Independent dominating sets in graphs of girth five via the semi-random method

Ararat Harutyunyan (Oxford), Paul Horn (Harvard), Jacques Verstraete (UCSD)

March 12, 2014
Introduction: The semi-random method

Independent dominating sets in graphs of girth five
   Main Idea of the proof
   The tools used in the proof

Conclusion
Let $2 \leq l < k < n$. Let $M(n, k, l)$ be the minimum size of a family $\mathcal{K}$ of $k$-element subsets of $\{1, \ldots, n\}$ such that every $l$-element subset is contained in at least one $A \in \mathcal{K}$.

**Fact:**

$M(n, k, l) \geq \frac{n^l}{k^l}$.

**Conjecture (Erdos-Hanani 1963):**

$M(n, k, l) \leq (1 + o(1)) \frac{n^l}{k^l}$. 
Hypergraph covering

- Let $2 \leq l < k < n$. Let $M(n, k, l)$ be the minimum size of a family $\mathcal{K}$ of $k$-element subsets of $\{1, \ldots, n\}$ such that every $l$-element subset is contained in at least one $A \in \mathcal{K}$.
- Fact: $M(n, k, l) \geq \binom{n}{l}/\binom{k}{l}$.
Hypergraph covering

Let $2 \leq l < k < n$. Let $M(n, k, l)$ be the minimum size of a family $\mathcal{K}$ of $k$-element subsets of $\{1, \ldots, n\}$ such that every $l$-element subset is contained in at least one $A \in \mathcal{K}$.

Fact: $M(n, k, l) \geq \binom{n}{l}/\binom{k}{l}$.

Conjecture (Erdos-Hanani 1963)

$M(n, k, l) \leq (1 + o(1))\binom{n}{l}/\binom{k}{l}$. 
Erdős-Hanani Conjecture

**Conjecture (Erdos-Hanani 1963)**

\[ M(n, k, l) \leq (1 + o(1))(\binom{n}{l}) / \binom{k}{l} \]

Conjecture (Erdos-Hanani 1963)

\[ M(n, k, l) \leq (1 + o(1)) \binom{n}{l} / \binom{k}{l} \]

Rodl’s Solution : the Rodl Nibble

Definition

Nibble: To bite in small bits.

The method of solution: Think algorithmically!

Build the covering family $K$ over many iterations.
Rodl’s Solution: the Rodl Nibble

Definition

Nibble: To bite in small bits.
Rodl’s Solution: the Rodl Nibble

- **Definition**
  - **Nibble**: To bite in small bits.
  - The method of solution: Think algorithmically!
Rodl’s Solution : the Rodl Nibble

I Definition

Nibble: To bite in small bits.

I The method of solution: Think algorithmically!
I Build the covering family $\mathcal{K}$ over many iterations.
A randomized algorithm

- Start $\mathcal{K} = \emptyset$.
- In each iteration, randomly pick few $k$-element sets to add to the covering family $\mathcal{K}$.
- Argue that in each iteration $\mathcal{K}$ does not grow too fast whp (1).
- Argue that in each iteration no $l$-element set is covered more than once whp (2).
- Deduce that with positive probability there is a choice of $k$-element sets to pick satisfying conditions (1) and (2).
- Condition on this good occurrence and... Repeat!
Semi-random: Not Random

- The algorithm is not actually random.
Semi-random: Not Random

- The algorithm is not actually random.
- Each iteration is **Deterministic**: we use probability to show that a good decision (nibble) exists.
Other fundamental results proved using the semi-random method

1. $G$ triangle-free graph, $\chi(G) = O(\Delta / \log \Delta)$. (Johansson 1995).
Other fundamental results proved using the semi-random method

- $G$ triangle-free graph, $\chi(G) = O(\Delta / \log \Delta)$. (Johansson 1995).
- Ramsey theory: $R(3, t) \sim t^2 / \log t$. (Kim 1996)
Other fundamental results proved using the semi-random method

- $G$ triangle-free graph, $\chi(G) = O(\Delta / \log \Delta)$. (Johansson 1995)
- Ramsey theory: $R(3, t) \sim t^2 / \log t$. (Kim 1996)
- Designs: “Existence” Conjecture (Keevash 2014)
The success of the method hinges on two concepts
1: Almost all random variables have a Normal-like distribution.
2: Almost all random variables are only locally dependent on each other.
Dominating sets

Definition
Graph $G = (V, E)$: set $S \subseteq V$ is a dominating set if every $v \in V - S$ is adjacent to a vertex in $S$. 
An old theorem


Graph $G$ with minimum degree $d$. Then $G$ has a dominating set of size at most $\frac{n(\log(d+1)+1)}{d+1}$.

Proof.
An old theorem


Graph $G$ with minimum degree $d$. Then $G$ has a dominating set of size at most $\frac{n(\log(d+1)+1)}{d+1}$.

Proof.

1. Put each vertex $v \in V$ in a set $X$ with probability $p$, independently.
An old theorem


Graph $G$ with minimum degree $d$. Then $G$ has a dominating set of size at most $\frac{n(\log(d+1)+1)}{d+1}$.

Proof.

1. Put each vertex $v \in V$ in a set $X$ with probability $p$, independently.

2. Set $Y_X :=$ vertices not dominated by $X$. Then $E[|Y_X|] \leq n(1-p)^{d+1}$.
An old theorem

**Theorem (Lovasz(1975), Payan(1974), Arnautov(1974))**

*Graph G with minimum degree d. Then G has a dominating set of size at most* \( \frac{n(\log(d+1)+1)}{d+1} \).

**Proof.**

1. Put each vertex \( v \in V \) in a set \( X \) with probability \( p \), independently.
2. Set \( Y_X := \) vertices not dominated by \( X \). Then \( E[|Y_X|] \leq n(1 - p)^{d+1} \)
3. \( X \cup Y_X \) is a dominating set with \( E[|X \cup Y_X|] \leq np + n(1 - p)^{d+1} \).
An old theorem


Graph $G$ with minimum degree $d$. Then $G$ has a dominating set of size at most $\frac{n(\log(d+1)+1)}{d+1}$.

Proof.

- Put each vertex $v \in V$ in a set $X$ with probability $p$, independently.
- Set $Y_X :=$ vertices not dominated by $X$. Then $E[|Y_X|] \leq n(1 - p)^{d+1}$
- $X \cup Y_X$ is a dominating set with $E[|X \cup Y_X|] \leq np + n(1 - p)^{d+1}$.
- Therefore, $\exists$ a dominating set of size $np + n(1 - p)^{d+1}$. Setting $p = \frac{\log(d+1)}{d+1}$ gives the result.
Independent dominating sets

**Definition**

$S \subset V(G)$ is called an *independent dominating set* if $S$ is both an independent set and a dominating set.

**Remarks**
Independent dominating sets

Definition

$S \subset V(G)$ is called an independent dominating set if $S$ is both an independent set and a dominating set.

Remarks

- A maximal independent set in $G$ is an independent dominating set.
The Main Result

Theorem (Horn, Verstraete, H. 2012)

Every $d$-regular graph of girth at least five has an independent dominating set of size at most $\frac{n(\log d + c)}{d}$, where $c$ is an absolute constant.

Remarks
The Main Result

Theorem (Horn, Verstraete, H. 2012)

Every \textit{d-regular} graph of girth at least five has an independent dominating set of size at most \(\frac{n(\log d + c)}{d}\), where \(c\) is an absolute constant.

Remarks

\begin{itemize}
  \item The previous method of proof will not work. Picking vertices with probability \(p = \frac{\log d}{d}\) will result in too many picked vertices being adjacent.
\end{itemize}
The Main Result

Theorem (Horn, Verstraete, H. 2012)

Every $d$-regular graph of girth at least five has an independent dominating set of size at most $\frac{n(\log d + c)}{d}$, where $c$ is an absolute constant.

Remarks

- The previous method of proof will not work. Picking vertices with probability $p = \log d / d$ will result in too many picked vertices being adjacent.

- **Idea:** pick vertices with smaller probability (roughly $1/d$), remove dominated vertices from the graph, and repeat.
A randomized algorithm

The proof uses the following randomized algorithm.

1. Build an independent dominating set by iterations.

2. During each iteration $t$, we randomly select each undominated vertex with probability $p = 1/d_t$, where $d_t$ will be roughly the average degree of the graph at time $t$. If two adjacent vertices were selected, un-select both of them.

3. Mark all the neighbors of the selected vertices as dominated. These vertices will not be selected at future iterations.

4. **Technical Trashcan:** Put each vertex $v$ not in the current dominating set or their neighborhood in a set $C$ with some probability $q(v)$. Purpose: to keep the undominated graph regular.
Sets and Random Variables

We have the following sets:
\(X_t:=\) the set vertices in \(G\) that still need to be dominated at time \(t\)
\(S_{t+1} :=\) the set of vertices selected to be in the independent dominating set at time \(t\)
\(Q_{t+1} :=\) the set of vertices put in the trashcan at time \(t\). These vertices will not be used to build the independent dominating set.

Also define real numbers:
\(n_t:=\) roughly the size of \(X_t\) we would expect at time \(t\)
\(d_t:=\) average degree of a vertex in \(X_t\) that we would expect at time \(t\).
Set \(S_0 = Q_0 = \emptyset, X_0 = V(G), d_0 = d, n_0 = n.\)
Sets and Random Variables

I Define $d_t = d \prod_{i=1}^{t} q_i$ and $n_t = n \prod_{i=1}^{t} q_i$, where $q_i \approx e^{-1/e}$.

I At time $t$, select each vertex in $X_t$ with probability $1/d_t$, independently. Let $S_{t+1}$ be the set of selected vertices in $X_t$ which have no selected neighbors.

I For each vertex $v \in X_t \setminus (S_{t+1} \cup \partial S_{t+1})$, we put $v$ in $Q_{t+1}$ with probability $q_{t+1}(v)$. $q_{t+1}(v)$ is defined so that $P(v \notin \partial S_{t+1})(1 - q_{t+1}(v)) = q_{t+1}$. 
Updating the sets

- $X_t$ is the set from which we can take vertices to build the independent dominating set at time $t$.
- $C_t$ is the set of vertices which will not be used to build the independent dominating set.
- How the sets are (roughly) updated:
  
  \[
  C_{t+1} = C_t \cup Q_{t+1} \\
  X_{t+1} = X_t \setminus (Q_{t+1} \cup S_{t+1} \cup \partial S_{t+1}).
  \]
Preserving the regularity of degrees

During each iteration, we need to ensure that $X_t$-degrees of vertices are all roughly the same.

Lower bounding the minimum degree allows us to claim that each randomly selected vertex dominates many vertices.

Upper bounding the maximum degree allows us to claim that most randomly picked vertices are not adjacent.

For a vertex $v \in X_t \cup C_t$, define the random variable $d_t(v)$ to be the number of neighbors of $v$ in $X_t$.

For a vertex $v \in X_t \cup C_t$, define the random variable $\gamma_t(v)$ to be the number of neighbors of $v$ in $C_t$. 
Preserving the regularity of degrees

During each iteration, we need to ensure that $X_t$-degrees of vertices are all roughly the same.
Preserving the regularity of degrees

- During each iteration, we need to ensure that $X_t$-degrees of vertices are all roughly the same.
- Lower bounding the minimum degree allows us to claim that each randomly selected vertex dominates many vertices. Upper bounding the maximum degree allows us to claim that most randomly picked vertices are not adjacent.
Preserving the regularity of degrees

During each iteration, we need to ensure that $X_t$-degrees of vertices are all roughly the same.

Lower bounding the minimum degree allows us to claim that each randomly selected vertex dominates many vertices. Upper bounding the maximum degree allows us to claim that most randomly picked vertices are not adjacent.

For a vertex $v \in X_t \cup C_t$, define the random variable $d_t(v)$ to be the number of neighbors of $v$ in $X_t$. 
Preserving the regularity of degrees

- During each iteration, we need to ensure that $X_t$-degrees of vertices are all roughly the same.

- Lower bounding the minimum degree allows us to claim that each randomly selected vertex dominates many vertices. Upper bounding the maximum degree allows us to claim that most randomly picked vertices are not adjacent.

- For a vertex $v \in X_t \cup C_t$, define the random variable $d_t(v)$ to be the number of neighbors of $v$ in $X_t$.

- For a vertex $v \in X_t \cup C_t$, define the random variable $\gamma_t(v)$ to be the number of neighbors of $v$ in $C_t$. 
The algorithm is semi-random

We show that at each iteration all of the following set of events hold simultaneously with positive probability:

\[ |d_{t+1}(v) - d_{t+1}| \leq \epsilon_{t+1} \quad \forall v \in X_{t+1} \cup C_{t+1} \]  

\[ \gamma_{t+1}(v) \leq 100 \epsilon_{t+1} \quad \forall v \in X_{t+1} \cup C_{t+1} \]  

\[ |C_{t+1}| \leq 200 \frac{\epsilon_{t+1} n_{t+1}}{d_{t+1}} \]  

\[ |S_{t+1} - \frac{n}{ed}| \leq 3 \max \left\{ \frac{\epsilon_{t+1} n_{t+1}}{d_{t+1}^2}, \frac{n_{t+1}}{\sqrt{d_{t+1} d}} \right\} \]  

\[ |X_{t+1} - n_{t+1}| \leq 20 \frac{n_{t+1}}{d_{t+1}}. \]

provided they hold at time \( t \).
How long is the algorithm be run?

- We run the algorithm until time $T \approx e \log d$.
- Since $|S_t| \approx \frac{n}{ed}$ for all $t$, $|\cup_{t=1}^{T} S_t|$, the total size of the selected vertices over the $T$ iterations, is $\approx n \log d / d$.
- Since $|X_t| \approx n_t \approx ne^{-t/e}$, $|X_T| \approx n / d$.
- Since $|C_t| \approx \frac{n_t}{d_t} = \frac{n}{d}$, then $|C_T \cup X_T| = O(n/d)$.
- Just pick a maximal independent set in $C_T \cup X_T$.
- There is an independent dominating set in $G$ of size at most $|\cup_{t=1}^{T} S_t \cup X_T \cup C_T| \leq \frac{n \log d}{d} + O(n/d)$. 
Preserving the property

At each step we want to show that the following set of events hold with positive probability:

\[ |d_{t+1}(v) - d_{t+1}| \leq \epsilon_{t+1} \quad \forall v \in X_{t+1} \cup C_{t+1} \quad (6) \]
\[ \gamma_{t+1}(v) \leq 100\epsilon_{t+1} \quad \forall v \in X_{t+1} \cup C_{t+1} \quad (7) \]
\[ |C_{t+1}| \leq 200 \frac{\epsilon_{t+1} n_{t+1}}{d_{t+1}} \quad (8) \]
\[ |S_{t+1} - \frac{n}{ed}| \leq 3 \max\{ \frac{\epsilon_{t+1} n_{t+1}}{d_{t+1}^2}, \frac{n_{t+1}}{\sqrt{d_{t+1}d}} \} \quad (9) \]
\[ |X_{t+1} - n_{t+1}| \leq 20 \frac{n_{t+1}}{d_{t+1}}. \quad (10) \]
Two main tools

There are two tools involved in showing this.

1. Show that each single event occurs with high probability.
   - Use concentration inequalities.

2. Argue that the events are only locally dependent.
   - Use the Lovasz Local Lemma to show that all events occur with positive probability.
Two main tools

There are two tools involved in showing this.

1. Show that each *single* event occurs with high probability.

2. Argue that the events are only locally dependent. Use the Lovasz Local Lemma to show that all events occur with positive probability.
Two main tools

There are two tools involved in showing this.

1. Show that each *single* event occurs with high probability.
2. Argue that the events are only locally dependent.
   - Use the Lovasz Local Lemma to show that all events occur with positive probability.
Two main tools

There are two tools involved in showing this.

1. Show that each *single* event occurs with high probability.
   - Use concentration inequalities.
2. Argue that the events are only locally dependent.
Two main tools

There are two tools involved in showing this.

1. Show that each *single* event occurs with high probability.
   - Use concentration inequalities.
2. Argue that the events are only locally dependent.
   - Use the Lovasz Local Lemma to show that *all* events occur with positive probability.
Martingale/Concentration Inequalities

Concentration Inequalities claim that often, under very weak conditions, one can claim that a random variable is strongly concentrated around its expected value.
Concentration Inequalities claim that often, under very weak conditions, one can claim that a random variable is strongly concentrated around its expected value.

Suppose $X = f(Z_1, Z_2, \ldots, Z_k)$ is a random variable that is a function of many independent random variables $Z_i$ with the property that changing each single $Z_i$ will have little impact on $X$. Then whp $X$ does not deviate too much from its mean.
Hoeffding-Azuma Inequality

Theorem (Hoeffding-Azuma Inequality)

Let $X = f(Z_1, ..., Z_l)$ where the $Z_i$ are independent random variables and suppose that changing the outcome of each single $Z_k$ can change $X$ by at most the amount $c_k$. Then $X$ satisfies

$$P[|X - E[X]| > t] \leq 2\exp\left\{-2t^2/l \sum_{k=1}^{l} c_k^2\right\}$$

for all $t > 0$. 

Theorem

Let $A_1, ..., A_m$ be a set of "bad" events in some probability space, and suppose that for some set $J_i \subset [n]$, $A_i$ is mutually independent of $\{A_j : j \notin J_i \cup \{i\}\}$. If there exist real numbers $\gamma_i \in [0, 1)$ such that $P(A_i) \leq \gamma_i \prod_{j \in J_i} (1 - \gamma_j)$, then

$$P(\cap_{i=1}^n A_i^c) \geq \prod_{i=1}^n (1 - \gamma_i) > 0.$$
Applying the concentration inequality

**Lemma** Let \( v \in X_{t+1} \) and \( d_t > K, K \) a large constant. Then

\[
P[|d_{t+1}(v) - d_{t+1}| > \epsilon_{t+1}] \leq d_t^{-100}.
\]

**Proof sketch**
Applying the concentration inequality

Lemma Let $v \in X_{t+1}$ and $d_t > K$, $K$ a large constant. Then

$$P[|d_{t+1}(v) - d_{t+1}| > \epsilon_{t+1}] \leq d_t^{-100}.$$  

Proof sketch

First, we show that $E[d_{t+1}(v)] \approx d_{t+1}$. This means we can concentrate around $d_{t+1}$ rather than $E[d_{t+1}(v)]$. 

For a vertex $u$, let $I_u$ be the indicator r.v. that $u$ is selected at time $t$ with probability $1/d_t$, and $J_u$ the indicator r.v. that $u$ is put in $Q_{t+1}$.

$d_{t+1}(v)$ is a function of r.v's $I_u$ and $J_u$.

Since girth $\geq 5$, then whp no single r.v $I_u$ and $J_u$ can affect $d_{t+1}(v)$ very much.
Applying the concentration inequality

**Lemma** Let $v \in X_{t+1}$ and $d_t > K$, $K$ a large constant. Then

$$P[|d_{t+1}(v) - d_{t+1}| > \epsilon_{t+1}] \leq d_t^{-100}.$$ 

**Proof sketch**

- First, we show that $E[d_{t+1}(v)] \approx d_{t+1}$. This means we can concentrate around $d_{t+1}$ rather than $E[d_{t+1}(v)]$.

- For a vertex $u$, let $I_u$ be the indicator r.v. that $u$ is selected at time $t$ with probability $1/d_t$, and $J_u$ the indicator r.v. that $u$ is put in $Q_{t+1}$.
Applying the concentration inequality

**Lemma** Let $v \in X_{t+1}$ and $d_t > K$, $K$ a large constant. Then

$$P[|d_{t+1}(v) - d_{t+1}| > \epsilon_{t+1}] \leq d^{-100}_t.$$

**Proof sketch**

- First, we show that $E[d_{t+1}(v)] \approx d_{t+1}$. This means we can concentrate around $d_{t+1}$ rather than $E[d_{t+1}(v)]$.
- For a vertex $u$, let $I_u$ be the indicator r.v. that $u$ is selected at time $t$ with probability $1/d_t$, and $J_u$ the indicator r.v. that $u$ is put in $Q_{t+1}$.
- $d_{t+1}(v)$ is a function of r.v’s $I_u$ and $J_u$. 

Since girth $\geq 5$, then whp no single r.v $I_u$ and $J_u$ can affect $d_{t+1}(v)$ very much.
Applying the concentration inequality

**Lemma** Let $v \in X_{t+1}$ and $d_t > K$, $K$ a large constant. Then

$$P[|d_{t+1}(v) - d_{t+1}| > \epsilon_{t+1}] \leq d_t^{-100}.$$

**Proof sketch**

- First, we show that $E[d_{t+1}(v)] \approx d_{t+1}$. This means we can concentrate around $d_{t+1}$ rather than $E[d_{t+1}(v)]$.
- For a vertex $u$, let $I_u$ be the indicator r.v. that $u$ is selected at time $t$ with probability $1/d_t$, and $J_u$ the indicator r.v. that $u$ is put in $Q_{t+1}$.
- $d_{t+1}(v)$ is a function of r.v’s $I_u$ and $J_u$.
- Since girth $\geq 5$, then whp no single r.v $I_u$ and $J_u$ can affect $d_{t+1}(v)$ very much.
Concluding Remarks

The upper bound in the theorem cannot be significantly improved: all the independent dominating sets in the random $d$-regular graph on $n$ vertices have size at least $n \log_d d - cn/d$ for some constant $c$. 
The upper bound in the theorem cannot be significantly improved: all the independent dominating sets in the random $d$-regular graph on $n$ vertices have size at least $\frac{n \log d}{d} - \frac{cn}{d}$ for some constant $c$. 
Relaxing the regularity condition in the theorem

- The regularity condition cannot be significantly improved: Take the graph that consists of the random graph $G_{n/2, 2d/n}$ and $\bar{K}_{n/2}$ where each vertex $v \in \bar{K}_{n/2}$ is connected to $d$ randomly chosen vertices in $G_{n/2, 2d/n}$.

- If $d$ is large, whp every vertex has degree at least $d$ and at most $3d$. We can remove a few edges to ensure that there are no triangles or 4-cycles.

- Every independent set in $G_{n/2, 2d/n}$ has size at most $\approx \frac{n \log d}{2d} \Rightarrow$ many vertices in $\bar{K}_{n/2}$ will be uncovered.
Relaxing the girth condition

The girth 5 condition cannot be improved: take the graph consisting of disjoint copies of $K_{d,d}$. 
An open question...

Is the following conjecture true?

**Conjecture**

*There exists an absolute constant $c$ such that any $n$-vertex $d$-regular graph with no cycles of length 4 has an independent dominating set of size at most $\frac{n(\log d + c)}{d}$.*
Thank You