

分子互作网络的拓扑结构

2016-5-10

小组成员及分工

- 文献翻译：石重庆、车玉隆、刘捷
- PPT制作：石重庆、车玉隆、刘捷
- PPT演讲：石重庆

Abstract

分子间的相互作用通常用网络模型表示，这种网络模型已经成为许多生物学领域的共同语言。初步研究表明，分子互作网络的拓扑结构与生物功能和演化有关。

Background

- 在过去的半个世纪里，我们在分子水平上对生命的理解有了巨大的进展。许多技术也在不断被改进，如微阵列（转录组学），质谱技术（蛋白质组学、代谢组学）和下一代测序技术（基因组学）。从系统生物学的角度上来看，类似的技术已被用来测量分子之间的相互作用，形成了所谓的相互作用组学。

这些分子相互作用的数据是分析、解释和预测生物现象的基础。分子的相互作用可以构成网络，并通过图来表示，之后就产生了网络生物学。

- 本文定义网络生物学为，研究用图来表示分子互作网络的拓扑结构，并且用这样的网络图来做生物的预测。

如下图便是一个有向图的例子

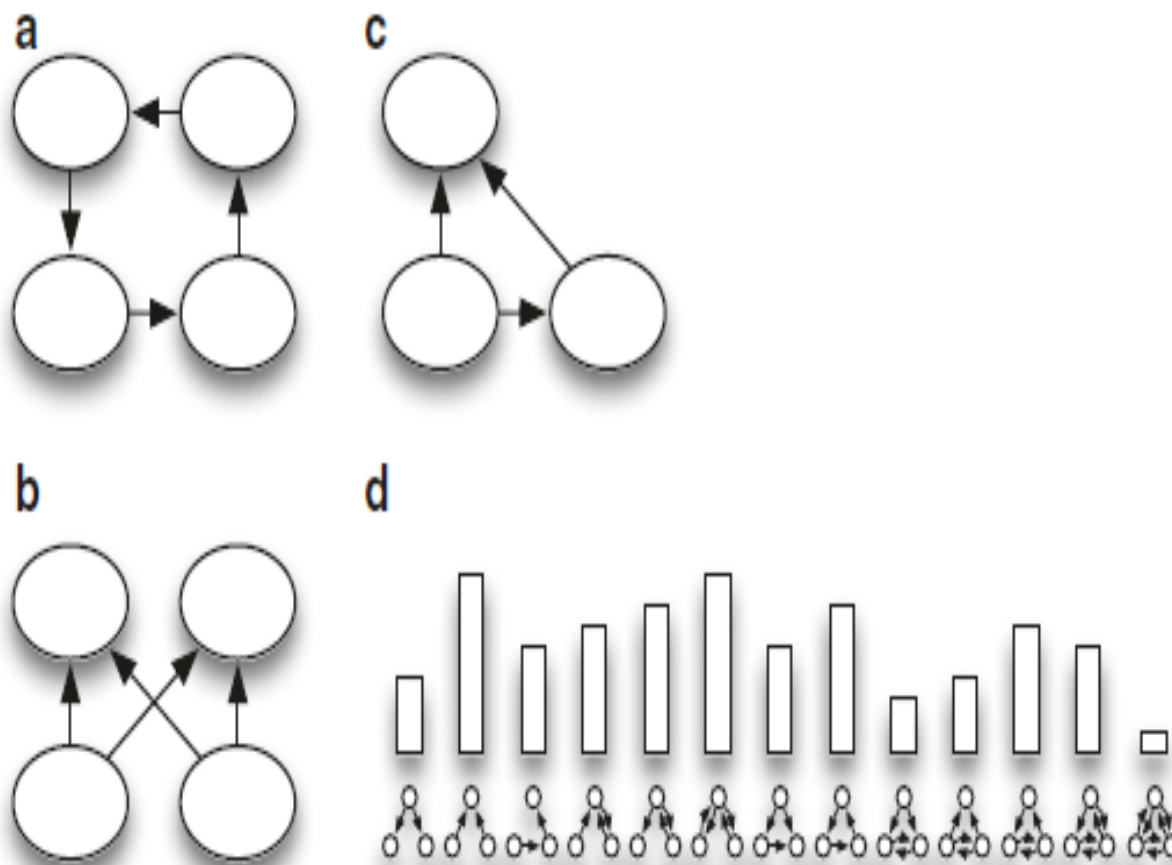


Figure 1 Some motifs thought to be overrepresented in molecular interaction networks. Arrowheads indicate link directionality. **(a)** A four-node feed-back motif. **(b)** A four-node bi-fan motif. **(c)** A three-node feed-forward motif. **(d)** Three-node motif signature for a network.

分子互作网络

分子生物学是研究所有涉及脱氧核糖核酸、核酸、蛋白质和代谢物的细胞过程的学科。从下图我们可以看见一些分子间的相互作用。

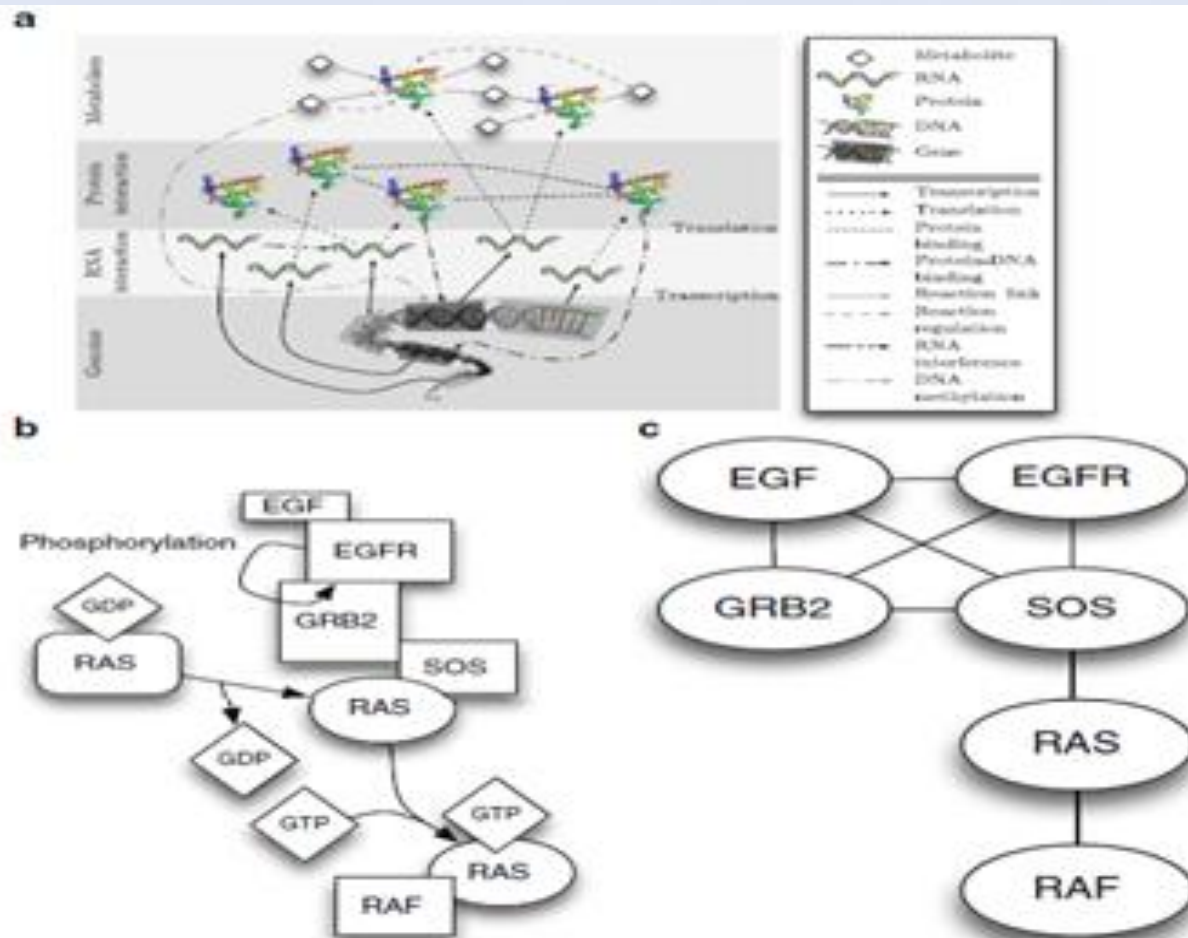


Figure 2 From biological models to networks. (a) Simple overview of molecular interactions in the cell. **(b)** Part of the MAPK/ERK pathway modeled as a network. **(c)** Homogenous protein interaction graph representation of part of the MAPK/ERK pathway.

- 生物学上分子互作网络的初步研究，配合在其他真正网络上的拓扑结构的观察。我们发现基因共表达网络，蛋白质互作网络、代谢网络和转录调控网络都含有无标度网络的性质。无标度网络具有严重的异质性，其各节点之间的连接状况（度数）具有严重的不均匀分布性。虽然各种随机网络模型都再现了一些分子网络的显著特性，但都被批评与分子网络其他重要的特性不一致。

描述性方法的限制

- 随着数据质量的提高，以及研究人员对旧发现的重新审视与解释，多层次视图的细节越来越有争议。
- 在模块水平上，我们发现的模块不像先前假设的那样，不能描述清楚。模块之间的连接似乎有很多，很难区分连接节点。
- 虽然分子网络中的通用模型比之前以为的要少，但是原来的多层次模型仍然是有用的，只不过需要一些修改。

显著性特征检测

- 显著性特征检测就是提取图像中重要的区域。根据显著性区域的应用不同，产生了很多不同的显著性区域定义。
- 显著特征检测的优点在于它的简单性。但这也是它的主要缺点。当选择一个错误的随机网络模型时，会使得原本不重要的功能显得重要。

聚类

- 在复杂的系统中，模块往往在内部高度连通，而与外部世界职工想几个连接。图聚类是一种通过把网络分成许多子网从而发现内部高度连接的模块的方法。大数据时代加快了聚类的进步，可以更有效地处理大数据集。

- 聚类是一个很有用的技术，可以用来了解一个分子网络的模块化结构。但需要注意的是，集群可能无法反应实际的生物模块。不准确的聚类可能是有不好的聚类标准所引起的。产生重叠集群的算法可能会导致给一个节点分配太多或太少的集群，目前仍然缺乏处理这类问题的技术。

中心节点和枢纽

- 网络生物学的早期研究发现表明，有些节点在分子互作网络更加重要。这种表现说明了一个物种的生存相比于其他方面更多地取决于一些重要的节点的存在。
- 比如说，P35基因是一个枢纽型基因，我们发现破坏人类的P35基因会导致癌症。
- 此外，代谢网络和转录调节网络都进一步支持了中心节点十分重要这一观点。

- 一些研究人员试图通过考虑中心节点的交互模式完善中心节点的概念：那些同时与许多互作对象互相作用的被称为party hubs，而那些与它们的对象依次相互作用的被称为date hubs。Party hubs被认为是全局协调员，连接网络模块的组件，而date hubs被认为是局部协调员，连接网络模块。但是，这种区别已经面临着得到的新的数据的挑战，新的数据并没有显示出中央节点之间的明显区别。

总体同源性

- 同源性原理是指生物系统相关的进化是结构性相似的。它的相反的结构相似性意味着有共同的性质，常被用来预测未知蛋白质和基因的功能。在网络中，拓扑结构的相似性同样可以用来推断功能的相似性。

数据挖掘在生物网络中的作用

- 数据挖掘技术已成功地应用于网络生物学，用来研究基因和蛋白质的生物学功能。
- 它更多地是对于子网络的研究而不是考虑整个生物网络的全局性质。这可以帮助我们推导出更具体的生物学结果。

关联推断

- 关联推断的原理是观察是否一个分子大多数的互作对象与某些特性相关联（如特定的生物学过程或分子功能），则该分子本身也可能与该特性相关联。关联推断已被用来给蛋白质分配功能，基于蛋白质互作网络中的大多数直接相邻的蛋白质（即互作对象）共享功能的原理。

- 一个未知的节点可以用它所属模块中最常见的功能来注释。
- 关联推断是一种简单而有效的技术，可以自然延伸到了网络。然而，当网络中的大多数分子的作用是已知的时候它才有效，因此限制了这项技术很好地研究物种。

邻域同调

- 在“Global homology”的部分，已经讨论了如何发现不同物种内的网络具有类似的结构特性。预测方法是使用拓扑学和生物学的相似性，在不同的网络中匹配相似的节点。一旦节点匹配，就可以预测未知的蛋白质或基因的功能。

- 此外，同源性是不是一个一对一的关系：许多节点可以匹配到许多节点。
- 全局比对的主要方法有2种：
 - 1. 聚类每个网络中的节点，计算集群的拓扑匹配分数（“匹配集群”）。
 - 2. 在局部相邻区域两两相似的不同网络和生物学标记可能两两相似的网络中选择节点组（“聚类匹配”）

差异分析

- 所谓的差异分析，就是发现网络结构的变化。基因/蛋白质表达的变化与分子网络中的中心节点存在关联，已被证明是疾病检测的可靠的标志物。如蛋白质互作网络中的中心节点在拓扑结构上的差异表达可以被用来诊断癌症。

生物特性的拓扑结构具有预测的能力

- 一般情况下，我们没有大尺度的良好的生物功能模型，预测方法通常适用于小群体的节点或子网络。也有例外情况，例如代谢网络。通量平衡分析（**FBA**）是一种常用的预测方法，例如可以用来研究代谢网络的稳定性。**FBA**可以考虑额外的物理约束条件，如热力学相互作用等。

- 最大的问题是将额外的生物知识纳入现有模型，对于任何给定的生物特性，我们很少有完整的数据。近期有两种构思，“可控性”和“可观测性”，有可能允许使用部分已知信息预测全局状态。可控性指的是可以“驱动”对系统的状态有很大的影响的节点，可观测性差不多是互补的，即适当注意一小部分的观测节点，这些节点的属性可以再现全局状态。这些技术都可以从有限的可用数据来预测全局属性。

conclusion

- 从全局分析到局部分析，这一趋势是由对网络中更多具体的生物学知识的需求引起的。调查结果表明，要么选择关于大型拓扑模型的无法验证的抽象假设，要么选择小规模的结果，而忽视了大型拓扑模型。局部技术的成功之处在于对生物学知识的拓扑观察，不在于对局部的关注。从这个起点上，我们看到了两个广阔的研究，以提高大型拓扑结构模式的解释力。

- 在理论研究方面，为了尽可能更好地再现真实的分子网络的拓扑结构，网络演化模型的改进是必要的。更好的网络演化模型可以更好地揭示的拓扑结构特征，可以让研究人员更专注于拓扑结果的解释，而且这些模型更便于生物学特征的观察。

- 在实际应用方面，提出了将预测技术应用到分子网络的各种部分当中，即多分辨率分析。
- 上述两个研究方向并不是发展网络生物学的唯一可能的路径。这表明这个新兴的领域仍然有很大的发展潜力。

Thank you