

# 基于高考录取成绩的院校竞争网络实证研究



王辰曦<sup>1</sup>, 张智成<sup>2</sup>, 袁 晨<sup>3</sup>, 蔡世民<sup>2\*</sup>

(1. 电子科技大学数学科学学院 成都 611731; 2. 电子科技大学大数据研究中心 成都 611731;  
3. 电子科技大学格拉斯哥学院 成都 611731)

**【摘要】**平行志愿录取规则与优质生源的紧缺使得院校在高考招生时存在激烈的竞争关系。从网络科学视角正确理解院校竞争关系,能够合理地指导考生进行志愿填报。利用 2019 年山西省高考理科录取成绩,通过改进的 Jaccard 相似度计算方法,构建院校竞争网络的邻接矩阵。基于复杂网络度量方法,实证研究院校竞争网络得到:1) 度分布存在一定的异质性,且具有较大的平均度;2) 簇度负相关表明院校竞争网络存在层次结构;3) 富人俱乐部连通性表明度数大于 375 的院校完全连通构成富人俱乐部;4) 利用基于节点标签与网络结构的社团划分算法,通过基尼不纯度评估社团内院校的同质性,进一步利用 Louvain 算法识别社团内院校存在层次化差异结构。这些实证结果刻画了院校竞争关系,在此基础上归纳了一些指导考生进行合理、分梯度志愿填报的建议和策略。

**关键词** 院校竞争网络; 社团检测; 复杂网络; 平行志愿

中图分类号 TP391 文献标志码 A doi:10.12178/1001-0548.2021207

## Empirical Research on College Competition Network Based on the Admission Scores of Colleges in Chinese Gaokao

WANG Chenxi<sup>1</sup>, ZHANG Zhicheng<sup>2</sup>, YUAN Chen<sup>3</sup>, and CAI Shimin<sup>2\*</sup>

(1. School of Mathematical Sciences, University of Electronic Science and Technology of China Chengdu 611731;  
2. Big Data Research Center, University of Electronic Science and Technology of China Chengdu 611731;  
3. Glasgow College, University of Electronic Science and Technology of China Chengdu 611731)

**Abstract** There is a very fierce competitive relationship among colleges because of the parallel application and the shortage of high-quality students in Chinese Gaokao. From the perspective of network science, correctly understanding such competitive relationship is helpful to guide candidates to apply for admission. This paper uses the 2019-year colleges' admission scores in Shanxi Province to construct the adjacency matrix of the competition network via the improved Jaccard similarity calculation method. Based on the measurements of complex network, the empirical research on the competition network shows that 1) there is the proper heterogeneity in the degree distribution and the larger average degree; 2) the negative correlation between the clustering coefficients and degrees of nodes suggests that the competitive network has a hierarchical structure; 3) the connectivity of rich club suggests that the colleges with a degree greater than 375 are fully connected and constitute the rich club; 4) using the community detection algorithm based on the colleges' metadata and network structure, the homogeneity of colleges in the community is evaluated through Gini Impurity. On this basis, The Louvain algorithm is further used to identify the hierarchical structure of colleges in the community. These empirical results clearly portray the competitive relationship among colleges, which enables us to bring great enlightenment for guiding candidates to apply for admission.

**Key words** college competition network; community detection; complex network; parallel application

高考招生录取制度采取平行志愿录取规则<sup>[1-2]</sup>, 优质生源的稀缺性加剧了不同院校对同一批次生源的争夺, 形成院校竞争关系(简称竞争关系)。同时, 考生在填报志愿时, 通常会依据“C9”“985”

“211”“双一流”“省部共建”等高校建设项目来区分院校的教育水平和实力差距, 使竞争关系存在层次性。然而, 在志愿填报策略分析中, 竞争关系的量化研究普遍未考虑这些因素<sup>[3-6]</sup>。因此, 采

收稿日期: 2021-05-25; 修回日期: 2021-10-11

作者简介: 王辰曦(1999-), 男, 主要从事复杂网络方面的研究。

\*通信作者: 蔡世民, E-mail: shimincai@uestc.edu.cn

用适当的定量分析方法研究竞争关系,有助于正确理解其内在结构特征与形成机制,对指导考生进行合理、分梯度地志愿填报起到重要启示作用<sup>[7]</sup>。

针对竞争关系这一问题,有学者分析了院校竞争力的影响因素,提出提升院校竞争力的方法<sup>[8-14]</sup>。如文献[11]基于波特五力模型提出的“竞争三力和影响二力”模型。文献[13]基于博弈论构建生源竞争的博弈模型,提出了正当提升竞争关系的方法;文献[14]基于元胞自动机进行建模仿真,得到了院校核心竞争力及附加条件对优质生源竞争力的影响。

网络科学是研究复杂网络系统规律的交叉科学<sup>[15-16]</sup>。从网络科学视角,院校及其竞争关系可以抽象地表示为院校竞争网络(简称竞争网络)。本文基于复杂网络度量方法,对竞争网络进行实证研究,定量分析拓扑结构特征,挖掘一般的网络统计特征与中尺度的社团结构特征。并利用 2019 年山西省高考录取成绩,基于改进的 Jaccard 相似度计算方法,计算院校之间录取成绩的相似性,定量表示竞争关系强度。

## 1 基于院校录取数据的竞争网络构建

### 1.1 院校录取数据描述

本文使用 2019 年山西省高考理科录取成绩,所选数据覆盖了山西省 1A、1B、2A 和 2B 批次的院校,总计 877 个数据项。1A、1B 对应本科第一批次,2A、2B 对应本科第二批次,两个批次的分数线分别为 507 分与 432 分。

每个数据项中除了院校名称外,还包含该院校在当前批次的(录取)最高分和最低分,以及院校标签,如“C9”“985”“211”“省部共建”“省

$$J(A, B) = \begin{cases} 0 & X_i \cap X_j = \emptyset \\ \frac{\min(s_{i\text{top}}, s_{j\text{top}}) - \max(s_{i\text{bottom}}, s_{j\text{bottom}}) + 1}{\max(s_{i\text{top}}, s_{j\text{top}}) - \min(s_{i\text{bottom}}, s_{j\text{bottom}}) + 1} & X_i \cap X_j \neq \emptyset \end{cases} \quad (3)$$

$J(A, B)$ 度量竞争关系强度,令 $J(A, B) = w_{ij} = w_{ji}$ 。通过遍历每对院校,利用改进 Jaccard 相似度度量方法,得到表示竞争关系的相似度矩阵 $W$ 。

### 1.3 竞争网络构建及度量方法

通过将竞争关系的相似度矩阵转换成邻接矩阵,实现竞争网络的构建。通过统计分析,81.6%的相似度矩阵元素是 0。图 1 给出非 0 相似度矩阵元素的概率分布,呈现中间高两端低的分布,与瑞利(Rayleigh)拟合的函数曲线相符,说明其偏离正态分布。

属”与“其他”。具体而言,“C9”包括 9 所院校及单独招生的医学院,共 13 个数据项。“985”共 34 个数据项。“211”共 82 个数据项。“省部共建”共 260 个数据项。“省属”共 485 个数据项。其余少部分、较特殊的院校将其标签设为“其他”,共 3 个数据项。值得注意的是,部分院校存在多重院校标签,本文按照院校标签表示的教育水平层次等级,一般优先考虑高等级。同时,部分省部共建或省属院校会分专业在不同批次进行招生,本文将其视作两个不同的数据项进行分析。

### 1.2 基于改进 Jaccard 相似度的竞争关系度量

竞争关系可以通过计算院校之间录取成绩的相似性度量。本文选择改进的 Jaccard 相似度实现相似性计算。Jaccard 相似度<sup>[17]</sup>用于比较有限样本集  $A$ 、 $B$  之间的相似性与差异性, Jaccard 系数值越大,样本相似性越高,即:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

本文根据院校录取数据的特殊性质改进 Jaccard 相似度。在每一项数据中,高校的最高分和最低分组成分数区间。两个院校分数区间,存在有交集和无交集两种情况。前者可按照交集长度与并集长度之比表示院校之间的相似性度量:

$$J(A, B) = \begin{cases} 0 & X_i \cap X_j = \emptyset \\ \frac{\text{len}(X_i \cap X_j)}{\text{len}(X_i \cup X_j)} & X_i \cap X_j \neq \emptyset \end{cases} \quad (2)$$

式中,  $\text{len}(\cdot)$ 表示集合的长度;  $X_i$ 表示院校 $i$ 的分数区间;  $\text{len}(X_i) = s_{i\text{top}} - s_{i\text{bottom}}$ ,其中 $s_{i\text{top}}$ 和 $s_{i\text{bottom}}$ 分别表示 $X_i$ 的最高分和最低分。因此,改进的 Jaccard 相似度可以表示为:

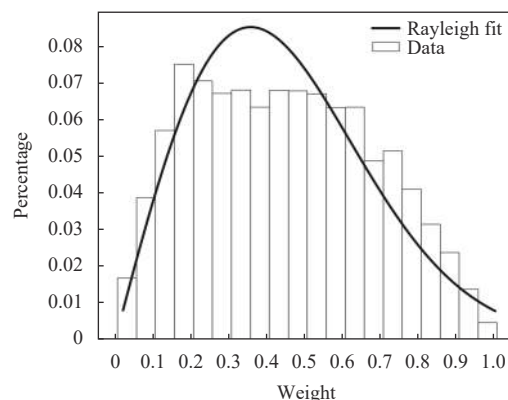


图 1 非 0 相似度矩阵元素的概率分布

由此可见, 竞争关系的相似度矩阵具有稀疏性。设定划分阈值为 0, 将相似度矩阵转换成表示竞争网络的邻接矩阵。邻接矩阵中的一个元素为 1 表示竞争网络中两个节点之间有连边。显然, 竞争网络是一个简化的无向、无权网络。表 1 给出了竞争网络的基本统计量。

表 1 竞争网络的基本统计量

T节点数 N/个	连边数 E/条	平均度 $\langle k \rangle$	最大度 $k_{\max}$	最小度 $k_{\min}$	平均聚类系数 $\langle C \rangle$	直径 D
877	71 677	163.46	465	1	0.79	7

如表 1 所示, 竞争网络的平均度相对较高, 主要原因是录取分数较低的院校 (简称底部院校) 数量较多, 而且它们的专业差异性较大。这些底部院校的分数区间较大, 使得它们之间的相似度较高, 连接较为紧密。如最大度节点是齐齐哈尔医学院, 度数达到了 465。该院校的特点是录取专业差异性较大, 应用心理学专业录取分数高达 496 分, 健康服务与管理专业录取分数却只有 430 分。它的分数区间几乎涵盖了本科第二批次招生院校的最大分数区间 (即是本科第一批次线与第二批次线的差值)。这些度数大的节点导致竞争网络的平均度较大。因此, 考生在填报志愿时, 在服从专业调剂条件下可以选择这些度数大的节点作为备选, 增加录取概率。同时, 院校之间的紧密连接使得竞争网络的平均聚类系数很高, 达到 0.79。这一结果表明院校之间倾向于构成社团, 即竞争网络可能存在显著的社团结构<sup>[18-19]</sup>。

除了上述复杂网络的基本统计量, 本文还从累积度分布、层次结构与富人俱乐部 (rich-club)<sup>[20-21]</sup> 3 个方面对竞争网络的统计特征进行实证研究。累积度分布  $P(k)$  表示度数不小于  $k$  的节点在整个网络中所占的比例,  $P(k) = \sum_{k' \geq k} P(k')$ 。它可以清晰刻画竞争网络的异质性特征。层次结构可以用簇度相关性表征。度为  $k_i$  的节点  $i$  的聚类系数  $C_i$  定义为  $C_i = 2E_i/k_i(k_i - 1)$ , 其中  $E_i$  是节点  $i$  的  $k_i$  个邻节点之间实际存在的边数, 即节点  $i$  的  $k_i$  个邻节点之间实际存在的邻居对的数目。 $k$  度节点的平均聚类系数 (簇度相关性) 为:

$$C(k) = \sum_{j=(i|k_i=k)} C_i/N_k \quad (4)$$

式中,  $N_k$  是度数为  $k$  的节点个数。在竞争网络中,

底部院校的部分节点拥有大量的边, 这些节点称为富节点 (rich nodes)。它们倾向于相互连接, 构成富人俱乐部。该统计特征可通过富人俱乐部连通性  $\Phi(r/N)$  刻画。 $\Phi(r/N)$  表示竞争网络中前  $r$  个度最大的节点之间, 实际存在的边数  $L$  与这  $r$  个节点之间总的可能存在的边数  $r(r-1)/2$  的比值<sup>[20-21]</sup>:

$$\Phi(r/N) = \frac{2L}{r(r-1)} \quad (5)$$

## 2 实证结果与讨论

### 2.1 竞争网络的统计特征

竞争网络的累积度分布如图 2 所示, 其中, 虚线指示节点度数为 100, 内嵌子图是基于一分一段表的累计人数变化, 其中的虚线指示本科第一批次分数线, 右边表示高分段, 左边表示低分段。竞争网络中度数超过 100 的节点数量开始变多, 度数小于 100 的节点数量很少。实证结果表明竞争网络具有一定的异质性特征, 反映底部院校的高聚集性及它们之间激烈的竞争关系。竞争网络的异质性特征主要由不同高考成绩的考生数量分布异质性决定。如图 2 中内嵌子图所示, 基于一分一段表<sup>[22]</sup> 的累计人数变化表明, 最高分与本科第一批次之间 (简称高分段, 虚线右边部分) 的考生数量 (约 3.2 万) 要远少于本科第一批次与本科第二批次之间 (简称低分段, 虚线左边部分) 的考生数量 (约 5.4 万), 且其分布也具有明显的异质性特征。

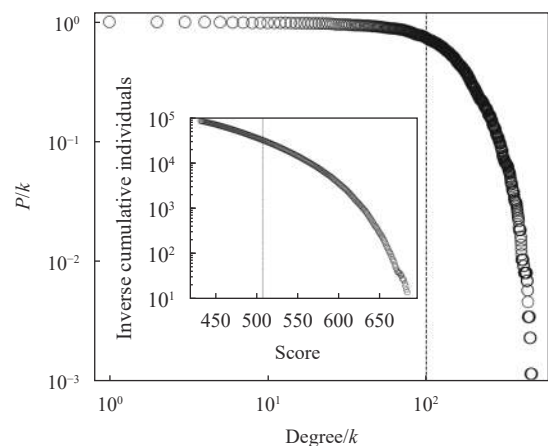


图 2 竞争网络的累积度分布

累积度分布的异质性表明大量低分段的考生在填报志愿时选择底部院校, 在院校招生人数限定的条件下, 他们考虑专业差异性, 使得对院校具有偏好选择。他们的偏好选择容易加大底部院校的分数区间, 形成较强的连接紧密性 (即形成高的相似

度), 加剧它们之间的竞争关系, 引起院校最低录取分的剧烈波动。同时, 少量高分段的考生填报志愿时, 基于一分一段表排名选择顶部院校(如 C9、985、211 院校), 使得它们的分数区间较小, 形成较弱的连接紧密性。由于底部院校最低录取分波动幅度大, 导致低分段考生更容易由于目标院校的最低录取分大幅上升而掉档, 同时也更有机会被最低录取分大幅下降的院校录取, 所以建议低分段考生增加填报少数往年录取分高于自身分数的院校。

竞争网络的簇度相关性如图 3 所示。竞争网络中节点度数越大, 聚类系数呈现逐渐减小的趋势。簇度负相关性表明竞争网络存在层次结构。相对而言, 大部分度数小的节点(如度数小于 100) 能够更好地聚类, 一般是录取成绩较高的优秀院校, 如 985、211 院校等; 度数较大的节点(如度数大于 400) 之间连接并不紧密, 往往聚类效果相对偏弱。然而, 竞争网络的高聚类性与层次结构, 表明院校倾向于构成层次社团结构, 本文将着重对竞争网络的社团结构进行分析。

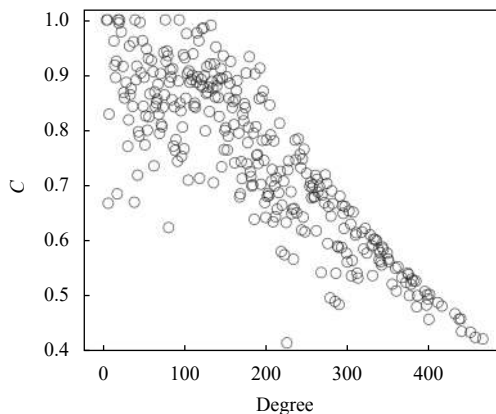


图 3 竞争网络的簇度相关性

竞争网络的富人俱乐部连通性如图 4 所示。具体而言, 当  $r/N = 7.3\%$  时,  $\Phi(r/N)$  非常接近于 1, 表明竞争网络存在富人俱乐部特征。特别是, 当  $r/N \leq 2.8\%$  时,  $\Phi(r/N) = 1$  表示度数在前 2.8% 的富节点形成的子图可以构成一个完全联通图。这个度数阈值为 375, 即度数大于 375 的节点均为竞争网络的富节点。表 2 列出竞争网络中富节点对应的院校, 其平均度 406.8。值得注意的是, 富节点之间相互连接紧密, 但是其连接的其余节点之间没有连边(即是富节点与其余节点没有形成闭环), 使得富节点的聚类系数相对不高。它们聚类系数波动在 0.4~0.5 之间。

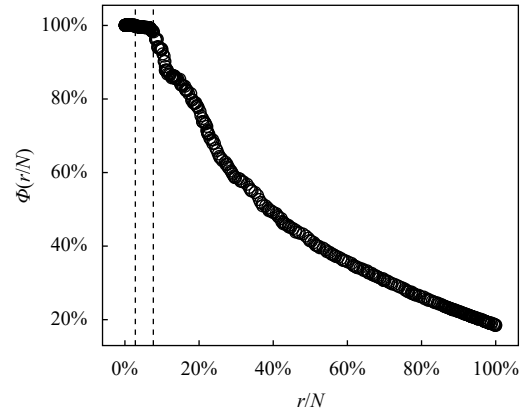


图 4 竞争网络的富人俱乐部连通性

表 2 全联通的富节点对应的院校

度	高校	度	高校	度	高校
375	湖北科技学院	394	沈阳医学院	415	湖北经济学院
377	沈阳工程学院	394	海南医学院	431	内蒙古工业大学
378	哈尔滨医科大学	396	山东第一医科大学	436	新疆医科大学
380	金陵科技学院	397	湖南农业大学	438	山东建筑大学
383	嘉兴学院	398	鲁东大学	439	福建中医药大学
384	浙江中医药大学	399	天津中医药大学	450	潍坊医学院
384	丽水学院	400	湖北医药学院	456	内蒙古医科大学
385	黑龙江中医药大学	410	广东海洋大学	465	齐齐哈尔医学院

由表 2 可知, 富节点中超过 50% 是医科类院校。它们的分数区间较大, 区间长度均超过 40, 且均处于 440~500 分数区间(低分段)附近。由于属于低分段, 这些医科类院校在招生时, 它们的专业相似度高、专业之间分差跨度大。这一特征导致报考其中某一院校的学生分数差距较大, 且这些同类型院校的分数区间重叠较大, 使得它们的竞争关系变得更加剧烈, 容易演变成富节点, 形成竞争网络的富人俱乐部特征。因此, 考生填报平行志愿时, 应该适当回避这类院校, 避免在平行志愿中重叠填报。

## 2.2 竞争网络的社团结构

竞争网络的中尺度如社团结构分析, 能够帮助考生深入地理解在不同层级院校招生时的竞争关系。本文基于节点元数据(如院校标签), 利用文献 [23] 提出的网络社团结构推断方法, 对竞争网络进行社区划分。假定一个具有  $N$  个节点与  $M$  个社团的竞争网络, 节点的标记  $u \in \{1, 2, \dots, N\}$ , 社团的标记  $s_u \in \{1, 2, \dots, M\}$ 。院校标签类别数量是  $K = 6$  的离散值, 分别表示“C9”“985”“211”“省部共建”“省属”与“其他”, 使得节点  $u$  的元数据标记  $\chi_u \in \{1, 2, \dots, K\}$ 。定义具有元数据  $\chi_i$  的节点  $i$  划分到社团  $s_j$  的概率为  $\gamma_{sx}$ , 社团之间存在连边的概率

为 $\theta_{st} = \theta_{ts}$ 。因此, 竞争网络的概率生成模型可以表示为<sup>[23]</sup>:

$$P(\mathbf{A}|\boldsymbol{\theta}, \boldsymbol{\Gamma}, \mathbf{X}) = \sum_{\mathbf{S}} P(\mathbf{A}|\boldsymbol{\theta}, \mathbf{S})P(\mathbf{S}|\boldsymbol{\Gamma}, \mathbf{X}) \quad (6)$$

式中,  $\mathbf{A}$ 表示一个 $N \times N$ 对称矩阵, 表征竞争网络的邻接矩阵;  $\mathbf{X}$ 表示一个 $N \times K$ 矩阵, 表征节点标记的元数据;  $\boldsymbol{\theta}$ 表示一个 $M \times M$ 对称矩阵, 包含参数 $\theta_{st}$ ;  $\boldsymbol{\Gamma}$ 表示一个 $M \times K$ 矩阵, 包含参数 $\gamma_{sx}$ ;  $\mathbf{S}$ 是一个 $N \times M$ 矩阵, 表征节点被分配的社团结构。将式(6)进行对数化, 应用 Jensen 不等式, 将其表达式变换为:

$$\begin{aligned} \log P(\mathbf{A}|\boldsymbol{\theta}, \boldsymbol{\Gamma}, \mathbf{X}) &\geq \sum_{\mathbf{S}} q(\mathbf{S}) \log \frac{P(\mathbf{A}|\boldsymbol{\theta}, \mathbf{S})P(\mathbf{S}|\boldsymbol{\Gamma}, \mathbf{X})}{q(\mathbf{S})} = \\ &\sum_{\mathbf{S}} q(\mathbf{S}) \log P(\mathbf{A}|\boldsymbol{\theta}, \mathbf{S}) + \sum_{\mathbf{S}} q(\mathbf{S}) P(\mathbf{S}|\boldsymbol{\Gamma}, \mathbf{X}) - \\ &\sum_{\mathbf{S}} q(\mathbf{S}) \log q(\mathbf{S}) \end{aligned} \quad (7)$$

式中,  $q(\mathbf{S})$ 表示任意节点分配到不同社团的概率分布, 满足 $\sum_{\mathbf{S}} q(\mathbf{S}) = 1$ 。在满足 $q(\mathbf{S}) =$

$\frac{P(\mathbf{A}|\boldsymbol{\theta}, \mathbf{S})P(\mathbf{S}|\boldsymbol{\Gamma}, \mathbf{X})}{\sum_{\mathbf{S}} P(\mathbf{A}|\boldsymbol{\theta}, \mathbf{S})P(\mathbf{S}|\boldsymbol{\Gamma}, \mathbf{X})}$ 时, 式(7)式右端取得最

大值<sup>[24]</sup>。因此, 在给定连接矩阵 $\mathbf{A}$ , 节点元数据 $\mathbf{X}$ 与划分的社团数量 $M$ , 本文可以利用期望最大化算法, 结合式(7)进行参数估计及社团结构划分: 1) 对 $\boldsymbol{\theta}$ 、 $\boldsymbol{\Gamma}$ 进行随机初始化, 计算 $q(\mathbf{S})$ ; 2) 固定 $q(\mathbf{S})$ , 最大化不等式(7); 3) 重复步骤1)和2), 直到 $\boldsymbol{\theta}$ 、 $\boldsymbol{\Gamma}$ 收敛稳定时,  $q(\mathbf{S}) =$

$\frac{P(\mathbf{A}|\boldsymbol{\theta}, \mathbf{S})P(\mathbf{S}|\boldsymbol{\Gamma}, \mathbf{X})}{\sum_{\mathbf{S}} P(\mathbf{A}|\boldsymbol{\theta}, \mathbf{S})P(\mathbf{S}|\boldsymbol{\Gamma}, \mathbf{X})} = \frac{P(\mathbf{A}, \mathbf{S}|\boldsymbol{\theta}, \boldsymbol{\Gamma}, \mathbf{X})}{P(\mathbf{A}|\boldsymbol{\theta}, \boldsymbol{\Gamma}, \mathbf{X})} = P(\mathbf{S}|\mathbf{A}, \boldsymbol{\theta}, \boldsymbol{\Gamma},$

$\mathbf{X})$ , 表示 $q(\mathbf{S})$ 等价于竞争网络的社团结构最优化划分的后验概率 $P(\mathbf{S}|\mathbf{A}, \boldsymbol{\theta}, \boldsymbol{\Gamma}, \mathbf{X})$ ; 4) 通过得到已收敛的 $q(\mathbf{S})$ , 获取任意节点分配到不同社团的概率, 选取最大概率对应的社团作为该节点的所属社团, 在此基础上得到竞争网络的(非重叠)社团结构划分。

同时, 社团内节点元数据并不完全一致, 导致社团内院校可能存在一定的层次化差异。本文采用基尼不纯度(Gini impurity, GI)来衡量社团内院校的层次化差异<sup>[25]</sup>。假设 $p(\chi)$ 表示标记元数据 $\chi$ 的节点在社团 $s$ 中出现的概率(或归一化频次), 则该社团的基尼不纯度定义为:

$$GI(s) = \sum_{\chi \in \{1, 2, \dots, K\}} p(\chi)(1 - p(\chi)) \quad (8)$$

GI取值越大, 表明社团内院校具有不同标签的混乱程度越高, 存在较为明显的层次化差异。通过预定划分竞争网络的社团数量 $M = 6$ (与标签类别数量一致), 利用上述算法得到竞争网络的社团结构, 如每个社团的GI值、规模、代表院校等, 如表3所示。

表3 基于节点元数据和网络结构的竞争网络社团划分

社团编号	GI值	社团规模	院校标签类别分布“C9”“985”			
			“211”	“省部共建”	“省属”	“其他”
1	0	2	2	0	0	0
2	0.198	9	8	1	0	0
3	0.711	366	3	33	79	139
4	0.404	251	0	0	69	181
5	0.421	116	0	0	35	81
6	0.213	190	0	0	23	167

由表3可知, 社团1、2、6具有较小的GI值, 表明构成社团的大部分院校具有相近的教育水平。如图5所示, 社团1由清华大学与北京大学构成, 社团2由部分C9院校(含医学院)与985院校中国人民大学构成。这些社团内院校具有相近的教育水平(即同质性), 它们在招生时存在较为激烈的竞争关系。而且, 由社团2可知, 同属于C9院校的哈尔滨工业大学与西安交通大学, 由于地理位置影响, 它们与其他C9院校的竞争关系相对较弱。社团6由大部分地方性的省属院校构成, 它们教育水平相对较弱。这些院校的分数区间靠近本科第二批次线, 具有较强的同质性导致激烈的竞争关系。

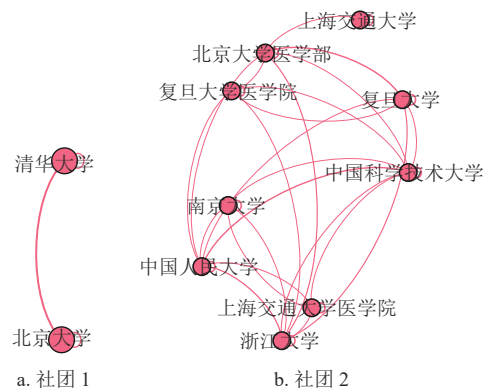


图5 竞争网络社团划分

同时, 社团4、5具有适中的GI值。它们主要由省部共建与省属院校混杂构成, 具有适当的教育

水平, 一般在本省范围内具有一定的知名度。特别值得注意的是, 社团 3 具有最高的 GI 值, 且社团规模最大、混杂度最高。它包括部分 985 院校、211 院校、省部共建、具有优势专业 (特指在本科第一批招生专业) 的省属院校, 以及其他特殊院校。虽然这些院校存在一定的同质性, 但是分数区间跨度较大, 其内部依然存在较为明显的层次化结构。Louvain 算法是基于模块度的经典社团划分算法, 能够快速划分网络社团<sup>[26]</sup>。同时, 考虑到再次使用网络社团结构推断方法可能无法细化社团 3 的子结构, 本文从模块度视角通过 Louvain 算法进一步分析该社团的层次化结构。社团 3 划分的子结构 GI 值、规模, 如表 4 所示。

表 4 基于 Louvain 算法的社团 3 划分

子结构编号	GI 值	子结构规模	院校标签类别分布			
			“C9”	“985”	“211”	“省部共建” “省属” “其他”
1	0.61	69	3	30	30	5, 1, 0
2	0.64	72	0	3	35	23, 10, 1
3	0.62	82	0	0	13	33, 36, 0
4	0.58	48	0	0	3	22, 22, 1
5	0.52	95	0	0	2	51, 41, 1

由表 4 可知, 所有子结构都具有相对适中的 GI 值, 每一个子结构包含两个大类院校, 且院校教育水平逐次降低以体现一定的层次结构。如子结构 1 主要由有教育水平较好的 985 与 211 院校构成 (如图 6 所示), 子结构 2 主要是排名适中的 211 院校与具有特色学科的省部共建院校构成。因此, 本文通过子结构中院校标签类别分布, 能够清晰地、科学地得到社团 3 内同质化院校的不同层次分布, 从而对合理、分梯度地志愿填报进行有效指导。

基于上述社团结构分析, 同一社团中同质化院校的竞争关系较为剧烈。这样会导致院校在招生政策上采取一些有利于考生的政策, 以此来吸引考生填报。但同样, 在对同质化院校进行填报时, 应充分考虑到社团内院校仍然存在分数区间的层次性。因此, 考生在进行志愿填报时, 应当依据社团结构划分的结果, 在异质化的社团之间进行初步的筛选, 并基于社团内部的层次结构, 利用高度同质化院校的竞争关系进行合理、分梯度地填报, 从而进入更优质的院校。

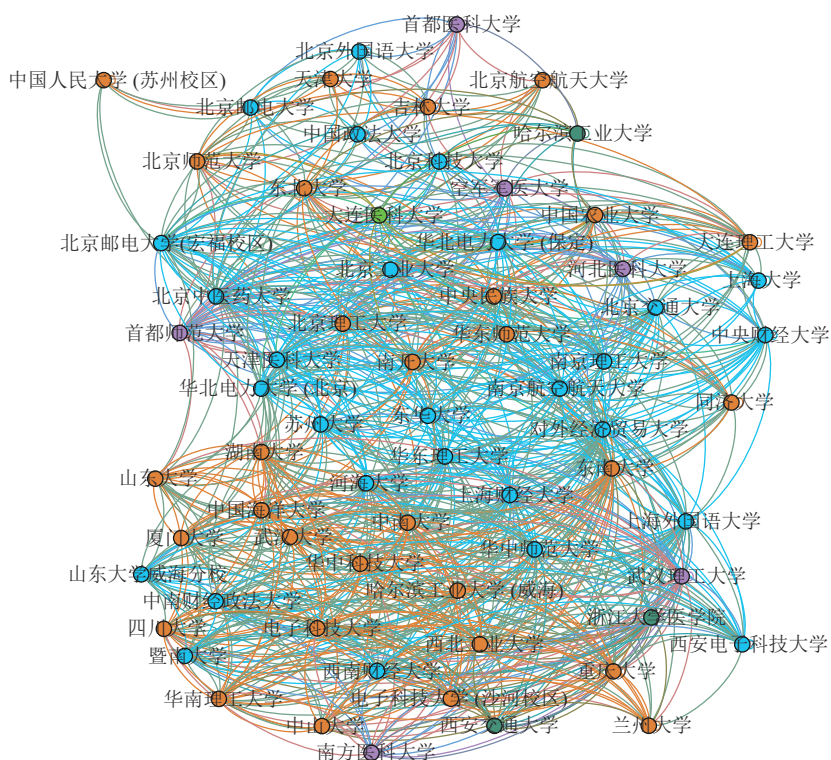


图 6 社团 3 划分的子结构 1 网络

### 3 结束语

本文应用复杂网络理论实证研究高考志愿填报过程中院校之间的竞争关系, 分析其存在的异质度

分布特征、层次结构特征、富人俱乐部以及社团结构特征。依据竞争网络结构 (院校竞争关系) 与节点 (院校) 的标签信息进行社团划分, 其结果表明

同一社团内院校在招生时存在较激烈的竞争关系,这种潜在的竞争关系将会影响院校的招生政策。进一步,本文通过 Louvain 算法发掘社团内竞争关系的层次结构特征,衡量社团内院校之间的竞争关系层次差异性,细化院校之间的竞争关系。

这些实证结果有助于从网络科学角度理解院校的竞争关系。在此基础上,本文归纳了一些志愿填报策略和建议:

1) 考生按照高考成绩,对应到社团划分得到的院校组群(即社团或社团子结构)。按照院校组群的层次化差异,可以合理、分梯度地填报平行志愿,且每个梯度对应某一个层级的院校组群。如社团3分解成5个层级的子结构,它们具有明显的梯度,考生可以适当选择2~3个梯度的院校组群,分别填报合适的目标院校。

2) 填报平行志愿时,要适当减少分数区间波动较大的院校,降低平行志愿滑档的风险。如医科类院校的分数区间每年波动较大,填报此类院校的考生也较容易滑档。

3) 如果考生的高考成绩处于本科第二批次的分数线附近,需要平衡目标院校的类型、地理位置等多种因素,结合“冲稳保”策略合理地填报平行志愿。

最后,考虑到不同院校专业录取规则与设置专业的差异性,本文不直接考虑(不同学校)同专业之间的竞争关系。但是,本文对院校竞争关系是基于院校分数区间进行度量的,分数区间的下限表示院校的录取成绩,分数区间的上限表示院校最优专业的录取成绩,分数区间一定程度上反映出同学校不同专业之间的竞争关系。因此,仅有上述最后一项建议中的“冲稳保”策略可适用于填报同院校专业,且建议勾选“服从专业调剂”。

### 参 考 文 献

- [1] WEN J. The historical evolution of enrollment policy for college entrance examination-based on the perspective of historical institutionalism[J]. *Journal of Educational Research and Policies*, 2020, 2(11): 23-28.
- [2] HUANG J. The educational purpose of the Chinese national college entrance examination[J]. *International Journal of New Developments in Education*, 2020, 2(6): 59-64.
- [3] LI Y, ZHANG X. Research on the volunteer filling in college entrance examination driven by information technology[J]. *Journal of Educational Research and Policies*, 2020, 2(6): 88-92.
- [4] WANG J, CHEN Y, HAN D, et al. New college entrance examination selection recommendation system based on fuzzy evaluation matrix[J]. *Computer Science and Application*, 2020, 10(2): 361-370.
- [5] LI J. Prediction model of college entrance examination score line[J]. *World Scientific Research Journal*, 2019, 5(12): 203-209.
- [6] WANG Y, LIU Y, LIU L. Research on the impact of intelligent decision-making system for voluntary filling in college entrance examination on senior high school students[C]//*Proceedings of 2019 3rd International Conference on Education, Management Science and Economics*. [S.l.]: Springer, 2019: 180-183.
- [7] GUO Z. College entrance examination volunteer filing research on related issues[C]//*Proceedings of 2019 International Conference on Education, Management, Economics and Humanities*. [S.l.]: DEStech, 2019: 28-32.
- [8] 谷进军,曹东方,蒋玉国. 高校招生生源竞争的策略[J]. *教育理论与实践*, 2005, 25(1): 9-10.  
GU J J, CAO D F, JIANG Y G. Strategy of student source competition of university recruiting students[J]. *Theory and Practice of Education*, 2005, 25(1): 9-10.
- [9] 于猛,杨爱民,尹星. 高校本科生源竞争力的提升策略[J]. *中国石油大学学报(社会科学版)*, 2017, 33(2): 98-102.  
YU M, YANG A M, YIN X. The strategies of improving competitiveness of university enrollment[J]. *Journal of China University of Petroleum (Edition of Social Sciences)*, 2017, 33(2): 98-102.
- [10] 黄风. 生源竞争与高校招生策略研究——以闽江学院为例[J]. *郑州铁路职业技术学院学报*, 2020, 32(4): 99-101.  
HUANG F. Study on student source competition and college enrollment strategy—Taking Minjiang college as an example[J]. *Journal of Zhengzhou Railway Vocational and Technical College*, 2020, 32(4): 99-101.
- [11] 曹伟,李笑晨. 波特五力模型应用于高职院校招生竞争力分析的可行性研究[J]. *高教论坛*, 2020(10): 68-70.  
CAO W, LI X C. Study on the feasibility of Porter's five forces model applied to the analysis of enrollment competitiveness of higher vocational colleges[J]. *Higher Education Forum*, 2020(10): 68-70.
- [12] 阎朝阳. 高校生源竞争与招生宣传策略[J]. *华中农业大学学报(社会科学版)*, 2002, 43(1): 92-95.  
YAN C Y. Student source competition and recruiting strategy in college enrollment[J]. *Journal of Huazhong Agricultural University (Social Sciences Edition)*, 2002, 43(1): 92-95.
- [13] 孟杰. 新高考背景下高校生源竞争的博弈分析[J]. *高教学刊*, 2020, 2020(11): 65-69.  
MENG J. Game model of college student source competition under the reform of college entrance examination[J]. *Journal of Higher Education*, 2020, 2020(11): 65-69.
- [14] 谢菲,侍旭. 基于元胞自动机的高校招生竞争演化研究[J]. *南京航空航天大学学报(社会科学版)*, 2015, 17(1): 29-34.  
XIE F, SHI X. Research on competitive college enrollment based on cellular automata[J]. *Journal of Nanjing University of Aeronautics and Astronautics (Social Science)*, 2015, 17(1): 29-34.

- [15] 周涛, 柏文洁, 汪秉宏, 等. 复杂网络研究综述[J]. *物理*, 2005, 34(1): 31-36.  
ZHOU T, BAI W J, WANG B H, et al. A brief review of complex networks[J]. *Physics*, 2005, 34(1): 31-36.
- [16] 周涛, 张子柯, 陈关荣, 等. 复杂网络研究的机遇与挑战[J]. *电子科技大学学报*, 2014, 43(1): 1-5.  
ZHOU T, ZHANG Z K, CHEN G R, et al. The opportunities and challenges of complex networks research[J]. *Journal of University of Electronic Science and Technology of China*, 2014, 43(1): 1-5.
- [17] JACCARD P. The distribution of the flora in the alpine zone[J]. *New Phytologist*, 1912, 11(2): 37-50.
- [18] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393(6684): 440-442.
- [19] HOLLAND P W, LEINHARDT S. Transitivity in structural models of small groups[J]. *Comparative Group Studies*, 1971, 2(2): 107-124.
- [20] ZHOU S, MONDRAGÓN R J. The rich-club phenomenon in the Internet topology[J]. *IEEE Communications Letters*, 2004, 8(3): 180-182.
- [21] GASPARINI M, IZQUIERDO J L C, CLARISÓ R, et al. Analyzing rich-club behavior in open source projects[C]//Proceedings of the 15th International Symposium on Open Collaboration. [S.l.]: ACM, 2019: 1-9.
- [22] 山西省招生考试管理中心. 2019年山西省普通高考成绩分段统计表 [EB/OL]. (2019-06-24). <http://www.sxkszx.cn/news/2019624/n739951587.html>.  
Shanxi Enrollment and Examination Management Center. Statistical table of general college entrance examination scores in 2019 in Shanxi province[EB/OL]. (2019-06-24). <http://www.sxkszx.cn/news/2019624/n739951587.html>.
- [23] NEWMAN M E J, CLAUSET A. Structure and inference in annotated networks[J]. *Nature Communications*, 2016, 7(1): 1-11.
- [24] BREIMAN L, FRIEDMAN J, STONE C J, et al. Classification and regression trees[M]. [S.l.]: CRC Press, 1984.
- [25] LERMAN R I, YITZHAKI S. A note on the calculation and interpretation of the Gini index[J]. *Economics Letters*, 1984, 15(3-4): 363-368.
- [26] BLONDEL V, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.

编辑 蒋晓

(上接第383页)

- [15] 孙铭才, 张秦, 陈光陆. 动态时间窗下的相控阵雷达自适应调度算法[J]. *雷达学报*, 2018, 7(3): 303-312.  
SUN M C, ZHANG Q, CHEN G L. Dynamic time window adaptive scheduling algorithm for the phased array radar[J]. *Journal of Radars*, 2018, 7(3): 303-312.
- [16] 李纪三, 侯娇, 班阳阳, et al. 旋转相控阵雷达资源调度策略研究[J]. *雷达与对抗*, 2018, 38(4): 5-10.  
LI J S, HOU J, BAN Y Y, et al. Study on resource scheduling strategy for rotating phased array radar[J]. *RADAR & ECM*, 2018, 38(4): 5-10.
- [17] 丁鹭飞, 耿富录, 陈建春. 雷达原理[M]. 第4版. 北京: 电子工业出版社, 2009.  
DING L F, GENG F L, CHEN J C. Radar principles[M]. 4rd ed. Beijing: Electronic Industry Press, 2009.
- [18] 刘一鸣, 盛文, 施端阳. 效费比视角下相控阵雷达多波束跟踪调度策略[J]. *西安电子科技大学学报*, 2019, 46(6): 155-162.  
LIU Y M, SHENG W, SHI D Y. Multi-beam tracking scheduling strategy for phased array radar based on the cost-effectiveness ratio[J]. *Journal of Xidian University*, 2019, 46(6): 155-162.
- [19] 高嘉乐, 邢清华, 梁志兵. 空天高速目标探测跟踪传感器资源调度模型与算法[J]. *系统工程与电子技术*, 2019(10): 2223-2251.  
GAO J L, XING Q H, LIANG Z B. Multiple sensor resource scheduling model and algorithm for high speed target tracking in aerospace[J]. *Systems Engineering and Electronics*, 2019(10): 2223-2251.
- [20] 杨善超, 田康生, 李宏权, 等. 综合优先级下反导预警相控阵雷达任务调度算法[J]. *兵工学报*, 2020, 41(2): 315-323.  
YANG S C, TIAN K S, LI H Q, et al. Comprehensive priority based task scheduling algorithm for anti-missile early warning phased array radar[J]. *Acta Armamentarii*, 2020, 41(2): 315-323.
- [21] 段毅, 谭贤四, 曲智国, 等. 基于分支定界法的相控阵雷达事件调度算法[J]. *电子学报*, 2019, 47(6): 1309-1315.  
DUAN Y, TAN X S, QU Z G, et al. Phased array radar task scheduling algorithm based on branch and bound method[J]. *Acta Electronica Sinica*, 2019, 47(6): 1309-1315.

编辑 税红