

# 基于社交媒体的城市旅游区特征分析 ——以苏州市为例

王雯夫<sup>1,2</sup> 陈子豪<sup>1,2</sup> 孙奇<sup>1,2</sup> 张毅<sup>1,2,†</sup>

1. 北京大学遥感与地理信息系统研究所, 北京 100871; 2. 空间信息集成与3S工程应用北京市重点实验室, 北京 100871; † 通信作者, E-mail: zy@pku.edu.cn

**摘要** 提出以面状旅游区为分析单元的新方法。首先从海量社交媒体数据中提取个体旅游时空行为, 然后基于行为计算城市的旅游区结构, 最后分析旅游区的游客行为特征和游客客源特征, 以旅游区为节点、旅游区之间的旅游流为关系构建旅游区网络, 并分析其结构特征。在对苏州市的实证研究中, 基于行为计算得到7个旅游区, 这一结果与苏州市旅游规划中的“一核一带三区”空间格局基本上一致。对苏州市旅游区的特征分析表明: 1) 古城和古镇是苏州市旅游的核心区域, 游客分散且距离远; 2) 苏州城区及附近的旅游区更能吸引跨旅游区的旅游行为; 3) 苏州旅游区网络已经形成多中心结构。通过实证研究, 验证了基于社交媒体数据计算城市旅游区结构和以面状旅游区为分析单元研究城市旅游的有效性, 为城市旅游研究提供新思路。

**关键词** 城市旅游; 社交媒体; 时空行为; 旅游区; 网络分析

## Feature Analysis of City's Tourism Districts Based on Social Media: A Case Study of Suzhou

WANG Wenfu<sup>1,2</sup>, CHEN Zihao<sup>1,2</sup>, SUN Qi<sup>1,2</sup>, ZHANG Yi<sup>1,2,†</sup>

1. Institute of Remote Sensing and Geographical Information System, Peking University, Beijing 100871; 2. Beijing Key Laboratory of Spatial Information Integration and Its Applications, Beijing 100871; † Corresponding author, E-mail: zy@pku.edu.cn

**Abstract** A new method is proposed using area feature-tourism districts as analysis unit. Firstly, spatial-temporal behaviors of individual tourists are extracted from social media data. Secondly, the city's tourism districts are extracted based on spatial-temporal behaviors. Finally, tourism districts are analyzed by 3 kinds of features—tourist activity features, tourist origin features and the structure features of tourism district network using tourism districts as nodes and tourist flow as edges. In the empirical research of Suzhou, 7 tourism districts are extracted based on spatial-temporal behaviors. The spatial structure of tourism districts is generally the same as “1-core-1-corridor-3-district” pattern in Suzhou tourism planning. The feature analysis of Suzhou tourism districts indicates that the Ancient City Tourism District and the Ancient Town Tourism District are the core tourism districts, which attracts tourists from various and distant origins. The tourism districts in and near Suzhou urban area attract more tourists. Suzhou tourism districts have already formed into a multi-core structure. This research shows the effectiveness of extracting city's tourism districts based on social media data and researching city's tourism with tourism districts as analysis unit, providing a new approach for research on urban tourism.

**Key words** urban tourism; social media; spatial-temporal behavior; tourism district; network analysis

基于游客行为的城市旅游研究对城市旅游规划、管理和政策制定具有指导意义, 是城市旅游研究的一个重要领域<sup>[1]</sup>, 其研究内容主要包括目的地

城市的客源特征、活动特征和旅游流网络特征等, 分析单元多为城市或景区, 数据源主要是传统的问卷调查数据, 近年来也广泛使用互联网用户生成数

据(user generated content, UGC)。在客源特征方面,国外较早开始对游客客源的理论研究<sup>[2-3]</sup>,其中城市客源是一个主要的关注点<sup>[4]</sup>。针对典型旅游城市,已经有一系列基于客源特征的实证研究<sup>[5-7]</sup>。在游客活动特征方面,主要关注游客运动模式<sup>[8]</sup>、路径<sup>[9-10]</sup>和热点区域<sup>[11]</sup>等。城市旅游流网络特征研究一般以城市中的吸引物为节点、旅游流为关系建立网络,通过网络分析,研究城市目的地的空间结构<sup>[12-13]</sup>。

随着互联网时代的来临,诸如Twitter, Flickr 和新浪微博等平台产生海量旅游相关的用户生成数据。利用这些数据,可以对游客的时空活动进行细致的刻画,为基于游客行为的城市旅游研究提供新的机遇<sup>[14]</sup>。近十年来,新数据被应用在基于游客行为的城市旅游研究中。在客源特征方面,一些研究开始使用微博等用户生成数据获取游客的客源地分布<sup>[15-16]</sup>,但方法创新较少。在游客活动特征方面,一些代表性的研究通过具有地理标签的照片数据获取游客活动的轨迹,研究热点旅游线路<sup>[17-18]</sup>和游客活动模式<sup>[19]</sup>等问题。在城市旅游流网络方面,以网络游记为主,成为旅游流的新数据源,但基本上仍以热门景点为节点<sup>[20-23]</sup>。

目前的研究基本上以城市整体或景区(吸引物)等点状要素为分析单元。根据旅游空间理论,城市旅游空间由点、线和面3类基本要素组成<sup>[24]</sup>。点要素即旅游节点,由相互联系的吸引物和接待服务设

施组成,通常是旅游活动最活跃的地方。线要素即旅游通道,连接目的地城市中的旅游节点。面要素(本文中称为旅游区)是旅游要素(吸引物和接待服务设施)聚集的空间载体,具有鲜明的旅游主题,是游客旅游行为和消费的主要场所。在全域旅游时代,旅游目的地的发展在空间上不再局限于热门吸引物,而是以一系列的吸引物为中心,形成整体形象突出、设施完善的综合性区域空间<sup>[25]</sup>,意味着城市中将形成多个主题鲜明、能满足多样性旅游需求与旅游期望的旅游区<sup>[26]</sup>。因此,以城市整体或景区为分析单元的城市旅游研究具有局限性,若以旅游区为分析单元,则能更好地反映全域旅游时代的城市旅游空间结构特征。

本文提出一种研究方法——使用社交媒体大数据,以旅游区为研究单元,进行基于用户行为的城市旅游研究。该方法的技术流程(图1)包括如下几个步骤:1)通过数据预处理,从原始社交媒体数据中提取外地游客的旅游时空行为;2)基于旅游时空行为,使用网络模型中的社区发现算法提取旅游区;3)计算旅游区的游客行为特征、游客客源特征和网络结构特征。

### 1 数据与预处理

本文以苏州市为研究对象,以新浪微博为数据源,使用微博API抓取2012—2013年在苏州市范围内发布的带有地理坐标的微博5399161条(其中签

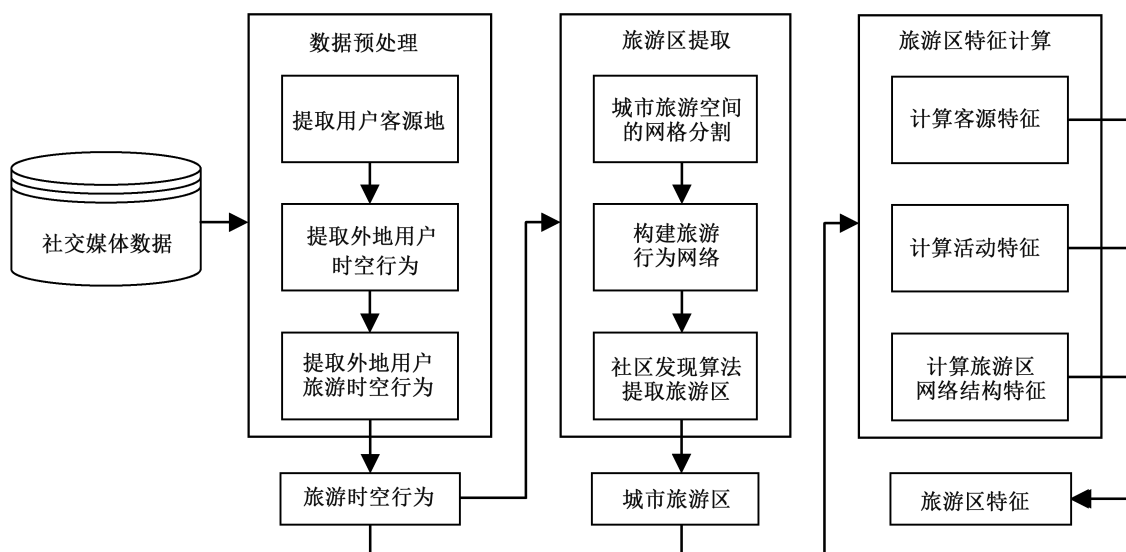


图1 城市旅游区特征分析流程  
Fig. 1 Workflow of feature analysis of city's tourism districts

到微博1707413条),涉及470041个用户。

通过以下步骤,对原始社交媒体信息(微博)进行预处理,从中提取外地游客在目的地城市的旅游时空行为。

首先提取用户的常居地信息。微博用户数据中的居住地信息多有缺失,我们将用户发布微博信息最多的城市作为用户的常居地城市(若不止一个发布微博数最多的城市,则选择用户最早发布微博的城市)。

然后,提取外地用户的时空行为。根据用户常居城市信息,确定目的地城市的外地用户,并从外地用户的所有地理微博中提取该用户在目的地城市的时空行为。一个时空行为 $B$ 表示为一系列时空四元组 $(u, t, g, c)$ (每个四元组对应一条微博信息)的有序集合:

$$B = (w_1, w_2, \dots, w_n), w_i = (u_i, t_i, g_i, c_i), i = 1, 2, \dots, n,$$

其中, $u$ 是用户标识符,一个时空行为的四元组都来自同一个用户; $t$ 是微博发送时间,一个时空行为中时空四元组的微博发送时间间隔不超过规定值,本研究中规定一个行为中的微博发送间隔不超过72小时; $g$ 是微博发布位置,表示为经纬度坐标,一个时空行为中的 $g$ 必须在同一个目的地城市内; $c$ 是微博文本内容。每个时空行为反映外地人在目的地城市一次活动的全过程。

最后,从时空行为中提取旅游时空行为。定义一个时空行为是旅游时空行为,当且仅当其行为四元组集合中至少存在一个四元组,它的位置( $g$ )在景区(吸引物)范围内。本研究选取的景区为苏州市的5A和4A级景区以及主要旅游风情商业街,共40个。本文提出一种景区语义标注算法,用来判断一个地理坐标是否属于某个景区。该算法的基本步骤如下:1)人工标注属于景区的POI,从而获得景区的签到地理微博,作为训练集;2)基于密度聚类DBSCAN算法<sup>[27]</sup>,计算景区的行为空间,算法参数需要根据实际情况确定;3)基于景区行为空间,判断地理坐标是否位于景区内。旅游时空行为反映外地人在目的地城市的一次旅游活动的全过程。

通过预处理,提取出290180个外地用户的565471次时空行为,并从中提取出168052个外地游客的221294次游客时空行为。为测试旅游行为提取的实际效果,人工对2814个随机选取的外地用户时空行为进行旅游-非旅游标注,作为测试集。

通过准确率( $P$ )和召回率( $R$ )两个指标<sup>[28]</sup>,对旅游行为提取结果进行评价,得到 $P=77.70\%$ , $R=85.37\%$ ,显示提取效果较好。

## 2 旅游区提取与特征分析

### 2.1 旅游区提取

Shao等<sup>[29]</sup>提出一种基于社交媒体大数据的旅游区计算方法,采用基于网络模型的社区发现算法来提取旅游区。该算法的基本原理如下。

根据定义,旅游区是旅游要素(吸引物和旅游服务设施)聚集的空间载体,具有明确的旅游主题,在功能上能够满足游客的综合需求<sup>[25]</sup>。旅游要素的聚集会导致游客在一段时间内的主要活动和消费在同一旅游区内进行<sup>[30]</sup>。受时间限制,游客一日内的行为主要在旅游区内的旅游要素间移动。因此,基于从社交媒体大数据中提取的游客旅游时空行为,根据其一日行为计算行为聚集的空间区域,可以得到旅游要素聚集的空间区域,即旅游区。

本文基于网络模型,利用社区发现算法提取旅游区。网络模型由节点之间的边构成,边的权重体现节点间联系的紧密程度。首先,将地理单元作为节点。为获取地理单元,基于规则网格方法,对目的地城市进行空间划分,每个网格(地理单元)的尺寸是500 m×500 m。然后,基于旅游时空行为中每一日的活动构建网格间的边,记网格个数为 $n$ ,网络的邻接矩阵 $R$ 定义如下:

$$R \in \mathbb{R}^{n \times n},$$

$$R_{ij} = \begin{cases} \text{同一天内网格 } i, j \text{ 的旅游行为数}, & i \neq j, \\ 0, & i = j. \end{cases}$$

最后,使用社区发现算法提取旅游区。社区发现算法将联系紧密的节点划分到同一社区中,不同社区节点之间的联系则较为松散<sup>[31]</sup>。由于游客一日内的活动主要在同一旅游区内进行,网络中同一旅游区内的网格联系紧密,不同旅游区之间的网格联系松散,故社区发现算法划分为同一社区的网格所占据的地理范围即为旅游区。我们采用fast-greedy社区发现算法,该算法在节点较多的网络上计算速度较快<sup>[32]</sup>。

图2为基于行为计算得到的苏州市旅游区结构。在计算结果中排除网格数少于50或网格未形成聚集的社区后,得到如下7个旅游区。

1) 古城旅游区(D1)。呈聚集状分布,空间范围

包括整个古城,并且向西延伸至丘山、运河和胥江一带。苏州最重要的旅游资源(古典园林、文物古迹和历史街区)集中在本区,5A级景区有拙政园、留园和虎丘山,4A级景区有狮子林、网师园、盘门、七里山塘、平江历史街区、寒山寺和西园戒幢律寺。D1区还包括观前街、十全街、石路、十梓街和凤凰街等苏州最主要的商业休闲街。

2) 金鸡湖旅游区(D2)。呈聚集状分布,空间范围从古城东部边缘到金鸡湖。D2区的核心吸引物金鸡湖是国家商务旅游示范区、5A级景区、中国最大的城市湖泊公园。

3) 高新区旅游区(D3)。大致分成两部分:①从古城西部边缘到苏州乐园,呈聚集状分布;②向西延伸至大阳山和太湖湿地,向北延伸至大白荡,以吸引物为中心呈散点状分布。D3区的主要景区有4A级景区苏州乐园、太湖湿地公园、中国刺绣艺术馆和大阳山国家森林公园。

4) 太湖旅游区(D4)。位于苏州太湖湖畔,覆盖吴中区的大部分。D4区是古城外另一个高等级吸引物的聚集地,分布形态已经基本上连成片。D4景区类型以山水自然风光为主,包括5A级的东山、旺山和穹窿山,4A级的西山、木渎、光福、石湖、天平山和天池山以及太湖国家旅游度假区。

5) 古镇旅游区(D5)。主要分布在昆山市以及靠近昆山的吴江区北部和吴中区东北角。除距离较远的木渎和沙溪古镇外,苏州其他4A级以上古镇都位于D5区,包括5A级的周庄和同里,4A级的角直、千灯和锦溪。因此,D5区的空间形态是按照古镇位置呈散点状分布。另外,D5区还包括昆山市区和同里附近的吴江区市区。

6) 阳澄湖旅游区(D6)。基本上分布在阳澄湖周边,涉及昆山市、相城区和工业园区3个县级行政区,空间上覆盖南岸工业园区的国家级阳澄湖半岛旅游度假区、西岸相城区的省级阳澄湖生态休闲旅游度假区和东岸巴城阳澄湖度假中心。

7) 沿长江旅游区(D7)。分布在苏州市沿长江的张家港、常熟和太仓3个县级市,空间形态上比较分散,根据行政区又可以划分为3个子区。

如图3所示,根据《苏州市旅游发展总体规划(2009—2020)》,苏州市的旅游区结构是“一核一带三区”。“一核”是主城区旅游发展极核心区,包括古城、工业园区和高新区三部分。“一带”是沿江休闲旅游带,主要包括张家港-常熟-太仓一线的沿江

江地区。“三区”是环太湖休闲度假旅游区、中部湖荡生态休闲旅游区和南部水乡古镇观光休闲旅游区。环太湖区以太湖国家旅游度假区为核心,包括吴中区、高新区环太湖地区和吴江区环太湖地区。中部湖泊区包括昆承湖、尚湖和阳澄湖。南部古镇区包括周庄、同里、角直、锦溪、千灯和黎里等水乡古镇。

通过对比发现,本文计算得出的旅游区与苏州市的旅游规划的结构基本上一致,D1,D2和D3分别对应“一核”中的古城、工业园和高新区,D4,D5和D6分别对应“三区”中的环太湖休闲度假旅游区、中部湖荡生态休闲旅游区和南部水乡古镇观光休闲旅游区,D7则对应“一带”的沿江休闲旅游带。同时,计算结果与规划之间存在一些差异:1)规划中的沿江旅游带是沿长江发展生态休闲旅游,而计算结果显示长江边的3个苏州县级市的旅游空间(D7)还是沿着交通线和城镇发展;2)规划中的中部湖荡旅游区包括常熟的尚湖和昆承湖以及横跨昆山、相城和工业园区的阳澄湖,计算结果中,尚湖和昆承湖与其他常熟旅游空间一起同属于沿长江旅游区(D7),阳澄湖发展为独立的阳澄湖旅游区(D6)。

## 2.2 旅游区特征分析

基于旅游时空行为,我们通过游客客源特征、游客活动特征和旅游区网络结构3类特征,对旅游区进行特征分析。其中,游客客源特征反映旅游区的游客客源结构,游客活动特征反映游客在旅游区的活动情况,旅游区网络结构特征反映旅游区之间的关系以及城市的旅游区结构。

### 2.2.1 游客客源特征

游客客源特征的计算基于根据客源地对游客的划分,从游客的客源地构成角度反映旅游区的特征。本文针对每个旅游区,计算旅游研究中常用的两个客源地特征:吸引力半径和客源地集中度<sup>[33]</sup>。

1) 吸引力半径(attraction radius, AR)是衡量分析单元吸引力的指标。AR值越大,分析单元的吸引范围越大,对游客的吸引力越强。根据 Smith<sup>[34]</sup>提出的中心地标准距离,AR的计算公式如下:

$$AR = \sqrt{\frac{\sum_{i=1}^n x_i^2 d_i^2}{\sum_{i=1}^n x_i^2}}$$

式中, $x_i$ 是第*i*个客源地的游客人数占游客总人数的比例, $d_i$ 是第*i*个客源地与分析单元的距离。本研究

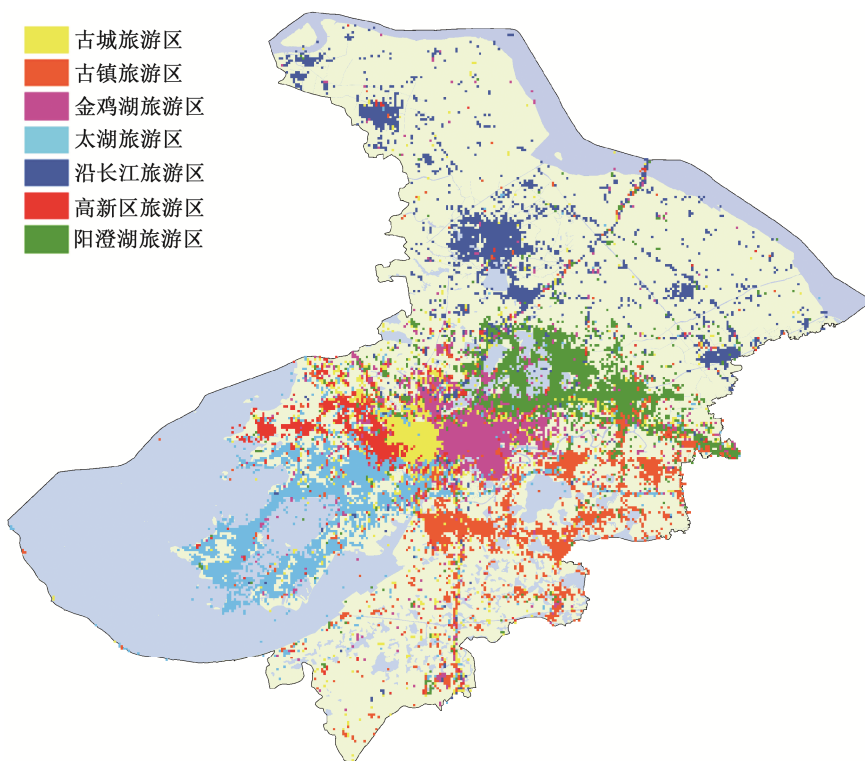
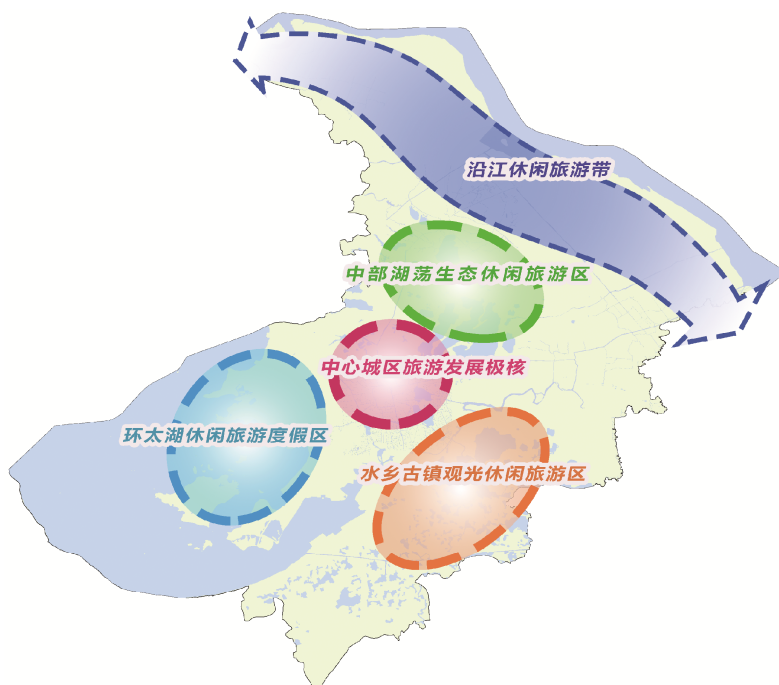


图 2 社区发现算法提取的苏州市旅游区

Fig. 2 Tourism districts extracted by applying community detection algorithm



根据《苏州市旅游发展总体规划(2009—2020)》修改

图 3 苏州市旅游区规划图

Fig. 3 Suzhou tourism districts in planning

中,客源地选取除港澳台外的中国31个省级行政区,其他客源地的游客不参与计算,  $d_i$ 取第*i*个省级行政区省会城市中心点到目的地城市中心点的直线距离。

2) 客源地地理集中度是衡量客源集中程度的指标,计算公式<sup>[35]</sup>如下:

$$G = 100 \times \sqrt{\sum_{i=1}^n \left(\frac{x_i}{T}\right)^2}$$

式中,  $G$ 是旅游区客源地的地理集中度,  $x_i$ 是第*i*个客源地的游客数量,  $T$ 是旅游区接待游客总数。  $G$ 值越接近100,游客越集中于少数客源地;反之则客源地越分散,目的地可以吸引更多客源地的游客。

旅游区的客源特征计算结果见表1。可以看出,古城和古镇的吸引半径都在300 km左右,明显超过其他旅游区(200 km以下)。同时,古城和古镇的地理集中度都在40左右,表明其客源地相对分散;其他旅游区的地理集中度都大于50,客源地相对集中。吸引半径最小、客源地最集中的旅游区是阳澄湖和太湖,表明其游客大多来自周边几个城市。

### 2.2.2 游客活动特征

旅游活动特征主要包括旅游偏好、重游率、旅游模式、旅游地选择和滞留时间等<sup>[8-12]</sup>。本文基于旅游时空行为体现的游客活动,对每个旅游区计算旅游人次、单旅游区模式旅游行为比例以及重游率3个游客活动特征。

1) 旅游人次:直接反映旅游区的旅游活动强度。基于提取的旅游行为,统计到达目的地城市每个旅游区的旅游行为次数,即所研究时段内的旅游人次。

2) 单旅游区模式旅游行为比例:城市旅游出行模式研究通常以景区或景点为单元<sup>[36-38]</sup>,本文以旅

游区为单元,将出行模式分为单旅游区和多旅游区两种模式。单旅游区模式旅游行为只在一个旅游区活动,多旅游区模式旅游行为则至少在两个旅游区活动。本文计算单旅游区模式旅游行为的占比,计算公式为

$$\begin{aligned} & \text{单旅游区模式旅游行为比例} \\ & = \text{单旅游区模式旅游行为数} / \text{旅游行为总数} \end{aligned}$$

该比例较高的旅游区称为孤立型旅游区,这类旅游区常作为游客一次旅游活动唯一的旅游区目的地;该比例较低旅游区称为依附型旅游区,这类旅游区很少作为游客的唯一旅游区目的地,在旅游活动中依附于其他旅游区;该比例居中的旅游区称为混合型旅游区,将其作为唯一目的地的旅游行为和跨旅游区的旅游行为次数相近。

3) 重游率:常用来反映地区或吸引物吸引游客反复旅游的能力<sup>[39]</sup>,游客的重游活动对旅游企业具有较大的意义<sup>[40]</sup>。本文将旅游区的重游率定义为所研究时段(2012—2013年)内平均每人到目的地城市旅游的次数,即

$$\text{重游率} = \text{旅游人次} / \text{旅游人数}。$$

旅游区的游客活动特征计算结果如表2所示,苏州市旅游区接待人次排序依次是古城、金鸡湖、古镇、环太湖、阳澄湖、高新区和沿长江。以商务旅游为主的金鸡湖重游率最高(1.28次),古镇最低(1.07次)。高新区的单旅游区模式旅游行为比例明显较低,只有27.62%,是明显的依附型旅游区;其他旅游区均在50%~70%之间,属于混合型旅游区,其中东北部的沿长江旅游区单旅游区模式旅游行为比例最高(67.77%),说明该旅游区的孤立程度最高。

### 2.2.3 网络结构特征

社会网络分析法是社会学中一种成熟的研究方

表1 旅游区客源特征

Table 1 Tourist origin features of tourism districts

旅游区	地理集中度	吸引半径/km
古城	39.4	322.98
古镇	43.0	283.33
金鸡湖	50.1	176.83
太湖	62.3	118.73
沿长江	56.3	131.12
高新区	50.5	159.99
阳澄湖	68.4	112.09

表2 旅游区游客活动特征

Table 2 Tourist activity features of tourism districts

旅游区	旅游人次	单旅游区模式旅游行为比例/%	重游率
古城	100942	61.62	1.18
古镇	37831	60.49	1.07
金鸡湖	52036	52.12	1.28
太湖	29636	57.42	1.14
沿长江	10300	67.77	1.11
高新区	12906	27.62	1.11
阳澄湖	32447	65.35	1.14

法,通过个体间的关系来研究个体特征或社群的整体特征<sup>[41]</sup>。本文利用以旅游区为节点的网络,通过社会网络分析方法,比较各旅游区的重要程度和地位,并研究目的地城市的空间结构。

以目的地城市的旅游区为节点,游客在旅游区之间的旅游流为纽带,构建城市旅游区网络。可以将其表示为一个加权无向图,记旅游区个数为  $m$ ,其邻接矩阵  $T$  定义如下:

$$T \in \mathbb{R}^{m \times m},$$

$$T_{i,j} = \begin{cases} \text{同时到达旅游区 } i \text{ 和 } j \text{ 的旅游行为数, } i \neq j, \\ 0, i = j. \end{cases}$$

基于旅游区网络,本文计算度中心度和核心-边缘特征两个指标。

1) 度中心度用来测量旅游区网络中旅游区的重要程度,有权无向图的度中心度计算公式<sup>[42]</sup>为

$$D_i = \sum_{j=1}^n x_{ij} (i \neq j),$$

式中,  $D_i$  是城市第  $i$  个旅游区的度中心度,  $x_{ij}$  是同时访问第  $i$  个和第  $j$  个旅游区的旅游行为数目。

2) 核心-边缘特征分析是将节点分为联系紧密的核心节点和联系稀疏的边缘节点,核心节点构成子图的联系越紧密,同时边缘节点构成子图的联系越稀疏,则网络的核心-边缘特征越明显<sup>[43]</sup>。本文使用 Hamming 算法进行核心-边缘计算,该算法是核心-边缘分析中常用的方式之一<sup>[44]</sup>。

城市旅游流网络研究中,度中心度反映研究单元在城市中的角色、位置以及重要程度;核心-边缘特征反映区域内旅游要素的不平衡性,对优化区域旅游结构、协同区域旅游关系具有指导意义<sup>[45]</sup>。我们以提取出的7个苏州市旅游区作为网络节点,以旅游区之间的旅游流为联系纽带,建立苏州市的旅游区网络,其邻接矩阵如表3所示,计算得到的旅游区网络结构特征如表4所示。从旅游区的网络

结构特征来看,古城旅游区的度中心度为50570,明显高于其他旅游区,表明在多旅游区旅游行为中,选择古城旅游区的频率最高,其次是金鸡湖(33822)和古镇(20341)。核心-边缘分析显示,苏州旅游核心区域是古城、古镇、金鸡湖和太湖旅游区,边缘区是沿长江、高新区和阳澄湖旅游区。其中,核心节点构成的子图密度为7749.2(密度为图中边的度的平均值<sup>[41]</sup>,根据表3计算得出),边缘节点构成的子图密度为789.3,仅为核心区的约1/10,说明旅游区之间形成明显的核心-边缘结构。核心旅游区的联系紧密,表明较多的游客在一次旅游行为中会游览这些旅游区中的多个,边缘旅游区则较为独立,难以吸引游客在一次旅游行为中同时到此游览。

### 3 结论

近年来新数据和新方法逐步应用在旅游学中,推动了旅游定量研究的发展。在城市旅游研究中主要以景区为分析单元,对全面认识和评价城市旅游综合发展具有较大的局限性。本文提出以城市旅游区为分析单元、以社交媒体大数据为数据源的新方法,丰富了旅游区尺度的城市旅游研究的数据源。基于空间聚类的景区语义标注等方法,实现从社交媒体大数据中提取城市个体旅游行为;基于社区发现算法,从旅游行为中提取行为聚集区,实现城市旅游区的提取;以城市旅游区为分析单元,分析了旅游区的游客行为特征、游客客源特征和网络结构特征。

在苏州的实证研究中,基于旅游行为计算得到7个旅游区,与规划中的“一核一带三区”基本上一致,表明苏州旅游的实际发展与规划高度一致,呈现旅游产业和要素聚集,表现出旅游行为的高度聚集。以旅游区为分析单元,分析了苏州市旅游区特征。从游客客源角度看,苏州市有两类旅游区,古

表3 旅游区网络的邻接矩阵  
Table 3 Adjacency matrix of tourism districts network

旅游区	古城	古镇	金鸡湖	太湖	沿长江	高新区	阳澄湖
古城	-	10347	18864	8041	1667	6652	4999
古镇	10347	-	3343	2097	652	1055	2847
金鸡湖	18864	3343	-	3803	948	2840	4024
太湖	8041	2097	3803	-	452	2252	1731
沿长江	1667	652	948	452	-	292	970
高新区	6652	1055	2840	2252	292	-	1106
阳澄湖	4999	2847	4024	1731	970	1106	-

表4 旅游区网络结构特征

Table 4 Network structure features of tourism districts

旅游区	度中心度	核心-边缘属性
古城	50570	核心
古镇	20341	核心
金鸡湖	33822	核心
太湖	18376	核心
沿长江	4981	边缘
高新区	14197	边缘
阳澄湖	15677	边缘

城和古镇是苏州市的核心,其游客客源地分散且距离远;其他旅游区的客源地则相对集中且距离近。从游客活动角度看,处于苏州城区及其附近的旅游区旅游人次较多,其中金鸡湖重游率较高,古镇旅游区的重游率则较低;高新区属于明显的依附型旅游区,其他旅游区则属于混合型。从网络结构角度看,古城、金鸡湖、古镇和太湖旅游区是苏州市的核心旅游区,苏州旅游区已形成多核的网络结构。

本文通过对苏州市的实证研究,证明基于社交媒体大数据对旅游区进行特征研究的方法是有效的。但是,社交媒体大数据普遍存在年龄、性别和区域等偏好,在基于此类数据进行旅游区特征分析时,需要进一步综合其他数据源进行评估和验证。

参考文献

[1] 郑嬿婷,陆林,章锦河,等. 近十年国外城市旅游研究进展. 经济地理, 2006, 26(4): 686-692

[2] Smith S L J. Dictionary of concepts in recreation and leisure studies. Westport: Greenwood Publishing Group, 1990

[3] Pearce D G. Tourism today: a geographical analysis. Harlow: Longman Scientific & Technical, 1995

[4] Law C M. Urban tourism: attracting visitors to large cities. London: Mansell Publishing, 1993

[5] Mckercher B. Segment transformation in urban tourism. Tourism Management, 2008, 29(6): 1215-1225

[6] 杨新军,马晓龙. 大西安旅游圈: 国内旅游客源空间分析与构建. 地理研究, 2004, 23(5): 695-704

[7] 万绪才,丁登山,马永立,等. 旅游客源市场结构分析——以南京市为例. 人文地理, 1998(3): 74-79

[8] Mckercher B, Gigi L. Movement patterns of tourists within a destination. Tourism Geographies, 2008, 10(3): 355-374

[9] 张骏,古风,卢凤萍. 城市旅游空间行为路径分析及优化——以南京市为例. 地理与地理信息科学, 2011, 27(1): 85-89

[10] Shoval N, Raveh A. Categorization of tourist attractions and the modeling of tourist cities: based on the co-plot method of multivariate analysis. Tourism Management, 2004, 25(6): 741-750

[11] Modsching M, Kramer R, Hagen K T, et al. Using location-based tracking data to analyze the movements of city tourists. Information Technology & Tourism, 2008, 10(1): 31-42

[12] Hsinyu S. Network characteristics of drive tourism destinations: an application of network analysis in tourism. Tourism Management, 2006, 27(5): 1029-1039

[13] 杨兴柱,顾朝林,王群. 南京市旅游流网络结构构建. 地理学报, 2007, 62(6): 609-620

[14] Lu W, Stepchenkova S. User-generated content as a research mode in tourism and hospitality applications: topics, methods, and software. Journal of Hospitality Marketing & Management, 2015, 24(2): 119-154

[15] 王茜雅,张建新,刘培学,等. 大数据视角下城市客源市场空间结构动态研究——以南京市为例. 资源开发与市场, 2018, 34(1): 77-82

[16] 贺小荣,李宗幸,李启明,等. 基于数字足迹的风景名胜区旅游者时空结构特征研究——以赴张家界景区的旅游者为例. 湖南师范大学自然科学学报, 2018, 41(1): 11-17

[17] Girardin F, Calabrese F, Fiore F D, et al. Digital footprinting: uncovering tourists with user-generated content. IEEE Pervasive Computing, 2008, 7(4): 36-43

[18] Ardizzone E, Miceli F D, La Cascia M, et al. Extracting touristic information from online image collections // Eighth International Conference on Signal Image Technology and Internet Based Systems. Naples: IEEE, 2013: 482-488

[19] Yang L, Wu L, Liu Y, et al. Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from Flickr. ISPRS International Journal of Geo-Information, 2017, 6(11): doi: 10.3390/ijgi6110345

[20] Xi Y L, Wang F, Wu B, et al. A social network analysis of overseas tourist movement patterns in Beijing: the impact of the Olympic games. International Journal of Tourism Research, 2012, 14(5): 469-484

[21] 王娟,胡静,贾焱焱,等. 城市旅游流的网络结构特征及流动方式——以武汉自助游为例. 经济地理, 2016, 36(6): 176-184

- [22] Tang J, Li J. Spatial network of urban tourist flow in Xi'an based on microblog big data. *Journal of China Tourism Research*, 2016, 12(1): 1-19
- [23] 张妍妍, 李君轶, 杨敏. 基于旅游数字足迹的西安旅游流网络结构研究. *人文地理*, 2014(4): 111-118
- [24] Dredge D. Destination place planning and design. *Annals of Tourism Research*, 1999, 26(4): 772-791
- [25] 厉新建, 张凌云, 崔莉. 全域旅游: 建设世界一流旅游目的地的理念创新——以北京为例. *人文地理*, 2013(3): 130-134
- [26] 卞显红. 城市旅游空间结构研究. *地理与地理信息科学*, 2003, 19(1): 105-108
- [27] Ester M, Kriegel H P, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise // *International Conference Knowledge Discovery and Data Mining*. Portland, 1996: 226-231
- [28] Powers D M. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2011, 2(1): 37-63
- [29] Shao H, Zhang Y, Li W. Extraction and analysis of city's tourism districts based on social media data. *Computers Environment & Urban Systems*, 2017, 65: 66-78
- [30] Pearce D G. Tourist districts in Paris: structure and functions. *Tourism Management*, 1998, 19(1): 49-65
- [31] Girvan M, Newman M E J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 2002, 99(12): 7821-7826
- [32] Clauset A, Newman M E, Moore C. Finding community structure in very large networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2004, 70(2): 066111
- [33] 保继刚, 郑海燕, 戴光全. 桂林国内客源市场的空间结构演变. *地理学报*, 2002, 57(1): 96-106
- [34] Smith S L J. *Tourism analysis: a handbook*. Harlow: Longman, 1995: 209-221
- [35] 保继刚, 楚义芳. *旅游地理学(修订版)*. 北京: 高等教育出版社, 1999: 16-23, 54-55
- [36] 赵荣, 王斌, 张结魁. 西安市国内游客旅游行为研究. *经济地理*, 2001, 32(增刊 1): 429-432
- [37] 黄潇婷. 基于时间地理学的景区旅游者时空行为模式研究——以北京颐和园为例. *旅游学刊*, 2009, 24(6): 82-87
- [38] Lu X, Wang C, Yang J M, et al. Photo2Trip: generating travel routes from geo-tagged photos for trip planning // *ACM International Conference on Multimedia*. Firenze: ACM, 2010: 143-152
- [39] 车自力, 杜忠潮. 森林公园游客结构及旅游行为特征比较研究——陕西省延安、太白山、褒河森林公园实证分析. *西北林学院学报*, 2008, 23(3): 211-215
- [40] 翁莉. 上海城郊景区旅游者重游行为分析. *旅游科学*, 2005, 19(2): 33-37
- [41] Scott J, Carrington P J. *The SAGE handbook of social network analysis*. London: SAGE Publications, 2011
- [42] Borgatti S P, Everett M G. A graph-theoretic perspective on centrality. *Social Networks*, 2006, 28(4): 466-484
- [43] Borgatti S P, Everett M G. Models of core/periphery structures. *Social Networks*, 2000, 21(4): 375-395
- [44] Veld D I, Lelyveld I V. Finding the core: network structure in interbank markets. *Journal of Banking & Finance*, 2014, 49: 27-40
- [45] 刘法建, 章锦河, 陈冬冬. 社会网络分析在旅游研究中的应用. *旅游论坛*, 2009, 2(2): 172-177