# THE LOVASZ LOCAL LEMMA, K-COLORING, AND APPLICATIONS IN GRAPH THEORY

CATHY WANG

Abstract. This paper introduces the Lovasz local lemma, which is a tool in the probabilistic method, and the concept of $k$-coloring. The paper then presents two applications of the tools in graph theory, lower bounds of Ramsey numbers and linear arboricity of $k$-regular graphs.

## Contents



## 1. Introductions

The probabilistic method is a nonconstructive method used in combinatorics. When proving that an object with a certain property exists, we can prove that the probability of finding an object with that property in a specified set of objects is strictly greater than zero. The tool is widely used in graph theory, which studies mathematical models that reflect pairwise relationship between objects. In this paper, we shall introduce a result called the Lovasz local lemma, which tells us that, as long as a set of events are "mostly" independent with each other, the probability that none of the events occur is positive. This lemma is useful for finding bounds for certain properties, and we shall apply this to two concepts in graph theory: the lower bound of Ramsey numbers and the upper bound of linear arboricity in regular graphs. Throughout the proofs, we shall also utilize the concept of $k$-coloring, which labels elements in a set using "colors".

## 2. The Lovasz Local Lemma

In this section, we shall introduce a proof of the Lovasz local lemma. The proof of the lemma utilizes some basic definitions of graph theory.

**Definition 2.1.** A *graph* $G\,(V,E)$ is a collections of vertices $V$ and edges $E$, where $E$ is a set of unordered pairs $(i,j)$ with $i,j \in V$, $i \neq j$. We say that two vertices $i,j \in V$ are adjacent if $(i,j) \in E$.

The local lemma utilizes the definition of a dependency graph, which relates a graph to random events.

**Definition 2.2.** A *dependency graph* on events $A_1, A_2, \cdots, A_n$ is a graph $G(V, E)$ with $V = [n]$ and each event $A_i$ is dependent with its neighbors $A_j \in \{A_j \mid (i,j) \in E\}$ and independent with its non-neighbors $A_j \in \{A_j \mid (i,j) \notin E\}$.

**Lemma 2.3** (Lovasz Local Lemma)**.** *Let $A_1, A_2, \cdots, A_n$ be events in a probability space $(\Omega, \mathcal{F}, P)$. Let $G(V, E)$ be a dependency graph on the events. If there exist real numbers $x_1, \cdots, x_n \in [0, 1)$ such that for all $i \in \mathbb{N}, i \leq n$, we have that $P(A_i) \leq x_i \cdot \prod_{(i,j) \in E} (1 - x_j)$, then,*

$$P\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) \geq \prod_{i=1}^{n} (1 - x_i)$$

*Proof.* We shall first prove by induction that for any $S \subset V = [n]$, $|S| < n$, and any $i \notin S$, we have

$$P\left(A_i \mid \bigwedge_{j \in S} \overline{A_j}\right) < x_i$$

(1) The base case is when $|S| = 0$. When $S = \emptyset$, we have that $\bigwedge_{j \in S} \overline{A_j} = \Omega$. Therefore,

$$P\left(A_i \mid \bigwedge_{j \in S} \overline{A_j}\right) = P(A_i) \leq x_i \cdot \prod_{(i,j) \in E} (1 - x_j) < x_i$$

(2) We shall next prove the inductive step. Suppose that the statement is true for all $S$ with $|S| < k$ for some $k \in \mathbb{N}$, $k < n$. Next, consider any $S_0 \subset V$ with $|S_0| = k$. Fix $i \notin S_0$. Let $S_1 = \{j \in S_0 | (i,j) \in E\}$ and $S_2 = S_0 \setminus S_1$. Thus, we have that $A_i$ is dependent with events corresponding with vertices in $S_1$ and independent with those in $S_2$. This leads to

$$P\left(A_i \mid \bigwedge_{j \in S_0} \overline{A_j}\right) = \frac{P\left(A_i \wedge \left(\bigwedge_{j \in S_1} \overline{A_j}\right) \mid \bigwedge_{j \in S_2} \overline{A_j}\right)}{P\left(\bigwedge_{j \in S_1} \overline{A_j} \mid \bigwedge_{j \in S_2} \overline{A_j}\right)}.$$

We know that, because $A_i$ is independent with all elements in $S_2$, we have that

$$\begin{aligned} P\left(A_i \wedge \left(\bigwedge_{j \in S_1} \overline{A_j}\right) \mid \bigwedge_{j \in S_2} \overline{A_j}\right) &\leq P\left(A_i \mid \bigwedge_{j \in S_2} \overline{A_j}\right) \\ &= P(A) \\ &\leq x_i \cdot \prod_{(i,j) \in E} (1 - x_j). \end{aligned}$$

Let $S_1 = \{j_1, j_2, \cdots, j_m\}$ for some $m \leq k$. We have that

$$\begin{aligned}
P\left(\bigwedge_{j\in S_1} \overline{A_j} \mid \bigwedge_{j\in S_2} \overline{A_j}\right) &= P\left(\overline{A_{j_1}} \wedge \overline{A_{j_2}} \wedge \cdots \wedge \overline{A_{j_m}} \mid \bigwedge_{j\in S_2} \overline{A_j}\right) \\
&= P\left(\overline{A}_{j_1} | \overline{A_{j_2}} \wedge \cdots \wedge \overline{A_{j_m}} \wedge \bigwedge_{j\in S_2} \overline{A_j}\right) \\
&\cdot P\left(\overline{A}_{j_2} | \overline{A_{j_3}} \wedge \cdots \wedge \overline{A_{j_m}} \wedge \bigwedge_{j\in S_2} \overline{A_j}\right) \\
&\cdots P\left(\overline{A}_{j_m} | \bigwedge_{j\in S_2} \overline{A_j}\right) \\
&= \left(1 - P\left(A_{j_1} | A_{j_2} \wedge \cdots \wedge A_{j_m} \wedge \bigwedge_{j\in S_2} \overline{A_j}\right)\right) \\
&\cdot \left(1 - P\left(A_{j_2} | A_{j_3} \wedge \cdots \wedge A_{j_m} \wedge \bigwedge_{j\in S_2} \overline{A_j}\right)\right) \\
&\cdots \left(1 - P\left(A_{j_m} | \bigwedge_{j\in S_2} \overline{A_j}\right)\right) \\
&\geq (1 - x_{j_1})(1 - x_{j_2}) \cdots (1 - x_{j_m}) \\
&\geq \prod_{(i,j)\in E} (1 - x_j).
\end{aligned}$$

where the second equation comes from the fact that

$$\begin{aligned}
&P(A_1 \mid A_2 \wedge A_3 \wedge \cdot A_n \wedge B) \cdot P(A_2 \mid A_3 \wedge \cdots \wedge A_n \wedge B) \cdots P(A_n \mid B) \\
&= \frac{P(A_1 \wedge A_2 \wedge \cdots \wedge A_n \wedge B)}{P(A_2 \wedge A_3 \wedge \cdots \wedge A_n \wedge B)} \cdot \frac{P(A_2 \wedge A_3 \wedge \cdots \wedge A_n \wedge B)}{P(A_3 \wedge \cdots \wedge A_n \wedge B)} \cdot \frac{P(A_n \wedge B)}{P(B)} \\
&= \frac{P(A_1 \wedge A_2 \wedge \cdots \wedge A_n \wedge B)}{P(B)} \\
&= P(A_1 \wedge A_2 \wedge \cdots \wedge A_n \wedge B \mid B) \\
&= P(A_1 \wedge A_2 \wedge \cdots \wedge A_n \mid B)
\end{aligned}$$

Therefore, $P\left(A_i \mid \bigwedge_{j\in S} \overline{A_j}\right) \leq x_i$.

Thus, we have that

$$\begin{aligned}
P\left(\bigwedge_{i=1}^{n} \overline{A}_i\right) &= (1 - P(A_1)) \cdot \left(1 - P\left(A_2 \mid \overline{A_1}\right)\right) \cdots \left(1 - P\left(A_n \mid \bigwedge_{i=1}^{n-1} \overline{A}_i\right)\right) \\
&\geq \prod_{i=1}^{n} (1 - x_i)
\end{aligned}$$

□

A direct corollary of the local lemma is its symmetric version.

**Corollary 2.4.** *Let $A_1, A_2, \cdots, A_n$ be events in a probability space $(\Omega, \mathcal{F}, P)$ such that each event is pairwise independent with all but at most $d$ events, and $P(A_i) < p$ for all $i \in \mathbb{N}$, $i \leq n$. If $ep(d+1) < 1$, then*

$$P\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) > 0$$

*Proof.* If $d = 0$, we have that $P\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) = (1-p)^n > 0$.

If $d > 0$, Let $G(V, E)$ be a dependency graph on the events $A_1, \cdots, A_n$. According to the local lemma, if we can find real numbers $x_1, \cdots, x_n \in (0,1)$ such that for all $i$, $P(A_i) \leq x_i \cdot \prod_{(i,j)\in E}(1-x_j)$, we have that $P\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) \geq \prod_{i=1}^{n}(1-x_i) > 0$.

Let $x_1 = \cdots = x_n = \dfrac{1}{d+1} \in (0,1)$. We have that $(1-x)^d = \left(1 - \dfrac{1}{d+1}\right)^d < \dfrac{1}{e}$. It follows that

$$P(A_i) < p < \frac{1}{e(d+1)} < x(1-x)^d = x_i \cdot \prod_{(i,j)\in E}(1-x_j)$$

Therefore, $P\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) > 0$. □

## 3. $k$-Colorings and Ramsey Theory

In this section, we shall introduce the concept of $k$-coloring. The tool is a function that assigns a finite set with different "colors", thereby catagorizing elements in the set. If the function assigns the colors randomly, we can prove the existence of certain types of set by calculating whether the probability of a certain kind of color arrangement is greater than 0. Here is a formal definition of $k$-coloring.

**Definition 3.1.** A *$k$-coloring* is a function $f : S \to \{0, 1, \cdots, k-1\}$ where $S$ is a finite set and the numbers $0, 1, \cdots, k-1$ each represent a color.

This is a very useful method for the Ramsey theory, which is built off the concept that, given a large enough number of people, there must exist a number of people who know each other or a number of people who are strangers to each other. Before we talk about how to apply the $k$-coloring method to Ramsey theory, we shall make this concept precise utilizing graph theory.

**Definition 3.2.** A graph $G'(V', E')$ is a *subgraph* of $G(V, E)$ if $V' \subset V$ and $E' \subset \{(i,j) \in E \mid i, j \in V'\}$. The subgraph $G'$ is denoted as $G[V']$ if $E = \{(i,j) \in E \mid i, j \in V'\}$.

**Definition 3.3.** A *complete graph* is a graph $G(V, E)$ where $E$ contains all possible edges. We have that $E = \{(i,j) \mid i, j \in V\}$ and $|E| = \binom{|V|}{2}$. A complete graph with $n$ vertices is denoted by $K_n$.

**Definition 3.4.** A *clique* of a graph $G(V,E)$ is a subset $S_1$ of $V$ such that $G[S_1]$ is a complete graph. An *independent set* of a graph $G(V,E)$ is a subset $S_2$ of $V$ such that no two vertices in $S$ are adjacent in $G$.

These concepts allow us to define the Ramsey number:

**Definition 3.5.** The *Ramsey number* $R(k,l)$ is the smallest number $n$ such that for all graphs with $n$ vertices, there must exist a clique of $k$ vertices or an independent set of $l$ vertices.

Ramsey theory states that all Ramsey numbers exist. We can show this by proving an upper bound for the Ramsey numbers.

**Theorem 3.6.** $R(k,l) \leq \binom{k+l-2}{k-1}$.

To do this, we shall first prove the inductive step:

**Lemma 3.7.** *$R(k,l) \leq R(k,l-1) + R(k-1,l)$.*

*Proof.* Let $G(V,E)$ be a graph such that $|V| = R(k,l-1)+R(k-1,l)$. Let $v \in V$. Because the total number of vertices that $v$ can be either adjacent or nonadjacent to is $|V|-1 = R(k,l-1)+R(k-1,l)-1$, we have that either $v$ is adjacent to a subset of vertices $V_1 \subset V$ such that $|V_1| = R(k-1,l)$ or $v$ is nonadjacent to a subset of vertices $V_2 \subset V$ such that $|V_2| = R(k,l-1)$.

Suppose the first case is true. We know that $G[S]$ contains either a clique of $k-1$ vertices or an independent set of $l$ vertices. If the former is true, because $v$ is adjacent to all vertices in the $k-1$ clique, we know that $G[S \cup \{v\}]$ a clique of $k$ vertices. If the latter is true, we have an independent set of $l$ vertices. The same argument holds for the second case.

From the above reasoning, $G$ must contain a clique of $k$ vertices or an independent set of $l$ vertices. Thus, $R(k,l) \leq R(k,l-1) + R(k-1,l)$. □

The proof for Lemma 3.6 is simple using Lemma 3.7:

*Proof.*

(1) The base case is when either $k$ or $l$ equals to 1. We know that $R(1,l) = R(k,1) = 1$ for all $k,l \in \mathbb{N}$ because any vertex is a clique of 1 vertex. We also know that $\binom{l-1}{0} = \binom{k-1}{k-1} = 1$. Thus, the statement is true for the base case.

(2) Let $k,l$ be positive integers such that $k,l > 1$. Suppose the theorem is valid for all $k',l'$ such that $k'+l' < k+l$. We have that

$$\begin{aligned} R(k,l) &\leq R(k,l-1) + R(k-1,l) \\ &\leq \binom{k+l-3}{k-1} + \binom{k+l-3}{k-2} \\ &= \binom{k+l-2}{k-1} \end{aligned}$$

□

*Remark* 3.8. A restatement of the Ramsey number using 2-coloring would be, $R(k,l)$ is the smallest number $n$ such that, given any complete graph $G(V,E)$ with $n$ vertices, every possible 2-coloring of $E$ contains either a complete subgraph $K_k$ whose edges are all colored with 0 or or a complete subgraph $K_l$ whose edges are all colored with 1.

$k$-coloring allows us to obtain a fairly straightforward proof for $R(3,3)$.

**Example 3.9.** $R(3,3) = 6$.

*Proof.* Consider the 2-coloring of $K_6([6], E)$, $f : E \to \{0,1\}$. Pick a vertex $1 \in [6]$ from the graph and consider the edges that extend from that vertex, $(1,2), (1,3), (1,4), (1,5), (1,6)$. Because the vertex is connected to 5 vertices, there must exist three vertices with edges that share the same color. Suppose without loss of generality that $f(1,2) = f(1,3) = f(1,4) = 0$ . Then, consider the edges that connect the vertices $2,3,4$. If, for example, $f(2,3) = 0$, the vertices $1,2,3$ form a monochromatic subgraph of color 0. Otherwise, the vertices $2,3,4$ form a monochromatic subgraph of color 1. Thus, $R(3,3) \leq 6$.

Then consider the 2-coloring of $K_5([5], E')$, $f : E' \to \{0,1\}$. If $f(1,2) = f(2,3) = f(3,4) = f(4,5) = f(5,1) = 0$ and $f(1,3) = f(1,4) = f(2,4) = f(2,5) = f(3,5) = 1$, we have that there does not exist a complete subgraph $K_3$ whose edges are monochromatic. Therefore, $R(3,3) > 5$.

Overall, $R(3,3) = 6$. □

## 4. Lower Bounds of Ramsey Numbers

We shall first prove a lower bound for diagonal Ramsey numbers utilizing $k$-coloring and probability.

**Theorem 4.1.** *$R(k,k) > 2^{\frac{k}{2}}$ for all $k \geq 3$.*

*Proof.* Consider a random 2-coloring of the edges of a graph $K_{\lfloor 2^{\frac{k}{2}} \rfloor}$ for some $k \geq 3$, $f : \lfloor 2^{\frac{k}{2}} \rfloor \to \{0,1\}$ such that $P(f(i,j) = 0) = P(f(i,j) = 1) = \frac{1}{2}$ for all $i, j \in \lfloor 2^{\frac{k}{2}} \rfloor$. It follows that, for any set $S$ with $k$ vertices, the probability that all edges on $K_{\lfloor 2^{\frac{k}{2}} \rfloor}[S]$ are monochromatic is $\frac{1}{2}^{\binom{k}{2}-1}$. We know that there exists $\binom{\lfloor 2^{\frac{k}{2}} \rfloor}{k}$ choices of $S$. Therefore, the probability of at least one set of $S$ is monochromatic is $\binom{\lfloor 2^{\frac{k}{2}} \rfloor}{k} \cdot \frac{1}{2}^{\binom{k}{2}-1}$. Because

$$\binom{\lfloor 2^{\frac{k}{2}} \rfloor}{k} \cdot \frac{1}{2}^{\binom{k}{2}-1} < 2^{1-\frac{k^2+k}{2}} \cdot \frac{2^{\frac{k^2}{2}}}{k!} < 1,$$

we have that the probability that none of the sets have a monochromatic coloring is greater than 0, meaning that there exists a graph with $\lfloor 2^{\frac{k}{2}} \rfloor$ vertices such that there neither exists a clique nor an independent set of $k$ vertices. Therefore, $R(k,k) > 2^{\frac{k}{2}}$ for all $k \geq 3$. □

An improved lower bound utilizes the symmetric local lemma.

**Theorem 4.2.** *If $e \cdot 2^{1-\binom{k}{2}} \cdot \binom{k}{2}\binom{n-2}{k-2} < 1$, then $R(k,k) > n$.*

*Proof.* Consider a random 2-coloring of the edges of $K_n$, $f : [n] \to \{0,1\}$ such that $P(f(i,j) = 0) = P(f(i,j) = 1) = \frac{1}{2}$ for all $i, j \in [n]$. Let $S_i$ be a set of $k$ vertices and $A_{S_i}$ be the event that $K_n[S_i]$ is monochromatic. We know that there exists a total of $\binom{n}{k}$ events possible.

For two events $A_{S_i}$ and $A_{S_j}$ to be dependent, they have to share an edge. Thus, they must share 2 vertices. Thus, we have that $d \leq \binom{k}{2}\binom{n-2}{k-2} - 1$. We also know that, just like in the proof of Lemma 4.1, $P(A_{S_i}) = 2^{1-\binom{k}{2}}$. Therefore, applying

Lemma 2.4, we know that, if there exists some $K_n$ such that there does not exist a monochromatic coloring, $n$ satisfies

$$e \cdot 2^{1-\binom{k}{2}} \cdot \binom{k}{2}\binom{n-2}{k-2} < 1$$

Therefore, $R(k,k) > n$. □

One could also obtain lower bounds of asymptotic Ramsey number using the Lovasz local lemma. Here, we will discuss the lower bound of Ramsey numbers that guarantee either a triangle or a independent set of size $l$.

**Theorem 4.3.** *There exists a constant $c$ such that $R(3,l) \geq c\frac{l^2}{log^2 l}$.*

*Proof.* Consider a random 2-coloring of the edges of $K_n$, $f : [n] \to \{0,1\}$ such that $P(f(i,j) = 0) = p$ and $P(f(i,j) = 1) = 1-p$ for all $i, j \in [n]$. Let $S_i$ be a set of 3 vertices and $A_{S_i}$ be the event that $K_n[S_i]$ is monochromatic. Let $T_j$ be a set of $l$ vertices and $B_{T_j}$ be the event that $K_n[T_j]$ is monochromatic. Therefore, we have that $P(A_{S_i}) = p^3$, $P(B_{T_j}) = (1-p)^l$, the number of $A_{S_i}$ is $\binom{n}{3}$ and the number of $B_{T_j}$ is $\binom{n}{l}$.

We shall create a dependency graph $G(V,E)$ consisting of events $A_{S_i}$ and $B_{T_j}$. For each vertex $A_{S_i}$ it is connected to at most $3n$ other $A_S$. This is because, in order for $A_{S_i}$ to be dependent with $A_S$, they must share one edge. It is also connected to at most $\binom{n}{l}$ $B_T$. This is a trivial bound given that there are only $\binom{n}{l}$ events of $B_T$. According to the same logic, for each vertex $B_{T_j}$, it is connected to at most $\binom{l}{2}n$ $A_S$ and at most $\binom{n}{l}$ $B_T$.

In order to apply the local lemma, all we need to do is find real numbers $x_1, \cdots, x_{\binom{n}{3}+\binom{n}{l}}$ such that for all $i \in \mathbb{N}, i \leq \binom{n}{3}+\binom{n}{l}$, $P(A_i) \leq x_i \cdot \prod_{(i,j)\in E}(1-x_j)$.

Because all $A_{S_i}$ are identical and all $B_{T_j}$ are identical, we can assign the same $x$ to all $A_{S_i}$ and the same $y$ to all $B_{T_j}$. Thus, we need to find $p, x, y \in [0,1)$ and $n \in \mathbb{N}$ that satisfy

$$p^3 \leq x(1-x)^{3n}(1-y)^{\binom{n}{l}}$$
$$(1-p)^l \leq y(1-x)^{\binom{l}{2}n}(1-y)^{\binom{n}{l}}$$

The process of finding the values is quite tedious and is a detour from our discussion of probabilistic methods, and therefore will be omitted from this paper. However, when $l \geq 20\sqrt{n}\log n$, we can find $p = \frac{1}{3\sqrt{n}}$, $x = \frac{1}{9n^{\frac{3}{2}}}$ and $y = \frac{1}{\binom{n}{l}}$ that satisfy the inequalities.

Thus, when $n \leq \frac{l^2}{(40\log l)^2}$, we have that $l \geq 20\sqrt{n}\log n$, and there either exists a monochromatic triangle or a monochromatic subgraph of $l$ vertices. Therefore, we have $c = \frac{1}{1600}$. □

*Remark* 4.4. This lower bound is fairly close to the best possible because $R(3,l) = \Theta\left(\frac{l^2}{\log^2 l}\right)$, the upper bound of which was proved by Ajtai, Kolmos, Sezmeredi, and the lower bound of which was proved by Jeong Han Kim.

## 5. Preliminaries of Linear Arboricity

We shall move on to another application of the probabilistic method in graph theory, which is linear arboricity. To reach the result we want, we must introduce some definitions.

We shall first define the concept of linear arboricity:

**Definition 5.1.** A *walk* in a graph $G(V,E)$ is a finite sequence $W = v_0e_1v_1 \cdots e_kv_k$ where $v_i \in V$ for all $0 \le i \le k$, $e_j \in E$ for all $1 \le j \le k$ and $e_j = (v_{j-1}, v_j)$. A *cycle* is a walk $W$ where $v_0 = v_k$. A *path* is a walk where $v_i \neq v_j$ for all $0 \le i < j \le k$. The

**Definition 5.2.** Two vertices $i, j$ of a graph $G$ are *connected* if there exists a walk between $i$ and $j$.

**Definition 5.3.** A *linear forest* is a graph in which every connected component is a path.

**Definition 5.4.** The *linear arboricity* $la(G)$ of a graph $G$ is the minimum number of linear forests in $G$, whose union is the set of all edges in $G$.

We shall next provide some more definitions related to the proofs in the next section:

**Definition 5.5.** The *degree* of a vertex $v$ of graph $G(V,E)$, $deg(v)$, is the number of edges connected to $v$.

**Definition 5.6.** A *d-regular graph* $G(V,E)$ is a graph where $deg(v) = d$ for all $v \in V$.

**Definition 5.7.** A *directed graph*, or a digraph, $D(V,E)$ is a collections of vertices $V$ and edges $E$, with $E$ being a set of ordered pairs $(i,j)$ with $i, j \in V$. The *indegree*, $d_D^-(v)$, of a vertex $v$ in graph $D$ is the number of elements in $\{(i,v) \mid (i,v) \in E\}$. The *outdegree*, $d_D^+(v)$, of a vertex $v$ in graph $D$ is the number of elements in $\{(v,j) \mid (v,j) \in E\}$. A *d-regular digraph* is a digraph $D(V,E)$ where $d_D^-(v) = d_D^+(v) = d$ for all $v \in V$. The linear arboricity of a directed graph $G$ is denoted as $dla(G)$.

**Definition 5.8.** The *directed girth* of a digraph $D(V,E)$ is the minimum length of a directed cycle in $D$.

**Definition 5.9.** A subgraph $G'(V',E')$ of a graph $G(V,E)$ is a *spanning subgraph* if $V' = V$.

**Definition 5.10.** A set of edges $M$ is a *matching* of a graph $G(V,E)$ if $M \subset E$ and no two edges in $M$ share a vertex. A matching is perfect if for every $v \in V$, there exists an edge in $M$ that connects $v$.

**Definition 5.11.** A graph $G(V,E)$ is *bipartite* if its vertices $V$ can be partitioned into two sets $X, Y$ such that each edge connects a vertex from $X$ and another vertex from $Y$.

**Definition 5.12.** Let $G(V,E)$ be a graph. Let $S \subset V$. The *neighboring set* of $S$ in $G$ is the set $\{v \mid u \in S, v \in V\ (u,v) \in E\}$. We denote this set $N(S)$.

## 6. Linear Arboricity

Now, we can introduce a conjecture about linear arboricity, which was brought forth in Akiyama, Exoo and Harary (1981):

**Conjecture 6.1.** *The linear arboricity of every $d$-regular graph, $la\,(G)$, is* $\dfrac{\lceil d+1 \rceil}{2}$.

This conjecture can be transformed into one that deals with directed graphs, because it is quite easy to see that the edge of any $2d$-regular graph can be oriented in a way such that the resulting digraph is a $d$-regular digraph, and a $(2d-1)$-regular graph is simply a subgraph of the $2d$-regular graph. Therefore, the digraph version of Lemma 6.1 can be stated as

**Conjecture 6.2.** *The linear arboricity of every $d$-regular digraph, $dla\,(G)$, is $d+1$.*

Unfortunately, the conjecture does not yet have a proof, but we shall prove an upper and lower bound for the linear arboricity.

The lower bound can be obtained without the probabilistic method.

**Theorem 6.3.** *For any $d$-regular graph $G\,(V,E)$, $la\,(G) \geq \frac{\lceil d+1 \rceil}{2}$.*

*Proof.* We know that each $d$-regular graph $G\,(V,E)$ has $\frac{|V|d}{2}$ edges, and the number of edges of any linear forest is at most $|V|-1$. Therefore,

$$la\,(G) \geq \frac{\frac{|V|d}{2}}{|V|-1} > \frac{d}{2}$$

Therefore, $la\,(G) \geq \dfrac{\lceil d+1 \rceil}{2}$. □

We shall move on to the upper bounds of linear arboricity of regular graphs. We can see from the proofs below that it is simpler to find upper bounds for the directed version. To obtain the bounds, we shall first prove that the conjecture holds for graphs with certain structures, specifically high girths, and then apply this result to general directed graphs.

**Theorem 6.4.** *Let $D\,(V,E)$ be a $d$-regular directed graph with directed girth $g \geq 8ed$. Then, we have that $dla\,(G) = d+1$.*

In order to prove the theorem, we need a few lemmas:

**Lemma 6.5.** *There exists a perfect matching for all $k$-regular bipartite graphs.*

*Proof.* We shall first prove that for any bipartite graph $G\,(V,E)$ with the partition $(X,Y)$, if for all $S \subset X$ we have $|N\,(S)| \geq |S|$, then there exists a matching for $G$ that connects all vertices in $X$ by induction:

1. When $|V| = 2$ and $|E| = 1$, we know that $V$ is partitioned into $(\{1\},\{2\})$. $E$ itself is a matching for $G$ that contains $\{1\}$.
2. Suppose the statement holds for all graphs with $|V| \leq i$ and $|E| < j$ for some $i, j \in \mathbb{N}$. Let $G\,(V,E)$ be any bipartite graph with $|V| = i$ and $|E| = j$. There are two cases of $G$:
   - (a) Suppose for all $S \subset X$, $|N\,(S)| > |S|$. Let $e \in E$. We have that $G'\,(V, E \setminus \{e\})$ still satisfies the conditions. Thus, there exists a matching $M$ in $G'$ which connects all vertices in $X$. Therefore, $M$ satisfies the conditions for $G$.

(b) Suppose that there exists some $S \subset X$ such that $|N(S)| = S$. Consider the subgraphs $G_1 = G[S \cup N(S)]$ and $G_2 = G[(X \setminus S) \cup (Y \setminus N(S))]$. We can check that the two subgraphs satisfy the condition:
   (i) For $G_1$, take any set $T \subset S$. By definition of neighboring sets, we know that $N_G(T) \subset N_G(S)$. Therefore, $N_{G_1}(T) = N_G(T)$. Because $|N_G(T)| \geq |T|$, we have that $|N_{G_1}(T)| \geq |T|$.
   (ii) For $G_2$, suppose that there exists some $T' \subset (X \setminus S)$ such that $|N_{G_2}(T')| < |T'|$. We know that $N_{G_2}(T') = N_G(T') \setminus N(S)$ and that $N(S \cup T') = N_G(T') \cup N_G(S)$. Therefore,
$$|N_G(S \cup T')| = |N_G(T') \cup N_G(S)| = |N_G(S)| + |N_{G_2}(T')|$$
   Because $|S| = |N_G(S)|$ and $|T'| > |N_{G_2}(T')|$, we have that
$$|S \cup T'| = |S| + |T'| > |N_G(S \cup T')|$$
   This contradicts with our assumption that for $S \cup T' \subset X$, $|N_G(S \cup T')| \geq |S \cup T'|$. Therefore, there does not exist some $T' \subset (X \setminus S)$ such that $|N_{G_2}(T')| < |T'|$.

   Therefore, there exists some matching $M_1, M_2$ for the subgraphs respectively. The set $M_1 \cup M_2$ is a matching for $G$ that connects all vertices in $X$.

Therefore, according to induction, if for all $S \subset X$ we have $|N(S)| \geq |S|$, then there exists a perfect matching for $G$.

Next, we shall consider a $k$-regular bipartite graph $G$ with the partition $(X, Y)$. Let $n_X, n_Y$ be the number of edges connecting to vertices in $X$ and $Y$ respectively. Because $G$ is bipartite, we know that $n_X = n_Y$. Because $G$ is $k$-regular, $n_X = k|X|$ and $n_Y = k|Y|$. Therefore, $|X| = |Y|$.

Let $S \subset X$. Let $E_1, E_2$ be the collection of edges that are incident with vertices in $S$ and $N(S)$ respectively. For each edge $e \in E_1$, we know that it is incident with a vertex in $N(S)$. Therefore, $e \in E_2$, and so $E_1 \subset E_2$. Because $G$ is $k$-regular, we have that $|E_1| = k|S|$ and $|E_2| = k|N(S)|$, which leads to that $|S| \leq |N(S)|$. Therefore, there exists a matching $M$ of the $k$-regular bipartite graph that contains all vertices in $X$. Because all edges in $M$ connect a vertex in $X$ and a vertex in $Y$, we have that $|N_M(X)| = |X| = |Y|$. Therefore, $N_M(X) = Y$. Thus, $M$ is a perfect matching.

□

**Lemma 6.6.** *The edges of a $k$-regular bipartite graph $G(V, E)$ can be decomposed into $k$ perfect matchings.*

*Proof.* We shall prove this lemma by induction.

(1) When $k = 1$, we know that $G$ is a perfect matching itself.
(2) Suppose that for some $i \in \mathbb{N}$, all $i$-regular bipartite graphs $G(V, E)$ can be decomposed into $i$ perfect matchings. Let $G_0(V, E)$ be an $(i+1)$-regular bipartite graph. According to Lemma 6.5, there exists a perfect matching $M$ for $G_0(V, E)$. Consider the graph $G'(V, E \setminus \{M\})$. Because $M$ is a perfect matching, $G'$ is a $i$-regular graph. Therefore, $G'$ can be decomposed into $i$ perfect matchings, $M_1, \cdots, M_i$. It follows that $G_0$ can be decomposed into $M, M_1, \cdots, M_i$, which are $i + 1$ perfect matchings.

Therefore, the edges of a $k$-regular bipartite graph $G(V, E)$ can be decomposed into $k$ perfect matchings. □

**Lemma 6.7.** *Let $D(V,E)$ be a $d$-regular digraph. It can be partitioned into $d$ pairwise 1-regular spanning subgraphs.*

*Proof.* Consider a graph $H(V',E')$ where $|V'| = 2|V|$. We can partition $V'$ into $X, Y$ such that $|X| = |Y| = |V|$. Let $f : X \to V, g : Y \to V$ be bijective correspondences between the sets of vertices. We define $E' = \{(u,v) \mid u \in X, v \in Y, (f(u), g(v)) \in E\}$. Therefore, there exists a bijection $h : E' \to E$ where $h(u,v) = (f(u), g(v))$. We also know that $H$ is $d$-regular. According to Lemma 6.6, $H$ can be decomposed into $d$ perfect matchings. Consider any perfect matching $M_i$ for $1 \le i \le d$. For every $u' \in V$, there exists one and only one $u \in X$ such that $f(u) = u'$. Therefore, there exists one and only one $v \in Y$ such that $(u,v) \in M$. This indicates that there exists one and only one $v' \in V$ such that $f(v) = v'$. Consider the graph $G_i(V, f(M_i))$. We have that $deg_{G_i}(u) = 1$. Thus, $G_i$ is a 1-regular spanning subgraph of $G$. Because $\{M_i\}$ are partitions of $H$ and $|M_i| = d$, we have that $\{G_i\}$ are $d$ pairwise 1-regular spanning subgraphs of $G$. □

**Lemma 6.8.** *Let $G(V,E)$ be a graph with maximum degree $d$. Let $V_1, V_2, \cdots, V_r$ be pairwise disjoint and that $V_1 \cup V_2 \cup \cdots \cup V_r = V$ such that $|V_i| \ge 2ed$ for all $1 \le i \le r$. Then, there exists a set of vertices $W \subset V$ where $W$ is an independent set, and contains one vertex from each $V_i$.*

*Proof.* We shall first prove that this is true if $|V_i| = \lceil 2ed \rceil$ for all $1 \le i \le r$. We shall pick a vertex $v_i$ from each $V_i$ randomly such that it forms a set $W$. The probability that any vertex $v$ is picked is $P(v) = \dfrac{1}{\lceil 2ed \rceil}$. Let $f$ be any edge $f \in E$. Let $A_f$ be the event that both vertices of the edge are in $W$. Let the two vertices be named $i, j$. We have that $P(A_f) \le P(i)P(j) = P^2(v) = \dfrac{1}{\lceil 2ed \rceil^2}$. We also know that the event $A_f$ is dependent on the events that correspond to the edges which connect vertices that belong to the same set as $i$ or $j$. This leads to that $A_f$ is dependent with fewer than $2\lceil 2ed \rceil d$ events. Because $e \cdot \dfrac{1}{\lceil 2ed \rceil^2} \cdot 2\lceil 2ed \rceil d \le 1$, we know from Lemma 2.4 that $P\left(\bigwedge \overline{A_f}\right) > 0$. Thus, there exists a set of vertices $W$ which contains one vertex from each $V_i$, such that there does not exist an edge between any of the vertices in $W$. Therefore, $W$ is an independent set.

If there exists some $V_i$ such that $|V_i| > \lceil 2ed \rceil$, we know that there exists a subgraph $V_i'$ of $V_i$ such that $|V_i'| = \lceil 2ed \rceil$. We can then consider the subgraph $G[\bigcup_{1 \le i \le r} V_i']$. The above tells us that there exists some $W \subset \bigcup_{1 \le i \le r} V_i$ that is an independent set containing one vertex from each $V_i'$. This $W$ satisfies that $W \subset V$, is an independent set, and contains one vertex from each $V_i$. □

We are now ready to prove Lemma 6.4.

*Proof.* As Lemma 6.7 shows, $D$ can be partitioned into $d$ pairwise 1-regular spanning subgraphs, $D_1, D_2, \cdots, D_d$. Because each $D_i$ is 1-regular, it consists of a union of $n_i$ disjoint cycles. Let $n = \sum_{i=1}^{d} n_i$. Consider the subgraphs $V_1, V_2, \cdots, V_n$ where $V_i$ is a disjoint cycle in one of the $D_i$. Because the directed girth of $D$ is greater or equal to $8ed$, we know that the number of edges in each subgraph is greater or equal to $8ed$.

Consider the line graph $H$ of $D$, where each edge in $D$ corresponds to a vertex in $H$, and vertices in $H$ are connected if the corresponding edges in $D$ join the same vertex. It is fairly obvious that $H$ can be partitioned into subgraphs $H_1, H_2, \cdots, H_n$ where $H_i$ is the line graph of $V_i$ for all $1 \leq i \leq n$. Therefore, $H_i$ satisfies $|H_i| \geq 8ed$. Because $H$ is $(4d-2)$-regular, we have that $|H_i| \geq 8ed > 2e(4d-2)$. According to 6.8, there exists a set of vertices $W$ where $W$ is an independent set, and contains one vertex from each $H_i$. Consider the line graph of $H[W]$. It represents a matching on $D$ where there exists one and only one edge in each cycle $V_i$ that belongs to the matching, which we shall denote as $M$. Therefore, $M$ is a linear forest. Consider the subgraphs $D_i \setminus M$. We can see that the original cycles in $D_i$ are now paths in $D_i \setminus M$. Thus, $D_i \setminus M$ is also a linear forest. Thus, the collection of subgraphs $M, D_1 \setminus M, D_2 \setminus M, \cdots, D_d \setminus M$ is a collection of linear forests in $D$, whose union is the set of all edges in $D$. Thus, $dla(D) \leq d+1$.

We also know that $D$ has $|V| \cdot d$ edges, and each linear forest can have at most $|V|-1$ edges, or there would exist a cycle. Therefore, $dla(D) \geq \dfrac{|V| \cdot d}{|V|-1} > d$.

Overall, $dla(D) = d+1$. □

To deal with $d$-regular graphs with small girths, we shall show that, if $d$ is sufficiently large, the edges of the graphs can be decomposed in a way such that the resulting graphs are almost regular graphs with high girths.

This requires us to specify how to decompose the graph using the following lemma.

**Lemma 6.9.** *Let $D(V,E)$ be a $d$-regular digraph where $d$ is sufficiently large. Let $p$ be an integer satisfying $10\sqrt{d} \leq p \leq 20\sqrt{d}$. Then, there exists a $p$-coloring of the vertices of $D$: $f: V \to \{0, 1, \cdots, p-1\}$ such that for each vertex $v \in V$ and each integer $i$ with $0 \leq i \leq p-1$, the numbers*

$$N^+(v,i) = |\{u \in V \mid (v,u) \in E, f(u) = i\}|$$

*and*

$$N^-(v,i) = |\{u \in V \mid (v,u) \in E, f(u) = i\}|$$

*satisfy*

$$\left|N^+(v,i) - \frac{d}{p}\right|, \left|N^-(v,i) - \frac{d}{p}\right| \leq 3\sqrt{\frac{d}{p}}\sqrt{\log d}$$

*Proof.* Let $f$ be any random $p$-coloring of the vertices of $D$ with colors labelled $0, 1, \cdots, p-1$ such that $f$ follows a uniform distribution on the $p$ colors. For every vertex $v$ and every integer $i$, let $A_{v,i}$ be the probability that $N^+(v,i)$ does not satisfy the above inequality and $B_{v,i}$ be the probability that $N^-(v,i)$ does not satisfy the above inequality. Because the probability of any given vertex being colored by $i$ is $\dfrac{1}{p}$ for all $0 \leq i < p$, we know that $N^+(v,i)$ and $N^-(v,i)$ follow a binomial distribution with expectation $\dfrac{d}{p}$ and standard deviation $\sqrt{d \cdot \dfrac{1}{p}\left(1 - \dfrac{1}{p}\right)}$. Therefore, according to Chernoff bound, we can estimate an upper bound of the probability that $A_{v,i}$ holds.

$$\begin{aligned}
P\left(A_{v,i}\right) &< \exp\left(-\frac{d}{p}\cdot\left(\frac{3\sqrt{\frac{d}{p}}\sqrt{\log d}}{\sqrt{d\cdot\frac{1}{p}\left(1-\frac{1}{p}\right)}}\right)^2\cdot\frac{1}{2}\right) \\
&< e^{-\frac{d}{p}\cdot\left(3\sqrt{\log d}\right)^2\cdot\frac{1}{2}} \\
&= \left(\frac{1}{d}\right)^{\frac{9}{2}\cdot\frac{d}{p}} \\
&< \left(\frac{1}{d^4}\right)^{\frac{d}{p}}
\end{aligned}$$

For $d$ sufficiently large, we have that $P\left(A_{v,i}\right) < \frac{1}{d^4}$. According to the same logic, we have that $P\left(B_{v,i}\right) < \frac{1}{d^4}$.

We know that each of the events $A_{v,i}, B_{v,i}$ is only dependent with events $A_{u,j}, B_{u,j}$ if $u$ has a common neighbor with $v$ for any $0 \leq j < p$. Thus, the dependency graph on $A_{v,i}$ and $B_{v,i}$ has a maximum degree not exceeding $(2d)^2 \cdot p$. Because

$$e\cdot\frac{1}{d^4}\cdot\left((2d)^2\cdot p+1\right)\leq 1$$

for sufficiently large $d$, we know from Lemma 2.4 that there exists a coloring where none of $A_{v,i}, B_{v,j}$ holds, which also satisfies the statement in the lemma. □

Having this knowledge, we can now deal with any digraph.

**Theorem 6.10.** *There exists some $c > 0$ such that for a $d$-regular digraph $G$ with sufficiently large $d$, we have that*

$$dla\left(G\right)\leq d+c\cdot d^{\frac{3}{4}}\cdot\sqrt{\log d}$$

*Proof.* Let $G\left(V,E\right)$ be a $d$-regular digraph. According to the Bertrand-Chebyshev theorem, there must exist a prime number between $10\sqrt{d}$ and $\leq 20\sqrt{d}$. Let $p$ be that prime number. According to Lemma 6.9, there exists a vertex $p$-coloring $f: V \to \{0, 1, \cdots, p-1\}$ such that for each vertex $v \in V$ and every integer $i$ with $0 \leq i \leq p-1$, we have that $\left|N^{-}v, i - \frac{d}{p}\right|, \left|N^{+}v, i - \frac{d}{p}\right| \leq 3\sqrt{\frac{d}{p}}\sqrt{\log d}$. For each $i$, let $G_i\left(V_i, E_i\right)$ be defined as $V_i = V$ and $E_i = \{(u,v) \in E \mid f\left(v\right) = \left(f\left(u\right)+i\right) \bmod p\}$. It follows that the maximum indegree $d_i^-$ and maximum outdegree $d_i^+$ are both less or equal to $\frac{d}{p} + 3\sqrt{\frac{d}{p}}$. Thus, each $G_i$ is a subgraph of a $\left(\frac{d}{p}+3\sqrt{\frac{d}{p}}\right)$-regular digraph with vertices $V$.

(1) For $i > 0$, we also have that the directed girth $g_i$ of $G_i$ is greater or equal to $p$. When $d$ is sufficiently large, we know that $g_i \geq p > 8e\left(\frac{d}{p} + 3\sqrt{\frac{d}{p}}\right)$. According to Lemma 6.4, we have that $dla\,(G_i) \leq \frac{d}{p} + 3\sqrt{\frac{d}{p}} + 1$.

(2) For $G_0$, we can break down the $\left(\frac{d}{p} + 3\sqrt{\frac{d}{p}}\right)$-regular digraph, which $G_0$ is a subgraph of, into $\frac{d}{p} + 3\sqrt{\frac{d}{p}}$ 1-regular spanning trees and break the spanning trees into halves to gain paths. Here, $dla\,(G_0) \leq 2\frac{d}{p} + 6\sqrt{\frac{d}{p}}$.

We can now add the inequalities and obtain

$$\begin{aligned} dla\,(G) &\leq (p-1)\,dla\,(G_i)\,(\text{for } i > 0) + dla\,(G_0) \\ &= d + 2\frac{d}{p} + 3\sqrt{pd}\sqrt{\log d} + 3\sqrt{\frac{d}{p}}\sqrt{\log d} + p - 1 \\ &< d + c \cdot d^{\frac{3}{4}} \cdot \sqrt{\log d} \end{aligned}$$

for some constant $c > 0$. □

## Acknowledgments

I would like to express my utmost gratitudes to my mentor, Matthew Scalamandre, for the suggestion of such an interesting topic and guidance offered through the paper writing process, both about the content and the writing style. I would also like to extend my thanks to Professor Peter May and Daniil Rudenko for organizing such a wonderful REU and apprentice program.

## References

[1] N. Alon and J. H. Spencer. The Probabilistic Method. John Wiley & Sons, Inc. 2015.

[2] J. A. Bondy and U. S. R. Murty. Graph Theory with Applications. Elsevier Science Publishing Co. Inc. 1982.

[3] Apporva Khare, Jan Vondrak, The Lovasz Local Lemma, Math 233A, Non-constructive methods in combinatorics, Spring, 2018, Stanford University (https://theory.stanford.edu/ jvondrak/MATH233A-2018/Math233-lec02.pdf)

[4] Michel Goemans, Chernoff bounds, and some applications, 18.310, February, 2015, Massachusetts Institute of Technology (http://math.mit.edu/ goemans/18310S15/chernoff-notes.pdf)