Chapter 1

Setting the Stage

Introductory words

This course goes under the umbrella of the “Topics in applied mathematics.” Hence, you should expect “applied” part and “mathematical” parts of the course. The general idea is to consider mathematical models of various real objects, which can be given a natural structure of a network (mathematically, of a graph, when we can specify the set of nodes and the set of edges, sometimes with some additional structure). A prototypical example is the World Wide Web (WWW), where the nodes are web pages and the edges are the links pointing to other web pages (this is an example of a directed graph; all the necessary mathematical definitions will be given below). Thinking of the structure of WWW network, many various questions can be asked, including: What are the distinct properties of this graph? Is this graph connected? If not, what can we say about the structure of the various components of this graph? What is the diameter of WWW? Is World Wide Web stable with respect to removing some of the nodes? One way to answer these questions is to collect the data of the web and analyze it. However, the size of WWW is such that it is hardly feasible to collect data about all the possible pages and links. More importantly, WWW is evolving, and collecting the data would mean constantly updating it, which is unrealistic. Therefore, another method is utilized — *mathematical modeling*. Instead of World Wide Web we consider a mathematical model of a network. Our model should reflect (at least some) salient features of the object under study, whereas the many otherwise important details must be disregarded to guarantee that at least some insight can be obtained out of the analysis of the model. How to do this for the web? Which features should be kept and which should be disregarded?

Here we face the problem that we need to model an object about which we do not have full knowledge. Here is a simple solution: Let us assume that this object is random! Think about the mathematical model of the coin toss. If we knew all the details about the coin position, coin material, and forces applied to the coin, we would be able to say exactly what is the outcome of the coin toss. We do not have this information, hence we build a probabilistic model which roughly says that if the coin is fair then in a long run of experiments in approximately half of the cases we will see heads and in approximately half of the cases we will see tails. There is significant vagueness in this description, but using the powerful machinery of the probability theory all the statements can be made rigorous; the agreement of the mathematical theory with observations confirms that the premises on which the probability theory is build can be used to describe the surrounding reality. Therefore, in this course we are going to build *probabilistic models of networks or graphs.*

In many cases the models will be built such that at least some of the features of the real objects are kept intact, and this is the contribution of the “applied” part of the course. But the goal of the course is *the analysis of mathematical models of random networks* and not the relations of these models with the world around us. We will touch on this latter subject, but we will never claim any “theorems” or “proofs” about the relations of our abstract models and the real world. Therefore, the “applied” part of the course will mostly consist of the fact that the type of analysis we will be doing is generally considered to be relevant for analysis of many physical, biological, chemical, economic, social (you can continue from here) systems; however any discussion of “a real system” in this course serves only as a toy example for our abstract models.

Having said all this, I would like first to set the stage for the course with an “applied” motivation to study complex networks — i.e., with a short answer to the question “What kind of networks we would like to model?” and also with a “mathematical” illustration of the power of the methods we will be using — i.e., with a short answer to the question “What kind of mathematics we will be using?” The student *should not* feel discouraged if some (or even all) of the ideas, methods, or descriptions are not clear at this point, we will return to most of them in a systematic way later; these two following sections only “wake up the appetite.”

1.1 What we would like to model: Web graph and other examples

This section tries to discuss in a concise form which features of the real world networks we would like to include in our mathematical models. It is clear that every particular network has its own unique features, but here we are interested mostly in very general features, which probably are true only approximately.

A good example to have in mind is the World Wide Web graph. Here are two examples of the web graph (actually, of a 2-neighborhood of some web pages; 2-neighborhood means that only the web pages at the distance no more than 2 links from the original one are shown). For the first one the starting node or vertex was my personal web page, whereas for the second one I start with the page `www.ndsu.edu`. The difference is obvious.

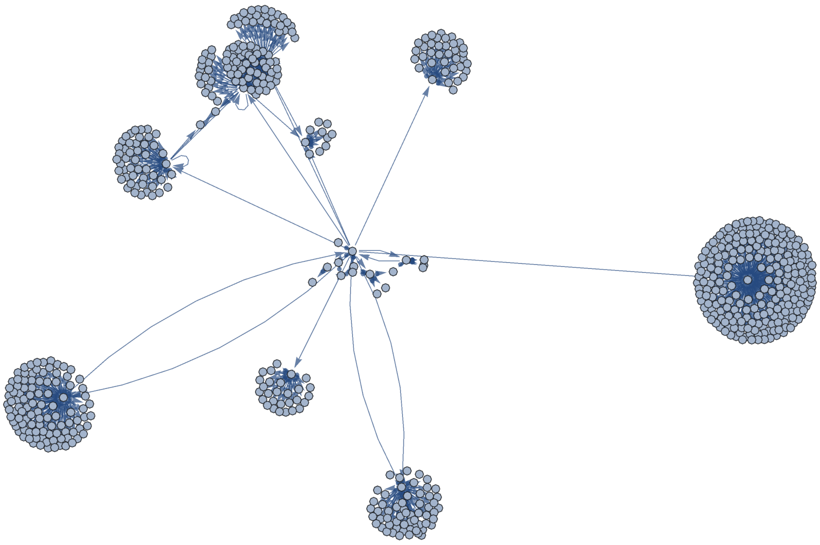

Figure 1.1: A tiny portion of the web graph (technically, a 2-neighborhood of my own web page https://www.ndsu.edu/pubweb/ novozhil/)

Here are the main properties of the web graph, which are shared by many other real world networks.

- *It is large.* Actually, the web graph is simply huge. And it is growing. Here are a few numbers: in 1998 it was estimated that there were at least 320 million web pages in late 1997; in 2005 it was reported that the web had around 11.5 billion pages. Another estimate from 2005 gives 53.7 billion pages, with 34.7 billion pages indexed by Google. A very recent number of pages indexed by Google is above 40 billion pages[1].

- *It is sparse.* This simply means that the number of edges is significantly smaller than it is theoretically possible to have for this enormous number of vertices. To define the sparseness mathematically, we will need to consider formalization of the web as a graph $G = (V, E)$, which consists of the set of vertices V and the set of edges E. The cardinalities of these sets are, respectively, $|V|$ and $|E|$ (the notation $|A|$ for a finite set A means the number of elements belonging to this set). The graph is sparse if $|E| = \mathcal{O}(|V|)$, i.e., in words, the number of edges has the same order as the number of vertices. Note that for many graphs (i.e., for the graph without multiple edges connecting the same vertices and without loops, i.e., without the edges that start and end at the same vertex)

[1] http://www.worldwidewebsize.com/

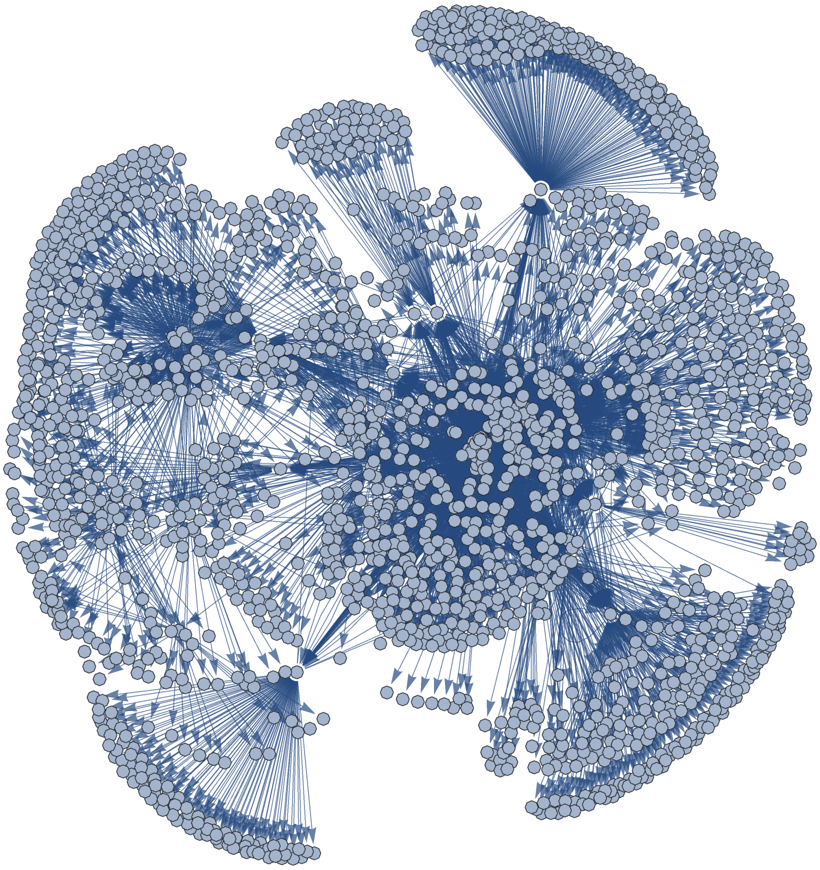

Figure 1.2: A tiny portion of the web graph (a 2-neighborhood of https://www.ndsu.edu)

this is not true since the maximal number of edges for a simple graph on n vertices is given by $n(n-1)/2 = \mathcal{O}(n^2)$ (can you explain why this particular formula is true?).

- *It is small world*. This means that despite the fact that there are so many vertices and "not so many edges" (see the previous points), for any pair of nodes it is possible to find a path (an ordered set of vertices connected by links) that is quite short. We'll get to the mathematical formalization in due course, but now just look at the notion of *the Erdős number*, which is illustrated in the figure below. Paul Erdős (1913 — 1996) was an extremely prolific mathematician, whose works are very relevant to our course. He is also famous in that most of his papers are written with co-authors. Now imagine *the collaboration graph*, whose nodes are the authors, and nodes u and v are connected with a link if u and v published a joint paper, on which they both are authors. Paul Erdős in this graph plays the rôle of *a hub*, i.e., of a node that has very many links (this node has *high degree*). For almost any mathematician (and also for physicists, biologists, etc) it is possible to find a path to Paul Erdős. By definition Erdős himself has the Erdős number 0. All his direct co-authors have the Erdős number 1, all people that co-authored a paper with those with the Erdős number 1 and did not co-authored a paper with Paul Erdős himself have the Erdős number 2, and so on. What do you think would be the Erdős number for a randomly chosen scientist? Since the co-authorship graph is definitely small world, this number is usually quite low. For instance, my Erdős number is 3 (here, an explanation is that I co-authored a paper with another hub of this network — biologist

Eugene Koonin from National Institutes of Health).

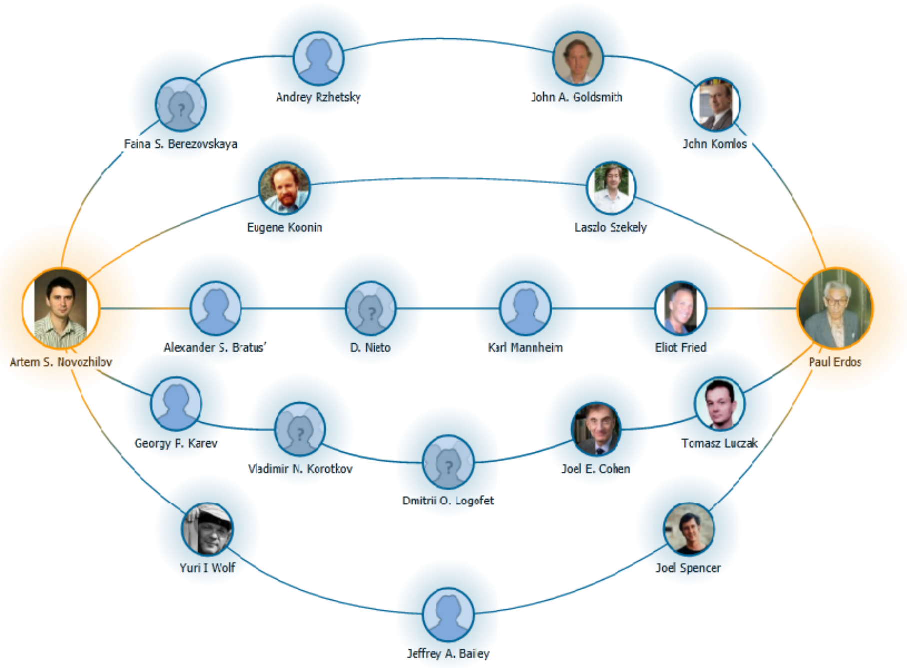

Figure 1.3: An example of a small subgraph of the collaboration graph. My Erdős number is 3. The figure is taken from *Microsoft Academic Search*

- *Power law degree distribution.* Degree of a given vertex of a network is defined as the number of vertices adjacent to this vertex (i.e., the number of directly linked nodes). Given a graph, we can always speak of its degree distribution: This is simply the number of vertices that have degree zero (no edges), number of vertices that have degree 1, number of vertices that have degree 2, and so on. So this degree distribution is a very useful statistical characteristic of a network, especially if a network is a large one, drawing of which is not an option. By definition, the network has a power law distribution if the number of vertices with degree k is proportional to $k^{-\gamma}$ for some parameter γ. Note that this implies that in double logarithmic scale this degree distribution must look like a straight line (see the figure below for an example).

- Real world networks have high *clustering coefficients*. In simple words it means that "a friend of a friend has a much higher chance to be my friend than a random person." Mathematically, it is a ratio of the number of triangles in the network to the potential number of triangles in it. It is highest for the full graph K_n and equal to 1 (graph is full if all possible edges are present), and equal to 0 for, e.g., a tree. The exact definitions will be given later.

Problem 1.1. Think of the following networks:

- *The collaboration graph*, which was described in the lecture.

- *The sexual relations graph.* Here the nodes are people, and the edges correspond to sexual relations.

- *The word graph.* The vertices are words, and two words are joined if they appear in the same sentence.

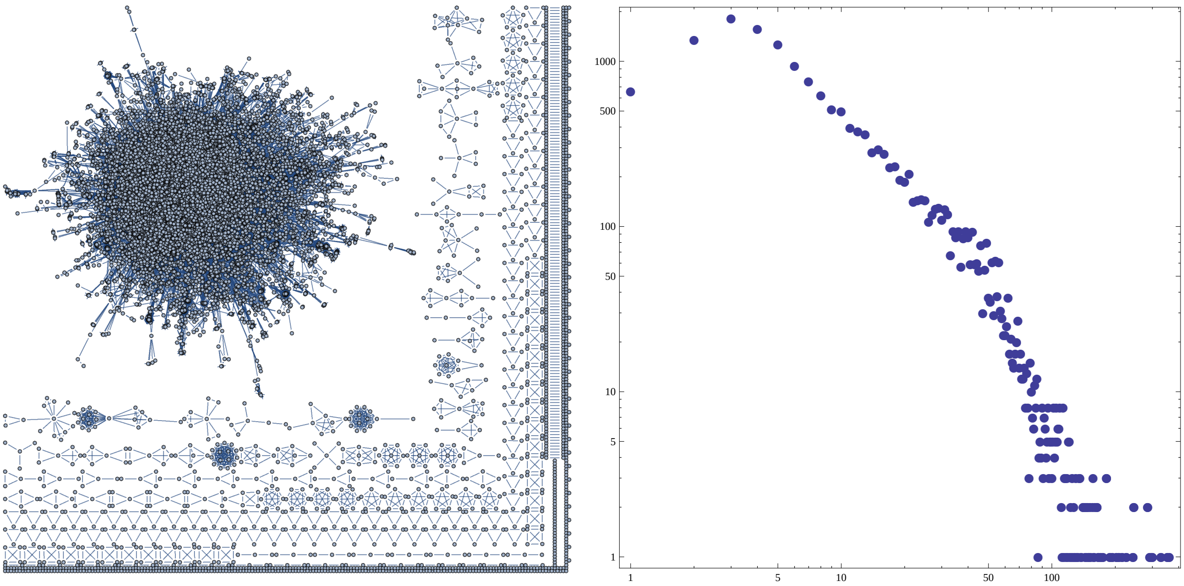

Figure 1.4: Collaboration graph for astrophysics (left) and its degree distribution in double logarithmic coordinates (right)

- *The transportation graph.* E.g., the vertices are the airports, and the edges correspond to the direct flights between these airports.
- *The movie actor networks.* All actors are nodes, and they are connected by an edge if they happened to be in the same movie.

Which of these graphs, in your opinion, have power law degree distribution? Which are sparse and which are small world? Which have higher clustering coefficients and have lower ones? Justify your answer heuristically.

Problem 1.2. Can you give your own examples of real-world networks, which can be modeled mathematically as (large, sparse, small world) graphs?

1.2 How we will model: The Ramsey numbers

The main technical device of the course is *the probabilistic method.* This section gives an example of this mathematical technique to analyze the *Ramsey numbers.* At the very beginning of this section I would like to note that this topic is only tangential related to the rest of the course, since most of the time we will be studying the properties of the random graphs for the sake of the random graphs themselves. Here we will use the ideas of the random graphs to study something else.

As a very informal introduction to this vast mathematical subject of Ramsey numbers, consider the following problem.

Problem 1.3. Prove that among any six people either there are three persons that know each other, or there are three persons that do not know each other. *Hint:* Pigeonhole principle.

To define the Ramsey numbers, consider coloring of the edges of a graph $G = (V, E)$ into red and blue (I will give all the necessary definitions of the notions I am going to use in the course in the following lectures. Here I assume just a common knowledge of a graph as can be visualized using nodes and links connecting some of these nodes. In particular, notation $G = (V, E)$ means that I consider graph G which has the set of nodes or vertices V and the set of edges E.)

The Ramsey number $R(s, t)$ is by definition the minimal natural number $n \in \mathbf{N}$ such that for any coloring of the full graph K_n (i.e., for the graph on n vertices, where every two vertices are connected by an edge) there exist either a full subgraph $K_s \subseteq K_n$ whose edges are all red, or a full subgraph $K_t \subseteq K_n$ whose edges are all blue.

For example, the problem above shows that $R(3, 3) \leq 6$. Actually, $R(3, 3) = 6$ (prove it). Also, one has $R(0, t) = 0$, $R(1, t) = 1$, $R(2, t) = t$, and $R(s, t) = R(t, s)$.

It is possible to show (left as an exercise) that for $s, t > 1$

$$R(s, t) \leq R(s - 1, t) + R(s, t - 1),$$

which implies that

$$R(s, t) \leq \binom{s + t - 2}{s - 1},$$

where

$$\binom{n}{k} := \frac{n!}{k!(n - k)!}$$

are the binomial coefficients. From here for the diagonal ($s = t$) Ramsey numbers

$$R(s, s) \leq \binom{2s - 2}{s - 1} \leq 2^{2s-2},$$

or, by using the Stirling formula

$$s! \sim \sqrt{2\pi s}\left(\frac{s}{e}\right)^s,$$

$$R(s, s) \leq \frac{2^{2s-2}}{\sqrt{\pi(s - 1)}}(1 + o(1)),$$

where $o(1)$ is any function that tends to zero when $s \to \infty$. These upper bounds for the Ramsey numbers show that they exist and their definition makes sense.

Problem 1.4.

1. Show that $R(3, 3) = 6$.
2. Show that $\binom{2n}{n} \leq 2^{2n}$.
3. Show that $R(s, t) \leq R(s - 1, t) + R(s, t - 1)$.
4. Show that if $R(s - 1, t)$ and $R(s, t - 1)$ are both even, then $R(s, t) \leq R(s - 1, t) + R(s, t - 1) - 1$.
5. Show that $R(s, t) \leq \binom{s+t-2}{s-1}$. Use the fact that $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$.

Problem 1.5.

1. Show that $1 + x \leq e^x$ for all x.
2. For $1 \leq k \leq n$
$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k.$$
3. Show for $1 \leq k \leq n$
$$\binom{n}{k} \leq \frac{n^k}{k!} \leq n^k.$$

To obtain a lower bound on the diagonal Ramsey number, we will need elementary information on *the classical probability model.* In short, the set of outcomes of an experiment is considered, I denote this set as $\Omega = \{\omega_1, \dots, \omega_k\}$, and ω_i, $i = 1, \dots, k$ are *outcomes.* This set satisfies three properties:

1. ω_i are pairwise disjoint (i.e., they cannot occur at the same time);
2. they form a full group of events (i.e., one of them always occurs);
3. $$\mathsf{P}(\{\omega_i\}) = \frac{1}{|\Omega|} = \frac{1}{k}\,,$$
   (i.e., they have equal chances to occur).

Now any *event* A associated with our experiment can be represented as a subset of Ω, $A \subseteq \Omega$, and its probability can be calculated (again, within the classical probability model) as

$$\mathsf{P}(A) = \sum_{\omega_i \in A} \mathsf{P}(\{\omega_i\}) = \frac{|A|}{|\Omega|}\,.$$

Here $\sum_{\omega_i \in A}$ means the sum throughout all the elements of the event (set) A.

Example 1.1. Consider an experiment with throwing a dice once. The outcomes are $\Omega = \{\omega_1, \dots, \omega_6\}$, where $\omega_i = \{i \text{ points were observed}\}$. Obviously $\mathsf{P}(\{\omega_i\}) = \frac{1}{6}$. Consider an event $A = \{\text{a prime number is observed}\}$. This event is $A = \{\omega_2, \omega_3, \omega_5\}$, and hence $\mathsf{P}(A) = \frac{3}{6} = \frac{1}{2}$.

The main result can be stated as

Theorem 1.2. *Let n and s be such that $\binom{n}{s} 2^{1-\binom{s}{2}} < 1$. Then $R(s,s) > n$.*

Before proving the theorem, consider a corollary, which gives a lower estimate for the diagonal Ramsey numbers. The notation $\lfloor q \rfloor$ means "the integer part of $q > 0$" or "the floor function."

Corollary 1.3. *$R(s,s) > \lfloor 2^{s/2} \rfloor$ for $s \geq 3$.*

Proof. Let $n = \lfloor 2^{s/2} \rfloor \leq 2^{s/2}$. Since

$$\binom{n}{s} = \frac{n!}{s!(n-s)!} = \frac{n(n-1)\dots(n-s+1)}{s!} \leq \frac{n^s}{s!} \leq \frac{2^{s^2/2}}{s!}\,,$$

one has

$$\binom{n}{s} 2^{1-\binom{s}{2}} \leq \frac{2^{s^2/2}}{s!} 2^{1-\frac{s(s-1)}{2}} = \frac{2^{1+s/2}}{s!} < 1,$$

because any exponent grows slower than the factorial. Hence, using the statement of the theorem above, we obtain the desired result. ■

Proof of Theorem 1.2. We need to show that $R(s,s) > n$, which means that we need to find a coloring of K_n into two colors such that any subgraph K_s would not have edges of the same color. Consider a random coloring of K_n. The total number of different colorings are $2^{\binom{n}{2}} =: N$, and these full graphs K_n with different colorings are our outcomes $\omega_1, \dots, \omega_N$. Hence, according to the classical probability model,

$$\mathsf{P}(\{\omega_i\}) = 2^{-\binom{n}{2}}.$$

Now consider an event A_S such that the full subgraph K_s on a fixed set of vertexes $S \subseteq V$ is either red or blue (i.e., all the edges either red or blue), and $s = |S|$. One has

$$\mathsf{P}(A_S) = \frac{2 \cdot 2^{\binom{n}{2} - \binom{s}{2}}}{2^{\binom{n}{2}}} = 2^{1-\binom{s}{2}},$$

here the numerator is obtained in the following way: Fix s vertices of our graph and assume that all the edges connecting these s vertices of the same color (blue or red, hence factor 2), there are $n-s$ vertices left, and there are $\binom{n}{2}-\binom{s}{2}$ edges about whose colors we did not make any assumptions. Hence we have $2^{\binom{n}{2}-\binom{s}{2}}$ coloring of these edges. Now to the main point. Consider a union of all possible events A_S:

$$\mathsf{P}\Big(\bigcup_{S\subseteq V} A_S\Big) \leq \sum_{S\subseteq V} \mathsf{P}(A_S) = \binom{n}{s} 2^{1-\binom{s}{2}} < 1$$

due to the assumption (I used $\mathsf{P}(A_1+A_2) \leq \mathsf{P}(A_1)+\mathsf{P}(A_2)$). Therefore, considering the complementary event $\overline{A} := \Omega \setminus A$ and using the property of probability that $\mathsf{P}(A)+\mathsf{P}(\overline{A}) = 1$ for any event A,

$$\mathsf{P}\Big(\overline{\bigcup_{S\subseteq V} A_S}\Big) > 0,$$

which literally means that there exists a coloring of K_n such that it is possible to find $S \subseteq V$ for which K_s on these S vertices is not mono-colored. ■

Hence we got two estimates (quite crude):

$$2^{s/2} < R(s,s) \leq 2^{2s}.$$

There are a lot of papers that improved these results, using quite sophisticated methods. But still there are a lot of open question. I.e., what is $R(5,5)$? No one knows[2].

[2]http://en.wikibooks.org/wiki/Combinatorics/Bounds_for_Ramsey_numbers