A de Bruijn - Erdős theorem and metric spaces

Ehsan Chiniforooshan[1][†] and Vašek Chvátal[2][‡]

[1] *Google, Inc.*
[2] *Canada Research Chair in Combinatorial Optimization, Concordia University, Montreal, Canada*

De Bruijn and Erdős proved that every noncollinear set of n points in the plane determines at least n distinct lines. Chen and Chvátal suggested a possible generalization of this theorem in the framework of metric spaces. We provide partial results in this direction.

1   Introduction

Two distinct theorems are referred to as "the de Bruijn - Erdős theorem". One of them [de Bruijn and Erdős, 1951] concerns the chromatic number of infinite graphs; the other [de Bruijn and Erdős, 1948] is our starting point:

> *Every noncollinear set of n points in the plane*
> *determines at least n distinct lines.*

This theorem involves neither measurement of distances nor measurement of angles: the only notion employed here is incidence of points and lines. Such theorems are a part of *ordered geometry* [Coxeter, 1961], which is built around the ternary relation of *betweenness*: point y is said to lie between points x and z if y is an interior point of the line segment with endpoints x and z. It is customary to write $[xyz]$ for the statement that y lies between x and z. In this notation, a *line* $\overline{uv}$ is defined — for any two distinct points u and v — as

$$\{p : [puv]\} \ \cup \ \{u\} \ \cup \ \{p : [upv]\} \ \cup \ \{v\} \ \cup \ \{p : [uvp]\}. \tag{1}$$

In terms of the Euclidean metric d, we have

$$[abc] \ \Leftrightarrow \ a, b, c \text{ are three distinct points and } d(a,b) + d(b,c) = d(a,c). \tag{2}$$

[†]Email: `chiniforooshan@alumni.uwaterloo.ca`
[‡]Email: `chvatal@cse.concordia.ca`

For an arbitrary metric space, equivalence (2) defines the ternary relation of *metric betweenness* introduced in [Menger, 1928] and further studied in [Blumenthal, 1970; Busemann, 1955; Chvátal, 2004]; in turn, (1) defines the line $\overline{uv}$ for any two distinct points u and v in the metric space. The resulting family of lines may have strange properties. For instance, a line can be a proper subset of another: in the metric space with points u, v, x, y, z and

$$d(u,v) = d(v,x) = d(x,y) = d(y,z) = d(z,u) = 1,$$
$$d(u,x) = d(v,y) = d(x,z) = d(y,u) = d(z,v) = 2,$$

we have

$$\overline{vy} = \{v, x, y\} \quad \text{and} \quad \overline{xy} = \{v, x, y, z\}.$$

Chen [Chen, 2006] proved that a classic theorem of ordered geometry, the Sylvester-Gallai theorem, generalizes in the framework of metric spaces (when lines in these spaces are defined differently than here). Chen and Chvátal [Chen and Chvátal, 2008] suggested that the de Bruijn - Erdős theorem, too, might generalize in this framework:

True or false? Every finite metric space (X, d)
where no line consists of the entire ground set X
determines at least $|X|$ distinct lines.

They proved that

- in every metric space on n points, there are at least $\lg n$ distinct lines or else some line consists of all n points.

We prove that

- in every metric space on n points, there are $\Omega((n/\rho)^{2/3})$ distinct lines, where ρ is the ratio between the largest distance and the smallest nonzero distance (Theorem 1);

- in every metric space induced by a connected graph on n vertices, there are $\Omega(n^{2/7})$ distinct lines or else some line consists of all n vertices (Corollary 1);

- in every metric space on n points where each nonzero distance equals 1 or 2, there are $\Omega(n^{4/3})$ distinct lines and this bound is tight (Theorem 3).

2   Lines in hypergraphs

A *hypergraph* is an ordered pair (X, H) such that X is a set and H is a family of subsets of X; elements of X are the *vertices* of the hypergraph and members of H are its *edges*. A hypergraph is called k-*uniform* if each of its edges consists of k vertices. The definition of lines in a metric space (X, d) depends only on the 3-uniform hypergraph $(X, H(d))$ where

$$H(d) = \{\{a, b, c\} : \ d(a,b) + d(b,c) = d(a,c) \text{ and } a \neq b, \ b \neq c\} :$$

the line $\overline{uv}$ equals $\{u, v\} \cup \{w : \{u, v, w\} \in H(d)\}$. This observation suggests extending the notion of lines in metric spaces to a notion of lines in 3-uniform hypergraphs: for any two distinct vertices u and v

in a 3-uniform hypergraph (X, H), the line $\overline{uv}$ is defined as $\{u, v\} \cup \{w : \{u, v, w\} \in H\}$. Now every metric space (X, d) and its associated hypergraph $(X, H(d))$ define the same family of lines.

Let $f(n)$ denote the smallest number of lines in a 3-uniform hypergraph on n vertices where no line consists of all n vertices and let $g(n)$ denote the smallest number of lines in a metric space on n points where no line consists of all n points. In this notation, $f(n) \leq g(n)$ for all n; Chen and Chvátal [Chen and Chvátal, 2008] proved that

$$\lg n \leq f(n) < c^{\sqrt{\lg n}}$$

for some positive constant c. (The proof of the lower bound is based on the observation that $w \notin \overline{uv}$ if and only if $v \notin \overline{uw}$, and so — unless some line contains all the vertices — the mapping that assigns to each vertex the set of lines containing it is one-to-one.) The upper bound on $f(n)$ does not rule out the possibility of $g(n) = n$: not all 3-uniform hypergraphs arise from metric spaces (X, d) as $(X, H(d))$. (It has been proved ([Chvátal, 2004; Chen, 2006]) that the hypergraph consisting of the seven vertices $0, 1, 2, 3, 4, 5, 6$ and the seven edges $\{i \bmod 7, (i+1) \bmod 7, (i+3) \bmod 7\}$ with $i = 0, 1, 2, 3, 4, 5, 6$ does not arise from any metric space. This 3-uniform hypergraph is known as the *Fano plane* or the *projective plane of order two*.)

We let K_4^3 denote the 3-uniform hypergraph with four vertices and four edges.

Lemma 1 *Let H be a 3-uniform hypergraph, let x be a vertex of H, and let T be a set of vertices of H such that (i) $x \notin T$ and (ii) there are no vertices u, v, w in T such that x, u, v, w induce a K_4^3 in H. Then H defines at least*

$$2^{-1/3}|T|^{2/3} - 2^{-2/3}|T|^{1/3}$$

distinct lines.

Proof: Let S denote a largest subset of T such that all the lines $\overline{xv}$ with $v \in S$ are identical. Now H defines at least $|T|/|S|$ distinct lines, which gives the desired conclusion with room to spare when $|S| \leq 2^{1/3}|T|^{1/3}$. We will prove that H defines at least $|S|(|S| - 1)/2$ distinct lines, which gives the desired conclusion when $|S| \geq 2^{1/3}|T|^{1/3}$. More precisely, we will prove that all the lines $\overline{uv}$ with $u, v \in S$ are distinct. For this purpose, consider any three pairwise distinct vertices u, v, w in S. Since $\overline{xu} = \overline{xv} = \overline{xw}$, all three of $\{x, u, v\}, \{x, u, w\}, \{x, v, w\}$ are edges of H; since x, u, v, w do not induce a K_4^3, it follows that $\{u, v, w\}$ is not an edge of H; since w is an arbitrary vertex in S distinct from u and v, the line $\overline{uv}$ intersects S in $\{u, v\}$.   □

3   Lines in metric spaces

Theorem 1 *In every metric space on n points such that $n \geq 2$, there are at least*

$$2^{-1/3} \left(\frac{n-1}{\lfloor \rho \rfloor} \right)^{2/3} - 2^{-2/3} \left(\frac{n-1}{\lfloor \rho \rfloor} \right)^{1/3}$$

distinct lines, where ρ is the ratio between the largest distance and the smallest nonzero distance.

Proof: Let the metric space be (X, d), let δ denote the smallest nonzero distance and let x be an arbitrary point of X. The $n - 1$ distances $d(x, u)$ with $u \neq x$ are distributed into buckets $[i\delta, (i+1)\delta)$ with

$i = 1, 2, \ldots, \lfloor \rho \rfloor$. It follows that there are a subset T of $X - \{x\}$ and a positive integer i such that

$$u \in T \;\Rightarrow\; i\delta \le d(x, u) < (i+1)\delta$$

and $|T| \ge (n-1)/\lfloor \rho \rfloor$. We will complete the proof by showing that the hypergraph $(X, H(d))$ satisfies the hypothesis of Lemma 1. For this purpose, consider arbitrary points u, v, w in T such that $\{x, u, v\}, \{x, u, w\}, \{x, v, w\} \in H(d)$; we will prove that $\{u, v, w\} \notin H(d)$. Since $\{x, u, v\} \in H(d)$ and $|d(x, u) - d(x, v)| < \delta \le d(u, v)$, we have $d(u, v) = d(x, u) + d(x, v)$; similarly, $d(u, w) = d(x, u) + d(x, w)$ and $d(v, w) = d(x, v) + d(x, w)$. Since

$$i\delta \le d(x, u), d(x, v), d(x, w) < (i+1)\delta \le 2i\delta,$$

we conclude that

$$2i\delta \le d(u, v), d(u, w), d(v, w) < 4i\delta,$$

and so $\{u, v, w\} \notin H(d)$. $\square$

4 Metric spaces induced by graphs

Every finite connected undirected graph induces a metric space, where the distance between vertices u and v is defined as the smallest number of edges in a path from u to v.

Theorem 2 *If, in a metric space (X, d) induced by a graph of diameter t, no line equals X, then there are at least $\sqrt{t/2}$ distinct lines.*

Proof: There are vertices $v_0, v_1, \ldots, v_t$ such that $d(v_i, v_j) = j - i$ whenever $0 \le i < j \le t$. Consider a largest set S of subscripts r such that all lines $\overline{v_r v_{r+1}}$ are equal. There are at least $t/|S|$ distinct lines; this gives the desired conclusion when $|S| \le \sqrt{2t}$. We will complete the argument by proving that there are at least $|S|/2$ distinct lines, which gives the desired conclusion when $|S| \ge \sqrt{2t}$.

Let u be any vertex outside the line $\overline{v_r v_{r+1}}$ with $r \in S$. We will prove that at least $|S|/2$ of the lines $\overline{u v_r}$ with $r \in S$ are pairwise distinct: more precisely, for every three subscripts i, j, k in S, at least two of the three lines $\overline{u v_i}, \overline{u v_j}, \overline{u v_k}$ are distinct.

By the triangle inequality and since $u \notin \overline{v_r v_{r+1}}$, we have

$$\left| d(u, v_r) - d(u, v_{r+1}) \right| \;<\; d(v_r, v_{r+1}) \quad \text{for all } r \text{ in } S;$$

since $d(v_r, v_{r+1}) = 1$, it follows that $d(u, v_r) = d(u, v_{r+1})$ for all r in S. Now consider any three subscripts i, j, k in S such that $i < j < k$. We have

$$d(v_i, v_j) + d(v_j, u) > d(v_{i+1}, v_j) + d(v_j, u) \ge d(v_{i+1}, u) = d(v_i, u),$$
$$d(u, v_i) + d(v_i, v_j) > d(u, v_i) + d(v_{i+1}, v_j) = d(u, v_{i+1}) + d(v_{i+1}, v_j) \ge d(u, v_j),$$

and so $v_j \in \overline{u v_i}$ if and only if $d(v_i, u) + d(u, v_j) = d(v_i, v_j)$. Similarly, $v_k \in \overline{u v_j}$ if and only if $d(v_j, u) + d(u, v_k) = d(v_j, v_k)$. Since

$$d(u, v_i) + d(u, v_k) \;=\; d(u, v_i) + d(u, v_{k+1}) \;\ge\; d(v_i, v_{k+1}) \;=\; k + 1 - i,$$

we have $d(u, v_i) > j - i$ or else $d(u, v_k) > k - j$. If $d(u, v_i) > j - i$, then $d(v_i, u) + d(u, v_j) > d(v_i, v_j)$, and so $v_j \notin \overline{uv_i}$ (and $v_i \notin \overline{uv_j}$), which implies $\overline{uv_i} \neq \overline{uv_j}$. If $d(u, v_{k+1}) > k - j$, then $d(v_j, u) + d(u, v_k) > d(v_j, v_k)$, and so $v_k \notin \overline{uv_j}$ (and $v_j \notin \overline{uv_k}$), which implies $\overline{uv_j} \neq \overline{uv_k}$. $\qquad \square$

Corollary 1 *If, in a metric space induced by a connected graph on n vertices, no line consists of all n vertices, then there are at least $2^{-8/7} n^{2/7}$ distinct lines.*

Proof: If the graph has diameter at most $2^{-9/7} n^{4/7}$, then the bound follows from Theorem 1; else it follows from Theorem 2. $\qquad \square$

5   Metric spaces where each nonzero distance is 1 or 2

By a 1-2 *metric space,* we mean a metric space where each nonzero distance is 1 or 2.

Theorem 3 *The smallest number $h(n)$ of lines in a 1-2 metric space on n points satisfies the inequalities*

$$(1 + o(1))\alpha n^{4/3} \leq h(n) \leq (1 + o(1))\beta n^{4/3}$$

with $\alpha = 2^{-7/3}$ and $\beta = 3 \cdot 2^{-5/3}$.

We say that points u, v in a 1-2 metric space are *twins* if, and only if, $d(u, v) = 2$ and $d(u, w) = d(v, w)$ for all w distinct from both u and v. Our proof of Theorem 3 relies on the following lemma, whose proof is routine.

Lemma 2 *If u_1, u_2, u_3, u_4 are four distinct points in a 1-2 metric space, then:*

(i) *if $d(u_i, u_j) = 1$ for all choices of distinct i and j, then $\overline{u_1 u_2} \neq \overline{u_3 u_4}$,*

(ii) *if $d(u_1, u_2) = 1$ and $d(u_3, u_4) = 2$, then $\overline{u_1 u_2} \neq \overline{u_3 u_4}$,*

(iii) *if $d(u_1, u_2) = d(u_3, u_4) = 2$ and u_4 has a twin other than u_3, then $\overline{u_1 u_2} \neq \overline{u_3 u_4}$.*

If u_1, u_2, u_3 are three distinct points in a 1-2 metric space, then:

(iv) *if $d(u_1, u_2) = d(u_2, u_3) = 1$ and u_1, u_3 are not twins, then $\overline{u_1 u_2} \neq \overline{u_2 u_3}$,*

(v) *if $d(u_1, u_2) = 1$, $d(u_2, u_3) = 2$, and u_3 has a twin other than u_2, then $\overline{u_1 u_2} \neq \overline{u_2 u_3}$,*

(vi) *if $d(u_1, u_2) = d(u_2, u_3) = 2$, then $\overline{u_1 u_2} \neq \overline{u_2 u_3}$.*

$\qquad \square$

Proof of Theorem 3. To see that $h(n) \leq (1 + o(1))\beta n^{4/3}$, consider the metric space where the ground set is split into pairwise disjoint groups of sizes as nearly equal as possible, every two points that belong to two different groups have distance 1, and every two points that belong to one group have distance 2. If each group includes at least three points, then $\overline{uv} = \overline{wx}$ if and only if either $\{u, v\} = \{w, x\}$ or else there are two distinct groups such that each of the sets $\{u, v\}$, $\{w, x\}$ has one element in each of these two groups. Consequently, when there are n points altogether and $(1 + o(1))2^{-1/3}n^{2/3}$ groups, there are $(1 + o(1))\beta n^{4/3}$ lines.

To prove that $h(n) \geq (1 + o(1))\alpha n^{4/3}$, consider an arbitrary 1-2 metric space (X, d) and write $n = |X|$. Let X_1 be any maximal subset of X that does not contain a pair of twins.

CASE 1: $|X_1| \geq n/2$. In this case, consider a largest set of distinct two-point subsets $\{u_i, v_i\}$ ($i = 1, 2, \ldots, s$) of X_1 such that

$$\overline{u_1 v_1} = \overline{u_2 v_2} = \ldots = \overline{u_s v_s}.$$

Since every two-point subset of X_1 determines a line, there are at least

$$\binom{|X_1|}{2} \cdot \frac{1}{s}$$

distinct lines; this gives the desired conclusion when $s \leq (n/2)^{2/3}$. We will complete the argument by proving that there are at least

$$\binom{s}{2} - 5$$

distinct lines, which gives the desired conclusion when $s \geq (n/2)^{2/3}$.

For this purpose, we may assume that $s \geq 5$. Part (iv) of Lemma 2 guarantees that the sets $\{u_i, v_i\}$ with $d(u_i, v_i) = 1$ are pairwise disjoint; part (vi) of Lemma 2 guarantees that the sets $\{u_i, v_i\}$ with $d(u_i, v_i) = 2$ are pairwise disjoint; part (ii) of Lemma 2 guarantees that each of the sets $\{u_i, v_i\}$ with $d(u_i, v_i) = 1$ meets each of the sets $\{u_i, v_i\}$ with $d(u_i, v_i) = 2$; now our assumption $s \geq 5$ guarantees that all s distances $d(u_i, v_i)$ are equal. We are going to prove that there are at least $s(s - 1)/2$ distinct lines: for every choice of subscripts i, j such that $1 \leq i < j \leq s$, there is a line L_{ij} such that

$$\{u_k, v_k\} \subseteq L_{ij} \iff k \in \{i, j\}.$$

SUBCASE 1.1: $d(u_1, v_1) = d(u_2, v_2) = \ldots = d(u_s, v_s) = 1$.

Since $\{u_j, v_j\} \subseteq \overline{u_j v_j} = \overline{u_i v_i}$ and $\{u_i, v_i\} \subseteq \overline{u_i v_i} = \overline{u_j v_j}$, we may assume (after switching u_j with v_j if necessary) that $d(u_i, u_j) = 2$ and $d(u_i, v_j) = d(u_j, v_i) = 1$. Now we may set $L_{ij} = \overline{u_i u_j}$: if $k \notin \{i, j\}$, then $u_j \in \overline{u_j v_j} = \overline{u_k v_k}$ implies that one of $d(u_j, u_k)$ and $d(u_j, v_k)$ equals 2, and so $\{u_k, v_k\} \not\subseteq \overline{u_i u_j}$.

SUBCASE 1.2: $d(u_1, v_1) = d(u_2, v_2) = \ldots = d(u_s, v_s) = 2$.

Since $\overline{u_1 v_1} = \overline{u_2 v_2} = \ldots = \overline{u_s v_s}$, the distance between any point in one of the sets $\{u_1, v_1\}$, $\{u_2, v_2\}$, $\ldots$, $\{u_s, v_s\}$ and any point in another of these s sets equals 1; it follows that we may set $L_{ij} = \overline{u_i u_j}$.

CASE 2: $|X_1| < n/2$. Write $X_2 = X - X_1$, consider a largest set S of points in X_2 such that

$$u, v \in S, u \neq v \implies d(u, v) = 1,$$

and write

$$
\begin{aligned}
E_1 &= \{\,\{u,v\}\colon u,v \in S,\ u \neq v\}, \\
E_2 &= \{\,\{u,v\}\colon u,v \in X_2,\ d(u,v) = 2\}.
\end{aligned}
$$

Since every vertex of X_2 has a twin (else it could be added to X_1), Lemma 2 guarantees that every two distinct pairs in $E_1 \cup E_2$ determine two distinct lines. We complete the argument by pointing out that

$$|E_1 \cup E_2| \geq (1 + o(1))\alpha n^{4/3} :$$

the famous theorem of Turán [Turán, 1941, 1954] guarantees that

$$
|S| \ \geq \ \frac{|X_2|^2}{2|E_2| + |X_2|},
$$

and so $|E_2| \ < \ \alpha n^{4/3}$ implies $|E_1| \ \geq \ (1 + o(1))\,\alpha n^{4/3}$. □

The lower bound of Theorem 3 can be easily improved through a more careful analysis of Case 2: a routine exercise in calculus shows that

$$
x \geq 3,\ y \geq 0 \ \Rightarrow \ \frac{1}{2} \cdot \left(\frac{x^2}{2y + x}\right)^2 + y \ \geq \ \beta x^{4/3} - \frac{x}{2}\,,
$$

and so $|E_1 \cup E_2| \geq (1 + o(1))\beta|X_2|^{4/3}$. Perhaps $h(n) = (1 + o(1))\beta n^{4/3}$.

Acknowledgements

References

Leonard M. Blumenthal. *Theory and applications of distance geometry*. Second edition. Chelsea Publishing Co., New York, 1970.

Herbert Busemann. *The geometry of geodesics*. Academic Press Inc., New York, N. Y., 1955.

Xiaomin Chen. The Sylvester-Chvátal theorem. *Discrete Comput. Geom.*, 35(2):193–199, 2006. ISSN 0179-5376. doi: 10.1007/s00454-005-1216-9. URL http://dx.doi.org/10.1007/s00454-005-1216-9.

Xiaomin Chen and Vašek Chvátal. Problems related to a de Bruijn-Erdős theorem. *Discrete Appl. Math.*, 156(11):2101–2108, 2008. ISSN 0166-218X. doi: 10.1016/j.dam.2007.05.036. URL http://dx.doi.org/10.1016/j.dam.2007.05.036.

Vašek Chvátal. Sylvester-Gallai theorem and metric betweenness. *Discrete Comput. Geom.*, 31(2):175–195, 2004. ISSN 0179-5376. doi: 10.1007/s00454-003-0795-6. URL http://dx.doi.org/10.1007/s00454-003-0795-6.

H. S. M. Coxeter. *Introduction to geometry*. John Wiley & Sons Inc., New York, 1961.

N. G. de Bruijn and P. Erdős. On a combinatorial problem. *Nederl. Akad. Wetensch., Proc.*, 51:1277–1279 = Indagationes Math. 10, 421–423 (1948), 1948.

N. G. de Bruijn and P. Erdős. A colour problem for infinite graphs and a problem in the theory of relations. *Nederl. Akad. Wetensch. Proc. Ser. A.* **54** = *Indagationes Math.*, 13:369–373, 1951.

Karl Menger. Untersuchungen über allgemeine Metrik. *Math. Ann.*, 100(1):75–163, 1928. ISSN 0025-5831. doi: 10.1007/BF01448840. URL http://dx.doi.org/10.1007/BF01448840.

P. Turán. On the theory of graphs. *Colloquium Math.*, 3:19–30, 1954.

Paul Turán. Eine Extremalaufgabe aus der Graphentheorie. *Mat. Fiz. Lapok*, 48:436–452, 1941.