2020 IEEE PES Innovative Smart Grid Technologies Conference,
Washington, DC, February 17-20, 2020

PMU Visibility Graphs

Laith Shalalfeh
Dept. of Energy Engineering
German Jordanian University
Amman, Jordan
laith.shalalfeh@gju.edu.jo

Paul Bogdan
Dept. of Electrical Engineering
University of Southern California
Los Angeles, California 90089
pbogdan@usc.edu

Edmond Jonckheere
Dept. of Electrical Engineering
University of Southern California
Los Angeles, California 90089
jonckhee@usc.edu

***Abstract*—In this paper, we map the PMU data to a graph via the visibility algorithm to generate the PMU visibility graph. Applying complex network analysis to this graph unravels several hidden features in the PMU data. The applications of the PMU visibility graphs in the real power grid include anomaly detection, topology identification, and state estimation. We show that the PMU visibility graph is not random and could not be modeled by the classical Erdős-Rényi model. On the contrary, the graph represents a scale-free network with a heavy-tailed degree distribution. So, we fit its degree distribution to a power-law distribution to find the best model and we perform goodness-of-fit analysis on the estimated distribution. Lastly, it is argued that power-law degree distribution of the PMU visibility graph reveals the nonstationarity and fractality in the PMU data.**

I. Introduction

Deciphering the unknown unknowns (UUs) (e.g., hidden statistical characteristics and conservation principles) that characterize the dynamics of complex systems promises to provide efficient algorithms for anomaly detection, self-healing and self-optimizing in engineered Artificial Intelligence (AI)-based systems. As a prominent example of an evolving complex system, engineering the autonomous smart grid requires a deep understanding of the real-time wide area monitoring (WAM) data (e.g., phasor measurement unit (PMU) signals) in order to provide a higher degree of observability, dependability, security and reliability while overcoming data science challenges (e.g., incomplete, heterogeneous, multi-modal and noisy data streams).

Since the introduction of first PMU in 2005, the PMU data has been used in several applications, like fault detection, protection and control, and load modeling. Due to the advantages of PMUs, an increasing number of PMUs are being deployed in the power grid all over the world. There are approximately 2000 PMUs installed across North America based on North American SynchroPhasor Initiative (NASPI) report [1].

Massive amount of data is currently collected by phasor measurement units and smart meters from all over the grid. The data is exceeding the traditional storage and processing capacities of the power utility. In [2] and [3], the authors point to the need of classifying the PMU data under the category of “Big Data". That is justifiable due to the high volume, high traffic, and variability of the collected data.

Extracting the UUs and the hidden laws of "healthy" power grid operation needs advanced data science techniques. This is an important stage to gain knowledge and insight about the PMU data [4]. As part of the development of the smart grid, it is urgent to build fast and novel algorithms capable of processing the data and help in decision making in several facets of the power system. That includes fault detection, predictive maintenance, transient stability, topology identification [5].

The hidden features and statistical characteristics in data can be revealed by transforming the time series to a graph or complex network [6][7]. In [7], the authors introduce the mapping of pseudoperiodic time series to complex network where the vertices are the cycles and the edges are abstracting the temporal correlation between cycles. A generalized mapping was defined in [6] by transforming the time series to a visibility graph. In the visibility graph, the vertices are the samples in the time series and the edges exist if their corresponding samples are line-of-sight visible despite the "walls" from the time samples due to the signal value samples.

We have recently shown existence of long-range dependence in PMU data using Detrended Fluctuation Analysis (DFA) [8] and Fractional ARIMA (ARFIMA) models [9]. In [10], the authors assert that visibility graph degree distribution of fractional Brownian motion (fBm) follows a power law. Moreover, a relationship between the Hurst exponent of fBm and scaling parameter of the power law was derived. However, the estimation of the scaling parameter of the degree distribution was inaccurate due to the use of continuous power-law formula instead of the discrete one. In addition, goodness-of-fit analysis was not performed to evaluate the power-law fit.

In this paper, we generate the visibility graphs of a large data set of PMU data (voltage magnitude, frequency, phase angle) and we show that their degree distributions have heavy tails. Then, we estimate the parameters of the best power-law fit using Maximum Likelihood Estimation (MLE). At the end, we perform goodness-of-fit analysis to test the plausibility of the power-law model.

The rest of the paper is organized as follows: in Sec. II, we introduce the visibility algorithm and apply it to PMU data. Sec. III shows the characteristics of the PMU visibility graph. In Sec. IV, we fit the degree distributions of the graphs to power-law distributions. Sec. V is the conclusion.

II. Natural Visibility Graph

In this section, we explain the criterion of mapping the time series to a graph using the visibility algorithm. We subsequently apply the algorithm on three sample time series of PMU data: voltage magnitude, frequency, and phase angle.

A. Visibility algorithm

The visibility algorithm transforms a time series into a graph by mapping time-stamped data points into specific vertices and drawing edges among these vertices if they have a direct link that does not intersect the height wall of any other data point in between.

Assuming a time series $x(t)$ of N data points, its corresponding graph consists of N vertices and any vertex pair (i, j) in the graph has an edge if the corresponding two data points, (t_i, x_i) and (t_j, x_j), are line-of-sight visible. The visibility between the two data points exists if all the data points (t_k, x_k) between data points i and j satisfy Eq. (1).

$$x_k < x_j + (x_i - x_j)\ \frac{t_j - t_k}{t_j - t_i} \tag{1}$$

To illustrate the visibility algorithm, assume we have a sample time series and its corresponding graph as in Fig. 1. All the data points, shown in Fig. 1(a), are mapped to vertices in the graph, as shown in Fig. 1(b). The links exist between any two data points if they are visible to each other by satisfying Eq. (1). These links are mapped to edges in the visibility graph. The resulting graph of the visibility algorithm is called natural visibility graph. This graph is invariant, undirected, and connected.

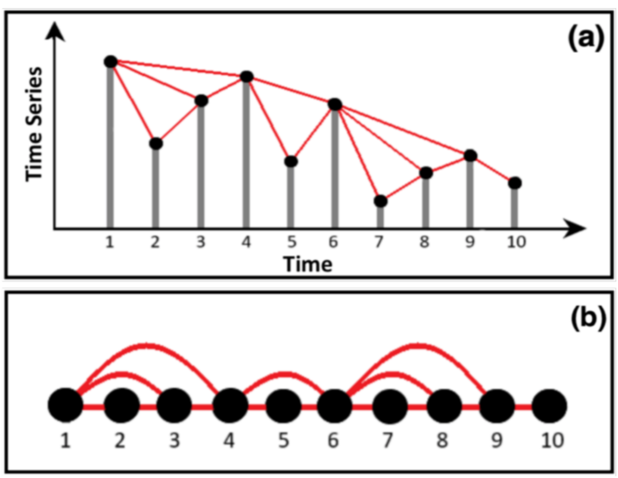

Fig. 1. From time series to a visibility graph

B. Visibility Graph of PMU data

Throughout this paper, we use real PMU data collected from the École Polytechnique Fédérale de Lausanne (EPFL) campus grid during the period 2014-2016 [11]. Three 100-sample time series of voltage magnitude (red), frequency (blue), and phase angle (green) are shown in Figs. 2 (a)-(c). The visibility algorithm was implemented by an open-source Fortran program [6] that computes the adjacency matrix of the visibility graph. The visibility graphs of three time series are shown in Figs. 2 (d)-(f).

The three visibility graphs of PMU data are undirected and connected graphs. The size of the vertex in the graph is proportional to the degree of the vertex. That means the large vertices have higher degrees than small ones. Moreover, the number inside the vertex refers to its corresponding data point in the time series.

In Fig. 2(a), we have the visibility graph of the voltage time series. The graph consists of 100 vertices and 412 edges. We can notice that vertices #55 and #75 have higher degrees compared to other vertices. That is justifiable because the sample #55 has high visibility of the samples to its right and sample #75 has high visibility of the samples to its left in the time series.

Fig. 2(b) shows the visibility graph corresponding to frequency time series. The graph has 100 vertices and 260 edges. Vertices #26, #54, and #72 in the graph have the highest degrees because of the high visibility of their corresponding samples in the time series. From the time series, it can be noticed that three samples have high values that enable them to have better visibility in both directions.

Fig. 2(c) shows the visibility graph corresponding to phase angle time series. The graph has 100 vertices and 334 edges. Vertices #5 and #7 have the highest degrees in the graph because their corresponding samples in the time series are located at the beginning of the series and have visibility to many samples to their right.

III. Characteristics of PMU Visibility Graph

We focus here on calculating the degree distribution of PMU visibility graph. Afterwards, we show that the PMU visibility graph looks like a scale-free network with a heavy-tailed degree distribution.

A. Random vs. Scale-Free Graphs

An undirected and connected graph, $G(V, E)$, has a set of vertices, V, and a set of edges, E. The degree, $k(v)$, of a vertex $v \in V$ represents the number of edges associated with that vertex. Furthermore, the probability that a vertex in the graph has a degree k is shown in Eq. (2). $P(k)$ is called the graph degree distribution. The degree distribution carries important information about the complexity of the graph structure.

$$P(k) = \frac{\#\text{ of vertices with degree } k}{\#\text{ of all vertices}} = \frac{N_k}{N} \tag{2}$$

In graph theory, a random graph is a graph that consists of vertices and edges where the presence of an edge between any two vertices is random [12]. The Erdős-Rényi [13] model $G(n, p)$ was the first to model random graphs where the presence of an edge between any two vertices has probability p. The degree distribution of Erdős-Rényi model is the binomial distribution which asymptotically converges to the Poisson distribution as the number of vertices (n) increases $(n \rightarrow \infty)$. In the Erdős-Rényi model, the vertices are more likely to have degrees around the mean (np) of the binomial distribution. The probability of having vertices with higher or smaller degrees decays as we move away from the mean.

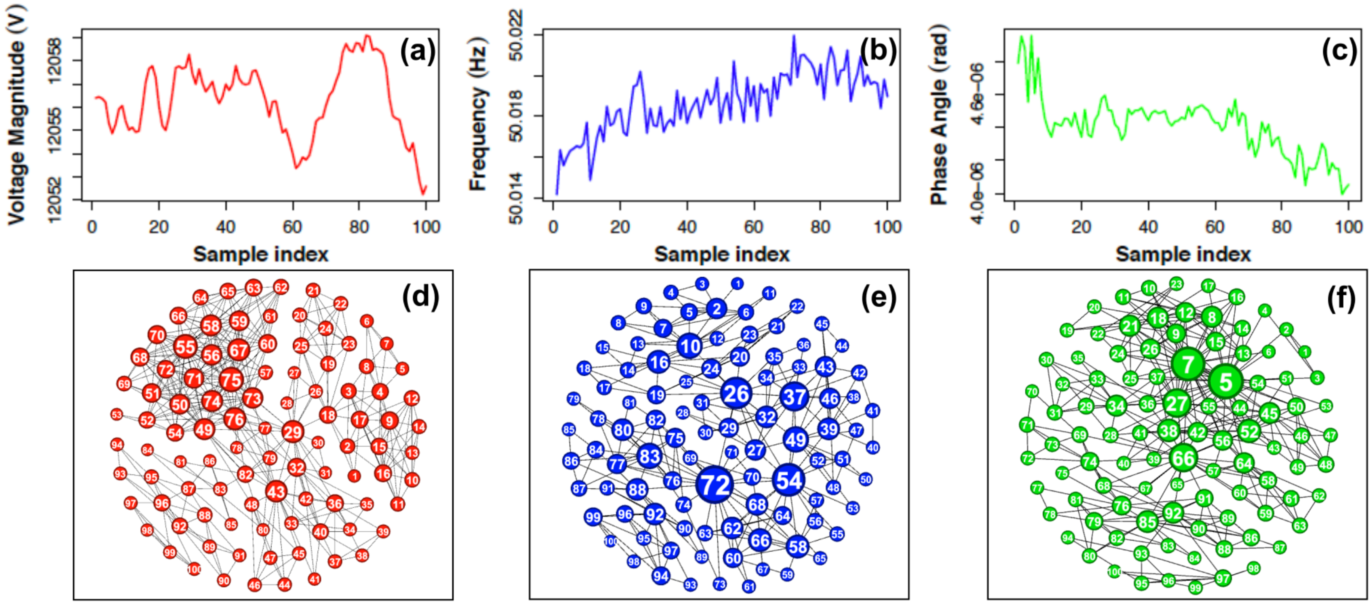

Fig. 2. The visibility graphs of three 100-sample time series of PMU data

Although the model was popular and able to model a few number of networks, it failed to model several other real-world networks. The degree distributions of these networks do not fit a binomial distribution, but they follow a power-law distribution ($p(k) \sim k^{-\alpha}$). The parameter α is called the scaling parameter.

Due to the shortcomings of the random graph model, the scale-free network was introduced by Barabási [14] to model graphs and networks that possess a power-law degree distribution. The scale-free network model has two important features, growth and preferential connectivity [14]. That means the new added vertices to the network prefer to connect to high-degree vertices rather than small-degree vertices. The high-degree vertices are known as hubs. These two properties yield the power-law behavior in the degree distribution.

B. Degree Distribution

To investigate the characteristics of PMU visibility graphs, we have chosen three 10^5-sample time series of PMU data from the EPFL campus grid, as shown in Figs. 3 (a)-(c). The linear and logarithmic plots of the degree distributions are shown in Figs. 3 (d)-(f). By ignoring the first and last vertices, the lowest degree of the visibility graph is 2 because each vertex is connected at least to its neighbor vertices.

The degree distributions indicate that the PMU visibility graphs have hybrid structure between random and scale-free graphs. The region of low-degrees in the distribution has vertices that have a tendency to connect randomly to other vertices similar to the random graph. These vertices correspond to the PMU data points with no large or sudden deviations from the rated value or the average in their neighborhoods. So, the vertices have low degrees because their data points do not have high visibility to other points in the series.

On the other hand, the right region of the degree distribution is a heavy-tail distribution that decays like a power law. The discrete form of a power-law distribution is defined by Eq. (3),

$$P(k) = \frac{k^{-\alpha}}{\zeta(\alpha, k_{min})}, \tag{3}$$

where $\zeta(.)$ is the Hurwitz zeta function as shown in Eq. (4),

$$\zeta(\alpha, k_{min}) = \sum_{n=0}^{\infty} (n + k_{min})^{-\alpha}. \tag{4}$$

α is the scaling parameter and k_{min} is the lower bound of the scaling region in the power law. The estimation of α and k_{min} of PMU visibility graphs is carried out in Sec. IV.

The tail of distribution represents the vertices that have high degrees and behave like hubs in scale-free graphs where other vertices have preference to connect to them. The corresponding data points of the hubs are part of large or sudden deviations from the rated value in the PMU data. These deviations enable them to have higher visibility compared to the other points in the series. The three degree distributions of PMU data are clearly having different scaling parameters (α) that could quantify the dynamics of the data.

IV. Scaling in PMU Visibility Graphs

We estimate the scaling parameter of the power law for several PMU visibility graphs. Then, we test the plausibility of the power-law hypothesis using goodness-of-fit analysis.

A. Estimation of the scaling parameter in a power-law distribution

It is clear that the degree distribution does not follow a power law over the full range of k. So, finding the range ($k > k_{min}$) in which the distribution closely follows a power law is necessary to have accurate estimation of the scaling

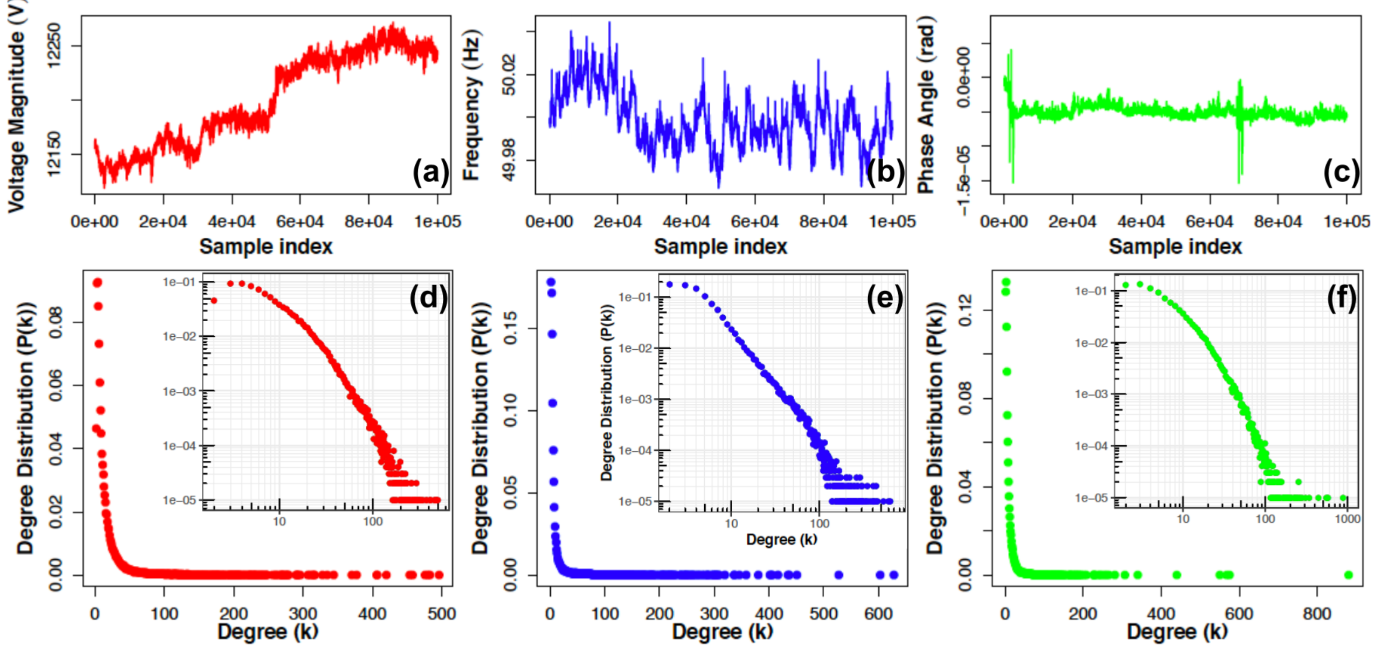

Fig. 3. Degree Distributions of PMU visibility graphs of three 100,000-sample time series

parameter. The empirical Complementary CDF (CCDF) of PMU visibility graph represents the probability of having a vertex with a degree higher than the degree k. The CCDF has also a power-law behavior that is more visually obvious in the empirical CCDF compared to the degree distribution.

The fitting of the empirical CCDF to a power-law distribution is performed via Maximum Liklihood Estimation (MLE) combined with Kolmogorov-Smirnov (KS) statistic [15]. Starting from $k_{min} = 2$, we use MLE to estimate the scaling parameter of the empirical CCDF for values of $k \geq k_{min}$. Subsequently, we calculate the maximum distance (KS statistic) between the empirical CCDF of the visibility graph and the fitted model. Then, we repeat these two steps for each k_{min} greater than 2. At the end, the best fitting power-law model is the one that has the minimum KS statistic among all fitted models. The estimated parameters (α and k_{min}) are corresponding to the model with the minimum KS statistic.

The CCDFs of the PMU visibility graphs for three sample time series (voltage, frequency, and angle) are shown in Figs. 4 (a)-(c). The best power-law fits of the three CCDFs are shown using a dotted line with scaling parameters of $\alpha_v = 4.37$, $\alpha_f = 3.55$, and $\alpha_a = 4.09$.

Using MLE and KS statistic, we estimate the scaling parameters of the power-law distributions for 1200 PMU visibility graphs of voltage magnitude, frequency, and phase angle. The densities of the scaling parameters for each variable are shown in Figs. 5 (a)-(c). The variability in the scaling parameters among different PMU data variables is related and proportional to the values of their Hurst exponents [10].

B. Goodness-of-fit analysis

To assess the validity of fitting the empirical CCDF of the PMU visibility graph to a power-law distribution, we investigate the statistical goodness-of-fit by comparing maximum distance (KS statistic) between the empirical CCDF and the postulated model to the distances of synthetic data sets drawn from the same postulated model.

TABLE I. Power-law fits of 1200 PMU visibility graphs (voltage, frequency, and phase angle)

Parameter	Voltage	Frequency	Angle
% of graphs with $p \geq 0.1$	20.5%	8.3%	13.5%
Scaling parameter ($\overline{\alpha}$)	4.8	3.7	3.9
Lower bound ($\overline{k}_{min}$)	97.8	72.5	40.2
# of points in tail ($\overline{n}_{tail}$)	1139.5	1629.4	2981

For each empirical CCDF, we generate 1000 synthetic data sets from the postulated model. Then, for each synthetic CCDF, we calculate the maximum distance between the synthetic data and its best power-law fit. At significance level of 0.10, we define the p-value to be the fraction of synthetic data with a distance higher than the distance of the empirical data. If $p < 0.1$, we can reject the null hypothesis and the power-law model is not a plausible fit for the empirical data. Otherwise, we can not reject the power-law model as a good fit of data.

We apply the goodness-of-fit test on three empirical CCDF of PMU visibility graphs as shown in Figs. 4 (a)-(c). After estimating the best model (α, k_{min}) of the three empirical distributions, we generate 1000 synthetic data sets to calculate the p-value. It seems the power-law fit is a plausible fit for the three distributions with p-values higher than 0.1.

We can apply similarly the goodness-of-fit test on 1200 PMU visibility graphs. The detailed results are shown in Table I. The results show that the power law could be a

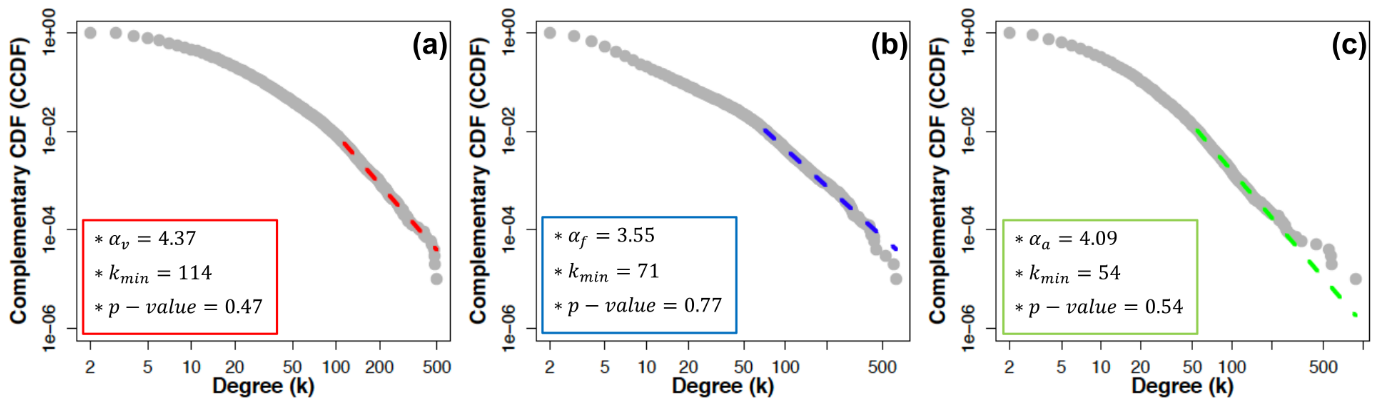

Fig. 4. Complementary Cumulative Distribution Functions of PMU visibility graphs of three 100,000-sample time series

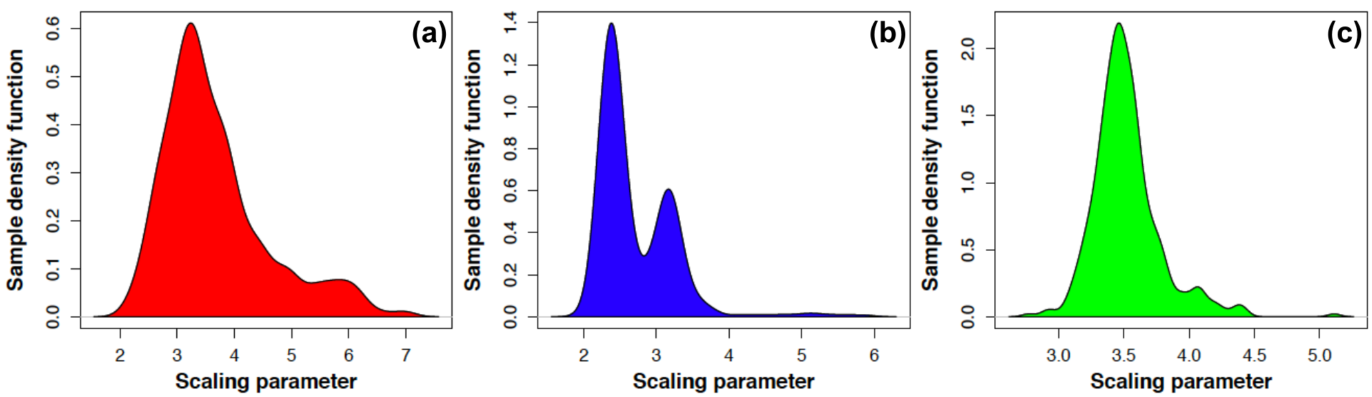

Fig. 5. Sample density functions of the scaling parameters of 1200 PMU visibility graphs of voltage, frequency, and angle

plausible fit for 20.5% of voltage visibility graphs, 8.3% of frequency visibility graphs, and 13.5% of angle visibility graphs. The low percentage of empirical distributions passing the goodness-of-fit test could be a result of the dwindles at the end of the degree distribution. In this region, the vertices are not represented by enough samples as result of the series size.

V. Conclusion

In this paper, we introduced the PMU visibility graph to study the hidden statistical properties in the PMU data. Then, we showed that the graph is scale-free with heavy-tailed distribution like a power law. The goodness-of-analysis indicated that the power law could be more plausible fit for the voltage and phase angle compared to the frequency. The degree distribution hybrid structure and the variability in the estimated scaling parameters among different PMU data variables reveals the nonstationary and fractal properties of the PMU data [10].

In the future, we will fit the distributions of the PMU visibility graphs to other heavy-tailed distributions (Log-normal, Weibull, α-stable, etc.). In addition, a rigorous statistical comparison will be performed between the different fitted model via the likelihood ratio test.

References

[1] C. Huang, F. Li, D. Zhou, J. Guo, Z. Pan, and Y. Liu. Data quality issues for synchrophasor applications part i: a review. *Journal of Modern Power Systems and Clean Energy*, 4(3):342–352, Jul 2016.

[2] Shyam R., Bharathi Ganesh H.B., Sachin Kumar S., Prabaharan Poornachandran, and Soman K.P. Apache spark a big data analytics platform for smart grid. *Procedia Technology*, 21:171 – 178, 2015.

[3] IBM Software. Managing big data for smart grids and smart meters. May 2012.

[4] Y. Cheng, K. Chen, H. Sun, Y. Zhang, and F. Tao. Data and knowledge mining with big data towards smart production. *Journal of Industrial Information Integration*, 9:1 – 13, 2018.

[5] Y. Zhang, T. Huang, and E. F. Bompard. Big data analytics in smart grids: a review. *Energy Informatics*, 1(1):8, Aug 2018.

[6] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño. From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105(13):4972–4975, 2008.

[7] J. Zhang and M. Small. Complex network from pseudoperiodic time series: Topology versus dynamics. *Phys. Rev. Lett.*, 96:238701, Jun 2006.

[8] L. Shalalfeh, P. Bogdan, and E. Jonckheere. Evidence of long-range dependence in power grid. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–5, July 2016.

[9] L. Shalalfeh, P. Bogdan, and E. Jonckheere. Modeling of pmu data using arfima models. In *2018 Clemson University Power Systems Conference (PSC)*, pages 1–6, Sep. 2018.

[10] L. Lacasa, B. Luque, J. Luque, and J. C. Nuño. The visibility graph: A new method for estimating the hurst exponent of fractional brownian motion. *EPL (Europhysics Letters)*, 86(3):30001, may 2009.

[11] PMU data from EPFL campus. http://nanotera-stg2.epfl.ch/. [Online; accessed 5-February-2019].

[12] Béla Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001.

[13] P. Erdős and A. Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290, 1959.

[14] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[15] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51(4):661–703, November 2009.