Hierarchical Clustering of Dynamical Networks Using a Saddle-Point Analysis

Mathias Bürger, Daniel Zelazo, and Frank Allgöwer

***Abstract*—This paper studies cluster synchronization in dynamical networks. A class of cooperative dynamical networks that exhibit clustering in their asymptotic behavior is analyzed. The network nodes are equipped with heterogeneous dynamics and interact with a nonlinear and saturated interaction rule. It is proven that cluster synchronization appears asymptotically independent of the initial conditions. The clustering behavior of the dynamic network is shown to correspond to the solution of a static saddle-point problem, enabling a precise characterization of the clustering structure. We show how the clustering structure depends on the relation between the underlying graph, the node dynamics, and the saturation level of the interactions. This interpretation leads to deeper combinatorial insights related to clustering, including a generalization of classical network partitioning problems such as the inhibiting bisection problem, the min $s-t$-cut problem, and hierarchical clustering analysis. The theoretical results are applied for the analysis of a test-case network, inspired by the IEEE 30-bus system.**

***Index Terms*—Cluster synchronization, dynamical networks, saddle-point optimization.**

I. Introduction

CLUSTERING, or cluster synchronization, is the phenomenon that in a network of dynamical systems, the network partitions into several groups and all systems within the same group agree upon a common state. Observed across diverse fields ranging from the brain sciences [1] to social networks [2], clustering has recently received significant attention [3].

In engineering systems, clustering may represent desirable or undesirable behaviors. The importance of clustering in engineering is most readily understood by the related problem of synchronization, where each agent in the network should agree upon a common value. For example, the synchronization of power networks is crucial for its stability, while frequency clustering in the network can lead to catastrophic failures. Consequently, for the design and the control of dynamical networks, it is imperative to develop analytic tools for understanding the mechanisms leading to this behavior.

For agreement and synchronization, an advanced control theory has been recently developed [4], including a variety of novel stability-analysis techniques ranging from passivity [5] to contraction analysis [6]. In contrast, the mechanisms leading to clustering are still not fully understood, and a control-theoretic approach to this phenomena is only beginning to emerge. For example, [7] and [8] consider how networks can be forced to cluster according to predefined structures using pinning control or adaptive interaction weights. The authors of [9] study different mechanisms leading to clustering in diffusively coupled networks, including structured dynamics, delays, and negative couplings. Some other models are proposed that exhibit clustering inherently by their dynamical properties without having the clustering structure specified *a priori*. One of the most celebrated clustering models is the "bounded confidence opinion dynamics" model [10], where clustering is caused by a state-dependent communication graph. Although the model itself is fairly simple, no theory currently exists that allows one to predict the resulting clustering structure. In [11] and [12], a clustering model is presented where the partitioning of the network is caused by different driving forces applied to the agents along with saturated interaction rules. The development of the aforementioned models is driven by an increasing interest in understanding the mechanisms leading to clustering.

Despite these advances, explaining how this behavior is related to the topological structure of the underlying graph remains an open problem. Problems such as these, known as (static) *community detection* problems in graphs [13], are only partially explained using the current models. In several applications it is, for example, of interest to detect which parts of the network are strongly connected (with respect to the dynamic behavior) and which connections are critical, in the sense that the network is likely to partition along them. Such an analysis is particularly difficult if the number of partitions is not specified. This motivates the search for *hierarchical clustering structures* in dynamical networks.

This paper aims to address aspects of the aforementioned open problems. We consider a general class of complex dynamic networks that exhibit clustering in their steady-state behavior. The distinguishing features in the model we adopt are: 1) the uncoupled node dynamics have distinct equilibria and 2) the interaction rules between neighboring agents are bounded. We show that the network synchronizes for sufficiently large saturation bounds, but partitions into clusters otherwise. To analyze the clustering structure precisely, we connect the asymptotic behavior of the dynamic network model to a specific convex static saddle-point optimization problem. We show that the solution of

Manuscript received June 10, 2011; revised November 12, 2011, March 28, 2012, and June 14, 2012; accepted June 21, 2012. Date of publication June 29, 2012; date of current version December 17, 2012. This work was supported in part by the German Research Foundation (DFG) within the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart and in part by the Priority Programme 1305 "Control Theory of Digitally Networked Dynamical Systems." Recommended by Associate Editor S. Mascolo.

M. Bürger and F. Allgöwer are with the Insitute for Systems Theory and Automatic Control, University of Stuttgart, Stuttgart 70550, Germany (e-mail: mathias.buerger@ist.uni-stuttgart.de; frank.allgower@ist.uni-stuttgart.de).

D. Zelazo is with the Faculty of Aerospace Engineering, Technion -Israel Institute of Technology, Haifa 32000, Israel (e-mail: dzelazo@technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2012.2206695

the saddle-point problem, corresponding to the Lagrange-dual of a network optimization problem with additional constraints on the dual variables, exhibits a clustered structure. This result allows us to exactly predict the clustering structure using computationally efficient methods. We show then that the clustering structure of the network is related to optimal graph partitioning: it maximizes the ratio of the "power imbalance" between the partitions over the number of edges connecting them. This result establishes a direct connection between dynamical clustering and graph partitioning. We show how our setup connects to the *inhibiting bisection problem*, used in the literature for power network analysis [14]. Furthermore, for a particular choice of node dynamics, the clustering structure of our method solves the min $s-t$-cut problem [15]. Finally, we illustrate our theoretical results on a test case example (i.e., the IEEE 30-bus power system). We use the saddle-point problem to analyze the hierarchical clustering structure of this network and perform dynamic simulations of our system model defined on the given graph structure.

The remainder of the paper is organized as follows. In Section II, some graph-theoretic results are reviewed and the notion of clustering and hierarchical clustering, as used in this paper, is defined. The dynamical network is presented in Section III. Following this, the static saddle-point problem is presented and analyzed in Section IV. The connection between the dynamical network and the static saddle-point problem is established in Section V. In Section VI, we discuss how the clustering structure of the dynamical network is connected to optimal graph partitioning and we relate our result to the inhibiting bisection problem and the min $s-t$-cut. The theory is applied to a test-case network in Section VII and concluding remarks are given in Section VIII.

Notation: The set of real numbers is denoted by $\mathbb{R}$. We write $\mathbb{R}_{\geq 0}$ for all non-negative real numbers. For a vector $x \in \mathbb{R}^{|\mathbf{V}|}$, its transpose is given by x' and the ith component by x_i. The ijth element of a matrix A is denoted by $[A]_{ij}$. The inner product of two vectors is denoted as $\langle x, y\rangle = x'y$, and the standard Euclidean norm is $\|x\| = \langle x, x\rangle^{1/2}$. We also sometimes use the infinity norm $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$ and the 1-norm $\|x\|_1 = \sum_{i=1}^n |x_i|$. The null space and range space of a matrix is denoted as $\mathcal{N}(A)$ and $\mathcal{R}(A)$, respectively. The boundary of a set Γ is denoted as $\partial\Gamma$ and the interior by $\mathrm{int}\Gamma$. The vector $\mathbb{1}_n$ is the vector of all ones with length n. We omit the subscript if the dimension of the vector is unambiguous from the context.

II. Preliminaries

Throughout this paper, we consider systems defined over undirected graphs [16]. A *graph* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ consists of a set of *nodes* $\mathbf{V} = \{v_1, \ldots, v_{|\mathbf{V}|}\}$ and a set of *edges* $\mathbf{E} = \{e_1, \ldots, e_{|\mathbf{E}|}\}$, describing the incidence relation between pairs of nodes. For the formulation of the results, it is, however, convenient to introduce an arbitrary orientation on the edges. The notation $v_i \sim v_j$ denotes that node v_i is connected (or adjacent) to node v_j. Equivalently, $e = (v_i, v_j) \in \mathbf{E}$ is the *directed* edge connecting v_i and v_j. A simple *path* is a sequence of distinct nodes so that consecutive nodes are adjacent to each other in the undirected graph, and each node is used once. A simple *cycle* in a graph is a path where the initial and terminal nodes are the same. A graph is *connected* if a path exists between any pair of nodes. We also use the convention that an isolated vertex is a connected graph.

A graph $\mathcal{G}' = (\mathbf{V}', \mathbf{E}')$ is a *subgraph* of $\mathcal{G}$ if $\mathbf{V}' \subseteq \mathbf{V}$ and $\mathbf{E}' \subseteq \mathbf{E}$; equivalently, we write $\mathcal{G}' \subseteq \mathcal{G}$. Subgraphs can be induced by either a node set or an edge set.[1] A disconnected graph can be expressed as the union of connected subgraphs; each connected subgraph is referred to as a *component* of $\mathcal{G}$. Throughout this paper, we follow the convention that boldfaced capital letters refer to sets, as in $\mathbf{V}$, and the script notation for graphs, as in $\mathcal{Q}$.

The *incidence matrix* $E(\mathcal{G}) \in \mathbb{R}^{|\mathbf{V}|\times|\mathbf{E}|}$ of the graph $\mathcal{G}$ with arbitrary orientation, is a $\{0, \pm 1\}$ matrix with the rows and columns indexed by the nodes and edges of $\mathcal{G}$ so that $[E(\mathcal{G})]_{ik}$ has value "$+1$" if node i is the initial node of edge k, "-1" if it is the terminal node, and "0" otherwise. This definition implies that for any graph, $\mathbb{1}'E(\mathcal{G}) = 0$. The *Laplacian matrix* $L(\mathcal{G})$ of the undirected graph $\mathcal{G}$ is defined as $L(\mathcal{G}) = E(\mathcal{G})E(\mathcal{G})'$. At times, we will refer to the *flow space* (*null space*) and the *cut space* (*range space*) of the incidence matrix, defined as $\mathcal{N}(E(\mathcal{G}))$ and $\mathcal{R}(E(\mathcal{G})')$, respectively [16]. The cycles in a graph provide an important characterization of the flow space.

Definition 2.1: A *signed path vector* $\zeta \in \mathbb{R}^{|\mathbf{E}|}$ of a connected graph $\mathcal{G}$ corresponds to a path so that the ith element of ζ takes the value "$+1$" if edge i is traversed positively, "-1" if traversed negatively, and "0" if the edge is not used in the path.

Theorem 2.2 ([16]): The flow space $\mathcal{N}(E(\mathcal{G}))$ of a connected graph $\mathcal{G}$ is spanned by all of the linearly independent signed path vectors corresponding to the cycles in $\mathcal{G}$.

We now provide some definitions related to graph partitioning and clustering.

Definition 2.3: A *cluster* $\mathcal{P}$ is a connected subgraph of $\mathcal{G}$ induced by a node set $\mathbf{P} \subseteq \mathbf{V}$.

Definition 2.4: A *p-Partition* of the graph $\mathcal{G}$ is a collection of node sets $\mathbb{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_p\}$ with $\mathbf{P}_i \subseteq \mathbf{V}$, $\cup_{i=1}^p \mathbf{P}_i = \mathbf{V}$, and $\mathbf{P}_i \cap \mathbf{P}_j = \emptyset$ for all $\mathbf{P}_i, \mathbf{P}_j \in \mathbb{P}$, so that each subgraph $\mathcal{P}_i$ induced by the node sets $\mathbf{P}_i$ is connected.

Each subgraph $\mathcal{P}_i$ induced by a p-partition is also a cluster. At times, we will also refer to the *p-cluster* of a graph to mean the set of subgraphs induced by a p-partition. For a connected graph $\mathcal{G}$, the union of all clusters induced by a partition will not reconstruct the original graph ($p \geq 2$); that is, $\cup_{i=1}^p \mathcal{P}_i \subset \mathcal{G}$. This is formalized by the definition of a *cutset*.

Definition 2.5: A *cutset* of the graph $\mathcal{G}$ is a set of edges whose deletion leads to an increase in the number of connected components in $\mathcal{G}$.

According to this definition, a cutset always induces a p-partition ($p \geq 2$). Similarly, any p-partition ($p \geq 2$) of a graph will induce a cutset. In this case, the cutset is defined as $\mathbf{Q} = \{(v_i, v_j) \in \mathbf{E} | v_i \in \mathbf{P}_k, v_j \in \mathbf{P}_l, \forall\, \mathbf{P}_k, \mathbf{P}_l \in \mathbb{P}, k \neq l\}$.

We denote the *set of all possible p-partitions* that can be formed in the graph $\mathcal{G}$ by $\mathbb{P}^p(\mathcal{G})$. Partitions of $\mathcal{G}$ are sometimes ordered in a *hierarchical* manner, with smaller partitions being contained in the larger partitions.

[1]For example, the subgraph $\mathcal{P} \subseteq \mathcal{G}$ induced by the node set $\mathbf{P} \subseteq \mathbf{V}$ is the graph $\mathcal{P} = (\mathbf{P}, \mathbf{E}')$, with $\mathbf{E}' = \{e = (v_i, v_j) | v_i, v_j \in \mathbf{P}, e \in \mathbf{E}\}$. Similarly, the subgraph $\mathcal{Q} \subseteq \mathcal{G}$ induced by the edge set $\mathbf{Q} \subseteq \mathbf{E}$ is the graph $\mathcal{Q} = (\mathbf{V}', \mathbf{Q})$, with $\mathbf{V}' \subseteq \mathbf{V}$ being the set of all nodes incident to the edges in $\mathbf{Q}$.

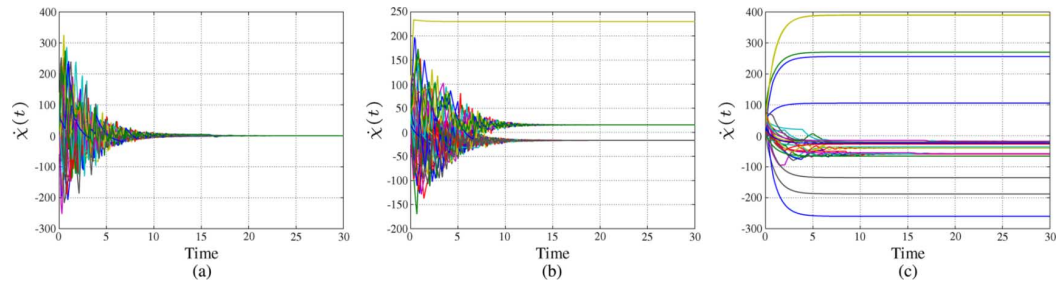


Fig. 1. Three simulations of the trajectories $\dot{\chi}(t)$ with different edge capacities. The network reaches agreement for (a) $\alpha_{[1]}$, (b) partitions into a small number of clusters for $\alpha_{[2]}$, or (c) separates almost all nodes for $\alpha_{[3]}$.

Definition 2.6: A partition $\mathbb{P}_i \in \mathbb{P}^k(\mathcal{G})$ is a *successor* of $\mathbb{P}_j \in \mathbb{P}^l(\mathcal{G})$ with $k > l$, denoted by $\mathbb{P}_i \succ \mathbb{P}_j$, if $\mathbb{P}_j$ can be formed by merging components from $\mathbb{P}_i$. We write $\mathbb{P}_i \succeq \mathbb{P}_j$ if $\mathbb{P}_i$ is either a successor or exactly identical to $\mathbb{P}_j$.

Throughout this paper, we associate scalar variables with each node and edge in a graph. For example, each component x_i of the vector $x \in \mathbb{R}^{|\mathbf{V}|}$ is associated with a node $v_i \in \mathbf{V}$. Similarly, each component z_i of a vector $z \in \mathbb{R}^{|\mathbf{E}|}$ is associated with an edge $e_i \in \mathbf{E}$. This can be used to provide an additional characterization of clusters and partitions of a graph.

Definition 2.7: A cluster $\mathcal{P}$ is in *agreement* if $\|x_k - x_l\| = 0$, for all $v_k, v_l \in \mathbf{P}$.

For a vector $x \in \mathbb{R}^{|\mathbf{V}|}$ defined on the nodes of $\mathcal{G}$ and a subgraph $\mathcal{P} = (\mathbf{P}, \mathbf{E}') \subseteq \mathcal{G}$, we write $x(\mathcal{P}) \in \mathbb{R}^{|\mathbf{P}|}$ to denote the vector of all components x_j associated with the nodes $v_j \in \mathbf{P}$. Similarly, for a vector $z \in \mathbb{R}^{|\mathbf{E}|}$ defined on the edges of $\mathcal{G}$, the vector $z(\mathcal{P})$ is the vector containing all variables associated with an edge of the graph $\mathcal{P}$. Using this notation, we can express the values of a cluster in agreement as $x(\mathcal{P}) = \beta \mathbb{1}_{|\mathbf{P}|}$, where β is the *cluster value*.

III. Dynamical Network Model for Clustering

We begin our study of clustering in networked dynamic systems by first presenting a general class of second-order, diffusively coupled systems. Models of this type are used for various applications including vehicle platoons [17] and power networks [18]. This class of model is also extensively used to study the phenomena of synchronization in networked systems [5], [19]. The model is based on a collection of dynamic systems that interact over an undirected graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ according to the second-order differential equation

$$\ddot{\chi}_i = -f_i(\dot{\chi}_i) + \sum_{e_k=(v_j,v_i)\in\mathbf{E}} \alpha_k \psi_k(\chi_j - \chi_i), \quad i \in \{1, \ldots, |\mathbf{V}|\} \tag{1}$$

where $\chi_i \in \mathbb{R}$ is the dynamic state of each agent in the ensemble. A first result of this paper is to show that under certain assumptions on the functions $f_i(\cdot)$ and $\psi_k(\cdot)$, the model (1) exhibits a *clustering* phenomena.

Assumption 3.1: The functions $f_i(\cdot)$ and $\psi_k(\cdot)$ in (1) satisfy the following properties:

- $(f_i(s) - f_i(\tilde{s}))(s - \tilde{s}) \geq \eta_i (s - \tilde{s})^2$, for some $\eta_i > 0$ and all $s, \tilde{s} \in \mathbb{R}$ (i.e., each $f_i(\cdot)$ is the gradient of a strongly convex function $F_i(s)$).
- for each node i, a unique s_i^{eq} exists for which $f_i(s_i^{\text{eq}})0 = 0$ and $s_i^{\text{eq}} \neq s_j^{\text{eq}}$, for all $i, j \in \mathbf{V}$.
- the coupling functions $\psi_k(s)$ vanish at the origin ($\psi_k(0) = 0$), monotonically increase, and are bounded as

$$\lim_{s\to\infty} \psi_k(s) = +1 \quad \text{and} \quad \lim_{s\to-\infty} \psi_k(s) = -1. \tag{2}$$

As the $\alpha_k > 0$, defined for each edge of the interaction graph $\mathcal{G}$, scales the normalized nonlinearity $\psi_k(\cdot)$, we refer to it as the *edge capacity*.

The important features distinguishing our model from general diffusively coupled networks are: 1) each node, if uncoupled from the network, will converge to a unique equilibrium and 2) the coupling functions $\psi_k(\cdot)$ are saturated. We show that these two properties will cause the dynamical network to cluster and we provide a novel explanation of the *mechanisms leading to this clustering phenomenon*.

As a motivating preview of the clustering behavior exhibited by the model (1) with Assumption 3.1, Fig. 1 shows the trajectories of a network with 30 nodes. We will discuss these simulations in more detail in Section VII, and introduce them here only to illustrate the observed clustering behavior. For each of the three simulations, we use identical capacities for all the edges in the graph $\alpha_k = \alpha_{[i]}, k \in \{1, \ldots, |\mathbf{E}|\}$, $i \in \{1, 2, 3\}$. The three simulation scenarios differ only in the choice of the edge capacities (i.e., $\alpha_{[1]} > \alpha_{[2]} > \alpha_{[3]}$). For a sufficiently large edge capacity $\alpha_{[1]}$, Fig. 1(a) shows that all nodes reach agreement on their "velocities" $\dot{\chi}$ (e.g., they form an exact 1-cluster). When reducing the edge capacity to $\alpha_{[2]}$, as shown in Fig. 1(b), we observe that the network forms an exact 3-cluster, comprised of two large clusters and a single isolated node. For a smaller edge capacity $\alpha_{[3]}$, the network partitions into a large number of clusters, as shown in Fig. 1(c).

This example suggests that the clustering behavior depends on three parameters of the system: 1) the interaction graph $\mathcal{G}$, 2) the local functions $f_i(\cdot)$ of each agent, and 3) the edge capacities α_k. We explain how the relation between these three properties leads to clustering. The dynamics (1), with Assumption 3.1, is a model system that explains clustering in diffusively coupled networks. A similar attempt has been made in [11] and [12], where first-order dynamics with saturated interactions are used to study the fundamental principles of clustering.

"Node-Edge" Model Representation: Although the dynamical network (1) is stated in a standard form, it is advantageous to represent it in an alternative form, which is inspired by [5].

Lemma 3.2: The model (1) with Assumption 3.1 is equivalent to the following second-order model:

$$\dot{x}(t) = -\nabla \mathbf{F}(x(t)) - E(\mathcal{G}) W \boldsymbol{\psi}(z(t)) \tag{3}$$

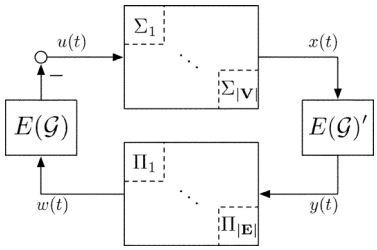


Fig. 2. Block diagram of the dynamical network model with node dynamics Σ_i and edge dynamics Π_k.

$$\dot{z}(t) = E(\mathcal{G})'x(t) \tag{4}$$

where $x(t) := \dot{\chi}(t)$, $z(t) := E(\mathcal{G})'\chi(t)$, $f_i(\cdot) = \nabla F_i(\cdot)$ $\boldsymbol{\nabla}\mathbf{F}(x(t)) = [\nabla F_1(x_1(t)), \ldots, \nabla F_{|\mathbf{V}|}(x_{|\mathbf{V}|}(t))]'$, $\boldsymbol{\psi}(z) = [\psi_1(z_1(t)), \ldots, \psi_{|\mathbf{E}|}(z_{|\mathbf{E}|}(t))]'$, and the diagonal matrix $W = \mathrm{diag}([\alpha_1, \ldots, \alpha_{|\mathbf{E}|}])$.

Proof: Each node in the network is associated with a node state, i.e., the "velocity" $x_i := \dot{\chi}_i$. The node states evolve according to

$$\Sigma_i : \dot{x}_i(t) = -\nabla F_i\left(x_i(t)\right) + u_i(t) \tag{5}$$

where $u_i(t) \in \mathbb{R}$ represents an external input due to the coupling effect of neighboring nodes. The relative velocity of neighboring nodes in the graph $\mathcal{G}$ can be captured in a new variable $y \in \mathbb{R}^{|\mathbf{E}|}$, defined as $y(t) = E(\mathcal{G})'x(t)$. Due to the diffusive coupling in (1), only the relative "position" of neighboring nodes is of importance. Therefore, we associate with each edge $e_k \in \mathbf{E}$ the *edge state* $z_k(t) \in \mathbb{R}$ that represents the relative "position," i.e., $z_k = \chi_j - \chi_i$, $\forall (j,i) \in \mathbf{E}$. The evolution of the relative positions and their influence on the nodes is given by the following dynamical system:

$$\Pi_k : \dot{z}_k(t) = y_k(t), \quad w_k(t) = \alpha_k \psi_k\left(z_k(t)\right). \tag{6}$$

The output $w_k(t)$ represents the influence of one edge to its incident nodes. The overall influence of (6) on the node dynamics (5) can be stated with the incidence matrix as

$$u(t) = -E(\mathcal{G})w(t). \tag{7}$$

Combining (5)–(7) yields the closed-loop dynamics given in Lemma 3.2. ∎

The block structure of the model (3) is illustrated in Fig. 2. The node-edge representation of the model given in Lemma 3.2, together with Assumption 3.1, is a key feature that will facilitate the analysis of the clustering phenomena. This is due to the surprising connection between this model and a *static saddle-point problem*. In the sequel, we explore this connection and use it to identify how the local dynamics of each agent, the underlying connection graph, and the edge capacities lead to clustering.

IV. Saddle-Point Problems and Static Network Clustering

The clustering phenomena in networks can be described by a static optimization problem that is intimately connected to the dynamical network (3). In particular, we examine the following static *max-min problem*, referred to as a *saddle-point problem*:

$$\max_{\mu \in \Gamma} \min_{\mathrm{x}} L(\mathrm{x}, \mu) = \sum_{i=1}^{|\mathbf{V}|} F_i(\mathrm{x}_i) + \mu' E(\mathcal{G})'\mathrm{x} \tag{8}$$

where $\mathrm{x} = [\mathrm{x}_1, \ldots, \mathrm{x}_{|\mathbf{V}|}]' \in \mathbb{R}^{|\mathbf{V}|}$ are decision variables associated with each node in the graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, and $\mu = [\mu_1, \ldots, \mu_{|\mathbf{E}|}]' \in \mathbb{R}^{|\mathbf{E}|}$ are variables associated with the edges in $\mathbf{E}$. The objective functions $F_i(\mathrm{x}_i)$ are the integral functions of $f_i(\cdot) := \nabla F_i(\cdot)$ appearing in the dynamics (5). We will sometimes abbreviate the notation writing $F(\mathrm{x}) := \sum_{i=1}^{|\mathbf{V}|} F_i(\mathrm{x}_i)$. Note that the vector x here is a static vector while the state vector $x(t)$ of the system (3) is dynamic.

The *constraint set* $\Gamma = \Gamma_1 \times \cdots \times \Gamma_{|\mathbf{E}|}$, with $\Gamma_k = [-\alpha_k, \alpha_k]$ is a box constraint. Throughout this paper, except if explicitly stated otherwise, we assume that $0 < \alpha_k < \infty$ and, thus, Γ is a compact and convex set.

A special instance of the problem (8) occurs when the variables μ_k are *unconstrained*, corresponding to $\Gamma = \mathbb{R}^{|\mathbf{E}|}$. In this case, (8) corresponds to the *Lagrange-dual problem* of a network optimization problem of the form

$$\min_{\mathrm{x}} \sum_{i=1}^{|\mathbf{V}|} F_i(\mathrm{x}_i) \quad \text{s.t.} \quad E(\mathcal{G})'\mathrm{x} = 0. \tag{9}$$

We denote (x^*, μ^*) as the solution to the problem (9), where μ^* are the dual variables of (9) associated with the constraints $E(\mathcal{G})'\mathrm{x} = 0$. The optimal primal solution x^* of (9) will always form a 1-cluster in agreement (i.e., $E(\mathcal{G})'\mathrm{x}^* = 0 \Rightarrow \mathrm{x}^* = \beta\mathbb{1}$), for some $\beta \in \mathbb{R}$. An optimal dual solution can then be obtained by the first-order optimality condition $\boldsymbol{\nabla}\mathbf{F}(\mathrm{x}^*) + E(\mathcal{G})\mu^* = 0$. This problem falls under a broad class of network optimization problems and, consequently, can be solved efficiently using a variety of methods. (See [20] for example.)

There is a game-theoretic interpretation of the saddle-point problem. A decision maker in each node v_i aims to minimize its individual objective function F_i; simultaneously, another decision maker, attached to an edge, penalizes any deviation between the decision variables of its incident nodes. For the problem (9), the dual variables associated with the constraints will force the decision makers on the nodes to reach an exact agreement on their values. However, the saddle-point problem formulation (8) does not allow the edge decision makers to arbitrarily penalize the deviation between neighboring agents (e.g., the penalty variables μ are restricted to be contained in the set Γ). This additional constraint has a strong impact on the structure of the primal solution x.

Saddle Points: Here, we provide some properties of the saddle points associated with (8).

Definition 4.1: A point $(\bar{\mathrm{x}}, \bar{\mu})$ is a *saddle point* of (8) if

$$L(\bar{\mathrm{x}}, \mu) \leq L(\bar{\mathrm{x}}, \bar{\mu}) \leq L(\mathrm{x}, \bar{\mu}), \quad \text{for all} \quad \mathrm{x} \in \mathbb{R}^n, \quad \mu \in \Gamma.$$

We will denote the *set of all saddle points* in the following text by $\mathbb{S} = \mathbb{X} \times \mathbb{M}$.

Lemma 4.2: The set of all saddle points $\mathbb{S}$ for (8) is nonempty.

Proof: The set Γ is nonempty, convex, and compact. The function $L(\mathrm{x}, \mu)$ is convex for each fixed $\mu \in \Gamma$, and concave for each $\mathrm{x} \in \mathbb{R}^n$. Furthermore, for some $\bar{\mu} \in \Gamma$ and $\beta \in \mathbb{R}$, the

level sets are nonempty and compact since each F_i is a strongly convex function. The statement follows from the *Saddle Point Theorem* [21, Prop. 4.7]. ■

Saddle points also admit some first-order optimality conditions. Let $(\bar{\mathrm{x}}, \bar{\mu})$ be a saddle point, then [22]

$$\boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}}) + E(\mathcal{G})\bar{\mu} = 0, \quad \text{and} \quad \bar{\mathrm{x}}' E(\mathcal{G})(\mu - \bar{\mu}) \leq 0, \quad \forall \mu \in \Gamma. \tag{10}$$

Lemma 4.3: Let $(\bar{\mathrm{x}}, \bar{\mu})$ be a saddle point of (8), then $\mathbb{X} = \{\bar{\mathrm{x}}\}$ and $\mathbb{M} = \{\mu \in \Gamma | \mu = \bar{\mu} + \nu, \nu \in \mathcal{N}(E(\mathcal{G}))\}$.

Proof: We first show the uniqueness of the saddle point in the x-coordinate. Suppose $(\bar{\mathrm{x}}, \bar{\mu})$ and $(\tilde{\mathrm{x}}, \tilde{\mu})$ are saddle points with $\bar{\mathrm{x}} \neq \tilde{\mathrm{x}}$. Then, $L(\bar{\mathrm{x}}, \bar{\mu}) = L(\tilde{\mathrm{x}}, \tilde{\mu})$ and, furthermore, $F(\bar{\mathrm{x}}) - F(\tilde{\mathrm{x}}) = \tilde{\mu}' E(\mathcal{G})' \tilde{\mathrm{x}} - \bar{\mu}' E(\mathcal{G})' \bar{\mathrm{x}}$. The first-order optimality conditions state that $\tilde{\mu}' E(\mathcal{G})' \tilde{\mathrm{x}} \geq \bar{\mu}' E(\mathcal{G})' \tilde{\mathrm{x}}$ and $E(\mathcal{G})\bar{\mu} = -\boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}})$, implying that

$$\begin{aligned} F(\bar{\mathrm{x}}) - F(\tilde{\mathrm{x}}) &= \tilde{\mu}' E(\mathcal{G})' \tilde{\mathrm{x}} - \bar{\mu}' E(\mathcal{G})' \bar{\mathrm{x}} \\ &\geq \bar{\mu} E(\mathcal{G})' \tilde{\mathrm{x}} - \bar{\mu}' E(\mathcal{G})' \bar{\mathrm{x}} = -\boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}})(\tilde{\mathrm{x}} - \bar{\mathrm{x}}). \end{aligned}$$

On the other hand, due to the strong convexity of F_i, we have

$$F(\tilde{\mathrm{x}}) - F(\bar{\mathrm{x}}) \geq \boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}})'(\tilde{\mathrm{x}} - \bar{\mathrm{x}}) + \frac{\eta}{2}\|\tilde{\mathrm{x}} - \bar{\mathrm{x}}\|^2. \quad \eta > 0.$$

This is a contradiction, proving that $\bar{\mathrm{x}} = \tilde{\mathrm{x}}$ and, thus, $\mathbb{X} = \{\bar{\mathrm{x}}\}$. The nonuniqueness of $\bar{\mu}$ remains to be shown. If $\mathcal{N}(E(\mathcal{G}))$ is nontrivial, then for any vector $\nu \in \mathcal{N}(E(\mathcal{G}))$, one has $L(\bar{\mathrm{x}}, \bar{\mu}) = L(\bar{\mathrm{x}}, \bar{\mu} + \nu)$. Any vector $\mu = \bar{\mu} + \nu \in \Gamma$ satisfies the saddle-point and first-order optimality conditions. ■

The result states that there is a unique vector $\bar{\mathrm{x}}$ at which a saddle point can be attained, whereas μ can be in a set $\mathbb{M}$. However, the set of all saddle points depends on the structure of the graph and, in particular, its flow space.

Lemma 4.4: For $\Gamma = \mathbb{R}^{|\mathbf{E}|}$, the set $\mathbb{M}$ contains more than one point if and only if $\mathcal{G}$ contains at least one cycle.

Proof: From Theorem 2.2, the flow space of $E(\mathcal{G})$ is nontrivial if and only if $\mathcal{G}$ contains at least one cycle. ■.

Observe that for $\Gamma = \mathbb{R}^{|\mathbf{E}|}$, the problem (8) is equivalent to the Lagrange dual for (9). The additional restrictions on Γ subsequently change the structure of $\mathbb{M}$. Note that there are a variety of computational methods readily available for saddle-point problems [23]. Furthermore, the problem proposed in (8) enables computationally tractable solutions by observing that the maximization component is over the Lagrange-dual function of a strictly convex program (9) with a box constraint on the dual variables.

Network Clustering: Having established the existence and uniqueness properties of the saddle points for (8), we now show how these solutions lead to clusters in the graph $\mathcal{G}$. First, we introduce the notion of a *saturated edge* in the graph.

Definition 4.5: An edge $e_k \in \mathbf{E}$ is said to be *saturated* if for all $\bar{\mu} \in \mathbb{M}$, $\bar{\mu}_k \in \partial\Gamma_k$ (e.g., $|\bar{\mu}_k| = \alpha_k$).

Note that, in general, $\bar{\mu}_k \in \partial\Gamma_k$ for a particular $\bar{\mu}$ does not imply that the edge is saturated. For an edge to be saturated, the constraint associated with that edge must be active for *all* possible saddle points in the set $\mathbb{M}$. The following lemma connects the definition of saturated edges to graph properties.

Lemma 4.6: Any cycle in $\mathcal{G}$ contains either zero or at least two saturated edges.

Proof: Assume by contradiction that edge e_k is the *only* saturated edge contained in a cycle with a corresponding signed path vector ζ. Then, $\zeta_k \neq 0$ and from Theorem 2.2, $\zeta \in \mathcal{N}(E(\mathcal{G}))$. From Lemma 4.3, a $\delta \in \mathbb{R}$ exists that is sufficiently small such that $\tilde{\mu} = \bar{\mu} + \delta\zeta \in \mathbb{M}$ and $\tilde{\mu}_k \in \mathrm{int}\Gamma_k$. But this contradicts the definition of a saturated edge. Therefore, e_k *cannot* be saturated. Thus, if a cycle contains a saturated edge, it must contain at least two saturated edges. ■

We now show that if the set $\mathbb{M}$ contains saturated edges, then there is a corresponding cutset for the graph comprised of those edges.

Lemma 4.7: The set of saturated edges in $\mathbb{M}$ forms a cutset for the graph.

Proof: First, assume that an edge e_k in the graph is saturated and is *not* contained in any cycle in $\mathcal{G}$. Then, $\overline{\mu}_k \in \partial\Gamma_k$ and its deletion from the graph must result in an increase in the number of components, thus forming a cutset. Now assume that a saturated edge e_k is contained in one or more cycles. Then, by Lemma 4.6, any cycle contains at least one other saturated edge. The deletion of two or more edges from a cycle results in an increase in the number of components in the graph and, thus, each saturated edge in a cycle is included in a cutset. ■

Lemma 4.7 makes a strong connection between the saddle points of (8), saturated edges, and cutsets. We are now able to state the main result of this section, relating clustering to saddle points.

Theorem 4.8 (Saddle-Point Clustering): Let $\mathbb{S} = \{\bar{\mathrm{x}}\} \times \mathbb{M}$ be the saddle points of (8), and let $\mathbf{Q} \subseteq \mathbf{E}$ be the set of saturated edges. Then, $\mathbf{Q}$ induces a p-partition $\mathbb{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_p\}$ and each cluster $\mathcal{P}_i$ induced by the set $\mathbf{P}_i$ is in exact agreement.

Proof: Let $(\bar{\mathrm{x}}, \bar{\mu}) \in \mathbb{S}$ be a saddle point with $\bar{\mu}_k \in \mathrm{int}\Gamma_k$ for all nonsaturated edges. Note that stating cluster $\mathcal{P}_i$ is in agreement and is equivalent to $E(\mathcal{P}_i)'\bar{\mathrm{x}}(\mathcal{P}_i) = 0$. Assume that in order to arrive at a contradiction, some $P_i \in \mathbb{P}$ exists such that $E(\mathcal{P}_i)'\bar{\mathrm{x}}(\mathcal{P}_i) \neq 0$. Denote $\mathcal{Q}$ as the subgraph induced by $\mathbf{Q}$. Function (8) can then be written as

$$\begin{aligned} L(\bar{\mathrm{x}}, \bar{\mu}) = \sum_{i=1}^{|\mathbf{V}|} F_i(\bar{\mathrm{x}}_i) + \sum_{j=1}^{p} \bar{\mu}(\mathcal{P}_j)' E(\mathcal{P}_j)' \bar{\mathrm{x}}(\mathcal{P}_j) \\ + \bar{\mu}(\mathcal{Q})' E(\mathcal{Q})' \bar{\mathrm{x}}(\mathcal{Q}). \end{aligned} \tag{11}$$

Since all clusters except $\mathcal{P}_i$ are assumed to be in agreement, the second summand of (11) can be written as $\sum_{e_i=(v_k,v_l)\in\mathcal{P}_i} \bar{\mu}_i(\bar{\mathrm{x}}_k - \bar{\mathrm{x}}_l)$. Assume without loss of generality that only the edge $e_k = (v_i, v_j)$ in $\mathcal{P}_i$ connects two nodes that are not in agreement with a positive difference (e.g., $\bar{\mathrm{x}}_i - \bar{\mathrm{x}}_j > 0$). Then, $\epsilon > 0$ exists such that $\bar{\mu}_k + \epsilon \in \Gamma_k$ and $\bar{\mu}_k(\bar{\mathrm{x}}_i - \bar{\mathrm{x}}_j) < (\bar{\mu}_k + \epsilon)(\bar{\mathrm{x}}_i - \bar{\mathrm{x}}_j)$. Let $\tilde{\mu}$ be the edge value after adding ϵ to only edge value $\bar{\mu}_k$ as described before. Then, $L(\bar{\mathrm{x}}, \bar{\mu}) \leq L(\bar{\mathrm{x}}, \tilde{\mu})$ contradicts the assumption that $(\bar{\mathrm{x}}, \bar{\mu})$ is a saddle point. Therefore, each cluster $\mathcal{P}_i$ must be in agreement. ■

With this theorem, we can directly draw a conclusion about the clustering structure of the network. The network will partition along the saturated edges contained in the saddle points of (8). Note that Theorem 4.8 can be used to express any saddle point $\bar{\mathrm{x}}$ in the form $x(\mathcal{P}_i) = \beta_i \mathbb{1}, i \in \{1, \ldots, p\}$, for some $\beta_i \in \mathbb{R}$. The clusters $\mathcal{P}_i$ are the connected components of the graph $\mathcal{G}$ after deleting all of the saturated edges.

Analysis of the Network Clustering: We are now prepared to characterize the network clustering. We first provide a necessary and sufficient condition for the network to fully synchronize. In the following equation, denote $\mathbb{S} = \{\bar{\mathrm{x}}\} \times \mathbb{M}$ again as the set

of saddle points for (8) and $\{\mathrm{x}^*\} \times \mathbb{M}^*$ as the primal and dual optimal solution sets to the network optimization problem (9).

Theorem 4.9 (Synchronization Condition): The solution $\bar{\mathrm{x}} \in \mathbb{X}$ of the saddle-point problem (8) forms a 1-cluster in agreement if and only if $\Gamma \cap \mathbb{M}^* \neq \emptyset$.

Proof: If $\Gamma \cap \mathbb{M}^* \neq \emptyset$, a $\mu^* \in \mathbb{M}^*$ exists which is also contained in Γ. Since μ^* is the optimal dual solution to (9), we know that the corresponding optimal primal solution is $\mathrm{x}^* = \beta \mathbb{1}$. This solution (x^*, μ^*) satisfies the first-order optimality conditions for (8) $\boldsymbol{\nabla}\mathbf{F}(\beta\mathbb{1}) + E(\mathcal{G})\mu^* = 0$ and $\beta\mathbb{1}'E(\mathcal{G})(\mu - \bar{\mu}) = 0\ \forall \mu \in \Gamma$. Thus, (x^*, μ^*) is also a saddle point of (8) and, therefore, $\bar{\mathrm{x}} = \beta\mathbb{1}$. Now suppose that $\bar{\mathrm{x}} = \beta\mathbb{1}$. There exists $\mu \in \Gamma$ such that $\boldsymbol{\nabla}\mathbf{F}(\beta\mathbb{1}) + E(\mathcal{G})\bar{\mu} = 0$. Since $\bar{\mathrm{x}}$ also satisfies the second condition of optimality for (9), $E(\mathcal{G})'\bar{\mathrm{x}} = \beta E(\mathcal{G})'\mathbb{1} = 0$. Thus, $(\bar{x}, \bar{\mu})$ is also a saddle point of (9) and, therefore, $\bar{\mu} \in \mathbb{M}^*$. This shows that $\Gamma \cap \mathbb{M}^* \neq \emptyset$ and concludes the proof. ∎

The importance of Theorem 4.9 is that it provides a condition for achieving a 1-cluster for (3) in terms of the solution of an unconstrained network optimization problem (9). Considering that there are many efficient algorithms for solving (9), checking whether the saddle-point problem achieves a 1-cluster is equivalent to solving the unconstrained problem.

We can now also characterize the agreement values of the clusters. Assume that the network forms a p-cluster $(\mathbf{P}_1, \ldots, \mathbf{P}_p)$. We have already shown that $\bar{\mathrm{x}}(\mathcal{P}_i) = \beta_i \mathbb{1}$. For notational simplicity, define $F_{\mathcal{P}_i}(\beta_i) = \sum_{j \in \mathcal{P}_i} F_j(\beta_i)$ and $\nabla F_{\mathcal{P}_i}(\beta_i) = \sum_{j \in \mathcal{P}_i} \nabla F_j(\beta_i)$. The notation is introduced to prepare an alternative representation of the first-order optimality condition (10) in terms of the network partitions. Recall also that each edge in the cutset $\mathbf{Q}$ that induces the p-cluster is saturated; each component of the vector $\bar{\mu}(\mathcal{Q})$ is either $+\alpha_k$ or $-\alpha_k$.

Consider the incidence matrix induced by the cutset $E(\mathcal{Q})$ and note that each row corresponds to a node of the graph. We can re-sort the rows so that those corresponding to nodes in the same partition are grouped together, and express the incidence matrix as $E(\mathcal{Q}) = [E_1(\mathcal{Q})' \cdots E_p(\mathcal{Q})']'$. where $E_i(\mathcal{Q})$ is a matrix formed by all rows of the matrix $E(\mathcal{Q})$ corresponding to nodes in $\mathbf{P}_i$.

Now we define the vectors $y_{\mathcal{P}_i} = E_i(\mathcal{Q})\bar{\mu}(\mathcal{Q}), i \in \{1, \ldots, p\}$. Note that for all edges in $\mathbf{Q}$ that are not incident to any nodes in $\mathbf{P}_i$, the corresponding value of $y_{\mathcal{P}_i}$ is 0. Otherwise, the corresponding value of $y_{\mathcal{P}_i}$ is $\pm\alpha_k$ for an edge k that is in the cutset and incident to nodes in $\mathbf{P}_i$.

Lemma 4.10: The agreement value for the cluster $\mathcal{P}_i$ is given by $\beta_i = \nabla F^*_{\mathcal{P}_i}(-y'_{\mathcal{P}_i}\mathbb{1})$, where $F^*_{\mathcal{P}_i}$ is the convex conjugate of the function $F_{\mathcal{P}_i}$.

Proof: We can rewrite the standard conditions of optimality using the cluster notation as

$$\nabla F_{\mathcal{P}_i}(\beta_i) + \bar{\mu}(\mathcal{P}_i)'E(\mathcal{P}_i)\mathbb{1} + y'_{\mathcal{P}_i}\mathbb{1} = 0, \quad i \in \{1, \ldots, p\}.$$

Note that the previous equation can be derived by taking the sum of all first-order optimality conditions corresponding to the nodes in the cluster $\mathcal{P}_i$. Considering now that $E(\mathcal{P}_i)'\mathbb{1} = 0$ and that the gradient of the convex conjugate function is the inverse of the gradient of the function ([24]), the statement follows directly. ∎

This result highlights that the cluster agreement value β_i depends only on the objective functions of the cluster nodes $F_{\mathcal{P}_i}$ and the edge capacities α_k of the edges separating the cluster from neighboring clusters. It is independent of the distribution of the dual variables $\bar{\mu}$ or the edge capacities α_k within a cluster.

V. Dynamic Network Clustering and Saddle-Point Problems

The results of Section IV provided a rigorous analysis of clustering within the framework of a static optimization problem. The main purpose of this paper is, however, to explain the clustering behavior observed in the dynamical network (1). A contribution of this work is to show how the static analysis can be related to the asymptotic behavior of the dynamical network. The connection between the static saddle-point problem and the dynamical network can be established considering the network in the node-edge representation (3). Recall, for the connection to the original model representation (1) that $x(t) = \dot{\chi}(t)$ as well as $w(t) = W\boldsymbol{\psi}(E(\mathcal{G})'\chi(t))$.

Theorem 5.1 (Convergence): Let the network (3) be defined according to Assumption 3.1 and let $\mathbb{S} = \{\bar{\mathrm{x}}\} \times \mathbb{M}$ be the set of saddle points of problem (8). Then, the trajectories $x(t)$ of (3) and $w(t)$ of (6) remain bounded and

$$\lim_{t\to\infty} x(t) \to \bar{\mathrm{x}}, \quad \lim_{t\to\infty} w(t) \to \mathbb{M}.$$

Proof: We assume throughout the proof without loss of generality that the vector $\bar{\mathrm{x}}$ corresponds to a p-partition in exact agreement of the network. The set $\{(x(t), w(t)) : x(t) = \bar{\mathrm{x}}, w(t) \in \mathbb{M}\}$ is invariant under the dynamics (3) as the following discussion shows. Within the set, the dynamics (3) satisfy

$$\dot{x}(t) = -\boldsymbol{\nabla}\mathbf{F}\left(x(t)\right) - E(\mathcal{G})w(t)|_{x(t)=\bar{\mathrm{x}}, w(t)\in\mathbb{M}} = 0$$

and, therefore, as a result of Lemma 4.3, $x(t) \equiv \bar{\mathrm{x}}$. In addition, for $x(t) \equiv \bar{\mathrm{x}}$, we have the dynamics (4) as $\dot{z}(t) = E(\mathcal{G})'\bar{\mathrm{x}}$. Contrary to the dynamics of (3), a saddle point $x(t) = \bar{\mathrm{x}}$ is, in general, not an equilibrium point for (4), since the vector $\bar{\mathrm{x}}$ corresponds to some p-partition, $\dot{z}_k = 0$ for all edges e_k connecting nodes within the same partition. Meanwhile, for edges e_k connecting two nodes, say v_i and v_j, in different partitions, $\dot{z}_k = \bar{\mathrm{x}}_i - \bar{\mathrm{x}}_j \neq 0$. Suppose now, without loss of generality, that $\bar{\mathrm{x}}_i - \bar{\mathrm{x}}_j > 0$. Then, the dynamics (4) are such that $\lim_{t\to\infty} \alpha_k\psi_k(z_k(t)) = \alpha_k$. Since by the first-order optimality conditions (10) in this situation $\bar{\mu}_k = \alpha_k > 0$, for all $\mu \in \mathbb{M}$, we can conclude the invariance of the desired set.

We now show that the set $\{(x(t), w(t)) : x(t) = \bar{\mathrm{x}}, w(t) \in \mathbb{M}\}$ is reachable from the dynamics (3). As the nonlinear functions $\psi(\cdot)$ need not attain their limit, the points on the boundary of Γ can, at best, be approached asymptotically. Note also that due to the structure of the dynamics (4) the state $z(t)$ is constrained to be such that $z(t) \in \mathcal{R}(E(\mathcal{G})')$ for all times. The following proposition shows that there exists at least one trajectory of $w(t)$ that asymptotically approaches a point in $\mathbb{M}$.

Proposition 5.2: Sequences $\{z^{[k]}\}_{k=1}^{\infty}$ and $\{w^{[k]}\}_{k=1}^{\infty}$ exist such that $z^{[k]} \in \mathcal{R}(E(\mathcal{G})')$ and $w^{[k]} := W\boldsymbol{\psi}(z^{[k]}) \to \mathbb{M}$ as $k \to \infty$.

The proof of the proposition is presented in the Appendix. The intuition behind the proof is that the nonlinear map $\psi(\cdot)$ is a homeomorphism, (i.e., $\mathbb{R}^{|\mathbf{E}|}$ is homeomorphic to $\mathrm{int}\Gamma$) [20]. Let the following $\bar{z}$ be the limit point of a sequence $\{z^{[k]}\}$ as considered in Proposition 5.2, i.e., such that $z^{[k]} \to \bar{z}$ implies

that $W\boldsymbol{\psi}(z^{[k]}) \to \mathbb{M}$ and let $\bar{\mu} \in \mathbb{M}$ be the convergence point of $W\boldsymbol{\psi}(z^{[k]})$.

To show convergence and boundedness of the trajectories of (3) to this set, we use a storage function $V(x,z)$ which decreases along the trajectories of (3)

$$V(x,z) = \frac{1}{2}\left\|x(t) - \bar{\mathrm{x}}\right\|^2 + \sum_{k=1}^{|\mathbf{E}|} \int_{\bar{z}_k}^{z_k(t)} (\alpha_k \psi_k(s) - \bar{\mu}_k)\, ds. \tag{12}$$

Clearly, $(1/2)\|x(t) - \bar{\mathrm{x}}\|^2$ is strictly positive for any $x \neq \bar{\mathrm{x}}$. The second part of (12) vanishes for $z = \bar{\mathrm{z}}$. To see this, define $h_k(z_k) := \int_{\bar{z}_k}^{z_k(t)} (\alpha_k \psi_k(s) - \bar{\mu}_k) ds$ and note that $h(\bar{\mathrm{z}}_k) = 0$. Also, the gradient of this function vanishes at the point $\bar{\mathrm{z}}_k$ (i.e., $(\partial h_k/\partial z_k)|_{z_k=\bar{\mathrm{z}}_k} = (\alpha_k \psi_k(\bar{\mathrm{z}}_k) - \bar{\mu}_k) = 0$). One can see now that the function $V(x,z)$ is positive (semi) definite since the second derivative is everywhere non-negative $\partial^2 h_k/\partial z_k^2 = \partial \psi_k/\partial z_k \geq 0$ due to the monotonic property of $\psi_k(z_k)$. We conclude that $V(x,z)$ is a positive semidefinite function and is a suitable storage function candidate. We consider the directional derivative of V along the trajectories

$$\begin{aligned}\dot{V} &= (x - \bar{\mathrm{x}})' \dot{x} + \sum_{k=1}^{|\mathbf{E}|} (\alpha_k \psi_k(z_k) - \bar{\mu}_k)\, \dot{z} \\ &= (x - \bar{\mathrm{x}})' \left(\boldsymbol{\nabla}\mathbf{F}(x) - E(\mathcal{G}) W \boldsymbol{\psi}(z)\right) \\ &\quad + \left(W\boldsymbol{\psi}(z) - \bar{\mu}\right)' E(\mathcal{G})' x. \end{aligned} \tag{13}$$

We now add zero to obtain

$$\begin{aligned}\dot{V} = (x - \bar{\mathrm{x}})' \left(-\boldsymbol{\nabla}\mathbf{F}(x) + \boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}}) - \boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}}) - E(\mathcal{G})' W \boldsymbol{\psi}(z)\right) \\ + \left(W\boldsymbol{\psi}(z) - \bar{\mu}\right)' E(\mathcal{G})' x.\end{aligned}$$

Using the first-order condition of optimality $\boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}}) = -E(\mathcal{G})\bar{\mu}$, we can write this as

$$\begin{aligned}\dot{V} = -(x - \bar{\mathrm{x}})' \left(\boldsymbol{\nabla}\mathbf{F}(x) - \boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}})\right) + \left(W\boldsymbol{\psi}(z) - \bar{\mu}\right)' E(\mathcal{G})' x \\ + (x - \bar{\mathrm{x}})' \left(E(\mathcal{G})\bar{\mu} - E(\mathcal{G}) W \boldsymbol{\psi}(z)\right).\end{aligned}$$

Due to strong convexity, $(x - \bar{\mathrm{x}})'(\boldsymbol{\nabla}\mathbf{F}(x) - \boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}})) \geq \eta(x - \bar{\mathrm{x}})'(x - \bar{\mathrm{x}})$, one obtains

$$\begin{aligned}\dot{V} \leq -\eta(x - \bar{\mathrm{x}})'(x - \bar{\mathrm{x}}) + (x - \bar{\mathrm{x}})' E(\mathcal{G}) \left(\bar{\mu} - W\boldsymbol{\psi}(z)\right) \\ + \left(W\boldsymbol{\psi}(z) - \bar{\mu}\right)' E(\mathcal{G})' x\end{aligned}$$

and, thus, $\dot{V} \leq -\eta(x - \bar{\mathrm{x}})'(x - \bar{\mathrm{x}}) - \bar{\mathrm{x}}' E(\mathcal{G})(\bar{\mu} - W\boldsymbol{\psi}(z))$.

Note that $\bar{\mathrm{x}}' E(\mathcal{G})$ is a vector with one entry for each edge of $\mathcal{G}$. The entry $[\bar{\mathrm{x}}' E(\mathcal{G})]_k$ of this vector is nonzero if and only if the edge e_k is saturated ($|\bar{\mu}| = \alpha_k$). Thus, we can write

$$\dot{V} \leq -\bar{\mathrm{x}}(\mathcal{Q}') E(\mathcal{Q}) \left(\bar{\mu}(\mathcal{Q}) - W(\mathcal{Q})\boldsymbol{\psi}(\mathcal{Q})\right) \tag{14}$$

with all entries of $\bar{\mu}(\mathcal{Q})$ having the value "$+\alpha_k$" or "$-\alpha_k$."

It remains to be shown that the right-hand side of (14) is nonpositive. The entries of the vector $(\bar{\mu}(\mathcal{Q}) - W(\mathcal{Q})\boldsymbol{\psi}(\mathcal{Q}))$ have the same sign as the corresponding entries of the vector $\bar{\mu}(\mathcal{Q})$. In particular, if $\bar{\mu}_k = +\alpha_k$ ($\bar{\mu}_k = -\alpha_k$) then $\bar{\mu}_k - \alpha_k \psi_k(z_k) > 0$ ($\bar{\mu}_k - \alpha_k \psi_k(z_k) < 0$) for all z_k. Note that the entries of the vector $E(\mathcal{Q})'\bar{\mathrm{x}}(\mathcal{Q})$ have the same sign as $\bar{\mu}(\mathcal{Q})$. This is a condition of optimality. If $\bar{\mu}_k [E(\mathcal{G})'\bar{\mathrm{x}}]_k \not\geq 0$, the solution $\bar{\mu}$ cannot be optimal since a simple change in the sign would increase the value of $L(\bar{\mathrm{x}}, \bar{\mu})$. Taking these observations into account, we conclude that each component $[\bar{\mathrm{x}}(\mathcal{Q})' E(\mathcal{Q})]_k [(\bar{\mu}(\mathcal{Q}) - W\boldsymbol{\psi}(\mathcal{Q}))]_k \geq 0$ for all z and consequently

$$\dot{V} \leq -\eta(x - \bar{\mathrm{x}})'(x - \bar{\mathrm{x}}). \tag{15}$$

We conclude with Barbalat's Lemma that all trajectories $x(t)$ remain bounded converge to the unique saddle point $\bar{\mathrm{x}}$ [25]. The largest invariant set contained in $\{x : x \equiv \bar{\mathrm{x}}\}$ is, in fact, $\{(x,z) : x = \bar{\mathrm{x}}\ W\boldsymbol{\psi}(z) = \bar{\mu} + \nu, \nu \in \mathcal{N}(E)\}$ and all trajectories will converge to this set. ■

The previous proof establishes the connection between the dynamical network (3) and the static saddle-point problem (8). It seems remarkable that the clustering in the saddle-point problem was caused by the fact that the Lagrange multipliers were constrained and could not compensate for the imbalance between the node objectives. The same mechanism leads to clustering in the dynamical network (3). The influence of the z-dynamics (4), which is driven by the state deviation of neighboring nodes, only has a bounded influence on the node dynamics (3) and can therefore not fully compensate for the inhomogeneous node dynamics. The connection between (3) and (8) provides, therefore, a novel interpretation along with a profound understanding of a mechanism leading to dynamical clustering.

VI. Hierarchical Analysis and Combinatorial Aspects of Clustering

We will in the following exploit the tight connection to the saddle-point problem to reveal how the clustering structure of our model is related to an *optimal graph partitioning*. For the clarity of the results, we focus the analysis on the particular situation where all edges in the network have the same capacity (i.e., $\alpha_k = \alpha$, for $k \in \{1, \ldots, |\mathbf{E}|\}$). When considering only identical edges, the edge capacity becomes a single parameter for the network behavior. We first observe the following result.

Lemma 6.1: For any two scalars $\alpha_1 > \alpha_2$, let $\mathbb{P}_{\alpha_1}$ and $\mathbb{P}_{\alpha_2}$ be the partitions induced by the solution of the saddle-point problem (8) with $\Gamma_{\alpha_i} = \{\mu \in \mathbb{R}^{|\mathbf{E}|} : |\mu_k| \leq \alpha_i\}$. Then, it holds that $\mathbb{P}_{\alpha_2} \succeq \mathbb{P}_{\alpha_1}$.

Proof: We first note that $\Gamma_{\alpha_1} \supset \Gamma_{\alpha_2}$. Let $\mathbb{M}_i$ denote the set of saddle-point solutions to (8) computed with Γ_{α_i}. It naturally holds that $\mathbb{M}_1 \supseteq \mathbb{M}_2$. Therefore, any edge that is saturated in $\mathbb{M}_1$ must be saturated in $\mathbb{M}_2$. On the other hand, edges can be saturated in $\mathbb{M}_2$, which are not saturated in $\mathbb{M}_1$. Thus, $\mathbb{P}_2$ can either be identical to $\mathbb{P}_1$ or it must be a successor $\mathbb{P}_2 \succ \mathbb{P}_1$. ■

Thus, for identical edges, a variation of the edge capacity uncovers a *hierarchical clustering structure* of the dynamical network (3). The complete hierarchical clustering structure can be detected, for example, by decreasing α from a large value to 0 and computing the corresponding saddle-point solutions $\bar{\mathrm{x}}$ repeatedly. Nodes that are separated only for very small α can be considered to be strongly connected to each other, whereas other nodes which are separated earlier for large α are only loosely connected. We will return to the idea of hierarchical clustering later on in Section VII, where we analyze the hierarchical structure of a power systems network. Let us note first another implication of Lemma 6.1.

Theorem 6.2 (Minimal Agreement Capacity): A *minimal agreement capacity* α_{min} exists, such that for any edge capacity $\alpha \geq \alpha_{\min}$, the network (3) reaches agreement and $\bar{\mathrm{x}} = \mathrm{x}^*$.

Proof: For a given edge capacity α, the constraint set for the saddle-point problem (8) is given by $\Gamma_\alpha = \{\mu : |\mu_k| \leq \alpha, k \in \{1, \ldots, |\mathbf{E}|\}\}$. Recall that the solution x^* to (8) forms a 1-cluster if and only if $\Gamma \cap \mathbb{M}^* \neq \emptyset$. Now for any $\alpha \geq \alpha_{\min}$, we have the set $\{\mu_{\min} : \mu_{\min} = \arg\min_{\mu \in \mathbb{M}^*} \|\mu\|_\infty\}$, which is a subset of the intersection $\Gamma_\alpha \cap \mathbb{M}^*$ (i.e., for any $\alpha \geq \alpha_{\min} = \min_{\mu \in \mathbb{M}^*} \|\mu\|_\infty$, the elements $\mu_{\min}$ are contained in the intersection). On the other hand, for all $\alpha < \alpha_{\min}$, the intersection $\Gamma_\alpha \cap \mathbb{M}^* = \emptyset$. According to Theorem 4.9, $\alpha_{\min}$ is thus the minimal agreement capacity. ∎

For the considered setup with identical capacities, the minimal agreement capacity $\alpha_{\min}$ can be computed as follows.

Corollary 6.3: Let $\mathbb{M}^*$ be the set of dual solutions to the network optimization problem (9). The minimal agreement capacity can be computed as

$$\alpha_{\min} := \min_{\mu \in \mathbb{M}^*} \|\mu\|_\infty. \tag{16}$$

Recall that the minimal agreement capacity provides the smallest saturation bounds for the interaction functions in the dynamics (3) for which the network reaches full agreement on the node states. Surprisingly, there is a direct connection between the minimal agreement capacity and weighted partitions of the graph.

Let us consider a network in agreement with agreement state x^* and interpret this network as a node-weighted graph, with weights corresponding to $\nabla F_i(\mathrm{x}^*)$. We consider now the combinatorial problem of partitioning the graph into 2-partitions. In order to compare different partitions, we need to define a measure which quantifies the *quality* of partitions.

Definition 6.4: Consider a 2-partition $\mathbb{P} = (\mathbf{P}_1, \mathbf{P}_2)$ with the corresponding cutset $\mathbf{Q}$. Let $\nabla F_{\mathcal{P}_1} := \sum_{i \in \mathcal{P}_1} \nabla F_i(\mathrm{x}^*) = \sum_{i \in \mathcal{P}_1} f_i(\mathrm{x}^*)$ $(\nabla F_{\mathcal{P}_2} := \sum_{i \in \mathcal{P}_2} \nabla F_i(\mathrm{x}^*))$ and assume without loss of generality that $\nabla F_{\mathcal{P}_1} \geq 0$. The quality of $\mathbb{P}$ is defined as

$$\Psi(\mathbb{P}) := \frac{\nabla F_{\mathcal{P}_1}}{|\mathbf{Q}|} \tag{17}$$

where $|\mathbf{Q}|$ denotes the number of edges in the cutset $\mathbf{Q}$.

Note that due to the optimality property, $\nabla F_{\mathcal{P}_1} = -\nabla F_{\mathcal{P}_2}$. The definition of the partition quality can be understood in the following way. The quality measures the "imbalance" between the two clusters over the number of edges connecting them. A high-quality partition is one where the two partitions are strongly imbalanced and only connected by a few edges. The problem of detecting the 2-partition of a network with maximal quality is a combinatorial problem. However, there is an intimate connection to the minimal agreement capacity, as the following duality-like relation shows.

Theorem 6.5 (Combinatorial Equivalence): Let $\mathbb{M}^*$ be the set of all solutions to the network problem (9) and let $\mathbb{P}^2(\mathcal{G})$ be the set of all possible 2-partitions of $\mathcal{G}$. Then

$$\min_{\mu \in \mathbb{M}^*} \|\mu\|_\infty = \max_{\mathbb{P} \in \mathbb{P}^2(\mathcal{G})} \Psi(\mathbb{P}).$$

Remark 6.6: The previous statement has the flavor of a duality relation. In fact, the dual to $\min_{\mu \in \mathbb{M}^*} \|\mu\|_\infty$ in the linear programming sense is $\max_u \nabla \mathbf{F}(\mathrm{x}^*)' u$, s.t. $\|E(\mathcal{G})' u\|_1 \leq 1$. It is discussed, for example, in [26] that a solution to the latter problem corresponds to a partition of the graph $\mathcal{G}$. Our result, however, states the relation of the linear program to the combinatorial problem more explicitly.

Proof: Before presenting the main idea of the proof, we need some preliminary discussion. Assume that a partition $\mathbb{P} = (\mathbf{P}_1, \mathbf{P}_2)$ and the corresponding cutset $\mathbf{Q}$ are given. For a given solution x^* to the network optimization problem (9), one directly obtains the vector $\nabla \mathbf{F}(\mathrm{x}^*)$, which is defined on the node space. Without loss of generality, one can apply a permutation, resorting the nodes, and write $\nabla \mathbf{F}(\mathrm{x}^*) = [\nabla \mathbf{F}(\mathcal{P}_1)', \nabla \mathbf{F}(\mathcal{P}_2)']'$, where $\nabla \mathbf{F}(\mathcal{P}_1) \in \mathbb{R}^{|\mathbf{P}_1|}$ is a vector consisting of all entries of $\nabla \mathbf{F}(\mathrm{x}^*)$ which correspond to nodes in the partition $\mathbf{P}_1$. With only a slight abuse of our conventional notation, we rewrite the incidence matrix of the subgraph induced by the cutset $\mathbf{Q}$ as $E(\mathcal{Q}) = [E_1(\mathcal{Q})' \; E_2(\mathcal{Q})']'$, where $E_1(\mathcal{Q})$ is formed by all rows corresponding to nodes in the partition $\mathbf{P}_1$. We now write the first-order optimality conditions for (9) in the form

$$\begin{bmatrix} \nabla \mathbf{F}(\mathcal{P}_1) \\ \nabla \mathbf{F}(\mathcal{P}_2) \end{bmatrix} + \begin{bmatrix} E(\mathcal{P}_1) & E_1(\mathcal{Q}) & 0 \\ 0 & E_2(\mathcal{Q}) & E(\mathcal{P}_2) \end{bmatrix} \begin{bmatrix} \mu(\mathcal{P}_1) \\ \mu(\mathcal{Q}) \\ \mu(\mathcal{P}_2) \end{bmatrix} = 0 \tag{18}$$

where $E(\mathcal{P}_1)$ and $E(\mathcal{P}_2)$ are the incidence matrices of the subgraphs induced by the partitions $\mathbf{P}_1$ and $\mathbf{P}_2$, respectively. Note that any $\mu \in \mathbb{R}^{|\mathbf{E}|}$, which satisfies the first-order condition (18), is an optimal solution to (9) and is thus in the set $\mathbb{M}^*$.

We aim to derive a condition relating $\Psi(\mathbb{P})$ and μ. To begin, multiply (18) from the left with the matrix

$$\frac{1}{|\mathbf{Q}|} \begin{bmatrix} \mathbb{1}'_{|\mathbf{P}_1|} & 0 \\ 0 & \mathbb{1}'_{|\mathbf{P}_2|} \end{bmatrix} \in \mathbb{R}^{2 \times n}$$

to obtain the condition

$$\begin{bmatrix} \frac{\nabla F_{\mathcal{P}_1}}{|\mathbf{Q}|} \\ \frac{\nabla F_{\mathcal{P}_2}}{|\mathbf{Q}|} \end{bmatrix} = -\frac{1}{|\mathbf{Q}|} \begin{bmatrix} \mathbb{1}'_{|\mathbf{P}_1|} E_1(\mathcal{Q}) \\ \mathbb{1}'_{|\mathbf{P}_2|} E_2(\mathcal{Q}) \end{bmatrix} \mu(\mathcal{Q}). \tag{19}$$

Let us shortly discuss the properties of the vector $\mathbb{1}'_{|\mathbf{P}_1|} E_1(\mathcal{Q})$. One can directly see that $\mathbb{1}'_{|\mathbf{P}_1|} E_1(\mathcal{Q}) \in \{-1, 1\}^{|\mathbf{Q}|}$, with $[\mathbb{1}'_{|\mathbf{P}_1|} E_1(\mathcal{Q})]_k = 1$ if edge $k \in \mathbf{Q}$ originates in partition $\mathbf{P}_1$ and -1, otherwise. Furthermore, it holds that $\mathbb{1}'_{|\mathbf{P}_1|} E_1(\mathcal{Q}) = -\mathbb{1}'_{|\mathbf{P}_2|} E_2(\mathcal{Q})$. Analogous to this vector, we define a new vector $c_{\mathcal{P}} \in \{-1, 0, 1\}^{|\mathbf{E}|}$, which extends $\mathbb{1}'_{|\mathbf{P}_1|} E_1(\mathcal{Q})$ to the complete edge space, representing the cutset in the following sense: $[c_{\mathcal{P}}]_k$ takes the value "+1" if edge k originates in $\mathbf{P}_1$, "−1" if the edge terminates in $\mathbf{P}_1$, and "0" if $k \notin \mathbf{Q}$. Note that $c_{\mathcal{P}}$ has exactly $|\mathbf{Q}|$ nonzero entries.

Now we can derive from (19) the following condition which must hold for all $\mu \in \mathbb{M}^*$:

$$\Psi(\mathbb{P}) = -\frac{1}{|\mathbf{Q}|} c'_{\mathcal{P}} \mu. \tag{20}$$

We are now ready to prove the main statement. Therefore, we define $\alpha_{\min} := \min_{\mu \in \mathbb{M}^*} \|\mu\|_\infty$ (as we defined the minimal agreement capacity) and $\Psi^* := \max_{\mathbb{P} \in \mathbb{P}^2(\mathcal{G})} \Psi(\mathbb{P})$ (i.e. $\Psi^* \geq \Psi(\mathbb{P})$ for all $\mathbb{P} \in \mathbb{P}^2(\mathcal{Q})$).

First, we use $\alpha_{\min}$ to provide an upperbound for the right-hand side of (20). For all $\mathbb{P} \in \mathbb{P}^2(\mathcal{G})$, $\Psi(\mathbb{P}) \leq (1/|\mathbf{Q}|)\|c_{\mathcal{P}}\|_1\|\mu\|_\infty = \alpha_{\min}$ and thus

$$\Psi^* \leq \alpha_{\min}. \tag{21}$$

Next, we provide a lower bound for Ψ^*. We assume, by contradiction, that $\Psi^* < \alpha_{\min}$. Then, for all $c_{\mathcal{P}}$ and all $\mu \in \mathbb{M}^*$

$$-c'_{\mathcal{P}}\mu < |\mathbf{Q}|\alpha_{\min}.$$

We can now choose $\tilde{\mu} = \arg\min_{\mu\in\mathbb{M}^*} \|\mu\|_\infty$. Since then for all edges l, $|\tilde{\mu}_l| \leq \alpha_{\min}$, it must be possible to choose for any 2-partition $\mathbb{P}$ in the graph $\mathcal{G}$, $-c'_{\mathcal{P}}\tilde{\mu} < |\mathbf{Q}|\alpha_{\min}$. This implies that no cutset $\mathbf{Q}$ exists so that $\tilde{\mu}_k = -\alpha_{\min}c_{\mathcal{P}k}$, $k \in \mathbf{Q}$, which, in turn, requires that every edge $l \in \mathbf{E}$ for which $|\tilde{\mu}_l| = \alpha_{\min}$ is contained in at least one cycle $\mathbf{C}$ with $|\tilde{\mu}_k| < \alpha_{\min}$ for all $k \in \mathbf{C} \setminus \{l\}$. With the same argumentation as used to prove Lemma 4.6, we can conclude that another vector must exist $\breve{\mu} \in \mathbb{M}^*$ such that $|\breve{\mu}_k| < \alpha_{\min}$ for all $k \in \mathbf{C}$. But this contradicts the assumption that $\alpha_{\min} = \min_{\mu\in\mathbb{M}^*} \|\mu\|_\infty$. Having shown a contradiction, we conclude that

$$\Psi^* \geq \alpha_{\min}. \tag{22}$$

The two conditions (21) and (22) finally lead to the desired conclusion $\alpha_{\min} = \Psi^*$. ∎

The connection between the minimal agreement capacity and optimal graph partitions has a direct consequence for the dynamical network (3).

Theorem 6.7 (Combinatorial Clustering Condition): The dynamical network (3) with identical edge capacities α is clustering (i.e., is forming a p-partition with $p \geq 2$) if and only if a 2-partition $\mathbb{P} \in \mathbb{P}^2(\mathcal{G})$ exists with a quality exceeding the edge capacity

$$\Psi(\mathbb{P}) > \alpha. \tag{23}$$

Proof: The statement follows directly from Theorem 6.2 and Theorem 6.5. ∎

We want to emphasize again that the given result explicitly relates the purely graph theoretic concept of optimal partitioning to the behavior of the dynamical network. The connection between the clustering structure of the network (3) is even stronger than the necessary and sufficient clustering condition defined in the previous theorem. To characterize the connection explicitly, we consider the following concept.

Definition 6.8: The *first-cut* $\mathbf{Q}_{fc}$ of the network $\mathcal{G}$ pertains to the edges that are saturated for $\alpha = \alpha_{\min}$.

The name "first-cut" emphasizes that this is the first set of edges to become saturated as α is decreased from a large value.

Corollary 6.9: Assume the first-cut $\mathbf{Q}_{fc}$ induces a 2-partition $\mathbb{P}_{fc}$, then

$$\mathbb{P}_{fc} = \arg\max_{\mathbb{P}\in\mathbb{P}^2(\mathcal{G})} \Psi(\mathbb{P}) = \arg\max_{\mathbb{P}\in\mathbb{P}^2(\mathcal{G})} \frac{\sum_{i\in\mathcal{P}_1} f_i(\mathrm{x}^*)}{|\mathbf{Q}|}. \tag{24}$$

This statement is a direct consequence of the proof for Theorem 6.5. It shows that the partition with the highest quality is the one which is "most likely" to appear in the clustering structure of the network. We can additionally characterize the first-cut $\mathbf{Q}_{fc}$ precisely in terms of the quality of all partitions in a network.

Proposition 6.10: Assume $\mathbb{P}_1, \ldots, \mathbb{P}_q$ are all 2-partitions maximizing $\Psi(\mathbb{P})$ in (17) and let $\mathbf{Q}_1, \ldots, \mathbf{Q}_q$ be the corresponding cut-sets. Then, the first-cut is given by $\mathbf{Q}_{fc} := \mathbf{Q}_1 \cup \cdots \cup \mathbf{Q}_q$.

Proof: Denote the optimal quality by Ψ^*. By assumption, we have $\Psi(\mathbb{P}_i) = \Psi^* = \alpha_{\min}$, $i \in \{1, \ldots, q\}$. Furthermore, $\tilde{\mu} \in \mathbb{M}^*$ exists with $|\tilde{\mu_k}| \leq \alpha_{\min}$, $k \in \mathbf{E}$, such that

$$\Psi(\mathbb{P}_i) = -\frac{1}{|\mathbf{Q}_i|} c'_{\mathcal{P}i}\tilde{\mu} \stackrel{!}{=} \alpha_{\min}.$$

For this to be true, it must hold that $|\tilde{\mu}_k| = \alpha_{\min}$ for all $k \in \mathbf{Q}_1 \cup \ldots \cup \mathbf{Q}_q$. ∎

The implications of this result are two-fold. On the one hand, it emphasizes how important the partitions with high quality are for the clustering structure. On the other hand, the first-cut can be considered as a computational means to find the optimal partitioning of the graph.

Clustering and Graph Partitioning Problems: To further illustrate the significance of the previous results, we show that the clustering structure of the network is inherently related to well-established partitioning problems. We show that there is a close relation of our results to two problems: the *inhibiting bisection problem* and the *min s-t-cut*.

First, we clarify the relation of our first-cut to the *inhibiting bisection problem* studied in [14] and [27]. The motivation for the mentioned work is to use graph partitioning methods for computing the critical lines in large-scale power networks. The authors formulate a variation of the classical bisection problem by searching for the cut separating the network into two groups which: 1) has the minimal cut size and 2) leads to the maximal power imbalance between the two clusters. The following result shows that the mathematical formulation of our problem is closely connected to the one studied in [27].

For the presentation of the result, we need the notion of *indicator vectors*. A vector $\xi \in \{-1, 1\}^n$ is said to be an indicator vector for a two partition $\mathbb{P} = (\mathbf{P}_1, \mathbf{P}_2)$ if $\xi_i = 1$ if $i \in \mathbf{P}_1$ and $\xi_i = -1$ if $i \in \mathbf{P}_2$. The set of indicator vectors for all possible 2-partitions is denoted by $\mathbb{I}_2$.

Lemma 6.11: Assume that the first-cut induces a 2-partition and let ξ^* be the indicator vector of this 2-partition. Then ξ^* is a minimizer of

$$\min_{\xi\in\mathbb{I}_2} \quad \alpha_{\min}\xi' L(\mathcal{G})\xi - 2\left|\nabla\mathbf{F}(\mathrm{x}^*)'\xi\right| \tag{25}$$

where $\alpha_{\min}$ is the minimal agreement capacity and $L(\mathcal{G})$ is the Laplacian of $\mathcal{G}$.

Remark 6.12: If $\nabla\mathbf{F}(\mathrm{x}^*)$ is interpreted as the vector of power supplies and demands on the nodes, then problem (25) corresponds to the inhibiting bisection problem studied in [27].

Proof: Given a 2-partition $\mathbb{P}$ and the corresponding indicator vector $\xi \in \mathbb{I}_2$, it follows that $|\mathbf{Q}| = (1/4)\xi' L(\mathcal{G})\xi$, and $|\sum_{i\in\mathcal{P}_1} \nabla F_i(\mathrm{x}^*)| = (1/2)|\nabla\mathbf{F}(\mathrm{x}^*)'\xi|$. Thus, we can rewrite the combinatorial problem formulated in Theorem 6.5 as

$$\max_{\xi\in\mathbb{I}_2} 2\frac{|\nabla\mathbf{F}(\mathrm{x}^*)\xi|}{\xi' L(\mathcal{G})\xi}.$$

Recall that Theorem 6.5 states that the value of the optimal 2-partition (which is induced by the first-cut) is $\Psi(\mathbb{P}_{fc}) = \alpha_{\min}$. Thus, we know that for the optimal cut and the corresponding indicator vector ξ^* it must hold that

$2(|\boldsymbol{\nabla}\mathbf{F}(\mathrm{x}^*)\xi^*|/\xi^{*'}L(\mathcal{G})\xi^*) = \alpha_{\min}$. The problem to solve is therefore the feasibility problem of finding $\xi \in \mathbb{I}_2$ so that $|\boldsymbol{\nabla}\mathbf{F}(\mathrm{x}^*)\xi| - \alpha_{\min}\xi' L(\mathcal{G})\xi = 0$.

We show now by contradiction that for all $\xi \in \mathbb{I}_2, \xi \neq \xi^*$ it holds that $|\boldsymbol{\nabla}\mathbf{F}(\mathrm{x}^*)\xi| - \alpha_{\min}\xi' L(\mathcal{G})\xi < 0$. Assume that $\bar{\xi} \in \mathbb{I}_2$ exists such that $|\boldsymbol{\nabla}\mathbf{F}(\mathrm{x}^*)\bar{\xi}| - \alpha_{\min}\bar{\xi}' L(\mathcal{G})\bar{\xi} > 0$. Then, it must also hold that $2(|\boldsymbol{\nabla}\mathbf{F}(\mathrm{x}^*)\bar{\xi}|/\bar{\xi}' L(\mathcal{G})\bar{\xi}) > \alpha_{\min}$. But this contradicts that $\alpha_{\min}$ is the optimal value of all partitions. The case with equality is excluded by the assumption that the first-cut induces a 2-partition, which implies that there is a unique partition with maximal quality. Keeping this in mind, we can again rewrite the original problem (24) as the new optimization problem

$$\max_{\xi \in \mathbb{I}_2} |\boldsymbol{\nabla}\mathbf{F}(\mathrm{x}^*)\xi| - \alpha_{\min}\xi' L(\mathcal{G})\xi \tag{26}$$

with a maximal value of exactly 0. This, however, is equivalent to the optimization problem formulated in the statement. ■

The connection to the inhibiting bisection problem is remarkable for two reasons. First, our setup generalizes the problem studied in [27] since it allows not only to compute the first-cut, but to detect the complete hierarchical clustering structure. Second, it connects the inhibiting bisection problem to a dynamical network model.

To further emphasize the relevance of our setup, we show that the clustering structure of our dynamical network model (3) corresponds, for a certain setup, to the min $s-t$-cut. The min $s-t$ cut is one of the most basic graph partitioning problems, dating back to Ford and Fulkerson [15].

Definition 6.13 (min-s-t cut): Given a sink node $s \in \mathbf{V}$ and a source node $t \in \mathbf{V}$, the min-$s-t$ cut is the cut with the minimal number of edges that separates s and t.

For a particular setup, our first-cut corresponds exactly to the min $s-t$-cut. The correspondence holds, however, only for a particular choice of the objective functions. To define the objective functions, we choose a constant $b \in \mathbb{R}_{>0}$ and set

$$F_i(x_i) = \begin{cases} \frac{1}{2}(x_s - b)^2 & \text{if } i = s \\ \frac{1}{2}(x_s + b)^2 & \text{if } i = t \\ \frac{1}{2}x_i^2 & \text{if } i \neq s, t. \end{cases} \tag{27}$$

Lemma 6.14: Consider the problem (8) with the objective functions defined as in (27). Let $\mathbf{Q}_{st,1}, \ldots, \mathbf{Q}_{st,q}$ be all minimal $s-t$-cuts. Then, the first-cut $\mathbf{Q}_{fc}$ is the union of *all* min $s-t$-cuts (i.e., $\mathbf{Q}_{fc} = \mathbf{Q}_{st,1} \cup \ldots \cup \mathbf{Q}_{st,q}$).

Proof: We note that the solution to (9) is $\mathrm{x}^* = 0$. Therefore, $\nabla F_s(\mathrm{x}^*) = -b$, $\nabla F_t(\mathrm{x}^*) = b$ and $\nabla F_i(\mathrm{x}^*) = 0$. This implies that for any 2-partition $\mathbb{P} = (\mathbf{P}_1, \mathbf{P}_2)$ that separates the nodes s and t (i.e. $s \in \mathbf{P}_1$ and $t \in \mathbf{P}_2$), the value $|\sum_{i \in \mathcal{P}_1} \nabla F_i(\mathrm{x}^*)| = b$. Any partition that does not separate s and t gives $|\sum_{i \in \mathcal{P}_1} \nabla F_i(\mathrm{x}^*)| = 0$. Theorem 6.5 states that the first-cut will maximize the quantity $\Psi(\mathbb{P})$. We know that the quantity $\Psi(\mathbb{P})$ is zero for any partition not separating s and t and it will thus be maximized by a partition separating the two nodes. We conclude that the first-cut is the cut maximizing $\Psi(\mathbb{P}) = b/|\mathbf{Q}_{\mathcal{P}}|$ over all cuts separating s and t. The maximum is achieved for the cut with a minimal number of edges (i.e., for the cut minimizing $|\mathbf{Q}_{\mathcal{P}}|$ which is the min $s-t$-cut). Proposition 6.10 follows that the first-cut will be the union of all min $s-t$ cuts. ■

It clearly matches the intuition that a dynamical network will cluster along the edges of the minimal cut, and we clarify here the assumptions under which this will occur. In summary, the previous results show that the clustering structure of our dynamical network model (3) is strongly connected to the properties of the underlying graph topology.

VII. Structural Analysis of the IEEE 30-bus Electric Power System

Dynamic clustering is a well-known instability phenomenon in power networks where, in the case of failures, the network partitions into groups, and generators within one group synchronize their frequencies but are out of synch with generators in other groups. Although often observed, it is a difficult problem to relate the graph topology and the expected clustering structure of power networks.

A standard dynamical model for power systems in the general form (1) that exhibits such clustering behavior for certain network configurations is the *structure preserving model* proposed by Bergen and Hill [18]

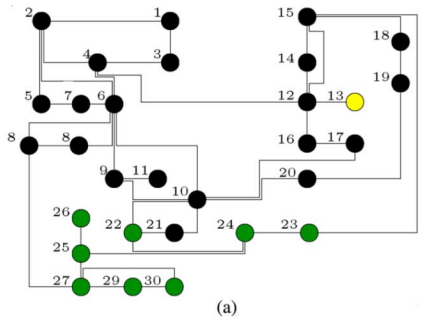
$$M_i\ddot{\delta}_i + D_i\dot{\delta}_i + \sum_{j=1}^{n} \alpha_{ij}\sin(\delta_i - \delta_j) = P_i. \tag{28}$$

Here, $\delta_i \in [0, 2\pi)$ is the rotor angle at bus i, M_i are the inertia constants of the generators, D_i are the damping coefficients of the load or generator i, and α_{ij} is a coefficient computed from the voltage of the incident buses and the impedance of the line and is symmetric (i.e., $\alpha_{ij} = \alpha_{ji}$). The power supply or demand at bus i is P_i.[2] The clustering structure observed in the model (28) is related to the network topology, the edge capacities α_{ij}, and the power supply. Identifying clustering structures based on the dynamical model (28) is considered a hard problem (see, e.g., [28]). The main difference between (1) and (28) is the coupling nonlinearity. One can therefore interpret the model (1) as a simplified model, which, in contrast to a linearized model, still maintains the ability to cluster.

From a *computational* perspective, the structural analysis of power systems has attained significant attention aiming to detect strongly connected components and critical cut-sets (see, e.g., [27] and [29]). In this example, we show that our analysis provides a bridge between purely graph theoretical analysis methods for power networks and the study of dynamical network models. We demonstrate how our results relate to the computation approach for identifying critical cut-sets proposed in [27]. Note that [27] considers a purely computational approach, whereas our analysis is directly connected to a dynamcial network model. In order to obtain comparative results, we use the same problem setup as [27] (i.e., a modified version of the IEEE 30-bus power network).

The graph topology and the considered network data are summarized in Fig. 3(a) and(b), respectively. In [27], the nodes {22, 23, 24, 25, 26, 27, 29, 30} as well as the single node {13} were identified to be loosely connected to the rest of the network. Following [27], we take P_i to be the power injection, and we

[2]We only sketch the power network model here since the purpose of this paper is to explain the mathematical basics of clustering and not to provide a detailed analysis of power systems.



(a)

Node i	P_i[MW]	Node i	P_i[MW]	Node i	P_i[MW]
1	145.3	11	0	21	-175
2	-7.3	12	-112	22	315.9
3	-24	13	410	23	310
4	-76	14	-62	24	-87
5	0	15	-82	25	0
6	0	16	-35	26	-35
7	-228	17	-90	27	469.1
8	-300	18	-32	28	0
9	0	19	-95	29	-24
10	-58	20	-22	30	-106

(b)

Fig. 3. Graph structure and network data of the IEEE 30-bus system. Node 13 is the single node which is separated by the first-cut. The nodes {22, 23, 24, 25, 26, 27, 29, 30} are separated by the second-cut. (a) Network structure. (b) Network data.

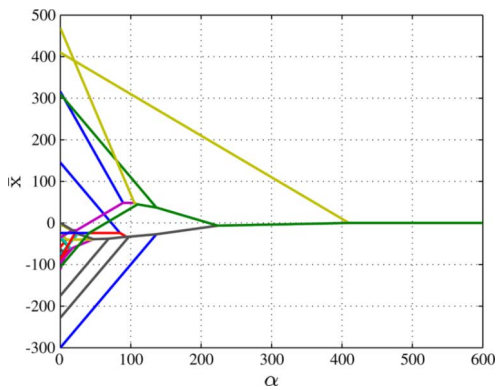


Fig. 4. Optimal solution $\bar{\mathrm{x}}$ to the saddle-point problem (8) as a function of the edge capacities α.

choose the damping as $D_i = 1$. To detect the clustering structure of the network using our novel saddle-point analysis, we consider identical edge capacities (i.e., $\alpha_i = \alpha$ for all $i \in \mathbf{E}$), which we vary continuously. Fig. 4 visualizes how the optimal solution of the saddle-point problem (8), $\bar{\mathrm{x}}$, varies as a function of the edge capacities α. Observe the characteristic *dendogram* structure of the function illustrated in Fig. 4. For very large values of α, the entire network is in agreement, while as α decreases, a partitioning of the network appears (represented by multiple values for $\bar{\mathrm{x}}$ for a fixed α in Fig. 4). This shows that our saddle-point problem provides a *hierarchical clustering* of the network. A deeper analysis of the clustering structure reveals that the first-cut partitions only a single node, node 13, from the network. The "second-cut," occurring at $\alpha \approx 223.5$, however, separates the network nodes {22, 23, 24, 25, 26, 27, 29, 30}. Our saddle-point analysis computes therefore the *same* critical cut-sets and strongly connected components as the approach proposed in [27]. Please note that our analysis provides significantly more information, since it uncovers not only one partition, but a complete hierarchical clustering structure.

We cannot expect our model to show exactly the same behavior as the power network model (28). However, our dynamical model (1) still captures the hierarchical clustering structure. Recall now the motivating example given in Section III and Fig. 1. The trajectories were generated using our simplified version of (28) with $\psi_k(\cdot) = \tanh(\cdot)$. For each simulation, the initial conditions $\dot{\chi}(0)$ were chosen randomly, whereas $\chi(0)$ is always set to zero. The edge capacities for each simulation are, respectively, $\alpha_{[1]} = 500$, $\alpha_{[2]} = 180$, and $\alpha_{[3]} = 20$. The simulation results clearly show that our dynamical model is a reasonable attempt to bridge the existing gap between graph-based network analysis and the widely observed cluster synchronization phenomenon. While this work does not claim to provide a formal analysis of the power network model (28), we believe the qualitative similarity with our model is of value. Indeed, the understanding of the simplified dynamical model with clustering behavior can provide new insights.

VIII. Conclusions

This paper examined the phenomena of clustering within a coupled dynamical system. The main principle leading to clustering, as considered in this paper, is as follows. Every node has an individual preference and is attracted to its neighbors by a bounded interaction force. If the interaction between two neighboring nodes is large enough, they will agree on a common value, if not, they will split. This intuitive idea leads to the development of a static saddle-point problem, which can explain clustering from a computational perspective. The saddle-point problem has an interpretation as a Lagrange-dual problem to a network optimization problem with additional constraints on the dual variables. This result provides a novel perspective on clustering and establishes a direct connection between graph-based, static clustering algorithms, and the behavior of dynamical networks.

Appendix

A. Proof of Proposition 5.2

Proof: We consider a convergent sequence $\{\bar{\mu}^{[k]}\}_{k=1}^{\infty}$ such that $\bar{\mu}^{[k]} \in \mathrm{int}\Gamma$ and $\lim_{k\to\infty} \bar{\mu}^{[k]} \in \mathbb{M}$. We define $\omega^{[k]} = -E(\mathcal{G})\bar{\mu}^{[k]}$ and note that due to the first-order saddle-point optimality conditions, $\omega^{[k]} \to \boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}})$ as $k \to \infty$. Consider now the sequence of optimization problems

$$\min_{z,u} \quad P(z) - \bar{\mu}^{[k]'} z, \quad \text{s.t.} \quad z = E(\mathcal{G})' u \tag{29}$$

where $P(z)$ is chosen such that $\nabla P(z) = W\psi(z)$. Any problem in the sequence has a finite optimal solution since, for any $\bar{\mu}^{[k]} \in \mathrm{int}\Gamma$, some $\tilde{z}^{[k]} \in \mathbb{R}^{|\mathbf{E}|}$ exists such that $\nabla P(\tilde{z}^{[k]}) - \bar{\mu}^{[k]} = W\boldsymbol{\psi}(\tilde{z}^{[k]}) - \bar{\mu}^{[k]} \stackrel{!}{=} 0$ and $\nabla^2 P(z^{[k]}) = W\nabla\boldsymbol{\psi}(z^{[k]}) \geq 0$ for all $z^{[k]}$. Therefore, the objective of the optimization problem is bounded from below and, thus, a finite optimal solution always exists.

The Lagrangian of the problem is defined as $\mathcal{L}(z,u,\delta) = P(z) - \bar{\mu}^{[k]'} z - \delta' z + \delta' E(\mathcal{G})' u$, and the KKT conditions then tell us that for any optimal point $(z^{[k]*}, u^{[k]*}, \delta^{[k]*})$,

$$\nabla\mathcal{L}(z,u,\delta)|_{z^{[k]*},u^{[k]*},\delta^{[k]*}} = 0.$$

This leads us to $\nabla_z\mathcal{L}(z,u,\delta)|_{z^{[k]*},u^{[k]*},\delta^{[k]*}} = \nabla P(z^{[k]*}) - \delta^{[k]*} - \bar{\mu}^{[k]} = 0$ from which we conclude $\delta^{[k]*} = W\psi(z^{[k]*}) -$

$\bar{\mu}^{[k]}$. In addition, $\nabla_u \mathcal{L}(z,u,\delta)|_{z^{[k]*},u^{[k]*},\delta^{[k]*}} = E\delta^{[k]*} = 0$ shows that

$$0 = E(\mathcal{G})W\boldsymbol{\psi}(z^*) - E(\mathcal{G})\bar{\mu} \Rightarrow 0 = E(\mathcal{G})W\boldsymbol{\psi}\left(z^{[k]*}\right) - \omega^{[k]}.$$

Together with $E(\mathcal{G})'u^{[k]*} = z^{[k]*}$, we can conclude that $z^{[k]*} \in \mathcal{R}(E^T)$. Since $\omega^{[k]} \to \boldsymbol{\nabla}\mathbf{F}(\bar{\mathrm{x}})$, for $k \to \infty$, the sequence $\{z^{[k]}\}$ converges such that $W\psi(z^{[k]*}) \to \mathbb{M}$ as $k \to \infty$. ■

REFERENCES

[1] M. Escalona-Morán, M. Cosenza, P. Guillén, and P. Coutin, "Synchronization and clustering in electroencephalographic signals," *Chaos, Solitons & Fractals*, vol. 31, no. 4, pp. 820–825, 2007.

[2] D. Lazer, A. Pentland, L. Adamic, S. Aral, and A.-L. Barabasi, "Computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.

[3] W. Lu, B. Liu, and T. Chen, "Cluster synchronization in networks of coupled nonidentical dynamical systems," *Chaos*, vol. 20, p. 013120, 2010.

[4] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Princeton, NJ: Princton Univ. Press, 2010.

[5] M. Arcak, "Passivity as a design tool for group coordination," *IEEE Trans. Autom. Control*, vol. 52, no. 8, pp. 1380–1390, Aug. 2007.

[6] P. DeLellis, M. di Bernardo, and G. Russo, "On QUAD, lipschitz, and contracting vector fields for consensus and synchronization of networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 3, pp. 576–573, Mar. 2011.

[7] W. Wu, W. Zhou, and T. Chen, "Cluster synchronization of linearly coupled complex networks under pinning control," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, no. 4, pp. 829–839, Apr. 2009.

[8] Z. Ma, Z. Liu, and G. Zhang, "A new method to realize cluster synchronization in connected chaotic networks," *Chaos*, vol. 16, p. 023103, 2006.

[9] W. Xia and M. Cao, "Clustering in diffusively coupled networks," *Automatica*, vol. 47, pp. 2395–2405, 2011.

[10] V. Blondel, J. M. Hendrickx, and J. N. Tsitsiklis, "On krauses multi-agent consensus model with state-dependent connectivity," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2586–2597, Nov. 2009.

[11] F. De Smet and D. Aeyels, "Clustering in a network of non-identical and mutually interacting agents," in *Proc. Roy. Soc. A*, 2009, vol. 465, pp. 745–768.

[12] D. Aeyels and F. De Smet, "Cluster formation in a time-varying multiagent system," *Automatica*, vol. 47, no. 11, pp. 2481–2487, 2011.

[13] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, pp. 75–174, 2010.

[14] A. Pinar, Y. Fogel, and B. Lesieutre, "The inhibiting bisection problem," Lawrence Berkeley National Laboratory, Berkeley, CA, 2006. [Online]. Available: http://escholarship.org/uc/item/0gt129fs, Tech. Rep.

[15] C. Ford and D. Fulkerson, "Maximal flow through a network," *Can. J. Math.*, vol. 8, pp. 399–404, 1956.

[16] C. Godsil and G. Royle, *Algebraic Graph Theory*. New York: Springer, 2001.

[17] D. Helbing, "Traffic and related self-driven many-particle systems," *Rev. Mod. Phys.*, vol. 73, pp. 1067–1141, Dec. 2001.

[18] A. Bergen and D. Hill, "A structure preserving model for power system stability analysis," *IEEE Trans. Power App. Syst.*, vol. PAS-100, no. 1, pp. 25–35, Jan. 1981.

[19] D. Zelazo and M. Mesbahi, "Edge agreement: Graph-theoretic performance bounds and passivity analysis," *IEEE Trans. Autom. Control*, vol. 56, no. 3, pp. 544–555, Mar. 2011.

[20] R. T. Rockafellar, *Network Flows and Monotropic Optimization*. Belmont, MA: Athena Scientific, 1998.

[21] D. Bertsekas, "Min common/max crossing duality: A geometric view of conjugacy in convex optimization," Lab. Inf. Dec. Syst., Tech. Rep. LIDS-P-2796, 2009, MIT.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[23] A. Nedic and A. Ozdaglar, "Subgradient methods for saddle-point problems," *J. Optimiz. Theory Appl.*, pp. 205–228, 2009.

[24] R. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1997.

[25] H. Khalil, *Nonlinear Systems*. Upper Saddle River, NJ: Prentice-Hall, 2002.

[26] G. Strang, "Duality in the classroom," *Amer. Math. Monthly*, vol. 91, pp. 250–254, 1984.

[27] B. Lesieutre, S. Roy, and A. Pinar, "Power system extreme event screening using graph partitioning," in *Proc. 38th North Amer. Power Symp.*, 2006, pp. 503–510.

[28] A. Aroposthatis, S. Sastry, and P. Varaiya, "Analysis of power-flow equation," *Int. J. Elect. Power Energy Syst.*, vol. 3, no. 3, pp. 115–126, 1981.

[29] E. Bompard, D. Wu, and F. Xue, "The concept of betweenness in the analysis of power grid vulnerability," in *Proc. Complexity Eng.*, Rome, Italy, 2010, pp. 52–54.

Mathias Bürger received his diploma degree in engineering cybernetics from the University of Stuttgart, Stuttgart, Germany, in 2009, where he is currently pursuing the Ph.D. degree in engineering cybernetics.

He was a Visiting Student at the University of Toronto, Toronto, ON, Canada, and a Visiting Scholar at Queen's University, Kingston, ON, Canada, and at the University of California, Santa Barbara. His research interests include distributed optimization and dynamical network analysis.

Daniel Zelazo received the B.Sc. and M.Eng. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1999 and 2001, respectively, and the Ph.D. degree in aeronautics and astronautics from the University of Washington, Seattle, in 2009.

In between earning the M.Sc. and Ph.D. degrees, he worked on audio compression algorithms at Texas Instruments, Tsukuba, Japan. From 2010 to 2012, he was a Research Associate with the Institute for Systems Theory and Automatic Control, University of Stuttgart, Stuttgart, Germany. Currently, he is an Assistant Professor in the Faculty of Aerospace Engineering, Technion, Israel Institute of Technology, Haifa, Israel. His research interests focus on networked dynamic systems, optimization, and graph theory.

Frank Allgöwer received the Ph.D. degree in chemical engineering from the University of Stuttgart, Stuttgart, Germany.

Currently, he is the Director of the Institute for Systems Theory and Automatic Control, University of Stuttgart. Prior to his present appointment, he held a Professorship in Electrical Engineering at ETH Zurich, Zurich, Switzerland. His main areas of interest are in cooperative control, predictive control, and systems biology.

Dr. Allgöwer received several recognitions for his work, including the prestigious Gottfried–Wilhelm-Leibniz prize of the Deutsche Forschungsgemeinschaft.