

Tractable Structure Learning in Radial Physical Flow Networks

Deepjyoti Deka*, Scott Backhaus*, and Michael Chertkov*

*Los Alamos National Laboratory, USA

Email: deepjyoti@lanl.gov, backhaus@lanl.gov, chertkov@lanl.gov

Abstract—Physical Flow Networks are different infrastructure networks that allow the flow of physical commodities through edges between its constituent nodes. These include power grid, natural gas transmission network, water pipelines etc. In such networks, the flow on each edge is characterized by a function of the nodal potentials on either side of the edge. Further the net flow in and out of each node is conserved. Learning the structure and state of physical networks is necessary for optimal control as well as to quantify its privacy needs. We consider radial flow networks and study the problem of learning the operational network from a loopy graph of candidate edges using statistics of nodal potentials. Based on the monotonic properties of the flow functions, the key result in this paper shows that if variance of the difference of nodal potentials is used to weight candidate edges, the operational edges form the minimum spanning tree in the loopy graph. Under realistic conditions on the statistics of nodal injection (consumption or production), we provide a greedy structure learning algorithm with quasi-linear computational complexity in the number of candidate edges in the network. Our learning framework is very general due to two significant attributes. First it is independent of the specific marginal distributions of nodal potentials and only uses order properties in their second moments. Second, the learning algorithm is agnostic to exact flow functions that relate edge flows to corresponding potential differences and is applicable for a broad class of networks with monotonic flow functions. We demonstrate the efficacy of our work through realistic simulations on diverse physical flow networks and discuss possible extensions of our work to other regimes.

Index Terms—Physical flow networks, monotonic flow, positive quadrant dependence, Spanning Tree, Graphical Models, Missing data, Computational Complexity

I. INTRODUCTION

Physical flow networks [1] form strategic components of modern society’s activities and help in the mass transport of energy and daily utilities from far off-generation points to end users through pipes/edges. Example of such networks include the power grid (flow of electricity), gas transmission network (delivery of gas for household usage and energy production), water pipelines (for drinking and industrial water) etc. Further, in the continuum domain (fluid approximation), discrete particular network like vehicular traffic can also be considered as a physical flow network. Over time, such networks have grown in size and become vital for the smooth functioning of most, if not all, activities - be it industry or household related. Efficient control and fast monitoring of the state of such flow networks is crucial for their real-time usage. Further, the advancement in smart active devices (energy meters, smart thermostats/heaters etc.) has led to efforts in

distributed but optimal control of flow networks. Fast topology (set of inconnections in the flow network) and usage (steady state of current vehicular traffic, statistics of power, gas, water consumption etc.) estimation are necessary tools to ensure the optimal application of decentralized schemes. Finally the assimilation of online communication tools for monitoring and signaling exchange in flow networks puts them at risk of cyber adversaries and disruptive agents. Low overhead algorithms to learn the topology and state of the network can help quantify the cyber risk associated with compromised set of measurements and help guide preventive measures and placement of secure devices.

Due to the diversity of transported commodities, flow networks vary in the analytical flow models. However such flow models often satisfy common laws. One, net flow at each node is conserved, i.e., the total flow injected into each node is equal to the sum outward flow on all edges connected to that node. Second, the flow on each edge is guided by the difference between potentials at the two nodes on either side of the edge. Examples of such potential include voltages in power networks, pressure in gas and water networks. Similarly, one can think of virtual potentials in traffic network fluid models. In most flow networks, the potential difference across an edge is expressed as a monotonic function of the flow on it, implying that the flow increases when the difference in potentials increases and vice versa. Note that the monotonic function may be non-linear. In this paper, we discuss topology learning algorithms for radial flow networks with monotonic flow functions. We focus specifically on radially structured networks as they appear frequently in different contexts. For example, electricity distribution grids [2] are known to operate in a radial topology due to economic and operational reasons [3]. Similarly, gas transmission networks [4], [5] and some water networks [6] have a tree-like operational configuration. However, the set of all permissible edges in the network (operational and open) form an underlying loopy graph with cycles. The radial operational configuration is achieved by restricting the flow to a subset of the permissible edges in the network as shown in Fig. 1. In certain networks like the electricity distribution grid, this radial configuration can be changed over a few hours by switching on and off edges (transmission line breakers) and needs to be estimated for control applications. Real-time meters on edges that relay information on current flow and operational status are often sparsely present. Even if the radial structure is static, third-

party applications may be interested in learning the topology using indirect (non-edge based) measurements as access to network structure information is often restricted. Hence, we analyze the problem of estimating the true operational structure using only empirical nodal potential measurements. Specifically, *we show that second moments of nodal potentials are sufficient to reconstruct the operational radial topology by a greedy algorithm.* Surprising, our learning scheme does not depend on the exact flow function (linear or non-linear) for the network as long as the flow function is monotonic. Thus, it has wide applicability for a variety of networks as mentioned in detail in subsequent sections. Note that brute force approaches to learn the topology is computationally prohibitive due to the exponential number of radial topologies that can be constructed from a dense loopy graph of permissible edges.

A. Prior Work

Past work in learning the structure of flow networks have generally focussed on specific applications. In particular, nodal measurement based structure estimation of power distribution networks is an area of active research. Researchers have used Markov random fields [7], signs in inverse covariance matrix of voltages (potentials) [8], envelope comparison based reconstruction methods [9]–[11] etc. Limited power flow measurements have been used to estimate the topology using maximum likelihood tests in [12]. In our prior work [3], [13], [14], iterative greedy learning schemes based on a linear power flow model have been used to determine the operating power grid, even in the presence of missing/unobserved nodes.

We are not aware of prior work in learning the structure of gas or water networks using information of nodal potentials. However there exist several efforts on different optimization (stochastic, robust etc.) and control (optimal, distributed etc.) schemes for these networks that depend on information of network structure and use nodal potentials as variables. These include geometric programming based optimization schemes specifically for radial gas [5] and water networks [6]. Recent work [15] has demonstrated the tractability of several robust optimization schemes in gas networks due to the monotonic nature of the function that relates edge flows to nodal potential differences. The learning algorithm in this paper shows that the monotonicity of the flow function makes structure learning using nodal potentials tractable as well. These algorithms can in turn enable optimization problems to be tackled without the prior knowledge of the underlying flow network as that can be easily estimated.

Aside from the mentioned work in flow networks, the Chow-Liu algorithm [16] uses a spanning tree algorithm for learning tree-structured graphical models that is based on the pairwise-factorization of the systems's mutual information. This is generalized in [17] to tree-structured graphical models with hidden/latent variables through the use of information distances as edge weights.

B. Contribution of This Work

Most of the previous work in learning structure of physical flow networks are limited to specific cases, in particular power grids. Further they assume linear flow function relating edge flow and nodal potentials. The fundamental contribution of this paper is to develop learning algorithms that are applicable for physical flow networks with monotonic flow functions that can be nonlinear and even distinct for each edge in the system. We show that under independent nodal injections, the variance of potential differences in such networks show provable trends that can guide greedy algorithms for structure learning. Our main algorithm uses variance in difference of nodal potentials as edge weight and identifies the operational structure by a spanning tree algorithm. In particular, the algorithm does not need any information of the flow function involved or nodal injection statistics. If the flow functions are known, the algorithm can be used to estimate the statistics of nodal injections. In essence, this work generalizes prior work [14] on linear power flow models to general radial networks with monotonic flow functions that are distinct and non-linear. The worst-case computational complexity of our algorithm is quasi-linear in the number of permissible edges in the network, which is efficient to learn the structure of large networks. We are not aware of any existing work that estimates the structure of general flow networks with non-linear flow functions. We demonstrate the performance of our algorithms through experiments on two test networks, one pertaining to power grid and the other to gas grid.

The rest of the manuscript is organized as follows. Section II introduces nomenclature and relations between injections, flows and potentials in physical flow networks through detailed example networks. We present key properties and trends in nodal potentials for flow networks in Section III. Design of the spanning tree based learning algorithm is given in Section IV. We also include a part on extensions of our work in Section IV. Simulation results of our learning algorithm on different example networks are presented in Section V. Finally, Section VI contains conclusions, extensions and discussion of future work.

II. FLOW MODELS FOR FLOW NETWORKS

We first provide the notation for the topology of the flow network.

Radial Structure: Mathematically, the overall physical flow network is represented as a loopy graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of nodes and \mathcal{E} is the set of all permissible edges. Nodes are denoted by alphabets (a, b, \dots) and edge between two nodes a, b by node pair (ab) . The ‘radial’ structure composed of operational edges is denoted by tree \mathcal{T} with nodes $\mathcal{V}_{\mathcal{T}}$ and operational edge set $\mathcal{E}_{\mathcal{T}} \subset \mathcal{E}$. We restrict our discussion to one operational tree as shown in Fig. 1 though our results hold for the case with multiple disjoint trees. Next, we list the flow equations for the network.

Flow Models: The flow network is characterized by three sets of variables: nodal injections, edge flows and nodal potentials. At each node a , flow is conserved and $P_a = \sum_{b:(ab) \in \mathcal{E}_{\mathcal{T}}} f_{ab}$

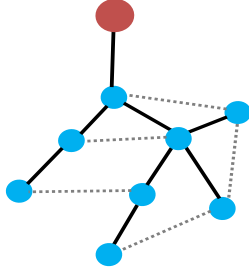


Fig. 1. A radial flow network with operational edges colored solid black. Dotted grey lines represent non-operational edges. The red node denotes the ‘reference’ node.

where P_a is the injection at a and f_{ab} is the flow from a to b on edge (ab) . In vector form, we write

$$P = M^T f \quad (1)$$

Here M is the node to edge incidence matrix in tree \mathcal{T} . Each edge (ab) in the network is represented by a row equal to $(e_a^T - e_b^T)$ in M . Here e_a denotes the standard basis vector with 1 at the a^{th} position. Note that $\mathbf{1}^T P = \mathbf{1}^T P M^T f = 0$ where $\mathbf{1}$ is the vector of all ones. Thus the network is ‘lossless’ and total flow is conserved. Next, the flow $f_{(ab)}$ on edge (ab) and potentials π_a and π_b at nodes a and b satisfy

$$\pi_a - \pi_b = g_{ab}(f_{ab}) \quad (2)$$

where g_{ab} is the monotonic flow function for edge (ab) and can be distinct for each edge. Further, the flows and injections are unchanged if all nodal potentials are increased/decreased by the same amount. Following standard practice [5], [18], one node’s potential can be considered as reference and potentials are measures relative to that of the reference node. The substation or node with largest production of power or gas is generally considered as the reference node. We give the following examples of lossless flow networks.

Power Distribution Grid: Distribution grid [2] is the final tier of the power grid that extends from the distribution substation to the end-users. Flows in the radial distribution grid are composed of active and reactive power flows that are related to nodal voltages according to Kirchoff’s laws. During stable operations, the line flows can be expressed by the following lossless approximation commonly termed as LinDistFlow model [19]–[21]:

$$P_a = \sum_{\substack{(ab) \in \mathcal{E}^{\mathcal{T}} \\ b \neq a}} f_{ab}^p, Q_a = \sum_{\substack{(ab) \in \mathcal{E}^{\mathcal{T}} \\ b \neq a}} f_{ab}^q \quad (3)$$

$$v_a^2 - v_b^2 = 2(r_{ab}f_{ab}^p + x_{ab}f_{ab}^q) \forall (ab) \in \mathcal{E}^{\mathcal{T}} \quad (4)$$

Here P_a (Q_a) is the nodal active (reactive) power injection at node a while v_a is the voltage magnitude. f_{ab}^p (f_{ab}^q) is the active (reactive) flow on edge (ab) and r_{ab} (x_{ab}) is the line resistance (reactance). From Eq. (4), it is clear that squares of voltage magnitudes (v^2) represent potentials here and the flow function that relates potentials to line flow is linear. Similar linear flow functions in this area include linear coupled (LC) AC power

flow model [3], [22] with complex voltages as potentials and resistive DC power flow models [18] with phase angles as potentials.

Natural Gas Transmission Network: In gas grids, natural gas is driven from generators to consumers (households, gas turbines) through pipelines [4], [5]. During steady state, gas flow is governed by the following relation:

$$P_a = \sum_{(ab) \in \mathcal{E}^{\mathcal{T}}} f_{ab} \quad (5)$$

$$\phi_a^2 - \phi_b^2 = \alpha_{ab} f_{ab} |f_{ab}| + \beta_{ab} \forall (ab) \in \mathcal{E}^{\mathcal{T}} \quad (6)$$

Here P_a denotes the nodal injection. f_{ab} is the gas flux (flow per unit length) from node a to b . ϕ_a is the pressure at node a and its second power represents the potential. The quantity β_{ab} is the pressure boost provided by the compressor on edge (ab) and is constant over short time intervals where changes in flow are driven by difference in nodal potentials. Further, α_{ab} represents a constant factor that depends on friction, length and diameter of the pipe (edge (ab)) as well as temperature, universal gas constant and gas compressibility [5]. Note that the flow function here is second-order but monotonic.

Radial Water Network: Water networks consists of pipes where nodal ‘head pressures’ at their ends control the direction and quantity of flow in them [6], [23]. The flow equations are non-linear and similar to Eqs. (5),(6) for gas networks, but with a different exponent (> 1) for flow. We omit describing them mathematically for brevity. Similarly, radial **traffic networks** also satisfy conservation of flow at each node and can be modelled in a similar way [24].

We use $\mu_{g(X)}$ and $\Omega_{g(X)}$ to denote the mean and variance of function g defined over random variable X . Similarly, $\Omega_{g(X)h(Y)}$ denotes the covariance (centered second moment) of functions g and h defined over random variables X and Y respectively. Here X and Y may be correlated. Thus

$$\begin{aligned} \mu_{g(X)} &= \mathbb{E}[g(X)], \quad \Omega_{g(X)} = \mathbb{E}[(g(X) - \mu_{g(X)})^2], \\ \Omega_{g(X)h(Y)} &= \mathbb{E}[(g(X) - \mu_{g(X)})(h(Y) - \mu_{h(Y)})] \end{aligned} \quad (7)$$

In the next Section, we derive algebraic properties of second moments of nodal potentials in radial networks using the flow functions. These properties will help derive our learning algorithms.

III. TRENDS IN SECOND MOMENT OF POTENTIALS IN RADIAL NETWORKS

Let tree \mathcal{T} denote the operational radial flow network with edge set $\mathcal{E}_{\mathcal{T}}$. Without a loss of generality, we assume that all edges are directed towards the reference node. We denote the unique path (sequence of edges) from any node a to the reference node in tree \mathcal{T} by $\mathcal{P}_a^{\mathcal{T}}$. The set of all nodes b such that path $\mathcal{P}_b^{\mathcal{T}}$ passes through node a is called the ‘descendant’ set $D_a^{\mathcal{T}}$ of node a . By definition, $a \in D_a^{\mathcal{T}}$. If $b \in D_a^{\mathcal{T}}$ and $(ba) \in \mathcal{E}_{\mathcal{T}}$, we term a as parent and b as its child. See Fig. 2 for an illustrative example.

Eqs. (1,2) represent the relation between injections (P), flows (f) and potentials (π) in the network. As stated in the

previous section, the potential at the reference node is fixed, while its injection is given by negative sum of all other nodal injections (due to lossless property). The number of degrees of freedom in the injection or potential vector is thus one less than the number of nodes in the system. Following standard practice, we analyze the ‘reduced’ network of only non-reference nodes with nodal potentials measured relative to that of the reference node. We remove the column corresponding to the reference node from the incidence matrix M and omit its injection and potential terms from the vectors P and π respectively. Abusing notation, we use M , P and π to refer to the reduced versions of their respective definitions in the remaining part of this paper. Note that the reduced incidence matrix M has full rank. As all edges are directed towards the reference node, the inverse M^{-1} has the following specific analytical structure [25](also see Fig. 2):

$$M^{-1}(a, r) = \begin{cases} 1 & \text{if edge } r \in \mathcal{P}_a^T \\ 0 & \text{if edge } r \notin \mathcal{P}_a^T \end{cases} \quad (8)$$

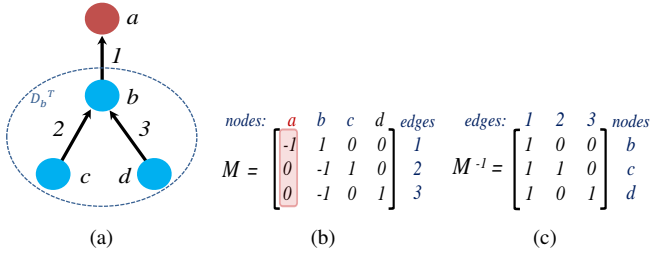


Fig. 2. (a) Radial network with four nodes (a, b, c, d) and three edges $(1, 2, 3)$ directed toward the reference node a . $\mathcal{P}_c^T = \{(cb), (ba)\}$. $\mathcal{D}_b^T = \{b, c, d\}$. b is the parent and c, d are its children nodes. (b) Reduced incidence matrix M derived by removing column corresponding to node a . (c) M^{-1} as per Eq. (8).

Using Eq. 1 with Eq. (8), the flow on edge (ab) with node a and its parent b satisfies:

$$f_{ab} = \sum_{c \in \mathcal{D}_a^T} P_c \quad (9)$$

Observe the flow relation in Eq. (2). Let k_1, k_2, \dots, k_r be the sequence of r intermediate nodes between a node k and its descendant a . Using a telescopic sum for nodal potentials, we have

$$\pi_k - \pi_a = \pi_k - \pi_{k_1} + \pi_{k_1} - \pi_{k_2} + \dots + \pi_{k_r} - \pi_a = \sum_{(cd) \in \mathcal{P}_a^T - \mathcal{P}_k^T} g_{cd}(f_{cd})$$

where $\mathcal{P}_a^T - \mathcal{P}_k^T$ consists of edges that lie in path from node a to k . For any two nodes a and b , we can find some node k on the path from a to b such that a, b are both descendants of k and $\mathcal{P}_b^T \cap \mathcal{P}_a^T = \mathcal{P}_k^T$. Here k may not be distinct from a or b . Note that $\mathcal{P}_b^T - \mathcal{P}_a^T = \mathcal{P}_b^T - \mathcal{P}_k^T$. Writing $\pi_a - \pi_b = (\pi_k - \pi_b) - (\pi_k - \pi_a)$ and using a telescopic sum, we get the following result.

Lemma 1. For two nodes a and b in the flow network

$$\pi_a - \pi_b = \sum_{(cd) \in \mathcal{P}_b^T - \mathcal{P}_a^T} g_{cd}(f_{cd}) - \sum_{(cd) \in \mathcal{P}_a^T - \mathcal{P}_b^T} g_{cd}(f_{cd}) \quad (10)$$

Before further analysis, we make the following assumption on probability distributions of different nodal injections as reported in literature [8], [13].

Assumption 1: Nodal injections at non-reference nodes in the network are independent.

Over short time intervals, this assumption is valid as injections are affected by changes/fluctuations in user behavior that are independent. In a subsequent section, we discuss extensions/scenarios where this assumption is relaxed. We now give the following definition for dependence between random variables that is well-studied in literature [26], [27].

Definition 1 [26], [27]: Two random variables are termed *Positive Quadrant Dependent (PQD)* if their probability distributions satisfies: $\mathbb{P}(X \leq x, Y \leq y) \geq \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$ for all x and y .

In other words, X and Y are PQD if larger (smaller) values of X are associated probabilistically with larger (smaller) values of Y . PQD for probability distributions of random variables can thus be thought as analogous to positive correlation for their second moments. Note that two independent random variables are PQD by definition. Further the following lemma holds:

Lemma 2. If X and Y are two independent random variables, then X and $X + Y$ are PQD.

The proof is listed in the appendix. Next, we state the following result without proof:

Lemma 3. [26, Theorem 2.4] Monotonic Functions of PQD random variables are positively correlated.

Using this we deduce the following result.

Theorem 1. Let $\mathcal{V}_1 \subset \mathcal{V}_2$ be nonempty sets of nodes in \mathcal{T} . Let $P_{\mathcal{V}_1} = \sum_{a \in \mathcal{V}_1} P_a$ and $P_{\mathcal{V}_2} = \sum_{a \in \mathcal{V}_2} P_a$. Then for any two flow functions g_i and g_j , $g_i(P_{\mathcal{V}_1})$ and $g_j(P_{\mathcal{V}_2})$ are positively correlated.

Proof: $P_{\mathcal{V}_1}$ and $P_{\mathcal{V}_2} - P_{\mathcal{V}_1}$ are independent as \mathcal{V}_1 and $\mathcal{V}_2 - \mathcal{V}_1$ are disjoint sets and nodal injections are independent. Using Lemma 2, $P_{\mathcal{V}_1}$ and $P_{\mathcal{V}_2} = P_{\mathcal{V}_1} + (P_{\mathcal{V}_2} - P_{\mathcal{V}_1})$ are PQD. The result follows from Lemma 3 as flow functions are monotonic. ■

We now analyze trends in second moments of nodal potentials using the preceding result. Denote the variance of potential difference $\pi_a - \pi_b$ as ϕ_{ab} . Using Eq. (7) and Lemma 1, we write ϕ_{ab} as follows:

$$\begin{aligned} \phi_{ab} &= \mathbb{E}[\pi_a - \pi_b - (\mu_{\pi_a} - \mu_{\pi_b})]^2 \\ &= \sum_{(jk), (st) \in \mathcal{P}_b^T - \mathcal{P}_a^T} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} + \sum_{(jk), (st) \in \mathcal{P}_a^T - \mathcal{P}_b^T} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} \\ &\quad - 2 \sum_{(jk) \in \mathcal{P}_b^T - \mathcal{P}_a^T, (st) \in \mathcal{P}_a^T - \mathcal{P}_b^T} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} \end{aligned} \quad (11)$$

If a is a descendant of b , $\mathcal{P}_b^T \subset \mathcal{P}_a^T$. Thus, $\mathcal{P}_b^T - \mathcal{P}_a^T$ is empty and Eq. (11) reduces to

$$\phi_{ab} = \sum_{(jk), (st) \in \mathcal{P}_a^T - \mathcal{P}_b^T} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} \quad (12)$$

Note that using Eq. (9), we can express flows on the right side of Eq. (12) and Eq. (11) in terms of the injections at

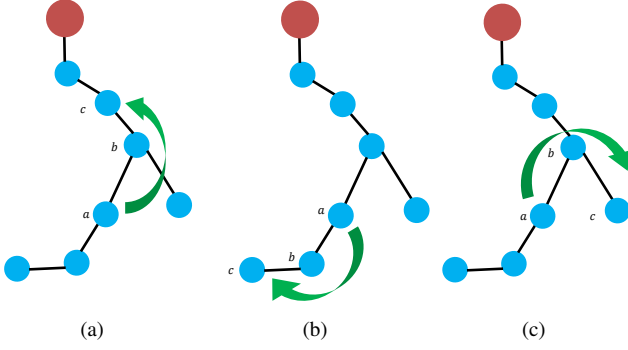


Fig. 3. Permissible configurations for nodes a , b and c when the path from a to c passes through b .

descendant nodes. The following theorem states a key trend that is observed in ϕ_{ab} for a radial flow network.

Theorem 2. Consider three nodes $a \neq b \neq c$ in the radial flow network such that the path from a to c passes through b . The variance of potential differences ϕ satisfies $\phi_{ab} < \phi_{ac}$.

Proof: As the path from a to c passes through b , there are three possible configurations for nodes a , b and c .

- 1) a is a descendant of node b , b is a descendant of c (see Fig. 3(a)).
- 2) c is a descendant of node b , b is a descendant of a (see Fig. 3(b)).
- 3) a and c are descendants of node b (see Fig. 3(c)).

To prove the theorem, we consider each case separately and prove the inequality $\phi_{ab} < \phi_{ac}$.

Case 1: Note that $\mathcal{P}_c^\mathcal{T} \subset \mathcal{P}_b^\mathcal{T} \subset \mathcal{P}_a^\mathcal{T}$. Thus,

$$\mathcal{P}_a^\mathcal{T} - \mathcal{P}_b^\mathcal{T} \subset \mathcal{P}_a^\mathcal{T} - \mathcal{P}_c^\mathcal{T}$$

. Using this in expression for ϕ_{ab} and ϕ_{ac} in Eq. (12) gives

$$\phi_{ac} - \phi_{ab} = \sum_{(jk),(st) \in \mathcal{P}_b^\mathcal{T} - \mathcal{P}_c^\mathcal{T}} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} + 2 \sum_{\substack{(jk) \in \mathcal{P}_a^\mathcal{T} - \mathcal{P}_b^\mathcal{T} \\ (st) \in \mathcal{P}_b^\mathcal{T} - \mathcal{P}_c^\mathcal{T}}} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} \quad (13)$$

Using Eq. (9), $f_{jk} = \sum_{r \in D_j^\mathcal{T}} P_r$ and $f_{st} = \sum_{r \in D_s^\mathcal{T}} P_r$. Note that for any two edges (jk) and (st) in $\mathcal{P}_a^\mathcal{T} - \mathcal{P}_c^\mathcal{T}$, $D_j^\mathcal{T} \subset D_s^\mathcal{T}$ or $D_s^\mathcal{T} \subset D_j^\mathcal{T}$ depending on which node is topologically nearer to a . If $(jk) \in \mathcal{P}_a^\mathcal{T} - \mathcal{P}_b^\mathcal{T}$, $(st) \in \mathcal{P}_b^\mathcal{T} - \mathcal{P}_c^\mathcal{T}$ then $D_s^\mathcal{T} \subset D_j^\mathcal{T}$. In either case, using Theorem 1, we have $\Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} > 0$. Thus, all terms in Eq. (13) are positive. Thus $\phi_{ab} < \phi_{ac}$.

Case 2: In this case, $\mathcal{P}_a^\mathcal{T} \subset \mathcal{P}_b^\mathcal{T} \subset \mathcal{P}_c^\mathcal{T}$. Following the analysis of Case 1, the expression of $\phi_{ac} - \phi_{ab}$ here becomes

$$\phi_{ac} - \phi_{ab} = \sum_{(jk),(st) \in \mathcal{P}_c^\mathcal{T} - \mathcal{P}_b^\mathcal{T}} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} + 2 \sum_{\substack{(jk) \in \mathcal{P}_c^\mathcal{T} - \mathcal{P}_b^\mathcal{T} \\ (st) \in \mathcal{P}_b^\mathcal{T} - \mathcal{P}_a^\mathcal{T}}} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})}$$

Using the same logic as Case 1, all covariance terms are positive valued and hence $\phi_{ab} < \phi_{ac}$.

Case 3: From Fig. 3(c) it is clear that the common edges on paths from nodes a and c to the reference node are the ones on the path from node b to the reference node. Thus

$\mathcal{P}_a^\mathcal{T} - \mathcal{P}_b^\mathcal{T} = \mathcal{P}_a^\mathcal{T} - \mathcal{P}_c^\mathcal{T}$ and $\mathcal{P}_c^\mathcal{T} - \mathcal{P}_a^\mathcal{T} = \mathcal{P}_c^\mathcal{T} - \mathcal{P}_b^\mathcal{T}$. Further, for any edge (jk) in $\mathcal{P}_a^\mathcal{T} - \mathcal{P}_b^\mathcal{T}$ and (st) in $\mathcal{P}_c^\mathcal{T} - \mathcal{P}_b^\mathcal{T}$, their respective descendant sets $D_j^\mathcal{T}$ and $D_s^\mathcal{T}$ are disjoint. By Eq. (9), flows f_{jk} and f_{st} are independent. The expression for ϕ_{ac} using Eq. (11) reduces to

$$\phi_{ac} = \sum_{(jk),(st) \in \mathcal{P}_a^\mathcal{T} - \mathcal{P}_b^\mathcal{T}} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} + \sum_{(jk),(st) \in \mathcal{P}_c^\mathcal{T} - \mathcal{P}_b^\mathcal{T}} \Omega_{g_{jk}(f_{jk})g_{st}(f_{st})} = \phi_{ab} + \phi_{cb} \quad (14)$$

where Eq. (14) follows from Eq. (12). As $\phi_{cb} > 0$, $\phi_{ab} < \phi_{ac}$. Thus, the statement holds as it is true for all configurations of a, b and c in the network. ■

In the next section, we use Theorem 2 to design our structure learning algorithm.

IV. STRUCTURE LEARNING WITH FULL OBSERVATION

The following theorem follows naturally from Theorem 2

Theorem 3. The set of operational edges in the radial flow network \mathcal{T} is given by the minimum spanning tree for the loopy graph of all permissible edges where each permissible edge (ab) is given weight $\phi_{ab} = \mathbb{E}[\pi_a - \pi_b - (\mu_{\pi_a} - \mu_{\pi_b})]^2$.

Proof: For each node a , the path to all nodes in the operational tree passes through one of its nearest neighbors (parent and children nodes). Using Theorem 2, weight ϕ_{ab} is, thus, minimum at the nearest neighbors in the tree. The spanning tree constructed using the operational edges thus has the minimum weight among all spanning trees formed from the set of permissible edges. ■

Algorithm 1: The algorithm for constructing the operational network is now straight forward. Using measurements for nodal potentials, permissible edges (ab) are given weights ϕ_{ab} and a spanning tree is constructed greedily by picking edges in the increasing order of their edge weights, while avoiding cycles. This is known as Kruskal's algorithm [28], [29]. If no information on permissible edges is available, then all potential node pairs are considered as permissible and the spanning tree is constructed from the complete graph (every node pair is connected). Note that no information on flow function or nature of probability distribution for individual nodal injections are necessary in Algorithm 1.

Algorithm 1 Structure Learning using Potential Measurements

Input: m potential measurements π for all nodes, set of all permissible edges \mathcal{E} (optional).

Output: Operational Edge set $\mathcal{E}_\mathcal{T}$.

- 1: Compute $\phi_{ab} = \mathbb{E}[(\pi_a - \mu_{\pi_a}) - (\pi_b - \mu_{\pi_b})]^2$ for all permissible edges
 - 2: Find minimum weight spanning tree from \mathcal{E} with ϕ_{ab} as edge weights.
 - 3: $\mathcal{E}_\mathcal{T} \leftarrow$ spanning tree edges
-

Algorithm Complexity: Kruskal's Algorithm learns the minimum spanning tree in quasi-linear time in the number of permissible edges in the system. The computational complexity of learning the operational tree is $O(|\mathcal{E}| \log |\mathcal{E}|)$ where \mathcal{E} is the set of all permissible edges. If no information on

set \mathcal{E} is available, then the complexity (worst-case) becomes $O(|\mathcal{V}|^2 \log |\mathcal{V}|)$ which is quasi-quadratic in $|\mathcal{V}|$, the number of nodes in the network.

We now discuss a few extensions of our algorithm to generalized cases.

Extension to Multiple Trees: Our learning algorithm and analysis can be immediately extended to networks with multiple operational trees. In each tree, one can denote one reference node and compute potentials relative to that. Potentials at multiple trees will be uncorrelated and can be separated into different groups before running Algorithm 1.

Learning Flow Functions/Statistics of Nodal Injection: Note that in our algorithm, no information on flow functions or nodal consumption statistics is necessary. However, if either one of them is known (flow function or statistics of injection) in addition to potential measurements, the other one can be estimated. To obtain this, Algorithm 1 is first used to learn the structure of the grid and then Eqs. (2), (9) can be used to recursively estimate the flow function or the injection statistics from the leaves up to the reference node.

Learning with Missing Nodes: This refers to the regime where a section of the nodes are measured and potential measurements for others are not available. To learn the structure in the presence of missing nodes, we need additional information pertaining to permissible flow functions and nodal injection statistics. In that case, a modification of Algorithm 1 can be proposed, where the available measurements of potentials are used to generate a spanning tree without the missing nodes. At the next level, Eqs. (11) and Eqs. (12) can be used recursively to identify the presence of missing nodes. We plan to expand on this aspect in a future work.

V. EXPERIMENTS

In this section, we discuss the performance of Algorithm 1 in learning the operational radial structure of flow networks using nodal potential measurements as input. To demonstrate the general nature of our work, we present simulation results on two radial networks: a power distribution grid with linear flow function (Fig. 4(a)), and a gas transmission grid with quadratic flow function (Fig. 4(b)). The power distribution grid [30], [31] consists of 30 nodes, while the gas grid [32] consists of 25 nodes. One node is denoted as the reference node with constant potential.

To conduct a simulation on either grid, we first generate injection samples at each non-reference node using an uncorrelated multivariate Gaussian distribution. Then flow equations (LinDisFlow Eqs. (3,4) for power grid and Eqs. (5,6) for gas grid) are used to derive input nodal potential measurements (squares of voltage magnitude for power grid and squares of pressures for gas grid). Further, fictitious edges (numbering 30 for power grid and 25 for gas grid) are introduced into the loopy set of permissible edges \mathcal{E} along with the true operational edges. This is done to observe the performance of structure estimation in Algorithm 1. The potential measurements and set \mathcal{E} are sent as input to Algorithm 1.

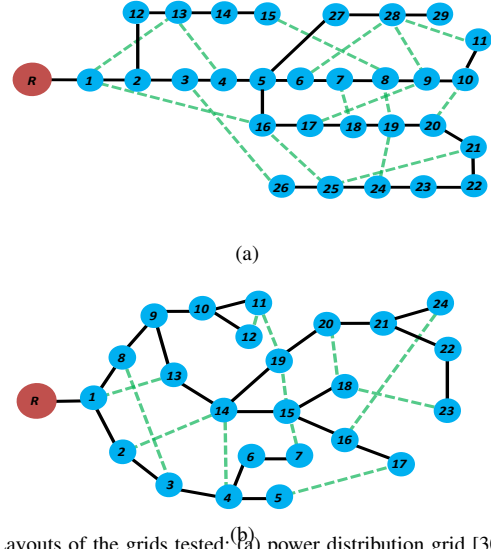
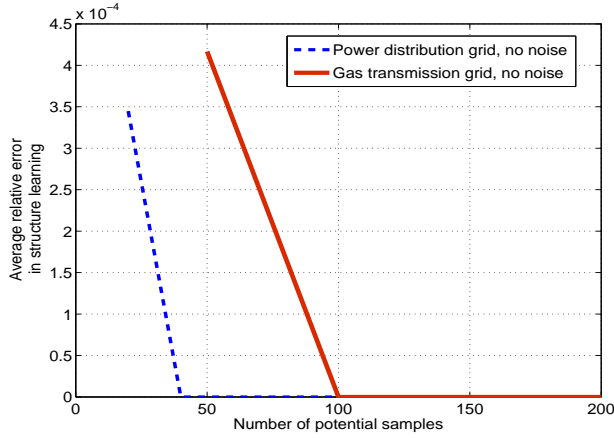


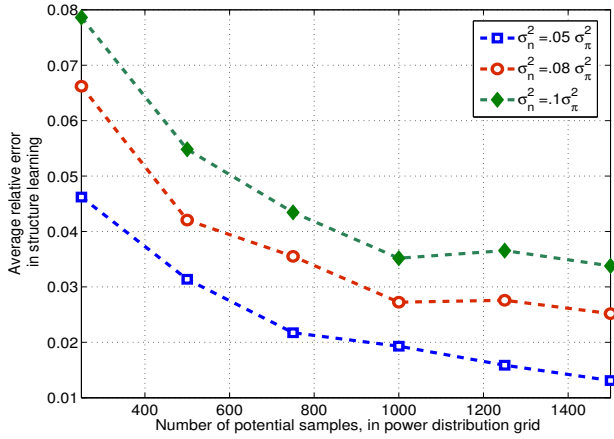
Fig. 4. Layouts of the grids tested: (a) power distribution grid [30], [31] (b) gas transmission grid [32]. The red and blue circles denote the reference and non-reference nodes respectively in either grid. Operational edges are colored solid black while some of the fictitious non-operational edges are denoted by dotted green lines.

We measure average errors produced in determining the true structure and express them relative to the number of operational edges. To demonstrate the performance for either grid, we plot the trend in average relative errors in Algorithm 1 versus the number of nodal potential samples available as input. We first consider the case with no measurement noise in Fig. 5(a). Notice that the performance is excellent and errors quickly decay to zero for both grids. In fact perfect recovery is observed for samples sizes greater than 100. All errors here are induced by finite sample sizes that lead to imperfect empirical estimation of ϕ_{ab} .

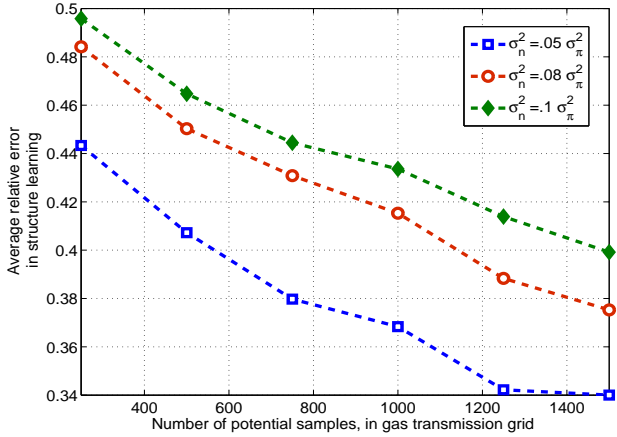
Next, we present performance of Algorithm 1 when potential samples are corrupted with additive noise. We consider the potential samples at each node to suffer from additive Gaussian noise of mean 0 and variance of value expressed as a fraction of the average variance of nodal potentials. In either grid, we consider three fractions (8×10^{-2} , 5×10^{-2} , 10^{-1}) to represent different levels of noise that are commensurate with noise suffered in off-the shelf measurement devices. Fig. 5(b) and Fig. 5(c) shows the performance with noise for the power and gas grids respectively. Note that the average fractional errors recorded for either grid go down with increase in the number of samples, though the decay is much slower than in Fig. 5(a) with no noise. Further, as expected, the errors increase with an increase in the noise variance. It can be observed that the error performance in power grid is significantly better than that in gas networks. This can be explained on the basis of quadratic flow functions in the latter which induce greater errors in empirical approximation of ϕ as compared to linear flow functions in the power distribution grid. We plan to theoretically analyze the error performance in detail in future work.



(a)



(b)



(c)

Fig. 5. Average fractional errors vs number of samples used in Algorithm 1 for (a) Power distribution and gas transmission networks with no measurement noise (b) Power distribution network with Gaussian noise (c) Gas transmission network with Gaussian noise. Different noise variances σ_n are taken as fractions (.05, .08, .1) relative to variance σ_π in nodal potential.

VI. CONCLUSIONS

Flow networks represent several key infrastructure including power grid, gas grid, water and residential heating networks.

Despite the diversity of network traffic, the flow in each network is driven by nodal potentials that are related to the edge flows by a class of nonlinear monotonic flow functions. This paper addresses the problem of estimating the structure of radial flow networks using measurements of nodal potentials. Using properties of positive quadrant dependent functions, we show that the variance of potential differences has provable ordering properties along the network edges. Based on this, a spanning tree based learning algorithm is proposed that can learn the network using only nodal potential statistics. The significant aspect of this algorithm is that it does not require any knowledge of the edge flow functions or specific marginal distributions of nodal injections. This work thus presents the first approach to learning general radial networks with nonlinear flows. The performance of our algorithms are demonstrated through simulations on test radial networks pertaining to a power system and a gas grid. We discuss extensions of our framework, including that with missing/unobserved nodes. Efficient learning of the network structure using potential measurements has application in control and optimization applications as well as in identifying the estimation capability of third parties possessed with limited information. In addition to the extensions mentioned in the paper, potential areas of future work include expanding the learning framework to lossy flows and loopy networks and understanding the sample complexity associated with learning in the presence of noisy measurements.

APPENDIX

Proof of Lemma 2: To prove X and $X + Y$ are PQD, we need to show

$$\begin{aligned} \mathbb{P}(X + Y \leq b, X \leq a) &\geq \mathbb{P}(X \leq a) \mathbb{P}(X + Y \leq b) \forall a, b \\ \Rightarrow \mathbb{P}(X + Y \leq b | X \leq a) &\geq \mathbb{P}(X + Y \leq b) \end{aligned} \quad (15)$$

Since we are dealing with physical random variables (power injection etc.), we assume that their probability distribution functions exist. Let ρ_X and ρ_Y denote the probability distribution functions (p.d.f.s) for X and Y respectively. The p.d.f. for $X + Y$ conditioned on $X \leq a$ is given by:

$$\rho_{X+Y|X \leq a}(z) = \int_{-\infty}^a \rho_{X+Y|X=x}(z) \rho_{X|X \leq a}(x) dx \quad (16)$$

$$= \frac{\int_{-\infty}^a \rho_Y(z-x) \rho_X(x) dx}{\int_{-\infty}^a \rho_X(x) dx} \quad (17)$$

$$\begin{aligned} \Rightarrow \mathbb{P}(X + Y \leq b | X \leq a) &= \frac{\int_{-\infty}^b \int_{-\infty}^a \rho_Y(z-x) \rho_X(x) dx dz}{\int_{-\infty}^a \rho_X(x) dx} \\ &= \frac{\int_{-\infty}^a \mathbb{P}(Y \leq b-x) \rho_X(x) dx}{\int_{-\infty}^a \rho_X(x) dx} \end{aligned} \quad (18)$$

Here, Eq. (16) follows from the chain rule of conditional probability. Eq. (17) uses the fact that the p.d.f. for $X + Y$ conditioned on $X = a$ is given by $\rho_{X+Y|X=a}(z) = \rho_Y(z-a)$ as X and Y are independent. Eq. (18) follows from changing the order of variables x and z under the integration. The right hand side of Eq. (18) represents the weighted average of

$\mathbb{P}(Y \leq b-x)$ with weight $\frac{\rho_X(x)}{\int_{-\infty}^a \rho_X(x)dx}$ in $(\infty, a]$ and 0 otherwise. The derivative of $\mathbb{P}(X+Y \leq b|X \leq a)$ with a is non-positive as shown below:

$$\begin{aligned} & \frac{d}{da} \frac{\int_{-\infty}^a \mathbb{P}(Y \leq b-x) \rho_X(x) dx}{\int_{-\infty}^a \rho_X(x) dx} \\ & \propto \mathbb{P}(X \leq a) \mathbb{P}(Y \leq b-a) - \int_{-\infty}^a \mathbb{P}(Y \leq b-x) \rho_X(x) dx \\ & \leq \mathbb{P}(X \leq a) \mathbb{P}(Y \leq b-a) - \mathbb{P}(Y \leq b-a) \int_{-\infty}^a \rho_X(x) dx \leq 0 \end{aligned} \quad (19)$$

The inequality holds as $\mathbb{P}(Y \leq b-x)$ is a decreasing function of x .

Thus $\mathbb{P}(X+Y \leq b|X \leq a)$ is non-increasing in a and hence proved that

$$\mathbb{P}(X+Y \leq b|X \leq a) \leq \lim_{a \rightarrow \infty} \mathbb{P}(X+Y \leq b|X \leq a) = \mathbb{P}(X+Y \leq b).$$

ACKNOWLEDGMENT

The authors thank S. Misra and A. Zlotnik at Los Alamos National Laboratory for providing information regarding the test gas network used for simulations in the paper.

REFERENCES

- [1] R. S. Dembo, J. M. Mulvey, and S. A. Zenios, "Or practicelarge-scale nonlinear network models and their application," *Operations Research*, vol. 37, no. 3, pp. 353–372, 1989.
- [2] R. Hoffman, "Practical state estimation for electric distribution networks," in *Power Systems Conference and Exposition, 2006. PSCE'06. 2006 IEEE PES*. IEEE, 2006, pp. 510–517.
- [3] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning and statistical estimation in distribution networks - part i," *arXiv preprint arXiv:1501.04131*, 2015.
- [4] S. Mokhtab and W. A. Poe, *Handbook of natural gas transmission and processing*. Gulf Professional Publishing, 2012.
- [5] S. Misra, M. W. Fisher, S. Backhaus, R. Bent, M. Chertkov, and F. Pan, "Optimal compression in natural gas networks: a geometric programming approach," *Control of Network Systems, IEEE Transactions on*, vol. 2, no. 1, pp. 47–56, 2015.
- [6] L. Sela Perelman and S. Amin, "Control of tree water networks: A geometric programming approach," *Water Resources Research*, vol. 51, no. 10, pp. 8409–8430, 2015.
- [7] M. He and J. Zhang, "A dependency graph approach for fault detection and localization towards secure smart grid," *Smart Grid, IEEE Transactions on*, vol. 2, no. 2, pp. 342–351, 2011.
- [8] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 1659–1664.
- [9] G. Cavarro, R. Arghandeh, A. von Meier, and K. Poola, "Data-driven approach for distribution network topology detection," *arXiv preprint arXiv:1504.00724*, 2015.
- [10] J. Peppanen, J. Grimaldo, M. J. Reno, S. Grijalva, and R. G. Harley, "Increasing distribution system model accuracy with extensive deployment of smart meters," in *PES General Meeting—Conference & Exposition, 2014 IEEE*. IEEE, 2014, pp. 1–5.
- [11] J. Peppanen, M. J. Reno, M. Thakkar, S. Grijalva, and R. G. Harley, "Leveraging ami data for distribution system model calibration and situational awareness," 2015.
- [12] R. Sevlian and R. Rajagopal, "Feeder topology identification," *arXiv preprint arXiv:1503.07224*, 2015.
- [13] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning and statistical estimation in distribution networks - part ii," *arXiv preprint arXiv:1502.07820*, 2015.
- [14] —, "Learning topology of the power distribution grid with and without missing data," *arXiv preprint arXiv:1603.01650*, 2016.
- [15] M. Vuffray, S. Misra, and M. Chertkov, "Monotonicity of dissipative flow networks renders robust maximum profit problem tractable: General analysis and application to natural gas flows," in *Decision and Control (CDC), IEEE 52nd Annual Conference on*. IEEE, 2015.
- [16] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *Information Theory, IEEE Transactions on*, vol. 14, no. 3, pp. 462–467, 1968.
- [17] M. J. Choi, V. Y. Tan, A. Anandkumar, and A. S. Willsky, "Learning latent tree graphical models," *The Journal of Machine Learning Research*, vol. 12, pp. 1771–1812, 2011.
- [18] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC Press, 2004.
- [19] M. Baran and F. Wu, "Optimal sizing of capacitors placed on a radial distribution system," *Power Delivery, IEEE Transactions on*, vol. 4, no. 1, pp. 735–743, Jan 1989.
- [20] —, "Optimal capacitor placement on radial distribution systems," *Power Delivery, IEEE Transactions on*, vol. 4, no. 1, pp. 725–734, Jan 1989.
- [21] —, "Network reconfiguration in distribution systems for loss reduction and load balancing," *Power Delivery, IEEE Transactions on*, vol. 4, no. 2, pp. 1401–1407, Apr 1989.
- [22] S. Bolognani and S. Zampieri, "On the existence and linear approximation of the power flow solution in power distribution networks," *Power Systems, IEEE Transactions on*, vol. 31, no. 1, pp. 163–172, 2016.
- [23] P. F. Boulos, K. E. Lansey, and B. W. Karney, *Comprehensive water distribution systems analysis handbook for engineers and planners*. American Water Works Association, 2006.
- [24] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, "On robustness analysis of large-scale transportation networks," in *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems—MTNS*, vol. 5, no. 9, 2010.
- [25] J. Resh, "The inverse of a nonsingular submatrix of an incident matrix," *IEEE Transactions on Circuit Theory*, vol. 10, p. 131.
- [26] M. Egozcue, L. F. Garcia, and W.-K. Wong, "On some covariance inequalities for monotonic and non-monotonic functions," *Journal of Inequalities in Pure and Applied Mathematics*, vol. 10, no. 3, pp. 1–7, 2009.
- [27] E. L. Lehmann, "Some concepts of dependence," *The Annals of Mathematical Statistics*, pp. 1137–1153, 1966.
- [28] J. B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proceedings of the American Mathematical Society*, vol. 7, no. 1, pp. 48–50, 1956.
- [29] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. The MIT Press, 2001.
- [30] U. Eminoglu and M. H. Hocaoglu, "A new power flow method for radial distribution systems including voltage dependent load models," *Electric Power Systems Research*, vol. 76, no. 13, pp. 106 – 114, 2005.
- [31] [Online]. Available: <http://www.dejazzer.com/reds.html>
- [32] A. Zlotnik, M. Chertkov, and S. Backhaus, "Optimal control of transient flow in natural gas networks," in *Decision and Control (CDC), IEEE 52nd Annual Conference on*. IEEE, 2015.