

综述

高阶网络统计指标综述*

刘波¹⁾³⁾ 曾钰洁¹⁾³⁾ 杨荣湄¹⁾ 吕琳媛^{2)1)3)†}

1) (电子科技大学基础与前沿研究院, 成都 610054)

2) (中国科学技术大学网络空间安全学院, 合肥 230026)

3) (电子科技大学长三角研究院(湖州), 湖州 313001)

(2024 年 2 月 17 日收到; 2024 年 4 月 10 日收到修改稿)

复杂网络是描述和理解现实世界中复杂系统的有力工具. 近年来, 为了更准确地描述复杂网络中的交互关系, 或者从高阶视角分析成对交互作用网络, 许多学者开始使用高阶网络进行建模, 并在研究其动力学过程中发现了与成对交互作用网络不同的新现象. 然而, 与成对交互作用网络相比, 高阶网络的研究相对较少; 而且, 高阶网络结构相对复杂, 基于结构的统计指标定义较为分散且形式不统一, 这些都给描述高阶网络的拓扑结构特征带来了困难. 鉴于此, 本文综述了两种最常见的高阶网络——超图和单纯形网络——常用的统计指标及其物理意义. 本文有助于加深对高阶网络的理解, 促进对高阶网络结构特征的定量化研究, 也有助于研究者在此基础上开发更多适用于高阶网络的统计指标.

关键词: 高阶网络, 超图, 单纯形网络, 统计指标**PACS:** 89.75.-k, 89.75.Fb, 02.10.Ox**DOI:** 10.7498/aps.73.20240270

1 引言

现实世界可被视为一个高度复杂的巨系统, 由各种不同类型的子系统构成, 包括社会系统^[1,2]、生物系统^[3,4]、电力系统^[5,6]和交通系统^[7,8]等. 为了全面深入地理解这些复杂系统, 网络科学应运而生^[9-11]. 网络科学的主要研究目的是找出这些复杂系统的共性, 定量和定性地刻画它们所具有的普适性规律, 从而为理解和研究复杂系统提供理论基础和实证应用途径^[12-14]. 为实现这一目的, 将复杂系统抽象成复杂网络通常是研究的第一步, 也是后续研究网络结构和动力学的基础.

传统的复杂网络通常将系统中的主体抽象为节点, 主体之间的成对作用关系抽象为连边. 然而, 对部分复杂系统而言, 只利用这种成对作用关系进

行建模并不合理. 例如, 一篇论文经常由多名科学家共同合作完成^[15,16], 完成一次新陈代谢需要多个反应物的参与^[17], 一个社交群组往往包含多个成员间的互动^[18,19]. 在这些情况下, 描述节点之间的作用关系需要超越成对关系, 建立具有高阶相互作用关系的网络模型, 即高阶网络^[20-23].

相较于成对交互作用网络, 高阶网络不仅能更加精确地描述复杂系统的结构, 而且可揭示系统中的一些新型动力学现象^[24-30]. 例如, 在同步动力学领域, 传统网络中的爆炸相变现象只在满足严格条件时才会发生^[21,31], 而在高阶网络中这种现象却能自发出现; 即使存在负耦合的情况下, 高阶相互作用仍能有效地维持系统的同步, 并且在系统中形成与网络结构有关的集群^[32]; 此外, 高阶相互作用还有助于产生多稳态^[32,33]、嵌合状态^[34,35]以及混沌动态^[32,33,36]. 在传播动力学领域, 传播过程可以在

* 科技创新 2030-“脑科学与类脑研究”重大项目青年科学家项目 (批准号: 2022ZD0211400)、国家自然科学基金重大项目 (批准号: T2293771) 和四川省杰出青年科学基金 (批准号: 2023NSFSC1919) 资助的课题.

† 通信作者. E-mail: linyuan.lv@ustc.edu.cn

网络的高阶结构上发生,个体不但会受到其最临近邻居的影响,还会受到它所参与的高阶交互关系对它的作用,并且高阶传播还具有非连续相变的特性^[28,37-40].

网络结构是研究复杂网络的基础,结合统计量有利于量化描述网络的结构特征.利用网络结构特征,可以衡量不同网络结构对动力学的影响,还可进行图分类^[41]、重要节点挖掘^[42-44]、链路预测^[45-47]、社团检测^[48,49]等信息挖掘任务;通过挖掘一些有共性的结构特征,也可以建立基于通用统计特性的模型网络^[50-53].然而,由于高阶网络的结构复杂,其统计指标也相对复杂,同一类型的统计指标往往还存在多种定义.因此,刻画高阶网络的结构特征具有一定挑战性.现有高阶网络研究主要集中于度相关的指标、聚集系数、节点之间的距离、中心性指标、熵和与代数拓扑相关的一系列指标.这些指标既有从成对交互网络统计指标中拓展出来的,也有高阶网络的特有指标.本文首先介绍了两种最常见的高阶网络——超图和单纯形网络的基本定义,然后综述了这两类高阶网络中常用的统计指标及其物理意义,最后对于高阶网络统计指标的使用场景进行了总结和展望.

2 高阶网络的基本概念

高阶网络考虑了两个及以上节点之间的高阶关系,能够更全面地捕捉节点之间的复杂关联和多元交互关系.高阶网络的两种主要表示形式为超图(hypergraph)和单纯形网络(simplicial network)^[20,54].

2.1 超图

相比于成对交互作用网络,超图可以表示点边

之间的多元关联关系.超图可以定义为 $\mathcal{H} = (V, E)$,其中 V 为节点集,每个节点被称为超节点(hypervertex); E 为边集,每条超边(hyperedge) $e \in E$ 由它所连接的节点组成,是节点集 V 的一个子集,即 $e \subseteq V$,并且一条超边可以由任意数量的节点组成.使用超图表示高阶网络不受闭包特性的限制,例如,超图 $\mathcal{H}$ 的一条超边 e , e' 是 e 的子集, e' 不一定是超图 $\mathcal{H}$ 的超边.如果一个超图中的每条边都恰好包含 d 个节点,则称此超图为 d -均匀超图(d -uniform hypergraph),表示为 $d\text{-}\mathcal{H}$.图1是两类超图的简单示例:图1(a)是一个普通超图,其中包含了11个超节点和7条超边;图1(b)是一个3-均匀超图,包含9个节点和4条超边,并且每条超边中都包含3个超节点.

目前,超图的研究已经渗透到多个学科领域,包括数据挖掘^[55]、社交网络分析^[56]、生物信息学^[57]和通信网络^[58]等.在处理多元关系数据时,超图展现出了其在表示复杂关系和高阶交互方面的优势.一些研究突破包括新型超图模型的开发,旨在更准确地捕捉实际系统的结构特征,并且基于超图的算法已被应用于社区发现^[59]、节点分类^[60]和网络演化分析^[61]等任务.此外,超图的研究还涉及到高阶图神经网络的发展,这是融合了超图理论和深度学习的前沿方向,其目标是提高数据的表示能力和预测精度^[62].

2.2 单纯形网络

单纯形网络是另一种常用的高阶网络表示方式,它基于代数拓扑的单纯复形(simplicial complex)思想,用单纯形(simplex)来表示高阶网络中的节点交互关系.因此,单纯形网络也可被认为是一个由若干个单纯形组成的单纯复形,其维度是其中最高阶单纯形的阶数.一个 k 阶单纯形由

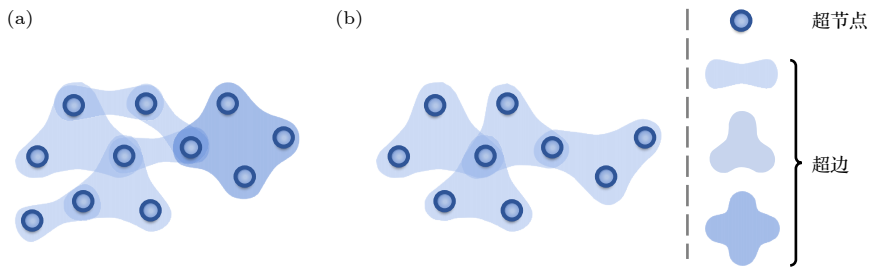


图1 两种不同类型超图的简单示例 (a) 一个拥有11个节点的超图; (b) 一个拥有9个节点的3-均匀超图

Fig. 1. A simple example of two different types of hypergraphs: (a) Simple hypergraph with 11 nodes; (b) 3-uniform hypergraph with 9 nodes.

$k+1$ 个节点组成, 例如节点为 0 阶单纯形, 边为 1 阶单纯形, 三角形为 2 阶单纯形, 四面体为 3 阶单纯形. 相比于超图, 单纯形网络有更严格的数学定义, 每个单纯形都要满足闭包性, 即一个 k 阶单纯形不仅包含 k 阶交互关系, 还包含组成该单纯形的 $k+1$ 个节点所能组成的所有低阶单纯形. 单纯形网络的构建方式主要有两种, 一种是将高阶交互数据构建为单纯形, 另一种是将传统成对交互作用网络中的团 (clique, k 阶团为 $k+1$ 个节点组成的全连通子图) 结构抽象为单纯形.

2.2.1 基于高阶数据的单纯形网络

利用高阶交互数据可以构建单纯形网络, 如图 2(a) 中每个时间戳对应的数据都表示一个单纯形, 利用这些数据可以构建出图 2(b) 所示的单纯形网络. 基于单纯形网络可以提取出一个骨架/投影网络 (skeleton/projected network), 该网络由原网络中所有的 0 阶单纯形和 1 阶单纯形组成, 是一个成对交互作用网络. 基于这个网络可以计算出所有的团结构, 借助单纯形网络中的高阶信息可以把这些团划分为开团 (open clique) 和闭团 (closed clique): 如果一个团的所有节点都在同一个单纯形内, 则该团为闭团, 否则为开团. 例如, 图 2(b) 中的 (1, 4, 5) 3 个节点组成了一个闭团, 对应图 2(a) 单纯形数据中的 t_2 时间戳; (5, 9, 10) 组成了一个开团, 因为这 3 个节点没有同时出现在任意一个单纯形中.

2.2.2 团复形网络

将成对交互作用网络中的团结构抽象为单纯形可以构建团复形 (clique complex). 对于给定的

一个由节点和连边组成的成对交互作用网络, 计算网络中所有的团结构, 再将每个 k 阶团视为 k 阶单纯形, 可以得到这个网络所对应的团复形网络, 也是一个单纯形网络. 如图 2(c) 所示的这个成对交互作用网络, 可以计算出其中包含 11 个 0 阶团 (节点)、16 个 1 阶团 (连边)、2 个 2 阶团 (三角形)、1 个 3 阶团 (四面体), 将这些团看作对应阶数的单纯形就可以得到如图 2(d) 所示的团复形网络. 团复形网络满足单纯形网络的所有性质, 同样满足闭包性, 并且其中任意两个单纯形的交集只有空集或团复形集合中的子集两种情况, 生成团复形网络的原始成对交互作用网络即为其骨架网络.

基于高阶数据的单纯形网络和团复形网络的主要区别在于: 基于高阶交互数据生成的单纯形, 其节点之间存在真实的高阶交互关系; 而团复形中的单纯形是基于对应阶数的团生成的, 即是基于一个成对交互作用网络生成的, 节点之间不一定存在真实的高阶交互关系. 因此, 两个网络的应用场景不同, 基于高阶数据的单纯形网络通常被用来探究具有高阶交互关系的复杂系统的特性和动力学; 团复形网络则用于为传统的成对交互作用系统提供更高阶的向量空间或高阶视角, 以研究高阶结构在网络中的作用和功能.

具体而言, 基于高阶数据的单纯形网络更多的是探索网络中的动力学行为的研究. 例如, 在高阶传播中, 研究者提出了多种单纯形网络上的传播模型 [28,29], 也有研究分析了高阶相互作用的引入为网络动力学带来的新特性 [63], 单纯形网络还可以用于研究网络中的振荡和同步现象 [26,64]. 此外,

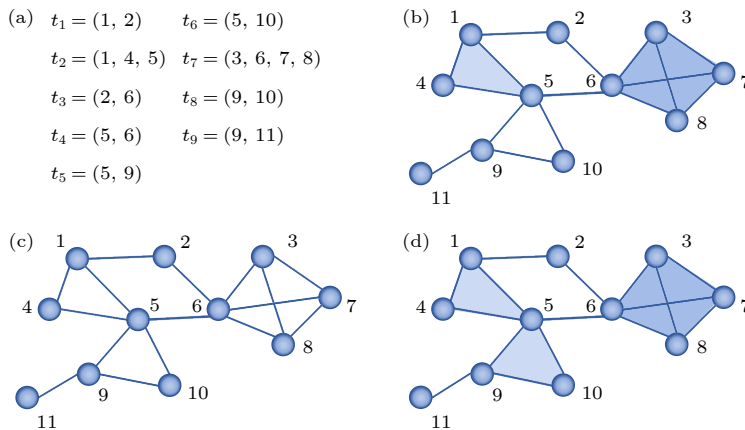


图 2 单纯形网络相关示意图 (a) 一组时序高阶交互数据; (b) 一个 11 节点的单纯形网络; (c) 基于图 (b) 中单纯形网络的骨架网络; (d) 一个 11 节点的团复形网络

Fig. 2. Correlation diagrams of the simplicial network: (a) A set of temporal higher-order interaction data; (b) a simplicial network with 11 nodes; (c) a skeleton network based on the simplicial network in Fig. 2(b); (d) a clique complex with 11 nodes.

在高级网络重构^[65]、高级链路预测^[66]等多种领域也有广泛的应用. 团复形网络常用于高级结构相关的研究, 包括团、洞等高级结构的计算^[67], 以及探索脑网络中的高级结构对大脑认知和信息传播的影响等^[68,69], 也有研究者提出团复形网络的生成模型, 进行高级网络建模和分析^[70]. 不过, 无论是用高级交互数据构建的单纯形网络, 还是由成对交互网络构建的团复形网络, 其最终的形式都属于单纯形网络, 后续介绍的基于单纯形网络的统计量均可使用.

3 基于超图的统计指标

3.1 指标概览

本节对超图的一些核心统计指标进行了全面的概述. 这些指标在超图分析和应用中起着至关重要的作用, 有助于更深入地理解超图的结构和特性. 表1列出了这些指标的概览, 涵盖了多个维度. 其中, 度相关指标是描述超图中节点连接情况的基础指标, 它反映了节点在超图中的直接连接情况, 对于分析超图的连通性和复杂性具有重要意义. 其次, 介绍了关于节点的局部聚集系数和关于网络整体的全局聚集系数, 能够揭示超图中节点群聚的紧密程度, 对于理解超图的结构特性具有重要意义. 此外, 本节介绍了考虑多方面因素的超图节点距离的度量方式, 以及反映连边或者网络连接紧密程度的密度指标. 另外, 曲率作为超图的一个几何特性指标, 能够揭示超图的形态和弯曲程度, 在超图嵌入和形状分析等领域具有广泛应用. 本节还介绍了几种可用于衡量超图中节点的重要性的影响力的中心性指标. 最后, 熵作为信息论中的一个重要概念, 也被引入到超图分析中, 通过描述超图中信息的不确定性和复杂性, 来量化超图结构的复杂性.

表1 基于超图的统计指标总结

Table 1. Summary of statistical indicators of the hypergraph.

指标类型	指标名称
度相关指标	度、超度、超边度、余平均度
聚集系数	节点的聚集系数、网络的聚集系数
距离相关指标	路径长度、超节点之间的距离
密度相关指标	超边密度、超图密度
曲率相关指标	Forman-Ricci曲率、Ollivier-Ricci曲率
中心性指标	度中心性、核心度中心性、接近中心性、介数中心性、特征向量中心性
熵相关指标	超图熵、超图的香农熵、加权超图的超图熵

这些指标为全面而深入地分析超图结构和特性提供了工具和方法.

3.2 度相关指标

3.2.1 度

度 (degree) 意在揭示节点和其他节点之间的关联程度, 超图中的度可以定义为超节点的邻居节点数量, 一个超节点的邻居节点被定义为与该超节点在同一个超边的超节点. 超节点的度可以通过超图的邻接矩阵计算得出, 即超节点 v_i 的度可以表示为

$$D(v_i) = \sum_{j=1}^N \text{sgn}(A_{ij}), \quad (1)$$

其中 N 为超图中的超节点数量; $\mathbf{A}$ 为超图的邻接矩阵, 如果超节点 i 和超节点 j 同时在 l 个超边内, $A_{ij} = l$, $l \in [0, M]$; sgn 为符号函数; M 为超图中的超边数量.

3.2.2 超度

在超图中, 超度 (hyperdegree) 可以被定义为该超节点所包含的超边数量, 可以通过超图的关联矩阵计算得出, 即超节点 v_i 的超度可以表示为

$$HD(v_i) = \sum_{j=1}^M H_{ij}, \quad (2)$$

其中 $\mathbf{H}$ 为超图的关联矩阵, 如果超节点 v_i 在超边 e_j 中 $H_{ij} = 1$, 反之 $H_{ij} = 0$.

3.2.3 超边度

在超图中可以根据超边中节点的数量定义超边度 (hyperedge degree), 也称为该超边的基, 超边 e_j 的超边度可以表示为

$$ED(e_j) = |e_j|. \quad (3)$$

3.2.4 余平均度

余平均度 (excess average degree) 用于反映节点的邻居节点的情况. 在成对交互作用网络中, 余平均度是计算节点的邻居节点的度平均值. 类似地, 在超图中, 余平均度被定义为节点的邻居节点的平均超度:

$$\langle dk_{nn} \rangle_i = \frac{1}{|\tau(i)|} \sum_{j \in \tau(i)} HD(v_j), \quad (4)$$

其中, $\tau(i)$ 为节点 v_i 的邻居节点集合, $HD(v_j)$ 为邻居节点 v_j 的超度.

3.3 聚集系数

聚集系数 (clustering coefficient) 是反映网络中节点的聚集程度的指标^[50], 分为节点的聚集系数和网络的聚集系数; 前者是一个局部指标, 后者是针对网络整体的全局指标. 在成对交互作用网络中, 局部聚集系数指节点的邻居节点之间实际存在连边的数量占可以存在的所有连边数量的比例. 如果一个节点的两个邻居节点之间存在连边, 则代表这 3 个节点之间存在一个三角形, 所以局部聚集系数反映的也是该节点与邻居节点之间存在三角形的情况. 而全局聚集系数则通过将所有节点的聚集系数求平均值或者通过社会学中的传递系数^[71]来计算.

利用这种思想可以在超图中计算聚集系数, 最简单的情况就是在 d -均匀超图中计算聚集系数. 此时 $C^d(v_i)$ 是超节点 v_i 的聚集系数, 网络中的对应全局聚集系数是 C_g^d ,

$$C^d(v_i) = \frac{|\{e_{i_1 i_2 \dots i_d} : v_{i_1}, v_{i_2}, \dots, v_{i_d} \in \tau(i), e_{i_1 i_2 \dots i_d} \in E\}|}{\binom{|\tau(i)|}{d}}, \quad (5)$$

$$C_g^d = \frac{1}{n} \sum_{i=1}^N C^d(v_i). \quad (6)$$

超图中计算聚集系数的难点之一就是两个超节点之间可能存在多条超边, 因此 3 个节点可能出现由多条不同超边组成的多个三角形结构. 两个节点之间已有的超边数量可以由其对对应邻接矩阵得到, 但是由于在非均匀超图中的超边的阶数规模无法估计, 进而节点之间可能存在的超边数量无法估计, 导致无法准确地计算出聚类系数. 一种解决方法就是把超节点之间的关系投影到成对关系网络中, 只判断两个超节点是否存在超边关系, 不考虑存在超边的阶数和数量, 即利用超图的邻接矩阵 $\mathbf{A}$ 即可计算:

$$C_{v_i} = \frac{\sum_{j,k=1}^n \text{sgn}(A_{ij}) \text{sgn}(A_{ik}) \text{sgn}(A_{jk})}{D(v_i)(D(v_i) - 1)}, \quad (7)$$

$$C_g = \frac{1}{n} \sum_{i=1}^n C_{v_i}. \quad (8)$$

由于将局部聚集系数求平均所获得的全局聚集系数可能并不准确, 因此另一种方法是根据社会学中的传递系数来计算. 该方法将全局聚集系数定义为网络中所有三角形的数量和 2 阶路径数量的比例. Estrada 和 Rodríguez-Velázquez^[72] 将此方法推广到超图中, 定义了一种聚集系数 $C_2(\mathcal{H})$:

$$\begin{aligned} C_2(\mathcal{H}) &= \frac{6 \times \text{超三角形的数量}}{2 \text{阶路径的数量}} \\ &= \frac{CW_3(\mathcal{H}) - 6t}{W_2(\mathcal{H}) - CW_2(\mathcal{H}) - 6t} \\ &= \frac{\sum_i \mu_3(i) - 6t}{W_2(\mathcal{H}) - \sum_i \mu_2(i) - 6t} \\ &= \frac{\sum_{s=1}^N \lambda_s^3 - 6t}{\sum_{s=1}^N \left[\left(\sum_{j=1}^N u_{is} \right)^2 - 1 \right] \lambda_s^2 - 6t}, \end{aligned} \quad (9)$$

其中, $CW_a(\mathcal{H})$ 表示的是网络中所有长度为 a 的闭合路径的数量; t 表示不满足超三角形条件的长度为 3 的闭合路径数量,

$$t = \sum_{j=1}^M (-1)^{j+1} \sum_{i_1, i_2, \dots, i_j} \binom{|\cap_{r=1}^j E_{i_r}|}{3};$$

$W_2(\mathcal{H})$ 表示所有长度为 2 的游走的数量; $\mu_c(i) = (A^c)_{ii} = \sum_{s=1}^N (u_{is})^2 \lambda_s^c$; u_{is} 为 A^c 的第 i 个特征向量中的第 s 个值, λ_s^c 为 A^c 的第 s 个特征值, c 为常数.

3.4 超节点之间的距离

在无向无权超图中, 路径 (path) 被定义为一个超节点序列 $P = \{v_1, v_2, \dots, v_p\}$, 其中每两个相邻的超节点 v_i 和 v_{i+1} 之间都存在一条超边, P 也被称为从超节点 v_1 到 v_p 之间的一条路径, 路径所涉及的超边的数量被称为该路径的长度 (length), 也可称为两个超节点之间的距离. 但是这种计算方法只考虑了超边的数量, 没有进一步考虑超边中超节点数量和超边之间重叠的超节点数量对度量超节点之间距离的影响.

在一些基于随机游走 (random walk) 的算法中, 研究人员单独考虑了将超边的基作为超边的权

重进行随机游走^[73,74]; 还有一些研究是基于超边之间相交超节点的数量进行随机游走^[75,76]. 最近, Vasilyeva 等^[77]提出了一个同时考虑超边的基和超边相交超节点的数量这两个因素的超节点距离指标, 他们将超图中的距离在超图映射的加权线图 (weighted line graph) 中进行计算, 加权线图是将超图中的超边当作节点, 加权线图中的连边表示两个超边之间存在重复超节点, 权重则为自行定义, 在下文会详细介绍. 图 3(b) 就是图 3(a) 对应的一个加权线图 $L(\mathcal{H})$, 图 3(a) 之中的 3 条连边在图 3(b) 中被当作节点, e_1 和 e_2 之间在图 3(a) 中有重复的节点, 所以图 3(b) 中它们之间存在连边. 接下来可以根据加权线图来定义超图中节点 v_i 和 v_k 的距离为

$$d^w(v_i, v_k) = 1 + \min_{e \in E_i, \tilde{e} \in E_k} d^L(e, \tilde{e}), \quad (10)$$

在这里 E_i 和 E_k 分别为节点 v_i 和 v_k 所涉及的超边的集合, $d^L(e, \tilde{e})$ 是超边 e 和 $\tilde{e}$ 在加权线图_{中的距离}, 其计算方式和含权成对交互网络中计算两点之间的距离的方式相同.

Vasilyeva 等^[77]考虑了两个超节点之间距离的 4 种情况: 1) 对于由成对交互作用网络生成的团作为超边的超网络来说, 节点在成对交互作用网络中的距离应该和在超图中的距离相等; 2) 当两个超节点之间经过超边的数量相同时, 超边的交点 (重叠的超节点) 数量越多距离应该越近, 如图 4(a) 中超节点 v_i 和 v_j 的距离要小于图 4(b) 中节点 v_i 和 v_j 的距离; 3) 当两个节点之间经过超边的数量相同时, 所涉及的超边中的节点越多, 距离应该越大, 如图 4(d) 中节点 v_i 和 v_j 的距离要大于图 4(c) 中

节点 v_i 和 v_j 的距离; 4) 两个节点之间涉及到的节点数量相同时, 超边数量越多, 距离应该越大, 如图 4(f) 中节点 v_i 和 v_j 的距离要大于图 4(e) 中节点 v_i 和 v_j 的距离. 根据这 4 种情况定义了加权线图_{中的权重}, 也就是超图中两条超边 e_i 和 e_j 之间的权重 w_{ij} :

$$w_{ij} = w_{ji} = w(|e_i \cup e_j|, |e_i \cap e_j|) = \frac{1}{3} \left(|e_i \cup e_j| + \frac{|e_i \cup e_j|}{|e_i \cap e_j|} \right) - 1. \quad (11)$$

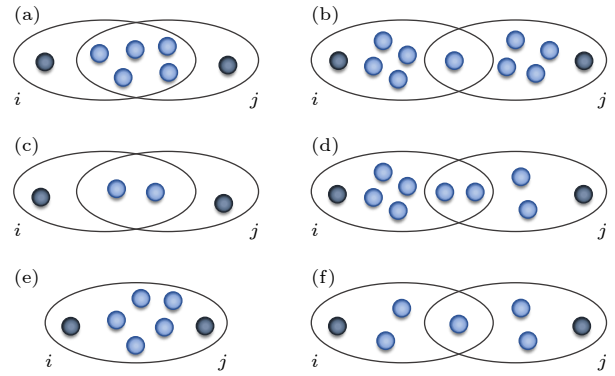


图 4 不同情形下两个超节点之间的路径示意图^[73]

Fig. 4. Diagram of the path between two hypernodes in different cases^[73].

和成对交互作用网络一样, 超节点之间也可能存在多条路径, 两个节点之间距离最短的路径被称作最短路径 (shortest path), 其长度为最短路径长度^[78]. 两个节点之间的最短路径长度是固定的, 但是其最短路径可能不唯一.

3.5 密度

密度 (density) 是理解网络的连接紧密程度以及节点之间相互关系的重要指标. 高密度网络中的节点更紧密地连接在一起, 而低密度网络中的连接相对稀疏. 在超图中, 同样可以用密度来衡量网络的紧密程度, 目前在超图中有如下关于密度的指标.

3.5.1 超边密度

超边密度 (hyperedge density) 是为了计算网络中某一节点所连接的超边在网络中的重叠程度. 在一个非均匀超图中, 当一个节点 v_i 所在的超边个数为 k 时, 它的超边密度定义为

$$D_h(i) = \frac{z_i}{z_{\max}}, \quad (12)$$

其中, z_i 为配位数 (coordination number), 为节点

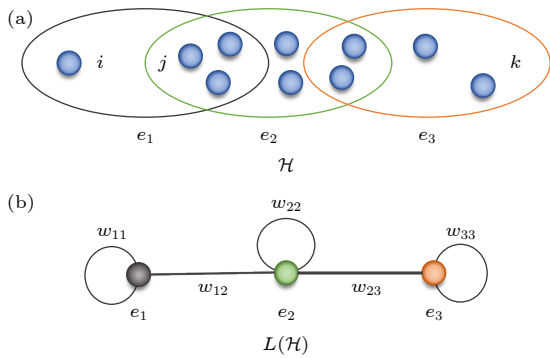


图 3 一个 3 条超边的超图和其对应的加权线图^[73]: (a) 一个 3 条超边的超图; (b) 图 3(a) 对应的加权线图

Fig. 3. A hypergraph with 3 hyperedges and its corresponding weighted line graph^[73]: (a) A hypergraph with 3 hyperedges; (b) a weighted line graph corresponding to Fig. 3(a).

v_i 直接连接的节点个数, 即节点 i 的度; 当节点所在的超边完全不重叠时, z 取最大值, 有 $z_{\max} = \sum_{j \in E_i} |e_j| - k$, E_i 为节点 v_i 所在超边的集合.

接下来, 可以进一步考虑一个节点所在的超边重叠最严重情况的配位数, 也就是配位数的下界 $z_{\min}$, 超边密度可以表示为 $D_{\text{dh}}(i)$:

$$D_{\text{dh}}(i) = \frac{z_{\max} - z_i}{z_{\max} - z_{\min}}. \quad (13)$$

由于非均匀超图的超边的基差异比较大, 所以这里讨论的是均匀超图中的配位数的上下界. 在 d -均匀超图中 $z_{\max} = (d-1)k$, 但是推导 $z_{\min}$ 的难度是随着 d 增加的, 所以目前仅考虑最常见的也是最简单的 3-均匀超图的配位数下界, $z_{\min} = \lceil (1 + \sqrt{8k+1})/2 \rceil$, 更高维的均匀超图建议直接使用 (10) 式进行计算. Zlatić 等 [79] 还讨论了一种特殊的 3-均匀超图的超边密度, 该超图中的超节点共分为 3 种类型, 每条超边都包含 3 种不同类型的节点, 此时可以推导出节点的配位数的下界, 对于 $q(q-1) < k \leq q^2$, $z_{\min} = 2q$; 对于 $q^2 < k \leq q(q+1)$, $z_{\min} = 2q+1$; q 取正整数. 无论是在均匀超图还是非均匀超图, 超图密度的取值范围都是 $[0, 1]$.

3.5.2 超图密度

超图密度 (hypergraph density) 是指已存在的超边数量与可能存在的所有超边数量之比. 对于 d -均匀超图来说, 超边中的节点数量均是 d 个, 超图的密度定义为

$$d(dH) = \frac{M}{\binom{d}{N}} = \frac{M \cdot d!}{N(N-1) \cdots (N-d+1)}. \quad (14)$$

对于非均匀超图来说, 由于超边的基不均匀, 可能存在的超边数量为 2^N , 密度为 $M/2^N$; 由于分母随着节点数量呈指数级增长, 远多于实际存在的连边, 因此其超图密度通常很小.

3.6 曲率

曲率 (curvature) 最初通过对描述光滑物体形状 (如曲线或曲面) 的函数进行二阶导数运算来计算. 在黎曼几何中, 曲率作为编码光滑流形的黎曼度量的几何不变量的张量, 是黎曼几何中的重点研究内容 [80]. 曲率在图论中也有着特殊意义, 比如在平面几何中, 曲率可以通过计算曲线在该点的切线角度的变化率来得到; 如果曲率越大, 那么该点的

切线角度变化越快, 曲线在该点也就越弯曲. 研究者将这些以往的曲率定义迁移到复杂网络中, 提出关于成对交互作用网络和超图的曲率定义.

3.6.1 Forman-Ricci 曲率

Forman-Ricci 曲率是针对边的统计指标, 在图中有如下应用. 首先, Forman-Ricci 曲率可用于捕捉网络的几何特征和拓扑结构. 它提供了一种衡量网络形状和连通性的方法, 通过计算网络中所有边的 Forman-Ricci 曲率并分析它们的分布, 可以了解网络的整体拓扑结构. 其次, Forman-Ricci 曲率还可用于研究网络中的连边重要性和社团发现. 通过与其他边缘度量指标如边缘介数中心性、分布度和嵌入度的相关性比较, 可以揭示网络中的关键连边和重要社团结构. 通过 Forman-Ricci 曲率的分析, 可以更好地了解复杂网络的结构、动态和演化, 从而在网络科学和实际应用中提供洞察和解决方案. 对于具有显著高或低 Forman-Ricci 曲率值的边, 它们可能在网络的连通性和信息流中扮演重要角色. 通过识别这些边, 可以了解哪些部分对维持网络整体结构和功能至关重要. 而对于网络的整体结构来说, 边的 Forman-Ricci 曲率总和或平均值可以作为衡量网络连通性的一个全局指标. 在一些情况下, 整个网络的 Forman-Ricci 曲率可以体现网络整体结构的情况, 如网络是否高度集群或者是否存在明显的社区结构. 在无向有权网络中, 一条连边 $\mathcal{E}$ 的 Forman-Ricci 曲率定义如下 [81]:

$$F(\mathcal{E}) = w_{\mathcal{E}} \left(\frac{w_i}{w_{\mathcal{E}}} + \frac{w_j}{w_{\mathcal{E}}} - \sum_{\mathcal{E}_h \in E_i \sim \mathcal{E}} \frac{w_i}{\sqrt{w_{\mathcal{E}} w_{\mathcal{E}_h}}} - \sum_{\mathcal{E}_l \in E_j \sim \mathcal{E}} \frac{w_j}{\sqrt{w_{\mathcal{E}} w_{\mathcal{E}_l}}} \right), \quad (15)$$

其中 $w_{\mathcal{E}}$ 表示边 $\mathcal{E}$ 的权重, w_i 和 w_j 表示边 $\mathcal{E}$ 所连接的两个节点 v_i 和 v_j 的权重, E_i 表示节点 i 所在超边的集合, $E_i \sim \mathcal{E}$ 则表示边集 E_i 中去除掉连边 $\mathcal{E}$. 如果在无向无权图中, 那么 (15) 式可以简化为

$$F(\mathcal{E}) = 2 - (D_i - 1) - (D_j - 1) = 4 - D_i - D_j, \quad (16)$$

其中 D_i 和 D_j 表示节点 v_i 和节点 v_j 的度.

而基于上述成对交互作用网络中曲率的定义, Leal 等 [82] 和 Eidi 等 [83] 分别对超图中的 Forman-Ricci 曲率进行了拓展. 他们根据上述定义, 认定一条超边 e 的 Forman-Ricci 曲率也和成对交互作用网络的计算思想一致, 并且由于一条超边 e 可以包

含多个顶点, 那么在无向有权超图中的公式可以拓展为

$$F(e) = w_e \left[\sum_{h \in e} \left(\frac{w_h}{w_e} - \sum_{e_l \sim h} \frac{w_h}{\sqrt{w_e w_{e_l}}} \right) \right]. \quad (17)$$

同样地, 如果所有顶点和超边都没有权重, 即在无向无权超图中, 那么 (17) 式可以简化为

$$\begin{aligned} F(e) &= \sum_{h \in e} [1 - (HD(v_h) - 1)] \\ &= 2|e| - \sum_{h \in e} HD(v_h). \end{aligned} \quad (18)$$

3.6.2 Ollivier-Ricci 曲率

Ollivier-Ricci 曲率^[84]是根据几何曲率的概念来定义的. 它的计算涉及到节点间的 Wasserstein 距离和概率分布. 在成对交互网络中, Ollivier-Ricci 曲率的计算方法如下: 首先, 对于一条边 e 两端的节点 v_u 和 v_v 定义一个概率分布. 例如, 这个分布可以是均匀分布, 即节点 v_u 和 v_v 将概率均匀分配给它自己和它的邻居. 接着使用 Wasserstein 距离计算节点 v_u 和 v_v 的概率分布间的距离. 这通常涉及到解决一个最优传输问题. Ollivier-Ricci 曲率可通过比较节点 v_u 和 v_v 的 Wasserstein 距离和其在图中的实际距离来计算. 具体来说, 其计算公式为

$$\kappa(e) = 1 - \frac{W(\mu_u, \mu_v)}{d(v_u, v_v)}, \quad (19)$$

其中, $W(\mu_u, \mu_v)$ 是节点 v_u 和 v_v 的概率分布 μ_u 和 μ_v 间的 Wasserstein 距离, $d(v_u, v_v)$ 是节点 v_u 和 v_v 间在图中的最短路径长度.

Wasserstein 距离, 也称为“Earth Mover’s Distance”(EMD), 是用来度量两个概率分布间差异的一种方法. 在数学上, Wasserstein 距离可以用多种方式定义, 其中一种常见的方法是基于线性规划来计算, 其计算公式为

$$\begin{aligned} &W(\mu_u, \mu_v) \\ &= \min_{\mathcal{E} \in \Pi(\mu_u, \mu_v)} \sum_{i \in N(u)} \sum_{j \in N(v)} d(v_i, v_j) \mathcal{E}(v_i, v_j), \end{aligned} \quad (20)$$

其中, $N(u)$ 和 $N(v)$ 分别为节点 v_u 和 v_v 的邻居, $d(v_i, v_j)$ 是网络中从节点 v_i 到节点 v_j 的最短路径长度, $\mathcal{E}(v_i, v_j)$ 表示从节点 v_i 到节点 v_j 的转移的概率值, $\Pi(\mu_u, \mu_v)$ 是所有可能的传输计划的集合, 这些计划满足 $\sum_j \mathcal{E}(v_i, v_j) = \mu_u(i)$ 和 $\sum_i \mathcal{E}(v_i, v_j) =$

$\mu_v(j)$. Wasserstein 距离度量了使两个概率分布相似所需的最小“工作量”或成本. Ollivier-Ricci 曲率指标越大, 说明这条边连接的两个节点在网络中的拓扑位置越紧密或越重要.

在有向超图 $\mathcal{H}^d = \{V, E^d\}$ 中, Ollivier-Ricci 曲率的定义如下^[85]: 任意一条有向超边 e^d 可表示为 $A = \{x_1, \dots, x_n\} \xrightarrow{e^d} B = \{y_1, \dots, y_m\}$, $(n, m) \leq |V|$, 即在超边 e^d 中, $A = \{x_1, \dots, x_n\}$ 和 $B = \{y_1, \dots, y_m\}$ 是两个节点集合, 超边的方向是从节点集 A 指向节点集 B , 节点集 A 中的每一个节点都指向节点集 B 中的节点. 那么该超边的 Ollivier-Ricci 曲率为

$$\kappa(e_d) := 1 - W(\mu_{A^{\text{in}}}, \mu_{B^{\text{out}}}), \quad (21)$$

其中, $\mu_{A^{\text{in}}}$ 为入节点的概率分布, 即:

$$\mu_{A^{\text{in}}} = \sum_{i=1}^n \mu_{v_i}^{\text{in}}, \quad \forall 1 \leq i \leq n, \quad \forall z \in V, \quad (22)$$

$$\mu_{v_i}^{\text{in}}(z) = \begin{cases} 0, & z = v_i \text{ \& } d_{v_i}^{\text{in}} \neq 0, \\ 1/n, & z = v_i \text{ \& } d_{v_i}^{\text{in}} = 0, \\ \sum_{z \in B} \frac{1}{nd_{v_i}^{\text{in}} |B|}, & z \neq v_i \text{ \& } z \in B, \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

其中, $d_{v_i}^{\text{in}}$ 表示节点 v_i 的入度.

同样地, $\mu_{B^{\text{out}}}$ 为出节点的概率分布, 其计算方法为

$$\begin{aligned} \mu_{B^{\text{out}}} &= \sum_{j=1}^m \mu_{v_j}^{\text{out}}, \quad \forall 1 \leq j \leq m, \quad z \in V(H), \\ \mu_{v_j}^{\text{out}}(z) &= \begin{cases} 0, & z = v_j \text{ \& } d_{v_j}^{\text{out}} \neq 0, \\ 1/m, & z = v_j \text{ \& } d_{v_j}^{\text{out}} = 0, \\ \sum_{z \in A} \frac{1}{md_{v_j}^{\text{out}} |A|}, & z \neq v_j \text{ \& } z \in A, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (24)$$

其中, $d_{v_j}^{\text{out}}$ 表示节点 v_j 的出度.

进而可以计算二者的 Wasserstein 距离 $W(\mu_{A^{\text{in}}}, \mu_{B^{\text{out}}})$:

$$\begin{aligned} &W(\mu_{A^{\text{in}}}, \mu_{B^{\text{out}}}) \\ &= \min_{\mathcal{E} \in \Pi(\mu_{A^{\text{in}}}, \mu_{B^{\text{out}}})} \sum_{u \rightarrow A} \sum_{B \rightarrow v} d(u, v) \mathcal{E}(u, v), \end{aligned} \quad (25)$$

其中 $u \rightarrow A$ 指的是指向节点集合 A 中任意一个节点的节点, 而 $B \rightarrow v$ 指的是被节点集合 B 中任意一个节点指向的节点. $d(v_u, v_v)$ 是从 v_u 到 v_v 的最短超路径的长度, $\mathcal{E}(v_u, v_v)$ 表示从顶点 v_u 转移到

顶点 v_v 的概率, $\Pi(\mu_{A^{\text{in}}}, \mu_{B^{\text{out}}})$ 是所有可能的传输计划的集合, 其需同时满足:

$$\begin{cases} \sum_{u \rightarrow A} \mathcal{E}(u, v) = \sum_{j=1}^m \mu_{y_j}^{\text{out}}(v), \\ \sum_{B \rightarrow v} \mathcal{E}(u, v) = \sum_{i=1}^n \mu_{x_i}^{\text{in}}(u). \end{cases} \quad (26)$$

如此一来便可计算超边的 Ollivier-Ricci 曲率. 与 Forman-Ricci 曲率相比, Ollivier-Ricci 曲率考虑的是基于 Wasserstein 距离 (最优传输成本) 的度量, 它在概念上更接近于黎曼几何中的经典 Ricci 曲率. 这种度量能够更全面地反映网络的全局结构特征. 它不仅仅考虑了局部邻域, 而且在计算节点间的 Wasserstein 距离时, 可以捕捉到网络更广泛区域的特性. 但是, Ollivier-Ricci 曲率的计算复杂度更高, 所以更不适合拓展到大型图上.

3.7 中心性指标

中心性 (centrality) 指标意在找到网络中重要或核心的节点, 此处将介绍多个超图的中心性指标, 其中描述的网络统一定义为一个拥有 M 条超边, N 个节点的超图 $\mathcal{H} = \{V, E\}$.

3.7.1 度中心性

节点的度中心性 (degree centrality) 是最简单直观的中心性指标, 它衡量节点在超图中的连接程度, 度中心性高的节点在超图中与其他节点的连接较多, 可能在信息传播、交互和合作等方面发挥重要作用. 在无权超图中, 节点的度中心性和成对交互作用网络中的度中心性类似, 可被定义为节点所拥有的邻居节点的数量, 可以通过超图的邻接矩阵进行计算 (详见 (1) 式), 也可以用其他的度相关指标进行定义 (详见 3.1 节).

在超图中, 两个相邻节点之间可以存在多条超边, 因此, 可以根据节点之间存在的超边数量定义节点中心性, 此时节点 i 的中心性 $C_d^h(i)$ 就定义为 [86]

$$C_d^h(i) = \sum_{k=1}^N A_{ij}. \quad (27)$$

进一步可以根据两个超节点之间更详细的关系定义超节点之间的权重, 进而定义加权的中心性指标, 在这种条件下, 加权度中心性定义为 [86]

$$C_d^h(i) = \sum_{j=1, j \neq i}^N \sum_{k=1, \{v_i, v_j\} \subset e_k}^M \omega_k, \quad (28)$$

其中, ω_k 是每条超边的权重. 超边的权重可以通过多种方式进行度量, 可以根据结构来定义, 例如, 超边的多重性 (超边在超图中的重复出现次数) 与超边的基 (超边度); 也可根据实际意义或需求来定义超边的权重.

除了依靠节点之间的连接状况, 还可以通过节点之间的连接属性来区分, 社会学将连接分为强连接和弱连接 [87], 其中, 强连接指的是紧密的、亲密的人际关系, 如家人、亲密的朋友, 这种连接关系中的成员通常互动频繁, 有共享的价值观和兴趣, 情感上比较亲近. 而弱连接指的是不太紧密或不太亲密的人际关系; 这种连接关系中的成员可能不经常互动, 共享的兴趣和价值观也可能较少, 但是在社会网络中起着重要的连接作用. 基于上述定义, Kapoor 等 [86] 将强连接度中心性 $C_{\text{sd}}^h(i)$ 和弱连接度中心性 $C_{\text{wd}}^h(i)$ 表示如下:

$$C_{\text{sd}}^h(i) = \sum_{j=1, j \neq i}^N \sum_{k=1, \{v_i, v_j\} \subset e_k}^L I_{w_k > \text{MEAN} + \text{SD}}, \quad (29)$$

$$C_{\text{wd}}^h(i) = \sum_{j=1, j \neq i}^N \sum_{k=1, \{v_i, v_j\} \subset e_k}^L I_{w_k < \text{MEAN} - \text{SD}}, \quad (30)$$

其中, MEAN 和 SD 表示所有权重的均值和标准差.

3.7.2 核心度中心性

核心度中心性 (coreness centrality) 是基于超图的核心分解而定义的指标. 超图的核心分解将超图分解为多个核心, 其中每个核心是一组相互连接的节点和超边. 核心度中心性高的节点位于超图的核心结构中, 起到关键的组织和连接作用. 在成对交互作用网络中, 核心度中心性表示图在 k -核分解 [88] 中分解到的最大层数, 扩展到超图上即在超图的 k -核分解中分解到的最大层数, 此指标决定了节点与超图核的邻近程度.

Mancastropa 等 [61] 提出了超核数的概念, 它是通过计算 k -核分解的扩展算法——超核分解得到的. 由于超边所含节点数不一样, 所以该分解会存在两个参数——超边的数量 k 和超边度 c , 简单来说, (k, c) -超核中的所有节点都至少属于 k 个超边度大于等于 c 的超边. 很容易得出, (k, c) -超核

包含 $(k+1, c)$ -超核和 $(k, c+1)$ -超核. 接着, 他们将节点 i 的 c -壳的索引定义为 $C_c(i) = k$, 使得节点 v_i 属于 (k, c) -超核, 但是不属于 $(k+1, c)$ -超核, 即节点 v_i 连接了 k 个超边度为 c 的超边. 那么 (k, c) -壳 $S(k, c)$ 就可以被定义为所有 $C_c(i) = k$ 的节点集合, 即所有连接了 k 个超边度为 c 的超边的节点集合. 继而可以刻画 k_{cax}^c 为使 $S(k, c)$ 不为空的 k 的最大值. 举例来说, 如图 5 中有 5 个节点和 5 条超边, 粉色的节点属于 2 条超边度为 2 的超边中, 所以它的 $C_2 = 2$. $(2, 2)$ -壳 $S(2, 2)$ 即为绿色和粉色的节点集合, 因为它们都只连接了 2 条超边度为 2 的超边. k_{cax}^2 就是使得 $S(k, 2)$ 不为空的 k 的最大值, 在图 5 中等于 2, 因为 $S(3, 2)$ 为空集, 没有节点连接 3 条超边度为 2 的超边. 因此, 超图的 g -核心度中心性的定义为

$$R(i) = \sum_{c=2}^M g(c) C_c(i) / k_{\text{cax}}^c, \quad (31)$$

其中 M 是超图中的最大超边度, $g(c)$ 是任意权函数, 它可以对高阶相互作用进行加权. 最简单的就是超边没有权重或权重相等, 即 $g(c) = 1$.

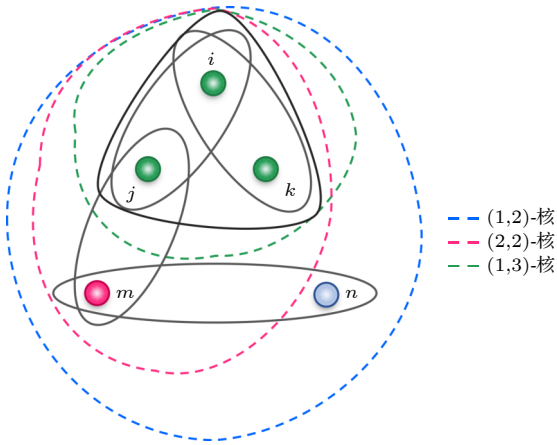


图 5 一个具有 5 条超边的超图的 k -核分解示意图^[61]
Fig. 5. Diagram of a k -core decomposition of a hypergraph with 5 hyperedges^[61].

3.7.3 接近中心性

Aksoy 等^[77] 提出了 s -调和接近中心性 (s -closeness centrality). 首先定义在超图上的游走形式为 s -游走, 并将其定义为一个节点序列, 这个节点序列中的邻接节点同在 s 个超边 (而在普通网络中, 两点之间最多只能有 1 条连边, 所以在普通网络只有 1-游走). 进而, 两条超边 f 和 g 之间的 s -距

离就可以定义为超边之间最短的 s -游走长度, 即:

$$d_s(f, g) = \begin{cases} \text{最短的 } s\text{-游走长度,} & \text{两条超边 } f \text{ 和 } g \text{ 之间存 } s\text{-游走,} \\ \infty, & \text{两条超边 } f \text{ 和 } g \text{ 之间不存 } s\text{-游走.} \end{cases} \quad (32)$$

接下来定义某一条超边 f 的 s -调和接近中心性为

$$C_f = \frac{|E_s| - 1}{\sum_{g \in E_s} d_s(f, g)}, \quad (33)$$

其中 $E_s = \{e \in E : |e| \geq s\}$, 即超边中节点个数大于等于 s 的超边.

3.7.4 介数中心性

介数中心性 (betweenness centrality) 衡量了节点在网络信息传播和通信中的重要性. 节点的介数中心性越高, 意味着该节点在网络中扮演着更重要的桥梁角色, 连接了许多关键路径, 从而使得信息在网络中更易传播和流动. 成对交互作用网络上的介数中心性通过计算节点在最短路径上出现的次数来衡量节点在网络中的重要性. Xiao^[89] 将该成对交互作用网络上的介数中心性概念推广到超图上. 他首先定义了超路径 (hyperpath) 的概念为节点和超边的交错路径, 例如 $v_1, e_1, v_2, e_2, \dots, v_q, e_q, v_{q+1}$ 是一条长度为 q 的超路径, 接着很自然地推广出超图上的介数中心性指节点在最短超路径上出现的次数:

$$B(v_i) = \frac{\sum_{j=1}^n \sum_{k=1}^n \frac{g_k(i)}{g_k}}{N(N-1)/2}, \quad j \neq k \neq i, j < k, \quad (34)$$

其中, g_k 代表从 v_j 到 v_k 的全部最短超路径的数量, 而 $g_k(i)$ 表示经过节点 v_i 的所有最短超路径的数量.

Lee 等^[90] 也扩展了成对交互作用网络的介数中心性到超图中. 依据他们的观点, 一个代表着团队之间联系的 3 阶团, 可以被解释为由顶点对组成的 3 个成对交互关系组. 基于此, 他们提出衡量超边的介数中心性 (h-BC). 为了计算这个指标, 首先需要把超图转换成二部图, 其中上层节点是超边所代表的顶点, 下层节点是原始网络中的节点. 接着, 构造一个最短路径树, 其中节点与超边所表示的节点依次相连. 那么此时超图中节点的介数中心性和

超边的介数中心性都可以依据传统的成对交互网络的介数中心性进行计算。

3.7.5 特征向量中心性

在成对交互作用网络中, 特征向量中心性基于这样的理念: 一个节点的重要性不仅取决于它的邻居节点的数量, 还取决于这些节点的重要性. 这个概念在 1972 年由 Phillip Bonacich 提出 [91], 并在社会网络分析中得到广泛应用. 特征向量中心性考虑了一个节点的邻居的重要性. 它基于这样的假设: 与重要节点相连接的节点也是重要的. 特征向量中心性是网络邻接矩阵的最大特征值 λ 对应的特征向量.

$$E_i = \frac{1}{\lambda} \sum_{t \in \tau(i)} E_t, \quad (35)$$

其中 $\tau(i)$ 为节点 v_i 的邻居节点集合.

而推广到超图中, 目前只有关于均匀超图的特征向量中心性的定义 [92]. 在 d -均匀超图 $d\text{-}\mathcal{H} = \{V, E\}$ 中, 节点 v_i 的特征向量中心性计算如下:

$$E_i = \frac{1}{\lambda} \sum_{\{i, j_1, \dots, j_{d-1}\} \in E} F(E_{j_1}, \dots, E_{j_{d-1}}), \quad (36)$$

其中节点 v_i 和节点 $v_{j_1}, \dots, v_{j_{d-1}}$ 同属于一条超边, F 是任意函数, 一种简单的方法是直接求和, 即 $F(x_1, \dots, x_{k-1}) = x_1 + \dots + x_{k-1}$. 当然也可以选择别的函数, 在这种情况下, 中心性的存在性和唯一性可以通过非线性 Perron-Frobenius 理论来证明 [93].

3.8 熵

在复杂网络中, 可以使用熵 (entropy) 度量网络结构的复杂性或信息量, 本节介绍几种适用于超图的熵. 熵由 Clausius [94] 提出, 后来 Claude Shannon 第一次将熵的概念引入到信息论中, 从而定义了香农熵 (Shannon entropy) [95]. 假设 $p = \{p_1, p_2, \dots, p_n\}$ 是一组概率分布, 并且满足 $0 \leq p_i \leq 1$ 和 $\sum_{i=1}^n p_i = 1$, 那么香农熵就被定义为

$$I(p) = - \sum_{i=1}^n (p_i \log p_i). \quad (37)$$

在超图中, 学者们基于香农熵提出了很多相关的超图熵定义. Bloch 和 Bretto [96] 则是利用关联矩阵 $\mathbf{I}$ 来计算超图的熵, 他们定义了一个新的矩阵

$\mathbf{L} = \mathbf{I}\mathbf{I}^t = ((|e_i \cap e_j|))_{i,j \in \{1 \dots M\}}$. 在这里 Bloch 和 Bretto 只给出了一种特殊超图的熵的计算方法——均匀且每条超边没有交集的超图. 由于每条超边没有交集且超边度一致, 故超图的秩 $r(\mathcal{H})$ (最大的超边度) 恒为一个常数 N/M , 进而 $\mathbf{L}$ 矩阵的迹为所有节点的超度之和 $\text{tr}(\mathbf{L}(\mathcal{H})) = \sum_{x \in V} d(x)$. 所以超图的香农熵可以定义为

$$\begin{aligned} I(\mathcal{H}) &= - \sum_{i=1}^n p_i \log p_i \\ &= - \sum_{e \in E} \frac{|e|}{\text{tr}(\mathbf{L}(\mathcal{H}))} \log \left(\frac{|e|}{\text{tr}(\mathbf{L}(\mathcal{H}))} \right) \\ &= - \frac{mr(\mathcal{H})}{N} \log \left(\frac{r(\mathcal{H})}{N} \right) = - \log \left(\frac{r(\mathcal{H})}{N} \right) \\ &= \log(N) - \log(r(\mathcal{H})) = \log(M). \end{aligned} \quad (38)$$

但是上述方法只能够计算特殊类型的超图, 而 Hu 等 [97] 利用顶点超度数的概率分布来拟合香农熵公式, 从而计算超图熵. 假设超图 $\mathcal{H}$ 中的节点超度序列为 $\{HD_1, HD_2, \dots, HD_n\}$, 那么超图熵则被定义为

$$\begin{aligned} I(\mathcal{H}) &= - \sum_{i=1}^n \left(\frac{HD_i}{\sum_{j=1}^n HD_j} \log \frac{HD_i}{\sum_{j=1}^n HD_j} \right) \\ &= \log \left(\sum_{i=1}^n HD_i \right) - \sum_{i=1}^n \left(\frac{HD_i}{\sum_{j=1}^n HD_j} \log HD_i \right). \end{aligned} \quad (39)$$

Wang 等 [98] 则构建了加权超图的超图熵, 它是通过计算节点的权重差来得到的. 对于加权超图来说, 定义 q_i 为每一个节点的权重离平均权重的距离, 则节点的权重差分布为

$$p_i = \begin{cases} \frac{q_i}{\sum_{j=1}^M q_j}, & q_i > 0, \\ \xi, & \text{otherwise,} \end{cases} \quad (40)$$

其中, ξ 表示比较小的正值, 以避免除到零. 得到权重的差分布之后, 就可以计算超图熵:

$$I(H) = - \sum_{i=1}^M p_i \log p_i. \quad (41)$$

4 基于单纯形网络的统计指标

4.1 指标概览

本节对单纯形网络的重要统计指标进行了综述, 指标概览如表 2 所列, 主要从度相关指标、路径和距离相关指标、中心性指标、聚集系数和拓扑不变量等多个维度展开. 首先, 度相关指标详细地综述了关于单纯形不同的邻接状态的捕捉方式, 并定义了对应度值的计算方法; 关于单纯形网络的路径, 本节介绍了两种游走的方式, 根据定义的路径可以进而计算出距离、离心率和直径等相关指标; 对于中心性指标, 本节也介绍了基于度定义的中心性指标以及结合路径的一些常见的中心指标的扩展; 关于聚集系数, 本节介绍了基于单纯形网络中节点的聚集系数和两种关于单纯形的聚类系数; 最后, 介绍了关于拓扑空间内描述单纯形特征的拓扑不变量在单纯形网络中的相关指标.

4.2 度相关指标

传统成对交互作用网络中, 度衡量的是节点之间的连接状态, 但单纯形网络中存在不同阶数的单纯形, 因此度的定义会相对复杂^[99–102], 本节将介绍单纯形网络中度的定义. 对于一个 0 阶单纯形 (节点), 它与另一个 0 阶单纯形是否邻接取决于二者之间是否存在 1 阶单纯形 (连边). 但是, 对于 1 阶单纯形来说, 它与另一个 1 阶单纯形是否邻接取决于二者是否有共同的 0 阶单纯形 (节点), 或者是否在同一个 2 阶单纯形 (三角形) 之中. 因此, 对于两个 k 阶单纯形 $s_i^{(k)}$ 和 $s_j^{(k)}$, 如果它们有一个公共的 $(k-1)$ 阶单纯形, 则称它们为下邻接, 记作 $s_i^{(k)} \sim_L s_j^{(k)}$; 如果它们是一个 $(k+1)$ 阶单纯形的两个面, 则称它们为上邻接, 记作 $s_i^{(k)} \sim_U s_j^{(k)}$.

定义一个 k 阶单纯形 $s^{(k)}$ 的下邻接度 $\deg_L(s^{(k)})$ 为该单纯形连接的 $(k-1)$ 阶单纯形的数量, 即为 $k+1$; 上邻接度 $\deg_U(s^{(k)})$ 为该单纯形作为一个 k 阶面的 $(k+1)$ 阶单纯形的数量; 基于这两个度的定义, 可以定义一个单纯形的度 $\deg(s^{(k)})$,

$$\begin{aligned}\deg(s^{(k)}) &= \deg_L(s^{(k)}) + \deg_U(s^{(k)}) \\ &= \deg_U(s^{(k)}) + k + 1.\end{aligned}\quad (42)$$

上述关于度的定义只是单纯形网络中最基本的度定义, 由于单纯形网络中的阶数比较复杂, 因此可以进一步定义一些关于不同阶数的度. 在上邻接和下邻接的基础上, 可以更加细致地定义网络中的任意一阶邻接关系, p 邻接. 主要分为上 p 邻接和下 p 邻接, 分别用 U_p 和 L_p 表示. 如果两个单纯形 $\sigma^{(k)} \sim_{U_p} \sigma^{(k')}$, 则一定存在一个 p 阶单纯形 $\tau^{(p)}$, 并且同时满足 $\sigma^{(k)} \subseteq \tau^{(p)}$ 和 $\sigma^{(k')} \subseteq \tau^{(p)}$. 如果两个单纯形 $\sigma^{(k)} \sim_{L_p} \sigma^{(k')}$, 则一定存在一个 p 阶单纯形 $\tau^{(p)}$, 并且同时满足 $\tau^{(p)} \subseteq \sigma^{(k)}$ 和 $\tau^{(p)} \subseteq \sigma^{(k')}$. 进而还可以定义两个单纯形的严格上 p 邻接和严格下 p 邻接, 分别用 U_{p^*} 和 L_{p^*} 来表示. 如果两个单纯形 $\sigma^{(k)} \sim_{L_{p^*}} \sigma^{(k')}$, 需要同时满足 $\sigma^{(k)} \sim_{L_p} \sigma^{(k')}$ 和 $\sigma^{(k)} \not\sim_{L_{p+1}} \sigma^{(k')}$. 如果两个单纯形 $\sigma^{(k)} \sim_{U_{p^*}} \sigma^{(k')}$, 需要同时满足 $\sigma^{(k)} \sim_{U_p} \sigma^{(k')}$ 和 $\sigma^{(k)} \not\sim_{U_{p+1}} \sigma^{(k')}$. 基于以上的邻接关系, 可以定义一个单纯形 $\sigma^{(k)}$ 的上 p 邻接度 ($\deg_U^p$)、下 p 邻接度 ($\deg_L^p$)、严格上 p 邻接度 ($\deg_U^{p^*}$) 和严格下 p 邻接度 ($\deg_L^{p^*}$):

$$\deg_L^p(\sigma^{(k)}) = \#\{\sigma^{(k')} : \sigma^{(k')} \sim_{L_p} \sigma^{(k)}\}, \quad (43)$$

$$\deg_U^p(\sigma^{(k)}) = \#\{\sigma^{(k')} : \sigma^{(k')} \sim_{U_p} \sigma^{(k)}\}, \quad (44)$$

$$\deg_L^{p^*}(\sigma^{(k)}) = \#\{\sigma^{(k')} : \sigma^{(k')} \sim_{L_{p^*}} \sigma^{(k)}\}, \quad (45)$$

$$\deg_U^{p^*}(\sigma^{(k)}) = \#\{\sigma^{(k')} : \sigma^{(k')} \sim_{U_{p^*}} \sigma^{(k)}\}. \quad (46)$$

上述定义中, 上邻接和下邻接度计算的是所有

表 2 基于单纯形网络的统计指标总结

Table 2. Summary of statistical indicators of the simplicial network.

指标类型	指标名称
度相关指标	上邻接度、下邻接度、度、上 p 邻接度、下 p 邻接度、严格上 p 邻接度、严格下 p 邻接度、上 (h, p) 邻接度、下 (h, p) 邻接度、 p 邻接度、最大 p 邻接度、最大单纯形度
路径和距离相关指标	s_k 游走、 p 游走、最短路径长度、离心率、直径
中心性指标	度中心性、特征向量中心性、Katz 中心性、接近中心性、介数中心性
聚集系数	聚集系数
拓扑不变量	贝蒂数、欧拉示性数

与 k 阶单纯形 $\sigma^{(k)}$ 上 p 邻接和下 p 邻接 k' 阶单纯形 $\sigma^{(k')}$ 的数量, 也可以在此基础上计算规定阶数范围的单纯形度, 定义 k 阶单纯形 $\sigma^{(k)}$ 的 (h, p) 度, 可以分为上 (h, p) 邻接度和下 (h, p) 邻接度, 记作 $\deg_L^{h,p}$, $\deg_U^{h,p}$, 同时也可以计算其对应的严格约束形式的度, 记作 $\deg_L^{h,p,*}$, $\deg_U^{h,p,*}$:

$$\deg_L^{h,p}(\sigma^{(k)}) = \#\{\sigma^{(k-h)} : \sigma^{(k-h)} \sim_{L_p} \sigma^{(k)}\}. \quad (47)$$

$$\deg_U^{h,p}(\sigma^{(k)}) = \#\{\sigma^{(k+h)} : \sigma^{(k)} \sim_{U_p} \sigma^{(k+h)}\}, \quad (48)$$

$$\deg_L^{h,p,*}(\sigma^{(k)}) = \#\{\sigma^{(k-h)} : \sigma^{(k-h)} \sim_{L_p} \sigma^{(k)} \ \& \ \sigma^{(k-h)} \not\sim_{L_{p+1}} \sigma^{(k)}\}, \quad (49)$$

$$\deg_U^{h,p,*}(\sigma^{(k)}) = \#\{\sigma^{(k+h)} : \sigma^{(k)} \sim_{U_p} \sigma^{(k+h)} \ \& \ \sigma^{(k)} \not\sim_{U_{p+1}} \sigma^{(k+h)}\}. \quad (50)$$

进一步地, 可以定义单纯形网络中更一般化的邻接关系及度—— p 邻接, 记作 A_p . 如果两个单纯形 $\sigma^{(k)} \sim_{A_p} \sigma^{(k')}$, 需要同时满足 $\sigma^{(k)} \sim_{L_p} \sigma^{(k')}$ 和 $\sigma^{(k)} \not\sim_{U_{p'}} \sigma^{(k')}$, 这里 $p' = k + k' - p$. 为了和传统的成对交互作用关系图对应, 当 $q = 0$ 时, 如果两个节点 $v_i \sim_{U_1} v_j$, 则 v_i 和 v_j 邻接. 可以基于 p 邻接定义最大 p 邻接, 记作 A_p^* . 如果两个单纯形 $\sigma^{(k)} \sim_{A_p^*} \sigma^{(k')}$, 除了满足 p 邻接的条件外, 还要满足对于任意一个和单纯形 $\sigma^{(k)}$ p 邻接的单纯形 $\sigma^{(k'')}$ 满足 $\sigma^{(k')} \not\sim \sigma^{(k'')}$. 基于此, 定义一个 k 阶单纯形 $\sigma^{(k)}$ 的 p 邻接度 $\deg_A^p$ 和最大 p 邻接度 $\deg_A^{p,*}$:

$$\deg_A^p(\sigma^{(k)}) = \#\{\sigma^{(k')} | \sigma^{(k)} \sim_{A_p} \sigma^{(k')}\}, \quad (51)$$

$$\deg_A^{p,*}(\sigma^{(k)}) = \#\{\sigma^{(k')} | \sigma^{(k)} \sim_{A_p^*} \sigma^{(k')}\}. \quad (52)$$

最后, 可以在单纯形网络中定义一个无参数的度指标, 即最大单纯形度, 记作 $\deg^*(\sigma^{(k)})$:

$$\deg^*(\sigma^{(k)}) = \deg_A^*(\sigma^{(k)}) + \deg_U^*(\sigma^{(k)}), \quad (53)$$

其中, $\deg_A^*(\sigma^{(k)}) = \sum_{p=0}^{k-1} \deg_A^{p,*}(\sigma^{(k)})$, $\deg_U^*(\sigma^{(k)}) = \sum_{h=1}^{\dim K - k} \deg_U^{h,(k+h)*}(\sigma^{(k)})$, $\dim K$ 为单纯形网络 K 的维度.

以图 2(b) 单纯形网络中的 1 阶单纯形 (1, 5) 为例, 记为 $\sigma^{(1)}$, 该单纯形网络的维度 $\dim K = 3$, $k = 1$, $p = 0$, $h = 1, 2$. 因此 $\deg^*(\sigma^{(1)}) = \deg_A^{0,*}(\sigma^{(1)}) + \deg_U^{1,2*}(\sigma^{(1)}) + \deg_U^{2,3*}(\sigma^{(1)}) = 4 + 1 + 0$. $\deg_A^{0,*}(\sigma^{(1)})$ 表示与 $\sigma^{(1)}$ 严格下 0 邻近但不是上邻接的单纯形 $\sigma^{(k')}$ 的数量, 有 4 个单纯形 (1, 2), (5, 6), (5, 9),

(5, 10); $\deg_U^{1,2*}(\sigma^{(1)})$ 代表的是与 $\sigma^{(1)}$ 严格上 2 邻接的 2 阶单纯形 (三角形) $\sigma^{(2)}$ 的数量, 有 1 个单纯形 (1, 4, 5); $\deg_U^{2,3*}(\sigma^{(1)})$ 表示与 $\sigma^{(1)}$ 严格上 3 邻接的 3 阶单纯形 (四面体) $\sigma^{(3)}$ 的数量, 这里没有符合要求的 3 阶单纯形. 更详细的关于高阶度的计算方法可以查询文献 [102].

4.3 路径和距离相关指标

计算网络中的最短路径, 首先从定义单纯形网络中的游走 (walk) 开始. 在单纯形网络中的研究主体不再只是节点, 而变成了单纯形, 本节主要介绍单纯形直接的游走和其相关统计指标 [100,102].

4.3.1 s_k 游走

Estrada 和 Ross [100] 定义了一种 s_k 游走, 用来描述 k 阶单纯形之间的游走, 其主要特点是两个同阶单纯形只能通过低一阶的单纯形进行游走. 具体来讲, s_k 游走在 $\sigma_1^{(k)}$ 和 $\sigma_r^{(k)}$ 两个 k 阶单纯形之间定义了一个序列 $\{\sigma_1^{(k)}, \sigma_1^{(k-1)}, \sigma_2^{(k)}, \sigma_2^{(k-1)}, \dots, \sigma_{r-1}^{(k-1)}, \sigma_r^{(k)}\}$, $\sigma_1^{(k)}$ 和 $\sigma_r^{(k)}$ 两个单纯形不存在上邻接关系. 在这里 k 阶单纯形之间的游走只能依赖于 $(k-1)$ 阶单纯形, 根据单纯形网络中的定义可知, 这个 $(k-1)$ 单纯形是近邻的两个 k 阶单纯形的面.

4.3.2 p 游走

文献 [86] 定义了一个相对 s_k 游走来说更加广义的游走序列—— p 游走, $\{\sigma_1^{(k_1)}, \sigma_1^{(p_1)}, \sigma_2^{(k_2)}, \sigma_2^{(p_2)}, \dots, \sigma_r^{(p_r)}, \sigma_{r+1}^{(k_{r+1})}\}$, 这里要求 $\sigma_i^{(k_i)} \sim_{A_{p_i}} \sigma_{i+1}^{(k_{i+1})}$, p 取 $\{p_1, p_2, \dots, p_r\}$ 中的最小值, r 为游走的长度, 如果两个单纯形之间有一条 p 游走, 则称这两个单纯形最大 p 连通. 当 $k_i = k$, $p_i = k-1$ 时, p 游走即为 s_k 游走. 也就是说, p 游走是一个更一般化的游走, s_k 游走是 p 游走的一种特殊形式, 因此, 下文介绍的基于路径的一系列统计指标均基于 p 游走, 如有需要, 通过设置参数就可转化为 s_k 游走.

4.3.3 最短路径长度

接下来可以基于游走定义单纯形网络中的最短路径, 对于两个任意阶数的单纯形 $\sigma_a^{(k)}$ 和 $\sigma_b^{(k)}$, 两者之间 p 游走的路径长度最小时, 该路径为最短路径, 其长度为两个单纯形之间的最短路径长度, 记为 $d_p(\sigma^{(k)}, \sigma^{(k')})$.

4.3.4 离心率

度量单纯形的离心率可以从该单纯形与其他单纯形的距离来考虑, 通常把该单纯形到其余单纯形的最短路径长度计算出来, 选取其中最大值来表示该单纯形的离心率, 下面介绍基于最短路径定义的离心率. 基于 p 游走, 可以定义一个 k 阶单纯形 $\sigma^{(k)}$ 的 p 离心率:

$$\epsilon_p(\sigma_a^{(k)}) = \max_{\sigma^{(k')} \in K_p} d_p(\sigma_a^{(k)}, \sigma^{(k')}). \quad (54)$$

4.3.5 直径

网络的直径 (diameter) 是一个描述全局特征的统计指标, 所有节点的离心率的最大值即为直径. 直径也可以理解为网络中所有单纯形两两之间的最短路径长度的最大值:

$$D_p = \max \epsilon_p(\sigma_a^{(k)}). \quad (55)$$

4.4 中心性指标

4.4.1 度中心性

通常, 定义度相关的中心性的思想是衡量节点与其他节点相连的程度. 最简单的方法是在传统成对交互作用网络中, 计算节点的度 (即与其相连的边的数量), 然后将其除以除了自身节点以外的所有节点数量. 可以遵循此中心思想和单纯形网络中度的定义, 来定义一系列度中心性指标^[102].

对于一个节点来说, 在单纯形网络里面是 0 阶单纯形, 是一个阶数最低的单纯形. 因此, 定义其度中心性只利用上邻接度即可, 记作 $C_{\deg_U^{h,h}}(v)$,

$$C_{\deg_U^{h,h}}(v) = \frac{\deg_U^{h,h}(v)}{\binom{f_0 - 1}{h}}, \quad (56)$$

其中, f_0 表示 0 阶单纯形的数量, 即节点数量. 可以发现, 当 $h = 1$ 时, 该统计指标是成对交互作用网络定义的度中心性.

除了节点的度中心性指标, 也可以在单纯形网络中进一步定义关于一个 k 阶单纯形的度中心性, 由于单纯形网络中有多种度定义, 本文介绍一种最常用的度中心性^[102],

$$C_{\deg^*}(\sigma^{(k)}) = \frac{\deg^*(\sigma^{(k)})}{\sum_{i=0}^{\dim K} f_i - 1}, \quad (57)$$

$\deg^*(\sigma^{(k)})$ 表示 k 阶单纯形 $\sigma^{(k)}$ 的最大单纯形度

(详见 (53) 式), f_i 表示第 i 阶单纯形的数量, $\dim K$ 表示单纯复形 K 的维数.

4.4.2 特征向量中心性

特征向量中心性捕捉的是一个节点的邻居节点的重要性^[103], 即如果一个节点的邻居节点是重要节点, 那么其自身也可能非常重要, 可以看作是度中心性的进一步考量. 计算特征向量中心性需要先构造邻接矩阵, 本文介绍的是基于 k 阶单纯形的特征向量中心性, 所以, 首先给出的是 k 阶单纯形的邻接矩阵 $A(k)_{ij}$ ^[100],

$$A(k)_{ij} = \begin{cases} 1, & \text{if } \sigma_i^{(k)} \sim_L \sigma_j^{(k)} \text{ and } \sigma_i^{(k)} \not\sim_U \sigma_j^{(k)} \\ 0, & \text{if } i = j \text{ or } \sigma_i^{(k)} \not\sim_L \sigma_j^{(k)} \text{ or } \sigma_i^{(k)} \sim_U \sigma_j^{(k)}. \end{cases} \quad (58)$$

第 i 个 k 阶单纯形 $\sigma_i^{(k)}$ 的特征向量中心性为矩阵 $A(k)_{ij}$ 主特征向量 (最大特征值对应的特征向量) 的第 i 个分量对应的值.

4.4.3 Katz 中心性

Katz 中心性意在解决特征向量中心性在有向网络中的局限性^[104], 主要思想是首先给每个节点赋予相同的中心性, 一般取值为 1, 这样就解决了只有出度没有入度的节点中心性为 0 的局限, 在这里第 i 个 k 阶单纯形的 Katz 中心性可以表示为^[100,105]

$$K_{k,i} = \left[\sum_{m=0}^{\infty} (\alpha^m \mathbf{A}_k^m) \mathbf{e} \right]_i, \quad (59)$$

这里 $\mathbf{A}_k$ 表示 k 阶单纯形的邻接矩阵; m 表示计算的迭代次数; $\mathbf{e}$ 表示单位向量; α 是一个常数, 其取值范围在 $(0, 1/\lambda_1)$, λ_1 为矩阵 $\mathbf{A}_k$ 的最大特征值. 当 $\alpha \leq \rho(\mathbf{A}_k)$ 时 $(\alpha^m \mathbf{A}_k^m)$ 可以收敛, 所以有

$$K_{k,i} = [(\mathbf{I} - \alpha \mathbf{A}_k)^{-1} \mathbf{e}]_i, \quad (60)$$

其中, $\mathbf{I}$ 表示单位矩阵, 进而根据文献^[105]可以推出用特征值和特征向量表示的 Katz 中心性,

$$K_{k,i} = \sum_l \sum_j \psi_{k,j}(i) \psi_{k,j}(l) \frac{1}{1 - \alpha \lambda_{k,j}}, \quad (61)$$

在这里 $\psi_{k,j}(i)$ 和 $\psi_{k,j}(l)$ 分别表示矩阵 $\mathbf{A}_k$ 的第 j 个向量的第 i 和 l 个分量, $\lambda_{k,j}$ 为该向量对应的特征值.

4.4.4 接近中心性

接近中心性是衡量网络中两个节点接近程度的物理量^[106], 在单纯形网络中可以用来衡量两个

单纯形之间的接近程度^[102]. 如果一个单纯形 $\sigma_i^{(k_i)} \in K_p$, 则有

$$C_{CL,p}(\sigma_i^{(k_i)}) := \frac{1}{\sum_{i \neq j} d_p(\sigma_i^{(k_i)}, \sigma_j^{(k_j)})}, \quad (62)$$

其中, K_p 表示由阶数大于等于 p 的单纯形组成的集合, 分母表示其他单纯形到 $\sigma_i^{(k_i)}$ 的距离之和.

4.4.5 介数中心性

介数中心性计算节点经过的最短路径的数量, 可以衡量一个节点起桥梁作用的程度^[107]. 在单纯形网络中, 如果一个单纯形满足 $\sigma_h^{(k_h)} \in \bar{K}_{p,l}^* \subseteq \bar{K}_p^*$, 则其介数中心性被定义为^[102]

$$C_{B,p}(\sigma_h^{(k_h)}) = \frac{2}{(Q_{p,l}^* - 1)(Q_{p,l}^* - 2)} \sum_{i \neq k \neq j} \frac{l_{ij,p}(\sigma_h^{(k_h)})}{l_{ij,p}}, \quad (63)$$

其中, $Q_{p,l}^*$ 表示第 l 个最大 p 连通分量 $\text{bar}K_{p,l}^*$ 中的单纯形数量, $l_{ij,p}$ 表示两个单纯形 $\sigma_i^{(k_i)}$ 和 $\sigma_j^{(k_j)}$ 之间最短 p 游走的总数, $l_{ij,p}(\sigma_h^{(k_h)})$ 表示两个单纯形 $\sigma_i^{(k_i)}$ 和 $\sigma_j^{(k_j)}$ 之间最短 p 游走中经过单纯形 $\sigma_h^{(k_h)}$ 的数量.

4.5 聚集系数

节点的聚集系数可以反映该节点的邻居节点之间的聚集程度, 是一个常用的局部统计指标. 在单纯形网络中, Maletić 等^[108] 定义了关于单纯形的聚集系数 $C'_{\sigma_i^k}$:

$$C'_{\sigma_i^k} = \sum_{\sigma_j^{k'} \in K_{nn}^i} \frac{\text{两个单纯形 } \sigma_i^k \text{ 和 } \sigma_j^{k'} \text{ 共享的面的数量}}{(1/2)(\text{两个单纯形 } \sigma_i^k \text{ 和 } \sigma_j^{k'} \text{ 的面的总数})}, \quad (64)$$

其中 K_{nn}^i 表示单纯形 σ_i^k 的邻居单纯形的集合.

Serrano 和 Gómez^[102] 在此基础上引进了最大 p 邻接度和严格上邻接度用来定义了基于节点的聚集系数 $C_s(v)$ 和基于单纯形的聚集系数 $C_s(\sigma^{(k)})$:

$$C_s(v) = \frac{v \text{ 的最大邻居单纯形之间的连边数量}}{\frac{\deg_U^*(V)(\deg_U^*(V) - 1)}{2}}, \quad (65)$$

$$C_s(\sigma^{(k)}) = \frac{\sigma^{(k)} \text{ 的最大邻居单纯形之间的连边数量}}{\frac{\deg^*(\sigma^{(k)})(\deg^*(\sigma^{(k)}) - 1)}{2}}. \quad (66)$$

4.6 拓扑不变量

4.6.1 贝蒂数

贝蒂数 (Betti number) 是代数拓扑中的一个重要概念^[22,109], 用于描述拓扑空间的拓扑不变量, 是拓扑数据分析中常用的统计指标. 贝蒂数表示单纯形网络中线性独立的 k ($k \geq 0$) 阶洞的数量, 记为 β_k , 其中, β_0 表示单纯形网络的连通分量数, β_1 表示网络中 1-洞的数量, β_2 表示网络中由团包围而成的 2-洞的数量, 等等. 通过计算贝蒂数 β_k , 可以了解网络的拓扑特征, 如它的空间结构是否连通, 是否有闭合的洞等, 计算公式如下:

$$\beta_k = m_k - r_k - r_{k+1}, \quad (67)$$

其中, m_k 为网络中 k 阶团的数量, r_k 为网络的 k 阶边界矩阵 B_k 的秩. 边界矩阵用来描述拓扑空间中单纯复形的边界信息, B_k 表示 $(k-1)$ 阶单纯形和 k 阶单纯形的关联, 如果 $(k-1)$ 阶单纯形是 k 阶单纯形的子集, 则矩阵对应位置的元素为 1, 否则为 0. $k=1$ 时, 边界矩阵 B_1 即网络的关联矩阵. 边界矩阵的秩可以用高斯消去法计算, 首先将边界矩阵化为行阶梯形矩阵, 行阶梯形矩阵中非零行的数量就是边界矩阵 B_k 的秩 r_k .

4.6.2 欧拉示性数

欧拉示性数 (Euler characteristic) 是代数拓扑中另一个重要的拓扑不变量^[22], 用于描述拓扑空间的特征. 在三维情况下, 欧拉示性数等于网络的节点数减去边数, 再加上三角形数量. 对于更高维的单纯复形, 欧拉示性数可以定义为交错和:

$$\chi = \sum_{k \geq 0} (-1)^k m_k, \quad (68)$$

其中, m_k 表示 k 阶团的数量. 通过欧拉-庞加莱公式, 可以将欧拉示性数和贝蒂数联系起来, 欧拉示性数等于贝蒂数的交错和, 即

$$\chi = \sum_{k \geq 0} (-1)^k \beta_k. \quad (69)$$

5 总结

本文详细阐述了超图和单纯形网络这两类高阶网络的基本概念, 并系统综述了在这两类网络中常用的统计指标及其物理意义. 高阶网络的统计指标具有广泛的应用场景. 它们能够更准确地量化描

述各类复杂系统的结构信息,进而揭示其内在的演化规律. 具体而言,在超图的实证分析中,统计指标的应用尤为重要. 由于超图中超边的基通常存在较大差异,如何更细致地考虑结构因素来定义或拓展相关统计指标,成为未来研究的一个重要方向. 同时,如何在大规模超图中实现这些统计指标的高效计算,也是亟待解决的问题.

单纯形网络的实证分析目前主要集中在网络的高阶度分布和高阶连通性研究上. 未来,可以基于代数拓扑开发更多统计指标,或提出新的基于代数拓扑理论的方法以深入分析单纯形网络的结构特性. 在网络信息挖掘方面,高阶统计指标如高阶距离的定义和高阶度的概念,可以进一步用于定义节点的中心性指标,从而评估网络中的关键节点. 这不仅可以用于超边或单纯形的预测,还可以利用这些统计特性进行高阶社团检测、图分类等任务. 在网络建模方面,可以构建保持网络特定统计特性的配置模型和零模型,或是根据演化特征构建生长模型. 目前面向超图建模的研究相对较多,基于单纯形结构的模型开发是一个值得探索的领域.

此外,高阶网络的结构信息与动力学关系的研究也具有重要意义. 利用高阶网络的结构信息或结构相关的模型,有助于我们更深入地理解高阶动力学. 例如,基于统计指标构造的中心性指标可用于研究高阶网络中的传播能力,而高阶网络的结构相关模型则可用于高阶同步优化等动力学过程的研究. 本文仅介绍了基于无向无权高阶网络的统计指标,因此,研究包含权重、有向和时序特征的高阶网络,将是一个极具挑战性的研究方向,对于这些网络的研究也有望提供更丰富、更深入的网络分析手段.

感谢中国科学技术大学范天龙博士对本文的讨论.

参考文献

- [1] Marin A, Wellman B 2011 *Social network analysis: An introduction* (London: SAGE publications) pp11–25
- [2] Kossinets G, Watts D J 2006 *Science* **311** 88
- [3] Alon U 2003 *Science* **301** 1866
- [4] Alm E, Arkin A P 2003 *Curr. Opin. Struct. Biol.* **13** 193
- [5] Bose A, Clements K A 1987 *Proc. IEEE* **75** 1607
- [6] Wu F F, Varaiya P 1999 *Int. J. Electr. Power Energy Syst.* **21** 75
- [7] Williams J C, Mahmassani H S, Herman R 1987 *Transp. Res. Rec.* **1112** 78

- [8] Verma T, Araujo N A, Herrmann H J 2014 *Sci. Rep.* **4** 5638
- [9] Strogatz S H 2001 *Nature* **410** 268
- [10] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D U 2006 *Phys. Rep.* **424** 175
- [11] Costa L D F, Rodrigues F A, Traverso G, Villas Boas P R 2007 *Adv. Phys.* **56** 167
- [12] Barabási A L 2013 *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **371** 20120375
- [13] Wang X F, Li X, Chen G R 2012 *Network Science: An Introduction* (Higher Education Press) p82 (in Chinese) [汪小帆, 李翔, 陈关荣 2012 网络科学导论 (高等教育出版社) 第82页]
- [14] Zhou T, Bai W J, Wang B H, Liu Z J, Yan G 2005 *Physics* **34** 31 (in Chinese) [周涛, 柏文洁, 汪秉宏, 刘之景, 严钢 2005 物理 **34** 31]
- [15] Courtney O T, Bianconi G 2017 *Phys. Rev. E* **95** 062301
- [16] Lung R I, Gaskó N, Suciu M A 2018 *Scientometrics* **117** 1361
- [17] Pearcy N, Crofts J J, Chuzhanova N 2014 *Int. J. Biol. Vet. Agric. Food Eng.* **8** 752
- [18] Mastrandrea R, Fournet J, Barrat A 2015 *PloS One* **10** e0136497
- [19] Stehlé J, Voirin N, Barrat A, et al. 2011 *PloS One* **6** e23176
- [20] Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J G, Petri G 2020 *Phys. Rep.* **874** 1
- [21] Battiston F, Amico E, Barrat A, et al. 2021 *Nat. Phys.* **17** 1093
- [22] Bianconi G 2021 *Higher-order Networks* (Cambridge: Cambridge University Press) pp7–45
- [23] Shi D, Chen G 2022 *Natl. Sci. Rev.* **9** nwac038
- [24] Zhao D, Li R, Peng H, Zhong M, Wang W 2022 *Chaos Solit. Fractals* **155** 111701
- [25] Wang W, Li W, Lin T, Wu T, Pan L, Liu Y 2022 *Appl. Math. Comput.* **420** 126793
- [26] Millán A P, Torres J J, Bianconi G 2020 *Phys. Rev. Lett.* **124** 218301
- [27] Lucas M, Cencetti G, Battiston F 2020 *Phys. Rev. Res.* **2** 033410
- [28] Iacopini I, Petri G, Barrat A, Latora V 2019 *Nat. Commun.* **10** 1
- [29] Chowdhary S, Kumar A, Cencetti G, Iacopini I, Battiston F 2021 *J. Phys.: Complex.* **2** 035019
- [30] Chen H Y, Xu T, Liu C, Zhang Z K, Zhan X X 2024 *Acta Phys. Sin.* **73** 038901 (in Chinese) [陈浩宇, 徐涛, 刘闯, 张子柯, 詹秀秀 2024 物理学报 **73** 038901]
- [31] Gómez-Gardenes J, Gómez S, Arenas A, Moreno Y 2011 *Phys. Rev. Lett.* **106** 128701
- [32] Kovalenko K, Dai X, Alfaro-Bittner K, Raigorodskii A, Perc M, Boccaletti S 2021 *Phys. Rev. Lett.* **127** 258301
- [33] Tanaka T, Aoyagi T 2011 *Phys. Rev. Lett.* **106** 224101
- [34] Zhang Y, Latora V, Motter A E 2021 *Commun. Phys.* **4** 195
- [35] Kundu S, Ghosh D 2022 *Phys. Rev. E* **105** L042202
- [36] Bick C, Ashwin P, Rodrigues A 2016 *Chaos* **26** 094814
- [37] Wang W, Wang Z X, Cai S M 2018 *Phys. Rev. E* **98** 052312
- [38] Guilbeault D, Becker J, Centola D 2018 *Complex Spreading Phenomena in Social Systems* (Cham: Springer) pp3–25
- [39] Wang W, Liu Q H, Liang J, Hu Y, Zhou T 2019 *Phys. Rep.* **820** 1
- [40] Wang D, Zhao Y, Luo J, Leng H 2021 *Chaos: Interdiscip. J. Nonlinear Sci.* **31** 053112
- [41] Wang Z H, Shen H W, Cao Q, Cheng X Q 2011 *J. Softw.* **33** 171 (in Chinese) [王兆慧, 沈华伟, 曹婧, 程学旗 2011 软件学报 **33** 171]

- [42] Lü L, Chen D, Ren X L, Zhang Q M, Zhang Y C, Zhou T 2016 *Phys. Rep.* **650** 1
- [43] Ren X L, Lü L Y 2014 *Chin. Sci. Bull.* **59** 1175 (in Chinese) [任晓龙, 吕琳媛 2014 科学通报 **59** 1175]
- [44] Li J, Liu Y, Wang W, Zhou T 2024 *Acta Phys. Sin.* **73** 048901 (in Chinese) [李江, 刘影, 王伟, 周涛 2024 物理学报 **73** 048901]
- [45] Lü L, Zhou T 2011 *Phys. A: Stat. Mech. Appl.* **390** 1150
- [46] Liu B, Yang R, Lü L 2023 *Chaos: Interdiscip. J. Nonlinear Sci.* **33** 083108
- [47] Lü L Y 2010 *J. Univ. Electron. Sci. Technol. China* **39** 651 (in Chinese) [吕琳媛 2010 电子科技大学学报 **39** 651]
- [48] Newman M E 2006 *Proc. Natl. Acad. Sci.* **103** 8577
- [49] Jiang Y, Jia C, Yu J 2013 *Phys. A: Stat. Mech. Appl.* **392** 2182
- [50] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [51] Barabási A L, Albert R 1999 *Science* **286** 509
- [52] Xu X K, Cui W K, Cui L Y, Xiao J, Shang K K 2019 *J. Univ. Electron. Sci. Technol. China* **48** 122 (in Chinese) [许小可, 崔文阔, 崔丽艳, 肖婧, 尚可 2019 电子科技大学学报 **48** 122]
- [53] Zeng Y, Liu B, Zhou F, Lü L 2023 *Entropy* **25** 1390
- [54] Bick C, Gross E, Harrington H A, Schaub M T 2023 *SIAM Rev.* **65** 686
- [55] Feng Y, You H, Zhang Z, Ji R, Gao Y 2019 *Proceedings of the AAAI Conference on Artificial Intelligence* **33** 3558
- [56] Zhu J, Zhu J, Ghosh S, Wu W, Yuan J 2018 *IEEE Trans. Netw. Sci. Eng.* **6** 801
- [57] Viñas R, Joshi C K, Georgiev D, Lin P, Dumitrescu B, Gamazon E R, Liò P 2023 *Nat. Mach. Intell.* **5** 739
- [58] Huang J, Zhang S, Yang F, Yu T, Prasad L N, Guduri M, Yu K 2023 *IEEE Trans. Consum. Electron.* **1** 1775
- [59] Ruggeri N, Contisciani M, Battiston F, De Bacco C 2023 *Sci. Adv.* **9** eadg9159
- [60] Wu H, Li N, Zhang J, Chen S, Ng M K, Long J 2024 *Pattern Recognit.* **146** 109995
- [61] Mancastroppa M, Iacopini I, Petri G, Barrat A 2023 *Nat. Commun.* **14** 6223
- [62] Gao Y, Feng Y, Ji S, Ji R 2022 *IEEE Trans. Pattern Anal. Mach. Intell.* **45** 3181
- [63] Li Z, Deng Z, Han Z, Alfaro-Bittner K, Barzel B, Boccaletti S 2021 *Chaos Solit. Fractals* **152** 111307
- [64] Gambuzza L V, Di Patti F, Gallo L, et al. 2021 *Nat. Commun.* **12** 1255
- [65] Wang H, Ma C, Chen H S, Lai Y C, Zhang H F 2022 *Nat. Commun.* **13** 3043
- [66] Benson A R, Abebe R, Schaub M T, Jadbabaie A, Kleinberg J 2018 *Proc. Natl. Acad. Sci.* **115** E11221
- [67] Shi D, Chen Z, Sun X, Chen Q, Ma C, Lou Y, Chen G 2021 *Commun. Phys.* **4** 249
- [68] Reimann M W, Nolte M, Scolamiero M, et al. 2017 *Front. Comput. Neurosci.* **11** 48
- [69] Sizemore A E, Giusti C, Kahn A, Vettel J M, Betzel R F, Bassett D S 2018 *J. Comput. Neurosci.* **44** 115
- [70] Kovalenko K, Sendiña-Nadal I, Khalil N, et al. 2021 *Commun. Phys.* **4** 43
- [71] Holland P W, Leinhardt S 1971 *Comp. Group Stud.* **2** 107
- [72] Estrada E, Rodríguez-Velázquez J A 2006 *Phys. A: Stat. Mech. Appl.* **364** 581
- [73] Carletti T, Battiston F, Cencetti G, Fanelli D 2020 *Phys. Rev. E* **101** 022308
- [74] Carletti T, Fanelli D, Lambiotte R 2021 *J. Phys.: Complex.* **2** 015011
- [75] Aksoy S G, Joslyn C, Marrero C O, Praggastis B, Purvine E 2020 *EPJ Data Sci.* **9** 16
- [76] Lu L, Peng X 2013 *Internet Math.* **9** 3
- [77] Vasilyeva E, Romance M, Samoylenko I, Kovalenko K, Musatov D, Raigorodskii A M, Boccaletti S 2023 *Entropy* **25** 923
- [78] Gao J, Zhao Q, Ren W, Swami A, Ramanathan R, Bar-Noy A 2014 *IEEE/ACM Trans. Netw.* **23** 1805
- [79] Zlatić V, Ghoshal G, Caldarelli G 2009 *Phys. Rev. E* **80** 036118
- [80] Bauer F, Hua B, Jost J, Liu S, Wang G 2017 *Modern Approaches to Discrete Curvature* (Cham: Springer) pp1–62
- [81] Samal A, Sreejith R, Gu J, Liu S, Saucan E, Jost J 2018 *Sci. Rep.* **8** 8650
- [82] Leal W, Restrepo G, Stadler P F, Jost J 2021 *Adv. Complex Syst.* **24** 2150003
- [83] Eidi M, Farzam A, Leal W, Samal A, Jost J 2020 *Theory Biosci.* **139** 337
- [84] Bauer F, Jost J, Liu S 2012 *Math. Res. Lett.* **19** 1185
- [85] Eidi M, Jost J 2020 *Sci. Rep.* **10** 12466
- [86] Kapoor K, Sharma D, Srivastava J 2013 *IEEE 2nd Network Science Workshop* New York, USA, April 29–May 1, 2013 p152
- [87] Granovetter M S 1973 *Am. J. Sociol.* **78** 1360
- [88] Dorogovtsev S N, Goltsev A V, Mendes J F F 2006 *Phys. Rev. Lett.* **96** 040601
- [89] Xiao Q 2013 *Res. J. Appl. Sci. Eng. Technol.* **5** 568
- [90] Lee J, Lee Y, Oh S M, Kahng B 2021 *Chaos: Interdiscip. J. Nonlinear Sci.* **31** 061108
- [91] Bonacich P 1972 *J. Math. Sociol.* **2** 113
- [92] Benson A R 2019 *SIAM J. Math. Data Sci.* **1** 293
- [93] Lemmens B, Nussbaum R 2012 *Nonlinear Perron-frobenius Theory* (Vol. 189) (Cambridge: Cambridge University Press) pp2–4
- [94] Clausius R 1879 *The Mechanical Theory of Heat* (Macmillan) pp327–365
- [95] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379
- [96] Bloch I, Bretto A 2019 *Discrete Geometry for Computer Imagery: 21st IAPR International Conference* Marne-la-Vallée, France, March 26–28, 2019 pp143–154
- [97] Hu D, Li X L, Liu X G, Zhang S G 2019 *Acta Math. Sin. Engl. Ser.* **35** 1238
- [98] Wang H, Xiao G, Yan Y, Suter D 2018 *IEEE Trans. Pattern Anal. Mach. Intell.* **41** 697
- [99] Goldberg T E 2002 *Sr. Thesis Bard Coll.* **6** 25
- [100] Estrada E, Ross G J 2018 *J. Theor. Biol.* **438** 46
- [101] Serrano D H, Hernández-Serrano J, Gómez D S 2020 *Chaos Solit. Fractals* **137** 109839
- [102] Serrano D H, Gómez D S 2020 *Appl. Math. Comput.* **382** 125331
- [103] Bonacich P 2007 *Soc. Netw.* **29** 555
- [104] Katz L 1953 *Psychometrika* **18** 39
- [105] Estrada E, Knight P A 2015 *A First Course in Network Theory* (Oxford: Oxford University Press) pp157–160
- [106] Okamoto K, Chen W, Li X Y 2008 *International Workshop on Frontiers in Algorithmics* Changsha China, June 19–21, 2008 pp186–195
- [107] Newman M E 2005 *Soc. Netw.* **27** 39
- [108] Maletić S, Rajković M, Vasiljević D 2008 *Computational Science-ICCS 2008: 8th International Conference* Kraków Poland, June 23–25, 2008 pp568–575
- [109] Shi D, Lü L, Chen G 2019 *Natl. Sci. Rev.* **6** 962

REVIEW

Fundamental statistics of higher-order networks: a survey^{*}Liu Bo¹⁾³⁾ Zeng Yu-Jie¹⁾³⁾ Yang Rong-Mei¹⁾ Lü Lin-Yuan^{2)1)3)†}¹⁾ (*Institute of Fundamental and Frontier Studies, University of Electronic Science and Technology of China, Chengdu 610054, China*)²⁾ (*School of Cyber Science and Technology, University of Science and Technology of China, Hefei 230026, China*)³⁾ (*Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou 313001, China*)

(Received 17 February 2024; revised manuscript received 10 April 2024)

Abstract

Complex networks serve as indispensable instruments for characterizing and understanding intricate real-world systems. Recently, researchers have delved into the realm of higher-order networks, seeking to delineate interactions within these networks with greater precision or analyze traditional pairwise networks from a higher-dimensional perspective. This effort has unearthed some new phenomena different from those observed in the traditional pairwise networks. However, despite the importance of higher-order networks, research in this area is still in its infancy. In addition, the complexity of higher-order interactions and the lack of standardized definitions for structure-based statistical indicators, also pose challenges to the investigation of higher-order networks. In recognition of these challenges, this paper presents a comprehensive survey of commonly employed statistics and their underlying physical significance in two prevalent types of higher-order networks: hypergraphs and simplicial complex networks. This paper not only outlines the specific calculation methods and application scenarios of these statistical indicators, but also provides a glimpse into future research trends. This comprehensive overview serves as a valuable resource for beginners or cross-disciplinary researchers interested in higher-order networks, enabling them to swiftly grasp the fundamental statistics pertaining to these advanced structures. By promoting a deeper understanding of higher-order networks, this paper facilitates quantitative analysis of their structural characteristics and provides guidance for researchers who aim to develop new statistical methods for higher-order networks.

Keywords: higher-order network, hypergraph, simplicial network, statistics**PACS:** 89.75.-k, 89.75.Fb, 02.10.Ox**DOI:** [10.7498/aps.73.20240270](https://doi.org/10.7498/aps.73.20240270)

^{*} Project supported by the STI 2030-Major Projects (Grant No. 2022ZD0211400), the Major Program of the National Natural Science Foundation of China (Grant No. T2293771), and the Science Fund for Distinguished Young Scholars of Sichuan Province, China (Grant No. 2023NSFSC1919).

[†] Corresponding author. E-mail: linyuan.lv@ustc.edu.cn

高阶网络统计指标综述

刘波 曾钰洁 杨荣湄 吕琳媛

Fundamental statistics of higher-order networks: a survey

Liu Bo Zeng Yu-Jie Yang Rong-Mei Lü Lin-Yuan

引用信息 Citation: *Acta Physica Sinica*, 73, 128901 (2024) DOI: 10.7498/aps.73.20240270

在线阅读 View online: <https://doi.org/10.7498/aps.73.20240270>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

识别高阶网络传播中最有影响力的节点

Identifying influential nodes in spreading process in higher-order networks

物理学报. 2024, 73(4): 048901 <https://doi.org/10.7498/aps.73.20231416>

基于高阶信息的网络相似性比较方法

Network similarity comparison method based on higher-order information

物理学报. 2024, 73(3): 038901 <https://doi.org/10.7498/aps.73.20231096>

具有双峰特性的双层超网络模型

Double-layer hypernetwork model with bimodal peak characteristics

物理学报. 2021, 70(1): 018901 <https://doi.org/10.7498/aps.70.20201065>

高阶耦合相振子系统的同步动力学

Collective dynamics of higher-order coupled phase oscillators

物理学报. 2021, 70(22): 220501 <https://doi.org/10.7498/aps.70.20211206>

相位可压缩相干态的高阶光子反聚束效应

Higher-order photon antibunching of phase-variable squeezed coherent state

物理学报. 2022, 71(19): 194202 <https://doi.org/10.7498/aps.71.20220574>

高阶Ablowitz-Ladik方程的局域波解及稳定性分析

Dynamics of localized wave solutions for a higher-order Ablowitz-Ladik equation

物理学报. 2020, 69(1): 010205 <https://doi.org/10.7498/aps.69.20191235>