

Network Effects on the Robustness of Dynamic Systems

Ketan Savla,¹ Jeff S. Shamma,² and Munther A. Dahleh³

¹Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, California 90089, USA; email: ksavla@usc.edu

²Computer, Electrical, and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia; email: jeff.shamma@kaust.edu.sa

³Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; email: dahleh@mit.edu

Annu. Rev. Control Robot. Auton. Syst. 2020.
3:115–49

First published as a Review in Advance on
November 8, 2019

The *Annual Review of Control, Robotics, and
Autonomous Systems* is online at
control.annualreviews.org

<https://doi.org/10.1146/annurev-control-091219-012549>

Copyright © 2020 by Annual Reviews.
All rights reserved

Keywords

robustness, networks, dynamics, transportation networks, electrical networks, economic networks

Abstract

We review selected results related to the robustness of networked systems in finite and asymptotically large size regimes in static and dynamical settings. In the static setting, within the framework of flow over finite networks, we discuss the effect of physical constraints on robustness to loss in link capacities. In the dynamical setting, we review several settings in which small-gain-type analysis provides tight robustness guarantees for linear dynamics over finite networks toward worst-case and stochastic disturbances. We discuss network flow dynamic settings where nonlinear techniques facilitate understanding the effect, on robustness, of constraints on capacity and information, substituting information with control action, and cascading failure. We also contrast cascading failure with a representative contagion model. For asymptotically large networks, we discuss the role of network properties in connecting microscopic shocks to emergent macroscopic fluctuations under linear dynamics as well as for economic networks at equilibrium. Through this review, we aim to achieve two objectives: to highlight selected settings in which the role of the interconnectivity structure of a network in its robustness is well understood, and to highlight a few additional settings in which existing system-theoretic tools give tight robustness guarantees and that are also appropriate avenues for future network-theoretic investigations.

ANNUAL
REVIEWS **CONNECT**

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

1. INTRODUCTION AND OVERVIEW

Robustness is the ability of a system to operate effectively under a range of different environmental conditions or in the face of possible disruptions. What is the impact of incidents on traffic flow? How does an energy grid respond to surges in demand? How do supply chain disruptions affect a production economy? Under what conditions can a communication network maintain quality of service? Of course, the issue of robustness is at the heart of control systems and a primary motivation for the introduction of feedback in dynamical systems.

While the concept of robustness is widely relevant to multiple domains, an application area of particular importance is networked systems, including physical networks; financial, economic, and social networks; and networks of people and systems. Indeed, tremendous advances in communication, computation, and sensing have led to renewed interest in networked systems, in that new technologies are introducing unprecedented interdependencies among people, devices, and infrastructure.

In settings where the operation of one component impacts the operation of other components, the presence of a network structure significantly affects the characterization of robustness. While each component of a networked system can experience its own disruptions, the impact on the overall system depends critically on the specific structure of the interconnectivity. Stated differently, the impact of the same component-level disruptions may range from inconsequential to critical, depending on the network structure. This article reviews selected results related to this issue of network effects on robustness.

Following an explanation of the notation in Section 2, Section 3 begins the discussion by considering the specific setting of physical flow over networks. There is a network of links, each with limited capacity, that must accommodate the overall flow. Disruptions take the form of reductions in capacity. The focus is on a static problem of whether a network has sufficient excess capacity to withstand the impact of an adversarial environment with a limited budget of capacity reduction, and on how to compute this excess capacity in an efficient manner.

Section 4 presents analysis of the robustness of feedback interconnections. The question is under what conditions a feedback system maintains stability (and, more generally, performance) in the presence of a specified family of possible model perturbations. The framework is motivated primarily by modeling of physical systems, where model simplifications are introduced for the sake of control design and analysis. The underlying network effect is reflected in the specifics of where model perturbations occur in a feedback interconnection. The discussion is separated based on whether the model perturbations are deterministic (i.e., a worst-case analysis) and with memory or are stochastic and memoryless. A motivating application for the latter is a communication network with unreliable links, and the main results there express tight conditions under which the resulting feedback system is stochastically stable.

Section 5 examines the nonlinear dynamic setting in the specific context of network flow dynamics under capacity constraint. Network flow is controlled in a distributed way, i.e., based on local conditions, as opposed to centralized control with full information on network conditions. Specific settings are presented to illustrate the impact of such limited information on network robustness to loss in capacity, and how to compensate for information constraint with additional control action.

Section 6 discusses robustness under cascading failure, under which the collapse of one component in a network propagates to other components, again as determined by an underlying network architecture. Here the motivation is to model the mechanism of failure and quantify network robustness. The specific settings include electrical networks, transport networks, and a contagion model.

Section 7 is motivated by the question of how random disruptions impact a network. A general perspective is that the effects of multiple sources of random disruptions average out to have a diminishing effect as the network grows. Alternatively, it may be that the network structure results in an amplification of such disruptions. These issues motivated the notion of systemic risk in financial networks. In the context of an interconnected production economy, Section 7 presents results that demonstrate that random component-level disruptions can indeed be amplified because of a network structure. This question is addressed in two ways, first in the size (variance) of the overall effect, and then in terms of the probability of an extreme event (as captured by a specific notion of risk). These results are static in that they apply to the equilibrium behavior of a dynamical system, and asymptotic in the sense that they capture the effects for progressively larger networks.

Finally, Section 8 also addresses the impact of disturbances but focuses on the transient behavior (i.e., not just equilibrium) of a dynamical system. The specific setting is the energy of the response of a linear system to a disruptive initial impulse. Again, the conclusions are asymptotic in terms of progressively larger systems (i.e., more states). The underlying network is captured within the specific structure of the dynamics.

These results illustrate multiple approaches that one can take in analyzing the robustness of networked systems—e.g., static versus dynamic models, deterministic versus stochastic disruptions, or asymptotically large versus fixed-size networks. Nonetheless, the common theme throughout is understanding the network effect of how conclusions depend on the underlying interdependency structure. Future research directions along these lines are also suggested at the end of each section.

2. NOTATIONS AND PRELIMINARIES

2.1. Miscellaneous

The sets of real, nonnegative real, and positive real numbers are denoted as \mathbf{R} , $\mathbf{R}_{\geq 0}$, and $\mathbf{R}_{> 0}$, respectively. \mathbf{N} denotes the set of natural numbers. For a set S , $|S|$ denotes its cardinality. \mathbf{R}^S , $\mathbf{R}_{\geq 0}^S$, and $\mathbf{R}_{> 0}^S$ are shorthand notations for $\mathbf{R}^{|S|}$, $\mathbf{R}_{\geq 0}^{|S|}$, and $\mathbf{R}_{> 0}^{|S|}$, respectively. The set of complex numbers is denoted as \mathbf{C} . For a vector $v \in \mathbf{C}$, v^H denotes the complex conjugate transpose. $[m]$ is shorthand for $\{1, 2, \dots, m\}$. For functions $f(\cdot)$ and $g(\cdot)$, we have $f(n) = O(g(n))$, when there exist constants C and n_0 such that $f(n) \leq Cg(n)$ for all $n \in \mathbf{N} > n_0$. If $f(n)$ is equal to $O(g(n))$, then $g(n)$ is equal to $\Omega(f(n))$. $f(n)$ equals $\Theta(g(n))$ when there exist constants C_1 , C_2 , and n_1 such that $C_1g(n) \leq f(n) \leq C_2g(n)$ for all $n \in \mathbf{N} > n_1$.

2.2. Probability Theory

$\mathbb{E}[X]$ and $\text{var}[X]$ denote, respectively, the expected value and variance of random variable X , and $\Phi(\cdot)$ denotes the cumulative distribution function of the Gaussian distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

A random variable X is said to exhibit tail risk (relative to the normal distribution) if $\lim_{\tau \rightarrow \infty} r_X(\tau) = 0$, where the τ -tail ratio of X ,

$$r_X(\tau) = \frac{\log \Pr(X < \mathbb{E}[X] - \tau \sigma_X)}{\log \Phi(-\tau)}, \quad 1.$$

is the probability that X deviates by at least τ standard deviations from its mean relative to a similar probability of deviation under the standard normal distribution. We say that X is light-tailed if $\mathbb{E}[\exp(bX)] < \infty$ for some $b > 0$; otherwise, we say that X is heavy-tailed. Every heavy-tailed random variable exhibits tail risk, but not all random variables with tail risks are heavy-tailed.

2.3. Matrix Theory

For a matrix A , $[A]_{ij}$ denotes its (i, j) th element, A_i denotes its i th row, $A(i)$ denotes its i th column, and $\rho(A)$ denotes its spectral radius. $\sigma(A)$ denotes the singular value of A , which is greater than or equal to other singular values of A . A matrix is called nonnegative if all of its entries are nonnegative. A nonnegative matrix A is said to have a Perron root λ_{PF} if λ_{PF} is a positive real number such that it is an eigenvalue of A and every other eigenvalue λ of A satisfies $|\lambda| < \lambda_{\text{PF}}$. Given compatible matrices A_1 and A_2 , $A_1 \circ A_2$ denotes their Hadamard (i.e., element-by-element) product. I_n is the $n \times n$ identity matrix. $\mathbf{0}_n$ and $\mathbf{1}_n$ denote vectors of all zeros and ones, respectively, of size n ; we omit the subscript on size when clear from the context. Given two vectors a and b of the same size, $a \leq b$ would imply entry-wise inequality. e_i denotes the column vector whose i th entry is one and other entries are zero; its size will be clear from the context.

2.4. Graph Theory

A graph is a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V} = \{v_1, \dots, v_n\}$ is the set of n nodes, \mathcal{E} is the set of edges, and $\mathcal{W} : \mathcal{E} \rightarrow \mathbf{R}$ assigns weights to edges. A directed edge from node i to node j is denoted by $(v_i, v_j) \in \mathcal{E}$; v_i is referred to as the tail node, and v_j is the head node. \mathcal{E}_v^+ and \mathcal{E}_v^- denote the set of edges outgoing from and incoming to node v , respectively (i.e., all the edges whose tail node and head node, respectively, are v). The node-edge incidence matrix $B \in \{\pm 1, 0\}^{\mathcal{V} \times \mathcal{E}}$ is defined such that B_{ve} is equal to -1 or $+1$ if v is the head or tail node of e , respectively, and equal to zero otherwise. An edge (v_i, v_j) is said to be incident onto edge (v_k, v_ℓ) if $v_j = v_k$. \mathcal{G} is called symmetric or undirected if $\mathcal{W}(v_i, v_j) \equiv w_{ij} = w_{ji} \equiv \mathcal{W}(v_j, v_i)$ for all $1 \leq i, j \leq n$. A directed path from v_i to v_j is an ordered sequence of vertices v_i, v_k, \dots, v_j such that any pair of consecutive vertices in the sequence is a directed edge. \mathcal{G} is said to be strongly connected if there exists a directed path from $v_i \in \mathcal{V}$ to $v_j \in \mathcal{V}$ for all $i, j \in [n]$, $i \neq j$, and is said to be weakly connected if the undirected version of \mathcal{G} is strongly connected. $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ is said to be induced by a matrix $A \in \mathbf{R}^{n \times n}$ if $|\mathcal{V}| = n$, $(i, j) \in \mathcal{E}$ if $[A]_{ij} \neq 0$, and $\mathcal{W}(v_i, v_j) = [A]_{ij}$.¹ Conversely, given a graph \mathcal{G} , there exists a matrix A that induces \mathcal{G} . Therefore, we use graph \mathcal{G} and the associated matrix A interchangeably to refer to the same object. Also, with a slight abuse of terminology, we use the terms graph and network interchangeably. Similarly, we use edge and link interchangeably. The degree of a node v in an undirected graph is the sum of the weights of the links at v . If the graph is unweighted, then the degree of a node is simply the number of edges at v . If all the nodes of an undirected graph have the same degree d , then the graph is called d -regular. An $(n-1)$ -regular graph with n nodes is called complete. On the other hand, we distinguish between in- and out-degrees for a directed graph. The out-degree of a node is defined to be equal to the sum of the weights of the edges outgoing from that node; the in-degree is defined similarly.

At times, we refer to large networks, by which we mean a sequence of networks $\{\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n, \mathcal{W}_n)\}_{n \in \mathbf{N}}$ with matrices $\{A_n\}_{n \in \mathbf{N}}$, where the topology of each network in the sequence is fixed but the network dimension (i.e., number of nodes) grows successively.

¹We use \mathcal{W} to denote both the map and the matrix whose entries are w_{ij} . When \mathcal{W} is induced by A , we use \mathcal{W} and A interchangeably.

2.5. Network Flow

Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, with two specific nodes $s, t \in \mathcal{V}$ designated to be the source and sink, respectively. Each link is associated with a flow variable; let $\{f_{ij} \geq 0\}_{(i,j) \in \mathcal{E}}$ be the vector of link-wise flows. We associate each link (i, j) with a flow capacity $c_{ij} > 0$; i.e., the flow variables are constrained to be $f_{ij} \leq c_{ij}$ for all $(i, j) \in \mathcal{E}$. The c_{ij} 's are to be distinguished from link weights w_{ij} 's. Additionally, f is constrained to satisfy flow conservation at every node $i \in \mathcal{V} \setminus \{s, t\}$:

$$\sum_{j: (j,i) \in \mathcal{E}} f_{ji} = \sum_{j: (i,j) \in \mathcal{E}} f_{ij}. \quad 2.$$

The value of flow f is defined to be equal to the difference between outflow and inflow at the source node, i.e., $\sum_{j: (s,j) \in \mathcal{E}} f_{sj} - \sum_{j: (j,s) \in \mathcal{E}} f_{js}$. The maximum-flow problem for a given \mathcal{G}, s , and t is to maximize the value of flows over all f satisfying the capacity and flow conservation constraints. This problem can be formulated as a linear program (e.g., see chapter 8 of Reference 1). Given its importance for several applications involving large networks, the development of computationally efficient algorithms for this problem and its variants continues to attract attention (e.g., see Reference 2).

Interestingly, the solution to the above maximum-flow problem is equal to the solution of a combinatorial optimization problem, the minimum $s - t$ cut problem, over the same data. The solution to the minimum $s - t$ cut problem is equal to the minimum among capacities of all cuts from s to t . A cut from s to t is a subset of links $\tilde{\mathcal{E}} \subset \mathcal{E}$ such that $s \in \tilde{\mathcal{E}}$ and $t \notin \tilde{\mathcal{E}}$. The capacity of cut $\tilde{\mathcal{E}}$ is the sum of the capacities of all links outgoing from $\tilde{\mathcal{E}}$. This equality between the solutions to the two problems is commonly referred to as the max-flow min-cut theorem.

The notion of residual capacity plays an important role in the characterization of robustness for network flow problems. Given a flow $f \in \mathbf{R}_{\geq 0}^{\mathcal{E}}$, the residual capacities of link $e \in \mathcal{E}$, node $v \in \mathcal{V}$, and the network are, respectively, defined to be $c_e - f_e$, $\sum_{e \in \mathcal{E}_v^+} (c_e - f_e)$, and the sum of the residual capacities of links outgoing from the minimum cut. This last item can be shown to be equal to the difference between the minimum-cut capacity and value of f .

In the presence of multiple sources and sinks, one can define a virtual super-source connected via infinite-capacity outgoing links to the individual sources, and similarly a virtual super-sink with infinite-capacity incoming links from individual sinks. One can then consider the maximum-flow problem from the super-source to the super-sink.

2.6. Input-Output Stability

Define

$$\mathbf{L}_2^n = \left\{ f : \mathbf{R}_+ \rightarrow \mathbf{R}_n \text{ such that (s.t.) } \int_0^\infty f^T(t)f(t) dt < \infty \right\}$$

and

$$\mathbf{L}_{2,e}^n = \left\{ f : \mathbf{R}_+ \rightarrow \mathbf{R}_n \text{ s.t. } \int_0^T f^T(t)f(t) dt < \infty, \text{ for all } T \in \mathbf{R}_+ \right\}.$$

For $f \in \mathbf{L}_2^n$, define

$$\|f\| = \left(\int_0^\infty f^T(t)f(t) dt \right)^{1/2}.$$

A mapping $M : \mathbf{L}_{2,e}^n \rightarrow \mathbf{L}_{2,e}^m$ is input–output stable if there exist $\alpha, \beta \geq 0$ such that

$$\|Mf\| \leq \alpha \|f\| + \beta, \forall f \in \mathbf{L}_2^n. \quad 3.$$

In case M is input–output stable and linear, define

$$\|M\| = \sup_{f \in \mathbf{L}_2^n} \frac{\|Mf\|}{\|f\|}.$$

2.7. Linear Dynamical Systems

The notation

$$G \sim \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right)$$

represents the linear time-invariant (LTI) system

$$\begin{aligned} \dot{x} &= Ax + Bu, & x(0) &= 0, \\ y &= Cx + Du, \end{aligned}$$

with associated input–output operation

$$y(t) = Du(t) + \int_0^t C e^{A(t-\tau)} Bu(\tau) d\tau.$$

If A is a stable matrix—i.e., all eigenvalues have strictly negative real parts—then

$$\|G\| = \sup_{\omega} \sigma_{\max}(D + C(j\omega I - A)^{-1}B), \quad 4.$$

where $\sigma_{\max}(\cdot)$ denotes the maximum singular value (see chapter 4 of Reference 3).

3. ROBUSTNESS OF FINITE NETWORKS: STATIC SETTING

Let us consider the robustness of network flow to loss in link capacities. Since the quantity of interest, i.e., flow, is associated with links, for brevity in notation, we switch the indices used for nodes in Section 2 to links. Accordingly, we denote link flows, weights, and capacities as $f = \{f_i\}_{i \in \mathcal{E}}$, $\mathcal{W} = \{\mathcal{W}_i\}_{i \in \mathcal{E}}$, and $c = \{c_i\}_{i \in \mathcal{E}}$, respectively. The starting point for robustness analysis is the max-flow min-cut theorem for the standard static formulation, as described in Section 2.5. Formally, consider the nominal scenario where an inflow of $\lambda > 0$ is routed from s to t . Because λ is less than the solution to the maximum-flow problem, there exists a feasible flow whose value is equal to λ . Let the link capacities be reduced by $\Delta \in [0, c]$. If $\|\Delta\|_1$ is less than the residual capacity of the unperturbed network, then there exists a new flow that is feasible for the perturbed network—i.e., it is link-wise less than $c - \Delta$ and satisfies Equation 2, with the same value λ . On the other hand, there exists a $\|\Delta\|_1$ infinitesimally greater than the network residual capacity, corresponding to reducing the capacity on links outgoing from the minimum cut that is greater than their respective residual capacities, under which such a feasible flow is not possible for the perturbed network. It is of interest to extend such robustness analysis to additional constraints on flow and in the presence of control.

3.1. Robustness of Electrical Networks

Consider the following setup motivated by electrical networks. To model the bidirectionality of electrical flow while continuing to adopt the directed graph formulation of network flow from Section 2, we do not constrain the entries of flow f to be nonnegative. In addition to Equation 2 (Kirchhoff's current law), the flow is constrained to also satisfy Ohm's law. Formally, $f \in \mathbf{R}^E$ is said to satisfy the physical constraints if there exist (voltage angles) $\phi \in \mathbf{R}^V$ such that $\Lambda = Bf$ and $f = \text{diag}(\mathcal{W})B^T\phi$, where $\Lambda \in \mathbf{R}^V$ is such that Λ_v is equal to λ or $-\lambda$ if v is the source or sink, respectively, and is equal to zero otherwise. This corresponds to interpreting f as a DC approximation to power flow, with \mathcal{W} being the negative of link susceptances. If \mathcal{G} is weakly connected, then such a unique f always exists and is given by

$$f(\mathcal{W}, \Lambda) = \text{diag}(\mathcal{W})B^T L^\dagger \Lambda, \quad 5.$$

where L^\dagger is the Moore–Penrose pseudoinverse of the weighted Laplacian $L := B\text{diag}(\mathcal{W})B^T$. Noting the linear dependence of $f(\mathcal{W}, \Lambda)$ on Λ , and hence on λ , one can easily see that, given c and \mathcal{W} , the maximum-flow problem in this setting can again be cast as a linear program. Let the solution be denoted as $\lambda^{\max}(\mathcal{W})$.

A generalization is when the link weights are flexible in a controlled manner: $\mathcal{W} \in [\underline{\mathcal{W}}, \overline{\mathcal{W}}]$, with $\overline{\mathcal{W}} \geq \underline{\mathcal{W}} \geq \mathbf{0}$. It is then of interest to study $\max_{\mathcal{W} \in [\underline{\mathcal{W}}, \overline{\mathcal{W}}]} \lambda^{\max}(\mathcal{W})$ and the corresponding argmax \mathcal{W} . The solution to this problem can be naturally interpreted in terms of robustness to perturbations of capacity under controllable link weights, in the same spirit as the max-flow min-cut theorem described above.

The technical challenge is due to the nonlinear dependence of f , and hence also of λ^{\max} , on \mathcal{W} . Ba & Savla (4) presented an incremental network reduction approach to reduce the complexity of this problem. In the classical network flow framework (see Section 2.5), replacing the network with a directed link from s to t whose capacity is the solution to the maximum-flow problem is equivalent from a capacity perspective. The notion of Thevenin equivalent resistance allows one to replace the electrical network having link-wise weights \mathcal{W} with an equivalent link with weight \mathcal{W}^{eq} . The equivalence then maps $[\underline{\mathcal{W}}, \overline{\mathcal{W}}] \subset \mathbf{R}_{\geq 0}^E$ to, say, $[\underline{\mathcal{W}}^{\text{eq}}, \overline{\mathcal{W}}^{\text{eq}}] \subset \mathbf{R}_{\geq 0}$, and it can be shown that $\underline{\mathcal{W}}$ and $\overline{\mathcal{W}}$ map to $\underline{\mathcal{W}}^{\text{eq}}$ and $\overline{\mathcal{W}}^{\text{eq}}$, respectively. The rewriting

$$\max_{\mathcal{W} \in [\underline{\mathcal{W}}, \overline{\mathcal{W}}]} \lambda^{\max}(\mathcal{W}) = \max_{\mathcal{W}^{\text{eq}} \in [\underline{\mathcal{W}}^{\text{eq}}, \overline{\mathcal{W}}^{\text{eq}}]} \lambda^{\max}(\mathcal{W}^{\text{eq}}), \quad 6.$$

with $\lambda^{\max}(\mathcal{W}^{\text{eq}}) := \max\{\lambda^{\max}(\mathcal{W}) : \mathcal{W} \in [\underline{\mathcal{W}}, \overline{\mathcal{W}}] \text{ s.t. the equivalent weight of } \mathcal{W} \text{ is } \mathcal{W}^{\text{eq}}\}$, suggests $\lambda^{\max}(\mathcal{W}^{\text{eq}})$ as the capacity associated with \mathcal{W}^{eq} . This notion of capacity lends itself to the following iterative solution of Equation 6 if \mathcal{G} is link reducible, i.e., if \mathcal{G} can be reduced to a single link by a sequence of series and parallel subnetwork reductions, as illustrated in **Figure 1**. Ba & Savla (4) showed that a particular quasi-concave property of capacity functions $\lambda^{\max}(\cdot)$ remains invariant across each of these reduction steps (see **Figure 1b**). Moreover, the interval over which the capacity function for a network achieves its maximum can be analytically related to the corresponding intervals of its subnetworks, thereby allowing an analytical solution to Equation 6. In other words, this procedure provides an analytical solution to the nonconvex problem in Equation 6 for link-reducible networks. For other \mathcal{G} , one can perform this reduction for each of the link-reducible subnetworks of \mathcal{G} , thereby reducing computational complexity (for an illustration, see **Figure 2**).

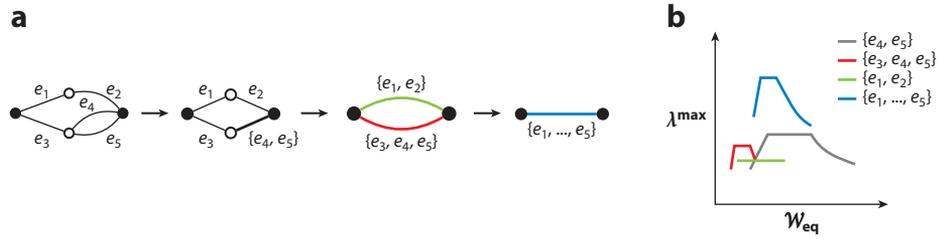


Figure 1

(a) Steps involved in incremental network reductions of a link-reducible graph between source and sink nodes (black circles). (b) Corresponding capacity functions. For example, the capacity function for the reduced network consisting of e_3 , e_4 , and e_5 from the original network is shown in red, while the capacity function for the entire network is shown in blue.

3.2. Future Research Directions

Extensions to other physical constraints, e.g., in AC power flow, natural gas, and water networks, along with relevant control actions, are natural directions to pursue. It is also of interest to pursue alternate techniques to tackle the resulting nonconvexity in capacity and robustness analysis. Several interesting properties of the optimization problem in Equation 6, such as sufficient conditions for the equivalence of local and global optimal solutions (as provided in chapter 3 of Reference 6), suggest possible directions.

4. ROBUSTNESS OF FINITE NETWORKS: LINEAR DYNAMICAL SETTING

We provide background on robustness analysis tools in Sections 4.1 and 4.2 and discuss application to network dynamics in Section 4.3 through an illustrative example.

4.1. Arbitrary Structured Uncertainties

The question of robust stability concerns whether a dynamical system maintains stability in the presence of a specified family of perturbations. A well-studied setting is illustrated in **Figure 3**,

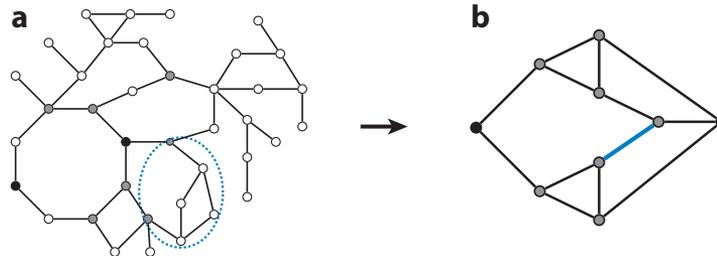


Figure 2

Network reduction for the IEEE 39 benchmark network (5) between source and sink nodes (black circles). Panel *a* shows the original network topology, and panel *b* shows the topology after reduction. The original nodes that remain in the reduced network are shown as gray circles. The sample link-reducible subnetwork enclosed by the dashed blue ellipse in panel *a* is reduced to the link shown in solid blue in panel *b*. Figure adapted from Reference 4.

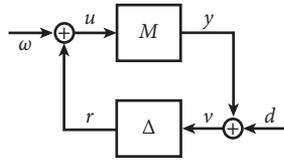


Figure 3

Robustness analysis configuration. The forward loop M is a known dynamical system. The feedback loop Δ is an unknown dynamical system that belongs to a family of potential candidates. The robustness analysis question is whether stability is maintained for all admissible Δ .

which represents the feedback equations

$$u = \omega + \Delta v =: \omega + r, \quad 7a.$$

$$v = d + Mu =: d + y. \quad 7b.$$

We assume that these equations are well posed in that for all $w, d \in \mathbf{L}_{2,e}^n$, there exist unique $u, v \in \mathbf{L}_{2,e}^n$ satisfying Equation 7. A further technical assumption is that the implied mapping $(w, d) \mapsto (u, v)$ is causal (7). Let $T[M, \Delta]$ denote the implied closed-loop input–output mapping, i.e.,

$$\begin{pmatrix} u \\ v \end{pmatrix} = T[M, \Delta] \begin{pmatrix} w \\ d \end{pmatrix}.$$

The setup in **Figure 3** is as follows:

- $M : \mathbf{L}_{2,e}^n \rightarrow \mathbf{L}_{2,e}^n$ is an input–output stable mapping satisfying

$$\|Mf\| \leq \alpha_M \|f\| + \beta_M, \forall f \in \mathbf{L}_{2,e}^n.$$

- $\Delta \in \mathbf{\Delta}$, where $\mathbf{\Delta}$ represents a family of input–output stable mappings satisfying

$$\mathbf{\Delta} = \{ \Delta : \mathbf{L}_{2,e}^n \rightarrow \mathbf{L}_{2,e}^n \text{ s.t. } \|\Delta f\| \leq \alpha_\delta \|f\| + \beta_\delta \}.$$

For a specific $\Delta \in \mathbf{\Delta}$, the feedback system shown in Equation 7 is closed-loop input–output stable if $T[M, \Delta]$ is input–output stable—i.e., there exist $\alpha, \beta \geq 0$ such that

$$\left\| \begin{pmatrix} u \\ v \end{pmatrix} \right\| = \left\| T[M, \Delta] \begin{pmatrix} w \\ d \end{pmatrix} \right\| \leq \alpha \left\| \begin{pmatrix} w \\ d \end{pmatrix} \right\| + \beta.$$

The feedback system in Equation 7 is robustly stable with respect to $\mathbf{\Delta}$ if $T[M, \Delta]$ is input–output stable for all $\Delta \in \mathbf{\Delta}$.

Our starting point is the classical small-gain theorem (8; for a recent survey, see 9).

Theorem 1. The closed-loop system shown in Equation 7 is robustly stable with respect to $\mathbf{\Delta}$ if $\alpha_M \alpha_\delta < 1$.

Note that the small-gain condition as stated is only a sufficient condition for robust stability. Also of interest is when the small-gain condition is necessary.

For this discussion, we restrict our attention to linear system models for the remainder of the section. A nonlinear setting where a small-gain condition is necessary is where M has fading

memory (10, 11). Define

$$\mathbf{\Delta}_{\text{LTI}} = \{ \Delta : \mathbf{L}_{2,e}^n \rightarrow \mathbf{L}_{2,e}^n \text{ s.t. } \Delta \text{ is LTI and } \|\Delta\| \leq 1 \}.$$

Theorem 2 (theorem 9.1 from Reference 3). Let M be an LTI system. The closed-loop system shown in Equation 7 is robustly stable with respect to $\mathbf{\Delta}_{\text{LTI}}$ if and only if $\|M\| < 1$.

The meaning of the small-gain condition being necessary is that if the small-gain condition is violated, i.e., $\|M\| \geq 1$, then there exists an admissible $\Delta \in \mathbf{\Delta}_{\text{LTI}}$ resulting in a closed loop $T[M, \Delta]$ that is not stable. An explicit construction is provided in theorem 9.1 of Reference 3. The general idea is as follows. Suppose that $M \sim \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right)$ and $\|M\| \geq 1$. Then, recalling Equation 4, there exists an ω^* such that

$$\sigma_{\max}(D + C(j\omega^*I - A)^{-1}B) \geq 1. \quad 8.$$

Define

$$\hat{M}(j\omega^*) = D + C(j\omega^*I - A)^{-1}B.$$

A consequence of Equation 8 is that there exists an $n \times n$ complex matrix Q with $\sigma_{\max}(Q) \leq 1$ such that

$$\det(I - Q\hat{M}(j\omega^*)) = 0. \quad 9.$$

Finally, one can construct an admissible $\Delta^* \in \mathbf{\Delta}$ with representation $\left(\begin{array}{c|c} A_\delta & B_\delta \\ \hline C_\delta & D_\delta \end{array} \right)$ such that

$$D_\delta + C_\delta(j\omega^*I - A_\delta)^{-1}B_\delta = Q.$$

The implication is that the closed-loop system will be unstable, in particular with a pole at $j\omega^*$, and Δ^* is a destabilizing perturbation.

The feedback configuration in **Figure 3** is constructed by isolating the effects of modeling errors. **Figure 4** shows an illustrative scenario. In this setup, P_o is a nominal plant model to be controlled by K_o . However, the family of possible plant models is $P_o + \Delta W$, where the system W acts as a dynamic weighting on the effects of modeling error $\Delta \in \mathbf{\Delta}_{\text{LTI}}$. Transforming **Figure 4** to **Figure 3** results in

$$M = -WK_o(I + P_oK_o)^{-1}.$$

Accordingly, a necessary and sufficient condition for robust stability is that

$$\|WK_o(I + P_oK_o)^{-1}\| < 1.$$

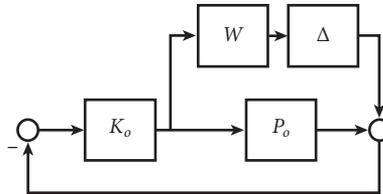


Figure 4

Additive error acting on a nominal plant model, P_o . The blocks P_o , K_o , and W are known dynamical systems. The block Δ represents a perturbation on the nominal P_o . The block W acts to normalize the impact of Δ . The robustness analysis question is whether stability is maintained for all admissible Δ .

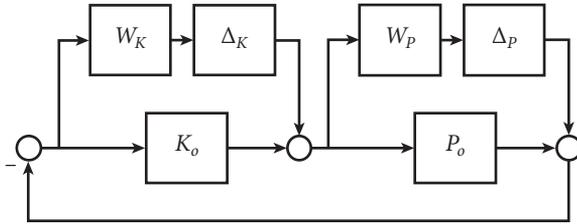


Figure 5

Additive error acting on both a nominal plant model, P_o , and nominal controller, K_o . The blocks P_o , K_o , W_P , and W_K are known dynamical systems. The blocks Δ_P and Δ_K represent perturbations on the nominal P_o and K_o , respectively. The blocks W_P and W_K act to normalize the impact of Δ_P and Δ_K , respectively. The robustness analysis question is whether stability is maintained for all admissible combined Δ_P and Δ_K .

Now consider the scenario illustrated in **Figure 5**. For convenience, let us assume that all mappings are single input/single output. As before, P_o is a nominal plant model with weighted additive error. However, there is now weighted additive error impacting the nominal controller, K_o , as well. Transforming **Figure 5** to the small-gain configuration of **Figure 3** results in

$$M = \begin{pmatrix} -W_P K_o (I + P_o K_o)^{-1} & W_P (I + K_o P_o)^{-1} \\ -W_K (I + P_o K_o)^{-1} & -W_K P_o (I + K_o P_o)^{-1} \end{pmatrix} \quad 10.$$

and

$$\Delta = \begin{pmatrix} \Delta_P & 0 \\ 0 & \Delta_K \end{pmatrix}. \quad 11.$$

We see that in the case where there is a graphical structure on the location of model perturbations, the resulting transformation to the form of **Figure 3** results in an associated restricted structure in Δ .

To analyze the consequences, define $\Delta_{LTI}^{\text{diag}} \subset \Delta_{LTI}$ to be the subset of diagonal stable LTI systems with norm less than one. While $\|M\| < 1$ remains a sufficient condition for robust stability with respect to $\Delta_{LTI}^{\text{diag}}$, it is no longer the case that this condition is necessary. As above, let us assume that $\|M\| \geq 1$ and that the condition shown in Equation 8 holds. Also as above, it is still possible to construct a complex matrix Q with $\sigma_{\max}(Q) \leq 1$ such that Equation 9 holds. However, this Q need not be diagonal, and so it may not be possible to construct an admissible destabilizing $\Delta \in \Delta_{LTI}^{\text{diag}}$ as above. In terms of controller synthesis, seeking $\|M\| < 1$ is unnecessarily restrictive.

The following special case of rank-one matrices illustrates the main idea. Define \mathbf{Q}^{full} to be the set of complex $n \times n$ matrices, and let $\mathbf{Q}^{\text{diag}} \subset \mathbf{Q}^{\text{full}}$ denote the subset of diagonal matrices.

Proposition 1. Let $X = ab^H$ be a rank-one matrix, defined by column vectors $a, b \in \mathbf{C}_n$. Then the following hold:

1. $\det(I - QX) \neq 0$ for all $Q \in \mathbf{Q}^{\text{full}}$ with $\sigma_{\max}(Q) \leq 1$ if and only if $\sigma_{\max}(X) = \|a\| \|b\| < 1$.
2. $\det(I - QX) \neq 0$ for all $Q \in \mathbf{Q}^{\text{diag}}$ with $\sigma_{\max}(Q) \leq 1$ if and only if $\sum_{i=1}^n |a_i| |b_i| < 1$.

Since in general

$$\sum_{i=1}^n |a_i| |b_i| \leq \|a\| \|b\|,$$

the restriction to diagonal Q matrices implies that larger X that fail the first condition in Proposition 1 can still satisfy the second. Indeed, depending on the specific structure of a and b , the difference can be significant.

These considerations led to the introduction of the structured singular value (12, 13; for an extensive discussion, see 14 and references therein). The definition is as follows. Let \mathbf{Q} denote a class of complex $n \times n$ matrices, where \mathbf{Q}^{full} and \mathbf{Q}^{diag} are specific examples. Now define²

$$\sigma_{\text{ssv}}(X; \mathbf{Q}) = \frac{1}{\inf\{\sigma_{\max}(Q) \text{ s.t. } Q \in \mathbf{Q} \text{ and } \det(I - QX) = 0\}}.$$

Note that $\sigma_{\text{ssv}}(\cdot; \mathbf{Q})$ must be defined with respect to a family of matrices. In the case of \mathbf{Q}^{full} , for any matrix X , the structured singular value defaults to the standard singular value, i.e.,

$$\sigma_{\text{ssv}}(X; \mathbf{Q}^{\text{full}}) = \sigma_{\max}(X).$$

By construction, the structured singular value can be used to derive necessary and sufficient conditions for robust stability in the case of structured perturbations. A lingering issue is its efficient computation. It is possible to compute an upper bound that is sometimes exact. The following result is representative.

Theorem 3 (from Reference 12). Let X be an $n \times n$ complex matrix. Then

$$\sigma_{\text{ssv}}(X; \mathbf{Q}^{\text{diag}}) \leq \inf_{D>0, \text{ diagonal}} \sigma_{\max}(DXD^{-1}).$$

Furthermore, in the case of $n \leq 3$, equality holds.

There is extensive literature addressing variations, extensions, and computational analysis of the main results of this section, including the case of slowly varying perturbations (15), time-varying perturbations (16, 17), and alternative system norms (18). Furthermore, the discussion above presented a small-gain approach, which is a special case of dissipativity analysis (19; for an extensive discussion of dissipativity for networked systems, see 20).

The main emphasis here is the implications of robustness analysis when there is an underlying graphical structure. To reinforce this point, consider now the scenario illustrated in **Figure 6**. This diagram models feedback control in the presence of modeling errors in the actuation (A) forward loop and sensing (S) feedback loop, both of which can be viewed as a consequence of controlling a system over a network with dynamic channels. Transforming **Figure 6** to the small-gain configuration of **Figure 3** results in

$$M = \begin{pmatrix} -W_A K_o P_o (I + K_o P_o)^{-1} & -W_A K_o (I + P_o K_o)^{-1} \\ W_S P_o (I + K_o P_o)^{-1} & -W_S P_o K_o (I + P_o K_o)^{-1} \end{pmatrix} \quad 12.$$

and

$$\Delta = \begin{pmatrix} \Delta_A & 0 \\ 0 & \Delta_S \end{pmatrix}. \quad 13.$$

In comparing the two scenarios of **Figures 5** and **6** when transformed to the standard configuration of **Figure 3**, we see that the analysis differs in comparing the resulting M in Equation 10 with that in Equation 12. The resulting Δ family in Equation 11 and that in Equation 13 are

² $\sigma_{\text{ssv}}(X; \mathbf{Q}) = 0$ whenever $\det(I - QX) = 0$ is not possible for any $Q \in \mathbf{Q}$.

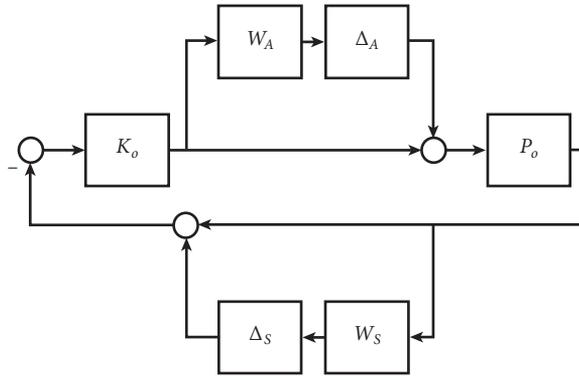


Figure 6

Modeling errors in the actuation and sensing channels. The blocks P_o , K_o , W_A , and W_S are known dynamical systems. The blocks Δ_A and Δ_S represent perturbations on (a) how the controller, K_o , acts on the plant, P_o , and (b) how the controller, K_o , measures the response of the plant, P_o , respectively. The blocks W_A and W_S act to normalize the impact of Δ_A and Δ_S , respectively. The robustness analysis question is whether stability is maintained for all admissible combined Δ_A and Δ_S .

effectively the same, despite the different origins. Accordingly, the robustness conclusions depend on the underlying graphical structure of where uncertainty enters into the overall system.

In closing this section, we also mention relatively later developments in robustness analysis that use the integral quadratic constraints framework (21). Enabled by advancements in related computational algorithms (e.g., see 22, 23), these techniques allow one to efficiently extend robustness analysis to several canonical nonlinearities and time-varying uncertainties in the Δ block.

4.2. Memoryless Stochastic Structured Uncertainties

Consider Equation 7 with d and ω as white second-order processes, and

$$\Delta(t) = \begin{bmatrix} \Delta_1(t) & & & \\ & \Delta_2(t) & & \\ & & \dots & \\ & & & \Delta_m(t) \end{bmatrix} \quad 14.$$

is a diagonal matrix of time-varying gains $\{\Delta_i\}_{i \in [m]}$, modeled as zero-mean random processes that are temporally independent but possibly mutually correlated (m is not necessarily the dimension of state-space realization of M in Equation 7). The entries of d can also possibly be mutually correlated, and so can the entries of ω . For simplicity, here we let all the quantities be real valued and let the covariance matrices of d and ω be time invariant (for extensions, see 24). In the above setting, the feedback system in Equation 7 (see **Figure 3**) is called mean-square (MS) stable if signals y and r have uniformly bounded variance sequences, i.e., if there exists a constant c such that

$$\max \{ \|\mathbb{E}[y(t)y^T(t)]\|_\infty, \|\mathbb{E}[r(t)r^T(t)]\|_\infty \} \leq c.$$

The next result (theorem 3.2 and section VIC of Reference 24) gives a tight condition on mean-square stability. In preparation for the result, let Σ_Δ and Σ_ω be, respectively, the covariance matrices of $[\Delta_1(t) \dots \Delta_m(t)]^T$ and ω , and let $\{M_i\}$ be the matrix-valued impulse response sequence of M .

Theorem 4. Let M be a stable (finite \mathcal{H}^2 norm), strictly causal, LTI system. The system defined by Equations 7 and 14 is mean-square stable if and only if $\rho(\mathbb{H}) < 1$, where the matrix-valued linear operator \mathbb{H} , also known as the loop-gain operator, is defined as

$$\mathbb{H}(X) := \Sigma_{\Delta} \circ \left(\sum_{t=0}^{\infty} M_t X M_t^T \right).$$

If $\rho(\mathbb{H}) > 1$ and Σ_{ω} is equal to the Perron eigenmatrix of \mathbb{H} , then the covariance $\mathbb{E}[u_t u_t^T]$ grows unbounded geometrically.

Section IVA of Reference 24 establishes that $\rho(\|M\|_2^2) < 1$ is also a necessary and sufficient condition for the mean-square stability of the system defined by Equations 7 and 14 when $\Sigma_{\Delta} = I$ (uncorrelated uncertainties), where $\|M\|_2^2$ is the matrix of squared \mathcal{H}^2 norms of subsystems of M . However, for correlated uncertainties, i.e., $\Sigma_{\Delta} \neq I$, section IVC of Reference 24 argues that such a condition involving only \mathcal{H}^2 norms of subsystems is not enough, and that one additionally needs other system metrics, such as the inner product between different subsystems' impulse responses.

The tight stochastic stability result in Theorem 4 is to be contrasted with techniques that approximate stochastic linear dynamics, in state-space form, using polynomial chaos expansions, and then project into a higher-dimensional deterministic linear dynamics. The standard Lyapunov argument is then used to analyze the stability of the approximate deterministic system (e.g., see 25, 26). On the other hand, these approximate techniques are applicable to more general parametric uncertainties.

Section VI of Reference 24 suggests that it might be convenient to compute $\rho(\mathbb{H})$ using state-space representation. If

$$M \sim \left(\begin{array}{c|c} A^{(0)} & B \\ \hline C & 0 \end{array} \right), \quad 15.$$

then $\rho(\mathbb{H})$ is equal to the largest real number λ such that the following linear matrix inequality has a feasible solution $X \geq 0$:

$$\lambda \left(X - A^{(0)} X A^{(0)T} \right) - B \left(\Sigma_{\Delta} \circ (C X C^T) \right) B^T = 0.$$

4.3. Application to Network Dynamics with Unreliable Links

The $M - \Delta$ framework in Section 4.1 and the associated small-gain techniques can be applied in the network context to handle specifics of where model perturbations occur in a feedback interconnection. An illuminating setup is in a state-space form such as

$$x(t+1) = \left(A^{(0)} + \sum_{j \in [m]} \Delta_j(t) A^{(j)} \right) x(t) + \omega(t), \quad 16.$$

where $A^{(0)}$ is interpreted as the nominal system description, and the parametric uncertainty in the nominal system is reflected by the m scalar quantities $\{\Delta_j\}_{j \in [m]}$. The structural knowledge about the uncertainty is contained in $\{A^{(j)}\}_{j \in [m]}$. The form of Equation 16 can be converted into Equation 7 with the structure of Δ as in Equation 14 (3).

In the setting of Section 4.2, $\{\Delta_j\}_{j \in [m]}$ are interpreted as multiplicative stochastic uncertainties. In this context, Equation 16 allows one to model network dynamics in the presence of unreliable

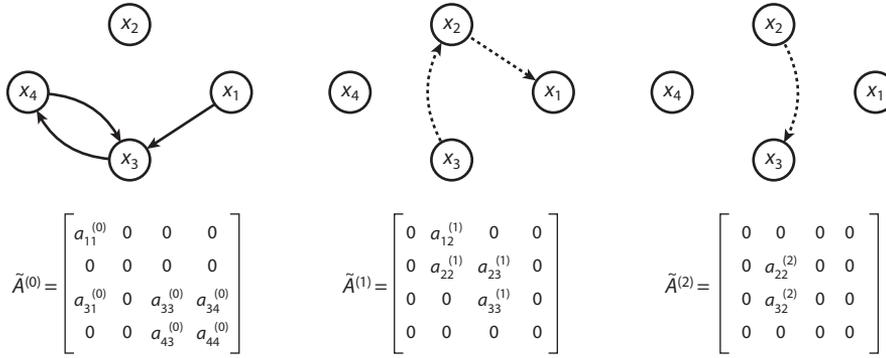


Figure 7

Illustration of the network dynamics in Equation 17. Reliable links are shown as solid lines, and unreliable links are shown as dashed lines.

links as follows. Let $x(t+1) = \tilde{A}^{(0)}x(t) + \omega(t)$ be the dynamics induced by reliable links. Let there be m links whose active status at t is given by binary variables $\{\gamma_j(t)\}_{j \in [m]}$, i.e., $\gamma_j(t) = 1$ if link j is active at time t and $\gamma_j(t) = 0$ otherwise. Let the unreliable links be associated with $\{\tilde{A}^{(j)}\}_{j \in [m]}$ (for an illustration, see **Figure 7**) such that the dynamics in the presence of reliable and unreliable links is

$$x(t+1) = \left(\tilde{A}^{(0)} + \sum_{j \in [m]} \gamma_j(t) \tilde{A}^{(j)} \right) x(t) + \omega(t). \quad 17.$$

Let $\Gamma(t) = [\gamma_1(t), \dots, \gamma_m(t)]^T$ be modeled as a multivariate Bernoulli random process that is temporally independent and has an identical distribution at all times. Let $\mathbb{E}[\Gamma(t)] \equiv \mu = [\mu_1, \dots, \mu_m]^T$, and let Σ_Γ denote the constant covariance. Let $\Delta_j(t) := \frac{\gamma_j(t) - \mu_j}{\sqrt{\Sigma_{\Gamma, jj}}}$, $j \in [m]$. Therefore, $\mathbb{E}[\Delta] = \mathbb{E}[(\Delta_1, \dots, \Delta_m)^T] = \mathbf{0}$, $\Sigma_{\Delta, jj} = 1$, for all $j \in [m]$, and Equation 17 can be rewritten in the form of Equation 16:

$$x(t+1) = \left(\underbrace{\tilde{A}^{(0)} + \sum_{j \in [m]} \mu_j \tilde{A}^{(j)}}_{A^{(0)}} + \sum_{j \in [m]} \Delta_j(t) \underbrace{\sqrt{\Sigma_{\Gamma, jj}} \tilde{A}^{(j)}}_{A^{(j)}} \right) x(t) + \omega(t).$$

Wang & Elia (27) used the above formulation to study the robustness of distributed averaging dynamics to link failures in the special case of diagonal Σ_Γ and diagonal Σ_ω , as follows. $\tilde{A}^{(0)}$ is equal to I . For every $j \in [m]$, $\tilde{A}^{(j)}$ is set as follows: Initialize $\tilde{A}^{(j)} = \mathbf{0}$, and for every link (i, k) associated with j , add $-e_i$ to the i th column and e_i to the k th column of $\tilde{A}^{(j)}$. Note that, in this case, all row sums of $A^{(0)}$ are equal to one, and, indeed, $\rho(A^{(0)}) = 1$. While Theorem 4 does not apply directly, Wang & Elia (27) provided a decomposition of the state into the conserved state and the deviation state. The dynamics of the deviation state is decoupled from that of the conserved state, and moreover, the spectral radius of its nominal dynamics is less than one and therefore amenable to Theorem 4.

4.4. Future Research Directions

We remarked at the beginning of Section 4.3 that the $M - \Delta$ framework can be used for tight robustness analysis of linear network dynamics in several deterministic settings. It would be

interesting to give a network-theoretic interpretation for the small-gain condition in these settings. Similarly, network-theoretic interpretations for the loop-gain operator and its Perron Frobenius eigenmatrix, which play an important role in characterizing the stochastic stability of linear network dynamics with multiplicative stochastic disturbances, would be of interest. The sufficient condition on Δ under which Theorem 4 holds true is to ensure a white-like property for the signals in the feedback loop. Such a property allows one to relate the covariance of input and output signals across blocks, analogously to Equation 3, and hence facilitate small-gain-type analysis. Extending the analysis to Δ , which is correlated in time, as also noted by Bamieh & Filo (24), would be a good step toward achieving the generality in Δ that is possible in the deterministic setting.

5. ROBUSTNESS OF FINITE NETWORKS: NONLINEAR DYNAMICAL SETTING

An interesting class of network dynamics is network flow dynamics. To formulate it, recall the notions of super-source and super-sink from Section 2.5. It is convenient to identify both of them with the same virtual node—say, 0. Add a virtual directed link $(0, v)$ from the virtual node to every source node $v \in \mathcal{V}$, and let this set be denoted as \mathcal{E}^{in} . Similarly, add a virtual directed link $(v, 0)$ from every sink node $v \in \mathcal{V}$ to the virtual node, and let this set be denoted as \mathcal{E}^{out} . Network flow dynamics corresponds to mass conservation:

$$\dot{x}_i = \underbrace{\sum_{j \in \mathcal{E} \cup \mathcal{E}^{\text{in}}} z_{ji}(x)}_{\text{inflow}} - \underbrace{\sum_{j \in \mathcal{E} \cup \mathcal{E}^{\text{out}}} z_{ij}(x)}_{\text{outflow}}, \quad i \in \mathcal{E}, \quad 18.$$

where $x_i \geq 0$ denotes the mass on link $i \in \mathcal{E}$, and $z_{ij} \geq 0$ is the flow from link $i \in \mathcal{E} \cup \mathcal{E}^{\text{in}}$ to link $j \in \mathcal{E} \cup \mathcal{E}^{\text{out}}$. Naturally, z_{ij} is zero if i is not incident to j . The z_{ij} 's depend on the state x , where this dependence also incorporates feedback control; the exact model depends on the setting. In all these settings, z_{ij} 's are constrained to be such that $z_{ij}(x) = 0$ for all j if $x_i = 0$, and $\sum_{j \in \mathcal{E} \cup \mathcal{E}^{\text{out}}} z_{ij}(x) = \lambda_i$ if $i \in \mathcal{E}^{\text{in}}$. The first constraint ensures that Equation 18 is a positive system, i.e., $x(t) \geq \mathbf{0}$ for all t , and the second constraint specifies the total external flow coming into source nodes.

It is natural to consider the decomposition $z_{ij}(x) = f_i(x)R_{ij}(x)$, where $f_i(x) \geq 0$ is the net outflow from $i \in \mathcal{E} \cup \mathcal{E}^{\text{in}}$ satisfying $f_i(x) = 0$ if $x_i = 0$ for all i and $f_i(x) \equiv \lambda_i$ for all $i \in \mathcal{E}^{\text{in}}$, and $R_{ij}(x) \in [0, 1]$ is the routing function satisfying $R_{ij}(x) \equiv 0$ if i is not incident to j and $\sum_j R_{ij}(x) \equiv 1$ for all i . Under these assumptions, Equation 18 can be written in vector form as

$$\dot{x} = (R^T(x) - I) f(x) + \lambda. \quad 19.$$

A simple linear version of Equation 19 is obtained by assuming a constant routing matrix, i.e., $R(x) \equiv R$, and letting the outflow from i be linearly increasing in x_i , i.e., $f_i(x) \equiv f_i(x_i) = b_i x_i$ for $b_i > 0$. Let H be the diagonal matrix whose entries are $\{b_i\}_{i \in \mathcal{E}}$. Since R is fixed, we can let the external flow arrive directly onto the links outgoing from the source nodes. Therefore, it is sufficient to consider Equation 19 restricted to links in \mathcal{E} as

$$\dot{x} = (R^T - I) Hx + \lambda, \quad 20.$$

where we use the same notation as in Equation 19 for brevity. The invertibility of $R^T - I$ follows from the connectivity of the underlying graph and the fact that R (restricted to \mathcal{E}) is row substochastic, with the entries of at least one row, corresponding to a sink, adding to strictly less than

one. Therefore, Equation 20 admits a unique equilibrium $x^* = H^{-1}(I - R^T)^{-1}\lambda$, whose stability analysis is straightforward.

A key practical consideration for network flow dynamics is to include link-wise capacity constraints by saturating link-wise outflows at $\{c_i\}_{i \in \mathcal{E}}$ (see Section 2.5). Introducing such a saturation in Equation 20 leads to a piecewise affine system, which can be written in the state-space form as

$$\begin{aligned} \dot{x} &= \underbrace{(R^T - I)H}_{A} x + \underbrace{(I - R^T)}_B u + \lambda, \\ y &= \underbrace{H}_C x, \\ u(t) &= \max\{0, y(t) - c\}, \end{aligned} \tag{21}$$

where c is the vector of link-wise capacities and \max is element-wise. Equation 21, which can be converted to a piecewise linear system through a simple change of variable along with a corresponding change of c to, say, \tilde{c} , can be interpreted as an LTI system in feedback with on-off nonlinearities (see **Figure 8**). Sufficient conditions for global stability and robustness analysis of piecewise linear systems have been developed (e.g., see 28 and references therein). To the best of our knowledge, these tools have not been applied to the specific piecewise linear system in Equation 21, but such an analysis will potentially face a few challenges. First, it remains to be seen how the sufficient condition for global asymptotic stability compares with the necessary condition given by the max-flow min-cut theorem in the single-sink setting. Second, the number of switching surfaces in Equation 21 grows exponentially with the number of links, thereby making the analysis computationally intense. Third, it is not clear how the analysis can be extended to other interesting nonlinear instances of Equation 18. In the remainder of this section, we present alternate nonlinear techniques for stability analysis of Equation 18 and robustness to perturbations in $\{c_i\}_{i \in \mathcal{E}}$, under a few setups for $\{z_{ij}(x)\} \in \mathcal{Z}$. In Section 5.1, we discuss how information constraint reduces robustness with respect to centralized control, and in Section 5.2, we discuss a setup where additional control action compensates for information constraint to maintain the same robustness as under centralized control.

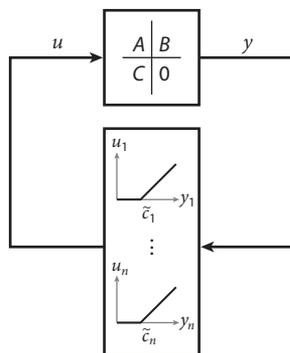


Figure 8

A linear time-invariant system in feedback with on-off nonlinearities as a simple abstraction for network flow dynamics under link-wise capacity constraints.

5.1. Robustness Under Information Constraint

We discuss two setups to illustrate different ways in which information constraint on the controller can affect network robustness. In the setup in Section 5.1.1, information constraint results in a loss of robustness, whereas there is no such loss in the setup in Section 5.1.2.

5.1.1. Loss of robustness under decentralized control. We return to the setup of Equation 19. We let $f_i(x) \equiv f_i(x_i)$ be strictly increasing and have capacity constraint $f_i(x_i) \leq c_i$ for all $i \in \mathcal{E}$. We simplify by letting $R_{ji}(x) \equiv R_{ki}(x) \equiv R_i(x)$ for any two links j and k incident to i —i.e., each node adds flow from immediate upstream links and instantaneously routes it among its immediate downstream links, without consideration for the source or destination of that flow. This naturally makes sense for a single-commodity flow. For simplicity in presentation, we assume a single source and single sink. To avoid triviality, we assume that, for every node $v \in \mathcal{V}$, there exists a directed path from the source node to v and from v to the sink node. For a single-commodity flow, extension to multiple sources and sinks follows from the construction in Section 2.5. Correspondingly, we shall remove the subscript and let λ denote the external flow to the only source node.

In the current setting, the stability and robustness properties of the network flow dynamics depend on the choice of routing functions R . Of particular interest is decentralized routing, where the routing control action for a link depends only on the state on i and on links having the same tail node as i , i.e., $R_i(x) \equiv R_i(\{x_j : j \text{ and } i \text{ have the same tail node}\})$. Consider a feasible flow f^* , i.e., satisfying the flow conservation and capacity constraints, and correspondingly $\{x_i^* := f_i^{-1}(f_i^*)\}_{i \in \mathcal{E}}$. Given such x^* and f^* , which exist if and only if λ is less than the minimum-cut capacity of the network, consider decentralized routing control satisfying (a) $R_i(x^*) = \frac{f_i^*}{\sum_j f_j^*}$ for all $i \in \mathcal{E}$, where the summation is over all links j that have the same tail node as i , and (b) $\frac{\partial R_i(x)}{\partial x_j} > 0$ for every two distinct links i and j having the same tail node. Item *a* implies that x^* is an equilibrium; item *b* implies that, if x_j increases but the states on other links that affect R_i remain the same, then more flow will be routed to every other link. This property helps to establish the global asymptotic stability of x^* if \mathcal{G} is acyclic, because item *b*, along with the increasing nature of $f_i(\cdot)$, implies that the Jacobian of Equation 18 under these conditions is a compartmental matrix, i.e., a matrix whose off-diagonal components are nonnegative and column sums are nonpositive. Such systems are known to possess a contraction principle (e.g., see 29).

The (decentralized) routing satisfying item *a* and, particularly, item *b* above is referred to as monotone routing. An example of monotone routing is

$$R_i(x) = \frac{f_i^* e^{\beta(x_i^* - x_i)}}{\sum_j f_j^* \exp(\beta(x_j^* - x_j))}, \quad \beta > 0,$$

which is the multinomial logit model at node v for discrete choice among links in \mathcal{E}_v^+ , when the utility function associated with $i \in \mathcal{E}_v^+$ is $x_i^* - x_i + \log f_i^*$. Indeed, monotone routing is interpreted in Reference 30 as an en route driver decision, with x^* being the nominal network state expected by the drivers during the trip. This nominal state is updated by the driver population at a slower timescale according to a best-response rule (e.g., see 31). It was shown in Reference 30 that the link (out)flows under such a multiscale update rule converge to a Wardrop equilibrium if \mathcal{G} is acyclic and if each $f_i(\cdot)$, $i \in \mathcal{E}$, is strictly concave in addition to being strictly increasing.

It was shown in References 32 and 33 that the monotone routing policies admit the following sharp characterization of the robustness of equilibrium x^* to a perturbation in capacity. Let $\Delta \in [0, c]$ denote the perturbation in capacity. Clearly, there exists a $\|\Delta\|_1$ that is (infinitesimally) greater

than the network residual capacity, such that at least one component of $x(t)$ grows unbounded for every (not necessarily decentralized) routing policy. Such a Δ , e.g., corresponds to reducing capacities on the links outgoing from the minimum cut by infinitesimally more than the residual capacities on them. However, if the routing policy is decentralized, as in monotone routing, then reducing capacities on the links outgoing from the node that has the smallest residual capacity by an amount infinitesimally greater than their respective residual capacities is sufficient to ensure that $x_i(t)$ grows unbounded at least for one $i \in \mathcal{E}$. Interestingly, under the monotone routing, if \mathcal{G} is acyclic, then for every $\|\Delta\|_1$ less than the minimum node residual capacity, there exists a globally asymptotically stable equilibrium, which is different from the equilibrium x^* associated with the nominal (i.e., nonperturbed) system. The existence of a new equilibrium was established in Reference 33 by a novel use of the monotonicity property of the underlying dynamics, and the stability then follows from the contraction principle, along the same lines as the stability of nominal dynamics.

In summary, if one interprets the minimum node residual capacity to be the robustness under decentralized routing, then this is in general strictly less than the network residual capacity, which is the robustness under a centralized routing policy. Section 5.2 discusses how allowing additional control action in this setting can prevent such a loss. We next discuss a setup where decentralized control does not cause a loss of robustness.

5.1.2. No loss of robustness under decentralized control. Let us fix the routing matrix; therefore, the decoupling of z_{ij} from Section 5.1.1 becomes $z_{ij}(x) = f_i(x)R_{ij}$. This allows us to consider multiple commodity flow between multiple sources and sinks. We remove \mathcal{E}^{in} and let the external flow arrive directly onto the links outgoing from the source nodes, similarly to Equation 20. In summary, in the current setting, Equation 18 becomes

$$\dot{x} = (R^T - I) f(x) + \lambda, \quad 22.$$

where $\lambda \in \mathbf{R}_{\geq 0}^{\mathcal{E}}$ is the external inflows at the links, and, with a slight abuse of notation, R denotes the routing submatrix restricted to links in \mathcal{E} . To avoid triviality, we assume that for every node $v \in \mathcal{V}$, there exist a directed path from at least one source node to v and a directed path from v to at least one sink node.

Motivated by scenarios involving allocation of a fixed service resource to multiple conflicting queues, e.g., at a signalized traffic intersection in an urban traffic network or at a router in a communication network, consider the following simple model for link outflows:

$$f_i(x) \equiv f_i(x_i) = \begin{cases} u_i(x)c_i & \text{if } x_i > 0 \\ \min\{u_i(x)c_i, \sum_j R_{ji}f_j(x)\} & \text{if } x_i = 0 \end{cases} \quad i \in \mathcal{E}, \quad 23.$$

where the summation is over all links $j \in \mathcal{E} \cup \mathcal{E}^{\text{in}}$ that are incident to i . The control input $u_i(x) \in [0, 1]$ can be interpreted as the fraction of times that the queue on i is serviced. Equation 23 ensures the positivity of Equation 22 in the current setting. The fixed resource feature is modeled by imposing

$$\sum_{i \in \mathcal{E}_v^-} u_i(x) \leq 1, \quad v \in \mathcal{V}. \quad 24.$$

By further imposing $u_i(x) \rightarrow 0^+$ as $x_i \rightarrow 0^+$ (and starting from an initial condition $x_0 > \mathbf{0}$), one can ensure that the $x_i = 0$ case in Equation 23 is never active, thereby also avoiding regularity issues with the right-hand side of Equation 22.

A necessary condition for the boundedness of $x(t)$, and hence also for the existence of equilibrium in the current setting, is $\sum_{i \in \mathcal{E}_v^-} \frac{f_i^*}{c_i} \leq 1$ for all $v \in \mathcal{V}$, where f^* is the flow induced by λ and R , i.e., $\lambda + (R^T - I)f^* = 0$, which implies $f^* = (I - R^T)^{-1}\lambda$. It is of interest to study stability conditions under natural controllers $u(x)$ and compare with the necessary condition. It is straightforward to see that, if the necessary condition is satisfied, then the open-loop controller $u_i(x) \equiv f_i^*/c_i$, $i \in \mathcal{E}$, satisfies Equation 24, and, under this condition, every $x > \mathbf{0}$ is an equilibrium of Equation 22. Besides being open loop, this controller requires information about λ and R and is therefore centralized. As in Section 5.1.1, it is of interest to study decentralized controllers of the form $u_i(x) \equiv u_i(x_j : j \text{ has the same head node as } i)$. Using monotonicity arguments similar to those in Section 5.1.1, Savla et al. (34) showed that, if \mathcal{G} is acyclic and if the necessary condition is satisfied, then under the following controller, there exists a unique $x^* > \mathbf{0}$ whose basin of attraction is $\mathbf{R}_{>0}^{\mathcal{E}}$:

$$u_i(x) = \frac{x_i}{\sum_j x_j + \beta}, \quad \beta > 0, \quad 25.$$

where the summation is over all j with the same head node as i . Extension of this result to cyclic \mathcal{G} is possible by establishing $\sum_{i \in \mathcal{E}} x_i \log(u_i(x) \frac{c_i}{f_i^*})$ as a Lyapunov function. The inspiration for this analysis comes from the literature on stochastic networks (e.g., see 35), where Equation 25 is interpreted as a proportionally fair controller for the current setting. It is remarkable to note that the decentralized controller in Equation 25 gives the same stability guarantee as the centralized (open-loop) controller without requiring information about (even local) λ , R , or c . The literature on back-pressure controllers (e.g., see 36) suggests that one can get the same stability guarantee by establishing $\sum_{i \in \mathcal{E}} x_i^2$ as a Lyapunov function. Such tight stability guarantees translate into tight guarantees on robustness to perturbation in link capacities. That is, for a perturbation in capacity, if there exists a centralized controller under which $x(t)$ remains bounded, then it also does so under the above decentralized controllers. Hence, there is no loss of robustness under decentralized control.

5.2. Compensating Information Constraint with Control Action for Robustness

Consider again the single-commodity-flow setting, with a single source and a single sink, from Section 5.1.1. Recall that, in Section 5.1.1, the outflow from a link was uncontrolled and determined by $\{f_i(x_i)\}_{i \in \mathcal{E}}$. Let us relax this feature but maintain the decentralization aspect. That is, we let $z_{ij}(x)$ be controlled and let $z_{ij}(x) \equiv z_{ij}(\{x_k : k = i \text{ or } k \text{ has the same tail node as } j\})$. An implication of the relaxation is that we extend the scope of control to schedule the outflow from a link in addition to determining the routing part. A natural extension of the monotonicity property from Section 5.1.1 to the current setting is to consider decentralized $\{z_{ij}\}$ that satisfy the following for all $i \in \mathcal{E} \cup \mathcal{E}^{\text{in}}$ and $j \in \mathcal{E} \cup \mathcal{E}^{\text{out}}$: (a) $\frac{\partial z_{ij}(x)}{\partial x_k} > 0$ for $k = i$ and for all k , except j , to which i is incident, and (b) $\frac{\partial \sum_j z_{ij}(x)}{\partial x_k} < 0$ for all k to which i is incident. These properties imply that, if x_k increases but the state on the other links that affect z_{ik} remains the same, then (a) more flow is sent to links other than k to which i is incident and (b) the total outflow $\sum_j z_{ij}(x)$ from link i decreases. With a slight abuse of terminology, we call the control policies $\{z_{ij}(x)\}$ satisfying these two properties monotone control policies (in contrast to the monotone routing policies in Section 5.1.1). An example of monotone control is

$$z_{ij}(x) = c_i (1 - e^{-\beta x_i}) \frac{e^{-\beta x_j}}{\sum_k e^{-\beta x_k}}, \quad \beta > 0,$$

where the summation is over $k = i$ and all k to which i is incident.

Through the use of tools similar to those described in Section 5.1.1, it was shown in Reference 37 that, under monotone control, if λ is less than the minimum-cut capacity, then there exists a globally asymptotically stable equilibrium x^* . An implication of this is that, for all perturbations to capacity $\|\Delta\|_1$ less than the network residual capacity, there exists a globally asymptotically stable (new) equilibrium under monotone control. Combining this with the fact that there exists Δ with $\|\Delta\|_1$ (infinitesimally) greater than the network residual capacity such that no controller (not necessarily decentralized) can prevent $x(t)$ from growing unbounded on at least one component implies that monotone control policies are maximally robust to perturbation in capacity among all (not necessarily decentralized) controllers. This robustness guarantee is stronger than the one on monotone routing policy in Section 5.1, which further required an acyclic assumption on \mathcal{G} . Moreover, the monotone routing is parameterized by equilibrium x^* and hence requires some kind of centralized coordination, whereas no such coordination is required in the current monotone control setting. The additional control action in the current setting in the form of scheduling, in conjunction with property b in the definition of monotone control given above, allows for forward and backward propagation of dynamics. Interestingly, such propagations are sufficient to compensate for the decentralization feature of the controller to give robustness performance as good as that of a centralized controller. On the other hand, the lack of scheduling action in the setting of Section 5.1 allows only forward propagation of dynamics, resulting in a gap in robustness guarantee with respect to a centralized controller.

A few extensions were also considered in Reference 37, the most notable being finite storage capacity on the links, i.e., $x(t) \leq B \in \mathbf{R}_{>0}^{\mathcal{E}}$. This work showed that, if λ is less than the minimum-cut capacity, which is a necessary condition for the boundedness of $x(t)$, there exists a globally asymptotic stable equilibrium under appropriate modification of the monotone controller. They also showed that, under such a modified controller, if the necessary condition does not hold true, there exists a cut such that all links outgoing from that cut hit their respective storage capacities (which are not necessarily equal) simultaneously. This possibly suggests graceful degradation in the unstable regime, i.e., maximizing the time until any link hits the storage capacity, which could cause link failure.

5.3. Future Research Directions

The discussion in the beginning of this section suggests a lack of understanding of the connection between analysis techniques for piecewise linear systems, which potentially are computationally intense and conservative, and the nonlinear analysis techniques that are proving to be effective in specific instances of nonlinear network dynamics. Comparing the techniques in these specific cases could be a good first step toward understanding the connection. It is expected to be challenging, however, to treat structural constraints such as distributed information and control actions in a similar fashion. Developing specialized tools to analyze the impact of such constraints on network robustness could be of independent interest. Some specific future directions along these lines are robustness analysis under general (e.g., multihop) information constraint in the setting of Section 5.1.1 and generalizing the substitutability between control actions and global information beyond the setup in Section 5.2.

6. ROBUSTNESS UNDER CASCADING FAILURE

The notion of capacity was introduced in Sections 3 and 5 by thresholding the link (out)flows. The capacities potentially define the boundary to failure of the corresponding link. The failure upon crossing the boundary implies a structural change in \mathcal{G} . The complete description then necessitates

specifying the jump map under such a structural change, i.e., the network flow dynamics immediately after the discontinuity induced by the failure. The jump could thereafter result in crossing of the capacity for some other link, potentially leading to a series of failures—a phenomenon known as cascading failure. Notationally, this is denoted as $\{\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t), \mathcal{W}(t)) : t = 0, 1, 2, \dots\}$, with $\mathcal{E}(0) \supset \mathcal{E}(1) \supset \dots$, where $\mathcal{G}(0)$ is the nominal graph structure.

Given the finite size of $\mathcal{G}(0)$, the process of cascading failure terminates after finite steps, either at a subnetwork with link-wise flows less than the respective link-wise capacities or at a subnetwork in which there is no directed path from the source node to the sink node. The former scenario corresponds to a new equilibrium for the network structure, and the latter corresponds to network failure. It is of interest to tightly characterize the set of perturbations under which the cascading failure terminates at a new equilibrium network structure. In Sections 6.1 and 6.2, we consider jumps governed by control and physical constraints on flow, along the lines of Sections 3 and 5. We then discuss a representative contagion model for cascading failure and present brief remarks at the end of the section on how it contrasts with the models in Sections 6.1 and 6.2.

6.1. Electrical Networks

Consider the setup of electrical networks from Section 3 with fixed link weights \mathcal{W} . The failure of a link—say, i —corresponds to setting $\mathcal{W}_i = 0$. The network flow after such a jump is again given by Equation 5 by setting the entries of \mathcal{W} , corresponding to the links that have failed, to be zero. A necessary and sufficient condition for cascading failure to terminate at t is $\lambda < \lambda^{\max}(\mathcal{W}(t))$ [for the definition of $\lambda^{\max}(\mathcal{W})$, see Section 3]. Therefore, if $\lambda^{\max}(\mathcal{W}(1)) \geq \lambda^{\max}(\mathcal{W}(2)) \geq \dots$, then a Δ causes network failure if and only if $\lambda > \lambda^{\max}(\mathcal{W}(1))$. However, $\{\lambda^{\max}(\mathcal{W}(t))\}_{t \in \mathbb{N}}$ is not monotonically decreasing in general. Therefore, in general, a Δ causes network failure if and only if $\lambda > \max_{t \in \mathbb{N}} \lambda^{\max}(\mathcal{W}(t))$. Due to the multistage nature of computing $\max_{t \in \mathbb{N}} \lambda^{\max}(\mathcal{W}(t))$, giving a tight characterization of the set of all Δ leading to network failure in the general case is computationally challenging.

Let us now extend the robustness analysis to the case where the above jump map is also influenced by control actions. Let us consider \mathcal{W} to be the control action, as in Section 3, and let $\mathcal{W}^* \in \operatorname{argmax}_{\mathcal{W}(0) \in [\underline{\mathcal{W}}, \overline{\mathcal{W}}]} \lambda^{\max}(\mathcal{W}(0))$ (a problem discussed in Section 3). If $\underline{\mathcal{W}} = \mathbf{0}$, then, with $\mathcal{W}(0) = \mathcal{W}^*$,

$$\lambda^{\max}(\mathcal{W}^*) \geq \lambda^{\max}(\mathcal{W}(1)) \geq \lambda^{\max}(\mathcal{W}(2)) \geq \dots,$$

where the inequalities follow from the fact that $\underline{\mathcal{W}} = \mathbf{0}$ implies that the choice of \mathcal{W}^* includes optimizing over all possible subnetworks of $\mathcal{G}(0)$ (recall that setting $\mathcal{W}_i = 0$ is equivalent to removing link i). The above monotonically decreasing relationship implies not only that \mathcal{W}^* is maximally robust but also that it is straightforward to tightly quantify its robustness. Formally, a Δ that removes the residual capacity $c_i - |f_i(\mathcal{W}^*, \Lambda)|$ from the link i for which this value is a minimum and does not perturb other links has the smallest $\|\Delta\|_1$ among all Δ that cause network failure. On the other hand, for all $\|\Delta\|_1 < \min_{i \in \mathcal{E}} c_i - |f_i(\mathcal{W}^*, \Lambda)|$, there is no link failure and hence no network failure.

An alternate setting is that of load-shedding control. In this setting, λ is controlled so that $\lambda(0) \geq \lambda(1) \geq \dots$, and $\mathcal{W}(t)$ is determined by the resulting cascading failure process, i.e., by setting entries of \mathcal{W} to zero for links that fail. To appreciate the effect of controlling λ , let us revisit the general scenario in the uncontrolled case, illustrated in **Figure 9a**. Every Δ such that $\lambda_0 := \lambda > \max_{t \in \mathbb{N}} \lambda^{\max}(\mathcal{W}(t))$ causes network failure. However, if one could choose $\lambda(k) < \max_{t \in \mathbb{N}} \lambda^{\max}(\mathcal{W}(t)) < \lambda$ for all $k \geq 1$, then the cascading process will terminate before network failure. Since λ is typically construed as a measure of network performance, this illustrates the

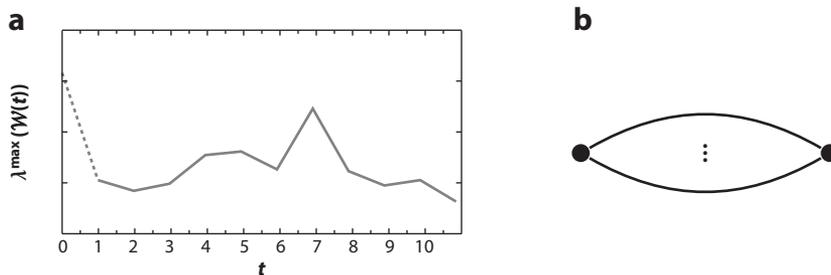


Figure 9

(a) Nonmonotonicity of $\lambda^{\max}(\mathcal{W}(t))$ of an electrical network during the cascading failure process. (b) A two-node network.

trade-off between performance and robustness. This trade-off is more pronounced when there is an additional requirement to terminate the cascading process within a given horizon. For example, in **Figure 9a**, if this horizon is 6, then the best performance is achieved for $\lambda(t) = \lambda^{\max}(\mathcal{W}(5))$ for all $t = 1, \dots, 6$, which is strictly less than $\max_{t \in \mathbb{N}} \lambda^{\max}(\mathcal{W}(t))$.

When selecting the best $\{\lambda(t)\}_t$ in the above example, we implicitly assumed that $\{\lambda^{\max}(\mathcal{W}(t))\}_t$ is independent of $\{\lambda(t)\}_t$. However, in general, this is true only for two-node networks (see **Figure 9b**). This is because, in general, $\lambda(k)$ affects the sequence $\{\mathcal{W}(t)\}_{t \geq k}$ and hence affects $\{\lambda^{\max}(\mathcal{W}(t))\}_{t \geq k}$. Therefore, computing optimal $\{\lambda(t)\}_t$ for the general case is challenging. Ba & Savla (38) presented a computational approach to solve this problem. Specifically, they constructed a finite abstraction in the form of a decision tree with states $(\lambda(t), \mathcal{W}(t))$, rooted at the initial condition $(\lambda(0), \mathcal{W}(0))$, with leaf nodes being desirable termination states of the cascade process. Edges correspond to one step of the cascade process. The edge costs are zero except if the edge is incoming into a leaf node, in which case the cost is equal to the negative of the value of λ associated with the leaf node. Among all the paths from the root to a leaf node, the one with the least cost gives an optimal $\{\lambda(t)\}_t$. This methodology applies to the multiple-sink, multiple-source scenario and is particularly well suited for the finite-horizon termination constraint.

6.2. Transport Networks

Let us now revisit the setup of transport networks from Sections 5.1.1 and 5.2, with the network flow dynamics immediately after failure at t determined by the routing policy restricted to residual graph $\mathcal{G}(t)$. In the centralized and static setting, a policy that routes over $\mathcal{G}(0)$ according to a feasible flow is maximally robust, with the network residual capacity of $\mathcal{G}(0)$ being the measure of robustness. This is because the minimum-cut capacity of a network is no less than that of any of its subnetworks. Analyzing the robustness of decentralized routing under hybrid dynamics obtained from the combination of flow dynamics from Sections 5.1.1 or 5.2 over a given network with dynamics of network structure under cascading failure is challenging. Savla et al. (39) considered the intermediate quasi-static regime, where, upon a jump, the flow equilibrates on every link instantaneously such that the flow ratio at every node is as determined by the decentralized routing policy implemented at that node. They provided an algorithm to synthesize decentralized routing policies that are maximally robust in this setting to perturbation processes $\Delta(t)$, where $\Delta(t)$ is the cumulative reduction in capacity on the links until time t , and $\Delta(t)$ is nondecreasing in time. [The robustness results in Section 5 extend to such a perturbation process by replacing $\|\Delta\|_1$ there with $\|\Delta\|_{1,\infty} := \lim_{T \rightarrow +\infty} \|\Delta(T)\|_1$.]

The algorithm of Savla et al. (39) implements a dynamic-programming-like computation over space backward from the sink node to the source node. At each stage, the algorithm computes the robust decentralized routing policy at a given node and the associated measure of robustness for the subnetwork downstream of that node and passes this information to its parent node. The routing policies synthesized by this algorithm, which performs only one iteration per node, can be shown to be maximally robust within the class of decentralized routing policies if the flow induced by it is monotonically nondecreasing on every active link during the cascading failure process. A simple sufficient condition for this is that the network is laterally symmetric about the source–sink pair in terms of both network structure and link capacities. Savla et al. (39) also provided a catalog of basis networks as well as rules to compose them into bigger networks, which are not necessarily laterally symmetric about the source–sink pair and yet ensure the monotone nondecreasing condition on flow.

6.3. Contagion

Consider the following threshold-based cascade model from Blume et al. (40), which has been studied extensively in the context of social and biological contagions. Let each node v choose a threshold $\ell(v)$ independently from a distribution μ on \mathbf{N} , where $\ell(v)$ represents the number of failed neighbors that v can withstand before v fails as well. The failure process works as follows. First, declare all nodes with a threshold of zero to have failed. Then, repeatedly check whether any node v that has not yet failed has at least $\ell(v)$ failed neighbors; if so, declare v to have failed as well and continue iterating. In this model, μ can be thought of as determining the distribution of levels of health of the nodes, hence implicitly controlling the way the failure process spreads.

For a given node v , let its failure probability be denoted as $r_\mu(v)$. Let $r_\mu^* = \sup_{v \in \mathcal{V}} r_\mu(v)$ be the maximum failure probability in \mathcal{G} . Here, r_μ^* can be viewed as a measure of robustness against cascading failures that operate under the threshold distribution μ and is also referred to as the μ risk of \mathcal{G} . It is of interest to understand the relationship between μ risk and the structure of the underlying graph.

While characterizing μ risk is challenging in general, certain graph classes lend themselves to useful insights. Within the class of d -regular graphs, denoted as \mathbb{G}_d , contrasting the clique K_{d+1} and the infinite complete d -ary tree T_d (for examples, see **Figure 10**) shows that the μ -risk-minimizing graph structure depends on the distribution μ . However, at each μ with $0 < \mu(0) < 1$, at least one of K_3 or T_2 achieves strictly lower μ risk than every other graph in $\mathbb{G}_2 \setminus \{K_3, T_2\}$. The behavior of μ risk on \mathbb{G}_d for $d > 2$ is complicated: (a) There are distributions μ for which K_{d+1} has strictly

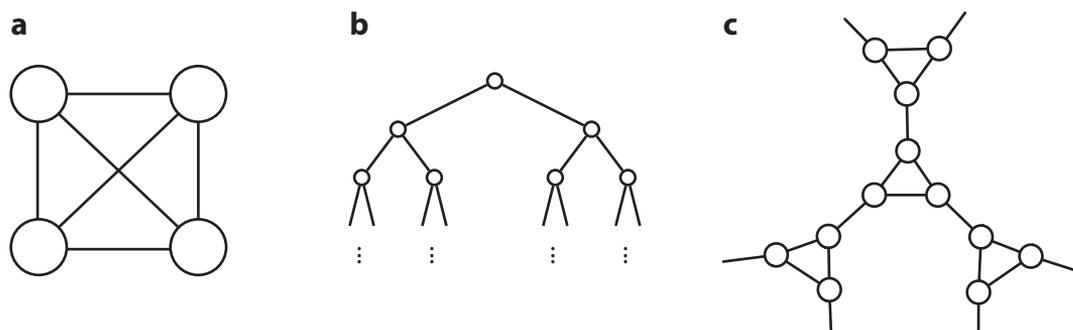


Figure 10

(a) Clique K_{d+1} for $d = 3$. (b) Infinite complete d -ary tree T_d for $d = 2$. (c) Tree of triangle ∇_d for $d = 3$.

lower μ risk than any other $\mathcal{G} \in \mathbb{G}_d$; (b) for every $\mathcal{G} \in \mathbb{G}_d$, there exists a $\mu_{\mathcal{G}}$ for which T_d has a strictly lower $\mu_{\mathcal{G}}$ risk than \mathcal{G} ; and (c) there exists a distribution μ for which the (d -regular) tree of triangles ∇_d (for an example, see **Figure 10b**), consisting essentially of a collection of disjoint triangles attached according to the structure of an infinite regular tree, has strictly lower μ risk than both K_{d+1} and T_d . ∇_d interpolates between the complete neighborhood diversification of T_d and the complete neighborhood closure of K_{d+1} .

The above results can be related to standard notions in relevant application domains. During epidemic disease, it is considered to be dangerous to belong to a large connected component, suggesting that K_{d+1} is the most robust network within \mathbb{G}_d . On the other hand, a principle in financial networks is that it is important to have diversity among one's neighbors, i.e., a lack of edges among one's neighbors, so that shocks are uncorrelated. This suggests that T_d is the most robust network in \mathbb{G}_d . These observations can be formalized by appropriate μ . Let $(\mu(0), \mu(1), \mu(2)) = (\varepsilon, x, 1 - \varepsilon - x)$ for some very small value of $\varepsilon > 0$ and $\mu(j) = 0$ for $j > 2$. If $x = 1 - \varepsilon$ (i.e., thresholds are either 0 or 1), then a node's failure probability is strictly increasing in the size of the component it belongs to, and so K_{d+1} uniquely minimizes the μ risk. On the other hand, there exists x strictly between 0 and $1 - \varepsilon$ such that x is very small but significantly larger than ε , so that thresholds of 1 are much more numerous than thresholds of 0. In this case, T_d is optimal, because, although failures are still rare, if a node u has connected neighbors v and w , then there is a nontrivial risk that v will have threshold 0 and w will have threshold 1, at which point v 's failure will ricochet off w and bring down u as well, even if u has the maximum (and most likely) threshold of 2. Therefore, in this scenario, it is safer to have no links among neighbors, even at the expense of producing very large connected components.

6.4. Future Research Directions

Very few formal stability and robustness analyses exist for cascading failure dynamics in physical networks. It would be natural to investigate whether the nonlinearities induced by cascading failure can be cast into canonical forms, similar in spirit to the capacity constraint in Section 5, so as to use tools from the piecewise linear or hybrid systems literature. In the setup in Sections 6.1 and 6.2, the vulnerability of a link is determined by its residual capacity with respect to the nominal flow, which could be interpreted as the equilibrium of the dynamics induced by control policy. Under this interpretation, robustness analysis is that of the dynamics. This is to be contrasted with the setting in Section 6.3, where the assignment of thresholds to nodes is independent of the initial network structure or the network formation process that leads to the initial interconnection structure. Modeling this dependency and evaluating the robustness of the network formation process will be an interesting direction to pursue.

7. ROBUSTNESS OF ASYMPTOTICALLY LARGE NETWORKS: STATIC SETTING

In Section 6.3, we discussed the interplay between network structure and failure thresholds at nodes in determining the robustness of financial networks under cascading failure. In this section, we discuss the interplay between network structure and interconnection weights in determining robustness under an equilibrium model in economic networks. Specifically, it is of interest to understand how idiosyncratic shocks at nodes translate into fluctuations of a meaningful aggregate quantity associated with the network. A central-limit-theorem-type argument suggests that if the shocks are independent, then fluctuations in the quantity corresponding to a simple summation of the shocks would decay proportional to $1/\sqrt{n}$. However, the interconnections induced by the

network can function as a potential propagation mechanism, under which the $1/\sqrt{n}$ decay may not hold true. Acemoglu et al. (41, 42) analyzed such propagation mechanisms and their dependence on network structure in the context of intersectoral economic networks.

Let the nodes in \mathcal{V} represent sectors that produce goods consumed by a representative household; the household provides a fixed inelastic cumulative unit labor for all the sectors. Let ℓ_i be the (fraction of) labor consumed by node i , and let x_i be the amount of goods produced by node i . The household utility function is of the Cobb–Douglas type—i.e., it is proportional to $\prod_{i \in [n]} x_i^{\eta_i}$, where $\eta_i > 0$ is i 's share in the household's utility function, normalized such that $\sum_{i \in [n]} \eta_i = 1$. The amount of goods produced by node i is modeled as

$$x_i = s_i^\alpha \ell_i^\alpha \prod_{j=1}^n x_{ij}^{(1-\alpha)w_{ij}}, \quad 26.$$

where x_{ij} is the amount of good j used in the production of good i ; $\alpha \in (0, 1)$ and $w_{ij} \in (0, 1)$ are the shares of labor and good j , respectively, in the input for production by i , satisfying $\sum_j w_{ij} = 1$ for all i ; and s_i is the idiosyncratic productivity shock to sector i . At the competitive equilibrium—i.e., where the household maximizes utility, the individual sectors maximize profits, and the labor and commodity market clears—the logarithm of the aggregate output of the network is given by

$$y = v^T \omega, \quad 27.$$

where $\omega = [\omega_1, \dots, \omega_n]^T$, with $\omega_i = \log(s_i)$, is the vector of microscopic shocks, and the i th component of the influence vector $v = [v^{(1)}, \dots, v^{(n)}]^T$, also referred to as the Domar weight of node i in economic network context, is given by

$$v^{(i)} = \sum_{j \in [n]} \eta_j L_{ji}, \quad 28.$$

where L_{ji} is the (j, i) element of the Leontief inverse $L = (I - (1 - \alpha)\mathcal{W})^{-1}$. That is, the logarithm of aggregate output is a linear combination of log node shocks with coefficients determined by the elements of the influence vector. The node shocks, and hence ω_i , are modeled to be random variables independent across the nodes. If $\mathbb{E}(\omega_i)$ is zero for all $i \in [n]$, then $\mathbb{E}(y)$ is also zero. It is of interest to study the standard deviation, also referred to as the aggregate volatility, of y ; this is done in Section 7.1. It is also of interest to study the aggregate output's τ -tail ratio (see Equation 1). A network is said to exhibit macroscopic tail risk if the aggregated output associated with it satisfies $\lim_{\tau \rightarrow \infty} r_y(\tau) = 0$, and this is studied in Section 7.2.

It can be shown that, for any finite network with n nodes, if $\{\omega_i\}_{i \in [n]}$ exhibit tail risks, then the network exhibits macroscopic tail risk as well. Therefore, for meaningful analysis, one considers a sequence of networks $\{\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n, \mathcal{W}_n)\}_{n \in \mathbb{N}}$ along with a collection of distributions of log node shocks $\{F_m\}_{i \in \mathcal{V}_n, n \in \mathbb{N}}$, and one studies aggregate volatility and tail risk as $n \rightarrow \infty$. We let the corresponding output vectors, influence vectors, and tail ratios be denoted as $\{y_n\}_{n \in \mathbb{N}}$, $\{v_n\}_{n \in \mathbb{N}}$, and $\{r_n(\tau_n)\}_{n \in \mathbb{N}}$, respectively.

7.1. Aggregate Volatility

Let the variances $\text{var}[\omega_m]$ be finite and uniformly strictly bounded away from zero to be able to focus on the effects of network structure as $n \rightarrow \infty$. Consider the special case when $\eta_i = 1/n$ for all $i \in [n]$. In this case, Equation 28 specializes to

$$v_n = \frac{1}{n} [I - (1 - \alpha)\mathcal{W}^T]^{-1} \mathbf{1}. \quad 29.$$

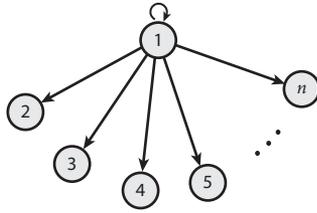


Figure 11

A network where one sector (node 1) is the only supplier.

For a given n , Equation 27 implies

$$(\text{var}[y_n])^{1/2} = \Theta(\|v_n\|_2). \quad 30.$$

It is of interest to study the dependence of volatility on n and compare against the $1/\sqrt{n}$ behavior, as would be implied by the naive application of the central limit theorem. One extreme case is the network depicted in **Figure 11**, for which $\|v_n\|_2 = \Theta(1)$ —i.e., volatility does not vanish even as $n \rightarrow \infty$. However, it is natural to expect that, in most realistic situations, volatility vanishes as $n \rightarrow \infty$. Indeed, one can show the following results that are complementary to Equation 30, when all the $\{\omega_{in}\}_{i,n}$ have the same variance σ^2 . If all $\{\omega_{in}\}_{i,n}$ are normally distributed, then $\frac{1}{\|v_n\|_2} y_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$. The same convergence result also holds true in general if there exists a cumulative distribution function \bar{F} such that $F_{in}(x) < \bar{F}(x)$ for $x < -a$ and $F_{in}(x) > \bar{F}(x)$ for $x > a$ for some $a > 0$, and if $\frac{\|v_n\|_\infty}{\|v_n\|_2} \rightarrow 0$. The last condition captures the dependence of network structure and weights through the influence vector v_n and implies that $\|v_n\|_\infty$, which captures the influence of the most central node, converges to zero faster than $\|v_n\|_2$.

A lower bound on volatility can be obtained in terms of the out-degrees $d_1^n, d_2^n, \dots, d_n^n$. Recall from Section 2 that $d_i^n := \sum_{j=1}^n w_{ij}$ for all $i \in [n]$. The lower bound is

$$(\text{var}[y_n])^{1/2} = \Omega\left(\frac{1 + \text{CV}_n}{\sqrt{n}}\right), \quad 31.$$

where the coefficient of variation CV_n measures the extent of asymmetry between nodes,

$$\text{CV}_n := \frac{1}{\bar{d}_n} \left(\frac{1}{n-1} \sum_{i=1}^n (d_i^n - \bar{d}_n)^2 \right)^{1/2},$$

where $\bar{d}_n = \sum_{i=1}^n d_i^n / n$ is the average out-degree. Equation 31 implies that asymmetry can cause the volatility to decay slower than $1/\sqrt{n}$. For example, for the network depicted in **Figure 11**, CV_n equals $\Theta(\sqrt{n})$, which then implies that the aggregate volatility is lower bounded by a constant for all values of n . More generally, if the network contains a dominant node whose degree grows linearly with n , then the aggregate volatility remains bounded away from zero. A complementary result holds true for networks whose degree sequences have heavier tails. Formally, a sequence of networks is said to have a power-law degree sequence if there exist a constant $\beta > 1$, a slowly varying function L satisfying $\lim_{t \rightarrow \infty} L(t)t^\beta = \infty$ and $\lim_{t \rightarrow \infty} L(t)t^{-\beta} = 0$ for all $\delta > 0$, and a sequence of positive numbers $\gamma_n = \Theta(1)$ such that, for all $n \in \mathbf{N}$ and all $k < d_{\max}^n = \Theta(n^{1/\beta})$, with d_{\max}^n being the maximum out-degree of \mathcal{G}_n , we have the empirical counter cumulative distribution function $P_n(k) \equiv \frac{1}{n} |\{i \in \mathcal{V}_n : d_i^n > k\}|$ satisfying $P_n(k) = \gamma_n k^{-\beta} L(k)$. For such networks with $\beta \in (1, 2)$, the aggregate volatility can be shown to be $\Omega(n^{-(\beta-1)/\beta-\delta})$ for arbitrary $\delta > 0$.

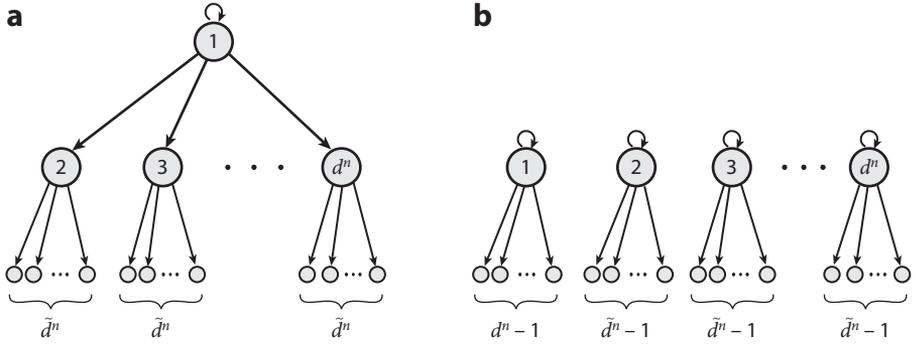


Figure 12

Networks with the same degree sequence but different aggregate volatilities.

The above results on the dependence of volatility on the degree properties of a network capture first-order effects of network structure and may not be sufficient to identify network properties under which the volatility decays slower than $1/\sqrt{n}$. For example, both the networks shown in **Figure 12** have the same degree sequence: a node labeled 1 with degree $d^n = \Theta(\sqrt{n})$; nodes $2, \dots, d^n$ with some degree \tilde{d}^n ; and other nodes with degree zero. However, the aggregate volatilities for the networks shown in **Figure 12a** and **Figure 12b** scale as $\Theta(1)$ and $\Theta(1/\sqrt[4]{n})$, respectively, independent of \tilde{d}^n . The first-order interconnections provide little or no information on the extent of cascade effects, whereby shocks to a node affect not only its immediate outgoing nodes but also outgoing nodes of those nodes, and so on. The second-order interconnectivity coefficient captures such effects:

$$\tau_2(\mathcal{W}_n) := \sum_{i=1}^n \sum_{j \neq i} \sum_{k \neq i, j} w_{ji}^n w_{ki}^n d_j^n d_k^n.$$

τ_2 takes higher values when high-degree nodes share incoming nodes with other high-degree nodes, as opposed to low-degree ones. The bound in Equation 31 can then be strengthened as

$$(\text{var}[y_n])^{1/2} = \Omega \left(\frac{1 + \text{CV}_n}{\sqrt{n}} + \frac{\sqrt{\tau_2(\mathcal{W}_n)}}{n} \right). \quad 32.$$

The effect of second-order interconnections can also be formalized in terms of a second-order degree sequence, where the second-order degree of a node i is defined as the weighted sum of the degrees of the nodes that are outgoing from i , i.e., $\sum_{j \in [n]} d_j^n w_{ji}^n$. If the second-order degree sequence associated with $\{\mathcal{G}_n\}_{n \in \mathbb{N}}$ has a power-law tail with shape parameter $\zeta \in (1, 2)$, then the aggregate volatility satisfies $(\text{var}[y_n])^{1/2} = \Omega(n^{-(\zeta-1)/\zeta-\delta})$ for any $\delta > 0$.

Equation 32 can be strengthened further to capture asymmetry in higher-order interconnections. Conversely, if the node degrees have limited variation, then the volatility decays at $1/\sqrt{n}$. Formally, for a sequence of balanced networks $\{\mathcal{G}_n\}_{n \in \mathbb{N}}$, i.e., ones satisfying $\max_{i \in [n]} d_i^n = \Theta(1)$, there exists $\bar{\alpha} \in (0, 1)$ such that for $\alpha \geq \bar{\alpha}$, $(\text{var}[y_n])^{1/2} = \Theta(1/\sqrt{n})$, where we recall from Equation 26 that α denotes the share of labor in the production of goods by nodes. **Figure 13** illustrates two simple examples of balanced networks.

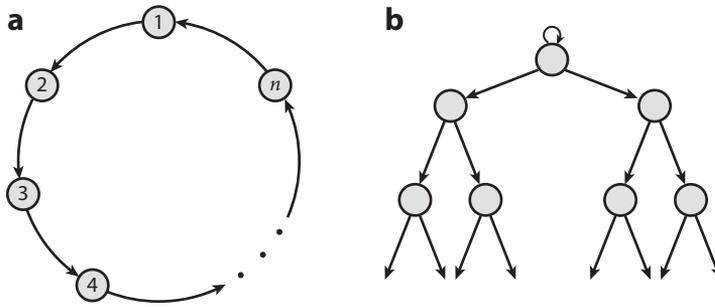


Figure 13

Two examples of balanced networks: (a) a ring and (b) a binary tree.

7.2. Macroscopic Tail Risks

Since the notion of macroscopic tail risk involves a $\tau \rightarrow \infty$ argument, the order in which τ and n are taken to infinity becomes crucial. We index τ in terms of n as $\{\tau_n\}_{n \in \mathbb{N}}$ to highlight the dependence of the rates at which the two limits are taken. The inspiration for the correct dependence of τ on n comes from the following result for simple economic networks, i.e., when $\alpha = 1$ and $\eta_i = 1/n$ for all $i \in [n]$ (see the formulation for Equation 26). If $\lim_{n \rightarrow \infty} \tau_n/\sqrt{n} = 0$, then $\lim_{n \rightarrow \infty} r_n(\tau_n) = 1$ for all light-tailed microscopic shocks; if $\lim_{n \rightarrow \infty} \tau_n/\sqrt{n} = \infty$, then there exist light-tailed microscopic shocks such that $\lim_{n \rightarrow \infty} r_n(\tau_n) = 0$. The latter in particular contradicts a standard argument under which $\{\omega_i\}_{i \in [n]}$ should have no aggregate impact as $n \rightarrow \infty$ for simple economic networks if the rate of growth of τ_n is fast enough, as noted by Acemoglu et al. (42). Accordingly, we say that a sequence of (not necessarily simple) networks exhibits macroscopic tail risks if $\lim_{n \rightarrow \infty} r_n(c\sqrt{n}) = 0$ for all $c > 0$.

The presence or absence of macroscopic tail risks is determined by the interaction between the extent of heterogeneity in the Domar weights in Equation 28 and the distribution of microscopic shocks $\{\omega_{in}\}$. For instance, if the microscopic shocks are normally distributed, then no sequence of networks exhibits macroscopic tail risks. In more interesting cases, a measure of the node dominance of a given network,

$$\delta_v^n := \frac{\|v_n\|_\infty}{\|v_n\|_2/\sqrt{n}}, \quad 33.$$

plays a key role in macroscopic tail risks. The normalization factor \sqrt{n} is meant to reflect that δ_v^n captures dominance relative to simple networks, where $\|v_n\|_\infty = 1/n$ and $\|v_n\|_2 = 1/\sqrt{n}$, and therefore the node dominance is one for simple networks.

Under exponential-tailed shocks, a sequence of networks exhibits macroscopic tail risks if and only if $\lim_{n \rightarrow \infty} \delta_v^n = \infty$. A shock has an exponential tail if its cumulative distribution function satisfies $\lim_{x \rightarrow \infty} \frac{1}{x} \log F(-x) = -\gamma$ for some $\gamma > 0$. For example, this is satisfied if, for some polynomial function $Q(x)$, $F(-x) = 1 - F(x) = Q(x)e^{-\gamma x}$ for all $x \geq 0$. While exponential-tailed shocks are light-tailed distributions, they exhibit tail risks. Therefore, heterogeneity in entries of the influence vector is essential not only in generating aggregate volatility, as in Section 7.1, but also in translating microscopic tail risk into macroscopic tail risks. However, the roles played in these two aspects are fundamentally distinct. For example, under exponential-tailed microscopic shocks, a sequence of networks for which $\lim_{n \rightarrow \infty} \delta_v^n/\sqrt{n} = 0$ and $\lim_{n \rightarrow \infty} \delta_v^n = \infty$ exhibits macroscopic tail risks, even though aggregate output is asymptotically normally distributed. As an illustration, consider the network in **Figure 14** with $\eta_i = 1/n$ for all $i \in [n]$. One can

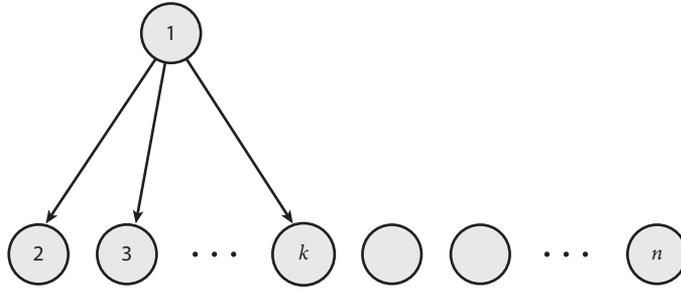


Figure 14

A network where one sector (node 1) is the only supplier and only a subset of $k - 1$ nodes require input.

verify that, in this case, $\lim_{n \rightarrow \infty} \delta_v^n = \infty$, and hence the network has macroscopic tail risk under exponential-tailed shocks if and only if $k \rightarrow \infty$ as $n \rightarrow \infty$. The latter could happen even if the only supply node is connected to a diminishing fraction of other nodes, e.g., $k = \log n$, which satisfies $\lim_{n \rightarrow \infty} k/n = 0$, which in turn can be shown to imply $\frac{\|v_n\|_\infty}{\|v_n\|_2} \rightarrow 0$ and hence no aggregate volatility from the discussion in Section 7.1.

The above results can be generalized to a larger subclass of light-tailed microscopic shocks, such as the ones with superexponential distributions with shape parameter $\nu \in (1, 2)$, in the sense that $\lim_{x \rightarrow \infty} \frac{1}{x^\nu} \log F(-x) = -\gamma$ for some $\gamma > 0$. For example, this is satisfied if $F(-x) = 1 - F(x) = Q(x) \exp(-\gamma x^\nu)$ for some polynomial function $Q(x)$. Such distributions exhibit tail risks while having tails that are lighter than that of the exponential distribution. Under microscopic shocks with superexponential tails with shape parameter $\nu \in (1, 2)$, a sequence of networks exhibits no macroscopic tail risk if $\liminf_{n \rightarrow \infty} \delta_v^n < \infty$, whereas it exhibits macroscopic tail risk if $\lim_{n \rightarrow \infty} \delta_v^n / n^{(\nu-1)/\nu} = \infty$.

As expected, one gets macroscopic tail risks for heavy-tailed shocks. Specifically, if the microeconomic shocks are Pareto (heavy-)tailed, i.e., if $\lim_{x \rightarrow \infty} \frac{1}{\log x} \log F(-x) = -\lambda$, with $\lambda > 2$, then any sequence of networks exhibits macroeconomic tail risks. This is because the likelihood that at least one node is hit with a large shock is high. However, as discussed above in this section, macroeconomic tail risks can emerge not only due to microshocks that are drawn from heavy-tailed distributions but also as a consequence of the interplay between relatively light-tailed distributions and heterogeneity in Domar weights. In fact, for a sequence of simple networks subject to Pareto-tailed shocks, there exists a sequence of networks subject to exponential-tailed shocks that exhibits an identical level of macroeconomic tail risks (42).

7.3. Future Research Directions

The results reviewed in this section depend on the linear relationship between Domar weights v and external shocks ω in Equation 27 and on the specific dependence of v on \mathcal{W} in Equation 28. This in turn depends on the underlying specific equilibrium setup. Acemoglu et al. (42) suggested practical settings under which \mathcal{W} is not independent of ω , which then induces a nonlinear relationship between v and ω . Investigating aggregate volatility and macroscopic tail risks under such nonlinearities would be interesting.

8. ROBUSTNESS OF ASYMPTOTICALLY LARGE NETWORKS: DYNAMICAL SETTING

The notion of aggregate output from Section 7 can be related to the following dynamics:

$$x(t+1) = Ax(t) + \omega \delta(0, t), \quad t \in \{0, 1, 2, \dots\}, \quad x(0) = 0, \quad 34.$$

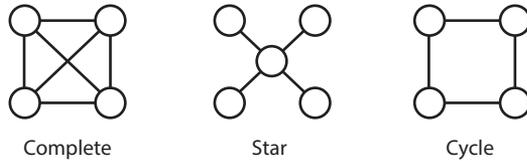


Figure 15

A few typical network topologies.

where $x(t)$ is the vector of state variables; A is the $n \times n$ state transition matrix, which is assumed to be Schur stable [i.e., it satisfies $\rho(A) < 1$]; $\delta(0, t)$ is the Kronecker delta function; and $\omega \in \mathbf{R}^n$ is exogenous to the system. Under this dynamics, since $x(t) = A^t \omega$, we have $x_\infty := \frac{1}{n} \sum_{t=0}^{\infty} \mathbf{1}^T x(t) = \frac{1}{n} \mathbf{1}^T (I + A + A^2 + \dots) \omega = \frac{1}{n} \mathbf{1}^T (I - A)^{-1} \omega$. Comparing with Equations 27 and 29, one can see that x_∞ is the aggregate output when each node's share in the household's utility is identical, i.e., $\eta_i = 1/n$ for all $i \in [n]$, and $A = (1 - \alpha)\mathcal{W}$.

Sarkar et al. (43) connected the macroscopic tail risk in x_∞ to the (identity) Gramian of A while generalizing A beyond $A = (1 - \alpha)\mathcal{W}$. The Gramian matrix of A , denoted as $P(A)$, is the one that satisfies $P = A^T P A + I$. Sarkar et al. (43) showed that if each of $\{\omega_i\}_{i \in [n]}$ is exponential-tailed with a continuous, symmetric probability density function with full support, in addition to satisfying $\mathbb{E}[\omega] = \mathbf{0}$ and $\mathbb{E}[\omega \omega^T] = I$, and if the network sequence $\{A_n\}_{n \in \mathbf{N}}$ is such that, for each n , (a) A_n is nonnegative; (b) A_n has a Perron root λ_{PF} , with a corresponding right eigenvector z satisfying $\max_i z_i / \min_i z_i = O(1)$; and (c) a scaled version of the associated Gramian satisfies $\|P(A_n / \sqrt{\lambda_{\text{PF}}})\|_1 = \Theta(1)$, then the network sequence does not have macroscopic tail risks. Accordingly, it can be shown that complete and cycle networks have no macroscopic tail risks, whereas star networks do (for an illustration of these network topologies, see **Figure 15**). Sarkar et al. (43) noted that conditions a and b above on the networks allow $A_n = \gamma \mathcal{W}_n$, with $\gamma \in (0, 1)$ and \mathcal{W}_n a row-stochastic matrix, as well as other A_n with $\|A_n\|_1 < 1$ and stable A_n with $\|A_n\|_1 > 1$. This is a generalization of the setup of Acemoglu et al. (42), where A_n is restricted to a specific class of A_n with $\|A_n\| < 1$, satisfying $A_n = (1 - \alpha)\mathcal{W}_n$, $\alpha \in (0, 1)$.

8.1. Disruption Energy

The Gramian also plays a role in characterizing network robustness metrics related to the energy of the network, $\sum_{t=0}^{\infty} x^T(t)x(t)$. It is of interest to study the behavior of this quantity under deterministic and stochastic shocks ω . Formally, the following quantities are of interest:

$$\mathbf{M}(A) = \sup_{\|\omega\|_2=1} \sum_{t=0}^{\infty} x^T(t)x(t), \quad \mathbf{E}(A) = \frac{1}{n} \mathbb{E}_\omega \left[\sum_{t=0}^{\infty} x^T(t)x(t) \right],$$

where, in the definition of $\mathbf{E}(A)$, $\mathbb{E}[\omega] = 0$ and $\mathbb{E}[\omega \omega^T] = I$. $\mathbf{M}(A)$ and $\mathbf{E}(A)$ are, respectively, referred to as the maximum and average disruption energy of A . These two quantities represent two different aspects of network robustness. Under a deterministic shock incident to the network, $\mathbf{M}(A)$ denotes the maximum energy that can propagate through the network, while, under a random shock, $\mathbf{E}(A)$ denotes expected energy that propagates through the network.

The Gramian gives a complete energy profile for a network. Sarkar et al. (43) showed that, for any A , $\mathbf{M}(A)$ is equal to $\sigma(P(A))$ and $\mathbf{E}(A)$ is equal to $\frac{1}{n} \text{trace}(P(A))$. The latter is also related to the notion of the \mathcal{H}_2 norm of a network, which measures the cumulative amplification of a shock due to network effects. Specifically, $\mathcal{H}_2(A)$ equals $\text{trace}(P(A))$. For any finite network, the two disruption energies are finite if $\rho(A) < 1$. Therefore, a meaningful approach to characterize

robustness is to consider a sequence $\{A_n\}_{n \in \mathbb{N}}$ of networks with $\rho(A_n) < 1$ for every n and study the scaling of disruption energies as the size $n \rightarrow \infty$.

Sarkar et al. (43) showed that, for large undirected networks A_n , the disruption energies are related to the spectral radius as $\mathbf{M}(A_n) = \frac{1}{1-\rho^2(A_n)}$ and $\mathbf{E}(A_n) = O(\frac{1}{1-\rho(A_n)})$. Specifically, for complete, star, and cycle graphs (see **Figure 15**), $\mathbf{M}(A_n)$ scales as $\Theta(n)$, $\Theta(1)$, and $\Theta(1)$, respectively, and $\mathbf{E}(A_n)$ scales as $\Theta(1)$, $\Theta(1)$, and $\Theta(n)$, respectively. On the other hand, for directed networks, the scaling of disruption energies may be independent of spectra. In particular, there exist vehicular platoon networks (e.g., see 44) for which $\mathcal{H}_2(A_n)$ is $\Omega(\exp(\delta n))$ for some $\delta > 0$, even if $\rho(A_n)$ is uniformly bounded away from one. These results suggest a possible advantage of undirected networks over directed networks. Indeed, Sarkar et al. (43) showed that making a large directed network A_n more undirected does not increase the two disruption energies in an order sense, and they provide examples where making directed network more undirected strictly reduces the \mathcal{H}_2 norm. The process of making a network more undirected is achieved by spectral balancing of its A_n . Formally, for $\epsilon \in (0, 1)$, an ϵ -balanced version of A_n is $A_n^\epsilon = (1 - \epsilon)A_n + \epsilon U_n \Gamma_n U_n^T$, where Γ_n is a diagonal matrix such that $\rho(\Gamma_n) \leq \rho(A_n)$ and U_n is obtained from the spectral decomposition of the Gramian of A_n : $P(A_n) = U_n D_n U_n^T$.

8.2. Future Research Directions

The dependence of macroscopic tail risks and disruption energies on a Gramian suggests a connection between these two notions from a network robustness perspective. However, the exact relationship remains to be established.

9. CONCLUSION

We have presented an overview of multiple approaches to analyzing the robustness of networked systems. The common theme is to highlight the role played by the underlying network structure in determining robustness.

For linear dynamics over networks with unreliable links, the interplay of the reliability of links and their role in the dynamics, as captured by the loop-gain operator of the associated covariance feedback system, determines whether the network dynamics could or could not arbitrarily amplify additive white noise. In an economic network setting, heterogeneity in the node degrees plays a key role in determining the rate at which the deviation in the aggregate equilibrium network output, due to independent shocks to individual nodes, decays to zero as the network size increases. On the other hand, the interplay between the distribution functions of the individual shocks and the heterogeneity in node dominance determines whether the aggregate output exhibits tail risk. The aggregate output in an economic network at equilibrium can be related to the transient of a related linear network dynamics subject to an initial shock. Such an abstraction allows one to relate macroscopic tail risks to the corresponding Gramian and to study tail risk beyond economic network settings. The Gramian also plays a key role in determining the scaling of network disruption energies, which serve as meaningful robustness metrics.

For physical flow over networks, the notions of link, node, and network residual capacities are key determinants of robustness to reduction in capacity. These quantities are relatively fast to compute in the static setting, when the flow is physically constrained only by Kirchhoff's law. In the presence of an additional constraint, such as Ohm's law in electrical networks, computing the relevant quantities is hard in general; however, the link-reducible structure of the network helps to considerably reduce the complexity. In the dynamical setting, network robustness is known to be influenced by information constraint and the nature of control actions. For instance, decentralized control does not cause a loss of robustness in certain scheduling scenarios, but it does so in certain

routing scenarios. In the latter case, adding scheduling to control action allows one to recover the loss in robustness.

The extension of network flow robustness to a cascading failure setting involves dynamic-programming-like computations, with considerable simplifications possible under a tree-like network structure and symmetry about a source–sink pair. For contagion-based cascading failure process, the maximum likelihood of failure among all nodes is determined by the interplay of the node threshold assignment and the network structure. Within regular graphs, the clique structure and the complete d -ary tree structure provide a clean way to analyze this interplay if the node degrees are two or if the node threshold values are upper bounded by two.

The selection of results and setups reviewed in the article naturally reflect our bias stemming from our work in this area. Therefore, the material in the article is to be interpreted as reflecting a subset of the growing literature on network robustness.

DISCLOSURE STATEMENT

K.S. has financial interest in Xtelligent Inc.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation CAREER Electrical, Communications, and Cyber Systems grant 1454729 and by funding from King Abdullah University of Science and Technology. The authors thank Bassam Bamieh for helpful discussions.

LITERATURE CITED

1. Korte B, Vygen J. 2002. *Combinatorial Optimization: Theory and Algorithms*. Berlin: Springer
2. Christiano P, Kelner JA, Madry A, Spielman DA, Teng SH. 2011. Electrical flows, Laplacian systems, and faster approximation of maximum flow in undirected graphs. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, pp. 273–82. New York: ACM
3. Zhou K, Doyle J, Glover K. 1996. *Robust and Optimal Control*. Upper Saddle River, NJ: Prentice Hall
4. Ba Q, Savla K. 2018. Robustness of DC networks under controllable link weights. *IEEE Trans. Control Netw. Syst.* 5:1479–91
5. Zimmerman RD, Murillo-Sánchez CE, Thomas RJ. 2011. MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans. Power Syst.* 26:12–19
6. Ba Q. 2018. *Elements of robustness and optimal control for infrastructure networks*. PhD Thesis, Univ. South. Calif., Los Angeles. <http://digitallibrary.usc.edu/cdm/compoundobject/collection/p15799coll40/id/475855/rec/1>
7. Willems J. 1970. *The Analysis of Feedback Systems*. Cambridge, MA: MIT Press
8. Desoer C, Vidyasagar M. 1975. *Feedback Systems: Input–Output Properties*. New York: Academic
9. Jiang ZP, Liu T. 2018. Small-gain theory for stability and control of dynamical networks: a survey. *Annu. Rev. Control* 46:58–79
10. Shamma JS. 1991. The necessity of the small-gain theorem for time-varying and nonlinear systems. *IEEE Trans. Autom. Control* 36:1138–47
11. Freeman RA. 2001. On the necessity of the small-gain theorem in the performance analysis of nonlinear systems. In *Proceedings of the 40th IEEE Conference on Decision and Control*, Vol. 1, pp. 51–56. Piscataway, NJ: IEEE
12. Doyle J. 1982. Analysis of feedback systems with structured uncertainties. *IEE Proc. D* 129:242–50
13. Safonov M. 1982. Stability margins of diagonally perturbed multivariable feedback systems. *IEE Proc. D* 129:251–56
14. Packard A, Doyle J. 1993. The complex structured singular value. *Automatica* 29:71–109

15. Poola K, Tikku A. 1995. Robust performance against time-varying structured perturbations. *IEEE Trans. Autom. Control* 40:1589–602
16. Megretski A. 1993. Necessary and sufficient conditions of stability: a multiloop generalization of the circle criterion. *IEEE Trans. Autom. Control* 38:753–56
17. Shamma JS. 1994. Robust stability with time-varying structured uncertainty. *IEEE Trans. Autom. Control* 39:714–24
18. Khammash M, Pearson JB. 1991. Performance robustness of discrete-time systems with structured uncertainty. *IEEE Trans. Autom. Control* 36:398–412
19. Willems JC. 2007. Dissipative dynamical systems. *Eur. J. Control* 13:134–51
20. Arcak M, Meissen C, Packard A. 2016. *Networks of Dissipative Systems: Compositional Certification of Stability, Performance, and Safety*. Cham, Switz.: Springer
21. Megretski A, Rantzer A. 1997. System analysis via integral quadratic constraints. *IEEE Trans. Autom. Control* 42:819–30
22. Nesterov Y, Nemirovskii A. 1994. *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia: Soc. Ind. Appl. Math.
23. Boyd S, El Ghaoui L, Feron E, Balakrishnan V. 1994. *Linear Matrix Inequalities in System and Control Theory*. Philadelphia: Soc. Ind. Appl. Math.
24. Bamieh B, Filo M. 2018. An input-output approach to structured stochastic uncertainty. arXiv:1806.07473 [cs.SY]
25. Fisher J, Bhattacharya R. 2009. Linear quadratic regulation of systems with stochastic parameter uncertainties. *Automatica* 45:2831–41
26. Kim KKK, Shen DE, Nagy ZK, Braatz RD. 2013. Wiener's polynomial chaos for the analysis and control of nonlinear dynamical systems with probabilistic uncertainties. *IEEE Control Syst. Mag.* 33(5):58–67
27. Wang J, Elia N. 2012. Distributed averaging under constraints on information exchange: emergence of Lévy flights. *IEEE Trans. Autom. Control* 57:2435–49
28. Gonçalves JM, Megretski A, Dahleh MA. 2003. Global analysis of piecewise linear systems using impact maps and surface Lyapunov functions. *IEEE Trans. Autom. Control* 48:2089–106
29. Sontag ED. 2010. Contractive systems with inputs. In *Perspectives in Mathematical System Theory, Control, and Signal Processing*, ed. JC Willems, S Hara, Y Ohta, H Fujioka, pp. 217–28. Berlin: Springer
30. Como G, Savla K, Acemoglu D, Dahleh MA, Frazzoli E. 2013. Stability analysis of transportation networks with multiscale driver decisions. *SIAM J. Control Optim.* 51:230–52
31. Sandholm WH. 2010. *Population Games and Evolutionary Dynamics*. Cambridge, MA: MIT Press
32. Como G, Savla K, Acemoglu D, Dahleh MA, Frazzoli E. 2013. Robust distributed routing in dynamical networks—part I: locally responsive policies and weak resilience. *IEEE Trans. Autom. Control* 58:317–32
33. Como G, Savla K, Acemoglu D, Dahleh MA, Frazzoli E. 2013. Robust distributed routing in dynamical networks—part II: strong resilience, equilibrium selection and cascaded failures. *IEEE Trans. Autom. Control* 58:333–48
34. Savla K, Lovisari E, Como G. 2013. On maximally stabilizing adaptive traffic signal control. In *2013 51st Annual Allerton Conference on Communication, Control, and Computing*, pp. 464–71. Piscataway, NJ: IEEE
35. Massoulié L. 2007. Structural properties of proportional fairness: stability and insensitivity. *Ann. Appl. Probab.* 17:809–39
36. Tassiulas L, Ephremides A. 1992. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Trans. Autom. Control* 37:1936–48
37. Como G, Lovisari E, Savla K. 2015. Throughput optimality and overload behavior of dynamical flow networks under monotone distributed routing. *IEEE Trans. Control Netw. Syst.* 2:57–67
38. Ba Q, Savla K. 2020. Computing optimal control of cascading failure in DC networks. *IEEE Trans. Autom. Control*. In press. <https://doi.org/10.1109/TAC.2019.2930232>
39. Savla K, Como G, Dahleh MA. 2014. Robust network routing under cascading failures. *IEEE Trans. Netw. Sci. Eng.* 1:53–66
40. Blume L, Easley D, Kleinberg J, Kleinberg R, Tardos É. 2011. Which networks are least susceptible to cascading failures? In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pp. 393–402. Piscataway, NJ: IEEE

41. Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A. 2012. The network origins of aggregate fluctuations. *Econometrica* 80:1977–2016
42. Acemoglu D, Ozdaglar A, Tahbaz-Salehi A. 2017. Microeconomic origins of macroeconomic tail risks. *Am. Econ. Rev.* 107:54–108
43. Sarkar T, Roozbehani M, Dahleh MA. 2019. Asymptotic network robustness. *IEEE Trans. Control Netw. Syst.* 6:812–21
44. Herman I, Martinec D, Hurák Z, Šebek M. 2014. Nonzero bound on Fiedler eigenvalue causes exponential growth of H-infinity norm of vehicular platoon. *IEEE Trans. Autom. Control* 60:2248–53