arXiv:2301.11965v1 [q-bio.MN] 27 Jan 2023

家谱网络的持久同调

Zachary M. Boyd[a,*]、Nick Callor[b]、Taylor Gledhill[c]、Abigail Jenkins[d]、Robert Snellman[e]、Benjamin Webb[f]、Raelynn Wonnacott[g]

[a]数学系，杨百翰大学，美国犹他州普罗沃 *84602*，*zach boyd@byu.edu*[b]数学系，杨百翰大学，美国犹他州普罗沃 *84602*，*n.b.callor@gmail.com*[c]数学系，杨百翰大学，美国犹他州普罗沃 *84602*，*gledhilltaylor2@gmail.com*[d]数学系，杨百翰大学，美国犹他州普罗沃 *84602*，*jenkins.abby@gmail.com*[e]数学系，杨百翰大学，美国犹他州普罗沃 *84602*，*snellman@mathematics.byu.edu*[f]数学系，杨百翰大学，美国犹他州普罗沃 *84602*，*bwebb@mathematics.byu.edu*[g]数学系，杨百翰大学，美国犹他州普罗沃 *84602*，*raelynnwo@gmail.com*

摘要

谱系网络（即家谱树）日益受到关注，目前已知的最大数据集如今包含超过十亿个个体。人们对家族史的兴趣也支撑着一个规模达85亿美元的产业，该产业预计将在7年内翻一番 [FutureWisereportHC-1137]。然而，人们对谱系网络复杂的网络性质，尤其是在大规模情况下，给予的数学关注仍然很少。

谱系网络的结构尤其值得关注，这是因为人们通常会形成结合，例如婚姻，而这些结合通常发生在直系家庭之外的个体之间。在包括其他社交网络在内的大多数其他网络中，关系形成的距离通常不存在类似限制。为研究这一点对谱系网络的影响，我们使用持久同调来识别并比较101个谱系网络和31个其他社交网络的结构。具体而言，我们引入了网络持久曲线这一概念，用于编码网络的持久区间集合。我们发现，与其他社交网络相比，谱系网络的持久曲线具有独特的结构。这种结构差异还延伸到谱系网络和社交网络的子网络中，这表明即使数据不完整，也可以利用持久同调对谱系网络进行有意义的分析。本文还介绍了谱系网络中的概念（例如共同祖先环）如何通过持久同调得到表示。随着公众对祖源研究的兴趣持续增长，我们预计持久同调工具将在家谱探索中变得越来越重要。

关键词：持久同调、家谱网络、社交网络、持久性曲线、瓶颈距离

1. 引言

对家谱网络的研究，即研究在连续世代中将父母与子女、配偶彼此联系起来的网络，正迅速受到关注。这既是因为系谱学广受欢迎，也因为它在遗传学 [1], 社会学 [2], 人口科学 [3], 和经济学 [4]中的应用。丰富的时间分辨数据日益可得，也推动了人们对系谱学的兴趣。例如，

*通讯作者

FamilySearch 已构建出一棵包含超过 14 亿个体的人类家谱，其依据是 22.1 亿份来源资料，其中包括 47.8 亿张图像 (https://www.familysearch.org/en/newsroom/company-facts)。DNA检测服务的普及，以及音频来源、地理标签、职业元数据和迁徙记录日益丰富，共同使家谱网络成为现存规模最大、信息最丰富且嵌入地理空间的时间网络之一。相关学术研究包括从文献中自动构建网络的方法 [5, 6], 婚姻模式分析 [4], 结构化种群建模、分枝过程 [7], 以及双连通分量 [2, 8]。我们尤其关注研究近期共同祖先距离的成果，包括理论研究和模拟研究（例如 [9, 3]）。越来越多的文献也将家谱网络用于遗传推断，例如 [1]。

与这些谱系研究相关，网络科学的一个主要目标是描述此类现实网络的结构。本文将持久同调作为分析和探索家谱网络结构的工具。粗略地说，持久同调是一种表示网络结构中空洞或间隙的方法，并能区分这些空洞对于整体网络结构的重要程度。持久同调可用于比较两个网络中的这些空洞，而无需要求各个顶点或边之间存在对应关系，甚至无需两个网络规模相同。其基本思想是用单形（点、边、三角形、四面体等）“填充”网络，并追踪网络在这一过程中的变化（详情请见第3节）。

持久同调在网络研究中的一些类似应用包括 [26], [28], [30], [27]。 [26] 研究的合作网络与本文用于比较的社交网络相似，不过我们的重点主要是将它们与家谱网络区分开来。 [28]和 [30] 都将持久同调技术应用于各种形式的一般随机网络。生成网络拓扑对象的方法也可以有所不同，例如 [27] 比较了三种方法。我们还推荐 [29] 和 [36] ，它们对应用持久同调的一般方法作了很好的概述。

在本文中，我们为每个网络构造拓扑表示的方法，遵循了上述文献中相同的一般模式。不过，我们也承认，编码此类信息的方法有很多种。 [32] 和 [33] 将信息编码为点云，而非图。 [34], 提出了更高维版本的持久同调，这或许可以纳入时变网络。最后， [35]中的表述或许能更好地分析受损或规模过大的数据集。

我们还希望介绍四个具体应用，以展示持久同调的多功能性。在这些应用中，持久同调用于识别数据中的结构空洞，并将这些空洞与底层网络中可识别的特征联系起来。我们希望强调的是后者。Robins 等人 [10] 表明，利用持久同调发现的空洞对应于多孔材料中的渗流球体。在 [11], 中，当若干神经元群组依次紧密连接，而非序列中的神经元对仅弱连接时，就会出现结构空洞。在这些神经网络中，持久同调提供了一种识别和分类不同序列的方法，同时还能量化这些连接的强度。 [31]中的应用通过构造适当的拓扑对象，提供了一种将传统遗传分析工具扩展至参数化数据集族的方法。最后， [12] 表明，结构空洞或间隙也可以代表更加抽象的概念。在这个例子中，研究显示持久空洞对应于音乐作品中的无调性。

直观地因此，与其他网络相比，家谱网络中的空洞或间隙应该大不相同，th

例如社交网络，因为家谱网络中的结合[1]（例如婚姻）通常发生在特定距离，而非通过三元闭包等其他机制形成。也就是说，形成结合的个体之间的距离通常不会太小或太大（见第2节）。相比之下，在其他社交网络中，新连接可以在任意距离形成，但往往距离较小 [13]。家谱网络与其他社交网络在网络增长方面的这种差异，会导致网络拓扑出现差异，并反映在网络的持久同调中。因此，持久同调是一种有用的描述性工具，可用于探索和建模家谱网络的结构。

在此，我们提出一种表示持久同调的新方法，称为持久曲线（见第 4 节）。许多家谱网络的持久性曲线彼此非常相似；更重要的是，家谱网络子集的持久性曲线，也就是抽样 家谱网络的持久性曲线，与未采样家谱网络的持久性曲线也相似（见第6节）。

为使我们对家谱网络的研究置于适当背景下，我们还研究了社交网络的持久同调。我们发现，对于所考察的社交网络，同样的结论也成立：社交网络的持久性曲线呈现出共同的模式，而社交网络与抽样社交网络的持久性曲线也彼此相似（见第6节）。我们还使用另一种比较持久同调的工具——瓶颈距离——验证了我们的分析。该工具也能够检测并区分家谱网络与其他社交网络之间不同的同调模式。

总之，我们作出以下贡献：

- 介绍持久曲线这一概念，并提出将其与瓶颈距离结合，用作分析一般网络的工具。
- 使用持久曲线和瓶颈距离，报告谱系网络独特的持久同调结构。
- 将这一结构与谱系学相关概念联系起来。
- 同样，报告社交网络独特的持久同调结构，并将其与谱系网络的结构进行比较。
- 提供证据，说明即使存在不完整数据，持久同调方法依然有效。这一点尤其重要，因为谱系数据往往（甚至可以说必然）并不完整。

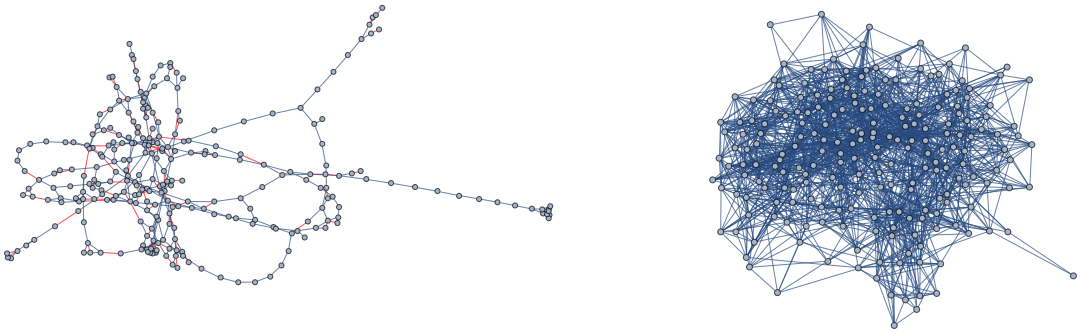
全文将家庭网络中的示例与其他社交网络进行对比，以从持久同调的角度突出谱系网络的独特特征。

本文的组织结构如下。第2节介绍谱系网络与社会网络。第3节定义网络的持久同调，并引入持久曲线的概念。第 4 节定义瓶颈距离，并说明如何利用这一距离和持久曲线比较网络。第5节介绍本研究所用的家谱和社会数据集，并在第6节给出实验结果。第6节还讨论如何利用持久同调表示社会网络和家谱网络的某些结构特征。第7节总结研究结果，并讨论将持久同调用作分析一般网络结构和恢复网络特征的工具。全文还会为每个引入的概念提供示例。

[1]为了涵盖本文讨论的各种相关关系，我们使用“结合”一词，不仅指合法婚姻和普通法婚姻，也包括其他一些关系，例如任何育有子女的关系。

2. 背景：谱系网络与社会网络

我们用图 $G = (V, E)$ 表示家谱网络，其中 $V = \{1, 2, \ldots, n\}$ 表示网络中的个体，而 E 表示（谱系）关系。这些关系包括亲子 边和配偶（或更一般地说，结合）边。为简便起见，这些边均视为无向边。我们注意到，家谱网络的结构通常被认为是



Tikopia 家谱网络　　宿舍社会网络

图 1：左：展示了 Tikopia 谱系网络中最大的连通分量，该分量由波利尼西亚 Tikopia 岛上1930年至2008年间的288个个体组成 [14]。亲子边以蓝色显示，结合边以红色显示。右：展示了澳大利亚国立大学校园宿舍社会网络中最大的连通分量，该分量由217个个体及其友谊关系组成 [15]。

呈现“树状”，因为谱系网络通常由一个个体及其父母、祖父母等构成，并忽略结合边。如果我们只纳入家庭的少数几个世代，所得结果就是一棵树，即一个连通无环图。然而，由于存在例如由双亲和子女构成的三角形（包含两条亲子边和一条结合边），完整的谱系网络并非树。鉴于这类环十分常见，而且是可能出现的最小环，我们将其称为平凡圈。另一种典型的家庭环，即由双亲和若干子女组成的家庭内部的环，是由双亲和两个子女构成的长度为四的环。

虽然亲缘环在家谱网络中无处不在，但它们并不是唯一可能形成的环。沿着某个人的祖先脉络追溯足够远，通常可以找到一个最近共同祖先，即其父亲和母亲的共同祖先。如果存在这样的祖先（通常确实存在），那么该家谱网络就包含一个非平凡环。我们将其称为共同祖先环，它只由亲子边构成。家谱网络中还可能通过结合形成其他非平凡环。例如，当一个家庭中的两兄弟姐妹分别与另一个家庭中的两兄弟姐妹结合时，就会形成“双重表亲”关系。由此产生联合环，即只包含结合边以及连接兄弟姐妹的亲子边的环。在家谱网络中，结合边和亲子边可以通过各种方式组合，形成复杂的非树结构（见左侧图 1）。

谱系网络的一个特有之处在于，结合边通常会在网络中的特定距离处形成。这里，若存在路径，则$d(i, j)$表示i 与j 这两个个体之间的最短路径距离；若不存在这样的路径，则距离为无穷大。在谱系网络中，我们将两个个体结合之前彼此间的距离称为这对伴侣的到结合的距离。由于文化、遗传及其他原因，这些距离通常不小，也就是说，往往大于四。因此，

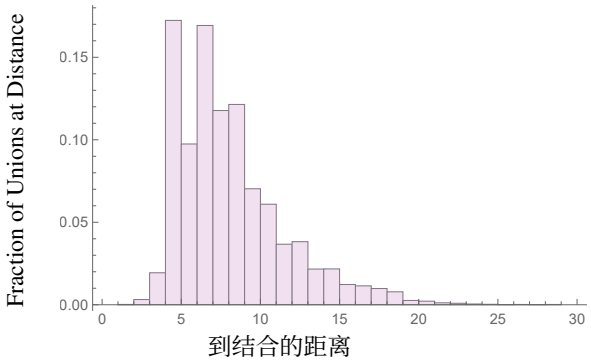


图2：此处展示了有限“到结合的距离”的直方图，数据取自 kinsources.net 上的 104 个家谱网络。每个柱子的高度表示在特定距离上形成的结合所占的比例。

家谱网络通常没有较短的非亲缘环，而且往往存在较长的扩展环。图2对此进行了说明，其中收集了附录表 2所列的104个公开可用家谱网络的到结合的距离数据。此处省略了家庭环，每个柱形的高度表示在特定距离处形成的结合所占的比例。值得注意的是，在距离小于五时形成的结合很少，而绝大多数距离介于5至10之间。

图3展示了家谱网络具有较长扩展环这一现象。左侧橙色部分显示的是圣马力诺谱系网络的环长度分布；该网络涵盖圣马力诺共和国从15世纪到19世纪末的人口 [14]。这个网络由28,586个个体组成，其中有7,146个长度为三的亲缘环和8,636个长度为四的亲缘环。为便于观察构成网络非亲缘环基的环的长度，图中省略了这些环。作为对比，蓝色部分显示了 Deezer欧洲社交网络中一组环基所含环的长度分布，该网络由28,281个个体组成。与家谱网络类似，社交网络也表示为一个图 $G = (V,E)$，其中顶点 V也代表个体。不同之处在于，社交网络中的边代表某种社交互动。Deezer网络是一个在线音乐流媒体平台，其社交网络代表欧洲使用该平台的个体，边则表示相互的用户关注关系。

值得注意的是，圣马力诺网络中长度小于 10 的非家族基环相对较少，而长度大于 30 的环却相当多。相比之下，Deezer 社交网络的基环长度分布要集中得多，大致介于 5 到 15 之间。

为了解这些环分布在多大程度上与相应网络的局部结构有关，我们分别将它们与这两个网络各自对应的配置模型的环分布进行比较。配置模型 是一种用于生成具有给定度序列 [16] 的随机网络的模型。我们取圣马力诺谱系网络和 Deezer 社交网络的度序列，分别创建十个版本，并确保每个版本都具有相同的度序列。将这些版本的圣马力诺和 Deezer 网络的基循环长度分布取平均后，结果如图3所示（中间和右侧分别以红色和绿色表示）。圣马力诺网络的环分布与配置模型生成的结果相当不同，而 Deezer 社交网络的环分布则与其配置模型预测的分布非常相似。这表明，Deezer 社交网络中的许多环结构主要由局部互动决定，而圣马力诺谱系网络中的环则受到形成该网络的非局部机制影响。这些机制包括，

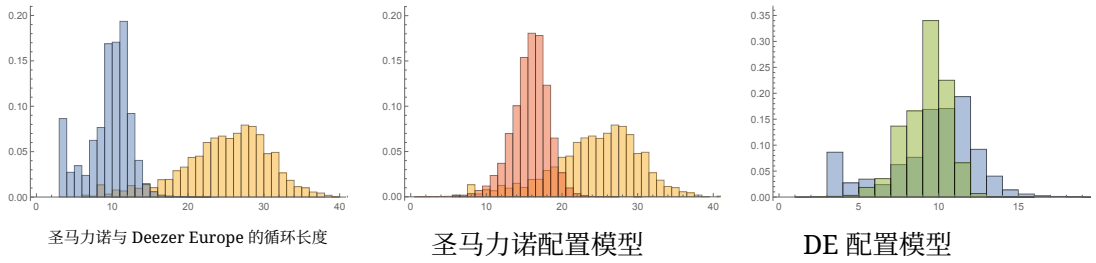


图3：左：橙色表示圣马力诺（SM）谱系网络中构成非家族循环长度基的环的长度分布。蓝色表示 Deezer Europe（DE）社交网络中所有环的循环长度分布。中：橙色再次表示圣马力诺谱系网络的基循环长度分布。红色表示在圣马力诺网络上进行十次（含自环和多重边）配置模型实现后，基循环长度的平均分布。由于配置模型生成的图与 SM 网络具有相同的度分布，该图表明，SM 中较长的环并非仅由度分布造成。右：蓝色再次表示 Deezer 社交网络的基循环长度分布。绿色表示在 Deezer 社交网络上进行十次配置模型实现后，基循环长度的平均分布。对于这个社交网络，循环长度分布大部分可以仅由度分布解释。

据推测，还有上文所述的非局部到结合的距离现象。

图3中圣马力诺家谱网络与 Deezer 社交网络的循环长度分布之间所呈现的关系，是我们在第5节中考察的谱系网络与社交网络的典型特征。这表明，循环长度分布是一种可用于区分家谱网络和社交网络的特征。具体而言，当我们考察两个环数量相近的网络时，家谱网络的循环长度分布比社交网络宽得多。然而，图3中用于计算循环长度分布的方法并未进一步揭示这一现象。这一局限促使我们运用持久同调的工具，它提供了描述和测量任意两个网络环之间关系的方法。这些方法所能获得的附加结构，有助于我们进一步区分家谱网络与社交网络的结构（见第 6.1 节），并将图3所展示的结构差异与分别生成家谱网络和社交网络的机制联系起来（见第6.3节）。

3. 网络的持久同调

持久同调提供了一种研究网络中环的方法。本文将从单纯同调的角度简要介绍持久同调。有关单纯同调的深入讲解，请参阅 [17] 的第2.1章。对于熟悉持久同调基础知识，或希望跳过以下技术讨论的读者，可以直接阅读第5节，我们将在该节讨论所分析的社会网络和家谱网络。

对于由图 $G = (V, E)$ 给出的网络，我们定义距离矩阵 $D(G) = [d_{ij}]$，其元素为 $d_{ij} = d(i, j)$，表示个体 i 与 j 之间最短路径的长度。对于距离矩阵 $D(G)$ 中出现的每个值 δ，我们按如下方式构造一个单纯复形 G_δ。0-单形的集合等同于 G 的顶点集合，其中每个0-单形都对应一个顶点。由于0-单形与顶点之间的区别纯属形式上的，我们将交替使用“0-单形”和“顶点”这两个术语，并且0-单形的编号与顶点相同。1-单形 E_δ 的集合对应于满足条件 $d(i, j) \leq \delta$ 的边集合 $\{i, j\}$，其中边 $\{i, j\}$ 对应于由 i 和 j 构成的1-单形

。再次强调，在此处区分二者对当前讨论并无必要，因此我们对1-单形和边使用相同的记号。不过，单纯复形G_δ也可能包含图G中没有对应代表元的对象，即n-单形，其中$n \geq 2$。对于每个整数$n \geq 2$，G_δ中的n-单形集合由所有满足条件$d(a_i, a_j) \leq \delta$的n-单形 $[a_0 a_1 \ldots\ldots a_n]$组成，其中$0 \leq i < j \leq n$。也就是说，若列出的每个顶点与$\sigma$中列出的每个顶点之间的距离都不超过$\delta$，则$G_\delta$包含一个$n$-单形 σ。

为简化后续定义，我们将G_δ的定义扩展到所有非负整数。对于$i \geq 0$，令δ_i为$D(G)$中满足$\delta_i \leq i$的最大项。令$G_i = G^{\delta_i}$。该定义与G_δ的构造确保了以下三个重要性质对所有G_i均成立。

1. 对于$i < j$，G_i是G_j的子复形，即G_i的每个单形都是G_j的单形。2. 对于$i \geq 1$，存在G_i的一个子复形，可将其与原图G对应起来。3. 由于G是有限的，令$M = \max_{ij} d(i, j)$，则对于所有$i \geq M$，$G_i = G_M$。

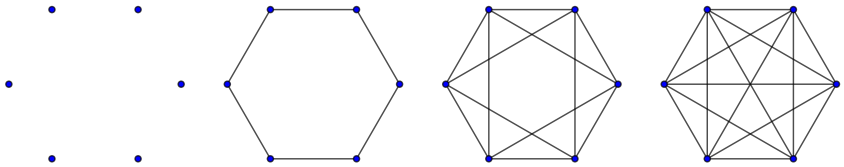


图4：例3.1中的六边形网络 $G = G_1$随着i从0增加到3而逐渐填充。由此得到的单纯复形 G_0, G_1, G_2, G_3从左到右依次展示。

例3.1。（六边形网络）考虑图*4(b)*所示的六边形网络 $G = (V,E)$，它有六个顶点，构成一个单独的环。该网络的距离矩阵为

$$D(G) = \begin{bmatrix} 0 & 1 & 2 & 3 & 2 & 1 \\ 1 & 0 & 1 & 2 & 3 & 2 \\ 2 & 1 & 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 & 1 & 2 \\ 2 & 3 & 2 & 1 & 0 & 1 \\ 1 & 2 & 3 & 2 & 1 & 0 \end{bmatrix}.$$

对于数值 $i = 0$、1、2、3，我们构造了四个单纯复形 G_0, G_1, G_2, 和 G_3，其中令 $G_i = (V_i, E_i)$。当 $i = 0$时，E_0 为空。因此，G_0 由六个顶点组成。当 $i = 1$ 时，集合 E_1 包含构成网络唯一环的六条边，因此 $G_1 = G$。该图没有平凡循环（即三角形），所以 G_1 不包含维度大于 *1* 的单形（即不存在 $n > 1$的 n-单形）。当 $i = 2$ 时，集合 E_2增加了另外六条边。此时我们还有八个平凡循环。每个环都是一个 *2*-单形的边界，因此 G_2 也包含这八个 *2*-单形。然而，这些 *2*-单形的任何子集都不能构成 *3*-单形的边界，所以 G_2 没有维度大于 *2* 的单形。当 $i = 3$ 时，集合 E_3包含 G 的顶点之间所有可能的边，因此所有可能的平凡循环都存在。此外，所有可能的 *2*-单形以及由此推得的所有可能的 n-单形也都存在于 G_3中。特别地，G_3 是一个带边界的 *6*-单形。由于 $M = 3$ 是我们在距离矩阵中看到的最大值，因此对于 $i \in \mathbb{Z}$、$i > 3$，$G_i = G_3$。

网络G的持久同调衡量了G_i的同调随i增大而发生的变化。如果在i的多个取值下都能识别出某些特征，我们就称这些特征持续存在。直观而言，源于实际网络结构的特征应当在许多i取值下持续存在，而由测量误差（即“噪声”）引起的特征则只会零星出现。稳定性定理（[18]的主定理）指出，如果网络测量误差有某个常数C作为上界，那么真实网络与含噪网络的持久同调之间的差异至多为C。我们将在第4.1节中更精确地阐述这一结论。

这里我们将用单纯同调给出持久同调的形式化定义，随后立即在网络的语境下给出等价定义。我们用$H_p(G_i)$表示单纯复形G_i的维度p单纯同调，其系数取自$\mathbb{Z}_2$，因为$H_p(X)$是一个$\mathbb{Z}_2$向量空间。

定义1。（第p阶持久同调）给定图G以及整数i、j，其中$0 \le i \le j$，令函数$\phi_{i,j}$：$H_p(G_i) \to H_p(G_j)$为包含映射$G_i \to G_j$所诱导的线性映射。G的第p阶持久同调$PH_p(G)$是有序对$(\{H_p(G_i)\}_{i\ge 0}, \{\phi_{i,j}\}_{0\le i<j})$。

我们在第4至6节中的分析只需要持久同调的前几个维度，就足以区分我们研究的谱系网络与社交网络。为了更好地理解持久同调所计算的内容，接下来我们将使用网络概念，为PH_0, PH_1, 和PH_2给出等价定义。我们还将说明这些定义如何应用于图4(b)中的六边形网络。（有关PH_0, PH_1, 和PH_2，请参见例3.3、3.4和3.5。）

定义 *2.*（诞生与消亡）设 $G = (V, E)$ 是一个网络，其单纯复形为 $G_0, G_1, G2, \cdots$。G 的 p 次持久同调给出了 $\phi_{i,j}$，它们连接 G_i 的 p 次同调与 G_j 的 p 次同调。假设已为每个 $H_p(G_i)$ 选定一组基，使得若 α 是 $H_p(G_i)$ 的一个基元素，则 $\phi_{i,j}(\alpha)$ 在 $H_p(G_j)$ 中要么是平凡元，要么是一个基元素。基元素 $\alpha \in H_p(G_j)$ 的诞生，是满足以下条件的最小指标 i：对于某个基元素 $\hat{\alpha} \in G_i$，有 $\alpha = \phi_{i,j}(\hat{\alpha})$。$\alpha$ 的死亡，是满足 $\phi_{j,k}(\alpha)$ 为平凡元的最小指标 k。

注 3.2。已经熟悉持久同调的读者会发现，前面的定义有些不寻常，尽管它与标准定义等价。我们采用这种处理方式，是为了减轻非专业读者的符号负担。对于其他一些持久同调定义，我们也采取了类似的做法。

我们将在以下定义中说明如何为 H_0, H_1, 和 H_2 选取这样的代表元。不过，有了这样的代表元，映射 $\phi_{i,j}$ 和 $\phi_{j,k}$ 就只是由包含映射 $G_i \subset G_j \subset G_k$ 诱导的同调映射。也就是说，如果 a 表示 $\alpha \in H_p(G_i)$，那么 a 也表示 $\phi_{i,j}(\alpha)$。持久同调基本定理保证，我们可以选取同一个代表元，使其对应于 $\alpha \in H_p(G_j)$、$\hat{\alpha} \in H_p(G_i)$，以及 $\phi_{j,k}(\alpha) \in H_p(G_k)$。于是，$\alpha$ 的诞生就是该代表元首次存在于其中的 G_i，而 α 的死亡则是该代表元首次在其中零伦、即同伦于一个平凡循环的 G_k。

定义 3。（表示持久同调：维度 0）设 $G = (V, E)$ 是一个网络，其顶点 $V = \{1, 2, \ldots, n\}$ 构成 k 个连通分量。则 $H_0(G_0) \cong \mathbb{Z}_2^n$，因此我们可以将 $H_0(G_0)$ 的基与全部 n 个顶点对应起来。同样地，我们可以选择 k 个顶点，每个连通分量中选取一个，用来表示 $H_0(G_i) \cong \mathbb{Z}_2^k$ 的基，其中 $i \ge 1$。因此，我们将 G 的顶点称为 $PH_0(G)$ 的代表元。（事实上，$PH_0(G)$ 是一个向量空间，其基元素是 *0*-单形形式和的等价类。）

例3.3。 现在考虑图*4*中六边形网络 G 的 PH_0（G），以及同一图中的 $G_0, G_1, G_2,$ 和 G_3。回顾一下，G 有六个不同的顶点，构成一个连通分量。如果我们对顶点进行任意编号，$V = \{1, 2, 3, 4, 5, 6\}$，则 H_0（G_0）$\cong \mathbb{Z}_2^6$，这等价于以 V 为基的 $\mathbb{Z}_2$ 上的向量空间。对于 $i > 0$，H_0（G_i）$\cong \mathbb{Z}_2$，这等价于以{1}为基的 $\mathbb{Z}_2$ 上的向量空间。对于任意 $v \in V$，由于 $i = 0$ 是我们首次看到 v 的时刻，因此我们称其为 v 的诞生 。在 $i = 1$时，由于我们已经从基中移除了除1 以外的所有顶点，因此称这是那五个 *0*-单形的死亡 。由于1 始终位于 G_i的基中，因此1 的死亡 被称为 ∞。

定义4。 ***(***表示持久同调：维度 ***1)*** 设 $G = (V, E)$ 是一个只有一个连通分量的网络。对于每个 $i \geq 0$, 我们可以将 $H_1(G_i)$ 的基与 G_i中的一组环 C_i 对应起来。持久同调基本定理允许我们选择这些环，使得对于 C_i中的任意环 σ，对于任何整数 $j \geq 0$，以下情况恰有一种成立：

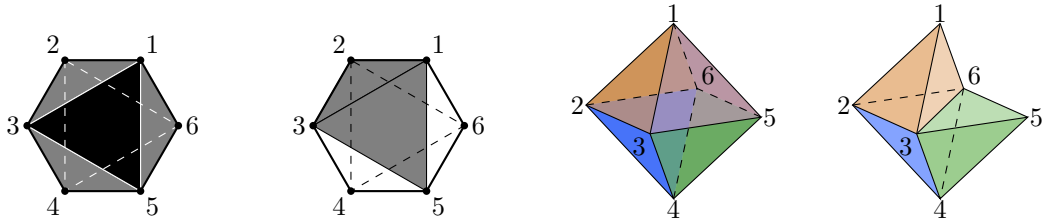
1. σ 在 G_j中不存在，此时 $j < i$；*2.* σ 在 G_j中是平凡的或零伦的，此时 $i < j$；*3.* σ 是 C_j中的一个环。

因此，我们将 $\bigcup_{i\geq 0} C_i$ 中的环称为 $PH_1(G)$的代表元。（再次说明，$PH_1(G)$ 实际上远不止这些。这些实际上是等价类的代表元，它们构成了 $PH_1(G)$作为向量空间的一组基。）

我们注意到C_0 始终为空，因为G_0中没有边。此外，对于所有$i \geq 0$，$\mathrm{rank}(H_1(G_i)) = |C_i|$ 。由于G_i 的构造，$PH_1(G)$的所有代表元都会出现在G_1中。可以将$PH_1(G)$的代表元视为代表“大型”环。更具体地说，如果一个环 σ 包含在$\bigcap_{s\leq i\leq t} C_i$中，那么它的直径至少为$t$，且至少有一对相邻顶点之间的距离为$s$。

例3.4. 现在我们来考察图*4(b)*中的六边形网络G的$PH_1(G)$ 。在图*4(a)*和图*4(b)*中，我们看到G_0 没有环，G_1 恰有一个环，并且G_1 中的环是非平凡的。在图*5(a)*和图*5(b)*中，我们标出了G_2，中的一些环，即图*5(a)*中的环*1,2,3,1*；*3,4,5,3*；*1,5,6,1*；和*1,3,5,1*，以及图*5(b)*中的环*1,2,3,5,1*。事实上，图*5(c)*表明G_2 是一个八面体，因此G_2 中的每个环要么是平凡的，要么是零伦的。最后，G_3 包含的环甚至比G_2，更多，例如*1,3,6,1*；但这些环全都是零伦的，因为G_3 还包含六个顶点所有可能的2-单形。因此，$PH_1(G)$ 只有一个代表元，即环*1,2,3,4,5,6,1*；它出现在G_1，中，所以我们说$t = 1$ 是该环的诞生 。该环在G_2，中是零伦的，因此$t = 2$ 是该环的死亡 。

现在我们转而关注 $PH_2(G)$，但为了表示 $PH_2(G)$，我们需要为诱导子图引入一些新的结构。在 G_i 中，三角形 [a b c] 是由三个顶点 a、b 和 c 组成的集合，它们在 G_i中构成一个平凡循环。也就是说，边 {a, b}、{b, c}和 {a, c} 都存在于 G_i中。G_i 中的闭曲面是由若干个互不相同的三角形组成的集合，其中对于集合中的每个 [a b c]，恰好有另一个三角形 [a b d] 也属于该集合。若 G_i 中的闭曲面所对应的 2-单形集合在 G_i中是零伦的，则该闭曲面是平凡的。也就是说，这个闭曲面被 G_i中的某些 3-单形集合“填充”。例如，图5(c)中的八面体是 G_2 中的非平凡闭曲面，因为 G_2中没有 3-单形。不过，在 G_3, 中，我们在距离为 3 的顶点之间添加了边。由此，我们得到若干个 3-单形，其中包括 [1 2 3 6]，[1 3 5 6], [3 4 5 6], 和 [2 3 4 6]。图5(d)展示了其中三个 3-单形，以说明 G_2 中的闭曲面如何由全部四个 3-单形填充。



(a)G_2 平凡圈　(b)G_2 零伦圈　(c)G_2 球面　(d)G_3 选择 3-单形

图 5 $_2$ 2 [1 2 3], : G 中单形和环的 [3 4 5], [1 5 6], [1 3 5] $_{1,2,3,5,1}$ $_2$ 示意图。左：四个平凡圈分别由单个 -单形填充：和。左中：一个非平凡但零伦的环，由两个 -单形 [1 2 3] 和 [1 3 5]填充。右中：全部八个 2-单形表示为正八面体的面。右：G_2 的闭曲面由四个 3-单形 [1 2 3 6], [1 3 5 6]（未显示）[3 4 5 6], [2 3 4 6]填充。

定义5. (表示持久同调：维度 **2)** 设 $G = (V,E)$ 是一个只有一个连通分量的网络。对于每个 $i \geq 0$，我们可以将 $H_2(G_i)$ 的基与 G_i中非平凡闭曲面的集合 S_i 对应起来。持久同调基本定理允许我们选择这些代表元，使得如果 σ 是 S_i中的一个闭曲面，那么对于任意整数 $j \geq 0$，以下情况恰有一种成立

1. σ 在 G_j中不存在，此时 $j < i$；*2.* σ 在 G_j中是平凡的，此时 $i < j$；*3.* σ是 S_j中的一个环。

因此，我们将 $\bigcup_{i\geq 0} S_i$ 中的闭曲面称为 $PH_2(G)$ 的代表元。

$PH_2(G)$ 的几何直觉与 $PH_1(G)$ 类似，都是为了识别 G中的大型“空洞”。如果$s \leq i \leq t$ S_i，那么 σ 是一个直径至少为 t 的闭曲面。s 的值较难描述，但与顶点密度有关。

例3.5。现在我们来考察例*3.1*中的六边形图 G 的 $PH_2(G)$ 。回忆例*3.4*可知，G_0 和 G_1 没有平凡环，因此不包含闭曲面。从图*5*中可以看出，G_2 恰好有一个闭曲面，而且它必定是非平凡的，因为其中没有 *3*-单形。最后，G_3 有许多闭曲面，但由于它包含六个顶点上所有可能的3-单形，这些闭曲面全都是平凡的。因此，$PH_2(G)$ 只有一个代表元，即 G_2中的八面体闭曲面。该曲面最初出现在 G_2, 中，因此 $t = 2$ 是它的诞生；该曲面在 G_3, 中被实体填充，因此 $t = 3$ 是它的死亡。

定义 6. (持久区间**)** 回顾一下，网络 G 的持久同调中代表元 $\sigma \in PH_p(G)$ （顶点、环或闭曲面）的诞生，是使 $\sigma \in G_i$ 成立的最小整数 i；而 σ 的死亡则是使 $\sigma \in G_{j-1}$ 成立且 σ 在 G_k 中对 $k \geq j$ 都是平凡的最大整数 j（如果这样的整数存在）。σ 的持久区间为 $[a,b)$，其中 a 和 b 分别是 σ 的诞生和死亡。这表示所有参数值 i 的集合：对于这些 i，对应于 σ 的等价类是 $H_p(G_i)$ 中的非平凡元素。σ 的持久性为 $b - a$。

例3.6. 现在我们完成对图*4(b)*中 G 的持久同调的讨论。回想例*3.3*，$PH_0(G)$ 有六个代表元。它们的诞生时间均为 $t = 0$。其中五个的死亡时间为 $t = 1$，另一个的死亡时间为 ∞。因此，$PH_0(G)$ 的持久区间为 $[0,1) \times 5$ 和$[0,\infty) \times 1$。

由例*3.4*可知，$PH_1(G)$有一个代表元，其诞生时间为 $t = 1$，死亡时间为 $t = 2$。因此，相应的持久区间为 $[1, 2)$。注意，环的直径为3，且每对相邻顶点之间的距离为1。这符合前文提到的观点：$PH_1(G)$ 的代表元表示“较大”的环。具体而言，σ 的直径至少为 σ的死亡时间，而 σ的诞生时间是相邻顶点之间的最大距离。

根据例*3.5*，$PH_2(G)$ 有一个代表元，其诞生时间为 $t = 2$，死亡时间为 $t = 3$。因此，该元素的持久区间为 $[2, 3)$。注意，对应顶点集的直径在G中为3。这也符合前文提到的观点：$PH_2(G)$ 能够识别G中的大型“空洞”。具体而言，σ 的死亡时间是 σ直径的下界。

根据定义 3、4、5 和 6 中选取的代表元，我们对有限无向无权图G的持久同调有以下三点观察：**(i)** 如果G 有n个顶点，那么$PH_0(G)$恰有n 个持久区间，其中恰有一个 $[0, \infty)$区间对应于每个连接的值$[0, 1)$区间。

(ii) 在维度 1 中，$PH_1(G)$描述原始网络中非平凡环的数量和大小。所有持久区间的形式均为 $[1,b)$，其中$b > 1$是某个整数。b 的值与相应环的直径有关。在我们研究过的网络中，我们注意到，$PH_1(G)$中的持久区间$[1,b)$对应一个简单环，其顶点数介于$3b - 2$ 和$3b$ 之间（含端点）。

（iii）在维度 2 中，我们在$PH_2(G)$中检测到的空洞揭示了环的非平凡交集。这类交集难以可视化，但粗略地说，只有当多个较大的环两两相交时，$PH_2(G)$中的代表元才可能形成。

4. 使用持久同调比较网络

本节将展示如何使用基于持久同调的方法来比较不同的网络。本文介绍的两种方法分别基于：(a) 瓶颈距离；(b) 给定网络集合的持久性曲线。(a) 和 (b) 都需要先计算持久区间，再分析这些区间之间的差异。

本节中，我们用两个网络来展示这些方法：图 1（左）中的 Tikopia 家谱网络，以及图4中的六边形网络。这两个网络的持久区间分别列于表 1。

维度	区间类型与持续性	
	Tikopia	六边形
0维	$[0, \infty) \times 8$, $[0, 1) \times 286$	$[0, \infty) \times 1$, $[0, 1) \times 1$
维度 1	$[1, 2) \times 16$, $[1, 3) \times 19$, $[1, 4) \times 5$, $[1, 5) \times 3$, $[1, 6) \times 2$, $[1, 7) \times 1$	$[1, 2) \times 1$
维度 2	$[2, 3) \times 4$, $[3, 4) \times 11$, $[4, 5) \times 12$, $[5, 6) \times 4$, $[6, 7) \times 5$, $[7, 8) \times 1$, $[8, 9) \times 1$	$[2, 3) \times 1$

表 1：展示了 Tikopia 家谱网络和六边形网络的持久区间。此处记号 $[a, b) \times k$ 表示该网络具有 k 个持久区间 $[a, b)$。相应的持久图见图 6，Tikopia 网络的相应持久曲线见图 7。

4.1. 持久图与瓶颈距离

表示持久区间的一种常见方法，是将它们绘制为 $\mathbb{R} \times (\mathbb{R} \cup \{\infty\})$ 中的点，这通常称为持久图。虽然这种可视化网络持久同调的方法无法显示某个持久区间出现的频率，但它能提供有关特定网络中会出现哪些类型的持久区间的信息。

定义7.（持久图）设 $PH_p(G)$ 是网络 G 的第 p 阶持久同调。$PH_p(G)$ 的持久图是 $\mathbb{R} \times (\mathbb{R} \cup \{\infty\})$ 中的一个点的多重集，定义如下。

- 对于每个具有持久区间 $[a,b)$ 的 $\sigma \in PH_p(G)$，我们将点 (a,b) 加入一次。
- 对于每个 $c \in \mathbb{R}$，我们将点 (c,c) 加入无穷多个副本。

请注意，我们加入点 (a,a)，以表示图 G 中在 $PH_p(G)$ 中被视为平凡的特征，例如恰由三个顶点组成的环。加入这些点是为了在持久图空间上定义有意义的度量。这里使用的度量称为瓶颈距离。

定义8.（瓶颈距离）设 S_1 和 S_2 分别是图 G 和 H 的持久图。令 η 遍历从 S_1 到 S_2 的所有双射。那么 S_1 与 S_2 之间的瓶颈距离 为

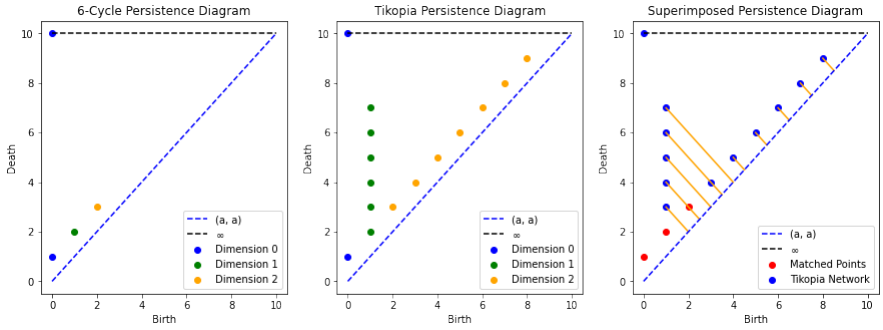
$$d_B(S_1, S_2) = \inf_{\eta} \sup_{x \in S_1} \|x - \eta(x)\|_\infty.$$

持久同调基本定理（在 [19], 中提出，并在 [36]和 [29]中得到了很好的阐释）保证，若两个图同构，则相应的持久图相等，因此瓶颈距离为 0。然而，非同构的图也有可能具有相同的持久图。

例4.1。（***Tikopia*** 网络与六边形网络之间的瓶颈距离）注意，*Tikopia* 家谱网络（见表 *1*）的持久区间包含我们在例*3.6*中考察的六边形网络的持久区间，后者是前者的子集。我们可以通过将六边形网络中的非平凡区间与 *Tikopia* 网络中的对应区间相匹配，在 *Tikopia* 网络和六边形网络的持久图之间构造一个双射。然后，将 *Tikopia* 网络中任何形如 $[a, b)$ 的额外区间映射到平凡区间 $[\frac{a+b}{2}, \frac{a+b}{2})$。（细心的读者可能会注意到，这并不能明显构成一个双射，但集合论中有一种标准方法可以对其进行调整，使其成为双射。）

此映射如图 *6*（右）所示。在这里，$[1, 7)$ 映射到 $[4, 4)$。由于这对点之间的距离比此双射中的任何其他点对都远，因此这两个网络的瓶颈距离至多为三，因为我们对所有可能的双射取下确界。反过来，六边形持久图中没有任何区间与 $[1, 7)$ 的距离小于 *3*，因此瓶颈距离至少为三。因此，这两个持久图的瓶颈距离恰好为 *3*。

假设两个网络各自都是连通的，并且都可以等距嵌入到 $\mathbb{R}^n$ 中。稳定性定理 [18]保证，如果这两个嵌入之间的豪斯多夫距离为 δ，那么相应持久图的瓶颈距离至多为 δ。例如，如果 PH_1 持久图相差 δ，那么在网络中尝试配对环时，对于任何等距嵌入，都至少有一对环在该嵌入中的距离为 δ。在第 6.1 节中，我们将这一思路应用于大量谱系网络与社交网络。



六边形网络持久性图　　Tikopia网络持久性图　　瓶颈双射

图 6：左：展示了图4(b)中六边形网络的持久图。中：展示了图 1（左）中 Tikopia 家谱网络的持久图。右：展示了六边形网络与 Tikopia 家庭网络的持久区间之间的瓶颈双射。橙色线条表示哪些点与形如（a,a）且满足$a \in \mathbb{R}$的点相匹配。

4.2. 持久性曲线

对于我们所考虑的网络数据，持久图掩盖了一个我们认为很重要的关键差异：持久区间的数量。一个简单的例子是，考虑形如$V = \{1, 2, \ldots\ldots、n\}$，且边形如 $\{i、i+1\}$（其中 $1 \le i < n$）的网络。对于 $n \ge 2$，任何这种类型的网络都将具有持久区间 $[0,1) \times (n-1)$ 和 $[0,\infty) \times 1$。然而，绘制持久图时，我们只能“看到”两个点：$(0,1)$ 和 $(0,\infty)$。

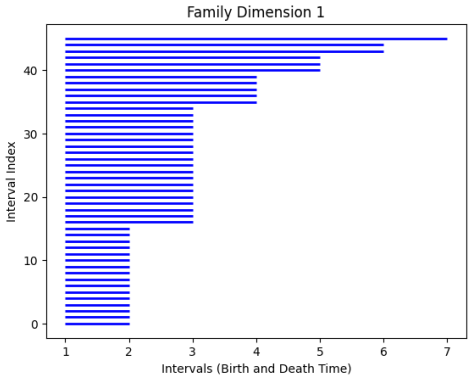
为解决这一局限，我们引入持久曲线 这一概念，作为可视化网络持久同调的新方法（见定义 9）。网络的持久曲线与持久图之间的区别在于，持久曲线还包含特定类型区间的数量。要绘制持久曲线，我们首先计算网络的持久区间，然后按持久性对给定维度的区间进行排序，并绘制成条形图。例如，在维度 1 中，Tikopia 家谱网络有十三个 $[1,2)$ 区间、十九个 $[1,3)$ 区间，等等；这些区间依次堆叠，如图 7（左）所示，从而形成我们称为条形码的图形。要绘制相应的持久曲线，则如图 7（右）所示，将每个后续条形的端点连接起来。

在一维情况下，由于我们所考虑的网络是无权、无向且连通的，因此所有持久区间的出生时间都从 1 开始。这意味着在这一维度中，所得条形图同时也是每个持久区间死亡时间的图示。对于具有不同出生时间的高维情况，我们还会绘制持久区间的长度；但为简单起见，我们像一维情况一样从 1 开始。

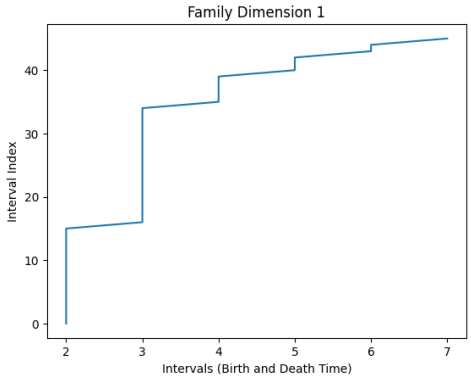
网络持久性曲线的形式化定义如下。

定义*9*。（持久性曲线）设 $G = (V,E)$ 是一个顶点集和边集均非空的网络。设$\{[a_j, b_j)\}$为每个 $\sigma_j \in PH_n(G)$（其中 $j \in \mathbb{N}$）的所有持久区间构成的集合。对于所有 $n \in \mathbb{N}$，持久性曲线 $PH_n(G)$ 是点集 $\{(b_j - (a_j - 1), j)\}$ 的线性插值，其中 $b_{j-1} - (a_{j-1} - 1) \le b_j - (a_j - 1)$。

将持久区间可视化为曲线，使我们能够以类似于持久图的方式比较不同网络的持久同调，同时保留不同的信息。具体而言，我们可以看到给定持久性下有多少个区间，而持久图仅表示是否存在这样的区间。下文中，我们通常会在同一坐标轴上绘制多个网络的持久曲线，以展示不同网络的持久同调有哪些差异（参见第6节）。



Tikopia 网络条形码



Tikopia 网络持久性曲线

图 7：左：展示了 Tikopia 家谱网络在维度 1 上的条形码。各条形由表 1 中给出的持久区间构成。右：展示了图 1 中蒂科皮亚网络对应的持久曲线。

5. 数据

本文考察的数据分为两类：谱系网络数据和其他社交网络数据。我们考察的谱系网络取自 [14], 中发现的97个谱系网络，其规模从$n = 17$ 到5,016个体不等。我们使用的社交网络数据取自 [20, 21, 22, 23]提供的27个不同社交网络，其规模从$n = 16$ 到2,539个体不等。（有关该数据集的完整说明，请参见附录中的表 2。）

尽管有许多规模更大的家谱网络和社交网络数据集可供使用，但计算持久区间所用算法在时间和空间复杂度方面的限制，使我们无法处理这些数据集。我们使用的程序名为*Ripser*（来自 Python 软件包 Ripser） [24], ，其时间和空间复杂度为$O((n+m)^3)$，其中n 是个体数量，m 是网络中的边数。网络中的单形数量为$n+m$。在我们研究的家谱网络中，单形数量介于$n + m = 41$ 至 15,735 之间；在我们研究的社交网络中，单形数量介于$n + m = 41$ 至 19,056 之间。

为了了解数据的完整或不完整会如何影响网络的持久区间，我们还考察了从若干规模大得多的家谱网络和社交网络中抽样得到的子网络。这些抽样网络通过随机选择一个只有一个邻居的个体（即度为 1 的顶点）来构建，然后从该个体开始执行广度优先搜索，找出网络中与该个体距离最近的 η 个体。由于 Ripser 在空间和计算能力方面的限制，我们选择 $600 \leq \eta \leq 3,000$，以确保能够计算这些抽样网络的持久区间。我们总共从四个不同的家谱网络和四个不同的社交网络中进行了抽样。这些网络分别是 Advogat、LastFM Asia、Deezer HU 和 Deezer RO 社交网络，以及表 2 中所示的 96–99 号家谱网络。我们分别从这些网络中抽样五次，共得到 20 个抽样家谱网络和 20 个抽样社交网络。我们从度为 1 的顶点开始广度优先搜索，是为了确保抽样网络既包含原始网络边界上的顶点，也包含其内部的顶点，从而更好地模拟所抽样的原始家谱网络和社交网络的结构。

除了我们研究的（i）谱系网络与社交网络以及（ii）这些网络的抽样版本之外，我们还研究了我们称为（iii）非典型家谱网络的网络。其中有一些

家谱网络似乎在创建时并未试图呈现全部家族关系，甚至连其中一部分也没有。例如，表 2 中列为非典型家谱网络 2 的美国总统网络，沿着总统之间最短的家谱路径构建，省略了无关关系。我们研究了若干此类非典型家谱网络；尤其是在持久同调方面，它们与我们研究的较标准的家谱网络形成对比。关于我们研究的（i）谱系网络与社交网络、（ii）抽样家谱网络与抽样社交网络，以及（iii）非典型家谱网络的介绍，见附录末尾。

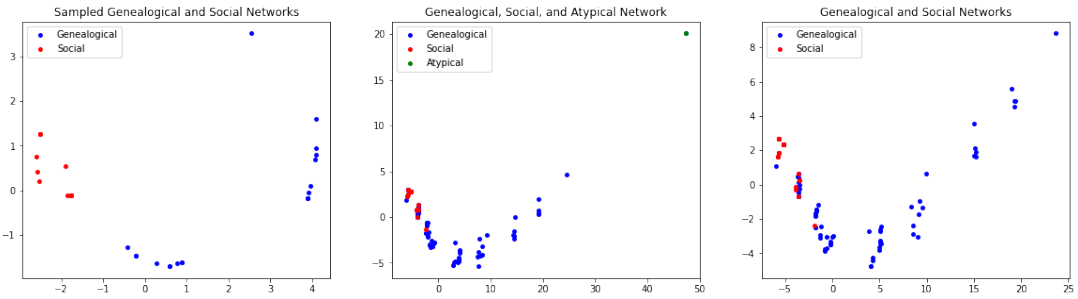


图 8：展示了网络之间瓶颈距离的主成分分析投影。左：展示了抽样的二十个家谱网络和抽样的社交网络之间各自的瓶颈距离。中：展示了我们所考察的家谱网络、社交网络和非典型家谱网络之间的瓶颈距离。右：仅展示了我们所考察的家谱网络和社交网络之间的中间面板中的瓶颈距离。

6. 结果

在此，我们使用第 4 节中定义的 (a) 瓶颈距离和 (b) 持久曲线来比较家谱网络与其他社交网络（分别参见定义 8 和定义 9）。对于略过第 3 节和第 4 节的读者，瓶颈距离根据两个网络持久同调的差异，给出它们之间的距离。持久曲线则提供了一种更详细地可视化这种差异的方法（参见图 7）。

6.1 使用瓶颈距离比较网络

在此，我们计算所考察的社交网络和家谱网络中每一对网络之间的瓶颈距离。为了可视化这些结果，我们使用主成分分析确定解释方差最大的两个分量，然后在 $\mathbb{R}^2$ 中绘制这些数据（见图 8）。

从图8的各个部分可以看出，家谱网络通常与社交网络相互分离，并形成易于区分的簇。对于抽样网络（左图），我们可以轻松区分谱系网络与社交网络，并识别出至少两个不同的谱系网络子类。然而，瓶颈距离在区分非抽样的家谱网络与社交网络方面效果较差（中图和右图）。例外的是非典型家谱网络，它们的持久区间与其他所有网络的差异足够显著，因此可以作为第三类网络加以区分（中图）。

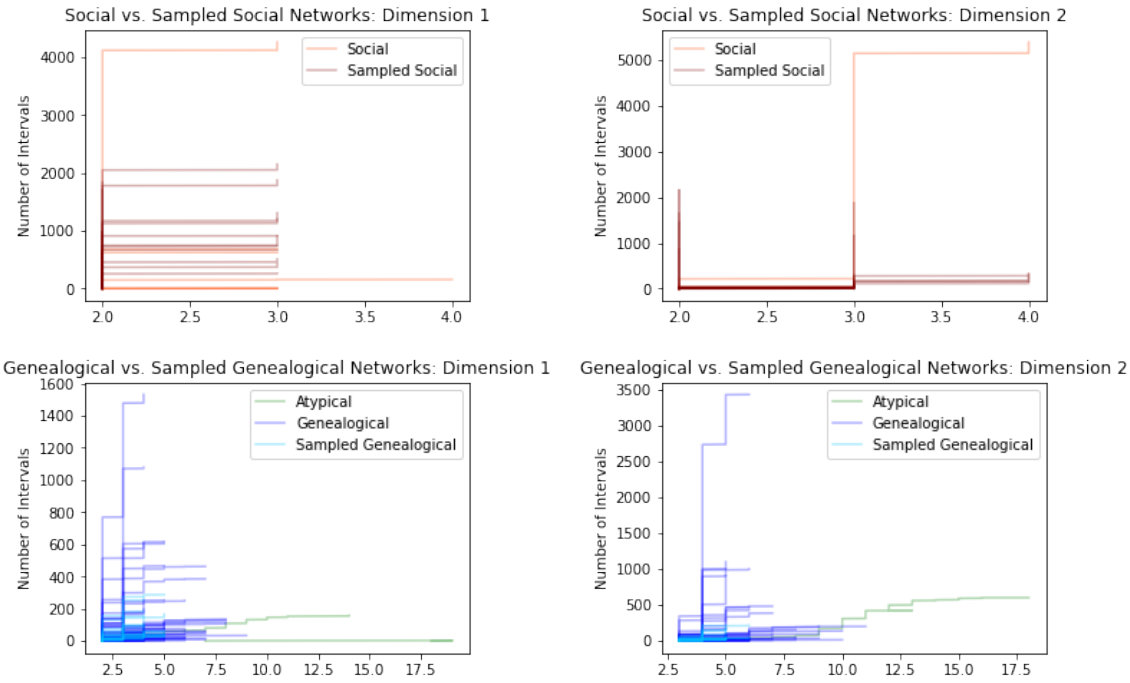


图 9：按维度和网络类型分组的完整网络与抽样网络的持久性曲线比较。上排：对社交网络进行抽样通常只会沿一个轴拉伸持久性曲线，而不会影响另一轴。下排：对家谱网络进行抽样通常会使持久性曲线沿两个轴收缩。总体而言，社交网络的平均斜率往往会在抽样后增大，而家谱网络的平均斜率则会减小。

6.2. 使用持久性曲线比较谱系网络与社会网络

持久曲线为比较网络提供了一种全新的替代方法。与瓶颈距离相比，使用这些曲线的优势在于，它们能更详细地呈现持久区间数量在不同网络之间的变化情况。这使我们能够更好地区分谱系网络与社交网络的结构，并分别观察谱系网络和社交网络各自共有的结构。

图 9 中分别以蓝色和红色显示了未经抽样的家谱网络和社交网络的持久性曲线。非典型家谱网络以绿色显示。社交网络在维度 1 和维度 2 上的持久性曲线都相当陡直。对于维度 1，这表明社交网络中的大多数环都接近平凡环；原因可能是它们的周长相对较小，也可能是它们可以分解为若干周长较小的环的并集。尤其是，大多数社交网络的最大死亡时间为 3（见定义 2），这意味着存在一个环基，其中各环的最大周长不超过 9。换言之，任何周长为 10 或以上的环都可以分解为较小环的并集。对于维度 2，持久性曲线的陡峭程度表明，某些顶点对之间存在许多条彼此不同但相似的路径。

在对比……中，家谱网络的持久曲线轮廓要平缓得多，

这表明大多数环都相当长，顶点对之间的“替代路径”也较少。极端情况下，非典型家谱网络在维度 1 上几乎是平坦的，这反映出这些非典型网络是有意构建的，环的数量极少。在维度 2 上，非典型网络呈现出与大多数典型家谱网络相似的斜率，但这些网络中的替代路径规模要大得多。这可能是由于有大量个体仅被加入网络以连接相距遥远的个体，例如总统。在典型的家谱网络中，这些个体之间的额外关系会使大型环得以分解，但在非典型家谱网络中，情况并非如此。

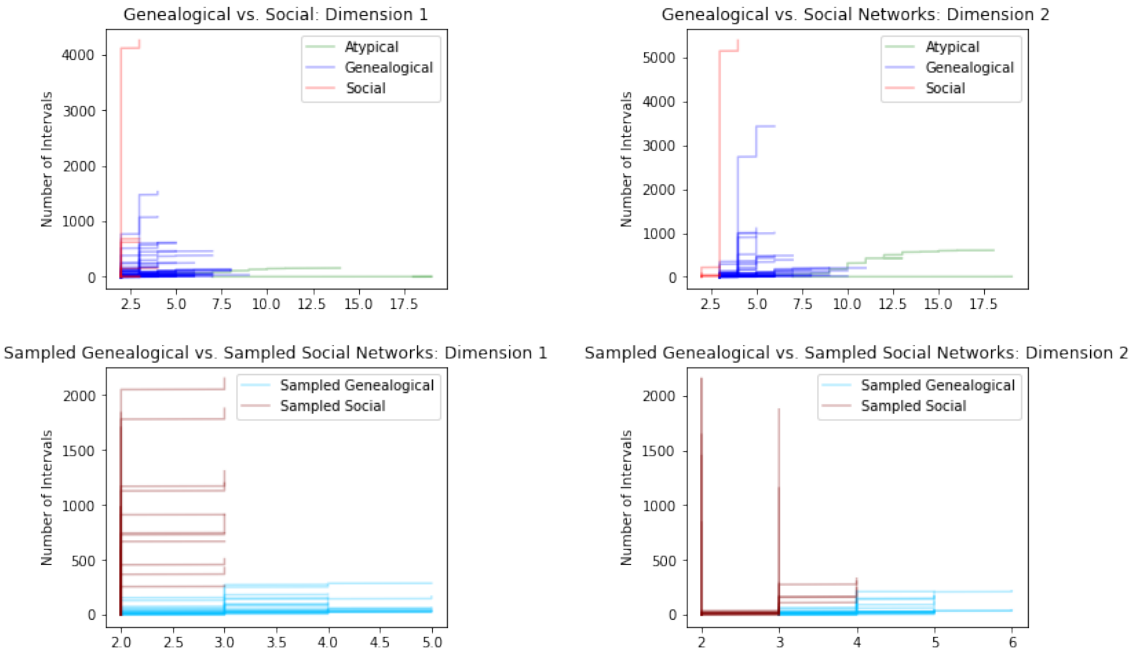


图 10：上排：按类型比较完整网络的持久性曲线。下排：按类型比较抽样网络的持久性曲线，不包括非典型家谱网络。在每个维度中，家谱网络的平均斜率通常低于社交网络的平均斜率。非典型家谱网络的平均斜率最低，总长度则大得多。与完整网络相比，抽样网络的平均斜率表现出更明显的这一趋势。

在图 10 中，我们可以看到抽样后的家谱网络和抽样后的社交网络的持久曲线，分别以蓝色和红色显示。非典型家谱网络以绿色显示。同样，社交网络在两个维度上的持久曲线都相当垂直，不过这些曲线没有未抽样社交网络的曲线那么高。这表明，对社交网络进行抽样时，它保留了相近比例的近乎平凡的环，但可能会失去许多顶点之间在维度 2 中出现的替代路径。相比之下，对于家谱网络，持久曲线表明，非常大的环完全消失，同时近乎平凡的环也按比例减少。在维度 2 中，家谱网络中替代路径的损失比社交网络更严重。因此，尽管抽样缩小了社交网络和家谱网络持久曲线的尺度，但两者在视觉上仍然迥异。

与瓶颈距离图一样，家谱网络和社交网络似乎聚集在一起，因为它们具有相似类型的持久曲线。事实上，无论网络是否经过抽样，情况都是如此。这表明，即使数据不完整，社交网络和家谱网络仍具有可区分的持久同调，至少在我们考察的尺度上如此。

值得一提的是，虽然瓶颈距离图在一定程度上向我们展示了谱系网络与社交网络有何 不同，但持久性曲线向我们展示了这些差异是什么 。不过，图 8 中的距离图确实具有简洁明了的优点，因此或许可以用来更快地识别网络之间那些不如谱系网络与社交网络之间的差异那样明显的区别。

6.3. 连接

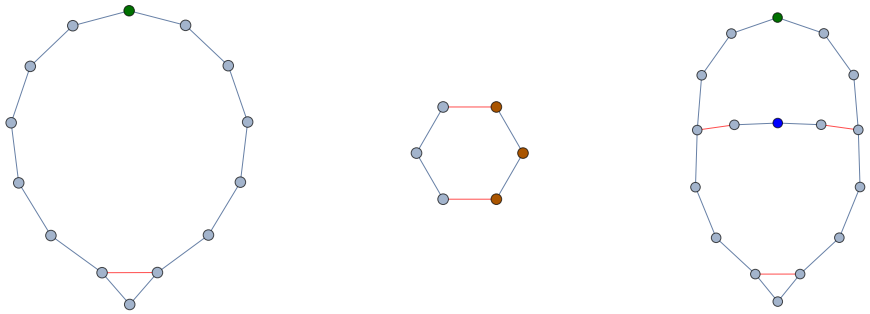
我们还可以利用持久同调研究网络的性质，例如连通分支的数量、环的典型大小，甚至数据中的“缺失连接”。对于谱系网络与社交网络，我们可以将这些数学概念转化为更熟悉的概念，例如家族群体或共同祖先。这也使我们能够通过将关于家庭或社交网络的标准假设转换为持久性的语言，对此类网络的持久同调提出猜想。

在 0维中，连通分支的数量决定了 $[0,\infty)$ 区间的数量，而不同顶点的总数等于 $[0,\infty)$ 区间的数量加上 $[0,1)$ 区间的数量。在谱系网络的语境中，每个连通分量都代表一个家族群体，该群体与其他家族群体之间不存在任何已知连接。因此，如果给定的家族网络确实是由亲属组成的单一“家庭”，那么它应当恰好有一个 $[0,\infty)$ 区间。在我们的 Tikopia 示例中，有八个$[0,\infty)$ 区间，每个区间都恰好对应于该谱系网络的一个连通分量。（请注意，图 1（左）仅展示了其中最大的连通分量。）在这个示例中，其他大多数“家族群体”实际上都是网络中没有关系边的个体。

在社交网络中，连通分支形成了可以称为朋友圈的群体。与家谱网络不同，社交网络中的边通常很少受到形成条件的限制。因此，对于这种情形下 $[0,\infty)$ 区间的数量，我们通常无法提出猜想。不过，按照第5节所述对任意网络进行采样，得到的新网络都将只有一个 $[0,\infty)$ 区间。

转到维度 1，这一维度中的持久区间描述了每个连通分量的内部结构。在足够大的家谱网络中，我们会看到三类特征，称为共同祖先、结合环和混合环。当一个个体的两个后代结合或共同生育子女时，就会出现共同祖先环 。我们用结合环 来指通过结合边以及连接两个兄弟姐妹的边形成环的情况。最后一种值得注意的环是混合环，它们由其他任何亲子边和结合边的组合形成，包括所有不属于严格意义上的共同祖先环或结合环的情况。这三类环在图 11 中有所展示，其中婚姻边以红色表示，亲子边以蓝色表示。图 11(a) 展示了一个共同祖先。图 11(b) 是结合环的一个例子：一个家庭中的两个兄弟姐妹分别与另一个家庭中的两个兄弟姐妹结合，其中每个家庭只显示了一位家长。图11(c) 给出了一个 θ-环的例子，它由一个共同祖先环和两个相互重叠的混合环组成。这个例子来自一个家庭的兄弟姐妹与另一个家庭的表亲结婚。从理论上说，这些环的长度可以任意，但文化规范会以不同方式影响每种环的典型规模和数量。记录实践和不完整数据也会影响这些环是否会出现在给定的数据集中。因此，描述这些环并了解相关文化，或许有助于发现记录数据中的错误。

反过来，了解高保真数据集中环的分布情况，有助于识别潜在的文化规范，并推断不完整数据集中哪些个体可能缺失。



(a) 共同祖先环　　(b) 联合环　　(c) θ-环

图11：左：一个共同祖先环。最上方的顶点是最下方顶点的共同祖先。水平红线表示婚姻，其他所有线都表示亲子边。中：一个联合环，具体来说，是第2节中描述的双重表亲情形。最左侧和最右侧的顶点分别是其相邻顶点的父母。两条水平红线表示婚姻边。右：由一个共同祖先环和两个重叠混合环构成的 θ-环。

由于许多文化会避免与近亲结婚，共同祖先环往往具有相当大的周长。在蒂科皮亚网络（见图 1 ）中，我们看到死亡值高达 7 的持久区间，对应于周长至少为 21 个体的环，这些环似乎是共同祖先环。这在一定程度上解释了为什么持久曲线如此平坦：网络中的极小共同祖先环相对较少，但它们的持久性非常高。更准确地说，如果到结合的距离（共同祖先环中的个体总数）为 n，那么该环的持久性为$\lfloor n/3 \rfloor$。不过，持久同调的代表元仅包含这些环的一个基，而不是所有可能的不同环。特别是，如果两个混合环各自都比共同祖先环短，那么一个较大的共同祖先环会分解为这两个混合环的并集，如图11(c)所示。持久同调反映的将是这两个较小环的大小，而不是较大共同祖先环的大小。值得注意的是，我们可以识别出基中实际选取的环，但我们使用的软件不提供这方面的信息，而且网络规模过大，无法手动识别这些环。

在社交网络中，我们发现持久性很高的环相当罕见。例如，要形成持久性为 3 的环，就需要一个周长至少为 9 的环，而且环上任意两个顶点之间都不存在更短的路径。小世界效应，或者更通俗地说，六度分隔现象，可能会限制社交网络的最大持久性。我们的示例数据集中也体现了这一点：除一个社交网络外，其余社交网络的最大持久性均为 3。

7. 结论

本文探讨谱系网络的持久同调结构。研究的动机在于观察到，家族联系往往形成于一个固定的中间距离范围内，这使谱系网络在同调性质上有别于大多数其他社交网络。我们还引入了持久曲线的概念，可用来概括和比较任何

网络的持久同调结构。我们还将共同祖先环等特定谱系结构与同调对象联系起来。

我们发现，即使存在不完整数据，同调分析在谱系研究中仍然有用。我们注意到，由于记录实践和数据不完整（这是现实家谱网络的普遍特征）造成的缺失数据，会限制给定数据集中出现的环的类型。因此，描述这些环并了解相关文化，可能有助于识别记录数据中的错误。反过来，了解高保真数据集中环的分布，有助于识别其背后的文化规范，并推断不完整数据集中哪些地方缺失了个体。

这项工作有几个值得探索的拓展方向。例如，我们的研究清楚地表明，确实有必要分析大型网络的持久同调——这类网络至少包含数万个节点，因为家庭形成通常发生在这一规模上。我们所依赖的 Ripser 库无法处理如此规模的网络。此外，我们非常希望创建能够反映人类家庭网络实际同调特征的随机图模型——我们团队在这方面的首次尝试相当成功，适用于包含数百个节点的网络 [25]。更广泛地说，我们需要对真实人类家庭网络进行建模。现有的所有数据源都只是存在偏差、有限且带有噪声的子网络，而谱系学界真正关注的是真实网络。在此背景下，信号去噪、图像修复和图外推等工具可能会有所帮助。最后，谱系网络的一个重要方面，是各种佐证文件／元数据与通过它们发现的连接之间的关系。例如，可以考虑在预算有限时采用最佳文档收集策略，或采用能够公平呈现少数群体信息的文档收集方式，因为这类信息往往代表不足。

8. 声明

8.1. 数据和材料的可获取性

本研究期间生成和/或分析的数据集链接见表 2。可在 https://github.com/AbigailJ32/The-persistent-homology-of- genealogical-networks 获取用于复现和扩展本研究的代码。

8.2. 利益冲突

作者声明不存在任何利益冲突。

8.3. 资金支持

ZB、BW 和 AJ 获得了 BYU CPMS CHIRP资助的支持。ZB 还获得了 NFS奖项#2137511、美国陆军研究办公室资助项目#W911NF-18-1-0244，以及詹姆斯·S·麦克唐纳基金会21世纪科学倡议—复杂系统学者奖资助项目#2200203的支持。BW 还获得了西蒙斯基金会资助项目#714015的支持。本文所载的观点和结论均为作者所有，不应被解释为代表美国陆军研究办公室或美国政府的明示或暗示的官方政策。尽管本文含有任何版权声明，美国政府仍有权为政府目的复制和分发本文的再版。

8.4. 作者贡献

实验设计：ZB、NC、BW、RW。实验实施：RF、RW。论文撰写：ZB、NC、TG、AJ、RS、BW、RW。所有作者均已阅读并批准最终稿。

8.5. 致谢

我们感谢 另外，感谢 Joseph Price 和 FamilySearch 工程研究团队提供的有益讨论。我们还感谢 Kolton Baldwin 帮助改进我们的代码和模拟。

9. 附录

这里列出了我们在持久同调计算中使用的谱系网络与社会网络（见第6节）。我们按网络类型区分数据集：友谊/熟人关系、社交媒体、合作/商业、疾病传播、信息共享、谱系网络，以及非典型谱系网络。我们还提供网络名称、网络中的顶点和边的数量，以及网络出处的引文。此外，特别感谢 Kolton Baldwin 为本文的数值模拟提供帮助。

表 2：社会网络和谱系网络数据集。

网络数据			
网络类型和名称	顶点	边	引文
社交网络			
友谊 和 **熟人关系**			
海豚	62	159	http://www-personal.umich.edu/~mejn/netdata/
扎卡里空手道俱乐部	34	78	http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/ucidata.htm#zachary
学生宿舍	217	2672	http://konect.cc/networks/moreno oz/
高地部落	16	58	http://konect.cc/networks/ucidata-gama/
七年级学生	29	376	http://konect.cc/networks/moreno seventh/
医生	241	1098	http://konect.cc/networks/moreno innovation/
高中	70	366	http://konect.cc/networks/moreno 高中/
荷兰大学	32	354	http://konect.cc/networks/moreno vdb/
Sampson修道院	25	322	http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:data:esna3:sampson
青少年健康	2539	12969	http://konect.cc/networks/moreno health/
Hamsterster朋友	2952	12534	http://konect.cc/networks/petster-hamster-friend/
社交网络1	32	220	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as1.net
社会网络 2	32	191	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as2.net
社会网络 5	32	90	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as5.net
社会网络 7	32	61	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as7.net
社会网络 8	32	79	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as8.net
社会网络 9	32	58	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as9.net
社交媒体			
高科技企业	33	124.5	https://networkrepository.com/soc-firm-hi-tech.php
Wiki-Vote	889	2.9K	https://networkrepository.com/soc-wiki-Vote.php
FB-PAGES-FOOD	620	2.1K	https://networkrepository.com/fb-pages-food.php
Advogato	6541	51127	http://konect.cc/networks/advogato/
LastFM Asia	7624	27806	https://snap.stanford.edu/data/feather-lastfm-social.html
Deezer HU	47538	222887	https://snap.stanford.edu/data/gemsec-Deezer.html
Deezer RO	41773	125826	https://snap.stanford.edu/data/gemsec-Deezer.html
合作&商业			
社交网络 4	32	218	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as4.net
社交网络 6	32	103	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as6.net
社交网络 11	32	83	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as11.net
社交网络 12	32	65	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as12.net
疾病传播			
塔罗交换	22	78	http://konect.cc/networks/moreno taro/

网络名称与类型	顶点	边	引文
信息共享			
社会网络 3	32	119	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as3.net
社会网络 10	32	80	http://vlado.fmf.uni-lj.si/pub/networks/doc/ECPR/assign.1/as10.net
家谱网络			
家谱网络 1	310	322	https://www.kinsources.net/kidarep/dataset-209-mowanjum-kalumburu.xhtml
家谱网络 2	303	537	https://www.kinsources.net/kidarep/dataset-2-mbuti-village-1957年-af03.xhtml
亲属网络3	371	718	https://www.kinsources.net/kidarep/dataset-58-奥吉布瓦人-1930年-nd07.xhtml
亲属网络4	795	1387	https://www.kinsources.net/kidarep/dataset-150-阿丘阿尔人-帕斯塔萨.xhtml
亲属网络5	636	1151	https://www.kinsources.net/kidarep/数据集-92-钦楚人-1940年-as02.xhtml
亲属网络6	782	1366	https://www.kinsources.net/kidarep/dataset-28-trio-20世纪60年代.xhtml
亲属网络7	128	202	https://www.kinsources.net/kidarep/dataset-23-shoshone-1880年-nd11.xhtml
亲属网络8	439	626	https://www.kinsources.net/kidarep/dataset-70-创世记.xhtml
谱系网络 9	244	481	https://www.kinsources.net/kidarep/dataset-66-waimiri-atroari.xhtml
谱系网络 10	410	746	https://www.kinsources.net/kidarep/dataset-240-kodiak.xhtml
谱系网络 11	337	572	https://www.kinsources.net/kidarep/dataset-51-wilcania.xhtml
谱系网络 12	216	378	https://www.kinsources.net/kidarep/dataset-22-阿伊努-1880年-as01.xhtml
谱系网络 13	77	134	https://www.kinsources.net/kidarep/dataset-69-slavey-1911-nd12.xhtml
家谱网络 14	815	1582	https://www.kinsources.net/kidarep/dataset-7-pakaa-nova.xhtml
家谱网络 15	20	28	https://www.kinsources.net/kidarep/dataset-38-wanindiljaugwa-1948-au06.xhtml
家谱网络 16	219	371	https://www.kinsources.net/kidarep/dataset-171-suya.xhtml
家谱网络 17	17	24	https://www.kinsources.net/kidarep/dataset-31-family.xhtml
家谱网络 18	168	221	https://www.kinsources.net/kidarep/数据集-14-labrador-inuit-1776年-nu02.xhtml
家谱网络 19	64	109	https://www.kinsources.net/kidarep/dataset-91-takamiut-1927年-64-nu03.xhtml
家谱网络 20	1423	3211	https://www.kinsources.net/kidarep/dataset-258-todas.xhtml
家谱网络 21	645	1097	https://www.kinsources.net/kidarep/dataset-65-igluligmiut-1961年-nu07.xhtml
家谱网络 22	4463	8416	https://www.kinsources.net/kidarep/dataset-115-charlevoix.xhtml
谱系网络 23	48	86	https://www.kinsources.net/kidarep/dataset-41-韦达人-1905-as04.xhtml
谱系网络 24	104	172	https://www.kinsources.net/kidarep/dataset-71-igluligmiut-1960-61-nu08.xhtml
谱系网络 25	1263	2021	https://www.kinsources.net/kidarep/dataset-223-samburu.xhtml
谱系网络 26	80	132	https://www.kinsources.net/kidarep/dataset-10-阿帕奇人-1932-nd01.xhtml
谱系网络 27	1269	2395	https://www.kinsources.net/kidarep/dataset-24-ayd-nl-yoruk-2005.xhtml
谱系网络 28	299	532	https://www.kinsources.net/kidarep/dataset-13-tory.xhtml
谱系网络 29	19	30	https://www.kinsources.net/kidarep/dataset-21-ngatatjara-1966-au04.xhtml
谱系网络 30	399	592	https://www.kinsources.net/kidarep/dataset-204-dogon-konsogu-donyu.xhtml
谱系网络 31	377	712	https://www.kinsources.net/kidarep/dataset-49-alyawarra-1971-au01.xhtml
谱系网络 32	1263	2021	https://www.kinsources.net/kidarep/dataset-223-samburu.xhtml
谱系网络 33	118	192	https://www.kinsources.net/kidarep/dataset-39-eyak-1890.xhtml
谱系网络 34	98	161	https://www.kinsources.net/kidarep/数据集-75-nunamiut-1885-nu11.xhtml
谱系网络 35	479	830	https://www.kinsources.net/kidarep/dataset-19-奥吉布瓦人-1949年-nd08.xhtml
谱系网络 36	1695	3206	https://www.kinsources.net/kidarep/dataset-103-tikuna-arara.xhtml
谱系网络 37	256	441	https://github.com/AbigailJ32/The-persistent-homology-of-genealogical-networks
谱系网络 38	798	1416	https://www.kinsources.net/kidarep/dataset-229-努库里尔马人-廷加.xhtml
谱系网络 39	738	1212	https://www.kinsources.net/kidarep/dataset-32-亚拉尔迪人.xhtml
谱系网络 40	525	855	https://github.com/AbigailJ32/The-persistent-homology-of-genealogical-networks
谱系网络41	619	1224	https://www.kinsources.net/kidarep/数据集-251-努尼瓦克.xhtml
谱系网络42	3008	6074	https://www.kinsources.net/kidarep/dataset-80-torshan.xhtml
谱系网络43	278	464	https://www.kinsources.net/kidarep/dataset-62-dogrib-1911-25-59-nd04.xhtml
谱系网络 44	105	172	https://www.kinsources.net/kidarep/dataset-5-konkama-1931-44-51-eu02.xhtml
谱系网络 45	240	395	https://www.kinsources.net/kidarep/dataset-158-tikar.xhtml
谱系网络 46	4178	7351	https://www.kinsources.net/kidarep/dataset-45-奥比杜斯.xhtml
谱系网络47	216	286	https://www.kinsources.net/kidarep/dataset-254-port-keats.xhtml
谱系网络48	147	242	https://www.kinsources.net/kidarep/dataset-78-pul-eliya-1954-simpler-version.xhtml
谱系网络49	277	516	https://www.kinsources.net/kidarep/数据集-213-萨尔米.xhtml
谱系网络50	330	622	https://www.kinsources.net/kidarep/dataset-73-parakana.xhtml
谱系网络51	35	53	https://www.kinsources.net/kidarep/dataset-81-gundangborn-1948-au02.xhtml
谱系网络52	48	76	https://www.kinsources.net/kidarep/dataset-84-hare-1956-nd05.xhtml

网络名称与类型	顶点	边	引文
亲属网络53	105	245	https://www.kinsources.net/kidarep/dataset-87-arara.xhtml
谱系网络54	116	220	https://www.kinsources.net/kidarep/数据集-89-Nunamiut-1960年-nu13.xhtml
谱系网络55	116	176	https://www.kinsources.net/kidarep/dataset-226-羯族.xhtml
谱系网络56	657	1166	https://www.kinsources.net/kidarep/dataset-27-纽恩加人.xhtml
谱系网络 57	659	1288	https://www.kinsources.net/kidarep/数据集-3-anuta-1972.xhtmlj
谱系网络 58	112	182	https://www.kinsources.net/kidarep/dataset-15-oodnadatta.xhtml
谱系网络 59	218	353	https://www.kinsources.net/kidarep/数据集-17-lainiovouma-1952-eu03.xhtml
亲属网络60	90	119	https://www.kinsources.net/kidarep/数据集-12-miwuyt-1967-au03.xhtml
亲属网络61	289	477	https://www.kinsources.net/kidarep/dataset-9-konkama-1951-eu01.xhtml
亲属网络62	1463	1969	https://www.kinsources.net/kidarep/dataset-306-nobles-ile-de-france-1000-1440.xhtml
亲属网络63	4109	6517	https://www.kinsources.net/kidarep/dataset-287-duu-rea.xhtml
亲属网络64	29	48	https://www.kinsources.net/kidarep/数据集-46-hatfields-and-mccoys.xhtml
亲属网络65	40	59	https://www.kinsources.net/kidarep/数据集-33-安马萨利克-1884-nu01.xhtml
谱系网络 66	294	441	https://www.kinsources.net/kidarep/dataset-18-tikopia-1930.xhtml
谱系网络 67	502	786	https://www.kinsources.net/kidarep/dataset-34-netsilik-1922-nu09.xhtml
谱系网络 68	83	126	https://www.kinsources.net/kidarep/dataset-8-semang-1924-50-as03.xhtml
谱系网络 69	95	157	https://www.kinsources.net/kidarep/dataset-4-shoshone-1860-nd10.xhtml
谱系网络 70	2588	5651	https://www.kinsources.net/kidarep/dataset-61-kelkummer.xhtml
谱系网络 71	88	144	https://www.kinsources.net/kidarep/dataset-77-apache-1935年-nd02.xhtml
谱系网络72	1513	2217	https://www.kinsources.net/kidarep/dataset-90-奥马哈人-1880年.xhtml
谱系网络73	3014	5454	https://www.kinsources.net/kidarep/dataset-128-ammonni.xhtml
谱系网络74	139	201	https://www.kinsources.net/kidarep/dataset-79-paiute-1880年-nd09.xhtml
谱系网络75	5016	10719	https://www.kinsources.net/kidarep/dataset-249-巴鲁亚人.xhtml
谱系网络76	125	202	https://www.kinsources.net/kidarep/dataset-242-特林吉特人.xhtml
谱系网络77	272	445	https://www.kinsources.net/kidarep/数据集-36-科珀人-1922-nu10.xhtml
谱系网络 78	378	609	https://www.kinsources.net/kidarep/dataset-52-apache-1936-nd03.xhtml
谱系网络 79	926	1951	https://www.kinsources.net/kidarep/dataset-68-苏鲁伊人.xhtml
谱系网络 80	706	1177	https://www.kinsources.net/kidarep/dataset-60-mbuti-forest-1957年-af02.xhtml
谱系网络 81	435	672	https://www.kinsources.net/kidarep/dataset-64-melombo.xhtml
谱系网络 82	128	114	https://www.kinsources.net/kidarep/dataset-164-凯因冈人.xhtml
谱系网络 83	169	275	https://www.kinsources.net/kidarep/dataset-11-top-of-the-mountain.xhtml
谱系网络 84	178	274	https://www.kinsources.net/kidarep/dataset-37-igluligmiut-1921-nu05.xhtml
谱系网络 85	87	111	https://www.kinsources.net/kidarep/dataset-216-蒂维人.xhtml
谱系网络 86	2049	4159	https://www.kinsources.net/kidarep/dataset-35-chuukese-1947-1940.xhtml
亲属网络87	868	980	https://www.kinsources.net/kidarep/dataset-20-saudi-royal-genealogy.xhtml
亲属网络88	2821	5079	https://www.kinsources.net/kidarep/dataset-30-manus-1929.xhtml
亲属网络89	454	980	https://www.kinsources.net/kidarep/dataset-74-阿拉维特人.xhtml
谱系网络 90	304	472	https://www.kinsources.net/kidarep/dataset-42-nunamiut-tareumiut-1900年-nu12.xhtml
谱系网络 91	367	671	https://www.kinsources.net/kidarep/dataset-48-wanindiljaugwa-1941年-au05.xhtml
谱系网络 92	3151	4289	https://www.kinsources.net/kidarep/数据集-54-feistritz-am-gael-1990年.xhtml
谱系网络 93	2975	5107	https://www.kinsources.net/kidarep/dataset-159-cocama-cocamilla.xhtml
谱系网络 94	585	1249	https://www.kinsources.net/kidarep/数据集-44-torres-strait.xhtml
谱系网络 95	334	530	https://www.kinsources.net/kidarep/dataset-6-igluligmiut-1949年-nu06.xhtml
谱系网络 96	9595	14988	https://www.kinsources.net/kidarep/dataset-93-sainte-catherine.xhtml
谱系网络 97	28586	51446	https://www.kinsources.net/kidarep/dataset-76-san-marino.xhtml
谱系网络 98	18645	32439	https://www.kinsources.net/kidarep/dataset-307-bwa-slam-biogsurvey.xhtml
谱系网络 99	8809	15643	https://www.kinsources.net/kidarep/dataset-194-kel-owey.xhtml
非典型家谱网络			
非典型家谱网络 1	429	705	（使用 FamilySearch.org 创建）
非典型家谱网络 2	2477	4015	https://www.kinsources.net/kidarep/dataset-56-us-presidents.xhtml

参考文献

[1] J. Kaplanis, A. Gordon, T. Shor, O. Weissbrod, D. Geiger, M. Wahl, M. Gershovits, B. Markus, M. Sheikh, M. Gymrek, G. Bhatia, D.G. MacArthur, A.L. Price, Y. Erlich, 《利用数百万亲属构成的家谱树对人口规模进行定量分析》，美国科学促进会，(2018) http://science.sciencemag.org/content/early/2018/02/28/science.aam9309。[2] K. Hamberger, M. Houseman, and D. R. White, 《亲属关系、阶级与社区》，载于《SAGE社会网络分析手册》，J. P. Scott and P. J. Carrington 编，Sage Publications Ltd., (2011), 第129–147页。[3] D. L. T. Rohde, S. Olson, and J. T. Chang, 《模拟所有现存人类的近期共同祖先》，《自然》, 431 (2004), 第562–566页。[4] J. Greenwood, N. Guner, G. Kocharkov, and C. Santos, 《与相似的人结婚：同类婚配与收入不平等》，Amer. Econ. Rev., 104 (2014), 第348–353页。[5] E. Malmi, A. Gionis, and A. Solin, 《通过计算推断的家谱网络揭示同类婚配的长期趋势》，载于《2018年万维网大会论文集 WWW 2018》，法国里昂，4月23-27日，(2018), 第883–892页, http://doi.acm.org/10.1145/3178876. 3186136。[6] G. Bloothooft, P. Christen, K. Mandemakers, and M. Schraagen, 《人口重建》，Springer, (2015)。[7] H. J. and C. A. Machado, 《结构化种群的研究——为一门困难且分歧重重的科学带来新希望》，Nat. Rev. Genet., 4 (2003), 第535–543页。[8] P. Hage and F. Harary, 《人类学中的结构模型》，剑桥大学出版社，剑桥，(1983)。[9] J. T. Chang, 《所有当代个体的近期共同祖先》，Adv. App. Prob., 31 (1999), 第1002–1026页，https://doi.org/10.1239/aap/1029955256。[10] V. Robins, M. Saadatfar, O. Delgado-Friedrichs, and A.P. Sheppard. 基于微型CT图像中多孔材料的拓扑持久性分析得到的渗流长度尺度。《水资源研究》Volume 52, Number 1, (2016), 第315-329页。[11] H. Lee, H. Kang, M.K. Chung, B. Kim, D.S. Lee. 从树状图角度研究持久脑网络同调。《IEEE医学影像汇刊》. Volume 31, Number 12, (2012), 第2267-2277页。[12] B. Mattia, B. Adriano, D.F. Barbara. 迈向音乐的拓扑指纹。《图像背景下计算拓扑学第六届国际研讨会论文集》. Volume 9667, (2016), 第 88-100 页。[13] S. Sintos, and P. Tsaparas. 利用强三元闭包刻画社交网络中的关系。《第20届ACM SIGKDD知识发现与数据挖掘国际会议论文集》，(2014), 第1466–1475页。[14] https://www.kinsources.net/browser/datasets.xhtml（最后访问于2022年6月21日）。

[15] 学生宿舍社交网络数据 http://konect.cc/networks/moreno oz/（最后访问于2022年1月20日）。[16] M.E.J. Newman. 网络中的模块度与社区结构，《美国国家科学院院刊》103(23): (2006), 第8577-8582页。[17] A. Hatcher. 《代数拓扑》，剑桥大学出版社，(2002)，美国马萨诸塞州剑桥。[18] D. Cohen-Steiner, H. Edelsbrunner, J. Harer. 持久图的稳定性。《离散与计算几何》37: (2007) 第103-120页。[19] A. Zomorodian,G. Carlsson. 计算持久同调。《离散与计算几何》33: (2005), 第249-274页。[20] http://konect.cc/networks/（最后访问于2022年6月10日）。[21] http://snap.stanford.edu/data/index.html#socnets（最后访问于2020年8月）。[22] http://networkrepository.com/soc.php（最后访问于2020年8月）。[23] http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:data:pajek:index（最后访问于2020年8月）。[24] Ripser Python软件包 https://anaconda.org/conda-forge/ripser（最后访问于2021年10月4日）。[25] R. Flores. 人类家庭网络建模，https://scholarsarchive.byu.edu/etd/9357/ (2021)。[26] C. J. Carstens, K. J. Horadam. 《合作网络的持久同调》，载于《工程中的数学问题》，第2013卷，文章编号815035，共7页，2013年。[27] G. Petri, M. Scolamiero, I. Donato, F. Vaccarino. 《网络与环：复杂网络的持久同调方法》，载于《2012年欧洲复杂系统会议论文集》，Gilbert T., Kirkilionis M., Nicolis G. (编). Springer复杂性会议论文集。Springer, 查姆。https://doi.org/10.1007/978-3-319-00395-5_15 (2013)[28] H. Kannan, E. Saucan, I. Roy, et al., 《通过离散莫尔斯理论研究无权复杂网络的持久同调》，《科学报告》9, 13817 (2019), 第1–18页。[29] M.E. Aktas, E. Akbas & A.E. Fatmaoui, 《网络的持久同调：方法与应用》。《应用网络科学》4, 61。https://doi.org/10.1007/s41109-019-0179-3 (2019)。[30] D. Horak, et al., 《复杂网络的持久同调》。《统计力学杂志》(2009)[31] A.N. Duman, H. Pirim, 《通过持久同调比较基因共表达网络》。《国际基因组学杂志》2018 (2018)[32] F. Chazal, L.J. Guibas, S.Y. Oudot,P. Skraba, 《黎曼流形中的基于持久性的聚类》。《美国计算机协会会刊》(JACM) 60(6), 41 (2013)[33] R. Vandaele, T. De Bie, Y. Saeys, 《局部拓扑数据分析：揭示趋近图结构拓扑的数据的全局结构》。载于《欧洲机器学习与数据库知识发现联合会议》。第19–36页。Springer (2018)

[34] *A.J. Blumberg & M.Lesnick*，《二参数持久同调的稳定性》，*arXiv https://arxiv.org/abs/2010.09628*（*2020*年）。[35] *N.A. Arafat*、*D. Basu*、*S. Bressan*，《图上的 ϵ-网诱导懒惰见证复形》，*arXiv https://arxiv.org/abs/2009.13071*（*2020*年）。[36] *N. Otter*、*M.A. Porter*、*U. Tillmann* 等，《持久同调计算路线图》，*EPJ Data Science*（*2017*年）