Phase transitions in discrete structures

The study of random discrete structures was pioneered by Paul Erdős and Alfréd Rényi in the 1950s/60s. Since that time, random structures (such as random graphs) have played a key role in combinatorics. But over the last decade, random discrete structures have been at the centre of a dramatic scientific development, involving **combinatorics**, **statistical mechanics**, **computational complexity** and **information theory**.

From the viewpoint of **combinatorics**, the key problem is to identify **phase transitions**, where a tiny change in the parameters entails a fundamental qualitative change in the overall outcome of the random experiment. The prime example of this is the emergence of the *"*giant component*"* (Erdős, Rényi 1959). More recently, progress has been made in studying phase transitions that are closely related to what physicists call "disordered systems", e.g., Achlioptas, Naor, Peres 2005.

In **statistical mechanics**, "disordered systems" comprise objects such as glasses that, in contrast to crystals, lack a rigid "ordered" structure. The mathematically rigorous study of disordered systems is notoriously difficult and, indeed, has been branded as a "challenge for mathematicians" by the probabilist Michel Talagrand. Random discrete structures serve as mathematical models ("mean-field models") of disordered systems. Over the past years, physicists have developed an ingenious, albeit non-rigorous, technique called the *cavity method* for the study of these models, see Mézard, Parisi, Zecchina: Science 2002. This has not only led to new, amazingly precise predictions on the location and nature of phase transitions but also to a new algorithm called *Survey Propagation* for prominent problems in computer science such as the notorious k-SAT problem.

In **computational complexity**, a fundamental question is why a very broad class of computational problems, the *NP-hard problems*, have withstood all attempts at developing efficient algorithms for constructing optimal solutions. In fact, the famous P≠NP-problem, one of the (open) "millennium problems" of the Clay mathematics institute, asks whether such efficient algorithms exist. In order understand why all currently known algorithms fail to solve NP-hard problems efficiently, mathematicians and computer scientists have been studying random models such as random *k*-SAT formulas (e.g., Kirkpatrick, Selman: Science 1994). An intriguing hypothesis is that in these models, the success of "local" algorithms is governed by a "dynamical" phase transition that resembles the glass transition from statistical mechanics, e.g., Achlioptas and Coja-Oghlan: FOCS 2008.

In **information theory***,* the goal is to devise codes that are not only efficient with respect to their various parameters, but that can also be decoded by an efficient algorithm. A broad family of candidates for meeting these criteria are Low Density Parity Check codes. These codes are, roughly speaking, based on random systems of linear equations.

My current work on random discrete structures is supported an ERC Starting Grant. Some of my prior work was supported by EPSRC grant EP/G039070/2.

Spectral methods and quasi-randomness

Numerous algorithms for combinatorial problems are based on **spectral methods**: the algorithm represents its input by a matrix and computes the eigenvalues and eigenvectors of the matrix in order to determine a solution. One particular application is graph partitioning, but the scope of spectral methods even encompasses problems such as *k*-SAT, the main benchmark problem in computational complexity: Coja-Oghlan, Goerdt, Lanka 2007. Although spectral methods are quite popular in practice, their theoretical understanding remains very limited – apart from the fact that most spectral algorithms used in practice have a terrible worst-case performance. Therefore, one of my research goals is to obtain a profound understanding of the relationship between combinatorial and spectral properties of graphs and other objects, and to exploit this information algorithmically.

In spectral graph theory the parameter that has attracted the most attention is the ***spectral gap***, i.e., the difference between the largest and the second largest eigenvalue (say, of the adjacency matrix). The spectral gap attains its largest possible value for random graphs *G(n,p)*, and conversely graphs with a large spectral gap are ***quasi-random***, i.e., (roughly speaking) they share many of the "global" combinatorial properties of random graphs. However, since "local" graph properties affect the spectral gap as well, it is not true that quasi-random graphs also have a large spectral gap (unless the graph is very dense). Hence, providing decent sufficient conditions for a large spectral gap is an open problem. We made a step towards a solution in Alon, Coja-Oghlan, Han, Kang, Rödl, Schacht 2010.

The notion of quasi-randomness is closely related to the concept of ***regular partitions***. A regular partition essentially is an approximation of a combinatorial object (e.g., a graph) by a bounded number of quasi-random objects. Here "bounded" means that the number only depends on the desired quality of the

approximation, but not on the size of the graph that we aim to approximate. (Of course, there are several ways to define precisely what it means to approximate a graph.) In the case of *dense* graphs, the notion of regular partitions is well-established and there are satisfactory algorithms for computing regular partitions. By contrast, for *sparse* graphs regular partitions are only known to exist for a limited class of graphs. Furthermore, all known algorithms for computing sparse regular partitions rely on very heavy machinery (including semidefinite programming). Therefore, my aim is to devise simpler, more efficient algorithms for computing regular partitions on as broad a class of inputs as possible. For recent progress towards this aim and applications in combinatorial optimisation see Coja-Oghlan, Cooper, Frieze 2009.

As indicated above, most of the spectral algorithms used in practice have a bad worst-case performance. Therefore, one way of understanding heuristic spectral methods (as are routinely used in various applications) is via ***probabilistic analysis***. The idea is to set up a meaningful probabilistic model of input instances and to analyse the algorithm's performance when applied to that model; frequently the model is identical to classes of benchmark instances for evaluating algorithms experimentally. Among other things, the probabilistic analysis of spectral algorithms requires analysing spectral properties of random matrices such as the spectral gap or the asymptotic distribution of the eigenvalues.

In summary, there is a significant gap between the practical success of spectral methods and our theoretical understanding, much like in the case of the simplex algorithm for linear programming. The research goal is to bridge this gap. On the one hand, this means devising enhanced spectral algorithms that combine the “robustness” of SDP with the efficiency of known spectral methods. On the other hand, we need to develop better techniques for analysing spectral methods. Furthermore, notions such as the spectral gap are closely related to quasi-randomness and regularity.

There is a Google Scholar profile and also a DBLP entry.