Problems related to a de Bruijn - Erdős theorem

Xiaomin Chen

Department of Computer Science,
Rutgers University, Piscataway, NJ 08854-8019, USA

Vašek Chvátal [1]

Canada Research Chair in Combinatorial Optimization,
Department of Computer Science and Software Engineering,
Concordia University, Montréal, Québec H3G 1M8, Canada

In memory of Leo Khachiyan

Abstract

De Bruijn and Erdős proved that every noncollinear set of n points in the plane determines at least n distinct lines. We suggest a possible generalization of this theorem in the framework of metric spaces and provide partial results on related extremal combinatorial problems.

Key words: combinatorial geometry, metric space, metric betweenness, extremal combinatorial problem

1 Lines in metric spaces

Two distinct theorems are referred to as "the de Bruijn - Erdős theorem". One of them [9] concerns the chromatic number of infinite graphs; the other [8] is our starting point: *Every noncollinear set of n points in the plane determines at least n distinct lines.*

[1] This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program and from the Natural Sciences and Engineering Research Council of Canada.

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* [7], 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 ρ, we have

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

For an arbitrary metric space, equivalence (2) defines the ternary relation of *metric betweenness* introduced in [13] and further studied in [2,4,6]; 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

$$\begin{aligned} \rho(u,v) = \rho(v,x) = \rho(x,y) = \rho(y,z) = \rho(z,u) = 1, \\ \rho(u,x) = \rho(v,y) = \rho(x,z) = \rho(y,u) = \rho(z,v) = 2, \end{aligned}$$

we have

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

Nevertheless, fragments of ordered geometry might translate to the framework of metric spaces. In particular, we know of no counterexample to the de Bruijn - Erdős theorem in this framework.

Question 1 *True or false? Every finite metric space (X, ρ) where no line consists of the entire ground set X determines at least $|X|$ distinct lines.*

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*. Our definition of lines in a metric space (X, ρ) depends only on the hypergraph $(X, H(\rho))$ where

$$H(\rho) = \{\{a, b, c\} : [abc]\} :$$

the line $\overline{uv}$ equals $\{u,v\} \cup \{p : \{u,v,p\} \in H(\rho)\}$. This observation leads us to extend the notion of lines in metric spaces to a notion of lines in hypergraphs: given an arbitrary hypergraph (X,H), we define the line $\overline{uv}$ — for any two distinct vertices u and v — as $\{u,v\} \cup \{p : \exists T\, (T \in H, \{u,v,p\} \subseteq T)\, \}$. Now every metric space (X,ρ) and its associated hypergraph $(X, H(\rho))$ define the same family of lines.

A hypergraph is called *k-uniform* if each of its edges consists of k vertices. All the hypergraphs $(X, H(\rho))$ are 3-uniform, but some 3-uniform hypergraphs do not arise from any metric space (X,ρ) as $(X, H(\rho))$: it has been proved ([6,5]) that the hypergraph consisting of the seven vertices $0,1,2,3,4,5,6$ and the seven edges

$$\{i \bmod 7,\ (i+1) \bmod 7,\ (i+3) \bmod 7\} \qquad (i = 0,1,2,3,4,5,6)$$

does not arise from any metric space. (This 3-uniform hypergraph is known as the *Fano plane* or the *projective plane of order two*.) Restricting the notion of lines to 3-uniform hypergraphs would bring about no loss of generality: for every hypergraph (X,H) there is a 3-uniform hypergraph $(X, H^{(3)})$ such that (X,H) and $(X, H^{(3)})$ define the same family of lines. Specifically,

$$H^{(3)} = \{S : |S| = 3 \ \text{ and } \ \exists T\, (T \in H, S \subseteq T)\, \}.$$

Let $m(n,k)$ denote the smallest number of lines in a hypergraph on n vertices where every line consists of at most k vertices. Showing that $m(n, n-1) \geq n$ would show that the answer to Question 1 is "true". However, as we are going to prove, $m(n,n-1)$ grows slower than every power of n.

Lemma 2 *If n, ℓ, a are positive integers such that $2 \leq n - \ell \leq a^{\ell}$, then*

$$m(n, n-1) \leq 2^{\ell} + \ell a.$$

PROOF. Write $P = \{1,2,\ldots,\ell\}$ and let A be a set of size a. By assumption, there is a set S of strings of length ℓ over alphabet A such that $|S| = n - \ell$ and such that, for each i in P, some two strings in S differ in their i-th position. For each choice of i in P and x in A, set

$$E_{ix} = \{i\} \cup \{x_1 x_2 \ldots x_{\ell} \in S : x_i = x\}.$$

Now consider all the lines $\overline{uv}$ in the hypergraph

$$(P \cup S,\ \{P, S\} \cup \{E_{ix} : i \in P, x \in A\}).$$

If $u, v \in P$, then $\overline{uv} = P$. If $u \in P$ and $v \in S$, then $\overline{uv} = E_{ux}$ with x the u-th character in v. If $u, v \in S$, then $\overline{uv} = S \cup P'$ with P' the set of positions in which u and v agree; P' is a proper (and possibly empty) subset of P. So the hypergraph has n vertices, none of its lines consists of all n vertices, and there are at most $1 + \ell a + (2^\ell - 1)$ lines. $\square$

Theorem 3 *There are positive constants n_0 and c such that*

$$n \geq n_0 \;\Rightarrow\; m(n, n-1) \leq c^{\sqrt{\ln n}}. \tag{3}$$

for all n.

PROOF. Let $\alpha, \beta, \gamma, \delta$ be arbitrary constants such that

$$0 < \alpha < 1 < \beta < \gamma < 2 < \delta.$$

There is a positive integer ℓ_0 such that

$$\ell \geq \ell_0 \;\Rightarrow\; \alpha\ell < \ell - 1,\; \beta^\ell < \gamma^\ell - 1,\; \ell\gamma^\ell < 2^\ell,\; 2^{\ell+1} < \delta^\ell.$$

We claim that (3) holds as long as

$$n \geq n_0 \;\Rightarrow\; n - \left\lceil \sqrt{\frac{\ln n}{\ln \beta}} \right\rceil \geq 2$$

and

$$\ln n_0 \geq \ell_0^2 \ln \beta, \quad \ln c \geq \frac{\ln \delta}{\alpha\sqrt{\ln \beta}}.$$

To justify this claim, consider an arbitrary n such that $n \geq n_0$ and set

$$\ell = \left\lceil \sqrt{\frac{\ln n}{\ln \beta}} \right\rceil, \quad a = \lfloor \gamma^\ell \rfloor.$$

Now $\ell \geq \ell_0$, $a > \beta^\ell$, and so $\ell \ln a > \ell^2 \ln \beta \geq \ln n$. Lemma 2 guarantees that

$$m(n, n-1) \leq 2^\ell + \ell a;$$

since

$$\ell < \frac{\ell - 1}{\alpha} < \frac{1}{\alpha}\sqrt{\frac{\ln n}{\ln \beta}}$$

we have

$$2^\ell + \ell a < 2^{\ell+1} < \delta^\ell < c^{\sqrt{\ln n}}.$$

$\square$

We do not know the order of growth of $m(n, n-1)$; our best lower bound is only logarithmic in n. (We follow the convention of letting lg stand for the logarithm to base 2.)

Theorem 4 $m(n, n-1) \geq \lg n$.

PROOF. Consider an arbitrary hypergraph with n vertices and m lines where no line consists of all n vertices. Let us observe that

$$\begin{array}{l}\text{for every two distinct vertices } u \text{ and } v, \\ \text{there is a line which includes } u \text{ and does not include } v:\end{array} \tag{4}$$

by assumption, some vertex w is not included in line $\overline{uv}$, and so no edge includes all three vertices u, v, w, and so line $\overline{uw}$ includes u and does not include v. For each vertex x, let S_x denote the set of all lines that include x. Property (4) guarantees that these n sets are all distinct, and so $n \leq 2^m$. $\square$

Actually, property (4) guarantees that the n sets S_x form an *antichain* in the sense that none of them is a subset of another. This observation allows a negligible improvement of the bound in Theorem 4: first, the classic result of Sperner ([15]) asserts that an antichain on a ground set of size m has at most

$$\binom{m}{\lfloor m/2 \rfloor}$$

sets; next, by Stirling's formula,

$$\binom{m}{\lfloor m/2 \rfloor} \sim \frac{2^m}{\sqrt{\pi m/2}};$$

finally, if $m = \lg n + \frac{1}{2} \lg \lg n + c$, then

$$\frac{2^m}{\sqrt{\pi m/2}} \sim 2^c (2/\pi)^{1/2} n.$$

It follows that for every positive ε there is an n_0 such that

$$n \geq n_0 \ \Rightarrow \ m(n, n-1) > \lg n + \frac{1}{2} \lg \lg n + \frac{1}{2} \lg \frac{\pi}{2} - \varepsilon.$$

Since $m(n, k)$ is a nonincreasing function of k, Theorem 4 guarantees that $m(n, k) \geq \lg n$ whenever $2 \leq k < n$. For small values of k, this bound can be much improved.

Theorem 5

$$m(n, k) \geq \frac{n(n-1)}{k(k-1)}$$

whenever $n \geq k \geq 2$.

PROOF. Consider an arbitrary hypergraph with n vertices and m lines where every line consists of at most k vertices. Trivially,

$$\text{for every two distinct vertices } u \text{ and } v, \text{ there is a line which includes both } u \text{ and } v. \tag{5}$$

Let P denote the set of all pairs $(L, \{u, v\})$ such that L is a line and u, v are two distinct vertices in L. On the one hand, every line includes at most k points, and so

$$|P| \leq m\binom{k}{2}.$$

On the other hand, property (5) guarantees that

$$|P| \geq \binom{n}{2}.$$

The lower bound on m follows by comparing the two bounds on $|P|$. □

When the value of k is fixed, the lower bound of Theorem 5 is asymptotically optimal:

Theorem 6

$$\lim_{n\to\infty} m(n,k) \cdot \frac{k(k-1)}{n(n-1)} = 1$$

whenever $k \geq 2$.

PROOF. Theorem 5 guarantees that

$$\liminf_{n\to\infty} m(n,k) \cdot \frac{k(k-1)}{n(n-1)} \geq 1\,.$$

In every k-uniform hypergraph (X, H) such that

$$\text{every two edges share at most one vertex,} \tag{6}$$

each line is either an edge or a set of two vertices that is not a subset of any edge, and so there are

$$|H| + \left(\binom{|X|}{2} - |H|\binom{k}{2}\right)$$

lines altogether. In particular, with $f(n, k)$ standing for the largest number of edges in a k-uniform hypergraph with n vertices and with property (6), we have

$$m(n,k) \le \binom{n}{2} - f(n,k)\left(\binom{k}{2} - 1\right);$$

Erdős and Hanani [11] proved that

$$\lim_{n\to\infty} f(n,k)\cdot\frac{k(k-1)}{n(n-1)} = 1;$$

it follows that

$$\limsup_{n\to\infty} m(n,k)\cdot\frac{k(k-1)}{n(n-1)} \le 1.$$

□

3 Closure-lines in hypergraphs and metric spaces

The *Sylvester-Gallai theorem* [16,10,7,3,12,14,6] asserts that every noncollinear finite set X of points in the plane includes two points such that the line passing through them includes no other point of X. This theorem does not translate to the framework of metric spaces along the simple lines of our Section 1: in the five-point example of that section, every line consists of three or four ponts. Nevertheless, it does translate to the framework of metric spaces in a circuitous way, which we are about to describe.

Let us call a set T of vertices in a hypergraph *affinely closed* if, and only if, every edge that shares at least two vertices with T is fully contained in T. For every set S of vertices, the intersection of all affinely closed supersets of S is an affinely closed set, which we will refer to as the *affine closure* of S and which we will denote by $\text{aff}(S)$. By *closure-lines* in the hypergraph, we shall mean all the sets $\text{aff}(\{u,v\})$ with u and v two distinct vertices; by closure-lines in a metric space (X,ρ), we shall mean closure-lines in its associated hypergraph $(X,H(\rho))$.

When X is a subset of a Euclidean space and ρ is the Euclidean metric, lines and closure-lines in (X,ρ) coincide: each of them is the intersection of X and the Euclidean line passing through two distinct points of X. One of us [6] conjectured and the other one [5] proved that the notion of closure-lines provides a translation of the Sylvester-Gallai theorem to the framework of metric spaces:

In every finite metric space, some closure-line includes either all the points of the ground set or only two of them.

The same notion falls far short of providing a translation of the de Bruijn - Erdős theorem to the framework of metric spaces:

Theorem 7 *For every integer n greater than 5, there is a metric space on n points where each closure-line consists of at most $n-2$ points and there are precisely 7 distinct closure-lines altogether.*

PROOF. Consider the metric space (X,ρ), where $X = \{x_k : 1 \le k \le n\}$ with

$$x_1 = (1,3), \quad x_2 = (2,4), \quad x_3 = (3,1), \quad x_4 = (4,2),$$
$$x_k = (k, n+5-k) \text{ whenever } 5 \le k \le n,$$

and

$$\rho((a_1,a_2),\,(b_1,b_2)) \;=\; |a_1-b_1| + |a_2-b_2|.$$

Since $H(\rho)$ consists of all $\{x_1,x_2,x_k\}$ with $5 \le k \le n$, all $\{x_3,x_4,x_k\}$ with $5 \le k \le n$, and all $\{x_i,x_j,x_k\}$ with $5 \le i < j < k \le n$, we have

$$\begin{aligned}
&\mathrm{aff}(\{x_1,x_2\}) = X - \{x_3,x_4\},\\
&\mathrm{aff}(\{x_3,x_4\}) = X - \{x_1,x_2\},\\
&\mathrm{aff}(\{x_i,x_j\}) = X - \{x_1,x_2,x_3,x_4\} \text{ whenever } 5 \le i < j \le n,\\
&\mathrm{aff}(\{x_i,x_j\}) = \{x_i,x_j\} \text{ whenever } 1 \le i \le 2 \text{ and } 3 \le j \le 4,\\
&\mathrm{aff}(\{x_i,x_j\}) = X - \{x_3,x_4\} \text{ whenever } 1 \le i \le 2 \text{ and } 5 \le j \le n,\\
&\mathrm{aff}(\{x_i,x_j\}) = X - \{x_1,x_2\} \text{ whenever } 3 \le i \le 4 \text{ and } 5 \le j \le n.
\end{aligned}$$

□

Finally, let $\overline{m}(n,k)$ denote the smallest number of closure-lines in a hypergraph on n vertices where every closure-line consists of at most k vertices. Our proof of Theorem 5 with "lines" replaced by "closure-lines" shows that

$$\overline{m}(n,k) \ge \frac{n(n-1)}{k(k-1)} \tag{7}$$

whenever $n \ge k \ge 2$; in turn, our proof of Theorem 6 with "lines" replaced by "closure-lines" yields the following conclusion.

Theorem 8

$$\lim_{n\to\infty} \overline{m}(n,k) \cdot \frac{k(k-1)}{n(n-1)} \;=\; 1$$

whenever $k \ge 2$.

The order of growth of $\overline{m}(n,k)$ is given by its lower bound (7):

Theorem 9 *There is a positive constant* c *such that*

$$\frac{n(n-1)}{k(k-1)} \;\le\; \overline{m}(n,k) \;\le\; c \cdot \frac{n(n-1)}{k(k-1)}$$

whenever $n \ge k \ge 2$.

PROOF. For every integer k greater than 1, Theorem 8 guarantees the existence of a constant c_k such that

$$\overline{m}(n,k) \le c_k \cdot \frac{n(n-1)}{k(k-1)} \quad \text{whenever } n \ge k. \tag{8}$$

With c any constant such that

$$c \ge 12 \quad \text{and} \quad c \ge c_k \text{ whenever } 2 \le k < 12,$$

we propose to show that, for every integer k greater than 1,

$$\overline{m}(n,k) \;\le\; c \cdot \frac{n(n-1)}{k(k-1)} \quad \text{whenever } n > k. \tag{9}$$

(Trivially, $\overline{m}(n,k) = 1$ whenever $2 \le n \le k$.) For this purpose, consider an arbitrary but fixed integer k greater than 1. If $k < 12$, then (9) follows from (8); if $k \ge 12$, then we will use induction on n to prove that $\overline{m}(n,k) \le cn^2/k^2$ whenever $n > k$.

Set

$$p = 2\left\lceil \frac{n+1}{k} \right\rceil$$

and note for a future reference that

$$4 \le p < 2\left(\frac{n+1}{k} + 1\right) \le \frac{4n}{k}\,.$$

Take a set X such that $|X| = n$, take a subset X_0 of X such that $|X_0| = p-1$, and partition $X - X_0$ into pairwise disjoint sets V_i $(1 \le i \le p)$ whose sizes are as nearly equal as possible. Since

$$\frac{k}{4} - 1 < \frac{n-(p-1)}{p} \le \frac{k}{2} - 1\,,$$

we have

$$2 \le \min|V_i| \le \max|V_i| \le \frac{k-1}{2}\,.$$

In some hypergraph (X_0, H_0), every closure-line consists of at most k vertices and there are precisely $\overline{m}(p-1,k)$ distinct closure-lines altogether. A theorem of Behzad, Chartrand, and Cooper, Jr. [1] guarantees that (the chromatic index of the complete graph K_{2s} is $2s-1$, and so) there is a mapping

$$\phi : \{S :\ S \subset \{1,2,\ldots,p\},\ |S| = 2\} \to X_0$$

with the following property:

for every i in $\{1,2,\ldots,p\}$ and for every w in X_0
there is precisely one j in $\{1,2,\ldots,p\}$ such that $\phi(\{i,j\}) = w$.

Set

$$H_1 = \{\{u,v,w\} :\ \text{there are } i \text{ and } j \text{ with } u \in V_i,\ v \in V_j,\ \phi(\{i,j\}) = w\},$$
$$H_2 = \{S :\ |S| = 3 \text{ and there is an } i \text{ with } S \subseteq V_i\},$$

and $H = H_0 \cup H_1 \cup H_2$. Since closure-lines in hypergraph (X,H) are

- all the closure-lines in hypergraph (X_0, H_0),
- all the sets $V_i \cup V_j \cup \{\phi(\{i,j\})\}$ such that $1 \le i < j \le p$, and
- all the sets V_i such that $1 \le i \le p$,

we have

$$\overline{m}(n,k) \le \overline{m}(p-1,k) + \binom{p}{2} + p\,.$$

If $p-1 > k$, then (as $p-1 < n/3$) the induction hypothesis guarantees that

$$\overline{m}(p-1,k) \le c\left(\frac{p-1}{k}\right)^2 < \frac{c}{9}\left(\frac{n}{k}\right)^2;$$

if $p-1 \le k$, then

$$\overline{m}(p-1,k) = 1 < \frac{c}{9}\left(\frac{n}{k}\right)^2;$$

finally,

$$\binom{p}{2} + p = \binom{p+1}{2} < 10\left(\frac{n}{k}\right)^2.$$

We conclude that

$$\overline{m}(n,k) \le \frac{c}{9}\left(\frac{n}{k}\right)^2 + 10\left(\frac{n}{k}\right)^2 \le c \cdot \frac{n^2}{k^2}\,.$$

□

Acknowledgements

We are grateful to Bahman Kalantari for a question which stimulated our development of Theorem 4.

References

[1] M. Behzad, G. Chartrand, and J.K. Cooper, Jr., The colour numbers of complete graphs, The Journal of the London Mathematical Society 42 (1967) 226–228.

[2] L.M. Blumenthal, Theory and Applications of Distance Geometry, Oxford University Press, Oxford, 1953.

[3] P. Borwein and W.O.J. Moser, A survey of Sylvester's problem and its generalizations, Aequationes Math. 40 (1990) 111–135.

[4] H. Busemann, The Geometry of Geodesics, Academic Press, New York, 1955.

[5] X. Chen, The Sylvester-Chvátal theorem, Discrete & Computational Geometry 35 (2006) 193–199.

[6] V. Chvátal, Sylvester-Gallai theorem and metric betweenness, Discrete & Computational Geometry 31 (2004) 175–195.

[7] H.S.M. Coxeter, Introduction to Geometry, Wiley, New York, 1961.

[8] N.G. de Bruijn and P. Erdős, On a combinatorial problem, Indagationes Mathematicae 10 (1948) 421–423.

[9] N.G. de Bruijn and P. Erdős, A colour problem for infinite graphs and a problem in the theory of relations, Indagationes Mathematicae 13 (1951) 369–373.

[10] P. Erdős, Three point collinearity, American Mathematical Monthly 50 (1943) Problem 4065, p. 65. Solutions in Vol. 51 (1944) 169–171.

[11] P. Erdős and H. Hanani, On a limit theorem in combinatorial analysis, Publicationes Mathematicae Debrecen 10 (1963) 10–13.

[12] P. Erdős and G. Purdy, Some extremal problems in combinatorial geometry, in: R.L. Graham, M. Grötschel, and L. Lovász (Eds.), Handbook of Combinatorics, North-Holland, Amsterdam, 1991, pp. 809–874.

[13] K. Menger, Untersuchungen über allgemeine metrik, Mathematische Annalen 100 (1928) 75–163.

[14] J. Pach and P.K. Agarwal, Combinatorial Geometry, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 1995.

[15] E. Sperner, Ein Satz über Untermengen einer endlichen Menge, Mathematische Zeitschrift 27 (1928) 544-548.

[16] J.J. Sylvester, Mathematical Question 11851, Educational Times 59 (1893) p. 98.