Enumerating Isolated Cliques in Synthetic and Financial Networks

Falk Hüffner Christian Komusiewicz Hannes Moser
Rolf Niedermeier

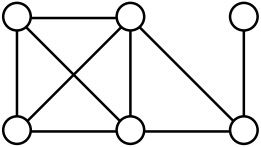
Friedrich-Schiller-Universität Jena
Institut für Informatik

August 22, 2008

Clique Enumeration

Input: An undirected graph $G = (V, E)$.

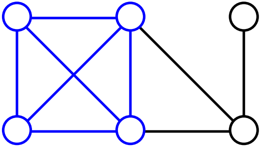
Task: Enumerate all maximal cliques, that is, all vertex subsets $C \subseteq V$ such that $G[C]$ is complete and there is no $C' \supset C$ such that $G[C']$ is complete.



Clique Enumeration

Input: An undirected graph $G = (V, E)$.

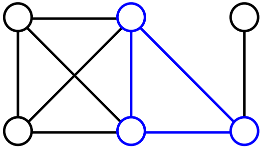
Task: Enumerate all maximal cliques, that is, all vertex subsets $C \subseteq V$ such that $G[C]$ is complete and there is no $C' \supset C$ such that $G[C']$ is complete.



Clique Enumeration

Input: An undirected graph $G = (V, E)$.

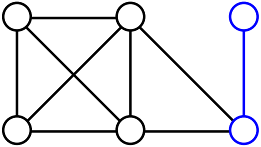
Task: Enumerate all maximal cliques, that is, all vertex subsets $C \subseteq V$ such that $G[C]$ is complete and there is no $C' \supset C$ such that $G[C']$ is complete.



Clique Enumeration

Input: An undirected graph $G = (V, E)$.

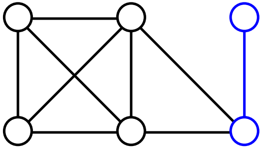
Task: Enumerate all maximal cliques, that is, all vertex subsets $C \subseteq V$ such that $G[C]$ is complete and there is no $C' \supset C$ such that $G[C']$ is complete.



Clique Enumeration

Input: An undirected graph $G = (V, E)$.

Task: Enumerate all maximal cliques, that is, all vertex subsets $C \subseteq V$ such that $G[C]$ is complete and there is no $C' \supset C$ such that $G[C']$ is complete.



Applications

- Computational finance [Boginski et al., Comput. Oper. Res., 2006]
- Biological networks [Chesler et al., Nature Genetics, 2005]
- Social networks, clustering in data mining [Makino & Uno, SWAT 2004]

Clique Enumeration

Maximal Clique Enumeration

- Simple model
- NP-hard [Garey & Johnson, 1979]
- up to $3^{n/3}$ cliques [Moon & Moser, Israel J. Math., 1965]

Clique Enumeration

Maximal Clique Enumeration

- Simple model
- NP-hard [Garey & Johnson, 1979]
- up to $3^{n/3}$ cliques [Moon & Moser, Israel J. Math., 1965]

Isolated Cliques

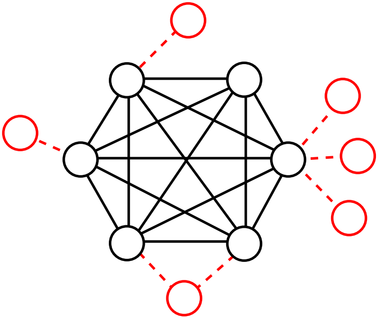
- More specific model
- More efficient enumeration algorithms

c-Isolation

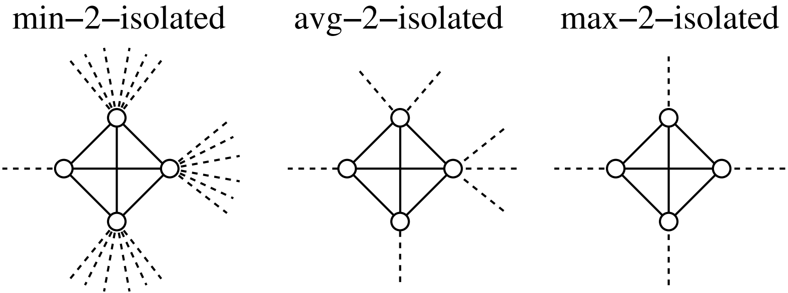
Definition [Ito, Iwama, Osumi, ESA 2005]

A vertex set S is called avg-c-isolated if on average the vertices in S have less than c neighbors outside of S.

Example: avg-2-isolation



Comparison of Isolation Concepts



Running times for enumeration of maximal c-isolated cliques

min-c-isolation	$O(2^c \cdot cm + nm)$
avg-c-isolation	$O(4^c \cdot c^4 m)$
max-c-isolation	$O(2.44^c \cdot cm)$

[Komusiewicz et al., COCOON 2007]

Maximality Test

Known avg-c-isolation algorithm:

- Algorithm enumerates $O(2^c)$ cliques
- Filter out nonmaximal cliques by pairwise comparison $\rightsquigarrow O(4^c)$

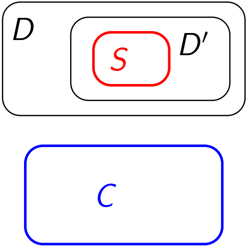
Idea

Determine maximality independently for each clique.

Maximality Test

If an avg-c-isolated clique C is not maximal, then there must be $S \subseteq V$ such that $C \cup S$ is an avg-c-isolated clique.

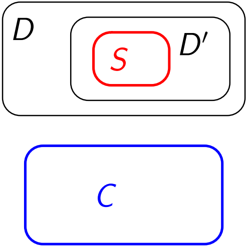
- $S \subseteq D$ with $D := \left(\bigcap_{v \in C} N(v)\right) \setminus C$



Maximality Test

If an avg-c-isolated clique C is not maximal, then there must be $S \subseteq V$ such that $C \cup S$ is an avg-c-isolated clique.

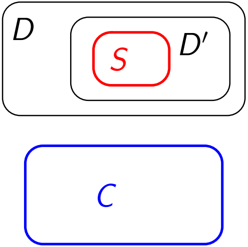
- $S \subseteq D$ with $D := \left(\bigcap_{v \in C} N(v)\right) \setminus C$
- $|D| < c$



Maximality Test

If an avg-c-isolated clique C is not maximal, then there must be $S \subseteq V$ such that $C \cup S$ is an avg-c-isolated clique.

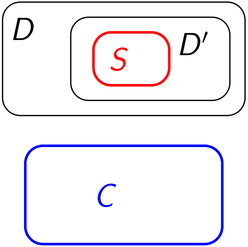
- $S \subseteq D$ with $D := \left(\bigcap_{v \in C} N(v)\right) \setminus C$
- $|D| < c$
- Enumerate maximal cliques D' in D



Maximality Test

If an avg-c-isolated clique C is not maximal, then there must be $S \subseteq V$ such that $C \cup S$ is an avg-c-isolated clique.

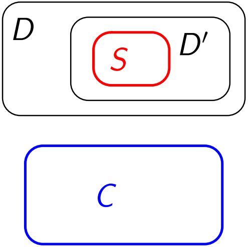
- $S \subseteq D$ with $D := \left(\bigcap_{v \in C} N(v)\right) \setminus C$
- $|D| < c$
- Enumerate maximal cliques D' in D
- Remaining task: is there a clique $S \subseteq D'$ such that $C \cup S$ is an avg-c-isolated clique?



Maximality Test

If an avg-c-isolated clique C is not maximal, then there must be $S \subseteq V$ such that $C \cup S$ is an avg-c-isolated clique.

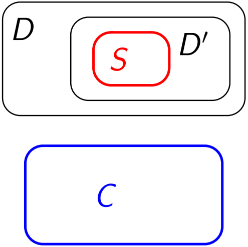
- $S \subseteq D$ with $D := \left(\bigcap_{v \in C} N(v)\right) \setminus C$
- $|D| < c$
- Enumerate maximal cliques D' in D
- Remaining task: is there a clique $S \subseteq D'$ such that $C \cup S$ is an avg-c-isolated clique?
- Remove vertices from D' in order of highest degree



Maximality Test

If an avg-c-isolated clique C is not maximal, then there must be $S \subseteq V$ such that $C \cup S$ is an avg-c-isolated clique.

- $S \subseteq D$ with $D := \left(\bigcap_{v \in C} N(v)\right) \setminus C$
- $|D| < c$
- Enumerate maximal cliques D' in D
- Remaining task: is there a clique $S \subseteq D'$ such that $C \cup S$ is an avg-c-isolated clique?
- Remove vertices from D' in order of highest degree



Theorem

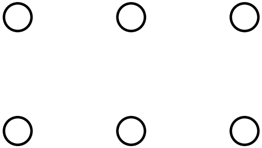
All maximal c-isolated cliques can be enumerated in $O(2.89^c \cdot c^2 \cdot m)$ *time.*

Synthetic Data: $G_{n,m,p}$ Model

$G_{n,m,p}$ Model [Behrisch & Taraz, Theoret. Comput. Sci., 2006]

Each of n vertices draws each of m *features* with probability p, and two vertices are connected by an edge iff they have at least one feature in common.

$m = 3$
$n = 6$

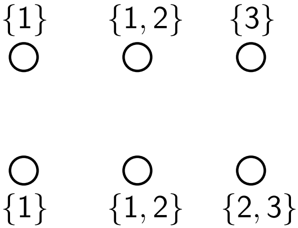


Synthetic Data: $G_{n,m,p}$ Model

$G_{n,m,p}$ Model [Behrisch & Taraz, Theoret. Comput. Sci., 2006]

Each of n vertices draws each of m *features* with probability p, and two vertices are connected by an edge iff they have at least one feature in common.

$m = 3$
$n = 6$

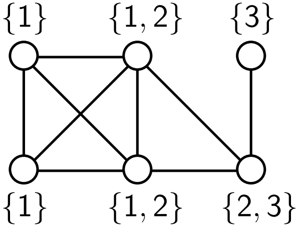


Synthetic Data: $G_{n,m,p}$ Model

$G_{n,m,p}$ Model [Behrisch & Taraz, Theoret. Comput. Sci., 2006]

Each of n vertices draws each of m *features* with probability p, and two vertices are connected by an edge iff they have at least one feature in common.

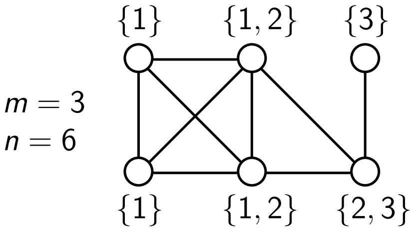
$m = 3$
$n = 6$



Synthetic Data: $G_{n,m,p}$ Model

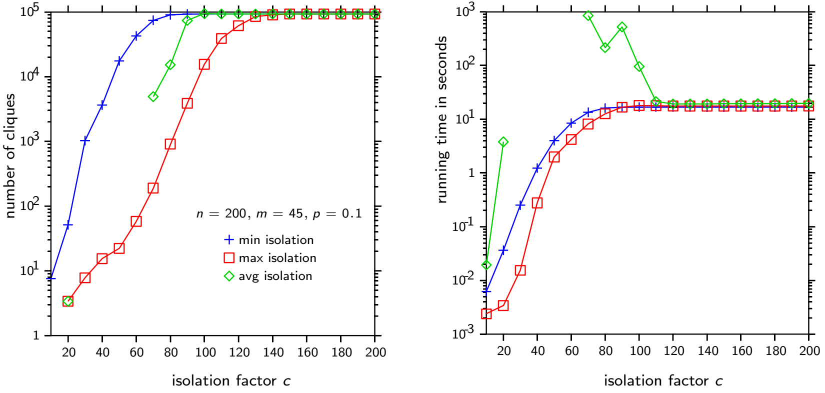
$G_{n,m,p}$ Model [Behrisch & Taraz, Theoret. Comput. Sci., 2006]

Each of n vertices draws each of m *features* with probability p, and two vertices are connected by an edge iff they have at least one feature in common.



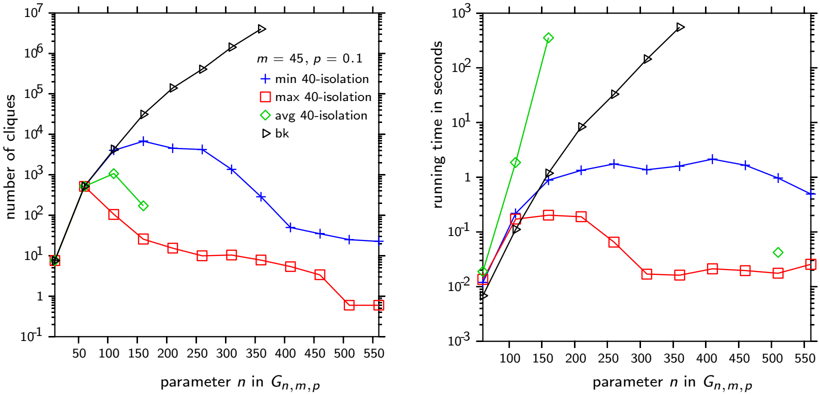
- Each feature induces a clique.
- Every nonempty intersection of feature cliques is a clique $\rightsquigarrow$ we obtain many cliques.

$G_{n,m,p}$ Model



number of cliques
isolation factor c
$n = 200$, $m = 45$, $p = 0.1$
min isolation
max isolation
avg isolation
running time in seconds
isolation factor c

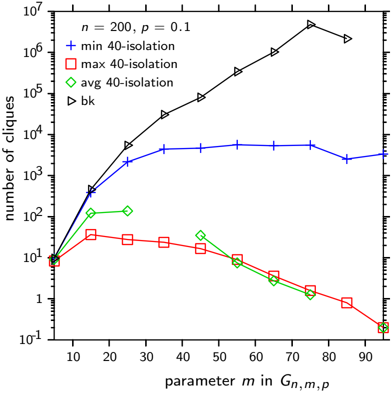
$G_{n,m,p}$ Model

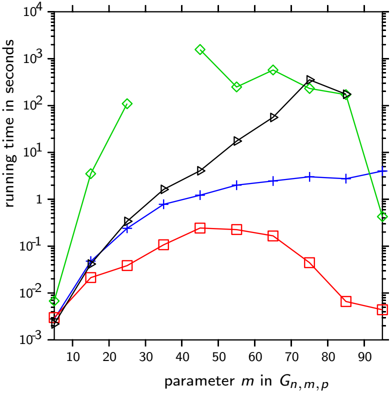


"bk" is an improved variant of the standard Bron–Kerbosch algorithm, which enumerates all maximal cliques.

[Koch, Theoret. Comput. Sci., 2001]

$G_{n,m,p}$ Model

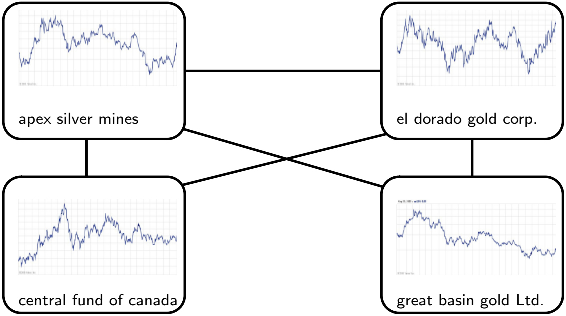

$n = 200$, $p = 0.1$
min 40-isolation
max 40-isolation
avg 40-isolation
bk
number of cliques
parameter m in $G_{n,m,p}$


running time in seconds
parameter m in $G_{n,m,p}$

Stock Data

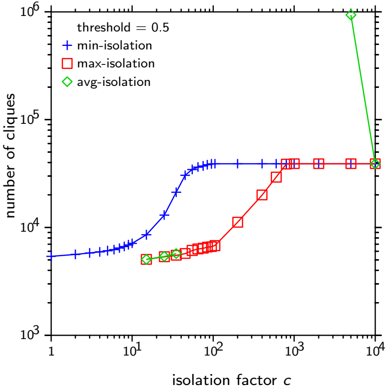
Market Graph [Mantegna & Stanley, 2000]

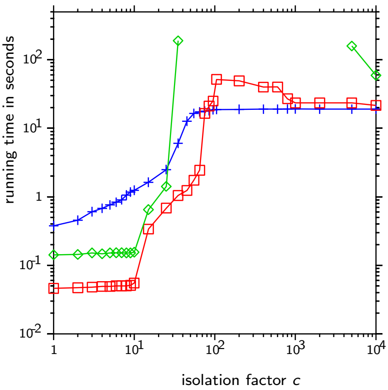
- Stocks are vertices.
- Two stocks are connected iff the correlation of the daily fluctuations of their prices exceeds some threshold.



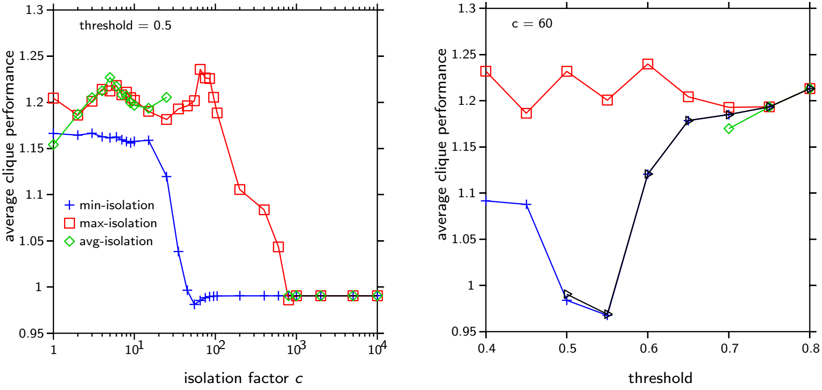
(chart data from 2008 Yahoo! Inc.)

Stock Data


threshold = 0.5
min-isolation
max-isolation
avg-isolation
number of cliques
isolation factor c
10^3
10^4
10^5
10^6
1
10^1
10^2
10^3
10^4


running time in seconds
isolation factor c
10^{-2}
10^{-1}
1
10^1
10^2
1
10^1
10^2
10^3
10^4

Stock Data


threshold = 0.5
average clique performance
isolation factor c
min-isolation
max-isolation
avg-isolation
c = 60
average clique performance
threshold

Summary

- Enumerating min- and max-isolated cliques is feasible over a very large range of instances and parameters.
- Sometimes even beats Bron–Kerbosch for enumerating all maximal cliques.
- Avg-isolation more limited.
- Isolation leads to “interesting” cliques, like, e.g., sets of stocks with unusual performance.

Other Applications

- Finding complexes in protein interaction networks.
- Finding communities in web graphs.
- Finding genres in music artist similarity networks.