Applied Energy 210 (2018) 1101–1112

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Topological properties of medium voltage electricity distribution networks

Sathsara Abeysinghe [a], Jianzhong Wu [a,*], Mahesh Sooriyabandara [b], Muditha Abeysekera [a], Tao Xu [c], Chengshan Wang [c]

[a] Institute of Energy, School of Engineering, Cardiff University, Cardiff CF24 3AA, United Kingdom
[b] Telecommunications Research Laboratory, Toshiba Research Europe Limited, 32 Queen Square, Bristol BS1 4ND, England, United Kingdom
[c] Tianjin University, China

HIGHLIGHTS

- Topological properties of medium voltage distribution networks were investigated.
- Graph theory and complex networks analysis techniques were used in the study.
- A novel depth dependent approach was developed to investigate topological properties.
- Key properties to characterise distribution networks were identified and quantified.

ARTICLE INFO

Article history:
Received 16 May 2017
Received in revised form 28 June 2017
Accepted 29 June 2017
Available online 12 July 2017

Keywords:
Complex network analysis
Electricity distribution networks
Low carbon technologies
Medium voltage
Topological properties

ABSTRACT

With a large penetration of low carbon technologies (LCTs) at medium voltage and low voltage levels, electricity distribution networks are undergoing rapid changes. Much research has been carried out to analyse the impact of employing LCTs in distribution networks based on either real or synthetic network samples. Results of such studies are usually case specific and of limited applicability to other networks. Topological properties of a distribution networks describe how different network components are located and connected, which are critical for the investigation of network performance. However, the number of network modelling and simulation platforms are limited in the open literature which can provide random realistic representations of electricity distribution networks. Thus, it is difficult to arrive to generalized and robust conclusions on impact studies of LCTs. As the initial step to bridge this gap, this paper studies the topological properties of real-world electricity distribution networks at the medium voltage level by employing the techniques from complex networks analysis and graph theory. The networks have been modelled as graphs with nodes representing electrical components of the network and links standing for the connections between the nodes through distribution lines. The key topological properties that characterize different types (urban and sub-urban) of distribution networks have been identified and quantified. A novel approach to obtain depth-dependent topological properties has also been developed. Results show that the node degree and edge length related graph properties are a key to characterize different types of electricity distribution networks and depth dependent network properties are able to better characterize the topological properties of urban and sub-urban networks.

1. Introduction

The increasing penetration of Distributed Energy Resources (DERs) together with load growth, the new requirements of decarbonisation, efficiency, security and quality of power supply and the deregulation of the electricity markets have significantly changed the traditional approaches to planning, design and operation of the electrical power system. The High Voltage (HV) transmission grid is usually the backbone of a national power system. Distribution networks are very large and topologically complicated systems which connect the HV transmission grid to end users. Electricity distribution networks are undergoing rapid changes in recent years. With numerous Low Carbon Technologies (LCTs), e.g. electric vehicles, wind and solar farms, introducing at the Medium Voltage (MV) and

* Corresponding author.
E-mail address: wuj5@cardiff.ac.uk (J. Wu).

http://dx.doi.org/10.1016/j.apenergy.2017.06.113
0306-2619/

Nomenclature

Abbreviations

CNA	complex networks analysis
DER	distributed energy resources
HV	high voltage
LCT	low carbon technologies
LV	low voltage
MV	medium voltage

Sets and matrices

A	adjacency matrix
D	distance matrix
E	edges set
G	graph
K	degree matrix
L	Laplacian matrix
V	vertices/nodes set

Parameters and variables

b_r	branching rate
C	average clustering coefficient
C_i	clustering coefficient of node i
circuit_id	circuit identification number
$d_{r,x}$	depth of node x from the root node r
D_f	fractal dimension
d_{max}	maximum depth
e_{avg}	average edge length
k_i	degree of a node i
k_{avg}	average degree of the network
l	level of the network
l_{avg}	average path length
L_{total}	total network length
M	total number of edges
m_e	number of edges in the edge length range e
N	total number of nodes
n_k	number of nodes with degree equals to k
$P(k)$	degree distribution
$P(e)$	edge length distribution
r	root node
x, y	coordinates of the nodes
$start_x$, $start_y$	coordinates of a starting point of an edge
end_x, end_y	coordinates of an end point of an edge
ρ	Pearson correlation coefficient
ε	box size used in the box counting method
$\lambda_G(i)$	number of edges between the neighbours of node i
$\tau_G(i)$	total number of edges that can exist among the neighbours of node i

Low Voltage (LV) levels, the distribution networks are of increasing importance [1].

Much research has been carried out worldwide to analyse and quantify the impacts of LCTs on electricity distribution networks. Most of these studies are based on real network samples [2–4] standard synthetic networks such as the IEEE test cases [5,6] or other representative test networks [7,8]. As a result, most reported analysis in the literature is only useful for evaluating a specific test case and conclusions made from such studies have limited applicability to other networks. Moreover, a large number of new methodologies and algorithms have been proposed recently to overcome the operational challenges of the electrical power system with the integration of smart grids technologies. The effectiveness of these methodologies were mostly tested or implemented on one or a few specific test networks that are not yet able to provide robust and generalized conclusions [9–11].

However, there is a clear need of providing generalized and robust conclusions on network studies for strategic decision-making and policy support. For example, it is important for policy making to have the ability of characterizing and quantifying how differently the urban networks and rural networks perform with different integration levels of a new LCT. A network modelling and simulation platform with the ability of providing statistically-similar realistic models of electricity distribution networks, can be used to make such generalized conclusions through a large number of simulation studies.

However, it is often difficult to produce a large amount of random, realistic models for most of the real-world networks including the electricity distribution networks. Identifying and quantifying the important statistical properties of different types of distribution networks is a key requirement when developing such random, realistic network models. Statistical properties of electric power networks can be categorized into two groups: topological properties and electrical properties. Both electrical and topological properties of electricity distribution networks have a significant impact on their overall network performance including voltage drops, power losses, network reliability and costs, etc. This paper focuses on investigating the topological properties of the real-world electricity distribution networks.

Statistical properties of power networks have been studied by many researchers. In the past, the interest of studying the statistical properties, primarily the topological properties of electrical power grid was mainly led by the major blackouts happened in North America [12,13], Italy [14], Europe [15] and few other power grids [16]. After these large-scale blackouts happened worldwide, researchers were seeking solutions for improving security and reliability of the power grid from different perspectives. Ref. [17] provides an overview of the security and privacy issues in Smart Grids and [18] developed a new control method for nonlinear dynamics of the power systems. Advances in statistical physics and complex network theory together with graph theory applications also have developed new areas of interest in vulnerability assessment in power systems.

Much literature is summarized in the recent survey conducted by G. A. Pagani on the review on ‘power grids as complex networks’ [19]. Majority of the studies focused on the HV networks and the work carried out in the MV and LV level are very limited. The main goal of many HV network studies was to find out the connection between the structure of the power grid and the risk probability based on well-known complex network models [20,21] such as random graph model [22], small-world network models [23] and scale-free network models [24]. The structure of HV grids is different from that of MV and LV grids. The HV transmission and sub-transmission is usually a meshed system, but distribution networks (MV and LV) are mainly with radial structures. Therefore, the research findings in the HV network analysis cannot be directly used in MV and LV networks.

In addition, a few new network models have been developed in the past decade some of them with some efforts in investigating the statistical properties of the power grid [25–30]. It was observed that the electricity distribution networks consist of self-repeating patterns, which is a key property in fractal structures. Fractal based models have been proposed for the power grid [26,29]. However, the main objective of these new developments was different from the objectives of previously discussed HV network studies, and they focused on generating test cases to evaluate smart grid technologies. Using a fractal based software tool, a statistical approach was introduced to investigate the technical and cost performance of alternative network design strategies for electricity distribution networks under different development scenarios [29]. Different ranges of the fractal parameter for different types of distribution networks have been used in their work. A complex networks based model for the integration of distributed energy systems in urban areas was developed in [30]. A simulation platform was developed that generates electrical power grid test cases with realistic topologies and electrical parameter settings, while providing an underlying statistical investigation for the statistical properties of real and synthetic power grids [25]. In [27], the need for generating synthetic graphs for electrical power networks was identified and a methodology was developed to produce random, realistic models for power networks. However these work [25,27] also fall into the HV level.

A comprehensive statistical study on the distribution (MV and LV) level of the power grid was carried out in [22], using a lot of real world data. In their work, a novel analysis of the power grids using statistical tools from the complex networks analysis (CNA) field is presented and the study focused on a methodology of integrating topological metrics with economic factors.

While all above-mentioned research demonstrates the value of investigating topological properties and generating random realistic network models of real world power networks, most of them were carried out for the HV level. Also, majority of the previous studies were supported by a limited set of real network data or a limited set of topological properties of the real networks were analysed. Hence, a wider statistical analysis supported by a large amount of real world data is required for electricity distribution networks.

The motivation behind this study comes from finding out answers to the following two questions; (i) what are the key topological properties that characterize the realistic nature of different types of electricity distribution networks? (ii) can we find an approach to efficiently generate the ensembles of random but realistic topologies similar to the real electricity distribution networks in order to conduct a large amount of simulation studies? The main contributions of this paper include: (1) identifying the key topological properties that are useful to characterize the topological structures of real world MV distribution networks; (2) quantifying the topological properties of real world MV urban and sub-urban networks using graph theory and complex networks analysis techniques; and (3) introducing a novel, depth dependent approach to better characterize the topological properties of the MV distribution networks. Fig. 1 shows the schematic overview of this paper.

The organization of this paper is as follows. Section 2 summarizes the definitions and formulations of the key statistics used to investigate the topological properties of electricity distribution networks. A thorough topological investigation of real distribution networks at MV level is presented and the results are discussed in Section 3. Section 4 provides the conclusions followed by some suggestions for future work.

2. Properties for investigation of network topologies

In order to understand and model a real world complex network such as electrical power grid, a right set of tools and techniques are required. Most widely used techniques and tools are coming from the fields of CNA and graph theory. Ignoring the 3-Dimensional features such as very tall buildings and elevation of the equipment, the electrical power grid can be considered as a 2-Dimensional grid composed of various elements such as transmission lines, distribution transformers and switchgears. A graph model can be easily constructed by taking into consideration the relationship between these elements. This section describes some of the essential definitions and formulations of the key statistics used in the topological investigation of the distribution networks. In parallel, the physical meanings of the graph related measures with reference to the electrical power networks are also described in this section. Fig. 2 summarizes the fundamental approach and the key topological features used in this study.

2.1. Graph properties

2.1.1. Basic graph properties

A graph G consists of a collection of vertices V (nodes) and a collection of edges E: $G = (V, E)$. With reference to an electricity distribution network, V-nodes set include substations, distribution transformers, switches, busbars and consumer locations. Edges set E, stands for the physical connections between the nodes through underground (UG) cables and overhead (OH) distribution line segments. Since the collected real network data includes the actual geographical location of the electrical components, the information about nodes are extracted with the x, y coordinates. The edges are represented using the start and end x, y coordinates of the nodes. For an electrical power network with N nodes and M edges,

$$V = [x, y]_{N\times 2} \tag{1}$$

$$E = [start_x, stary_y, end_x, end_y, circuit\_id]_{M\times 5} \tag{2}$$

In real networks, very often two or more circuits share the same towers. In that case, to distinguish them a fifth dimension called *circuit_id* is added to E.

Connectivity between nodes in the graph is represented by an Adjacency Matrix 'A' using the unique node identifiers. As the electrical power grid can be considered as an undirected graph, the adjacency matrix of a power grid becomes a symmetric $N \times N$ matrix. The element A_{ij} becomes 1 if there exists a link between nodes i and j, otherwise A_{ij} equals to 0 (3).

$$A_{ij} = \begin{cases} 1, & if(i,j) \in E \\ 0, & if(i,j) \notin E \end{cases} \tag{3}$$

The distance matrix 'D' is defined for a graph using the edge lengths between nodes. The element D_{ij} becomes the length of the edge e_{ij} if there exists an edge between nodes i and j. Otherwise D_{ij} is equal to zero.

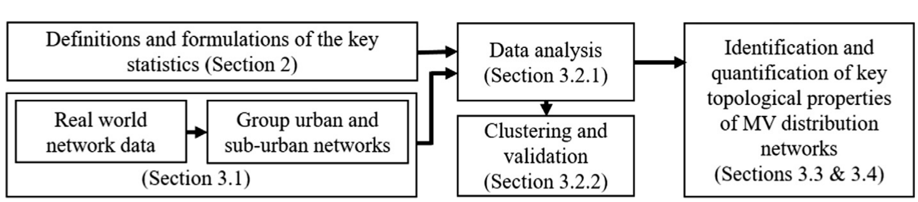


Fig. 1. Schematic overview of the study.

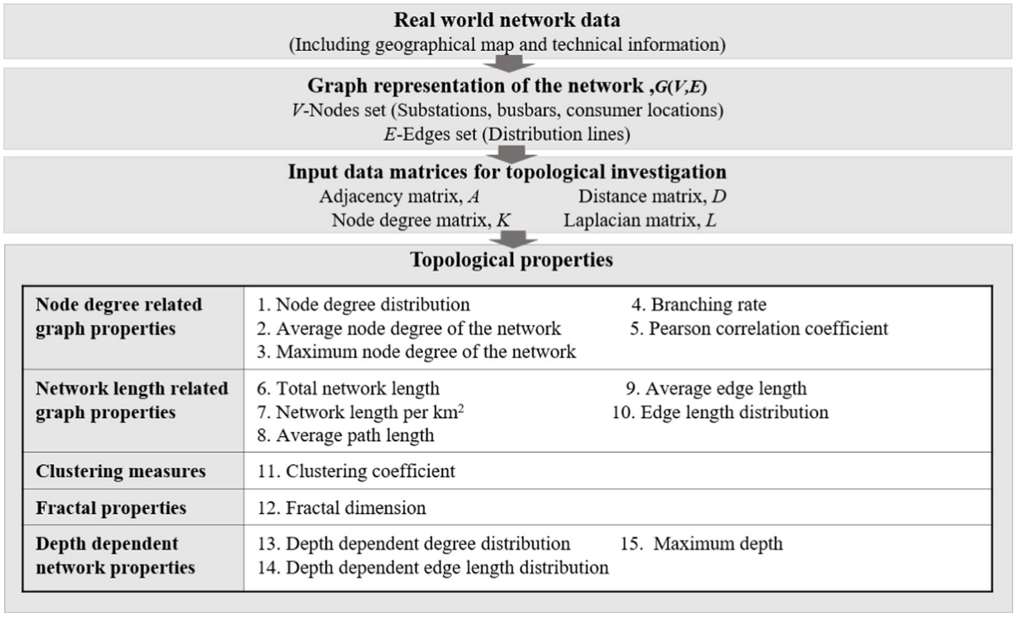


Fig. 2. Feature extraction.

$$D_{ij} = \begin{cases} \surd\{V(i,1)-V(j,1)\}^2+\{V(i,2)-V(j,2)\}^2, & if\,(i,j)\in E \\ 0, & if\,(i,j)\notin E \end{cases} \tag{4}$$

The degree of a node i in a graph which is denoted as k_i, is the number of edges incident to that node and is obtained using the adjacency matrix. The values obtained for the node degrees are used to construct the Degree Matrix 'K' which is a $N \times N$ diagonal matrix.

$$k_i = \sum_{j=1}^{N} A_{ij} \tag{5}$$

$$K_{ij} = \begin{cases} k_i, & if\ i=j \\ 0, & otherwise \end{cases} \tag{6}$$

Laplacian matrix 'L' is also useful in obtaining the graph properties of the power networks which is written as,

$$L = K - A \tag{7}$$

A connected component in graph theory refers to a set of vertices in a graph that are linked to each other by paths. For instance, a radial 10 kV network supplied by one 33 kV/10 kV substation can be considered as one connected component (ignoring the connections to the main grid from the 33 kV side of the main supply point). According to this definition, the number of connected components in a radial electricity distribution network of a certain voltage level (in a given area) is equal to the number of main grid supply points in the network. From the graph theory definitions, the number of connected components in a graph is equal to the number of times 0 appears as an eigenvalue in the Laplacian matrix of the graph.

Most of the graph-related properties of the networks described below are derived using the above basic graph properties *A, D, K,* and *L*.

2.1.2. Node degree related graph properties

Degrees can be used to identify the key components in a network. In power networks, nodes such as substations have a high node degree compared to the other nodes. Also, degrees, and notably degree distributions can be used to derive information on the structure of a network. For example, if most vertex degrees are the same the network is more or less a regular network in which vertices have equal roles. The degree distribution $P(k)$ of a network is defined as the fraction of nodes in the network with degree k. If the total number of nodes in the network is N and n_k of them have degree k, $P(k)$ is defined in Eq. (8).

$$P(k) = \frac{n_k}{N} \tag{8}$$

Researches have observed that the HV electrical transmission grids have heavy-tailed degree distributions. A network is said to be scale-free when the degree distribution of the network follows a power law, resulting in few nodes having many edges and many nodes having few edges [31]. The power law relationship between $P(k)$ and k is as shown in Eq. (9) where, γ is a parameter whose value is typically ≥ 1.

$$P(k) \propto k^{-\gamma} \tag{9}$$

In vulnerability assessment studies on the HV transmission grid, this scale free property has been broadly analysed. However, for weakly meshed and radial networks in distribution level a power law distribution for $P(k)$ has not been observed. But still it is worth to observe the behaviour of $P(k)$ for similar types of distribution level networks. The average node degree k_{avg}, of a graph G is also an important measure about the structure of the network. If $k_{avg} > 2$ the network has a meshed structure.

$$k_{avg} = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2M}{N} \tag{10}$$

Another node degree related measure is the branching rate b_r, which gives an indication of how much a given network tends to branch out. For instance, the urban distribution networks tend to branch out more compared to the rural distribution networks. Branching rates are usually different for different types of networks (rural, urban), for different voltage levels, and for different locations, i.e. close to the supply points or close to the customer points.

$$b_r = \frac{\text{Number of nodes with degree} \geqslant 3}{\text{Total number of node sin the network}} \tag{11}$$

A network is said to show assortative mixing if there exists a correlation between nodes of similar degree. Assortativity property of the networks is examined in terms of node degrees using the Pearson correlation coefficient ρ:

$$\rho = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2} \tag{12}$$

where j_i, k_i are the degrees of the vertices at the ends of the ith edge, with $i = 1, \ldots, M$. A positive value for ρ indicates correlation between nodes with similar degree and negative values for ρ indicates the relationship between nodes of different degree which is called disassortative mixing [32].

2.1.3. Network length related graph properties

In power system terminology, network length of an MV distribution network refers to the total length of the overhead and underground electricity distribution line segments. The network length is a critical parameter for electrical power networks since it impacts on a number of technical and economic factors such as the voltage drops, power losses and the cost of cables/overhead lines in the network. It is also important when describing the network topology realistically. An edge can be referred to a feeder section in an electrical distribution network. Total network length L_{total}, average edge length e_{avg} that is obtained by dividing the total network length by the total number of edges, and the average path length l_{avg}, are some of these length related network measures.

A path is a sequence of edges from one vertex to another. Length of a path is the addition of the lengths of all the edges in the path. The geodesic distance between nodes v_1 and v_2, denoted as $l(v_1, v_2)$ is the length of the shortest path between v_1 and v_2. Diameter of a network is defined as the longest graph geodesic between any two graph vertices v_1, v_2 of a graph. Then the average path length is defined as the average length along the shortest paths for all possible pairs of network nodes. It has been widely used as a measure of the efficiency of information or mass transport on a network.

$$l_{avg} = \frac{1}{N(N-1)}\sum_{i \neq j} l(v_i, v_j) \tag{13}$$

Similar to the degree distribution, the edge length distribution $P(e)$ of a network is defined as the fraction of edges in the network, with length e. Here, length e represents a range of edge lengths. If the total number of edges in the network is M and m_e of them fall into the length range e, $P(e)$ is expressed as:

$$P(e) = \frac{m_e}{M} \tag{14}$$

Some network models assume equal spacing between nodes. However, this is not true for most of the real networks. Therefore, the edge length distribution is an important feature to be considered in geographical network modelling.

2.1.4. Clustering of the nodes

In graph theory, clustering coefficient C is a measure of the degree to which nodes in a graph tend to cluster together. The clustering coefficient is defined as the average of the clustering coefficient for each node C_i [15]. C_i is expressed as the ratio of number of edges between the neighbours of node i, $\lambda_G(i)$ to the total number of edges that can exist among neighbours of node i, $\tau_G(i)$ [23]. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to cluster into groups. However, clustering coefficient of a radial network is zero. Therefore, in order to analyse the MV and LV radial networks this measure is not very useful. But in HV network studies this measure has been widely used.

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{15}$$

$$C_i = \frac{\lambda_G(i)}{\tau_G(i)} \tag{16}$$

2.1.5. Fractal properties

A few researchers have observed that the real world electricity distribution networks consist of self-repeating patterns across all scales [26]. This property is called the self-similarity. Self-similarity is a typical property of fractals. Fractal dimension is another important property of fractals that provides a statistical index of complexity of a fractal pattern with the scale at which it is measured. It has also been characterized as a measure of the space-filling capacity. If this property of a fractal needs to be explained, the box counting method provides the practical solution for that. In this method, the fractal object (e.g. image of the network layout) is covered with boxes with equal sides ε, and find how the number of boxes $n(\varepsilon)$ which include the fractal object changes with the box size [33]. A network is said to be fractal if the box counting dimension exists for that network. The box-counting dimension D_f is defined as:

$$D_f = lim_{\varepsilon \to 0} \frac{\log n(\varepsilon)}{\log\left(\frac{1}{\varepsilon}\right)} \tag{17}$$

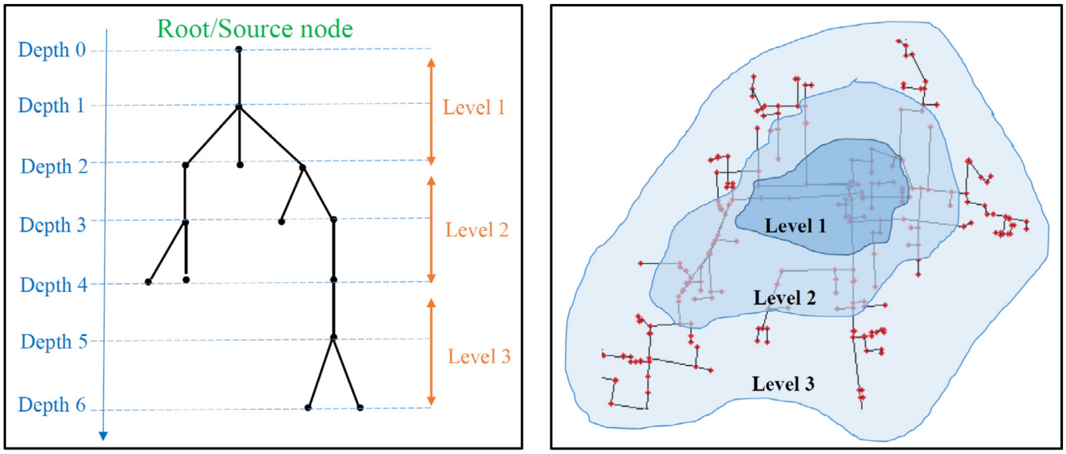
2.2. Depth property

Analysing the depth dependent network properties is an effective approach in both feature identification and the network model development. A simple connected graph with no cycles is called a tree. Mostly the electricity distribution networks at MV and LV levels have radial or tree like structures. The depth d, of a node x, in a tree is defined as the number of edges n_e, from the root node r, to the node x (Fig. 3(a)).

$$d_{r,x} = n_e(r, x) \tag{18}$$

It was observed that the power networks have considerably different graph related properties at different depths of the networks. In this case, according to the depth of the nodes network properties are observed; closer to the supply point (level 1), at the middle level (level 2) and at the furthest away area from the supply point (level 3) respectively. The different types (rural/urban, MV/LV) of networks usually have different depth dependent properties. Electrical power grid is an evolving network, with new nodes and edges added with time. Similar to most of the real-world networks, the development and evolution of the electrical power network is closely defined by the factors such as geographical environment, population distribution, social and economic development. Due to these factors, different networks may have observable topological differences. For example, in rural networks consumer locations tend to aggregate in a more clustered fashion with large open areas dedicated to farms and green spaces, while in urban networks consumer locations are usually evenly distributed. The distribution of consumer settlements also defines the distribution of feeder lengths.

The depth of a node from the given root node can be obtained using Dijkstra shortest path algorithm [34]. The HV/MV substation is considered as the root of a MV radial network. Maximum depth d_{max} is the number of edges along the longest path from the root node down to the farthest leaf node. The network is divided into several levels along the depths as shown in Fig. 3(a). The number of levels chosen, can be varied. In this paper, the network is divided



(a) (b)

Fig. 3. (a) The concept of depth of a node. (b) The idea of the levels of a network along the depth.

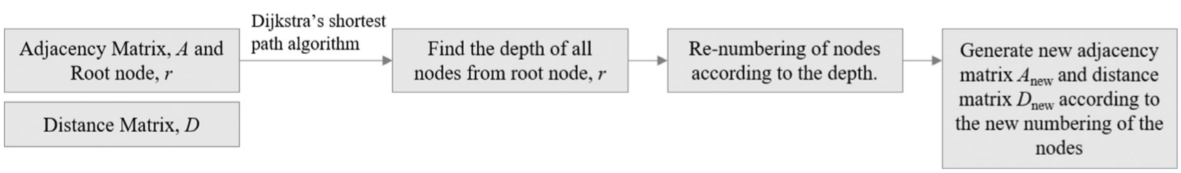


Fig. 4. Algorithm to obtain depth dependent network properties.

into three levels l_1, l_2 and l_3 along the depth. Fig. 3(b) illustrates segmenting a radial network into levels.

$$l = \begin{cases} l_1; & 0 < d \leqslant \frac{d_{max}}{3} \\ l_2; & \frac{d_{max}}{3} < d \leqslant \frac{2d_{max}}{3} \\ l_3; & \frac{2d_{max}}{3} < d \leqslant d_{max} \end{cases} \tag{19}$$

In order to obtain the depth dependent properties for a tree like graph, the adjacency 'A' and distance 'D' matrices can be re-organized as 'A_{new}' and 'D_{new}' following the new node identifiers given according to the depth of the node (Fig. 4). For example, the root/source node is now numbered as 'node 1' and the nodes immediately connected to the root node takes the next consecutive numbers for their node identifiers. Submatrices of the A_{new} and D_{new} are used to derive the depth dependent degree and edge length distributions of the network. For example, the edge length distribution of the network at level 1 (l_1) is obtained from the values in submatrix D_{l1} of D_{new}, using the basic definition in (14). Similarly, the values in submatrices D_{l2} and D_{l3} are used to obtain the edge length distributions of level 2 and level 3 respectively. The idea is illustrated in Fig. 5.

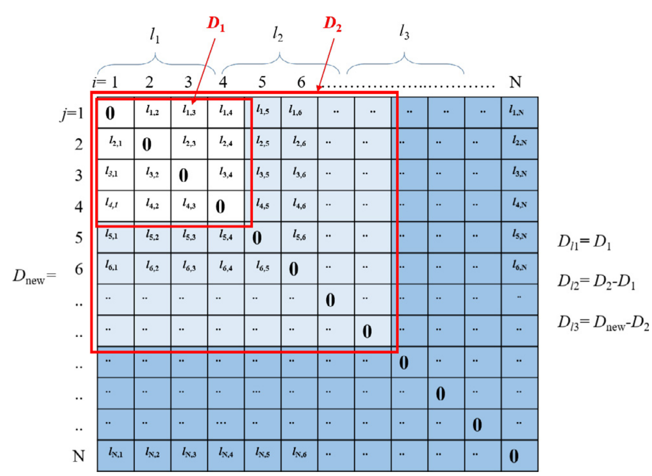


Fig. 5. New distance matrix (node numbers 1 to N are given according to the depth of the node from its root node).

3. Topological properties of the real-world electricity distribution networks

3.1. Power grid data for topological investigation

An investigation of the topological properties of the real-world networks was conducted using real network data that was collected from China, covering urban and sub-urban areas. These data include the detailed technical and geographical information of transmission, sub-transmission, and distribution level networks and also the population data of the supplied areas of the networks. Since the present work is mainly focused on MV level networks, only the 10 kV level network information was extracted. All selected networks have a radial structure.

A graph representation of each radial network component was obtained. A graph representation obtained from the geographical layout of Network 1 under study, is shown in Fig. 6(a). The red[1] nodes represent the 10 kV level consumers, 10 kV/400 V distribution transformers and busbars. The green circle represents the main grid supply point of the 10 kV network.

Table 1 summarizes the basic information available for the selected 30 networks at the 10 kV level. The networks were

[1] For interpretation of color in Figs. 6 and 10, the reader is referred to the web version of this article.

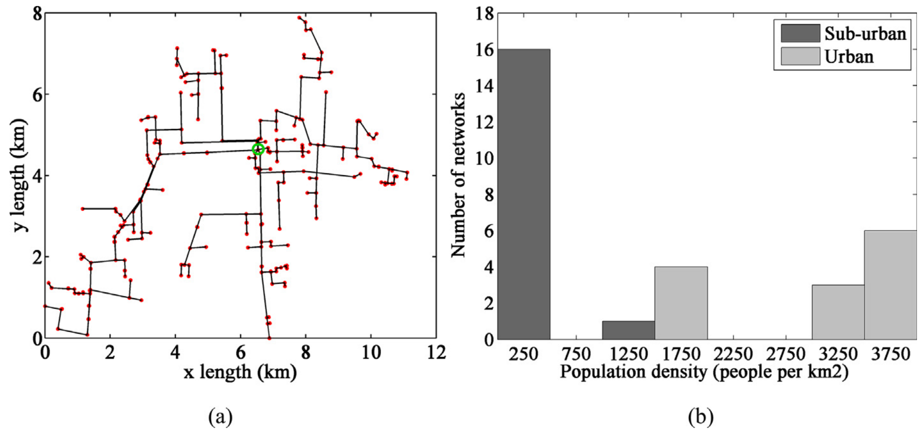


Fig. 6. (a) Graph representation of network 1 (b) Distribution of population densities of the networks.

Table 1
Basic network information of 30 networks at 10 kV level.

Network ID	Area (km^2)	Population density (/km^2)	Total network length (km)	Number of nodes	Number of edges	Network ID	Area (km^2)	Population density (/km^2)	Total network length (km)	Number of nodes	Number of edges
1	66.2	482	76.4	254	253	**16**	101	256	103	377	376
2	142	328	88.5	399	398	**17**	44.3	1054	84.1	384	383
3	64.5	405	64.2	285	284	**18**	9.8	1939	28.3	234	233
4	86.2	377	80.6	350	349	**19**	9.2	1935	29.6	237	236
5	33.8	376	35.5	136	135	**20**	10	1930	34.2	331	330
6	108.5	256	78.8	204	203	**21**	8	1750	34.2	205	204
7	86.2	229	66.8	226	225	**22**	14	3200	38	328	327
8	93.8	228	81.5	246	245	**23**	16	3200	52.5	400	399
9	69	464	90.6	308	307	**24**	14	3200	31.2	267	266
10	44.3	361	32.1	100	99	**25**	7	3600	28.7	227	226
11	78.5	369	61.1	196	195	**26**	7.5	3600	29	217	216
12	84.1	473	87.3	321	320	**27**	7	3600	17.5	115	114
13	69.1	411	71.6	220	219	**28**	7.5	3600	14.7	114	113
14	80.9	449	54.2	173	172	**29**	7.5	3600	16.1	153	152
15	125.4	364	95.6	351	350	**30**	8	3600	30.4	265	264

categorized as sub-urban and urban depending on the population density. According to the Demographic Yearbook 2013 by United Nations the definition of 'urban' for the cities in China is defined as the areas with population density higher than 1500 people per square kilometre [35]. The distribution of the population density of the networks under study is shown in Fig. 6(b).

According to Fig. 6(b) out of 30 10 kV networks, 17 networks are in the sub-urban category while the other 13 networks fall into the urban category.

3.2. Topological analysis

3.2.1. Quantification of the topological properties

Table 2 presents the results of the basic topological investigation of the above 30 networks at 10 kV level.

In order to compare the results in Table 2, probability distributions of the topological properties of both sub-urban and urban networks were obtained. Fig. 7 shows the comparative probability distribution plots for the two types of networks, arranged back to back on the *x*-axis (probability of occurrence). For one topological property, the same bin size and the same number of bins were used to generate the probability distributions of both types of networks.

It can be observed from Fig. 7 (sub-graphs with letter A) that, some of the topological properties such as, nodes per km^2, network length per km^2, average edge length and average path length are able to clearly characterize the topological differences of the two types of networks. The ranges of the variation of the above four properties are summarised in Table 3.

Even though the probability distributions of branching rate, maximum node degree, fractal dimension and maximum depth have some noticeable differences (Probability distribution plots of the two network types are biased into different directions) of the two types of networks, still there are some overlapping of the values (sub-graphs with letter B). According to Fig. 7 (sub-graphs with letter C), properties such as Average node degree and Pearson correlation coefficient do not give clear information to characterize urban and sub-urban networks.

3.2.2. Validation using clustering

Clustering is defined as the grouping of similar objects. A simple procedure to classify the data set in Table 2 through a number of clusters was carried out using the *k*-means clustering algorithm. The *k*-means clustering aims to partition a number of observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. Each observation is a *d*-dimensional real vector. Euclidian distances are used to calculate the distance from the observation to the mean. In this study, it was assumed that the number of clusters are known. The implementation of the *k*-means algorithm is as follows; (1) make initial guesses for the means m_1, m_2, ... m_k. (2) use the estimated means to classify the samples into clusters. (3) for *i* from 1 to *k*, replace m_i with the mean of all of the samples for cluster *i* (4) repeat steps 2 and 3 until there are no changes in any mean [36,37].

Fig. 8 illustrates the *k*-means cluster analysis procedure of the present study. According to the above explanation of *k*-means clustering the 30 networks used in the study represent 30 observa-

Table 2
Graph related properties of the 30 networks at 10 kV level.

Network ID	Number of nodes per 1 km^2	Network length per 1 km^2 (km)	Average node degree	Branching rate	Maximum node degree	Pearson correlation coefficient	Average edge length (km)	Average path length (km)	Fractal dimension	Max depth
1	4	1.154	1.992	0.324	7	0.534	0.302	7.408	1.301	28
2	3	0.623	1.995	0.313	5	0.559	0.222	10.356	1.270	54
3	4	0.995	1.993	0.365	4	0.457	0.226	7.551	1.275	35
4	4	0.934	1.994	0.377	7	0.516	0.231	6.445	1.331	28
5	4	1.052	1.985	0.346	4	0.421	0.263	5.404	1.266	23
6	2	0.726	1.990	0.363	5	0.504	0.388	8.418	1.261	26
7	3	0.775	1.991	0.292	3	0.776	0.297	8.518	1.273	34
8	3	0.868	1.992	0.350	4	0.356	0.332	9.694	1.276	35
9	4	1.316	1.994	0.361	5	0.477	0.295	8.179	1.323	38
10	2	0.724	1.980	0.330	3	0.394	0.324	4.546	1.228	25
11	2	0.779	1.990	0.352	4	0.597	0.313	7.625	1.270	24
12	4	1.038	1.994	0.336	6	0.616	0.273	7.983	1.319	35
13	3	1.036	1.991	0.341	3	0.600	0.327	9.840	1.285	38
14	2	0.670	1.988	0.358	4	0.353	0.315	6.513	1.266	27
15	3	0.762	1.994	0.288	7	0.475	0.273	9.792	1.296	51
16	4	1.020	1.995	0.358	8	0.612	0.274	9.452	1.319	37
17	9	1.899	1.995	0.299	9	0.363	0.220	6.820	1.344	36
18	24	2.885	1.991	0.350	7	0.538	0.121	2.879	1.304	23
19	26	3.222	1.992	0.371	7	0.682	0.126	3.388	1.302	38
20	33	3.418	1.994	0.390	7	0.606	0.104	2.641	1.336	30
21	26	4.275	1.990	0.263	9	0.502	0.168	3.738	1.281	22
22	23	2.711	1.994	0.348	8	0.499	0.116	3.788	1.324	45
23	25	3.281	1.995	0.370	8	0.678	0.132	4.370	1.341	29
24	19	2.230	1.993	0.397	9	0.101	0.117	3.415	1.352	33
25	30	3.828	1.991	0.370	6	0.633	0.127	4.494	1.353	40
26	29	3.867	1.991	0.373	6	0.580	0.134	3.578	1.341	29
27	15	2.336	1.983	0.365	7	0.390	0.154	2.583	1.327	20
28	15	1.960	1.982	0.272	6	0.491	0.130	2.232	1.319	20
29	20	2.152	1.987	0.366	7	0.542	0.106	2.053	1.369	31
30	35	4.048	1.992	0.343	10	0.226	0.115	2.507	1.344	21

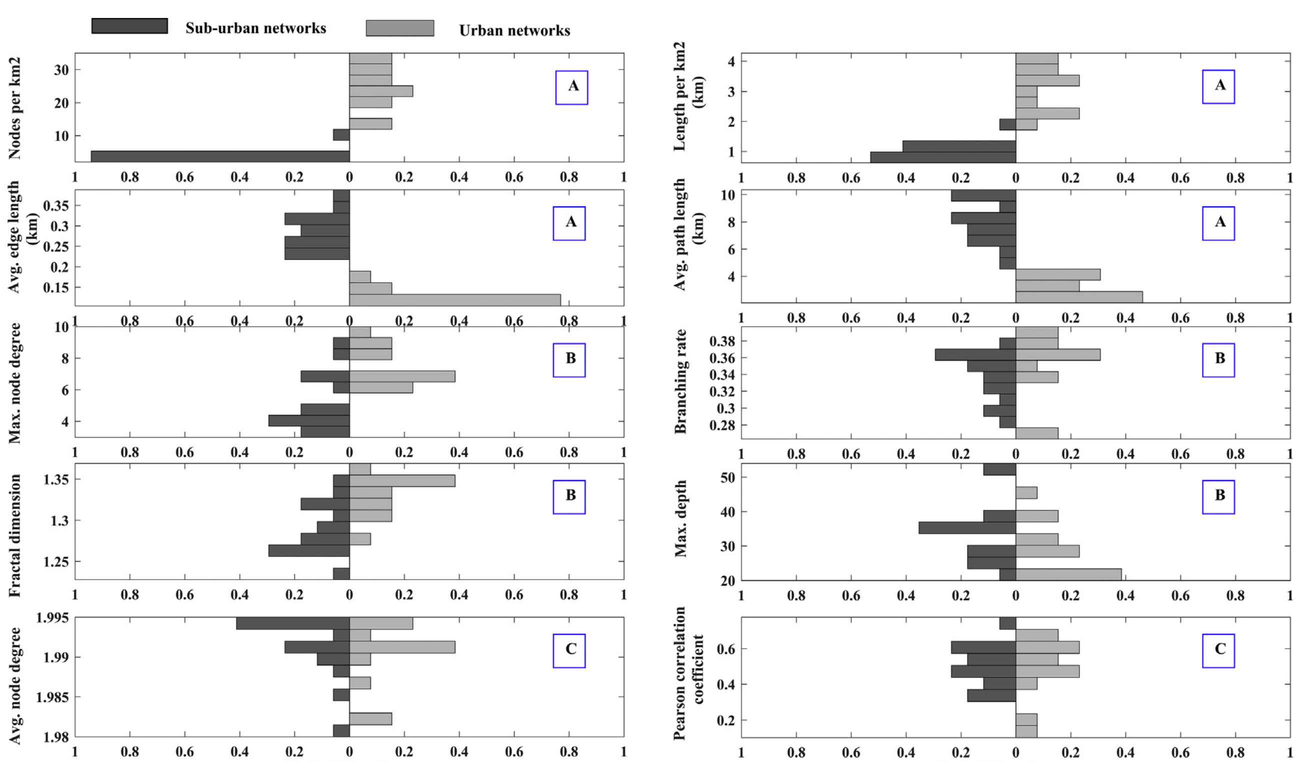


Fig. 7. Comparison of the topological properties of urban and sub-urban networks.

tions, and each observation is a 10-dimensional real vector. The 10 dimensions (cluster variables) are the 10 topological properties listed in Table 2. The input data matrix for the *k*-means algorithm was formed using different subsets of the properties from Table 2 to find out which subsets of the parameters together can effectively characterize the two network types (urban and sub-urban).

Table 3
The ranges of variations of the topological properties of sub-urban and urban distribution networks.

Topological property	The range of variation	
	Sub-urban networks	Urban networks
Nodes per km^2	2–9	15–35
Network length per km^2	0.6–2.0 km	2.0–4.2 km
Average edge length	0.2–0.4 km	0.1–0.2 km
Average path length	4.5–10.5 km	2.0–4.5 km

Therefore, it was assumed that the number of clusters k is known for the data set ($k = 2$). Two clusters are the urban and sub-urban networks. Then, the k-means algorithm was used to group the 30 networks into two clusters. Trial 10 and trial 3 were used as examples for the discussion of the clustering results in this section. In the trial 10 all the 10 topological parameters in Table 2 were used as cluster variables. The grouping done by the clustering in trial 10 exactly followed the sub-urban and urban classification done by the population density parameter. Hence trial 10 was used to validate the urban, sub-urban classification of the networks.

The results shown in Fig. 9 are graphical representations of the case where, different subsets with 3 parameters were chosen as the cluster variables (trial 3). In the first two cases (Fig. 9 (a) and (b)), the selected sets of cluster variables were able to group the network sample into two clusters accurately, as defined by the population density of the networks. However, the third set of cluster variables shown in Fig. 9(c) did not cluster the network sample into the right groups. Some of the sub-urban type networks were fallen into the urban category and also the data points in each cluster seemed to be much more dispersed that the previous two cases. This observation explains the importance of feature selection when characterizing different network types.

3.2.3. Identification of key topological features

From the results so far, it is evident that the node degree and edge length related topological measures are a key to characterize different types of networks (most of the above listed topological properties are related with node degrees and edge lengths). Therefore, apart from using the single value properties to express degree and length related graph properties of the networks (Table 2), investigating the degree distributions and edge length distributions have clear benefits as they can also be used to capture most of the above single value properties.

Fig. 10(a) and (b) show the edge length distributions of sub-urban and urban networks respectively. Light blue and light red curves in both figures represent the edge length distribution of a single network. The dashed dark blue and dark red lines show the average curves of the edge length distributions of all the networks in the corresponding figure. The average curve was obtained by taking into consideration the edge lengths in all the networks of one type as one set and by getting the probability of occurrences for the whole set.

The curves of average edge length distribution of sub-urban and urban types of networks are compared in Fig. 10(c). It was observed that edge length distributions of both types of networks follow negative exponential patterns. The edge length distribution of urban networks has a faster decay compared to the sub-urban networks and this observation explains that the urban networks have a considerably higher fraction of shorter edge lengths compared to the sub-urban networks.

Similarly, degree distributions of the sub-urban and urban networks are shown in Fig. 11(a) and (b) respectively. The average degree distribution curves for the two types of the networks are compared in Fig. 11(c). The curves do not follow any well-known distribution. However, it is noticeable that the number of nodes with a degree 2 in most of the urban networks is less than the

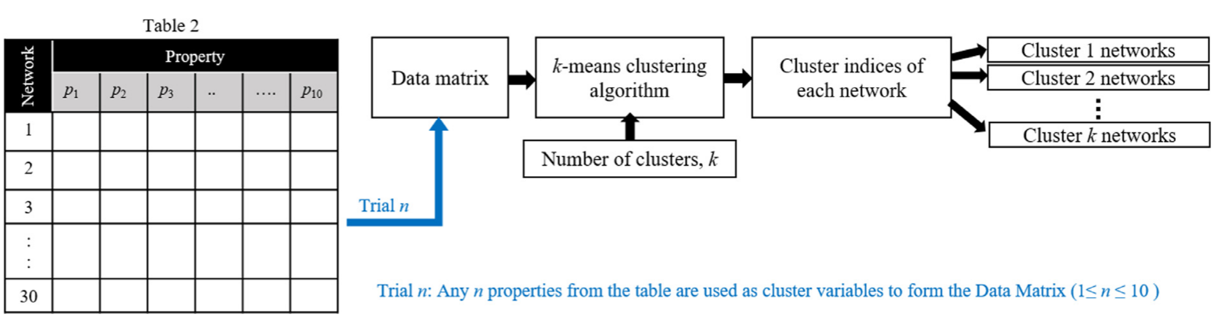


Fig. 8. The k-means cluster analysis.

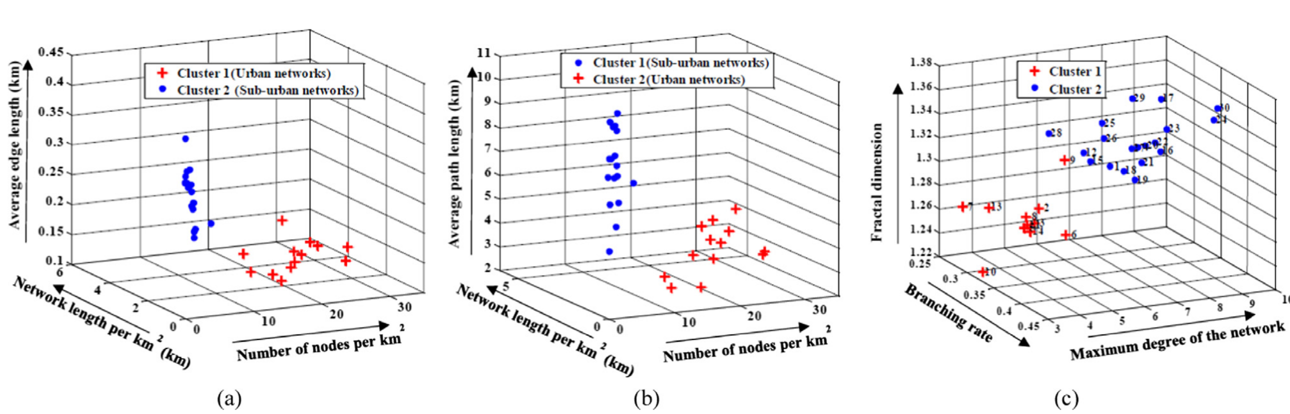


Fig. 9. Cluster assignments.

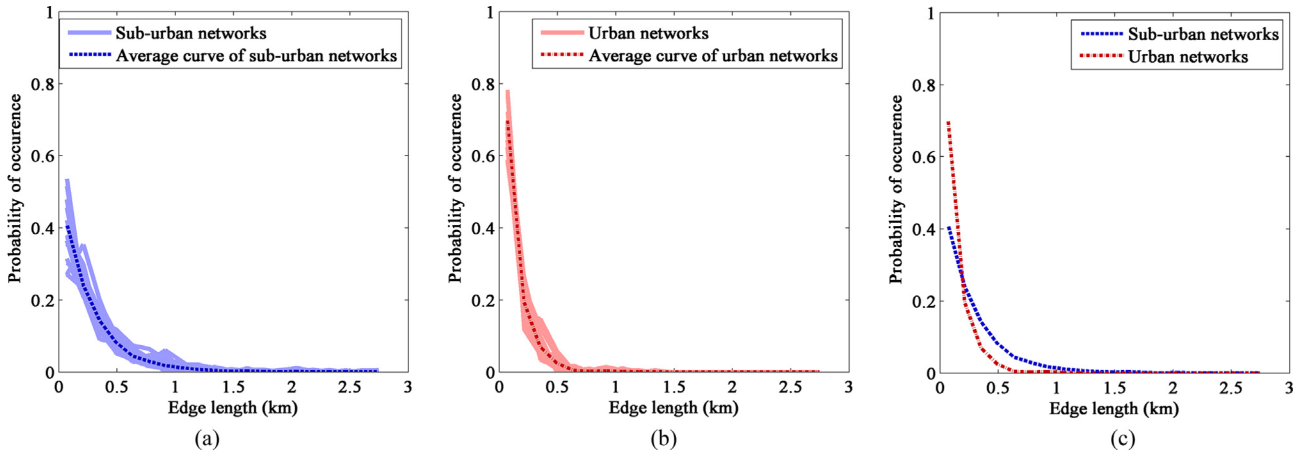


Fig. 10. Edge length distributions of; (a) sub-urban networks, (b) urban networks and (c) comparison of edge length distributions of sub-urban and urban networks.

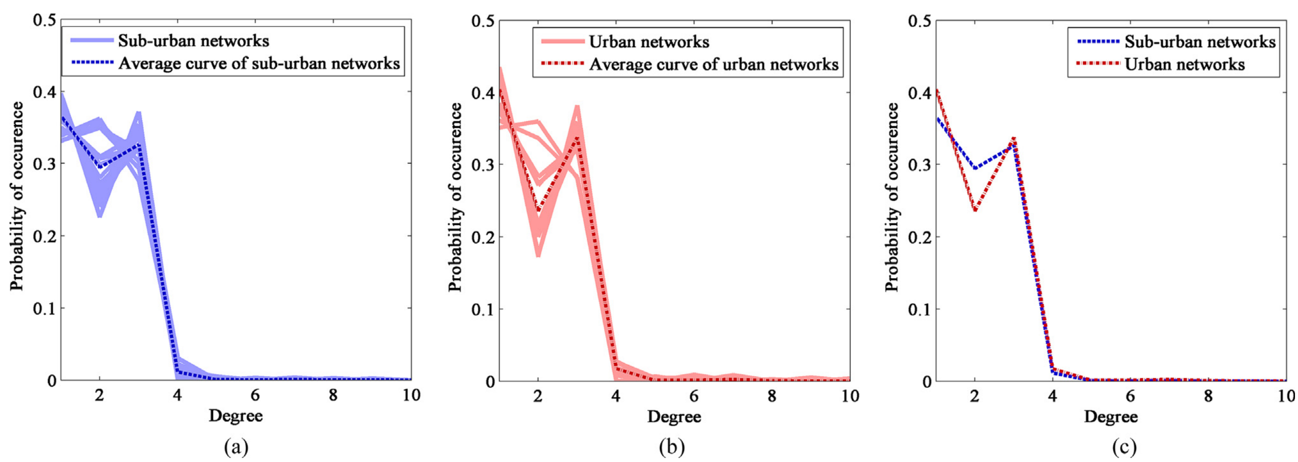


Fig. 11. Node degree distributions of; (a) sub-urban networks, (b) urban networks and (c) comparison of node degree distributions of sub-urban and urban networks.

sub-urban networks. This implies that the urban networks tend to have more branches (nodes with degree ≥ 3) and leaf nodes (nodes with degree 1) than sub-urban networks. Also, the maximum degree observed in sub-urban networks is 7 while maximum degree of the urban networks is up to 10.

3.3. The depth dependent topological properties

An investigation for the depth dependent topological properties was conducted with the same set of real world network data. Since the length and degree related measures play a critical role in describing the topology of a network, the depth dependent degree distributions and the edge length distributions were thoroughly investigated.

Fig. 12(a)–(c) show the results of depth dependent analysis of one sub-urban type network (Network 1 in Table 2). From Fig. 12 (a) it can be observed that in Network 1, the edge length distributions of all the three ‘levels’ approximately follow negative exponential distributions. However, the maximum edge length and the total number of edges in each level has been reduced when

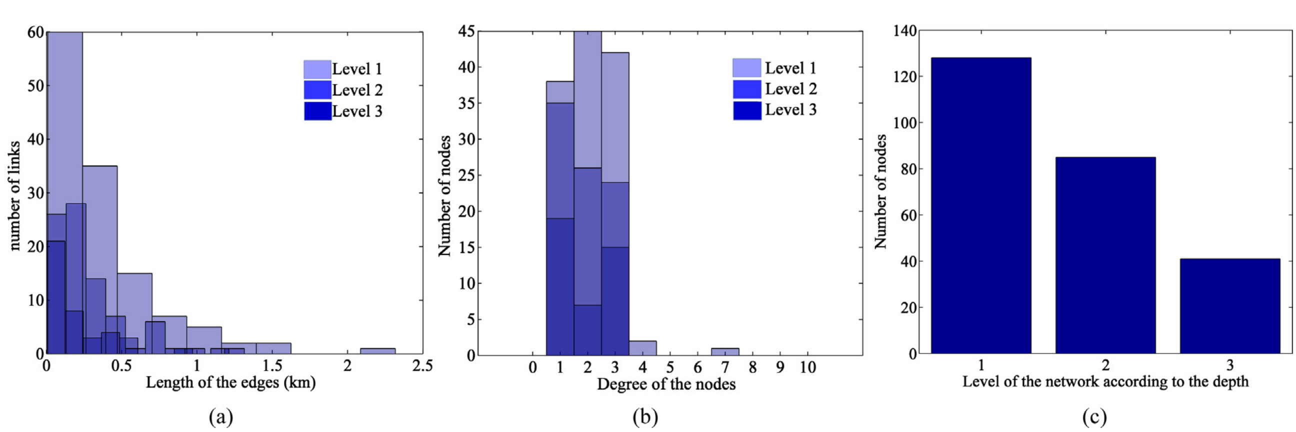


Fig. 12. (a) Depth dependent edge length distribution of Network 1. (b) Depth dependent degree distribution of Network 1. (c) Distribution of nodes among the levels of Network 1.

Table 4
Number of nodes in different levels of the networks.

Level	Number of nodes in different levels of sub-urban networks (Network IDs 1–17)																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Level 1	128	185	122	150	52	102	93	106	172	39	66	164	94	84	174	180	214
Level 2	85	159	109	130	53	80	74	100	107	38	86	117	80	69	104	136	130
Level 3	41	54	54	70	31	22	59	40	29	23	44	40	46	20	73	61	40
Total no. of nodes	254	398	285	350	136	204	226	246	308	100	196	321	220	173	351	377	384

Level	Number of nodes in different levels of urban networks (Network IDs 18–30)												
	18	19	20	21	22	23	24	25	26	28	27	29	30
Level 1	70	107	175	70	158	171	108	96	106	47	58	74	101
Level 2	96	100	95	99	121	175	112	89	88	44	38	48	121
Level 3	68	30	61	36	49	54	47	42	23	23	19	31	43
Total no. of nodes	234	237	331	205	328	400	267	227	217	114	115	153	265

going towards level 3 from the level 1 of the network. Fig. 12(b) shows the degree distributions of the three levels of the network 1. Fig. 12(c) shows the distribution of the nodes among the levels of the network 1. It was observed that the number of nodes in each level has been reduced when moving from level 1 to level 3 of the network.

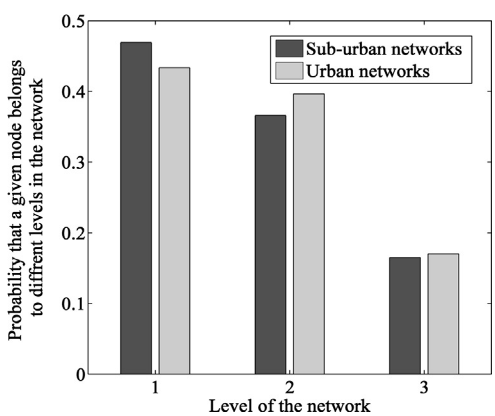


Fig. 13. Comparison of the distribution of the nodes among the levels of sub-urban and urban networks.

Table 4 provides the information regarding the distribution of the number of nodes in different levels of urban and sub-urban networks used in the study. Fig. 13 compares and summarizes the information in Table 4. For instance, the average probability that a given node in an urban network belongs to level 1 was obtained by dividing the total number of nodes in level 1 by the total number of nodes in all urban networks. From Fig. 13 it was identified that in both types of networks the number of nodes in each level has been reduced when going away from the source node. According to Fig. 13, the fraction of nodes in levels 2 and 3 of urban networks are slightly higher than that of the sub-urban networks. This is due to the higher density of the distribution of consumers in urban areas, compared to that of the sub-urban areas.

Fig. 14(a) shows a comparison of the edge length distributions of sub-urban and urban networks at their levels 1, 2 and 3. Each curve represents the average variation of all the networks of one type (e.g. Average variation of 17-sub-urban and average variation of 13-urban networks). When comparing the results with the sub-urban networks it can be observed that at the same level, the urban networks have faster decaying negative exponential pattern and a shorter 'maximum edge length' in their edge length distributions. Also for both types of networks, the maximum edge length that can be observed in level 1 has been reduced when going away from the source node towards level 2 and level 3 (Table 5), and the exponential decaying of the distribution has also become faster.

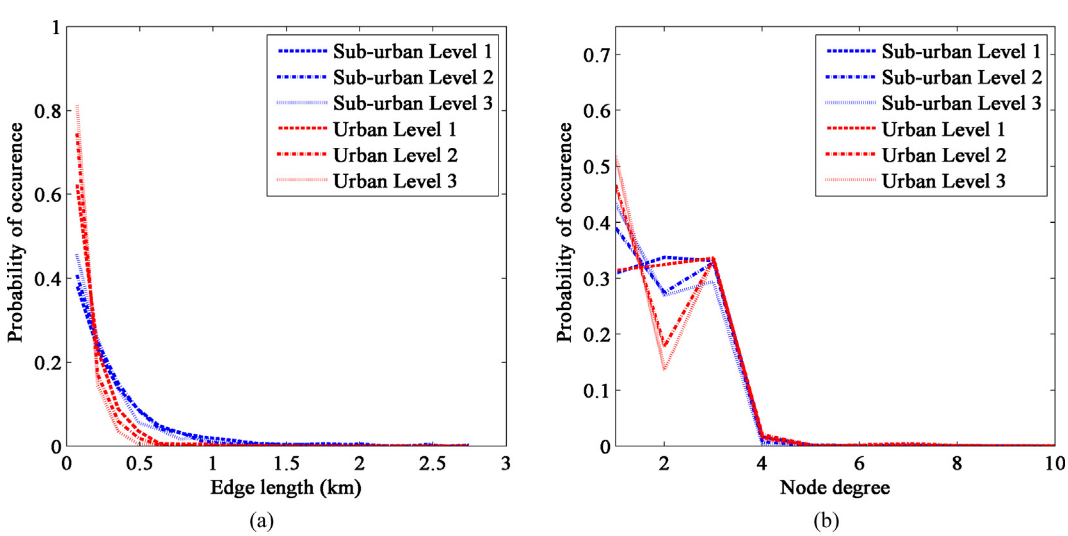


Fig. 14. (a) Comparison of edge length distributions of sub-urban and urban networks at different levels. (b) Comparison of degree distributions of sub-urban and urban networks at different levels.

Table 5
Maximum value for edge length in sub-urban and urban networks at different levels of the network.

Level	Maximum edge length (km)	
	Sub-urban networks	Urban networks
1	2.9	1.4
2	2.0	0.9
3	1.6	0.7

Similarly, a comparison of the depth dependent network analysis for the degree distributions is shown in Fig. 14(b). Compared to the information delivered by the degree distribution curves in Fig. 11, the depth dependent analysis provides detailed information regarding the network structure. It can be observed that in both types of networks, close to the supply point the network is less branched and when going away from the supply node branching (nodes with degree ≥ 3) and the fraction of leaf nodes (nodes with degree = 1) have increased. Comparison of the degree distribution of the same level in two types of networks shows that the urban networks tend to have a strong depiction of the above discussed property than that of the sub-urban networks.

4. Conclusions and future work

Investigating topological properties of real world electricity distribution networks have great benefits of providing random, realistic test network models for various simulation studies. In the present study an in-depth topological investigation on real world MV electrical power networks was presented.

Results of the real world network investigation showed that (1) node degree and edge length related graph properties are fundamental in characterizing the topological structures of the radial type sub-urban and urban electricity distribution networks; (2) results from the clustering approach showed the importance of feature selection when characterizing different types of distribution networks; (3) the depth dependent approach was able to better capture the topological features at different depth levels of the networks. Results from the depth dependent analysis showed that urban and sub-urban types of electricity distribution networks have different graph related properties at different depth levels of the networks.

Similar investigation will be conducted in future for MV rural networks and for LV networks of different types (urban, sub-urban, and rural). These results will be used to build a network generation tool which can produce statistically similar, random-realistic network topologies.

Acknowledgements

The authors would like to thank the Toshiba TRL and the Cardiff University's President Scholarship for their financial and technical support. This project is partially funded by EPSRC OPEN (EP/K006274/1), JUICE (EP/P003605/1) and MISTRAL (EP/N017064/1) projects.

Data Statement-"Information on the data underpinning the results presented here, including how to access them, can found in the Cardiff University data catalogue at http://doi.org/10.17035/d.2017.0040458448".

References

[1] Pavić I, Capuder T, Kuzle I. Low carbon technologies as providers of operational flexibility in future power systems. Appl Energy 2016;168:724–38.
[2] Fernandez LP, Román TGS, Cossent R, Domingo CM, Frías P. Assessment of the impact of plug-in electric vehicles on distribution networks. IEEE Trans Power Syst 2011;26:206–13.
[3] Chin Ho T, Chin Kim G. Impact of grid-connected residential PV systems on the malaysia low voltage distribution network. In: IEEE 7th international power engineering and optimization conference (PEOCO), 2013; 2013. p. 670–5.
[4] Abeysinghe S, Nistor S, Wu J, Sooriyabandara M. Impact of electrolysis on the connection of distributed generation. Energy Proc 2015;75:1159–64.
[5] Clement-Nyns K, Haesen E, Driesen J. The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. IEEE Trans Power Syst 2010;25:371–80.
[6] Fu X, Chen H, Cai R, Yang P. Optimal allocation and adaptive VAR control of PV-DG in distribution networks. Appl Energy 2015;137:173–82.
[7] Kabir MN, Mishra Y, Ledwich G, Xu Z, Bansal RC. Improving voltage profile of residential distribution systems using rooftop PVs and battery energy storage systems. Appl Energy 2014;134:290–300.
[8] Navarro-Espinosa A, Mancarella P. Probabilistic modeling and assessment of the impact of electric heat pumps on low voltage distribution networks. Appl Energy 2014;127:249–66.
[9] Long C, Wu J, Thomas L, Jenkins N. Optimal operation of soft open points in medium voltage electrical distribution networks with distributed generation. Appl Energy 2016;184:427–37.
[10] Boroojeni KG, Amini MH, Iyengar SS, Rahmani M, Pardalos PM. An economic dispatch algorithm for congestion management of smart power networks. Energy Syst 2016;7:1–25.
[11] Amini MH, Nabi B, Haghifam MR. Load management using multi-agent systems in smart distribution network. In: 2013 IEEE power & energy society general meeting. Vancouver, BC; 2013. p. 1–5.
[12] Chassin DP, Posse C. Evaluating North American electric grid reliability using the Barabási-Albert network model. Phys A: Stat Mech Appl 2005;355:667–77.
[13] Albert R, Albert I, Nakarado GL. Structural vulnerability of the North American power grid. Phys Rev E 2004;69:025103.
[14] Crucitti P, Latora V, Marchiori M. A topological analysis of the Italian electric power grid. Phys A: Stat Mech Appl 2004;338:92–7.
[15] Rosas-Casals M, Valverde S, SolÉ RV. Topological vulnerability of the european power grid under errors and attacks. Int J Bifurcation Chaos 2007;17:2465–75.
[16] Andersson G, Donalek P, Farmer R, Hatziargyriou N, Kamwa I, Kundur P, et al. Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. IEEE Trans Power Syst 2005;20:1922–8.
[17] Boroojeni KG, Amini MH, Iyengar S. Overview of the security and privacy issues in smart grids. Smart grids: security and privacy issues: Springer; 2017. p. 1–16.
[18] Vahdati PM, Kazemi A, Amini MH, Vanfretti L. Hopf bifurcation control of power system nonlinear dynamics via a dynamic state feedback controller-part I: Theory and Modeling. IEEE Trans Power Syst 2017;32:3217–28.
[19] Pagani GA, Aiello M. The power grid as a complex network: a survey. Phys A: Stat Mech Appl 2013;392:2688–700.
[20] Chen G, Dong ZY, Hill DJ, Zhang GH, Hua KQ. Attack structural vulnerability of power grids: a hybrid approach based on complex networks. Phys A: Stat Mech Appl 2010;389:595–603.
[21] Arianos S, Bompard E, Carbone A, Xue F. Power grid vulnerability: a complex network approach. Chaos: an Interdisciplinary. J Nonlinear Sci 2009;19:013119.
[22] Erdős P, Rényi A. On the strength of connectedness of a random graph. Acta Math Hung 1961;12:261–7.
[23] Watts DJ, Strogatz SH. Collective dynamics of 'small-world'networks. Nature 1998;393:440–2.
[24] Barabási A-L, Albert R. Emergence of scaling in random networks. Science 1999;286:509–12.
[25] Wang Z, Scaglione A, Thomas RJ. Generating statistically correct random topologies for testing smart grid communication and control networks. IEEE Trans Smart Grid 2010;1:28–39.
[26] Barakou F, Koukoula D, Hatziargyriou N, Dimeas A. Fractal geometry for distribution grid topologies. PowerTech, 2015 IEEE Eindhoven. IEEE; 2015. p. 1–6.
[27] Hu J. Cluster-and-connect: an algorithmic approach to generating synthetic electric power network graphs. Arizona State University; 2015.
[28] Domingo CM, Román TGS, Sánchez-Miralles Á, González JPP, Martínez AC. A reference network model for large-scale distribution planning with automatic street map generation. IEEE Trans Power Syst 2011;26:190–7.
[29] Chin Kim G, Pudjianto D, Djapic P, Strbac G. Strategic assessment of alternative design options for multivoltage-level distribution networks. IEEE Trans Power Syst 2014;29:1261–9.
[30] Fichera A, Frasca M, Volpe R. Complex networks for the integration of distributed energy systems in urban areas. Appl Energy 2017;193:336–45.
[31] Barabási BA-L, Bonabeau E. Scale-free. Scientific American; 2003.
[32] Newman ME. Assortative mixing in networks. Phys Rev Lett 2002;89:208701.
[33] Block A, Von Bloh W, Schellnhuber H. Efficient box-counting determination of generalized fractal dimensions. Phys Rev A 1990;42:1869.
[34] Chen J-C. Dijkstra's shortest path algorithm. J Formalized Math 2003;15:144–57.
[35] Demographic Yearbook (Table 6) btUN. 2013.
[36] MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley symposium on mathematical statistics and probability, vol. 1: Statistics. Berkeley, Calif.: University of California Press; 1967. p. 281–97.
[37] Al-Wakeel A, Wu J, Jenkins N. k-means based load estimation of domestic smart meter measurements. Appl Energy 2017;194:333–42.