



VOLUNTEERED GEOGRAPHIC INFORMATION USERS CONTRIBUTIONS PATTERN AND ITS IMPACT ON INFORMATION QUALITY

Amir Forati, Rina Ghose
Autocarto 2020

INTRODUCTION

Introduction

Research
Objectives,
Literature
Review

Study Area

Analysis

Conclusion

References

- Recent advances in technology, cause an increase in the number of individuals involved in the production and sharing of geospatial data.
- VGI is a crowdsourced spatial data produced by a group of nonprofessional users
 - ✓ VGI is free.
 - ✓ Available to all people.
 - ✓ More updated.
 - ❖ It lacks reliable quality assessment.
 - The fact that the VGI was provided voluntarily by lay people makes the quality and reliability of these data challenging.

OBJECTIVES

Introduction ●

Research Objectives, Literature Review ●

Study Area ●

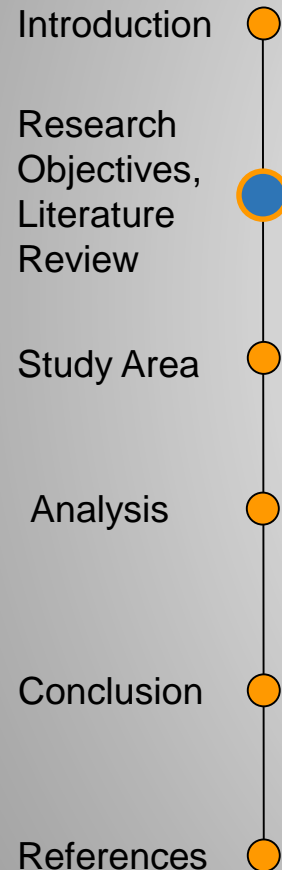
Analysis ●

Conclusion ●

References ●

- How is OSM participation in overall?
- Where is the highest participation ? Why ?
- Is participation level even whole over the world ?
- How local users participates in OSM ?
- What makes VGI based map different from authoritative (professional) maps ?

LITERATURE REVIEW



A review of user's participation history allows us to provide criteria based on the level of commitment and reliability of users to measure the quality of VGI.

- Rehrl et al. proposed a conceptual model for analyzing the activities of each participant (Rehrl et al. 2013).
 - if a user makes a correction in an area with high level of participation and other users do not correct his/her correction, then that user can be considered as a trusted user.

LITERATURE REVIEW

Introduction



Research Objectives, Literature Review



Study Area



Analysis



Conclusion

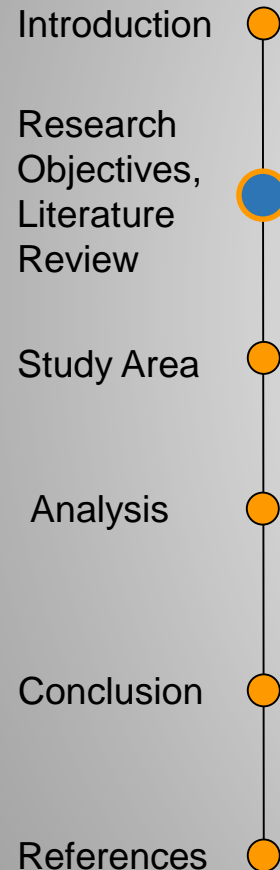


References



- Many studies have shown that the participation of users in VGI platforms are in a way that a limited number of contributors provide a large amount of information, and the rest of users make up only a few percentages of the total system contributions.
- Mooney and Corcoran in 2012 conducted research on OSM dataset in the UK, they concluded that 84% of the total edits were made by only 12% of the total users .
- So, this 12% of users compete and have a high participation in system, and we can accredit them (Mooney & Corcoran, 2012).

LITERATURE REVIEW



- Bégin and research fellows conducted research on a Canadian region, found that more than 95% of the area's information was provided by just three users. (Bégin, Devillers, and Roche 2013).
- Neis and Zipf conducted research on the OSM and obtained statistical findings that could reflect OSM status and the behavior of its users (Neis and Zipf 2012).

LITERATURE REVIEW

Introduction ●

Research Objectives, Literature Review ●

Study Area ●

Analysis ●

Conclusion ●

References ●

- Their results show that:
- 38% of OSM users have done at least one edit, and only 5% of all OSM users are actively involved in updating and editing OSM information.
- 30% of OSM users had made an edit on the day they entered the system.
- Only 5% of users entered more than 1,000 points, and 62% of users had not entered any points and just used the OSM.

LITERATURE REVIEW

Introduction



Research Objectives, Literature Review



Study Area



Analysis



Conclusion



References



- in terms of distribution, the users are limited to specific areas.
 - Only 27% of all OSM users reside outside of the European continent, and about 26% of all OSM users reside only in Germany. Asia continent share is 8% of the total OSM users.
- according to the mentioned research, a limited number of users enter most of the data, and according to the user history, we can measure their data quality through accreditation.
- But the problem arises when it comes to assessing the quality of VGI in low-density areas or features with limited number of versions.

STUDY AREA

Introduction

Research Objectives, Literature Review

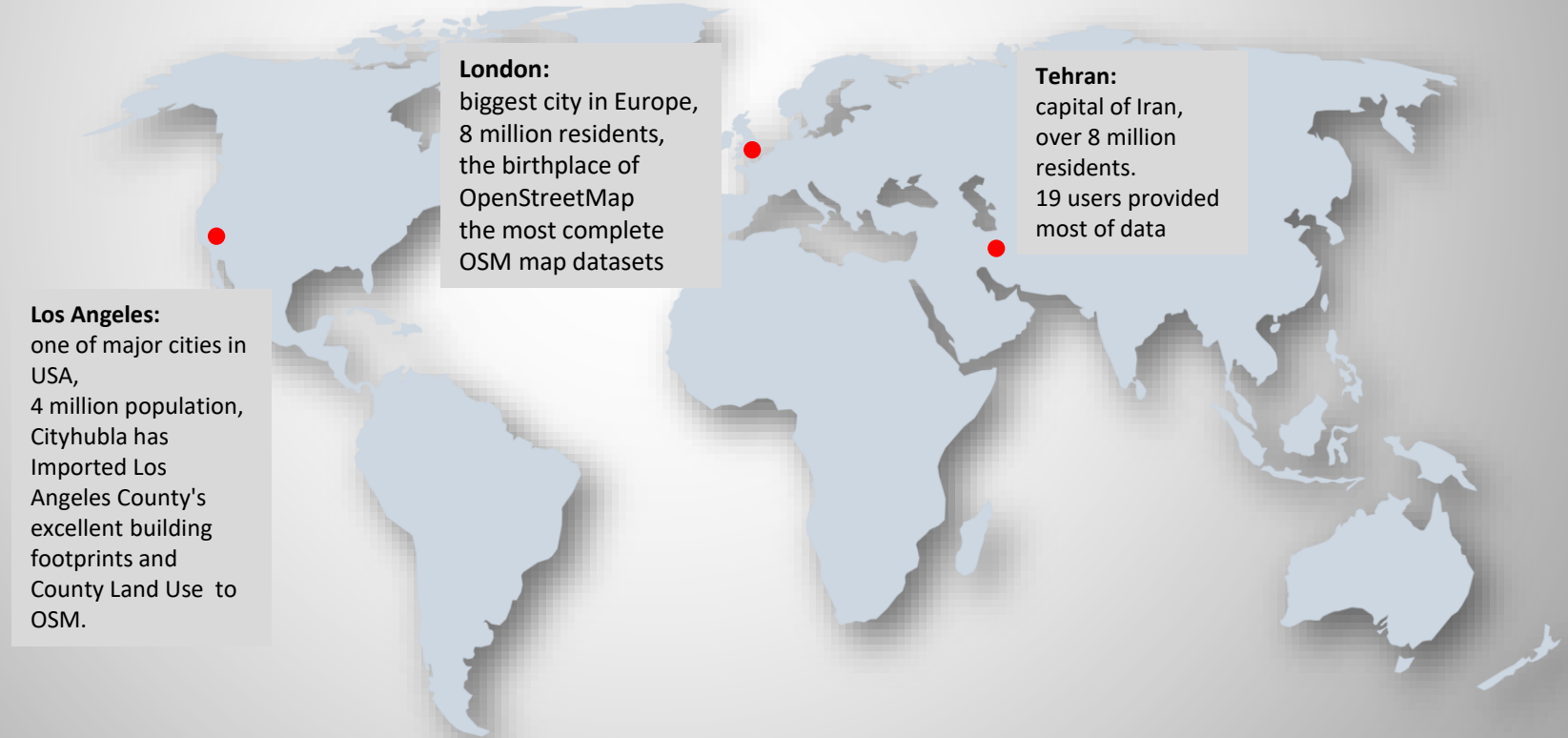
Study Area

Analysis

Conclusion

References

- a comparison between level of participation in three different megacities: London, Tehran and Los Angeles



STUDY AREA

Introduction

Research Objectives, Literature Review

Study Area

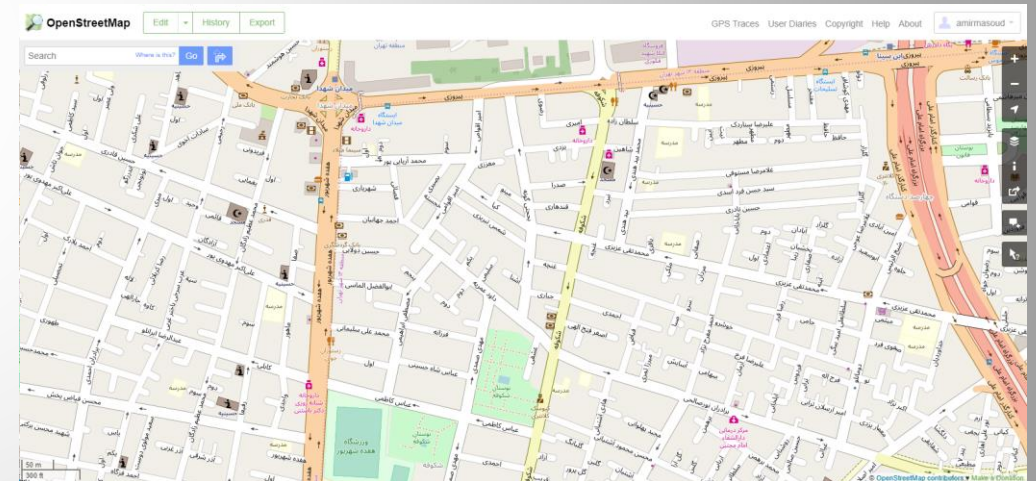
Analysis

Conclusion

References

- Wealthy and well-educated neighborhoods in London and Los Angeles were selected as a comparison source for an old and medium income neighborhood in Tehran.

West Kensington is an area of West London. It is best known as home to the Olympia Exhibition Centre and the Queen's Club. West Kensington is a cosmopolitan enclave. The area also has a large student population, as it is close to Imperial College London, Charing Cross Hospital, South Kensington, Hammersmith and other Central London Universities.



STUDY AREA

Introduction

Research Objectives, Literature Review

Study Area

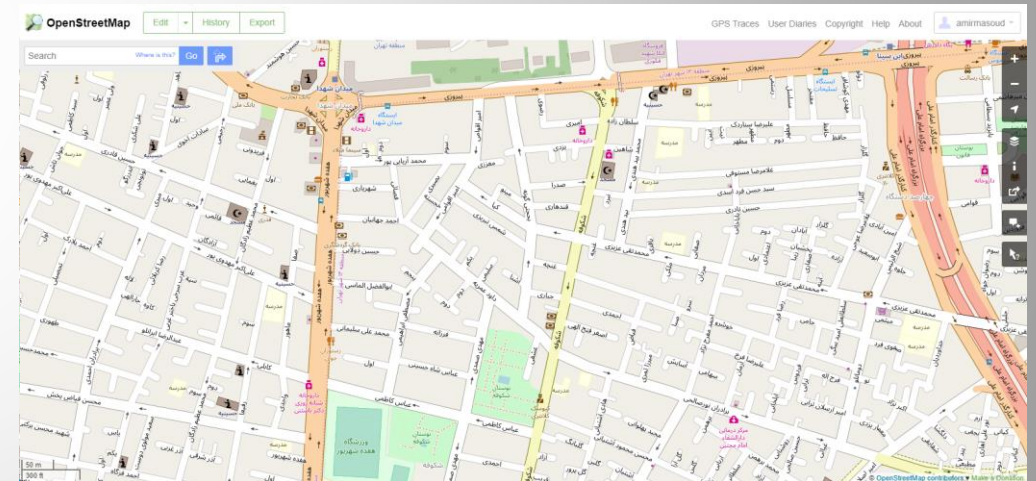
Analysis

Conclusion

References

- Wealthy and well-educated neighborhoods in London and Los Angeles were selected as a comparison source for an old and medium income neighborhood in Tehran.

Shekoufeh is a commercial and residential neighborhood in center of Tehran, Iran. It is one of oldest neighborhood of Tehran, with diverse ethnicity, roughly medium income level, and conservative traditional population.



STUDY AREA

Introduction

Research Objectives, Literature Review

Study Area

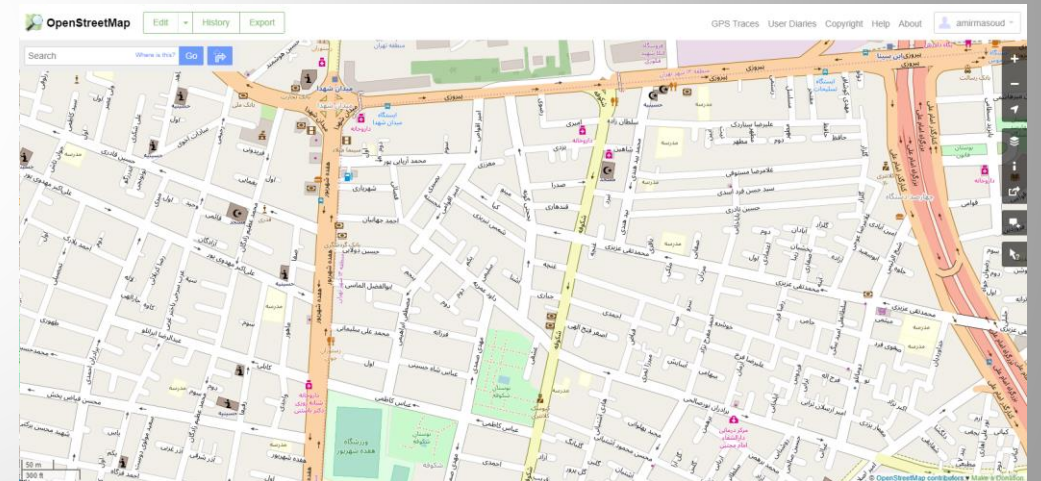
Analysis

Conclusion

References

- Wealthy and well-educated neighborhoods in London and Los Angeles were selected as a comparison source for an old and medium income neighborhood in Tehran.

Westwood is a commercial and residential neighborhood in the northern central portion of the Westside region of Los Angeles, California. It is the home of the University of California, Los Angeles (UCLA). With generally young and moderately diverse ethnically, with a generally high level of income and education population



COMPARISON METRICS

Introduction

Research
Objectives,
Literature
Review

Study Area

Analysis

Conclusion

References

- OSM dataset is consist of:
 - objects and tags, objects are map features and tags describe specific features of map elements (nodes, ways, or relations) or changesets
- Metrics to assess the level of contribution in each neighborhood.
 - 1. Number of Tags per Object,**
 - shows the level of completeness of map features and it implicitly conveys the level of local knowledge in the dataset as tags are mostly provided by local users.
 - 2. Number of Objects per Users,**
 - shows level of user's participation and how and in what degree local knowledge exists in the dataset.
 - 3. Number of Users in a Square Kilometer,**
 - conveys the participation level in the area.

PARTICIPATION ANALYSIS OF THREE NEIGHBORHOODS

Introduction

Research
Objectives,
Literature
Review

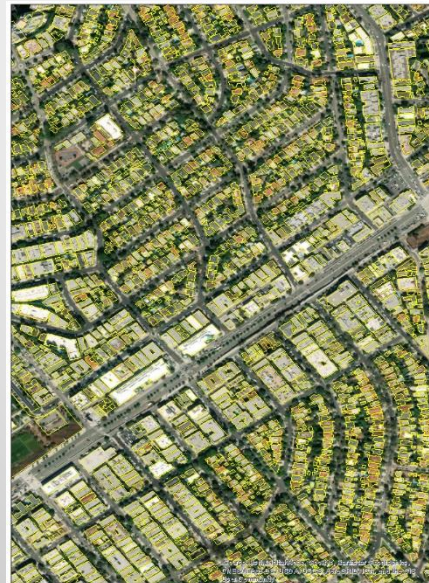
Study Area

Analysis

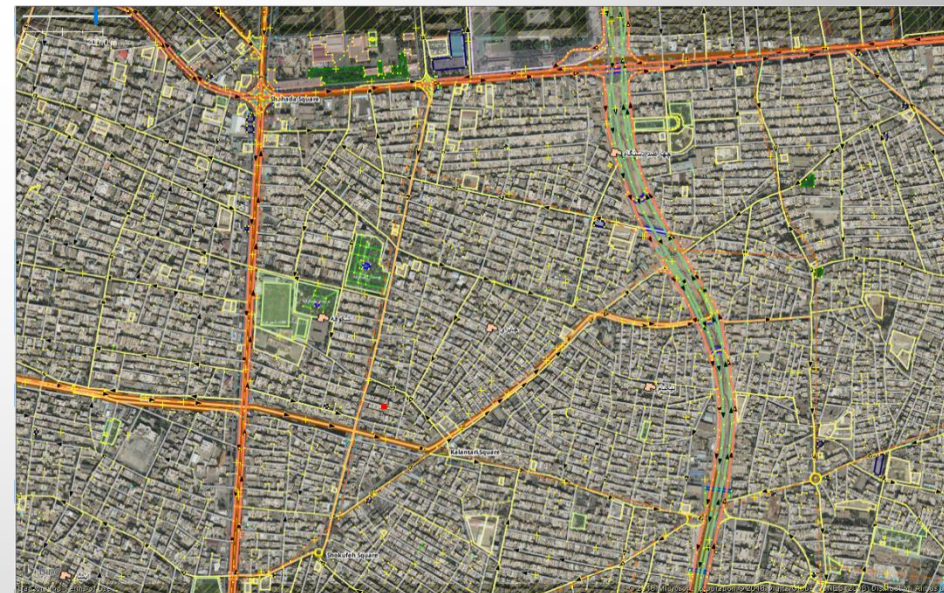
Conclusion

References

City	Neighborhood	objects	tags	users	tags/objects	objects/users	users per km ²
L.A.	Westwood	57063	174	148	0.003	385.5608108	45.96
London	West Kensington	16306	289	168	0.017	97.05952381	63.63
Tehran	Shekoufeh	7888	87	83	0.011	95.03614458	22.80



L.A. OSM map



Tehran OSM map

NUMBER OF OBJECTS

Introduction

Research Objectives, Literature Review

Study Area

Analysis

Conclusion

References

City	Neighborhood	objects	tags	users	tags/objects	objects/users	users per km ²
L.A.	Westwood	57063	174	148	0.003	385.5608108	45.96
London	West Kensington	16306	289	168	0.017	97.05952381	63.63
Tehran	Shekoufeh	7888	87	83	0.011	95.03614458	22.80

- Westwood, LA has the most number of objects, 57063 objects. it is because of two major import of data by Cityhubla that has Imported Los Angeles County's excellent building footprints dataset and also Los Angeles County Land Use to OSM. Thus, LA OSM dataset has considerable amount of data at very good level of completeness. then West Kensington, London has 16306 and Shekoufeh, Tehran has 7888. actually, all blocks and buildings are existing in LA OSM map but in Tehran only major buildings and streets are existing

NUMBER OF TAGS PER OBJECT

Introduction ●

Research Objectives, Literature Review ●

Study Area ●

Analysis ●

Conclusion ●

References ●

City	Neighborhood	objects	tags	users	tags/objects	objects/users	users per km ²
L.A.	Westwood	57063	174	148	0.003	385.5608108	45.96
London	West Kensington	16306	289	168	0.017	97.05952381	63.63
Tehran	Shekoufeh	7888	87	83	0.011	95.03614458	22.80

- The **Number of tags per object** of West Kensington is 0.017, then Shekoufeh is 0.011 and in third place Westwood has 0.003 tags per each object. It means London OSM map is more detailed and complete and then Tehran's OSM map has acceptable level of tags per object in comparison with London, which means it progressed well in this manner, Los Angeles tags per object is very low at 0.003 that means 1000 objects on map have only 3 tags, there is considerable gap between London tags per object and Los Angeles, it might be backed to the Cityhubla project, but it highlights the need for more local knowledge to incorporated into LA OSM map.

NUMBER OF OBJECTS PER USERS

Introduction

Research Objectives, Literature Review

Study Area

Analysis

Conclusion

References

City	Neighborhood	objects	tags	users	tags/objects	objects/users	users per km ²
L.A.	Westwood	57063	174	148	0.003	385.56	45.96
London	West Kensington	16306	289	168	0.017	97.06	63.63
Tehran	Shekoufeh	7888	87	83	0.011	95.04	22.80

- **Number of objects per users** implies the level of user's participation and how and in what degree local knowledge exists in the dataset. In Shekoufeh, Tehran, each user provides about 95.04 objects then each Londoner provide about 97.06 object, then comparison between these two number shows that London and Tehran have approximately the same level of local knowledge in their OSM dataset, in contrast, each Westwood user provides 385.56 objects. it significantly differs from Tehran and London. So, LA OSM dataset needs more local knowledge to be integrated with.

NUMBER OF USER IN A SQUARE KILOMETER

Introduction ●

Research Objectives, Literature Review ●

Study Area ●

Analysis ●

Conclusion ●

References ●

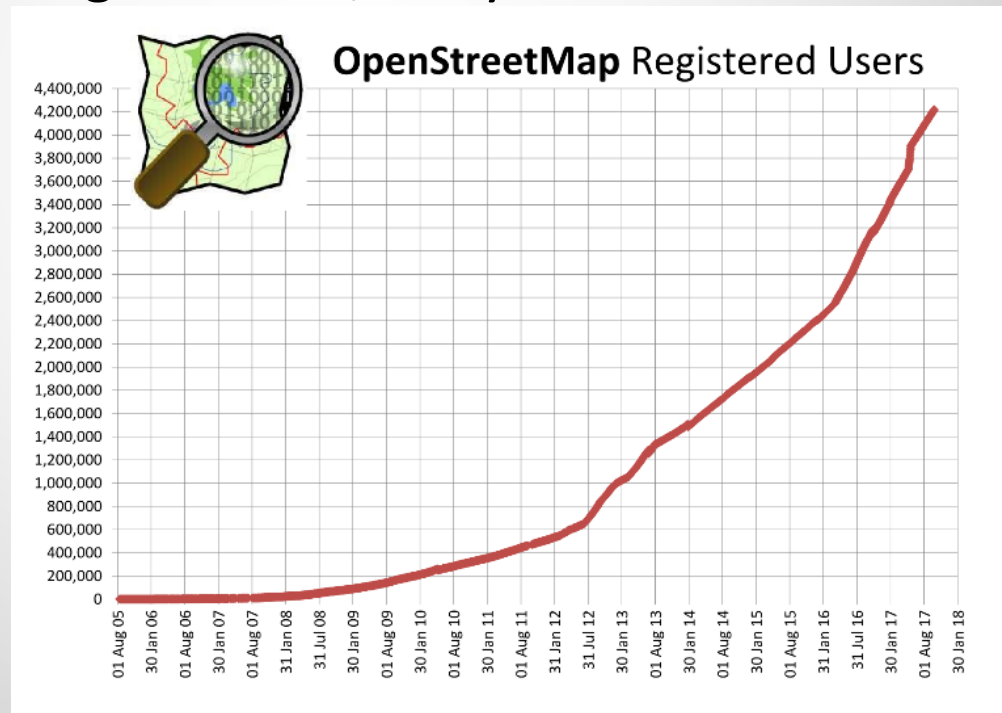
City	Neighborhood	objects	tags	users	tags/objects	objects/users	users per km ²
L.A.	Westwood	57063	174	148	0.003	385.5608108	45.96
London	West Kensington	16306	289	168	0.017	97.05952381	63.63
Tehran	Shekoufeh	7888	87	83	0.011	95.03614458	22.80

- the higher number of users involved in correcting a specific feature, the higher reliability of that feature. This criterion is based on Many eyes principle *“If something is visible to many people then, collectively, they are more likely to find errors in it.”* Therefore, **number of user in a square kilometer** conveys the participation level in the neighborhood, and also quality of data. London has the most number of participators in a square kilometer, 63 participators, then LA has 45 users per each square kilometer and Tehran has 22. It shows that the level of participation in Tehran is very low in comparison with London, it might be because of no OSM based services offered in Iran and OSM has low reputation in Iran.

CURRENT OSM STATUS

- Introduction
- Research Objectives, Literature Review
- Study Area
- Analysis
- Conclusion
- References

- The number of users of VGI systems is growing dramatically, as in 2012, only 500,000 OSM members were in place, and at the beginning of 2018, they reached more than 4,200,000 people.



OSM NEW CONTRIBUTORS PER MONTH

Introduction

Research Objectives, Literature Review

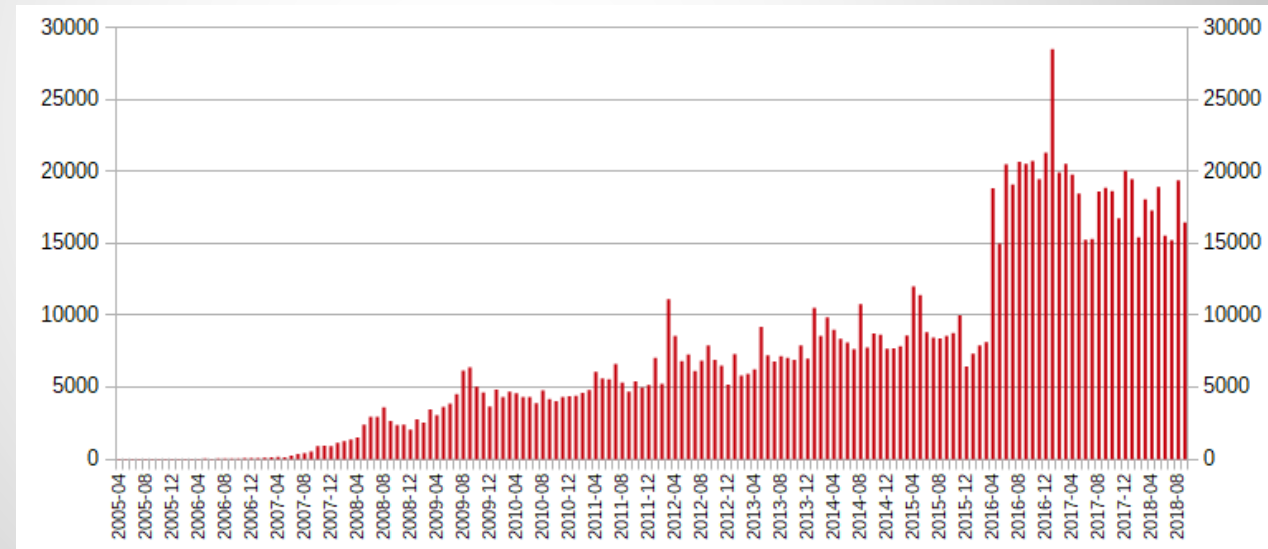
Study Area

Analysis

Conclusion

References

- Each month, the number of users of OSM increases, and this increment increases month-by-month.
- a massive amount of new users register into the system and begins to generate, correct, or use VGI systems



- So, VGI quality assessment methods should have the ability to assess quality of the information generated by new users

VGI QUALITY ASSESSMENT METHODS

Introduction

Research Objectives, Literature Review

Study Area

Analysis

Conclusion

References

- Various methods have been designed to determine the quality of VGI.
 - based on the concepts of trustworthiness and reputation of users
- The underlying assumption:
 - the VGI system has a high level of participation.
- As discussed earlier, nearly 90% of the information in VGI systems is provided by 10% of the total users. Hence, the existing methods analyze the 90% of the data and ignore the remaining 10% of the information.
- **What is the remaining 10% of the information ?**

AMOUNT OF LOCAL KNOWLEDGE

Introduction

Research Objectives, Literature Review

Study Area

Analysis

Conclusion

References

City	Neighborhood	Objects	Total users	Percentages of users who produced 90% of whole dataset
L.A.	Westwood	57063	148	5.50%
London	West Kensington	16306	168	19.54%
Tehran	Shekoufeh	7888	83	19.27%

- There are 7888 objects in Shekoufeh neighborhood that provided by 83 users, 3 users created 70% of whole dataset.
- 5.50% of users in LA provided 90% of whole dataset and about 20% of users in London and Tehran, provided 90% of whole dataset.
 - the implication is that local user provide about 10% percent of whole dataset.
 - The discrepancy between L.A., London and Tehran is because of Cityhubla project

LOCAL USERS

Introduction

Research Objectives, Literature Review

Study Area

Analysis

Conclusion

References

- **Shekoufeh,**
 - **3 users** have provided **70% of the information**, but the features which need local information have been created by about **47 users**, **21** of whom only have done one edit in the system.
- **West Kensington,**
 - **8 users** provided 80% of whole dataset and 54 users provided features which need local knowledge, 30 users have done only 1 edits in the system.
- **Westwood,**
 - 90% of whole dataset was provided by 7 users and 82 users provided local knowledge-based features, and 36 of them have created only one feature/tag.

USERS WITH ONE EDIT PERFORMANCE

Introduction

Research Objectives, Literature Review

Study Area

Analysis

Conclusion

References

Neighborhood	User ID	Number of Edits	Edit
Westwood	Vilmos80	1	amenity=fuel
Westwood	celosia	1	name=Shamshiri Grill Persian restaurant
Westwood	gnnmotors	1	name=G&N Motors MBZ Certified Mercedes-Benz Service & Repair
Westwood	lordwimp	1	shop=convenience
Westwood	soccerman_04	1	Updated a cafe
West Kensington	cantor34	1	underground station name
West Kensington	davidearl	1	uk_postcode_centroid=W14
West Kensington	earthloop	1	traffic lights here
West Kensington	ehm1806	1	Adding detail to supermarkets
West Kensington	nlehuby	1	brewery= beer Fullers > Fuller's
Shekoufeh	Morteza Farahani	1	Created 2 books shops, a artwork, and a beauty shop
Shekoufeh	Mostafadingo	1	Created a dentist
Shekoufeh	Nasser Tahani	1	Traffic zone split
Shekoufeh	amirhossianzareasadi	1	Updated a government office and a pharmacy
Shekoufeh	tehsh	1	it,s correct name is Zarkar.

VGI VERSUS AUTHORITATIVE MAPS

Introduction

Research
Objectives,
Literature
Review

Study Area

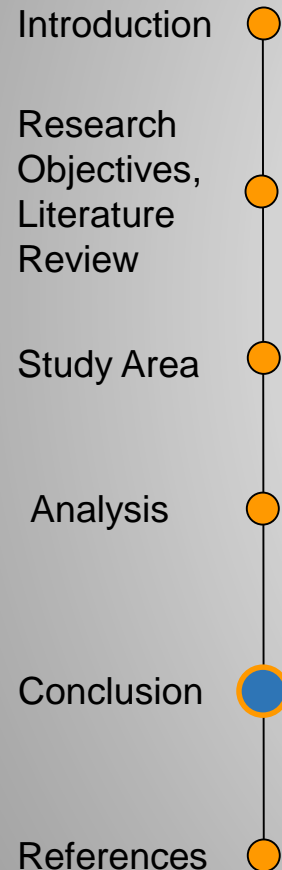
Analysis

Conclusion

References


- The main issue, which makes authoritative maps provided by national surveying organizations different from maps generated through VGI, is the local information.
- factually 90% of the information in these two types of maps are common, and it can be said that the primary and most important VGI, are parts of these maps that are produced by users with local knowledge.
- So, there is a considerable weakness if a quality assessment method of VGI does not have the ability to assess all data, especially data with a limited number of versions.

CONCLUSION



- Non-responsiveness of existing methods for the quality measurement of VGI systems with low level of participation, as well as the inability of existing methods for the quality analysis of VGI generated by new users, highlights the need to provide a solution to overcome these limitations.
- Scholars who want to develop an algorithm to assess the quality of VGI should take features which are created by new user/user with limited performance history and features with limited number of version into account.

REFERENCES



Introduction • Bégin, D., Devillers, R., & Roche, S. (2013, May). Assessing volunteered geographic information (VGI) quality based on contributors' mapping behaviours. In *Proceedings of the 8th international symposium on spatial data quality ISSDQ* (pp. 149-154).

Research Objectives, Literature Review • Forati, A. M., & Karimipour, F. A VGI Quality Assessment Method for VGI based on Trustworthiness.

• Forati, A., Soleimani, S., Karimipour, F., & Malek, M. R. (2015). INCLUDING USERS' SEMANTICS IN EVALUATING THE CREDIBILITY OF CROWDSOURCED LANDSCAPE DESCRIPTIONS. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3), 35.

Study Area • Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 37(4), 682-703.

Analysis • Mooney, P., & Corcoran, P. (2012). Characteristics of heavily edited objects in OpenStreetMap. *Future Internet*, 4(1), 285-305.

• Neis, P., & Zipf, A. (2012). Analyzing the contributor activity of a volunteered geographic information project—The case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1(2), 146-165.

Conclusion • Rehrl, K., Gröechenig, S., Hochmair, H., Leitinger, S., Steinmann, R., & Wagner, A. (2013). A conceptual model for analyzing contribution patterns in the context of VGI. In *Progress in location-based services* (pp. 373-388). Springer, Berlin, Heidelberg.

References •

QUESTION AND ANSWER

Introduction ●

Research
Objectives,
Literature
Review ●

Study Area ●

Analysis ●

Conclusion ●

References ●

