Tangled Paths: A Random Graph Model from Mallows Permutations

(Extended abstract)

Jessica Enright[*]   Kitty Meeks[†]   William Pettersson[‡]   John Sylvester[§]

Abstract

We introduce the random graph $\mathcal{P}(n, q)$ which results from taking the union of two paths of length $n \geqslant 1$, where the vertices of one of the paths have been relabelled according to a Mallows permutation with real parameter $0 < q(n) \leqslant 1$. This random graph model, the tangled path, goes through an evolution: if q is close to 0 the graph bears resemblance to a path, and as q tends to 1 it becomes an expander. In an effort to understand the evolution of $\mathcal{P}(n, q)$ we determine the treewidth and cutwidth of $\mathcal{P}(n, q)$ up to log factors for all q. We also show that the property of having a separator of size one has a sharp threshold. In addition, we prove bounds on the diameter, and vertex isoperimetric number for specific values of q.
DOI: `https://doi.org/10.5817/CZ.MUNI.EUROCOMB23-000`

Introduction

Given two graphs G, H on a common vertex set $[n] = \{1, \ldots, n\}$, and a permutation σ on $[n]$, it is natural to consider the following graph

$$\mathsf{layer}(G, \sigma(H)) = ([n], \ E(G) \cup \{\sigma(x)\sigma(y) : xy \in E(H)\}),$$

which is the union of two graphs where the second graph has been relabelled by a permutation σ. This general and powerful construction has featured previously in the literature in

---
[*]School of Computing Science, University of Glasgow, UK. E-mail: `jessica.enright@glasgow.ac.uk`.
[†]School of Computing Science, University of Glasgow, UK. E-mail: `kitty.meeks@glasgow.ac.uk`.
[‡]School of Computing Science, University of Glasgow, UK. E-mail: `william.pettersson@glasgow.ac.uk`.
[§]Dept. of Computing Science, University of Liverpool, UK. E-mail: `john.sylvester@liverpool.ac.uk`.

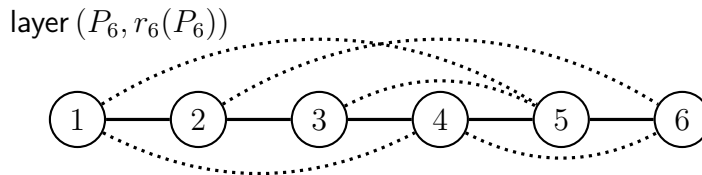| n	v_n	r_n
1	1	1
2	2	12
3	1	312
4	3	3142
5	2	35142
6	5	351462

layer $(P_6, r_6(P_6))$

Figure 1: The table on the left gives the sequences of permutations (r_n) generated by the sequence (v_n) for $i = 1, \ldots, 6$. On the right we have a tangled path generated by r_6, where the edges of $r_6(P_6)$ are dotted.

the contexts of constructions and decompositions of graphs [1, 4, 22]. Let P_n be the path on $[n]$ and S_n be the set of all permutations on $[n]$. Consider the following scenario: one must choose a permutation $\sigma \in S_n$ with the goal of making $\mathsf{layer}(P_n, \sigma(P_n))$ as different from a path as possible. There are several parameters one may use to measure the difference between a connected graph G and a path; for example one may look at the diameter $\mathsf{diam}(G)$ or the vertex isoperimetric number $\mathsf{i}(G)$, as the path is extremal for these parameters. The treewidth $\mathsf{tw}(G)$ which, broadly speaking, measures how far (globally) the graph is from being a tree [14], is another natural candidate. It is fairly easy to see that given two or more paths one can build a grid-like graph (see [7, Lemma 8] for more details), and such a graph would have treewidth and diameter $\Theta(\sqrt{n})$. If we choose a permutation uniformly at random, then as a consequence of a result of Kim & Wormald [12, Theorem 1], with high probability the resulting graph is a bounded degree expander. Thus, in this case, the graph $\mathsf{layer}(P_n, \sigma(P_n))$ has treewidth $\Theta(n)$ and diameter $\Theta(\log n)$, so by these parameters it is essentially as far from a path as a sparse graph can be.

The example above shows that even restricting the input graphs to paths can produce rich classes of graphs. Having seen what happens for a uniformly random permutation, one may ask about the structure of $\mathsf{layer}(P_n, \sigma(P_n))$ when σ is drawn from a distribution on S_n that is not uniform. One of the most well known non-uniform distributions on S_n is the Mallows distribution, introduced by Mallows [17] in the late 1950's in the context of statistical ranking theory. Recently it has been the subject of renewed interest for other settings [6, 3, 11], and as an interesting and natural model to study in its own right [2, 10, 20]. The distribution has a parameter q which, roughly speaking, controls the amount of disorder in the permutation.

Mallows Permutations. For real $q > 0$ and integer $n \geqslant 1$, the (n, q)-*Mallows measure* $\mu_{n,q}$ on S_n is given by

$$\mu_{n,q}(\sigma) = \frac{q^{\mathrm{Inv}(\sigma)}}{Z_{n,q}} \qquad \text{for any } \sigma \in S_n,$$

where $\mathrm{Inv}(\sigma) = |\{(i,j) : i < j \text{ and } \sigma(i) > \sigma(j)\}|$ is the number of inversions in the permutation σ and $Z_{n,q}$ is given explicitly by the following formula [2, Equation (2)]:

$$Z_{n,q} = \prod_{i=1}^{n} \left(1 + q + \cdots + q^{i-1}\right) = \prod_{i=1}^{n} \frac{1 - q^i}{1 - q}.$$

When $q \to 0$, the distribution $\mu_{n,q}$ converges weakly to the degenerate distribution on the identity permutation. We extend $\mu_{n,q}$ to $q = 0$ by setting $\mu_{n,0}$ to be the probability measure assigning 1 to the identity permutation. On the other hand if $q = 1$ then $\mu_{n,1}$ is the uniform measure on S_n. One can see that $\sigma \sim \mu_{n,q}$ has distribution $\mu_{n,1/q}$ when reversed.

A key feature of Mallows permutations is that they can be constructed by a simple procedure from a sequence of independent random variables. For any $q > 0$, the *Mallows Process* gives a sequence (r_n) such that $r_n \sim \mu_{n,1/q}$, for each $n \geqslant 1$. Furthermore each r_n is constructed from r_{n-1} by inserting n at a position in r_{n-1} sampled via a simple distribution that is independent from $r_1, \ldots, r_{n-1}$. Several desirable properties of Mallows permutations can be deduced from this construction, see [10, Sec. 2] for more details.

The Tangled Paths Model. We study the random graph distribution induced by $\mathsf{layer}(P_n, \sigma(P_n))$, where $\sigma \sim \mu_{n,q}$ and $0 \leqslant q := q(n) \leqslant 1$. From now on we call this graph distribution the *tangled path* model and denote it by $\mathcal{P}(n, q)$. Thus a random graph $\mathcal{P}(n, q)$ has vertex set $[n]$ and (random) edge set $E(P_n) \cup \{\sigma(i)\sigma(i+1) : i \in [n-1]\}$, where $\sigma \sim \mu_{n,q}$. We restrict to $q \in [0, 1]$ as reversing the permutation does not affect our construction (up to a relabelling). We also identify any multi-edges created as one edge, however this detail is not important for any of our results. This paper will focus on $\mathcal{P}(n, q)$; as we have seen already combining paths can give rise to interesting and varied graphs, and the Mallows permutation gives our model a parameter q which, roughly speaking, increases the 'tangled-ness' of the graph.

We see, from above, that $\mathcal{P}(n, 0)$ is a path and $\mathcal{P}(n, 1)$ is an expander with high probability by [12, Theorem 1]. Our ultimate aim is to understand the structure of $\mathcal{P}(n, q)$ for intermediate values of q, and this paper takes the first steps in this direction. Informally, one aspect of this is knowing when $\mathcal{P}(n, q)$ stops looking 'path-like'; we show that if $q < 1$ is fixed the diameter is linear (Theorem 3), and there is a sharp threshold for having a single vertex cut at $q_c = 1 - \frac{\pi^2}{6 \log n}$ (Theorem 2). For $q \to 1$ sufficiently fast, it makes more sense to measure the complexity of the internal structure of $\mathcal{P}(n, q)$ by how much it differs from a tree. In this direction we show that, up to logarithmic factors, the treewidth [14] of $\mathcal{P}(n, q)$ grows at rate $(1-q)^{-1}$ (Theorem 4) until the graph becomes an expander at around $q = 1 - \frac{1}{n \log n}$ (Theorem 1), indicating that, in the sense of treewidth, the complexity of the structure grows smoothly with q. This behaviour contrasts with the binomial/Erdős-Rényi random graph [8] where the treewidth increases rapidly from being bounded by a constant, to $\Theta(n)$ as the average degree rises from below one to above one [15].

Aside from this model being natural, motivation for this line of study comes from practical algorithmic applications. Many real-world systems – including social, biological
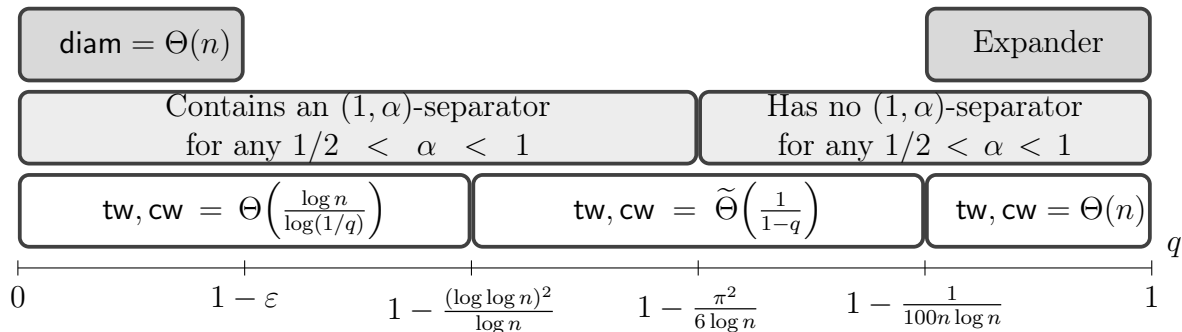
Figure 2: The diagram above gives a pictorial representation of our results. All results above hold with high probability, and we say that $f(n) = \widetilde{\Theta}(g(n))$ if $f(n) = \mathcal{O}(g(n)\log g(n))$ and $f(n) = \Omega(g(n)/\log g(n))$.

and transport networks – involve qualitatively different types of edges, where each type of edge generates a "layer" with specific structural properties [13, 19]. For example, when modelling the spread of disease in livestock, one layer of interest arises from physical adjacency of farms, and so is determined entirely by geography. A second epidemiologically-relevant layer could describe the pairs of farms which share equipment: this is no longer fully determined by geography, but will nevertheless be influenced by the location of farms, as those that are geographically close are more likely to cooperate in this way. It is known that algorithmically useful structure in individual layers of a graph is typically lost when the layers are combined adversarially [7]. The present work can be seen as an attempt to understand the structure of graphs generated from two simple layers which are both influenced to some extent by a shared underlying "geography". In this setting the treewidth tw is a natural parameter as many NP-hard problems become tractable when parametrised by tw [5, Ch. 7].

Our Results

In what follows, the integer $n \geqslant 1$ denotes the number of vertices in the graph (or elements in a permutation) and $q := q(n)$, the parameter of the Mallows permutation (or related tangled path), is a real valued function of n taking values in $[0, 1]$. We say a sequence of events $\mathcal{E}_n$ occurs *with high probability* (w.h.p.) if $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$. Throughout log is base e. See Figure 2 for a summary of our results.

A graph G is a *vertex-expander* if there exists a fixed real $c > 0$ such that any set $S \subseteq V$ with $|S| \leqslant \lceil n/2 \rceil$ is adjacent to at least $c|S|$ vertices in $V \backslash S$. As mentioned above, when $q = 1$ the permutation is uniform, and so the fact that $\mathcal{P}(n, 1)$ is an expander follows from [12, Theorem 1]. Our first result shows that for q sufficiently close to 1, this still holds.

Theorem 1. *If* $q \geqslant 1 - \frac{1}{100n \log n}$, *then w.h.p.* $\mathcal{P}(n, q)$ *is a bounded degree vertex-expander.*

For an integer $s \geqslant 1$ and real $1/2 \leqslant \alpha < 1$ we say G has an (s, α)-*separator* if there is a vertex subset S with $|S| \leqslant s$ such that $G \backslash S$ can be partitioned into two disjoint sets of at most $\alpha|V|$ vertices with no crossing edges. Balanced separators (e.g. $\alpha = 2/3$) are useful for designing divide and conquer algorithms, in particular for problems on planar graphs [16]. Balanced separators have intimate connections to notions of sparsity for graphs [18].

Observe that, for any fixed $1/2 < \alpha < 1$, if G is a vertex expander then there exists a $c > 0$ such that G has no (cn, α)-separator. At the other extreme, the path has an $(1, \alpha)$-*separator*. We show that for $\mathcal{P}(n, q)$ this 'path-like' property disappears around $q_c = 1 - \frac{\pi^2}{6 \log n}$.

Theorem 2. *For any fixed real $1/2 < \alpha < 1$ we have*

$$\lim_{n \to \infty} \mathbb{P}(\, \mathcal{P}(n, q) \ has \ a \ (1, \alpha)\text{-}separator\,) = \begin{cases} 0 & if \ \frac{\pi^2}{6(1-q)} - \log n + \frac{\log \log n}{2} \to \infty \\ 1 & if \ \frac{\pi^2}{6(1-q)} - \log n + \frac{5 \log \log n}{2} \to -\infty \end{cases}.$$

We say that q_0 is *sharp threshold* for a graph property $\mathfrak{P}$ if for any $\varepsilon > 0$ w.h.p. $\mathcal{P}(n, p) \notin \mathfrak{P}$ for any $p \leqslant q_0(1 - \varepsilon)$, and $\mathcal{P}(n, r) \in \mathfrak{P}$ for any $r \geqslant q_0(1 + \varepsilon)$, see [9]. Theorem 2 is quite precise as it determines the second order term in the threshold up to a constant, showing that the property of having a cut vertex has a sharp threshold of width $\mathcal{O}\left(\frac{\log \log n}{(\log n)^2}\right)$. Theorem 2 is established by finding first and second moment thresholds for the property. Positive correlation between cuts suggests this result cannot be significantly improved using standard methods alone.

The diameter $\mathsf{diam}(G)$ of a graph G is the greatest distance between any pair of vertices. Theorem 1 implies that $\mathsf{diam}(\mathcal{P}(n, q)) = \mathcal{O}(\log n)$ when q is close to 1. On the other hand $\mathsf{diam}(\mathcal{P}(n, 0)) = n - 1$ as it is a path; by applying bounds on the number of cut vertices, we show this holds (up to a constant) for any fixed $q < 1$.

Theorem 3. *Let $0 \leqslant q < 1$ be any function of n bounded away from 1. Then there exists a constant $c > 0$ such that w.h.p., $\mathsf{diam}(\mathcal{P}(n, q)) \geqslant cn$.*

The treewidth $\mathsf{tw}(G)$ of a graph G is the minimum size (minus one) of the largest vertex subset (i.e. bag) in a tree decomposition of G, minimised over all such decompositions [14]. The cutwidth $\mathsf{cw}(G)$ is the greatest number of edges crossing any real point under an injective function $\varphi : V \to \mathbb{Z}$, minimised over all φ. It is known that for any graph G we have $\mathsf{tw}(G) \leqslant \mathsf{cw}(G)$, however there may a multiplicative discrepancy of order up to n. We show there is at most only a constant factor discrepancy for $\mathcal{P}(n, q)$ for certain ranges of q and give bounds for all q which are tight up to log factors.

Theorem 4. *If there exists a real constant $\kappa > 0$ such that $0 \leqslant q \leqslant 1 - \kappa \cdot \frac{(\log \log n)^2}{\log n}$, then there exist constants $0 < c_1, C_2 < \infty$ such that w.h.p.*

$$c_1 \cdot \left(\sqrt{\frac{\log n}{\log(1/q)}} + 1 \right) \leqslant \mathsf{tw}(\mathcal{P}(n, q)) \leqslant C_2 \cdot \left(\sqrt{\frac{\log n}{\log(1/q)}} + 1 \right). \tag{1}$$

Furthermore, there exists some $c_3 > 0$ such that if $1 - \frac{(\log \log n)^2}{\log n} \leqslant q \leqslant 1 - \frac{1}{100n \log n}$, then w.h.p.

$$\frac{c_3}{1-q} \cdot \log\left(\frac{1}{1-q}\right)^{-1} \leqslant \mathsf{tw}(\mathcal{P}(n,q)) \leqslant \min\left\{\frac{5}{1-q} \cdot \log\left(\frac{1}{1-q}\right), 2n\right\}. \tag{2}$$

In addition, if $q \geqslant 1 - \frac{1}{100n \log n}$ then w.h.p.

$$\frac{n}{50} \leqslant \mathsf{tw}(\mathcal{P}(n,q)) \leqslant 2n. \tag{3}$$

The same upper and lower bounds in (1), (2), and (3) also hold for the cutwidth $\mathsf{cw}(\mathcal{P}(n,q))$.

Observe that if $q \to 1$ then $\log(1/q) \approx 1 - q$ and so when $q = 1 - \Theta\left((\log \log n)^2/\log n\right)$ we have $\sqrt{\log(n)/\log(1/q)} \approx -\log(1-q)/(1-q)$. Therefore, the first two upper bounds for cutwidth are equal up to constants for this value of q. Thus, for this q, the upper bound for the cutwidth given in (2) is tight and the lower bound for treewidth is off by a multiplicative factor of order $(\log \log n)^2$.

Proof Sketch of Theorem 4. The rough strategy for the lower bounds in (1) and (2) is as follows:

(i) relate containing a k-vertex expander as a minor in $\mathcal{P}(n,q)$, and thus $\Omega(k)$-treewidth, to a property of the underlying Mallows permutation or the random sequence generating it,

(ii) show this property holds, for a suitable k, with high probability by utilising the (asymptotic) independence of elements in a Mallows permutation or the sequence generating it.

However, the properties sought and method for controlling the probabilities in (1) and (2) differ slightly.

For Step (i), of the lower bound in (1), we show that if r_k and r_n are generated by sequences $x = x_1, \ldots, x_k$ and $y = y_1, \ldots, y_n$ respectively via the Mallows process, and x is contained in y as a consecutive sub-sequence, then $P_n \cup r_n(P_n)$ contains $P_k \cup r_k(P_k)$ as a minor. To prove (2) we instead show that if a permutation $\pi \in S_n$ contains a permutation $\sigma \in S_k$ as a consecutive pattern then $P_n \cup \pi(P_n)$ contains $P_k \cup \sigma(P_k)$ as a minor. In particular, both relations hold in the case where $P_k \cup \sigma(P_k)$ and $P_k \cup r_k(P_k)$ are expanders.

For Step (ii), the lower bound in (1) is shown using the second moment method to a given consecutive sub-sequence of k inputs occur w.h.p., whereas for (2) we use independence of permutations induced by disjoint intervals of elements in a Mallows permutation to show a given consecutive pattern occurs.

We now give a proof sketch for the upper bounds on $\mathsf{cw}\left(\mathcal{P}(n,q)\right)$ in (1) and (2). To begin, we fix the ordering $\varphi : [n] \to [n]$ in the definition of cutwidth to be the identity map.

That is, we order the vertices of $\mathcal{P}(n,q)$ along the line with respect to the ordering of the vertices given by the un-permuted path P_n. We then bound the number of edges crossing any vertex i by showing that not too many elements with values $j > i$ are inserted next to elements with values less than i by the Mallows process. To do this we show that, for $b = \Theta(\frac{\log n}{1-q})$ and some suitable $L \geqslant \ell$, within any consecutive sequence of L steps of the Mallows process (with generating inputs v_i) the following events hold with high probability:

(i) no insert position v_i has value greater than b,

(ii) after L steps the leftmost b places will each contain an element added added at most L steps ago,

(iii) there are at most ℓ values of v_i greater than ℓ.

The events (i) and (iii) ensure that not too many long edges are created from new entries being added far away from the left-hand end of the process. The event (ii) is a little bit more subtle but key to the success of our approach as it ensures that the left-hand end of the permutation grown by the Mallows process cannot retain entries that were inserted long ago, again preventing long edges caused by new elements lying next to old ones. If these three events hold then we can show that the number of edges crossing any vertex under the identity map is $\mathcal{O}(\ell)$. Optimising the choice of L and ℓ then gives the upper bounds in (1) and (2).

Open Problems

One could study the effect of q in $\mathcal{P}(n,q)$ on almost any graph property of interest for sparse graphs. One fundamental problem is to determine the number of edges in $\mathcal{P}(n,q)$ (recall that we disregard multi-edges). This deceptively non-trivial problem is related to clustering of consecutive numbers in Mallows permutations [21]. It would also be interesting to close the gap for treewidth by obtaining tight bounds for all q.

Theorem 2 proves that $q = 1 - \pi^2/(6\log n)$ is a sharp threshold for containing a single vertex whose removal separates the graph into two macroscopic components. A key open problem is to determine if there is a notion of monotone property in the setting of tangled paths which guarantees the existence of a threshold (or even a sharp threshold). One candidate feature (for a property to be monotone with respect to) is the number of inversions in the permutation generating $\mathcal{P}(n,q)$. However, one issue with parameterizing by the number of inversions is the fact that the tangled paths generated by $\sigma = (\sigma_1, \ldots, \sigma_n)$ and its reverse $\sigma^R = (\sigma_n, \ldots, \sigma_1)$ are isomorphic, but the number of inversions may differ greatly as $\mathrm{Inv}(\sigma^R) = \binom{n}{2} - \mathrm{Inv}(\sigma)$.

References

[1] Bogdan Alecu, Vadim Lozin, Daniel A. Quiroz, Roman Rabinovich, Igor Razgon, and Viktor Zamaraev. The treewidth and pathwidth of graph unions, 2022.

[2] Nayantara Bhatnagar and Ron Peled. Lengths of monotone subsequences in a Mallows permutation. *Probab. Theory Related Fields*, 161(3-4):719–780, 2015.

[3] Mark Braverman and Elchanan Mossel. Sorting from noisy information, 2009.

[4] Yijia Chen and Jörg Flum. On the ordered conjecture. In *Proceedings of the 27th Annual IEEE Symposium on Logic in Computer Science, LICS 2012*, pages 225–234. IEEE Computer Society, 2012.

[5] Marek Cygan, Fedor V. Fomin, Lukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michal Pilipczuk, and Saket Saurabh. *Parameterized Algorithms*. Springer, 2015.

[6] Persi Diaconis and Arun Ram. Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques. *Michigan Math. J.*, 48:157–190, 2000.

[7] Jessica Enright and Kitty Meeks. Two dichotomies for model-checking in multi-layer structures. *CoRR*, abs/1710.08758, 2017.

[8] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[9] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124(10):2993–3002, 1996.

[10] Alexey Gladkich and Ron Peled. On the cycle structure of Mallows permutations. *Ann. Probab.*, 46(2):1114–1169, 2018.

[11] Alexander E. Holroyd, Tom Hutchcroft, and Avi Levy. Mallows permutations and finite dependence. *Ann. Probab.*, 48(1):343–379, 2020.

[12] Jeong Han Kim and Nicholas C. Wormald. Random matchings which induce Hamilton cycles and hamiltonian decompositions of random regular graphs. *J. Comb. Theory, Ser. B*, 81(1):20–44, 2001.

[13] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *J. Complex Networks*, 2(3):203–271, 2014.

[14] Ton Kloks. *Treewidth, Computations and Approximations*, volume 842 of *Lecture Notes in Computer Science*. Springer, 1994.

[15] Choongbum Lee, Joonkyung Lee, and Sang-il Oum. Rank-width of random graphs. *J. Graph Theory*, 70(3):339–347, 2012.

[16] Richard J. Lipton and Robert Endre Tarjan. A separator theorem for planar graphs. *SIAM Journal on Applied Mathematics*, 36(2):177–189, 1979.

[17] Colin L. Mallows. Non-null ranking models. I. *Biometrika*, 44:114–130, 1957.

[18] Jaroslav Nesetril and Patrice Ossona de Mendez. *Sparsity - Graphs, Structures, and Algorithms*, volume 28 of *Algorithms and combinatorics*. Springer, 2012.

[19] Shai Pilosof, Mason A Porter, Mercedes Pascual, and Sonia Kéfi. The multilayer nature of ecological networks. *Nature Ecology & Evolution*, 1(4):1–9, 2017.

[20] Ross G. Pinsky. Permutations avoiding a pattern of length three under mallows distributions. *Random Struct. Algorithms*, 58(4):676–690, 2021.

[21] Ross G. Pinsky. Clustering of consecutive numbers in permutations under Mallows distributions and super-clustering under general p-shifted distributions. *Electronic Journal of Probability*, 27:1 – 20, 2022.

[22] Daniel A Quiroz. Chromatic and structural properties of sparse graph classes. *PhD thesis, The London School of Economics and Political Science (LSE)*, 2017.