arXiv:2512.24961v1 [math.CO] 31 Dec 2025

无向（超）图中结构等价与正则等价的几何刻画

Marzieh Eidi[1,2] 和 Nina Otter[3,41] 可扩展数据分析与人工智能中心（ScaDS.AI）以及[2] 马克斯·普朗克数学科学研究所（MPI MIS），莱比锡，德国[3] DataShape，Inria-Saclay，法国[4] 奥赛数学实验室，巴黎-萨克雷大学，法国

摘要

图中顶点之间的相似性概念，例如结构等价和正则等价，是复杂网络科学中聚类工具的主要组成部分之一。我们将结构等价和正则等价推广到无向超图，并通过邻域图和 Olivier-Ricci 曲率，对无向图和超图中的结构等价与正则等价进行刻画。我们的刻画为这些相似性概念带来了新的认识，为探索它们开辟了新的途径。这些刻画还能够构造一个可能很广泛的正则划分族，从而为一项迄今在计算上颇具挑战性的任务提供了新的方法。

目录

1 引言 21.1 相关工作.3

2 图的背景 3

2.1 图上的等价关系.4 2.2 Ollivier-Ricci 曲率.5 2.3 邻域图.7

3 无向图的结构等价与正则等价的几何刻画 7

3.1 n-邻域图的连通分量.7 3.2 结构等价.10 3.3 正则等价.11

4 无向超图的情形 15

5 图和超图的近似结构等价 21 6 结论与开放问题 24

7 致谢 **24**

1 引言

社会网络分析是社会科学的一个分支，其特点是采用结构主义方法，着重研究社会行动者之间的关系，而非行动者本身。该领域至少自20世纪30年代以来便不断发展，历史概述见 [Fre04] ，主要方法和问题概述见 [WF94] 。该领域的一个主要问题是如何定义某个社会行动者在系统中所占据的位置。人们提出了不同的等价关系，以描述两个行动者占据相似社会位置的含义。最早提出的是结构等价这一概念：如果两个行动者以相同的方式与相同的行动者互动，那么他们就是结构等价的。这个概念过于严格，在实践中并不实用，因此人们通过更灵活的概念对其加以放宽：一方面，人们希望研究近似结构等价的度量，以衡量两个行动者在多大程度上结构等价；另一方面，人们也寻求更灵活的等价关系，其中最灵活的是正则等价。如果两个行动者正则等价，那么他们各自互动的行动者也彼此等价。

正则等价比结构等价更加灵活，但在实践中计算起来也更加困难。满足正则等价性质的划分集合构成一个格，人们已提出多种算法来计算该格中的某些元素，例如 REGE [BE93], ，这是一种通过迭代细化划分并返回最大正则等价的算法。近年来，也有研究提出了计算近似正则划分的算法 [STTV25]。不同算法及其挑战的概述见 [EB19]。在超图中，情况则更具挑战性，因为超图刻画了两个或更多参与者之间的交互。近年来，尽管从图推广到超图的方法取得了进展[BGHS23, ZHS07], ，但超图结构等价和正则等价概念及其算法的发展基本上仍是一片空白。

我们的工作提出了一种基于 Ollivier–Ricci 曲率的结构等价几何刻画。Ollivier-Ricci 曲率 [Oll09] 是度量测度空间中 Ricci 曲率的一种推广，其定义方式是通过最优传输距离比较邻域上的概率测度。直观而言，如果两个顶点之间路径上的邻域只需付出较低的传输成本便可相互映射，那么该路径上的曲率就是正的。因此，曲率是刻画结构等价中邻域重叠的天然工具，也使我们能够针对超图定义不同类型的结构等价度量，这些度量基于能够捕捉不同局部关联模式的不同类型的随机游走。

在另一条并行研究路线中，我们提出了一种方法，可基于图和超图的路径邻域图计算某些正则划分（即满足正则等价性质的划分），以及超——

图，从而据我们所知，首次提供了计算超图某些正则划分的算法。

我们的主要贡献如下：

- 我们为超图定义了弱结构等价、强结构等价、弱正则等价和强正则等价的概念，将图中的结构等价和正则等价加以推广（参见定义 4.2）。
- 我们为图中的结构等价（定理 3.4）和超图中的结构等价（定理 4.9）提供了基于曲率的刻画。
- 我们提出了一种方法，通过相关的 n-路径邻域图 G_n 的连通分量，计算图（定理 3.6 和定理 3.7）及超图（推论 4.13((iii))-((iv)) 和推论 4.14）中的一类正则划分。
- 在第 5 节中，我们提出了一种基于曲率的度量，用于近似图和超图中的结构等价，并将其与余弦相似度联系起来。余弦相似度是图中近似结构等价的主要度量之一。

1.1 相关工作

据我们所知，曲率与结构等价或正则等价之间的联系此前尚未得到研究。我们的工作很大一部分在于将基于路径的邻域图的连通分量结构与正则类或结构类联系起来。至少自 20 世纪 80 年代以来，基于路径的邻域图就已得到研究，参见 [Sim83、BGK19], ；但与基于游走的邻域图相比，它们鲜为人知。后者因与图的邻接矩阵的幂有关，在复杂网络科学及相关领域中得到了广泛研究。尤其是，它们与 Ollivier-Ricci 曲率之间的联系已在 [BJL13]中得到研究。

据我们所知，超图正则等价的推广仅在 Borgatti 的博士论文 [Bor89], 中有所探讨；另见综述文章[EB93], ，其中 Borgatti 基于正则着色提出了一种等价的正则等价概念。

据我们所知，文献此前尚未探讨基于邻域图和曲率计算正则划分的算法。因此，我们的方法与现有的计算划分迭代细化的方法有根本区别 [BE93、STTV25]。

2 图的背景知识

本节回顾与无向图的结构等价和正则等价、这些等价关系的近似概念、图上的 Ricci 曲率以及邻域图相关的基本概念。

定义 2.1。 一个**无向图** G 是一个二元组(V, E)，其中 V 是一个非空集合，且$E \subset \mathcal{P}_{\leq 2}(V)$。我们称 V **顶点** 为顶点，称 E **边**为边。我们称一条

基数为1 的边 $e \in E$ 为**环**。如果存在一条边 e，使得 $x, y \in e$，则两个顶点 $x, y \in V$ 互为**邻居**。顶点 x 的**度** d_x 是包含该顶点的边的数量。

因此我们注意到，对我们而言，无向图可以包含自环，但任意两个顶点之间至多允许有一条边。

定义 *2.2.* 设 $G = (V, E)$ 为一个无向图。两个顶点 u、$v \in V$ 之间的游走是边 e_1、...、e_n 以及顶点 x_0、x_1、...、x_{n-1}、x_n 构成的序列 $u = x_0 e_1 x_1 e_2 \dots x_{n-1} e_n x_n = v$，满足对于所有 $i = 0$、...、$n-1$，都有 x_i 和 $x_{i+1} \in e_{i+1}$。从同一顶点出发并终止于该顶点，且除该顶点外没有其他顶点重复出现的游走称为环。我们还将包含 k 个不同顶点的环称为 k-环。于是，如果一个环是 k-环，其中 k 为奇数（分别为偶数），则称该环为奇环（分别为偶环）。如果对于 V 中的任意一对顶点，都存在连接它们的游走，则无向图 $G = (V, E)$ 是连通的。图 $G = (V, E)$ 的子图是指满足 $V' \subset V$ 和 $E' \subset E$ 的图 $G' = (V', E')$。无向图 $G = (V, E)$ 的连通分量是 G 的一个连通子图 G'，且不包含于 G 的任何其他连通子图中。完全图是这样一种图：其中任意两个顶点之间都由一条边连接。

2.1 图上的等价关系

定义 2.3。给定一个无向图 G，以及其顶点集合上的等价关系 $\sim$，若该等价关系满足以下条件，则称其为：

- **结构等价**：对所有顶点 a、b，均有 $a \sim b$ 当且仅当：$\{a, c\}$ 是 G 中的一条边，当且仅当 $\{b, c\}$ 是 G 中的一条边。
- **正则等价**：对所有顶点 a、b，均有 $a \sim b$ 当且仅当：如果 $\{a, c\}$ 是 G 中的一条边，则存在 d，使得 $\{b, d\}$ 是一条边且 $c \sim d$。

给定图 G 的顶点集合的一个划分，若其满足结构等价关系（分别为正则等价关系）的性质，则称其为**结构**（分别为**正则划分**）划分。同样地，我们将这种划分的块称为 G 的**结构**（分别为**正则类**。

我们注意到，对于给定的图 G，至多存在两个结构等价，其中一个是仅由单元素块组成的划分。每个结构等价都是正则等价。另一方面，并非每个正则等价都是结构等价；事实上，一般而言，给定图的顶点集合上存在许多不同的划分，它们都满足正则等价的性质。给定图的正则划分构成一个按包含关系排序的格。该格具有一个极大元；对于无向图而言，它总是由仅含一个块的平凡划分给出；同时还具有一个极小元，即每个块都是单元素的划分 [WR83]。由于极大划分在我们的工作中发挥着重要作用，我们采用以下术语：

定义 2.4。如果一个集合的划分恰好包含一个块，我们就称其为**平凡的**。

在实际中，图中的顶点很少结构等价。因此，人们试图找到一些度量方法，量化两个顶点在多大程度上不具备结构等价性。文献中已经提出了许多此类概念。这里我们回顾余弦相似度，它是主要使用的概念之一。

定义 2.5。 设 x、 y 是无向图 G中的两个顶点。 x 和 y 的**余弦相似度**定义如下：

$$\sigma_{x,y} = \frac{\eta_{xy}}{\sqrt{d_x}\sqrt{d_y}},$$

其中， n_{xy} 是 x和 y的公共邻居数量。

我们注意到，尽管正则等价具有更大的灵活性，人们也研究近似正则等价的概念。有关正则等价主要度量指标的概述，请参阅 [WF94, Section 12.4.4] 。

2.2 Ollivier-Ricci 曲率

2007 年，Ollivier [Oll07] 将 Ricci 曲率推广到了配备随机游走的度量空间，而图和超图就是其中的特例。这里我们先给出 Ollivier 提出的定义，再给出其在图上的具体形式；超图的情形将在第 4 节讨论。

定义 2.6。 设$(X$、 $d)$ 是一个完备可分的度量空间，并配有一个 *Borel*σ-代数。在 X 上的**随机游走** μ 是定义在 X 上的一族概率测度 μ_x ，对每个$x \in X$均有一个测度，且每个测度 μ_x 关于 x 可测并具有有限一阶矩。给定两个概率测度 μ_x、 μ_y， μ_x与 μ_y 之间的**运输距离** 或**1-Wasserstein 距离** 为

$$W_1(\mu_x, \mu_y) := \inf_{\mathcal{E} \in \Pi(\mu_x, \mu_y)} \int_{(x,y) \in X \times X} d(x,y) d\mathcal{E}(x,y)$$

其中 Π（μ_x、 μ_y）是投影到 $X \times X$ 和 μ_x 、 μ_y上的概率测度集合。

定义 2.7。 设（X、 d）为配备随机游走 μ ： $=$ （μ_x）$_{x \in X}$ 的度量空间，且x、 $y \in X$ 为两个不同的点。***x*、 *y*的 Ollivier-Ricci 曲率（ORC）为：**

$$\kappa(x,y) := 1 - \frac{W_1(\mu_x, \mu_y)}{d(x,y)}.$$

要为无向图定义 Ollivier-Ricci 曲率，通常将最短路径距离作为度量。

定义 2.8。 设 $G =$ （V、 E）为一个无向图。V 中两个顶点之间的**路径**是指顶点和边均各不相同的游走。路径 γ 的**长度**是其包含的边数，即$\mathrm{length}(\gamma)$ 。**最短路径**（或**组合距离**）是由 d_{SP}： $V \times V \to \mathbb{N}$ 定义的函数

$$\mathrm{d}_{\mathrm{SP}}(u,v) = \min_{\gamma} \mathrm{length}(\gamma),$$

其中，最小值取遍所有连接 γ 与 u 的路径 v。

我们注意到，在图的几何研究中，通常会考虑限制较少的路径概念，其中允许重复经过某些顶点或边。虽然这些概念不适用于刻画结构等价或正则等价，因为重复经过顶点或边会给等价关系添加冗余信息，但我们在此回顾它们的定义，并将在适当之处明确讨论采用这些替代概念时所产生的差异。

定义 2.9。 设 $G = (V, E)$ 为一个无向图。顶点 $u, v \in V$ 之间的**非回溯游走** 是指一种不允许出现模式 $x_i e_{i+1} x_{i+1} e_{i+1} x_i$的游走，也就是说，游走不能立即返回先前访问过的顶点。

为了定义无权图中任意两个顶点的 Ollivier-Ricci 曲率，我们通过给每个顶点指定一个概率测度来定义随机游走：该测度基于从该顶点随机跳转到其相邻邻点的过程：

定义 2.10 （图上的 Ollivier-Ricci 曲率）。 设 $G = (V, E)$ 为一个无自环的连通无向图。对于每个顶点 $x \in V$，定义**邻点测度** μ_x如下：

$$\mu_x(z) = \begin{cases} \dfrac{1}{d_x} & \textit{if } z \textit{ is a neighbour of } x\,, \\ 0 & \textit{otherwise.} \end{cases}$$

运输距离或**1-Wasserstein 距离**介于两个邻点测度 μ_x 和 μ_y 之间，其值由以下最优传输问题的解给出：

$$W_1(\mu_x, \mu_y) = \min_{\xi \in \Pi(\mu_x, \mu_y)} \sum_{u,v \in V} d(u, v)\, \xi(u, v),$$

其中 $\Pi(\mu_x, \mu_y)$ 是 μ_x 与 μ_y之间所有耦合构成的集合，也就是说，

$$\Pi(\mu_x, \mu_y) = \big\{\, \xi : V \times V \to [0, 1] \;\big|\; \sum_{v \in V} \xi(u, v) = \mu_x(u),\ \sum_{u \in V} \xi(u, v) = \mu_y(v) \big\}.$$

于是，对于同一连通分量中的一对顶点 x 和 y，**Ollivier–Ricci 曲率** 定义为

$$\kappa(x, y) := 1 - \frac{W_1(\mu_x, \mu_y)}{d_{SP}(x, y)}\,.$$

我们注意到，我们在此给出的图的 Ollivier-Ricci 曲率定义最早由 [Oll07], 和 [LY 10]研究。“懒惰”版本的邻点随机游走中，随机游走以给定概率 α 不离开其起始顶点；该版本由 [YL11]提出。我们在此考虑的版本对应于情形 $\alpha = 0$。在 [YL11]中，作者研究了对[Oll07, LY10], 中提出的图的 Ricci 曲率概念的一种修正，其中他们考虑当 $\alpha \to 1$时的极限。

直观地说，可以把 μ_x 的值想象成沙子，把 μ_y 的值想象成孔。由于 μ_x 和 μ_y 都是概率测度，它们的总质量相等，均为一。因此，最优传输的核心思想就是确定将沙子运送到孔中的最高效（即成本最低）方式。

2.3 邻域图

我们刻画中的一个主要组成部分是与无向图或超图相关联的邻域图。

定义 2.11。 令 $G = (V, E)$ 为一个无向图。令 $n \in \mathbb{N}$ 为正整数。 $\boldsymbol{G}$ 的第 $\boldsymbol{n}$**个邻域图**是图 $G_n = (V, E_n)$ ，其顶点为 G的顶点，边由 G中长度为 n 的路径给出：

$$E_n = \{\{u, v\} \mid \ there\ exist\ a\ path\ between\ u\ and\ v\ \ of\ length\ n\}\,.$$

我们注意到，尽管基于游走的邻域图可能是一个更为熟悉的概念，因为它们与图的邻接矩阵的幂直接相关，但基于路径的邻域图至少自 20 世纪 80 年代起就已受到研究，参见 [Sim83、BGK19]。

由于在适当情况下，我们将讨论基于路径的不同放宽方式所得到的邻域图之间的一些差异，因此我们在此给出定义：

定义 2.12。 令 $G = (V, E)$ 为一个无向图。令 $n \in \mathbb{N}$ 为正整数。 $\boldsymbol{G}$ 的第 $\boldsymbol{n}$**个非回溯邻域图**是图 $G_n^{\mathrm{NB}} = (V, E_n^{\mathrm{NB}})$，其顶点为 G 的顶点，边由 G中长度为 n 的非回溯游走给出：

$$E_n^{\mathrm{NB}} = \{\{u, v\} \mid \text{存在一条长度为 } n\text{的非回溯游走连接 } u \text{ 和 } v \qquad \}\,.$$

第 $\boldsymbol{n}$**阶游走邻域图** $\boldsymbol{G}$ 是图 $G_n^{\mathrm{walk}} = (V, E_n^{\mathrm{walk}})$ ，其顶点为 G 的顶点，边由 G中长度为 n 的游走给出：

$$E_n^{\mathrm{walk}} = \{\{u, v\} \mid \text{存在一条长度为 } n\}\text{.的游走连接 } u \text{和 } v$$

我们注意到，由三种不同的路径概念（即游走、非回溯游走和路径）构造的邻域图通常并不相同。根据构造，对于任意 $n \geq 2$ ，相应的边集满足以下包含关系：

$$E_n \subseteq E_n^{\mathrm{NB}} \subseteq E_n^{\mathrm{walk}}\,. \tag{1}$$

3 无向图的结构等价与正则等价的几何刻画

本节研究无向图的结构等价与正则等价。首先，我们讨论 n-邻域图的一些性质，这些性质将在本节其余部分发挥作用。

3.1 n 邻域图的连通连通分支

随着 n 增大， n-邻域图 G_n 可能会很快变得不连通。例如，当 n 超过图的直径时，在具有 m 条边的星形图中很容易观察到这一现象： G_2 已经有两个连通分量，而 G_3则由 $m + 1$ 个孤立顶点组成。我们首先给出 G_2的连通分量数的上界。

引理 3.1. 设 G 为一个连通图，则 G_2 至多有两个连通分量。更精确地说：

1. 如果 G是非二分图（即包含一个奇圈），那么 G_2 是连通的。

2. 如果 G 是二分图，且顶点集 $V = V_0 \sqcup V_1$被划分为两个部分，那么 G_2 恰有两个连通分量，其顶点集分别为 V_0 和 V_1。

证明。 注意，若 G 是二分图，其顶点集划分为 V_0 和 V_1，则根据定义，属于 V_0 的节点之间没有边，属于 V_1的节点之间也没有边。由于 G 是连通的， V_0中任意两个顶点之间都存在一条路径， V_1中任意两个顶点之间也同样如此。此类路径每一步都必然在两个划分部分之间交替，因此长度为偶数。反之， V_0 中的给定节点与 V_1 中的给定节点之间的路径必然具有奇数长度。因此， V_0 和 V_1 是 G_2的两个连通分量的顶点集，因此 G_2 恰有两个连通分支。

如果 G 不是二分图，我们证明 G_2 是连通的，也就是说， G_2中任意两个顶点之间都存在一条路径。令 x 和 y 为 G中的任意顶点。由于 G 是连通的，它们之间至少存在一条路径。如果这条路径的长度为偶数，那么 x 和 y 位于 G_2的同一个连通分量中。

反之，假设在 G中， x 和 y 之间不存在偶数长度的路径。那么它们之间必定存在一条奇数长度的路径；将这条路径写作

$$x = x_1, x_2, \ldots, x_k, x_{k+1} = y,$$

其中 k 为奇数。由于 x 和 y之间不存在偶数长度的路径，给定的奇数长度路径不能经过奇圈上的任何边；否则，沿奇圈的其余部分行进，就能得到一条连接 x 和 y的偶数长度路径。由于 G 是非二分图，它至少包含一个奇圈。令 α 为该奇圈上的一个顶点，使得经过 y 、从 x_k 到 α 的路径至少包含该奇圈的一条边（示例见图 3.1）。

由于这个环的长度为奇数，因此 x_k 和 α之间存在两条不同的路径，一条长度为偶数（称为路径 1），另一条长度为奇数（称为路径 2）。因此， α 在 G_2中与 x_k 连通。此外，由于 x_k 和 y 在 G中相邻， α 也在 G_2 中与 y 连通（沿着一条经过修改的路径 $2'$，该路径由路径 2 删除 G中的边 $\{x_k, y\}$ 后得到），因此，x_k 和 y 位于 G_2的同一个连通分量中。由于 x和 x_k 由一条长度为偶数的路径连接，它们也属于同一个连通分量。因此， x 和 y 属于 G_2的同一个连通分量。由于 x 和 y 是任意选取的，因此 G_2 是连通的。

□

有人可能会想知道，随着 n变化，G_n 的连通分量数量是否会增加或减少。一般而言，两种情况都不会发生，正如下一个例子所示。如果 n 超过图的直径， G_n 就会变得不连通；当n足够大时，它由一组孤立顶点构成。

例3.2。 我们考虑图*2*中的图，并注意到由此产生的 n邻域图恰好有四种不同类型：

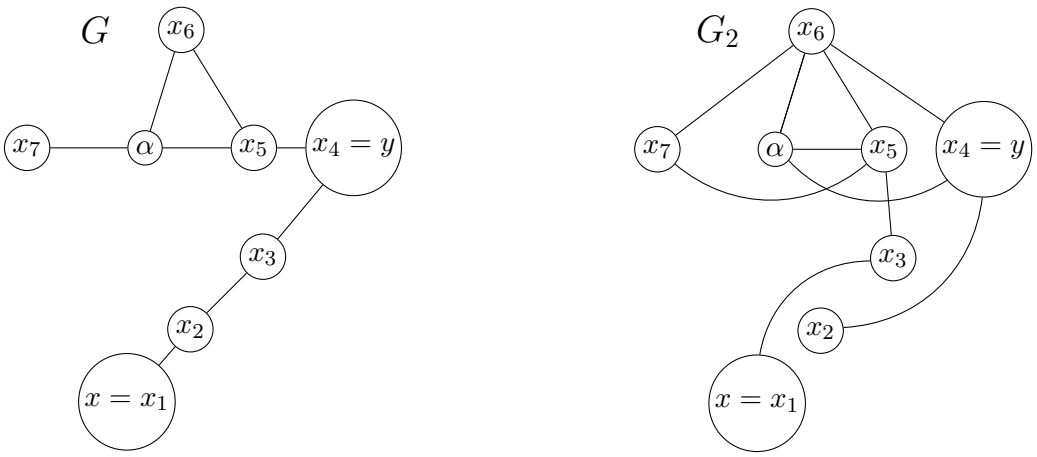

图1：引理 3.1 证明中针对G 非二分图的构造示意图。

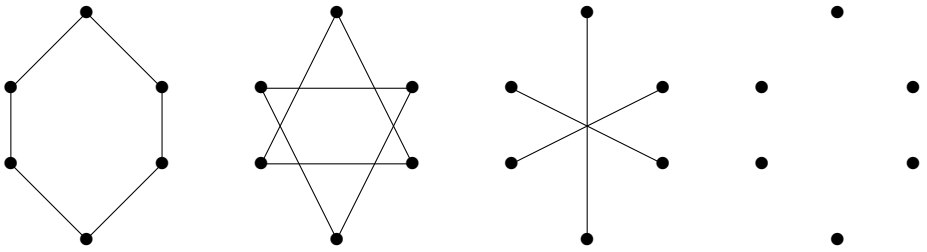

图2： 作为 G的 n-邻域图出现的四种不同类型的图 .

$$G_n = \begin{cases} G, \ n = 1 \ or \ 5 \\ G_2, \ n = 2 \ or \ 4 \\ G_3, \ n = 3 \\ G_6, \ n \geq 6 \end{cases}$$

因此，我们可以看到，随着 n 增加， G_n 的连通分支数在1、2、3和6 之间变化。

我们注意到，根据公式 1，连通分量的数量满足

$$\#\mathrm{Comp}\big(G_n^{\mathrm{walk}}\big) \ \leq \ \#\mathrm{Comp}\big(G_n^{\mathrm{NB}}\big) \ \leq \ \#\mathrm{Comp}(G_n)\,.$$

换言之，由游走定义的 n-邻域图至少与其非回溯对应图同样连通，而非回溯对应图又至少与 n-邻域图同样连通。值得注意的是，非回溯 n-邻域图与 n-邻域图的一个关键区别在于它们处理环的方式。前者中，游走可以任意多次经过长度至少为三的环；而后者中，

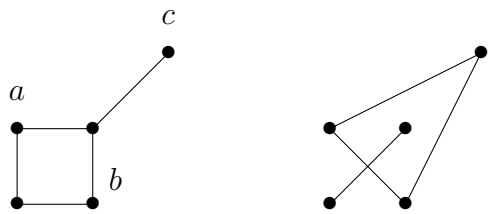

图 3：（左）图 G 及其（右）2-邻域图 G_2。尽管 G_2 的所有连通分量都是完全的，但节点 a、b、c 上的连通分量并不导出 G的结构类。

路径不能多次经过同一个环。因此，当 $n = 2,3$ 时，我们有 $G_n^{\mathrm{NB}} = G_n$。

3.2 结构等价

在这里，我们开始研究 n-邻域图与结构划分之间的关系。为此，我们首先固定一些术语：

定义3.3。设 $n \geq 2$，并考虑无向图 G的 n-邻域图 G_n 。由 G_n 的连通分量**导出** 的类是 V 的一个划分的块，该划分由属于同一连通分量的顶点给出。

定理 3.4。设 G 是一个不含环的连通无向图。则有：

（*i*） G 中的两个顶点 v、v' 结构等价，当且仅当 (v, v') 的 *ORC* 等于 1。（*ii*） G 的结构类是 G_2的连通分量的完全子图的顶点集。

证明。(i) 根据定义，当且仅当两个顶点 x 和 y 拥有相同的邻居时，它们才是结构等价的。这意味着，对于上文定义的 x 和 y， μ_x 和 μ_y 相同，因此它们之间的 1-Wasserstein 距离等于 0。反过来，$\kappa(x,\ y) = 1$ 意味着 μ_x 和 μ_y 之间的 1-Wasserstein 距离等于 0；由于 W_1 是度量， $\mu_x = \mu_y$，因此 x 和 y 恰好拥有相同的邻居。

(ii) 如果 x 和 y 结构等价，它们就恰好拥有相同的邻居，尤其属于 G_2 的同一个连通分量。我们需要证明，对于同一个结构类中的所有顶点，它们之间所有可能的边都存在于 G_2 中。假设 x 和 y 结构等价。如果它们之间没有边，这意味着在 G 中， x 和 y 之间不存在长度为 2 的路径，这与它们至少有一个共同邻居这一事实相矛盾。

□

需要指出的是，命题 (ii) 的逆命题一般并不成立；也就是说，即使 G_2 的一个连通分量是完全图，也不一定意味着它诱导出一个结构类，如图 3 所示。然而，结合定理 3.4 中的命题 (i) 和 (ii)，我们得到以下刻画：

推论 3.5. 连通分量 G_2 的完全子图诱导出 G 的结构类，当且仅当同一子图中任意两对顶点的 *ORC* 均等于1，且该子图是满足此性质的极大完全子图（即不存在更大的完全子图，使其中所有顶点对的 *ORC* 均等于 1。）

如果我们再次考虑图 3 中的例子，可以看到，顶点 a、b、c 构成的连通分量并不导出一个结构类，而由边 $\{a, b\}$ 构成的完全子图则可以。事实上，这是 $\{a, b, c\}$ 中 ORC 等于 1 的唯一一对顶点。

我们注意到，根据上述刻画， G_2 的连通分量只能为 G的结构类提供有限的信息。接下来我们将看到，如果改为考虑更一般的正则等价， G_2 的连通分量总能导出 G的正则类。最后请注意，若两个顶点 x 和 y 结构等价，它们不能是邻居，并且必须至少有一个共同邻居，因此它们必然属于 G_2的同一个连通分量。我们将在第 3.3 节看到，排除这一要求后，如何超越 G_2，进而考虑一般 n的 G_n。

3.3 正则等价

现在我们转向正则等价。如第 2.1 节所述，给定图 G的顶点集 V 上的正则划分集合构成一个格。如何高效计算该格中的元素，一直是备受关注的问题。本节将探讨格中的元素与 n-邻域图的连通分量之间的联系。需要指出的是，一般而言， G 的非平凡正则划分不一定与任意 n对应的 G_n 的连通分量相同。例如，在一条含有三条边的线图中，可以将两个端点置于同一个正则等价类中，并将两个内部顶点置于另一个正则等价类中，尽管这并不反映任何 G_n的连通分量结构。不过，正如本节所示，在若干情况下， G_n 的连通分量结构可以得出 G的正则划分。我们先从G_2的情形开始。

定理 3.6。 G_2 的连通分量诱导出 G的正则类。此外，如果 G_2不连通，即有两个连通分量，那么这些连通分量就是一个正则划分的块，且该划分并不真包含于任何非平凡划分中。

证明。 对于第一部分，假设 x 和 y 是 G_2 的同一个连通分量中的两个顶点。不失一般性地，我们可以假设 x 和 y 通过 G_2 中的一条边相连。根据 G_2 的定义，这意味着在 G 中存在一个顶点 α，使得 α 同时是 x 和 y 的邻点。现在，对于 G 中 x 的任意邻点 z，如果 $z = \alpha$，那么由于 α 也是 y 的邻点，正则等价类的要求显然得到满足。如果 $z \neq \alpha$，那么 z 和 α 通过 G_2 中的一条边相连，因此它们属于同一个连通分量。这就证明了该陈述的第一部分。

接下来我们注意到，如果 G_2 不连通，那么由 G_2 的连通分量所诱导的类构成一个具有两个块的划分。根据构造，唯一比该划分更粗的划分是平凡划分。

□

我们注意到，前一个结果不能直接推广到任意 G_n 的 n；也就是说，一般而言， G_n 的连通分量并不产生正则类，即使 G_n 恰好具有两个连通分量也是如此，图4所示的示例说明了这一点。

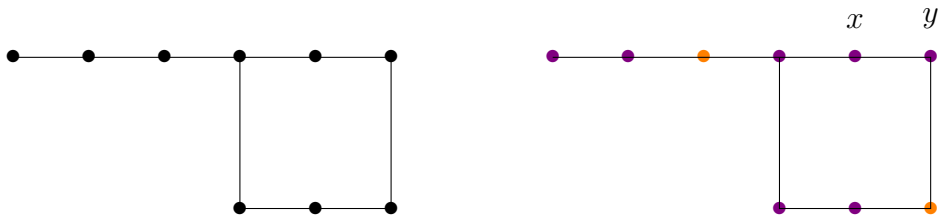

图4：图 G （左），以及由 G_4 的两个连通分量诱导的节点集合的两个块划分；我们将同一块中的顶点着以相同颜色（右）。 G_4 的两个连通分量并未在 G 上诱导出正则类：节点 x 和 y 属于同一个类，然而，尽管 y 通过一条边与一个橙色节点连通，却不存在从 x 到橙色节点的边。

尽管如此，在增加没有顶点的度为1这一假设后，我们可以将该结果推广。

定理 3.7。 设 G 为无向图。如果 G 中的所有顶点的度至少为 2 ，那么 G_n 的连通分量将在 G 上诱导出正则类。

证明。 假设节点 a 和 b 位于 G_n 的同一个连通分量中，并且不失一般性地假设它们在 G_n 中通过一条边相连。因此，在 G 中存在一条长度为 n 的路径连接它们，例如： a， x_1， $\ldots$， x_{n-1}， b。现在，考虑 a 的任意一个邻点，例如 α。如果 $\alpha \neq x_1$，那么 $\{\alpha, x_{n-1}\}$ 是 G_n 中的一条边，其中 x_{n-1} 是 b 在 G 中的一个邻点。如果 $\alpha = x_1$，由于 $d_b \geq 2$ ，存在 b 的一个邻点 β ，使得 $\beta \neq x_{n-1}$ ，并且 $\{x_1, \beta\}$ 是 G_n 中的一条边。根据正则等价的定义，这证明了该断言。 □

注3.8。 顶点的度条件并不是 G_n 的连通分量诱导出 G 的正则类的必要条件。例如，考虑图5中的图。 G_3 的连通分量在 G 上诱导出一个正则划分，然而顶点a 在 G 中的度为 1 。

现在，聚焦于二阶邻域图 G_2，我们提出以下问题：

(∗) 我们能否从 G_2 的连通分量中得到其他非平凡的正则类?

放宽推论 3.5 中的极大性条件可以得到一种答案。

推论3.9。 G_2 的连通分量的完全子图满足：每个子图中所有顶点对在 G 中的曲率均等于1 ，则这些完全子图在 G 中诱导出正则类。

证明。 设 x 和 y 是 G_2 的一个连通分量中的完全子图内的两个顶点，且 $(x,\ y)$ 在 G 中的曲率等于 1。这意味着它们在 G 中的邻居相同，因此该子图诱导出一个正则等价类。

□

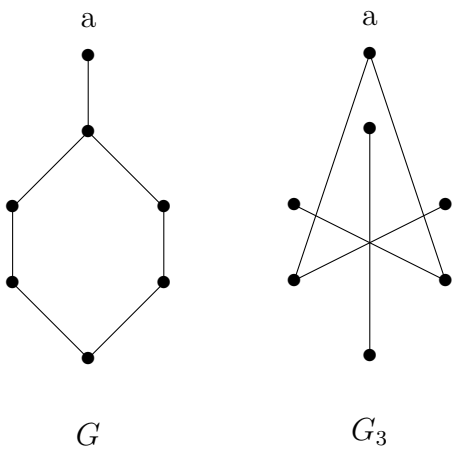

图5：一个图 G，其顶点的度为1，且其中 G_3 的连通分支给出正则划分。

因此，根据推论3.9，我们可以通过将 G_2 的各连通分量中的完全子图所诱导的块作为划分，得到 G 的一个正则划分；其余顶点则各自分配到一个单元素块中。我们在图6中给出一个例子。请注意，在所得划分中，属于同一类的任意两个顶点之间，在G中都没有边相连。事实上，这种划分可视为给定结构划分的细化（在图6中，这指的是左下方的第一个划分）。

下面的命题通过对定理 3.6 稍作修改，为我们的问题（$*$）提供了另一种解答。

命题 3.10. 设 G' 是从 G 中删除每个三角形至多一条边后得到的图，设 G'_2 是 G' 的 2 邻域图。那么，G'_2 的连通分量在原图 G 中诱导出正则类。

证明。我们要证明：如果 $\{x, y\}$ 是 G'_2 中的一条边，且 z 是 G 中 x 的任意一个邻点，那么 G 中存在一个 y 的邻点（例如 z'），使得 z 和 z' 属于 G'_2 的同一个连通分量。根据构造，如果 $\{x, y\}$ 是 G'_2 中的一条边，那么存在 α，使得 $\{x, \alpha\}$ 和 $\{\alpha, y\}$ 是 G'（以及 G）中的边。现在有两种可能：要么 G 中 x 和 y 之间有一条边（即 α，x，y 构成一个三角形），要么不存在这样的边（不存在这样的三角形）：

- 如果在 G 中，x 和 y 之间有一条边，那么 x 的任意一个邻点可能是 α、y，也可能两者都不是。在第一种和第二种情况下，正则类的条件显然成立。在最后一种情况下，将 x 的邻点记为 α'。那么在 G'_2 中，α 和 α' 之间有一条边，因此 α 和 α' 属于 G'_2 的同一个连通分量。由于 α 是 G 中 y 的邻点，这就证明了该结论。

- 如果在 G 中 x 和 y 之间没有边，那么 x 的任意一个邻点都可能是 α，也可能是一个不同于 α 和 y 的顶点 α'。证明与前一项中的证明类似，这里略去。

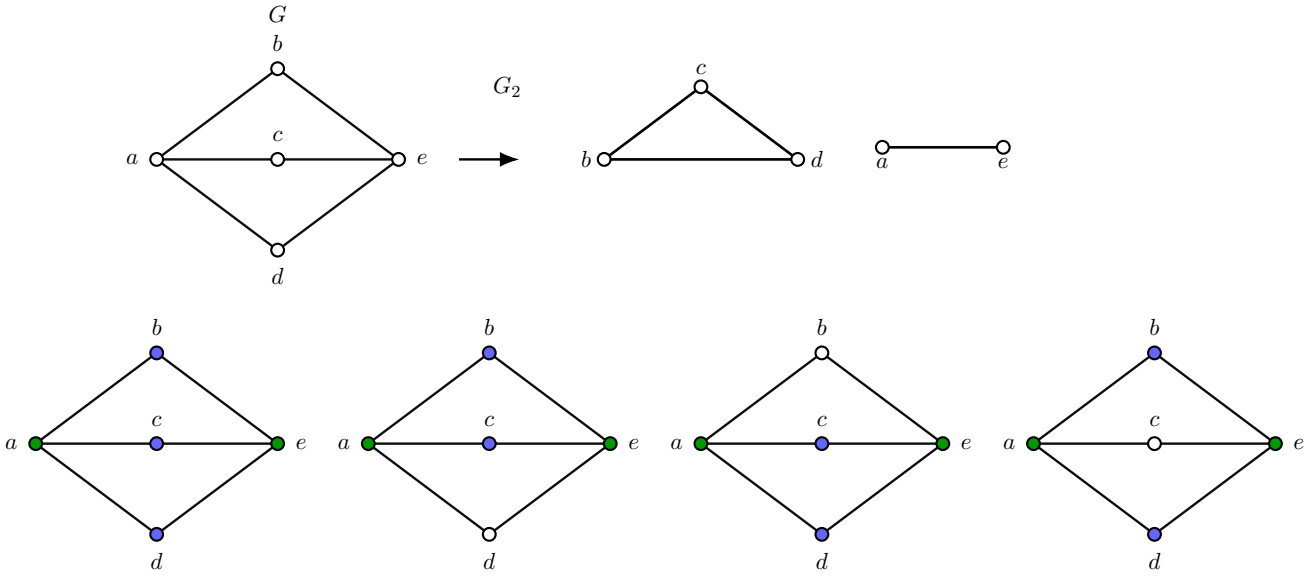

图6：上：图 G（左）及其 G_2（右），后者分裂为 $\{b, c, d\}$ 上的一个三角形和右侧的一条边 $\{a, e\}$ 。下： G_2 各连通分支的完全子图，其中每对顶点的曲率均为 1（在 G 中），它们在 G 中诱导出正则类；颜色相同的顶点属于同一个正则类。

□

我们已在定理 3.6 中看到， G_2 的连通分量在 G 上诱导出正则划分。现在，若 G 是非二分图，根据引理 3.1，我们知道这个正则划分是平凡划分。于是，命题 3.10 提供了一种方法，在奇圈由三角形构成的情况下，从 G_2 得到非平凡正则划分。因此，我们可以问，对于任意奇圈，是否也能得到类似的结论。确实可以，但代价是在可删除的边上增加一些限制。

命题 3.11. 设 G 是一个连通无向图，且没有度为 1 的顶点。设 G' 是从 G 中至多删除一条 G 中的 k-环的边（$k \geq 3$）后得到的图，并且删边后不会出现度为 1 的顶点（换言之， G' 中没有度为 1 的顶点）。设 G'_{k-1} 是 G' 的$(k-1)$邻域图。那么， G'_{k-1} 的连通分量在原图 G 中诱导出正则类。

证明。设 $\{x, y\}$ 是 G'_{k-1}中的一条边。设 α 是 G中 x 的一个邻点。我们需要证明，存在 G 中 y 的一个邻点 β，使得 α 和 β 位于 G'_{k-1}的同一个连通分量中。根据构造， x 和 y之间存在一条长度为 $k-1$ 的路径，例如 x， x_1，……， x_{k-2}， y。分两种情况：（i）若 $\alpha = x_1$，则由于 y 在G'中的度至少为 2，存在一个不同于 x_{k-2}（在 G'中）的 y 的邻点 β，因此 α， β 位于G'_{k-1}的同一个连通分量中。（ii）若 $\alpha \neq x_1$，则 α 和 x_{k-2} 位于 G'_{k-1}的同一个连通分量中。 □

4 无向超图的情形

据我们所知，结构等价和正则等价的概念尚未推广到超图等高阶组合结构。尽管这两个概念都可以直接推广到超图（见定义 4.2），主要难点在于如何计算满足正则性条件的划分。一位作者 [MOR25] 正在当前研究中开发一种基于余代数理论的框架，以便为某一类有向超图高效计算正则等价。本文将部分几何刻画推广到无向超图。

定义 4.1。 无**向超图** $\mathcal{H}$ 是一个二元组 (V, H)，其中 V 是一个非空集合，而 $H \subset \mathcal{P}(V)$ 是 V 的子集组成的集合。我们将 V **称为** $\boldsymbol{H}$ 的顶点，将 H **称为** $\boldsymbol{H}$ 的超边。超边的**基数** $|e|$ 是 e 中元素的数量。我们将基数等于2 的超边称为**边**，将基数等于1 的超边称为**环**。如果存在超边 e 使得 $u, v \in e$，则两个顶点 u, v 互为**邻居**。顶点的**度** d_v 是包含顶点 $v \in V$ 的超边数量。

接下来我们介绍结构等价和正则等价的两种不同概念。在弱版本中，我们只考虑给定两个顶点的邻居，而不考虑将它们与相邻顶点连接起来的关系类型。在强版本中，我们还会考虑关系的类型。

定义 4.2. 考虑一个无向超图 $\mathcal{H} = (V, H)$ 及其顶点集合上的一个等价关系 $\sim$。我们称 $\sim$为：

- **弱结构等价**，如果对所有顶点 $u, v \in V$，都有 $u \sim v$ 当且仅当：$\{u, a\} \subset e$ 存在某条超边 e，当且仅当 $\{v, a\} \subset e'$ 存在某条超边e'。换言之，u, v 恰有相同的邻居。

- **强结构等价**，如果对所有顶点 $u, v \in V$，都有 $u \sim v$ 当且仅当：$\{u, a\} \subset e$ 存在某条超边 e，当且仅当 $\{v, a\} \subset e'$ 存在某条超边e'（即，$\sim$ 是弱结构等价），并且还满足 $|e| = |e'|$。换言之，u, v 通过基数相同的超边与相同的邻居相连。

- **弱正则等价** 当且仅当对于所有顶点 u, v，我们有 $u \sim v$：若$\{u, a\} \subset e$对于某条超边 e成立，则存在 $a' \in V$和 $e' \in H$，使得 $\{v, a'\} \subset e'$和 $a \sim a'$。换言之，u, v 的邻居彼此等价。

- **强正则等价** 是指对于所有顶点 u、v，$u \sim v$ 当且仅当：若$\{u, a\} \subset e$对于某条超边 e成立，则存在 $a' \in V$和 $e' \in H$，使得 $\{v, a'\} \subset e'$且 $a \sim a'$（即，$\sim$ 是弱正则等价），此外还满足 $|e| = |e'|$。换言之，u、v 与邻点相连，而这些邻点通过基数相同的超边彼此等价。

如果存在定义在 V 上的等价关系，且该关系是强（分别为弱）结构等价（分别为正则等价），则称两个顶点 u、$v \in V$ 是**强**（分别为**弱</style）结构**（分别为**正则**）**等价**的。

注 4.3。我们注意到，如果两个顶点是强结构等价的，那么它们必然具有相同的度。反之，如果两个顶点具有相同的邻居和相同的度，则它们不一定通过具有相同基数的超边连接到相同的邻居。因此，可以设想由此得到结构等价的一种中间概念。关于超图中几种不同的结构等价可能定义的更深入讨论，参见注 *4.12*。

现在出现了一个主要问题：我们能否将第 3 节中的结果推广到超图？要回答这个问题，我们首先需要将一些概念加以推广。

定义 4.4。设 $\mathcal{H} = (V, E)$ 为一个无向超图。两个顶点 u 和 $v \in V$ 之间的**游走**是超边 $u = x_0 e_1 x_1 e_2 \dots x_{n-1} e_n x_n = v$、**...**、$e_n$ 与顶点 x_0、x_1、**...**、x_{n-1}、x_n 所组成的序列，满足对于所有 $i = 0$、**...**、$n-1$，都有 x_i、$x_{i+1} \in e_{i+1}$。起点和终点为同一顶点的游走称为**环**。**路径**是所有顶点和超边均不相同的游走。路径 γ 的**长度**是其中包含的超边数量，即$\mathrm{length}(\gamma)$。**最短路径距离**是由 $d_{\mathrm{SP}}\colon \ V \times V \to \mathbb{N}$ 定义的函数：

$$\mathrm{d_{SP}}(u, v) = \min_{\gamma} \mathrm{length}(\gamma)\,,$$

其中，最小值是在 u 与 v之间的所有路径 γ 上取得的。

定义 4.5。设 n 为正整数。超图$\mathcal{H}$ 的 ***n*-邻域图**是图 H_n，其顶点集与 $\mathcal{H}$ 的顶点集相同，边集由长度为 n的超边路径给出：

$$H_n = \{\{u, v\} \mid \text{存在一条长度为 } n \text{ 的超边路径，连接 } u \text{ 与 } v\}\,.$$

我们将超图 $\mathcal{H}$对应的**图**称为 H_1。

要将 Ollivier-Ricci 曲率推广到超图，可以考虑多种方法。这些方法涉及对概率测度及其间距离的推广，其中一些推广方法会考虑超边的基数。与图类似，曲率可以针对任意两个连通顶点 x, y 定义；更一般地，也可以针对由一条或多条超边连接的任意顶点集定义。与图类似，我们只考虑属于同一连通分量的超图顶点 x, y 上的 $k(\mu_x, \mu_y)$。为了将随机游走推广到超图，我们考虑 [CDR23]中的两种方法：

定义 4.6。我们在无向超图 $\mathcal{H} = (V, H)$ 的顶点集合 V 上定义以下随机游走：

- **等节点 *(EN)*** 随机游走：我们从顶点 x 的邻点中均匀随机地选择一个 z：$\mu_x(z)$：$= \frac{1}{|N(x)|}$ 其中 $N(x) \coloneqq \{z\colon d_{\mathrm{SP}}(x, z) = 1\}$

- **等边 *(EE)*** 随机游走：首先以均匀随机的方式选择一条超边 e，满足 $|e| \geq 2$，然后选择一个顶点 $z \in e - \{x\}$，两者均为均匀随机选择： $\mu_x(z) := \frac{1}{deg\ x} \frac{1}{()-|\{x \in e : |e|=1\}|} \sum_{x,z \in e} \frac{1}{|e|-1}$

选定其中一种随机游走后，我们使用与之前相同的曲率定义：

定义 4.7。 设 x、 y 是无向超图 $\mathcal{H} = (V, H)$同一连通分量中的两个顶点。设 $\{\mu_x\}_{x \in V}$ 为上述定义的两种随机游走之一。我们定义

$$\kappa(x, y) := 1 - \frac{W_1(\mu_x, \mu_y)}{d_{\mathrm{SP}}(x, y)}.$$

当概率测度 $\{\mu_x\}_{x \in V}$ 由 *EN* 随机游走定义时，我们称 $\kappa(x, y)$为**EN-ORC** ，即(x, y) 的 *EN-ORC*；当它们由 *EE*随机游走定义时，我们称$\kappa(x, y)$ 为**EE-ORC** ，即(x, y)的 *EE-ORC*。

等边随机游走会纳入超边的基数，而等顶点随机游走只记录邻接关系。

在本节中，我们假设超图 $\mathcal{H} = (V, H)$没有自环，即所有超边均满足 $|e| \geq 2$。

首先，我们注意到，弱等价可以完全类比图的情形推导出来，从而使基于图的结果自然扩展到超图中的结构等价和正则等价。注意， H_1 涵盖了 $\mathcal{H}$各顶点之间的所有邻接关系。因此，只要关注连接是否存在，而不考虑其类型、数量或基数， $\mathcal{H}$ 和 H_1 中的等价关系对于结构等价和正则等价都会产生相同的类。根据弱等价的定义，我们得到：

引理 4.8。 $\mathcal{H}$ 中的两个顶点是弱结构等价的（分别为弱正则等价的），当且仅当它们在 H_1中是结构等价的（分别为正则等价的）。

定理 4.9。 设 $\mathcal{H} = (V, H)$ 是一个无向超图。则：

(i) x 和 y 中的两个顶点是弱结构等价的，当且仅当(x, y) 的 *EN-ORC* 等于1。***(ii)*** 如果 x 和 y 中的两个顶点是强结构等价的，那么(x, y)的 *EE-ORC* 等于1。

证明。 为说明第一种情况，我们回顾一下，在弱等价中，无论连接超边（超边）的基数如何， z 是 x 的邻点，当且仅当它是 y 的邻点；我们应证明，点对 (x, y) 的 EN-ORC 为 1。对于 EN-游走，对于每个这样的 z，根据定义 $\mu_x(z) = \mu_y(z)$，因此 $W_1(\mu_x, \mu_y) = 0$ ，或者等价地，$\kappa(x, y) = 1$。反之，如果 $\kappa(x, y) = 1$，这意味着 $\mu_x = \mu_y$ ，特别是它们具有完全相同的邻居。

对于第二种情况，如果两个顶点 x 和 y 强结构等价，那么它们通过基数相同的超边连接到相同的邻居，因此它们的 EE 随机游走相同。这意味着 $W_1(\mu_x, \mu_y) = 1$ ，特别是它们的 EE-ORC 等于 1。

□

例 4.10。 考虑图 7（左）中的超图。两个顶点 x 和 y 拥有相同的邻居、度数相同，且弱结构等价，但并非强结构等价。通过计算（x, y）的曲率即可看出：我们有，（x, y）的 *EE-ORC* 不等于1，因此根据定理 *4.9 (ii)*，x 和y 并非强结构等价。另一方面，（x, y）的 *EN-ORC* 等于1，因此 x 和 y 弱结构等价。对于后一种情况，也可以等价地考虑与 $\mathcal{H}$ 关联的图中（x, y）的 *ORC*（见图 *7*（右））。

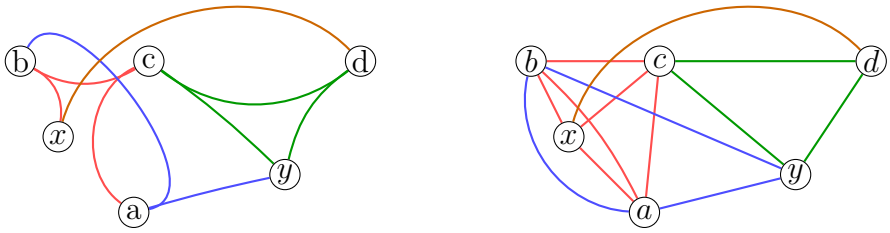

图 7：无向超图（左），其中四条超边分别以红色、绿色、蓝色和橙色标示，以及它的关联图 H_1（右）。

我们注意到，定理 4.9 的第 (ii) 项中的逆命题并不成立，而且无法通过我们在此考察的两种 Ollivier-Ricci 曲率，完整刻画超图的强结构等价。只需考察这样一个例子便可轻易看出这一点：两个顶点拥有相同的三个邻点，但它们与这些邻点之间要么通过一条超边相连，要么通过三条边相连（见图 8）。在这个例子中，对于 x 和 y，无论采用 EE 还是 EN 随机游走，三个顶点 a，b，c的概率都是 $1/3$。一方面，这引出了一个问题：是否可以采用另一种随机游走，使顶点 x 和 y 能够彼此区分。我们将在第 6 节进一步讨论这一问题。另一方面，我们注意到，尽管尚无法完整刻画，EE-ORC 仍可作为衡量近似强结构等价的有用指标，详见第 5 节。

最后，我们注意到，通过施加更多假设，可以得到定理 4.9 中命题 (ii) 的以下逆向推论：

引理 4.11。 设 x 和 y 是超图的两个顶点。假设

1. 对于 x 和 y的每个公共邻居 z，恰有一条包含$\{x, z\}$ 的超边，且恰有一条包含 $\{y, z\}$ 的超边；*2.* $d_x = d_y$。

若 EE-ORC$(x, y) = 1$，则 x 和 y 强结构等价。

证明。 由于 EE-ORC$(x, y) = 1$，我们有 $W(\mu_x, \mu_y) = 0$，因此

$$\mu_x = \mu_y.$$

特别地，μ_x 和 μ_y 具有相同的支撑，这意味着它们有相同的邻居。设 z 是 x 和 y的任意一个公共邻居。根据假设，

存在唯一一条包含 $\{x,z\}$ 的超边 e_{xz}，也存在唯一一条包含$\{y,z\}$的超边 e'_{yz}。根据 EE 游走的定义，我们有

$$\mu_x(z)=\frac{1}{d_x}\cdot\frac{1}{|e_{xz}|-1},\qquad \mu_y(z)=\frac{1}{d_y}\cdot\frac{1}{|e'_{yz}|-1}.$$

由于 $\mu_x=\mu_y$ 和 $d_x=d_y$，因此

$$|e_{xz}|=|e'_{yz}|.$$

由于 z 是任意的，这个等式对 x 和 y的每个公共邻居 z 都成立。因此， x 和 y 通过基数相同的超边连接到相同的邻居，这意味着 x 和 y 是强结构等价的。

□

注 4.12（超图中结构等价概念的层级）。 上述引理的证明依赖于相当强的局部假设，最值得注意的是，连接一个顶点与其每个邻点的超边具有唯一性，且各顶点的度相等。

这些假设足以确保 *EE*测度 $\mu_x=\mu_y$相等会导致局部超边基数完全匹配，但它们并非 *EE*随机游走或 *EE-ORC* 定义所固有的条件。这促使我们引入一系列强度递增的结构等价概念，以反映可从 μ_x 和基于曲率的量中恢复的不同层次的组合信息。我们指出，在以下所有等价概念中， x 和 y 的邻居均相同，这意味着这些概念都强于弱结构等价。

(i) EE 度量等价。 如果EE 度量等价，则称两个顶点 x 和 y 满足

$$\mu_x=\mu_y.$$

这是本文所讨论的最弱等价概念，其特征恰好是满足条件$\mathrm{EE\text{-}ORC}(x,y)=1$。它刻画了相对于单步 *EE* 扩散的等价性，但不能唯一确定其底层的超边关联结构。

(ii) 加权邻域等价。 如果对于每个顶点 z，加权邻域等价 ，则称顶点 x 和 y 满足

$$\sum_{x,z\in e}\frac{1}{|e|-1}=\sum_{y,z\in e}\frac{1}{|e|-1}.$$

我们将这一概念称为加权邻域等价，因为它通过比较分配给每个邻点的超边诱导权重总和来比较顶点；这些权重由所有关联超边的$(|e|-1)^{-1}$ 形式贡献聚合而成。在无环假设下，当 $d_x=d_y$时，根据 *EE* 随机游走的定义，这一概念与 *EE* 度量等价一致。

（***iii***）多重集结构等价。 顶点 x 和 y 是多重集结构等价 的，当且仅当它们拥有相同的邻居，并且对于每个公共邻居 z，多重集

$$\big\{|e|:\ e\in E,\ \{x,z\}\subseteq e\big\}\quad and\quad \big\{|e|:\ e\in E,\ \{y,z\}\subseteq e\big\}$$

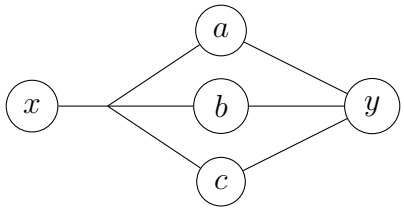

图 8：顶点 x 和 y 拥有相同的邻居、不同的度，并且是弱结构等价的。此外，我们有 $(x,\ y)$ 的 EE-ORC 等于 1，但 x 和 y 并不强结构等价。

都相同。这一概念保留了与每个邻居关联的超边基数的完整谱，同时允许存在多个超边，且不要求在单个超边之间建立双射。它严格强于 *EE* 度量等价，但弱于完整的组合关联匹配。

(iv) 强结构等价。 我们注意到，在最初定义的这一概念中，x 和 y 的局部超边关联模式在完全组合意义下是一致的；具体而言，连接 x 和 y 与每个公共邻居的匹配超边（包含重数）相同，并且保持基数不变。这是本文考虑的最强概念，对应于关联星的局部同构。

各概念之间的关系。 上述定义构成了一条自然的蕴含链：

$$\text{强等价} \implies \text{多重集合等价} \implies \text{加权邻域等价} \implies \textit{EE}\ \text{度量等价。}$$

一般而言，反向蕴含均不成立。上述引理表明，在唯一超边关联和度相同等附加假设下，*EE* 度量等价足以恢复强结构等价。没有这些假设时，基于曲率的量无法区分诱导出相同 *EE* 转移测度的不同局部关联模式。

我们在此不再进一步探讨这些概念的性质，而将系统研究留待未来工作。

我们通过给出图相关结果在超图上的两个直接推广来结束本节。首先，考虑 H_2，我们得到以下结果：

推论 4.13。 设 $\mathcal{H}$ 是一个无向超图。则有：

(i) $\mathcal{H}$ 的弱结构类是 H_2 的连通分量中完全子图的顶点集。*(ii)* 当且仅当同一子图中每一对顶点的 *EN-ORC* 曲率在 $\mathcal{H}$ 中均等于 1，且该子图是满足这一性质的极大子图时，H_2 的连通分量中的完全子图才会诱导出 $\mathcal{H}$ 的弱结构类。*(iii)* H_2 的连通分量会诱导出 $\mathcal{H}$ 的弱正则类。此外，如果 H_2 有两个连通分量，这些连通分量会诱导出一个弱正则划分，且该划分不被任何非平凡弱正则划分真包含。

(iv) 若 H_2 的连通分量中的完全子图满足每个子图内任意一对顶点的 *EN-ORC* 在 $\mathcal{H}$ 中均等于 1，则这些完全子图会诱导出 $\mathcal{H}$ 的弱正则类。

证明。 该证明可直接由引理 4.8 和定理 3.4（对应 (i) 和 (ii)）、定理 3.6（对应 (iii)）以及推论 3.9（对应 (iv)）得出。 □

以下是定理 3.7 向超图的直接推广：

推论 4.14。 设 $\mathcal{H}$ 是无向超图。若每个顶点至少有两个邻居，则 H_n 的连通分量诱导出 $\mathcal{H}$ 的弱正则类。

证明。 该证明可直接由引理引理 4.8 和定理 3.7 得出。 □

5 图和超图的近似结构等价

正如我们所见，结构等价是一个限制性很强的概念。因此，在实践中，人们会使用近似结构等价的度量，例如余弦相似度，见第 2.5 节。我们的几何刻画提出了另一种近似结构等价的度量，它不仅适用于图，也适用于超图：我们知道，图中的两个顶点 $x,\ y$ 结构等价，当且仅当 $(x,\ y)$ 的 Ollivier-Ricci 曲率等于 1。这一结果可直接推广到超图：弱结构等价可通过 EN-ORC 刻画，而 EE 度量等价可通过 EE-ORC 刻画（见定理 4.9 和注 4.12）。放宽这些等式后，我们便得到近似结构等价的自然度量：对应曲率越接近 1，这两个顶点就越结构等价。本节旨在更好地理解 Ollivier–Ricci曲率为无向图和超图提供的信息类型。

我们首先做一些观察。

- 当且仅当 x, y 结构等价时，曲率 $\kappa(x, y)$ 等于 1。我们注意到，在这种情况下，$d_{\mathrm{SP}}(x,\ y)\ \neq 1$。类似地，对于超图中的顶点，弱结构等价对应 EN-ORC，而 EE-测度等价对应 EE-ORC（后者见注 4.12）。

- 对于图和超图，如果 $d_{\mathrm{SP}}(x, y)\ > 2$，那么 x 和 y 没有共同邻居，我们称它们为**完全不相似**。

- 在图中，如果 $d_{\mathrm{SP}}(x, y)\ = 2$： $-1 < \kappa(x, y)\ \leq 1$。因此，这对顶点的曲率越接近 1，它们在结构上就越相似；曲率越接近 -1，它们就越不相似。对于超图中的顶点 x 和 y，同样的结论也成立，其中弱等价（分别为 EE 度量等价）基于 EN-ORC（分别基于 EE-ORC ）。

- 如果 $d_{\mathrm{SP}}(x,y) = 1$：$-2 < \kappa(x,y) < 1$；与前一种情况类似，曲率越接近 1（在这种情况下，它永远不会等于 1），这两个顶点就越相似；反之，这个数越接近 -2，它们就越不相似。将 $\kappa(x,y)$ 替换为 EN-ORC 或 EE-ORC，可分别得出超图中弱等价和 EE 度量等价的类似结论。

因此，对于距离不超过 2 的顶点，曲率是介于 -2 和 1 之间的一个数。另一方面，余弦相似度是介于 0 和 1 之间的一个数。我们注意到，余弦相似度已针对图提出；直接将其推广到超图通常会得到不同类型的度量，因为对于超图，顶点的度通常小于其相邻顶点的数量。对于超图，我们认为最值得关注的是直接考察由注 4.12 中讨论的随机游走等价概念所给出的近似等价度量。我们将这些随机游走等价概念的探索留待未来工作。

曲率与余弦相似度之间有何关联？为建立图中这两个概念之间的联系，假设我们想以最优方式将沙子从 μ_x 移动到 μ_y 孔中。我们注意到，取决于 x 和 y之间的距离，一些沙子和孔可能已经重合；正如 [JL14, EJ20], 中所述，这部分的数量等于 $\frac{\eta_{xy}}{d_x \vee d_y}$，其中 $d_x \vee d_y$是 x 和 y的度的最大值。重要的是，总存在一种最优运输方案，不移动已经位于孔中的沙子。正如 [EJ20], 中所观察到的，对于两个相邻顶点 x 和 y，有 $W_1(\mu_x, \mu_y) = \sum_{i=0}^{3} i\mu_i$，其中 μ_i 是在最优运输方案中需要移动距离为 i的沙子的数量；如果 x 和y 不是邻点，我们只需将求和的范围扩展到 $d(x, y) + 2$，因为这是沙子到孔的最大距离。由于存在一种最优运输方案，能让已经位于孔中的沙子保持不动，因此对于这样的运输方案，我们有 $\mu_0 = \frac{\eta_{xy}}{d_x \vee d_y}$。这一观察使我们能够得到余弦相似度与 Ollivier-Ricci 曲率之间的如下关系：

Pro命题 5.1. 设 x、y 是图中任意一对不同的顶点。我们有 ：

$$3\frac{\sigma_{xy}\sqrt{d_x d_y}}{d_x \vee d_y} - 2 \leq \kappa(x,y) \leq 1 - \frac{1-\sigma_{xy}}{d(x,y)}\,.$$

若对余弦相似度取不同的值并代入这些不等式，可得 $\kappa(x,y) = 1$ 对于 $\sigma_{xy} = 1$，此外：

- 若 $\sigma_{xy} = 0$，则有 $-2 \leq \kappa(x,y) \leq 1 - \frac{1}{d_x(x,y)}$；[1]
- 若 $\sigma_{xy} = s$ 满足 $0 < s < 1$，则有 $3s\frac{\sqrt{d_x d_y}}{d_x \vee d_y} - 2 \leq \kappa(x,y) \leq 1 - \frac{1-s}{d_x(x,y)}$.

证明。首先证明上界。设 μ_i 表示在一个最优运输方案中，跨越距离 i 移动的质量。于是我们有

$$W_1(\mu_x, \mu_y) = \sum_{i>0} i\,\mu_i \geq \sum_{i>0} \mu_i = 1 - \mu_0\,,$$

[1]对于图，这可由三角不等式推出，参见 [BCL+18，引理 4.1]。更一般地，正如 [BCL+18] 中所指出的，这可由康托罗维奇-鲁宾斯坦距离在质量减除下的不变性推出 [Vil03，推论 1.16]。

因为 $\sum_{i\geq 0}\mu_i = 1$。不失一般性，我们可以假设 μ_0 等于两个概率测度共有的质量，因为存在一种运输计划可以使这部分质量保持不动。

因此，

$$\begin{aligned} W_1(x,y) &\geq 1-\mu_0 \\ &= 1-\frac{\eta_{xy}}{d_x \vee d_y} \\ &= 1-\frac{\sigma_{xy}\sqrt{d_x d_y}}{d_x \vee d_y} \\ &\geq 1-\sigma_{xy}\,, \end{aligned}$$

其中，最后一个不等式成立是因为 $\frac{\sqrt{d_x d_y}}{d_x \vee d_y} \leq 1$。因此，我们得到

$$\kappa(x,y) = 1-\frac{W_1(\mu_x,\mu_y)}{d(x,y)} \leq 1-\frac{1-\sigma_{xy}}{d(x,y)}\,.$$

对于下界，我们首先观察到：

$$W_1(\mu_x,\mu_y) = \sum_{i=1}^{d(x,y)+2} i\,\mu_i \leq (d(x,y)+2)\sum_{i=1}^{d(x,y)+2}\mu_i = (d(x,y)+2)(1-\mu_0)\,.$$

因此，我们得到以下下界：

$$\begin{aligned} \kappa(x,y) = 1-\frac{W_1(\mu_x,\mu_y)}{d(x,y)} &\geq 1-\frac{(d(x,y)+2)(1-\mu_0))}{d(x,y)} \\ &= 1-(d(x,y)+2)\frac{(1-\frac{\sigma_{xy}\sqrt{d_x d_y}}{dx\vee d_y})}{d(x,y)} \\ &= \frac{\frac{\sigma_{xy}\sqrt{d_x d_y}}{dx\vee d_y}(d(x,y)+2)-2}{d(x,y)}\,. \end{aligned}$$

此外，我们指出，若对于每条边 $\{x,y\}$ 都有 $\kappa(x,y) \geq \kappa_0$，则对于任意一对顶点，都有 $\kappa(x,y) \geq \kappa_0$（这是 W_1 的三角不等式的直接推论）。因此，我们可以取 $d(x,y) = 1$，从而得到下界

$$\kappa(x,y) \geq 3\frac{\sigma_{xy}\sqrt{d_x d_y}}{dx \vee d_y} - 2$$

对于任意 $x \neq y \in V$。

最后，我们注意到，如果 $\sigma_{xy} = 1$，那么必有 $d_x = d_y$ 和 $d(x,y) = 2$。因此，下界简化为 **1**，上界也同样如此，于是得到 $1 \leq \kappa(x,y) \leq 1$。 □

对命题 5.1 的陈述，另一种解读方式是：它给出了两个顶点的曲率的易于推导的界。由于这些界仅由公共邻居所占的比例决定，因此可以用非常简单的方式理解。例如，如果两个顶点完全不相似，那么可以说曲率至多为 $1 - \frac{1}{d_x}(\cdot, y)$。特别地，对于相邻顶点，曲率必定为零或负值。

6 结论与开放问题

在这项工作中，我们将图和超图上的 Ollivier-Ricci 曲率与随机游走联系起来，并将其与结构等价和正则等价的概念相联系。我们认为，这项工作只是朝这个方向迈出的第一步，而我们建立的联系也引出了许多有趣的问题。其中一个问题涉及超图上的随机游走。我们通过曲率研究超图结构等价的刻画，凸显了一个事实：目前已有的超图随机游走概念无法辨别超图中的局部邻接模式；例如，所有现有的随机游走概念都会为图 8 中的节点 x 和 y 赋予相同的值。我们可以设想这种区分可能很重要的应用场景，例如社会系统建模，其中超边可以表示共同出席活动，而随机游走则可以编码一个人在同一活动期间与另一个人交谈的概率。在这种情形下，节点 x 和 y 的情况截然不同，因此应当通过另一种类型的随机游走更好地加以刻画。

在将曲率刻画推广到超图时，我们注意到，基于随机游走的不同类型等价关系更适合捕捉不同的局部关联相似模式。关于超图这些新型等价关系的系统研究，我们留待未来工作开展。

最后，我们注意到，正则等价和 Ollivier-Ricci 曲率近期已应用于机器学习；正则等价已用于表示学习 [TCW+18]，而 Ollivier-Ricci 曲率则应用于研究图神经网络中的瓶颈和过度压缩 [TCE+25]。因此，这引出了一个问题：本文所揭示的联系，能为几何深度学习中的问题带来哪些新的见解?

7 致谢

我们感谢马克斯·普朗克学会（MPG）与法国国家科学研究中心（CNRS）之间的 SALTO 交流计划，以及 Fondation Math´ematiques Jacques Hadamard 提供的资助。

参考文献

[BCL+18] D. P. Bourne, D. Cushing, S. Liu, F. M¨unch, and N. Peyerimhoff. 图的奥利维耶–里奇闲置函数。*SIAM J. Discret. Math.*，32(2):1408–1424，2018年1月。

[BE93] Stephen P. Borgatti 和 Martin G. Everett。计算正则等价的两种算法。Social Networks，15(4):361–376，1993年。[BGHS23] C. Bick、E. Gross、H.A. Harrington 和 M.T. Schaub。什么是高阶网络？SIAM Review，65(3)，2023年。[BGK19] B. Breˇsar、N. Gastineau 和 S. et al. Klavˇzar。积图的精确距离图。Graphs and Combinatorics，35:1555–1569，2019年。[BJL13] Frank Bauer、J¨urgen Jost 和 Shiping Liu。奥利维耶 -Ricci 曲率与归一化图拉普拉斯算子的谱，2013 年。[Bor89] Stephen Peter Borgatti。图、超图和矩阵中的正则等价。博士论文，加州大学欧文分校，1989年。[CDR23] Corinna Coupette、Sebastian Dalleiger 和 Bastian Rieck。超图的奥利维耶-Ricci 曲率：统一框架，2023年。[EB93] Martin Everett 和 Stephen Borgatti。正则等价：一般理论。Journal of Mathematical Sociology，19:29–52，1993年。[EB19] Martin Everett 和 Steve Borgatti。计算正则等价：实践与理论问题。17(1):31–42，2019年。出版方版权：© 2019，Uni- verza v Ljubljani。[EJ20] Marzieh Eidi 和 J¨urgen Jost。有向超图的奥利维耶-Ricci 曲率。Scientific Reports，10(1):12466，2020年。开放获取。[Fre04] Linton Freeman。社会网络分析的发展：科学社会学研究。2004年。[JL14] J¨urgen Jost 和 Shiping Liu。图上的奥利维耶的 Ricci 曲率、局部聚类与曲率维数不等式。Discrete & Computational Geometry，51(2):300–322，2014年。[LY10] Yong Lin 和 Shing-Tung Yau。Mathematical Research Letters，17(2):343–356，2010年。[MOR25] Nima Motamed、Nina Otter 和 Emily Roff。社会系统的余代数分析。2025年。准备中。[Oll07] Yann Ollivier。度量空间上马尔可夫链的 Ricci 曲率，2007年。[Oll09] Yann Ollivier。度量空间上马尔可夫链的 Ricci 曲率。Journal of Functional Analysis，256(3):810–864，2009年。[Sim83] Slobodan K Simi´c。线图和 n-距离图的图方程。Publications de l’Institut de Math´ematiques de Beograd，33(47):203–216，1983年。

[STTV25] G. Squillace、M. Tribastone、M. Tschaikowski 和 A. Vandin。通过划分细化实现近似正则等价。《应用网络科学》，10，2025。[TCE+25] Diaaeldin Taha、James Chapman、Marzieh Eidi、Karel Devriendt 和 Guido Montufar。揭开关系结构中拓扑消息传递的奥秘：单纯形消息传递中过度压缩的案例研究。载于第十三届国际学习表征会议 *(ICLR)*，2025。[TCW+18] Ke Tu、Peng Cui、Xiao Wang、Philip S. Yu 和 Wenwu Zhu。基于正则等价的深度递归网络嵌入。载于第 *24* 届 *ACM SIGKDD* 国际知识发现与数据挖掘会议论文集，KDD '18，第 2357–2366 页，纽约，NY，美国，2018。计算机协会。[Vil03] C. Villani。最优传输专题。美国数学会，2003。[WF94] Stanley Wasserman 和 Katherine Faust。《社会网络分析：方法与应用》。《社会科学中的结构分析》。剑桥大学出版社，1994。[WR83] Douglas R. White 和 Karl P. Reitz。关系网络上的图同态与半群同态。*Social Networks*，5(2)：193–234，1983。[YL11] Shing-Tung Yau Yong Lin、Linyuan Lu。图的里奇曲率。《东北数学杂志》*(2)*，63(4)：605–627，2011。[ZHS07] Dengyong Zhou、Jiayuan Huang 和 Bernhard Sch¨olkopf。利用超图学习：聚类、分类与嵌入。载于《神经信息处理系统进展》，第 19 卷，第 1601–1608 页，2007。