Provide a Real-World Graph Suitable for the Mathematical Optimization of Communication Networks

Markus Prossegger
Department of Engineering and IT
Carinthia University of Applied Sciences
Klagenfurt, Austria
m.prossegger@cuas.at

***Abstract*—One of the main reasons of the hesitant expansion of fiber optic communication networks is the large financial expense for the excavation work. While the cost for (passive) hardware components and the fiber optics cable itself have fallen during the last decade, the cable layout work is still the cost driver of network construction projects. The most promising approach for a valid cost estimation is to use state-of-the-art network simulation and optimization techniques in the field of operations research. This study examines an automated process of generating a network graph that closes the missing link between the real-world and mathematical optimization of a communication network. The constructed graph is based on heterogeneous spatial data and weighted with real-world construction costs. It is then used to solve the minimum Steiner tree problem as typical optimization problem for the modeling and optimization of communication networks. While the applied mathematical model is studied in detail, the quality of the result as well as the runtime performance of the optimization algorithm is heavily dependent on the complexity and validity of the input graph. Based on a general format, the normalized geobasisdata, an initial graph, is constructed. This graph is then used as input into our rule-based system to select and weight the edges to be in the final graph. The optimization results of the experiments on real-world data, prove the effectiveness and efficiency of the proposed approach.**

Keywords*—*geographic information; spatial data; communication network optimization; normalized geobasisdata.

I. Introduction

This study is an extension of [1] and examines the generation of network graphs, weighted with real-world construction costs to allow a valid optimization of communication networks. A typical network optimization problem is to connect a given number of customers by a wired network at a minimal expense. This state-of-the-art optimization problem is known as the Steiner tree problem (STP) that searches a cost minimal tree connecting terminals in a weighted network graph described in [2] and [3].

The generation of graphs, representing road networks, is subject of different approaches (e.g., the extraction of road networks from satellite images as described in [4] or the extraction of road intersections from raster maps described in [5]). The generation of real-world graphs, allowing the optimization of communication networks, is understudied.

To enable the mathematical simulation and optimization of a wired telecommunication network, the spatial data has to be converted into a weighted network graph, which is the missing link between the real world and the mathematical modeling. This graph is consisting of vertices (i.e., points-of-interest) and pair wise joining edges (representing network construction costs) between them. It is generated using spatial polygon data describing the land use on the one hand, and spatial line- and point-objects, describing existing infrastructure on the other hand. Based on this spatial data originating from a number of hybrid sources, a rule-based expert system is used to construct a network graph as vital input in subsequent mathematical models.

We focus on generating a weighted graph that can be used in different optimization models within the scope of wired telecommunication network construction. An instance of such a model using weighted graphs is the simulation and optimization engine of fiber optic communication networks described in [6]. Here, the land use polygons are divided into a grid, called cost raster. Each cell in this grid is representing the averaged underground construction costs determined by network constructors. While using a cost raster is a feasible way to generate a network graph representing network construction costs, a more sophisticated approach is needed to generate graphs by taking into consideration all kinds of real world information and being able to be computed in a reasonable time.

The present paper is divided into a preliminaries section, the section dedicated to the definition and origin of normalized geobasisdata, details and quality of our approach followed by experimental results and the conclusion.

II. Preliminaries

The subsequent simulation and optimization algorithms require undirected graphs as the fundamental data structure. The graph $G = (V, E, d)$ consists of $n = |V|$ vertices and $m = |E|$ edges. The distance of an edge $e_{ij} \in E$ connecting the two vertices $i \in V$ and $j \in V$ is given as a cost function $d_{ij} : E \to \mathbb{R}$, where $\mathbb{R}$ is the set of all real numbers. When calculating the shortest path or the minimum Steiner Tree as typical routing problems, the distance d can be the Euclidean distance. The more sophisticated algorithms use travel time as the distance between two vertices. In case of scenarios considering the cable layout of wired networks, the distances have to be construction costs, such as underground work or the costs for building cable poles.

The goal of optimizing the cable layout of a communication network is to find a connected sub graph $S = (V_s, E_s)$ in G, connecting the terminals (i.e., access objects) $T \subseteq V$ such that the sum of edge weights $\sum_{e \in E_s} d_e$ is minimal. Verticals from the set $V \setminus T$ are called Steiner nodes.

The two-dimensional geographic data originate from hybrid sources, thus these data need to be prepared to serve as the basis for the construction of network graphs. There are three main sources for the geographical data:

(a) The geographic information system (GIS).
(b) The network information system (NIS).
(c) The digital cadastral map (DKM).

Each of the above listed items is needed for the construction of a consistent data source.

A. GIS

In our case, the geographic information system includes typical information used in marketing scenarios and strategic decisions. It is important to know where the potential customers are located. The data showing the population density or the number of households collected in a population census are incorporated in the GIS as well. The GIS contains statistical data aggregated from public sources together with information gathered by the prosecuting company itself. The most important information for a network construction company is the information about the location of potential, private or public customers as well as the expected benefit.

The network operator knows the exact location and the return on investment of the current customers, but not for potential customers. The marketing division uses market surveys and other statistical data to predict the location of potential customers and the likely yearly sales.

B. NIS

The network information system contains the information regarding all hardware components of the communication network as well as all logical links between these components. Typically, the NIS contains the most important business secrets of a network operating company. The following gives a list of the typical content of a NIS:

- Current and former customers.
- Network components.
- Physical cabling plan.
- Logical interconnection plan.

C. DKM

The Digital Cadastral Map is part of the official boundaries cadastre, which is the binding evidence of all parcel's boundaries. The DKM contains all public and private property and is typically available nationwide. Furthermore, it documents the type of land use of each parcel as well as buildings. Similar information is held in layers and together they form the DKM (a comprehensive interface description can be found in [7]):

- Boundaries.
- Parcel numbers.
- Types of land use (building land, forest, running water, standing water, etc.).
- Buildings.
- Control and Boundary points.

Formerly available only as an analog hard copy, the DKM was not only digitized but also enhanced using other official sources like Orthophotos and partition plans. Due to this reason, the quality of the digital map exceeds the quality of the analog version but it may include historical failures as well. There is a list of papers describing the aspects of spatial data quality [8], [9], [10], [11] as well as an ISO Standard regarding the quality of spatial data [12].

While the spatial accuracy is acceptable in most of mathematical simulation and optimization scenarios, the topological quality of the input data has to be ensured. There is some work proposed to identify spatial inconsistencies and incorrect object classifications using either manually defined spatial integrity constraints [13], [14], [15] or an automatic and incremental approach using decision trees proposed in [16] and improved in [17].

The DKM is used as one of the basic input into our approach and has to be normalized together with the other spatial input data.

III. Normalized Geobasisdata

The subsequent graph generation is designed as a completely automated process without the need of any user interaction. Due to this fact, the input data are stored in a predefined digital map format and the spatial objects must meet a set of conditions. In our approach, we have decided to use the ESRI Shapefile [18] as the digital map format. This open format is widely used and supported and stores spatial geometry and attributes as elements representing points, lines, and polygons.

The Normalized Geobasisdata (NGB) format [19] is an add-on to the ESRI Shapefile specifying the minimum qualitative and logical requirements of the spatial objects. It was developed in order to allow the automated generation of weighted network graphs based on any two-dimensional spatial data that represent surface data (i.e., land uses) in form of polygons at least. If the hybrid spatial data fulfill the specified NGB format, they qualify as input into the graph generation process.

The NGB format was originally developed for a simulation model dealing with the layout planning of a fiber-optics communication network in the year 2009. Since then it has been continuously adapted to the specific requirements of individual projects.

The majority of mathematical simulation and optimization models dealing with cable infrastructure planning or routing in general rely on spatial data as the main input source to stay real world compatible. Table I represents the spatial information to be covered in the NGB format.

TABLE I
NGB OBJECT TYPES AND THEIR SPATIAL REPRESENTATION.

Id	Object class	Spatial representation
a	Project area	Polygon
b	Land use	Polygon
c	Usable (own or third-party) infrastructure	Polyline
d	Infrastructure points	Point
e	Access points	Point

The polygon describing the project area (a) as well as each polygon describing the land use (b) must show the following characteristics (in addition to some mandatory attributes described below):

- Valid and closed polygon.
- No crossing edges.
- Degree-two vertices only.
- No overlapping or equality with other polygons (see Figure 1).

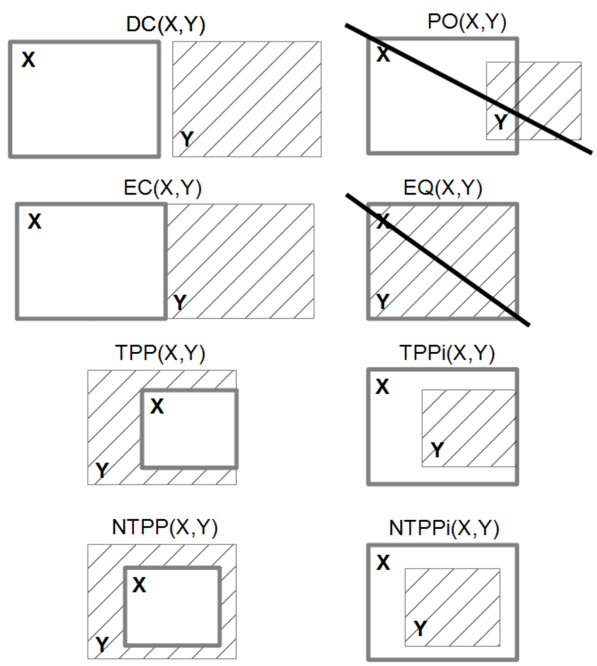


Fig. 1. All existing relations between pairwise polygons. EQ (equal) and PO (partly overlapping) relations are not allowed in NGB; Based on Region Connection Calculus (RCC-8) [20].

A polyline referred to as usable infrastructure (c) can be associated with own infrastructure (e.g., the copper cabling in the access domain) or third-party infrastructure (e.g., leased lines). The spatial object must be a valid polyline. In case of any attributive restriction in the accessibility, there must be at least two infrastructure points assigned; hence the infrastructure can only be accessed via one of the two points (an open line infrastructure can be accessed at the masts only).

The infrastructure points (d) are attributive assigned to exactly one polyline of the usable infrastructure and represent any type of infrastructure objects that can be localized on exactly one position (e.g., shafts to access pipes, masts, hardware like splitters or routers, etc.).

The last object class is the class of access points (e). These are points representing the terminals in a following optimization. We distinguish between existing access points that are currently supplied by one or a group of connected usable infrastructure polylines, and potential access points that are not yet connected.

The presented approach of the graph generation makes use of geometric algorithms and algorithms from operations research. Thus, ambiguous relations or even gaps between spatial objects cannot be allowed. This distinguishes between common geodatabase systems and spatial data in the NGB format, because there are no tolerances allowed.

Two points meant to be on the same position have to share the same coordinates. Furthermore, there are general specifications, which cover the notation, the coordinate system, the locale, default attribute values, and a list of common abbreviations.

The next section describes our graph generation approach, that is based on spatial input data in NGB format.

IV. GRAPH GENERATION

The process that we follow to generate a weighted network graph consists of three consecutive stages (see Figure 4 for the individual results):

- NGB preprocessing and enhancement.
- Generation of the candidate graph.
- Running the rule-based system.

In the following section, further details to the stages will be given.

A. NGB preprocessing and enhancement

The preprocessing and enhancement of the input data are fully automated processes. As long as the input data fulfill the requirements in the NGB format, the generation of a weighted network graph will succeed. Moreover, the quality and usability of the generated graph are crucial in terms of topological errors within the spatial data. Furthermore, the succeeding process of assigning the correct weights to all edges in the graph is sensitive to the correct spatial classification.

To ensure a valid real-world graph, the input data are validated running the decision tree approaches described in [16] and [17]. Based on error free spatial data covering provincial, rural, suburban, and urban areas, a representative decision tree was constructed. Both approaches use this decision tree to validate the input data. The process will output warnings in case

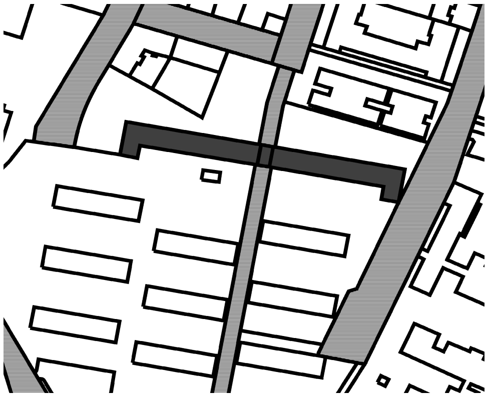


Fig. 2. The street (hatched area) is erroneously disconnected due to the building (dark filled area) that covers the thoroughfare.

of topological errors and reclassify spatial objects according to the decision tree. An example of an automatically identified error can be seen in Figure 2. The missing information about the thoroughfare can lead to a poor solution or even to an insoluble routing problem in consequence of the isolated street.

The validation is followed by the enhancement of the input data. Since underground work in crossroads areas and the subsequent obstruction in traffic should be avoided, the roadways are supplemented by polygons representing crossroads areas. Each crossing of at least two center lines of street polygons is identified and replaced by a polygon classified as crossroad area. The process runs automatically and produces results (see Figure 3), eminently suitable in the generation of valid candidate graphs.

B. Generation of the candidate graph

The goal of any graph-based mathematical optimization is to connect the access points to a given or new access network. As a result of the graph generation, each spatial object will be connected with other objects in the candidate graph. In the approach we follow three consecutive steps:

- Construction of the initial candidate graph.
- Classification of the graph edges.
- Assign real-world construction and activation costs.

Algorithm 1 describes the construction of the initial candidate graph that is used as an input into the rule-based system. Starting with the polygons representing the land use objects, the outer border and hole edges are collected and added to the graph. It is extended by the edges representing existing infrastructure considering restrictions in the accessibility. Additional projections originating from access points as well as artificial polygon crossings ensure a connected graph. The edges are classified with the corresponding land use and infrastructure running Algorithm 2.

Subsequently the classified edges are weighted with real-world construction and activation costs running Algorithm 3. The exemplary construction costs with respect to land use classifications and averaged activation costs of existing communication infrastructure are shown in Tables II and III, respectively. The values are representing typical costs of underground work including the surface reconstruction, with the tendency to avoid airport and buildings as well as the crossing of railways, rivers, and lakes. Since the network constructing companies have their own empirical values, the construction and activation costs can be specified individually.

(a)

(b)

Fig. 3. (a) Original NGB street polygons and (b) enhanced by crossroads area polygons.

The rule-based system is applied to the generated and weighted candidate graph to significantly reduce the dimension (i.e., number of vertices and edges).

C. Running the rule-based system

The rule-based system is based on the expert knowledge of network constructors. It is applied to the generated candidate graph to test for qualified edges. Each of the following questions will be answered with yes or no and determine the appearance of the edge in the final graph (edge is rejected, if all answered with no):

Algorithm 1 : Generation of the candidate graph

Import all polygons P.
Import all infrastructure polylines L.
Import all infrastructure points I.
Import all access points A.
Import specifications regarding additional crossings.

for all polygons $p \in P$ **do**
  Create edges representing the border of p.
  if p encloses an access point $a \in A$ **then**
    Create (orthogonal) projections from a to p.
  end if
  if p should be enhanced with crossings **then**
    Create a crossing all x meter.
  end if
end for
for all polylines $l \in L$ **do**
  if l has restricted access at two or more points $i_{1..n}$ **then**
    Create edges between the points $i_{1..n}$ representing l.
  else
    Create edges from the polyline l.
  end if
end for
for all infrastructure points $i \in I$ **do**
  if i hits any created edge e **then**
    Split e into e_1 and e_2 at location i.
  else
    Create (orthogonal) projections to connect i.
  end if
end for

TABLE II
EXEMPLARY CONSTRUCTION COSTS WITH RESPECT TO LAND USE CLASSIFICATION.

Land use (extract)	Construction costs [€/meter]	
	along border	crossing
Agricultural land	50	100
Airport	1,000	1,000
Building	1,000	1,000
Building land	300	600
Freeway	140	500
Parkland	80	160
Plantation	200	400
Railway	100	1,000
River	100	2,000
Lake	250	2,000
Street	90	300
Vineyard	250	500
Woodland	400	400

1) The edge is needed to ensure a connected graph.
2) The edge is part of the existing infrastructure.
3) The edge is not part of any land use to be filtered (compare Figure 5).
4) The edge is part of a good (cost-efficient) possibility to connect spatial objects (access/infrastructure points).

TABLE III
EXEMPLARY ACTIVATION COSTS WITH RESPECT TO INFRASTRUCTURE CLASSIFICATION.

Infrastructure (extract)	Activation costs [€/meter]
Leased line	12
Own fiber optic cable (buried)	1,50
Own fiber optic cable (open line)	2

Algorithm 2 : Classification of edges

Import all edges E that are member in graph G.
Import all polygons P.
Import all infrastructure polylines L.

for all edges $e_i \in E$ **do**
  if e_i intersects any other edge $e_j \in E$, where $e_i \neq e_j$ **then**
    Split e_i and e_j at the intersection point.
    Add splitted edges to set E.
    Remove e_i and e_j of set E.
  end if
end for
for all edges $e \in E$ **do**
  if e is surrounded by polygon $p \in P$ **then**
    Assign $landuse(p)$ as classification to edge e.
    Label e as crossing edge.
  else if e is on border of polygon $p_i \in P$ and $p_j \in P$ **then**
    Assign $landuse(p_i)$ and $landuse(p_j)$ as classification to edge e.
    Label e as border edge.
  end if
  if e is part of infrastructure $i \in I$ **then**
    Add $type(i)$ as classification to edge e.
  end if
end for

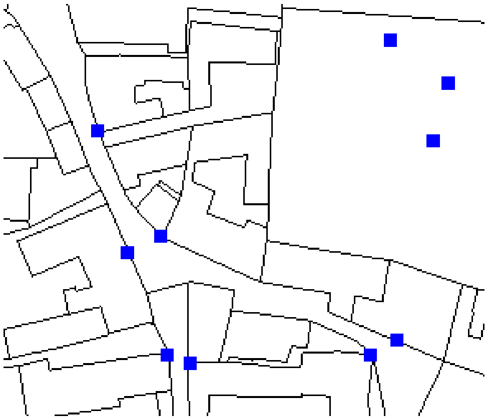
Algorithm 3 : Weighting of edges

Import classified edges E that are member in graph G.
Import construction costs C_L from Table II.
Import activation costs C_I from Table III.

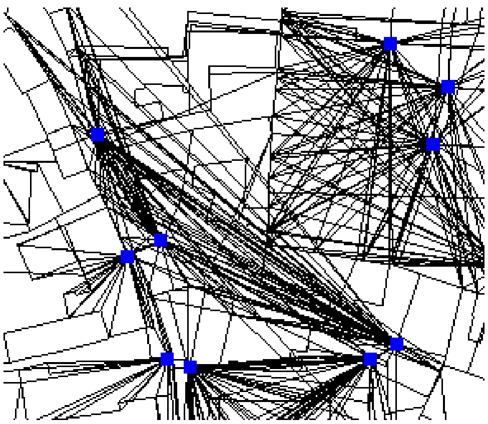
for all edges $e \in E$ **do**
  Select the cheapest classification c of edge e considering the border/crossing label.
  Assign $length(e) \times c$ as weight to e.
end for

While for the first three questions the answer is quite easy to find, the last question requires running Algorithm 4 to be answered. The idea is to keep edges that allow the cost-efficient connection of spatial point objects regardless of the direction. Crossing edges will be discarded, if following the border is cheaper.

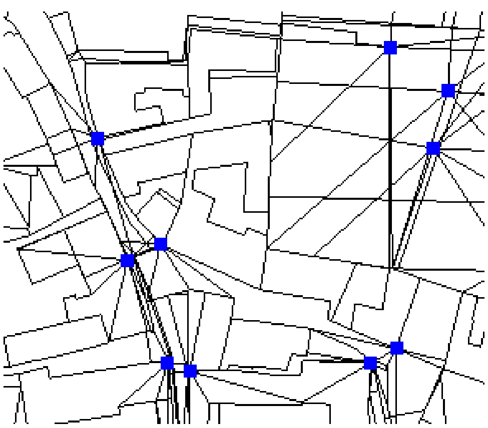
After identifying all edges that are members in the final graph, it can be used in the mathematical optimization of a



(a) Input data in NGB format



(b) Candidate graph



(c) Generated graph

Fig. 4. (a) Preprocessed and enhanced input data, (b) the candidate graph according to Algorithm 1, and (c) the resulting generated graph. Access points are shown as squares.

wired communication network.

V. Quality

The quality of the mathematical optimization result is heavily dependent on the complexity and validity of the generated graph. The main factors determining the quality of the generated network graph are:

(a) Unfiltered graph

(b) Filtered graph

Fig. 5. (a) Unfiltered graph and (b) filtered by agricultural land uses.

Algorithm 4 : Cost-efficiency heuristics

Import weighted edges E that are member in graph G.
Import access points A.
Import infrastructure points I.
Import polygons P.

for all point objects $o \in A \cup I$ **do**
  Select all incident edges $E_o \subset E$ of o.
  Cluster E_o into subsets representing $90°$ sectors.
  Keep the lowest weighted edge $e \in E_o$ in each sector.
end for
for all polygon $p \in P$ **do**
  Select all edges $E_p \subset E$, crossing p or following the border of p.
  for all edges $e \in E_p$ connecting the vertices v_i and v_j **do**
    Discard e, if a shorter path between v_i and v_j is found in E_p.
  end for
end for

(a) The usability.
(b) The real-world correlation.
(c) The correct cost assignment.

The (a) usability is primarily determined by the subsequent optimization algorithm. The dimension of the weighted graph (i.e., the number of vertices and edges) must be small enough to allow the application of heuristics or optimal algorithms but at the same time it also must be large enough to be non-restrictive. So a crossing of a river is a potentially unwanted scenario (due to the expense) but has to be made possible to reach isolated areas.

The (b) real-world correlation of the generated graph is equal to the correlation of the spatial data, if:

- all boundaries of parcels and buildings,
- all infrastructure, and
- all customers are contained and accessible.

To ensure the (c) correct weighting of the edges, the system is supplied with average network construction costs, broken down to activation/usage costs for infrastructure and excavation costs for each available land use. For the latter the granularity of excavation costs can be chosen freely, depending on the spatial information available. Due to the costs for a surface reconstruction or applying for official permits the crossing of a bitumenized street can be more expensive than to dig up a horizontal shaft along the boundary.

In the next section, we will describe the outcome of experiments we ran using real-world spatial data.

VI. Experimental Results

A series of experiments was run to evaluate the performance of the described graph generating approach as well as determining the quality of the weighted network graph. The algorithms were implemented using the object-oriented programming language C# (respectively the functional programming language F# implementing the rule-based system) from the .NET Framework version 4.5.1. All experiments were executed on a standard personal computer with Intel Core i5-2400 CPU, 8GB RAM, and 64-bit architecture.

A. Performance

To analyze the performance of the graph generation approach, we selected spatial data from four different classification areas:

- Urban;
- Suburban;
- Rural, and
- Provincial.

Table IV provides the classification of the exemplary selected input data together with the dimensions.

The dimension and runtime of the generated and weighted network graphs can be seen in Table V.

As expected, the runtime of the approach corresponds to the number of spatial objects enclosed in the area. For the simple reason that the graph needs to be generated only once, the duration is reasonable and acceptable.

TABLE IV
Selected classifications and dimension of enclosed objects.

Area	# Polygon Objects	# Line Objects	# Point Objects
urban	20,569	59,408	33,090
suburban	18,497	30,997	16,361
rural	5,255	21,850	11,800
provincial	792	1,076	508

TABLE V
Dimension and runtime of the generated graphs.

Area	# Edges	# Vertices	Runtime [sec]
urban	346^3	228^3	3,600
suburban	263^3	176^3	2,043
rural	139^3	88^3	882
provincial	17^3	14^3	519

B. Quality

The spatial input data selected for the quality experiments originates from the Austrian Digital Cadastral Map and covers a suburban area of about $6km^2$. It represents sparsely as well as densely populated areas to provide the most significant quality evaluation. To allow a self-sufficient comparison of the generated graphs, there is no existing infrastructure given. So the cabled communication network has to be build from scratch. The 42 points objects to be connected (i.e., Access points) are randomly selected centroids of buildings. Figure 6 shows the preprocessed and enhanced NGB of the area.

The generated graphs used in the experiments have been verified to contain all polygon boundaries (no filtering or aggregation of edges) and that there are no isolated parts. The weighting of the edges was done using average excavation costs per meter with respect to the underlying land use.

A fully connected graph (fully mesh all vertices) represents the best possible real-world correlation (because nearly every path is present) but the least quality with respect to the applicability. Considering only the polygon boundary, the number of vertices is about $13,300$ in our example. A fully connected graph would have 88^6 edges and therefore is not applicable.

We choose the best approximation as benchmark reference: a graph containing all polygon boundaries enhanced by edges building a fully connected subgraph of all access objects. Next to this we have generated three more graphs using our proposed approach.

The following list of network graphs were generated to allow an estimation of the quality:

- $Benchmark$ ... Polygon boundaries and fully connected access objects - Figure 7(a).
- $Fix150$ ... Generated graph with additional street crossings each $150m$ - Figure 7(b).
- $Fix5$ ... Generated graph with additional street crossings each $5m$ - Figure 7(c).
- $Adaptive$ ... Generated graph with adaptive crossings. The distance between two crossings is dependent on the construction costs of the underlying land use parcel.

Fig. 6. Spatial data of the project area representing the borders of land use polygons, street polygons (gray filled) and the access objects to be connected (black squares).

(a) $Benchmark$

(b) $Fix150$

(c) $Fix5$

(d) $Adaptive$

Fig. 7. The graphs used in the experiments: (a) benchmark reference, generated graphs with fixed street crossings of (b) $150m$ and (c) $5m$, and (d) the graph with adaptive crossings.

(a) $Benchmark$

(b) $Fix150$

(c) $Fix5$

(d) $Adaptive$

Fig. 8. The graphs superimposed by the optimized minimal Steiner tree running scenario (a).

Interpolated value between $10m$ (most expensive) and $500m$ - Figure 7(d).

Table VI is listing the dimension of the graphs. In case of the graph $Adaptive$, the number of edges and vertices is varying with the applied cost-set in the weighting process.

TABLE VI
DIMENSION OF THE GENERATED GRAPHS.

Graph	# Edges	# Vertices	# Access Objects
$Benchmark$	247^3	130^3	42
$Fix150$	17^3	13^3	42
$Fix5$	63^3	40^3	42
$Adaptive$	85^3 ($\pm 10^3$)	51^3 ($\pm 7^3$)	42

We used the defined graphs to approximate the minimal Steiner tree using the famous minimum spanning tree heuristics described in [21]. It has a worst case time complexity of $O(|T||V|^2)$ and guarantees a optimization result of no more than $2(1 - \frac{1}{l})$ times the result of the optimal Steiner tree (l represents the number of leaves in the optimal tree).

The following three different artificial scenarios (cost-triggered due to different cost-sets used in the weighting) have been applied:

(a) Each type of land use has its own cost value per meter. Edges on the boundary of a parcel and edges crossing a parcel show equal costs per meter.
(b) Avoid crossings of parcels. The costs per meter doubles for crossing edges.
(c) Force minimal usage of edges representing street crossings and avoid crossings of other land use parcels. The costs per meter doubles for crossing edges. The costs for edges crossing the street are multiplied by ten.

Table VII shows the optimized construction costs of a wired telecommunication network build from scratch. The costs of all edges in the Steiner tree are summarized in the third column. The last column is giving the deviation from the generated benchmark graph that is set as reference. In case of scenario (c) the deviation of graph $Fix5$ and $Adaptive$ is even negative. Hence, both the graphs are better suited for the optimization of a wired communication network than the complex benchmark graph.

Figure 8 is visualizing the optimized trees.

As shown in the experimental results, the proposed graph generation approach can be applied to generate weighted network graphs that are usable in mathematical optimization of network construction costs. The construction and activation costs are subject to change and should be updated prior to the edge weighting, to ensure a correct overall cost assignment. The real-world correlation is validated comparing the optimization results with a nearly fully connected graph.

TABLE VII
OPTIMIZED CONSTRUCTION COSTS

Scenario	Graph	Cost [Euro]	Deviation [%]
(a)	$Benchmark$	209,588	0
	$Fix150$	344,333	64.29
	$Fix5$	256,246	22.26
	$Adaptive$	242,198	15.56
(b)	$Benchmark$	306,502	0
	$Fix150$	426,692	39.21
	$Fix5$	346,584	13.08
	$Adaptive$	337,765	10.2
(c)	$Benchmark$	1,196,125	0
	$Fix150$	1,310,881	9.59
	$Fix5$	1,183,349	-1.07
	$Adaptive$	1,159,338	-3.08

VII. CONCLUSION

In this paper, an approach for the generation and weighting of a network graph has been proposed. Introducing a normalization format called NGB to support a fully automated process of generating graphs using a wide range of hybrid spatial data as the input. The process of generating a candidate graph followed by the classification and weighting of the edges produces a valid and computable real-world graph.

The experiments show that the approach is effective and efficient and that the weighted graph can be used as basic input into mathematical optimization algorithms. As a future investigation, we intend to explore ways to improve the adaptivity of the approach to further reduce the dimension of the generated graphs.

REFERENCES

[1] M. Prossegger, "Generation of a weighted network graph based-on hybrid spatial data," in *Proceedings GEOProcessing 2013, The Fifth International Conference on Advanced Geographic Information Systems, Applications, and Services*, 2013, pp. 120–124.

[2] R. D. Hwang, F. and P. Winter, *The Steiner Tree Problem*. North-Holland, 1992.

[3] A. Ivanov and A. Tuzhelin, *Minimal Networks: The Steiner Problem and its Generalizations*. CRC Press, 1994.

[4] O. Tuncer, "Fully automatic road network extraction from satellite images," in *Recent Advances in Space Technologies, 2007. RAST '07. 3rd International Conference on*, June 2007, pp. 708–714.

[5] Y.-Y. Chiang and C. A. Knoblock, "Automatic extraction of road intersection position, connectivity, and orientations from raster maps," in *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '08. New York, NY, USA: ACM, 2008, pp. 22:1–22:10.

[6] P. Bachhiesl, M. Prossegger, H. Stoegner, J. Werner, and G. Paulus, "Cost optimal implementation of fiber optic networks in the access net domain," in *International Conference on Computing, Communications and Control Technologies*, 2004, pp. 334–349.

[7] "Katastralmappe SHP Schnittstellenbeschreibung, Version 2.0.1." BEV - Bundesamt fuer Eich- und Vermessungswesen, 2012.

[8] R. Wang and S. D.M., "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, pp. 5–34, 1996.

[9] L. Leo Pipino, Y. W. Lee, and R. Y. Wang, "Data quality assessment," *Commun. ACM*, vol. 45, no. 4, pp. 211–218, 2002.

[10] B. K. Kahn, D. M. Strong, and R. Y. Wang, "Information quality benchmarks: product and service performance," *Commun. ACM*, vol. 45, no. 4, pp. 184–192, Apr. 2002.

[11] A. Jakobsson and F. Vauglin, "Status of data quality in european national mapping agencies," in *Proceedings of the 20th International Cartographic Conference*, vol. 4, 2001, pp. 2875–2883.

[12] "19113 Geographic Information - Quality Principles," in *ISO/TC 211*. International Organization for Standardization (ISO), 2002.

[13] T. Ubeda and M. J. Egenhofer, "Topological error correcting in gis," in *Proceedings of the 5th International Symposium on Advances in Spatial Databases*, ser. SSD '97. London, UK, UK: Springer-Verlag, 1997, pp. 283–297.

[14] K. A. V. Borges, C. A. Davis, Jr., and A. H. F. Laender, "Database integrity," J. H. Doorn and L. C. Rivero, Eds. Hershey, PA, USA: IGI Publishing, 2002, ch. Integrity constraints in spatial databases, pp. 144–171.

[15] M. Mostafavi, G. Edwards, and R. Jeansoulin, "An ontology-based method for quality assessment of spatial data bases," in *Proceedings fo the Third International Symposium on Spatial Data Quality*, vol. 28, 2004, pp. 49–66.

[16] M. Prossegger and A. Bouchachia, "Incremental identification of topological errors in spatial data," in *The 17th International Conference on Geoinformatics*, Aug. 2009, pp. 1–6.

[17] M. Prossegger and A. Bouchachia, "Incremental semi-automatic correction of misclassified spatial objects," in *Adaptive and Intelligent Systems*, ser. Lecture Notes in Computer Science, A. Bouchachia, Ed. Springer Berlin Heidelberg, 2011, vol. 6943, pp. 16–25.

[18] "ESRI Shapefile Technical Description," in *An ESRI White Paper*, July 1998.

[19] M. Prossegger, "Normalized Geobasisdata (NGB) - technical requirements v.3.0," FHplus Project Netquest, Carinthia University of Applied Sciences, Tech. Rep., October 2012.

[20] A. David, Z. Cui, and A. Cohn, "A spatial logic based on regions and connection," in *Proc. KR-92*, 1992, pp. 165–176.

[21] L. Kou, G. Markowsky, and L. Berman, "A fast algorithm for steiner trees," *Acta Informatica*, vol. 15, no. 2, pp. 141–145, 1981.