Preface

Within a couple of months in 2003, in the Theory Group of Microsoft Research in Redmond, Washington, three questions were asked by three colleagues. Michael Freedman, who was working on some very interesting ideas to design a quantum computer based on methods of algebraic topology, wanted to know which graph parameters (functions on finite graphs) can be represented as partition functions of models from statistical physics. Jennifer Chayes, who was studying internet models, asked whether there was a notion of "limit distribution" for sequences of graphs (rather than for sequences of numbers). Vera T. Sós, a visitor from Budapest interested in the phenomenon of quasirandomness and its connections to the Regularity Lemma, suggested to generalize results about quasirandom graphs to multitype quasirandom graphs. It turned out that these questions were very closely related, and the ideas which we developed for the answers have motivated much of my research for the next years.

Jennifer's question recalled some old results of mine characterizing graphs through homomorphism numbers, and another paper with Paul Erdős and Joel Spencer in which we studied normalized versions of homomorphism numbers and their limits. Using homomorphism numbers, Mike Freedman, Lex Schrijver and I found the answer to Mike's question in a few months. The method of solution, the use of graph algebras, provided a tool to answer Vera's. With Christian Borgs, Jennifer Chayes, Lex Schrijver, Vera Sós, Balázs Szegedy, and Kati Vesztergombi, we started to work out an algebraic theory of graph homomorphisms and an analytic theory of convergence of graph sequences and their limits. This book will try to give an account of where we stand.

Finding unexpected connections between the three questions above was stimulating and interesting, but soon we discovered that these methods and results are connected to many other studies in many branches of mathematics. A couple of years earlier Itai Benjamini and Oded Schramm had defined convergence of graph sequences with bounded degree, and constructed limit objects for them (our main interest was, at least initially, the convergence theory of dense graphs). Similar ideas were raised even earlier by David Aldous. The limit theories of dense and bounded-degree graphs have lead to many analogous questions and results, and each of them is better understood thanks to the other.

Statistical physics deals with very large graphs and their local and global properties, and it turned out to be extremely fruitful to have two statistical physicists (Jennifer and Christian) on the (informal) team along with graph theorists. This put the burden to understand the other person's goals and approaches on all of us, but at the end it was the key to many of the results.

Another important connection that was soon discovered was the theory of property testing in computer science, initiated by Goldreich, Goldwasser and Ron several years earlier. This can be viewed as statistics done on graphs rather than on numbers, and probability and statistics became a major tool for us.

One of the most important application areas of these results is extremal graph theory. A fundamental tool in the extremal theory of dense graphs is Szemerédi's Regularity Lemma, and this lemma turned out to be crucial for us as well. Graph limit theory, we hope, repaid some of this debt, by providing the shortest and most general formulation of the Regularity Lemma ("compactness of the graphon space"). Perhaps the most exciting consequence of the new theory is that it allows the precise formulation of, and often the exact answer to, some very general questions concerning algorithms on large graphs and extremal graph theory. Independently and about the same time as we did, Razborov developed the closely related theory of flag algebras, which has lead to the solution of several long-standing open problems in extremal graph theory.

Speaking about limits means, of course, analysis, and for some of us graph theorists, it meant hard work learning the necessary analytical tools (mostly measure theory and functional analysis, but even a bit of differential equations). Involving analysis has advantages even for some of the results that can be stated and proved purely graph-theoretically: many definitions and proofs are shorter, more transparent in the analytic language. Of course, combinatorial difficulties don't just disappear: sometimes they are replaced by analytic difficulties. Several of these are of a technical nature: Are the sets we consider Lebesgue/Borel measurable? In a definition involving an infimum, is it attained? Often this is not really relevant for the development of the theory. Quite often, on the other hand, measurability carries combinatorial meaning, which makes this relationship truly exciting.

There were some interesting connections with algebra too. Balázs Szegedy solved a problem that arose as a dual to the characterization of homomorphism functions, and through his proof he established, among others, a deep connection with the representation theory of algebras. This connection was later further developed by Schrijver and others. Another one of these generalizations has lead to a combinatorial theory of categories, which, apart from some sporadic results, has not been studied before. The limit theory of bounded degree graphs also found very strong connections to algebra: finitely generated infinite groups yield, through their Cayley graphs, infinite bounded degree graphs, and representing these as limits of finite graphs has been studied in group theory (under the name of sofic groups) earlier.

These connections with very different parts of mathematics made it quite difficult to write this book in a readable form. One way out could have been to focus on graph theory, not to talk about issues whose motivation comes from outside graph theory, and sketch or omit proofs that rely on substantial mathematical tools from other parts. I felt that such an approach would hide what I found the most exciting feature of this theory, namely its rich connections with other parts of mathematics (classical and non-classical). So I decided to explain as many of these connections as I could fit in the book; the reader will probably skip several parts if he/she does not like them or does not have the appropriate background, but perhaps the flavor of these parts can be remembered.

The book has five main parts. First, an informal introduction to the mathematical challenges provided by large networks. We ask the "general questions" mentioned above, and try to give an informal answer, using relatively elementary mathematics, and motivating the need for those more advanced methods that are developed in the rest of the book.

The second part contains an algebraic treatment of homomorphism functions and other graph parameters. The two main algebraic constructions (connection matrices and graph algebras) will play an important role later as well, but they also shed some light on the seemingly completely heterogeneous set of "graph parameters".

In the third part, which is the longest and perhaps most complete within its own scope, the theory of convergent sequences of dense graphs is developed, and applications to extremal graph theory and graph algorithms are given.

The fourth part contains an analogous theory of convergent sequences of graphs with bounded degree. This theory is more difficult and less well developed than the dense case, but it has even more important applications, not only because most networks arising in real life applications have low density, but also because of connections with the theory of finitely generated groups. Research on this topic has been perhaps the most active during the last months of my work, so the topic was a "moving target", and it was here where I had the hardest time drawing the line where to stop with understanding and explaining new results.

The fifth part deals with extensions. One could try to develop a limit theory for almost any kind of finite structures. Making a somewhat arbitrary selection, we only discuss extensions to edge-coloring models and categories, and say a few words about hypergraphs, to much less depth than graphs are discussed in parts III and IV.

I included an Appendix about several diverse topics that are standard mathematics, but due to the broad nature of the connections of this material in mathematics, few readers would be familiar with all of them.

One of the factors that contributed to the (perhaps too large) size of this book was that I tried to work out many examples of graph parameters, graph sequences, limit objects, etc. Some of these may be trivial for some of the readers, others may be tough, depending on one's background. Since this is the first monograph on the subject, I felt that such examples would help the reader to digest this quite diverse material.

In addition, I included quite a few exercises. It is a good trick to squeeze a lot of material into a book through this, but (honestly) I did try to find exercises about which I expected that, say, a graduate student of mathematics could solve them with not too much effort.