FINAL TECHNICAL REPORT

DE-SC0002051

Robust Optimization for Connectivity and Flows in Dynamic Complex Networks

Principal Investigator Information

Institution: Oklahoma State University, College of Engineering, Architecture and Technology, 201 ATRC, Stillwater, OK 74078

Principal Investigator: Dr. Balabhaskar Balasundaram, School of Industrial Engineering and Management, 322 Engineering North, Stillwater, OK 74075, Phone: (405) 744-6055,
E-mail: baski@okstate.edu

DOE/Office of Science Program Office: Advanced Scientific Computing Research
DOE/Office of Science Program Office Technical Contact: Alexandra Landsberg

Subaward Information

Oklahoma State University issued subawards to Texas A&M University for work to be performed under the direction of Co-PI Dr. Sergiy Butenko and to the University of Florida for work to be performed under the direction of Co-PIs Dr. Vladimir Boginski and Dr. Stanislav Uryasev.

Contents

1 Project Background and Motivation **1**

2 Technical Progress **1**

2.1 Generic Problem Statement and Overview of the Algorithmic Approach 2

2.2 Resilient Network Design: The Minimum Spanning k-Core Problem 4

2.3 Robust Flows: The Fixed Charge Network Flow Problem 7

2.4 Cluster Detection: Hereditary Clique Relaxations 9

2.5 Cluster Detection: The Maximum k-Club Problem 11

3 Key Accomplishments **12**

4 Research Transition **13**

5 Comprehensive List of Project Outcomes **13**

5.1 Refereed Journal Articles 13

5.2 Refereed Book Chapters 15

5.3 Refereed Conference Proceedings 15

5.4 Doctoral Dissertations 15

5.5 Masters Theses 16

5.6 Presentations 16

Executive Summary

Objective. The goal of this project was to study robust connectivity and flow patterns of complex multi-scale systems modeled as networks. Networks provide effective ways to study global, system level properties, as well as local, multi-scale interactions at a component level. Numerous applications from power systems, telecommunication, transportation, biology, social science, and other areas have benefited from novel network-based models and their analysis. Modeling and optimization techniques that employ appropriate measures of risk for identifying robust clusters and resilient network designs in networks subject to uncertain failures were investigated in this collaborative multi-university project.

Potential Impact. In many practical situations one has to deal with uncertainties associated with possible failures of network components, thereby affecting the overall efficiency and performance of the system (e.g., every node/connection has a probability of partial or complete failure). Some extreme examples include power grid component failures, airline hub failures due to weather, or freeway closures due to emergencies. These are also situations in which people, materials, or other resources need to be managed efficiently. Important practical examples include rerouting flow through power grids, adjusting flight plans, and identifying routes for emergency services and supplies, in the event network elements fail unexpectedly. Solutions that are robust under uncertainty, in addition to being economically efficient, are needed. This project has led to the development of novel models and methodologies that can tackle the optimization problems arising in such situations. A number of new concepts, which have not been previously applied in this setting, were investigated in the framework of the project. The results can potentially help decision-makers to better control and identify robust or risk-averse decisions in such situations.

Method of Approach. Formulations and optimal solutions of the considered problems need to capture uncertainty and risk using appropriate probabilistic, statistical and optimization concepts. The main difficulty arising in addressing these issues is the dramatic increase in the computational complexity of the resulting optimization problems. This project studied novel models and methodologies for risk-averse network optimization- specifically, network design, network flows and cluster detection problems under uncertainty. The approach taken was to incorporate a quantitative risk measure known as conditional value-at-risk that is widely used in financial applications. This approach presents a viable alternate modeling and optimization framework to chance-constrained optimization and mean-variance optimization, one that also facilitates the detection of risk-averse solutions.

Investigators and Affiliation.
(PI) Dr. B. Balasundaram, Oklahoma State University
(Co-PI) Dr. S. Butenko, Texas A&M University
(Co-PI) Dr. V. Boginski, University of Florida
(Co-PI) Dr. S. Uryasev, University of Florida

1 Project Background and Motivation

The DOE workshop on *Mathematical Research Challenges in Optimization of Complex Systems* reports with great emphasis, the role of *optimization* and *risk analysis* in the study complex systems [22]. Enhancing the mathematical knowledge-base for modeling and optimization of complex energy systems will impact applications in nuclear and fossil fuel based power generation; design and control of power grids; as well as growing applications in smart grids and sustainable energy. Uncertainty quantification is not only central to prediction and simulation of such complex systems, but it is also critical to enable the incorporation of appropriate risk measures in the optimization of complex systems.

This project focuses on network optimization models that are widely used and regarded as effective high-level models of complex systems. Specifically, this project investigated *conditional value-at-risk (CVaR) optimization models and algorithms for network flow, network design and cluster detection problems in stochastic networks.* The distinguishing characteristics of this project can be summarized as follows.

1. A vast majority of current literature addressing uncertainty in network flow and design problems restrict their attention to uncertainty in demands, capacity or cost parameters leaving the network deterministic and certain. This project investigated selected problems from these broad families of network optimization problems focusing specifically on uncertainty in the network structure itself, i.e., nodes/arcs have probabilities for failure.

2. Quantification of risk via the concept of CVaR and its advantages are well-known in financial applications. The results of this project are among the first of their kind in extending this concept to study "risk-averse" network flow, network design and cluster detection problems. Models for the loss functions that balance computational and practical considerations were developed and used to quantify downside risk of high losses using CVaR in network optimization.

3. New exact approaches based on decomposition and integer programming, as well as purely combinatorial algorithms were developed to solve the resulting large-scale mathematical optimization problems. This study also advanced the polyhedral results and complexity results of the deterministic versions of the considered combinatorial optimization problems which is central to the overall solution procedure.

2 Technical Progress

In this section we summarize the major research contributions of this project and highlight key innovations that resulted from this grant. Details can be found in published articles, and those in the final stages of preparation will be available as prepublication manuscripts after Jan 2014. A comprehensive list of major project outcomes is provided in Section 5.

The overall objective of this project was to investigate theory and algorithms to solve the CVaR-constrained version of selected network optimization problems when the edges in the graph can fail independently according to a probability distribution. The computational methods developed (for the most part) assume an available set of scenarios which are spanning subgraphs of the random graph. Hence, the independence assumption in edge failures is strictly not required from a computational perspective as long as sampling from the joint distribution is possible. Likewise, node failures are also included in this setting if we assume that sampling from the joint node-arc failure distribution is possible since the presence of isolated nodes (ones that have failed, and

hence have no incident edges) does not alter the solution feasibility or interpretation in the network optimization problems considered in this study.

2.1 Generic Problem Statement and Overview of the Algorithmic Approach

Let $\tilde{G} = (V, \tilde{E})$, denote the (nonuniform) random graph and we call $G = (V, E)$ the (deterministic) support graph in which every edge with a positive probability to exist is included. The sample space Ω contains all spanning subgraphs of G. We denote the probability measure by $\mathbb{P} : 2^{\Omega} \longrightarrow [0, 1]$. The vector of indicator random variables $\mathbf{y} : \Omega \longrightarrow \{0, 1\}^{|E|}$ denote the existence of edges with $\mathbb{P}\{\mathbf{y}_e = 1\} = p_e$. We refer to $G^s \in \Omega$ for $s = 1, \ldots, N$, or equivalently, the realizations of $\mathbf{y}$ denoted by $\mathcal{N} = \{y^1, \ldots, y^N\}$, as the *scenarios* and we denote the probability of scenario s realizing by $\pi^s = \mathbb{P}\{\mathbf{y} = y^s\}$.

Further, let $Q(G)$ denote the collection integral vectors that satisfy the specified network property on the support graph. For instance, if we are interested in the network design problem $Q(G) \subseteq \{0, 1\}^{|E|}$ and $x \in Q(G)$ would denote a subset of edges- a feasible design that satisfies a desired property (e.g., spanning k-cores). Similarly, if we are interested in the cluster detection problem $Q(G) \subseteq \{0, 1\}^{|V|}$ and $x \in Q(G)$ would be an incidence vector of a subset of vertices that satisfy the required cluster definition (e.g. k-plex, k-defective clique, k-club). Alternately, $Q(G) \subseteq \mathbb{R}_+^{|E|}$ if we are interested in flows that are feasible for the support graph G. Naturally, E is also assumed to be directed for the network flow problem. For the problems of interest to us, $Q(G)$ has a linear formulation in discrete or continuous variables.

Under uncertainty, a solution that was feasible for the support graph can become infeasible in some scenarios. The approach here is to quantify losses under uncertainty induced infeasibility of the solution. Denote by $L(x, \mathbf{y})$, a *loss function* that measures the infeasibility of $x \in Q(G)$ under random edge failures. Given a threshold $\alpha \in (0, 1)$, α-CVaR$[L(x, \mathbf{y})]$ measures the conditional expectation of tail losses where the tail is defined by the α-percentile (α-VaR) of the loss distribution. When α is large, say 0.9, we thereby quantify the downside risk of high losses using CVaR and limit it using a user-specified parameter C in the optimization model.

It is possible for an optimal solution to a CVaR-constrained formulation to be infeasible with higher probability since it is not explicitly controlled. By contrast a chance-constrained optimization formulation requires that the solution be feasible with high probability ($> \alpha$). But since α-VaR$[L(x, \mathbf{y})] \leq \alpha$-CVaR$[L(x, \mathbf{y})] \leq C$, with high probability the losses will not exceed the user-specified tolerance C. The CVaR-constrained optimization framework is appropriate for settings were the user is less sensitive to infeasibility per se and more concerned about scenarios that result in heavy losses (i.e., very high infeasibility- measured as cumulative or maximum constraint violation). In other words, this approach assumes that the user can manage infeasibility as long as it is limited or contained. Chance-constrained optimization on the other hand is appropriate when the distinction between feasibility and infeasibility is dichotomous and the "amount" of infeasibility is irrelevant. The following is a general-purpose and compact representation of the problems of interest:

$$\min\Big\{w^T x \ : \ x \in Q(G), \ \alpha\text{-CVaR}[L(x, \mathbf{y})] \leq C\Big\}, \tag{1}$$

where $Q(G) = \Big\{x \in Z_+^d : Ax \leq b\Big\}$, d is an appropriate dimension (usually, $|V|$ or $|E|$), and $w \in \mathbb{Z}^d$ is the cost vector for the network design and network flow problems and is a vector of -1s for the cluster detection problem where we are interested in maximizing the cardinality of the detected cluster. To solve this problem, we focused on the following technical issues:

1. Investigated tight convex relaxations of $Q(G)$ by studying the deterministic version of the

problem, specifically polyhedral combinatorics;

2. Investigated general approaches for developing convex loss functions that will ensure the convexity of $\{x \in \mathbb{R}^d \ : \ \alpha\text{-CVaR}[L(x, \mathbf{y})] \leq C\}$ and enable effective computational techniques;

3. Investigated reformulation and decomposition techniques and developed effective exact algorithms including those that are purely combinatorial for solving the deterministic as well as CVaR-constrained optimization problems.

Bounding CVaR in an optimization model can be equivalently achieved by bounding the function [15],

$$F_\alpha(x, \zeta) = \zeta + \frac{1}{1-\alpha} \sum_{s=1}^{N} \pi^s [L(x, y^s) - \zeta]^+. \tag{2}$$

CVaR is convex in x if $L(x, \mathbf{y})$ is convex in x. Since the loss functions employed in this study are convex and piecewise linear in x, the aforementioned equivalence implies that the optimization problem (1) can be reformulated without explicitly knowing either the distribution of $L(x, \mathbf{y})$, or an expression for CVaR. The following formulation is the Rockafellar and Uryasev reformulation (RUR) of (1) using function 2.

$$\textbf{(RUR)} \qquad \min_{x \in Q(G),\ \zeta \in \mathbb{R}} \left\{ w^T x \ : \ \zeta + \frac{1}{1-\alpha} \sum_{s=1}^{N} \pi^s [L(x, y^s) - \zeta]^+ \leq C \right\} \tag{3}$$

An obvious challenge working with RUR is the number of scenarios in Ω, since $N = 2^{|E|}$ if the edge failures are independent. One approach here is to approximate Ω through uniform random sampling to produce a set of equally likely samples $\mathcal{S}$, in which case one can substitute $F_\alpha(x, \zeta)$ with its scenario-based approximation

$$\tilde{F}_\alpha(x, \zeta) = \zeta + \frac{1}{(1-\alpha)|\mathcal{S}|} \sum_{s \in \mathcal{S}} [L(x, y^s) - \zeta]^+. \tag{4}$$

With this approximation, the CVaR-constrained optimization problem based on $\mathcal{S}$ is a large-scale mixed integer program (MIP) once the term $[L(x, y^s) - \zeta]^+$ and the loss function $L(x, y^s)$ have been linearized. If $\mathcal{S}$ is manageable, it could be solved using a general-purpose MIP solver. Pertinently, the quality of the approximation $\tilde{F}_\alpha(x, \zeta)$ based on $\mathcal{S}$ would benefit from using larger $\mathcal{S}$. This project assumes that a discretely distributed set of scenarios are available as input to the decomposition algorithm, either as a result of sampling, or by using all possible scenarios, or other scenario generation approaches [9] extrinsic to the algorithm in the computational studies. The focus of this study was on algorithm design and development as opposed to statistical inference.

The following subsections describe the results of this study on selected network problems to provide an overview of the technical progress achieved. Problems were chosen to balance multiple criteria: relevance to DoE (potential application to power grids, bioinformatics); node variables (cluster detection) and edge variables (network design, flow) based formulations; hereditary (s-plexes, s-defective cliques) and nonhereditary (k-clubs, quasi-cliques) problems were considered.

2.2 Resilient Network Design: The Minimum Spanning k-Core Problem

The well-known *survivable network design problem* studied by Steiglitz [33] and others [21] investigating telecommunication network design requires a specified number of redundant edge or vertex disjoint paths between every pair of vertices, or between those in a designated set. This ensures that communication is not interrupted if a link or node fails destroying a communication path. Critical communication and transmission networks deployed in adverse environments like a disaster or a battlefield have to not only survive limited failures, but also preserve functionality such as 2-hop distances between nodes to remain mission capable. This motivated the notion of *resilient network design*, wherein not only are redundant disjoint paths required, but those with short lengths must be present. This section summarizes the results on a new combinatorial optimization problem introduced in this study for applications that require a resilient network design.

Given a graph $G = (V, E)$, denote vertex connectivity by $\kappa(G)$, the diameter by $diam(G)$ and the minimum degree by $\delta(G)$. A graph G is called a *k-core* if $\delta(G) \geq k$. The k-core concept was introduced in *social network analysis* [29] to identify regions of the social network containing *cohesive subgroups* [41]. The connectivity and diameters of k-cores were investigated in [29] from which the following corollary of particular interest to this study was derived.

Corollary 1 *For $r \geq 2$ and $k = \lceil \frac{n+r-2}{2} \rceil$, if $G = (V, E)$ is a k-core on n vertices then, $diam(G) \leq 2$ and $\kappa(G) \geq r$. Furthermore, for any $v \in V$,*

1. *$G - v$ is a $(k-1)$-core;*
2. *$\kappa(G - v) \geq r - 1$;*
3. *$diam(G - v) \leq 2$.*

By Corollary 1, if k is fixed to be $\lceil \frac{n+r-2}{2} \rceil$, the graph obtained by deleting a vertex from k-core G will have vertex-connectivity at least $r - 1$, and preserve the diameter bound of at most 2. In particular, when $r = 2$ and $k = \lceil \frac{n}{2} \rceil$, G is a 2-connected, diameter-2, k-core which upon single vertex (or edge) deletion still has diameter two, i.e., a resilient two-hop network design.

Consider a graph $G = (V, E)$, not necessarily complete, where V denotes the set of vertices of the network to be designed and E represents the set of edges E that can be created. Further suppose the cost c_e of creating an edge $e \in E$ and an appropriately chosen non-negative integer k are available. Consider the following optimization problem:

$$\min_{x \in \{0,1\}^{|E|}} \Big\{ \sum_{e \in E} c_e x_e \mid \sum_{e \in \gamma(v)} x_e \geq k, \ \forall \ v \in V \Big\}, \tag{5}$$

where, $\gamma(v) \subseteq E$ denotes the set of edges incident at v. A feasible solution to this formulation is an incidence vector of a subset of edges, which along with V represents some spanning subgraph of $G = (V, E)$ with minimum degree at least k, i.e., a spanning k-core of G. Hence, (5) formulates *the minimum spanning k-core problem (MSkCP)*, which is to identify an $E^* \subseteq E$ such that $G^* = (V, E^*)$ is a k-core and the total cost of the edges in E^* is a minimum. The MSkCP can be solved in polynomial time by reduction to the maximum weighted b-matching problem [20].

The motivation behind resilient network design problems is the expectation that some of the links or nodes in the network could fail. Hence, it is reasonable to not only build resilience through

redundancy in the design, but also employ optimization models that explicitly capture the uncertainty associated with the failure network components. In this setting, the degree of vertex $v \in V$ in any design specified by the binary vector $x \in \{0,1\}^{|E|}$ is a random variable given by,

$$d(v) = \sum_{e \in \gamma(v)} x_e \mathbf{y}_e.$$

Here, $\gamma(v)$ denotes the set of edges that are incident at v in support graph G. The *degree deficiency* quantified as $[k - d(v)]^+ = \max\{k - d(v), 0\}$, is a measure of loss due to uncertainty at vertex v, leading to the following two loss functions.

$$L^1(x, \mathbf{y}) = \sum_{v \in V} [k - \sum_{e \in \gamma(v)} x_e \mathbf{y}_e]^+ \tag{6}$$

$$L^2(x, \mathbf{y}) = \max_{v \in V} [k - \sum_{e \in \gamma(v)} x_e \mathbf{y}_e]^+ \tag{7}$$

Both loss functions quantify the violation of the minimum degree requirement due to edge failures; L^1 measures the *cumulative degree deficiency* which is proportional to the average degree deficiency and L^2 measures the *maximum degree deficiency*. Both loss functions are piecewise linear and convex with respect to decision vector x. The following results from this study show that a pseudo-polynomial number of samples are sufficient to approximate CVaR within any desired tolerance for the loss functions we use.

Theorem 1 *Consider the loss function L^1 along with its corresponding scenario-based CVaR function $\tilde{F}_\alpha(x, \zeta)$. For any fixed $\epsilon, \beta \in \mathbb{R}$ such that $\epsilon \geq 0$, and $\beta \in (0, 1]$, the required number of scenarios to guarantee that $\mathbb{P}\{|\tilde{F}_\alpha(x, \zeta) - F_\alpha(x, \zeta)| \leq \epsilon\} \geq 1 - \beta$ is $O(|V|^2 k^2)$.*

Theorem 2 *Consider the loss function L^2 along with its corresponding scenario-based CVaR function $\tilde{F}_\alpha(x, \zeta)$. For any fixed $\epsilon, \beta \in \mathbb{R}$ such that $\epsilon \geq 0$, and $\beta \in (0, 1]$, the required number of scenarios to guarantee that $\mathbb{P}\{|\tilde{F}_\alpha(x, \zeta) - F_\alpha(x, \zeta)| \leq \epsilon\} \geq 1 - \beta$ is $O(k^2)$.*

The next significant result of this study is the development of a reformulation that is more amenable to decomposition and leading to a branch-and-cut approach to solve the CVaR-constrained MSkCP based on $\mathcal{S}$:

$$\min_{x \in \mathcal{P},\ \zeta \in \mathbb{R}} \Big\{ \sum_{e \in E} c_e x_e \mid \zeta + \frac{1}{(1-\alpha)|\mathcal{S}|} \sum_{s \in \mathcal{S}} (L(x,\ y^s) - \zeta)^+ \leq C \Big\}, \tag{8}$$

where $\mathcal{P} = \Big\{ x \in \{0,1\}^{|E|} \mid \sum_{e \in \gamma(v)} x_e \geq k, v \in V \Big\}$ and C is a user specified bound.

Following the pioneering work done by Rockafellar and Uryasev [28, 27], several notable developments on computational aspects of bounding or minimizing CVaR have appeared in recent literature [15, 16, 39, 10, 40, 34, 12, 14, 13]. In particular, Künzi-Bay and Mayer [16] study CVaR minimization with linear loss functions in financial applications. They reformulate the CVaR minimization problem as a two-stage stochastic programming problem with recourse [4] and specialize the L-shaped method [36] for their particular structure. A central result in their study [16] is a polyhedral characterization of the CVaR constraint under a *linear loss function* through exponentially many constraints.

Theorem 3 ([16]) *Define $\mathcal{Q}_1$ and $\mathcal{Q}_2$ as following.*

$$\begin{aligned}\mathcal{Q}_1 &:= \{(x,\zeta) \mid \zeta + \frac{1}{(1-\alpha)|\mathcal{S}|}\sum_{s\in\mathcal{S}}[L(x,\ y^s) - \zeta]^+ \leq C\}\\ \mathcal{Q}_2 &:= \bigcap_{\mathcal{A}\subset\mathcal{S}} \{(x,\zeta) \mid \zeta + \frac{1}{(1-\alpha)|\mathcal{S}|}\sum_{s\in\mathcal{A}}[L(x,\ y^s) - \zeta] \leq C\}\end{aligned}$$

with the sum defined as zero for $\mathcal{A} = \emptyset$ in $\mathcal{Q}_2$. Then $\mathcal{Q}_1 = \mathcal{Q}_2$.

Note the absence of $[\cdots]^+$ in $\mathcal{Q}_2$. Theorem 3 was introduced in the context of linear loss functions, while true for any loss function, the set $\mathcal{Q}_2$ is a polyhedron when the loss function is linear (and $\mathcal{S}$ is finite). We refer to the following as the Künzi-Bay and Mayer reformulation (KBMR) of the CVaR-constrained MSkCP.

$$\textbf{(KBMR)} \quad \min_{x\in\mathcal{P},\ \zeta\in\mathbb{R}} \Big\{ \sum_{e\in E} c_e x_e \mid \zeta + \frac{1}{(1-\alpha)|\mathcal{S}|}\sum_{s\in\mathcal{A}}[L(x,\ y^s) - \zeta] \ \leq\ C,\ \forall\ \mathcal{A}\subset\mathcal{S}\Big\} \qquad (9)$$

In KBMR, the single nonlinear CVaR constraint in (8) is replaced with exponentially many constraints. Similar to RUR, KBMR is also an MIP when the convex piecewise linear loss function is also linearized using additional variables. Although KBMR has $O(2^{|\mathcal{S}|})$ constraints compared to RUR which has $O(|\mathcal{S}|)$, this reformulation enables the adoption of a row-generation framework such as Benders decomposition [3]. This generic row-generation framework can be described as follows. A master relaxation of KBMR with a subset of the constraints and is solved to optimality with optimal solution (x^*, ζ^*). Let, $\mathcal{A}^* = \{s \in \mathcal{S} \mid L(x^*, y^s) - \zeta^* > 0\}$. If the constraint for $\mathcal{A}^*$ is satisfied, then no violated constraint exists; (x^*, ζ^*) is optimal to KBMR and x^* is optimal to the CVaR-constrained MSkCP. Otherwise, the violated constraint corresponding to $\mathcal{A}^*$ must be added to the master relaxation and the process repeated until an optimal solution is found or the master relaxation becomes infeasible.

This sequential cutting plane method is correct, and finite since $\mathcal{S}$ is finite and no subset of $\mathcal{S}$ will be repeated [16]. It was also found to be computationally effective in [16], which is due at least in part to the following characteristics:

1. The $\mathcal{A}^*$ that is found corresponds to the most violated KBMR constraint (if one exists).
2. Since the variables were continuous and the loss function linear in [16], the master relaxation solved in each major iteration was a large-scale linear program. Since x is binary in this study, using this approach as-is requires the solution of an MIP in each major iteration as the master relaxation.
3. Since the loss function $L(.)$ considered in [16] is linear, the violated KBMR constraint that is added is linear and no additional variables are required to linearize $L(.)$ as it would be necessary when the loss function is piecewise linear as in this study.

This study developed two key computational ideas to design an effective exact algorithm to solve the CVaR-constrained MSkCP. First, the reformulation ideas of Künzi-Bay and Mayer [16] and Haneveld and van der Vlerk [11] were extended. Specifically, an equivalent reformulation for the piecewise linear loss functions L^1 and L^2 that uses linear constraints was developed (but much

more than $2^{|\mathcal{S}|}$ are used). This also enabled the detection of the most violated extended-KBMR constraint as before. The details of this reformulation can be found in [20]. Secondly, the binary restriction on x variables discourages the use of a sequential cutting plane method in favor of a branch-and-cut (BC) algorithm. This study integrates the reformulation and decomposition ideas into the BC, leading to an integer programming-based approach for the CVaR-constrained MSkCP, in the same vein as those developed recently for chance constrained optimization in [18, 31].

The computational advances made for this problem through this project are directly applicable to any CVaR constrained optimization problem where the loss function is derived as cumulative or maximum constraint violation. Hence, the main results of this study can be leveraged well beyond the context in which they were developed. Furthermore, it was observed that much larger instances could be solved to optimality under a CVaR constraint in [20] compared to a chance-constrained version of MSkCP in [19]. This further reinforces the basic premise of this project that CVaR optimization presents a viable alternative to chance-constrained and robust optimization in applications where this modeling framework is appropriate. The details excluded from this section can be found in the paper that will be submitted for publication in Jan 2014 [20] to a *Math. Programming, Ser. B* special issue on *Integer Programming Under Uncertainty* and the manuscript will be available immediately after upon request.

2.3 Robust Flows: The Fixed Charge Network Flow Problem

Fixed charge network flow (FCNF) problems attract a lot of attention among researchers and practitioners, since a variety of problems, such as single source problem, lot-sizing, network design, investment decision, pricing policy, scheduling, and other problems that arise in the supply chain, logistics, transportation science, and information networks can be modeled as a FCNF problem. Uncertainties in fixed charge network flow problems have been mainly addressed in literature from the viewpoint of uncertain demands and costs. Although these parameters are clearly important, another crucial issue that needs to be addressed in many real-world situations is the presence of potential disruptions/failures of network arcs, which are uncertain by nature and may be caused by natural or man-made factors.

In the context of transportation, telecommunication and power transmission network infrastructure, uncertain disruptions may be caused by severe weather conditions (winter storms, extreme heat) or by network elements failing due to overload or wear. The main difference in the problem setup with multiple uncertain arc failures from the previously considered problem formulations with uncertain demands and/or costs is the fact that the topology of the network itself is deterministic which creates challenges in terms of defining "optimal robust flow assignments"; since the obtained optimal solutions need to be not only cost-efficient, but also robust with respect to potential multiple arc failures. These failures can cause the loss of flow through a network, and the questions that need to be rigorously addressed are: 1) How to quantify the risks of collateral losses of flow associated with uncertain arc failures? 2) How to efficiently minimize or restrict these losses in the framework of large-scale fixed charge network flow problems?

In general, an instance of the MINIMUM COST FLOW PROBLEM (MCF) is represented by a directed graph $G = (V, E)$, where each directed edge (arc) $(i, j) \in E$ has the associated cost c_{ij} per unit of flow along this edge, as well as the capacity λ_{ij} denoting the maximum amount of flow that can traverse the edge (i, j). For each node $i \in V$, d_i denotes the demand/supply of node i. In practice, it is often required to pay a certain fixed setup cost (fixed charge) in order to have a nonzero flow through a certain arc. Let s_{ij} denote the fixed setup cost for arc (i, j). Then if x_{ij} is the flow

through this arc, the total cost associated with this arc is

$$f_{ij}(x_{ij}) = \begin{cases} c_{ij}x_{ij} + s_{ij}, & x_{ij} \in (0, \lambda_{ij}] \\ 0, & x_{ij} = 0. \end{cases}$$

The standard deterministic FIXED CHARGE NETWORK FLOW PROBLEM (FCNF) can then be formulated as

$$\textbf{(FCNF-IP)} \quad \min \quad \sum_{(i,j)\in E} (c_{ij}x_{ij} + s_{ij}z_{ij}) \tag{10}$$

s.t.

$$\sum_{\{j:(i,j)\in E\}} x_{ij} - \sum_{\{j:(j,i)\in E\}} x_{ji} = d_i, \ \forall\ i \in V, \tag{11}$$

$$0 \le x_{ij} \le \lambda_{ij} z_{ij}, \ \forall\ (i,j) \in E, \tag{12}$$

$$z_{ij} \in \{0,1\}, \ \forall\ (i,j) \in E. \tag{13}$$

Due to the fact that multiple arcs may fail under the assumptions of our model, with these failures being uncertain, we need to introduce the loss function that quantifies the losses of flow associated with arc failures. As before, let $\mathbf{y}$ be a random vector representing the uncertain arc failures in the network. In the models below, we assume that each arc can either be fully functional or completely fail. In this case, each component $\mathbf{y}_{ij}$ of the vector $\mathbf{y}$ is a Bernoulli random variable defined as follows:

$$\mathbf{y}_{ij} := \begin{cases} 1, \text{ with probability } p_{ij}, \\ 0, \text{ with probability } 1 - p_{ij}, \end{cases}$$

where p_{ij} is the probability of failure for the arc (i,j). Furthermore, based on the distributions of $\mathbf{y}_{ij}$, one can generate a set of *failure scenarios* $s = 1, 2, \ldots, N$, where the values y^s_{ij} are defined for each scenario as follows:

$$y^s_{ij} := \begin{cases} 1, \text{ if arc } (i,j) \text{ fails under scenario } s, \\ 0, \text{ otherwise.} \end{cases} \tag{14}$$

Then, in the simplest case the random variable $L(x, \mathbf{y})$ representing the total loss can be defined as the following linear expression:

$$L(x, \mathbf{y}) = \sum_{(i,j)\in E} \beta_{ij} x_{ij} \mathbf{y}_{ij}. \tag{15}$$

Note that in this definition we introduce the coefficients β_{ij} corresponding to each arc (i,j), which can be viewed as either *arc importance coefficients* or *rerouting cost coefficients*. The values of arc importance coefficients may differ depending on specific applications and provide a certain degree of flexibility for the quantitative expression of the collateral losses associated with the failures of different arcs. This interpretation is appropriate, for instance, in the case of the classical transportation problem setup with fixed costs and uncertain arc failures; however, in a more general case, these coefficients can also be viewed as rerouting cost coefficients that show cost of flow rerouting to an alternative path. Exact computation of these coefficients for specific network topologies and

all failure scenarios might be challenging; however, some estimations of these coefficients can be obtained from historical data. In this case one can assume that rerouting cost may be different for different arcs and linearly proportional to the original flow x_{ij}. In the following discussion, without loss of generality, we set $\beta_{ij} = 1$, $\forall\ (i,j) \in E$; however, the approaches developed will continue to hold for arbitrary values of β_{ij}. Then we can model the CVaR-constrained fixed charge network flow problem using the Rockafellar-Uryasev reformulation as shown below.

$$\textbf{(CVaR FCNF)} \qquad \min \sum_{(i,j)\in E} (c_{ij}x_{ij} + s_{ij}z_{ij}) \tag{16}$$

s.t.

$$\sum_{\{j:(i,j)\in E\}} x_{ij} - \sum_{\{j:(j,i)\in E\}} x_{ji} = d_i,\ \forall\ i \in V, \tag{17}$$

$$0 \le x_{ij} \le \lambda_{ij} z_{ij},\ \forall\ (i,j) \in E,$$

$$\eta + \frac{1}{1-\alpha} \sum_{s=1}^{N} \pi_s t_s \le C, \tag{18}$$

$$t_s \ge L(x, y^s) - \eta,\ \forall\ s = 1, 2, \ldots, N, \tag{19}$$

$$t_s \ge 0,\ \forall\ s = 1, 2, \ldots, N. \tag{20}$$

$$z_{ij} \in \{0, 1\},\ \forall\ (i,j) \in E, \tag{21}$$

$$\eta \in \mathbb{R} \tag{22}$$

$$\tag{23}$$

Note that the exact decomposition based BC algorithm developed for the CVaR-constrained MSkCP is also applicable in this setting. In fact, the KBMR reformulation itself would be adequate since the loss function employed is linear in x. However, both deterministic and CVaR-based versions of FCNF problems are NP-hard, and this study approached this problem with a problem-specific heuristic to solve the CVaR-constrained FCNF. Specifically, an approach previously developed for the deterministic FCNF problem was extended to the CVaR-constrained version and numerical studies demonstrate that the proposed heuristic often produces near-optimal solutions for problem instances much larger than those previously addressed in the literature (in both deterministic and stochastic settings). The paper describing these results was published in *Journal of Combinatorial Optimization* [32] and listed in Section 5.

2.4 Cluster Detection: Hereditary Clique Relaxations

Increasing interest in studying *community structures*, or *clusters* in complex networks arising in various applications has led to a large and diverse body of literature introducing numerous graph-theoretic models relaxing certain characteristics of the classical clique concept. This project analyzed the elementary clique-defining properties implicitly exploited in the available *clique relaxation* models and proposed a taxonomic framework that not only allows to classify the existing models in a systematic fashion, but also yields new clique relaxations of potential practical interest. These results were published in [26].

In particular, some basic structural properties of several of the considered models may dictate the choice of methods for solving the corresponding optimization problems. This structural analysis is centered around the concept of *heredity*, which could be thought of as a dynamic property, since it describes the characteristics of a graph undergoing a change, i.e., vertex addition or removal. Heredity is defined with respect to a graph property Π. A graph property Π is said to be *hereditary on induced subgraphs*, if for any graph G with property Π the deletion of any subset of vertices

does not produce a graph violating Π. The presence of heredity on induced subgraphs implies certain properties that may help streamlining the study of the corresponding optimization problems. In particular, it turns establishing the computational intractability of the problem into a simple exercise of checking several basic facts about the property Π. Namely, a property Π is called *nontrivial* if it is true for a single-vertex graph and is not satisfied by every graph, and Π is called *interesting* if there are arbitrarily large graphs satisfying Π.

We observe that heredity on induced subgraphs is the foundational property for some of the most successful combinatorial algorithms for the maximum clique problem, which can be generalized to solve any other maximum RELAXED CLIQUE problem based on relaxed clique-defining properties that are hereditary on induced subgraphs. We exploit this observation to develop effective exact algorithms for the maximum s-plex and s-defective clique problems in [35]. Furthermore, heredity is preserved in the case with uncertain edge failures under appropriate risk measures (VaR, CVaR), meaning that the proposed algorithms can be extended to compute large robust clusters under uncertainty. This approach is explored in [42], which also proposes mixed integer programming formulations for the problem of computing a maximum clique subject to CVaR constraints for different loss functions.

Even though the properties defining some other clique relaxation models do not posses heredity, they have closely related characterizations that can also be utilized in designing solution methods. We defined these dynamic properties of *weak heredity, quasi-heredity*, and *k-heredity* as follows. A graph property Π is said to be *weakly hereditary*, if for any graph $G = (V, E)$ with property Π all subsets of V demonstrate the property Π in G. A graph property Π is said to be *quasi-hereditary*, if for any graph $G = (V, E)$ with property Π and for any size $0 < r < |V|$, there exists some subset $R \subset V$ with $|R| = r$, such that $G[V \setminus R]$ demonstrates property Π. Finally, a graph property Π is said to be *k-hereditary on induced subgraphs*, if for any graph G with property Π the deletion of any subset of vertices with up to k vertices does not produce a graph violating Π.

Note that weak heredity considers whether a certain property is still applicable for all subsets in the original graph, as opposed to heredity on the induced subgraph. On the other hand, quasi-heredity essentially requires the existence of a sequence of vertices such that their removal in this sequence preserves, at every step of the vertex removal process, the property in the remaining subgraph. However, property Π may not be preserved for every sequence of vertex deletions. It also follows that heredity implies both weak heredity and quasi-heredity, whereas the latter two do not appear to have any definitive relation.

A systematic and rigorous analysis of cohesiveness properties of clique relaxations led us to introduce a new concept of k-community, which can be viewed as a stronger version of k-core. This structure proved to be extremely useful for solving the maximum clique and vertex coloring problems on very large-scale real life networks [37]. Due to the size of such networks and the intractability of the considered problems, previously developed exact algorithms may not be directly applicable. The proposed community-based approaches aim to reduce the network instances to a size that is tractable for existing solvers, while preserving optimality. Experimental results on real life graphs (collaboration networks, P2P networks, social networks, etc.) show the proposed procedures to be effective by finding, for the first time, exact solutions for instances with over 18 million vertices [37]. In addition, a network clustering algorithm based on k-community was competitive with other approaches presented in DIMACS Implementation Challenge [38].

In addition, in [30] we introduced and studied the maximum ratio clique problem, which is defined as follows. Given a simple undirected graph $G = (V, E)$, where each vertex $i \in V$ is assigned two non-negative rational weights, a_i and b_i, the maximum ratio clique problem (MRCP) is to find a

maximal clique C in G that maximizes the quantity $\frac{\sum_{i \in C} a_i}{\sum_{i \in C} b_i}$. A fractional objective arises in network-based location models, such as establishment of wind farms. To minimize the effects of wind speed variability, it is required that wind farms be located at places that have a negative correlation with each other in terms of wind speed over time. However, the costs of setting up wind farms might vary widely depending on factors like the size of the farm, the height of wind turbines, and geographical factors. High wind speeds are required to get the maximum possible power output in a wind farm. The problem of finding appropriate locations for wind farms that maximize the overall energy output per dollar invested can be modeled as the problem of finding a maximal clique, where the vertices of the graph represent different possible wind farm configurations for the prospective locations of interest, and the edges correspond to pairs of locations with negatively correlated wind speeds. Then a maximum ratio maximal clique has the highest value of the ratio of the net wind speed to the net cost. In [30], we developed and tested several exact algorithms for the maximum ratio clique problem.

2.5 Cluster Detection: The Maximum k-Club Problem

Consider a simple, undirected graph $G = (V, E)$ on n vertices, and let $d_G(u, v)$ denote the length of a shortest path between vertices u and v in G. Denote the diameter as before by $diam(G) = \max\{d_G(u, v) : u, v \in V\}$. A subset of vertices S forms a k-clique if $d_G(u, v) \leq k$ for all $u, v \in S$ [17] and it is called a k-club if $d_{G[S]}(u, v) \leq k$ for all $u, v \in S$, i.e., $diam(G[S]) \leq k$. By definition, 1-clique and 1-club correspond to the classical clique- the gold standard for cluster models. Clearly, every k-club is also a k-clique as $d_{G[S]}(u, v) \leq k \Rightarrow d_G(u, v) \leq k$. The converse however, is not true.

The k-clique model was introduced in social network analysis to model *cohesive social subgroups*. The k-club model was introduced to overcome the drawback of k-cliques which allowed members outside the cohesive subgroup to be used on the short paths between members inside the group. The concept of k-club overcomes this weakness of k-cliques by bounding pairwise distances in the induced subgraph. As a result, every subset of a k-club (for $k \geq 2$) is not necessarily k-club, although every subset of a k-clique continues to satisy the k-clique definition. The k-club property is not hereditary on induced subgraphs as the diameter of the induced subgraph could increase upon vertex deletion whereas the k-clique property is weakly hereditary as pairwise distances in the original graph are unaffected by vertex deletions.

Note that $S \subseteq V$ is a k-clique in G, if and only if S is a clique in the power graph G^k, which contains an edge for every distinct $u, v \in V$ such that $d_G(u, v) \leq k$. Hence, the maximum k-clique problem for any k is equivalent to the maximum clique problem on the corresponding power graph. The k-clubs, however, due to their nonhereditary nature, pose significant new challenges to mathematical analysis despite the seemingly minor difference in the definition. This project undertook a study of this problem with an eye towards future investigation into a CVaR-constrained version of this problem. During the project duration, some significant complexity and polyhedral results were obtained that we summarize next.

A direct consequence of the nonhereditary nature of k-clubs is the following intractability result established by PI Balasundaram and his doctoral student supported by this grant [24]. This is to the best of our knowledge, the only known complexity result on testing the inclusionwise maximality of any nonhereditary property, and this question has remained open since the introduction of the problem to the optimization community in 2000 [5].

Theorem 4 ([24]) *Given a graph $G = (V, E)$, a fixed integer $k \geq 2$, and a nonempty k-club D, verifying whether D is an inclusionwise maximal k-club is NP-complete.*

In order to find a maximal (by inclusion) k-clique, we can start with a set that is initially empty and add a new vertex to this set so that the resulting set is a k-clique. We repeat this until no such vertex exists, and the current set is a maximal k-clique. But this simple approach does not work for k-clubs which are not hereditary. Unlike hereditary properties, nonexistence of a single "addable" vertex outside the current set is not equivalent to maximality. Furthermore, it is possible that no maximal k-clique in a graph forms a k-club (see [2]) implying that even complete enumeration of all maximal k-cliques can fail to identify a single k-club. As a result, well-known combinatorial algorithms for the maximum clique problem like the Carraghan-Pardalos algorithm [7] and the Östergård algorithm [23] do not directly extend to the maximum k-club problem for $k \geq 2$. A new combinatorial branch-and-bound algorithm was hence developed in [24]. These results were published in *Discrete Optimization* and is listed in Section 5.

The polyhedral combinatorics of the special case- *the 2-club polytope* (convex hull of characteristic binary vectors of 2-clubs in the graph) was also investigated as a part of this grant leading to two major results. First, a much broader class of facet defining inequalities for the 2-club polytope were discovered that include all previously known facets as special cases [25] and its separation problem was also shown to be NP-hard. Second, the polytope that is described using all such inequalities along with non-negativity was shown to be the 2-club polytope of acyclic graphs producing a complete minimal description for this class. Furthermore, the describing system was also shown to be totally dual integral. These polyhedral results are part of the article that is in its final stages of preparation for submission to *Mathematical Programming* in January 2014; a manuscript will be available upon request after that time. As part of his doctoral dissertation supported by this grant, Foad Mahdavi Pajouh, doctoral student of PI Balasundaram who developed the aforementioned results also investigated new dual decomposition techniques to solve the CVaR-constrained maximum 2-club problem. The details of this work are available in his dissertation and it is listed in Section 5. This work is still in progress as we work to extend the approach to general k under CVaR constraints before an article is developed for publication.

3 Key Accomplishments

This project focused on a family of combinatorial optimization problems on stochastic networks that sought risk-averse solutions to network flow, network design, and cluster detection problems by employing a coherent measure of downside risk that is widely used in financial applications like portfolio allocation and optimization. The technical contributions of this project add to the growing knowledge-base on modeling and optimization of complex networks and uncertainty quantification; two major thrust areas of the ASCR program.

The overarching mission of the US DoE to ensure energy security through transformative science and technology solutions relies heavily on fundamental mathematical advances that facilitate a better understanding of complex networked systems that govern and control the generation, storage and distribution of energy from renewable and nonrenewable sources. The big-data analytics needed for transformative impact will benefit from this project both in terms of the high level concepts investigated (network models of complex systems, modeling losses and quantifying risk) and the mathematical theory and algorithms developed with support from this grant. The following is a brief summary of accomplishments that were enabled by this grant.

Research.

- 15 refereed journal papers published or accepted; 1 submitted and 4 in preparation for submission in January 2014

- 4 book chapters and 2 conference proceedings published
- 4 doctoral dissertations completed, 2 in progress and 1 masters thesis completed

Education and Outreach.

- PI Balasundaram developed a doctoral course on stochastic optimization with an emphasis on optimization of conditional value-at-risk at Oklahoma State University
- PI Balasundaram gave an invited seminar in the special lectures series at the Department of Computational and Applied Mathematics, Rice University (Nov 2011)
- PI Balasundaram, Co-PI Boginski and Co-PI Butenko served on the advisory board of the "Systems and Optimization Aspects of Smart Grid Challenges" conference at the University of Florida (April 2011)

4 Research Transition

- Our work on the decomposition algorithms for the optimization of CVaR has led us to explore delayed and optimal scenario generation techniques as an alternative to exterior sampling.
- A community-based network clustering algorithm by Verma and Butenko [38] was included in 10-th DIMACS Implementation Challenge (http://www.cc.gatech.edu/dimacs10/).
- PI Balasundaram is a co-PI on the Oklahoma Transportation Center equipment grant titled *Oklahoma Center for Transportation and Logistics Research, Education and Outreach* (PI: Ingalls, co-PIs: Kamath, Balasundaram, Liu, 7/1/2011–6/30/2013, $150,000). This grant supported the purchase of a high-performance computing cluster "Cimarron" (10 high performance compute nodes with dual quad core Intel E5620 processors; 6 have 96 GB RAM and 4 have 144 GB RAM). The research supported by the DoE grant was a great driver for the use high performance computing and several of the publications resulting from the DoE grant also utilized the Cimarron cluster for numerical experiments.
- PI Balasundaram's research work on the DoE grant was included among the work that was identified as those to benefit from an enhanced university-wide HPC infrastructure upgrade in a proposal by the OSU HPC director Dr. Dana Brunson who received a National Science Foundation (MRI) grant to acquire the current "Cowboy" cluster in use at Oklahoma State University. "Cowboy" consists of 254 compute nodes, each with dual hex-core Intel Xeon E2620 "Sandy Bridge" processors and 32 GB memory for an aggregate total peak speed of 49 TFLOPs (8x our previous supercomputer).

5 Comprehensive List of Project Outcomes

The following is a comprehensive list of all publications, presentations, M.S. theses and Ph.D. dissertations that were published (or nearing completion) that were at least partially supported by this grant. Graduate student names are underlined.

5.1 Refereed Journal Articles

1. F. Mahdavi Pajouh, B. Balasundaram and I. V. Hicks. On the 2-club polytope of graphs. In preparation for submission to *Mathematical Programming, Series A*, Jan 2014.

2. J. Ma, F. Mahdavi Pajouh, B. Balasundaram and V. Boginski. The minimum spanning k-core problem with bounded CVaR under probabilistic edge failures. In preparation for submission to *Mathematical Programming, Series B*, Special Issue on *Integer Programming Under Uncertainty*, Jan 2014.

3. K. Pavlikov, A. Veremyev, V. Boginski and S. Uryasev. Identifying Critical Nodes in Coupled Interdependent Networks with Uncertain Failure Propagations using Conditional Value-at-Risk. In preparation for submission to *Journal of Complex Networks*, Jan 2014.

4. O. Yezerska, S. Butenko, and V. Boginski. Detecting robust cliques in graphs with uncertain edge failures. In preparation.

5. A. Verma, A. Buchanan, and S. Butenko. Solving the maximum clique and vertex coloring problems on very large sparse networks. Submitted to *INFORMS Journal on Computing*.

6. Z. Miao, B. Balasundaram and E. L. Pasiliao. An exact algorithm for the maximum probabilistic clique problem. To appear in *Journal of Combinatorial Optimization*, DOI: 10.1007/s10878-013-9699-4.

7. F. Mahdavi Pajouh, Z. Miao and B. Balasundaram. A branch-and-bound approach for maximum quasi-cliques. To appear in *Annals of Operations Research*, DOI 10.1007/s10479-012-1242-y.

8. A. Buchanan, J. S. Sung, V. Boginski, and S. Butenko. On connected dominating sets of restricted diameter. To appear in *European Journal of Operational Research*, DOI: 10.1016/j.ejor.2013.11.036.

9. J. Pattillo, Y. Wang, and S. Butenko. Approximating 2–cliques in unit disk graphs. To appear in *Discrete Applied Mathematics*, DOI: 10.1016/j.dam.2013.10.017.

10. S. Sethuraman and S. Butenko. The maximum ratio clique problem. To appear in *Computational Management Science*, DOI: 10.1007/s10287–013–0197–z. **Runner–up, Best Student Paper Award, CMS–2013 Conference**.

11. O. Shirokikh, A. Sorokin, and V. Boginski. A note on transmission switching in electric grids with uncertain line failures. To appear in *Energy Systems*, DOI: 10.1007/s12667-013-0087-z.

12. S. Trukhanov, C. Balasubramaniam, B. Balasundaram and S. Butenko. Algorithms for detecting optimal hereditary structures in graphs, with application to clique relaxations. *Computational Optimization and Applications*, 56(1):113–130, 2013.

13. J. Pattillo, N. Youssef, and S. Butenko. On clique relaxation models in network analysis. *European Journal of Operational Research*, 226(1): 9–18, 2013.

14. S. Shahinpour and S. Butenko. Algorithms for the maximum k–club problem in graphs. *Journal of Combinatorial Optimization*, 26: 520–554, 2013.

15. F. Mahdavi Pajouh, B. Balasundaram and O. Prokopyev. On characterization of maximal independent sets via quadratic optimization. *Journal of Heuristics, Special Issue on Unconstrained Quadratic Binary Optimization*, 19(4):629–644, 2013.

16. S. Butenko, O. Yezerska and B. Balasundaram. Variable objective search. *Journal of Heuristics, Special Issue on Unconstrained Quadratic Binary Optimization*, 19(4):697–709, 2013.

17. A. Sorokin, V. Boginski, A. Nahapetyan, and P.M. Pardalos. Computational risk management techniques for fixed charge network flow problems with uncertain arc failures. *Journal of Combinatorial Optimization*, 25: 99–122, 2013.

18. J. Pattillo, A. Veremyev, S. Butenko, and V. Boginski. On the maximum quasi-clique problem. *Discrete Applied Mathematics*, 161: 244–257, 2013.

19. F. Mahdavi Pajouh and B. Balasundaram. On inclusionwise maximal and maximum cardinality k-clubs in graphs. *Discrete Optimization*, 9(2):84–97, 2012.

20. A. Verma and S. Butenko. A distributed approximation algorithm for the bottleneck connected dominating set problem. *Optimization Letters*, 6: 1583–1595, 2012.

5.2 Refereed Book Chapters

1. B. Balasundaram and F. Mahdavi Pajouh. Graph theoretic clique relaxations and applications. *In:* P. M. Pardalos, D-Z. Du and R. Graham (Eds) *Handbook of Combinatorial Optimization, 2nd Edition*, pages 1559–1598. Springer, 2013, DOI 10.1007/978-1-4419-7997-1_9.

2. A. Verma and S. Butenko. Network clustering via clique relaxations: a community–based approach. In: *Graph Partitioning and Graph Clustering.* Ed. by D. A. Bader, H. Meyerhenke, P. Sanders, and D. Wagner. American Mathematical Society, pp.125–136, 2013.

3. S. Shahinpour and S. Butenko. Dinstance–based clique relaxations in networks: s–cliques and s–clubs. In: *Models, Algorithms, and Technologies for Network Analysis.* Ed. by B. I. Goldengorin, V. A. Kalyagin, and P. M. Pardalos. Vol. 59. Springer Proceedings in Mathematics & Statistics. New York: Spinger Science+Business Media, pp.149–174, 2013.

4. J. Pattillo, N. Youssef, and S. Butenko. Clique relaxations in social network analysis. In: *Handbook of Optimization in Complex Networks: Communication and Social Networks.* Ed. by M. T. Thai and P. M. Pardalos. New York: Spinger Science + Business Media, pp.143–162, 2012.

5.3 Refereed Conference Proceedings

1. J. Ma and B. Balasundaram. Solving chance-constrained spanning k-core problem via decomposition and integer programming. In: Proceedings of the 2013 Industrial and Systems Engineering Research Conference (ISERC 2013). **Recipient of the 2013 ISERC Best Paper Award (Operations Research track) from the IIE Operations Research Division.**

2. Z. Miao and B. Balasundaram. Cluster detection in large-scale social networks using k-plexes. In: Proceedings of the 2012 Industrial and Systems Engineering Research Conference (ISERC 2012).

5.4 Doctoral Dissertations

1. Juan Ma (B. Balasundaram, Co-chair, In progress, December 2014), Dissertation: *Resilient network design under uncertainty: CVaR and chance-constrained optimization*

2. Oleksandra Yezerska (S. Butenko, Chair, In progress, December 2014), Dissertation: *Detecting robust clusters in networks.*

3. Je Sang Sung (S. Butenko, Chair, December 2013), Dissertation: *Generalized Domination in Graphs with Applications in Wireless Networks*, currently with BNSF Railway Co.

4. Foad Mahdavi Pajouh (B. Balasundaram, Chair, August 2012), Dissertation: *Polyhedral combinatorics, complexity & algorithms for k-clubs in graphs*, currently a Research Assistant Professor at University of Florida (REEF).

5. Alexey Sorokin (V. Boginski, Co-Chair, August 2012), Dissertation: *Modeling and optimization techniques for ensuring resiliency in heterogeneous networked systems*, currently with Innovative Scheduling, Inc., Gainesville, FL.

6. Anurag Verma (S. Butenko, Chair, August 2012), Dissertation: *Network Based Approaches for Clustering and Location Decisions*, currently with United Airlines.

5.5 Masters Theses

1. Juliana Bright (B. Balasundaram, Chair, Fall 2011), Thesis: *Robust shortest paths under uncertainty using conditional value-at-risk*, currently a Ph.D. student at University of Arkansas

5.6 Presentations

1. Solving chance-constrained spanning k-core problem via decomposition and integer programming. 2013 Industrial and Systems Engineering Research Conference (ISERC), May 18-22, 2013, San Juan, Puerto Rico, (Presented by coauthor J. Ma).

2. S. Sethuraman and S. Butenko, "The maximum ratio clique problem, 10th International Conference on Computational Management Science, Montréal, May 1–3, 2013.

3. A. Buchanan, J. S. Sung, S. Butenko, and E. L. Pasiliao, An integer programming approach for k-connected m-dominating sets, 1st Annual Meeting of the Mathematical Modeling and Optimization Institute, Shalimar, FL, July 29 – August 1st, 2013.

4. S. Butenko, and J. Pattillo, and Y. Wang,"On distance-based clique relaxations in unit disk graphs", 1st Annual Meeting of the Mathematical Modeling and Optimization Institute, Shalimar, FL, July 29 – August 1st, 2013.

5. C. Balasubramaniam and S. Butenko, "On the minimum k-connected subgraph", 1st Annual Meeting of the Mathematical Modeling and Optimization Institute, Shalimar, FL, July 29 – August 1st, 2013.

6. J. S. Sung, A. Buchanan, V. Boginski, and S. Butenko, On connected dominating sets of restricted diameter, INFORMS Annual Meeting, Minneapolis, MN, October 6–9, 2013.

7. A. Buchanan, J. S. Sung, and S. Butenko, An integer programming approach for k-connected m-dominating sets, INFORMS Annual Meeting, Minneapolis, MN, October 6–9, 2013.

8. Y. Wang, S. Butenko, and J. Pattillo, "On distance-based clique relaxations in unit disk graphs", INFORMS Annual Meeting, Minneapolis, MN, October 6–9, 2013.

9. C. Balasubramaniam and S. Butenko, "On the minimum k-connected subgraph", INFORMS Annual Meeting, Minneapolis, MN, October 6–9, 2013.

10. S. Butenko, "Clique relaxations in unit disk graphs", NATO Advanced Research Workshop on Examining Robustness and Vulnerability of Critical Infrastructure Networks, Kyiv, Ukraine, June 3–5, 2013.

11. V. Boginski, "Diameter-based Robust Clusters in Networks", NATO Advanced Research Workshop on Examining Robustness and Vulnerability of Critical Infrastructure Networks, Kyiv, Ukraine, June 3–5, 2013.

12. V. Boginski, A. Veremyev, S. Butenko, P. Krokhmal, J. Pattillo, and D. Jeffcoat, "Identifying large dense clusters in networks: exact optimal solutions and asymptotic phase transitions", SIAM Workshop on Network Science, San Diego, CA, July 7–8, 2013.

13. S. Butenko, "Clique relaxations in network analysis" (poster presentation), SIAM Workshop on Network Science, San Diego, CA, July 7–8, 2013.

14. Algorithms for computing optimal clique relaxation structures that are hereditary. 2012 INFORMS Annual Meeting, Oct 14–17, 2012, Phoenix, AZ (Presented by coauthor C. Balasubramaniam).

15. On chance-constrained minimum spanning k-core problem under probabilistic edge failure. 2012 INFORMS Annual Meeting, Oct 14–17, 2012, Phoenix, AZ (Presented by coauthor J. Ma).

16. Cluster detection in very large-scale social networks using k-plexes. 2012 Industrial and Systems Engineering Research Conference (ISERC), May 19-23, 2012, Orlando, FL (Presented by coauthor Z. Miao).

17. The maximum k-club problem. 2012 INFORMS Optimization Society Meeting, Feb 24–26, 2012, University of Miami, School of Business Administration, Coral Gables, FL.

18. Butenko, S., "Clique relaxation models in network analysis", NSF Center for Research on Complex Networks, Texas Southern University, Houston, TX, November 8, 2012.

19. Balasubramaniam, C., B. Balasundaram, S. Butenko, and S. Trukhanov, "Algorithms for computing optimal clique relaxation structures that are hereditary", INFORMS Annual Meeting, Phoenix, AZ, October 14–17, 2012.

20. Ertem, Z., and S. Butenko, "Graph theoretical analysis of brain networks", INFORMS Annual Meeting, Phoenix, AZ, October 14–17, 2012.

21. Sethuraman, S., and S. Butenko, "The maximum ratio maximal clique problem", INFORMS Annual Meeting, Phoenix, AZ, October 14–17, 2012.

22. Shahinpour, S., S. Butenko, and Z. Ertem, "A global optimization algorithm for the maximum biclique problem in large–scale networks", INFORMS Annual Meeting, Phoenix, AZ, October 14–17, 2012.

23. Sung, J.S., and S. Butenko, "Minimum selective dominating set problem", INFORMS Annual Meeting, Phoenix, AZ, October 14–17, 2012.

24. Yezerska, O., V. Boginski, and S. Butenko, "Detecting robust cliques in graphs with uncertain edge failures", INFORMS Annual Meeting, Phoenix, AZ, October 14–17, 2012.

25. Buchanan, A., A. Verma, and S. Butenko, "Maximum clique problem on very large scale sparse networks", 21st International Symposium on Mathematical Programming, Berlin, Germany, August 19–24, 2012.

26. Butenko, S., "Clique relaxation models in networks: theory, algorithms, and applications", Summer School on Operational Research and Applications, Laboratory of Algorithms and Technologies for Networks Analysis, Nizhny Novgorod, Russia, May 5–6, 2012.

27. Butenko, S., and A. Verma, "A community based approach to network clustering", 4th International Conference on Dynamics of Information Systems, Gainesville, FL, February 20–22, 2012.

28. Verma, A., and S. Butenko, "Network clustering via clique relaxations: a community based approach", 10th DIMACS Implementation Challenge – Graph Partitioning and Graph Clustering, Atlanta, GA, February 13–14, 2012.

29. Detecting Low-Diameter Clusters in Graphs. Colloquium / Special Lectures Series, Department of Computational and Applied Mathematics, Nov 28, 2011, Rice University, Houston, TX.

30. On the CVAR-constrained maximum 2-club problem. INFORMS 2011, Nov 13-16, 2011, Charlotte, NC (Presented by coauthor E. Moradi).

31. CVaR-constrained minimum spanning k-core problem for robust network design. INFORMS 2011, Nov 13-16, 2011, Charlotte, NC (Presented by coauthor J. Ma).

32. New polyhedral results for the 2-club polytope of graphs. INFORMS 2011, Nov 13-16, 2011, Charlotte, NC (Presented by coauthor F. Mahdavi Pajouh).

33. On characterization of maximal independent sets via quadratic optimization. INFORMS 2011, Nov 13-16, 2011, Charlotte, NC (Presented by coauthor F. Mahdavi Pajouh).

34. Conditional value-at-risk based approaches to robust network flow, connectivity and design problems. 2011 DOE Applied Mathematics Program Meeting, October 17–19, 2011, Washington, D.C. (Poster presentation).

35. Survivable network design by spanning k-cores. Systems and Optimization Aspects of Smart Grid Challenges, April 28-30, 2011, Gainesville, FL.

36. Butenko, S., "Clique relaxation models in network analysis", First International Conference on Network Analysis, Gainesville, FL, December 14–16, 2011

37. Pattillo, J., S. Butenko, and Y. Wang, "2-Cliques on unit disk graphs", INFORMS Annual Meeting, Charlotte, NC, November 13–18, 2011.

38. Shahinpour, S., and S. Butenko, "Heuristic algorithms for solving the maximum k–club problem in graphs", INFORMS Annual Meeting, Charlotte, NC, November 13–18, 2011.

39. Verma, A., and S. Butenko, "A 6-approximate algorithm for the k–bottleneck connected dominating set problem", INFORMS Annual Meeting, Charlotte, NC, November 13–18, 2011.

40. Yezerska, O., V. Boginski, and S. Butenko, "Maximum k-plex problem in networks under uncertainty", INFORMS Annual Meeting, Charlotte, NC, November 13–18, 2011.

41. Youssef, N., J. Pattillo, and S. Butenko. "Taxonomy of clique relaxations", INFORMS Annual Meeting, Charlotte, NC, November 13–18, 2011.

42. Butenko, S., "Clique Relaxation Models in Networks: Theory, Algorithms, and Applications", Summer School on Achievements and Applications of Contemporary Informatics, Mathematics, and Physics, Kyiv, Ukraine, August 11, 2011.

43. Boginski, V., S. Butenko, J. Pattillo, and A. Veremyev, "Finding quasi–cliques in networks", INFORMS North–Eastern Conference, Amherst, MA, May 6–7, 2011.

44. Verma, A., and S. Butenko, "Maximum clique problem on very large scale sparse networks", INFORMS North–Eastern Conference, Amherst, MA, May 6–7, 2011.

45. Youssef, N., and S. Butenko, "Clique relaxation models in social network analysis", INFORMS North–Eastern Conference, Amherst, MA, May 6–7, 2011.

46. Youssef, N., S. Butenko, and V. Boginski, "Risk–based network design facing node disruptions", 3rd International Conference on Dynamics of Information Systems, Gainesville, FL, February 16–18, 2011.

47. New developments on the maximum k-club problem in graphs. INFORMS Annual Meeting 2010, Nov 7–10, 2010, Austin, TX (Presented by coauthor F. Mahdavi Pajouh).

48. Using conditional value-at-risk in the shortest path problem to find robust paths under arc failures. INFORMS Annual Meeting 2010, Nov 7–10, 2010, Austin, TX (Presented by coauthor J. Bright).

49. Variable objective search. INFORMS Annual Meeting 2010, Nov 7–10, 2010, Austin, TX (Presented by coauthor O. Yezerska).

50. Conditional value-at-risk based approaches to robust network flow, connectivity and design problems. 2010 DOE Applied Mathematics Program Meeting, May 3–5, 2010, Berkeley, CA (Poster presentation).

51. Minimum spanning k-core problem for robust network design. 2010 INFORMS Optimization Society Meeting, Feb 26–28, 2010, University of Florida, Gainesville, FL.

52. Degree-based relaxations of cliques, independent sets and coloring. International Conference on the Dynamics of Information Systems (DIS) 2010, February 3–5, 2010, Gainesville, FL.

53. Butenko, S., "Bottleneck connected dominating set problem in ad hoc wireless networks", Computational & Applied Mathematics Colloquium/Special Lectures, Rice University, Houston, TX, September 20, 2010.

54. J. Pattillo, V. Boginski, S. Butenko, and A. Veremyev, "The maximum p/q-clique problem", INFORMS Annual Meeting, Austin, TX, November 7–10, 2010.

55. S. Shahinpour and S. Butenko, "An efficient heuristic algorithm for finding maximum k-club in an undirected graph", INFORMS Annual Meeting, Austin, TX, November 7–10, 2010.

56. A. Verma and S. Butenko, "An approximation algorithm for the k–BCDS problem in wireless sensor networks", INFORMS Annual Meeting, Austin, TX, November 7–10, 2010.

57. Yezerska, O., S. Butenko, and B. Balasundaram, "Variable objective search", 3rd Yalta Optimization Conference: "Network Science", August 2–4, 2010.

58. Verma, A., and S. Butenko, "A distributed approximation algorithm for the bottleneck connected dominating set problem" , 3rd Yalta Optimization Conference: "Network Science", August 2–4, 2010.

59. Verma, A., S. Butenko, and S. Kahruman–Anderoglu, "Approximate algorithms for bottleneck connected domination in wireless sensor networks", 2nd International Conference on Dynamics of Information Systems, Destin, FL, February 3–5, 2010.

60. A computational study of the maximum k-club problem. INFORMS Annual Meeting 2009, Oct 11–14, 2009, San Diego, CA (Presented by coauthor F. Mahdavi Pajouh).

61. Butenko, S., J. Chen, and S. Trukhanov, "On equivalence of some scale–reduction techniques for the maximum weight independent set problem", INFORMS Annual Meeting, San Diego, CA, October 11–14, 2009.

References

[1] Y. Asahiro, E. Miyano, and K. Samizo. Approximating maximum diameter-bounded subgraphs. In Alejandro Lopez-Ortiz, editor, *LATIN 2010: Theoretical Informatics*, volume 6034 of *Lecture Notes in Computer Science*, pages 615–626. Springer Berlin / Heidelberg, 2010.

[2] B. Balasundaram, S. Butenko, and S. Trukhanov. Novel approaches for analyzing biological networks. *Journal of Combinatorial Optimization*, 10(1):23–39, 2005.

[3] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.

[4] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, NY, 1997.

[5] J.-M. Bourjolly, G. Laporte, and G. Pesant. Heuristics for finding k-clubs in an undirected graph. *Computers & Operations Research*, 27:559–569, 2000.

[6] J.-M. Bourjolly, G. Laporte, and G. Pesant. An exact algorithm for the maximum k-club problem in an undirected graph. *European Journal Of Operational Research*, 138:21–28, 2002.

[7] R. Carraghan and P. Pardalos. An exact algorithm for the maximum clique problem. *Operations Research Letters*, 9:375–382, 1990.

[8] W. Cook, W. Cunningham, W. Pulleyblank, and A. Schrijver. *Combinatorial Optimization*. John Wiley and Sons, New York, 1998.

[9] J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming. *Mathematical programming*, 95(3):493–511, 2003.

[10] C. I. Fábián. Handling cvar objectives and constraints in two-stage stochastic models. *European Journal of Operational Research*, 191(3):888 – 911, 2008.

[11] W. Haneveld and M. van der Vlerk. Integrated chance constraints: Reduced forms and an algorithm. *Computational Management Science*, 3:245–269, 2006. 10.1007/s10287-005-0007-3.

[12] L. J. Hong and G. Liu. Simulating sensitivities of conditional value at risk. *Management Science*, 55:281–293, February 2009.

[13] P. Huang and D. Subramanian. Iterative estimation maximization for stochastic linear programs with conditional value-at-risk constraints. *Computational Management Science*, 9(4):441–458, 2012.

[14] P. Huang, D. Subramanian, and J. Xu. An importance sampling method for portfolio cvar estimation with gaussian copula models. In *Proceedings of the Winter Simulation Conference*, WSC '10, pages 2790–2800. Winter Simulation Conference, 2010.

[15] P. Krokhmal, J. Palmquist, and S. Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4:43–68, 2002.

[16] A. Künzi-Bay and J. Mayer. Computational aspects of minimizing conditional value-at-risk. *Computational Management Science*, 3:3–27, 2006.

[17] R. D. Luce. Connectivity and generalized cliques in sociometric group structure. *Psychometrika*, 15(2):169–190, 1950.

[18] J. Luedtke. An integer programming and decomposition approach to general chance-constrained mathematical programs. In F. Eisenbrand and F. Shepherd, editors, *Integer Programming and Combinatorial Optimization*, volume 6080 of *Lecture Notes in Computer Science*, pages 271–284. Springer Berlin / Heidelberg, 2010.

[19] J. Ma and B. Balasundaram. Solving chance-constrained spanning k-core problem via decomposition and integer programming. Proceedings of the 2013 Industrial and Systems Engineering Research Conference (ISERC 2013), 2013.

[20] J. Ma, F. Mahdavi Pajouh, B. Balasundaram, and V. Boginski. The minimum spanning k-core problem with bounded CVaR under probabilistic edge failures, 2014. In preparation.

[21] C. L. Monma and D. F. Shallcross. Methods for designing communications networks with certain two-connected survivability constraints. *Operations Research*, 37(4):531–541, 1989.

[22] Department of Energy Workshop. Mathematical research challenges in optimization of complex systems. Available at: `http://www.courant.nyu.edu/ComplexSystems/literature/ComplexSystemsReport.pdf`, 2006.

[23] P. R. J. Östergård. A fast algorithm for the maximum clique problem. *Discrete Applied Mathematics*, 120:197–207, 2002.

[24] F. Mahdavi Pajouh and B. Balasundaram. On inclusionwise maximal and maximum cardinality k-clubs in graphs. *Discrete Optimization*, 9(2):84–97, 2012.

[25] F. Mahdavi Pajouh, B. Balasundaram, and I. V. Hicks. On the 2-club polytope of graphs. In preparation.

[26] J. Pattillo, N. Youssef, and S. Butenko. On clique relaxation models in network analysis. *European Journal of Operational Research*, 226(1): 9–18, 2013.

[27] R. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.

[28] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41, 2000.

[29] S. B. Seidman. Network structure and minimum degree. *Social Networks*, 5:269–287, 1983.

[30] S. Sethuraman and S. Butenko. The maximum ratio clique problem. *Computational Management Science*. To appear, DOI: 10.1007/s10287–013–0197–z.

[31] S. Shen, J. C. Smith, and S. Ahmed. Expectation and chance-constrained models and algorithms for insuring critical paths. *Management Science*, 56(10):1794–1814, 2010.

[32] A. Sorokin, V. Boginski, A. Nahapetyan, and P. M. Pardalos. Computational risk management techniques for fixed charge network flow problems with uncertain arc failures. *Journal of Combinatorial Optimization*, 25(1):99–122, 2013.

[33] K. Steiglitz, P. Weiner, and D. Kleitman. The design of minimum-cost survivable networks. *IEEE Transactions on Circuit Theory*, 16(4):455–460, 1969.

[34] D. Subramanian and P. Huang. An efficient decomposition algorithm for static, stochastic, linear and mixed-integer linear programs with conditional-value-at-risk constraints. Technical Report RC24752, IBM Research Report, Feb 2009.

[35] S. Trukhanov, C. Balasubramaniam, B. Balasundaram and S. Butenko. Algorithms for detecting optimal hereditary structures in graphs, with application to clique relaxations. *Computational Optimization and Applications*, 56(1):113–130, 2013.

[36] R. Van Slyke and R.J-B. Wets. L-shaped linear programs with applications to control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17:638–663, 1969.

[37] A. Verma, A. Buchanan, and S. Butenko. Solving the maximum clique and vertex coloring problems on very large sparse networks. Submitted to *INFORMS Journal on Computing*.

[38] A. Verma and S. Butenko. Network clustering via clique relaxations: a community–based approach. In: *Graph Partitioning and Graph Clustering*. Ed. by D. A. Bader, H. Meyerhenke, P. Sanders, and D. Wagner. American Mathematical Society, pp.125–136, 2013.

[39] W. Wang. *Sample Average Approximation of Risk-Averse Stochastic Programs*. PhD thesis, Georgia Institute of Technology, 2007.

[40] W. Wang and S. Ahmed. Sample average approximation of expected value constrained stochastic programs. *Operations Research Letters*, 36(5):515 – 519, 2008.

[41] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, New York, 1994.

[42] O. Yezerska, S. Butenko, and V. Boginski, "Detecting robust cliques in graphs with uncertain edge failures", In preparation.