In the academic year 2007–2008 I worked on numerous problems concerning percolation, extremal graph theory, entropy, and models of real-world graphs, and published a good many papers. In what follows below, first I shall list some of the papers I published that year and will give the titles of various manuscripts I finished, and then I shall give brief descriptions of some of the results.

Papers Published


Highly connected monochromatic subgraphs, *Discrete Math.* 308 (2008), 1722–1725. (with A. Gyárfás)


Percolation on dual lattices with \(k\)-fold symmetry, *Random Structures and Algorithms* 32 (2008), 463–472. (with O. Riordan)


Pentagons vs. triangles, *Discrete Math.* 308 (2008), 4332–4336. (with E. Győri)


Relevant Manuscripts


Sentry selection in wireless networks (with P. Balister, A. Sarkar and M. Walters)

Hereditary properties of combinatorial structures (with J. Balogh and R. Morris)

Percolation on dense graph sequences (with C. Borgs, J. Chayes and O. Riordan)

$k$-nearest neighbour critical constant (with P. Balister, A. Sarkar and M. Walters)

On the structure of almost all graphs without fixed forbidden subgraphs (with J. Balogh and M. Simonovits)

Projections, entropy and sumsets (with P. Balister)

Random graphs and branching processes (with O. Riordan)

Sparse graphs: metrics and random models (with O. Riordan)

Bond percolation with attenuation in high dimensional random Voronoi tilings (with P. Balister)

Extending the Erdös-Stone theorem to $r$-cliques (with V. Nikiforov)
Descriptions of Some Results

Counting Regions With Bounded Surface Area. To state our results, we need some definitions. An \(r\)-cube \(C\) is an \(r\)-dimensional unit cube in \(\mathbb{R}^d\) with vertices in \(\mathbb{Z}^d\), i.e. a set of the form

\[ C = C(a, I) = \{ x \in \mathbb{R}^d : x_i = a_i \text{ for } i \notin I, \ a_i \leq x_i \leq a_i + 1 \text{ for } i \in I \}, \]

where \(a = (a_1, \ldots, a_d) \in \mathbb{Z}^d\) and \(I\) is a subset of \(\{1, \ldots, d\}\) of size \(r\).

An \(r\)-dimensional cubical complex (or \(r\)-complex) \(B\) is a finite union of \(r\)-cubes in \(\mathbb{R}^d\); a complex is rooted if it contains the cube \(C_r = C(0, \{1, \ldots, r\})\). The volume \(|B|\) of a complex \(B\) is the number of \(r\)-cubes in \(B\).

We define the boundary \(\partial C\) of a cube \(C\) to be the \((r - 1)\)-complex which is the union of the \(r\) pairs of faces \(C((a_1, \ldots, a_i, \ldots, a_d), I \setminus \{i\})\) and \(C((a_1, \ldots, a_i + 1, \ldots, a_d), I \setminus \{i\})\) for \(i \in I\). Furthermore, we define the boundary \(\partial B\) of the complex \(B = \bigcup_{i=1}^{n} C_i\) to be the \((r - 1)\)-complex which contains each \((r - 1)\)-cube that occurs in an even number of boundaries \(\partial C_i\). A complex \(B\) is closed if \(\partial B = \emptyset\). Define the surface area of \(B\) to be the volume of the boundary \(|\partial B|\).

We say that an \(r\)-complex \(B\) is connected if it is connected via its \((r - 1)\)-dimensional faces. More formally, let \(G\) be the graph with vertices equal to the component \(r\)-cubes of \(B\) and two vertices joined by an edge when these cubes share a common \((r - 1)\)-dimensional face. Then \(B\) is connected precisely when \(G\) is connected.

The number of \(d\)-dimensional cubical complexes with a given volume or surface area is interesting in its own right; however it also has applications to the Ising model in \(d\) dimensions, where the convergence of the low temperature expansion is dependent on the number of Peierls contours, i.e., the number of connected boundaries of rooted cubical complexes.

In 1998, Lebowitz and Mazel proved that there are between \((C_1 d)^n/2d\) and \((C_2 d)^{64n/d}\) complexes containing a fixed cube with connected boundary of \((d - 1)\)-volume \(n\). This result tells us much about this number of complexes, but still leaves a factor 128 in the exponent between the lower bound and the upper bound. In this paper Balister and I proved the considerably better bounds \((C_3 d)^{n/d}\) and \((C_4 d)^{2n/d}\), and so reduced the factor to 2. Nevertheless, the true order of the exponent is still a mystery. Furthermore, we showed that there are \(n^{n/(2d(d-1)) + o(1)}\) connected complexes containing a fixed cube with (not necessarily connected) boundary of volume \(n\).
**Line-of Sight Percolation.** Frieze, Kleinberg, Ravi and Debany proposed the following random graph as a model of an *ad hoc* network in an environment with (regular) obstructions. Given positive integers $n$ and $\omega$, let $G = \mathbb{Z}^2(\omega)$ be the graph with vertex set $[n] \times [n]$ in which two vertices $(x_1, y_1)$ and $(x_2, y_2)$ are joined if $x_1 = x_2$ and $|y_1 - y_2| \leq \omega$ or $y_1 = y_2$ and $|x_1 - x_2| \leq \omega$. Let $V$ be a random subset of $[n] \times [n]$ obtained by selecting each point $(x, y)$ with probability $p$, independently of the other points, and let $G[V]$ denote the subgraph of $G$ induced by the vertices in $V$. We write $G_{n,\omega,p}$ for $G[V]$, and $G_{\omega,p}$ for the infinite random graph defined in the same way but starting from $\mathbb{Z}^2$ rather than $[n] \times [n]$.

A natural interpretation of $G_{n,\omega,p}$ is as follows: sensors are dropped onto a random selection of crossroads in a regularly laid out city (or planted forest); two sensors can communicate if they are within distance $\omega$ and the line of sight between them is not blocked by a building (tree). The graph $G_{n,\omega,p}$ indicates which pairs of sensors can communicate directly, so we would like to know when $G_{n,\omega,p}$ has a giant component, and roughly how large it is. Alternatively, taking the point of view of percolation theory, we would like to study percolation in $G_{\omega,p}$, i.e., to know for which choices of the parameters $G_{\omega,p}$ has an infinite component, i.e., we should like to determine the critical probability $p_c(\omega)$ of percolation.

In a paper with Janson and Riordan, I have extended recent results of Frieze et al to $\lim_{\omega \to \infty} \omega p_c(\omega) = \log(3/2)$. We have also proved analogues of this result on the $n$-by-$n$ grid and in higher dimensions, the latter involving interesting connections to Gilbert’s continuum percolation model. To prove our results, we explored the component of the origin in a certain non-standard way, and showed that this exploration is well approximated by a certain branching process.

**Barrier Coverage.** Deriving the critical density needed to achieve coverage (or connectivity) of the geometric random graph defined by a random collection of discs (or other domains) is a fundamental problem. At a density *lower than critical*, with high probability, the network *does not* provide coverage (or connectivity), and, at a density *higher than critical*, with high probability, the network *does* provide coverage; hence, the term critical. Such conditions, however, are asymptotic in nature. Since in real life the deployment regions are always finite, such conditions are not too useful in practice.

Another major limitation of most of the results so far is that they limit themselves to thick deployment regions such as disks and squares,
so are not applicable to thin strips. The fact that percolation does not occur for thin regions is often cited as a primary reason for avoiding thin strips, as in Dousse, Baccelli, and Thiran’05. People who do consider rectangles (like Peserico and Rudolph’06) place a lower bound on the width to length ratio. When sensors are deployed in thin strips, such as when deploying along international borders to detect intrusion, or around forests to detect fire, no existing work can be used to derive the density of sensors needed for achieving coverage or connectivity.

If the goal of deploying sensors is to detect moving objects and phenomena (which is often the case in thin strip deployments), then the model of barrier coverage, as introduced by Kumar, Lai, and Arora’05, may be a more appropriate model than the widely studied full coverage. Barrier coverage ensures that no moving object or phenomenon can cross the barrier of sensors without being detected, whereas full-coverage ensures that every point in the deployment region is covered.

Further, when sensors are deployed for barrier coverage, achieving s-t connectivity, which ensures that a connected path exists between the two far ends of a thin strip, may be more appropriate than achieving full connectivity, which requires that every sensor be connected to every other sensor. The fact that some sensors may not be connected to the base station does not compromise the barrier coverage guarantee; all events can still be detected and communicated to the base station even if the base station is located at a far end.

The dotted line indicates a separating path — either a possible path of an undetected intruder, or a path disconnecting s from t in the network.

We have derived reliable estimates for the density needed to achieve coverage and connectivity in thin strips for four basic models of coverage and connectivity. We have developed a novel definition of break (a disruption in connectivity) that is critical in solving the problems of barrier coverage, s-t connectivity, and full connectivity in thin strips, all three of which are harder than full coverage.

For descriptions of our work, see

We also have a conference publication:

Yet another paper, with even more precise results and detailed mathematical proofs, is in preparation: unfortunately, there are numerous rather unpleasant difficulties on the way.

The critical constant for coverage in the $k$-nearest neighbor model. Let $\mathcal{P}$ be a Poisson process of intensity one in a square $S_n$ of area $n$. For a fixed integer $k$, we join every point of $\mathcal{P}$ to its $k$ nearest neighbours, creating an undirected random geometric graph $G_{S_n,k} = G_{n,k}$ in which every vertex has degree at least $k$. The connectivity of these graphs was first studied by Xue and Kumar; their results were greatly improved by Balister, Sarkar, Walters and me. It is not hard to see that $G_{n,k}$ becomes connected around $k = \Theta(\log n)$, and we proved that if $k(n) \leq 0.3043 \log n$ then the probability that $G_{n,k(n)}$ is connected tends to zero as $n \to \infty$, while if $k(n) \geq 0.5139 \log n$ then the probability that $G_{n,k(n)}$ is connected tends to one as $n \to \infty$. However, we were unable to prove the natural conjecture that there exists a critical constant $c_{\text{crit}}$ such that for $c < c_{\text{crit}},$

$$\mathbb{P}(G_{n,\lfloor c \log n \rfloor} \text{ is connected}) \to 0$$

and for $c > c_{\text{crit}},$

$$\mathbb{P}(G_{n,\lfloor c \log n \rfloor} \text{ is connected}) \to 1$$

as $n \to \infty$. By a careful analysis of the structure of the random geometric graph we have managed to prove this conjecture.

Clique percolation. For the past ten years or so, much work has been done on large-scale real-world graphs. Most of this work is of an experimental nature – although an important aspect of this work is to construct viable mathematical models of these networks, there is much less rigorous mathematical work on these models. This paper is a contribution to the body of rigorous mathematical results.
More precisely, in this paper we shall continue the work of Derényi, Palla and Vicsek, who in 2005 introduced clique percolation in order to study the dense, highly interconnected parts of real-world graphs, often referred to as communities, modules, clusters of cohesive groups. These groups can correspond to a variety of structures, including multi-protein functional units in molecular biology, sets of closely coupled stocks, industrial sectors in economy, groups of people, etc.

To define a clique percolation model, starting with a random graph $G$ generated by some rule, form an auxiliary graph $G'$ whose vertices are the $k$-cliques of $G$, in which two vertices are joined if the corresponding cliques share $k - 1$ vertices. Derényi, Palla and Vicsek considered in particular the case where $G = G(n, p)$, and found heuristically the threshold for a giant component to appear in $G'$. In this paper, Riordan and I give a rigorous proof of this result, as well as many of its extensions. The model turns out to be very interesting due to the essential global dependence present in $G'$.

**Percolation on dense graph sequences.** In this paper Borgs, Chayes, Riordan and I consider ‘percolation’ on arbitrary finite graphs: our aim is to determine the percolation threshold for an arbitrary sequence of dense graphs $(G_n)$. (As usual, by ‘percolation’ on a finite graph one understands the appearance of a giant component understood in some sensible way.) Let $\lambda_n$ be the largest eigenvalue of the adjacency matrix of $G_n$, and let $G_n(p_n)$ be the random subgraph of $G_n$ obtained by keeping each edge independently with probability $p_n$. We show that the appearance of a giant component in $G_n(p_n)$, corresponding to percolation on infinite graphs, has a sharp threshold at $p_n = 1/\lambda_n$. In fact, we have proved much more: if $(G_n)$ converges to an irreducible limit, then the density of the largest component of $G_n(c/n)$ tends to the survival probability of a multi-type branching process defined in terms of this limit. Here the notions of convergence and limit are those of Borgs, Chayes, Lovász, Sós and Vesztergombi.

In addition to using basic properties of convergence, in this paper we have made heavy use of the methods Janson, Riordan and I used in our long paper on random graph models. In particular, we used multi-type branching processes to study the emergence of a giant component in a very broad family of sparse inhomogeneous random graphs.

**Sentry selection in wireless networks.** For several years now, Balister, Hänggi, Sarkar, and Walters have been investigating the possibility of partitioning a set of randomly placed sensors ‘covering’ a certain
domain into many groups of ‘sentries’ that can guard the entire domain. By using combinatorial and geometric methods combined with sophisticated probabilistic tools, we have proved sharp results that are essentially best possible.

A network of wireless sensors is deployed in a large area (the sensing region). Each sensor can detect any event occurring within a certain domain in its vicinity. Suppose that we wish to devise a rota system so that each sensor can sleep for most of the time, for example, to extend battery life. A natural way of doing this would be to partition the set of sensors into $k$ groups, and arrange that only the sensors in group $\ell$ are active in the $\ell^{th}$ time slot. After $k$ time slots have expired, we repeat the process. In order to detect an event occurring anywhere and at any time, it is necessary that the sensors in each group themselves form a single cover of the sensing region. What density of sensors (equivalently, how large sensing regions) would make likely that such a partitioning is possible?

In the case we consider, the sensor locations are points of a unit intensity Poisson process, and each sensor can observe the points within distance $r$. Thus our question becomes: for fixed $k$, how large should $r$ be to ensure that the sensors can be partitioned into $k$ groups, each of which covers the sensing region? We call this the problem of sentry selection, since each of the groups is a group of sentries keeping watch over the region while the others are sleeping.

Note that coverage and partitionability are very different. In the figures below, coverage is by large disks, so that only some arcs are shown. The solid discs are exactly 2-covered, and everything is covered at least twice: nevertheless, the system cannot be split into two subcovers.

To state one of our main results, let $\mathcal{P}$ be a Poisson process of intensity one in the infinite plane $\mathbb{R}^2$. We surround each point $x$ of $\mathcal{P}$ by the open disc of radius $r$ centred at $x$. Now let $S_n$ be a fixed disc of area $n$, and let $C_r(n)$ be the set of discs which intersect $S_n$. Write $E^k_r$ for the
event that $C_r(n)$ is a $k$-cover of $S_n$, and $F_r^k$ for the event that $C_r(n)$ may be partitioned into $k$ disjoint single covers of $S_n$. Among many other results, we prove that $\mathbb{P}(E_r^k \setminus F_r^k) \leq \frac{c_k}{\log n}$, and that this result is best possible. We also give improved estimates for $\mathbb{P}(E_r^k)$.

Finally, here is a hitting time version of the main result. One by one throw disks (of radius 1, say) at random on a large square, and stop this process when every point of the square is covered at least $k$ times. Then, with probability tending to one (with the size of the square), the set of disks can be partitioned into $k$ groups, each of which covers the entire square.

**Sparse graphs: metrics and random models.** In 2007, in a one 120-page long paper, Janson, Riordan and I introduced a very general family of random graph models, producing inhomogeneous random graphs with $\Theta(n)$ edges. Roughly speaking, there is one model for each kernel, i.e., each symmetric measurable function from $[0, 1]^2$ to the non-negative reals, although the details are much more complicated, to ensure the exact inclusion of many (perhaps even most) of the recent models for large-scale real-world networks.

A different connection between kernels and random graphs arises in the recent work of Borgs, Chayes, Lovász, Sós, Szegedy and Vesztergombi, published in numerous papers composed of over two hundred pages. They introduced several natural metrics on dense graphs (graphs with $n$ vertices and $\Theta(n^2)$ edges), showed that these metrics are equivalent, and gave a description of the completion of the space of all graphs with respect to any of these metrics in terms of graphons, which are essentially kernels. One of the most appealing aspects of this work is the message that sequences of inhomogeneous quasi-random graphs are in a sense completely general: any sequence of dense graphs contains such a subsequence. Alternatively, their results show that certain natural models of dense inhomogeneous random graphs (one for each kernel) cover the space of dense graphs: there is one model for each point of the completion, producing graphs that converge to this point.

In this paper, Riordan and I have investigated to what extent the results above for dense graphs can be generalized to graphs with $o(n^2)$ edges. Although many of the definitions extend in a simple way, the connections between the various metrics, and between the metrics and random graph models, turn out to be much more complicated than in the dense case. We have managed to prove many partial results, and have stated even more conjectures and open problems, whose resolution
would greatly enhance the currently rather unsatisfactory theory of metrics on sparse graphs.

**Sparse random graphs with clustering.** This paper, with Janson and Riordan, can be viewed as a continuation of the long paper we wrote in 2007, in which we introduced a general model of sparse random graphs with independence between the edges. The aim of this paper is to present an extension of this model in which the edges are far from independent, and to prove several results about this extension. The basic idea is to construct the random graph by adding not only edges but also other small graphs. In other words, we first construct an inhomogeneous random hypergraph with independent hyperedges, and then replace each hyperedge by a (perhaps complete) graph.

Although flexible enough to produce graphs with significant dependence between edges, this model is nonetheless mathematically tractable. Indeed, we find the critical point where a giant component emerges in full generality, in terms of the norm of a certain integral operator, and relate the size of the giant component to the survival probability of a certain (non-Poisson) multi-type branching process. While our main focus is the phase transition, we also study the degree distribution and the numbers of small subgraphs. We illustrate the model with a simple special case that produces graphs with power-law degree sequences with a wide range of degree exponents and clustering coefficients.