Clique Relaxations in Complex Networks: Foundations and Algorithms

Sergiy Butenko (Texas A&M University)
Baski Balasundaram (Oklahoma State University)
Vladimir Boginski (University of Central Florida)

TEXAS A&M UNIVERSITY

Brief summary

- FA9550-12-1-0103 “Clique Relaxations in Biological and Social Networks: Foundations and Algorithms”
- PI: Sergiy Butenko; co-PIs: Balabhaskar Balasundaram and Vladimir Boginski
- Dates: July 1, 2012 – June 30, 2015
- 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.

Network-based analysis of big data

Big data arising in various complex systems can be conveniently modeled using networks/graphs:

- components of the complex system – vertices
- pairwise interactions between different components – edges

Network-based analysis allows to capture some global structural properties of the system and predict overall trends in its dynamics.

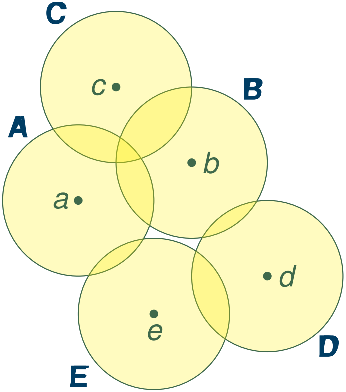
Clusters in networks

In many applications one is interested in detecting/designing cohesive clusters

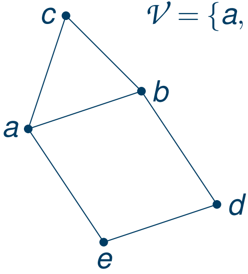
- Communication/sensor networks
- Transportation/supply networks
- Power grid
- Biological networks
- Social networks
- Financial networks
- Electronic warfare

Geometric graphs

A unit-disk graph (UDG) can be defined as the intersection graph of closed disks of equal (e.g., unit) diameter.



$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$
$$\mathcal{V} = \{a, b, c, d, e\}$$



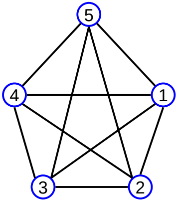
Social networks

A social network is described by $G = (V, E)$ where V is the set of "actors" and E is the set of "ties".

- actors are people and a tie exists if two people know each other.
- actors are wire transfer database records and a tie exists if two records have the same *matching field*.
- *Cohesive subgroups* are "closely knit groups" in a social network.
- *Social cohesion* is often used to explain and develop sociological theories.

Cliques

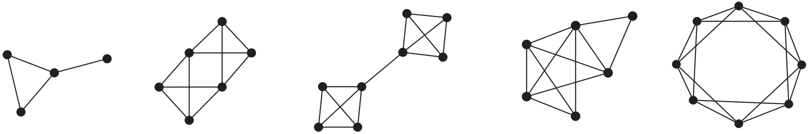
- *Etymology:* The term *clique* originates from Old French *cliquer* meaning *make a noise*
- *WordNet dictionary definition:* an exclusive circle of people with a common purpose
- Luce and Perry (1949): social clique – a group of people that know (are friends of) all other people in the group

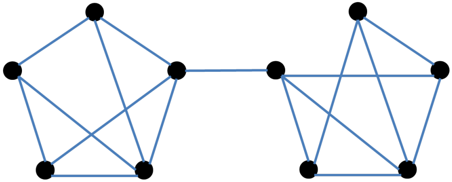


Clusters in real-life networks

Cliques may be overly restrictive for practical purposes

1xg0 (immune sys.) 1p9m (signaling) 1dxr (photosynthesis) 1ruz (viral protein) 1kw6 (lyase)





Alternatives to clique

$G = (V, E)$. $S \subseteq V$ is

- s**-clique** if $d_G(v, v') \leq s$, for any $v, v' \in S$ (Luce 1950)
- s**-club** if $diam(G[S]) \leq s$ (Alba 1973, Mokken 1979)
- s**-plex** if $\delta(G[S]) \geq |S| - s$ (Seidman & Foster 1978)
- s**-defective clique** if $G[S]$ has at least $\binom{|S|}{2} - s$ edges (Yu et al. 2006)
- k**-core** if $\delta(G[S]) \geq k$ (Seidman 1983)
- k**-block** if $\kappa(G[S]) \geq k$ (Moody & White 2003)
- γ**-quasi-clique** if $\rho(G[S]) \geq \gamma$ (Abello et al. 2002)
- (λ, γ)**-quasi-clique** if $\delta(G[S]) \geq \lambda(|S| - 1)$ and $\rho(G[S]) \geq \gamma$ (Brunato et al. 2008)
- ...

TEXAS A&M UNIVERSITY

Alternative clique definitions

(a) Vertices are **distance *one*** away from each other

(b) Vertices induce a subgraph of **diameter *one***

(c) Every ***one*** vertex forms a **dominating set**

(d) **Degree:** Each vertex neighbors all vertices

(e) **Density:** Vertices induce a subgraph that has all possible edges

(f) **Connectivity:** need to be remove all vertices to obtain a disconnected induced subgraph

Defining clique relaxations

We can define clique relaxations by

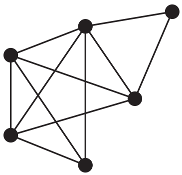
(i) **restricting** a **violation** of an elementary clique-defining property
or by

(ii) **ensuring** the **presence** of an *elementary clique-defining property*

Relative relaxations

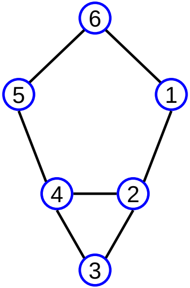
Vertices induce a subgraph that has **the fraction** γ of all possible edges – **γ-quasi-clique**

1ruz (viral protein)



.7-quasiclique

Regular and weak relaxations



- $\{2,3,4\}$ is a 1-club ... the "regular" clique
- $\{1,2,4,5,6\}$ is a 2-club
- $\{1,2,3,4,5\}$ is a 2-clique but NOT a 2-club
- maximality of a 2-club is harder to test

s-clique appears to be a **weaker** cluster than *s*-club

Weak clique relaxations

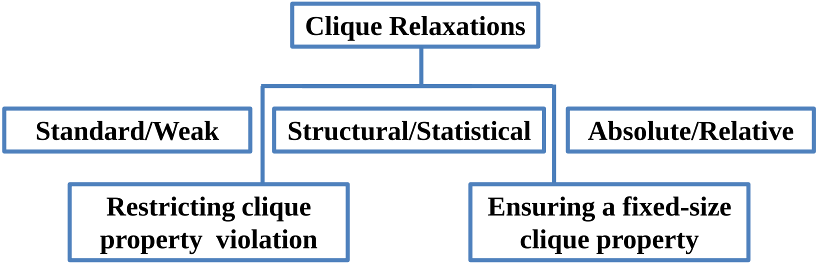
Distance-based: ***s*-clique (weak *s*-club)**

Vertices in S are **distance at most** s away from each other **in** G.

Connectivity-based: **weak *k*-block**

Any two vertices in **S** have **at least** k vertex-independent **paths** between them **in** G

Clique relaxations taxonomy



J. Pattillo, N. Youssef, and S. Butenko. On clique relaxation models in network analysis. *European Jour. of Oper. Res.*, 226: 9–18, 2013.

Additional elementary clique-defining properties

A subset of vertices C is a clique in G if and only if one of the following conditions hold:

(g) Independence number $\alpha(G[C]) = 1$;

(h) Vertex cover number $\tau(G[C]) = |C| - 1$;

(i) Chromatic number $\chi(G[C]) = |C|$;

(j) Clique cover number $\bar{\chi}(G[C]) = 1$;

(k) Edge connectivity number $\lambda(G[C]) = |C| - 1$.

Higher order clique relaxations

***Simple* Higher Order Relaxations**: relaxing multiple elementary clique-defining properties simultaneously

(λ, γ)*-quasiclique*: Each vertex is connected to *at least* $\lambda(|S|-1)$ vertices, and the induced subgraph has at least *the fraction* γ of all possible edges.

***k-Hereditary* Higher Order Relaxations**: connectivity *embedded* into the definition

k-hereditary s-club: The induced subgraph is not only an s-club, but also the removal of up to k vertices still preserves the s-club property.

Cohesiveness properties

$S \subseteq V$	Diameter	Dominating Set	Minimum Degree	Edge Density	Connectivity
Clique	"one"	"one"	"all"	"one"	"all"
s-club	s	$\lvert S\rvert - 1$	1	$\frac{2}{\lvert S\rvert}$	1
s-plex	s	s	$\lvert S\rvert - s$	$1 - \frac{s-1}{\lvert S\rvert-1}$	$\lvert S\rvert - 2s + 2$
k-core	d'_k	$\lvert S\rvert - k$	k	$\frac{k}{\lvert S\rvert-1}$	$2k + 2 - \lvert S\rvert$
γ-quasi-clique	d_γ	$\lvert S\rvert$	$\left\lceil \gamma\binom{\lvert S\rvert}{2} - \binom{\lvert S\rvert-1}{2} \right\rceil$	γ	$\left\lceil \gamma\binom{\lvert S\rvert}{2} - \binom{\lvert S\rvert-1}{2} \right\rceil$
k-block	$\left\lfloor \frac{\lvert S\rvert-2}{k} + 1 \right\rfloor$	$\lvert S\rvert - k$	k	$\frac{k}{\lvert S\rvert-1}$	k

$$d'_k = \max\left\{ \left\lceil \frac{|S|}{k+1} \right\rceil, 3\left(\left\lfloor \frac{|S|-z}{k+1} \right\rfloor - 1 \right) + z, z \in \{0,1,2\} \right\}$$

$$d_\gamma = \left\lfloor |S| + \frac{1}{2} - \sqrt{\gamma|S|^2 - (2+\gamma)|S| + \frac{17}{4}} \right\rfloor.$$

Structiural properties

Definition (Heredity)

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.

Definition (Weak heredity)

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.

Structiural properties

Definition (Quasi-heredity)

A graph property Π is said to be *quasi-hereditary*, if for any graph $G = (V, E)$ with property Π and for any size $0 \leq r < |V|$, there exists some subset $R \subset S$ with $|R| = r$, such that $G[S \setminus R]$ demonstrates property Π.

Definition (k-Heredity)

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.

Hereditary clique relaxations

- *s*-plex, *s*-defective clique, *s*-bundle

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: 113–130, 2013.

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

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

Weakly hereditary clique relaxations

Weak clique relaxations:

- s-clique
- weak k-block

Optimization problems for these models can be reduced to maximum clique in corresponding auxiliary graphs.

Quasi-hereditary clique relaxations

- quasi-clique

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

A. Veremyev, O. A. Prokopyev, S. Butenko, and E. L. Pasiliao. Exact MIP-based approaches for finding maximum quasi-cliques and dense subgraphs. *Computational Optimization and Applications*, to appear, DOI 10.1007/s10589-015-9804-y.

Z. Miao and B. Balasundaram. Lagrangian dual bounds for the maximum quasi-clique problem. Working paper.

Non-hereditary clique relaxations

- *s*-club and variations

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 et al., Spinger Science + Business Media, 2013, pp.149–174.

E. Moradi and B. Balasundaram. Finding a maximum *k*-club using the *k*-clique formulation and canonical hypercube cuts. *Optimization Letters*, to appear, DOI 10.1007/s11590-015-0971-7.

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

Non-hereditary clique relaxations

- *s*-club and variations

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

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

Non-hereditary clique relaxations

- k-cores, k-blocks and variations

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

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: 349–362, 2014.

J. Ma and B. Balasundaram. On the chance-constrained minimum spanning k-core problem. Working paper.

Non-hereditary clique relaxations

- connectivity, clustering coefficient based

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: 178–188, 2015.

Y. Wang, A. Buchanan, and S. Butenko. On imposing connectivity constraints in integer programs. Submitted to *Mathematical Programming*, June 2015. Available online at http://www.optimizationonline. org/DB HTML/2015/02/4768.html

Z. Ertem, A. Veremyev, and S. Butenko. Detecting large cohesive subgroups with high clustering coefficients in social networks. *Social Networks*, conditionally accepted.

Scale-reduction techniques

Using scale reduction techniques based on clique relaxations in conjunction with Östergård's max clique algorithm the clique number was obtained on all graphs in the SNAP database and 10-th DIMACS Implementation Challenge (graphs with up to $\approx$18.5 million vertices)

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: 164–177, 2015.

INFORMS Connect President's Pick for May 2015.

Cliques revisited

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: 1611–1617, 2014.

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

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

Fractional objective

Definition

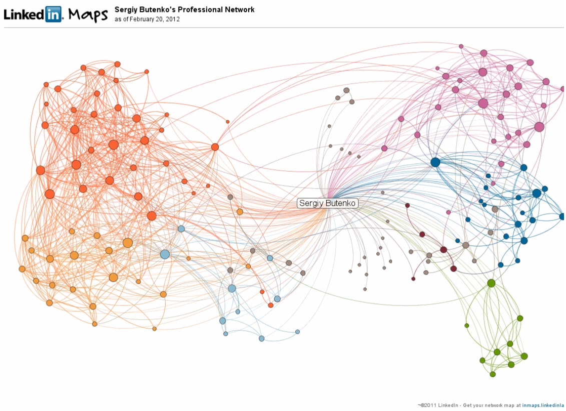
Given a simple undirected graph $G = (V, E)$, where each vertex $i \in V$ is assigned two non-negative rational weights, a_i and b_i, the maximum ratio clique problem (MRCP) is to find a maximal clique C in G that maximizes the quantity $\frac{\sum_{i \in C} a_i}{\sum_{i \in C} b_i}$.

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

Runner-up, best student paper award at the CMS-2013 conference in Montreal

Clustering

Partitioning a set of entities into ‘natural groups’ (clusters)



Fortunato’s 2010 survey has over 3,700 citations (Google Scholar).

k-Community clustering

- Introduced a general purpose clustering algorithm based on clique relaxations.
- Do not aim to optimize any standard performance measure.
- Using *k*-community as a structure does well for a number of clustering quality measures.
- Enhancements to the basic algorithm can be designed according to requirements.

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, 2013, pp.125–136.

Extensions: Analysis of heuristics

"Club sandwich"

$$\omega(G) \leq \Delta(G) + 1 \leq \bar{\omega}_2(G),$$

$\Delta(G)$ is the largest degree, $\omega_2(G)$ is the 2-club number

Definition

A heuristic is said to be *provably best* for an optimization problem if, assuming $\mathcal{P} \neq \mathcal{NP}$, there is no polynomial-time algorithm that always finds a better solution (when one exists).

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

"The whole is more than the sum of its parts."
–Aristotle (384-322 BC)

Thank you!