Generalized Distances Between Rankings

Ravi Kumar
Sergei Vassilvitskii

Yahoo! Research

Evaluation

How to evaluate a set of results?

- Use a Metric! NDCG, MAP, ERR, ...

Evaluation

How to evaluate a set of results?

-Use a Metric! NDCG, MAP, MRR, ...

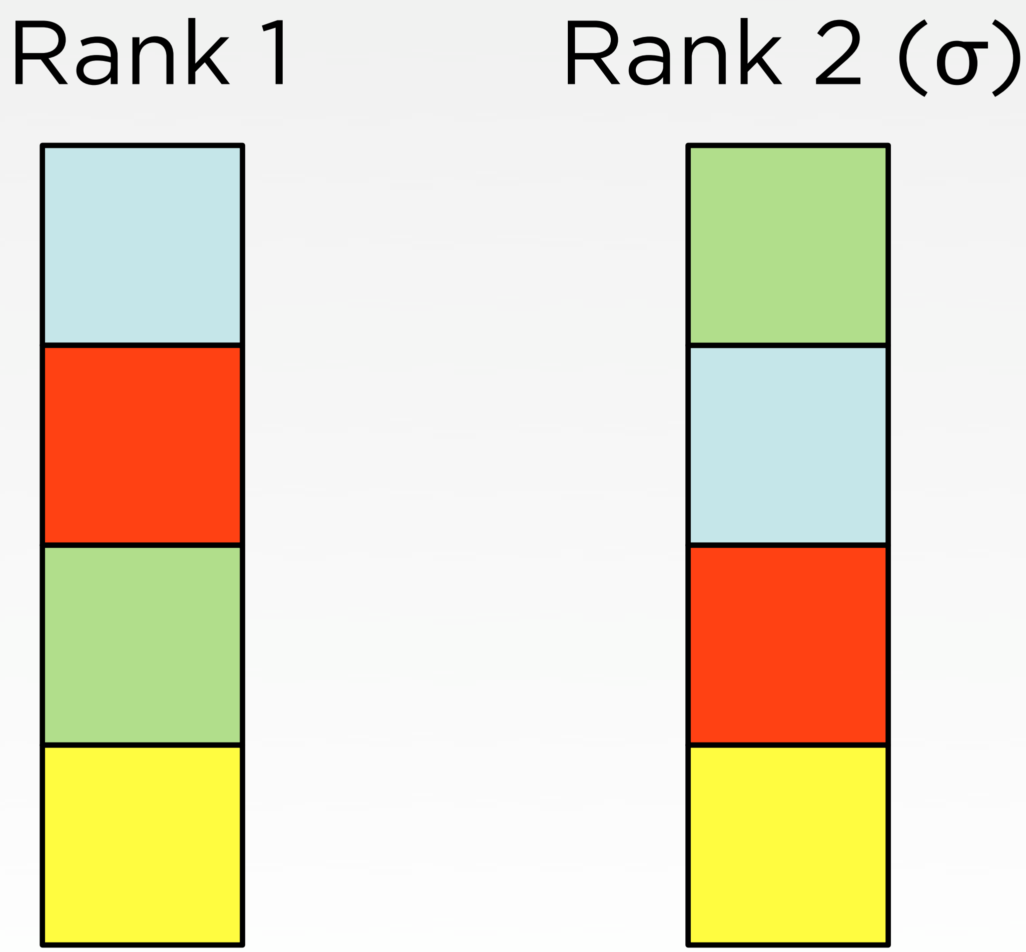
How to evaluate a measure?

1. Incremental improvement
   - Show a problem with current measure
   - Propose a new measure that fixes that (and only that) problem
2. Axiomatic approach
   - Define rules for good measures to follow
   - Find one that follows the rules

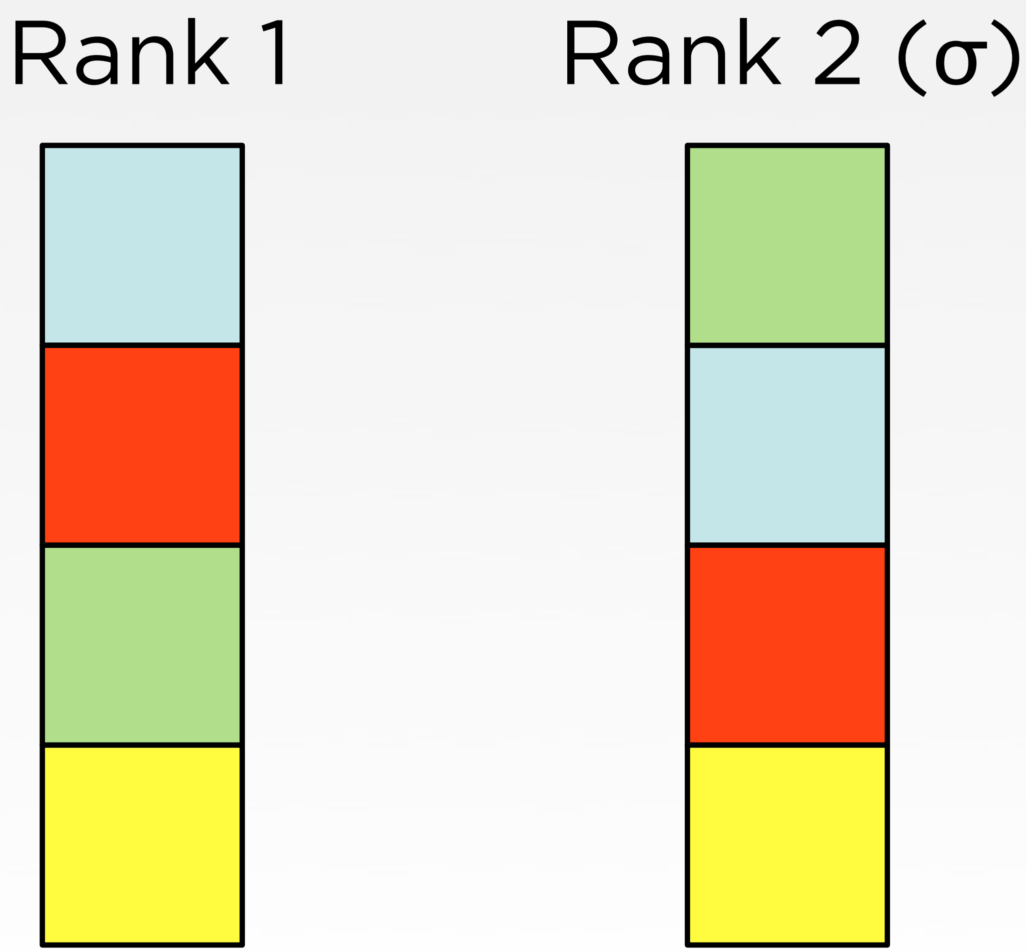
Desired Properties

- *Richness*
  - Support element weights, position weights, etc.
- *Simplicity*
  - Be simple to understand
- *Generalization*
  - Collapse to a natural metric with no weights are present
- *Satisfy Basic Properties*
  - Scale free, invariant under relabeling, triangle inequality…
- *Correlation with other metrics*
  - Should behave similar to other approaches
  - Allows us to select a metric best suited to the problem

Kendall's Tau

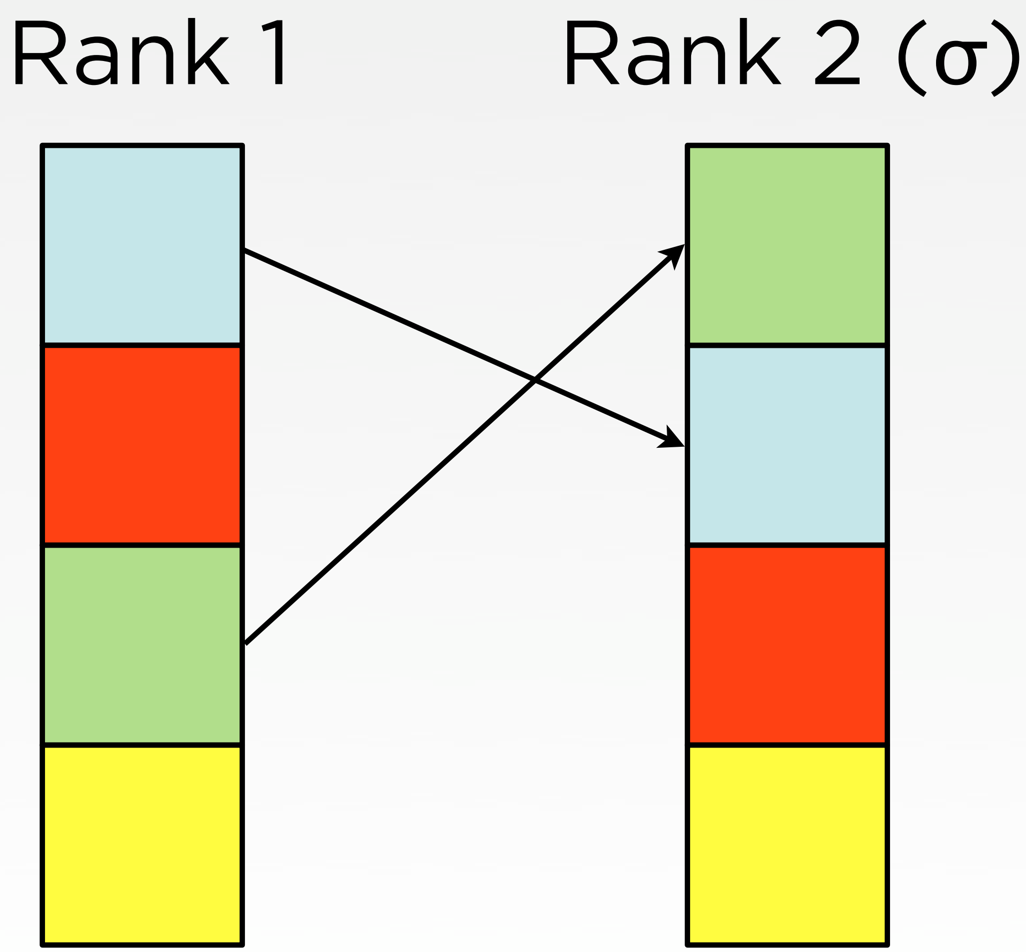


Kendall’s Tau



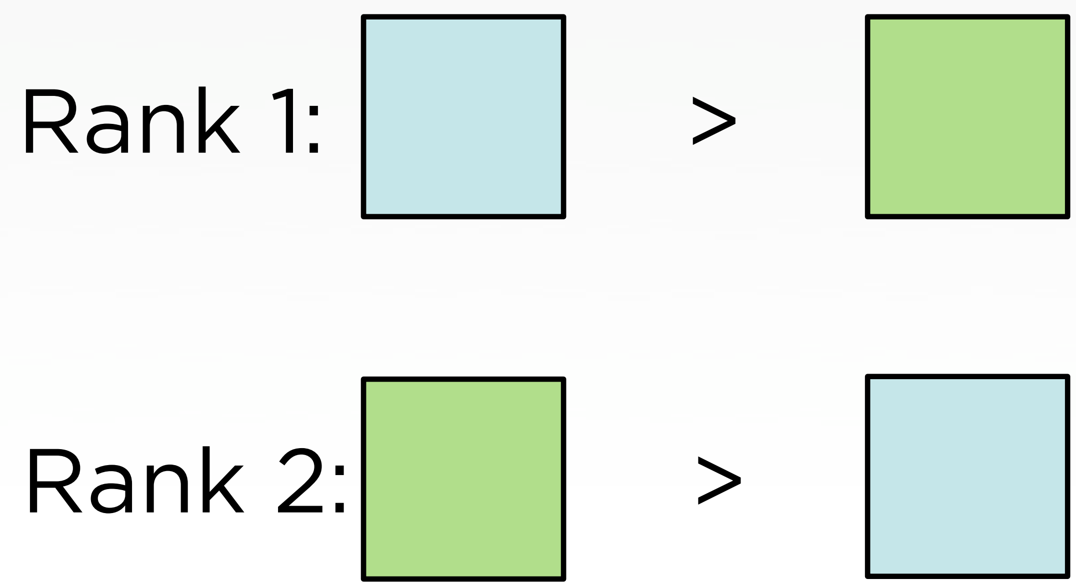
An Inversion: A pair of elements i and j such that $i > j$ and $\sigma(i) < \sigma(j)$.

Kendall's Tau

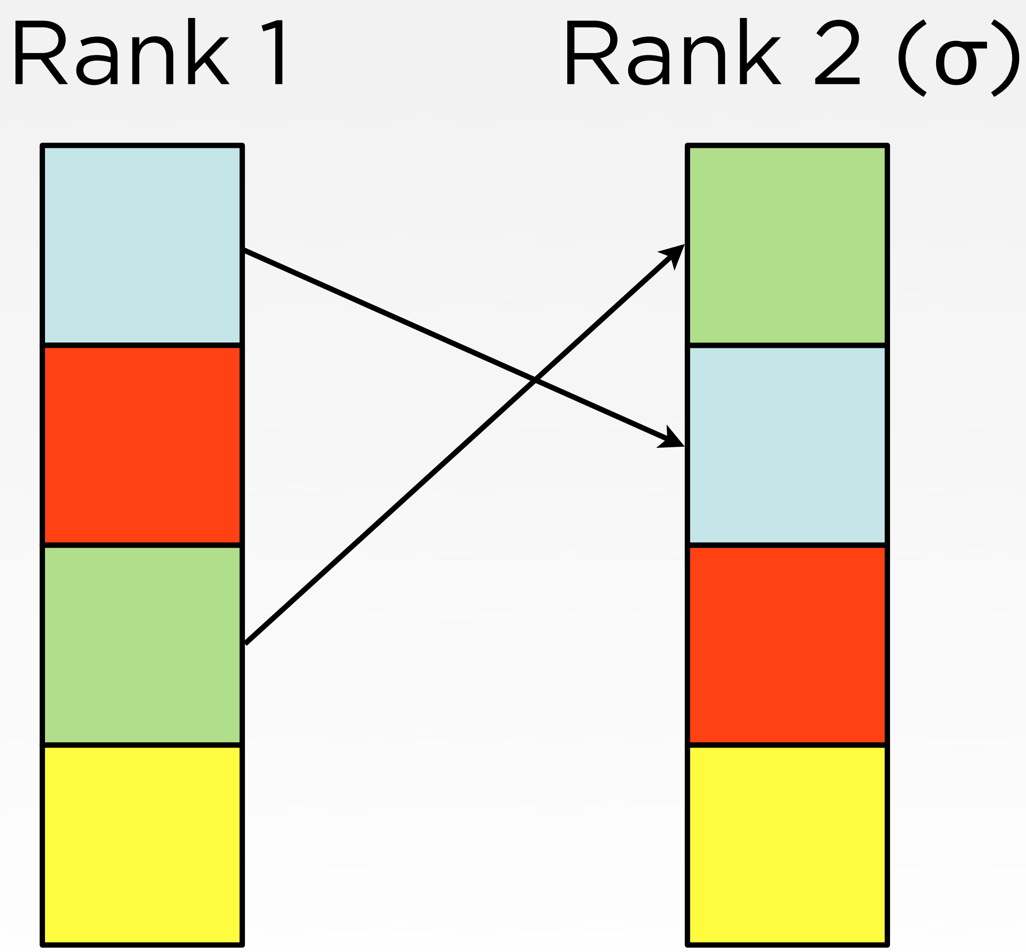


An Inversion: A pair of elements i and j such that $i > j$ and $\sigma(i) < \sigma(j)$.

Example:



Kendall's Tau



An Inversion: A pair of elements i and j such that $i > j$ and $\sigma(i) < \sigma(j)$.

Kendall's Tau:

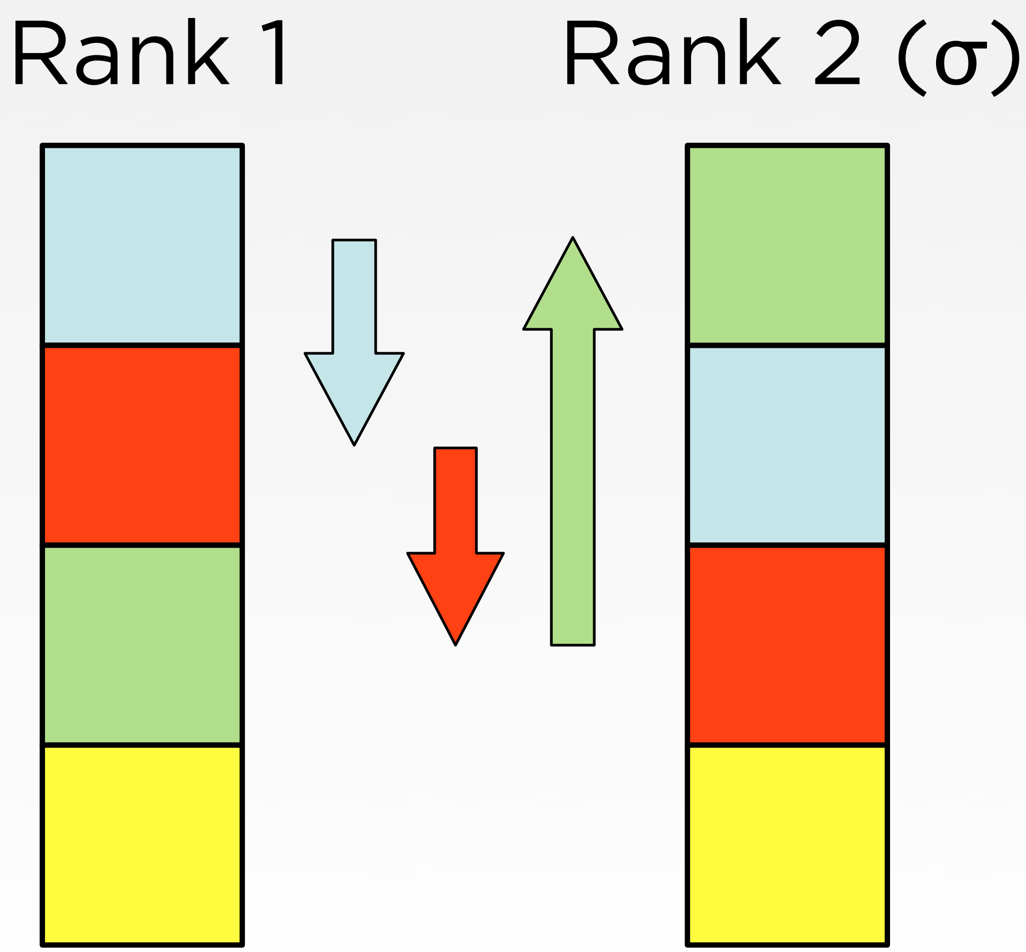
Count total number of inversions in σ.

$$K(\sigma) = \sum_{i<j} \mathbf{1}_{\sigma(i)>\sigma(j)}$$

Example: Inverted pairs: (,) , (,)

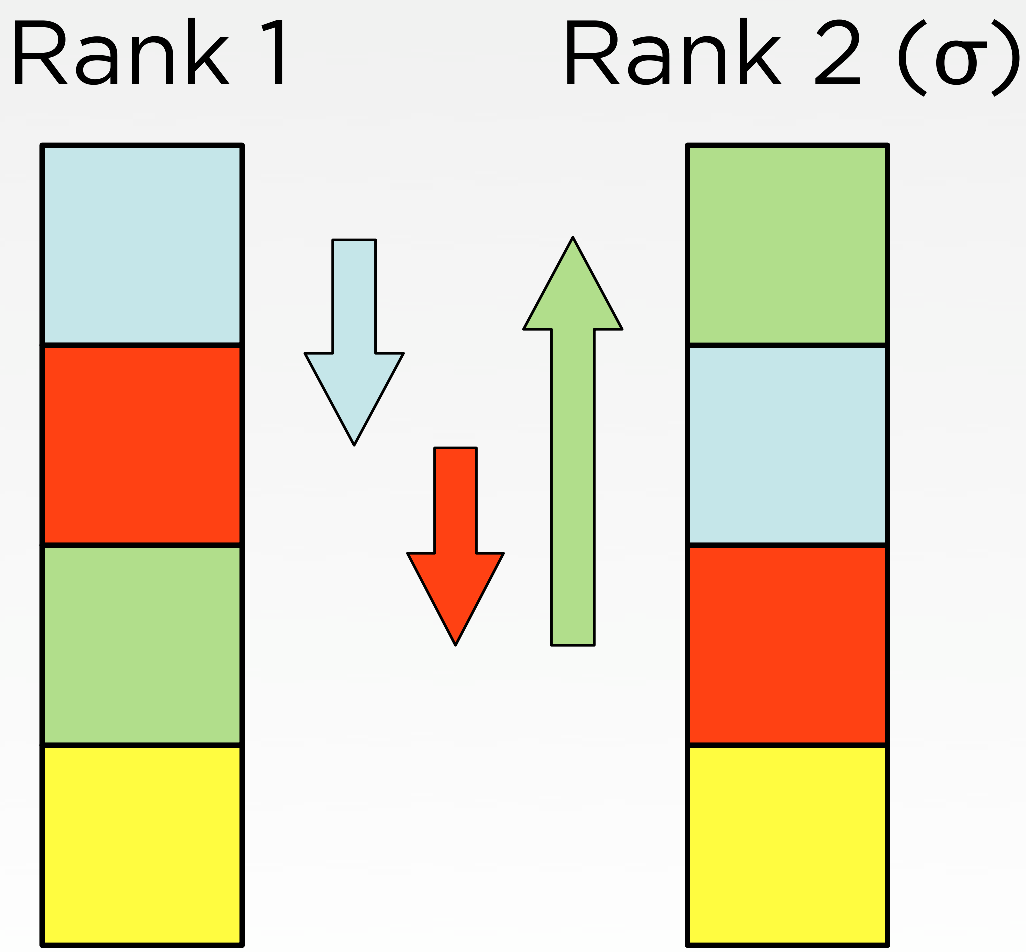
Kendall's Tau: 2

Spearman's Footrule



Displacement: distance an element i moved due to σ = $|i - \sigma(i)|$.

Spearman's Footrule



Displacement: distance an element i moved due to σ = $|i - \sigma(i)|$.

Spearman's Footrule:

Total displacement of all elements:

$$F(\sigma) = \sum_i |i - \sigma(i)|$$

Example: Total Displacement = 1 + 1 + 2 = 4

Kendall vs. Spearman Relationship

Diaconis and Graham proved that the two measures are robust:

$$\forall\sigma \qquad K(\sigma) \leq F(\sigma) \leq 2K(\sigma)$$

Thus the rotation (previous example) is the worst case.

Weighted Versions

How to incorporate weights into the metric?

- Element weights
  - swapping two important elements vs. two inconsequential ones
- Position weights
  - swapping two elements near the head vs. near the tail of the list
- Pairwise similarity weights
  - swapping two similar elements vs. two very different elements

Element Weights

Swap two elements of weight w_i and w_j . How much should the inversion count in the Kendall's tau?

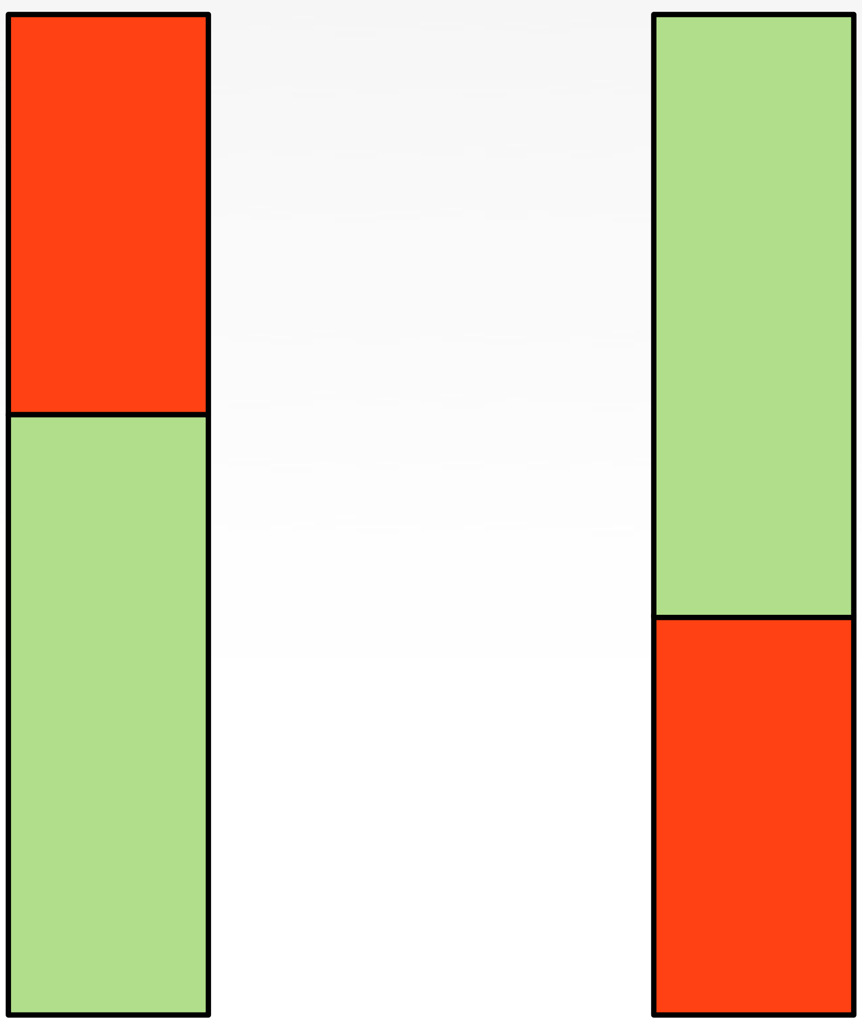
- Average of the weights $\frac{w_i + w_j}{2}$?
- Geometric average of the weights: $\sqrt{w\ w}$?
- Harmonic average of the weights: $\frac{1}{\frac{1}{w} + \frac{1}{w}}$?

- Some other monotonic function of the weights?

Element Weights

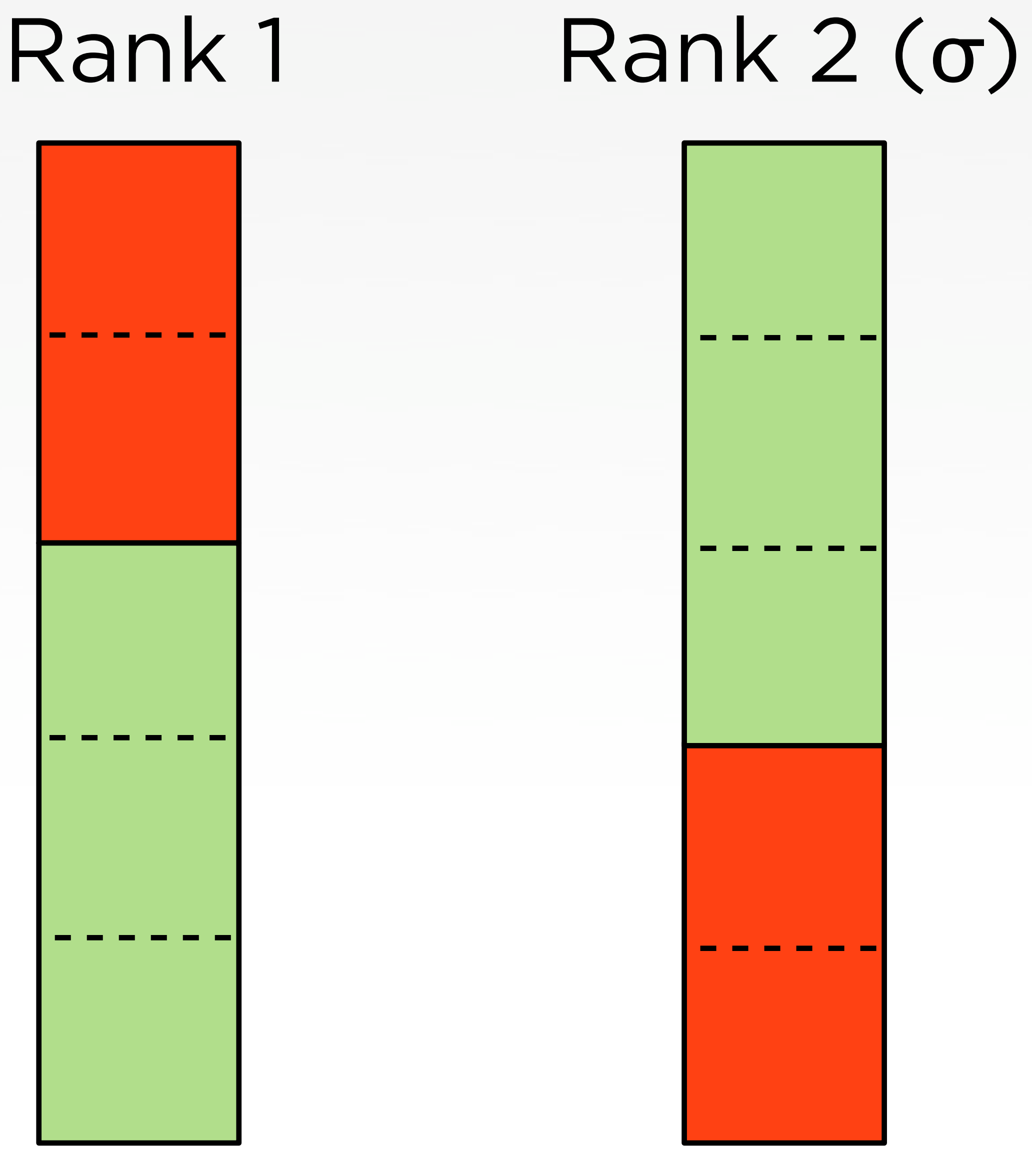
Swap two elements of weight w_i and w_j . How much should the inversion count in the Kendall's tau?

Rank 1 Rank 2 (σ)



Element Weights

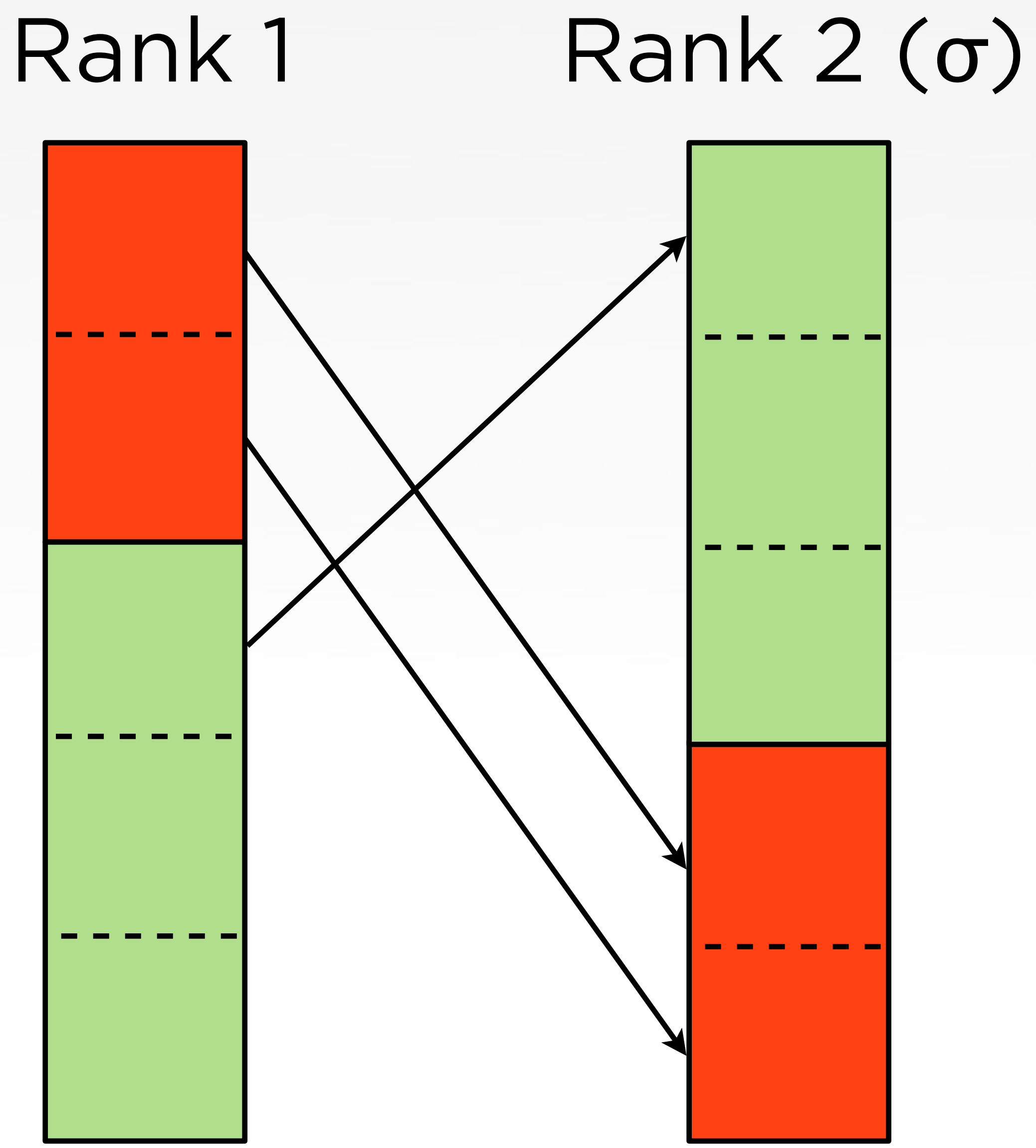
Swap two elements of weight w_i and w_j . How much should the inversion count in the Kendall's tau?



Treat element i as a collection of w_i subelements of weight 1.

Element Weights

Swap two elements of weight w_i and w_j. How much should the inversion count in the Kendall's tau?

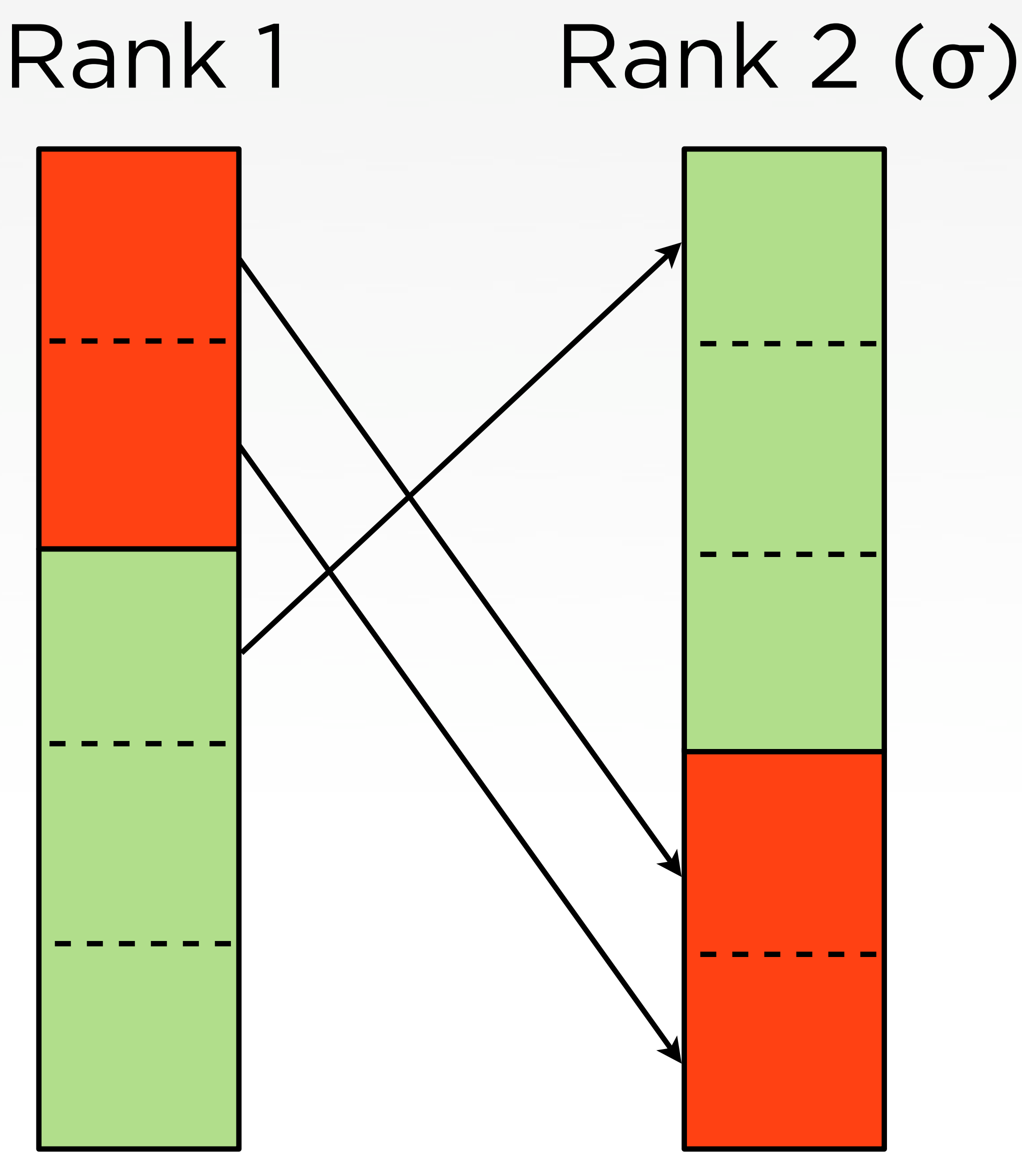


Treat element i as a collection of w_i subelements of weight 1.

The subelements remain in same order

Element Weights

Swap two elements of weight w_i and w_j . How much should the inversion count in the Kendall's tau?



Treat element i as a collection of w_i subelements of weight 1.

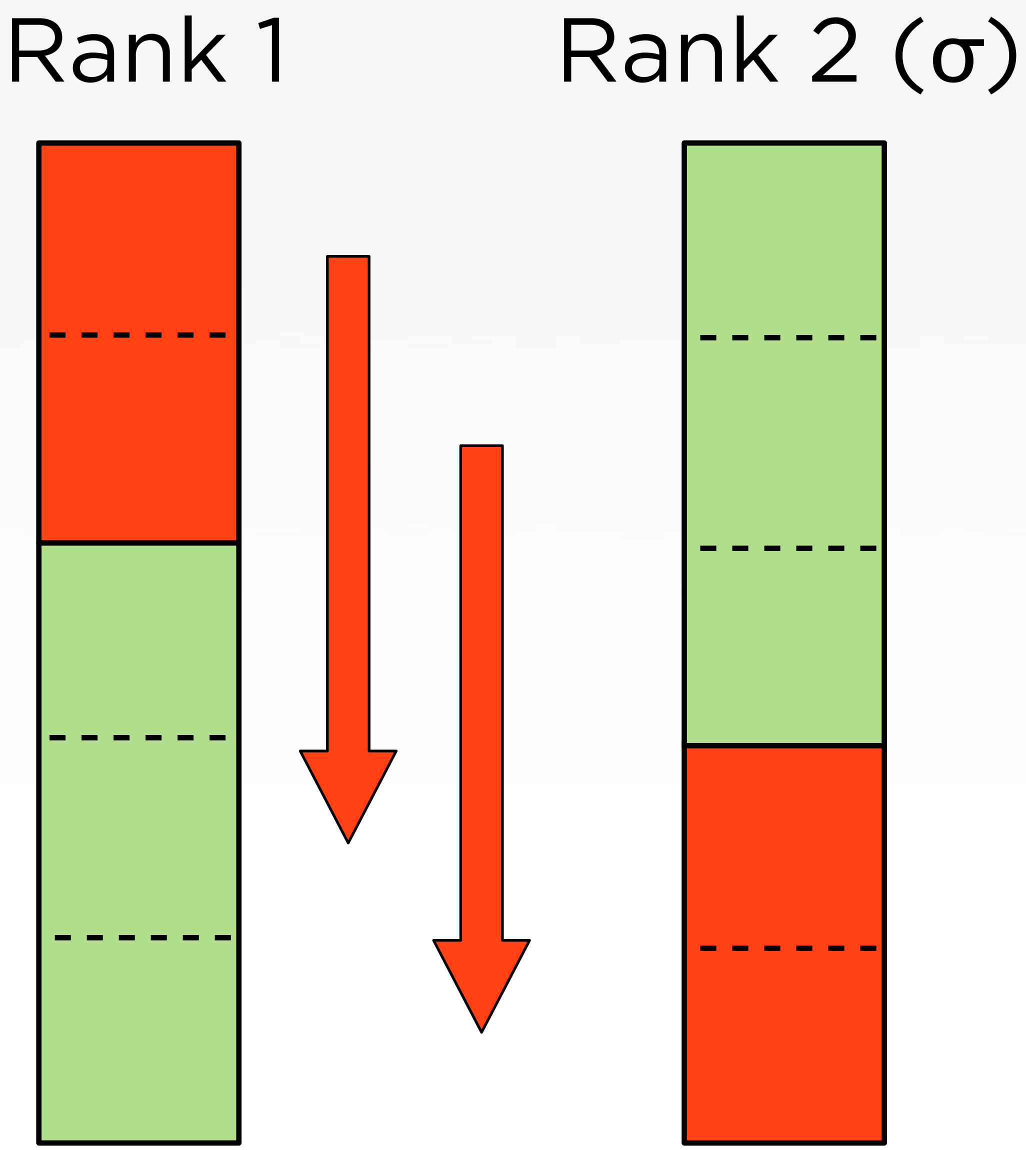
The subelements remain in same order

Then: The total number of inversions between subelements of i and j : $w \ w$

Define: $K_w(\sigma) = \sum_{i<j} w_i w_j \mathbf{1}_{\sigma(i)>\sigma(j)}$

Element Weights

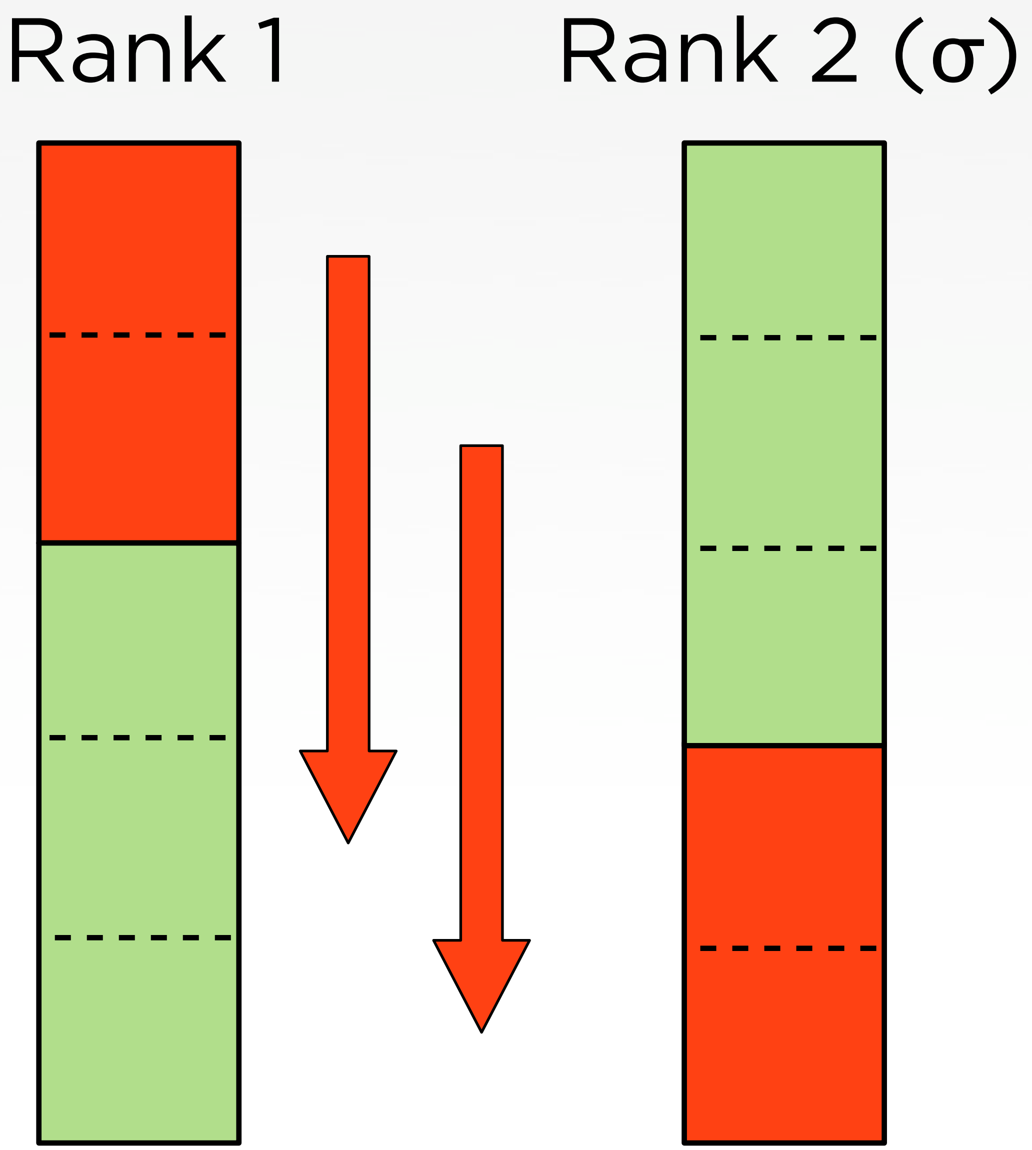
Using the same intuition, how do we define the displacement and the Footrule metric?



Each of the w_i subelements is displaced by: $\left|\sum_{j<i} w_j - \sum_{\sigma(j)<\sigma(i)} w_j\right|$.

Element Weights

Using the same intuition, how do we define the displacement and the Footrule metric?



Each of the w_i subelements is displaced by: $|\sum_{j<i} w_j - \sum_{\sigma(j)<\sigma(i)} w_j|$.

Therefore total displacement for element i: $w_i|\sum_{j<i} w_j - \sum_{\sigma(j)<\sigma(i)} w_j|$.

Weighted Footrule Distance:

$$F_w(\sigma) = \sum_i w_i|\sum_{j<i} w_j - \sum_{\sigma(j)<\sigma(i)} w_j|$$

Kendall vs. Spearman Relationship

The DG Inequality extends to the weighted case:

$$\forall\sigma \qquad K_w(\sigma) \leq F_w(\sigma) \leq 2K_w(\sigma)$$

Rotation remains the worst case example.

Position Weights

How should we differentiate inversions near the head of the list versus those at the tail of the list?

- Let δ_i be the cost of swapping element at position i-1 with one at position i.

- In typical applications: $\delta_2 \geq \delta_3 \geq \ldots \geq \delta_n$

(DCG sets $\delta_i = \frac{1}{\log i} - \frac{1}{\log i + 1}$)

- Let $p_i = \sum_{j=2}^{i} \delta_j$, and $\bar{p}_i(\sigma) = \frac{p_i - p_{\sigma(i)}}{i - \sigma(i)}$ be the average cost of per swap charged to element i.

Position Weights

- Let $p_i = \sum_{j=2}^{i} \delta_j$, and $\bar{p}_i(\sigma) = \frac{p_i - p_{\sigma(i)}}{i - \sigma(i)}$ be the average cost of per swap charged to element i.

We can treat $\bar{p}_i(\sigma)$ as if they were element weights, and define:

Position Weights

- Let $p_i = \sum_{j=2}^{i} \delta_j$, and $\bar{p}_i(\sigma) = \frac{p_i - p_{\sigma(i)}}{i - \sigma(i)}$ be the average cost of per swap charged to element i.

We can treat $\bar{p}_i(\sigma)$ as if they were element weights, and define:

Kendall's Tau: $K_\delta(\sigma) = \sum_{i<j} \bar{p}_i(\sigma)\bar{p}_j(\sigma)\mathbf{1}_{\sigma\ i\ >_\sigma\ j}$

Footrule: $F_\delta(\sigma) = \sum_i \bar{p}_i(\sigma) | \sum_{j<i} \bar{p}_j(\sigma) - \sum_{\sigma(j)<\sigma(i)} \bar{p}_j(\sigma)|$

Conclude:

$$\forall\sigma \qquad K_\delta(\sigma) \leq F_\delta(\sigma) \leq 2K_\delta(\sigma)$$

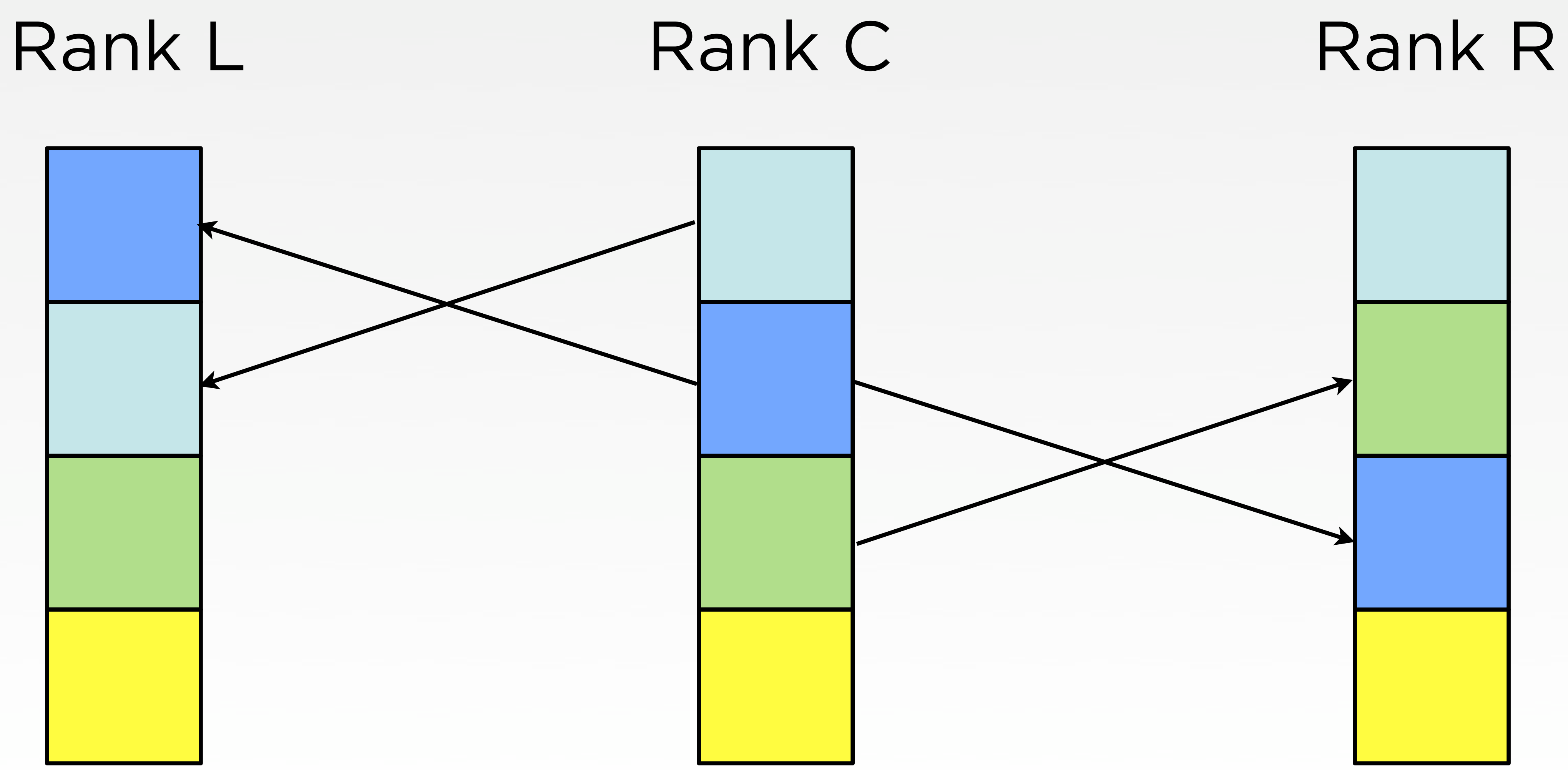
Element Similarities

Element weights: model cost of important versus inconsequential elements.

Position weights model different cost of inversions near the head or tail of list

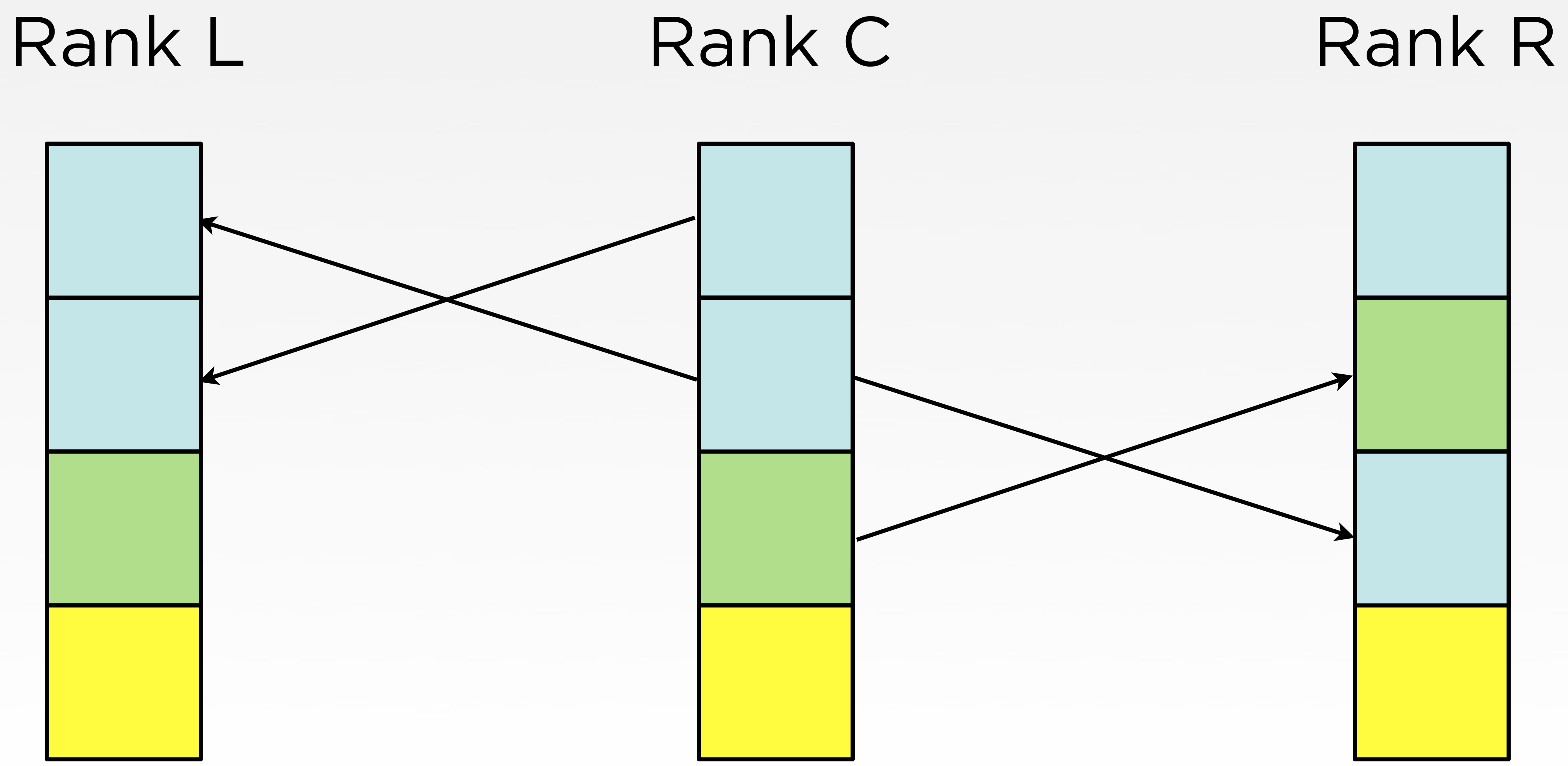
How to model the cost of swap similar elements versus different elements.

Element similarities



With identical element and position weights is L or R better?

Element similarities



With identical element and position weights is L or R better?

In the extreme case L and C are identical, even though an inversion occurred

Modeling Similarities

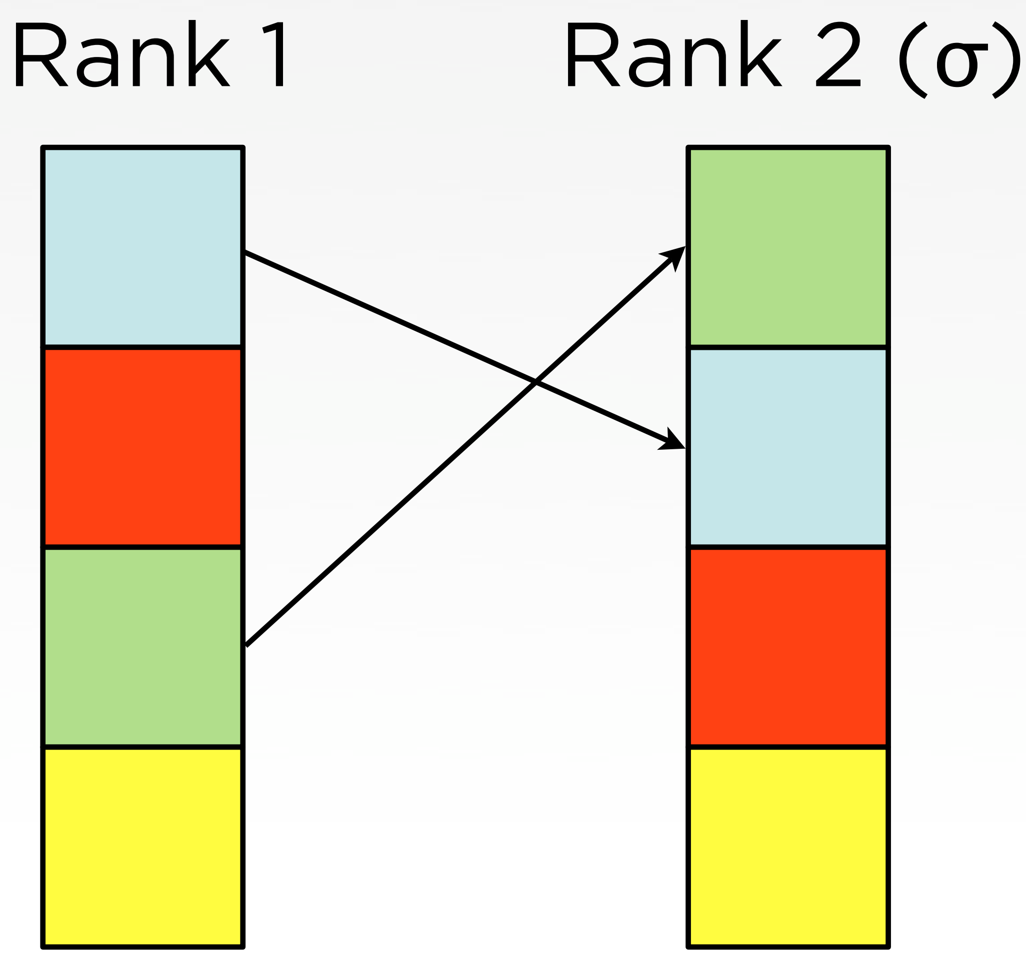
For two elements i and j let D_{ij} denote the distance between them.

We assume that $D : [n] \times [n]$ forms a metric (follows triangle inequality).

Modeling Similarities

For two elements i and j let D_{ij} denote the distance between them.

We assume that $D : [n] \times [n]$ forms a metric (follows triangle inequality).

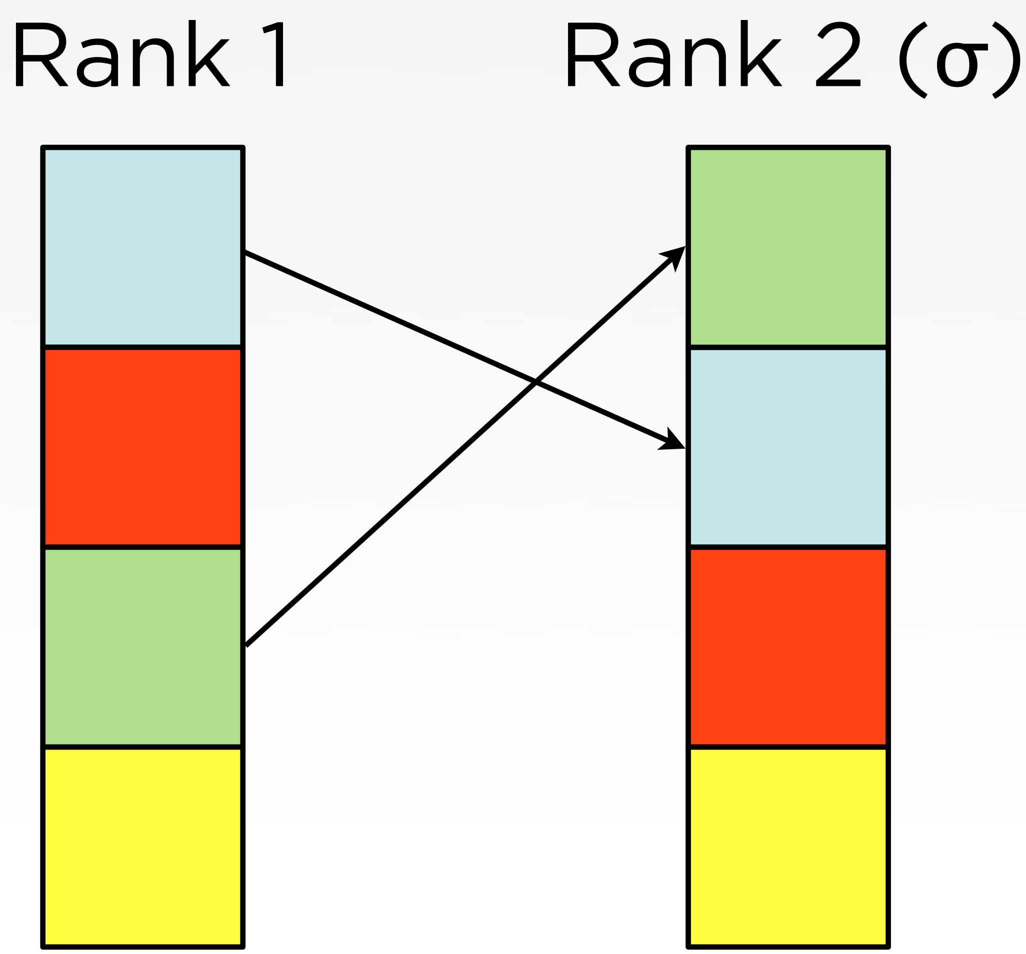


To define Kendall's Tau: scale each inversion by the distance between the inverted elements.

Modeling Similarities

For two elements i and j let D_{ij} denote the distance between them.

We assume that $D : [n] \times [n]$ forms a metric (follows triangle inequality).



To define Kendall's Tau: scale each inversion by the distance between the inverted elements.

In the example:

K(σ) = D(□ ,□) + D(□ ,□)

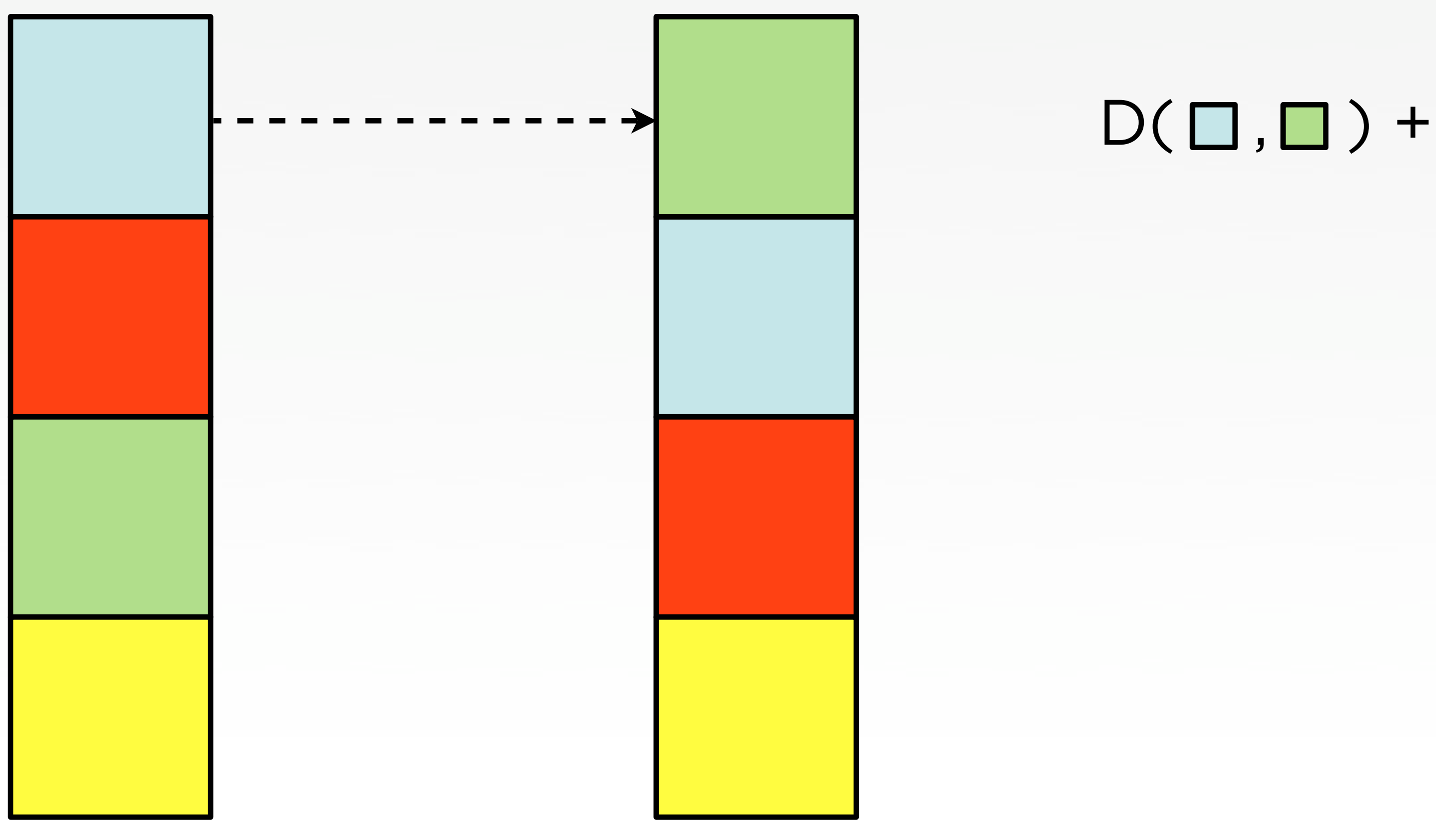
Generally:

$$K_D(\sigma) = \sum_{i<j} D_{ij} \mathbf{1}_{\sigma(i)>\sigma(j)}$$

Footrule with similarities

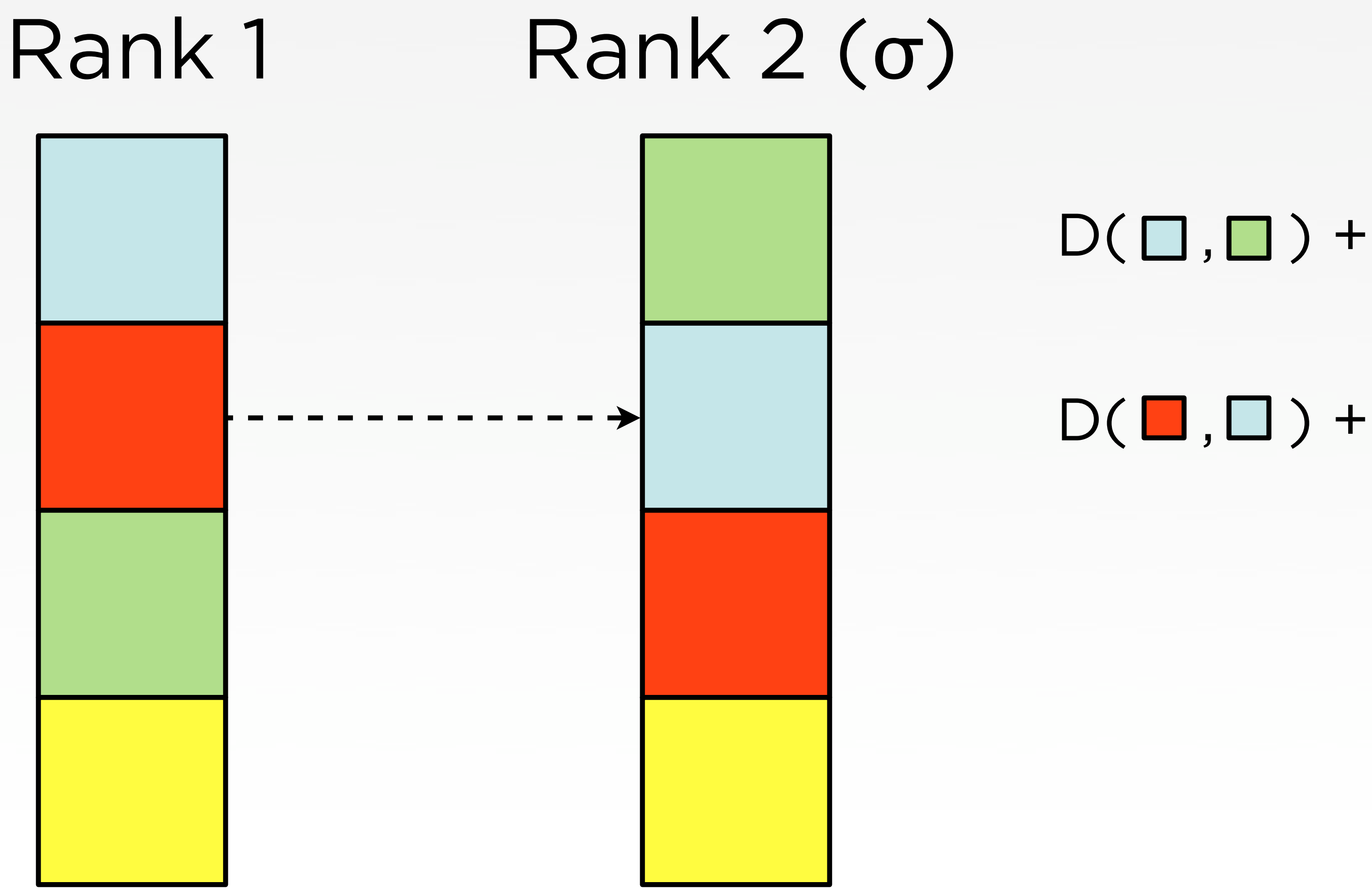
Defining Footrule with similarities

Rank 1 Rank 2 (σ)



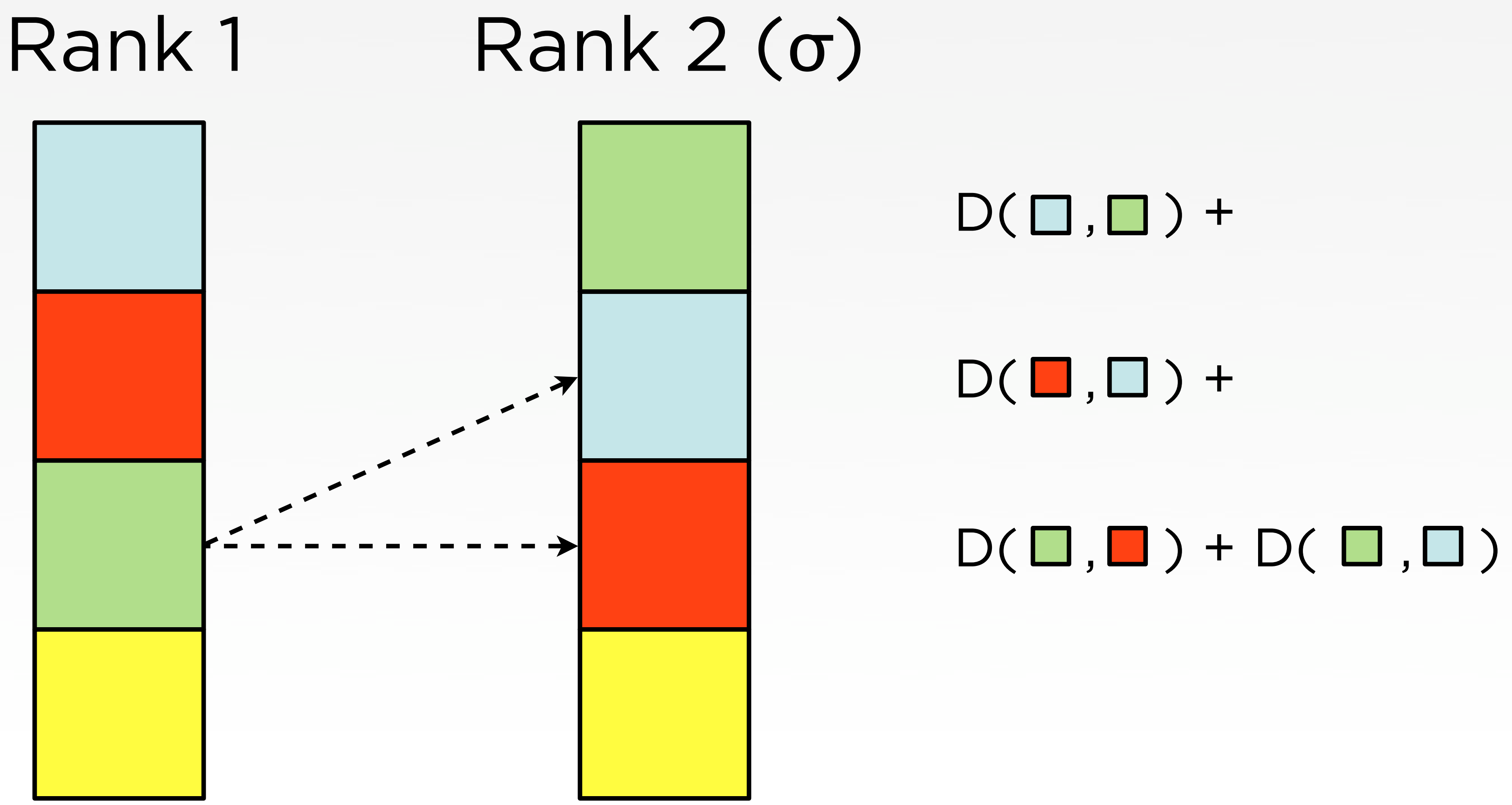
Footrule with similarities

Defining Footrule with similarities



Footrule with similarities

Defining Footrule with similarities



Formally: $F'_D(\sigma) = \sum_i \left| \sum_{j<i} D_{ij} - \sum_{\sigma(j)<\sigma(i)} D_{ij} \right|$

Kendall vs. Spearman Relationship

The DG Inequality extends to this case as well:

$$\forall\sigma \qquad \frac{1}{3}K_D(\sigma) \leq F_D(\sigma) \leq 3K_D(\sigma)$$

There are examples where:

$$F_D(\sigma) = 3K_D(\sigma)$$

We conjecture that:

$$K_D(\sigma) \leq F_D(\sigma)$$

Combining All Weights

We can combine element, position and similarity weights all into :

$$K^* = \sum_{i<j} w_i w_j \bar{p}_i \bar{p}_j D_{ij} \mathbf{1}_{\sigma(i)>\sigma(j)}$$

and

$$F^*(\sigma) = \sum_i w_i \bar{p}_i \left| \sum_{j<i} w_j \bar{p}_j D_{ij} - \sum_{\sigma(j)<\sigma(i)} w_j \bar{p}_j D_{ij} \right|$$

Evaluation

Evaluation of K^* and F^* :

Richness:

Captures element, position weights, element similarities

Simplicity:

you decide

Generalization:

If all weights are 1 collapse to classical K and F.

Basic Properties:

Scale free, right invariant, satisfy triangle inequality.

Correlation:

Always within a factor of 3 of each other.

More on Robustness

Rank Aggregation: Given a set of rankings, find one that best summarizes them.

Using K the problem is NP-hard

Using F the problem has a simple solution

Alternatively:

Using F* the problem appears daunting

Using K* the problem has a simple approximation algorithm

Knowing that F and K (as well as F* and K*) are close to each other allows us to select the easiest metric to work with.

Evaluating Robustness

Dataset:

A set of clicks on 80,000 Y! search queries from 09/2009.

Each query with at least 1000 total clicks

Rank 1: Yahoo! Search order

Rank 2: Order by the number of clicks at each position

Evaluating Robustness

Dataset:

A set of clicks on 80,000 Y! search queries from 09/2009.

Each query with at least 1000 total clicks

Rank 1: Yahoo! Search order

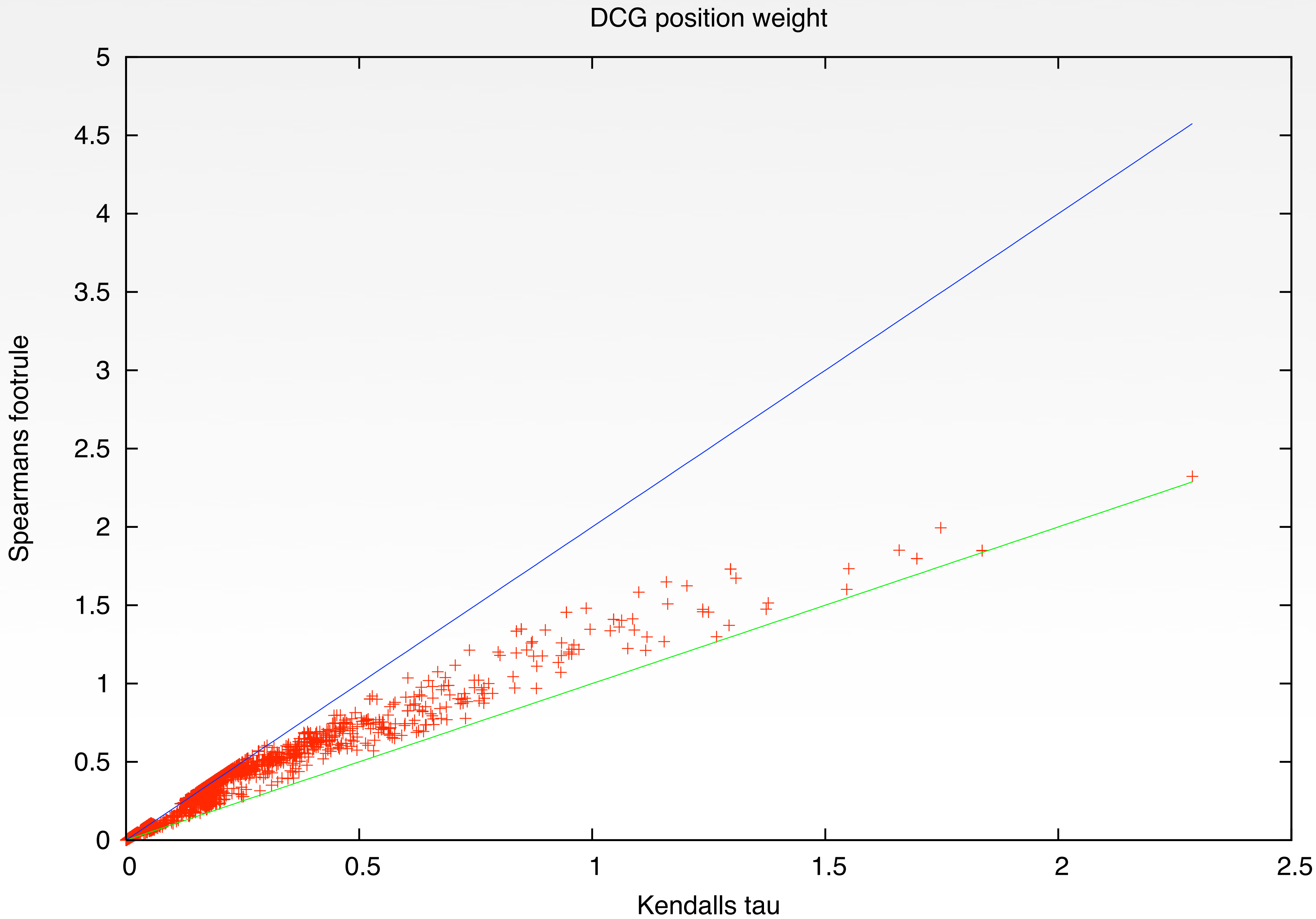
Rank 2: Order by the number of clicks at each position

Element weights set arbitrarily to 1

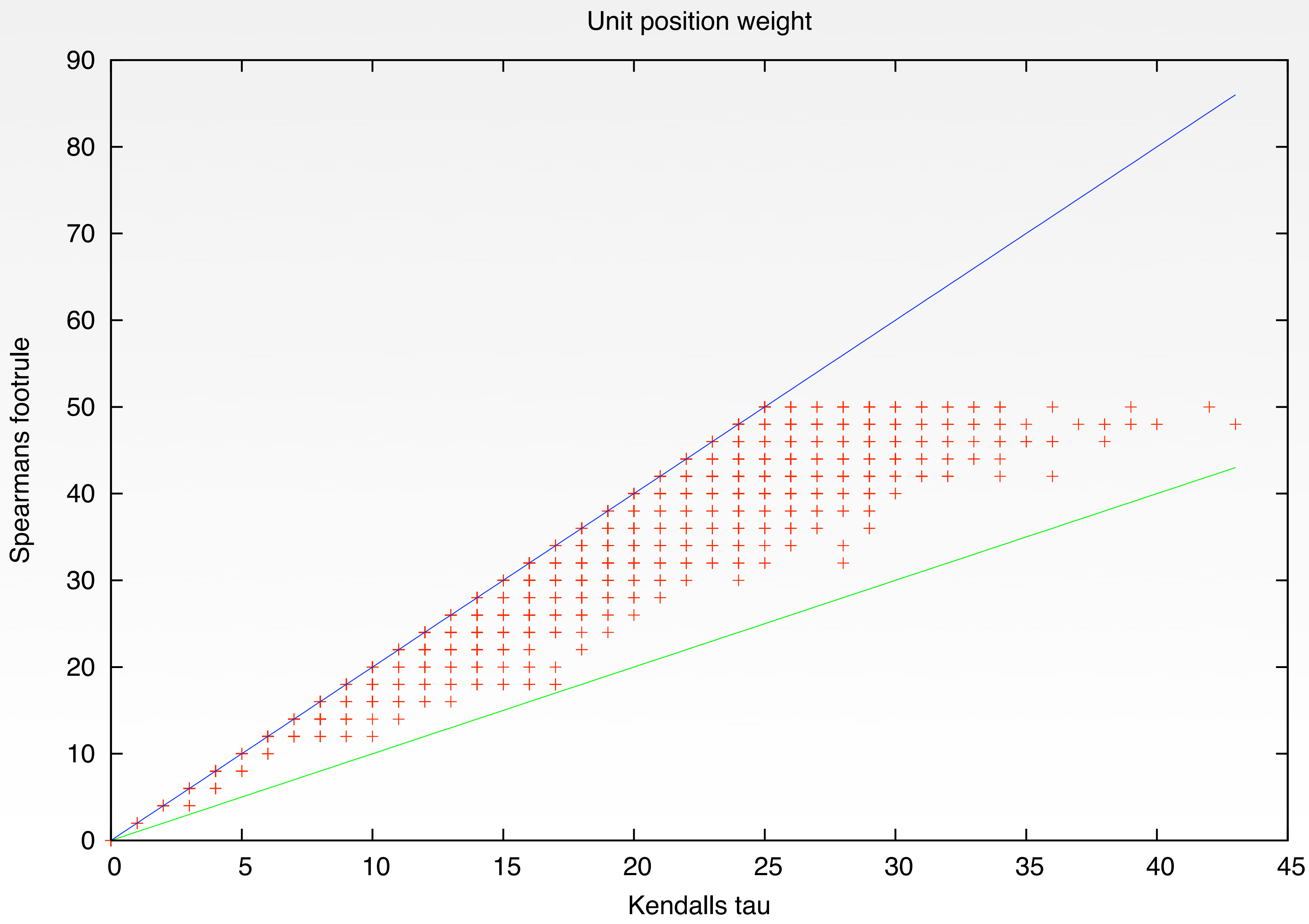
Position weights set:

- DCG: $\delta_i = \frac{1}{\log i} - \frac{1}{\log i + 1}$
- UNIT: $\delta_i = 1$

Evaluating Robustness (DCG)



Evaluating Robustness (Unit)



Conclusion

What makes a good metric?

Categorized the different kinds of weights:

- Element weights
- Position weights
- Similarity weights

Introduced new K* and F* measures and showed near-equivalence

Open Questions:

Express: MAP, ERR, NDCG, others in this framework

Thank You

{sergei, ravikumar} @ yahoo-inc.com