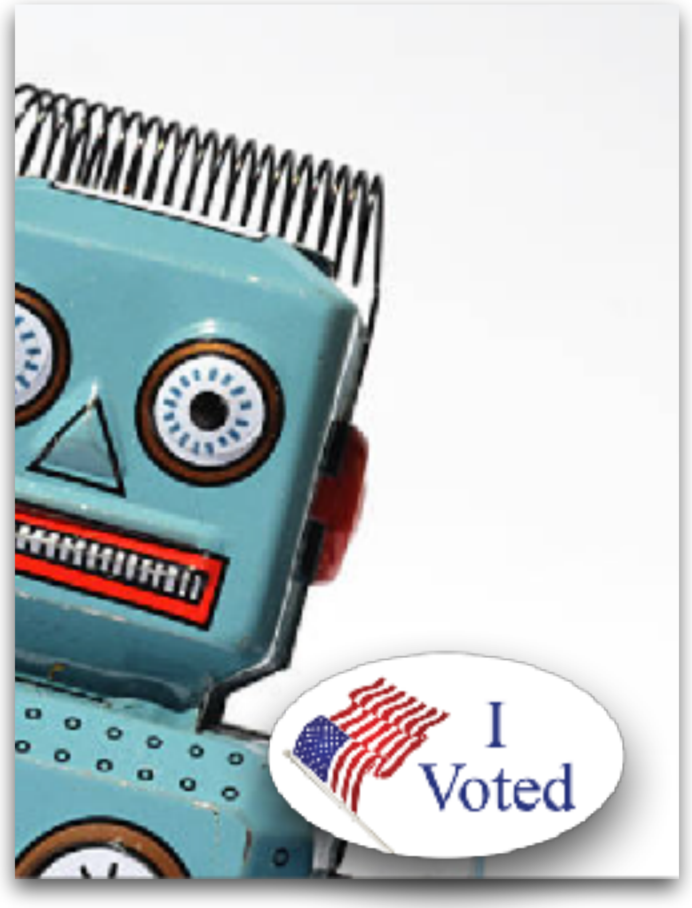
Modeling Democracy

Lecture 13 - **Applied modeling pays off!**
(**Learning blocs and slates, choosing voting rules**)



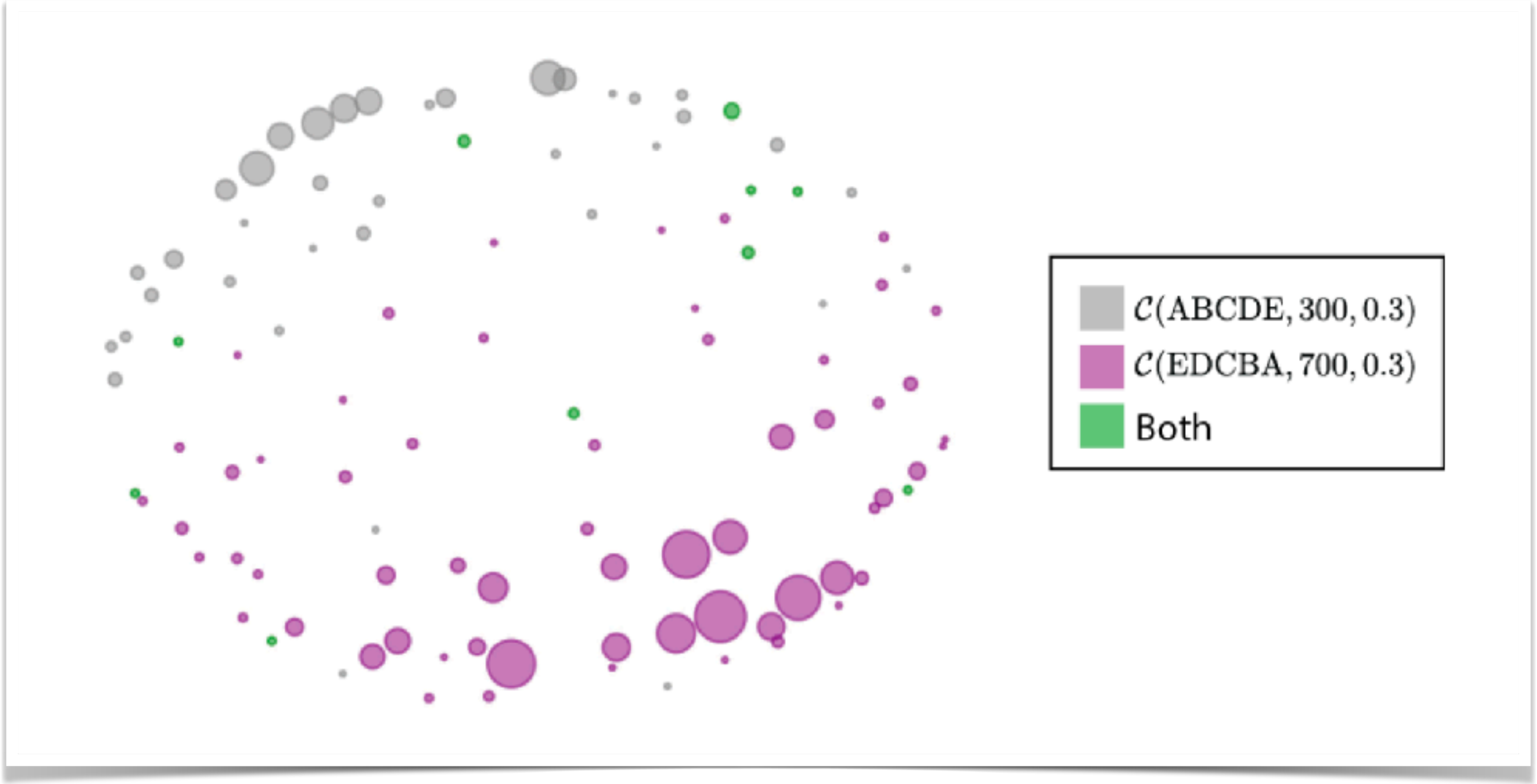
Moon Duchin, Winter 2026 — Wed Feb 18

we have ballot distances
— let's cluster voters!

Synthetic tests

Let's make random profiles with "ground truth" clusters via mixtures.

$P \sim \mathscr{C}(\sigma, N, p)$ — each of N voters takes $X \sim \mathrm{Geom}(p)$ swap steps from ballot σ



Ways to cluster:

- 2-means (heuristic L^2 min)
  - Borda or H2H
- 2-Kemeny (exact L^1 min)
  - all ballots (B/H)
  - cast ballots (B/H)
  - coordinates (B/H)

All give *exactly correct* cluster centers, only differ by assignment of points to closest centers.

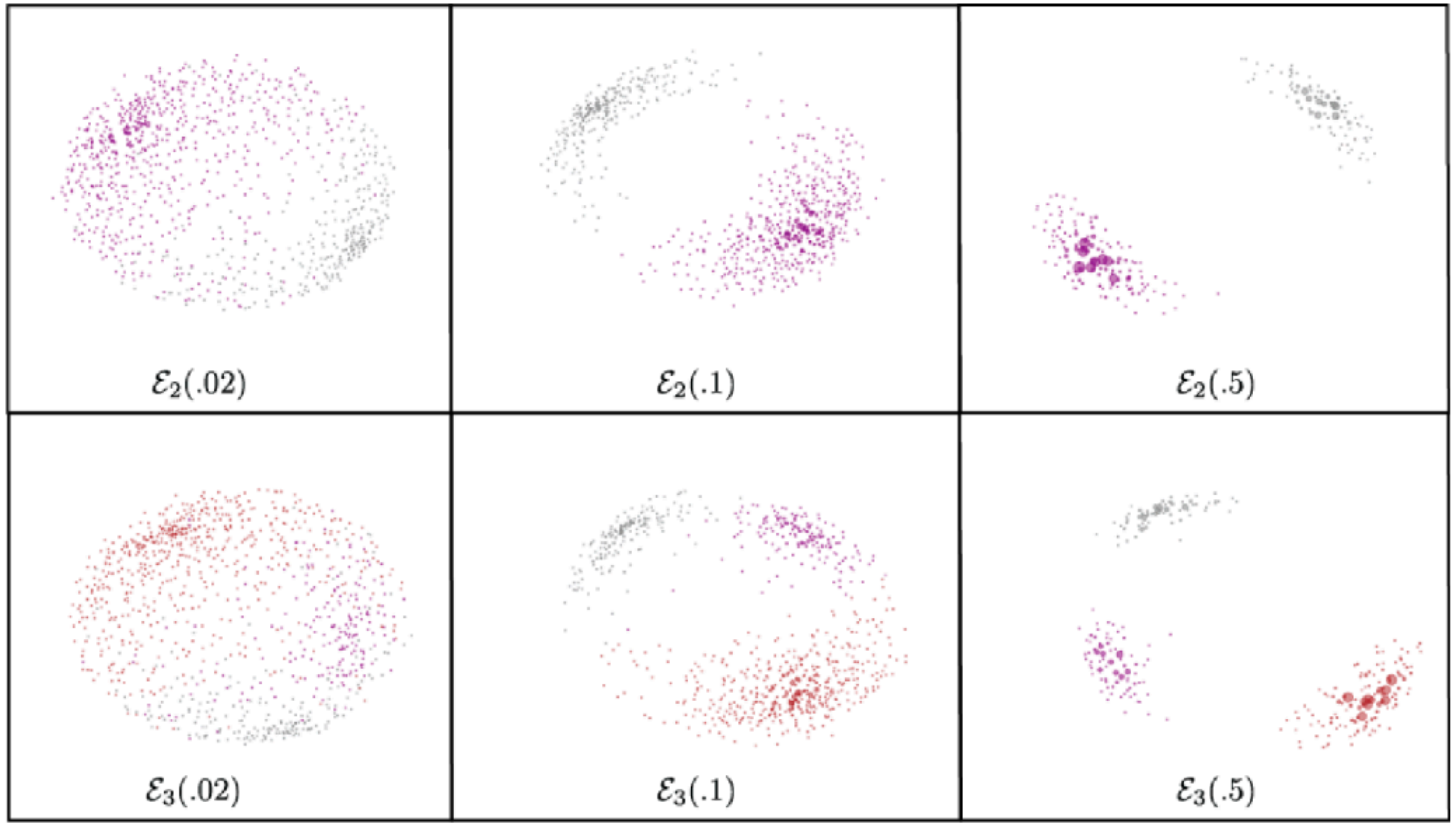
very robust!

L^2 clusters only 1% different than L^1

Note: biLip bounds imply clustering stability between B and H for sufficiently polarized elections!

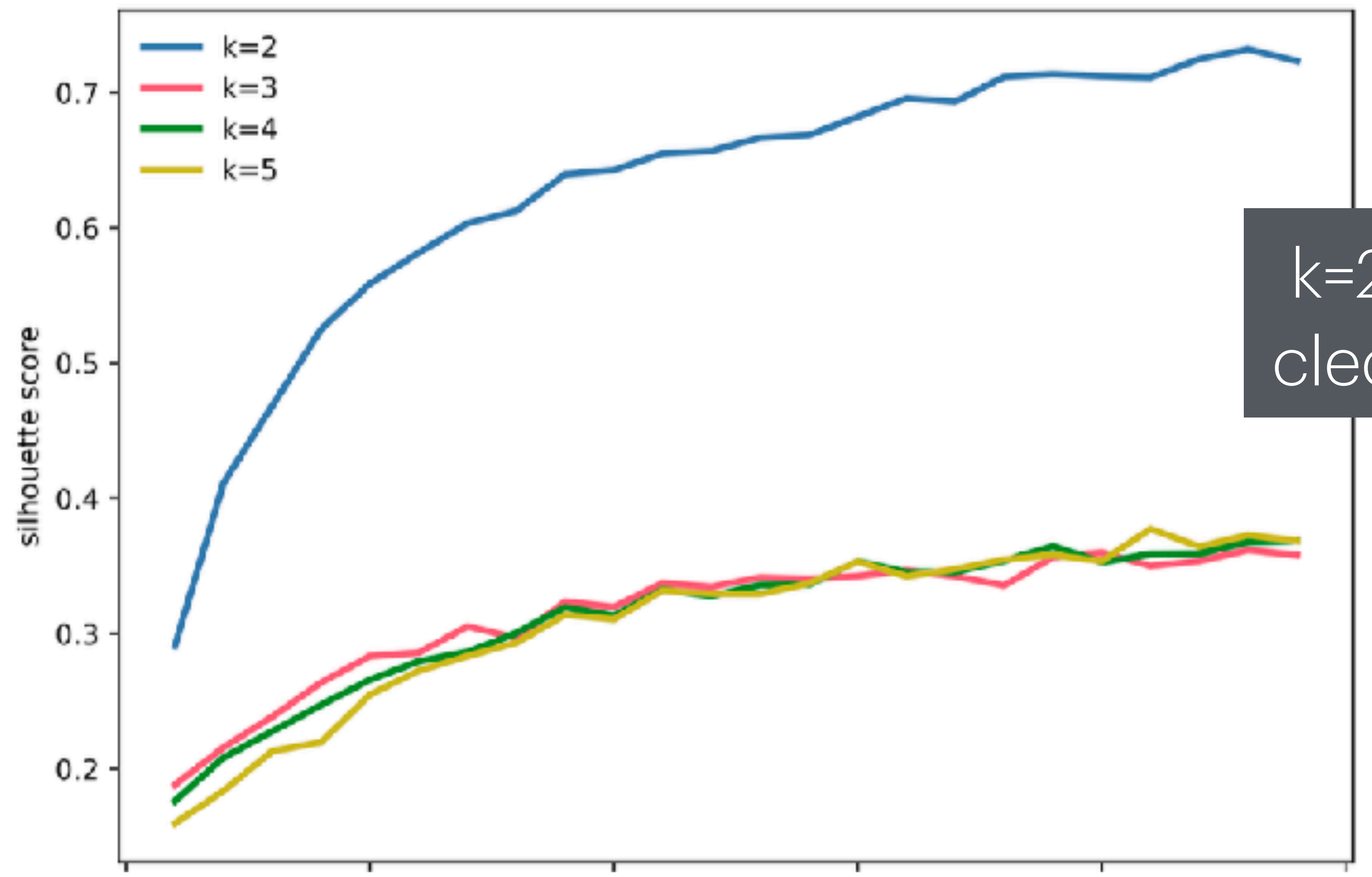
Experiments with 2-3 clusters

Silhouette score: within-cluster dists vs closest other cluster

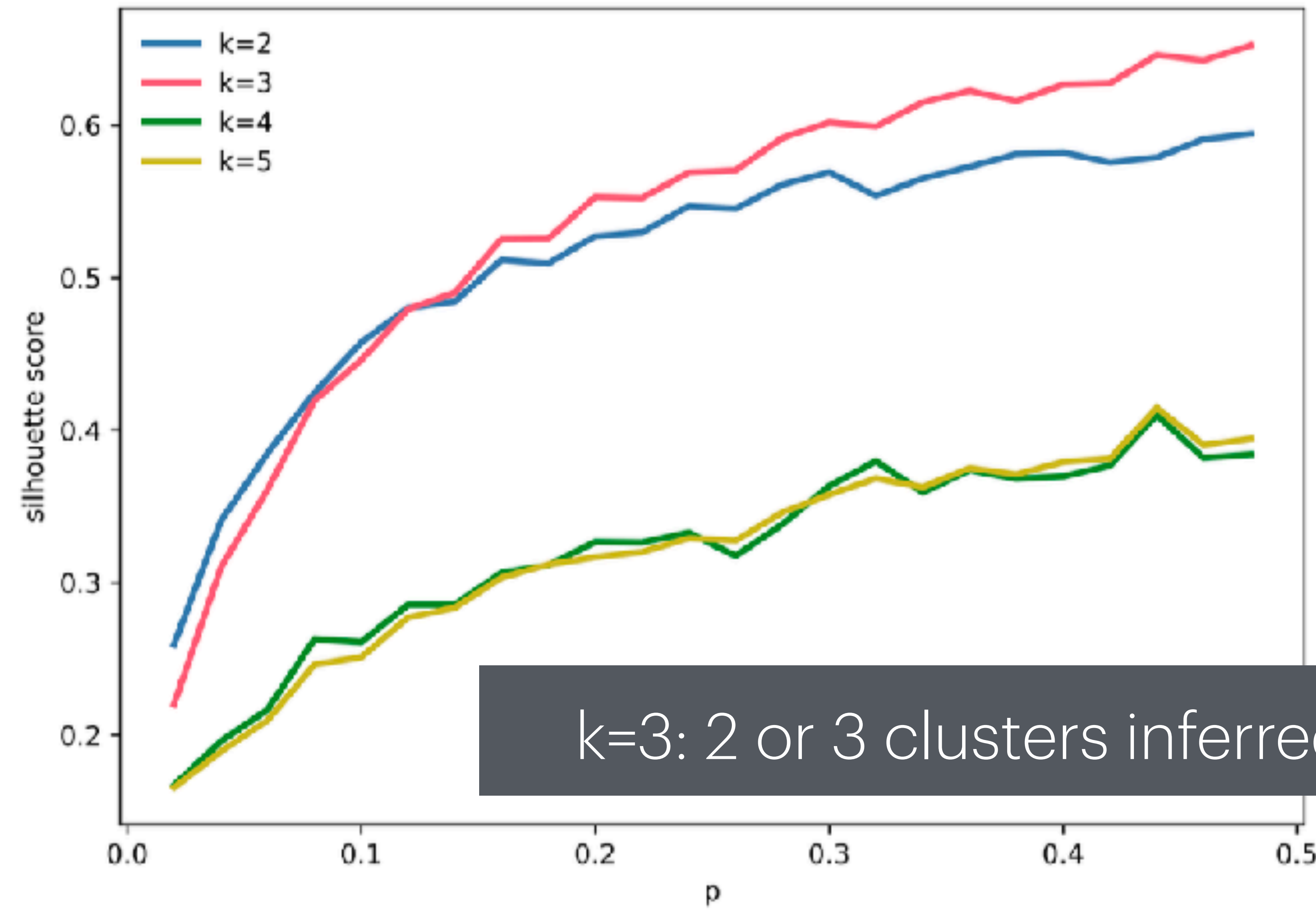


$\mathcal{E}_2(p) = C(\text{ABCDEFGHI}, 300, p) \cup C(\text{HGEIFCBAD}, 700, p);$

$\mathcal{E}_3(p) = C(\text{ABCDEFGHI}, 200, p) \cup C(\text{DFEAHBGCI}, 200, p) \cup C(\text{HIGDEFCBA}, 600, p)$



k=2: clear



k=3: 2 or 3 clusters inferred

Real data: Scottish STV

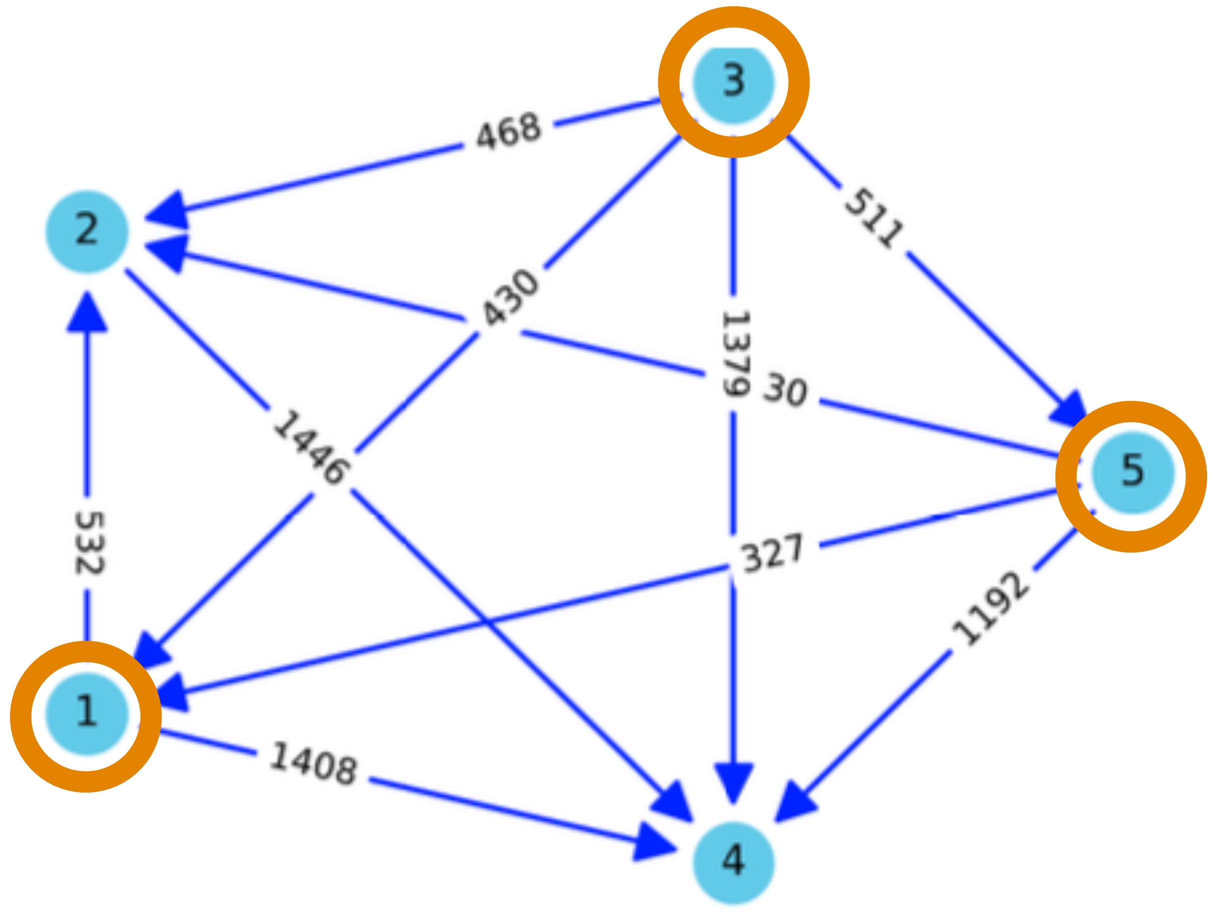
- Scotland uses STV (rankings to winner set) for local government elections: 32 localities conducting city council-style elections every 5 years, back to 2007.

- 1070 real-world preference profiles from Scotland curated by David McCune and MGGG

- Example, in Renfrewshire Ward 2, 2022, 5 candidates ran for 3 seats.

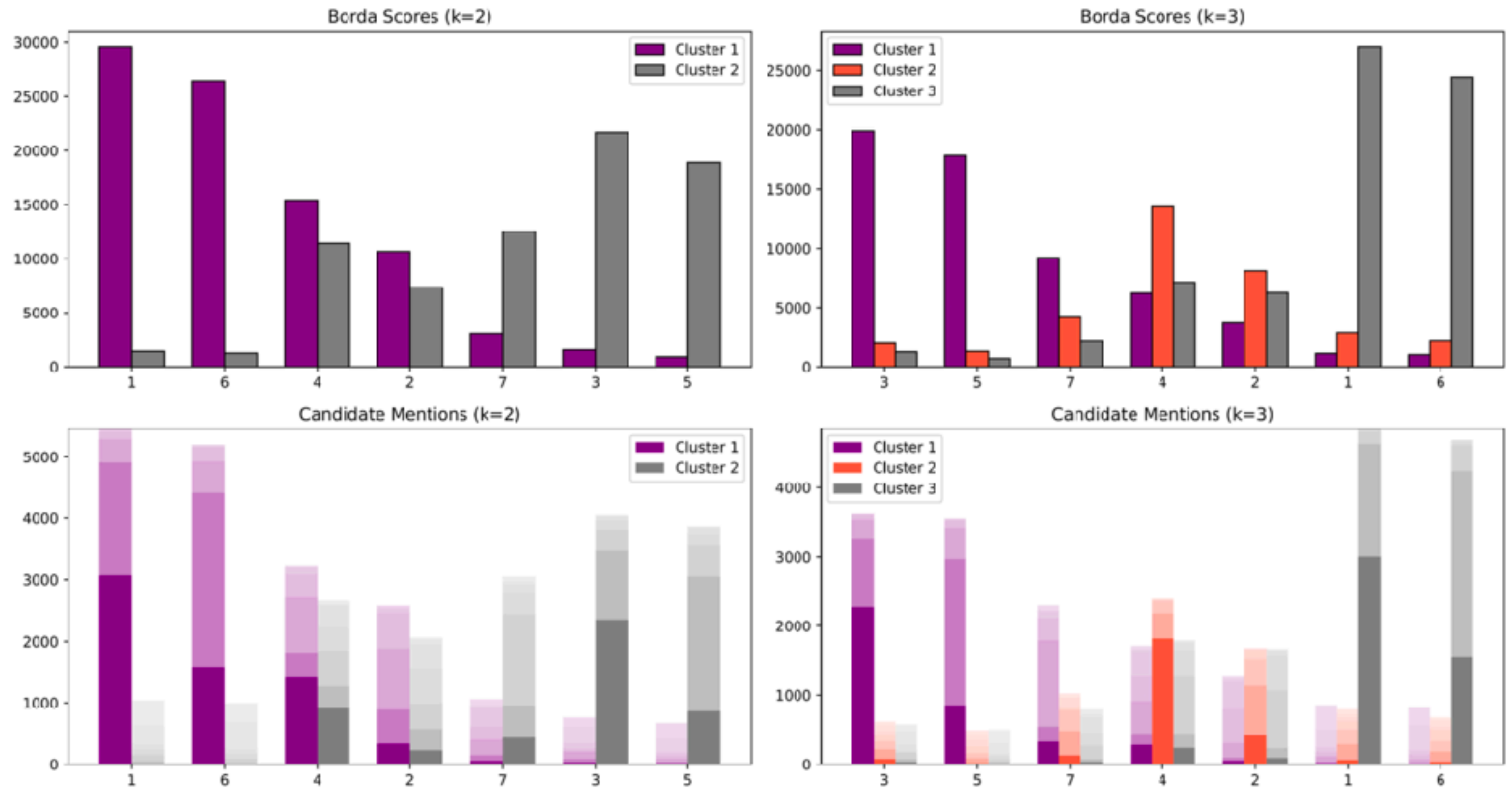
Renfrewshire Ward 2 (2022)

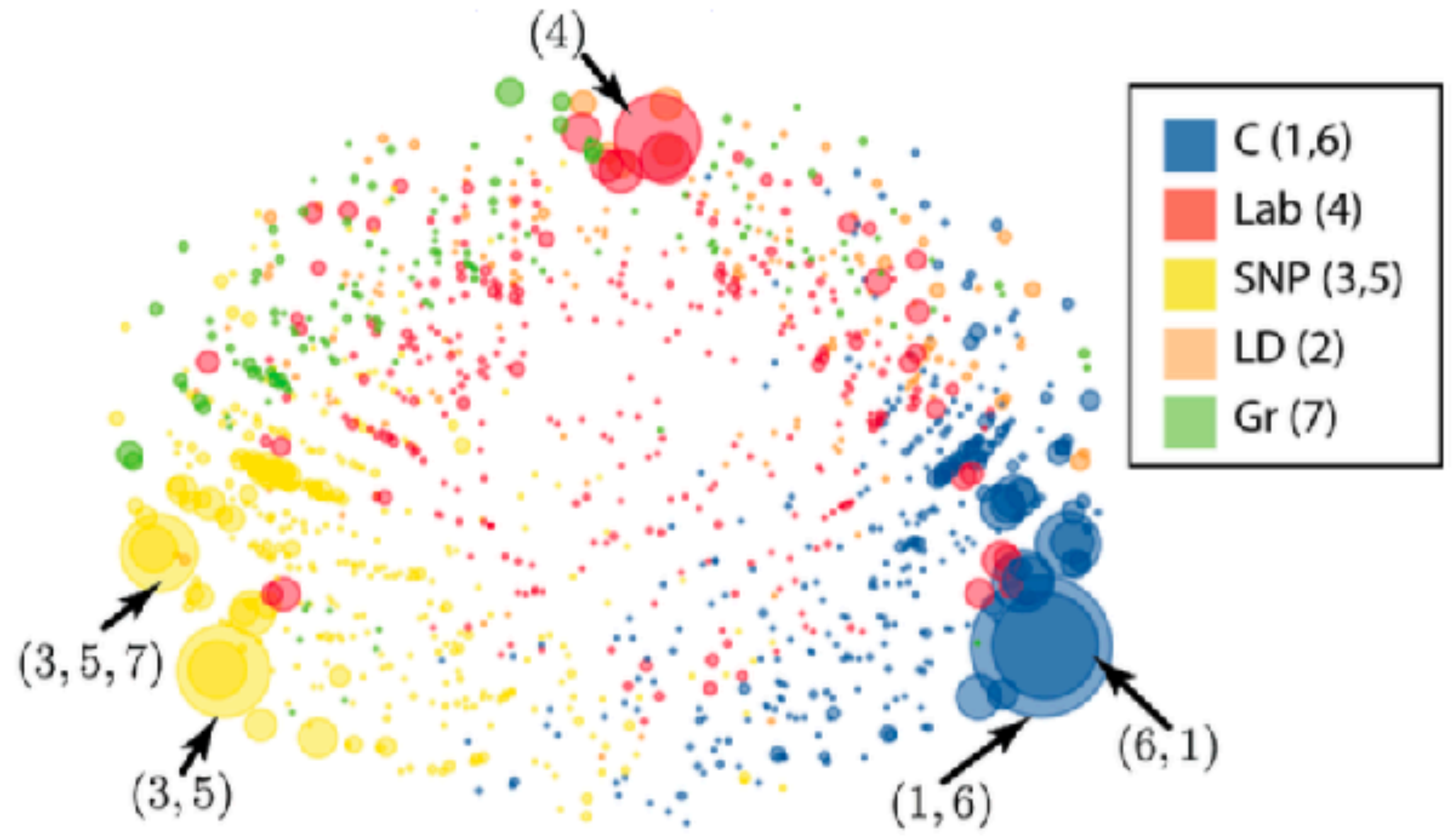
1: Grady (Labour)
2: Hughes (Labour)
3: McEwan (SNP)
4: Nelson (Con)
5: Paterson (SNP)

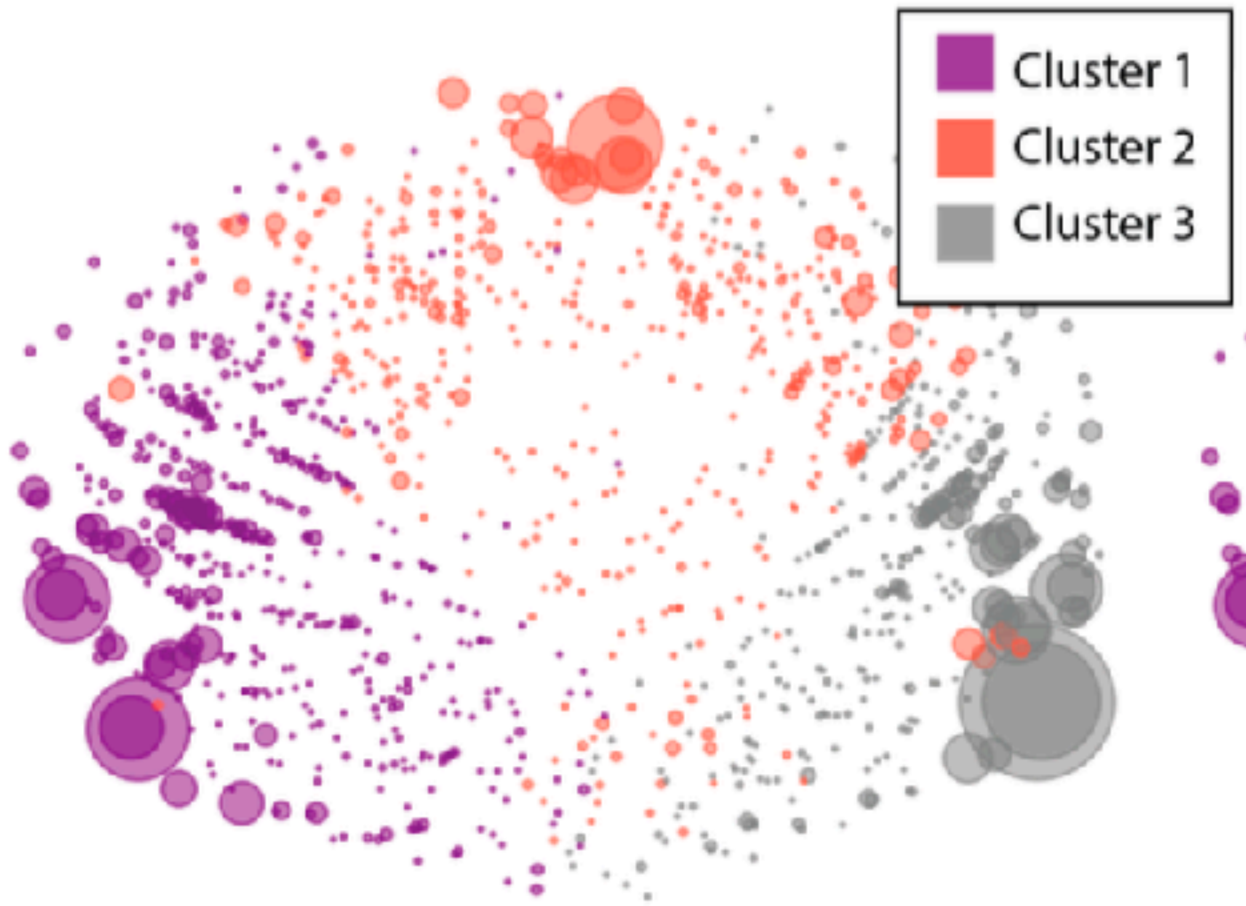


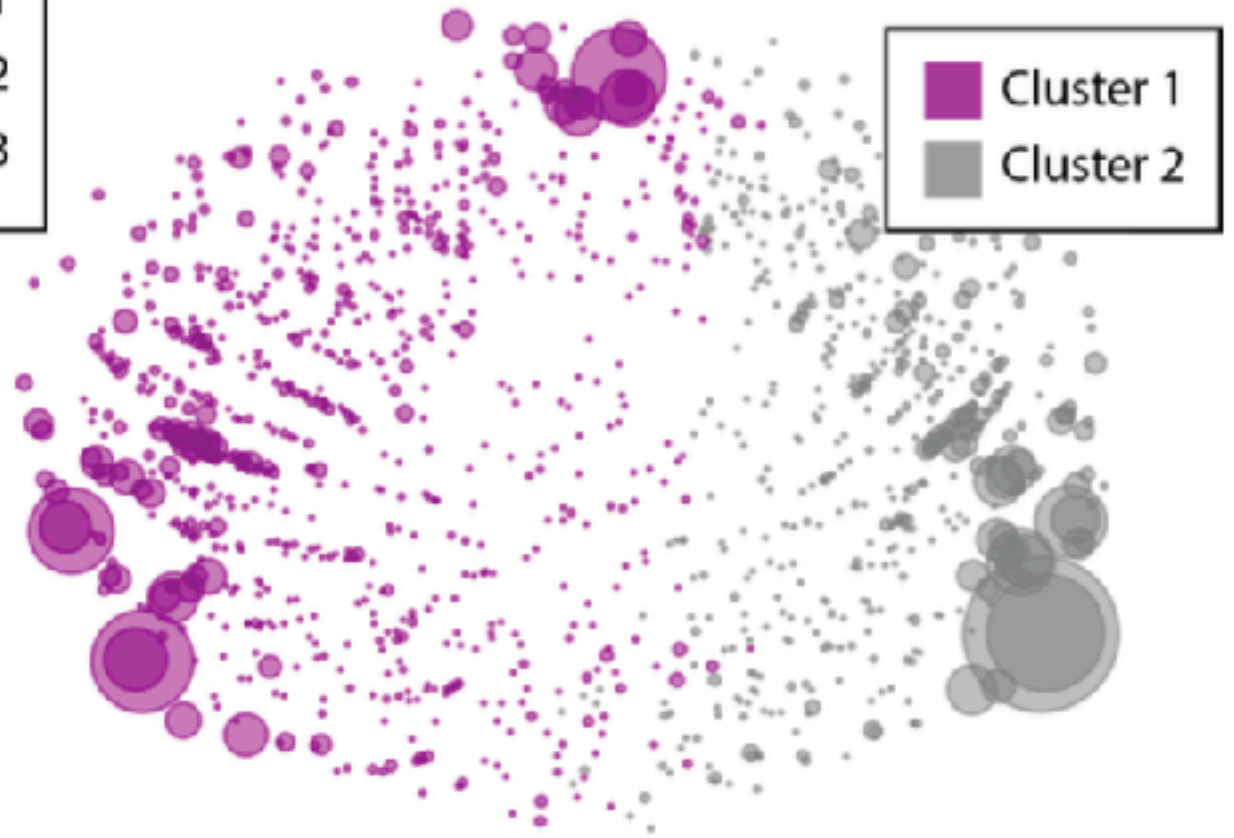
Pentland Hills

7 candidates; 4 winners; 11,315 cast ballots
967 bullet votes; 1431 complete ballots; avg length 3.2
18 ballots cast over 100x with (1,6) receiving 11.8% support;
660 ballots cast 1x; 1238 types cast out of 8659 possible









k=2: All ballot methods give one center (1,6)

Second center (3, 5, 7, 4), (3, 5, 7), or (3, 5, 4)

k=3: (1, 6), (3, 5, 7), (2, 4) by all methods

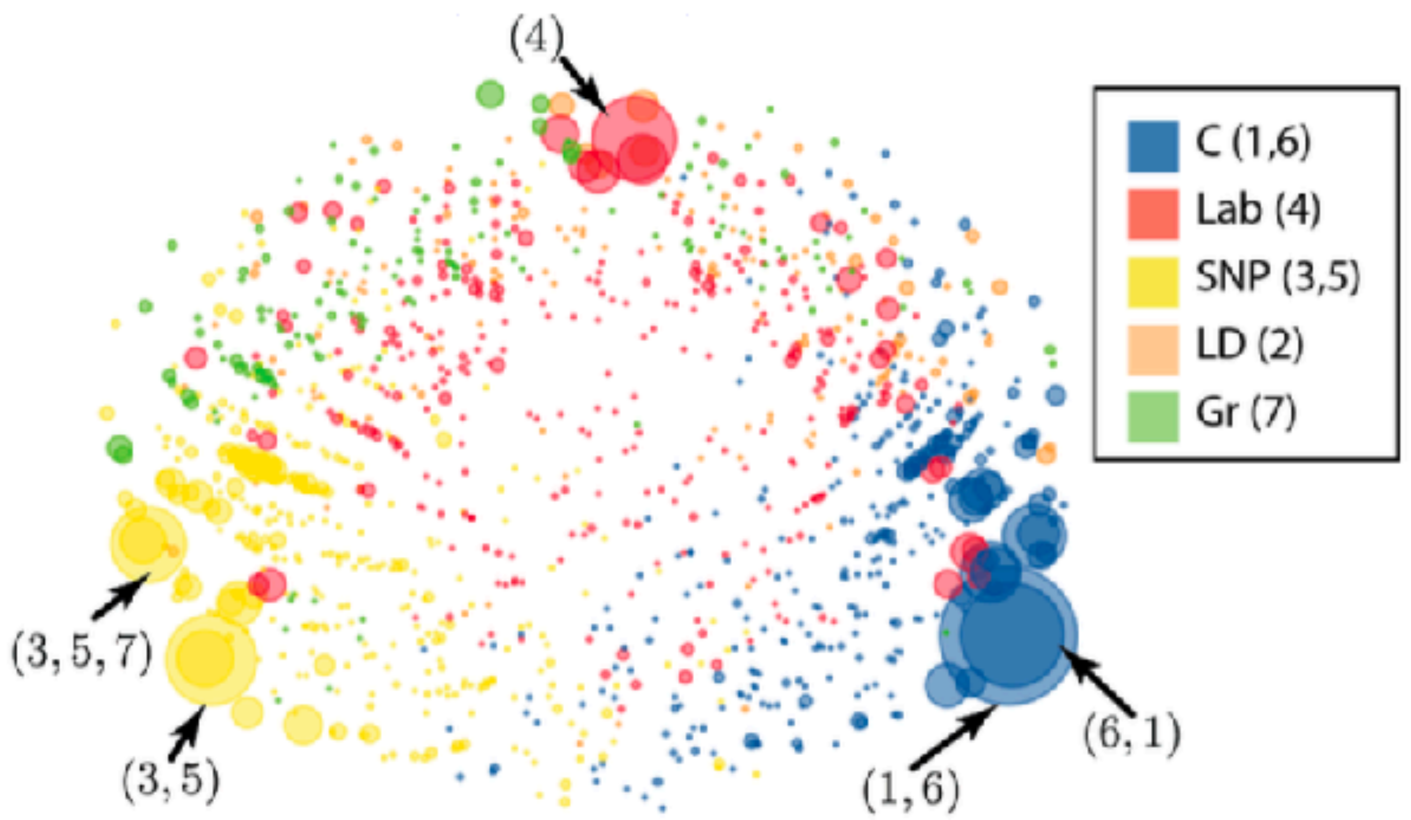
WINNERS: 1, 3, 4 in first round, then 6 in second

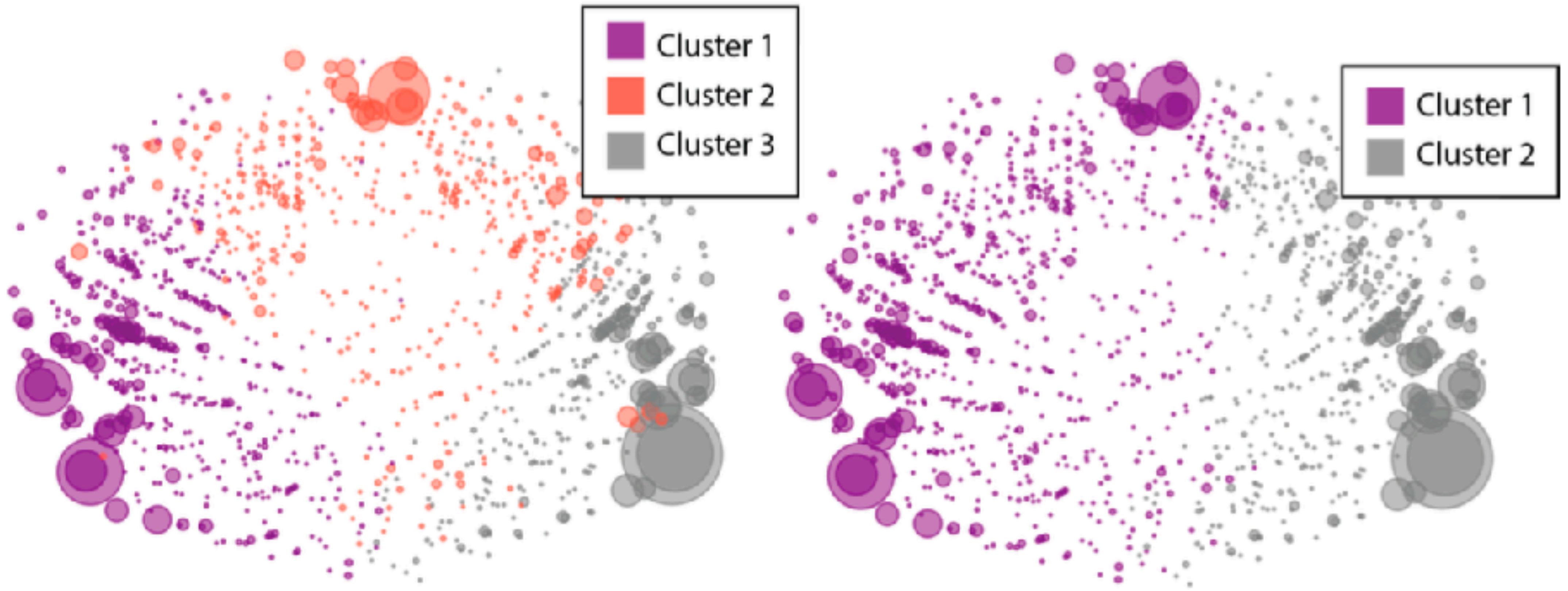
Pentland Hills

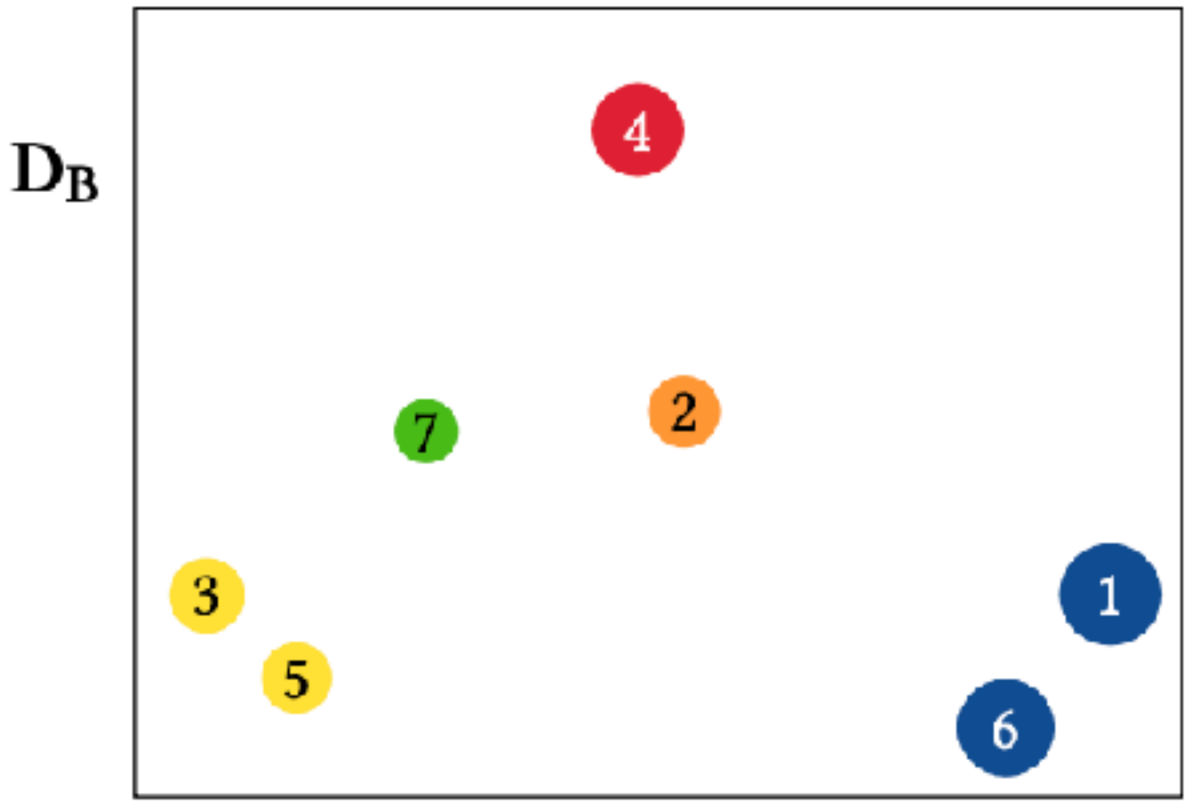
Can also use voter preferences to learn slates directly.

$$D_B(A_i, A_j) := \mathrm{avg}_\sigma \left|\mathfrak{b}(\sigma)_i - \mathfrak{b}(\sigma)_j\right|.$$

Candidate dissimilarity by rank difference

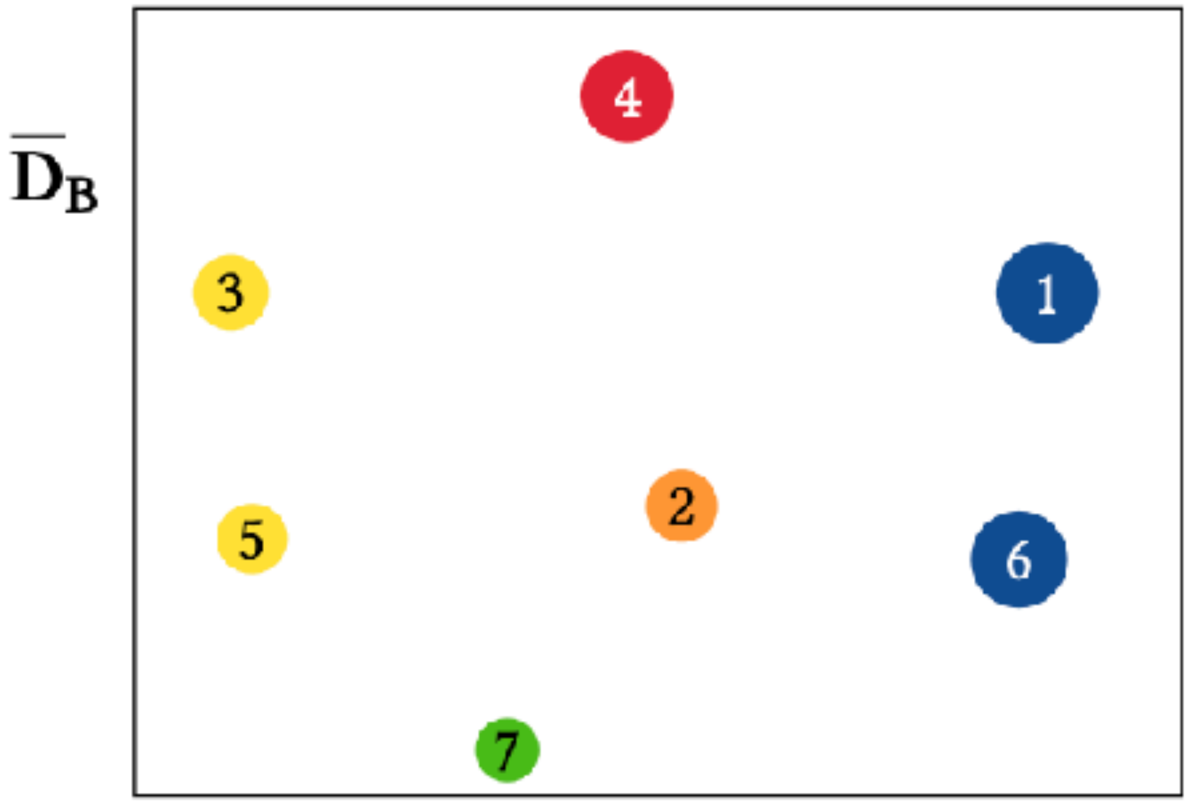






$k = 2$: Centers 1 and 7 are optimal, giving slates $\{1, 6\}, \{2, 3, 4, 5, 7\}$

$k = 3$: Centers 1, 2, and 5 are optimal, giving slates $\{1, 6\}, \{2, 4, 7\}, \{3, 5\}$



$k = 7 : \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}$
$k = 6 : \{1, 6\}, \{2\}, \{3\}, \{4\}, \{5\}, \{7\}$
$k = 5 : \{1, 6\}, \{2\}, \{3, 5\}, \{4\}, \{7\}$
$k = 4 : \{1, 6\}, \{2, 7\}, \{3, 5\}, \{4\}$
$k = 3 : \{1, 6\}, \{2, 4, 7\}, \{3, 5\}$
$k = 2 : \{1, 6\}, \{2, 3, 4, 5, 7\}$
$k = 1 : \{1, 2, 3, 4, 5, 6, 7\}$

k=2: All ballot methods give one center (1,6)
Second center (3, 5, 7, 4), (3, 5, 7), or (3, 5, 4)

k=3: (1, 6), (3, 5, 7), (2, 4) by all methods

Agglomeration through d_B point clouds

$$\overline{D}_B(A_i, A_j) := \mathrm{avg}_\sigma\, \mathrm{avg}_{\tau \in \overline{\sigma}} \left|\tau(i) - \tau(j)\right|.$$

now let’s compare voting rules

Comparing rules

- We will use Scottish profiles and compare several voting rules that output rankings:

- SCORING RULES — score, then rank by score

Borda	(5,4,3,2,1)
3-approval	(1,1,1,0,0)
2-approval	(1,1,0,0,0)
SNTV / Plurality	(1,0,0,0,0)

- STV RANKING

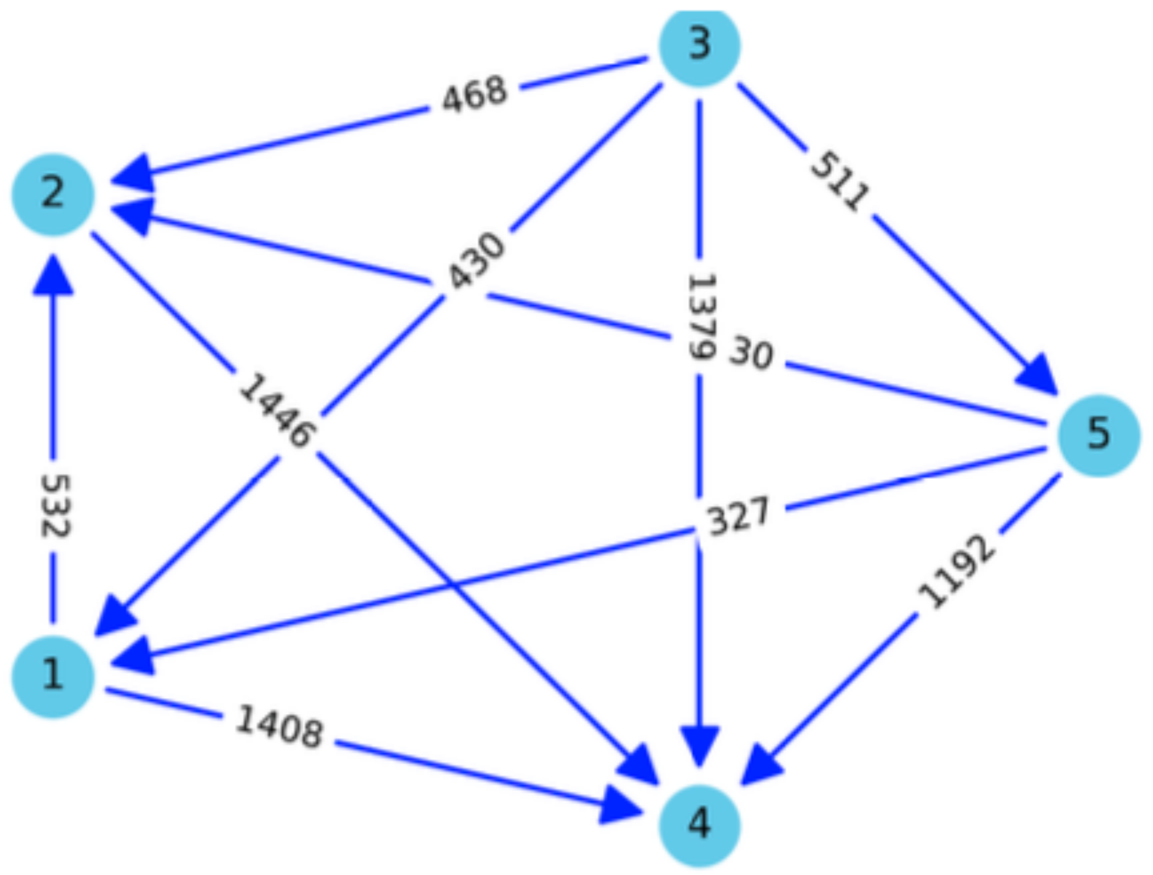
$$\begin{pmatrix} ELEC1 \\ ELEC2 \\ \vdots \\ ELIM2 \\ ELIM1 \end{pmatrix}$$

Renfrewshire Ward 2 (2022)

1: Grady (Labour)
2: Hughes (Labour)
3: McEwan (SNP)
4: Nelson (Con)
5: Paterson (SNP)

Renfrewshire results

Borda	3-app	2-app	SNTV	STV
3	3	3	3	3
5	2	5	1	1
1	5	1	5	5
2	1	2	4	4
4	4	4	2	2



Definitions of axiom defect

joint work with Rock, Sana, Wells

$$\rho_{\mathrm{IIA}}(f,P) := 1 - \frac{\sum_{C \in \mathcal{C}} d_{\mathrm{swap}}(f(P^C), f(P)^C)}{m\binom{m-1}{2}}.$$

if you disqualify a candidate **before** or **after** running a voting rule, you'll see two rankings of the m-1 other candidates. How different are those rankings, on average?

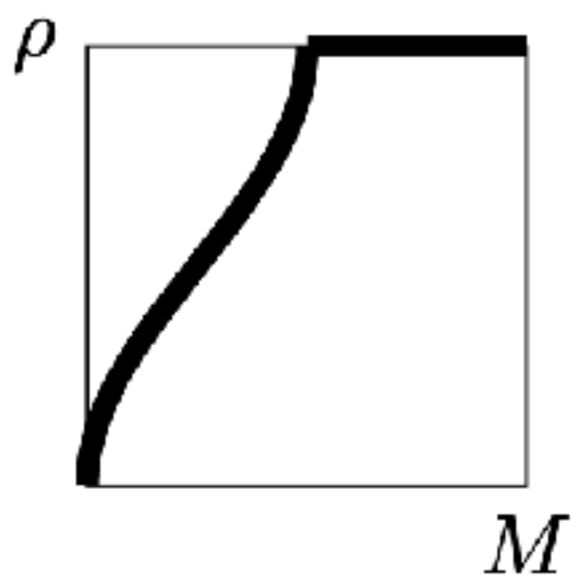
IIA axiom demands they always be the same, which would make this score 1 all the time.

$\rho_{\mathrm{IIA}}(f,P) = 1$ for all P $\iff$ f satisfies classical IIA

"worst match rate"

$$M(f,P) := \min_{A \succ_{f(P)} B} \frac{\#\{i \in \mathcal{V} : A \succ_i B\}}{n}.$$

$$\rho_{\mathrm{UM}}(f,P) := \begin{cases} \frac{2}{\pi}\arcsin\sqrt{2M}, & M(f,P) < 1/2 \\ 1, & M(f,P) \geq 1/2, \end{cases}$$



This ρ_{UM} score is set up to be 1 if every comparison observed in f(P) matches with at least half the voters: $M \geq 1/2$.

It's 0 if there's some comparison in f(P) that **everyone** disagrees with! — so 0 flags a unanimity violation.

$\rho_{\mathrm{UM}}(f,P) > 0$ for all P $\iff$ f satisfies classical UF

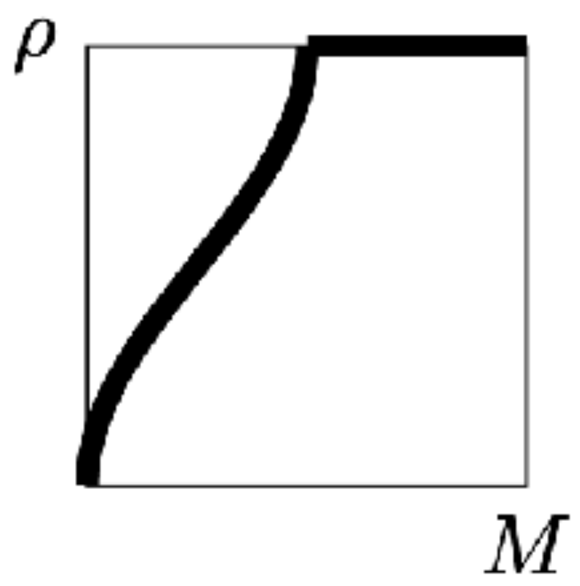
Definitions of axiom defect

joint work with Rock, Sana, Wells

$$\rho_{\mathrm{IIA}}(f,P) := 1 - \frac{\sum_{C\in\mathcal{C}} d_{\mathrm{swap}}(f(P^C), f(P)^C)}{m\binom{m-1}{2}}.$$

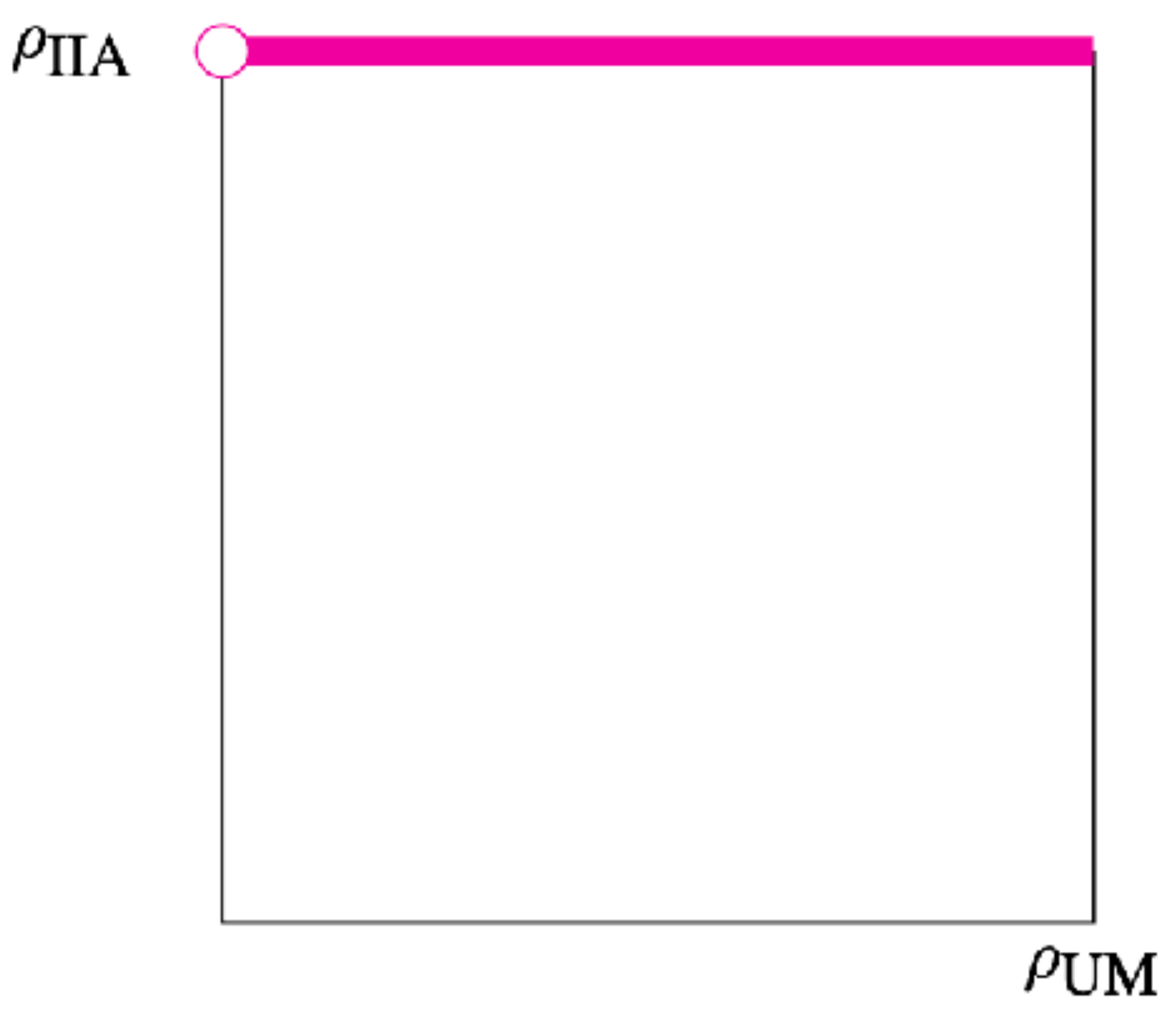
$$M(f,P) := \min_{A \succ_{f(P)} B} \frac{\#\{i\in\mathcal{V} : A \succ_i B\}}{n}.$$

$$\rho_{\mathrm{UM}}(f,P) := \begin{cases} \frac{2}{\pi}\arcsin\sqrt{2M}, & M(f,P) < 1/2 \\ 1, & M(f,P) \geq 1/2, \end{cases}$$



Visual Arrow's Theorem:

A ranking rule that always hits the **pink** strip ($\rho_{\mathrm{UM}} > 0,\ \rho_{\mathrm{IIA}} = 1$) must be **Dictatorship**

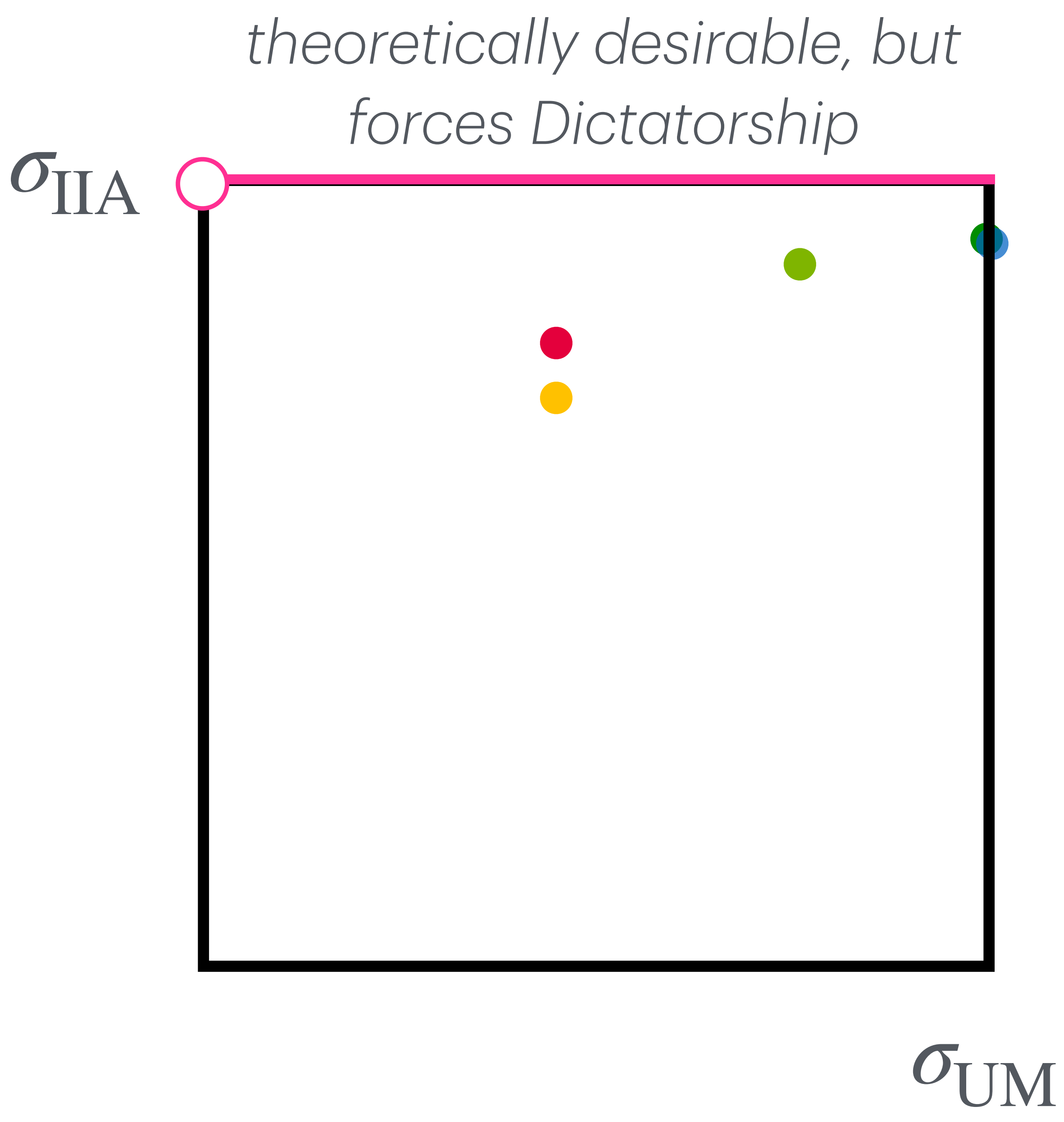


Examples

Renfrewshire results

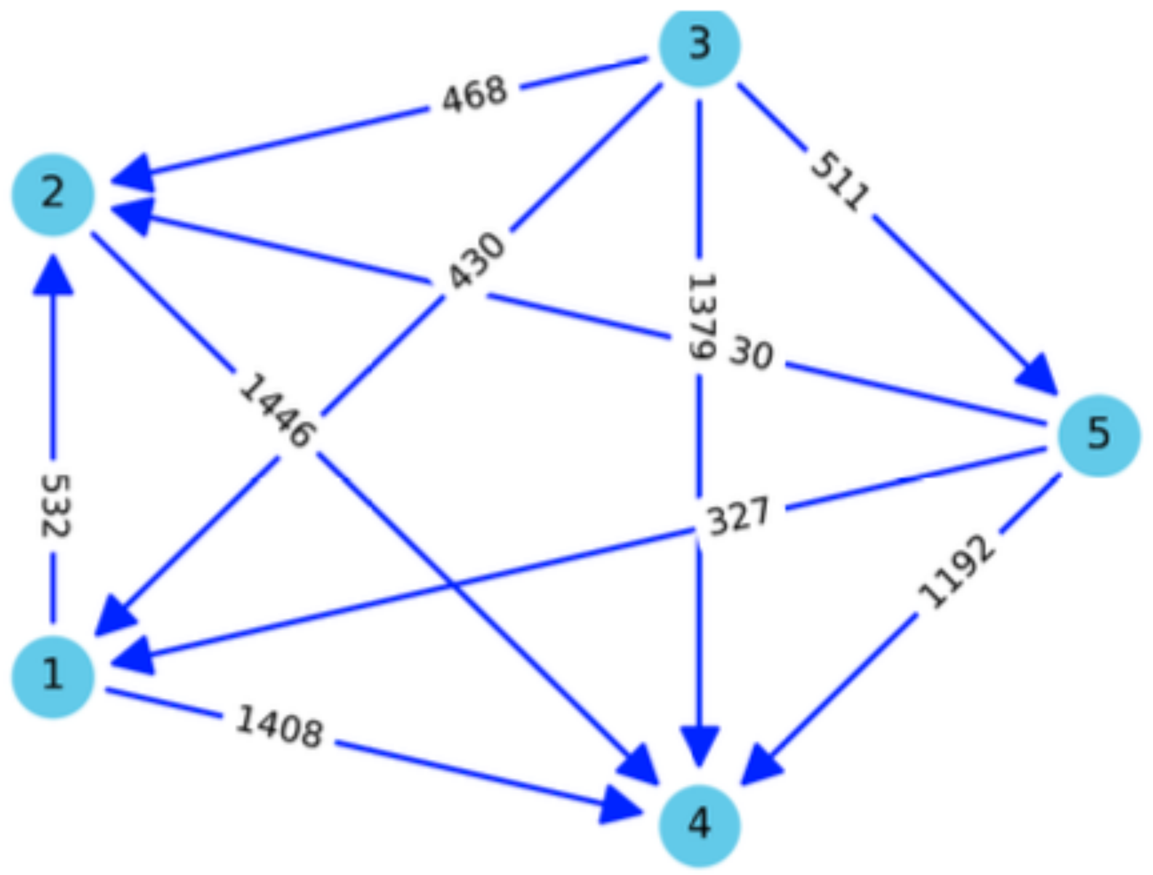
Borda	3-app	2-app	SNTV	STV
$\begin{pmatrix}3\\5\\1\\2\\4\end{pmatrix}$	$\begin{pmatrix}3\\2\\5\\1\\4\end{pmatrix}$	$\begin{pmatrix}3\\5\\1\\2\\4\end{pmatrix}$	$\begin{pmatrix}3\\1\\5\\4\\2\end{pmatrix}$	$\begin{pmatrix}3\\1\\5\\4\\2\end{pmatrix}$

IIA	0.93	0.90	0.93	0.80	0.73
UM	1	0.75	1	0.44	0.44



Renfrewshire Ward 2 (2022)

1: Grady (Labour)
2: Hughes (Labour)
3: McEwan (SNP)
4: Nelson (Con)
5: Paterson (SNP)



PWCG and topological sort

- We'll use Pairwise Comparison Graphs (tournament graphs)

- A full ranking $\pi \in \mathcal{S}(\mathcal{C})$ is a **topological sort** of a directed graph if for every pair of candidates, π respects the pairwise comparison

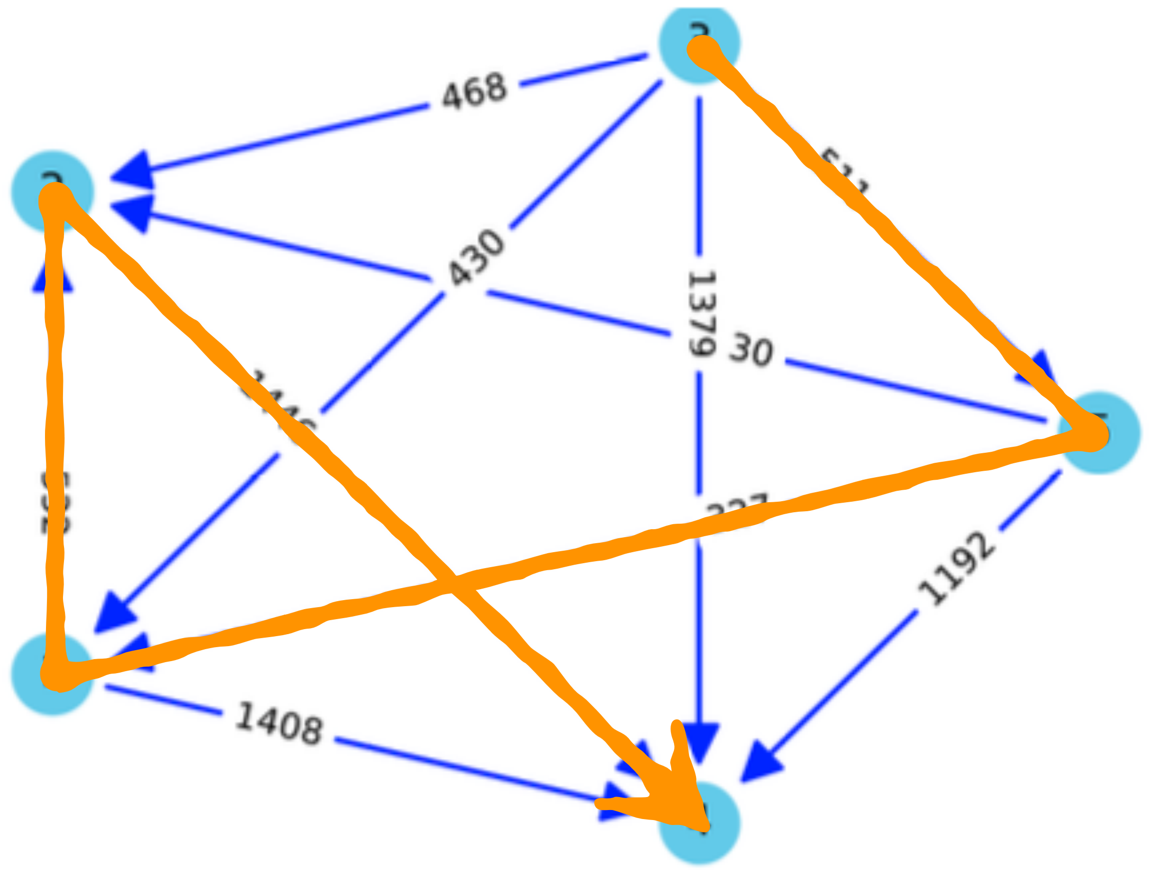
- Clearly a topological sort exists for a profile P iff there is a total Condorcet order (in this case, 3>5>1>2>4)
  - If sort exists, only $f(P) = \pi$ achieves $\sigma_{\mathrm{UM}} = 1$.
  - If no sort exists, then no rule can achieve $\sigma_{\mathrm{UM}} = 1$.

- **Lemma**: A `GreedySort` procedure (polynomial time) gives UM-optimal ranking.

- Observation: can't do analogous greedy sort for IIA, which depends on not only $f(P)$ but all $f(P^C)$ as well.

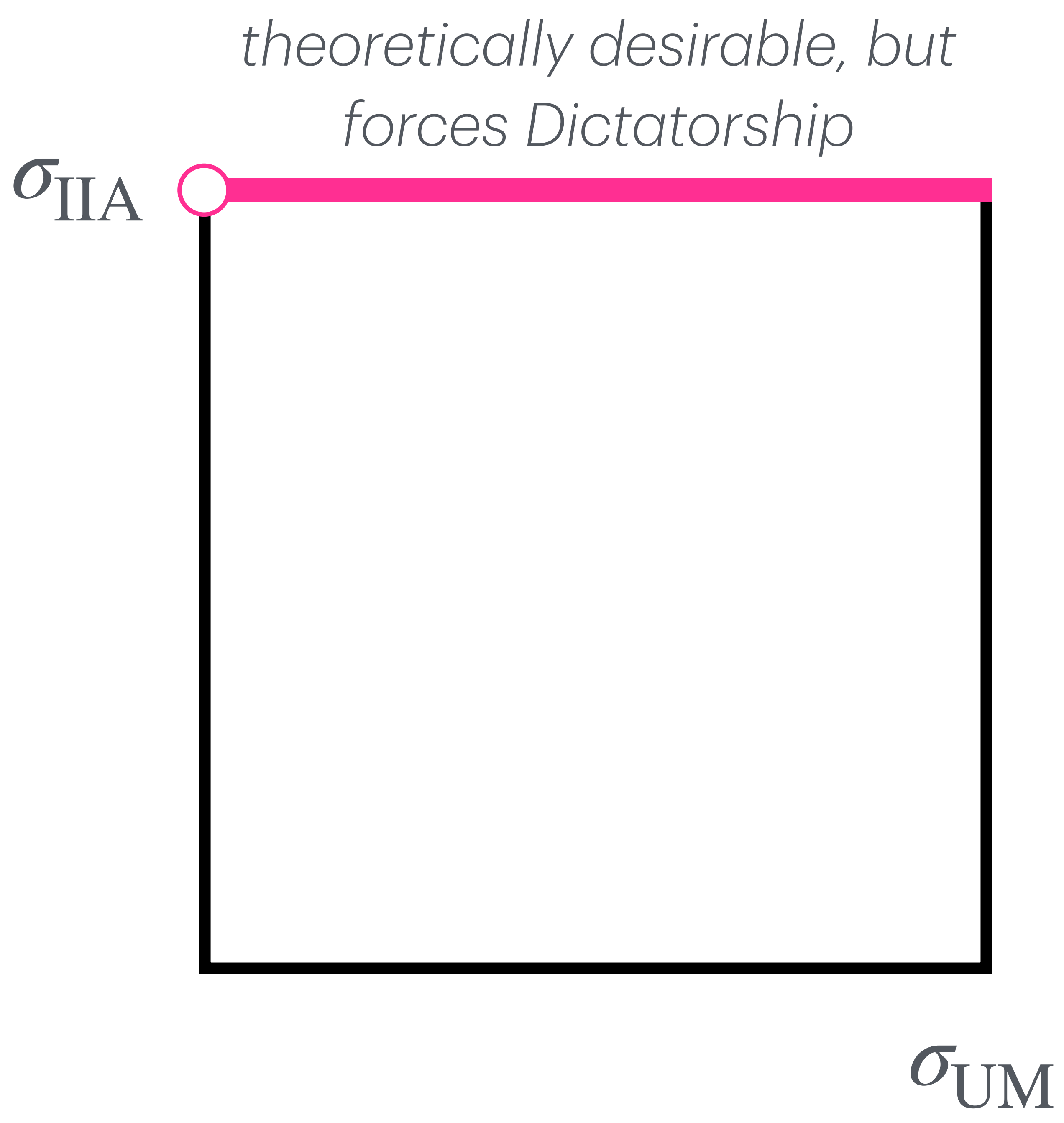
Renfrewshire Ward 2 (2022)

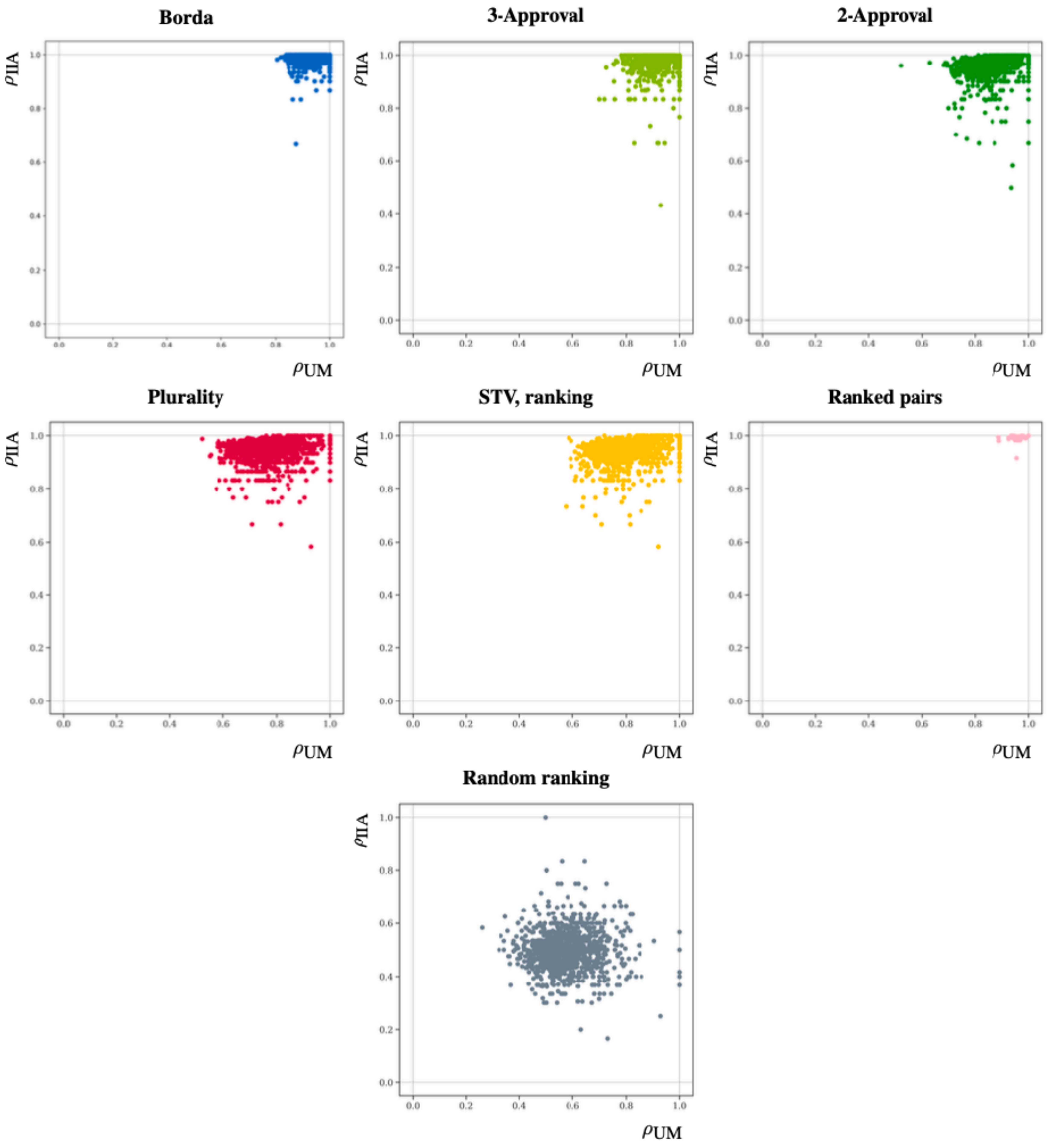
1: Grady (Labour)
2: Hughes (Labour)
3: McEwan (SNP)
4: Nelson (Con)
5: Paterson (SNP)



Scottish empirics

- 1025 out of 1070 Scottish elections have a topological sort (!!!) — "Condorcet all the way down"
- thus you can achieve $\sigma_{\text{UM}} = 1$ almost all the time



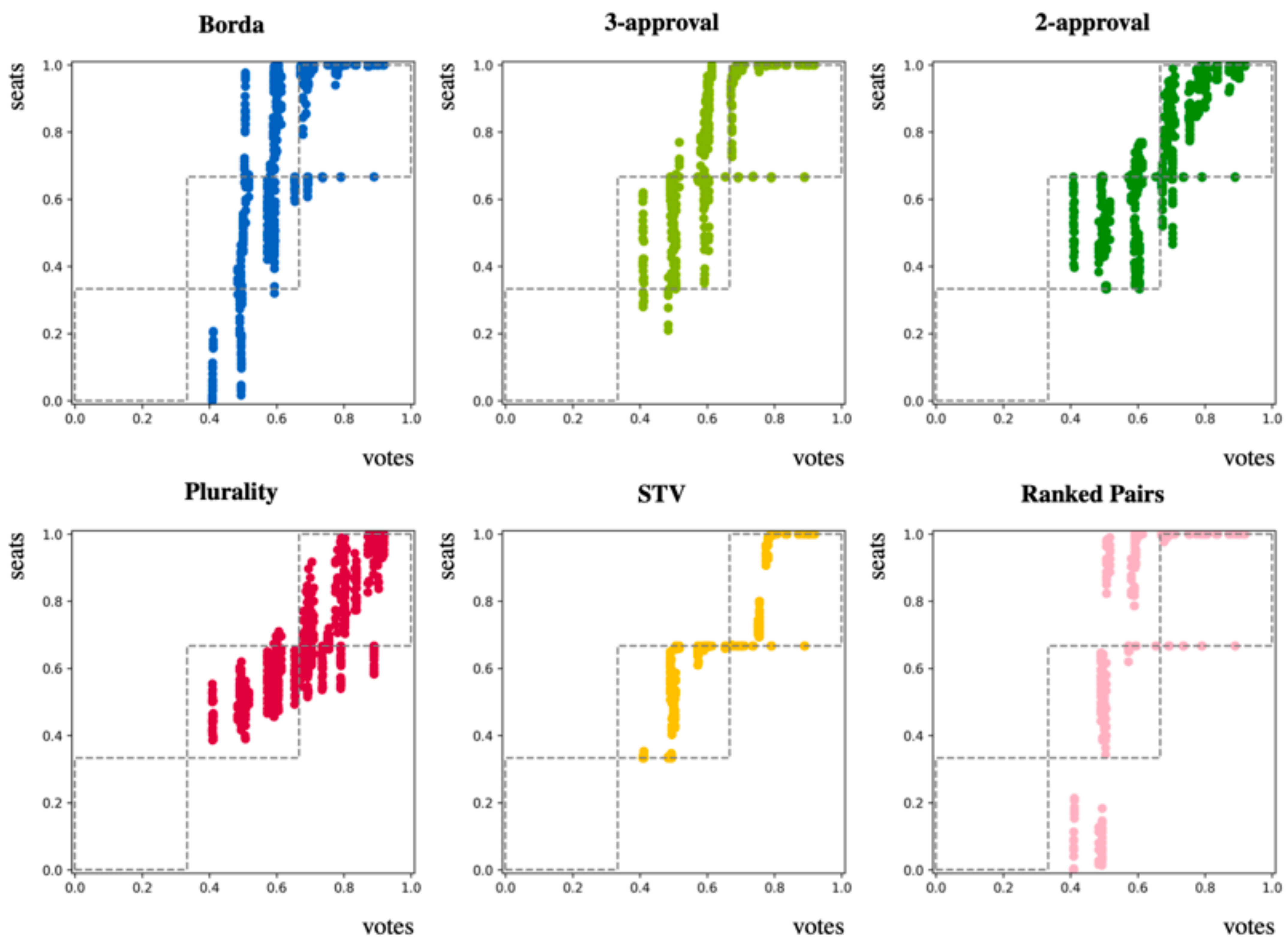


So what about **fairness**?

- Stability and majoritarianism seem like fully reasonable goals for LLM alignment, say, so our work says that Borda is a strong choice and Ranked Pairs is superlative.

- HOWEVER, consider this scenario: Group A (of candidates) has 40% support, Group B has 30% support, with the rest scattered. *Topsort often puts all A above all B in such cases.* In Renfrewshire, this would put both SNP candidates as the top two, while other methods have SNP/Labour split, satisfying many more people.

- So in polarized elections, optimizing for these metrics may be fundamentally **anti-proportional**— it will frequently allow one bloc to sweep the winners. If you think proportionality is important for fairness, then IIA and U/M may be counterproductive. And this may well be the case for electoral representation!



84,000 Bradley-Terry profiles
run through rules to elect 3 winners

Portland and NYC

NYC: $N = 1{,}071{,}730$

“Condorcet ladder”

Zohran Mamdani $\rightarrow$ Brad Lander $\rightarrow$ Adrienne Adams $\rightarrow$ Andrew Cuomo $\rightarrow$ $\rightarrow$ Zellnor Myrie $\rightarrow$ Scott Stringer $\rightarrow$ Michael Blake $\rightarrow$ Jessica Ramos $\rightarrow$ $\rightarrow$ Whitney Tilson $\rightarrow$ Selma Bartholomew $\rightarrow$ Paperboy Love Prince

	Borda				**3-Approval**				**2-Approval**			
	ρ_{IIA}	ρ_{UM}	σ_{IIA}	σ_{UM}	ρ_{IIA}	ρ_{UM}	σ_{IIA}	σ_{UM}	ρ_{IIA}	ρ_{UM}	σ_{IIA}	σ_{UM}
NYC	1	0.9532	1	1	0.9909	0.8185	1	1	0.9909	0.7410	1	1
D1	0.9968	0.8669	0.9394	0.9523	0.9970	0.8669	0.9242	0.9523	0.9950	0.8669	0.9773	0.9523
D2	0.9993	0.8297	0.9524	1	0.9972	0.7818	1	0.9977	0.9970	0.8297	0.9643	1
D3	0.9985	0.8575	1	1	0.9983	0.7897	1	1	0.9973	0.7596	1	1
D4	0.9987	0.9586	1	1	0.9986	0.9565	1	1	0.9978	0.8348	0.9722	0.8348

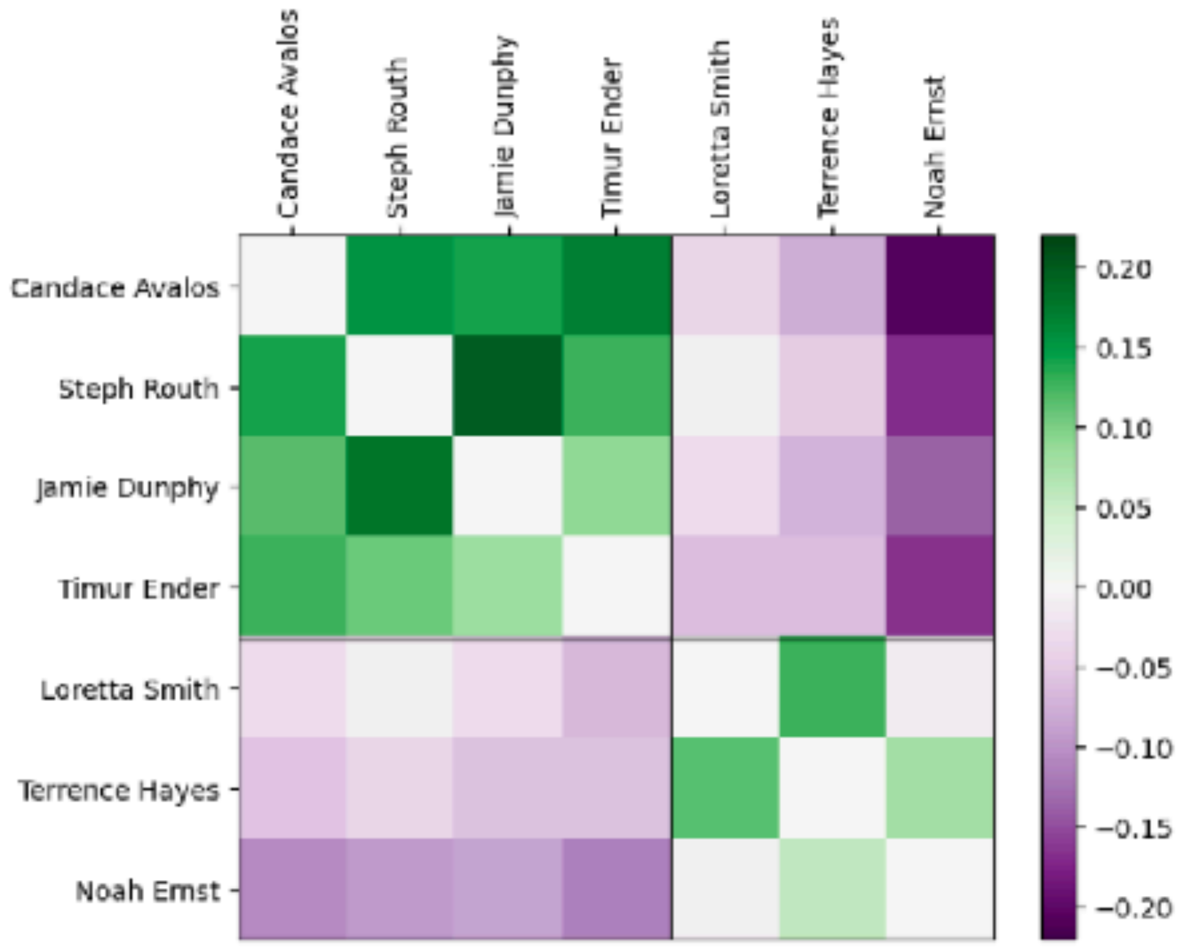
	Plurality				**STV / IRV**				**Ranked pairs**			
	ρ_{IIA}	ρ_{UM}	σ_{IIA}	σ_{UM}	ρ_{IIA}	ρ_{UM}	σ_{IIA}	σ_{UM}	ρ_{IIA}	ρ_{UM}	σ_{IIA}	σ_{UM}
NYC	0.9621	0.7410	1	1	0.9606	0.7410	1	1	1	1	1	1
D1	0.9935	0.6890	1	0.8401	0.9695	0.6890	0.9545	0.8401	1	1	1	1
D2	0.9975	0.7823	1	1	0.9842	0.7823	1	1	1	1	1	1
D3	0.9961	0.7387	1	1	0.9886	0.7387	1	1	0.9999	0.9612	1	1
D4	0.9954	0.7673	0.9769	0.7673	0.9868	0.7948	1	0.8348	1	1	1	1

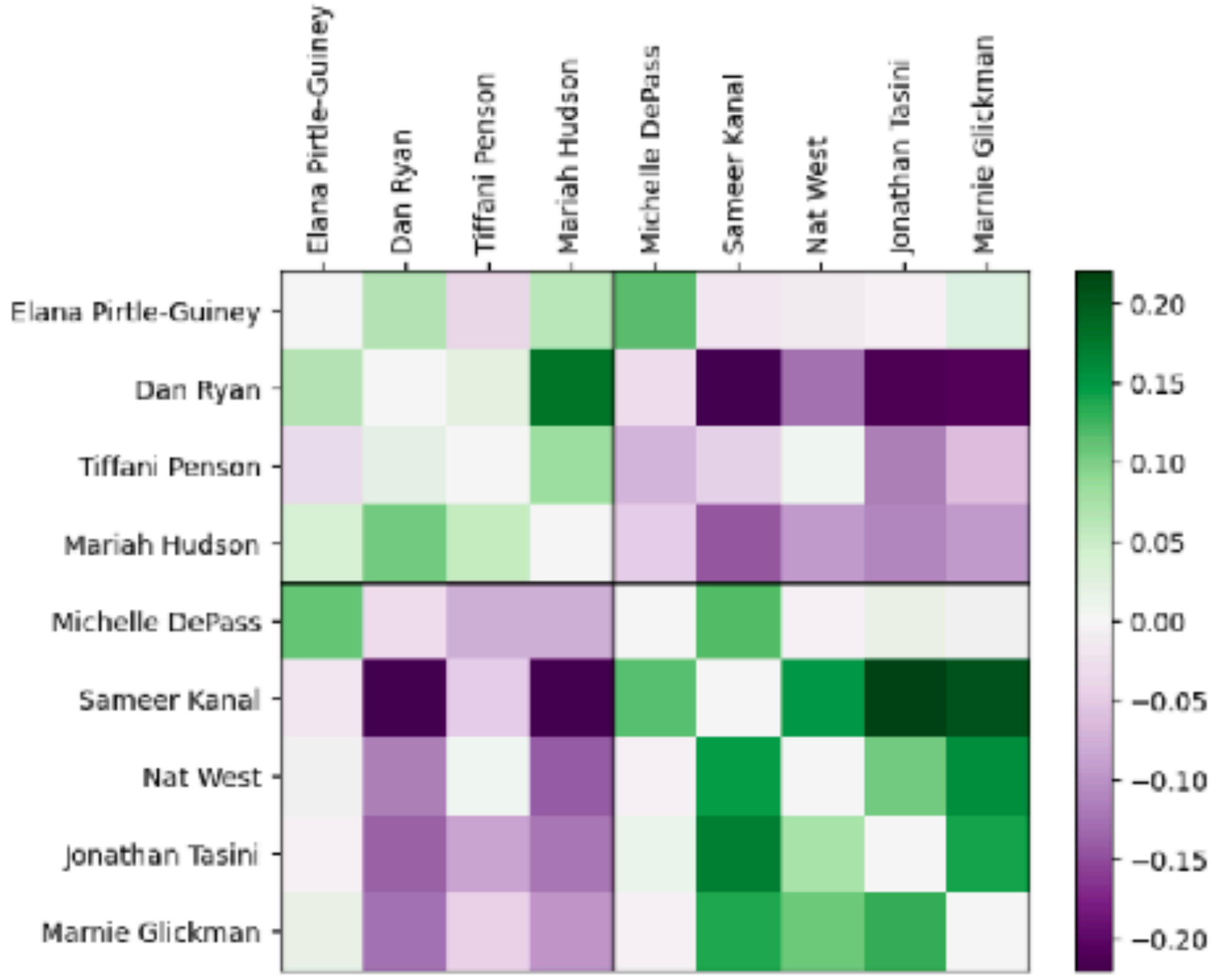
Mamdani, Lander ranked 639,693 and 631,838 times, respectively — Cuomo just 489,682.

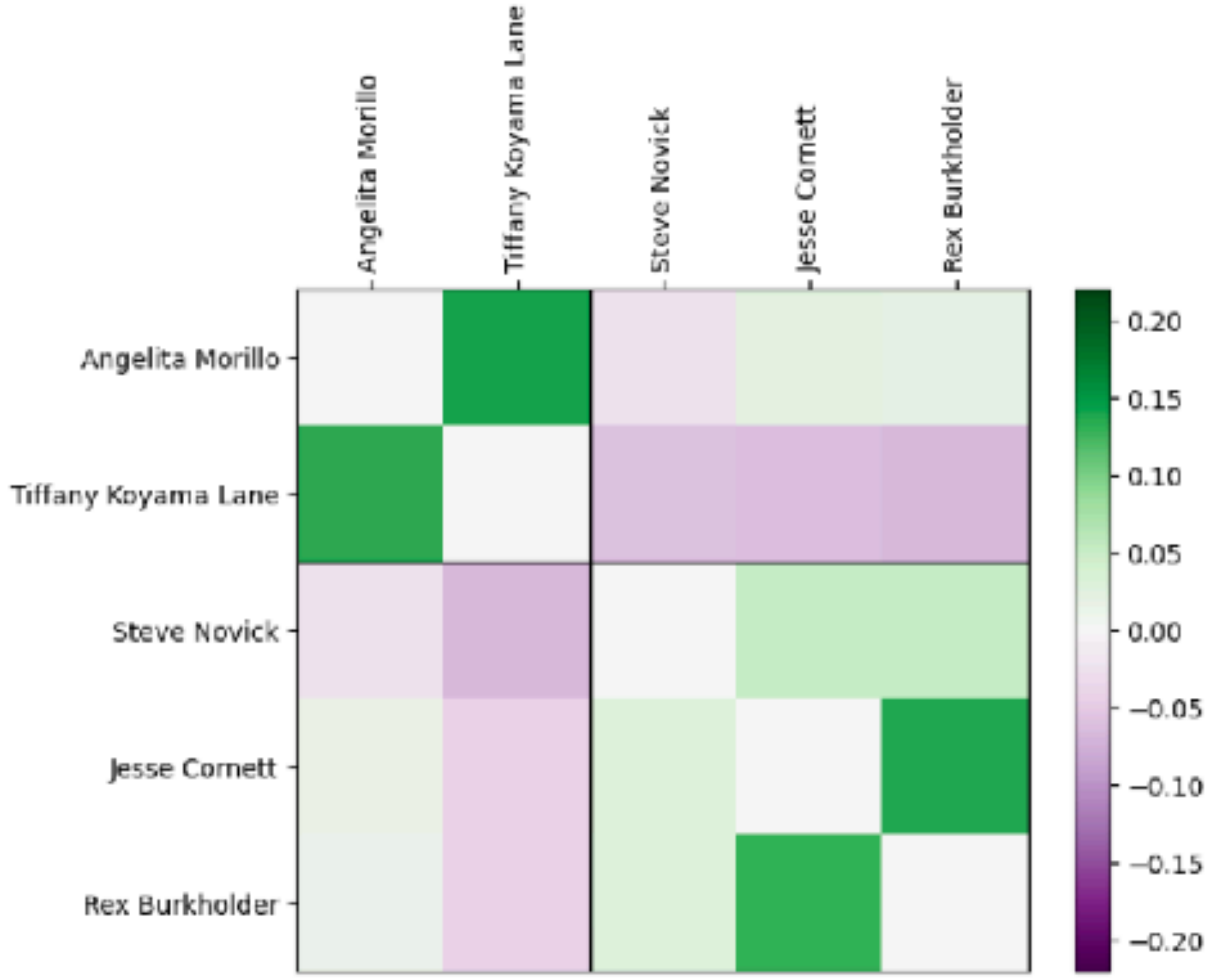
Cuomo not **viable**.

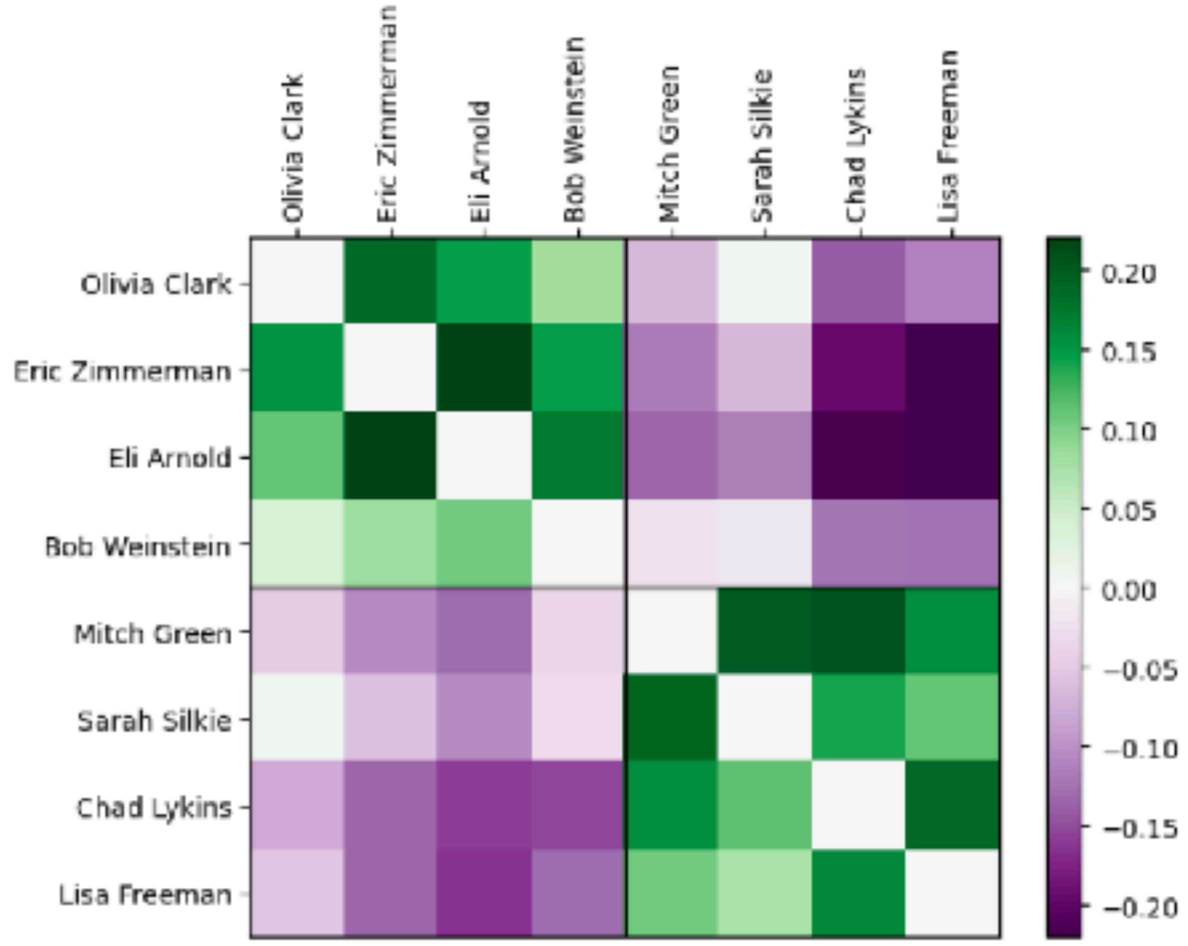
De facto progressive slate Mamdani>Lander>Adams>Myrie> Ramos — in fact listed in that order by Working Families Party!

Portland's blocs and slates









Portland	D1 ARDE vs. SHE	D2 PRPH vs. DKWTG	D3 MK vs. NCB	D4 CZAW vs. GSLF
First-place share	.604	.534	.537	.627
Mentions share	.631	.463	.455	.551
Borda share	.632	.475	.488	.575
Borda top-3 share	.627	.498	.523	.606
Borda	ARS- 2/3	RPK- 2/3	NKM- 2/3	CZA- 3/3
3-Approval	ARS- 2/3	RPD- 2/3	NKM- 2/3	CZA- 3/3
2-Approval	ARS- 2/3	RPK- 2/3	NMK- 2/3	CZG- 2/3
Plurality	ASD- 2/3	PRK- 2/3	NMK- 2/3	CGA- 2/3
STV	ASD- 2/3	KPR- 2/3	NMK- 2/3	CGZ- 2/3
Ranked pairs	ARD- 3/3	PRK- 2/3	NMK- 2/3	CZA- 3/3

As expected, the methods with **best** UM and IIA scores—Ranked Pairs, then Borda, then 3-approval—are **worst** for proportionality.