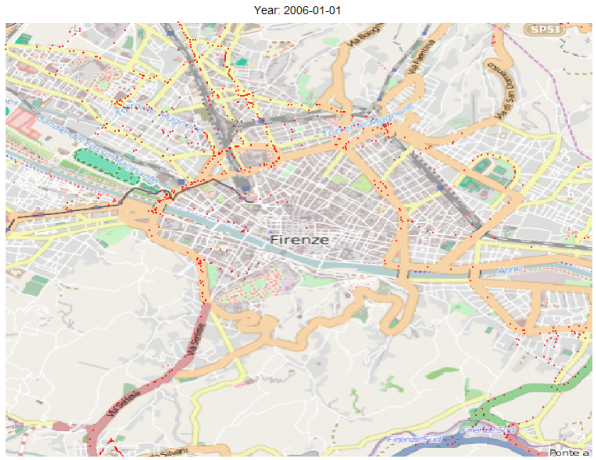


Monitoring data consistency in OpenStreetMap using its spatial features and tags semantics

Alfonso Crisci - IBIMET CNR
Maurizio Napolitano - DCL FBK Trento
Francesca De Chiara - DCL FBK Trento
Cristian Consonni - DCL FBK Trento

George Kingsley Zipf
(PD, https://commons.wikimedia.org/wiki/File:George_Kingsley_Zipf_1917.jpg)

Consiglio Nazionale delle Ricerche
Istituto di Biometeorologia

FONDAZIONE BRUNO KESSLER

Backgrounds #osm3words

OpenstreetMap is a *language* of free representation of real geographical entities to build visual patterns called maps where user communities works in participatory style.

https://www.flickr.com/photos/cainmark/4859235622/

Aims build a local areal approach

Do metrics exist to manage OSM spatial and textual informative complexity?

Which are the candidates?

Targets

- to build customized guidelines for thematic mapping
- to help areal OSM fill gapping strategies
- detect spatial and informative gaps

Parameters looking up the most interesting

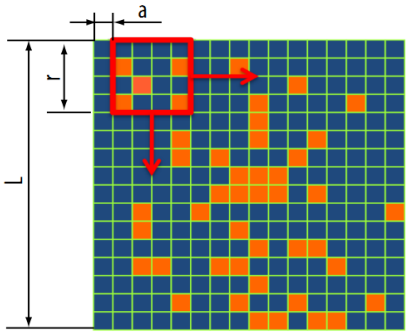
- Fractal Dimension Is it possible to measure spatial complexity of OSM feature?
- Lacunarity Is it possible to identify where OSM contributions have spatials gaps and how they change over time?
- Textual informative density Is semantics of textual descriptors (keys and tags) informative?
- Diversity and Dissimilarity Is it possible to detect semantic differences among different areas/communities at various spatial scales?

What's Lacunarity : measure of spatial pattern voids

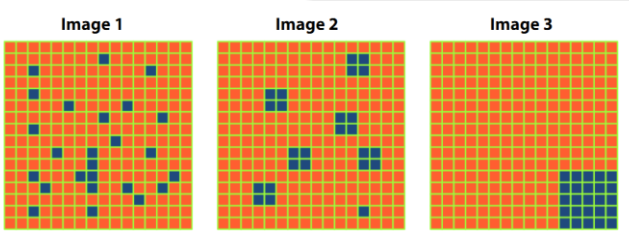
Lacunarity

It is a pattern design analytical tool and can be defined as a complementary measure of fractal dimension.
It allows to distinguish spatial patterns through the analysis of their gap (pixel void) distribution at different scales.
Is rotational invariant ma as function of the scale.

$$\Lambda(r) = \frac{Z_Q^{(2)}(r)}{[Z_Q^{(1)}(r)]^2}$$



Gliding Box lacunarity
(Allain Cloitre 1991)



$$(\lambda_1 > \lambda_2 > \lambda_3)$$

Same image complexity different lacunarity

Images & definitions
Marco Diego DOMINIETTO
ETH Zurich
Multimodality Approach To Study
The Fractal Physiology Of Tumor Angiogenesis

Information Entropy & Zipf plot: semantic analisys of OSM 's wordsets

Textual information density

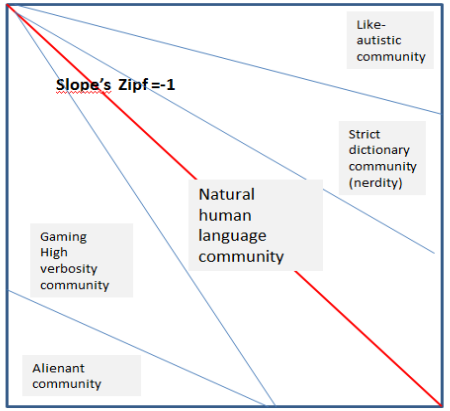
From OSM data is possible to retrieve textual corpus (set of words) of *keys, tags, keyvalue*) for every bounded area. Two action are possible:

Zipf plot : Description of word set in terms of distribution of terms. Rare terms detection.

Information entropy : to detect indirectly textual information density (Shannon entropy)

http://en.wikipedia.org/wiki/Entropy_%28information_theory%29

Zipf's graph & Ortolani's areas



Zipf's law states that given some corpus of natural language utterances, the frequency of any word is inversely proportional to its rank in the frequency table. http://en.wikipedia.org/wiki/Zipf%27s_law

x is the rank of a word in the frequency table;
y is the total number of the word's occurrences (frequency).

Tools analitical framework for OSM data

Osmconvert
Osmfilter

Urbanisation Regime and Environmental Impact: Analysis and Modelling of Urban Patterns, Clustering and Metamorphoses
GDAL lacunarity and fractal dimensions
Spatial-tools library
Christian Kaiser
http://github.com/christiankaiser/spatial-tools

R *packages*

tm & ZipfR & LanguageR & qdap &wordcloud

Openstreetmap & Osmar & fractaldim

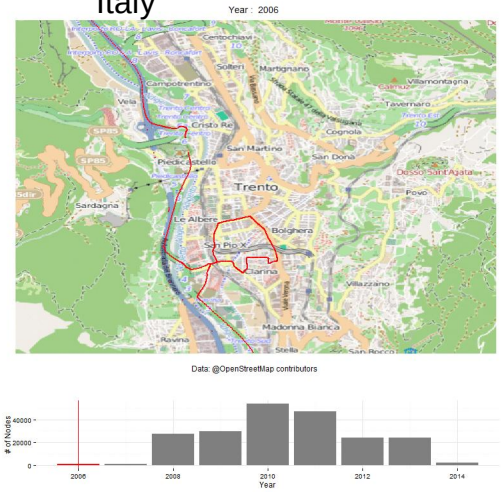
raster & ***rgdal*** & ***spatstat***

Nepal Civic Hacker @prabhasp
http://prabhasp.github.io/OSMTimeLapseR

http://github.com/alfcrisci/osm_analitics

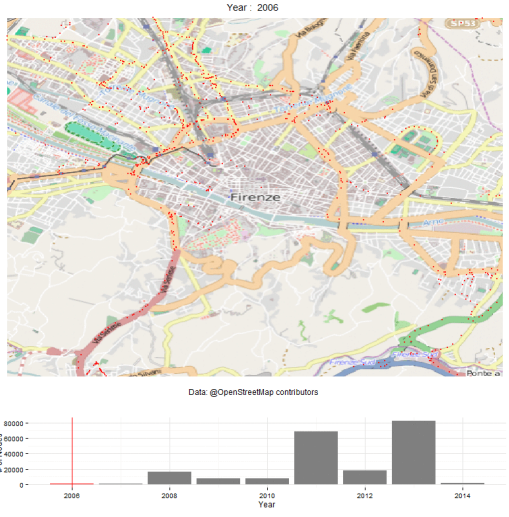
Areas Zoom.level 12 Scale 1:150,000 Admin-centre centered

Trento Northern Italy



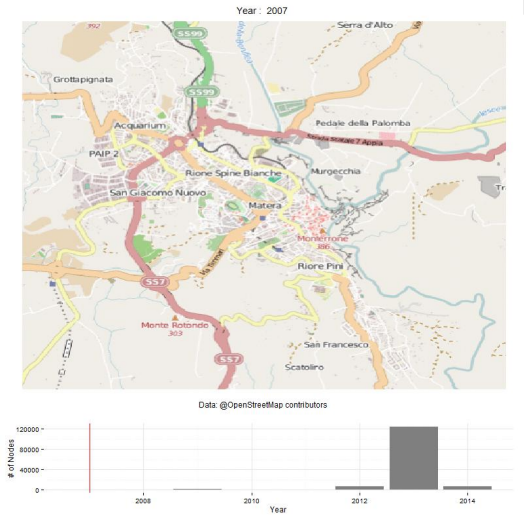
Medium city
Large Community
High density of features

Florence Central Italy



Large urban area
Large Community
High density of features

Matera Southern Italy



Small urban area
Young Community
Recent mapping

OSMTimeLapseR

OSM History Data preview

Raster Density Maps

A. Feature density
B. Users density
(at least one edit)

A. Version Count density

A. Local complexity
Fractal dimension isoentropic method
Davies and Hall (1999)

Lexical Analysis

a. Zipf plot keys
b. Wordcloud keys
c. Histogram keys/ N_users
d. Venn diagram keys/user
e. Clustering users by key

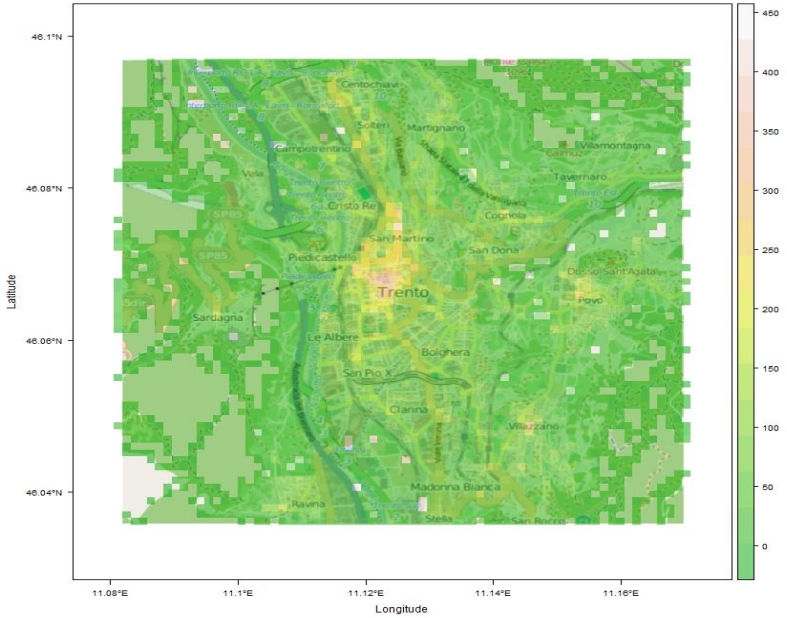
Tag Lexical Analysis

a. Zipf plot of selected key-values
b. Lexical diversity by keys
c. Treemap users by key
d. Treemap values by key
e. Word-network of user by keys

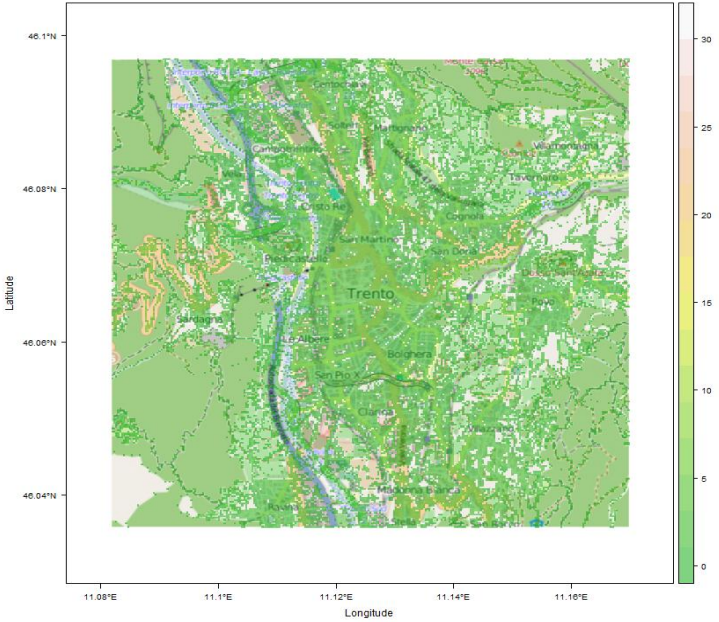
Temporal Evolution

I. Year Feature amount
II. Year Lacunarity index

Aerial view Trento spatial resolution 20 m

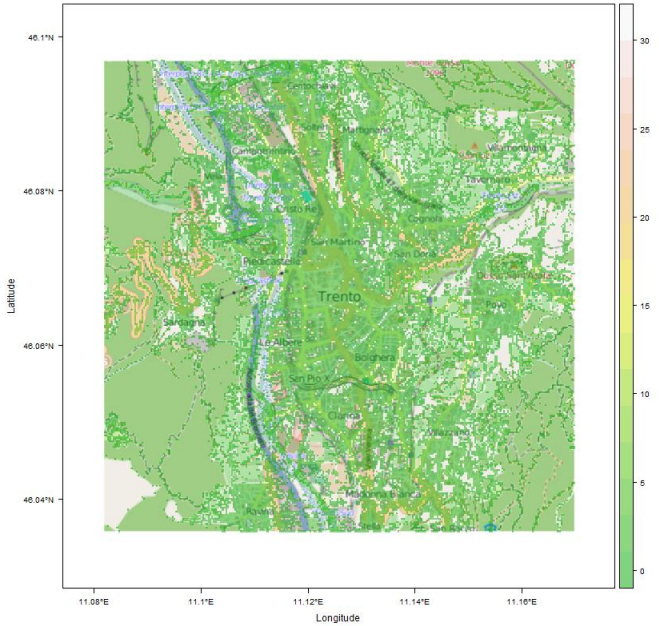


Feature density

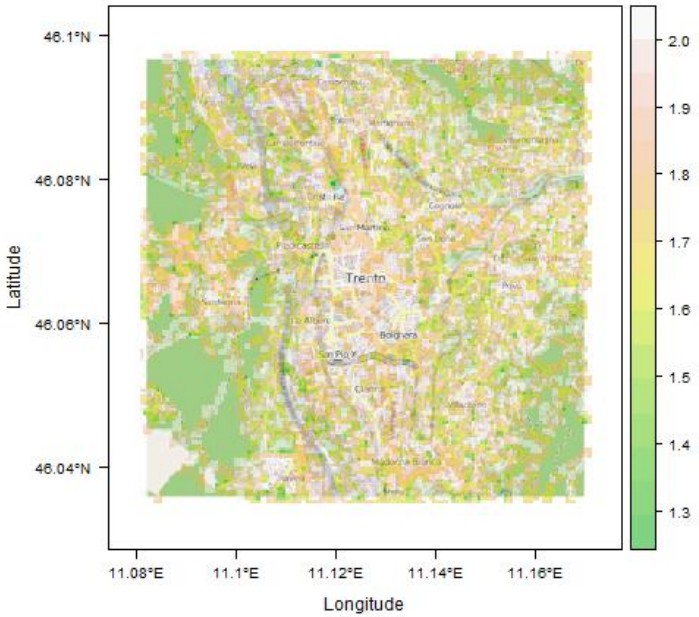


Users density

Aerial view Trento spatial resolution 20 m

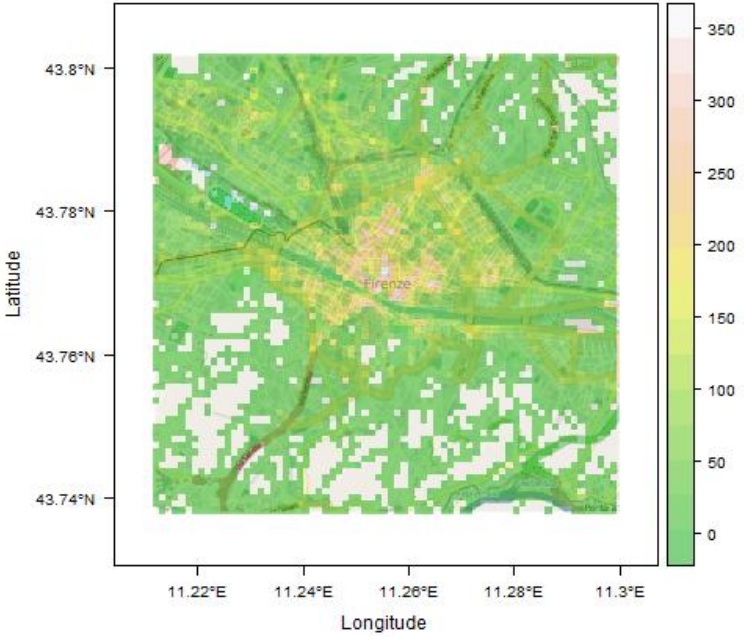


Version Count density

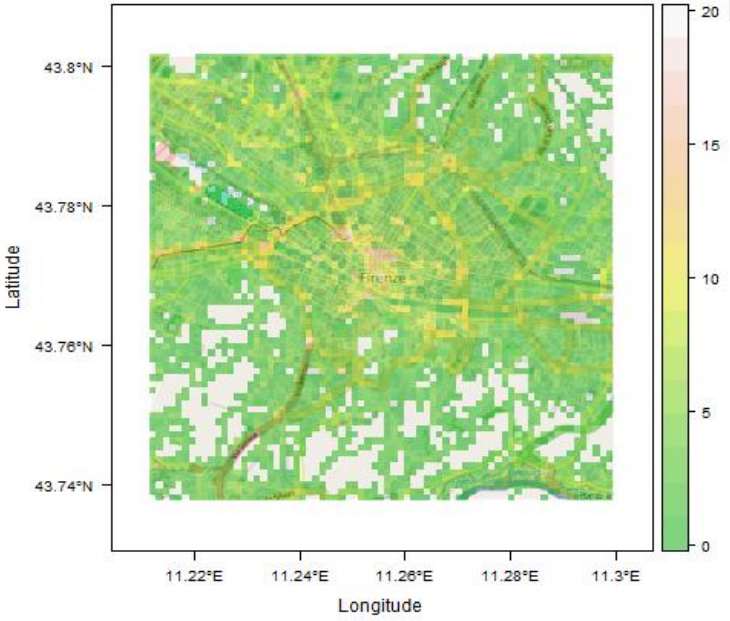


Local complexity
(pixel-area where complexity is lower <2)

Aerial view Florence spatial resolution 100 m

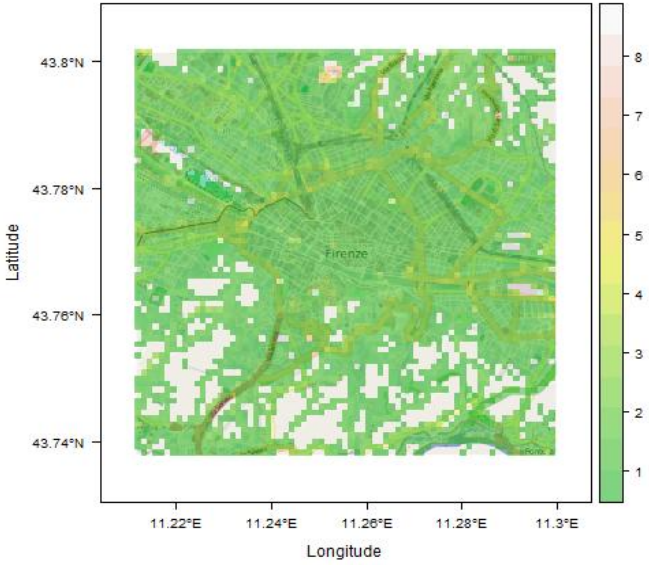


Feature density

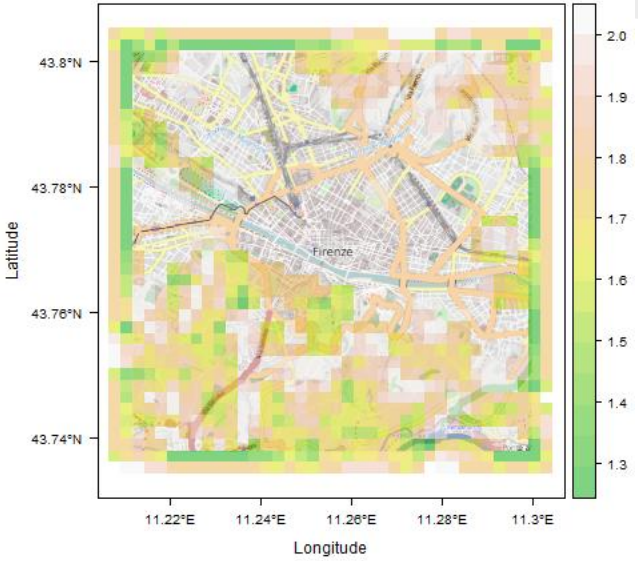


Users density

Aerial view Florence spatial resolution 100 m

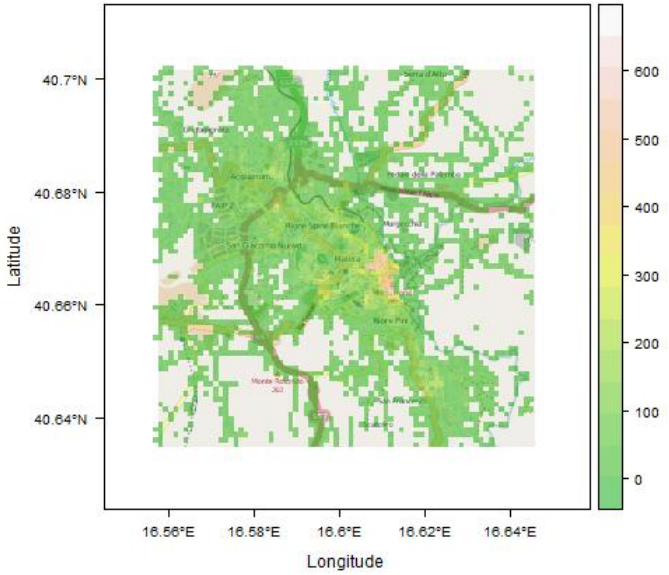


Version Count density

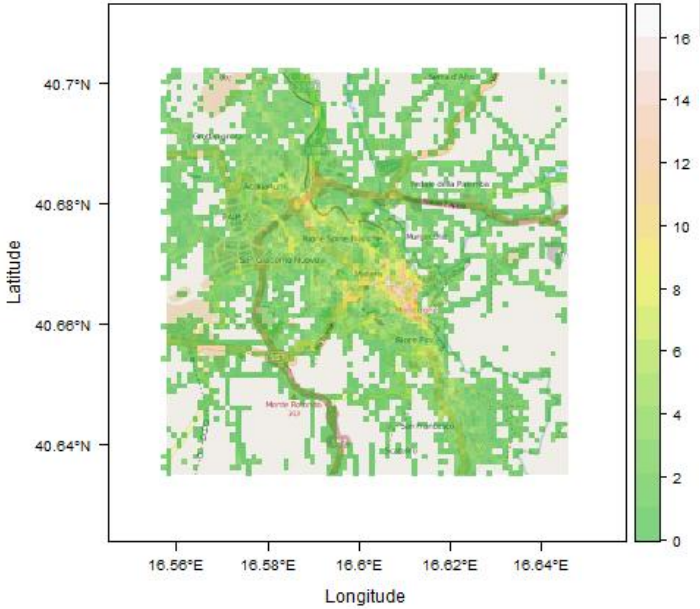


Local complexity
(pixel-area where complexity is lower <2)

Aerial view Matera spatial resolution 100 m

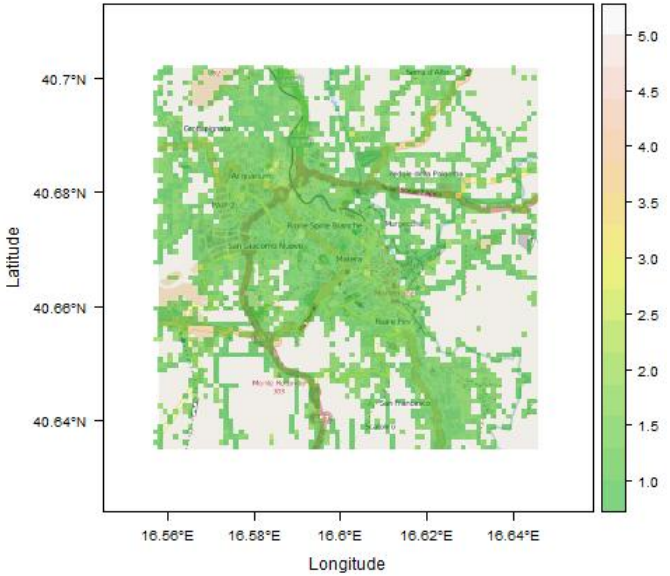


Feature density

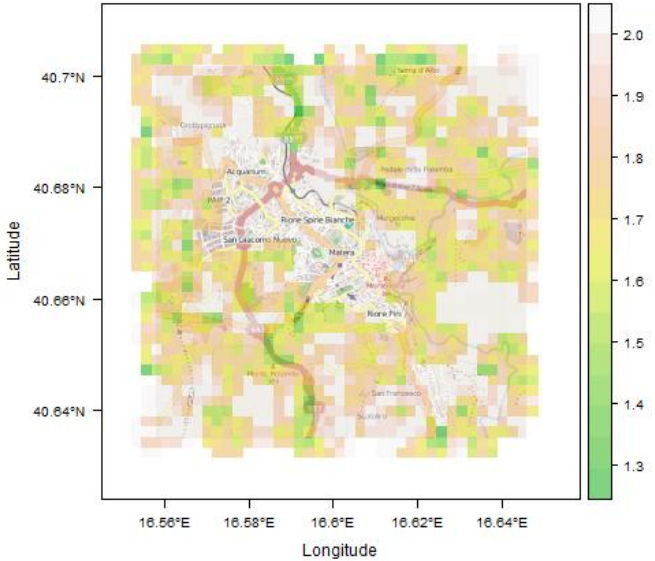


Users density

Areal view Matera spatial resolution 100 m

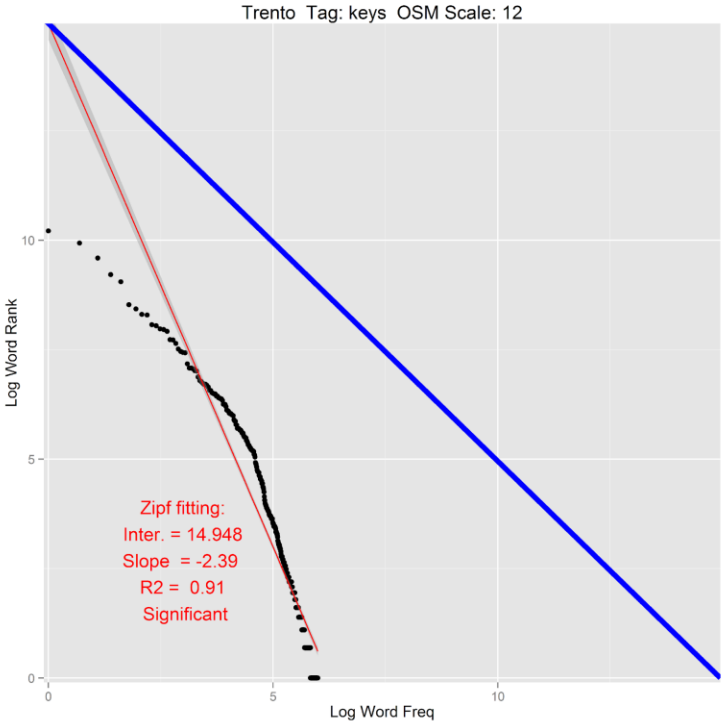


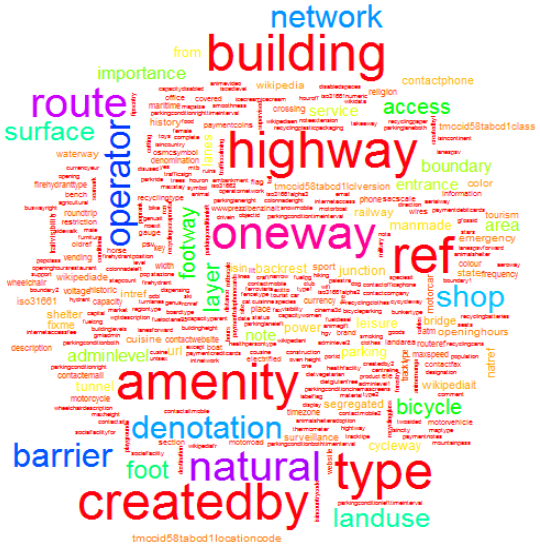
Version Count density



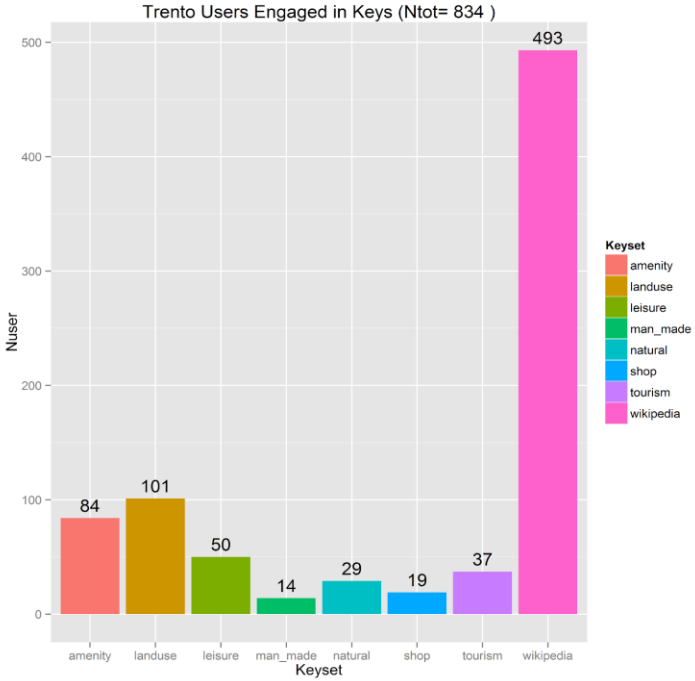
Local complexity
(pixel-area where complexity is lower <2)

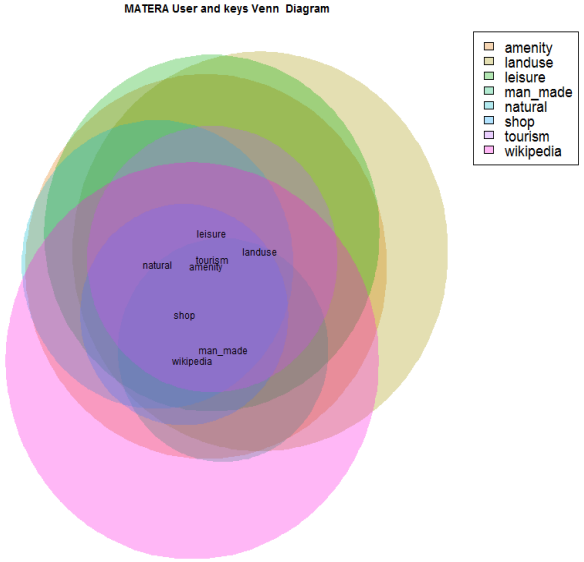
Lexical analysis (1) Trento



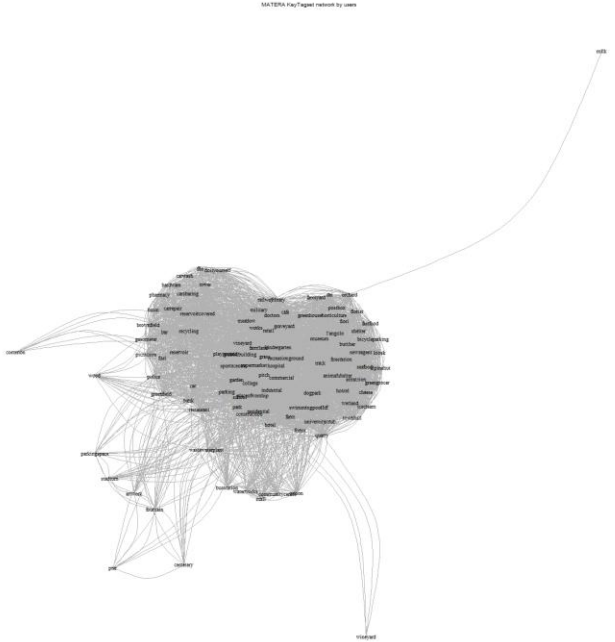


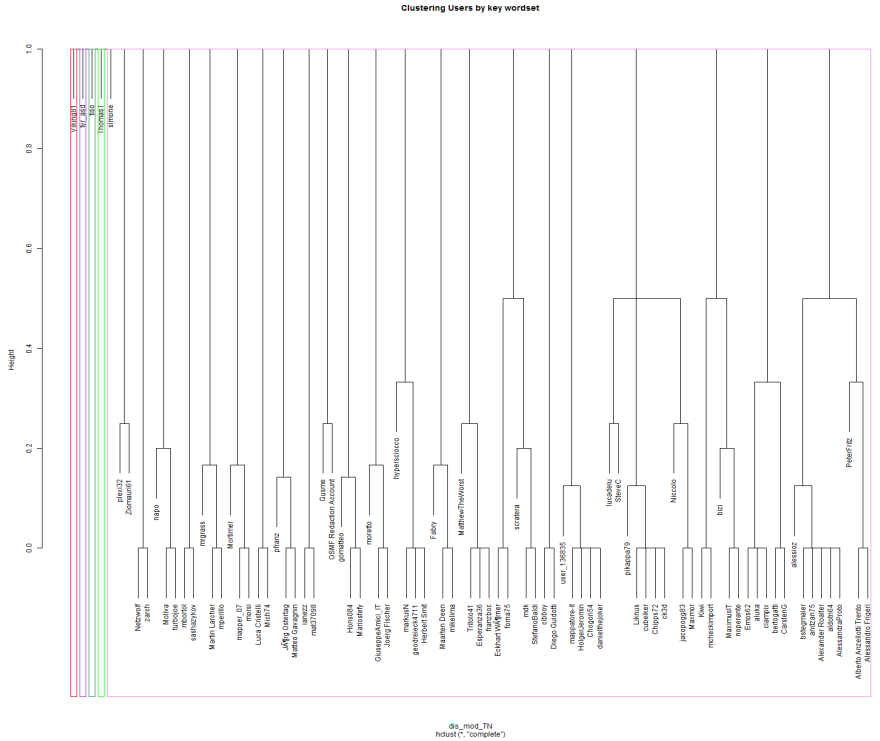
Lexical analysis (2) Trento



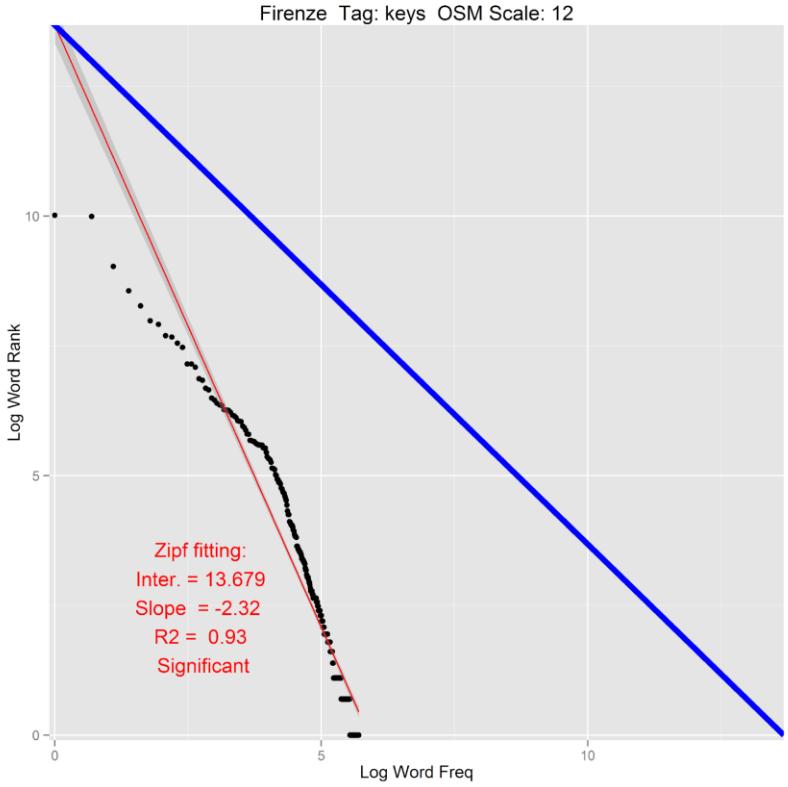


Lexical analysis (2) Trento

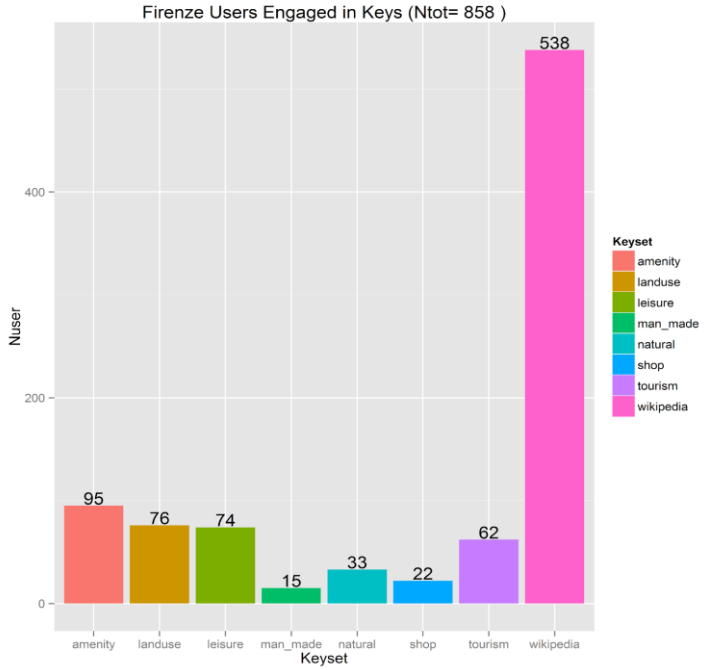


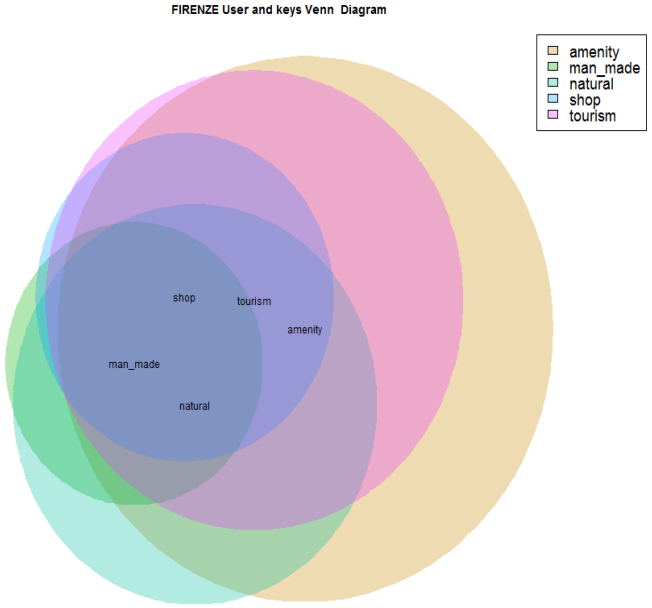


Lexical analysis (1) Florence

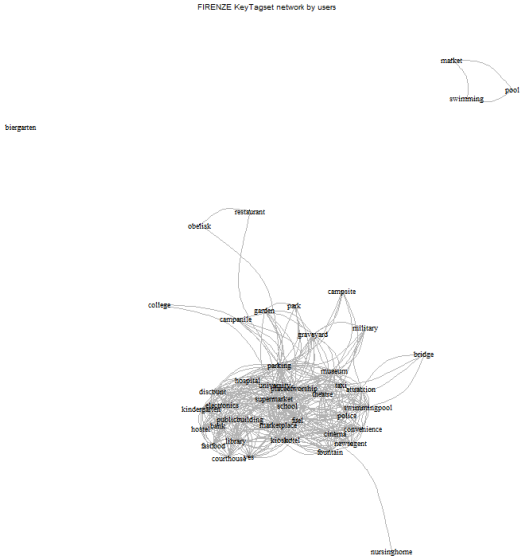


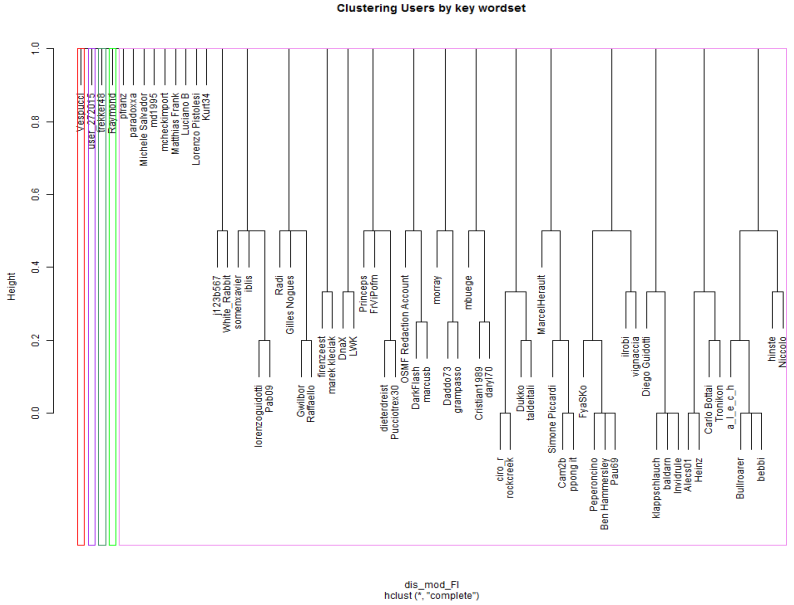
Lexical analisys (2) Florence



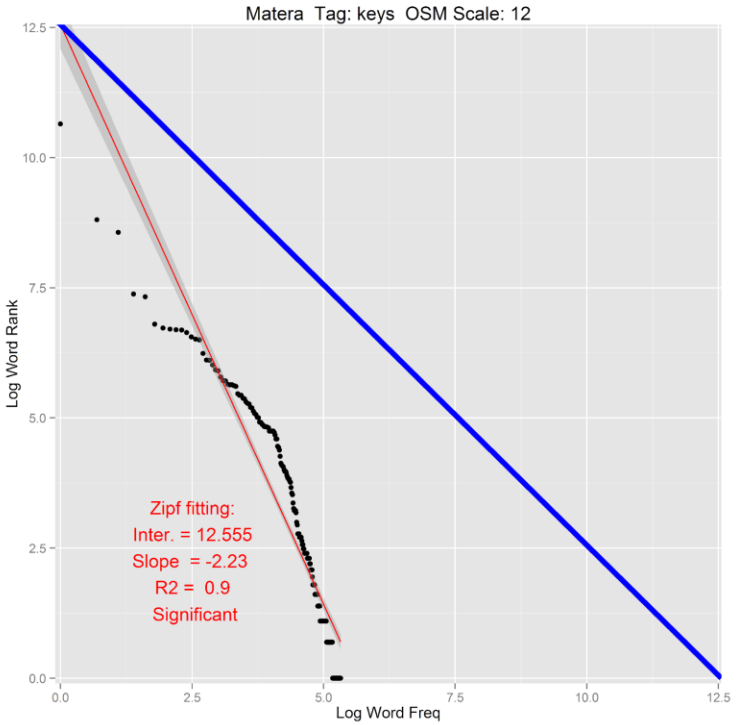


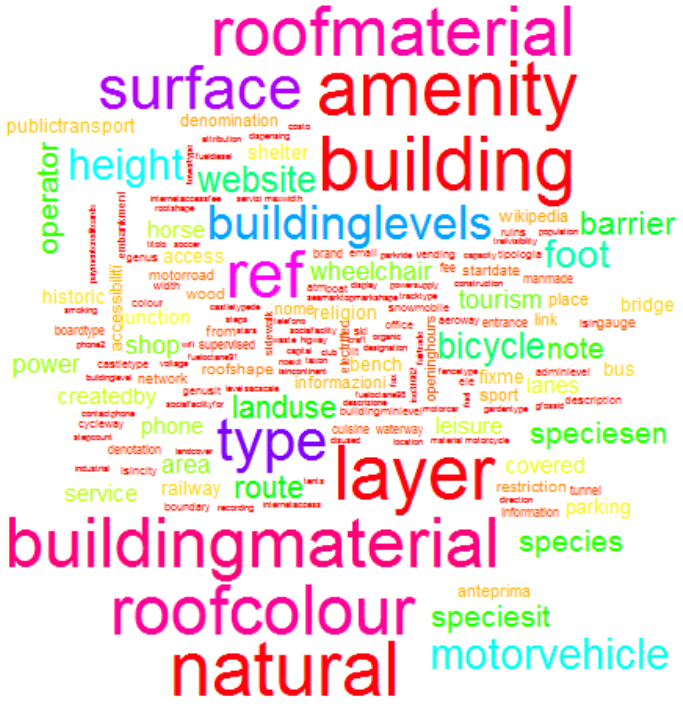
Lexical analysis (2) Firenze



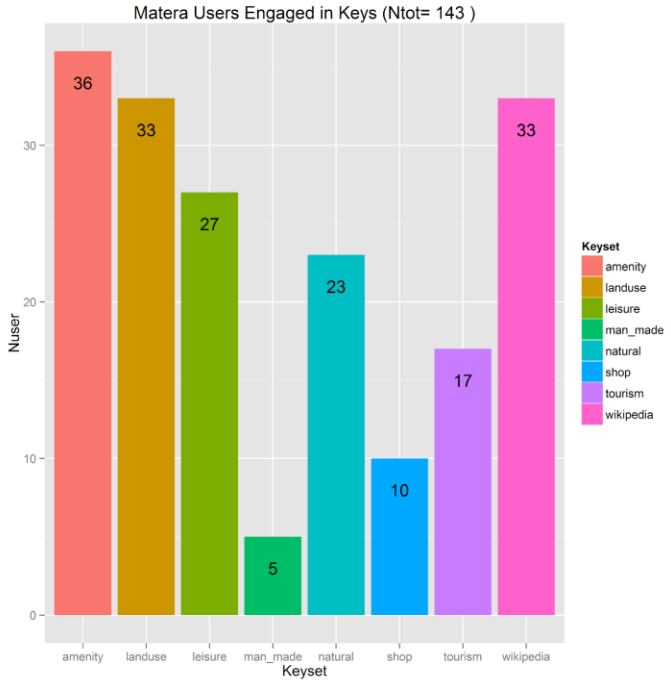


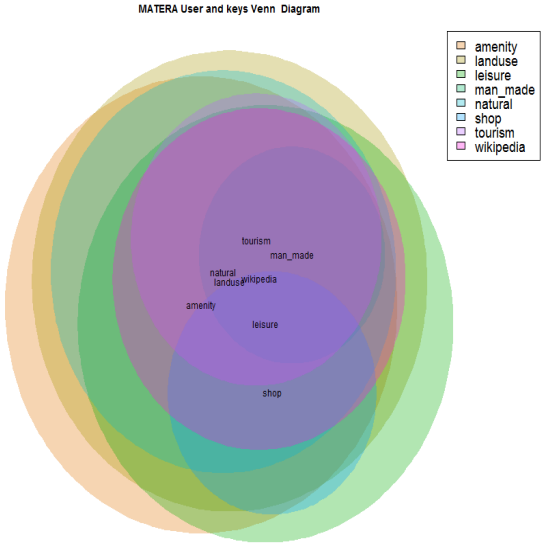
Lexical analysis (1) Matera



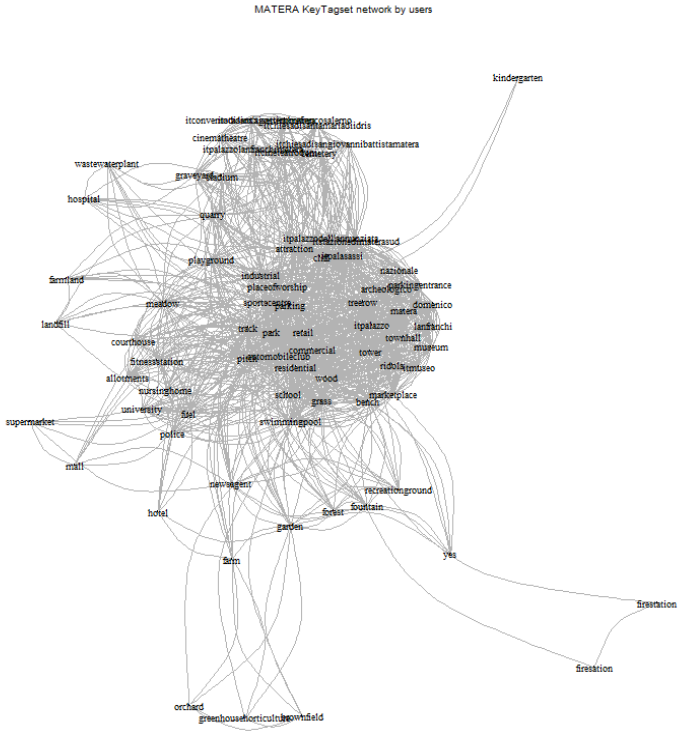


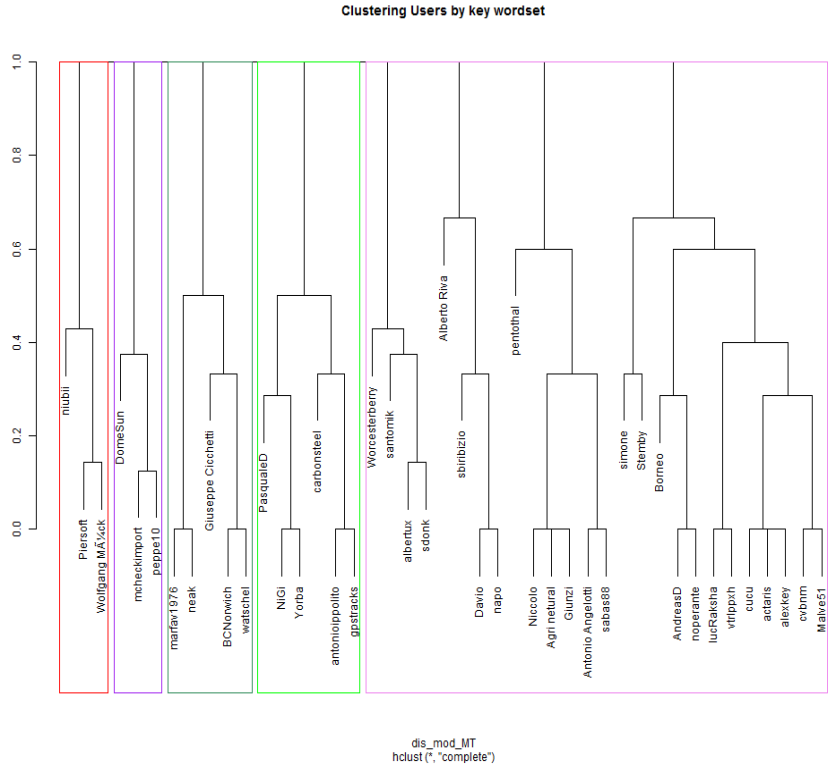
Lexical analysis (2) Matera



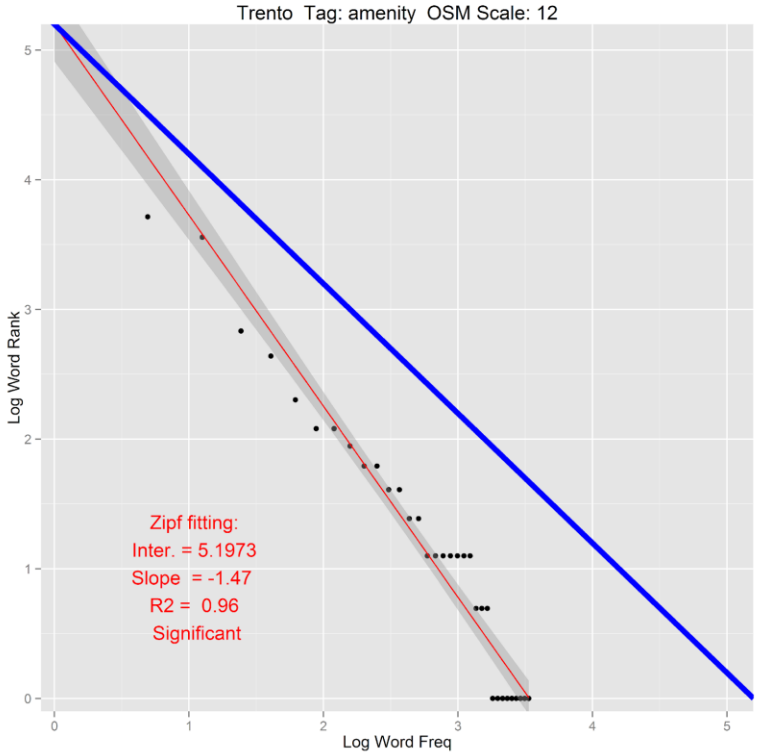


Lexical analysis (2) Matera





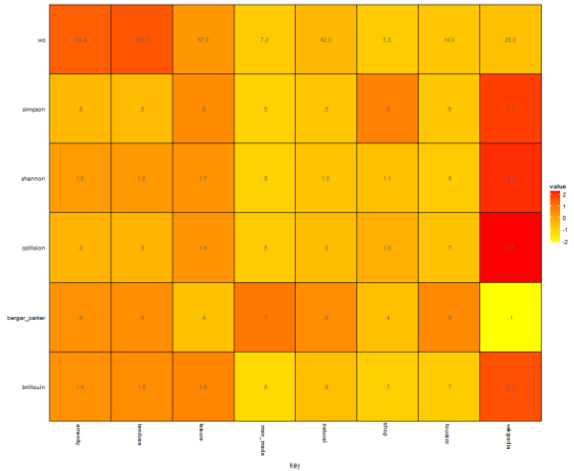
Tags insight Amenity in Trento



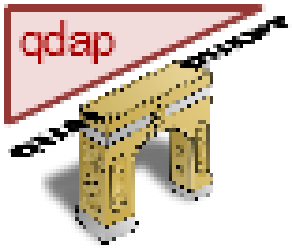
Information entropy Diversity

Using several diversity index corpus of values for keys is possible to see the different use of tags in cities

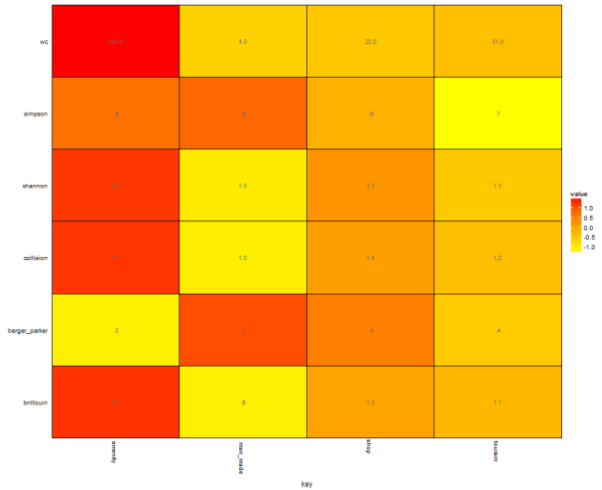
MATERA



"shop","amenity","tourism","man_made","natural","leisure","landuse","wikipedia"



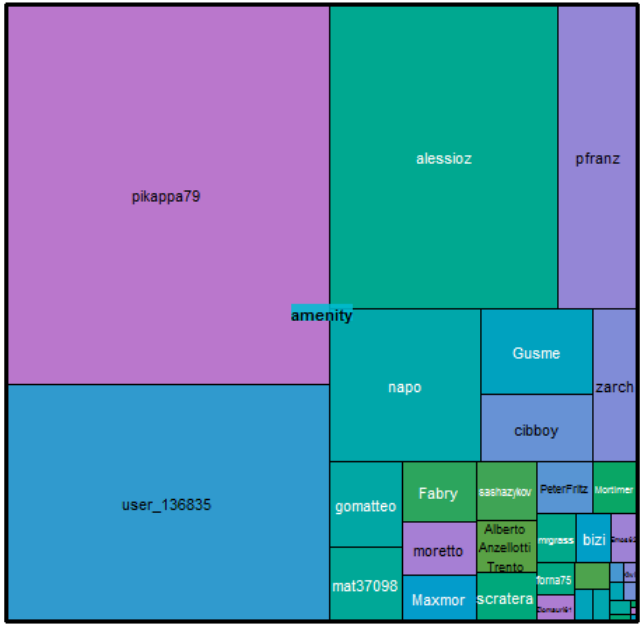
FIRENZE



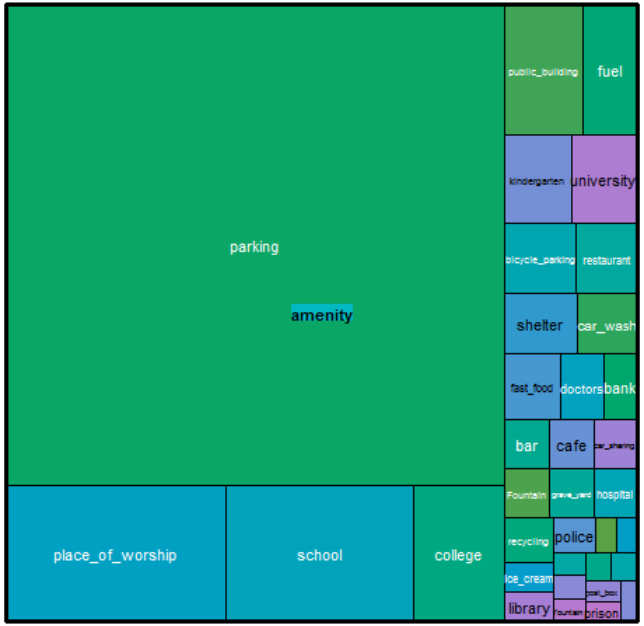
"shop","amenity","tourism","man_made"

Tags insight Amenity in Trento

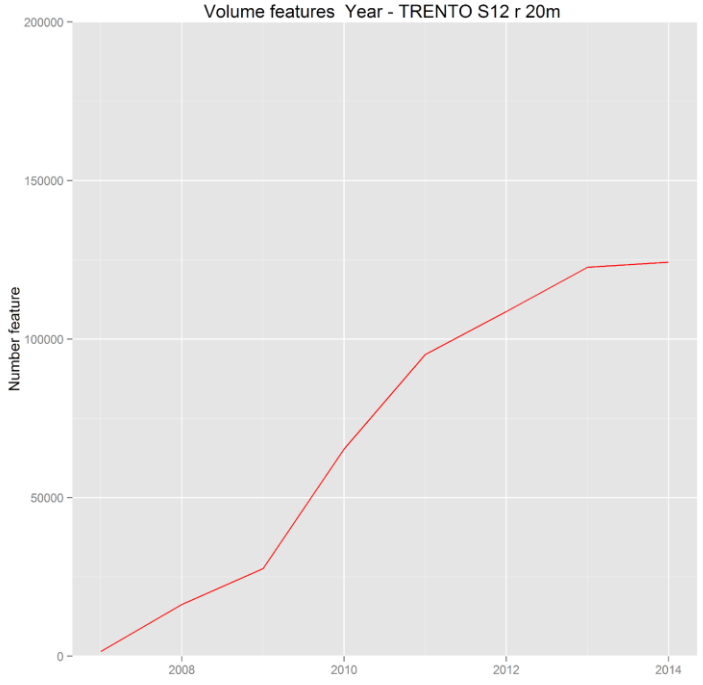
Trento amenity TreeMap Key - User

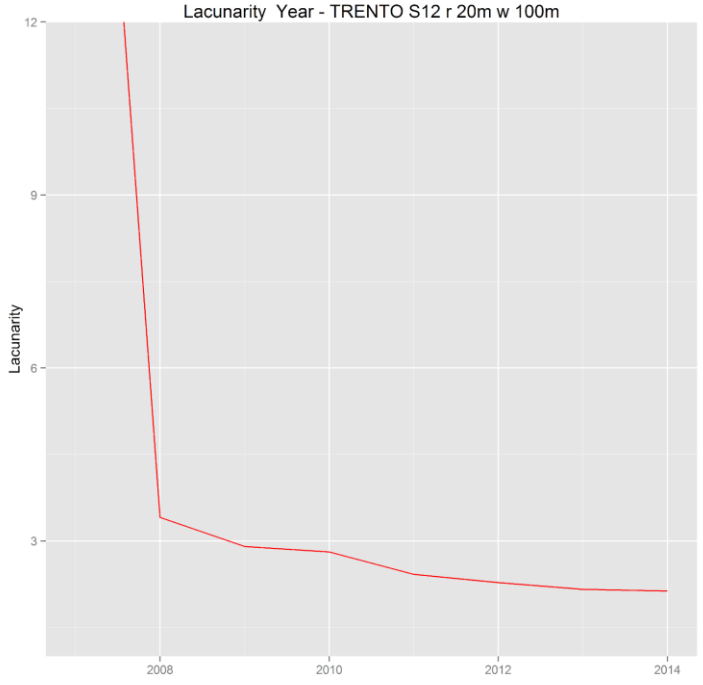


Trento amenity TreeMap Key with Value

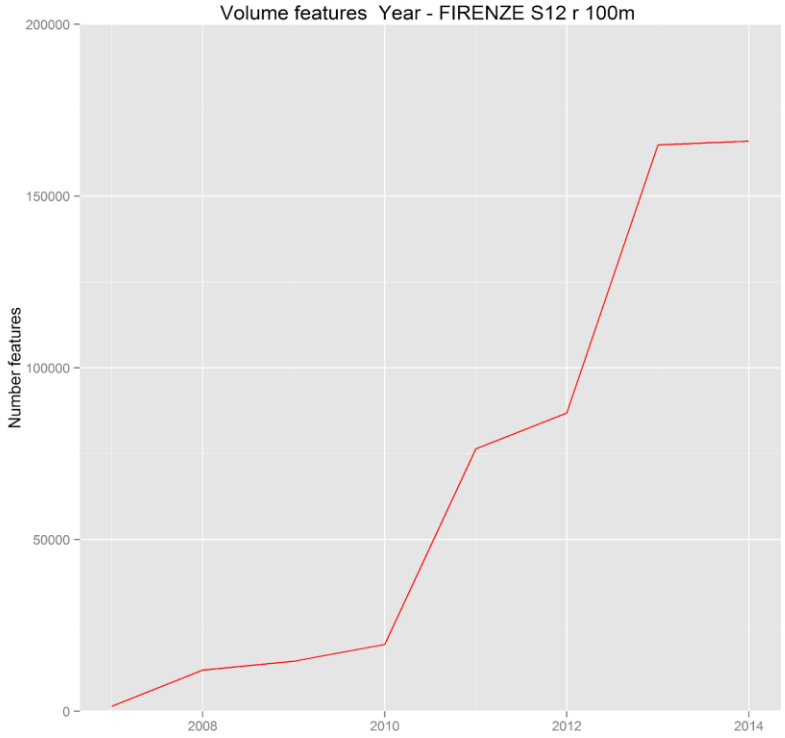


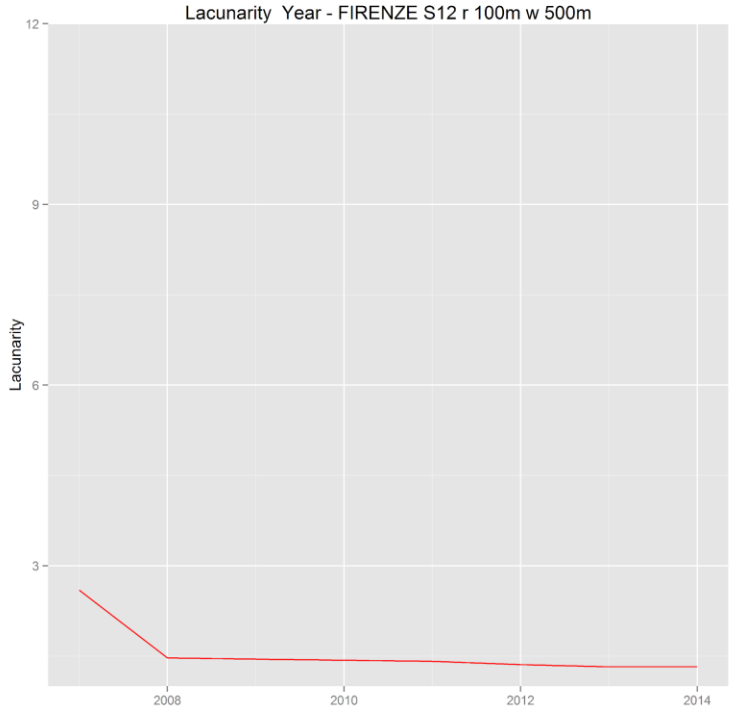
Temporal view Trento



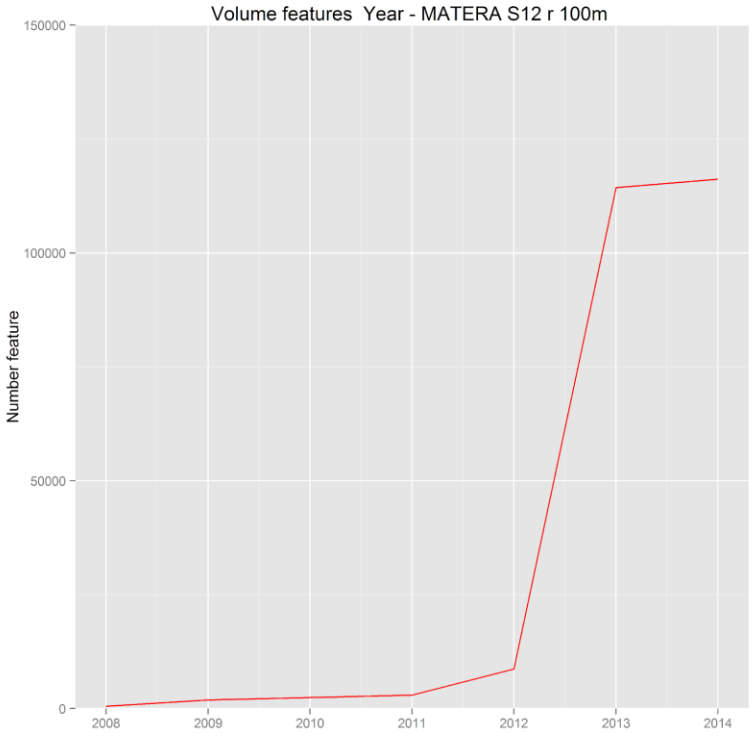


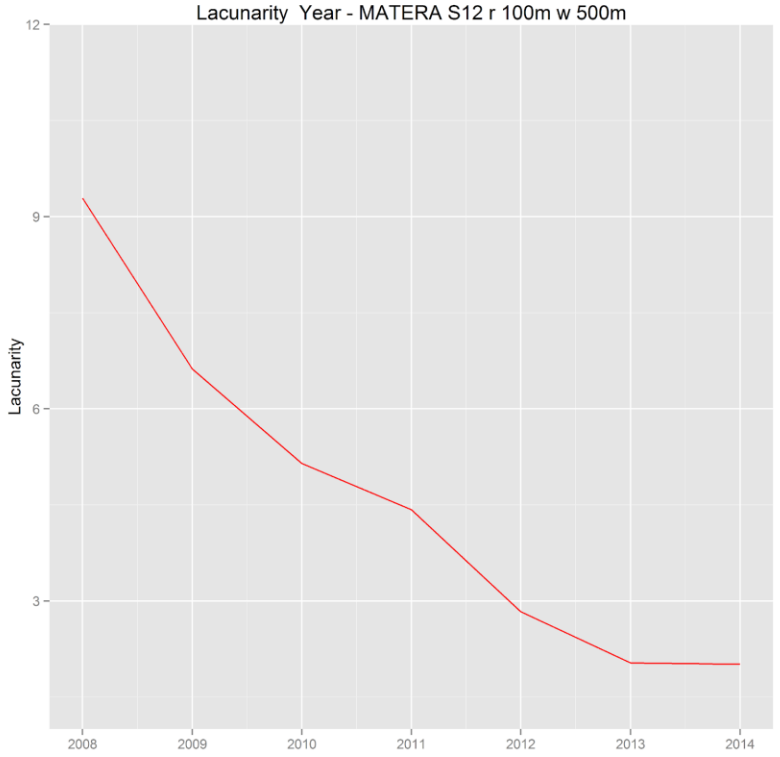
Temporal view Firenze





Temporal view Matera





Main Findings

• Spatial complexities in OSM for a specific area could be detected and monitored in space and time by using complexity metric.

•The lacunarity decay show well the OSM informativity growth but its reliability depends by the spatial scale used. In densely mapped areas small resolutions are required (20 m or 10 m).

•Lacunarity thresholds for OSM quality assessment needs further investigations in relation to the zoom level involved and the keys (tags) monitored.

•Local fractal dimension indicates well where are area with a low complexity.

Main Findings

•Lexical statistical frameworks works with OSM data and describe their informativity and the differences that exist among areal communities.

•Textual informativity parameters show the general terms' abundancy in OSM and demonstrate that is a really rich informative environment .

• Areal keyset, and only in certain tags, follow a natural language distribution (Zipf's law emerges!) and integrating Information entropy analysis for different spatial scales (zoom level) is possible to infer on information suitability of the area done by OSM users' community.

•All kind of investigation must ever to take into account population and user density

Conclusions

- Spatial and textual “complexity” parameters seems promising tools to help the assessment of data quality in specific area.

- They main role is to quantify the amount but need to be linked with other areal metrics (population & mappers density, OSM feature density).

- The need is to define proper metrics linked to these parameters presented ……to create osm services as well.

- Suggestions are welcome!

Contacts

Thank you!
Contacts:

Alfonso Crisci
mail: :
a.crisci@ibimet.cnr.it

alfcrisci@gmail.com

@alfcrisci

Download presentation

http://www.slideshare.net/alfcrisci/sot-m-eu2014crisci

Appendix Fractal dimension: measure of spatial complexity state

A **fractal dimension** is a ratio providing a statistical index of complexity comparing how detail in a pattern (strictly speaking, a fractal pattern) changes with the scale at which it is measured.
http://en.wikipedia.org/wiki/Fractal_Dimension

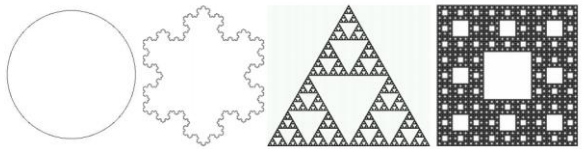
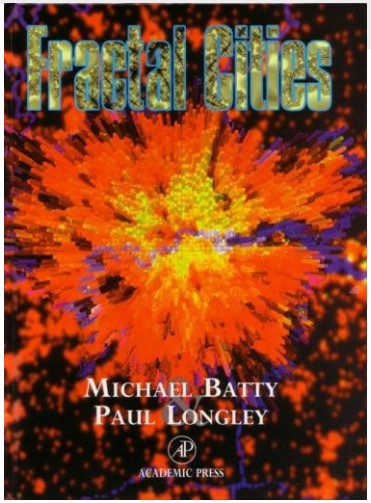


Figure 4.3: Circle (FD=1.000), snow-flake (FD=1.262), Sierpinsky triangle (FD=1.585) and Sierpinsky carpet (FD=1.892) show different levels of complexity.

Images
Marco Diego DOMINIETTO
ETH Zurich
Multimodality Approach To Study The Fractal Physiology Of Tumor Angiogenesis



We need much better statistics that pertain to the different kinds of dynamics and their variation over time and space. (Batty,1994)

Batty, M., and Longley, P. (1994). Fractal Cities: A Geometry of Form and Function, Academic Press, San Diego, CA, at www.fractalcities.org