THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF MATHEMATICS

Approximate Ramsey Theory

XINGYU ZHANG
SPRING 2015

A thesis
submitted in partial fulfillment
of the requirements
for the baccalaureate degree
in Mathematics
with honors in Mathematics

Reviewed and approved* by the following:

Jan Reimann
Assistant Professor of Mathematics
Thesis Supervisor

Victoria Sadovskaya
Associate Professor of Mathematics
Honors Adviser

*Signatures are on file in the Schreyer Honors College.

Abstract

Ramsey theory is concerned with the question: Given two graphs G_1 and G_2, how large does n have to be so that any two coloring of K_n, the complete graph on n vertices, with red and blue either contains a red G_1 or a blue G_2 subgraph? This number is denoted as the Ramsey number for G_1 and G_2, notationally written as $R(G_1, G_2)$. Any graph can be represented as an adjacency matrix. Using matrix norms such as the edit norm and the cut norm, we can define a meaningful notion of distance between two graphs. We investigate Ramsey numbers for subgraphs which are a certain distance away from a complete graph using the edit norm and the cut norm.

Table of Contents

Acknowledgements

I would like to thank my thesis supervisor, Dr. Jan Reimann, for his help and support as well as my honors adviser, Dr. Victoria Sadovskaya, for reviewing this thesis.

Chapter 1

Graph Distances

1.1 Introduction to Graphs

For sections 1.1, 1.2, 1.3, the proofs and ideas are largely from Lovasz's book *Large Networks and Graph Limits* [1]. I also added a few examples and observations to the ideas in the book.

To begin, we consider several possible ways of measuring the distance between two graphs. We want to define distance because we would like a good way to measure how similar two graphs are to each other.

First, what is a graph? A graph G is an ordered pair of vertices and edges, and edges are 2-element subsets of vertices. Typically, the set of vertices is denoted as V and the set of edges is denoted as E. There are generalizations of graphs such as directed graphs where an edge is an *ordered* pair of vertices, or hypergraphs where the edges are any arbitrary subsets of vertices (so they could have more than 2 elements). However, we will not deal with these generalizations in this paper; we will focus on the standard definition of a graph.

Generally, as notation, we denote the edge connecting v_i and v_j by $v_i v_j$.

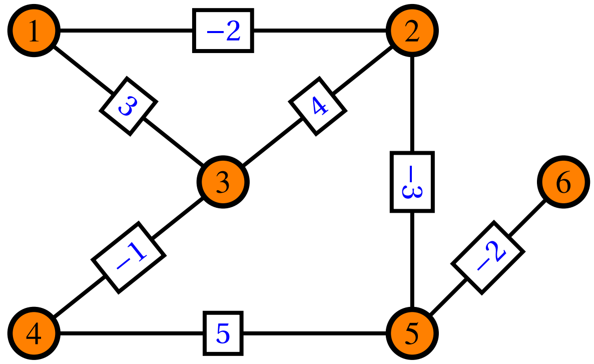
There is one type of generalization we will consider however, and this is the concept of weighted graph. Weighted graphs are graphs where each edge is assigned a specific numeric value. Typically, a weighted graph refers to a graph where numeric values are assigned to the edges, but sometimes the vertices of a graph could be assigned weights, or even both the vertices and the edges could be assigned weights.

To start off, we note that there is a way to represent a graph in the form of a matrix. Let $G = (V, E)$ be a weighted graph with weights $c : E \to \mathbb{R}$ with the vertex set $V = \{v_1, v_2, \ldots, v_n\}$. Then the weighted adjacency matrix (sometimes known as the weighted incidence matrix) A is defined as follows:

$$A_{ij} = \begin{cases} c(v_i v_j), & v_i v_j \in E \\ 0, & \text{otherwise} \end{cases}$$

When $A_{ii} \neq 0$, there is a loop at vertex v_i, which is an edge that connects a vertex to itself. We can represent an unweighted graph as a weighted graph by putting $c(v_i v_j) = 1$, $\forall v_i v_j \in E$.

To make some of these concepts more concrete, here is an example of a weighted graph and its corresponding weighted adjacency matrix:



$$A = \begin{pmatrix} 0 & -2 & 3 & 0 & 0 & 0 \\ -2 & 0 & 4 & 0 & -3 & 0 \\ 3 & 4 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 5 & 0 \\ 0 & -3 & 0 & 5 & 0 & -2 \\ 0 & 0 & 0 & 0 & -2 & 0 \end{pmatrix}$$

Notice that adjacency matrices for undirected weighted graphs are always symmetric. This is because if v_i is adjacent to v_j then v_j is adjacent to v_i.

1.2 Norms of Matrices

Now that we have defined what a graph is, we want to establish a useful way of comparing two graphs. Given two graphs, we want to be able to assign a meaningful number to measure how closely they resemble each other. This leads us to consider graph distances.

Since we can represent a graph as a matrix, we can assign a norm to a graph simply by assigning a norm to the corresponding adjacency matrix. There are several ways to do this.

A very simple norm is the l_1 norm which is defined as:

$$||A||_1 = \frac{1}{n^2} \sum_{i,j=1}^{n} |A_{ij}|$$

This norm simply adds up the absolute values of all entries in the matrix and divides the sum by the number of entries (we want to divide by n^2 because we don't want to give too much preference to larger graphs). Therefore, $||A||_1$ is the average of absolute values of the entries. Additional norms include l_2:

$$||A||_2 = \sqrt{\frac{1}{n^2} \sum_{i,j=1}^{n} A_{ij}^2}$$

which adds the squares of all entries of the matrix, divides it by the size of the matrix and then takes the square root. This is the quadratic mean of the entries. Finally, l_∞ norm is

$$||A||_\infty = \max_{i,j=1} |A_{ij}|$$

which picks the maximum absolute value of an entry of the matrix.

The norm that we will be concerned with the most is the cut norm. Here $[n]$ denotes the vertex set:

$$||A||_\square = \frac{1}{n^2} \max_{S,T \subseteq [n]} |\sum_{i \in S, j \in T} A_{ij}|$$

One observation is that if we take the adjacency matrix of an unweighted, simple (meaning no loops) graph, then the cut norm simply gives us $\frac{2|E|}{n^2}$ by taking $S = [n]$ and $T = [n]$. Also, we notice that for an unweighted, simple graph, the l_1 norm similarly gives us $\frac{2|E|}{n^2}$ because each edge contributes 2 to the matrix. Similarly, for a weighted graph with all positive weights, the cut norm will be $\frac{1}{n^2}\sum_{ij} c(v_i v_j)$ and this is the same as the l_1 norm.

How do we understand the cut norm intuitively? Consider a weighted graph G with both positive and negative weights and let its vertex set be $V = \{v_1, v_2, \ldots, v_n\}$. Now take a copy of G, call it G' with vertices $V' = \{w_1, w_2, \ldots, w_n\}$. Define a graph H on the vertex set $V \cup V'$ by defining the edges of H such that it contains the edges in G and G' and includes $v_i w_j$ if $v_i v_j$ is an edge in G. Choose a subgraph of G and a subgraph of G' and let them be denoted as H and H' respectively. Now consider the sum of the weights of the edges between H and H', let this sum of weights be denoted as W. The cut norm is the maximum W by choosing appropriate H and H'.

In the example before, $||A||_1 = \frac{40}{6^2} = \frac{10}{9}$, $||A||_2 = \sqrt{\frac{136}{6^2}} = \frac{\sqrt{34}}{3}$, $||A||_\infty = 5$, and $||A||_\square = \frac{7}{18}$.

The l_1 norm bounds the cut norm: $||A||_\square \leq ||A||_1$. This is true because by the triangle inequality,

$$\max_{S,T \subseteq [n]} |\sum_{i \in S, j \in T} A_{ij}| \leq \max_{S,T \subseteq [n]} \sum_{i \in S, j \in T} |A_{ij}| \leq \sum_{i,j}^{n} |A_{ij}|$$

In addition, we have the relation: $||A||_1 \leq ||A||_2$. This is just showing that the quadratic mean is greater than or equal to the arithmetic mean. In particular, we have to show that for any numbers $A_1, \ldots, A_N$,

$$(\frac{1}{N}\sum_{i=1}^{N} A_i^2)^{1/2} \geq \frac{1}{N}\sum_{i=1}^{N} |A_i|$$

$$\Leftrightarrow (\sum_{i=1}^{N} |A_i|)^2 \leq N \sum_{i=1}^{N} A_i^2$$

This follows from the Cauchy-Schwarz inequality. Recall that the inequality states:

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right)$$

Using the inequality where $x_i = A_i$ and $y_i = 1$, we get

$$(A_1^2 + A_2^2 + \ldots + A_N^2)N \geq (A_1 + A_2 + \ldots A_N)^2$$
$$\Rightarrow \left(\frac{A_1 + A_2 + \ldots + A_N}{N}\right)^2 \leq \frac{A_1^2 + \ldots A_N^2}{N}$$

Finally, we also have that $||A||_2 \leq ||A||_\infty$. Assume WLOG that $|A_{11}| = \max_{i,j}\{A_{ij}\}$. Then this identity follows immediately since

$$\sum_{i,j=1} A_{i,j}^2 \leq \sum_{i,j=1}^{n} |A_{11}^2| = n|A_{11}|^2$$

Thus, $\sqrt{\frac{1}{n^2}\sum_{i,j=1}^{n} A_{ij}^2} \leq \max_{i,j=1}|A_{ij}|$.

Therefore, we proved that $||A||_\Box \leq ||A||_1 \leq ||A||_2 \leq ||A||_\infty$.

Here is a more complicated example. Suppose A is an $n \times n$ matrix with entries that are independent random ± 1 with probability $\frac{1}{2}$. Then it is clear that $||A||_1 = ||A||_2 = ||A||_\infty = 1$. Now interpreting A_{ij} to be a random variable where the i, j entry of A assumes 1 or -1 with probability $\frac{1}{2}$ each, then the expectation of A_{ij} is 0. Let X denote the random variable $\sum_{i\in S, j\in T} A_{ij}$. Then, $E(X) = 0$. Also, $Var(A_{ij}) = E(A_{ij}^2) - (E(A_{ij})^2 = E(A_{ij}^2) = 1$. In addition, because the random variables A_{ij} are independent, the variance of the sum of these random variables is the sum of the variances. Thus, $Var(X) = Var(\sum_{i\in S, j\in T} A_{ij}) = \sum_{i\in S, j\in T} Var(A_{ij}) = \sum_{i\in S, j\in T} 1 = O(n^2)$. Then, since $Var(X) = E(X^2) - E(X)^2 \Rightarrow O(n^2) = E(X^2)$, $E(|X|) = O(n)$. But what is an upper bound for the cut norm?

We use Hoeffding's Inequality ([2]):

Theorem: Let $X_1, \ldots, X_n$ be independent random variables and $P(X_i \in [a_i, b_i]) = 1$, $1 \leq i \leq n$. Let $\overline{X} = \frac{1}{n}(X_1 + \ldots + X_n)$. Then

$$P(|\overline{X} - E(\overline{X})| \geq t) \leq 2\exp(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2})$$

In our example, A_{ij} are all independent random variables and $a_i = -1$, $b_i = 1$, $1 \leq i \leq n$. In addition, let $t = 4n^{-\frac{1}{2}}$. Then,

$$\begin{aligned} P(||A||_\Box \geq 4n^{-\frac{1}{2}}) &\leq 2\exp(-\frac{2n^2(4n^{-1/2})^2}{2^2 n}) \\ &= 2\exp(-\frac{32n}{4n}) \\ &= 2e^{-8} \end{aligned}$$

Thus, with very high probability, $||A||_\Box < 4n^{-\frac{1}{2}}$.

1.3 Two Graphs on Same Set of Nodes

Now that we have a working definition of different matrix norms, we can begin to describe the distance between two graphs. For the most part, we are concerned with describing the distance between two graphs on the same vertex (node) set.

Suppose G and G′ be two unweighted graphs on the same vertex set $[n]$. By taking any of the norms in section 1.2, the norm of the difference of their adjacency matrices defines a distance between the graphs. This would be a matrix consisting of only 1's, 0's, and −1's. Suppose we are subtracting the adjacency matrix of G′ from that of G. Let the resulting matrix be A. Now $A_{ij} = 1$ if $v_i v_j$ is an edge of G but not an edge of G′, $A_{ij} = 0$ if both G and G′ have an edge $v_i v_j$ or if both don't, and $A_{ij} = -1$ if $v_i v_j$ is not an edge of G but is an edge of G′.

Let $S, T \subseteq V$. We define $e_G(S,T)$ to be the number of edges with one endnode in S and the other endnode in T. Then, the cut distance between graphs G and G′ is defined as:

$$d_\square(G,G') = ||A_G - A'_G||_\square = \max_{S,T\subseteq V} \frac{|e_G(S,T) - e'_G(S,T)|}{n^2}$$

Notice that $||A_G||_\square = d_\square(G,\emptyset) = \frac{1}{n^2}\max_{S,T\subseteq V} e_G(S,T)$, that is by taking G′ to be the empty graph on n vertices.

Consider the example we had before: Suppose G and G′ are two independent random graphs on $[n]$ with edge probability $1/2$. Let A_G and A'_G be the adjacency matrices for G and G′ respectively and let $B := A_G - A'_G$. Then B_{ij} is equal to 1 with probability $1/4$, -1 with probability $1/4$, and 0 with probability $1/2$. Then $d_\square(G,G') = ||B||_\square$. Again, using Hoeffding's Inequality,

$$P(||B||_\square \geq Cn^{-1/2}) \leq 2\exp(-\frac{2n^2(Cn^{-1/2})^2}{4n}) = 2e^{-C/2}$$

Hence, given large C, we can conclude that $d_\square(G,G') \leq Cn^{-1/2}$ with large probability. It's curious to note that $P(||B||_\square \geq Cn^{-1/2}) \leq 2e^{-C/2}$ no matter what n is. Therefore, as n tends to infinity, $d_\square(G,G')$ tends to 0. Hence, random graphs will look more and more "alike."

Weights could be assigned to vertices as well as edges. Let the vertex set be $\{v_1,\ldots,v_n\}$. Let $\alpha_i(G)$ denote the nodeweight of vertex v_i in graph G and $\alpha_G := \sum_{i=1}^n \alpha_i(G)$.

We can extend our definition of cut distance to weighted graphs as well. Suppose G and G′ are still defined on the same vertex set, but G and G′ are weighted graphs with possibly different nodeweights and edgeweights. Let β_{ij} denote the edgeweight of edge $v_i v_j$ in graph G. In addition, let $\alpha_i = \frac{\alpha_i(G)}{\alpha_G}$ denote the relative nodeweight of vertex v_i with respect to the total nodeweight. Let α'_i and β'_{ij} denote the respective values for graph G′.

We define the cut distance of these weighted graphs to be:

$$d_\square(G,G') = \frac{\sum_{i\in V}|\alpha_i - \alpha'_i| + \max_{S,T\subseteq V}|\sum_{i\in S, j\in T}(\alpha_i\alpha_j\beta_{ij} - \alpha'_i\alpha'_j\beta'_{ij})|}{n^2}$$

In the special case when $\alpha_i = 1$ and $\alpha'_i = 1$ for all $1 \le i \le n$, this reduces to:

$$d_\square(G,G') = \frac{\max_{S,T\subseteq V}|\sum_{i\in S, j\in T}(\beta_{ij} - \beta'_{ij})|}{n^2}$$

Now, this is precisely $||A_G - A'_G||_\square$ where A_G and A'_G denote the weighted adjacency matrices for G and G′. In the even more specific case when the edgeweights are 0 or 1, this definition reduces to the usual cut distance for unweighted graphs.

1.3.1 Edit Distance

We will introduce another very simple way to measure the distance between two graphs: The Edit Distance. Suppose we have two graphs G and H on the same node set. Intuitively, the edit distance measures the number of edges that must be added or deleted from G in order to get the graph H. In addition, we want to scale the distance by the size of the graphs. The formal definition of the edit distance between graphs G and H is

$$d_1(G,H) = \frac{|E(G)\triangle E(H)|}{n^2}$$

Here, $\triangle$ represents the symmetric difference between E(G) and E(H).

For weighted graphs, the edit distance can be generalized accordingly. The edit distance is simply the sum of the absolute values of the difference of weights between the graphs. Suppose the edge weights for G are a_{ij} and the edge weights for H are b_{ij}. Then the edit distance is defined as

$$d_1(G,H) = \frac{\sum_{i,j\in V}|a_{ij} - b_{ij}|}{n^2}$$

It's easy to see that the edit distance is simply $d_1(G,H) = ||A_G - A_H||_1$. We have already proved from before that $||A||_\square \le ||A||_1$, thus it follows easily that the cut distance between any two graphs can never exceed the edit distance.

1.4 Computing the Cut Norm

Given a matrix, is there an easy way to determine its cut norm? Since the cut norm involves testing all subsets of a given set, a “brute force” approach would have a large running time. Is there a fast algorithm by which we can find a matrix's cut norm?

Now we are going to take a slight detour to analyze the task of algorithmically computing the cut norm of a matrix, in particular the computational complexity of the problem. We are going to refer to the problem of computing the cut norm of a matrix as CUT-NORM. We are considered specifically with the decision problem: Given a constant $k \in \mathbb{R}$, is the cut norm greater than k? If so, the algorithm should return "yes" and if not, then "no." Consider first, the brute force approach for computing the cut norm, written using pseudocode:

```
maxvalue=0
for any two subsets S,T of [n],
        newvalue=cut norm of S,T
        maxvalue=max{maxvalue, newvalue}
end
if maxvalue > k
        return YES
else
        return NO
```

The problem with this approach is that we must run the for loop an exponential number of times. Given that there are 2^n subsets of an n element base set, we must run the loop $2^n * 2^n = 4^n$ times to cover all of the cases. Thus the running time for this algorithm is $O(4^n)$.

Obviously this is unfortunate. Can we improve this algorithm at all?

One observation to make is that if any row or column is all positive, then we will definitely include it in our solution and if any row or column is all negative, then we will definitely leave it out in our solution. So for some matrices, we can reduce the number of subsets to check by removing these obvious cases.

I will now briefly introduce the concepts of NP-Completeness and NP-Hardness. If a problem can be solved in polynomial time, we say that the problem is in the complexity class P. There is a further class of problems which are classified as NP-complete, and this means that these problems are currently not shown to belong in P but these problems are reducible to each other. What this means is that if you can solve one of these problems, then you can effectively solve the rest. The most famous problem in theoretical computer science is the P = NP problem. If we can show that one of these NP-complete problems has a polynomial time algorithm, then they all have polynomial time algorithms. Most scholars in this field believe that P ≠ NP, hence they believe that there is no fast algorithm to solve NP-complete problems. A problem is NP-Hard if every problem in NP can be reduced to that problem.

I will now introduce a graph theoretic interpretation of CUT-NORM. Suppose we're given an $n \times n$ matrix A. Let G_1 and G_2 be two independent sets of n vertices with vertex sets $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ respectively. We will construct a weighted graph G as follows: if $a_{ij} \neq 0$, then let $v_i w_j$ be an edge in G with weight a_{ij}. The graph problem of CUT-NORM is finding subsets $S \subseteq G_1$ and $T \subseteq G_2$ such that the sum of the weights between S and T is maximized.

Proposition 1: The graph problem of CUT-NORM is equivalent to the matrix problem of CUT-NORM.

Proof. Suppose I solved the graph problem of CUT-NORM, i.e, I found subsets $S \subseteq G_1$ and $T \subseteq G_2$ such that the sum of the weights of the edges between S and T are maximized; let this value be k. Then k is also our matrix CUT-NORM. If not, then there exists $S', T' \subseteq [n]$ such that $\max_{S',T' \subseteq [n]} |\sum_{i \in S', j \in T'} A_{ij}| > k$. But then $S' \subseteq G_1$ and $T'' \subseteq G_2$ would have a greater sum of weights between them, thus, k was not the maximum, a contradiction.

Similarly, if k is our matrix CUT-NORM satisfied with sets $S, T \subseteq [n]$, then k is also our graph CUT-NORM. If not, then there exists subsets $S' \subseteq G_1$ and $T' \subseteq G_2$ such that the sum of the weights of the edges from S' to T' is greater than k. However, this means that taking the corresponding S' and T' of the rows and columns of our matrix would result in a greater CUT-NORM than k, a contradiction. □

Now the graph problem of CUT-NORM is similar to finding the maximum cut of a graph except the cut doesn't have to partition the entire vertex set of G, but it only has to partition a subset of the vertex set of G. Thus, intuitively, the CUT-NORM problem appears to be harder than the MAX-CUT problem. We know from existing literature that the MAX-CUT problem is NP-Complete, hence if we can find a reduction of MAX-CUT to CUT-NORM, then we will prove that CUT-NORM is NP-Hard.

Proposition 2: CUT-NORM is NP-Hard ([3])

Proof. Here is a way to reduce MAX-CUT to CUT-NORM. Suppose you are given a weighted (with positive weights), undirected graph $G = (V, E)$ where $|V| = n$ and $|E| = m$ and you want to find the weighted maximum cut of this graph. Let A be the weighted adjacency matrix of G. We will construct a new matrix B with size $2m \times n$ such that the cut norm of B is exactly the maximum cut of G.

Let $\{e_1, e_2, \dots, e_m\} = E$ and let $w_1, \dots, w_m$ be the corresponding weights for the edges. Suppose edge e_i connects vertices v_j and v_k. Consider the two rows $2i-1$ and $2i$. Pick one of the two rows arbitrarily and let the j^{th} entry of that row be w_i and the k^{th} entry be $-w_i$ and for the other row, let the j^{th} entry be $-w_i$ and the k^{th} entry be w_i. Let the rest of the entries of the two rows be 0. The result is our matrix B.

Fix a subset S of the vertex set (this is the same as picking a subset of the columns of B). Now, suppose we want to find a subset T of the rows of B such that $|\sum_{i \in T, j \in S} A_{ij}| := W$ is maximized. Which rows of B will T contain?

Some observations:

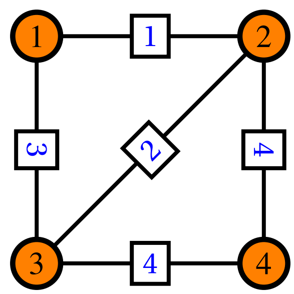
1. If e_i is an edge connecting vertices within S, then the entries of the rows $2i-1$ and $2i$ sum up to 0. Hence, whether or not T contains these rows is irrelevant.
2. If e_i is an edge from S to $\overline{S}$, then only one of the rows $2i-1$ and $2i$ is in T, which would contribute w_i to W (the other row would add the negative of the weight of the edge, which we do not want).

Hence, the maximal subset of the rows of B which maximizes W, denoted T, would contain one row from rows $\{2i-1, 2i\}$, where e_i is not an edge within S, and this would contribute w_i to W. Thus, W is precisely the sum of the weights of the edges from S to $\overline{\text{S}}$.

Thus, it follows that the cut norm of matrix B gives the maximum value of the cut of graph G.

Notice that the construction of matrix B takes a polynomial number of steps since B has size $2mn \leq (n-1)n^2 = \text{O}(n^2)$. Thus, this is a polynomial time reduction of MAX-CUT to CUT-NORM, and hence, CUT-NORM is NP-Hard. □

Here is an example of the construction of the matrix B in the proof before. Suppose we are given the graph:



Then the corresponding adjacency matrix is:

$$\begin{pmatrix} 0 & 1 & 3 & 0 \\ 1 & 0 & 2 & 4 \\ 3 & 2 & 0 & 4 \\ 0 & 4 & 4 & 0 \end{pmatrix}.$$

The matrix B is then:

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -3 & 0 & 3 & 0 \\ 3 & 0 & -3 & 0 \\ 0 & 0 & 4 & -4 \\ 0 & 0 & -4 & 4 \\ 0 & 4 & 0 & -4 \\ 0 & -4 & 0 & 4 \\ 0 & 2 & -2 & 0 \\ 0 & -2 & 2 & 0 \end{pmatrix}.$$

With the proof that CUT-NORM is NP-Hard, we know that there cannot exist a fast algorithm to compute the cut norm for an arbitrary matrix. This makes it very difficult to find the actual cut norm for most matrices.

Chapter 2

Ramsey Theory

2.1 Background on Ramsey Theory

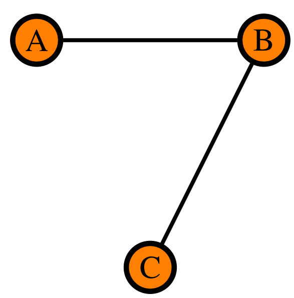
We are going to turn our attention to the second part of the paper now, which deals with Ramsey Theory. A reference for this section is *Graph Theory* by Diestel [4].

A complete graph on n vertices is a graph where any two vertices is connected by an edge and this graph is denoted as K_n. The graph K_n has $\binom{n}{2} = \frac{n(n-1)}{2}$ edges.

Suppose you are at a party consisting of N people. As a mathematician, you ask yourself a question: Is there a group of k people here at this party who all know each other or a group of k people who all don't know each other? Answering this question is essentially the study of Ramsey Theory.

We currently know basic facts such as the fact that in a party of 6 people, there are either 3 people who mutually know each other or 3 people who mutually don't know each other, but this fact isn't true in a party of 5 people. In a party of 18 people, there is either a group of 4 who know each other or a group of 4 who do not know each other.

We can ask this question within graph theory by associating people to vertices and "knowing each other" to edges. For example, if Adam knows Bob and Bob knows Charlie, but Adam doesn't know Charlie, the corresponding graph is (where A stands for Adam, B stands for Bob, and C stands for Charlie):



The problem can also be represented in an equivalent way in terms of graph edge coloring, which will be the representation we will primarily use. An edge coloring of a graph is an assignment of colors to the edges of a graph. For the purposes here, we only need two colors, say red and blue. If an edge is colored red, we assign it the meaning that the two people associated with the edge know each other, and if an edge is colored blue, then the two people do not know each other.

Thus, using this representation, the fact that in a party of 6 people, either 3 people know each other or 3 people do not, is equivalent to the statement that any edge coloring of K_6 with 2 colors will either contain a red K_3 or a blue K_3.

We denote by $R(m,n)$ the smallest integer such that a 2-edge coloring of $K_{R(m,n)}$ will either contain a red K_m or a blue K_n. $R(m,n)$ is known as a Ramsey number.

A natural question is whether $\mathrm{R}(m,n)$ exists. Note that it is sufficient to prove that $\mathrm{R}(m,m)$ exists for all m. This is because if we let $m \geq n$, then if we are guaranteed to either have a red K_m or a blue K_m, then we also would either have a red K_m or a blue K_n.

Theorem: $\mathrm{R}(m,m)$ exists for all m ([5]).

Proof. Let $G = K_n$ be our graph and suppose it has n vertices, where $n \geq 2^{2m}$. Suppose further that a random 2-coloring of the edges of G is given. Pick a vertex at random and call it v_1. Consider the colors of the edges which contain v_1 and let c_1 be the more frequent color (if the colors are equally frequent, choose one arbitrarily). Let V_1 be the set of vertices which is incident to v_1 with an edge of color c_1.

Now pick a vertex at random within V_1 and call it v_2. Then consider the colors of the edges which contain v_2 within V_1 and let c_2 denote the more frequent color. Let V_2 be the subset of vertices in V_1 which is incident to v_2 by an edge of color c_2.

Continuing this process until we run out of vertices, we get a sequence of vertices $V' := \{v_1, v_2, \ldots, v_p\}$ with colors $c_1, c_2, \ldots, c_p$ for some p. Either red or blue appears more often in the sequence $c_1, \ldots, c_p$, assume red. Then the subset of vertices in V' which correspond to the red form a red clique of size at least $p/2$.

At each step of the process, we are removing at most half of the vertices. Since $n \geq 2^{2m}$, we are guaranteed to either have a red or blue clique of size $\frac{2m}{2} = m$. This proves that $\mathrm{R}(m,m)$ exists for all m and is at most 2^{2m}. □

2.2 Important Theorems in Ramsey Theory

We mentioned in the previous section that in a party of 6, either 3 people know each other or 3 people don't. The proof is similar to the existence proof of Ramsey numbers. However, before we begin, we must define one more graph:

Definition: A path is a graph $P = (V, E)$ where $V = \{x_1, x_2, \ldots, x_k\}$, $E = \{x_1x_2, x_2x_3, \ldots, x_{k-1}x_k\}$ and the x_i are all distinct. A cycle is a path with the additional edge x_kx_1. A graph which consists of only one cycle with n vertices is denoted as C_n.

Theorem: $\mathrm{R}(3,3) = 6$.

Proof. First we will prove that $\mathrm{R}(3,3) \leq 6$. 2-color K_6 with colors red and blue. Pick a vertex v_1 from K_6. Of the 5 edges incident to v_1, there are at least 3 that are red or 3 that are blue. Assume without loss of generality that it is red and (v_1, v_2), (v_1, v_3), (v_1, v_4) are red edges. Now the edges among $\{v_2, v_3, v_4\}$ must all be blue, otherwise we can form a red K_3. But if they were all blue, then $\{v_2, v_3, v_4\}$ form a blue K_3.

To prove that $R(3,3) \geq 6$, all we need to do is find a 2-coloring of K_5 such that there is no red K_3 or blue K_3. Simply color a C_5 subgraph of K_5 with red and color the remaining edges with blue. The result has no red K_3 and no blue K_3. □

There is also a simple recursive relation involving Ramsey theory which is easy to prove.

Theorem: $R(m,n) \leq R(m-1,n) + R(m,n-1)$.

Proof. We start with a complete graph of size $R(m-1,n) + R(m,n-1)$ and suppose a 2-coloring of its edges with red and blue is given. Pick a vertex v at random. Let V_1 be the set of vertices which are incident to v by a red edge and let V_2 be the set of vertices which are incident to v by a blue edge. Now either $|V_1| \geq R(m-1,n)$ or $|V_2| \geq R(m,n-1)$ by the pigeonhole principle. Assume $|V_1| \geq R(m-1,n)$. If V_1 contains a blue K_n subgraph, then we're done. If V_1 contains a red K_{m-1} subgraph, then this subgraph together with v is a red K_m subgraph, and we're done. The same argument holds when $|V_2| \geq R(m,n-1)$. □

The Ramsey question can naturally be generalized to include three or more colors and to hypergraphs. However, we won't be considering these generalizations here. We are concerned, however, with one generalization, which is when the subgraphs are not necessarily complete graphs. For example, we can ask what is $R(G,H)$, where G and H are given subgraphs? Thus the question is, what is the smallest complete graph such that any 2-coloring gives us a red G or a blue H? To illustrate our next idea, we need one more definition.

Definition: A **tree** is an undirected simple (no loops) graph that is connected and contains no cycles.

Some work has already been done on this problem. In fact, the Ramsey number of a tree and a complete graph is known exactly. The proof is not difficult, but we first have to introduce some lemmas.

Definition: Let C be a set of colors. A vertex coloring of a graph $G = (V,E)$ is a map $c : V \to C$ where $c(v) \neq c(w)$ whenever v and w are adjacent (connected by an edge).

Basically, a vertex coloring means that we assign colors to the vertex set of a graph such that adjacent vertices are colored with different colors. The minimum number of colors needed to color vertex color a graph G is denoted as the chromatic number and is written notationally as $\chi(G)$.

Definition: The degree of a vertex v is the number of edges which are incident to v. The minimum degree of a graph is the minimum degree of the vertices of the graph.

Here are some basic lemmas in graph theory which we will not prove:

Lemma: Every graph G has a subgraph of minimum degree at least $\chi(G) - 1$.

Lemma: If T is a tree and G is a graph with minimum degree at least $|T|-1$, then T is a subgraph of G.

Theorem (Chvatal): If s and t are positive integers and T is a tree of order t, then $R(T,K_s)=(s-1)(t-1)+1$ (source: [4]).

Proof. First we will prove that $R(T,K_s) \geq (s-1)(t-1)+1$ by finding a 2-coloring of a complete graph of $(s-1)(t-1)$ vertices that doesn't contain T or K_s. Take $s-1$ copies of K_{t-1} and color this red. Notice that there cannot be a red tree with order t since there is no set of t vertices that are connected to each other by red edges. Now color the remaining edges blue. There cannot be a blue K_s here. Hence, $R(T,K_s)>(s-1)(t-1)$.

Now we will prove that $R(T,K_s) \leq (s-1)(t-1)+1$. In order to prove this, I can prove that if the complement of a graph G of order $(s-1)(t-1)+1$ doesn't contain a K_s, then G doesn't contain a tree of order t (this follows from the alternate formulation of the Ramsey problem explained before). So let G be such a graph or order $(s-1)(t-1)+1$ whose complement doesn't contain a K_s. This implies that in a vertex coloring of G, at most $s-1$ vertices can have the same color (if s vertices have the same color, then the complement would contain a K_s). Thus, the graph contains at least $\frac{(s-1)(t-1)+1}{s-1}$ colors so the chromatic number for G must be at least t. Using the first lemma, G must have a subgraph H with a minimum degree of at least $t-1$. By the second lemma, G must contain a tree of order t. Hence, if the complement of G does not contain K_s, then G must contain a tree of order t. The proof is done. □

There is another theorem of similar nature concerning cycles and K_4 given in a paper by Sheng, Ru, and Min [6]:

Lemma: $R(C_n,K_3)=2(n-1)+1$ (source: [7]).

Theorem: $R(C_n,K_4)=3(n-1)+1$ for $n \geq 4$.

Proof. First we will prove that $R(C_n,K_4)>3(n-1)$. This is evident by taking three copies of K_{n-1}. There cannot be a cycle of size n and the complement cannot contain a K_4.

Now we will prove that $R(C_n,K_4) \leq 3(n-1)+1$. It's been proven in other papers that this result is true for $n=4$ and $n=5$ so we can assume that $n>5$. We will prove this by induction. Since the base case has already been proven, we will assume the induction hypothesis that $R(C_{n-1},K_4)=3(n-2)+1$. To prove this, we assume that G is a graph of order $3(n-1)+1$ that does not contain a C_n nor does its complement contain a K_4 (in other words, it does not contain a 4-element independent set).

Using the fact that $R(C_n,K_3)=2(n-1)+1$ and our induction hypothesis that $R(C_{n-1},K_4)=3(n-2)+1$, we can conclude that since $|G|=3(n-1)+1$ and does not contain C_n or a 4-element independent set, it must contain a C_{n-1} and a 3-element independent set. Let the 3-element independent set be $X:=\{x_1,x_2,x_3\}$ and the cycle be $\{v_1,\ldots,v_{n-1}\}$. The independent set and the cycle can be assumed to be disjoint because if we remove the $n-1$ vertices of our cycle, we still have

$2(n-1)+1$ vertices in our graph G, among them must contain a 3-element independent set.

Because we assume that G has no 4-element independent set, we can assume that every vertex in the cycle is connected to a vertex in X (if not, then a 4-element independent set would emerge). By the pigeonhole principle and because we assume $n > 5$, we know that $n-1 > 3$, hence there is at least one vertex in X which connects to more than one vertex in the cycle. We can assume without loss of generality that x_1 is connected to both v_i and v_j.

Now $v_{i+1}v_{j+1} \notin$ E because otherwise $(v_{i+1}v_{j+1}v_{j+2}\dots v_i x_1 v_j v_{j-1}\dots v_{i+1})$ is a cycle of length n, a contradiction (we added three edges and removed two). In addition, notice that no vertex in X can connect to two adjacent vertices in the cycle. This is because then a cycle of length n can be built because you added two edges and removed one. By this fact, the fact that $n > 5$, and the fact that if x_1 is adjacent to $v_i v_j$, $v_{i+1}v_{j+1} \in$ E (as proved already), we have $x_1 v_{j+2} \notin$ E otherwise, $\{v_{i+1}, v_{j+1} x_1 v_{j+3}\}$ will form an independent set.

Now v_{j+2} is adjacent to a vertex in X and it cannot be x_1, we can assume that this vertex is x_2. Now, $x_2 v_{i+1} \notin$ E since otherwise a cycle of length n can be formed. Also because no vertex in X can be adjacent to two consecutive vertices in the cycle, $x_2 v_{j+1} \notin$ E.

We have proved that $v_{i+1}v_{j+1} \notin$ E, $x_2 v_{j+1} \notin$ E, $x_2 v_{i+1} \notin$ E, and $x_1 v_{i+1} \notin$ E, $x_1 v_{j+1} \notin$ E (no vertex in X is adjacent to two consecutive vertices in cycle). Also obviously $x_1 x_2 \notin$ E since they are part of an independent set. Thus, $x_1 x_2 v_{i+1} v_{j+1}$ form a 4-element independent set, a contradiction.

We conclude that $R(C_n, K_4) \leq 3(n-1)+1$ so $R(C_n, K_4) = 3(n-1)+1$.

□

There are many improvements on this result over the years. In 1976 there was a conjecture by Erdos, Faudree, Rousseau, and Schelp that $R(C_n, K_m) = (n-1)(m-1)+1$ for $n \geq m \geq 3$. Over the years, many specific cases of this conjecture have been proved: for $n > 3 = m$, $n \geq 4 = m$, $n \geq 5 = m$, $n \geq 6 = m$, $n \geq 7 = m$, and more. "Small Ramsey Numbers" by Radziszowski [8] was a reference I used for the results.

2.3 Known Ramsey Numbers and Observations

Calculating Ramsey numbers is very difficult in general. For example, we know that $R(3,3) = 6$, and we know that $R(4,4) = 18$, but we only know $R(5,5)$ lies somewhere in between 43 and 49. Thus, finding exact values for Ramsey numbers even for small cases is already very difficult.

We make a few observations. From "Small Ramsey Numbers" [8], we have a lot of data about different Ramsey numbers. In particular, there is a conjecture that $R(C_n, K_m) = (n-1)(m-1)+1$ for $n \geq m \geq 3$. Now let's compare this to $R(K_n, K_m)$. Notice that there are $\binom{n}{2} = \frac{n(n-1)}{2}$ edges in K_n and there are n edges in C_n. Thus,

$$
\begin{aligned}
d_1(C_n, K_n) &= \frac{n^2 - n - n}{2n^2} \\
&= \frac{n^2 - 2n}{2n^2} \\
&= \frac{1}{2} - \frac{1}{n}
\end{aligned}
$$

Thus, the edit distance between K_n and C_n is bounded by $\frac{1}{2}$. We also notice that $R(C_n, K_m)$ increases polynomially, but $R(K_n, K_m)$ increases much faster than that. Hence, we noticed that just bounding the edit distance between two graphs does not bound the difference of their Ramsey numbers.

In addition we made another observation: $R(C_4, K_n) \leq R(C_3, K_n)$ for $n \geq 7$. Thus, it seems as if the structure of the graphs G and H in the computation of $R(G, H)$ has influence in the Ramsey number. C_4 in this case is much more "sparse" than C_3, hence it contains fewer graphs of size 4 as subgraphs than C_3.

2.4 Further Plans

There are many questions to be asked about the relation between Ramsey numbers and graph distances. For example, if I know the Ramsey number of two graph G and G′, and I know that H and H′ are a "small" distance away from G and G′ respectively, can we deduce anything about the Ramsey number for H and H′? In addition, how exactly does the structure of the graphs impact the Ramsey number?

Bibliography

[1] L. Lovasz, *Large Networks and Graph Limits*. American Mathematical Society, 2012.

[2] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American statistical association*, vol. 58, no. 301, pp. 13–30, 1963.

[3] A. A. Gittens, "Topics in randomized numerical linear algebra," 2013.

[4] R. Diestel, *Graph Theory*. Springer, 2010.

[5] F. P. Ramsey, "On a Problem of Formal Logic," *Proceedings of the London Mathematical Society. Third Series*, vol. S2-30, no. 1, p. 264, 1928.

[6] H. Y. R. Y. J. Sheng and Z. K. Min, "The value of the ramsey number," *Australasian Journal of Combinatorics*, 1999.

[7] R. J. Faudree and R. H. Schelp, "All ramsey numbers for cycles and graphs," *Discrete Mathematics*, vol. 8, pp. 313–329, 1974.

[8] S. P. Radziszowski, "Small ramsey numbers," pp. 19–20, 2014.

[9] M. Katz and J. Reimann, *An Introduction to Ramsey Theory*. Book, in preparation.

Academic Vita

Xingyu Zhang

(267) · 274 · 7887
xxz5058@psu.edu

4 Periwinkle Place Newtown, PA
18940

1400 Martin Street Apt. 2028 State College, PA
16802

EDUCATION

Penn State University Park — Expected: May 2015
Schreyer Honors College
B.S. in Mathematics

RESEARCH EXPERIENCE

Honors Thesis — May 2013 - August 2013, September 2014 - Present

- Researched under Professor Jan Reimann
- Investigated generalizations of Ramsey theory for graphs that are a certain distance from complete, where the distance is measured by the cut norm

Finite Dimensional Frame Theory — June 2014 - August 2014

- Part of Central Michigan Math REU
- Worked under Professor Yeonhyang Kim and worked with Hong Suh and Rachel Domagalski
- Studied the structure of tight finite dimensional frames

Parametrically Excited Waves on a Viscous Fluid — January 2012 - August 2012

- Researched under Professor Diane Henderson
- Studied the relationship between viscosity and surface tension to natural frequency and wave decay rate.
- Performed lab experiments in the Pritchard Fluid Mechanics Lab
- Obtained Mathematica coding experience

Special Educational Experiences

MASS program — Fall 2012

- MASS is a semester-long intensive math program at Penn State

Budapest Semesters in Mathematics — Spring 2014

- Participant in mathematics study abroad program in Budapest