Detecting Cohesive Subgraphs

Benjamin McClosky and Illya V. Hicks

Department of Computational and Applied Mathematics
Rice University, Houston, TX.

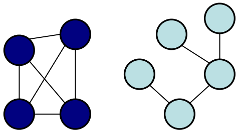
March 25, 2008

Outline

Graphs

$G = (V, E)$

- vertex set V is finite
- edges $E \subseteq \{uv \ : \ u, v \in V\}$
- undirected



Example 1: Modularity in gene co-expression networks

- vertices represent genes
- $uv \in E$ if expression of gene u has high correlation with expression of gene v

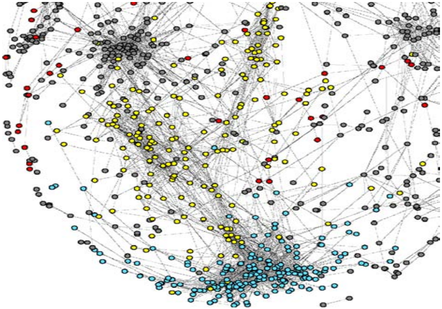
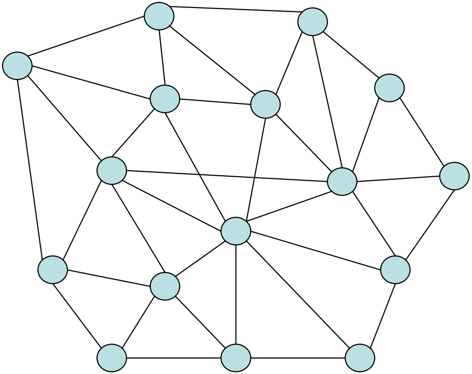


Figure: Carlson, Zhang, Fang, Mischel, Howrvath, and Nelson. *Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks*, BMC Genomics 2006, 7:40.

Modularity in gene co-expression networks



Cohesive subgraphs: Completeness and cliques

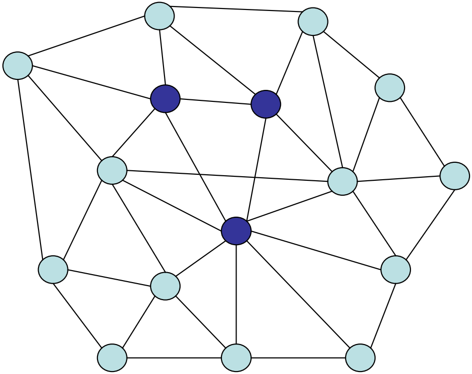


Figure: $\omega(G)$:= max cardinality of a clique

All vertex pairs are adjacent (restrictive).

Cohesive subgraphs: k-plexes

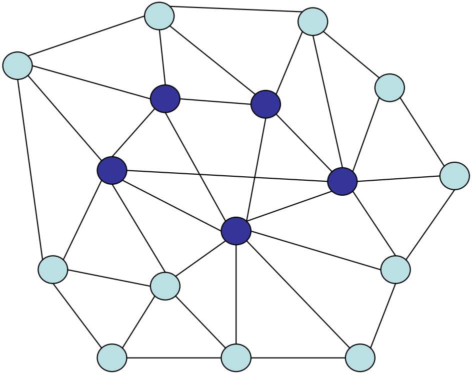


Figure: $\omega_k(G) :=$ max cardinality of a k-plex

User-defined level of mutual adjacency (a relaxation).

A general notion of graph cohesion

Definition (Seidman and Foster 1978)

Fix an integer $k \geq 1$. $K \subseteq V$ is a *k-plex* if

$$deg_{G[K]}(v) \geq |K| - k \quad \text{for all} \quad v \in K.$$

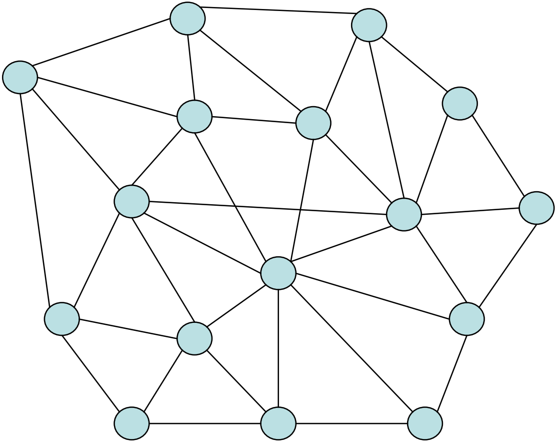
- 1-plexes are complete graphs
- k-plexes relax the structure of complete graphs

Example 2: Social Networks

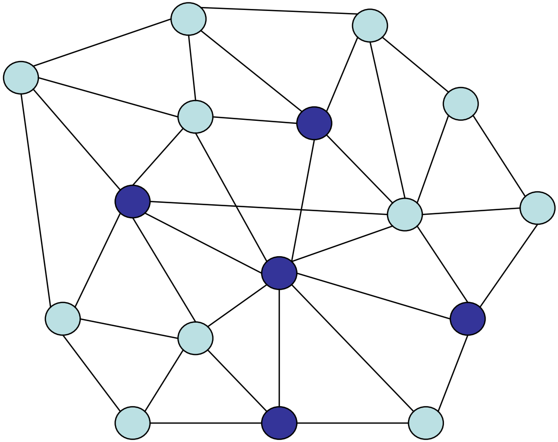
- vertices are people
- edges represent specific types of relations or interdependencies
    - values
    - financial exchange
    - friendship or kinship
    - conflict
    - disease transmission

Moody, James, and Douglas R. White (2003). "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups." American Sociological Review 68(1):103-127.

Social Networks



Other relaxations: k-club



k-plexes

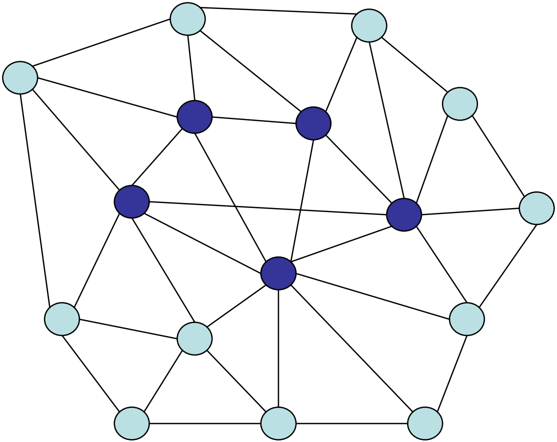


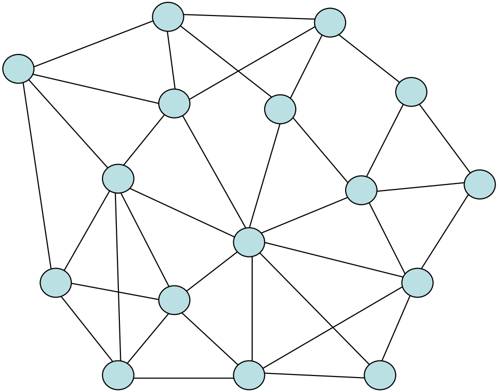
Figure: $\omega_2(G)$:= max cardinality of a 2-plex

Example 3: Retail location

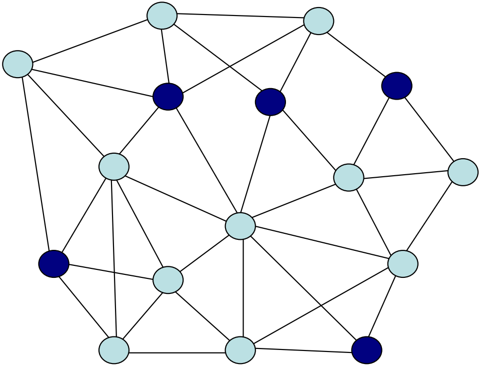
A successful company plans to open many new outlets.

- vertices represent potential locations
- research indicates that stores closer than x miles will compete for customers (market cannibalism)
- $uv \in E$ if location u is within x miles of location v

Retail location

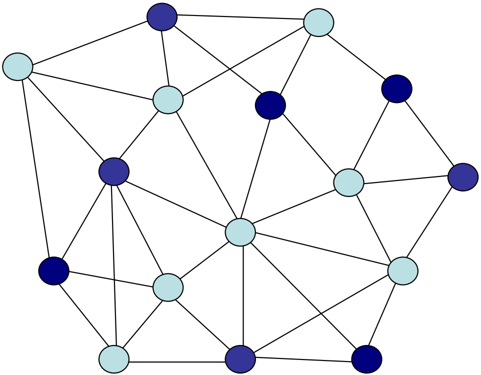


Stable sets



All vertex pairs are non-adjacent.

Co-k-plexes



User-defined level of non-adjacency (a relaxation).

Definition

Definition (Seidman and Foster 1978)

Fix an integer $k \geq 1$. $S \subseteq V$ is a *co-k-plex* if

$$deg_{G[S]}(v) \leq k - 1 \quad \text{for all} \quad v \in S.$$

- co-1-plexes are isolated vertices (stable sets)
- co-k-plexes are degree-bounded subgraphs

Cohesion and sparsity

Detecting cohesive subgraphs (k-plexes) is computationally equivalent to detecting sparse subgraphs (co-k-plexes).

Why?

Cohesive in G...

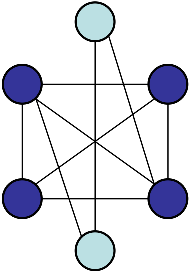


Figure: $G = (V, E)$

...is sparse in $\bar{G}$

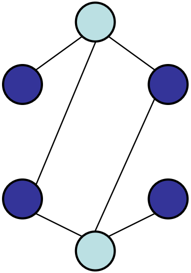


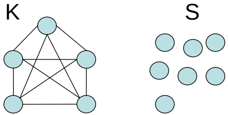
Figure: $\bar{G} = (V, \bar{E})$, where $e \in \bar{E} \Leftrightarrow e \notin E$.

Previous Work

- Seidman and Foster (1978)
  - introduced k-plexes in context of social network analysis
  - derived basic properties
- Balasundaram, Butenko, Hicks, and Sachdeva (2006)
  - established NP-completeness of Maximum k-plex
  - studied the k-plex polytope

Outline

Upper bound on the size of cohesive subgraphs



If $K \subseteq V$ is complete and $S \subseteq V$ is a stable set, then

$$|K \cap S| \leq 1.$$

Consequently, if V partitions into stable sets $S_1, \ldots, S_m$, then

$$|K| = |K \cap V| = |K \cap (\cup_{i=1}^{m} S_i)| = \sum_{i=1}^{m} |K \cap S_i| \leq \sum_{i=1}^{m} 1 = m.$$

Graph coloring and the chromatic number

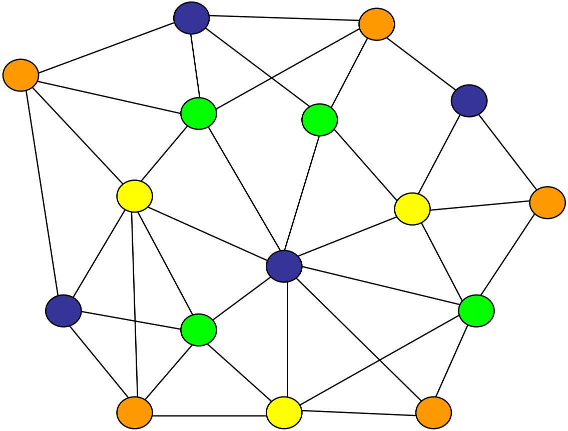


Figure: $\omega(G) \leq \chi(G)$

Analogously...

If $K \subseteq V$ is a k-plex (cohesive) and
$S \subseteq V$ is a co-k-plex (sparse), then

$$|K \cap S| \leq 2k - 2 + k \ mod \ 2.$$

Consequently, if V partitions into co-k-plexes $S_1, \ldots, S_m$, then

$$|K| = |K \cap (\cup_{i=1}^{m} S_i)| = \sum_{i=1}^{m} |K \cap S_i| \leq m(2k - 2 + k \ mod \ 2).$$

Co-k-plex coloring and the co-k-plex chromatic number

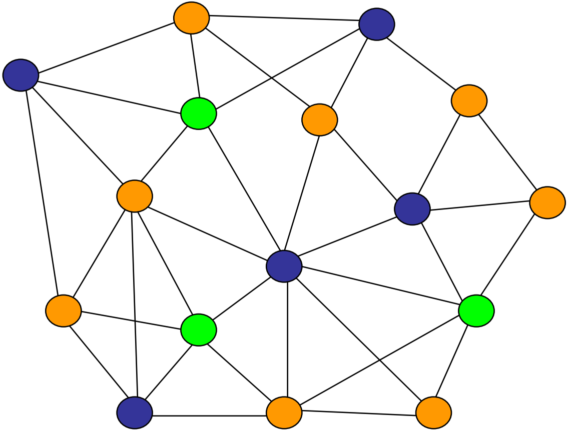


Figure: $\omega_2(G) \leq \chi_2(G)$

Computational Results (2.2 GHz Dual-Core AMD Opteron processor with 3 GB of memory)

Table: Coloring Results

G	$\chi_2(G)$	seconds	$\chi_3(G)$	seconds	$\chi_4(G)$	seconds
brock200-1	83	0.1	139	0.1	167	0.0
brock400-2	152	0.7	272	0.1	320	0.2
brock800-2	224	1.7	400	2.6	535	1.6
c-fat200-1	15	0.0	20	0.0	21	0.0
c-fat500-1	22	0.1	23	0.0	24	0.0
C125.9	84	0.0	116	0.0	122	0.0
hamming6-2	32*	0.0	59	0.0	61	0.0
hamming8-2	128*	0.1	231	0.1	251	0.1
johnson8-2-4	10	0.0	18	0.0	19	0.0
johnson16-2-4	34	0.0	76	0.0	95	0.0
johnson32-2-4	75	1.0	224	0.5	299	0.3
keller4	44	0.1	90	0.0	111	0.0
MANN-a9	37	0.0	42	0.0	45	0.0
p-hat300-1	35	0.0	63	0.0	89	0.0
p-hat700-1	68	0.4	124	0.3	169	0.5
p-hat1500-1	125	4.0	230	6.3	326	4.4
san200-0.7-2	57	0.0	113	0.0	144	0.1

* optimal

Lower bound on the size of cohesive subgraphs

If we can find a k-plex $K \subseteq V$, then

$$|K| \leq \omega_k(G).$$

For a lower bound, use local search to find feasible k-plexes.

Computational Results (2.2 GHz Dual-Core AMD Opteron processor with 3 GB of memory)

Table: Lower Bound Results

G	$\omega_2(G)$	seconds	$\omega_3(G)$	seconds	$\omega_4(G)$	seconds
brock200-1	25	1	27	1	31	1
brock400-2	27	2	31	2	35	2
brock800-2	22	15	26	15	29	15
c-fat200-1	12*	2	12*	2	12*	2
c-fat500-1	14*	20	14*	19	14*	19
C125.9	42	0	47	0	54	0
hamming6-2	32*	0	32*	0	32	0
hamming8-2	128*	0	128*	0	128	0
johnson8-2-4	4	0	8*	0	9*	0
johnson16-2-4	8	0	16	0	18	0
johnson32-2-4	16	2	32	2	36	2
keller4	15*	1	18	1	20	1
MANN-a9	22	0	30	0	36*	0
p-hat300-1	9	4	11	4	12	4
p-hat700-1	10	33	13	33	16	32
p-hat1500-1	13	202	14	204	16	204
san200-0.7-2	26	1	36	1	48	1

* optimal

Outline

Max k-plex Algorithm: Type 1*

- $V := \{v_1, ..., v_n\}$
- $S_i := \{v_i, ..., v_n\}$ for $1 \leq i \leq n$

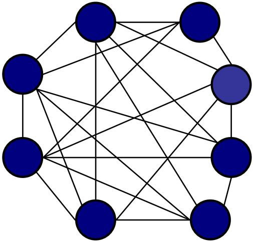
for i : 1 to n

  Search S_i for largest k-plex containing v_i.

end

*Applegate and Johnson; Carraghan and Pardalos

Max k-plex Algorithm: Type 1



Max k-plex Algorithm: Type 1

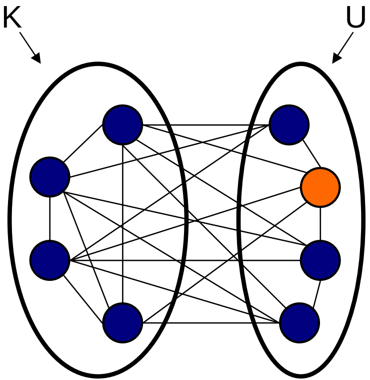


Figure: $U = \{v \in V \setminus K \ : \ K \cup \{v\}$ is a k-plex$\}$.

Max k-plex Algorithm: Type 1

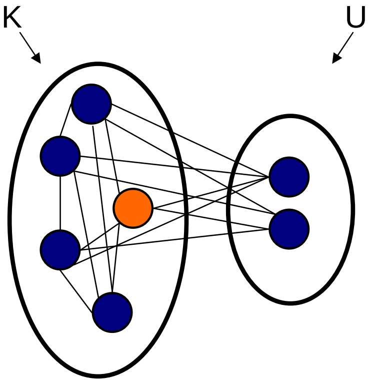


Figure: $U = \{v \in V \setminus K \ : \ K \cup \{v\}$ is a k-plex$\}$.

Computational Results (2.2 GHz Dual-Core AMD Opteron processor with 3 GB of memory)

Table: k-plex1 Results

G	$\omega_2(G)$	seconds	$\omega_3(G)$	seconds	$\omega_4(G)$	seconds
brock200-1	25	-	28	-	31	-
brock400-2	27	-	31	-	35	-
brock800-2	22	-	26	-	29	-
c-fat200-1	12	2	12	12	12	378
c-fat500-1	14	24	14	393	14	-
C125.9	42	-	49	-	56	-
hamming6-2	32	0	32	-	36	-
hamming8-2	128	1	128	-	128	-
johnson8-2-4	5	0	8	0	9	1
johnson16-2-4	10	-	16	-	18	-
johnson32-2-4	21	-	32	-	36	-
keller4	15	-	19	-	22	-
MANN-a9	26	103	36	4	36	592
p-hat300-1	10	107	12	-	14	-
p-hat700-1	12	-	13	-	16	-
p-hat1500-1	13	-	14	-	16	-
san200-0.7-2	26	-	36	-	48	-

- exceeded 3600 second time limit

Max k-plex Algorithm: Type 2*

- $V := \{v_1, ..., v_n\}$
- $S_i := \{v_i, ..., v_n\}$ for $1 \leq i \leq n$
- $c_k(i) := \omega_k(G[S_i])$

for $i : (n-1)$ to 1
    Search S_i for largest k-plex containing v_i.
    $c_k(i) \in \{c_k(i+1), c_k(i+1)+1\}$.
end

*Östergård

Max k-plex Algorithm: Type 2

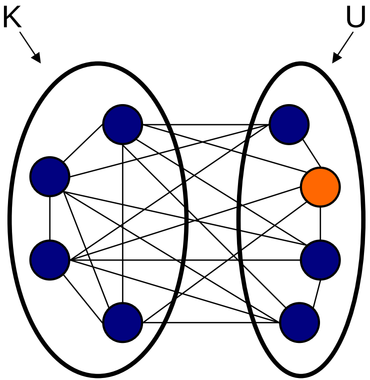


Figure: $j := min\{i \ : \ v_i \in U\}; \quad U \subseteq S_j \Rightarrow \omega_k(G[U]) \leq c_k(j).$

Computational Results (2.2 GHz Dual-Core AMD Opteron processor with 3 GB of memory)

Table: *k*-plex2 Results

G	$\omega_2(G)$	sec.	$\omega_3(G)$	sec.	$\omega_4(G)$	sec.
brock200-1	23	-	24	-	26	-
brock400-2	22	-	23	-	23	-
brock800-2	18	-	20	-	21	-
c-fat200-1	12	0	12	0	12	18
c-fat500-1	14	0	14	8	14	1234
C125.9	34	-	37	-	39	-
hamming6-2	32	0	32	1	40	951
hamming8-2	128	1	102	-	44	-
johnson8-2-4	5	0	8	0	9	0
johnson16-2-4	10	-	15	-	18	-
johnson32-2-4	21	-	24	-	25	-
keller4	15	913	21	-	16	-
MANN-a9	26	0	36	2	36	141
p-hat300-1	10	5	12	416	13	-
p-hat700-1	13	383	13	-	13	-
p-hat1500-1	12	-	14	-	13	-
san200-0.7-2	24	-	34	-	46	-

- exceeded 3600 second time limit

k-plex2 wins

Table: Results Summary

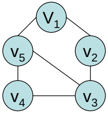
Algorithm	$k = 2$	$k = 3$	$k = 4$	Total
k-plex1-noBounds	13	8	5	26
k-plex1	16	7	5	28
k-plex2	19	14	11	44

Outline

Linear inequalities

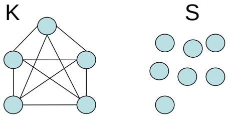
Let $G = (V, E)$, $S \subseteq V$, and $n = |V|$.

Consider the n-dimensional binary vector x^S where $x^S_v = 1$ if $v \in S$ and $x^S_v = 0$ otherwise.



Represent the stable set $S = \{v_1, v_3\}$ as $x^S = [1, 0, 1, 0, 0]^T$.

Linear inequalities



If $S \subseteq V$ is a stable set and $K \subseteq V$ is complete in G, then

$$\sum_{v \in K} x_v^S \leq 1$$

is a *valid inequality*.

Polyhedra and linear programming

Each valid inequality defines a halfspace in $\mathbf{R}^n$.

The intersection of all such halfspaces defines the polytope

$$P := \{x \in \mathbf{R}^n \ : \ Ax \leq b\}.$$

The linear program

$$max_{x \in P} \sum_{v \in V} x_v$$

determines the largest stable set in G.

Analogously...

Inequalities for co-k-plexes define the co-k-plex polytope

$$P_k := \{x \in \mathbf{R}^n \;:\; Ax \leq b\}.$$

The linear program

$$max_{x \in P} \sum_{v \in V} x_v$$

determines the largest co-k-plex in G.

Defining P_k

Definition

A *facet* is a valid inequality which must be present in any linear defining system $Ax \leq b$ (necessity).

The facets together form a defining system (sufficiency).

We focused on finding facets for the co-2-plex polytope.

Co-2-plexes are subgraphs with degree at most one.

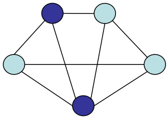
2-plexes

Theorem (McClosky and Hicks, Balasundaram et al.)

If K is a maximal 2-plex in G such that $|K| > 2$, then

$$\sum_{v \in K} x_v \leq 2$$

is a facet for $P_2(G)$.



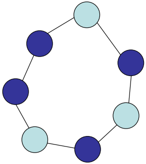
Cycles

Theorem (McClosky and Hicks)

If C^n is a chordless cycle such that $n > 4$ and $n \not\equiv 0$ mod 3, then

$$\sum_{v \in V(C^n)} x_v \leq \left\lfloor \frac{2n}{3} \right\rfloor$$

is a facet for $P_2(C^n)$.



Webs

Definition

For fixed integers $n \geq 1$ and p, $1 \leq p \leq \left\lfloor \frac{n}{2} \right\rfloor$,
the web $W(n,p)$ has vertices $V = \{1, ..., n\}$ and edges

$$E = \{(i,j) \mid j = i + p, ..., i + n - p;\ \forall\ i \in V\}.$$

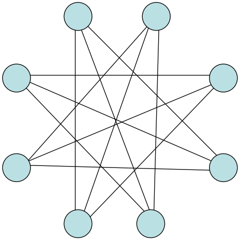


Figure: W(8,3)

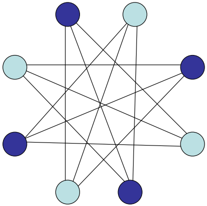
Web, cont.

Theorem (McClosky and Hicks)

If gcd($n, p+1$)=1 and $p < \left\lfloor \frac{n}{2} \right\rfloor$, then

$$\sum_{v \in V(W(n,p))} x_v \leq p+1$$

is a facet for $P_2(W(n,p))$.



Outline

Summary

- Algorithmic
  - co-k-plex coloring
  - k-plex heuristics
  - exact algorithms
- Polyhedral
  - linear description of the co-2-plex polytope

Future Work: Algorithmic

- exact co-k-plex coloring
- k-plex heuristics

Future Work: Polyhedral

- find facets for co-k-plex polyhedra ($k \geq 3$)
- computational study on facets we found

References

- Carlson, Zhang, Fang, Mischel, Howrvath, and Nelson. *Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks*, BMC Genomics 2006, 7:40.
- S. B. Seidman and B. L. Foster. *A graph theoretic generalization of the clique concept*. Journal of Mathematical Sociology, 6:139-154, 1978.
- Moody, James, and Douglas R. White (2003). "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups." American Sociological Review 68(1):103-127.
- B. Balasundaram, S. Butenko, I. Hicks, and S. Sachdeva. *Clique Relaxations in Social Network Analysis: The Maximum k-plex Problem.* Submitted, January 2006

Questions