预和管理提供支持. 例如基于时间差异网络的计算方法(Chen et al., 2017)通过收集不同时间点的样本数据来构建时间序列的差异网络, 还有基于动态网络生物标志物(DNM)的方法(Liu et al., 2015)通过分析时间序列数据中的多样本信息, 构建动态网络并降低噪声影响, 以预警复杂系统的关键转变, 见表 1.

具体来说, 基于时间差异网络的计算方法通过构建差异网络序列, 并使用隐马尔可夫模型(HMM)来训练和识别临界状态, 提出了不一致性得分(I-score)作为量化临界点的指标, 并通过模拟数据和疾病的基因表达数据集来验证方法的有效性; DNM 方法侧重于分析系统的动态特性和概率分布的转变, 通过“分布嵌入”策略将观察到的状态变量(带有大噪声)转换为概率分布变量(带有小噪声), 从而使得传统的基于 DNB 或 CSD 的方法在变换后的高维数据中变得有效.

表1 基于多样本数据的动态网络生物标志物方法

Table 1 Dynamic network biomarker method based on multi sample data

方法	输入数据类型	量化指标	模型	实验数据集
时域差异网络方法	基因表达数据的时间序列	不一致性得分 (I-score)	基于隐马尔可夫模型(HMM)	模拟数据和3种疾病的真实数据集
基于动态网络生物标志物(DNM)的方法	多维的时间序列数据	DNM的综合分数	无模型	3个真实数据集

3.2 基于单样本数据的动态网络生物标志物方法

随着精准医疗的发展,对个体化治疗方案的需求日益增长,并且在某些情况下,如罕见病或特定条件下的疾病研究,可能难以获得大量的样本数据,同时随着高通量测序技术的进步,单个样本可以产生大量的生物信息数据,因此基于单样本数据的DNB算法是很有必要的.基于单样本DNB的算法和模型有很多,在这里简要概述两种——利用网络信息的和利用单个病例样本与健康参考样本集之间的差异的DNB方法,见表2.

表2 基于单样本数据的动态网络生物标志物方法

Table 2 Dynamic network biomarker method based on single sample data

方法	输入数据类型	量化指标	模型	实验数据集
景观动态网络生物标志物(l-DNB)	单样本和参考样本的基因表达数据、基因相互作用信息	全局DNB得分(l-DNB)	无模型	4个真实数据集
单样本节点熵(SNE)	单样本和参考样本的基因表达数据、蛋白质-蛋白质相互作用(PPI)网络	局部SNE得分	单样本节点熵(SNE)模型	4个真实数据集
基于多模优化机制和个性化动态网络生物标志物理论的模型(MMPDNB)	个体患者的基因表达数据、基因突变数据、参考基因相互作用网络信息	最小化PDNB模块中基因的数量	MMPDNB模型	2个真实数据集
动态网络生物标志物单样本得分(DNB-S score)	单个病例样本的基因表达数据、一组控制样本的基因表达数据	DNB-S得分	无模型	3个真实数据集
单样本DNB(sDNB)	单个病例样本的基因表达数据、一组控制样本的基因表达数据	sDNB得分	无模型	2个真实数据集
单样本Kullback-Leibler散度(sKLD)	单个病例样本的基因表达数据、一组控制样本的基因表达数据	sKLD得分	无模型	数值模拟和6个真实数据集

3.2.1 利用网络信息探测临界状态特征 基于网络信息和单样本数据的DNB算法能够充分利用分子关联网络的拓扑信息,并且可以利用高通量数据,如基因表达谱数据,能够挖掘和分析丰富的动态和高维信息,为疾病研究提供了新的视角.例如,景观动态网络生物标志物(l-DNB)(Liu et al., 2019b)和基于单样本节点熵(SNE)的个体样本模型(Han et al., 2020)都利用了全局网络信息,通过综合局部网络的结构改变信息来计算全局网络得分,分数的显著变化指示着系统接近临界状态,从而为复杂疾病的恶化提供早期预警信号.

具体来说, l-DNB方法使用给定样本的基因表达数据,构建一个单样本网络(SSN),对于数据集中的每个基因,计算其局部模块的局部DNB得分.通过汇总所有基因的局部DNB得分,形成一个景观(landscape),从中可以识别出得分最高的 k 个基因作为潜在的DNB成员; SNE模型首先通过PPI网络信息构建全局模板网络,再将全局模板网络分割成多个局部网络,并对每个局部网络中的基因表达数据基于参考样本集中的表达值进行高斯分布拟合,接着根据单个样本的基因表达数据为每个局部网络构建一个状态向量,最后用状态向量计算每个局部网络的局部单样本节点熵(local SNE),并将所有局部网络中局部

SNE得分最高的前5%的总和作为全局单样本节点熵(global SNE).

MMPDNB模型(Liang et al., 2022)在构建过程中也大量利用了网络信息, 通过结合个性化动态网络生物标志物理论和多模态优化机制, 为癌症的精准医疗提供了一个新的视角. 具体来说, 研究者使用患者配对样本的基因表达数据、基因突变数据和已知的网络拓扑信息来构建个性化基因相互作用网络(PGIN), 接着定义两个优化目标函数: 最小化个性化动态网络生物标志物(PDNB)模块中基因的数量(f_1), 以及最大化模块的早期预警信号得分(f_2), 最后使用多模进化算法(EA)搜索多个PDNB模块. MMPDNB模型在疾病预警和治疗药物靶点的发现方面展现出显著的优势.

通过生物医学实验, 1-DNB方法预警了流感病毒感染的严重症状, 并在多个癌症数据集(如肺腺癌、肾细胞癌和甲状腺癌等)中识别了疾病恶化前的临界状态; SNE模型通过模拟数据和来自TCGA的4个不同癌症数据集识别了与癌症进展相关的关键基因和新的预后生物标志物, 还揭示了DNA损伤修复途径中非特异性基因网络的转换可能是推动癌症进展的主要驱动力, 为癌症治疗干预提供了新的靶基因和潜在的药物目标; MMPDNB模型通过在模拟数据和4种不同癌症的TCGA数据集上的实验, 检测了癌症发展过程中的关键阶段, 并识别了位于重要癌症相关信号通路上游的动态网络生物标志物, 还发现了一些新的预后生物标志物.

3.2.2 利用单个病例样本与健康参考样本集之间的差异探测临界信号 动态网络生物标志物单样本评分(DNB-S score)方法(Liu et al. 2014b)、基于单样本的动态网络生物标志物(sDNB)(Liu et al., 2017)、基于单样本的Kullback-Leibler散度(sKLD)度量(Zhong et al., 2020)等方法, 都是把大量正常状态的样本(例如健康人样本)作为参考/对照样本集合(reference/control sample set), 通过比较单个待测样本(single case sample)的分子表达模式与参考样本集的整体表达模式之间的差异, 从而评估待测样本是否处于疾病的临界状态.

不同的是, DNB-S score方法是基于DNB理论和DNB与非DNB分子表达之间的差异分布的, 通过Kullback-Leibler散度(K-L散度)来量化单个样本的DNB-S得分, 从而可靠地检测单个样本中的前病状态; sDNB方法基于差异关联信息而非差异表达信息, 使用控制样本集为参考, 计算单样本表达偏差(sED), 单样本皮尔逊相关系数(sPCC), 并通过层次聚类算法分解网络为多个模块, 最终选择具有最高分数的模块作为潜在的sDNB模块; sKLD方法是通过参考样本集对每个基因的表达数据拟合出一个高斯分布, 即参考分布P, 然后将单个待测样本加入参考样本集合中, 针对每个基因构建一个受扰动的分布Q, 最后通过计算sKLD分数来评估参考分布P和受扰动分布Q之间的差异, 从而为复杂疾病的恶化提供早期预警信号. DNB-S score、sDNB和sKLD等方法都是无模型的方法, 不依赖于大规模样本学习, 避免了过拟合问题, 使得它们在样本数量有限的情况下依然能够有效地工作.

3.3 结合信息论的动态网络生物标志物方法

信息论提供了量化信息变化的工具, 如信息增益、熵和互信息等, 见表3. 通过分析信息增益, 我们可以确定哪些生物标志物对于疾病预警具有更高的重要性. 在特定网络信息增益(SNIG)方法(Peng et al., 2023)中, 信息增益这一概念被用来量化和分析微生物群落网络的结构变化, SNIG方法通过监测网络信息增益的变化, 可以识别这些临界状态的到来. 具体来说, SNIG方法通过构建微生物关联网络(SSAN), 提取局部网络, 并计算中心物种及其邻近物种的节点概率和条件概率, 进而计算网络流熵和累积信息增益(CIG), 最终通过计算全局网络中的平均扰动信息增益(序列数据和代谢物数据)来识别结肠癌发展过

表3 结合信息论的动态网络生物标志物方法

Table 3 Dynamic network biomarker method based on information theory

方法	输入数据类型	量化指标	实验数据集
特定网络信息增益(SNIG)	序列数据和代谢物数据	序列数据和代谢物数据	3个真实数据集
互信息加权熵(MIWE)	批量和单细胞表达数据	MIWET值	1组数值模拟数据集和4个真实的生物数据集

程中的关键状态, 为疾病的早期诊断和干预提供重要依据。

互信息能够量化基因或变量之间的依赖关系, 通过计算生物标志物之间的互信息, 有助于揭示生物标志物之间的复杂相互作用, 从而提高对疾病状态转变的预警精度。例如, 互信息加权熵(MIWE)的方法(Xie et al., 2024)是通过量化分子动态差异, 能够在系统接近临界状态时提供更精确的预警信号。熵作为一种度量不确定性的指标, MIWE方法通过加权熵的方式, 可以更准确地捕捉和表达数据中的关键特征和它们的相互作用。具体来说, MIWE方法首先将基因表达值转换为概率分布, 并在不同时间点构建基于互信息的网络, 然后从全局网络中提取局部网络, 其中包含中心基因及其一阶邻居, 接着计算局部网络中邻居基因的微分熵, 并根据每个局部网络的权重计算全局网络的互信息加权熵(MIWET)。在系统达到临界状态时, MIWET值会显著增加, 从而为即将发生的状态变化提供预警信号。

3.4 基于网络熵的动态网络生物标志物方法

局部网络熵关注网络中特定区域或子集的行为, 允许对网络的局部结构和功能进行更细致的分析, 通过分析局部网络熵, 可以识别网络中的功能模块或社区结构, 这些模块可能在特定的生物学过程中起关键作用, 见表4。

表4 基于网络熵的动态网络生物标志物方法
Table 4 Dynamic network biomarker method based on network entropy

方法	输入数据类型	量化指标	实验数据集	仅适用于单样本
基于状态转移的局部网络熵(SNE)	基因表达数据和蛋白质相互作用网络	局部网络熵	2个真实数据集	否
局部网络熵(LNE)	单个样本的基因表达数据	LNE得分	10个真实数据集	是
时间差异网络流熵(TNFE)	高维组学数据的时间序列	网络流熵(NFE)	多种真实疾病数据集	否
定向网络排名评分(DNRS)	基因表达数据和蛋白质相互作用网络的时间序列	DNRS得分	模拟数据集和6个真实数据集	否
单样本景观熵(SLE)	单样本的生物分子表达数据和一组参考样本数据	全局网络熵	模拟数据集和4个真实数据集	是
景观网络熵(LNE)	地理区域网络数据和每日新病例的时间序列数据	全局网络熵	6个真实数据集	否

例如研究者们提出的基于状态转移的局部网络熵(SNE)方法(Liu et al., 2012), 基于信息流动的局部网络熵(LNE)方法(Liu et al., 2022), 时间差异网络流熵(TNFE)(Gao et al., 2022)等, 都是通过量化局部网络的结构变化来捕捉系统在临界状态时网络特征的方法。不同的是, SNE方法是利用蛋白质-蛋白质相互作用网络的局部结构信息, 通过计算局部网络中状态转移的熵来量化网络的动态变化。通过在基因表达数据集上的应用, SNE识别了肝癌和急性肺损伤病程中的临界状态, 并识别出了与临界状态密切相关的生物分子网络。LNE方法则将基因表达映射到蛋白质-蛋白质相互作用网络并计算每个基因的局部网络熵, 再通过分析单个样本的基因表达数据相对于健康参考样本集的局部网络熵的变化来识别疾病进展中的临界状态。通过分析TCGA数据库的十种癌症数据集, LNE识别了癌症进展的临界状态, 并且发现了与患者预后相关的新型生物标志物O-LNE和P-LNE, 同时揭示了这些标志物在癌症生物学过程中的作用和网络动态变化。而TNFE是一个基于数据驱动的无模型的计算方法, 它通过比较局部网络流熵(NFE)和条件网络流熵(条件NFE)的差异来计算时间差异网络流熵(即TNFET得分), 从而量化分子网络的波动, 从而进一步分析疾病恶化风险和识别药物靶标。

有向网络等级评分(DNRS)也是一个无模型的计算方法(Zhong et al., 2022), 它通过分析时间序列基因表达数据和蛋白质相互作用网络来构建时间特定的有向网络, 并利用个性化PageRank方法计算每个基

因节点的局部 DNRS 得分, 通过比较不同时间点的得分并结合统计检验来识别显著变化的临界点. DNRS 方法应用于模拟数据集和真实生物数据集, 验证了该方法在检测生物系统中临界转变点的能力, 并通过对关键基因的功能和通路富集分析, 揭示了这些临界点背后的生物学机制, 展现了 DNRS 在预测和理解复杂生物过程动态变化中的潜力.

一般直接计算全局网络的复杂度较高, 大都是把网络分片化, 先计算局部网络特征, 再综合起来计算全局的一个指标. 在疾病预测和生物标志物发现中, 全局网络熵可以用于识别系统级的临界转变, 而局部网络熵可以用于识别这些转变中起关键作用的特定生物分子或通路. 例如单样本景观熵 (SLE) (Liu et al., 2020b) 和景观网络熵 (LNE) (Liu et al., 2021) 的计算方法就都同时结合了局部网络熵和全局网络熵来分析. SLE 和 LNE 方法都是先基于输入数据来构建局部网络并计算局部网络熵, 然后进行加权求和得到全局网络熵, 最后根据全局网络熵的变化来识别疾病进展中的临界转变点. 与前述方法不同之处在于, SLE 和 LNE 方法加权的方式不同, SLE 对每个局部网络计算节点的表达水平之间的相关性, 并基于这些相关性计算局部熵值, 再将单个样本的局部熵值与参考样本集的局部熵值进行比较计算差异熵值, 最后将所有局部网络的差异熵值进行加权求和, 从而得到全局网络熵; LNE 则对于每个地区节点, 根据与其相邻地区的病例数据来计算局部网络的熵值, 然后计算局部网络中节点的病例数据的标准差, 最后将所有局部网络的熵值和标准差信息进行加权求和, 从而得到全局网络熵.

除此之外, SLE 和 LNE 都是无模型的计算方法, 不依赖于特定的动力学模型或参数化的形式来分析系统的行为, 可以在没有复杂模型构建的情况下, 从数据中提取有价值的信息, 并应用于多种生物学和医学研究场景. 例如: 研究者通过多个数值实验验证了 SLE 方法的准确性, 展示 SLE 分数在系统接近临界点时的急剧增加, 除此之外还在流感病毒感染、肺腺癌、急性肺损伤和前列腺癌 4 个真实数据集上应用 SLE 方法, 识别了疾病进展的临界点和早期预警信号; LNE 方法检测到了中国湖北省、日本关东地区、西欧、意大利半岛、韩国和美国 17 个州等地区 COVID-19 爆发的早期预警信号. 除此之外, 研究者还基于 LNE 方法开发了一个网络工具, 用于实时监测包括 COVID-19 和流感在内的传染病爆发. 该工具提供了全球和局部 LNE 指数, 以便于实时监测和控制传染病的传播.

4 动态网络生物标志物的几类应用场景

4.1 识别与预警代谢类疾病的临界状态

代谢性疾病是一类涉及身体代谢过程异常的疾病, 这些异常可能影响能量的生产、转换和利用, 以及身体对营养物质的处理. 其中, 糖尿病是一种以高血糖为特征的慢性代谢疾病, 主要原因是胰岛素分泌不足或者身体对胰岛素的反应减弱, 导致血糖水平无法得到有效控制, 因此糖尿病及其并发症的早期诊断对于预防和管理至关重要. 动态网络生物标志物 (DNB) 方法能够帮助监测糖尿病的进展和治疗效果 (见表 5).

例如利用非肥胖糖尿病 (NOD) 小鼠作为 1 型糖尿病 (T1D) 的模型, 结合 DNB 的理论, 从基因表达的变化中识别出了两个关键的生物分子网络 (Liu et al., 2013b). 通过应用 DNB 理论和方法 (Li et al., 2014), 对 2 型糖尿病 (T2D) 的多组织基因表达数据进行了分析, 揭示了在 T2D 发展过程中, 肝脏、脂肪和肌肉组织存在着与胰岛素抵抗和严重炎症反应有关的两个不同的关键状态.

此外, 通过分析糖尿病相关的生物标志物网络, 研究人员可能发现新的治疗靶点, 有助于开发新的

表 5 识别与预警代谢类疾病的临界状态

Table 5 Identify and warn the critical state of metabolic diseases

疾病	数据	结果
1 型糖尿病 (T1D)	非肥胖糖尿病 (NOD) 小鼠的胰腺淋巴结基因表达谱	检测 1 型糖尿病 (T1D) 的早期预警信号, 并预测疾病在急剧恶化前的临界点
2 型糖尿病 (T2D)	2 型糖尿病 (T2D) 大鼠模型的多组织、多阶段基因表达谱分析	识别了在 2 型糖尿病发展和进展过程中肝脏、脂肪和肌肉组织特异性的早期预警信号, 并揭示了这些信号在疾病早期阶段的分子机制

药物和治疗方法, 以更有效地控制糖尿病及其并发症. 例如, 研究者们识别出的两个 DNB 分别预示着胰岛周围炎和高血糖的发生, 这两个阶段是 T1D 发病过程中的关键转折点(Liu et al., 2013b). 这些发现不仅为 T1D 的早期诊断提供了新的生物标志物, 而且增进了我们对 T1D 发病机制的理解, 为未来的预防和治疗策略提供了重要的科学依据.

4.2 识别与预警肿瘤疾病的临界状态

DNB 理论关注于疾病发生发展状态的动态变化, 特别是关键的临界点, 通过分析与肿瘤相关的动态网络生物标志物的变化, DNB 有助于在肿瘤发展的早期阶段捕捉到关键的生物分子变化, 从而为患者提供更及时的治疗, 见表 6.

表 6 识别与预警肿瘤疾病的临界状态
Table 6 Identify and warn the critical state of tumor diseases

疾病	数据	结果
肝细胞癌(HCC)	木鸭肝炎病毒/c-myc 转基因小鼠的时间序列蛋白质组数据	揭示了慢性炎症向 HCC 转变的关键时期和分子机制(Li et al., 2017)
肝细胞癌(HCC)	HCCLM3-RFP 小鼠模型的基因表达数据	预测 HCC 的转移, 揭示新的治疗靶点, 作为预后生物标志物(Yang et al., 2018)
乳腺癌	乳腺癌 MCF-7 细胞的基因表达数据	预测乳腺癌细胞分化前的关键状态, 揭示新的治疗靶点(Chen et al., 2015)
乳腺癌	经过他莫昔芬治疗的雌激素受体阳性 MCF-7 乳腺癌细胞的时间序列 mRNA 数据	识别乳腺癌细胞内分泌抵抗过程中的关键转变点, 作为预后生物标志物(Liu et al., 2019a)
结直肠癌(CRC)	28 位结直肠癌 (CRC) 患者的癌组织和相邻正常组织表达数据	识别结直肠癌发展中的关键转变点和早期预警信号, 发现了 12 个潜在的药物靶点(Liu et al., 2020a)
结直肠癌(CRC)	两个结直肠癌 (CRC) 数据集	验证了该方法在检测疾病关键状态和识别 DNB 方面的有效性(Chen et al., 2015)
结直肠癌(CRC)	结直肠癌单细胞 RNA 测序 (scRNA-seq) 数据	确定 FOS/JUN 为结直肠癌(CRC)中预恶化上皮细胞亚群的生物标志物(Huang et al., 2023)
肾透明细胞癌(KIRC)	KIRC 的 RNA 测序数据, 包括肿瘤和肿瘤邻近样本的基因表达数据及相应的临床信息	识别了在 KIRC 中促进或抑制这些转变过程的关键基因(Liu et al., 2024)
多种肿瘤	食管癌、结肠腺癌、肾透明细胞癌和肺腺癌等 4 种肿瘤的 RNA 测序数据	基于单样本马尔可夫流熵方法, 在肿瘤数据集中验证了方法有效性, 同时还识别了两种新的预后生物标志物(Liu et al., 2023)
肿瘤的免疫检查点	肿瘤体积变化、治疗前后的基因表达模式、免疫细胞浸润情况等数据	生物标志物可以在治疗开始后不久用于区分响应者和非响应者, 并有助于识别新的治疗靶点(Lesterhuis et al., 2017)

例如, 通过分析木鸭肝炎病毒/c-myc 转基因小鼠的时间序列蛋白质组数据, 利用 DNB 模型揭示了慢性炎症向肝细胞癌(HCC)转变的关键时期和分子机制(Li et al., 2017). 通过分析 HCCLM3-RFP 小鼠模型的基因表达数据, 使用 DNB 方法发现了能够预测 HCC 的转移关键的生物标志物——CALML3(Yang et al., 2018). 通过分析 MCF-7 细胞系的基因表达数据, 使用 DNB 方法识别了乳腺癌细胞分化前的关键状态(Chen et al., 2015). 通过分析经过他莫昔芬治疗的雌激素受体阳性 MCF-7 乳腺癌细胞的时间序列 mRNA 数据, 应用 DNB 方法识别了乳腺癌细胞产生药物抵抗过程中的关键转变点(Liu et al., 2019a). 利用 28 位结直肠癌(CRC)患者的癌组织和相邻正常组织表达数据, 通过构建个体化蛋白质相互作用网络和应用动态网络分析, 识别了结直肠癌发展中的关键转变点和早期预警信号(Liu et al., 2020a). 通过构建基于肠道微生物组数据的特定网络信息增益(SNIG)方法, 然后通过数值模拟和两个结直肠癌数据集的应用, 验证了

SNIG 方法在检测疾病关键状态和识别 DNBs 方面的有效性等(Peng et al., 2023).

DNB 还可以揭示新的治疗靶点, 为开发新的治疗方法提供理论基础. 例如, 使用 DNB 方法发现了 CALML3 可作为预测和治疗 HCC 的潜在靶点(Yang et al., 2018). 研究者们发现, 识别出的 DNB 基因与癌症相关的细胞死亡调控和信号通路有关, 可能可以作为乳腺癌的早期预警和治疗靶点, 为乳腺癌的早期诊断提供了新的策略(Chen et al., 2015). 使用 DNB 方法可以识别上皮细胞恶化的关键阶段, 并确定 FOS/JUN 为结直肠癌(CRC)中预恶化上皮细胞亚群的生物标志物(Huang et al., 2023). 采用数据驱动的能量景观算法可以分析肾透明细胞癌(KIRC)的演变过程, 并在癌症发生过程中识别到关键基因 KRT4 和 MMP3, 在癌症恶化过程中识别到关键基因 CALCA 和 NR0B2(Liu et al., 2024). 这说明 DNB 框架下的各种分析方法对于发现新的生物标志物、药物靶点和癌症治疗策略具有重要意义.

DNB 不仅可以用于疾病的早期诊断, 还可以作为预后生物标志物, 帮助医生评估治疗效果和预警疾病进展, 这对于调整治疗方案和改善患者预后具有重要意义. 例如, 通过临床数据的生存分析, 验证了 DNB 基因与乳腺癌患者预后不良之间的关联(Liu et al., 2019a), 这一发现有助于在乳腺癌治疗中及时调整治疗方案, 尤其是在患者可能对特定药物产生抵抗的情况下. 单样本马尔可夫流熵(sMFE)方法(Liu et al., 2023)利用推断的直接相互作用网络和 DNB 概念来表征单个样本的扰动, 展示了其检测 4 种癌症中可能存在的临界状态的能力, 并提供了两种新的预后生物标志物, O-sMFE 和 P-sMFE, 以促进患者的个性化预后.

DNB 还可以应用在免疫检查点阻断治疗中, 通过识别那些在治疗前后表现出显著变化的基因、蛋白质或其他生物分子, 进而更准确地预测哪些患者可能对治疗产生响应, 从而实现个性化医疗和优化治疗方案(Lesterhuis et al., 2017). 研究者讨论了癌症免疫检查点阻断治疗中的生物标志物, 强调了开发能够预测治疗反应的动态生物标志物的重要性. 临床应用显示, 尽管免疫检查点阻断治疗在某些癌症患者中取得了成功, 但大多数患者并未响应, 因此迫切需要可靠的预测性生物标志物(Lesterhuis et al., 2017).

4.3 应用于多种生物过程的分析

最近, 多个研究团队还将 DNB 应用于许多生物过程的分析, 见表 7. 例如基于 SH-SY5Y 的细胞模型, 使用 DNB 方法确定了 α -突触核蛋白(α -Syn)在帕金森疾病的发展中开始病理性聚集的关键时间点(Zhong et al., 2023), 揭示了 MAPKAPK2 有潜力成为与 α -Syn 病理性聚集相关疾病的早期诊断生物标志物. 通过动态网络生物标志物(DNBs)来识别和预测偏头痛发作的临界点(Dahlem et al., 2015), 研究者提出偏头痛的先兆症状可能是大脑达到临界点的早期警告信号, 而非直接的触发因素, 为非侵入性成像技术提供了新的研究方向. 在细胞分化的早期阶段, 基因表达的细胞间变异性显著增加, 通过应用 DNB 方法揭示了一组基因, 它们在分化过程中的关键转变点表现出动态变化, 同时研究者还强调了在单细胞水平上理解细胞分化过程的重要性, 并指出分化可能是一个由底层分子网络动态行为所驱动的过程(Richard et al., 2016). 通过 DNB 方法也可以分析烟草植物感染烟草蚀纹病毒的转录组数据, 以识别从轻度到严重疾病的早期预警信号(Tarazona et al., 2019). 通过拉曼光谱技术获取 T 细胞在不同时间点的光谱数据, 并运用 DNB 方法识别出了 T 细胞激活的临界状态(Haruki et al., 2022). 基于 DNB 理论构建的复合指数, 可以用于捕捉网络中节点群体在临近临界转变点(即癫痫发作起始)时的动态特性, 从而分析和识别在癫痫发作过程中首先表现出动态变化的脑区子网络(Runfola et al., 2023). 在细胞命运决定的背景下, DNB 可以被用来监测和分析细胞状态的转变, 特别是在多能祖细胞向特定细胞系分化的过程中. 通过单细胞分辨率基因表达分析揭示了血细胞祖细胞在分化为红细胞或白细胞系时, 会先经历一个高维吸引子状态的不稳定化阶段, 随后通过临界状态转变实现细胞命运的决定(Mojtahedi et al., 2016). 该研究提出了一种新的定量指标来预警这种高维状态空间中的临界转变, 并观察到“叛逆细胞(rebellious cells)”的出现, 证实了祖细胞吸引子状态的不稳定化模型. 这一发现为理解细胞如何在动态系统中做出命运选择提供了新的理论框架, 并可能有助于预警和控制细胞状态的转变. 通过这些应用, DNB 理论及方法为研究各类生物过程提供了一个系统的视角, 有助于揭示在复杂生物系统中发生的动态变化和关键调控机制. DNB 方法特别适用于那些涉及多个相互作用组分和复杂反馈回路的生物过程.

表7 应用于多种生物过程的分析
Table 7 Applied to the analysis of various biological processes

疾病/生物过程	数据	结果
帕金森疾病中 α -突触核蛋白的病理聚集过程	人类神经母细胞瘤细胞系(SH-SY5Y)的RNA测序数据, 健康人和帕金森病患者的大脑组织和外周血样本的表达谱数据	鉴定了 α -突触核蛋白病理聚集前的关键时间点, 表明其可能作为 α -突触核蛋白病理聚集相关疾病的早期诊断生物标志物(Zhong et al. ,2023)
偏头痛	数值模拟数据	数学模型能够帮助识别偏头痛发作前在特定神经和自主神经系统子网络中出现的大范围和相关联的波动, 这些波动可以作为偏头痛的动态网络生物标志物(Dahlem et al. ,2015)
细胞分化过程	鸡红细胞前体细胞的基因表达数据	在细胞分化的临界点之前, 细胞间的分子变异性显著增加, 且这种变异性是由基因表达的DNB方法识别的, 表明分化过程是一个动态的、非线性的分子事件序列(Mojtahedi et al. ,2016; Richard et al. ,2016)
烟草植物感染烟草蚀纹病毒	烟草植物感染烟草蚀纹病毒的转录组数据	识别了从轻度到严重烟草蚀纹病转变过程中的早期预警信号(Tarazona et al. ,2019)
拉曼光谱分析	T细胞激活过程中的拉曼光谱数据	检测到T细胞激活过程中的临界状态, 并识别出与过渡状态相关的DNB拉曼位移(Haruki et al. ,2022)
癫痫	模拟和临床立体脑电图(SEEG)数据	地识别了局灶性癫痫的致病区(EZ)(Runfola et al. ,2023)

5 结 语

传统的生物标志物在生物和医学领域中扮演着关键角色, 它们能够揭示生物体的特征并用于疾病的诊断, 这些标志物通常用于区分不同的健康状态或疾病状态. 然而, 这些传统的生物标志物, 无论是分子层面的还是网络层面的, 往往缺乏在疾病发展早期识别关键转变点的能力. 换句话说, 它们更擅长于标识已经发生的状态变化, 而不是预警即将到来的状态转变. 动态网络生物标志物(DNB)代表了一种创新的概念和方法论, 它专注于检测疾病发展过程中的状态转换前的临界状态. 后续发展的基于单样本数据的DNB改进方法允许对单个样本的网络进行分析, 这意味着可以为每个患者提供定制化的疾病预警和治疗策略. 通过将信息论中的概念与DNB结合, 可以增强对疾病早期状态的识别能力, 提高疾病预警的准确性和可靠性. 网络熵是一种衡量网络复杂性和不确定性的指标, 结合网络熵的方法可以探测到疾病早期的微小变化, 这些变化可能是传统DNB难以捕捉的. DNB方法还可以通过挖掘组学数据中的动态信息精确地量化疾病即将发生的临界点. 最近还应用于多种生物过程的分析、癌症免疫检查点阻断和识别细胞命运决定的临界点. 总之, DNB的核心优势在于其能够捕捉到复杂生物过程或具有突变特性的疾病在关键转变发生之前的微妙变化, 因此基于DNB的方法可以提高对疾病临界点的检测能力, 提供更早期的预警信号, 改善个体化医疗, 处理小样本数据, 并揭示复杂网络的动态特性, 从而在疾病预警和治疗中发挥重要作用.

参考文献:

BARGAJE R, TRACHANA K, SHELTON M N, et al, 2017. Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells[J]. Proc Natl Acad Sci USA, 114(9): 2271-2276.

- BRETT T, AJELLI M, LIU Q H, et al, 2020. Detecting critical slowing down in high-dimensional epidemiological systems[J]. *PLoS Comput Biol*, 16(3): e1007679.
- CHEN E I, YATES J R, 2006. Maspin and tumor metastasis[J]. *IUBMB Life*, 58(1): 25–29.
- CHEN L, LIU R, LIU Z P, et al, 2012. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers[J]. *Sci Rep*, 2: 342.
- CHEN L, WANG R, LI C, et al, 2010. Modeling biomolecular networks in cells: structures and dynamics[M]. London: Springer-Verlag.
- CHEN P, LI Y, LIU X, et al, 2017. Detecting the tipping points in a three-state model of complex diseases by temporal differential networks[J]. *J Transl Med*, 15(1): 217.
- CHEN P, LIU R, CHEN L N, et al, 2015. Identifying critical differentiation state of MCF-7 cells for breast cancer by dynamical network biomarkers[J]. *Front Genet*, 6: 252.
- DAHLEM M A, KURTHS J, FERRARI M D, et al, 2015. Understanding migraine using dynamic network biomarkers[J]. *Cephalalgia*, 35(7): 627–630.
- DAKOS V, van NES E H, D'ODORICO P, et al, 2012. Robustness of variance and autocorrelation as indicators of critical slowing down[J]. *Ecology*, 93(2): 264–271.
- DRAKE J M, GRIFFEN B D, 2010. Early warning signals of extinction in deteriorating environments[J]. *Nature*, 467(7314): 456–459.
- GAO R, YAN J, LI P, et al, 2022. Detecting the critical states during disease development based on temporal network flow entropy [J]. *Brief Bioinform*, 23(5): bbac164.
- GARDINER C W, 1985. Handbook of stochastic methods for physics, chemistry and the natural sciences[M]. Berlin: Springer-Verlag.
- HAN C, ZHONG J, HU J, et al, 2020. Single-sample node entropy for molecular transition in pre-deterioration stage of cancer[J]. *Front Bioeng Biotechnol*, 8: 809.
- HARUKI T, YONEZAWA S, KOIZUMI K, et al, 2022. Application of the dynamical network biomarker theory to Raman spectra [J]. *Biomolecules*, 12(12): 1730.
- HIRATA Y, BRUCHOVSKY N, AIHARA K, 2010. Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer[J]. *J Theor Biol*, 264(2): 517–527.
- HORSTHEMKE W, LEFEVER R, 1984. Noise-induced transitions: Theory and applications in physics, chemistry, and biology [M]. Berlin: Springer-Verlag.
- HUANG X, HAN C, ZHONG J, et al, 2023. Low expression of the dynamic network markers FOS/JUN in pre-deteriorated epithelial cells is associated with the progression of colorectal adenoma to carcinoma[J]. *J Transl Med*, 21(1): 45.
- KAMBHU J, WEIDMAN S, KRISHNAN N, 2007. New directions for understanding systemic risk: A report on a conference cosponsored by the Federal Reserve Bank of New York and the National Academy of Sciences [M]. Washington, DC: National Academies Press.
- KOIZUMI K, OKU M, HAYASHI S, et al, 2020. Suppression of dynamical network biomarker signals at the predisease state (Mibyō) before metabolic syndrome in mice by a traditional Japanese medicine (kampo formula) bofutsushosan[J]. *Evid Based Complement Altern Med*, 2020: 9129134.
- LENTON T M, 2011. Early warning of climate tipping points[J]. *Nat Clim Change*, 1: 201–209.
- LENTON T M, HELD H, KRIEGLER E, et al, 2008. Tipping elements in the Earth's climate system[J]. *Proc Natl Acad Sci*, 105(6): 1786–1793.
- LESTERHUIS W J, BOSCO A, MILLWARD M J, et al, 2017. Dynamic versus static biomarkers in cancer immune checkpoint blockade: Unravelling complexity[J]. *Nat Rev Drug Discov*, 16(4): 264–272.
- LI M, ZENG T, LIU R, et al, 2014. Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: Study of type 2 diabetes by cross-tissue analysis[J]. *Brief Bioinform*, 15(2): 229–243.
- LI M, LI C, LIU W X, et al, 2017. Dysfunction of PLA2G6 and CYP2C44-associated network signals imminent carcinogenesis from chronic inflammation to hepatocellular carcinoma[J]. *J Mol Cell Biol*, 9(6): 489–503.
- LIANG J, LI Z W, YUE C T, et al, 2022. Multi-modal optimization to identify personalized biomarkers for disease prediction of

- individual patients with cancer[J]. *Brief Bioinform*, 23(5): bbac254.
- LITT B, ESTELLER R, ECHAUZ J, et al, 2001. Epileptic seizures may begin hours in advance of clinical onset: A report of five patients[J]. *Neuron*, 30(1): 51–64.
- LIU J, DING D, ZHONG J, et al, 2022. Identifying the critical states and dynamic network biomarkers of cancers based on network entropy[J]. *J Transl Med*, 20(1): 254.
- LIU J, LI C, 2024. Data-driven energy landscape reveals critical genes in cancer progression[J]. *NPJ Syst Biol Appl*, 10(1): 27.
- LIU J, TAO Y, LAN R, et al, 2023. Identifying the critical state of cancers by single-sample Markov flow entropy[J]. *PeerJ*, 11: e15695.
- LIU L, SHAO Z, LV J, et al, 2020a. Identification of early warning signals at the critical transition point of colorectal cancer based on dynamic network analysis[J]. *Front Bioeng Biotechnol*, 8: 530.
- LIU R, AIHARA K, CHEN L, 2013a. Dynamical network biomarkers for identifying critical transitions and their driving networks of biologic processes[J]. *Quant Biol*, 1(2): 105–114.
- LIU R, CHEN P, AIHARA K, et al, 2015. Identifying early-warning signals of critical transitions with strong noise by dynamical network markers[J]. *Sci Rep*, 5: 17501.
- LIU R, CHEN P, CHEN L, 2020b. Single-sample landscape entropy reveals the imminent phase transition during disease progression[J]. *Bioinformatics*, 36(5): 1522–1532.
- LIU R, LI M, LIU Z P, et al, 2012. Identifying critical transitions and their leading biomolecular networks in complex diseases [J]. *Sci Rep*, 2: 813.
- LIU R, WANG J, UKAI M, et al, 2019a. Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers[J]. *J Mol Cell Biol*, 11(8): 649–664.
- LIU R, WANG X D, AIHARA K, et al, 2014a. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers[J]. *Med Res Rev*, 34(3): 455–478.
- LIU R, YU X, LIU X, et al, 2014b. Identifying critical transitions of complex diseases based on a single sample [J]. *Bioinformatics*, 30(11): 1579–1586.
- LIU R, ZHONG J, HONG R, et al, 2021. Predicting local COVID-19 outbreaks and infectious disease epidemics based on landscape network entropy[J]. *Sci Bull*, 66(22): 2265–2270.
- LIU X, CHANG X, LENG S, et al, 2019b. Detection for disease tipping points by landscape dynamic network biomarkers[J]. *Natl Sci Rev*, 6(4): 775–785.
- LIU X, CHANG X, LIU R, et al, 2017. Quantifying critical states of complex diseases using single-sample dynamic network biomarkers[J]. *PLoS Comput Biol*, 13(7): e1005633.
- LIU X, LIU R, ZHAO X M, et al, 2013b. Detecting early-warning signals of type 1 diabetes and its leading biomolecular networks by dynamical network biomarkers[J]. *BMC Medl Genomics*, 6(Suppl 2): S8.
- MATURANA M I, MEISEL C, DELL K, et al, 2020. Critical slowing down as a biomarker for seizure susceptibility[J]. *Nat Commun*, 11(1): 2172.
- MOJTAHEDI M, SKUPIN A, ZHOU J, et al, 2016. Cell fate decision as high-dimensional critical state transition[J]. *PLoS Biol* 14(12): e2000640.
- MUKHERJI R, 2013. Ideas, interests, and the tipping point: Economic change in India[J]. *Rev Int Polit Econ*, 20(2): 363–389.
- OLDE RIKKERT M G M, DAKOS V, BUCHMAN T G, et al, 2016. Slowing down of recovery as generic risk marker for acute severity transitions in chronic diseases[J]. *Crit Care Med*, 44(3): 601–606.
- PENG X, GAO R, REN J, et al, 2023. Specific network information gain for detecting the critical state of colorectal cancer based on gut microbiome[J]. *Brief Bioinform*, 25(1): bbad465.
- QUAX R, KANDHAI D, SLOOT P M A, 2013. Information dissipation as an early-warning signal for the Lehman Brothers collapse in financial time series[J]. *Sci Rep*, 3: 1898.
- RICHARD A, BOULLU L, HERBACH U, et al, 2016. Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process[J]. *PLoS Biol*, 14(12): e1002585.
- RIETKERK M, BASTIAANSEN R, BANERJEE S, et al, 2021. Evasion of tipping in complex systems through spatial pattern formation[J]. *Science*, 374(6564): eabj0359.

- RIETKERK M, DEKKER S C, de RUITER P C, et al, 2004. Self-organized patchiness and catastrophic shifts in ecosystems[J]. *Science*, 305(5692): 1926–1929.
- RUNFOLA C, SHEHEITLI H, BARTOLOMEI F, et al, 2023. In pursuit of the epileptogenic zone in focal epilepsy: A dynamical network biomarker approach[J]. *Commun Nonlinear Sci Numer Simul*, 117: 106973.
- SARKAR S, SINHA S K, LEVINE H, et al, 2019. Anticipating critical transitions in epithelial-hybrid-mesenchymal cell-fate determination[J]. *Proc Natl Acad Sci*, 116(52): 26343–26352.
- SCHEFFER M, BASCOMPTE J, BROCK W A, et al, 2009. Early-warning signals for critical transitions [J]. *Nature*, 461(7260): 53–59.
- SCIUTO A M, PHILLIPS C S, ORZOLEK L D, et al, 2005. Genomic analysis of murine pulmonary tissue following carbonyl chloride inhalation[J]. *Chem Res Toxicol*, 18(11): 1654–1660.
- SHI J, AIHARA K, CHEN L, 2021. Dynamics-based data science in biology[J]. *Natl Sci Rev*, 8(5): nwab029.
- SHI J, LI T, CHEN L, 2016. Towards a critical transition theory under different temporal scales and noise strengths[J]. *Phys Rev E*, 93(3): 032137.
- STAPELBERG N J C, NEUMANN D L, SHUM D, et al, 2019. Health, pre-disease and critical transition to disease in the psycho-immune-neuroendocrine network: Are there distinct states in the progression from health to major depressive disorder? [J]. *Physiol Behav*, 198: 108–119.
- TANAKA G, TSUMOTO K, TSUJI S, et al, 2008. Bifurcation analysis on a hybrid systems model of intermittent hormonal therapy for prostate cancer[J]. *Phys D Nonlinear Phenom*, 237(20): 2616–2627.
- TANAKA H, OGISHIMA S, 2015. Network biology approach to epithelial-mesenchymal transition in cancer metastasis: Three stage theory[J]. *J Mol Cell Biol*, 7(3): 253–266.
- TARAZONA A, FORMENT J, ELENA S F, 2019. Identifying early warning signals for the sudden transition from mild to severe tobacco etch disease by dynamical network biomarkers[J]. *Viruses*, 12(1): 16.
- TREDICCE J R, LIPPI G L, MANDEL P, et al, 2004. Critical slowing down at a bifurcation[J]. *Amer J Phys*, 72(6): 799–809.
- van de LEEMPUT I A, WICHERS M, CRAMER A O J, et al, 2014. Critical slowing down as early warning for the onset and termination of depression[J]. *Proc Natl Acad Sci*, 111(1): 87–92.
- van NES E H, SCHEFFER M, 2007. Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift[J]. *Amer Nat*, 169(6): 738–747.
- VENEGAS J G, WINKLER T, MUSCH G, et al, 2005. Self-organized patchiness in asthma as a prelude to catastrophic shifts[J]. *Nature*, 434(7034): 777–782.
- VIERA A J, 2011. Predisease: When does it make sense? [J]. *Epidemiol Rev*, 33: 122–134.
- WANG R, DEARING J A, LANGDON P G, et al, 2012. Flickering gives early warning signals of a critical transition to a eutrophic lake state[J]. *Nature*, 492(7429): 419–422.
- XIE Y, PENG X, LI P, 2024. MIWE: Detecting the critical states of complex biological systems by the mutual information weighted entropy[J]. *BMC Bioinformatics*, 25(1): 44.
- YANG B, LI M, TANG W, et al, 2018. Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma[J]. *Nat Commun*, 9(1): 678.
- ZHONG J, HAN C, WANG Y, et al, 2022. Identifying the critical state of complex biological systems by the directed-network rank score method[J]. *Bioinformatics*, 38(24): 5398–5405.
- ZHONG J, LIU R, CHEN P, 2020. Identifying critical state of complex diseases by single-sample Kullback-Leibler divergence[J]. *BMC Genomics*, 21(1): 87.
- ZHONG Z, LI J, ZHONG J, et al, 2023. MAPKAPK2, a potential dynamic network biomarker of α -synuclein prior to its aggregation in PD patients[J]. *NPJ Parkinsons Dis*, 9(1): 41.