PRETTY APPLICATIONS

OF THE

PROBABILISTIC METHOD

Adrian Bondy

Université Lyon 1 and Université Paris 6

POINTS AND LINES

P: finite set of points in the plane

line L: maximal set of collinear points

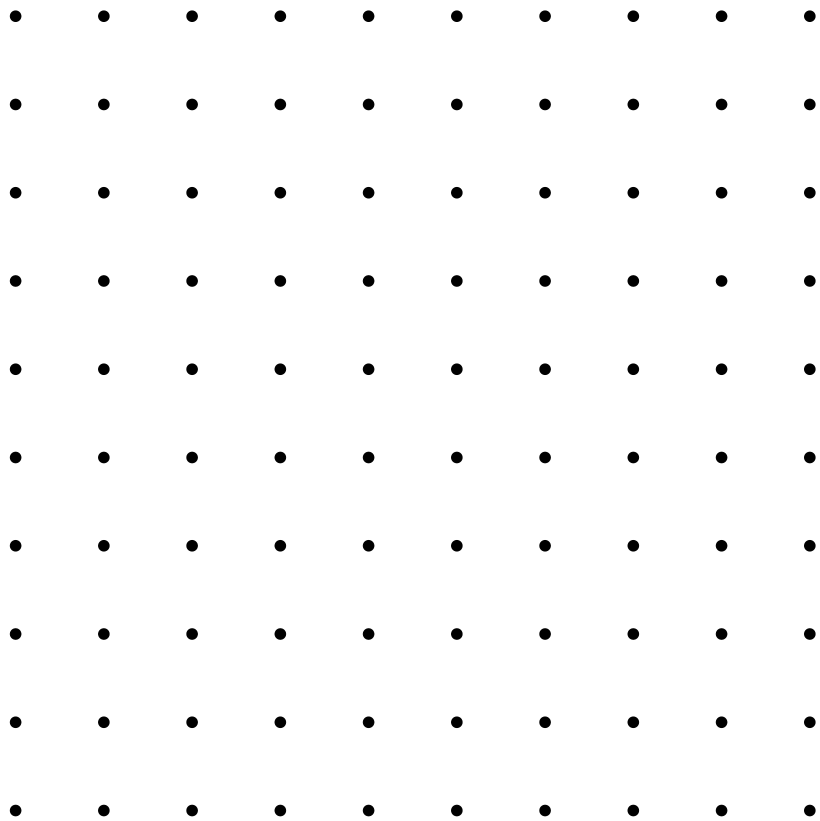
Theorem (Gallai 1944). *If P is a set of at least two points in the plane, not all collinear, then there exists a two-point line.*

Proof (Kelly 1958). Consider a point-line pair (p, L) where $p \notin L$ and p is as close to L as possible. Then L is a two-point line.

How about lines with many points?

Theorem (Szemerédi and Trotter 1983). *If P is a set of n points, then the number of lines containing more than k points is less than $32n^2/k^3$, provided that $1 \leq k \leq 2\sqrt{2n}$.*

THEOREM (Szemerédi and Trotter 1983). *If P is a set of n points, then the number of lines containing more than k points is less than $32n^2/k^3$, provided that $1 \leq k \leq 2\sqrt{2n}$.*



DISCRETE PROBABILITY SPACE

discrete probability space (Ω, P):

- sample space: finite set Ω
- probability function: mapping $P : \Omega \longrightarrow [0, 1]$ satisfying

$$\sum_{\omega \in \Omega} P(\omega) = 1$$

Examples

- random graphs on n vertices with edge probability p

  Ω: set of all *labelled* graphs on n vertices

  $P(G) := p^m (1-p)^{\binom{n}{2}-m}$, where $m := e(G)$

  This space is denoted $\mathcal{G}(n,p)$.

- random induced subgraphs with vertex probability p

  Ω: set of all *induced* subgraphs F of a graph G on n vertices

  $P(F) := p^k (1-p)^{n-k}$, where $k := v(F)$

EVENTS

event in (Ω, P): subset A of Ω.

probability of event A in (Ω, P): $P(A) := \sum_{\omega \in A} P(\omega)$

Examples

- Any property can be regarded as an event, namely the subset of Ω having the given property. We can thus speak of the probability of a property.
- For a set S of k vertices in $\mathcal{G}(n, p)$, denote by A_S the event that S is a stable set. Then $P(A_S) = p^{\binom{k}{2}}$.

Counting Sieve

If A and B are events in a probability space (Ω, P), then

$$P(A \cup B) \leq P(A) + P(B)$$

Thus if $P(A) + P(B) < 1$, then some element of Ω has neither property A nor property B.

RANDOM VARIABLES

random variable: mapping $X : \Omega \to \mathbb{R}$

Examples

- indicator random variable of an event A:

$$X(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

- For a set S of k vertices in $\mathcal{G}(n,p)$, denote by X_S the indicator random variable of A_S, the event that S is stable. Then $X := \sum_S X_S$, where the sum is taken over all sets S of k vertices, is the number of stable sets of cardinality k.

Notation

For a random variable X and a real number t,

$$P(X = t) := \sum_{\omega \in \Omega} \{P(\omega) : \; X(\omega) = t\}$$

Likewise,

$$P(X < t), \qquad P(X \leq t), \qquad P(X \geq t), \qquad P(X > t)$$

Example

- If X is the number of stable sets of cardinality k in $\mathcal{G}(n, p)$, then $P(X \geq 1)$ is the probability $\alpha \geq k$.

EXPECTATION

expectation of random variable X:

$$E(X) := \sum_{\omega \in \Omega} X(\omega) P(\omega)$$

Equivalently,

$$E(X) := \sum_{t \in \mathbb{R}} t P(X = t)$$

Example

- If X is the indicator random variable for an event A, then $E(X) = P(A)$. In particular, $E(X_S) = P(A_S) = p^{\binom{k}{2}}$.

LINEARITY OF EXPECTATION

Expectation is a *linear* function:

- $E(aX) = aE(X)$
- $E(X + Y) = E(X) + E(Y)$

Examples

- If X is the number of stable sets of cardinality k in $\mathcal{G}(n,p)$, then $X = \sum_S X_S$, where the X_S are indicator random variables. Therefore

$$E(X) = E(\sum_S X_S) = \sum_S E(X_S) = \sum_S p^{\binom{k}{2}} = \binom{n}{k} p^{\binom{k}{2}}$$

- Likewise, if X is the number of cycles of length k in $\mathcal{G}(n,p)$, then

$$E(X) = \frac{(n)_k}{2k} p^k$$

where $(n)_k := n(n-1)\cdots(n-k+1)$

CROSSING NUMBER

crossing number of graph G: minimum number $cr(G)$ of pairs of crossing edges in plane embedding of G

planar graph: one with crossing number zero

Euler's Formula

If G is a simple planar graph, then $m \leq 3n - 6$.

Corollary

If G is a simple graph, then

$$cr(G) \geq m - 3n$$

Theorem (Ajtai, Chvátal, Newborn and Szemerédi 1982). *If G is a simple graph on n vertices and m edges, where $m \geq 4n$, then*

$$cr(G) \geq \frac{1}{64}\frac{m^3}{n^2}$$

Proof (Alon 1992).

- $\widetilde{G}$: planar embedding of G with $cr(G)$ crossings
- F: random induced subgraph of G with vertex probability p
- $\widetilde{F}$: corresponding embedding of F

- X: number of vertices of F
- Y: number of edges of F
- Z: number of crossings of $\widetilde{F}$

By Corollary to Euler's Formula

$$Z \geq cr(F) \geq Y - 3X$$

By linearity of expectation

$$E(Z) \geq E(Y) - 3E(X)$$

Now

$$E(X) = pn, \qquad E(Y) = p^2 m, \qquad sE(Z) = p^4 cr(G)$$

Hence

$$p^4 cr(G) \geq p^2 m - 3pn$$

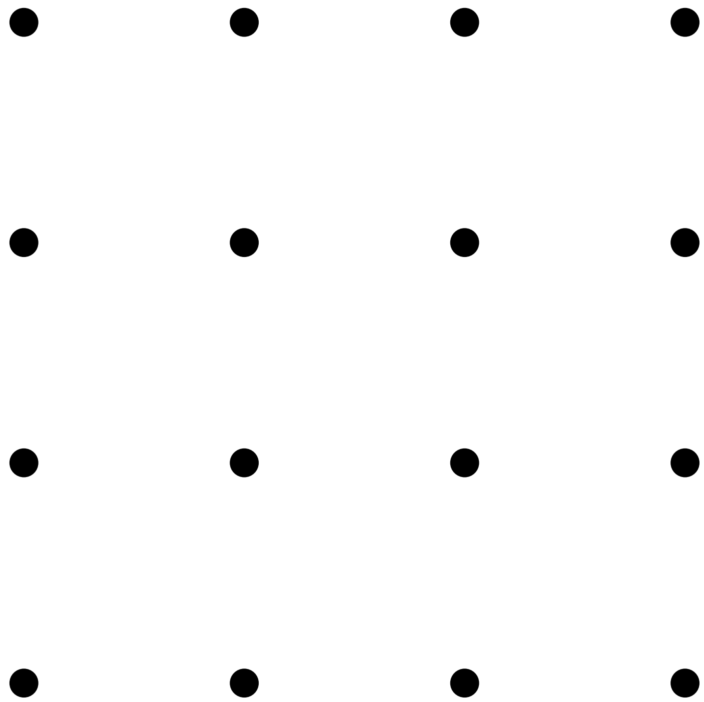
Dividing both sides by p^4 and setting $p = 4n/m$

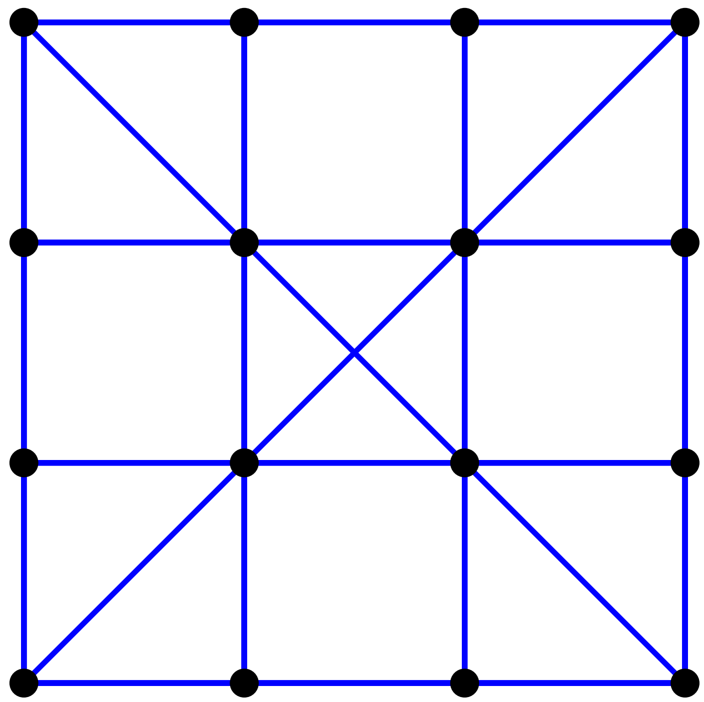
$$cr(G) \geq \frac{pm - 3n}{p^3} = \frac{n}{(4n/m)^3} = \frac{1}{64}\frac{m^3}{n^2}$$

THEOREM (Szemerédi and Trotter 1983). *If P is a set of n points in the plane, then the number ℓ of lines containing more than k points is less than $32n^2/k^3$, provided that $1 \leq k \leq 2\sqrt{2n}$.*

PROOF (Székely 1997). Define a simple graph G as follows:

- $V(G) := P$
- $E(G)$: segments between consecutive points on lines containing more than k points





Theorem (Szemerédi and Trotter 1983). *If P is a set of n points in the plane, then the number ℓ of lines containing more than k points is less than $32n^2/k^3$, provided that $1 \leq k \leq 2\sqrt{2n}$.*

Proof (Székely 1997). Define a simple graph G as follows:

- $V(G) := P$
- $E(G)$: segments between consecutive points on lines containing more than k points

Then

$$v(G) = n, \qquad e(G) \geq k\ell, \qquad cr(G) \leq \binom{\ell}{2}$$

- If $e(G) \geq 4n$, then by the Crossing Lemma

$$\binom{\ell}{2} \geq cr(G) \geq \frac{(k\ell)^3}{64n^2}$$

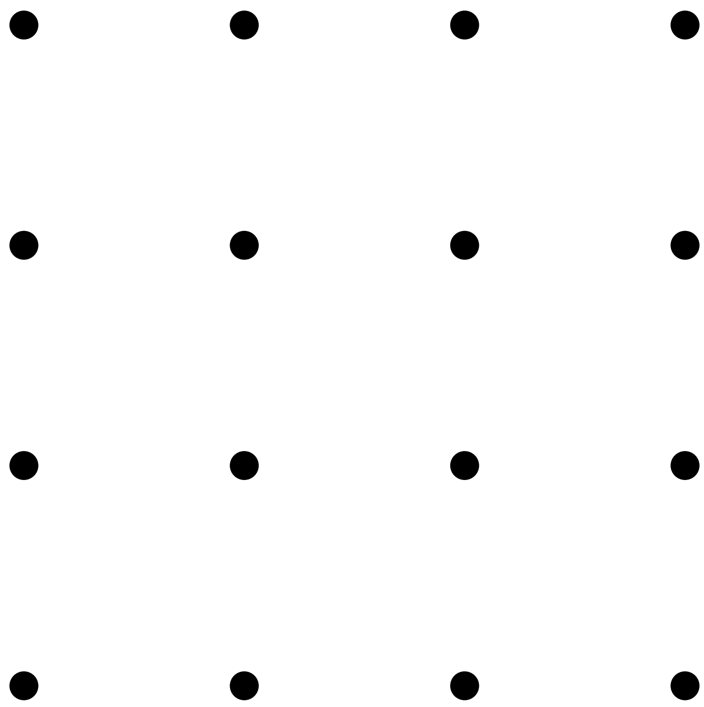
- If $e(G) < 4n$, then $k\ell < 4n$.

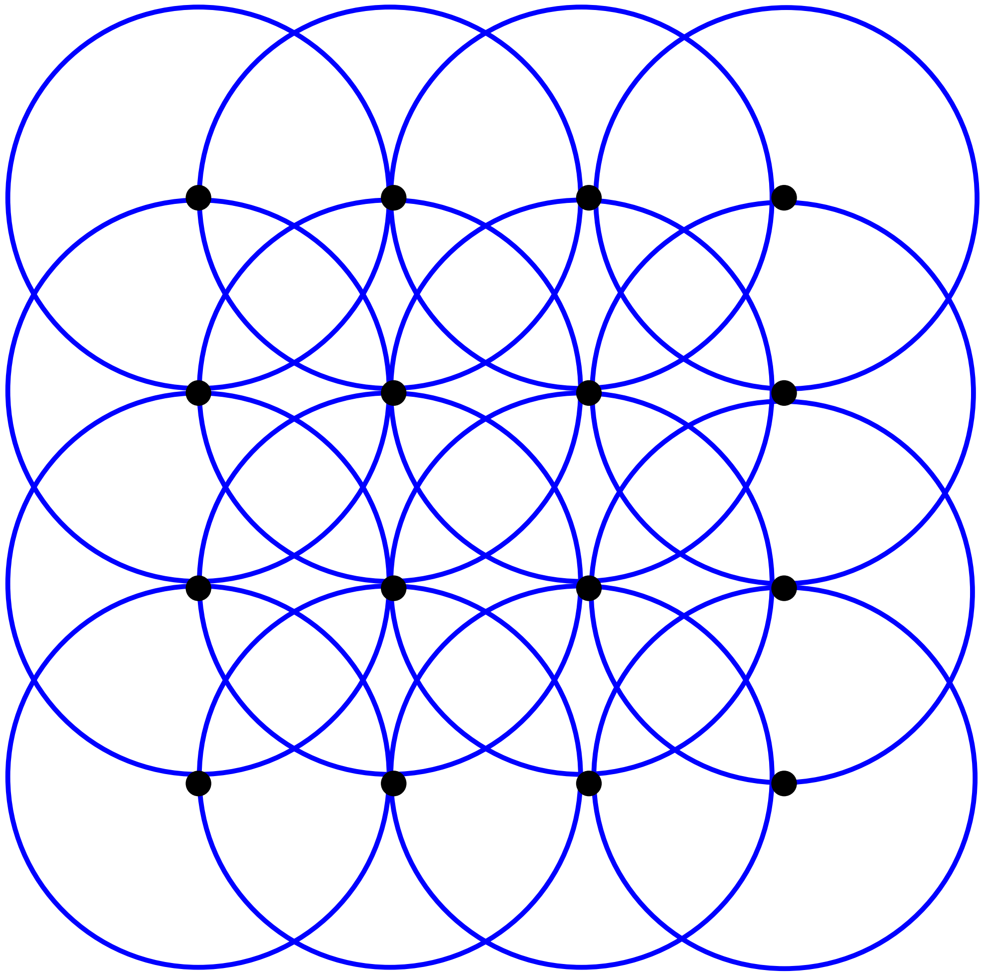
In both cases, $\ell \leq 32n^2/k^3$.

Theorem (Spencer, Szemerédi and Trotter 1984). If P is a set of n points in the plane, then the number k of pairs of points of P at unit distance is less than $n^{4/3}$.

Theorem (Spencer, Szemerédi and Trotter 1984). If P is a set of n points in the plane, then the number k of pairs of points of P at unit distance is less than $n^{4/3}$.

Proof (Székely 1997).
Draw a unit circle around each point of P.





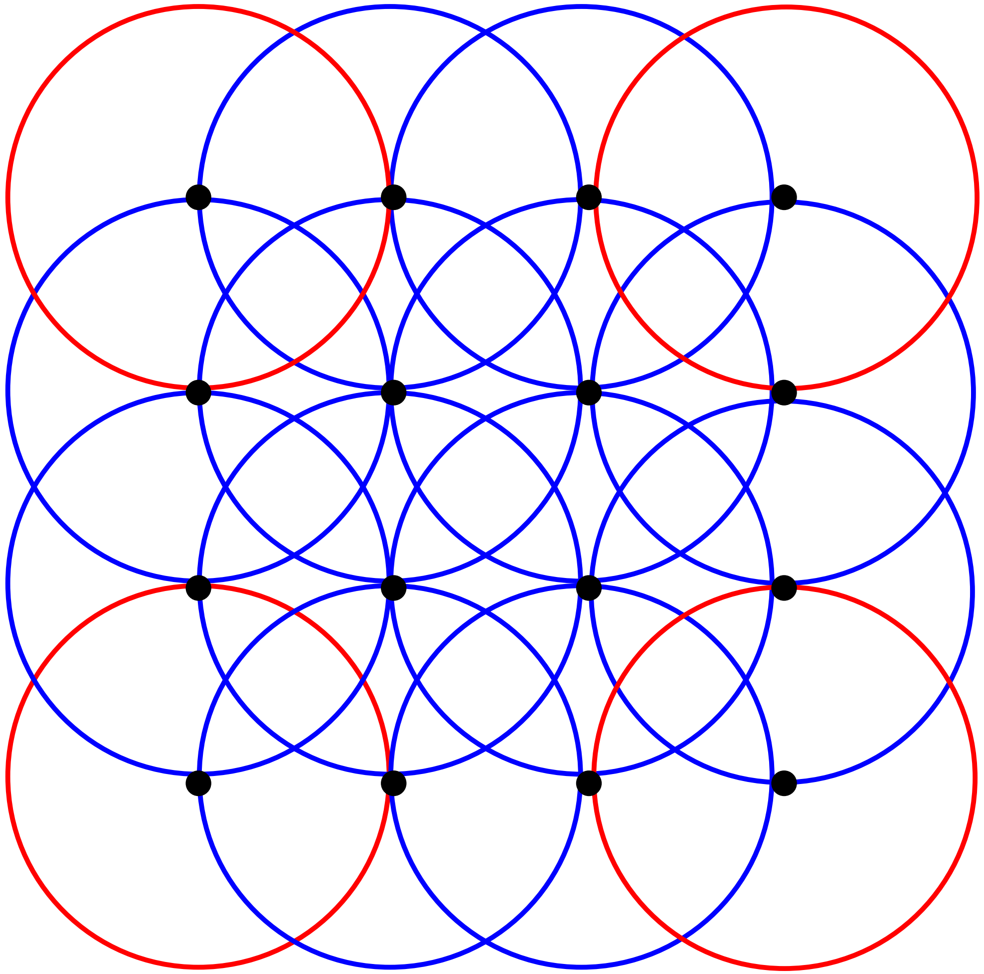
Theorem (Spencer, Szemerédi and Trotter 1984). If P is a set of n points in the plane, then the number k of pairs of points of P at unit distance is less than $n^{4/3}$.

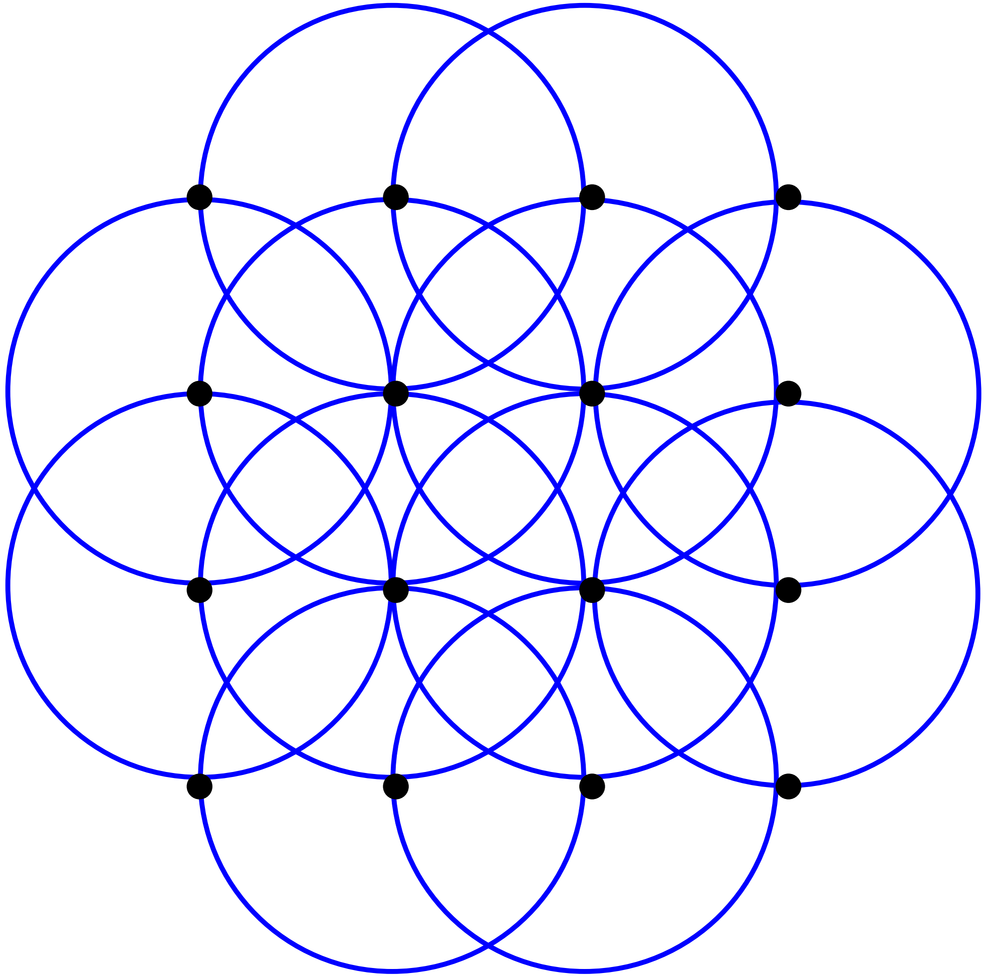
Proof (Székely 1997).
Draw a unit circle around each point of P.
If n_i is the number of circles which include exactly i points of P, then

$$\sum_{i=0}^{n-1} n_i = n \quad \text{and} \quad \sum_{i=0}^{n-1} i n_i = 2k$$

Define a graph H as follows:

- $V(H) := P$
- $E(H)$: arcs between consecutive points on circles containing at least three points of P





Theorem (Spencer, Szemerédi and Trotter 1984). If P is a set of n points in the plane, then the number k of pairs of points of P at unit distance is less than $n^{4/3}$.

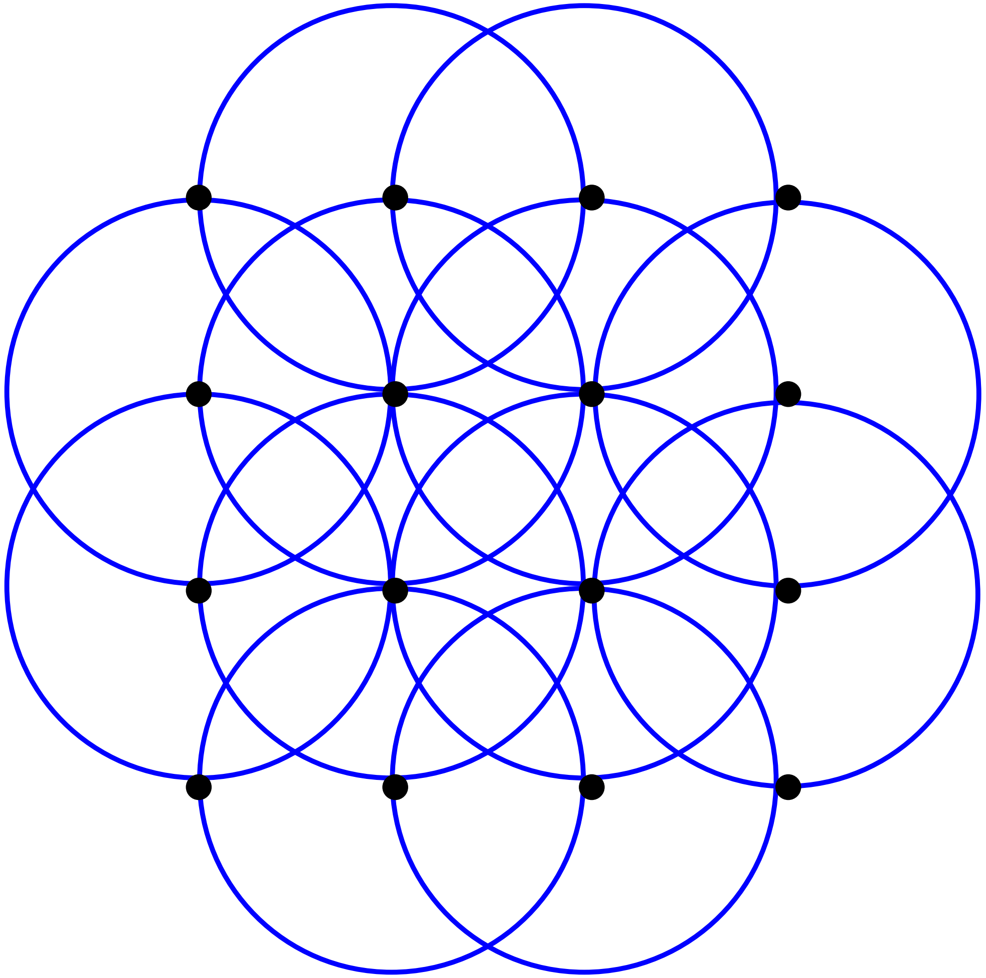
Proof (Székely 1997).
Draw a unit circle around each point of P.
If n_i is the number of circles passing which include exactly i points of P, then

$$\sum_{i=0}^{n-1} n_i = n \qquad \text{and} \qquad \sum_{i=0}^{n-1} i n_i = 2k$$

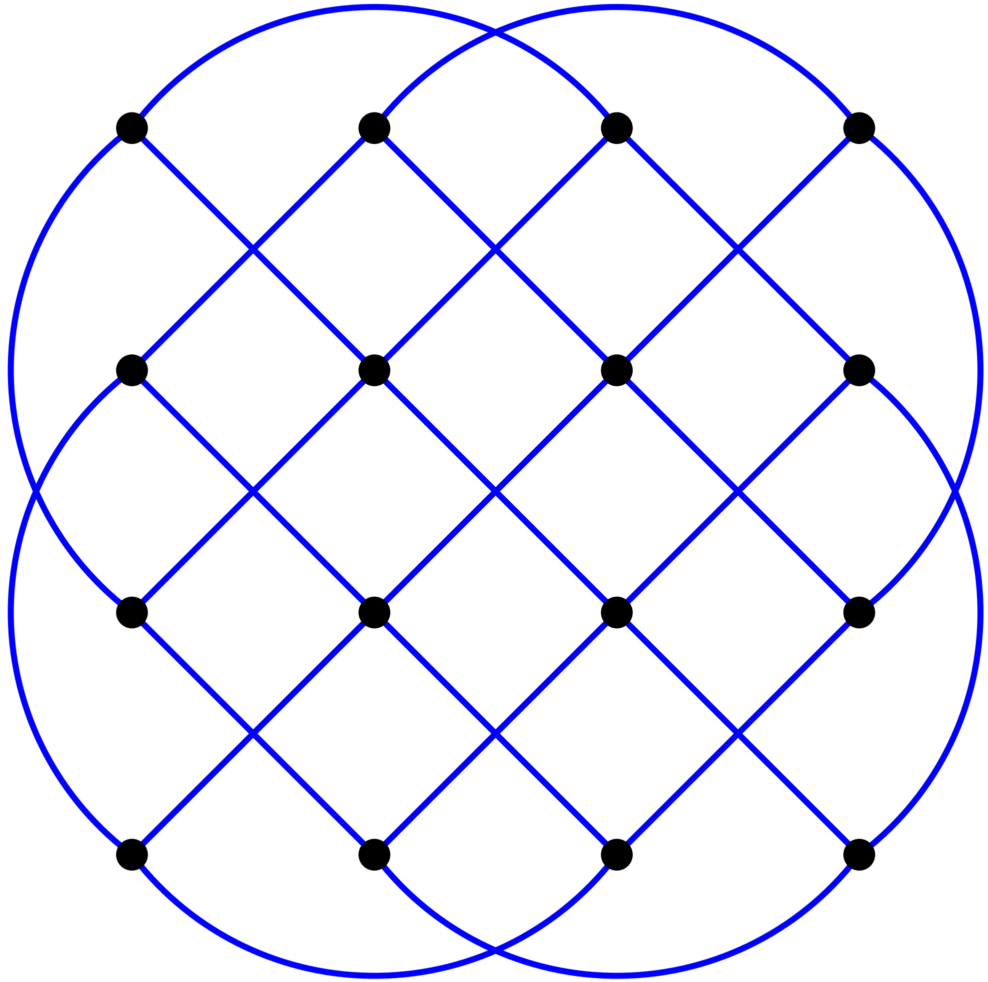
Each circle which includes i points of P contributes i edges to H.
Therefore

$$e(H) \;=\; \sum_{i=3}^{n-1} i n_i \;=\; 2k - n_1 - 2n_2 \;\geq\; 2k - 2n$$

Some pairs of vertices of H might be joined by two parallel edges.



Some pairs of vertices of H might be joined by two parallel edges. Delete one of each pair of parallel edges to get a simple graph G.



Some pairs of vertices of H might be joined by two parallel edges.

Delete one of each pair of parallel edges to get a simple graph G.

Because any two circles generate at most two crossings,

$$v(G) = n, \qquad e(G) \geq k - n, \qquad cr(G) \leq n(n-1)$$

- If $e(G) \geq 4n$, then by the Crossing Lemma

$$n(n-1) \geq cr(G) \geq \frac{(k-n)^3}{64n^2}$$

- If $e(G) < 4n$, then $k < 5n$.

In both cases, $k < n^{4/3}$.

GIRTH AND CHROMATIC NUMBER

Theorem (Mycielski 1955). *For any positive integer k, there exists a triangle-free k-chromatic graph G_k.*

Inductive Construction:

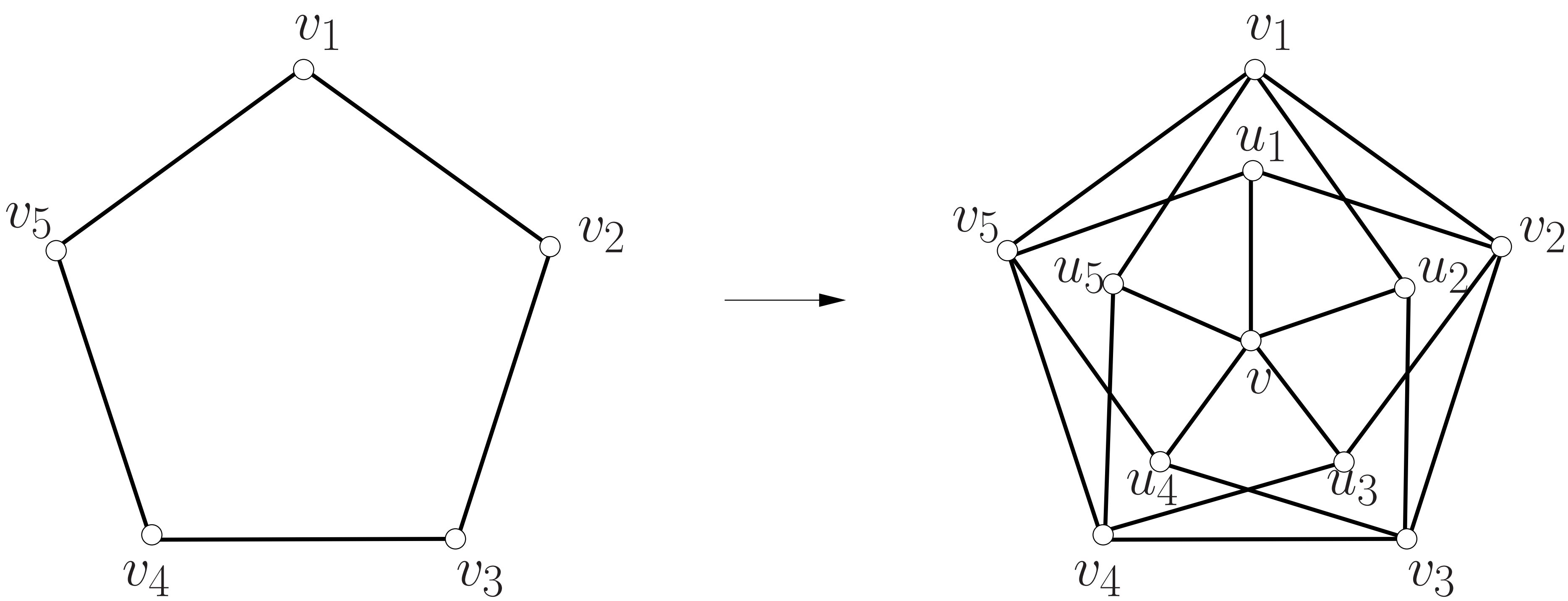
For $k = 1$ and $k = 2$, let $G_1 = K_1$ and $G_2 = K_2$.

Suppose that G_k is a triangle-free graph with chromatic number k, where $k \geq 2$. Let $V(G_k) = \{v_1, v_2, \ldots, v_n\}$.

Construct G_{k+1} from G_k as follows:

- add $n + 1$ new vertices $u_1, u_2, \ldots, u_n, v$
- for $1 \leq i \leq n$, join u_i to the neighbours of v_i in G_k, and to v.

Thus G_3 is the 5-cycle and G_4 the Grötzsch graph:



G_3 G_4

It is easy to see that G_{k+1} has no triangles:

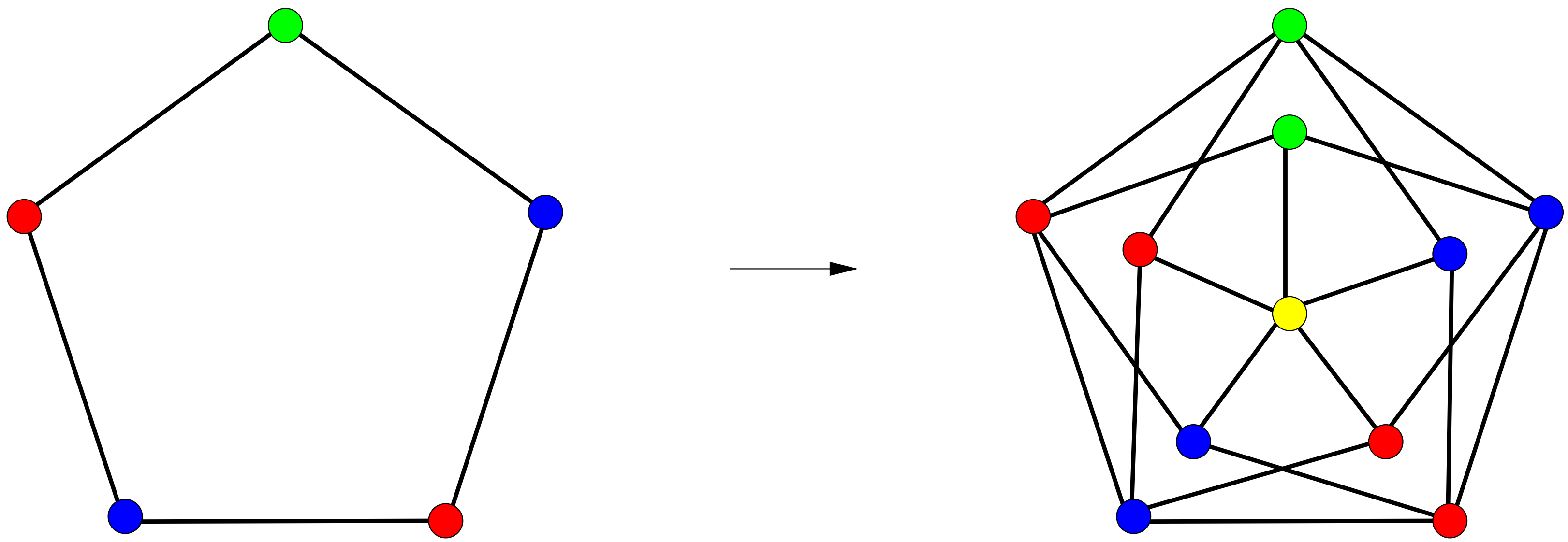
- Vertex v lies in no triangle.
- A triangle can contain at most one vertex of each stable set $\{u_i, v_i\}$.
- A triangle can contain at most one vertex of the stable set $\{u_1, u_2, \ldots, u_n\}$ and at most two vertices of the triangle-free graph G_k.
- But if $\{u_i, v_j, v_k\}$ induces a triangle, then so does $\{v_i, v_j, v_k\}$.

G_{k+1} *is* $(k+1)$*-colourable*:

- Consider a k-colouring of G_k.
- Assign the colour of v_i to u_i, $1 \leq i \leq n$.
- Assign a new colour to v.

G_{k+1} is $(k+1)$-colourable:

- Consider a k-colouring of G_k.
- Assign the colour of v_i to u_i, $1 \le i \le n$.
- Assign a new colour to v.



G_{k+1} is not k-colourable:

- Consider a hypothetical k-colouring of G_{k+1}.
- The restriction of this colouring to G_k is a k-colouring of G_k.
- Since G_k is k-chromatic, for each colour j, there exists a vertex v_i of colour j which is adjacent in G_k to vertices of every other colour.
- Because u_i has precisely the same neighbours in G_k as v_i, the vertex u_i also has colour j.
- Therefore, each of the k colours appears on at least one of the vertices u_i.
- No colour is now available for the vertex v.

MARKOV'S INEQUALITY

MARKOV'S INEQUALITY. *Let X be a nonnegative random variable and t a positive real number. Then*

$$P(X \geq t) \leq \frac{E(X)}{t}$$

PROOF.

$$\begin{aligned} E(X) = \sum_s sP(X = s) &\geq \sum_{s \geq t} sP(X = s) \\ &\geq \sum_{s \geq t} tP(X = s) = tP(X \geq t) \end{aligned}$$

THEOREM (Erdős 1959). *For each integer k, there is a graph with girth at least k and chromatic number at least k.*

PROOF.

- X: number of cycles of length at most $k-1$ in $\mathcal{G}(n,p)$
- Y: number of stable sets of cardinality $t+1$ in $\mathcal{G}(n,p)$
- By Linearity of Expectation,

$$E(X) = \sum_{i=3}^{k-1} \frac{(n)_i}{2i} p^i < \sum_{i=1}^{k-1} (np)^i = \frac{(np)^k - 1}{np - 1}$$

$$E(Y) = \binom{n}{t+1} (1-p)^{\binom{t+1}{2}} < n^{t+1} e^{-p\binom{t+1}{2}} = \left(n e^{-pt/2}\right)^{t+1}$$

- Setting $p := n^{-(k-1)/k}$ and $t := \lceil 4\, n^{(k-1)/k} \ln n \rceil$,

$$E(X) < \frac{n-1}{n^{1/k}-1} \qquad \text{and} \qquad E(Y) < (ne^{-2\ln n})^{t+1} = n^{-t-1}$$

- By Markov's Inequality,

$$P(X > n/2) \; < \; \frac{E(X)}{n/2} = o(1) \quad \text{and} \quad P(Y \geq 1) \; \leq \; E(Y) = o(1)$$

- Choose n large enough to guarantee that

$$P(X > n/2) < 1/2 \qquad \text{and} \qquad P(Y \geq 1) < 1/2$$

- By the Counting Sieve, there is a graph G on n vertices with at most $n/2$ circuits of length at most k and no stable set of cardinality greater than t.

- Delete one vertex from each circuit of G of length less than k.
- The resulting graph G^* has girth at least k, at least $n/2$ vertices, and no stable set of cardinality greater than t, so

$$\chi(G^*) \geq \frac{v(G^*)}{\alpha(G^*)} \geq \frac{n}{2t} \sim \frac{n^{1/k}}{8 \ln n}$$

- Choose n large enough to guarantee that $\chi(G^*) \geq k$.

References

N. Alon and J. Spencer, *The Probabilistic Method*, Third Edition, Wiley Interscience, 2008.

B. Bollobás, *Random Graphs*, Second Edition. Cambridge Studies in Advanced Mathematics **73**. Cambridge University Press, Cambridge, 2001.

J.A. Bondy and U.S.R. Murty, *Graph Theory*, Graduate Texts in Mathematics **244**, Springer, 2008.

S. Janson, T. Luczak and A. Rucinski, *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

M. Molloy and B.A. Reed, *Graph Colouring and the Probabilistic Method*, Springer, 2002.