# Predicting the Evolution Process of Infrastructure Networks With an NSIPA Link Prediction Method

Lan Mei, Disheng Tang, Tianyu Wang, Wenbo Du, Yongxiang Xia, and Xianbin Cao

***Abstract*—Infrastructure networks, such as power grids, have been developing rapidly. There is a growing demand for predicting the future evolution process of power grids in the planning of infrastructure systems. As the structure of power grids evolves to be more complex, the grids can be studied under complex network theory by extracting structural properties. Accordingly, the prediction of the evolution process of power grids can be regarded as a link prediction problem. To correspond with the actual self-organization regulations of power grids, a novel link prediction model based on preferential attachment is proposed in this brief. After experimenting on several real world power grids, results show that our model outperforms six other popular models in various kinds of realistic networks. It is demonstrated as a possible methodology for solving the prediction problem of power grid evolution.**

***Index Terms*—Power grid, complex network, link prediction, preferential attachment.**

## I. Introduction

INFRASTRUCTURE networks, such as power grids, have been developing rapidly in modern society, with a growing structural complexity and a requirement of reliability [1], [2]. As this trend will continue in foreseeable future, it is crucial to predict the future evolution process of power grids, which can provide guidance for the planning process of infrastructure systems to meet the design principle of power grid stability and high power quality [3].

The complex evolution process of power grids can be simplified as a problem of adding new transmission lines between current stations. To facilitate study of the rapidly developing power grids, a novel perspective is to apply complex network theory to power grid structural characteristic analysis, which regards power grids as complex networks and extracts their structural properties [4]–[6]. The method of regarding real world networks as complex networks can also be applied in biological networks [7], contributing to an insightful analysis into biological field. As the evolution of complex networks tends to show self-organization characteristics [8]–[10], current structural properties of power grids can be utilized to reveal regulations of the future evolution process [11]. As an example, the preferential attachment mechanism is at the heart of self-organization across social and natural sciences [10]. Therefore, complex network theory can be adopted to solve the prediction problem of power grid evolution.

Research has been focused on predicting possible links in complex network theory, constituting a branch of link prediction problem [12]. Link prediction problem aims at estimating the connection likelihood of node pairs, including the prediction of both unknown links and future possible links [12], [13]. Through link prediction methods, a score is assigned to each unknown link, with a higher score representing a higher connection possibility [13], [14]. In the evolution process of power grids, there is also a demand of predicting the location of newly-constructed transmission lines. Accordingly, it is feasible to score the unknown links in power grids using link prediction methods, regarding the links with higher scores (or predicted to be existed) as more likely to form in future evolution process. In general, link prediction puts forward a solution to the prediction problem of power grid evolution.

Current methods of link prediction can be applied to power grids by analyzing different structural features [15]–[19]. To make the evolution process more consistent with practical circumstances, a prediction method should utilize features or mechanisms that correspond with actual power grids as closely as possible. Previous research has found that power grids tend to show some scale-free properties of the degree or betweenness distribution [4], [8]. As a result, we improve the original neighbor set information (NSI) model [18] by considering the preferential attachment mechanism and design a new neighbor set information based on preferential attachment (NSIPA) model, in order to fit with the actual properties of power grids and obtain more accurate prediction results. The prediction method can put forward links more likely to be connected in future evolution process of power grids. With cost factors involving distance between stations and resource consumption remaining reasonable, this method can provide guidance for construction of new transmission lines in real world power systems. This brief is organized as follows. In part II, the link prediction problem in power grids is introduced, and our new NSIPA model is proposed to solve the problem. In part III, the experimental framework and setup is described. Finally, the performance of our model is

Manuscript received November 28, 2018; revised February 5, 2019; accepted February 22, 2019. Date of publication March 7, 2019; date of current version November 1, 2019. This work was supported by the National Natural Science Foundation of China under Grant 61573310. This brief was recommended by Associate Editor H. H.-C. Iu. *(Corresponding author: Yongxiang Xia.)*

L. Mei, D. Tang, and T. Wang are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: wenbodu@buaa.edu.cn).

W. Du and X. Cao are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China, and also with the Hefei Innovation Research Institute, Beihang University, Hefei 230013, China.

Y. Xia is with the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: xiayx@zju.edu.cn).

Digital Object Identifier 10.1109/TCSII.2019.2903558

1549-7747 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

tested on several power grids and other kinds of real world networks.

## II. Problem Statement and a New Solution

### A. Link Prediction Problem in Power Grids

In this brief, power grids can be analyzed under complex network theory. A power grid can be considered as an undirected network $G(V, E)$, where $V$ is the set of generators, bus bars and loads, and $E$ is the set of transmission lines. In complex network perspective, $V$ and $E$ represent node set and link set, respectively. Self-connections and multi-connections are neglected.

In link prediction problem, given a non-existent link $L_{xy} \in U - E$, where $x, y \in V$ and $U$ is the universal possible link set, the goal is to provide a score $s_{xy}$ to evaluate the connecting likelihood. For each specific predictor, all non-existent links can be ranked by their scores. A node pair with a higher score is considered to be more likely to have a candidate link [20], [21].

To verify the prediction performance of different predictors, the observable links are randomly divided into two sets, *i.e.,* the training set $E^T$ and the probe set $E^P$. We regard $E^T$ as known topological information and $E^P$ as a set only used for testing the algorithms. Accordingly there are $E^T \bigcup E^P = E$ and $E^T \bigcap E^P = \emptyset$.

### B. Preferential Attachment Mechanism in Power Grids

Current methods that utilize different structural features can be applied to the link prediction problem of power grids [15]–[19]. Still, most of previous literature neglects the preferential attachment nature of real world networks [8], [22], [23]. In many real world networks, the connecting possibilities of nodes follow the preferential attachment mechanism. The mechanism, first proposed by Albert and Barabási in their scale-free network model [8], can be used in analyzing a large proportion of real word networks.

Power grids tend to show preferential attachment nature, at least partially. Previous research has discovered some scale-free properties of the degree or betweenness distribution of U.S. power grids [4], [8], [24]. It is also illustrated that the growth of power grids at least partially follows the preferential attachment mechanism [25]. Therefore it is natural to introduce the preferential attachment mechanism in prediction of power grids.

### C. Deviation of a New Method

Here, we put forward a novel link prediction method, a neighbor set information model based on preferential attachment (NSIPA).

On the basis of the neighbor set information (NSI) prediction index in [18], NSIPA introduces a weight coefficient to represent the effect of preferential attachment. Central to the preferential attachment mechanism is a correlation between the probability $\Pi_i$ of a new node connecting to the existing node $i$ and the degree $k_i$ of node $i$, presented as

$$\Pi_i = \frac{k_i}{\sum_j k_j}. \tag{1}$$

In a network $G(V, E)$, recording the number of nodes as $|V| = N$ and the number of edges as $|E| = M$, the weight coefficient $w_{xy}|G$ indicating the effect of preferential attachment on the connecting probability of node pair $(x, y)$ in NSIPA can be calculated as

$$w_{xy}|G = c \cdot \left[ \frac{1}{N} \cdot \frac{k_x + 1}{\sum_{j \in V}(k_j + 1)} + \frac{1}{N} \cdot \frac{k_y + 1}{\sum_{j \in V}(k_j + 1)} \right], \tag{2}$$

where $c$ is the normalized coefficient to satisfy the following constraint that

$$\sum_{x \neq y, x \in V, y \in V} w_{xy}|G = 1. \tag{3}$$

With concepts from network science, it can be deduced that

$$w_{xy}|G = \frac{k_x + k_y + 2}{(N-1) \cdot (2M + N)}. \tag{4}$$

Considering common neighbors, the prediction score of $L_{xy}$ is denoted as $s_{xy}^{\Omega} = \sum_{\omega \in \Omega} (I(L_{xy}^1) - I(L_{xy}^1|\omega)) - I(L_{xy}^1)$, which is similar to MI index. However, $p(L_{xy}^1)$ and $p(L_{xy}^1|\omega)$ are modified by multiplying the weight coefficient in (4) that

$$p(L_{xy}^1) = \frac{|E|}{|V|(|V|-1)/2} \cdot \frac{k_x + k_y + 2}{(N-1) \cdot (2M + N)}, \tag{5}$$

$$p\left(L_{xy}^1|\omega\right) = \frac{N_{\triangle\omega}}{N_{\triangle\omega} + N_{\wedge\omega}} \cdot \frac{k_{x\omega} + k_{y\omega} + 2}{(N_\omega - 1) \cdot (2M_\omega + N_\omega)}, \tag{6}$$

where $G_\omega(V_\omega, E_\omega)$ is the sub-network composed of the neighbors of $\omega(\Gamma(\omega))$ and links $L_{xy}$ $(x, y \in \Gamma(\omega))$. In (6), $k_{x\omega}$ and $k_{y\omega}$ are defined as degree of $x$ and $y$ in sub-network $G_\omega(V_\omega, E_\omega)$, without considering the links outside $G_\omega$. Similarly, considering the links across neighbor sets, $p(L_{xy}^1|l_{st})$ is modified as

$$p\left(L_{xy}^1|l_{st}\right) = \frac{N_{\square st}}{N_{\square st} + N_{\sqcap st}} \cdot \frac{k_{x\omega} + k_{y\omega} + 2}{(N_\omega - 1) \cdot (2M_\omega + N_\omega)}. \tag{7}$$

With the definitions $I(L_{xy}^1) = -\log(p(L_{xy}^1))$ and $I(L_{xy}^1|\omega) = -\log(p(L_{xy}^1|\omega))$ in information theory, NSIPA model of link prediction can be defined as

$$s_{xy}^{NSIPA} = -I(L_{xy}^1|\Omega) - \lambda I\left(L_{xy}^1|P_{xy}\right), \tag{8}$$

where

$$I(L_{xy}^1|\Omega) = I\left(L_{xy}^1\right) - \sum_{\omega \in \Omega}\left(I\left(L_{xy}^1\right) - I\left(L_{xy}^1|\omega\right)\right), \tag{9}$$

$$I(L_{xy}^1|P_{xy}) = I\left(L_{xy}^1\right) - \sum_{\omega \in \Omega}\left(I\left(L_{xy}^1\right) - I\left(L_{xy}^1|l_{st}\right)\right), \tag{10}$$

in which $p(L_{xy}^1)$, $p(L_{xy}^1|\omega)$ and $p(L_{xy}^1|l_{st})$ are calculated in (5), (6) and (7), different from those in NSI index.

## III. Experimental Framework and Setup

### A. Supervised Learning Framework

Link prediction can be regarded as a binary classification problem, and thus can be solved under supervised learning framework [14], [26]. Liben-Nowell and Kleinberg first applied supervised learning framework to link prediction problems in 2003 [12]. Under supervised learning framework,

network-structure-based prediction indices can be used to train different classifiers according to machine learning algorithms. These classifiers can then be used to predict the presence of unknown links. As the supervised learning framework can treat any topological property or any other link prediction measure as a training index [14], it is possessed of a strong applicability and can be easily studied by machine learning software. In this brief, we also adopt the supervised learning framework as in [12] and [26]–[28].

In the supervised learning framework, the existed links in current network are regarded as positive samples, each labeled as "1". Meanwhile, missing links that cannot be observed are regarded as negative samples, each labeled as "0". If there are more non-existent links than edges, all $|E|$ edges are marked as positive samples, and $|E|$ non-existent links are extracted as negative samples. The whole sample set is composed of both positive and negative samples.

Then, 10-fold cross validation method [29] is applied to randomly divide the whole sample set into 10 parts. One part is selected to be the probe set, while the remaining 9 parts constitute the training set. The above process is repeated 10 times to ensure that each part has been treated as probe set for once. In training process, all links existed in probe set are recognized as non-existent to avoid possible data leakage.

In the supervised learning part of our experiment, Waikato Environment for Knowledge Analysis (WEKA) [30] is used as the learning platform, and C4.5 (J48) decision tree algorithm is adopted because it is an optimal choice of machine learning algorithms in the link prediction problem [26]. The minimum number of instances in each leaf node is set as 10, while other parameters keep the default values in WEKA.

### B. Link Prediction Methods for Comparison

In previous studies, link prediction methods involve considerations of different network features. We select 3 common CN-based methods and 3 recently-proposed IT-based methods to compare with our NSIPA model in the experiments. Methods for comparison include Common Neighbors (CN), Adamic-Adar (AA) [15], Resource Allocation (RA) [16], Mutual Information (MI) [17], Allocated Mutual Information (AMI) and Neighbor Set Information (NSI) [18]. Here we turn the original NSIA index [19] into AMI index by only considering the common neighbor feature in order for the convenience when comparing it with the MI index.

### C. Evaluation Metrics

In this brief, two widely used evaluation metrics, AUC (area under the receiver operating characteristic curve) and Accuracy, are used to quantify the performance of prediction methods.

In link prediction problem under supervised learning framework, the prediction results can be classified into four categories, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). "True" or "False" denotes the accuracy of prediction, while "Positive" or "Negative" denotes the result of prediction. For instance, TP denotes the case that a positive sample is truly predicted to be positive. The True Positive Rate (TPR) and False Positive Rate (FPR) are defined as

$$TPR = \frac{TP}{TP + FN}, \tag{11}$$

$$FPR = \frac{FP}{FP + TN}. \tag{12}$$

With the definitions above, AUC metric and Accuracy metric are illustrated as follows.

*AUC Metric:* AUC metric is defined as the area under ROC (Receiver Operating Characteristic curve). It is also known as the probability of the predicted score of a randomly-chosen positive sample being higher than the predicted score of a randomly-chosen negative sample [31]. The classifier will perform better if the AUC is larger.

*Accuracy Metric:* Accuracy is defined as the proportion of correctly predicted instances in all instances [28], presented as

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}. \tag{13}$$

TABLE I
THE STRUCTURAL PARAMETERS OF THE POWER GRID NETWORKS

| Network \ Index | $N$ | $M$ | $C$ | $r$ | $\langle k \rangle$ | $H$ |
|---|---|---|---|---|---|---|
| USPower | 4941 | 6594 | 0.107 | 0.004 | 2.669 | 1.450 |
| EUPower | 1494 | 2156 | 0.106 | -0.121 | 2.886 | 1.370 |
| AusPower | 912 | 1211 | 0.123 | -0.056 | 2.656 | 1.616 |
| ENTPower | 7893 | 9235 | 0.062 | -0.011 | 2.340 | 1.261 |

$N$ and $M$ denote the number of nodes and links respectively. $C$ represents the average clustering coefficient. $r$ stands for the assortative coefficient. $\langle k \rangle$ denotes the average degree. $H$ represents the degree heterogeneity, defined as $H = \frac{\langle k^2 \rangle}{\langle k \rangle}$.

## IV. RESULTS

### A. Power Grid Data

The power grids used as data sets, denoted as USPower [32], EUPower [33], AusPower [34] and ENTPower [35], are electrical grids of the western of U.S., the continental European electricity system, Australia's largest electricity transmission network, and the Gridkit extract of the ENTSO-E interactive map of the European power system (including North Africa and the Middle East to a limited extent), respectively. Their structural parameters are listed in Table I. It is noticeable that the structural features of different power grids are similar, such as their average clustering coefficient, average degree and degree heterogeneity.

### B. Experimental Results and Analysis

First, NSIPA model is tested on 4 power grid data sets. To better evaluate the performance of the NSIPA model in link prediction, we compare NSIPA with other 6 indices. We test the NSI and NSIPA indices by adjusting $\lambda$ from 0 to 1.0 with a difference of 0.1, and choose $\lambda = 0.1$ which is demonstrated as the best choice in NSI [18] and shows an overall good performance in NSIPA. Results under different metrics are shown in Table II and Table III.

In all 4 power grids, NSIPA provides an overall best performance. Owing to the preferential attachment nature of

TABLE II
PREDICTION RESULTS OF SEVEN METHODS ON POWER GRIDS MEASURED BY AUC METRIC. EACH VALUE IS OBTAINED BY AVERAGING RESULTS OF 10 IMPLEMENTATIONS IN 10-FOLD CROSS-VALIDATION. THE BEST PERFORMANCE IS PRESENTED IN BOLD FONT. THE STANDARD DEVIATIONS OF 10 IMPLEMENTATIONS ARE PRESENTED INSIDE PARENTHESES

| Network\ AUC | CN | AA | RA | MI | AMI | NSI λ=0.1 | NSIPA λ=0.1 |
|---|---|---|---|---|---|---|---|
| USPower | 0.5863 (0.006) | 0.5863 (0.006) | 0.5863 (0.006) | 0.5509 (0.004) | 0.5509 (0.004) | 0.6051 (0.006) | **0.6079 (0.006)** |
| EUPower | 0.6001 (0.008) | 0.6001 (0.008) | 0.6001 (0.008) | 0.5418 (0.008) | 0.5418 (0.008) | 0.6288 (0.017) | **0.6306 (0.015)** |
| AusPower | 0.6245 (0.018) | 0.6245 (0.018) | 0.6245 (0.018) | 0.5617 (0.014) | 0.5617 (0.014) | 0.6229 (0.016) | **0.6338 (0.008)** |
| ENTPower | 0.5596 (0.006) | 0.5596 (0.006) | 0.5596 (0.006) | 0.5154 (0.002) | 0.5154 (0.002) | 0.5703 (0.005) | **0.5780 (0.007)** |

TABLE III
PREDICTION RESULTS OF SEVEN METHODS ON POWER GRIDS MEASURED BY ACCURACY METRIC. EACH VALUE IS OBTAINED BY AVERAGING RESULTS OF 10 IMPLEMENTATIONS IN 10-FOLD CROSS-VALIDATION. THE BEST PERFORMANCE IS PRESENTED IN BOLD FONT. THE STANDARD DEVIATIONS OF 10 IMPLEMENTATIONS ARE PRESENTED INSIDE PARENTHESES

| Network\ Accuracy (%) | CN | AA | RA | MI | AMI | NSI λ=0.1 | NSIPA λ=0.1 |
|---|---|---|---|---|---|---|---|
| USPower | 58.644 (1.112) | 58.636 (1.123) | 58.636 (1.123) | 55.096 (1.053) | 55.096 (1.053) | 60.525 (1.144) | **60.767 (1.089)** |
| EUPower | 59.997 (2.692) | 59.997 (2.692) | 59.997 (2.692) | 54.177 (3.013) | 54.177 (3.013) | 62.850 (3.303) | **62.919 (2.938)** |
| AusPower | 62.468 (2.534) | 62.468 (2.534) | 62.468 (2.534) | 56.194 (2.969) | 56.194 (2.969) | 62.388 (2.148) | **63.339 (2.349)** |
| ENTPower | 55.988 (1.008) | 55.988 (1.008) | 55.988 (1.008) | 51.527 (1.281) | 51.527 (1.281) | 57.011 (1.326) | **57.997 (1.430)** |

power grids described in the second part, NSIPA provides better prediction performance than other indices in all 4 networks under both AUC and Accuracy. Though the advantage of NSIPA is limited by real world restrictions of power grids, such as distance between stations and maximum power load of stations, the improvements are still acceptable in link prediction field.

To prove the universality of NSIPA model, it is tested on other networks apart from power grids, which are listed as follows, Email [36], EPA [37], Everglades [37], Karate [38], Kohonen [37], Router [39], SciMet [37] and SmaGri [37] (in SmaGri, only the giant component is picked). Unlike power grids, these networks contain different structural features. For instance, the degree heterogeneity of EPA, Router, Kohonen and SmaGri are higher ($H>5$), while those of Email, Everglades, Karate and SciMet are lower ($H<3$). Results under different metrics in these networks are shown in Table IV and Table V.

When $\lambda = 0.1$, in networks such as EPA, Router, Kohonen and SmaGri, NSIPA apparently wins over other indices. Such an obvious increase of prediction accuracy is due to the scale-free nature and the high degree heterogeneity of these networks.

On the other hand, in networks with low degree heterogeneity, the differences between NSIPA and the optimal index apart from NSIPA are relatively small comparing to the standard deviations. However, NSIPA still provides acceptable performance comparing to the optimal index apart from NSIPA. It is also noticeable that NSIPA performs better on

TABLE IV
PREDICTION RESULTS OF SEVEN METHODS ON OTHER NETWORKS MEASURED BY AUC METRIC. EACH VALUE IS OBTAINED BY AVERAGING RESULTS OF 10 IMPLEMENTATIONS IN 10-FOLD CROSS-VALIDATION. THE BEST PERFORMANCE IS PRESENTED IN BOLD FONT. THE STANDARD DEVIATIONS OF 10 IMPLEMENTATIONS ARE PRESENTED INSIDE PARENTHESES

| Network\ AUC | CN | AA | RA | MI | AMI | NSI λ=0.1 | NSIPA λ=0.1 |
|---|---|---|---|---|---|---|---|
| Email | 0.8279 (0.014) | 0.8279 (0.014) | 0.8279 (0.014) | 0.8263 (0.013) | 0.8294 (0.012) | 0.8563 (0.013) | **0.8640 (0.010)** |
| EPA | 0.5846 (0.005) | 0.5857 (0.004) | 0.5857 (0.004) | 0.5721 (0.003) | 0.5705 (0.004) | 0.7661 (0.009) | **0.8922 (0.008)** |
| Everglades | 0.6492 (0.035) | 0.6449 (0.033) | 0.6341 (0.027) | 0.6722 (0.029) | 0.6604 (0.025) | **0.7306 (0.029)** | 0.7283 (0.033) |
| Karate | 0.5927 (0.107) | 0.5621 (0.093) | 0.6034 (0.127) | 0.5814 (0.124) | 0.5740 (0.117) | 0.5658 (0.108) | **0.6042 (0.081)** |
| Kohonen | 0.7939 (0.005) | 0.8103 (0.006) | 0.8105 (0.006) | 0.7842 (0.005) | 0.7831 (0.006) | 0.8451 (0.013) | **0.8615 (0.009)** |
| Router | 0.5603 (0.005) | 0.5608 (0.005) | 0.5608 (0.005) | 0.5578 (0.006) | 0.5579 (0.006) | 0.6311 (0.011) | **0.7371 (0.009)** |
| SciMet | 0.7830 (0.005) | 0.7830 (0.005) | 0.7830 (0.005) | 0.7753 (0.005) | 0.7738 (0.005) | **0.8740 (0.012)** | 0.8647 (0.008) |
| SmaGri | 0.8020 (0.021) | 0.8239 (0.016) | 0.8161 (0.016) | 0.8378 (0.018) | 0.8287 (0.015) | 0.8166 (0.020) | **0.8382 (0.020)** |

TABLE V
PREDICTION RESULTS OF SEVEN METHODS ON OTHER NETWORKS MEASURED BY ACCURACY METRIC. EACH VALUE IS OBTAINED BY AVERAGING RESULTS OF 10 IMPLEMENTATIONS IN 10-FOLD CROSS-VALIDATION. THE BEST PERFORMANCE IS PRESENTED IN BOLD FONT. THE STANDARD DEVIATIONS OF 10 IMPLEMENTATIONS ARE PRESENTED INSIDE PARENTHESES

| Network\ Accuracy (%) | CN | AA | RA | MI | AMI | NSI λ=0.1 | NSIPA λ=0.1 |
|---|---|---|---|---|---|---|---|
| Email | 82.777 (1.279) | 82.777 (1.279) | 82.777 (1.279) | 82.316 (1.185) | 82.429 (1.312) | **84.227 (1.001)** | 84.098 (1.323) |
| EPA | 58.453 (1.000) | 58.548 (1.004) | 58.548 (1.004) | 56.873 (1.141) | 56.868 (1.050) | 76.475 (0.908) | **86.810 (0.692)** |
| Everglades | 64.915 (3.861) | 64.350 (3.270) | 63.559 (3.138) | 66.610 (3.016) | 64.859 (2.299) | **73.164 (2.750)** | 72.881 (3.501) |
| Karate | 61.048 (9.413) | 55.714 (7.319) | **62.191 (12.65)** | 59.905 (14.03) | 57.905 (12.35) | 55.333 (9.638) | 54.667 (8.327) |
| Kohonen | 79.389 (0.632) | 79.661 (0.670) | 79.645 (0.670) | 78.000 (0.680) | 77.874 (0.701) | 83.109 (0.826) | **84.122 (0.611)** |
| Router | 56.024 (1.005) | 56.072 (0.970) | 56.072 (0.970) | 55.784 (0.938) | 55.792 (0.950) | 62.847 (1.117) | **71.820 (0.993)** |
| SciMet | 78.301 (0.647) | 78.301 (0.647) | 78.301 (0.647) | 77.274 (0.652) | 77.284 (0.618) | **84.550 (0.781)** | 83.745 (0.590) |
| SmaGri | 80.156 (2.185) | 82.346 (1.683) | 81.560 (1.651) | 81.560 (1.730) | 81.627 (1.804) | 81.460 (2.196) | **82.430 (2.019)** |

power grids than on the other four networks with similar low degree heterogeneity, which reflects that the improvements of NSIPA in power grids are acceptable, and NSIPA is suitable for prediction in power grids.

Altogether, NSIPA model is tested on 4 power grids and a variety of other networks. As a result, NSIPA is proved to display a better overall performance in various kinds of real world networks, especially in power grid networks and other networks with high degree heterogeneity. When degree heterogeneity is relatively low, NSIPA can still perform at least equally well comparing to the optimal index in most of the networks. In conclusion, NSIPA can achieve advantage in power grids and other networks with scale-free nature.

## V. DISCUSSION

In this brief, a NSIPA link prediction model is put forward to solve the prediction problem in the evolution of power

grids. Our NSIPA model provides better prediction results in power grids and various other networks, especially those with scale-free nature. Meanwhile, our NSIPA model shows similar computing efficiency to the original NSI index, for NSIPA only introduces a constant weight coefficient on the basis of NSI, making their complexity belonging to the same order of magnitude. In our experiments, the consuming time of NSIPA model is almost the same as NSI. Considering the improvement and the similar computing efficiency, NSIPA can provide a possible solution to the prediction problem of power grids evolution, and benefit the planning process of infrastructure systems. Besides, our model can be applied in traditional fields of link prediction, including facilitating the study of network models and their evolution mechanisms [20], predicting future hashtags [40] and providing recommendation services [41] in social networks, and predicting possible interactions in biological networks [7].

Moreover, there is still room for improvement. Utilizing the property of preferential attachment, NSIPA displays acceptable scale of improvements in power grids. To obtain a more obvious improvement, further research could be carried out on combining network science with real world properties of power grids, like introducing a coefficient representing the geographical distance between stations or the real world properties of stations such as maximum power load. With more real world properties considered, the model can be modified to show better results.

## REFERENCES


[1] A. Monti and F. Ponci, "Power grids of the future: Why smart means complex," in *Proc. Complex. Eng.*, Rome, Italy, 2010, pp. 7–11.

[2] M. G. Simoes *et al.*, "A comparison of smart grid technologies and progresses in Europe and the U.S.," *IEEE Trans. Ind. Appl.*, vol. 48, no. 4, pp. 1154–1162, Jul./Aug. 2012.

[3] D. Ciechanowicz, D. Pelzer, B. Bartenschlager, and A. Knoll, "A modular power system planning and power flow simulation framework for generating and evaluating power network models," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2214–2224, May 2017.

[4] G. A. Pagani and M. Aiello, "From the grid to the smart grid, topologically," *Physica A Stat. Mech. Appl.*, vol. 449, no. 1, pp. 160–175, May 2016.

[5] B. Liu, Z. Li, X. Chen, Y. Huang, and X. Liu, "Recognition and vulnerability analysis of key nodes in power grid based on complex network centrality," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 3, pp. 346–350, Mar. 2018.

[6] Z. Chen, J. Wu, Y. Xia, and X. Zhang, "Robustness of interdependent power grids and communication networks: A complex network perspective," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 1, pp. 115–119, Jan. 2018.

[7] M. Gosak *et al.*, "Network science of biological systems at different scales: A review," *Phys. Life Rev.*, vol. 24, pp. 118–135, Mar. 2018.

[8] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, Jan. 2002.

[9] M. Perc and A. Szolnoki, "Coevolutionary games—A mini review," *Biosystems*, vol. 99, no. 2, pp. 109–125, Feb. 2010.

[10] M. Perc, "The Matthew effect in empirical data," *J. Roy. Soc. Interface*, vol. 11, no. 98, Sep. 2014, Art. no. 20140378.

[11] M. Jalili and M. Perc, "Information cascades in complex networks," *J. Complex Netw.*, vol. 5, no. 5, pp. 665–693, Oct. 2017.

[12] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," *J. Assoc. Inf. Sci. Tech.*, vol. 58, no. 7, pp. 1019–1031, May 2007.

[13] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Physica A Stat. Mech. Appl.*, vol. 390, no. 6, pp. 1150–1170, 2011.

[14] V. Martínez, F. Berzal, and J.-C. Cubero, "A survey of link prediction in complex networks," *ACM Comput. Surveys*, vol. 49, no. 4, pp. 1–69, Feb. 2017.

[15] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," *Soc. Netw.*, vol. 25, no. 3, pp. 211–230, 2003.

[16] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information," *Eur. Phys. J. B*, vol. 71, no. 4, pp. 623–630, 2009.

[17] F. Tan, Y. Xia, and B. Zhu, "Link prediction in complex networks: A mutual information perspective," *PLoS ONE*, vol. 9, no. 9, Sep. 2014, Art. no. e107056.

[18] B. Zhu and Y. Xia, "An information-theoretic model for link prediction in complex networks," *Sci. Rep.*, vol. 5, Sep. 2015, Art. no. 13707.

[19] P. Pei, B. Liu, and L. Jiao, "Link prediction in complex networks based on an information allocation index," *Physica A Stat. Mech. Appl.*, vol. 470, pp. 1–11, Mar. 2017.

[20] W.-Q. Wang, Q.-M. Zhang, and T. Zhou, "Evaluating network models: A likelihood analysis," *Europhys. Lett.*, vol. 98, no. 2, 2012, Art. no. 28004.

[21] Z. Liu, Q.-M. Zhang, L. Lü, and T. Zhou, "Link prediction in complex networks: A local naïve Bayes model," *Europhys. Lett.*, vol. 96, no. 4, 2011, Art. no. 48007.

[22] M.-Y. Wang, G. Yu, and D.-R. Yu, "The scale-free model for citation network," in *Proc. IEEE Int. Conf. Intell. Comput. Intell. Syst.*, Xiamen, China, 2010, pp. 773–776.

[23] H. Li and J. Lu, "A novel scale-free network model with accelerating growth," in *Proc. IEEE Int. Symp. Circuits Syst.*, Taipei, Taiwan, 2009, pp. 1693–1696.

[24] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the North American power grid," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 2, Jan. 2004, Art. no. 025103.

[25] K. Wang, "Research on structural complexity and vulnerability of power grids based on complex network theory," Ph.D. dissertation, Dept. Power Syst. Autom., Huazhong Univ. Sci. Technol., Wuhan, China, 2011.

[26] M. Fire *et al.*, "Link prediction in social networks using computationally efficient topological features," in *Proc. IEEE 3rd Int. Conf. Privacy Security Risk Trust IEEE 3rd Int. Conf. Soc. Comput.*, Boston, MA, USA, 2011, pp. 73–80.

[27] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *Proc. SDM Workshop Link Anal. Counterterrorism Security*, 2006, pp. 1–10.

[28] N. Torabi, H. Shakibian, and N. M. Charkari, "An ensemble classifier for link prediction in location based social network," in *Proc. 24th Iran. Conf. Elect. Eng. (ICEE)*, Shiraz, Iran, 2016, pp. 529–532.

[29] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell.*, vol. 2. Montreal, QC, Canada, Aug. 1995, pp. 1137–1143.

[30] M. Hall *et al.*, "The WEKA data mining software: An update," *SIGKDD Explor. Newslett.*, vol. 11, no. 1, pp. 10–18, Jun. 2009.

[31] T. Fawcett, "An introduction to ROC analysis," *Pattern Recogn. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[32] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.

[33] N. Hutcheon and J. W. Bialek, "Updated and validated power flow model of the main continental European transmission network," in *Proc. IEEE Grenoble Conf.*, Grenoble, France, 2013, pp. 1–5.

[34] A. Xenophon and D. Hill, *Geospatial Modelling of Australia's National Electricity Market—Dataset (Version v1.3)*, Zenodo, Geneva, Switzerland, 2018. [Online]. Available: http://doi.org/10.5281/zenodo.1326942

[35] B. Wiegmans, *GridKit Extract of ENTSO-E Interactive Map*, Zenodo, Geneva, Switzerland, 2016. [Online]. Available: http://doi.org/10.5281/zenodo.55853

[36] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 72, no. 2, Aug. 2005, Art. no. 027104.

[37] V. Batagelj and A. Mrvar. (2006). *Pajek DataSets*. [Online]. Available: http://vlado.fmf.uni-lj.si/pub/networks/data/

[38] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. 452–473, Dec. 1977.

[39] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP topologies with rocketfuel," *IEEE/ACM Trans. Netw.*, vol. 12, no. 1, pp. 2–16, Feb. 2004.

[40] S. Martinčić-Ipšić, E. Mocibob, and M. Perc, "Link prediction on Twitter," *PLoS ONE*, vol. 12, no. 7, 2017, Art. no. e0181079.

[41] M. Jalili, Y. Orouskhani, M. Asgari, N. Alipourfard, and M. Perc, "Link prediction in multiplex online social networks," *Roy. Soc. Open Sci.*, vol. 4, no. 2, 2017, Art. no. 160863.