特定領域研究「新世代の計算限界 ―その解明と打破―」
全体会議 2005.6.16–17

孤立したクリークの線形時間列挙

伊藤大雄
京都大学 大学院 情報学研究科

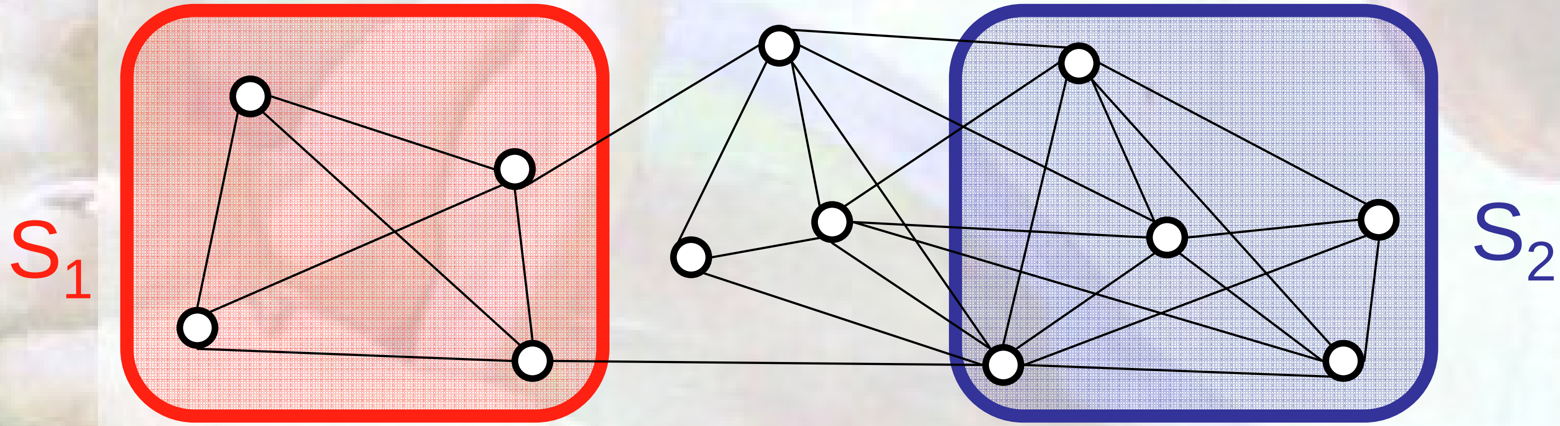
（共同研究者：岩間一雄・大隅剛史）

クリークと列挙

- 最大クリーク問題
  - 近似度Ω$(n^{1-\varepsilon})$ [Hastad 99]
- 極大クリークの列挙
  - O(nm)時間／個[Tsukiyama, et al. 77]
  - O(Δ^4)時間／個[Uno03]（Δは最大次数)
  - しかし指数($O(3^{n/3})$)個存在[Moon, et al. 65]

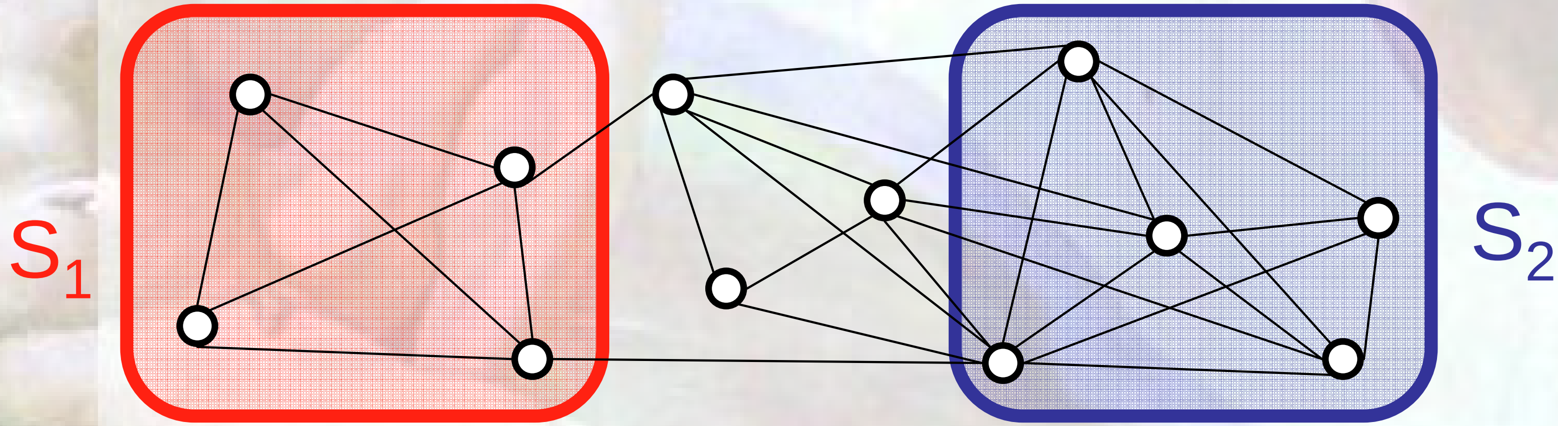
Why Cliques?

- Inside: Densely connected.
- Inside-Outside: Sparsely connected.



Isolated cliques

- Let $c>0$ be a constant. A clique $S \subseteq V$ with k vertices is an c-isolated clique if $|E(S)|<ck$. ($E(S)$={edges between S and V-S}.)
- 1-isolated cliques = isolated cliques.



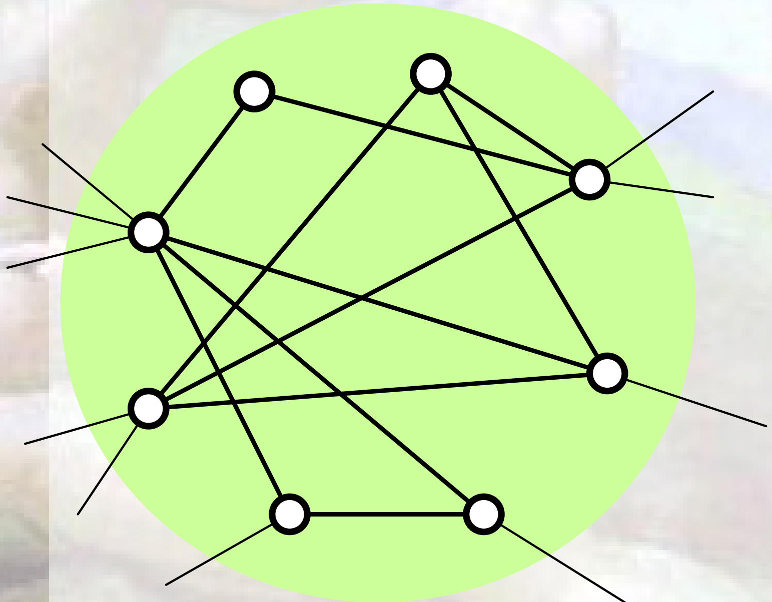
$|E(S_1)=2<4 \Rightarrow$ isolated clique

$|E(S_2)=9 \geqq 5 \Rightarrow$ non isolated clique

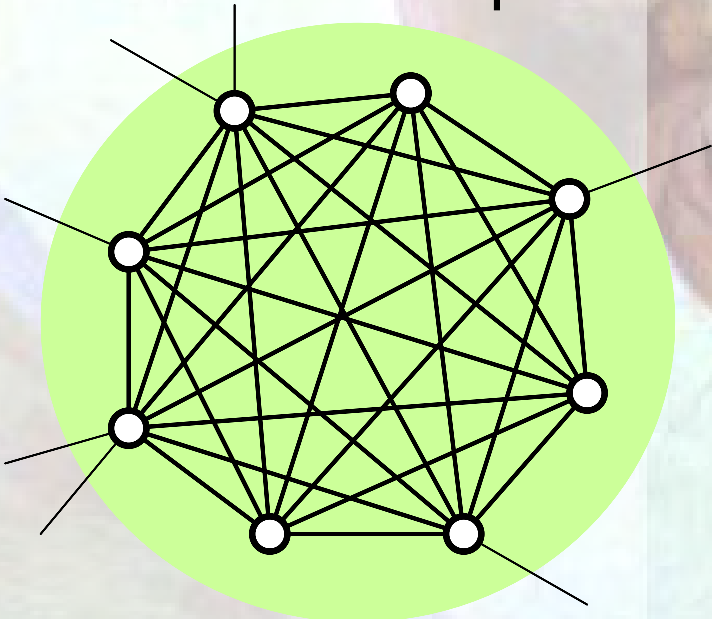
Related work

- Flake, Lawrence, Giles (2000)
  - Community $S \subset V$: $|E(v,S)| > |E(v,V-S)| \ \forall \ v \in S$.

Community of Flake, et al.



Isolated clique



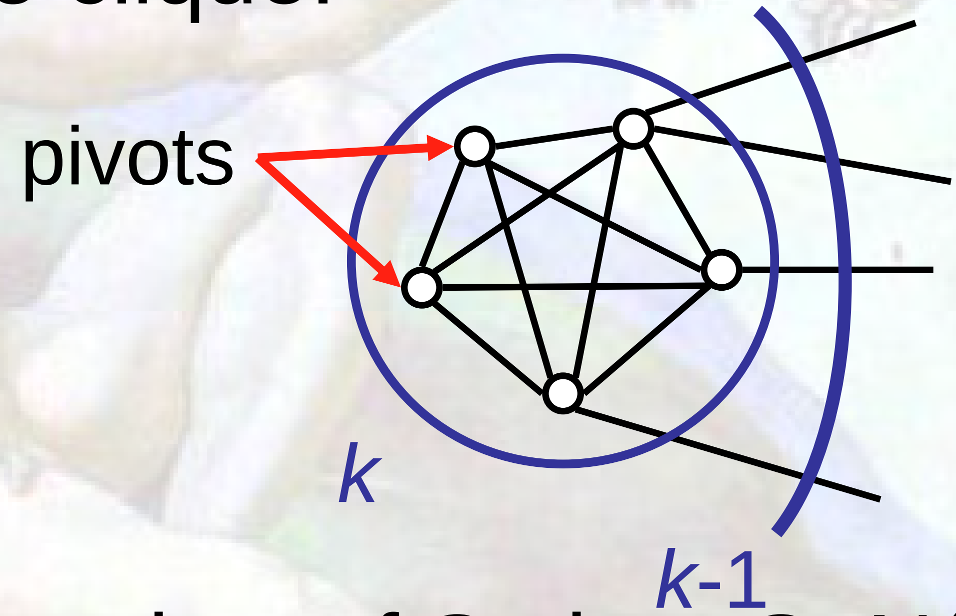
Preliminary Result

- Theorem 0.1. All isolated cliques can be enumerated in linear time.

- Corollary 1. The # of isolated cliques is O(*m*) for any graph.

Observation for Proving Theorem 0.1

- Lemma 1. An isolated clique has a vertex (called a pivot) that has no outgoing edge from the clique.



- If v is the pivot of S, then $S = N(v) \cup \{v\}$.

Strategy of enumeration

- Check each vertex whether or not it can be a pivot.
- Sort and renumber all vertices as $d(v_1) \leqq d(v_2) \leqq \ldots \leqq d(v_n)$.
- If $\exists$ $j>i$, $v_j \in N(v_i)$ (adjacent vertices of v_i), v_i can be ignored (we can consider that the vertex having the minimum index in a isolated clique is the pivot): Test (a).

Observation 2

- Lemma 2. If S is an isolated clique and v is the pivot, then $\sum_{w \in S} d(w) < (d(v)+1)^2$. (1)

- Proof. Let $k=|S|=d(v)+1$. $\sum_{w \in S} d(w) < k(k-1)+k = k^2 = (d(v)+1)^2$. □

Observation 2

- Lemma 3. If v has the minimum indices in $S=N(v)\cup\{v\}$ and S satisfies (1), then $d(w)<2d(v)+1 \quad \forall w\in S$.

- Proof. Let $k=|S|=d(v)+1$. If $d(w)\geqq 2d(v)+1=2k-1$ for a $w\in S$, $\sum_{w\in S} d(w)\geqq (k-1)^2+2k-1=k^2 =(d(v)+1)^2$, contradiction. □

Strategy 2

- If v passes Test (a), we check whether $S=N(v) \cup \{v\}$ satisfies (1): Test (b). (This can be done in $O(d(v))$ time.)
- If N(v) passes this test, $d(w)<2d(v)+1=O(d(v))$ $\forall w \in N(v)$ from Lemma 3.

Observation 3

- Lemma 4. If $S=\{v=w_1, ..., w_k\}$ is an isolated clique and v is the pivot of S ($d(w_1) \leqq ... \leqq d(w_k)$), then $S_i=\{w_1, ..., w_i\}$ has at most $i-1$ outgoing edges from S.
- Proof. Assume that $|E(S_i, V-S)| \geqq i$. Then $d(w_i) \geqq d(v)+1$, and hence $d(vj) \geqq d(v)+1$ for all $j=i+1, ..., k$. Therefore $\Sigma_{w \in S} d(w) = \Sigma_{w \in Si} d(w) + \Sigma_{v \in S-Si} d(w) \geqq i+(k-i)=k$, contradiction.

Strategy 3

- Assume that v passed Tests (a) and (b) (primal tests).
- Let $S=N(v)\cup\{v\}=\{v=w_1, w_2, ..., w_k\}$ $(d(w_1)\leqq d(w_2)\leqq \dots \leqq d(w_k))$.
- Clique test: From i=1 to k,
  - check whether (1) w_i is adjacent to $w_1, ..., w_{i-1}$ (i.e., $S_i=\{w_1, ..., w_i\}$ is a clique) and
  - (2) S_i has at most i-1 outgoing edges from S.
  - If not, *v* is not a pivot and then skipped (finish checking v).

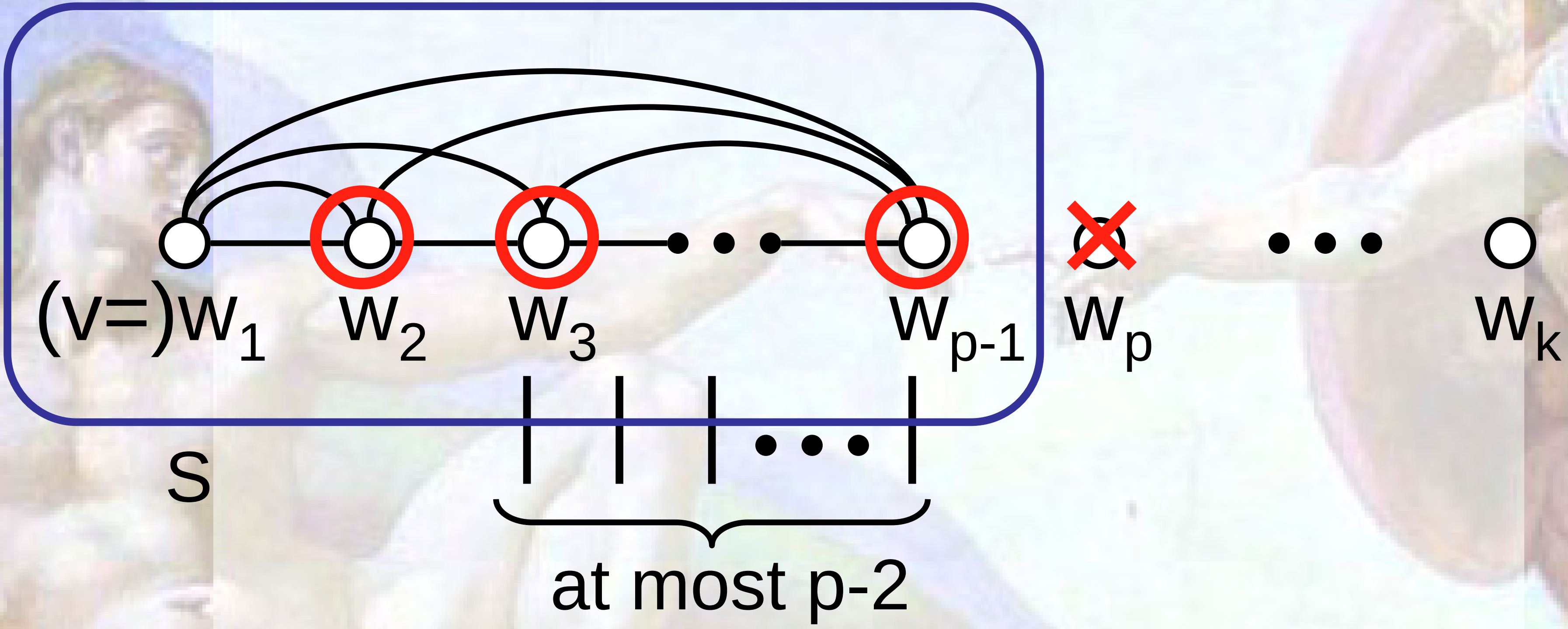
Running time

- Sorting vertices by their degrees: O(m).
- Primal tests: O(d(v)) for each $v \in V$, i.e., O(m).
- Clique test: Assume the test breaks at w_p.
  $d(w_1)+d(w_2)+\ldots+d(w_p)=O(w_p^2) \rightarrow O(m^2)$?

$\rightarrow$ More precise estimation!

Running time (Cont.)

- Assume the test is done until w_p. ($k=d(v_1)+1$)



By v: $O(d(w_1)+d(w_2)+\dots d(w_p))=O(pk)$

By other pivots: $O((p-1)d(w_{p-1}))=O(pk)$

} $O(pk)$

→ Amortize as $O(k)=d(w_i)$ for each vertex in S.

Running time (Cont.)

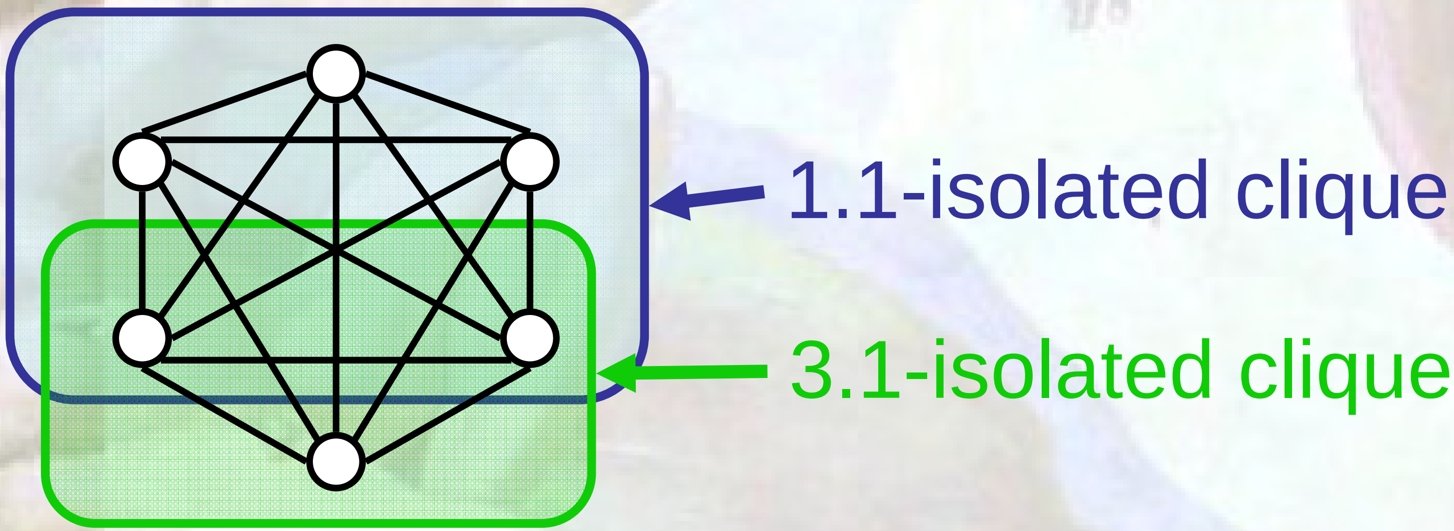
- Sorting vertices by their degrees: O(m).
- Primal tests: O(d(v)) for each $v \in V$, i.e., O(m).
- Clique test: O(d(v)) for each $v \in V$, i.e., O(m).

→ O(m). ∙∙∙ Theorem 0.1 is proved.

Extension

- Theorem 0.1. All isolated cliques can be enumerated in linear time.

- For general c-isolated cliques?



- Maximal ones are important.

Results

- Theorem 1. All maximal *c*-isolated cliques of a graph with *n* vertices can be enumerated in $O(c^5 2^{2c} m)$ time.

- Corollary 1. For any constant *c*, all maximal *c*-isolated cliques can be enumerated in linear time.
- Corollary 2. For any $c=O(\log n)$, all maximal *c*-isolated cliques of a graph with n vertices can be enumerated in polynomial time.

Results (cont.)

- Theorem 2. Let *c*, *x*, and *y* are functions of *n* s.t. $c=xy$. There is a graph with *m* edges for which the # of maxmal *c*-isolated cliques is $\Omega((2x^y/c^2)m)$.

⇩

- Cor. 3. If $c=\omega(1)$, there is a graph with *n* vertices consisting of super-linear # of maximal *c*-isolated cliques.
- Cor. 4. If $c=\omega(\log n)$, there is a graph with *n* vertices consisting of super-polynomial # of maximal *c*-isolated cliques.

Results (cont.)

- Cor. 1. If $c=O(1)$, all maximal c-isolated cliques can be enumerated in linear time.
- Cor. 3. If $c=\omega(1)$, there is a graph consisting of super-linear # of maximal c-isolated cliques.

- $c=\Theta(1)$ is the tight bound for enumerating all maximal c-isolated cliques in linear time.

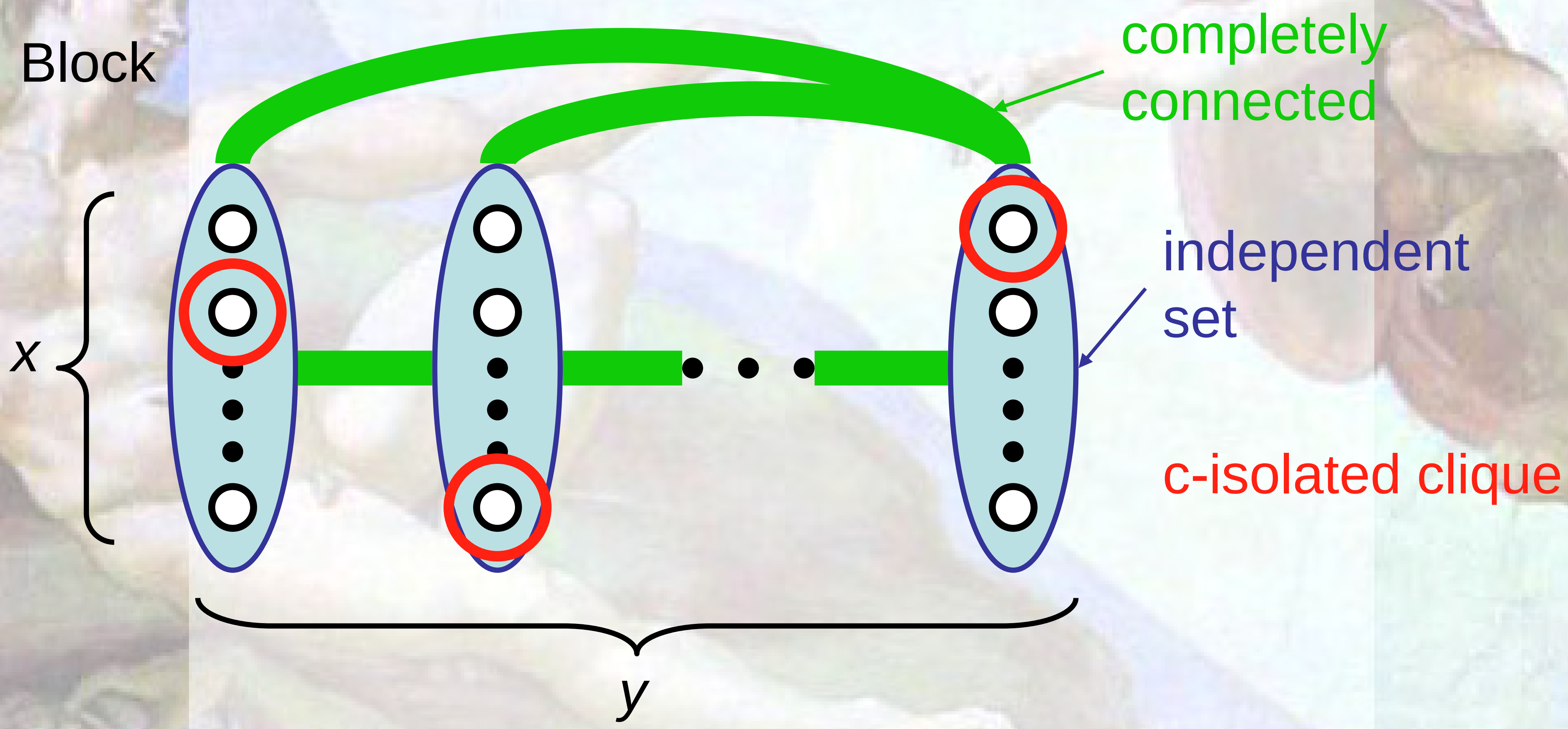
Results (cont.)

- Cor. 2. If $c=O(\log n)$, all maximal c-isolated cliques can be enumerated in polynomial time.
- Cor. 4. If $c=\omega(\log n)$, there is a graph consisting of super-polynomial # of maximal c-isolated cliques.

- $c=\Theta(\log n)$ is the tight bound for enumerating all maximal c-isolated cliques in polynomial time.

Proof of Theorem 2

- Theorem 2. Let *c*, *x*, and *y* are functions s.t. $c=xy$. There is a graph with *m* edges for which the # of maxmal *c*-isolated cliques is $\Omega((2x^y/c^2)m)$.



Proof of Corollaries 3 and 4

- If $c=\omega(1)$, then by letting x=2, y=c/2 $(2x^y/c^2)m$ becomes super-linear.
- If $c=\omega(\log n)$, then by letting $x=c/\log n$, $y=\log n$ $(2x^y/c^2)m$ is super-polynomial.

Other Results: Pseudo-Clieques

- Let $\alpha(k)$ and $\beta(k)$ are functions. Pseudo-Clique $PC(\alpha, \beta)$ is a vertex-proper-subset $S \subset V$ ($|S|=k$) s.t.
- $\mathrm{av}_{v \in S}\, d_{G(S)}(v) \geqq \alpha(k)$ and
- $\min_{v \in S}\, d_{G(S)}(v) \geqq \beta(k)$.

Results for PC

- Theorem 3. Suppose $f(k)=\Omega(1)$ and $0<\varepsilon<1$ is a constant.
  - There is a graph including super-poly. # of maximal isolated PC($k-f(k),k^{\varepsilon}$).
  - There is a graph including super-poly. # of maximal isolated PC($k-k^{\varepsilon},k/f(k)$).
- Proposition 1. All maximal isolated PC(α,c_1k) and PC($k-c_2,k^{\varepsilon}$) are enumerated in poly. time for constant $c_1<1$ and $c_2\geqq 1$.

Results for PC (Cont.)

- Theorem 4. There is a graph including super-poly. # of maximal isolated PC(k-$(\log k)^{1+\varepsilon}$,k/$(\log k)^{1+\varepsilon}$) for any $0<\varepsilon$.

- Theorem5. All maximal isolated PC(k-logk,k/logk) can be enumerated in poly. time.

Summary

- Introduce f-isolated cliques with parameter function c.
- All c-isolated cliques can be enumerated in linear time for any constant c.
- $c=\Theta(1)$ is the tight bound of linear time enumeration.
- All c-isolated cliques can be enumerated in poly. time if $c=O(\log n)$.
- $c=\Theta(\log n)$ is the tight bound of poly. time enumeration.