

Graphs with few paths of prescribed length between any two vertices

David Conlon*

Abstract

We use a variant of Bukh's random algebraic method to show that for every natural number $k \geq 2$ there exists a natural number ℓ such that, for every n , there is a graph with n vertices and $\Omega_k(n^{1+1/k})$ edges with at most ℓ paths of length k between any two vertices. A result of Faudree and Simonovits shows that the bound on the number of edges is tight up to the implied constant.

1 Introduction

Given a graph H , let $\text{ex}(n, H)$ be the maximum number of edges in an H -free graph on n vertices. The classic Erdős–Stone–Simonovits theorem [9, 10] gives a satisfactory first estimate for this function, showing that

$$\text{ex}(n, H) = \left(1 - \frac{1}{\chi(H) - 1} + o(1)\right) \binom{n}{2},$$

where $\chi(H)$ is the chromatic number of H . For bipartite H , this gives the bound $\text{ex}(n, H) = o(n^2)$. While more precise estimates are known, a number of notoriously difficult open problems remain.

The most intensively studied case is when $H = K_{s,t}$, the complete bipartite graph with parts of order s and t . In this case, a famous result of Kővári, Sós and Turán [13] shows that $\text{ex}(n, K_{s,t}) = O_{s,t}(n^{2-1/s})$ whenever $s \leq t$. This bound was shown to be tight for $s = 2$ by Erdős, Rényi and Sós [8] and for $s = 3$ by Brown [5]. For higher values of s , it is only known that the bound is tight when t is sufficiently large in terms of s . This was first shown by Kollár, Rónyai and Szabó [12], though the construction was improved slightly by Alon, Rónyai and Szabó [1], who showed that there are graphs with n vertices and $\Omega_s(n^{2-1/s})$ edges containing no copy of $K_{s,t}$ with $t = (s - 1)! + 1$. An alternative construction, with a slightly weaker bound on t , was also given by Blagojević, Bukh and Karasev [3].

Very recently, Bukh [6] found a simple, elegant method for showing that the Kővári–Sós–Turán bound is tight for t sufficiently large in terms of s , fusing the algebraic techniques used in all previous constructions with an application of the probabilistic method. In this paper, we adapt his method to make progress on an equally stubborn problem.

When $H = C_{2k}$, the cycle with $2k$ vertices, a result of Erdős (see [4]) shows that $\text{ex}(n, C_{2k}) = O_k(n^{1+1/k})$. Since $C_4 = K_{2,2}$, the result of Erdős, Rényi and Sós mentioned above shows that this bound is tight for $k = 2$. For $k = 3$ and 5 , constructions matching the upper bound were found by Benson [5] and Singleton [18] and later by Wenger [19], Lazebnik and Ustimenko [15] and Mellinger and Mubayi [17]. For general k , the best known lower bound on $\text{ex}(n, C_{2k})$ is due to Lazebnik, Ustimenko and Woldar [16] but does not match the upper bound.

*Mathematical Institute, Oxford OX2 6GG, United Kingdom. E-mail: david.conlon@maths.ox.ac.uk. Research supported by a Royal Society University Research Fellowship.

If we let $\theta_{k,\ell}$ be the graph consisting of ℓ internally disjoint paths of length k , each with the same endpoints, we see that $\theta_{k,2} = C_{2k}$ and so the problem of determining $\text{ex}(n, \theta_{k,\ell})$ generalises the problem of determining $\text{ex}(n, C_{2k})$. This problem was first studied by Faudree and Simonovits [11], who showed that $\text{ex}(n, \theta_{k,\ell}) = O_{k,\ell}(n^{1+1/k})$ for all k and $\ell \geq 2$. The lower bounds for $\text{ex}(n, C_{2k})$ show that this bound is tight when $k = 2, 3$ or 5 . Additionally, a result of Mellinger and Mubayi [17] shows that it is tight for $k = 7$ and $\ell \geq 3$. In this paper, we generalise these results, showing that the upper bound is tight for every k , provided that ℓ is sufficiently large. This may be seen as an analogue of the Kollár–Rónyai–Szabó result for the cycle-free problem.

Theorem 1 *For any natural number $k \geq 2$, there exists a natural number ℓ such that*

$$\text{ex}(n, \theta_{k,\ell}) = \Omega_k(n^{1+1/k}).$$

More generally, we will show that for any natural number $k \geq 2$ there exists a natural number ℓ such that, for every n , there is a graph with n vertices and $\Omega_k(n^{1+1/k})$ edges with at most ℓ (not necessarily disjoint) paths of length k between any two vertices. Though heuristics suggest that it might be possible to take $\ell = k^{O(k^2)}$, our method falls well short of this. We have therefore made no systematic attempt to optimise (or even compute) the value of ℓ arising from our proof. Throughout the paper, we will use standard asymptotic notation, with subscripts indicating that the implied constants depend on these parameters.

2 Preliminaries

Let q be a prime power and let \mathbb{F}_q be the finite field of order q . We will consider polynomials in t variables over \mathbb{F}_q , writing any such polynomial as $f(X)$, where $X = (X_1, \dots, X_t)$. We let \mathcal{P}_d be the set of polynomials in X of degree at most d , that is, the set of linear combinations over \mathbb{F}_q of monomials of the form $X_1^{a_1} \cdots X_t^{a_t}$ with $\sum_{i=1}^t a_i \leq d$. By a random polynomial, we just mean a polynomial chosen uniformly from the set \mathcal{P}_d . One may produce such a random polynomial by choosing the coefficients of the monomials above to be random elements of \mathbb{F}_q .

We will now determine the probability that a randomly chosen polynomial from \mathcal{P}_d passes through a given set of points. For one point, the probability is $1/q$, as shown by the following simple result of Bukh [6]. We include the proof for completeness.

Lemma 1 *If f is a random polynomial from \mathcal{P}_d , then, for any fixed $x \in \mathbb{F}_q^t$,*

$$\mathbb{P}[f(x) = 0] = 1/q.$$

Proof: Let $\mathcal{Q}_d = \{f \in \mathcal{P}_d : f(0) = 0\}$, that is, the collection of polynomials in \mathcal{P}_d with zero constant term. Since every $f \in \mathcal{P}_d$ can be written as $g + h$ where $g \in \mathcal{Q}_d$ and h is a constant, we may sample a random element of \mathcal{P}_d by adding a random element g of \mathcal{Q}_d and a random element h of \mathbb{F}_q . Since, for any fixed choice of g , there is only one choice out of q for h such that $f(x) = 0$, the result follows. \square

The following result shows that once q and d are sufficiently large, the probability that a randomly chosen polynomial from \mathcal{P}_d contains each of m distinct points is exactly $1/q^m$. That is, the events are independent. This observation is again due to Bukh and corresponds to Lemma 4 of his paper [6], though we state and prove it in greater generality.

Lemma 2 Suppose that $q > \binom{m}{2}$ and $d \geq m - 1$. Then, if f is a random polynomial from \mathcal{P}_d and x_1, \dots, x_m are m distinct points in \mathbb{F}_q^t ,

$$\mathbb{P}[f(x_i) = 0 \text{ for all } i = 1, \dots, m] = 1/q^m.$$

Proof: To begin, we choose elements $a_2, \dots, a_t \in \mathbb{F}_q$ such that $x_{i,1} + \sum_{j=2}^t a_j x_{i,j}$ is distinct for all $i = 1, \dots, m$. To see that this is possible, note that there are exactly $\binom{m}{2}$ equations

$$x_{i,1} + \sum_{j=2}^t a_j x_{i,j} = x_{i',1} + \sum_{j=2}^t a_j x_{i',j},$$

each with at most q^{t-2} solutions (a_2, \dots, a_t) . Therefore, since the total number of choices for (a_2, \dots, a_t) is q^{t-1} and $q^{t-1} > q^{t-2} \binom{m}{2}$, we can make an appropriate choice.

We now consider \mathcal{P}'_d , the set of polynomials of degree at most d in Z , where $Z_1 = X_1 + \sum_{j=2}^t a_j X_j$ and $Z_j = X_j$ for all $2 \leq j \leq t$. Since this change of variables is an invertible linear map, \mathcal{P}'_d is identical to \mathcal{P}_d , so it will suffice to show that a randomly chosen polynomial from \mathcal{P}'_d passes through all of the points z_1, \dots, z_m corresponding to x_1, \dots, x_m with probability q^{-m} . Note that, by our choice above, $z_{i,1} \neq z_{i',1}$ for any $1 \leq i < i' \leq m$.

For any f in \mathcal{P}'_d , we may write $f = g + h$, where h contains all monomials of the form Z_1^j for $j = 0, 1, \dots, m - 1$ and g contains all other monomials. For any fixed choice of g , there is exactly one choice of h such that $f(z_i) = 0$ for all $i = 1, \dots, m$, namely, the unique polynomial of degree at most $m - 1$ which takes the value $-g(z_i)$ at $z_{i,1}$ for all $i = 1, 2, \dots, m$, where uniqueness follows from the fact that the $z_{i,1}$ are distinct. Since there were q^m possible choices for h , the result follows. \square

We also need to note some basic facts about affine varieties over finite fields. If we write $\overline{\mathbb{F}}_q$ for the algebraic closure of \mathbb{F}_q , a variety over $\overline{\mathbb{F}}_q$ is a set of the form

$$W = \{x \in \overline{\mathbb{F}}_q^t : f_1(x) = \dots = f_s(x) = 0\}$$

for some collection of polynomials $f_1, \dots, f_s : \overline{\mathbb{F}}_q^t \rightarrow \overline{\mathbb{F}}_q$. We say that W is defined over \mathbb{F}_q if the coefficients of these polynomials are in \mathbb{F}_q and write $W(\mathbb{F}_q) = W \cap \mathbb{F}_q^t$. We say that W has complexity at most M if s, t and the degrees of the f_i are all bounded by M . Finally, we say that a variety is absolutely irreducible if it is irreducible over $\overline{\mathbb{F}}_q$, reserving the term irreducibility for irreducibility over \mathbb{F}_q of varieties defined over \mathbb{F}_q .

The first result we will need is the Lang–Weil bound [14] relating the dimension of a variety W to the number of points in $W(\mathbb{F}_q)$. It will not be necessary to give a formal definition for the dimension of a variety, though some intuition may be gained by noting that if $f_1, \dots, f_s : \overline{\mathbb{F}}_q^t \rightarrow \overline{\mathbb{F}}_q$ are generic polynomials then the dimension of the variety they define is $t - s$.

Lemma 3 Suppose that W is a variety over $\overline{\mathbb{F}}_q$ of complexity at most M . Then

$$|W(\mathbb{F}_q)| = O_M(q^{\dim W}).$$

Moreover, if W is defined over \mathbb{F}_q and absolutely irreducible, then

$$|W(\mathbb{F}_q)| = q^{\dim W} (1 + O_M(q^{-1/2})).$$

We will also need the following standard result from algebraic geometry (see, for example, [7]), which says that if W is an irreducible affine variety over an algebraically closed field and g is a polynomial whose zero set intersects W then either W is contained in the zero set of g or its intersection with this zero set has smaller dimension.

Lemma 4 *Suppose that W is an absolutely irreducible variety over $\overline{\mathbb{F}}_q$ of complexity at most M and $\dim W \geq 1$. Then, for any polynomial $g : \overline{\mathbb{F}}_q^t \rightarrow \overline{\mathbb{F}}_q$, $W \subseteq \{x : g(x) = 0\}$ or $W \cap \{x : g(x) = 0\}$ is a variety of dimension less than $\dim W$.*

The final ingredient we require is again essentially due to Bukh [6] and says that if W is a variety which is defined over \mathbb{F}_q , then there is a finite collection of absolutely irreducible varieties Y_1, \dots, Y_s , each of which is defined over \mathbb{F}_q , such that $\cup_{i=1}^s Y_i(\mathbb{F}_q) = W(\mathbb{F}_q)$. Since it is less standard than the previous two lemmas, we will include the proof. We will repeatedly use the fact that a variety X is defined over \mathbb{F}_q if and only if it is fixed by the Frobenius automorphism taking any element x to x^q .

Lemma 5 *Suppose that W is a variety over $\overline{\mathbb{F}}_q$ of complexity at most M which is defined over \mathbb{F}_q . Then there are $O_M(1)$ absolutely irreducible varieties Y_1, \dots, Y_s , each of which is defined over \mathbb{F}_q and has complexity $O_M(1)$, such that $\cup_{i=1}^s Y_i(\mathbb{F}_q) = W(\mathbb{F}_q)$.*

Proof: We begin by splitting W into irreducible components, noting that the number and complexity of these components, each of which is defined over \mathbb{F}_q , is bounded by a function of the complexity of W . Let X be an irreducible component of W . If X is absolutely irreducible, we set it aside as one of our Y_i . If X is not absolutely irreducible, we let X_1, \dots, X_r be the absolutely irreducible components of X , noting that the number and complexity of these components is again bounded by a function of the complexity of W . Since X is defined over \mathbb{F}_q , the Frobenius automorphism $x \mapsto x^q$ acts on its components, permuting X_1, \dots, X_r . Moreover, this action is transitive. This is because the union of the sets in any orbit is fixed by the Frobenius automorphism and so defined over \mathbb{F}_q . Therefore, if the action were not transitive, X would be reducible over \mathbb{F}_q . By transitivity, we have that $X(\mathbb{F}_q) \subseteq X_i$ for all $i = 1, \dots, r$. But then $X(\mathbb{F}_q) \subseteq X_1 \cap \dots \cap X_r$, which is a variety of lower dimension than W that is defined over \mathbb{F}_q and has complexity bounded by a function of the complexity of W . We may now repeat the entire procedure with $X_1 \cap \dots \cap X_r$. Since the dimension is bounded below by zero, this iteration must eventually terminate. \square

3 The construction

Let $t = r = 2k$, $d = kr$, $N = q^k$ and suppose that q is sufficiently large. Let $f_1, \dots, f_{k-1} : \mathbb{F}_q^k \times \mathbb{F}_q^k \rightarrow \mathbb{F}_q$ be independent random polynomials in \mathcal{P}_d . We consider the bipartite graph G between two copies U and V of \mathbb{F}_q^k , each of order $N = q^k$, where (u, v) is an edge of G if and only if

$$f_1(u, v) = \dots = f_{k-1}(u, v) = 0.$$

Since f_1, \dots, f_{k-1} were chosen independently, Lemma 1 tells us that the probability a given edge (u, v) is in G is $q^{-(k-1)}$. Therefore, the expected number of edges in G is $q^{-(k-1)}N^2 = N^{1+1/k}$.

Suppose now that w_1 and w_2 are two fixed vertices in G and let S be the set of paths of length k between them. We will be interested in estimating the r -th moment of $|S|$. To begin, we note that

$|S|^r$ counts the number of ordered collections of r (possibly overlapping or identical) paths of length k in G between w_1 and w_2 . Since the total number of edges m in any given collection of r paths is at most $kr = d$ and q is sufficiently large, Lemma 2 tells us that the probability that particular collection of paths is in G is $q^{-(k-1)m}$, where we again used the fact that f_1, \dots, f_{k-1} are chosen independently. Within the complete bipartite graph between U and V , let $P_{r,m}$ be the number of ordered collections of r paths, each of length k , from w_1 to w_2 whose union has m edges. Then

$$\mathbb{E}[|S|^r] = \sum_{m=1}^{kr} P_{r,m} q^{-(k-1)m}.$$

To estimate $P_{r,m}$ and hence $\mathbb{E}[|S|^r]$, we will show that if the union of r paths, each of length k , has m edges then the number of vertices n other than w_1 and w_2 satisfies $kn \leq (k-1)m$. To see this, suppose that p_1, \dots, p_r are paths of length k whose union has m edges. We consider the paths in sequence, letting n_i be the number of vertices in $p_i \setminus p_1 \cup \dots \cup p_{i-1}$. If $n_i \neq 0$, it is easy to see that adding p_i to the union adds at least $n_i + 1$ edges. Therefore, summing over all i , we see that $m \geq n + r'$, where r' is the number of non-zero n_i . Since $r' \geq m/k$, this implies the required inequality. Therefore, $P_{r,m} \leq N^{(k-1)m/k}$ and

$$\mathbb{E}[|S|^r] = \sum_{m=1}^{kr} P_{r,m} q^{-(k-1)m} \leq \sum_{m=1}^{kr} N^{(k-1)m/k} q^{-(k-1)m} = \sum_{m=1}^{kr} q^{(k-1)m} q^{-(k-1)m} \leq \sum_{m=1}^{kr} 1 = kr.$$

By Markov's inequality, we may conclude that

$$\mathbb{P}[|S| \geq s] = \mathbb{P}[|S|^r \geq s^r] \leq \frac{\mathbb{E}[|S|^r]}{s^r} \leq \frac{kr}{s^r}.$$

We now note that the set of paths S is a subset of $T(\mathbb{F}_q)$, where

$$T = \{(x_1, \dots, x_{k-1}) \in \overline{\mathbb{F}_q}^{k(k-1)} : f_i(w_1, x_1) = f_i(x_2, x_1) = \dots = f_i(x_{k-1}, w_2) = 0 \text{ for } i = 1, \dots, k-1\}.$$

Note that, depending on which sides of the bipartition contain w_1 and w_2 , the order of the two variables in many of these equations may need to be reversed. However, this makes little difference to what follows, so we will assume that the orders are as given above.

If $T(\mathbb{F}_q)$ were equal to S , we could apply Lemma 5 from [6] to show that S is either bounded by a constant or quite large and then use the corollary of Markov's inequality proved above to show that there are very few pairs (w_1, w_2) for which S is large. Unfortunately, $T(\mathbb{F}_q)$ may contain degenerate walks as well as the paths we are interested in, so we must somehow take these into account.

If $T(\mathbb{F}_q)$ contains a degenerate walk $w_1, x_1, \dots, x_{k-1}, w_2$, it must be the case that either $w_1 = x_b$ for some $1 \leq b \leq k-1$ or $x_a = x_b$ for some $1 \leq a < b \leq k-1$ or $x_a = w_2$ for some $1 \leq a \leq k-1$. This naturally leads us to consider the collections of sets

- $T_{0b} = T \cap \{(x_1, \dots, x_{k-1}) : w_1 = x_b\}$ for $1 \leq b \leq k-1$,
- $T_{ab} = T \cap \{(x_1, \dots, x_{k-1}) : x_a = x_b\}$ for $1 \leq a < b \leq k-1$ and
- $T_{ak} = T \cap \{(x_1, \dots, x_{k-1}) : x_a = w_2\}$ for $1 \leq a \leq k-1$.

Since T is defined over \mathbb{F}_q and has complexity bounded in terms of k , Lemma 5 tells us that there are $O_k(1)$ absolutely irreducible varieties Y_1, \dots, Y_s , each of which is defined over \mathbb{F}_q and has complexity $O_k(1)$, such that $\cup_{i=1}^s Y_i(\mathbb{F}_q) = T(\mathbb{F}_q)$. If $\dim Y_i \geq 1$, Lemma 4 tells us that either there exist a and b such that $Y_i \subseteq T_{ab}$ or the dimension of $Y_i \cap T_{ab}$ is smaller than the dimension of Y_i for all a and b . If $Y_i \subseteq T_{ab}$ for some a and b , the component does not contain any non-degenerate paths and may be removed from consideration. If instead the dimension of $Y_i \cap T_{ab}$ is smaller than the dimension of Y_i for all a and b , the Lang–Weil bound, Lemma 3, tells us that for q sufficiently large

$$|S| \geq |Y_i(\mathbb{F}_q)| - \sum_{a,b} |Y_i \cap T_{ab}(\mathbb{F}_q)| \geq q^{\dim Y_i} - O_k(q^{\dim Y_i - \frac{1}{2}}) - O_k(q^{\dim Y_i - 1}) \geq \frac{q}{2}.$$

On the other hand, if $\dim Y_i = 0$ for every Y_i which is not contained in some T_{ab} , Lemma 3 tells us that $|S| \leq \sum |Y_i(\mathbb{F}_q)| = O_k(1)$, where the sum is taken over all i for which $\dim Y_i = 0$.

Putting everything together, we see that there exists a constant c_k , depending only on k , such that either $|S| \leq c_k$ or $|S| \geq q/2$. Therefore, by the consequence of Markov’s inequality noted earlier,

$$\mathbb{P}[|S| > c_k] = \mathbb{P}[|S| \geq q/2] \leq \frac{kr}{(q/2)^r}.$$

We call a pair of vertices (w_1, w_2) bad if there are more than $\ell = c_k$ paths between them. If we let B be the random variable counting the number of bad pairs, we have, since $r = 2k$,

$$\mathbb{E}[B] \leq N^2 \cdot \frac{kr}{(q/2)^r} = O_k(q^{2k-r}) = O_k(1).$$

We now remove a vertex from each bad pair to form a new graph G' . Since each vertex has degree at most N , the total number of edges removed is at most BN . Hence, the expected number of edges is

$$N^{1+1/k} - \mathbb{E}[B]N = \Omega_k(N^{1+1/k}).$$

Therefore, there is a graph with at most $2N$ vertices and $\Omega_k(N^{1+1/k})$ edges such that no two vertices have more than $\ell = c_k$ paths of length k between them. As stated, this result only holds when q is a prime power and $N = q^k$. However, it is a simple matter to use Bertrand’s postulate to show that the same conclusion holds for all N .

Acknowledgements. I would like to thank Boris Bukh for a number of valuable comments on an earlier draft of this paper.

References

- [1] N. Alon, L. Rónyai and T. Szabó, Norm-graphs: variations and applications, *J. Combin. Theory Ser. B* **76** (1999), 280–290.
- [2] C. T. Benson, Minimal regular graphs of girths eight and twelve, *Canad. J. Math.* **18** (1966), 1091–1094.
- [3] P. V. M. Blagojević, B. Bukh and R. Karasev, Turán numbers for $K_{s,t}$ -free graphs: topological obstructions and algebraic constructions, *Israel J. Math.* **197** (2013), 199–214.

- [4] A. J. Bondy and M. Simonovits, Cycles of even lengths in graphs, *J. Combin. Theory Ser. B* **16** (1974), 97–105.
- [5] W. G. Brown, On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* **9** (1966), 281–285.
- [6] B. Bukh, Random algebraic construction of extremal graphs, arXiv:1409.3856 [math.CO].
- [7] D. Bump, **Algebraic geometry**, World Scientific Publishing Co., Inc., River Edge, NJ, 1998.
- [8] P. Erdős, A. Rényi and V. T. Sós, On a problem of graph theory, *Studia Sci. Math. Hungar.* **1** (1966), 215–235.
- [9] P. Erdős and M. Simonovits, A limit theorem in graph theory, *Studia Sci. Math. Hungar.* **1** (1966), 51–57.
- [10] P. Erdős and A. H. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52** (1946), 1087–1091.
- [11] R. J. Faudree and M. Simonovits, On a class of degenerate extremal graph problems, *Combinatorica* **3** (1983), 83–93.
- [12] J. Kollár, L. Rónyai and T. Szabó, Norm-graphs and bipartite Turán numbers, *Combinatorica* **16** (1996), 399–406.
- [13] T. Kővári, V. T. Sós and P. Turán, On a problem of K. Zarankiewicz, *Colloq. Math.* **3** (1954), 50–57.
- [14] S. Lang and A. Weil, Number of points of varieties in finite fields, *Amer. J. Math.* **76** (1954), 819–827.
- [15] F. Lazebnik and V. A. Ustimenko, Explicit construction of graphs with an arbitrary large girth and of large size, *Discrete Appl. Math.* **60** (1995), 275–284.
- [16] F. Lazebnik, V. A. Ustimenko and A. J. Woldar, A new series of dense graphs of high girth, *Bull. Amer. Math. Soc.* **32** (1995), 73–79.
- [17] K. E. Mellinger and D. Mubayi, Constructions of bipartite graphs from finite geometries, *J. Graph Theory* **49** (2005), 1–10.
- [18] R. R. Singleton, On minimal graphs of maximum even girth, *J. Combin. Theory* **1** (1966), 306–332.
- [19] R. Wenger, Extremal graphs with no C^4 's, C^6 's, or C^{10} 's, *J. Combin. Theory Ser. B* **52** (1991), 113–116.