doi:10.3969/j.issn.1672-4623.2020.09.032

利用社交媒体数据探测细粒度城市子区域

武善梅[1]，王艳东[1,2*]，乔梦玲[1]，邵世维[3]，刘　辉[3]
（1. 武汉大学 测绘遥感信息工程国家重点实验室，湖北 武汉 430079；2. 地球空间信息技术协同创新中心，湖北 武汉 430079；3. 武汉市国土资源和规划信息中心，湖北 武汉 430014）

摘　要：对城市中人类活动子区域的精细探测，能为城市资源合理配置和科学管理提供有效依据。提出了一种使用社交媒体数据细粒度分割城市空间区域的方法，基于带有地理位置的社交媒体数据，利用大量用户移动数据构建空间交互网络，利用复杂网络的社区发现算法探测城市子区域。实验结果表明，利用社交媒体数据能较好地探测出反映人们真实活动空间的细粒度城市区域，为城市管理提供参考。

关键词：社交媒体；空间交互；社区发现；城市子区域
中图分类号：P208　　**文献标志码**：B　　**文章编号**：1672-4623（2020）09-0122-04

随着城市发展进程的不断加快，城市的科学化管理也面临着巨大挑战。城市的分区管理通常依据自上而下的行政区域划分，但该方式并不能真正反映城市居民在空间上的活动规律，不利于城市规划、高效资源配置的实施。对城市空间进行合理的子区域探测，理解和掌握自组织演变下的城市空间结构，有利于解决城市问题中的资源配置问题，对建设智慧城市有重大意义[1]。城市居民的移动规律潜在反映了城市空间结构信息，因此可基于居民空间活动规律对城市区域进行划分。

得益于信息技术的发展，获取大规模长时间序列的人类移动定位轨迹大数据得以实现，为基于群体在空间上的移动规律，自下而上地研究城市空间结构带来了新的契机[2]。同时，结合人类移动定位数据，网络分析理论的发展也为研究城市空间结构提供了新的方法[3-4]。Barthélemy M[5]等将复杂网络理论应用于移动定位数据，从而进行空间网络分析。网络中的社区[6]发现算法能将网络划分为内部联系紧密、外部联系稀疏的社区结构，实现城市空间的区域划分。目前用于研究城市空间结构的人类移动定位大数据源主要包括手机数据、公共交通刷卡数据、出租车数据以及社交媒体数据等。基于这些数据，结合网络分析方法，学者们取得了较好的研究成果，如 Sobolevsky S[7]等利用一系列全国范围内的手机数据构建了大规模的人类交互网络，并结合网络社区发现算法揭示了与现有行政区相似的、有凝聚力的城市分区结构；ZHONG C[8]等利用不同时期的地铁智能卡数据构建了网络，计算了网络节点属性并划分了网络社区结构，分析了新加坡城市中心、枢纽以及社会经济分区的动态变化；GUO D S[9]等利用出租车轨迹数据构建了空间网络，并提出了一种在空间约束上进行优化的网络社区发现算法，将网络划分为揭示空间移动的地理区域；YIN J J[10]等利用社交媒体数据，结合 Infomap 网络社区发现算法探测了国家尺度和城市尺度下的城市区域边界。

本文提出了一种使用社交媒体数据细粒度分割城市空间区域的方法。通过带有地理位置的社交媒体数据构建空间交互网络，结合复杂网络中社区发现算法划分空间网络，进而揭示城市子区域结构。实验结果表明，相比于单一出行数据，社交媒体数据的用户活动类型更丰富，能将城市空间划分为细粒度内部活动紧密的区域，为城市资源配置优化和科学管理提供依据。

1　数据来源与研究方法

1.1　数据来源

本文将新浪微博社交媒体数据作为研究人类空间交互的载体，实现对空间区域的划分。社交媒体数据主要通过新浪微博开放 API 接口获取，包括微博的创建时间、空间位置、文本内容、用户信息、微博 ID 等。本文收集了武汉市 13 个月带有地理位置的微博数据，共计 3 006 158 条记录，来自 787 530 个微博用户，时间为 2017 年 7 月—2018 年 8 月。微博数据的分布存在空间异质性，市中心分布较多、郊区分布较少，如图 1 所示，因此本文将数据分布较多的武汉市都市区作为研究区域，包括江岸、江汉、硚口、汉阳、武昌、青山、洪山 7 个城区以及江夏、黄陂、新洲、东西湖、蔡甸、汉南 6 个远城区的部分区域。

收稿日期：2019-05-10。
项目来源：国家重点研发计划资助项目（2016YFB0501403）；国家自然科学基金资助项目（41271399）；测绘地理信息公益性行业科研专项经费资助项目（201512015）。
（* 为通讯作者）

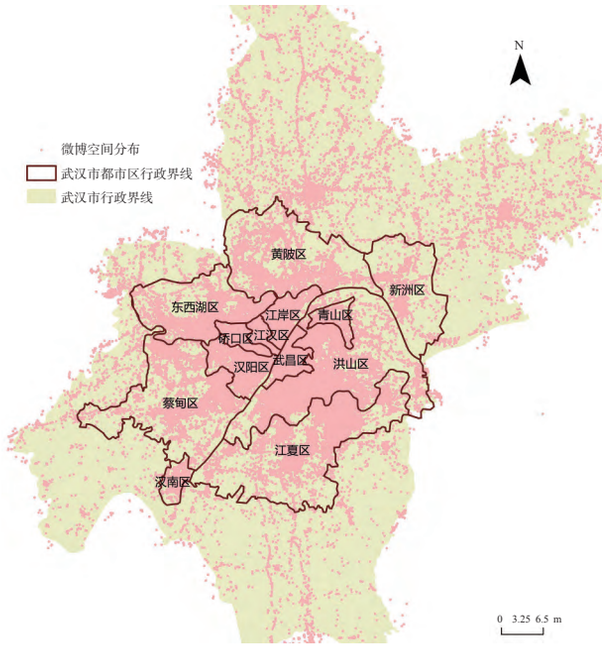

图 1　原始微博数据分布和研究区域

此外，本文还结合武汉市行政区划边界、主要道路干线、火车站、汽车站、机场、学校以及地铁线路等一些城市基本要素数据，对区域划分结果进行了分析。

1.2　研究方法

基于大量人类移动定位数据可以挖掘出群体的空间交互模式 [11]，进而挖掘城市空间结构信息。本文利用大量微博用户长时间序列在空间上的移动数据构建了空间交互网络，并采用复杂网络中的社区发现算法划分了城市区域。

1）空间交互网络的构建。本文设计了一种利用带地理位置信息的社交媒体数据构建空间交互网络的方法：将城市划分为规则的格网，以空间格网为基本研究单元，即将格网作为网络中的节点；在微博移动数据中，时间相邻的两个记录可看作是一次移动，移动跨越两个格网则作为两个格网的一次交互，对应的两个网络节点之间存在一条有向边，若两次记录位置均位于同一个格网，则作为无效的移动流。本文设置规则格网的大小为 400 m × 400 m，其主要原因在于：①居民在 1 km 以内的出行方式大多为步行，若格网大小超过 1 km，则社交媒体数据包含的该类出行可能被当作无效边处理掉，因此格网大小需小于 1 km；②为保证步行起始点位置落在不同的空间格网上构成有效的空间交互，格网大小选择 400 m、500 m 较为适宜；③参考其他微博数据相关研究 [12]。

根据上述方法，构建一个代表空间交互的加权有向网络 $G=\{V, E, W\}$，将全部用户轨迹对应的所有格网集合作为网络节点集 $V=\{v_1, v_2, v_3, \cdots, v_j\}$；将格网间所有的有效交互作为网络的边集 $E=\{e_1, e_2, e_3, \cdots, e_m\}$；将同一条边包含的交互数量作为边的权重 $W=\{w_1, w_2, w_3, \cdots, w_m\}$。以用户 i 为例，其移动数据记录为 $U_i=\{(p_{i,1}, t_{i,1}), (p_{i,2}, t_{i,2}), (p_{i,3}, t_{i,3}), \cdots, (p_{i,n}, t_{i,n})\}$，其中 $(p_{i,n}, t_{i,n})$ 为用户 i 在 $t_{i,n}$ 时刻于 $p_{i,n}$ 位置上发布了其第 n 条微博，$t_{i,p} > t_{i,q}$（$1 \leqslant q < p \leqslant n$），可看作为用户 i 的移动轨迹；再将移动记录映射到空间格网集 D，则得到用户 i 移动轨迹的格网集 $D_i=\{d_{i,1}, d_{i,2}, d_{i,3}, \cdots, d_{i,n}\}$，其中 $(p_{i,1}, t_{i,1})$ 对应格网 $d_{i,1}$；然后在用户 i 格网集合的基础上，根据有效移动的定义得到其边集 $E_i=\{e_{i,1}, e_{i,2}, e_{i,3}, \cdots, e_{i,p}\}$。

2）细粒度城市子区域探测。利用社交媒体数据构建的空间交互网络隐藏着城市商业、社交、通勤、娱乐、政治活动等信息，反映在城市结构上形成空间联系紧密的簇 [10]，通过社区发现算法划分网络结构，找到簇的边界，则可实现城市区域的精细探测。本文采用 Map Equation [13] 社区发现算法，其能较好地适用于加权有向网络。该算法从信息熵的角度出发，考虑流的方向，能发现网络中精细的社区结构，算法公式为：

$$\lg(M) = H(P) + \sum_{i=1}^{m} P_i H(p)_i$$

$$= -p\sum_{i=1}^{m} P_i \log P_i - \sum_{i=1}^{m} P_i \sum_{k=1}^{M_i} \frac{p_k}{P_i} \log \frac{p_k}{P_i}$$

$$P_i = \sum_{k} p_k$$

式中，$H(P)$ 为子区域之间的信息熵；$\sum_{i=1}^{m} P_i H(p)_i$ 为子区域加权信息熵；P_i 为用户访问子区域 m 的概率；p_k/P_i 为子区域 M_i 中的节点 k，即空间单元 k 被访问的概率。M_i 达到稳定的标准是信息熵达到最小。划分的网络社区结构具有社区内部节点联系紧密、社区之间节点联系稀疏的特点，因此能直观地展示微博用户在区域内部和区域间的交互强弱情况，更好地理解城市子区域划分结果。

3）城市子区域探测结果优化。由于没有考虑地理空间因素的影响，仅使用社交媒体用户移动的交互网络结构探测城市子区域将出现一些噪声点 [14]，因此需对结果进一步处理。城市子区域初步探测结果存在子区域边缘的离散格网点嵌入其他子区域内部的情况，分析其原因可能是只考虑了网络节点中的流关系，而实际情况中这些离散的格网可能受其邻近的格网影响更大，因此本文将其作为噪声处理。本文利用 DBSCAN 算法对探测的城市子区域进行去噪处理。DBSCAN 是一种基于密度相连的聚类算法，能较好地识别噪声，主要包括 EPS 邻域大小和 EPS 邻域内最小对象数量 MinPts 两个参数。根据研究区和节点格网

的空间尺度等具体情况，设置适宜的 EPS 邻域大小和 MinPts，从而达到较好的去噪效果，优化探测结果。

2　城市子区域探测实例

本文基于武汉市带有地理位置的微博数据，利用本文提出的方法探测城市子区域。实验前需先对数据进行预处理，包括去除数量过少、随机性强的用户数据、广告用户、重复记录等，最终采用的数据集包括 1 722 086 条微博，来自 168 676 个用户；构建的网络包括 11 400 个节点，535 884 条有效边，权重之和为 1 130 649。基于上述空间交互网络，本文划分了网络社区结构，并结合实际地理位置信息对探测结果进行了优化，得到城市子区域的划分结果。本文设置 DBSCAN 聚类 EPS 邻域大小为 400 m，MinPts 为 3。最终，得到了武汉市都市区的 193 个子区域，范围覆盖 7 051 个空间格网单元，如图 2 所示。局部放大如图 3 所示。

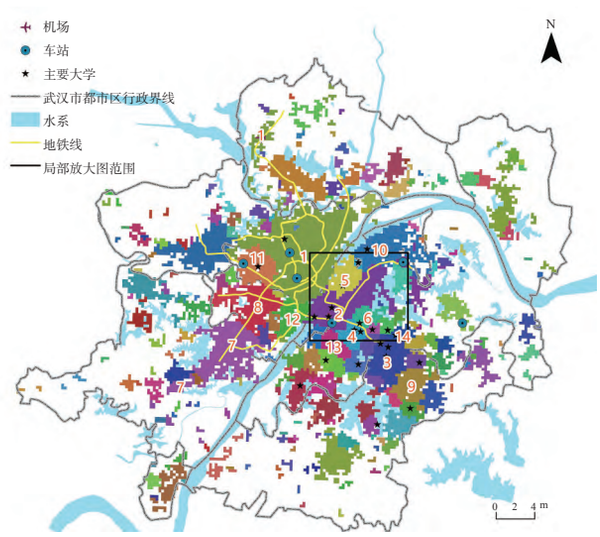

图 2　武汉市都市区子区域探测结果图

图 3　子区域探测结果局部放大图

本文结合武汉市行政区划边界、主要道路干线、火车站、汽车站、机场、学校以及地铁线路等城市基本要素信息，对武汉市区域划分结果进行分析，并分析了利用社交媒体数据划分细粒度城市子区域的可行性。由图 2、3 可知，本文方法从整体上实现了对武汉市较精细且符合真实人类活动情况的空间区域划分，子区域探测结果可分为两类。一类是面积较大、服务范围较广的区域，如面积最大的 1 号区域基本包括了江汉区和江岸区，且其区域边界与行政区划边界较吻合，其成因在于，整个 1 号区域（图 2）内部地铁线路发达、交通便利、居民活动范围较大，空间分布大体上沿着地铁线圈定区域向外延伸；探测出的 2 号区域（图 3）整体沿内部地铁线狭长分布，居民在空间活动上较便利，打破了行政区边界的约束；1 号、2 号、5 号等区域之间部分边界以长江为分界线，不存在区域打破水系约束的情况，说明水系等天然屏障在空间活动范围上对居民有巨大影响。另一类是面积较小、更加细粒度的城市子区域，如探测出的最细空间粒度范围为 3 个格网单元（图 3 中 48 号区域），覆盖了中南财经政法大学首义校区，在较大的 2 号区域内被当作独立区域划分出来；图 3 中的 6 号、26 号、21 号、14 号、4 号和 33 号等均为较精细的城市区域，主要为大学校区，分别为武汉大学、武汉体育学院、中国地质大学、华中科技大学、华中师范大学和武汉理工大学（余家头校区）。从结果分析整体来看，对于范围较大的区域，居民在空间上的交互更多体现的是行政区划、交通以及水系等天然屏障的影响；而对于细粒度的城市区域空间，更多体现的是人们日常生活形成的社区。

通过对武汉市区域划分结果的分析可知，基于社交媒体数据构建的网络，利用社区发现算法能根据区域内部和区域间交互性的强弱，实现对城市空间较为精细的子区域探测。在城市空间划分问题上，本文将社交媒体数据与其他数据进行了对比分析，采用出租车数据构建的空间网络（时间为 2013 年 12 月 31 日—2014 年 3 月 30 日，共 7 653 个节点、535 373 条有效边，与社交媒体网络数量较一致）划分的城市子区域结果如图 4 所示。对比社交媒体数据和出租车数据探测结果可以发现，二者差异显著，特别是图 4 中 2 号、5 号、8 号区域，利用社交媒体数据的空间交互能探测出较精细的区域结构，而在出租车数据结果中，则将其划分为粗粒度的子区域。分析其原因可能是相较于其他数据源，社交媒体数据涵盖的用户活动类型更丰富，在工作、购物、娱乐、锻炼等日常生活

中均会产生数据，这对刻画城市空间子区域非常有利；而地铁、出租车等交通出行数据，只能反映单方面居民交通出行在空间交互上的情况，采用单一出行数据，对城市空间划分的粒度较粗[14]，不能区分细粒度下的空间差异。因此，在利用单一数据源研究城市子区域划分时，选用能涵盖多方面信息的社交媒体数据在精细空间尺度上表现较好，进而能为城市管理提供合理的参考。

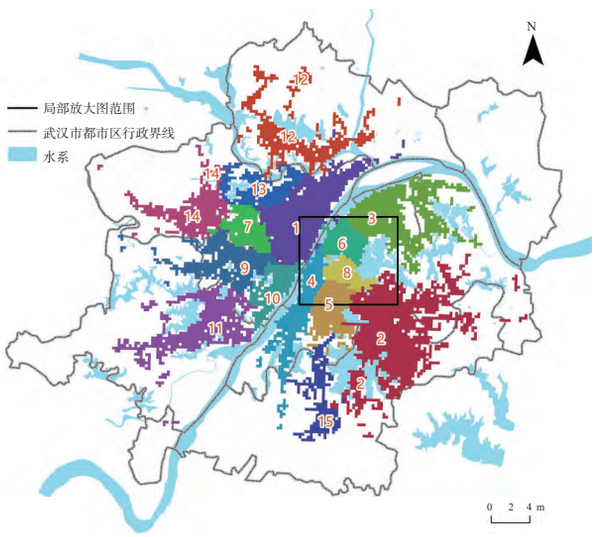

图 4　出租车数据划分城市子区域结果图

3　结　语

带有地理位置的社交媒体数据能记录用户在空间上的移动轨迹，进而能通过这种移动定位大数据探测出城市群体移动规律中蕴含的城市结构信息。本文通过获取大量社交媒体用户长时间序列的移动数据构建了用户在城市空间上的交互网络，再利用复杂网络中的社区发现算法得到区域内部交互紧密、区域间交互稀疏的城市子区域结构。区域探测结果表明，大学校区及其周边等内部一致性强、交互关系相对周边更紧密的细粒度区域能被正确划分出来；地铁等发达交通使得居民在城市空间上的移动更便利，因此也存在空间上居民活动聚集范围较大的区域；得到的子区域边界部分与行政区边界以及水系边界吻合，说明居民的空间移动在一定程度上受行政区划分、水系等天然屏障的影响；城市子区域边界也有跨越行政边界的部分，这可能是由于交通等方面原因，使居民在空间移动上能打破行政划分的约束。综上所述，社交媒体数据能综合城市空间交互上多方面的因素影响，能对城市进行较精细的区域划分，进而为城市资源合理配置和科学管理提供参考。

参考文献

[1] 闵忠荣, 杨贤房. 城市空间结构优化与城市空间管制区划: 以南昌市为例[J]. 现代城市研究, 2011(3): 43-47
[2] 杨喜平, 方志祥. 移动定位大数据视角下的人群移动模式及城市空间结构研究进展[J]. 地理科学进展, 2018, 37(7): 880-889
[3] GAO S, LIU Y, WANG Y L, et al. Discovering Spatial Interaction Communities from Mobile Phone Data[J]. Transactions in GIS, 2013, 17(3): 463-481
[4] Kallus Z, Barankai N, Szüle J, et al. Spatial Fingerprints of Community Structure in Human Interaction Network for an Extensive Set of Large-scale Regions[J]. Plos One, 2015, 10(5): 126 713
[5] Barthélemy M, Flammini A. Modeling Urban Street Patterns[J]. Physical Review Letters, 2008, 100(13): 138 702
[6] Kjaergaard M, Brander S, Poulsen F M. Small but Slow World: How Network Topology and Burstiness Slow down Spreading[J]. Physical Review E Statistical Nonlinear and Soft Matter Physics, 2011, 83(2): 25 102
[7] Sobolevsky S, Szell M, Campari R, et al. Delineating Geographical Regions with Networks of Human Interactions in an Extensive Set of Countries[J]. Plos One, 2013, 8(12): 81 707
[8] ZHONG C, HUANG X, Batty M, et al. Detecting the Dynamics of Urban Structure Through Spatial Network Analysis[J]. International Journal of Geographical Information Science, 2014, 28(11): 2 178-2 199
[9] GUO D S, JIN H, GAO P, et al. Detecting Spatial Community Structure in Movements[J]. International Journal of Geographical Information Science, 2018, 32(1): 1-22
[10] YIN J J, Soliman A, YIN D D, et al. Depicting Urban Boundaries from a Mobility Network of Spatial Interactions: a Case Study of Great Britain with Geo-located Twitter Data[J]. International Journal of Geographical Information Systems, 2017, 31(7): 1 293-1 313
[11] LIU Y, SUI Z W, KANG C G, et al. Uncovering Patterns of Inter-urban Trip and Spatial Interaction from Social Media Check-in Data[J]. Plos One, 2014, 9(1): 86 026
[12] WANG Y D, WANG T, Tsou M H, et al. Mapping Dynamic Urban Land Use Patterns with Crowdsourced Geo-tagged Social Media (Sina-Weibo) and Commercial Points of Interest Collections in Beijing, China[J]. Sustainability, 2016, 8(11): 1 202
[13] Rosvall M, Axelsson D, Bergstrom C T. The Map Equation[J]. European Physical Journal Special Topics, 2009, 178(1): 13-23
[14] LIU X, GONG L, GONG Y X, et al. Revealing Travel Patterns and City Structure with Taxi Trip Data[J]. Journal of Transport Geography, 2015, 43: 78-90

第一作者简介：武善梅，硕士研究生，研究方向为时空数据挖掘与分析。