A Crowdsourcing Approach for the Inference of Distribution Grids

Pezhman Nasirifard
Technical University of Munich
p.nasirifard@tum.de

Jose Rivera
Technical University of Munich
j.rivera@tum.de

Qunjie Zhou
Technical University of Munich
qunjie.zhou@tum.de

Klaus Bernd Schreiber
Technical University of Munich
klaus.schreiber@tum.de

Hans-Arno Jacobsen
Technical University of Munich

ABSTRACT

Maintaining a complete and up-to-date model of the distribution grid is a challenging task, and the scarcity of open models represents a significant bottleneck for researchers in this area. In this work, we address these challenges by introducing a crowdsourcing framework for the collection of open data on distribution grid devices and an algorithm to infer the topological model of the distribution grids. We use the crowd and smartphones to collect an image and the geographical position of power distribution grid devices. Since power distribution lines are usually underground and cannot be mapped, we use spatial data analytics on the collected data in combination with other open data sources to infer the topology of the distribution grid. This paper describes and evaluates our crowdsourcing and inference approach. To evaluate our approach, we organized and conducted a crowdsourcing campaign to map and infer a sizeable district in Munich, Germany. The results are compared with the ground truth of the distribution system operator. Our field experiments show that using the crowd to recognize power distribution elements, a precision of up to 82% and a recall of up to 65% can be obtained. The numerical evaluation of our inference algorithm demonstrates that the model we inferred based on the acquired official DSO grid dataset achieves a power length accuracy of 88% compared to the ground truth. These results confirm our approach as a practical method to infer real power distribution grid models.

CCS CONCEPTS

• **Hardware** → **Energy distribution**; • **Information systems** → **Crowdsourcing**;

KEYWORDS

Power grids, Power distribution, Distribution grid inference, Geographic information systems, Crowdsourcing

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
e-Energy '18, June 12–15, 2018, Karlsruhe, Germany
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5767-8/18/06…$15.00
https://doi.org/10.1145/3208903.3208927

1 INTRODUCTION

Over the past few years, the majority of the international community has grown increasingly committed to the reduction of greenhouse gases, especially CO2 emissions [5]. Since the electric power industry is responsible for producing a significant portion of the CO2 emissions [2], researchers and practitioners have proposed several approaches to address these issues by facilitating further integration of renewable resources and introducing new electrical devices such as electric vehicles, and local storage units [29]. However, before implementing the proposed solutions, the practicality and stability of the approaches should be comprehensively evaluated based on the actual distribution grid models. Nevertheless, the majority of studies are based on standardized test feeders, such as the IEEE test feeders [12], the PNNL feeders [14] and CIGRE test feeders [33], which are considerably simplified models and fail to reflect the complexity, geographic features, and limitations of real individual power grids. Although some distribution system operators (DSOs) maintain digitilized models of their grids, the operators do not publicly publish the grid models due to security and legal reasons. Additionally, in several cases, the grid data, especially for distributions grids, is either incomplete or outdated. Also, periodically collecting and updating the grid data is time-consuming, intrusive and a significant financial burden for operators [6].

In this paper, we introduce a non-intrusive crowdsourcing framework for collecting and inferring distribution grid models. The crowdsourcing approach considerably reduces the cost, effort, and time required for gathering grid data by distributing the data collection tasks among the crowd. Furthermore, to improve the quality of the collected grid data, we merge the crowdsourced grid data with the extracted distribution grid elements from free and publicly available OpenStreetMap (OSM) data [24]. Finally, we propose an approach for inferring a distribution grid topological model for a particular region based on the position of grid devices and the spatial features of the area.

To capture, analyze and model the medium and low voltage grid models the power industry has been utilizing geographic information systems (GIS) for a long time. However, no utility manages to maintain an exhaustive and most up-to-date model of their grids [16]. The novelty of our work lies in implementing a complete crowdsource framework for gradually and consistently collecting valid and verified grid data which we use for inferring accurate grid models. In this work, we also explain how to conduct a crowdsourcing campaign based on our framework. We use the result of the crowdsourcing campaign to evaluate the performance of the participants and confirm the practicality of crowdsourced grid data in comparison to official grid data provided by the distribution grid operator.

We recognize the following contributions in this paper:

(1) We introduce a crowdsourcing framework for collecting, verifying, storing and publishing power grid elements.
(2) We design and conduct a crowdsourcing campaign with several participants, where the evaluation of the acquired data verifies crowdsourcing as a practical approach to collecting and maintaining the grid data.
(3) We propose an inference approach for generating distribution grid models of an area based on the geographical position of power grid elements, consumer endpoints, and the pathway’s structure of the area.

We structure the rest of the paper as follows: In Section 2, we review the previous works on crowdsourcing and inference approaches of the distribution grids which we build on. In Section 3, we introduce the platform we built for executing the crowdsourcing event. Then, we discuss our crowdsourcing framework and describe the insights and results derived from the conducted crowdsourcing campaign in Section 4, where we also explain the surveyed feedback of the participants and evaluate the quality of the crowdsourced grid data. In Section 5, we describe and evaluate the inference approach for generating distribution grid models followed by Section 6, where we discuss the results and limitations of our work. In Section 7, we provide the concluding remarks and describe the future works.

2 RELATED WORK

Crowdsourcing is a time-efficient and cost-effective method for collecting detailed and highly accurate geographical data, including electrical grid elements [9, 26]. However, there has not been a coherent crowdsourcing approach for collecting electrical grid devices. A potential resource of crowdsourced power-related data is OSM, which uses a community approach to locate and map physical structures of an area. As of January 2018, OSM contains about 16 million components marked with power-related tags all around the globe [25]. However, the majority of the power-related OSM data are transmission level elements. Medium and low voltage grid devices are scarce in OSM. Furthermore, the OSM community often tags the power-related components with wrong values due to the lack of expert knowledge, or just errors, e.g., some transformers are marked as cable cabinets. An approach for crowdsourcing grid data and integrating the collected data with OSM data is present in [30]. We extend the previous works by improving the crowdsourcing approach through designing a robust crowdsourcing framework as a supplement to OSM grid data to take advantage of all available datasets and also to record grid elements for inferring distribution models. Moreover, we provide an evaluation of the approach against a ground truth, which was missing in the literature.

The studies on inferring topological models of distribution grids are somewhat limited. Although several studies propose approaches for inferring topological models of transmission grids based on complex network theories [1, 4, 10, 31], these approaches are not applicable to distribution grids. The main reasons are that the distribution grid components and structures are inherently different, the number of devices in distribution grids is more extensive than transmission grids, and the structures tend to be more complicated than in transmission grids. These challenges make distribution networks much more difficult to map and to build accurate models. Furthermore, in contrast to transmission grid elements, in many countries a significant portion of distribution level grid elements are underground, e.g., in Germany, 73 percent of the medium voltage and 87 percent of low voltage cables are buried [37]. As a result, locating the accurate geographical position of grid components and their characteristics can be very challenging and require several assumptions and background knowledge. Nevertheless, some studies propose intrusive methods for inferring distribution grid topologies based on the interaction among grid devices. In [11], the authors propose an approach for decentralized inference of distribution grids based on the communication among a set of autonomous intelligent agents, on an overlay network. In [3], they estimate the grid topology by applying correlation analysis on the voltage amplitude measurements of grid endpoints. However, none of these approaches consider geographical characteristics, and they require detailed information of grid and interaction with grid elements, which we can not apply to crowdsourced grid data. One similar area to our inference challenge is planning optimal and cost-efficient distribution grid systems based on the location of expected consumers, where applying genetic algorithm and graph theory are accepted methods [13, 34]. Therefore, in this work, we use spatial analysis and graph theory for inferring distribution models when the location of the grid devices, consumers, and structure of the region is known. We differentiate ourselves from previous works in the combination of crowdsourcing and inference approaches to infer real power distribution grids. Also, our method is non-intrusive, i.e., we only require device locations and no grid measurements. Finally, in contrast to other works, we conduct field experiments and compare our results to the ground truth of the distribution grid operator.

3 OPENGRIDMAP CROWDSOURCING PLATFORM

Our crowdsourcing campaign heavily makes use of the OpenGridMap (OGM) project [23]. The OGM project offers a platform for collecting, organizing, and openly publishing a broad range of transmission and distribution grid data and models. Also, OGM provides the researchers and practitioners with a crowdsourcing platform for collecting high, medium and low voltage level grid devices. The OGM extracts and combines the power-related grid data of OSM with the verified submissions of volunteers. As an example, electrical utility crews can use the OGM platform to record the continuously changing electrical grid to maintain the most recent information of the grid.

The OGM crowdsourcing platform consists of two primary components, including a smartphone application and a web application. Figure 1 displays the OGM Android application available free of charge on the Google PlayStore [21]. The participants of the crowdsourcing activity are required to download and install this application on their phone. Afterward, the participants follow a simple procedure for submitting the grid element upon identifying the element in their surroundings. First, the participants should

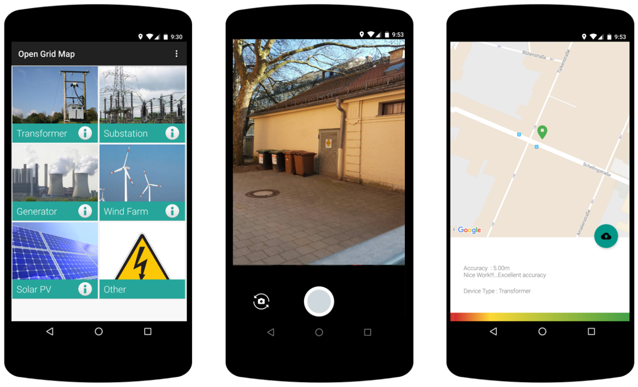


Figure 1: OpenGridMap Android application.

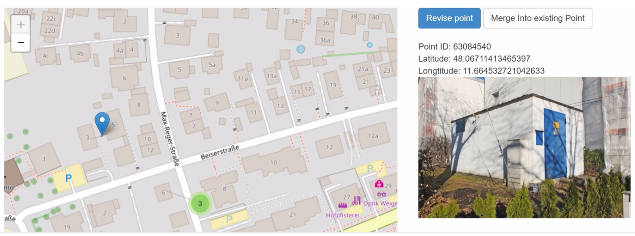


Figure 2: OpenGridMap web application.

select the type of the discovered grid device. Then, the participants take a picture of the grid device with the application and review the location of the device on the map obtained from the location service of the smartphone. Since the location service may not be accurate, the participants can edit the location of the grid device manually. Finally, participants submit the recorded grid element to the OGM servers either immediately or when the smartphone has access to a WiFi connection.

On the OGM web application, the expert in the loop reviews and accepts the submitted grid elements. The expert has the option of examining the grid element, correcting the assigned metadata such as the type of device and merge the device with existing devices to avoid duplicate entries. For example, Figure 2 illustrates a submission of a transformer. After the review, the expert accepts and publishes the grid element to the OGM platform.

4 CROWDSOURCING DISTRIBUTION GRID DATA

In the following section, we describe the designed and developed crowdsourcing framework for collecting distribution grid devices. We also performed a crowdsourcing campaign in the Freimann district of Munich, Germany. A comparison to the ground truth reveals that crowdsourcing is a practical method for mapping distribution grid devices.

4.1 Crowdsourcing Framework

Crowdsourcing makes an appealing option for collecting distribution grid data because the grid devices are widely distributed, and their positions are previously unknown [15]. Nevertheless, we require a consistent framework to clarify and divide up the data collection activity into smaller precise and realizable tasks which several participants can accomplish independently and in parallel. We base our crowdsourcing framework on the collective intelligence framework developed by Malone et al. at the MIT's Center for Collective Intelligence [15]. The Malone's framework consists of four elements, also known as "genes," which are required for recognizing the building blocks of collective intelligence. The four elements are described as four fundamental questions of "Who, Why, What and How," which we utilize to structure our crowdsourcing method.

The detailed description of the Malone's framework is out of the scope of this paper. According to the Malone's framework, we identify the following requirements and attributes of an organized crowdsourcing campaign for collecting distribution grid devices:

- We require the crowdsourcing movement to be relevant anywhere on the planet, independent of the geographical location and participants' previous knowledge and training. Therefore, as a medium for collecting the grid data, we utilize a custom designed smartphone application with a simplified data collection procedure. Due to the high prevalence of smartphones, the application can be used by any participants without any previous training required.
- Since any practical crowdsourcing activity profoundly depends on the number of participants, providing an appropriate incentive for the participants is crucial. Therefore, we present the participants with a detailed description of the project and its objectives, emphasizing the project's potential for reducing the greenhouse gases and integration of renewable energy resources, presumably increasing the intrinsic joy of the participants for engaging in such a community.
- The distribution grid devices have a wide range of designs and types. However, we require collecting only specific devices such as transformers and cable cabinets. Therefore, we provide the participants with a protocol, explicitly defining the type of necessary grid elements and also the tasks which they need to fulfill. Because we presume the participants do not have any previous training and they may catch wrong grid devices, we verify the validity of the collected grid elements after the crowdsourcing campaign with the help of an expert in the loop.
- To break down the crowdsourcing activity into recognizable smaller tasks, we divide the data collection region into smaller subareas which we assign each subarea to a group of participants. However, to prevent any potential duplicate recordings, the expert in the loop monitors the elements based on their position and removes the duplicated entries.

Accordingly, any crowdsourcing event that we manage follows a precise procedure. First, before beginning the crowdsourcing event, we hold a preliminary meeting with all participants to describe the objectives of the event and build a group of two persons. Each group receives a package, containing the group protocol, a stamped letter describing the intentions of the crowdsourcing campaign that they present to the police or other security personals in case of any inquiry. We also provide an agenda that use visual examples to explain the conventional design, signs, and characteristics of the wanted grid devices specific to the area.

In each group, one person has the role of navigator and the other data collector. The group navigator has the responsibility of filling the group's protocol and navigating the group through the area during the crowdsourcing event. The collector follows the navigator and carefully monitors the area for the requested power devices, and upon identifying a new grid device, the collector snapshots and submits the element by using the OGM smartphone application. The group protocol, completed by the navigator, contains a printed map of the area assigned to the group. The navigator marks the streets and paths which are covered by the group. Furthermore, in case of failing to inspect some parts of the area, due to time limitation or lack of accessibility, the navigator marks the missing parts on the protocol's map and documents the reason. To prevent losing any collected grid data due to any unexpected OGM platform failures, the navigator keeps a list of discovered transformers on the protocol's map. Also, the navigator keeps a record of the number of identified transformers and cable cabinets in the protocol.

To increase the safety of the participants, we introduce a few obligatory rules which all participants are required to follow. We prohibit participants from trespassing any military, industrial or private properties. Only power devices which are observable from the streets and public areas should be recorded. Furthermore, we provide the participants with phone numbers of the event's organizers, which they can reach in case of emergency.

4.2 Crowdsourcing Campaign in Munich Freimann

To measure the quality of our framework, we organized and conducted a crowdsourcing campaign in the German city of Munich's Freimann district at 9th of May 2017 [20]. We selected the Freimann district because we received the official distribution grid data from *Stadtwerke München* (SWM) [17], the Munich city utilities and DSO, which we use as the base for our evaluation.

Initially, we divided up the Freimann district into several subareas with approximately equal areas as Figure 3 displays. However, since we recruited only 22 participants for the crowdsourcing event, we covered eleven areas with two persons assigned to each area, and we gave each group 90 minutes to perform the mapping in the designated area. Although we did not record the exact distance each group traversed, we intended to cover 4 kilometers of routes on average by each group. Furthermore, since participants installed the application on their phones, their participation did not enforce any initial cost on us. Although the OGM crowdsourcing platform is capable of storing any distribution grid devices, for the sake of simplicity, we asked the participants to collect only transformers and cable cabinets in their area. In the agenda, we provided detailed information about the transformers and cables cabinets used by SWM, the Munich's DSO. We should mention that after the end of crowdsourcing event, we offered an incentive to the best performing group.

After the event, we used a questionnaire to survey the overall crowdsourcing experience of all participants. We include the complete result of the survey in the appendix and briefly explain the most interesting responses. In general, the results are in the

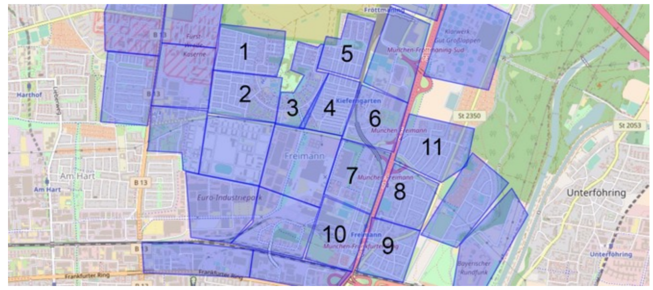


Figure 3: Freimann district data collection areas.

affirmative upper third in all indicators, confirming that the participants are satisfied with the organization and execution of the OGM crowdsourcing platform and our framework. Furthermore, for the significant majority of the participants, the community spirit is a more valuable incentive than monetary prizes, when deciding to join the event, indicating that our community-spirit-oriented incentives were attractive to the participants. Although 75% of the participants expressed willingness to participate in such a crowdsourcing event again, we must remark that recruiting a large number of participants is challenging. We recruited 70% percent of the participants from the course instructed by one of the event organizers, and the rest of the participants were informed by their friends, and no participant discovered the event from the public Facebook event we had created two months before the event. Therefore, we suggest investing enough time and publicity for gathering the participants before any crowdsourcing event. Finally, the survey reveals that identifying distribution grid devices is a challenging task because the grid devices are well-hidden. However, since we only conducted the crowdsourcing event in Munich, we can not argue that this difficulty extends to other urban or rural areas on the planet. Nevertheless, the effort and time required for mapping an area depends on the complexity and density of the area and located grid devices more than the size of the area.

4.3 Evaluation of the Crowdsourced Grid Data

After the crowdsourcing event, our experts in the loop use the OGM web application to verify the submitted devices. The grid elements which are not classified correctly, such as telecommunication cabinets that are incorrectly marked as cable cabinets, are removed from the dataset. The results show that on average 75% of the collected devices are correctly identified by the participants, and Area 8 has the best result with 96% accuracy because Group 8 strictly adhered to the agenda and captured only devices with detectable SWM signs. Figure 4 summarizes the number of correctly verified devices, including the cable cabinets and transformers combined, and also the number of rejected devices from all eleven areas.

After the verification, we use the OGM platform to merge the grid data collected during the crowdsourcing event with the existing OGM grid data. As mentioned in the previous section, the OGM grid data consists of extracted power-related OSM data combined with a few potential submissions by other volunteers since the beginning of the project. During merging, we removed the duplicate submissions of the same grid devices that have identical type and

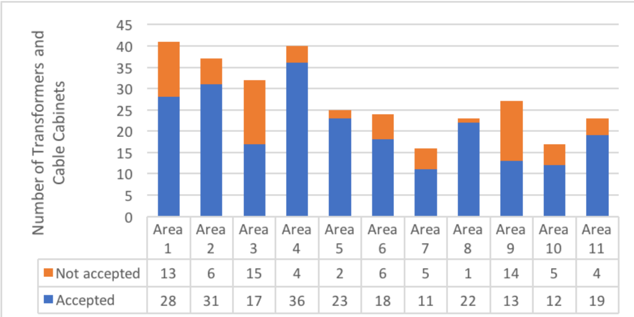


Figure 4: The number of accepted and rejected grid devices in each area.

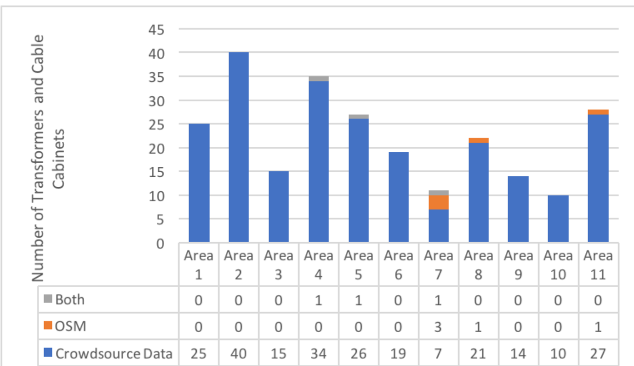


Figure 5: OGM grid data separated based on their origin.

position. We carry on the rest of the evaluation by using the merged grid data which we refer to as OGM grid data. The reason for using the merged data is that one of the contributions of the crowdsourcing approach is in being a supplement to the extracted OSM grid data and it is more crucial to evaluate the quality of aggregated publicly available grid data. However, at the time of conducting the crowdsourcing event, hardly any of the OGM transformers or cable cabinets originate from OSM or any volunteers, and the crowdsourcing event creates the significant majority of the grid elements shown in Figure 5.

To evaluate the accuracy and validity of the collected grid data, we compare the OGM grid data from Freimann subareas one to eleven with the official DSO grid data in the corresponding subareas, that we acquired from SWM. First, we examine the DSO grid elements and discover that for a few transformers there are multiple identical entries which are overlapping on the map. Therefore, we merge the overlapping DSO transformers into one before evaluating the OGM grid elements. Then, we compare the number of OGM grid elements with the number of DSO grid elements in each area without any constraints on the distance, meaning that we do not enforce any maximum distance threshold between the exact geographical position of the OGM grid element with its corresponding DSO grid device. On average, the OGM grid data contains 60% of DSO grid elements, and Area 4 reports the best accuracy of 82%, as Figure 6 summarizes and compares the number of extracted OGM

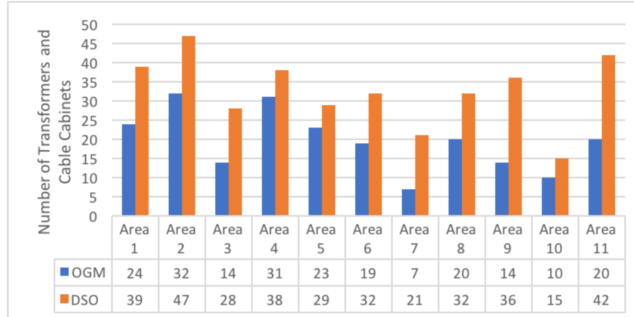


Figure 6: The number of transformers and cable cabinets from OGM and DSO datasets.

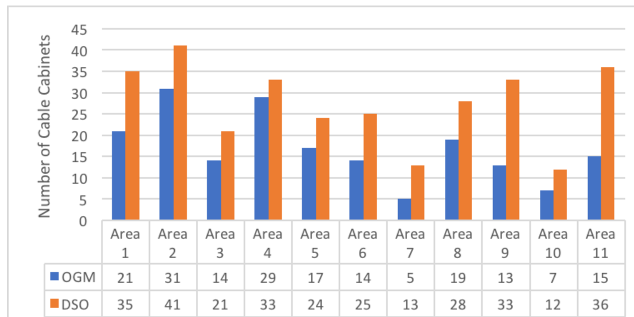


Figure 7: The number of cable cabinets from OGM and DSO datasets.

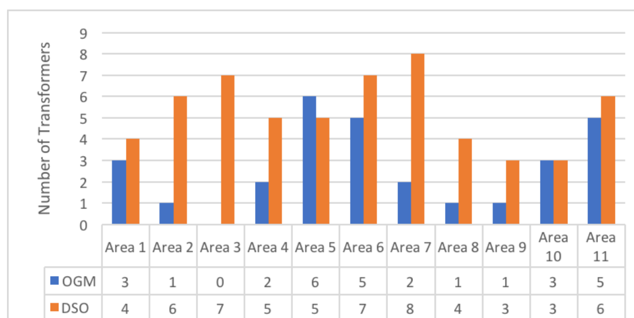


Figure 8: The number of transformers from OGM and DSO datasets.

and DSO transformers and cable cabinets. Furthermore, in more detail, our comparison reports a 61% coverage of DSO cable cabinets with Area 4 offering the best precision of 88%. The comparison of OGM transformers to DSO transformers show a 50% coverage, with Area 5 and 10 having 100% accuracy. Figure 7 and Figure 8 displays the comparison between OGM and DSO grid data in the number of cable cabinets and transformers, respectively.

For the second phase of evaluation, we define an eight meters maximum distance threshold between the OGM grid element and corresponding DSO element. As an example, Figure 9 displays the OGM and DSO grid elements on the map, where we only count

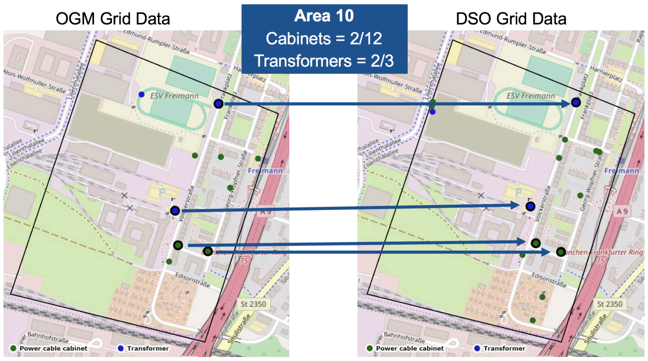


Figure 9: OGM and DSO grid devices in the Freimann subarea 10.

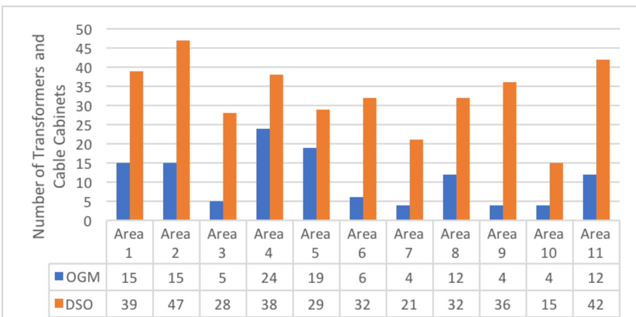


Figure 10: The number of transformers and cable cabinets from OGM and DSO datasets based on an eight meters distance threshold.

OGM elements as valid, if there exists an identical DSO element with the same type, within an eight meters proximity of the OGM element. With this constraint, we observe that on average 33% of DSO elements are covered by the OGM dataset, where Area 5 has the highest coverage of 66%, as Figure 10 compares the number of OGM and DSO transformers and cables cabinets in each area. In more detail, the results report a 35% coverage of DSO cable cabinets with Area 4 showing the highest 70% coverage. The results also show a 26% coverage of DSO transformers, where Area 10 reports the best accuracy of 67%. Figure 11 and Figure 12 illustrate the number of recognized cable cabinets and transformers, respectively.

Given the fact that we mapped a large area of Freimann district in 90 minutes and with the participation of 22 people and collected 230 valid transformers and cable cabinets, we argue that crowd-sourcing is an effective approach for fast and cost-efficient data collection. Although none of the participants had expert knowledge, our experts in the loop verified the correctness of 75% of the collected devices and the high precision (the number of true positives over the number of true positives and false positives) and recall (the number of true positives over the number of true positives and false negatives) rate of Area 5 (the group with the best result) [36], as Table 1 displays, confirm the practicality of crowd intelligence. However, the participants training and motivation is necessary for consistent performance. Nevertheless, the limited precision of the

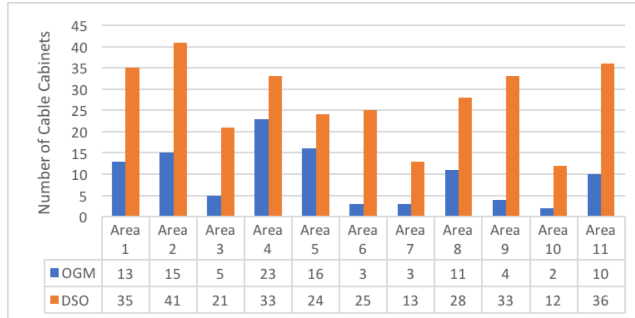


Figure 11: The number of cable cabinets from OGM and DSO datasets based on an eight meters distance threshold.

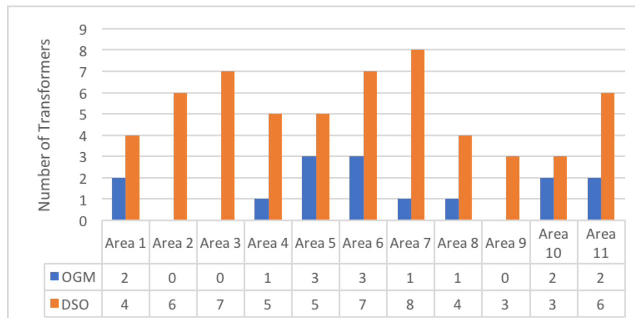


Figure 12: The number of transformers from OGM and DSO datasets based on an eight meters distance threshold.

Table 1: Precision and Recall of Area 5

Area 5	Precision	Recall
Transformers + cable cabinets	82.61%	65.52%
Transformers	94.12%	66.67%
Cable cabinets	50%	60%

location services of the smartphones and the difficulty of detecting and locating well-hidden grid devices are the primary source of inaccuracy, as the varying accuracy of different areas and groups shows.

5 INFERENCE OF THE DISTRIBUTION GRID

In the following section, we introduce our approach for inferring distribution grid models based on the geographical position of grid devices and consumers' endpoints. Furthermore, we discuss the ground truth which we derive from the official DSO grid data, that we received from SWM, and finally, we evaluate the accuracy of two inferred models compared to the ground truth model.

5.1 Distribution Network Inference based on Grid Data

We propose an approach for inferring distribution grid models of an area based on the grid data of the area. In other words, we

Figure 13: The extraction of the start and end points and intersection points of roads.

infer a model of the low voltage underground power cables located within a specific area based on the position of transformers, cable cabinets and consumer endpoints such as buildings and also the structure of pathways in the area. We base our inference method on the primary assumption that the majority of underground cables are installed along the roads and pathways. Therefore, we take advantage of the complete and freely available road information of the area from OSM. Inference Algorithm 1 describes the method we develop for heuristically inferring a Minimum Spanning Tree (MST) as the distribution grid model. We consider the provided position of grid devices and consumers as target nodes taking into account the structure of the roads. Each road consists of several nodes representing a line, which we simplify to enforce the inference of MST along the roads. Therefore, we select the start and end points of the roads and the intersection point of each road pair. To avoid including duplicate points in our dataset, we use a set ensuring the existence of only one copy of any point. As an example, Figure 13 display the roads as blue lines that we give as an input to the inference algorithm, but only the filtered red points are used for the generation of MST.

Afterward, we determine the projection of the target nodes on the nearest road. As an example, Figure 14 displays target nodes in blue dots, their projection as green nodes and the edge between target and projection nodes as red lines. We merge the projected nodes and filtered road nodes to create a base graph containing edges between every pair of nodes in the union of the two sets where we consider the distance between nodes in meters as the weight of the edge. Then, we use the base graph to infer an MST, and finally, add the edges between target nodes and projected nodes to the inferred MST and return the tree as the distribution grid model. We use projected nodes instead of target nodes for creating the base graph because otherwise, we could not create a clean MST along the roads.

The inference engine is implemented in Python with the help of several packages, including but not limited to, NumPy [19], Shapely [32], and NetworkX [18] for creating and manipulating complex networks. We store our grid data on a PostgreSQL [8] database with PostGIS [27] extension enabled. For visualizing and inspecting the data on the area's map, which we import from OSM, we also use the open source geographical information system QGIS [35].

Algorithm 1: Distribution Grid Inference Approach

```
InferDistributionGrid (TargetNodes, Roads)
    input : TargetNodes, the set of transformers, cabinets
            and buildings.
    input : Roads, the line geometry of roads.
    output: GridModel, A minimum spanning tree
            representing the distribution grid.
    filteredRoadNodes = set() ;
    foreach road_i ∈ Roads do
        foreach road_j ∈ Roads do
            filteredRoadNodes.add(road_i[0], road_j[0]) ;
             // Roads start points
            filteredRoadNodes.add(road_i[−1], road_j[−1]) ;
             // Roads end points
            if roadsAreIntersecting(road_i, road_j) then
                intersectionPoint =
                  findIntersectionPoint(road_i, road_j) ;
                filteredRoadNodes.add(intersectionPoint) ;
    projectedNodeSet =
      projectTargetNodesOnRoad(TargetNodes, Roads) ;
    mergedNodeSet =
      filteredRoadNodes ∪ projectedNodeSet ;
    baseGraph = Graph() ;
    baseGraph.addNodes(mergedNodeSet) ;
    foreach node_i ∈ mergedNodeSet do
        foreach node_j ∈ mergedNodeSet do
            if edge(node_i, node_j) not in baseGraph AND i! = j
             then
                edgeWeight =
                  getDistanceInMeters(node_i, node_j) ;
                baseGraph.addEdge(node_i, node_j, edgeWeight)
    gridModel =
      generateMinimumSpanningTree(baseGraph) ;
    projectedEdges =
      makeProjectedNodeTargetNodeEdge(TargetNodes) ;
    gridModel = gridModel ∪ projectedEdges ;
    return gridModel
```

Furthermore, our source code is open-source and freely available [22].

5.2 Ground Truth Model of Freimann District

To evaluate the quality of the inference algorithm, first, we generate the ground truth of the distribution model of Freimann based on the official DSO grid data. We need to construct the compatible ground truth models because the acquired DSO grid information is in shapefile format (shape format, shape index format, and attribute format) [7] which are not compatible with our inferred grid models. Furthermore, the DSO data require cleaning due to some data inconsistency, data duplication, and errors. After importing the DSO grid data of the Freimann district into our database,

Figure 14: Projection of the target node on the nearest roads with an edge between them.

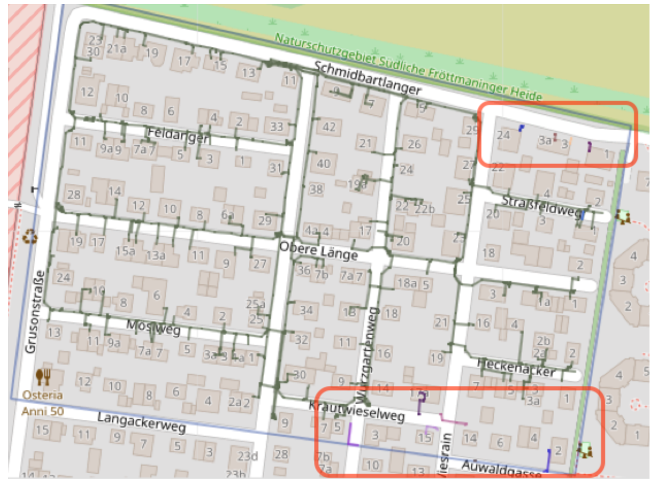


Figure 15: Freimann subarea ground truth subgraphs.

including the location of transformers, cable cabinets, consumer connections and location of underground cables, we perform a few steps of data cleaning. For some transformers, cable cabinets and consumer nodes, there exist duplicate copies of nodes which are either overlapping or located within a few meters of each other. We merge these nodes by defining a maximum distance of one to five meters between them. The reason for using a varying range as distance threshold is that we inspect the data visually to find the best threshold based on the type of the node. Afterward, we inspect the transformers, cable cabinets and consumer endpoints to detect the ones which are disconnected from the grid due to the absence of a connection to any neighboring cables. We connect these isolated nodes by connecting the nodes to the closest cable located within a node's five meters proximity, and if such a cable does not exist, the node is removed from the dataset. In the end, we review any remaining separated cable which is not connected to any node or another cable at any endpoints and remove the isolated cables from the dataset.

After cleaning and preprocessing, we use the cleaned data to create a graph representing the distribution grid of the area. However, we only use the cables from the DSO dataset, since all cables are presumably connected to either endpoints or other cables. We build the base ground truth graph by iterating through each DSO cable line and retrieving the geographical representation of the line from the database. Then, we convert the line's data into a set of nodes with an edge connecting the points which are in the row behind each other. Afterward, we select the largest connected subgraph as the model representing the distribution model of the area. We follow this approach for creating the ground truth model because the cable line data is stored according to EPSG:31468 Projection [28] which requires conversion into the fundamental longitude and latitude we use. As an example, Figure 15 displays the ground truth graph of Freimann subarea with the largest connected subgraph is identified with green edges, and the smaller marked disconnected subgraphs are discarded. Figure 16 shows the created ground truth model of Freiman subareas one to five with a total cable length of 46484 meters.

5.3 Evaluation of the Inference Approach

To evaluate the accuracy of the inference algorithm, we infer two separate models based on the DSO and OGM transformers and cable cabinets data of Freimann and compare the models with the generated ground truth model. Since our inference algorithm also requires the information of roads and consumer endpoints of the area, we integrate the grid data with the extracted related OSM data, including the structure of the roads and the position of residential and commercial buildings. The OGM grid data that we use for generating models are the verified grid devices which we evaluated in the Subsection 4.3. Figure 17 and Figure 18 illustrate the inferred models of the DSO and OGM grid data, respectively. We generate the ground truth model, and the DSO inferred model, both based on the acquired DSO data of the area. However, the difference is that the ground truth is created based on the structure of previously known underground cables, whereas the inferred DSO grid model generates the structure of the underground cables based on the position of transformers and cable cabinets.

The visual comparison of the models reveals high coverage of the ground truth model with the inferred models. Figure 19 displays the two ground truth, and DSO inferred models overlapping on the map. Furthermore, we compare the models based on the length of the models' inferred cables. Table 2 summarizes the calculated length of cable for each inferred model, reporting the 88% coverage of ground truth model by the DSO inferred model and 75% coverage by the crowdsourced OGM grid model. The results indicate that the inference approach is capable of inferring accurate grid models. However, the lower availability of grid data in the OGM dataset as compare to the DSO dataset indicates that the quality of the model heavily depends on the availability of distribution grid elements in the area.

6 DISCUSSION AND LIMITATIONS

The results of the campaign and inference approach demonstrate that crowdsourcing is a practical, fast and cost-efficient approach for collecting grid data of an area which can be used for generating

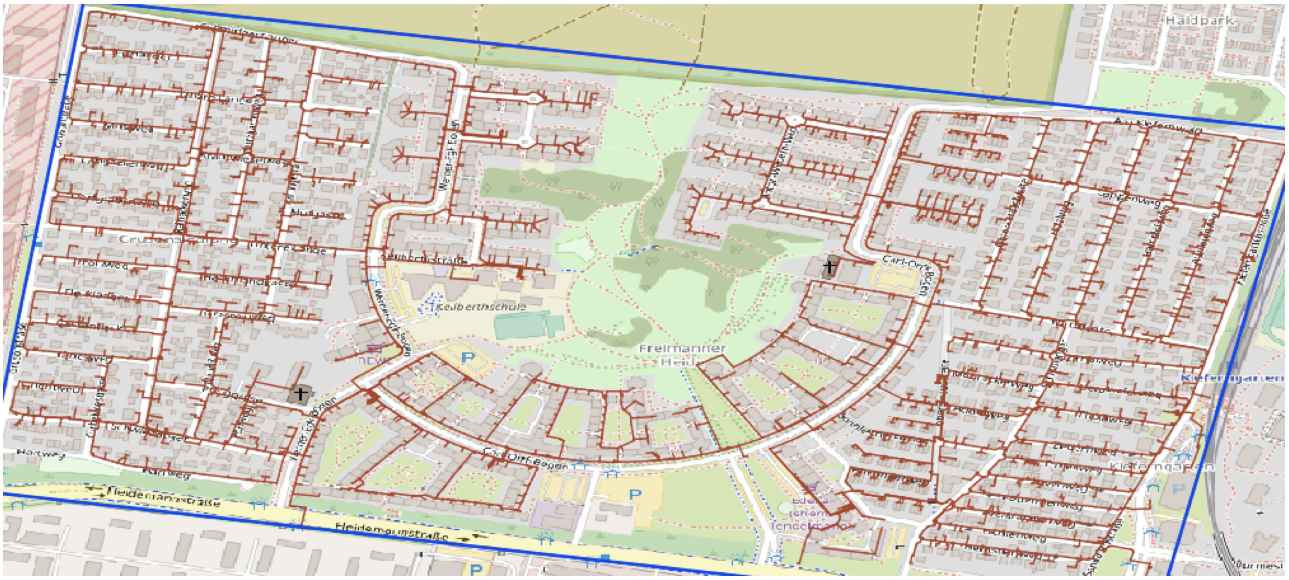


Figure 16: Ground truth model of the Freimann subarea.

Figure 17: The DSO inferred model of the Freimann subarea.

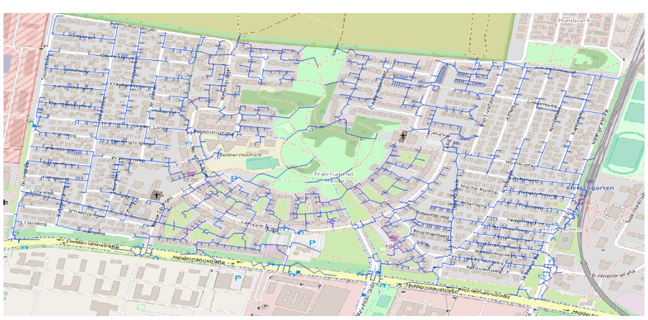


Figure 18: The OGM inferred model of the Freimann subarea.

accurate distribution grid models. Although the inferred models can be used for simple academic simulation studies, they lack the sufficient accuracy to be used by TSOs and DSOs for power infrastructure control purposes. To improve the quality of inferred distribution models, we need to improve the quality and the availability of crowdsourced grid data as well as the inference approach. To improve the quality of grid data, we require increasing the accuracy of collection devices, invest in training of the participants and the integration of grid data and models from various official resources. Although these approaches can be beneficial, they introduce new financial and legislation burdens.

Table 2: Comparison of Ground Truth Model with DSO and OGM Inferred Models

Grid Model	Cable Length (m)	Coverage
DSO Inferred Model	40840	87.86%
OGM Inferred Model	34866	75.01%

To improve the accuracy of the inference approach, we require considering the exceptional inference cases. For example, the discussed inference algorithm does not make a difference between the various types of roads and pathways and assumes the existence of underground cables along any paths such as pedestrian ways and dirt roads, but in reality, the cables are often not installed along the dirt roads. Furthermore, MSTs are a sub-optimal solution when inferring distribution grid models, because DSOs often implement loops in the distribution systems for increasing the resilience and reliability of the power grid. Therefore, we suggest investigating more complex network theory approaches or power flow-based network design approaches.

Furthermore, we should mention that, although several OSM power-related elements contain useful information such as voltage level, several other essential grid characteristics such as the line's thermal parameters are missing, and acquiring such information requires expert knowledge of the local distribution grid. Therefore, we limit ourselves in this work to inferring the topological model of the distribution grids.

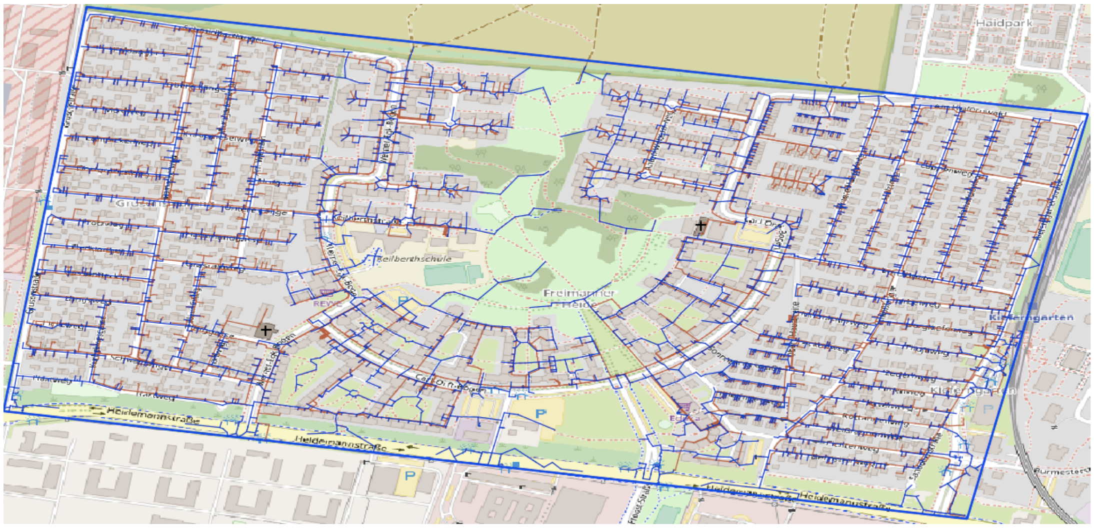


Figure 19: Overlapping ground truth and the DSO inferred model.

7 CONCLUSIONS

In this work, we introduced a crowdsourcing framework developed with the help of the OGM platform which we used for conducting a crowdsourcing campaign. We also proposed an inference approach for generating distribution grid models based on the position of grid devices, consumer endpoints, and roads. The results of the crowdsourcing event report a precision of up to 82% and a recall of up to 65% depending on the performance of the participants. Furthermore, the evaluation of the inferred distribution grid model based on the official complete DSO grid dataset shows a power length accuracy of 88% compared to the ground truth. Our results confirm crowdsourcing as an efficient and beneficial data collection approach for distribution grid device mapping, which in combination with an inference algorithm can provide a practical method to obtain realistic distribution grid models.

As future work, we will work on improving the quality of data collection methods and inference approach by integrating additional data sources and methods, such as automatic detection of distribution grid elements and also automatic improvement and classification of grid data by utilizing deep learning approaches. Furthermore, we intend to investigate the applicability of our crowdsourcing approach and interference algorithm to the North American distribution grid, where the grid elements are more spread out over larger geographical areas, and the gird models are also less accurate in comparison with the German grid. For improving the precision of inference algorithm on the crowdsourced data, we plan to consider the use of more complex graph structures, genetic algorithm based methods and spatial clustering approaches of smart meters data.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Carsten Sperl from Stadtwerke München for his contribution to the project. Most importantly, we would like to thank all the contributors that have helped crowdsource geographical data with the help of OpenGridMap project. This research was supported by the German Federal Ministry of Education and Research Grant (BMBF 01IS12057) and the Alexander von Humboldt Foundation.

REFERENCES

[1] R. Albert, I. Albert, and G. L. Nakarado. 2004. Structural Vulnerability of the North American Power Grid. *Phys. Rev. E* 69 (2004). Issue 2.
[2] The World Bank. 2014. CO2 Emissions from Electricity and Heat Production. (2014). Accessed: 2018-01-03.
[3] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato. 2013. Identification of Power Distribution Network Topology via Voltage Correlation Analysis. In *52nd IEEE Conference on Decision and Control*. 1659–1664.
[4] B. Cloteaux. 2013. Limits in Modeling Power Grid Topology. *2013 IEEE 2nd Network Science Workshop (NSW)* (2013), 16–22.
[5] United Nations Treaty Collection. 2015. Paris Agreement. (2015).
[6] Esri. 2012. *Enterprise GIS and the Smart Electric Grid*. Technical Report.
[7] ESRI ESRI. 1998. Shapefile Technical Description. *An ESRI White Paper* (1998).
[8] PostgreSQL Global Development Group. 2018. PostgreSQL. (2018). https://www.postgresql.org/ Accessed: 2018-01-10.
[9] C. Heipke. 2010. Crowdsourcing Geospatial Data. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, 6 (2010), 550–557.
[10] P. Hines, S. Blumsack, E. C. Sanchez, and C. Barrows. 2010. The Topological and Electrical Structure of Power Grids. In *2010 43rd Hawaii International Conference on System Sciences*. 1–10.
[11] N. Honeth, A. Saleem, K. Zhu, L. Vanfretti, and L. Nordström. 2012. Decentralized Topology Inference of Electrical Distribution Networks. In *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*. 1–8.
[12] W.H. Kersting. 1991. Radial Distribution Test Feeders. *Power Systems, IEEE Transactions on* 6, 3 (1991), 975–985.
[13] V. Kumar, R. Krishan, and Y. R. Sood. 2013. Optimization of Radial Distribution Networks using Path Search Algorithm. *International Journal of Electronics and Electrical Engineering* 1, 3 (2013), 182–187.
[14] Pacific Northwestern National Laborattories. 2008. *Modern Grid Initiative Distribution Taxonomy Final Report*. Technical Report.
[15] T. W. Malone, R. Laubacher, and C. Dellarocas. 2009. Harnessing Crowds: Mapping the Genome of Collective Intelligence. (2009).
[16] B. Meehan. 2013. *Modeling electric distribution with GIS*. Esri Press Redlands.
[17] Stadtwerke München. 2018. Munich City Utilities. (2018). https://www.swm.de/ Accessed: 2018-01-05.
[18] NetworkX. 2018. NetworkX Package. (2018). https://networkx.github.io/ Accessed: 2018-01-03.
[19] NumPy. 2018. NumPy Package. (2018). http://www.numpy.org/ Accessed: 2018-01-03.
[20] City of Munich. 2018. Munich Freimann District. (2018). http://www.muenchen.de/stadtteile/freimann.html Accessed: 2018-01-05.
[21] OpenGridMap. 2018. OpenGridMap Android Application. (2018). https://play.google.com/store/apps/details?id=tanuj.opengridmap Accessed: 2018-01-03.

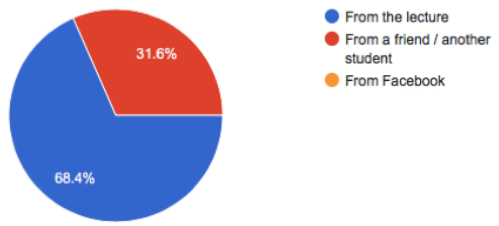


Figure 20: How did you find out about the crowdsourcing event? (19 Responses)

[22] OpenGridMap. 2018. OpenGridMap Distribution Inference Engine. (2018). https://github.com/OpenGridMap/power-grid-detection Accessed: 2018-01-03.

[23] OpenGridMap. 2018. OpenGridMap Homepage. (2018). http://opengridmap.com/ Accessed: 2018-01-03.

[24] OpenStreetMap. 2017. OpenStreetMap Wiki. (2017). https://wiki.openstreetmap.org/wiki/Main_Page Accessed: 2018-01-03.

[25] OpenStreetMap. 2018. OpenStreetMap Power Tag Info. (2018). https://taginfo.openstreetmap.org/keys/power Accessed: 2018-01-03.

[26] G. Paolacci, J. Chandler, and P. G. Ipeirotis. 2010. Running Experiments on Amazon Mechanical Turk. (2010).

[27] PostGIS. 2018. Spatial and Geographic Objects for PostgreSQL. (2018). https://postgis.net/ Accessed: 2018-01-05.

[28] Spatial Reference. 2018. EPSG:31468 Projection. (2018). http://spatialreference.org/ref/epsg/31468/ Accessed: 2018-01-11.

[29] J. Rivera, C. Goebel, and H. A. Jacobsen. 2017. Distributed Convex Optimization for Electric Vehicle Aggregators. *IEEE Transactions on Smart Grid* 8, 4 (2017), 1852–1863.

[30] J. Rivera, C. Goebel, D. Sardari, and H. A. Jacobsen. 2015. *OpenGridMap: An Open Platform for Inferring Power Grids with Crowdsourced Data.* Springer International Publishing, 179–191.

[31] J. Rivera, J. Leimhofer, and H. A. Jacobsen. 2017. OpenGridMap: Towards Automatic Power Grid Simulation Model Generation from Crowdsourced Data. *Computer Science - Research and Development* 32, 1 (2017), 13–23.

[32] Shapely. 2018. Shapely Package. (2018). https://pypi.python.org/pypi/Shapely Accessed: 2018-01-03.

[33] K. Strunz, R.H. Fletcher, R. Campbell, and F. Gao. 2009. Developing Benchmark Models for Low-Voltage Distribution Feeders. In *Power Energy Society General Meeting, 2009. PES '09. IEEE.* 1–3.

[34] C. G. Taroco, E. G. Carrano, R. H. C. Takahashi, and O. M. Neto. 2012. A Faster Genetic Algorithm for Substation Location and Network Design of Power Distribution Systems. In *2012 IEEE Congress on Evolutionary Computation.*

[35] QGIS Development Team. 2018. QGIS. (2018). https://www.qgis.org/en/site/ Accessed: 2018-01-04.

[36] K. M. Ting. 2010. *Precision and Recall.* Springer US, 781–781.

[37] S. von Roon, M. Sutter, F. Samweber, and K. Wachinger. 2014. Netzausbau in Deutschlan. (2014). http://www.kas.de/wf/doc/kas_38837-544-1-30.pdf Accessed: 2018-01-03.

A CROWDSOURCING EVENT SURVEY RESULTS

After the crowdsourcing campaign, we surveyed the participants to evaluate their overall experience. In the following, we include the complete list of questions and the participants' response.

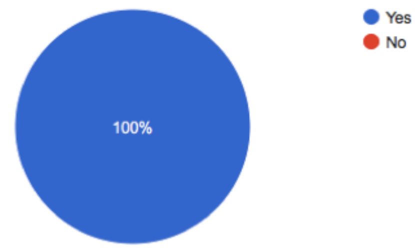


Figure 21: Do you know what the OGM application is used for? (20 Responses)

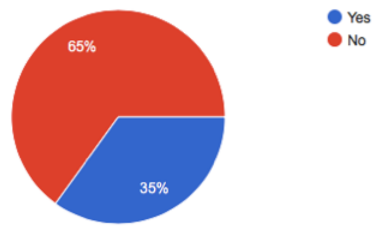


Figure 22: Have you used the OGM application before this campaign? (20 Responses)

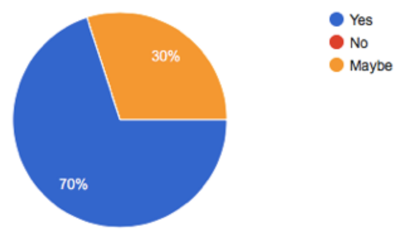


Figure 23: Would you like to use the OGM application to help crowdsourcing in the future? (20 Responses)

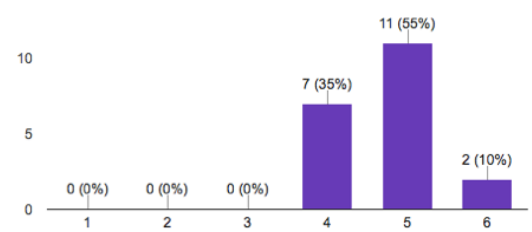


Figure 24: How useful did you find the OGM application for collecting data? (20 Responses)

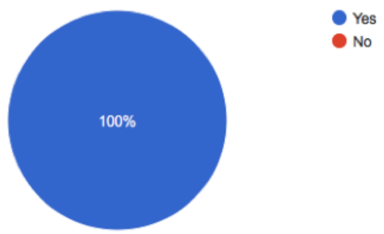


Figure 25: Do you have a clear understanding of the task? (20 Responses)

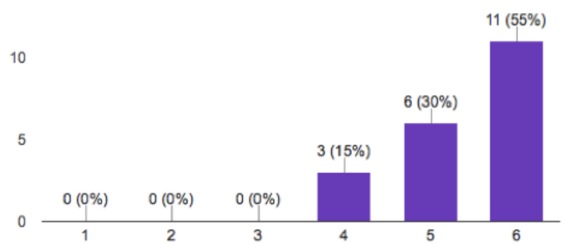


Figure 26: How was the organization of the campaign? (20 Responses)

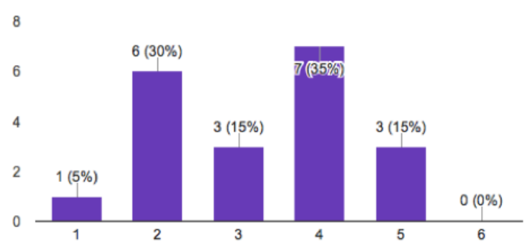


Figure 27: How easy was it to find devices within your area? (20 Responses)

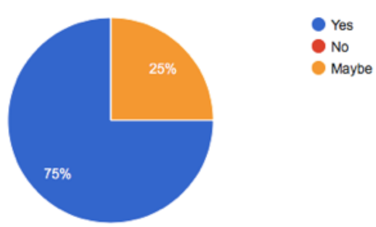


Figure 28: Did you have a clear understanding of the different devices to be mapped? (20 Responses)

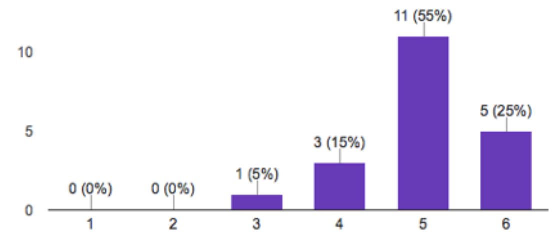


Figure 29: How was the experience of this crowdsourcing campaign? (20 Responses)

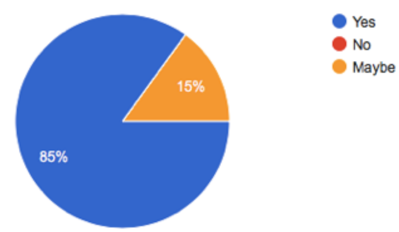


Figure 30: Do you think crowdsourcing is meaningful for scientific research? (20 Responses)

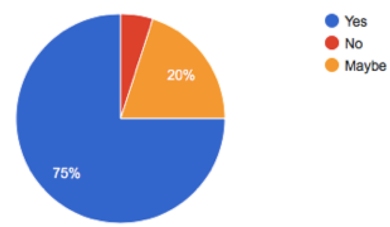


Figure 31: Are you willing to help in crowdsourcing of data in the future? (20 Responses)

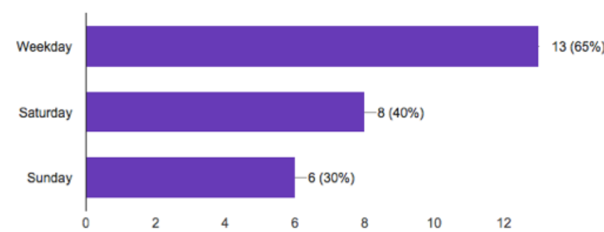
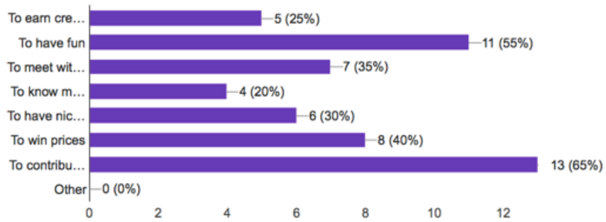


Figure 32: What days do you prefer to participate in a similar campaign in the future? (20 Responses)



Motivation
To earn credits of the course
To have fun
To meet with friends
To know more about power grids
To have a nice time
To win prizes
To contribute to the project objectives
Other

Figure 33: What is the motivation for you to participate in such a crowdsourcing campaign? (20 Responses)