AFRL-AFOSR-VA-TR-2016-0001

Clique Relaxations in Biological and Social Network Analysis Foundations and Algorithms

Sergiy Butenko
TEXAS ENGINEERING EXPERIMENT STATION COLLEGE STATION

10/26/2015
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*
06-10-2015

2. REPORT TYPE
Final Report

3. DATES COVERED *(From - To)*
July 1, 2012 - June 30, 2015

4. TITLE AND SUBTITLE
Clique Relaxations in Biological and Social Network Analysis: Foundations and Algorithms

5a. CONTRACT NUMBER
FA9550-12-1- 0103

5b. GRANT NUMBER
FA9550-12-1- 0103

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Butenko, Sergiy
Balasundaram, Balabhaskar
Boginski, Vladimir

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dept. of Industrial and Systems Engineering, Texas A&M University, 4037 Emerging Technologies Build., College Station, TX 77843-3131, Phone: 979-458-2333, Fax: 979-458-4299, Email: butenko@tamu.edu
Sub-contracts: Oklahoma State University, University of Florida

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Optimization and Discrete Mathematics
Mathematics, Information and Life Sciences
Air Force Office of Scientific Research

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
For public-A

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The objective of this project is to provide a unifying theoretical and computational framework for the study of clique relaxation models arising in biological and social networks. This project examines the elementary clique-defining properties inherently exploited in the available clique relaxation models and proposes 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. Based on the proposed taxonomy, a comprehensive study of the resulting optimization problems is carried out, aiming to study the cohesiveness properties of various clique relaxation aiming to assist researchers in selecting the most appropriate model for a particular application of interest; explore the fundamental properties of the clique relaxation models of interest that are responsible for the computational complexity of the corresponding optimization problems, and exploit these properties in designing appropriate computational tools for solving the problems in question; identify robust clique relaxation structures of practical interest.

15. SUBJECT TERMS
Optimization, Complex Networks, Network Analysis

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU	12	Sergiy Butenko
					19b. TELEPHONE NUMBER *(Include area code)* (979) 458-2333

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18
Adobe Professional 7.0

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Clique Relaxations in Biological and Social Networks: Foundations and Algorithms

Contract/Grant #: FA9550-12-1-0103

Final Project Report

07/01/2012 – 06/30/2015

Sergiy Butenko, Ph.D. (PI)
Texas A&M University
Department of Industrial and Systems Engineering
4037 Emerging Technologies Building, College Station, TX 77843-3131
Tel: (979) 458-2333
Fax: (979) 458-4299
Email: butenko@tamu.edu

Balabhaskar Balasundaram, Ph.D. (Co-PI)
Oklahoma State University
School of Industrial Engineering & Management
322 Engineering North, Stillwater, OK 74078
Tel: (405) 744-6055
Fax: (405) 744-4654
Email: baski@okstate.edu

Vladimir Boginski, Ph.D. (Co-PI)
University of Florida
Research & Engineering Education Facility
1350 N Poquito Road, Shalimar, FL 32579
Tel.: (850) 833-9355 ext 240
Fax: (850) 833-9366
Email: boginski@reef.ufl.edu

Abstract

The objective of this project is to provide a unifying theoretical and computational framework for the study of clique relaxation models arising in biological and social networks. Increasing interest in studying *community structures*, *functional modules*, or *clusters* in complex networks arising in biological and social networks, as well as in various other applications, has lead to a large and diverse body of literature introducing numerous graph-theoretic models, which relax certain characteristics of the classical clique concept. This project examines the elementary clique-defining properties inherently exploited in the available *clique relaxation* models and proposes 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. Based on the proposed taxonomy, a comprehensive study of the resulting optimization problems is carried out, aiming to address the following issues.

- Systematically studying the cohesiveness properties of various clique relaxation aiming to assist researchers in selecting the most appropriate model for a particular application of interest.
- Exploring the fundamental properties of the clique relaxation models of interest that are responsible for the computational complexity of the corresponding optimization problems, and exploiting these properties in designing appropriate computational tools for solving the problems in question.
- Identifying robust clique relaxation structures of minimum cost that are of practical interest.

The applications of direct interest to the Air Force include the design, management and defense of complex networks, information technology, and other applications as described in *complex networked systems*, *robust decision making*, and *socio-cultural modeling of effective influence* discovery challenge thrusts.

Summary of Research Contributions

The main research contributions resulting from the project are summarized below, including the list of papers supported by this project that have been finalized and are published/accepted for publication or submitted/ready for submission.

The unifying framework for investigation of clique relaxation models in this project was established in [1]. This paper analyzes the elementary clique-defining properties implicitly exploited in the available *clique relaxation* models and proposes 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. According to the proposed taxonomy, clique relaxation structures were classified into regular/weak, absolute/relative models that either restrict a violation of an elementary clique defining property or enforce a fixed-size clique property in the corresponding definition. Higher-order relaxations, which consider more than one elementary clique defining property simultaneously, have also been introduced, and rigorous bounds on the cohesiveness properties guaranteed by the so-called *canonical* clique relaxations have been established. With respect to dynamic structural properties, the analysis focused on properties related to the concept of *heredity.* It was observed that the presence of the heredity in induced subgraphs property is essential for developing effective combinatorial branch-and-bound strategies for detection of maximum clique relaxation structures. The concepts of weak heredity, quasi-heredity, and k-heredity have been introduced and their algorithmic implications have been discussed. These developments defined the further sequence of steps taken towards achieving the project's objectives, as described next.

Hereditary clique relaxations

Papers [2–8] study *hereditary clique relaxations*, which, in addition to clique itself, include s-plex, s-defective clique, s-bundle, and s-stable cluster models arising in social and biological networks. The presence of heredity allows to develop effective scale-reduction procedures and exact algorithms for these models [2]. As a result, it is demonstrated that real-life problem instances with millions of vertices and edges can be solved to optimality [3,4]. Theoretical explanation of such a success in solving very large instances of the maximum clique problem, which is notorious for its intractability, is provided in [5]. The maximum ratio clique problem is introduced and studied for the first time in [6]. The fractional objective is motivated by practical consideration, as one is often interested in maximizing benefit/cost ratio when searching for a cluster. Easily computable analytical lower bounds on the clique number are studied in [7], and connections of the maximum clique problem to the potential energy principles in networked systems are discussed in [8].

Non-hereditary clique relaxations

New approaches for the edge density based quasi-clique model, which does not posses heredity but is quasi-hereditary, are developed in [9, 10]. One of the most interesting and practically important non-hereditary clique relaxations is the distance-based s-club model, which is defined as a subset of nodes inducing a subgraph of diameter at most s. Due to the lack of heredity, computing even maximal by inclusion s-clubs is an NP-hard problem. A survey of recent progress in solving the maximum s-club problem and its weaker counterpart, the maximum s-clique problem, is given in [11]. An exact approach for the maximum s-club problem that takes advantage of the close relation to s-cliques is proposed in [12]. A special case of 2-cliques is considered in [13], where a 0.5-approximation algorithm for this problem in unit-disk graphs is proposed. Dominating clubs, modeling low-diameter virtual backbones in networks, are studied in [14]. Complexity of gap recognition results for clubs inspired a new general theoretical framework for characterizing "provably best" construction heuristics for hard combinatorial optimization problems [15]. A new clique relaxation structure, α-cluster, based on the concept of a clustering coefficient is introduced and studied in the context of social networks in [16].

Robustness considerations

Several approaches are used to address robustness considerations in this project. The main focus is on ensuring connectivity, which is perhaps the most essential requirement that can be imposed on a robust network cluster. The connected subgraph polytope of a graph is studied in detail in [17]. Approaches for finding maximum subgraphs with relatively large vertex connectivity are proposed in [18]. An effective integer programming approach for computing fault-tolerant connected dominating sets is developed in [19]. This approach performs much better than previously published methods. Minimization version of k-core and related problems is suggested as a way of obtaining robust clusters and tight upper bounds on the clique number in [20] and [21], respectively. Finally, stochastic risk-averse versions of the spanning k-core, maximum clique, and maximum 2-club problems are investigated in [22], [23], and [24], respectively.

References

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

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

[3] A. Verma, A. Buchanan, and S. Butenko. Solving the maximum clique and vertex coloring problems on very large sparse networks. *INFORMS Journal on Computing* 27 (2015), 164–177.

[4] Z. Miao and B. Balasundaram. Approaches for finding cohesive subgroups in large-scale social networks via maximum k-plex detection. Under review in *Networks*, August 2014.

[5] A. Buchanan, J. L. Walteros, S. Butenko, and P. M. Pardalos. Solving maximum clique in sparse graphs: an $O(nm + 2^{d/4})$ algorithm for d-degenerate graphs. *Optimization Letters* 8 (2014), 1611–1617.

[6] S. Sethuraman and S. Butenko. The maximum ratio clique problem. *Computational Management Science* 12 (2015), 197–218.

[7] V. Stozhkov, G. Pastukhov, V. Boginski, and E. L. Pasiliao. New analytical lower bounds on the clique number of a graph. Submitted October 2015.

[8] A. Veremyev, V. Boginski, E.L. Pasiliao. Potential energy principles in networked systems and their connections to optimization problems on graphs. *Optimization Letters* 9 (2015) 585–600.

[9] Z. Miao and B. Balasundaram. Lagrangian dual bounds for the maximum quasi-clique problem. In preparation. Anticipated submission, December 2015.

[10] F. Mahdavi Pajouh, Z. Miao, and B. Balasundaram. A branch-and-bound approach for maximum quasi-cliques. *Annals of Operations Research* 216 (2014), 145–161.

[11] 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, 2013, pp.149–174.

[12] E. Moradi and B. Balasundaram. Finding a maximum k-club using the k-clique formulation and canonical hypercube cuts. Conditionally accepted at *Optimization Letters*, October 2015.

[13] J. Pattillo, Y. Wang, and S. Butenko. Approximating 2-cliques in unit disk graphs. *Discrete Applied Mathematics* 166 (2014), 178–187.

[14] A. Buchanan, J. Sung, V. Boginski, and S. Butenko. On connected dominating sets of restricted diameter. *European Journal of Operational Research* 236 (2014), 410–418.

[15] S. Kahruman-Anderoglu, A. Buchanan, S. Butenko, and O. Prokopyev. On provably best construction heuristics for hard combinatorial optimization problems. *Networks* (2015). To appear, DOI 10.1002/net.21620.

[16] Z. Ertem, A. Veremyev, and S. Butenko. Detecting large cohesive subgroups with high clustering coefficients in social networks. Under review at *Social Networks*, submitted December 2014.

[17] Y. Wang, A. Buchanan, and S. Butenko. On imposing connectivity constraints in integer programs. Submitted, June 2015. Available online at http://www.optimization-online.org/DB_HTML/2015/02/4768.html

[18] A. Veremyev, O.A. Prokopyev, V. Boginski, and E.L. Pasiliao. Finding maximum subgraphs with relatively large vertex connectivity. *European Journal of Operational Research* 239 (2014) 349–362.

[19] A. Buchanan, J. Sung, S. Butenko, and E. L. Pasiliao. An integer programming approach for fault-tolerant connected dominating sets. *INFORMS Journal on Computing* 27 (2015), 178–188.

[20] C. Balasubramaniam and S. Butenko. On robust clusters of minimum cardinality in networks. *Annals of Operations Research* (2015). To appear, DOI 10.1007/s10479-015-1992-4.

[21] C. Balasubramaniam, S. Butenko, and B. Balasundaram. On upper bounds for the maximum clique problem. Submitted, October 2015.

[22] J. Ma and B. Balasundaram. On the chance-constrained minimum spanning k-core problem. In preparation. Anticipated submission, December 2015.

[23] O. Yezerska, S. Butenko, and V. Boginski. Detecting robust cliques in graphs subject to uncertain edge failures. Under review at *Annals of Operations Research*, submitted June 2015.

[24] F. Mahdavi Pajouh, E. Moradi, and B. Balasundaram. Detecting large risk-averse 2-clubs in graphs with random edge failures. Under review at *Annals of Operations Research*, submitted December 2014.

Appendix: Abstracts of Papers

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

Abstract

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 paper analyzes the elementary clique-defining properties implicitly exploited in the available *clique relaxation* models and proposes 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. Some basic structural properties of several of the considered models are identified that may facilitate the choice of methods for solving the corresponding optimization problems. In addition, bounds describing the cohesiveness properties of different clique relaxation structures are established, and practical implications of choosing one model over another are discussed.

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

Abstract

The general node deletion problem can be stated as follows: Given a graph property Π, find the minimum number of nodes whose deletion results in a graph satisfying property Π. In 1978, Yannakakis established that if the property Π is *hereditary* on induced subgraphs, *nontrivial*, and *interesting*, the resulting node deletion problem is NP-hard. Several well-known graph properties such as requiring the graph to be complete, planar, perfect, and bipartite, among others, meet the conditions of this result, highlighting its scope. This paper proposes a general purpose exact algorithmic framework to solve such NP-hard node deletion problems, and investigates key algorithm design and implementation issues that will help tailor the general framework for specific graph properties. The performance of the algorithms so derived for *the maximum s-plex problem*, and *the maximum s-defective clique problem*, is assessed through a computational study. These problems are used in graph-based data mining applications, specifically in social and biological network analysis.

3. A. Verma, A. Buchanan, and S. Butenko. Solving the maximum clique and vertex coloring problems on very large sparse networks. *INFORMS Journal on Computing* 27 (2015), 164–177.

Abstract

This paper explores techniques for solving the maximum clique and vertex coloring problems on very large-scale real life networks. 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 approaches aim to reduce the network instances to a size that is tractable for existing solvers, while preserving optimality. Two clique relaxation structures are exploited for this purpose. In addition to the known k-core structure, a newly introduced clique relaxation, *k-community*, is used to further reduce the instance size. 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.

4. Z. Miao and B. Balasundaram. Approaches for finding cohesive subgroups in large-scale social networks via maximum k-plex detection. Under review in *Networks*, August 2014.

Abstract

A k-plex is a clique relaxation introduced in social network analysis to model cohesive social subgroups. Several exact algorithms and heuristic approaches to find a maximum size k-plex in the graph have been developed recently for this NP-hard problem. This article develops a basic greedy randomized adaptive search procedure (GRASP) for the maximum k-plex problem, with a key improvement in the design of the construction procedure that alleviates a drawback observed in multiple past studies. In existing construction heuristics, k-plexes found for smaller values of parameter k are sometimes not found for larger k even though they are feasible, and inferior solutions are found instead. We identify the reasons behind this behavior and address those in our new construction procedure. We then show that an existing exact algorithm for solving this problem on power-law graphs can be considerably enhanced by using GRASP and "peeling," a problem-specific preprocessing procedure. The overall approach is able to solve the problem to optimality on very large-scale social networks with over 100,000 vertices to those with over 4.8 million vertices. These are several orders of magnitude larger than the largest real-life social networks on which this problem has been solved to optimality in current literature.

5. A. Buchanan, J. L. Walteros, S. Butenko, and P. M. Pardalos. Solving maximum clique in sparse graphs: an $O(nm + 2^{d/4})$ algorithm for d-degenerate graphs. *Optimization Letters* 8 (2014), 1611–1617.

Abstract

We describe an algorithm for the maximum clique problem that is parameterized by the graph's degeneracy d. The algorithm runs in $O(nm + nT_d)$ time, where T_d is the time to solve the maximum clique problem in an arbitrary graph on d vertices. The best bound as of now is $T_d = O(2^{d/4})$ by Robson. This shows that the maximum clique problem is solvable in $O(nm)$ time in graphs for which $d \leq 4 \log_2 m + O(1)$. The analysis of the algorithm's runtime is simple; the algorithm is easy to implement when given a subroutine for solving maximum clique in small graphs; it is easy to parallelize. In the case of Bianconi-Marsili power-law random graphs, it runs in $2^{O(\sqrt{n})}$ time with high probability. We extend the approach for a graph invariant based on common neighbors, generating a second algorithm that has a smaller exponent at the cost of a larger polynomial factor.

6. S. Sethuraman and S. Butenko. The maximum ratio clique problem. *Computational Management Science* 12 (2015), 197–218.

Abstract

This paper introduces a fractional version of the classical maximum weight clique problem, the *maximum ratio clique problem*, which is to find a maximal clique that has the largest ratio of benefit and cost weights associated with the clique's vertices. NP-completeness of the decision version of the problem is established, and three solution methods are proposed. The results of numerical experiments with standard graph instances, as well as with real-life instances arising in finance and energy systems, are reported.

7. V. Stozhkov, G. Pastukhov, V. Boginski, and E. L. Pasiliao. New analytical lower bounds on the clique number of a graph. Submitted October 2015.

Abstract

This paper proposes three new analytical lower bounds on the clique number of a graph and compares these bounds with those previously established in the literature. Two proposed bounds are derived from the well-known Motzkin-Straus quadratic programming formulation for the maximum clique problem. Theoretical results on the comparison of various bounds are established. Computational experiments were performed on various random graph models such as Erdos-Renyi model for uniform graphs and a generalized graph model for power-law graphs that simulate graphs with different densities and assortativity coefficients. Computational results suggest that the proposed new analytical bounds improve the existing ones on many graph instances.

8. A. Veremyev, V. Boginski, E.L. Pasiliao. Potential energy principles in networked systems and their connections to optimization problems on graphs. *Optimization Letters* 9 (2015) 585–600.

Abstract

We introduce the concepts of "gravitational" and "elastic" potential energy in the context of networks and investigate their implications on interpreting certain well-known graph-theoretic concepts and optimization problems. In the case of gravitational potential energy, we treat the nodes of a graph as "particles" with masses and consider the potential energy of gravitational interactions between them. We prove that the maximum clique in a graph represents the minimum gravitational potential energy structure. This result yields a natural physics-based interpretation of the well- known Motzkin-Strauss formulation for the maximum clique problem. In the case of elastic potential energy, we assume that graph nodes are located on an elastic surface, graph edges represent "springs" connecting the nodes on a surface, and consider potential energy corresponding to elastic deformations of springs. We show that under certain reasonable assumptions on surface deformation, the second-smallest and the largest eigenvalues of the graph Laplacian matrix are essentially the minimum and the maximum elastic potential energy of a graph, respectively. Moreover, the associated eigenvectors correspond to node heights in the minimum and maximum energy node configurations. Note that the second-smallest eigenvalue of the graph Laplacian is widely known as algebraic connectivity, and it turns out that this concept also has a simple and intuitive physics-based interpretation.

9. Z. Miao and B. Balasundaram. Lagrangian dual bounds for the maximum quasi-clique problem. In preparation. Anticipated submission, December 2015.

Abstract

A γ-quasi-clique in a simple undirected graph refers to a subset of vertices that induces a subgraph with edge density at least $\gamma \in [0, 1]$. When $\gamma = 1$, this definition corresponds to a classical clique. When $\gamma < 1$, it relaxes the requirement of all possible edges in the clique definition. Quasi-clique has been used to detect dense clusters in graph-based data mining, especially in large-scale, error-prone data sets in which clique model can be overly restrictive. The maximum γ-quasi-clique problem which detects a maximum cardinality γ-quasi-clique from a given graph, can be formulated as a mathematical program with a linear objective function and a single quadratic constraint in binary variables. This article investigates the Lagrangian dual of the formulation, and develops an upper bounding technique based on approximately solving the Lagrangian dual. The upper-bounds based on our approach is significantly tighter in comparison to the current state-of-the-art mixed-integer programming solvers employing two well-known formulations for this problem. We establish the strength of the relaxation mathematically and via numerical experiments on DIMACS benchmark instances.

10. F. Mahdavi Pajouh, Z. Miao, and B. Balasundaram. A branch-and-bound approach for maximum quasi-cliques. *Annals of Operations Research* 216 (2014), 145–161.

Abstract

Detecting quasi-cliques in graphs is a useful tool for detecting dense clusters in graph-based data mining. Particularly in large-scale data sets that are error-prone, cliques are overly restrictive and impractical. Quasi-clique detection has been accomplished using heuristic approaches in various applications of graph-based data mining in protein interaction networks, gene co-expression networks, and telecommunication networks. Quasi-cliques are not hereditary, in the sense that every subset of a quasi-clique need not be a quasi-clique. This lack of heredity introduces interesting challenges in the development of exact algorithms to detect maximum cardinality quasi-cliques. The only exact approaches for this problem are limited to two mixed integer programming formulations that were recently proposed in the literature. The main contribution of this article is a new combinatorial branch-and-bound algorithm for the maximum quasi-clique problem.

11. 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, 2013, pp.149–174.

Abstract

The concept of the clique, originally introduced as a model of a cohesive subgroup in the context of social network analysis, is a classical model of a cluster in networks. However, the ideal cohesiveness properties guaranteed by the clique definition put limitations on its applicability to situations where enforcing such properties is unnecessary or even prohibitive. Motivated by practical applications of diverse origins, numerous *clique relaxation models*, which are obtained by relaxing certain properties of a clique, have been introduced and studied by researchers representing different fields. *Distance-based* clique relaxations, which replace the requirement on pairwise distances to be equal to 1 in a clique with less restrictive distance bounds, are among the most important such models. This chapter surveys the up-to-date progress made in studying two common distance-based clique relaxation models called s-clique and s-club, as well as the corresponding optimization problems.

12. E. Moradi and B. Balasundaram. Finding a maximum k-club using the k-clique formulation and canonical hypercube cuts. Conditionally accepted at *Optimization Letters*, October 2015.

Abstract

Detecting low-diameter clusters is an important graph-based data mining technique used in social network analysis, bioinformatics and text-mining. Low pairwise distances

within a cluster can facilitate fast communication or good reachability between vertices in the cluster. Formally, a subgraph of diameter at most k is called a k-club. For low values of the parameter k, this model offers a graph-theoretic relaxation of the clique model that formalizes the notion of a low-diameter cluster. Using a combination of graph decomposition and model decomposition techniques, we demonstrate how the fundamental optimization problem of finding a maximum size k-club can be solved optimally on large-scale benchmark networks that are available in the public domain. Our approach circumvents the use of complicated formulations of the maximum k-club problem in favor of a simple relaxation based on necessary conditions, combined with canonical hypercube cuts introduced by Balas and Jeroslow.

13. J. Pattillo, Y. Wang, and S. Butenko. Approximating 2-cliques in unit disk graphs. *Discrete Applied Mathematics* 166 (2014), 178–187.

Abstract

This paper studies distance-based clique relaxations in unit disk graphs arising in wireless networking applications. Namely, a 2-clique is a subset of nodes with pairwise distance at most two in the graph, and a 2-club is a subset of nodes inducing a subgraph of diameter two. It is shown that in a unit disk graph any 2-clique is 4-dominated and any 2-club is 3-dominated. The former observation is used to develop a $\frac{1}{2}$-approximation algorithm for the maximum 2-clique problem in unit disk graphs. Moreover, this also implies polynomial solvability of the minimum dominating set problem in unit disk graphs of diameter two, whereas the same problem is shown to be hard in general diameter-two graphs. The paper also poses several related open questions of interest.

14. A. Buchanan, J. Sung, V. Boginski, and S. Butenko. On connected dominating sets of restricted diameter. *European Journal of Operational Research* 236 (2014), 410–418.

Abstract

A connected dominating set (CDS) is commonly used to model a virtual backbone of a wireless network. To bound the distance that information must travel through the network, we explicitly restrict the diameter of a CDS to be no more than s leading to the concept of a dominating s-club. We prove that for any fixed positive integer s it is NP-complete to determine if a graph has a dominating s-club, even when the graph has diameter $s+1$. As a special case it is NP-complete to determine if a graph of diameter two has a dominating clique. We then propose a compact integer programming formulation for the related minimization problem, enhance the approach with variable fixing rules and valid inequalities, and present computational results.

15. S. Kahruman-Anderoglu, A. Buchanan, S. Butenko, and O. Prokopyev. On provably best construction heuristics for hard combinatorial optimization problems. *Networks* (2015). To appear, DOI 10.1002/net.21620.

Abstract

In this paper, a heuristic is said to be *provably best* if, assuming $\mathcal{P} \neq \mathcal{NP}$, no other heuristic always finds a better solution (when one exists). This extends the usual notion of "best possible" approximation algorithms to include a larger class of heuristics. We illustrate the idea on several problems that are somewhat stylized versions of real-life network optimization problems, including the maximum clique, maximum k-club, minimum (connected) dominating set, and minimum vertex coloring problems. The corresponding provably best construction heuristics resemble those commonly used within popular metaheuristics. Along the way, we show that it is hard to recognize whether the clique number and the k-club number of a graph are equal, yet a polytime computable function is "sandwiched" between them. This is similar to the celebrated Lovász function wherein an efficiently computable function lies between two graph invariants that are $\mathcal{NP}$-hard to compute.

16. Z. Ertem, A. Veremyev, and S. Butenko. Detecting large cohesive subgroups with high clustering coefficients in social networks. Under review at *Social Networks*, submitted December 2014.

Abstract

Clique relaxations are used in classical models of cohesive subgroups in social network analysis. *Clustering coefficient* was introduced more recently as a structural feature characterizing small-world networks. Noting the similarities of the concepts of cohesive subgroups and small-world networks, this paper introduces a new clique relaxation, α-*cluster*, defined by enforcing a lower bound α on the clustering coefficient in the corresponding induced subgraph. Two variations of the clustering coefficient are considered, namely, the local and global clustering coefficient. Certain structural properties of α-clusters are analyzed and mathematical optimization models for determining α-clusters of the largest size in a network are developed and validated using several real-life social network instances. In addition, a network clustering algorithm based on local α-clusters is proposed and successfully tested.

17. Y. Wang, A. Buchanan, and S. Butenko. On imposing connectivity constraints in integer programs. Submitted, June 2015. Available online at http://www.optimization-online.org/DB_HTML/2015/02/4768.html

Abstract

In many network applications, one searches for a connected subset of vertices that exhibits other desirable properties. To this end, this paper studies the connected subgraph polytope of a graph, which is the convex hull of subsets of vertices that induce a connected subgraph. Much of our work is devoted to the study of two nontrivial classes of valid inequalities. The first are the a, b-separator inequalities, which have been successfully used to enforce connectivity in previous computational studies. The second are the indegree inequalities, which have previously been shown to induce all nontrivial facets for trees. We determine the precise conditions under which these inequalities induce facets and when each class fully describes the connected subgraph polytope. Both classes of inequalities can be separated in polynomial time and admit compact extended formulations. However, while the a, b-separator inequalities can be lifted in linear time, it is NP-hard to lift the indegree inequalities.

18. A. Veremyev, O.A. Prokopyev, V. Boginski, and E.L. Pasiliao. Finding maximum subgraphs with relatively large vertex connectivity. *European Journal of Operational Research* 239 (2014) 349–362.

Abstract

We consider a clique relaxation model based on the concept of relative vertex connectivity. It extends the classical definition of a k-vertex-connected subgraph by requiring that the minimum number of vertices whose removal results in a disconnected (or a trivial) graph is proportional to the size of this subgraph, rather than fixed at k. Consequently, we further generalize the proposed approach to require vertex-connectivity of a subgraph to be some function f of its size. We discuss connections of the proposed models with other clique relaxation ideas from the literature and demonstrate that our generalized framework, referred to

as f-vertex-connectivity, encompasses other known vertex-connectivity-based models, such as s-bundle and k-block. We study related computational complexity issues and show that finding maximum subgraphs with relatively large vertex connectivity is NP-hard. An interesting special case that extends the R-robust 2-club model recently introduced in the literature, is also considered. In terms of solution techniques, we first develop general linear mixed integer programming (MIP) formulations. Then we describe an effective exact algorithm that iteratively solves a series of simpler MIPs, along with some enhancements, in order to obtain an optimal solution for the original problem. Finally, we perform computational experiments on several classes of random and real-life networks to demonstrate performance of the developed solution approaches and illustrate some properties of the proposed clique relaxation models.

19. A. Buchanan, J. Sung, S. Butenko, and E. L. Pasiliao. An integer programming approach for fault-tolerant connected dominating sets. *INFORMS Journal on Computing* 27 (2015), 178–188.

Abstract

This paper considers the minimum k-connected d-dominating set problem, which is a fault-tolerant generalization of the minimum connected dominating set (MCDS) problem. Three integer programming formulations based on vertex-cuts are proposed (depending on whether $d < k$, $d = k$, or $d > k$) and their integer hulls are studied. The separation problem for the vertex-cut inequalities is a weighted vertex-connectivity problem and is polytime solvable, meaning that the LP relaxation can be solved in polynomial time despite having exponentially many constraints. A new class of valid inequalities – r-robust vertex-cut inequalities – is introduced and is shown to induce exponentially many facets. Finally, a lazy-constraint approach is shown to compare favorably with existing approaches for the MCDS problem (the case $k = d = 1$), and is in fact the fastest in literature for standard test instances. A key subroutine is an algorithm for finding an inclusion-wise minimal vertex-cut in linear time. Computational results for $k = d = 2, 3, 4$ are provided as well.

20. C. Balasubramaniam and S. Butenko. On robust clusters of minimum cardinality in networks. *Annals of Operations Research* (2015). To appear, DOI 10.1007/s10479-015-1992-4.

Abstract

This paper studies two clique relaxation models, k-blocks and k-robust 2-clubs, used to describe structurally cohesive clusters with good robustness and reachability properties. The minimization version of the two problems are shown to be hard to approximate for $k \geq 3$ and $k \geq 4$, respectively. Integer programming formulations are proposed and a polyheral study is presented. The results of sample numerical experiments on several graph instances are also reported.

21. C. Balasubramaniam, S. Butenko, and B. Balasundaram. On upper bounds for the maximum clique problem. Submitted, October 2015.

Abstract

In this paper, we develop strong and easily computable upper bounds for the maximum clique problem that take advantage of continuous relaxations of integer programming formulations for the minimum k-core problem and its diameter-restricted version. We show that the proposed upper bounds are better than those obtained from the standard relaxations of the clique polytope. We also report the results of computational experiments with several classes of graph instances and provide a comparison with some existing bounds from the literature.

22. J. Ma and B. Balasundaram. On the chance-constrained minimum spanning k-core problem. In preparation. Anticipated submission, December 2015.

Abstract

A graph in which the degree of every vertex is at least k is called a k-core. If the parameter k is sufficiently large relative to the number of vertices, a k-core is guaranteed to possess 2-hop reachability between all pairs of vertices. Furthermore, it is guaranteed to preserve pairwise 2-hop distances under single vertex deletion. Hence, the notion of a k-core can be used to produce 2-hop survivable network designs, specifically to design inter-hub networks. Formally, given an edge-weighted graph, the minimum spanning k-core problem seeks to find a spanning subgraph of the given graph with minimum total edge weight that is also a k-core. For any fixed k, this problem is equivalent to a generalized graph matching problem, and can be solved in polynomial time. This article focuses on a chance constrained version of the minimum spanning k-core problem under probabilistic edge failures. We first establish the complexity of this problem by proving that it is NP-hard. Subsequently, we conduct a polyhedral study to identify faceting inducing inequalities, and then develop a strengthened formulation. The quality of bounds produced by the strengthened formulation, and its impact on solving this problem is demonstrated through a computational study.

23. O. Yezerska, S. Butenko, and V. Boginski. Detecting robust cliques in graphs subject to uncertain edge failures. Under review at *Annals of Operations Research*, submitted June 2015.

Abstract

This paper develops and compares several heuristic approaches, as well as an exact combinatorial branch-and-bound algorithm, for detecting maximum robust cliques in graphs subjected to multiple uncertain edge failures. The desired robustness properties are enforced using Conditional Value-at-Risk (CVaR) risk measure. The proposed heuristics are adaptations of the well-known tabu search and GRASP methods, whereas the exact approach is an extension of Östergård's algorithm for the maximum clique problem. The results of computational experiments on DIMACS graph instances are reported.

24. F. Mahdavi Pajouh, E. Moradi, and B. Balasundaram. Detecting large risk-averse 2-clubs in graphs with random edge failures. Under review at *Annals of Operations Research*, submitted December 2014.

Abstract

Detecting large 2-clubs in biological, social and financial networks can help reveal important information about the structure of the underlying systems. In large-scale networks that are error-prone, the uncertainty associated with the existence of an edge between two vertices can be modeled by assigning a failure probability to that edge. Here, we study the problem of detecting large "risk-averse" 2-clubs in graphs subject to probabilistic edge failures. To achieve risk aversion, we first model the loss in 2-club property due to probabilistic edge failures as a function of the decision (chosen 2-club cluster) and randomness (graph structure). Then, we utilize the conditional value-at-risk (CVaR) of the loss for a given decision as a quantitative measure of risk for that decision, which is bounded in the model. More precisely, the problem is modeled as a CVaR-constrained single-stage stochastic program. The main contribution of this article is a new decomposition algorithm based on a Benders decomposition scheme, which outperforms an algorithm based on an existing decomposition idea, on a test-bed of randomly generated instances, and real-life biological and social networks.

1.

1. Report Type

Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.

butenko@tamu.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

979-458-2333

Organization / Institution name

Texas A&M University

Grant/Contract Title
The full title of the funded effort.

Clique Relaxations in Biological and Social Networks: Foundations and Algorithms

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0103

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Sergiy Butenko

Program Manager
The AFOSR Program Manager currently assigned to the award

Fariba Fahroo

Reporting Period Start Date

07/01/2012

Reporting Period End Date

06/30/2015

Abstract

The objective of this project is to provide a unifying theoretical and computational framework for the study of clique relaxation models arising in biological and social networks. This project examines the elementary clique-defining properties inherently exploited in the available clique relaxation models and proposes 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. Based on the proposed taxonomy, a comprehensive study of the resulting optimization problems is carried out, aiming to study the cohesiveness properties of various clique relaxation aiming to assist researchers in selecting the most appropriate model for a particular application of interest; explore the fundamental properties of the clique relaxation models of interest that are responsible for the computational complexity of the corresponding optimization problems, and exploit these properties in designing appropriate computational tools for solving the problems in question; identify robust clique relaxation structures of practical interest.

Distribution Statement
This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

AFD-070820-035.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

FA9550-12-1-0103.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

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

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

A. Verma, A. Buchanan, and S. Butenko. Solving the maximum clique and vertex coloring problems on very large sparse networks. INFORMS Journal on Computing 27 (2015), 164-177.

A. Buchanan, J. L. Walteros, S. Butenko, and P. M. Pardalos. Solving maximum clique in sparse graphs: an O(nm + 2^{d/4}) algorithm for d-degenerate graphs. Optimization Letters 8 (2014), 1611-1617.

S. Sethuraman and S. Butenko. The maximum ratio clique problem. Computational Management Science 12 (2015), 197-218.

A. Veremyev, V. Boginski, E.L. Pasiliao. Potential energy principles in networked systems and their connections to optimization problems on graphs. Optimization Letters 9 (2015) 585-600.

F. Mahdavi Pajouh, Z. Miao, and B. Balasundaram. A branch-and-bound approach for maximum quasi-cliques. Annals of Operations Research 216 (2014), 145-161.

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, 2013, pp.149-174.

J. Pattillo, Y. Wang, and S. Butenko. Approximating 2-cliques in unit disk graphs. Discrete Applied Mathematics 166 (2014), 178-187.

A. Buchanan, J. Sung, V. Boginski, and S. Butenko. On connected dominating sets of restricted diameter. European Journal of Operational Research 236 (2014), 410-418.

S. Kahruman-Anderoglu, A. Buchanan, S. Butenko, and O. Prokopyev. On provably best construction heuristics for hard combinatorial optimization problems. Networks (2015). To appear, DOI 10.1002/net.21620.

A. Veremyev, O.A. Prokopyev, V. Boginski, and E.L. Pasiliao. Finding maximum subgraphs with relatively large vertex connectivity. European Journal of Operational Research 239 (2014) 349-362.

A. Buchanan, J. Sung, S. Butenko, and E. L. Pasiliao. An integer programming approach for fault-tolerant

connected dominating sets. INFORMS Journal on Computing 27 (2015), 178-188.

C. Balasubramaniam and S. Butenko. On robust clusters of minimum cardinality in networks. Annals of Operations Research (2015). To appear, DOI 10.1007/s10479-015-1992-4.

Changes in research objectives (if any):

Change in AFOSR Program Manager, if any:

Extensions granted or milestones slipped, if any:

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Oct 08, 2015 12:41:25 Success: Email Sent to: butenko@tamu.edu