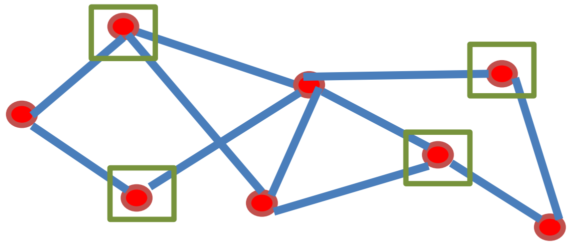
Independent sets in sparse graphs

[B'15]: Bansal

[BGG]: Bansal, Anupam Gupta, Guru Guruganesh

Independent set Problem

Given graph G find the largest independent set (size = α(G))



Approx. ratio (Alg) = $\max_G \alpha(G)$/ Alg(G)

Notoriously hard: $\Omega(n^{0.999\ldots})$ [Hastad 96]

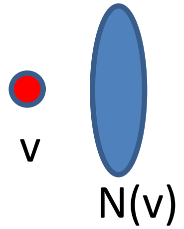
Best approx: $n/\log^3 n$ [Feige' 04]

Our Focus: max. degree = d (avg. degree d suffices)

(d+1) approximation trivial [Greedy $\geq \frac{n}{d+1}$]

Pick each vertex independently with prob. 1/(2d).

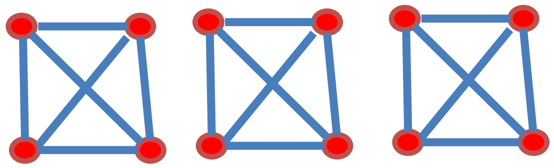
Retain v if no conflicts



Sparse graphs

In general n/d bound is tight

n/d disjoint copies of K_d



IP: $\max \sum_i x_i$ s.t. $x_i + x_j \leq 1$ if $(\mathrm{i},\mathrm{j}) \in E$ $x_i \in \{0,1\}$

LP relaxation useless: Ω(d) integrality gap (each $x_i = 1/2$)

First o(d) guarantee: O(d/ log log d)

[Bopanna, Halldorsson '94] [Ajtai, Erdos, Kolmos, Szemeredi' 81]

Current best: d (log log d) / log d using SDPs

[Halperin' 02 , Alon Kahale' 96 + Vishwanathan, Halldorsson'98]

Hardness

$\Omega(d/\log^2 d)$ assuming UGC [Austrin-Khot-Safra'11]
$\Omega(d/\log^4 d)$ assuming P $\neq$ NP. [Chan'13]

Hardness only for small d
(For d=n: best approx O(n/$\log^3 n$) [Feige' 04])

Right answer: $d/\log^2 d$ or $d/\log d$?

(i) SDP seems to not help beyond d log log d/log d
(ii) Ramsey theoretic barrier to showing $> d/\log^2 d$ hardness

Generic Approach

To get d/k approximation (say k = log d)

If OPT $\leq$ n/k (easy, Alg returns greedy n/d)

OPT > n/k must return something non-trivial

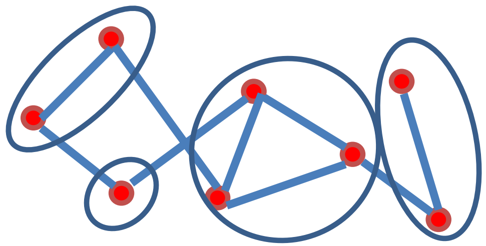
(i) Certify α(G) < n/k

(ii) If cannot certify, find non-trivial solution.

Certifying α(G) is small

Clique cover: V $= S_1 \cup \cdots \cup S_t$ (each S_i a clique)

$\alpha(G) \leq \chi(\overline{G})$



Theta number: $\alpha(G) \leq \vartheta(G) \leq \chi(\overline{G})$ (SDP captures cliques)

If $\vartheta(G) >$ n/k, find non-trivial independent set.

Ramsey theory: No large clique $\Rightarrow$ large independent sets

Ramsey theory (bounded degree graphs)

[Ajtai, Komlos, Szemeredi' 80]: If K_3-free, $\alpha(G) \geq \left(\frac{n}{d}\right) \log d$
(celebrated result; pioneered nibble method)

Tight: Random graph G(n,d/n) (girth $\approx \log n$)

Several other proofs known by now.

Johansson'96a: $\chi(G) \leq \frac{d}{\log d}$ [unpublished; Molloy-Reed'02]
(several nice ideas)

But K_3 is special: Even K_4-free case, much more challenging

Ramsey theory (bounded degree graphs)

K_r – free

[Ajtai, Erdos, Komlos, Szemeredi' 82]: $\alpha(G) \geq c_r \left(\frac{n}{d}\right) \log \log d$

[Shearer'95]: Beautiful entropy method (non-algorithmic)

$\alpha(G) \geq c_r \left(\frac{n}{d}\right) \frac{\log d}{\log \log d}$ (major question: remove log log d)

$c_r \approx \frac{1}{r}$ (get trivial bound for r > log d / log log d)

[Johansson'96b]: Never published, impossible to find, never verified …

Much stronger: $\chi(G) = O\left(\frac{r\, d \log \log d}{\log d}\right)$ & Algorithmic!

What does Ramsey give us?

Suppose SDP = n/r

(≈ each vertex contributes 1/r, say K_r-free)

Shearer: $\alpha(G) \geq \frac{1}{r}\left(\frac{n}{d}\right)\frac{\log d}{\log\log d}$

Integrality gap = $\frac{SDP}{\alpha(G)} \leq d\frac{\log\log d}{\log d}$

Same as given by Halperin, Alon-Kahale, … (perhaps not so surprising)

Can we combine Halperin + Ramsey?

Our Results

Thm[B'15]: $O(\log^4 d)$ levels of SA+ hierarchy, integrality gap $\approx \frac{d}{\log^2 d}$

(Entropy method: does not give a algorithmic result)

Thm[B'15]: n^c $\mathrm{quasipoly}(\mathrm{d})$ time, d/log d approx.

(beats Halperin by log log d)

Use O(log d) levels of SA+ to simulate AEKS'82 + Halperin

Thm[BGG] : Can get ≈ d/ $\log^2 d$ approximation in n^c *exp(d)* time.

(Use Johansson instead of entropy method, but need d levels of SA+)

(Johansson described in BGG)

Our Results

Thm [BGG] : Standard SDP, integrality gap $\approx \; d / \log^{3/2} d$

Non-algorithmic. Extend Shearer's result.

[Shearer'95]: $\alpha(G) \geq \left(\frac{n}{d}\right) \frac{\log d}{r \;\; \log \log d}$

Thm[BGG]: For any r, $\alpha(G) \; \geq \; \left(\frac{n}{d}\right) \sqrt{\log d / \log r}$

Eg. can set r = $\log^{100} d$

Thm [BGG]: LP based ≈ d/log d approximation (SA d-levels)

Rest of the talk

- SA+ -> better guarantees
- Shearer's entropy method
- Sketch of our K_r-free result
- Nibble Methods

See main idea behind each

Standard SDP formulation

SDP (vector program): vector v_i for vertex i.

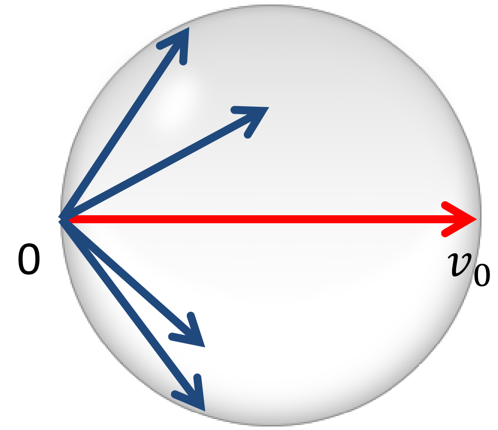
Intended solution: $v_i = v_0$ if i chosen, 0 otherwise.

Max $\Sigma_i \; v_i \cdot v_i$

$v_i \cdot v_j = 0 \qquad \text{if } (i,j) \in E$

$v_i \cdot v_0 = v_i \cdot v_i$

$v_0 \cdot v_0 = 1$



$$\left(v_i - \frac{v_0}{2}\right) \cdot \left(v_i - \frac{v_0}{2}\right) = \frac{v_0 \cdot v_0}{4} = \frac{1}{4}$$

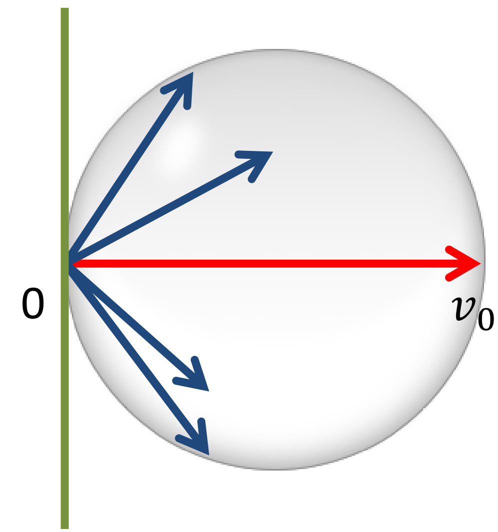
Exercise: Total contribution of a clique ≤ 1

Halperin's Rounding

Halperin: If SDP value = c n log log d/ log d,
gives independent set of size (n/d) $(\log d)^{2c-0.5}$

Sketch: Consider subspace $v_0^{\perp}$
Let $w_i = \text{Proj}(v_i)$ on $v_0^{\perp}$ $\quad (w_i = v_i - a_i v_0)$

If v_i and v_j orthogonal, then w_i and w_j
are negatively correlated
KMS rounding: Get advantage over random



k= log d/log log d

$\frac{d}{\log^2 d}$-approx with d levels of SA+ hierarchy

Hierarchies

Automatic way to strengthen LP or SDP relaxation by adding constraints (Lovasz Schrijver, Sherali Adams, Lasserre, …)

k-levels: $\approx n^k$ variables

n-levels = exact integer program

$$\max \sum x_{i,1}$$
$$x_{i,1} + x_{j,1} \leq 1 \quad (\mathrm{i, j}) \in E$$
$$x_{i,0} + x_{i,1} = 1$$

Distributional view: $x_{i,1}$= Pr[vertex i is picked] (marginals)

For an edge (i,j), cannot pick simultaneously

All-half solution: For edge (i,j) have (1,0) or (0,1) with prob. 0.5

Hierarchies

Automatic way to strengthen LP or SDP relaxation by adding constraints (Lovasz Schrijver, Sherali Adams, Lasserre, …)

k-levels: $\approx n^k$ variables

n-levels = exact integer program

$\max \sum x_{i,1}$

$x_{i,1} + x_{j,1} \leq 1 \quad (\mathrm{i,j}) \in E$

$x_{i,0} + x_{i,1} = 1$

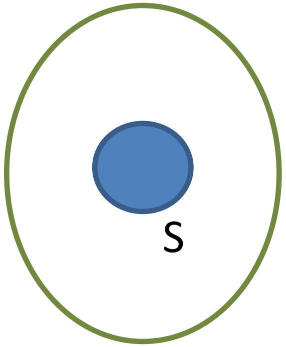
Variables: $x_{\{S,\alpha\}} \quad \alpha \in \{0,1\}^{|S|}$

Various constraints

For any subset of ≤ k vertices S,

1) A distribution over valid 0-1 assignments to S.

2) For S and T, distributions are consistent over $S \cap T$.



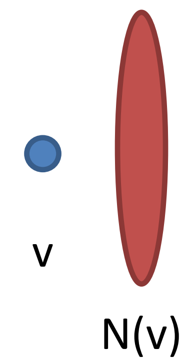
Strengthen SDP with d-levels of SA

Let $x_i = v_i \cdot v_i$ (contribution of i to objective)

If Val $\leq n/\log^2 d$, done.
Assume each $x_i \geq 1/(2\log^2 d)$ (delete i if not)

Claim: For any v, neighborhood N(v) is $O(\log^3 d)$ colorable.

Proof: $|N(v)| \leq d$, SA solution defines distribution over independent sets of N(v), s.t. each vertex lies in $\geq 1/(2\log^2 d)$ fraction of sets.



Alon'96: For locally-r colorable graphs

$$\alpha(G) \geq c\left(\frac{n}{d}\right)\left(\frac{\log d}{\log r}\right)$$

Recall SDP ≈ n/log d

Make algorithmic via Johansson

Locally-r-colorable graphs [Alon]

Goal: $\alpha(G) \geq c\left(\frac{n}{d}\right)\left(\frac{\log d}{\log r}\right)$

Proof: Pick set S uniformly from all independent sets.

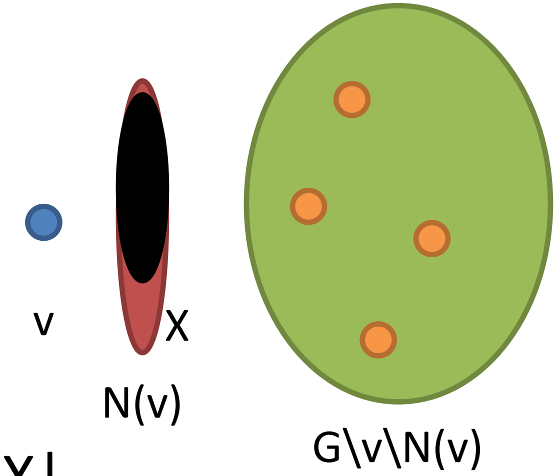
$p_v = E_S[\,|S \cap \{v\}|]$ and $Y_v = E_S[\,|S \cap N(v)|]$

E[S] $= \sum_v p_v$ and $E[S] \geq \frac{1}{d}\sum_v Y_v$

Suffices to show : $p_v + \frac{Y_v}{d} \geq 2c\left(\frac{1}{d}\right)\frac{\log d}{\log r}$ for all v.

In fact will show this holds for every conditioning of $S \setminus (v \cup N(v))$

$$p_v + \frac{Y_v}{d} \geq 2c\left(\frac{1}{d}\right)\frac{\log d}{\log r}$$



Condition on $S \cap G\backslash \mathrm{v} \backslash \mathrm{N}(v)$

Let X be vertices still available in N(v), and x = |X|

To extend independent set to G

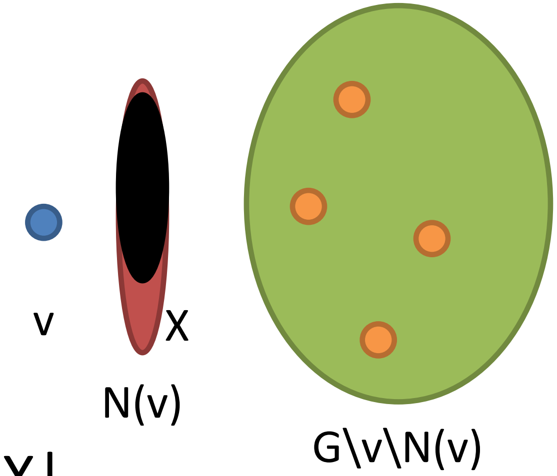
1) Either pick v (only)
2) Pick an independent set in X (and not v)

Let $2^{\epsilon x}$ be the number of independent sets in X.

$$\frac{1}{2^{\epsilon x}+1} + \frac{1}{d}\, s\, \frac{2^{\epsilon x}}{2^{\epsilon x}+1}$$

s = avg. size of independent set in X

$$p_v + \frac{Y_v}{d} \geq 2c\left(\frac{1}{d}\right)\frac{\log d}{\log r}$$



Condition on $S \cap G\backslash \mathrm{v} \backslash \mathrm{N}(v)$

Let X be vertices still available in N(v), and x = |X|

Let $2^{\epsilon x}$ be the number of independent sets in X.

Fact: $2^{\epsilon x}$ subsets in a universe of size x have size $\geq c\ \epsilon x/(\log 1/\epsilon)$

$$\frac{1}{2^{\epsilon x}+1} + \frac{1}{d}\, c\, \frac{\epsilon x}{\log 1/\epsilon}\ \frac{2^{\epsilon x}}{2^{\epsilon x}+1}$$

If $2^{\epsilon x} \leq \sqrt{d}$ done.

Otherwise, $\frac{\epsilon x}{\log 1/\epsilon} \geq \frac{½ \log d}{\log r}$ (as $\epsilon \geq \frac{1}{r}$)

Reducing the number of levels

Don't really need locally-r-colorable

Suffices if every subset S of N(v)
with $|S| \leq r \log^2 d$ is fractionally r-colorable

For us, $r = \log^2 d$. So $O(\log^4 d)$ levels of SA+ suffice

Gives an existential proof.
Open: Can Johansson's work with this weaker condition?

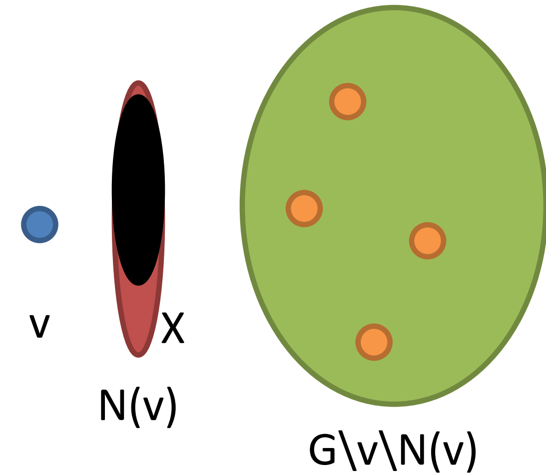
Shearer's K_r-free proof

K_r- free graph has an independent set of size $\approx$ $x^{1/r}$

Off-diagonal Ramsey: R(s,t) $\approx$ $\binom{s+t}{s}$

So $\geq \exp(x^{1/r})$ independent sets in X.

Apply to entropy method (with $\epsilon x \geq x^{1/r}$)



Some algebra: $\alpha(G) \geq \frac{1}{r}\left(\frac{n}{d}\right)\frac{\log d}{\log\log d}$

Our idea to extend to higher r: Do a more refined counting of independent sets.

Puzzle

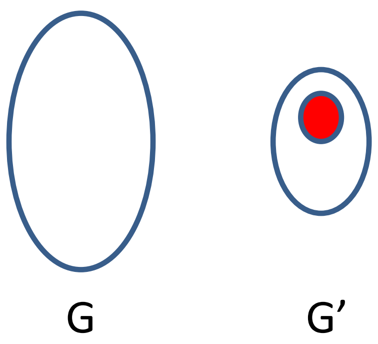
For K_3-free graphs $\alpha(G) \geq \frac{n}{d} \log d$

Number of independent sets $\geq 2^{\alpha(G)}$, at most $\binom{n}{\alpha(G)} \approx d^{\frac{n}{d} \log d}$

Where is the truth?

Thought experiment: Sample each vertex with $p = \frac{1}{\sqrt{d}}$.

Get G': $N' \approx pn \qquad d' \approx pd \qquad \text{so} \quad \frac{n'}{d'} \approx \frac{n}{d}$



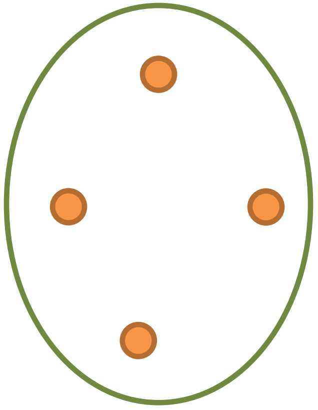
So $\alpha(G') \geq \left(n' \frac{\log d'}{d'}\right)$ = (n/d) (½ log d)

Key: Now, G must contain many such sets

A word about nibble methods

Nibble Methods (Johansson)

Coloring with s << d colors.

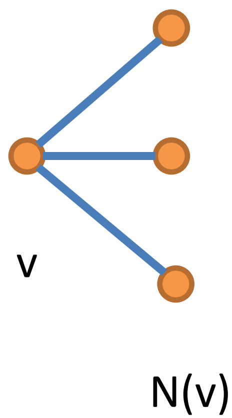


Each vertex maintains list of available colors.
Initially L(v) = {1,…,s}
Each round:
Some vertices activate, pick a random color from list.
Neighbors update their lists.

Thought experiment: If s = 2 d / log d
Each neighbor of v picks independently

Pr [color i left for v] = $\left(1-\frac{1}{s}\right)^{d} = \frac{1}{\sqrt{d}}$

Enough colors will be left for v.

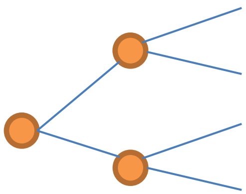


Nibble Methods (Johansson)

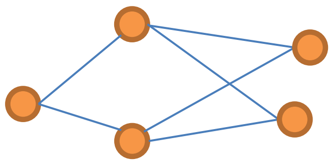
Coloring with s << d colors.

Each vertex maintains list of available colors.
Initially L(v) = {1,…,s}
Each round:
Some vertices activate, pick a random color from list.
Neighbors update their lists.

Kim'95: Girth 5, lists of neighbors of v ≈ independent



Triangle-free: Not true (new ideas: entropy, energy)
Johansson'96a: Never published; Molloy Reed book



K_r free, locally r-colorable
(new ideas: modifies the process) Johansson'96b

Open questions

Is SDP integrality gap d/ $\log^2 d$?

Best upper bound on $\alpha(G)$ for K_r-free graphs: $\alpha(G) \leq \left(\frac{n}{d}\right) \frac{\log d}{\log r}$

Matching tight (lower) bound will imply the result

Obtain an algorithmic $d\,/\log^{3/2} d$ approximation for SDPs (extend Johansson's approach)?

Using SA+, d/$\log^2 d$ approximation in n^c quasi-poly(d) time.

Questions!