

# 基于社交网络签到数据的城市空间 相互作用和节点吸引力研究

贺泽亚<sup>1</sup> 吴必虎<sup>1,†</sup> 刘瑜<sup>2</sup>

1. 北京大学城市与环境学院旅游研究与规划中心, 北京 100871; 2. 北京大学遥感与地理信息系统研究所, 北京 100871;

† 通信作者, E-mail: tigerwu@urban.pku.edu.cn

**摘要** 基于社交网络大数据的研究视角, 选取全国348个城市之间一年的跨城市社交媒体地理位置签到数据, 采取优化粒子群(PSO)方法, 使用引力模型, 逆向推导该系统中空间相互作用的距离衰减函数以及各城市的节点吸引力。通过引入经济发展水平、产业结构、人口规模和结构、旅游竞争力、教育水平5个方面的12项变量, 经过因子分析和回归分析, 探究这些变量对节点吸引力的影响作用。结果表明, 社交媒体签到系统中的交互流量符合距离衰减的幂律函数, 与其他交互系统相比, 其距离衰减系数偏小, 说明全国尺度的城市间人的移动受距离影响不明显。对全国城市节点吸引力及其排名的进一步分析发现, 与旅游竞争力、城市发展成熟度、人口规模这几个维度相关的因子对社交媒体签到系统中的城市节点吸引力有显著的影响。研究结论将为更好地理解人类签到和移动行为, 为进一步了解复杂网络系统中节点吸引力的内涵做出一定的理论和实际贡献。

**关键词** 地理位置签到; 空间相互作用; 引力模型; 节点吸引力; 优化粒子群法(PSO)

**中图分类号** P208

## Study of Spatial Interaction and Nodal Attractions of Municipal Cities in China from Social Media Check-in Data

HE Zeya<sup>1</sup>, WU Bihu<sup>1,†</sup>, LIU Yu<sup>2</sup>

1. The Center for Recreation and Tourism Research, College of Urban and Environmental Sciences, Peking University, Beijing 100871;

2. Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871;

† Corresponding author, E-mail: tigerwu@urban.pku.edu.cn

**Abstract** To investigate the spatial interaction effect and nodal attractions of cities, a set of inter-city social network location-based check-in data with a time span of one year among 348 municipal cities in China is examined with a PSO (Particle Swarm Optimization) method and the gravity model. Twelve variables related with economic development, industrial structure, population scale and structure, tourism competitiveness and educational level are introduced to further investigate their influences on nodal attractions of cities. The results indicate a distance decay effect which is relatively weaker than in other systems, suggesting that human mobility at the regional level is less sensitive to the change in geographic distance. A close examination of the nodal attractions suggests variables related to the cities' tourism competitiveness, maturity of development and population scale significantly influence the value of nodal attractions. This article will serve as a stepping-stone for a better future understanding of human travel pattern, check-in behaviors and the real meaning of nodal attractions in some complicated networks.

**Key words** location-based; check-in; spatial interaction; gravity model; nodal attraction; PSO

Harris 等<sup>[1]</sup>在《The nature of cities (城市本质)》一文中指明,城市本质是城市的内部组织关系。Taylor 等<sup>[2]</sup>将城市之间的关系称为“城市的第二本质”。20 世纪 90 年代以来,随着全球化和信息化的加速,任何区域或城市的发展不再是在一个封闭的系统内进行。从全球、国家或区域层面来看,城市的发展已经跨越自身界限,通过交通、信息等各种网络将各个城市及其设施紧密地联系在一起。城市之间的关系变得比以往更加重要。城市网络中的空间相互作用通常以各种“要素流”的形式呈现。国际上早期的研究主要通过通讯、交通等基础设施的流量(如航班数量、港口吞吐量、公路车流量、铁路交通量等)测度来研究城市网络交互<sup>[3-5]</sup>。20 世纪 80 年代后,以互联网为代表的信息技术在全球迅速发展,基于电子邮件<sup>[6]</sup>、互联网传输量<sup>[7]</sup>、网络宽带流量<sup>[8]</sup>、科学知识创新合作数量<sup>[9]</sup>、专利创新合作数量<sup>[10]</sup>等途径的城市空间相互作用的研究不断涌现,重构了全球城市的空间网络形态<sup>[11]</sup>。由于虚拟的信息网络、实体的地理空间与人类活动之间的相互影响越来越深远<sup>[12-13]</sup>,虚拟网络与实体网络空间的关系已成为地理学研究的重点课题<sup>[14]</sup>。其中,由“地理空间”与“网络空间”交融而成的“地理网络空间”(Geocyberspace)<sup>[15]</sup>以及网络信息地理学(Cybergeography)成为研究热点<sup>[16]</sup>。

通过对大量相关文献进行总结可以发现,区别于其他拓扑网络,作为空间网络之一的城市网络具有强烈的空间属性,其空间嵌入(spatially embedded)特征对网络形态和结构有重要的影响,节点空间相互作用的可能性或强度随着距离的增加呈规律性的衰减。这种城市之间的距离可以用真实的空间距离或交通时间、交通成本等其他抽象距离来表示<sup>[17]</sup>。不同系统中城市空间相互作用往往具有不同的节点吸引力属性和距离衰减属性,探究某系统中城市空间网络的节点属性和空间相互作用的距离衰减效应,对理解城市在这个系统中体现的规模位序和人类群体移动特征(collective human mobility patterns)有重要的作用。

随着智能设备的逐渐发展,带有时空信息的数据在过去几年中得到爆炸式的增长。一些带有签到(地理位置标记)功能的大型社交平台,例如 Facebook, Twitter, Flickr, Foursquare, 微博和街旁等,各自拥有数以亿计的活动用户,其用户生成的带有特定位置属性的信息被 Goodchild<sup>[18]</sup>定义为

volunteered geographical information (VGI), 并且已经成为一种非常重要的大数据类型<sup>[19]</sup>。这些社交平台上的签到数据,既有虚拟网络中的社交属性,又对应着实际的地理空间。这些数据为社会经济环境、城市结构<sup>[20]</sup>、旅游管理<sup>[21]</sup>、智能交通<sup>[22]</sup>等方面的研究提供了更微观的素材<sup>[23]</sup>,补充了传统研究中在探究空间相互作用方面数据和方法的不足<sup>[24]</sup>。

本文选取时间跨度为一年的社交媒体用户生成的海量跨城市地理签到大数据进行处理和计算。利用优化粒子群的计算方法,使用引力模型逆向推导,得到城市空间相互作用的距离衰减函数,解释其揭示的人类移动特征。在此基础上,计算该系统中全国 348 个城市节点的吸引力值和排序,并进行空间呈现。通过因子分析和多元回归,研究复杂网络系统中节点吸引力的影响因素,从一个新颖的视角加深对网络信息地理学和地理网络空间的认知。

## 1 研究方法

### 1.1 研究模型

随着距离的增长,空间相互作用的强度不断减弱,呈现距离衰减效应。在宏观和集聚层面,在其他变量(如人口)相对稳定的前提下,距离衰减体现为地理要素(如城市)间相互作用各类流量(如客流量)与距离的负相关关系。这种流量与距离的关系可用以下公式表示:

$$G_{ij} = k \frac{P_i P_j}{F(d_{ij})}$$

其中,  $k$  为常数系数;  $G_{ij}$  表示  $i$  与  $j$  两地的空间相互作用,如交通流、信息流等;  $P_i$  和  $P_j$  为两地区的节点吸引力(nodal attraction);  $d_{ij}$  为  $i$  与  $j$  两地之间的距离相关因素,可以表现为距离、旅行成本、旅行时间等;  $F(d_{ij})$  为距离衰减函数,通常有幂律、指数和高斯函数等形式。

两种被广泛采用的交互模型为引力模型和最大熵模型<sup>[25]</sup>。由于最大熵模型的距离衰减函数为负指数形式:

$$F(d_{ij}) = e^{-\alpha d_{ij}} \quad (\alpha > 0),$$

在其他条件不变的情况下,最大熵模型拟合的空间相互作用流量会随着距离增长过快地衰减,因而适用于较短程的空间尺度和封闭的系统。由于引力模

型的空间衰减函数为幂律形式:

$$F(d_{ij}) = d_{ij}^{-\beta} \quad (\beta > 0),$$

因此与负指数形式相比较,其相互作用流量随着距离的增长衰减较慢,因而适用于大尺度的空间相互作用系统<sup>[26]</sup>。引力模型中的指数 $\beta$ 为距离衰减的摩擦系数,反映距离对流量 $G_{ij}$ 的影响程度,其值越大,距离对空间相互作用的影响越大。通常情况下,实体的交通、物流等流量的距离衰减系数大于虚拟的信息、网络流量等。在全国尺度和开放系统的前提下,较适用于城市间交互作用的引力模型的基本形式为

$$G_{ij} = k \frac{P_i P_j}{d_{ij}^\beta} \quad (1)$$

对引力模型较好的使用方法,是根据交互量和距离,逆向推导出距离衰减函数和节点吸引力<sup>[19]</sup>,有解析法和模拟法两大类。其中,逆向解析可以运用线性规划法(linear programming, LP)<sup>[27]</sup>和目标优化法<sup>[28]</sup>,通过引入误差项,并利用建立约束条件优先级对误差进行优化,解析出距离衰减系数以及节点吸引力对数值。另一种逆向解析的方式是对地理对象间交互等式运用代数法(algebraic method, AM)<sup>[29]</sup>或简单代数法(simplified algebraic method, SAM)<sup>[30]</sup>进行一定的代数运算,解析不同距离衰减系数值下节点的吸引力值。

Xiao 等<sup>[31]</sup>首次将优化粒子群算法引入重力模型的逆向求解中,把估算节点吸引力的问题转化为优化问题。在有  $n$  个地理要素的空间系统中,将节点吸引力 $P$ 的每组解视为一个粒子,假设共有  $m$  个粒子,每个粒子  $X_i$  可表示为有  $n$  个维度的向量, $X_i$  在不同维度上的值代表不同地理要素的节点吸引力 $P$ ,

$$X_i = \{P_1, P_2, P_3, \dots, P_n\}。$$

粒子  $i$  在初始时都有一个设定的位置  $X_i^0$  和速度  $V_i^0$ ,在经过反复迭代计算后的  $t$  时刻,粒子  $i$  的位置和速度分别为  $X_i^t$  和  $V_i^t$ 。粒子  $i$  的迭代过程中,在所有的解  $\{X_i^1, X_i^2, X_i^3, \dots, X_i^t\}$  中会有一个拟合最优的解  $Pbest_i$ ,而在所有粒子的最优解  $\{Pbest_1, Pbest_2, Pbest_3, \dots, Pbest_m\}$  中也会出现一个全局拟合最优的解  $Gbest$ 。每次迭代后,  $Pbest_i$  和  $Gbest$  也会从历次计算的结果中选择最优来更新其结果。每个粒子从

第  $t$  时刻到  $t+1$  时刻,会根据式(2)和(3)来迭代更新其位置  $X_i^t$  和速度  $V_i^t$  至  $X_i^{t+1}$  和  $V_i^{t+1}$ 。

$$V_i^{t+1} = wV_i^t + c_1r_1(Pbest_i^t - X_i^t) + c_2r_2(Gbest^t - X_i^t), \quad (2)$$

$$X_i^{t+1} = X_i^t + V_i^{t+1}, \quad (3)$$

式(2)中,  $w$  为惯性权重;  $c_1$  和  $c_2$  为加速系数,通常取相等的值来平衡  $Pbest$  和  $Gbest$  的影响效果,  $r_1$  和  $r_2$  为 0~1 之间的随机数<sup>[31]</sup>。根据 Van den Bergh 等<sup>[32]</sup>的研究,  $w$  取 0.73,  $c_1$  和  $c_2$  取 1.50 时,粒子群模拟运算总能导致收敛的轨迹。本文也据此设定惯性权重值和加速系数值。粒子的拟合优度(goodness of fit, GOF)以基于粒子  $i$  的解正向推算出来的流量预测值和真实的流量值拟合的  $R^2$  来衡量。根据距离衰减系数  $\beta$  的经验取值范围 0~2,通过每次以 0.1 为幅度输入不同的  $\beta$  值,经过多次的调试计算,选择流量的观测值和预测值拟合的 GOF 最高时的  $\beta$  值。

Xiao 等<sup>[31]</sup>也在不同网络密度和不同数量地理对象的条件下,对 LP, SAM 和 PSO 三种方法进行比较,发现随着交互网络密度的下降和地理对象的增加,用 LP 和 SAM 方法解析出的节点吸引力值正向推断的交互流量预测值与真实交互流量值之间的拟合优度 GOF 呈不同程度的下降趋势。由于需要将缺省的流量值用非零常数值代替, SAM 算法在稀疏网络的条件下表现较差。LP 算法在稀疏网络和多地理对象的条件下仍具有一定的稳定性,但有明显的下降趋势。PSO 算法在各种条件下计算出的结果都有较好的稳定性, GOF 一直稳定在 0.9 上下<sup>[31]</sup>。综合来看, PSO 模拟算法是一种较好的应对较大规模和稀疏网络的逆向求解方法。

## 1.2 研究数据

本文选取 Liu 等<sup>[19]</sup>在 2014 年的研究中使用过的数据,进行进一步分析。该数据搜集于中国著名的基于地理位置的社交网站(Locational-Based Social Network Service, LBSNS)街旁,选取从 2011 年 9 月至 2012 年 9 月间近 52.1 万活跃注册用户产生的地理位置数据。对该数据经过阈值值的处理,筛选掉其中虚假签到的记录后,留下约 2350 万条签到记录。这些签到记录所在的地理位置分布在全国 370 个城市(包括 4 个直辖市, 332 个地级行政单元, 13

个县级行政单元, 香港、澳门以及台湾的 19 个城市), 西藏自治区的日喀则市和青海省的海南藏族自治州在这一时间段没有地理位置签到记录。在这些数据中, 有 45.6% 的用户曾经有过在两个及两个以上城市间签到的记录。这些跨城市的签到记录以交互的形式统计(如 A, B 两地之间的交互量为 A 地到 B 地与 B 地到 A 地两个方向起点至终点流量的叠加)。城市被抽象为空间点坐标, 城市间的距离(km)用其点坐标间的球面距离表示。城市间的空间相互作用流量和距离可以表示为矩阵的形式, 表 1 和 2 选取其中的 10 个城市进行矩阵展示。

空间网络的密度  $D$  可以用下式所表示:

$$D = 2E / [V(V - 1)], \quad (4)$$

其中,  $E$  代表网络中的连接数,  $V$  代表交互节点数量,  $D$  值范围在  $[0, 1]$  之间。  $D$  越趋近 1, 表示网络系统

的密度越大, 越倾向于均质的网络;  $D$  越趋近 0, 网络越稀疏, 系统的中心性越突出。根据样本数据计算的交互网络密度为 0.22, 属于较稀疏的网络。依据 Xiao 等<sup>[31]</sup>在不同网络密度和不同节点数量条件下进行的算法比较, 本文拟采取 PSO 算法, 通过不同的距离衰减系数来进行重力模型的逆向拟合。

## 2 计算结果

### 2.1 距离衰减系数

根据距离衰减系数的经验取值范围 0~2, 本文从 0.1 开始, 以每次 0.1 的增幅来输入  $\beta$  值, 直至 2.0, 通过寻找最佳的 GOF 值来确定合适的  $\beta$  值。从图 1 中  $\beta$  输入值与拟合的 GOF 之间关系可以发现, 随着  $\beta$  值增长, GOF 值呈类似抛物线的形状。当  $\beta$  取值约为 0.8 时, GOF 达到最高值 0.99。这一衰减系数与文献中其他类型数据的衰减系数(如钱币数据

表 1 城市间交互量矩阵  
Table 1 Matrix of cities' spatial interactions

城市	北京	苏州	杭州	南京	广州	香港	嘉兴	无锡	厦门
上海	70261	137847	95936	36741	18033	35742	43273	36550	31652
北京		2558	6050	5308	6621	6080	714	1258	3578
苏州			2928	4574	357	359	1657	7664	353
杭州				1901	1092	717	11088	1204	1050
南京					628	424	411	3343	731
广州						3348	84	182	1198
香港							87	166	388
嘉兴								454	127
无锡									177

表 2 城市间距离矩阵  
Table 2 Matrix of cities' distances

城市	上海	北京	苏州	杭州	南京	广州	香港	嘉兴	无锡
北京	1067								
苏州	85	1028							
杭州	165	1123	121						
南京	270	898	189	237					
广州	1212	1889	1161	1048	1132				
香港	1224	1959	1179	1063	1170	119			
嘉兴	86	1091	63	79	236	1127	1140		
无锡	114	999	34	136	157	1161	1183	92	
厦门	821	1722	797	676	846	513	467	745	809

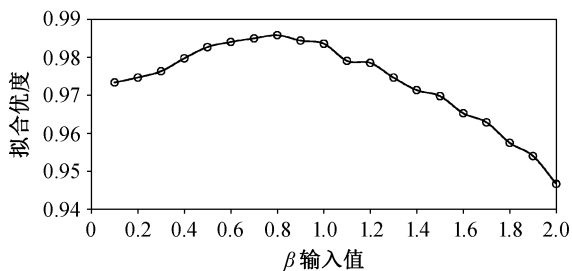


图 1 以 0.1 为间隔的  $\beta$  输入值下返回的 GOF 值  
Fig. 1 Returned GOF of entered  $\beta$  at 0.1 intervals

( $\beta=1.59\pm 0.02$ )<sup>[33]</sup>、出租车数据( $\beta=2.5$ )<sup>[34]</sup>、城市内签到数据( $\beta=1.5$ )<sup>[35]</sup>、航空数据( $\beta=0.87$ )<sup>[31]</sup>等)相比偏小,表明社交媒体签到数据代表的城市空间相互作用受距离的影响更小。这可能是由于与其他人群相比,使用社交媒体签到功能的人群受时空限制较少,也可能是由于社交媒体用户记录自己中长距离出行行为的概率更大。

### 2.2 节点吸引力

图 2 中利用几何分级法,将 PSO 算法模拟的全国城市节点吸引力划分 11 个等级,用不同大小的圆点表示,空白的地区表示缺少数据值。在排除流量值空缺的城市对后,用模拟的城市节点吸引力正

向推测的流量值与实际流量值之间的拟合优度 GOF 达到 0.99,说明引力模型能够较好地反映城市空间相互作用的模式。可以发现,较发达的地区(如 4 个直辖市、东南沿海地区及内陆的一些旅游城市)呈现较高的节点吸引力。

### 2.3 节点吸引力的影响因素

为了更好地理解图 2 反映的节点吸引力特征以及内涵,有必要对其影响因素进行更深入的分析。Shen<sup>[29]</sup>指出,求解出节点吸引力后,可以将节点吸引力作为因变量,建立线性回归方程,寻找可以解释节点吸引力内涵的其他自变量。Xiao 等<sup>[31]</sup>基于航空网络交互数据,发现城市人口规模、收入水平、第三产业比重、旅游营收等因素对城市节点吸引力有显著影响。总结有关节点吸引力分析的文献可知,不同系统中的节点吸引力有着不同类型的影响因素。通常,涉及的因素有经济发展水平、产业比重、交通发达水平、人口规模和人口教育水平(或其代表的信息技术水平)等,使用的函数为多元线性形式和幂函数乘积形式。虽然文献中未出现对社交网络签到系统中节点吸引力的研究,但是 Liu 等<sup>[19]</sup>在对距离衰减函数和人类移动特征的研究中

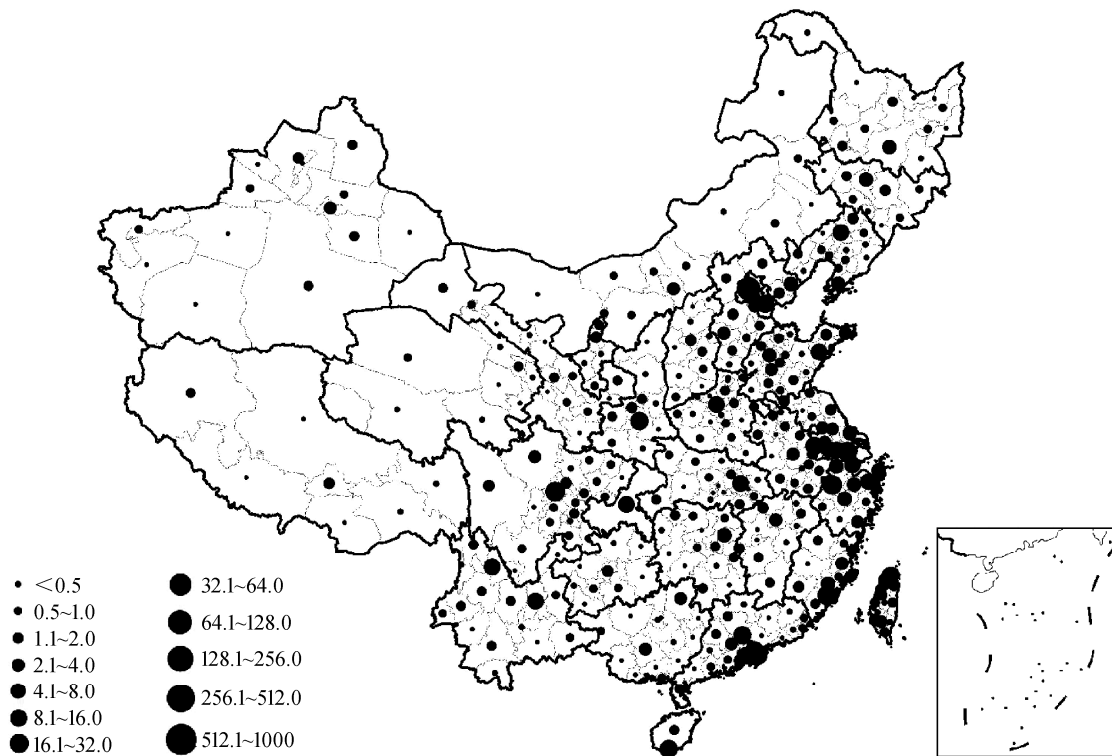


图 2 全国城市节点吸引力  
Fig. 2 Nodal attractions of municipal cities in China

提到, 社交媒体签到系统中的节点吸引力可能与城市人口规模、旅游吸引力和信息技术水平存在一定关系。节点吸引力  $P$  通常表示成一系列有关变量幂函数的乘积形式, 如  $P_i = A_i^\alpha B_i^\omega$  [36]。为便于回归分析, 需将该等式两边取对数, 变形为  $\text{Ln}P_i = \alpha \text{Ln}A_i + \omega \text{Ln}B_i$ 。本文基于文献总结, 选取经济发展水平、旅游竞争力、产业人口比重、人口规模和结构和教育水平 5 个方面的 12 项变量(表 3)来探究影响节点吸引力的因素。为了方便计算, 对表 3 中的变量都做了对数化处理。由于这一阶段选取的变量仅含 336 个城市的完整信息(在原来的样本中去除了香港、澳门、台湾以及海南省部分城市的信息), 因此下面分析的样本数也从 348 个缩减为 336 个。

检验发现, 选取的 12 个变量 KMO 值为 0.760, 较适宜做因子分析。用主成分提取法提取公因子, 通过分析碎石图(图 3)以及方差贡献率和累计方差贡献率(图 4)发现, 若选取特征根大于 1 的因子解, 只能提取 3 个因子, 解释原始变量的 82.18%; 若选择 5 个因子, 可以解释原始变量的 90.57%, 虽然后两个因子的方差贡献率较低, 但保留它们可以还原较多的原始变量信息; 若选取 6 个因子, 虽然可以更大程度(93.25%)地解释原始变量, 但第 6 个因子

的因子载荷普遍较低, 缺乏解释力。因此, 本文提取 5 个因子进行分析。

表 4 显示经过方差极大法旋转后的 5 个因子的因子载荷分布。可以发现, 因子 1 是与高等文化教育普及率、第三产业人口比重、经济发展和消费水平和劳动人口比重高度相关的因素, 将其命名为综合发展成熟度。因子 2 代表与旅游竞争力相关的因素, 因子 3 代表与基础教育水平相关的因素, 因子 4 代表与工业化水平相关的因素, 因子 5 代表与总体人口规模相关的因素。

以提取的 5 个因子为自变量, 对因变量进行回归分析, 结果如表 5 所示。由于回归结果不显著, 因子 3 (基础教育水平)被剔除。回归的调整系数  $R^2$  为 0.515, 说明模型的整体拟合效果适中, 选取的变量和提取的因子能够解释一部分节点吸引力的内涵, 但由于系统的复杂性, 有一部分节点吸引力的内涵未能解释。 $p$  值和  $t$  值反映选定的因子总体较显著的影响作用(回归系数显著非零)。调整  $R^2$  表明模型的拟合优度一般, 导致这一结果出现的可能原因是网络中局部的交互流量过于稀疏, 导致部分节点的吸引力的估算值与真实值差别较大。我们选取节点吸引力排前 200 位的城市进行回归计算, 拟合

表 3 考察城市节点吸引力涉及的变量  
Table 3 Related variables in investigating nodal attractions of cities

变量	解释
因变量	城市节点吸引力
	根据 PSO 算法模拟出的城市节点吸引力 $P$ 的值
	(a) 人均 GDP
	2011 年各城市常住人口人均 GDP
	(b) 人均消费水平
	2011 年各城市常住人口人均消费水平
	(c) 常住人口规模
	2011 年各城市常住人口
	(d) 第二产业人口比重
	2010 年各城市第二产业人口占常住人口比例
	(e) 第三产业人口比重
	2010 年各城市第三产业人口占常住人口比例
	(f) 劳动人口比重(15~59 岁)
	2010 年各城市劳动人口占常住人口比例
自变量	(g) 成人受教育率
	2010 年各城市成人受教育率
	(h) 中等教育接受率
	2010 年各城市中等教育接受率
	(i) 高等教育接受率
	2010 年各城市高等教育接受率
	(j) 观光旅游竞争力
	2011 年各城市观光旅游竞争力
	(k) 休闲旅游竞争力
	2011 年各城市休闲旅游竞争力
	(l) 综合旅游竞争力
	2011 年各城市综合旅游竞争力

说明: (a) 和 (b) 数据来源于各省市 2012 年统计年鉴、各城市 2011 年统计公报、2010 年第六次全国人口普查分县数据; (c) 数据来源于各省市 2012 年统计年鉴、各城市 2011 统计公报、GDP 总量和 2010 年第六次全国人口普查分县数据中常住人口总量的计算值; (d)~(i) 数据来源于中国 2010 年第六次人口普查分县数据; (j)~(l) 数据来源于 2012 年大地风景咨询集团发布的行业研究报告《中国城市旅游投资竞争力研究报告》。

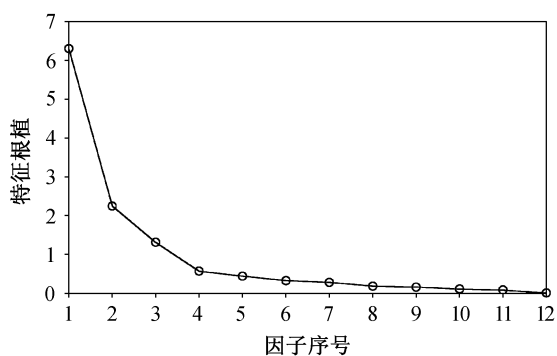


图 3 碎石图  
Fig. 3 Scree plot

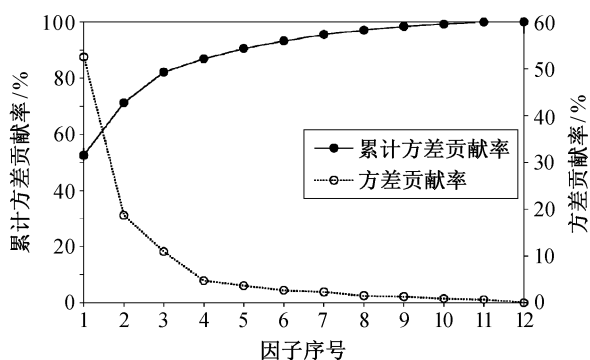


图 4 方差贡献率与累计方差贡献率  
Fig. 4 Variance and cumulative variance

优度调整  $R^2$  提升到 0.617。这说明一些排名靠后的城市节点吸引力模拟值存在误差。值得注意的是,这部分模拟误差不是由算法导致的,而是由于部分

城市间的相互作用流量未被观测到引起的。部分样本因子分析和回归分析的过程中同样提取 5 个因子,回归结果的显著性水平和回归系数与总体样本的计算结果差别不大。

对因子回归的  $p$  值和标准化回归系数进行分析后发现,与基础教育水平相关的因子 3 回归结果不够显著,因此被剔除;与旅游竞争力、综合发展成熟度以及总体人口规模相关的因子 2、因子 1 和因子 5 对节点吸引力有显著的正向影响;与城市工业化水平相关的因子 4 的显著性水平较低,且回归系数的绝对值比前三者低很多。这表明,在社交媒体签到系统中,旅游竞争力因素对节点吸引力的影响最高,城市综合发展成熟度因素次之,人口规模因素的影响比前两者低很多,工业化水平的影响最不显著。可能的解释为:使用社交媒体签到功能的用户可能更愿意前往旅游竞争力高的城市,或更愿意在这些地方记录他们的位置信息;由于具备较好的文化教育水平、服务业水平和经济发展水平,发展成熟度较高的城市可能拥有更多使用社交媒体签到功能记录其出行的用户,或吸引其他地区的用户更愿意前往并记录他们到该地的行为;拥有较大人口基数的城市在一定程度上拥有更大的用户量以及与其他城市间更大的交互流量。这表明社交媒体签到系统与其他空间交互系统的节点吸引力内涵有明显不同,旅游竞争力已经取代人口规模成为主要的影响因素<sup>[37]</sup>。

表 4 旋转后的因子载荷矩阵  
Table 4 Rotated factor matrix

变量	因子				
	1	2	3	4	5
(a) 人均 GDP	0.764	0.143	0.186	0.485	0.012
(b) 人均消费水平	0.689	0.306	0.078	0.522	0.240
(c) 观光旅游竞争力	0.352	0.876	0.153	-0.007	0.147
(d) 休闲旅游竞争力	0.013	0.884	0.190	0.289	0.135
(e) 综合旅游竞争力	0.235	0.936	0.152	0.132	0.147
(f) 常住人口规模	-0.123	0.479	0.276	0.071	0.797
(g) 第二产业人口比重	0.315	0.350	0.473	0.660	0.050
(h) 第三产业人口比重	0.790	0.179	0.180	0.340	-0.188
(i) 劳动人口比重(15~59 岁)	0.873	0.018	0.257	-0.080	0.044
(j) 成人受教育率	0.215	0.170	0.906	0.101	0.083
(k) 中等教育接受率	0.190	0.184	0.904	0.142	0.143
(l) 高等教育接受率	0.896	0.245	0.089	0.047	-0.081

表 5 因子多元回归结果  
Table 5 Results of factor multiple regression

模型	$R^2$	调整 $R^2$	因子	标准化回归系数	$t$	$p$
1	0.241	0.239	常数项		6.436	0.000
			因子 2 (旅游竞争力)	0.491	10.261	0.000
2	0.455	0.452	常数项		7.585	0.000
			因子 2 (旅游竞争力)	0.491	12.094	0.000
			因子 1 (发展成熟度)	0.463	11.393	0.000
3	0.511	0.507	常数项		7.993	0.000
			因子 2 (旅游竞争力)	0.491	12.774	0.000
			因子 1 (发展成熟度)	0.463	12.006	0.000
			因子 5 (总体人口规模)	0.236	6.120	0.000
4	0.521	0.515	常数项		8.061	0.000
			因子 2 (旅游竞争力)	0.491	12.852	0.000
			因子 1 (发展成熟度)	0.463	12.107	0.000
			因子 5 (总体人口规模)	0.236	6.172	0.000
			因子 4 (工业化水平)	0.098	2.569	0.011

### 3 结论

本文基于地理学经典引力模型,对社交网络签到数据这一用户生成的大数据进行分析。区别于以往研究中使用引力模型时对已知项和未知项的假设和求解,本研究选取目前较为先进的优化粒子群(PSO)方法,以空间相互作用流量为已知项,不预先假设空间衰减系数 $\beta$ ,逆向求解出节点吸引力和空间衰减系数的最优值。然后,根据城市相关统计数据,对该节点吸引力值进行因子分析和回归分析,进一步解释社交媒体签到系统中城市节点吸引力的内涵,结论如下。

1) 在对空间衰减系数 $\beta$ 的推导过程中发现,在经验取值范围(0~2)内,随着输入 $\beta$ 输入值的上升,预测的相互作用流量值与观测的相互作用流量值间的拟合优度有着先上升后下降的趋势。在本文样本数据的计算中,当 $\beta$ 取值为0.8时,拟合优度达到最高,此时的 $\beta$ 值能较好地解释全局空间衰减效应。根据以往学者的研究<sup>[31,33-35]</sup>,这一 $\beta$ 值明显低于城市内的空间衰减系数以及基于其他数据(如航空、钱币、其他签到系统)的距离衰减系数,表明社交媒体签到系统反映的人类活动更少地受距离衰减效应影响。

2) 通过总结文献和分析计算出的节点吸引力结

果,找出经济发展水平、旅游竞争力、产业人口比重、人口规模和结构、教育水平5个方面的12项变量,对节点吸引力的影响因素进行分析。经过对数变换后,对变量进行因子分析,经过方差极大旋转后,提取与综合发展成熟度、旅游竞争力、基础教育水平、工业化水平、总体人口规模相关的5个因子,解释了逾90%的原始变量。将这5个因子作为自变量对因变量(对数)进行多元回归分析,发现与旅游竞争力、综合发展成熟度、总体人口规模以及工业化水平相关的因素对节点吸引力有较显著的正向影响。

3) 因子回归分析结果进一步显示,旅游竞争力、综合发展成熟度以及总体人口规模3个因子的回归系数和显著性远高于工业化水平这一因子。其中,旅游竞争力因素的回归系数最高,说明其影响程度最大;综合发展成熟度因子的标准化回归系数稍低于旅游综合竞争力;总体人口规模因子的标准化回归系数明显低于前两者,表明与其他系统(如航空系统<sup>[31]</sup>)不同的是,社交媒体签到系统更明显地受旅游竞争力和城市综合发展因素而非人口规模的影响。

4) 因子的回归分析整体上拟合效果一般,调整 $R^2$ 为0.52。这说明可能存在:①由于局部城市观测到的流量过少或缺省,导致模拟出的节点吸引力与

实际规模偏离; ② 地理位置签到系统中的节点吸引力是一个综合了较多变量的函数值, 现有的变量不足以解释其全部内涵。通过选取节点吸引力较大的城市, 进行部分样本回归分析, 模型的拟合优度上升 0.1 左右, 调整  $R^2$  达到 0.62。这表明上述两方面的原因都存在。但是, 在全局和部分样本回归分析中, 旅游竞争力、综合发展成熟度和总体人口规模这 3 个因子的标准化回归系数均十分显著, 表明这些因素的确存在较为明显的影响。总的来说, 节点吸引力可以表达为

$$P_i = A_i^\alpha B_i^\gamma C_i^\delta D_i^\varepsilon \dots$$

的幂律乘积形式, 其中,  $P_i$  表示城市  $i$  的节点吸引力,  $A_i$ ,  $B_i$  和  $C_i$  分别代表与旅游竞争力、城市综合发展成熟度和总体人口规模相关的因素,  $D_i$  以及省略的部分代表可能存在的其他影响因素,  $\alpha$ ,  $\gamma$ ,  $\delta$ ,  $\varepsilon$ , ... 分别代表这些因素的影响程度。

## 4 讨论

基于本文结论可以看出, 社交媒体签到这一结合了虚拟网络与实际空间网络的系统可能存在以下特征。

1) 其相对偏小的距离衰减系数  $\beta$  可能表明, 社交媒体签到数据表征的人类空间移动更少地受距离作用的影响。这可能是由于与其他人群相比, 使用社交媒体签到功能的人群较少受时空限制, 也可能是由于社交媒体用户更大概率地记录了自己的长距离出行行为。更确切的解释需要对签到功能的使用者进行深入的问卷和访谈研究才能得出。

2) 地理位置签到行为可能与部分旅游行为和经济行为之间存在高度重合, 使用社交媒体签到功能的用户可能更愿意前往旅游竞争力高的城市, 或在更愿意这些地方记录他们的位置。发展成熟度也会在较大程度上影响社交网络地理位置签到功能的使用者数量和其签到行为, 发展成熟度较高的城市可能由于其较好的文化教育水平、服务业水平和经济发展水平, 拥有更多使用社交媒体签到功能的用户, 或吸引其他地区的用户更愿意前往并记录他们到该地的行为。

由于传统统计数据的数据、尺度和规模限制, 以往很长时间的的相关研究都受到时间和空间的局限。基于位置的社交网络的兴起以及用户生成的

“大数据”的积累, 带来新的机遇去研究更大尺度、更海量的对象和更微观的个体。本文选取的样本为 2011—2012 年逾 20 万用户生成的跨城市签到记录, 这正是全球签到类系统使用最热门的时期, 用户生成数据的质量较高, 此后, 签到系统的使用热度略有下降。本文中数据的使用揭示了 2011—2012 年间地理空间距离和地理要素对人类签到活动的影响。虽然这一地理位置签到平台已经有可观的用户积累, 但其样本只能代表特定的用户群体, 且有签到行为的用户只是所有用户中的一部分, 所以本文的研究结论有一定的局限性。本研究得出的结论也能反映使用社交媒体签到功能的群体与其他群体之间空间行为的差异性, 亦或能反映用户使用社交媒体签到功能自我报告的群体空间行为与其他类型空间行为的差异。

由于地理位置签到行为产生较晚, 在心理学方面的解释亟待发展。现有的研究揭示出, 因其所在的社交平台, 地理位置签到行为具有强烈的社交属性, 且含“炫耀”成分, 并伴随更高的自我感知价值(perceived value)<sup>[38]</sup>。然而, 现有研究较多地对签到行为的社会性和自我性做解读, 较少从签到地点(反映为网络系统中的节点)的角度理解其在旅行者认知中的价值。本文的结果将为更好地理解人类签到和移动行为, 进一步了解复杂网络系统中节点吸引力的内涵做出一定的理论和实际贡献。如果在此基础上对这部分用户进行人口统计学、心理学及行为学研究, 便可以获得对签到这一类既有社交性又有空间性行为的更深入理解。

此外, 现有的数据格式是将两地之间的空间相互作用流量统计为来往的流量之和, 将所反映的节点吸引力和出行力视为等价。然而, 从旅游的角度出发, 在推拉(push and pull)理论<sup>[39]</sup>的支持下, 节点的吸引力和出行力并不完全等价。式(1)通常会变形为

$$G_{ij} = k \frac{E_i A_j}{d_{ij}^\beta},$$

$E_i$  和  $A_j$  分别代表作为客源地的出游力和作为目的地的吸引力两个方面。如果能够获得更精确的客源地-目的地空间流量数据进行分析, 则可以从旅游方面对社交媒体签到系统中的空间相互作用以及节点属性的内在含义得到更加深刻的认识<sup>[40]</sup>。

## 参考文献

- [1] Harris C D, Ullman E L. The nature of cities. *Annals of the American Academy of Political and Social Science*, 1945, 242(1): 7-17
- [2] Taylor P J, Derudder B. World city network: a global urban analysis. *International Social Science Journal*, 2004, 31(4): 641-642
- [3] Goetz A R. Air passenger transportation and growth in the US urban system, 1950-1987. *Growth and Change*, 1992, 23(2): 217-238
- [4] Taaffe E J. The urban hierarchy: an air passenger definition. *Economic Geography*, 1962, 38(1): 1-14
- [5] Vickerman R W. The regional impacts of Trans-European networks. *The Annals of Regional Science*, 1995, 29(2): 237-254
- [6] Batty M. *Urban information networks: the evolution and planning of computer-communications infrastructure*. Sydney: Halsted Press, 1991
- [7] Mitchelson R L, Wheeler J O. The flow of information in a global economy: the role of the American urban system in 1990. *Annals of the Association of American Geographers*, 1994, 84(1): 87-107
- [8] Townsend A M. Network cities and the global structure of the Internet. *American Behavioral Scientist*, 2001, 44(10): 1697-1716
- [9] Matthiessen C W, Schwarz A W. World cities of scientific knowledge: systems, networks and potential dynamics. an analysis based on bibliometric indicators. *Urban Studies*, 2010, 47(9): 1879-1897
- [10] Sonn J W, Storper M. The increasing importance of geographical proximity in knowledge production: an analysis of US patent citations, 1975-1997. *Environment and Planning A*, 2008, 40: 1020-1039
- [11] Graham S, Marvin S. *Telecommunications and the city: electronic spaces, urban places*. London: Routledge, 1996: 211-236
- [12] 甄峰, 张敏, 刘贤腾. 全球化、信息化对长江三角洲空间结构的影响. *经济地理*, 2004, 24(6): 748-752
- [13] 甄峰, 魏宗财, 杨山, 等. 信息技术对城市居民出行特征的影响研究——以南京为例. *地理研究*, 2009, 28(5): 1307-1317
- [14] 王波, 甄峰, 席广亮, 等. 基于微博用户关系的网络信息地理研究——以新浪微博为例. *地理研究*, 2013, 32(2): 380-391
- [15] H·巴凯斯, 路紫. 从地理空间到地理网络空间的变化趋势. *地理学报*, 2000, 55(1): 104-110
- [16] Dodge M. *Cybergeography*. *Environment and Planning B: Planning and Design*, 2001, 28(1): 1-2
- [17] Bullock S, Barnett L, Di Paolo E A. Spatial embedding and the structure of complex networks. *Complexity*, 2010, 16(2): 20-28
- [18] Goodchild M F. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 2007, 69(4): 211-221
- [19] Liu Y, Sui Z, Kang C, et al. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PloS ONE*, 2014, 9(1): e86026
- [20] Ahas R, Mark Ü. Location based services — new challenges for planning and public administration?. *Futures*, 2005, 37(6): 547-561
- [21] Tiru M, Kuusik A, Lamp M L, et al. LBS in marketing and tourism management: measuring destination loyalty with mobile positioning data. *Journal of Location Based Services*, 2010, 4(2): 120-140
- [22] Yuan J, Zheng Y, Zhang C, et al. T-drive: driving directions based on taxi trajectories // *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Savannah: ACM, 2010: 99-108
- [23] Kang C, Liu Y, Wu L. Delineating intra-urban spatial connectivity patterns by travel-activities: a case study of Beijing, China // *Proceedings of The 23rd International Conference on Geoinformatics*. Wuhan, 2014: 1-7
- [24] Liu Y, Liu X, Gao S, et al. Social sensing: a new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 2015, 105(3): 512-530
- [25] Wilson A G. The use of entropy maximizing models in the theory of trip distribution, mode split and route split. *Journal of Transport Economics and Policy*, 1969, 3(1): 108-126
- [26] 刘瑜, 龚俐, 童庆禧. 空间交互作用中的距离影响及定量分析. *北京大学学报(自然科学版)*, 2014,

- 50(3): 526–534
- [27] Schrage L. Linear, integer, and quadratic programming with LINDO. 3rd ed. Palo Alto: The Scientific Press, 1986
- [28] O’Kelly M E, Song W, Shen G. New estimates of gravitational attraction by linear programming. *Geographical Analysis*, 1995, 27(4): 271–285
- [29] Shen G. Estimating nodal attractions with exogenous spatial interaction and impedance data using the gravity model. *Papers in Regional Science*, 1999, 78(2): 213–220
- [30] Shen G. Reverse-fitting the gravity model to inter-city airline passenger flows by an algebraic simplification. *Journal of Transport Geography*, 2004, 12(3): 219–234
- [31] Xiao Y, Wang F, Liu Y, et al. Reconstructing gravitational attractions of major cities in china from air passenger flow data, 2001–2008: a particle swarm optimization approach. *The Professional Geographer*, 2013, 65(2): 265–282
- [32] Van den Bergh F, Engelbrecht A P. A study of particle swarm optimization particle trajectories. *Information Sciences*, 2006, 176(8): 937–971
- [33] Brockmann D, Hufnagel L, Geisel T. The scaling laws of human travel. *Nature*, 2006, 439: 462–465
- [34] Jiang B, Yin J, Zhao S. Characterizing the human mobility pattern in a large street network. *Physical Review E*, 2009, 80: 021136
- [35] Noulas A, Scellato S, Lambiotte R, et al. A tale of many cities: universal patterns in human urban mobility. *PloS ONE*, 2012, 7(5): e37027
- [36] 保继刚. 引力模型在游客预测中的应用. *中山大学学报(自然科学版)*, 1992, 31(4): 133–136
- [37] 刘瑜, 康朝贵, 王法辉. 大数据驱动的人类移动模式和模型研究. *武汉大学学报(信息科学版)*, 2014, 39(6): 660–666
- [38] Wang S S, Stefanone M A. Showing off? Human mobility and the interplay of traits, self-disclosure, and Facebook check-ins. *Social Science Computer Review*, 2013, 31(4): 437–457
- [39] Dann G M S. Anomie, ego-enhancement and tourism. *Annals of Tourism Research*, 1977, 4(4): 184–194
- [40] 吴必虎. 上海城市游憩者流动行为研究. *地理学报*, 1994, 49(2): 117–127