Finding the most influential group in a complex network

Foad Mahdavi Pajouh

Assistant Professor
Management Science and Information Systems Department
University of Massachusetts Boston
foad.mahdavi@umb.edu

UMASS BOSTON

Based on

M. Rysz, F. Mahdavi Pajouh and E. L. Pasiliao. *Finding clique clusters with the highest betweenness centrality*. European Journal of Operational Research, 271(1):155-164, 2018.

UMASS BOSTON

Outline

UMASS BOSTON

Outline

UMASS BOSTON

Central groups in network systems

- An important concept in analyzing complex systems modeled as networks involves identifying components/members that are "influential" or "central".

Central groups in network systems

- An important concept in analyzing complex systems modeled as networks involves identifying components/members that are "influential" or "central".
- The vast majority of the literature emphasizes centrality of individual members. An important extension involves finding the most central groups embedded in a network (Everett and Borgatti (1999); Everett and P. (2005); Ni et al. (2011)).

UMASS BOSTON

Central groups in network systems

- An important concept in analyzing complex systems modeled as networks involves identifying components/members that are "influential" or "central".
- The vast majority of the literature emphasizes centrality of individual members. An important extension involves finding the most central groups embedded in a network (Everett and Borgatti (1999); Everett and P. (2005); Ni et al. (2011)).
- This setting becomes particularly relevant when one additionally imposes a structural property Π (e.g., a clique) that the selected set/group must satisfy, depending on the application.

UMASS BOSTON

The most central set problem (MCSP)

The most central set problem (MCSP) must satisfy multiple criteria:

- An inter-vertex structure defined by Π;
- It is the most central such set in the network.

The most central set problem (MCSP)

The most central set problem (MCSP) must satisfy multiple criteria:

- An inter-vertex structure defined by Π;
- It is the most central such set in the network.

Several relevant considerations:

- Does the group centrality measures consider the interactions among the set members, interactions between set members and non-member vertices, and/or interactions exclusively between non-member vertices?
- Whether Π itself affects centrality? For instance: property Π does not affect group closeness centraltiy, which measures the shortest distances between a set and the non-member vertices.
- Whether a larger set is more central than any of its subsets? This holds profound theoretical and algorithmic implications when solving the MCSP.

UMASS BOSTON

Most betweenness-central clique problem (MBCP)

Definition 1

Given a simple undirected connected graph $G = (V, E)$, a set $D \subseteq V$, and a pair of vertices $i, j \in V$ $(i \neq j)$, let σ_{ij} and $\sigma_{ij}(D)$ denote the total number of shortest paths between i and j in G, and the number of shortest paths between i and j in G that intersect set D, respectively.

Most betweenness-central clique problem (MBCP)

Definition 1

Given a simple undirected connected graph $G = (V, E)$, a set $D \subseteq V$, and a pair of vertices $i, j \in V$ $(i \neq j)$, let σ_{ij} and $\sigma_{ij}(D)$ denote the total number of shortest paths between i and j in G, and the number of shortest paths between i and j in G that intersect set D, respectively.

Definition 2

Given sets $D \subseteq V$ and $H \subseteq V$, the betweenness centrality of D with respect to H (denoted by $C_H(D)$) is defined as

$$C_H(D) = \sum_{i,j \in H, i<j} \frac{\sigma_{ij}(D)}{\sigma_{ij}}. \tag{1}$$

If $H = V$, then $C_V(D)$ is referred to as the betweenness centrality of set D and the subscript V is removed from the notation for simplicity.

Most betweenness-central clique problem (MBCP)

Definition 3

A set $D \subseteq V$ is a clique, if $G[D]$ is a complete subgraph, i.e., there is an edge between any pair of vertices in D.

Most betweenness-central clique problem (MBCP)

Definition 3

A set $D \subseteq V$ is a clique, if $G[D]$ is a complete subgraph, i.e., there is an edge between any pair of vertices in D.

Definition 4

The most betweenness-central clique problem (MBCP):

$$\max\left\{C(D) : D \subseteq V \text{ is a clique}\right\}. \tag{2}$$

UMASS BOSTON

Potential applications

To identify a group that intercepts large portion of information and quickly disseminates it among its members, it becomes of interest to find a "tightly knit" group with high group betweenness centrality.

UMASS BOSTON

Potential applications

To identify a group that intercepts large portion of information and quickly disseminates it among its members, it becomes of interest to find a "tightly knit" group with high group betweenness centrality.

Overall, such frameworks hold promise in developing strategies for product marketing purposes, corporate information sharing, internet traffic monitoring, and so on.

UMASS BOSTON

Potential applications

- To detect components that are critical for preserving the flow of electricity through power grids (Barabasi and Albert (1999)).

Potential applications

- To detect components that are critical for preserving the flow of electricity through power grids (Barabasi and Albert (1999)).
- To identify central positions within in-task networks who can accelerate the speed of acquiring information necessary for task completion (Wu et al. (2008)).

Potential applications

- To detect components that are critical for preserving the flow of electricity through power grids (Barabasi and Albert (1999)).
- To identify central positions within in-task networks who can accelerate the speed of acquiring information necessary for task completion (Wu et al. (2008)).
- To improve insight about firms' innovation potential relative to their embeddedness in inter-firm relations networks (Gilsing et al. (2008)).

UMASS BOSTON

Potential applications

- To detect components that are critical for preserving the flow of electricity through power grids (Barabasi and Albert (1999)).
- To identify central positions within in-task networks who can accelerate the speed of acquiring information necessary for task completion (Wu et al. (2008)).
- To improve insight about firms' innovation potential relative to their embeddedness in inter-firm relations networks (Gilsing et al. (2008)).
- To identify "gatekeepers" in social and communication networks who control the information flow through the network (Borgatti (2005); David and Jon (2010)).

UMASS BOSTON

Potential applications

- To detect components that are critical for preserving the flow of electricity through power grids (Barabasi and Albert (1999)).
- To identify central positions within in-task networks who can accelerate the speed of acquiring information necessary for task completion (Wu et al. (2008)).
- To improve insight about firms' innovation potential relative to their embeddedness in inter-firm relations networks (Gilsing et al. (2008)).
- To identify "gatekeepers" in social and communication networks who control the information flow through the network (Borgatti (2005); David and Jon (2010)).
- To identify elements in Internet traffic and scientific collaboration networks whose existence is critical for preserving the network's performance (Ishakian et al. (2012)).

UMASS BOSTON

Previous work

- Among recent theoretical developments, several articles put emphasis on mixed-integer programming formulations for detecting groups of a given cardinality with the highest centrality.

Previous work

- Among recent theoretical developments, several articles put emphasis on mixed-integer programming formulations for detecting groups of a given cardinality with the highest centrality.
- Vogiatzis et al. (2015):
  - They proposed models for finding cliques of a given size with the highest level of degree, closeness as well as several versions of betweenness centrality.
  - The authors considered "optimistic" and "pessimistic" variants, where vertices reward or penalize the shortest paths intersecting the selected clique, respectively.

UMASS BOSTON

Previous work

- Veremyev et al. (2016):
  - They introduced several additional models for finding betweenness-central groups of a predefined size.
  - The authors implemented the following variants of betweenness centrality into their models: distance-scaled, stress, proximal source/target, bounded-distance and pair-weighted.
  - They also utilized the relations between bounded-distance betweenness and group betweenness centralities to develop an algorithm for identifying sets with *approximately* highest group betweeness centrality.
  - This approach was implemented to find central sets that both exhibit and do not exhibit an assigned cohesiveness property.

Our contributions

- The *most betweenness-central clique problem* is introduced and studied.

Our contributions

- The *most betweenness-central clique problem* is introduced and studied.
- The decision version of this problem is proven to be NP-complete.

Our contributions

- The *most betweenness-central clique problem* is introduced and studied.
- The decision version of this problem is proven to be NP-complete.
- An analytical bounding scheme for the centrality of a maximal clique is proposed.

Our contributions

- The *most betweenness-central clique problem* is introduced and studied.
- The decision version of this problem is proven to be NP-complete.
- An analytical bounding scheme for the centrality of a maximal clique is proposed.
- A combinatorial branch-and-bound algorithm for solving this problem is developed.

UMASS BOSTON

Our contributions

- The *most betweenness-central clique problem* is introduced and studied.
- The decision version of this problem is proven to be NP-complete.
- An analytical bounding scheme for the centrality of a maximal clique is proposed.
- A combinatorial branch-and-bound algorithm for solving this problem is developed.
- Results of numerical experiments with real-life and random graphs are provided.

UMASS BOSTON

Outline

Theoretical properties of betweenness centrality

Lemma 1 (Rysz et al. (2017))

Given sets $D \subseteq V$, $H \subseteq V$, and $D' \subseteq D$, we have

1. $C_H(D') \leq C_H(D)$ *(i.e. centrality measure $C(\cdot)$ is monotone).*
2. $C_H(D) \leq C_H(D \setminus D') + C_H(D')$.

UMASS BOSTON

Theoretical properties of betweenness centrality

Lemma 1 (Rysz et al. (2017))

Given sets $D \subseteq V$, $H \subseteq V$, and $D' \subseteq D$, we have

1. $C_H(D') \leq C_H(D)$ *(i.e. centrality measure $C(\cdot)$ is monotone).*
2. $C_H(D) \leq C_H(D \setminus D') + C_H(D')$.

Corollary 1 (Rysz et al. (2017))

There exists a maximal clique in G that is an optimal solution to the MBCP.

Theoretical properties of betweenness centrality

Definition 5 (Rysz et al. (2017))

Given vertices $i, j \in V$ $(i \neq j)$, and sets $D_1, D_2 \subseteq V$, let $\sigma_{ij}(D_1, D_2)$ denote the total number of shortest paths between i and j in G whose intersection with set D_1 is equal to set D_2.

Theoretical properties of betweenness centrality

Definition 5 (Rysz et al. (2017))

Given vertices $i, j \in V$ $(i \neq j)$, and sets $D_1, D_2 \subseteq V$, let $\sigma_{ij}(D_1, D_2)$ denote the total number of shortest paths between i and j in G whose intersection with set D_1 is equal to set D_2.

Lemma 2 (Rysz et al. (2017))

Given a simple undirected graph $G = (V, E)$, a clique $D \subseteq V$, vertices $i, j \in V \setminus D$, and vertices $k, l \in D$, we have

1. $\sigma_{ij}(D, \{k, l\}) = \sigma_{ij}(\{k, l\}, \{k, l\})$.
2. $\sigma_{ij}(D, \{k\}) = \sigma_{ij}(\{k\}) - \sum_{u \in D \setminus \{k\}} \sigma_{ij}(\{k, u\}, \{k, u\})$.
3. $\sigma_{ij}(D) = \sum_{u \in D} \sigma_{ij}(\{u\}) - \sum_{u,v \in D, u<v} \sigma_{ij}(\{u, v\}, \{u, v\})$.

Theoretical properties of betweenness centrality

Lemma 3 (Rysz et al. (2017))

Given a clique $D \subseteq V$, and a vertex $k \in V \setminus D$ that forms a clique with set D, we have

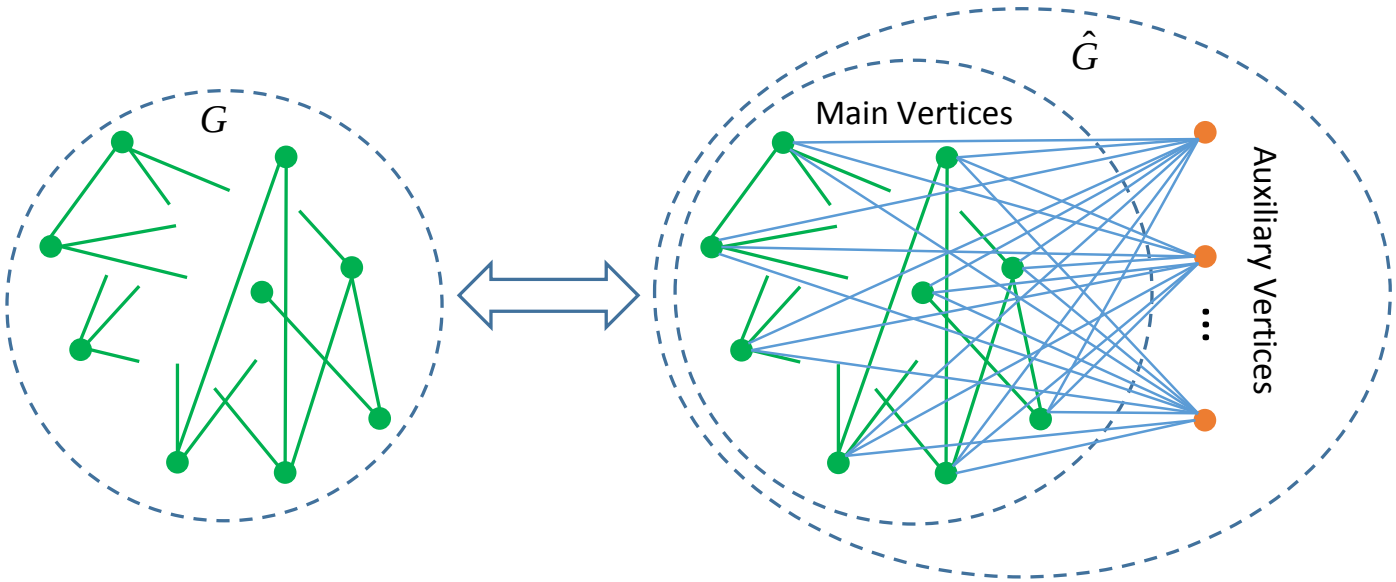
$$C_{V\setminus(D\cup\{k\})}(D\cup\{k\}) = C_{V\setminus D}(D) - \sum_{j\in V\setminus(D\cup\{k\})} \frac{1}{\sigma_{kj}}[\sum_{u\in D}\sigma_{kj}(\{u\})] + \sum_{i,j\in V\setminus(D\cup\{k\}),i<j} \frac{1}{\sigma_{ij}}[\sigma_{ij}(\{k\}) - \sum_{u\in D}\sigma_{ij}(\{k,u\},\{k,u\})].$$

Computational complexity of MBCP

Theorem 1 (Rysz et al. (2017))

The decision version of the MBCP is NP-complete.

Reduction from Maximum Clique Problem:



Outline

UMASS BOSTON

Search tree structure

- Each node t of the search tree consists of a set of fixed vertices (F_t) and a set of candidate vertices (U_t).

Search tree structure

- Each node t of the search tree consists of a set of fixed vertices (F_t) and a set of candidate vertices (U_t).
- Set F_t contains all vertices that are part of the maximal clique being constructed along the unique search path originating at the root node of the search tree and ending at the present node t.

Search tree structure

- Each node t of the search tree consists of a set of fixed vertices (F_t) and a set of candidate vertices (U_t).
- Set F_t contains all vertices that are part of the maximal clique being constructed along the unique search path originating at the root node of the search tree and ending at the present node t.
- Set U_t contains all vertices in $V \setminus F_t$ that can be added to set F_t to from a larger clique that contains clique F_t.

Search tree structure

- Each node t of the search tree consists of a set of fixed vertices (F_t) and a set of candidate vertices (U_t).
- Set F_t contains all vertices that are part of the maximal clique being constructed along the unique search path originating at the root node of the search tree and ending at the present node t.
- Set U_t contains all vertices in $V \setminus F_t$ that can be added to set F_t to from a larger clique that contains clique F_t.
- Each parent node t will have $|U_t|$ child nodes, each one corresponding to the selection of a branching vertex $i \in U_t$ and to moving i from set U_t to set F_t.

Search tree structure

- Each node t of the search tree consists of a set of fixed vertices (F_t) and a set of candidate vertices (U_t).
- Set F_t contains all vertices that are part of the maximal clique being constructed along the unique search path originating at the root node of the search tree and ending at the present node t.
- Set U_t contains all vertices in $V \setminus F_t$ that can be added to set F_t to from a larger clique that contains clique F_t.
- Each parent node t will have $|U_t|$ child nodes, each one corresponding to the selection of a branching vertex $i \in U_t$ and to moving i from set U_t to set F_t.
- Based on empirical observations, a *depth-first-search (DFS)* strategy for creating the nodes of the search tree is utilized.

Search tree structure

- Each node t of the search tree consists of a set of fixed vertices (F_t) and a set of candidate vertices (U_t).
- Set F_t contains all vertices that are part of the maximal clique being constructed along the unique search path originating at the root node of the search tree and ending at the present node t.
- Set U_t contains all vertices in $V \setminus F_t$ that can be added to set F_t to from a larger clique that contains clique F_t.
- Each parent node t will have $|U_t|$ child nodes, each one corresponding to the selection of a branching vertex $i \in U_t$ and to moving i from set U_t to set F_t.
- Based on empirical observations, a *depth-first-search (DFS)* strategy for creating the nodes of the search tree is utilized.
- When creating a child node of a node t (node $t+1$), a previously unselected branching vertex $i \in U_t$ with the maximum value of $C_{V\setminus(F_t\cup\{i\})}(\{i\})$ is chosen.

UMASS BOSTON

Upper-bounding approach

Definition 6

A feasible coloring of a graph is an assignment of colors to its vertices in a way that no two adjacent vertices receive the same color.

Upper-bounding approach

Definition 6

A feasible coloring of a graph is an assignment of colors to its vertices in a way that no two adjacent vertices receive the same color.

Proposition 1

Given a search tree node t with a nonempty candidate set U_t, the betweenness centrality of any maximal clique containing clique F_t is less than or equal to

$$(|F_t|+\chi(U_t))(\frac{|F_t|+\chi(U_t)-1}{2}+|V|-|F_t|-\chi(U_t))+C_{V\setminus F_t}(F_t)+\sum_{r=1}^{\chi(U_t)} c_r, \quad (3)$$

where $\chi(U_t)$ is the number of colors used in a feasible coloring of $G[U_t]$, and vector $\mathbf{c}=[c_r \in \mathbb{Q}]$ $(r=1,...,|U_t|)$ contains $C_{V\setminus(F_t\cup\{i\})}(\{i\})$ for all $i \in U_t$ in decreasing order of their value.

Upper-bounding approach

- To calculate $\chi(U_t)$, a greedy approximate heuristic (Tomita et al. (2010)) is used to obtain a feasible coloring of $G[U_t]$.

Upper-bounding approach

- To calculate $\chi(U_t)$, a greedy approximate heuristic (Tomita et al. (2010)) is used to obtain a feasible coloring of $G[U_t]$.
- Denoting the parent node of node t by node p, the values of $C_{V\setminus F_t}(F_t)$ and $C_{V\setminus(F_t\cup\{i\})}(\{i\})$ for all $i \in U_t$ are calculated by updating the values of $C_{V\setminus F_p}(F_p)$ and $C_{V\setminus(F_p\cup\{i\})}(\{i\})$ for all $i \in U_t$ as follows.

Upper-bounding approach

- To calculate $\chi(U_t)$, a greedy approximate heuristic (Tomita et al. (2010)) is used to obtain a feasible coloring of $G[U_t]$.
- Denoting the parent node of node t by node p, the values of $C_{V\setminus F_t}(F_t)$ and $C_{V\setminus(F_t\cup\{i\})}(\{i\})$ for all $i \in U_t$ are calculated by updating the values of $C_{V\setminus F_p}(F_p)$ and $C_{V\setminus(F_p\cup\{i\})}(\{i\})$ for all $i \in U_t$ as follows.
  - Let k denote the branching vertex in child node t. Since set $F_p \cup \{k\}$ forms a clique, Lemma 3 can be used to calculate $C_{V\setminus F_t}(F_t)$ by updating $C_{V\setminus F_p}(F_p)$.

Upper-bounding approach

- To calculate $\chi(U_t)$, a greedy approximate heuristic (Tomita et al. (2010)) is used to obtain a feasible coloring of $G[U_t]$.
- Denoting the parent node of node t by node p, the values of $C_{V\setminus F_t}(F_t)$ and $C_{V\setminus(F_t\cup\{i\})}(\{i\})$ for all $i \in U_t$ are calculated by updating the values of $C_{V\setminus F_p}(F_p)$ and $C_{V\setminus(F_p\cup\{i\})}(\{i\})$ for all $i \in U_t$ as follows.
  - Let k denote the branching vertex in child node t. Since set $F_p \cup \{k\}$ forms a clique, Lemma 3 can be used to calculate $C_{V\setminus F_t}(F_t)$ by updating $C_{V\setminus F_p}(F_p)$.
  - The values of $C_{V\setminus(F_t\cup\{i\})}(\{i\})$ for all $i \in U_t$ can be calculated by:

$$C_{V\setminus(F_t\cup\{i\})}(\{i\}) = C_{V\setminus(F_p\cup\{i\})}(\{i\}) - \sum_{j\in V\setminus(F_t\cup\{i\})} \frac{\sigma_{kj}(\{i\})}{\sigma_{kj}}, \quad \forall i \in U_t. \tag{4}$$

Upper-bounding approach

- To calculate $\chi(U_t)$, a greedy approximate heuristic (Tomita et al. (2010)) is used to obtain a feasible coloring of $G[U_t]$.
- Denoting the parent node of node t by node p, the values of $C_{V\setminus F_t}(F_t)$ and $C_{V\setminus(F_t\cup\{i\})}(\{i\})$ for all $i \in U_t$ are calculated by updating the values of $C_{V\setminus F_p}(F_p)$ and $C_{V\setminus(F_p\cup\{i\})}(\{i\})$ for all $i \in U_t$ as follows.
  - Let k denote the branching vertex in child node t. Since set $F_p \cup \{k\}$ forms a clique, Lemma 3 can be used to calculate $C_{V\setminus F_t}(F_t)$ by updating $C_{V\setminus F_p}(F_p)$.
  - The values of $C_{V\setminus(F_t\cup\{i\})}(\{i\})$ for all $i \in U_t$ can be calculated by:

$$C_{V\setminus(F_t\cup\{i\})}(\{i\}) = C_{V\setminus(F_p\cup\{i\})}(\{i\}) - \sum_{j\in V\setminus(F_t\cup\{i\})} \frac{\sigma_{kj}(\{i\})}{\sigma_{kj}}, \quad \forall i \in U_t. \tag{4}$$

  - The values of σ_{ij}, $\sigma_{ij}(\{u\})$ and $\sigma_{ij}(\{v,w\},\{v,w\})$ for all $i,j \in V$ $(i \neq j)$, $u \in V$, and $(v,w) \in E$ used in Lemma 3 and Equation (4) are calculated and stored before starting the proposed CBB algorithm.

Combinatorial branch-and-bound (CBB) algorithm

Algorithm 1: Depth-first-search CBB algorithm for MBCP

Initialize: $t \leftarrow 0$; $U_0 \leftarrow V$; $F_0 \leftarrow \emptyset$; $F^* \leftarrow \emptyset$; $C^* \leftarrow 0$;
while $t \geq 0$ **do**
  while $U_t \neq \emptyset$ **do**
    select vertex $i \in \arg\max\{C_{V\setminus(F_t\cup\{i\})}(\{i\}) : i \in U_t\}$;
    $U_t \leftarrow U_t \setminus \{i\}$;
    $F_{t+1} \leftarrow F_t \cup \{i\}$;
    $U_{t+1} \leftarrow \{v \in U_t : \{v\} \cup F_{t+1}$ is a clique$\}$;
    if $U_{t+1} = \emptyset$ **then**
      if $C(F_{t+1}) > C^*$ **then**
        $C^* \leftarrow C(F_{t+1})$;
        $F^* \leftarrow F_{t+1}$;
    else
      if $(|F_{t+1}| + \chi(U_{t+1}))(\frac{|F_{t+1}| + \chi(U_{t+1}) - 1}{2} + |V| - |F_{t+1}| - \chi(U_{t+1})) + C_{V\setminus F_{t+1}}(F_{t+1}) + \sum_{r=1}^{\chi(U_{t+1})} c_r > C^*$ **then**
        $t \leftarrow t + 1$;
  $t \leftarrow t - 1$;
return C^*, F^*;

Outline

Equivalent linear 0–1 program (Veremyev et al. (2016))

$$\max \quad \sum_{i,j \in V, i<j} \frac{\sum_{P_t \in \mathcal{P}_{ij}} y_{ij}^t}{\sigma_{ij}} \tag{5a}$$

$$\text{s. t.} \quad y_{ij}^t \leq \sum_{q \in P_t} x_q, \quad \forall i,j \in V,\ i<j,\ \forall P_t \in \mathcal{P}_{ij}, \tag{5b}$$

$$x_i + x_j \leq 1, \quad \forall (i,j) \in \overline{E}, \tag{5c}$$

$$x_i \in \{0,1\},\ \ y_{ik} \in [0,1],\ \ \forall i,j \in V,\ \ i<j,\ \ \forall P_t \in \mathcal{P}_{ij}, \tag{5d}$$

where $\mathcal{P}_{ij}$ denotes the set of all shortest paths between $i,j \in V$, the binary variable x_i determines whether vertex $i \in V$ is included in the optimal clique, and y_{ij}^t indicates whether the path $P_t \in \mathcal{P}_{ij}$ coincides with the optimal clique.

Setup of numerical experiments

- Several real-life graphs obtained from the DIMACS library and Network Repository (Rossi and Ahmed (2015)).

Setup of numerical experiments

- Several real-life graphs obtained from the DIMACS library and Network Repository (Rossi and Ahmed (2015)).
- Randomly generated graphs with $|V| = 150$:
  - *Uniform random graphs* (Erdös and Rényi (1959)): Notation ER-150-$|E|$ is used, where $|E| = 1000, 1500, 2000, 2500, 3000$.
  - *Power-law random graphs* (Albert and Brabasi (2002)): Notation BA-150-k is used, where k represents the number of edges emanating from each vertex, and $k = 20, 25, 30, 35, 40$.
  - *Watts-Strogatz random graphs* (Watts and Strogatz (1998)): Notation WS-150-$|E|$ is used, where rewiring probability is 0.1 and $|E| = 750, 1500, 2250, 3000, 3750$.

Setup of numerical experiments

- Several real-life graphs obtained from the DIMACS library and Network Repository (Rossi and Ahmed (2015)).
- Randomly generated graphs with $|V| = 150$:
  - *Uniform random graphs* (Erdös and Rényi (1959)): Notation ER-150-$|E|$ is used, where $|E| = 1000, 1500, 2000, 2500, 3000$.
  - *Power-law random graphs* (Albert and Brabasi (2002)): Notation BA-150-k is used, where k represents the number of edges emanating from each vertex, and $k = 20, 25, 30, 35, 40$.
  - *Watts-Strogatz random graphs* (Watts and Strogatz (1998)): Notation WS-150-$|E|$ is used, where rewiring probability is 0.1 and $|E| = 750, 1500, 2250, 3000, 3750$.
- In cases when the underlying graph is disconnected, the largest connected component is considered.

UMASS BOSTON

Setup of numerical experiments

- Several real-life graphs obtained from the DIMACS library and Network Repository (Rossi and Ahmed (2015)).
- Randomly generated graphs with $|V| = 150$:
  - *Uniform random graphs* (Erdös and Rényi (1959)): Notation ER-150-$|E|$ is used, where $|E| = 1000, 1500, 2000, 2500, 3000$.
  - *Power-law random graphs* (Albert and Brabasi (2002)): Notation BA-150-k is used, where k represents the number of edges emanating from each vertex, and $k = 20, 25, 30, 35, 40$.
  - *Watts-Strogatz random graphs* (Watts and Strogatz (1998)): Notation WS-150-$|E|$ is used, where rewiring probability is 0.1 and $|E| = 750, 1500, 2250, 3000, 3750$.
- In cases when the underlying graph is disconnected, the largest connected component is considered.
- All computations were conducted on an Intel Xenon 3.5GHz PC with 64GB RAM. A computation time limit of 7200 seconds was imposed.

UMASS BOSTON

Setup of numerical experiments

- Several real-life graphs obtained from the DIMACS library and Network Repository (Rossi and Ahmed (2015)).
- Randomly generated graphs with $|V| = 150$:
  - *Uniform random graphs* (Erdös and Rényi (1959)): Notation ER-150-$|E|$ is used, where $|E| = 1000, 1500, 2000, 2500, 3000$.
  - *Power-law random graphs* (Albert and Brabasi (2002)): Notation BA-150-k is used, where k represents the number of edges emanating from each vertex, and $k = 20, 25, 30, 35, 40$.
  - *Watts-Strogatz random graphs* (Watts and Strogatz (1998)): Notation WS-150-$|E|$ is used, where rewiring probability is 0.1 and $|E| = 750, 1500, 2250, 3000, 3750$.
- In cases when the underlying graph is disconnected, the largest connected component is considered.
- All computations were conducted on an Intel Xenon 3.5GHz PC with 64GB RAM. A computation time limit of 7200 seconds was imposed.
- The CBB algorithm was coded in C++ and formulation (5) was solved with the CPLEX 12.6 Mixed Integer Programming solver.

UMASS BOSTON

Numerical results

Graph Name	$\|V\|$	$\|E\|$	CPLEX			CBB		
			Time (s)	Objective	BnB Nodes	Time (s)	Objective	BnB Nodes
494-bus	494	586	—	$-\infty$	0	0.23	45.06	528
662-bus	662	906	—	$-\infty$	0	0.28	89.31	673
adjnoun.clq	112	425	22.36	3.23	0	0.03	3.23	194
bcspwr05	443	590	—	10.00	0	0.11	45.02	457
bio-yeast	1458	1948	—	$-\infty$	0	3.88	326.24	1498
brock200-2	200	9876	—	$-\infty$	0	61.68	2.81	58658
brock200-4	200	13089	—	$-\infty$	0	2458.20	3.73	1710496
c500.9	500	112332	—	$-\infty$	0	—	23.84	138345
c.-m.[1]	453	2025	—	85.31	0	3.09	85.31	1027
c.-n.[2]	297	2148	708.82	36.36	0	3.78	36.36	1517
chesapeake	39	170	2.40	0.39	0	0.00	0.39	108
dolphins	62	159	5.18	0.88	0	0.02	0.88	86
e.-f.-b.[3]	128	2137	379.97	2.60	7	0.66	2.60	1421
email	1133	5451	—	$-\infty$	0	15.26	68.79	1708
Erdos971	472	1314	—	$-\infty$	0	6.04	0.86	2118
football	115	613	90.62	1.16	67	0.28	1.16	829
g.-p.-4.[4]	200	17910	—	$-\infty$	0	—	6.96	1459152
inf-euroroad	1174	1417	—	$-\infty$	0	2.10	121.03	1273
jazz	198	2742	—	8.54	21	123.85	8.77	37691
karate	34	78	2.11	0.39	3	0.00	0.39	94
keller4	171	9435	—	$-\infty$	0	1101.60	2.11	1078699
lesmis	77	254	16.29	2.13	5	0.09	2.13	531
netscience	1589	2742	—	$-\infty$	0	17.38	2.21	4331
p-hat300-1	300	10933	—	$-\infty$	0	30.27	4.08	15913
p-hat300-2	300	21928	—	$-\infty$	0	—	10.74	1354836
p-hat1500-1	1500	284923	NA	$-\infty$	0	6817.70	1.62	224139
p-hat1500-2	1500	568960	NA	$-\infty$	0	3172.14	1.54	192446
polblogs	1490	16715	NA	$-\infty$	0	459.64	1.80	18205
polbooks	105	441	83.43	2.21	11	0.08	2.21	307
USAir	332	2126	821.14	49.05	0	96.57	49.05	13064

[1]celegans-metabolic, [2]celegans-neural, [3]eco-foodweb-baydry, [4]gen200-p0.9-44.b

UMASS BOSTON

Numerical results

Graph Name	$\|V\|$	$\|E\|$	CPLEX			CBB		
			Time (s)	Objective	BnB Nodes	Time (s)	Objective	BnB Nodes
ER-150-1000	150	1000	917.55	1.06	127	0.19	1.06	458
ER-150-1500	150	1500	1266.27	1.08	157	0.20	1.08	475
ER-150-2000	150	2000	1465.26	1.10	361	0.66	1.10	1310
ER-150-2500	150	2500	4217.55	1.20	594	0.98	1.20	1783
ER-150-3000	150	3000	—	1.25	578	1.65	1.27	2809
BA-150-20	150	2600	913.32	3.82	0	1.00	3.82	1210
BA-150-25	150	3125	2086.38	4.25	0	1.33	4.25	1711
BA-150-30	150	3600	6511.62	4.25	0	5.73	4.25	5148
BA-150-35	150	4025	—	$-\infty$	0	2.62	4.91	2458
BA-150-40	150	4400	—	$-\infty$	0	10.98	5.08	8510
WS-150-750	150	750	320.36	1.58	105	0.34	1.58	772
WS-150-1500	150	1500	2054.80	2.18	81	1.62	2.18	2251
WS-150-2250	150	2250	—	2.37	1	10.53	2.41	9353
WS-150-3000	150	3000	—	2.89	3	13.32	2.89	10508
WS-150-3750	150	3750	—	2.94	0	53.99	2.97	38865

UMASS BOSTON

Numerical results

Graph Name	$\|V\|$	$\|E\|$	BnB Nodes	Bound fathoms (%)	Feasibility fathoms (%)
494-bus	494	586	528	54.2	44.7
662-bus	662	906	673	61.8	37.6
adjnoun.clq	112	425	194	53.6	43.8
bcspwr05	443	590	457	55.6	43.8
bio-yeast	1458	1948	1498	36.5	63.4
brock200-2	200	9876	58658	90.2	7.0
brock200-4	200	13089	1710496	92.6	4.9
c500.9	500	112332	138345	90.3	5.1
c.-m.[1]	453	2025	1027	60.5	37.7
c.-n.[2]	297	2148	1517	71.1	27.6
chesapeake	39	170	108	38.9	50.9
dolphins	62	159	86	50.0	44.2
e.-f.-b.[3]	128	2137	1421	58.7	37.9
email	1133	5451	1708	56.8	42.2
Erdos971	472	1314	2118	7.1	78.1
football	115	613	829	50.5	28.5
g.-p.-4.[4]	200	17910	1459152	93.0	4.3
inf-euroroad	1174	1417	1273	61.1	36.9
jazz	198	2742	37691	85.4	7.7
karate	34	78	94	30.9	59.6
keller4	171	9435	1078699	76.9	19.7
lesmis	77	254	531	55.7	34.3
netscience	1589	2742	4331	32.6	46.5
p-hat300-1	300	10933	15913	78.7	18.7
p-hat300-2	300	21928	1354836	91.7	6.2
p-hat1500-1	1500	284923	224139	88.7	10.7
p-hat1500-2	1500	568960	192446	92.6	7.0
polblogs	1490	16715	18205	51.9	44.7
polbooks	105	441	307	51.8	39.4
USAir	332	2126	13064	80.3	13.2

[1]celegans-metabolic, [2]celegans-neural, [3]eco-foodweb-baydry, [4]gen200-p0.9-44.b

UMASS BOSTON

Numerical results

Graph Name	$\|V\|$	$\|E\|$	BnB Nodes	Bound fathoms (%)	Feasibility fathoms (%)
ER-150-1000	150	1000	458	47.6	45.4
ER-150-1500	150	1500	475	64.6	29.7
ER-150-2000	150	2000	1310	65.0	29.1
ER-150-2500	150	2500	1783	70.9	24.2
ER-150-3000	150	3000	2809	76.7	18.8
BA-150-20	150	2600	1210	77.0	19.9
BA-150-25	150	3125	1711	71.5	26.0
BA-150-30	150	3600	5148	66.4	30.3
BA-150-35	150	4025	2458	67.9	29.8
BA-150-40	150	4400	8510	55.3	41.7
WS-150-750	150	750	772	56.0	28.2
WS-150-1500	150	1500	2251	72.7	15.8
WS-150-2250	150	2250	9353	74.7	13.7
WS-150-3000	150	3000	10508	79.8	11.5
WS-150-3750	150	3750	38865	83.5	9.2

UMASS BOSTON

Directions for future research

- Studying the problem of detecting the most central clusters in the context of other clique relaxation models and centrality measures.

UMASS BOSTON

Directions for future research

- Studying the problem of detecting the most central clusters in the context of other clique relaxation models and centrality measures.
- Investigating the problem of detecting the most central clusters under uncertainty, where graph components can fail with some probability.

Directions for future research

- Studying the problem of detecting the most central clusters in the context of other clique relaxation models and centrality measures.
- Investigating the problem of detecting the most central clusters under uncertainty, where graph components can fail with some probability.
- Developing pre-processing and decomposition schemes to solve the MBCP on large-scale real-life networks.

UMASS BOSTON

Thank you

THANK YOU
ANY QUESTIONS?

- M. Rysz, F. Mahdavi Pajouh and E. L. Pasiliao. *Finding clique clusters with the highest betweenness centrality*. European Journal of Operational Research, 271(1):155-164, 2018.

UMASS BOSTON