

## 基于 $K$ -means 聚类方法的电力通信模型自动生成方法及其优化

金鑫<sup>1</sup>, 杨亮<sup>1</sup>, 金成明<sup>2</sup>, 苏国华<sup>3</sup>, 孙磊<sup>3</sup>

(1. 中央财经大学信息学院, 北京 100081; 2. 国网辽宁省电力有限公司信息通信分公司, 辽宁 沈阳 110006;

3. 北京国电通网络技术有限公司, 北京 100070)

**摘 要:** 为了在电力通信模型的建立过程中生成一个比较合理和强壮的初始通信模型, 在传统的 Waxman 方法的基础上, 运用  $K$ -means 的聚类思想生成节点, 并对所生成的拓扑网络进行脆弱性分析和信息孤岛检测, 然后对于网络中的脆弱节点进行补强。仿真实验结果显示这种结合  $K$ -means 和脆弱性分析的拓扑生成算法在生成对意外风险具有较强抗性的电力网络拓扑方面具有比较好的效果。

**关键词:**  $K$ -means 聚类; 电力通信; 网络拓扑; 脆弱性分析

**中图分类号:** TP391; TP393

**文献标识码:** A

## Method to create and optimize original electric power communication network based on $K$ -means

JIN Xin<sup>1</sup>, YANG Liang<sup>1</sup>, JIN Cheng-ming<sup>2</sup>, SU Guo-hua<sup>3</sup>, SUN Lei<sup>3</sup>

(1. Information School, Central University of Finance and Economics, Beijing 100081, China;

2. Liaoning Power Supply Company ICT Branch of State Grid Corporation, Shenyang 110006, China;

3. Beijing State Power Communication Network Technology Company, Beijing 100070, China)

**Abstract:** The major contribution was using the  $K$ -means to create nodes instead of traditional Waxman means in order to get a suitable and strong original electric power communication network. After the nodes were created, a vulnerability analysis was given on the grid of the network topology and check that if it has a network islanding, at last the weak nodes would be reinforced. A simulation about this algorithm shows that this algorithm which combines  $K$ -means and vulnerability can create a electric power communication networks with better resistance of risks.

**Key words:**  $K$ -means clustering, power communication, network topology, vulnerability analysis

### 1 引言

电力通信网络拓扑模型是一种基于真实电力通信网络的模拟, 是当下研究电力通信网络的一种重要手段。由于电力通信网络存在多样性和复杂性, 影响拓扑模型模拟稳定性的重要因素就是研究过程中所选择的拓扑模型。用尽量能反映真实电力通信网络的模型作为对真实电力通信环境的一种抽象, 是取得正确结果的必要条件。在电力通信模型的生成过程中, 要求研究设计的电力通信模型能适用于某个特殊节点发生故障或被人工干预场景

下的情形。这样可以使生成的电力通信拓扑模型尽量能反映真实应用的情况。为了给电力系统提供必要的支持, 保证安全、经济地发供电, 确保电能的合理分配, 保证电力指标的质量, 以方便系统事故的防止和及时处理, 达到电力系统的集中管理和统一调度, 一个能够提供特殊保障性服务的通信系统是非常必要的。而具有可靠的保障性服务的通信系统取决于电力通信的物理结构和拓扑模型。目前, 关于电力通信网络模型的研究, 大多是针对其可靠性或者脆弱性等对已有的具体电力通信网络模型进行分析, 如文献[1,2]等。对于生成初始的电力通信

收稿日期: 2016-08-28

基金项目: 国家自然科学基金资助项目 (No.U1509214); 国网科技部基金资助项目 (No.SGTYHT/14-JS-188)

**Foundation Items:** The National Natural Science Foundation of China (No.U1509214), Technology Project of State Grid Corporation of China (No.SGTYHT/14-JS-188)

模型的研究则包括最小生成树的 Kruskal 算法、最短路径算法和最大流量算法等, 如文献[3,4]等。本文将通过一种在 Waxman 算法基础上引进  $K$ -means 聚类算法的方法来进行电力通信模型的初始生成, 这样生成的电力通信网络拓扑模型相较上述方法, 生成的网络比较灵活多样, 对现实的电力网络模拟得更好。而且在生成初始网络后, 会对网络的脆弱性进行分析, 并根据分析结果对网络中的脆弱节点添加冗余链路, 实现网络的进一步优化, 使生成的电力通信网络结构更加合理, 具有更强的顽健性。

## 2 Waxman 随机模型

1988 年, Waxman 首先提出了一个随机网络拓扑生成模型, 后来被网络通信界广泛接受, 此模型生成随机网络拓扑的思想是: 假设网络规模即节点数为  $N$ , 然后按照式(1)的概率  $P_e$  来决定 2 个节点  $u$ 、 $v$  之间是否存在一条直连链路<sup>[5,6]</sup>。

$$P_e(u, v) = \beta \exp \frac{-L(u, v)}{\alpha} \quad (1)$$

其中, 参数  $L$  为所有拓扑中节点的距离的最大值,  $L(u, v)$  为  $u$  与  $v$  的几何距离,  $\alpha$ 、 $\beta$  是调节此拓扑结构命令的参数。

在实际应用中, Waxman 算法在节点数较多时, 平均节点度远大于最佳值  $4^{[6]}$ , 因此, 在网络规模较大的时候, 原始的 Waxman 算法已经不能满足实际的需要。

为了解决上述问题, SALAMA 于 1996 年对 Waxman 模型进行了改进, 他提出的 SALAMA 模型可以用来控制网络拓扑结构节点度的模型, 从而为此种问题的解决提供了一个可行的思路。文献[6]比较详细地介绍了此模型的算法, 在此模型中引进了参数  $\bar{e}$  用于控制网元的节点度。当某个节点的节点度较高时, 其在拓扑结构中的连接几率将有所下降。

综上所述, 原始 Waxman 方法及其改进方法 SALAMA 都可以按照一定的规律生成网络拓扑, 但是存在着以下几个比较明显的缺点。一是网络节点都是完全随机生成的, 没有间距和疏密的控制, 生成的拓扑可能非常复杂而难以辨认。二是由生成公式可知, 2 个点距离越远, 那么两点之间产生连接的概率就越小, 可能导致网络孤岛出现, 即最后生成的拓扑结构可能并非全部连通而存在一个或多个节点完全孤立, 与其他任何节点都没有联系, 文献[7]对这种情景有比较详细的描述。

而且在 Waxman 这种比较原始的节点生成算法中, 如果建模者要产生  $n$  个节点, 就使用随机函数直接生成所需要的  $n$  个节点, 理论上所有节点的生成位置完全随机, 这种随机性使产生的节点数量较少时可能出现相距过近或过远的情况, 一旦出现这种情况, 建模者就必须将现有生成节点全部删除再重新进行节点的初始化, 这样会带来时间和计算资源上的浪费。

## 3 基于 $K$ -means 聚类方法的电力通信网络拓扑结构生成算法

针对上述缺点, 本文尝试利用  $K$ -means 聚类方法来产生网元节点。 $K$ -means 聚类是指将数据对象划分为  $K$  个类的过程, 是一种非常经典的聚类算法, 文献[8]对其初始思想以及近些年的发展有比较详尽的介绍。 $K$ -means 聚类思想是先随机生成  $K$  个聚类中心, 然后逐个将需要分类的每个聚类样本按最小距离准则分配给  $K$  个聚类中心中的某一个, 所有聚类样本分配完毕之后, 求各个聚类中心的样本均值向量, 作为新的聚类中心, 如此循环进行, 直达到某个终止条件。由于这  $K$  个中心是通过均值法聚类得到的, 因此不会出现 2 个或多个网元节点之间相距很近甚至重叠的现象。

采用  $K$ -means 聚类方法的生成电力通信网络拓扑的算法描述如下<sup>[8]</sup>。

**Step1** 设原始网元节点为  $\{x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(m)}\}$ , 这些网元节点是完全无序的、没有被标记的。

**Step2** 初始化  $k$  个随机节点  $\{u_1, u_2, u_3, \dots, u_k\}$ 。

**Step3** 计算所有原始节点和初始化的随机网元节点间的距离, 然后根据计算结果, 列出距离每个随机节点最近的节点, 计算方法如式(2)所示。

$$c^{(i)} = \arg \min \|x^{(i)} - u_j\|^2 \quad (2)$$

**Step4** 求出所有和找出的节点最近的原始节点距离的均值, 计算方法如式(3)所示。

$$u_j = \frac{\sum_{i=1}^m I\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m I\{c^{(i)} = j\}} \quad (3)$$

**Step5** 式(2)和式(3)不断迭代, 直到所有的  $u$  都变化不大了, 聚类过程就完成了, 所得到的  $u - u_k$  就是所要求的聚类中心。

在电力通信网络的实际构建中, 假如要求产生

$n$  个节点的网络拓扑, 可先在平面上生成  $x$  个节点, 并使  $x$  值远大于  $n$  值 (比如  $x=100n$ ), 然后用  $K$ -means 方法求出这  $n$  个聚类中心, 多次迭代后, 这  $n$  个聚类中心就是所要求产生的  $n$  个节点的电力通信网络拓扑, 然后计算这  $n$  个节点链路之间的距离, 得到一个  $n \times n$  的链路矩阵, 最后依据 Waxman 方法计算出一个概率数值以决定 2 个节点间是否应该产生联系。考虑到电力系统的实际工作过程中大量小型设备、网元 (如家庭电表、小型变压器等) 均以当地的变电站、大型变压器等为运转中心, 这种聚类是符合实际情况的。

#### 4 电力通信网络拓扑的脆弱性分析及优化

在上节中构建了基于  $K$ -means 的电力网络初始拓扑结构后, 通过分析此网络中节点的脆弱性 (vulnerability), 来对这个电力通信网络进行进一步的优化。

##### 4.1 脆弱性分析

电力网络通常包括电力设备和电力线路 2 个部分, 电力设备可看作拓扑中的节点, 电力线路可看作拓扑中生成的链路。如果电力设备或线路因遭受到外界因素或自身因素的影响而失效, 电力网络的传输性能必然会受到一定程度的影响。文献[9,10]是不同领域应用复杂网络知识进行分析网络脆弱性的具体例子。

如果要以更加量化的指标来考察拓扑中某一节点失效时对网络的影响, 设置一个定量的参数指标来进行评价是必要的。本文将采用基于“网络全局效能”生成的“网络效能函数  $E(G)$ ”来作为某一节点失效时对网络全局影响力的评价指标<sup>[11]</sup>。

网络效能函数  $E(G)$  可以直接找出一个网络中“最脆弱”的节点, 只需要关注它的计算值就可以找出网络拓扑中最脆弱的节点并对其进行加强。网络效能函数  $E(G)$  的具体计算方法如式(4)所示。

$$E(G) = \frac{1}{N(N-1)} \sum_{i,j \in G, i \neq j} \varepsilon_{ij} = \frac{1}{N(N-1)} \sum_{i,j \in G, i \neq j} d_{ij} \quad (4)$$

式(4)的含义是: 在具有  $N$  个节点的网络  $G$  中,  $d_{ij}$  为网络中连接节点  $i$ 、 $j$  之间的最短路径长度, 则定义  $\varepsilon = \frac{1}{d_{ij}}$  为节点  $i$  和  $j$  之间的效能, 此时整个

网络的平均效能计算方法即如式(4)所示。

可以使用诸如 DIJKSTRA 算法这种典型的最

短路径算法计算出  $d_{ij}$ , 进而求得  $E(G)$ 。

对网络  $G$  而言, 脆弱节点分析过程如下。

**Step1** 计算网络  $G$  中各节点的度数, 并将所得结果从大到小进行排序。

**Step2** 选择度数最大的节点, 将此节点和所有与此节点相连的链路从  $G$  中去除, 得到剩余网络  $G'$ 。

**Step3** 计算  $G'$  的传输效能  $E(G')$ 。

对于每一个去除节点后的计算结果进行排序, 造成  $E(G)$  降落最大, 即令  $E(G')$  最小的节点, 即为此网络拓扑的最脆弱的节点。

##### 4.2 电力通信网络优化

为了对此网络拓扑结构中的脆弱节点进行进一步的补强和优化, 增强网络的顽健性和对意外的抵抗性, 本文将生成一个链路集, 通过增加一定数量的冗余链路来实现优化, 同时为避免节点度过高, 会设置一个阈值, 如果某个节点所连链路过多, 会强行跳过当前节点的优化, 进行下一个节点的分析, 这种方法是避免出现某些“中枢节点”的出现, 这些“中枢节点”连接了过多的链路, 一旦出现故障将导致全网瘫痪。

对于脆弱的部位, 本文以添加冗余链路的方式加以优化。添加冗余链路方式算法描述如下。

**Step1** 计算网络中每个节点的介数  $b(n)$ , 节点介数为网络中所有最短路径中经过该节点的路径的数目占最短路径总数的比例。

**Step2** 根据实际需要, 对于最脆弱的一个或几个网元节点, 基于其节点介数, 为其添加冗余链路。

**Step3** 之后再计算新生成拓扑网络的效能函数  $E$ , 若使本节点的  $E(G)$  降落幅度有所下降, 则说明所添加冗余链路成功加强了网络的顽健性使网络更加强壮; 如果  $E(G)$  的降落幅度反而上升了, 则说明添加的链路有问题, 将取消刚才添加的冗余链路, 并重新返回 Step2 中生成一条新的冗余链路继续进行测试。

为防止死循环和过于复杂的网络拓扑出现, 本算法设置了一定的节点度作为终止条件, 即某脆弱节点添加冗余链路后, 节点度超过设定的值后, 此冗余链路仍将取消, 并且不再自动添加其他冗余链路, 而进行下一个脆弱节点的计算。

#### 5 实验评估

通过实验来分析 Waxman 方法和  $K$ -means 方法生成电力通信网络的效果, 实验软件采用的是

Matlab7.0, 实验硬件环境为 I7-4710 四核八线程 CPU, 64 位 Win7 系统, 2 GB 内存。

### 5.1 生成网络拓扑结构对比

在 Matlab 中分别使用 2 种方法模拟生成 10 个网元和 25 个网元的拓扑结构, 并加以对比得出结论。

Waxman 方法生成的 10 个节点的网络拓扑结构和 K-means 聚类方法生成的 10 个节点的网络拓扑结构对比如图 1 所示。

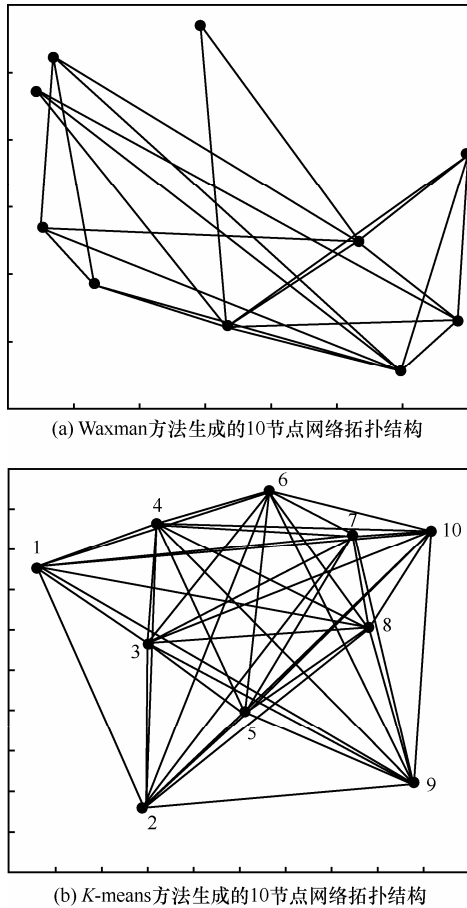
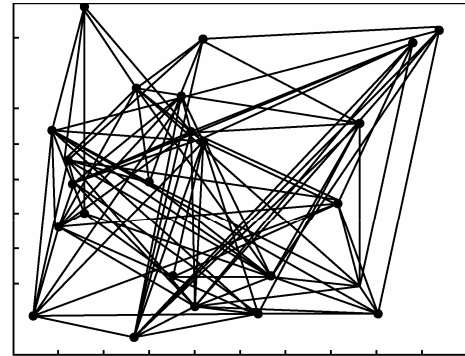


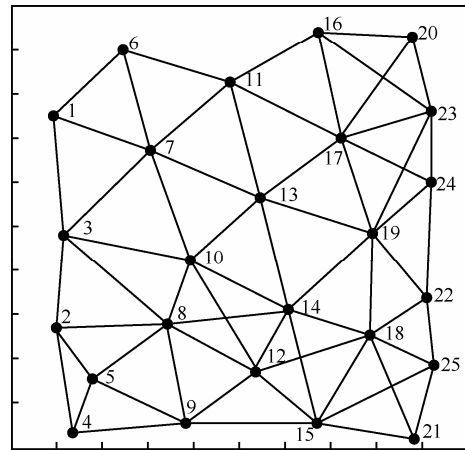
图 1 原始 Waxman 方法生成的拓扑结构 (10 节点) 与 K-means 方法生成的拓扑结构 (10 节点) 对比

Waxman 方法生成的 25 个节点的网络拓扑结构和 K-means 聚类方法生成的 25 个节点的拓扑结构对比如图 2 所示。

可以看到, 在节点总数较少 (10 个左右) 的时候, Waxman 方法与 K-means 方法生成的拓扑结构图复杂度相差不大, 但随着节点的增加 (25 个), Waxman 方法生成拓扑的复杂度有了显著的提高, 不利于网络拓扑的实际构建, 而 K-means 方法生成的聚类节点度仍然较低, 边交叉现象较少。本次实验中, 设定节点度最高为 7, 可以看到 8 和 17 是 2 个关键脆弱节点, 从而被添加了冗余链路。



(a) Waxman 方法生成的 25 节点网络拓扑



(b) K-means 方法生成的 25 节点网络拓扑

图 2 原始 Waxman 方法生成的拓扑结构 (25 节点) 与 K-means 方法生成的拓扑结构 (25 节点) 对比

由上述对比可以看出, Waxman 随机模型在电力通信网元数量较少时表现不错, 但随着网元的增多, 其复杂度也以几何程度上升, 如果通过调低式 (1) 中的参数  $\alpha$ 、 $\beta$  来降低复杂度, 又可能导致信息孤岛的出现 (即生成的网元不与任何其他网元连接或者某几个网元组成的系统与其他网元相互孤立) 而 K-means 方法则解决了在网元数量较多时出现的此种问题。

### 5.2 脆弱性对比

在原有 100 个节点电力通信网络中网元逐个失效的情况下, 根据式 (4), 分析原始 Waxman 方法、未进行冗余链路添加的 K-means 方法和本文最终提出的算法的网络效能函数, 模拟最终结果如图 3 所示。

可以看出, 本文算法在降低电力网络脆弱性方面, 相较于原始的 Waxman 方法和单纯不添加冗余链路的 K-means 方法, 有着比较大的优势, 其网络全局效能函数初始值比较高, 且随着网络中失效节点的逐渐增多, 其网络全局效能函数初始值依然能领先于这 2 种方法。

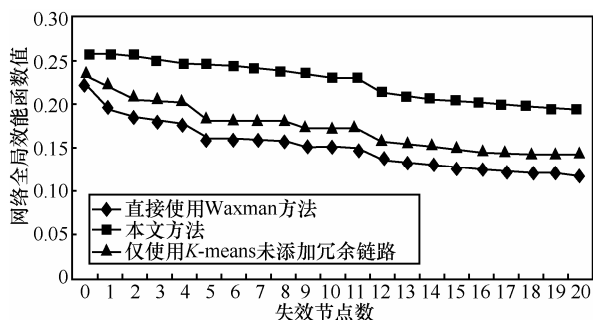


图 3 100 个节点下原始 Waxman 方法、未进行冗余链路添加的 K-means 方法和本文方法的网络全局效能函数值对比

因此，本文通过上述模拟实验认为，本文算法在电力通信网络的生成方面，相较于原始的 Waxman 方法，有了较大的改进，具体体现为生成的网络比较简洁、交叉少，且脆弱性较 Waxman 方法生成的模型低。

### 6 结束语

本文主要提出了电力通信模型的一种自动生成方法及其基于脆弱性方面的改进，具体来说，是在原有的 Waxman 方法上进行了一些改进，加入了 K-means 方法以生成初始电力通信网络拓扑，然后再对此拓扑中的节点进行了脆弱性分析，对于脆弱性较高的节点添加了冗余链路以增强其顽健性，最后生成一个脆弱性较低的电力通信网络拓扑。对比传统的 Waxman 方法，这种 K-means 聚类方法生成的网络拓扑边数简洁，交叉现象较少，而且在抵抗电力网络节点中的某个节点意外失效的情况时表现要更好。

### 参考文献:

[1] 赵璟. 基于业务可靠性的电力通信网规划算法的设计及应用[D]. 北京邮电大学, 2015.  
ZHAO J. Design and implementation of planing algorithm of electric power communication network based on business reliability[D]. Beijing University of Posts and Telecommunications, 2015.

[2] 张倩倩. 电力通信资源管理中拓扑可视化的研究与应用[D]. 华北电力大学, 2014.  
ZHANG Q Q. Research and application of network topology visualization for power communication resource management[D]. North China Electric Power University, 2014.

[3] 李慧娟, 丁宝, 董成明, 等. 基于最小生成树的电力通信系统时钟同步网规划算法[J]. 电力学报, 2012(2): 136-139.  
LI H J, DING B, DONG C M, et al. Clock synchronization network planning algorithm of electric power telecommunication system based on minimum spanning tree[J]. Journal of Electric Power, 2012(2): 136-139.

[4] 黄宗君. 基于最小生成树理论的配电网故障恢复算法[D]. 四川大

学, 2004.  
HUANG Z J. A distribution network reconfiguration algorithm based on least weight tree for service restoration after faluts[D]. Sichuan University, 2004.

[5] WAXMAN B M. Routing of multipoint connections[J]. IEEE Journal of Selected Areas in Communication, 1988, 6(9): 1617-1622.

[6] 周灵. Waxman-Salama 模型网络拓扑生成算法设计与实现[J]. 湖南理工学院学报(自然科学版), 2008(2): 40-42.  
ZHOU L. Design and implement of random network topology using Waxman-salama model[J]. Journal of Human Institute of Science and Technology(Natural Sciences), 2008(2): 40-42.

[7] 马天琳. 复杂电力网络中连锁故障异常现象分析与故障防御[D]. 华北电力大学, 2014.  
MA T L. Abnormal phenomena and defence of cascading failures in complex power grids[D]. North China Electric Power University, 2014.

[8] JAIN A K. Data clustering: 50 years beyond K-means[J]. Pattern Recognition Letters, 2010, 31(8): 651-666.

[9] 杨博, 刘大有, 刘金弟, 等. 复杂网络聚类方法[J]. 软件学报, 2009(1): 54-66.  
YANG B, LIU D Y, LIU J D, et al. Complex network clustering algorithms[J]. Journal of Software, 2009(1): 54-66. ,

[10] 郑啸, 陈建平, 邵佳丽, 等. 基于复杂网络理论的北京公交网络拓扑性质分析[J]. 物理学报, 2012, 19: 95-105.  
ZHENG X, CHEN J P, SHAO J L, et al. Analysis on topological properties of Beijing urban public transit based on complex network theory[J]. Acta Phys Sin, 2012, 19: 95-105.

[11] 周苗, 杨家海, 刘洪波, 等. Internet 网络拓扑建模[J]. 软件学报, 2009(1): 109-123.  
ZHOU M, YANG J H, LIU H B, et al. Modeling the complex internet topology[J]. Journal of Software, 2009(1): 109-123.

### 作者简介:



金鑫 (1974-), 男, 内蒙古乌海人, 中央财经大学教授, 主要研究方向为商务智能。

杨亮 (1989-), 男, 满族, 河北承德人, 中央财经大学硕士生, 主要研究方向为大数据应用。

金成明 (1981-), 男, 吉林省吉林市人, 国网辽宁省电力有限公司信息通信分公司工程师, 主要研究方向为电力信息化。

苏国华 (1986-), 男, 山东聊城人, 北京国电通网络技术有限公司助理工程师, 主要研究方向为电力云计算、电力通信大数据分析和电力办公基础软件。

孙磊 (1987-), 男, 江苏盐城人, 北京国电通网络技术有限公司工程师, 主要研究方向为电力云计算、电力通信大数据分析和电力办公基础软件。