On improving data quality and topology in vector spatial data

Nina Solomakhina*,†, Thomas Hubauer* and Silvio Becher*
* Siemens AG, Munich, Bavaria, 81739, Germany
Email: fname.lname@siemens.com
† École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
Email: fname.lname@epfl.ch

***Abstract*—Data quality is an important issue for a spatial data, especially for topological relations between geographical features. Errors and inconsistencies found in Geographical Information System (GIS) data often misrepresent topological structure of the dataset and, therefore, geoprocessing and spatial analysis (e.g., network analysis) do not yield reliable results. The focus of this paper is to identify and correct topological errors in vector spatial data, in network data in particular. We present the method for identifying and correcting dangling line in datasets aiming to reconstruct incorrect topological relations between lines and other features. We tested the proposed approach on the real-world energy network data.**

Keywords–spatial data; networks; topology; topology errors

I. INTRODUCTION

Spatial data is very diverse. It comes in different formats and types: satellite or aerial photographs, hand drawn or printed maps, raster or vector graphics files and other numerous formats. Before performing further data mining and exploring procedures, it is necessary to assess whether data is suitable for their application, e.g., whether satellite and photo images require noise removal procedures or tables and Geographical Information System (GIS) files require duplicate removal procedures.

Digitized GIS data has two major formats - raster and vector graphics. Broadly speaking, raster graphics typically uses a grid of colored pixels to build the image, whereas vector graphics uses points, lines and simple geometric shapes. Vector graphics data is quite widespread and is represented by numerous formats, such as shapefiles (one of the most popular spatial data formats), GML (XML-like grammar developed by the Open Geospatial Consortium (OGC)), KML/KMZ (extension of XML for spatial data developed by Google) and others. The prevalence of vector formats for spatial data can easily be explained by its numerous advantages, including compatibility with relational databases and possibilities to easily scale and combine vector layers or update data. For instance, in comparison with raster data, vector format allows more efficient encoding of a topology and hence offers more analysis capabilities for networks, such as roads, rivers, rails and energy networks.

However, data seldom comes clean and accurate, and this statement holds for geospatial data as well. It might be inaccurate or outdated, and, as consequence, the topological structure of vector data can be corrupted leading to incorrect encoding of geographical features. In this paper, we discuss on topological errors in vector data and, in particular, on one of the most frequent problems in the network data: incorrect connections to line features. Further, we demonstrate and evaluate proposed methods on a real-world geographical data.

The rest of the paper is organized as follows: the next section introduces data quality and topology for vector spatial data. Overview of related work is provided in Section III. In Section IV, we propose a method for correcting topological errors in data and further in Section V we evaluate proposed method on our use case data. Section VI concludes the paper.

II. DATA QUALITY AND TOPOLOGY IN GIS VECTOR DATA

In order to keep spatial data as accurate and complete as possible, a set of general quality criteria was defined [1]. These criteria are called *the elements of spatial data quality*:

1) Lineage - the history of the dataset, i.e., how was this data derived and how was the data transformed and processed;
2) Positional accuracy - a measure of accuracy of absolute and a relative positions of geographic features in the dataset;
3) Attribute accuracy - a measure of accuracy of quantitative and qualitative attributes of geographical features;
4) Completeness - a measure of whether all geographical features and their attributes were included in the set and, if otherwise, selection criteria which attributes were omitted;
5) Logical consistency - compliance with the structure of data model, absence of apparent contradictions in data;
6) Semantic accuracy - correct encoding of geographical features, i.e., the difference between geographical features in a given data set and in reality;
7) Temporal information - validity period for a given data set, dates of its observation and any updates performed;

Poor-quality data does not conform to one or several elements of quality. For example, irresponsible documentation affects lineage and temporal information quality; map transformations and generalizations cause attribute and semantic inaccuracies. Other typical sources for deficiencies in data quality elements include data collection, data conversions and transfer between different formats and coordinate systems.

Insufficient data quality is especially critical for vector data since its topology can be disturbed. *A geospatial topology* enforces rules concerning relationships between geospatial features representing real-world objects. These rules are called *topology rules* [2]. They are formulated using spatial predicates such as `Contains`, `Covers`, `Disjoint`, `Intersects`, and others. The geospatial topology determines and preserves relationships between geographical features. For instance, in road or telecommunication datasets topology is what makes

 ISBN: 978-1-61208-469-5

a set of lines to be a network. It is essential for spatial data analysis, e.g., for querying or routing. Different types of topologies are distinguished, depending on the feature classes presenting in the dataset, for example, the arc-node topology defines relations between lines. Similarly, the polygon topology determines relations between polygons [3]. According to the type of the topology and data model requirements appropriate topology rules are defined. For example, both buildings and climatic zones can be encoded by multipolygons, but those representing buildings are allowed to have gaps between them, whereas multipolygons representing climatic zones are forbidden to have void areas between them. Errors in data may lead to violations of these topology rules, incorrect definitions of relationships between features, and, therefore, failure to meet data quality criteria. Such errors are called *topological errors*.

As it was mentioned above, vector data is especially suitable for networks, such as roads, electricity grids and others. Similarly, for each network dataset there are corresponding topology rules defined by a data model, requirements and further characteristics of data. However, some topological errors are typical for all kinds of networks, including: dangling lines, i.e., not precise connection of lines to the other features. These errors occur quite frequently in network data breaking its topology and, as one of the consequences, corrupting results of data analysis. In this paper, we concentrate on dangling lines and propose techniques for connecting them to the ending points correctly.

III. Related work

There was a lot of research on spatial data quality since 1990s, when the geographic information science took its roots. Also for vector spatial data there exist various methods of detection and correction of topological errors.

In general, all features shall be checked for violating defined quality criteria, topology rules or any other restrictions set by the data model. There are two main possibilities to conduct spatial data quality assessment and improvement in practice: using GIS or Computer Aided Design (CAD) software. CAD platforms provide an environment supplying graphic operators and algorithms for data processing, such as checking intersections, creating features, etc. Authors of [4], [5], [6], [7] and other works present systems that detect and correct wide range of topological errors operating objects in CAD environment. Some modules also treat positional inaccuracies and logical inconsistencies, such as identifier duplications [4], and semantic inaccuracies, such as self-intersections of features [7] .

In the first place, GIS is a system for storing and displaying geospatial information. However, present-day GIS software often offers some analysis functionalities, including checking validity of the topology in data. GRASS GIS [8], QGIS [9], ArcGIS [10] and other similar software packages find and fix errors in two- and three-dimensional data. For example, ArcGIS allows to choose from 28 topology rules and detects features that violate these rules [11]. It is important to mention though, that in GIS software checking validity functionality is often aimed rather for faster rendering and simplification then for an efficient data analysis.

One of the most important concepts underlying topological error detection and correction is a *tolerance gap* (also called

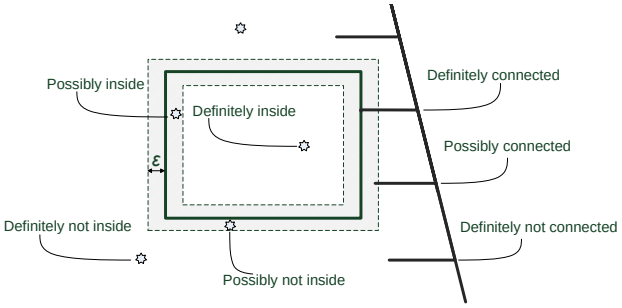


Figure 1. Epsilon-bounded tolerance gaps.

an error band, *a search radius*, or *an epsilon-bounded error region*) which is defined as an area around the feature expanded by epsilon in all directions from the boundary of the feature [12]. Its purpose is to access the cartographic error in feature relations. For example, for each polygon in a dataset with a polygon-point topology epsilon-bounded tolerance gap allows to separate surrounding features in four groups (see Figure 1): (i) definitely lying inside of the polygon, (ii) possibly lying inside of the polygon, (iii) possibly not lying inside of the polygon, (iv) definitely not lying inside of the polygon. Similarly, in case of the polygon-line topology, tolerance gap allows to find possibly connected lines (see also Figure 1). The value of epsilon shall be application dependent and mark out as many of doubtful features as possible. Tolerance gaps are widely used for detection and correction of dangling lines, slivering polygons and other topological errors [12], [3], [13], [5]. Additionally, error bands are used in probabilistic and fuzzy logic approaches [14], [15], [16] for spatial topology, where error band is introduced as an uncertainty in the boundary of features.

However, epsilon-bounded tolerance gaps suggest to connect dangles to the closest feature around. It might be incorrect choice leading to semantic inaccuracies, in case the object, encoded by dangling line, is connected to the other object rather than closest one. Purpose of our work was to correct numerous dangling features in real-world data and to rebuild network topology as accurate as possible. Existing systems mostly use the epsilon-bounded tolerance gap method for dangling lines and, therefore, produced semantic inaccuracies in the dataset. In order to avoid these inaccuracies, we propose a novel method for correcting dangling lines. In this method we suggest to respect the network structure, distinguish features that are already connected to the network from the features that are not yet connected. This method can be especially relevant for road, utility, telecommunication and other network. We also introduce error band in our method as an aid for correcting dangles. In further sections we detail our method for detection and correction topological errors in vector data.

IV. Data quality improvement

Vector data tends to dominate in network and other applications, where it is important to analyze relations between features such as connectivity and adjacency. However, it might not be possible to yield reliable analysis results due to poor data quality affecting the topological structure of the dataset.

One of the common data quality discrepancies for vector spatial data are positional inaccuracies of features. Positional

 ISBN: 978-1-61208-469-5

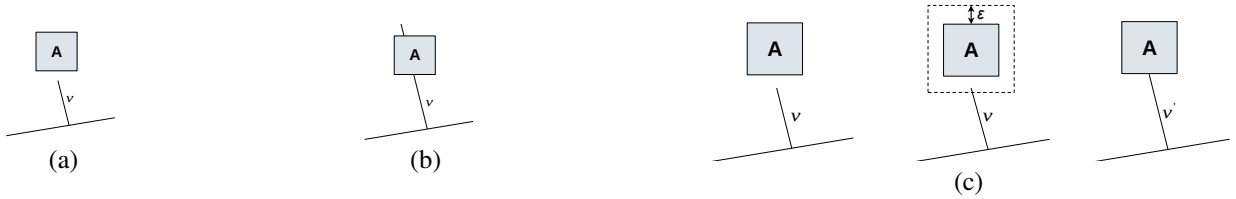


Figure 2. Dangling lines: a) undershoot, b) overshoot, c) correcting dangle using a tolerance gap

accuracy shows whether the geographical position of a feature corresponds to the real-world position of the object it represents. While constructing the dataset and transforming between formats and coordinate systems, geographical characteristics of the feature may be affected. Transformations that lead to positional inaccuracies might not only comprise transformations between coordinate systems, but also map generalizations and transformations of attribute format, e.g., rounding coordinate values.

As it was mentioned in Section II, there are two types of positional accuracy: (i) absolute, that defines an absolute geographical position of a feature, (ii) relative, that defines a position of a feature with respect to the other features. Dangling lines are an example of a relative positional inaccuracy. Line is called dangling, if its beginning or ending point does not agree with any other features. Typically, line features are encoded as a pair of points, multiline features - as a sequence of points. The first point is called *the beginning of the line* and the last point is *the ending of the line*. However, for our approach it is not significant, whether the beginning or the ending of a dangling line is not connected properly. Therefore, throughout the next sections we say *endpoint of a line* without specifying whether it is the first or the last point in the sequence of coordinates encoding the line.

Dangling lines are also often called undershoots or overshoots indicating on the type of displacement of the feature. Sometimes this topology error may be cleared by introducing tolerance gaps. Figures 2a, 2b illustrates examples of an undershoot and an overshoot correspondingly for a line-polygon topology. Figure 2c illustrates the process of restoring the connection by introducing a tolerance gap around polygon A, checking the containment relationship between the tolerance gap of the polygon and the endpoint of the line v, which allows to conclude that line v is possibly connected to polygon A, and finally building a corrected line v'.

In Figure 3 we schematically illustrate several possible situations when simply introducing tolerance gaps is not enough. Positional and relative inaccuracies in different layers may superpose and lead to a situation similar to the one shown in Figure 3a, when the endpoint of the conduit is not reached by the tolerance gap with the defined ε. In this case, we suggest using a tolerance gap for the endpoint of a line and a stepwise increment of ε as shown in Figure 4a. However, increment of ε shall be limited in order not to produce false connections. Figures 3b and 3c show cases, when the line v_2 is possibly connected to several polygons. We suggest two further actions to remove the uncertainty:

1) Build a line with the same slope and offset as the initial line or as the corresponding segment of an initial multiline. Among all candidates choose the feature that lies on the line and is the closest to the endpoint of the dangling line. According to this technique, Figures 3b and 4b shows the connection of the dangle v_2 to the polygon B, since the continuation of a line segment v_2 intersects it.
2) Among all candidates, filter out those features that are already connected to other lines. This remark, however, depends on topology rules specific to the data, but nevertheless is true for many common types of networks. According to the technique shown in previous point, in Figures 3c and 4c polygon A is the endpoint of the dangle. However, filtering out polygons A and C that are both connected to other lines, we choose polygon B out of all candidates.

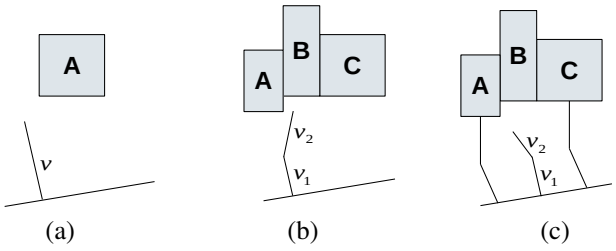


Figure 3. Other possible occurrences of dangling lines.

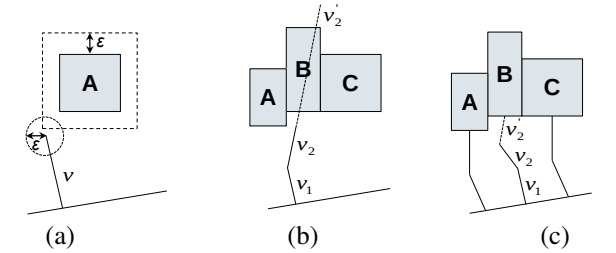


Figure 4. Correction of dangling lines for cases introduced in Figure 3.

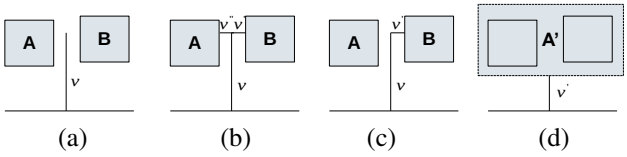


Figure 5. Different possibilities of treating dangling line in case it is unclear to which feature it is connected to.

Candidates for a correct connection of a dangle usually are determined by an exhaustive search, as in [4] and other works. However, it might be computationally intensive in case of a

large dataset. We suggest to reduce search of the candidates by a part of a dataset, i.e., analogously with an epsilon value defining the width of tolerance gaps, we suggest to define an Ω value determining a search area around the dangling feature, i.e., as an envelope around the dangle expanded by Ω in all directions.

One of the indecisive cases that can not be solved using technique suggested above is shown in Figure 5a. In this case, both candidate polygons A and B are not connected to another lines, equidistant from the dangling line v and do not lie on the line of the dangle. The best solution in this case is to consult a domain expert who knows the data and can exactly point out the correct relationship between features. Otherwise, there are the following possibilities how to treat this problem: (i) leave it as it is (see Figure 5a), (ii) connect both polygons (see Figure 5c), (iii) choose (ot guess) which polygon to connect (on Figure 5a polygon B is chosen), or (iv) merge two polygons into a new bigger polygon A' connected to the endpoint (see Figure 5d).

Each of these methods has its pros and cons. For example, the methods (ii) and (iii) are suitable if there are few such cases, but if there are a lot of them, it creates an abundance of synthetically introduced connections and, as a consequence, might result in geospatial topology which contains a lot of semantic and positional inaccuracies, and as a consequence has nothing to do with real-world topology. The latter method is not suitable, when these polygons represent different objects in real world and have completely different characteristics. The semantic accuracy of the dataset is affected by merging two polygons into one. On the other hand, ignoring dangling features might lead to oversimplification and logical inconsistencies, for example, when we ignore a dangling line that is supposed to connect buildings to an electricity substation, and there are no other substations in the network delivering electricity to consumers. The choice of a particular method depends on a field of application, type of data, task at hand and quality criteria for a particular dataset.

```
1:  for i = 0 to n do
2:     if ∄f ∈ F s.t. beginning of l_i ∈ L connected to f and
       ∄l ∈ L, l ≠ l_i s.t. beginning of l_i ∈ L connected to l
       then
3:        L' ← l_i
4:     else if ∄f ∈ F s.t. ending of l_i ∈ L connected to f and
       ∄l ∈ L, l ≠ l_i s.t. ending of l_i ∈ L connected to l then
5:        L' ← l_i
6:     end if
7:  end for
8:  for i = 0 to m do
9:     if ∃l ∈ L such that f_i ∈ F connected to l then
10:       F' ← f_i
11:    end if
12: end for
```

Figure 6. Inspecting lines and features in dataset, building L', F' sets .

We suggest algorithms shown in Figures 6 and 7 summarizing techniques we introduced above. We use the algorithm in Figure 6 to process all data and to determine spatial relationships between features. Let $\{L\}_{i=1}^{n}$ be a set of line features and $\{F\}_{i=1}^{m}$ be a set of point and polygon features.

```
1:  for all l' ∈ L' do
2:     for all f ∈ F s.t. f is contained in E(l', Ω) do
3:        if ending of l' is contained in e(f, ε) then
4:           F(l') ← f
5:        end if
6:     end for
7:     if |F(l')| = 1 then
8:        return connection l' to f
9:     else if |F(l')| > 1 and |F(l') \ F'| = 1 then
10:       g = F(l') \ F'
11:       return connection l' to g
12:    else if |F(l')| > 1 then
13:       for all f in F(l') do
14:          if extension of l' crosses f and f ∉ F' then
15:             return connection l' to f
16:          else if extension of l' crosses f and |F(l')\F'| = ∅
             then
17:             return connection l' to f
18:          else if f ∉ F' and f is the closest to l' then
19:             return connection l' to f
20:          end if
21:       end for
22:    else if F(l') = ∅ then
23:       repeat
24:          increase ε
25:       until F(l') ≠ ∅ or ε < threshold
26:    end if
27: end for
```

Figure 7. Detecting and correcting dangling lines in a vector spatial dataset.

The algorithm builds sets $L' \subseteq L$ of dangling lines and $F' \subseteq F$ of features having connection to lines. The first *for* loop iterates over all n line features in the dataset; l_i denotes the current line. During the loop the *if* conditionals on lines 2 and 4 check, whether there are no feature f from the set of all features F and no line l from the set of lines L that are connected to a beginning or to an ending of l_i. If it is the case for at least one endpoint of the line l_i, it is added to a set of dangling lines L' as shown on lines 3 and 6 of the algorithm. The second *for* loop iterates over set F of features and checks whether there exists line l that is connected to this feature. If yes, feature f_i is added to a set F' of connected features. Thus, the set F' of features connected to lines and the set $F \backslash F'$ of unconnected features are built. These two procedures are separated in the pseudo code in Figure 6 for an easier understanding where do sets L' and F' come from. However, set F' can be built during the first *for*-loop.

Algorithm listed in Figure 7 performs data cleaning. It consists of one *for* loop that iterates over a set L' of dangling lines and attempts to connect it to a feature. Firstly, we introduce values Ω, ε that defines a size of an envelope $E(l', \Omega)$ around any dangling line l' and an epsilon-bounded tolerance gap $e(f, \varepsilon)$ around any feature f. In a nested *for* loop we iterate over features from $E(l', \Omega)$ and building a set $F(l')$ of features that are possibly connected to dangle l' using epsilon-bounded tolerance gap method. The cardinality of the set $F(l')$ determines the next actions. If $F(l')$ has only one element, then a connection between this element and the endpoint of l' is restored (see *if* conditional on lines 7-9). In case $F(l')$ is not empty and has more than one member, we

TABLE I. Networks size in one of the districts in Geneva canton in Switzerland

	Lines	Points and Polygons
Electricity	524	338
District heating	164	120
Water treatment	821	263
Buildings	-	407
Σ	1509	1128

apply techniques elaborated above. In particular, we use a *for* loop on lines 13-21 to search for the nearest feature that is not the member of the set F' and crossed by the extension line with the same slope and offset as the line l'. If all candidate features are not the members of the set F', we connect l' to the feature that is the nearest in crossed by the extension line. Finally, if $F(l')$ is empty, we increase tolerance value ε and repeat the search for candidates, unless ε can not be increased anymore (see *if* conditional on lines 22-26). Note, that in Figure 7 we process lines with dangling endings, in case of dangling beginning of a line procedures are similar.

V. Case study: urban energy networks

A. Data cleaning

We applied data cleaning techniques introduced above to urban energy networks data. In this section, we describe this data and demonstrate the result of application of geoprocessing procedures.

The authors of this paper work in the European project "CI-NERGY" [1], which aims to develop urban decision making and operational optimization software tools to minimize non-renewable energy use in cities. In particular, the authors' expertise lies in the area of analysis of energy networks. Spatial data plays a crucial role since it provides a topology of the network, precise geographical positions of network equipment and consumers as well as connections between them. We perform routing, breadth-first, depth-first and other algorithms on the spatial data. Therefore data of sufficient quality is especially crucial to gain as precise layout of the network as possible and to produce meaningful results of analysis. One of the case studies in the project is the canton of Geneva, located in the south-western corner of Switzerland. Geneva energy networks data was provided by SIG Geneva[2].

We evaluated our methods on one of the districts in canton Geneva, Table I provides a short overview of its network size. The provided spatial data is stored in ESRI shapefile format and consists of a building layer and network layers, typically three layers per network. Buildings are represented as polygon and multipolygon features. Points, polygons and multipolygons depict installations and other equipment in networks, and lines and polylines depict conduits and pipes connecting those installations and buildings. However, the data included some positional inconsistencies caused by data losses and errors during conversion from internally used format to commonly used ESRI shapefiles. In particular, it concerned network conduits layers representing connection of the buildings and other objects to networks, which resulted in dangling features, not precisely connected to networks. Based on this information, we concluded that positional accuracy of buildings is not disturbed and dangling features and imprecise connections are caused by inconsistencies in network conduits spatial data rather than in building layers. Moreover, different network layers suffered from different displacement of features. Such, electricity network was the least affected and in most cases missing connections could be restored using tolerance gap technique with ε value not increasing 5-10 meters, whereas for the water network large ε values were needed in order to find features possibly connected to dangling lines. In dense districts of the city it resulted in large search sets and a need to choose which feature to connect.

[1]The CINERGY, Smart cities with sustainable energy systems Marie Curie Initial Training Network (ITN) project: http://ci-nergy.eu/About.html

[2]SIG: Swiss supplier of local energy services http://http://www.sig-ge.ch/

TABLE II. Comparison of number of buildings connected to the network

	Electricity	DH	Water
Uncleaned data	137	4	37
Proposed technique	139	37	93
Clients in the database	140	37	101

We implemented data cleaning and graph construction methods in Java. We aimed to create a module that would be independent from existing GIS or CAD software and could detect and process dangles in datasets with line-line, line-point and line-polygon relationships. We used GeoTools Java library [22] for manipulation with shapefiles and geometries.

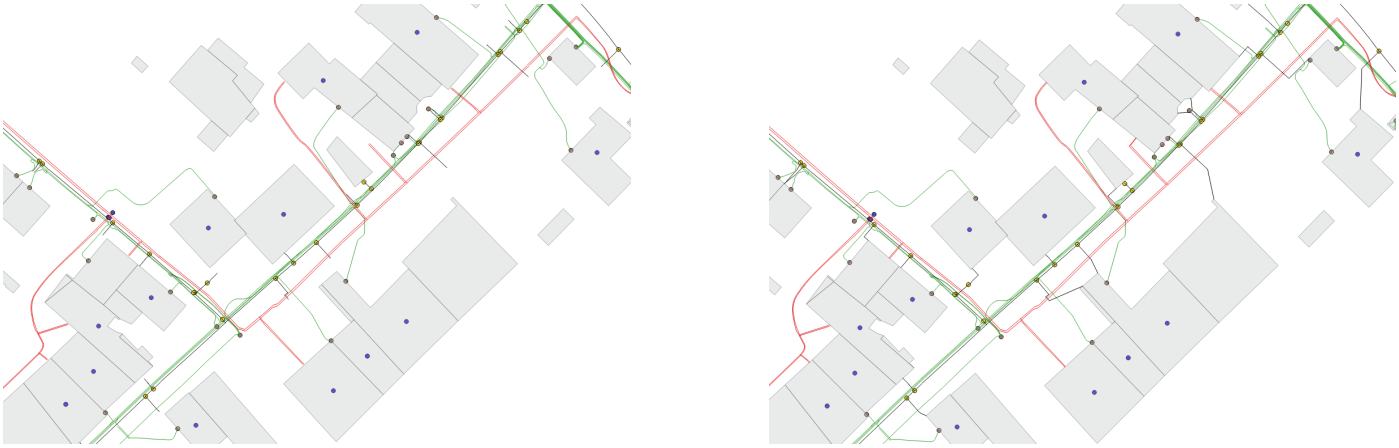
Figure 8 illustrates the results of application of data cleaning procedures to the real-world data. Electricity lines are shown as green lines, water pipes - black, district heating pipes - red. We used QGIS software for visualization of shapefiles [9].

B. Evaluation

Apart from the geospatial data we have also received aggregated consumption data from our city partners in Geneva. That allows us to evaluate our approach. We apply path search algorithm on the networks before and after data cleaning and compare, which buildings are connected to the network and which are listed in the client database. Dangling lines result in the absence of the path between buildings, that are connected to the network. In Table II we compare results of search of connected buildings before and after data cleaning. Electricity grid data had a sufficient quality, being almost completely connected and containing only a couple of dangling lines. On the contrary, district heating and water networks had poor quality and in most cases lines representing pipes did not connect to the polygons representing buildings. For electricity and water, method returned a very good result without false negatives and false positives, connecting correct clients to the network in the spatial data. In case of the district heating our procedure resulted in a complete connectedness of the network and correction of each dangling line. For the water network unfortunately it was not the case, as it contains multiple occurrences of a situation shown in Figure 5a. Therefore, out of 101 building that shall be connected to the water network only 37 were correctly connected in the initial dataset and 93 were correctly connected using procedures described in this paper.

VI. Conclusion

In this paper, we considered data quality in vector data, and, in particular, data quality inconsistencies corrupting geospatial



a) b)

Figure 8. Correcting dangling features: a) initial data containing not precisely connected features, b) Result of application of cleaning procedures

relations between features. These relations are especially important for application of spatial analysis algorithms. We considered one of the most frequent topological error in topologies of types line-line, line-point and line-polygon and suggested a method for their correction. This method allowed to restore lost connections in urban utility network data. Accurate topology is essential for a network analysis, that we work on using a graph-based model representing the geospatial topology of networks. However, as for each data cleaning algorithm, there is a danger of overcorrecting data and thus ending up with even worse data quality than before. In future, we are planning to improve our method taking into account further topology rules, domain knowledge and other characteristics of the dataset.

ACKNOWLEDGMENTS

The authors thank colleagues from SIG Geneva who provided case study data that greatly assisted our research. Moreover, the first author gratefully acknowledges the European Commission for providing financial support during conduct of research under FP7-PEOPLE-2013 Marie Curie Initial Training Network CI-NERGY project with Grant Agreement Number 606851.

REFERENCES

[1] S. C. Guptill and J. L. Morrison, Elements of spatial data quality. Elsevier, 2013.

[2] K.-t. Chang, Introduction to geographic information systems. McGraw-Hill Higher Education Boston, 2006.

[3] H. Hansen and L. Grondal, “Methods for cross-referencing, consistency check and generalisation of spatial data,” SENSOR Project Deliverable Report 5.1.2, 2006.

[4] M. Siejka, M. Ślusarski, and M. Zygmunt, “Correction of topological errors in geospatial databases,” International Journal of Physical Sciences, vol. 8, no. 12, 2013, pp. 498–507.

[5] S. S. Maraş, H. H. Maraş, B. Aktuğ, E. E. Maraş, and F. Yildiz, “Topological error correction of gis vector data,” International Journal of the Physical Sciences, vol. 5, no. 5, 2010, pp. 476–583.

[6] H. Gu, T. R. Chase, D. C. Cheney, D. Johnson et al., “Identifying, correcting, and avoiding errors in computer-aided design models which affect interoperability,” Journal of Computing and Information Science in Engineering, vol. 1, no. 2, 2001, pp. 156–166.

[7] A. A. Mezentsev and T. Woehler, “Methods and algorithms of automated cad repair for incremental surface meshing.” in IMR, 1999, pp. 299–309.

[8] “GRASS: Geographic resources analysis support system,” https://grass.osgeo.org/, accessed: 2016-02-20.

[9] “QGIS: A free and open source geographic information system,” http://www.qgis.org, accessed: 2016-02-20.

[10] “ArcGIS: Commercial GIS software application,” https://www.arcgis.com, accessed: 2016-02-20.

[11] ArcGIS, “ArcGIS geodatabase topology rules,” http://help.arcgis.com/en/arcgisdesktop/10.0/help/001t/pdf/topology_rules_poster.pdf, accessed: 2016-02-20.

[12] M. Blakemore, “Part 4: Mathematical, algorithmic and data structure issues: Generalisation and error in spatial data bases,” Cartographica: The International Journal for Geographic Information and Geovisualization, vol. 21, no. 2-3, 1984, pp. 131–139.

[13] G. Klajnšek and B. Žalik, “Merging polygons with uncertain boundaries,” Computers & geosciences, vol. 31, no. 3, 2005, pp. 353–359.

[14] M. Schneider, “Fuzzy spatial data types for spatial uncertainty management in databases.” Handbook of research on fuzzy information processing in databases, vol. 2, 2008, pp. 490–515.

[15] W. Shi and K. Liu, “A fuzzy topology for computing the interior, boundary, and exterior of spatial objects quantitatively in gis,” Computers & Geosciences, vol. 33, no. 7, 2007, pp. 898–915.

[16] X. Tong, T. Sun, J. Fan, M. F. Goodchild, and W. Shi, “A statistical simulation model for positional error of line features in geographic information systems (gis),” International Journal of Applied Earth Observation and Geoinformation, vol. 21, 2013, pp. 136–148.

[17] M. Neteler and H. Mitasova, Open source GIS: a GRASS GIS approach. Springer Science & Business Media, 2013, vol. 689.

[18] A. S. Analyst, “Advanced gis spatial analysis using raster and vector data,” An ESRI White Paper, ESRI (Environmental Systems Research Institute), Redlands, USA, 2001.

[19] “OSRM: routing engine for shortest paths in road networks,” http://project-osrm.org/, accessed: 2016-02-20.

[20] “pgRouting: an extension for PostGIS and PostgreSQL providing geospatial routing functionality,” http://pgrouting.org/, accessed: 2016-02-20.

[21] “Flowmap: a software package for analyzing and displaying spatial flow data,” http://flowmap.geo.uu.nl/, accessed: 2016-02-20.

[22] “GeoTools: The open source Java GIS toolkit,” http://geotools.org/, accessed: 2016-02-20.

 ISBN: 978-1-61208-469-5