Erdös's conjecture on multiplicities of complete subgraphs

Frantisek Franek

Algorithms Research Group
Department of Computing & Software
McMaster University
Hamilton, Ontario
L8S 4L7, Canada

Consider a random colouring of K_n by two colours and "calculate" the number of monochromatic t-cliques:

$$2\binom{n}{t}\left(\frac{1}{2}\right)^{\binom{t}{2}} = \binom{n}{t}\frac{2}{2^{\binom{t}{2}}}$$

$$\textit{proportion} \quad \frac{\#\ of\ monochrom.\ t-cliques}{\#\ of\ all\ t-subsets} = \frac{\binom{n}{t}\frac{2}{2^{\binom{t}{2}}}}{\binom{n}{t}} = \frac{2}{2^{\binom{t}{2}}} = 2^{1-\binom{t}{2}}$$

The probability that a randomly choosen t-subset is a monochromatic clique is

$$2^{1-\binom{t}{2}}$$

Consider a graph G of order n

$k_t(G) =$ **# of t-cliques in G**

$k_t(n) = \min\{k_t(G) + k_t(\overline{G}) : |G| = n\}$

$c_t(n) = \frac{k_t(n)}{\binom{n}{t}}$

The probabiliy that G "contains" a given t-clique =

quotient of # of all graphs on $n-t$ vertices and # of all graphs

on n **vertices** $= \dfrac{2^{\binom{n-t}{2}}}{2^{\binom{n}{2}}} \leq 2^{-\binom{t}{2}}$ **for** $n \geq t$

Therefore $k_t(G) \leq \binom{n}{t} 2^{-\binom{t}{2}}$ **and so** $k_t(G)+k_t(\overline{G}) \leq 2\binom{n}{t}2^{-\binom{t}{2}} = \binom{n}{t}2^{1-\binom{t}{2}}$

$c_t(n) \leq 2^{1-\binom{t}{2}}$

From ramsey theory $\frac{t!}{4^{t^2}} \leq c_t(n)$ **when** $n \gg t$**.**

Thus

$$\frac{t!}{4^{t^2}} \leq c_t(n) \leq 2^{1-\binom{t}{2}}$$

when $n \gg t$**.**

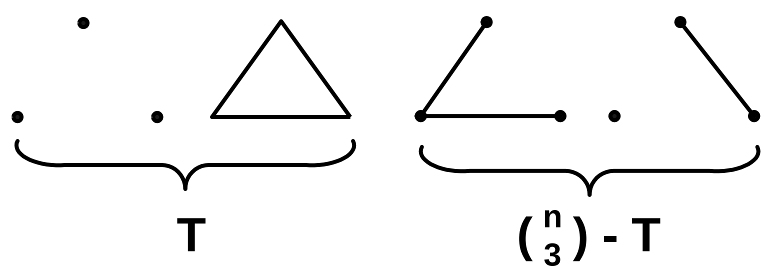
For a given t**,** $\{c_t(n)\}$ **is increasing, so** $c_t = \lim\limits_{n \to \infty} c_t(n)$ **exists.**

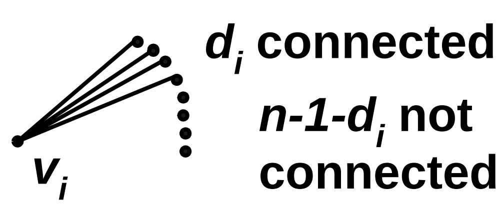
Erdös (1962) conjectured (sort of) that for $t \geq 2$,

$$c_t = \lim_{n \to \infty} c_t(n) = 2^{1-\binom{t}{2}}$$

Trivially true for $t = 2$:
$k_2(G) + k_2(\overline{G}) = \binom{|G|}{2}$**, and thus** $c_2(n) = 1$ **and hence** $c_2 = 1$**.**

Also true for $t = 3$:
from Goodman's work (1959)





$$2\left(\binom{n}{3}-T\right) = \sum_{i=1}^{n}(n-1-d_i)d_i = \sum_{i=1}^{n}(n-1)d_i - \sum_{i=1}^{n}d_i^2 \le (n-1)2e - \frac{1}{n}\left(\sum_{i=1}^{n}d_i\right)^2 \le$$

e **is the # of edges, and using Schwartz's inequality**

$$\le (n-1)2e - \frac{1}{n}4e^2 \le (n-1)\frac{n^2}{2} - \frac{n^3}{4} \le \frac{n^3}{2} - \frac{n^3}{4} = \frac{n^3}{4}$$

using maximizing value for

$$e = \frac{n(n-1)}{4} \approx \frac{n^2}{4}$$

$$\binom{n}{3} - T \leq \frac{n^3}{8}$$

$$\frac{T}{\binom{n}{3}} \geq 1 - \frac{\frac{n^3}{8}}{\binom{n}{3}} = 1 - \frac{\frac{n^3}{8}}{\frac{n(n-1)(n-2)}{6}} \approx$$

$$1 - \frac{\frac{n^3}{8}}{\frac{n^3}{6}} = \frac{1}{4}$$

The concjecture can naturally be modified for "counting" of other than complete subgraphs in other than complete graphs:

Erdös, Moon (1964) showed it true for "counting" complete bipartite subgraphs in bipartite graphs.

Sidorenko (1986-1993) showed it true for "cycles" in complete

graphs, and false for some incomplete subgraphs.

Giraud (1977) using geometrical interpretation and computational methods showed $c_4 > \frac{1}{46}$.

How does it relate to ramsey theory?

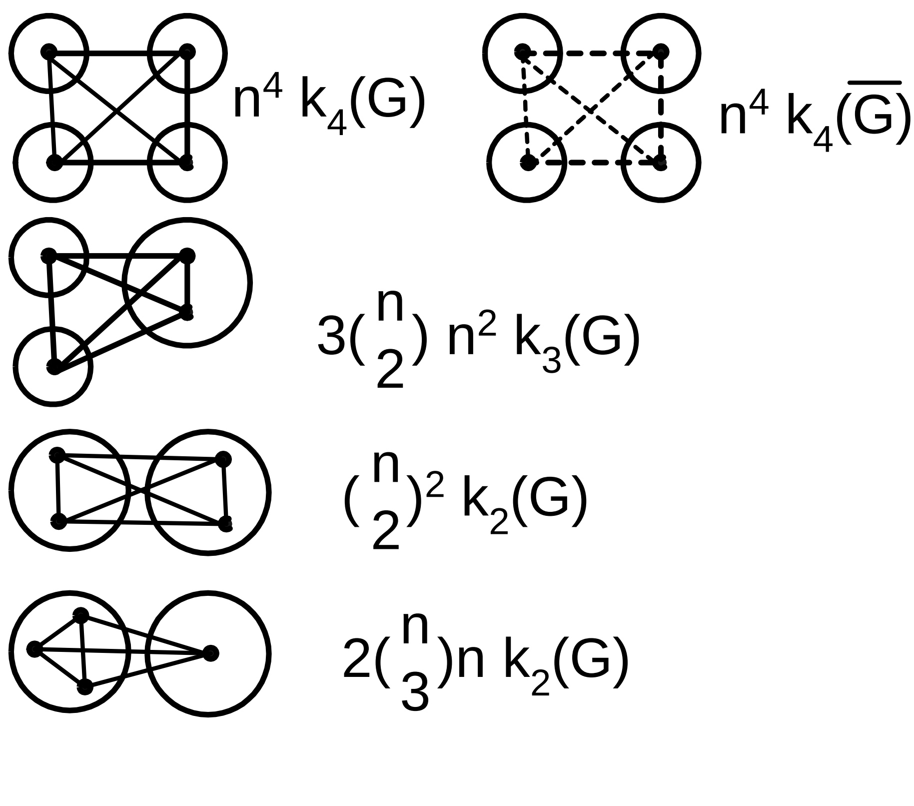
What must be the smallest order of graph G that any colouring by two colours yields a monochromatic K_t (diagonal Ramsey number $r(t,t)$)?

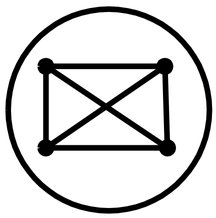
The knowledge of c_t can improve the bounds of $r(t,t)$ (communicated to me by Rödl, more about it at the end of the talk).

Pseudorandom or Quasirandom graphs - graphs that in some ways "behave" like random graphs. Introduced in many ways (Chung, Graham, Wilson 1989, Thomason 1985, 1987).
Frankl, Rödl, Wilson (1988), Thomason (1985) showed that Erdös's conjecture holds true for pseudorandom graphs, and hence not good to look for counterexamples there.

The following argument occurred in Thomason (1987), but I heard it from Rödl earlier, possibly folklore?
Consider a graph G. "Blow up" every vertex v of G to nv, a set

of verteces of size n. If v, u is an edge of G, then every vertex from nv forms an edge with every vertex from nu. Also, every two verteces from nv form an edge. The resulting graph is nG of order $n|G|$. Let us count 4-cliques in nG:





$\binom{n}{4}\, k_1(G)$

and thus

$$c_4 \le \lim_{n\to\infty} \frac{k_4(nG)+k_4(\overline{n}G)}{\binom{n|G|}{4}} = \sigma_4(G) =$$
$$\frac{24(k_4(G)+k_4(\overline{G}))+36k_3(G)+14k_2(G)+|G|}{|G|^4}$$

Can be generalized for any t, thus $c_t = \sigma_t(G)$ for any graph G. And thus a counterexample to the conjecture my be obtained by finding a suitable G.

Early 1980's: Rödl, Graham, Šiňajová unsuccesfully tried computer search for such a (small) G (computational limitations, begun with "pseudorandom" like

trying to improve on it).

Thomason (1987, the main paper 1989) disaproved the conjecture; found suitable G's constructed as "orthogonal towers" of certain vectors in orthogonal geometries V_t^+ (V_t^-) of dimension $2t$ over $\mathbf{F_2}$ with maximal (minimal) Witt indexes.

Thomason obtained the following upper bounds:

$$c_4 < 0.967{\cdot}2^{-5}; c_5 < 0.906{\cdot}2^{-9}$$

$$c_t < 0.936{\cdot}2^{1-\binom{t}{2}}, t \geq 6$$

In the late 1980's, Rödl and Franek set to find out why the

computational search failed and provide a simpler counterexample(s), possibly with better upper bound(s).

1989 (appeared 1992) Franek and Rödl showed that the conjecture for $t = 4$ is true for "nearly quasirandom" graphs, or graphs obtained from quasirandom by small "perturbations". That explained to some degree (besides computational limitations) the earlier failure, for in essence they were searching among "nearly quasirandom" graphs. The idea to make more "robust" perturbations in the search panned out and in 1993 Franek and Rödl provided a sim-

ple Cayley graph of order 2^{10} as a counterexample for $t = 4$ with the same upper bound as Thomason.

A sequence of graphs $\mathcal{R} = \{R_n\}_{n=0}^{\infty}$ *is a* **pseudorandom sequence** *iff for all but* $o(|V(R_n)|)$ *vertices* $u \in V(R_n)$, $d(u) = |N(u)|$ *satisfies* $\left| d(u) - \frac{|V(R_n)|}{2} \right| < o(|V(R_n)|)$, *and for all but* $o\left(\binom{|V(R_n)|}{2}\right)$ *pairs of vertices* $u, v \in V(R_n)$, *the size* $d(u,v)$ *of their common neighborhood* $N(u) \cap N(v)$ *satisfies* $\left| d(u,v) - \frac{|V(R_n)|}{4} \right| < o(|V(R_n)|)$.

Pseudorandom graphs have the following property

Theorem *Let $\mathcal{R} = \{R_n\}$ be a pseudorandom sequence of graphs, then there exists a sequence of positive reals $\{\varepsilon_n\}$ so that $\varepsilon_n \to 0$ as $n \to \infty$ and so that for every $V \subset V(R_n)$, $|V| \geq \varepsilon_n |V(R_n)|$, $\left(\frac{1}{2} - \varepsilon_n\right)\binom{|V|}{2} < e < \left(\frac{1}{2} + \varepsilon_n\right)\binom{|V|}{2}$, where e is the number of edges of R_n induced on a set V.*

For a graph $D = (V, E)$ and $U \subset V$ let $\delta_D(U) = \frac{E \cap [U]^2}{\binom{|U|}{2}}$ denote the edge density of the subgraph in-

duced on U. For a sequence $\mathcal{D} = \{D_n\}$ and $0 < p \leq 1$ let $p\mathcal{D} = \{pD_n\}$ be any sequence with the following property: $V_n = V(pD_n) = V(D_n)$, and there exists $\varepsilon_n \to 0$ such that $\left|\delta_{pD_n}(U) - p\delta_{D_n}(U)\right| < \varepsilon_n$ as $n \to \infty$ for any $U \subset V_n$, $|U| > \varepsilon_n |V_n|$. We can think of pD as a graph obtained from the graph D by flipping a p-biased coin for each edge of D, if the heads shows up the edge is left there, otherwise the edge is removed.

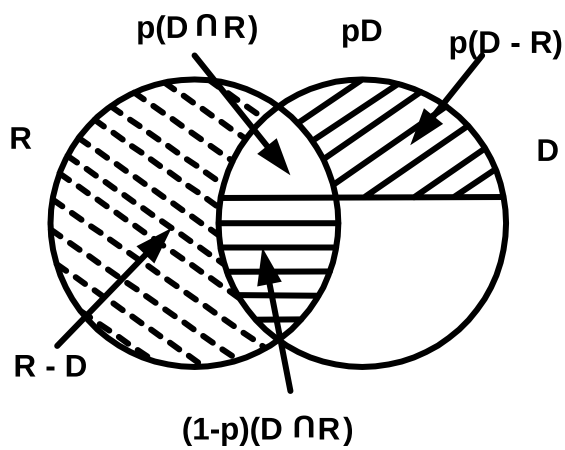
For a sequence $\mathcal{G} = \{G_n\}$ of graphs with $|V(G_n)| \to \infty$ as $n \to \infty$, let $d(\mathcal{G}) = lim\ inf\ d(G_n)$.

Let $\mathcal{H} = \{H_n\}$ be an arbitrary sequence of graphs and let $\mathcal{R} = \{R_n\}$ be a pseudorandom sequence with $V(R_n) = V(H_n) = V_n$ for all n. Let $D_n = R_n \div H_n$ be a graph whose edges are formed by all pairs one needs to change to obtain H_n from R_n (i.e. $E(D_n)$ is formed by symmetric difference $E(H_n) \div E(R_n)$). It follows that $H_n = R_n \div D_n$ as well. Suppose that we will not carry all the "changes" corresponding to D_n to obtain H_n from R_n but only "changes" on a "random" subgraph pD_n of D_n. This way we obtain a graph sequence $\{p(R_n, D_n)\} = \{R_n \div pD_n\}$. More

formally $p(R_n, D_n)$ is a graph sequence that satisfies:

- **there exists a sequence $\{\varepsilon_n\}$ of positive reals such that $\varepsilon_n \to 0$ and for every $U \subset V_n$, $|U| > \varepsilon_n |V_n|$, $|\delta_{p(R_n,D_n)}(U) - \delta_{R_n - D_n}(U) - (1-p)\delta_{R_n \cap D_n}(U) - p\delta_{D_n - R_n}(U)| < \varepsilon_n$.**

The diagram bellow shows the relative position of edge sets of R, D, pD, and $p(R, D)$.



If V, W are disjoint sets of vertices of G, then $e(V, W)$ denotes the number of edges of G with one endpoint in V and the other in W. $\delta(V, W) = \frac{e(V,W)}{|V| \cdot |W|}$ is the edge-density between V and W. If $\varepsilon > 0$, we say that V,W is an ε-uniform pair if $|\delta(V, W) - \delta(V', W')| < \varepsilon$ whenever $V' \subset V$ and $|V'| \geq \varepsilon \cdot |V|$, and $W' \subset W$ and $|W'| \geq \varepsilon \cdot |W|$.

Let t be a positive integer. $\vec{x}$ is a t-vector if it is a vector with t^2 real valued entries $x_{i,j}$, $1 \leq i, j \leq t$ and so that $x_{i,j} = x_{j,i}$. $B_t = \{\vec{x} \in R^{t^2} : \vec{x}$ is a t-vector & $|x_{i,j}| \leq 1$ *for all* $1 \leq i, j \leq t\}$.

Let G be a graph. Let $\varepsilon > 0$, and let t be a positive integer. We say that a t-vector $\vec{x}$ ε-represents graph G iff the vertex set of G can be partitioned into t disjoint classes $A_1, ..., A_t$ so that $\big||A_i| - |A_j|\big| \leq \mathbf{1}$ for all $1 \leq i, j \leq t$, and all but $t^2\varepsilon$ pairs $\{A_i, A_j\}$, are ε-uniform, and where $\delta(A_i, A_j) = \frac{1}{2}(1+x_{i,j})$ for all $1 \leq i, j \leq t$, $i \neq j$, and $\delta(A_i, A_i) = \delta(A_i)$ for all $1 \leq i \leq t$. If $\mathcal{G}$ is an infinite sequence of graphs and $\vec{x}$ is a t-vector, we say that $\vec{x}$ represents sequence $\mathcal{G}$ iff there is a sequence of positive reals $\{\varepsilon_n\}$ so that $\varepsilon_n \to 0$ and $\vec{x}$ ε_n-represents G_n, for every n.

We use t-vectors as representatives of sequences of graphs. For technical reasons the coordinates of t-vectors are not edge-densities directly, but edge-densities transformed by $p_{i,j} = \frac{1}{2}(1+x_{i,j})$. Henceforth B_t defined above is the part of R^{t^2} which is meaningful for us. Note also that the origin then represents pseudorandom graphs as $p_{i,j} = \frac{1}{2}$ corresponds to $x_{i,j} =$ 0.

Theorem. *A t-vector $\vec{x}$ represents a pseudorandom sequence iff $\vec{x} = \vec{o}$.*

We need a few polynomials in t^2 variables:

$$C_4(\vec{x}) = \frac{1}{2^6 \cdot t^4} \sum_{1 \leq i,j,k,l \leq t} [(1+x_{i,j})(1+x_{i,k})(1+x_{i,l})(1+x_{j,k})(1+x_{j,l})(1+x_{k,l}) + (1-x_{i,j})(1-x_{i,k})(1-x_{i,l})(1-x_{j,k})(1-x_{j,l})(1-x_{k,l})]$$

$$D(\vec{x}) = \frac{6}{2^5 \cdot t^4} \sum_{1 \leq i,j,k,l \leq t} [(1+x_{i,j})(1+x_{i,k})(1+x_{i,l})(1+x_{j,k})(1+x_{j,l}) + (1-x_{i,j})(1-x_{i,k})(1-x_{i,l})(1-x_{j,k})(1-x_{j,l})]$$

$$c(\vec{x}) = \frac{3}{2^5 \cdot t^4} \left(4t \sum_{1 \leq i,j,k \leq t} x_{i,j} x_{j,k} + \sum_{1 \leq i,j,k,l \leq t} x_{i,j} x_{k,l} \right)$$

$$b(\vec{x}) = \frac{3}{2^5 \cdot t^4} \left(\sum_{1 \leq i,j,k,l \leq t} x_{i,j} x_{i,l} x_{j,k} x_{k,l} + 4 \sum_{1 \leq i,j,k,l \leq t} x_{i,j} x_{i,l} x_{j,l} x_{k,l} \right)$$

$$a(\vec{x}) = \frac{1}{2^5 \cdot t^4} \sum_{1 \leq i,j,k,l \leq t} x_{i,j} x_{i,k} x_{i,l} x_{j,k} x_{j,l} x_{k,l}$$

Lemma.

(a) Let $\{\varepsilon_n\}$ *be an infinite sequence of positive reals so that* $\varepsilon_n \to 0$*.*

Let $\{t_n\}$ *be an infinite sequence of positive integers so that* $t_n \to \infty$*. Let* $\{G_n\}$ *be an infinite sequence of graphs. Let for each* n*,* $\vec{x}_n$ *be a* t_n*-vector such that it* ε_n*-represents graph* G_n*. Then* $\lim_{n\to\infty} c_4(G_n) = \lim_{n\to\infty} C_4(\vec{x}_n)$*, and* $\lim_{n\to\infty} d(G_n) = \lim_{n\to\infty} D(\vec{x}_n)$*.*

(b) Let a t*-vector* $\vec{x}$ *represent a graph sequence* $\mathcal{G}$*. Then* $d(\mathcal{G}) = D(\vec{x})$*.*

Lemma. *For any* t*-vector* $\vec{x}$*,*

$$C_4(\vec{x}) = \frac{1}{32} + c(\vec{x}) + b(\vec{x}) + a(\vec{x})$$

$$D(\vec{x}) = \frac{3}{8} + 4\Big(2c(\vec{x}) + b(\vec{x})\Big)$$

Lemma. *For any t-vector $\vec{x} \in B_t$, $|a(\vec{x})| \leq \frac{1}{32}$.*

Lemma. *For any t-vector $\vec{x}$, $c(\vec{x}) \geq 0$.*

Theorem. *Let $\mathcal{G}$ be a sequence of graphs. Then $d(\mathcal{G}) \geq \frac{3}{8}$ and equality holds if and only if $\mathcal{G}$ is a pseudorandom sequence.*

Lemma. *$D(\vec{x})$ is strictly minimal for $\vec{x} = \vec{o}$.*

Corollary. *For any t-vector $\vec{x}$, $2c(\vec{x}) + b(\vec{x}) \geq 0$. The equality is attained if and only if $\vec{x} = \vec{o}$.*

Lemma. *For any $\lambda > \frac{3}{8}$ there is μ_λ, $0 < \mu_\lambda \leq 1$, so that for any pos-*

itive integer t *and for any* $\vec{u} \in B_t$ *with* $D(\vec{u}) \geq \lambda$, $f_{\vec{u}}(\mu) = a(\vec{u})\mu^6 + b(\vec{u})\mu^4 + c(\vec{u})\mu^2 \geq \frac{1}{8}(\lambda - \frac{3}{8})\mu^4$ *for any* $\mu \in [0, \mu_\lambda]$.

Main theorem. *For every* $\lambda > \frac{3}{8}$ *there exists* p_λ, $0 < p_\lambda \leq 1$, *such that for every pseudorandom sequence of graphs* $\mathcal{R} = \{R_n\}$, *and for every sequence of graphs* $\mathcal{D} = \{D_n\}$ *with* $d(\mathcal{R} \div \mathcal{D}) \geq \lambda$, *if* $c_4(p(\mathcal{R}, \mathcal{D}))$ *exists, then* $c_4(p(\mathcal{R}, \mathcal{D})) \geq \frac{1}{32} + \frac{1}{8}(\lambda - \frac{3}{8})p^4$ *whenever* $0 < p \leq p_\lambda$.

It is not surprising (see the definition of how a a t-vector $\vec{x}$ ε-represents a graph) that the proof of the main theorem heavily relies on

Szemerédi's Regularity Lemma. *Given $\varepsilon > 0$, and a positive integer l. Then there exist positive integers $m = m(\varepsilon, l)$ and $n = n(\varepsilon, l)$ with the property that the vertex set of every graph G of order $\geq n$ can be partitioned into t disjoint classes $A_1, ..., A_t$ such that*

(a) $l \leq t \leq m$,

(b) $\big||A_i| - |A_j|\big| \leq 1$
for all $1 \leq i, j \leq t$,

(c) All but at most $t^2\varepsilon$ *pairs* A_i, A_j, $1 \leq i, j \leq t$, *are* ε*-uniform.*

Franek-Rödl's counterexample $\mathcal{G}_{n,F}$**: the set of vertices consists of all subsets of** $\{1, 2, ..., n\}$**, the cardinality family** F **is a subset of** $\{1, 2, ..., n\}$**,** $X, Y \subseteq \{1, 2, ..., n\}$

form an edge iff $|X \triangle Y| \in F$**. (**$\triangle$ **denotes symmetric difference)**

We can't calculate $k_t(\mathcal{G}_{n,F})$ **directly, but we can calculate (computer generate) ordered sequences** $\langle x_1, ..., x_{t-1} \rangle$ **so that** $|x_i|, |x_i \triangle x_j| \in F$**.**

Lemma. $k_{t+1}(\mathcal{G}_{n,F}) = \frac{2^n}{(t+1)!} s_t(n, F)$.

The parameters n **and** F **were used for the search, while** t **was fixed at 4. We found many counterexamples for** $n = 10$ **and** 11**.**

$n = 10$**,** $F = \{1, 3, 4, 7, 8, 10\}$ **give** $c_4 < 0.967501 \cdot 2^{-5}$ **and in** $\mathcal{G}_{10,F}$**, approximately** $\frac{1}{12}$ **of all pairs are**

edges, while $\frac{11}{12}$ of them are non-edges.

Jagger, Šťovíček, Thomason (1996) simplified the original Thomason's proof and discussed the problem from the extremal graph theory point of view investigating all kind of subgraphs rather than just cliques. They also obtained improved upper bounds $c_5 \le 0.8801{\cdot}2^{-9}$ and $c_6 \le 0.7641{\cdot}2^{-14}$.

Thomason (1997) described a method how to computer generate counterexamples for K_4 and other graphs. Interestingly, none improved on the upper bound.

The method could not be really appplied to K_t for higher t.

Franek (1997, a note to appear) extended Franek-Rödl's approach to $t = 5, 6$ - search infeasable, so worked with a fixed graph $\mathcal{G}_{10,\{1,3,4,7,8,10\}}$ that was found to be "good" for $t = 4$:

$c_5 < 0.885834{\cdot}2^{-9}$ **and**
$c_6 < 0.744514{\cdot}2^{-14}$

Interestingly, a referee of the note claimed to use the method and the graph an obtained $c_7 \leq 0.715527{\cdot}2^{-20}$.

As mentioned before, the knowl-

edge of c_t can lead to improvements of the bounds of Ramsey numbers:

- **Weak Rödl's conjecture:** $c_t 2^{\binom{t}{2}} \to 0$ **(this improves the bounds somehow)**
- **Strong Rödl's conjecture:** $c_t 2^{\binom{t}{2}} \to 0$ **exponentially fast (this improves the bounds exponentially)**

☺ ☺ ☺ ☺

All Franek-Rödl's papers on the topic as well these slides can be viewed at the web site www.cas.mcmaster.ca/∼franek