



中国在线音乐社会的地域多样性分析

耿 屿, 韩筱璞*

(杭州师范大学阿里巴巴复杂科学研究中心 杭州 311121)

【摘要】该文利用网易云音乐的数据集, 分析了中国在线音乐社会的文化结构。分析侧重于用户所关注音乐的地域多样性, 以及定义在地域多样性基础上的用户偏好多样性。计算了各地区的用户所关注音乐的地域多样性和平均偏好多样性, 挖掘了这些多样性同各地经济发展水平之间的相关关系, 观察到经济发达地区的用户更可能关注小众音乐。此外, 基于各地区用户的个体地域多样性分布的差异, 识别出中国当代的两个主要的音乐文化圈, 并通过分析地域多样性同该地区的居民来源多样性的关系来挖掘移民带来的文化影响。这些发现为探测中国的社会文化结构从音乐视角提供了一系列全局性的认知。

关键词 文化结构; 音乐文化圈; 在线音乐社会; 地域多样性

中图分类号 TP311; N94 **文献标志码** A **doi**:10.12178/1001-0548.2019075

Region Diversity Analysis for Chinese Online Music Society

GENG Yu and HAN Xiao-pu*

(Alibaba Research Center for Complexity Sciences, Hangzhou Normal University Hangzhou 311121)

Abstract In this paper, we analyze the cultural pattern of Chinese online music society using the dataset collected from NetEase Cloud Music. Our analysis focuses on the region diversity of user's favorite music, and average preference diversities defined on the region diversity. We calculate the region diversity and the average preference diversity of users at each region, and analyze the correlation between these diversities and the development level of economy, and find that users in the developed regions have higher probability in following niche music. And also, we detect two major musical cultural circles from the differences on the distribution of user's region diversity of each area, and analyze the correlation between region diversity and the diversity on the source of residents to investigate the cultural impact of migration. These findings provides series of global feature on Chinese cultural structure from the perspective of music.

Key words cultural structure; musical cultural circles; online music society; region diversity

音乐, 始终是人类文化的主要展示方式之一, 在人类群体文化演化中扮演着极为重要的角色^[1]。一个人的音乐偏好往往是个体心理特征同外部环境协同影响的结果。在对音乐偏好的研究中, 早期侧重于对个体的音乐感知和音乐偏好同其心理特征之间关系的识别, 并以此为桥梁^[2-8], 延伸至对音乐自身特征、音乐文化的分析等方面^[1, 9]。

近年来, 涌现出一批基于各种音乐大数据的研究。这些研究除了进一步深入挖掘认知风格对音乐偏好的影响外, 还涉及到了音乐感知和音乐偏好的年龄性别效应^[6, 10-11]、影响音乐感知的各类社会性因素等众多方面^[6, 12]。例如, 文献 [13] 基于 5 个主

要音乐类别对美国的城市音乐听众进行调查, 发现人们的音乐偏好同社会经济阶层之间具有强相关性; 文献 [14] 基于音乐类别对虾米音乐用户进行多样性分析, 讨论了各音乐类型之间的文化距离; 文献 [15] 通过神经网络手段训练词嵌入模型, 来挖掘音乐流派所对应的社会阶层。此外还有大量研究讨论和分析了影响音乐作品流行的各类社会因素^[16-18] 和高流行度音乐所具有的特征^[19-20]。

值得注意的是, 在以往这些研究中, 音乐类型或音乐风格始终被作为核心的分类标签。换句话说, 这些研究常常高度依赖于准确的音乐标签信息。但是, 由于当代音乐类型之间的高度融合性,

收稿日期: 2019-03-25; 修回日期: 2019-05-21

基金项目: 国家自然科学基金(61873081, 11622538, 61673150, 61673151); 浙江省基础公益研究计划(LGF18F030007); 浙江省自然科学基金(LR16A050001); 浙江省大学生科技创新活动计划暨新苗人才计划(2019R426067)

作者简介: 耿屿(1992-), 男, 主要从事信息经济与社会复杂性分析方面的研究。

通信作者: 韩筱璞, E-mail: xp@hznu.edu.cn

尽管已经有了多类手段来对海量音乐进行特征提取, 但其准确性依然有着相当大的提高空间^[21-23]。

本文尝试采用另外一种方法从音乐角度来认知社会的文化结构, 不考虑具体的音乐类型, 而是基于在线音乐社会大数据, 直接根据音乐受众的地理位置信息, 通过计算其地域多样性, 来比较挖掘在线音乐社会的文化结构特征。该方法的基本假设是, 从社会文化角度, 把具有完全相同的受众群体的不同音乐类型视作同一类, 即把音乐类型定义在用户群体基础上, 由此来摆脱对音乐标签信息的依赖。本文的研究显示出了这一方法的有效性。

1 数据

1.1 数据来源

本文所使用的数据是 2016 年 9 月 7 日—2016 年 9 月 27 日期间通过网易云音乐所提供的 API 接口获取的网易云音乐的歌单信息。网易云音乐 (<http://music.163.com>, NCM) 是中国最大的在线音乐社区之一。其用户可以通过计算机或手机上的客户端软件, 从其在线音乐库中搜索、收听、下载和评论数百万首的音乐, 编辑歌单并在社交网络上分享他们的感受。该产品自 2013 年 4 月发布以来, 已吸引了数千万注册用户, 庞大的用户群体建立起数以亿计的歌单。

1.2 数据描述

从其 API 接口获取歌单数据的收集规则是: 根据歌单 ID 顺序依次检索, 对每 200 万个歌单 ID, 连续抓取其中起始的 20 万个。由于网易云音乐的歌单 ID 是根据其创建时间依次排序的, 这样得到的是一个占总量比例约为 10%、可以均匀覆盖整个用户群的样本, 总共包含 30 562 590 个歌单。

这些歌单包括两种类型: 一种是每个用户的“我喜欢的音乐”歌单 (以下简称为“红心歌单”), 该类歌单为用户第一次将某首音乐标记为“喜欢”时自动创建, 并且每个用户仅有唯一一个该类歌单, 因此该类歌单是与每一个用户严格对应的。另一种类型是普通歌单, 可以由用户自由创建, 每个用户可以创建不超过 1000 个普通列表。该数据集中一共包含了 4499 164 个红心歌单, 其余为普通歌单。在这些红心歌单中, 歌单非空而且用户名显示非“用户已注销”者共计 3028 351 个, 对应同等数量的用户, 这些歌单就是本文的研究分析对象。

每个歌单所包含的信息包括其创建者的个人信

息 (id、性别、所在省份和城市的地区代码)、列表的创建时间以及列表中每种音乐的信息 (id、标题、专辑等)。

2 数据分析

2.1 地域多样性定义

本文重点从多样性角度来挖掘网络音乐社区的空间结构特征。多样性是进行社会文化结构分析的常用统计量^[14, 24]。参考文献 [25] 所提出的方法, 采用 Shannon 信息熵来定义音乐的受众的地域多样性。其具体计算方法如下:

对于某音乐 i , 其受众的地域分布的信息熵 H_i^M 可以被描述为 p_{ij} 的函数:

$$H_i^M = - \sum_j p_{ij} \log(p_{ij}) \quad (1)$$

式中, p_{ij} 指音乐 i 的用户在地区 j 总的受众中的相对比例:

$$p_{ij} = \frac{n_{ij}}{N_j} \left(\sum_k \frac{n_{ik}}{N_k} \right)^{-1} \quad (2)$$

式中, n_{ik} 是该音乐位于地区 k 的受众数量; N_k 是地区 k 的用户总数。音乐 i 的受众地域多样性被定义为:

$$D_i^M = \frac{H_i^M}{\log(A)} \quad (3)$$

式中, A 为数据集所涉及的地区的总数, 该多样性的值介于 0~1。其值为 0, 说明该音乐的受众仅仅局限在一个地区中; 其值为 1, 说明该音乐的受众均匀地散布在各个地区。

进一步可以定义地区 j 的用户的个体地域多样性为当地所有用户的个体地域多样性的平均值:

$$D_j^A = \frac{\sum_l D_l^U}{N_j} \quad (4)$$

其中,

$$D_l^U = \frac{\sum_m^{L_l} D_m^M}{L_l} \quad (5)$$

为用户 l 的个体地域多样性, 表征的是用户 l 的红心歌单中所有音乐的地域多样性 D^M 的平均值, L_l 为该用户的红心歌单列表长度。较高的地域多样性 D^A 意味着该地区的用户倾向于关注在国内广泛传播的大众化音乐, 而低地域多样性则意味着该地区的用户更多关注小范围流行的音乐或者一些

本地化的音乐。

2.2 各地区用户关注音乐的地域多样性统计

以省(包含直辖市、自治区、特别行政区)为地区区分单位,计算了每个省份用户的地域多样性值,如图 1 所示。总的来说,平均地域多样性最高的地区出现在河南、安徽、江西等中东部省份,而西部地区,如新疆、西藏、青海等地区,还有香港、台湾等地区,其地域多样性则相对较低,显示出西部地区的用户对于全国流行性的音乐的关注相对较少,更多关注流行范围较小的音乐,显露出其音乐文化具有较强的地域性特征。

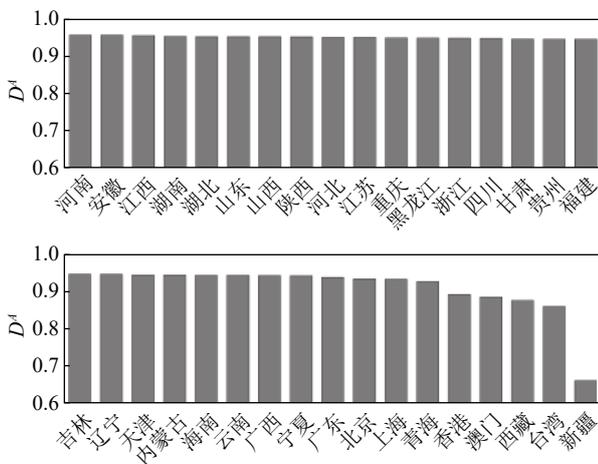


图 1 各地区用户关注音乐的地域多样性

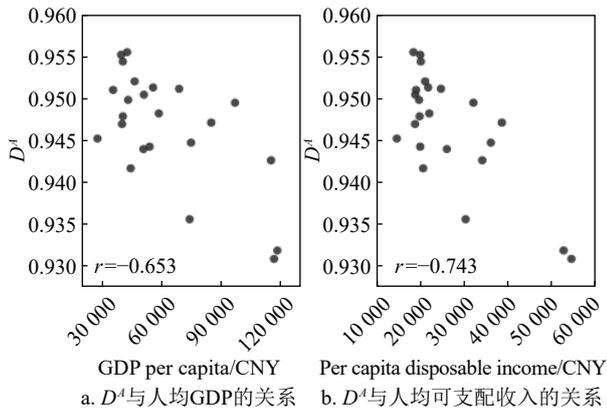


图 2 23 个省市的用户所关注音乐的平均地域多样性 D^1 与其人均 GDP、人均可支配收入的相关关系

同时,北京、上海和广东等经济较发达地区也有着相对较低的平均地域多样性,暗示出地域多样性可能同区域经济发展水平有着密切关系。为了挖掘这一关系,本文从国家统计局网站 (<http://data.stats.gov.cn/>) 获取了 2015 年大陆各省市自治区的人均国民生产总值 (GDP) 和人均可支配收入数据。排除了香港、澳门、台湾和 5 个民族自治区,以及

青海、云南、贵州这 3 个少数民族人口比例超过 20% 的省份,统计了其余 23 个地处中国大陆地区而且汉族占人口绝对多数的省市的平均地域多样性,同他们的人均国民生产总值、人均可支配收入的相关关系。如图 2 所示,该相关性呈现出较强的负相关特征。特别是同人均可支配收入的负相关关系更为强烈,其 Pearson 相关系数 r 为 -0.743 (显著性 $P=4.79 \times 10^{-5}$)。

2.3 各地区用户的偏好多样性

为了观察各地区用户所关注音乐的差异,基于以上地域多样性,定义了各地区用户的偏好多样性。该偏好分布多样性所描述的是,对于一个用户所关注的音乐的地域多样性的多元化程度。将地域多样性的定义范围 $[0, 1]$ 等分为 B 个区间,则用户 i 的偏好多样性可以定义为:

$$D_i^p = \frac{-\sum_j v_{ij} \log(v_{ij})}{\log(B)} \quad (6)$$

式中, v_{ij} 是地域多样性落在第 j 个区间中的歌曲中,在用户 i 的红心歌单音乐中所占的相对比例,表示为:

$$v_{ij} = \frac{m_{ij}}{M_j} \left(\sum_k \frac{m_{ik}}{M_k} \right)^{-1} \quad (7)$$

式中, m_{ik} 是用户 i 的红心歌单音乐中地域多样性落在第 k 个区间中的音乐数量; M_k 是所有用户的红心歌单中地域多样性处于第 k 个区间的音乐总数。对于该偏好多样性,如果某用户所关注的音乐的地域多样性全部集中在某一个区间,则其值为 0; 在各区间的分布越均匀,其值越趋近于 1。

设置 $B=10$, 计算出每个省的用户偏好多样性 D^p 的平均值,如图 3 所示。总体上,与图 1 所示各省份用户关注音乐的平均地域多样性相比,图 2 所呈现趋势大致相反,西部地区 and 少数大城市的用户具有更高的平均偏好多样性,说明这些地区的用户在其所接受的音乐的地域多样性方面有着较高的多元化程度。

同图 2 所示结果一样,同样分析了这 23 个省市的平均偏好多样性,同其人均国民生产总值和人均可支配收入的相关关系。该相关关系如图 4 所示,其相关性分别为 0.620 (显著性 $P=1.60 \times 10^{-3}$) 和 0.711 (显著性 $P=1.43 \times 10^{-4}$), 同图 2 所示的平均地域多样性趋势相反,平均偏好多样性正相关于同人均可支配收入,暗示出在经济

较为发达的地区, 其用户对音乐的选择总体更加倾向于多元化。同样, 相比人均 GDP, 平均偏好多样性与人均可支配收入的相关性更强。这意味着, 影响用户音乐选择的, 主要是该地区居民的实际收入水平。

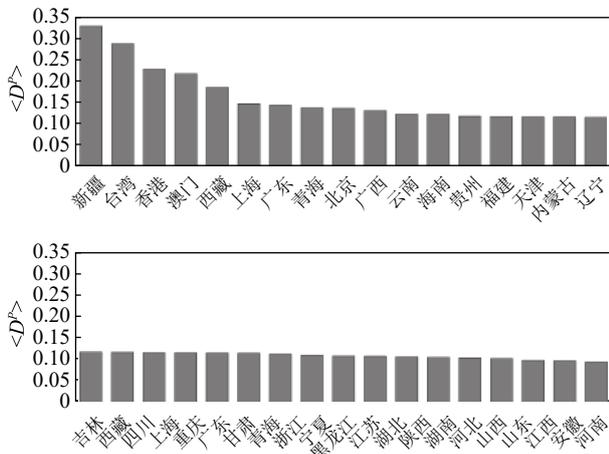


图3 各地区用户的平均偏好多样性 $\langle D^p \rangle$

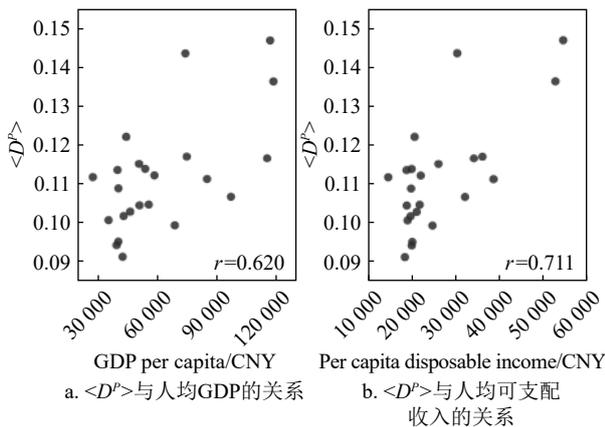


图4 23个省市的用户的平均偏好多样性 $\langle D^p \rangle$ 与其人均GDP、人均可支配收入的相关关系

2.4 各地区用户的个体地域多样性分布

进一步计算了各个省市用户的个体地域多样性 D^u 的分布。其中, 几个代表性省市的分布如图5所示。大部分省市, 其用户个体的地域多样性的分布满足非常相似的单调上升趋势: 大多数用户的个体地域多样性趋近于1, 表示他们主要关注在全国范围内广泛流行的音乐; 而多样性很低的用户比例很少。这一特征在以河南和江西为代表的高地域多样性、低平均偏好多样性的省份中尤其明显, 其个体地域多样性较低的用户比例非常低。而广东、北京、上海等地域多样性较低、平均偏好多样性较高的省市, 其个体地域多样性较低的用户比例稍高一些。由此可见, 导致地域多样性和平均偏好多样性

差异的主要来源是具有较低个体地域多样性的用户的比例不同。考虑到这些多样性同各地经济发展水平的相关性, 可以发现, 经济发展水平较高的地区, 用户较大可能偏好于低地域多样性的小众音乐, 显露出经济发展水平同用户音乐偏好的个性化之间的关系。

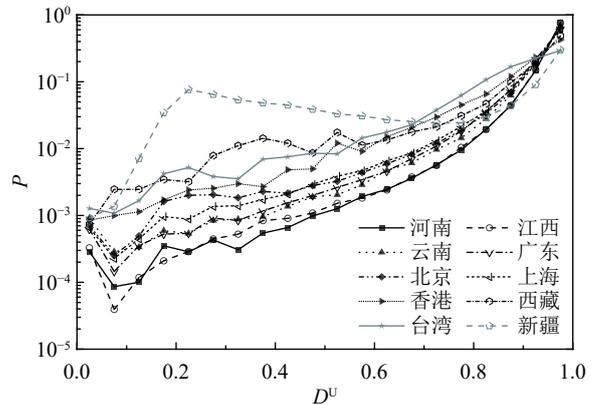


图5 几个代表性地区的用户个体地域多样性 D^u 的分布

此外, 在用户个体的地域多样性的总体上升趋势方面的一致性, 说明在音乐偏好方面, 全国范围内大部分地区之间并不存在强烈的差异, 基本可以归入同一个音乐文化圈。

值得注意的是, 西藏和新疆等地域多样性较低的地区, 其用户个体地域多样性的分布呈现出一定程度的非单调性。特别是新疆, 其分布呈现明显的双峰特征: 除了占据多数的个体地域多样性趋近1的群体之外, 还有大量的用户, 其个体地域多样性相当低, 仅分布在0.2~0.4之间, 如图5所示, 各地区按其平均地域多样性顺序由高到低的排序, 从左到右, 由上到下进行排列。从图中可看出这些用户倾向关注具有强烈地域性的音乐, 而很少关注全国范围内流行的音乐, 显露出他们可以被归类于一个较小的音乐文化圈。同时, 台湾作为一个平均地域多样性也相对较低的地区, 其个体地域多样性在0.8~0.9之间的用户比例相对较高, 显露出同全国大部分用户所属的文化圈的轻微偏离。

2.5 音乐地域多样性与居民来源多样性的关系

一般而言, 人们音乐偏好的形成往往是流行风潮和历史文化记忆的共同结果。近期有研究观察到用户的所居住地域及居住地迁移对其音乐偏好的显著影响^[26]。由于移民会带来其来源地的音乐文化, 这可能影响到移民输入地的音乐偏好多样性, 因此本文对各地区用户所关注音乐的地域多样性同该地

区的居民来源多样性的关系进行了考察。

从国家统计局网站获取到国家统计局人口与就业统计司发布的 2015 年全国 1% 人口抽样调查数据, 其中涉及到移民的信息有, 中国大陆地区各省份的现住居民数量 (1% 抽样), 以及这些现住居民中 5 年前定居于各个外省的数量。根据这一数据, 定义了各个省份的居民来源多样性。

假设省份 i 的现住居民中, 5 年前定居于省份 j 的比例为 s_{ij} (即省份 i 在最近 5 年里来自省份 j 的移民所占人口比例; 其中 j 的取值范围也包括 i 自身, 表示省份 i 的原住居民比例), 则该省 i 的居民来源多样性可以定义为:

$$D_i^s = \frac{-\sum_j s_{ij} \log(s_{ij})}{\log(C)} \quad (8)$$

式中, $C=31$, 为中国大陆地区的省市自治区总数。计算了各省的居民来源多样性 D^s 的值, 总体而言, 北京、上海和广东等覆盖一线城市的省份居民多样性较高, 居民多样性最低的地区分布在黑龙江、安徽、湖南、河南等省份。

进一步, 考察各地区 (23 个省市) 的居民来源多样性 D^s 与各省用户所关注音乐的平均地域多样性 D^d 的相关关系。考虑到移民来源较为多元化的地区, 由于其关注的音乐应该存在同各个移民来源地的用户的所关注音乐存在关联, 使得其应该具有较高的地域多样性, 因此对这一相关性的预期是正相关。然而, 实际统计发现, D^s 与 D^d 呈现出较强的负相关特征, 如图 6 所示, 其 Pearson 相关系数为 -0.745 (显著性 $P=4.54 \times 10^{-5}$), 说明居民来源较为多元化的地区, 其用户所关注的音乐反而显露出较强的本地化特征。这一现象同本文的预期结果完全相悖。

考虑到居民来源多样性 D^s 同样与当地的经济水平密切相关 (同各省人均可支配收入的相关性为 0.887), 通过计算 D^s 与 D^d 的一阶偏相关系数, 来剔除居民来源多样性同人均可支配收入的相关关系影响。该一阶偏相关系数为 -0.276 , 依然呈现为负相关, 但并不显著 (显著性 $P=0.213$)。尽管该负相关特性并不显著, 但仍然显示出, 经济发展同用户音乐偏好的关系之强烈, 几乎可以完全覆盖移民所带来的来自其来源地的音乐文化的影响。

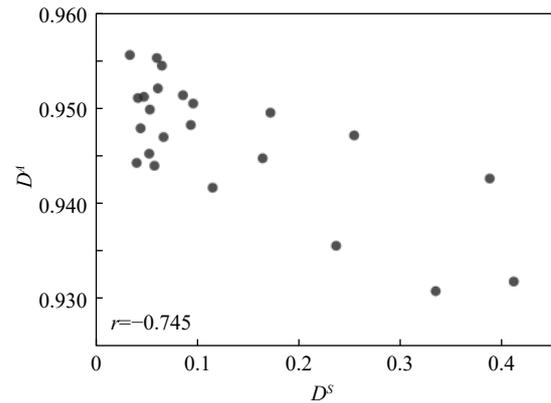


图 6 23 个省市的用户所关注音乐的的平均地域多样性 D^d 与其居民来源多样性 D^s 的相关关系

3 结束语

综上所述, 本文通过对网易云音乐的大量歌单数据进行基于地域多样性的分析, 挖掘了各省市地区的用户所关注音乐的差异。本文发现, 绝大多数用户大致可以归入两个音乐文化圈: 一个音乐文化圈的用户覆盖了全国的大多数用户; 另一个较小的文化圈主要由新疆地区的部分用户构成。同时也观察到, 经济发展水平同用户音乐偏好之间的关系: 经济发展程度较高的地区, 用户对小众音乐的偏好倾向更为强烈。这些发现, 从音乐角度揭示了中国当代社会文化结构的部分特征。

最后需要说明的是, 从方法论的角度, 本文对音乐文化的分析没有像以往研究那样去直接涉及音乐本身的特征 (例如对音乐风格的分析等^[20]), 而是将文化的区分完全定义在接受文化的用户之上。这种方法使得对社会文化的研究可以脱离具体文化元素的局限, 进而通过建立“文化”同社会网络的直接联系来进行。这一方法在社会文化分析领域可能有着更大的应用潜力。

参 考 文 献

- [1] CLAYTON M, HERBERT T, MIDDLETON R. The cultural study of music: A critical introduction[M]. New York: Routledge, 2003.
- [2] RENTFROW P J, GOSLING S D. The do re mi's of everyday life: The structure and personality correlates of music preferences[J]. Journal of Personality & Social Psychology, 2003, 84(6): 1236-1256.
- [3] NORTH A C. Individual differences in musical taste[J]. American Journal of Psychology, 2010, 123(2): 199-208.
- [4] RENTFROW P J, GOLDBERG L R, LEVITIN D J. The structure of musical preferences: A five-factor model[J]. Journal of Personality and Social Psychology, 2011, 100(6): 1139-1157.

- [5] GREENBERG D M, BARON-COHEN S, STILLWELL D J, et al. Musical preferences are linked to cognitive styles[J]. *PloS One*, 2015, 10(7): e0131151.
- [6] BONNEVILLEROUSSY A, RENTFROW P J, XU M K, et al. Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood[J]. *Journal of Personality & Social Psychology*, 2013, 105(4): 703-717.
- [7] LANGMEYER A, GUGLHÖR-RUDAN A, TARNAI C. What do music preferences reveal about personality? A cross-cultural replication using self-ratings and ratings of music samples[J]. *Journal of Individual Differences*, 2012, 33(2): 119-130.
- [8] HALLAM S, CREECH A, MCQUEEN H. Pupils' perceptions of informal learning in school music lessons[J]. *Music Education Research*, 2018, 20(2): 213-230.
- [9] DOWLING W J, TIGHE T J. Psychology and music: The understanding of melody and rhythm[M]. London: Psychology Press, 2014.
- [10] RAWLINGS D, CIANCARELLI V. Music preference and the five-factor model of the NEO personality inventory[J]. *Psychology of Music*, 1997, 25(2): 120-132.
- [11] KALIA A. Music was better back then: When do we stop keeping up with popular music?[EB/OL]. (2015-04-22). <https://skynetandebert.com/2015/04/22/music-was-better-back-thenwhen-do-we-stop-keeping-up-with-popular-music/>.
- [12] ROY W G, DOWD T J. What is sociological about music?[J]. *Annual Review of Sociology*, 2010, 36(1): 183-203.
- [13] MELLANDER C, FLORIDA R, RENTFROW P J, et al. The geography of music preferences[J]. *Journal of Cultural Economics*, 2018(4): 1-26.
- [14] LI H, HAN X P, LÜ L, et al. Measuring diversity of music tastes in online musical society[J]. *International Journal of Modern Physics C*, 2018, 29(5): 1840006.
- [15] KOZLOWSKI A C, TADDY M, EVANS J A. The geometry of culture: Analyzing meaning through word embeddings[J]. *American Sociological Review*, 2019, 84(5): 905-949.
- [16] GODART F. The power structure of the fashion industry: Fashion capitals, globalization and creativity[J]. *International Journal of Fashion Studies*, 2014, 1(1): 39-55.
- [17] CARMEL D, ZWERDLING N, GUY I, et al. Personalized social search based on the user's social network[C]// Proceedings of the 18th ACM Conference on Information and Knowledge Management. [S.l.]: ACM, 2009: 1227-1236.
- [18] UZZI B, SPIRO J. Collaboration and creativity: The small world problem[J]. *American Journal of Sociology*, 2005, 111(2): 447-504.
- [19] ASKIN N, MAUSKAPF M. What makes popular culture popular? Product features and optimal differentiation in music[J]. *American Sociological Review*, 2017, 82(5): 910-944.
- [20] SHIN S, PARK J. On-chart success dynamics of popular songs[J]. *Advances in Complex Systems*, 2017, DOI: 10.1142/S021952591850008X.
- [21] FRIBERG A, SCHOONDERWALDT E, HEDBLAD A, et al. Using listener-based perceptual features as intermediate representations in music information retrieval[J]. *The Journal of the Acoustical Society of America*, 2014, 136(4): 1951-1963.
- [22] PACHET F, AUCOUTURIER J J. Improving timbre similarity: How high is the sky[J]. *Journal of Negative Results in Speech and Audio Sciences*, 2004, 1(1): 1-13.
- [23] POHLE T, PAMPALK E, WIDMER G. Evaluation of frequently used audio features for classification of music into perceptual categories[C]// Proceedings of the Fourth International Workshop on Content-Based Multimedia Indexing. [S.l.]: [s.n.], 2005: 162.
- [24] PARK M, WEBER I, NAAMAN M, et al. Understanding musical diversity via online social media[C]// In Proceedings of the 9th International AAAI Conference on Web and Social Media. [S.l.]: AAAI Press, 2016: 308-317.
- [25] EAGLE N, MACY M, CLAXTON R. Network diversity and economic development[J]. *Science*, 2010, 328(5981): 1029-1031.
- [26] WAY S F, GIL S, ANDERSON I, et al. Environmental changes and the dynamics of musical identity[EB/OL]. [2019-02-10]. <https://arxiv.org/pdf/1904.04948.pdf>.

编辑 叶芳