Root Cause Localization on Power Networks

Zhen Chen, Kai Zhu and Lei Ying
School of Electrical, Computer and Energy Engineering
Arizona State University
Tempe, AZ, United States
Email: {zchen113, kzhu17, lei.ying.2}@asu.edu

***Abstract*—Given information about cascading failures on a power network, localization of the root cause is of great importance in accelerating the process of recovering the power network from cascading failures. In order to locate the root cause, we need to answer two questions: (1) Which information can we use to locate the root cause of cascading failures? (2) What algorithm can locate the root cause effectively? In this paper, we present our recent results towards answering these two questions. In particular, we developed a root-cause localization algorithm by exploiting the concept of correlation network, which can describe the diffusion. The simulations based on different power networks are provided, in which we show our algorithm outperforms some other existing algorithms.**

Keywords—Power network, information source detection.

I. Introduction

The power grid is a critical infrastructure of our society. The failure of the power grid network will have catastrophic impacts on water supplies, transportation networks, communications networks, and almost every other aspect of our daily life. For post-cascade fault diagnostics, important questions shall be answered including: How can we locate the source of fault on the power network? Which kind of information can we utilize to locate the initially malfunctioning part that causes the cascading failures? The answers to these questions are crucial for recovering the power network from cascading failures and even stopping the cascade.

The properties of cascading failures in the power grid have been studied in the literature. For instance, [1], [2], [3], [4] use the DC power flow model to model the cascading failures in the power network. A different method is to assume that the failure of a node or a transmission line leads to the failures of other nodes or transmission lines nearby with some probability, such as [5], [6], [7]. However, there are a few works on locating the root cause of the cascading failures. In [8], the authors used the information from the communication network to find the root cause of cascading failures in the power network.

In this paper, we relate the problem of locating the root cause of cascading failures in power networks to source localization in networks, which has been studied in the literature. There are some effective methods proposed in the literature, which aim at detecting the information source, such as the rumor centrality estimator in [9], [10], the sample-path-based estimator in [11], [12], the eigenvalue-based estimator in [13] and the dynamic message-passing algorithm in [14]. However, the root-cause localization problem addressed in this paper is different from the information source detection problem. In social networks, the information generally propagates through topological connections while the cascading failures in power networks are the result of electrical interactions restricted by Kirchhoff's and Ohm's laws. In addition, it has been pointed out in [15] that topological models may result in misleading conclusions in power networks.

II. Approach

In this section, the approach used to locate the root cause is introduced.

A. Correlation Network

In power networks, the outage of a transmission line can result in power flow reallocation, which leads to overload and outage of other transmission lines. The spread of cascading failures in power networks is governed by Kirchhoff's and Ohm's laws, which is very different from social networks where information diffusion is primarily determined by the network topology. For instance, in the four-bus system shown in Figure 1, the outage of transmission line $(2,3)$ can cause the outage of line $(1,4)$, but these two transmission lines do not connect with each other.

Without a good understanding of the propagation of cascading failures on power networks, it is impossible to locate the root cause accurately. Unfortunately, the network topology does not accurately capture the dependency of the transmission lines, so we adopt the correlation network proposed in [16], which models the influence of one transmission line to the others in power networks.

To establish the correlation network, the first step is to check the $N-1$ contingencies of the power network to extract the following matrix:

$$\begin{bmatrix} \Delta P_{11} & \Delta P_{12} & \Delta P_{13} & \cdots & \Delta P_{1n} \\ \Delta P_{21} & \Delta P_{22} & \Delta P_{23} & \cdots & \Delta P_{2n} \\ \Delta P_{31} & \Delta P_{32} & \Delta P_{33} & \cdots & \Delta P_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Delta P_{n1} & \Delta P_{n2} & \Delta P_{n3} & \cdots & \Delta P_{nn} \end{bmatrix}, \tag{1}$$

where n is the number of transmission lines, $\Delta P_{kl}(k \neq l)$ is the change of power flow on line l caused by the outage of line k, and $\Delta P_{kk} = 0$ for $k = 1, \ldots, n$. Then the correlation network is defined to be an undirected complete graph in which each node represents a transmission line and the weight between two nodes k and l is

$$w_{kl} = \frac{1}{\frac{\Delta P_{kl}}{\sum_{i=1}^{n} \Delta P_{il}} + \frac{\Delta P_{lk}}{\sum_{i=1}^{n} \Delta P_{ik}}} \tag{2}$$

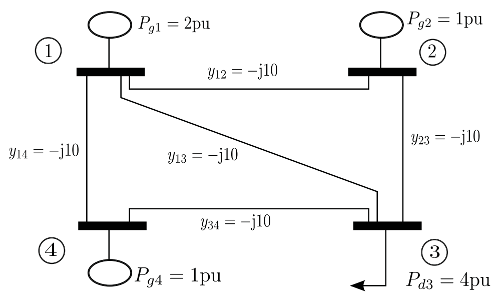


Fig. 1: This is a four-bus network, where y_{ij} is the admittance of line (i, j), P_{gi} is the power injection of bus i and P_{dj} is the power demand of bus j. By using the DC power flow model, we obtain $P^o_{12} = 0.25$, $P^o_{13} = 1.5$, $P^o_{14} = 0.25$, $P^o_{23} = 1.25$ and $P^o_{34} = -1.25$, where P^o_{ij} is the power flow on line (i, j) when the system is in steady-state operation. We assume the capacity of each transmission line is $c_{ij} = (1+\beta)|P^o_{ij}|$, where $\beta = 1$. If the power flow on the transmission line reaches its capacity, it will be tripped. Therefore, if line $(2, 3)$ fails first, according to the DC power flow model, we have lines $P_{12} = -1, P_{13} = 2.333, P_{14} = 0.667$ and $P_{34} = -1.666$, where P_{ij} represents the value of power flow on line (i, j) after line $(2, 3)$ failed. Therefore, line $(1, 2)$ and line $(1, 4)$ will exceed their capacity and be tripped next.

for each edge (k, l). Since $\frac{\Delta P_{kl}}{\sum_{i=1}^{n} \Delta P_{il}}$ is the ratio between the change of power flow of line l caused by the outage of line k and the sum of the change of power flow of line l caused by the outage of each transmission line, the weight can be viewed as the impact of line k on line l. Similarly, $\frac{\Delta P_{lk}}{\sum_{i=1}^{n} \Delta P_{ik}}$ measures line l's impact on k. Therefore, the weight between two nodes in the correlation network can be used to measure the correlation (of failures) between two transmission lines in the original power network. We use this correlation network as the underlying network and model the cascading failures on the power network as the information diffusion on the correlation network.

B. Infection Spreading Tree & Cost Function

Considering an information diffusion on the correlation network $G(\mathcal{V}, \mathcal{E})$, for each edge $(v, u) \in \mathcal{E}$, node v infects node u means the outage of the transmission line v in the power network result in the outage of the transmission line u. Then we define $\mathcal{T}$ to be the infection spreading tree, which specifies the sequence of infection, define $\boldsymbol{t}$ to be a $|\mathcal{T}|$-dimensional vector, which specifies the infection time of each node on the infection spreading tree, and define $\boldsymbol{\tau}$ to be the vector of infection time of partial infected nodes, which can be observed. The information we can utilize to locate the root cause is the correlation network $G(\mathcal{V}, \mathcal{E})$, the set of infected nodes (i.e., tripped transmission lines), I, and the timestamps $\boldsymbol{\tau}$ of partial infected nodes.

Given a spreading tree $\mathcal{T}$ and infection time vector $\boldsymbol{t}$, for each edge $(v, u) \in \mathcal{E}(\mathcal{T})$, we assume that the time it takes for node v to infect node u follows a truncated Gaussian distribution with mean μ_{vu} and variance σ^2_{vu} by considering that the time cannot be negative. Since the weight w_{vu} for edge (v, u) measures the mutual impact between v and u and the smaller w_{vu} is, more likely that the outage of one node will cause the outage of the other, we assume

$$\mu_{vu} = \sigma^2_{vu} \propto w_{vu}. \tag{3}$$

Therefore, the means and variances of different edges can be different, which makes the homogeneous assumption in [17] cannot be adopted in this paper. Then given a spreading tree $\mathcal{T}$, the probability density associated with infection time vector $\boldsymbol{t}$ is

$$\begin{aligned} f_{\mathcal{T}}(\boldsymbol{t}) = \prod_{(v,u)\in\mathcal{E}(\mathcal{T})} & \frac{1}{Z_{vu}\sqrt{2\pi}\sigma_{vu}} \\ & \exp(-\frac{(t_u - t_v - \mu_{vu})^2}{2\sigma^2_{vu}}), \end{aligned} \tag{4}$$

where

$$Z_{vu} = 1 - \phi(\frac{-\mu_{vu}}{\sigma_{vu}}) \tag{5}$$

is the normalization constant that depends on μ_{vu} and σ^2_{vu}. Thus, the log-likelihood is

$$\begin{aligned} \log f_{\mathcal{T}}(\boldsymbol{t}) = & -\sum_{(v,u)\in\mathcal{E}(\mathcal{T})} \log(Z_{vu}\sqrt{2\pi}\sigma_{vu}) \\ & - \sum_{(v,u)\in\mathcal{E}(\mathcal{T})} \frac{1}{2\sigma^2_{vu}}(t_u - t_v - \mu_{vu})^2. \end{aligned} \tag{6}$$

Therefore, we define cost of the spreading tree $\mathcal{T}$ given infection time $\boldsymbol{t}$ to be

$$\begin{aligned} C_{\mathcal{T}}(\boldsymbol{t}) = \sum_{(v,u)\in\mathcal{E}(\mathcal{T})} & \log(Z_{vu}\sqrt{2\pi}\sigma_{vu}) \\ & + \frac{1}{2\sigma^2_{vu}}(t_u - t_v - \mu_{vu})^2. \end{aligned} \tag{7}$$

For each infected node s in the network, the cost of the node is

$$C(s) = \min_{(\mathcal{T}_s, \boldsymbol{t}_s)\in T_s(I,\boldsymbol{\tau})} C_{\mathcal{T}_s}(\boldsymbol{t}_s), \tag{8}$$

where $T_s(I, \boldsymbol{\tau})$ is the set of all possible pairs of infection tree, T, and infection time vector, $\boldsymbol{t}$, which satisfy the observation I and $\boldsymbol{\tau}$, starting from source s. We will choose the node with the smallest cost as the root cause of the cascading failures. According to Theorem 1 in [17], the problem described in (8) is NP-hard. In next section, a greedy algorithm of finding the minimum cost for each infected node will be introduced, which is a variation of the greedy algorithm proposed in [17].

C. Greedy Algorithm

In this section, we present a greedy algorithm, named SP, to build an infection spreading tree starting from a specific node and to assign infection time to each node in order to minimize the cost. The algorithm uses the shortest paths of the correlation network to build the infection spreading tree. Define M to be the number of nodes with timestamps observed and $\boldsymbol{x}$ to be the vector of all infected nodes, where the first M elements x_i $(1 \leq i \leq M)$ represent the nodes with timestamps observed. Assume $t_{x_i} \leq t_{x_j}$ $(1 \leq i \leq j \leq M)$, where t_{x_i} is the infection time of node x_i. Then the algorithm of calculating the cost of an infected node, s, is described in Algorithm 1.

Compared with the algorithm used in [17], in Step 3, we choose the shortest path to add to the spreading tree instead of choosing the path with the smallest cost. According to equation (7), the cost brought by each edge, (v,u), in the infection tree consists of two parts: $\log(Z_{vu}\sqrt{2\pi}\sigma_{vu})$ and $\frac{1}{2\sigma_{vu}^2}(t_u - t_v - \mu_{vu})^2$. Unlike the homogeneous diffusion model used in [17], different edges in the correlation network have different weights, so the first part in (7) cannot be ignored when we compare the costs of different spreading trees. Here according to (5) and (10), we have

$$\log(Z_{vu}\sqrt{2\pi}\sigma_{vu}) = \log((1-\phi(-\sqrt{t_{\text{avg}}w_{vu}}))\sqrt{2\pi t_{\text{avg}}w_{vu}}),$$

whose value increases as the increase of w_{vu}. Therefore, by utilizing the shortest path to form the infection tree in Step 3, we aim to minimize the cost brought by the first part in the cost function. Then, in Step 4, when assigning infection time to the nodes on the newly added infection path, we aim at minimizing the cost caused by the second part in the (7).

Algorithm 1 SP

1: Estimate μ_{vu} and σ_{vu}^2 by using $\boldsymbol{\tau}$. Define

$$l_{vu} = \min_{p\in P(v,u)} \sum_{(a,b)\in p} w_{ab},$$

where $P(v,u)$ is the set of all possible paths between node v and node u and $(a,b)\in p$ means edge (a,b) belongs to path p. Define the normalized average infection time to be

$$t_{\text{avg}} = \frac{\sum_{t_v,t_u\in\boldsymbol{\tau},v\neq u}|t_v - t_u|}{\sum_{t_v,t_u\in\boldsymbol{\tau},v\neq u} l_{vu}}. \quad (9)$$

According to (3), we assume that

$$\mu_{vu} = \sigma_{vu}^2 = w_{vu}t_{\text{avg}}. \quad (10)$$

2: Construct the initial infection spreading tree, $\mathcal{T}_s$, by including node, s, and set the initial cost $C(s)$ to be zero and $k=1$.

3: If x_k is in $\mathcal{T}_s$, $k=k+1$ and repeat Step 3. Otherwise, node x_k is added to the current spreading tree $\mathcal{T}_s$, by adding the path

$$p_k^\star = \operatorname*{argmin}_{p\in P'(v,x_k),\forall v\in V(\mathcal{T}_s)} \sum_{(a,b)\in p} w_{ab} \quad (11)$$

into the spreading tree, where $P'(v,x_k)$ is defined to be the set of all modified paths from v to x_k. Each modified path from v to x_k has to satisfy the following conditions:

- it does not include any node on the spreading tree $\mathcal{T}_s$, except node v;
- it does not include node x_j for any $k < j \leq M$ in vector $\boldsymbol{x}$.

The cost is updated as follows:

$$C(s) = C(s) + \sum_{(v,u)\in p_k^\star} \log(Z_{vu}\sqrt{2\pi}\sigma_{vu}). \quad (12)$$

4: Assume the other endpoint besides x_k of path $p_k^\star$ is g. Assign infection time to each node, h, on path $p_k^\star$ by following the rule:

$$t_h = t_{x_1} - \sum_{\substack{(a,b)\in p_{hx_k},\\ p_{hx_k}\subset p_k^\star}} \mu_{ab}, \text{ if } k=1 \text{ and } s\neq x_1, \quad (13)$$

otherwise,

$$t_h = t_g + \frac{t_{x_k} - t_g}{\sum_{(m,n)\in p_k^\star}\mu_{mn}} \sum_{\substack{(a,b)\in p_{gh},\\ p_{gh}\subset p_k^\star}} \mu_{ab}, \text{ if } k\leq M, \quad (14)$$

and

$$t_h = t_g + \sum_{\substack{(a,b)\in p_{gh},\\ p_{gh}\subset p_k^\star}} \mu_{ab}, \text{ if } k > M, \quad (15)$$

where $p_{gh}\subset p_k^\star$ represents that path p_{gh} is part of path $p_k^\star$ between node g and h. The cost is further updated to

$$C(s) = C(s) + \sum_{(v,u)\in p_k^\star} \frac{1}{2\sigma_{vu}^2}(t_u - t_v - \mu_{vu})^2 \quad (16)$$

and $k=k+1$.

5: If $k < |I|$, go back to Step 3. Otherwise, stop.

D. Cost-based and tree-based ranking

Among nodes $x_1, ..., x_M$, only x_1 is possible to be the source, so we set $C(x_i) = \infty$ $(i = 2, ..., M)$. For other infected nodes, s, we construct the spreading tree and calculate the cost, $C(s)$. We rank all the infected nodes by using the following two methods. Assume the node with the smallest cost is $s^\star$, and the infection spreading tree starting from $s^\star$ obtained by SP is $\mathcal{T}_{s^\star}$. For any transmission line v, define $t_v^\star$ to be the infection time of v, which is obtained by SP, on tree $\mathcal{T}_{s^\star}$.

1) Cost-based ranking (CR): Rank the transmission lines in an ascending order of the cost.

2) Tree-based ranking (TR): Set $t_{x_i}^\star = \infty$ $(i = 2, ..., M)$. Rank all transmission lines in an ascending order of their infection time on tree $\mathcal{T}_{s^\star}$.

III. Experimental Evaluation

In this section, we present performance evaluations of our algorithm on the IEEE 300 bus system and the electricity transmission network of Great Britain.

A. DC power flow model

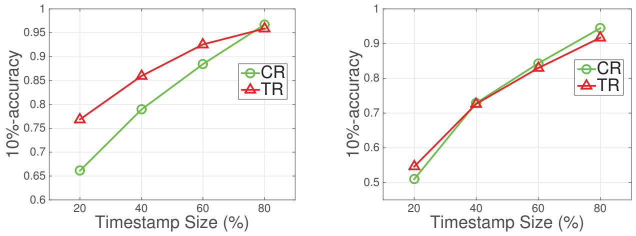
In our simulations, we use the DC power flow model to calculate the power flow on each transmission line, which is widely used as an approximation for more realistic AC power flow model. Define P_i to be the active power injection and θ_i to be the voltage angle at bus i. Then we have

$$P_i = \sum_{j=1}^{N} B_{ij}(\theta_i - \theta_j), \quad (17)$$

in which B_{ij} is the reciprocal of the reactance between bus i and bus j. The active power flow through transmission line L, between buses i and j, is

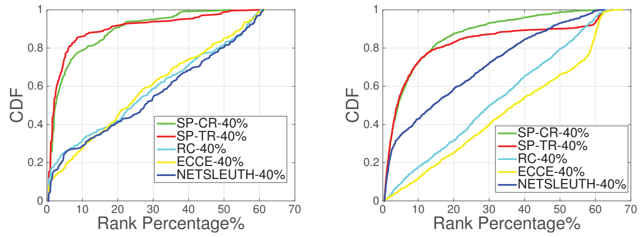
$$P_L = \frac{1}{X_L}(\theta_i - \theta_j), \quad (18)$$

where X_L is the reactance of line L. Therefore, if we have the data of the power injections of all buses and the reactances



(a) IEEE 300 bus power system (b) The Great Britain power network

Fig. 2: The probability that the first tripped transmission line is ranked among the top 10%.



(a) IEEE 300 bus power system (b) The Great Britain power network

Fig. 3: The empirical CDF of the rank of the first tripped transmission line when the failure time of 40% of tripped transmission lines is observed.

of all transmission lines of a power network, we can use DC power flow model to calculate the power flow on each transmission line.

B. Cascading failures model

We adopt the cascading failure model proposed in [4]. Each transmission line L has a capacity c_L. The cascading failures model in [4] consists of the following steps:

1) Step 1: At time step $t = 0$, a failure of some transmission line occurs.
2) Step 2: Adjust the total demand to equal the total supply by decreasing the demand (supply) by the same factor at all loads (generators).
3) Step 3: Use the DC power flow model to calculate the power flow P_L on each transmission line L.
4) Step 4: Compute a moving average
$$P_L^t = \alpha|P_L| + (1-\alpha)P_L^{t-1}, \quad (19)$$
where P_L^{t-1} is the moving average at previous time slot.
5) Step 5: Remove all lines which satisfy $P_L^t > c_L$. If at least one line is removed, go to Step 2. Otherwise, stop.

C. Simulation Results

In this section, we evaluate the performance of our algorithm on two power systems: IEEE 300 bus power system and the Great Britain power network.

- The IEEE 300 bus power system was developed by the IEEE Test Systems Task Force under the direction of Mike Adibi in 1993 [18]. This system consists of 69 generation buses, 231 load buses with a total 411 transmission lines.
- The Great Britain power network consists of 2224 nodes, which includes 394 generation buses, and 3207 transmission lines [19].

In our simulations, we set $\alpha = 1$, where α is a parameter in (19). Since we do not know the capacity of each transmission line in both of these two power systems, we assume the capacity of each line L, c_L, satisfies that $c_L = (1+\beta)|P_L^o|$, where β is a constant and P_L^o is the power flow on transmission line L when the power system is in the steady state with no transmission line is tripped. The simulation results are presented in Figure 2 and Figure 3. From Figure 2a and Figure 2b, we can see that the top 10% accuracy increases with the number of nodes with timestamps. The top-10% accuracy is ≥ 0.85 when 60% timestamps are known in the IEEE 300 bus power system and is ≥ 0.8 in the Great Britain Power network. We further compared our algorithms with existing algorithms for source localization including rumor centrality (RC) [9], Jordan center (ECCE) [11] and NETSLEUTH [13]. The result is in Figure 3, from which we can see that our algorithms outperform others.

IV. Conclusion

In this paper, we studied the problem of locating the root cause of cascading failures in power networks. We utilize the novel correlation network to describe the diffusion of cascading failures and developed a cost based algorithm.

Acknowledgment

Research supported in part by ARO grant W911NF-13-1-0279

References

[1] I. Dobson, B. Carreras, V. Lynch, and D. Newman, "An initial model fo complex dynamics in electric power system blackouts," in *Proc. of the 34th Annu. Hawaii Int. Conf. on System Sciences*, Jan 2001, pp. 710–718.

[2] J. Chen, J. S. Thorp, and I. Dobson, "Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model," *Int. J. of Elect. Power & Energy Syst.*, vol. 27, no. 4, pp. 318–326, 2005.

[3] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, "Critical points and transitions in an electric power transmission model for cascading failure blackouts," *Chaos: An Interdisciplinary J. of Nonlinear Sci.*, vol. 12, no. 4, pp. 985–994, 2002.

[4] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman, "Power grid vulnerability to geographically correlated failures - analysis and control implications," in *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, April 2014, pp. 2634–2642.

[5] D. P. Chassin and C. Posse, "Evaluating north american electric grid reliability using the barabási–albert network model," *Physica A: Statistical Mechanics and its Applicat.*, vol. 355, no. 2, pp. 667–677, 2005.

[6] Z. Kong and E. M. Yeh, "Resilience to degree-dependent and cascading node failures in random geometric networks," vol. 56, no. 11, pp. 5533–5546, 2010.

[7] H. Xiao and E. M. Yeh, "Cascading link failure in the power grid: A percolation-based analysis," in *IEEE Int. Conf. on Commun. Workshops*. IEEE, 2011, pp. 1–6.

[8] A. Das, J. Danerjee, and A. Sen, "Root cause analysis of failures in interdependent power-communication networks," in *The Military Commun. Conf.* IEEE, 2014.

[9] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?" *IEEE Trans. Inf. Theory*, vol. 57, pp. 5163–5181, Aug. 2011.

[10] ——, "Rumor centrality: a universal source detector," in *SIGMETRICS Perform. Eval. Rev.*, New York, NY, USA, 2012, pp. 199–210.

[11] K. Zhu and L. Ying, "Information source detection in the SIR model: A sample path based approach," in *Proc. Information Theory and Applications Workshop (ITA)*, Feb. 2013.

[12] ——, "A robust information source estimator with sparse observations," *Computational Social Networks*, vol. 1, no. 1, pp. 1–21, 2014.

[13] B. A. Prakash, J. Vreeken, and C. Faloutsos, "Spotting culprits in epidemics: How many and which ones?" in *IEEE Int. Conf. Data Mining (ICDM)*, Brussels, Belgium, 2012, pp. 11–20.

[14] A. Y. Lokhov, M. Mézard, H. Ohta, and L. Zdeborová, "Inferring the origin of an epidemic with a dynamic message-passing algorithm," *Phys. Rev. E*, vol. 90, p. 012801, Jul 2014.

[15] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Do topological models provide good information about electricity infrastructure vulnerability?" *Chaos: An Interdisciplinary J. of Nonlinear Sci.*, vol. 20, no. 3, 2010.

[16] X. Zhang, F. Liu, R. Yao, X. Zhang, S. Mei, Z. Zhang, and X. Li, "Identification of key transmission lines in power grid using modified k-core decomposition," in *Int. Conf. on Electric Power and Energy Conversion Syst. (EPECS)*. IEEE, 2013, pp. 1–6.

[17] K. Zhu, Z. Chen, and L. Ying, "Locating contagion sources with partial timestamps," *Arxiv preprint arXiv:1412.4141*, 2014.

[18] R. Christie, "300 bus power flow test case [online]," 1993, available at http://www.ee.washington.edu/research/pstca/pf300/pg_tca300bus.htm.

[19] W. Bukhsh and K. McKinnon, "Network data of real transmission networks [online]," April 2013, available at http://www.maths.ed.ac.uk/optenergy/NetworkData/fullGB/.