# 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

## KT-distance (between two votes $v$ and $w$)

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

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

Example:

$$v : a > b > c$$
$$w : c > a > b$$

$$\begin{aligned} \text{KT-dist}(v, w) &= d_{v,w}(a, b) + d_{v,w}(a, c) + d_{v,w}(b, c) \\ &= 0 + 1 + 1 \\ &= 2 \end{aligned}$$

# 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

| | | |
|---|---|---|
| $v_1$ : | $a > b > c$ | KT-dist$(r_{con}, v_1) = 0$ |
| $v_2$ : | $a > c > b$ | KT-dist$(r_{con}, v_2) = 1$ because of $\{b, c\}$ |
| $v_3$ : | $b > c > a$ | KT-dist$(r_{con}, v_3) = 2$ because of $\{a, b\}$ and $\{a, c\}$ |
| $r_{con}$ : | $\mathbf{a} > \mathbf{b} > \mathbf{c}$ | 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?

| | | | | | |
|---|---|---|---|---|---|
| vote 1: | a | > | b | > | c |
| vote 2: | a | > | c | > | b |
| vote 3: | b | > | c | > | a |
| Kemeny consensus: | a | > | b | > | c |

Kemeny score $= 0+1+2 = 3$

Kemeny winner: a

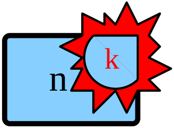
# Known results

- KEMENY SCORE is NP-complete (even for 4 votes)
  [DWORK ET AL., WWW 2001]
- KEMENY WINNER is $P_{\parallel}^{NP}$-complete
  [E. HEMASPAANDRA ET AL., TCS 2005]

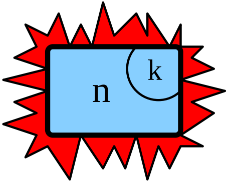
Algorithms:

- 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 $k$




**instead of**




**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.

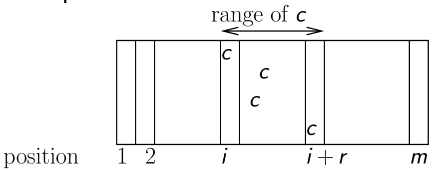
# Parameterizations of Kemeny Score

| | |
|---|---|
| Number of votes $n$ [Dwork et al. WWW 2001] | NP-c for $n = 4$ |
| Number of candidates $m$ | $O^*(2^m)$ |
| Kemeny score $k$ | $O^*(1.53^k)$ |

# Parameterizations of Kemeny Score

| | |
|---|---|
| Number of votes $n$ [Dwork et al. WWW 2001] | NP-c for $n = 4$ |
| Number of candidates $m$ | $O^*(2^m)$ |
| Kemeny score $k$ | $O^*(1.53^k)$ |

Further "structural" parameters:

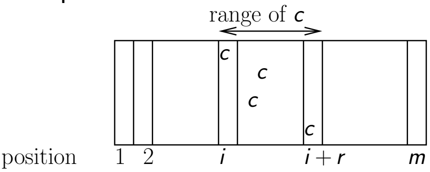



| | |
|---|---|
| Maximum range $r_m := \max_{c \in C} \text{range}(c)$ | $O^*(32^{r_m})$ |
| Average range $r_a$ | NP-c for $r_a \geq 2$ |

# Parameterizations of Kemeny Score

| | |
|---|---|
| Number of votes $n$ [Dwork et al. WWW 2001] | NP-c for $n = 4$ |
| Number of candidates $m$ | $O^*(2^m)$ |
| Kemeny score $k$ | $O^*(1.53^k)$ |

Further "structural" parameters:




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

Average KT-distance

# Average KT-distance

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

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $>$ | $c$ | $>$ | $b$ | $>$ | $e$ | $>$ | $d$ | $>$ | $f$ | $\dots$ |
| $b$ | $>$ | $a$ | $>$ | $c$ | $>$ | $d$ | $>$ | $e$ | $>$ | $f$ | $\dots$ |
| $b$ | $>$ | $c$ | $>$ | $a$ | $>$ | $e$ | $>$ | $f$ | $>$ | $d$ | $\dots$ |

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

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $>$ | $b$ | $>$ | $c$ | $>$ | $d$ | $>$ | $e$ | $>$ | $f$ | | $\dots$ |
| $b$ | $>$ | $c$ | $>$ | $d$ | $>$ | $e$ | $>$ | $f$ | $>$ | $a$ | | $\dots$ |
| $a$ | $>$ | $b$ | $>$ | $c$ | $>$ | $d$ | $>$ | $e$ | $>$ | $f$ | | $\dots$ |

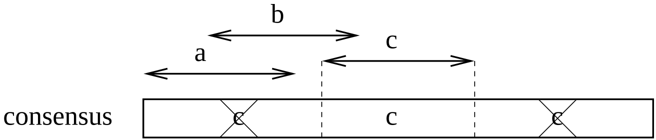
$\Rightarrow$ check size of parameter and then use appropriate strategy

# 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**

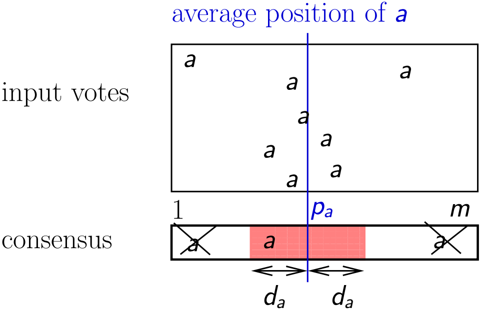
making use of the fact that every candidate can be "forgotten" or "inserted" at a certain position.

# 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 $p_a(c) - d_a < l(c) < p_a(c) + d_a$.



# 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.

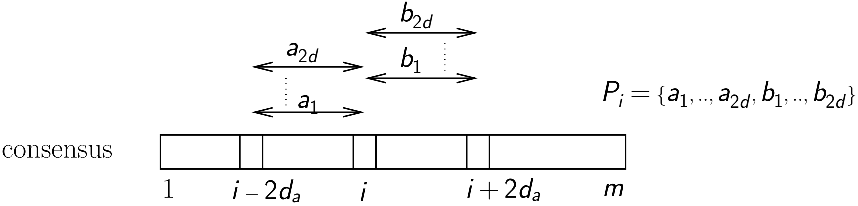
# Number of candidates per position

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

**Lemma**

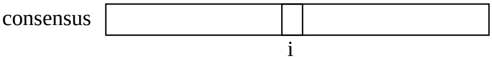
Let $d_a$ be the average KT-distance of an election $(V, C)$. For a position $i$, we have $|P_i| \leq 4 \cdot d_a$.

**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 $i - 2d_a$.

# Dynamic programming




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

**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 $i$.

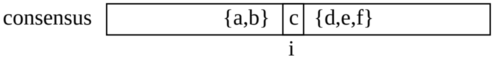
# Dynamic programming table

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 $i$

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




$P'_i = \{a, b\}$

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

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

consensus $\{a, b\}$ | $c$ | $\{d, e, f\}$
i

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**

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

| | Kemeny Score |
|---|---|
| Number of votes $n$ [Dwork et al. WWW 2001] | NP-c for $n = 4$ |
| Kemeny score $k$ | $O^*(1.53^k)$ |
| Number of candidates $m$ | $O^*(2^m)$ |
| Maximum range of candidate positions $r$ | $O^*(32^r)$ |
| Average range of candidate positions $r_a$ | NP-c for $r_a \geq 2$ |
| Average KT-distance $d_a$ | $O^*(16^{d_a})$ |

# 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?