Research Statement

Natalie Behague

My research interests are in extremal and probabilistic combinatorics. My work so far encompasses several different problems, all variations on the usual extremal combinatorics theme of looking at the existence of certain substructures or properties of a family of combinatorial objects (graphs, hypergraphs, automata) as the objects grow large. A particular pattern that emerges from my work is a use of intricate constructions, which occur not only as counterexamples, but are also used to show positively that various things are possible and as tools to bound probabilities.

I was awarded a PIMS postdoctoral research fellowship to work at the University of Victoria with Natasha Morrison and Jonathan Noel. Prior to this, I spent a year as a postdoctoral research fellow at Toronto Metropolitan University (formerly Ryerson University) with the Graphs at Ryerson group, including Paweł Prałat and Anthony Bonato. I completed my PhD at Queen Mary University of London in 2020 under the supervision of Robert Johnson.

I am the sole author of two papers and I have co-authored a further seven papers with thirteen different collaborators, including prominent researchers as well as other early-career researchers. My network has expanded rapidly in a relatively short time. I have been invited to give a total of twelve external seminar talks at ten different institutions and three conference talks, as well as attending numerous conferences as a contributor. I have also attended four highly selective workshops, either by invitation (Workshop on Graph Theory and Combinatorics in Thuringia, Cross-community Collaborations in Combinatorics workshop) or by application (Graduate Research Workshop in Combinatorics, Young Researchers in Combinatorics workshop).

1 Overview of Significant Research Contributions

My work so far encompasses several different problems under the broad umbrella of extremal and probabilistic combinatorics. The different areas I have worked on are summarised briefly below and then in more depth in the indicated subsections.

Subsection 1.1 *Two Saturation Problems.* Tuza's famous conjecture on the saturation number $sat(F, n)$ states that for r-uniform hypergraphs F the value $\frac{\mathrm{sat}(F,n)}{n^{r-1}}$ converges. I answered a question of Pikhurko concerning the asymptotics of the saturation number for families of hypergraphs, proving in particular that $\frac{\mathrm{sat}(\mathcal{F},n)}{n^{r-1}}$ need not converge if $\mathcal{F}$ is a family of r-uniform hypergraphs. I, together with Johnston, Letzter, Morrison and Ogden, also proved that the rainbow saturation number, a natural generalization of the saturation number to edge-coloured graphs, is linear. This answered a question of Girão, Lewis and Popielarz.

Subsection 1.2 *1-factorizations of the Hypercube.* The existence or non-existence of perfect 1-factorizations has been studied for various families of graphs, with perhaps the most famous open problem in the area being Kotzig's conjecture which states that the complete graph K_{2n} has a perfect 1-factorization. I focused on another well-studied family of graphs: the hypercubes Q_d. I answered almost fully the question of how close (in some particular sense) to perfect a 1-factorization of the hypercube can be.

Subsection 1.3 *Subgraphs of Semi-random graphs.* The semi-random graph process is a single-player game. Starting with an empty graph on n vertices, in each round a vertex u is presented at random and the player selects an edge uv to add to the graph. In our work, the objective of the player is to construct with high probability a subgraph isomorphic to a fixed graph H in as few rounds as possible. Trent Marbach, Paweł Prałat and Andrzej Ruciński and I proved the conjecture that asymptotically almost surely one can construct H in t rounds for any $t \gg n^{(d-1)/d}$, where $d \geq 2$ is the degeneracy of H. We also generalized this to a semi-random s-uniform hypergraph process.

Subsection 1.4 *Černý's Conjecture.* An automaton on n states is synchronizing if there is a word that maps all n states onto the same state. Černý's conjecture on the length of the shortest such word is probably the most famous open problem in automata theory. With Robert Johnson, I considered the closely related question of determining the minimum length of a word that maps k states onto a single state. We found a simple combinatorial argument that improves the upper bound on the minimum length of such a word, and improved the bound even further in the case $k = 4$ or 5.

Subsection 1.5 *Probabilistic Zero-forcing.* Probabilistic zero-forcing is a model for rumour-spreading across a network, where each vertex infects its neighbours with a probability proportional to how many of its neighbours are already infected. With Trent Marbach and Paweł Prałat, I studied the speed of propogation on grids and hypercubes.

Subsection 1.6 *Common Pairs of Graphs.* A graph H is said to be *common* if the number of monochromatic labelled copies of H in a red/blue edge colouring of a large complete graph is asymptotically minimized by a random colouring with an equal proportion of each colour. Natasha Morrison and Jonathan Noel and I extended this notion to an asymmetric setting in a very natural way. That is, we define a pair (H_1, H_2) of graphs to be $(p, 1-p)$-common if a particular linear combination of the density of H_1 in red and H_2 in blue is asymptotically minimized by a random colouring in which each edge is coloured red with probability p and blue with probability $1 - p$. We extended many of the results on common graphs to this asymmetric setting and we obtained several novel results for common pairs of graphs with no natural analogue in the symmetric setting, using tools such as graph limits, convexity and flag algebras. There are many natural and interesting open problems around this new definition.

1.1 Two Saturation Problems

For a fixed graph H we say that a graph G is *H-saturated* if it does not contain H as a subgraph, but adding any extra edge creates a copy of H as a subgraph. The *saturation number* of H, denoted by $\mathrm{sat}(n, H)$, is the minimum number of edges in an H-saturated graph G on n vertices.

The saturation number can be thought of as a sort of opposite to the Turán extremal number $\mathrm{ex}(n, H)$ that counts the maximum number of edges in an H-free graph on n vertices. Since a maximal H-free graph is H-saturated, $\mathrm{ex}(n, H)$ equivalently counts the maximum number of edges in an H-saturated graph on n vertices. The saturation number in contrast counts the minimum number of edges among such graphs and is an interesting counterpoint to the

Turán number – the saturation number is in many ways less well-behaved. For example, we know that the Turan density $\lim_{n\to\infty} \mathrm{ex}(F,n)/n^2$ exists. Kászonyi and Tuza [20] proved that $\mathrm{sat}(n,H) = O(n)$ and Tuza [26] conjectured that $\mathrm{sat}(F,n)/n$ always tends to a limit as n tends to infinity, but this conjecture is still open.

The definition of saturation extends naturally to families of graphs. Pikhurko [24] disproved a strengthening of Tuza's conjecture by finding a finite family $\mathcal{F}$ of graphs such that $\mathrm{sat}(\mathcal{F},n)/n$ does not converge as n tends to infinity. Pikhurko then asked whether a similar behaviour can occur for families of r-uniform hypergraphs.

I showed [3] that the answer to this question is yes. The proof is by construction of a family $\mathcal{F}$ which grows in size with r. By adjusting this construction I also demonstrated that we can reduce the size of the forbidden family to contain just four hypergraphs, bringing it closer to Tuza's conjecture which concerns single hypergraphs.

Saturation can also be generalised to edge-coloured graphs, a generalisation that was first considered by Hanson and Toft [18] and studied further by Barrus, Ferrara, Vandenbussche and Wenger [2] and Girão, Lewis, and Popielarz [13] for rainbow edge-colourings.

An edge-colouring of a graph is *rainbow* if every edge is assigned a distinct colour. An edge-coloured graph G is *H-rainbow saturated* if it does not contain a rainbow copy of H, but the addition of any non-edge in any colour creates a rainbow copy of H. The *rainbow saturation number* of H, denoted by $\mathrm{rsat}(n,H)$, is then the minimum number of edges in an H-rainbow-saturated edge-coloured graph on n vertices.

Girão, Lewis, and Popielarz [13] conjectured that, like the usual saturation number, the rainbow saturation number of any non-empty graph H is at most linear in n. In collaboration with Tom Johnston, Shoham Letzter, Natasha Morrison and Shannon Ogden, I have recently proved this conjecture [6]. In addition, we also found an improved upper bound on the rainbow saturation number of the complete graph, disproving a conjecture of Girão, Lewis, and Popielarz. We are continuing to work on a related open problem on rainbow saturation for proper colourings

1.2 1-Factorizations of the Hypercube

A 1-factorization of a graph H is a partition of the edges of H into disjoint perfect matchings $\{M_1, M_2, \dots, M_n\}$, also known as 1-factors. Let $\mathcal{M} = \{M_1, M_2, \dots, M_n\}$ be such a 1-factorization. We say that $\mathcal{M}$ is a perfect factorization if every pair $M_i \cup M_j$ with i, j distinct forms a Hamilton cycle. A 1-factorization $\mathcal{M}$ is called *semi-perfect* if $M_1 \cup M_i$ forms a Hamilton cycle for all $i \neq 1$.

The existence or non-existence of perfect 1-factorizations has been studied for various families of graphs, with perhaps the most famous open problem in the area being Kotzig's conjecture which states that the complete graph K_{2n} has a perfect 1-factorization. I focused on another well-studied family of graphs: the hypercubes Q_d.

Craft [1] conjectured that for every integer $d \geq 2$ there is a semi-perfect 1-factorization of Q_d. This was proved independently by Gochev and Gotchev [14] and by Královič and Královič [22] in the case where d is odd, and settled for d even by Chitra and Muthusamy [11]. Gochev and Gotchev in fact went further and defined $\mathcal{M}$ to be k-semi-perfect if $M_i \cup M_j$ forms a Hamilton cycle for every $1 \leq i \leq k$ and $k+1 \leq j \leq d$. They proved that there is a k-semi-perfect factorization of Q_d whenever k and d are both even with $k < d$.

It turns out there is no perfect 1-factorization of the hypercube, which is a corollary of a result due to Laufer [23]. An analysis of Laufer's proof reveals that a k-semi-perfect factorization is the best we can hope for: the matchings must split into two classes with no two matchings in the same class forming a Hamilton cycle. In light of this observation, the only remaining question is whether for any k and d there is a k-semi-perfect factorization of Q_d. In my work [4] I resolved the question almost completely, with only the case $k = 3$ and $d = 6$ left unresolved.

1.3 Subgraphs of Semi-random Graphs and Hypergraphs

The semi-random graph process is a single-player game that begins with an empty graph on n vertices. In each round, a vertex u is presented to the player independently and uniformly at random. The player then adaptively selects a vertex v and adds the edge uv to the graph. For a fixed monotone graph property, the objective of the player is to force the graph to satisfy this property with high probability in as few rounds as possible.

In collaboration with Marbach, Prałat and Ruciński [8] I worked on this problem where the player is aiming to construct a subgraph isomorphic to an arbitrary, fixed graph G. Let $\omega = \omega(n)$ be any function tending to infinity as $n \to \infty$. Ben-Eliezer, Hefetz, Kronenberg, Parczyk, Shikhelman and Stojakovič [10] proved that asymptotically almost surely one can construct G in less than $n^{(d-1)/d}\omega$ rounds where $d \geq 2$ is the degeneracy of G. It was also proved that the result is sharp for $G = K_{d+1}$, that is, asymptotically almost surely it takes at least $n^{(d-1)/d}/\omega$ rounds to create K_{d+1}. Moreover, these authors conjectured that their general upper bound is sharp for all graphs G.

We proved this conjecture [8]. Prałat, Ruciński and I also have an ongoing project on a natural generalization of the process to s-uniform hypergraphs, the semi-random hypergraph process in which $r \geq 1$ vertices are presented at random, and the player then selects $s - r \geq 1$ vertices to form an edge of size s. Our results for graphs easily generalise to hypergraphs when $r = 1$; the threshold for constructing a fixed s-uniform hypergraph G is, again, determined by the degeneracy of G. However, when $r \geq 2$; thresholds are not even known for complete hypergraphs. We have general upper and lower bounds, and we can calculate the exact threshold for some sparse hypergraphs. We are working on proving tightness for our upper bound, starting with the case of complete hypergraphs.

1.4 Černý's Conjecture

An automaton consists of a finite set of states – usually $[n]$ – and a finite set of *mappings*, which are functions from the set of states to itself. We are interested in the results of applying a sequence of mappings to the set of states – we call such a sequence of mappings a *word* of the automaton. We say that a word is a *reset word* if it sends every state to the same point, and we call an automaton *synchronizing* if it has a reset word.

The most famous and long-standing open problem on synchronizing automata is Černý's conjecture.

Conjecture 1.1 (Černý's Conjecture)**.** *Suppose an automaton on n states is synchronizing. Then there exists a reset word of length at most $(n-1)^2$.*

We considered the following question: what is the minimum m such for any synchronizing automaton Ω on $[n]$ there is some set of k states and a word w length at most m sending that set to a singleton? We call this minimum the *k-set rendezvous time*, denoted by $\mathrm{rdv}(k, n)$. Note that finding $\mathrm{rdv}(n, n)$ is equivalent to solving Černý's conjecture. This question was studied by Gonze and Jungers [15] in the special case $k = 3$. They proved that $\mathrm{rdv}(3, n)$ is bounded below by $n + 3$ and bounded above by $\approx 0.1545n^2 + O(n)$, using an approach based on linear programming.

Robert Johnson and I [5] found a simple argument for general k almost halving the upper bound on the minimum length of a word sending k states to a single state. We further improved the upper bound on the minimum length of a word sending 4 states to a singleton from $0.5n^2$ to $\approx 0.459n^2$, and the minimum length sending 5 states to a singleton from n^2 to $\approx 0.798n^2$. In contrast to the result of Gonze and Jungers [15], our methods are purely combinatorial in nature. We are hopeful that the tools we used could be extended further and combined with new ideas to give improved upper bounds on $\mathrm{rdv}(k, n)$ for $k > 3$. We conjectured that $\mathrm{rdv}(3, n)$ is in fact linear in n. Any techniques involved in a proof of this conjecture may well generalise to give improved bounds on the k-set rendezvous time $\mathrm{rdv}(k, n)$ and potentially $\mathrm{rdv}(n, n)$, the Černý bound itself.

1.5 Probabilistic Zero Forcing

Zero forcing is a deterministic iterative graph colouring process in which vertices are coloured either blue or white, and in every round, any blue vertices that have a single white neighbour force these white vertices to become blue. Zero forcing was initially formulated to bound a problem in linear algebra known as the min-rank problem [17]. In addition to this application to mathematics, zero forcing also models many real-world propagation processes such as rumour spreading. However, the deterministic nature of zero forcing may not fit the chaotic nature of real-life situations and as such, probabilistic zero forcing has also been proposed and studied. In probabilistic zero forcing, given a graph G, a set of blue vertices Z, and vertices $u \in Z$ and $v \in V(G) \setminus Z$ such that $uv \in E(G)$, in a given time step, vertex u will force vertex v to become blue with probability $(|N[u] \cap Z|) / \deg(u)$, where $N[u]$ is the closed neighbourhood of u.

With Trent Marbach and Paweł Prałat [7], I studied probabilistic zero forcing on hypercubes and grids. We showed that the propagation time for the hypercube Q_d is asymptotically almost surely $\Theta(d)$, which is tight. The propagation time for the $m \times n$ grid is $\leq (1 + 10^{-7} + o(1))(m + n)/2$, which is very close to the lower bound (and conjectured true value) $(1 + o(1))(m + n)/2$. The proof for the grids requires the introduction of a Markov chain and an analysis using Chernoff–Heoffding bounds for Markov chains.

1.6 Common Pairs of Graphs

Ramsey's Theorem says for a given graph H, we can find an integer N such that every red/blue graph on $\geq N$ vertices contains a monochromatic H. The closely related Ramsey multiplcity problem asks instead for the minimal number of monochromatic copies of H in a red/blue edge-colouring of a large complete graph. Goodman [16] found the Ramsey multiplicity of the triangle by proving that the number of monochromatic triangles is minimized by an unbiased random edge colouring. Inspired by this, a graph H is called *common* if, like the triangle, the number of monochromatic labelled copies of H in a red/blue edge colouring of a large complete graph is asymptotically minimized by an unbiased random colouring.

Ramsey multiplicity and common graphs have been studied extensively following Goodman. Erdős [12] conjectured that all complete graphs are common. This was disproved much later by Thomason [25] who showed that K_4 is not common. Moreover, Jagger, Šťovíček and Thomason [19] extended this to show that any graph containing a copy of K_4 is not common.

Natasha Morrison, Jonathan Noel and I extended this notion of commonness very naturally to an asymmetric setting [9]. Take a pair of graphs H_1 and H_2. Suppose we wish to minimise c_1 times the number of red copies of H_1 plus c_2 times the number of blue copies of H_2 in a red/blue edge colouring of a large complete graph. We say the pair is $(p, 1-p)$-*common* if this count is asymptotically minimised by a random red/blue colouring where edges are red with probability p and blue with probability $1 - p$. It transpires that there is only one natural definition of what

c_1 and c_2 should be in terms of p, $e(H_1)$ and $e(H_2)$.

We generalised many of the results on common graphs to this asymmetric setting, including a proof that if H_1 contains K_4 then the pair (H_1, H_2) is not $(p, 1-p)$-common for any p. In addition, we obtain several novel results for common pairs of graphs with no natural analogue in the symmetric setting. This work uses a wide range of techniques, including graph limits, entropy methods (allowing us to take one $(p, 1-p)$-common pair of graphs (F_1, F_2) and build others) and flag algebras (to prove that (C_4, C_5) is $(1/2, 1/2)$-common and $(1/3, 2/3)$-common, but not $(.52, .48)$-common – in particular, the set of p for which the pair is $(p, 1-p)$-common is something non-trivial).

There are many compelling open problems on this area, a dozen of which we state in our paper. The following question is very natural and particularly intriguing.

Conjecture 1.2. *Let H_1 and H_2 be graphs. The set $\{p \in (0,1) : (H_1, H_2)$ is $(p, 1-p)$-common$\}$ is always an interval.*

One of our other conjectures related to the idea of strongly common graphs. We call a graph H *strongly common*, if for any graphons W_1 and W_2 such that $W_1 + W_2 = 1$, we have

$$t(H, W_1) + t(H, W_2) \geq t(K_2, W_1)e(H) + t(K_2, W_2)e(H).$$

Any graph that is Sidorenko is strongly common, and any strongly common graph is common. Goodman's work [16] shows that a triangle is strongly common. We proved that C_5 is strongly common and conjectured that all odd cycles are. This conjecture was proved very recently by Kim and Lee [21] using the method of Schur convexity. We intend to further explore the use of this valuable and new-to-us technique and to test the limits of what can be proved by using it in conjunction with our other methods.

2 Future Collaboration

The mathematics department at Warwick University is one of the strongest in the world, and I would particularly relish the opportunity to work with Richard Montgomery, who has myriad exciting results in extremal combinatorics and graph factorizations in particular, including a recent celebrated proof of Ringel's long-standing conjecture. The research project 'Spanning Subgraphs in Graphs' focuses on the two inter-related areas of spanning subgraphs in random graphs and in coloured graphs, encompassing many fascinating combinatorial problems.

I have prior experience working on spanning subgraphs through my work on 1-factorizations. Moreover, my recent research on finding the threshold for subgraph emergence in semi-random graphs involved probabilistic tools commonly also used on purely random graphs, as well as extra arguments to deal with the player strategy, and I have further familiarity with probabilistic methods from my research into probabilistic zero-forcing.

In addition to Richard Montgomery, there are other researchers at Warwick university with whom I share a number of interests, including Oleg Pikhurko, whose results on graph saturation I built on directly in my own work, and whose expertise on dense graph limits and the associated techniques would be invaluable for any continuing work on common pairs of graphs or other similar projects. I would potentially be interested in taking advantage of the DIMAP program to work with colleagues in the computer science department. Dmitry Chistikov and Marcin Jurdzinski for example have experience working on problems related Černý's conjecture, and I would hope that by pooling our respective expertise we could make further progress towards solving the conjecture.

References

[1] Dan Archdeacon. Problems in topological graph theory. `http://www.cems.uvm.edu/TopologicalGraphTheoryProblems/perfectq.htm`, 1995.

[2] Michael D. Barrus, Michael Ferrara, Jennifer Vandenbussche, and Paul S. Wenger. Colored saturation parameters for rainbow subgraphs. *J. Graph Theory*, 86(4):375–386, 2017.

[3] Natalie Behague. Hypergraph saturation irregularities. *Electron. J. Combin.*, 25(2):Paper 2.11, 13, 2018.

[4] Natalie Behague. Semi-perfect 1-factorizations of the hypercube. *Discrete Mathematics*, 342(6):1696–1702, 2019.

[5] Natalie Behague and Robert Johnson. Synchronizing times for k-sets in automata. *Electron. J. Combin.*, 29(3):Paper No. 3.41, 23, 2022.

[6] Natalie Behague, Tom Johnston, Shoham Letzter, Natasha Morrison, and Shannon Ogden. The rainbow saturation number is linear. `https://arxiv.org/abs/2211.08589`, 2022.

[7] Natalie Behague, Trent G. Marbach, and Paweł Prałat. Tight bounds on probabilistic zero forcing on hypercubes and grids. *Electron. J. Combin.*, 29(1):Paper No. 1.3, 20, 2022.

[8] Natalie Behague, Trent G. Marbach, Pawel Pralat, and Andrzej Rucinski. Subgraph games in the semi-random graph process and its generalization to hypergraphs. `https://arxiv.org/abs/2105.07034`, 2021.

[9] Natalie Behague, Natasha Morrison, and Jonathan A. Noel. Common pairs of graphs. `https://arxiv.org/abs/2208.02045`, 2022.

[10] Omri Ben-Eliezer, Dan Hefetz, Gal Kronenberg, Olaf Parczyk, Clara Shikhelman, and Miloš Stojaković. Semi-random graph process. *Random Structures & Algorithms*, 56(3):648–675, 2020.

[11] Vaithiyalingam Chitra and Appu Muthusamy. A note on semi-perfect 1-factorization and Craft's conjecture. *Graph Theory Notes N. Y.*, 64:58–62, 2013.

[12] P. Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, 53:292–294, 1947.

[13] António Girão, David Lewis, and Kamil Popielarz. Rainbow saturation of graphs. *J. Graph Theory*, 94(3):421–444, 2020.

[14] Vasil S. Gochev and Ivan S. Gotchev. On k-semiperfect 1-factorizations of Q_n and Craft's conjecture. *Graph Theory Notes N. Y.*, 58:36–41, 2010.

[15] François Gonze and Raphaël M. Jungers. On the synchronizing probability function and the triple rendezvous time for synchronizing automata. *SIAM J. Discrete Math.*, 30(2):995–1014, 2016.

[16] A. W. Goodman. On sets of acquaintances and strangers at any party. *Amer. Math. Monthly*, 66:778–783, 1959.

[17] AIM Minimum Rank-Special Graphs Work Group. Zero forcing sets and the minimum rank of graphs. *Linear Algebra Appl.*, 428(7):1628–1648, 2008.

[18] D. Hanson and B. Toft. Edge-colored saturated graphs. *J. Graph Theory*, 11(2):191–196, 1987.

[19] C. Jagger, P. Šťovíček, and A. Thomason. Multiplicities of subgraphs. *Combinatorica*, 16(1):123–141, 1996.

[20] L. Kászonyi and Zs. Tuza. Saturated graphs with minimal number of edges. *J. Graph Theory*, 10(2):203–210, 1986.

[21] Jang Soo Kim and Joonkyung Lee. Extended commonality of paths and cycles via schur convexity. `https://arxiv.org/abs/2210.00977`, 2022.

[22] Rastislav Královič and Richard Královič. On semi-perfect 1-factorizations. In *Structural information and communication complexity*, volume 3499 of *Lecture Notes in Comput. Sci.*, pages 216–230. Springer, Berlin, 2005.

[23] P. J. Laufer. On strongly Hamiltonian complete bipartite graphs. *Ars Combin.*, 9:43–46, 1980.

[24] Oleg Pikhurko. Results and open problems on minimum saturated hypergraphs. *Ars Combin.*, 72:111–127, 2004.

[25] A. Thomason. Graph products and monochromatic multiplicities. *Combinatorica*, 17(1):125–134, 1997.

[26] Zsolt Tuza. Extremal problems on saturated graphs and hypergraphs. *Ars Combin*, 25:105–113, 1988.