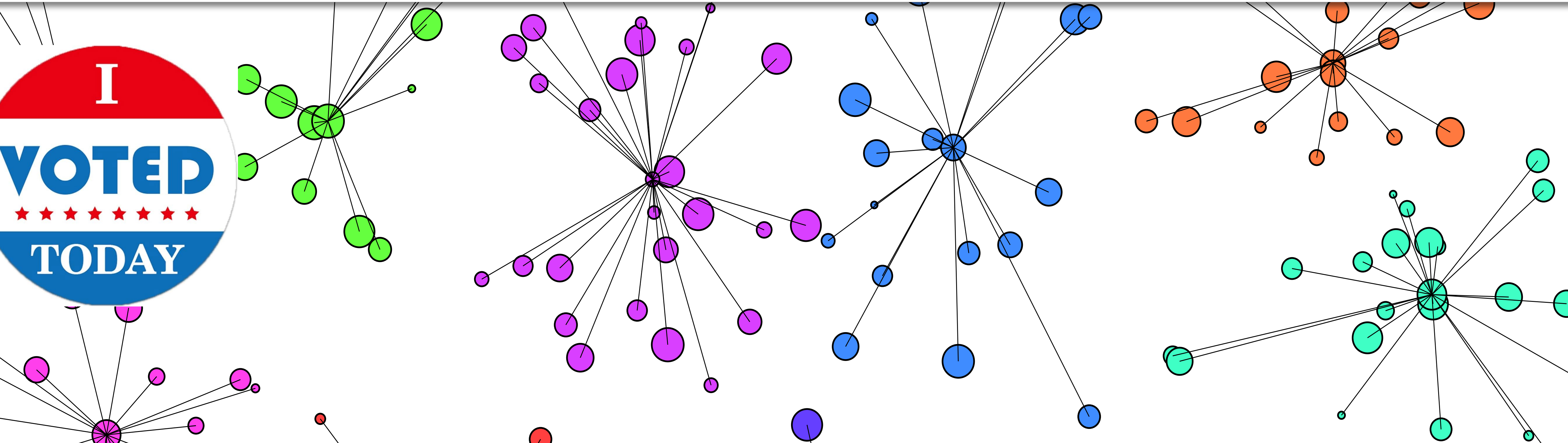
Ballot Clustering Algorithms

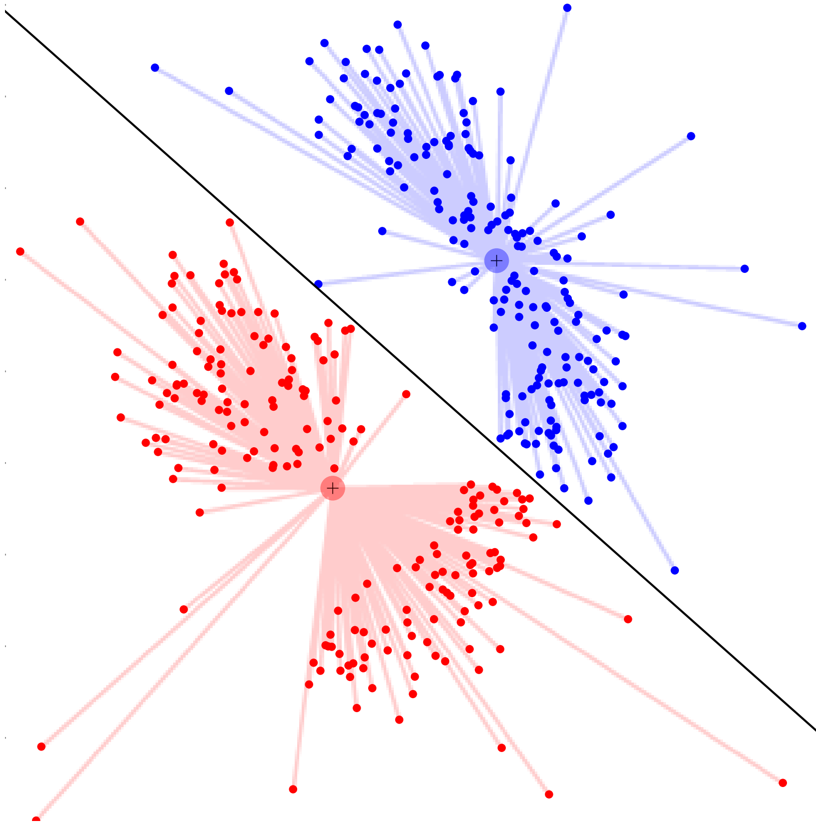
Kristopher Tapp, Saint Joseph's University

joint with Moon Duchin & David Shmoys



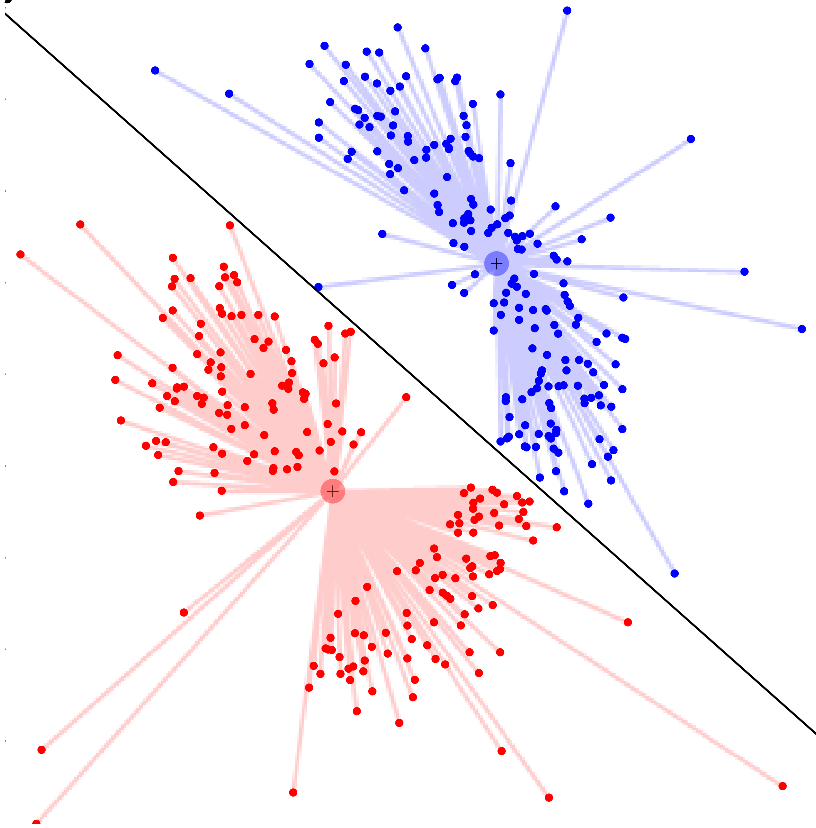
There are standard algorithms to cluster points of Euclidean space.

What about ranked choice ballots?



Edinburgh Ward 2 (Petland Hills): 7 candidates, 11315 ballots, 1238 distinct ballots.

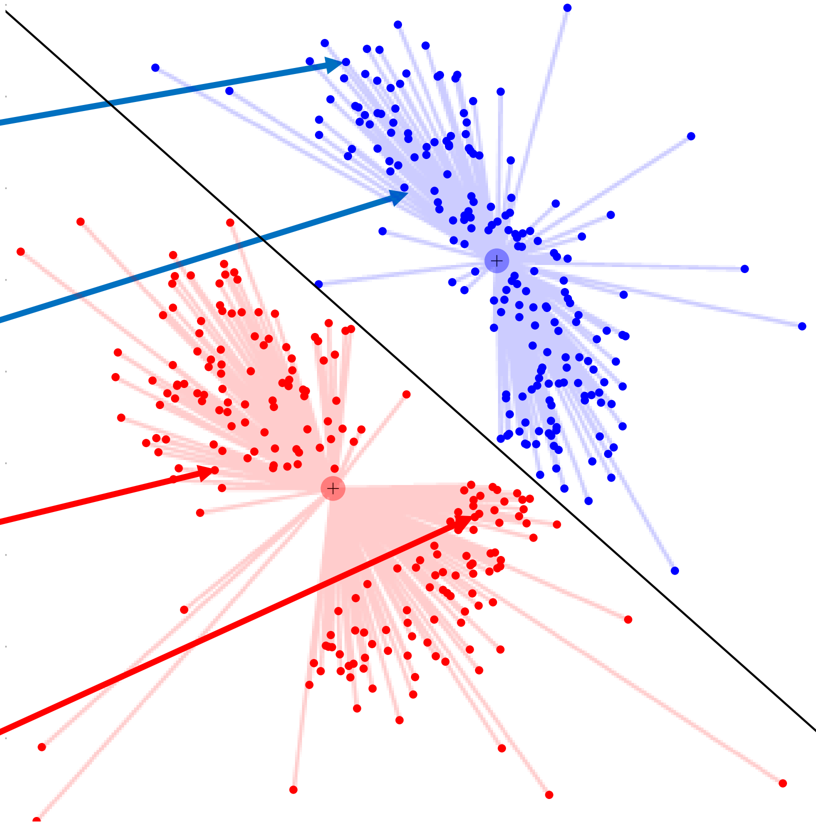
1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).
173 votes for (4, 2).
152 votes for (6, 1, 4).
144 votes for (5, 3, 7).
136 votes for (1, 6, 4, 2).
136 votes for (3, 5, 4)…



1=Graeme Bruce (C), 2 = Emma Farthing (LD), 3 = Neil Gardiner (SNP), 4 = Ricky Henderson (Lab), 5 = Ernesta Noreikiene (SNP), 6= Susan Webber (C), 7 = Evelyn Weston (Grn).

PROXY CLUSTERING: Associate each ballot to a proxy point of Euclidean space, and perform a standard clustering algorithm on the proxies.

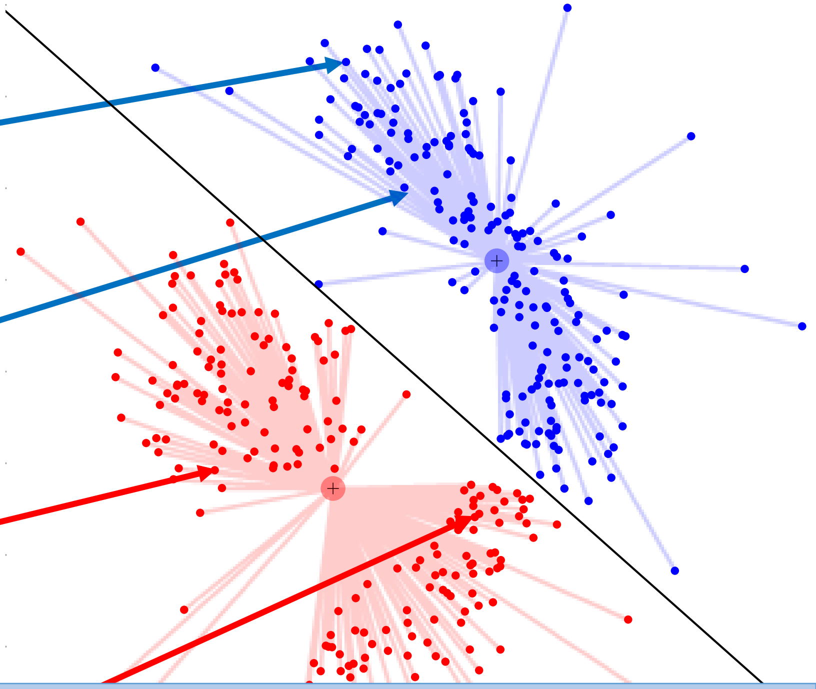
1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).
173 votes for (4, 2).
152 votes for (6, 1, 4).
144 votes for (5, 3, 7).
136 votes for (1, 6, 4, 2).
136 votes for (3, 5, 4)….



1=Graeme Bruce (C), 2 = Emma Farthing (LD), 3 = Neil Gardiner (SNP), 4 = Ricky Henderson (Lab), 5 = Ernesta Noreikiene (SNP), 6= Susan Webber (C), 7 = Evelyn Weston (Grn).

PROXY CLUSTERING: Associate each ballot to a proxy point of Euclidean space, and perform a standard clustering algorithm on the proxies.

1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).
173 votes for (4, 2).
152 votes for (6, 1, 4).
144 votes for (5, 3, 7).
136 votes for (1, 6, 4, 2).
136 votes for (3, 5, 4)....



* Any number of clusters is allowed.

* The distance between proxies must correspond to some natural measurement of ballot similarity.

SLATE CLUSTERING: Find the partition of the candidates into two slates A,B such that the ballots are most starkly divided into “A>B types” and “B>A types”. Partition the ballots accordingly.

SLATE CLUSTERING: Find the partition of the candidates into two slates A,B such that the ballots are most starkly divided into “A>B types” and “B>A types”. Partition the ballots accordingly.

1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).
173 votes for (4, 2).
152 votes for (6, 1, 4).
144 votes for (5, 3, 7).
136 votes for (1, 6, 4, 2).
136 votes for (3, 5, 4)….

SLATE A:
3 = Neil Gardiner (SNP)
5 = Ernesta Noreikiene (SNP)
7 = Evelyn Weston (Grn)

SLATE B:
1 = Graeme Bruce (C)
2 = Emma Farthing (LD)
4 = Ricky Henderson (Lab)
6= Susan Webber (C)

SLATE CLUSTERING: Find the partition of the candidates into two slates A,B such that the ballots are most starkly divided into “A>B types” and “B>A types”. Partition the ballots accordingly.

1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).
173 votes for (4, 2).
152 votes for (6, 1, 4).
144 votes for (5, 3, 7).
136 votes for (1, 6, 4, 2).
136 votes for (3, 5, 4)….

SLATE A:
3 = Neil Gardiner (SNP)
5 = Ernesta Noreikiene (SNP)
7 = Evelyn Weston (Grn)

SLATE B:
1 = Graeme Bruce (C)
2 = Emma Farthing (LD)
4 = Ricky Henderson (Lab)
6= Susan Webber (C)

- Always forms 2 clusters
- Good for studying polarized elections

SLATE CLUSTERING: Find the partition of the candidates into two slates A,B such that the ballots are most starkly divided into "A>B types" and "B>A types". Partition the ballots accordingly.

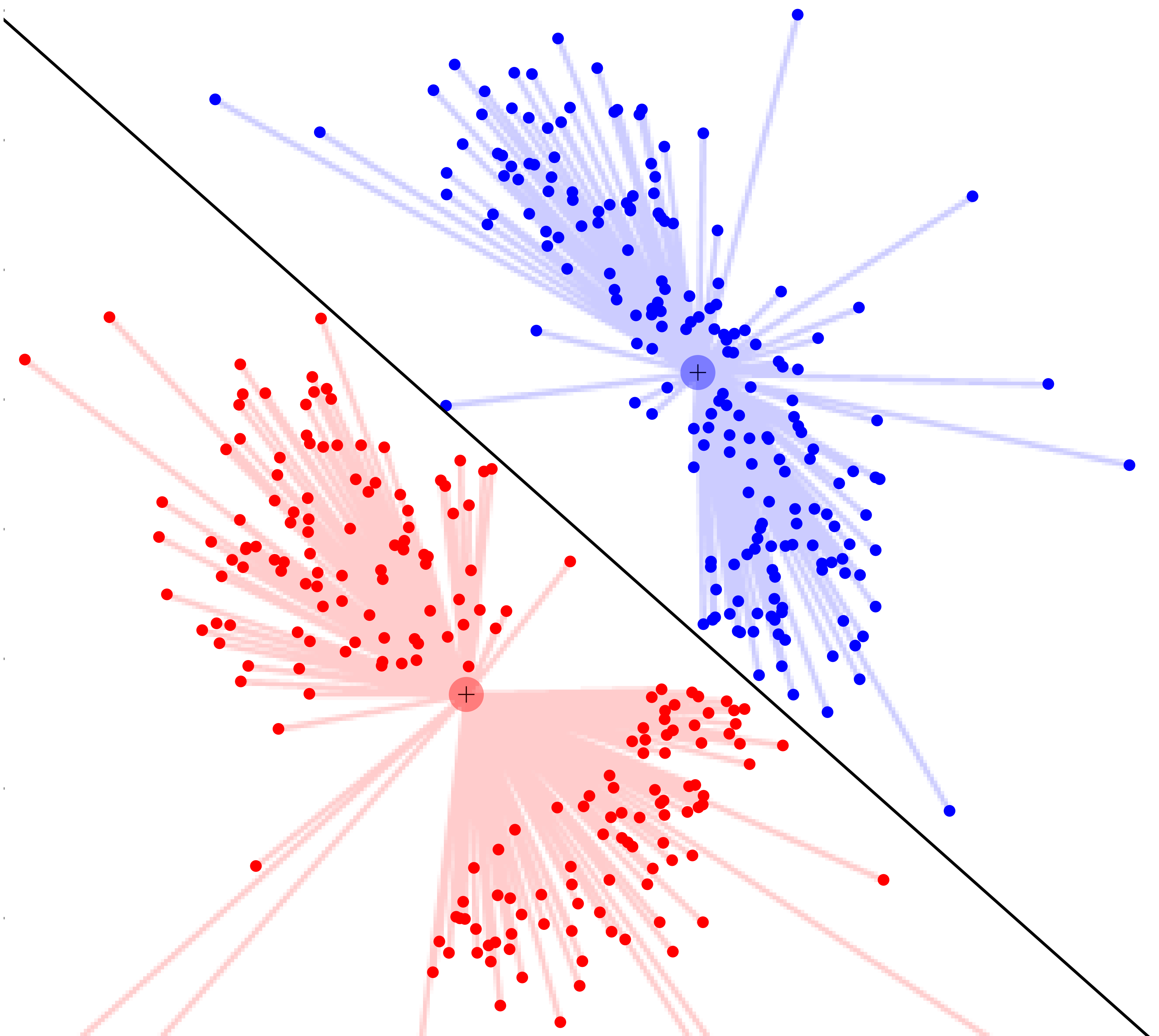
1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).
173 votes for (4, 2).
152 votes for (6, 1, 4).
144 votes for (5, 3, 7).
136 votes for (1, 6, 4, 2).
136 votes for (3, 5, 4)….

SLATE A:
3 = Neil Gardiner (SNP)
5 = Ernesta Noreikiene (SNP)
7 = Evelyn Weston (Grn)

SLATE B:
1 = Graeme Bruce (C)
2 = Emma Farthing (LD)
4 = Ricky Henderson (Lab)
6= Susan Webber (C)

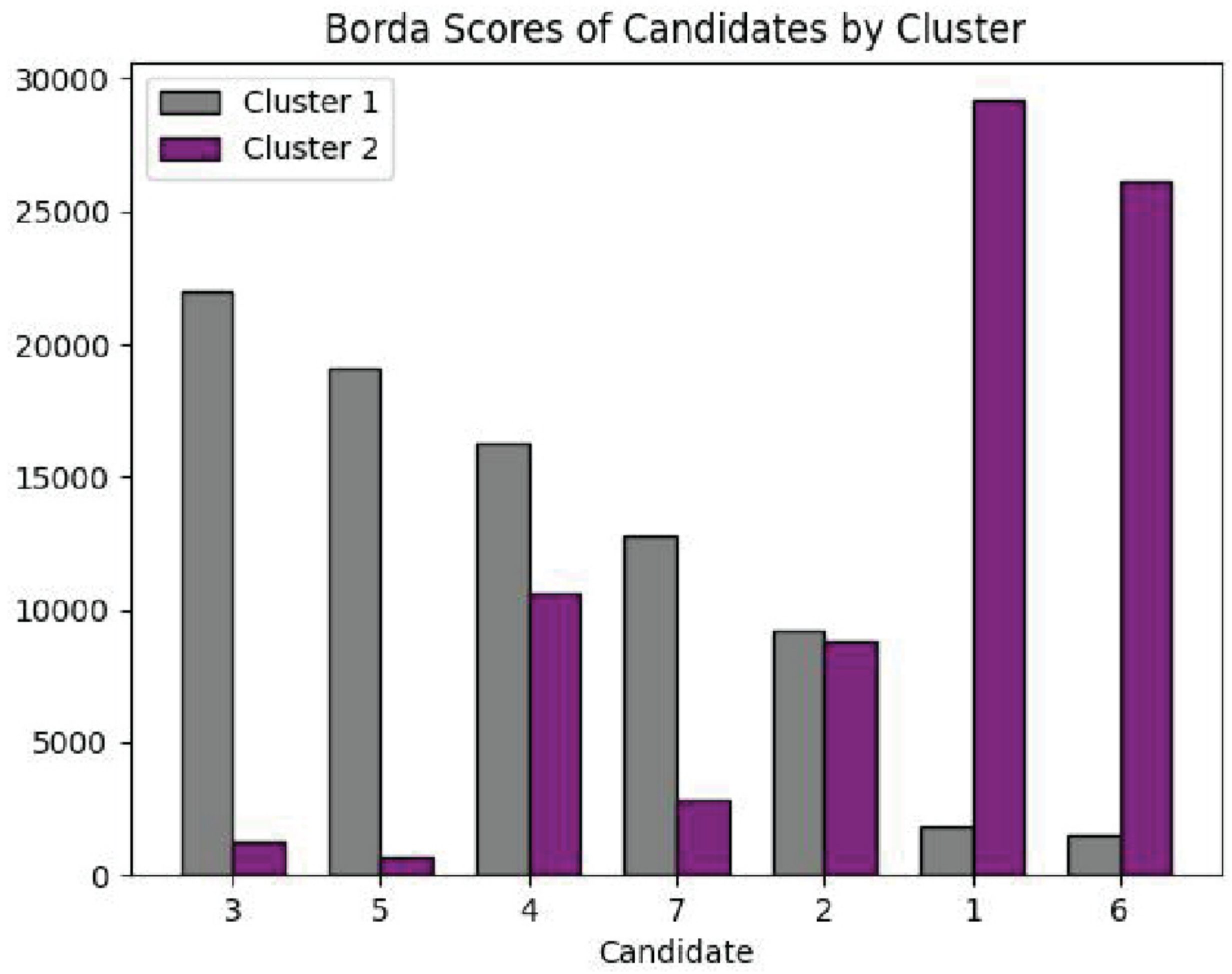
- Always forms 2 clusters
- Good for studying polarized elections
- How should we sort (3,2,4,7)?

Visualizing clusters



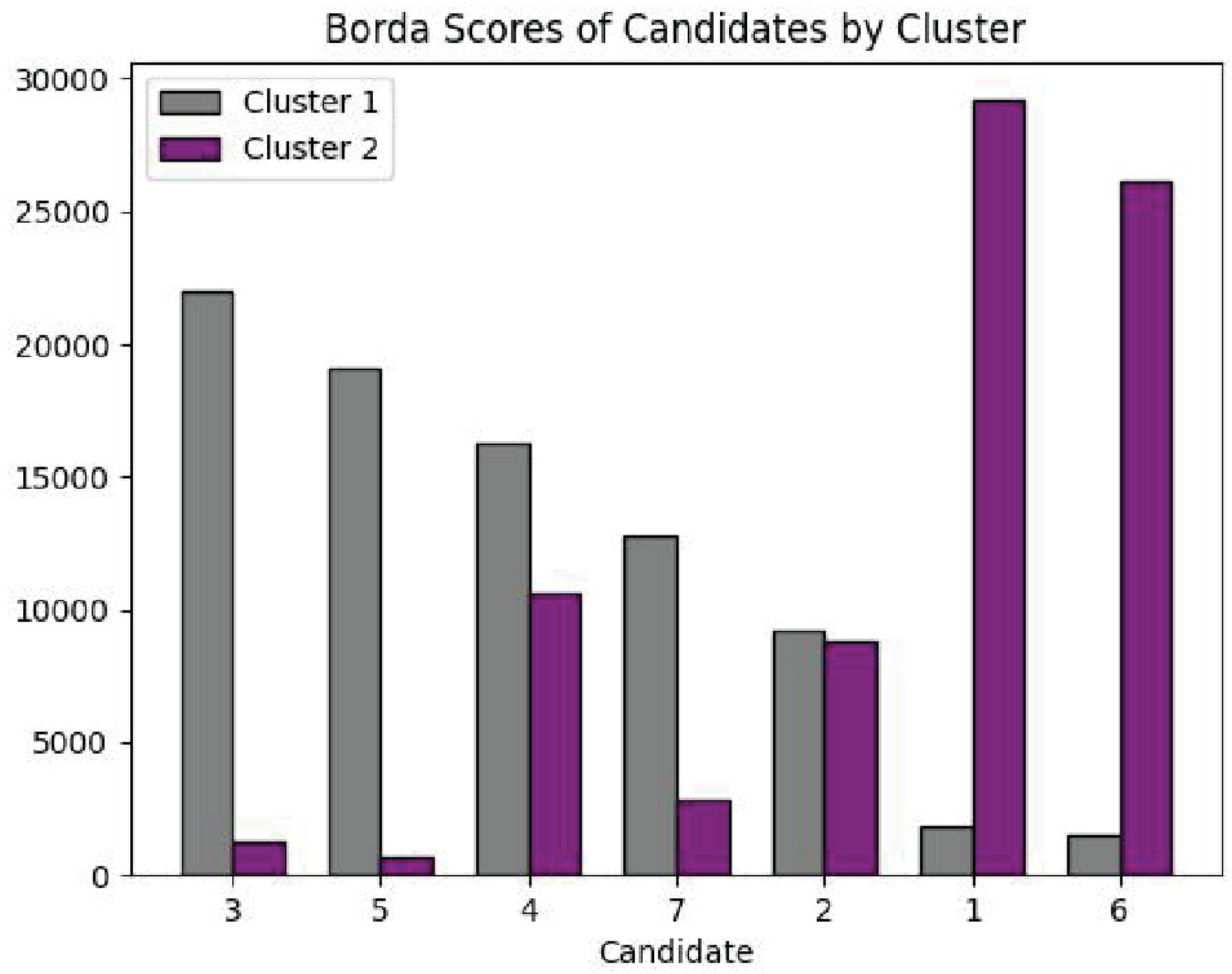
Visualizing clusters

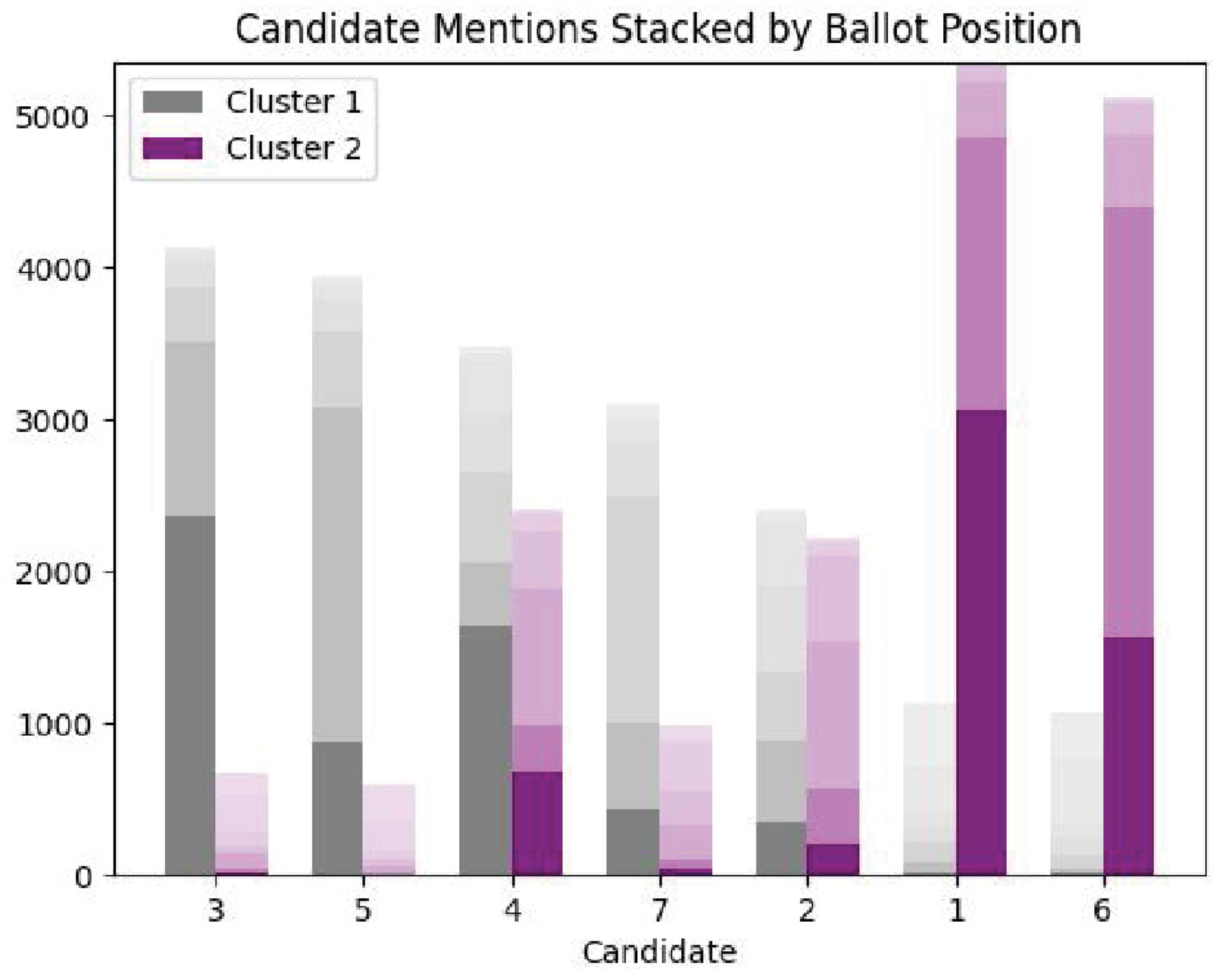
Proxy clustering (with Borda proxies and k-means)



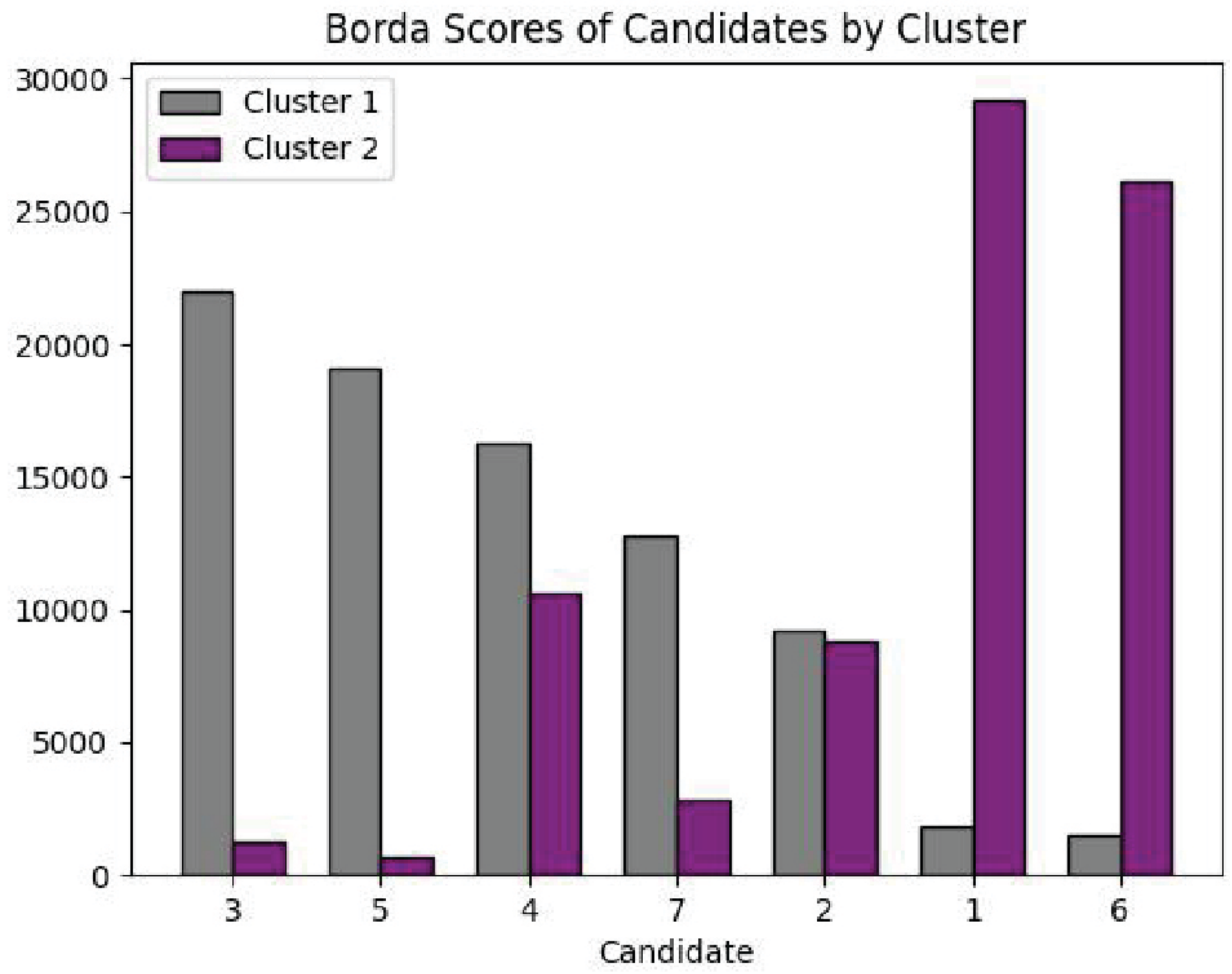
Visualizing clusters

Proxy clustering (with Borda proxies and k-means)

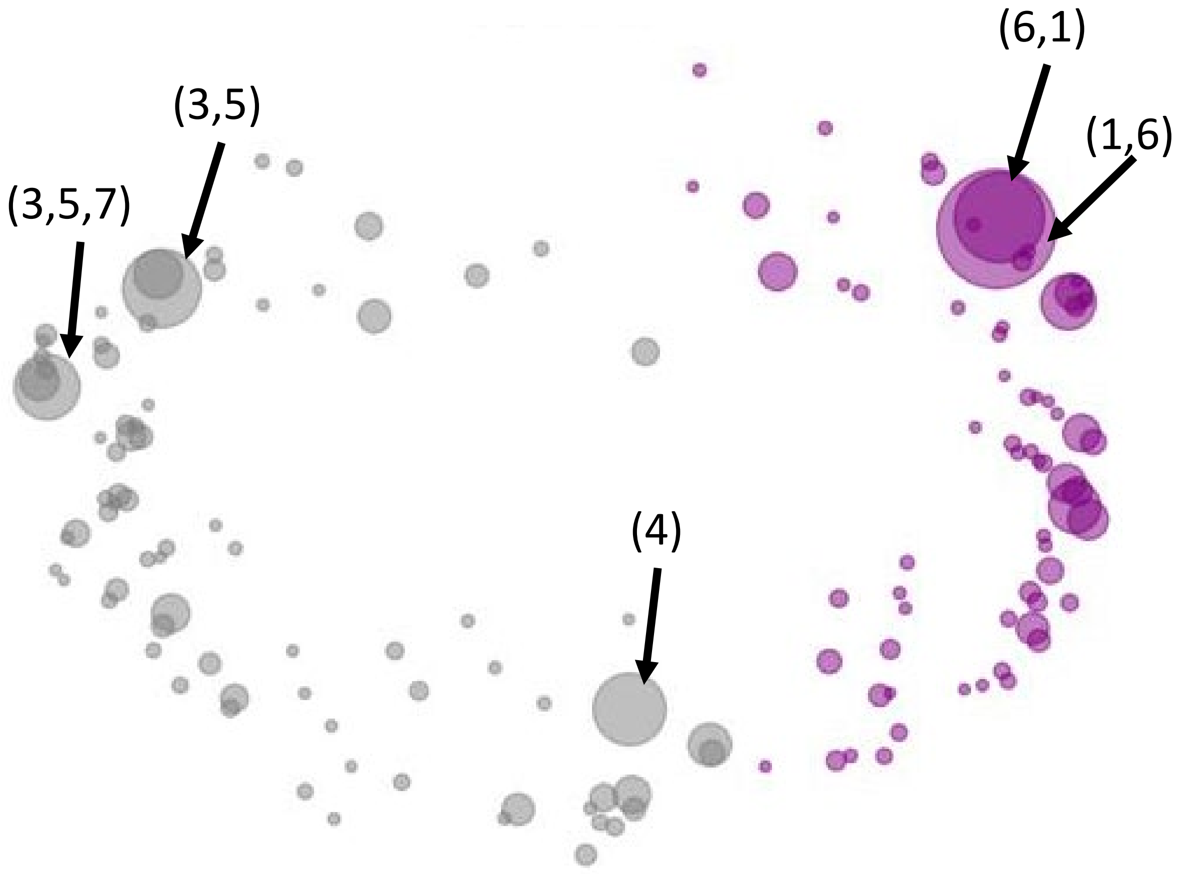




Visualizing clusters

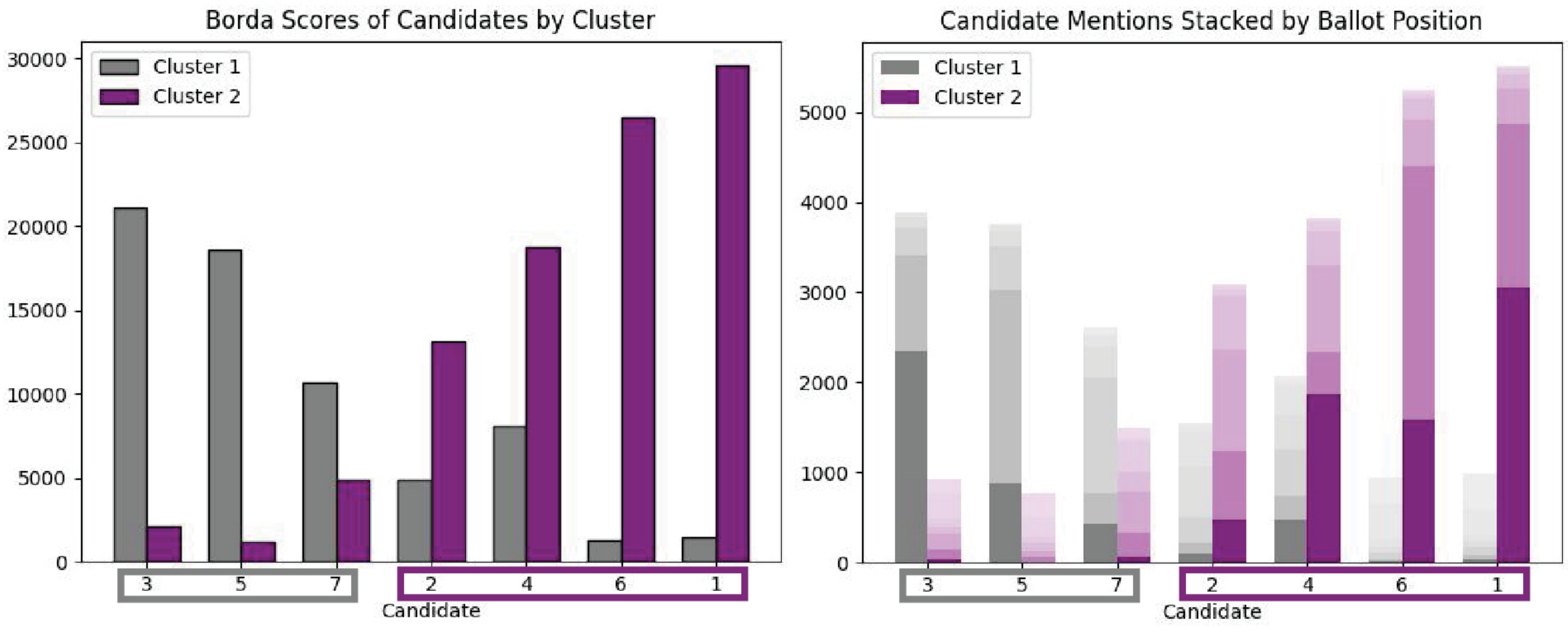


MDS Plot of ballots with more than 10 votes.



Visualizing clusters

Slate clusters



Metric space wishes

GOAL: Find a *good* metric on Ω_n = the set of possible ballots on n candidates.

WISH LIST:

Metric space wishes

GOAL: Find a *good* metric on Ω_n = the set of possible ballots on n candidates.

WISH LIST:

- Interpretable as a natural and intuitive “similarity” of the ballots.

Metric space wishes

GOAL: Find a *good* metric on Ω_n = the set of possible ballots on n candidates.

WISH LIST:

- Interpretable as a natural and intuitive "similarity" of the ballots.
- Realizable as the distance between proxy points in coordinate space *(Euclidean or Manhattan)*

Metric space wishes

GOAL: Find a *good* metric on Ω_n = the set of possible ballots on n candidates.

WISH LIST:

- Interpretable as a natural and intuitive "similarity" of the ballots.
- Realizable as the distance between proxy points in coordinate space.
- Generalizable to ballots that allow ties.

like: $(\{A_1, A_2\}, \{B_1, B_2, B_3\}, \{C_1, C_2, C_3\})$

Unmentioned candidates are group last

Metric space wishes

GOAL: Find a *good* metric on Ω_n = the set of possible ballots on n candidates.

WISH LIST:

- Interpretable as a natural and intuitive "similarity" of the ballots.
- Realizable as the distance between proxy points in coordinate space.
- Generalizable to ballots that allow ties.

like: $(\{A_1, A_2\}, \{B_1, B_2, B_3\}, \{C_1, C_2, C_3\})$

Unmentioned candidates are group last

or like: SLATE A > SLATE B.

Metric space wishes

GOAL: Find a *good* metric on Ω_n = the set of possible ballots on n candidates.

WISH LIST:

- Interpretable as a natural and intuitive "similarity" of the ballots.
- Realizable as the distance between proxy points in coordinate space.
- Generalizable to ballots that allow ties.
- Based on familiar concepts like head-to-head comparisons or Borda points.

Metric space wishes

GOAL: Find a *good* metric on Ω_n = the set of possible ballots on n candidates.

WE'LL STUDY TWO METRICS:

- d_H = *head-to-head distance*

- d_B = *Borda distance*

Metric space wishes

GOAL: Find a *good* metric on Ω_n $=$ the set of possible ballots on n candidates.

WE'LL STUDY TWO METRICS:

- d_H = *head-to-head distance*
  = ballot graph distance
  = weighted disagreement count
  = distance between head-to-head proxies

- d_B = *Borda distance*
  = *shift count*
  = *distance between Borda proxies*
  = *distance in augmented ballot graph*

Metric space wishes

GOAL: Find a *good* metric on $\Omega_n =$ the set of possible ballots on n candidates.

WE'LL STUDY TWO METRICS:

- d_H = *head-to-head distance*
  = ballot graph distance
  = weighted disagreement count
  = distance between head-to-head proxies

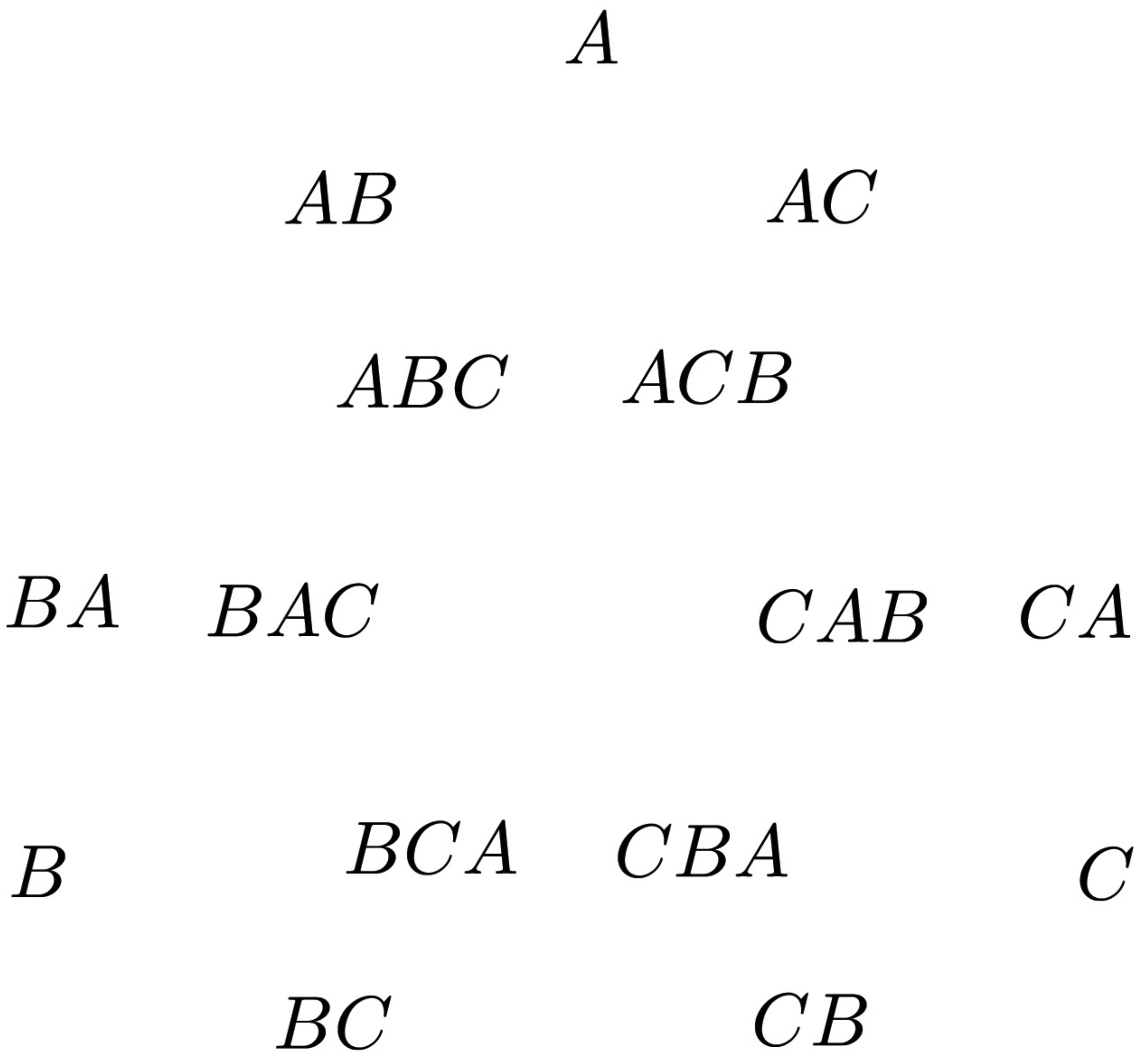
- d_B = *Borda distance*
  = *shift count*
  = *distance between Borda proxies*
  = *distance in augmented ballot graph*

And we'll show that d_H and d_B are very similar.

The Ballot Graph

The Ballot Graph

- Nodes = all possible ballots



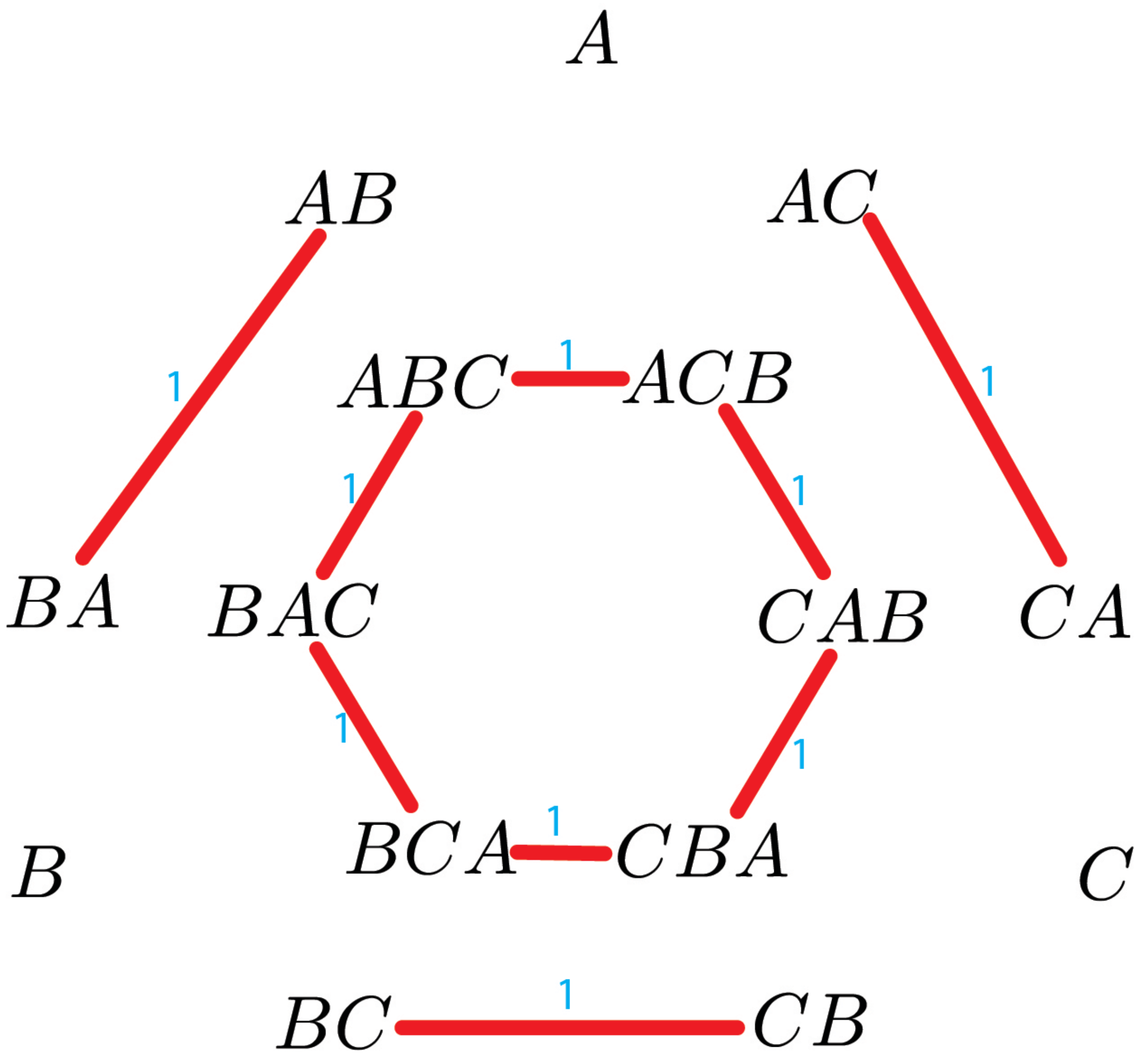
The ballot graph for Ω_3

The Ballot Graph

- Nodes = all possible ballots
- An edge of weight 1 if related by an adjacent swap.

Example:

ABCDEF —1— ABDCEF in Ω_7



The ballot graph for Ω_3

The Ballot Graph

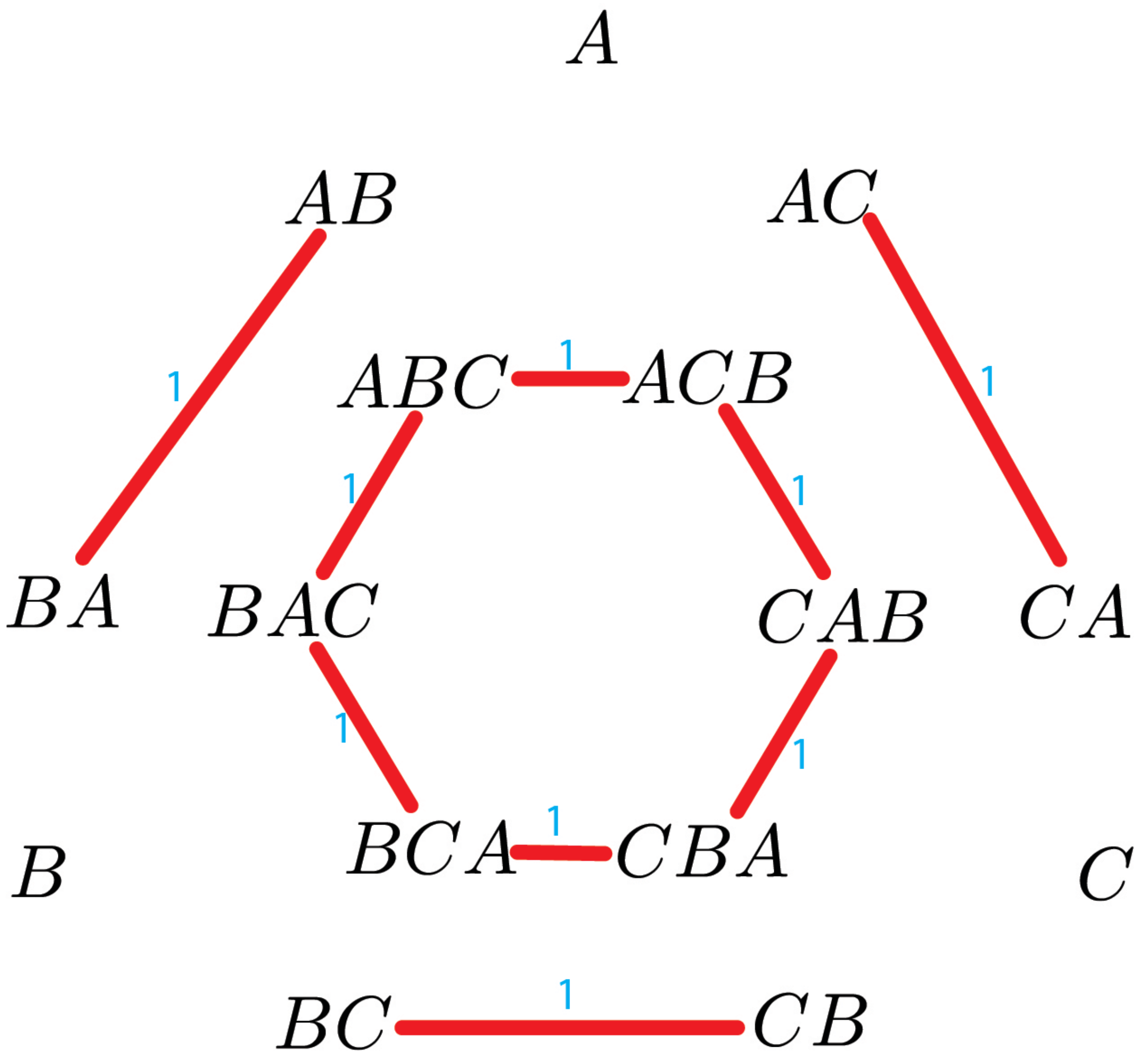
- Nodes = all possible ballots
- An edge of weight 1 if related by an adjacent swap.
- An edge if related by removing/adding final candidate.

  Weight = $\frac{\#\, missing}{2}$

Example:

AB**** --- ABC*** in Ω_6

has weight $\frac{3}{2}$



The ballot graph for Ω_3

The Ballot Graph

- Nodes = all possible ballots
- An edge of weight 1 if related by an adjacent swap.
- An edge if related by removing/adding final candidate.

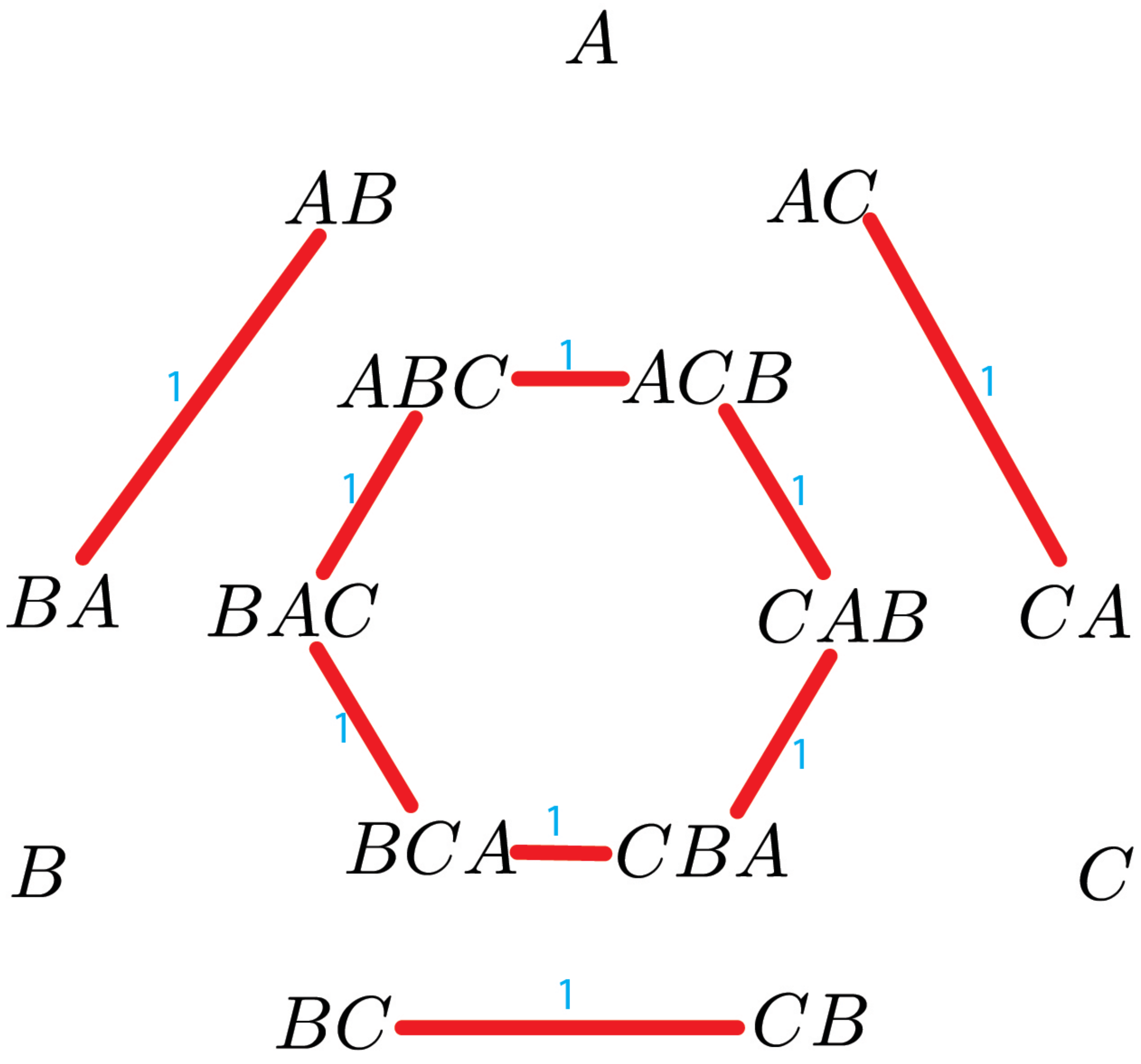
  Weight = $\frac{\#\ missing}{2}$

Example:

3 head-to-head changes

AB**** --- ABC*** in Ω_6

has weight $\frac{3}{2}$



The ballot graph for Ω_3

The Ballot Graph

- Nodes = all possible ballots
- An edge of weight 1 if related by an adjacent swap.
- An edge if related by removing/adding final candidate.

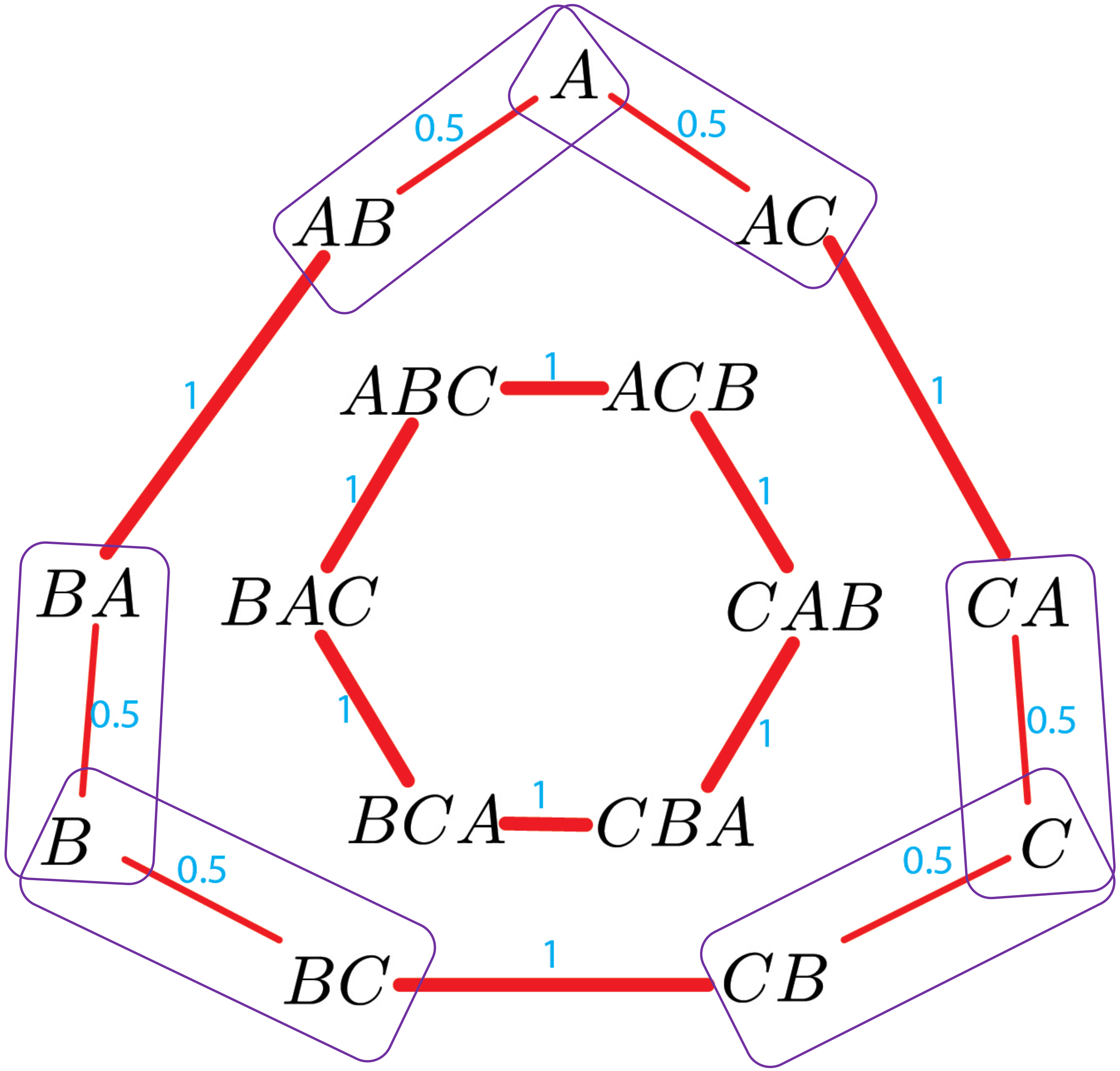
  Weight = $\frac{\#\, missing}{2}$

Example:

3 head-to-head changes

AB**** --- ABC*** in Ω_6

has weight $\frac{3}{2}$



The ballot graph for Ω_3

The Ballot Graph

- Nodes = all possible ballots
- An edge of weight 1 if related by an adjacent swap.
- An edge if related by removing/adding final candidate.

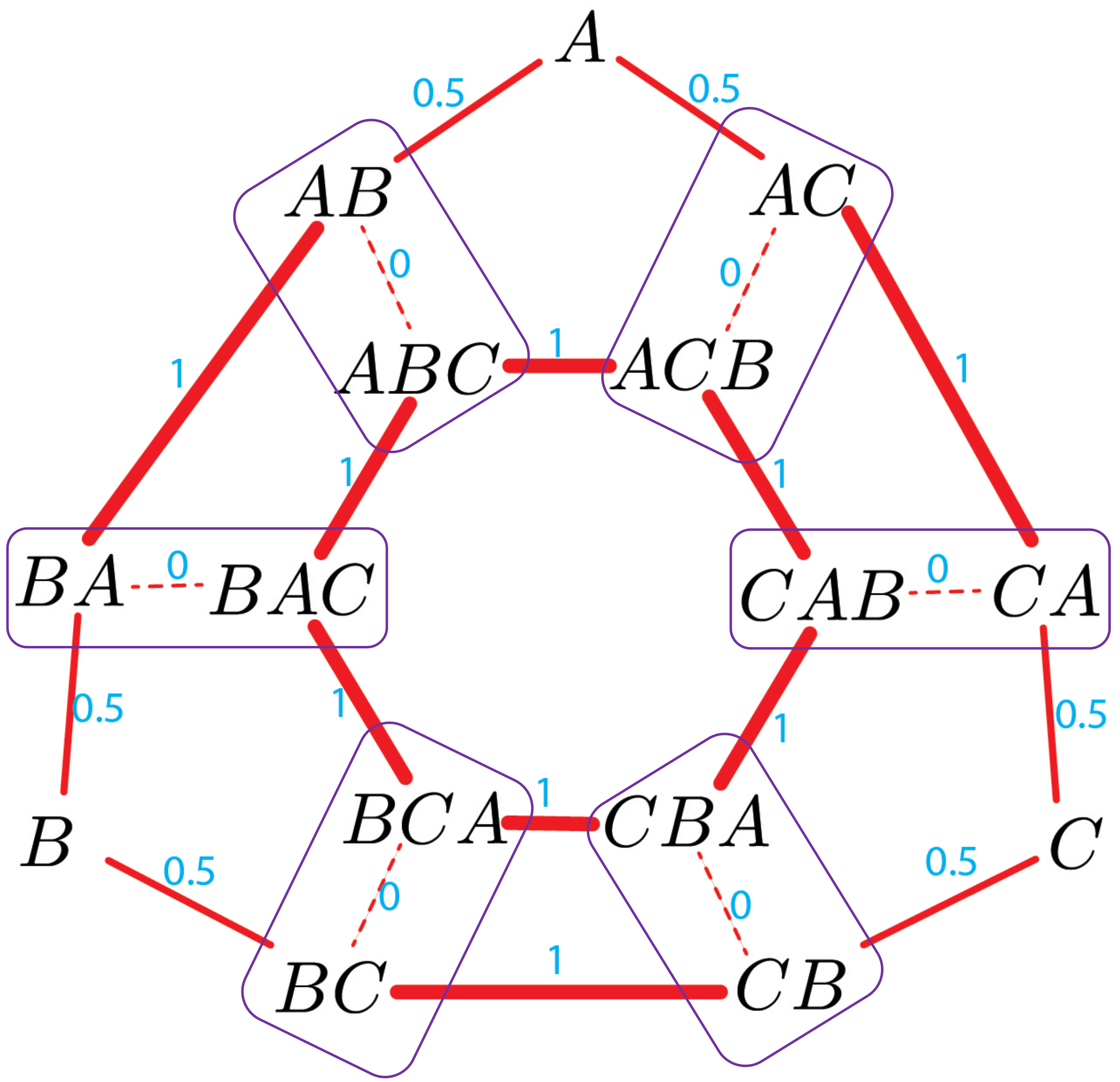
  Weight = $\frac{\#\, missing}{2}$

Example:

3 head-to-head changes

AB**** --- ABC*** in Ω_6

has weight $\frac{3}{2}$



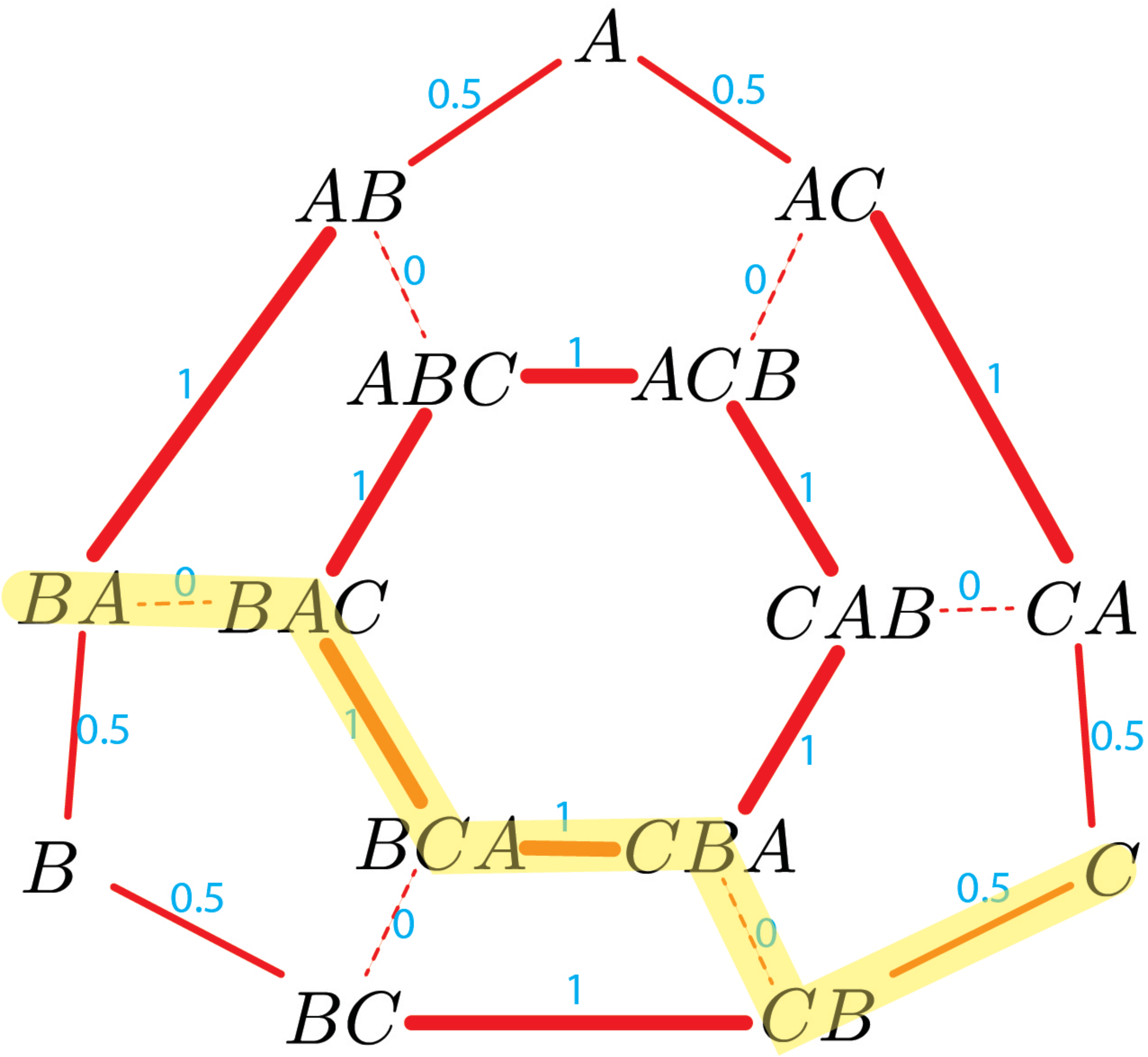
The ballot graph for Ω_3

The Ballot Graph

- Nodes = all possible ballots
- An edge of weight 1 if related by an adjacent swap.
- An edge if related by removing/adding final candidate.

  Weight = $\frac{\#\, missing}{2}$

Distance from **BA** to **C** = 2.5



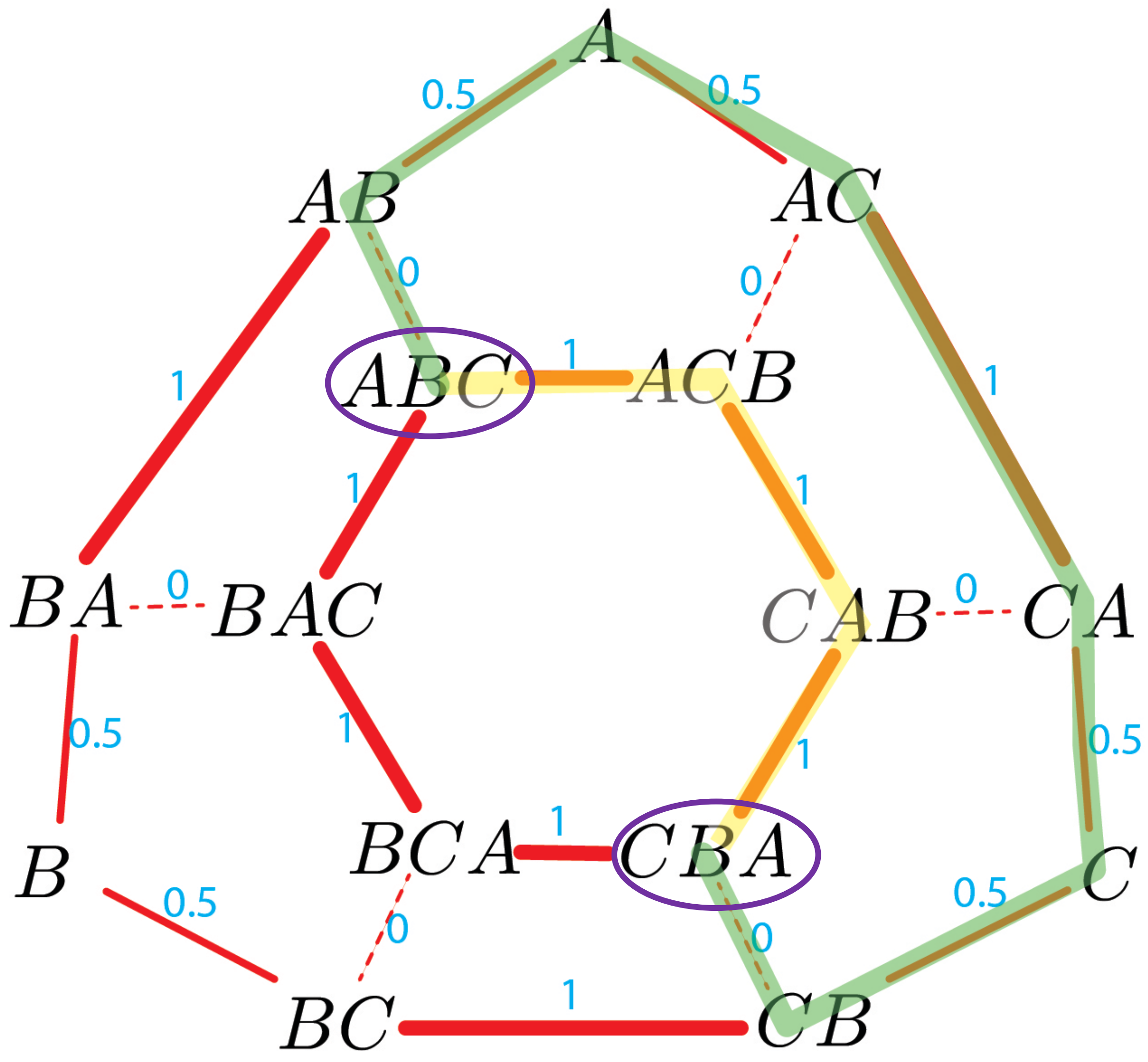
The ballot graph for Ω_3

The Ballot Graph

- Nodes = all possible ballots
- An edge of weight 1 if related by an adjacent swap.
- An edge if related by removing/adding final candidate.

  Weight = $\frac{\#\, missing}{2}$

Between a *complete* ballot and its reversal, the bullet path ties.



The ballot graph for Ω_3

The weighted disagreement count

The weighted disagreement count

Consider the ballots **ACB** and **AE** in Ω_5.
Among the 10 comparisons:

AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

The weighted disagreement count

Consider the ballots **ACB** and **AE** in Ω_5.
Among the 10 comparisons:

AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

- Strong disagreements have weight 1.

The weighted disagreement count

Consider the ballots **ACB** and **AE** in Ω_5.
Among the 10 comparisons:

AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

- Strong disagreements have weight 1.
- Weak disagreements have weight $\frac{1}{2}$.

The weighted disagreement count

Consider the ballots **ACB** and **AE** in Ω_5.
Among the 10 comparisons:

AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

- Strong disagreements have weight 1.
- Weak disagreements have weight $\frac{1}{2}$.

So, distance from **ACB** to **AE** $= strong + \frac{1}{2} \cdot weak = 2 + \frac{1}{2} \cdot 4 = 4.$

The weighted disagreement count

Consider the ballots **ACB** and **AE** in Ω_5.
Among the 10 comparisons:

AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

- Strong disagreements have weight 1.
- Weak disagreements have weight $\frac{1}{2}$.

So, distance from **ACB** to **AE** $= strong + \frac{1}{2} \cdot weak = 2 + \frac{1}{2} \cdot 4 = 4.$

(We'll see that the $\frac{1}{2}$ weighting convention makes this a valid metric.)

The head-to-head proxy distance

The head-to-head proxy distance

KEY:
$.5$ = win
$-.5$ = lose
0 = tie

In Ω_5, head-to-head proxies are in 10-dimensional space.
The proxy of **ACB** is:

	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
ACB	.5	.5	.5	.5	$-.5$	.5	.5	.5	.5	0

The head-to-head proxy distance

KEY:
$.5$ = win
$-.5$ = lose
0 = tie

In Ω_5, head-to-head proxies are in 10-dimensional space.
The proxy of **ACB** is:

	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
ACB	.5	.5	.5	.5	−.5	.5	.5	.5	.5	0

The head-to-head proxy distance

KEY:
$.5$ = win
$-.5$ = lose
0 = tie

In Ω_5, head-to-head proxies are in 10-dimensional space.
The proxy of **ACB** is:

	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
ACB	.5	.5	.5	.5	−.5	.5	.5	.5	.5	0

The head-to-head proxy distance

In Ω_5, head-to-head proxies are in 10-dimensional space.
The proxy of **ACB** is:

KEY:
$.5$ = win
$-.5$ = lose
0 = tie

	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
ACB	.5	.5	.5	.5	−.5	.5	.5	.5	.5	0

The head-to-head proxy distance

In Ω_5, head-to-head proxies are in 10-dimensional space.
The proxies of **ACB** and **AE** are:

KEY:
$.5$ = win
$-.5$ = lose
0 = tie

	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
ACB	.5	.5	.5	.5	−.5	.5	.5	.5	.5	0
AE	.5	.5	.5	.5	0	0	−.5	0	−.5	−.5
dif =	0	0	0	0	-.5	.5	1	.5	1	-.5

$d_H(\boldsymbol{ACB}, \boldsymbol{AE}) =$ the Manhattan distance between their proxies = $2 + \frac{4}{2} = 4.$

The head-to-head proxy distance

Proposition 4.1. *For* $\mathcal{B}_1, \mathcal{B}_2 \in \Omega_n$*, the following are equivalent definitions of* $d_H(\mathcal{B}_1, \mathcal{B}_2)$.

1. *The ballot graph distance*
2. *The Manhattan distance between their head-to-head proxies.*
3. *The weighted disagreement count.*

The head-to-head proxy distance

Proposition 4.1. *For $\mathcal{B}_1, \mathcal{B}_2 \in \Omega_n$, the following are equivalent definitions of $d_H(\mathcal{B}_1, \mathcal{B}_2)$.*

1. *The ballot graph distance*
2. *The Manhattan distance between their head-to-head proxies.*
3. *The weighted disagreement count.*

The proposition generalizes to ballots that allow ties, like this

	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
$(\{\boldsymbol{A},\boldsymbol{E}\},\{\boldsymbol{B},\boldsymbol{C},\boldsymbol{D}\})$	.5	.5	.5	0	0	0	$-.5$	0	$-.5$	$-.5$

The head-to-head proxy distance

Proposition 4.1. *For $\mathcal{B}_1, \mathcal{B}_2 \in \Omega_n$, the following are equivalent definitions of $d_H(\mathcal{B}_1, \mathcal{B}_2)$.*

1. *The ballot graph distance*
2. *The Manhattan distance between their head-to-head proxies.*
3. *The weighted disagreement count.*

The proposition generalizes to ballots that allow ties, like this

	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
$(\{A, E\}, \{B, C, D\})$	.5	.5	.5	0	0	0	$-.5$	0	$-.5$	$-.5$

The head-to-head proxy distance

Proposition 4.1. *For $\mathcal{B}_1, \mathcal{B}_2 \in \Omega_n$, the following are equivalent definitions of $d_H(\mathcal{B}_1, \mathcal{B}_2)$.*

1. *The ballot graph distance*
2. *The Manhattan distance between their head-to-head proxies.*
3. *The weighted disagreement count.*

The proposition generalizes to ballots that allow ties, like this

	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
$(\{A, E\}, \{B, C, D\})$	.5	.5	.5	0	0	0	−.5	0	−.5	−.5

The head-to-head proxy distance

Proposition 4.1. *For $\mathcal{B}_1, \mathcal{B}_2 \in \Omega_n$, the following are equivalent definitions of $d_H(\mathcal{B}_1, \mathcal{B}_2)$.*

1. *The ballot graph distance*
2. *The Manhattan distance between their head-to-head proxies.*
3. *The weighted disagreement count.*

How is this defined for generalized ballots?

The proposition generalizes to ballots that allow ties, like this

	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
$(\{A, E\}, \{B, C, D\})$	.5	.5	.5	0	0	0	$-.5$	0	$-.5$	$-.5$

The generalized ballot graph

Half the product of the sizes of the merged sets

$$(\{A_1, A_2\}, \{B_1, B_2, B_3\}, \{C_1, C_2\}, \{D\}) \overset{\frac{6}{2}}{\text{———}} (\{A_1, A_2\}, \{B_1, B_2, B_3, C_1, C_2\}, \{D\})$$

The generalized ballot graph

Half the product of the sizes of the merged sets

$$(\{A_1, A_2\}, \{B_1, B_2, B_3\}, \{C_1, C_2\}, \{D\}) \overset{\frac{6}{2}}{\text{———}} (\{A_1, A_2\}, \{B_1, B_2, B_3, C_1, C_2\}, \{D\})$$

Adjacent candidate swaps have weight 1, as before:

$$(\{A_1, A_2\}, \{B\}, \{C\}, \{D\}) \overset{\frac{1}{2}}{\text{———}} (\{A_1, A_2\}, \{B, C\}, \{D\}) \overset{\frac{1}{2}}{\text{———}} (\{A_1, A_2\}, \{C\}, \{B\}, \{D\})$$

The generalized ballot graph

Half the product of the sizes of the merged sets

$$(\{A_1, A_2\}, \{B_1, B_2, B_3\}, \{C_1, C_2\}, \{D\}) \overset{\frac{6}{2}}{\text{———}} (\{A_1, A_2\}, \{B_1, B_2, B_3, C_1, C_2\}, \{D\})$$

Adjacent candidate swaps have weight 1, as before:

$$(\{A_1, A_2\}, \{B\}, \{C\}, \{D\}) \overset{\frac{1}{2}}{\text{———}} (\{A_1, A_2\}, \{B, C\}, \{D\}) \overset{\frac{1}{2}}{\text{———}} (\{A_1, A_2\}, \{C\}, \{B\}, \{D\})$$

and truncation has the same weight as before:

$$(\{A_1, A_2\}, \{B_1, B_2\}, \underbrace{\{C_1, C_2, C_3, C_4\}}_{\text{Missing from ballot}}) \overset{\frac{3}{2}}{\text{———}} (\{A_1, A_2\}, \{B_1, B_2\}, \{C_1\}, \underbrace{\{C_2, C_3, C_4\}}_{\text{Missing from ballot}})$$

The Borda proxy distance

The Borda proxy distance

In Ω_5, Borda proxies are in 5-dimensional space.
The proxies of **ACB** is:

KEY:

____	____	____	____	___
4	3	2	1	0

	A	B	C	D	E
ACB	4	2	3	0	0

The Borda proxy distance

In Ω_5, Borda proxies are in 5-dimensional space.
The proxies of **ACB** and **AE** are:

KEY:

___	___	___	___	___
4	3	2	1	0

	A	B	C	D	E
ACB	4	2	3	0	0
AE	4	0	0	0	3
$dif\ =$	*0*	*2*	*3*	*0*	*-3*

$d_B(\boldsymbol{ACB}, \boldsymbol{AE}) =$ the Manhattan distance between their proxies $= 8$.

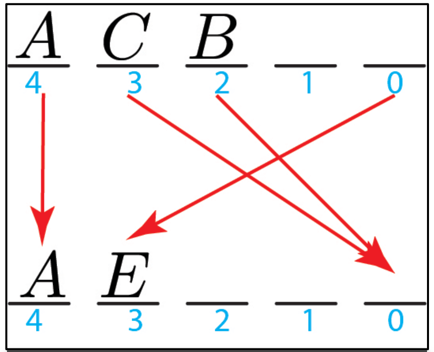
The Borda proxy distance

In Ω_5, Borda proxies are in 5-dimensional space.
The proxies of **ACB** and **AE** are:

KEY:

___	___	___	___	___
4	3	2	1	0

	A	B	C	D	E
ACB	4	2	3	0	0
AE	4	0	0	0	3
$dif =$	*0*	*2*	*3*	*0*	*-3*



"Shift counting"

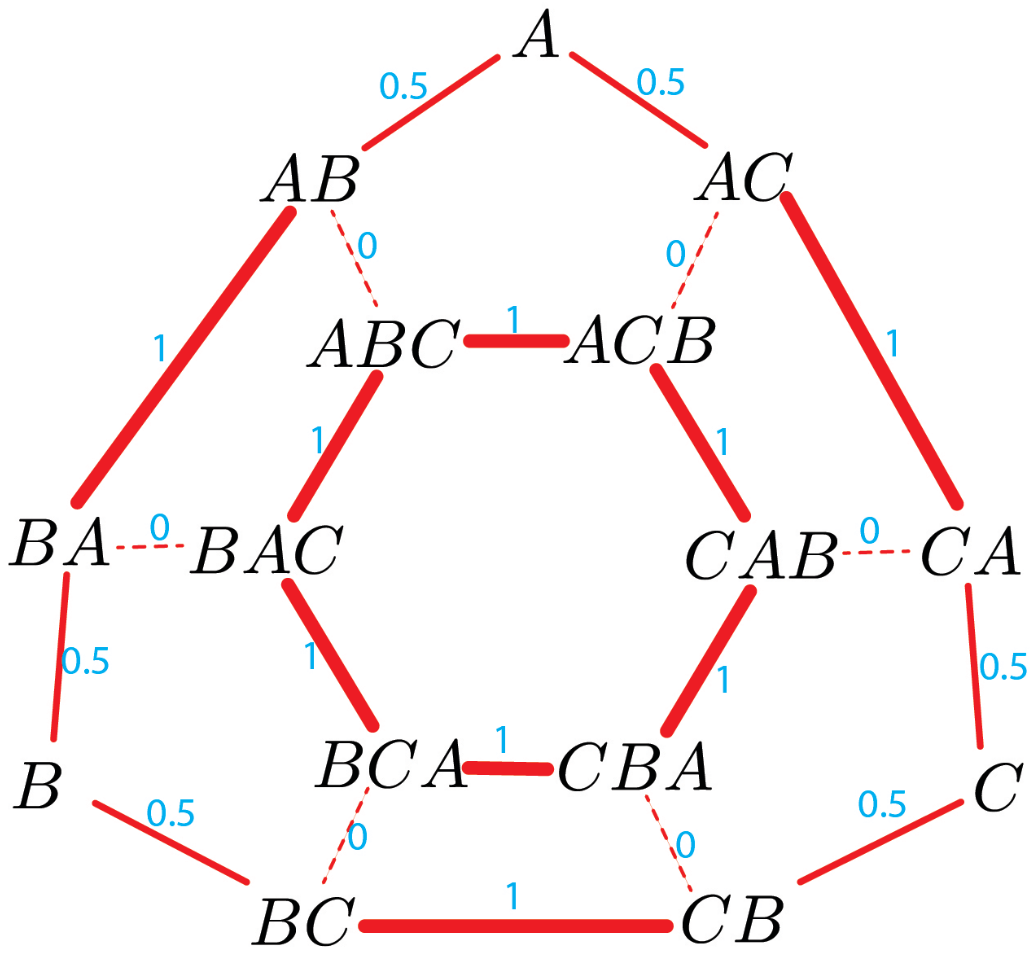
$d_B(\boldsymbol{ACB}, \boldsymbol{AE}) =$ the Manhattan distance between their proxies $= 8$.

Borda vs. head-to-head distance

- $1 < \frac{d_B}{d_H} \leq 2.$

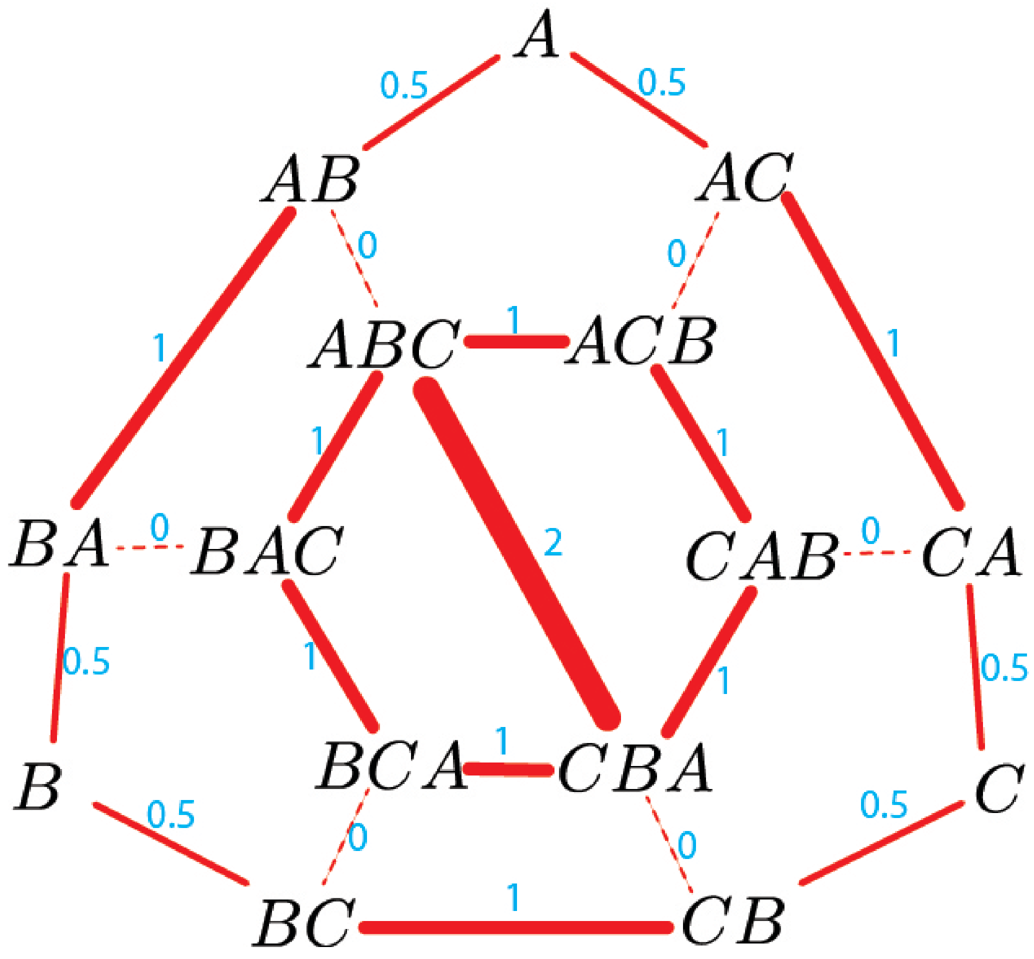
Borda vs. head-to-head distance

- $1 < \frac{d_B}{d_H} \leq 2.$
- $d_B = 2\ d_H$ for adjacent pairs.



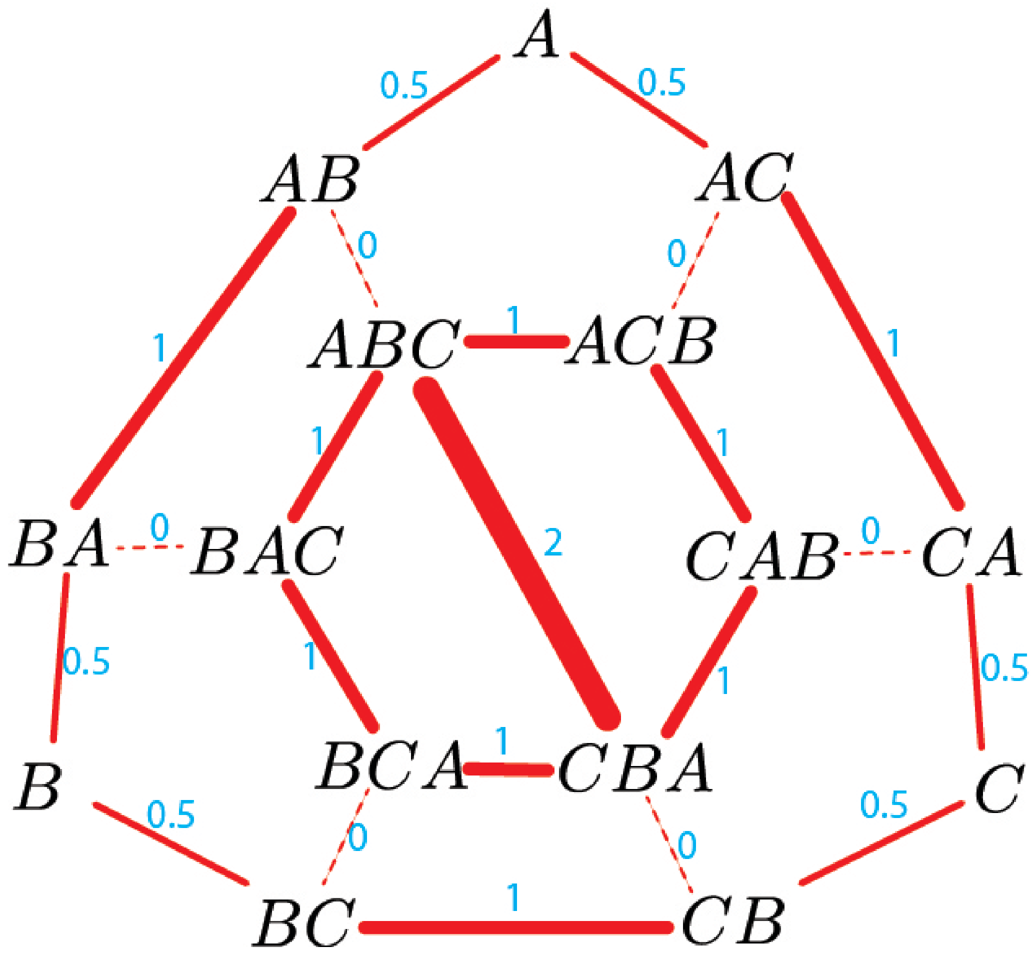
Borda vs. head-to-head distance

- $1 < \frac{d_B}{d_H} \leq 2.$
- $d_B = 2\ d_H$ for adjacent pairs.
- After adding an edge for each non-adjacent swap, d_B equals twice the graph distance.



Borda vs. head-to-head distance

- $1 < \frac{d_B}{d_H} \leq 2.$
- $d_B = 2\ d_H$ for adjacent pairs.
- After adding an edge for each non-adjacent swap, d_B equals twice the graph distance.
- *d_B penalizes strong disagreements a bit less than 2 d_H.*



Borda vs. head-to-head distance

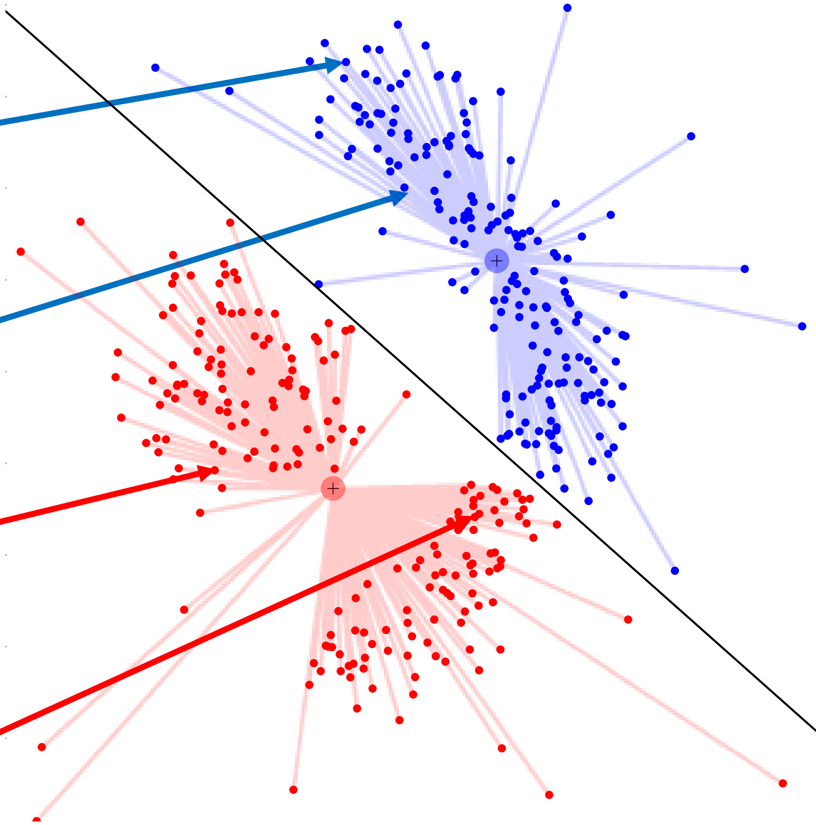
$$2\,d_H \;=2\cdot(strong)+1\cdot\;(weak)$$

$$d_B \;=\alpha\cdot(strong)+1\cdot(weak)$$

For $\alpha\in[1,2]$ that depends on the pair of ballots.

PROXY CLUSTERING: Associate each ballot to a proxy point of Euclidean space, and perform a standard clustering algorithm on the proxies.

1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).
173 votes for (4, 2).
152 votes for (6, 1, 4).
144 votes for (5, 3, 7).
136 votes for (1, 6, 4, 2).
136 votes for (3, 5, 4)....



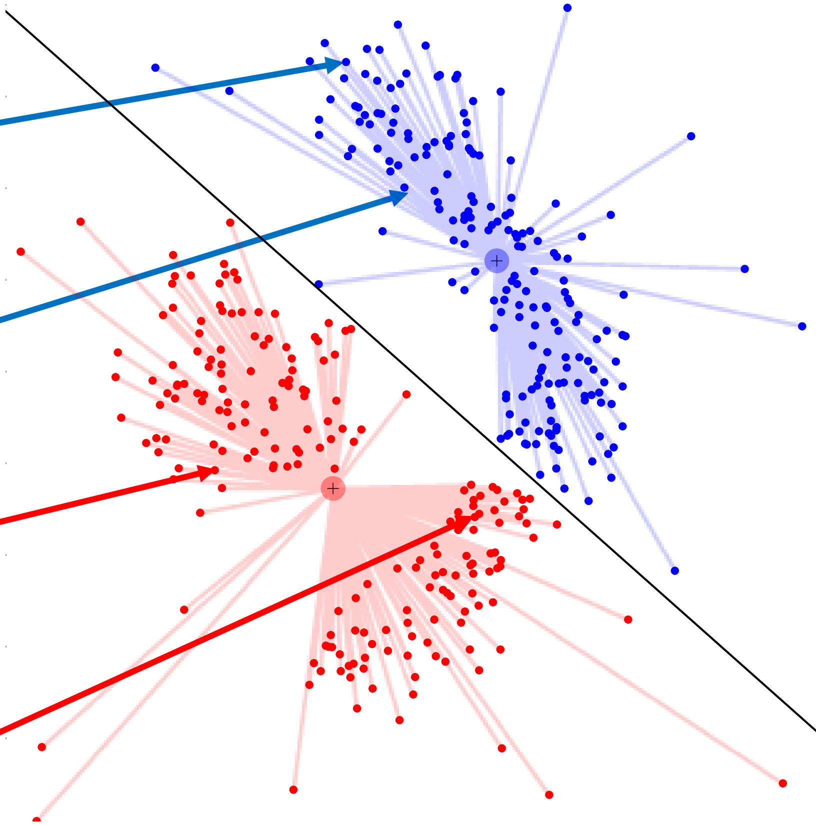
PROXY CLUSTERING: Associate each ballot to a proxy point of Euclidean space, and perform a standard clustering algorithm on the proxies.

1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).
173 votes for (4, 2).
152 votes for (6, 1, 4).
144 votes for (5, 3, 7).
136 votes for (1, 6, 4, 2).
136 votes for (3, 5, 4)....

Borda or head-to-head proxies?

k-means or k-medoids?

All choices yield about the same clusters ($\sim 4\%$)



Only slight differenced between the methods
(averaged over elections from 17 wards of Edinburgh)

Proxy cluster methods

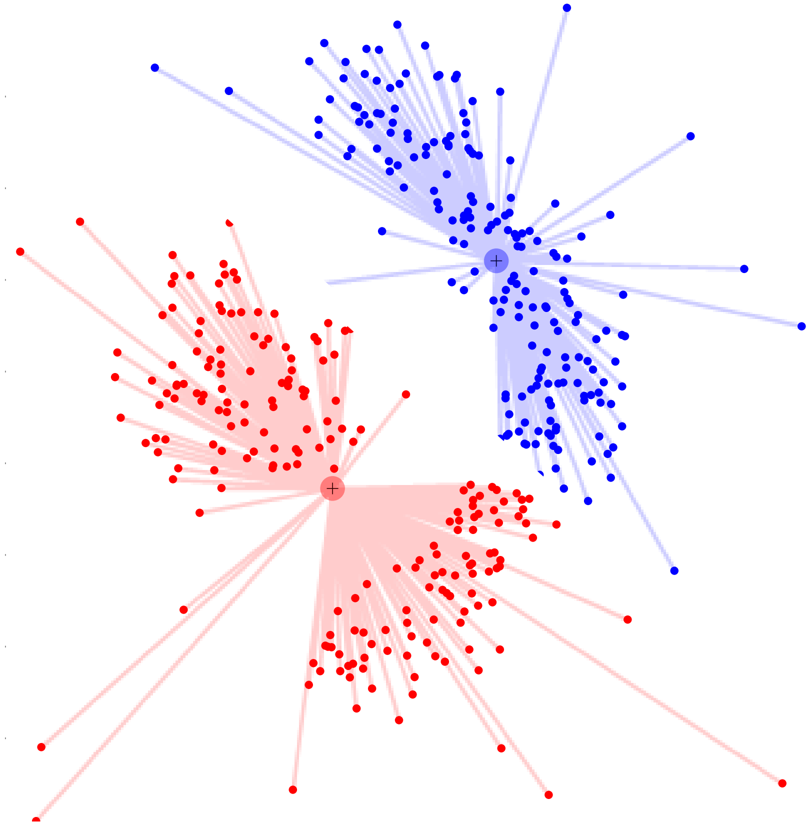
	MeanB	MeanH	MedoB	MedoH	Slate	Random
MeanB	0	0.014	0.052	0.039	0.077	0.431
MeanH	-	0	0.050	0.036	0.081	0.431
MedoB	-	-	0	0.032	0.076	0.429
MedoH	-	-	-	0	0.077	0.428
Slate	-	-	-	-	0	0.429
Random	-	-	-	-	-	0.459

k-means

- Center = centroid (center of mass)
- Minimizes summed squared L^2 distances of data points to their centers.

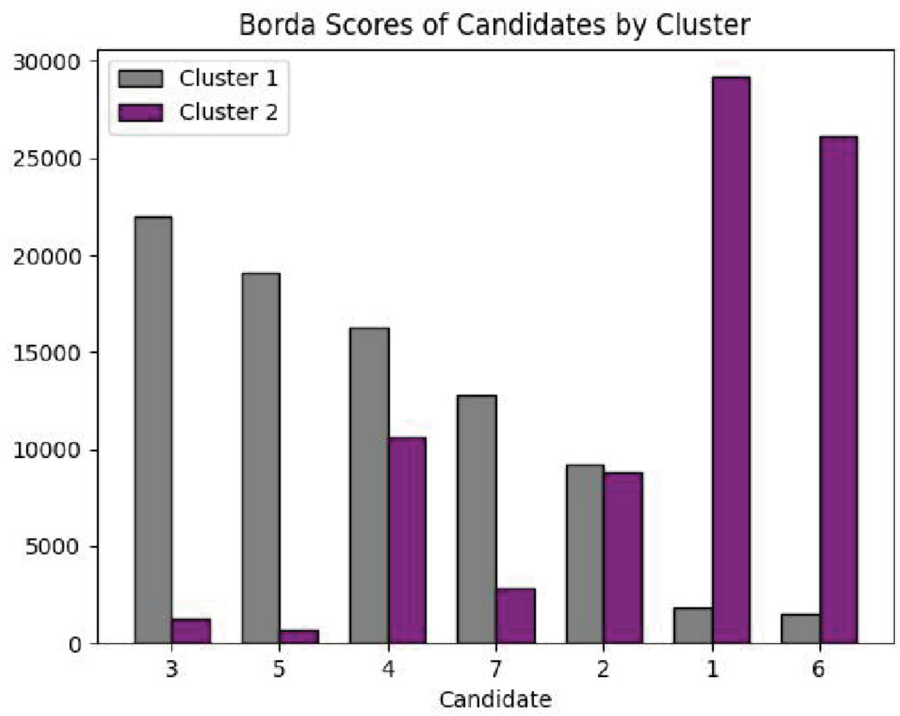
k-medoids

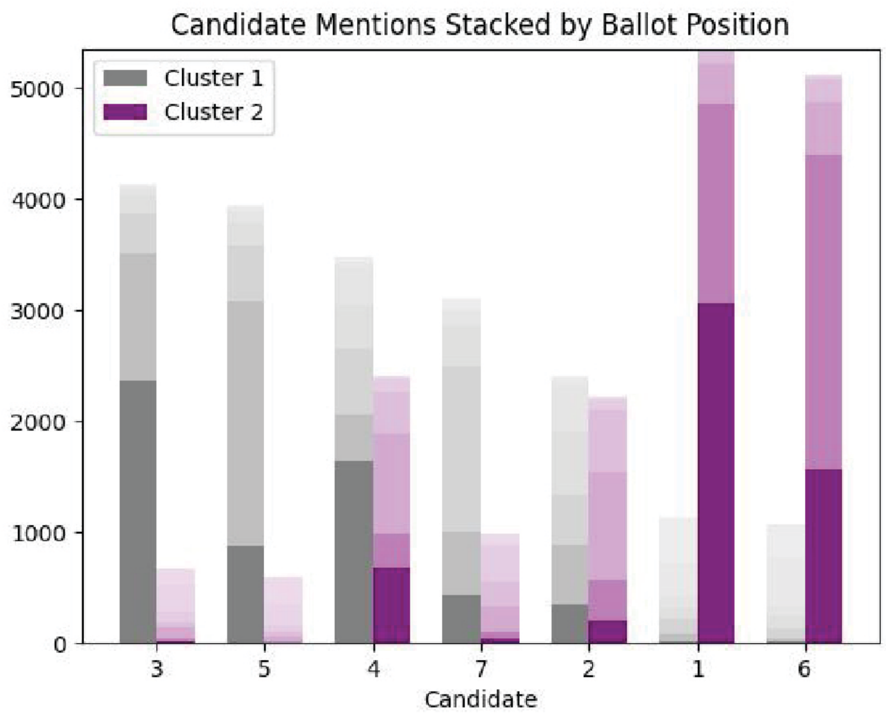
- Centers are data points
- Minimizes summed distance of data points to their center with respect to arbitrary metric (we use L^1).



2-medoid clusters with head-to-head proxies.

Medoids: (3,5,7) and (1,6)





SLATE CLUSTERING: Find the partition of the candidates into two slates A,B such that the ballots are most starkly divided into “A>B types” and “B>A types”. Partition the ballots accordingly.

1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).
173 votes for (4, 2).
152 votes for (6, 1, 4).
144 votes for (5, 3, 7).
136 votes for (1, 6, 4, 2).
136 votes for (3, 5, 4)….

SLATE A:
3 = Neil Gardiner (SNP)
5 = Ernesta Noreikiene (SNP)
7 = Evelyn Weston (Grn)

SLATE B:
1 = Graeme Bruce (C)
2 = Emma Farthing (LD)
4 = Ricky Henderson (Lab)
6= Susan Webber (C)

SLATE CLUSTERING: Find the partition of the candidates into two slates A,B such that the ballots are most starkly divided into “A>B types” and “B>A types”. Partition the ballots accordingly.

1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).

SLATE A:
3 = Neil Gardiner (SNP)
5 = Ernesta Noreikiene (SNP)
7 = Evelyn Weston (Grn)

SLATE B:
1 = Graeme Bruce (C)
2 = Emma Farthing (LD)
4 = Ricky Henderson (Lab)
6= Susan Webber (C)

METHOD: Exhaustively try all partitions to find the one with the best score.

SCORE = the sum of the distances of the ballots
to “A>B” or “B>A” (whichever is closest)

SLATE CLUSTERING: Find the partition of the candidates into two slates A,B such that the ballots are most starkly divided into “A>B types” and “B>A types”. Partition the ballots accordingly.

1342 votes for (1, 6).
759 votes for (6, 1).
578 votes for (3, 5).
494 votes for (4).
403 votes for (3, 5, 7).
285 votes for (1, 6, 2).
254 votes for (1, 6, 4).
219 votes for (5, 3).

SLATE A:
3 = Neil Gardiner (SNP)
5 = Ernesta Noreikiene (SNP)
7 = Evelyn Weston (Grn)

SLATE B:
1 = Graeme Bruce (C)
2 = Emma Farthing (LD)
4 = Ricky Henderson (Lab)
6= Susan Webber (C)

METHOD: Exhaustively try all partitions to find the one with the best score.

SCORE = the sum of the distances of the ballots
to “A>B” or “B>A” (whichever is closest)

	A_1A_2	B_1B_2	B_1B_3	B_2B_3	A_1B_1	A_1B_2	A_1B_3	A_2B_1	A_2B_2	A_2B_3
$A_1B_2A_2$	.5	-.5	0	.5	.5	.5	.5	.5	-.5	.5
$\{A_1, A_2\} > \{B_1, B_2, B_3\}$	0	0	0	0	.5	.5	.5	.5	.5	.5
$dif\ =$	*.5*	*-.5*	*0*	*.5*	*0*	*0*	*0*	*0*	*1*	*0*

METHOD: Exhaustively try all partitions to find the one with the best score.

SCORE = the sum of the distances of the ballots
to “A>B” or “B>A” (whichever is closest)

	A_1A_2	B_1B_2	B_1B_3	B_2B_3	A_1B_1	A_1B_2	A_1B_3	A_2B_1	A_2B_2	A_2B_3
$A_1B_2A_2$	.5	-.5	0	.5	.5	.5	.5	.5	-.5	.5
$\{A_1, A_2\} > \{B_1, B_2, B_3\}$	0	0	0	0	.5	.5	.5	.5	.5	.5
$dif \ =$	*.5*	*-.5*	*0*	*.5*	*0*	*0*	*0*	*0*	*1*	*0*

not as relevant

or just use this part

METHOD: Exhaustively try all partitions to find the one with the best score.

SCORE = the sum of the distances of the ballots
to "A>B" or "B>A" (whichever is closest)

Slate clusters