Problems that I would like Somebody to Solve

Sam Spiro

December 2, 2022

Contents

1 What is this?

This is a (very informal) collection of problems that are of personal interest to me, most of which are lesser known problems. Feel free to reach out to me if you have any questions regarding these questions and/or if you spot any glaring typos. I also **strongly** recommend reaching out to me if you start seriously working on any of these problems so that I don't accidentally work on it at the same time (which has happened before!).

Organization. Section 2 contains the main problems that I'm interested in. Section 3 consists of some smaller problems, many of which are relatively elementary to state and might end up having very simple solutions. Section 4 contains the "Hall of Fame" list of solutions to past problems.

The order that the problems appear is essentially chronological. Note that problems can jump between being labeled "main" or "small" depending on my current mood, and I may remove problems from the list if I start actively working on them again.

Acknowledgments. We thank Zachary Chase for correcting some small typos.

2 The Main Problems

2.1 Coloring mod p

Given a graph G and an integer p, we say that $I \subseteq G$ is an *independent set mod* p if every vertex in the induced graph $G[I]$ has degree $0 \mod p$. For example, independent sets are always independent sets mod p. We define the *mod* p *independence number* $\alpha_p(G)$ to be the size of a largest independent set mod p. Similarly we define the *mod* p *chromatic number* $\chi_p(G)$ to be the smallest integer k such that there exists a partition $V_1 \cup \cdots \cup V_k$ of $V(G)$ such that V_i is an independent set mod p for all i.

Conjecture 2.1. *For all primes p, there exists a constant $C = C(p)$ such that for all graphs G, $\chi_p(G) \le C$.*

It's quite plausible that the conjecture is true without having to restrict to primes, but focusing on primes is probably a good place to start since one can most easily use algebraic techniques in this case.

Gallai proved that Conjecture 2.1 is holds with $C = 2$ when $p = 2$, see [13] for a simple proof, as well as [8] for two other proofs written in a different language [1] Caro, Krasikov, and Roditty [2] proved a weaker version of Conjecture 2.1, showing that G can be partitioned into C induced subgraphs $G[V_1], \ldots, G[V_C]$ such that $e(G[V_i]) \equiv 0 \mod p$ for all i. Ferber, Hadiman and Krivelevich [7] showed that there exists a C such that almost every graph has $\chi_p(G) \le C$.

Overall Conjecture 2.1 seems pretty hard, and there are a couple of weaker versions of this conjecture that might be provable.

Conjecture 2.2. *For all primes p, there exists a constant $C = C(p)$ such that for all graphs G, $\alpha_p(G) \ge |V(G)|/C$.*

Conjecture 2.3. *For all primes p, there exists a constant $C = C(p)$ such that for all graphs G, one can partition $V(G)$ into C sets $V_1 \cup \cdots \cup V_C$ such that no $G[V_i]$ contains a vertex of degree $1 \mod p$.*

[1] This reference gives three proofs that there exists a solution to the "Lights Out!" game. It is relatively easy to show that this implies the stated result by considering a graph G' with a leaf attached to each vertex.

It also natural to conjecture this for $-1 \mod p$, since in both cases we know the result holds for $p = 2$.

Lastly, we note that a trivial lower bound on the $C(p)$ in Conjecture 2.1 is $C(p) \geq p$ by considering $G = K_p$. However, for odd p one can prove that we must have $C(p) \geq p+1$ (there are many examples; the simplest is to take a circulant graph on $2p+2$ vertices such that every vertex has degree $p+1$). It would be interesting to know if one could find constructions which give significantly stronger bonds.

2.2 Small Quasikernels

Let D be a digraph. Given a set S, we define $N^+(S) = \bigcup_{v \in S} N^+(v)$, where $N^+(v)$ is the out-neighborhood of v. We say that a set $K \subseteq V(D)$ is a *kernel* of D if (1) $N^+(K) \cap K = \emptyset$ (that is, K is an independent set of the underlying graph of D), and (2) $N^+(K) \cup S = V(D)$ (that is, every vertex is either in K or can be reached by a vertex in K in one step).

Not every digraph has a kernel (take any directed cycle of odd length), but it is not too hard to prove that every digraph has a *quasikernel*. This is a set $Q \subseteq V(D)$ such that (1) $N^+(Q) \cap Q = \emptyset$ and such that (2) $N^+(N^+(Q)) \cup N^+(Q) \cup Q = V(D)$. That is, it is an independent set such that every vertex can be reached from Q in at most two steps.

Given that every digraph has a quasikernel, it is natural to ask how small of a quasikernel one can find. One quickly realizes that it can be quite large: any source of D must belong to a quasikernel of D. Thus the most natural setting to consider is when D has no sources, and in this case the following was conjectured by Erdős and Székely.

Conjecture 2.4. *Every digraph D with no sources has a quasikernel of size at most $|V(D)|/2$.*

Some progress has been made, see [11], but overall the conjecture is far from being resolved.

2.3 C_4-free Subgraphs of Random Hypergraphs

Given a hypergraph H and a family of hypergrpahs $\mathcal{F}$, we define $\mathrm{ex}(H, \mathcal{F})$ to be the maximum number of edges in an $\mathcal{F}$-free subgraph of H. We're particularly interested in the case when $H = G^r_{n,p}$, the random r-uniform hypergraph obtained by keeping each edge of K^r_n independently and with probability p and when $\mathcal{F}$ is a family of r-partite r-graphs.

Perhaps the simplest non-trivial case of this problem is when we consider C_4-free subgraphs of the random graph $G_{n,p}$. This problem was essentially solved by Füredi [9], and later two more solutions were given by Morris and Saxton [14] (who essentially solved the problem for both graph cycles and complete bipartite graphs). The problem in this section is concerned about extending these results to hypergraphs C_4's; which remains an elusive problem despite having multiple proofs in the graph setting. There are many ways one can define what it means for a hypergraph to be a "C_4", below we consider two common notions.

Let C^3_4 be the 3-uniform loose 4-cycle, which can be defined by having edges

$$\{1, 2, 3\}, \{3, 4, 5\}, \{5, 6, 7\}, \{7, 8, 1\}.$$

That is, it's obtained from the graph C_4 by inserting an extra vertex into each edge. A standard deletion argument shows that, for example,

$$\mathbb{E}[\mathrm{ex}(G^3_{n,n^{-2/3}}, C^3_4)] = \Omega(n^{4/3}),$$

and work of Mubayi and Yepremyan [15] shows[2]

$$\mathbb{E}[\mathrm{ex}(G^3_{n,n^{-2/3}}, C^3_4)] \leq n^{13/9+o(1)}.$$

Problem 2.5. *Improve either of these bounds for* $\mathbb{E}[\mathrm{ex}(G^3_{n,n^{-2/3}}, C^3_4)]$.

Mubayi and Yepremyan [15] conjecture that the lower bound from the deletion argument is essentially best possible, and they provide some evidence that points in this direction. On the other hand, Nie, Verstraëte, and myself [16] proved that the analogous problem for loose triangles has a stronger lower bound than what you get from a deletion argument, which suggests that improvements might be possible for all loose cycles. Overall it's very unclear what the correct answer should be here.

In another direction, we say that a 3-uniform hypergraph F is a Berge C_4 if it has four edges e_1, e_2, e_3, e_4 and if there exist four distinct vertices v_1, v_2, v_3, v_4 with $v_i \in e_i \cap e_{i+1}$ for all i (with the indices written cyclically). We let $\mathcal{B}^3_4$ denote the set of 3-uniform Berge C_4's.

Verstraëte and I [20] showed

$$\mathbb{E}[\mathrm{ex}(G^3_{n,p}, \mathcal{B}^3_4)] \geq p^{1/4} n^{3/2-o(1)} \text{ for } p \gg n^{-2/3},$$

and that $\mathbb{E}[\mathrm{ex}(G^3_{n,n^{-2/3}}, \mathcal{B}^3_4)] \leq p^{1/6} n^{3/2+o(1)}$ for $p \gg n^{-2/3}$; and I believe that using a significantly more complicated argument I can prove

$$\mathbb{E}[\mathrm{ex}(G^3_{n,p}, \mathcal{B}^3_4)] \leq p^{1/5} n^{3/2+o(1)} \text{ for } p \gg n^{-2/3}.$$

Problem 2.6. *Improve either of these bounds for* $\mathbb{E}[\mathrm{ex}(G^3_{n,p}, \mathcal{B}^3_4)]$.

We note that improving the lower bound of Problem 2.5 is strictly easier than improving the lower bound of Problem 2.5, and if the lower bound of Problem 2.5 is tight (as conjectured by [15]), then it would be easier (in principle) to show that the lower bound for Problem 2.5 is tight.

2.4 Maximal Independent Sets of Clique-free Graphs

We say that a set of vertices $I \subseteq V(G)$ of a graph G is a *maximal independent set*, or simply an MIS, if I is an independent set but $I \cup \{v\}$ is not an independent set for any $v \notin I$. Let $m_t(n,k)$ denote the maximum number of MIS's of size k that an n-vertex K_t-free graph can have.

[2] They proved bounds in a much larger range, but this is the point where the gap between the bounds is largest.

We initiated the study of $m_t(n,k)$ together with He and Nie [10] (and we refer the reader to our paper for motivation of this particular problem). We stated a lot of open problems about this function in our paper, any of which I would be thrilled to see solved. Here we emphasize two of these problems.

Our first problem concerns upper bounding the number of MIS's in triangle-free graphs.

Problem 2.7. *Prove that there exists an integer $k \geq 5$ and real number $\epsilon > 0$ such that*

$$m_3(n,k) = O(n^{k-2-\epsilon}).$$

We conjectured that in fact $m_3(n,k) = \Theta(n^{\lfloor k/2 \rfloor})$ for all $k \geq 5$, and we implicitly proved $m_3(n,k) = O(n^{k-2})$ for $k \geq 5$. Thus this problem asks to improve our upper bound, which our conjectured lower bound suggests should be very far from tight as is.

The next problem concerns K_4-free graphs. In this setting we proved $m_4(n,3) \geq n^{2-o(1)}$ and that $m_4(n,3) = O(n^2)$.

Problem 2.8. *Determine whether the $o(1)$ term in the lower bound for $m_4(n,3)$ mentioned above is necessary or not.*

I believe that this $o(1)$ should be necessary. In fact, I believe that $m_4(n,3)$ should be equal (up to constants) to the maximum number of edges of an n-vertex graph which is such that every edge is contained in a unique triangle (determining this quantity is often referred to as the Ruzsa-Szemerédi problem). One approach that would give this stronger result is the following.

Problem 2.9. *Show that if G is an n-vertex K_4-free graph with "many" (e.g. $n^{2-\epsilon}$) MIS's of size 3, then $\chi(G) = O(1)$.*

If this were true then one could essentially convert the problem of working with K_4-free graphs to working with tripartite graphs, and in this case we proved that the maximum number of 3-MIS's is essentially the solution to the Ruzsa-Szemerédi problem. We note that it's easy to prove that if G is an n-vertex *triangle*-free graphs with at least one MIS of size k that $\chi(G) \leq k+1$, and in particular for K_4-free graphs one may need much fewer than $n^{2-\epsilon}$ MIS's to guarantee a bounded chromatic number.

3 Smaller Problems

3.1 Slow Tribonacci Walks

Given a triple of positive integers (w_1, w_2, w_3), recursively define $w_k = w_{k-1} + w_{k-2} + w_{k-3}$ for $k \geq 4$. We say that (w_1, w_2, w_3) is an *n-tribonacci walk* if $w_s = n$ for some s. There are infinitely many n-tribonacic walks, e.g. those of the form $(42, n, x)$ for any x. To make things more interesting, we say that (w_1, w_2, w_3) is an *n-slow tribonacci walk* if $w_s = n$ with s as large as possible. For example, $(1,1,1)$ and $(42,3,42)$ are both 3-tribonacci walks, but only the first one is slow. Let $p(n)$ denote the number of n-slow tribonacci walks. For example, it's easy to check that $p(3) = 1$, and $p(1) = p(2) = \infty$

Question 3.1. *Does there exist some absolute constant c such that either $p(n) = \infty$ or $p(n) \leq c$ for all n?*

If we instead look at Fibonacci walks (which are defined using the Fibonacci recurrence $w_k = w_{k-1} + w_{k-2}$), then one can show that $p(n) \leq 2$ for all $n \geq 2$ [4]. More generally if one looks at walks following a two-term recurrence of the form $w_k = \alpha w_{k-1} + \beta w_{k-2}$ with $\alpha, \beta \geq 1$ relatively prime, then $p(n) \leq \alpha^2 + 2\beta - 1$ for all but finitely many n [19].

Computational data made it easy to conjecture the correct answer for two-term recurrences, but the situation is less clear for slow tribonacci walks. For example, $p(61) = 9$, which is fairly large given how small 61 is.

I'll note that there are many other interesting problems related to the behavior of slow recurrences that were left unanswered in [4, 19]. However, as is often the case in number theory, these relatively easy to state questions are likely very difficult to solve. This being said, I do think that this tribonacci problem is tractable.

3.2 An Adversarial Chernoff Bound

Update: A positive answer to this problem has been found; I will update further with details.

Persi Diaconis, Ron Graham, Xiaoyu He, and myself [5] proved the following.

Theorem 3.2 ([5]). *Let X_i be independent Bernoulli random variables with $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1 - p$. Let $S_t = \sum_{i \leq t} X_i$. There exist absolute constants c_0, c_1 such that for all $\lambda > 0$ and integers $k_1 \geq k_0 \geq 2\lambda^{-1}$,*

$$\Pr[\exists t \in [k_0, k_1] : |S_t - pt| \geq \lambda pt] \leq \frac{c_0 k_1}{k_0} e^{-c_1 \lambda^3 p k_0}.$$

That is, with high probability, for every t in the interval $[k_0, k_1]$, every partial sum S_t is close to its expectation. This is immediate for any given value of t by the Chernoff bound (since each S_t is a binomial random variable), but it does *not* follow from the Chernoff bound and a naive application of the union bound (this gives a bound like $k_1 e^{-\lambda^2 p k_0}$, which is much weaker if k_0 is very large).

Question 3.3. *Does the bound of Theorem 3.2 hold with λ^2 instead of λ^3?*

Note that λ^2 would be best possible because this is what one gets if $k_0 = k_1$. While the statement of Theorem 3.2 is fairly technical, the proof itself only required a slightly clever application of the union bound together with the Chernoff bound, so my hope is that more sophisticated probabilistic tools can be used to solve Question 3.3 without too much difficulty.

Secretly I'm interested in this because it would improve upon the error term for our main result in [5], but also I just think it's of independent interest to determine how much concentration one can get for an "adversarial" binomial distribution.

3.3 Saturation Games

For a family of graphs $\mathcal{F}$, we say that a graph G is $\mathcal{F}$-saturated if G contains no graph of $\mathcal{F}$ as a subgraph, but adding any edge to G would create a subgraph of $\mathcal{F}$. The $\mathcal{F}$-saturation game consists of two players, Max and Mini, who alternate turns adding edges to an initially empty graph G on n vertices (say with Max starting), with the only restriction being that G is never allowed to contain a subgraph that lies in $\mathcal{F}$. The game ends when G is $\mathcal{F}$-saturated. The payoff for Max is the number of edges in G when the game ends, and Mini's payoff is the opposite of this. We let $\mathrm{sat}_g(n, \mathcal{F})$ denote the number of edges that the graph in the $\mathcal{F}$-saturation game ends with when both players play optimally, and we call this quantity the game $\mathcal{F}$-saturation number.

Bounding $\mathrm{sat}_g(\mathcal{F}; n)$ seems to be pretty hard in general, and even the original problem of determining $\mathrm{sat}_g(n, C_3)$ is still wide open. See [17] for further history and known bounds. In [17] I proved $\mathrm{sat}_g(n, \mathcal{C}_\infty^o \setminus \{C_3\}) = O(n)$, where $\mathcal{C}_\infty^o$ is the set of all odd cycles. I also proved (somewhat indirectly) that $\mathrm{sat}_g(n, \mathcal{C}_\infty^o \setminus \{C_{2k+1}\}) = \Omega(n^2)$ for $k \geq 3$. Given this, it is natural to ask the following.

Problem 3.4. *Prove non-trivial bounds on* $\mathrm{sat}_g(n, \mathcal{C}_\infty^o \setminus \{C_5\})$*, where* $\mathcal{C}_\infty^o = \{C_3, C_5, C_7, \ldots\}$*.*

I'd be happy to have even an $\omega(n)$ lower bound or any non-trivial asymptotic upper bound for this problem. Possibly a more tractable problem is the following.

Problem 3.5. *Prove non-trivial bounds on* $\mathrm{sat}_g(n, C_k)$ *for* $k > 3$*.*

For odd k, I proved an asymptotic upper bound of $\mathrm{sat}(n, C_k) \leq \frac{4}{27}n^2 + o(n^2)$. In [3] the authors proved a non-trivial lower bound for C_4 if you play a "bipartite" version of the game, but I'd still like to see bounds proved in the original setting. Lastly, I'd like to know the following.

Question 3.6. *Does there exist a finite set of (odd) cycles* $\mathcal{C}$ *such that* $\mathrm{sat}_g(n, \mathcal{C}) = O(n)$*?*

3.4 Card Guessing with Adversarial Shuffling

Consider the following game. We start with a deck of mn cards consisting of n different card types each appearing m times (e.g. $m = 4, n = 13$ corresponds to a standard deck of cards). First, one of the players (Shuffler) shuffles the deck however they'd like. Then the other player (Guesser) sequentially guesses what the top card of the deck is. After each guess, the Guesser is told only whether their guess was correct or not, and then the top of the card is discarded. This game is called the *offline partial feedback model*, and the score at the end of the game is equal to the number of times Guesser correctly guesses a card type. One can also consider the *online partial feedback model* where Shuffler is allowed to reshuffle the remaining cards in the deck each time Guesser makes a guess.

Question 3.7. *Assuming* $n \gg m$*, can Guesser play in the offline partial feedback model so that they get* $m + \omega(1)$ *points in expectation? Can they play in the offline partial feedback model so that they get* $m + \Omega(1)$ *points in expectation?*

Simple strategies that Guesser can use in either model are to either guess a single card type each round, or to randomly guess a card type each round. Both strategies give Guesser m points in expectation regardless of Shuffler's strategy. In [18] I came up with a strategy giving at least $m + 1/2$ points in the offline model (and an easy adaptation of the argument gives $m + e - 2$), as well as a strategy giving just a smidge more than m in the online model; but the situation is pretty pitiful overall.

Note that in [5] we showed that if Shuffler shuffles the deck uniformly at random, then Guesser can not do better than $m + O(m^{3/4} \log m)$ points in expectation, and in [6] we showed that there exists a strategy in this setting giving $m + \Omega(m^{1/2})$ points in expectation. Thus these provide some reasonable benchmarks on how well one might be able to do here.

4 Hall of Fame

Note: solving any of my "Main Problems" guarantees addition to this list. Solutions to "Small Problems" may or may not guarantee addition to the list depending on how emotionally invested I was in the problem/how non-trivial and elegant the solution ends up being.

4.1 Main Problems

A list of people who have successfully solved any of my main problems.

- Wang and Zhao [21] for solving my original conjecture on ballot permutations; and Lin, Wang, and Zhao [12] for solving an even stronger version!

4.2 Small Problems

- Alon and Kravitz [1] for solving my problem with Greg Patchell about the number of CAT's one can pack into a cube filled with letters (and the extension to arbitrary words of distinct letters!).

References

[1] Noga Alon and Noah Kravitz. Cats in cubes. *arXiv preprint arXiv:2211.14887*, 2022.

[2] Y Caro, I Krasikov, and Y Roditty. Zero-sum partition theorems for graphs. *International Journal of Mathematics and Mathematical Sciences*, 17(4):697–702, 1994.

[3] James M Carraher, William B Kinnersley, Benjamin Reiniger, and Douglas B West. The game saturation number of a graph. *Journal of Graph Theory*, 85(2):481–495, 2017.

[4] Fan Chung, Ron Graham, and Sam Spiro. Slow fibonacci walks. *Journal of Number Theory*, 210:142–170, 2020.

[5] Persi Diaconis, Ron Graham, Xiaoyu He, and Sam Spiro. Card guessing with partial feedback. *arXiv preprint arXiv:2010.05059*, 2020.

[6] Persi Diaconis, Ron Graham, and Sam Spiro. Guessing about guessing: Practical strategies for card guessing with feedback. *arXiv preprint arXiv:2012.04019*, 2020.

[7] Asaf Ferber, Liam Hardiman, and Michael Krivelevich. On subgraphs with degrees of prescribed residues in the random graph. *arXiv preprint arXiv:2107.06977*, 2021.

[8] Rudolf Fleischer and Jiajin Yu. A survey of the game "lights out!". In *Space-efficient data structures, streams, and algorithms*, pages 176–198. Springer, 2013.

[9] Zoltán Füredi. Random ramsey graphs for the four-cycle. *Discrete Mathematics*, 126(1-3):407–410, 1994.

[10] Xiaoyu He, Jiaxi Nie, and Sam Spiro. Maximal indpendent sets of clique-free graphs. *arXiv preprint arXiv:2107.09233*, 2021.

[11] Alexandr Kostochka, Ruth Luo, and Songling Shan. Towards the small quasi-kernel conjecture. *arXiv preprint arXiv:2001.04003*, 2020.

[12] Zhicong Lin, David G.L. Wang, and Tongyuan Zhao. A decomposition of ballot permutations, pattern avoidance and gessel walks. *arXiv preprint arXiv:2103.04599*, 2021.

[13] László Lovász. *Combinatorial problems and exercises*, volume 361. American Mathematical Soc., 2007.

[14] Robert Morris and David Saxton. The number of $c_{2\ell}$-free graphs. *Advances in Mathematics*, 298:534–580, 2016.

[15] Dhruv Mubayi and Liana Yepremyan. Random turán theorem for hypergraph cycles. *arXiv preprint arXiv:2007.10320*, 2020.

[16] Jiaxi Nie, Sam Spiro, and Jacques Verstraete. Triangle-free subgraphs of hypergraphs. *arXiv preprint arXiv:2004.10992*, 2020.

[17] Sam Spiro. Saturation games for odd cycles. *The Electronic Journal of Combinatorics*, pages P4–11, 2019.

[18] Sam Spiro. Online card games. *arXiv preprint arXiv:2106.11866*, 2021.

[19] Sam Spiro. Slow recurrences. *Journal of Number Theory*, 218:370–401, 2021.

[20] Sam Spiro and Jacques Verstraëte. Counting hypergraphs with large girth. *arXiv preprint arXiv:2010.01481*, 2020.

[21] David GL Wang and Tongyuan Zhao. The peak and descent statistics over ballot permutations. *arXiv preprint arXiv:2009.05973*, 2020.