



POLITECNICO
MILANO 1863

Un approccio open source per la valutazione intrinseca di accuratezza tematica, accuratezza temporale, aggiornamento e lignaggio di OSM

Francesco Frassinelli, Marco Minghini, Maria Antonia Brovelli
Dipartimento di Ingegneria Civile e Ambientale

Merge-IT 2018 – 24 marzo 2018, Torino

Chi sono?

Francesco Frassinelli (Frafra) – <https://frafra.eu>

- Studente di ingegneria geoinformatica al PoliMi
Questo è l'inizio del mio progetto di tesi, pensato sia per la comunità, sia per i ricercatori
- Utente OpenStreetMap dal 2008
- Coordinatore OSM Lombardia e vicepresidente PoliMappers
- Amante del software libero (Linux, Python, Bash, ...)
 - Utente GNU/Linux da 15+ anni, fondatore di ViGLug
 - Amministratore della comunità italiana di Fedora, Ambassador, Maintainer, supporto, correzione bug...

Obiettivo principale

Stimare la qualità dei dati di OpenStreetMap
al fine di ridurre le barriere che ne impediscono
l'adozione

Differenti criteri

Accuratezza spaziale, completezza, temporale,
semantica, lignaggio...

Come stimare la qualità dei dati di OpenStreetMap?

Estrinseco

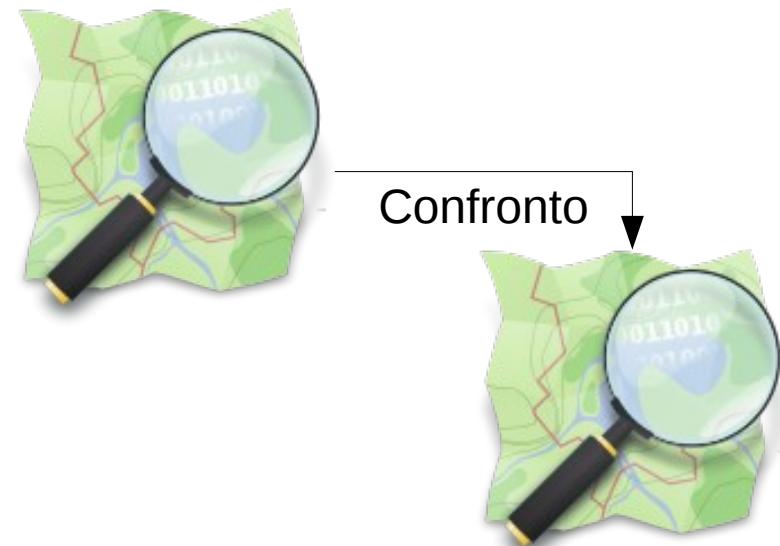
Confrontare OSM ad un dataset esterno



Esempio: cartografia ufficiale

Intrinseco

Confrontare OSM a se stesso, senza utilizzare altre risorse



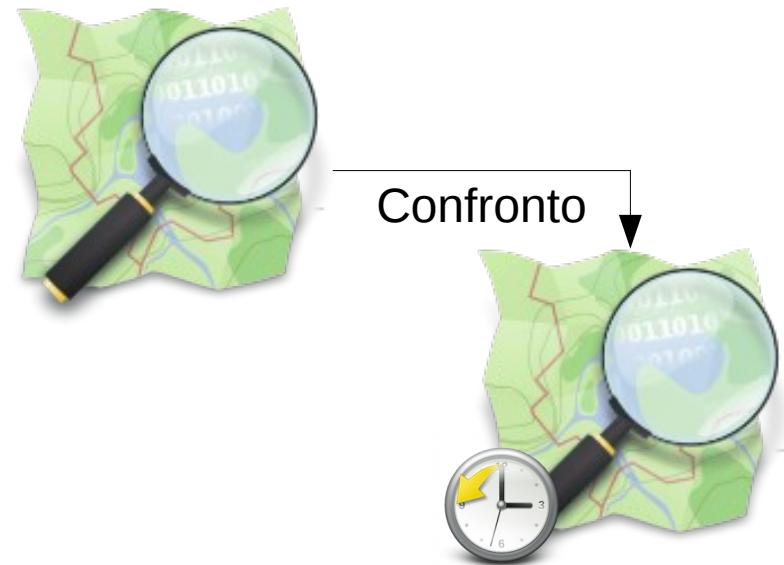
Esempio: violazione uso tag

Qualità temporale

Obiettivi

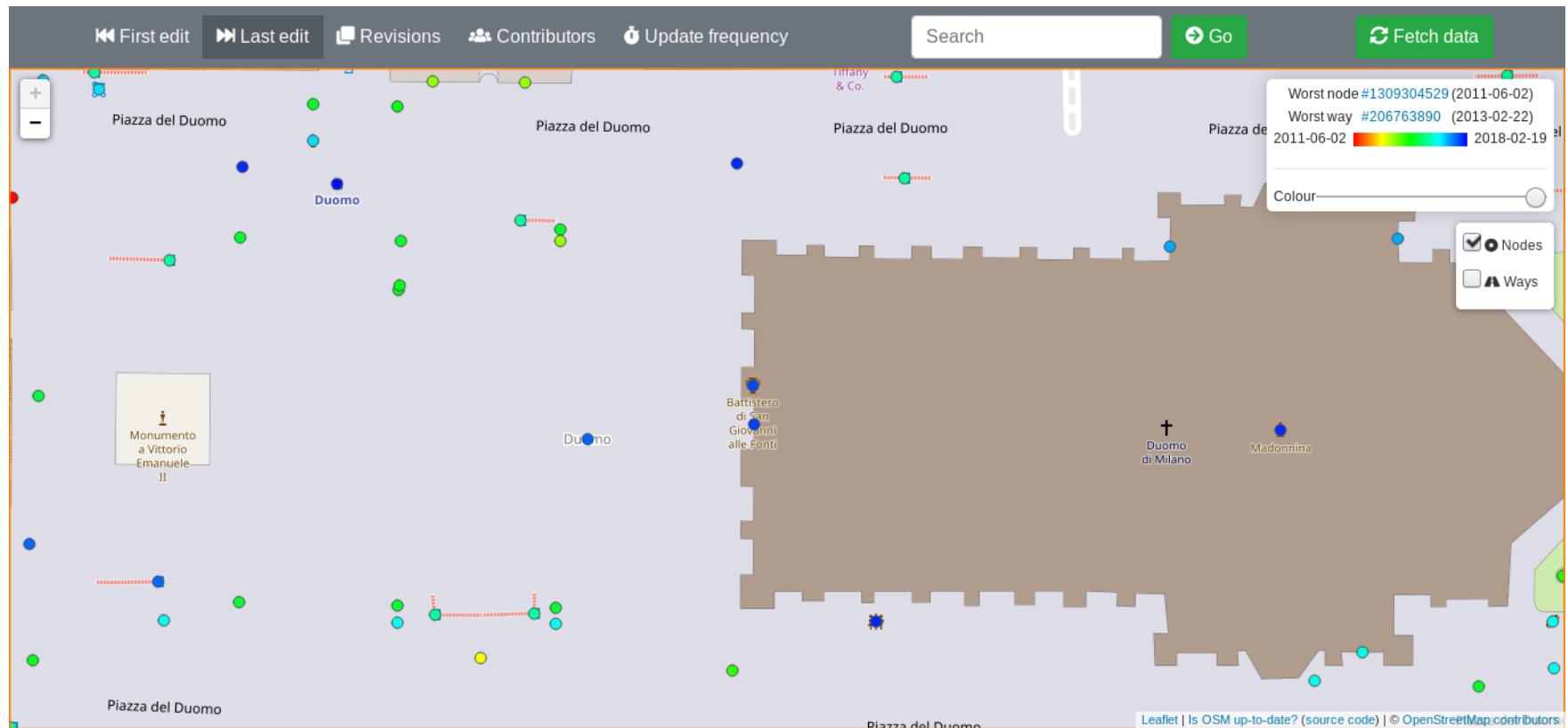
- Realizzazione webapp per stima aggiornamento, accuratezza temporale, lignaggio
- Estensione dell'analisi al caso italiano (dati aggregati per comune)
- Accuratezza semantica e trend nell'aggiunta dei tag

Studio dell'evoluzione del dato
(gardening, ...)

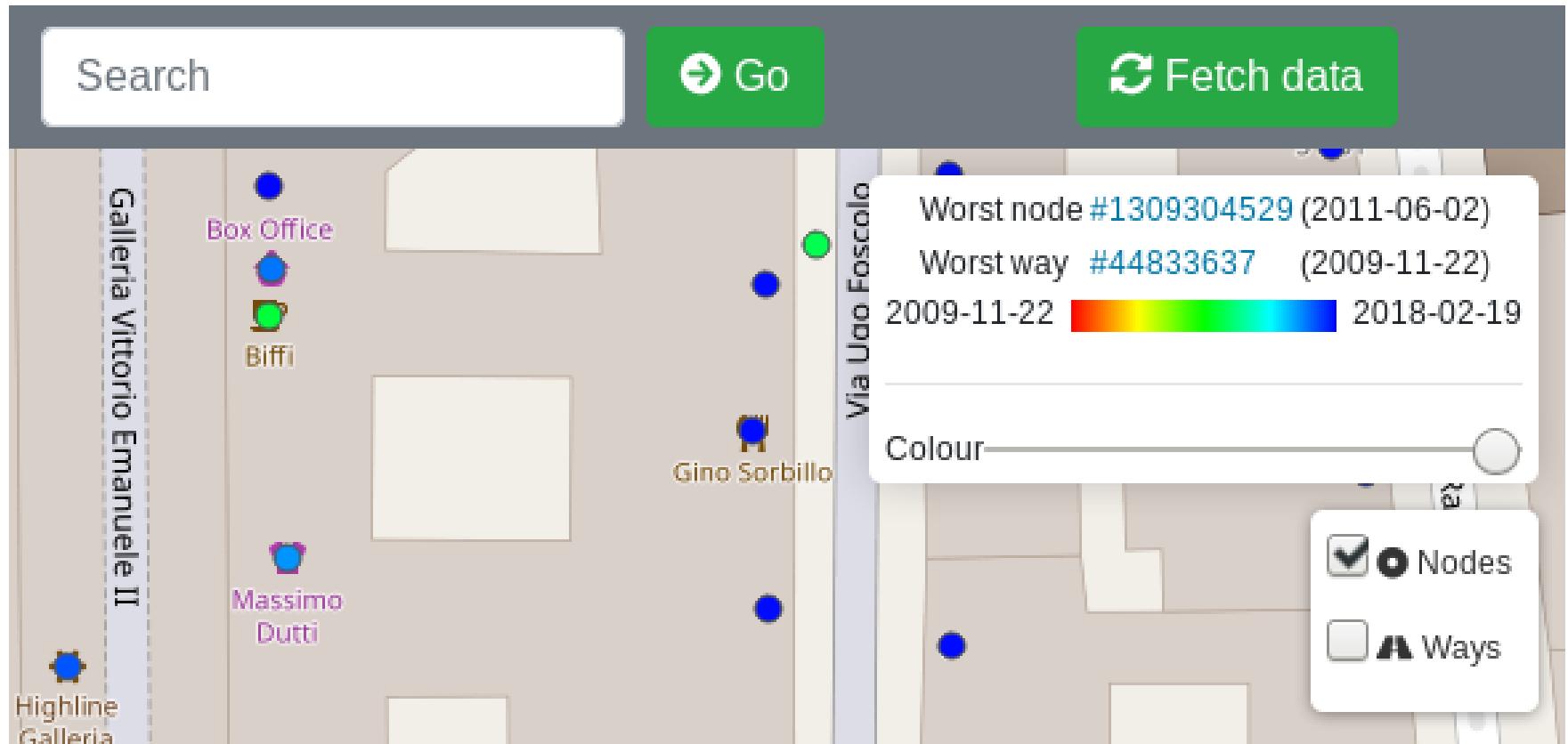


Is OSM up-to-date? - Schermata principale

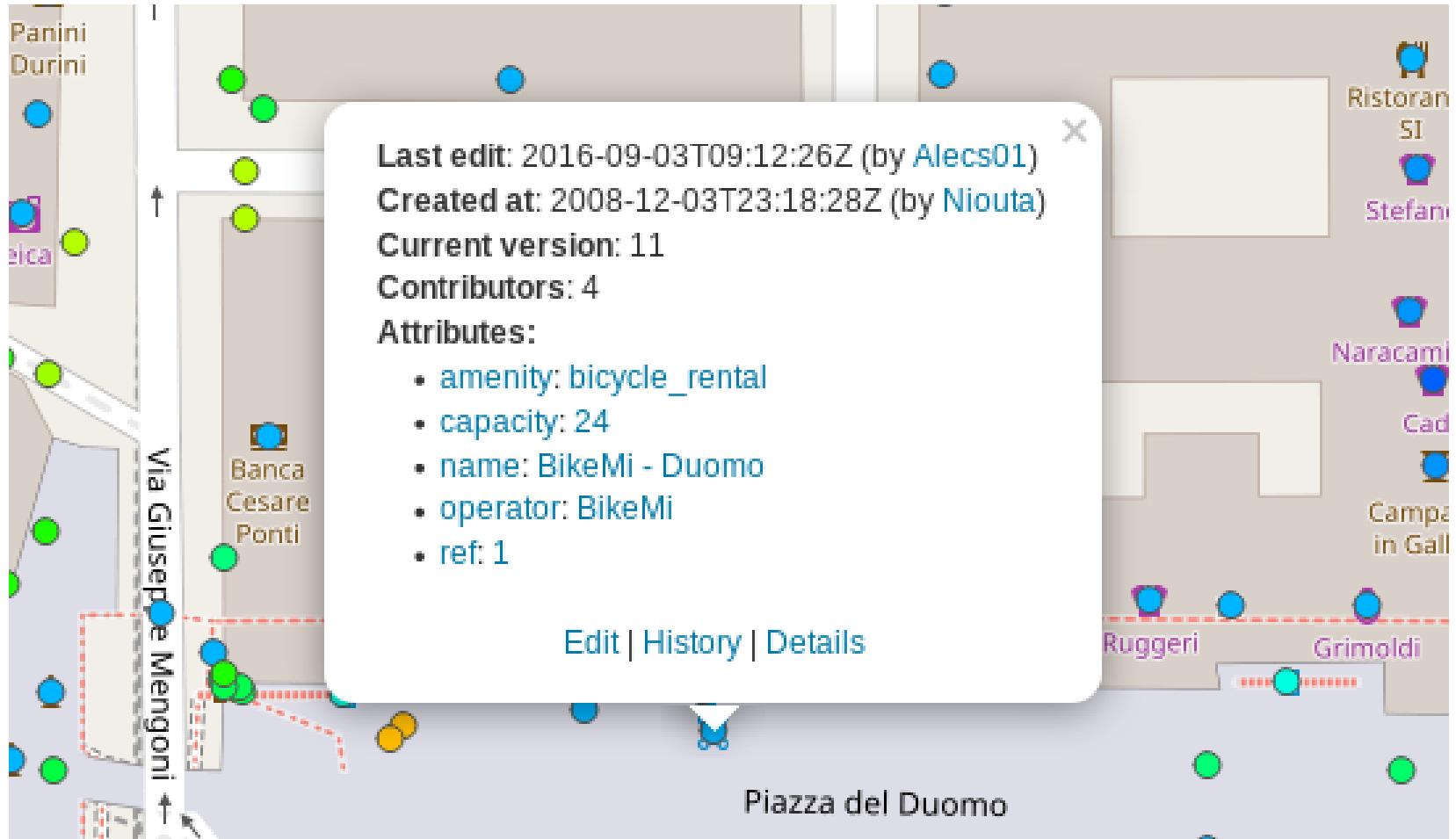
<https://is-osm-updated.frafra.eu>



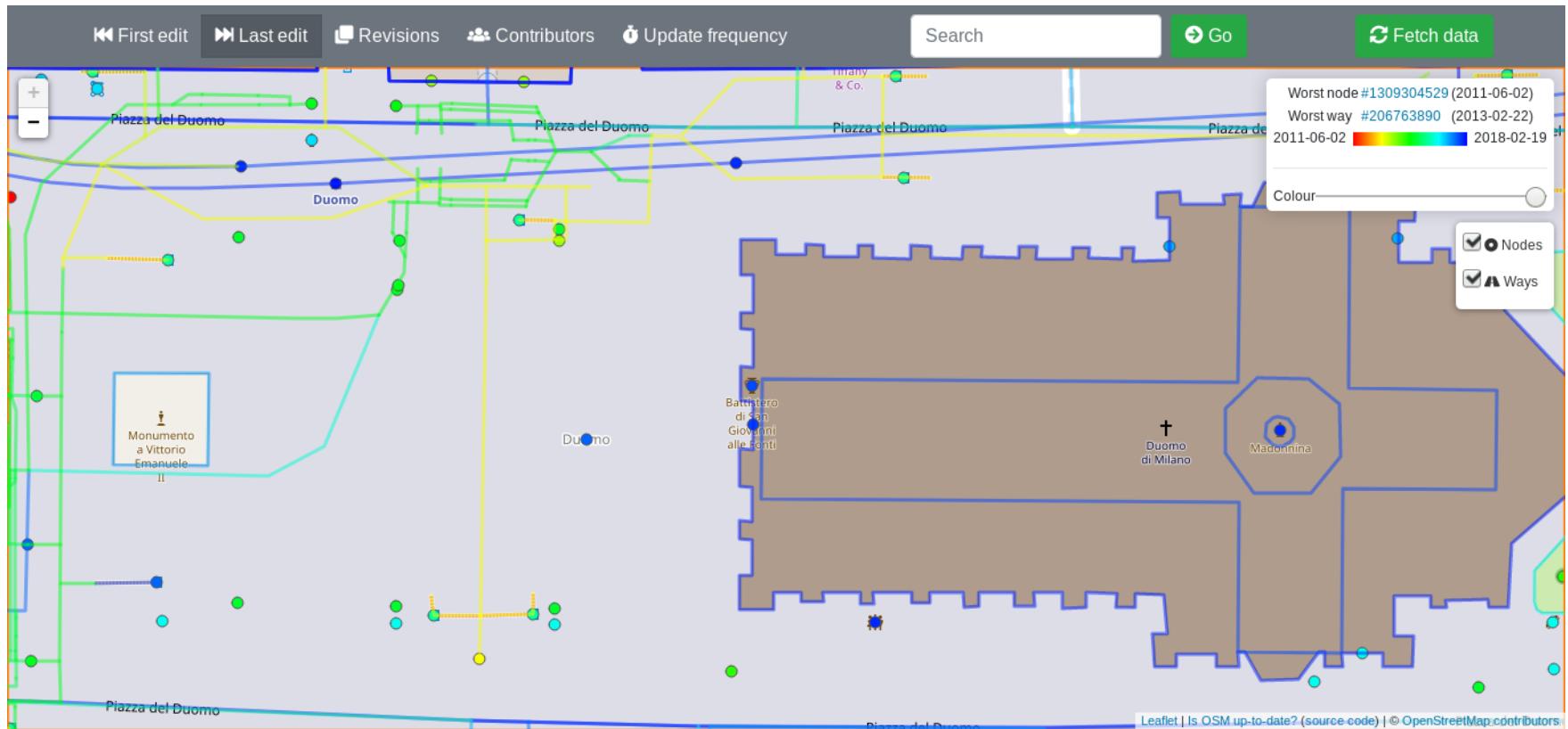
Is OSM up-to-date? - Controlli, ricerca e legenda



Is OSM up-to-date? - Controlli, ricerca e legenda



Is OSM up-to-date? - Schermata principale (nodi e vie)



Is OSM up-to-date? - Criteri di classificazione

First edit Last edit Revisions Contributors Update frequency

- Prima modifica
- Ultima modifica
- Numero di revisioni (massimo numero di versione)
- Numero di contributori
- Frequenza di aggiornamento

Is OSM up-to-date? - Tecnologie utilizzate

OSM API v0.6

- Caricamento
./map?bbox=...
- ./history: gruppi
di 100 elementi

Librerie JS/CSS

-  Leaflet
-  Bootstrap
- Font Awesome 

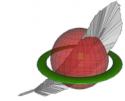


- HUG 
- Dugong: HTTP pipelining
- asyncio, ...

uWSGI

- REST API (HUG)
- File statici

Spatialite



- OSM XML → SQL
- Processamento...
- SQL → GeoJSON



- Deploy rapido
- Immagine basata su



Is OSM up-to-date? - Riferimenti

Applicazione: <https://is-osm-updated.frafra.eu>

Codice sorgente: <https://github.com/frafra/is-osm-updated>

Wiki: https://wiki.openstreetmap.org/wiki/Is_OSM_up-to-date



Analisi aggregata sul territorio italiano

Quali nodi vengono considerati?

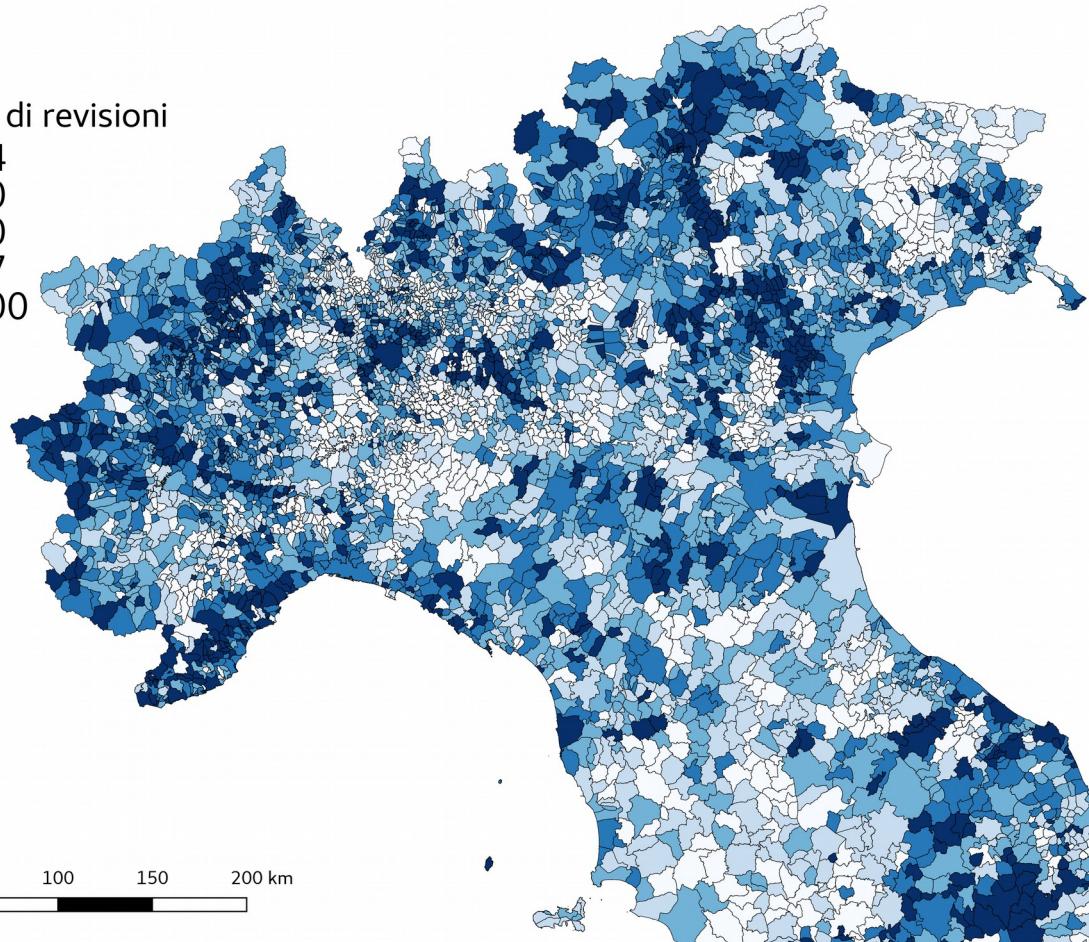
- Qualsiasi nodo che si distingue dal precedente per almeno un tag aggiunto, modificato o rimosso
 - Variazioni legate alla sola posizione sono ignorate
- Non c'è distinzione tra nodi isolati o appartenenti a percorsi, poligoni o relazioni
- Non sono considerati i nodi cancellati

Numero medio di revisioni – Nord Italia

Legenda

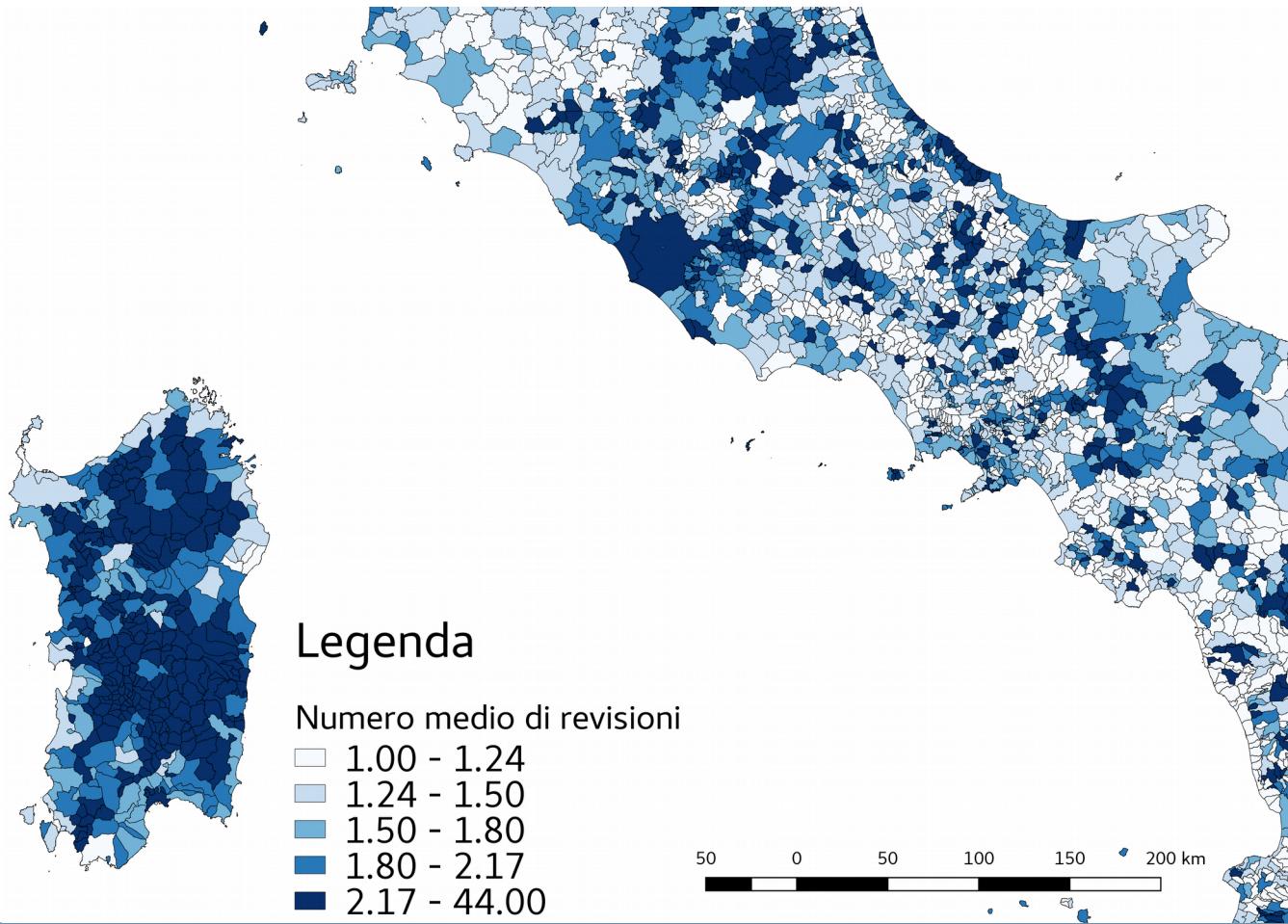
Numero medio di revisioni

- 1.00 - 1.24
- 1.24 - 1.50
- 1.50 - 1.80
- 1.80 - 2.17
- 2.17 - 44.00



50 0 50 100 150 200 km

Numero medio di revisioni – Centro/Sud Italia

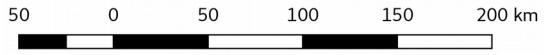


Numero medio di revisioni – Sud Italia

Legenda

Numero medio di revisioni

- 1.00 - 1.24
- 1.24 - 1.50
- 1.50 - 1.80
- 1.80 - 2.17
- 2.17 - 44.00

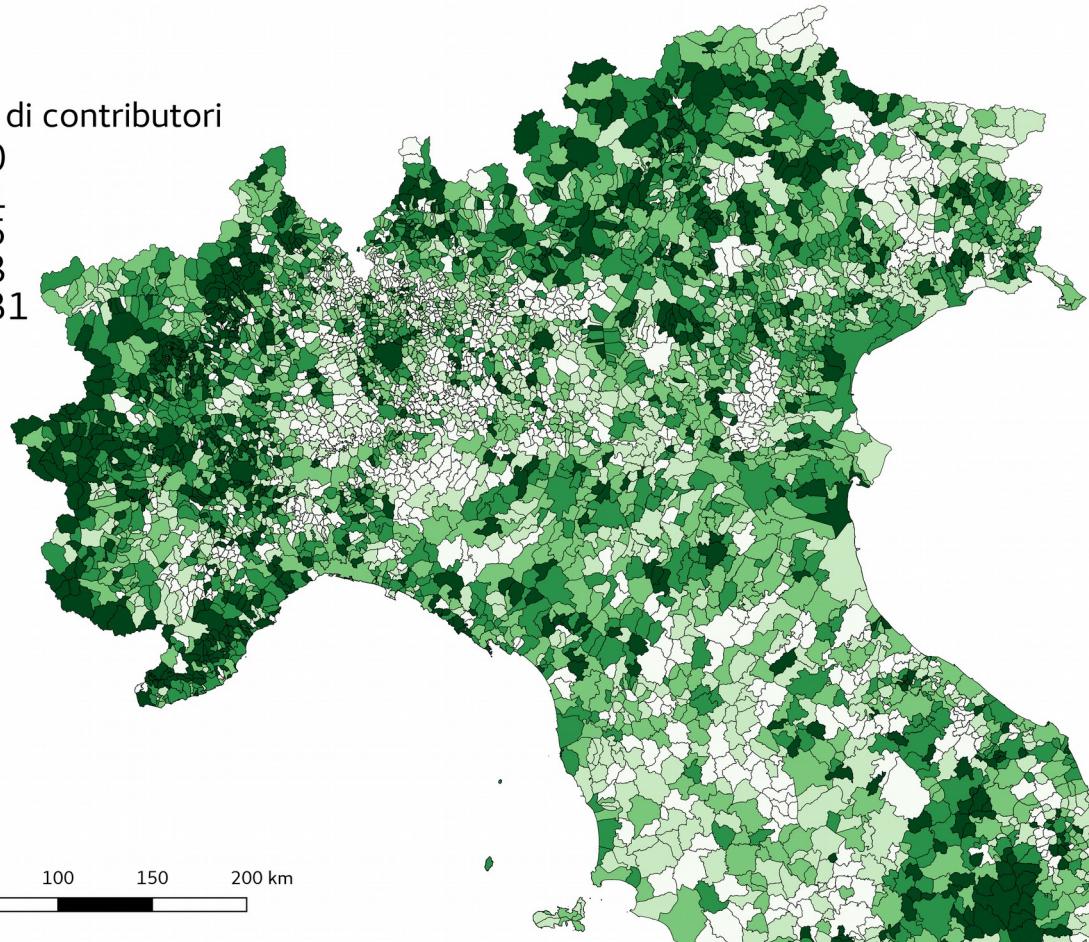


Numero medio di contributori – Nord Italia

Legenda

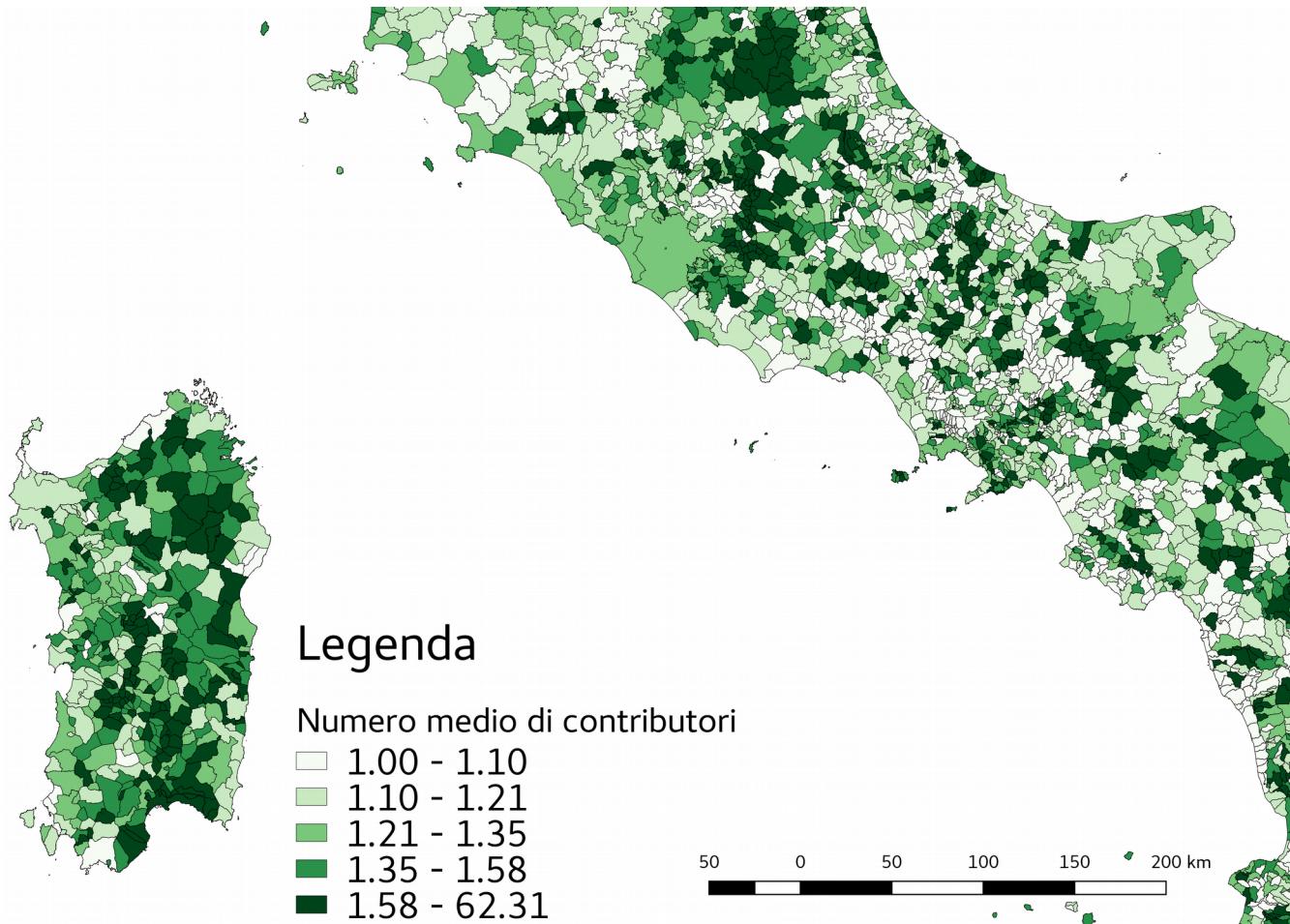
Numero medio di contributori

- 1.00 - 1.10
- 1.10 - 1.21
- 1.21 - 1.35
- 1.35 - 1.58
- 1.58 - 62.31



50 0 50 100 150 200 km

Numero medio di contributori – Centro/Sud Italia

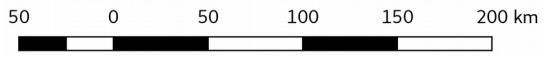


Numero medio di contributori – Centro/Sud Italia

Legenda

Numero medio di contributori

- 1.00 - 1.10
- 1.10 - 1.21
- 1.21 - 1.35
- 1.35 - 1.58
- 1.58 - 62.31

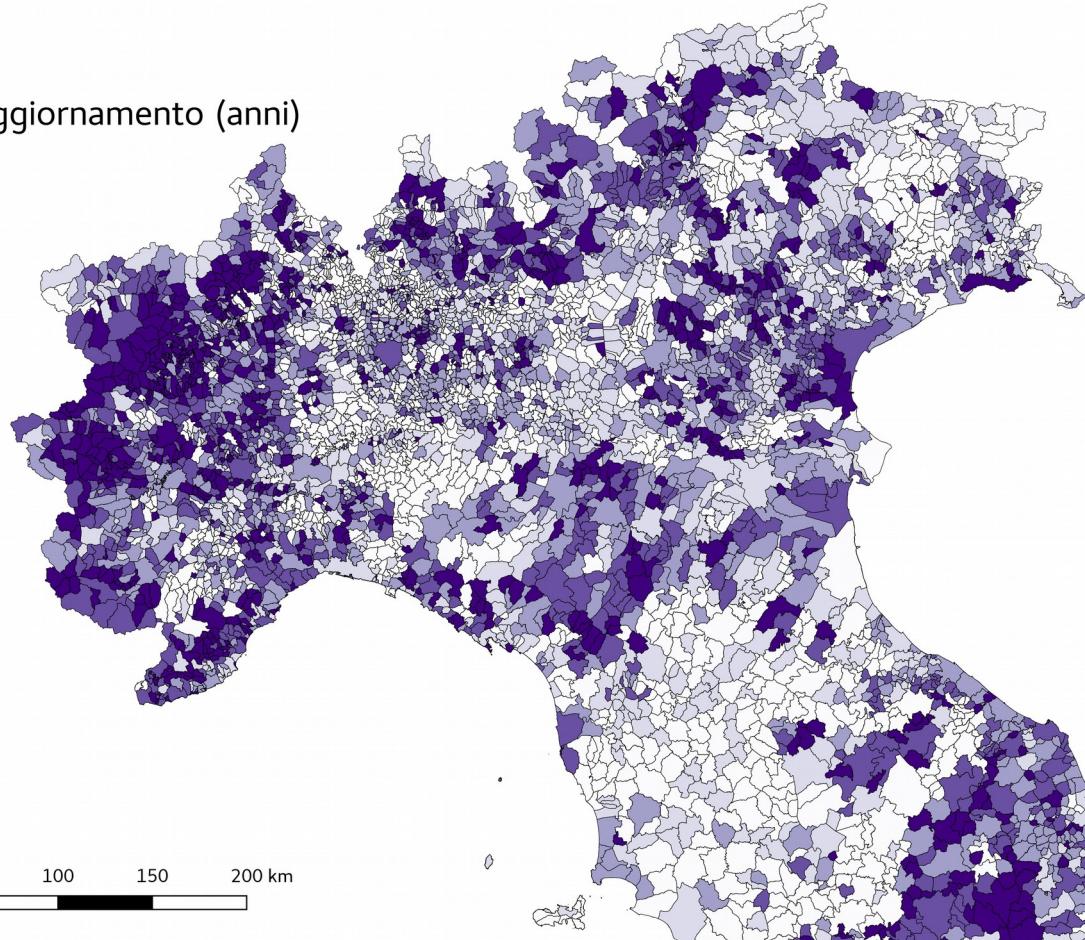


Frequenza media di aggiornamento – Nord Italia

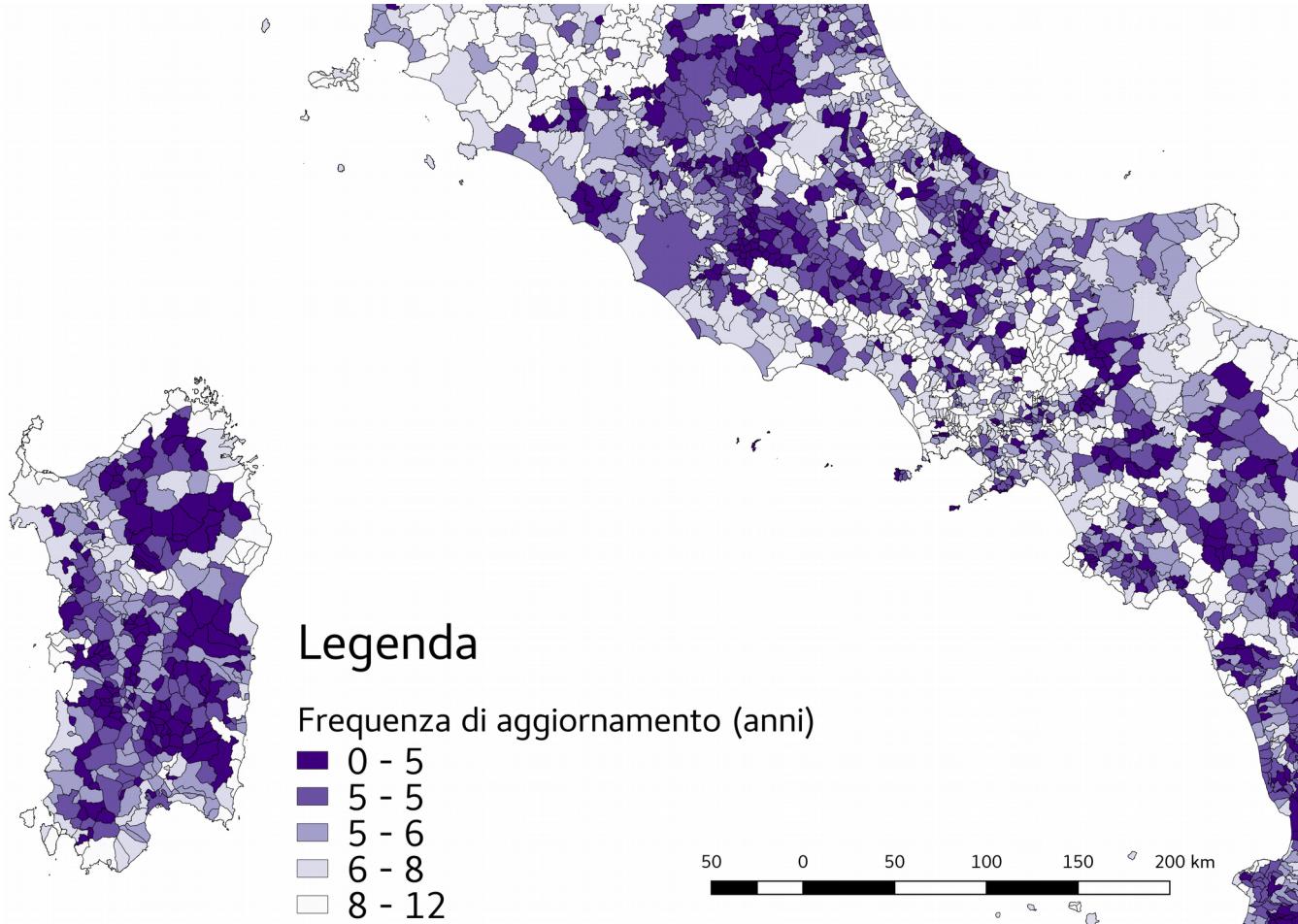
Legenda

Frequenza di aggiornamento (anni)

- 0 - 5
- 5 - 5
- 5 - 6
- 6 - 8
- 8 - 12



Frequenza media di aggiornamento – Centro/Sud Italia

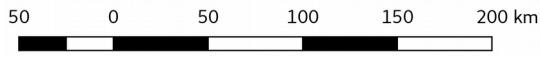


Frequenza media di aggiornamento – Sud Italia

Legenda

Frequenza di aggiornamento (anni)

- 0 - 5
- 5 - 5
- 5 - 6
- 6 - 8
- 8 - 12

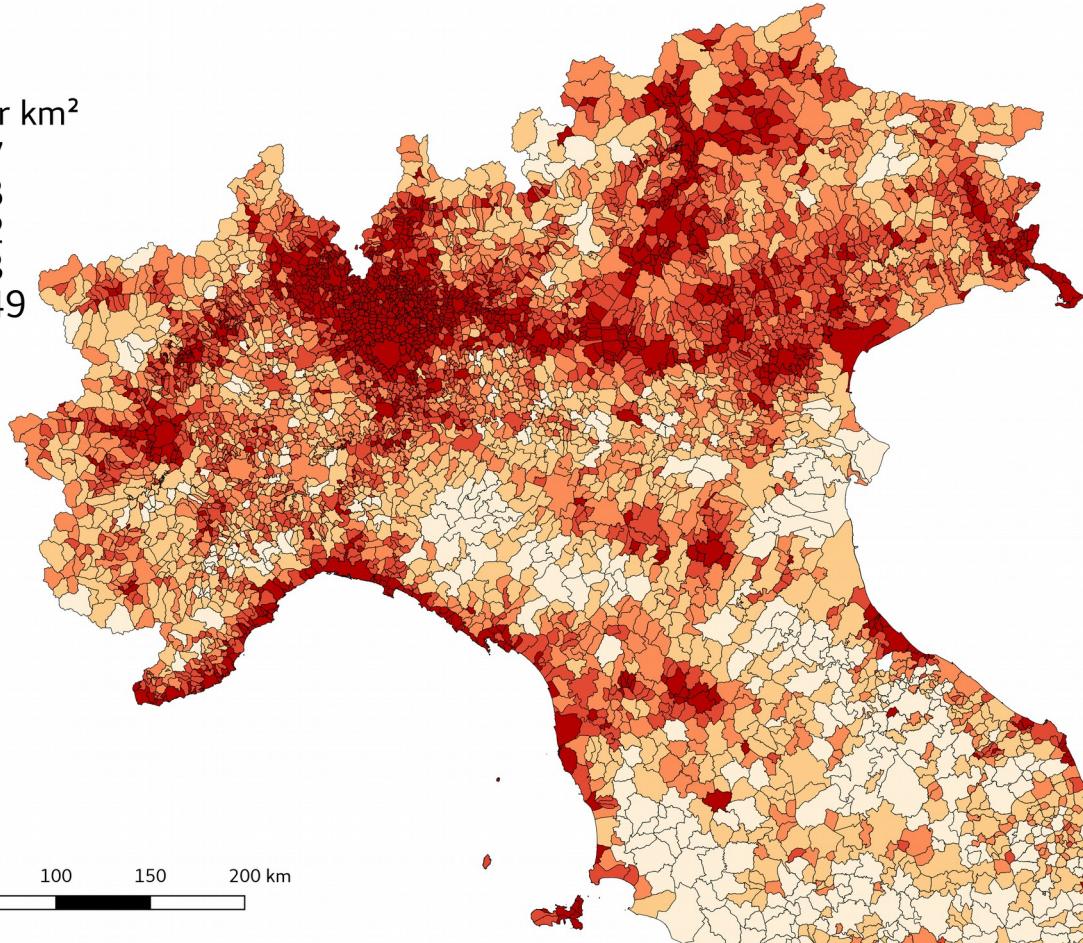


Contributori per km² – Nord Italia

Legenda

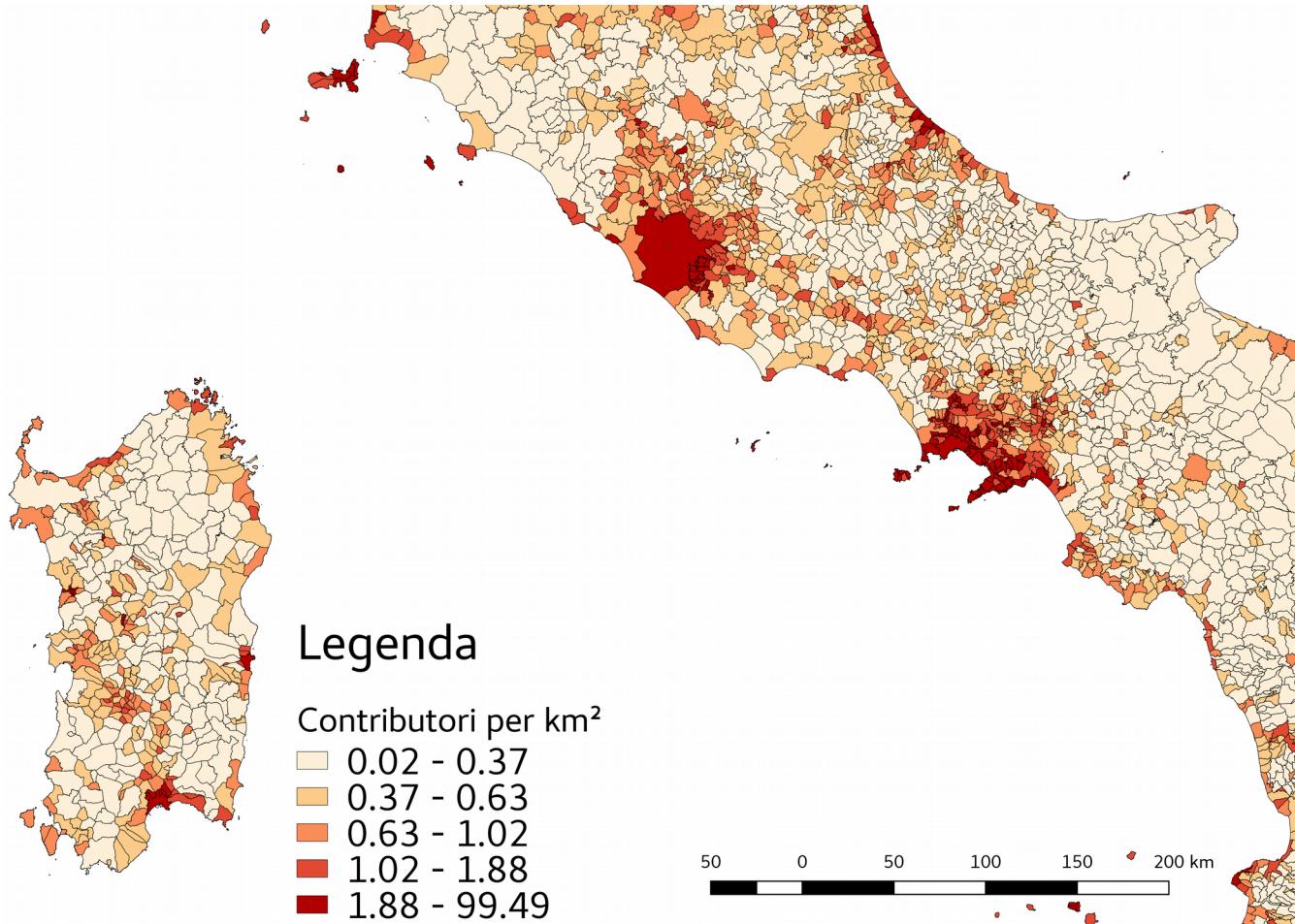
Contributori per km²

- 0.02 - 0.37
- 0.37 - 0.63
- 0.63 - 1.02
- 1.02 - 1.88
- 1.88 - 99.49



50 0 50 100 150 200 km

Contributori per km² – Centro/Sud Italia

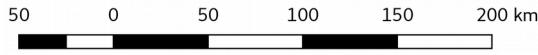


Contributori per km² – Centro Italia

Legenda

Contributori per km²

- 0.02 - 0.37
- 0.37 - 0.63
- 0.63 - 1.02
- 1.02 - 1.88
- 1.88 - 99.49



Aspetti tecnici relativi all'analisi aggregata dei dati

Tecnologie utilizzate

- Python 3
- Osmium
- SQLite + Spatialite
- QGIS

Tempi di processamento

- Conversione italy.pbf: 45'
- Associare ad ogni nodo un comune: ~2 ore
- Tabella statistiche: 2'

Possibili migliorie

- ORC anziché PBF
 - Spark+Hive
 - PostgreSQL+Citus

Accuratezza semantica

È possibile stabilire se un nodo è stato classificato in maniera incompleta o in maniera errata indipendentemente dalle regole prescritte nella wiki, studiandone l'evoluzione nel tempo?

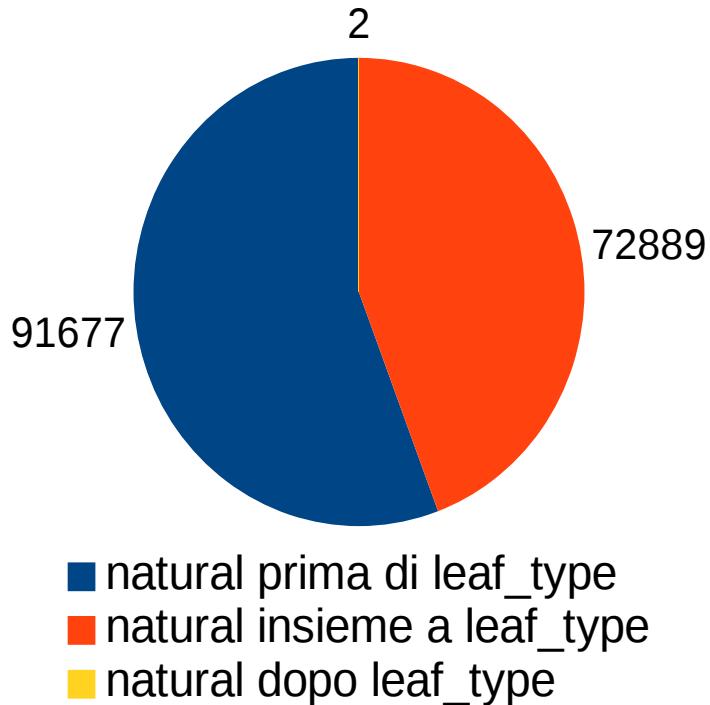
Dati presi in considerazione:

- Nodi presenti in `italy.pbf`
- 100 tag più popolari

Trend nei tag – Tag A aggiunto prima del tag B

Rapporto natural/leaf_type

~165.000 nodi aventi sia natural sia leaf_type



Potenziali errori

- 5 volte leaf_type appare senza natural

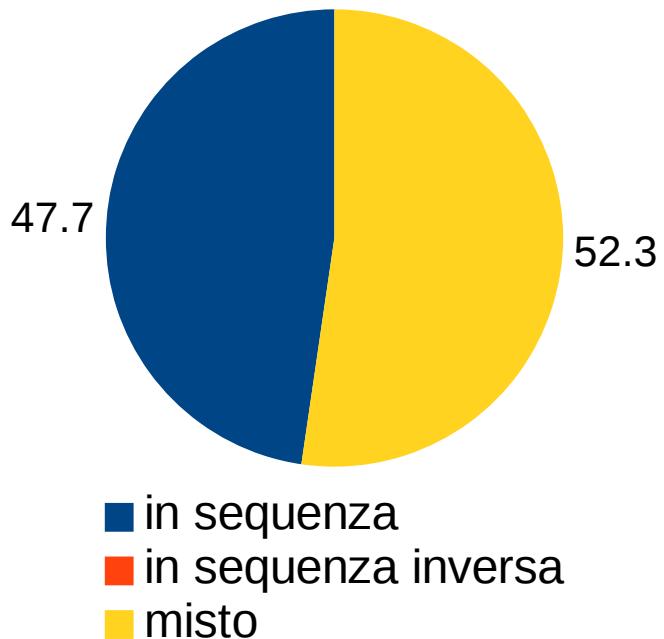
Statistiche generali

- 134 coppie in cui A viene prima di B almeno il 50 % delle volte
- 61 coppie appaiono più di 1000 volte

Trend nei tag – Tag A aggiunto prima del tag B e B prima di C

Rapporto tourism, parking, opening_hours

~600 nodi aventi i tre tag



Statistiche generali

- 69 terne in cui A viene prima di B e B viene prima di C almeno una volta su 3
- 19 terne appaiono più di 500 volte
- 43 % delle terne termina con “wikidata”

Conclusioni

Stato attuale

- Webapp pronta all'uso, utilizzabile anche da linea di comando
- Sono stati esplorati differenti approcci per convertire, filtrare e gestire i dati temporali di OpenStreetMap
 - Pochi strumenti supportano dati con stesso ID e revisioni differenti
 - Problemi di performance
- Lavoro preliminare sulla combinazione tra analisi semantica/ aggregata e temporale

Conclusioni e prospettive future

Is OSM up-to-date?

- Dato aggregato/tile
- Esportazione geoJSON
- Utilizzo del planet

Analisi aggregata

- Metriche pesate sul tipo di oggetti (esempio: i cestini di solito non vengono modificati)
- Comprensione dei trend locali (esempio: Sardegna; coste...)

Analisi semantica

- Correlazione con uso del suolo e densità di popolazione
- Attributi way
- ...

Le analisi aggregate e semantiche sono solo all'inizio: ogni suggerimento e critica è ben accetta!

Indice di qualità unico?



POLITECNICO
MILANO 1863

Domande?

fraph24@gmail.com - <https://frafra.eu>

Francesco Frassinelli, Marco Minghini, Maria Antonia Brovelli

Dipartimento di Ingegneria Civile e Ambientale

Merge-IT 2018 – 24 marzo 2018, Torino