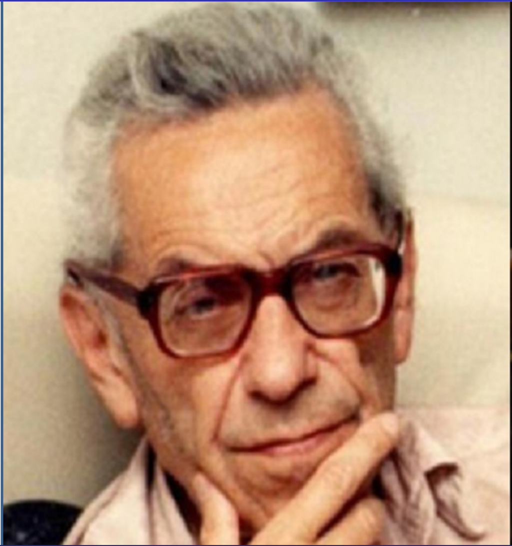
The Erdős-Hajnal Conjecture, structured non-linear graph-based hashing and b-matching anonymization via perfect matchings

Krzysztof Choromanski

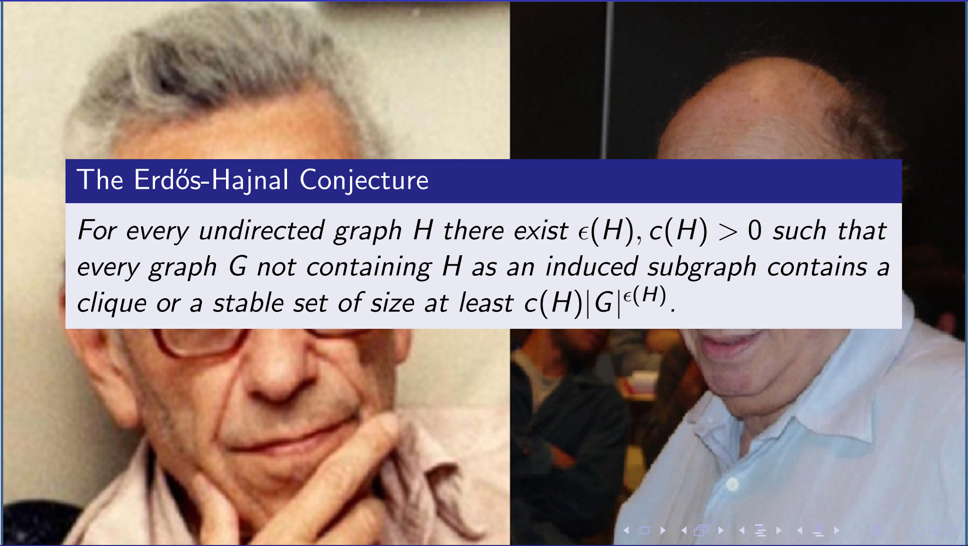
Google Research, New York City

October 19 2015

The Conjecture - Polynomial Phenomenon



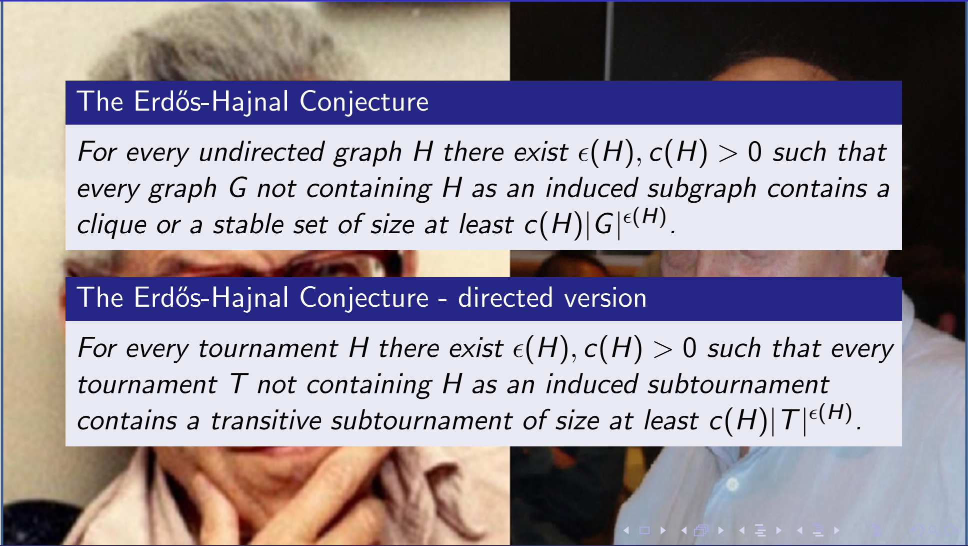
The Conjecture - Polynomial Phenomenon



The Erdős-Hajnal Conjecture

For every undirected graph H there exist $\epsilon(H), c(H) > 0$ *such that every graph G not containing H as an induced subgraph contains a clique or a stable set of size at least* $c(H)|G|^{\epsilon(H)}$.

The Conjecture - Polynomial Phenomenon



The Erdős-Hajnal Conjecture

For every undirected graph H there exist $\epsilon(H), c(H) > 0$ such that every graph G not containing H as an induced subgraph contains a clique or a stable set of size at least $c(H)|G|^{\epsilon(H)}$.

The Erdős-Hajnal Conjecture - directed version

For every tournament H there exist $\epsilon(H), c(H) > 0$ such that every tournament T not containing H as an induced subtournament contains a transitive subtournament of size at least $c(H)|T|^{\epsilon(H)}$.

The Conjecture - Polynomial Phenomenon

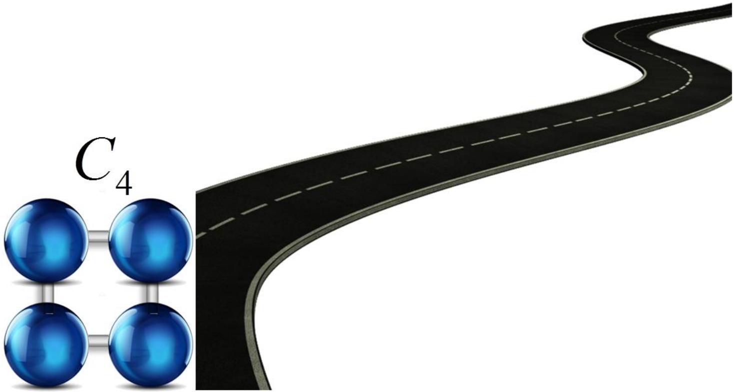
The Erdős-Hajnal Conjecture

For every undirected graph H there exist $\epsilon(H), c(H) > 0$ such that every graph G not containing H as an induced subgraph contains a clique or a stable set of size at least $c(H)|G|^{\epsilon(H)}$.

The Erdős-Hajnal Conjecture - directed version

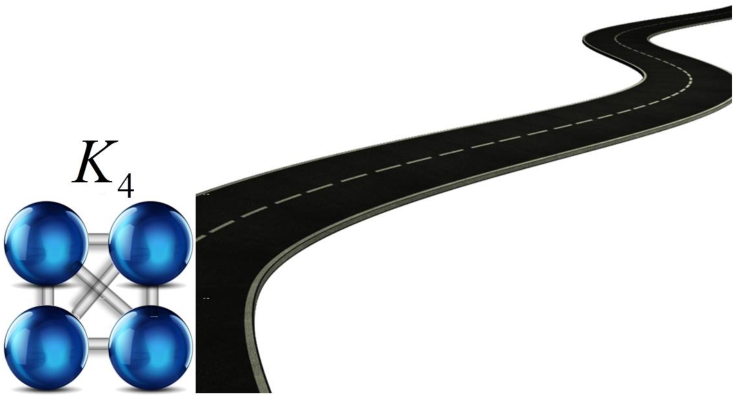
For every tournament H there exist $\epsilon(H), c(H) > 0$ such that every tournament T not containing H as an induced subtournament contains a transitive subtournament of size at least $c(H)|T|^{\epsilon(H)}$.

What was known...

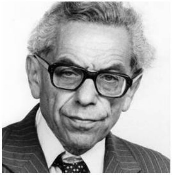


What was known...

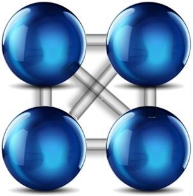
What was known...

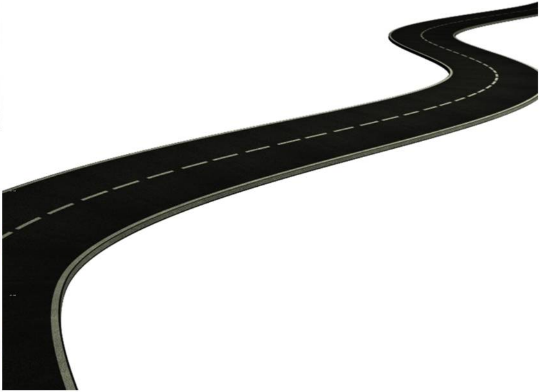


What was known...

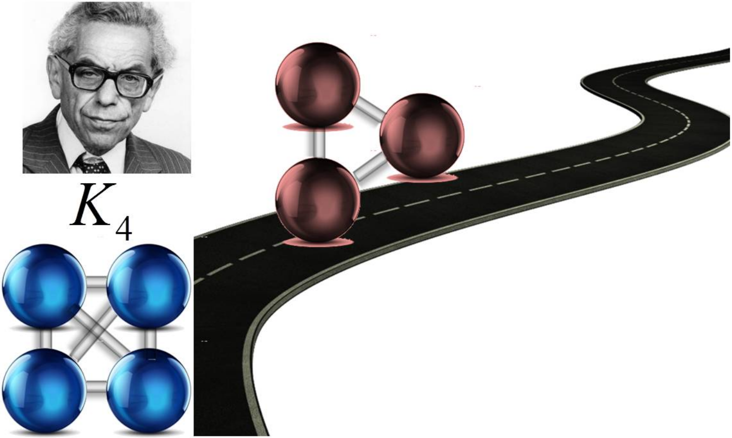


K_4

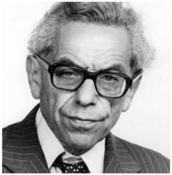




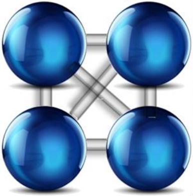
What was known...

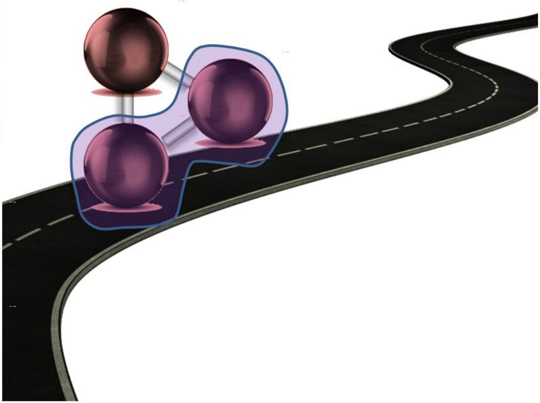


What was known...

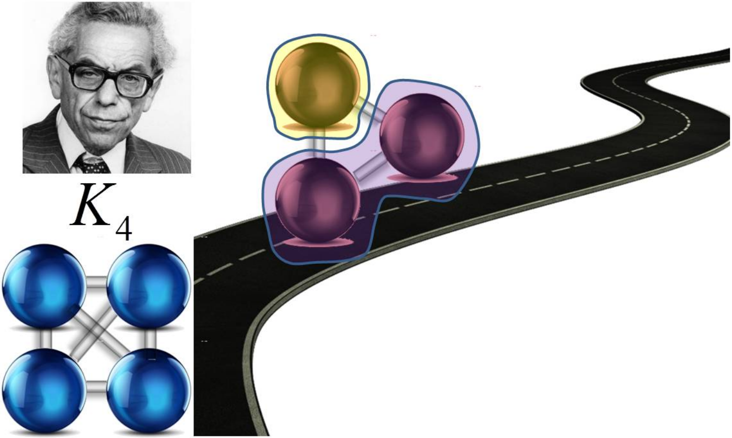


K_4

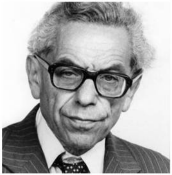




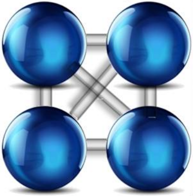
What was known...

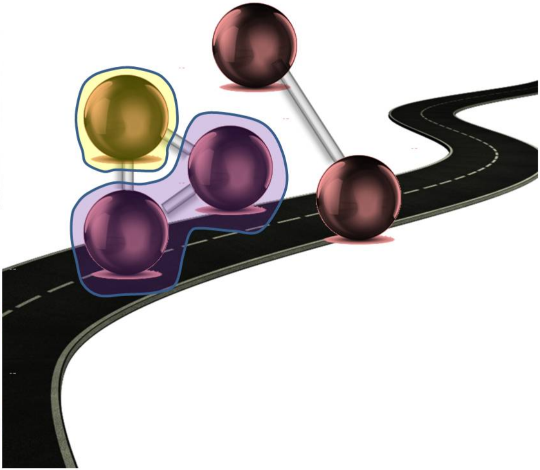


What was known...

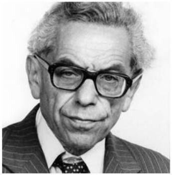


K_4

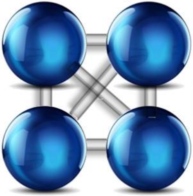


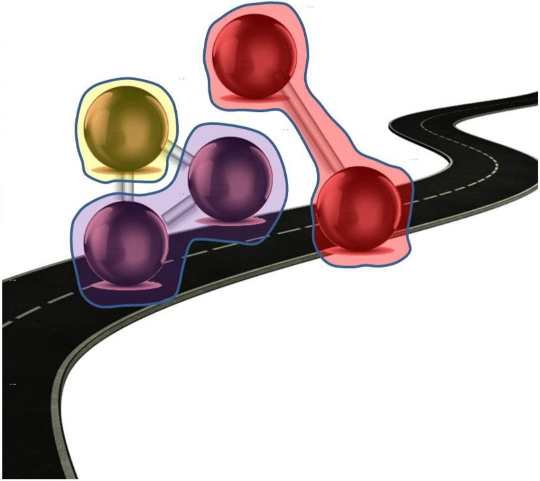


What was known...

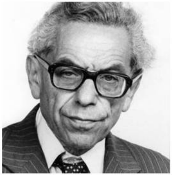


K_4

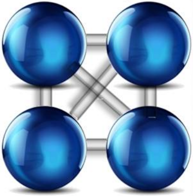


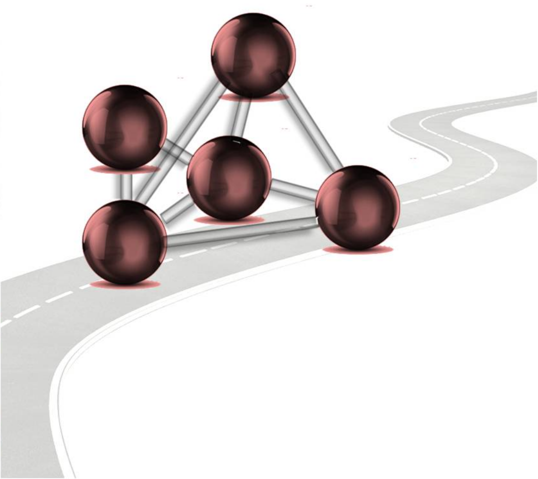


What was known...

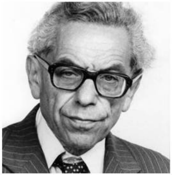


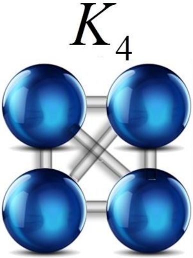
K_4

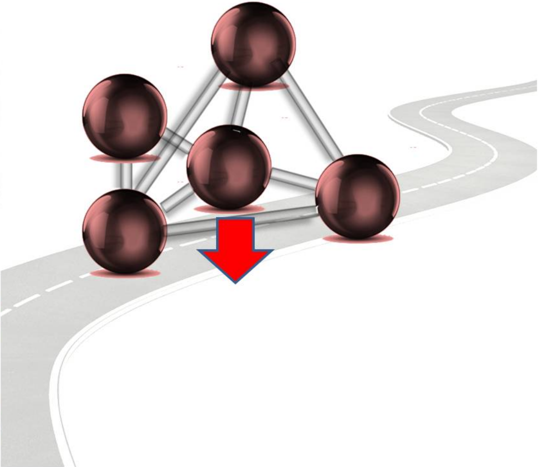




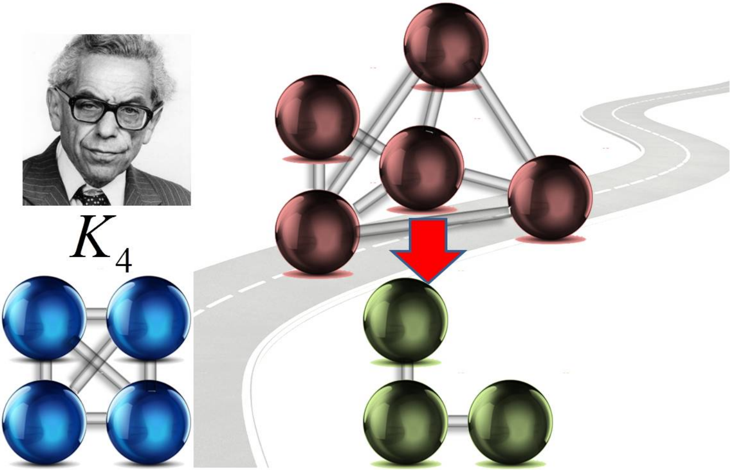
What was known...



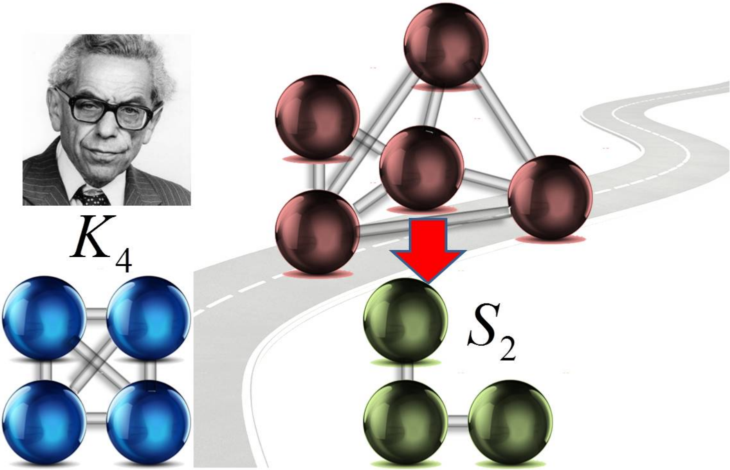




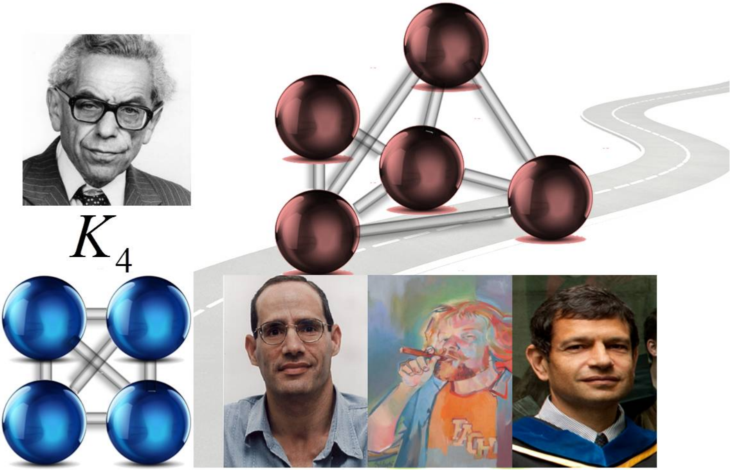
What was known...



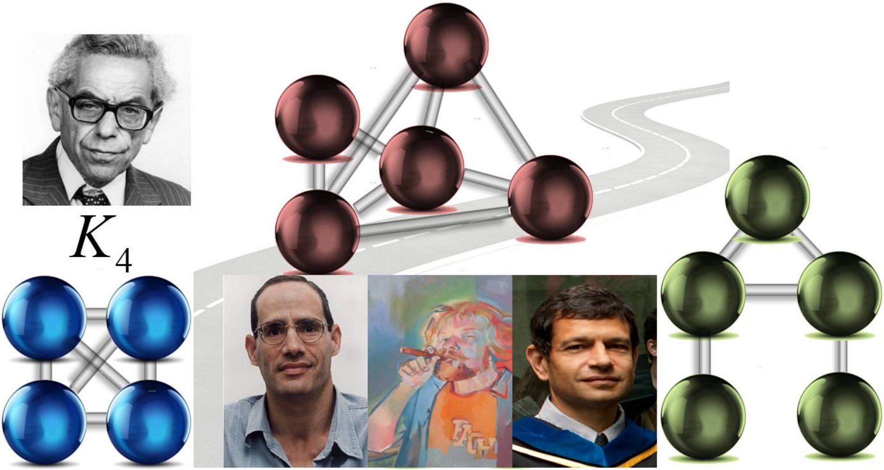
What was known...



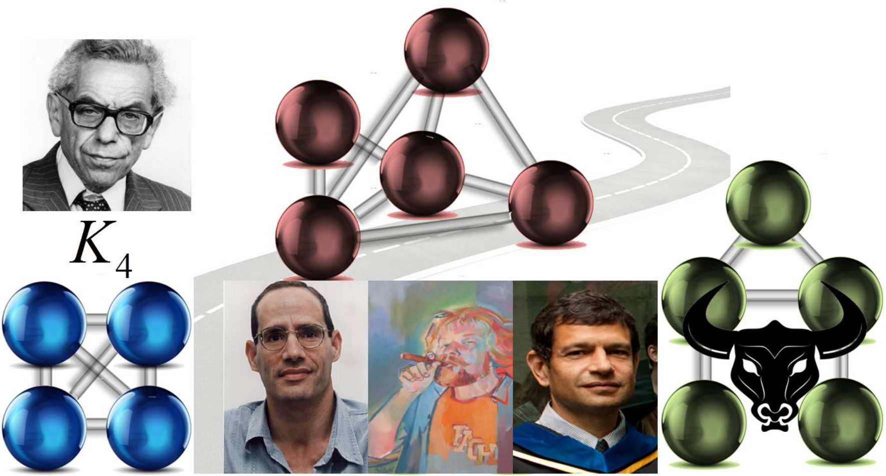
What was known...



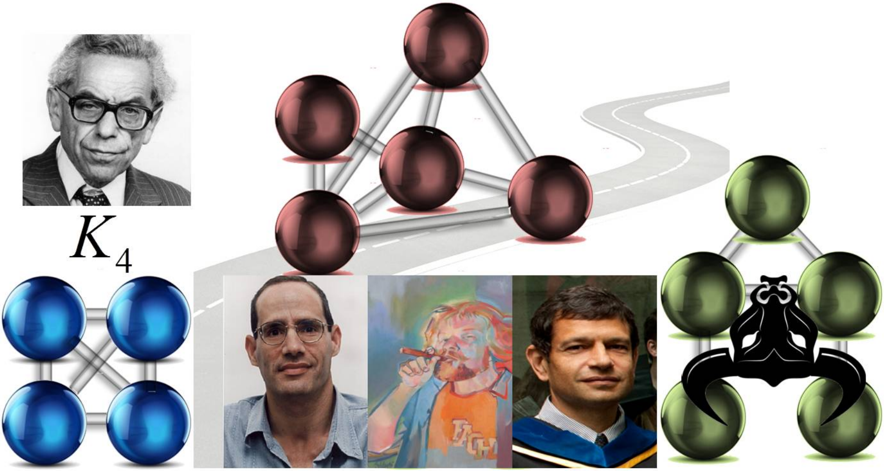
What was known...



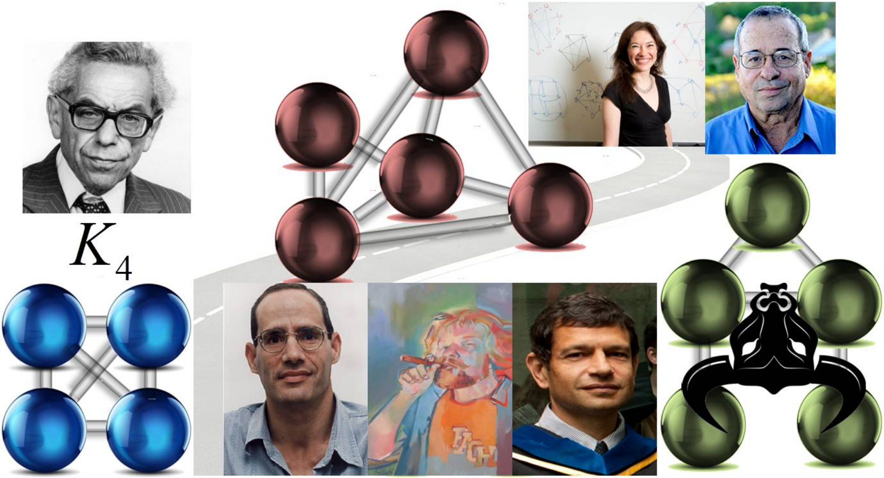
What was known...



What was known...

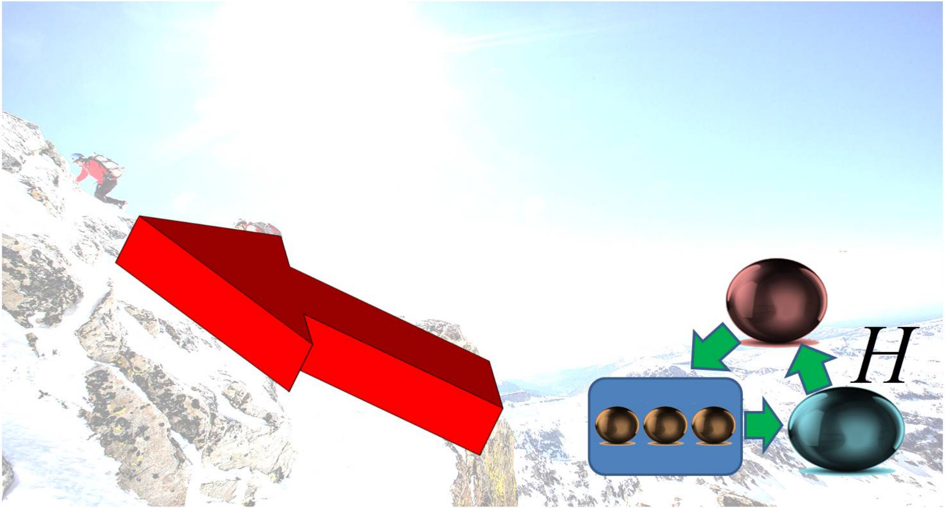


What was known...

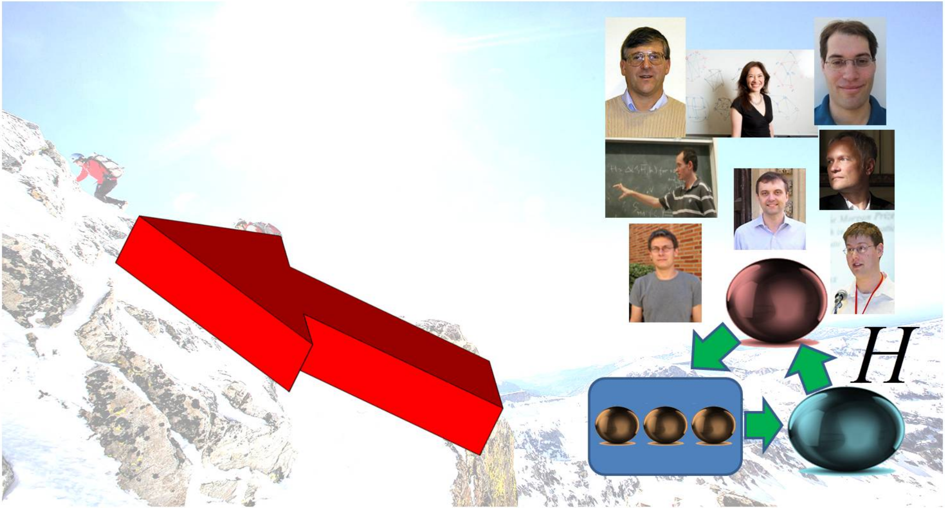


Directed case revolution

Directed case revolution



Directed case revolution



Tournaments satisfying the Conjecture in the linear sense

Definition

Tournament H is a celebrity if there exists $c(H) > 0$ such that every H-free n-vertex tournament contains a transitive subtournament of order at least $c(H)n$.

Theorem (Berger, Choromanski, Chudnovsky, Fox, Loebl, Scott, Seymour, Thomasse '11)

Tournament H is a celebrity iff either:

- *it is not strongly connected and is of the form $H_1 \implies H_2$, where H_1, H_2 are celebrities or,*
- *is strongly connected and is of the form $\Delta(1, T_k, D)$, where D is a celebrity and T_k is a transitive tournament on k vertices.*

Tournaments satisfying the Conjecture in the linear sense

Definition

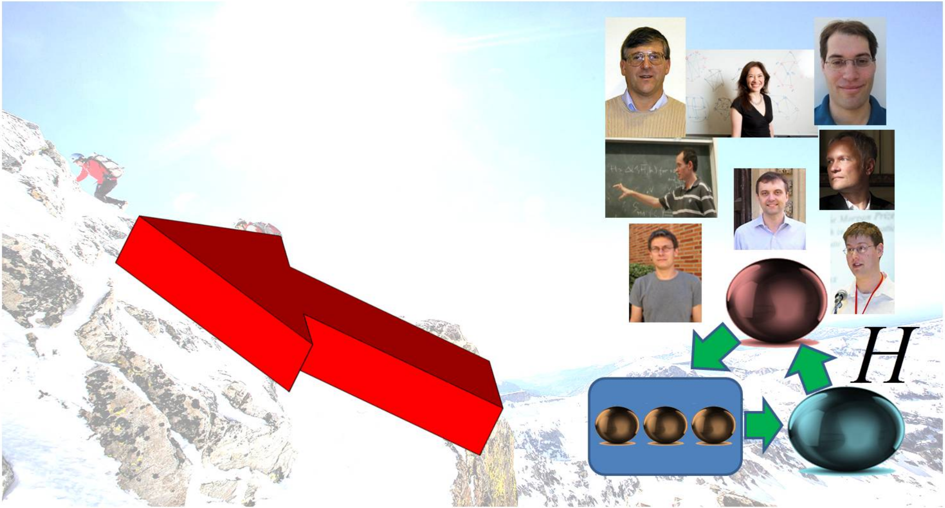
A dichromatic number $\chi(T)$ of the tournament T is the smallest number of colors that can be used to color its vertices in such a way that there does not exist a monochromatic directed cycle.

Theorem (Berger, Choromanski, Chudnovsky, Fox, Loebl, Scott, Seymour, Thomasse '11)

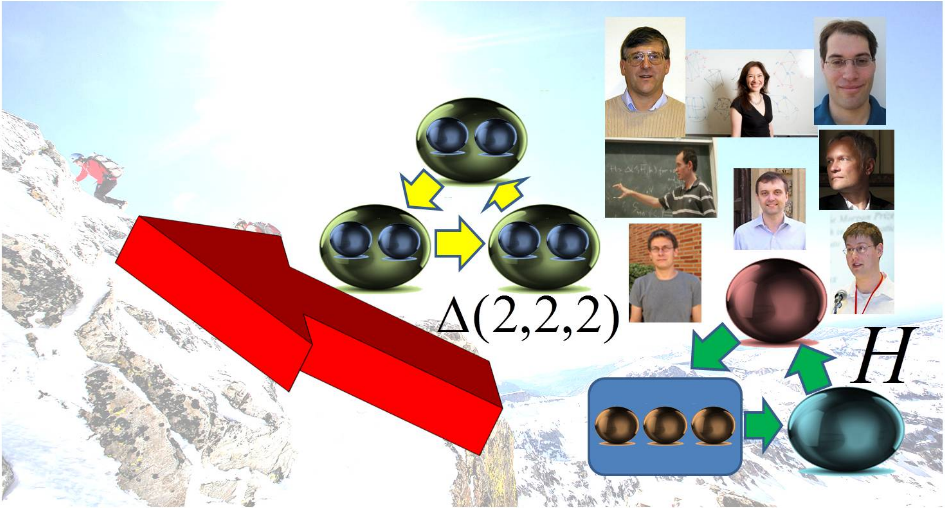
Tournament H is a celebrity iff there exists $d(H)$ such that every H-free tournament T satisfies:

$$\chi(T) \leq d(H).$$

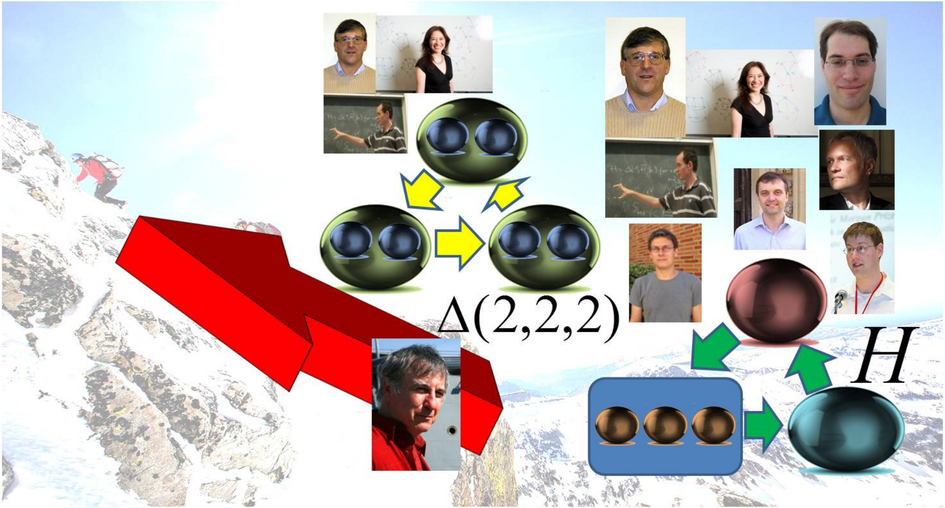
Directed case revolution



Directed case revolution



Directed case revolution



Tournaments satisfying the Conjecture in the pseudolinear sense

Definition

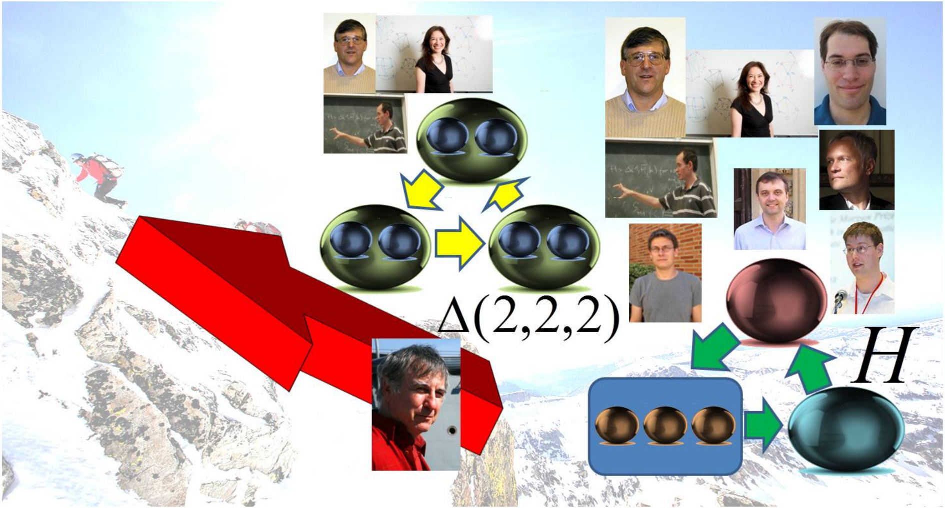
Tournament H is a pseudocelebrity if it is not a celebrity, but there exist $c(H), d(H) > 0$ such that every n-vertex H-free tournament T satisfies: $\chi(T) \leq c(H) \log^{d(H)}(n)$.

Theorem (Choromanski, Chudnovsky, Seymour '12)

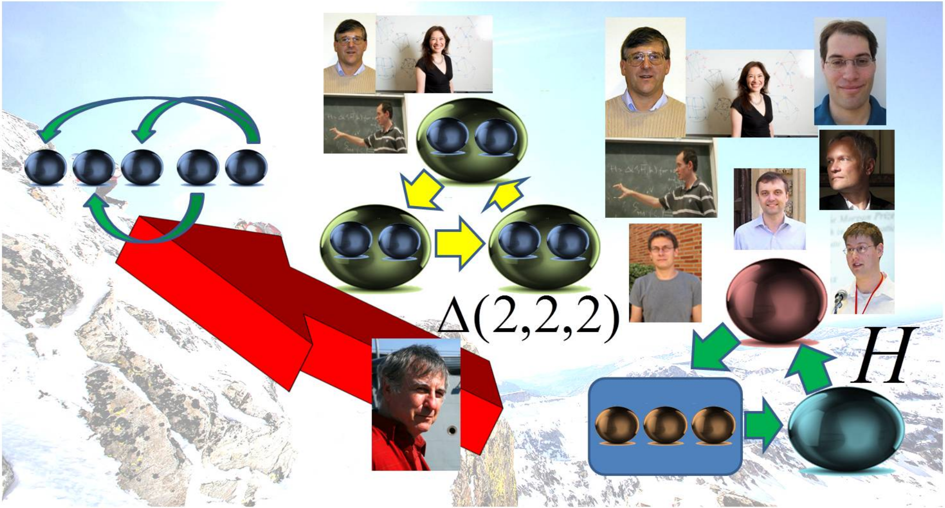
Tournament H is a pseudocelebrity if it is of the form:

- $H_1 \implies H_2$, *where both H_is are pseudocelebrities or one is a celebrity and the other one is a pseudocelebrity or*
- $\Delta(1, T_k, H)$ *or* $\Delta(2, T_k, T_k)$, *where H is a pseudocelebrity.*

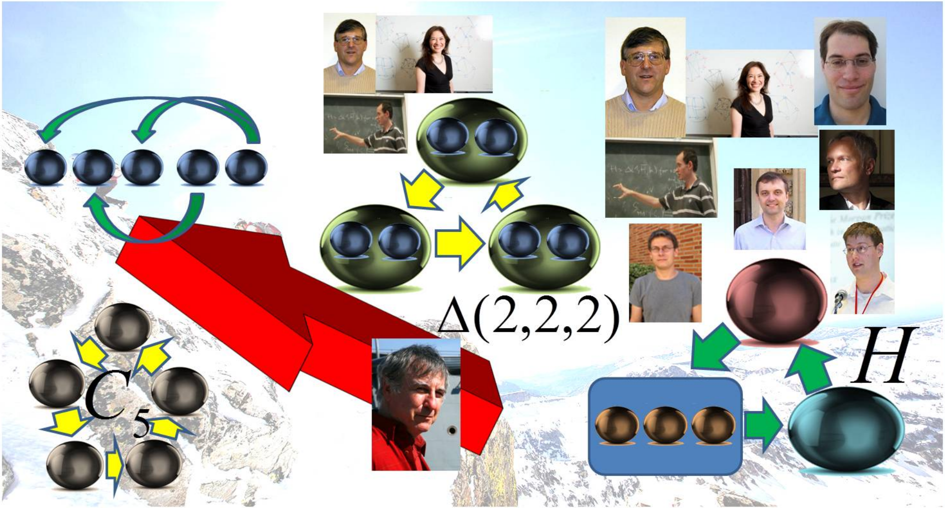
Directed case revolution



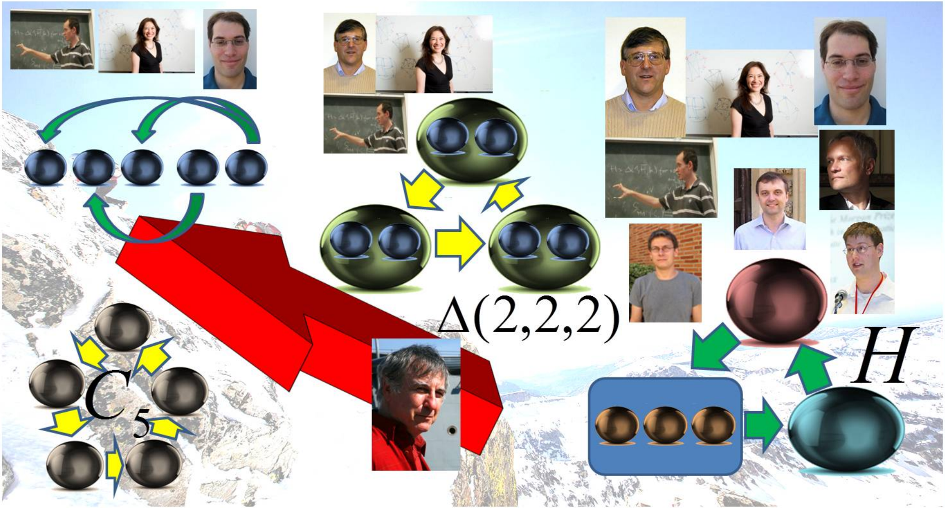
Directed case revolution



Directed case revolution

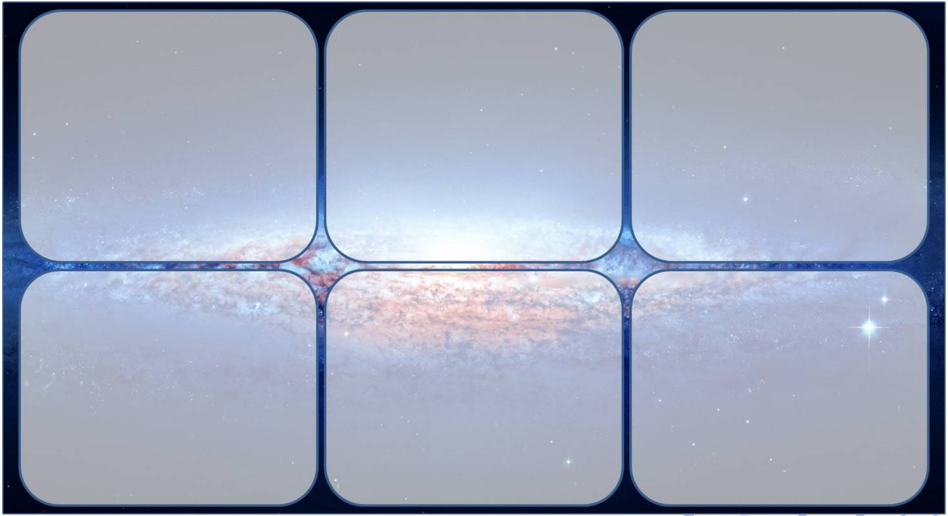


Directed case revolution

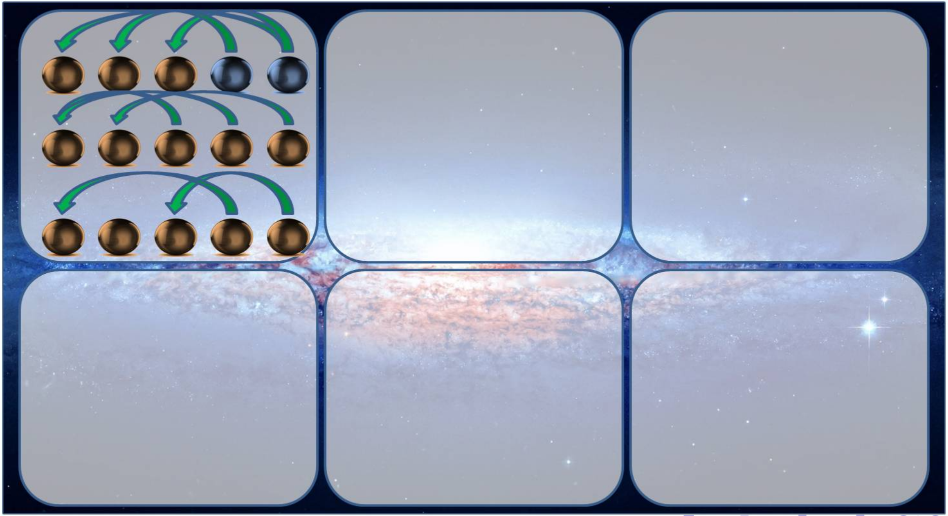


Galaxies, constellations, nebulae...

Galaxies, constellations, nebulae...



Galaxies, constellations, nebulae...



Infinitely many prime Erdős-Hajnal tournaments

Theorem (Berger, Choromanski, Chudnovsky '12)

Every galaxy satisfies the Erdős-Hajnal Conjecture. In particular, every directed path satisfies the Conjecture.

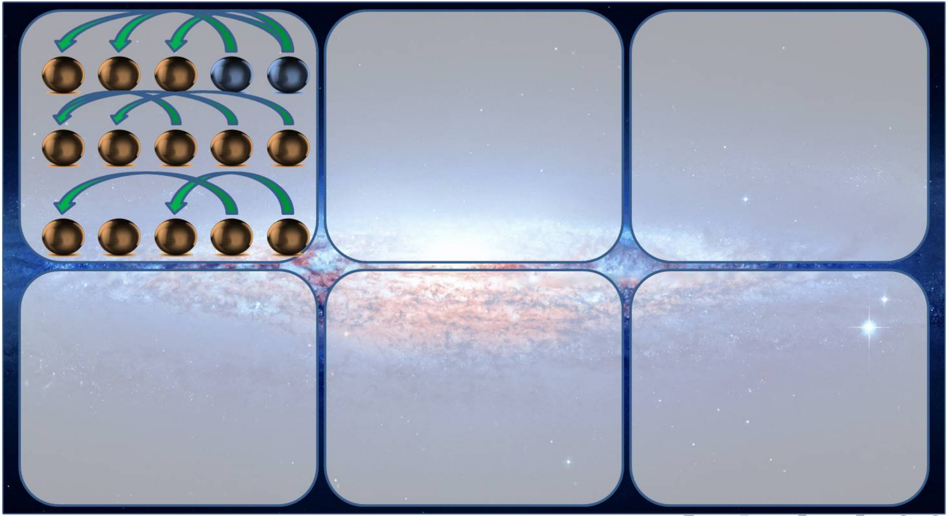
Theorem (Choromanski '12)

Tournament C_5 satisfies the Conjecture.

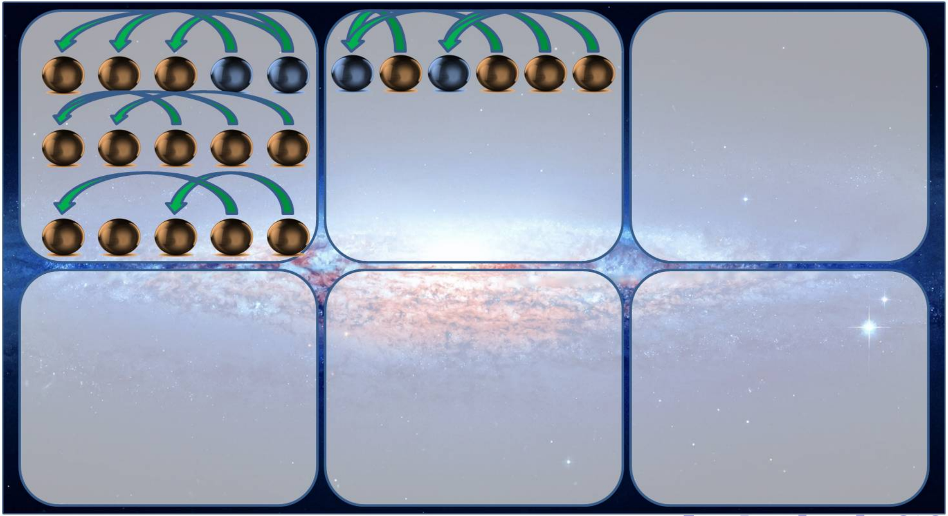
Corollary

Every tournament on at most five vertices satisfies the Conjecture.

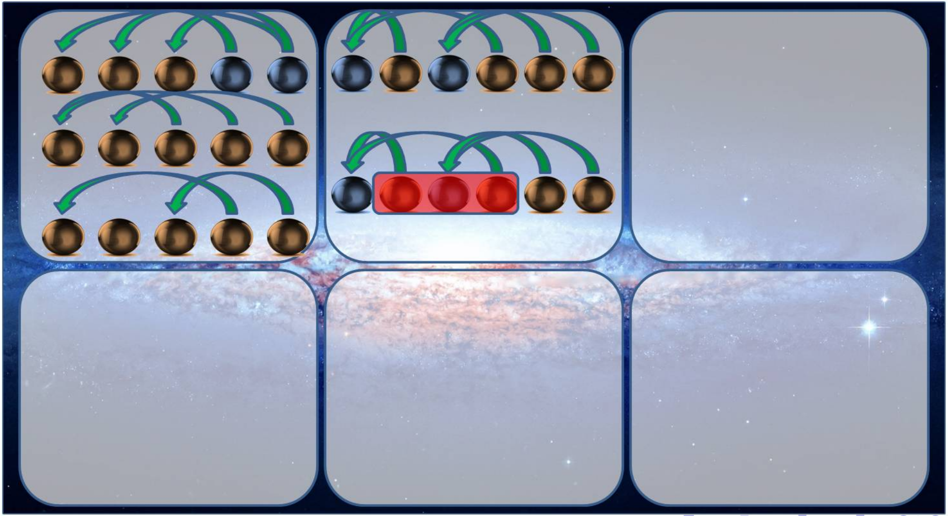
Galaxies, constellations, nebulae...



Galaxies, constellations, nebulae...



Galaxies, constellations, nebulae...

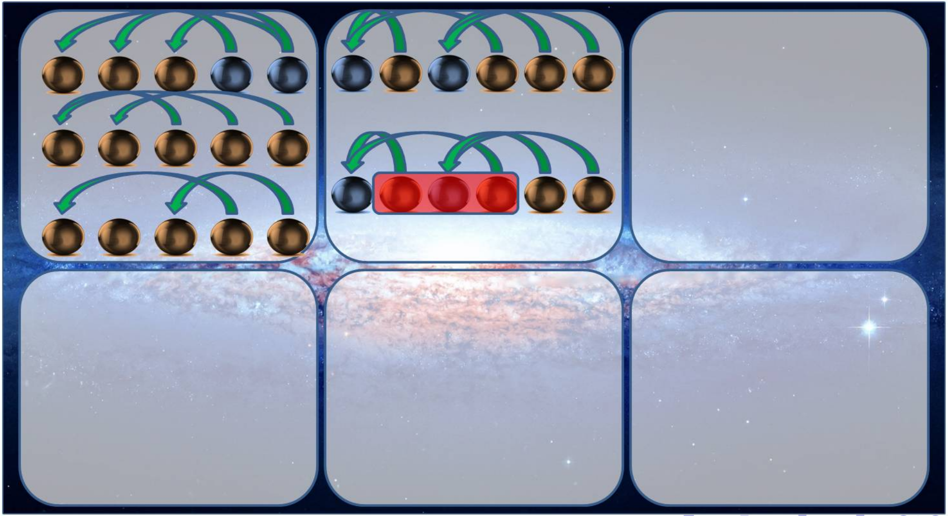


Going beyond galaxies

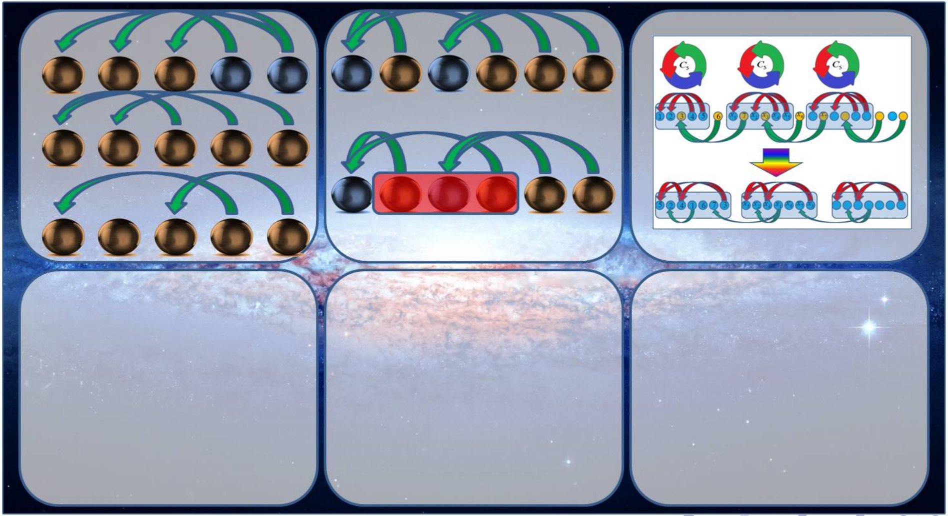
Theorem (Choromanski '12)

Every constellation satisfies the Erdős-Hajnal Conjecture.

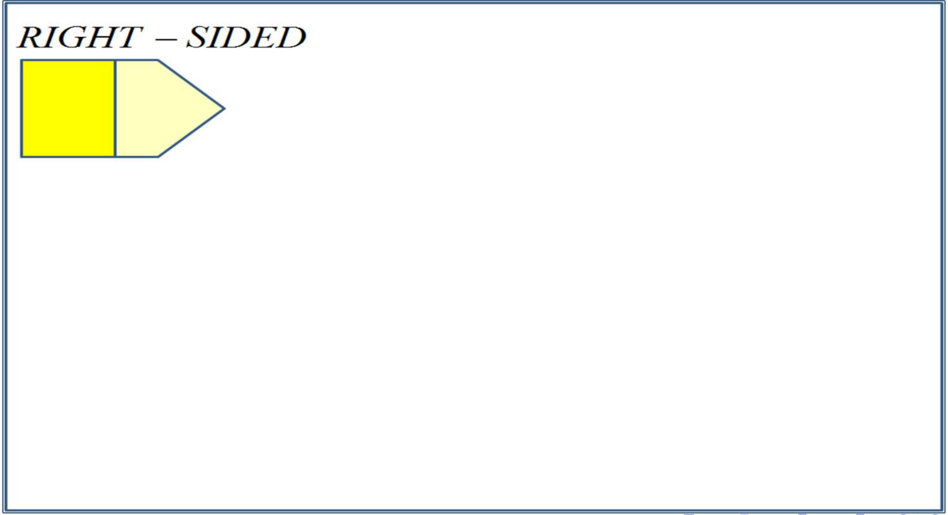
Galaxies, constellations, nebulae...



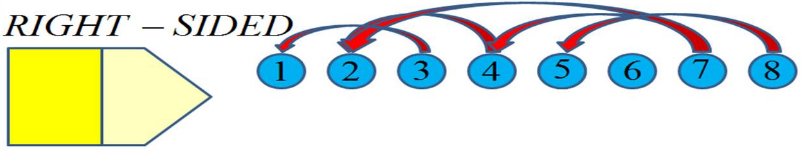
Galaxies, constellations, nebulae...



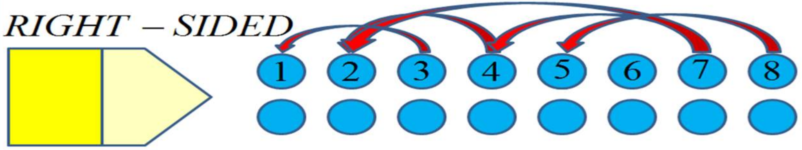
Combining tournaments...



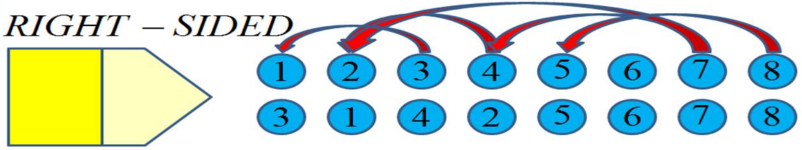
Combining tournaments...



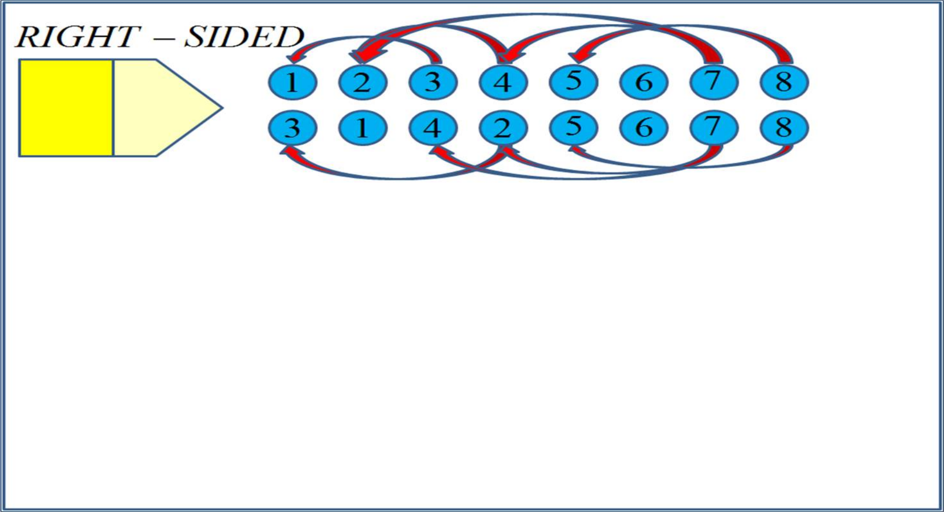
Combining tournaments...



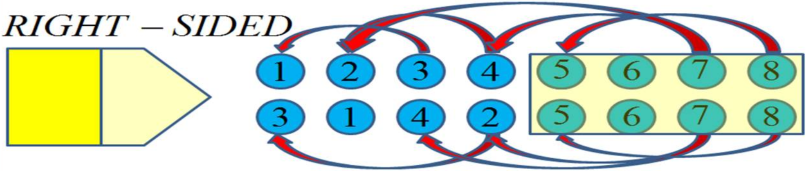
Combining tournaments...



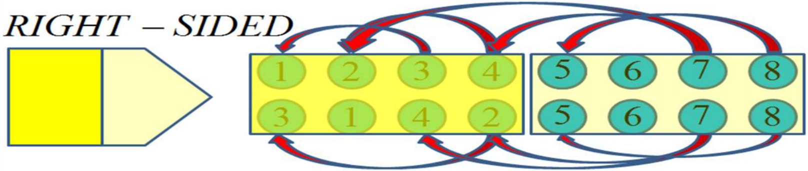
Combining tournaments...



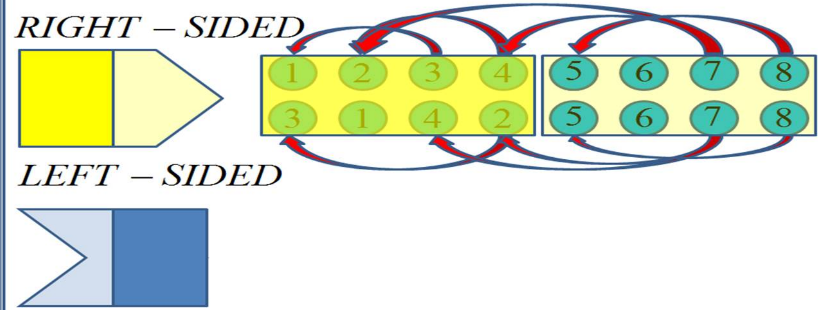
Combining tournaments...



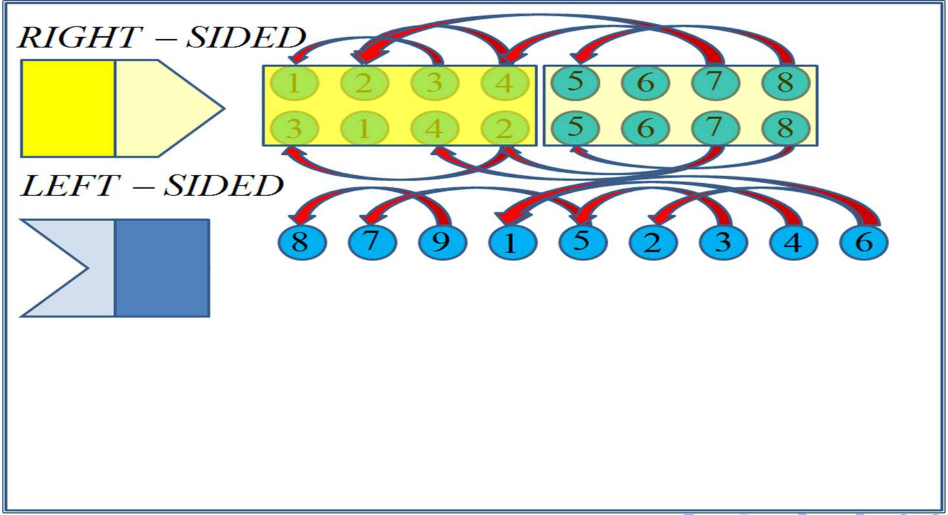
Combining tournaments...



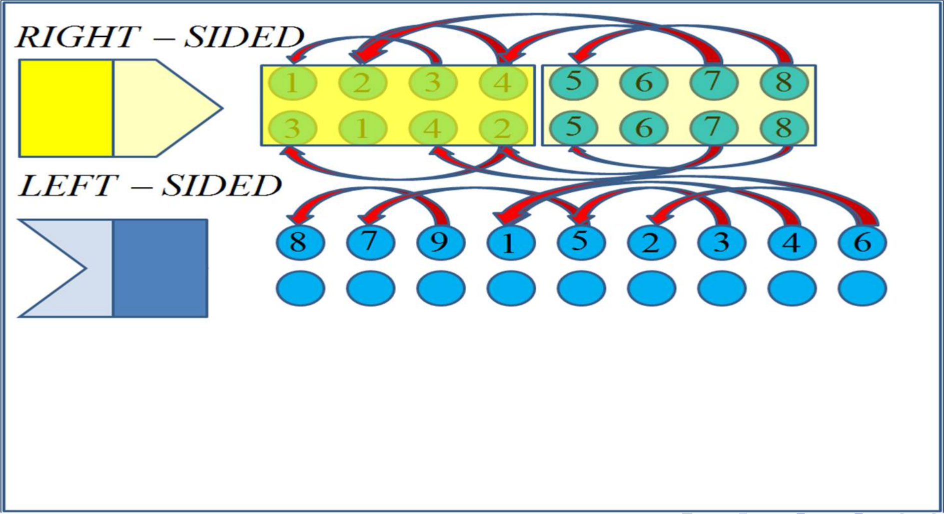
Combining tournaments...



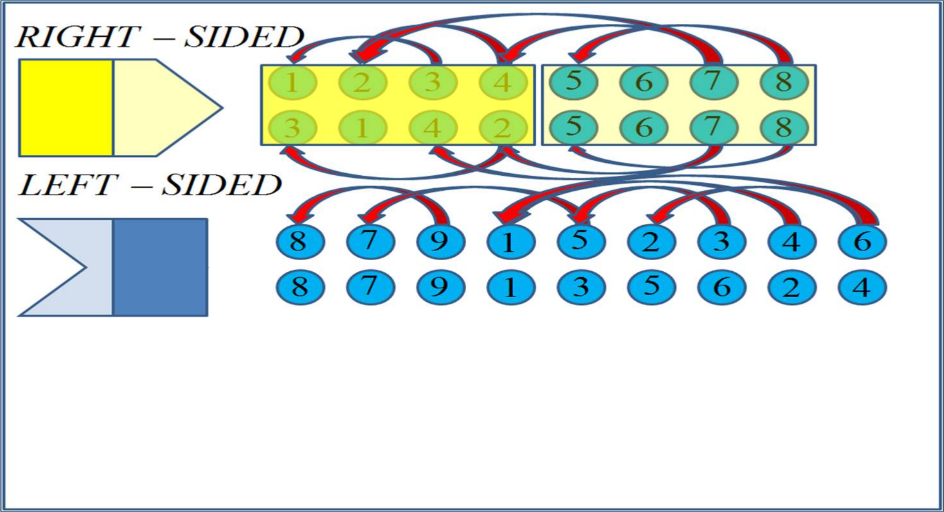
Combining tournaments...



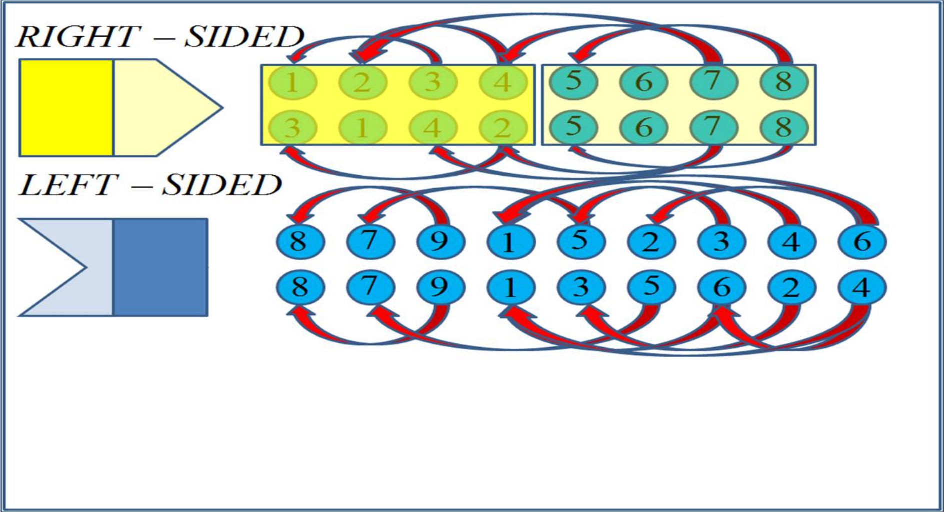
Combining tournaments...



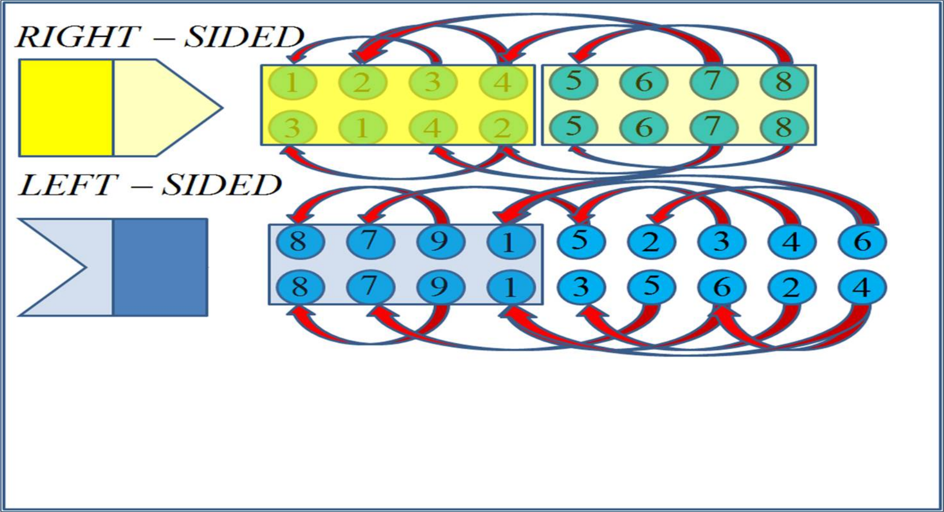
Combining tournaments...



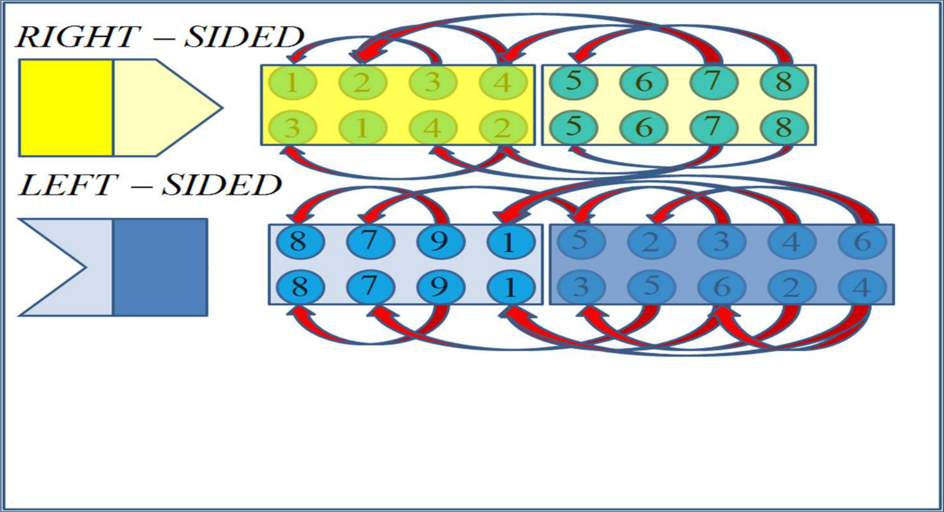
Combining tournaments...



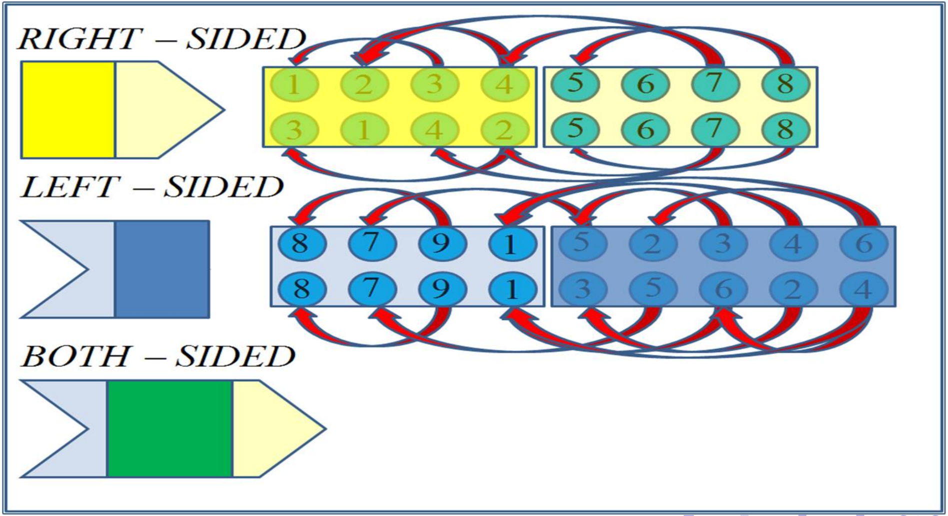
Combining tournaments...



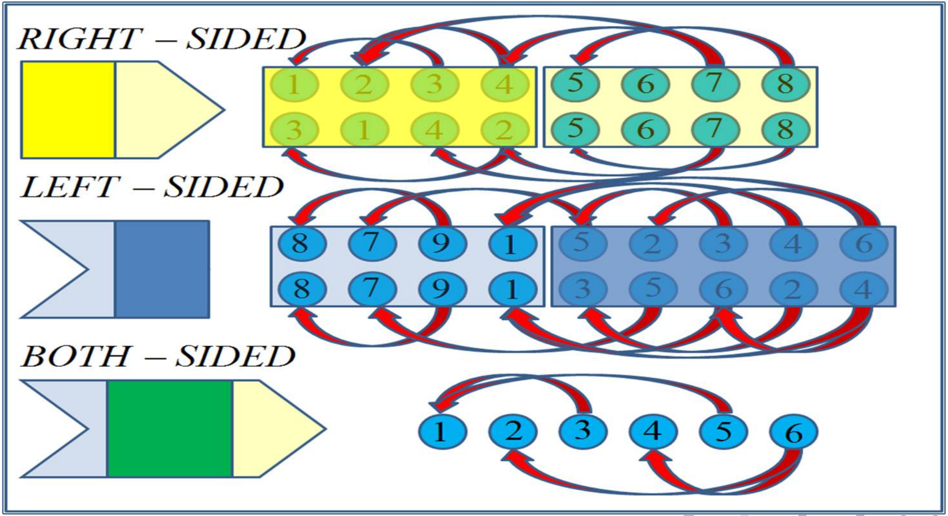
Combining tournaments...



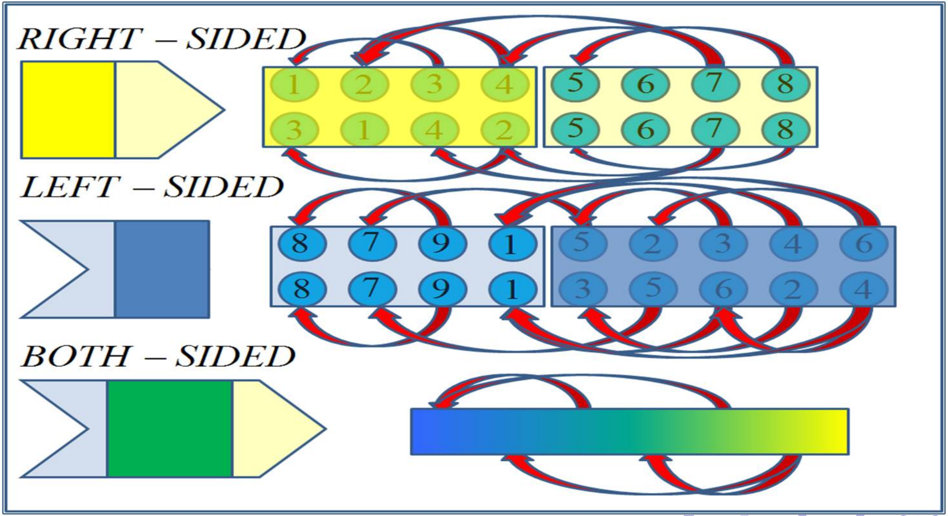
Combining tournaments...



Combining tournaments...

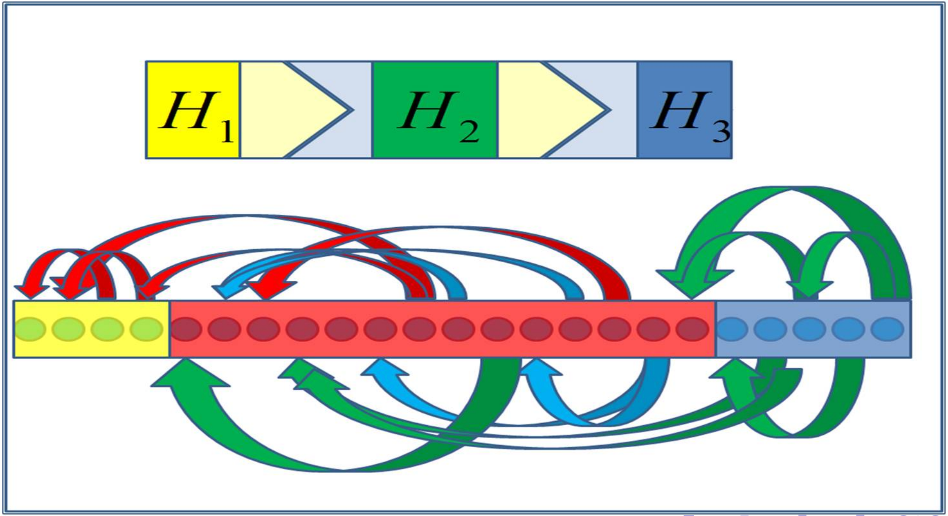


Combining tournaments...

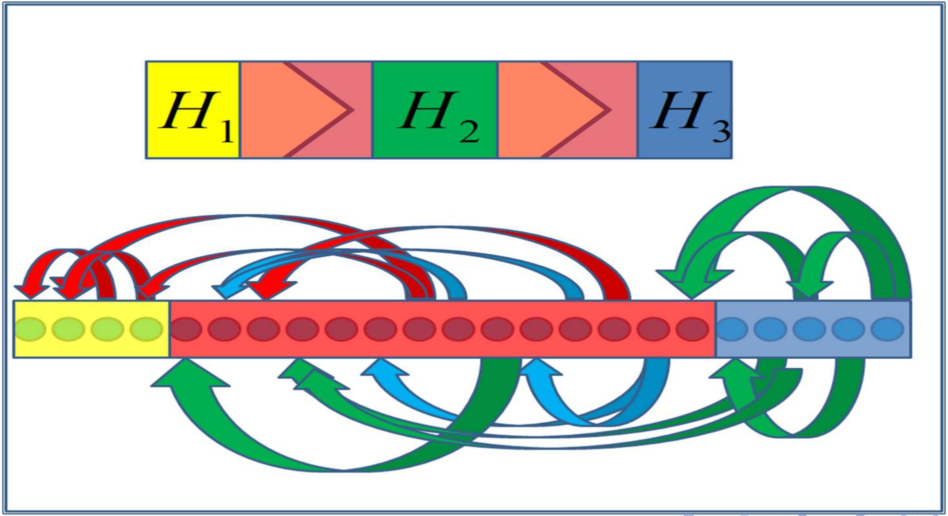


Combining tournaments...

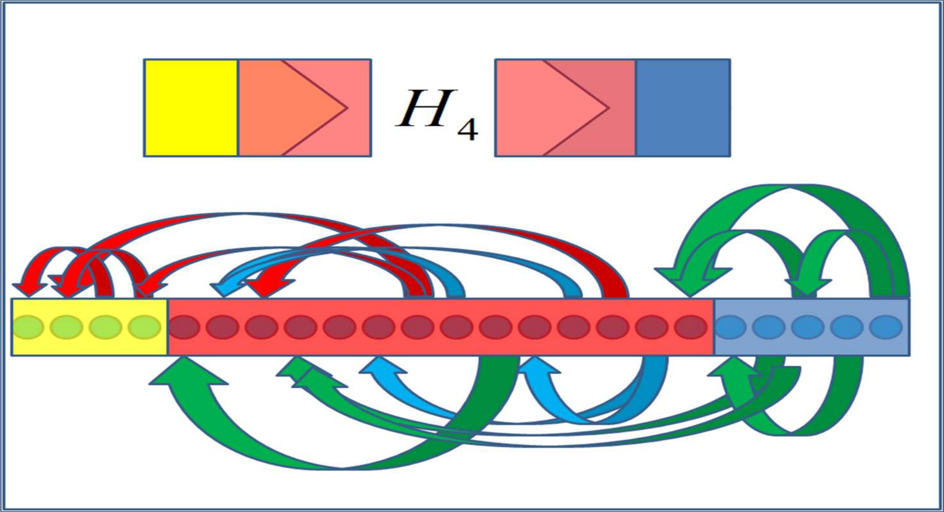
Combining tournaments...



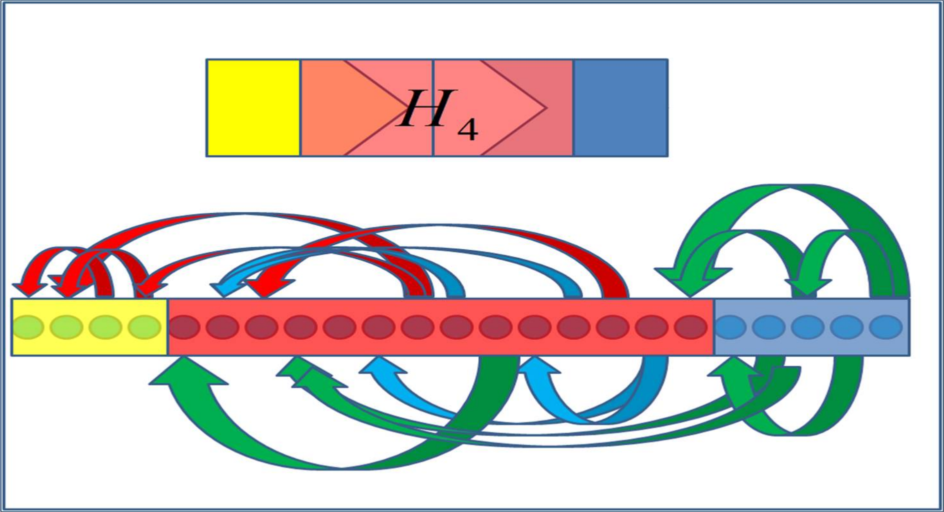
Combining tournaments...



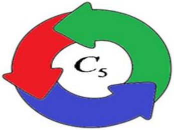
Combining tournaments...



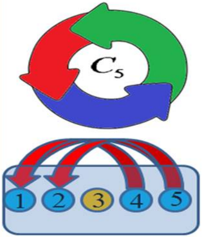
Combining tournaments...



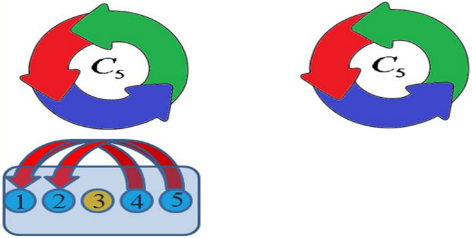
Combining tournaments...



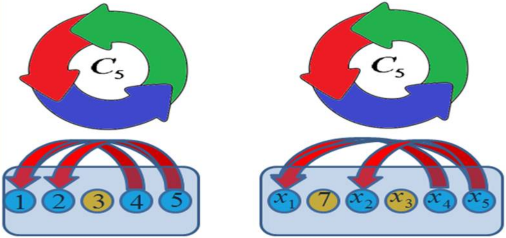
Combining tournaments...



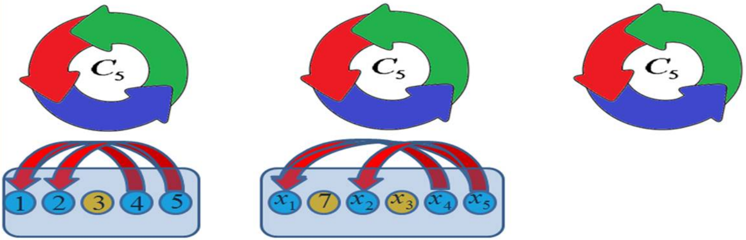
Combining tournaments...



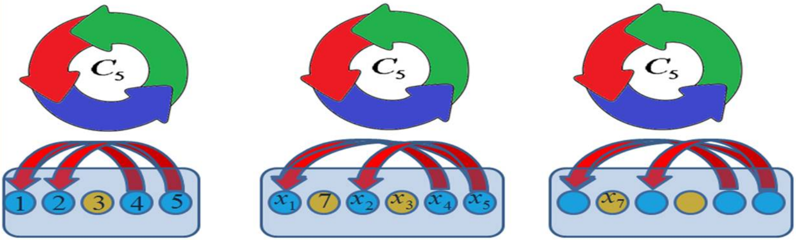
Combining tournaments...



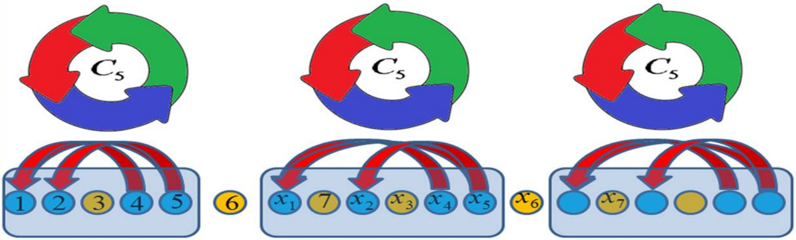
Combining tournaments...



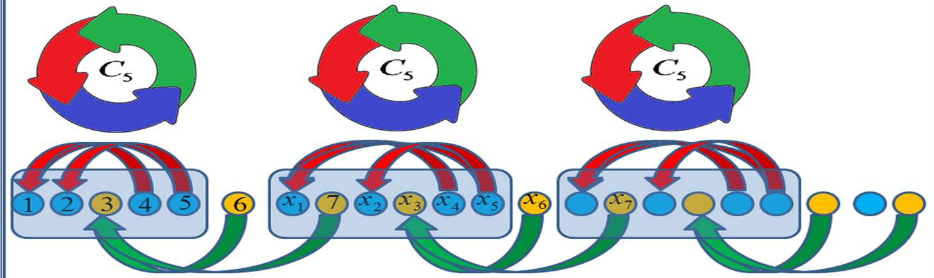
Combining tournaments...



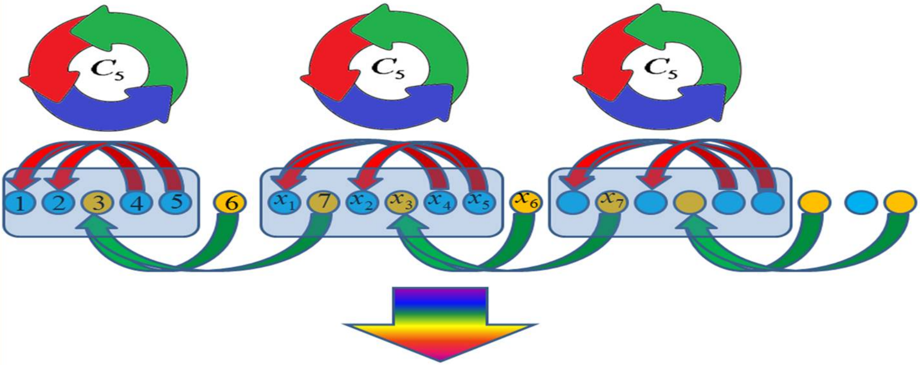
Combining tournaments...



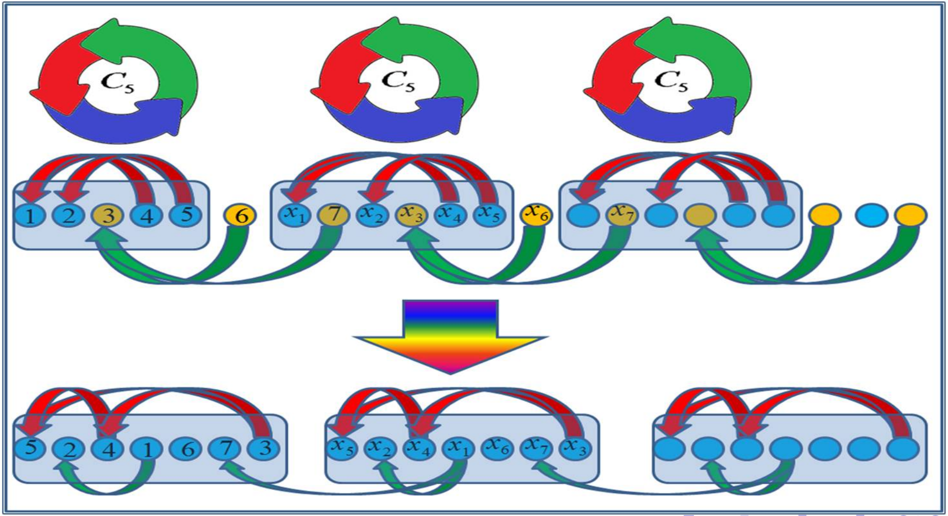
Combining tournaments...



Combining tournaments...



Combining tournaments...

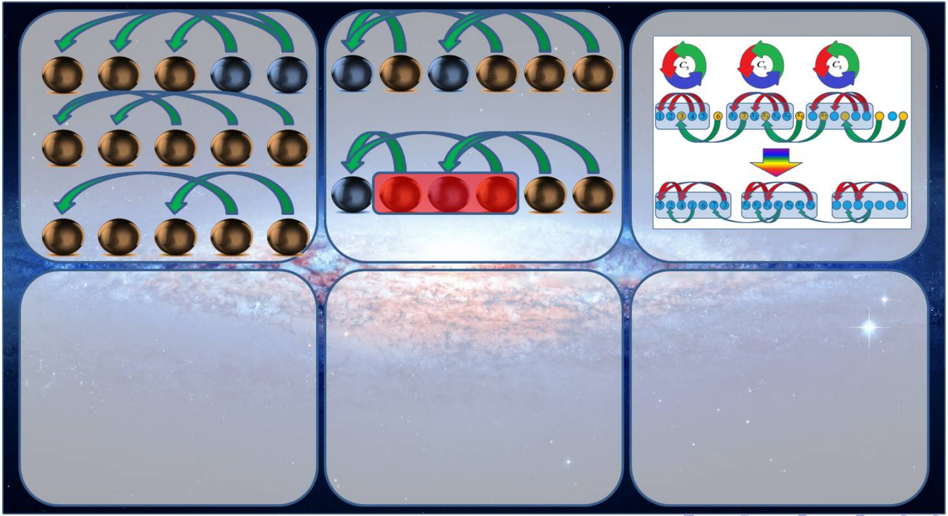


Combining tournaments...

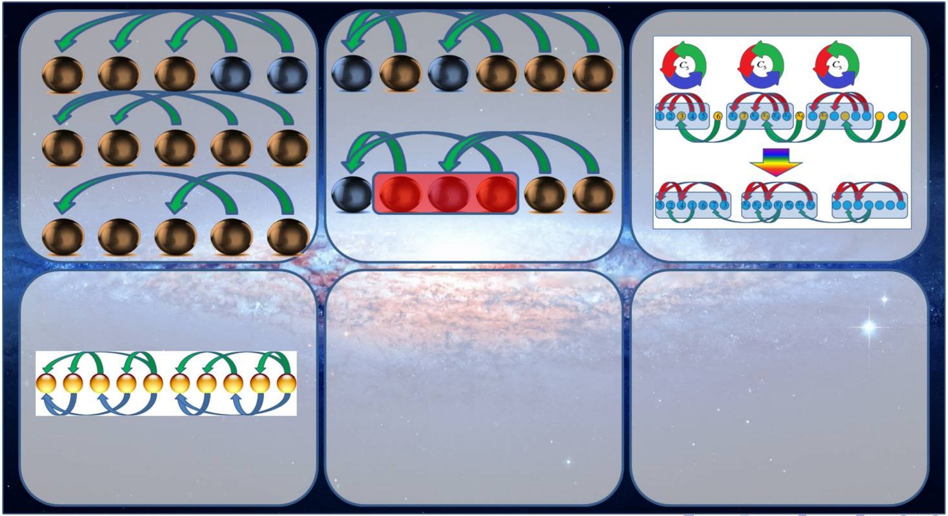
Theorem (Choromanski '13)

There exists a generic procedure for constructing larger prime tournaments satisfying the conjecture from smaller ones.

Galaxies, constellations, nebulae...



Galaxies, constellations, nebulae...



Nebulae...

Definition

Tournament is a **nebula** if it has an ordering of vertices under which the graph of backward edges is a collection of vertex disjoint stars.

Nebulae...

Definition

Tournament is a **nebula** if there exists its ordering of vertices under which the graph of backward edges is a collection of vertex disjoint stars.

Definition

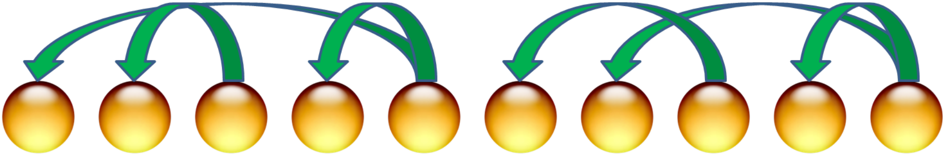
Tournament is a **left/right nebula** if it has an ordering of vertices under which the graph of backward edges is a collection of vertex disjoint left/right stars.

Nebulae...

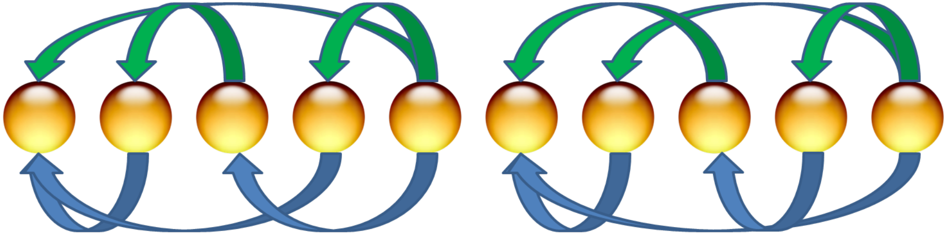
Conjecture

Let N_l be a left nebula and N_r be a right nebula. Then there exists $\epsilon(N_l, N_r) > 0$ such that every $\{N_l, N_r\}$-free n-vertex tournament contains a transitive subtournament of order at least $n^{\epsilon(N_l, N_r)}$.

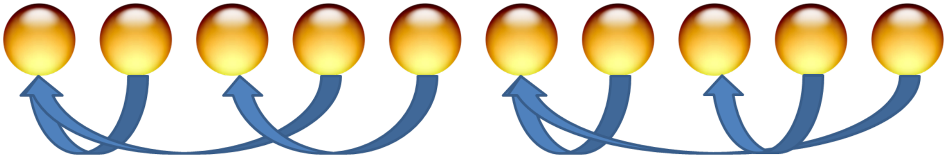
Nebulae of small stars



Nebulae of small stars



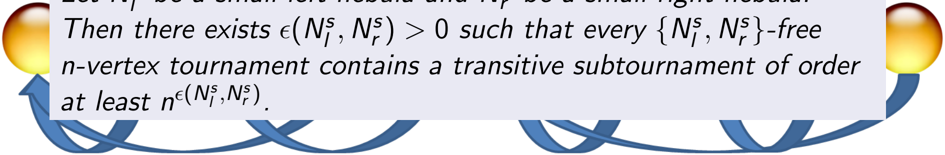
Nebulae of small stars



Nebulae of small stars

Theorem (Choromanski '14)

Let N^s_l be a small left nebula and N^s_r be a small right nebula. Then there exists $\epsilon(N^s_l, N^s_r) > 0$ such that every $\{N^s_l, N^s_r\}$-free n-vertex tournament contains a transitive subtournament of order at least $n^{\epsilon(N^s_l, N^s_r)}$.



Hardcore nebulae...

Definition

Let S be a left/right star. We call the set of vertices of S other than its first and last vertex a **core**.

Definition

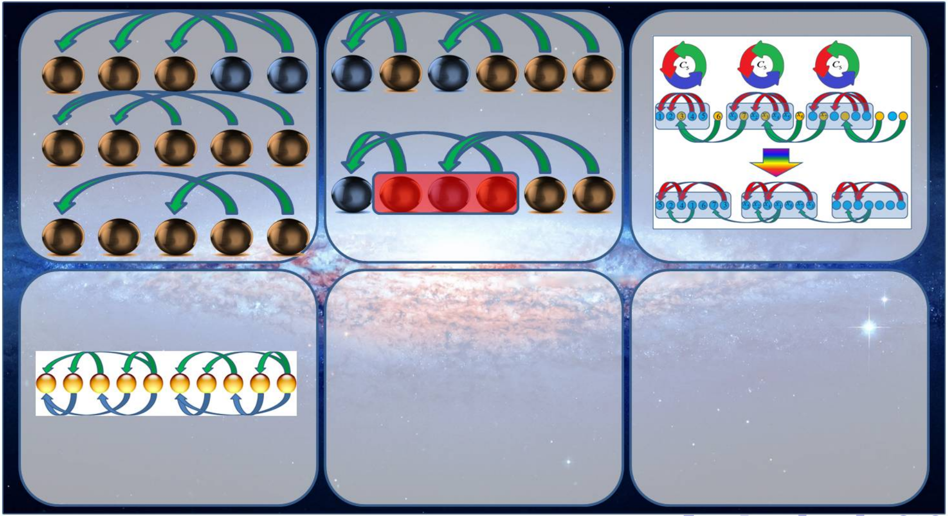
A tournament is called a **hardcore nebula** if it is a collection of vertex-disjoint left and right stars such that the only vertices between core vertices of any given star in the collection are core vertices.

Hardcore nebulae...

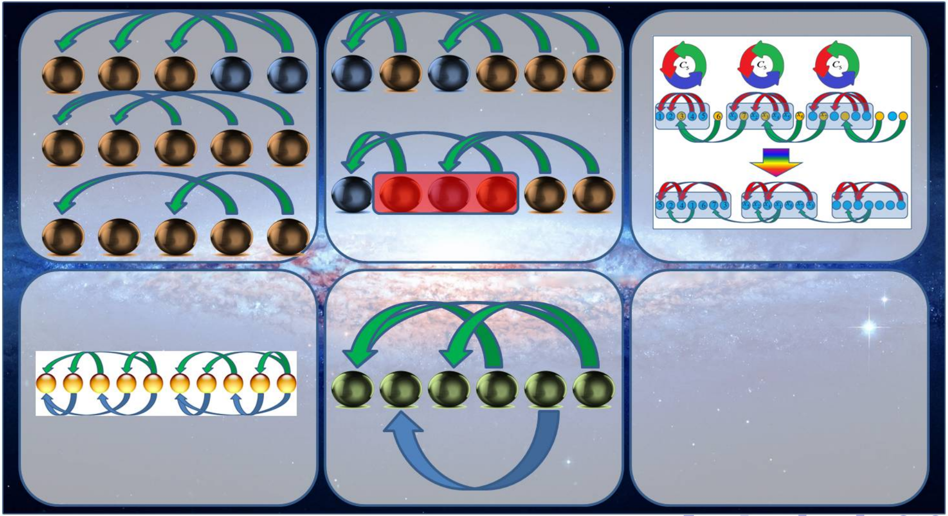
Theorem (Choromanski '14)

Let HN_l be a left hardcore nebula and HN_r be a right hardcore nebula. Then there exists $\epsilon(HN_l, HN_r) > 0$ such that every $\{HN_l, HN_r\}$-free n-vertex tournament contains a transitive subtournament of order at least $n^{\epsilon(HN_l, HN_r)}$.

Galaxies, constellations, nebulae...



Galaxies, constellations, nebulae...



On the Erdős-Hajnal Conjecture for six-vertex tournaments

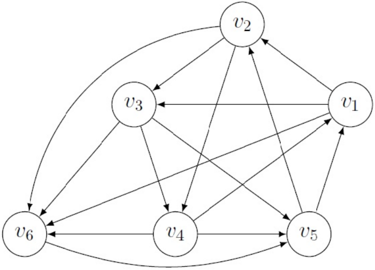


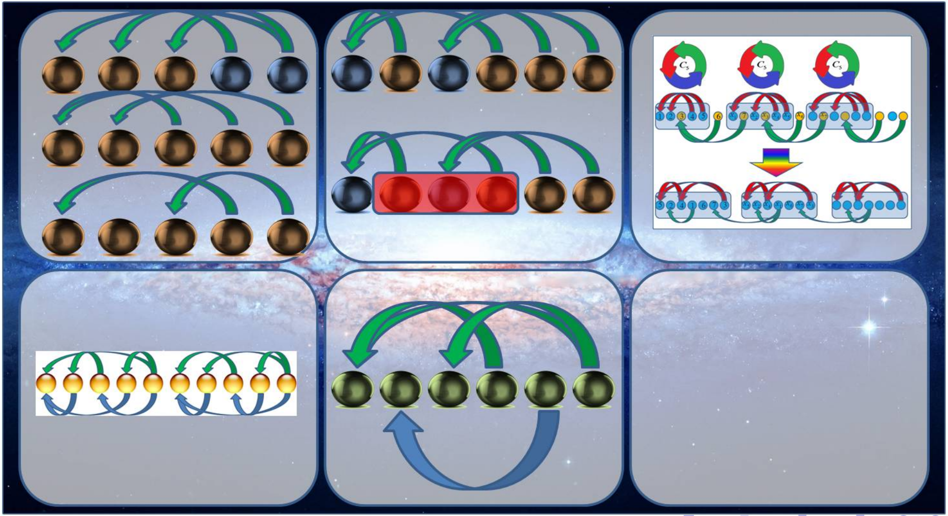
Figure: Tournament L_1 on the left and tournament L_2 on the right. Both are obtained from C_5 by adding one extra vertex.

On the Erdős-Hajnal Conjecture for six-vertex tournaments

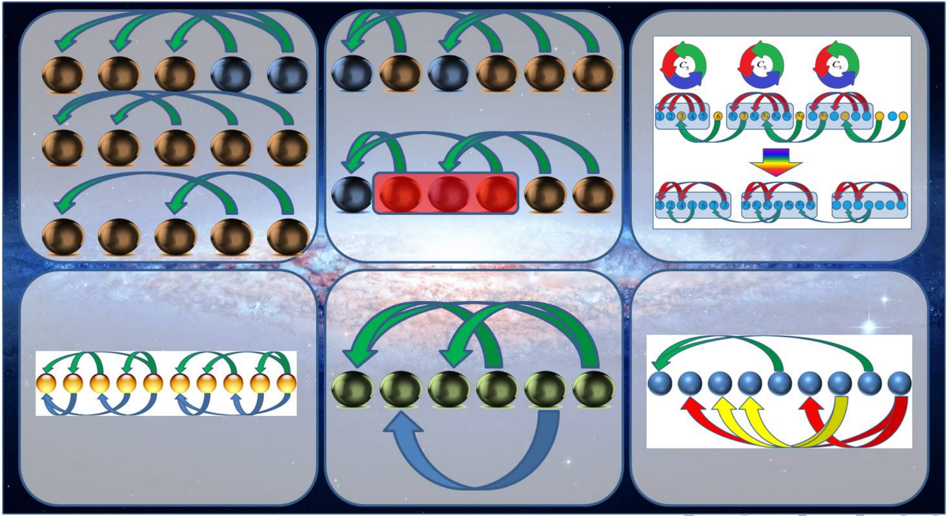
Theorem (Berger, Choromanski, Chudnovsky '15)

Every tournament on six vertices other than K_6 satisfies the Erdős-Hajnal Conjecture.

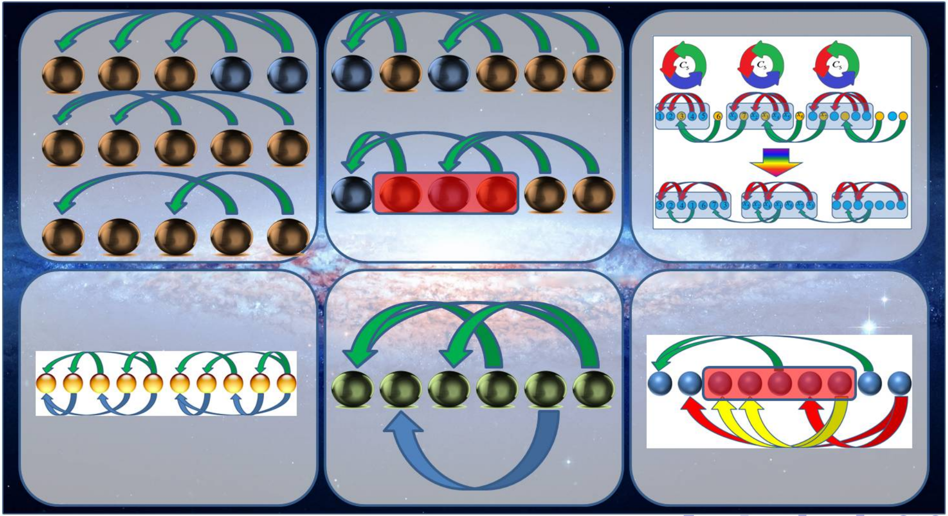
Galaxies, constellations, nebulae...



Galaxies, constellations, nebulae...



Galaxies, constellations, nebulae...



Ultraconstellations - main results

Theorem (Choromanski '15)

Let H be an ultraconstellation and let θ_H be its ultraconstellation ordering of vertices. Then there exists $\epsilon(H) > 0$ such that every $\{(H, \theta_H), (H, \theta_H^c)\}$-free ordered tournament (T, θ_T) contains a transitive subtournament of order at lest $|T|^{\epsilon(H)}$.

Ultraconstellations - main results

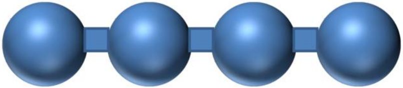
Corollary 1

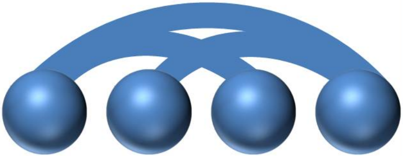
Gives the proof of the standard directed version of the Conjecture for the class of tournaments that contains as special cases all known infinite families of prime tournaments satisfying the Conjecture and defined by a single ordering.

Corollary 2

Implies all known results regarding excluding pairs of prime tournaments.

Undirected setting - excluding H and H^c

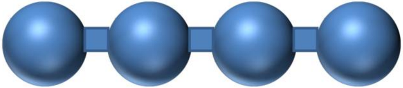


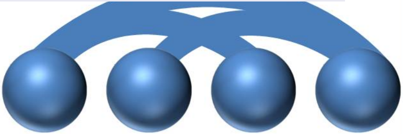


Undirected setting - excluding H and H^c

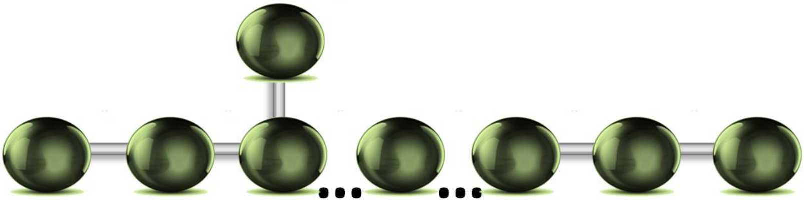
Theorem (Bousquet, Lagoutte, Thomasse '13)

Let $k, l > 0$. Define the class $\mathcal{H}_{k,l}$ of tournaments as those tournaments that are $\{P_k, P_l^c\}$-free, where P_k is a path of k vertices and P_l^c is an antipath of l vertices. Then $\mathcal{H}_{k,l}$ has polynomial-size transitive subtournaments.

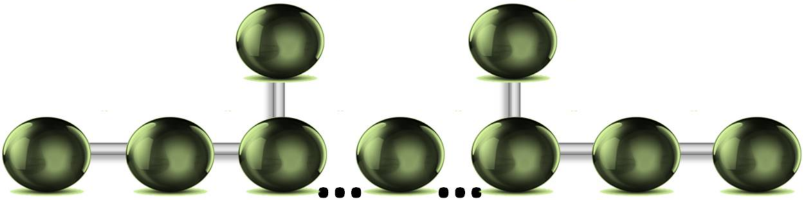




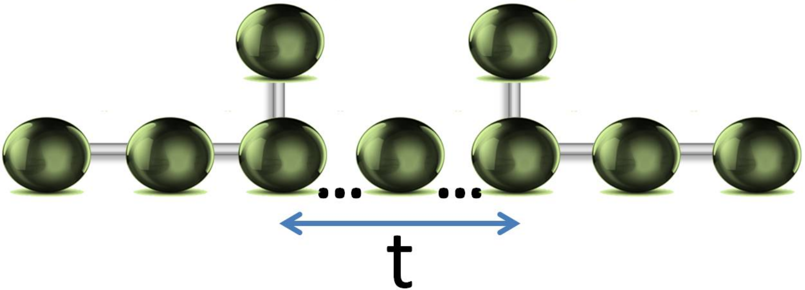
Hooks



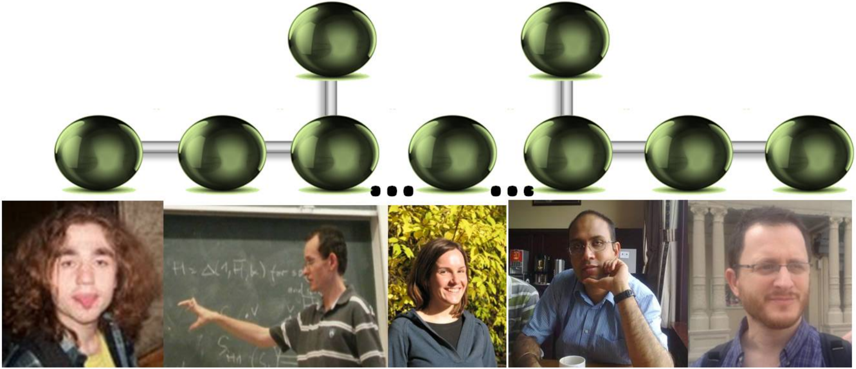
Excluding double-hooks



Excluding double-hooks



Excluding double-hooks



Excluding double-hooks

Theorem (Choromanski, Falik, Liebenau, Patel, Pilipczuk '15)

Let H_t be an double t-hook. Then for every m there exists $\epsilon(m)$ such that every $\{H_t, H_t^c : t = m, m+1, ...\}$-free undirected n-vertex graph G contains a clique or a stable set of size at least $n^{\epsilon(m)}$.

Excluding double-hooks

Theorem (Choromanski, Falik, Liebenau, Patel, Pilipczuk '15)

Let H_t be an double t-hook. Then for every m there exists $\epsilon(m)$ such that every $\{H_t, H_t^c : t = m, m+1, \ldots\}$-free undirected n-vertex graph G contains a clique or a stable set of size at least $n^{\epsilon(m)}$.

Corollary

For every hook H there exists $\epsilon(H) > 0$ such that every $\{H, H^c\}$-free undirected n-vertex graph G contains a clique or a stable set of size at least $n^{\epsilon(H)}$ (that extends the result of Bousquet, Lagoutte and Thomasse).

Excluding double-hooks

Theorem (Choromanski, Falik, Liebenau, Patel, Pilipczuk '15)

Let H_t be an double t-hook. Then for every m there exists $\epsilon(m)$ such that every $\{H_t, H_t^c : t = m, m+1, \ldots\}$-free undirected n-vertex graph G contains a clique or a stable set of size at least $n^{\epsilon(m)}$.

Corollary

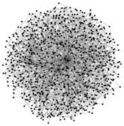
For every tree H on at most six vertices there exists $\epsilon(H) > 0$ such that every $\{H, H^c\}$-free undirected n-vertex graph G contains a clique or a stable set of size at least $n^{\epsilon(H)}$.

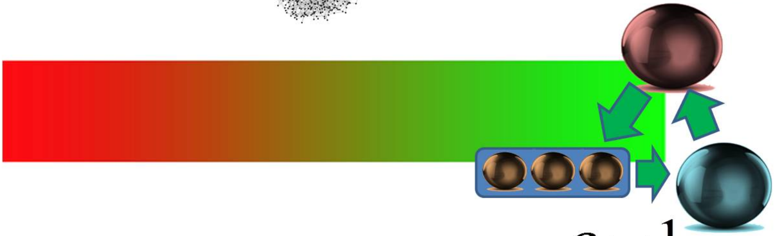
Asymptotics of the EH coefficients

Asymptotics of the EH coefficients

$\varepsilon = 1$

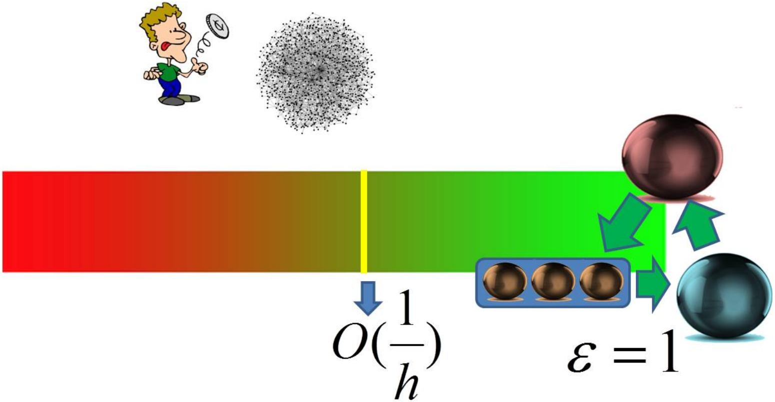
Asymptotics of the EH coefficients





$$\varepsilon = 1$$

Asymptotics of the EH coefficients



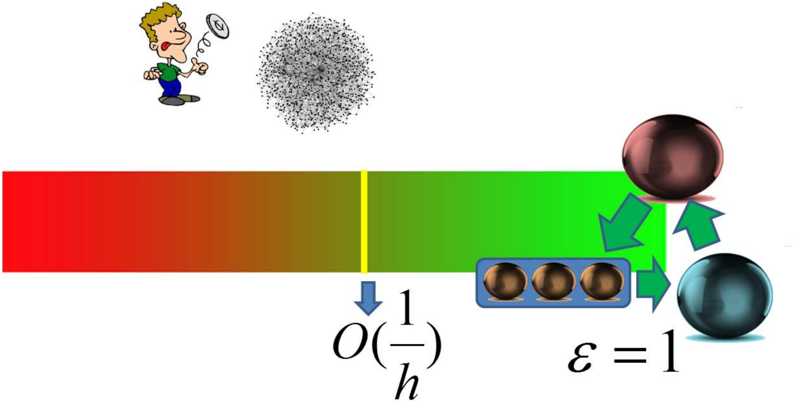
EH-coefficients of random tournaments

Theorem (Choromanski '10)

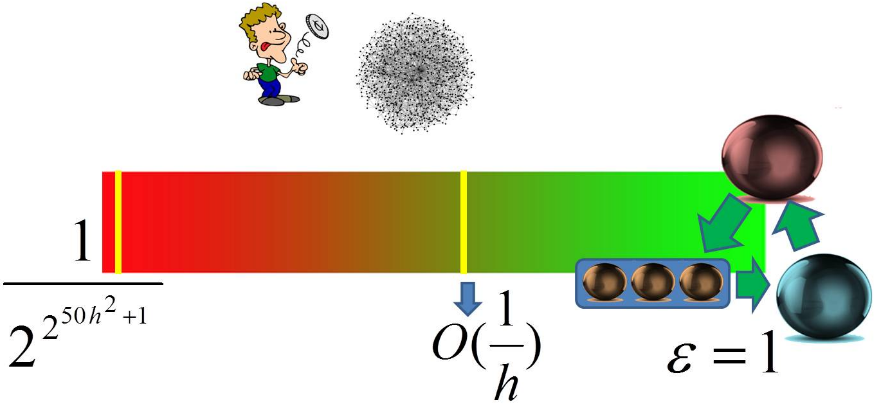
There exists $\eta > 0$ such that if we denote by $H^{n,\eta}$ the set of all n-vertex tournaments H with $\epsilon(H) \leq \frac{4}{|H|}(1 + \frac{\eta\sqrt{\log(|H|)}}{\sqrt{|H|}})$, and by H^n the set of all n-vertex tournaments then

$$\lim_{n \to \infty} \frac{|H^{n,\eta}|}{|H^n|} = 1.$$

Asymptotics of the EH coefficients



Asymptotics of the EH coefficients

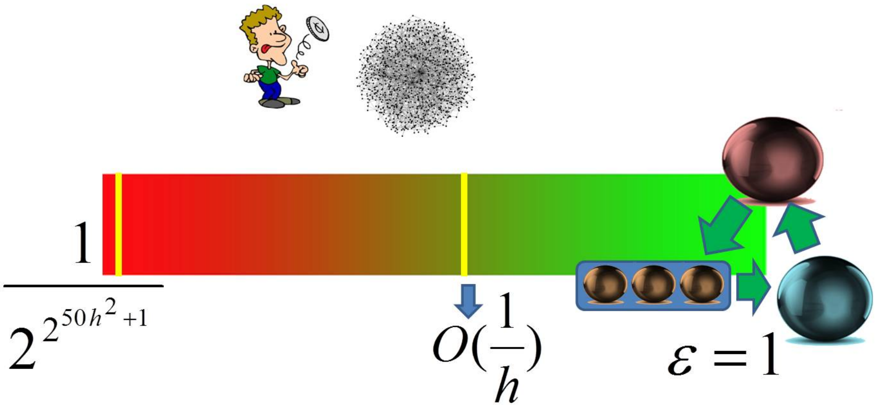


Freed from the Regularity Lemma

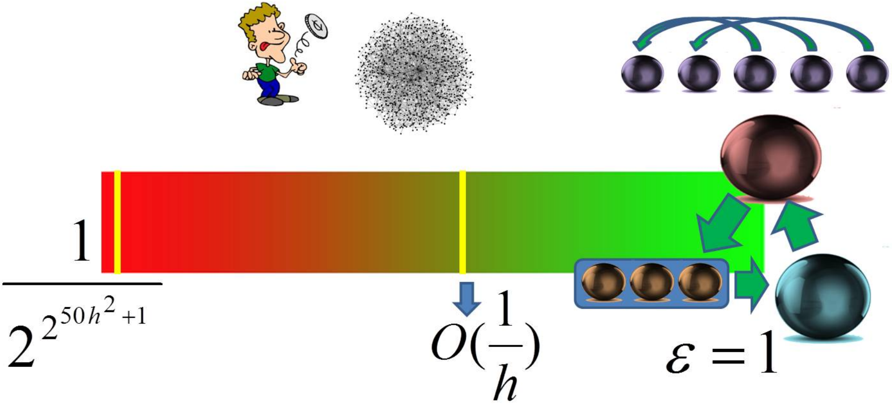
Theorem (Choromanski, Jebara '13)

Every known prime tournament H satisfying the Conjecture satisfies also: $\epsilon(H) \geq \frac{1}{2^{2^{50|H|^2+1}}}$.

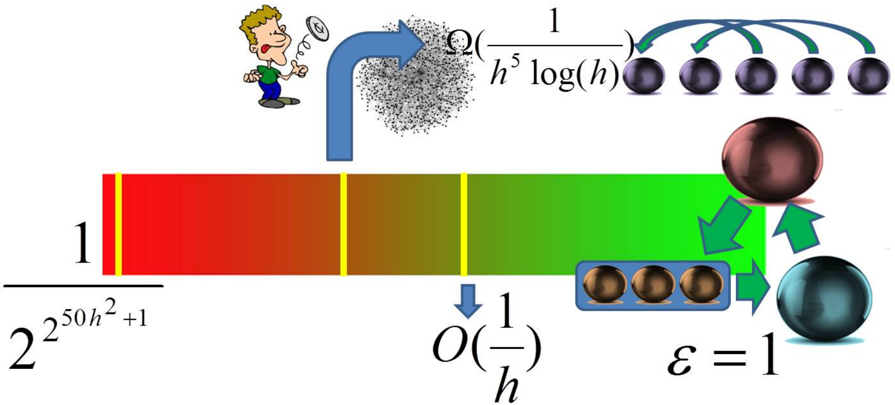
Asymptotics of the EH coefficients



Asymptotics of the EH coefficients



Asymptotics of the EH coefficients



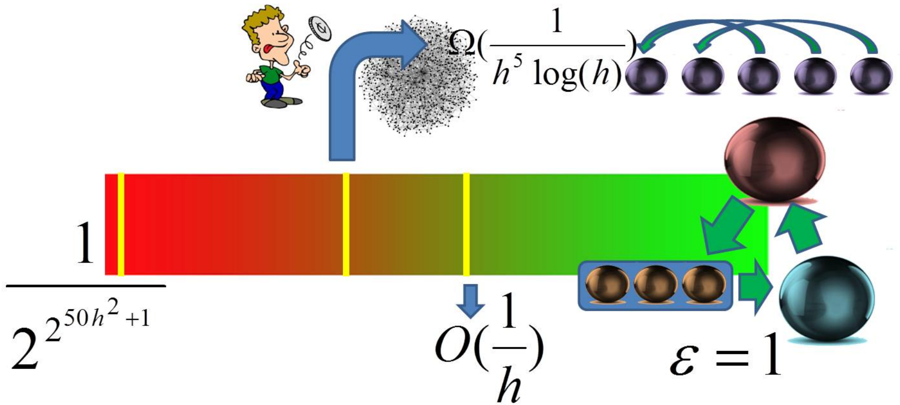
Polynomial EH-coefficients

Theorem (Choromanski '14)

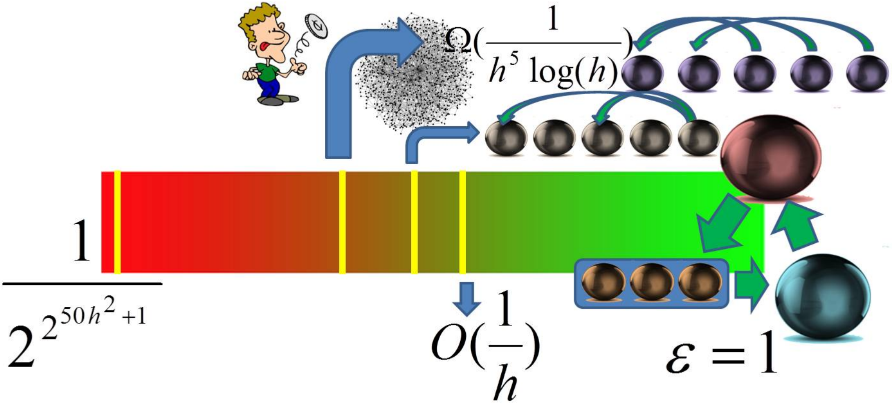
There exists $C > 0$ such that every known prime tournament H satisfying the Conjecture satisfies also:

$$\epsilon(H) \geq \frac{C}{|H|^5 \log(|H|)}.$$

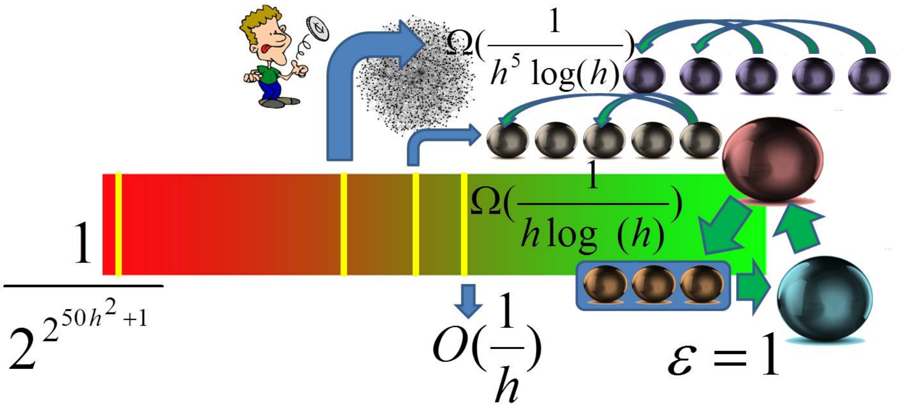
Asymptotics of the EH coefficients



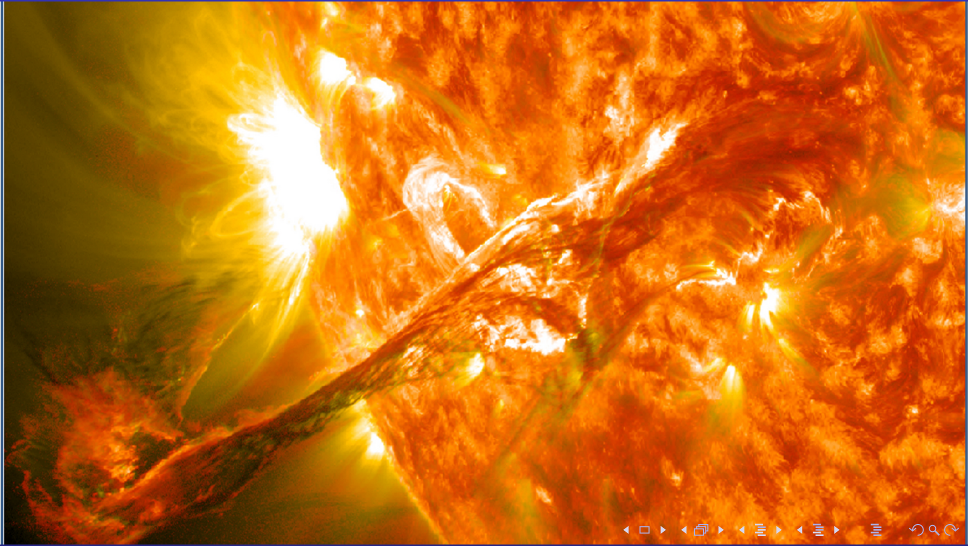
Asymptotics of the EH coefficients



Asymptotics of the EH coefficients



Tight bounds on EH coefficients for stars



Tight bounds on EH coefficients for stars

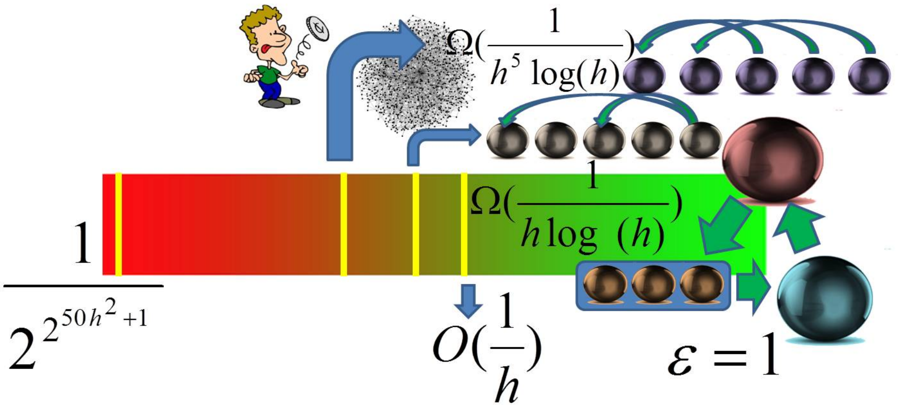
Theorem (Choromanski '12)

For every star H there exist $C_1, C_2 > 0$ such that

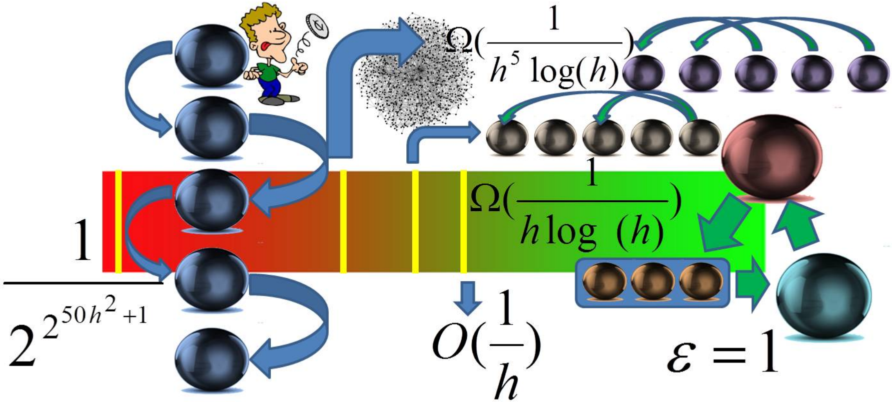
$$\frac{C_1}{h\log(h)} \leq \epsilon(H) \leq \frac{C_2\log(h)}{h},$$

where $h = |H|$.

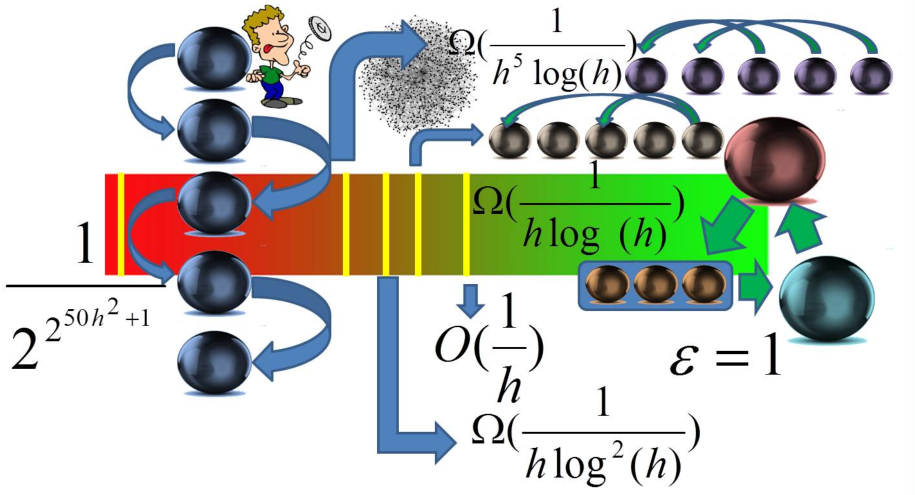
Asymptotics of the EH coefficients



Asymptotics of the EH coefficients



Asymptotics of the EH coefficients



Tight bounds on EH coefficients for directed paths

Theorem (Choromanski '15)

For every directed path P_h there exist $C_1, C_2 > 0$ such that

$$\frac{C_1}{h \log^2(h)} \leq \epsilon(H) \leq \frac{C_2 \log(h)}{h}.$$

Partition numbers and EH-coefficients

Theorem (Choromanski '12)

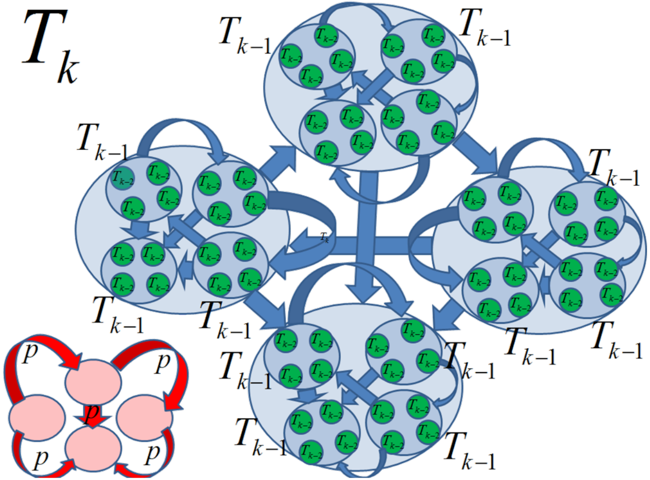
There exists $C_1 > 0$ such that for a tournament H the following holds:

$$\epsilon(H) \leq C_1 \frac{\log(\log(p(H)))}{\log(p(H))}.$$

There exists $C_2 > 0$ such that if H is prime then the following holds:

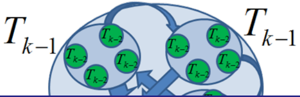
$$\epsilon(H) \leq C_2 \frac{\log(|H|)}{|H|}.$$

Powers of graphs



Small homogeneous sets imply small EH-coefficients

T_k

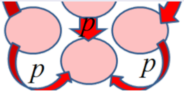


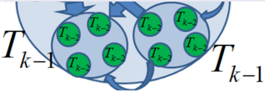
Theorem (Choromanski '12)

If H is a tournament without homogeneous sets of size larger than $\frac{\sqrt{|H|}}{2}$ then

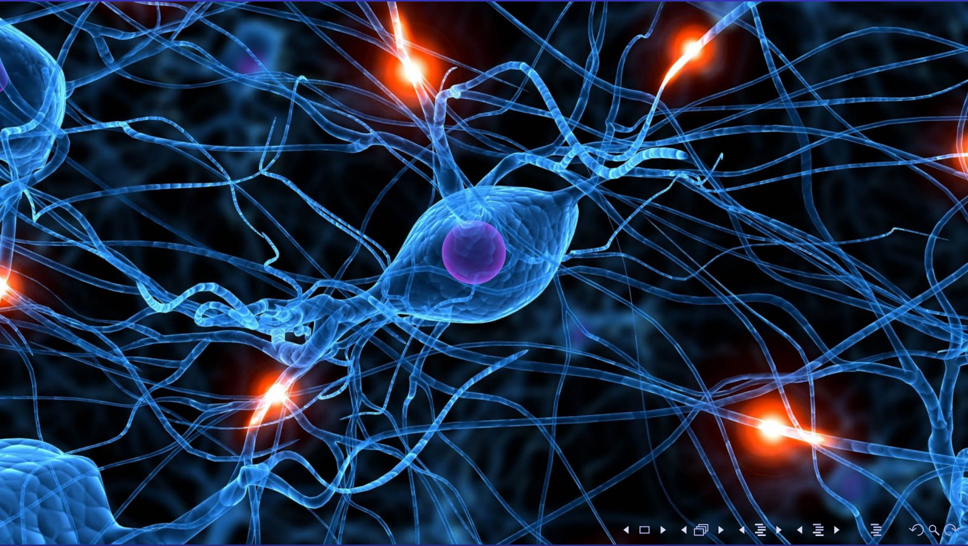
$$\limsup_{p(H)\to\infty} \frac{\epsilon(H)}{\frac{\log(p(H))}{p(H)^{\frac{1}{2}-\delta}}} < \infty,$$

for every $\delta > 0$.

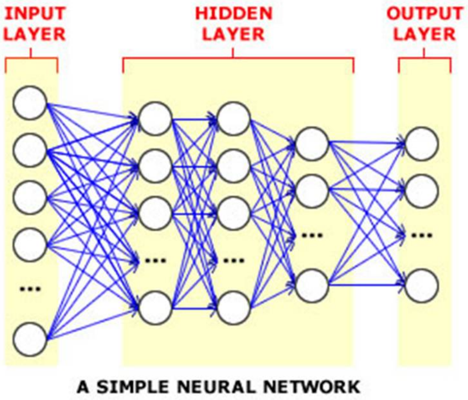




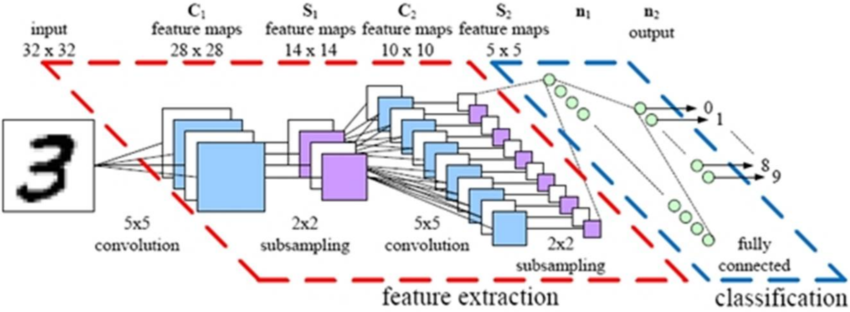
Neural networks - an overview



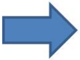
Neural networks - an overview



Neural networks - an overview



Neural networks - an overview



x

Neural networks - an overview

$$
x \rightarrow
\begin{bmatrix}
a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} \\
a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} & a_{19} \\
a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} & a_{29} \\
a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} & a_{39} \\
a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} & a_{49} \\
a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} & a_{59} \\
a_{60} & a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} & a_{69} \\
a_{70} & a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} & a_{78} & a_{79} \\
a_{80} & a_{81} & a_{82} & a_{83} & a_{84} & a_{85} & a_{86} & a_{87} & a_{88} & a_{89} \\
a_{90} & a_{91} & a_{92} & a_{93} & a_{94} & a_{95} & a_{96} & a_{97} & a_{98} & a_{99}
\end{bmatrix}
$$

Neural networks - an overview

$$x \rightarrow \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} & a_{19} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} & a_{29} \\ a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} & a_{39} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} & a_{49} \\ a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} & a_{59} \\ a_{60} & a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} & a_{69} \\ a_{70} & a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} & a_{78} & a_{79} \\ a_{80} & a_{81} & a_{82} & a_{83} & a_{84} & a_{85} & a_{86} & a_{87} & a_{88} & a_{89} \\ a_{90} & a_{91} & a_{92} & a_{93} & a_{94} & a_{95} & a_{96} & a_{97} & a_{98} & a_{99} \end{bmatrix} \rightarrow z$$

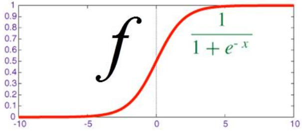
Neural networks - an overview

$$x \Rightarrow \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} & a_{19} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} & a_{29} \\ a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} & a_{39} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} & a_{49} \\ a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} & a_{59} \\ a_{60} & a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} & a_{69} \\ a_{70} & a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} & a_{78} & a_{79} \\ a_{80} & a_{81} & a_{82} & a_{83} & a_{84} & a_{85} & a_{86} & a_{87} & a_{88} & a_{89} \\ a_{90} & a_{91} & a_{92} & a_{93} & a_{94} & a_{95} & a_{96} & a_{97} & a_{98} & a_{99} \end{bmatrix} \Rightarrow z \quad \begin{pmatrix} f(z_0) \\ f(z_1) \\ f(z_2) \\ f(z_3) \\ f(z_4) \\ f(z_5) \\ f(z_6) \\ f(z_7) \\ f(z_8) \\ f(z_9) \end{pmatrix}$$

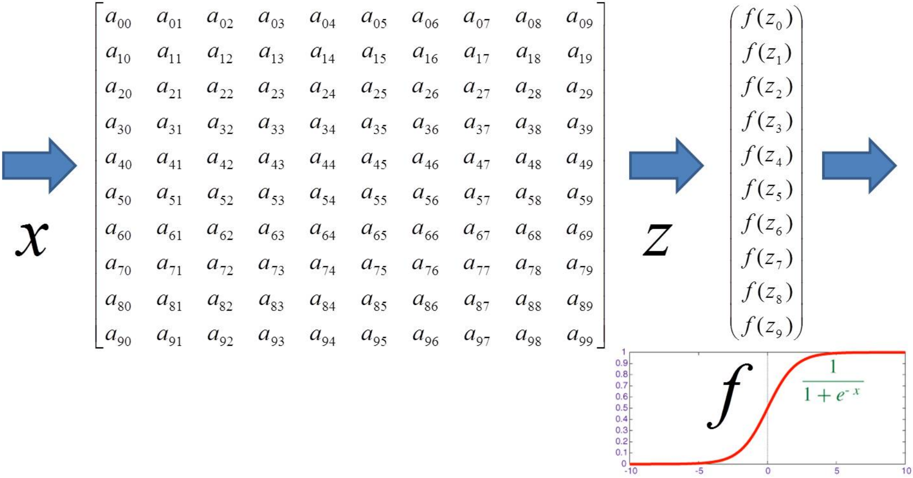
$$f$$

Neural networks - an overview

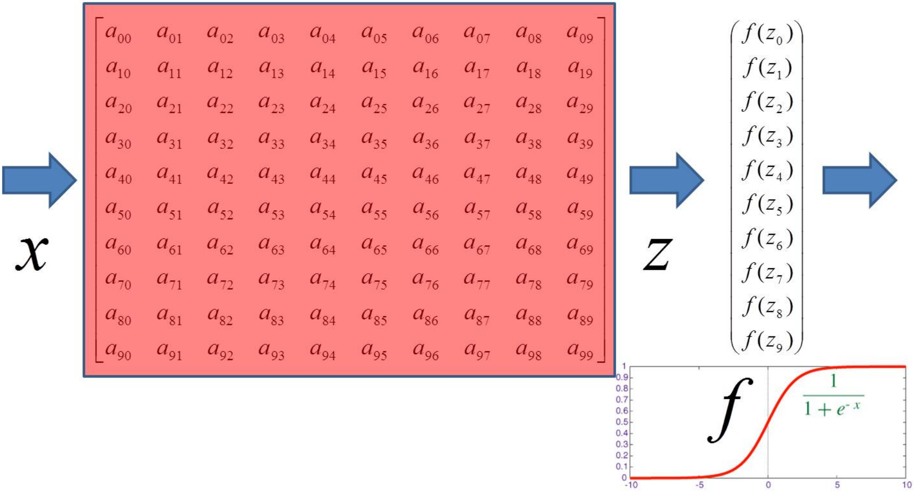
$$x \rightarrow \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} & a_{19} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} & a_{29} \\ a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} & a_{39} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} & a_{49} \\ a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} & a_{59} \\ a_{60} & a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} & a_{69} \\ a_{70} & a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} & a_{78} & a_{79} \\ a_{80} & a_{81} & a_{82} & a_{83} & a_{84} & a_{85} & a_{86} & a_{87} & a_{88} & a_{89} \\ a_{90} & a_{91} & a_{92} & a_{93} & a_{94} & a_{95} & a_{96} & a_{97} & a_{98} & a_{99} \end{bmatrix} \xrightarrow{z} \begin{pmatrix} f(z_0) \\ f(z_1) \\ f(z_2) \\ f(z_3) \\ f(z_4) \\ f(z_5) \\ f(z_6) \\ f(z_7) \\ f(z_8) \\ f(z_9) \end{pmatrix}$$



Neural networks - an overview

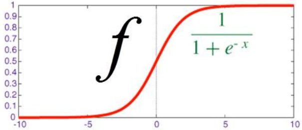


Introducing structured approach

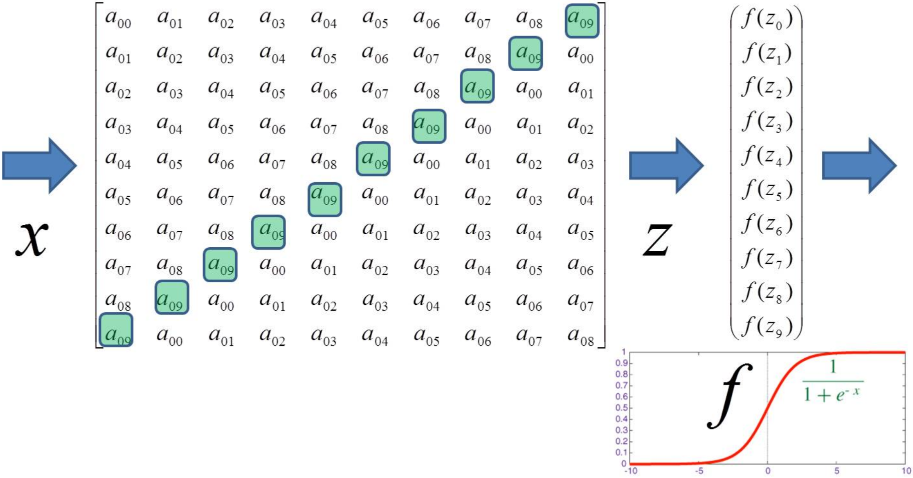


Introducing structured approach

$$x \rightarrow \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} \\ a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} \\ a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} \\ a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} \\ a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} \\ a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} \\ a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} \\ a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} \\ a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} \\ a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} \end{bmatrix} \rightarrow z \rightarrow \begin{pmatrix} f(z_0) \\ f(z_1) \\ f(z_2) \\ f(z_3) \\ f(z_4) \\ f(z_5) \\ f(z_6) \\ f(z_7) \\ f(z_8) \\ f(z_9) \end{pmatrix} \rightarrow$$

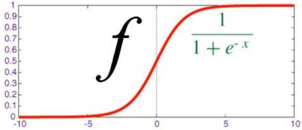


Introducing structured approach

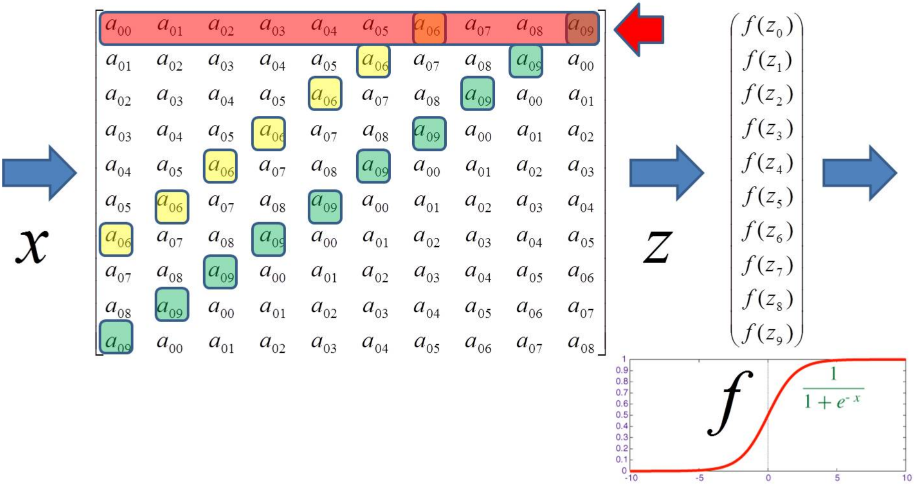


Introducing structured approach

$$x \Rightarrow \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} \\ a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} \\ a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} \\ a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} \\ a_{04} & a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} \\ a_{05} & a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} \\ a_{06} & a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} \\ a_{07} & a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} \\ a_{08} & a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} \\ a_{09} & a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} & a_{06} & a_{07} & a_{08} \end{bmatrix} \Rightarrow z \Rightarrow \begin{pmatrix} f(z_0) \\ f(z_1) \\ f(z_2) \\ f(z_3) \\ f(z_4) \\ f(z_5) \\ f(z_6) \\ f(z_7) \\ f(z_8) \\ f(z_9) \end{pmatrix} \Rightarrow$$



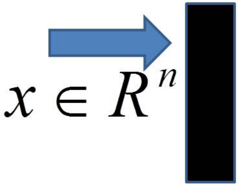
Introducing structured approach



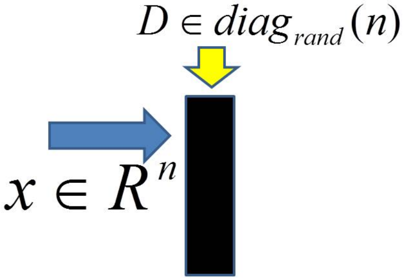
Structured nonlinear hashing with PHDs

$x \in R^n$

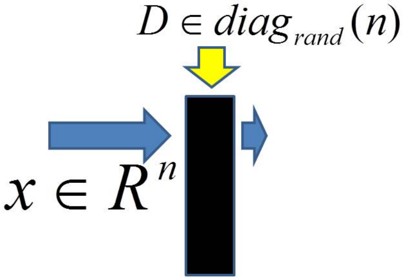
Structured nonlinear hashing with PHDs



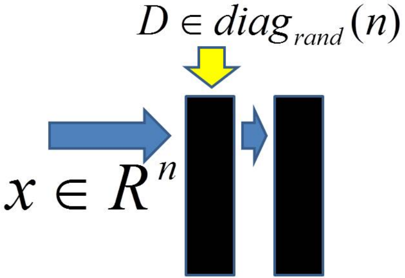
Structured nonlinear hashing with PHDs



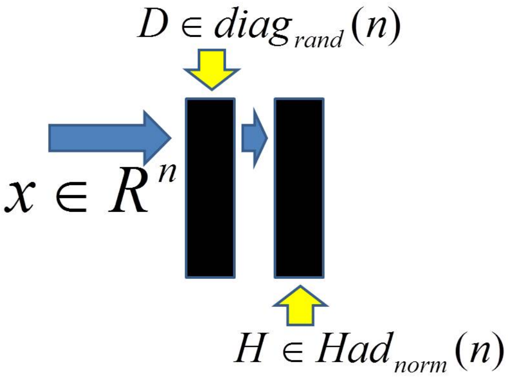
Structured nonlinear hashing with PHDs



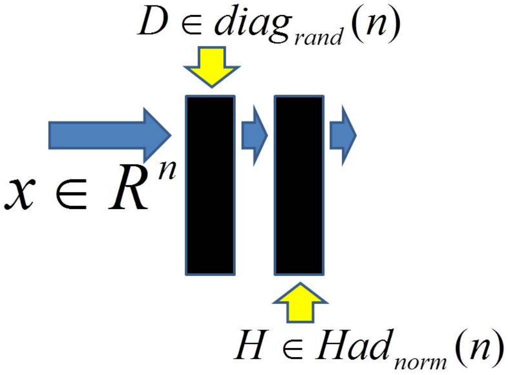
Structured nonlinear hashing with PHDs



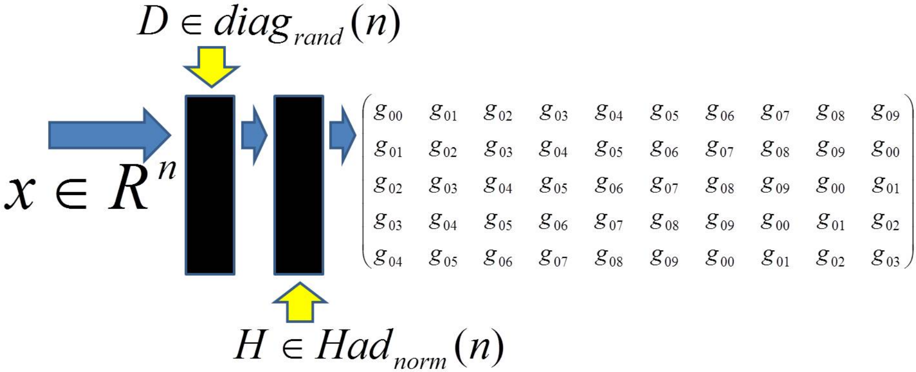
Structured nonlinear hashing with PHDs



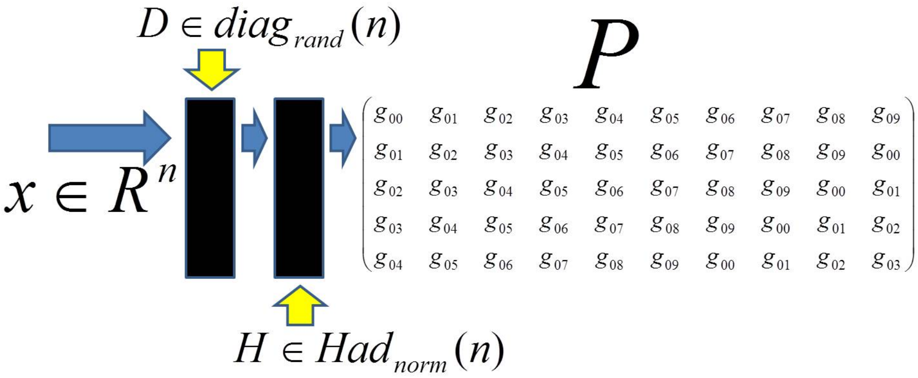
Structured nonlinear hashing with PHDs



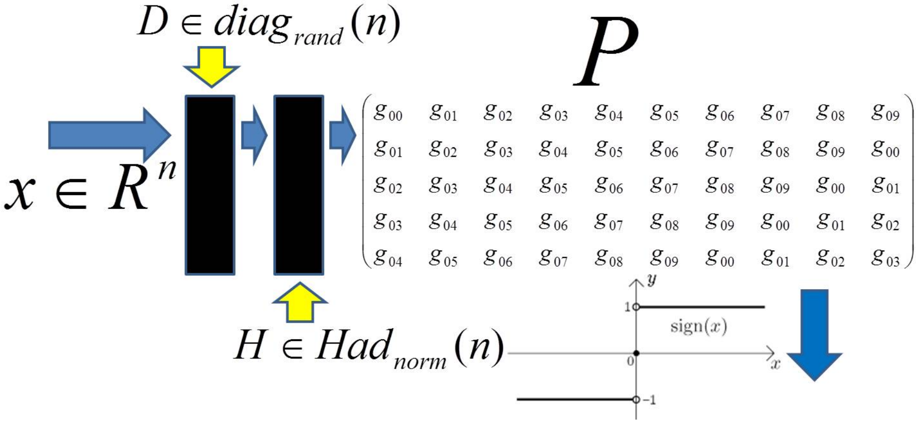
Structured nonlinear hashing with PHDs



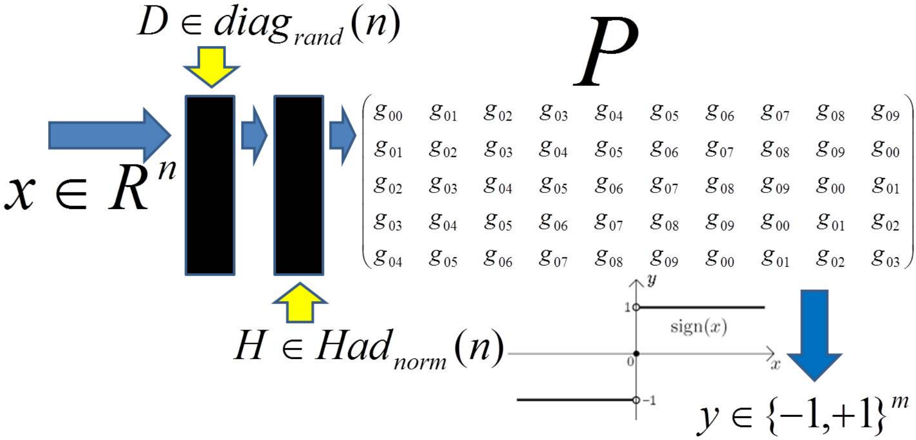
Structured nonlinear hashing with PHDs



Structured nonlinear hashing with PHDs



Structured nonlinear hashing with PHDs



Ψ-regular structured gaussian matrices

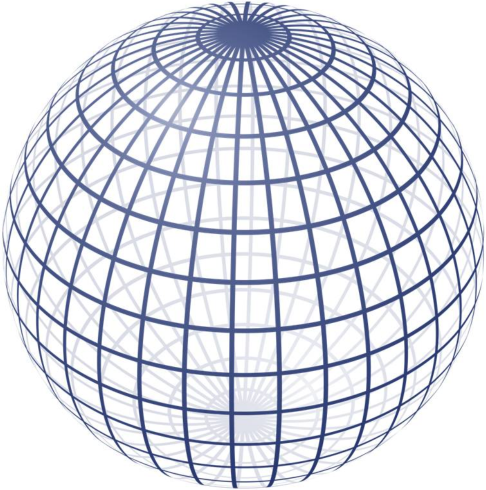
Matrix $\mathcal{P}$ is Ψ-regular random matrix if it has the following form

$$\begin{pmatrix} \sum_{l \in \mathcal{S}_{1,1}} g_l & \cdots & \sum_{l \in \mathcal{S}_{1,j}} g_l & \cdots & \sum_{l \in \mathcal{S}_{1,n}} g_l \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{l \in \mathcal{S}_{i,1}} g_l & \cdots & \sum_{l \in \mathcal{S}_{i,j}} g_l & \cdots & \sum_{l \in \mathcal{S}_{i,n}} g_l \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{l \in \mathcal{S}_{k,1}} g_l & \cdots & \sum_{l \in \mathcal{S}_{k,j}} g_l & \cdots & \sum_{l \in \mathcal{S}_{k,n}} g_l \end{pmatrix}$$

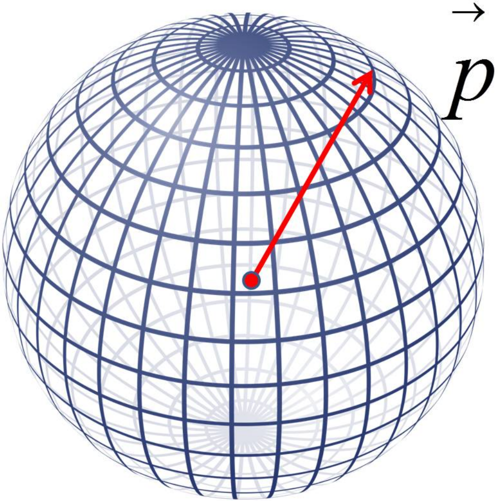
where $S_{i,j} \subseteq \{1,...,t\}$, $|S_{i,1}| = ... = |S_{i,n}|$, $S_{i,j} \cap S_{i,u} = \emptyset$ for $j \neq u$, and furthermore:

- for a fixed column $\mathcal{C}$ of $\mathcal{P}$ and fixed $l \in \{1,...,t\}$ random variable g_l appears in at most $\Psi + 1$ entries from $\mathcal{C}$.

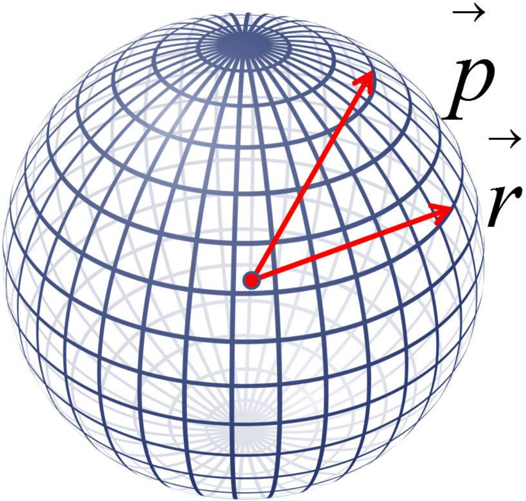
Preserving angles with structured gaussian matrices



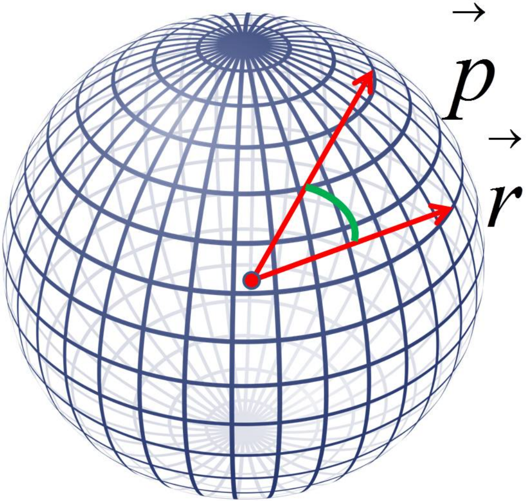
Preserving angles with structured gaussian matrices



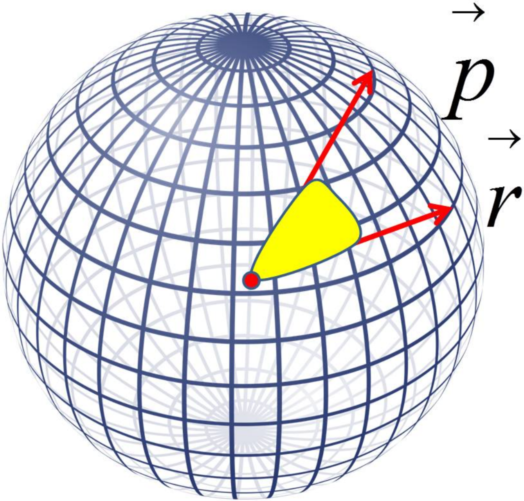
Preserving angles with structured gaussian matrices



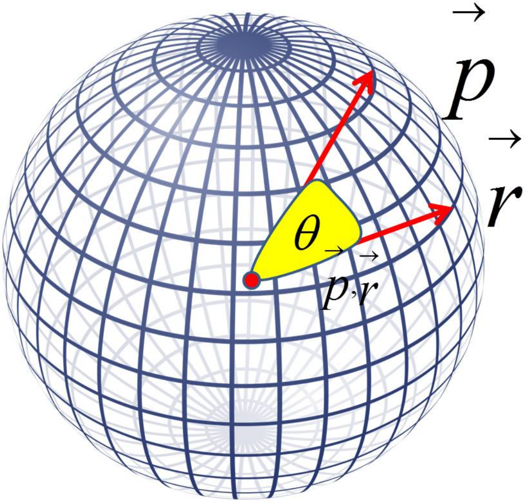
Preserving angles with structured gaussian matrices



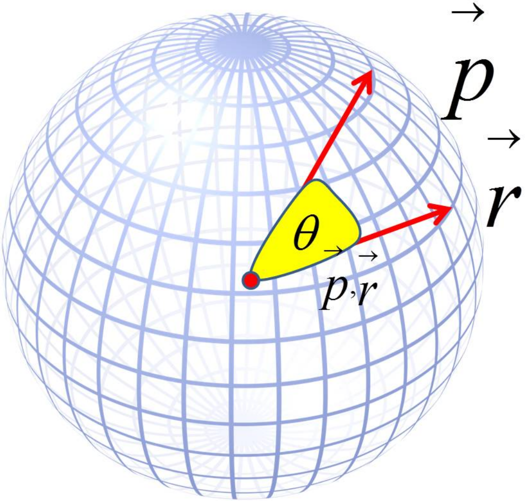
Preserving angles with structured gaussian matrices



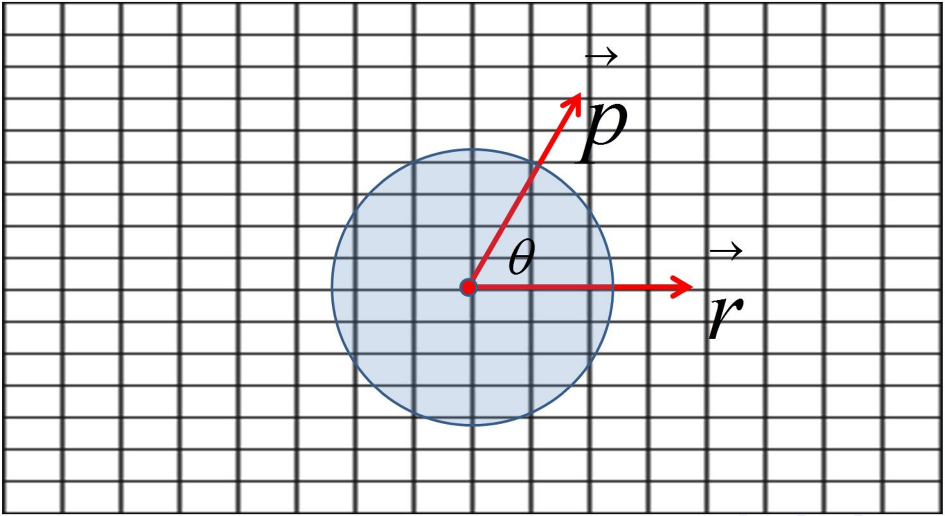
Preserving angles with structured gaussian matrices



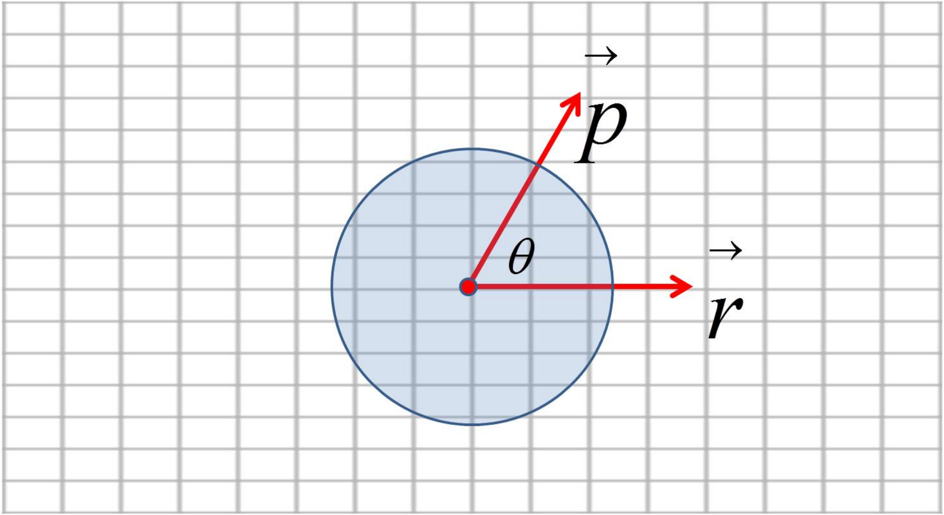
Preserving angles with structured gaussian matrices



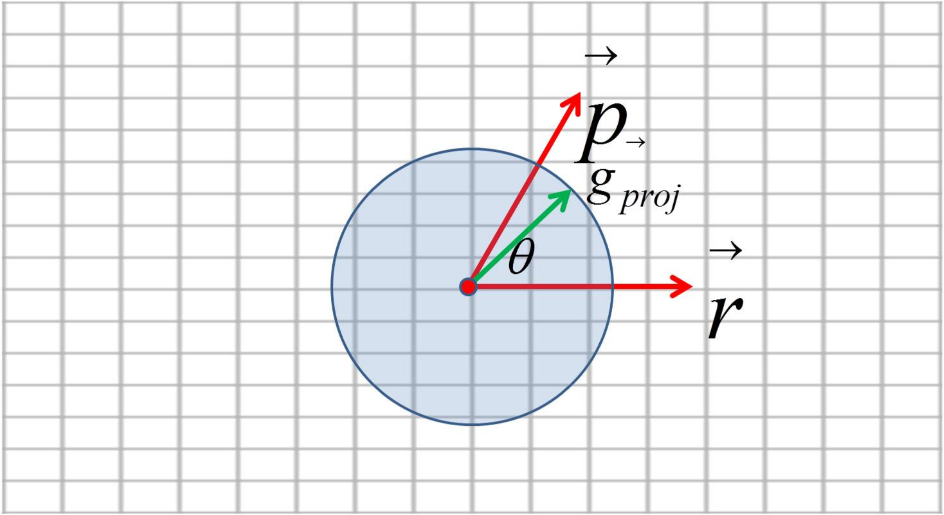
Preserving angles with structured gaussian matrices



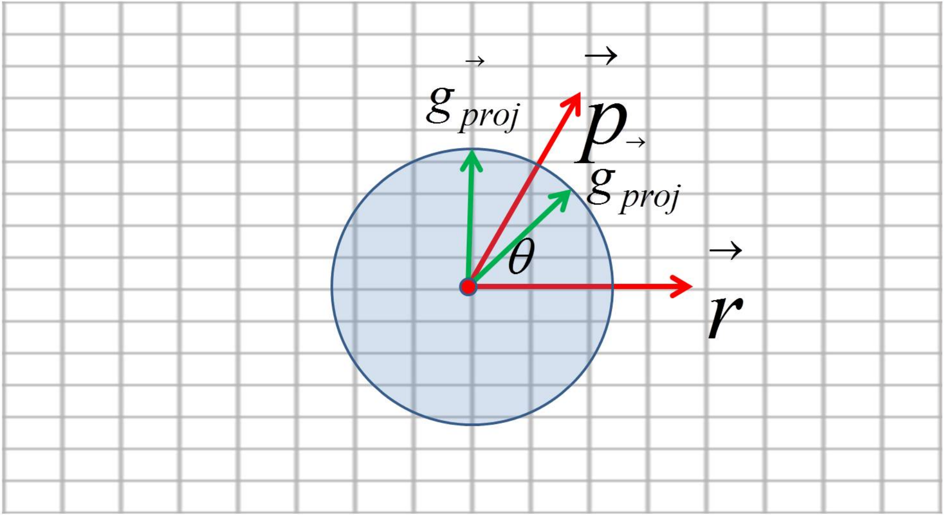
Preserving angles with structured gaussian matrices



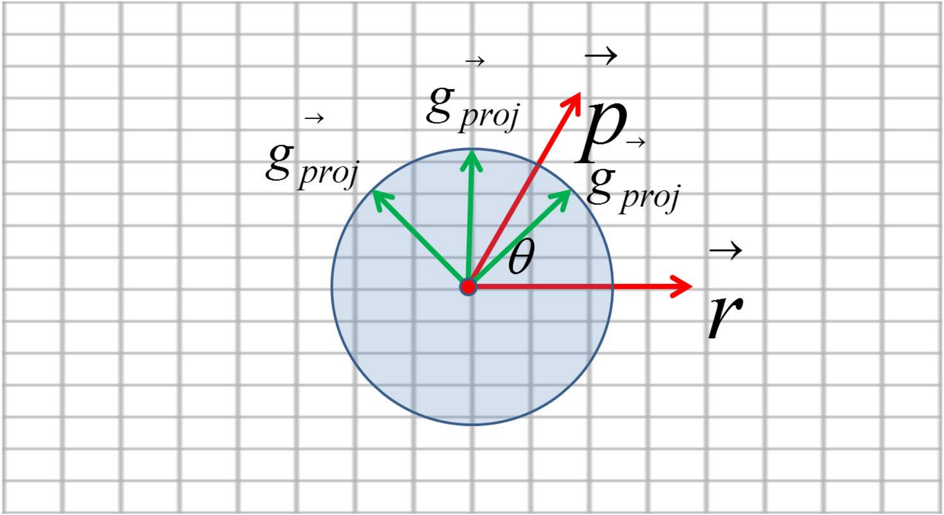
Preserving angles with structured gaussian matrices



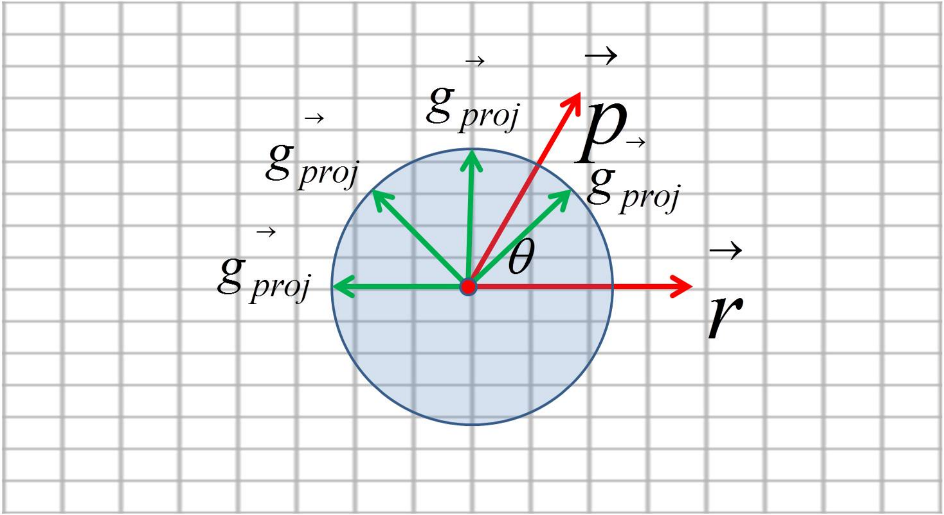
Preserving angles with structured gaussian matrices



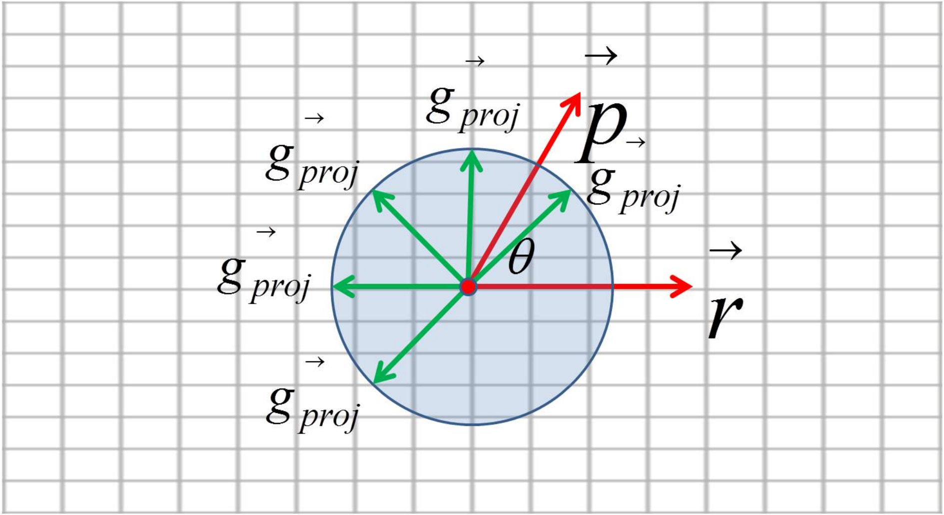
Preserving angles with structured gaussian matrices



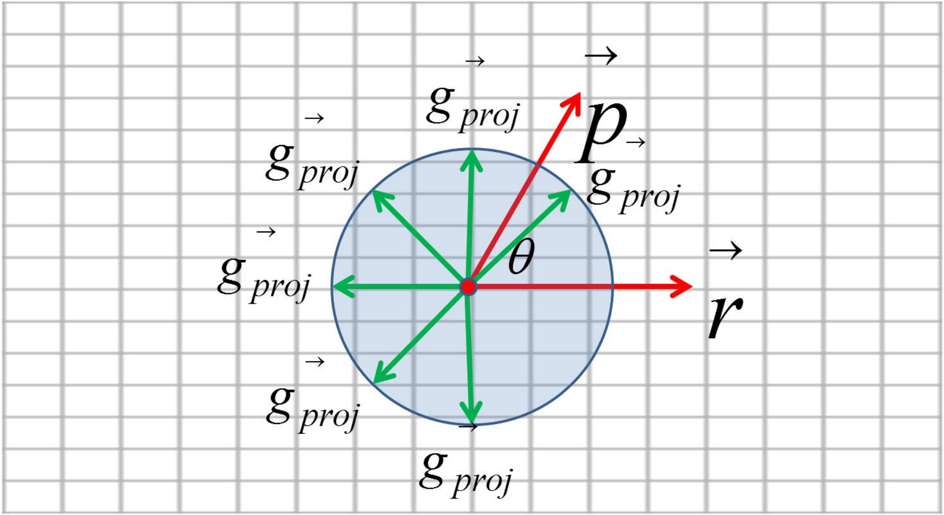
Preserving angles with structured gaussian matrices



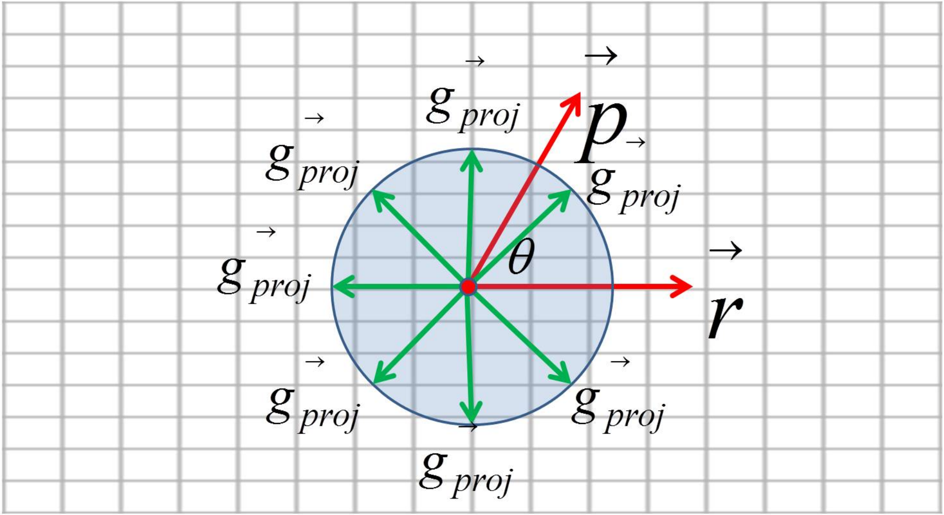
Preserving angles with structured gaussian matrices



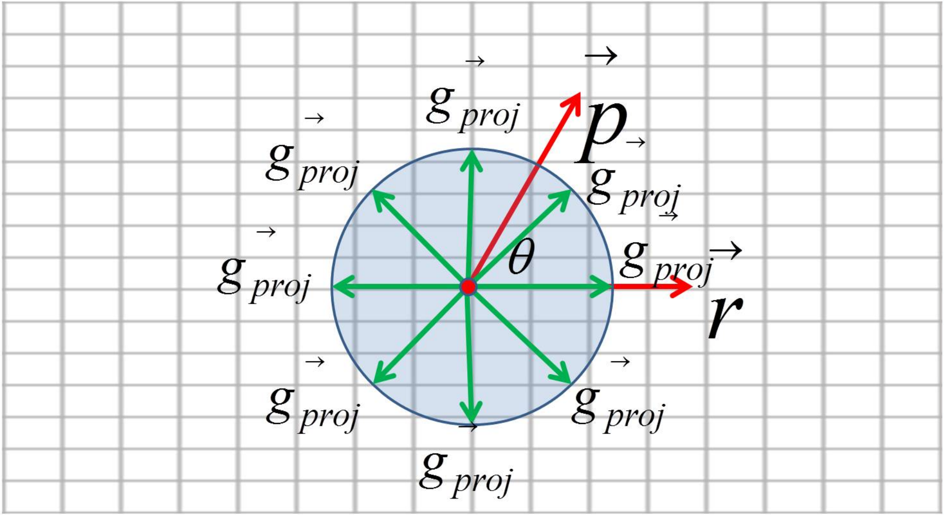
Preserving angles with structured gaussian matrices



Preserving angles with structured gaussian matrices



Preserving angles with structured gaussian matrices



Coloring graphs of structured matrices

Let us fix two rows of $\mathcal{P}$ of indices $1 \leq k_1 < k_2 \leq k$ respectively. We define a graph $\mathcal{G}_{\mathcal{P}}(k_1, k_2)$ as follows:

- $V(\mathcal{G}_{\mathcal{P}}(k_1, k_2)) = \{\{j_1, j_2\} : \exists l \in \{1, ..., t\} s.t. g_l \in \mathcal{S}_{k_1,j_1} \cap \mathcal{S}_{k_2,j_2}, j_1 \neq j_2\}$,
- there exists an edge between vertices $\{j_1, j_2\}$ and $\{j_3, j_4\}$ iff $\{j_1, j_2\} \cap \{j_3, j_4\} \neq \emptyset$.

Definition

Let $\mathcal{P}$ be a Ψ-regular matrix. We define the $\mathcal{P}$-chromatic number $\chi(\mathcal{P})$ as:

$$\chi(\mathcal{P}) = \max_{1 \leq k_1 < k_2 \leq k} \chi(\mathcal{G}(k_1, k_2)).$$

Coloring graphs of structured matrices

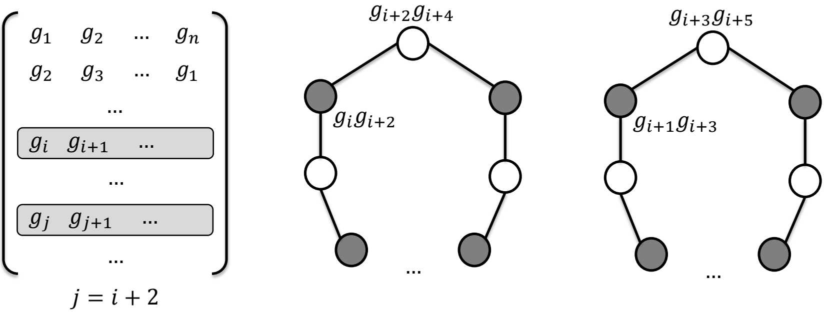


Figure: Structured graph for the circulant matrix - a set of disjoint cycles.

Coloring and concentration results

Theorem (Choromanski '15)

Take the extended Ψ-regular hashing model $\mathcal{M}$. Let N be the size of the dataset. Denote by k the size of the hash and by n the dimensionality of the data. Let $f(n)$ be an arbitrary positive function. Then for every $a, \epsilon > 0$ the following is true:

$$\mathbb{P}\left(\left|\tilde{\theta}^n_{p,r} - \frac{\theta_{p,r}}{\pi}\right| \leq \epsilon\right) \geq \left[1 - 4\binom{N}{2} e^{-\frac{f^2(n)}{2}} - 4\chi(\mathcal{P})\binom{k}{2} e^{-\frac{2a^2 t}{f^4(t)}}\right] \Lambda,$$

where $\Lambda = 1 - \frac{1}{\pi}\sum_{j=\frac{\epsilon k}{2}}^{k} \frac{1}{\sqrt{j}}(\frac{ke}{j})^j \mu^j (1-\mu)^{k-j} + 2e^{-\frac{\epsilon^2 k}{2}}$ *and*

$\mu = \frac{8k(a\chi(\mathcal{P}) + \Psi\frac{f^2(n)}{n})}{\theta_{p,r}}$.

Coloring and concentration results

Corollary

Take the extended Ψ-regular hashing model $\mathcal{M}$. Assume that the projection matrix $\mathcal{P}$ is Toeplitz gaussian. Let N be the size of the dataset. Denote by k the size of the hash and by n the dimensionality of the data. Then for every $\epsilon > 0$ the following is true:

$$\mathbb{P}\left(\left|\tilde{\theta}^{n}_{p,r} - \frac{\theta_{p,r}}{\pi}\right| \leq k^{-\frac{1}{3}}\right) \geq \left[1 - O\left(\frac{N^2}{n^{4.5}} + k^2 e^{-\Omega\left(\frac{n^{\frac{1}{3}}}{\log^2(n)}\right)}\right)\right]\Lambda,$$

where $\Lambda = \left[1 - \left(\frac{k^7}{n}\right)^{\frac{1}{3}}\right]$.

Coloring and concentration results

Theorem (Choromanski '15)

Take the short Ψ*-regular hashing model* $\mathcal{M}$*, where* $\mathcal{P}$ *is a Toeplitz gaussian matrix. Denote by* k *the size of the hash. Then the following is true*

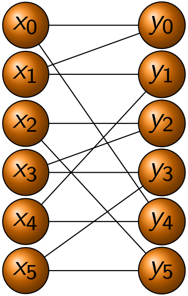
$$Var(\tilde{\theta}^{n}_{p,r}) \leq \frac{1}{k}\frac{\theta_{p,r}(\pi - \theta_{p,r})}{\pi^2} + (\frac{\log(k)}{k^2})^{\frac{1}{3}},$$

and thus for any $c > 0$*:*

$$\mathbb{P}\left(\left|\tilde{\theta}^{n}_{p,r} - \frac{\theta_{p,r}}{\pi}\right| \geq c\left(\frac{\sqrt{\log(k)}}{k}\right)^{\frac{1}{3}}\right) = O\left(\frac{1}{c^2}\right).$$

Adaptive anonymity with b-matchings

username				
Anna	1	0	0	0
Robert	0	0	0	0
Paul	0	0	1	1
Peter	1	0	1	1
Carl	1	1	0	0
Olaf	0	1	1	1



				key
*	0	0	0	172
*	*	0	0	236
*	0	1	1	672
*	*	1	1	229
1	*	0	0	761
0	*	1	1	298

Figure: The b-matching k-anonymity. The comparability graph is not a disjoint union of complete bipartite graphs. The parameters of the model are: $n = 6$, $f = 4$, $k = 2$. Presented solution achieves $\#(*) = 8$. The standard k-anonymity would achieve $\#(*) = 10$.

Combinatorics of adaptive anonymity via b-matching

Definition

Let $G(A, B)$ be a bipartite graph with color classes: A, B, where $|A| = |B| = n$. For a vertex $v \in V(G(A, B))$ we denote by $N(v)$ the set of its neighbours in $G(A, B)$. For a subset $S \subseteq V(G(A, B))$ we denote: $N(S) = \cup_{v \in S} N(v)$.

Definition

A *perfect matching* in the graph G is the set of its pairwise vertex-disjoint edges that cover all its vertices.

Hall's Theorem

Bipartite graph $G(A, B)$ has a perfect matching if and only if $|N(S)| \geq |S|$ for every $S \subseteq A$.

Definitions again...

Definition

Assume that $G(A, B)$ has a perfect matching. Let M be some fixed canonical matching in $G(A, B)$. Then for $S \subseteq A$ we denote $m(S) = \cup_{s \in S} m(s)$, where $(s, m(s)) \in M$.

Definition

Let $G(A, B)$ be a bipartite graph with $|A| = |B| = n$ and let M be its canonical matching. We say that a set $S \subseteq V(A)$ is *closed* if $N(S) = m(S)$.

Simple theorems

Lemma

If $G(A, B)$ is an arbitrary d-regular graph and the adversary does not know in advance any edges of the matching he is looking for then every person is d-anonymous.

Simple theorems

Lemma

If $G(A, B)$ is an arbitrary d-regular graph and the adversary does not know in advance any edges of the matching he is looking for then every person is d-anonymous.

Proof:

It suffices to prove that for every edge e of $G(A, B)$ there exists a perfect matching in $G(A, B)$ that uses e. This is a direct implication of Hall's Theorem. You keep finding matchings one by one, removing edges of the matchings found so far from the graph.

Simple theorems - sustained attack with d-anonymity

Lemma

If $G(A, B)$ is clique-bipartite d-regular graph and the adversary knows in advance c edges of the matching then every person is $(d - c)$-anonymous.

Simple theorems - sustained attack with d-anonymity

Proof:
Follows immediately from the following lemma:

Lemma

Assume that $G(A, B)$ is clique-bipartite d-regular graph (i.e. it is a union of disjoint complete bipartite graphs). Denote by M some perfect matching in $G(A, B)$. Let C be some subset of the edges of M and let $c = |C|$. Fix some vertex $v \in A$ not matched in C. Then there are at least $(d - c)$ edges adjacent to v such that for each edge e like that there exists some perfect matching M^e in $G(A, B)$ that uses both e and C.

Main Result - Adversary with extra Knowledge

Theorem (Choromanski '11-15)

Let G(A,B) be a k-regular bipartite graph with color classes: A and B. Assume that $|A| = |B| = n$. Denote by M some perfect matching M in $G(A,B)$. Let C be some subset of the edges of M and let $c = |C|$. Take some $\xi \geq c$. Denote $\hat{n} = n - c$. Fix any function $\phi : N \rightarrow R$ satisfying $\forall_k (\xi \sqrt{2k + \frac{1}{4}} < \phi(k) < k)$. Then for all but at most

$$\delta = \frac{2ck^2 \hat{n} \xi (1 + \frac{\phi(k) + \sqrt{\phi^2(k) - 2\xi^2 k}}{2\xi k})}{\phi^3(k)(1 + \sqrt{1 - \frac{2\xi^2 k}{\phi^2(k)}})(\frac{1}{\xi} - \frac{c}{\phi(k)} + \frac{k(1 - \frac{c}{\xi})}{\phi(k)})} + \frac{ck}{\phi(k)}$$

vertices $v \in A$ not matched in C the following holds:

Main Result - Adversary with extra knowledge

Theorem (Choromanski '11-15)

The size of the set of edges e adjacent to v and with the additional property that there exists some perfect matching M^v in $G(A, B)$ that uses e and edges from C is at least $(k - c - \phi(k))$.

Main Result - proof

Definition

Take a bipartite graph $G_{del} = G(A_{del}, B_{del})$ with color classes A_{del}, B_{del}, obtained from $G(A, B)$ by deleting all the vertices of C. For a vertex $v \in A_{del}$ and an edge e adjacent to it in G_{del} we say that e is *bad in respect to* v if there is no perfect matching in $G(A, B)$ that uses e and all the edges from C.

Definition

We say that a vertex $v \in A_{del}$ is *bad* if there are at least $\phi(d)$ edges bad with respect to v.

Main Result - proof

Lemma

For every edge e which is bad with respect to a vertex v there exists a closed set S_v^e such that $v \notin S_v^e$ and v is adjacent to some vertex in $m(S_v^e)$.

Definition

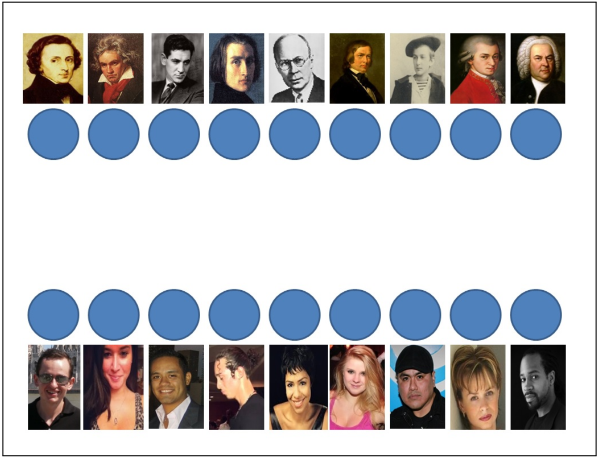
Fix some bad vertex v and some set E of its bad edges of size $\phi(d)$. Let $S_v^E = \bigcup_{e \in E} S_v^e$. Note that S_v^E is closed as a sum of closed sets. We also have: $v \notin S_v^E$. Besides every edge from E touches some vertex from $m(S_v^E)$. We say that the set S is $\phi(d)$-bad with respect to a vertex $v \in A_{del} - S$ if it is closed and there are $\phi(d)$ bad edges with respect to v that touch S. So we conclude that S_v^E is $\phi(d)$-bad with respect to v.

Main Result - proof

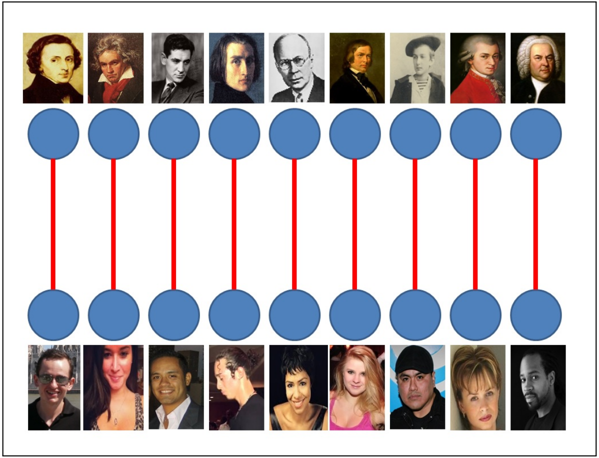
Definition

Denote by S_v^m a minimal $\phi(d)$-bad set with respect to v.

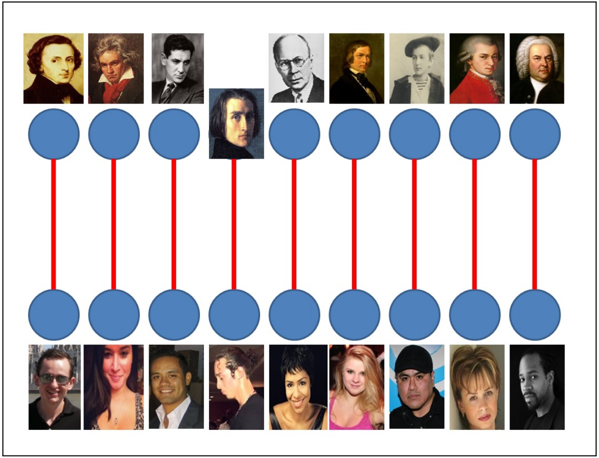
Main Result - proof



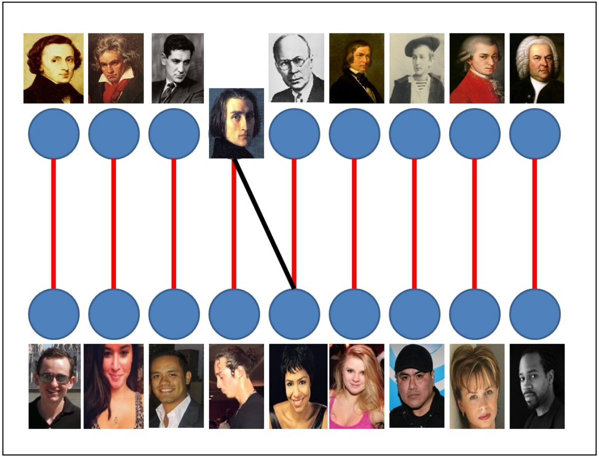
Main Result - proof



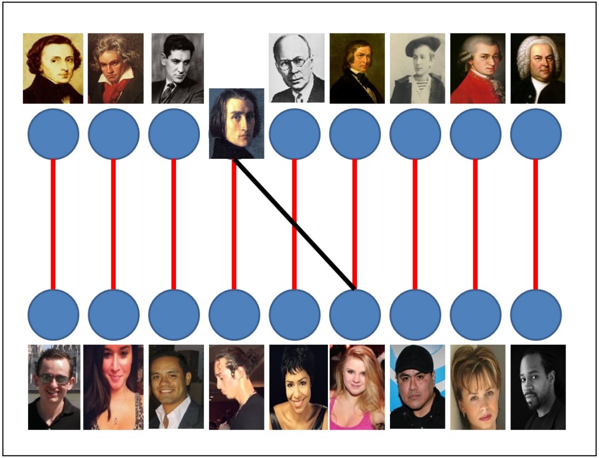
Main Result - proof



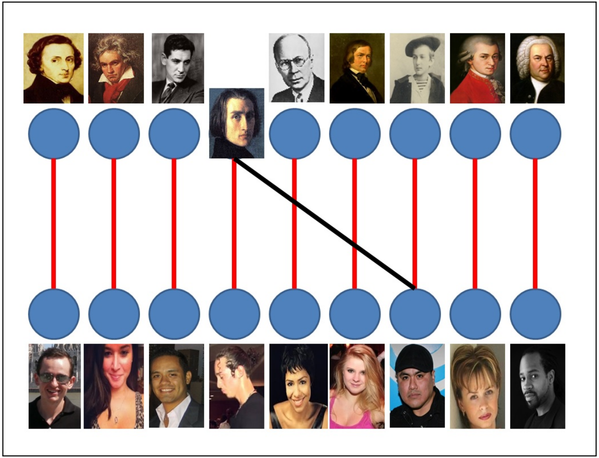
Main Result - proof



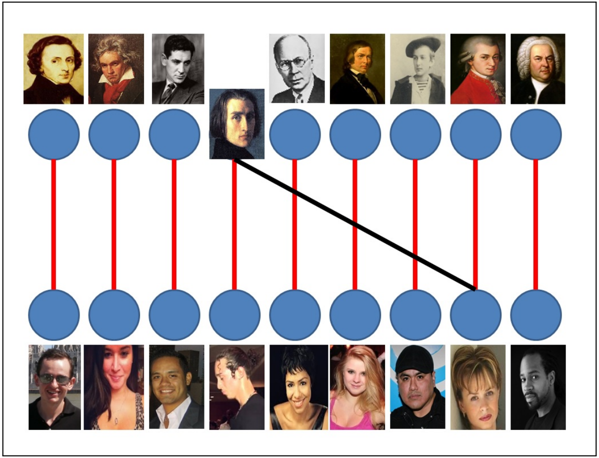
Main Result - proof



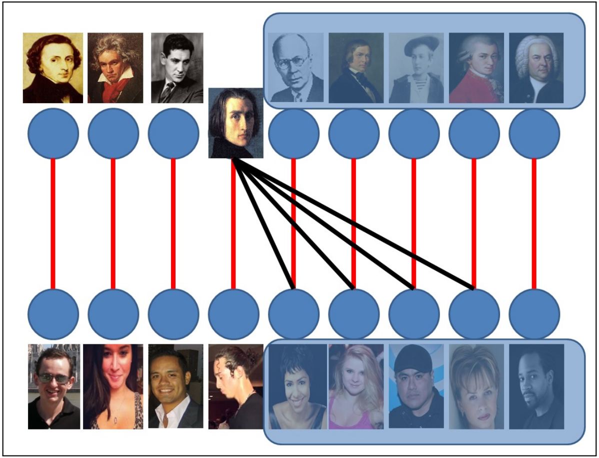
Main Result - proof



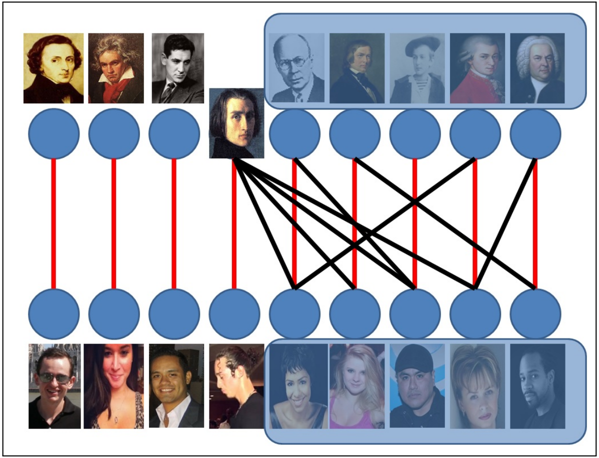
Main Result - proof



Main Result - proof



Main Result - proof

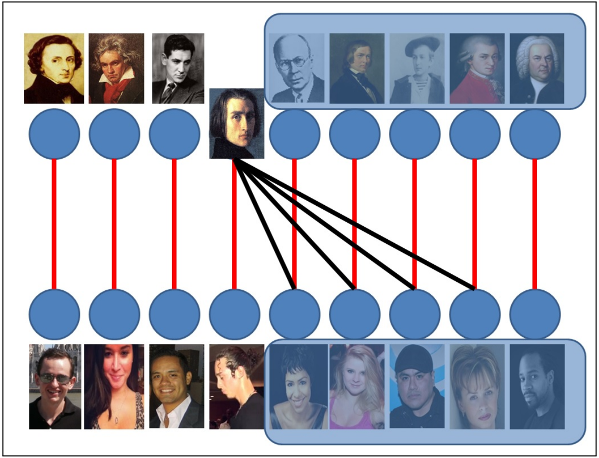


Main Result - proof

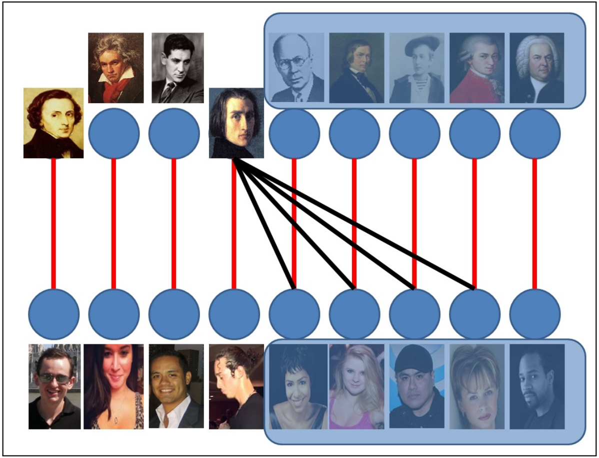
Lemma

Let v_1, v_2 be two bad vertices. If $v_2 \in S^m_{v_1}$ then $S^m_{v_2} \subseteq S^m_{v_1}$.

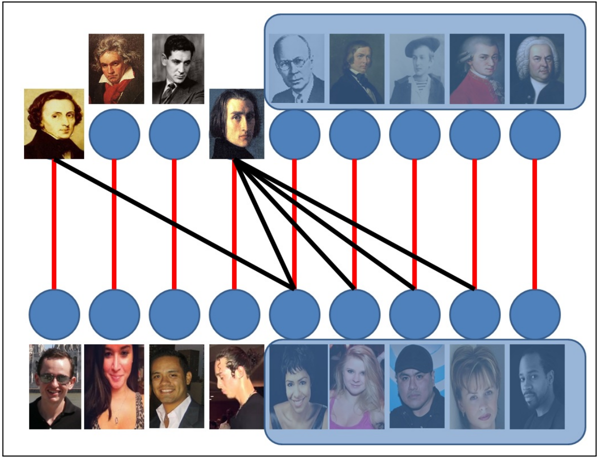
Main Result - proof



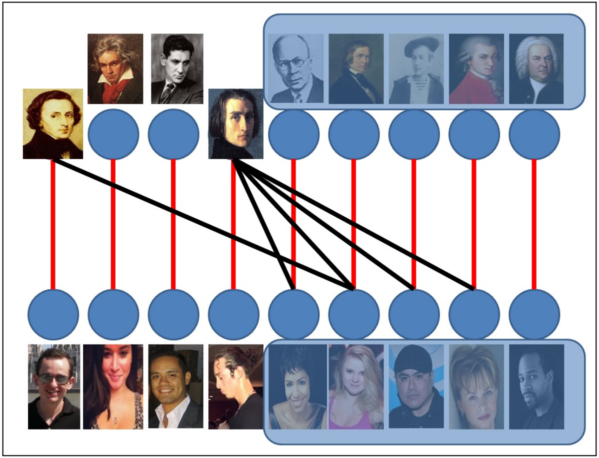
Main Result - proof



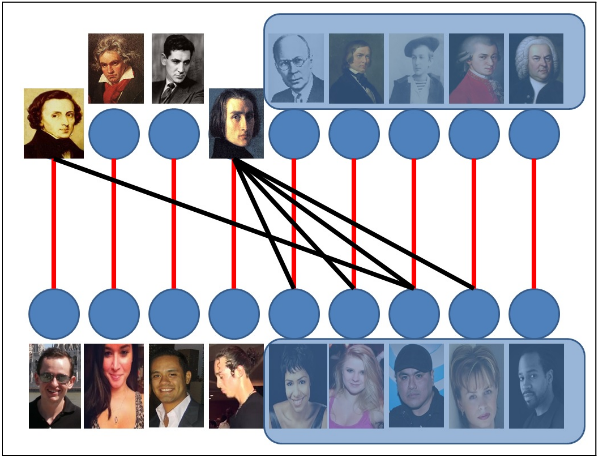
Main Result - proof



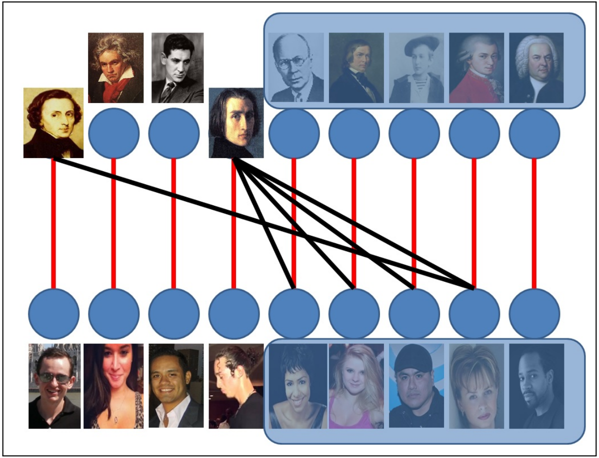
Main Result - proof



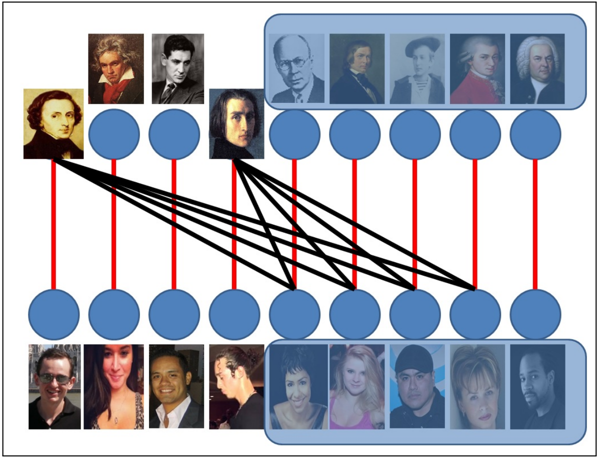
Main Result - proof



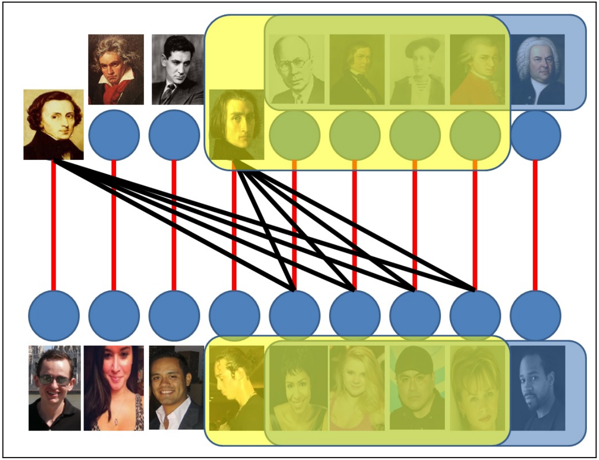
Main Result - proof



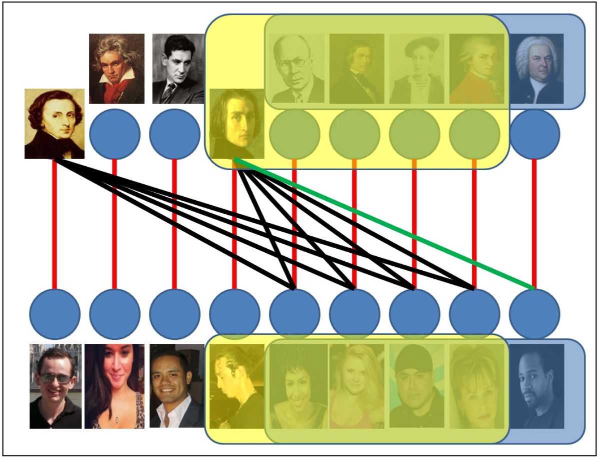
Main Result - proof



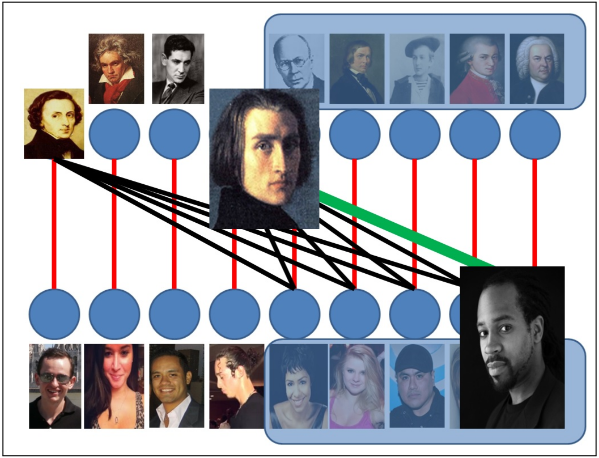
Main Result - proof



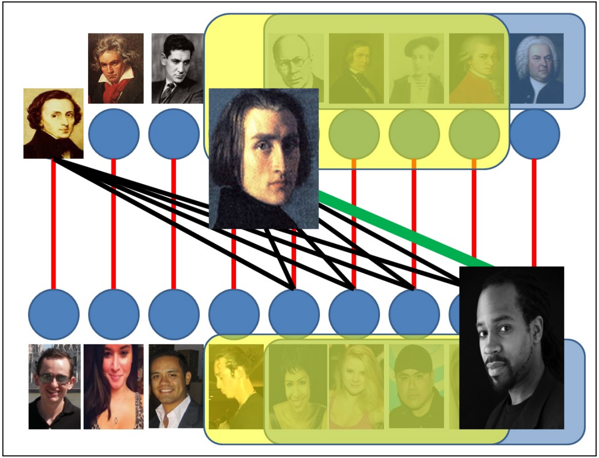
Main Result - proof



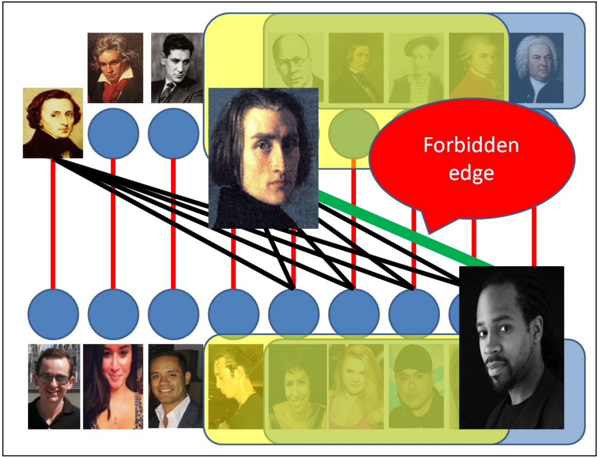
Main Result - proof



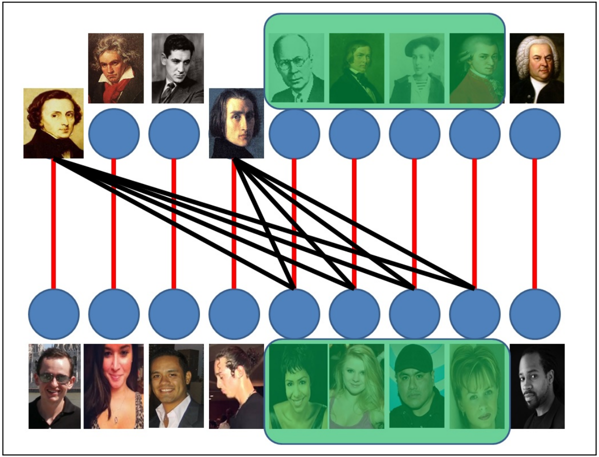
Main Result - proof



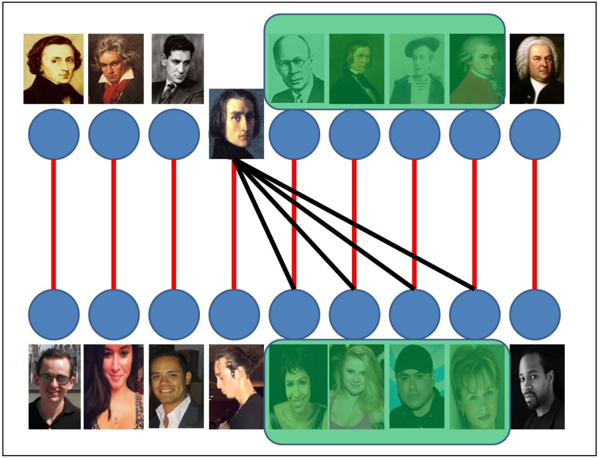
Main Result - proof



Main Result - proof



Main Result - proof



Main Result - proof

Lemma

Denote $P = \{S_v^m : v \in X\}$*. As a poset with an ordering induced by the inclusion relation, it does not have antichains of size larger than* $\frac{cd}{\phi(d)}$.

Main Result - proof

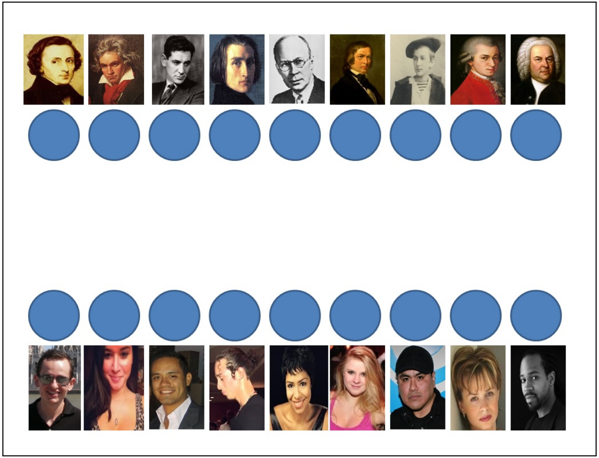
Lemma

Denote $P = \{S_v^m : v \in X\}$*. As a poset with an ordering induced by the inclusion relation, it does not have antichains of size larger than* $\frac{cd}{\phi(d)}$.

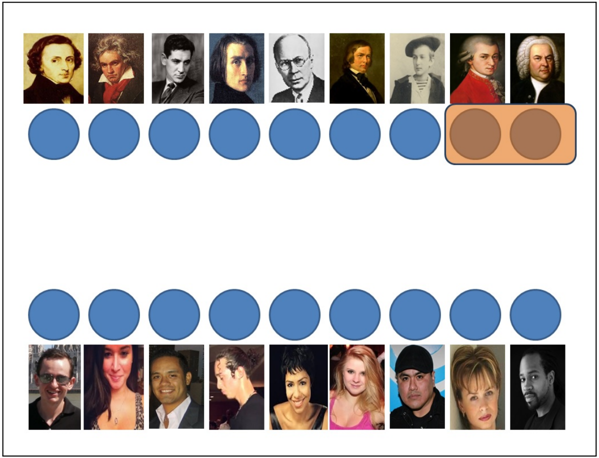
Corollary

Using Dilworth's lemma about chains and antichains in the poset and the previous lemma we can conclude that a set $P = \{S_v^m : v \in A\}$ *has a chain of length at least* $\frac{\hat{n}\phi(d)}{cd}$.

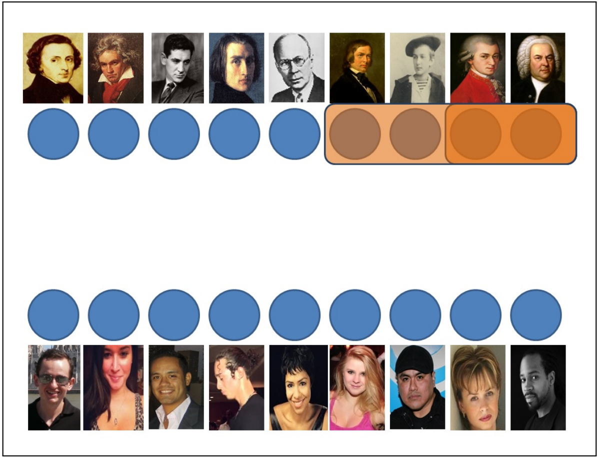
Main Result - proof



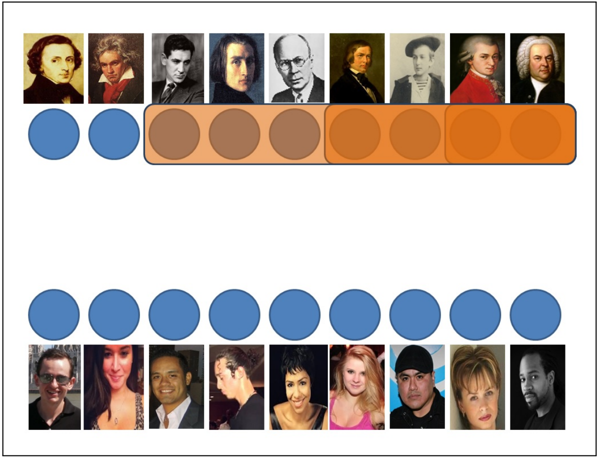
Main Result - proof



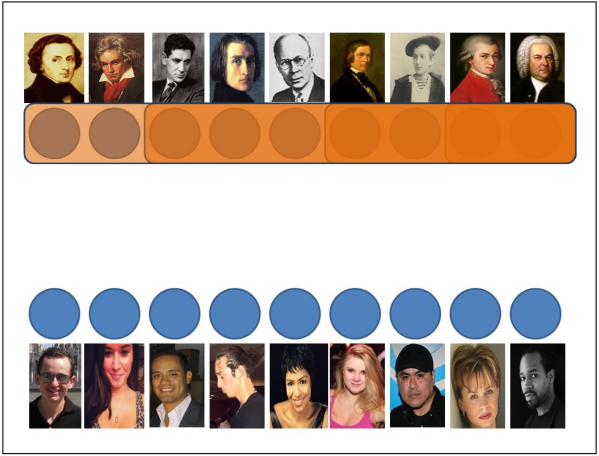
Main Result - proof



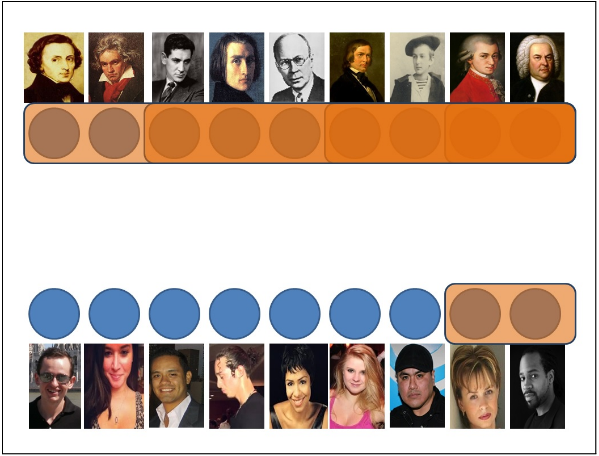
Main Result - proof



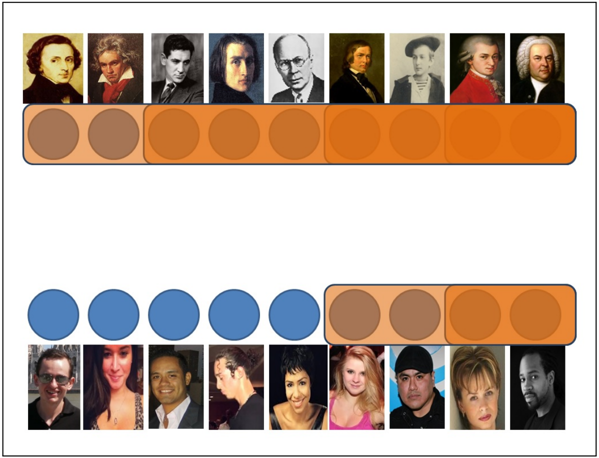
Main Result - proof



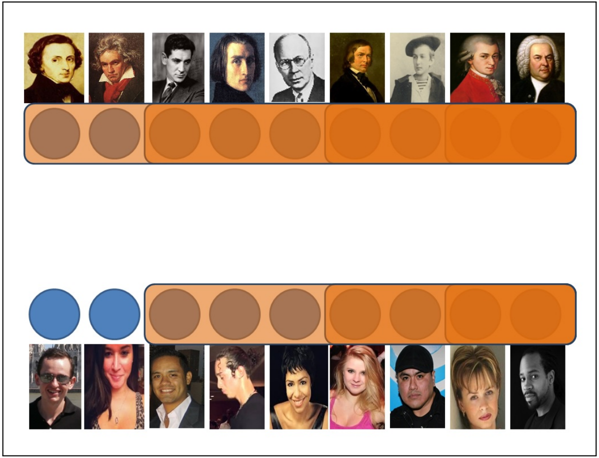
Main Result - proof



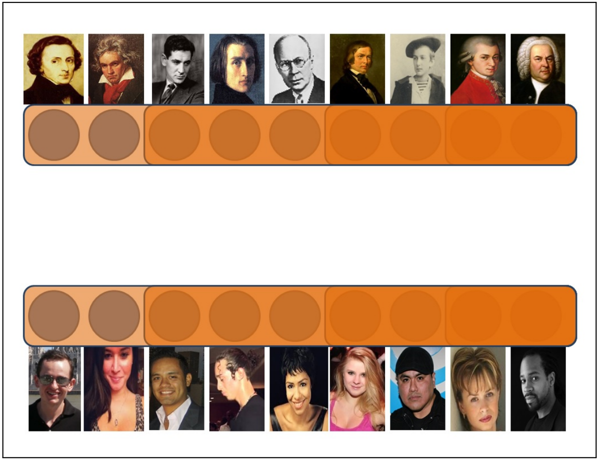
Main Result - proof



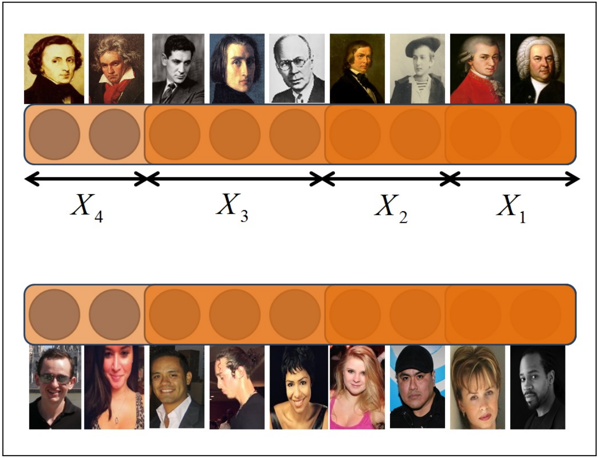
Main Result - proof



Main Result - proof



Main Result - proof



Main Result - proof

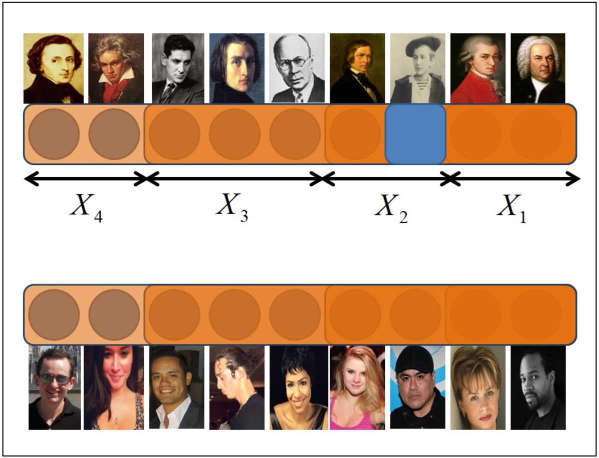
Gap Lemma

If $|X_i| > c$ then $|X_i| \geq \frac{\phi(k)}{c} - c$.

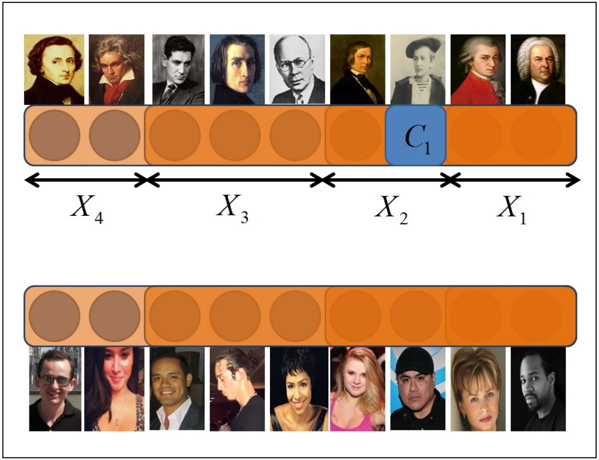
Short subsequences of small values

For every i and $l > \frac{\phi(d) - \sqrt{\phi^2(d) - 2\xi^2 d}}{\xi}$ in the sequence $(X_{i+1}, ..., X_{i+l})$ there exists at least one element of size more than c.

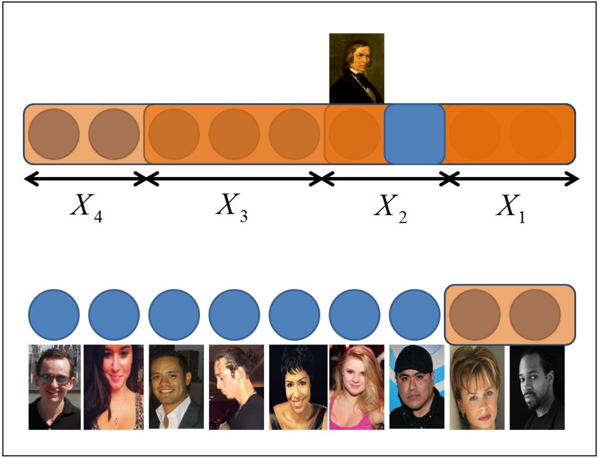
Main Result - proof - Gap Lemma



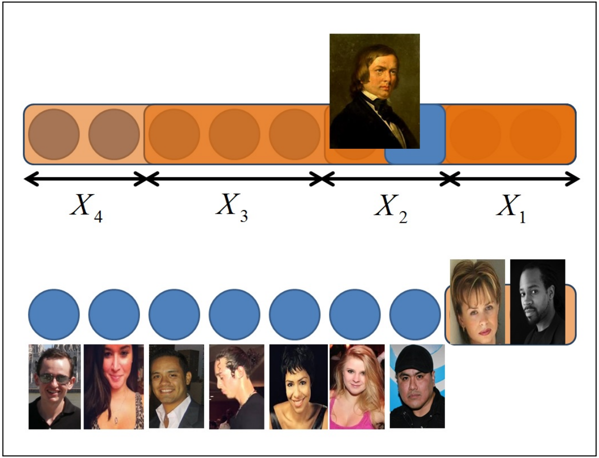
Main Result - proof - Gap Lemma



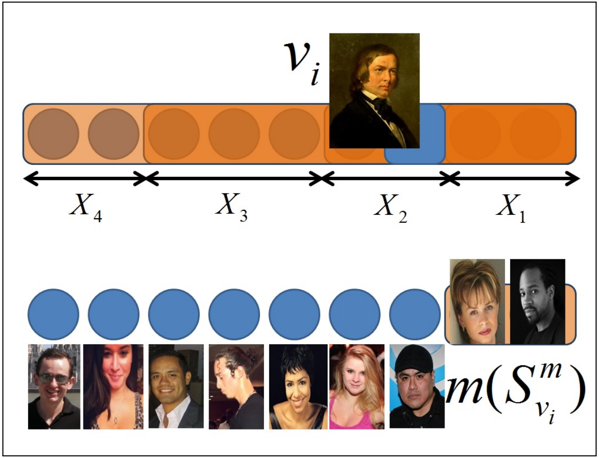
Main Result - proof - Gap Lemma



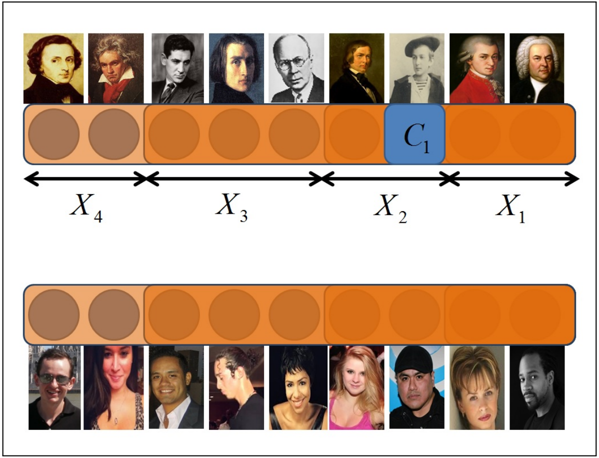
Main Result - proof - Gap Lemma



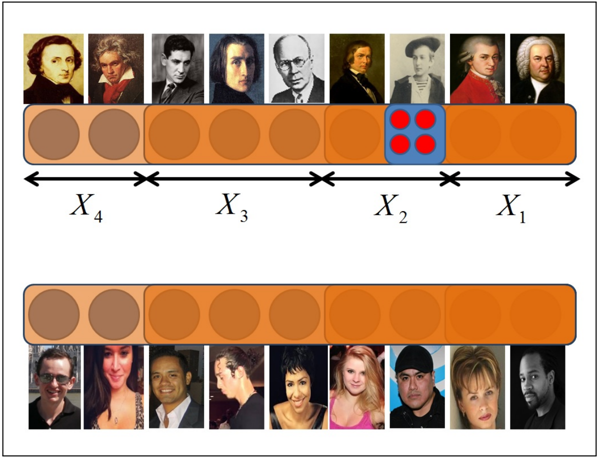
Main Result - proof - Gap Lemma



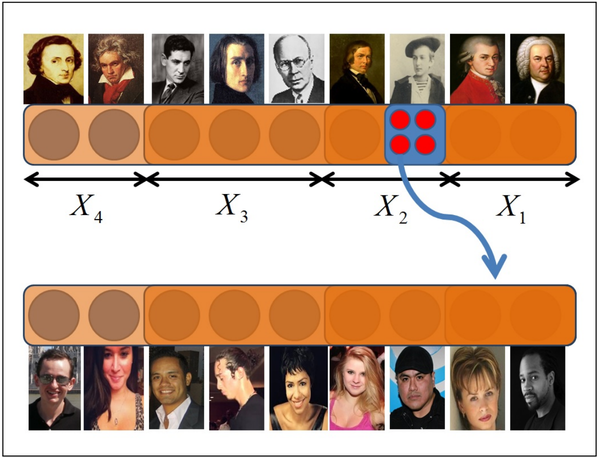
Main Result - proof - Gap Lemma



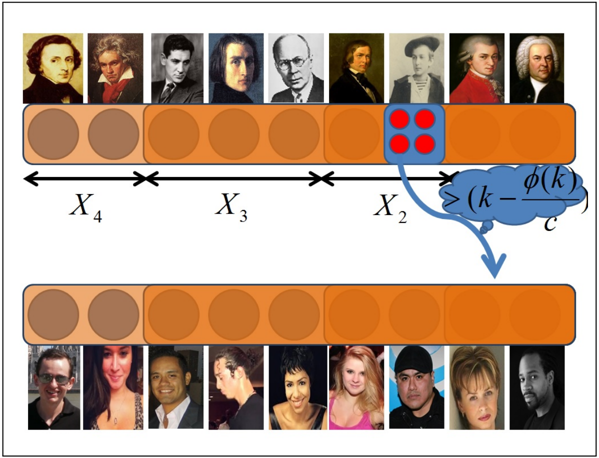
Main Result - proof - Gap Lemma



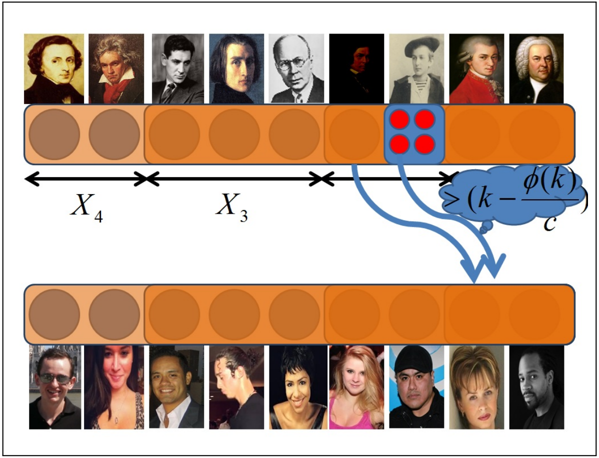
Main Result - proof - Gap Lemma



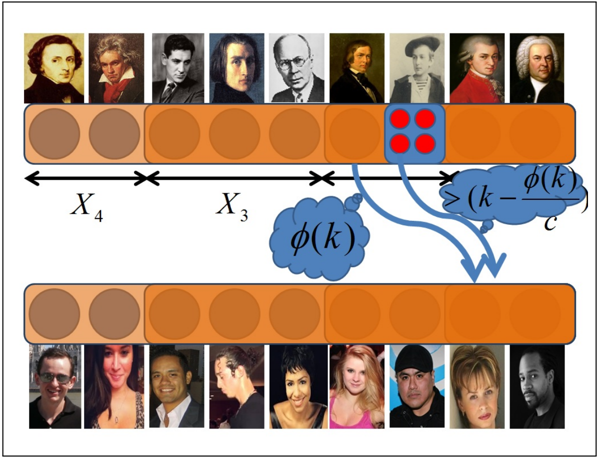
Main Result - proof - Gap Lemma



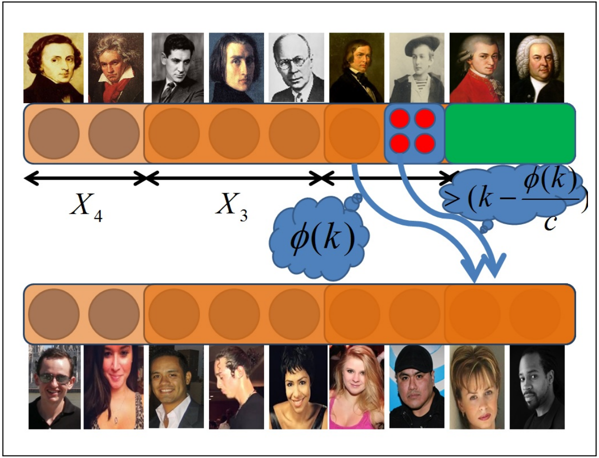
Main Result - proof - Gap Lemma



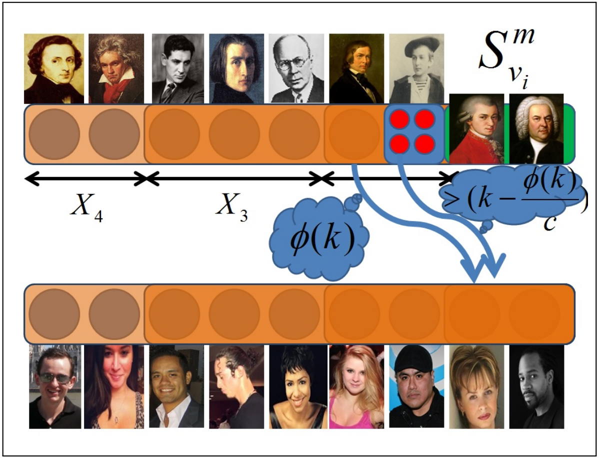
Main Result - proof - Gap Lemma



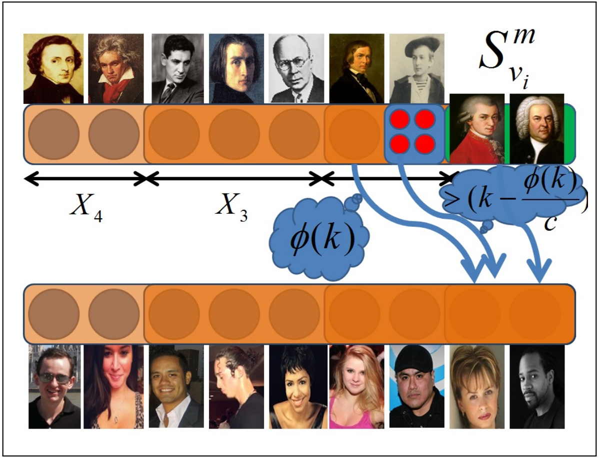
Main Result - proof - Gap Lemma



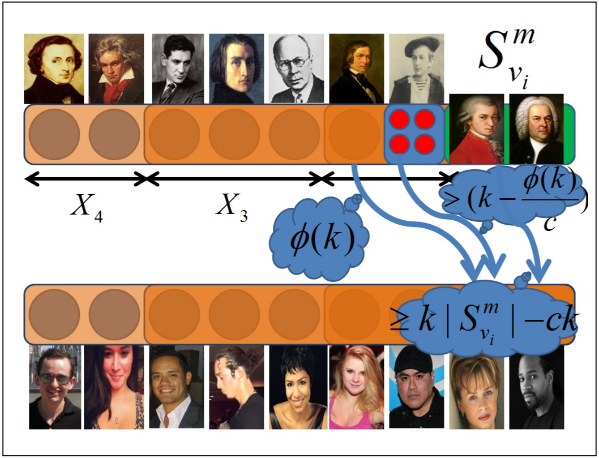
Main Result - proof - Gap Lemma



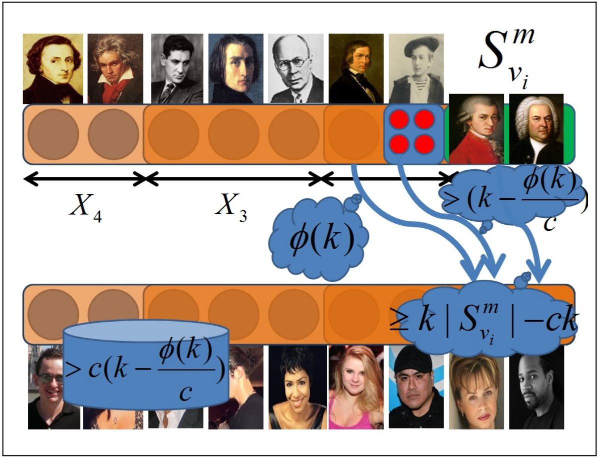
Main Result - proof - Gap Lemma



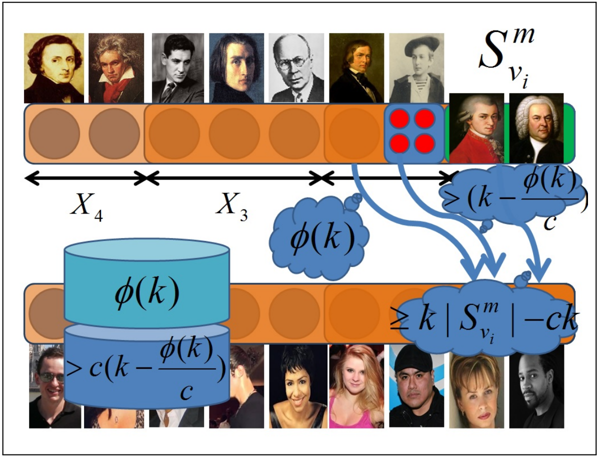
Main Result - proof - Gap Lemma



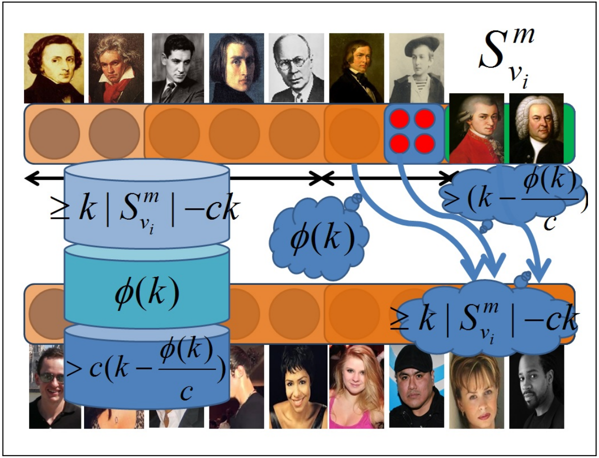
Main Result - proof - Gap Lemma



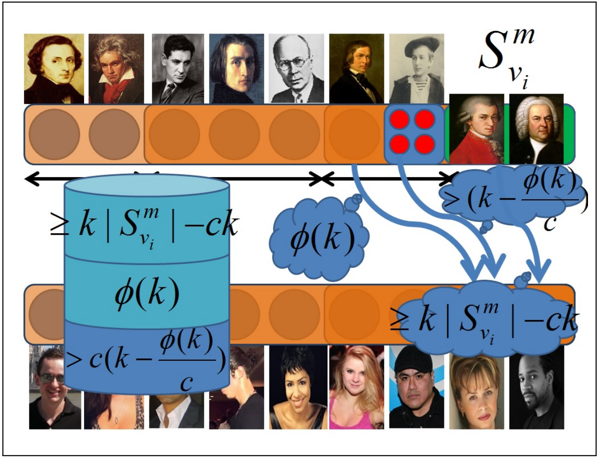
Main Result - proof - Gap Lemma



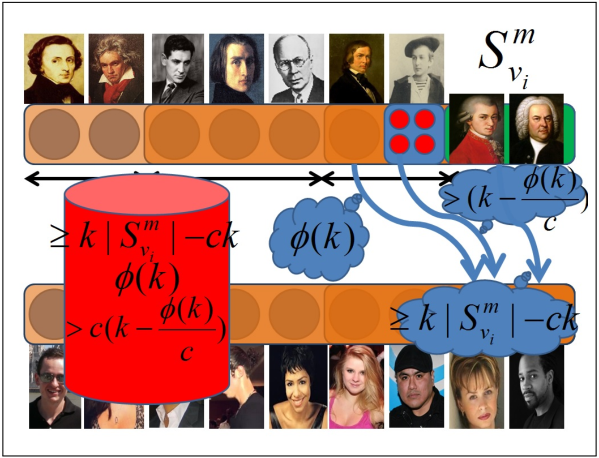
Main Result - proof - Gap Lemma



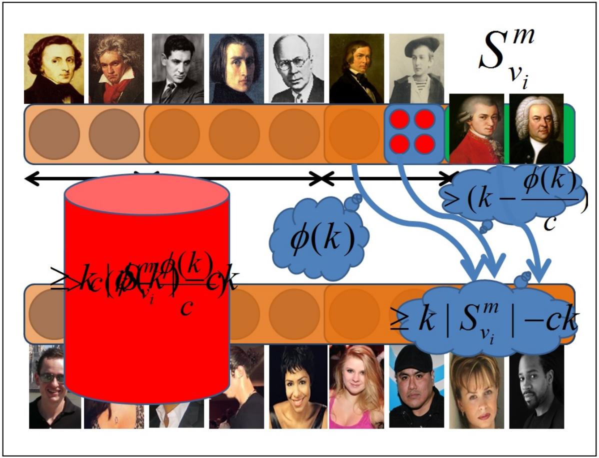
Main Result - proof - Gap Lemma



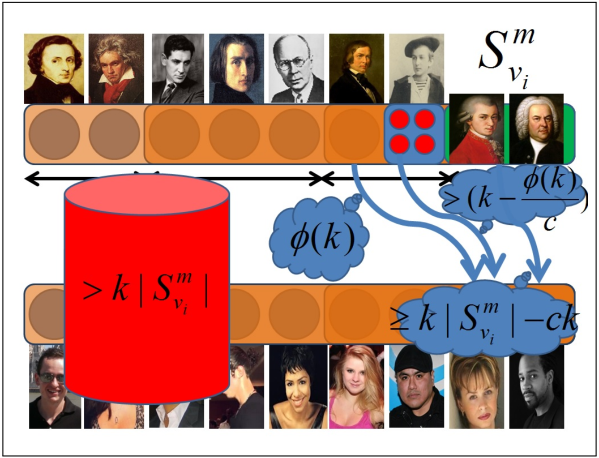
Main Result - proof - Gap Lemma



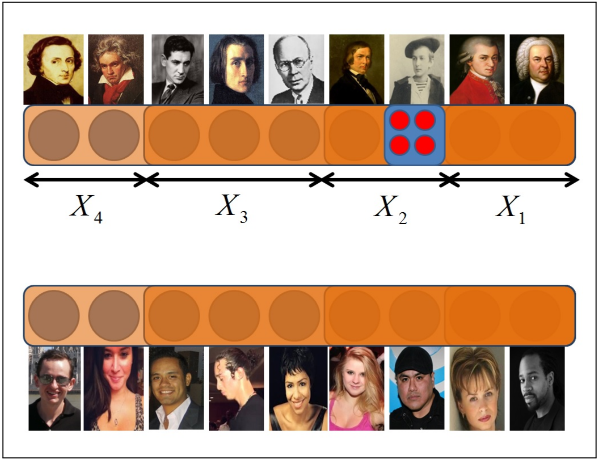
Main Result - proof - Gap Lemma



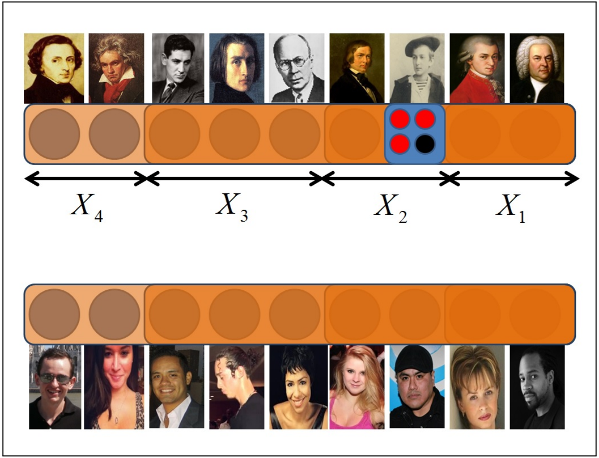
Main Result - proof - Gap Lemma



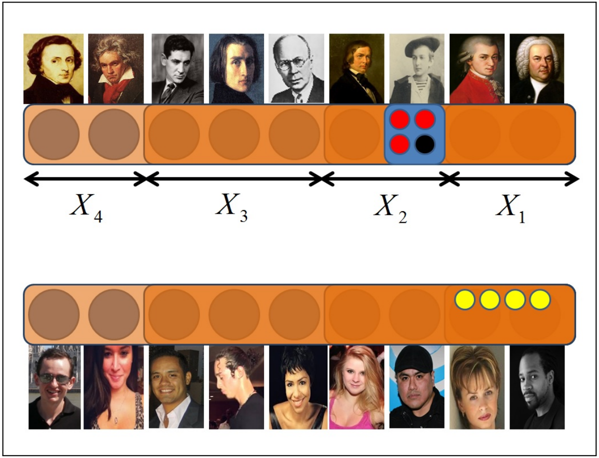
Main Result - proof - Gap Lemma



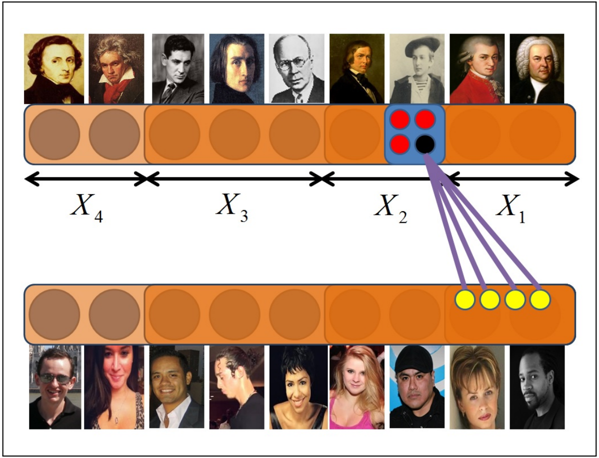
Main Result - proof - Gap Lemma



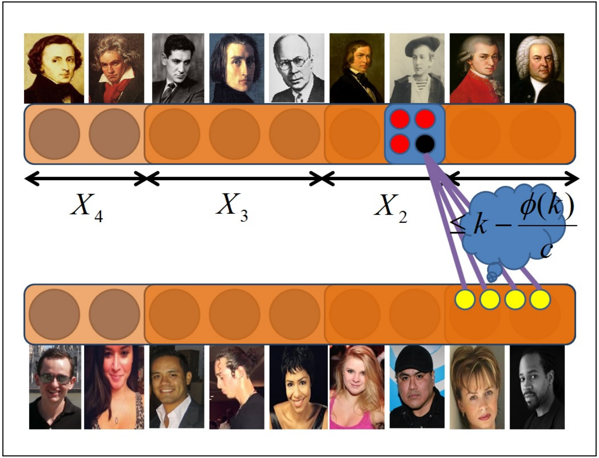
Main Result - proof - Gap Lemma



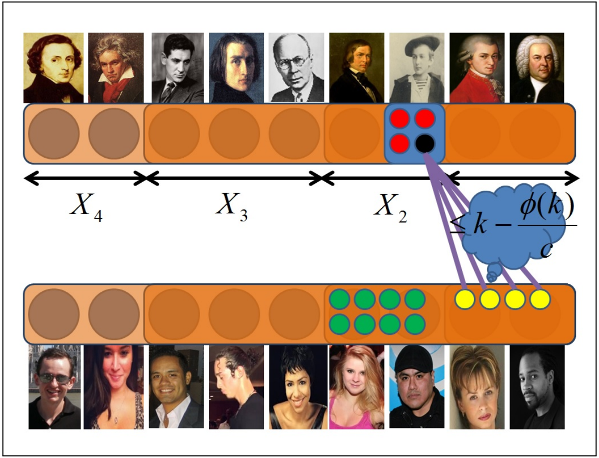
Main Result - proof - Gap Lemma



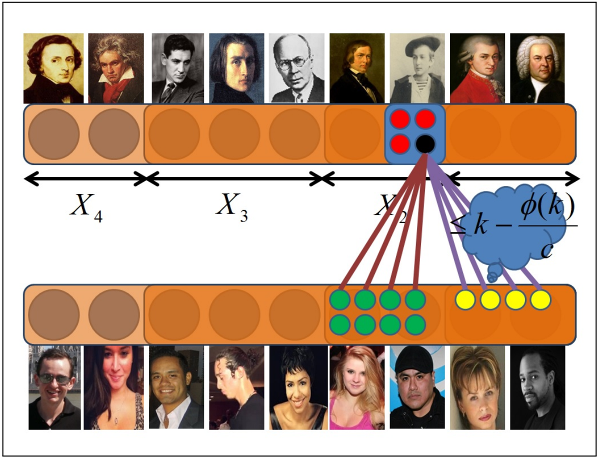
Main Result - proof - Gap Lemma



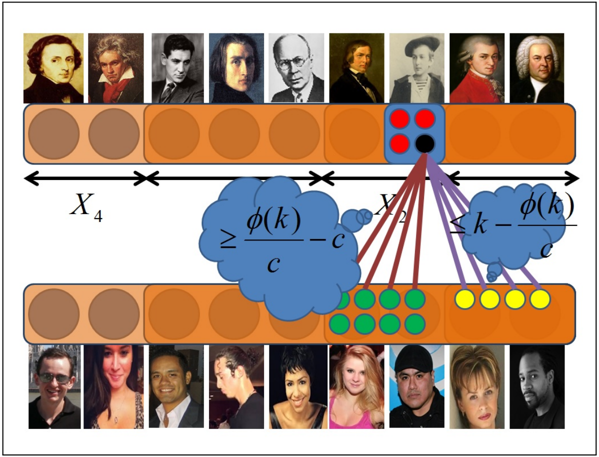
Main Result - proof - Gap Lemma



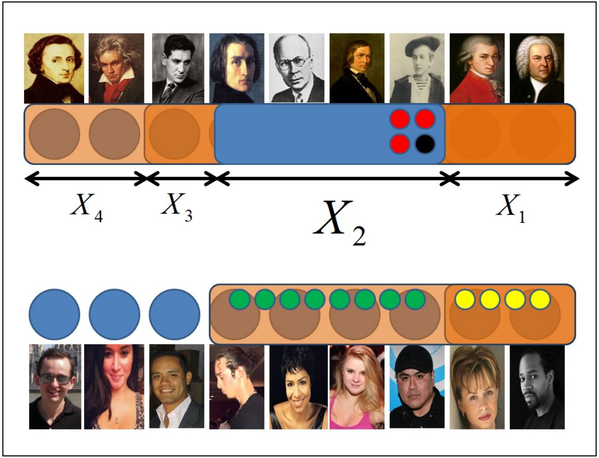
Main Result - proof - Gap Lemma



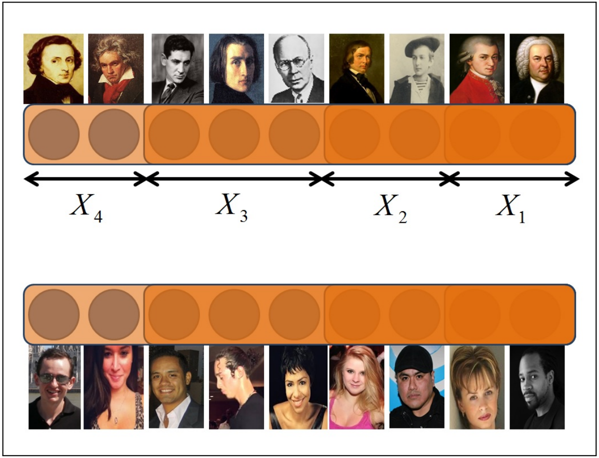
Main Result - proof - Gap Lemma



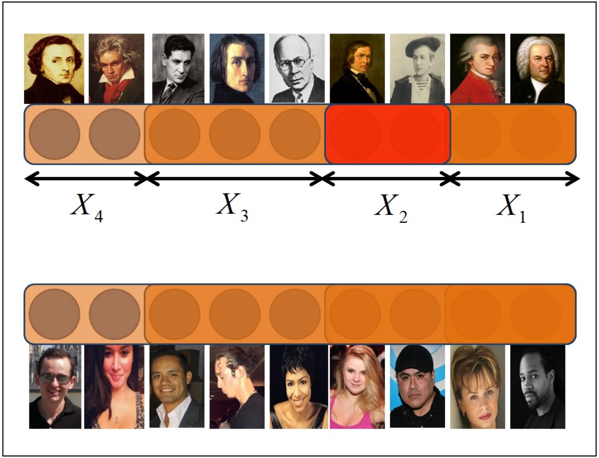
Main Result - proof - Gap Lemma



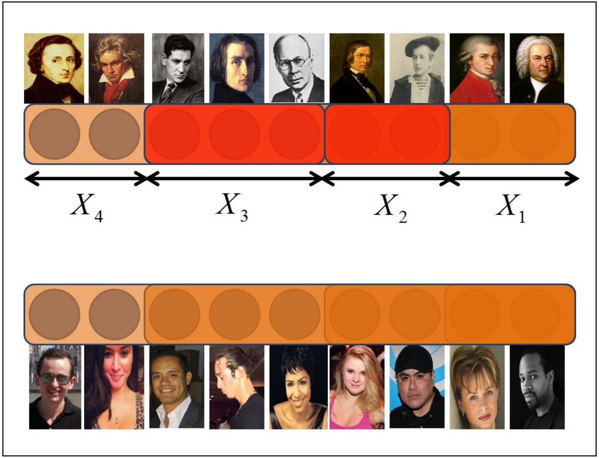
Main Result - proof - short subsequences



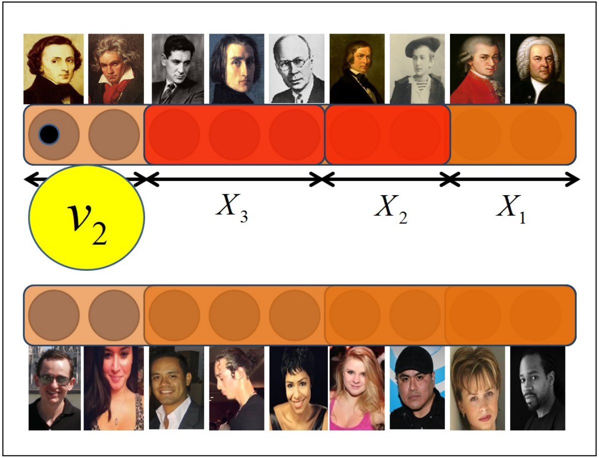
Main Result - proof - short subsequences



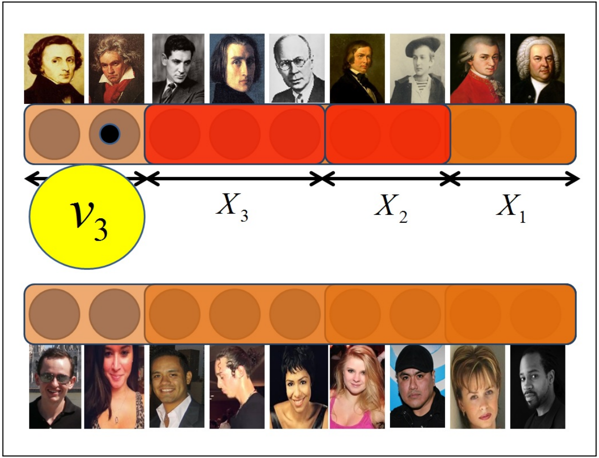
Main Result - proof - short subsequences



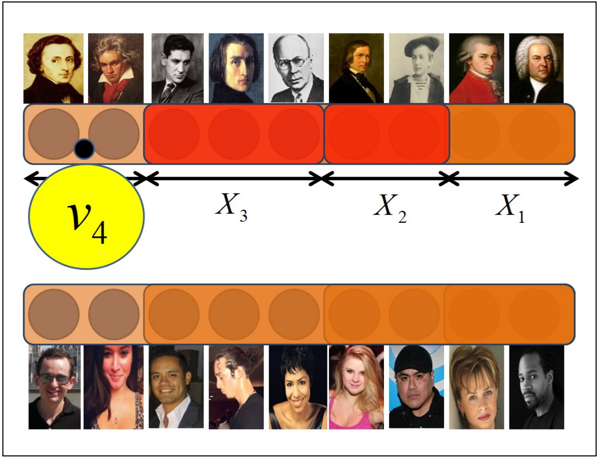
Main Result - proof - short subsequences



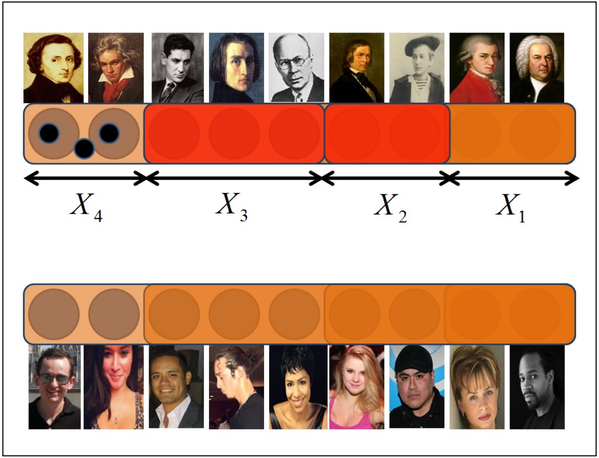
Main Result - proof - short subsequences



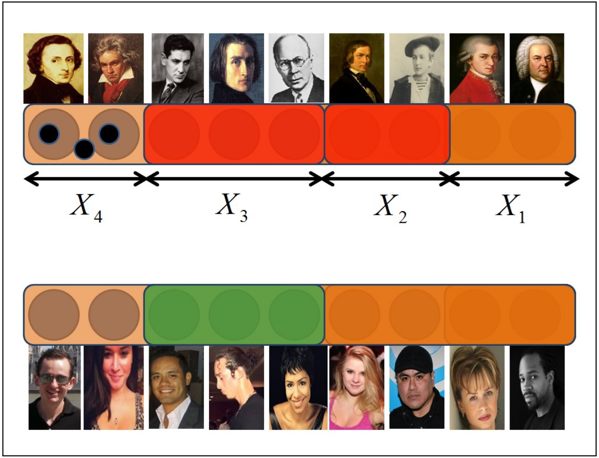
Main Result - proof - short subsequences



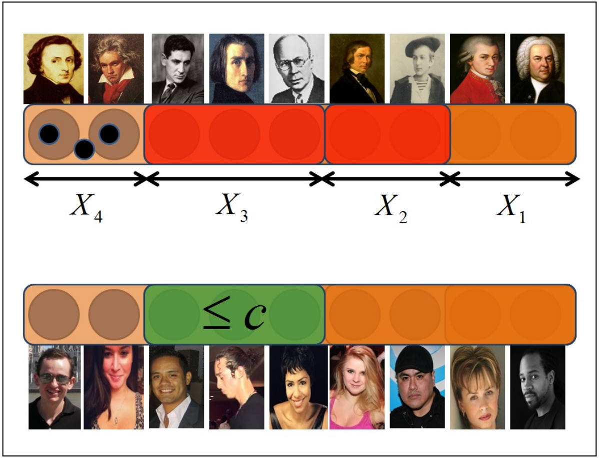
Main Result - proof - short subsequences



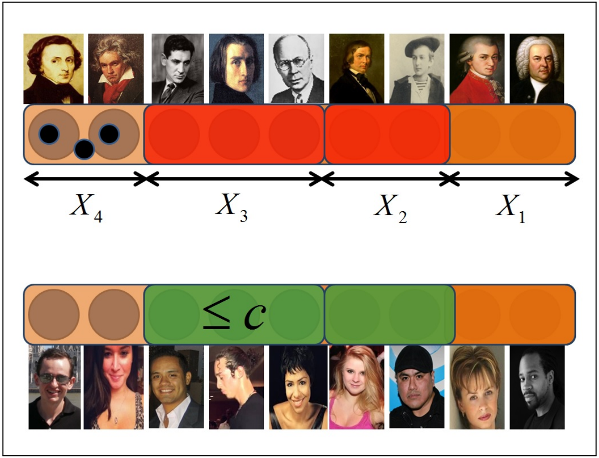
Main Result - proof - short subsequences



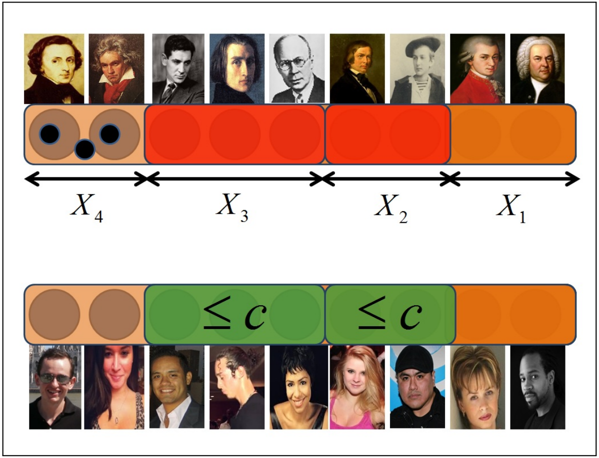
Main Result - proof - short subsequences



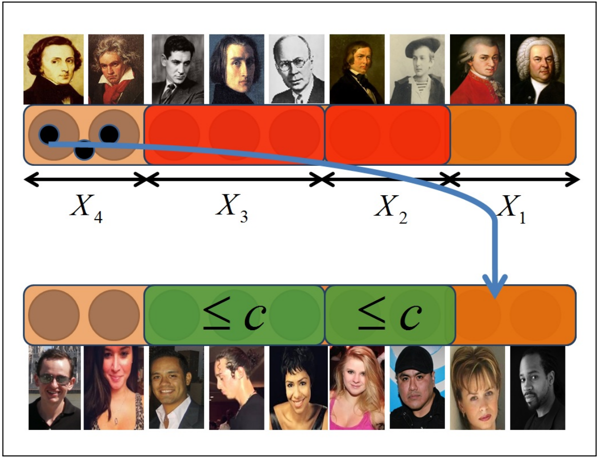
Main Result - proof - short subsequences



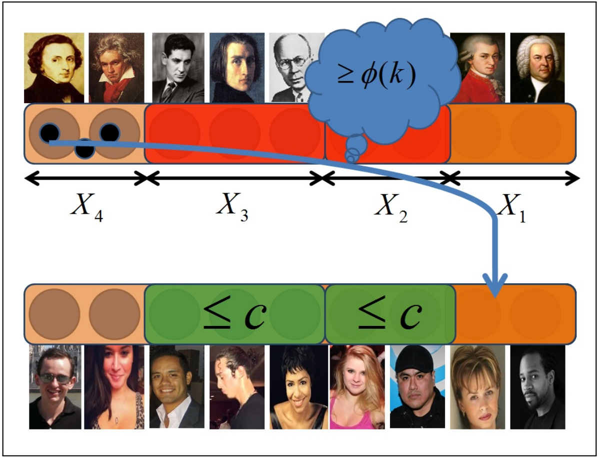
Main Result - proof - short subsequences



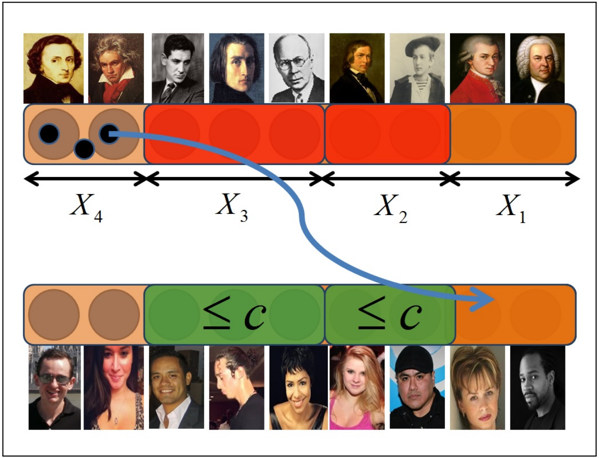
Main Result - proof - short subsequences



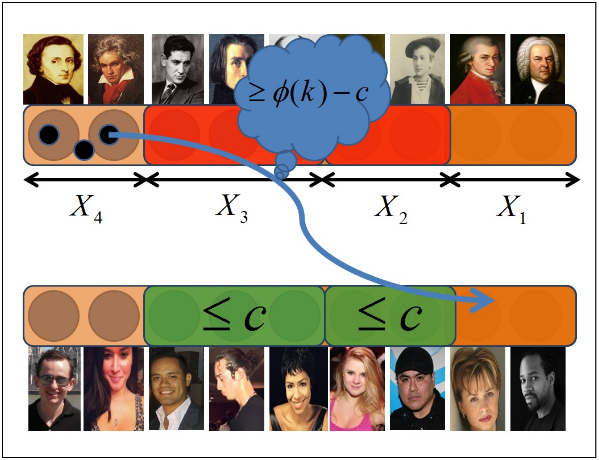
Main Result - proof - short subsequences



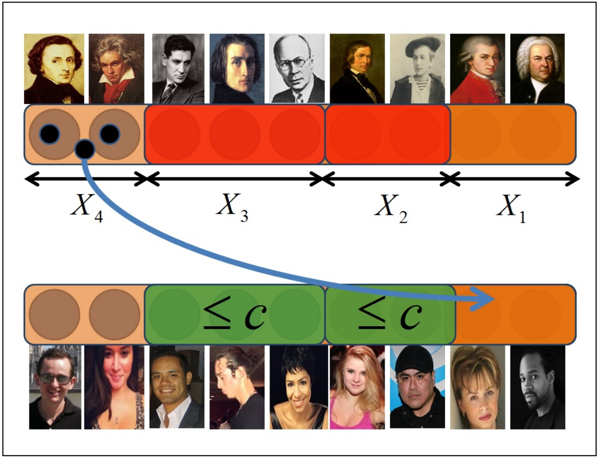
Main Result - proof - short subsequences



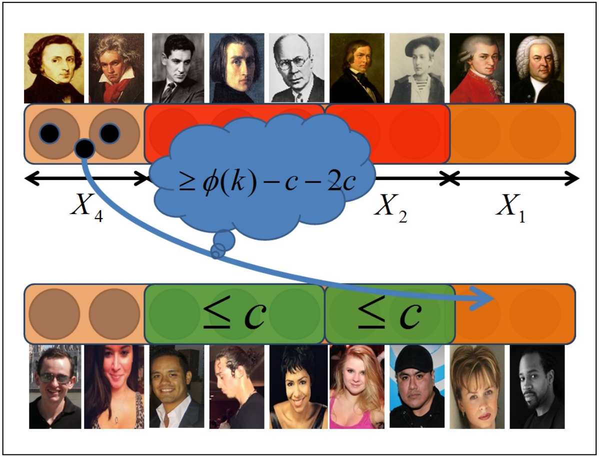
Main Result - proof - short subsequences



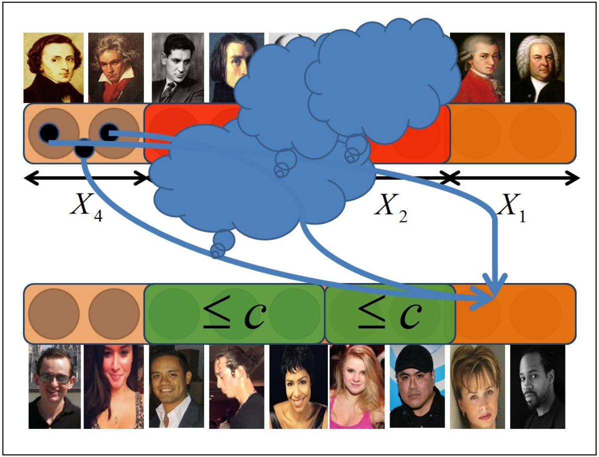
Main Result - proof - short subsequences



Main Result - proof - short subsequences



Main Result - proof - short subsequences



Main Result - proof - short subsequences