Counting Independent sets in Graphs

Xinyi XU

SN: 14020529

Supervised by Dr. J. Talbot

A report submitted for the fourth year project
of
Msci Mathematics with Economics
of
University College London.

Department of Mathematics
University College London

March 8, 2018

Abstract

In this project, we will discuss an algorithm introduced by Kleitman and Winston, which encodes independent sets of a given graph in an invertible manner. Using this algorithm, we can give bounds on the number of independent sets in various graphs. Many problems in discrete mathematics can be solved or simplified by counting independent sets in relevant graphs and this algorithm has a number of interesting applications. We will also discuss some applications and in particular, we will derive bounds for counting sum-free sets in $\{1, 2, ..., n\}$, discuss the relationship between a graph and its eigenvalues of its adjacency matrix, counting C_4-free graphs and prove Roth's theorem.

Contents

Chapter 1

Introduction

1.1 Why are we interested in independent sets

In mathematics, we see quite a lot of problems which can be considered as graph theory problems. For examples the famous Königsberg seven bridges problem and the four color theorem.

Definition 1. *A graph is an ordered pair $G = (V, E)$ where V is the vertex set $\{v_1, v_2, ..., v_n\}$ with set of edges $E \subseteq \binom{V}{2}$. Two vertices $v, w \in V$ are adjacent if there is an edge $vw \in E$.*

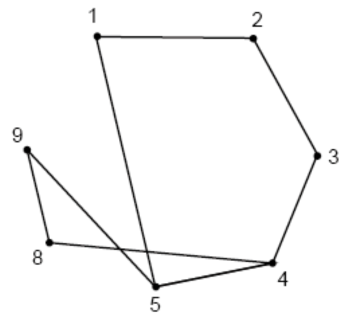
Definition 2. *The order of a graph is $|V|$, i.e. the number of vertices in G.*

Definition 3. *The neighborhood of a vertex $v \in V(G)$ is the set $N_G(v)$ of all vertices in $V(G)$ adjacent to v in G. The neighborhood of a subset $W \subseteq V(G)$ is the set $N_G(W)$ of all vertices in $V(G)$ adjacent to at least one vertex in W. The degree of a vertex v, $d(v)$ is the size of its neighborhood.*

Definition 4. *An independent set I in a graph $G = (V, E)$ is a subset of vertices in V that contains no edges.*

Many questions in combinatorics can be phrased in terms of counting independent sets. We will first see an easy example to help us understand the definitions and algorithms in this project (which will be referred further in relevant sections). We will then use our understanding to discuss some more complicated applications.

Example 1. *Let $A \subseteq \mathbb{N}$ be given, we want to count subsets $A' \subseteq A$ such that the difference between each pair of numbers in A' is not a square number. We can define a graph $G = (A, E)$ where $ab \in E$ if and only if $|a - b|$ is a square. Hence number of A' is the number of independent sets in G.*
See the graph where $A = \{1, 2, 3, 4, 5, 8, 9\}$, edges built with rules above.



Some of its independent sets are: $\{1, 4\}$, $\{4, 9\}$, $\{1, 3, 8\}$, $\{1, 3, 9\}$. □

Now we consider counting sum-free sets in a subset of $\mathbb{Z}$. We say a set of integers A is *sum-free* if there is no $x, y, z \in A$ so that $x + y = z$. Cameron and Erdös asked (see [1]) the question: how many sum-free subsets can we find in $[n] = \{1, 2, ..., n\}$? Note that any subset only consisting of odd elements or only elements larger than $\frac{n}{2}$ is sum-free. In fact, these are the largest examples.

Lemma 1. *If $A \subseteq [n]$ is sum-free then $|A| \leq \lceil \frac{n}{2} \rceil$.*

Proof. (By contradiction) Assume that we have a sum-free subset $A \subseteq [n]$, $A = \{x_1, x_2, ..., x_k\}$ where $k \geq \lceil \frac{n}{2} \rceil + 1$. Without loss of generality assume $x_1 < x_2 < ... < x_k$, then $\{x_k - x_1, x_k - x_2, ..., x_k - x_{k-1}\} \subseteq [n] \setminus A$, so there are at least $k - 1$ elements in $[n] \setminus A$. Hence there must be at least $\lceil \frac{n}{2} \rceil + 1 + \lceil \frac{n}{2} \rceil = 2 \lceil \frac{n}{2} \rceil + 1 > n$ elements in $[n]$. Contradiction. □

Let $SF(n)$ denote the number of sum-free sets in $[n]$. Consider a maximum sized sum-free set in $[n]$, such as the set $\mathcal{O}$ of odd integers. Since all subsets of $\mathcal{O}$ are also sum-free, we have $SF(n) \geq 2^{\lceil n/2 \rceil}$. Cameron and Erdös conjectured that $SF(n) = O(2^{n/2})$. Soon afterwards, Alon ([1]) proved a slightly weaker bound of $SF(n) \leq 2^{(1/2+o(1))n}$ by converting the question into a problem of counting independent sets in graphs.

We now introduce a theorem proved by Sapozhenko ([16], [17]) and derive Alon's bound from Sapozhenko's bound. We'll prove Sapozhenko's bound in Section 3.1.

Definition 5. *A graph G is d-regular if every vertex in $V(G)$ is of the degree d.*

Theorem 1. ([16], [17]) *There is a constant C so that all n-vertex d-regular graph G satisfy $i(G) \leq 2^{\left(1+c\sqrt{\log(d)/d}\right)\frac{n}{2}}$.*

Theorem 2. ([1]) *The set $[n] = \{1, 2, ..., n\}$ has at most $2^{(1/2+o(1))n}$ sum-free subsets.*

Proof. (of Theorem 2)

Let $n \in \mathbb{N}$ be given. The number of sum-free sets containing less than $n^{2/3}$ integers strictly smaller than $\frac{n}{2}$ (i.e. in $[\lceil \frac{n}{2} \rceil - 1]$) is small as we will now show.

Let $A \subseteq [n]$ be a sum-free set. We consider those sum-free sets containing less than $n^{2/3}$ integers in $[\lceil \frac{n}{2} \rceil - 1]$: the number of possibly combination of choices from $[\lceil \frac{n}{2} \rceil - 1]$ is less than $\left(\frac{n}{2}\right)^{n^{2/3}}$, and the total number of subsets of $[n] \setminus [\lceil \frac{n}{2} \rceil - 1]$ is $2^{n/2+1}$. Therefore we have:

$$\#\left\{A \subseteq [n] : \left|A \cap \left\{1, ..., \left\lceil \frac{n}{2} \right\rceil - 1\right\}\right| < n^{2/3}\right\} \leq \left(\frac{n}{2}\right)^{n^{2/3}} 2^{n/2+1} = 2^{\frac{n}{2}+o(n)}.$$

Consider those sets with at least $\left\lfloor n^{2/3} \right\rfloor$ elements from $[\lceil \frac{n}{2} \rceil - 1]$. For any such set X we define S_X as the smallest $\left\lfloor n^{2/3} \right\rfloor$ elements in X. Fix a sum-free set $S \subseteq [\lceil \frac{n}{2} \rceil - 1]$ of size $\left\lfloor n^{2/3} \right\rfloor$, we want to count the number of sum-free sets B with $S_B = S$, we denote the family of all such sum-free B by $\mathcal{B}(S)$.

Here we introduce a graph G_S and convert the above problem to a problem of counting independent sets:

$$V(G_S) = [n],\ E(G_S) = \{xy : \exists s \in (S \cup -S) \quad x + s \equiv y \mod n\}.$$

The graph is $2|S|$-regular as every $|s|$ is strictly smaller than $n/2$ so S and $-S$ provide different residue classes modulo n.

Claim. For all sum-free set $B \in \mathcal{B}(S)$, $B \setminus S$ is independent in G_S.

Proof of the claim. (by contradiction) Suppose not, then there will be integers $x, y \in B \setminus S$ and $s \in (S \cup -S)$ such that $x + s \equiv y \mod n$ by our definition of G_S.

Note that $x + s \equiv y \mod n$ implies $x + s = y - kn$ and then $kn = y - x - s$ for some integer k. Since $|s| \in S = S_B$, and S consists of the smallest $\lfloor n^{2/3} \rfloor$ elements in B, we have $1 \leq |s| < x, y \leq n$. So we have $1 \leq |y - x| \leq n - (|s| + 1)$ and $|y - x - s| \leq |y - x| + |s| \leq n - 1$. No matter $s \in S$ or $s \in -S$ we must have $k = 0$ so $x + s = y$. Thus B is not sum-free, contradiction. $\square$

Let $S \subseteq \{1, ..., \lceil n/2 \rceil - 1\}$ defined as before, then there are no more than $i(G_S)$ (i.e. number of independent sets in G_S) sum-free sets B satisfying $S_B = S$. There are at most $\binom{n/2}{n^{2/3}}$ choices of S and for each fixed S, we can add at most $i(G_S)$ different subsets to it to obtain a sum-free set with at least $\lfloor n^{2/3} \rfloor$ elements from $[\lceil \frac{n}{2} \rceil - 1]$. That is, number of sum-free sets with at least $\lfloor n^{2/3} \rfloor$ integers from $\{1, ..., \lceil \frac{n}{2} \rceil - 1\}$ is at most $\binom{n/2}{n^{2/3}} i(G_S)$.

Hence we need a suitable estimate of $i(G_S)$ to get the appropriate bound for number of sum-free sets. We apply Theorem 1 and take $d = 2\lfloor n^{2/3} \rfloor$.

Adding the number of sum-free sets with less than $\lfloor n^{2/3} \rfloor$ elements from $[\lceil \frac{n}{2} \rceil - 1]$ and those with at least $\lfloor n^{2/3} \rfloor$ elements from $[\lceil \frac{n}{2} \rceil - 1]$ together, we have

$$\begin{aligned}
SF(n) &\leq \left(\frac{n}{2}\right)^{n^{2/3}} 2^{n/2+1} + \binom{n/2}{n^{2/3}} 2^{1 + C\sqrt{\log\left(2\lfloor n^{2/3} \rfloor\right) / \left(2\lfloor n^{2/3} \rfloor\right)} \frac{n}{2}} \\
&\leq \left(\frac{n}{2}\right)^{n^{2/3}} 2^{n/2+1} + \binom{n/2}{n^{2/3}} 2^{\left(1 + O\left(n^{-1/3}\sqrt{\log n}\right)\right)\frac{n}{2}} \\
&\leq 2^{n^{2/3}\log_2(n/2) + n/2 + 1} + \left(\frac{n}{2}\right)^{n^{2/3}} 2^{\left(1 + O\left(n^{-1/3}\sqrt{\log n}\right)\right)\frac{n}{2}} \\
&= 2^{O\left(n^{-1/3}\log n\right)\frac{n}{2} + \frac{n}{2} + 1 - 1} + 2^{O\left(n^{-1/3}\log n\right)\frac{n}{2} + \left(1 + O\left(n^{-1/3}\sqrt{\log n}\right)\right)\frac{n}{2}} \\
&= 2^{\left(1 + O\left(n^{-1/3}\log n\right)\right)\frac{n}{2}} \\
&= 2^{\left(1/2 + O\left((\log n)/(n^{-1/3})\right)\right)n} \\
&= 2^{(1/2 + o(1))n}
\end{aligned}$$

as required. $\square$

1.2 Notation used in this project

In this project, unless otherwise stated, we use notations as follows:

$\#B := |B|$ i.e. the number of elements in B.

$\mathcal{P}(S) := \{S' : S' \subseteq S\}$.

$G :=$ graph with vertex set $V(G)$ and edge set $E(G)$.

$\deg_G(v, A) :=$ degree of v in the subgraph of G induced by vertex set A

$\Delta(G) := \max_{v \in V(G)} d(v)$, $\delta(G) := min_{v \in V(G)} d(v)$.

$e_G(A) := \#\{edges\ in\ G\ induced\ by\ vertex\ set\ A\}$.

$e_G(A, B) = \#\{edges\ between\ vertex\ sets\ A\ and\ B\}$.

$G[A] :=$ subgraph of G induced by $A \subseteq G$.

$G(v_1, v_2, ..., v_k) := G\,[A \setminus \{v_1, v_2, ..., v_k\}]$.

$I :=$ an independent set in G.

$\mathcal{I}(G) :=$ family of all independent sets in G.

$i(G) :=$ cardinality of $\mathcal{I}(G)$.

$\alpha(G) := \max_{I \in \mathcal{I}(G)} |I|$.

$i(G, m) := |\{I \in \mathcal{I}(G) : |I| = m\}|$ where $m \in \mathbb{N}$.

1.3 Upper bounds for independent sets

Counting independent sets for small order graphs is relatively easy, but we are also interested in large order graphs.

For a fixed but arbitrary graph G, we have the following inequality:

$$2^{\alpha(G)} \leq i(G) \leq \sum_{m=0}^{\alpha(G)} \binom{|V(G)|}{m}$$

Where the first inequality holds because all subsets of any largest independent set are independent. The second inequality holds because $i(G)$ is smaller than the number of all subsets with cardinality smaller than $\alpha(G)$.

As we noted earlier, any subset of an independent set is independent. Recall the method of proving the upper bound for counting sum-free sets in Section 1.1: we count sum-free sets with fixed 'smallest q' elements and sum over possible combinations of those smallest elements.

For counting independent sets in a more general sense, we may approach it by counting possible ways to extend from some fixed smallest independent elements and some over possible combinations of those elements. If we can describe all independent sets so that those independent sets with same 'smallest q' elements are in a same family and the subset consisting of those 'smallest q' elements is code-able, then by finding and encoding a collection of suitable subsets, we may derive a better upper bound of counting independent sets in that graph.

In the rest of this project, we will discuss an algorithm first introduced by Kleitman and Winston ([10, 11]), which provides a method for encoding each independent set of a graph in an invertible manner (so that we can recover the independent set from its unique code and hence derive an upper bound of counting independent sets from possible codes). Then with the KW-algorithm, we will give bounds on the number of independent sets for some particular graphs. We will also look at some interesting applications of the KW-algorithm as mentioned in the abstract.

Chapter 2

The Kleitman Winston Algorithm

In this chapter we describe the KW-algorithm for graphs and explain how it works.

2.1 The Algorithm

Definition 6. *Given $A \subseteq V(G)$, a max-degree-ordering of A is a total ordering with the property:*
Each vertex $w_i \in A$ is of maximum degree in the subgraph $G(w_1, ..., w_{i-1})$ induced by the vertex set $A \setminus \{w_1, ..., w_{i-1}\}$.

Now we introduce a method to find a max-degree-ordering of a vertex set A. Assume an initial total ordering is given. Among those vertices of maximum degree, choose the vertex which is first in the initial order, denote by w_1. Then delete w_1 from A and recursively apply this process to find the next vertex until all vertices are deleted from A. The max-degree-ordering w.r.t. given initial order is built by the order of the deleted vertices.

Formally, let $V = \{v_1, v_2, ..., v_n\}$ be the vertex set with initial order of a graph G and suppose $A \subseteq V$. We obtain the max-degree-ordering of A as follows:

(a_1) Build $G[A]$, find vertex $w_1 := v_{i_1} \in A$,
where $i_1 = \min\{j : \ v_j \in A \ and \ d_{G[A]}(v) = \Delta(G[A])\}$,

(b_1) Delete w_1, denote the resulting graph by $G[A_1]$;

When $G[A_{k-1}] = G[A \setminus \{w_1, ..., w_{k-1}\}]$ is given,

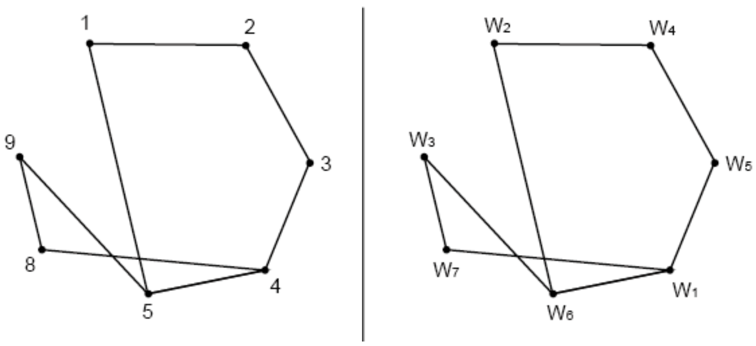
(a_k) Find vertex $w_k := v_{i_k} \in V(G[A_{k-1}])$,
where $i_k = \min\{j : \ v_j \in A \ and \ d_{G[A_{k-1}]}(v) = \Delta(G[A_{k-1}])\}$,

(b_k) Delete w_k, denote the result by $G[A_k]$.

Repeat until we find $w_{|A|}$.

Output: $(w_1, w_2, ..., w_{|A|})$ is the max-degree-ordering w.r.t. given initial order. □

Example 2. *Recall the graph we constructed in Example 1. We now find its max-degree-ordering using above method.*
Take $A = V = \{1, 2, 3, 4, 5, 8, 9\}$.



The left graph is of the initial order and the right graph gives the max-degree-ordering by $(w_1, w_2, ..., w_7)$. □

Then we can introduce the KW-algorithm. As mentioned in Section 1.3, we want to describe all independent sets so that those independent sets with a same fixed structure are in one family and the structure is code-able.

Idea of the KW-algorithm is to reorder the elements in an independent set I, then for a given integer q, find the unique code (w.r.t. the full vertex set) for the 'first q' elements, finally output this code and the subset of 'last $|I| - q$' elements in I.

Note that we should be able to recover the 'first q' elements from its code and union to recover the independent set. Also note that for every integer

q, independent sets with same ‘first q’ elements have the same outputting code. Therefore we can sort all independent sets into different families w.r.t. their ‘first q’ elements.

We begin with the algorithm.

Algorithm 1. *Given a graph G and an independent set $I \in \mathcal{I}(G)$, let an integer $q \le |I|$ also be given. Let $A = V(G)$ and $S = \emptyset$. Recursively applying the following steps for $s = 1, 2, ..., q$ (note that we reorder all remaining elements in A in the beginning of each round):*

(a) Let $(w_1, w_2, ..., w_{|A|})$ be the max-degree-ordering of A w.r.t. given initial order,

(b) Find $w_{j_s} \in A$ where $j_s = min\{j : w_j \in I\}$, move w_{j_s} from A to S,

(c) Delete vertices $w_{j_1}, ..., w_{j_s-1}$ and $N_G(w_{j_s}) \cap A$ from A.

Recursively applying (a) $\rightarrow$ (c) to the graph q times.

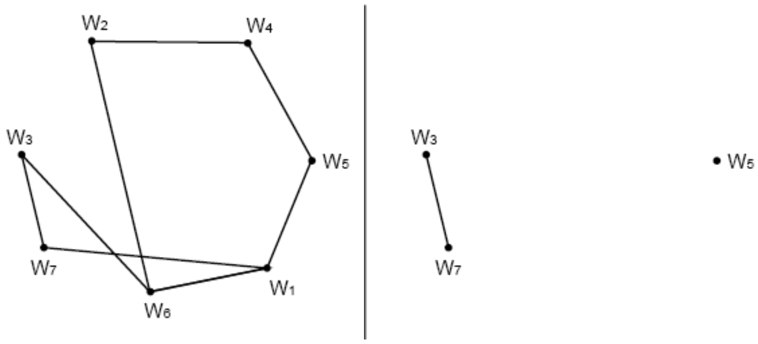
Output: $(j_1, j_2, ..., j_q)$ and $A \cap I$. $\square$

Example 3. *Then we apply the algorithm to the graph G constructed in Example 1. Take independent set $I = \{1, 3, 8\} = \{w_2, w_5, w_7\}$ and $q = 1$.*

Recall the max-degree-ordering obtained in Example 2, so $A = \{w_1, ..., w_7\}$ and $S = \emptyset$.

Follow the steps in the algorithm, $j_1 = 2$ so we first move w_2 from A to S, then delete vertices w_1 and $N_G(w_2) \cap A = \{w_4, w_6\}$ from A.

See below, the graph in left is the max-degree-ordered graph induced by vertex set $\{w_1, ..., w_7\}$ and the right one is $G[A]$ after step (c) i.e. the subgraph induced by $\{w_3, w_5, w_7\}$.

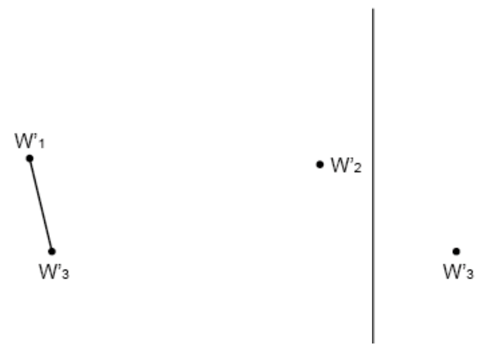


We stop here as $q = 1$. Output: $(j_1) = (2)$ and $A \cap I = \{w_5, w_7\}$. $\square$

Example 4. *Use the same graph constructed in Example 1. Independent set* $I = \{1, 3, 8\} = \{w_2, w_5, w_7\}$*,* $q = 2$*.*

The first round of the algorithm is as in Example 3 and we continue with $A = \{w_3, w_5, w_7\}$*,* $S = \{w_2\}$*. Reorder elements in* A *to get a max-degree-ordering:* $w'_1 = w_3, w'_2 = w_5, w'_3 = w_7$*.*

Here $j_2 = 2$ *so we move vertex* $w'_2 = w_5$ *from* A *to* S *and delete vertex* $w'_1 = w_3$ *and* $N_G(w'_2) \cap A = \emptyset$ *from* A*.*



Here the left graph is the max-degree-ordered graph induced by vertex set $\{w'_1, w'_2, w'_3\}$ *and the right one shows* $G[A]$ *after step (c), which is the graph induced by* $\{w'_3\}$*.*

We stop here as $q = 2$*. Output:* $(j_1, j_2) = (2, 2)$ *and* $A \cap I = \{w'_3\} = \{w_7\}$*.*
□

If $q = 3$ *and take the same independent set, we will output* $(j_1, j_2, j_3) = (2, 2, 1)$ *and* $A \cap I = \emptyset$*.* □

Denote the set A after the s^{th} round by $A(j_1, j_2, ..., j_s)$ and the set S after the s^{th} round by $S(j_1, j_2, ..., j_s)$. Note that $A(j_1, j_2, ..., j_s)$ consists of vertices we can add back to $S(j_1, j_2, ..., j_s)$ and keep it independent and will not change the order of 'first s' elements. In the above Example 3 and Example 4, we see that $A(2) = \{w_3, w_5, w_7\}$, $A(2, 2) = \{w'_3\} = \{w_7\}$ and $S(2) = \{w_2\}$, $S(2, 2) = \{w_2, w'_2\} = \{w_2, w_5\}$. Moreover $A(2, 2, 1) = \emptyset$, $S(2, 2, 1) = \{w_2, w_5, w_7\}$.

2.2 How the KW-algorithm works

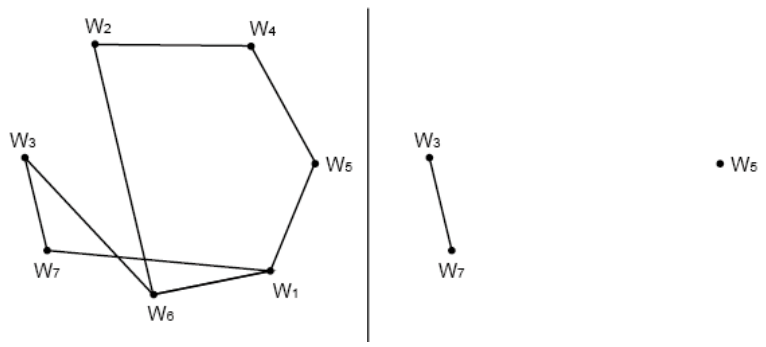
We can recover the original independent set from our output. In particular, $S(j_1, ..., j_q)$ provides the 'first q' elements in I w.r.t. the max-degree-ordering and $A(j_1, ..., j_q)$ consists of vertices that are not adjacent to any vertices in S. So the remaining $|I| - q$ elements of I belongs to $A(j_1, ..., j_q)$ and we have

$$I = S(j_1, ..., j_q) \cup (A(j_1, ..., j_q) \cap I).$$

We see $S(j_1, ..., j_q)$ and $A(j_1, ..., j_q)$ are uniquely determined by $(j_1, ..., j_q)$. Moreover, we can reconstruct the subset of 'smallest q' elements $S(j_1, ..., j_q)$ and the subset of remaining possible elements $A(j_1, ..., j_q)$ from the sequence $(j_1, ..., j_q)$.

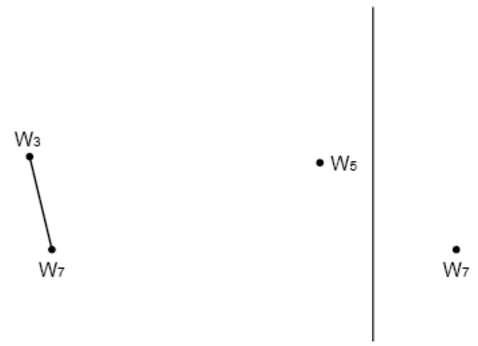
Example 5. *Given sequence* $(2, 2)$ *and the graph* G *in Example 1, we now try to recover the relevant subsets* S *and* A *based on it. The basic idea is to go through the KW-algorithm, but rather than finding the vertex in a given independent set who appears first in the max-degree-ordering, we choose the vertex in our max-degree-ordering from the order indicated in the sequence.*

We start with $A = \emptyset$ *and* $S = \emptyset$. *We first find the max-degree-ordering w.r.t. the initial ordering (as mentioned in Example 2).*



Here $j_1 = 2$ *so we choose the second element in the max-degree-ordering, i.e.* w_2. *So we move* w_2 *from* $V(G)$ *to* S *and delete vertices before* w_2 *(i.e.* w_1*) and* $N_G(w_2) = \{w_4, w_6\}$ *from* $V(G)$.

Then we find the graph induced by $V(G) \setminus \{w_1, w_2, w_4, w_6\}$ and find the max-degree-ordering of $\{w_3, w_5, w_7\}$ with the original initial order, which is still $V(G') = \{w_3, w_5, w_7\}$, i.e. order unchanged.



Here $j_2 = 2$ so we move the second vertex from $V(G')$ to S, which is w_5, and delete vertices before w_5 (i.e. w_3) and $N_G(w_5) = \emptyset$ from $V(G')$. So we have $V(G'') = \{w_7\}$.

As the sequence ends here, we output $S = \{w_2, w_5\}$ and $A = V(G'') = \{w_7\}$.
$\square$

Recall the KW-algorithm.
Note that $j_1 + j_2 + ... + j_q \leq |V(G)| - |A(j_1, j_2, ..., j_q)|$ because we remove at least j_s elements from A in the s^{th} round of the algorithm.

By the definition of $i(G, m)$ and the algorithm, for all $m, q \in \mathbb{N}$ such that $m \geq q$ we have

$$i(G, m) \leq \sum_{(j_s)} i\left(G\left[A(j_1, j_2, ..., j_q), m - q\right]\right) \leq \sum_{(j_s)} \binom{|A(j_1, j_2, ..., j_q)|}{m - q}. \quad (2.1)$$

The first inequality holds because $A(j_1, j_2, ..., j_q)$ contains all vertices not adjacent any vertex in $S(j_1, j_2, ..., j_q)$, and summing over all possible sequences (j_s) ensures all possible independent sets were counted by the middle formula. The second inequality holds trivially: it just indicates that the number of all possible subsets of $A(j_1, j_2, ..., j_q)$ with cardinality $m - q$ is larger than the number of independent sets in $A(j_1, j_2, ..., j_q)$ with cardinality $m - q$.

Let $n = |V(G)|$, we will treat independent sets in G with different cardinality separately. In particular, we first consider independent sets I with $|I| < q$ and then consider I with $|I| \geq q$. Then we have

$$i(G) \leq \sum_{m=0}^{q-1} \binom{n}{m} + \sum_{(j_s)} i\left(G[A(j_1, j_2, ..., j_q)]\right) \leq \sum_{m=0}^{q-1} \binom{n}{m} + \sum_{(j_s)} 2^{|A(j_1, j_2, ..., j_q)|}.$$

Where the middle formula is the number of all subsets of size 0 to $q-1$ plus number of possible independent sets with at least q elements (by summing $i(G, m)$ over m), so the first inequality holds. The second inequality holds because the most right formula counts all subsets of $A(j_1, j_2, ..., j_q)$.

We claim that each round in the loop of the KW-algorithm shrinks $|A|$ by some factor between 0 and 1.

To see this, let $A^- = A(j_1, ..., j_{s-1}) \setminus \{v_1, ..., v_{j_s-1}\}$. Denote $deg_G(v, A^-) :=$ $deg(v)$ in the subgraph of G induced by A^-, hence

$$|N_G(v_{j_s}) \cap A^-| = \max_{v \in A^-} deg_G(v, A^-) \geq \frac{2e_G(A^-)}{|A^-|}.$$

The equality holds because of the max-degree-ordering, and the inequality part holds as the maximum degree is larger or equal to average degree in A^-.

Find $\beta \in (0, 1]$ so that $e_G(A^-) = \beta\binom{|A^-|}{2}$ then

$$\frac{2e_G(A^-)}{|A^-|} = \beta(|A^-| - 1).$$

As v_{j_s} is of the maximum degree in A^-, we have $|N_{A^-}| \leq \beta(|A^-| - 1)$.

As we remove vertices $v_1, ..., v_{j_s}$ and $N_{A^-}(v_{j_s})$ from A in the s^{th} round of the algorithm, and $|A^-| = |A(j_1, ..., j_{s-1})| - (j_s - 1)$, we hence have

$$|A(j_1, ..., j_{s-1})| - |A(j_1, ..., j_s)| \geq j_s + \beta(|A^-| - 1) \geq \beta|A(j_1, ..., j_{s-1})|.$$

Thus $|A(j_1, ..., j_s)| \leq (1-\beta)|A(j_1, ..., j_{s-1})|$ and we conclude that each round of the main loop shrinks A by a factor of at most $1 - \beta$.

Lemma 2. ([12]) *Let G be a graph with $|V(G)| = n$, and assume $q \in \mathbb{Z}$, $R \in \mathbb{R}$, $\beta \in (0, 1]$ with $R \geq e^{-\beta q}n$. Suppose that $e_G(U) \geq \beta\binom{|U|}{2}$ for all subset $U \subseteq V(G)$ with $|U| \geq R$. Then for all integer $m \geq q$ we have*

$$i(G, m) \leq \binom{n}{q}\binom{R}{m-q}.$$

Proof. Note that $\binom{n}{q}$ counts all possible sequences for $(j_1, j_2, ..., j_q)$ (as $j_1 + j_2 + ... + j_q \leq n$, we consider it as inserting q gaps between 1 and n). If we have

$$|A(j_1, ..., j_{s-1})| \leq R \quad (*),$$

then the Lemma follows from the inequality (2.1). Now we prove $(*)$.

(By contradiction) Suppose we have a sequence $(j_1, j_2, ..., j_q)$ such that

$$|A(j_1, ..., j_{s-1})| > R,$$

then by assumptions in the lemma we have

$$e_G(A(j_1, ..., j_{s-1})) \geq \beta \binom{|A(j_1, ..., j_{s-1})|}{2}.$$

As all vertices in A are max-degree-ordered, denote $A^- = A(j_1, ..., j_{s-1}) \setminus \{v_1, ..., v_{j_s-1}\}$ as before, we have

$$e_G(A^-) \geq \beta \binom{|A^-|}{2}$$

holds for all s. Therefore the algorithm shrinks A by a factor of at most $1 - \beta$ in each round. Then

$$|A(j_1, j_2, ..., j_q)| \leq (1 - \beta)^q n \leq e^{-\beta q} n \leq R,$$

which contradicts to $|A(j_1, ..., j_{s-1})| > R$. Thus there is no such sequence and lemma is proved. □

Lemma 3. ([5, 15]) *Let G be a graph with $|V(G)| = n$, assume $q \in \mathbb{Z}$, $R, D \in \mathbb{R}$ and they satisfy $R + qD \geq n$. Suppose $2e_G(U) \geq D|U|$ for all subset $U \subseteq V(G)$ with $|U| \geq R$. Then there is a collection $\mathcal{S}$ of q-element subsets of $V(G)$ and two well-defined mappings:*

$$g : \mathcal{I} \to \mathcal{S}, \quad f : \mathcal{S} \to \mathcal{P}(V(G)),$$

so that $|f(S)| \geq R$ for all $S \in \mathcal{S}$ and

$$g(I) \subseteq I \subseteq (f(g(I)) \cup g(I)) \quad \forall (I \in \mathcal{I}(G) \text{ with } |I| \geq q). \tag{2.2}$$

Proof. We will obtain such mappings by running the KW-algorithm over I. In particular, define

$$g(I) = S(j_1, ..., j_q), \quad f(g(I)) = A(j_1, ..., j_q)$$

and then $\mathcal{S} = \{g(I) : I \in \mathcal{I}\}$. By this definition, we will have the formula (2.2). It suffices to check $|f(g(I))| \leq R$.

(By contradiction) Assume there is a sequence $(j_1, ..., j_q)$ so that

$$|A(j_1, ..., j_q)| > R,$$

therefore

$$|A^-| = |A(j_1, ..., j_{s-1}) \setminus \{v_1, ..., v_{j_s-1}\}| > R, \quad s = 1, 2, ..., q.$$

Then as the average degree of A^- is at least D, we have $2e_G(A^-) \geq D|A^-|$. So each round of the algorithm removes at least $D+1$ elements from A (the vertex itself plus its neighborhood). Therefor we have

$$|A(j_1, ..., j_q)| \leq n - Dq \leq R,$$

which contradicts to our assumption. So no such sequence exist and Lemma proved. $\square$

As the conclusion of this section, we can see that Lemma 3 is stronger than Lemma 2. Because the existence of f and g implies the bounds of $i(G, m)$ in Lemma 2. We also note the assumptions of those lemmas are interchangeable (i.e. if the q, R-pair works for one lemma, it works for another).

Chapter 3

Applications

In this chapter, we will discuss some applications of the KW-algorithm.

3.1 Independent sets in regular graphs

Granville conjectured ([1]) that, for an n-vertex d-regular graph G we have $i(G) \leq 2^{(1+o(1))n/2}$ where $o(1) \to 0$ as $d \to \infty$. A few years later, Alon proved that actually $i(G) \leq 2^{(1+O(d^{-0.1}))n/2}$ for n-vertex d-regular graph G a few years later.

Theorem 3. ([8, 10]) *For every n-vertex d-regular G we have*

$$i(G) \leq i(K_{d,d})^{n/2d} = (2^{d+1} - 1)^{n/2d}.$$

Recall the Theorem 1 by Sapozhenko we mentioned in Section 1.1. We will use Theorem 3 above to prove Theorem 1.

Theorem 1. There is a constant C so that any n-vertex d-regular graph G satisfies $i(G) \leq 2^{\left(1+c\sqrt{\log(d)/d}\right)n/2}$.

Proof. Let G be an n-vertex d-regular graph. We determine the bound of $i(G)$ by estimating $i(G, m)$ and summing over m. We assume d and n are sufficiently large and consider two cases.

Case 1: $m \leq n/10$.

As for all a and b we have

$$\binom{a}{b} = \frac{a!}{(a-b)!b!} \leq \frac{a^b}{b!} \leq \left(\frac{a}{b}\right)^b \left(\frac{b^b}{b!}\right) \leq \left(\frac{a}{b}\right)^b \sum_{n=0}^{\infty} \frac{b^n}{n!} = \left(\frac{ea}{b}\right)^b,$$

we see $\binom{a}{b} \leq (ea/b)^b$ is true for all a and b. Hence we have

$$i(G,m) \leq \binom{n}{n/10} \leq (10e)^{n/10} \leq 2^{0.48n}. \tag{3.1}$$

Case 2: $m > n/10$, let $B \subseteq V(G)$ and note that

$$\begin{aligned} d|B| = \sum_{v \in B} deg(v) = 2e(B) + e(B, B^c) &\leq 2e(B) + \sum_{v \in B^c} deg(v) \\ &= 2e(B) + d(n - |B|). \end{aligned}$$

That is, total degree of $v \in B$ in the original graph G is twice of edges induced by $v \in B$ plus number of edges between B and its complement. The number between B and its complement is at most the total degree of vertices in B^c, which is $d(n - |B|)$.

Note that $d|B| \leq 2e_G(B) + d(n - |B|)$ is equivalent to $e(B) \geq d|B| - dn/2$.

Then we apply Lemma 2. Assume $\beta > 10/n$ and let $q = \lceil 1/\beta \rceil$, let $R = \frac{n}{2} + \frac{\beta n^2}{2d}$, for $|B| \geq R$ we have

$$e(B) \geq \frac{d}{2}(2|B| - n) \geq \frac{d}{2}(2R - n) \geq \frac{\beta n^2}{2} \geq \beta \binom{|B|}{2},$$

and

$$e^{-\beta q} n \leq e^{-1} n \leq R.$$

Hence by Lemma 2, for every m with $m \geq \lceil n/10 \rceil \geq q$,

$$\begin{aligned} i(G,m) \leq \binom{n}{q} \binom{\frac{n}{2} + \frac{\beta n^2}{2d}}{m-q} &\leq \left(\frac{en}{q}\right)^q \binom{\frac{n}{2} + \frac{\beta n^2}{2d}}{m-q} \\ &\leq (e\beta n)^{\lceil 1/\beta \rceil} \binom{\frac{n}{2} + \frac{\beta n^2}{2d}}{m-q}. \end{aligned} \tag{3.2}$$

Then we sum (3.1) and (3.2) together over m to get:

$$\begin{aligned}
i(G) &= \sum_{k\le\lfloor n/10\rfloor} 2^{o.48n} + \sum_{k\ge\lceil n/10\rceil} (e\beta n)^{\lceil 1/\beta\rceil} \binom{\frac{n}{2}+\frac{\beta n^2}{2d}}{k-q} \\
&\le 2^{0.49n} + (e\beta n)^{\lceil 1/\beta\rceil} \sum_{k\ge\lceil n/10\rceil} \binom{\frac{n}{2}+\frac{\beta n^2}{2d}}{k-q} \\
&\le 2^{0.49n} + (e\beta n)^{\lceil 1/\beta\rceil} \sum_{x=0} \binom{\frac{n}{2}+\frac{\beta n^2}{2d}}{x} \\
&\le 2^{0.49n} + 2^{\lceil 1/\beta\rceil \log_2(e\beta n)} 2^{\frac{n}{2}+\frac{\beta n^2}{2d}} \\
&= 2^{0.49n} + 2^{\lceil 1/\beta\rceil \log_2(e\beta n)+\frac{n}{2}+\frac{\beta n^2}{2d}}.
\end{aligned}$$

Let $\beta = \frac{\sqrt{d\log(d)}}{n}$ (we have $\sqrt{d\log(d)} > 10$ for large d, so $\beta > 10/n$), then

$$\begin{aligned}
&\frac{n}{2} + \frac{\beta n^2}{2d} + \lceil 1/\beta\rceil \log_2(e\beta n) \\
&= \frac{n}{2} + \frac{n\sqrt{d\log(d)}}{2d} + \left\lceil \frac{n}{\sqrt{d\log(d)}} \right\rceil \log_2\left(e\sqrt{d\log(d)}\right) \\
&\le \frac{n}{2}\Big(1+\sqrt{\frac{\log(d)}{d}}\Big) + \left(\frac{n}{\sqrt{d\log(d)}}+1\right) \log_2\left(e\sqrt{d\log(d)}\right) \\
&\le \frac{n}{2}\Big(1+\sqrt{\frac{\log(d)}{d}}\Big) + \left(\frac{2n}{\sqrt{d\log(d)}}\right) \log_2\left(e\sqrt{d\log(d)}\right) \\
&= \frac{n}{2}\left(1+\sqrt{\frac{\log(d)}{d}}\left(1+\frac{4}{\log(d)}\log_2\left(e\sqrt{d\log(d)}\right)\right)\right).
\end{aligned}$$

I.e. $i(G) \le 2^{\left(1+C\sqrt{\log(d)/d}\right)\frac{n}{2}}$ for constant $C = 1 + \frac{4}{\log(d)}\log_2\left(e\sqrt{d\log(d)}\right)$. $\square$

3.2 Independent sets in regular graphs without small eigenvalues

Definition 7. *A graph G is bipartite if there is a partition of $V(G) = A\cup B$ such that A and B are independent sets. A graph G is complete bipartite if $\forall x \in A$, $y \in B$, $xy \in E(G)$, in that case, we also denote G by $K_{|A|,|B|}$.*

For a bipartite graph G its clear that $\alpha(G) \ge n/2$, hence by taking subsets of its largest independent set, $V(G)$ contains at least $2^{n/2}$ independent subsets.

Therefore the bound proved in Theorem 1 is essentially best possible for bipartite G. Its natural to ask: for graphs which are far from bipartite, can we improve the upper bound of counting independent sets?

In this section we introduce an approach to count independent sets in regular graphs using the adjacency matrix, which is also an application of Lemma 2.

Definition 8. *Let a graph G be given, assume $V(G) = \{v_1, v_2, ..., v_n\}$. The adjacency matrix M_G of G is $M_G = (m_{i,j})_{1 \leq i,j \leq n}$ where $m_{i,j} = 1$ if $v_i v_j \in E(G)$ and $m_{i,j} = 0$ if $v_i v_j \notin E(G)$.*

Note that for an n-vertex graph G, M_G is a symmetric real valued $n \times n$ matrix hence it has n real eigenvalues. Denote these eigenvalues by $\lambda_1, \lambda_2, ..., \lambda_n$ where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$. If G is d-regular then the largest eigenvalue of M_G is d with eigenvector $(1, 1, ..., 1)^T$. Denote $\lambda := \max\{|\lambda_2|, |\lambda_n|\}$ as the largest absolute value of all other eigenvalues, and $\lambda(G) := \lambda_n$ as the smallest eigenvalue. So $\lambda(G) \geq -\lambda$.

It was first proved by Hoffman ([7]) that $\alpha(G) \leq \frac{-\lambda(G)}{d-\lambda(G)}$ and later strengthened by Alon and Chung ([3]), who established the following relationship.

Lemma 4. *For an n-vertex d-regular graph G and for all $A \subseteq V(G)$ we have $2e(A) \geq \frac{d}{n}|A|^2 - \frac{\lambda}{n}|A|(n - |A|)$.*

For graphs that are far from bipartite (not bipartite even if we remove a small number of edges), Alon and Rödl ([4]) first proved that if graph G satisfies $\lambda(G) >> -d$, then it has far fewer than $2^{n/2}$ independent sets. We will prove the following theorem (which gives a similar estimate as in [4]) using Lemma 4 and Lemma 2. It was originally proved in [2].

Theorem 4. *Let G be an n-vertex d-regular graph. For all positive ϵ, there is a constant C so that if $m \geq Cn/d$ then $i(G, m) \leq \binom{(\frac{\lambda}{d+\lambda}+\epsilon)n}{m}$.*

Proof. (of Theorem 4) Let G be an n-vertex d-regular graph, fix a positive ϵ and λ is defined as above. If $\frac{\lambda}{d+\lambda} + \epsilon \geq 1$ then there is nothing to prove as clearly $i(G, m) \leq \binom{n}{m}$. So we can assume $\frac{\lambda}{d+\lambda} + \epsilon < 1$.

Choose an arbitrary subset $U \subseteq V(G)$ so that $|U| \geq \left(\frac{\lambda}{d+\lambda} + \frac{\epsilon}{2}\right) n$. According to Lemma 4 we have

$$\begin{aligned}
2e(U) &\geq \frac{d}{n}|A|^2 - \frac{\lambda}{n}|A|(n-|A|) \\
&= \frac{|U|}{n}((d+\lambda)|U| - \lambda n) \\
&\geq \frac{|U|}{n}\left((d+\lambda)\left(\frac{\lambda}{d+\lambda} + \frac{\epsilon}{2}\right) n - \lambda n\right) \\
&\geq \frac{\epsilon d}{2}|U| \geq \epsilon d \frac{|U|}{2}\frac{|U|-1}{n} = \frac{\epsilon d}{n}\binom{|U|}{2}.
\end{aligned}$$

Next we apply Lemma 2. Let $\beta = \frac{\epsilon d}{2n}$, $q = \left\lceil \frac{\log(2/\epsilon)}{\epsilon}\frac{2n}{d} \right\rceil$, $R = (\frac{\lambda}{d+\lambda} + \frac{\epsilon}{2})n$. So $e^{-\beta q} \leq \frac{\epsilon}{2} \leq R$ and we have $i(G,m) \leq \binom{n}{q}\binom{R}{m-q}$ for every $m \geq q$.

Note that $m \leq \alpha(G) \leq \frac{-\lambda(G)}{d-\lambda(G)}n \leq \frac{\lambda}{d+\lambda}n$ as $-\lambda(G) \leq \lambda$. Denote $r(t) = \binom{n}{t}\binom{R}{m-t}$ then $i(G,m) \leq r(q)$ and note that

$$\frac{r(t+1)}{r(t)} = \binom{n}{t+1}\binom{R}{m-(t+1)}\binom{n}{t}^{-1}\binom{R}{m-t}^{-1} \leq \frac{2m}{\epsilon(t+1)}.$$

Hence by the inequalities $a! > (a/e)^a$ and $\binom{a}{c} \geq (a/b)^c\binom{b}{c}$ (which hold for all $a \geq b \geq c \geq 0$) we have

$$\begin{aligned}
i(G,m) \leq r(q) = \prod_{t=0}^{q-1}\frac{r(t+1)}{r(t)}r(0) &\leq \frac{(2m)^q}{\epsilon^q q!}\binom{R}{m} \\
&\leq \left(\frac{2em}{\epsilon q}\right)^q \left(\frac{R}{R+\epsilon n/2}\right)^m \binom{R+\epsilon n/2}{m}.
\end{aligned}$$

Note that

$$\sqrt[m]{\left(\frac{2em}{\epsilon q}\right)^q \left(\frac{R}{R+\epsilon n/2}\right)^m} = \left(\frac{2em}{\epsilon q}\right)^{q/m}\frac{R}{R+\epsilon n/2},$$

and

$$\frac{R}{R+\epsilon n/2} = \frac{\lambda/(d+\lambda)+\epsilon/2}{\lambda/(d+\lambda)+\epsilon} = 1 - \frac{\epsilon}{2(\lambda/(d+\lambda)+\epsilon)} \leq 1 - \frac{\epsilon}{2}.$$

Find sufficiently large K (depending on ϵ) and constant $C \geq K\left\lceil \frac{2\log(2/\epsilon)}{\epsilon}\right\rceil$. For $m \geq Cn/d \geq Kq$ we have

$$\left(\frac{2em}{\epsilon q}\right)^{q/m} \leq \left(\frac{2Ke}{\epsilon}\right)^{1/K},$$

so

$$\left(\frac{2em}{\epsilon q}\right)^q \left(\frac{R}{R+\epsilon n/2}\right)^m \leq \left(\left(\frac{2Ke}{\epsilon}\right)^{1/K}\left(1-\frac{\epsilon}{2}\right)\right)^m \leq 1$$

and finally,

$$i(G,m) \leq \binom{R+\epsilon n/2}{m} = \binom{\left(\frac{\lambda}{d+\lambda}+\epsilon\right)n}{m}$$

as required. □

Now we prove Lemma 4.

Proof. For a given n-vertex d-regular graph G, let $A \subseteq V(G)$. Let M_G denote its adjacency matrix and $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$ be its second eigenvalue. Define $f : V(G) \to R$ by $f(v_i) = -1/|A|$ if $v_i \in A$ and $f(v_i) = -1/(n-|A|)$ if $v_i \notin A$. As $\sum_{i=1}^n f(v_i) = 0$, the vector $(f_i) = (f(v_i))$ is orthogonal to the all 1 vector, the eigenvector of the largest eigenvalue of M_G. Hence (f_i) is the eigenvector of another eigenvalue of M_G, which is at most λ. We conclude that $|M_G(f_i)(f_i)| \leq \lambda(f_i)(f_i)$. Note that

$$\begin{aligned}
M_G(f_i)(f_i) &= 2\sum_{v_iv_j\in E(G)} f(v_i)f(v_j) \\
&= \sum_{v_iv_j\in E(G)} \left(f(v_i)^2 + f(v_j)^2 - (f(v_i)-f(v_j))^2\right) \\
&= d\sum_{i=1}^n f(v_i)^2 - \sum_{v_iv_j\in E(G)} (f(v_i)-f(v_j))^2 \\
&= d\left(\frac{1}{|A|}+\frac{1}{n-|A|}\right) - e(A, V(G)\setminus A)\left(\frac{1}{|A|}+\frac{1}{n-|A|}\right)^2.
\end{aligned}$$

So we have

$$\begin{aligned}
\left|d\left(\frac{1}{|A|}+\frac{1}{n-|A|}\right) - e(A, V(G)\setminus A)\left(\frac{1}{|A|}+\frac{1}{n-|A|}\right)^2\right| &\leq \lambda(f_i)(f_i) \\
&\leq \lambda\left(\frac{1}{|A|}+\frac{1}{n-|A|}\right),
\end{aligned}$$

thus

$$\left|e(A, V(G)\setminus A) - \frac{d}{n}|A|(n-|A|)\right| \leq \lambda\left(\frac{1}{|A|}+\frac{1}{n-|A|}\right).$$

Since G is d-regular we have $2e(A) + e(A, V(G)\setminus A) = d|A|$ so $e(A, V(G)\setminus A) = d|A| - 2e(A)$, therefore

$$\left|d|A| - 2e(A) - \frac{d}{n}|A|(n-|A|)\right| \leq \lambda\left(\frac{1}{|A|}+\frac{1}{n-|A|}\right).$$

I.e. $\left|2e(A) - \frac{d}{n}|A|^2\right| \leq \frac{\lambda}{n}|A|(n - |A|)$.

So $2e(A) \geq \frac{d}{n}|A|^2 - \frac{\lambda}{n}|A|(n - |A|)$ as required. $\square$

3.3 Counting C_4-free graphs

In this section we introduce another application of Lemma 2. This is the main result of the paper originally introduced the KW-algorithm ([11]).

Definition 9. *A graph G is C_4-free if it does not contain any cycle of length 4. Denote $ex(n, C_4)$ as the maximum number of edges in an n-vertex C_4-free graph. Denote $f_n(C_4)$ as the number of (labeled) C_4-free graphs on vertex set $\{1, 2, ..., n\}$.*

As the subgraph of any C_4-free graph is C_4-free, we have

$$2^{ex(n,C_4)} \leq f_n(C_4) \leq \sum_{m=0}^{ex(n,C_4)} \binom{\binom{n}{2}}{m} = 2^{\Theta(ex(n,C_4)\log n)}$$

and hence

$$ex(n, C_4) \leq \log_2 f_n(C_4) \leq O(ex(n, C_4) \log n).$$

Deriving from a classical result of Kövári, Sós, and Turán ([9]) and a construction by Brown ([6]) implies the following:

$$ex(n, C_4) = \left(\frac{1}{2} + o(1)\right) n^{3/2}.$$

Erdös asked whether $\log_2 f_n(G) = (1 + o(1))ex(n, G)$. Later in the following theorem, Kleitman and Winston ([11]) partly answered the question by showing an upper bound of $\log_2 f_n(C_4)$.

Theorem 5. *There is a positive constant C such that* $\log_2 f_n(C_4) \leq Cn^{3/2}$.

Proof. Let $G_i = G[\{v_1, ..., v_i\}]$, we construct an n-vertex graph G in the following way:

- Let $V(G_1) = \{v_1\}$, $E(G_1) = \emptyset$.
- Attach v_i to G_{i-1} recursively where $deg_{G_i}(v_i) \leq \delta(G_{i-1}) - 1$ for $2 \leq i \leq n$.

According to our ordering above we have $\delta(G_{i-1}) \geq deg_{G_i}(v_i) - 1$ and note that G is C_4-free if and only if G_i is C_4-free for all i.

Let integers d and i be given with $d \leq i$, denote $g_i(d)$ as the number of ways to attach a d-degree vertex v_{i+1} to G_i where both of them are C_4-free. We have $g_i(d) \leq \binom{i}{d}$. Denote $g_i = \max\{g_i(d) : d \leq i\}$, as there are $n!$ ways to order $[n]$ and for each ordering there are $n!$ choices of $d_2, d_3, ..., d_n$, we have $f_n(C_4) \leq (n!)(n!) \prod_{i=2}^{n} g_i$. We now introduce a claim from which we will deduce the theorem.

Claim. There exists a constant D such that $g_n \leq \exp(D\sqrt{n})$ for all n.

Proof of the claim. Wlog assume n is large. For the case $d \leq \sqrt{n}/\log n$ we have

$$g_n(d) \leq \binom{n}{d} \leq \binom{n}{\sqrt{n}/\log n} \leq (e\sqrt{n}\log n)^{\sqrt{n}/\log n} \leq \exp\sqrt{n}.$$

From now we assume $d < \sqrt{n}/\log n$. Let G be an n-vertex C_4-free graph with $\delta(G) \geq d-1$ and we construct graph H by $V(H) = V(G)$ and $E(H) = \{xy : xz, yz \in E(G) \ for \ some \ z \in V(G)\}$. Note that adding v to G will give a C_4-free graph if and only if $N_H(v)$ is an independent set. Hence the number of C_4-free extensions with d-degree vertex of graph G is upper bounded by $i(H, d)$. We apply Lemma 2 to estimate $i(H, d)$.

We first show that a subgraph of H has a high density of edges if it is induced by a large subset of vertices. For each $B \subseteq V(H)$ we have

$$\sum_{z \in V(G)} \deg_G(z, B) = \sum_{x \in B} \deg_G(x) \geq |B|\delta(G) \geq (d-1)|B|.$$

As G is C_4-free, any edges xy in H corresponds to a unique vertex z where $xz, yz \in E(G)$, we have

$$e_H(B) = \sum_{z \in V(G)} \binom{deg_G(z, B)}{2} \geq n \binom{\sum_{z \in V(G)} \deg(z, B)/n}{2}.$$

So if $|B| \geq \frac{2n}{d-1}$, we have

$$e_H(B) \geq n\frac{(d-1)|B|}{2n}\left(\frac{(d-1)|B|}{n} - 1\right) \geq \frac{(d-1)^2}{2n}\binom{|B|}{2}.$$

Finally we use Lemma 2. Let $R = \frac{2n}{d-1}$, $\beta = \frac{(d-1)^2}{2n}$, and $q = \lceil 3(\log n)^3 \rceil$. $\beta q \geq \log n$ as $d > \sqrt{n}/\log n$ and n is large, and hence $e^{-\beta q} n \leq 1 \leq R$.

Therefore follow from Lemma 2 and $\sup\{(e/x)^x : x > 0\} = e$ we have

$$\begin{aligned} i(H,d) \leq \binom{n}{q}\binom{\frac{2n}{d-1}}{d-q} \leq \exp(4\log^4 n)\left(\frac{2en}{(d-q)^2}\right)^{d-q} &\leq \sup_{k>0}\left(\frac{e\sqrt{n}}{k}\right)^{2k} \\ &= e^{2\sqrt{n}}. \end{aligned}$$

□

Thus $\log_2 f_n(C_4) \leq \log_2((n!)^2 \exp(D(\sqrt{n-1}+\cdots+\sqrt{1}))) \leq Cn^{3/2}$ for some constant C. □

3.4 Roth's theorem in integer sets

In this section we will use Lemma 3 to prove the existence of 3-term arithmetic progression in large integer sets.

Definition 10. *We say a subset A of integers is δ-Roth for a positive δ if every subset $B \subseteq A$ with at least $\delta|A|$ elements contains a 3-term arithmetic progression.*

Theorem 6. (Roth's Theorem [14]) *For each positive δ, there exists an integer n_0 such that $[n]$ is $\delta -$ Roth for all $n > n_0$.*

The following result was proved by Kohayakawa, Luczak, and Rödl ([13]), which shows the existence of smaller and sparser δ-Roth sets.

Theorem 7. *For each positive δ, there is a constant C such that if $C\sqrt{n} \leq m \leq n$, then $P($uniformly chosen random m-element subset of $[n]$ is δ-Roth$)$ converges to 1 as n goes to ∞.*

Theorem 7 can be deduced from the following theorem, which is derived from Roth's theorem and Lemma 3.

Theorem 8. ([5, 18]) *For each positive ϵ, there is a constant D such that $D\sqrt{n} \leq m \leq n \Rightarrow \#\{A \subseteq [n] : |A| = m,\ A \text{ contains no 3-term AP}\} \leq \binom{\epsilon n}{m}$.*

We first prove Theorem 7 using Theorem 8.

Proof. Let a positive δ be given, and let $\epsilon = \delta/6$. Find constant D as described in Theorem 8. Let $C = Ddelta$ and $C\sqrt{n} \leq m \leq n$. As $\lceil \delta m \rceil \geq \delta C\sqrt{n} \geq D\sqrt{n}$ and $C\sqrt{n} \leq n$, Theorem 8 implies that

$$\#\mathcal{A} =: \#\{A \subseteq [n] : |A| = \lceil \delta m \rceil\ and\, A\, contains\, no\, 3-term\, AP\} \leq \binom{\epsilon n}{\lceil \delta m \rceil}.$$

Chose $R \subset [n]$ uniformly randomly with $|R| = m$, then

$$\begin{aligned}
P(R\,is\,not\,\delta - Roth) &= P(R \supseteq A\,for\,some\,A \in \mathcal{A}) \\
&\leq \sum_{A \in \mathcal{A}} P(R \supseteq A) \\
&= \sum_{A \in \mathcal{A}} \frac{\binom{n-|A|}{m-|A|}}{\binom{n}{m}} \\
&\leq \sum_{A \in \mathcal{A}} \left(\frac{m}{n}\right)^{|A|} \\
&= |\mathcal{A}|(m/n)^{\lceil \delta m \rceil} \\
&\leq \binom{\epsilon n}{\lceil \delta m \rceil} \frac{m}{n}^{\lceil \delta m \rceil} \quad by\,Theorem\,8 \\
&\leq \left(\frac{\epsilon e n}{\lceil \delta m \rceil} \frac{m}{n}\right)^{\lceil \delta m \rceil} \quad as\,\epsilon = \delta/6 \\
&\leq 2^{-\delta m}.
\end{aligned}$$

Therefore $P(R\,is\,\delta\text{-}Roth) \to 1$ as $n \to \infty$. $\square$

We can deduce Theorem 8 from the following theorem, which is a consequence of Roth's theorem.

Theorem 9. ([14, 19]) *For each positive δ, there is an integer n_0 and a positive β so that: if $n \geq n_0$, then each integer set A with $|A \cap [n]| > \delta n$ contains at least (βn^2) 3-term APs.*

Definition 11. *Let V be a finite set and $\mathcal{H} \subseteq \mathcal{P}(V)$, $\mathcal{H}$ is called hypergraph and each element in $\mathcal{H}$ can be considered as a termed edge.*

Now we prove Theorem 8 using Theorem 9.

Proof. Let $\epsilon > 0$ be given, let $\delta = \epsilon/2$ and apply Theorem 9 to get corresponding n_0 and β. Suppose $n \geq n_0$, let B be an arbitrary subset of $[n]$ and m, n' integers. Define

$$a(B, m) = \#\{I \subseteq B : |I| = m,\ I\,contains\,no\,3 - term\,AP\},$$

$$a(n', m) = max\{a(B, m) : B \subseteq [n], |B| = n'\}.$$

Aim: show there exists a C (depending only on ϵ), such that $m \geq C\sqrt{n}$ implies $a([n], m) = a(n, m) \leq \binom{\epsilon n}{m}$. We can prove it by $a(n', m) \leq \binom{n'}{m}$ and the following claim.

Claim. If $n' \geq \frac{\epsilon n}{2}$ and $m \geq 2\lfloor\sqrt{n}\rfloor$, then

$$a([n], m) \leq 2\binom{n}{\lfloor\sqrt{n}\rfloor}^2 a\left(n' - \lceil\beta n/12\rceil, m - 2\lfloor\sqrt{n}\rfloor\right).$$

We first derive our aim using the above claim.

Let $K = \lceil(12-6\epsilon)/\beta\rceil$ and suppose $m \geq \sqrt{n}$. We apply the claim recursively for K times to obtain

$$\begin{aligned} a(n,m) &\leq 2\binom{n}{\lfloor\sqrt{n}\rfloor}^2 a\left(n - \lceil\beta n/12\rceil, m - 2\lfloor\sqrt{n}\rfloor\right) \\ &\leq \left(2\binom{n}{\lfloor\sqrt{n}\rfloor}^2\right)^2 a\left(n - 2\lceil\beta n/12\rceil, m - 2\left(2\lfloor\sqrt{n}\rfloor\right)\right) \\ &\quad \cdots\cdots \\ &\leq 2^K\binom{n}{\lfloor\sqrt{n}\rfloor}^{2K}\binom{\epsilon n/2}{m - 2K\lfloor\sqrt{n}\rfloor} \\ &\leq 2^K\binom{2Kn}{2K\lfloor\sqrt{n}\rfloor}\binom{\epsilon n/2}{m - 2K\lfloor\sqrt{n}\rfloor}. \end{aligned} \tag{3.3}$$

Denote $r(t) := 2^K\binom{2Kn}{t}\binom{\epsilon n/2}{m-t}$, i.e. replace $2K\lfloor\sqrt{n}\rfloor$ of the right hand side of (3.3) by t. We can assume $m < \epsilon n/4$ as otherwise by Roth's theorem $[n]$ is $(\epsilon/4)$-Roth for sufficiently large n and hence $a(n,m) = 0$. By elementary calculation we have

$$\frac{r(t+1)}{r(t)} = \frac{2Kn - t}{t+1}\frac{m-t}{\epsilon n/2 - m + t + 1} \leq \frac{2Knm}{(t+1)(\epsilon n/2 - m)} \leq \frac{8Km}{\epsilon(t+1)}.$$

Let $T = 2K\lfloor\sqrt{n}\rfloor$, then

$$\begin{aligned} a(n,m) \leq r(T) &= \prod_{t=0}^{T-1}\left(\frac{r(t+1)}{r(t)}\right) r(0) \\ &\leq 2^K\frac{(8Km)^T}{\epsilon^T T!}\binom{\epsilon n/2}{m} \\ &\leq 2^K\left(\frac{8eKm}{\epsilon T}\right)^T\left(\frac{1}{2^m}\right)\binom{\epsilon n}{m}. \end{aligned}$$

If D is sufficiently large (depending on K and ϵ), then for every m with $m \geq D\sqrt{n} \geq D/(2k)T$ we have

$$2^{K/m}\left(\frac{8eKm}{\epsilon T}\right)^{T/m} \leq 2,$$

and hence

$$a(n,m) \leq 2(1/2)^m \binom{\epsilon n}{m} \leq \binom{\epsilon n}{m}$$

as required in the theorem.

Now we prove the claim we used.

Proof of the claim. Let $\mathcal{H}$ be the 3-uniform hypergraph with vertex set $[n]$ and all 3-term APs as termed edges. Let B be an arbitrary subset of $[n]$ with n' elements, then $e_{\mathcal{H}} \geq \beta n^2$. Let $Z := \{v \in B : deg_{\mathcal{H}}(v, B) > \beta n\}$, i.e. the set of numbers in $[n]$ contained in more than βn 3-term APs. We have $|Z| \geq \beta n$ as the maximum degree of elements in $\mathcal{H}$ is at most $2n$.

Then we estimate the number of m-element subset A of B without 3-term APs.

For those A with $|A \cap Z| < \sqrt{n}$, we can partition A into A_1, A_2 where $|A_1| = \lfloor \sqrt{n} \rfloor$ and $A_2 \subseteq B \setminus Z$. There are at most $\binom{n}{\lfloor \sqrt{n} \rfloor} a\left(n' - \lceil \beta n \rceil, m - \lfloor \sqrt{n} \rfloor\right)$ such A.

For those A with $|A \cap Z| \geq \sqrt{n}$, we will apply Lemma 3 to get a suitable upper bound. Let W be an arbitrary subset of Z and consider the auxiliary graph G_W with vertex set B and edges $\{\{x, y\} : x, y, z \in \mathcal{H}\ for\ some\ z \in W\}$. As there are at most three such z for each pair of $\{x, y\}$, we have $3e(G_W) \geq |W|\beta n$ and the maximum degree of elements in G_W is no more than $3|W|$. Hence for an arbitrary subset $U \subseteq B$ with $|U| \geq n' - \beta n/12$ we have

$$e_{G_W}(U) \geq e(G_W) - |B \setminus U|\Delta(G_W) \geq \frac{\beta n|W|}{3} - \frac{\beta n}{12} 3|W| = \frac{\beta n|W|}{12}. \quad (3.4)$$

We observe that if $I \cup W$ does not contain a 3-term AP, then I is independent in G_W.

Let $w = \lfloor \sqrt{n} \rfloor$ and fix a subset $W \subseteq Z$ with $|W| = w$. We will find the number of I with $m - w$ elements so that $I \cup W$ is an m-element subset of B and contains no 3-term APs. By our observation I is am independent subset of G_W. Then we apply Lemma 3.

Let $G = G_W$, $q = \lfloor \sqrt{n} \rfloor$, $R = n' - \lceil \beta n/12 \rceil$ and $D = \beta w/6$ in Lemma 3, then $\mathcal{S}$ is the set of q-element sets satisfying corresponding g and f. The assumption of Lemma is satisfied by 3.4. As $I \cap f(g(I))$ contains no 3-term APs for each I, the number of I, denoted by E_W satisfies

$$E_W \leq \sum_{S \in \mathcal{S}} a(f(S), m - w - q) \leq \binom{n}{q} a(R, m - w - q)$$

as $|S| = q$ and $|f(S)| \leq R$.

Finally we conclude that

$$\begin{aligned}
a(B,m) &\leq \binom{n}{\lfloor\sqrt{n}\rfloor} a(n' - \lceil \beta n \rceil, m - \lfloor\sqrt{n}\rfloor) + \sum_{W \subseteq Z : |W| = w} E_W \\
&\leq \binom{n}{\lfloor\sqrt{n}\rfloor}^2 a(n' - \lceil \beta n \rceil, m - 2\lfloor\sqrt{n}\rfloor) \\
&\quad + \binom{n}{w}\binom{n}{q} a(n' - \lceil \beta n/12 \rceil, m - 2\lfloor\sqrt{n}\rfloor) \\
&\leq 2\binom{n}{\lfloor\sqrt{n}\rfloor}^2 a(n' - \lceil \beta n/12 \rceil, m - 2\lfloor\sqrt{n}\rfloor),
\end{aligned}$$

which proves the claim as B was an arbitrary n'-element subset of $[n]$.

$\square$

Bibliography

[1] N. Alon. Independent sets in regular graphs and sum-free subsets of finite groups. *Israel J. Math.*, 73:247–256, 1991.

[2] N. Alon, J. Balogh, R. Morris, and W. Samotij. Counting sum-free sets in abelian groups. *Israel J. Math*, 199:309–344, 2014.

[3] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks. *Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986)*, 72:15–19, 1988.

[4] N. Alon and V. Rödl. Sharp bounds for some multicolor ramsey numbers. *Combinatorica*, 25:125–141, 2005.

[5] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. *J. Amer. Math. Soc.*, to appear.

[6] W. G. Brown. On graphs that do not contain a thomsen graph. *Canad. Math. Bull.*, 9:281–285, 1966.

[7] A. J. Hoffman. On eigenvalues and colorings of graphs. *Graph Theory and its Applicationss (Proc. Advanced Sem., Math. Research Center, Univ. of Wisconsin, Madison, Wis., 1969) Academic Press, New York*, pages 79–91, 1970.

[8] J. Kahn. An entropy approach to the hard-core model on bipartite graphs. *Combin. Probab. Comput.*, 10:219–237, 2001.

[9] T. Kövari, V. T. Sós, and P. Turán. On a problem of k. zarankiewicz. *Colloquium Math.*, 3:50–57, 1954.

[10] D. J. Kleitman and K. J. Winston. The asymptotic number of lattices. *Ann. Discrete Math.*, 6:243–249, 1980.

[11] D. J. Kleitman and K. J. Winston. On the number of graphs without 4-cycles. *Discrete Math.*, 41:167–172, 1982.

[12] Y. Kohayakawa, S. Lee, V. Rödl, and W. Samotij. The number of sidon sets and the maximum size of sidon sets contained in a sparse random set of integers. *Random Structures Algorithms*, to appear.

[13] Y. Kohayakawa, T. Luczak, and V. Rödl. Arithmetic progressions of length three in subsets of a random set. *Acta Arith.*, 75:133–163, 1996.

[14] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[15] W. Samotij. Counting independent sets in graphs. *European J. Combinatorics*, pages 5–18, Aug 2015.

[16] A. A. Sapozhenko. On the number of independent sets in extenders. *Diskret. Mat.*, 13:56–62, 2001.

[17] A. A. Sapozhenko. The cameron-erdös conjecture. *Dokl. Akad. Nauk*, 393:749–752, 2003.

[18] D. Saxton and A. Thomason. Hypergraph containers. *[math.CO]*, 2012.

[19] P. Varnavides. On certain sets of positive density. *J. London Math. Soc.*, 34:315–320, 2010.