引文格式：安佳燕,杨敏,徐之俊,等. 基于贡献者活跃模式的OSM可持续分析[J]. 地理空间信息, 2025, 23(4):1-6. doi:10. 3969/j.issn. 1672-4623.2025.04.001

基于贡献者活跃模式的OSM可持续分析

安佳燕[1]，杨　敏[1]，徐之俊[2]，张　翔[3]
（1. 武汉大学 资源与环境科学学院，湖北 武汉 430079；2. 湖北省测绘成果档案馆，湖北 武汉 430074；
3. 中山大学 测绘科学与技术学院，广东 珠海 519082）

摘　要：目前，OpenStreetMap（OSM）在基于位置的服务、城市规划、旅游和灾害管理方面发挥着越来越重要的作用。在此情况下，OSM项目的可持续性问题值得重点关注。因此使用OSM变更集（ChangeSet）数据，从贡献者行为出发，探索性分析了贡献者活跃水平发展变化情况。研究结果表明，目前OSM发展可持续性良好，未出现衰退现象，同时还发现：①贡献者过渡到更高活跃度所需的时间相当短（当年甚至当月就能达到后一活跃度），尤其是在后期；②活跃贡献者通常更快完成转化，且其整个生命期内活跃性普遍较高，这表明了其“天生的，而非后天的”的特点；③活跃是一种间歇性行为。虽然活跃度高的贡献者更专注，但极大多数贡献者的活跃度会降低，最终归于沉寂。

关键词：OSM；可持续分析；活跃模式；贡献行为
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2025）04-0001-06

OSM Sustainability Analysis Based on Contributor Activity Patterns

AN Jiayan[1], YANG Min[1], XU Zhijun[2], ZHANG Xiang[3]
(1. School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China;
2. Province Surveying Mapping Production Archives of Hubei, Wuhan 430074, China;
3. School of Geospatial Engineering and Science, Sun Yat-Sen University, Zhuhai 519082, China)

Abstract: At present, OpenStreetMap (OSM) is playing an increasingly important role in location-based services, urban planning, tourism and disaster management. In this context, it is worth focusing on the sustainability issues of OSM projects. In this paper, we use OSM ChangeSet data to analyze the developmental changes in contributor activity levels from the perspective of contributor behavior. The results show that the sustainability of OSM development is good at present, and there is no recession phenomenon. ①The time required for a contributor to transition to a higher activity level is quite short (reaching the latter activity level in the same month), especially in the later stages. ②Active contributors usually complete the conversion faster and are generally more active throughout their lifespan, which demonstrates their “natural, not acquired” characteristics. ③Active is an intermittent behavior. While contributors with high activity levels are more focused, a very large majority of contributors become less active and eventually fall silent.

Key words: OSM, sustainability analysis, active pattern, contribution behavior

OpenStreetMap（OSM）具有覆盖范围广、质量提升明显、语义丰富性高、时效性强等特点，是当前应用最为广泛的VGI开放地理数据平台之一[1-2]。与传统地理数据相比，OSM具有现势性高和更新及时等优点，在灾难消息传递与灾后重建等方面可发挥重大作用[3-9]。诸如Humanitarian等[10]人道主义组织利用OSM数据在全球范围内开展人道主义救援、地区建设等工作。国家测绘机构、专业测绘部门和数据提供商对OSM数据的接受度也在不断提高，多国官方数据导入至OSM，大量政府机构主导项目也将OSM数据作为数据来源之一[11]。维基百科、亚马逊、Facebook、Flickr等商业网站也将OSM作为数据源，以提供地理定位等服务。

随着对OSM数据依赖性的不断增加，同为众包项目的维基百科已经不可逆的衰退[12]，OSM的可持续性问题变得至关重要。我们试图从贡献者行为的角度来探讨OSM的可持续性问题。研究分析了维基百科贡献者的行为模式，发现活跃用户“生来活跃”，即用户的活跃性在其生命周期早期即有所显现[13]。研究发现OSM用户贡献偏好性也呈现出“生来的，而非后天的”的特征[14]。OSM贡献者具有高度流动性，无论曾经多么活跃的贡献者最终都会停止贡献[12]。虽然从基本统计数据来看，OSM社区整体上似乎很活跃，但OSM存在着严重的贡献不均衡现象，OSM项目的大多

项目来源：国家自然科学基金资助项目（42271458、41671384）。
第一作者简介：安佳燕（1999—），硕士研究生，主要研究方向为志愿者地理信息（VGI），E-mail：anjiayan@whu.edu.cn。

数数据由极少部分用户贡献[15-16]，这有助于我们了解OSM数据的来源。

本文基于OSM官方变更集数据（OSM ChangeSet），对贡献者分别进行了静态和动态分类，从OSM贡献者行为的角度出发，探索性分析贡献者“生命周期”内活跃度的变化情况，挖掘OSM贡献者编辑行为的动态规律，从而进一步揭示OSM项目的可持续性发展趋势。

1 研究数据、方法与目标

1.1 ChangeSet数据与预处理

OSM变更集数据集记录了自2005年4月起OSM注册用户任意一次编辑活动的基本信息，仍在不断更新。它包括贡献者的信息和此次编辑活动的信息，如用户ID、编辑数量、编辑时间等。本文以ChangeSet数据集为研究的基础数据（时段为2005-04—2021-07），导入至本地Postgre SQL数据库进行处理分析。经过数据清洗、转换后，根据研究需要，将数据时间粒度转化为月、年粒度。

1.2 OSM贡献者分级

为刻画OSM贡献者的贡献行为规律，本文采取了动、静态分级相结合的方式，以分析不同群体贡献者在生命期内活跃变化的情况。

1.2.1 静态分级

与其他在线社区类似，OSM项目也存在贡献者参与不均衡的现象。利用用户参与不均衡的1-9-90规则[12]，根据贡献者总编辑数量，对贡献者进行了静态分级：

1）1%贡献者：总编辑数量在前1%的贡献者，贡献约占总编辑数量的86.6%。

2）9%贡献者：总编辑数量在1%~10%的贡献者，贡献约占总编辑数量的11.8%。

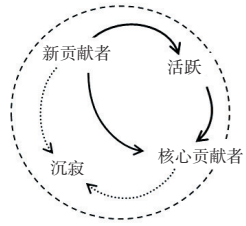
3）90%贡献者：总编辑数量在10%~100%的贡献者，贡献约占总编辑数量的1.6%。

1.2.2 动态分级

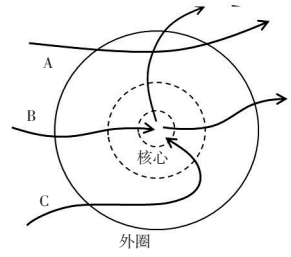
为保证项目持续发展，OSM必须不断吸引新用户。一部分新用户必须从项目外围走向核心，而另一部分用户则将最终离开核心。图1是贡献者生命周期内活跃轨迹的概念图。如图1b所示，在OSM项目中，部分贡献者生命周期内活跃水平一直较低（A）；部分贡献者一开始就表现出较高的活跃水平，成为核心用户（B）；还有部分贡献者活跃水平逐渐提高直至成为核心用户（C），但所有的贡献者最终都会离开。

按照贡献者月编辑数量，本文定义了5个活跃等级，注册用户（Register，REG）、首次编辑者（first editor，FE）、新编辑者（new editor，NE）、活跃编辑者（active editor，AE）、高度活跃编辑者（highly active editor，HAE）。OSM项目中，用户参与贡献即成为FE；NE偏向于为静态类别，当累计编辑数量超过100即成为NE；为避免90%贡献者成为HAE，我们定义月编辑数量超过10 000即转化为HAE；AE的标准则设定为月编辑数量超过1 000，这样可使大部分1%和9%贡献者转化成为AE。

在OSM的生命周期中，贡献者可以在不同等级之间转化，也可以停留在某一类别中，或者保持沉寂。我们试图根据人员流动的速度和比例来衡量其活跃轨迹的动态变化，从而进一步反映OSM项目的内部情况与可持续性。



a 生命期内循环



b OSM贡献者活跃等级转化

图1 OSM贡献者活跃轨迹概念图

1.3 可持续性度量指标

1）贡献者由等级G转化为等级$G+1$的转化时间：

$$T_{G+1}-T_G \tag{1}$$

式中，T_{G+1}为贡献者首次转化为等级$G+1$的时间；T_G为贡献者首次转化为等级G的时间。

2）贡献者由等级G转化为等级$G+1$的转化率：

$$\frac{N_{G+1}}{N_G} \tag{2}$$

式中，N_G为截至统计日期前，该时间点新增FE中最终转化成为等级G的贡献者数量；N_{G+1}为截至统计日期前该时间点新增FE中最终转化成为等级$G+1$的

贡献者数量。

此外，我们还将分析：①贡献者活跃的次数；②相邻活跃的间隔时间；③活跃持续时间，以了解贡献者在首次活跃后的情况。

1.4　研究目标

我们通过分析静态与动态用户分类下的长时序编辑活动，对OSM项目的可持续性进行评估和预测，具体研究问题包括：

1）在OSM的不同发展阶段用户的转化时间、转化率如何？有无显著差别？

2）不同静态活跃度贡献者，即前1%、9%、90%用户的转化率和转化时间如何？有无显著差别？活跃用户是天生的吗？

3）AE与HAE后会持续活跃吗？保持活跃情况如何？

本文将主要对这3个问题进行研究分析。实验主要分成用户转化情况分析（2.2节）与活跃用户的活跃贡献情况分析（2.3节）两部分。

2　实验结果与讨论

2.1　各活跃等级贡献者数量变化情况

2005年4月至2021年7月，超过七百万的注册用户中，仅有约22%贡献者进行了编辑，成为了FE，这表明OSM超过3/4的注册用户为“沉默的大多数”，从未进行过贡献。FE中，占比约32.6%的NE贡献了99.8%的数据；占比分别为10.3%和2%的AE和HAE贡献数据量分别占总数据的98.5%和92.7%。通过对OSM用户编辑次数进行统计分析，发现OSM存在参与不均现象。本文从贡献者编辑数量出发，分析结果也验证了这一结论，OSM的绝大多数贡献来自极少数活跃贡献者。

以月为粒度，图2为各动态活跃等级贡献者新增数量。为便于探索，我们将其分为了3个阶段。2005—2009年为OSM发展早期增长阶段，2007年之前注册的贡献者属于“创新者”；2007—2009年，OSM的普及等因素的影响，OSM贡献者数量明显增加。2009—2016年为平稳增长阶段，贡献者增速放缓。2016年前后为快速增长阶段，贡献者增量突增，后趋于平稳。

总体而言，我们发现OSM仍在源源不断地吸引着新的贡献者，且较为稳定。图2中4条曲线相似性很高（皮尔逊系数 $p>0.8$，在0.01水平上显著相关），

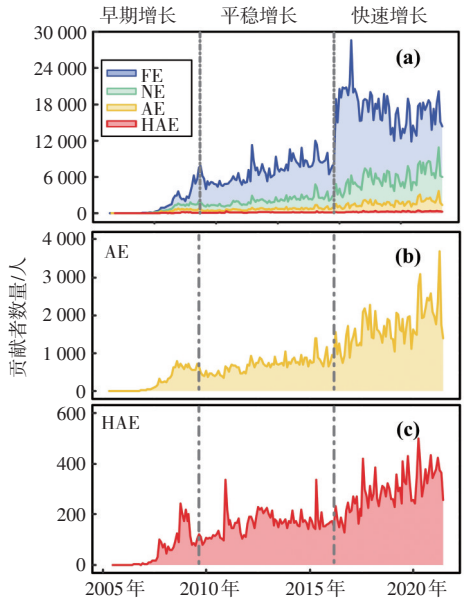


图2　各等级月新增贡献者数量

并均呈现出上升趋势。为了解不同动态活跃等级间的转化情况，我们进行了定量分析。

2.2　贡献者活跃等级转化

2.2.1　转化时间分布分析

本文将研究4种基本转换类型：REG–FE、FE–NE、NE–AE和AE–HAE。图3依次为4个活跃等级贡献者REG–FE等级转化时间累积分布图（CDF）。我们发现，超过60%的贡献者均在注册当天进行首次编辑，近85%的贡献者在一月之内会进行编辑，超过95%的贡献者在1 a内完成转化。若1 a内新注册用户没有进行首次编辑，之后会进行编辑贡献的概率显著降低。

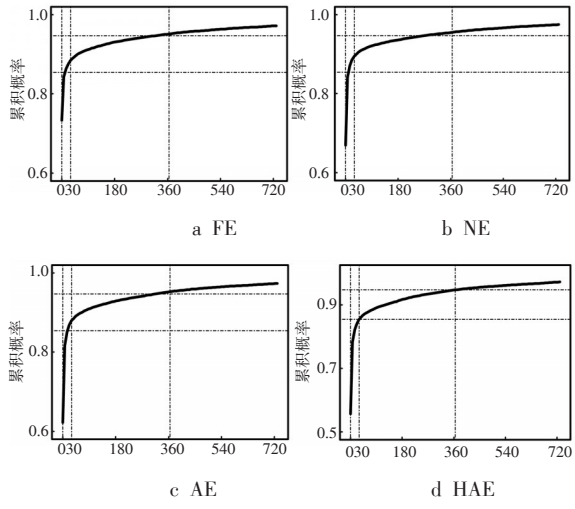


图3　各等级用户REG–FE转化时间CDF图/d

我们还发现，活跃度较高的用户（即高级用户）在较短时间内进行首次编辑的概率相对较低。例如，HAE的编辑者在注册当天进行首次编辑的人数（56%）少于FE的编辑者（73%），参见图3a和3d。

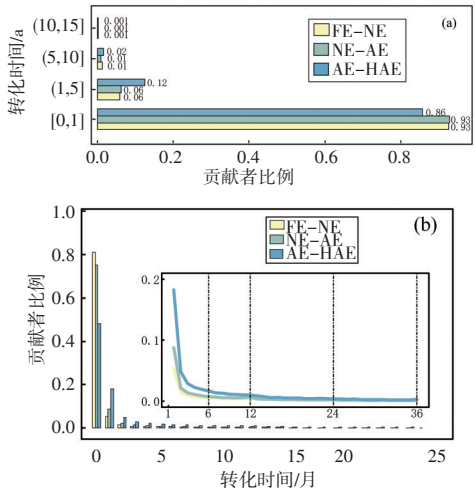


图4　用户等级转化时间图

这种现象需要进一步研究。

2.2.2　动态转化时间对比分析

为研究贡献者转化情况随时间的变化情况，我们以月为粒度进行了分析，使用了第一分位数、中位数、第三分位数3个指标，进行分组统计，并在放大插图中加入了误差线。由图4、5发现，在OSM早期阶段（2007年前），由于用户基数较小，转化时间波动明显；后期OSM发展稳定阶段，用户数量增多，转化时间趋于稳定也更具有研究意义。对比来看，FE–NE、NE–AE 2种转化类型趋势更为相似，贡献者越来越倾向于当月转化为更高等级。相较而言，AE–HAE转化类型转化时间明显更长，即使到OSM发展的后期阶段其转化时间平均也需3个月左右。

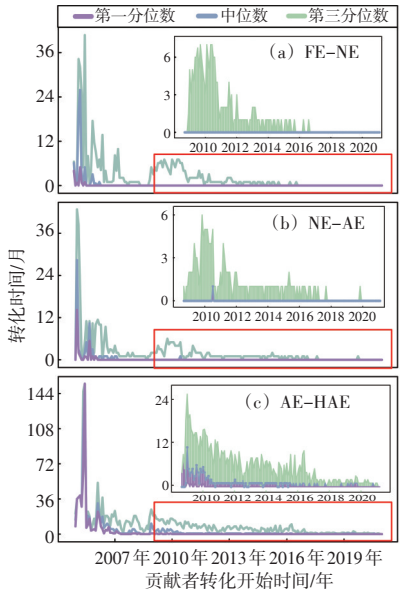


图5　用户转化时间分位数折线图

2.2.3　静态转化时间对比分析

本文将转化时间分析与贡献者静态活跃等级相结合，

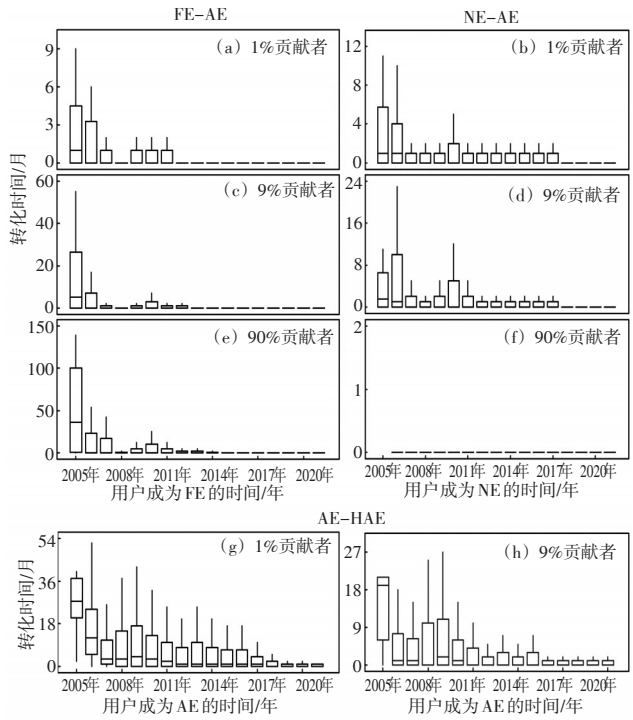


图6　1–9–90用户转化时间箱型图

如图6所示。整体来看，3种静态活跃等级贡献者的转化时间均呈现出不断减小的趋势。在OSM发展的后期阶段，四分位数趋向于0，这表明贡献者在当月即转化成为更高活跃等级。横向对比来看，同一活跃等级贡献者向较高活跃等级的转化时间（如AE–HAE）普遍高于向较低活跃等级（如NE–AE）的时间。纵向对比来看，同种转化类型下，总贡献较多的贡献者（如1%贡献者）转化时间明显少于贡献较少的贡献者（如9%贡献者），这种趋势在OSM初期已很明显。

研究中通过对OSM贡献者编辑行为的时空行为分析，发现贡献者在贡献地区与类型上有明显的偏好性，且这种偏好是“天生的，而非后天的”[14]。我们的研究表明OSM贡献者在活跃性上也表现出类似的性质，即贡献者活跃度“天生的，而非后天的”。从个人活跃轨迹来看，若贡献者一段时间内没有完成转化，后续完成转化的可能性将大幅降低，即若贡献者生命期早期不活跃，后续变得活跃的可能性较小。从不同活跃度贡献者对比来看，整体呈现出高活跃度的贡献者（总编辑数量多），其转化为更高动态活跃等级的速度更快，即整个生命期活跃性均较高。

2.2.4　转化率分析

转化率表现出贡献者转化成为更高动态活跃等级的比例，如图7所示。3种转化类型的转化率总体趋势相似。在OSM早期阶段，转化率波动较大，且转化率很高，至2009年左右转化率迅速下降。这与研

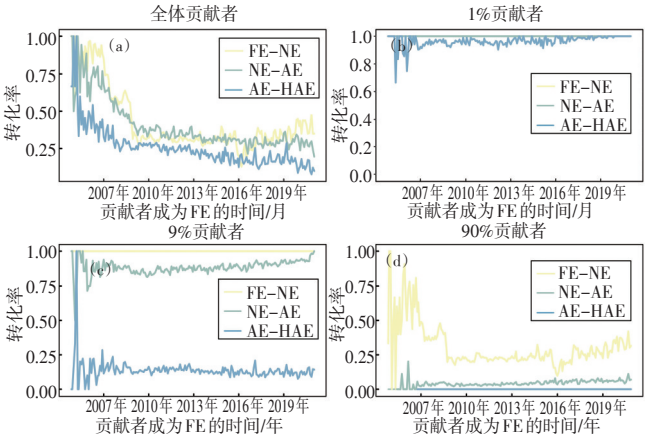


图7　OSM贡献者转化率折线图

究中提出的“创新者”和“早期使用者”时期相对应[14]。2009年之后转化类波动和缓，趋于平稳。FE–NE类型转化率呈先减后增的趋势，稳定在40%左右。NE–AE类型转化率则稳定在30%左右。AE–HAE类型在OSM发展后期，转化率也略呈减小趋势。

不同静态活跃等级贡献者转化率对比来看，除去早期波动外，三类用户转化率差异比较稳定。1%用户全部完成了FE–NE、NE–AE和AE–HAE的转化，结果转化为HAE。9%用户中，均完成了FE–NE的转化，绝大部分转化为了AE，15%~20%的用户转化为了HAE。而90%用户中，20%~40%的用户转化为了NE，不到10%的用户转化为了AE，没有转化为HAE的贡献者。

2.3　活跃贡献者活跃情况

为了解贡献者留存情况，本文聚焦于AE、HAE两类贡献者，就活跃次数、两次活跃间隔时间、活跃持续时间进行探索性分析。

2.3.1　活跃/高度活跃次数

图8为贡献者活跃/高度活跃次数的分布情况。整体来看，近60%的AE和超过40%的HAE只活跃/高度活跃一次，多次（>50次）活跃贡献者比例极小。这表明活跃度较高的贡献者中存在近半的偶发性活跃贡献者，贡献者活跃/高度活跃之后活跃度显著降低。对比来看，HAE偶发性活跃用户比例低，活跃次数也更多，即活跃持续度更高。

2.3.2　用户相邻活跃/高度活跃时间间隔

图9为用户相邻2次活跃/高度活跃的时间间隔的累积分布图。由图可知，贡献者相邻2次成为AE的时间间隔相当短，超过70%的案例在一个月之内，90%在半年之内，接近95%在1 a之内。HAE的时间间隔比AE的时间间隔稍短，但两者之间的差异不大，再

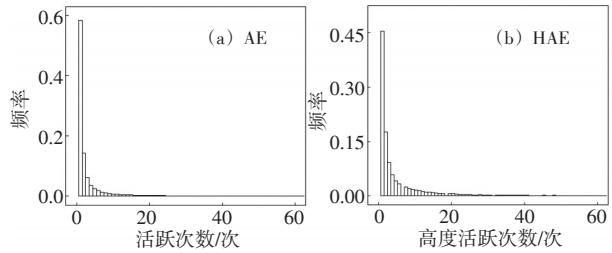


图8　AE & HAE活跃/高度活跃次数频率直方图

次表明HAE活跃持续度更高。

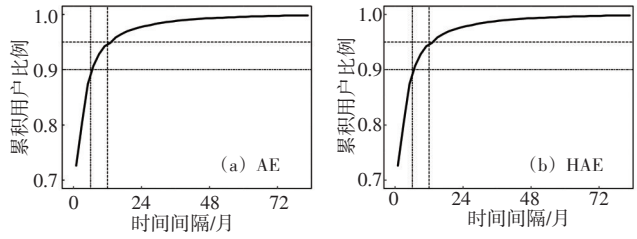


图9　AE & HAE相邻活跃时间间隔CDF图

2.3.3　用户活跃持续时间

为研究AE和HAE活跃的持续时间，我们将活跃标准统一，定义贡献者月编辑数量超过1 000为活跃；我们选取1 a为阈值来定义持续活跃，即当2次相邻活跃之间的时间间隔小于1 a时，我们认为贡献者在此期间为持续活跃状态。

图10为AE和HAE持续活跃时间累积分布图。AE中，超过75%的贡献者活跃持续时间少于1 a，接近85%的用户活跃持续时间少于2 a；HAE中超过55%的用户活跃持续时间小于1 a，接近75%的用户活跃持续时间小于2 a。

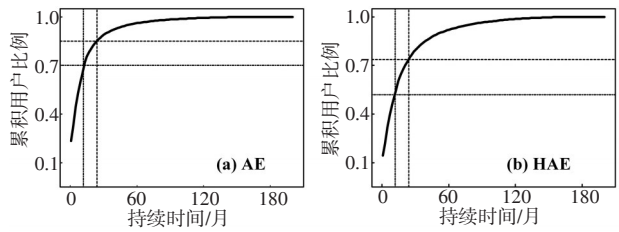


图10　AE & HAE活跃持续时间比例图

综上所述，我们的研究结果表明HAE是更为专注的贡献者，他们保持活跃的时间明显长于AE。然而，活跃期超过5 a的贡献者极少。这也证实了所有贡献者最终会停止参与贡献[12]。

3　结论与展望

本文分析了2005—2021年贡献者编辑活动的发展变化，以及编辑者在不同OSM职业阶段转换速率，旨在以一种内部视角评估OSM项目的发展阶段、衰败以及可持续性问题。研究结果表明：

1）超过一半的贡献者在一个月内会转化成为更

高的活跃等级。如果没有在1 a内完成，之后转化的可能性显著降低。过渡到更高活跃等级（如HAE）需要更长的时间。此外，加入OSM时间较晚的贡献者完成转化所需的时间较短，这似乎表明随着时间的推移，OSM贡献者越来越容易做出贡献。与之相反的是，随着时间的推移，贡献者向更高等级的转化率不断下降，但在OSM发展的后期趋于稳定。这表明OSM截至目前具有良好的可持续性。

2）相比9%和90%的贡献者，总贡献量（编辑数量）排名前1%的贡献者转化为更高活跃等级所需的时间最短。一般来说，大多数贡献者都是在当月转化为更高活跃等级，这意味着OSM中的活跃贡献者是“天生的，而不是后天培养的”。

3）活跃是一种间歇性行为。虽然活跃度高的贡献者更专注，但极大多数贡献者的活跃度都会降低，最终归于沉寂。

总体来看，我们发现OSM仍然处于"健康状态"，因为OSM几乎在2005—2021年的所有阶段都呈现出内容和人员的加速增长，而没有像维基百科那样出现明显的衰退。为了OSM的可持续性发展，本研究建议通过设计机制等吸引新来者迅速进入项目核心，鼓励核心成员在项目中停留更长时间，同时应该减少限制新用户的各种质量管控规则。

参考文献

[1] Heipke C . Crowdsourcing Geospatial Data[J]. Isprs Journal of Photogrammetry & Remote Sensing, 2010, 65(6): 550-557

[2] 张攀,周晓光,赵肆江. 基于VGI的境外地理空间数据获取[J]. 地理空间信息,2017,15(3):48-51

[3] Fohringer J , Ransch D D, Kreibich H , et al. Social Media as an Information Source for Rapid Flood Inundation Mapping[J]. Natural Hazards & Earth System Sciences, 2015, 15(12):2 725-2 738

[4] Sakaki T, Okazaki M, Matsuo Y. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors[C]// Proceedings of the 19th International Conference on World Wide Web,2010(10):851-860

[5] 陈梓,罗年学,高涛. 基于VGI的台风灾情评估研究[J]. 测绘与空间地理信息,2016,39(10):33-34

[6] Ahmed A , Hochmair H H , Sreten C . Analyzing the Effect of Earthquakes on Open StreetMap Contribution Patterns and Tweeting Activities[J]. Geo Spatial Information Science, 2018(12):1-18

[7] Poiani T H , Rocha R , Degrossi L C , et al. Potential of Collaborative Mapping for Disaster Relief: A Case Study of OpenStreetMap in the Nepal Earthquake 2015[C]// Hawaiian International Conference on System Sciences-HICSS , 2015

[8] 王艳东,魏广泽,刘波,等. 众源矢量路网融合系统的设计与实现[J]. 地理空间信息,2021,19(6):1-5

[9] 李莹,涂志德,刘艳芳,等. 基于时间序列相似度的城市功能区识别研究[J]. 地理空间信息,2021,19(1):22-29

[10] Soman S , Beukes A , Nederhood C , et al. Worldwide Detection of Informal Settlements Via Topological Analysis of Crowdsourced Digital Maps[J]. International Journal of Geo-information, 2020, 9(11):685

[11] Haklay M, Antoniou V, Basiouka S, et al. Crowdsourced Geographic Information Use in Government[M]. World Bank Publications, 2014

[12] Halfaker A , Geiger R S , Morgan J T , et al. The Rise and Decline of an Open Collaboration System How Wikipedia's Reaction to Popularity Is Causing Its Decline[J]. American Behavioral Scientist, 2013, 57(5):664-688

[13] Panciera K, Halfaker A, Terveen L .Wikipedians Are Born, not Made: A Study of Power Editors on Wikipedia[C]// ACM International Conference on Supporting Group Work. University of Minnesota, 2009.DOI:10.1 145/1 531 674.153 1682

[14] Thebault-Spieker J, Hecht B, Terveen L .Geographic Biases Are 'Born,not Made': Exploring Contributors' Spatiotemporal Behavior in OpenStreetMap[C]//Proceedings of the 2018 ACM Conference on Supporting Groupwork.Association for Computing Machinery,2018.DOI:10.1145/314 833 0.3 148 350

[15] Neis P , Zipf A . Analyzing the Contributor Activity of a Volunteered Geographic Information Project The Case of Open StreetMap[J]. ISPRS International Journal of Geo-Information,2012,1(3):146-165

[16] 杨岸然. 基于历史数据的志愿者地理信息数据质量生成机制研究[D]. 长沙:国防科学技术大学,2019