RESEARCH ON THE APPLICATION OF NODE IMPORTANCE ASSESSMENT BASED ON HITS ALGORITHM IN POWER GRID PLANNING

GAOSHAN FU, XIANG YIN, YUE GAO, DAN MENG, AND LIANG CHEN

Abstract. Power grid planning needs to be strong and effective as the world’s energy environment shifts to include more renewable energy sources and smart technology. This study explores the use of the HITS (Hyperlink-Induced Topic Search) algorithm to apply node importance assessment in the context of power grid planning. The HITS method provides a new way of looking at the importance of nodes in power grid networks. It was initially developed for online link analysis. The first section of the paper offers a thorough analysis of the power grid planning techniques now in use, highlighting the crucial role that nodes play in guaranteeing flexible and resilient systems. Next, the HITS method is modified and used in power grid networks, taking dependability, interaction, and node centrality into account. As part of the research process, a mathematical model that combines the HITS method with important variables unique to power grid planning is developed. On real-world power grid datasets, simulation tests are carried out to evaluate the algorithm’s performance in identifying nodes that are critical to fault tolerance, overall performance, and system stability. The study’s findings go beyond conventional power grid planning techniques by providing a sophisticated method of evaluating node relevance that is in line with the dynamic and interdependent character of contemporary energy networks. The results aid in the infrastructure optimization of the power grid, allowing planners and managers to better prioritize expenditures, increase resilience to disturbances, and make it easier to integrate energy from renewable sources smoothly.

Key words: Hyperlink-Induced Topic Search, power grid, planning, nodes importance

1. Introduction. Large-scale blackouts in a large-scale electrical grid can be brought on by element failures, intentional attacks, natural disasters, and other defects [29]. Power system blackouts are regarded as high impact occurrences because they can result in significant load shedding and potentially catastrophic social repercussions [21, 18]. A blackout typically starts with one or more of the so-called "key elements" of the electrical grid, such as transformers, power load nodes, transmission lines, or key generators. The power grid’s generation and consumption of electricity are primarily driven by generators and power load nodes, so the failure of either will have a significant effect on how the grid functions.

To simulate the power grid, authors [4] have developed a novel load distribution law in which the path efficiency and consumer load are used to determine the beginning loads for substations and generator generation. It is stated how important power load nodes and generator nodes are. Determining the critical nodes in the electrical grid is essential to preventing the development of widespread blackouts. Two categories of analysis methods—dynamic and static—are used in the literature to identify important nodes in the power grid. Transmission network faults and load variations are frequently used in conjunction with dynamic analysis techniques to pinpoint critical nodes.

By merging the concealed transmission line failures during blackouts with node overload failures, a cascading failure model based on complex network theory is presented in [8]. Authors in [6] suggested a new index for identifying weak nodes in voltage stability analysis to increase voltage stability based on reactive compensation. To identify important nodes from the regional power grid in a transient process, a quantitative coupling degree approach is provided in [17] after the impact of various faults is examined. Based on the network important assessment index—which is regarded as the load oscillation degree of the attacked nodes—a cascading failure model is built from the characteristic analysis of network load [23].
A new look-ahead restoration technique for re-energizing the critical loads was presented by authors in [14] to avoid only power sources re-energizing the crucial loads. Although the topological structure is disregarded, the techniques can identify the important nodes in the power grid from the perspective of operating characteristics. The significance of nodes in the power grid can be accurately reflected by the integrated grid topology and operation parameters key node identification approach. Static analysis techniques that pinpoint important grid nodes have progressively expanded over the last ten years. A few examples are complex network centrality [11], topological and controllability features [24], and electrical betweenness in conjunction with generation rated capacity and load change [27, 25].

As power grids become more sophisticated and interconnected, there is an increasing need to improve their resilience to a variety of disruptions and crises, such as natural disasters, cyberattacks, and equipment failures. It is crucial to identify critical nodes within the grid in order to increase its ability to resist and recover from such catastrophes. Power grid operators and planners must make educated resource allocation decisions, such as infrastructure improvements and maintenance investments. Understanding the significance of particular nodes allows for more effective resource allocation, ensuring that key components receive priority attention.

The main contribution of the proposed method is given below:
1. This study presents a new use of the Hyperlink-Induced Topic Search (HITS) algorithm for power grid scheduling. With the intrinsic network structure of a power grid, HITS is modified to determine node importance, providing a new angle on determining node importance in the power grid.
2. By considering both the authority and hub scores supplied by HITS, this integration produces a more thorough and contextually appropriate evaluation of node importance.
3. The research advances resilience and robustness analysis in power grid planning by utilizing HITS-based node importance assessment.
4. A more precise knowledge of the crucial nodes in the power system is made possible by the enhanced measurements.

Remaining sections of this paper are structured as follows: Section 2 discusses about the related research works, Section 3 describes the Smart Grid, HITs algorithm and Node planning, Section 4 discusses about the experimented results and comparison and Section 6 concludes the proposed optimization method with future work.

2. Related Works. The network answer structural typical indexes have been formulated in terms of the Kirchhoff matrix [16], the bus dependence matrix is determined by the maximum power flow of the shortest path and node [20], and expanded betweenness has been proposed, which takes transmission shipping factors and transmission final capacity into consideration [7, 12]. Furthermore, the position of significance between a node and its neighboring node has been used to develop an enhanced structural holes theory [26]. Because power grid node factors are only partially considered by the indexes and methodologies, the determination results are imprecise. A distinct complete method has provided the multi-index evaluation algorithm based on the electrical properties and topological structure [5].

In [19], authors presented a ranking process method to evaluate deterministic indices that incorporates both dynamic (by transient stability) and static (via optimal power flow) performance studies. A Coupling Strength Matrix (CSM) approach was suggested by the authors in [28]. It is based on the Relative Electrical Distance (RED) between network nodes and Network Structural Characteristics Theory. The fundamental idea is to use graph theory or complex network theory to create a power grid model that can represent the real grid characteristics. Next, indexes are created to help locate significant nodes in the grid. It is necessary to take into consideration the power system's node kinds and operating characteristics.

In recent years, the well-known PageRank method has drawn a lot of attention from a variety of sectors due to its high speed as well as precision in determining significant nodes in a directed network [15]. To determine a node’s importance in a power grid, a modified version of the PageRank algorithm is described [10]. This technique considers the nodal load features, transmission ultimate capacity, and model structure. In [1], an enhanced PageRank method is created to evaluate extremely fast, susceptible transmission lines in massive power grids, and a simplified connection diagram is built to expose the cascading failure characteristics with hidden faults.

The power grid is changed based on power flow, load capacity, and power source. The modified sorting
method PageRank, known as hypertext, induced topic selection (HITS), is presented to find key nodes in [13]. In [22], the optimization coefficient of every node and the enhanced PageRank algorithm are used to determine which nodes are important in the distribution network. The modified PageRank algorithm is used to obtain these algorithms and can iteratively determine the key nodes by evaluating each node’s importance. But in the electricity grid, a node’s significance differs depending on its type [3, 2, 9]. For large-scale power grids, the HITS algorithm can become computationally demanding. Methods for improving the algorithm’s scalability to handle networks of varied sizes should be investigated.

Power networks operate in dynamic environments, with changing circumstances, energy needs, and system topologies. Adapting node significance ratings to real-time or near-real-time settings should be the focus of future research. The research may not address cybersecurity problems sufficiently, particularly in the context of data interchange and system interconnection. Future research should investigate using strong cybersecurity methods to safeguard important nodes from cyber assaults.

3. Proposed Methodology. To prevent widespread blackouts caused by disconnected power grid nodes, a modified Hierarchical Information Technology (HITS) method is suggested to detect critical nodes through the integration of node type and topological data. Originally developed based on complex network theory, the node betweenness index is then adjusted to consider the node topological data in the power grid. Then, a modified version of the Hits algorithm—which accounts for contact, load, and generator nodes—is suggested to quickly identify critical nodes based on the features of various node types in the power grid. In figure 3.1 shows the architecture diagram of proposed method.

3.1. Power Grid Model. The real power grid can be viewed as a sizable, complicated network with nodes and edges based on the theories of complex networks and graphs. Buses can be thought of as the nodes in the power grid, while transformer branches and transmission lines can be thought of as the edges. Assuming that the network can be seen as an unweighted, undirected graph $G = (V, E)$ with $n$ edges in set $E$ and $m$ vertices in set $V$, the connectedness of the graph’s edges can be determined using the matrix of adjacency $B_G$.

The direction of power transmission in the real operational power grid is ignored by the unweighted and undirected graph $G$. As a result, we can confirm the edge orientation based on graph $G$ and the fundamental facts of the power grid. Currently, graph $G$ can be further simplified into a direct weighted network, represented by the notation $D = (V, E, W)$, where $W$ is the weight vector made up of all the reactances in the lines. Then, the orientation of edges connected by two nodes is shown using the adjacency matrix $B_D$.

3.2. HITs Algorithm. Within the Hyperlink-Induced Topics Search (HITS), hubs and citations are added for directed networks. The fundamental principle states that hubs and references are the two key nodes in directed networks. A significant hub effectively links to numerous significant sources. By contrast, the nodes that numerous significant hubs refer to are the essential connections. Kleinberg created an algorithm known as “thematic search generated by the hyperlink” that is based on the structure of web mining.

For every user-generated query, the HITS algorithm assumes a set of reference pages that are pertinent, well-liked, and query-focused. Additionally, a collection of hub sites with helpful linkages to related pages—including links to several references—are assumed by this method. According to the HITS, the web is a directional graph $G (V, E)$, where $V$ is a collection of vertices that represent pages and $E$ is a collection of edges that are connected
by links. A link from page p to page q is represented as a directed link (p, q). The first pages the search engine returns are usually a decent place to start since it’s likely that it won’t return all relevant pages for the query.

Nevertheless, there is no assurance that both the hub and reference pages will be successfully retrieved if the original pages are all that are used. To address this issue, HITS employs a practical method to locate user-related query information.

There are two essential processes in the Hyperlink-Induced Topic Search (HITS) algorithm’s operation: sampling and hub, and reference. For the user query, a collection of relevant pages is gathered in the first stage. Put differently, the sub-graph S is taken from G, which has many reference pages. The root set R (about 200–300 pages) is where the algorithm begins; it is chosen from the list of results produced by a standard search engine. From R, one can obtain the set S. It should be mentioned that most of the strongest references are found in this very tiny, densely referenced S t. Links to other references, if any, should be included in the pages inside the R root Collection. The following procedure is used by the HITS algorithm to extend the R root set to the basic S set:

1. The set of all of R’s roots is the input; the base set S is the output.
2. To begin, assume that the set S and the set R are equal.
3. For every p ∈ S, follow steps 3 through 5.
4. Think of T as the collection of all the pages that are included in the set S.
5. Think of F as a collection of pages that make references to S.
6. Treat all or a portion of S = S + T as part of F.
7. Eliminate every link sharing the same domain.
8. Get S back.

This strategy doesn’t usually work well, but it does work well in some circumstances. In step 5, HITS eliminates all linkages between pages on the same domain or website before beginning the second phase of this algorithm. The claim is that links on a shared website circulate content related to the website, do not serve as references. Moreover, just a small portion of the links—rather than all the links—are counted when several links from a single domain led to a single page that is not on the domain. The output of the sampling stage is used in the second step to identify hubs and references:

Baseline set is the input, while the hub and standard sets are the output.

1. Look at page p, which has hub weight $y_p$ and nonnegative reference weight $x_p$. References are pages with a comparatively high reference weight ($x_p$). In a similar manner, hubs are defined as pages having a high $y_p$ hub weight.
2. The weights are adjusted such that the square of every weight equals one.
3. The value of $x_p$ is adjusted for page p to equal the total of all the $y_p$ weights of all the q pages that are connected to p.
4. When a link from page 5 is made to any page q, the value of $y_p$ is changed to equal the total of that pages’ $x_p$ weights.
5. The algorithm goes back to step 2 if the output conditions are not met. There are two sets of pages: references, which have the highest $x_p$ weights, and hubs, which have the highest $y_p$ weights.

Hubs and citations are given appropriate weights. A strong reference is one that is cited by a significant number of well-regarded hubs. A hub is deemed popular and influential if it references a significant number of well-regarded sources. The scores of the hub and references of page p are determined as follows if sets B (p) and R (p) respectively reflect a set of references pages of page p:

$$x_p = \sum_{q \in B(p)} y_p$$  \hspace{1cm} (3.1)

$$y_p = \sum_{q \in B(p)} x_p$$  \hspace{1cm} (3.2)

As seen in Figure 3.2, the hub and reference locations are computed. Pages are rated according on their hub and points of reference.
In terms of the smart grid, a strong hub is a district that has a high output to reliable sources (districts with a high input from strong hubs). One way to specify the scores of a hub (\(h \rightarrow\)) and references (\(\rightarrow a\)) repeatedly; alternatively, the dominant special vectors of can be used to determine the hub matrix \(A A^T\) and the reference matrix \(A^T A\). Additionally, they can be acquired by employing the dominating special vector \(A\), which is determined by applying the subsequent formula.

\[
A = \begin{pmatrix}
0 & A \\
A^T & 0
\end{pmatrix}
\]  \hspace{1cm} (3.3)

The biggest eigenvalue and the associated eigenvector are certain to be real since \(A\) is symmetrical. \(A\) becomes a \(2n \times 2n\) matrix if it is a \(n \times n\) matrix. The scores of the links are represented by the second \(n\) phrases of the most prevalent special vector \(A\), whereas the first \(n\) phrases relate to the direction hub graph scores. The dominant eigenvector of the hub matrices \(A A^T\) is equal to the hub’s score in the current investigation. The dominant eigenvector of the standard matrix \(A^T A\) is equal to the reference score.

Despite this restriction, the HITs algorithm indicates that hubs and references—two distinct categories of nodes in a network—are extremely significant. Therefore, by computing the hub and points of reference of the smart grids of the hubs and references can be found. Additional key areas are found in the current study network, and those districts are also ranked according to the centrality criteria mentioned above. Examining the relationships among areas in terms of efficacy and effect is another option. Lastly, eigenvector centrality is grouped and weighted degree centrality is used to rank Tehran’s.

3.3. HITs algorithm applied to Power Grid. Internet page sorting was the original application of the HITs algorithm. Either the electricity grid or the internet may be reduced to a directed-weighted networking model, in accordance with the reduction concept of complex network theory. Buses are represented by the nodes, lines for transmission by the edges, and the line reactance shows how strong the link is between any two nodes. Table 3.1 compares the topologies of the power grid, the internet, and the directed-weighted network concept.

Table 3.1: Power Grid Comparison

<table>
<thead>
<tr>
<th>Directed-Weighted Network</th>
<th>Internet</th>
<th>Power Grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node</td>
<td>Web Page</td>
<td>Bus or Substation</td>
</tr>
<tr>
<td>Edge</td>
<td>Hyperlink</td>
<td>Line or Transformer</td>
</tr>
<tr>
<td>Information Value</td>
<td>Link Relationship</td>
<td>Power Flow</td>
</tr>
<tr>
<td>Degree</td>
<td>Number of Visits</td>
<td>Generator or Load Capacity</td>
</tr>
</tbody>
</table>

Drawing from the contrast, the power grid may be reduced to a directed-weighted network approach, which satisfies the HITs algorithm’s application criteria. Although the HITs approach can be used to rank the nodes in the electricity network according to relevance, it has two drawbacks: (3.1) it ignores the electrical properties that
exist among nodes. (3.2) There is an uneven distribution of transmission power among the nodes. Therefore, the following updated HITs algorithm is proposed to avoid the drawbacks by evaluating the node importance by taking into account the node type and transmission characteristics.

Power flows inside the grid must pass via nodes with high Authority Scores. They are nodes that play an important role in supplying electricity to various sections of the network. Nodes with high Hub Scores are critical grid links. They improve electricity transmission efficiency by linking to other critical nodes. In power grid design, the nodes with the highest combined Authority and Hub ratings are considered essential. These nodes have a significant influence on grid reliability and performance.

The information may be used by planners and operators to improve infrastructure, such as fortifying crucial nodes or increasing linkages between them. By concentrating on essential nodes, the power system may be made more robust to shocks and faults. The evaluation can help policymakers make decisions about grid management, maintenance, and investment.

4. Result Analysis. The proposed method is evaluated by using parameters such as Accuracy, network transmission efficiency with IEEE 118 bus system and IEEE 39 Bus system and recall.

Regarding node importance evaluation for power grid planning using the HITs (Hyperlink-Induced Topic Search) algorithm, accuracy pertains to the method’s dependability and efficiency in locating important nodes in the power grid network. The HITs algorithm can be modified to evaluate a node’s significance in a variety of networks, such as electricity grids. It was initially created for web link analysis.

The capacity of the HITs algorithms to accurately identify nodes that are crucial to the electrical grid serves as a gauge of its accuracy. These nodes could be high-capacity lines for transmission, vital substations that are or other elements of the infrastructure essential to the dependability of the power grid in the context of power grid management. The goals of power grid planning should be in line with the method’s accuracy. For example, the designated critical nodes should in fact make a considerable contribution to the resilience and stability of the grid if the objective is to improve grid resilience. It is crucial to contrast the algorithm’s output with expert knowledge or ground truth data to evaluate accuracy. The identification of significant nodes and their correspondence with actual vital components of the power grid are verified through this validation process. In figure 4.1 shows the accuracy of proposed method.

"Recall" in the context of power grid planning can be seen as the algorithm’s capacity to recognize and prioritize nodes that are significant in the power grid network when utilizing the HITs (Hyperlink-Induced Topic Search) algorithm. Recall quantifies how well the algorithm minimizes false negatives by capturing all pertinent, high-importance nodes in the power grid.

Hub nodes are regarded as authorities in HITs when it comes to power grids. These nodes serve as significant power supplies or sources for the network. Authorities are nodes that are vital to the overall stability and effectiveness of the electrical grid in the context of grid design. These could be important substations,
significant power producing facilities, or vital connecting points. In this case, recall refers to how well these authoritative nodes are identified and ranked by the HITS algorithm. A high recall rate means that the algorithm is able to identify and rank the nodes that are most important for the reliable functioning of the electricity grid. In figure 4.2 shows the evaluation of recall.

Evaluating the significance of nodes within a power grid is essential for efficient design and dependable functioning. Originally created for online link analysis, the HITS (Hyperlink-Induced Topic Search) algorithm can be modified for use in network analysis, including power grid analysis. Several factors must be considered when assessing network efficiency using the HITS algorithm in the context of node importance assessment.

Nodes in HITS are given hub ratings and authority. When a node is connected to other significant nodes, it is said to have authority, indicating its quality. High authority score nodes—i.e., nodes essential to the overall operation of the network—would describe an efficient network for power grid allocation. Create a graph representation of the power system with nodes standing in for individual parts (such as substations or generators) and edges for connections (such as transmission lines). Application of the HITS algorithm is based on this network. Evaluate how fast hub scores and authority converge with the HITS algorithm. Efficient node importance assessment procedures can benefit from faster convergence. In figure 4.3 shows the network transmission efficiency of IEEE 39 Bus system.

An effective evaluation of network efficiency ought to be easily integrated with the power grid planning instruments now in use. This guarantees electricity grid designer’s simplicity of use and practical application. Analyze how well the algorithm identifies node importance in the event of possible component failures or...
deliberate attacks. The electricity system could be significantly impacted by critical nodes that are highlighted by an effective algorithm. Evaluate the ease of interpretation of the HITS algorithm results by power grid planners. Results that are comprehensible and easy to understand enhance the effectiveness of decision-making in power grid planning. Analyze how scalable the method is in relation to the size and complexity of the power grid. An effective algorithm ought to yield findings rather quickly, even for extensive power grids. In figure 4.4 shows the evaluation of IEEE 118 bus system.

5. Conclusion. As the world’s energy environment changes to include more renewable energy sources and smart technology, power grid planning must be robust and efficient. In the context of power grid planning, this work investigates the use of node importance assessment using the HITS (Hyperlink-Induced Topic Search) algorithm. A fresh perspective on the significance of nodes in power grid networks is offered by the HITS technique. It was first created for link analysis on the internet. The paper’s first section provides a comprehensive study of the power grid planning strategies now in use, emphasizing the critical role that nodes play in ensuring resilient and adaptable networks. Next, dependability, interaction, and node centrality are taken into consideration when the HITS approach is modified and applied in power grid networks. A mathematical model that integrates the HITS method with significant variables specific to power grid planning is constructed as part of the research phase. Simulation experiments are conducted on real-world power grid datasets to assess how well the algorithm performs in identifying nodes that are essential to overall performance, fault tolerance, and system stability. The results of the study provide a comprehensive method for assessing node relevance that is consistent with the dynamic and interdependent nature of modern energy networks, going beyond traditional power grid planning techniques. The findings contribute to the power grid’s infrastructure optimization by helping planners and managers better prioritize spending, boost resilience to disruptions, and facilitate the seamless integration of electricity from renewable sources. Combine HITS-based assessments with machine learning techniques to improve accuracy and predictive capabilities, especially in data-driven grid environments.

REFERENCES


Edited by: Rajanikanth Aluvalu
Special issue on: Evolutionary Computing for AI-Driven Security and Privacy:
Received: Dec 10, 2023
Accepted: Jan 4, 2024