The global effects of local graph structure

Alex Scott (University of Oxford)

Nankai Symposium, August 2021

A fundamental question about mathematical objects is how local structure and global structure interact. What are the large-scale effects of local structure? How are global properties reflected in local structure? The relationship between local and global structure is subtle, and is an important theme in many fields (combinatorics, statistical physics, complexity, geometric group theory, model theory, and across much of applied mathematics and computer science).

Our focus here will be on combinatorics, and in particular on the local/global structure of graphs. Perhaps the first question is: does a graph need to have any structure at all? On other words, can a graph be totally disordered or random? The simplest (or most highly ordered) graphs are complete graphs and graphs with no edges. A *complete subgraph* of a graph G is a set of vertices, every pair of which is joined by an edge; an *independent set* is a set of vertices, no pair of which is adjacent. It is then natural to ask: if G is a large graph, must it contain a large complete subgraph or independent set?[1]

This question was answered by Frank Ramsey [15], who showed in 1930 that every infinite graph contains an infinite complete subgraph or independent set. The finite version of this result (which is equivalent by compactness) is the following:

Theorem 1 *(Ramsey's Theorem) For every* $k \geq 1$ *there is an integer* $R(k)$ *such that every graph with at least* $R(k)$ *vertices contains a complete subgraph or independent set of size* k.

This elementary but foundational result shows that 'large' graphs contain 'large' homogeneous structures. But how large is large? A quantitative version of Ramsey's Theorem was proved in 1933 by Erdős and Szekeres [10], who showed

---

[1] Note that we have to include both complete subgraphs and independent sets as possibilities, as G might itself be a complete graph, or it might have no edges.

the 'off-diagonal' result that every graph with at least $\binom{r+s-2}{r-1}$ vertices contains a complete subgraph with r vertices or an independent set with s vertices. A straightforward calculation gives the following corollary:

Theorem 2 *Every graph on n vertices contains a complete subgraph or independent set of size at least* $c_1 \log n$*.*

Can this be improved? It turns out that this bound is within a constant factor of being best possible. In fact, by considering random graphs, it can be shown that most graphs do not have cliques or independent sets of more than logarithmic size. This was shown by Paul Erdős [8], in one of the earliest applications of the probabilistic method.

Theorem 3 *For some* $c_2 > 0$ *and every* $n \geq 2$*, there are graphs on n vertices do not contain a complete subgraph or independent set of size more than* $c_2 \log n$*.*

In fact, this bound is quite simple to prove: consider a random graph on n vertices, in which each possible edge is present independently with probability $1/2$. The expected number of independent sets or complete subgraphs on k vertices is $2\binom{n}{k}2^{-\binom{k}{2}} < (n2^{-(k-1)/2})^k$, which tends to 0 if $k > c_2 \log n$ (where c_2 is an appropriately chosen constant). By Markov's Inequality, the probability that there is an independent set or complete subgraph of size more than $c_2 \log n$ also tends to 0.

Thus most graphs have complete subgraphs and independent sets of only logarithmic size. But what if we are given some information about the graph? In particular, how does the picture change if there is some restriction on the local structure of a graph? It turns out that the picture is then very different.

In order to discuss this, we must first fix what we mean by 'local structure'. There are two natural notions of substructure for graphs:

- A *subgraph* of G is any graph obtained from G by deleting vertices and edges.
- An *induced subgraph* of G is any graph obtained by deleting vertices (and only edges that are incident with deleted vertices).

For example, every graph is a subgraph of some complete graph; but every induced subgraph of a complete graph is complete. Our interest here will be in *induced subgraphs*.

So, what happens if we constrain the local structure, by looking at graphs that do not contain some specific graph H as an induced subgraph? Let us say that a graph is *H-free* if it does not contain an induced copy of H. Erdős and Hajnal [9] conjectured in 1989 that H-free graphs exhibit a radically different behaviour: the size of independent sets or complete subgraphs jumps from *logarithmic* to *polynomial*.

Conjecture 4 *(Erdős and Hajnal, 1989) For every graph H, there is a constant $c = c(H) > 0$ such that the following holds: for $n \geq 2$, every H-free graph with n vertices has a complete subgraph or independent set with at least n^c vertices.*

The Erdős-Hajnal Conjecture has become one of the central conjectures in graph theory, and has inspired a substantial body of research. Erdős and Hajnal proved that the conjecture holds with the weaker bound $e^{c\sqrt{\log n}}$ in place of n^c, but this remains the best result for general graphs H.

For what graphs H has the the Erdős-Hajnal Conjecture been proved? Embarrassingly few. . . .

- The conjecture holds when H is a complete graph (this follows easily from the quantitative version of Ramsey's theorem).
- The conjecture holds when H is three-edge path (this gives a well understood class of graphs known as *cographs*). This follows from a simple structural argument. A more complex structural argument shows the conjecture for the five-vertex graph known as the *bull* (Chudnovsky and Safra [4]).
- If it's true for H then it is also true for $\overline{H}$ (this follows easily from the statement of the conjecture).

An influential paper of Alon, Pach and Solymosi [1] considered the effects of *substitution*. The operation of *substituting* a graph H_1 into a vertex v of a graph H_2 replaces v by a copy of H_1, whose vertices are adjacent to all the neighbours of v in G_2. This was a natural operation to investigate, as it had previously played an important role in the investigation of perfect graphs: Lovász [13, 14] showed that the class of perfect graphs is closed under substitution en route to proving the Weak Perfect Graph Conjecture. Using very different methods, Alon, Pach and Solymosi showed the following.

Theorem 5 *(Alon, Pach and Solymosi, 2001) The class of graphs H for which Erdős-Hajnal holds is closed under substitution.*

Until recently, the only graphs H for which the Erdős-Hajnal Conjecture was known were the graphs listed above, together with graphs that can be obtained from them by repeated applications of complementation and substitution. A particularly frustrating open case was the five-vertex cycle C_5. This has been the focus of significant attention over the last thirty years. Indeed, it was specifically highlighted by Erdős and Hajnal [9] and also by Gyárfás [11].

Recently, there was some progress:

Theorem 6 *(Chudnovsky, Scott, Seymour, Spirkl [7]) The Erdős-Hajnal Conjecture holds when H is a cycle of length 5.*

The next target is clear: up to complementation, the five-vertex path P_5 is the only graph on five vertices for which the Erdős-Hajnal Conjecture is not known.

Another line of development that has been very fruitful recently has been to consider the effects of forbidding more than one induced subgraph. A *hereditary class* of graphs is a class of graphs that is closed under taking induced subgraphs. Two important properties of hereditary classes are the following:

- A hereditary class $\mathcal{G}$ of graphs has the *Erdős-Hajnal property* if there is $c > 0$ such that every $G \in \mathcal{G}$ has an independent set or complete subgraph with at least $|G|^c$ vertices.
- A hereditary class $\mathcal{G}$ of graphs has the *strong Erdős-Hajnal property* if there is $\delta > 0$ such that every $G \in \mathcal{G}$ with at least two vertices has disjoint sets A, B of at least δn vertices such that the pair (A, B) is either complete or anticomplete (either all edges between A and B are present or none are).

The Erdős-Hajnal Conjecture says that the class of H-free graphs has the Erdős-Hajnal property. It is not hard to show that the strong Erdős-Hajnal property imples the Erdős-Hajnal property.

So when does the strong Erdős-Hajnal property hold for the class of H-free graphs? By considering random graphs, it is easily shown that H must be a forest (take sparse random graph, $p \sim \log n/n$). Unfortunately, the same argument (using very dense random graphs) shows that the complement $\overline{H}$ must also be a forest. But then H has at most four vertices, and the conjecture is already known for these cases.

However, the situation is more interesting if we forbid more than one induced subgraph. The starting point here is a result of Bousquet, Lagoutte and Thomassé [2], which states that for every positive integer t, the class of graphs G such that neither G nor its complement contains a t-vertex path as an induced subgraph satisfies the strong Erdős-Hajnal property. It follows that the Erdős-Hajnal property holds if we exclude two graphs: *both* a path on t vertices *and* its complement.

This was extended by Choromanski, Falik, Liebenau, Patel and Pilipczuk [3] and then further by Liebenau, Pilipczuk, Seymour and Spirkl [12] who conjecturerd that the path could be replaced by any tree. This conjecture has now been proved.

Theorem 7 *(Chudnovsky, Scott, Seymour and Spirkl [5]): Let T be a tree. Then the class of graphs G such that neither G nor $\overline{G}$ contains an induced copy of T satisfies the strong Erdős-Hajnal property.*

It follows that, if we forbid both a tree and the complement of a tree as induced subgraphs, then we get complete subgraphs or independent sets of polynomial size (the Erdős-Hajnal Conjecture says that it would be enough to forbid just one of these two induced subgraphs). Furthermore, since we *must* exclude both a tree and the complement of a tree to get the strong Erdős-Hajnal property, this result characterizes hereditary classes satisfying the strong Erdős-Hajnal property.

For more general H, there is an analogue of Theorem 7 where we exclude subdivisions of H. A *subdivision* of a graph H is any graph obtained by added new vertices along the edges of H (or equivalently replacing some of the edges by paths).

Theorem 8 *(Chudnovsky, Scott, Seymour and Spirkl [6]) For any graph H, the class of graphs G such that neither G nor $\overline{G}$ contains an induced subdivision of H satisfies the strong Erdős-Hajnal property*

There isn't space here for any detailed discussion of the proofs, but it seems likely that any successful strategy for attacking the conjecture will combine structural understanding with (pseudo)random techniques (indeed, many proofs already begin by applying Szemerédi's Regularity Lemma [17], often by means of a valuable result of Rödl [16]).

In any case, it is clear that the relationship between local and global structure is still somewhat mysterious. The Erdős-Hajnal Conjecture remains a central problem in graph theory, and an intriguing challenge.

References

[1] N. Alon, J. Pach and J. Solymosi, Ramsey-type theorems with forbidden subgraphs, *Combinatorica* **21** (2001), 155–170.

[2] N. Bousquet, A. Lagoutte, and S. Thomassé, The Erdős-Hajnal conjecture for paths and antipaths, *J. Combinatorial Theory, Ser. B*, **113** (2015), 261–264.

[3] K. Choromanski, D. Falik, A. Liebenau, V. Patel, and M. Pilipczuk, Excluding hooks and their complements, *Electronic J. Combinatorics* **25** (2018) P3.27.

[4] M. Chudnovsky and M. Safra, The Erdős-Hajnal conjecture for bull-free graphs, *J. Combinatorial Theory, Ser. B*, **98** (2008), 1301–1310.

[5] M. Chudnovsky, A. Scott, P. Seymour and S. Spirkl, Pure pairs. I. Trees and linear anticomplete pairs, *Advances in Mathematics* **375** (2 December 2020), 107396.

[6] M. Chudnovsky, A. Scott, P. Seymour and S. Spirkl, Pure pairs. II. Excluding all subdivisions of a graph, *Combinatorica* **41** (2021), 379–405.

[7] M. Chudnovsky, A. Scott, P. Seymour and S. Spirkl, Erdős-Hajnal for graphs with no 5-hole, *arXiv:2102.04994*

[8] P. Erdős, Some remarks on the theory of graphs, *Bulletin of the American Mathematical Society* **53** (1947), 292–294.

[9] P. Erdős and A. Hajnal, Ramsey-type theorems, *Discrete Applied Math.* **25** (1989), 37–52.

[10] P. Erdős and G. Szekeres, A combinatorial problem in geometry, *Compositio Mathematica* **2** (1935), 463–470.

[11] A. Gyárfás, Reflections on a problem of Erdős and Hajnal, *in: The Mathematics of Paul Erdős*, (R. L. Graham and J. Nešetřil, eds.), Algorithms and Combinatorics **14**, Volume II, Springer-Verlag, Heidelberg, 1997, 93–98.

[12] A. Liebenau, M. Pilipczuk, P. Seymour and S. Spirkl, Caterpillars in Erd˝os-Hajnal, *J. Combinatorial Theory, Ser. B* **136** (2019), 33–43.

[13] L. Lovász, Normal hypergraphs and the perfect graph conjecture, *Discrete Mathematics* **2** (1972), 253–267.

[14] L. Lovász, Perfect Graphs, *in: Selected Topics in GraphTheory, Volume 2* (L. W. Beineke, R. J. Wilson, eds.), Academic Prcess, London-New York, 1983, 55–87.

[15] F.P. Ramsey, On a Problem of Formal Logic, *Proceedings of the London Mathematical Society* (1930), 264–286.

[16] V. Rödl, On universality of graphs with uniformly distributed edges, *Discrete Math.* **59** (1986), 125–134.

[17] E. Szemerédi, Regular partitions of graphs, *Problèmes combinatoires et théorie des graphes* (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Colloq. Internat. CNRS, vol. 260, Paris: CNRS, pp. 399–401.