Lecture 15: Pigeons and patterns

15.1 Pigeonholes

The simplest way to show that patterns must appear uses *the pigeonhole principle.*

The Pigeonhole Principle

If there are more pigeons than pigeonholes, one
pigeonhole must contain at least two pigeons.

The following application is based on contents of a webpage at the University of Manchester Institute of Science and Technology.

Some multiple of 12345 has a decimal expansion which
can be written using only the digits 0 and 1.

Why is this true? Consider the 1234**6** numbers $A_1, A_2, A_3, \ldots A_{12346}$ defined by the formula $A_n = \overbrace{111\ldots111}^{n\text{ 1's}}$ (in decimal notation). Further define the R_n's to be the remainders gotten from dividing the A_n's by 12345. These integer remainders are scattered between 0 and 12344, a total of 12345 possibilities. These possibilities are the *pigeonholes.*

There are 12346 remainders: the *pigeons.* The pigeonhole principle implies that at least two of these R_n's are equal. So we may suppose that there are remainders R_a and R_b where $a > b$ so that $R_a = R_b$. Then $A_a - A_b$ is divisible by 12345 because its remainder, the difference $R_a - R_b$, is 0. What is $A_a - A_b$? It must be $\overbrace{111\ldots111}^{(a-b)\text{ 1's}}\overbrace{000\ldots000}^{b\text{ 0's}}$

It isn't too difficult to write a computer program which implements the scheme outlined above. For example, here's a multiple of 13 which is all 0's and 1's: $85470 \cdot 13 = 11\,11110$. But the number found may not be the smallest answering the question: $77 \cdot 13 = 1001$ is also a solution which is quite a bit smaller! More critically, the process outlined above shows that certain numbers exist, but it doesn't find them. A direct search for the multiple of 12345 using the idea above could involve manipulation of numbers with 12345 digits, which is somewhat intimidating, even for a computer.

Exercise Find an explicit multiple of 12345 which has only 0's and 1's in its decimal expansion.

15.2 M&M's

Here's a question involving pattern inevitability taken from a webpage at Illinois State.

The colors of M&M's are red, green, brown, yellow, blue, and orange.
How many M&M's would you have to grab from a package to guarantee
that you have grabbed at least three of one color?

Certainly a "random" three M&M's taken could have the same color. Let's analyze that probability. Here we use information supplied by Mars, Inc. of Hackettstown, New Jersey.

Brown
Fact: 30% of all M&M's are brown.

Consequence: if we choose three different M&M's independently, then $(.3)^3 = .027$ of the time we'd have three brown M&M's.

Yellow & Red
Facts: 20% of all M&M's are yellow; 20% of all M&M's are red.
Consequences: the chance of getting three yellow M& M's is $(.2)^3 = .008$ and the chance of getting three red M& M's is the same.

Green & Orange & Blue
Facts: 10% of all M&M's are green; the same is true for orange and for blue.
Consequences: The chance of getting three greens is $(.1)^3 = .001$, and this is the same as the chance of getting three oranges or three blues.

Let's assume that the manufacturer has supplied correct information and that the selection of the three M&M's is independent. Then the event of selecting three of the same color splits into six disjoint events, one for each color. The probability of selecting three of the same color is therefore $.027 + 2(.008) + 3(.001) = .046$, roughly one chance in twenty. This reasoning is, of course, probabilistic. Suppose we wanted to be certain of getting M&M's with the same color.

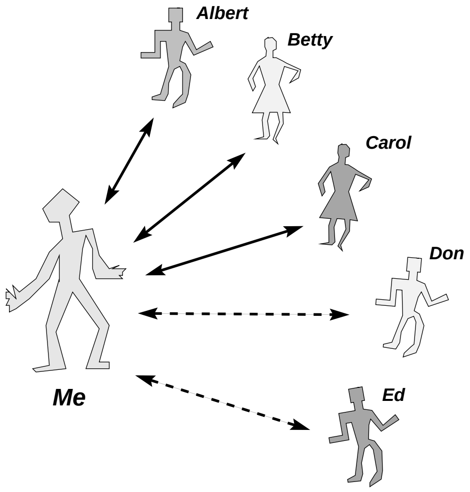
We could grab more. If we took 7 M&M's, then at least one color would appear twice. This is a consequence of the Pigeonhole Principle: the pigeonholes would be the color of the candy taken, and the pigeons would be each M&M. Taking 7 M&M's with 6 colors available is the same as having 7 "pigeons" (the 7 chosen M&M's) and 6 pigeonholes, one hole for each color chosen: two M&M's of any 7 must have the same color. Similar reasoning shows that any group of **13** M&M's must have three of the same color.

Exercise Carefully compute the probability of three M&M's of one color being grabbed in a random handful of N, where $3 \le N \le 13$. For example, the probability is .8 when $N = 12$. The probability will grow from .046 ($N = 3$) to 1 ($N = 13$).

The pigeonhole principle can produce mandatory "structure" in randomness.

Lecture 16: Friends, strangers, and coloring graphs

16.1 Mutual friends and mutual strangers

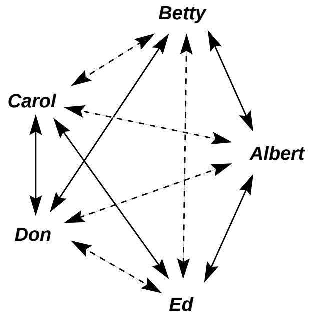


Let's suppose that any pair of people in the world are either mutual friends or mutual strangers. Take any six people in the world. Must there be a group of these people (more than two!) who are all mutual friends or mutual strangers?

Suppose I (in the lovely picture: "*Me*") am one of the people, and the others are *Albert*, *Betty*, *Carol*, *Don*, and *Ed*. Some of these five people are strangers to me, and some are my friends. By the pigeonhole principle, at least three of these must be one of these two alternatives (strangers to me or friends of mine). Let me assume that three are friends, and that *Albert* and *Betty* and *Carol* are these friends.

Consider *Albert* and *Betty* and *Carol* who could all be unacquainted with each other. So these people then would be mutual strangers. If this assumption is false, at least one pair of them, say *Albert* and *Betty*, would be friends. Then *Albert* and *Betty* and *Me* would be three mutual friends.

We could continue backtracking the assumptions that were made. What would happen if there were three people who were strangers to *Me*? Similar reasoning will again provide three mutual friends or three mutual strangers.



What if we considered a collection of five people? Look at the diagram displayed. The solid lines indicate that the people are mutual friends, while the dashed lines imply they are mutual strangers (better thought: "friends who haven't met"). Check carefully: each of the 10 relationships is shown. There is no triple of mutual friends or mutual strangers.

Here is the major conclusion:

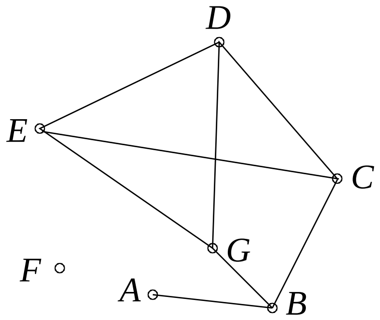
> *Any* group of six people must include a triple of mutual friends or a triple of mutual strangers but there can be groups of five people having no triple of mutual friends and no triple of mutual strangers.

16.2 Graphs and coloring four

We just considered what could happen at a party with a group of people. We were interested in whether a small number of them were mutual friends or strangers. We are actually studying *graphs**.

---

* Mathematicians tend to overload simple words, making them serve several distinct meanings. The objects called graphs here are *not* the same as the graphs considered in analytic geometry, such as "the graph of $y = x^2$." Sorry about this.



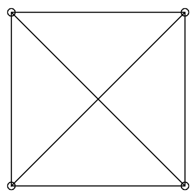
Here a graph will be a collection of vertices (the plural form of "vertex") and edges between pairs of vertices. One graph is shown here. The vertex set of this graph has 7 elements, $\{A, B, C, D, E, F, G\}$; the edge set has 8 elements, $\{(A, B), (B, C), (B, G), (C, D), (C, E), (D, E), (D, G), (E, G)\}$.

The vertex B is a member of three edges, and we say that the degree of B is 3: $\deg(B) = 3$. The degree of F is 0. The picture is a symbolic representation of the graph so the edges (C, E) and (D, G) don't really meet or intersect.

The concept of a graph has many variations (directed graph: edges with an orientation; weighted graph: edges with associated costs; a multigraph: more than one edge may connect vertices, etc.). There is also much specialized terminology.

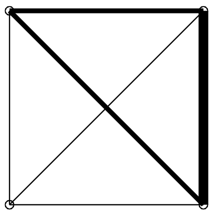
Many simple questions having difficult answers have been asked about graphs. One of the most famous optimization problems asks about graphs. Suppose there is a list of cities and distances between them is given. Finding a tour visiting each of the cities with minimum travel distance is the Traveling Salesman Problem. This is certainly a difficult (NP!) problem. It has <u>many</u> applications*.

Here we will look at complete graphs. A graph is *complete* if every pair of distinct vertices is connected by an edge. The complete graph on n vertices is called K_n.



K_4

K_4 is shown at the left. It has 4 vertices and it has $3+2+1 = 6$ edges. The picture at the right shows this sum: 3 counts the edges connecting a first vertex to all the other edges (edges drawn with standard thickness); 2 comes from connecting another vertex to all vertices except the first (edges drawn with increased thickness); 1 (the most heavily drawn edge) is from the next-to-last vertex, connecting it to the last vertex.



1+2+3=6

More generally, the number of edges in K_n is $\frac{n(n-1)}{2}$. You can see this in several ways: by looking at the pairs of vertices (n times $n-1$, since no vertex is connected to itself) and dividing by 2 (since the order of the pairs doesn't matter). Or you can count downwards and add: $n + (n-1) + (n-2) + \ldots + 2 + 1 = \frac{n(n-1)}{2}$.

We will **color** the edges of K_n's. Here I want to use only two colors, say **red** (indicated with ———) and **blue** (indicated with - - - -). Since each edge can be colored one of these **2** colors, K_n can be colored $2^{(\text{\# of decisions})} = 2^{n(n-1)/2}$ different ways. In the previous section, we actually verified the following statement:

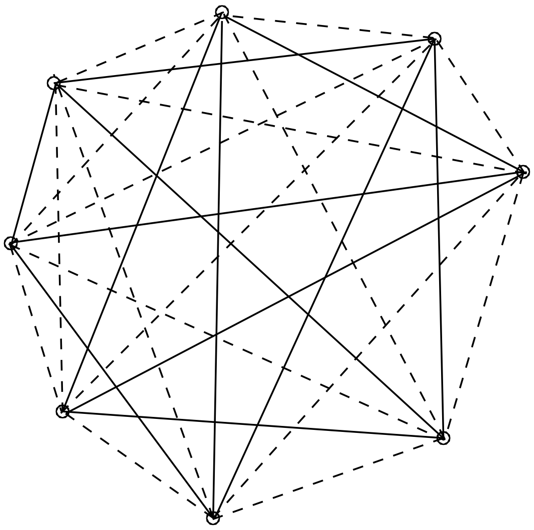
> If $n \geq 6$ then every 2-colored K_n contains a monochromatic triangle (K_3). If $n \leq 5$, there are colorings of K_n's not containing a monochromatic triangle.

Is there a similar statement about K_4's? That is, can we find a threshold number N so that there is <u>some</u> 2-colored K_{N-1} which contains *no* monochromatic K_4 and so that <u>every</u> 2-colored K_N *must* contain at least one monochromatic K_4? Statements of this type are not easy to check because the number of possible 2-colorings is very large.

---

* See `http://www.math.princeton.edu/tsp/` for example.

For example, here's one 2-colored K_8. It has 28 edges and *no* monochromatic K_4. To check this claim, look at all of the possible K_4's and see if they have one color. How many K_4's are there? First find the number of distinct 4-tuples of vertices: you can choose the first vertex 8 ways, the second 7, then 6, and then 5. The result is 1680. But we get the same K_4 if these 4 vertices are chosen in any order. The number of orders of 4 objects is $4 \cdot 3 \cdot 2 \cdot 1 = 24$. We need to check *only* 70 different K_4's*. You may know this is the binomial coefficient $\binom{8}{4}$.



Check all 70!

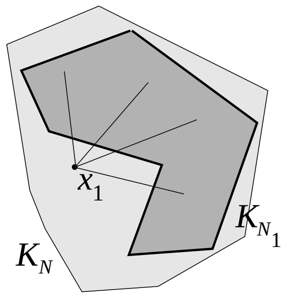
How **hard** is it to check for monochromatic K_4's in a 2-colored K_N? We might need to check the coloring of every K_4 in the given K_N. There are $\binom{N}{4}$ different K_4's: that is, we just pick 4 vertices and look at the 6 edges connecting these vertices. The number of ways of choosing 4 objects from N objects is $\binom{N}{4} = \frac{N(N-1)(N-2)(N-3)}{1\cdot 2\cdot 3\cdot 4} = \frac{N^4-6N^3+11N^2-6N}{24}$. This is a polynomial of degree 4. We could search one 2-colored K_N for monochromatic K_4's with a program having four "nested" loops. This is not really too bad. For example, when $N = 15$, there are 1,365 K_4's. But ... to check that *every* 2-colored K_N has a monochromatic K_4, we would need to check $2^{N(N-1)/2}$ different 2-colored K_N's. When $N = 15$, there are $2^{105} = 40\,56481\,92073\,03340\,84789\,45025\,72032 \approx 4 \cdot 10^{31}$ different colorings to check. A direct check is not feasible even with huge computer resources (the return of P versus NP!). We have two tasks:

I Find N large enough so that *every* 2-colored K_N has a monochromatic K_4. It isn't even clear that there is such an N, and therefore exhaustive searching may not succeed.

II For suitable N's, exhibit 2-colored K_N's without any monochromatic K_4's. Certainly $N = 8$ works here (the example above!). But can larger N's be used and still *not necessarily* have monochromatic K_4's? Even here searching be impractical.

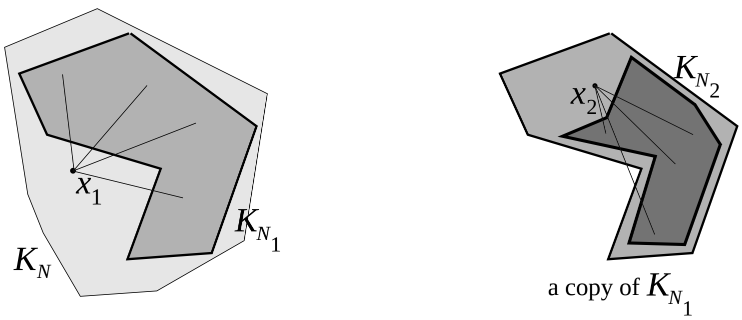
How can we find a monochromatic K_4? Let's start with a very BIG 2-colored K_N and try to find a K_4 with only one color. We search for a monochromatic K_4 in a greedy way. First, pick *any* vertex: call it x_1. Look at the edges coming out of x_1: there are $N-1$ of them. Some are red and some are blue. There are at least $\frac{1}{2}(N-1)$ edges of one color, because if the number of edges of both colors were $< \frac{1}{2}(N-1)$, then the total number of edges would be less than $N-1$. Let's see: call that color C_1, and collect all the vertices which connect to x_1 with edges of that color. Look at the complete graph (with assigned edge colorings!) on those other vertices. It is a K_{N_1} whose internal edges are colored red and/or blue. The "and/or" is used to emphasize that we have no control over the internal edges of K_{N_1}. Also, $N_1 \geq \frac{1}{2}(N-1)$.

* $70 = \frac{1680}{24}$.



Here is an attempt to illustrate what has just been described. The lighter colored polygonal region "is" K_N. Then x_1 is inside K_N, and is connected by edges of all one color to K_{N_1}, which takes up at least half of K_N. We don't know whether the edges connecting the vertices of K_{N_1} to x_1 are red or blue, only that they are all the same color.

Now pick a vertex x_2 in K_{N_1}. Look at the edges coming out of x_2 in K_{N_1}: there are $N_1 - 1$ of them. There are at least $\frac{1}{2}(N_1 - 1)$ edges of one color, because if the number of edges of <u>both</u> colors were $< \frac{1}{2}(N_1 - 1)$, then the total number of edges would be less than $N_1 - 1$. Call that color C_2. I am *not* asserting it is the same as C_1! Collect all the vertices in K_{N_1} which are at the other end of these edges coming out of x_2. Look at the complete graph (with edge colorings!) on those other vertices. It is a K_{N_2} whose internal edges again are colored red and/or blue. Also, $N_2 \geq \frac{1}{2}(N_1 - 1)$.



This picture tries to show the "evolution" from K_N to K_{N_1} to K_{N_2}. Symbolically we could write:

$$x_1 \text{ in } K_N \stackrel{C_1}{\Rightarrow} K_{N_1} \text{ and } x_2 \text{ in } K_{N_1} \stackrel{C_2}{\Rightarrow} K_{N_2}$$

where each time the number of vertices of the $K_{\text{something}}$ is at least one-half the number of vertices of the previous graph. We'll worry about the specific N needed later.

Let's continue the process. We get

$$x_1 \stackrel{C_1}{\Rightarrow} K_{N_1} \text{ and } x_2 \stackrel{C_2}{\Rightarrow} K_{N_2} \text{ and } x_3 \stackrel{C_3}{\Rightarrow} K_{N_3} \text{ and } x_4 \stackrel{C_4}{\Rightarrow} K_{N_4} \text{ and } x_5 \stackrel{C_5}{\Rightarrow} K_{N_5}$$

where we stop after five evolutions or reductions. Each of the colors (the C_j's) is either red or blue, and there are five of them. The pigeonhole principle again implies that at least one color (say **red** for specificity) must appear at least three times (since $2 \cdot 2 = 4 < 5$ both can't appear less than three times!). Again, let us assume for specificity that C_2, C_3, and C_5 are red. The logic will be similar for any other triple.

Now we will discover a red K_4 in this K_N. Take y in K_{N_5}. Since C_5 is red, (x_5, y) is red. But the creation of the new graphs is by shrinking, so y is also in K_{N_3} and so (x_3, y) is red. Similarly, y is in K_{N_2} and (x_2, y) is red. Since x_5 is in K_{N_4}, it must be in K_{N_3} and therefore (x_3, x_5) is red. (x_2, x_3) and (x_2, x_5) are also red with the same arguments. So the vertices $\{x_2, x_3, x_5, y\}$ and the associated edges give a monochromatic (in this case, red) K_5.

We must go back and check on the size of N. Since K_{N_5} is not empty (we need a y in it!), the size of K_{N_5} is at least 1. Then K_{N_4} should be at least 3, since we create K_{N_5} by taking out x_5 and being greedy. K_{N_3} must have at least 7 vertices, since we take out x_4 and are greedy. Backwards again: K_{N_2} must have at least 15 vertices. And again: K_{N_1} must have at least 31 vertices. And finally: K_N must have at least 63 vertices.

If $N \geq 63$, then every 2-colored K_N contains a monochromatic K_4.

A logical scheme following this outline verifies a more general statement.

If $N \geq 2^{2(n-1)} - 1$, then every 2-colored K_N contains a monochromatic K_n.

Lecture 17: Ramsey and five

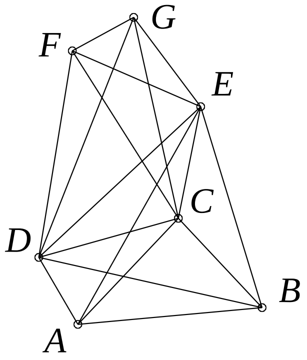
17.1 Definitions and $\mathcal{R}_5$

The n^{th} Ramsey number, which will be called $\mathcal{R}_n$ here, is the "threshold number" for n. It is the smallest integer so that every 2-colored K_N with $N \geq \mathcal{R}_n$ must have a monochromatic K_n. The procedure outlined above actually shows that $\mathcal{R}_n$ exists, which is *not* obvious, and that $\mathcal{R}_n \leq 2^{2(n-1)} - 1$.

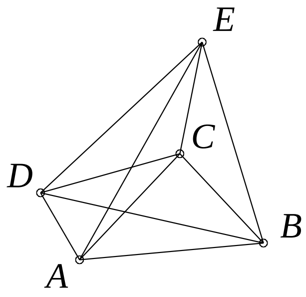
We have shown that $\mathcal{R}_4 \leq 63$ and our earlier example shows that $\mathcal{R}_4 \geq 9$. In fact, $\mathcal{R}_4$'s exact value is known: 18. This takes more work, though.

The exact value of $\mathcal{R}_5$ is *not* known*. I will show you an *underestimate* of $\mathcal{R}_5$ using a technique which is ludicrously simple (maybe) and fiendishly clever (really). It is relatively new and has become a powerful method for analyzing problems in many areas of theoretical computer science. It is also useful in aspects of theoretical physics, chemistry, and bioinformatics (computational biology).

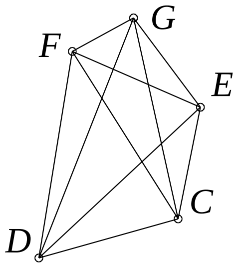
Let's begin with an *uncolored* K_N. Let's flip coins to color the edges of this K_N. So we flip a fair coin (equal probability for heads and tails) repeatedly. Color an edge red if the coin lands heads, and blue if it shows a tail. Suppose that S is one of the K_5's in K_N. What's the chance that S has all edges colored blue? S has 10 edges either because $\binom{5}{2} = 10$ or because $1+2+3+4 = 10$. If the flips are fair and independent, the chance that S is all blue is $\frac{1}{2^{10}}$. Of course, S could also have all of its edges colored red, with the same probability. Therefore the chance that S is monochromatic is $\frac{1}{2^9} = \frac{1}{512}$. So far this is exactly like our analysis of trying to get three M&M's of one color. We computed the chance of the event that the three M&M's were green, then the chance they were red, etc., and finally we added these chances. This was simple because there was no overlapping of the random choices involved in these events. That is, there was no interaction between the events. But this situation is more complicated. An example may indicate the complexity.



Seven vertices with some edges in a K_N



This is S_1



This is S_2

What's shown is supposed to be seven vertices in a very big K_N. I haven't drawn every edge in order to keep the picture simpler. One K_5 is S_1 with vertex set $\{A, B, C, D, E\}$, and another is S_2 with vertex set $\{C, D, E, F, G\}$. Each K_5 has 10 edges. S_1 and S_2

* The best current information seems to be that $\mathcal{R}_5$ is between 43 and 49.

share three edges: (C, D), (C, E), and (D, E). The chance that S_1 is monochromatic is $\frac{1}{2^9}$ and the chance that S_2 is monochromatic is $\frac{1}{2^9}$. A careful computation shows that the chance that *both* of them are monochromatic is $\frac{1}{2^{16}}$. So the chance that at least one is monochromatic is less than the sum of the chances that each of them is monochromatic:

$$P(S_1 \text{ is monochromatic}) + P(S_2 \text{ is monochromatic}) = \\ P(\text{At least one is monochromatic}) + P(\text{Both are monochromatic})\,.$$

But we don't need to be so careful if we want a quick overestimate. We will forget the "fine structure" here and only use this idea, which overcounts:

> The chance that at least one K_5 is monochromatic will be overestimated by *the sum* of the chances that *any* K_5 is monochromatic.

The number of K_5's in K_N is $\binom{N}{5} = \frac{N(N-1)(N-2)(N-3)(N-4)}{5!} = \frac{N^5-10N^4+35N^3-50N^2+24N}{120}$. Each one separately has monochromatic probability $\frac{1}{2^9} = \frac{1}{512}$. So we can *overestimate* the probability of having at least one monochromatic K_5 by* $\frac{N^5-10N^4+35N^3-50N^2+24N}{61440}$. When $N = 11$, this number is approximately .90234, less than 1. If every 2-colored K_{11} had a monochromatic K_5, this number would have to be 1. So there must be some 2-colored K_{11} having no monochromatic K_5, and $\mathcal{R}_5 \geq 12$.

This underestimate of $\mathcal{R}_5$, showing the existence of a 2-colored K_{11} with no monochromatic K_5, is an example of "the probabilistic method". If this all seems clear, try the following exercise.

Exercise Find an explicit 2-coloring of K_{11} which has no monochromatic K_5.

Comment There are 462 K_5's in a K_{11}, and there are $2^{55} = 36\,02879\,70189\,63968 \approx 3 \cdot 10^{21}$ 2-colorings of K_{11}. This search space is very large. How can you find the 2-coloring requested? An exhaustive search may not quickly produce an example. Our computations assert that a "random" 2-coloring will be satisfactory about 10% of the time (actually more because we got an overestimate). Here's one strategy: first write a program to check for the presence of monochromatic K_5's in a 2-colored K_{11}. Then create and test some random 2-colorings of K_{11} (you'll need at least a reasonable source of randomness!). You may fail to get a satisfactory coloring, but as you choose more and more random colorings, your chance of success *at least once* is close to certainty. That's because repeated failure for t attempts has approximate probability $\leq (.9)^t$ which, as t grows, approaches 0 rapidly. For example, there's less than 1 chance in 40,000 of failure in 100 random attempts to create a suitable coloring. A direct search with carefully structured colorings, starting with all red edges, etc., may take a while to report success: your search may be stuck in an untypical corner of the exponentially large sample space, where there are many monochromatic K_5's. "Random" will likely be faster than "deterministic" in this problem!

17.2 Thanks, a further remark, and a charming story

Conversations with David Galvin, János Komlós, and Jason Tedor helped me prepare this material. Dr. Komlós is a faculty member in mathematics and computer science at

---

* $61440 = 120 \cdot 512$.

Rutgers and is one of the world's leading authorities on probabilistic methods. Mr. Galvin and Mr. Tedor are currently graduate students in mathematics at Rutgers.

The methods given here for over- and underestimates of Ramsey numbers $\mathcal{R}_n$ are, for n large, quite close to the best currently known! They can be used to show that $\mathcal{R}_n$ is always between 4^n and $\left(\sqrt{2}\right)^n$.

Joel Spencer (in [6]) tells this anecdote about Paul Erdős, who was one of the greatest and strangest mathematicians of the twentieth century (see [8] and [9]). Erdős essentially invented the probabilistic method.

> Erdős asks us to imagine an alien force, vastly more powerful than us, landing on Earth and demanding the value of $\mathcal{R}_5$ or they will destroy our planet. In that case, he claims, we should marshall all our computers and all our mathematicians and attempt to find the value. But suppose, instead, that they ask for $\mathcal{R}_6$. In that case, he believes, we should attempt to destroy the aliens.

17.3 Bibliography

[1] Alexandre V. Borovik and Elena V. Bessonova, an undated webpage of the University of Manchester Institute of Science and Technology:
`http://www.ma.umist.ac.uk/avb/Pigeon.html`

[2] Roger Day, a webpage from a summer 2001 course at Illinois State University:
`http://www.math.ilstu.edu/~day/courses/old/305/pigeonholesamples.html`

Current information about Ramsey numbers is available here.

[3] Stanislaw Radziszowski, Dynamic Survey on *Small Ramsey Numbers*, 2001, *Electronic Journal of Mathematics*: `http://www.combinatorics.org/Surveys/`

There may be information about Ramsey theory and Ramsey numbers in introductory books about combinatorics and discrete mathematics, but the material is rather new. The following texts are the standard professional references on Ramsey theory.

[4] Noga Alon and Joel Spencer, *The Probabilistic Method*, second edition, 2000, John Wiley (320 pages, $90, an advanced text).

[5] Ronald L. Graham, Joel H. Spencer, and Bruce Rothschild, *Ramsey Theory*, second edition, 1990, John Wiley (200 pages, $145*, an advanced text).

[6] Joel Spencer, *Ten Lectures on the Probabilistic Method*, 1987, Society for Industrial and Applied Mathematics (a paperback, 78 pages, $29, perhaps a bit more accessible than the other two references).

There are several advanced computer science texts on randomized algorithms. One reviewer writes this about the text below: "For computer scientists, this is *the* reference work in randomized algorithms, by now a major paradigm of algorithms design."

[7] Rajeev Motwani and Prabhakar Raghavan, *Randomized Algorithms*, 1995, Cambridge University Press (476 pages, $55).

* This price is not a misprint.

Here are two books about Paul Erdős, a great mathematician with a highly unusual life:

[8] Paul Hoffman, *The Man Who Loved Only Numbers: The Story of Paul Erdős and the Search for Mathematical Truth*, 1999, Little Brown (a paperback, 320 pages, $10).

[9] Bruce Schechter, *My Brain Is Open: The Mathematical Journeys of Paul Erdős*, 2000, Touchstone Books (a paperback, 224 pages, $10).