On fault-tolerant low-diameter clusters in graphs

Yajun Lu

Assistant Professor
Department of Management & Marketing, Jacksonville State University
ylu@jsu.edu

Joint work with: Hosseinali Salemi

Department of Industrial & Manufacturing Systems Engineering, Iowa State University

Baski Balasundaram and Austin Buchanan

School of Industrial Engineering & Management, Oklahoma State University

October 25, 2021

Outline

Outline

Why does fault-tolerant cluster matter?

Source: nizamtaher.wordpress.com

What is an r-robust s-club?

Graph $G = (V, E)$ and positive integer r, s:

What is an r-robust s-club?

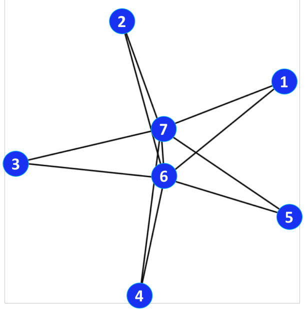
Graph $G = (V, E)$ and positive integer r, s:

- $S \subseteq V$ is called an r-robust s-club if there are at least r vertex-disjoint paths of length at most s in $G[S]$ between every distinct pair of vertices in S (Veremyev and Boginski, 2012).

What is an r-robust s-club?

Graph $G = (V, E)$ and positive integer r, s:

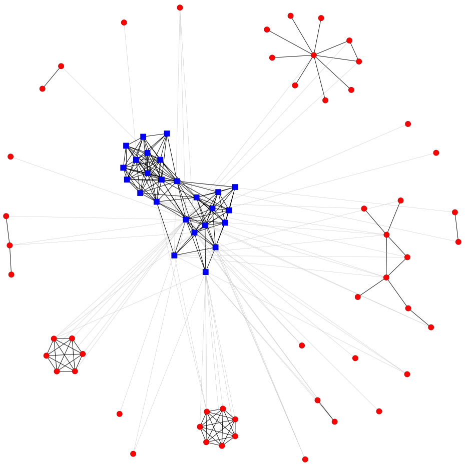
- $S \subseteq V$ is called an r-robust s-club if there are at least r vertex-disjoint paths of length at most s in $G[S]$ between every distinct pair of vertices in S (Veremyev and Boginski, 2012).



Maximum 2-robust 2-club in *H. pylori*

- Consider vertices $\{1, 2\}$
  - Path: $1 - 6 - 2$
  - Path: $1 - 7 - 2$

A maximum 3-robust 3-club in real-life network `lesmis`



Blue vertices visualized as square dots form a 3-robust 3-club; image was generated using the igraph package (Csardi and Nepusz, 2006)

Outline

Complexity

- **Problem:** Maximum r-robust s-club problem (MRCP)
- **Input:** Graph $G = (V, E)$ and positive integers $r \geq 2, s \geq 2$
- **Output:** An r-robust s-club of maximum cardinality

Complexity

- **Problem:** Maximum r-robust s-club problem (MRCP)
- **Input:** Graph $G = (V, E)$ and positive integers $r \geq 2, s \geq 2$
- **Output:** An r-robust s-club of maximum cardinality

Proposition 1 (Komusiewicz et al. (2019))

The decision version of the maximum r-robust 2-club problem is NP-complete for every fixed positive integer $r \geq 2$.

JSU JACKSONVILLE STATE UNIVERSITY

Complexity

- **Problem:** Maximum r-robust s-club problem (MRCP)
- **Input:** Graph $G=(V,E)$ and positive integers $r \geq 2, s \geq 2$
- **Output:** An r-robust s-club of maximum cardinality

Proposition 1 (Komusiewicz et al. (2019))

The decision version of the maximum r-robust 2-club problem is NP-complete for every fixed positive integer $r \geq 2$.

However, the complexity of the decision version of MRCP is not addressed by this result for every fixed positive integers $r \geq 2$ and $s \geq 3$.

NP-Hardness of Optimization

Theorem 1

The decision version of the MRCP is NP-complete for every pair of fixed integers $s \geq 2$ and $r \geq 2$, even on graphs with domination number one.

NP-Hardness of Optimization

Theorem 1

The decision version of the MRCP is NP-complete for every pair of fixed integers $s \geq 2$ and $r \geq 2$, even on graphs with domination number one.

Corollary 1

For every pair of fixed integer $r \geq 2$, the decision version of the MRCP remain NP-complete,

1. *on 4-chordal graphs for every fixed integer $s \geq 1$, and*
2. *on chordal graphs for every fixed even integer $s \geq 2$.*

NP-Hardness of Verification

Problem: IS r-ROBUST s-CLUB (positive integers s, r)
Question: Given a graph $G = (V, E)$ and a subset $S \subseteq V$, is S an r-robust s-club in G?

Theorem 2

IS r-ROBUST s-CLUB *is NP-complete for every fixed integer $s \geq 5$ and arbitrary positive integer r.*

NP-Hardness of Verification

Problem: IS r-ROBUST s-CLUB (positive integers s, r)
Question: Given a graph $G = (V, E)$ and a subset $S \subseteq V$, is S an r-robust s-club in G?

Theorem 2

IS r-ROBUST s-CLUB *is NP-complete for every fixed integer $s \geq 5$ and arbitrary positive integer r.*

Theorem 3

IS r-ROBUST s-CLUB *is NP-complete for every fixed integer $r \geq 2$ and arbitrary positive integer s.*

Integer Programing (IP) formulations of MRCP

- Veremyev and Boginski (2012) formulated the maximum r-robust 2-club problem.
- Almeida and Carvalho (2014) developed an IP formulation for the maximum r-robust 3-club problem, but no numerical experiments were reported in that work.
- No IP formulations exist for the MRCP when $r \geq 2, s \geq 4$.

Definition and Notation

Definition 1 (Salemi and Buchanan (2020); Lovász et al. (1978))

Given a pair of non-adjacent vertices u and v in graph $G = (V, E)$, a subset of vertices $C \subseteq V \setminus \{u, v\}$ is called a length-s u, v-separator if $d_{G-C}(u, v) > s$.

Notations:

- $\mathscr{C}_{uv}(G - uv)$ denotes the collection of all length-s u, v-separators in $G - uv$.
- $\mathbb{1}_E(u, v) = 1$ if $uv \in E$ and 0 otherwise.

JSU

Cut-like IP Formulation for the MRCP

Let $x_i = 1$ if and only if vertex i is included in the r-robust s-club.

$$\max \sum_{i \in V} x_i \tag{1a}$$

$$\text{s.t.} \quad (r - \mathbb{1}_E(u,v))(x_u + x_v - 1) \leq \sum_{i \in C} x_i \qquad \forall C \in \mathscr{C}_{uv}(G - uv), \forall uv \in \binom{V}{2} \tag{1b}$$

$$x_i \in \{0,1\} \qquad \forall i \in V. \tag{1c}$$

Cut-like IP Formulation for the MRCP

Let $x_i = 1$ if and only if vertex i is included in the r-robust s-club.

$$\max \sum_{i \in V} x_i \tag{1a}$$

$$\text{s.t.} \quad (r - \mathbb{1}_E(u,v))(x_u + x_v - 1) \leq \sum_{i \in C} x_i \qquad \forall C \in \mathscr{C}_{uv}(G - uv), \forall uv \in \binom{V}{2} \tag{1b}$$

$$x_i \in \{0,1\} \qquad \forall i \in V. \tag{1c}$$

Theorem 4

Given a graph $G = (V, E)$ and parameter $s \in \{2,3,4\}$, a subset of vertices S is an r-robust s-club if and only if its characteristic vector x^S satisfies the constraints of formulation (1).

JSU
JACKSONVILLE STATE UNIVERSITY

Formulation Strength

Proposition 2

The cut-like formulation (1) has a tighter LP relaxation than that of the Veremyev and Boginski (2012) formulation of the maximum r-robust 2-club problem when $r \geq 2$.

Formulation Strength

Proposition 2

The cut-like formulation (1) has a tighter LP relaxation than that of the Veremyev and Boginski (2012) formulation of the maximum r-robust 2-club problem when $r \geq 2$.

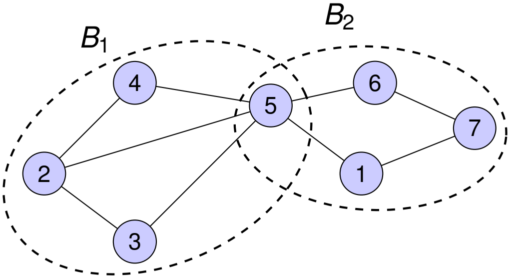
Proposition 3

The LP relaxations of the formulation of the maximum r-robust 3-club problem proposed by Almeida and Carvalho (2014) and the cut-like formulation (1) strengthened by inequalities $x_u + x_v \leq 1, \forall uv \in \binom{V}{2} \mid \rho_s(G; u, v) \leq r - 1$ are incomparable.

Outline

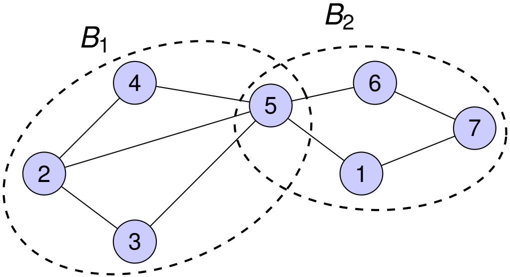
Blocks in graphs

A block is a subset of vertices that induces maximal biconnected subgraph.



Blocks in graphs

A block is a subset of vertices that induces maximal biconnected subgraph.



A block decomposition covers the graph using blocks. The example graph above decomposes into two blocks.

Block decomposition principle

Lemma 1

Let $G = (V, E)$ and $B_1, \ldots, B_t$ be its blocks. Consider an r-robust s-club S, then there exists a block B_i such that $S \subseteq V(B_i)$ for every $r \geq 2$.

Block decomposition principle

Lemma 1

Let $G = (V, E)$ and $B_1, \ldots, B_t$ be its blocks. Consider an r-robust s-club S, then there exists a block B_i such that $S \subseteq V(B_i)$ for every $r \geq 2$.

- A block decomposition of a graph $G = (V, E)$ can be found in $O(|V| + |E|)$ time (Hopcroft and Tarjan, 1973).

Recursive Block Decomposition Algorithm

Algorithm 1: Recursive Block Decomposition for the MRCP

Input: A graph $G = (V, E)$.
Output: A maximum cardinality r-robust s-club K.

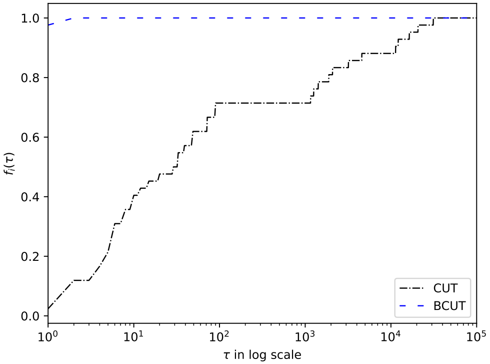
find the block decomposition $\mathscr{B}$ of G
$K \leftarrow$ a heuristic solution (Algorithm 2) of MRCP on the largest block in $\mathscr{B}$
while *a block* $D \in \arg\max\{|\hat{D}| : \hat{D} \in \mathscr{B}, |\hat{D}| > |K|\}$ *exists* **do**
  $\mathscr{B} \leftarrow \mathscr{B} \setminus \{D\}$
  preprocess block D by vertex peeling (Algorithm 3) using solution K
  find the block decomposition $\mathscr{F}$ of D
  if $|\mathscr{F}| = 1$ **then**
    $\hat{K} \leftarrow$ a maximum r-robust s-club in D
    if $|\hat{K}| > |K|$ **then**
      $K \leftarrow \hat{K}$
  else
    $\mathscr{B} \leftarrow \mathscr{B} \cup \mathscr{F}$
return K

Outline

Computational Experiments

- **Goal:**
  - Assessing the Cut-Like Formulations
  - Assessing the Recursive Block Decomposition
- **Test-bed:** Real-life networks from the 10th DIMACS Implementation Challenge on Graph Clustering (a collection of social and biological networks)
- **Software:** Gurobi™ Optimizer v9 and implemented in C++
- **Hardware:** 64-bit Linux® compute node with with dual intel® Skylake 6130 processors and 96 GB RAM

Assessing the Effectiveness of Recursive Block Decomposition When $s = 2$



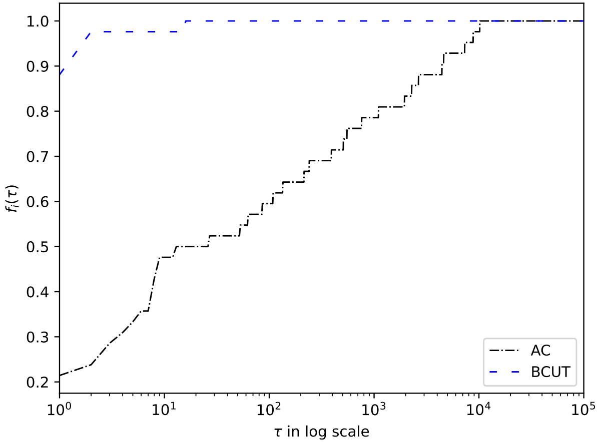
Performance profiles (Dolan and Moré, 2002; Gould and Scott, 2016) based on the wall-clock running times of solvers CUT and BCUT for the maximum r-robust 2-club problem when $r \in \{2,3,4\}$.

- CUT: Cut-Like IP Formulation + Branch-and-Cut Algorithm
- BCUT: Cut-Like IP Formulation + Branch-and-Cut Algorithm + Recursive Block Decomposition Algorithm

Wall-clock running times in seconds by CUT and BCUT in solving the maximum r-robust 2-club problem

			$r = 2$		$r = 3$		$r = 4$	
Graph	n	m	CUT	BCUT	CUT	BCUT	CUT	BCUT
karate	34	78	0.03	0.03	0.03	0.00	0.02	0.02
dolphins	62	159	0.06	0.03	0.06	0.00	0.05	0.04
lesmis	77	254	0.08	0.00	0.08	0.00	0.09	0.00
polbooks	105	441	0.14	0.03	0.13	0.03	0.20	0.03
adjnoun	112	425	0.15	0.03	0.18	0.02	0.39	0.03
football	115	613	0.10	0.11	0.07	0.00	0.07	0.00
jazz	198	2742	0.19	0.06	0.20	0.05	0.20	0.04
celegans	453	2025	1.41	0.02	1.38	0.02	1.11	0.03
email	1133	5451	109.48	7.38	38.12	0.53	13.40	0.28
polblogs	1490	16715	22.39	5.25	56.15	7.69	61.49	6.61
netscience	1589	2742	22.64	0.00	19.97	0.01	15.23	0.01
power	4941	6594	625.26	0.50	53.24	0.02	41.31	0.00
hep-th	8361	15751	1299.56	0.69	1284.72	0.28	897.76	0.07
PGP	10680	24316	1479.27	0.71	LPNS	0.22	3074.40	0.10

Assessing the Cut-Like Formulation For the MRCP When $s = 3$



Performance profile based on the wall-clock running times of solvers AC and BCUT for the maximum r-robust 3-club problem when $r \in \{2,3,4\}$.

- AC: AC formulation by Almeida and Carvalho (2014) + Recursive Block Decomposition
- BCUT: Cut-Like IP Formulation + BC Algorithm + Recursive Block

Wall-clock running times in seconds for solving the maximum r-robust 3-club problem

Graph	n	m	Best objective $r=2$	Best objective $r=3$	Best objective $r=4$	Wall-clock running time $r=2$ AC	$r=2$ BCUT	$r=3$ AC	$r=3$ BCUT	$r=4$ AC	$r=4$ BCUT
karate	34	78	21	11	9	0.03	0.01	0.01	0.01	0.07	0.01
dolphins	62	159	22	14	7	0.34	0.04	0.35	0.06	0.17	0.13
lesmis	77	254	35	25	21	0.00	0.00	0.00	0.00	0.00	0.00
polbooks	105	441	39	31	24	0.80	0.03	4.29	0.05	1.56	0.03
adjnoun	112	425	63	47	31	9.63	0.04	15.26	0.03	54.63	0.14
football	115	613	40	27	17	94.02	0.70	55.85	0.89	9.94	1.34
jazz	198	2742	158	145	136	65.95	0.06	438.29	0.06	276.30	0.06
celegans	453	2025	234	141	99	121.50	0.16	1627.88	0.16	1142.98	0.13
email	1133	5451	138	88	66	**45.76%**	403.02	**125.42%**	**9.09%**	**243.33%**	1130.71
polblogs	1490	16715	672	605	557	MEM	1.36	MEM	1.58	MEM	1.85
netscience	1589	2742	24	21	20	0.02	0.32	0.10	0.02	0.08	0.01
power	4941	6594	17	12	12	0.82	0.32	0.26	0.02	0.00	0.00
hep-th	8361	15751	52	38	32	**8%**	16.80	76.80	0.71	0.18	0.18
PGP	10680	24316	239	170	124	1.08	1.12	231.37	0.42	1815.75	0.41

Outline

Concluding remarks

- Develop cut-like IP formulations for the MRCP when $s \in \{2, 3, 4\}$.
- Establish complexity results of the decision version of the MRCP.
- Devise BC algorithms for the MRCP when $s \in \{2, 3, 4\}$.
- Recursive block decomposition algorithm is effective for solving the MRCP.
- Our computational studies include the first reported numerical results for the MRCP when $s \in \{3, 4\}$.
- The results also extend to the ”hereditary” counterpart.

Manuscript

Code

THANK YOU

Q & A

ylu@jsu.edu
http://yajunlu.com

Reference I

M. T. Almeida and F. D. Carvalho. An analytical comparison of the LP relaxations of integer models for the k-club problem. *European Journal of Operational Research*, 232(3):489–498, 2014.

G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. URL `http://igraph.org`.

E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.

N. Gould and J. Scott. A note on performance profiles for benchmarking software. *ACM Transactions on Mathematical Software (TOMS)*, 43(2):1–5, 2016.

J. Hopcroft and R. Tarjan. Algorithm 447: Efficient algorithms for graph manipulation. *Communications of the ACM*, 16(6):372–378, 1973.

C. Komusiewicz, A. Nichterlein, R. Niedermeier, and M. Picker. Exact algorithms for finding well-connected 2-clubs in sparse real-world graphs: Theory and experiments. *European Journal of Operational Research*, 275(3):846–864, 2019.

L. Lovász, V. Neumann-Lara, and M. Plummer. Mengerian theorems for paths of bounded length. *Periodica Mathematica Hungarica*, 9(4):269–276, 1978.

Reference II

H. Salemi and A. Buchanan. Parsimonious formulations for low-diameter clusters. *Mathematical Programming Computation*, 12(3):493–528, 2020. doi: 10.1007/s12532-020-00175-6. URL https://doi.org/10.1007/s12532-020-00175-6.

A. Veremyev and V. Boginski. Identifying large robust network clusters via new compact formulations of maximum *k*-club problems. *European Journal of Operational Research*, 218(2):316–326, 2012.

A heuristic for finding an r-robust s-club

Algorithm 2: A heuristic for finding an r-robust s-club

Input: A graph $G = (V, E)$.
Output: An r-robust s-club S.

create compatibility graph $G^c \leftarrow (V, E^c)$, where $E^c := \left\{ ij \in \binom{V}{2} \mid \rho_s(G; i, j) \geq r \right\}$
$S \leftarrow$ a maximal clique in G^c
while $S \neq \emptyset$ **do**
  $\tau_i \leftarrow 0, \forall i \in S$
  for $ij \in \binom{S}{2}$ **do**
    if $\rho_s(G[S]; i, j) \leq r - 1$ **then**
      $\tau_i \leftarrow \tau_i + 1$
      $\tau_j \leftarrow \tau_j + 1$
  $v \leftarrow \arg\max_{i \in S} \tau_i$
  if $\tau_v \geq 1$ **then**
    $S \leftarrow S \setminus \{v\}$
  else
    return S

Vertex peeling based on an r-robust s-club of size ℓ

Algorithm 3: Vertex peeling based on an r-robust s-club of size ℓ

Input: A graph $G = (V, E)$ and a lower bound ℓ.
Output: Preprocessed graph G.

```
repeat
    G ← the r-core of G
    S ← ∅
    for v ∈ V(G) do
        if |N_G^s(v)| < ℓ or |T_v| < ℓ then
            S ← S ∪ {v}
    if S ≠ ∅ then
        G ← G − S
until S = ∅
return G
```

where $T_v := \{u \in N_G^s(v) \mid \rho_s(G; v, u) \geq r\}$