Parameterized Complexity of Voting Systems

Nadja Betzler
Friedrich-Schiller-Universität Jena

joint work with
Michael R. Fellows, Jiong Guo, Rolf Niedermeier,
and Frances A. Rosamond

Theorietag, Februar 2009, Ulm
Applications

Voting scenarios:

- political elections
- committees: decisions about job applicants, grant proposals
- meta search engines, recommender systems
- daily life: choice of restaurant
Applications

Voting scenarios:
- political elections
- committees: decisions about job applicants, grant proposals
- meta search engines, recommender systems
- daily life: choice of restaurant

Different goals:
- single winner
- set of winners
- ranking of all candidates
- decisions on several (dependent) subjects
How to determine a winner?

Election

Set of votes $V$, set of candidates $C$.

A vote is a ranking (total order) over all candidates.
How to determine a winner?

**Election**

Set of votes $V$, set of candidates $C$.

A vote is a ranking (total order) over all candidates.

many different kinds of **voting systems**

2 important groups:

**positional scoring protocols**

Borda, plurality, . . .

**based on pairwise head-to-head contests**

Condorcet, Copeland, sequential majority voting, . . .
Computational problems

Examples:

**Computation of a/the winner**

**“Behavior” of voting systems**
- Strategic voting/Manipulation (Gibbarth-Satterwaith Theorem)
- Electoral control
- Bribery, lobbying, . . .

**Dealing with incomplete information**
Possible Winner and Necessary Winner
Kemeny ranking

Election

Set of votes $V$, set of candidates $C$.
A vote is a ranking (total order) over all candidates.

Example: $C = \{a, b, c\}$

- vote 1: $a > b > c$
- vote 2: $a > c > b$
- vote 3: $b > c > a$

How to aggregate the votes into a “consensus ranking”? 
KT-distance (between two votes $v$ and $w$)

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

where $d_{v,w}(c, d)$ is 0 if $v$ and $w$ rank $c$ and $d$ in the same order, 1 otherwise.

Example:

$v : a > b > c$

$w : c > a > b$

$$KT\text{-}dist(v, w) = d_{v,w}(a, b) + d_{v,w}(a, c) + d_{v,w}(b, c)$$
$$= 0 + 1 + 1$$
$$= 2$$
Kemeny Consensus

Kemeny score of a ranking \( r \):
sum of KT-distances between \( r \) and all votes

Kemeny consensus \( r_{con} \):
a ranking that minimizes the Kemeny score

\[ \begin{align*}
  v_1 : & \quad a > b > c \quad \text{KT-dist}(r_{con}, v_1) = 0 \\
  v_2 : & \quad a > c > b \quad \text{KT-dist}(r_{con}, v_2) = 1 \quad \text{because of } \{b, c\} \\
  v_3 : & \quad b > c > a \quad \text{KT-dist}(r_{con}, v_3) = 2 \quad \text{because of } \{a, b\} \text{ and } \{a, c\}
\end{align*} \]

\[ r_{con} : \quad a > b > c \quad \text{Kemeny score: } 0 + 1 + 2 = 3 \]
Motivation

Applications:

- ranking of web sites (meta search engines), spam detection
  [Dwork et al., WWW 2001]
- databases
  [Fagin et al., SIGMOD, 2003]
- bioinformatics
  [Jackson et al., IEEE/ACM Transactions on Computational Biology and Bioinformatics 2008]

Only voting system that is
- neutral,
- consistent, and
- Condorcet.
Decision problems

**Kemeny Score**

*Input:* An election \((V, C)\) and a positive integer \(k\).

*Question:* Is the Kemeny score of \((V, C)\) at most \(k\)?

**Kemeny Winner**

*Input:* An election \((V, C)\) and a distinguished candidate \(c\).

*Question:* Is there a Kemeny consensus in which \(c\) is at the “best” position?

\[
\begin{align*}
\text{vote 1:} & \quad a > b > c \\
\text{vote 2:} & \quad a > c > b \\
\text{vote 3:} & \quad b > c > a \\
\text{Kemeny consensus:} & \quad a > b > c
\end{align*}
\]

Kemeny score = 0 + 1 + 2 = 3
Kemeny winner: a
Known results

- **Kemeny Score** is NP-complete (even for 4 votes)
  [Dwork et al., WWW 2001]
- **Kemeny Winner** is $P^{NP}$-complete
  [E. Hemaspaandra et al., TCS 2005]

Algorithms:

- randomized factor 11/7-approximation
  [Ailon et al., J. ACM 2008]
- factor 8/5-approximation
  [van Zuylen and Williamson, WAOA 2007]
- PTAS [Kenyon-Mathieu and Schudy, STOC 2007]
- Heuristics; greedy, branch and bound
  [Davenport and Kalagnanam, AAAI 2004],
  [Conitzer et al. AAAI, 2006]
Given an NP-hard problem with input size $n$ and a parameter $k$

**Basic idea:** Confine the combinatorial explosion to $k$

**Definition**

A problem of size $n$ is called *fixed-parameter tractable* with respect to a parameter $k$ if it can be solved exactly in $f(k) \cdot n^{O(1)}$ time.
## Parameterizations of Kemeny Score

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of votes $n$</td>
<td>$n$ [Dwork et al. WWW 2001]</td>
</tr>
<tr>
<td>NP-c for $n = 4$</td>
<td>$O^*(2^m)$</td>
</tr>
<tr>
<td>Number of candidates $m$</td>
<td></td>
</tr>
<tr>
<td>Kemeny score $k$</td>
<td>$O^*(1.53^k)$</td>
</tr>
</tbody>
</table>

Further "structural" parameters:

- Position range
- Maximum range
- Average range
- Average KT-distance

Max range $r_m := \max_{c \in C} \text{range}(c) \leq O^*(32r_m)$
### Parameterizations of Kemeny Score

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of votes $n$</td>
<td>$\text{NP-c for } n = 4$ $O^*(2^m)$</td>
</tr>
<tr>
<td>Number of candidates $m$</td>
<td>$O^*(1.53^k)$</td>
</tr>
<tr>
<td>Kemeny score $k$</td>
<td></td>
</tr>
</tbody>
</table>

Further “structural” parameters:

- **Maximum range** $r_m := \max_{c \in C} \text{range}(c)$ $O^*(32^{r_m})$
- **Average range** $r_a$ $\text{NP-c for } r_a \geq 2$

**Diagram:**

Position 1 2 $i$ $i+r$ $m$

Range of $c$
### Parameterizations of Kemeny Score

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of votes $n$</td>
<td>$\text{NP-c for } n = 4$</td>
</tr>
<tr>
<td>Dwork et al. WWW 2001</td>
<td>$O^*(2^m)$</td>
</tr>
<tr>
<td>Number of candidates $m$</td>
<td>$O^*(1.53^k)$</td>
</tr>
<tr>
<td>Kemeny score $k$</td>
<td></td>
</tr>
</tbody>
</table>

Further "structural" parameters:

- **Maximum range** $r_m := \max_{c \in C} \text{range}(c)$ $O^*(32^{r_m})$
- **Average range** $r_a$ $\text{NP-c for } r_a \geq 2$
- **Average KT-distance**

![Diagram](image.png)

- **Position**: $1$ $2$ $i$ $i + r$ $m$
- **Range of $c$**
Recall: The KT-distance between two votes is the number of inversions or "conflict pairs".

**Definition**

For an election \((V, C)\) the average KT-distance \(d_a\) is defined as

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

In the following, we show that KEMENY SCORE is fixed-parameter tractable with respect to the "average KT-distance".
Complementarity of parameterizations

- Number of candidates \( m \): \( O^*(2^m) \)
- Maximum range \( r \) of candidate positions in the input votes: \( O^*(32^r) \)
- Average distance of the input votes: \( O^*(16^d_a) \)

\((m \geq r \), but corresponding algorithm has a better running time\)

Example 1: small range, large number of candidates and average distance

\[ \begin{align*}
  a &> c > b > e > d > f &\ldots \\
  b &> a > c > d > e > f &\ldots \\
  b &> c > a > e > f > d &\ldots
\end{align*} \]

⇒ check size of parameter and then use appropriate strategy

Example 2: small average distance, large number of candidates and range

\[ \begin{align*}
  a &> b > c > d > e > f &\ldots \\
  b &> c > d > e > f > a &\ldots \\
  a &> b > c > d > e > f &\ldots
\end{align*} \]
Basic idea

Average distance $d_a$.

Crucial observation

In every Kemeny consensus every candidate can only assume a number of consecutive positions that is bounded by $2 \cdot d_a$.

Dynamic programming

making use of the fact that every candidate can be “forgotten” or “inserted” at a certain position.
Crucial observation

Let the average position of a candidate $c$ be $p_a(c)$.

Lemma

Let $d_a$ be the average KT-distance of an election $(V, C)$. Then, in every optimal Kemeny consensus $l$, for every candidate $c \in C$ we have $p_a(c) - d_a < l(c) < p_a(c) + d_a$. 

average position of $a$

input votes

consensus

$\begin{bmatrix}
   \text{a} & \text{a} & \text{a} \\
   \text{a} & \text{a} & \text{a} \\
   \text{a} & \text{a} & \text{a} \\
   \text{a} & \text{a} & \text{a} \\
\end{bmatrix}$
Crucial observation

Let the average position of a candidate $c$ be $p_a(c)$.

**Lemma**

Let $d_a$ be the average KT-distance of an election $(V, C)$. Then, in every optimal Kemeny consensus $l$, for every candidate $c \in C$ we have $p_a(c) - d_a < l(c) < p_a(c) + d_a$.

Idea of proof:

1. “The Kemeny score of $(V, C)$ is smaller than $d_a \cdot |V|$.”
   We show that one of the input votes has this Kemeny score.

2. Contradiction: Assume a candidate has a position outside the given range. Then, we can show that the Kemeny score is greater than $d_a \cdot |V|$, a contradiction.
Number of candidates per position

For a position $i$, let $P_i$ denote the set of candidates that can assume $i$ in an optimal consensus.

**Lemma**

Let $d_a$ be the average KT-distance of an election $(V, C)$. For a position $i$, we have $|P_i| \leq 4 \cdot d_a$.

**Proof:** Position “range” of every candidate is at most $2 \cdot d_a$. Every candidate of $P_i$ must have a position smaller than $i + 2d_a$ and greater than $i - 2d_a$. 

$$P_i = \{a_1, \ldots, a_{2d}, b_1, \ldots, b_{2d}\}$$
consensus

\[ P_i = \{a, b, c, d, e, f\} \]

**Observation:**
For any position \( i \) and a subset \( P_i \) of candidates that can assume \( i \):

- One candidate of \( P_i \) must assume position \( i \) in a consensus.
- Every other candidate of \( P_i \) must be either left or right of \( i \).
Dynamic programming table

Position $i$, a candidate $c \in P_i$, a subset of candidates $P'_i \subseteq P_i \setminus \{c\}$

**Definition**

$T(i, c, P'_i) :=$ optimal partial Kemeny score if $c$ has position $i$ and all candidates of $P'_i$ have positions smaller than $i$

$P_i = \{a, b, c, d, e, f\}$

<table>
<thead>
<tr>
<th>consensus</th>
<th>{a,b}</th>
<th>c</th>
<th>{d,e,f}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P'_i = {a, b}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Computation of partial Kemeny scores:

- Overall Kemeny score can be decomposed (just a sum over all votes and pairs of candidates)
- Relative orders between $c$ and all other candidates are already fixed
Running time

$n$ votes
$m$ candidates

$P_i = \{a, b, c, d, e, f\}$

<table>
<thead>
<tr>
<th>consensus</th>
<th>{a, b}</th>
<th>c</th>
<th>{d, e, f}</th>
</tr>
</thead>
</table>

We have $|P_i| \leq 4d_a$, thus there are at most $2^{4d_a}$ subsets of $P_i$.

$\Rightarrow$ Table size is bounded by $16^{d_a} \cdot \text{poly}(n, m)$.

Theorem

\textsc{Kemeny Score} can be solved in $O(n^2 \cdot m \log m + 16^d \cdot (16d^2 \cdot m + 4d \cdot m^2 \log m \cdot n))$ time with average KT-distance $d_a$ and $d := \lceil d_a \rceil$. 
Overview of parameterized complexity

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of votes</strong> $n$ [Dwork et al. WWW 2001]</td>
<td>NP-c for $n = 4$</td>
</tr>
<tr>
<td><strong>Kemeny score</strong> $k$</td>
<td>$O^*(1.53^k)$</td>
</tr>
<tr>
<td><strong>Number of candidates</strong> $m$</td>
<td>$O^*(2^m)$</td>
</tr>
<tr>
<td><strong>Maximum range of candidate positions</strong> $r$</td>
<td>$O^*(32^r)$</td>
</tr>
<tr>
<td><strong>Average range of candidate positions</strong> $r_a$</td>
<td>NP-c for $r_a \geq 2$</td>
</tr>
<tr>
<td><strong>Average KT-distance</strong> $d_a$</td>
<td>$O^*(16^{d_a})$</td>
</tr>
</tbody>
</table>
Outlook

- Average distance: investigate typical values.
- Improve the running time for the parameterizations “average distance” and “maximum candidate range”.
- Implementation of the algorithms.
- Consider generalizations like incomplete votes and ties.
- NP-completeness of Kemeny Score with 3 votes?