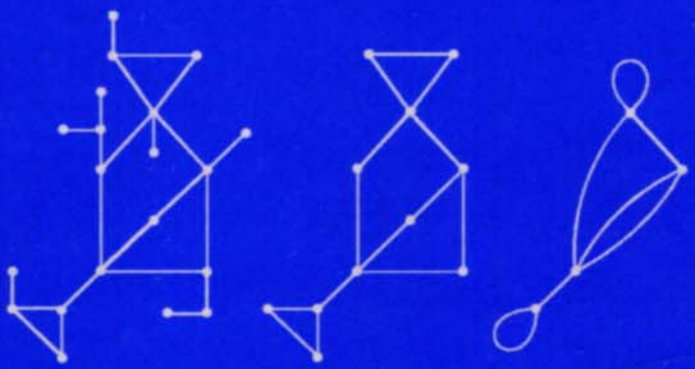
Wiley-Interscience Series in Discrete Mathematics and Optimization

RANDOM GRAPHS



SVANTE JANSON
TOMASZ ŁUCZAK
ANDRZEJ RUCIŃSKI

This page intentionally left blank

Random Graphs

WILEY-INTERSCIENCE
SERIES IN DISCRETE MATHEMATICS AND OPTIMIZATION

A complete list of titles in this series appears at the end of this volume.

Random Graphs

SVANTE JANSON
Uppsala University
Sweden

TOMASZ ŁUCZAK
Adam Mickiewicz University
Poland

ANDRZEJ RUCINSKI
Emory University
Atlanta, Georgia

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging-in-Publication Data is available.

ISBN 0-471-17541-2

10 9 8 7 6 5 4 3 2 1

Preface

The theory of random graphs originated in a series of papers published in the period 1959–1968 by two outstanding Hungarian mathematicians, Paul Erdős and Alfred Rényi. Over the forty years that have passed since then, the theory has developed into an independent and fast-growing branch of discrete mathematics, located at the intersection of graph theory, combinatorics and probability theory, with applications to theoretical computer science, reliability of transportation and communication networks, natural and social sciences and to discrete mathematics itself. Aside from applications, random graphs continue to serve as nontrivial, but simple enough models for other, more complex random structures, paving the road for more advanced theories.

In the early days, the literature on the subject was scattered around several probabilistic, combinatorial and general mathematics journals. In the late seventies, Béla Bollobás became the leading scientist in the field and contributed dozens of papers, which gradually made up a framework for his excellent, deep and extensive monograph *Random Graphs*, printed in 1985. The appearance of that book stimulated the research even further, shaping up a new theory.

Two other ingredients that added to this trend were the ongoing series of international conferences on random graphs and probabilistic methods in combinatorics held biennially in Poznań, Poland, since 1983, and the journal, *Random Structures and Algorithms*, launched by Wiley in 1990. Both have established a forum for the exchange of ideas and cooperation in the theory of random graphs and related fields.

It is not accidental then that tremendous progress has been made since 1985. Over the last decade several new, beautiful results have been proved and numerous fine techniques and methods have been introduced. Our goal is to present many of these new developments, including results on threshold functions (Ch. 1), small subgraphs (Ch. 3), generalized matchings (Ch. 4), phase transition (Ch. 5), limit distributions (Ch. 6), chromatic number (Ch. 7), partition and extremal properties (Ch. 8), Hamiltonian cycles in random regular graphs (Ch. 9), and zero-one laws (Ch. 10). We emphasize new techniques and tools such as the martingale, Talagrand and correlation inequalities (Ch. 2), the orthogonal decomposition (Ch. 6), the Regularity Lemma of Szemerédi (Ch. 8), the Contiguity Theorem (Ch. 9), and the analysis of variance (Ch. 9).

In a sense, our book can be viewed as an update on Bollobás's 1985 book. However, the topics selected for the book reflect the interest of its authors and do not pretend to exhaust the entire field. In fact, in order not to duplicate Bollobás's work, we do not include subjects which are covered there, on which only a little progress has been made. In particular, we have no sections on degree sequences, long paths and cycles, automorphisms, and the diameter. Moreover, we restrict ourselves to the main core of the theory and focus on the basic models of random graphs, making no attempt to present such rapidly developing areas as random walks on graphs, randomized algorithms or complexity of Boolean functions. Likewise, we exclude random cubes, directed graphs and percolation.

It has been our goal to make the book accessible to graduate students in mathematics and computer science. This has led to simplifications of some statements and proofs, which, we hope, result in better clarity of exposition. The book may be used as a textbook for a graduate course or an honors course for undergraduate senior mathematics and computer science majors. Although we do not provide problems and exercises separately, we often leave to the reader to complete parts of proofs or to provide proofs of results analogous to those proven. These instances, marked by the parenthetic phrase "(Exercise!)", can easily be picked up by the instructor and turned into homework assignments. The prerequisites are limited to basic courses in graph theory or combinatorics, elementary probability and calculus. We believe that the book will also be used by scientists working in the broad area of discrete mathematics and theoretical computer science. It is both an introduction for newcomers and a source of the most recent developments for those working in the field for many years.

We would like to thank several friends and colleagues, without whom this book would be *a.a.s.* worse than it is. Among those whose insightful remarks and suggestions led to improvements of earlier drafts are: Andrzej Czygrinow, Dwight Duffus, Ehud Friedgut, Johan Jonasson, Michał Karoński, Yoshiharu Kohayakawa, Michael Krivelevich, Justyna Kurkowiak, Jiří Matoušek, Brendan Nagle, Yuejian Peng, Joanna Polcyn, Vojtěch Rödl, Jozef Skokan, Joel Spencer, Edyta Szymańska, Michelle Wagner, and Julie White.

Special thanks are due to Penny Haxell and Izolda Gorgol. Penny spent several days correcting our English. Without her tedious work the text would probably need subtitles to be understood by an American reader. Izolda generously exercised her editing skills providing us with electronic files of all figures.

Jessica Downey, the Wiley editor, earned our deep appreciation for her continuous enthusiasm and support of the project.

Finally, the three authors would like to thank each other for patience, mutual encouragement and persistence in negotiations, the compromising effect of which is now in your hands.

SVANTE JANSON
TOMASZ ŁUCZAK
ANDRZEJ RUCIŃSKI

Uppsala, Poznań, and Atlanta

This page intentionally left blank

Contents

This page intentionally left blank

1

Preliminaries

1.1 MODELS OF RANDOM GRAPHS

The notion of a random graph originated in a paper of Erdős (1947), which is considered by some as the first conscious application of the probabilistic method. It was used there to prove the existence of a graph with a specific Ramsey property.

The model introduced by Erdős is very natural and can be described as choosing a graph at random, with equal probabilities, from the set of all $2^{\binom{n}{2}}$ graphs whose vertex set is $[n] = \{1, 2, \ldots, n\}$. In other words, it can be described as the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of all graphs with vertex set $[n]$, $\mathcal{F}$ is the family of all subsets of Ω, and for every $\omega \in \Omega$

$$\mathbb{P}(\omega) = 2^{-\binom{n}{2}}.$$

This probability space can also be viewed as the product of $\binom{n}{2}$ binary spaces. In simple words, it is a result of $\binom{n}{2}$ independent tosses of a fair coin, where "turning up heads" means "drawing an edge".

Generally speaking, a random graph is a graph constructed by a random procedure. In accordance with standard definitions in probability theory, this is formalized by representing the "random procedure" by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the "construction" by a function from the probability space into a suitable family of graphs. The *distribution* of a random graph is the induced probability distribution on the family of graphs; for many purposes this is the only relevant feature of the construction and we usually do not distinguish between different random graphs with the same distribution. Indeed, it is

often convenient to define a random graph by specifying its distribution; that is, we specify a family of graphs and a probability distribution on it. Note, however, that it is not sufficient to formally define a random graph as a probability distribution only, as is sometimes done in the literature; an important case in which this would not do is when several random graphs are considered at once, for example, in the two-round exposure described at the end of this section.

The word "model" is used rather loosely in the theory of random graphs. It may refer to a specific class of random graphs, defined as above, or perhaps to a specific distribution. Usually, however, there is also a parameter involved which measures the size of the graphs and typically it tends to infinity; there may also be other parameters. Needless to say, the whole theory of random graphs is thus asymptotic in its nature.

Two basic models

Nowadays, among several models of random graphs, there are two basic ones, the binomial model and the uniform model, both originating in the simple model introduced by Erdős (1947). In this book we will mainly restrict ourselves to studying these two models.

Given a real number p, $0 \leq p \leq 1$, the *binomial random graph*, denoted by $\mathbb{G}(n,p)$, is defined by taking as Ω the set of all graphs on vertex set $[n]$ and setting

$$\mathbb{P}(G) = p^{e_G}(1-p)^{\binom{n}{2}-e_G} ,$$

where $e_G = |E(G)|$ stands for the number of edges of G. It can be viewed as a result of $\binom{n}{2}$ independent coin flippings, one for each pair of vertices, with the probability of success (i.e., drawing an edge) equal to p. For $p = 1/2$ this is the model of 1947. However, most of the random graph literature is devoted to cases in which $p = p(n) \to 0$ as $n \to \infty$.

The binomial model is a special case of a *reliability network*. In this more general model, Ω is the family of all spanning subgraphs of a given graph F and $\mathbb{P}(G) = p^{e_G}(1-p)^{e_F-e_G}$. By a *spanning subgraph* we mean a graph G such that $V(G) = V(F)$ and $E(G) \subseteq E(F)$. Thus, in a reliability network, the edges of a given graph (network) are independently destroyed, each with failure probability $1-p$. One can generalize this model even further, by allowing different probabilities of failure at different edges. (Binomial models are sometimes called *Bernoulli*.)

Taking $F = K_n$, the complete graph on n vertices, we obtain the model $\mathbb{G}(n,p)$. Taking $F = K_{m,n}$, the complete bipartite graph (here either m is a function of n, or they are two independent parameters, typically both tending to infinity), we obtain the bipartite random graph $\mathbb{G}(m,n,p)$. Other popular models, not discussed here, are those in which the initial graph F is the hypercube or the $n \times n$ square lattice. The reliability network based

on the infinite square lattice belongs to percolation theory (Grimmett 1992a) which too, as all infinite models, is beyond the scope of this book.

The main advantage of the binomial model $\mathbb{G}(n,p)$ is the independence of presence of edges, but the drawback is that the number of edges is not fixed; it varies according to a binomial distribution with expectation $\binom{n}{2}p$. If one conditions on the event that $|E(\mathbb{G}(n,p))| = M$, then a uniform space is obtained. This space can be defined directly.

Given an integer M, $0 \le M \le \binom{n}{2}$, the ***uniform random graph***, denoted by $\mathbb{G}(n,M)$, is defined by taking as Ω the family of all graphs on the vertex set $[n]$ with exactly M edges, and as $\mathbb{P}$ the uniform probability on Ω,

$$\mathbb{P}(G) = \binom{\binom{n}{2}}{M}^{-1}, \qquad G \in \Omega.$$

This model, closely related to enumerative combinatorics, was apparently considered already in 1939 in an unpublished work of Erdős and Whitney on the connectedness of almost all graphs with n vertices and about $M = \frac{1}{2}n \log n$ edges. This was the model used throughout by Erdős and Rényi in their series of papers between 1959 and 1968, which gave rise to the theory of random graphs. (For an account of the contents of these eight fundamental papers, see Karoński and Ruciński (1997).)

The two basic models are in many cases asymptotically equivalent, provided $\binom{n}{2}p$ is close to M (see Section 1.4).

The uniform random graph $\mathbb{G}(n,M)$ belongs to a broad family of uniform random graphs defined by taking the uniform distribution over a family of graphs $\mathcal{F}$. The pioneering model from Erdős (1947) belongs here too, with $\mathcal{F}$ being the family of all graphs on a given set of vertices. Other popular models of this type are random trees (not studied in this book), where $\mathcal{F}$ is the family of all n^{n-2} trees on n labeled vertices, and random r-regular graphs (see Chapter 9), where $\mathcal{F}$ is the family of all graphs on n vertices of equal degree r, provided nr is even. We will use $\mathbb{G}(n,r)$ to denote a uniform random r-regular graph. It may look dangerous to use the notation $\mathbb{G}(n,p)$, $\mathbb{G}(n,M)$ and $\mathbb{G}(n,r)$ for three different things: What is $\mathbb{G}(n,1)$? In practice, however, the correct meaning is always clear from the context. (As for the three models: $\mathbb{G}(n,p)$ with $p = 1$, $\mathbb{G}(n,M)$ with $M = 1$, and $\mathbb{G}(n,r)$ with $r = 1$, each one is rather dull.)

Both the binomial and the uniform model have their counterparts for directed graphs. Besides these, there are interesting, natural random directed graphs which do not have analogues in the undirected case. Let us mention the k-out model, in which every vertex independently chooses k out-neighbors (including or excluding itself); the case of random mappings (i.e., $k = 1$) is well studied (Kolchin 1986, Aldous and Pitman 1994). Random tournaments, in which every edge of a complete graph assumes randomly one of the two possible orientations, have a broad literature too (Moon 1968, Gruszka, Łuczak and Ruciński 1996, Andersson 1998).

There are still other random graphs which do not fall into either category (binomial or uniform). For instance, in some reliability networks the vertices but not the edges are destroyed. Furthermore, some random graphs result from more complex probabilistic experiments, and here the sky is the limit. Restricted random graph processes constitute an interesting class of such experiments, but we should better define the unrestricted case first.

Random graph processes

In general, a *random graph process* is a stochastic process that describes a random graph evolving in time. In other words, it is a family $\{\mathbb{G}(t)\}_t$ of random graphs (defined on a common probability space) where the parameter t is interpreted as time; the time can be either discrete or continuous. The processes studied here will have a fixed vertex set (typically $[n]$), and they will start without any edges and grow monotonically by adding edges according to some rule but never deleting any.

A simple and important random graph process $\{\mathbb{G}(n,M)\}_M$ (sometimes called *the random graph process*) was introduced by Erdős and Rényi (1959) and has been well studied since then. It begins with no edges at time 0 and adds new edges, one at a time; each new edge is selected at random, uniformly among all edges not already present. Hence this random graph process is a Markov process, with time running through the set $\{0,1,\ldots,\binom{n}{2}\}$. The M-th stage of this process can be identified with the uniform random graph $\mathbb{G}(n,M)$. The process, however, allows one to study the random graph $\mathbb{G}(n,M)$ as it evolves with M growing from 0 to $\binom{n}{2}$. For example, a typical result, meaningful only for random graph processes, says that, with probability approaching 1 as $n \to \infty$, the very edge which links the last isolated vertex with another vertex makes the graph connected (Bollobás and Thomason (1985); see also Bollobás (1985)).

A related *continuous time random graph process* can be defined by assigning a random variable T_e to each edge e of the complete graph K_n, such that the $\binom{n}{2}$ variables T_e are independent with a common continuous distribution, and then defining the edge set of $\{\mathbb{G}(t)\}_t$ to consist of all e with $T_e \le t$. Clearly, the resulting random graph $\{\mathbb{G}(t)\}_{t_0}$ at a fixed time t_0 can be identified with the binomial random graph $\mathbb{G}(n,p)$, where $p = \mathbb{P}(T_e \le t_0)$. Furthermore, since almost surely no two values of the random variables T_e coincide, we may define $T_{(i)}$ as the random time at which the i-th edge is added. Then, by symmetry, $\mathbb{G}(T_{(i)})$ is the uniform random graph $\mathbb{G}(n,i)$, and the sequence $\{\mathbb{G}(T_{(i)})\}$ for $i = 1,\ldots,\binom{n}{2}$, equals the ordinary random graph process $\{\mathbb{G}(n,M)\}_M$ defined above. Hence, this continuous time random graph process is a joint generalization of the binomial random graph, the uniform random graph and the standard discrete time random graph process.

Clearly, different choices of the distribution of T_e affect the model only trivially, by a change in the time variable. The continuous time evolving

model was introduced by Stepanov (1970) with T_e exponentially distributed; we prefer the uniform distribution over the interval $[0, 1]$, in which case $p = \mathbb{P}(T_e \leq t) = t$, $0 \leq t \leq 1$. Thus, we may unambiguously use the notation $\{\mathbb{G}(n, t)\}_t$.

Recently, a number of *restricted random graph processes* have been studied. In general, such a process can be defined as a random graph process in which edges are chosen one by one uniformly from a dynamically modified set of available pairs of vertices until this set becomes empty. More formally, consider a Markov chain of random edge sets $E_0 = \emptyset, E_1, \ldots, E_s$, where $E_i = \{e_1, \ldots, e_i\}$ and e_i is chosen uniformly from a set A_i which depends only on the set E_{i-1}.

In one of these restricted models, studied by Ruciński and Wormald (1992), the maximum degree is bounded from above by a given integer d. Thus, the set A_i contains only those pairs whose addition to the set E_{i-1} does not create a vertex of degree $d + 1$. The graph at the end of the process may not be d-regular, though it is shown to be so with probability approaching 1. See also Wormald (1999a), where, moreover, further related processes are defined and studied.

Another restricted process is studied by Erdös, Suen and Winkler (1995), in which it is not allowed to create a triangle. In this model it is even an open problem to determine the length of a typical process, measured by the number of edges in the final graph. It is only known that with high probability the process takes more than $c_1 n^{3/2}$ but fewer than $c_2 n^{3/2} \log n$ steps, where c_1 and c_2 are positive constants. Recently, this result was generalized to a wide class of forbidden subgraphs by Osthus and Taraz (2000+).

By forbidding cycles, one obtains a process which creates a non-uniform random tree (Aldous 1990), while forbidding components with more than one cycle leads to a random graph which still is to be studied.

Random subsets

The two basic models of random graphs fall into the framework of random subsets of a set. Monotonicity, equivalence and threshold behavior of the probabilities of properties of random graphs can often be proved at no extra cost in this general setting. Other principal examples of random subsets of a set include random sets of integers and random hypergraphs. In the remaining sections of this chapter (as well as in parts of Chapter 2) we will mainly study this more general random set framework. For an arbitrary set X and an integer k, let $[X]^k$ stand for the family of all k-element subsets of X. If $X = [n]$, we will simplify this notation to $[n]^k$.

Let Γ be a finite set, $|\Gamma| = N$, let $0 \leq p \leq 1$ and $0 \leq M \leq N$. Then the random subset Γ_p of Γ is obtained by flipping a coin, with probability p of success, for each element of Γ to determine whether the element is to be included in Γ_p; the distribution of Γ_p is the probability distribution on $\Omega = 2^\Gamma$ given by $\mathbb{P}(F) = p^{|F|}(1-p)^{|\Gamma|-|F|}$ for $F \subseteq \Gamma$. Similarly, let Γ_M be

a randomly chosen element of $[\Gamma]^M$; that is, Γ_M has the uniform distribution $\mathbb{P}(F) = \binom{N}{M}^{-1}$ for $F \in [\Gamma]^M$.

Taking $\Gamma = [n]^2$ we obtain the two basic models of random graphs defined above, $\mathbb{G}(n,p)$ and $\mathbb{G}(n,M)$.

The binomial model Γ_p can be generalized to $\Gamma_{p_1,\dots,p_N}$, where the element i is included with probability p_i, independently for all $i = 1,\dots,N$.

Two-round exposure

The two-round exposure is a successful proof technique applicable to the binomial model. It relies on viewing Γ_p as a union of two independent random subsets Γ_{p_1} and Γ_{p_2}, where p_1 and p_2 are such that $p = p_1 + p_2 - p_1p_2$. (It is easy to see that this union indeed is distributed as Γ_p - Exercise!) In the special case of random graphs we first generate a random graph $\mathbb{G}(n,p_1)$ and then, independently, another random graph $\mathbb{G}(n,p_2)$ on the same vertex set. By replacing double edges by single ones, we obtain $\mathbb{G}(n,p)$.

An argument typically used in applications of the two-round exposure can be expressed in the following general form. Let $\mathbb{P}_1$ be the probability distribution associated with Γ_{p_1}, and let $\mathbb{P}_F$ be the conditional probability in Γ_p under the condition $\Gamma_{p_1} = F$. Then for any two families $\mathcal{A}$ and $\mathcal{B}$ of subsets of Γ

$$\mathbb{P}(\mathcal{A}) \geq \sum_{F \in \mathcal{B}} \mathbb{P}_F(\mathcal{A})\,\mathbb{P}_1(F) \geq \mathbb{P}_{F_0}(\mathcal{A})\,\mathbb{P}_1(\mathcal{B}), \tag{1.1}$$

where F_0 minimizes the probability $\mathbb{P}_F(\mathcal{A})$ over all $F \in \mathcal{B}$. Thus, knowing that $\mathbb{P}_1(\mathcal{B}) \to 1$, in order to prove that also $\mathbb{P}(\mathcal{A}) \to 1$, it is enough to show that $\mathbb{P}_F(\mathcal{A}) \to 1$, for every $F \in \mathcal{B}$. In practice, computing the last probability means fixing an instance of $\Gamma_{p_1} \in \mathcal{B}$ and throwing in new elements independently with probability p_2 (the second round of exposure).

1.2 NOTES ON NOTATION AND MORE

Graph theory

All graphs are simple and undirected, unless otherwise stated. We use standard notation for graphs. For example, $V(G)$ is the vertex set of a graph G, $E(G)$ is the edge set, $v_G = |V(G)|$ is the number of vertices and $e_G = |E(G)|$ is the number of edges; for typographical reasons we sometimes write the latter two as $v(G)$ and $e(G)$. In this book the *size* of G always means $v(G)$ (and not $e(G)$ as sometimes used by other authors). However, we also will call $v(G)$ the *order* of G.

Moreover, let $d(G) = e_G/v_G$ be the *density* and $m(G) = \max_{H \subseteq G} d(H)$ the *maximum density* of G. (Note that $d(G)$ equals half the average degree of G, and that some authors define $d(G)$ as the average degree, which is twice

our value.) Another measure of the density of a graph G, ranging between 0 and 1, is defined as $\rho(G) = e(G)/\binom{v(G)}{2}$. (It is sometimes called the *relative density* of G.)

Furthermore, $\delta(G)$ is the *minimum degree*, $\Delta(G)$ is the *maximum degree*, $\chi(G)$ is the *chromatic number*, $D(G) = \max_{H \subseteq G} \delta(H)$ is the *degeneracy number*, $\alpha(G)$ is the *stability number* (the size of the largest stable, or independent, set of vertices), and $\mathrm{aut}(G)$ is the number of automorphisms of G.

We let $N(v) = N_G(v)$ denote the *neighborhood* of a vertex v in G, that is, the set $\{w \in V(G) : vw \in E(G)\}$. Its size is called the *degree* of v and is denoted by $\deg(v) = \deg_G(v)$. Similarly, if $S \subseteq V(G)$, its neighborhood $N_G(S) = \bigcup_{v \in S} N_G(v) \setminus S$ is the set of all vertices outside S adjacent to at least one vertex in S. Moreover, we let $\overline{N}_G(v) = N_G(v) \cup \{v\}$ and $\overline{N}_G(S) = N_G(S) \cup S$ denote the corresponding *closed neighborhoods*, which include v and S, respectively.

Any graph without edges will be called *empty*, while the graph with no vertices (and thus no edges) will be called the *null graph* and denoted by $\emptyset$.

Some special graphs are: the complete graph K_n on n vertices, the complete bipartite graph $K_{m,n}$ on $m+n$ vertices, the cycle C_k with k vertices, and the path P_k with k edges and thus $k+1$ vertices. A star is any graph $K_{1,n}$, $n \geq 0$. We let jG denote the union of j vertex-disjoint copies of G. A matching is a forest consisting of isolated edges only (i.e., a graph of the form jK_2, $j \geq 0$).

If G is a graph and $V \subseteq V(G)$, then $G[V]$ denotes the restriction of G to V, defined as the graph with vertex set V and edge set $E(G) \cap [V]^2$; similarly, if $E \subseteq [V(G)]^2$, $G[E]$ denotes the graph with vertex set $V(G)$ and edge set $E(G) \cap E$. A subgraph of G of the type $G[V]$ is called *induced* or *spanned* by V, while a subgraph of the type $G[E]$ is called *spanning*. The number of edges in the subgraph $G[V]$ is sometimes denoted by $e_G(V) = e(V)$, while for two disjoint subsets $A, B \subset V(G)$, the quantity $e_G(A,B)$ counts the number of edges of G with one endpoint in A and the other in B.

By a *copy of a given graph* G inside another graph F we mean any, not necessarily induced, subgraph of F which is isomorphic to G. If the subgraph happens to be induced, we call it an *induced copy* of G.

Although we define our random graphs as labelled, we are mainly interested in properties that are independent of the labelling, that is, properties that depend on the isomorphism type only. Such properties are called *graph properties*. (In contrast, "vertex 1 is isolated" is not a graph property; such properties will occasionally be studied too.)

Probability

We use $\mathrm{Bi}(n,p)$, $\mathrm{Be}(p) = \mathrm{Bi}(1,p)$, $\mathrm{Po}(\lambda)$ and $\mathrm{N}(\mu,\sigma^2)$ to denote the binomial, Bernoulli, Poisson and normal distributions, respectively. We further write $X \in \mathcal{L}$, meaning that X is a random variable with distribution $\mathcal{L}$ (e.g., $X \in$

$\mathrm{N}(0,1)$). The distribution of a random variable X is occasionally denoted by $\mathcal{L}(X)$.

We denote by $\mathbf{1}[\mathcal{E}]$ the indicator function of the event $\mathcal{E}$, which equals 1 if $\mathcal{E}$ occurs and 0 otherwise. We will often consider random variables that are the indicator functions of some events; such random variables will be called *indicator* or *zero-one random variables*. They clearly have Bernoulli distributions with $p = \mathbb{P}(\mathcal{E})$, where $\mathcal{E}$ is the corresponding event.

The expected value and the variance of a random variable X (if they exist) will be denoted by $\mathbb{E}\,X$ and $\operatorname{Var} X$, respectively. Thus, the well-known *Chebyshev's inequality*, which will be frequently used throughout the book, can be stated in the following, standard form. If $\operatorname{Var} X$ exists, then

$$\mathbb{P}(|X - \mathbb{E}\,X| \geq t) \leq \frac{\operatorname{Var} X}{t^2}, \qquad t > 0. \tag{1.2}$$

Similarly, *Markov's inequality* states that, if $X \geq 0$ a.s., then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}\,X}{t}, \qquad t > 0. \tag{1.3}$$

We denote the covariance of two random variables X and Y by $\operatorname{Cov}(X, Y)$. Recall that the variance of a (finite) sum of random variables is given by $\operatorname{Var}(\sum_i X_i) = \sum_i \sum_j \operatorname{Cov}(X_i, X_j)$.

The conditional expectation of X given an event $\mathcal{E}$ is denoted by $\mathbb{E}(X \mid \mathcal{E})$. We similarly write $\mathbb{E}(X \mid Y_1, \ldots, Y_k)$ for the conditional expectation of X given some random variables $Y_1, \ldots, Y_k$; note that this conditional expectation is a function of $(Y_1, \ldots, Y_k)$ and thus itself a random variable. When using martingales (Section 2.4), we will more generally denote by $\mathbb{E}(X \mid \mathcal{G})$ the conditional expectation of X given a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$.

Quite frequently our proofs will rely on the elementary *law of total probability* which states that for any partition of the probability space $\Omega = \mathcal{E}_1 \cup \mathcal{E}_2 \ldots$ and any random variable X defined on Ω,

$$\mathbb{E}\,X = \sum_i \mathbb{E}(X \mid \mathcal{E}_i)\,\mathbb{P}(\mathcal{E}_i)\,.$$

In particular, if $X = \mathbf{1}[\mathcal{E}]$, then $\mathbb{P}(\mathcal{E}) = \sum_i \mathbb{P}(\mathcal{E} \mid \mathcal{E}_i)\,\mathbb{P}(\mathcal{E}_i)$.

If $X_1, X_2, \ldots$ are random variables and a is a constant, we say that X_n *converges in probability* to a as $n \to \infty$, and write $X_n \xrightarrow{p} a$, if $\mathbb{P}(|X_n - a| > \varepsilon) \to 0$ for every $\varepsilon > 0$; see, for example, Gut (1995, Chapter VI).

One similarly defines $X_n \xrightarrow{p} Y$, where Y is another random variable, but then Y and every X_n have to be defined on the same probability space; this can be reduced to the preceding case, since $X_n \xrightarrow{p} Y$ if and only if $X_n - Y \xrightarrow{p} 0$.

Let $X_1, X_2, \ldots$ and Z be random variables. We say that X_n *converges in distribution* to Z as $n \to \infty$, and write $X_n \xrightarrow{d} Z$, if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(Z \leq x)$ for every real x that is a continuity point of $\mathbb{P}(Z \leq x)$ (Billingsley 1968, Gut 1995).

If $X_1, X_2, \ldots$ and Z are integer-valued then, equivalently, $X_n \xrightarrow{d} Z$ if and only if $\mathbb{P}(X_n = k) \to \mathbb{P}(Z = k)$ for every integer k.

Note that convergence in distribution is really a property of the distributions of the random variables and does not require the variables to be defined on the same probability space. Nevertheless, it is customary (and convenient) to talk about convergence of random variables. We also use hybrid notation such as $X_n \xrightarrow{d} \mathrm{N}(0,1)$, which means $X_n \xrightarrow{d} Z$ for some (and thus every) random variable $Z \in \mathrm{N}(0,1)$.

An important special case is one in which Z is a (non-random) real constant. It is easily shown that convergence in distribution to a constant is the same as convergence in probability, that is, $X_n \xrightarrow{d} a$ if and only if $X_n \xrightarrow{p} a$ for $a \in \mathbb{R}$. A useful fact is that if $X_n \xrightarrow{d} Z$ and $Y_n \xrightarrow{p} a$, where a is a constant, then $X_n + Y_n \xrightarrow{d} Z + a$ and $Y_n X_n \xrightarrow{d} aZ$ (*Cramér's theorem*), see, for example, Gut (1995, Theorem VI.7.5).

The definition of convergence in distribution extends to random vectors with values in $\mathbb{R}^k$ for every fixed k; this is also expressed as *joint convergence in distribution* of the components of the vectors. A powerful method for extending results on the real random variables to the vector-valued ones is known as the *Cramér–Wold device* (Billingsley 1968, Theorem 7.7). It states that $(X_{n1}, \ldots, X_{nk}) \xrightarrow{d} (Z_1, \ldots, Z_k)$ if and only if $\sum_i t_i X_{ni} \xrightarrow{d} \sum_i t_i Z_i$ for every sequence of real numbers $t_1, \ldots, t_k$. For more details, as well as for the convergence of random variables with values in even more general spaces, see Billingsley (1968).

Remark 1.1. Convergence in distribution does *not*, in general, imply convergence of the sequence of means or variances. However, in many specific applications we find that these sequences do, in fact, converge to the mean and variance of the limit distribution.

Asymptotics

We will often use the following standard notation for the asymptotic behavior of the relative order of magnitude of two sequences of numbers a_n and b_n, depending on a parameter $n \to \infty$. The same notation is also used in other situations, for example, for functions of a variable ε that tends to 0. We will often omit the phrase "as $n \to \infty$" when there is no risk of confusion. For simplicity we assume $b_n > 0$ for all sufficiently large n.

- $a_n = O(b_n)$ as $n \to \infty$ if there exist constants C and n_0 such that $|a_n| \le C b_n$ for $n \ge n_0$, i.e., if the sequence a_n/b_n is bounded, except possibly for some small values of n for which the ratio may be undefined.

- $a_n = \Omega(b_n)$ as $n \to \infty$ if there exist constants $c > 0$ and n_0 such that $a_n \ge c b_n$ for $n \ge n_0$. If $a_n \ge 0$, this is equivalent to $b_n = O(a_n)$.

- $a_n = \Theta(b_n)$ as $n \to \infty$ if there exist constants $C, c > 0$ and n_0 such that $cb_n \le a_n \le Cb_n$ for $n \ge n_0$, i.e., if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$. This is sometimes expressed by saying that a_n and b_n are of the *same order of magnitude.*
- $a_n \asymp b_n$ if $a_n = \Theta(b_n)$.
- $a_n \sim b_n$ if $a_n/b_n \to 1$.
- $a_n = o(b_n)$ as $n \to \infty$ if $a_n/b_n \to 0$, i.e., if for every $\varepsilon > 0$ there exists n_ε such that $|a_n| < \varepsilon b_n$ for $n \ge n_\varepsilon$.
- $a_n \ll b_n$ or $b_n \gg a_n$ if $a_n \ge 0$ and $a_n \asymp o(b_n)$.

Since most results in this book are asymptotic, we will be frequently assuming in the proofs that n is sufficiently large, sometimes without explicitly saying so.

Probability asymptotics

We say that an event $\mathcal{E}_n$, describing a property of a random structure depending on a parameter n, holds *asymptotically almost surely* (abbreviated *a.a.s.*), if $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$.

Remark 1.2. In many publications on random structures the phrase "almost surely" or a.s. is used. However, we wish to reserve that phrase for what it normally means in probability theory, i.e. that the probability of an event equals exactly 1. It seems that the first paper where the phrase *a.a.s.* and not *a.s.* was used is Shamir and Upfal (1981). (Some authors use the phrase "almost every" or a.e. which we reject for the same reason as "almost surely". Others write "with high probability", or **whp**.)

When discussing asymptotics of random variables, we avoid expressions like "$X_n = O(1)$ *a.a.s.*" or "$X_n = o(1)$ *a.a.s.*", which may be ambiguous, since they combine two asymptotic notions. As a substitute we give probabilistic versions of some of the symbols above, denoting them with a subscript p or C. Let X_n be random variables and a_n positive real numbers. We then define:

- $X_n = O_p(a_n)$ as $n \to \infty$ if for every $\delta > 0$ there exist constants C_δ and n_0 such that $\mathbb{P}(|X_n| \le C_\delta a_n) > 1 - \delta$ for every $n \ge n_0$.
- $X_n = O_C(a_n)$ as $n \to \infty$ if there exists a constant C such that *a.a.s.* $|X_n| \le Ca_n$.
- $X_n = \Theta_p(a_n)$ as $n \to \infty$ if for every $\delta > 0$ there exist constants $c_\delta > 0$, $C_\delta > 0$ and n_0 such that $\mathbb{P}(c_\delta a_n \le X_n \le C_\delta a_n) > 1 - \delta$ for every $n \ge n_0$.
- $X_n = \Theta_C(a_n)$ as $n \to \infty$ if there exist positive constants c and C such that *a.a.s.* $ca_n \le X_n \le Ca_n$.

- $X_n = o_p(a_n)$ as $n \to \infty$ if for every $\varepsilon > 0$, *a.a.s.* $|X_n| < \varepsilon a_n$.

Note that $X_n = O_C(a_n)$ implies $X_n = O_p(a_n)$, but not conversely; indeed, $X_n = O_C(a_n)$ if and only if the constant C_δ in the definition of O_p can be chosen independently of δ. For example, any sequence X_n of identically distributed random variables is $O_p(1)$, but such a sequence is $O_C(1)$ only if the common distribution has support in a finite interval.

Similarly, $X_n = \Theta_C(a_n)$ implies $X_n = \Theta_p(a_n)$, but not conversely. On the other hand, $X_n = o_p(a_n)$ implies $X_n = O_C(a_n)$.

Remark 1.3. It is easy to verify (Exercise!) that $X_n = O_p(a_n)$ if and only if for every function $\omega(n) \to \infty$, $|X_n| \le \omega(n)a_n$ *a.a.s.* Similarly, $X_n = o_p(a_n)$ if and only if for some function $\omega(n) \to \infty$, $|X_n| \le a_n/\omega(n)$ *a.a.s.*

Such notation with an unspecified sequence $\omega(n)$ is common in publications on random structures, but we believe that the equivalent notation O_p and o_p is clearer.

It is an immediate consequence of the definitions (Exercise!) that $X_n = o_p(a_n)$ if and only if $X_n/a_n \xrightarrow{p} 0$. Conversely, $X_n \xrightarrow{p} a$ if and only if $X_n = a + o_p(1)$ (and $X_n \xrightarrow{p} Y$ if and only if $X_n = Y + o_p(1)$).

Remark 1.4. The symbol O_p can also be expressed by equivalent standard probabilistic concepts. In fact, a sequence X_n is *bounded in probability*, or *tight*, if $X_n = O_p(1)$. Hence, $X_n = O_p(a_n)$ if and only if the sequence X_n/a_n is bounded in probability (or tight).

Dependency graphs

Let $\{X_i\}_{i\in\mathcal{I}}$ be a family of random variables (defined on a common probability space). A *dependency graph* for $\{X_i\}$ is any graph L with vertex set $V(L) = \mathcal{I}$ such that if A and B are two disjoint subsets of $\mathcal{I}$ with $e_L(A,B) = 0$, then the families $\{X_i\}_{i\in A}$ and $\{X_i\}_{i\in B}$ are mutually independent.

Dependency graphs will be used several times in this book. They are particularly useful when they are sparse, meaning that there is a lot of independence in the family $\{X_i\}$.

Example 1.5. In a standard situation, there is an underlying family of independent random variables $\{Y_\alpha\}_{\alpha\in\mathcal{A}}$, and each X_i is a function of the variables $\{Y_\alpha\}_{\alpha\in A_i}$ for some subset $A_i \subseteq \mathcal{A}$. Let $\mathcal{S} = \{A_i : i \in \mathcal{I}\}$. Then the graph $L = L(\mathcal{S})$ with vertex set $\mathcal{I}$ and edge set $\{ij : A_i \cap A_j \neq \emptyset\}$ is a dependency graph for the family $\{X_i\}_{i\in\mathcal{I}}$ (Exercise!).

Example 1.6. As a special case of the preceding example, let $\{H_i\}_{i\in\mathcal{I}}$ be given subgraphs of the complete graph K_n and let X_i be the indicator that H_i appears as a subgraph in $\mathbb{G}(n,p)$, that is, $X_i = \mathbf{1}[H_i \subseteq \mathbb{G}(n,p)]$, $i \in \mathcal{I}$. Then $L(\mathcal{S})$, with $\mathcal{S} = \{E(H_i) : i \in \mathcal{I}\}$, is a natural dependency graph with edge set $\{ij : E(H_i) \cap E(H_j) \neq \emptyset\}$ (Exercise!).

Remark 1.7. In particular, if L is a dependency graph for $\{X_i\}$, then two variables X_i and X_j are independent unless there is an edge in L between i and j. Note, however, that this is only a necessary condition, and does not imply that L is a dependency graph (Exercise!).

Remark 1.8. Another context, outside the scope of this book, in which dependency graphs are used is the Lovász Local Lemma (Erdős and Lovász (1975); see also Alon and Spencer (1992)). There it actually suffices to use a slightly weaker definition, considering only singletons B in the definition above.

Remark 1.9. In our applications, there exists a natural dependency graph, but it should be observed that, in general, there is no canonical choice and the dependency graph is not unique, even if it is required to be minimal (Exercise!).

The subsubsequence principle

It is often convenient to use the well-known *subsubsequence principle*, which states that if for every subsequence of a sequence there is a subsubsequence converging to a limit a, then the entire sequence must converge to the same limit. This holds for sequences of real numbers, vectors, random variables (both for convergence in probability and for convergence in distribution) and, in general, for sequences in any topological space.

For example, this means that if we want to prove a limit theorem for $\mathbb{G}(n,p)$, we may without loss of generality assume that an expression such as $n^a p^b$ converges to some $c \leq \infty$ (provided, of course, that the result we want to prove does not depend on the limit c).

We will be using this principle throughout the book (see, e.g., the proof of Proposition 1.15), sometimes without explicitly mentioning it.

And finally ...

The base of all logarithms is e, unless specified otherwise.

1.3 MONOTONICITY

A family of subsets $\mathcal{Q} \subseteq 2^\Gamma$ is called *increasing* if $A \subseteq B$ and $A \in \mathcal{Q}$ imply that $B \in \mathcal{Q}$. A family of subsets is *decreasing* if its complement in 2^Γ is increasing, or, equivalently, if the family of the complements in Γ is increasing. A family which is either increasing or decreasing is called *monotone*. A family $\mathcal{Q}$ is *convex* if $A \subseteq B \subseteq C$ and $A, C \in \mathcal{Q}$ imply $B \in \mathcal{Q}$. We identify properties of subsets of Γ with the corresponding families of all subsets having the property; we thus use the same notation and terminology for properties.

In the special case in which $\Gamma = [n]^2$, any family $\mathcal{Q} \subseteq 2^\Gamma$ is a family of graphs and, if it is closed under isomorphism, it can be identified with a graph property. Some examples of increasing graph properties are "being connected", "containing a triangle" and "having a perfect matching". Automatically, the negations of all of them are decreasing. Natural decreasing graph properties include "having at least k isolated vertices", "having at most k edges" and "being planar". The property of "having exactly k isolated vertices" is an example of a convex but not monotone property, whereas "the largest component is a tree" is not even convex (Exercise!).

It is reasonable to expect that the probability of a random set falling into an increasing family of sets gets larger when the (expected) size of the random set does. This is indeed the case. Lemma 1.10 below appeared first in Bollobás (1979).

Lemma 1.10. *Let $\mathcal{Q}$ be an increasing property of subsets of Γ, $0 \le p_1 \le p_2 \le 1$ and $0 \le M_1 \le M_2 \le N$. Then*

$$\mathbb{P}(\Gamma_{p_1} \in \mathcal{Q}) \le \mathbb{P}(\Gamma_{p_2} \in \mathcal{Q})$$

and

$$\mathbb{P}(\Gamma_{M_1} \in \mathcal{Q}) \le \mathbb{P}(\Gamma_{M_2} \in \mathcal{Q}).$$

Proof. To prove the first inequality we employ a simple version of the two-round exposure technique (see Section 1.1). Let $p_0 = (p_2 - p_1)/(1 - p_1)$. Then Γ_{p_2} can be viewed as a union of two independent random subsets, Γ_{p_1} and Γ_{p_0}. As then $\Gamma_{p_1} \subseteq \Gamma_{p_2}$ and $\mathcal{Q}$ is increasing, the event "$\Gamma_{p_1} \in \mathcal{Q}$" implies the event "$\Gamma_{p_2} \in \mathcal{Q}$", completing the proof of first inequality.

For the second inequality, we construct a *random subset process* $\{\Gamma_M\}_M$, similar to the random graph process defined in Section 1.1, by selecting the elements of Γ one by one in random order. Clearly, Γ_M can be taken as the M-th subset in the process. Then $\Gamma_{M_1} \subseteq \Gamma_{M_2}$, and, as in the first part, the event "$\Gamma_{M_1} \in \mathcal{Q}$" implies the event "$\Gamma_{M_2} \in \mathcal{Q}$", which completes the proof. ■

Trivially, each monotone property is convex. In a special case this can be, in a sense, reversed: if $\mathcal{Q}$ is convex, and for some M we have $[\Gamma]^M \subseteq \mathcal{Q}$ then, for $M' < M$, $\mathcal{Q}$ behaves like an increasing property, and in particular $\mathbb{P}(\Gamma_{M'} \in \mathcal{Q}) \le \mathbb{P}(\Gamma_{M''} \in \mathcal{Q})$ for all $M' \le M'' \le M$ (Exercise!). Similarly, for $M'' > M$, $\mathcal{Q}$ can be treated as decreasing. A probabilistic version of this simple observation is stated in the next lemma.

Lemma 1.11. *Let $\mathcal{Q}$ be a convex property of subsets of Γ, and let M_1, M, M_2 be three integer functions of N satisfying $0 \le M_1 \le M \le M_2 \le N$. Then*

$$\mathbb{P}(\Gamma_M \in \mathcal{Q}) \ge \mathbb{P}(\Gamma_{M_1} \in \mathcal{Q}) + \mathbb{P}(\Gamma_{M_2} \in \mathcal{Q}) - 1.$$

Hence, if $\mathbb{P}(\Gamma_{M_2} \in \mathcal{Q}) \to 1$ as $N \to \infty$, then $\mathbb{P}(\Gamma_{M_1} \in \mathcal{Q}) \le \mathbb{P}(\Gamma_M \in \mathcal{Q}) + o(1)$. In particular, if $\mathbb{P}(\Gamma_{M_i} \in \mathcal{Q}) \to 1$ as $N \to \infty$, $i = 1, 2$, then $\mathbb{P}(\Gamma_M \in \mathcal{Q}) \to 1$.

Proof. The following simple proof was observed by Johan Jonasson (personal communication). It is easily seen that a convex property Q is the intersection of an increasing property Q' and a decreasing property Q''. (Exercise! - Note that the converse is obvious.) Thus

$$\begin{aligned}\mathbb{P}(\Gamma_M \in Q) &\geq \mathbb{P}(\Gamma_M \in Q') + \mathbb{P}(\Gamma_M \in Q'') - 1 \\ &\geq \mathbb{P}(\Gamma_{M_1} \in Q') + \mathbb{P}(\Gamma_{M_2} \in Q'') - 1 \\ &\geq \mathbb{P}(\Gamma_{M_1} \in Q) + \mathbb{P}(\Gamma_{M_2} \in Q) - 1. \end{aligned}$$

■

1.4 ASYMPTOTIC EQUIVALENCE

In this section we examine the *asymptotic equivalence* of the two models Γ_p and Γ_M; recall that this includes the random graphs $\mathbb{G}(n,p)$ and $\mathbb{G}(n,M)$ as a special case. Our goal is to establish conditions under which convergence of $\mathbb{P}(\Gamma_p \in Q)$ implies convergence of $\mathbb{P}(\Gamma_M \in Q)$ to the same limit and vice versa. One expects such equivalence when M is near Np. Since Γ_p is a mixture of Γ_M's for different M, the above implication is more straightforward in the direction from the uniform to the binomial model and then does not require any restriction on Q. The only tools we use are the elementary law of total probability and Chebyshev's inequality. Most results in this section are based on Łuczak (1990a); in the case in which the limit is one they already appeared in Bollobás (1979, 1985).

Let $\Gamma(n)$ be a sequence of sets of size $N(n) = |\Gamma(n)| \to \infty$. (In the example of main concern to us, *viz.* random graphs, $\Gamma(n) = [n]^2$ and thus $N(n) = \binom{n}{2}$.) We further consider a property Q of subsets of these sets; formally the property corresponds to a sequence $Q(n) \subseteq 2^{\Gamma(n)}$ of families of subsets of $\Gamma(n)$, $n = 1, 2, \ldots$. Finally, $p(n)$ is a given sequence of real numbers with $0 \leq p(n) \leq 1$, and $M(n)$ is a sequence of integers with $0 \leq M(n) \leq N(n)$. We usually omit the argument n and write Γ, N, Q, p and M; moreover, we let $q = 1 - p$.

Proposition 1.12. *Let Q be an arbitrary property of subsets of $\Gamma = \Gamma(n)$ as above, $p = p(n) \in [0,1]$ and $0 \leq a \leq 1$. If for every sequence $M = M(n)$ such that $M = Np + O(\sqrt{Npq})$ it holds that $\mathbb{P}(\Gamma_M \in Q) \to a$ as $n \to \infty$, then also $\mathbb{P}(\Gamma_p \in Q) \to a$ as $n \to \infty$.*

Proof. Let C be a large constant and define (for each n)

$$\mathcal{M}(C) = \{M : |M - Np| \leq C\sqrt{Npq}\}.$$

Furthermore, let M_* be the element M of $\mathcal{M}(C)$ that minimizes $\mathbb{P}(\Gamma_M \in \mathcal{Q})$. By the law of total probability,

$$\begin{aligned}\mathbb{P}(\Gamma_p \in \mathcal{Q}) &= \sum_{M=0}^{N} \mathbb{P}(\Gamma_p \in \mathcal{Q} \mid |\Gamma_p| = M)\, \mathbb{P}(|\Gamma_p| = M) \\ &= \sum_{M=0}^{N} \mathbb{P}(\Gamma_M \in \mathcal{Q})\, \mathbb{P}(|\Gamma_p| = M) \\ &\geq \sum_{M \in \mathcal{M}(C)} \mathbb{P}(\Gamma_{M_*} \in \mathcal{Q})\, \mathbb{P}(|\Gamma_p| = M) \\ &= \mathbb{P}(\Gamma_{M_*} \in \mathcal{Q})\, \mathbb{P}\big(|\Gamma_p| \in \mathcal{M}(C)\big).\end{aligned}$$

By assumption, $\mathbb{P}(\Gamma_{M_*} \in \mathcal{Q}) \to a$, and using Chebyshev's inequality (1.2), we also have $\mathbb{P}\big(|\Gamma_p| \notin \mathcal{M}(C)\big) \leq \operatorname{Var}|\Gamma_p| / \big(C\sqrt{Npq}\big)^2 = 1/C^2$. Consequently,

$$\liminf_{n\to\infty} \mathbb{P}(\Gamma_p \in \mathcal{Q}) \geq a \liminf_{n\to\infty} \mathbb{P}\big(|\Gamma_p| \in \mathcal{M}(C)\big) \geq a(1 - C^{-2}).$$

Similarly, if M^* maximizes $\mathbb{P}(\Gamma_M \in \mathcal{Q})$ among $M \in \mathcal{M}(C)$,

$$\mathbb{P}(\Gamma_p \in \mathcal{Q}) \leq \mathbb{P}(\Gamma_{M^*} \in \mathcal{Q}) + \mathbb{P}\big(|\Gamma_p| \notin \mathcal{M}(C)\big) \leq \mathbb{P}(\Gamma_{M^*} \in \mathcal{Q}) + C^{-2},$$

and

$$\limsup_{n\to\infty} \mathbb{P}(\Gamma_p \in \mathcal{Q}) \leq a + C^{-2}.$$

The result follows by letting $C \to \infty$. ∎

In the other direction no asymptotic equivalence can be true in such generality. The property of containing exactly M edges serves as a simplest counterexample (Exercise!). However, the additional assumption of monotonicity of $\mathcal{Q}$ suffices.

Proposition 1.13. *Let $\mathcal{Q}$ be a monotone property of subsets of $\Gamma = \Gamma(n)$ as above, $0 \leq M \leq N$ and $0 \leq a \leq 1$. If for every sequence $p = p(n) \in [0,1]$ such that $p = M/N + O\big(\sqrt{M(N-M)/N^3}\big)$ it holds that $\mathbb{P}(\Gamma_p \in \mathcal{Q}) \to a$, then $\mathbb{P}(\Gamma_M \in \mathcal{Q}) \to a$ as $n \to \infty$.*

Proof. We consider only the case in which $\mathcal{Q}$ is increasing (the decreasing case is similar). Let C be a large constant, $p_0 = M/N$, $q_0 = 1 - p_0$, and define $p_+ = \min(p_0 + C\sqrt{p_0 q_0/N}, 1)$ and $p_- = \max(p_0 - C\sqrt{p_0 q_0/N}, 0)$. Arguing as in the proof of Proposition 1.12 and using Lemma 1.10, we have

$$\begin{aligned}\mathbb{P}(\Gamma_{p_+} \in \mathcal{Q}) &\geq \sum_{M' \geq M} \mathbb{P}(\Gamma_{M'} \in \mathcal{Q})\, \mathbb{P}(|\Gamma_{p_+}| = M') \\ &\geq \mathbb{P}(\Gamma_M \in \mathcal{Q})\, \mathbb{P}(|\Gamma_{p_+}| \geq M) \\ &\geq \mathbb{P}(\Gamma_M \in \mathcal{Q}) - \mathbb{P}(|\Gamma_{p_+}| < M) \qquad (1.4)\end{aligned}$$

and similarly

$$\mathbb{P}(\Gamma_{p_-} \in \mathcal{Q}) \leq \mathbb{P}(\Gamma_M \in \mathcal{Q}) + \mathbb{P}(|\Gamma_{p_-}| > M). \tag{1.5}$$

The cases $M = 0$ and $M = N$ are trivial (Exercise!), so we may further assume $1 \leq M \leq N-1$, and thus $Np_0q_0 = M(N-M)/N \geq 1/2$. Since $|\Gamma_{p_-}|$ has the binomial distribution with mean Np_- and variance

$$Np_-(1-p_-) \leq M(1 - p_0 + C\sqrt{p_0q_0/N}) \leq Np_0q_0 + C\sqrt{Np_0q_0},$$

Chebyshev's inequality (1.2) yields, with $\delta(C) = C^{-2} + \sqrt{2}C^{-1}$,

$$\mathbb{P}(|\Gamma_{p_-}| > M) \leq \frac{Np_-(1-p_-)}{(Np_0 - Np_-)^2} \leq \frac{Np_0q_0 + C\sqrt{Np_0q_0}}{C^2Np_0q_0} \leq \delta(C),$$

and similarly $\mathbb{P}(|\Gamma_{p_+}| < M) \leq \delta(C)$. Since

$$\lim_{n\to\infty} \mathbb{P}(\Gamma_{p_+} \in \mathcal{Q}) = \lim_{n\to\infty} \mathbb{P}(\Gamma_{p_-} \in \mathcal{Q}) = a$$

by assumption, the inequalities (1.4) and (1.5) yield

$$a - \delta(C) \leq \liminf_{n\to\infty} \mathbb{P}(\Gamma_M \in \mathcal{Q}) \leq \limsup_{n\to\infty} \mathbb{P}(\Gamma_M \in \mathcal{Q}) \leq a + \delta(C),$$

and the result follows by letting $C \to \infty$, which implies $\delta(C) \to 0$. ∎

Remark 1.14. In the above proof one can relax the monotonicity of $\mathcal{Q}$ and instead require only that in the range $M' = M + O(\sqrt{M(N-M)/N})$

$$\mathbb{P}(\Gamma_{M'} \in \mathcal{Q}) \leq \mathbb{P}(\Gamma_M \in \mathcal{Q}) + o(1)$$

for $M' \leq M$, and

$$\mathbb{P}(\Gamma_{M'} \in \mathcal{Q}) \geq \mathbb{P}(\Gamma_M \in \mathcal{Q}) + o(1)$$

for $M' \geq M$. By Lemma 1.11, these conditions are satisfied whenever $\mathcal{Q}$ is convex and for some M' with $M' - M \gg \sqrt{M(N-M)/N}$, it holds that $\lim_{n\to\infty} \mathbb{P}(\Gamma_{M'} \in \mathcal{Q}) = 1$ (Exercise!).

The next result simplifies Proposition 1.13 for $a = 1$ by showing that for convex properties $\mathcal{Q}$, we have *a.a.s.* $\Gamma_M \in \mathcal{Q}$ provided *a.a.s.* $\Gamma_{M/N} \in \mathcal{Q}$.

Proposition 1.15. *Let $\mathcal{Q}$ be a convex property of subsets of Γ and let $0 \leq M \leq N$. If $\mathbb{P}(\Gamma_{M/N} \in \mathcal{Q}) \to 1$ as $n \to \infty$, then $\mathbb{P}(\Gamma_M \in \mathcal{Q}) \to 1$.*

Proof. We assume for simplicity that $M(N-M)/N \to \infty$, leaving the cases in which M or $N - M$ is bounded to the reader (Exercise!). (Note that the subsubsequence principle implies that it suffices to consider these three cases - Exercise!)