M400 MSci Project - Discrete Isoperimetric Inequalities

by Henry Liu

Session 1999 - 2000
## Contents

0 Introduction .......................................................... 1

1 Combinatorial ideas .................................................. 3
   1.1 Harper’s vertex isoperimetric theorem .......................... 3
   1.2 The infinite grid and the finite grid ............................ 8
   1.3 The weighted cube .............................................. 17

2 Probabilistic ideas ................................................... 25
   2.1 Concentration of measure ........................................ 25
   2.2 Martingale techniques .......................................... 27

3 Product graphs and eigenvalue methods ........................... 36
   3.1 Products of graphs .............................................. 36
   3.2 Eigenvalue techniques .......................................... 41

4 The chromatic number of random graphs .......................... 48
   4.1 Random graphs, cliques and independent sets .................. 48
   4.2 The chromatic number of a.e. $G_{n,p}$ .......................... 53

A Appendix A  
   Pre-requisites from C395 (Graph Theory and Combinatorics) .......... 61

B Appendix B  
   Pre-requisites from C327 (Measure Theory) and C393 (Probability) .. 63

References ..................................................................... 69
0 Introduction

Discrete isoperimetric inequalities are analogous to the following classical result: *among all regions on the plane with a given area, the circle has the smallest perimeter*. In our situation, our metric space is a graph $G$, with the graph distance. For a subset $A$ of the vertex set $V(G)$ of $G$, the ‘area of $A$’ is the number of vertices in $A$ $(= |A|)$. The most common notion for the ‘perimeter of $A$’ is the (vertex) boundary of $A$, which is the set of all vertices of $G$ within a distance of 1 from $A$ (an alternative notion of ‘perimeter’ is the edge boundary, which will not be discussed here). More generally, if we replace ‘1’ by any $t \geq 0$, we have the $t$-boundary of $A$.

Hence our problem is to determine an inequality giving a lower bound for size of the $t$-boundary of $A$, given $|A|$. Such an inequality is a discrete isoperimetric inequality on $G$. As in the continuous case, for $A$ to have a small $t$-boundary, we expect $A$ to be a ‘nice’ set (ie: not scattered about, nor has an irregular shape, but ‘tightly packed together’). Ideally, we would like the inequality to be the best possible. However, in most situations, all we can hope for is a good estimate.

These discrete isoperimetric problems, as they are known, have a relatively short history (1960’s). The earliest methods of solving them were long and cumbersome. Recently, more and more elegant methods have been discovered to solve them. In this dissertation, the first three chapters will each present one type of these methods. We shall discuss purely combinatorial ideas, probabilistic ideas, and eigenvalue techniques.

Discrete isoperimetric inequalities have since become a very important branch of combinatorics. Not only that they answer many natural questions about graphs, but they also have a numerous amount of applications. In chapter 4, we look at a particular example concerning the theory of random graphs. We aim to apply the results obtained to one of the most notorious problems in the theory of random graphs: to determine the chromatic number of almost every random graph. Another notable area where the application of isoperimetic inequalities are useful is geometric functional analysis.

The pre-requisites for this dissertation are mainly the courses C395 (Graph Theory and Combinatorics) (for chapters 1, 3 and 4), C393 (Probability)/C327 (Measure Theory) (for chapters 2 and 4) along with a few basic principles of linear algebra (eg: M221, Algebra 3, for chapter 3). The reader is referred to appendices A and B for details whenever the definitions of these basic concepts are omitted, and are denoted by †.
The conventions used throughout this dissertation are fairly standard. Every notable result is numbered, and the symbol □ either denotes the end of the proof of a result, or indicates that no proof is given (along with a reference).

Figure 1  Structure of the dissertation
1 Combinatorial ideas

The common approach in a combinatorial proof of an isoperimetric inequality is the use of compressions. Loosely speaking, for a graph $G$ and $A \subset V(G)$, a compression is an operator on $A$ that replaces $A$ by another subset $A' \subset V(G)$ with the same size, and a smaller boundary. One then hopes to apply compressions in different ways to end up with a set $B \subset V(G)$ which can be seen that it is ‘almost’ the set with the smallest boundary.

Let us begin with some formal definitions.

Let $G$ be a graph. For $x, y \in V(G)$, let $d(x, y)$ denote the graph distance (usually called the Hamming distance), ie: the length of the shortest path from $x$ to $y$ in $G$. Here, we always let $G$ be a connected graph, and so $d(x, y)$ always exists. Unless otherwise stated, we write $A \subset G$ for $A \subset V(G)$. For $A \subset G$, we let $d(A, y) = \inf\{d(x, y) : x \in A\}$. Then for $t \geq 0$, the $t$-boundary of $A$ is

$$A(t) = \{y \in V(G) : d(A, y) \leq t\}.$$ 

Thus every vertex of $A(t) \subset G$ can be joined to some vertex of $A$ by a path of length at most $t$. We write $\partial A$ for $A(1)$, and call this the boundary of $A$.

So an isoperimetric problem is as follows. If $A \subset G$ and $|A| = m$, find a function $f_G(m, t)$, such that $|A(t)| \geq f_G(m, t)$. Often, $f_G(m, t)$ can only be a good estimate. Such an inequality is an isoperimetric inequality on $G$.

1.1 Harper’s vertex isoperimetric theorem

Let us now proceed with an important example: the discrete cube $Q_n$. We often label the vertices of $Q_n$ by elements of $\mathcal{P}(X)$, where $X$ is a set of size $n$. We often take $X = \{1, 2, \ldots, n\}$ and sometimes write $X = [n]$. Any subset of $\mathcal{P}(X)$ is a set system on $X$. For convenience, we can omit brackets and commas for elements of $\mathcal{P}(X)$. So for example, $\{1, 2, 3\}$ can be written as 123. We then join $x \in V(Q_n)$ to $y \in V(Q_n)$ iff for some $i \in X$, either $x \cup \{i\} = y$ or $y \cup \{i\} = x$. In other words, $x$ and $y$ are adjacent iff $|x \triangle y| = 1$. Alternatively, we can think of the vertices of $Q_n$ as 0 - 1 sequences of length $n$, with two vertices adjacent iff there is exactly one place where the two 0 - 1 sequences differ. In some situations, this is the preferred notion. We have $|Q_n| = 2^n$. Define

\footnote{See Appendix A for definitions}
1. Combinatorial ideas

\[ X^{(r)} = \{ x \in P(X) : |x| = r \} \], and \[ X^{(\leq r)} = \{ x \in P(X) : |x| \leq r \} \]; similarly for ‘\( \geq \)’, ‘\(<\)’, and ‘\(>\)’ in place of ‘\( \leq \)’. Any set of the form \( X^{(\leq r)} \) is called a **Hamming ball**.

\[ 0 \leq m \leq 2^n \], how shall we choose \( A \subset Q_n \) with \( |A| = m \) such that \( |\partial A| \) is minimal? Obviously we would like to ‘pack the elements of \( A \) as tightly as possible’. For example, for \( Q_4 \) and \( m = 5 \), one would choose a point with its 4 neighbours for \( A \), such as \( X^{(\leq 1)} \). Further experiment suggests that if \( m = |X^{(\leq r)}| \) for some \( r \), then one would choose \( A = X^{(\leq r)} \). If \( |X^{(\leq r)}| < m < |X^{(\leq r+1)}| \) for some \( r \), shall we take \( X^{(\leq r)} \subset A \subset X^{(\leq r+1)} \)?

This suggests that we should define an **ordering** for the elements of \( P(X) \), the simplicial order, in such a way that ‘initial segments of the ordering have the smallest boundary’. If \( x, y \in P(X) \), let \( x \) precede \( y \), written \( x < y \), if either \( |x| < |y| \) or \( |x| = |y| \) and \( \min(x \triangle y) \in x \), i.e. \( i \in x, i \not\in y \), where \( i \) is the least element of \( X \) for which \( x \) and \( y \) differ. For example, the simplicial ordering on \( Q_4 \) is \( \{ \emptyset, 1, 2, 3, 4, 12, 13, 14, 23, 24, 34, 123, 124, 134, 234 \} \).

Then Harper’s theorem says that ‘initial segments of simplicial ordering are best for \( |\partial A| \)’. Precisely:

**Theorem 1 (Harper’s theorem, 1966)** Let \( A \subset Q_n \) with \( |A| = m \) and let \( I \subset Q_n \) be the first \( m \) elements of \( Q_n \) in simplicial order. Then \( |\partial A| \geq |\partial I| \). Moreover, if \( |A| \geq \sum_{k=0}^{r+1} \binom{n}{k} \), then \( |\partial A| \geq \sum_{k=0}^{r+1} \binom{n}{k} \).

The ‘trick’ in proving Harper’s theorem is the idea of **compressions**. This allows us to avoid direct calculations of \( |\partial I| \) or \( |\partial A| \), which are often horrible unless \( m = |X^{(\leq r)}| \). Instead, ‘compressing \( A \)’ means replacing \( A \) by a set \( A' \) such that \( |A'| = |A| \) and \( |\partial A'| \leq |\partial A| \), and \( A' \) looks more similar to \( I \) than \( A \) did. By repeatedly compressing in different ways, we

![Figure 2](attachment:image.png)

*Figure 2* The discrete cube \( Q_4 \)
hope to end up with a ‘compressed set’ \( B \) for which we can see directly that \( |\partial B| \geq |\partial I| \).

We now formally define these compression operators.

Let \( A \subset \mathcal{P}(X) \) be a set system on \( X \), and \( 1 \leq i \leq n \). The \textit{i-sections} of \( A \) are the set systems on \( X \setminus \{i\} \) given by

\[
A_i^- = \{ x \in \mathcal{P}(X \setminus \{i\}) : x \subset A \} \subset Q_{n-1}^{(i^-)}, \\
A_i^+ = \{ x \in \mathcal{P}(X \setminus \{i\}) : x \cup \{i\} \subset A \} \subset Q_{n-1}^{(i^+)}.
\]

where \( Q_{n-1}^{(i^-)}, Q_{n-1}^{(i^+)} \) are each copies of \( Q_{n-1} \) labelled by sets of \( \mathcal{P}(X \setminus \{i\}) \).

For example, if \( A = \{1, 3, 12, 124, 134\} \subset \mathcal{P}([4]) \), then

\[
A_{2^-} = \{1, 3, 134\} \subset \mathcal{P}([4] \setminus \{2\}) = Q_3^{(2^-)}, \\
A_{2^+} = \{1, 4, 14\} \subset \mathcal{P}([4] \setminus \{2\}) = Q_3^{(2^+)}.
\]

Then the simplicial ordering on \( \mathcal{P}(X \setminus \{i\}) \) is just the simplicial ordering on \( \mathcal{P}(X) \) restricted to elements in \( \mathcal{P}(X \setminus \{i\}) \). We then define the \textit{i-compression} of \( A \) to be the set system \( C_i(A) \subset \mathcal{P}(X) \) by giving its \( i \)-sections

\[
C_i(A)_i^- = \text{The first } |A_{i^-}| \text{ points in simplicial ordering on } \mathcal{P}(X \setminus \{i\}), \\
C_i(A)_i^+ = \text{The first } |A_{i^+}| \text{ points in simplicial ordering on } \mathcal{P}(X \setminus \{i\}).
\]

Thus for example, if \( A = \{1, 3, 12, 24, 124, 134\} \subset \mathcal{P}([4]) \), then

\[
C_2(A)_2^- = \{\emptyset, 1, 3\} \subset \mathcal{P}([4] \setminus \{2\}) = Q_3^{(2^-)}, \\
C_2(A)_2^+ = \{\emptyset, 1, 3\} \subset \mathcal{P}([4] \setminus \{2\}) = Q_3^{(2^+)},
\]

so that \( C_2(A) = \{\emptyset, 1, 3\} \cup \{2, 12, 23\} \). We say that \( A \) is \textit{i-compressed} if \( C_i(A) = A \).

\begin{center}
\begin{tikzpicture}
\node (A) at (0,0) [draw=black, circle, thick, inner sep=1mm, fill=white] {\(A \subset Q_n\)};
\node (A+1) at (2,2) [draw=black, circle, thick, inner sep=1mm, fill=white] {\(A_i^+ \subset Q_{n-1}^{(i^+)}\)};
\node (C+1) at (5,2) [draw=black, circle, thick, inner sep=1mm, fill=white] {\(C_i(A)_i^+ \subset Q_{n-1}^{(i^+)}\)};
\node (C+2) at (8,0) [draw=black, circle, thick, inner sep=1mm, fill=white] {\(C_i(A) \subset Q_n\)};
\node (A-1) at (2,-2) [draw=black, circle, thick, inner sep=1mm, fill=white] {\(A_i^- \subset Q_{n-1}^{(i^-)}\)};
\node (C-1) at (5,-2) [draw=black, circle, thick, inner sep=1mm, fill=white] {\(C_i(A)_i^- \subset Q_{n-1}^{(i^-)}\)};
\draw[->] (A) to (A+1);
\draw[->] (A+1) to (C+1);
\draw[->] (A) to (C+1);
\draw[->] (A+1) to (C+2);
\draw[->] (C+1) to (C+2);
\draw[->] (A) to (A-1);
\draw[->] (A-1) to (C-1);
\draw[->] (A-1) to (A);
\draw[->] (A-1) to (C+2);
\end{tikzpicture}
\end{center}

\textbf{Figure 3} The \( i \)-compression of \( A \subset Q_n \)

We are now ready to give a beautiful proof of Harper’s theorem, due to Kleitman.

\textbf{Proof of Harper’s theorem} Clearly,

\[
|C_i(A)| = |C_i(A)_i^-| + |C_i(A)_i^+| = |A_i^-| + |A_i^+| = |A|.
\]
We first aim to show that $|\partial C_i(A)| \leq |\partial A|$. For convenience, write $C$ for $C_i(A)$. To show that $|\partial C| \leq |\partial A|$, by (1), it is enough to show that $|\partial C_i(A)| \leq |\partial A|$ and $|\partial C_i(A)| \leq |\partial A|$. 

Now by the definition of the boundary, we have:

$$\partial A_{i-} = \partial(A_{i-}) \cup A_{i+}, \quad (2)$$

$$\partial C_{i-} = \partial(C_{i-}) \cup C_{i+}, \quad (3)$$

Since the contribution to points in $(\partial A)_{i-}$ from $Q_{n-1}^{(i-)}$ comes from $\partial(A_{i-})$ (note that this ‘$\partial$’ is taken within $Q_{n-1}^{(i-)}$), and from $Q_{n-1}^{(i+)}$ comes from $A_{i+}$ (see figure 4); likewise for $C$.

Since $|C_{i-}| = |A_{i-}|$, and $C_{i-}$ is an initial segment of simplicial ordering on $\mathcal{P}(X \setminus \{i\})$, one can easily check by induction on $n$ that $|\partial(C_{i-})| \leq |\partial(A_{i-})|$ (certainly the theorem is trivial for $n = 1$, so that the induction does start). Also, we have $|C_{i+}| = |A_{i+}|$. Clearly, if $C_{i-}$ is an initial segment of simplicial order, then so is $\partial(C_{i-})$. Then since $\partial(C_{i-})$ and $C_{i+}$ are both initial segments of simplicial order on $\mathcal{P}(X \setminus \{i\})$, they are nested, ie: either $\partial(C_{i-}) \subset C_{i+}$ or $C_{i+} \subset \partial(C_{i-})$. By (3), $|\partial(C_{i-})| = \max\{|\partial(C_{i-})|, |C_{i+}|\}$. So by (2), either $|\partial(C_{i-})| = |\partial(A_{i-})| \leq |\partial(A_{i-})| \leq |\partial(A)|_{i-}$, or $|\partial(C_{i-})| = |C_{i+}| = |A_{i+}| \leq |\partial(A)|_{i-}$. Thus $|\partial(C_{i-})| \leq |\partial(A)|_{i-}$.

By symmetry between $Q_{n-1}^{(i-)}$ and $Q_{n-1}^{(i+)}$, an identical argument gives $|\partial C_{i+}| \leq |\partial A_{i+}|$. Thus we have $|\partial C| \leq |\partial A|$ with $|C| = |A|$, as required.

Now define a sequence $A_0 = A, A_1, \ldots$ of set systems on $\mathcal{P}(X)$ as follows. If $A_j$ is $i$-compressed for all $i$, then stop the sequence at $A_j$. Otherwise, if there is an $i$ such that $A_j$ is not $i$-compressed, define $A_{j+1} = C_i(A_j)$, and continue this inductively. This sequence of compressions must end at some $A_k$, because clearly, if the compression operator $C_i$ moves
1.1 Harper’s vertex isoperimetric theorem

a point of $A_j$, it must move the point to a position which is earlier in the simplicial order. This cannot happen infinitely often.

Hence the set system $B = A_k$ satisfies $|B| = |A|$, $|\partial B| \leq |\partial A|$, and $B$ is $i$-compressed for all $i$. But is it necessarily true that $B = I$? If so, then the proof of Harper’s theorem would be complete.

Unfortunately, this is not true. For example, $B = \{\emptyset, 1, 2, 12\} \subset P(\{3\})$ is $i$-compressed for $i = 1, 2, 3$, but $B$ is not an initial segment of simplicial order on $P(\{3\})$.

Luckily, we have the following lemma.

**Lemma 2** Let $B \subset P(X)$ be $i$-compressed for all $i$ which is not an initial segment of simplicial order on $P(X)$. Then either $n$ is odd and

$$B = X^{(< n/2)} \setminus \{(n + 3)/2, (n + 5)/2, \ldots, n\} \cup \{1, 2, \ldots, (n + 1)/2\},$$

or $n$ is even and

$$B = X^{(< n/2)} \cup \{x \in X^{(n/2)} : 1 \in x\} \setminus \{(n/2) + 2, (n/2) + 3, \ldots, n\} \cup \{2, 3, \ldots, (n/2) + 1\}.$$

So lemma 2 tells us that, for each $n$, there is at most one possibility for $B$ which is not an initial segment of simplicial order. If we know lemma 2, the proof of Harper’s theorem will be complete, because each of the two examples in lemma 2 has a greater boundary than the corresponding initial segment of simplicial order with the same size, namely, $X^{(< n/2)}$ or $X^{(< n/2)} \cup \{x \in X^{(n/2)} : 1 \in x\}$, respectively.

**Proof of lemma 2** Let $B$ be $i$-compressed for all $i$, but not an initial segment of the simplicial order. Then there are sets $x, y \in P(X)$ with $x < y$ in simplicial order, $x \notin B$ and $y \in B$.

Can we have $i \notin x$ and $i \notin y$ for some $i$? No, because then $x, y \in B_i$, so that $B_i$ is not an initial segment of the simplicial order, contradicting the fact that $B$ is $i$-compressed. Likewise, we cannot have $i \in x$ and $i \in y$. This holds for all $i$, so this implies that each $i$ belongs to either $x$ or $y$. Hence $y^c = X \setminus y = x$. This means that the only $z < y$ such that $z \notin B$ is $x$ (as such a $z$ satisfies $z = y^c$), and the only $z > x$ such that $z \in B$ is $y$ (as such a $z$ satisfies $z = x^c$). So $z \in B$ whenever $z < x$ and $z \notin B$ whenever $z > y$. If $x < z < y$, then we cannot have $z \in B$ (since $z > x$), and we also cannot have $z \notin B$ (since $z < y$): a contradiction.

$$\cdots \quad x \quad \ldots \quad y \quad \cdots$$

$$\in B \quad \cdots \notin B$$
1. Combinatorial ideas

So \( B = \{ z \in \mathcal{P}(X) : z \leq y \} \setminus \{ x \} \), where \( y \) is the immediate successor of \( x \), and \( y^c = x \).

So if \( n \) is odd, then \(|x| = |y| + 1\), so that \( x \) is the last set of \( X^{((n-1)/2)} \). If \( n \) is even, then we have \(|x| = |y| = n/2\), and \( x, y \in X^{(n/2)} \). We see that \( x = \{1, (n/2) + 2, (n/2) + 3, \ldots, n\} \), \( y = \{2, 3, \ldots, (n/2) + 1\} \) are the only sets satisfying the conditions, since this is the only occasion that 1 belongs to the symmetric difference of successive sets in \( X^{(n/2)} \), and we require such a symmetric difference to be \( X \) (since \( y^c = x \)). This completes the proof of lemma 2, and hence of Harper’s theorem.

From Harper’s theorem, one can deduce the following result concerning \( t \)-boundaries.

**Corollary 3** Let \( A \subset Q_n \) with \(|A| \geq \sum_{r=0}^{n} \binom{n}{r}\). Then \(|A_{(t)}| \geq \sum_{k=0}^{t} \binom{n}{k}\), \( \forall t = 0, 1, \ldots \).

**Proof** \( t = 0 \) is trivial. \( t = 1 \) is Harper’s theorem. If \(|A_{(0)}| \geq \sum_{k=0}^{t} \binom{n}{k}\), then \(|A_{(s+1)}| = |\partial (A_{(s)})| \geq \sum_{k=0}^{r+s+1} \binom{n}{k}\) by Harper’s theorem. Result follows by induction. \( \square \)

What happens in corollary 3 when \( n \) is large? We can use estimates on the tail of the binomial distribution (see, eg, [5, ch. 1]) to get the following:

**Corollary 4** Let \( A \subset Q_n \) with \(|A| \geq 2^{n-1}\). Then \( \forall t = 0, 1, \ldots \),

\[
|A_{(t)}| / 2^n \geq 1 - \exp(-2t^2/n). \tag{4}
\]

\( \square \)

Sometimes, an inequality in the form of corollary 3 or corollary 4 is called Harper’s inequality. An inequality in the form of (4) has an enormous amount of applications. We shall see some of these applications in chapters 2 and 4.

### 1.2 The infinite grid and the finite grid

We now turn our attention to the isoperimetric problems on the infinite grid and the finite grid, on which we hope to apply similar compression ideas.

The \( n \)-dimensional infinite grid is the graph with vertices labelled by vectors in \( n \) dimensions with non-negative integer coordinates. We join \( x \) and \( y \) iff for some \( i \), \(|x_i - y_i| = 1\) and \( x_j = y_j \) for all \( j \neq i \). We denote the infinite grid by \( \mathbb{Z}^n = \{ x \in \mathbb{Z}^n : x_i \geq 0, \forall i \} \). We will define product graphs in chapter 3, but we remark for now that \( \mathbb{Z}^n \) is the product of \( n \) copies of the infinite path\(^1\) \( \mathbb{Z}_+ \).

\( \text{\(^1\)}\)See appendix A for definition
1.2 The infinite grid and the finite grid

If we think about compressing a set \( A \subset \mathbb{Z}^n_+ \), we would again like to define an ordering on \( \mathbb{Z}^n_+ \). Then we would try take a partition of \( A \) (like the \( i \)-sections for \( Q_n \)), and compress \( A \) in the same way as before. To do this, we need to introduce a fair amount of terminology.

Let \( \{e_1, \ldots, e_n\} \) denote the standard basis of \( \mathbb{Z}^n_+ \). A set \( A \subset \mathbb{Z}^n_+ \) is a down-set if \( \forall x \in A \) and \( y \in \mathbb{Z}^n_+ \) such that \( y_i \leq x_i, \forall i \), we have \( y \in A \). Hence \( A \) is a down-set iff whenever \( x \in A \) and \( x - e_i \in \mathbb{Z}^n_+ \), we have \( x - e_i \in A \) (see figure 6).

If \( x \in \mathbb{Z}^n_+ \), define the positive support of \( x \) by \( x_+ = \{1 \leq i \leq n : x_i > 0\} \) and the negative support of \( x \) by \( x_- = \{1 \leq i \leq n : x_i < 0\} \).

For a set \( I \subset X = \{1, \ldots, n\} \), let \( \bar{I} = X \setminus I \). Define \( \mathbb{Z}^{I,n}_+ = \{x \in \mathbb{Z}^n_+ : x_i = 0, \forall i \in \bar{I}\} \). So in particular, \( \mathbb{Z}^{I,n}_+ = \{x \in \mathbb{Z}^n_+ : x_i = 0\} \) for \( 1 \leq i \leq n \).

For \( A \subset \mathbb{Z}^n_+ \), \( x \in \mathbb{Z}_+^{I,n} \), define the \( i \)-section of \( A \) at \( x \) by \( A_i(x) = \{\lambda \in \mathbb{Z}_+ : x + \lambda e_i \in A\} \). This means that to obtain \( A_i(x) \), we take the line in \( \mathbb{Z}^n_+ \) in the direction of \( e_i \) through \( x \), then take the points on this line that belong to \( A \), and then take their \( i \)th coordinates.

\[ Z_{+}^{i,n} = \{x : x_i = 0\} \]

\[ A_i(x) = \{2, 5, 6\} \]

Figure 5 The infinite grid \( \mathbb{Z}^2_+ \)

Figure 6 A down set

Figure 7 The \( i \)-section of \( A \) at \( x \)
Now we can consider the isoperimetric problem on $\mathbb{Z}_n^+$. This time, which sets $A \subset \mathbb{Z}_n^+$ with a given size have minimum boundary? Obviously we want to pack the vertices tightly around the origin. A little experiment on $\mathbb{Z}_2^+$ seems to suggest that sets ‘bounded below by $x_1 + x_2 = r$ as tightly as possible’, for some $r$, are best. So we claim that for $A \subset \mathbb{Z}_n^+$, $A$ has the minimum boundary when it is ‘bounded under some simplex $\sum x_i = r$ as tightly as possible’. Therefore, we shall define an ordering on $\mathbb{Z}_n^+$, whose initial segments look like these ‘best sets’ for $A$.

Let $x, y \in \mathbb{Z}_n^+$. Define the simplicial ordering on $\mathbb{Z}_n^+$ by letting $x < y$ if either $\sum x_i < \sum y_i$, or $\sum x_i = \sum y_i$ and for some $j$, we have $x_j > y_j$ and $x_i = y_i$, $\forall i < j$.

For example, the simplicial ordering on $\mathbb{Z}_3^+$ is

$$\{(0,0,0), (1,0,0), (0,1,0), (0,0,1), (2,0,0), (1,1,0), (1,0,1), (0,2,0), (0,1,1), (0,0,2), (3,0,0), \ldots\}.$$ 

For $m = 0, 1, \ldots$, define $\partial^{(n)}(m)$ to be the size of the boundary of the first $m$ points in the simplicial order on $\mathbb{Z}_n^+$.

We aim to show that initial segments of simplicial order have the smallest boundary. That is, $|\partial A| \geq \partial^{(n)}(|A|)$ for all $A \subset \mathbb{Z}_n^+$. For example, for $A \subset \mathbb{Z}_2^+$ with $|A| = 12$, we would take the set in figure 8 to get $|\partial A| \geq 18$, which is the best possible.

![Figure 8](image)

Now we define our compressions. For $A \subset \mathbb{Z}_n^+$ and $1 \leq i \leq n$, define $C_i(A) \subset \mathbb{Z}_n^+$, the $i$-compression of $A$, by giving its $i$-sections

$$C_i(A)_i(x) = \{0,1,\ldots,|A_i(x)|-1\}, \quad x \in \mathbb{Z}_{i,n}^+, A_i(x) \neq \emptyset.$$ 

So we may think of an $i$-compression as if $A$ is ‘falling in the direction of $-e_i$ and jumbled up on $\mathbb{Z}_{i,n}^+$’ (see figure 9). We say that $A$ is $i$-compressed if $C_i(A) = A$. Note that $A$ is $i$-compressed iff $\forall x \in A$, $x - e_i \in \mathbb{Z}_n^+ \Rightarrow x - e_i \in A$. Clearly, $A$ is $i$-compressed $\forall i$ iff $A$ is a down-set. Note also that $|C_i(A)| = |A|$. 

1. Combinatorial ideas
1.2 The infinite grid and the finite grid

We are now in a position to prove the isoperimetric problem on the infinite grid, due to Wang and Wang. We will do this in a series of lemmas.

**Lemma 5** Let $A \subset \mathbb{Z}_+^n$.

(i) If $1 \leq i \leq n$, then $|\partial C_i(A)| \leq |\partial A|$. (ii) There is a down-set $B \subset \mathbb{Z}_+^n$ with $|B| = |A|$ and $|\partial B| \leq |\partial A|$.

**Proof** (i) Write $C$ for $C_i(A)$ for convenience. To show that $|\partial C| \leq |\partial A|$, it is enough to show that $|\partial C| \leq |\partial A|$ for each $x \in \mathbb{Z}_+^{i,n}$, since $|\partial A| = \sum_{x \in \mathbb{Z}_+^{i,n}} |\partial A|$, and similarly for $C$.

For any $x \in \mathbb{Z}_+^{i,n}$, we have

$$\partial A_i(x) = \partial A_i(x) \cup \bigcup \left \{ A_i(y) : d(y, x) = 1, y \in \mathbb{Z}_+^{i,n} \right \},$$

with the ‘$\partial$’ on the r.h.s. taken over $\mathbb{Z}_+$. This is because if $z \in (\partial A)_i(x)$, either $d(z, z') \leq 1$ for some $z' \in A_i(x)$, or $d(z, z'') = 1$, where for some $y$ with $d(y, x) = 1$, $z'' \in A_i(y)$ (see figure 10). A similar expression holds with $C$ in place of $A$.

We have the sets $\partial(C_i(x))$ and the sets $C_i(y), d(y, x) = 1$, are initial segments of $\mathbb{Z}_+$, so they are nested. So by (5) (with $C$ in place of $A$)

$$|\partial C_i(x)| = \max \left \{ |\partial C_i(x)|, \max \left \{|C_i(y)| : d(y, x) = 1, y \in \mathbb{Z}_+^{i,n} \right \} \right \}. \quad (6)$$
1. Combinatorial ideas

So by (5) and (6), either \(|(\partial C)_i(x)| = |C_i(y)| = |A_i(y)| \leq |(\partial A)_i(x)|\), for some \(y \in \mathbb{Z}^n_+\) with \(d(y, x) = 1\), or \(|(\partial C)_i(x)| = |(\partial C_i(x))| = |A_i(x)| + 1 \leq |(\partial A_i(x))| \leq |(\partial A)_i(x)|\), if \(C_i(x) \neq \emptyset\); clearly this would not be the case if \(C_i(x) = \emptyset\).

So \(|(\partial C)_i(x)| \leq |(\partial A)_i(x)|\) as required. Hence \(|\partial C| \leq |\partial A|\).

(ii) The set \(C_i(C_i(A))\) is \(i\)-compressed. This can easily be checked by looking at the \(i\)th and \(j\)th coordinates of the points of \(A\), while keeping the rest fixed. So setting \(B = C_n(C_{n-1}(\ldots C_1(A) \ldots))\), \(B\) is \(i\)-compressed for all \(i\), so that \(B\) is a down-set. Also, \(|B| = |A|\), since \(|C_i(A)| = |A|\) for all \(i\). So by (i), we have \(|\partial B| \leq |\partial A|\).

So having got \(A\) down to a down-set \(B\), how can we get from \(B\) to a set with a smaller boundary? For example, in \(\mathbb{Z}^2_+\), how can we get from a rectangular set to the triangular set (initial segment of simplicial order) that we want (see figure 11)?

![Figure 11](image-url)

Naturally, we need to compress our down-set \(B\) in the direction of some carefully chosen arbitrary vector \(v\), so that the boundary would further decrease. Hopefully then, one can get to the initial segment of simplicial order, and show that the boundary cannot further decrease. The compression operator \(C_i\) compresses a set in the direction of \(-e_i\). We now wish to define a compression operator \(C_v\) that compresses a set in the direction of a vector \(v \in \mathbb{Z}^n\). Obviously, we need \(v_- \neq \emptyset\).

Write \(\mathbb{Z}^{v,n}_+\) for \(\{x \in \mathbb{Z}^n_+ : x + v \notin \mathbb{Z}^n_+\}\). So \(x \in \mathbb{Z}^{v,n}_+\) iff \(\exists i \in v_-, 0 \leq x_i \leq -v_i - 1\) (as then \(x_i + v_i \leq -1\)). It is easy to verify that as \(x\) ranges over \(\mathbb{Z}^{v,n}_+\), the lines \(\{x - \lambda v : \lambda \in \mathbb{Z}_+\} \cap \mathbb{Z}^n_+\) form a partition of \(\mathbb{Z}^{v,n}_+\). We wish to compress a set in the direction of these lines. In a similar way to \(C_i\), for \(A \subset \mathbb{Z}^n_+\), \(x \in \mathbb{Z}^{v,n}_+\), define the \(v\)-section of \(A\) at \(x\) by \(A_v(x) = \{\lambda \in \mathbb{Z}_+ : x - \lambda v \in A\}\). Define the \(v\)-compression of \(A\) by giving its \(v\)-sections

\[
C_v(A)_v(x) = \{0, 1, \ldots, |A_v(x)| - 1\}, \quad x \in \mathbb{Z}^{v,n}_+, A_v(x) \neq \emptyset.
\]
1.2 The infinite grid and the finite grid

We say that $A$ is $v$-compressed if $C_v(A) = A$. So $A$ is $v$-compressed iff $\forall x \in A$ and $x + v \in \mathbb{Z}_+^n$, we have $x + v \in A$.

It is clear that we want to choose $v$ so that $C_v$ does not 'move the points of a set too far apart'. For example, we do not want $v$ to have a large positive component (see figure 13).

We have the following two important results, both with some key properties of $v$. Lemma 6 is a somewhat tedious, but strong result, and whose proof tells us why $v$ is chosen to have the two imposed properties. Lemma 7 is a fairly straightforward result that relates the $v$-compressions to the simplicial ordering on $\mathbb{Z}_+^n$.

**Lemma 6** Let $A \subset \mathbb{Z}_+^n$ be a down-set, and $v \in \mathbb{Z}^n$ with $v_- \neq \emptyset$, $v_+ \neq \emptyset$. Suppose that

(i) for each $i \in v_+$, $A$ is $(v - e_i)$-compressed,

(ii) for each $i \in v_-$, there is a $j \in v_+$ such that $A$ is $(v + e_i - e_j)$-compressed.

Then $C_v(A)$ is a down-set, and $|\partial C_v(A)| \leq |\partial A|$.

**Proof** Write $C$ for $C_v(A)$ for convenience. To show that $C$ is a down-set, we shall prove that for all $y \in C$ and $y - e_i \in \mathbb{Z}_+^n$, we have $y - e_i \in C$. There are unique $x \in \mathbb{Z}_+^{v,n}$, $\lambda \in \mathbb{Z}_+$ such that $y = x - \lambda v$. So we have $|C_v(x)| \geq \lambda + 1$, and $|A_v(x)| = |C_v(x)| \geq \lambda + 1$.

If $i \notin v_- \cup v_+$, then $x - e_i \in \mathbb{Z}_+^{v,n}$ (since $x_i - e_i = x_i - e_i$ and $(x - e_i)_j = x_j \forall j \neq i$). Since $A$ is a down-set, $|A_v(x - e_i)| \geq |A_v(x)|$, and thus $|C_v(x - e_i)| \geq \lambda + 1$. Hence $y - e_i \in C$, since $C_v(x - e_i)$ is an initial segment of $\mathbb{Z}_+$ and $y - e_i = (x - e_i) - \lambda v$. 

![Figure 12 A v-compression](image)

![Figure 13](image)
If \( i \in v_- \) and \( x - e_i \in \mathbb{Z}_+^n \), then \( x_i \geq 1 \) so \( x - e_i \in \mathbb{Z}_+^{v,n} \), and the same argument applies.

If \( x_i = 0 \), then \( x - e_i - v \in \mathbb{Z}_+^{v,n} \). Since \( A \) is a down-set, \( |A_v(x - e_i - v)| + 1 \geq |A_v(x)| \), so that \( |C_v(x - e_i - v)| \geq \lambda \). Hence as before, \( y - e_i = (x - e_i - v) - (\lambda - 1)v \in C \).

If \( i \in v_+ \), then \( x - e_i \in \mathbb{Z}^{v,n}_+ \) (same argument as before). Set \( \nu = \max \{ \mu : x - \mu v \in \mathbb{Z}_+^n \} \).

If \( (x - \nu v) - e_i \in \mathbb{Z}_+^n \), then \( A \) is a down-set \( \Rightarrow |A_v(x - e_i)| \geq |A_v(x)| \), and the same argument again applies. If \( (x - \nu v) - e_i \notin \mathbb{Z}_+^n \), then, since \( A \) is a down-set,

\[
x - \mu v \in A \Rightarrow (x - \mu v) - e_i \in A \quad \text{for} \quad 0 \leq \mu \leq \nu - 1.
\]

Also, condition (i) gives

\[
x - \mu v \in A \Rightarrow (x - \mu v) + (v - e_i) \in A \quad \text{for} \quad 1 \leq \mu \leq \nu.
\]  

(7) If \( |A_v(x)| \leq \nu \), then \( 0 \leq \mu' \leq \nu \) with \( \mu' \notin A_v(x) \). So (7) and (8) together induce the injection \( \psi : A_v(x) \to A_v(x - e_i) \) given by \( \psi(\mu) = \mu \) for \( 0 \leq \mu < \mu' \) and \( \psi(\mu) = \mu - 1 \) for \( \mu' < \mu \leq \nu \). Hence \( |A_v(x - e_i)| \geq |A_v(x)| \) again (see figure 14). If \( |A_v(x)| = \nu + 1 \), then \( A_v(x) = \{0, 1, \ldots, \nu\} = C_v(x) \), so \( A_v(x - e_i) = C_v(x - e_i) \). In every case, we have \( y - e_i \in C \).

Hence \( C \) is a down-set.

\[\begin{array}{c}
\begin{array}{c}
\mathbb{Z}_+^{v,n} \\

\end{array}
\end{array}\]

Figure 14

Now to prove that \( |\partial C| \leq |\partial A| \), fix \( x \in \mathbb{Z}_+^{v,n} \). It is enough to show that \( |(\partial C)_v(x)| \leq |(\partial A)_v(x)| \). Since \( A \) is a down-set, contribution to points in \( (\partial A)_v(x) \) come from \( v \)-sections at \( x \) and at points whose ‘generating lines are 1 below in each direction’.

That is, \( (\partial A)_v(x) = A_v(x) \cup \bigcup_{i=1}^{n} N_i(A) \), where

\[
N_i(A) = \begin{cases} 
A_v(x - e_i) & \text{if} \quad x - e_i \in \mathbb{Z}_+^{v,n}, \\
A_v(x - e_i - v) + 1 & \text{if} \quad x - e_i - v \in \mathbb{Z}_+^{v,n}, \\
\emptyset & \text{otherwise},
\end{cases}
\]

where \( S + 1 = \{ s + 1 : s \in S \} \). For example, in \( \mathbb{Z}_+^3 \) (see figure 15):
A similar expression holds for $C$ in place of $A$ (as $C$ is also a down-set). This time, the set $C_v(x)$ and the sets $C_v(x-e_i)$ are all initial segments of $\mathbb{Z}_+$, while the sets $C_v(x-e_i-v)+1$ are all initial segments of $\mathbb{Z}_+ \setminus \{0\}$. Thus

$$|(\partial C)_v(x)| = \max \left\{ |C_v(x)|, \max_i \{|C_v(x-e_i)|, |C_v(x-e_i-v)+1|+1\} \right\}. \quad (9)$$

From (9), if $|(\partial C)_v(x)| \neq |C_v(x-e_i-v)+1|+1$, $\forall i$, then we would be done by the same argument as in lemma 5(i). If $|(\partial C)_v(x)| = |C_v(x-e_i-v)+1|+1 = |C_v(x-e_i-v)|+1$ for some $i$, then it is enough to show that $|(\partial A)_v(x)| \geq |A_v(x-e_i-v)|+1$. In this case, $x-e_i-v \in \mathbb{Z}_+^{0,n}$ and $A_v(x-e_i-v) \neq \emptyset$, and this means that $i \in v_-$. Now

$$\mu \in A_v(x-e_i-v) \Rightarrow \mu + 1 \in (\partial A)_v(x), \quad (10)$$

and by condition (ii), for some $j \in v_+$, we have

$$(x-e_i-v) - \mu v \in A \quad \Rightarrow \quad (x-e_i-v) - \mu v + (v+e_i-e_j) = x-e_j - \mu v \in A$$

$$\Rightarrow \quad x - \mu v \in \partial A$$

$$\Rightarrow \quad \mu \in (\partial A)_v(x). \quad (11)$$

From (10) and (11), define the injection $\phi : A_v(x-e_i-v) \rightarrow (\partial A)_v(x)$ by $\phi(\mu) = \mu$. Then $\nu+1 \in (\partial A)_v(x)$, but $\nu+1 \notin \phi(A_v(x-e_i-v))$, where $\nu = \max \{ \mu : (x-e_i-v) - \mu v \in A \}$ (see figure 16). Thus $|(\partial A)_v(x)| \geq |A_v(x-e_i-v)|+1$ as required. \qed
Lemma 7 Let \( V = \{ v \in \mathbb{Z}^n : v_-, v_+ \neq \emptyset \} \) and either \( \sum v_i < 0 \) or \( \sum v_i = 0 \) and \( \min v_+ < \min v_- \). Then a down-set \( A \subset \mathbb{Z}_+^n \) is an initial segment of simplicial order on \( \mathbb{Z}_+^n \) iff \( A \) is \( \nu \)-compressed \( \forall v \in V \).

Proof Fix arbitrary \( x \in A, v \in V, \) with \( x+v \in \mathbb{Z}_+^n \). If \( \sum v_i < 0 \), then \( \sum (x+v)_i < \sum x_i \). If \( \sum v_i = 0 \) and \( \min v_+ < \min v_- \), then \( \sum (x+v)_i = \sum x_i \) and \( \min \{ i : (x+v)_i > x_i \} < \min \{ i : x+v_i < x_i \} \). So if \( A \) is an initial segment of simplicial order, then \( x+v \in A \), and \( A \) is \( \nu \)-compressed.

Conversely, suppose that \( A \) is \( \nu \)-compressed \( \forall v \in V \). Fix \( x \in A \). We shall show that \( \forall y < x \) in simplicial order on \( \mathbb{Z}_+^n, \ y \in A \). If \( \sum y_i < \sum x_i \), then \( (y-x)_- \neq \emptyset \). So either \( y_i \leq x_i, \forall i \) (when \( (y-x)_+ = \emptyset \)), or \( y-x \in V \), ie: \( y \in A \) in both cases. If \( \sum y_i = \sum x_i \) and for some \( j, y_j > x_j \) and \( y_i = x_i, \forall i < j \), then \( (y-x)_-, (y-x)_+ \neq \emptyset \), \( \sum (y-x)_i = 0 \) and \( \min (y-x)_+ < \min (y-x)_- \). So again we have \( y-x \in V \), and \( y \in A \).

\[ \square \]

Theorem 8 (Wang and Wang, 1977) If \( A \subset \mathbb{Z}_+^n \), then \( |\partial A| \geq \partial^{(n)}(|A|) \).

Proof By lemma 5, it is enough to prove the theorem when \( A \) is a down-set. Define a sequence of sets \( A_0 = A, A_1, \ldots \) as follows. If for some \( s, A_s \) is \( \nu \)-compressed \( \forall v \in V \), where \( V \) is as in lemma 7, then stop the sequence at \( A_s \). Otherwise, we claim that \( \exists u \in V \) such that \( A_s \) is not \( \nu \)-compressed, and \( u \) satisfies conditions (i) and (ii) of lemma 6 with \( A_s \) in place of \( A \). Choose any \( u \in \{ v : \sum |v_i| \) is minimal among the \( v \in V \) for which \( A_s \) is not \( \nu \)-compressed \}. Then \( A_s \) is \( \nu \)-compressed \( \forall v \in V \) with \( \sum |v_k| + 1 \leq \sum |u_k| \). So \( \forall i \in u_+ \), clearly \( u-e_i \in V \), and \( \sum |(u-e_i)_k| + 1 \leq \sum |u_k| \), so (i) holds. For \( i \in u_- \), choose \( j = \max u_+ \). Then it is easily seen that \( u + e_i - e_j \in V \cup \{0 \} \), and \( \sum |(u + e_i - e_j)_k| + 1 \leq \sum |u_k| - 1 \), and so (ii) holds. Set \( A_{s+1} = C_u(A_s) \), and continue inductively.

The sequence \( A_0, A_1, \ldots \) has to end at some \( A_t \). It is easy to see that \( \forall v \in V \), if \( C_v \) moves a point, it must move it to a position which is earlier in the simplicial order. This cannot happen infinitely often. So the set \( A_t \) is a down-set, with \( |A_t| = |A|, |\partial A_t| \leq |\partial A| \), and is \( \nu \)-compressed \( \forall v \in V \). So by lemma 7, \( A_t \) is an initial segment of simplicial order on \( \mathbb{Z}_+^n \).

Now we turn our attention to the isoperimetric problem on the finite grid. Here, as in any subsequent problems concerning the finite grid, we adopt the unusual, auxiliary notation \( [k] = \{ 0, 1, \ldots, k-1 \} \). The \( n \)-dimensional finite grid \( [k]^n \) is the induced subgraph of \( \mathbb{Z}_+^n \), whose vertices have coordinates in \( [k] \). Notice that \( [2]^n \) is the discrete cube \( Q_n \). We have analogous notations for objects associated with \( \mathbb{Z}_+^n \) for \( [k]^n \), just with \( \mathbb{Z}_+ \) replaced by \([k]\).
The weighted cube

For example, \([k]^I = \{x \in [k]^n : x_i = 0, \forall i \in I\}\).

The isoperimetric problem on the infinite grid is, in its own right, interesting. However, it is generally not quite as powerful as the result for the finite grid, in terms of applications. Nevertheless, its proof does provide many key ideas to the proof of the result for the finite grid.

We define the simplicial order on \([k]^n\) to be the simplicial order on \(\mathbb{Z}^n\) restricted to \([k]^n\). For \(m = 0, 1, \ldots\), define \(\partial_k^{(n)}(m)\) to be the size of the boundary of first \(m\) points in simplicial order on \([k]^n\). Then surprisingly, the isoperimetric inequality on \([k]^n\) is:

\[\text{Theorem 9} \quad \text{If } A \subset [k]^n, \text{ then } |\partial A| \geq \partial_k^{(n)}(|A|).\]

Theorem 9 is surprising in a sense that, the outer edges of \([k]^n\) make no difference compared to theorem 8. The proof of theorem 9 is analogous to the one used for the infinite grid, but more tedious, especially for an analogue of lemma 6. See [14] for a full proof.

Since the boundary of an initial segment of simplicial order is again an initial segment, one can easily deduce the following result concerning \(t\)-boundaries (by a similar induction argument to that of corollary 3):

\[\text{Corollary 10} \quad \text{Let } A, I \subset [k]^n, \text{ where } I \text{ is the first } |A| \text{ elements of } [k]^n \text{ in simplicial order. Then } \forall t = 0, 1, 2, \ldots, |A_{(t)}| \geq |I_{(t)}|.\]

With corollary 10, the following problem is of particular interest. For two sets \(A, B \subset [k]^n\) of given sizes, ‘how far apart can they be?’ In other words, how big can \(d(A, B) = \inf \{d(x,y) : x \in A, y \in B\} \) be?

\[\text{Corollary 11} \quad \text{There exist sets } A, B \subset [k]^n \text{ with } |A| = r, |B| = s \text{ and } d(A, B) \geq d \text{ iff } d(I, J) \geq d, \text{ where } I, J \subset [k]^n \text{ are the first } r \text{ and the last } s \text{ elements of } [k]^n \text{ in simplicial order, respectively.}\]

Corollary 11 is obvious with a little thought. It is actually a result related to what we will discuss in chapter 3.

1.3 The weighted cube

In this section, we focus on what is perhaps the most important graph in probabilistic combinatorics, the weighted cube.

For \(0 < p < 1\), \(X = [n]\), the weighted cube consists of the graph of \(Q_n\), equipped with the probability measure\(^1\) \(P : \mathcal{P}(X) \rightarrow [0, 1] \) given by \(P(x) = p^{|x|}(1-p)^{n-|x|}\) (ie: \(p\) is the

\(^1\)See Appendix B for definition
Combinatorial ideas

(constant) probability of any \( i \in X \) being chosen). For a set system \( A \subset \mathcal{P}(X) \), \( \mathcal{P} \) induces \( \mathcal{P}(A) = \sum_{x \in A} \mathcal{P}(x) \). We write \( Q_n(p) \) for this space. Thus \( Q_n(1/2) \) is just \( Q_n \) equipped with the uniform distribution.

More generally, for \( r_1, \ldots, r_n > 0 \), the weighted cube with ratios \( r_1, \ldots, r_n \) consists of \( Q_n \), equipped with the measure\(^1\) \( w : \mathcal{P}(X) \to \mathbb{R} \) given by \( w(x) = \prod_{i \in x} r_i \). For a set system \( A \subset \mathcal{P}(X) \), \( w \) induces \( w(A) = \sum_{x \in A} w(x) \), the weight of \( A \). Note that \( w(\emptyset) = 1 \) and

\[
Q_n(r_1, \ldots, r_n) \text{ for this space.}
\]

For the case \( r_1 = \cdots = r_n = q \), setting \( p = w/(1+q)^n \) and \( p = q/(1+q) \) gives us \( Q_n(p) \). \( Q_n(p) \) has numerous applications in probabilistic combinatorics. We will see an application of \( Q_n(p) \) to the theory of random graphs in chapter 4.

We aim to tackle the isoperimetric problem on \( Q_n(r_1, \ldots, r_n) \), and then we can deduce the result for \( Q_n(p) \). The isoperimetric problem on \( Q_n(r_1, \ldots, r_n) \) is as follows. For \( A \subset Q_n(r_1, \ldots, r_n) \) with \( w(A) \) fixed, which \( A \) has minimal \( w(\partial A) \)?

We aim to show that, among all down-sets \( A \) of given size, Hamming balls are still the best. This is somehow very surprising, in a sense that it is a generalisation of Harper’s theorem. Because we have many parameters \( r_1, \ldots, r_n \), we require an operator on the subsets which is analogous to the compression operator, the symmetrisation operator. Roughly speaking, symmetrisation is just a different style of compressing our subsets, in a somewhat ‘refined’ manner.

So how may we go about compressing our sets? This time, the proof that we used for Harper’s theorem is hopeless. We see that, for a down-set \( A \) of given weight, there may be very few set systems with the same weight as \( A \). For this reason, we aim to employ a ‘trick’, by defining a stronger notion of set systems: fractional set systems. These allow us to consider \( \text{weight} \) to be a continuous function, hence providing a better ‘freedom of movement’ for our compression (symmetrisation) operators. We then have to correspondingly define a new notion for boundary, in such a way that we can prove a stronger result.

A fractional set system, or simply a system on \( X \), is a function \( f : \mathcal{P}(X) \to [0,1] \). Intuitively, \( f \) measures the ‘fraction’ of each set \( x \in \mathcal{P}(X) \) that is in some set system \( A \subset \mathcal{P}(X) \), ie: \( f(x) = \alpha \) means that ‘\( \alpha \) of \( x \) is in \( A \)’. So this is a generalisation of a set system: if \( f(\mathcal{P}(X)) \subset \{0,1\} \), then we naturally identify \( f \) with a set system \( A \) by \( A = f^{-1}(1) \) (in this case, \( f \) is a non-fractional set system). We say that \( f \) is monotone decreasing, or

---

\(^1\)See Appendix B for definition
simply monotone, if $x \subseteq y$ implies $f(x) \geq f(y)$. Note that if $f$ is non-fractional, then $f$ is monotone iff the corresponding set system $A$ is a down-set. We define the weight of $f$ by $w(f) = \sum_{x \in \mathcal{P}(X)} f(x)w(x)$.

For our situation, we define $\partial f$, the boundary of the system $f$ by

$$\partial f(x) = \begin{cases} 1 & \text{if } f(x) > 0, \\ \max\{f(y) : |y \Delta x| = 1\} & \text{if } f(x) = 0. \end{cases}$$

So if $f(\mathcal{P}(X)) \subseteq \{0,1\}$ and $f$ identified with the set system $A = f^{-1}(1)$, then $\partial f$ is identified with $\partial A$, the usual boundary of $A$.

A system $f$ of the form

$$f(x) = \begin{cases} 1 & \text{if } |x| < r \\ \alpha & \text{if } |x| = r \\ 0 & \text{if } |x| > r \end{cases}$$

where $0 \leq r \leq n$ and $\alpha \in [0,1]$, is a fractional Hamming ball, or just simply a ball. Clearly, for any $0 \leq v \leq w(\mathcal{P}(X))$, there is a unique ball $b^v$ with $w(b^v) = v$. Precisely, given $v$, choose the largest $r$ such that $w(X^{(\leq r-1)}) \leq v$ and set $b^v(x) = 1$ for $x \in X^{(\leq r-1)}$. Then set $\alpha = (v - w(X^{(\leq r-1)})) / \binom{n}{r} w(x)$, where $x \in X^{(r)}$, and set $b^v(x) = \alpha$ for $x \in X^{(r)}$. Finally, set $b^v(x) = 0$ for $x \in X^{(\geq r+1)}$. We see that the boundary of a ball is again a ball.

So now we may sharpen our task of tackling the isoperimetric problem. We now aim to show that, among all monotone systems, balls have the smallest boundary. In other words, if $f$ is a monotone system of weight $v$, then $w(\partial f) \geq w(\partial b^v)$. So this will be a stronger result than our original problem, and we can deduce the original problem by letting $f$ be non-fractional. The proof is due to Bollobás and Leader [15]. We start by considering the case $n = 2$.

**Proposition 12** Let $f$ be a monotone system on $[2]$. If $w(f) = v$, then $w(\partial f) \geq w(\partial b^v)$.

**Proof** If $f(1) = f(2) = 0$ then $b^v = f$. If $f(12) > 0$ then $\partial f \equiv 1$ so that $w(\partial f) = w(\mathcal{P}([2])) \geq w(\partial b^v)$. So w.l.o.g., assume $f(12) = 0$ and $f(1) > 0$.

If $f(2) = 0$, then $w(\partial f) = 1 + r_1 + r_2 w(f)$. We cannot have $b^v(12) > 0$, otherwise $w(f) = w(b^v) = 1 + r_1 + r_2 + r_1 r_2 w(12) > f(\emptyset) + r_1 f(1) + 0 + 0 = w(f)$: a contradiction. So either $w(b^v) = b^v(\emptyset)$, for which

$$w(\partial b^v) = 1 + (r_1 + r_2) b^v(\emptyset) \leq 1 + r_1 + r_2 b^v(\emptyset) = 1 + r_1 + r_2 w(b^v),$$

or $w(b^v) = 1 + (r_1 + r_2) b^v(1)$, for which

$$w(\partial b^v) = 1 + r_1 + r_2 + r_1 r_2 b^v(1) \leq 1 + r_1 + r_2 w(b^v).$$
Thus $w(\partial f) \geq w(\partial b^v)$.

If $f(2) > 0$, then $w(\partial f) = 1 + r_1 + r_2 + r_1 r_2 \max\{f(1), f(2)\}$. A similar argument as above gives $b^v(12) = 0$. If $w(b^v) = b^v(\emptyset)$ then $w(\partial b^v) = 1 + (r_1 + r_2) b^v(\emptyset) \leq w(\partial f)$. If $w(b^v) = 1 + (r_1 + r_2) b^v(1)$, then

$$w(\partial b^v) = 1 + r_1 + r_2 + r_1 r_2 b^v(1).$$

Now $1 + (r_1 + r_2) b^v(1) = w(b^v) = w(f) = f(\emptyset) + r_1 f(1) + r_2 f(2)$, so

$$(r_1 + r_2) b^v(1) \leq r_1 f(1) + r_2 f(2) \leq (r_1 + r_2) \max\{f(1), f(2)\},$$

so $b^v(1) \leq \max\{f(1), f(2)\}$. So again, this gives $w(\partial f) \geq w(\partial b^v)$.

So having proved the result for $n = 2$, we now consider the problem for general $n$. We aim to define another compression operator: the *symmetrisation operator*, which this time acts on our systems. Symmetrisation is based on the fact that we have the result for $n = 2$, so that our analogy to the concept of *sections* from earlier would be to partition $Q_n$ into $2^{n-2}$ copies of $Q_2$. We then replace the weight of our system on each of these $Q_2$ by the corresponding ball of the same weight, and compress in a similar way as before.

To put this into technical terms, let $f$ be a system on $X = [n]$, $1 \leq i < j \leq n$, and $x \in \mathcal{P}(X \setminus \{i, j\})$. Define the *ij-section* of $f$ at $x$ to be the system $f_{ij|x}$ on $\{i, j\}$ given by

$$f_{ij|x}(y) = f(x \cup y), \quad y \in \mathcal{P}(\{i, j\}).$$

Thus $f_{ij|x}$ is just $f$ restricted to the copy of $Q_2$ with vertices $\{x, x \cup \{i\}, x \cup \{j\}, x \cup \{i, j\}\}$. For example, figure 17 shows the case $i = 1$, $j = 2$, $n = 4$:
Now define a system $S_{ij}(f)$ on $X$, the \textit{ij-symmetrisation} of $f$, by giving its \textit{ij-sections}

$$S_{ij}(f)_{ij|x} = b^{w(f)_{ij|x}}, \quad x \in \mathcal{P}(X \setminus \{i, j\}).$$

We say that $f$ is \textit{ij-symmetrised} if $S_{ij}(f) = f$. Note that if $f$ is monotone, then so is $S_{ij}(f)$. To see this, a sufficient condition for monotonicity is: if $x, y \in \mathcal{P}(X)$ with $x \subset y$, $|x \Delta y| = 1$, then $f(x) \geq f(y)$. So the assertion is obvious if $x$ and $y$ belong to the same copy of $Q_2$. If $x$ and $y$ belong to different copies of $Q_2$, then $y = x \cup \{k\}$ for some $k \in X \setminus \{i, j\}$. If $z$ is the 'ground set' (ie: of smallest cardinality) of the copy of $Q_2$ containing $x$, then $z \cup \{k\}$ is the ground set of the copy of $Q_2$ containing $y$. Then we have $r_k w\left(f_{ij|z}\right) \leq w\left(f_{ij|z\cup\{k\}}\right)$ by monotonicity between $f_{ij|z}$ and $f_{ij|z\cup\{k\}}$, which, after symmetrisation, one can check that this implies $f(x) \geq f(y)$.

With this, we have the following key result.

**Lemma 13** Let $f$ be a monotone system on $X$, and $1 \leq i < j \leq n$. Then $w(S_{ij}(f)) = w(f)$ and $w(\partial S_{ij}(f)) \leq w(\partial f)$.

**Proof** First, we have

$$w(f) = \sum_{x \in \mathcal{P}(X \setminus \{i, j\})} \sum_{y \in \mathcal{P}\{i, j\}} w(x \cup y)f(x \cup y)$$

$$= \sum_{x \in \mathcal{P}(X \setminus \{i, j\})} \sum_{y \in \mathcal{P}\{i, j\}} w(x)w(y)f_{ij|x}(y)$$

$$= \sum_{x \in \mathcal{P}(X \setminus \{i, j\})} w(x)w(f_{ij|x}), \quad (12)$$

and this implies that $w(S_{ij}(f)) = w(f)$. Now write $s$ for $S_{ij}(f)$ for convenience. By (12), it is enough to show that $\forall x \in \mathcal{P}(X \setminus \{i, j\})$, we have $w((\partial s)_{ij|x}) \leq w((\partial f)_{ij|x})$. Now fix an arbitrary $x \in \mathcal{P}(X \setminus \{i, j\})$. We have

$$(\partial f)_{ij|x} = \partial(f_{ij|x}) \lor \bigvee \left\{f_{ij|y} : y \in \mathcal{P}(X \setminus \{i, j\}), |y \Delta x| = 1\right\}, \quad (13)$$

where $\lor$ denotes pointwise maximum. Derivation of (13) is analogous to the one used in lemma 5(i). A similar expression holds with $f$ replaced by $s$. Now the system $\partial(s_{ij|y})$ and the systems $s_{ij|y}$ with $y \in \mathcal{P}(X \setminus \{i, j\})$, $|y \Delta x| = 1$, are each balls, and hence \textit{nested}, ie: for any pair $g_1$, $g_2$ of these systems, either $g_1 \subset g_2$ or $g_2 \subset g_1$. So

$$w((\partial s)_{ij|x}) = \max\left\{w(\partial(s_{ij|x})), \max\left\{w(s_{ij|y}) : y \in \mathcal{P}(X \setminus \{i, j\}), |y \Delta x| = 1\right\}\right\}. \quad (14)$$

From (14), if $w((\partial s)_{ij|x}) = w(\partial(s_{ij|x}))$, then $w((\partial s)_{ij|x}) \leq w((\partial f)_{ij|x})$ by proposition 12 and (13). If $w((\partial s)_{ij|x}) = w(s_{ij|y})$ for some $y$ as in (14), then $w((\partial s)_{ij|x}) \leq w((\partial f)_{ij|x})$
Theorem 14 (Bollobás and Leader, 1991) Let $f$ be a monotone system on $X$. If $w(f) = v$, then $w(∂f) ≥ w(∂v)$.

Proof The result is obvious if $n = 1$, so assume that $n ≥ 2$. In this case, we see that a system $s$ which is $ij$-symmetrised $∀ i, j$ must be a ball. Roughly speaking, this is because if $x ∈ X^{(r)}$, then $s(x) = s(x \setminus \{i\} ∪ \{j\})$ for some $i, j$ (by looking at the $ij$-section at $x \setminus \{i\}$). Repeating this for different $i, j$, we can get to any $r$-set, so that $s$ is constant on $X^{(r)}$. $s$ is a ball follows from the fact that each $ij$-section of $s$ is a ball.

So theorem 14 follows if we can find a system $g$ such that $w(g) = w(f)$, $w(∂g) ≤ w(∂f)$, and $g$ is $ij$-symmetrised $∀ i, j$. We consider the following compactness argument. The idea is that, starting with our given system $f$, we ‘whittle’ all the possible systems down to a single system $g$ with the desired properties.

Let $G = \{ g : \mathcal{P}(X) → [0, 1] : g$ is monotone, $w(g) = w(f)$, and $w(∂g) ≤ w(∂f) \}$. Then $f ∈ G$, so $G ≠ ∅$. The condition $w(∂g) ≤ w(∂f)$ implies that $G$ a compact set\footnote{See Appendix B for definitions} isomorphic to a subset of the product space $[0, 1]^{\mathcal{P}(X)}$. By lemma 13, we see that if $g ∈ G$, then $S_{ij}(g) ∈ G, ∀ i, j$.

For $0 ≤ r ≤ n$, let $w^{(r)}(g) = \sum_{|x|=r} w(x)g(x)$. Construct $H ⊂ G$ as follows. Choose the largest $r$ such that $v − \sum_{k=0}^{r-1} w^{(k)}(g) ≥ 0$. Then define $H ⊂ G$ so that $g ∈ H$ iff $g$ has the form:

$$g(x) = \begin{cases} 1 & \text{if } |x| < r \\ 0 & \text{if } |x| > r \end{cases}$$

and with the constraint $\sum_{|x|=r} w(x)g(x) = v − \sum_{k=0}^{r-1} w^{(k)}(g)$ (see figure 18). Hence $∀ g ∈ H$, each $ij$-section $g_{ij|x}$ is either a ball, or satisfies $g_{ij|x}(∅) = 1, g_{ij|x}(ij) = 0$. We see that this implies that $g ∈ H ⇒ S_{ij}(g) ∈ H, ∀ i, j$.

Now since $w$ is a continuous function on the compact space $[0, 1]^{\mathcal{P}(X)}$, we can choose $g ∈ H$ such that $w(g^{1/2})$ is maximal. We claim that this $g$ is $ij$-symmetrised $∀ i, j$. Assume that $1 ≤ i ≤ j ≤ n$ and $x ∈ \mathcal{P}(X \setminus \{i, j\})$ such that $s_{ij|x} ≠ g_{ij|x}$, where $s = S_{ij}(g)$. Then $g_{ij|x}(∅) = 1, g_{ij|x}(ij) = 0$. Set $a = g_{ij|x}(i), b = g_{ij|x}(j) ≠ a, c = s_{ij|x}(i) = s_{ij|x}(j)$. Then $r_1a + r_2b = (r_1 + r_2)c$. By the Cauchy-Schwarz inequality:
1.3 The weighted cube

\[ r_i a^{1/2} + r_j b^{1/2} = (r_i a)^{1/2} r_i + (r_j b)^{1/2} r_j < (r_i a + r_j b)^{1/2} (r_i + r_j)^{1/2} = (r_i + r_j)^{1/2}. \]

We have a strict inequality since \( a \neq b \). Thus \( w(g_{ij}^{1/2}) < w(s_{ij}^{1/2}) \) for some \( x \in P(X \setminus \{i, j\}) \), and (12) implies that \( w(g^{1/2}) < w(s^{1/2}) \). This contradicts the maximality of \( w(g^{1/2}) \). □

Now letting \( f \) be non-fractional, we get the following.

**Corollary 15** Let \( A \subset Q_n(r_1, \ldots, r_n) \) be a down-set. If \( w(A) \geq w(X(\leq r)) \), then

\[ w(\partial A) \geq w(X(\leq r + 1)). \]

In particular, when \( r_1 = \cdots = r_n = q \) and \( p = q/(1+q) \), then for a down-set \( A \subset Q_n(p) \) with \( P(A) \geq P(X(\leq r)) \), we have

\[ P(\partial A) \geq P(X(\leq r + 1)). \]

□

By an inductive argument on this we can immediately obtain the corresponding result about \( t \)-boundaries.

**Corollary 16** Let \( A \subset Q_n(r_1, \ldots, r_n) \) be a down-set. If \( w(A) \geq w(X(\leq r)) \), then for \( t = 0, 1, 2, \ldots \), we have \( w(A(t)) \geq w(X(\leq r + t)) \). □
By considering estimates on the tail of the binomial distribution (see [5, ch. 1]), we have

**Corollary 17** Let $0 < p < 1$, $q = 1 - p$ and $(pqn)^{1/2} \leq t \leq \min\{pqn/10, (pqn)^{2/3}/2\}$. If $A \subset Q_n(p)$ is a down-set with $\mathbb{P}(A) \geq 1/2$, then

$$
\mathbb{P}(A_{(t)}) \geq 1 - \frac{1}{(pqn)^{1/2}} \exp\left(-t^2/2pqn\right).
$$

To end this chapter, we shall just briefly remark that, although these combinatorial ideas can yield exact isoperimetric inequalities, they often fail to work for most graphs. Firstly, not every graph has a simplicial ordering defined on it so that ‘initial segments have the smallest boundaries’. Then, even if we were able to define a simplicial ordering, and were able to arrive at a ‘compressed set’, we have already seen that, depending on the graph, there is a particular method to compare the boundary of a compressed set with that of a set in simplicial order. These methods were illustrated in lemma 2, lemma 7 and theorem 14 for the graphs that we have considered. For these reasons, exact isoperimetric inequalities for many natural graphs are still unknown.
2 Probabilistic ideas

We now aim to approach isoperimetric problems using ideas from probability theory. With a rich amount of theory available to us (many of these are given in Appendix B), we can derive valuable estimates to isoperimetric inequalities on a much wider class of graphs.

2.1 Concentration of measure

A starting idea behind the ‘concentration of measure phenomenon’ is the following. We consider a sequence of graphs \( \{G_n\}_{n=1}^{\infty} \), and find a good estimate for an isoperimetric inequality for half-size sets on each \( G_n \). We then investigate the behaviour of these estimates as \( n \to \infty \). Our interest is that, as \( n \to \infty \), \( t \)-boundaries of half-size sets are almost as large as \( |G_n| \) itself.

We begin by characterising our sequence \( \{G_n\}_{n=1}^{\infty} \).

Let \( G \) be a graph with diameter \( D = \text{diam}(G) \equiv \sup \{d(x, y) : x, y \in V(G)\} \). For \( 0 < \varepsilon < 1 \), define the concentration function by

\[
\alpha(G, \varepsilon) = \sup \left\{ 1 - \left| A(\varepsilon D) \right| / |G| : A \subset G, |A|/|G| \geq 1/2 \right\}.
\]

So \( \alpha(G, \varepsilon) \) measures the largest possible fraction of \( |G| \) of the sizes of the complements of \( A(\varepsilon D) \) in \( G \), with \( |A|/|G| \geq 1/2 \). Thus if \( \alpha(G, \varepsilon) \) is small, then half-size sets of \( G \) have large neighbourhoods.

A sequence of graphs \( \{G_n\}_{n=1}^{\infty} \) is a Lévy family if \( \forall \varepsilon, \alpha(G_n, \varepsilon) \to 0 \) as \( n \to \infty \). It is a concentrated Lévy family if \( \exists C_1, C_2 > 0 \) such that \( \forall n, \varepsilon, \alpha(G_n, \varepsilon) \leq C_1 \exp(-C_2 \varepsilon n^{1/2}) \). It is a normal Lévy family if \( \exists C_1, C_2 > 0 \) such that \( \forall n, \varepsilon, \alpha(G_n, \varepsilon) \leq C_1 \exp(-C_2 \varepsilon^2 n) \).

As a prime example, consider \( \{Q_n\}_{n=1}^{\infty} \). By corollary 4, we have:

**Proposition 18** \( \{Q_n\}_{n=1}^{\infty} \) is a normal Lévy family with exponent \( C_2 = 2 \).

Indeed, we merely notice that \( \text{diam}(Q_n) = n \), so set \( t = \varepsilon n \) in corollary 4. Note however, that there may be a problem if \( \varepsilon n \) is not an integer, but this can be overcome by a suitable choice of \( C_1 \), such as \( e^6 \).

In fact, for any normal or concentrated Lévy family, the magnitude of the constant \( C_1 \) is not important, just as long as it exists.

So proposition 18 tells us that, if we take a half-size subset \( A \subset Q_n \), then \( A(\varepsilon n) \) will contain all apart from exponentially few points of \( Q_n \).
Having seen how these concentration properties work on $Q_n$, let us now look at one very important property of normal Lévy families: to relate them to a class of ‘well behaved’ functions. We can naturally turn any graph $G$ into a probability space\footnote{See Appendix B for definitions} by giving it the uniform distribution. So for $A \subset G$, we have the probability measure $\mathbb{P}(A) = |A|/|G|$.

A function $f : V(G) \to \mathbb{R}$ is called Lipschitz with constant $L$ if $\forall x,y \in V(G)$, $|f(x) - f(y)| \leq Ld(x,y)$. When $L = 1$, we simply call $f$ Lipschitz. A real number $M_f$ is a Lévy mean for $f$ if $\mathbb{P}(f \leq M_f) \geq 1/2$, and $\mathbb{P}(f \geq M_f) \geq 1/2$. For discontinuous measurable functions\footnote{See Appendix B for definitions}, $M_f$ always exists, but it need not be unique.

**Theorem 19** Let $\{G_n\}_{n=1}^\infty$ be a normal Lévy family with constants $C_1, C_2$, and let $\text{diam}(G_n) = D_n$. Let $f$ be a Lipschitz function on $G_n$, with Lévy mean $M_f$. Then

$$\mathbb{P}(|f - M_f| > \varepsilon D_n) \leq 2C_1 \exp(-C_2\varepsilon^2 n).$$

**Proof** Let $A = \{x \in V(G_n) : f(x) \leq M_f\}$ and $B = \{x \in V(G_n) : f(x) \geq M_f\}$. By definition of $M_f$, we have $\mathbb{P}(A) \geq 1/2$, so $1 - \mathbb{P}(A_{(\varepsilon D_n)}) \leq \alpha(G_n, \varepsilon) \leq C_1 \exp(-C_2\varepsilon^2 n)$. Similarly, $\mathbb{P}(B) \geq 1/2$, so $1 - \mathbb{P}(B_{(\varepsilon D_n)}) \leq C_1 \exp(-C_2\varepsilon^2 n)$. So

$$1 - \mathbb{P}(A_{(\varepsilon D_n)} \cap B_{(\varepsilon D_n)}) = 1 - (\mathbb{P}(A_{(\varepsilon D_n)}) + \mathbb{P}(B_{(\varepsilon D_n)}) - \mathbb{P}(A_{(\varepsilon D_n)} \cup B_{(\varepsilon D_n)})) \leq 2C_1 \exp(-C_2\varepsilon^2 n) - 1 + \mathbb{P}(A_{(\varepsilon D_n)} \cup B_{(\varepsilon D_n)}) \leq 2C_1 \exp(-C_2\varepsilon^2 n).$$

Now let $x \in A_{(\varepsilon D_n)}$. So $\exists y \in A$ with $d(x, y) \leq \varepsilon D_n$. Since $f$ is Lipschitz, $|f(x) - f(y)| \leq \varepsilon D_n$. Since $f(y) \leq M_f$, we have $f(x) \leq \varepsilon D_n + M_f$. Similarly, if $x \in B_{(\varepsilon D_n)}$, then $f(x) \geq -\varepsilon D_n + M_f$. So $f(x) \geq -\varepsilon D_n + M_f$. So $\mathbb{P}(|f - M_f| > \varepsilon D_n) = 1 - \mathbb{P}(|f(x) - M_f| \leq \varepsilon D_n) \leq 1 - \mathbb{P}(A_{(\varepsilon D_n)} \cap B_{(\varepsilon D_n)})).$ \hfill $\square$

Theorem 19 tells us that, if a function $f$ is rather well-behaved (Lipschitz) and $n$ is large, then $f$ is very sharply concentrated around its Lévy mean, apart from an exponentially small set.

In an opposite direction to theorem 19, if any Lipschitz function on a graph $G$ is sharply concentrated about its Lévy mean, then we can correspondingly find a good isoperimetric inequality for $G$.

**Theorem 20** Let $G$ be a graph. Suppose that whenever $f$ is a Lipschitz function on $G$, with Lévy mean $M_f$, we have $\mathbb{P}(|f - M_f| > t) \leq \beta$. If $A \subset G$ with $\mathbb{P}(A) \geq 1/2$, then $\mathbb{P}(A_{(t)}) \geq 1 - \beta$.

**Proof** Setting $f = d(x, A)$, then $f$ is Lipschitz with $M_f = 0$. So $\mathbb{P}(A_{(t)}) \leq \beta$. \hfill $\square$
2.2 Martingale techniques

We have seen that when a function on a graph is rather well-behaved (ie: Lipschitz) and the graph satisfies a good isoperimetric inequality, then the values of the function are sharply concentrated about some number. Such a situation also happens if our function (random variable) is the sum of many small random variables which are close to being independent (eg: Bernoulli random variables).

We start by looking at an early and very useful result in this direction: Azuma’s inequality. It is based on the simple fact that the exponential function is convex.

**Lemma 21** Let \( X_1, X_2, \ldots, X_m \) be random variables such that \( |X_i| \leq 1 \) \( \forall i \), and \( E \left( \prod_{i=1}^k X_{j_i} \right) = 0 \) \( \forall j_1 < \cdots < j_k \) and \( k \geq 1 \), where \( E \) denotes expectation. Then

\[
E \left( \exp \left( t \sum_{i=1}^m c_i X_i \right) \right) \leq \exp \left( \frac{t^2}{2} \sum_{i=1}^m c_i^2 \right).
\]

**Proof** Since \( \exp (tc_i X_i) \) is a convex function in \( X_i \), and \( |X_i| \leq 1 \)

\[
\exp (tc_i X_i) \leq \frac{\exp (tc_i) - \exp (-tc_i) (X_i - 1) + \exp (tc_i)}{2} = X_i \sinh(tc_i) + \cosh(tc_i).
\]

Since \( E \left( \prod_{i=1}^k X_{j_i} \right) = 0, \forall j_1 < \cdots < j_k \) (*), we have

\[
E \left( \exp \left( t \sum_{i=1}^m c_i X_i \right) \right) = E \left( \prod_{i=1}^m \exp (tc_i X_i) \right) \leq E \left( \prod_{i=1}^m (X_i \sinh(tc_i) + \cosh(tc_i)) \right) = \prod_{i=1}^m \cosh(tc_i) = \prod_{i=1}^m \sum_{r=0}^{\infty} \frac{(tc_i)^{2r}}{(2r)!}
\]

\[
\leq \prod_{i=1}^m \sum_{r=0}^{\infty} \frac{(tc_i)^{2r}}{2^r r!} = \exp \left( \frac{t^2}{2} \sum_{i=1}^m c_i^2 \right).
\]

Before we prove Azuma’s inequality, we require one of the most important inequalities in probability.

**Lemma 22 (Markov’s inequality)** Let \( X \) be a random variable, \( g : \mathbb{R} \to [0, \infty] \), where \( g \) is (Borel) measurable and increasing, and \( c \in \mathbb{R} \). Then

\[
E(g(X)) \geq g(c)P(X \geq c).
\]

**Proof** See, eg. [4, p. 80] or [28, p. 59].

\( \square \)

\( ^1 \)See Appendix B for definitions
Theorem 23 (Azuma’s inequality, 1967) Let $X_1, X_2, \ldots, X_m$ be random variables such that $|X_i| \leq 1 \forall i$, $E\left(\prod_{i=1}^k X_{j_i}\right) = 0 \forall j_1 < \cdots < j_k$ and $k \geq 1$. Then $\forall a, c_1, \ldots, c_m \in \mathbb{R}$ with $a > 0$

$$
P\left(\sum_{i=1}^m c_i X_i \geq a\right) \leq \exp\left(-\frac{a^2}{2 \sum_{i=1}^m c_i^2}\right),$$

$$
P\left(\sum_{i=1}^m c_i X_i \leq -a\right) \leq \exp\left(-\frac{a^2}{2 \sum_{i=1}^m c_i^2}\right).$$

Proof Set $g(X) = e^{tX}$ ($t > 0$), $X = \sum_{i=1}^m c_i X_i$ in Markov’s inequality. Then, via lemma 21

$$
P\left(\sum_{i=1}^m c_i X_i \geq a\right) \leq e^{-ta} E\left(e^{t \sum_{i=1}^m c_i X_i}\right) \leq e^{-ta} \exp\left(\frac{t^2}{2} \sum_{i=1}^m c_i^2\right).$$

Now set $t = a/\sum_{i=1}^m c_i^2$ to get

$$
P\left(\sum_{i=1}^m c_i X_i \geq a\right) \leq \exp\left(-\frac{a^2}{2 \sum_{i=1}^m c_i^2}\right).$$

To get the second inequality, replace $c_i$ by $-c_i$. □

We have the following equivalent formulation of Azuma’s inequality.

Theorem 24 Let $X_0, \ldots, X_m$ be random variables such that $E\left(\prod_{i=1}^k (X_{j_i} - X_{j_i-1})\right) = 0 \forall j_0 < \cdots < j_k$ and $|X_i - X_{i-1}| \leq c_i \forall i$. Then $\forall a > 0$

$$
P\left(X_m \geq X_0 + a\right) \leq \exp\left(-\frac{a^2}{2 \sum_{i=1}^m c_i^2}\right),$$

$$
P\left(X_m \leq X_0 - a\right) \leq \exp\left(-\frac{a^2}{2 \sum_{i=1}^m c_i^2}\right).$$

Proof Replace $X_i$ by $(X_i - X_{i-1})/c_i$ in theorem 23, and conversely. □

Having obtained the version of Azuma’s inequality that we want (theorem 24), we now describe what is perhaps the most natural sequence of random variables satisfying the conditions of theorem 24: martingales.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A sequence $\{\mathcal{F}_i : i \geq 0\}$ of sub-σ-fields of $\mathcal{F}$ is a filtration if $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}$. A sequence of random variables $X_0, X_1, \ldots$ on $(\Omega, \mathcal{F}, P)$ is a martingale (relative to $(\{\mathcal{F}_i\}, P)$) if

(i) each $X_i$ is $\mathcal{F}_i$-measurable;

(ii) $E(|X_i|) < \infty, \forall i$.

See Appendix B for definitions.
(iii) $\mathbb{E}(X_{i+1} | \mathcal{F}_i) = X_i$ (a.s.) $\forall i$.

Note that the above definition holds in very general situations, and the conditional expectation in (iii) has a deep theory when the $X_i$ are not discrete (see [28, ch. 9]). Here, we aim to simplify slightly for discrete random variables†. One usually takes $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

In combinatorics, since we very often consider the random variables $\{X_i\}_{i=0}^n$ to be discrete, $\{X_i\}_{i=0}^n$ is a martingale iff the following conditional expectation‡ holds

$$\mathbb{E}(X_{i+1} | X_0 = s_0, X_1 = s_1, \ldots, X_i = s_i) = s_i$$

for all $s_0, s_1, \ldots, s_i \in \mathbb{R}$, whenever $P(X_0 = s_0, X_1 = s_1, \ldots, X_i = s_i) > 0$.

For much of the rest of this chapter, we will consider $P$ to be the uniform probability measure on $\Omega$.

We are interested in $\Omega$ as a finite set, so that our filtration is a finite sequence $\{\mathcal{F}_i\}_{i=0}^m$, and we consider $\mathcal{F} = \mathcal{F}_m = \mathcal{P}(\Omega)$. We can then naturally identify each sub-$\sigma$-field $\mathcal{F}_i$ with a partition $\mathcal{P}_i = \{A_1, \ldots, A_q\}$ of $\Omega$ into the atoms of $\mathcal{F}_i$ (ie: $\forall l, A_l \in \mathcal{F}_i$ and for any $\emptyset \subsetneq B \subsetneq A_l$, $B \notin \mathcal{F}_i$). Thus $\mathcal{P}_0 = \{\Omega\}$, the trivial partition and $\mathcal{P}_m = \{\{\omega\} : \omega \in \Omega\}$, the discrete partition of $\Omega$ into singletons. We say that a partition $\mathcal{P}$ refines a partition $\mathcal{P}'$ if $\forall A \in \mathcal{P}$, $A \subseteq B$ for some $B \in \mathcal{P}'$. We write this as $\mathcal{P}' \prec \mathcal{P}$. So we have $\mathcal{P}_i \prec \mathcal{P}_{i+1}, \forall i$.

Therefore, we naturally identify a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_m$ with a sequence of partitions $\mathcal{P}_0 \prec \mathcal{P}_1 \prec \cdots \prec \mathcal{P}_m$.

Another very useful way of looking at this is to naturally consider a sequence of nested equivalence relations. We naturally identify a sub-$\sigma$-field $\mathcal{F}_i$ or a partition $\mathcal{P}_i$ with an equivalence relation $\equiv_i$ on $\Omega$ by $x \equiv_i y$ iff $x$ and $y$ belong to the same set of $\mathcal{P}_i$. Then the sequence $\{\equiv_i\}_{i=0}^m$ becomes a nested sequence of equivalence relations on $\Omega$, ie: $x \equiv_{i+1} y \Rightarrow x \equiv_i y$, for all $i$ (see figure 19)

Figure 19 Partitions of $\Omega$

†See Appendix B for definitions
Now let \( f : \Omega \rightarrow \mathbb{R} \) be any function, and let \( P_0 \prec P_1 \prec \ldots \prec P_m \) be a sequence of partitions, with \( P_0 \) trivial and \( P_m \) discrete. Define the functions \( X_0, X_1, \ldots, X_m : \Omega \rightarrow \mathbb{R} \) by setting \( X_i(\omega) \) to be the average of \( f \) on \( A \), where \( \omega \in A \in P_i \) (see figure 20). Then \( X_0 \) is constant, the average of \( f \) on \( \Omega \), and \( X_m = f \). It is easy to show that \( \{X_i\}_{i=0}^m \) becomes a martingale, and we say that \( \{X_i\}_{i=0}^m \) is the martingale determined by \( f \) and \( P_0 \prec P_1 \prec \ldots \prec P_m \). Conversely, we can recover the function \( f (= X_m) \) and the sequence of partitions \( P_0 \prec P_1 \prec \ldots \prec P_m \) from the martingale \( \{X_i\}_{i=0}^m \).

\[
X_i(\omega) = \frac{\sum_{z \in A} f(z)}{|A|}
\]

**Figure 20** Definition of \( X_i \)

In fact, if \( P \) is not necessarily uniform, the ‘averaging’ becomes taking the conditional expectation as described in (15).

From (15), or directly, it is easy to see that \( \forall j_0 < \ldots < j_k \)

\[
\mathbb{E} \left( X_{j_k} \prod_{i=1}^{k-1} (X_{j_i} - X_{j_{i-1}}) \right) = \mathbb{E} \left( X_{j_{k-1}} \prod_{i=1}^{k-1} (X_{j_i} - X_{j_{i-1}}) \right).
\]

Indeed, let \( P_{j_{k-1}} = \{A_1, \ldots, A_q\} \) and \( P_{j_k} = \left\{ B_{1}^{(i)}, \ldots, B_{m_i}^{(i)} \right\}_{i=1}^{q} \), where \( A_i = \bigcup_{h=1}^{m_i} B_{h}^{(i)} \).

Write \( s_l \) for \( \prod_{i=1}^{k-1} (X_{j_i} - X_{j_{i-1}})(\omega_l) \) for any \( \omega_l \in A_l \), so that \( \{s_l\}_{l=1}^{q} \) are all the possible values that can be taken by \( \prod_{i=1}^{k-1} (X_{j_i} - X_{j_{i-1}}) \) (see figure 21).

**Figure 21**

\[
X_{j_{k-1}} \prod_{i=1}^{k-1} (X_{j_i} - X_{j_{i-1}}) \quad X_{j_k} \prod_{i=1}^{k-1} (X_{j_i} - X_{j_{i-1}})
\]
2.2 Martingale techniques

Let $X_n$ be a martingale. Then

$$
E \left( X_{j_k} \prod_{i=1}^{k-1} (X_{j_i} - X_{j_{i-1}}) \right) = \sum_{i=1}^{q} \sum_{r=1}^{n_i} \left( \sum_{z \in B_r^{(i)}} f(z) \right) s_i \left( \frac{|B_r^{(i)}|}{|\Omega|} \right)
$$

$$= \sum_{i=1}^{q} \left( \sum_{z \in A_i} f(z) \right) s_i \left( \frac{|A_i|}{|\Omega|} \right)
$$

$$= \sum_{i=1}^{q} \left( \sum_{z \in A_i} f(z) \right) s_i \left( \frac{|A_i|}{|\Omega|} \right)
$$

$$= \prod_{i=1}^{k-1} (X_{j_i} - X_{j_{i-1}}).
$$

So

$$E \left( \prod_{i=1}^{k} (X_{j_i} - X_{j_{i-1}}) \right) = 0.
$$

Thus every martingale satisfies the conditions of theorem 24. So we have:

**Theorem 25** Let $X_0, X_1, \ldots, X_m$ be a martingale, with $|X_i - X_{i-1}| \leq c_i \forall i$. Then $\forall a > 0$

$$P \left( X_m \geq X_0 + a \right) \leq \exp \left( -a^2 / \sum_{i=1}^{m} c_i^2 \right),
$$

$$P \left( X_m \leq X_0 - a \right) \leq \exp \left( -a^2 / \sum_{i=1}^{m} c_i^2 \right).
$$

So now, theorem 25 tells us that the random variable $X_m$ is highly concentrated about its mean $X_0$. To get good bounds in theorem 25, we would like $\sum_{i=1}^{m} c_i^2$ to be small. So we want to know: how can we define partitions $\mathcal{P}_0 \prec \cdots \prec \mathcal{P}_m$ on $\Omega$ to keep $\sum_{i=1}^{m} c_i^2$ small?

An intuitive view may be helpful. Take $f = X_m$ to be Lipschitz, so that $|f(x) - f(y)| \leq d(x, y)$. To minimise $\sum c_i^2$, we would somehow like the length of our sequence of partitions $\mathcal{P}_0 \prec \cdots \prec \mathcal{P}_m$ to be rather short. However, if this length is too short, this will force some $|X_i - X_{i-1}|$ to be large (in some sense, if $\mathcal{P}_i$ is too much a refinement of $\mathcal{P}_{i-1}$, then averaging over more subsets can force the difference $|X_i - X_{i-1}|$ to be large). So forcing $c_i$ to be larger yields a summand of $c_i^2$ which makes the sum $\sum c_i^2$ worse (ie: larger).

In an ideal situation, we would like, if possible, to choose our sequence $\mathcal{P}_0 \prec \cdots \prec \mathcal{P}_m$ so that we could take $c_i = 1$ for all $i$. So the ‘rate of refinement’ of $\mathcal{P}_0 \prec \cdots \prec \mathcal{P}_m$ has to be ‘gradual’. This will apparently make more sense when we analyse the situation on $Q_n$.

Along this line, it was in [26] that Schechtman introduced the notion of the length of a metric space. Let $(\Omega, d)$ be a finite metric space, equipped with the uniform probability measure $P(A) = |A|/|\Omega|$. Then $(\Omega, d)$ has length at most $L$ if
(i) there are \(c_1, \ldots, c_m > 0\) with \((\sum c_i^2)^{1/2} = l\).

(ii) there is a sequence of partitions \(\mathcal{P}_0 \prec \cdots \prec \mathcal{P}_m\) of \(\Omega\), where \(\mathcal{P}_0\) is trivial and \(\mathcal{P}_m\) is discrete, such that whenever \(A, B \in \mathcal{P}_i\), with \(A \cup B \subseteq C\) for some \(C \in \mathcal{P}_{i-1}\), then \(\mathbb{P}(A) = \mathbb{P}(B)\), and there is a bijection \(\phi: A \to B\) with \(d(x, \phi(x)) \leq c_i \forall x \in A\) (see figure 22).

\[
\begin{array}{c}
\Omega \\
C \\
\mathcal{P}_{k-1}
\end{array} \prec \begin{array}{c}
\Omega \\
A \quad B \\
\mathcal{P}_k
\end{array}
\]

Figure 22

Thus a space with a small length gives a good inequality in theorem 25. Furthermore, if we let \(X_m = f\) be Lipschitz, then we have the following main result:

**Theorem 26** Let \((\Omega, d)\) be a finite metric space with length at most \(l\). Let \(f: \Omega \to \mathbb{R}\) be Lipschitz. Then

\[
\mathbb{P}(f \geq \mathbb{E}(f) + a) \leq \exp\left(-a^2/2l^2\right),
\]

\[
\mathbb{P}(f \leq \mathbb{E}(f) - a) \leq \exp\left(-a^2/2l^2\right).
\]

**Proof** Let \(\mathcal{P}_0 \prec \cdots \prec \mathcal{P}_m\) be partitions and \(c_1, \ldots, c_m > 0\) such that \((\Omega, d)\) has length at most \(l\). By theorem 25, it is enough to show that the martingale \(X_0, \ldots, X_m\) determined by \(f\) and \(\mathcal{P}_0 \prec \cdots \prec \mathcal{P}_m\) satisfies \(|X_i - X_{i-1}| \leq c_i\) for all \(i\).

Fix any \(x \in \Omega\) and \(1 \leq i \leq m\). Let \(A \in \mathcal{P}_i\) and \(C \in \mathcal{P}_{i-1}\) be the sets with \(x \in A\), \(x \in C\). Let \(B_1, \ldots, B_s \in \mathcal{P}_i\) be the sets such that \(C = A \cup \bigcup_{j} B_j\), \(A \neq B_j\) and \(B_j \neq B_k\) for all \(j, k\). Then

\[
X_i(x) = \sum_{z \in A} f(z)/|A|, \quad X_{i-1}(x) = \sum_{z \in C} f(z)/|C|.
\]
2.2 Martingale techniques

Since \(|A| = |B_j|\) for all \(j\)

\[
X_{i-1}(x) = \left( \sum_{z \in A} f(z) + \sum_{j=1}^{s} \sum_{z \in B_j} f(z) \right) / |C|
\]

\[
= \left( \sum_{z \in A} f(z) / |A| + \sum_{j=1}^{s} \sum_{z \in B_j} f(z) / |B_j| \right) |A| / |C|
\]

\[
= \frac{1}{s+1} \left( \sum_{z \in A} f(z) / |A| + \sum_{j=1}^{s} \sum_{z \in B_j} f(z) / |B_j| \right).
\] (16)

For each \(j\), there exists a bijection \(\phi : A \rightarrow B_j\) such that \(d(z, \phi(z)) \leq c_i, \forall z \in A\). Together with the Lipschitz condition of \(f\), we have

\[
\left| \sum_{z \in A} f(z) / |A| - \sum_{z \in B_j} f(z) / |B_j| \right| \leq \sum_{z \in A} |f(z) - f(\phi(z))| / |A|
\]

\[
\leq \sum_{z \in A} d(z, \phi(z)) / |A|
\]

\[
\leq c_i.
\] (17)

(16) and (17) imply

\[
|X_i(x) - X_{i-1}(x)| = \left\| \sum_{z \in A} f(z) / |A| - \frac{1}{s+1} \left( \sum_{z \in A} f(z) / |A| + \sum_{j=1}^{s} \sum_{z \in B_j} f(z) / |B_j| \right) \right\|
\]

\[
= \frac{1}{s+1} \left| \frac{1}{s+1} \sum_{z \in A} f(z) / |A| - \sum_{j=1}^{s} \sum_{z \in B_j} f(z) / |B_j| \right|
\]

\[
\leq \frac{1}{s+1} \sum_{j=1}^{s} \left| \sum_{z \in A} f(z) / |A| - \sum_{z \in B_j} f(z) / |B_j| \right|
\]

\[
\leq c_i.
\]  □

So now, having obtained theorem 26, we can now combine it with theorem 20 to find reasonably good isoperimetric inequalities on various families of graphs, and prove that they are normal Lévy families. All we have to do is to calculate the length of the graph as a metric space (and that this has to be small).

**Theorem 27** \(\{Q_n\}_{n=1}^{\infty}\) is a normal Lévy family with exponent \(1/2\).

**Proof** By theorem 20, via theorem 20, we just have to show that \(Q_n\) has length at most \(n^{1/2}\).

Considering the elements of \(Q_n\) as 0-1 sequences of length \(n\), let the partition \(\mathcal{P}_i\) of
$Q_n$ be induced by the equivalence relation $\equiv$, where $x = (x_1, \ldots, x_n) \equiv y = (y_1, \ldots, y_n)$ iff $x_j = y_j$ $\forall j \leq i$. Thus $\mathcal{P}_0 \prec \cdots \prec \mathcal{P}_n$, with $\mathcal{P}_0$ trivial and $\mathcal{P}_n$ discrete.

For $A, B \in \mathcal{P}_i$, $A \neq B$ and $A \cup B \subseteq C$ for some $C \in \mathcal{P}_{i-1}$, w.l.o.g. let

\[ A = \{x \in \{0, 1\}^n : x_j = a_j \text{ for } j < i \text{ and } x_i = 0\}, \]
\[ B = \{x \in \{0, 1\}^n : x_j = a_j \text{ for } j < i \text{ and } x_i = 1\}, \]

for some $a_1, \ldots, a_{i-1} \in \{0, 1\}$. Then $|A| = |B| = 2^{n-k}$, so $P(A) = P(B)$. Let $\phi : A \rightarrow B$ be the bijection that ‘changes the $i$th term of $x \in A$', ie: $\phi(x) = y$, where $x_i = 0$, $y_i = 1$ and $x_j = y_j \ \forall j \neq i$. Then $\phi$ satisfies $d(x, \phi(x)) \leq 1 \ \forall x \in A$. So we can take $c_i = 1 \ \forall i$, and hence $Q_n$ has length at most $n^{1/2}$. □

Remarks (i) We note that $n^{1/2}$ is the best possible value. If we were able to have fewer than $n + 1$ partitions, then at some refinement $\mathcal{P}_{i-1} \prec \mathcal{P}_i$, we have mutually distinct $A, B, C \in \mathcal{P}_i$ with $A \cup B \cup C \subseteq D$ for some $D \in \mathcal{P}_{i-1}$ (and so $|A| = |B| = |C|$). From this, it can be shown that some bijection between some pair in $\{A, B, C\}$ must have $c_i \geq 2$, and thus this creates a summand in $\sum c_i^2$ that actually increases the sum. In some sense, such a sequence of refinements is ‘not gradual enough’.

(ii) Theorem 27 gives a slightly weaker isoperimetric inequality on $Q_n$ than Harper’s inequality in corollary 4. Here, we have lost a factor of 4 in the exponent. This is simply because corollary 4 came from an estimate on the exact isoperimetric inequality on $Q_n$, so it is much more accurate. In general, this so-called Schechtman’s method of combining theorem 26 and theorem 20 works best on metric spaces of small length. But often it can only yield a rather weak isoperimetric inequality. For example, for the weighted cube $Q_n(p)$, a more general form of Schechtman’s method (with $P$ non-uniform: see [7]) gives

$$P(A(t)) \geq 1 - e^{-2t^2/n}$$

(18)

for $P(A) \geq 1/2$. Comparing this to corollary 17 (induced from the exact isoperimetric inequality), we have lost a factor of $1/pq$ in the exponent. Thus if $p$ is close to 0 or 1, then (18) is much worse than corollary 17.

(iii) In general, Schechtman’s method can be extended to any metric probability space $(\Omega, \mathcal{F}, P, d)$, even non-discrete ones. All the notions (eg: expectation, martingales, etc) will have to be generalised. See [7], [24] and [26] for variations.

Let us end this chapter by mentioning another well-known example of a graph with a small length: the symmetric group $S_n$. This is the graph formed by the group of permutations of $\{1, \ldots, n\} = [n]$, equipped with the uniform probability measure, ie:
\[ P(A) = |A|/n! \text{ for } A \subseteq S_n. \] For \( \rho, \sigma \in S_n \), the metric on \( S_n \) is
\[ d(\rho, \sigma) = |\{ k : \rho(k) \neq \sigma(k) \}|. \]

With this set-up, we have the following.

**Theorem 28** \( \{S_n\}_{n=1}^\infty \) is a normal Lévy family with exponent \( 1/8 \).

**Proof** By Schechtman's method, we have to show that \( S_n \) has length at most \( 2n^{1/2} \). Let the partition \( \mathcal{P}_i \) of \( S_n \) be induced by the equivalence relation \( \equiv_i \), where for \( \rho, \sigma \in S_n \), \( \rho \equiv_i \sigma \) iff \( \rho(j) = \sigma(j) \) \( \forall j \leq i \). Then \( \mathcal{P}_0 \prec \mathcal{P}_1 \prec \cdots \prec \mathcal{P}_{n-1} \), with \( \mathcal{P}_0 = \{S_n\} \) and \( \mathcal{P}_{n-1} \) discrete. So a set \( A \in \mathcal{P}_i \) has the form
\[ A = \{ \rho \in S_n : \rho(j) = a_j, 1 \leq j \leq i \} \]
for some distinct, fixed \( a_1, \ldots, a_i \in [n] \). So if \( A, B \in \mathcal{P}_i \) with \( A \neq B \) and \( A \cup B \subseteq C \) for some \( C \in \mathcal{P}_{i-1} \), then
\[ A = \{ \rho \in S_n : \rho(j) = a_j, 1 \leq j \leq i-1, \rho(i) = a \}, \]
\[ B = \{ \rho \in S_n : \rho(j) = a_j, 1 \leq j \leq i-1, \rho(i) = b \}, \]
for distinct, fixed \( a_1, \ldots, a_{i-1}, a, b \in [n] \). Thus \( |A| = |B| = (n-i)! \). Now let \( \tau \) be the transposition \((ab)\), and \( \phi : A \to B \) be given by \( \phi(\rho) = \tau \rho \). Then \( \phi \) is a bijection with \( d(\rho, \phi(\rho)) = 2 \forall \rho \in A \). So in fact, this shows that \( S_n \) has length at most \( 2(n-1)^{1/2} \), which is sufficient for theorem 28. \( \square \)

**Remark** Notice the slightly curious feature that for \( d \) defined above, \( d(\rho, \sigma) \neq 1 \) for any \( \rho, \sigma \in S_n \). Some authors have used a slight variation for \( d \), such as the ‘normalised Hamming metric’ \( d(\rho, \sigma) = |\{ k : \rho(k) \neq \sigma(k) \}|/n \) as described in [24]. Such a variation can yield a different exponent for the normal Lévy family \( \{S_n\}_{i=1}^\infty \).
3 Product graphs and eigenvalue methods

In this chapter, we are interested in isoperimetric inequalities on products of (connected) graphs. We aim to approach the problem in two different ways. The first approach is an application of Wang and Wang’s result for the finite grid (theorem 9, chapter 1.2). In the second approach, we introduce the idea of considering the eigenvalues of a certain matrix related to our graph, and show how this can solve the problem.

3.1 Products of graphs

Let $G$ and $H$ be graphs. The product graph $G \times H$ is the graph whose vertex set is $V(G) \times V(H)$, with $(g, h)$ and $(g', h')$ adjacent if either $g = g'$ and $hh' \in E(H)$, or $h = h'$ and $gg' \in E(G)$. More generally, $\prod_{i=1}^{n} G_i$ is the product of the graphs $G_1, \ldots, G_n$, where the vertices $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ are adjacent if $\exists i, x_i y_i \in E(G_i)$ and $x_j = y_j, \forall j \neq i$. We write $G^n$ for the $n$-fold product $G \times \cdots \times G$. Note that $\prod_{i=1}^{n} G_i$ contains $|\prod_{i=1}^{n} G_i| / |G_i|$ copies of $G_i, \forall i$ (or infinitely many if for some $j \neq i$, $|G_j| = \infty$). For example, figure 23 shows the case when $G_1 = C_3, G_2 = C_4$ (where $C_k$ is the cycle of length $k$). 

![Figure 23 The product graph $C_3 \times C_4$](image)

As mentioned earlier, $\mathbb{Z}_n^+$ is the product of $n$ copies of the infinite path $\mathbb{Z}_+$. Also, $[k]^n$ is the product of $n$ copies of $P_{k-1}$, the path of length $k - 1$.

For a set $A \subset \prod_{i=1}^{n} G_i$, $1 \leq i \leq n$, and $x \in \prod_{j \neq i} G_j$, the $i$-section of $A$ at $x$ is $A_i(x) = \{g \in G_i : (x_1, \ldots, x_{i-1}, g, x_{i+1}, \ldots, x_n) \in A\}$.

Now let $G_1, \ldots, G_n$ be graphs, with $|G_i| = k_i$. For $A \subset \prod_{j=1}^{n} G_j$ and $1 \leq i \leq n$, define the compression operator $\chi_i$ mapping subsets of $\prod_{j=1}^{n} G_j$ into subsets of $P_{k_i-1} \times \prod_{j \neq i} G_j$.

$^1$See Appendix A for definitions
(\(P_{k_i-1}\) is considered to be the \(i\)th factor of \(P_{k_i-1} \times \prod_{j \neq i} G_j\)) by giving its \(i\)-sections

\[
\chi_i(A_i)(x) = \{0,1,\ldots,|A_i(x)| - 1\}, \quad x \in \prod_{j \neq i} G_j, A_i(x) \neq \emptyset
\]

We have \(|\chi_i(A)| = |A|\). Moreover, we have the following.

**Lemma 29** Let \(G_1, \ldots, G_n\) be connected graphs\(^1\), \(1 \leq i \leq n\), and \(A \subset \prod_{j=1}^n G_j\). Then \(|\partial \chi_i(A)| \leq |\partial A|\).

**Remark** This is a stronger result than lemma 5(i).

**Proof** Write \(C\) for \(\chi_i(A)\) for convenience. To show that \(|\partial C| \leq |\partial A|\), it is enough to show that \(|(\partial C)_i(x)| \leq |(\partial A)_i(x)|\), \(\forall x \in \prod_{j \neq i} G_j\).

Fix \(x \in \prod_{j \neq i} G_j\). By an exact same argument as in lemma 5(i), we have

\[
(\partial A)_i(x) = \partial(A_i(x)) \cup \bigcup \left\{ A_i(y) : d(y,x) = 1, y \in \prod_{j \neq i} G_j \right\},
\]

and a similar expression holds with \(C\) in place of \(A\). The set \(\partial(C_i(x))\) and the sets \(C_i(y)\), \(d(y,x) = 1\) are initial segments of \(\mathbb{Z}_+\), so they are nested. Since \(|C_i(y)| = |A_i(y)|\), \(\forall y\), with an exact same argument as in lemma 5(i), it is enough to show that \(|\partial(C_i(x))| \leq |\partial(A_i(x))|\).

In this case, \(C_i(x) \neq \emptyset\). Thus clearly, \(|\partial(C_i(x))| = \min\{|A_i(x)| + 1, |G_i|\}\). Since \(G_i\) is a connected graph and \(A_i(x) \neq \emptyset\), we have \(|\partial(A_i(x))| \geq \min\{|A_i(x)| + 1, |G_i|\}\). Thus \(|\partial(C_i(x))| \leq |\partial(A_i(x))|\). \(\square\)

We can now prove the following isoperimetric inequality for product graphs.

**Theorem 30** Let \(G_1, \ldots, G_n\) be connected graphs of order \(k\), and \(A \subset \prod_{j=1}^n G_j\). Then \(|\partial A| \geq \partial_k^{(n)}(|A|)\).

**Proof** Let \(B = \chi_n(\chi_{n-1}(\cdots \chi_1(A) \cdots))\). By lemma 29, \(|B| = |A|\) and \(|\partial B| \leq |\partial A|\).

Since \(B \subset [k]^n = P_{k-1}^n\), by theorem 9 we have \(|\partial B| \geq \partial_k^{(n)}(|B|)\). It follows that \(|\partial A| \geq \partial_k^{(n)}(|A|)\). \(\square\)

Now define \(B_k^{(n)}(r) = \{x \in [k]^n : \sum x_i \leq r\}\) and \(b_k^{(n)}(r) = |B_k^{(n)}(r)|\). We have the following result for \(t\)-boundaries.

**Corollary 31** Let \(G_1, \ldots, G_n\) be connected graphs of order \(k\), \(A \subset \prod_{j=1}^n G_j\), and \(|A| \geq b_k^{(n)}(r)\). Then \(\forall t = 0,1,\ldots, |A(t)| \geq b_k^{(n)}(r + t)\).

**Proof** \(t = 0\) is trivial. For \(t = 1\),

\[
|A(1)| \geq \partial_k^{(n)}(|A|) \geq \partial_k^{(n)}(b_k^{(n)}(r)) = b_k^{(n)}(r + 1).
\]

\(^1\)See Appendix A for definition
If $|A(s)| \geq b_k^{(n)}(r + s)$, then

$$|A(s+1)| = |\partial (A(s))| \geq \partial_k^{(n)} (|A(s)|) \geq \partial_k^{(n)} (b_k^{(n)}(r + s)) = b_k^{(n)}(r + s + 1).$$

Result follows by induction.

Since the isoperimetric inequality on the finite grid (theorem 9) is the best possible, clearly theorem 30 and corollary 31 are the best possible isoperimetric inequalities for a general product of graphs of order $k$. But since these graphs are arbitrary, one can expect that the inequalities of theorem 30 and corollary 31 can be poor for certain graphs.

To make the situation a lot more interesting, Bollobás and Leader [14] considered the problem when the diameters of the $G_j$ are known. To find a good estimate for large $n$ in corollary 31, we need to find a good estimate for $b_k^{(n)}(r)$. Here we aim to use a probabilistic argument.

**Remark** If we throw a fair die with $k$ faces marked from 0 to $k-1$, $n$ times and independently, then $[k]^n$ is precisely the sample space with all the possible sequences of outcomes. Moreover, each simplex $\sum x_i = r$ consists of all the $n$-sequences with a sum of $r$. Thus $b_k^{(n)}(r)$ consists of all the $n$-sequences with sum $\leq r$.

**Lemma 32** Let $\varepsilon > 0$. Then

$$b_k^{(n)}([n(k-1)(1/2 - \varepsilon)]) \leq \exp \left( -6\varepsilon^2 n \frac{k-1}{k+1} \right) k^n.$$

Since $S = \{ x : \sum x_i = n(k-1)/2 \}$ is the simplex passing the centre of $[k]^n$ (and thus ‘bisecting $[k]^n$ into two identical parts’), $\{ x : \sum x_i = [n(k-1)(1/2 - \varepsilon)] \}$ is a simplex which lies strictly below $S$. So $b_k^{(n)}([n(k-1)(1/2 - \varepsilon)])$ has at most $k^n/2$ points (see figure 24). Thus lemma 32 claims that in fact, the proportion of $b_k^{(n)}([n(k-1)(1/2 - \varepsilon)])$ to $k^n$ is exponentially small.

**Figure 24** Definition of $b_k^{(n)}([n(k-1)(1/2 - \varepsilon)])$
3.1 Products of graphs

Proof of lemma 32 For convenience, we shift \([k]^n\) so that the centre is at the origin. So we can define the random variables \(X_1, \ldots, X_n\) which are identical, independent, and each with a uniform distribution on \(\{(-k+1)/2, (-k+3)/2, \ldots, (k-1)/2\}\) (thus in our die throwing experiment, \(X_i\) is the \(i\)th throw, but the possible outcomes are now \(\{(-k+1)/2, (-k+3)/2, \ldots, (k-1)/2\}\). Let \(X = \sum X_i\). We see that

\[
b_k^{(n)}([n(k-1)(1/2-\varepsilon)]) = k^n P(X \geq \varepsilon n(k-1)),
\]

where \(P\) is the uniform probability measure on \([k]^n\). Roughly, figure 25 explains (20).

![Figure 25](image)

We now find an estimate for the r.h.s. of (20). By Markov’s inequality, for any \(u > 0\)

\[
\exp(u\varepsilon n(k-1)) P(X \geq \varepsilon n(k-1)) \leq \mathbb{E} \left( e^{uX} \right) = \left( \mathbb{E} \left( e^{uX_1} \right) \right)^n = \left( \sum_{j=0}^{\infty} \frac{(uX_1)^j}{j!} \right)^n = \left( \sum_{j=0}^{\infty} \frac{u^j}{j!} \mathbb{E}(X_1^j) \right)^n
\]

(21)

For \(j = 0, 1, \ldots,\) we have

\[
\mathbb{E}(X_1^j) = \sum_{m=0}^{k-1} \frac{1}{k} \left( \frac{-k+1+2m}{2} \right)^j = \begin{cases} 
\frac{1}{2^k} \sum_{m=0}^{k-1} (-k+1+2m)^j & \text{for } j \text{ even,} \\
0 & \text{for } j \text{ odd.}
\end{cases}
\]

Now compute \(\mathbb{E}(X_1^j)\). Using the formulae \(1^2 + 3^2 + \cdots + (2s-1)^2 = s(2s+1)(2s-1)/3\) (for \(k\) even) and \(2^2 + 4^2 + \cdots + (2s)^2 = 2s(s+1)(2s+1)/3\) (for \(k\) odd), we obtain \(\mathbb{E}(X_1^j) = (k^2 - 1)/12\) in both cases. From this, one can check by induction on \(j\) that

\[
\frac{\mathbb{E}(X_1^{2j})}{(2j)!} \leq \frac{1}{j!} \left( \frac{k^2 - 1}{24} \right)^j, \quad j = 1, 2, \ldots.
\]

(22)
From (21) and (22), we have
\[
\exp(u\varepsilon n(k-1))\mathbb{P}(X \geq \varepsilon n(k-1)) \leq \left( \sum_{j=0}^{\infty} \frac{u_j^j}{j!} \mathbb{E}(X_j^k) \right)^n
\]
\[
= \left( \sum_{j=0}^{\infty} \frac{u_j^2 j}{(2j)!} \mathbb{E}(X_j^k) \right)^n
\]
\[
\leq \left( \sum_{j=0}^{\infty} \frac{u_j^2 j}{j!} \left( \frac{k^2 - 1}{24} \right)^j \right)^n
\]
\[
= \exp \left( nu^2(k^2 - 1)/24 \right)
\]
\[
\Rightarrow \mathbb{P}(X \geq \varepsilon n(k-1)) \leq \exp \left( nu^2(k^2 - 1)/24 - u\varepsilon n(k-1) \right) \tag{23}
\]

Now for \(k \neq 1\), \(nu^2(k^2 - 1)/24 - u\varepsilon n(k-1)\) is a quadratic in \(u\), with a minimum at \(u = \frac{12}{k+1}\varepsilon\). So setting \(u = \frac{12}{k+1}\varepsilon\) in (23) and via (20):
\[
b_k^n(|n(k-1)(1/2 - \varepsilon)|) \leq k^n \exp \left( n \left( \frac{12\varepsilon}{k+1} \right)^2 \frac{k^2 - 1}{24} - \frac{12\varepsilon}{k+1} n(k-1) \right)
\]
\[
= k^n \exp \left( -6t^2 \frac{n}{(k^2 - 1)} \right).
\]
\[\square\]

Now combining corollary 31 and lemma 32, we have

**Theorem 33 (Bollobás and Leader, 1991)** Let \(G_1, \ldots, G_n\) be connected graphs of order \(k\), and \(A \subset \prod_{j=1}^n G_j\) with \(|A|/k^n \geq 1/2\). Then for \(t = 0, 1, \ldots\)
\[
|A(t)|/k^n \geq 1 - \exp \left( - \frac{6t^2}{n(k^2 - 1)} \right).
\]

**Proof** Since \(|A|/k^n \geq 1/2\)
\[
|A| \geq b_k^n(|n(k-1)/2 - 1|),
\]
since \(|n(k-1)/2 - 1| < n(k-1)/2\). Now apply corollary 31 and lemma 32
\[
|A(t)| \geq b_k^n(|n(k-1)/2 - 1| + t)
\]
\[
\Rightarrow |[k]^n \setminus A(t)| \leq b_k^n(|n(k-1)/2 - 1| - t)
\]
\[
= b_k^n \left( \left| n(k-1) \left( \frac{1}{2} - \frac{t}{n(k-1)} \right) \right| \right)
\]
\[
\leq \exp \left( -6 \frac{k - 1}{k + 1} \left( \frac{t}{n(k-1)} \right)^2 n \right) k^n
\]
\[
= \exp \left( - \frac{6t^2}{n(k^2 - 1)} \right) k^n.
\]
(24) can be verified easily, and so the result follows. \[\square\]

Now having obtained theorem 33, we immediately have the following.
3.2 Eigenvalue techniques

Theorem 34 Let $G$ be a connected graph of order $k$, with $\text{diam}(G) = D$. Then $\{G^n\}_{n=1}^{\infty}$ is a normal Lévy family with exponent $6D^2/(k^2 - 1)$.

Now, by an application of Schechtman’s method described in chapter 2 on $G^n$, it is possible to consider the length of $G^n$, and this gives the following

Theorem 35 $\{G^n\}_{n=1}^{\infty}$ is a normal Lévy family with exponent $1/64$.

If we now compare the bounds of theorem 34 and theorem 35, we see that if $D = \text{diam}(G)$ is large, and $k = |G|$, then the bound $6D^2/(k^2 - 1)$ is much better than the bound $1/64$. This is because theorem 34 is based on the exact isoperimetric inequality on the finite grid $[k]^n$, and the larger $D$ is, the closer $G$ is to a path. Equally, if $D$ is small, then the bound of $1/64$ is better.

3.2 Eigenvalue techniques

Spectral graph theory is the study of the eigenvalues of a certain matrix associated with a graph. It has a long history, and provides a lot of information about the structure and properties of a graph. In this section, we will assume a fair amount of terminology in linear algebra, most of which are discussed in the course M221 (Algebra 3).

Let $G$ be a graph with vertices $v_1, \ldots, v_k$. Define the adjacency matrix $A_G$ of $G$ (w.r.t. $v_1, \ldots, v_k$) to be the $k \times k$ matrix with elements given by

$$a_{ij} = \begin{cases} 1 & \text{if } v_iv_j \in E(G), \\ 0 & \text{if } v_iv_j \notin E(G). \end{cases}$$

For example

$$G = \begin{array}{ccc} v_1 & v_2 \\ v_3 & v_4 & v_5 \end{array}$$

has $A_G = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$

Figure 26 The adjacency matrix

From now on, we will only be interested in the case when $G$ is connected. Obviously, $A_G$ depends on the labelling of the vertices, but our consideration will only come down to an inner product (to be defined). For this, it makes no difference whichever adjacency matrix we use.
Define the matrix $Q = Q_G = \text{diag}(\deg(v_1), \ldots, \deg(v_k)) - A_G$. $Q$ is sometimes called the (minus) Laplace operator for $G$. The spectrum of $Q$, $\sigma(Q)$, is the set of the eigenvalues of $Q$.

Define $L^2(G)$ to be the space of all functions $f: V(G) \to \mathbb{R}$, equipped with the inner product $\langle f, g \rangle = \sum_{v \in V(G)} f(v)g(v)$ and the norm $\|f\| = \|f\|_2 = (\langle f, f \rangle)^{1/2}$ induced by it (hence $f \in L^2(G)$ implies that $\|f\| < \infty$).

We begin with some important basic results.

Lemma 36 Consider the quadratic form $\langle Qf, f \rangle$ defined for $f \in L^2(G)$.

(i) $\langle Qf, f \rangle$ is a positive semi-definite form on $L^2(G)$, with $\langle Qf, f \rangle = 0$ iff $f$ is constant,

(ii) $Q$ has $0$ as the smallest eigenvalue which is simple,

(iii) if $f$ is orthogonal to the constants, and $\lambda_1$ is the second smallest eigenvalue of $Q$, then

$$\langle Qf, f \rangle \geq \lambda_1 \|f\|^2.$$  (25)

Proof (i) Write $D = \text{diag}(\deg(v_1), \ldots, \deg(v_k))$, $A = A_G$ for convenience.

$$\langle Qf, f \rangle = \sum_{v \in G} (Df)(v)f(v) - \sum_{v \in G} (Af)(v)f(v)$$

$$= \sum_{v \in G} \deg(v)(f(v))^2 - \sum_{i=1}^k \left( \sum_{v_i v_j \in E(G)} f(v_i) \right) f(v_i)$$

$$= \sum_{v \in G} \deg(v)(f(v))^2 - 2 \left( \sum_{v_i v_j \in E(G)} f(v_i)f(v_j) \right)$$

$$= \sum_{v_i v_j \in E(G)} (f(v_i) - f(v_j))^2 \geq 0.$$  (26)

Since $G$ is connected, $f(v_i)$ appears at least once in (26), $\forall i$. It follows that $\langle Qf, f \rangle = 0$ iff $f$ is constant.

(ii) Let $\lambda \in \sigma(Q)$. Then $Qf = \lambda f$, and (i) gives $\lambda = \langle Qf, f \rangle / \|f\|^2 \geq 0$, for some $f \neq 0$ (The expression $\langle Qf, f \rangle / \|f\|^2$ is known as the Rayleigh quotient). Since $f = \text{constant}$ iff $Qf = 0$, $0$ is an eigenvalue whose eigenfunctions are the constants. $0$ is simple since the space of constant functions has dimension $1$.

(iii) Since $Q$ is symmetric, $L^2(G)$ has an orthonormal basis of eigenfunctions corresponding to the eigenvalues of $Q$. Let $\sigma(Q) = \{\lambda_i\}_{i=0}^{k-1}$ with $\lambda_0 = 0$ and $\lambda_i > 0 \ \forall i \geq 1$. 


3.2 Eigenvalue techniques

Let \(\{e^{(i)}\}_{i=0}^{k-1}\) be an orthonormal basis of eigenfunctions for \(L^2(G)\), (with \(e^{(0)}\) constant). Then \(f = \sum_{i \geq 0} \gamma_i e^{(i)}\) for some \(\gamma_0, \ldots, \gamma_{k-1} \in \mathbb{R}\). Thus

\[
Qf = Q\left(\sum_{i \geq 0} \gamma_i e^{(i)}\right) = \sum_{i \geq 1} \gamma_i \lambda_i e^{(i)}
\]

\[
\Rightarrow \langle Qf, f \rangle = \sum_{i \geq 1} \lambda_i \gamma_i^2 \geq \lambda_1 \sum_{i \geq 1} \gamma_i^2 = \lambda_1 \|f\|^2,
\]

since \(f\) is orthogonal to \(e^{(0)}\) and the \(e^{(i)}\) are orthogonal, \(\|f\|^2 = \sum_{i \geq 1} \gamma_i^2\). \hfill \square

From now on, we will write \(\lambda_1 = \lambda_1(G)\) for the second smallest eigenvalue of \(Q_G\).

Now we are ready to consider the isoperimetric problem on product graphs again. The following approach is due to Alon and Milman [1]. Roughly speaking, the strategy is to consider two disjoint vertex subsets \(A, B\) of a graph \(G\). We then examine the relationship between the distance between \(A\) and \(B\), the sizes of \(A\), \(B\) and \(G\), and \(\lambda_1 = \lambda_1(G)\).

We start with following result, variations of which have been investigated by many authors.

**Proposition 37** Let \(A, B \subset G\) for a graph \(G\) of order \(k\), with \(A \cap B = \emptyset\). Let \(d = d(A, B)\), \(a = \mathbb{P}(A) = |A|/k\), \(b = \mathbb{P}(B) = |B|/k\) and \(E = E(G)^\dagger\). Let \(E_A, E_B\) be the edge sets of the induced subgraphs by \(A, B\), respectively. Then

\[
\lambda_1 k \leq \frac{1}{d^2} \left(\frac{1}{a} + \frac{1}{b}\right) (|E| - |E_A| - |E_B|).
\]

**Proof** Define \(g \in L^2(G)\) by

\[
g(v) = \frac{1}{a} - \frac{1}{d} \left(\frac{1}{a} + \frac{1}{b}\right) \min\{d(v, A), d\}.
\]

If \(v \in A\), then \(g(v) = 1/a\), and if \(v \in B\), then \(g(v) = -1/b\). If \(d(u, v) = 1\), then

\[
|g(u) - g(v)| = \frac{1}{d} \left(\frac{1}{a} + \frac{1}{b}\right) |\min\{d(v, A), d\} - \min\{d(u, A), d\}| \leq \frac{1}{d} \left(\frac{1}{a} + \frac{1}{b}\right),
\]

since clearly, \(|\min\{d(v, A), d\} - \min\{d(u, A), d\}| \leq 1\) if \(d(u, v) = 1\). Now set \(\alpha = \sum g(v)/k\) and \(f = g - \alpha\), so that \(\sum f(v) = 0\), ie: \(f\) is orthogonal to the constants, and \(f\) satisfies (25).

Via (26), we have

\[
\lambda_1 \|f\|^2 \leq \langle Qf, f \rangle = \sum_{v_i, v_j \in E} (f(v_i) - f(v_j))^2 = \sum_{v_i, v_j \in E} (g(v_i) - g(v_j))^2 = \sum_{v_i, v_j \in E \setminus (E_A \cup E_B)} (g(v_i) - g(v_j))^2 \leq \frac{1}{d^2} \left(\frac{1}{a} + \frac{1}{b}\right)^2.
\]  

\(\dagger\)See Appendix A for definition
Also
\[
\lambda_1 ||f||^2 \geq \lambda_1 \left( \sum_{v \in A} ((g - \alpha)(v))^2 + \sum_{v \in B} ((g - \alpha)(v))^2 \right)
\]
\[
= \lambda_1 \left( \frac{1}{a} - \alpha \right)^2 ka + \left( \frac{1}{b} + \alpha \right)^2 kb \right)
\]
\[
= \lambda_1 k \left( \frac{1}{a} + \frac{1}{b} + \alpha^2 a + \alpha^2 b \right)
\]
\[
\geq \lambda_1 k \left( \frac{1}{a} + \frac{1}{b} \right). \tag{28}
\]

Now combine (27) and (28) to get the result. \(\square\)

**Corollary 38** Let \(G\) be a graph of order \(k\), and let \(\delta = \delta(G)\) be the minimum degree\(^1\) of \(G\). Then \(\lambda_1 \leq k\delta/(k - 1)\).

**Proof** Choose \(v \in V(G)\) such that \(\deg(v) = \delta\). Now apply proposition 37 with \(A = \{v\}, B = V(G) \setminus \{v\}, d = 1\) and \(|E| - |E_A| - |E_B| = \delta\). \(\square\)

**Corollary 39** With the same set-up as proposition 37, let \(\Delta = \Delta(G)\) be the maximum degree\(^1\) of \(G\). If \(d > 1\), then \(b \leq (1 - a)/(1 + (\lambda_1 d^2)/\Delta)\).

**Proof** Since \(d > 1\), every edge of \(E \setminus (E_A \cup E_B)\) has an endpoint at one of the \(k - ka - kb\) vertices of \(V(G) \setminus (A \cup B)\). It follows that \(|E| - |E_A| - |E_B| \leq k(1 - a - b)\Delta\). Now apply proposition 37 to get
\[
\lambda_1 k \leq \frac{1}{d^2} \left( \frac{1}{a} + \frac{1}{b} \right) k(1 - a - b)\Delta \leq \frac{1}{d^2 \alpha b} k(1 - a - b)\Delta. \tag{29}
\]
The result follows from (29). \(\square\)

**Theorem 40 (Alon and Milman, 1985)** With \(G, A, B, a, b, \lambda_1\), and \(\Delta\) as in corollary 39, suppose that \(d(A, B) > d \geq 1\) (\(d\) is not necessarily an integer). Then
\[
b \leq (1 - a) \exp \left( - \log(1 + 2a)[(\lambda_1/2\Delta)^{1/2}d] \right). \tag{30}
\]
In particular, if \(a \geq 1/2\), then for \(t \geq 1\),
\[
\mathbb{P} (A_{(t)}) \geq 1 - \frac{1}{2} \exp(-t(\lambda_1/2\Delta)^{1/2} \log 2). \tag{31}
\]

**Proof** Let \(\mu = (2\Delta/\lambda_1)^{1/2}\). By corollary 38, if \(k \geq 2\), then
\[
\mu \geq ((k - 1)/k)^{1/2}(2\Delta/\delta)^{1/2} \geq (2 - 2/k)^{1/2} \geq 1.
\]

\(^1\)See Appendix A for definitions
Let $m = [d/\mu]$, so $d (A_{(j\mu)}, V(G) \setminus A_{((j+1)\mu)}) = s > \mu \geq 1$, for $0 \leq j < m$ (see figure 27). For $C \subset G$, let $c_t = \mathbb{P} (C_t) = |C_t|/k$. By corollary 39

$$1 - a_{(j+1)\mu} \leq \frac{1 - a_{j\mu}}{1 + (\lambda_1/\Delta) a_\mu s^2} \leq \frac{1 - a_{j\mu}}{1 + (\lambda_1/\Delta) a_\mu s^2} = \frac{1 - a_{j\mu}}{1 + 2a},$$

(32)

for all $0 \leq j < m$.

\[\text{Figure 27}\]

Multiplying the $m$ inequalities of the form of (32), we get

$$1 - a_{m\mu} \leq (1 - a) \left( \frac{1}{1 + 2a} \right)^m = (1 - a) \exp(-m \log(1 + 2a)).$$

(33)

Since $B \subset V(G) \setminus A_{(m\mu)}$, (30) follows from (33). \[\Box\]

So now, having obtained theorem 40, we can use it to show that $\{G^n\}_{n=1}^\infty$ is a concentrated Lévy family. To formalise this, we need the following.

**Lemma 41** Let $G_1, \ldots, G_n$ be connected graphs, where $G_i$ has $\text{diam}(G_i) = D_i$, $\Delta(G_i) = \Delta_i$, and $\lambda_1(G_i) = \Lambda_i$. Let $G = \prod G_i$. Then

(i) $\text{diam}(G) = \sum D_i$,

(ii) $\Delta(G) = \sum \Delta_i$,

(iii) $\lambda_1(G) = \min \Lambda_i$.

**Proof** (i) For each $i$, suppose $u_i, v_i \in V(G_i)$ has $d(u_i, v_i) = D_i$. Then it is easy to see that $\text{diam}(G) = d((u_1, \ldots, u_n), (v_1, \ldots, v_n)) = \sum D_i$.

(ii) For each $i$, suppose $v_i \in V(G_i)$ has $\text{deg}(v_i) = \Delta_i$. Then it is easy to see that $\Delta(G) = \text{deg}((v_1, \ldots, v_n)) = \sum \Delta_i$. 


(iii) By induction, it is enough to consider the case \( n = 2 \). Firstly, let \( A = \{ a_{ij} \} \) be an \( m \times m \) matrix and \( B = \{ b_{ij} \} \) be an \( n \times n \) matrix. The (right) Kronecker product of \( A \) and \( B \) is the \( mn \times mn \) matrix \( A \otimes B \) given by

\[
A \otimes B = \begin{pmatrix}
a_{11}B & \cdots & a_{1m}B \\
\vdots & \ddots & \vdots \\
a_{m1}B & \cdots & a_{mm}B
\end{pmatrix}.
\]

Note that \( \otimes \) is not commutative, but it is both left and right distributive over +.

Now let \( G_1 = G, G_2 = H, V(G) = \{ u_1, \ldots, u_m \} \), and \( V(H) = \{ v_1, \ldots, v_n \} \). We aim to compute \( Q = Q_{G \times H} \) in the form

\[
\text{diag}(\deg(u_1, v_1), \ldots, \deg(u_1, v_n), \ldots, \deg(u_m, v_1), \ldots, \deg(u_m, v_n)) - A_{G \times H}. \tag{34}
\]

So when \( Q \) is in the form of (34), we can think of it as a block decomposition into \( m^2 \) blocks of \( n \times n \) matrices, \( \{ B_{ij} \}_{i,j=1}^{m} \), say.

\[
Q = \begin{pmatrix}
B_{11} & \cdots & B_{1m} \\
\vdots & \ddots & \vdots \\
B_{m1} & \cdots & B_{mm}
\end{pmatrix}^n.
\]

Now \( B_{ii} \) corresponds to the ‘\( i \)th copy of \( H \)’ in \( G \times H \), for \( 1 \leq i \leq m \). It is clear that \( B_{ii} = Q_H + \deg(u_i)I_n \). Also, for \( i \neq j, 1 \leq i, j \leq m \),

\[
B_{ij} = \begin{cases} 
-1I_n & \text{if } u_iu_j \in E(G), \\
0I_n & \text{if } u_iu_j \not\in E(G).
\end{cases}
\]

Thus

\[
Q = I_m \otimes Q_H + \text{diag}(u_1, \ldots, u_m) \otimes I_n - A_G \otimes I_n \\
\Rightarrow Q = Q_G \otimes I_n + I_m \otimes Q_H. \tag{35}
\]

Now from [22, p. 411, 412], the spectrum of the r.h.s. of (35) is precisely \( \sigma(G) + \sigma(H) \).

Since \( G \) and \( H \) are both connected, we have \( \lambda_1(G \times H) = \min\{ \lambda_1(G), \lambda_1(H) \} \). \( \square \)

Combining theorem 40 and lemma 41, we have:

**Theorem 42** Let \( G \) be a connected graph, with \( \text{diam}(G) = D \) and \( \Delta(G) = \Delta \). Then \( \{ G^n \}_{n=1}^{\infty} \) is a concentrated Lévy family with exponent \( C_2 = (\lambda_1/2\Delta)^{1/2}D \log 2 \).

**Proof** Apply (31) and lemma 41 with \( \text{diam}(G^n) = n \text{diam}(G), \Delta(G^n) = n\Delta(G) \) and \( \lambda_1(G^n) = \lambda_1(G) \). \( \square \)
3.2 Eigenvalue techniques

To end this chapter, we remark that it is still very much an open problem to find good isoperimetric inequalities on $G^n$ with given $\text{diam}(G)$. The best isoperimetric inequalities on many of the simplest examples are still not known. For example, the best isoperimetric inequality on $K^n_k$, where $K_k$ is the complete graph of order $k$, is not known.

\[^{1}\text{See Appendix A for definition}\]
4 The chromatic number of random graphs

In this chapter, we focus on one of the most notorious problems where the applications of isoperimetric inequalities and martingales are important. The problem is: to determine the asymptotic behaviour of the chromatic number of almost every random graph, as the order goes to infinity.

4.1 Random graphs, cliques and independent sets

The theory of random graphs was founded by Erdős and Rényi in the late 1950s. Although it is one of the youngest branches of graph theory, its importance nowadays is second to none. The theory became a powerful tool in proving the existence of a graph (or else a similar structure) with a desired property, by proving that the probability of its existence has a positive measure (over some suitable probability space). This gives us no clue as to how to actually construct such a graph. This idea is the so-called probabilistic method. It can be unexpectedly applied to other branches of mathematics, such as number theory and Fourier analysis.

The study of random graphs is also of great interest in its own right. There are many models of random graphs, but here we will only be interested in the model $G(n,p)$, which we now define.

The model $G(n,p)$ consists of all possible graphs on $n$ vertices, where each of the $N = \binom{n}{2}$ edges exists (independently) with probability $p$, $0 < p < 1$ ($p$ may depend on $n$). We write $G_{n,p}$ for an element of $G(n,p)$, and call it a random graph of $G(n,p)$. We have a probability measure on $G(n,p)$, with the probability of a random graph $G_{n,p}$ with $m$ edges being $p^m q^{N-m}$, where $q = 1 - p$. Also, we see that $|G(n,p)| = 2^N$. If $V$ is the vertex set of $K_n$, we write $V^{(2)}$ for the set of all edges of $K_n$ (ie: the set all pairs of vertices).

The other model of random graphs frequently studied is $G(n,M)$, and there are some other less trivial models, such as $G_{k-out}(n)$ and $G_{r-reg}(n)$. See [5, ch. 2] or [7] for details.

Naturally, we can identify $G(n,p)$ with the weighted cube $Q_N(p)$, namely, we take an enumeration of $V^{(2)}$ by $\{1, \ldots , N\}$, and identify this with the $[N]$ associated with $Q_N(p)$ (or else, consider 0-1 sequences of length $N$). In particular, $G(n,1/2)$ can be identified with the discrete cube $Q_N$. So a random graph $G_{n,p}$ is identified with a vertex of $Q_N(p)$. Hence any real-valued function, or a graph invariant on $G(n,p)$, becomes a random variable.
We aim to show that certain graph invariants are ‘almost constant’, in a ‘highly concentrated’ sense, thus giving us particular properties that ‘almost all’ the random graphs have. In proving such a concentration result, we will consistently adopt Landau’s $O, o, \sim$ notation\(^\dagger\) in our proofs. We will look at two types of methods to tackle such a problem. The first is the so-called second moment method, which is essentially the application of Chebyshev’s inequality, and in turn, we have to estimate the expectation, variance, and in general, higher moments of our graph invariants. This method can only yield a bound for the probability of failure of $O(n^{-c})$ for any constant $c$, and often, such a polynomial bound can be inadequate. So our second idea is to apply the isoperimetric inequalities and martingale inequalities on $Q_N(p)$ and $Q_N$, derived in chapters 1 and 2, to our graph invariants. Here, we can obtain exponentially small bounds for the probability of failure.

As a first striking example, let us consider the clique number. For a graph $G$, a complete subgraph of maximal order is a clique of $G$, and we write $\text{cl}(G)$, the clique number, for the order of a clique of $G$. Also, an independent set of $G$ is an empty subgraph of $G\(^\dagger\)$, and we write $\text{ind}(G)$, the independence number, for the order of a maximal independent set of $G$. Note that if $G^c$ is the complement of $G$ in $K_{|G|}$, then $\text{ind}(G) = \text{cl}(G^c)$ (see figure 28).

Now suppose that $\Omega_n$ is a probability space of graphs of order $n$. Let $Q$ be a property of graphs (eg: $Q$ is the property ‘$\text{cl}(G) \leq n/2$’; note that an ‘interesting property’ $Q$ often depends on $n$, and possibly on other parameters that $\Omega_n$ may have, such as $p$ in $\mathcal{G}(n,p)$).

\(^\dagger\)See Appendices A and B for definitions
The chromatic number of random graphs

We say that almost every (a.e.) graph in $\Omega_n$ has the property $Q$ if $P(G \in \Omega_n : G has Q) \to 1$ as $n \to \infty$. So here, we are considering $\Omega_n = \mathcal{G}(n, p)$.

Matula (1970, 1972, 1976) was the first to observe that when $p$ is constant, the distribution of the clique numbers of the random graphs in $\mathcal{G}(n, p)$ is highly concentrated, i.e. almost every $G_{n, p}$ has one of a few possible values for the clique number. Claims to this phenomenon were further strengthened by Grimmett and McDiarmid (1975) in [19] and by Bollobás and Erdős (1976) in [13].

This problem was tackled by the second moment method, which we now describe. Let $X$ be a non-negative graph invariant on $\mathcal{G}(n, p)$. We aim to find an arbitrarily small upper bound for $P(X = 0)$, or else show somehow that $X$ is large or non-zero for a.e. $G_{n, p}$. From Markov's inequality, the expected value $E(X)$ on its own, can rarely be useful. We require a stronger inequality which involves the variance of $X$. If $\mu = E(X)$, we define the variance of $X$ by

$$\text{var}(X) = \sigma^2(X) = E((X - \mu)^2) = E(X^2) - \mu^2.$$ 

$\sigma = \sigma(X) \geq 0$ is the standard deviation of $X$. The variance roughly measures the amount that $X$ tends to deviate from $\mu$.

**Lemma 43 (Chebyshev's inequality)** Let $c \geq 0$ and $X \in L^2(\mathbb{P})$. Then

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2},$$

where $\mu = E(X)$, and $\sigma^2 = \sigma^2(X) = \text{var}(X)$. In particular

$$P(X = 0) \leq P(|X - \mu| \geq \mu) \leq \frac{\sigma^2}{\mu^2}.$$  

**Proof** Apply Markov’s inequality with $g(x) = x^2$ and $X = |Y - \mu|$. \hfill \Box

Thus Chebyshev’s inequality gives us a more promising bound that relies on more information about the distribution of $X$.

Before we look at the main theorem concerning the distribution of $\text{cl}(G_{n, p})$, we require a lemma that allows us to calculate certain expectations.

**Lemma 44** Let $F$ be a fixed graph on at most $n$ vertices. Let $X_F = X_F(G_{n, p})$ be the number of subgraphs of $G_{n, p}$ isomorphic$^\dagger$ to $F$. Then

$$E(X_F) = N_F p^{E(F)},$$

where $N_F$ is the number of subgraphs of $K_n$ isomorphic to $F$.

More generally, if $\mathcal{F} = \{F_1, F_2, \ldots\}$ is a family of fixed graphs, $X_{F_i}$ is the number of

$^\dagger$See Appendix B for definitions
4.1 Random graphs, cliques and independent sets

subgraphs of $G_{n,p}$ isomorphic to $F_i$, and $X_F$ is the total number of subgraphs of $G_{n,p}$, each isomorphic to some $F_i$, then $X_F = \sum X_{F_i}$ and

$$E(X_F) = \sum N_{F_i} p^{E(F_i)}.$$

In particular, if $X_s = X_s(G_{n,p})$ is the number of complete subgraphs of order $s$ in $G_{n,p}$, then

$$E(X_s) = \binom{n}{s} p^{\binom{s}{2}}.$$

The proof of lemma 44 requires writing $X_F$ as a sum of indicator functions\(^\dagger\), details of which we do not discuss here. See [12, p. 218, 219].

With lemma 43 and lemma 44, we can now show that $\text{cl}(G_{n,p})$ is highly concentrated. The following result says that in fact, for a.e. $G_{n,p}$, $\text{cl}(G_{n,p})$ can take one of at most two values.

**Theorem 45** Let $0 < p < 1$ be constant. Then the clique number of a.e. $G_{n,p}$ is $r$ or $r + 1$, for some $r = r(n,p)$ with $1 \leq r \leq n - 1$.

**Proof** As before, let $X_s = X_s(G_{n,p})$ denote the number of $K_s$ subgraphs in $G_{n,p}$. So $E(X_s) = \binom{n}{s} p^{\binom{s}{2}}$. Since $E(X_1) = n$ and $E(X_n) = p^{\binom{n}{2}} < 1$, we can choose $r = r(n,p)$, with $1 \leq r \leq n - 1$, to be the largest natural number such that

$$E(X_r) = \binom{n}{r} p^{\binom{r}{2}} \geq \log n. \quad (36)$$

We claim that for a.e. $G_{n,p}$, $\text{cl}(G_{n,p}) = r$ or $\text{cl}(G_{n,p}) = r + 1$. Note that (36) implies that

$$r = 2 \log_b n + O(\log \log n), \quad (37)$$

where as usual, $\log_a n = \log n / \log a$. Thus the assertion is equivalent to proving that

$$P(X_{r+2} \geq 1) \to 0, \quad (38)$$

$$P(X_r \geq 1) \to 1. \quad (39)$$

Write $b = 1/p$ for convenience. Applying estimates induced by *Stirling’s formula\(^\dagger\), it can be shown that (eg: see [5, ch. 11] or [12, ch. 7])

$$\frac{n}{\log_b n} < \frac{n}{r} < p^{-r/2} < n. \quad (40)$$

\(^\dagger\)See Appendix B for definitions.
Also, by the definition of $r$, we have $\mathbb{E}(X_{r+1}) < \log n$. Via (40),

$$P(X_{r+2} \geq 1) \leq \mathbb{E}(X_{r+2}) = \left(\frac{n}{r+2}\right)p^{\binom{r+2}{2}}\mathbb{E}(X_{r+1}) = \frac{n-r-1}{r+2}p^{r+1}\mathbb{E}(X_{r+1})$$

$$< \frac{n \log^2 n}{3} \frac{n^2}{n^2} \log n < \frac{1}{3 \log^2 b} \frac{n^{3/4}}{n} \rightarrow 0$$

since $\log n < n^{1/4}$ for $n$ sufficiently large. This proves (38).

To prove (39), since $\mu_r = \mathbb{E}(X_r) \geq \log n \rightarrow \infty$, so by Chebyshev’s inequality, it is enough to prove that $\sigma_r/\mu_r \rightarrow 0$, where $\sigma_r = \sigma(X_r)$.

We have to calculate $\sigma_r$. So we have to calculate $\mathbb{E}(X_r^2)$, which is the expected number of ordered pairs of complete graphs of order $r$ in $G_{n,p}$ (repetitions allowed). We aim to calculate the number of such pairs with $k$ vertices in common, then sum from $k = 0$ to $k = r$. In $K_n$, there are $\binom{n}{r}$ subgraphs $K_r$. Fix one of these, say $K'_r$. Within $K'_r$, there are $\binom{r}{k}$ subsets of $k$ vertices, and in $V(K_n) \setminus V(K'_r)$, there are $\binom{n-r}{r-k}$ sets of $r-k$ vertices, provided that $n-r \geq r-k$, or else we ignore the cases when $n-r < r-k$. Hence there are $\binom{n}{k}\binom{n-r}{r-k}$ ordered pairs of $K_r$ with $k$ vertices in common (see figure 29). If $k = 0$ or 1, the number of edges spanned by such a pair of $K_r$ is $2\binom{r}{2}$. If $k \geq 2$, this number is $2\binom{r}{2} - \binom{k}{2}$.

$$\begin{align*}
\sum_{k=0}^{r} \binom{n}{r} \binom{r}{k} \binom{n-r}{r-k} p^{2\binom{r}{2} - \binom{k}{2}}.
\end{align*}$$

So by lemma 44

$$\mathbb{E}(X_r^2) = \sum_{k=0}^{r} \binom{n}{r} \binom{r}{k} \binom{n-r}{r-k} p^{2\binom{r}{2} - \binom{k}{2}},$$

whenever $\binom{n-r}{r-k}$ and $\binom{k}{2}$ exist (ie: by convention, we take $\binom{n-r}{r-k} = 0$ if $n-r < r-k$ and $\binom{k}{2} = 0$ if $k < 2$). We can assume that $r \geq 2$ as the case $r = 1$ is clearly trivial.

The following results can be verified (where the binomial coefficients exist)

$$\begin{align*}
\binom{n}{r} &= \sum_{k=0}^{r} \binom{r}{k} \binom{n-r}{r-k},
\end{align*}$$

and

$$\begin{align*}
\binom{r}{k} \binom{n-r}{r-k} / \binom{n}{r} &= 2 \frac{r!^2}{k!(r-k)!^2} n^{-k}.
\end{align*}$$
\[ \frac{\sigma_r^2}{\mu_r^2} = \frac{\mathbb{E}(X_r^2) - \mu_r^2}{\mu_r^2} \]
\[ = \left( \sum_{k=0}^{r} \binom{n}{r} \binom{r}{k} \binom{n-r}{r-k} p^2(\binom{r}{2}) - \binom{n}{r} p^2(\binom{r}{2}) \right) / \left( \binom{n}{r} \right)^2 p^2(\binom{r}{2}) \]
\[ = \sum_{k=2}^{r} \binom{r}{k} \binom{n-r}{r-k} \left( p^{-\binom{r}{2}} - 1 \right) / \left( \binom{n}{r} \right) \]
\[ \leq 2 \sum_{k=2}^{r} \frac{r!^2}{k!(r-k)!^2} n^{-k} p^{-\binom{r}{2}} = 2 \sum_{k=2}^{r} \eta_k. \]

Rough calculations show that for \( 3 \leq k \leq r - 1 \), we have \( \eta_k \leq \eta_3 + \eta_{r-1} \). Thus
\[ \frac{\sigma_r^2}{\mu_r^2} \leq 2(\eta_2 + \eta_r) + 2r(\eta_3 + \eta_{r-1}). \]  \hfill (44)

Now for \( n \) sufficiently large, we have (via (40))
\[ 2\eta_2 \leq r^4 n^{-2} p^{-1} = O(n^{-1}), \]
\[ 2\eta_r = 2r!n^{-r} p^{-\binom{r}{2}} \leq 2 \binom{n}{r}^{-1} p^{-\binom{r}{2}} = 2/\mu_r = o(1), \]
\[ 2r\eta_3 \leq 2r^7 n^{-3} p^{-3} = O(n^{-2}), \]
\[ 2r\eta_{r-1} = 2r^2 n! n^{-r} p^{-r(r-1)/2} = 2r^2 n p^{-1} \eta_r < 2r^4 n^{-1} p^{-1} \eta_r = O(n^{-1/2}). \]

Putting these estimates into (44) gives
\[ \frac{\sigma_r^2}{\mu_r^2} = O(n^{-1/2}) + o(1) = o(1), \]
proving (39).

\[ \square \]

4.2 The chromatic number of a.e. \( G_{n,p} \)

Given a graph \( G \), a vertex \( k \)-colouring, or simply a \( k \)-colouring of \( G \), is a surjective function \( c : V(G) \to \{1, 2, \ldots, k\} \) such that \( c(x) \neq c(y) \) if \( xy \in E(G) \). In other words, we colour the vertices in a way that adjacent vertices receive different colours. The chromatic number of \( G \), \( \chi(G) \), is the minimal integer \( k \) such that \( G \) has an \( k \)-colouring. For example, in figure 30, \( \chi(E_n) = 1 \), \( \chi(K_n) = n \), \( \chi(P_n) = 2 \) \( (n \geq 1) \), \( \chi(C_n) = 2 \) if \( n \) is even, \( \chi(C_n) = 3 \) if \( n \) is odd, \( \chi(G) = 2 \) if \( G \) is a forest \( (E(G) \neq \emptyset) \), \( \chi(K_{ab}) = 2 \) \( (E(K_{ab}) \neq \emptyset) \), and so on.
The chromatic number of random graphs

So how may one determine \( \chi(G) \)? In general, that is not an easy task. We do have the trivial inequality \( \chi(G) \geq \text{cl}(G) \), but this inequality can be poor. Erdős has shown that \( \chi(G) \) can be arbitrarily large even when \( \text{cl}(G) \) is small. Another lower bound for \( \chi(G) \) is seen to be

\[
\chi(G) \geq \frac{|G|}{\text{ind}(G)}.
\]  

(45) follows from the fact that in a \( \chi(G) \)-colouring, every colour class (i.e. set of vertices with the same colour) is an independent set.

We are interested in the behaviour of \( \chi(G_{n,p}) \). Here, we will only consider the case when \( p \) is fixed. We aim to show that, as \( n \to \infty \), \( \chi(G_{n,p}) \) is about the same for a.e. \( G_{n,p} \). So this is another example where a graph invariant is almost constant, and one that has a very interesting history.

By (37), \( \text{cl}(G_{n,p}) \sim 2 \log_d n \). Since the complement of \( G_{n,p} \) in \( K_n \) is a random graph \( G_{n,q} \), where \( q = 1 - p \), the distribution of \( \text{ind}(G_{n,p}) \) is precisely the distribution of \( \text{cl}(G_{n,q}) \). Thus \( \text{ind}(G_{n,p}) \sim 2 \log_d n \), where \( d = 1/q \). Hence (45) implies the following.

**Theorem 46** For a.e. \( G_{n,p} \)

\[
\chi(G_{n,p}) \geq (1 + o(1)) \frac{n}{2 \log_d n}.
\]

\[ \square \]
4.2 The chromatic number of a.e. $G_{n,p}$

Note that the inequality $\chi(G) \geq cl(G)$ only yields a lower bound for $\chi(G_{n,p})$ of $O(\log n)$, which is much weaker than theorem 46.

What about an upper bound? It had been conjectured for many years that this is equal to the lower bound of theorem 46. The common way of finding an upper bound has been to analyse the result of a colouring algorithm. The so-called greedy algorithm (see [12, p. 146] for description, [13] for method) does produce a colouring of a.e. $G_{n,p}$ that requires just twice the number of colours of the lower bound of theorem 46:

\[
(1 + o(1)) \frac{n}{\log_d n} \leq \chi(G_{n,p}) \leq (1 + o(1)) \frac{n}{\log_d n}.
\]

(46) was first noticed by Matula, and minor improvements, only in the $o(1)$ term, were made by Grimmett and McDiarmid [19], and Bollobás and Erdős [13]. In 1987, Shamir and Spencer [27] made a breakthrough by using martingale inequalities to show that the distribution of $\chi(G_{n,p})$ is ‘highly concentrated in a short interval’. This came from the following result which played a vital role around the time of the breakthrough. It is based on what is sometimes referred to as Doob’s martingale process, which is essentially a variation of what we have already seen in chapter 2. The result is in fact an analogue of theorem 26, and hence of Azuma’s inequality.

**Theorem 47** Let $S_0 = \emptyset \subset S_1 \subset \cdots \subset S_m = V^{(2)}$ and let $f : \mathcal{G}(n, p) \to \mathbb{R}$ be such that whenever $E(G) \triangle E(H) \subset S_i \setminus S_{i-1}$, we have $|f(G) - f(H)| \leq c_i$. Set $s = \sum_{i=1}^m c_i^2$. Then for $a > 0$ we have

\[
P(|f - E(f)| \geq a) \leq 2 \exp(-a^2/2s).
\]

**Proof** We naturally identify $S_0 = \emptyset \subset S_1 \subset \cdots \subset S_m = V^{(2)}$ with a sequence of nested equivalence relations $\equiv_0, \ldots, \equiv_m$ on $\mathcal{G}(n, p)$ as follows. For $G, H \in \mathcal{G}(n, p)$, we set $G \equiv_i H$ iff $E(G) \cap S_i = E(H) \cap S_i$. This induces the sequence of partitions $\mathcal{P}_0 \prec \mathcal{P}_1 \prec \cdots \prec \mathcal{P}_m$ on $\mathcal{G}(n, p)$, with $G$ and $H$ belonging to the same set of $\mathcal{P}_i$ iff $G \equiv_i H$. Clearly, $\forall i$, the sets of $\mathcal{P}_i$ have the same size, namely, $2^{N-|S_i|}$.

Suppose $A, B \in \mathcal{P}_i$, $A \neq B$, with $A \cup B \subseteq C$ for some $C \in \mathcal{P}_{i-1}$. Then for some fixed $E_i, F_i \subseteq S_i$, we have $A = \{G_{n,p} : E(G_{n,p}) \cap S_i = E_i\}$ and $B = \{G_{n,p} : E(G_{n,p}) \cap S_i = F_i\}$ (see figure 31). We see that $E_i \triangle F_i \subset S_i \setminus S_{i-1}$. Indeed, suppose $E_i \not\subseteq F_i$ and let $e \in E_i \setminus F_i \subseteq S_i$. So $e \in E(G)$ for some $G \in A$, and $e \not\in E(H)$ for some $H \in B$. But $A \cup B \subseteq C \in \mathcal{P}_{i-1}$, so $E(G) \cap S_{i-1} = E(H) \cap S_{i-1}$, and it follows that $e \not\in S_{i-1}$. Now switch the roles of $E_i$ and $F_i$ (if necessary) to get $E_i \triangle F_i \subset S_i \setminus S_{i-1}$. 


The chromatic number of random graphs

Now since $|A| = |B|$, for $G \in A$ define the bijection $\phi : A \rightarrow B$ by ‘replacing $E_i$ by $F_i’$, ie: define $H = \phi(G) \in B$ by $E(H) = (E(G) \setminus E_i) \cup F_i$. Then $\phi$ satisfies

$$P(A') = \frac{P(\phi(A'))}{P(B)} \quad \forall \ A' \subset A. \quad (47)$$

Moreover, $E(G) \triangle E(\phi(G)) \subset S_i \setminus S_{i-1} \forall G \in A$. Thus $|f(G) - f(\phi(G))| \leq c_i$.

From here, via (47), it is essentially a variation of the argument used in theorem 26 to derive the required inequality (where here, $P$ is not necessarily uniform and the martingale notion is more general).

So how do we make use of theorem 47? How do we choose the sets $S_0, \ldots, S_m$? We would need to think about the graph invariant whose concentration we wish to establish. But there are two natural choices for $S_0, \ldots, S_m$. In what is known as the vertex exposure of $K_n$, we take an enumeration of the vertices by $\{1, \ldots, n\}$ and take $S_i$ to be the complete subgraph spanned by $\{1, \ldots, i+1\}$, so that $|S_i| = \frac{i+1}{2}$ and $m = n-1$. In the edge exposure of $K_n$, we take an enumeration of the edges of $K_n$ by $\{1, \ldots, N\}$ and take $S_i$ to be the first $i$ edges of the enumeration, so that $|S_i| = i$ and $m = N$. Often, it is helpful to consider the vertex exposure as a subsequence of the edge exposure. For example, for $n = 4$:

**Figure 32**
4.2 The chromatic number of a.e. $G_{n,p}$

Taking the vertex exposure of $K_n$ for $S_0, \ldots, S_m$ and $f = \chi(G_{n,p})$, we aim to show that $\chi(G_{n,p})$ is tightly concentrated. We say that a sequence $X(G_{n,p})$ where $p = p(n)$ is concentrated in width $w = w(n, p)$ if for some $u = u(n, p)$, we have

$$\mathbb{P}(u \leq X(G_{n,p}) \leq u + w) \rightarrow 1.$$ (48)

With (48), we have the following result by Shamir and Spencer:

**Theorem 48 (Shamir and Spencer, 1987)** For $p$ fixed and a function $\omega(n) \to \infty$ arbitrarily slowly, there are sequences $\{a_n\}$ and $\{b_n\}$ such that $b_n \leq a_n + \omega(n)n^{1/2}$ and $a_n \leq \chi(G_{n,p}) \leq b_n$ for a.e. $G_{n,p}$.

**Proof** Take $S_0, \ldots, S_m$ to be the vertex exposure of $K_n$. If $G, H \in \mathcal{G}(n, p)$ satisfy $E(G) \triangle E(H) \subset S_i \setminus S_{i-1}$, then $G$ and $H$ differ only in some edges incident with the $i$th vertex, so certainly, $|\chi(G) - \chi(H)| \leq 1$ (see figure 33).

![Figure 33](image)

Now apply theorem 47 with $s = n - 1$:

$$\mathbb{P}\left(|\chi(G_{n,p}) - E(\chi)| \geq \omega(n)n^{1/2}/2\right) \leq 2\exp(-\omega^2(n)/8) = o(1).$$

So we can take $a_n = E(\chi) - \frac{1}{2}\omega(n)n^{1/2}$ and $b_n = E(\chi) + \frac{1}{2}\omega(n)n^{1/2}$ to satisfy the conditions of the theorem. \hfill \Box

So theorem 48 tells us that $\chi(G_{n,p})$ is concentrated in width $n^{1/2}\omega(n)$. This means that, for arbitrarily large $n$ and arbitrarily small width, we can choose $\omega(n)$ to go to infinity so slowly, so that the concentration falls within the given width. Notice that theorem 48 tells us nothing about the location of the concentration, nor does it imply a possible improvement of the inequalities of (46).

In the same year, Matula [25] used an intelligent algorithm to reduce the interval in (46) to one-third of the length:

**Theorem 49 (Matula, 1987)** For a.e. $G_{n,p}$

$$(1 + o(1)) \frac{n}{2 \log_2 n} \leq \chi(G_{n,p}) \leq \left(\frac{2}{3} + o(1)\right) \frac{n}{\log_2 n}.$$ \hfill \Box
About a year later, Bollobás [8] used martingale inequalities in a different way to show that for a.e. $G_{n,p}$, $\chi(G_{n,p})$ is in fact equal to the lower bound of (46):

**Theorem 50 (Bollobás, 1988)** For a.e. $G_{n,p}$

$$\chi(G_{n,p}) = (1 + o(1)) \frac{n}{2 \log_d n}.$$ 

In theorem 48, if we were to find the location of the short interval $(a_n, b_n)$, we would have to estimate $\mathbb{E}(\chi(G_{n,p}))$, the value where $\chi(G_{n,p})$ is very sharply concentrated. Unfortunately, that is not easy.

In fact, to prove theorem 50, we can apply theorem 47 in a completely different way than the one that gave theorem 48. We can show that for a.e. $G_{n,p}$, $\chi(G_{n,p})$ is only a little larger than the lower bound of (46).

More precisely, the plan to prove theorem 50 is as follows. Instead of considering $\chi(G_{n,p})$ as our graph invariant, we consider $Y_r(G_{n,p})$, the number of independent sets of order $r$ in $G_{n,p}$. Consider $s = \mathbb{E}(\text{ind}(G_{n,p}))$, the expectation of the independence number of $G_{n,p}$ (so $s$ is about $2 \log_d n$). If we let $r$ be slightly less than $s$, we can apply theorem 47 to show that $Y_r(G_{n,p})$ is highly concentrated about $\mathbb{E}(Y_r)$, with probability exponentially close to 1. In turn, this means that for a.e. $G_{n,p}$, every vertex subset of fairly large order (in particular, much larger than $2 \log_d n$), say of $O(n/\log_d n)$, must contain an independent set of order $r$, if $\mathbb{E}(Y_r)$ is about $O(n^2/\log^4 n)$ (see figure 34). We will then apply a colouring argument to show that for every such $G_{n,p}$, $\chi(G_{n,p}) \leq n/r + n/\log^2 n = (1 + o(1))n/r$. It remains to verify that $r \sim s$, and this essentially solves the problem.

![Figure 34](image)

First, we establish the concentration result. We shall in fact consider $Y_r(G_{n,p})$, the number of complete graphs of order $r$, rather than $Y_r'(G_{n,p})$, since we have already studied complete graphs in section 4.1 to some extent. Then to get our result, just take complementations thereafter. We have the following result for $Y_r(G_{n,p})$. 
4.2 The chromatic number of a.e. $G_{n,p}$

**Lemma 51** Let $0 < p < 1$ be fixed and let $Y = Y(G_{n,p})$ be the number of complete graphs of order $r$ in $G_{n,p}$. If $E(Y) = \binom{n}{r}p^G = n^\alpha = O\left(\frac{n^2}{\log n}\right)$ and $\alpha \geq 1$, then for $0 \leq c \leq 1$

$$\mathbb{P}(Y \leq (1-c)n^\alpha) \leq \exp\left(-\left(c^2 + o(1)\right)n^{2\alpha/2}\right). \quad (49)$$

**Proof** We aim to apply theorem 47. Suppose we were to take the finest graduation for $S_0 = \emptyset \subset S_1 \subset \cdots \subset S_m = V(G)$, i.e. the edge exposure, in order to keep the martingale differences to be small. Unfortunately, not even this choice will work: if $E(G) \triangle E(H) \subset S_i \setminus S_{i-1}$, i.e. if $G$ and $H$ differ in at most one edge, $|Y(G) - Y(H)|$ may be as large as $n_{r-2}^\alpha$, which is too large (i.e. if we insert one edge, we can create as many as $n_{r-2}^\alpha$ new $K_r$).

The trick is to introduce another graph invariant related to $Y$. We choose a graph invariant so that in adding an edge to a certain graph, the corresponding martingale difference does not change much. So define $Z(G_{n,p})$ to be the maximal number of $K_r$ which are edge-disjoint. Taking the same graduation and letting $(Z_i)_{i=0}^N$ be the corresponding martingale, we have $|Z_i - Z_{i-1}| \leq 1$ since $Z$ changes by at most 1 when we add an edge. It can be shown that $E(Z) = (1 + o(1))n^\alpha$. So by theorem 47

$$\mathbb{P}(Y \leq (1-c)n^\alpha) \leq \mathbb{P}(Z \leq (1-c)n^\alpha) \leq \exp\left(-\left(c^2 + o(1)\right)n^{2\alpha/2N}\right).$$

**Remark** In fact, we can also prove a similar result to lemma 51 by using the isoperimetric inequalities on $Q_N(p)$ (corollary 17, Bollobás and Leader) and on $Q_N$ (corollary 4, Harper) from earlier, instead of Azuma’s inequality. For example, if $p = 1/2$, we can find a bound for $\mathbb{P}(Z = 0)$ as follows. It can be shown that $\mathbb{P}(Z \geq n^\alpha/2) \geq 1/2$. Let $A = \{G_{n,1/2} : Z(G_{n,1/2}) \geq n^\alpha/2\}$, so that $\mathbb{P}(A) \geq 1/2$. So $Z \geq 1$ in $A_{n^\alpha/2}$ and by Harper’s inequality,

$$\mathbb{P}(Z = 0) \leq \exp\left(-2\left(n^\alpha/2\right)^2/N\right) \leq \exp\left(-n^{2\alpha/2}\right).$$

A similar bound can be obtained from Bollobás and Leader’s result for any $0 < p < 1$.

So now having obtained lemma 51, it is essentially a straightforward argument to prove the chromatic number problem.

**Proof of theorem 50** By theorem 46, we need to show that for a.e. $G_{n,p}$, we have $\chi(G_{n,p}) \leq (1 + o(1))n/2\log_d n$. For $k \leq n$, write $E(k,s) = E_p(k,s) = \binom{k}{s}$ for the expected number of independent sets of size $s$ in $G_{k,p} \in \mathcal{G}(k,p)$, i.e. in a subgraph of $G_{n,p} \in \mathcal{G}(n,p)$ of order $k$. Setting $s_0 = \max\{s : E(n,s) \geq 1\}$, it is easy to see that $s_0 = 2\log_d n - 2\log_d \log n + O(1)$. Now set $s_1 = \lfloor s_0 - 6\log_d \log n \rfloor$ and $n_0 = \lfloor n/\log^2 n \rfloor$. 
Then it can be shown that \( \mathbb{E}(n_0, s_1) \geq n_0^2 \).

Let \( \Omega = \{ G_{n,p} : G_{n,p} \text{ has } n_0 \text{ vertices not containing an independent set of size } s_1 \} \).

Since we can find \( p_0 \geq p \) such that \( \mathbb{E}_{p_0}(n_0, s_1) = n_0^2 / \log^5 n \), lemma 51 gives that the probability that \( G_{n,p} \) contains a vertex set of size \( n_0 \) not containing an independent set of size \( s_1 \) is at most \( \exp\left(-n_0^2 / \log^{11} n\right) \). Thus

\[
\mathbb{P}(\Omega) \leq \exp\left(-n_0^2 / \log^{11} n\right) \left(\frac{n}{n_0}\right) = o(1). \tag{50}
\]

(50) can be seen from, say, the inequality \( \left(\frac{n}{n_0}\right) \leq \left(\frac{en}{n_0}\right)^{n_0} \). So for \( G_{n,p} \in \mathcal{G}(n,p) \setminus \Omega \), colour it as follows. Remove \( s_1 \)-element independent sets successively and give each one a distinct colour. When we are left with less than \( n_0 \) vertices, give each of these vertices a distinct colour (see figure 35).

![Figure 35](image-url)

So altogether

\[
\chi(G_{n,p}) \leq \left\lfloor \frac{n - n_0}{s_1} \right\rfloor + n_0 \leq \frac{n}{s_1} + n_0
\]

\[
= \frac{n}{2\log_d n} \left(1 + o(1)\right) + O\left(\frac{n}{\log^2 n}\right)
\]

\[
= \frac{n}{2\log_d n} \left(1 + o(1)\right).
\]

This happens for a.e. \( G_{n,p} \), so theorem 50 is proved. \( \Box \)

To finish this chapter, we shall briefly mention that there are several extensions to the chromatic number problem, namely, when \( p = p(n) \). One case that has been widely investigated is when \( p = n^{-\alpha} \) for \( 0 < \alpha < 1 \) (or sometimes in a shorter interval). Again, with the aid of Doob’s martingale process, similar concentration results for \( \chi(G_{n,p}) \) in small widths can be derived. See [8] and [27] for details. It is also possible to consider the problem on the other models of random graphs mentioned at the start of this chapter. See [7].
A Appendix A
Pre-requisites from C395 (Graph Theory and Combinatorics)

A.1. Graphs

• A graph is a pair $G = (V, E)$, where $V = V(G)$ is the set of vertices of $G$ and $E = E(G)$ is the set of edges of $G$ joining certain pairs of vertices. We do not allow more than one edge joining any two vertices, nor an edge looping from a vertex back to itself.

• Main examples of graphs (figure 36)

<table>
<thead>
<tr>
<th>Graph Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty graph $E_n$</td>
<td>$</td>
</tr>
<tr>
<td>Complete graph $K_n$</td>
<td>$</td>
</tr>
<tr>
<td>Cycle $C_n$ $(n \geq 3)$</td>
<td>$</td>
</tr>
<tr>
<td>Path $P_n$</td>
<td>$</td>
</tr>
</tbody>
</table>

Figure 36 Examples of graphs ($n = 6$ in all cases)

• Two vertices $x, y \in V(G)$ are adjacent iff $xy \in E(G)$. For $x \in V(G)$, the neighbourhood of $x$ is $\Gamma(x) = \{y \in V(G) : x$ and $y$ are adjacent$\}$. More generally, for $A \subseteq V(G)$, the neighbourhood of $A$ is $\Gamma(A) = \bigcup_{x \in A} \Gamma(x)$.

A.2. Subgraphs

• For a graph $G = (V, E)$, $G' = (V', E')$ is a subgraph of $G$ if $V' \subseteq V$ and $E' \subseteq E$.

• $G'$ is the induced subgraph spanned by the vertices of $V'$ if whenever $x, y \in V'$ and $xy \in E$, we have $xy \in E'$. In this case, we write $G' = G[V']$, since $V'$ determines $G'$.  

A.3. Paths and connectedness

- For a graph $G$ and $x, y \in V(G)$, a path from $x$ to $y$ is a sequence of distinct vertices $x_1, \ldots, x_k \in V(G)$ with $x_1 = x$, $x_k = y$ such that (if possible) $x_i x_{i+1} \in E(G)$, for all $1 \leq i \leq k - 1$.

- A shortest path from $x$ to $y$ is a path from $x$ to $y$ (if it exists) with $k$ minimal.

- $G$ is connected if $\forall x, y \in V(G)$, there exists a path from $x$ to $y$ in $G$.

A.4. Vertex degrees

- For a graph $G$ and $x \in V(G)$, the degree of $x$ is $\deg(x) = |\{y : xy \in E(G)\}| = |\Gamma(x)|$.

- The minimum degree of $G$ is $\delta(G) = \min_{x \in V(G)} \deg(x)$, and the maximum degree of $G$ is $\Delta(G) = \max_{x \in V(G)} \deg(x)$.

A.5. Isomorphic graphs

- Graphs $G = (V, E)$ and $H = (V', E')$ are isomorphic if there exists a bijection $\phi : V \to V'$ such that $xy \in E$ iff $\phi(x)\phi(y) \in E'$. 

B Appendix B
Pre-requisites from C327 (Measure Theory) and C393 (Probability)

B.1. $\sigma$-fields and Borel sets

- For a set $\Omega$, a family $\mathcal{F}$ of subsets of $\Omega$ is a $\sigma$-field (on $\Omega$) if
  
  (i) $\emptyset, \Omega \in \mathcal{F}$,
  
  (ii) $A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$,
  
  (iii) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

- If $S$ is a family of subsets of $\Omega$, then the $\sigma$-field generated by $S$ is
  
  $\sigma(S) = \bigcap \{ \mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field on } \Omega \text{ containing } S \}$.

Thus $\sigma(S)$ is the smallest $\sigma$-field on $\Omega$ containing $S$.

- The $\sigma$-field of subsets of a topological space $\Omega$ (see B.11. for definition) generated by the family of open subsets of $\Omega$ is the Borel $\sigma$-field, $\mathcal{B}(\Omega)$. Its elements are the Borel sets. The most important case is when $\Omega = \mathbb{R}$, where ‘open sets’ can be replaced by open intervals, or closed intervals, etc.

B.2. Measurable spaces, measures

- A measurable space is a pair $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a $\sigma$-field on the set $\Omega$. The sets of $\mathcal{F}$ are called $\mathcal{F}$-measurable.

- A measure on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ (note that $\infty$ is included) such that
  
  (i) $\mu(\emptyset) = 0$,

  (ii) whenever $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint sets, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

B.3. Measure space and probability space

- A measure space is a triple $(\Omega, \mathcal{F}, \mu)$, where $\mu$ is a measure on $(\Omega, \mathcal{F})$. The sets of $\mathcal{F}$ are called $\mu$-measurable.

- $\mu = \mathbb{P}$ is a probability measure if $\mathbb{P}(\Omega) = 1$. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.
B.4. Events

• For a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), \(\Omega\) is called the sample space, \(\omega \in \Omega\) is a sample point, or an outcome, and an event is an element of \(\mathcal{F}\), ie: it is an \(\mathcal{F}\)-measurable subset of \(\Omega\).

• For events \(A\) and \(B\) with \(\mathbb{P}(B) > 0\), the conditional probability that \(A\) occurs given that \(B\) occurs is defined by
  \[
  \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.
  \]

• A statement \(S\) about outcomes is said to be true almost surely (a.s.) with probability 1 if \(F = \{\omega : S(\omega) \text{ is true}\} \in \mathcal{F}\) and \(\mathbb{P}(F) = 1\).

B.5. Measurable functions, random variables

• For measurable spaces \((\Omega, \mathcal{F})\) and \((\Xi, \mathcal{G})\), a function \(X : \Omega \rightarrow \Xi\) is \(\mathcal{F}\)-measurable if \(X^{-1}(S) \in \mathcal{F}\) \(\forall S \in \mathcal{G}\).

• If \(X : \Omega \rightarrow \mathbb{R}\), then \(X\) is measurable if \(X^{-1}(S) \in \mathcal{F}\) \(\forall S \in \mathcal{B}(\mathbb{R})\).

• In particular, if \(\Omega\) is a topological space, then \(X : \Omega \rightarrow \mathbb{R}\) is Borel measurable, or Borel, if \(X^{-1}(S) \in \mathcal{B}(\Omega) \forall S \in \mathcal{B}(\mathbb{R})\).

• For a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), \(X : \Omega \rightarrow \mathbb{R}\) is a random variable if \(X\) is measurable.

B.6. Independence

• For a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), events \(A\) and \(B\) are independent if
  \[
  \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).
  \]

• Sub-\(\sigma\)-fields \(\mathcal{G}_1, \mathcal{G}_2, \subset \mathcal{F}\) are independent if for all \(A_1 \in \mathcal{G}_1, A_2 \in \mathcal{G}_2\)
  \[
  \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).
  \]

• More generally, sub-\(\sigma\)-fields \(\mathcal{G}_1, \mathcal{G}_2, \ldots \subset \mathcal{F}\) are independent if whenever \(G_i \in \mathcal{G}_i\) \((i \in \mathbb{N})\) and \(i_1, \ldots, i_n\) are distinct, then
  \[
  \mathbb{P}\left(\bigcap_{k=1}^n G_{i_k}\right) = \prod_{k=1}^n \mathbb{P}(G_{i_k}).
  \]

• A random variable \(X\) on \(\Omega\) generates a sub-\(\sigma\)-field
  \[
  \sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\},
  \]
  ie: \(\sigma(X)\) is the smallest sub-\(\sigma\)-field that makes \(X\) measurable.
• The random variables \( X_1, X_2, \ldots \) are independent if the \( \sigma \)-fields \( \sigma(X_1), \sigma(X_2), \ldots \) are independent.

### B.7. Integration

• If \( (\Omega, \mathcal{F}, P) \) is a probability space and \( A \) is an event, the indicator function of \( A \) is the function \( I_A : \Omega \rightarrow \mathbb{R} \) given by

\[
I_A(\omega) = \begin{cases} 
1 & \omega \in A, \\
0 & \omega \notin A.
\end{cases}
\]

• For a random variable \( X \) on \( \Omega \), define \( X^+ = 0 \lor X \) and \( X^- = 0 \lor (-X) \), where \( \lor \) denotes pointwise maximum. Thus \( X = X^+ - X^- \).

• Arithmetic operations on \( \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\} \) are defined naturally. So \( (\pm \infty) - (\pm \infty) \) is not defined, but \( 0 \cdot (\pm \infty) \) is defined as 0.

• If \( X \) is a simple and non-negative random variable on \( \Omega \), so that \( X = \sum_{i=1}^{n} c_i I_{A_i} \) for some \( c_1, \ldots, c_n \in [0, \infty) \) and \( A_1, \ldots, A_n \in \mathcal{F}, \) define

\[
\int_{\Omega} X \, dP = \sum_{i=1}^{n} c_i P(A_i).
\]

• If \( X \) is a non-negative random variable on \( \Omega \), define

\[
\int_{\Omega} X \, dP = \sup \int_{\Omega} \phi \, dP
\]

where sup is taken over all non-negative, simple random variables \( \phi \) on \( \Omega \) with \( \phi \leq X \) pointwise.

• If \( X : \Omega \rightarrow \overline{\mathbb{R}} \), define \( \int_{\Omega} X \, dP = \int_{\Omega} X^+ \, dP - \int_{\Omega} X^- \, dP \), if the r.h.s. is defined.

### B.8. Expectation

• Let \( (\Omega, \mathcal{F}, P) \) be a probability space. A random variable on \( \Omega \) is integrable over \( \Omega \) if \( \int_{\Omega} X \, dP \) is finite, ie: converges.

• If \( X \) is an integrable random variable over \( \Omega \), define the expectation, or expected value of \( X \) by

\[
\mathbb{E}(X) = \int_{\Omega} X \, dP.
\]
• If $X$ is a random variable on $\Omega$, the \textit{distribution function} of $X$ is the function $F_X : \mathbb{R} \to [0, 1]$ given by $F_X(t) = \mathbb{P}(X \leq t)$.

If $f_X$ is a real function such that

$$F_X(t) = \int_{-\infty}^{t} f_X(x) \, dx$$

$\forall t$, then $f_X$ is a \textit{(probability) density function} for $X$. $f_X$ is not unique, but if $g_X$ is another density function for $X$, then $f_X = g_X$ on $\mathbb{R}$, except for a Lebesgue null set (a set $E \subset \mathbb{R}$ is \textit{Lebesgue null}, i.e., has \textit{Lebesgue measure zero} if $\forall \varepsilon > 0$, there are open intervals $I_1, I_2, \ldots$ such that $E \subset \bigcup_{i=1}^{\infty} I_i$ and $\sum_{i=1}^{\infty} \text{length}(I_i) < \varepsilon$).

• If $X$ is a random variable on $\Omega$ with density function $f_X$, it can be shown that

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) \, dx$$

(see, eg, [28, p. 68]).

• For random variables $X$ and $Y$ on $\Omega$, $E$ satisfies

\begin{itemize}
  \item[(i)] $E(X) \geq 0$ if $X \geq 0$,
  \item[(ii)] $E(aX + bY) = aE(X) + bE(Y) \forall a, b \in \mathbb{R}$,
  \item[(iii)] $E(X) = 1$ if $X \equiv 1$.
\end{itemize}

\textbf{B.9. $L^p$ spaces}

• Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $1 \leq p < \infty$. Define $L^p = L^p(\mathbb{P}) = L^p(\Omega, \mathcal{F}, \mathbb{P})$ to be the set of all random variables $X$ on $\Omega$ such that

$$\int_{\Omega} |X|^p \, d\mathbb{P} < \infty.$$ 

For any such $X$, define the \textit{$p$-norm} of $X$ by

$$\|X\|_p = \left( \int_{\Omega} |X|^p \, d\mathbb{P} \right)^{1/p}.$$ 

• If $X, Y \in L^1$ are independent random variables, then $XY \in L^1$ and

$$E(XY) = E(X)E(Y).$$

\textbf{B.10. Expectation and conditional expectation for discrete random variables}

• Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable $X$ on $\Omega$ is a \textit{discrete random variable} if it takes a countable set of values.
• The (probability) mass function of a discrete random variable $X$ on $\Omega$ is the function $f_X : \mathbb{R} \to [0, 1]$ given by
  \[ f_X(x) = \mathbb{P}(X = x). \]
  Thus $f_X$ is a density function for $X$.

• Thus the expectation of a discrete random variable $X$ on $\Omega$ with mass function $f_X$ is
  \[ E(X) = \sum_{\{x : f_X(x) > 0\}} xf_X(x) \]
  whenever this sum is absolutely convergent.

• Let $X, Y$ be discrete random variables on $\Omega$. The conditional (probability) mass function of $Y$ given $X = x$ is defined by
  \[ f_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x) \]
  whenever $\mathbb{P}(X = x) > 0$.

More generally, if $X_1, \ldots, X_n, Y$ are discrete random variables on $\Omega$, the conditional (probability) mass function of $Y$ given $X_i = x_i, 1 \leq i \leq n$, is defined by
  \[ f_{Y|X_1,\ldots,X_n}(y|x_1,\ldots,x_n) = \mathbb{P}(Y = y|X_1 = x_1, \ldots, X_n = x_n) \]
  whenever $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) > 0$.

• The conditional expectation of $Y$ given $X_i = x_i, 1 \leq i \leq n$ is the random variable $Z : \Omega \to \mathbb{R}$ given by
  \[ Z(\omega) = E(Y|X_1 = x_1, \ldots, X_n = x_n) = \sum_y y f_{Y|X_1,\ldots,X_n}(y|x_1,\ldots,x_n), \]
  if $X_i(\omega) = x_i$ and the sum on the r.h.s. is absolutely convergent. We often write $Z(X_1, \ldots, X_n) = E(Y|X_1, \ldots, X_n)$.

B.11. Topological spaces and compact sets

• A topology $T$ on a set $\Omega$ is a family of subsets of $\Omega$ such that
  
  (i) $\emptyset, \Omega \in T,$
  (ii) $U_1, \ldots, U_n \in T \Rightarrow \bigcap_{i=1}^n U_i \in T,$
  (iii) $\{U_{\alpha} : \alpha \in A\} \subset T \Rightarrow \bigcup_{\alpha \in A} U_{\alpha} \in T$ for all index sets $A$. 
• As above, a pair \((\Omega, T)\) is a topological space. A set \(U \in T\) is an open set and its complement in \(\Omega\), \(\Omega \setminus U\) is a closed set.

• A subset \(E\) of a topological space \((\Omega, T)\) is a compact set if every family of open sets of \(T\) which covers \(E\) (ie: whose union contains \(E\)), has a finite sub-family which covers \(E\).

• For sets \(A\) and \(B\), \(A^B\) is the space of all functions from \(B\) to \(A\). Equivalently, \(A^B\) is a product space with index set \(B\). If \(A\) is compact (relative to some topological space), then \(A^B\) is a product of compact spaces, (and hence compact; cf: Tychonov’s theorem).

• In particular, if \(X = \{n\}\), the space \([0, 1]^{P(X)}\) is precisely the space of all \(2^n\)-vectors with coordinates in \([0, 1]\), ie: the compact product space \([0, 1]^{2^n}\).

B.12. Landau’s \(O, o, \sim\) notation

• Let \(f(x)\) and \(g(x)\) be real functions.

  (i) \(f(x) = O(g(x))\) as \(x \to \infty\) if \(\exists K, x_0\) such that \(|f(x)| < K|g(x)|\ \forall x > x_0\).

  (ii) \(f(x) = o(g(x))\) as \(x \to \infty\) if \(f(x)/g(x) \to 0\) as \(x \to \infty\). In particular, \(f(x) = o(1)\) as \(x \to \infty\) if \(f(x) \to 0\) as \(x \to \infty\).

  (iii) \(f(x) \sim lg(x) (l \neq 0)\) as \(x \to \infty\) if \(f(x)/g(x) \to l\) as \(x \to \infty\). We say that \(f(x)\) is asymptotic to \(lg(x)\).

B.13. Stirling’s formula

• A weak form of Stirling’s formula is

\[ n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \]

A sharper form is

\[ \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq e^{1/12n} \sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \]

In most cases, it suffices to consider \(n! \geq 2\sqrt{n(n/e)^n} \geq (n/e)^n\). Thus

\[ \binom{n}{k} \leq \left(\frac{n^k}{k!}\right) \leq \frac{1}{2\sqrt{k}} \left(\frac{en}{k}\right)^k \leq \left(\frac{en}{k}\right)^k. \]
References

[1] Alon, N. and Milman, V.D., $\lambda_1$, isoperimetric inequalities for graphs, and super-
concentrators, *J. Combinatorial Theory* (B) **38** (1985), 73-88.


(1967), 357-367.

xii + 593 pp.


+ 177pp.


graphs, in *Random Graphs ’87* (Karoński, M., Jaworski, J., and Ruciński, A., eds.),


81-97.

[12] Bollobás, B., *Modern Graph Theory*, Graduate Texts in Mathematics 184,

Soc.*, **80** (1976), 419-427.


[27] Shamir, E. and Spencer, J., Sharp concentration of the chromatic number of random graphs $G_{n,p}$, Combinatorica 7 (1987), 121-129.