Topology-aware indexing system for Urban Knowledge

(Article begins on next page)

Topology-aware indexing system for Urban Knowledge

Alessio Antonini
Department of Computer Science
University of Turin
Turin, Italy
antonini@di.unito.it

Lucia Lupi
Department of Computer Science
University of Turin
Turin, Italy
lupi@di.unito.it

Guido Boella
Department of Computer Science
University of Turin
Turin, Italy
boella@di.unito.it

Claudio Schifanella
Department of Computer Science
University of Turin
Turin, Italy
schi@di.unito.it

Stefania Buccoliero
Department of Computer Science
University of Turin
Turin, Italy
stefania.buccoliero@gmail.com

***Abstract*— Maps are being widely used as tools for presenting or retrieving information with spatial attributes. Existing map-based applications do not use the full potential of digital maps and geographical data: social media are disconnected from the underlying geographical entities; maps as visualization tools do not use the urban topology to cluster point of interest; maps as input systems are intrinsically ambiguous. This work presents a topology-aware indexing system supporting a new metaphor for a real integration between social media and digital maps. The methodology and technical solutions required to build and populate the indexing table starting from OpenStreetMap spatial primitives are introduced.**

Keywords*— *urban knowledge; geographical index; urban computing; web digital maps

I. Introduction

Web digital maps can be used to provide information context, as filtering system and as input system for spatial media, but there is still no connection between digital maps and contents. In particular, map scaling system can be extended to content, making maps work as recommender system, nevertheless current approaches do not exploit the geographical information behind maps to select information according to the theory of relevance associated to map scales. From another perspective, the use of digital maps as visualisation tools is limited by the issue of representing a large number of point of interest (POIs). Considering maps as input system crashes against the ambiguity of pinpointing mechanisms, because there is no evidence to decide what a user is indicating among many urban entities underlying a click (building, city block, neighbourhood, city, etc.).

The recent growth of map-based tools is indicating an increasing importance of spatial aspects of media to deal with the rising amount of information available on the web. Thus, this contribution addresses the general problem raising from the current lack of insight about spatial features of social media. Starting from the classical theory of maps, we are going to tackle these following questions:

1) How can we enrich urban entities with the knowledge of urban topology?
2) Can we use map features to build an index system based on urban topology for urban entities?
3) How a spatial indexing system can support intelligent visualization of urban entities?

We addressed the problem related to the misuse of digital maps, starting over from the visualisation theory of relevance characterising traditional maps, where the visualization of information is specific for each scale level, as well as the representation of geographical components.

We used these principles to structure a topology-aware indexing system from crowdsourced data, by introducing new geographical units in between the city level and the buildings level and associating them to a zoom interval.

In order to have spatial continuity of geographical units and make possible a continuous scaling from the indoor to the city level in our indexing table, we extracted the missing entities from crowd geographical data available in OSM about the city of Turin, Italy. We started from the minimum geographical units such us city blocks and open spaces.

In urban theory, city blocks are the key element in the perception of urban spaces because, at that level, physical and social elements are integrated as complex primary components defining the character of the city [1,2,3]. An important part of the urban knowledge about public and private functions, individual and collective organization, spatial hierarchy, meanings attribution, ownership and places identity is related to the characteristics of city blocks and open spaces. For this reason, we focused our attention on this scale level in order to identify automatically this type of geographical unit.

After, we reviewed the data input metaphor based on pinpointing mechanism, shifting to the connection between contents and geographical units based on dropping content into an area. The proposed indexing system supports an information visualisation system based on areas and city components, overcoming the current clustering approaches based on spatial grids.

We tested the index in a real-world application: a map-based component for the "WeGovNow!" project aimed to collect social information at urban level.

This contribution is aimed to present our approach to the problem and the experimental result. The rest of this paper is organized as follows. Section two presents the state of the art, highlighting the visualisation theory of relevance characterising traditional maps and how we used its principles to structure our index, the existing standards in spatial indexing, and the relation between location and context in current map-based web applications. In section three, we present the methodology used to obtain continuous geographical units at different level starting from OSM spatial primitives. We discuss also the current metaphor of pinpointing on the map used in data visualisation proposing a new perspective enriching spatial data with context information represented, retrieved, and visualised considering the corresponding relevance layer for each urban entity. Section four presents the technical solutions required to build and populate the indexing system and we discuss the computation time of an experiment on OpenStreetMap data.

II. State of the Art

The knowledge about the city can be seen as a coherent system of social and geo-graphical information about urban entities. Firstly, users' contributions should be explicitly consistent with urban entities (objects, activities, scale of interest) and the geographical context should enrich social information. As second step, knowledge about urban entities should be built merging spatial attributions and users' contributions, by using a correlation mechanism consistent with the city structure. In other words, contents must be aggregated if they are related to the same entity, not just to the same topic, distance, users or other metrics. Last, in order to support an urban knowledge system, we need to provide a coherent context through a topology-wise representation of urban entities.

GIS (Geographical Information System) are the main systems designed for spatial knowledge, with many important sectorial applications such as logistics, emergency, spatial planning managed by expert users. Recently, most of the available non-technical spatial information comes from crowd-based platforms such as OpenStreetMap. These web platforms (web GIS) are simplified version of standard GIS, easy enough for untrained users, but still aimed to collect geographical entities and not social information about city spaces and activities.

Nowadays, new challenges come from social media and map-based applications where the focus is not about geographical entities but about social information with spatial attributes. The use of location and location-aware features is becoming unavoidable in order to improve recommender, filtering systems and data visualisation of almost every day digital tools. Finding new intuitive solutions for data visualisation reflecting common perception and use of space will lead general users to access and share meaningful information not only on physical components, but also on intangible and un representable elements that define the corpus of the urban knowledge.

A. Applying classical maps principles in map-based web applications

What does it mean to connect map and contents? For instance, a political world atlas shows nations despite their size, population or political influence, nations are unnaturally coloured in order to highlight their boundaries, and there are no orographic references or any other element describing their morphology. Connecting map with contents means discarding useless information in favour of highlighting what is most important for the map goal. In other words, a map communicates applying a relevance theory to a set of geographical information.

Now, let us consider road guides: books for traveling by car mostly used before the spreading of digital navigation systems. Road guides are organized in tables at local and national scale. National tables represent only highways, state streets and main railways, while secondary and local streets are shown in local tables. Local and national tables share the same goal and look very similar but their content are different because the scale is a way to select information. Scales correspond to theories of relevance about map contents.

Traditional maps are one of the oldest ways to express relevant information on a locality base [4]. Thus, they can be saw as recommender systems ante litteram, because they always present a synthetic representation of reality at different scales with different purposes [5]. Maps can be used to represent many aspects of reality [6], combining and defining different semantics through visual signs [5]. "Maps mediate the inner mental world and outer physical world. They help us make sense of the universe at different scales, from galaxies to DNA, and connect the abstract with the concrete by overlaying meanings onto that world ... they help us remember what is important and explore possible configuration of the unknown" [7].

We are going to consider a specific aspect of maps, which remains implicit in digital maps for the web: the scale. Scale is a broad concept investing technical schemas, road charts, urban plans, and many other fields. Which aspects we should take into account considering digital maps for the web? Which scale are we considering?

It is possible to consider three typologies of scales: cartographic scale, geographic scale, and operational scale [8]. Cartographic scale considers the relationship between distance on a map and distance on the ground [8]. Geographic scale considers the geographical structure of social interactions [9]. Operational scale corresponds to the level at which relevant processes operates [8]. We are going to consider operational and geographical scales because the geographical data we can access: OpenStreetMap is mostly focused on space functions and uses [10]. Size, level, and relation [11] define the geographical scale. In our scenario, the size is controlled by users through zooming; the level - local, national, regional – is bounded to relevance layers, considering administrative levels, neighbourhoods, and city blocks at different zooming level; the relation is a mix including space, place, and environment, which is represented by OSM entities plus POIs.

Technically speaking, digital maps are web viewers of pyramidal images, composed by raster or vector tiles organized in spatial grids and zoom levels. We are considering the common use of digital maps in social media applications: as visualization system for POIs as markers. Digital maps can usually be used with different representations such as satellite images and road maps (tile sources). Each tile source may implement a different relevance theory (one for each scale) trough styling and rendering. For instance, starting from OSM data it is possible to provide completely different maps.

Digital maps are filtering systems where the map window is called bounding box and could be used to filter contents outside the map boundaries. Along with the bounding box, each zoom level defines a map status, with an integer between 1 (planet) and 18 (street) or 22 (indoor). Zooming a map can result in changing its scale, but there is not a fixed relation between scales and zoom level [12]. Usually, when a digital map change scale its contents are not affected, or in other words, the visualization of map contents does not change moving from the street level to national level. This make it difficult to read because of the overlap of elements and the disconnection among signs, meanings and relevance relation.

We extend the use of digital maps from using their bounding box to filter information to a coherent information management system extending scale effects to map contents.

B. Geographical Indexing Systems

An indexing system is a data structure supporting information representation and retrieval. Spatial indexing organizes contents according with their spatial feature: their position in a 2d or 3d space. Which features indexing systems should address? Are spatial features optimal for representing urban information?

Spatial indexing is nowadays a standard feature or a common extension to standard databases, for instance such as r-trees. There are several solutions specifically tailored for performance, space constrains, scalability, and domain application constrains. Recent improvements and extensions of spatial indexing are oriented to time and space performance in order to support real time complex queries for the new generation of location-based applications, avoiding introducing any insight about urban topology [13].

New hybrid approaches have been proposed in social media applications where the system architectures need to process several dimensions at the same time and be very fast in order to support real time applications. Hybrid because those indexing systems combine physical indexes with automatic extraction of features, for instance combining space, time, and text analysis.

A common scenario in urban context requires to cross users geographical profile with social media in order to identify anomalies from usual patterns (new visited locations) in real time, suggesting media tagging and the short route to home, at the same time. In this scenario, users will not put much more effort than a click on a map to input a location. This assumption can be extended to most of social media where the spatial information is provided as GPS coordinates, as an address or as a marker input.

Summarizing, the problem we are addressing falls in knowledge management of POIs (indexing, retrieval, recommendation, and visualization). A point of interest (POI) is an information unit enriched with punctual spatial information but it can have several other dimensions such as categories time. POIs index structures need to support very hard applications like location-aware user-based context-dependent recommenders [14, 15].

We are considering information about urban entities where the spatial feature is strictly punctual, since the application domain is not a web GIS platform, but a location-based social network oriented to the public sphere, or in other words a civic social network. This kind of application is not focused on mapping places, but on supporting local stakeholders and their activities in order to represent the social knowledge associated shared urban entities, mostly places.

We are proposing a domain-wise solution considering specifically urban information and knowledge management systems about urban entities. In detail, we consider urban knowledge as a subset of social knowledge about physical urban entities such as streets, squares, buildings, parks but also as conceptual constructs as neighbourhoods and city blocks. These entities are social, but they cannot be found looking into users' behaviour or text analysis because they belong to a shared environment built by the interaction of users with physical components.

Moreover, considering a continuous space leads to a severe misconception of POIs similarity. In urban scenario, the distance among POIs is not just a metric correlation, because what matters is if POIs are related to similar geographical entities. We adopt this as leading principle of the technical solutions we are presenting in this contribution.

C. Spatial Context in Location-based Applications

Everyday examples of web applications using digital maps are Google Maps, Booking, Airbnb, TripAdvisor. These systems provide an internal search engine, a time category based filtering system, and recommendation system based on search queries, filters, popularity, voting, etc. The spatial context of information is usually provided through a digital map, which is used to display results, manipulate the context, and zooming or panning the map. These applications are commonly described as location-based social network [15]. In almost every location-based social networks, maps are not main access point to the information and remain "hidden" as extra features, despite being very intuitive and powerful as filtering and recommendation systems.

There are other specific applications based on spatialized contents but requiring anonymous recommender systems - not based on user profile nor on popularity – such as platforms addressing public information for public purposes such ad We-government processes, promotion and support to public initiatives, and collaborative urban planning. For instance, civic media and civic social networks such as FirstLife [16,17] require avoiding standard user-based strategies for knowledge management. In this perspective, a map perfectly fits as solution to the representation of knowledge context, because maps are in general beyond the private sphere [18]: maps can recommend what is relevant from a general perspective letting users in control expressing an objective information context (the same for each user).

Considering mass crowd map-based applications, the main issue is a huge amount of dynamic unstructured softly linked data, generated by and meant for common untrained users. There are several techniques, algorithms and tools for map visualization in GIS, but most of them cannot support real time applications in web clients because of performance and space issues. Nevertheless, there are many important projects about big data, data stream and digital maps: vector tiles and rendering technologies server (WebGL). Vector tiles contains information for rendering tiles at client side - raw data (GeoJSON or TopoJSON) or SVG - instead of a raster image. Vector tiles enables a deeper experience with digital maps, where users can access to map contents directly as open structure. WebGL libraries for maps can render real time stream of information making digital maps powerful monitoring tools.

The challenge is to combine the current technologies used in all this kind of applications to build map-based social networks where crowd sourced information are input directly inside geographical entities and visualized real-time by all connected clients. A theory combining dynamically the input data with the up-to-date map status is missing, but is required to make this scenario real. In addition, the information context provided by existing web mapping tools is not location-aware and this create an incoherence between interpretation of data spatial attributes and the exchanged information. Moreover, the information context is not explicit, users cannot control it, and applications do not support multiple representations or support interpretation charts of POIs [19].

III. METHODOLOGY

A. From Spatial Primitives to Geographical Units

We considered OpenStreetMap (OSM) as a cartographic source [20]. OSM supports three types of spatial primitives: nodes, punctual information defined by latitude and longitude; ways, a collection of nodes shaping a geometry (polygon, line); relations, an array of ways of which the first is the outer geometry (external area) and the others are holes if there are. For instance, a tree is represented as a node; streets, buildings, parks are ways; countries, complex structures, cities and other administrative areas are relations.

OSM entities level pass from physical objects to administrative areas, without having an explicit definition of empty spaces and intermediate units, as for instance city sectors, sub-districts, neighbourhoods and city blocks. Nevertheless, urban knowledge is organised using these type of untracked entities, difficult to map because they are neither institutional nor only physical. Indeed, according to common experience, essential public services are often available at a sub-district level, most of everyday activities happen at a neighbourhood level, pedestrians use city blocks and spatial discontinuities for orientation because of their recognisability, and so on. Relevant information at the city scale are different from data regarding buildings or green areas, or from local news involving resident of a specific city sector, etc.

City sectors, neighbourhoods and city blocks are the geographical units needed to organize spatial information taking into account both the way to use the space by users and the possibility of creating clusters of relevant information for each level. A topology-aware indexing system considering these levels must be structured in progressive continuous scales ensuring the spatial continuity of the map for data representation.

The progressive scaling has been designed by establishing additional scale levels between the city level and the street level, respectively corresponding to the level 12-13 and to the level 18-19 of the zoom interval of digital maps, ranging from 1 (world) to 22 (indoors). According to the theory of relevance of maps used for this kind of indexing system (with the additional constraint of not having overlapping elements), we have different elements at each scale selected from the spatial information dataset. For this reason, we identified geometries corresponding to the geographical units listed before: city blocks and open spaces, neighbourhoods, city sectors. The extraction of geographical units from the cartography has been done using the OSM spatial primitives to calculate them automatically starting from their geometrical properties. We describe the technical aspects of this procedure in section 4.

In order to ensure the spatial continuity, we calculated the missing geometries associated to the new scale level we introduced by intersecting the geometry of the city level (administrative boundary) with the geometry of components belonging to the street level (buildings, parks, etc.), or in general, by intersecting upper and lower level geometries. The vertical connection is resulting because each geographical unit is contained in the unit at the upper scale.

B. Pinpointing VS dropping

Considering the relationship between points of interest and their placement on the map, we can observe that POIs are usually not connected to the relevant corresponding geographical unit in map-based applications because of the common method to ass spatial information to contents on digital maps: the pinpointing of markers on flat raster images. Pinpointing a marker is a very intuitive and easy way to interact with maps and only a few platforms support other geometric features (polygons, polylines).

Nowadays, vector maps are replacing their raster version giving the possibility to look which spatial entities compose maps, but still the connection between markers and map is the coordinate system and not the relevance of information. Indeed, analysing the results of some community mapping workshops we carried out in Turin, the experimental evidence is that only a minimal part of POIs is actually connected to physical objects at the street level as buildings. This is because the common knowledge of urban spaces is often referred to upper level entities, such as the built environment of a neighbourhood, the maintenance of local streets, the availability of services in the district, etc. This kind of information cannot be properly represented with the pinpointing mechanism, retrieved and synthetically visualised for users.

In order to disambiguate users pointing, we propose a switch of metaphor from “pinpointing” to “dropping” a marker on map geographical units. Even if a raster map-based application cannot identify single spatial entities, users can “see” map data and place POIs considering the vertical ordering of elements: a river is “below” a bridge, a building is “above” a parks, etc. Geographical units work as “buckets” one in other, according to their dimension and vertical order, where users drop their contents.

Given this assumption and the relevance theory about map scales, we can build an indexing system based on topology and support the data collection assisting users in connecting POIs by pre-sets oriented to link contents and urban components according to the current zoom level. We can discriminate between relevant information at the scale of city blocks and those relevant at lower or upper scales, considering which contents are dropped inside or outside the area corresponding to their geographical unit and selecting for visualization only POIs connected to the current scale of interest of users. [fig1].

In this way, we can enrich information related to urban entities contextualizing and connecting them with a specific area corresponding to the observation level, distinguishing relevant contents to visualize and offering the correct information context to users.

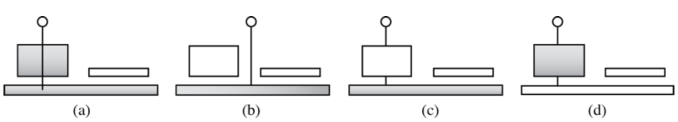

Fig. 1. (a) ambiguous pointing on overlapping elements; (b) pinning the lower entity explicitly requires to point the interspace among higher entities; (c) and (d) disambiguation of POIs position by using different geographical units at different scales.

C. Information representation, retrieval, and visualisation

In order to implement an indexing system based on relevance layers, we connected the zooming level of dynamic maps to a scaling system. Later we created the scale layers extracting the right geographical units for the zoom levels using the rules we defined from the map styles. After, we populate an index table instantiating the relevance layers (scales). In the end, each POI is connected to each unit in the index table when they intersect.

Fig. 2. On the left, a classical marker clustering visualisation; on the right the result of the index-ing system, connecting POIs and spatial entities; in the middle an intermediate step.

We have developed a viewer calling the indexing system at each change of the map, showing the GeoJSON with the object in the relevance layer containing POIs. The geographical units are rendered accordingly to the POIs number: from blue to green to yellow and red. The viewer queries our indexing system tracking the relations among areas and POIs. In particular, we consider a “geographic” distance: POIs are closer if they point the same geographic unit and they are apart if not.

At any level, we have the POIs geographic-based aggregation making possible to show aggregated data instead of multitudes of single POIs. This mechanism result in a content aggregation overcoming the current clustering methodologies for POIs, based on spatial grids, geometric and non-geographic, independent from topology and context characteristics related to the perception and use of space by users.

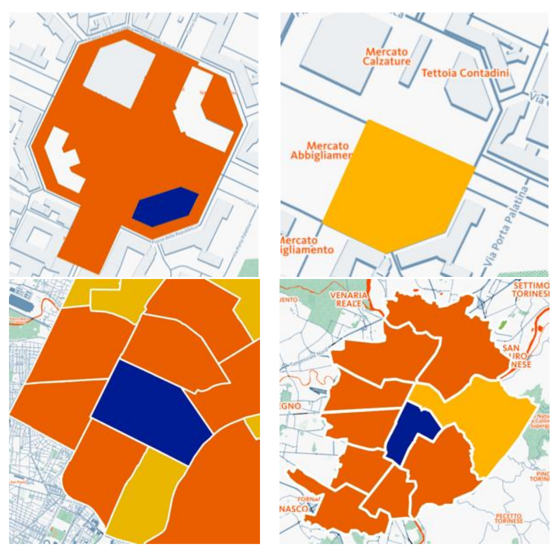

Fig. 3. The top-left screenshot shows a market square where there are POIs differently distributed among market buildings; following a square section, sub-districts and districts.

Our indexing system automatically filters the information on a geographical basis providing a clear representation of content without overlapping markers and presenting a manageable amount of information to users. In addition, a topology-aware indexing system can be used to structure recommender system based on real proximity, connecting POIs and lived spaces.

IV. Technical Solutions

A. *Structure of the Topology Aware Index System*

We structured our index system considering a selection of OSM layers from the nation to the municipality level, adding new geographical unit as districts, sub-districts, and neighbourhoods, important to form and organize the urban knowledge and for the representation of relevant information to the user. We introduced also the level of the minimum geographical unit corresponding to city blocks and open spaces. Then, we considered on the same level elements belonging to different OSM layers defined as buildings, waterway, leisure, land-use, highway, etc. having in common the fact of being specific spatial objects identified as unitary and distinct from others and labelled with a category name. Lastly, we added the indoor level. We associated a zoom interval to each level and a rendering priority order.

In the indexing table, each geometry is associated to an ID following the index structure described before. We extract from OSM the entities with their corresponding levels to fill an indexing table.

POIs are indexed using the intersect operation between POIs and geographical units listed in the index table, and populate a relation table used for answering to queries. We defined an API selecting all the records from the index table residing in the bounding box and having a zoom interval including the zoom level selected by the user. Given the list of records, the API combines spatial entities with POIs in the relation table creating a GeoJSON feature collection of spatial entities, which have as property the list of the connected POIs. The aggregation of POIs is done runtime using PostGIS spatial index on the field geometry.

An OSM dump, the index table and the relation table running on Postgres/PostGIS database compose our system. The API is implemented using Express, a REST framework for nodejs.

B. *Authors and Affiliations*

According to the basics of the spatial theory information, we consider the geometry of city blocks as polygons delineated by street segments. [25] We found an algorithm to extract city blocks and open spaces from the street network finding all the smallest cycles in the street network. Given a street graph where street segments are edges and crosses are nodes, for each couple of connected nodes there is a city block if exists an alternate shortest path and the city block is the union of the edge and the alternate shortest path.

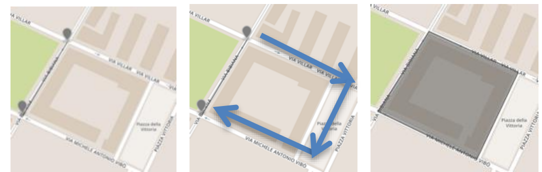

Fig. 4. Given a couple of adjacent nodes, we calculate the shortest alternative path. The block is the union between the edge and the shortest path.

We created the street graph by using OSM2PGRouting and PGRouting, a routing library for Postgres. During the data extraction, we found many issues caused about data itself. We ended up to pre-processing data for several reasons such as: to extract the labels from OSM, which are expressed with a unstable combination of 64 columns and an of an undetermined list of tags; to fix streets that are incomplete and wrong; to avoid bugs in the DBMS methods for geometries. In particular, we re-moved fake crosses and loopholes caused by dead ends and broken streets.

The first version of the algorithm has been used PGRouting k-shortest to find the two shortest routes (the edge plus the first alternative). The algorithm took almost a full week to calculate the city blocks of Turin (almost all). We discovered that this first version was not complete because the way we were handling the list of nodes needs to be checked. The algorithm we present is a revised complete version that we are still tuning using the graph database neo4j. The new version does not consider couple of connected nodes but triples *"L"*:

1. Streets are sliced in segments among crosses
2. Streets are scanned for all "L", two segments connected by a cross
3. For the two extremes of each "L" the second shortest path is being searched
4. If we found a shortest path
    a. The union of the "L" and the shortest path is added to the list of blocks
    b. Each other "L" in the new block is being removed from the search space

Generating the triples to check from the graph is quite fast and, since Neo4j is an index free native graph database, the shortest path computation is usually below the second. The new version is complete, but it may generate alternative blocks. In order to refine the results, we use two heuristic rules to discard unwanted blocks:

1. if a block X contains one or more existing blocks, discard X
2. if a block X is contained in an existing block Y, swap X and Y

The conversion of streets into a graph is made using PGRouting, a Postgres library for routing, and OSM2PGRouting, a tool converting OSM streets in nodes (crosses) an edges (street sections). OSM2PGRouting setup can consider street direction and labels. In order to build the dataset for the experimentation, we have ignored the street hierarchical labels (highways or local does not matter) and the street direction. Considering the street hierarchy, the presented algorithm can be used to calculate higher-level areas as urban sectors.

The automatic extraction of city blocks and open spaces from maps allows us to cluster spatial information at the level of the minimum geographical unit, offering a visualization of data that is topological aware and suited to the common perception of space.

C. Spatial continuity

As we stated in section 3, geometrical entities at each scale should define a continuous space. Considering one more time the street level, we can calculate the missing geometrical entities by difference with geometrical entities at higher level, in this case with the minimum geographical units (city blocks and open spaces). Technically speaking, this very simple operation can be generalised and applied at each index level as follows:

for each level L_i with geometry g_i

1. retrieve geographical entities E intersecting g within L_{i+1}
2. calculate the geometry $q = g_i – E$
   add q to L_{i+1}

This mechanism can be used to fill the missing geographical units. There are some evident constraints: 1) the higher level needs to be continuous, 2) it has to be applied top down and 3) if a lower level geometry falls within two higher level geometries its shape has to be considered two times. In figure 5 an example.

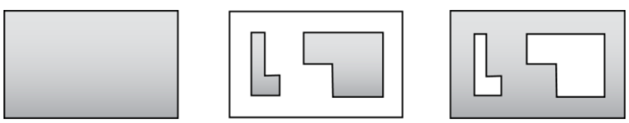

Fig. 5. The missing spatial unit (right) are the result of the intersection of an higher level geometry (left) with the lower level geometries (center).

D. Figures and Tables

The testing was done on the street network of the city of Turin, Italy, by using the OSM cartography. We considered all geometries related to the street categories (primary, secondary, local and pedestrian) in order to calculate the minimum geographical units defined as polygons delineated by street segments. The algorithm output is of 13.134 geometries and 39 holes (mostly parks and riverbanks), manually corrected in QGIS. Surprisingly, most of entries were street sections rather than city blocks or open spaces. Indeed, the algorithm found 6.314 city blocks/open spaces and 6.820 street sectors [fig.6].

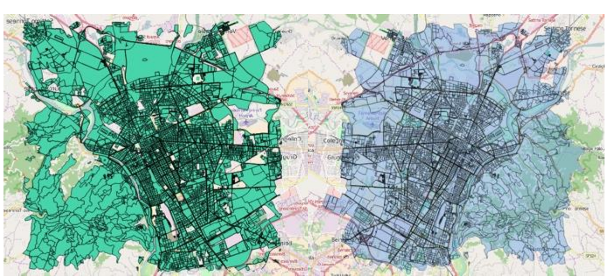

Fig. 6. Algorithm output without correction on the left, manually corrected in QGIS on the right.

The official cartography of the municipality of Turin reports 4.321 minimum geographical units and we used it to check the algorithm output and to isolate street sections. Crossing the municipality data with our results, we isolated 6.314 units, 1.993 blocks more than the official cartography, generated due to pedestrian paths and secondary streets splitting green areas and open spaces. [fig.7]

Fig. 7. On the left, geometries from the official map. In the centre, a detail of the map. On the right, the result of the intersect operation with our data.

Summarizing, our results are more detailed in comparison with the official map in in case of green areas and empty spaces, but we still have 6820 street sections that we do not want to consider as relevant. This unwanted entities included in the list of minimum geographical units are not due to the algorithm but to the representation of the street network itself: the topology considers not only streets but also routes, directions, splits, pedestrian crosses, and the street viability in general because the data source from OMS is meant for navigation purposes. Changing the data source with a simplified version of the street graph should solve the problem.

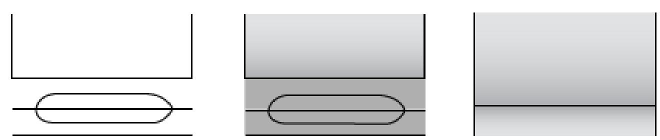

Fig. 8. The street graph (left) leads to unwanted geometries (center). A simplified graph can solve the problem without changing the current algorithm (right).

The whole process of importing and converting data took almost one month of computation: 6 days for graph creation, converting OSM data to PGRouting format and from PGRouting format to Neo4j; 6 days for the computation of the actual units; 10 days to check duplicates and exporting geometries from Neo4j graph; 1 hour for empty spaces computation.

V. CONCLUSIONS AND FUTURE WORKS

In this contribution, we addressed the issue of how to enrich social media about urban entities. We proposed to connect them with geographical entities presenting a topology-aware indexing system for POIs. For each index level, POIs are selected using a scale theory of relevance. Besides, in order to achieve spatial and scale continuity, we developed an algorithm to extract the minimum geographical units from a street network, and a mechanism combining higher and lower level entities to extract geometries of “empty” spaces.

The next step to achieve the full-scale continuity is to calculate neighbourhoods, difficult to extract because of the lack of administrative borders. In addition, these intermediate geographical units between city block and city sectors are mostly dependent from social factor and environmental characteristics. We will exploit the presented mechanism to identify city sectors, after the homogeneous areas in terms of relationships between built and empty spaces, and lastly neighbourhoods intersecting the previous two. Furthermore, we need to define scale rules behind a smooth scale transaction [21] to use the topology-aware indexing system and the related visualisation system in real world applications.

Currently, we are using the indexing system in the project “WeGovNow!”, a European platform for We-government we are building with several international partners for a H2020 project (http://wegovnow.eu). In this case, the urban data platform we are developing is meant to bridge local institutions, privates and other local players through sharing and co-producing public information about cities involved in the project. In this scenario, we are building a new visualisation system for huge amounts of POIs without relying on user profiling and ranking. The goal is to make users work at the right scale: low level for neighbourhoods and higher level to explore the city.

REFERENCES

[1] Batty, Michael, and Paul A. Longley. *Fractal cities: a geometry of form and function*. Academic Press, 1994.

[2] Krier, L. "The reconstruction of the city." La reconstruction de la ville européenne: architecture rationnelle= The reconstruction of the European city: rational architecture, Archives d’Architecture Moderne, Brussels (1978).

[3] Trancik, R. Finding lost space: theories of urban design. John Wiley & Sons, 1986.

[4] Shahaf, Dafna, et al. "Information cartography: creating zoomable, large-scale maps of information." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013.

[5] Keates, John Stanley. Understanding maps. Routledge, 2014.

[6] MacEachren, Alan M. How maps work: representation, visualization, and design. Guilford Press, 1995.

[7] Okada, Alexandra, Simon Buckingham Shum, and Tony Sherborne, eds. Knowledge Cartography: software tools and mapping techniques. Springer, 2014.

[8] Lam, Nina Siu-Ngan, and Dale A. Quattrochi. "On the issues of scale, resolution, and fractal analysis in the mapping sciences*." The Professional Geographer 44.1 (1992): 88-98.

[9] Smith, Neil. "Contours of a spatialized politics: homeless vehicles and the production of geographical scale." Social text 33 (1992): 55-81.

[10] Marston, Sallie A. "The social construction of scale." Progress in human geography 24.2 (2000): 219-242.

[11] Howitt, Richard. "Scale as relation: musical metaphors of geographical scale." Area 30.1 (1998): 49-58.

[12] Carral, David, et al. "An ontology design pattern for cartographic map scaling." The Semantic Web: Semantics and Big Data. Springer Berlin Heidelberg, 2013. 76-93.

[13] Balasubramanian, Lakshmi, and M. Sugumaran. "A state-of-art in R-tree variants for spatial indexing." International Journal of Computer Applications 42.20 (2012): 35-41.

[14] Griesner, Jean-Benoît, Talel Abdessalem, and Hubert Naacke. "POI Recommendation: Towards Fused Matrix Factorization with Geographical and Temporal Influences." Proceedings of the 9th ACM Conference on Recommender Systems. ACM, 2015.

[15] Liu, Bin, and Hui Xiong. "Point-of-Interest Recommendation in Location Based Social Networks with Topic and Location Awareness." SDM. Vol. 13. 2013.

[16] Lupi, L, et al. "Back to Public: rethinking the public dimension of institutional and private initiatives on an urban data platform." IEEE International Smart Cities Conference ISC2 2016, accepted paper.

[17] Antonini, A. et al. “Civic social network: a challenge for co-production of contents about common urban entities”. CollabTech 2016, accepted paper.

[18] Bao, Jie, Yu Zheng, and Mohamed F. Mokbel. "Location-based and preference-aware recommendation using sparse geo-social networking data." Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, 2012.

[19] Sannella, M. J. 1994. *Constraint Satisfaction and Debugging for Interactive User Interfaces*. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.

[20] Mordechai (Muki) Haklay and Patrick Weber. 2008. “OpenStreetMap: User-Generated Street Maps”. IEEE Pervasive Computing 7, 4 (October 2008), 12-18. DOI=http://dx.doi.org/10.1109/MPRV.2008.80

[21] Girres, J. F., and G. Touya. "Cartographic generalization Aware of Multiple Representations." Proceedings of the 8th International Conference on Geographic Information Science, Vienna, Austria. 2014.