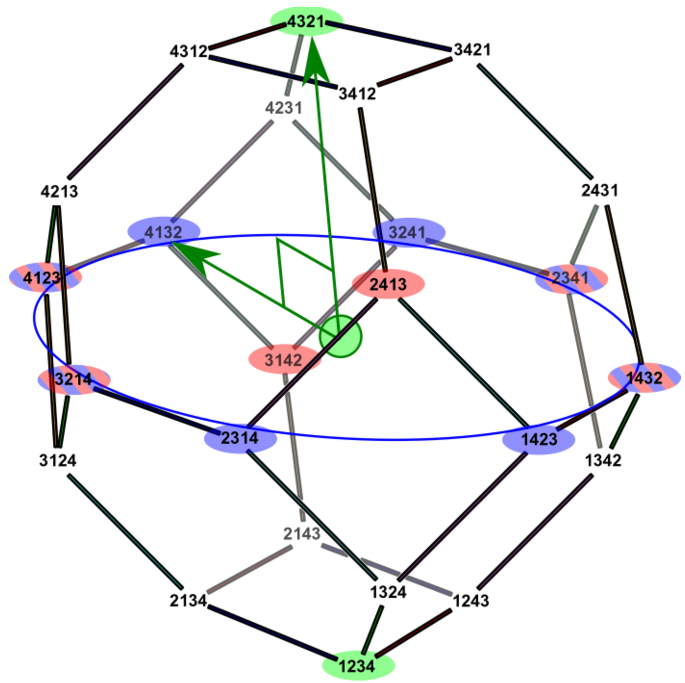
Two Equators of the Permutohedron



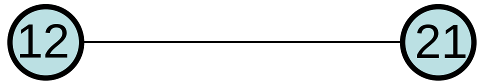
Joshua Cooper

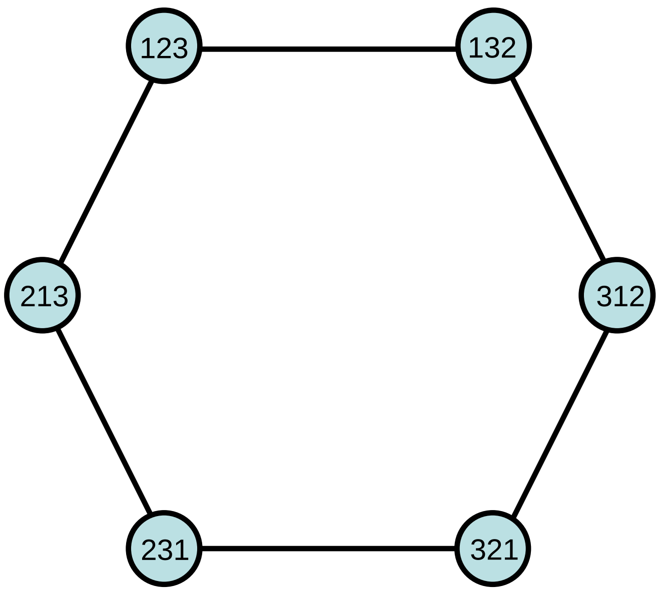
with Rory Mitchell (Nvidia), Eibe Frank (U Waikato), & Geoffrey Holmes (U Waikato)

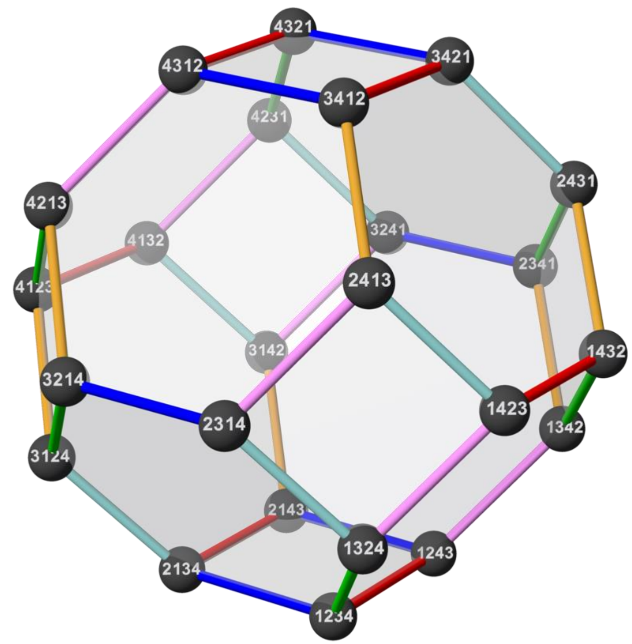
Permutation Patterns 2021

June 15, 2021

Definition. The order-n *permutohedron* is the convex polytope whose vertices are all n-vectors whose coordinates are a permutations of $\{1, \ldots, n\}$.



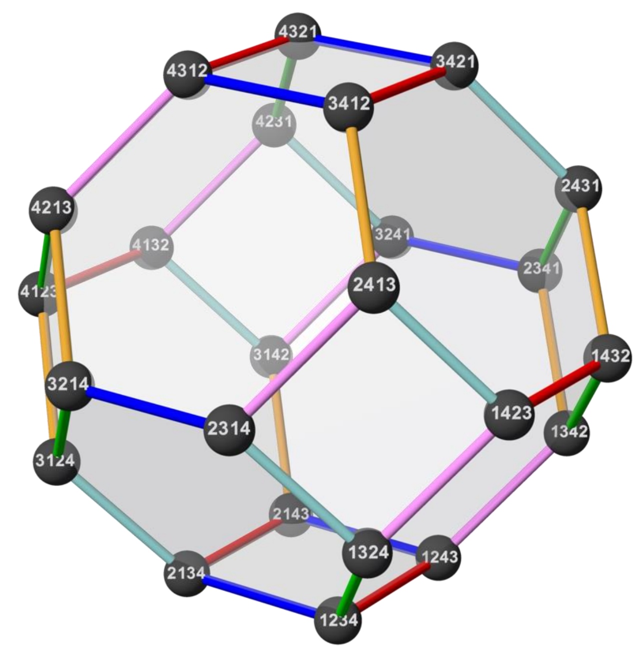




Definition. The order-n *permutohedron* is the convex polytope whose vertices are all n-vectors whose coordinates are a permutations of $\{1, \ldots, n\}$.

Facts:

- $n!$ vertices
- There are $n - 1$ edges containing each vertex σ, ending at each σ' which differs from σ by an adjacent transposition.
- Thus, its 1-skeleton is the Hasse diagram of the weak Bruhat order on $\boldsymbol{S}_n$ (and a Cayley graph)
- Since sum of coordinates is always $\binom{n+1}{2}$, dimension is $n - 1$.



Definition. The order-n *permutohedron* is the convex polytope whose vertices are all n-vectors whose coordinates are a permutations of $\{1, \ldots, n\}$.

Facts:

- $n!$ vertices
- There are $n - 1$ edges containing each vertex σ, ending at each σ' which differs from σ by an adjacent transposition.
- Thus, its 1-skeleton is the Hasse diagram of the weak Bruhat order on $\boldsymbol{S}_n$ (and a Cayley graph)
- Since sum of coordinates is always $\binom{n+1}{2}$, dimension is $n - 1$.

Estimating feature importances is a fundamental problem in machine learning (ML). In general, NP-hard.

Adopting every feature one at a time *over all orderings/permutations* and computing the average impact on their marginal values can be informative – this is exactly their *Shapley values*.

Shapley value: a way to measure how important different features are, *taking into account the impact on multi-factor/coalitional importance*

Considering all possible permutations of the n features is prohibitively expensive, we ask: can one obtain a quasi-random set of permutations that estimates well?

To ensure a set is well-distributed, one could sample a random set of (nearly) *orthogonal* permutations. However, the *kernel* matters a lot here.

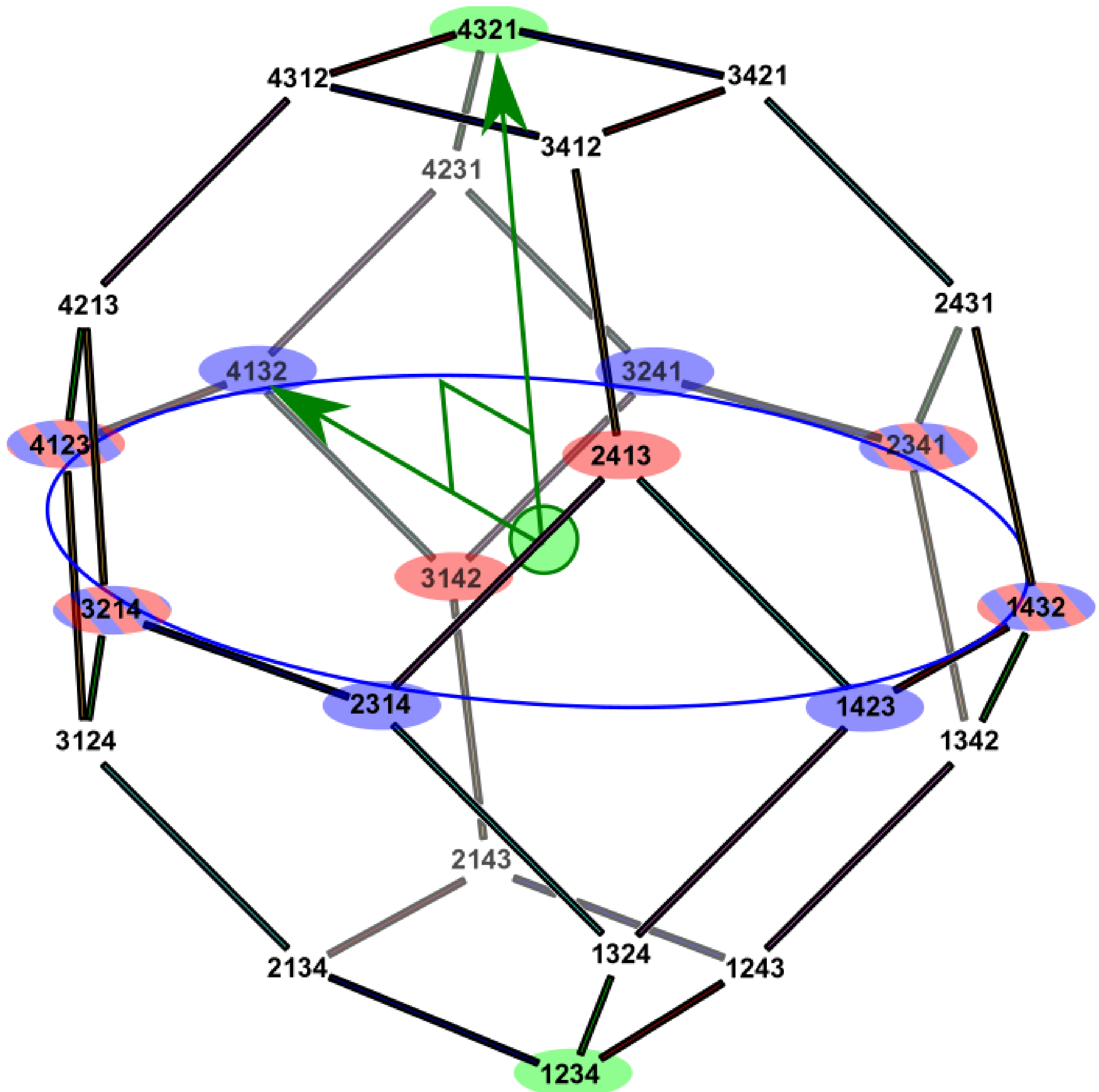
Definition. Given two permutations σ and σ', the Kendall τ kernel is given by

$$K_\tau(\sigma, \sigma') = 1 - \frac{2 \cdot \mathrm{inv}(\sigma^{-1}\sigma')}{\binom{n}{2}}$$

Note: $K_\tau(\sigma, \sigma') = 1$ iff $\sigma = \sigma'$, $K_\tau(\sigma, \sigma') = -1$ iff σ is the reverse of σ', and $K_\tau(\sigma, \sigma') = 0$ iff $\sigma^{-1}\sigma'$ is at the middle level of the Bruhat order.

"combinatorial equator"

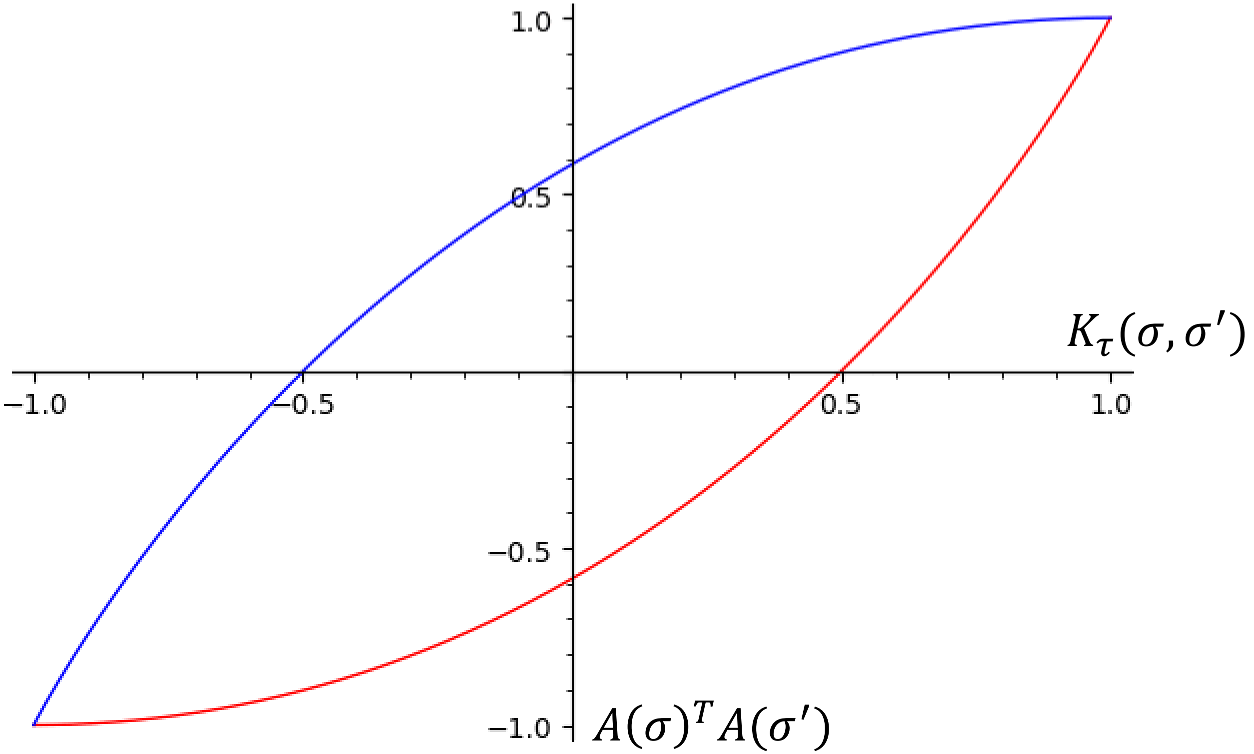
*Choosing orthogonal vectors w.r.t. Kendall τ is slow/hard***.** Maybe it's almost the same as the **geometric equator**? That is, project σ and σ' by the affine map $A(\cdot)$ that centers & normalizes the permutohedron, and seek dot-product orthogonality.


4321
4312
3421
3412
4231
4213
2431
4132
3241
4123
2413
2341
3142
3214
1432
2314
1423
3124
1342
2143
2134
1324
1243
1234

Theorem (Mitchell-C-Frank-Holmes '21+).

$$-2+3K_\tau(\sigma,\sigma')+4\left(\frac{1-K_\tau(\sigma,\sigma')}{2}\right)^{\frac{3}{2}} \le A(\sigma)^T A(\sigma')+O(n^{-1}) \le 2+3K_\tau(\sigma,\sigma')-4\left(\frac{1+K_\tau(\sigma,\sigma')}{2}\right)^{\frac{3}{2}}$$

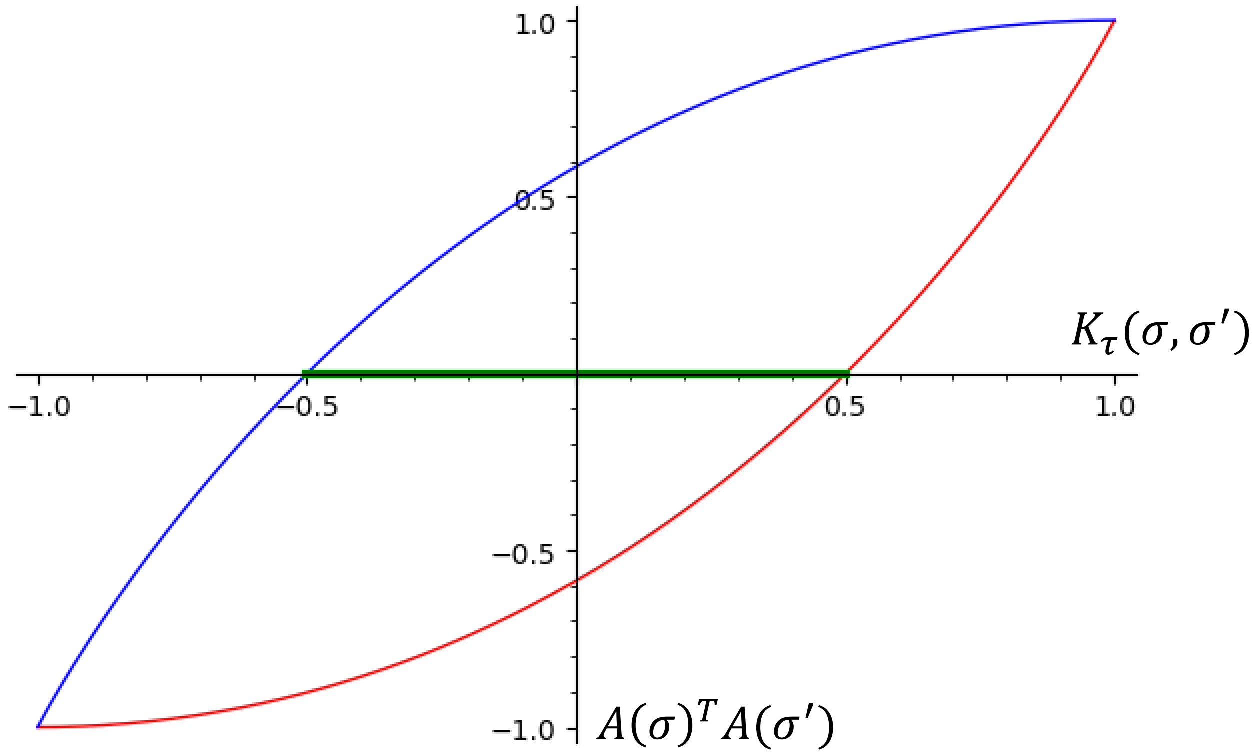
dot product of the permutations



Theorem (Mitchell-C-Frank-Holmes ’21+).

$$-2 + 3K_\tau(\sigma,\sigma') + 4\left(\frac{1-K_\tau(\sigma,\sigma')}{2}\right)^{\frac{3}{2}} \leq A(\sigma)^T A(\sigma') + O(n^{-1}) \leq 2 + 3K_\tau(\sigma,\sigma') - 4\left(\frac{1+K_\tau(\sigma,\sigma')}{2}\right)^{\frac{3}{2}}$$

Corollary. If $A(\sigma)^T A(\sigma') = o(1)$, then $|K_\tau(\sigma,\sigma')| \leq \frac{1}{2} + o(1)$.



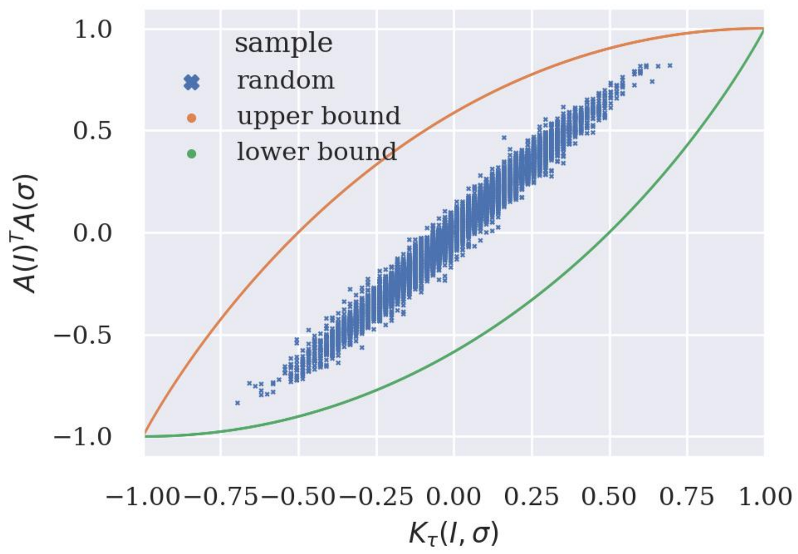
Theorem (Mitchell-C-Frank-Holmes ’21+).

$$-2 + 3K_\tau(\sigma,\sigma') + 4\left(\frac{1-K_\tau(\sigma,\sigma')}{2}\right)^{\frac{3}{2}} \le A(\sigma)^T A(\sigma') + O(n^{-1}) \le 2 + 3K_\tau(\sigma,\sigma') - 4\left(\frac{1+K_\tau(\sigma,\sigma')}{2}\right)^{\frac{3}{2}}$$

Corollary. If $A(\sigma)^T A(\sigma') = o(1)$, then $|K_\tau(\sigma,\sigma')| \le \frac{1}{2} + o(1)$.

Said another way: A permutation near the geometric equator of the permutohedron has between ¼ and ¾ of the $\binom{n}{2}$ possible inversions.

Is this tight?



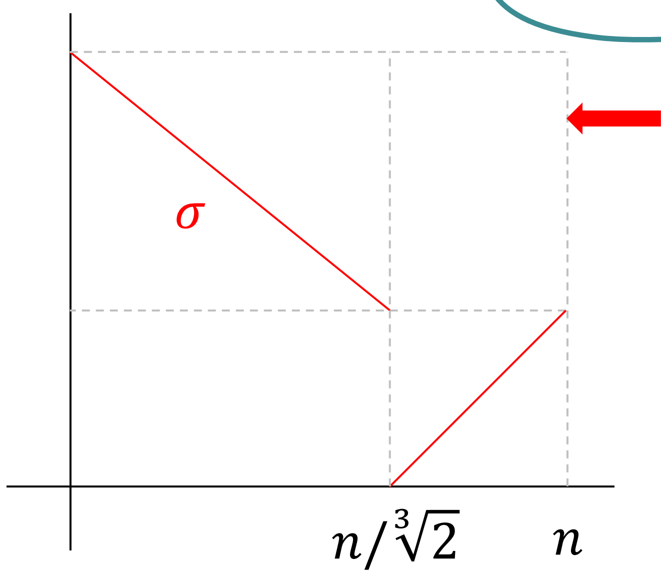
Theorem (Mitchell-C-Frank-Holmes ’21+).

$$-2 + 3K_\tau(\sigma,\sigma') + 4\left(\frac{1-K_\tau(\sigma,\sigma')}{2}\right)^{\frac{3}{2}} \le A(\sigma)^T A(\sigma') + O(n^{-1}) \le 2 + 3K_\tau(\sigma,\sigma') - 4\left(\frac{1+K_\tau(\sigma,\sigma')}{2}\right)^{\frac{3}{2}}$$

Corollary. If $A(\sigma)^T A(\sigma') = o(1)$, then $|K_\tau(\sigma,\sigma')| \le \frac{1}{2} + o(1)$.

Said another way: A permutation near the geometric equator of the permutohedron has between ¼ and ¾ of the $\binom{n}{2}$ possible inversions.

Is this tight?



This permutation is on the geometric equator

$$\mathrm{inv}(\sigma) \approx n^2 2^{-\frac{5}{3}}$$
$$\approx 63\% \cdot \binom{n}{2} < 75\% \binom{n}{2}$$
$$K_\tau(\sigma,\mathrm{id}) \approx -.26 > -1/2$$

감사합니다
dziękuję
תודה
شكرا
Grazie
Danke
ευχαριστώ
Dalu
Obrigado
Thank You
Köszönöm
Tack
Спасибо
Dank
Gracias
谢谢
Merci
Seé
ありがとう