

Article

Comparison of Open Source Power Grid Models—Combining a Mathematical, Visual and Electrical Analysis in an Open Source Tool

Wilko Heitkoetter ^{*}, Wided Medjroubi , Thomas Vogt  and Carsten Agert 

DLR Institute of Networked Energy Systems, Carl-von-Ossietzky-Str. 15, 26129 Oldenburg, Germany; Wided.Medjroubi@dlr.de (W.M.); T.Vogt@dlr.de (T.V.); Carsten.Agert@dlr.de (C.A.)

* Correspondence: wilko.heitkoetter@dlr.de

Received: 16 October 2019; Accepted: 7 December 2019; Published: 11 December 2019



Abstract: Power grid models are important in relation to several topics and applications, especially the modelling, optimisation and extension of electrical grids. The significance of grid models is heightened by the increase in renewable energy generation and the challenges associated with its integration into the power grid. However, despite their crucial importance, grid models have generally not been made publicly available for scientific studies or technical analyses. Little information has been published about either the details and methods used in the derivation of these models, or their input and output data. Recently, several projects were initiated in an effort to address this by developing open source grid models and associated data. These projects used different approaches and methods, but most are based on the OpenStreetMap database. The goal of this paper is to compare the different available grid models on the basis of the structure and derivation methods used. Therefore, a novel combination of a graph-theoretical, Geographic Information System (GIS)-based and power-related comparison level is introduced using the open source tool AutoGridComp, which was developed by the authors. The grid models considered in this study are the Scientific Grid Model (SciGRID), GridKit and open street map Transmission Grid Model (osmTGmod) models for Germany.

Keywords: grid models; open source; comparison; characterisation; GIS; network criteria; visualisation; power-related completeness

1. Introduction

The goal of energy system modelling is to model and simulate energy systems using different scenarios, as well as analysing their impact and results. This field gained increasing interest due to the integration of renewable energy sources in the energy supply, motivated by environmental, political and economic factors [1,2]. Energy system models constitute powerful tools, which help policy- and decision-makers to understand the effects and implication of policies on future energy systems. The insights provided by energy models are the subject of many technical papers, studies and reports, for example dealing with the cost of renewable energy [3], the stability of the German transmission network [4] and the Ten-Year Development Plan for the European transmission network [5].

Despite the importance of energy system modelling, the availability and accessibility of energy-related datasets is an issue, as such data is in general not publicly available [6]. Reliable data such as generation capacities, electrical loads estimation and locations of renewable energy plants are not available for scientific studies and research activities [7–9]. This extends to data of power networks, which are the subject of this publication. In particular cases, non-disclosure agreements can be established between Transmission Network Operators (TSOs) or national network agencies (such

as the German Network Agency BNetzA (Bonn, Germany) for research institutions to access power grid models and their corresponding datasets. However, the results obtained using this data are not open for publication, thus transparency, comparability and reproducibility of scientific research is not possible [10]. The issue of data accessibility is not only important for scientific purposes but also for policy- and decision-makers, as many studies are directly commissioned by different ministries and governmental bodies. Initiatives from the German Federal Ministry for Education (BMBF) and the Federal Ministry of Economic Affairs and Energy (BMWi), which finance projects dealing with open source and open data, are an indicator for the growing interest of decision makers in opening energy models and data.

Such an approach is the Scientific Grid Model (SciGRID) project in which an open source grid model SciGRID (For more information about the SciGRID project visit the project webpage: www.scigrd.de) [10,11] was developed. The SciGRID model uses the OpenStreetMap (OSM) database (OpenStreetMap: the free open source editable map of the world. For more information visit www.openstreetmap.org) to extract and abstract the power grid. The osmTransmission Grid Model (osmTGmod) (For more information visit: www.github.com/wupperinst/osmTGmod) model, developed at the Wuppertal Institute for Climate, Environment and Energy in cooperation with Flensburg University of Applied Sciences, also uses OSM as a database, but relies on a broader range of data types. Another example is OpenGridMap platform created by Rivera et al. [12], which extends the data from OpenStreetMap with information collected by the use of an OpenGridMap mobile app (Version 1.0, Technical University of Munich, Munich, Germany), developed by the authors.

Open source grid models have already been used for power flow calculations, yielding realistic results [13–16]. However, those models are not validated yet and differ in their structure. Therefore, it is necessary to first compare them to reference models of grid operators and second to explain the structural differences between the open source models by variations in the derivation processes.

Different approaches have been used to characterise and compare power grid models. Barthelemy [17] investigates the spatial aspects of networks and classifies power grids as planar networks. Matke et al. [11] use a graph-theoretical decomposition to characterise the structure of the German power grid. Clusters are identified that may be used in reduced models for optimisation problems addressing system design and operation. Crucitti et al. [18] conduct a topological analysis of the Italian power grid. Using a model for cascading failures, the authors determine the structural vulnerability of the grid. Cotilla-Sanchez, Hines et al. [19,20] describe the structure of the three North American electric power interconnections, from a topological and electrical connectivity perspective. The topology of the analysed networks is compared to those of random [21], preferential-attachment [22] and small-world [23] graphs by calculating mathematical network criteria distributions. For studying the electrical connectivity, the authors propose a new method that uses electrical distances instead of geographic connections.

In this contribution, we compare the open source power grid models SciGRID, GridKit and osmTGmod and address the following research questions. First, we discuss what the general differences in the model derivation processes and the resulting topologies are. Then, we apply a novel multi-criteria approach and combine a mathematical, visual and power-related comparison. Therefore, we investigate how the distributions of the network criteria differ between the models. Next, the grid models are plotted in an interactive grid map and we examine to what extent the topologies are spatially corresponding. The results of the graph-theoretical analysis are added to the map and we check whether this combination can help explaining specific findings in the network criteria distributions. Finally, the power-related completeness of the grid models is compared.

We will use the term power grid models interchangeably to represent models used to derive grid data and also to define the data outputs of power grid models, unless stated otherwise. Power grid models output represent power grids topologies and are defined by the nodes (or vertices) and edges (or links) of their network representation. For transmission networks, which are the subject of the

present contribution, the power grid nodes represent electrical substations and the power grid edges represent the transmission lines connecting them.

The remaining part of the paper proceeds as follows. Section 2 covers the methodology and introduces the open source tool AutoGridComp (AutoGridComp can be downloaded from the code repository, <https://github.com/wheitkoetter/AutoGridComp>, or the QGeographic Information System (QGIS) plugin repository, <https://plugins.qgis.org/plugins/AutoGridComp/>.) for the automated comparison of power grid models that was developed and applied in this work. Section 3 is dedicated to results of the multi-criteria comparison of the SciGRID, GridKit and osmTGmod grid models. A conclusion is presented in Section 4 and Appendices A and B contain further information on the implementation of the power grid comparison tool.

2. Methods

This contribution aims to compare the open source power grid models SciGRID, GridKit and osmTGmod on a visual, mathematical (or structural or network) and power-related level. The methodology of these three levels of comparison are described in Sections 2.1–2.3. For the automated comparison and the coupling of the different levels of comparison we developed the open source tool AutoGridComp, which is introduced in Section 2.4.

2.1. GIS-Based Visual Comparison

The straightforward approach to characterising power grids is to visually inspect their topologies. Visual inspection provides a first qualitative picture of the grid topology by showing for example, how nodes are connected or where nodes and edges are clustered. Geo-referenced power grid models can be visualised by geographic information systems (GIS). For this purpose, we used the open source application QGIS (QGIS A Free and Open Source Geographic Information System: www.qgis.org). For the comparison of two power grid models, the topologies were plotted on top of each other. Nodes and edges were identified in the resulting map that are contained in both power grid models and those that are only contained in one of the regarded models.

2.2. Mathematical Network Criteria

A more elaborated and quantitative strategy to analyse and compare power grids is to consider their graph structure. One approach for such an analysis is to compute the values and distributions of network criteria. Such measures allow for a better understanding of power grid network structures, their properties and how it can affect their operation [20,24–26]. In this work, the network criteria degree, betweenness centrality and clustering coefficient [27] were determined for the considered power grid models. In the following paragraphs, definitions for a graph in general, as well as the applied network criteria, are given.

We define a graph as $G = (V, L)$, with the set V of vertices v_i and the set L of links $l_{ij} = \{v_i, v_j\}$ [27]. Multiple links between vertices as well as self-edges are not considered in this publication. The first applied criterion, the degree d of a vertex v_i is defined as the number of links l_{ij} that connect it to adjacent vertices $v_i \neq v_j$. The second criterion, the betweenness centrality, is defined for a link l_{ij} as follows:

$$bc_{ij} = \sum_{v_s \neq v_i, v_j \neq v_t} \frac{\sigma_{st}(l)}{\sigma_{st}}, \quad (1)$$

where σ_{st} is the total number of shortest paths from vertex v_s to vertex v_t and $\sigma_{st}(l)$ is the total number of shortest paths from vertex v_s to vertex v_t passing by the link l_{ij} [11,28]. For a better comparison of

different networks/grids, the betweenness centrality is normalised by the total number of all shortest paths present in the graph/network of interest as follows:

$$bc_{ij,normalised} = \frac{bc_{ij}}{n(n-1)/2}, \quad (2)$$

where n is the number of vertices in the graph G . The results presented below refer to the normalised form of the betweenness centrality.

The third criterion, the local clustering coefficient of a vertex $v_i \in V$, is defined as:

$$c_i = \frac{2 | \{l_{mn} : v_m \in A_i \wedge v_n \in A_i\} |}{n_i(n_i - 1)}, \quad (3)$$

where $A_i = \{v_j : l_{ij} = \{v_i, v_j\} \in L\}$ and $n_i = |A_i| \equiv$ the number of adjacent vertices. The local clustering coefficient can be interpreted as the number of pairs of adjacent vertices that are connected, divided by the total number of pairs of adjacent vertices.

The described network criteria were determined for all nodes, respectively edges of the considered power grid models. Next, the frequency distribution and the cumulative frequency distribution of the network criteria results were determined. A cumulative distribution function (CDF) of a random variable X , evaluated at x , is the probability that X will correspond to a value less than or equal to x [29].

2.3. Power-Related Completeness

As the power grid models were generated from the crowd sourced database OpenStreetMap, the completeness of power-related information may be limited. We determined the share of nodes and edges of the considered power grid models that are equipped with the information voltage level, number of cables, number of wires and frequency. For more information on the electrical properties of open source power grid models, refer to [10].

2.4. Implementation and Coupling of Comparison Levels

The three comparison levels mentioned above are implemented and coupled in the developed AutoGridComp tool so that different properties of power networks can be investigated and related to each other. In this contribution we present the coupling of the mathematical and visual comparison. After the execution of the mathematical analysis of the grid models, each node and edge is associated with values for the considered network criteria. These results are then added to the visual map representation of the grid topology in QGIS. As shown in Section 3.3, this allows for a sound interpretation of specific findings in the distributions of the network criteria and how to relate them to the power grid's topology [30,31]. The developed AutoGridComp tool can be used as a standalone python tool or a plugin for QGIS. To determine the mathematical network criteria, functions from the Python package NetworkX (NetworkX package webpage: <https://networkx.github.io/>) have been adapted and used in AutoGridComp. For the visual representation QGIS functions were used, e.g., QgsVectorLayer() for building geo-referenced vector layers of the power grid. More information on the implementation of the open source comparison tool are provided in Appendix A. A flowchart of the comparison process is presented in Appendix B.

3. Results

This section is dedicated to the results of the qualitative and quantitative comparison of the open source grid models SciGRID, GridKit and osmTGmod. First, we give an overview of general similarities and differences in the derivation processes of the three grid models in Section 3.1. This allows for an understanding of the structural differences in the resulting grid topologies of these models. Second, in Section 3.2, the topologies are compared from a visual perspective. In Section 3.3, a graph theoretical comparison is conducted and the results are

added to the map representation of the topologies. Finally, Section 3.4 deals with the power related completeness of the grid models. For more detailed descriptions of the models, derivation processes and the power-related data available in OSM, please refer to [10] and the SciGRID model webpage (SciGRID model and project webpage: <http://scigrid.de/>), as well as the GridKit (GridKit repository: <https://github.com/bdw/GridKit>) and osmTGmod (osmTGmod repository: <https://github.com/wupperinst/osmTGmod>) github repositories.

3.1. General Differences in the Grid Model Derivation Approaches and Resulting Topologies

An overview of important features and simplifications used in deriving the three models is given in Table 1.

Table 1. Overview of the derivation processes used in the Scientific Grid Model (SciGRID), GridKit and open street map Transmission Grid Model (osmTGmod) grid models.

Criterion	SciGRID	GridKit	osmTGmod
data source	OpenStreetMap	OpenStreetMap	OpenStreetMap
OSM data type used	relations	nodes, ways	nodes, ways, relations
heuristics	no	yes	yes
nb. of nodes *	468	1849	811
nb. of edges *	774	2166	1054
cross-borders included	yes	no	no
Edge/links connections	relation information	implicitly derived	both (relation information, implicitly derived)
complexity of topology	low	high	medium

* topology of the German extra high voltage grid (220 and 380 kV) (grid models generated from OSM data downloaded on 15 January 2017).

Note that the most striking difference between the grid models is the number of nodes and edges in the resulting topologies. This is mainly due to the utilised OSM data types in the different approaches and the methods for creating the edges of the network, as well as the level of abstraction considered by the models. In OSM, power data are represented by the OSM types nodes, ways and relations [10]. Line-carrying towers and electrical poles are represented by nodes. Ways represent transmission lines, as well as the outlines of substations. Electrical circuits, which comprise, e.g., multiple substations and transmission lines, are represented by the OSM type relation.

The SciGRID grid model is only based on the OSM power relations and does not apply heuristics. This results in a significantly lower number of nodes and edges, as well as a lower complexity than the GridKit and the osmTGmod models. The GridKit model does not make use of power relations but instead derives the edges of the network from the power-tagged OSM nodes and ways using heuristics. This leads to the creation of so-called auxiliary nodes and auxiliary edges by the spatial algorithms used in the model. The osmTGmod approach primarily uses the available power relations to abstract the electrical grid and adds missing information from available power tagged nodes and ways. As a result, the number of nodes and edges and the complexity of the topology is in between the one of the SciGRID and GridKit topologies. In the open_eGo (open_eGo repository: <https://github.com/openego>) project, the osmTGmod approach was enhanced by the coverage of the high voltage level (110 kV). When both the high and extra high voltage levels of the German transmission grid are considered, the number of nodes in the osmTGmod grid topology amounts to more than 11,000 [15] nodes.

In summary, the use of SciGRID can be recommended when a deterministic approach is preferred and no heuristics are desired. If a high completeness of the topology is required, the GridKit and osmTGmod models are advantageous. The quality of OSM data may vary between different countries or regions [32–34]. One needs to be careful when deriving power grids for geographical regions where

OSM power relation data are not available in sufficient amounts. In this case, the GridKit approach is well-suited because it only relies on OSM nodes and ways data types.

In the next section, it will be shown, how the obtained topologies differ from a graph and a visual perspective due to the differences in the derivation processes and the assumptions used in each model. Since the SciGRID and GridKit approaches differ most, only these two models will be analysed.

3.2. GIS-Based Visual Comparison

In this section we investigate to what extent the SciGRID and GridKit topologies correspond to each other, using the GIS-based representation of the models. Figure 1 shows the topologies of the two models for the German extra high voltage grid.

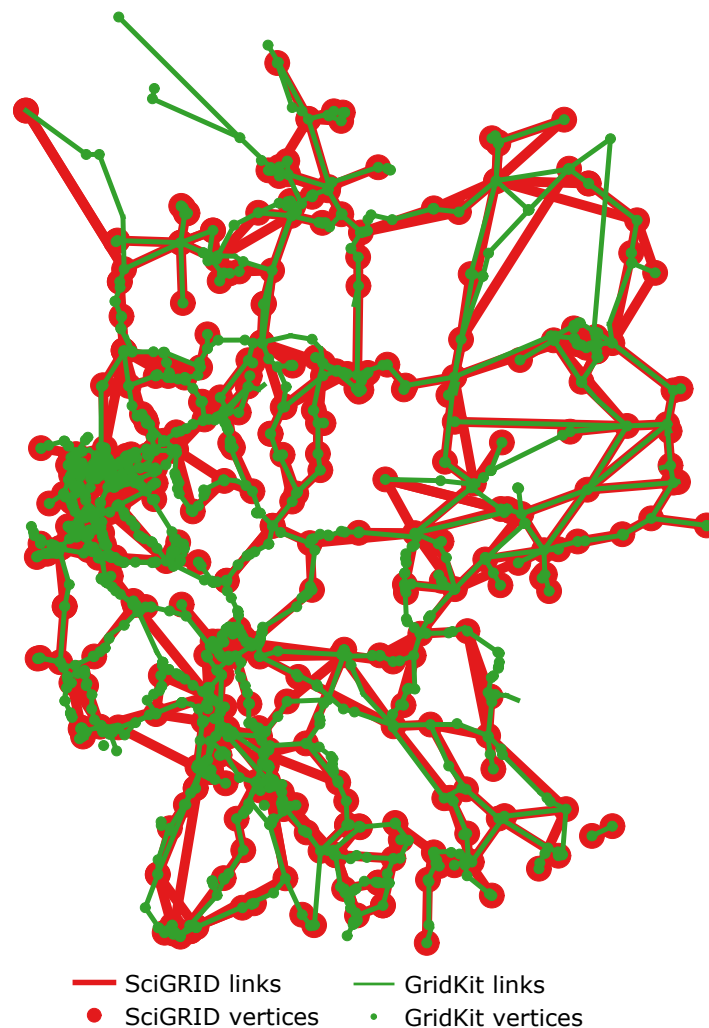


Figure 1. Visual comparison of the SciGRID and GridKit models for the German extra high voltage grid.

In general, the SciGRID and GridKit models show a high accordance. However, the routes of the links are more detailed in the GridKit model due to the introduction of auxiliary vertices and edges. Due to the higher degree of abstraction, the routes are less detailed in the SciGRID model. There are several links that are only contained in the GridKit model, but not in the SciGRID model. This can be explained by the fact that SciGRID only relies on relation information, which is less complete.

3.3. Coupled Mathematical and Visual Comparison

In this section, the results of the graph-theoretical comparison of the SciGRID and GridKit models of the German extra high voltage grid are presented. First, the distributions of the network criteria degree, betweenness centrality and the clustering coefficient are described. Then, the network criteria results are added to the GIS-based representation of the topology to explain specific findings in the distributions.

3.3.1. Degree

The degree frequency distribution of the vertices in the SciGRID and GridKit model of the Germany transmission grid are shown in Figure 2a.

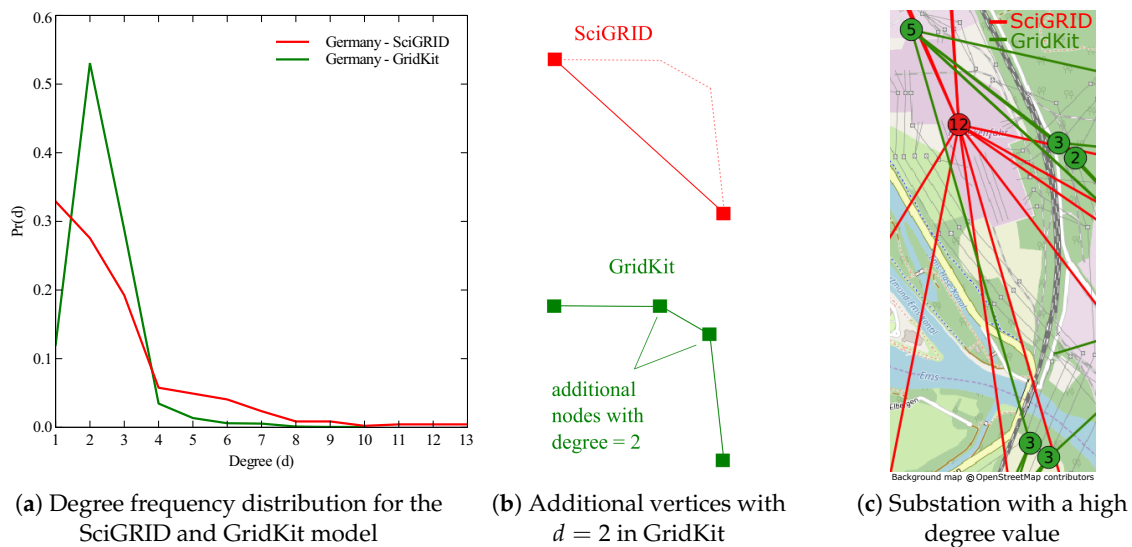


Figure 2. Degree for the SciGRID and GridKit models.

The frequency peak for the GridKit model is at a degree value of $d = 2$, while for the SciGRID model the peak is at $d = 1$. This difference is caused by the creation of the auxiliary vertices in the GridKit model, which have a degree of two, as depicted in Figure 2b. The maximum degree value in the GridKit model is 10, whereas there are also vertices with a degree value of $d = 11$, $d = 12$ and $d = 13$ in the SciGRID model. This difference can be explained by the higher degree of abstraction of the SciGRID model and is demonstrated exemplarily in Figure 2c for the substation at the nuclear power station in the Emsland region in Germany (the degree values are indicated by the numbers in the circles). While the substation is abstracted by one vertex with $d = 12$ in the SciGRID model, it is represented by multiple vertices with lower degree values in the GridKit model.

3.3.2. Edge Betweenness Centrality

Figure 3a shows the cumulative edge betweenness centrality (bc) distribution for both the SciGRID and GridKit transmission grids for Germany. It can be seen that for $bc < 10^{-3}$ and $bc > 10^{-2}$ the distributions are similar for both models. In the range $10^{-3} < bc < 10^{-2}$, the CDF of the edge betweenness centrality are different in both models. This can be seen as jumps that occur at dissimilar betweenness centrality values for each model. One such jump is marked with a black arrow in Figure 3a.

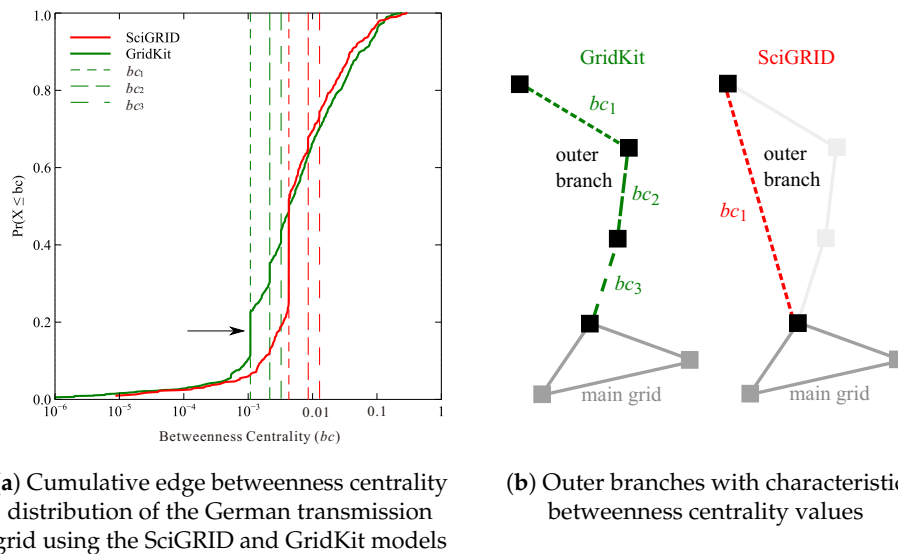


Figure 3. Edge betweenness centrality for the SciGRID and GridKit models.

To identify the reason for these jumps, the interactive map of the power grid is used. Therein, the links of the grid which have betweenness centrality values corresponding to the jumps are highlighted (see Figure 4). On the interactive map, it can be seen that all the links with betweenness centrality values corresponding to the jumps (in yellow colour) are part of grid branches situated at the edge of the power grid, hereinafter referred to as outer branches. An example for such a branch is schematically depicted in Figure 3b for a better understanding. In the GridKit model this outer branch consists of three lines. However, in the SciGRID model and due to the abstraction process, the outer branch is composed of only one line. Using this information, the authors derived an equation (Equation (4)) to calculate a normalised betweenness centrality value where the total number of vertices n in the grid, as well as an outer edge index labelled i , are used. The relationship between the outer edge index (i) and the betweenness centrality at the jumps is explained for the GridKit and the SciGRID models, as follows:

$$bc_i = \frac{i \times n - i^2}{n(n-1)/2}, \quad i = 1, 2, 3, \dots \quad (4)$$

For the outermost link, i is set to 1 in Equation (4). The outer link is part of all shortest paths going from the outermost vertex (labelled outer vertex) to all the remaining vertices in the grid. In Equation (4), the numerator contains $-i^2$ because self-connection for edges is not considered here. As described above, the denominator contains the total number of shortest paths in the network. The resulting betweenness centrality value matches exactly to the first jump in the curve of the distribution (see Figure 3a) for GridKit.

The betweenness centrality value for the link, marked with bc_2 (the before-last link) in Figure 3b, is calculated using Equation (4) and this time setting $i = 2$. The resulting $2n$ in the numerator reflects the fact that the two outermost vertices (one them part of edge bc_2) are both connected to all other vertices of the grid by the shortest path. This means that a value of four 2×2 (for the two edges) needs to be subtracted in the numerator, again because self-connecting edges are not considered. In this way, the betweenness centrality value for the second jump in the distribution for GridKit is derived. The value for the third jump in the distribution is calculated using Equation (4) and setting $i = 3$.

It can be noted that the first jump in the distribution for the GridKit model is larger than the second and third jumps. The reason for this is that there exist outer branches containing only one link. This is also the case for most of the outer branches in the SciGRID model. Due to the stronger degree of abstraction in SciGRID, nearly no outer branch consists of more than one link. For that reason, there

is a much larger first jump in the betweenness centrality distribution, compared to GridKit and the jumps which follow are significantly smaller.

Furthermore, in Figure 3a there are links having an even lower betweenness centrality value (blue colour), than the outermost edges. In the GridKit model, those edges account for 11% of all links and in the SciGRID model they represent 25% of all edges. These findings somehow contradict the expectation that the outer branches have the lowest betweenness centrality values.

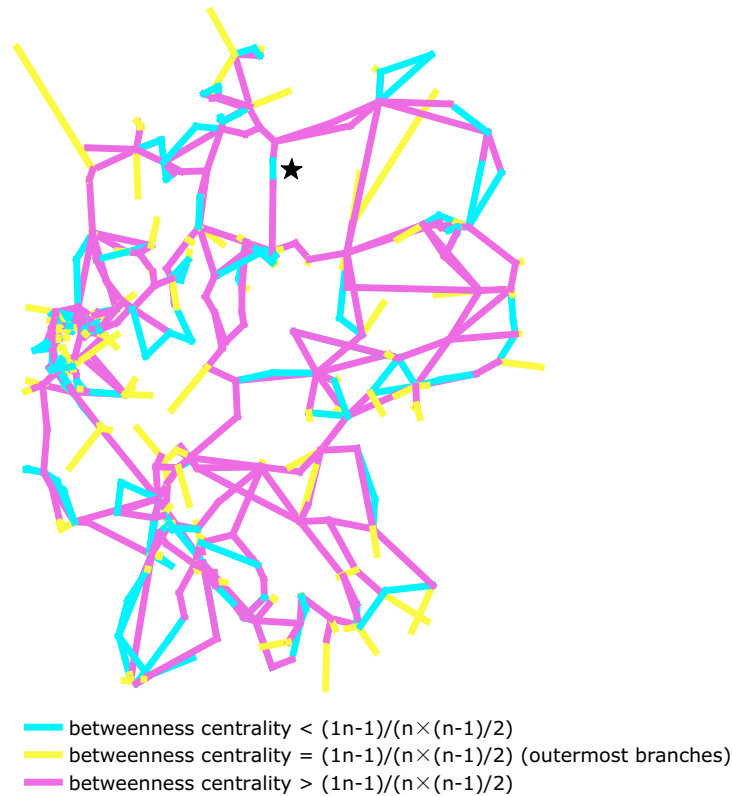


Figure 4. Edge betweenness centrality values in the SciGRID model for the German extra high voltage network; Star symbol indicates an edge with a low betweenness centrality value, despite a central position in the grid.

In the following paragraph, a possible explanation for this finding will be given. In Figure 4, the geo-referenced topology of the German power grid is shown obtained using the SciGRID model. We distinguish between the outer branches and links with a higher or lower betweenness centrality. Most of the links having a lower betweenness centrality than the outer branches are located at the edge of the power grid. However, they are connected to the grid on both ends and do thus differ from outer branches as defined in Figure 3b. However, on close inspection, some of the links with the lowest betweenness centrality have a central position in the power grid. An example link is marked with a star symbol in Figure 4. The low betweenness centrality value of this edge can be explained as follows. Firstly, there exist parallel links in the west (left) and east (right) of the considered link, which connect the vertices in the northern part of grid with those in the centre and southern part of the grid. Secondly, the shortest paths, which connect the vertices at the ends of the considered link with the rest of the network, may also head into northern and southern direction, not passing by the considered link which explains the low value of the betweenness centrality of this edge.

3.3.3. Clustering Coefficient

The distribution of the clustering coefficients for the vertices of SciGRID and GridKit grids for Germany is shown in Figure 5a. It can be seen that most of the vertices in both power grids have a clustering coefficient of zero. Those vertices can be divided into three groups. The first group is formed by the outermost vertices (belonging to the outer branches), as depicted in Figures 3b and 5b. These vertices have only one neighbour and thus no pair of adjacent neighbours. The second group consists of the vertices that are on an edge that does not branch and thus have only one pair of neighbours, which is however not connected. The vertices of the third group have multiple pairs of neighbours, but also none of these neighbours is connected.

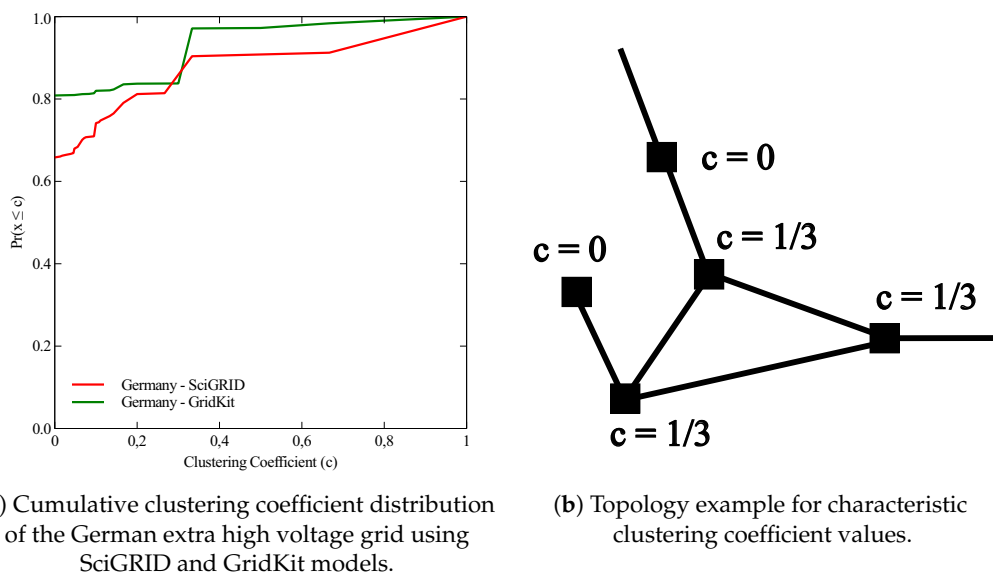


Figure 5. Clustering Coefficient for the SciGRID and GridKit models.

Furthermore, Figure 5a shows that the share of vertices with a clustering coefficient of zero is 15% larger in the GridKit dataset than in the SciGRID dataset. This is due to the lower degree of abstraction of GridKit and the presence of the auxiliary nodes which have a clustering coefficient of zero (belonging to the second group as defined above). Note that in both models, a high share of vertices has a clustering coefficient of 1/3. This is mainly caused by vertices forming triangles inside the network as shown in Figure 5b. The vertices, which are connected to the triangle and one other branch, have a clustering coefficient of 1/3, because they have three pairs of neighbours, of which one is connected. In the GridKit data the share of vertices with a clustering coefficient of 1/3 is about 5 percentage points higher than in the SciGRID grid, because additional triangles are introduced at the intersections of the power lines.

3.4. Power Related Completeness

This subsection deals with the completeness of information on the electrical properties of the grid models. The SciGRID, GridKit and osmTGmod models of the German extra high voltage grid (grid models generated from OSM data downloaded on 15 January 2017) were compared by applying the developed AutoGridComp tool. For more details on the modelling of the electrical properties of the transmission grid using open data, refer to [10].

Figure 6a shows the power related completeness for the grid vertices, which represent, e.g., the electrical substations. The osmTGmod model has the highest completeness with 100% for both, the voltage information and the frequency information. The completeness for the SciGRID model is 92% for the voltage and 71% for the frequency information. This is due to the fact that in SciGRID

missing information from OSM is not added. GridKit has a completeness of 94% for the voltage and 65% for the frequency information.

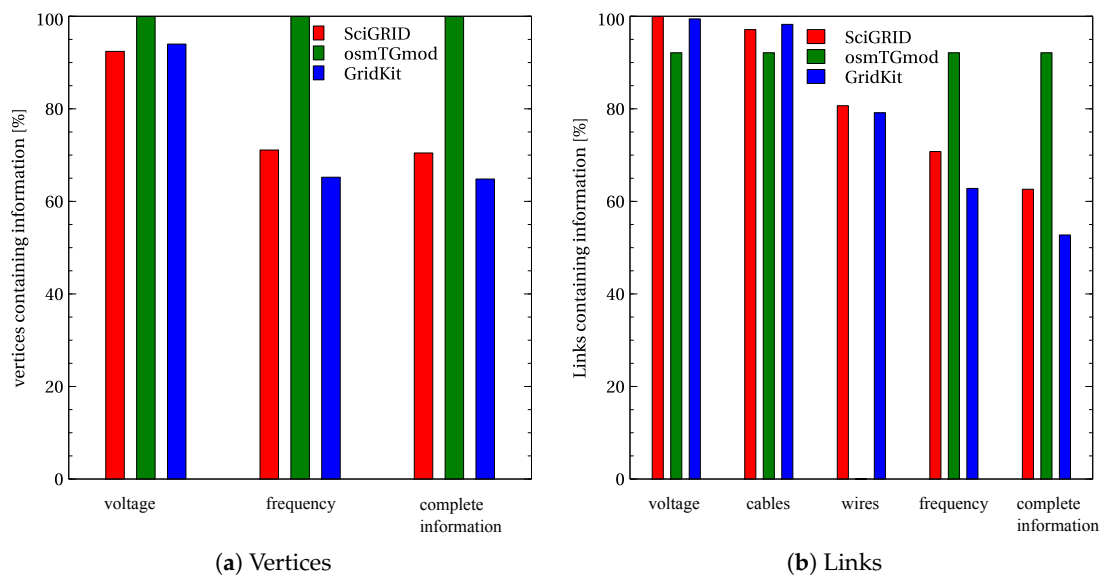


Figure 6. Completeness of the electrical information of the SciGRID, osmTGmod and GridKit model for the German extra high voltage grid.

The results for the links are shown in Figure 6b. For both the information on the voltage and the number of cables per transmission line, SciGRID and GridKit have a higher completeness than osmTGmod. Considering the frequency information, the osmTGmod model is most complete. The completeness of the information on the number of wires per transmission cable is approx. 80% for SciGRID and GridKit. For osmTGmod, no value is given in Figure 6b, because the information on the wires was not directly obtained from OSM. Instead, in osmTGmod it is assumed that all 220 kV transmission lines have two wires per cable and all 380 kV lines have four wires per cable [35]. Links having overall complete information account for 92% in osmTGmod, 62% in SciGRID and 52% in GridKit.

The higher power-related completeness of the osmTGmod and GridKit model in comparison to the SciGRID model can be explained as follows. The osmTGmod and GridKit approaches apply heuristics to complete the missing information, while SciGRID does not use heuristics.

4. Conclusions and Outlook

The qualitative comparison of the open source power grid modelling approaches SciGRID, GridKit and osmTGmod reported in this paper showed that they mainly differ in the OpenStreetMap data types used, as well as the applied heuristics. This leads to different numbers of nodes and edges in the resulting topologies. A novel approach was presented that combines a mathematical, geo-referenced and power-related comparison of power grid models. The results of the graph-theoretical analysis were highlighted in an interactive grid map. For the open source grid models considered in this study, the combined approach allows for sound interpretations of distinctive features, e.g., jumps in network criteria distributions.

These findings can help distinguish between features in the network distributions that are caused by the model derivation process and actual properties of the power grid. Such a primary stage analysis may be followed by a more advanced characterisation of the power grids, e.g., by searching for small-world properties [23] or community detection [11]. The results reported here also offer recommendations for improving grid model derivation methods. The GridKit approach creates auxiliary nodes and edges in order to derive the network model. As shown above, this has a strong influence on the network criteria distributions. We recommend that the algorithm be changed in

order to avoid this, while maintaining the high degree of information completeness integrated from OpenStreetMap.

The comparison tool, AutoGridComp, which was developed for this study, is provided under an open source license. Thus, it can be used by other researchers in endeavours such as comparing open source power grid models with the official grid models of grid operators. Such a validation is required for assessing the positional accuracy [36] and completeness [37] of the OpenStreetMap data.

Author Contributions: Conceptualisation, W.H. and W.M.; Methodology, W.H.; Software, W.H.; Supervision, W.M.; Writing—original draft, W.H. and W.M.; Writing—review and editing, W.H., W.M., T.V. and C.A.

Funding: This work builds upon previous research that was conducted as part of the SciGRID project. The SciGRID project was supported by the German Federal Ministry of Education and Research (BMBF) under the funding code 03SF0471, within the funding initiative Zukunftsfähige Stromnetze.

Acknowledgments: The first author gratefully acknowledges the financial support provided by the Foundation of German Business (sdw) through a PhD scholarship.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AutoGridComp	Automatic Grid Model Comparison
BMBF	German Federal Ministry for Education
BMWi	German Federal Ministry of Economic Affairs and Energy
CDF	Cumulative Density Function
csv	comma separated values
DLR	German Aerospace Centre
GIS	Geographic Information System
nb.	number
OSM	OpenStreetMap
osmTGmod	OSM Transmission Grid Model
QGIS	Quantum GIS
SciGRID	Scientific Grid Model
TSO	Transmission System Operator

Appendix A. Implementation

The python tool AutoGridComp is implemented as a plugin for the open source geographic information system QGIS, to allow for a geo-referenced and interactive visualisation of the grid models. QGIS is written in C++, but the majority of its classes are python-enabled. This is made possible through the use of SIP, the extension module generator that outputs C++ code [38]. To integrate the AutoGridComp plugin into QGIS, template files were generated using the QGIS Plugin Builder and these files were then further enhanced. For ease of use, a GUI was created using Qt Designer. As shown in Figure A1, the information entered in the GUI is handled via the dialog files `ui_autogridcomp_dialog_base.ui`, `ui_autogridcomp_dialog_base.py` and `autogridcomp_dialog.py`.

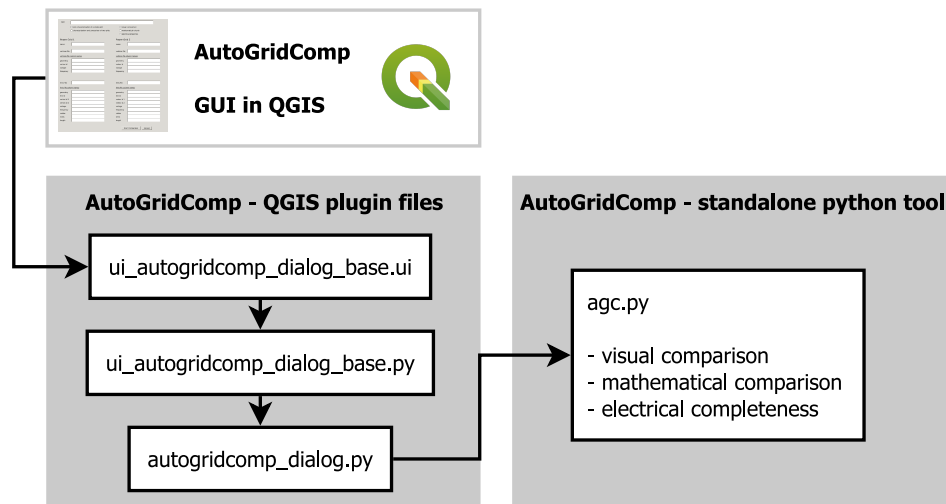


Figure A1. Implementation of the AutoGridComp tool.

Next, the information is passed to the file `agc.py`, which contains the main plugin functionality. The file `agc.py` can also be run as a standalone python tool, however the input information must be directly entered into the python file in that case. The AutoGridComp tool can be downloaded at the QGIS plugin repository, <https://plugins.qgis.org/plugins/AutoGridComp/>, or the code repository: <https://github.com/wheitkoetter/AutoGridComp>.

Appendix B. Comparison Process

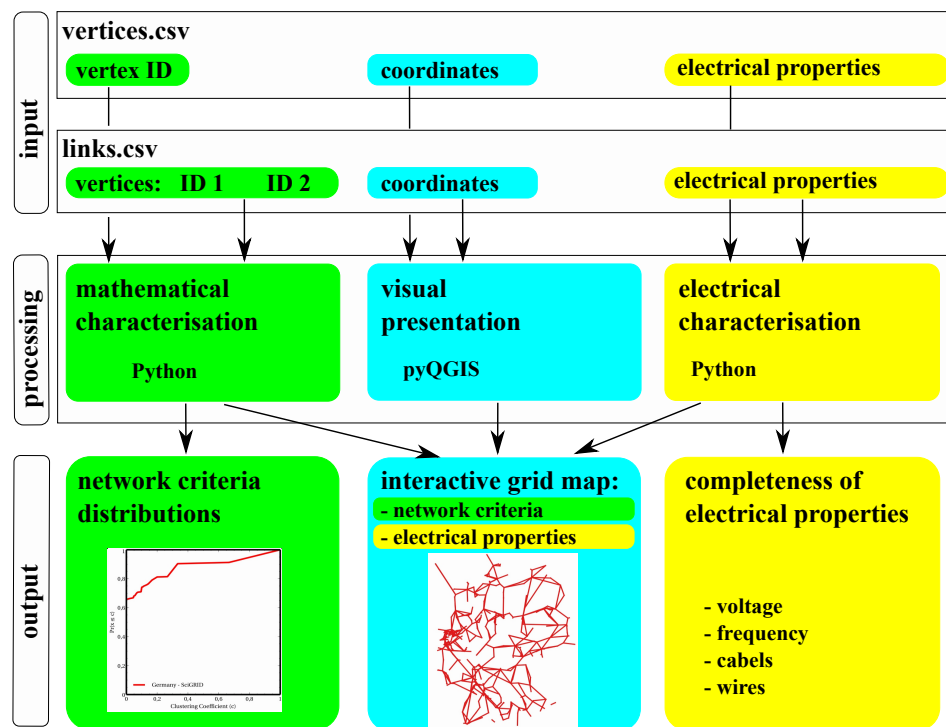


Figure A2. Flowchart of the AutoGridComp tool: three levels of comparison are possible: mathematical (in green), visual (in blue) and electrical (in yellow). Each level is shown with the input data needed, the different processing tasks it involves and its output.

References

1. Bazmi, A.A.; Zahedi, G. Sustainable energy systems: Role of optimization modeling techniques in power generation and supply—A review. *Renew. Sustain. Energy Rev.* **2011**, *15*, 3480–3500. [[CrossRef](#)]
2. Grubb, M.; Köhler, J.; Anderson, D. Induced technical change in energy and environmental modeling: Analytic approaches and policy implications. *Ann. Rev. Energy Environ.* **2002**, *27*, 271–308. [[CrossRef](#)]
3. Agora Energiewende. *The Cost of Renewable Energy: A Critical Assessment of the Impact Assessments Underlying the Clean Energy for All Europeans-Package*; Discussion Paper; Agora Energiewende: Berlin, Germany, 2017.
4. Schnettler, A.; Fuchs, B.; Roehder, A.; Mittelstaedt, M.; Massmann, J.; Natemeyer, H. *Studie zu Aspekten Der Elektrischen Systemstabilität Im Deutschen Übertragungsnetz Bis 2023*; Institut für Hochspannungstechnik: Aachen, Germany, 2015. (In German)
5. ENTSO-E. *Ten-Year Network Development Plan 2016*; ENTSO-E: Brussels, Belgium, 2016.
6. Pfenninger, S.; DeCarolus, J.; Hirth, L.; Quoilin, S.; Staffell, I. The importance of open data and software: Is energy research lagging behind? *Energy Policy* **2017**, *101*, 211–215. [[CrossRef](#)]
7. Wiese, F.; Bökenkamp, G.; Wingenbach, C.; Hohmeyer, O. An Open Source Energy System Simulation Model as an Instrument for Public Participation in the Development of Strategies for a Sustainable Future. *Wiley Interdiscip. Rev. Energy Environ.* **2014**, *3*, 490–504. [[CrossRef](#)]
8. Egerer, J.; Gerbaulet, C.; Ihlenburg, R.; Kunz, F.; Reinhard, B.; von Hirschhausen, C.; Weber, A.; Weibezahn, J. *Electricity Sector Data for Policy-Relevant Modeling: Data Documentation and Applications to the German and European Electricity Markets*; Data Documentation 72; German Institute for Economic Research: Berlin, Germany, 2014.
9. Zhou, Q. Cross-Border Congestion Management in the Electricity Market. Ph.D. Thesis, Durham University, Durham, UK, 2003. Available online: <http://etheses.dur.ac.uk/1263> (accessed on 10 September 2019).
10. Medjroubi, W.; Müller, U.P.; Scharf, M.; Matke, C.; Kleinhans, D. Open Data in Power Grid Modelling: New Approaches Towards Transparent Grid Models. *Energy Rep.* **2017**, *3*, 14–21. [[CrossRef](#)]
11. Matke, C.; Medjroubi, W.; Kleinhans, D.; Sager, S. Structure Analysis of the German Transmission Network Using the Open Source Model SciGRID. In *Advances in Energy System Optimization, Proceedings of the First International Symposium on Energy System Optimization (ISESO 2015), Heidelberg, Germany, 9–10 November 2015*; Springer: Berlin, Germany, 2017; pp. 177–188.
12. Rivera, J.; Leimhofer, J.; Jacobsen, H.A. OpenGridMap: Towards automatic power grid simulation model generation from crowdsourced data. *Comput. Sci. Res. Dev.* **2016**. [[CrossRef](#)]
13. Müller, U.P.; Cusmann, I.; Wingenbach, C.; Wendiggensen, J. AC Power Flow Simulations within an Open Data Model of a High Voltage Grid. In *Advances and New Trends in Environmental Informatics*; Springer: Berlin, Germany, 2017; pp. 181–193.
14. Wienholt, L.; Müller, U.; Bartels, J. Optimal Sizing and Spatial Allocation of Storage Units in a High-Resolution Power System Model. *Energies* **2018**, *11*, 3365. [[CrossRef](#)]
15. Müller, U.P.; Schachler, B.; Scharf, M.; Bunke, W.D.; Günther, S.; Bartels, J.; Pleßmann, G. Integrated techno-economic power system planning of transmission and distribution grids. *Energies* **2019**, *12*, 2091. [[CrossRef](#)]
16. Rohden, M.; Jung, D.; Tamrakar, S.; Kettemann, S. Cascading failures in ac electricity grids. *Phys. Rev. E* **2016**, *94*, 032209. [[CrossRef](#)] [[PubMed](#)]
17. Barthélemy, M. *Spatial Networks*; Springer: Berlin, Germany, 2014.
18. Crucitti, P.; Latora, V.; Marchiori, M. A topological analysis of the Italian electric power grid. *Phys. A Stat. Mech. Appl.* **2004**, *338*, 92–97. [[CrossRef](#)]
19. Cotilla-Sanchez, E.; Hines, P.D.; Barrows, C.; Blumsack, S. Comparing the topological and electrical structure of the North American electric power infrastructure. *IEEE Syst. J.* **2012**, *6*, 616–626. [[CrossRef](#)]
20. Hines, P.; Blumsack, S.; Sanchez, E.C.; Barrows, C. The topological and electrical structure of power grids. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences (HICSS)*, Honolulu, HI, USA, 5–8 January 2010; pp. 1–10.
21. Erdős, P.; Rényi, A. On random graphs I. *Publ. Math. (Debr.)* **1959**, *6*, 290–297.
22. Barabási, A.L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [[CrossRef](#)] [[PubMed](#)]
23. Watts, D.J.; Strogatz, S.H. Collective dynamics of ‘small-world’ networks. *Nature* **1998**, *393*, 440–442. [[CrossRef](#)] [[PubMed](#)]

24. LaRocca, S.; Johansson, J.; Hassel, H.; Guikema, S. Topological performance measures as surrogates for physical flow models for risk and vulnerability analysis for electric power systems. *Risk Anal.* **2015**, *35*, 608–623. [[CrossRef](#)] [[PubMed](#)]
25. Hines, P.; Cotilla-Sanchez, E.; Blumsack, S. Topological models and critical slowing down: Two approaches to power system blackout risk analysis. In Proceedings of the 2011 44th Hawaii International Conference on System Sciences (HICSS), Kauai, HI, USA, 4–7 January 2011; pp. 1–10.
26. Wang, Z.; Scaglione, A.; Thomas, R.J. Electrical centrality measures for electric power grid vulnerability analysis. In Proceedings of the 2010 49th IEEE Conference on Decision and Control (CDC), Atlanta, GA, USA, 15–17 December 2010; pp. 5792–5797.
27. Newman, M. *Networks: An Introduction*; Oxford University Press Inc.: New York, NY, USA, 2010; pp. 1–2.
28. Lu, L.; Zhang, M. Edge Betweenness Centrality. In *Encyclopedia of Systems Biology*; Springer: New York, NY, USA, 2013; pp. 647–648.
29. Gentle, J.E. *Computational Statistics*; Springer: Berlin, Germany, 2009; Volume 308.
30. Cuffe, P.; Keane, A. Novel quality metrics for power system diagrams. In Proceedings of the 2016 IEEE International Energy Conference (ENERGYCON), Leuven, Belgium, 4–8 April 2016; pp. 1–5.
31. Cuffe, P.; Keane, A. Visualizing the electrical structure of power systems. *IEEE Syst. J.* **2017**, *11*, 1810–1821. [[CrossRef](#)]
32. Haklay, M. How Good is OpenStreetMap Information? A Comparative Study of OpenStreetMap and Ordinance Survey datasets for London and the rest of England. *Environ. Plan. B* **2010**, *37*, 682–703. [[CrossRef](#)]
33. Arsanjani, J.; Mooney, P.; Zipf, A.; Schauss, A. Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets. In *OpenStreetMap in GIScience*; Springer: Berlin, Germany, 2015; pp. 37–58.
34. Çakmak, H.K.; Maass, H.; Bach, F.; Kühnapfel, U.; Hagenmeyer, V. Ein Ansatz zur automatisierten Erstellung umfangreicher und komplexer Simulationsmodelle für elektrische Übertragungsnetze aus OpenStreetMap-Daten. *AT-Automatisierungstechnik* **2015**, *63*, 911–925. [[CrossRef](#)]
35. Scharf, M. *Entwicklung Eines Modells des Deutschen Übertragungsnetzes auf Basis der Offenen Geodatenbank OpenStreetMap*; Flensburg University of Applied Sciences: Flensburg, Germany, 2015.
36. Zielstra, D.; Zipf, A. Quantitative studies on the data quality of OpenStreetMap in Germany. In Proceedings of the Sixth International Conference on Geographic Information Science (GIScience 2010), Zurich, Switzerland, 14–17 September 2010; pp. 20–26.
37. van Oort, P.A. *Spatial Data Quality: From Description to Application*; Wageningen Universiteit: Wageningen, The Netherlands, 2006.
38. Sherman, G. *The PyQGIS Programmer's Guide: Extending QGIS 2.x with Python*; Locate Press: Chugiak, AK, USA, 2014.

