

Reliability of Network Maps

The invention arises from the field of mapping physical networks, in particular electrical distribution grids.

5

Management of an electrical distribution grid requires performing many tasks, in particular responding quickly to repair the facilities in case of an incident, connecting new nodes to the existing grid, performing real-time balancing of power on the grid in order to limit the risks of local overload, and planning for operations for improving the grid in anticipation of future needs and constraints. Conventionally, Geographical Information Systems (GIS) are one type of tool available to grid managers for performing such tasks. Such GIS normally contain maps of the electric distribution grid. The maps contain data organized so as to represent, the most faithfully possible, the physical grid and facilities that constitute it. In practice, the data for the GIS must be entered and updated manually each time needed or that an action on the grid is taken. Human errors lead to inaccuracies and inconsistencies in the GIS relative to the physical reality of the grids. Hence, each difference between the physical distribution grid and the information relating to it can lead to inefficient or counterproductive incidents and operations.

15

Often, the detection of such errors arises by chance. For example, an operator in the field notices an error in a map and the signal. Often, repairing such an error requires the travel of a second operator to precisely identify the error, characterize it and finally correct it. Keeping the GIS data up-to-date is complex, costly and inefficient. Data verification is rarely implemented. For some electrical grids, the data are not cataloged and organized by computer. The GIS are nonexistent and still to be created.

25

Further, the numerous GIS have a structure which is not standardized and are therefore not directly compatible with each other. In situations where several people from the sector must cooperate, the exchange of information is complex.

30 The invention aims to improve the situation.

A method implemented by computer is proposed for constructing a map of an electrical distribution grid comprising at least one source node including a plurality of outputs and a plurality of consumer nodes each supplied by a single output. To each pair of nodes there corresponds a link connecting the two nodes to each other. Each of the consumer nodes is able

35

to communicate with the source node by at least one communication route through the electrical distribution grid. Each route is formed of:

- a single link directly connecting the source node to the consumer nodes; or
- a series of links indirectly connecting the source node to the consumer node through at least one other consumer node forming a relay node.

The method comprises:

- a) during at least one observation period and for each pair of consumer nodes, counting the number of times that the corresponding link forms part of a communication route between the source node and consumer node;
- 10 b) building a graph in which each vertex corresponds to a consumer node and the weight of each edge between two vertices is assigned depending on the number of times the link which connects the corresponding pair of consumer nodes forms part of a communication route established between the source node and a consumer node;
- c) partitioning the vertices of the graph into groups by applying algorithms arranged such that:
 - 15 - two vertices between which the weight of the edge is large and/or two vertices sharing a high number of common neighbors tending to be classified in a single group; whereas
 - two vertices classified in distinct groups tending to have a low or zero link weight and/or having few common neighbors.
- 20 Such a process serves to automate the construction of maps, by creation of new maps, or even by verification and/or correction of existing maps. Most errors can be detected preventively, without waiting for possible detection by an operator sent to the field. By detecting errors quickly, the simplest ones can be quickly corrected and the correction of the more complex ones can be planned in advance. Identified errors can be corrected before harming the management of
- 25 the grid. In most cases, the correction of the errors themselves can be automated, or can at least be implemented without the involvement of an operator in the field being necessary. The method can be implemented, for example, by means of existing communicating devices in the grid. In this case, the installation of dedicated equipment is superfluous. For example, communicating meters, known under the brand name "Linky" are being deployed in France.
- 30 Such meters comprise a communication module able to implement the method, with each meter corresponding to a consumer node, sometimes called "Delivery Point" (DP).

According to another aspect, a system is proposed comprising at least one concentrator and one processing unit, where the processing unit includes memory able to store at least one map of an

35 electric distribution grid. The electrical distribution grid comprises at least one source node

including a plurality of outputs and a plurality of consumer nodes each supplied by a single output. To each pair of nodes there corresponds a link connecting the two nodes to each other. Each of the consumer nodes is able to communicate with the source node by at least one communication route through the electrical distribution grid. Each route is formed of:

- 5 - a single link directly connecting the source node to the consumer nodes; or
- a series of links indirectly connecting the source node to the consumer node through at least one other consumer node forming a relay node.

The concentrator is arranged for:

- a) during at least one observation period and for each pair of consumer nodes, counting the
- 10 number of times that the corresponding link forms part of a communication route between the source node and consumer node;

The processing unit is arranged for receiving the counting data and

- b) building a graph in which each vertex corresponds to a consumer node and the weight of each edge between two vertices is assigned depending on the number of times the link which
- 15 connects the corresponding pair of consumer nodes forms part of a communication route established between the source node and a consumer node;
- c) partitioning the vertices of the graph into groups by applying algorithms arranged such that:
 - two vertices between which the weight of the edge is large and/or two vertices sharing a high number of common neighbors tending to be classified in a single group; whereas
 - 20 - two vertices classified in distinct groups tending to have a low or zero link weight and/or having few common neighbors.

- According to another aspect, a computer program is proposed able to be used in a method for construction of a map such as the methods defined in the present. According to another aspect, a
- 25 nonvolatile computer readable storage medium is proposed on which a program is recorded comprising instructions for the implementation of one of the methods proposed in the present.

The following features can optionally be implemented. They can be implemented independently of each other or in combination of ones with others:

- 30 - The weights are assigned depending on the number of times that the corresponding link forms part of a route for a constant number of routes between each source node and consumer node pair. In this way, the representativeness of the observation period can be improved, including if some consumer nodes are more active than others by establishing more communication routes
- 35 during the observation period than during a typical functional period.

- At the end of counting, an intermediate random sampling operation is done and the partitioning operations are done based on the resulting samples; the succession of sampling, graph construction and partitioning operations are the subject of an iterative loop. In that way
5 the measurement biases are reduced.

- The communication between the nodes is established according to a distributed architecture by Power Line Carrier on the electric distribution grid. The presence of a dedicated communication network is then superfluous. The electrical distribution networks provided with communication
10 modules can be made compatible with little or no equipment modifications.

- The method further comprises:

- d) comparing the composition of the groups resulting from assignment of consumer nodes to outputs according to a pre-existing map;
- 15 e) deducing from the comparison the accuracy of the map or the existence of at least one error;
- f) in case of deducing at least one error, correcting the map based on the composition of the resulting groups.

This way the existing maps can be verified, corrected and/or updated. The partitioning into groups can be combined and crossed with other data already contained in the existing maps.
20

- The partitioning includes implementation of a dividing algorithm, and agglomerating algorithm, an optimization algorithm by density verification or a combination of such algorithms. Such algorithms can be implemented in large part by automation while making it possible to achieve a high reliability of the map information.
25

Other features, details and advantages of the invention will appear upon reading the following detailed description and analyzing the attached drawings, on which:

- Figure 1 is a partial representation of the data from a map to be verified/corrected;
- Figures 2 and 3 each show a functional diagram of an embodiment of at least a part of one
30 method according to the invention;
- Figures 4 and 5 each show a graph prepared during a method according to the invention;
- Figures 6, 7 and 8 show details of a method according to the invention;
- Figure 9 shows a partition implemented according to the method;
- Figures 10, 11 and 12 each show a portion of an embodiment according to the invention;
- 35 - Figure 13 shows a portion of an embodiment according to the invention; and

- Figures 14 and 15 show graphs obtained during a method according to the invention.

The following drawings and description, for the most part, contain elements of some kind. They can therefore not only serve to better understand the present invention, but also contribute to the definition thereof, as applicable.

In the remainder, the term map designates a set of data representative of the mutual organization of the equipment of the network such as geographic and/or connection data. The map, or neighborhood layout, from Figure 1 is a sample visual representation of a part of the data from one map. The purpose of such a representation is essentially to make understanding easier. A network map can comprise other data than that shown on the neighborhood layout, in particular data not strictly geographical such as nominal powers associated with each node, features relating to cabling present, or even technical features of each piece of equipment.

In the example from Figure 1, the map 100 comprises three source nodes 1 each comprising a plurality of outputs 3, or lines, and consumer nodes 5. In the example described here, a source node 1 corresponds to a transformer located between the medium voltage transport network (or type A high-voltage HVA) and the low voltage (LV) distribution network downstream. For example, each of the source nodes 1 comprises between 1 and 10 lines. In general, each line extends in a neighborhood or street for supplying the consumer nodes 5 located there. The Low Voltage (LV) network chosen as an example has a tree structure. The vertex of the tree structure corresponds to a source node 1, here an HVA/LV transformer substation, supplying a set of outputs 3. Each output 3 is extended by a line itself made up of a succession of segments. The consumer nodes 5 form distributed branching points along the segments.

The consumer nodes 5 substantially correspond to "Delivery Points" (PL). Each consumer node 5 forms an interface between the low-voltage LV distribution grid and the secondary electrical grid, generally private. In practice, the consumer node 5 is materialized by equipment, including a meter specifically for accounting for the electrical consumption of the secondary grid for billing purposes. By extension, consumer nodes 5 can be called meters.

The map 100 associates each source node 1 and each consumer node 5 with a geographic position, for example a postal address, and/or a user/substation identifier. Several consumer nodes 5 can be associated with a shared user identifier and/or a shared geographic position.

Several nodes 1, 5 associated respectively with an identifier can further be associated with a common geographic position, for example for multi-dwelling units or dense areas.

For domestic use, such as a residence or business, one meter supplied with single phase is generally sufficient. For some uses, in particular industrial, a three-phase current supply may be necessary. In the examples described here, each output (or line) is three-phase consistent with the usage in France. The applicant observed that the transmission of data by Power Line Carrier (PLC) between two distinct phases of a single output remains possible. As a variant, the lines can be single phase.

10

The map 100 further comprises associations between consumer nodes 5 and outputs 3. This latter type of association is particularly useful information for managing power distribution in the grid.

15 Thus, according to the map 100 shown in Figure 1:

- the consumer nodes 5 represented by a cross are associated with the output 3A;
- the consumer nodes 5 represented by a square are associated with the output 3B;
- the consumer nodes 5 represented by an empty circle are associated with the output 3C;
- the consumer nodes 5 represented by a triangle are associated with the output 3D;
- 20 - the node represented by a warning cone is cataloged in the map 100 but is not associated with any output; and
- the node represented by an exclamation point is not cataloged at all in the map 100.

In theory, the associations contained in the map 100 should correspond to physical reality in the field. Each consumer node 5 should be supplied with electric power by the output 3 of the supply node 1 with which it is associated in the map 100. Just the same, the map 100 contains errors and some information is missing.

In the present context, deliberately simplistic compared to real situations, the errors and missing information are known and identified. Further, the visual representation of the cabling on a map allows some errors and their corrections to be found instinctively. In practice, the quantity of data and their complexity make such a detection burdensome, even impossible, without targeted search and without the assistance of computer tools.

30

In the example from Figure 1 and according to the map 100, the consumer node 5_1 is supplied by the output 3A (represented by a cross). It involves an error: the consumer node 5_1 is in reality supplied by the output 3E.

5 In the example from Figure 1 and according to the map 100, the consumer node 5_2 (represented by a warning cone) is not supplied, which could correspond for example to a meter that is installed but not yet connected. An operator must therefore *a priori* verify the connection on-site for entering the pertinent data in the map 100.

10 In the example from Figure 1 and according to the map 100, the consumer node 5_3 (represented by a square) is supplied by the output 3B. It involves an error: the consumer node 5_3 is in reality supplied by the output 3C.

In the example from Figure 1 and according to the map 100, the consumer node 5_4 does not
15 exist (because not cataloged; represented by an exclamation point). It involves an error: the consumer node 5_4 exists and is supplied by one of the outputs 3.

A first known method for error detection/correction consists of comparing the theoretical output 3 of the consumer node 5 to the theoretical output 3 of consumer nodes 5 geographically near
20 the first one. This supposes that the data used for the comparison are available, reliable, complete and precise. These assumptions are rarely satisfied. In particular, such a method does not give probative results in dense areas or areas having undergone many changes that are not yet transcribed in the map 100.

25 For the consumer nodes 5_1 , 5_2 and 5_4 , the attachment to the source node 1 is incorrect or missing. For the consumer node 5_3 , the attachment to the source node 1 is correct but the attachment to the output 3 is incorrect. This last kind of error can be considered as more difficult to detect and correct. For example, an analysis based on the geographic position of the node does not make it possible to determine reliably the output 3 through which it is supplied.

30 Each of the consumer nodes 5 comprises a communication module, or communication interface, able to communicate with other communication modules via the electrical distribution grid. For example, smart meters or AMM for “*Automated Meter Management*” meters are able to exchange data with a concentrator. The concentrators are generally arranged near source nodes
35 1. The concentrator collects data from several AMM.

- In the example described here, the concentrators in turn send information to a common, remote processing unit, for example via a mobile phone network. Thus, the data-processing unit comprises a processor operationally coupled with memory for implementing at least in part the method defined in the present. As a variant, at least a portion of the data processing can be implemented by the concentrators and/or locally by equipment near the concentrators. For example, counting operations can be done by the concentrators, and then the data relating to the counting are sent to the processing unit.
- 10 In some cases, analyzing and comparing voltage variations recorded by various consumer nodes 5 make it possible to identify correspondences and therefore to determine which consumer nodes 5 are connected to a common source 3. However, voltage changes measured on a consumer node 5 can result from activity on the corresponding private network and/or other private networks supplied by the same output 3. The reliability of such an analysis is therefore
- 15 low. Further, it requires frequent measurements at the same time for all-consumer nodes 5. Further, this involves long-term storage of precise, dated consumption measurements associated with each consumer node. Such data storage and processing could serve to deduce precise information on the usage by each consumer. Making such a usage of private data easier by technical means is generally not desirable. In other words, the automated collection of such data
- 20 could easily be diverted from the intended use (verification of the content of the GIS) for use which would compromise the confidentiality and protection of personal information. Some standards and regulations tend to restrict setting up computer systems which could be easily diverted for such purposes.
- 25 Among the AMM developed, some comprise communication modules operating by Power Line Carrier (PLC). Thus, the data are exchanged by the electric grid itself without the need for providing a dedicated communication network between the source nodes 1 and the consumer nodes 5. The communication network physically uses the cabled electrical grid.
- 30 In some cases, a mesh type communication technology is used: a consumer node 5 is able to communicate with a neighboring consumer node 5. For example, “Linky” meters deployed in France are equipped with PLC modules operating with the “G3” protocol. The G3 protocol is specifically suited to the features of electrical grids and in particular makes it possible to create a decentralized (or distributed) communication network. Thus, the consumer nodes 5 farthest
- 35 from the source node 1, or terminal nodes, can effectively exchange data with some other

consumer nodes 5, or switching nodes, serving as relays/routers. In other words, the information is sent from neighbor to neighbor from one consumer node 5 to a neighboring consumer node 5 all the way to the source node 1. In this way the inevitable attenuations of the PLC signal over long distances can be overcome. Further, this allows an automatic adaptation of the communication routes: two nodes can communicate with each other several times via diverse routes as a function of situations and in particular as a function of the availability of the equipment.

The communication routes between the consumer nodes 5 are preferably established between consumer nodes 5 connected to one shared output 3.

In the remainder, the route connecting a source node 1 to a terminal consumer node 5 is called the “communication route.” The routes are referenced $9_{i,j}$, with the index i identifying the source node 1 and the index j identifying the consumer node 5. For one pair of a source node 1 and a terminal consumer node 5, several different routes $9_{i,j}$ can exist. Each route 9 is formed of:

- a single link $7_{i,j}$ (or connection) directly connecting the source node 1 to the consumer nodes 5; or
- a series of links $7_{i,x}, 7_{x,y}, 7_{y,z}, 7_{z,j}$ indirectly connecting the source node 1 to the consumer node 5 through at least one other consumer node $5_x, 5_y, 5_z$ forming relay nodes. In the present context two routes $9_{i,j}$ connecting the same node pair i, j by different links 7 are distinguished from each other.

A link $7_{x,y}$ is defined as a subset of a route 9 and a path connecting a first node (source node 1 or consumer node 5) to a second node (source node 1 or consumer node 5). In a network, for each pair of nodes x, y there corresponds a link $7_{x,y}$ connecting the nodes x, y to each other. Each link $7_{x,y}$ is made up of the succession of one or more segments of the electric grid. In practice, two links $7_{x,y}$ made up of different segments of the electric grid can be used for sending data between the first node x and the second node y . In the present context, these differences are not considered and the distinction is not made between two links $7_{x,y}$ established by different segments of the network. In other words, it is considered that there is a single link $7_{x,y}$ for each pair of nodes x, y . A link $7_{x,y}$ can connect two nodes physically remote from each other, including if there are other nodes present between the first two. For example, a route $9_{i,j}$ can comprise links $7_{x,y}$ which are superposed at least in part on the physical network such that, on one cable segment, the data can pass in two opposite directions for a single route $9_{i,j}$.

In a first part of the method, the graph is prepared. The graph is representative of the links $7_{x,y}$ used while establishing communication routes $9_{i,j}$.

In an operation 201 from the first part of the method, the number of times that each link $7_{x,y}$, $7_{x,z}$, $7_{y,z}$ forms a part of a communication route $9_{i,j}$ between a source node 1_i and a terminal consumer node 5_j is counted. The operation 201, counting, is implemented over an observation time and for each pair of consumer nodes x, y ; y, z ; x, z .

In an operation 202 for the first part of the method, a graph is built in which each vertex corresponds to a consumer node $5_x, 5_y, 5_z$ and the weight $p_{x,y}, p_{y,z}, p_{x,z}$ of each edge between two vertices is assigned depending on the number of times the link $7_{x,y}, 7_{y,z}, 7_{x,z}$ which connects the pair of consumer nodes $5_x, 5_y$; $5_x, 5_z$; $5_y, 5_z$ forms part of a communication route 9 established between a source node 1 and a terminal consumer node 5.

The longer the observation and counting time, the more the resulting graph is representative of the variety of communication routes $9_{i,j}$ established and the greater are the computer resources necessary for processing.

In a first embodiment, the operation 202 is implemented considering all of the results (of counting) for the preceding operation 201. The resulting graph is considered to be “complete.” The operation 202 therefore has no sampling. The remainder of the method can be implemented with this single complete graph. In this case, it is appropriate to select the observation length and period so as to limit the amount of data to be processed as a function of the available computer resources while assuring the representativeness of the observations.

In a second embodiment, the operation 202 is implemented considering only a part of the monitored communication routes 9 in operation 201. In other words, a set of communication routes $9_{i,j}$ is extracted and only the counts from corresponding links $7_{x,y}$ are retained at the end of counting. In order to improve the representativeness of the graph, the sampling is preferably random. In this way, the quantity of data to process to construct the graphs can be selected *a posteriori*. The sampling can also be temporal: several counts are done over distinct temporal ranges. The resulting graphs are considered as “sampled.” The remainder of the method can be implemented with several graphs obtained after sampling.

In order to improve the representativeness of the graph, and independent of the preceding two embodiments, the number of routes $9_{i,j}$ for each terminal consumer node j for which the links $7_{x,y}$ are counted is made the most homogeneous possible, and preferably equal, for example by sampling. In other words, the number of routes monitored during the operation 201 is preferably
 5 identical for each terminal consumer node 5_j .

In the resulting graph, the greater the frequency of use of the link $7_{x,y}$ between two nodes x,y , the greater the strength of the connection uniting the two nodes x,y . Figure 4 shows a sample of a representation of a resulting graph, here based on communication routes under PLC G3
 10 protocol. The central node corresponds to the single source node 1 of the test network (shown as a triangle on Figure 1). The source node 1 is strongly linked to most of the consumer nodes 5. This is explained by the specific positioning of the source node 1 at the center of the tree network. In the examples described here, the source nodes 1 are ignored during construction of the graph. The result of eliminating the source node 1 from Figure 4 is shown in Figure 5. This
 15 can be done by effective elimination of the vertex corresponding to the source node 1 or by not counting the links $7_{x,y}$ for which x or y corresponds to the source node 1 during the operation 201.

In some situations, consumer nodes 5 can have links only with the source node 1. This generally
 20 corresponds to the case of nodes geographically near the source node 1. The corresponding established communication routes 9 are all made of the single link $7_{i,j}$ directly between the source node 1_i and said terminal consumer node 5_j , without a consumer node 5 forming a relay. In this case, the vertices of such consumer nodes 5 are isolated (not connected to other vertices) after elimination of the vertex for the source node 1. Such nodes cannot be localized. The
 25 corresponding vertices are therefore also eliminated from the graph.

The resulting graph can be represented in the form of the following matrix G , for which the index 1 corresponds to the source node 1 and is therefore absent and the indexes 2 to N correspond respectively to each of the $N-1$ consumer nodes 5, excluding isolated nodes:

30

$$G = \begin{bmatrix} p_{2,2} & \cdots & p_{2,N} \\ \vdots & \ddots & \vdots \\ p_{N,2} & \cdots & p_{N,N} \end{bmatrix}$$

In a second part of the method, the consumer nodes 5 are distributed in groups by partitioning the vertices of the graph. The partitioning is implemented by applying group detection algorithms arranged to tend towards the following two objectives:

- two vertices between which the weight $p_{x,y}$ of the edge is large and/or two vertices sharing a high number of common neighbors belong to a single group; whereas
- two vertices belonging to distinct groups have a low or zero link weight and have few common neighbors. In other words, the algorithms are arranged for defining groups, or communities, for which the nodes 5 of a single group are strongly connected with each other and the intergroup connections are weak or nonexistent. Each node 5 is assigned to a single group.

10

Examples of algorithms are described below. The algorithms can be classified into three types:

- dividing algorithms detecting the intergroup links and eliminating them to emphasize the groups;
- agglomerative algorithms gathering nodes or groups of nodes having similarities into a common group;
- algorithms for optimization by density verification: objective functions are maximized for defining groups of nodes on the basis of density criteria.

15

In a first example of the method, the second part of the method comprises implementation of an agglomerative algorithm. The second part is, for example, applied to a plurality of sampled graphs obtain by implementing the first part of the method according to the second embodiment. The agglomerative algorithm comprises for example the Louvain algorithm. The first sample method is shown by the flowchart in Figure 2.

20

The group detection algorithm is applied to a plurality of sample graphs. Each sample graph is obtained by random selection of communication routes 9 observed over one month and for which the links 7 are counted.

25

An identical number of routes 9 for each terminal node 5_j is selected for counting the links 7. In that way, the processing of the terminal nodes 5_j is equitable and the base for observation remains as diversified as possible. If some terminal nodes 5 are more active than others, that does not create a representativeness bias (a heterogeneity of established routes for each terminal node 5 has no consequence). Meeting such a criterion can be compared to a resampling, or "Bootstrap," for a statistical inference. The number of routes 9 that are chosen for monitoring for counting the links 7 is itself a hyper-parameter: by playing on this number, it is possible to

30

35

select the representativeness of the graph and the processing resources (computers) which are necessary.

Partitioning by Louvain algorithm is based on the maximization of a metric estimating the quality of the partition achieved, called “modularity.” The modularity of a partition measures the difference between the proportion of links internal to the group and the same quantity in an equivalent graph but defined randomly in which no group structure is expected. The modularity Q of a partition $P=\{c_1, \dots, c_k\}$ is defined as follows:

$$Q = \frac{1}{2m} \sum_{i,j} \left[p_{i,j} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

10

where $p_{i,j}$ is the weight associated with the link between the nodes i and j ;

$k_i = \sum_j p_{i,j}$ is the sum of the weights of the links from the node i ;

$\delta(c_i, c_j) = 1$ if and only if $c_i=c_j$.

The factor $\frac{1}{2m}$ is added in order to normalize the possible values of Q in the interval $[-1, 1]$, with

15 $m = \sum_{i,j} p_{i,j}$.

The modularity Q is positive for a graph and a partition for which there is no link (or an edge with zero weight) between two nodes of different groups but at least one link between two nodes of the same group. The modularity Q is negative for a graph and a partition for which there is no link between two nodes of the same group but at least one link between two nodes of different groups. The modularity Q tends to zero when a graph in which the arcs (or edges) are randomly distributed. In the example described here, it is considered that a graph partition is significant (pertinent) if the modularity Q is greater than or equal to 0.3. The selection of such a threshold is based on the literature, and in particular on the following document: M. E. J. Newman, *Fast algorithm for detecting community structure in networks*, Phys. Rev. E 69, 066133 (2004). As a variant, various thresholds are chosen.

One example of application of the Louvain algorithm to a graph is shown in Figure 6. The application of the Louvain algorithm can be decomposed in the following way.

30

In an initial state (shown in bottom left of Figure 6, drawn from M. E. J. Newman and M. Girvan, *Finding and evaluating community structure in networks*, Phys. Rev. E 69, 026113

(2004) in Newman and Girvan (2004)), the group number is equal to the number of nodes/vertices of the graph. Each node is randomly classified in a group.

Next, in a first sub operation, for each node i , the resulting modularity Q improvement is estimated by removing this node i from the group thereof and placing it in the group of the node j . The node i is moved into the group for which the modularity Q improvement is a maximum, if and only if the improvement is positive. The node i is then “reallocated.” The estimation-reallocation series is reiterated so long as the modularity Q of the partition is improved. A graph with optimized modularity Q results (shown in top of Figure 6).

Next, in the second sub operation, the graph with optimized modularity Q is transformed into a graph of groups (shown in the bottom at the center of Figure 6):

- The vertices each representing one node of a single group are merged into a single vertex representing said group so as to obtain a graph in which each vertex now represents a group.

- The edges between two nodes belonging to two distinct groups are merged into one edge now connecting the two vertices corresponding to said two groups. That way, each edge now represents the intergroup connections. During this merger, the weight of each edge between two groups-vertices is obtained by summing the weight of each merged edge.

- The edges internal to the groups (between the nodes of a single group) are transformed into a loop with the weight equal to the sum of the weights of the internal edges.

The application of the Louvain algorithm is reiterated from the random classification of nodes in the group so long as it serves to increase the modularity Q . The result of the second iteration is shown in bottom right of Figure 6.

The application of the algorithm forms a heuristic in the meaning of combinatorial optimization. The resulting partition (shown by the resulting graph) depends on the initialization (random) of the partition. Thus, the resulting partition can vary between two successive executions of the algorithm. The application of the algorithm is therefore preferably reiterated several times on various sample graphs in order to improve the stability of the result.

The second part of the method can comprise the construction of a similarity matrix M .

The similarity matrix M is square and has a dimension equal to the number of vertices of the graph, meaning the number of consumer nodes 5 for the map 100 from which the number of

nodes not connected to the remainder of the graph are taken away as described above. Since the graph is not directed, the similarity matrix M is symmetric. The similarity matrix M is constructed by iterative loops, as is shown in Figure 2 and described in detail below.

- 5 In an operation 301, the similarity matrix M is initialized to zero. The iteration, it , counter is initialized to one.

Then, an algorithm for detection of groups on a randomly constructed graph is implemented in one iteration, for example according to the Louvain method described above.

10

The results of the detection of groups done in the current iteration, it , are stored in the matrix M . In an operation 302, a group detection matrix D_{it} is constructed for a randomly built graph. The dimension of the matrix D_{it} is identical to that of the similarity matrix M . For example, the matrix D_{it} takes the form of a square matrix for which each line and each column correspond to one node. Each term has the value “1” when the nodes corresponding to the row-column intersection belong to a common group and otherwise has the value “0”. And in operation 303, the matrix D_{it} is added to the matrix M . In other words, on each iteration of group detection algorithm on a graph, the corresponding degree matrix D_{it} is added to the similarity matrix M .

15

- 20 After construction, the similarity matrix M is used to estimate the probability that two consumer nodes belong to a common group.

After the operation 303, it is checked whether the criterion for stopping the iterations is reached. Here, the stopping criterion has the form of a number it_max of iterations to use. If this is not the case, the iteration counter is incremented by one and then operations 302 and 303 are iterated again on a new randomly built graph.

25

In an operation 305, at the end of the loop, the matrix M is normalized. The normalization of the matrix M comprises a division by the number of iterations done above (M/it_max).

30

The normalized similarity matrix M can be seen as the adjacency matrix of a graph for which the weight of the links corresponds to the probability that the two nodes belong to a single group.

In the present context, two nodes belonging to the same group can be considered as being electrically linked to the same feeder 3 of the source node 1, and not only to the same source node 1.

- 5 A final partitioning can optionally be obtained by an additional operation 306 by applying a group detection algorithm, such as the Louvain algorithm, to the normalized similarity matrix M. Each consumer node 5 is thus placed in a group, which represents a structure of nodes highly connected by communicating links and by a single electrical connection at a feeder of the transformer substation.

10

In a second example of the method, the second part of the method comprises implementation of a dividing algorithm. This second part is, for example, applied to a single complete graph obtained by the first part of the method. The dividing algorithm comprises for example an algorithm for calculation of betweenness centrality, for example that of Girvan-Newman (or
15 “edge betweenness”). The second sample method is shown by the flowchart in Figure 3.

In the initial operation 401, a graph obtained according to the first part of the method, here a single complete graph, is prepared.

- 20 Then, in an operation 402, an algorithm for detection of groups on the complete graph is used, for example the Girvan-Newman algorithm which will be described later, so as to obtain a partition of nodes into groups.

- Then, an operation of correction 403 of the partitions is used: each partition of the graph is
25 corrected by application of a nearest neighbor algorithm (described in detail below) to each node for determining the most likely membership in a group of consumer nodes 5. Next, the pertinence of the regrouping of nodes into a group is measured by the conductance (operation 404). In an operation 405, the nodes of the group with the best connectivity are detached from the graph of the routes.

30

The loop is reiterated on the thus simplified graph: the remaining nodes are again cut into groups by application of the group detection algorithm (repetition of operation 402) and so on. The iterations are stopped when a stopping criterion is reached, or when only one single group remains.

35

The example of the Girvan-Newman (“Edge Betweenness”) algorithm is now going to be described in detail.

5 Instead of determining the central links indispensable to the structure of the group, the “edge betweenness” dividing algorithm is intended to detect the links connecting groups of nodes with each other. In that way, instead of constructing groups by addition of strong links to a set of non-connected nodes, the groups are revealed by successive elimination of intergroup links. The figure 7 shows the identification of the link B_{ij} having the largest betweenness measure.

- 10 To identify the intergroup links, meaning those corresponding to a high betweenness centrality, several ways are known:
- shortest path betweenness: determining the shortest paths between two nodes of the graph and counting the number of times each link is taken;
 - random walk betweenness: establish random paths between two nodes, count the number of
 - 15 times that each link is taken and add this number for each pair of nodes of the graph;
 - current-flow betweenness: calculate the absolute value of the current between two nodes, summed over all the pairs of source and sink nodes of the graph.

- In M. E. J. Newman and M. Girvan, *Finding and evaluating community structure in networks*,
 20 Phys. Rev. E 69, 026113 (2004), the dividing algorithms (“edge betweenness”) can be defined in the following way:
- i. calculate the betweenness centrality scores of all the links in the graph;
 - ii. find the link with the highest score and eliminated from the network;
 - iii. recalculate the betweenness centrality for all remaining links;
 - 25 iv. repeat step ii so long as high betweenness links can be eliminated without degrading the groups determined in the previous iterations or as long as the criterion of maximization of the modularity function is satisfied.

Figure 8 is an example of groups identified by successive elimination of bold links; here there
 30 are four groups. Figure 9 shows a tree diagram (or cluster tree) which can be obtained after implementing a dividing algorithm. The ends of the branches (bottom Figure 9) represent the nodes of the graph. The tree represents a connected network cut into smaller and smaller groups, along a path descending the tree. The dashed line indicates the distribution of the nodes into four communities and the number of selected groups.

The nearest neighbor algorithm is now going to be described. After simplification of the graph by elimination of the smallest conductance group, each node is reassigned to the most probable group following identification of the group from the nearest neighbors thereof (or the most similar nodes).

5

This algorithm comprises the following operations:

- For each node:

- i. Identify the neighboring nodes (those with which at least one link is shared);
- ii. For each neighbor node, calculate the level of common neighbors (see below);
- 10 iii. Assign the node the group of the closest neighbors.

The similarity between two nodes is measured by the Jaccard index, such as defined for example in Linyuan Lu and Tao Zhou, *Link prediction in complex network: A survey*, Phys. A 390 (2011):

15

$$S(X, Y) = \frac{\Gamma(X) \cap \Gamma(Y)}{\Gamma(X) \cup \Gamma(Y)}$$

where $\Gamma(X)$ and $\Gamma(Y)$ are the set of neighbors of the node X and Y respectively.

The numerator represents the list of neighbors common to the nodes X and Y. The denominator represents the complete list of all their neighbors. The index S measures the level of common neighbors and is included between 0 and 1. The higher the level, the greater is the probability of the two nodes X and Y belonging to a single group.

Figure 10 shows an example of two nodes u and v having three common neighbors. Figure 11 shows a graph in which the initial assignment of two nodes (identified by arrows) is incorrect. The implementation of the nearest neighbor algorithm serves to reassign these two nodes to the group of the nearest neighbors thereof. Figure 12 shows a graph in which the initial assignment of a node (identified by a narrow) seems incorrect. Here it is chosen to reassign a node to a group if and only if more than half of the neighboring nodes belong to a single group. The use of the nearest neighbor algorithm here therefore leads to leaving the node assigned to the initial group even if more than half of the neighboring nodes are assigned to two other groups.

The partitioning coming from the aforementioned operations can be improved by applying successively the edge betweenness algorithm (operation 402) on the nodes belonging to the least dense groups. In other words, the operation 401 can itself be an iterative loop. The nodes belonging to the most relevant group are determined by the measurement of the conductance, meaning the fraction of links leaving the community over the total volume of links from the nodes of the group. The lower the conductance, the less the nodes of the group share relations with nodes from other groups.

The following is the formula for the conductance Φ by group i :

$$\Phi_i = \frac{C_s}{2m_s + C_s}$$

where C_s is the number of links extending outside the group (towards nodes from other groups), and

m_s is the number of links for which at least one of the ends is a node of group i .

Figure 13 shows a succession of iterations for group detection. In the first iteration, the group labeled 3 has the lowest conductance. The nodes in the links from group 3 are therefore eliminated from the graph. In the second iteration, group 2 has the lowest conductance. The nodes and the links from group 2 are therefore in turn eliminated from the graph. At the end of the last generation N , 10 groups were successively identified and eliminated. Four nodes remain unconnected to the others and are therefore considered as isolated.

Figure 14 shows the results of a test by the applicant on communication routes created according to the PLC G3 protocol with 60 consumer nodes. The left part shows the resulting graph at the end of the first part of the method (observation phase and graph construction). The right part shows the results obtained after application of the group detection algorithm: three groups were identified and have respectively the conductances 0.06, 0.13 and 0.13.

Figure 15 shows the comparison of a graph established for the nodes where the initial assignment of the groups corresponds to the cartographic data of an existing GIS (at left in the figure) and the same graph after correction and reassignment of groups according to the method. In this example, the GIS comprises data about the assignment of nodes (or meters) to various feeders 3 of the source node 1, here a low-voltage transformer substation. In the left part, the

three different shapes used for representing the consumer nodes 5 respectively indicate the feeder 3 to which each node 5 is thought to be connected according to the information from the GIS.

- 5 In the trial done by the applicant, verifications were done in the field to identify possible errors with certainty:
- seven nodes incorrectly cataloged in the GIS as connected to the feeder 3 corresponding to the “triangle” sign were in reality connected to the feeder 3 corresponding to the “square” sign; and
 - seven nodes incorrectly cataloged in the GIS as connected to the feeder 3 corresponding to the
- 10 “triangle” sign were in reality connected to the feeder 3 corresponding to the “circle” sign.

As the left part of the figure shows, the method allows identifying the nodes 5 for which the assignment is incorrect on the basis of observations of the communication routes established in the network of nodes. Beyond the detection of errors itself, the method serves to determine how

15 to correct them. In other words, it is possible to identify the correct feeder 3 assignment for each incorrectly referenced consumer node 5.

During its tests, the applicant applied the method to 81 source nodes 1 and 7698 consumer nodes 5. The following is the distribution of the cases studied:

- 20 - 39 source nodes 1 connected to fewer than 50 consumer nodes 5;
- 19 source nodes 1 connected to more than 50 consumer nodes 5 and fewer than 100 consumer nodes 5;
- 23 source nodes 1 connected to more than 100 consumer nodes 5; and
- 78% of the source nodes 1 studied comprise between 2 and 6 feeders 3.

25 The incorrect or missing results seem essentially dependent on the number of consumer nodes 5 per group and the number of feeders per source node 1. In fact the smaller the number of consumer nodes 5 per feeder, the more the margins of error of the group detection algorithms increase.

30 For these tests, field verification served to clear any doubt and to detect any error at the outcome of the application of the method. A 96% confirmation or valid correction rate was reached. For 4% of the consumer nodes 3, the method was not able to identify an assigned feeder or incorrectly identified and assigned feeder.

35

In the preceding examples, the results of the methods, meaning the partitioning into groups of nodes, are used to verify, and optionally correct, pre-existing maps. In variants, the existing maps contain little reliable information or even don't exist at all. Such a method can be implemented for creating such a map without starting from a pre-existing map. It is understood
5 here that "construction" of a map means creating a map from nothing or else creating a map by verification and/or correction of a pre-existing map.

The invention is not limited to the method and system examples described above, only as examples, but it encompasses all variants that could be considered by the person skilled in the
10 art in the scope of the protection sought.

Claims

1. A method implemented by computer is proposed for constructing a map (100) of an electrical distribution grid comprising at least one source node (1) including a plurality of
5 outputs (3) and a plurality of consumer nodes (5) each supplied by a single output (3) where for each pair of nodes ($5_x, 5_y$) there corresponds a link ($7_{x,y}$) connecting the two nodes ($5_x, 5_y$) to each other.

where each of the consumer nodes (5_j) is able to communicate with the source node (1_i) by at least one communication route ($9_{i,j}$) through the electrical distribution grid,

10 each route ($9_{i,j}$) is formed of:

- a single link ($7_{i,j}$) directly connecting the source node (1_i) to the consumer nodes (5_j); or
- a series of links ($7_{i,x}, 7_{x,y}, 7_{y,z}, 7_{z,j}$) indirectly connecting the source node (1_i) to the consumer node (5_j) through at least one other consumer node ($5_x, 5_y, 5_z$) forming a relay node, the method comprising:

15 a) during at least one observation period and for each pair of consumer nodes ($5_x, 5_y; 5_x, 5_z; 5_x, 5_j; 5_y, 5_z; 5_y, 5_j; 5_z, 5_j$), counting (201) the number of times that the corresponding link ($7_{x,y}, 7_{x,z}, 7_{x,j}, 7_{y,z}, 7_{y,j}, 7_{z,j}$) forms part of a communication route ($9_{i,j}$) between the source node (1_i) and a consumer node (5_j);

b) build (202) a graph in which each vertex corresponds to a consumer node ($5_x, 5_y, 5_z, 5_j$) and
20 the weight ($p_{x,y}, p_{x,z}, p_{x,j}, p_{y,z}, p_{y,j}, p_{z,j}$) of each edge between two vertices is assigned depending on the number of times the link ($7_{x,y}, 7_{x,z}, 7_{x,j}, 7_{y,z}, 7_{y,j}, 7_{z,j}$) which connects the pair of consumer nodes ($5_x, 5_y; 5_x, 5_z; 5_y, 5_z$) forms part of a communication route (9) established between a source node (1) and a terminal consumer node (5).

c) partitioning the vertices of the graph into groups by applying algorithms arranged such that:

- 25
- two vertices between which the weight (p) of the edge is large and/or two vertices sharing a high number of common neighbors tending to be classified in a single group; whereas
 - two vertices classified in distinct groups tending to have a low or zero link weight (p) and/or having few common neighbors.

30 2. The method according to claim 1, wherein the weights (p) are assigned depending on the number of times that the corresponding link (7) forms part of a route (9) for a constant number of routes ($9_{i,j}$) between each source node (1_i) and consumer node (5_j) pair.

35 3. The method according to one of the preceding claims, wherein, at the end of counting, an intermediate random sampling operation is done and the partitioning operations are done based

on the resulting samples; the succession of sampling, graph construction and partitioning operations are the subject of an iterative loop.

4. The method according to one of the preceding claims, wherein the communication
5 between the nodes (1, 5) is established according to a distributed architecture by Power Line Carrier on the electric distribution grid.

5. The method according to one of the preceding claims, further comprising:

d) comparing the composition of the groups resulting from assignment of consumer nodes (5) to
10 outputs (3) according to a pre-existing map (100);

e) deducing from the comparison the accuracy of the map (100) or the existence of at least one error;

f) in case of deducing at least one error, correcting the map (100) based on the composition of the resulting groups.

15

6. The method according to one of the preceding claims, wherein the partitioning includes implementation of a dividing algorithm, and agglomerating algorithm, an optimization algorithm by density verification or a combination of such algorithms.

20 7. The system comprising at least one concentrator and one processing unit, where the processing unit includes memory able to store at least one map (100) of an electric distribution grid, the electrical distribution grid comprising at least one source node (1) including a plurality of outputs (3) and a plurality of consumer nodes (5) each supplied by a single output (3) where for each pair of nodes ($5_x, 5_y$) there corresponds a link ($7_{x,y}$) connecting the two nodes ($5_x, 5_y$) to
25 each other.

where each of the consumer nodes (5_j) is able to communicate with the source node (1_i) by at least one communication route ($9_{i,j}$) through the electrical distribution grid, each route ($9_{i,j}$) is formed of:

- a single link ($7_{i,j}$) directly connecting the source node (1_i) to the consumer nodes (5_j); or

30 - a series of links ($7_{i,x}, 7_{x,y}, 7_{y,z}, 7_{z,j}$) indirectly connecting the source node (1_i) to the consumer node (5_j) through at least one other consumer node ($5_x, 5_y, 5_z$) forming a relay node,

where the concentrator is arranged for:

a) collecting and sending the processing unit data from counting the number of times that each link ($7_{x,y}, 7_{x,z}, 7_{x,j}, 7_{y,z}, 7_{y,j}, 7_{z,j}$) corresponding to each pair of consumer notes ($5_x, 5_y; 5_x, 5_z; 5_x,$

$5_j; 5_y, 5_z; 5_y, 5_j; 5_z, 5_j)$ forms a part of the communication route $(9_{i,j})$ between the source node (1_i) and a consumer node (5_j) during at least one observation period,

where the processing unit is arranged for receiving the counting data and

- b) build (202) a graph in which each vertex corresponds to a consumer node $(5_x, 5_y, 5_z, 5_j)$ and the weight $(p_{x,y}, p_{x,z}, p_{x,j}, p_{y,z}, p_{y,j}, p_{z,j})$ of each edge between two vertices is assigned depending on the number of times the link $(7_{x,y}, 7_{x,z}, 7_{x,j}, 7_{y,z}, 7_{y,j}, 7_{z,j})$ which connects the pair of consumer nodes $(5_x, 5_y; 5_x, 5_z; 5_y, 5_z)$ forms part of a communication route (9) established between a source node (1) and a terminal consumer node (5).

c) partitioning the vertices of the graph into groups by applying algorithms arranged such that:

- 10 - two vertices between which the weight (p) of the edge is large and/or two vertices sharing a high number of common neighbors tending to be classified in a single group; whereas
 - two vertices classified in distinct groups tending to have a low or zero link weight (p) and/or having few common neighbors.

- 15 8. A computer program comprising instructions for implementing the method according to one of claims 1 to 6 when this program is executed by a processor.

9. A nonvolatile computer readable storage medium on which a program is recorded comprising instructions for the implementation of the method according to one of claims 1 to 6.