

# 基于持续同调的在线社交网络传播研究

邱吕琳, 张志坚, 蔡光程, 钟 慧

(昆明理工大学 理学院, 云南 昆明 650500)

**摘 要:** 为了对在线社交网络进行全面的可视化分析, 运用持续同调的计算方法对在线社交网络传播进行计算与分析。首先, 将在线社交网络看作距离子空间, 构造出该空间的不同子复形; 然后, 通过计算单纯复形的同调得到相应的贝蒂数和持续性图, 从而基于贝蒂数和持续性图获取社交网络的拓扑特征; 最后, 分析几个不同类型在线社交网络的传播特点, 将持续同调应用于在线社交网络传播, 从而得出在线社交网络的持续同调信息, 并提供一种描述在线社交网络传播的新方法。

**关键词:** 在线社交网络; 持续同调; 贝蒂数; 持续性图; 传播

**DOI:** 10.11907/rjdk.212151

**中图分类号:** TP391.1

**文献标识码:** A

**开放科学(资源服务)标识码(OSID):**

**文章编号:** 1672-7800(2022)002-0073-08



## Research on Online Social Network Communication Based on Persistent Homology

QIU Lyu-lin, ZHANG Zhi-jian, CAI Guang-cheng, ZHONG Hui

(School of Science, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract:** In order to conduct a comprehensive visual analysis of online social network, the method of persistent homology was used to calculate and analyze the communication of online social network. The online social network is regarded as a distance subspace and different subcomplex of this space are constructed. Then the corresponding Betti number and persistence graph are obtained by calculating the homology of simple complex, and the topological characteristics of social network are obtained based on Betti number and persistence diagram. Finally, the communication characteristics of several different types of online social networks are analyzed, and persistent homology is applied to online social network communication. In this way, the persistent homology information of online social networks is obtained, and a new method to describe the communication of online social networks is provided.

**Key Words:** online social network; persistent homology; Betti number; persistence diagram; spread

## 0 引言

随着当今社会经济与科技的飞速发展, 互联网已成为一个重要的新媒体, 在人们的学习、工作、生活等各个方面扮演着不可或缺的角色。社交网络以人际关系网为基础, 近年来发展迅速, 正逐渐融入人们的日常生活并发挥重要影响, 极大丰富了人们的各种需求。与此同时, 在线社交网络使得人们不再受到时间、空间限制, 可快捷接收外界信息, 并实时参与热点话题讨论, 发表自己的观点。在线社交网络在传播方面已显示出强大的影响力, 例如, “澳洲

森林山火”“全球新冠病毒”等新闻事件在微博不断实时发布消息等。在线社交网络信息传播对于生活和社会发展有着深远影响, 因此受到了研究者们广泛关注。

近年来在线社交网络信息传播研究取得了许多进展, 使人们能够从信息传播角度进一步理解社交网络的属性和规则。本文旨在分析在线社交网络的信息传播现象, 了解信息传播规律, 并根据相关规律进行引导, 从而更好地进行信息传播, 同时最大限度减少虚假消息和不良消息的传播。要想获取信息传播规律, 需要分析社交网络图。然而, 目前在分析社交网络图时, 采用最短路径、介数中心性、聚类系数和度分布等度量指标研究与认识在线社交网

**收稿日期:** 2021-08-25

**基金项目:** 国家自然科学基金项目(11461037); 云南省教育厅科学研究基金项目(2017ZZX133); 云南省高校联合青年项目(2017FH001-116)

**作者简介:** 邱吕琳(1996-), 女, 昆明理工大学理学院硕士研究生, 研究方向为社交网络; 张志坚(1980-), 男, 昆明理工大学理学院讲师, 研究方向为社交网络; 蔡光程(1965-), 男, 昆明理工大学理学院教授, 研究方向为数字图像处理、科学计算; 钟慧(1998-), 女, 昆明理工大学理学院硕士研究生, 研究方向为社交网络。本文通讯作者: 张志坚。

络结构。虽然这些指标足以提供网络图在特定方面的信息,但无法提供在线社交网络的全面特征。在线社交网络存在太多不确定信息,如网络节点分布很难确定,且在线社交网络的节点数量特别庞大,会随着时间不断变化,属于动态网络,所以对在线社交网络进行全面可视化是一个很有意义的问题。因此,本文从拓扑角度分析在线社交网络的拓扑结构,对在线社交网络进行全面分析。拓扑学中的持续同调为社交网络图中存储信息的全面可视化表示提供了一种新方法。

大量研究致力于理解在线社交网络的相关热点问题,主要研究方法已从早期社会科学研究领域的实证研究转向基于小数据集的理论研究与验证。本文主要利用持续同调方法和代数拓扑的一些知识提取在线社交网络的拓扑特征,并基于这些特征分析在线社交网络传播。

在线社交网络分析可总结为3个方面:在线社交网络结构特征和演化机制,在线社交网络群体行为形成和互动规律,在线社交网络信息传播规律和演化机制。在早期社会学与图论的研究基础上,Kempe等<sup>[1]</sup>提出的线性阈值模型LTM(Linear Threshold Model)和独立级联模型ICM(Independent Cascade Model)以及借鉴医学研究的传染病模型SIR(Susceptible Infected Recovered Model)<sup>[2]</sup>成为描述影响力传播的主要基础模型,这些模型描述了社交网络中影响力传播的随机性与累积性。后来许多研究者根据具体传播问题的特点扩展了基本模型,以反映传播模型的具体传播特性。目前,Aggarwal<sup>[3]</sup>已从数据挖掘的角度对社区发现、影响力计算、沟通可视化等方面进行了总结;Kwak等<sup>[4]</sup>基于Twitter的数据,对用户之间的关系、用户影响以及Twitter上话题的传播进行了实证分析;Fang等<sup>[5]</sup>介绍了在线社交网络相关研究,并从网络结构特征与演化机制、群体行为形成与互动规律、信息传播规律与演化机制3个维度提出进一步的研究方向。在在线社交网络信息传播方面,Guille等<sup>[6]</sup>从话题检测、传播建模以及识别有影响力的传播者3个方面进行了介绍,为社交网络中围绕信息扩散的工作提供全面的分析指导;Zinoviev<sup>[7]</sup>从传播者、传播路径和传播机制出发,提出博弈论模型来理解信息传播,并对已有研究进行总结。

鉴于在线社交网络的网络拓扑结构及特征,许多研究者提出利用“网络复杂性分析”的方法论进行相关问题分析与研究。Watts等<sup>[8]</sup>提出一个小世界网络模型,该模型更好地描述了从常规网络到随机网络的过渡;Barabdisi等<sup>[9]</sup>提出BA无标度网络模型,并根据模型的增长算法和优先连接特性演示了BA无标度网络演化过程。根据社交网络的基本功能和特点,田燕等<sup>[10]</sup>描述了不同类型在线社交网络的目的、特点及典型代表,分析了在线社交网络的拓扑结构和信息传输特点;候梦男等<sup>[11]</sup>提出一种融合拓扑势的层次化社区发现算法;Claire等<sup>[12]</sup>将网络图转换为持续同调中的条形码格式,对照各种效用度量来观察图之间关键特征的相关性。因此,使用持续同调分析网络拓扑结构以得到网络的传播变化,同样是一个挑战。

近年来,利用拓扑性质和单纯复形、持续同调性等方法对复杂网络,特别是加权网络的研究取得了很大进展。Carstens等<sup>[13]</sup>描述了作者合作网络的持续同调性,将拓扑学中的单纯复形方法应用于网络,通过对贝蒂数的定量比较解释了合作网络的持续同调性,并通过分析发现持续同调性在区分无标度网络和随机网络方面也起着重要作用;Dlotko等<sup>[14]</sup>提出一种简化的持续同调计算复形方法,从而加深了对拓扑单纯复形及其计算方法的理解;Cerri等<sup>[15]</sup>研究了多维持持续同调理论中贝蒂数的连续性质;Maletic等<sup>[16]</sup>采用拓扑单纯复形法构造网络,并对复杂网络的持续同调性进行分析与研究;Kahle<sup>[17]</sup>探讨随机网络的同调性与网络连通性之间的关系,为进一步理解网络拓扑性质与网络连通性之间的关系作了理论准备。

本文旨在采用一种不同的方法对社交网络进行概括,通过持续同调性表示在线社交网络图的传播特性,使用算法计算持续同调得到相应的贝蒂数和持续性图,并获取社交网络的拓扑特征,分析其传播特点。

## 1 持续同调相关理论

由于同胚对两个空间的拓扑性质相同,本文主要运用代数拓扑知识构造出该空间的同胚型。在研究一个社交网络时,常把社交网络中的各个用户看作节点,用户之间的关系看作边。换言之,可将社会网络中的节点视为点云,用持续同调方法构造一个由一系列简单复形组成的空间,并用构造的空间逼近社会网络空间。持续同调性为人们提供了一种在不减少维数情况下刻画数据全貌的方法。把数据放在原始高维空间中得到数据里的群集与环结构个数,这种计算拓扑特征的方法为网络图分析提供了一种新的效用度量。

### 1.1 图

在图论中,将图 $G$ 定义为一个有序二元组 $(V, E)$ 。其中, $V$ 称为顶集, $E$ 称为边集, $E$ 与 $V$ 不相交<sup>[18]</sup>。顶集的元素被称为顶点,表示对象;边集的元素被称为边,表示对象之间的关系,有不同类型的图表示顶点之间的不同关系。在无向图中,边对称地连接两个顶点;在有向图中,边不对称地连接两个顶点。如果顶点之间的关系有相互作用的强度,可用加权网络表示这种类型的关系。

有 $n$ 个顶点的图 $G$ 可用 $n \times n$ 个邻接矩阵表示,对于未加权图,若有 $G_{ij} = 1$ ,则表示顶点 $i$ 到顶点 $j$ 有一条边。如果顶点 $i$ 与顶点 $j$ 之间没有边,则 $G_{ij} = 0$ 。如果图随时间变化,则该图被称为动态图,图中的顶点和边会随着时间的变化不断删除或添加。在在线社交网络图中,用户被看作节点,用户之间的关系被看作边。

### 1.2 持续同调

#### 1.2.1 单纯复形与单纯同调

通常,单纯复形<sup>[19]</sup>是一种抽象代数结构,由点、边、三角形、四面体和高维多面体组成。0维单形是一个顶点,1

维单形是边, 2维单形是三角形, 3维单形是四面体(见图1), 直到第 $k$ 维, 其结构取决于单纯复形的维度。

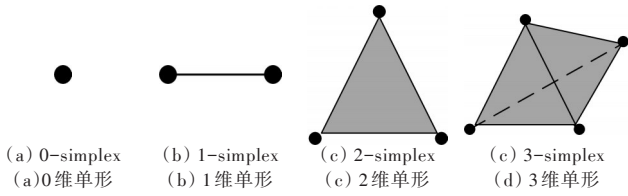


Fig. 1 Simplex  
图1 单形

**定义1** 设 $K$ 是 $n$ 维欧氏空间 $R$ 中单形的有限集合, 称 $K$ 为单纯复形, 简称复形(见图2), 则有: ①若为 $K$ 中的单形, 则其任意面都属于 $K$ ; ②若 $s$ 和 $s'$ 为 $K$ 中任意两个单形, 则 $s \cap s'$ 或者是空集, 或者是 $s$ 与 $s'$ 的一个公共面<sup>[20]</sup>。

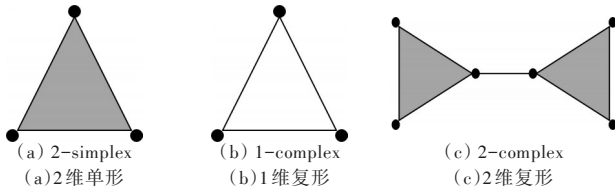


Fig. 2 Complex  
图2 复形

在单纯复形中, 把这些孔视为以不同维数的单纯形为界的空洞。在0维中, 其是连接分量; 在1维中, 其是以边(1维单纯形)为界的环; 在2维中, 其是以三角形为界的孔……在 $i$ 维中, 其是以 $i$ 维单纯形为界的孔。单纯同调是指寻找单纯复形中空洞的方法, 下面定义两种特殊类型的链, 即圈和边界<sup>[21]</sup>。

单纯形包含复形中最高维单形的所有面, 也即是说, 如果存在一个二维复形(具有最高维单形的单纯复形是一个二维单纯形, 即三角形), 则该复形还包含所有维数低于它的面(如边和顶点)。在同调理论中,  $i$ 链是一个具有整数系数的单纯复形 $K$ 所有 $i$ 维单形的和。 $K$ 的所有 $i$ 链集合都用 $C_i(K)$ 表示, 即 $C_i(K)$ 表示单纯复形上的 $i$ 维链群。对其作边缘同态映射:

$$\partial: C_i(K) \rightarrow C_{i-1}(K) \quad (1)$$

其中,  $C_{i-1}(K)$ 为 $K$ 的所有 $i$ 维链,  $C_i(K)$ 通过此映射后得到 $\partial$ 元的同态核是 $i$ 维链群 $C_i(K)$ 的子群, 称为维闭链群, 记作 $Z_i(K)$ 。带有方向向量 $[v_0, v_1, v_2, \dots, v_n]$ 的 $n$ 维单纯形 $K$ 的边界( $\partial(K)$ )为:

$$\partial(K) = \sum_{i=0}^n (-1)^i [v_0, v_1, v_2, \dots, \hat{v}_i, \dots, v_n] \quad (2)$$

其中, 第 $i$ 个向量被从序列中移除, 单个顶点的边界为0,  $\partial([v_i]) = 0$ , 且边界的边界为0。

$C_i(K)$ 中的所有边界, 即经过同态映射 $\partial: C_{i+1}(K) \rightarrow C_i(K)$ 得到的同态像是 $i$ 维链群 $C_i(K)$ 的子群, 也是 $i$ 维闭链群 $Z_i(K)$ 的子群(因为边缘都是封闭的), 称为 $i$ 维边缘闭链群, 简称 $i$ 维边缘边界群, 记作 $B_i(K)$ 。则定义单

纯复形 $K$ 的 $i$ 维同调群为:

$$H_i(K) = \frac{Z_i(K)}{B_i(K)} \quad (3)$$

第 $i$ 个连通数 $b_i$ 定义为 $H_i$ 的维度,  $b_i = \dim(H_i)$ 。

### 1.2.2 持续同调计算

对于有限的一组点 $X$ (如点云数据), 同调性不能给出感兴趣的信息。 $\beta_0$ 给出了连通分支数目, 但这只是点的数目, 其他所有贝蒂数都是零, 因为集合中没有其他维度的洞。因此, 可在数据上构建一个最大 $\varepsilon$ 值的单纯复形, 而不是构建多个不同 $\varepsilon$ 参数的单纯复形, 然后将其组合成一个序列。记录所有点对之间的距离, 并了解何值使得每对点形成边。所有隐藏在这些值中的单纯复形形成了一个连贯的滤流, 即为一系列嵌套的复形。因此, 本文不去处理点集, 而定义复形滤流为: 由不断增加的比例参数 $\varepsilon$ 生成的单纯复形序列。

在这样的结构中, 一些洞可能出现, 然后消失, 这些同源特征的持续性可被认为是数据集特征。在一个滤流中, 可以记录洞的出生, 即洞出现的时间, 以及洞的死亡, 即洞消失的时间。持续同源性的本质是对这些同源特征的出生和死亡进行持续性地追踪, 每个同源特征的寿命可表示为一个区间, 其中区间的起点与终点分别对应同源特征的出生和死亡。对于给定数据集, 可通过持久性条形码(PB)记录这些间隔。同样地, 持久性条形码可通过持久性图(PD)表示。下面定义贝蒂数<sup>[22]</sup>, 并构建 Vietoris-Rips (VR) 复形。

持续同调计算的理论基础是刻画网络拓扑结构, 通过网络拓扑结构变化分析网络社交的变化是有效、合理的。构建社交网络, 计算相应的网络拓扑同调群和贝蒂数相关参数, 进一步分析网络拓扑特性。这些同调空间可用空间维数来表征, 因此引入贝蒂数来量化拓扑不变量, 同时其也是拓扑空间的拓扑不变量。单纯复形 $K$ 的 $n$ 个贝蒂数表示为 $\beta_n$ :

$$\beta_n = \dim H_n = \text{rank} H_n = \text{rank} Z_n - \text{rank} B_n \quad (4)$$

$b_k(X) = \dim(H_k(X))$ , 即对于非负整数 $k$ , 定义空间 $X$ 的第 $k$ 个贝蒂数 $b_k(X)$ 为 $X$ 第 $k$ 个同调群 $H_k(X)$ 的秩(线性无关生成子数)。

添加或删除网络节点时, 相应同调群和贝蒂数都会发生变化。在网络拓扑图中, 零维贝蒂数表示网络上的连通分支数, 一维贝蒂数表示网络上的环(一维孔)数, 二维贝蒂数表示网络图中的空腔数。贝蒂数越大, 拓扑含义越复杂。

对于 Vietoris-Rips (VR) 复形, 其是一种抽象的代数结构, 能够表示欧氏空间 $E^n$ 中的一组点。假设 $G = (V, E)$ 是一个无向加权图, 权重值为 $W: V \times V \rightarrow R$ 。对于 $\forall \delta \in R$ , 1维骨架 $G_\delta = (V_\delta, E_\delta) \subset G$ 被定义为 $G$ 的子图, 其中 $V_\delta = V$ , 其边界集 $E_\delta = E$ 只包含权重小于或等于 $\delta$ 的边。然后, 对于 $\forall \delta \in R$ , 本文将 VR 复形定义为1维骨架 $G_\delta, Cl(G_\delta)$ 的团复



形,并将 Vietoris-Rips 过滤定义为:

$$\{Cl(G_\delta) \rightarrow Cl(G_{\delta'})\}_{0 \leq \delta \leq \delta'} \quad (5)$$

也即是说,在此过滤过程中,从顶点集开始,之后对边的权重从最小权重  $\omega_{\min}$  到最大权重  $\omega_{\max}$  进行排序,并让参数  $\delta$  从  $\omega_{\min}$  增加到  $\omega_{\max}$ 。在每一步中添加相应的边,得到阈值子图  $G_\delta$  的复形,因此产生了 Vietoris-Rips 过滤结构。

### 1.2.3 持续性图

随着 Rips 复形的构建,连续的间隔一起组成持续性图。持续性图是滤流的点图,当一个洞出现时,持续性图中的一个点会被标记出,该洞即具有包围开放空间的边缘结构。持续性图的横坐标代表开始时间,标志着一个洞的诞生;纵坐标代表洞消失的地方。该间隔通常被称为特征持续时间,持续性图可定性描述持续同调。

## 2 基于持续同调的社交网络传播特征

从根本上说,持续同调是用来检测空间的一种方法,该方法理想的,因为随着复形的增加,贝蒂数趋于稳定,即会生成稳定的持续性图。当距离  $\delta$  很小时,会有一些断开的组件。当  $\delta$  足够长时,该图成为一个连接图,断开的组件组合成一个。因此,所有持续性图都有一个无限持续下去的持续性点。对应在网络图中,当距离较小时用户紧密相连,从信息迅速传播到最后信息传播完成,形成一个稳定的圈。

运用代数拓扑中持续同调的知识对在线社交网络构造 VR 复形,将在线社交网络边数据集生成距离矩阵,再使用 perseus 软件计算得出贝蒂数和持续性图,从而分析在线社交网络的传播特征。具体方法如下:

### 2.1 贝蒂数持续性图计算算法

利用图描述社交网络时,节点表示社交网络中的用户,边表示用户之间的关系,有关系则为“1”,没有关系则为“0”。社交网络中关注的是节点间的连接关系,并不强调坐标位置。因此,本文首先利用 Floyd 算法计算得到距离矩阵,再将距离矩阵导入 Perseus 软件,最后通过 Perseus 软件计算得到社交网络图的贝蒂数和持续性图。

#### 2.1.1 数据处理

由于在线社交网络往往较为庞大,为方便进行持续同调相关计算,本文使用随机游走算法对数据进行处理,在有效保护网络的同时,创建一个可行的数据集。随机游走算法通常被视为一些随机过程的基本模型,并被视为马尔可夫过程<sup>[12]</sup>。随机游走算法的基本思想是从一个顶点或一系列顶点游走一个图,在任何顶点中,遍历者将以概率  $1 - a$  移动到此顶点的相邻顶点,并以概率  $a$  随机跳到图中任何顶点, $a$  称为跳跃发生概率。每次行走后都会得到一个概率分布,此概率分布描述了图中每个顶点被访问的概率。使用此概率分布作为下一次行走的输入,并迭代该过程。当满足一定条件时,概率分布趋于收敛,得到平稳的

概率分布。一维的随机游走可定义如下:每过一个单位时间,游走者从数轴位置  $x$  出发,以固定概率随机向左或向右移动一个单位。不妨将  $n$  时刻游走者的位置记为  $L_n$ ,则有:

$$L_n = x + X_1 + X_2 + \cdots + X_n \quad (6)$$

其中,  $X_1, X_2, \cdots, X_n$  为相互独立的随机变量,满足:

$$P(X_i = 1) = p = 1 - P(X_i = -1) \quad (7)$$

在本文中,伪随机数生成器从原始在线社交网络图中选择一个节点作为起点,接下来将相同节点作为第一个节点添加到新的图  $G'$  中。此后,程序查看连接到原始图中当前节点的边,并随机选择一个进行遍历,同时将相同的边添加到新的图  $G'$  中。整个过程经历了多次迭代,直到新构建的图  $G'$  完全包含一部分节点,本文取得的节点数为 100。

#### 2.1.2 距离矩阵形成

此时初始化一个邻接矩阵作为辅助,即在给定一组节点时,一个邻接矩阵表示其边关系。该矩阵表示在具有“1”和“0”的一对节点之间是否存在一条边,“1”表示两节点间存在一条边,“0”表示两节点间不存在边。然后使用 Floyd 算法,通过 Python 计算得出距离矩阵<sup>[23]</sup>。下面简单介绍 Floyd 算法以及距离矩阵。

Floyd 算法又称为插点法,是一种利用动态规划的思想寻找给定加权图中多源点之间最短路径的算法<sup>[23]</sup>。首先,从任意一条单边路径开始,所有两点之间的距离是边的权,这里设所有边的权值为 1。如果两点之间没有边相连,则权为无穷大。对于每一对顶点  $u$  和  $v$ ,看看是否存在一个顶点  $w$ ,使得从  $u$  到  $w$  再到  $v$  比已知路径更短,如果是则对其进行更新。用邻接矩阵  $G$  表示图,如果从  $v_i$  到  $v_j$  有路可达,则  $G_{ij} = d$ ,  $d$  表示该路的长度,否则  $G_{ij} = \infty$ 。定义一个矩阵  $D$  用来记录所插入点的信息,  $D_{ij}$  表示从  $v_i$  到  $v_j$  需要经过的点,初始化  $D_{ij} = j$ 。把各个顶点插入图中,比较插点后的距离与原来的距离,  $G_{ij} = \min(G_{ij}, G_{ik} + G_{kj})$ 。如果  $G_{ij}$  的值变小,则  $D_{ij} = k$ 。由此便可得到任意两点间的距离信息,一个距离矩阵是一个包含一组点两两之间距离的矩阵(即二维数组),所以可将距离信息转换成距离矩阵。

#### 2.1.3 持续性图生成

对数据进行处理后,首先运用 Floyd 算法对处理得到的 100 个点计算得出距离矩阵,然后利用持续同调计算软件计算得出各维数的持续性间隔以及各维贝蒂数,最后得到持续性图(见图 3)。

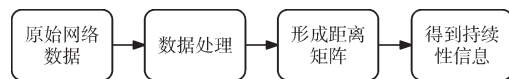


Fig. 3 Steps to generate a persistence diagram

图 3 持续性图生成步骤

### 2.2 小型社交网络持续性信息获取示例

已知两个小型社交网络,节点与节点之间的关系如图 4、图 5 所示。

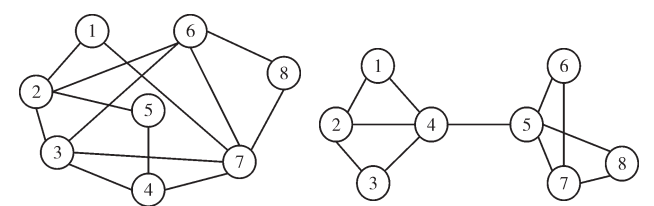


Fig. 4 Small network diagram a

Fig. 5 Small network diagram b

图 4 小型网络图 a

图 5 小型网络图 b

利用 Floyd 算法计算得到两个小型网络图的  $8 \times 8$  距离矩阵,分别如下:

$$\begin{pmatrix} 0 & 1 & 2 & 2 & 2 & 2 & 1 & 2 \\ 1 & 0 & 1 & 2 & 1 & 1 & 2 & 2 \\ 2 & 1 & 0 & 1 & 2 & 1 & 1 & 2 \\ 2 & 2 & 1 & 0 & 1 & 2 & 1 & 2 \\ 2 & 1 & 2 & 1 & 0 & 2 & 2 & 3 \\ 2 & 1 & 1 & 2 & 2 & 0 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 0 & 1 \\ 2 & 2 & 2 & 2 & 3 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 2 & 1 & 2 & 3 & 3 & 3 \\ 1 & 0 & 1 & 1 & 2 & 3 & 3 & 3 \\ 2 & 1 & 0 & 1 & 2 & 3 & 3 & 3 \\ 1 & 1 & 1 & 0 & 1 & 2 & 2 & 2 \\ 2 & 2 & 2 & 1 & 0 & 1 & 1 & 1 \\ 3 & 3 & 3 & 2 & 1 & 0 & 1 & 2 \\ 3 & 3 & 3 & 2 & 1 & 1 & 0 & 1 \\ 3 & 3 & 3 & 2 & 1 & 2 & 1 & 0 \end{pmatrix}$$

(8)

此时通过贝蒂数持续性图计算方法得出两个社交网络的贝蒂数分别如表 1、表 2 所示。

Table 1 Betti number of small network diagram a

表 1 小型网络图 a 的贝蒂数

子复形个数	零维贝蒂数	一维贝蒂数	二维贝蒂数	三维贝蒂数
0	8	0	0	0
1	1	0	0	0
2	1	0	0	35

Table 2 Betti number of small network diagram b

表 2 小型网络图 b 的贝蒂数

子复形个数	零维贝蒂数	一维贝蒂数	二维贝蒂数	三维贝蒂数
0	8	0	0	0
1	1	0	0	0
2	1	0	0	25
3	1	0	0	35

在网络拓扑图中,零维贝蒂数表示网络上的连通分支数,一维贝蒂数表示网络上的环(一维孔)数,二维贝蒂数表示网络图中的空腔数,贝蒂数越大,拓扑含义越复杂。

对于图 4,一开始它有 8 个连通分支(即一开始的 8 个节点),更高维的贝蒂数为零;形成第一个子复形时,零维贝蒂数为 1,即连通分支数为 1,一维贝蒂数为 2,即有 2 个圈,没有更高维的贝蒂数;形成第二个复形时,零维贝蒂数为 1,没有一、二维的贝蒂数,三维贝蒂数为 5,即更高一维的三维空间有 5 个空洞;形成第三个复形时,零维贝蒂数为 1,没有一、二维的贝蒂数,三维贝蒂数为 35,即更高一维的三维空间有 35 个空洞;没有再形成更多复形。

对于图 5,同样一开始它有 8 个连通分支,更高维的贝蒂数为零;形成第一个子复形时,零维贝蒂数为 1,即连通分支数为 1,没有更高维的贝蒂数;形成第二个复形时,零维贝蒂数为 1,没有一、二维的贝蒂数,三维贝蒂数为 25,即

更高一维的三维空间有 25 个空洞;形成第三个复形时,零维贝蒂数为 1,没有一、二维的贝蒂数,三维贝蒂数为 35,即更高一维的三维空间有 35 个空洞;没有再形成更多复形。

图 4 和图 5 是两个小型网络图,都拥有 8 个节点,从图中可以看出,图 4 相对于图 5 节点与节点之间的联系更紧密。所以图 4 各个节点之间的距离比图 5 更近,如果是节点之间相互传播,图 4 相对于图 5 也更为容易,而这一点在持续同调计算得出的贝蒂数中得到了验证。表 1 的贝蒂数总的来说少于表 2 的贝蒂数。

小型网络图的持续性图如图 6 所示。

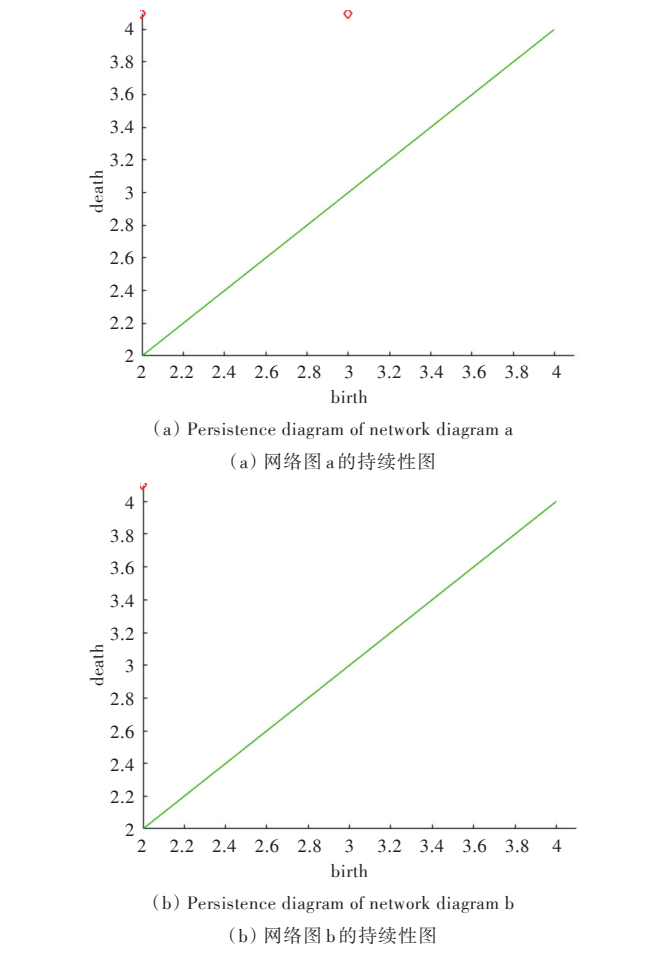


Fig. 6 Persistence diagram of small network diagram

图 6 小型网络图的持续性图

其中,横坐标表示持续性点的出生时间,纵坐标表示持续性点的死亡时间。

对于持续性图,具有无限持续性点的出生时间被绘制为红色菱形,即在网络图 a 的持续性图中有出生时间为 2 和 3 的两个持续性点,网络图 b 的持续性图中只有 1 个出生时间为 3 的持续性点。

通过示例可以看出:该方法充分捕捉了网络拓扑特征,可计算出网络拓扑信息,得到的拓扑信息与网络图的图像一致,并且可根据拓扑信息画出持续性图;节点间联系更紧密的网络图更容易传播,且其贝蒂数更大,传播更容易扩散;节点间联系更紧密的网络图的持续性点数量多于稀疏的网络图;贝蒂数在零维、一维和二维中表现出持

续同调性。

### 3 数值实验

选取 SNAP<sup>[24]</sup> 数据库中 Facebook、Deezer 和 Wiki-Vote 的 3 个社交网络图,每个图都提供了边列表(每个边列表数据由两个节点 ID 组成)。所有包含的数据集都满足社交网络标准,即一个简单、未加权的网络图。具体来说,Facebook 网络图代表匿名社交圈人与人之间的关系,Deezer 网络图表示在线音乐网站中在线社交网络的信息,而 Wiki-Vote 网络图是由世界各地的志愿者合作编写的免费百科全书。上述社交网络来自于大众广泛使用的社交媒体平台,可使本文能更好地模拟持续同调的实际应用。

对每个在线社交网络的研究,主要是分析网络所对应的持续性图。通过对每个使得持续性图稳定区间的分析与研究,可得到社交网络空间的同调信息。这些同调信息可帮助获取在线社交网络的空间特点,从而分析其传播变化。

#### 3.1 网络图比较

Facebook、Deezer 和 Wiki-Vote 网络数据集都包含了节点数据和边数据,通过边数据获取节点间的关联信息,若形成边则为“1”,若没有形成边则为“0”,由此生成网络图如图 7-图 9 所示。

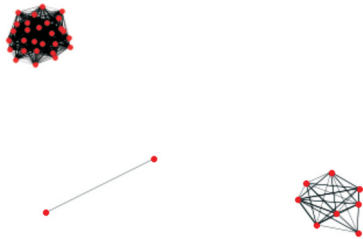


Fig. 7 Facebook network diagram  
图 7 Facebook 网络图

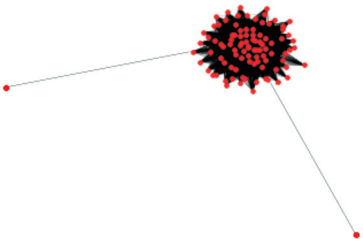


Fig. 8 Wiki-Vote network diagram  
图 8 Wiki-Vote 网络图

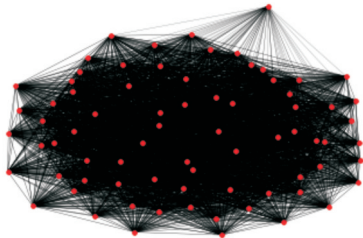


Fig. 9 Deezer network diagram  
图 9 Deezer 网络图

图 7-图 9 分别是 Facebook、Wiki-Vote 以及 Deezer 3 个网络在经过数据处理后 100 个节点的网络图,从图中可以看出,Facebook 网络图分成了 3 个团,且各个团中各节点相互之间联系紧密,具有很强的互联性,这也对应表现出 Facebook 各个用户之间的联系。Wiki-Vote 与 Deezer 两个社交网络的网络图则呈现出一个节点被其他多个节点连接,出现了类似于环状结构,这与 3 个网络的类型有一定联系。在现实世界中,Facebook 网络图代表匿名社交圈的人际关系,每个用户之间都存在相互联系的可能,所以图中所有节点之间始终具有很强的互连性;Wiki-Vote 网络图是用多种语言编写而成的网络百科全书,其基于维基技术的多语言百科全书式协作计划,所以其一个节点连接到多个端节点,形成环状;Deezer 网络图代表法国在线音乐网站,可提供乐队与歌曲搜索及音乐播放等功能,通常在图的中心创建一个环,而边界节点保持连接。

#### 3.2 贝蒂数比较

根据节点以及边关系,对 Facebook、Deezer 和 Wiki-Vote 3 个社交网络分别创建距离矩阵,通过持续同调计算,分别生成 3 个网络图的零维贝蒂数、一维贝蒂数及二维贝蒂数(见表 3-表 5),分析并比较其传播规律。

Table 3 Facebook Betti number

表 3 Facebook 贝蒂数

子复形个数	零维贝蒂数	一维贝蒂数	二维贝蒂数
0	100	0	0
1	1	10	1
2	1	0	0
...	...	...	...
9	1	0	0

Table 4 Wiki-Vote Betti number

表 4 Wiki-Vote 贝蒂数

子复形个数	零维贝蒂数	一维贝蒂数	二维贝蒂数
0	100	12	6
1	1	19	0
2	1	0	0
3	1	0	0
4	1	0	0

Table 5 Deezer Betti number

表 5 Deezer 贝蒂数

子复形个数	零维贝蒂数	一维贝蒂数	二维贝蒂数
0	100	0	0
1	1	10	3
2	1	0	0
3	1	0	0
4	1	0	0

从表 3 数据中可以看出,当没有形成子复形时,Facebook 网络对应的零维贝蒂数为 100,即连通分支数量为 100(本文选取 100 个节点),更高维的贝蒂数都为 0;形成第一个子复形时,零维贝蒂数为 1,即连通分支数为 1,一维贝蒂



数为10,即有10个圈,二维贝蒂数为1,即有1个洞,没有更高维的贝蒂数;形成第二个复形时,零维贝蒂数为1,没有一、二维的贝蒂数;之后随着子复形的增加,零维、一维、二维贝蒂数均没有发生变化;直到第9个子复形形成,之后没有再形成更多复形。

同理,从表4的数据中,一开始零维贝蒂数都为100,即连通分支数量为100,一维贝蒂数为12,二维贝蒂数为6;当第二个子复形形成时,零维贝蒂数变为1,而一、二维贝蒂数分别为19和0;当第三个子复形形成时,零维贝蒂数依然为1,而一、二维贝蒂数变为0;该情况一直持续到第四个子复形形成,之后没有再形成更多复形。

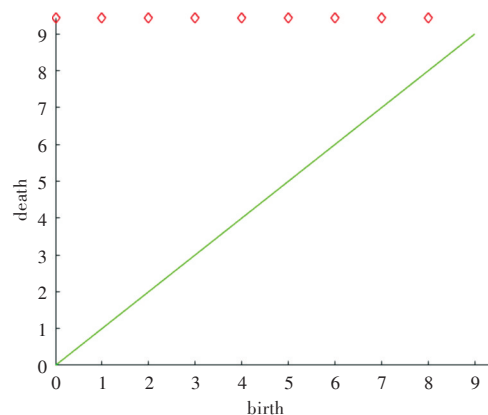
表5的数据情况与表3类似,一开始零维贝蒂数都为100,第二个子复形形成时,零维贝蒂数变为1,而一、二维贝蒂数分别为10和3;当第三个子复形形成时,零维贝蒂数依然为1,没有一、二维贝蒂数;直到第四个子复形形成,之后没有再形成更多复形。

从表中3组数据可以看出,当没有子复形形成时,3个网络图对应的零维贝蒂数都为100,即连通分支数量为100(本文选取100个节点)。Facebook网络和Deezer网络的一、二维贝蒂数都为0,而在Wiki-Vote网络中,一维贝蒂数为12,二维贝蒂数为6,这是由于Wiki-Vote网络是一个百科全书网络。从网络图中可以发现,它有一些圈和洞产生,即一、二维贝蒂数。随着子复形个数的增加,连通分支的个数逐渐稳定为1,而更高维的圈、洞个数逐渐变为0,此时意味着传播结束。从整个贝蒂数表格数据中可以看出,Facebook网络和Deezer网络在第二个复形形成时的一维贝蒂数都小于Wiki-Vote网络,说明Facebook网络和Deezer网络的传播更容易扩散,更有利于传播,而Wiki-Vote网络不容易激活很多节点,传播力相对弱一些,传播更为困难。因此,可利用贝蒂数来刻画在线社交网络的传播。

### 3.3 持续性图比较

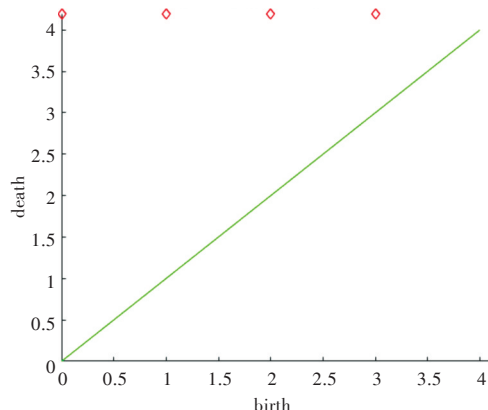
根据节点以及边关系,对3个社交网络分别创建距离矩阵,由此生成3个网络图的持续性图(见图10),分析比较其传播变化规律。

对于持续性图,具有无限持续性点的出生时间被绘制为红色菱形。从3个持续性图可以看出,Facebook网络的持续性图有9个红色菱形,即有9个无限持续性点,与上述贝蒂数中的子复形个数相对应;Wiki-Vote网络的持续性图有4个红色菱形,即有4个无限持续性点;Deezer网络的持续性图同样有4个红色菱形,即有4个无限持续性点。对于持续性点,图结构越复杂,其持续性点也会更多。从以上3个持续性图可以看出,Facebook网络的无限持续性点多于Wiki-Vote网络和Deezer网络,说明Facebook网络比Deezer网络和Wiki-Vote网络更利于传播,而Deezer网络和Wiki-Vote网络的传播力相对弱一些,传播更为困难。说明Wiki-Vote网络和Deezer网络相对于Facebook网络更稳定,这也很好地验证了上述小型网络图示例得出的结论。因此,持续性图也可用来刻画在线社交网络的传播。



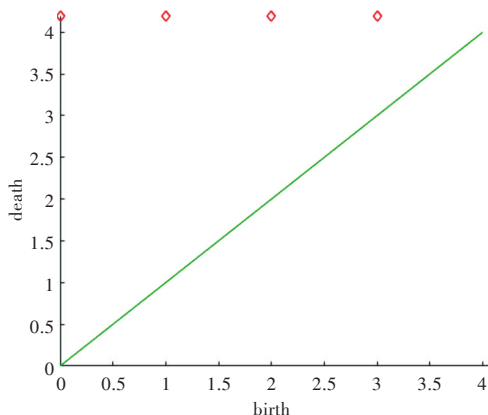
(a) Persistence diagram for Facebook

(a) Facebook 网络的持续性图



(b) Persistence diagram for Wiki-Vote

(b) Wiki-Vote 网络的持续性图



(c) Persistence diagram for Deezer

(c) Deezer 网络的持续性图

Fig. 10 Persistence diagram of online social networks

图10 在线社交网络的持续性图

## 4 结语

社交网络传播是在线社交网络的重要研究内容,本文主要运用持续同调方法构造一系列单纯复形来逼近社交网络空间,得到社交网络的同调信息,再将这些信息写入持续性图中,通过贝蒂数和持续性图的稳定信息分析在线社交网络传播。在一定程度下,尽管在线社交网络会发生变化(在线社交网络本就属于动态网络),对本文要得到的

同调信息也几乎没有影响。从本文得出的结论中,发现持续同调性可作为描述社交网络图传播的一种手段。根据已有数据显示,该方法充分捕捉了网络拓扑特征,为在线社交网络图描述传播提供了有效的新方法。在后续工作中将采用更大的数据集,并根据在线社交网络得到的持续同调对在线社交网络传播进行更深入研究。

#### 参考文献:

- [1] KEMPE D, KLEINBERG J. Maximizing the spread of influence through a social network[J]. *Theory of Computing*, 2015, 11: 105-147.
- [2] RICHARDSON M, DOMINGOS P. Mining knowledge-sharing sites for viral marketing[C]//*Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002: 61-70.
- [3] AGGARWAL C C. An introduction to social network data analytics [M]. Berlin: Springer, 2011.
- [4] KWAK H, LEE C, PARK H, et al. What is Twitter, a social network or a news media[C]//*Proceedings of the 19th International Conference on World Wide Web*, 2010: 591-600.
- [5] FANG B, JIA Y, HAN Y, et al. A survey of social network and information dissemination analysis[J]. *Chinese Science Bulletin*, 2014, 59 (32): 4163-4172.
- [6] GUILLE A, HACID H, FAVRE C, et al. Information diffusion in online social networks: a survey[J]. *SIGMOD Record: Special Interest Group on Management Data*, 2013, 42(2): 17-28.
- [7] ZINOVIEV D. A game theoretical approach to broadcast information diffusion in social networks[C]//*Boston: The 44th Annual Simulation Symposium*, 2011.
- [8] WATTS D J, STROGATZ S H. Collective relaxation dynamics of small-world networks[J]. *Physical Review E Statistical Nonlinear and Soft Matter Physics*, 2015, 91(5): 1-11.
- [9] BARABÁSI A-L, ALBERT R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509-512.
- [10] TIAN Y, ZHANG X G. Review of topological structure and information dissemination mechanism of online social network[J]. *Journal of Changsha University*, 2016(2): 73-75.  
田燕, 张新刚. 在线社交网络的拓扑结构与信息传播机制综述[J]. *长沙大学学报*, 2016(2): 73-75.
- [11] HOU M N, WANG Z X, HE J, et al. Hierarchical community discovery algorithm for social network on topology potential[J]. *Computer Engineering and Applications*, 2019, 55(1): 56-63.  
侯梦男, 王志晓, 何婧, 等. 融合拓扑势的社交网络层次化社区发现算法[J]. *计算机工程与应用*, 2019, 55(1): 56-63.
- [12] CLAIRE S L, TIANCHONG G, FENG L. Using persistent homology to represent online social network graphs[C]//*2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems*, 2017: 555-559.
- [13] CARSTENS C J, HORADAM K J. Persistent homology of collaboration networks[J]. *Mathematical Problems in Engineering*, 2013(6): 211-244.
- [14] DLOTKO P, WAGNER H. Simplification of complexes for persistent homology computations[J]. *Homology, Homotopy and Applications*, 2014, 16(1): 49-63.
- [15] CERRI A, FERRI M, FROSINI P, et al. Betti numbers in multidimensional persistent homology are stable functions[J]. *Mathematical Methods in the Applied Sciences*, 2013, 36(12): 1543-1557.
- [16] MALETIC S, HORAK D. Persistent homology of complex networks[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2009 (3): 4-21.
- [17] KAHLE M. Topology of random clique complexes[J]. *Discrete Mathematics*, 2009, 309: 1658-1671.
- [18] Happig. Basis of graph theory [EB/OL]. [https://blog.csdn.net/qq\\_44691917/article/details/104189417](https://blog.csdn.net/qq_44691917/article/details/104189417).  
Happig. 图论基础 [EB/OL]. [https://blog.csdn.net/qq\\_44691917/article/details/104189417](https://blog.csdn.net/qq_44691917/article/details/104189417).
- [19] MEHMET E A, ESRA A, AHMED E F. Persistence homology of networks: methods and applications [J]. *Applied Network Science*, 2019, 4(1): 1-28.
- [20] ZHAN H S, LI G X, ZHOU L H. Multiresolution representation of vector map based on simplicial complexes[J]. *Journal of Xidian University*, 2003, 30(5): 682-687.  
詹海生, 李广鑫, 周利华. 基于单纯复形的矢量地图多分辨率表示方法[J]. *西安电子科技大学学报*, 2003, 30(5): 682-687.
- [21] rxlfng. Simple complex and homology group [EB/OL]. <https://blog.csdn.net/rxlfng/article/details/111065611>.  
rxlfng. 单纯复形与同调群 [EB/OL]. <https://blog.csdn.net/rxlfng/article/details/111065611>.
- [22] EDELSBRUNNER H, LETSCHER D, ZOMORODIAN A. Topological persistence and simplification[J]. *Discrete and Computational Geometry*, 2002, 28(4): 511-533.
- [23] SHI W R, WANG K. Floyd algorithm for the shortest path planning of mobile robot[J]. *Chinese Journal of Scientific Instrument*, 2009, 30 (10): 2088-2092.  
石为人, 王楷. 基于 Floyd 算法的移动机器人最短路径规划研究[J]. *仪器仪表学报*, 2009, 30(10): 2088-2092.
- [24] JURE L, ANDREJ K. SNAP graph library: stanford large network dataset collection [EB/OL]. <http://snap.stanford.edu/data>.

(责任编辑: 黄健)