Rank Aggregation and Kemeny Voting

Rolf Niedermeier
FG Algorithmics and Complexity Theory
Institut für Softwaretechnik und Theoretische Informatik
Fakultät IV
TU Berlin
Germany

MAIN SOURCES OF THIS TALK


EXAMPLE: SELECT A PLACE FOR PHD STUDY

Choose between the following places:

- TU Berlin (B),
- MIT (M),
- Oxford University (O),
- Tsinghua University (T),
- ETH Zurich (Z).

Selection based on various criteria, leading to different rankings:
**Example: Select a Place for PhD Study**

Choose between the following places:

- TU Berlin (B),
- MIT (M),
- Oxford University (O),
- Tsinghua University (T),
- ETH Zurich (Z).

Selection based on various criteria, leading to different rankings:

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameterized Complexity</td>
<td>$B \succ O \succ M \succ T \succ Z$</td>
</tr>
<tr>
<td>Salary</td>
<td>$Z \succ O \succ M \succ T \succ B$</td>
</tr>
<tr>
<td>Practicing English</td>
<td>$M \succ O \succ B \succ Z \succ T$</td>
</tr>
<tr>
<td>Cultural activities</td>
<td>$B \succ T \succ Z \succ M \succ O$</td>
</tr>
</tbody>
</table>

**Goal:** Aggregate the given rankings (that is, permutations) into a median ranking.
### Pairwise Comparisons and Voting

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameterized Complexity</td>
<td>B ≻ O ≻ M ≻ T ≻ Z</td>
</tr>
<tr>
<td>Salary</td>
<td>Z ≻ O ≻ M ≻ T ≻ B</td>
</tr>
<tr>
<td>Practicing English</td>
<td>M ≻ O ≻ B ≻ Z ≻ T</td>
</tr>
<tr>
<td>Cultural activities</td>
<td>B ≻ T ≻ Z ≻ M ≻ O</td>
</tr>
</tbody>
</table>

**Condorcet and Kemeny:**

- **Condorcet Winner:** A candidate who wins against all other candidates in pairwise comparisons. A Condorcet winner does not always exist, but is unique if it exists!
### Pairwise Comparisons and Voting

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Parameterized Complexity</th>
<th>Salary</th>
<th>Practicing English</th>
<th>Cultural activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameterized Complexity</td>
<td>B ≻ O ≻ M ≻ T ≻ Z</td>
<td>Z ≻ O ≻ M ≻ T ≻ B</td>
<td>M ≻ O ≻ B ≻ Z ≻ T</td>
<td>B ≻ T ≻ Z ≻ M ≻ O</td>
</tr>
</tbody>
</table>

**Condorcet and Kemeny:**

- **Condorcet Winner:** A candidate who wins against all other candidates in pairwise comparisons. A Condorcet winner does not always exist, but is unique if it exists!

- **Kemeny:** Determine consensus ranking that minimizes the total sum of the number of “inversions” to the given rankings...

Always yields a Condorcet winner if it exists.
## On Condorcet Winner Determination

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameterized Complexity</td>
<td>B ≻ O ≻ M ≻ T ≻ Z</td>
</tr>
<tr>
<td>Salary</td>
<td>Z ≻ O ≻ M ≻ T ≻ B</td>
</tr>
<tr>
<td>Practicing English</td>
<td>M ≻ O ≻ B ≻ Z ≻ T</td>
</tr>
<tr>
<td>Cultural activities</td>
<td>B ≻ T ≻ Z ≻ M ≻ O</td>
</tr>
</tbody>
</table>
## On Condorcet Winner Determination

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameterized Complexity</td>
<td>B ≻ O ≻ M ≻ T ≻ Z</td>
</tr>
<tr>
<td>Salary</td>
<td>Z ≻ O ≻ M ≻ T ≻ B</td>
</tr>
<tr>
<td>Practicing English</td>
<td>M ≻ O ≻ B ≻ Z ≻ T</td>
</tr>
<tr>
<td>Cultural activities</td>
<td>B ≻ T ≻ Z ≻ M ≻ O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pairs of candidates</th>
<th># votes: x ≻ y</th>
<th># votes: y ≻ x</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x, y) = (B, O)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(x, y) = (B, M)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(x, y) = (B, T)</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>(x, y) = (B, Z)</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>(x, y) = (O, M)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(x, y) = (O, T)</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>(x, y) = (O, Z)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(x, y) = (M, T)</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>(x, y) = (M, Z)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(x, y) = (T, Z)</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Marie Jean Antoine Nicolas Caritat, Marquis de Condorcet 1743-1794
## On Condorcet Winner Determination

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Ranking</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameterized Complexity</td>
<td>B ≻ O ≻ M ≻ T ≻ Z</td>
<td></td>
</tr>
<tr>
<td>Salary</td>
<td>Z ≻ O ≻ M ≻ T ≻ B</td>
<td></td>
</tr>
<tr>
<td>Practicing English</td>
<td>M ≻ O ≻ B ≻ Z ≻ T</td>
<td></td>
</tr>
<tr>
<td>Cultural activities</td>
<td>B ≻ T ≻ Z ≻ M ≻ O</td>
<td></td>
</tr>
</tbody>
</table>

### Pairs of candidates

<table>
<thead>
<tr>
<th>Pairs of candidates</th>
<th># votes: $x ≻ y$</th>
<th># votes: $y ≻ x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(x, y) = (B, O)$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$(x, y) = (B, M)$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$(x, y) = (B, T)$</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$(x, y) = (B, Z)$</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$(x, y) = (O, M)$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$(x, y) = (O, T)$</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$(x, y) = (O, Z)$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$(x, y) = (M, T)$</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$(x, y) = (M, Z)$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$(x, y) = (T, Z)$</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

No Condorcet winner!
## Winner Determination in Kemeny Voting

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameterized Complexity</td>
<td>B $\succ$ O $\succ$ M $\succ$ T $\succ$ Z</td>
</tr>
<tr>
<td>Salary</td>
<td>Z $\succ$ O $\succ$ M $\succ$ T $\succ$ B</td>
</tr>
<tr>
<td>Practicing English</td>
<td>M $\succ$ O $\succ$ B $\succ$ Z $\succ$ T</td>
</tr>
<tr>
<td>Cultural activities</td>
<td>B $\succ$ T $\succ$ Z $\succ$ M $\succ$ O</td>
</tr>
</tbody>
</table>

Determine consensus ranking that minimizes the total sum of the number of inversions to the given rankings...

John George
## Winner Determination in Kemeny Voting

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameterized Complexity</td>
<td>B ≻ O ≻ M ≻ T ≻ Z</td>
</tr>
<tr>
<td>Salary</td>
<td>Z ≻ O ≻ M ≻ T ≻ B</td>
</tr>
<tr>
<td>Practicing English</td>
<td>M ≻ O ≻ B ≻ Z ≻ T</td>
</tr>
<tr>
<td>Cultural activities</td>
<td>B ≻ T ≻ Z ≻ M ≻ O</td>
</tr>
</tbody>
</table>

Determine consensus ranking that minimizes the total sum of the number of inversions to the given rankings...

⇝ Two (out of 18) optimal consensus ranking with "score" 16:

- $B ≻ O ≻ M ≻ Z ≻ T$
- $O ≻ M ≻ B ≻ T ≻ Z$

---

Kemeny Score: KT-Distance

Kendall Tau distance (between two votes \( v \) and \( w \))

\[
\text{KT-dist}(v, w) = \sum_{\{c, d\} \subseteq C} d_{v,w}(c, d),
\]

where \( d_{v,w}(c, d) = \begin{cases} 
0 & \text{if } v \text{ and } w \text{ rank } c \text{ and } d \text{ in the same order}, \\
1 & \text{otherwise.} 
\end{cases} \)

Example:

\( v: a > b > c \)
\( w: b > c > a \)

\[
\text{KT-dist}(v, w) = d_{v,w}(a, b) + d_{v,w}(a, c) + d_{v,w}(b, c) \\
= 1 + 1 + 0 \\
= 2
\]
Kemeny Score (Rank Aggregation):

**Input:** An set of rankings over the same candidate set and a positive integer \( k \).

**Question:** Is there a ranking \( r \) with Kemeny score at most \( k \), that is, the sum of KT-distances of \( r \) to all input rankings is at most \( k \)?
Central Problem: Rank Aggregation

Kemeny Score (Rank Aggregation):

Input: An set of rankings over the same candidate set and a positive integer $k$.

Question: Is there a ranking $r$ with Kemeny score at most $k$, that is, the sum of KT-distances of $r$ to all input rankings is at most $k$?

Applications:

- Ranking of web sites (using meta search engines)
- Sport competitions
- Databases
- Bioinformatics
Some Results for Kemeny Score

Complexity:

• NP-complete (even for four votes)
  Bartholdi, Tovey and Tick, Social Choice and Welfare 1989,
  Dwork, Kumar, Naor, and Sivakumar, WWW 2001

Algorithms:

• factor 8/5-approximation, randomized: factor 11/7
  van Zuylen and Williamson, WAOA 2007,
  Ailon, Charikar, and Newman, JACM 2008

• PTAS
  Kenyon-Mathieu and Schudy, STOC 2007

• exact algorithms, heuristics, branch and bound, and experiments
  Davenport and Kalagnanam, AAAI 2004,
  Conitzer, Davenport, and Kalagnanam, AAAI 2006,
  Schalekamp and van Zuylen, ALENEX 2009,
  Ali and Meilä, Mathematical Social Sciences, 2012
THE LEITMOTIF OF PARAMETERIZED ALGORITHMICS

Formally: “Two-dimensional analysis of complexity”:

NP-hard problem $X$: Input size $n$ and problem parameter $k$.

If there is an algorithm solving $X$ in time

$$f(k) \cdot n^{O(1)}$$

then $X$ is called fixed-parameter tractable (FPT):
**THE LEITMOTIF OF PARAMETERIZED ALGORITHMICs**

Formally: “Two-dimensional analysis of complexity”:

NP-hard problem $X$: Input size $n$ and problem parameter $k$.

If there is an algorithm solving $X$ in time

$$f(k) \cdot n^{O(1)},$$

then $X$ is called **fixed-parameter tractable (FPT)**:
Completeness program developed by Downey and Fellows (1999).

Presumably fixed-parameter intractable

\[
\text{FPT} \subseteq \text{W[1]} \subseteq \text{W[2]} \subseteq \ldots \subseteq \text{W[P]} \subseteq \text{XP}
\]
Completeness program developed by Downey and Fellows (1999).

Presumably fixed-parameter intractable

\[ \text{FPT} \subseteq \text{W[1]} \subseteq \text{W[2]} \subseteq \ldots \subseteq \text{W[P]} \subseteq \text{XP} \]

"Function battle" concerning allowed running time:

\[ \text{FPT: } f(k) \cdot n^{O(1)} \quad \text{vs} \quad \text{XP: } f(k) \cdot n^{g(k)} \]
Completeness program developed by Downey and Fellows (1999).

Presumably fixed-parameter intractable


“Function battle” concerning allowed running time:

\[
\text{FPT: } f(k) \cdot n^{O(1)} \quad \text{vs} \quad \text{XP: } f(k) \cdot n^{g(k)}
\]

Assumption: FPT \neq W[1]

For instance, if W[1]=FPT then 3-SAT for a Boolean formula \( F \) with \( n \) variables can be solved in \( 2^{o(n)} \cdot |F|^{O(1)} \) time.
# Parameterized Complexity of Kemeny Score

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Compl.</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of votes $n$</td>
<td>NP-c</td>
<td>1 for $n = 4$</td>
</tr>
<tr>
<td>Number of candidates $m$</td>
<td>FPT</td>
<td>$2O^*(2^m)$</td>
</tr>
<tr>
<td>Kemeny score $k$</td>
<td>FPT</td>
<td>$3O^*(2^{O(\sqrt{k})})$</td>
</tr>
<tr>
<td>Max. range of cand. pos. $r_m$</td>
<td>FPT</td>
<td>$2O^*(32r_m)$</td>
</tr>
<tr>
<td>Avg. range of cand. pos. $r_a$</td>
<td>NP-c</td>
<td>2 for $r_a \geq 2$</td>
</tr>
<tr>
<td>Avg. KT-distance $d_a$</td>
<td>FPT</td>
<td>$4O^<em>(5.823d_a), \ 3O^</em>(2^{O(\sqrt{d_a})})$</td>
</tr>
<tr>
<td>Partial kernel:</td>
<td></td>
<td>$5 \frac{16}{3} \cdot d_a$ candidates</td>
</tr>
<tr>
<td>Max. KT-distance $d_m$</td>
<td>FPT</td>
<td>$4O^<em>(4.829d_m), \ 3O^</em>(2^{O(\sqrt{d_m})})$</td>
</tr>
</tbody>
</table>

1. Dwork, Kumar, Naor, Sivakumar, WWW 2001
3. Karpinski and Schudy, ISAAC 2010
4. Simjour, IWPEC 2009
5. Betzler, Guo, Komusiewicz, and N., JCSS 2011; Betzler, Bredereck, and N., Manuscript of long version of IPEC 2010
Partial Kernelization

View Kemeny Score as a two-dimensional problem with dimensions “number $n$ of votes” and “number $m$ of candidates.

**Basic idea:**
Shrink instance into an equivalent smaller instance
- by polynomial-time executable data reduction rules such that
- the size of one “problem dimension” (that is, the number $m$ of candidates here) only depends on the parameter.
PARTIAL KERNELIZATION

View Kemeny Score as a two-dimensional problem with dimensions “number $n$ of votes” and “number $m$ of candidates.

Basic idea:
Shrink instance into an equivalent smaller instance

- by polynomial-time executable data reduction rules such that
- the size of one “problem dimension” (that is, the number $m$ of candidates here) only depends on the parameter.

Recall:

- Kemeny Score is NP-hard for $n = 4$ and
- Kemeny Score is fixed-parameter tractable with respect to $m$. ($O^*(2^m)$ dynamic programming algorithm.)
**PARTIAL KERNEL FOR KEMENY SCORE**

Idea based on 3/4-majority relations:

- Find candidate pairs that are in the same relative order in at least 3/4 of the votes.
- Their relative order in every Kemeny consensus is then fixed analogously.
**PARTIAL KERNEL FOR KEMENY SCORE**

Idea based on $3/4$-majority relations:

- Find candidate pairs that are in the same relative order in at least $3/4$ of the votes.
- Their relative order in every Kemeny consensus is then fixed analogously.

**Definition**

A candidate $c$ is **non-dirty** if for every other candidate $c'$ either $c' \geq 3/4 \ c$ or $c \geq 3/4 \ c'$. Otherwise $c$ is **dirty**.
**PARTIAL KERNEL FOR KEMENY SCORE**

**Idea based on 3/4-majority relations:**

- Find candidate pairs that are in the same relative order in at least 3/4 of the votes.
- Their relative order in every Kemeny consensus is then fixed analogously.

**Definition**

A candidate $c$ is **non-dirty** if for every other candidate $c'$ either $c' \geq \frac{3}{4} c$ or $c \geq \frac{3}{4} c'$. Otherwise $c$ is **dirty**.

**Lemma**

For a non-dirty candidate $c$ and candidate $c' \in C \setminus \{c\}$:

- If $c \geq \frac{3}{4} c'$, then $c > c'$ in every Kemeny consensus.
- If $c' \geq \frac{3}{4} c$, then $c' > c$ in every Kemeny consensus.
**Partial Kernel for Kemeny Score**

Idea based on $3/4$-majority relations:

- Find candidate pairs that are in the same relative order in at least $3/4$ of the votes.
- Their relative order in every Kemeny consensus is then fixed analogously.

**Definition**
A candidate $c$ is **non-dirty** if for every other candidate $c'$ either $c' \geq_{3/4} c$ or $c \geq_{3/4} c'$. Otherwise $c$ is **dirty**.

**Lemma**
For a non-dirty candidate $c$ and candidate $c' \in C\setminus\{c\}$:
- If $c \geq_{3/4} c'$, then $c > c'$ in every Kemeny consensus.
- If $c' \geq_{3/4} c$, then $c' > c$ in every Kemeny consensus.

**Data Reduction Rule**
If there is a non-dirty candidate $c$, then delete $c$ and partition the instance into two subinstances accordingly.
Reduction rules using “majority relations”

<table>
<thead>
<tr>
<th>Precendence</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$ &gt; $a_2$ &gt; $a_3$ &gt; $c$ &gt; $b_1$ &gt; $b_2$</td>
<td></td>
</tr>
<tr>
<td>$a_3$ &gt; $a_2$ &gt; $c$ &gt; $a_1$ &gt; $b_2$ &gt; $b_1$</td>
<td></td>
</tr>
<tr>
<td>$a_1$ &gt; $c$ &gt; $a_2$ &gt; $b_2$ &gt; $b_1$ &gt; $a_3$</td>
<td></td>
</tr>
<tr>
<td>$a_2$ &gt; $a_3$ &gt; $a_1$ &gt; $b_1$ &gt; $b_2$ &gt; $c$</td>
<td></td>
</tr>
</tbody>
</table>

Further (extended) rule:

Data reduction based on non-dirty sets of candidates.

$\frac{3}{4} a_i \geq c$ and $c \geq \frac{3}{4} b_i$ \implies$

in every Kemeny consensus:

$\{a_1, a_2, a_3\}$ > $c$ > $\{b_1, b_2\}$
**Reduction rules using “majority relations”**

\[ a_1 > a_2 > a_3 > c > b_1 > b_2 \]
\[ a_3 > a_2 > c > a_1 > b_2 > b_1 \]
\[ a_1 > c > a_2 > b_2 > b_1 > a_3 \]
\[ a_2 > a_3 > a_1 > b_1 > b_2 > c \]

\[ a_i \geq \frac{3}{4} c \text{ and } c \geq \frac{3}{4} b_i \]

\[ \Rightarrow \]

in every Kemeny consensus:

\[ \{a_1, a_2, a_3\} > c > \{b_1, b_2\} \]

\[ a_1 > a_2 > a_3 \]
\[ a_3 > a_2 > a_1 \]
\[ a_1 > a_2 > a_3 \]
\[ a_2 > a_3 > a_1 \]

\[ c \]
\[ b_1 > b_2 \]
\[ c \]
\[ b_2 > b_1 \]
\[ c \]
\[ b_2 > b_1 \]
\[ c \]
\[ b_1 > b_2 \]
Reduction rules using “majority relations”

\[
\begin{align*}
    a_1 &> a_2 > a_3 > c > b_1 > b_2 \\
    a_3 &> a_2 > c > a_1 > b_2 > b_1 \\
    a_1 &> c > a_2 > b_2 > b_1 > a_3 \\
    a_2 &> a_3 > a_1 > b_1 > b_2 > c
\end{align*}
\]

\[
\begin{align*}
    \text{Further (extended) rule:} \\
    \text{Data reduction based on non-dirty sets of candidates...}
\end{align*}
\]
Reduction rules using “majority relations”

\[
\begin{align*}
&a_1 > a_2 > a_3 > c_1 > c_2 > b_1 > b_2 \\
&a_3 > a_2 > c_2 > c_1 > a_1 > b_2 > b_1 \\
&a_1 > c_1 > c_2 > a_2 > b_2 > b_1 > a_3 \\
&a_2 > a_3 > a_1 > b_1 > b_2 > c_2 > c_1
\end{align*}
\]

\[a_i \geq \frac{3}{4} \text{ and } c_j \geq \frac{3}{4} \text{ for } b_i\]

\[
\Rightarrow \quad \text{in every Kemeny consensus:}
\]

\[
\{a_1, a_2, a_3\} > \{c_1, c_2\} > \{b_1, b_2\}
\]
### Reducation Rules Using "Majority Relations"

<table>
<thead>
<tr>
<th>a₁</th>
<th>a₂</th>
<th>a₃</th>
<th>c₁</th>
<th>c₂</th>
<th>b₁</th>
<th>b₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₃</td>
<td>a₂</td>
<td>c₂</td>
<td>c₁</td>
<td>a₁</td>
<td>b₂</td>
<td>b₁</td>
</tr>
</tbody>
</table>

\[ a_i \geq \frac{3}{4} \quad \text{and} \quad c_j \geq \frac{3}{4} \]

in every Kemeny consensus:

\[ \{a₁, a₂, a₃\} > \{c₁, c₂\} > \{b₁, b₂\} \]

**Three subinstances (one for the non-dirty set):**

\[
\begin{align*}
& a₁ > a₂ > a₃ & & c₁ > c₂ & & b₁ > b₂ \\
& a₃ > a₂ > a₁ & & c₂ > c₁ & & b₂ > b₁ \\
& a₁ > a₂ > a₃ & & c₁ > c₂ & & b₂ > b₁ \\
& a₂ > a₃ > a₁ & & c₂ > c₁ & & b₁ > b₂
\end{align*}
\]
**Reduction rules using “majority relations”**

In every Kemeny consensus:

\[
\{a_1, a_2, a_3\} > \{c_1, c_2\} > \{b_1, b_2\}
\]

Three subinstances (one for the non-dirty set):

- \(a_1 > a_2 > a_3\)
- \(c_1 > c_2\)
- \(b_1 > b_2\)

- \(a_3 > a_2 > a_1\)
- \(c_2 > c_1\)
- \(b_2 > b_1\)

- \(a_1 > a_2 > a_3\)
- \(c_1 > c_2\)
- \(b_2 > b_1\)

- \(a_2 > a_3 > a_1\)
- \(c_2 > c_1\)
- \(b_1 > b_2\)

Such sets can be found in polynomial time.
AVERAGE KT-DISTANCE AS PARAMETER FOR KEMENY SCORE

Parameter: average KT-distance between the input votes

\[ d_a := \frac{2}{n(n-1)} \cdot \sum_{\{u,v\} \subseteq V} \text{KT-dist}(u, v). \]
Parameter: average KT-distance between the input votes

\[ d_a := \frac{2}{n(n-1)} \cdot \sum_{\{u,v\} \subseteq V} \text{KT-dist}(u,v). \]

**Theorem**

A Kemeny Score instance with average KT-distance \( d_a \) can be reduced in polynomial time to an equivalent instance with less than \( \frac{16}{3} \cdot d_a \) candidates.

In parameterized terms: Kemeny Score yields a partial kernel with \( \frac{16}{3} \cdot d_a \) candidates.
WHAT ABOUT OTHER MAJORITIES, WHY 3/4?

Lemma
For a non-dirty candidate $c$ and candidate $c' \in C \setminus \{c\}$:
If $c \geq 3/4 c'$, then $c > c'$ in every Kemeny consensus.
If $c' \geq 3/4 c$, then $c' > c$ in every Kemeny consensus.

Observation
Lemma does not hold when we replace $3/4$ by any smaller value. We can construct counterexamples where lemma does not hold.
What about other majorities, why $3/4$?

**Lemma**

For a non-dirty candidate $c$ and candidate $c' \in C \setminus \{c\}$:

- If $c \geq 3/4 \ c'$, then $c > c'$ in every Kemeny consensus.
- If $c' \geq 3/4 \ c$, then $c' > c$ in every Kemeny consensus.

**Observation**

Lemma does not hold when we replace $3/4$ by any smaller value. We can construct counterexamples where lemma does not hold.

As to $>2/3$-majorities...:

- Kemeny Score is polynomial-time solvable if there are no dirty candidates;
WHAT ABOUT OTHER MAJORITIES, WHY 3/4?

Lemma
For a non-dirty candidate \( c \) and candidate \( c' \in C \setminus \{c\} \):
If \( c \geq \frac{3}{4} \ c' \), then \( c > c' \) in every Kemeny consensus.
If \( c' \geq \frac{3}{4} \ c \), then \( c' > c \) in every Kemeny consensus.

Observation
Lemma does not hold when we replace 3/4 by any smaller value. We can construct counterexamples where lemma does not hold.

As to \( > \frac{2}{3} \)-majorities...:
- Kemeny Score is polynomial-time solvable if there are no dirty candidates;
- quadratic partial kernel with respect to the number of dirty candidates;
WHAT ABOUT OTHER MAJORITIES, WHY 3/4?

Lemma
For a non-dirty candidate $c$ and candidate $c' \in C \setminus \{c\}$:
If $c \geq \frac{3}{4} c'$, then $c > c'$ in every Kemeny consensus.
If $c' \geq \frac{3}{4} c$, then $c' > c$ in every Kemeny consensus.

Observation
Lemma does not hold when we replace $3/4$ by any smaller value. We can construct counterexamples where lemma does not hold.

As to $>\frac{2}{3}$-majorities...:
- Kemeny Score is polynomial-time solvable if there are no dirty candidates;
- quadratic partial kernel with respect to the number of dirty candidates;
- open: is there a partial linear kernel with respect to the number of dirty candidates?
COUNTEREXAMPLE AGAINST USING 5/7-MAJORITIES

2 votes: \[ x > y > a > b > c > d > e > f \]
3 votes: \[ a > b > c > d > e > f > x > y \]
2 votes: \[ y > a > b > c > d > e > f > x \]

Remarks:

- Similar (a little more technical) counterexamples can be found for every majority ratio in \([2/3, 3/4]\].
- For majority ratios \(s \leq 2/3\), the \(\geq s\)-majority relation is not necessarily transitive...
Counterexample Against Using 5/7-Majorities

2 votes: \( x > y > a > b > c > d > e > f \)

3 votes: \( a > b > c > d > e > f > x > y \)

2 votes: \( y > a > b > c > d > e > f > x \)

- \( x \) is non-dirty according to the \( \geq \frac{5}{7} \)-majority, since "\( x > y \)" in five out of seven votes and "\( \{a,b,c,d,e,f\} > x \)" in five out of seven votes.
COUNTEREXAMPLE AGAINST USING 5/7-MAJORITIES

2 votes: \( x > y > a > b > c > d > e > f \)

3 votes: \( a > b > c > d > e > f > x > y \)

2 votes: \( y > a > b > c > d > e > f > x \)

- \( x \) is non-dirty according to the \( \geq_{5/7} \)-majority, since “\( x > y \)” in five out of seven votes and “\( \{a, b, c, d, e, f\} > x \)” in five out of seven votes.

- Although \( x \geq_{5/7} y \), the only ranking with minimum Kemeny score is: \( y > a > b > c > d > e > f > x \)
**Counterexample Against Using 5/7-Majorities**

2 votes: \( x > y > a > b > c > d > e > f \)

3 votes: \( a > b > c > d > e > f > x > y \)

2 votes: \( y > a > b > c > d > e > f > x \)

- \( x \) is non-dirty according to the \( \geq_{5/7} \)-majority, since \( x > y \) in five out of seven votes and \( \{a, b, c, d, e, f\} > x \) in five out of seven votes.
- Although \( x \geq_{5/7} y \), the only ranking with minimum Kemeny score is: \( y > a > b > c > d > e > f > x \)

Remarks:
- Similar (a little more technical) counterexamples can be found for every majority ratio in \( ]2/3, 3/4[ \).
- For majority ratios \( s \leq 2/3 \), the \( \geq_s \)-majority relation is not necessarily transitive...
Definition
A candidate $c$ beating every other candidate in at least half of the votes, that is, $c \geq \frac{1}{2} c'$ for every candidate $c' \neq c$, is called weak Condorcet winner.
**Definition**

A candidate $c$ beating every other candidate in at least half of the votes, that is, $c \geq \frac{1}{2} c'$ for every candidate $c' \neq c$, is called **weak Condorcet winner**.

A weak Condorcet winner takes the first position in at least one Kemeny consensus (Condorcet property).
**Definition**
A candidate $c$ beating every other candidate in at least half of the votes, that is, $c \geq \frac{1}{2} c'$ for every candidate $c' \neq c$, is called **weak Condorcet winner**.

A weak Condorcet winner takes the first position in at least one Kemeny consensus (Condorcet property).

**Reduction Rule**
If there is a weak Condorcet winner in an election provided by a Kemeny Score instance, then delete this candidate.
**Data Reduction Based on Condorcet**

**Definition**
A candidate $c$ beating every other candidate in at least half of the votes, that is, $c \geq \frac{1}{2} c'$ for every candidate $c' \neq c$, is called **weak Condorcet winner**.

A weak Condorcet winner takes the first position in at least one Kemeny consensus (Condorcet property).

**Reduction Rule**
If there is a weak Condorcet winner in an election provided by a Kemeny Score instance, then delete this candidate.

A **Condorcet loser** is defined analogously. Again, this rule can be extended to a rule searching for “Condorcet winner/loser sets”...
**Effectiveness of Condorcet Rules**

Example:

\[
\begin{align*}
\text{a}_1 & > \text{a}_2 > \text{a}_3 > \text{c}_1 > \text{c}_2 > \text{b}_1 > \text{b}_2 \\
\text{a}_3 & > \text{a}_2 > \text{c}_2 > \text{c}_1 > \text{a}_1 > \text{b}_2 > \text{b}_1 \\
\text{a}_1 & > \text{c}_1 > \text{c}_2 > \text{a}_2 > \text{b}_2 > \text{b}_1 > \text{a}_3 \\
\text{a}_2 & > \text{a}_3 > \text{a}_1 > \text{b}_1 > \text{b}_2 > \text{c}_2 > \text{c}_1
\end{align*}
\]

\[
a_i \geq \frac{3}{4} c_j \text{ and } c_j \geq \frac{3}{4} b_i
\]

\[
\Rightarrow
\]

\{ \text{a}_1, \text{a}_2, \text{a}_3 \} \text{ is a Condorcet winner set, and } \\
\{ \text{b}_1, \text{b}_2 \} \text{ is a Condorcet loser set.}
**Effectiveness of Condorcet Rules**

Example:

\[
\begin{align*}
& a_1 > a_2 > a_3 > c_1 > c_2 > b_1 > b_2 \\
& a_3 > a_2 > c_2 > c_1 > a_1 > b_2 > b_1 \\
& a_1 > c_1 > c_2 > a_2 > b_2 > b_1 > a_3 \\
& a_2 > a_3 > a_1 > b_1 > b_2 > c_2 > c_1 \\
\end{align*}
\]

\[ a_i \geq \frac{3}{4} c_j \text{ and } c_j \geq \frac{3}{4} b_i \]

\[
\Rightarrow
\]

\[
\{a_1, a_2, a_3\} \quad \text{is a Condorcet winner set, and}
\]

\[
\{b_1, b_2\} \quad \text{is a Condorcet loser set.}
\]
Effectiveness of Condorcet Rules

Example:

\[ a_1 > a_2 > a_3 > c_1 > c_2 > b_1 > b_2 \]
\[ a_3 > a_2 > c_2 > c_1 > a_1 > b_2 > b_1 \]
\[ a_1 > c_1 > c_2 > a_2 > b_2 > b_1 > a_3 \]
\[ a_2 > a_3 > a_1 > b_1 > b_2 > c_2 > c_1 \]

\[ \{a_1, a_2, a_3\} \] is a Condorcet winner set, and
\[ \{b_1, b_2\} \] is a Condorcet loser set.

Fact:
The rule searching for Condorcet sets are at least as effective as the majority-based rules.
Such sets can be found in polynomial time.
DATA REDUCTION RULES APPLIED

Running time comparison for four data reduction rules:
non-dirty candidates $<^1$ Condorcet candidates $<^2$ non-dirty sets $<^1$ Condorcet sets

Heuristic combination of the data reduction rules

1. If exists, eliminate a non-dirty candidate.
2. Otherwise, if exists, eliminate a Condorcet candidate.
3. Otherwise, if exists, eliminate a non-dirty set.
4. Otherwise, if exists, eliminate a Condorcet set.

$^1$empirical, $^2$provable
Running time comparison for four data reduction rules:
non-dirty candidates $<^1$ Condorcet candidates $<^2$ non-dirty sets $<^1$ Condorcet sets

Heuristic combination of the data reduction rules

1. If exists, eliminate a non-dirty candidate.
2. Otherwise, if exists, eliminate a Condorcet candidate.
3. Otherwise, if exists, eliminate a non-dirty set.
4. Otherwise, if exists, eliminate a Condorcet set.

<table>
<thead>
<tr>
<th>instance</th>
<th>Condorcet set alone</th>
<th>heuristic combination above</th>
</tr>
</thead>
<tbody>
<tr>
<td>blues</td>
<td>0.84 sec</td>
<td>0.10 sec</td>
</tr>
<tr>
<td>gardening</td>
<td>0.95 sec</td>
<td>0.11 sec</td>
</tr>
<tr>
<td>classical guitar</td>
<td>1.89 sec</td>
<td>0.18 sec</td>
</tr>
</tbody>
</table>

$^1$ empirical, $^2$ provable
## Reduction of Metasearch Engine Data

Four votes: Google, Lycos, MSN Live Search, and Yahoo!
top 1000 hits each, candidates that appear in all four rankings

<table>
<thead>
<tr>
<th>search term</th>
<th>cand.</th>
<th>sec.</th>
<th>red. inst. solved/unsol.</th>
</tr>
</thead>
<tbody>
<tr>
<td>aff. action</td>
<td>127</td>
<td>0.41</td>
<td>&gt; 41 &gt;</td>
</tr>
<tr>
<td>alcoholism</td>
<td>115</td>
<td>0.21</td>
<td>[115]</td>
</tr>
<tr>
<td>architecture</td>
<td>122</td>
<td>0.47</td>
<td>&gt; 12 &gt; [30] &gt; 17 &gt;</td>
</tr>
<tr>
<td>blues</td>
<td>112</td>
<td>0.16</td>
<td>&gt; 9 &gt;</td>
</tr>
<tr>
<td>cheese</td>
<td>142</td>
<td>0.39</td>
<td>&gt; 6 &gt;</td>
</tr>
<tr>
<td>class. guitar</td>
<td>115</td>
<td>1.12</td>
<td>&gt; 7 &gt; [50] &gt; 35 &gt;</td>
</tr>
<tr>
<td>Death Valley</td>
<td>110</td>
<td>0.25</td>
<td>&gt; 7 &gt; [30] &gt; 8 &gt;</td>
</tr>
<tr>
<td>field hockey</td>
<td>102</td>
<td>0.21</td>
<td>&gt; 26 &gt; [20] &gt; 4 &gt;</td>
</tr>
<tr>
<td>HIV</td>
<td>115</td>
<td>0.26</td>
<td>&gt; 5 &gt; [7] &gt; 20 &gt;</td>
</tr>
<tr>
<td>lyme disease</td>
<td>153</td>
<td>2.61</td>
<td>&gt; 97 &gt;</td>
</tr>
<tr>
<td>rock climbing</td>
<td>102</td>
<td>0.12</td>
<td>[102]</td>
</tr>
<tr>
<td>Shakespeare</td>
<td>163</td>
<td>0.68</td>
<td>&gt; 10 &gt; [25] &gt; 6 &gt;</td>
</tr>
<tr>
<td>telecomm.</td>
<td>131</td>
<td>2.28</td>
<td>&gt; 109 &gt;</td>
</tr>
</tbody>
</table>
Strongest (fastest) empirical results with combination of our data reduction rules and an ILP formulation...
Exact Solutions using Data Reduction & ILPs

Strongest (fastest) empirical results with combination of our data reduction rules and an ILP formulation...

ILP formulation of Kemeny Score\(^1\) using

- \(C\) for the set of candidates;
- coefficients \(#_{a>b}\) for the number of rankings having “\(a > b\)”;
- binary variables \(x_{a>b}\) if “\(a > b\)” in a Kemeny consensus.

\[
\text{minimize } \sum_{\{a,b\} \subseteq C} #_{a>b} \cdot x_{a>b} + #_{b>a} \cdot x_{b>a}
\]
Strongest (fastest) empirical results with combination of our data reduction rules and an ILP formulation...

**ILP formulation of Kemeny Score**$^1$ using

- $C$ for the set of candidates;
- coefficients $\#_{a>b}$ for the number of rankings having “$a > b$”;
- binary variables $x_{a>b}$ if “$a > b$” in a Kemeny consensus.

\[
\text{minimize } \sum_{\{a,b\} \subseteq C} \#_{a>b} \cdot x_{a>b} + \#_{b>a} \cdot x_{b>a}
\]

subject to

for all $\{a, b\} \subseteq C$: $x_{a>b} + x_{b>a} = 1$
**Exact Solutions using Data Reduction & ILPs**

Strongest (fastest) empirical results with combination of our data reduction rules and an ILP formulation...

**ILP formulation of Kemeny Score¹** using

- $C$ for the set of candidates;
- coefficients $\#_{a>b}$ for the number of rankings having “$a > b$”;
- binary variables $x_{a>b}$ if “$a > b$” in a Kemeny consensus.

**minimize** $\sum_{\{a,b\} \subseteq C} \#_{a>b} \cdot x_{a>b} + \#_{b>a} \cdot x_{b>a}$

**subject to**

for all $\{a, b\} \subseteq C$: $x_{a>b} + x_{b>a} = 1$

for all $\{a, b, c\} \subseteq C$: $x_{a>b} + x_{b>c} + x_{c>a} \geq 1$

- First conditions are to ensure that either “$a > b$” or “$b > a$” (for fixed $a$ and $b$);
- second conditions are to ensure transitivity.

¹ [Conitzer, Davenport, Kalagnanam, AAAI 2006]
Observations:

- In our experiments, no combinatorial (fixed-parameter) algorithm for exactly solving Kemeny Score could compete with the ILP-based solver (gurobi).
- Instances with hundreds of candidates can be solved within few seconds.
- Data reduction used as preprocessing led to significant speedups when compared to using the ILP alone.
SPEEDUP OF ILP THROUGH DATA REDUCTION

Rolf Niedermeier, TU Berlin ()
Rank Aggregation and Kemeny Voting
CONCLUDING REMARKS

- Key feature of our data reduction: Break instances into smaller, independent parts.
- Execution order of data reduction rule execution has significant impact on efficiency.
- “Cascading effects” of data reduction rules not well understood.
CONCLUDING REMARKS

• Key feature of our data reduction:
  Break instances into smaller, independent parts.
• Execution order of data reduction rule execution has significant impact on efficiency.
• “Cascading effects” of data reduction rules not well understood.

Some Challenges:
• Improved analysis using Condorcet rules?
• Linear partial kernel for “s-majorities” with $s < 3/4$?
• Our data reduction rules do not apply to “constraint rankings” where the input also contains some candidate pairs whose relative ordering in the consensus ranking is already fixed...
CONCLUDING REMARKS

- Key feature of our data reduction: Break instances into smaller, independent parts.
- Execution order of data reduction rule execution has significant impact on efficiency.
- “Cascading effects” of data reduction rules not well understood.

Some Challenges:
- Improved analysis using Condorcet rules?
- Linear partial kernel for “s-majorities” with $s < 3/4$?
- Our data reduction rules do not apply to “constraint rankings” where the input also contains some candidate pairs whose relative ordering in the consensus ranking is already fixed...

Merci!