arXiv:2302.09017v1 [math.CO] 17 Feb 2023

随机图的多邻复形的渐近拓扑

´WOJCIECH MATYSIAK 和 JAN SPALINSKI

Abstract. 我们引入图的多邻居复形，它是一种单纯复形，其中单形是图的一个子集，且这些顶点拥有足够多的共同邻点。我们研究此类复形在 Erd˝os–Renyi 随机图中的渐近同调性质，并得到若干消失与非消失结果。我们利用这一构造对含噪合成点云进行拓扑数据分析分类，取得了与标准方法相当的良好准确率。仅有一个邻点的情形此前已由 Mathew Kahle 研究过。

1. Introduction

本文旨在引入一族与（随机）图相关联的滤过单纯复形，并研究其性质。我们的研究始于 M. Kahle 的论文 [11]。

给定一个图 G 和一个正整数 m，多邻居复形 $N_m(G)$ 的一个 q-单形是 G 中由 $q+1$ 个顶点组成且至少有 m 个共同邻点的集合。$m=1$ 的情形已在 [11] 中研究。

本研究的动机源于拓扑数据分析（TDA，参见 G. Carlsson [4], Carlsson 和 A. Zomorodian [5] 以及 H. Edelsbrunner、D. Letscher 和 Zomorodian [7]）。从一个点云（严格来说，是度量空间的有限子集）和一个邻近参数出发，可以构造一族单纯复形。利用这族复形，可以获得点云的同调不变量（持久同调）。这类复形最自然的构造是 Rips-Vietoris（又称团）复形。这种构造用于数据分析的一个缺陷在于所涉及的复形规模庞大（仅顶点数量就与点云中的元素数量相同）。我们的构造可概括为以下复形的参数化版本：

2020数学主题分类。主要分类：55N31；次要分类：05C80, 62R40。关键词与短语。随机图，随机复形，图的邻域复形，多邻居复形，同调，拓扑数据分析。

邻域复形，其中额外参数（m ——邻点数量）反映了给定单形周围图的密度。

本文结构如下。第 2 节介绍多邻居复形，并给出几个简单示例。第 3 节（第 4 节）分别介绍关于随机图的多邻居复形约化同调的若干消失（非消失）定理。第 5 节介绍多邻居复形在拓扑数据分析中的一种应用。点云可以通过邻近参数 r和密度参数 m以自然的方式确定一个双分次持久图。我们的实验以拓扑熵作为欧几里得空间中含噪点云的分类器，结果表明，与邻域复形、Vietoris–Rips复形或Alpha复形相比，多邻居复形能使随机森林分类器获得更高的准确率。

C. Previte 的博士论文 [4]探讨了图的邻域复形这一思想的所有可能变体。

2. The multineighbor complex of a graph

在本节中，我们介绍并研究图的多邻居复形。图 G 是指无向图，其顶点记为 $\mathbf{V}(G)$，边记为 $\mathbf{E}(G)$。

在下面的定义中， m 扮演密度参数的角色，而q 是单形的维数。下方的图 1 展示了一个每个顶点至少有 3 个邻点的图。

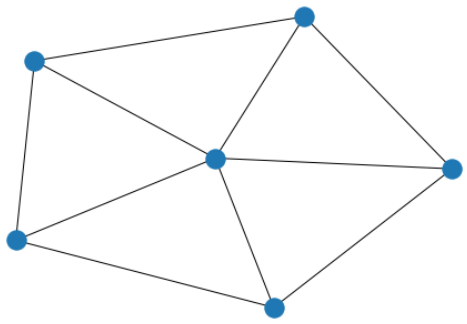


Figure 1. 每个顶点至少有 3 个邻点的图

定义。 设 G 为一个图，且 m 和 q 为非负整数。图 G的 *The multineighbor complex* 记作 $N_m(G)$，其定义如下：若一组 $q+1$ 个顶点在 G中至少有 m 个共同邻点，则将其视为一个 q-单形。

当 $m = 1$ 时，我们得到通常定义的邻域复形，该复形在 [11], 中被研究，记作 $N(G)$。最简单的例子是具有 n 个顶点 K_n的完全图，其多邻点复形是维数为 $n - 1$的标准单形的一个骨架。

给定一个单纯复形 L，我们令 $L^{(i)}$ 表示 L 的 i 维骨架，即由 L 中维数至多为 i 的单形组成的集合。

引理 1。令 K_n 为以 n 个顶点构成的完全图。K_n 的多邻居复形是以 K_n 的顶点为顶点、维数为 $n-1$ 的标准单形的 $n-1-m$ 维骨架

$$N_m(K_n) = (\Delta^{n-1})^{(n-1-m)}$$

证明。我们先明确描述前两种情形。若 $m = 0$，则 K_n 中每个由 $q+1$ 个顶点组成的集合都张成一个单形，因此多邻居复形就是 Δ^{n-1}，即由 K_n 的顶点所构成的 n 元集合的所有非空子集组成的这个单纯复形。接下来，若 $m = 1$，则一个 $q+1$ 元组在 K_n 中有邻居，当且仅当 $q + 1 < n$，或 $q < n - 1$。因此，对于 $m = 1$，我们得到 Δ^{n-1} 的 $n - 2$ 骨架。一般而言，由于该图是完全图且具有 n 个顶点，一个（$q + 1$）元组有邻居，当且仅当 $q + 1 + m \leq n$。结论得证。□

图2和图3展示了两个简单图的多邻居复形的1-骨架，其中邻点数分别为 $m = 1$和$m = 2$。

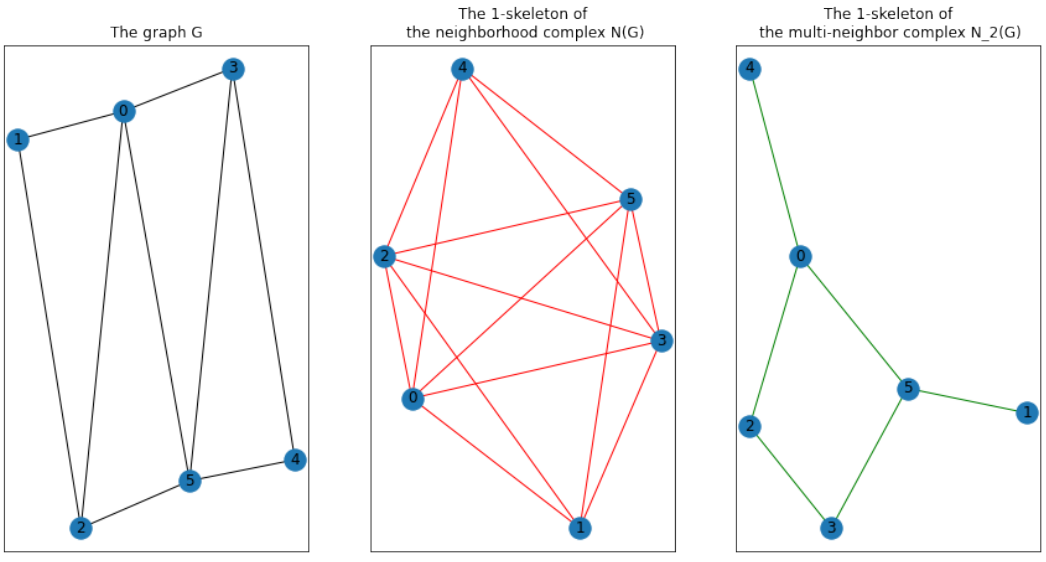


FIGURE 2. 图以及具有 $m = 1$ 和 $m = 2$ 的多邻居复形的1-骨架

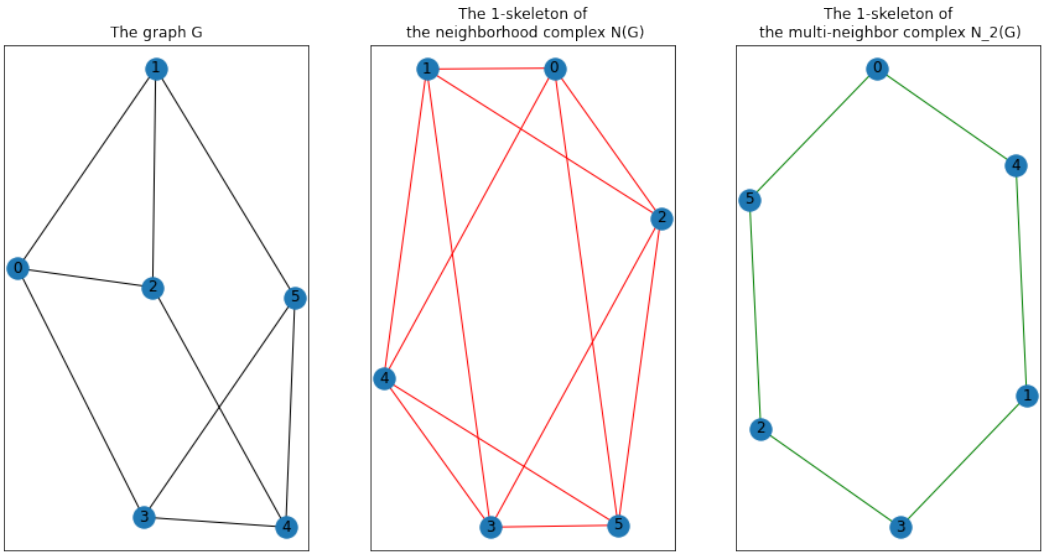


FIGURE 3. 一个图以及具有 $m = 1$ 和 $m = 2$ 的多邻域复形的 1-骨架

3. VANISHING THEOREMS FOR REDUCED HOMOLOGY

我们将需要A. Bjorner的如下结果（它是D. Quillen的序同伦定理的一个推论）。回顾一下，给定一个偏序集 P，序复形 $\Delta(P)$是一个单纯复形，其顶点为 P 的元素，其 q 单形为 P中元素构成的 $q+1$ 元链： $p_0 < p_1 < \cdots < p_q$。给定一个单纯复形 L，我们令 $P(L)$为其按包含关系排序的面的偏序集。可以验证， $\Delta(P(L))$是 L的重心细分，因此这两个单纯复形同胚。操作 $P(\cdot)$和 $\Delta(\cdot)$使我们能够将单纯复形和偏序集视为拓扑空间模型中本质上等价的对象。特别地，当我们考虑偏序集 P的某个拓扑概念时，这意味着我们是在考虑单纯复形 $\Delta(P)$的底层空间的该概念。

推论（A. Bjorner，第26页，推论11.8）： 设 P 为一个偏序集，且 $f : P \to P$ 为一个保序映射，使得

$$f(x) \geq x \qquad for\ all \qquad x \in P$$

则 f在 P 和 $f(P)$之间诱导出一个同伦等价。

此外，若对于所有 $x \in P$均有 $f^2(x) = f(x)$，则 $f(P)$ 是 P的强形变收缩。

我们将多次用到以下定理，它是 [11]中 Kahle引理2.3 的多邻居类比。

定理1. 设 H 为一个图， a 和 b 为正整数， m 为满足 $1 \leq m \leq a+b-2$的整数。若 H 不包含完全

二部图 $K_{a,\ b}$，则 $||N_m\ (H)\ ||$ 强形变收缩到一个维数至多为 $a+b-m-2$ 的复形。

证明。我们将证明其逆否命题。因此，假设 $||N_m(H)||$不能变形收缩到一个维数至多为 $a+b-m-2$的复形，因而每个变形收缩的维数都至少为 $a+b-m-1$。

对于一个顶点 $x \in \mathbf{V}(H)$，令 $\Gamma(x)$ 为 x的邻点集合。若 X是 $\mathbf{V}(H)$ 的一个子集，我们定义

$$\Gamma(X) = \bigcap_{x \in X} \Gamma(x)$$

容易看出， Γ 是从 $\mathbf{V}(H)$ 的所有子集构成的集合到其自身的一个序反转映射。因此， Γ^2 是一个序保持映射。这里有一个简单的例子。考虑图 H ，如图 4 所示。

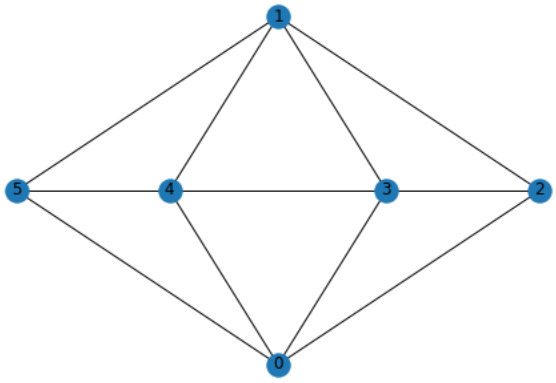


FIGURE 4. 用于描述该Γ 构造的图 H

We have $\Gamma(\{0\}) = \{2,3,4,5\}$ and $\Gamma(\{2,3,4,5\}) = \{0,1\}$.

令 X 为 $\mathbf{V}$ H 的任意子集。由于 X 中的每个元素都与 $\Gamma(X)$ 中的每个元素相连，我们得出 $X \subset \Gamma^2(X)$。由于 Γ反转包含关系，我们还有 $\Gamma^3(X) \subset \Gamma(X)$。在前一个公式中令 $X = \Gamma(Y)$，得到 $\Gamma(Y) \subset \Gamma^3(Y)$。

我们得出，对于所有 $X \subset \mathbf{V}(H)$，均有 $\Gamma^3(X) = \Gamma(X)$。

令 $P = P(N_m(H))$ 表示 $N_n(H)$ 的单形偏序集，而这些单形又是 $\mathbf{V}(H)$ 的某些子集。

令 $v : P \to P$ 为由下式定义的映射

$$v(X) = \Gamma^2(X) \cap N_m(H)$$

下面我们证明 $v^2 = v$。首先证明 $v(X) \subset v^2(X)$。令 $Y = v(X)$。我们有

$$Y \subset N_m(H) \qquad \text{and} \qquad Y \subset \Gamma^2(Y)$$

因此

$$Y \subset N_m(H) \cap \Gamma^2(Y) = v(Y)$$

回想起 $Y = v(X)$，我们有

$$v(X) \subset v(v(X)) = v^2(X)$$

接下来，我们证明反向包含关系，即 $v^2(X) \subset v(X)$。

首先，注意到 v 是单调的，也就是说

$$X \subset Y \subset N_m(H) \implies v(X) \subset v(Y)$$

这是因为我们知道，对于 $X \subset Y$， $\Gamma^2(X) \subset \Gamma^2(Y)$，因此

$$v(X) = \Gamma^2(X) \cap N_m(H) \subset \Gamma^2(Y) \cap N_m(H) = v(Y)$$

因此我们有，

$$\begin{aligned} v^2(X) &= \Gamma^2\left(\Gamma^2(X) \cap N_m(H)\right) \cap N_m(H) \\ &\subset \Gamma^2\left(\Gamma^2(X)\right) \cap N_m(H) \\ &= \Gamma^2(X) \cap N_m(H) = v(X) \end{aligned}$$

这就完成了对 $v^2(X) = v(X)$ 的证明。

根据 Bjorner 中的推论 11.8， [3], $v(P)$ 是 P 的强形变收缩。

根据证明开头所作的假设， $v(P)$ 的维数至少为 $a+b-m-1$。因此，它至少包含一个该维数的单形，例如一个维数为 q 且满足 $q \geq a+b-m-1$ 的单形。

$v(P)$ 中的这类单形是 $N_m(H)$ 中形如以下形式的严格递增单形序列：

$$X_1 \subsetneq X_2 \subsetneq \cdots \subsetneq X_{q+1} \qquad (*)$$

由于 X_1 非空且包含关系都是严格的， X_i 至少包含 $1 \leq i \leq q+1$的 i 个顶点。

令 $Y_i = \Gamma(X_i)$，则有

$$Y_1 \supsetneq Y_2 \supsetneq \cdots \supsetneq Y_{q+1} \qquad (**)$$

由于 $\Gamma^3 = \Gamma$，上面的包含关系立即可知是严格的。这里需要回顾一下，$(*)$ 是 $v(P)$ 中的一个单形，因此每个 X_i 都具有 $X_i = \Gamma^2(U_i)$ 的形式，其中某个 U_i 属于 P。

由于 X_{q+1} 是 $N_m(H)$ 中的一个单形，H 中至少有 m 个顶点位于 X_{q+1} 之外，并且与 X_{q+1} 的每个顶点相连。

由于序列 $(**)$ 是严格的，且 Y_{q+1} 至少包含 m 个元素，因此 Y_{q+1-i} 至少包含 $m+i$ 个顶点。

现在需要考虑两种情况：

(1) $m \leq$ 最大值$\{a, b\}$,

(2) $m >$ 最大值$\{a, b\}$。

情形 1。 不失一般性，假设 $m \leq b$。

设定 $i = b - m$，我们看到 Y_{q+1-i} 至少有 $m + (b - m)$ 个顶点。

回想一下我们对 q 的选择，可知必有

$$q + 1 - i \geq a + b - m - 1 + 1 - (b - m) = a$$

因此， X_{q+1-i} 至少有 a 个顶点，这些顶点与 Y_{q+1-i} 的顶点一起在 H 中张成一个完全二部图。这与定理陈述中的假设矛盾。

情形 2. 现在我们有 $m > a$ 和 $m > b$，且 X_{q+1} 的每个顶点都与 Y_{q+1} 的每个顶点相连，而它们各自至少有 m 个顶点（因为它们属于 $N_m(H)$）。因此， X_{q+1} 和 Y_{q+1} 的顶点在 H 中张成一个完全二部子图，其两侧子集的基数分别大于 a 和 b。这再次与定理陈述中的假设矛盾。

□

令 $G(n,\ p)$ 为一个有 n 个顶点的随机图，其中每条边独立于其他边，以 p 的概率出现（即这是一个 Erd˝os– R´enyi 随机图 [8]）。回顾一下，若一个序列a_n 是 **o(1)**，则表示当 n 趋于无穷时，它的极限为零。若一个对象 X 渐近几乎必然具有性质P，我们称之为渐近几乎必然（缩写为 a.a.s.），即 $\lim_{n\to\infty} \mathbb{P}(X \in P) = 1$。

引理 2. 设 m 为一个固定的正整数，且 $p = p(n)$ 为一个取值于区间 $[0, 1]$ 的单调序列。若

$$\binom{n}{i} \left(1 - p^{mi}\right)^{\binom{n-i}{m}} = \mathbf{o}(1)$$

则 $N_m(G(n, p))$ 渐近几乎必然是 $(i - 2)$-连通的。

证明。若每个由 i 个顶点组成的集合都张成一个面，则称单纯复形为 i-邻点的；这一性质意味着该单形是 $(i-2)$-连通的。在 $G(n,\ p)$ 中，给定的 i 个顶点集合以 p^{mi} 的概率拥有给定的 m 个顶点作为邻点。因此，其反面发生的概率为 $1 - p^{mi}$。于是，给定的i 个顶点集合在 $G(n,\ p)$ 中没有一组 m 个邻点的概率等于 $(1 - p^{mi})^{\binom{n-i}{m}}$。因此，最后，没有任何 i 个顶点的集合拥有 m 个邻点的概率至多为

$$\binom{n}{i} \left(1 - p^{mi}\right)^{\binom{n-i}{m}}$$

□

命题 1. 若 $p = \frac{1}{2}$ 且 $\varepsilon > 0$，则渐近几乎必然 $H_{l-2}(N_m G(n, p)) = 0$对于 $l \leq \left(1 - \frac{\varepsilon}{m}\right) \log_2 n$。

证明。 接下来，我们将使用以下著名不等式:

$$1-x \leq \exp(-x), \qquad x \in \mathbb{R}$$
$$\frac{n^k}{k^k} \leq \binom{n}{k}, \qquad 1 \leq k \leq n$$

根据上述引理，只需证明

$$\binom{n}{l}\left(1-p^{ml}\right)^{\binom{n-l}{m}} = \mathbf{o}(1)$$

我们有

$$\begin{aligned}
\binom{n}{l}\left(1-p^{ml}\right)^{\binom{n-l}{m}} &\leq n^l \exp\left(-\left(\frac{1}{2}\right)^{ml}\binom{n-l}{m}\right)\\
&\leq n^l \exp\left(-\frac{1}{2^{(m-\varepsilon)\log_2 n}}\binom{n-l}{m}\right)\\
&\leq n^l \exp\left(-\frac{1}{n^{(m-\varepsilon)}}\binom{n-l}{m}\right)\\
&\leq n^l \exp\left(-n^{(-m+\varepsilon)}\cdot\frac{(n-l)^m}{m^m}\right)\\
&= \exp\left(l\ln(n)-\left(\frac{1}{m^m}\right)n^{(-m+\varepsilon)}(n-l)^m\right)\\
&= \exp\left(l\ln(n)-\left(\frac{1}{m^m}\right)\left[n^\varepsilon - m\,l\,n^{\varepsilon-1}+\ldots\right]\right)\\
&= \mathbf{o}(1)
\end{aligned}$$

最后一个等式是由 $\lim_{x\to\infty} x^\varepsilon \ln x = \infty$ 这一事实得到的。

□

命题 2。 如果 $p=\frac{1}{2}$ 和 $\varepsilon>0$，那么渐近几乎必然有 $H_{l-m+1}(N_m G(n,p)) = 0$ 对于 $l \geq (4+\varepsilon)\log_2 n$。

证明。 令 $j=k=\left\lceil \frac{l}{2} \right\rceil$。在 $G(n,p)$ 中找到一个 (j,k)-二部子图的概率上界为

$$\binom{n}{j}\binom{n}{k}\left(\frac{1}{2}\right)^j\left(\frac{1}{2}\right)^k$$

我们将证明，对于给定的 l（因此也适用于 j 和 k），上述表达式为 **o(1)**。

$$\begin{aligned}\binom{n}{j}\binom{n}{k}\left(\frac{1}{2}\right)^{j}\left(\frac{1}{2}\right)^{k} &\leq n^{j}n^{k}\frac{1}{2^{jk}}\\ &\leq n^{j+k}\frac{1}{2^{jk}}\\ &\leq n^{l+2}\frac{1}{2^{jk}}\\ &\leq n^{l+2}\frac{1}{2^{\frac{l^2}{4}}}\\ &\leq n^{l+2}\frac{1}{\left(n^{4+\varepsilon}\right)^{\frac{l}{4}}}\\ &\leq n^{l+2}\frac{1}{\left(n^{l+\varepsilon l/4}\right)}\\ &\leq n^{2-\frac{\varepsilon}{4}l}=\mathbf{o}(1)\end{aligned}$$

根据定理1， $N_m(G(n,p))$ 变形收缩到一个维数至多为

$$j+k-m-2\leq\frac{l}{2}+1+\frac{l}{2}+1-m-2=l-m$$

因此， $N_m(G(n,\ p))$ 的同调在 $l-m+1$ 及以上维数中消失。

□

命题3。 令 l 为非负整数且 $p=n^{\alpha}$，其中$\alpha\in\left[\begin{matrix}-1\\ l+2\end{matrix},\ 0\right]$。则渐近几乎必然有 $H_l(N_m(G(n,\ p)))\ =0$。

证明。 证明过程包括为在 $N_m(G(n,\ p))$ 中找到一个 $l+2$ 单形的概率求一个上界。我们有

$$\begin{aligned}\binom{n}{l+2}\left(1-p^{m(l+2)}\right)^{\binom{n-(l+2)}{m}} &\leq n^{l+2}\exp\left(-p^{m(l+2)}\binom{n-(l+2)}{m}\right)\\ &\leq n^{l+2}\exp\left(-n^{m\alpha(l+2)}\binom{n-(l+2)}{m}\right)\\ &\leq n^{l+2}\exp\left(-n^{m\alpha(l+2)}\,\frac{(n-(l+2))^m}{m^m}\right)\\ &= n^{l+2}\exp\left(-n^{m\alpha(l+2)}\,(n-(l+2))^m\frac{1}{m^m}\right)\\ &= n^{l+2}\exp\left(-n^{m\alpha(l+2)}\,\left(n^m-m(l+2)n^{m-1}+\dots\right)\frac{1}{m^m}\right)\\ &= n^{l+2}\exp\left(\left(-n^{m(\alpha(l+2)+1)}+m(l+2)n^{m(\alpha(l+2))}+\dots\right)\frac{1}{m^m}\right)=\mathbf{o}(1)\end{aligned}$$

由上面的引理可知， $N_m(G(n,\ p))$ 渐近几乎必然是 l-连通的，因此其 l 维约化同调消失。

□

命题4。 设 l 和 m 为正整数，且 $p=n^{\alpha}$，其中

(a) *for* l *even* $\alpha\in\left[-2,\dfrac{-4}{l+2}\right]$,

(b) *for* l *odd* $\alpha\in\left[-2,\dfrac{-4(l+2)}{(l+1)(l+3)}\right]$,

˜那么，渐近几乎必然有 $H_l(N_m(G(n,p)))=0$。

证明。 我们采用 Kahle [11] 第384页推论2.5证明中的论证，并应用定理1以得到该结果。□

4. Non-vanishing theorems for reduced homology

现在我们开始证明，在满足关于 p的适当条件时，对于某些随机图 $G(n,\ p)$，多邻居复形的某些同调群在渐近意义下不消失。完全图构成的子图通常称为团。

定义。设 k 和 m 为正整数。$X_{k,m}$ 图是顶点集为

$$\{u_1,\dots,u_k\}\cup\{v_{1,l},\dots,v_{k,l}\}_{l=1,\dots,m}$$

且 $\{u_1,\dots,u_k\}$ 构成一个团， u_i 与 $v_{j,l}$ 相邻，其中 $i\neq j$和 $l=1,\dots,m$。

图 $X_{3,2}$ 如图5所示。

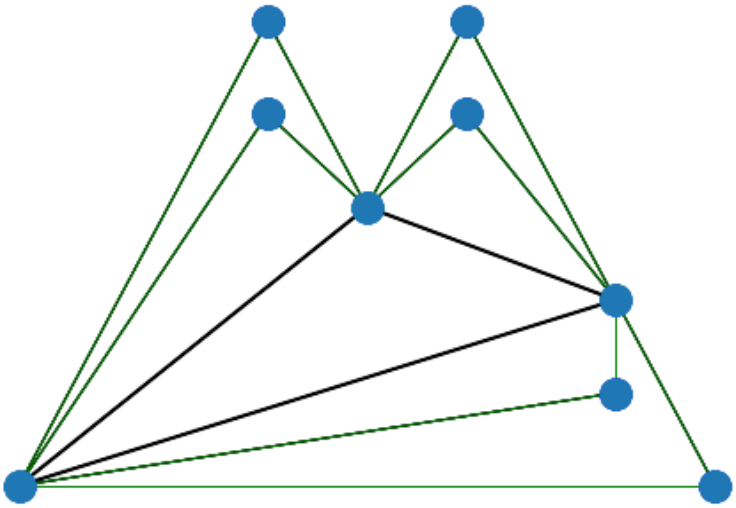


FIGURE 5. The$X_{3,\ 2}$图

命题 5. 设 m 和 k 为整数，且满足 $m \geq 1$ 和 $k \geq 3$。如果图 H 有一个极大团，无法扩展为一个 $X_{k,m}$子图，那么 $||N_m(H)||$可收缩到球面 $\mathbb{S}^{k-2}$

证明。 *</style}*证明在于注意到，中给出的论证在此处同样可以重复，只需将那里所考虑的图 替换为图 ， m。 □

当然，该结果的应用需要采用不同的论证，因为需要确定另一类子图的缺失。

命题6。 设 $p = \frac{1}{2}$ 、 $\varepsilon > 0$ 和 k 为满足以下不等式的整数

$$\left(\frac{2m+2}{2m+1} + \varepsilon\right) \log_2 n < k < (2-\varepsilon) \log_2 n$$

那么渐近几乎必然有

$$\tilde{H}_k(N_m(G(n,p))) \neq 0$$

证明。由 B. Bollobas [2], 可知， $G(n,\ p)$ 包含阶数属于区间 k 内的所有极大团

$$(1+\varepsilon) \log_2 n < k < (2-\varepsilon) \log_2 n$$

还需证明，在我们的假设下，渐近几乎必然不存在 $X_{k,\ m}$子图。

找到一个$X_{k,\ m}$ 子图的概率上界如下：

$$
\begin{aligned}
[(m+1)k]!\binom{n}{(m+1)k}\left(\frac{1}{2}\right)^{\frac{(2m+1)k(k-1)}{2}} & \\
&\leq n^{(m+1)k}\left(\frac{1}{2}\right)^{\frac{(2m+1)k(k-1)}{2}} \\
&= n^{(m+1)k}2^{-\frac{(2m+1)k(k-1)}{2}} \\
&= n^{(m+1)k}n^{\log_n 2[-\frac{(2m+1)k(k-1)}{2}]} \\
&= n^{(m+1)k}n^{[-\frac{(2m+1)k(k-1)}{2\log_2 n}]} \\
&= n^{[\frac{2(m+1)k\log_2 n-(2m+1)k(k-1)}{2\log_2 n}]} \\
&= n^{k[\frac{2(m+1)\log_2 n-(2m+1)(k-1)}{2\log_2 n}]} \\
&\leq n^{k[\frac{2(m+1)\log_2 n-(2m+1)\left(\left[\frac{2m+2}{2m+1}+\varepsilon\right]\log_2 n-1\right)}{2\log_2 n}]} \\
&= n^{k[\frac{-(2m+1)\varepsilon\log_2 n+(2m+1)}{2\log_2 n}]}
\end{aligned}
$$

因此， n 的指数阶为

$$\frac{-(2m+1)k\varepsilon}{2}$$

因此，当 n 趋于无穷大时，该界为 **o(1)**。

□

回顾一下，具有 v 个顶点和 e 条边的图，其密度定义为 $\lambda = \frac{e}{}$。如果一个图的密度大于或等于其每个子图的密度，则称它是平衡的。

引理 3. 图$X_{k,m}$ 是平衡的，其密度为

$$\lambda = \frac{(2m+1)(k-1)}{2(m+1)}$$

证明。 显然，我们有

$$\mathbf{V}(X_{k,m}) = (m+1)k$$

$$\mathbf{E}(X_{k,m}) = \frac{k(k-1)}{2} + mk(k-1) = \frac{(2m+1)k(k-1)}{2}$$

注意到，图 $X_{k,\ m}$ 是通过依次添加阶为 $k-1$ 的顶点，从含有 k 个顶点的平衡完全图得到的。以这种方式得到的图的密度构成如下序列

$$\lambda_n = \frac{\frac{k(k-1)}{2} + n(k-1)}{k+n}$$

通过常规计算可知，这是一个严格递增的序列。只需注意：如果移除团中的一个顶点，则剩余图的密度将小于移除团外顶点之一时的密度，即可完成论证。□

根据 N. Alon 和 J. Spencer [1], 中的定理 4.4.2，我们知道，对于密度为 λ 的平衡图 H， $G(n,\ p)$ 包含 H 作为子图的尖锐阈值为

$$p = n^{-1/\lambda}$$

也就是说，若 $p = n^{\alpha}$ 且 $\alpha > -\frac{1}{\lambda}$，则 $G(n,\ p)$ 渐近几乎必然包含 H 作为子图；若 $\alpha < -\frac{1}{\lambda}$，则 $G(n,\ p)$ 渐近几乎必然不包含 H 作为子图。

命题 7. 若 $p = n^{\alpha}$ 满足

$$\frac{-2}{k+1} < \alpha < \frac{-2(m+1)}{(2m+1)(k+1)}$$

那么渐近几乎必然有 $H_k(N_m(G(n,p))) \neq 0$。

证明。 由于 K_{k+2} 和 $X_{k+2,\ m}$ 均为平衡的，根据命题前面的说明可知， $G(n,\ p)$ 渐近几乎必然包含阶数为 $k+2$ 的团，但不包含 $X_{k,\ m}$ 子图。为了应用上面的命题 5，还需证明将一个给定的阶数为 $k+2$ 的团扩展为阶数为 $k+3$ 的团的概率为 $\mathbf{o(1)}$。事实上，此概率的上界为

$$np^{k+2} = n \cdot n^{\alpha(k+2)} \leq n \cdot n^{\frac{-2(m+1)(k+2)}{(2m+1)(k+1)}} = n^{\frac{-k-2m-3}{(2m+1)(k+1)}} = \mathbf{o}(1)$$

□

5. Applications to topological data analysis

给定一个有限度量空间 $\mathbb{X}$ （例如某个欧几里得空间中的点云），我们可以如下构造一个双滤过复形。设 r 为正实数， m 为正整数。度量空间 $\mathbb{X}$确定了一个图的滤过族 $G(\mathbb{X},\ r)$，其中 $\mathbb{X}$ 中的点是顶点，且当两个顶点之间的距离小于 r时，它们之间有一条边。这就是 Rips 复形的一维骨架。若 $r_1 < r_2$，则存在一个自然映射

$$G(\mathbb{X}, r_1) \to G(\mathbb{X}, r_2)$$

对于每个正整数 m，我们可以考虑多邻居复形$N_m(G(\mathbb{X},\ r))$；对于固定的 r，它关于 m构成一个递减滤过。

现在可以考虑三种类型的持久性：

在保持 m 不变的情况下，关于 r；

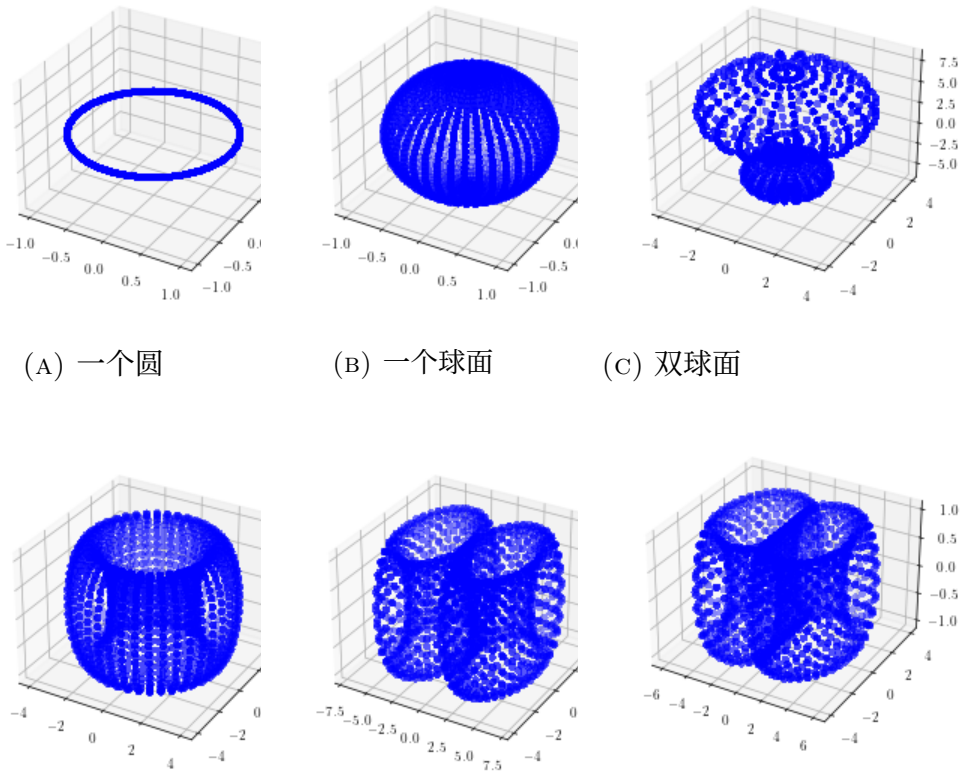
在保持 r 不变的情况下，关于 m ：

•使用 G. Carlsson 和 A. Zomorodian [6] 的多维持久性，同时考察这两个变量。

下面我们探讨上述第一种可能性。下述过程是将 Giotto TDA 教程使用 *Vietoris-Rips* 持久性和持久性熵进行拓扑特征提取, [9]调整后应用于我们的情境。

主要区别在于，为了计算点云的持久同调，我们使用多邻居复形，而非 Vietoris-Rips 复形。对于我们所考虑的噪声极大的示例，使用多邻居复形时，随机森林分类器的预测效果远胜于使用 Vietoris-Rips 复形时的效果。为计算多邻居复形相关的持久同调，我们使用 Gudhi TDA 库的单形流 [10]。

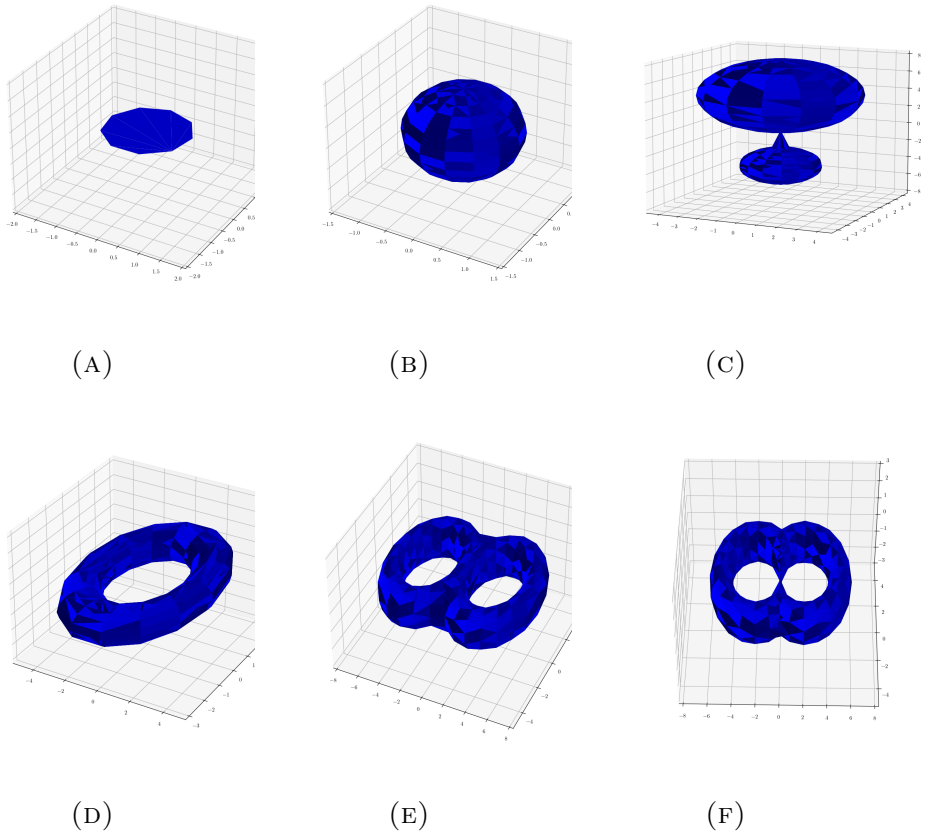
考虑 $\mathbb{R}^3$ 的以下六个子集：一个圆、一个球面、一个不同半径球面的楔和、一个环面，以及两种椒盐卷饼曲面。



(A) 一个圆　(B) 一个球面　(C) 双球面

(D) 环面　(E) 椒盐卷饼曲面（第一种）　(F) A 椒盐卷饼曲面（第二种类型）

FIGURE 6. 无噪声点云



(A) (B) (C)

(D) (E) (F)

FIGURE 7. 无噪声点云的 Alpha复形表示

除圆之外，每一个都是 $\mathbb{R}^2$ 中一个矩形点阵在连续映射下的像。这些 $\mathbb{R}^3$ 子集的图示显示在图 6 中。

现在，我们考虑每种形状的 8 组含噪采样（对每个坐标独立添加幅度为 1 的噪声；每个点云的直径不超过 6）。圆的情形如图7所示。

这样，我们得到 48 个点云。对每个点云，我们按如下方式构造一个过滤图：顶点是点云中的点。边的添加方式与 Rips 复形的构造相同。接下来，对于集合 $\{1, 2, 3, 4, 5\}$ 中的 m 值，我们构造（过滤的）多邻居复形，并计算其持久同调（使用 Gudhi TDA软件包）。带噪声的环面离散采样及其持久图如图8所示。

在每个同调维数中，我们计算持久图的熵。定义如下。假设非对角线上的

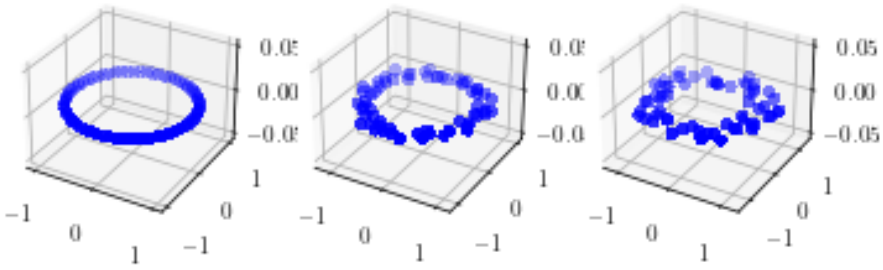


FIGURE 8. 具有不同噪声量的圆的离散采样

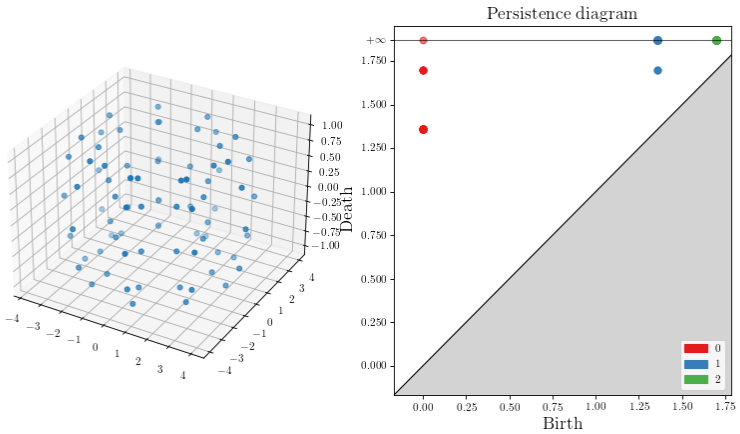


FIGURE 9. 带噪声的环面采样及其持久性图

给定同调维数的持久图中的点构成集合：

$$\{(b_i, d_i) \mid i = 1, \ldots, n\}$$

也就是说，存在 n 个持久同调类，其出生时间为 b_i 、死亡时间为 d_i，我们只考虑满足 $b_i < d_i$ 的那些类。我们为每个同调类关联一个数

$$p_i = \frac{d_i - b_i}{\sum_{i=1}^{n}(d_i - b_i)}$$

这些数均为正数，且总和为 1。我们用概率论和统计学中的常见公式定义熵：

$$E = -\sum_{i=1}^{n} p_i \ln p_i$$

如果以某种方式改变点云，使得单个持久类向对角线移动并与之合并，那么熵将以连续的方式发生变化（由 $\lim_{x\to 0+} x\ln\ x = 0$ 可知）。

同调维数 0、1 和 2 中的熵向量被用作点云的分类器。这些数据被输入随机森林分类器。二十次重复实验的平均分类准确率列于下表中。第一列给出所考察点云中的点数。接下来的五列给出所考察多邻居复形中的邻点数量（从 1 到 5）。最后两列分别显示 Vietoris–Rips复形和 Alpha复形的结果。

点数m=1点	m=2	m=3	m=4	m=5	Vietoris–Rips	Alpha复形
36	0.87	0.91	0.88	0.93	0.89	0.69	0.74
64	0.87	0.89	0.90	0.88	0.93	0.52	0.75

6. Conclusions

我们已经证明，随机图的多邻居复形是邻域复形的一种有用扩展。它具有若干与单邻居情形相似的同调性质，但其中一些性质需要采用截然不同的证明方法（尤其是定理1和命题6中描述的性质）。在第6节中，我们将多邻居复形应用于三维欧几里得空间中含噪点云的分类。我们的结论是，若使用与点云相关联的复形的拓扑熵作为分类器，与邻域复形、Vietoris–Rips复形或Alpha复形相比，多邻居复形具有更强的判别能力。

References

[1] N. Alon 和 J. Spencer，《概率方法》，J. Wiley，第4版，2016年。[2] B. Bollobas，《随机图》，第二版，剑桥大学出版社，2001年。[3] A. Bj¨orner，《拓扑方法》，载于《组合学手册》第2卷，1996年，第1819—1872页。[4] G. Carlsson，《拓扑与数据》，《美国数学会通报》第46卷，2009年，第255-308页。

[5] G. Carlsson 和 A. Zomorodian，《计算持久同调》，《离散与计算几何》33（2），2005年，第247-274页。[6] G. Carlsson 和 A. Zomorodian，《多维持久性理论》，《离散与计算几何》42（2009），第71–93页。[7] H. Edelsbrunner、D. Letscher 和 A. Zomorodian，《拓扑持久性与简化》，《离散与计算几何》28，2002年，第511-533页。[8] P. Erd˝os 和 A. R´eney，《随机图》I，Debrecen数学出版物6（1959），第290–297页。[9] G. Tauzin、U. Lupo、L. Tunstall、J. Burella P´erez、M. Caorsi、A. M. Medina-Mardones、A. Dassatti、K. Hess；《用于机器学习和数据探索的拓扑数据分析工具包》，《机器学习研究杂志》22(39):1-6，2021年；https://www.jmlr.org/papers/v22/20-325.html[10] 用于拓扑数据分析和高维几何理解的Gudhi库：https://gudhi.inria.fr/[11] M. Kahle，《随机图的邻域复形》，组合理论杂志A辑114（2007），380–387页。[12] C. Previte，《图的D-邻域复形》，博士论文，科罗拉多州立大学，2014年。

Faculty of Mathematics and Information Science, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland电子邮箱地址：wojciech.matysiak@gmail.com电子邮箱地址：j.spalinski@mini.pw.edu.pl