

Guide to Numerical Experiments on Elections in Computational Social Choice

Piotr Faliszewski
AGH University
Kraków, Poland



Niclas Boehmer
Harvard Univ.
USA



Łukasz Janeczko
AGH University
Poland



Andrzej Kaczmarczyk, AGH Univ.
Poland



Grzegorz Lisowski
AGH University
Poland



Grzegorz Pierczyński
AGH University
Poland



Simon Rey
U. of Amsterdam,
The Netherlands



Stanisław Szufa
AGH University
Poland



Dariusz Stolicki
Jagiellonian Uni.
Kraków, Poland



Tomasz Wąs
Universite Paris
Dauphine, France



No.	Title	Authors	Experiments
41	New Approximation for Borda Coalitional Manipulation AAMAS-2017	O. Keller, A. Hassidim, N. Hazon	O1: Impartial Culture 8 samples/datapoint: {5, 15, 25, 35, 45, 55}x{6, 12, 24} Note: m+1 candidates, k manipulators, n = 2k (non-manipulators)
42	Majority Graphs of Assignment Problems and Properties of Popular Random Assignments AAMAS-2017	F. Brandt, J. Hofbauer, M. Suderland	O1: Euclidean 2D, Impartial Culture 10000 samples/datapoint: {1, 7}x{1, 7} Euclidean: Unclear (but most probably uniform) 2D $([0, 1]^2)$ O2: exhaustive Unspecified number of samples/datapoint: {1, 4}x{1, 4} Note: The paper is about application of majority graphs from voting theory to matching.
43	Divide and Conquer: Using Geographic Manipulation to Win District-Based Elections AAMAS-2017	Y. Lewenberg, O. Lev, J. Rosenschein	O1: Mallows Mixture 1000 samples/datapoint: {6}x{1000000} Mallows: Complicated mixture simulating division into urban and rural areas, with components whose dispersion parameter depends on the location in a 2D space (this is geographical space, not ideology-one, like in Euclidean models). Central preferences of the mixtures were (apparently) hand-designed. O2: Real-Life (beyond PrefLib) Unspecified number of samples/datapoint: {10}x{?}, {11}x{?} Note: 2015 UK and Israeli elections. Data sets are reported to be at http://votes20.gov.il/ and http://www.electoralcommission.org.uk/our-work/ but neither of the pages was available while recording this entry. Generally, this is not exactly ordinal data (the experiment was recorded in the database for the sake of completeness)
44	Bisimulations for Verifying Strategic Abilities with an Application to ThreeBallot AAMAS-2017	F. Belardinelli, R. Condurache, C. Dima, W. Jamroga, A. Jones	Note: Has journal version
45	Proxy Voting for Better Outcomes AAMAS-2017	G. Cohensius, S. Mannor, R. Meir, E. Meiron, A. Orda	O1: PrefLib 1000 samples/datapoint: {16}x{?} A1: PrefLib (from ordinal data) 1000 samples/datapoint: {16}x{?} A2: specific own model (from ordinal data) 1000 samples/datapoint: {?}x{0, 100}
46	Manipulation of Hamming-based Approval Voting for Multiple Referenda and Committee Elections AAMAS-2017	N. Barrot, J. Lang, M. Yokoo	A1: Impartial Culture, Impartial Culture (Variant) (from ordinal data) 10000 samples/datapoint: {3, 4, 5}x{25} Note: Under IC each voter approves each candidate with probability 0.5. The authors also use biased model, where they choose p1 and p2 as two randomly selected probabilities. 40% of the voters approve each candidate with probability p1, 40% approve each candidate with probability p2, and 20% approve each candidate with probability 0.5. A2: Impartial Culture (from ordinal data) Unspecified number of samples/datapoint: {5}x{5, 70}



An Election

v_1 : 🐼 > 🐳 > 🐱

v_2 : 🐳 > 🐱 > 🐼

v_3 : 🐼 > 🐱 > 🐳

v_4 : 🐱 > 🐼 > 🐳

$$E = (C, V)$$

$$C = \{ \text{🐼}, \text{🐳}, \text{🐱} \}$$

$$V = (v_1, v_2, v_3, v_4)$$

An Election

v_1 :  >  > 

v_2 :  >  > 

v_3 :  >  > 

v_4 :  >  > 

$$E = (C, V)$$

$$C = \{ \img alt="panda" data-bbox="258 793 283 833" , \img alt="whale" data-bbox="293 793 318 833" , \img alt="cat" data-bbox="328 793 353 833" \}$$

$$V = (v_1, v_2, v_3, v_4)$$

Also an election

v_1 : {  ,  }

v_2 : {  ,  }

v_3 : {  ,  ,  }

v_4 : {  ,  }

We mostly focus on
the ordinal setting

An Election

- v_1 : 🐼 > 🐳 > 🐱
- v_2 : 🐳 > 🐱 > 🐼
- v_3 : 🐼 > 🐱 > 🐳
- v_4 : 🐱 > 🐼 > 🐳

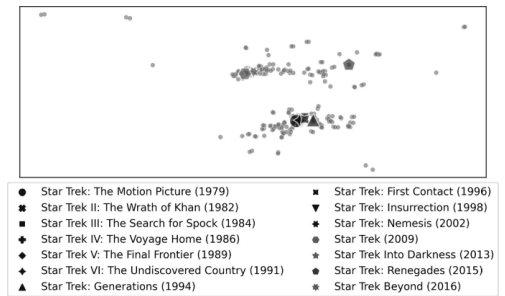
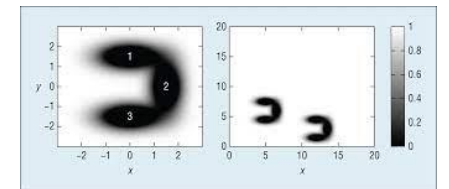
Winner Determination

Result Modification/Analysis

Normative Properties



New Rules, New Settings

Applications



- Star Trek: The Motion Picture (1979)
- Star Trek II: The Wrath of Khan (1982)
- Star Trek III: The Search for Spock (1984)
- ◆ Star Trek IV: The Voyage Home (1986)
- ◆ Star Trek V: The Final Frontier (1989)
- ◆ Star Trek VI: The Undiscovered Country (1991)
- ▲ Star Trek: Generations (1994)
- Star Trek: First Contact (1996)
- ▼ Star Trek: Insurrection (1998)
- Star Trek: Nemesis (2002)
- Star Trek (2009)
- ◆ Star Trek Into Darkness (2013)
- ◆ Star Trek: Renegades (2015)
- ★ Star Trek Beyond (2016)

An Election

- v_1 :  >  > 
- v_2 :  >  > 
- v_3 :  >  > 
- v_4 :  >  > 

Largely studied
theoretically

Why not do
experiments?

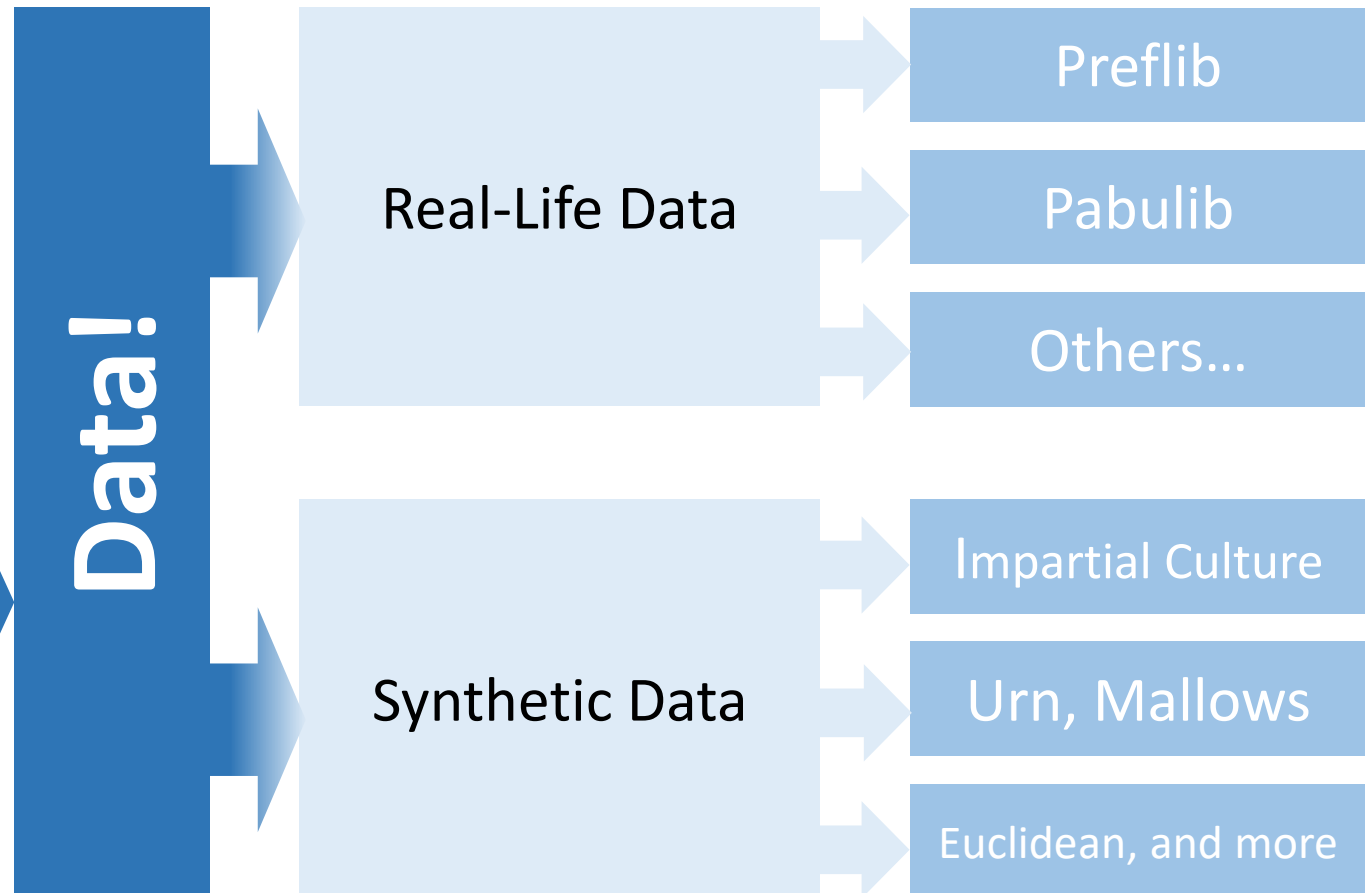
Benefits of Experiments

- More complex settings
- More precise results
 - Exact running time vs asymptotic running time
- Observe actual phenomena instead of merely predicting their possibility
 - Condorcet winners often exist
 - No-show paradox is/is-not a problem
 - Voting rules do/do-not give very different results

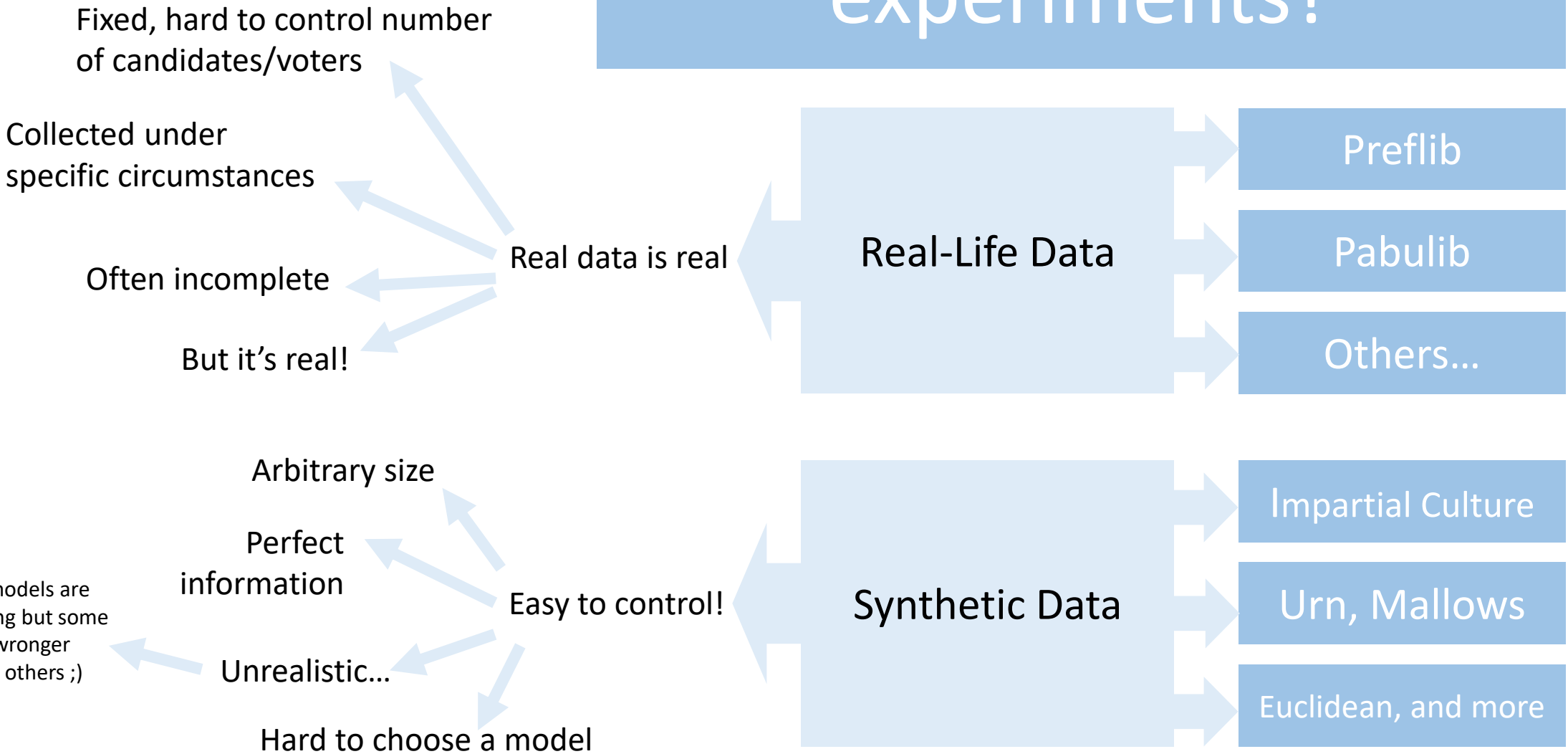
Problems with Experiments

- They don't generalize
- May be misleading
- Some insights are impossible to get experimentally
- You never really know...

Why not do experiments?



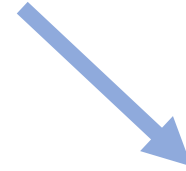
Why not do experiments?



Guide to Experiments



A city guide: How were experiment ran so far?



A „how to” guide suggesting how experiments should be done

Idea: Collect as many COMSOC papers about elections as possible, see what they do, draw conclusions!

Collecting the Data

Papers

- AAI, AAMAS, IJCAI
- 2010—2023
- Downloaded all the papers using the XML file from DBLP (September 2023)

Screening Process

- Automated script looking for election- and experiment-related keywords
 - `election, vote, ballot`
 - `experiment, empirical, simulation`
- Manual check of the shortlist
- E.g., IJCAI-23:
 - 846 papers
 - Script shortlisted 41
 - Manual check retained 7

Collecting the Data

Easy false positives:

- A passing remark about voting in an unrelated paper
- *Election* as part of *selection*
- NLP papers studying political debates (e.g., twitter messages from US presidential campaigns)
- Majority voting in ensemble learning
- Voting used to aggregate some data (but not useful to us)

Screening Process

- Automated script looking for election- and experiment-related keywords
 - `election, vote, ballot`
 - `experiment, empirical, simulation`
- Manual check of the shortlist
- E.g., IJCAI-23:
 - 846 papers
 - Script shortlisted 41
 - Manual check retained 7

Collecting the Data

Less easy false positives:

- Papers focussing on 2 candidates only (a lot of liquid democracy works, many works about electing parliaments)
- Papers not discussing experiments, but simply suggesting that someone else should.
- Papers about topics related to voting, but not similar enough to make it to the final database

Screening Process

- Automated script looking for election- and experiment-related keywords
 - `election, vote, ballot`
 - `experiment, empirical, simulation`
- Manual check of the shortlist
- E.g., IJCAI-23:
 - 846 papers
 - Script shortlisted 41
 - Manual check retained 7

Collecting the Data: Experiments

- Each paper may contain several experiments
- The notion of an experiment is not well-defined and paper-specific
- For our purposes, a single experiment regards identically generated data of given sizes

Two experiments in a paper, one in the Guide:

- The paper first studies running time of an algorithm on IC elections, for 10 candidates and 10 to 100 voters.
- Then it looks for the probability of a Condorcet winner in such data
- In the Guide this is a single experiments because it uses the same data.

Collecting the Data: Experiments

- Each paper may contain several experiments
- The notion of an experiment is not well-defined and paper-specific
- For our purposes, a single experiment regards identically generated data of given sizes

One experiment in a paper, one in the guide:

- The paper studies the running time of an algorithm on IC elections, for 10 candidates and 10 to 100 voters, and for 10 to 100 candidates and 50 voters.
- The Guide would contain a single experiment, whose size would be described as:

$\{10\} \times [10, 100], [10, 100] \times \{50\}$

Collecting the Data: Experiments

- Each paper may contain several experiments
- The notion of an experiment is not well-defined and paper-specific
- For our purposes, a single experiment regards identically generated data of given sizes

One experiment in a paper, two in the guide:

- The paper studies the running time of an algorithm on IC elections, for 10 candidates and 10 to 1000 voters, and then for 20 candidates and 10 to 100 voters on Mallows elections
- The Guide would contain two experiments

Basic Statistics

- Papers: 160
 - 130 ordinal
 - 35 approval
 - **Puzzle?**
- Experiments: 250
 - 211 ordinal
 - 46 approval
- Authors: 273 (+/-)

P. Faliszewski --> 26 paper(s) (18 ordinal, 8 approval)
P. Skowron --> 14 paper(s) (8 ordinal, 6 approval)
N. Talmon --> 14 paper(s) (11 ordinal, 3 approval)
M. Lackner --> 12 paper(s) (3 ordinal, 9 approval)
S. Szufa --> 11 paper(s) (8 ordinal, 3 approval)
A. Procaccia --> 8 paper(s) (ordinal)
A. Slinko --> 8 paper(s) (7 ordinal, 1 approval)
N. Boehmer --> 7 paper(s) (ordinal)
N. Mattei --> 7 paper(s) (5 ordinal, 2 approval)
N. Shah --> 7 paper(s) (6 ordinal, 1 approval)
L. Xia --> 7 paper(s) (ordinal)
C. Boutilier --> 6 paper(s) (ordinal)
U. Endriss --> 6 paper(s) (4 ordinal, 2 approval)
J. Lang --> 6 paper(s) (3 ordinal, 3 approval)
O. Lev --> 6 paper(s) (ordinal)
D. Peters --> 6 paper(s) (4 ordinal, 2 approval)
T. Walsh --> 6 paper(s) (ordinal)
R. Bredebeck --> 5 paper(s) (4 ordinal, 1 approval)
M. Brill --> 5 paper(s) (2 ordinal, 3 approval)
E. Elkind --> 5 paper(s) (3 ordinal, 2 approval)
R. Meir --> 5 paper(s) (3 ordinal, 3 approval)
R. Niedermeier --> 5 paper(s) (4 ordinal, 1 approval)
J. Rosenschein --> 5 paper(s) (ordinal)
F. Rossi --> 5 paper(s) (ordinal)
H. Aziz --> 4 paper(s) (ordinal)
F. Brandt --> 4 paper(s) (ordinal)
I. Caragiannis --> 4 paper(s) (ordinal)
S. Kraus --> 4 paper(s) (ordinal)
Y. Lewenberg --> 4 paper(s) (ordinal)
S. Nath --> 4 paper(s) (ordinal)
K. Sornat --> 4 paper(s) (2 ordinal, 2 approval)
A. Wilczynski --> 4 paper(s) (ordinal)

Basic Statistics

- Papers: 158
 - 125 ordinal
 - 35 approval)
 - **Puzzle?**
- Experiments: 250
 - 204 ordinal
 - 46 approval
- Authors: 273 (+/-)

P. Faliszewski --> 26 paper(s) (18 ordinal, 8 approval)

P. Skowron --> 14 paper(s) (8 ordinal, 6 approval)

N. Talmon --> 14 paper(s) (11 ordinal, 3 approval)

M. Lackner --> 12 paper(s) (3 ordinal, 9 approval)

S. Szufa --> 11 paper(s) (8 ordinal, 3 approval)

A. Procaccia --> 8 paper(s) (ordinal)

A. Slinko --> 8 paper(s) (7 ordinal, 1 approval)

N. Boehmer --> 7 paper(s) (ordinal)

N. Mattei --> 7 paper(s) (5 ordinal, 2 approval)

N. Shah --> 7 paper(s) (6 ordinal, 1 approval)

L. Xia --> 7 paper(s) (ordinal)

C. Boutilier --> 6 paper(s) (ordinal)

U. Endriss --> 6 paper(s) (4 ordinal, 2 approval)

J. Lang --> 6 paper(s) (3 ordinal, 3 approval)

O. Lev --> 6 paper(s) (ordinal)

D. Peters --> 6 paper(s) (4 ordinal, 2 approval)

T. Walsh --> 6 paper(s) (ordinal)

R. Bredebeck --> 5 paper(s) (4 ordinal, 1 approval)

M. Brill --> 5 paper(s) (2 ordinal, 3 approval)

E. Elkind --> 5 paper(s) (3 ordinal, 2 approval)

R. Meir --> 5 paper(s) (3 ordinal, 3 approval)

R. Niedermeier --> 5 paper(s) (4 ordinal, 1 approval)

J. Rosenschein --> 5 paper(s) (ordinal)

F. Rossi --> 5 paper(s) (ordinal)

H. Aziz --> 4 paper(s) (ordinal)

F. Brandt --> 4 paper(s) (ordinal)

I. Caragiannis --> 4 paper(s) (ordinal)

S. Kraus --> 4 paper(s) (ordinal)

Y. Lewenberg --> 4 paper(s) (ordinal)

S. Nath --> 4 paper(s) (ordinal)

K. Sornat --> 4 paper(s) (2 ordinal, 2 approval)

A. Wilczynski --> 4 paper(s) (ordinal)

Basic Statistics

- Papers: 158
 - 125 ordinal
 - 35 approval)
 - **Puzzle?**
- Experiments: 250
 - 204 ordinal
 - 46 approval
- Authors: 273 (+/-)

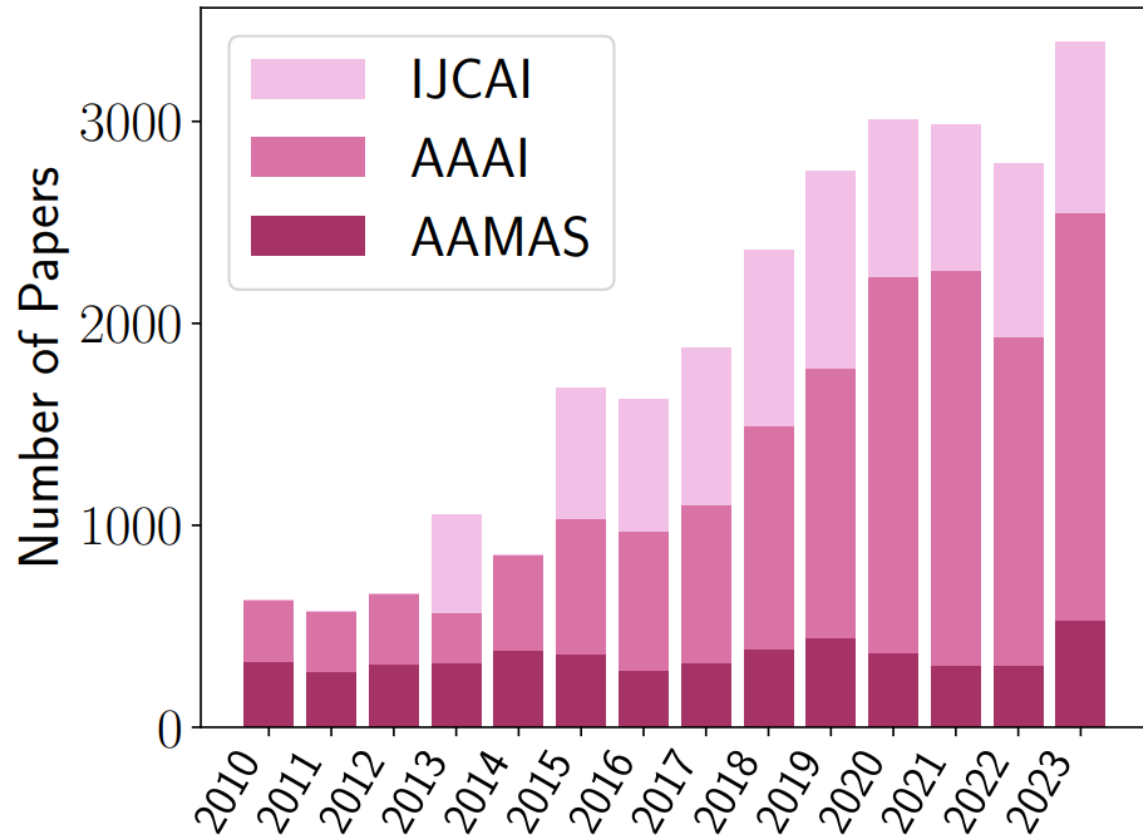
P. Faliszewski --> 26 paper(s) (18 ordinal, 8 approval)
P. Skowron --> 14 paper(s) (8 ordinal, 6 approval)
N. Talmon --> 14 paper(s) (11 ordinal, 3 approval)
M. Lackner --> 12 paper(s) (3 ordinal, 9 approval)
S. Szufa --> 11 paper(s) (8 ordinal, 3 approval)
A. Procaccia --> 8 paper(s) (ordinal)
A. Slinko --> 8 paper(s) (7 ordinal, 1 approval)
N. Boehmer --> 7 paper(s) (ordinal)
N. Mattei --> 7 paper(s) (5 ordinal, 2 approval)
N. Shah --> 7 paper(s) (6 ordinal, 1 approval)
L. Xia --> 7 paper(s) (ordinal)
C. Boutilier --> 6 paper(s) (ordinal)
U. Endriss --> 6 paper(s) (4 ordinal, 2 approval)
J. Lang --> 6 paper(s) (3 ordinal, 3 approval)
O. Lev --> 6 paper(s) (ordinal)
D. Peters --> 6 paper(s) (4 ordinal, 2 approval)
T. Walsh --> 6 paper(s) (ordinal)
R. Bredebeck --> 5 paper(s) (4 ordinal, 1 approval)
M. Brill --> 5 paper(s) (2 ordinal, 3 approval)
E. Elkind --> 5 paper(s) (3 ordinal, 2 approval)
R. Meir --> 5 paper(s) (3 ordinal, 3 approval)
R. Niedermeier --> 5 paper(s) (4 ordinal, 1 approval)
J. Rosenschein --> 5 paper(s) (ordinal)
F. Rossi --> 5 paper(s) (ordinal)
H. Aziz --> 4 paper(s) (ordinal)
F. Brandt --> 4 paper(s) (ordinal)
I. Caragiannis --> 4 paper(s) (ordinal)
S. Kraus --> 4 paper(s) (ordinal)
Y. Lewenberg --> 4 paper(s) (ordinal)
S. Nath --> 4 paper(s) (ordinal)
K. Sornat --> 4 paper(s) (2 ordinal, 2 approval)
A. Wilczynski --> 4 paper(s) (ordinal)

Basic Statistics

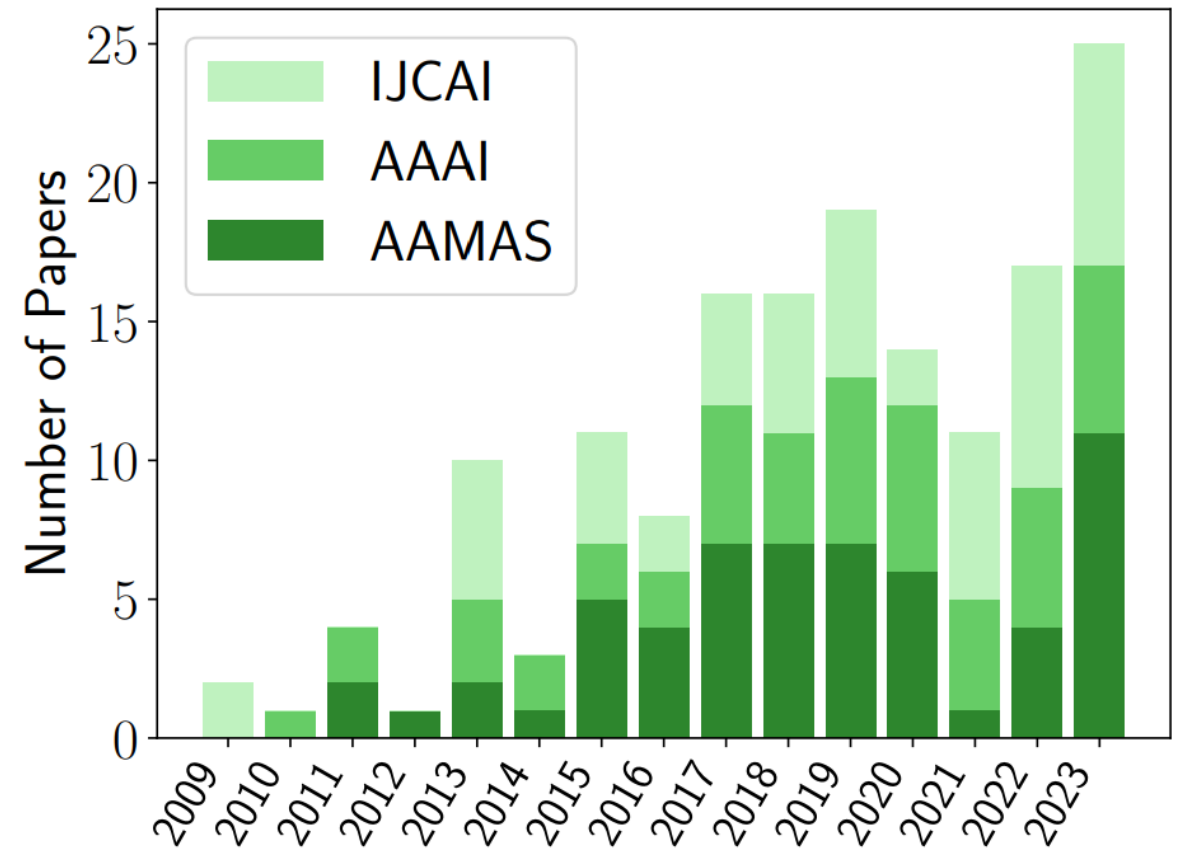
- Papers: 158
 - 125 ordinal
 - 35 approval)
 - **Puzzle?**
- Experiments: 250
 - 204 ordinal
 - 46 approval
- Authors: 273 (+/-)

P. Faliszewski --> 26 paper(s) (18 ordinal, 8 approval)
P. Skowron --> 14 paper(s) (8 ordinal, 6 approval)
N. Talmon --> 14 paper(s) (11 ordinal, 3 approval)
M. Lackner --> 12 paper(s) (3 ordinal, 9 approval)
S. Szufa --> 11 paper(s) (8 ordinal, 3 approval)
A. Procaccia --> 8 paper(s) (ordinal)
A. Slinko --> 8 paper(s) (7 ordinal, 1 approval)
N. Boehmer --> 7 paper(s) (ordinal)
N. Mattei --> 7 paper(s) (5 ordinal, 2 approval)
N. Shah --> 7 paper(s) (6 ordinal, 1 approval)
L. Xia --> 7 paper(s) (ordinal)
C. Boutilier --> 6 paper(s) (ordinal)
U. Endriss --> 6 paper(s) (4 ordinal, 2 approval)
J. Lang --> 6 paper(s) (3 ordinal, 3 approval)
O. Lev --> 6 paper(s) (ordinal)
D. Peters --> 6 paper(s) (4 ordinal, 2 approval)
T. Walsh --> 6 paper(s) (ordinal)
R. Brederick --> 5 paper(s) (4 ordinal, 1 approval)
M. Brill --> 5 paper(s) (2 ordinal, 3 approval)
E. Elkind --> 5 paper(s) (3 ordinal, 2 approval)
R. Meir --> 5 paper(s) (3 ordinal, 3 approval)
R. Niedermeier --> 5 paper(s) (4 ordinal, 1 approval)
J. Rosenschein --> 5 paper(s) (ordinal)
F. Rossi --> 5 paper(s) (ordinal)
H. Aziz --> 4 paper(s) (ordinal)
F. Brandt --> 4 paper(s) (ordinal)
I. Caragiannis --> 4 paper(s) (ordinal)
S. Kraus --> 4 paper(s) (ordinal)
Y. Lewenberg --> 4 paper(s) (ordinal)
S. Nath --> 4 paper(s) (ordinal)
K. Sornat --> 4 paper(s) (2 ordinal, 2 approval)
A. Wilczynski --> 4 paper(s) (ordinal)

Experiments on Elections in COMSOC

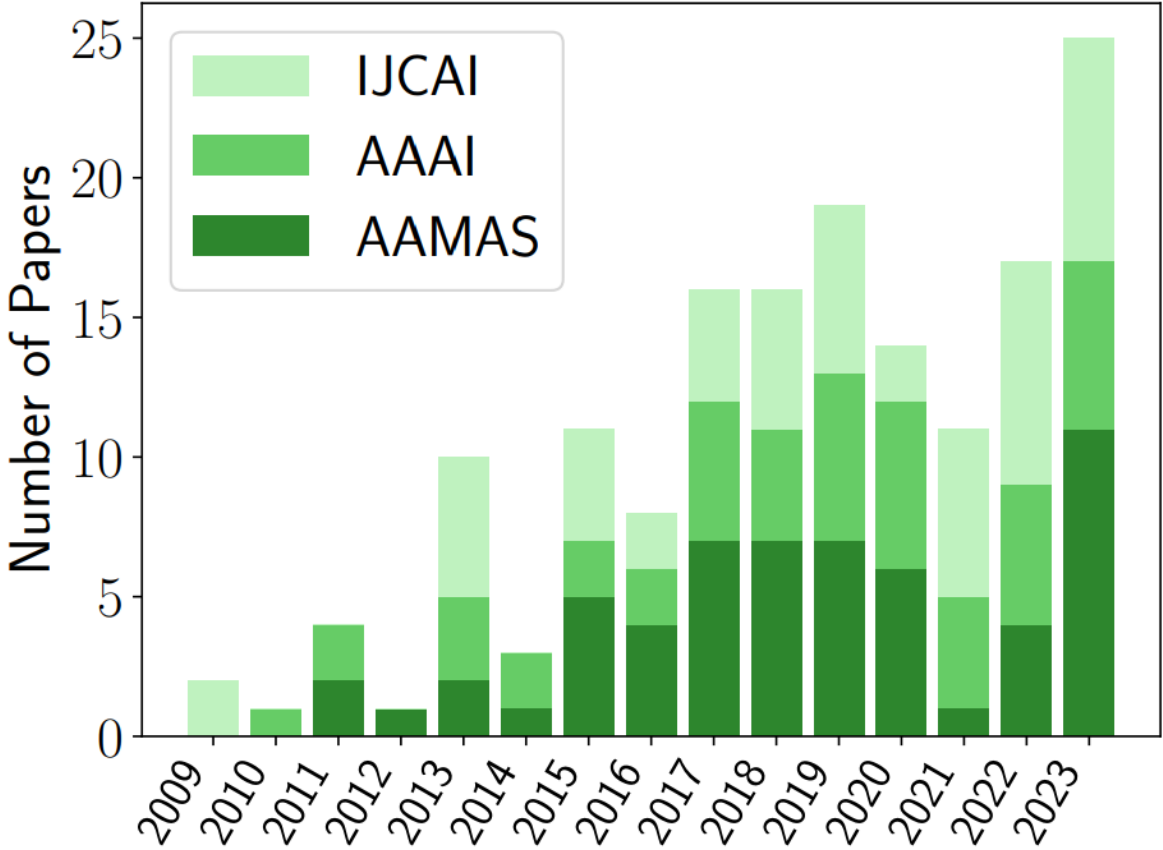


Papers in recent AI conferences

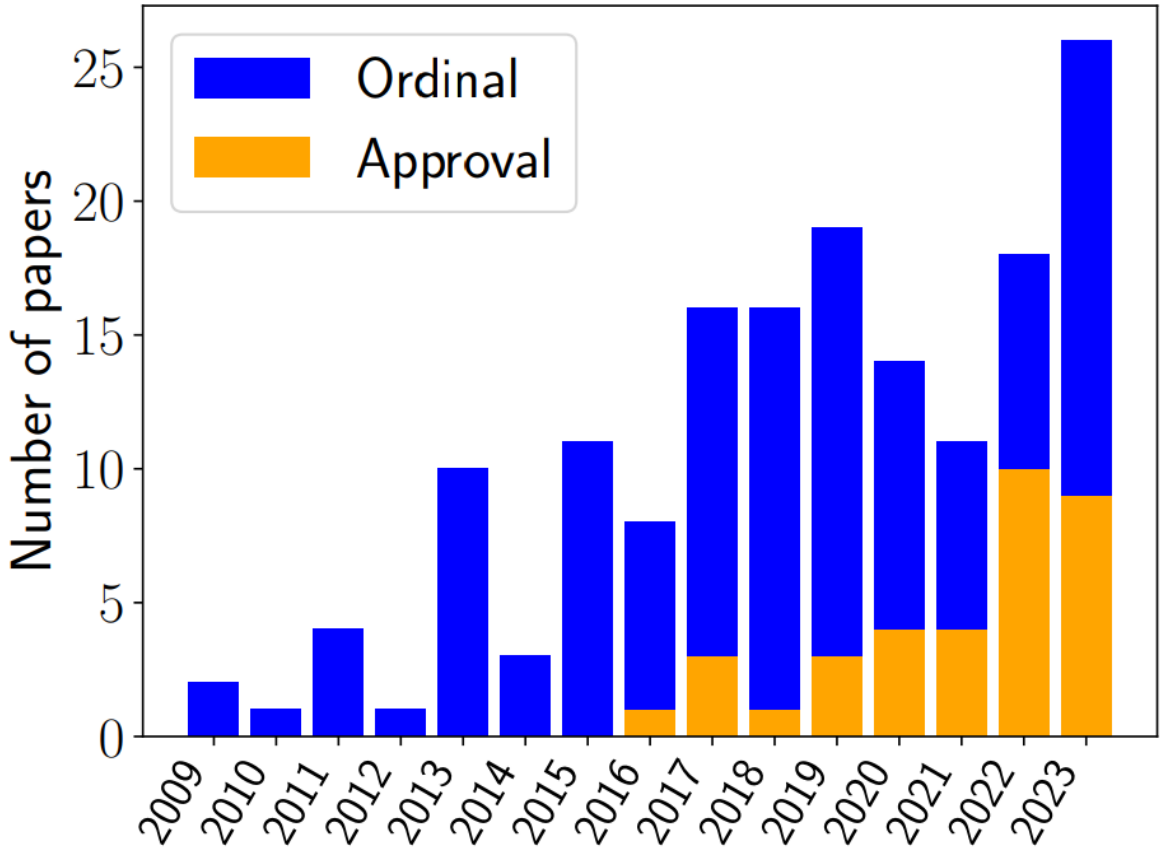


Papers in recent AI conferences that include experiments on elections*

Experiments on Elections in COMSOC



Papers in recent AI conferences that include experiments on elections*



Ordinal preferences versus approval (as covered in the papers)

So Experiments Happen

Are They Any Good?

What Elections to Study?



Structure of
the preference
orders?



Reasonable
numbers of
candidates and
voters?

What Elections to Study?

Structure of the preference orders?

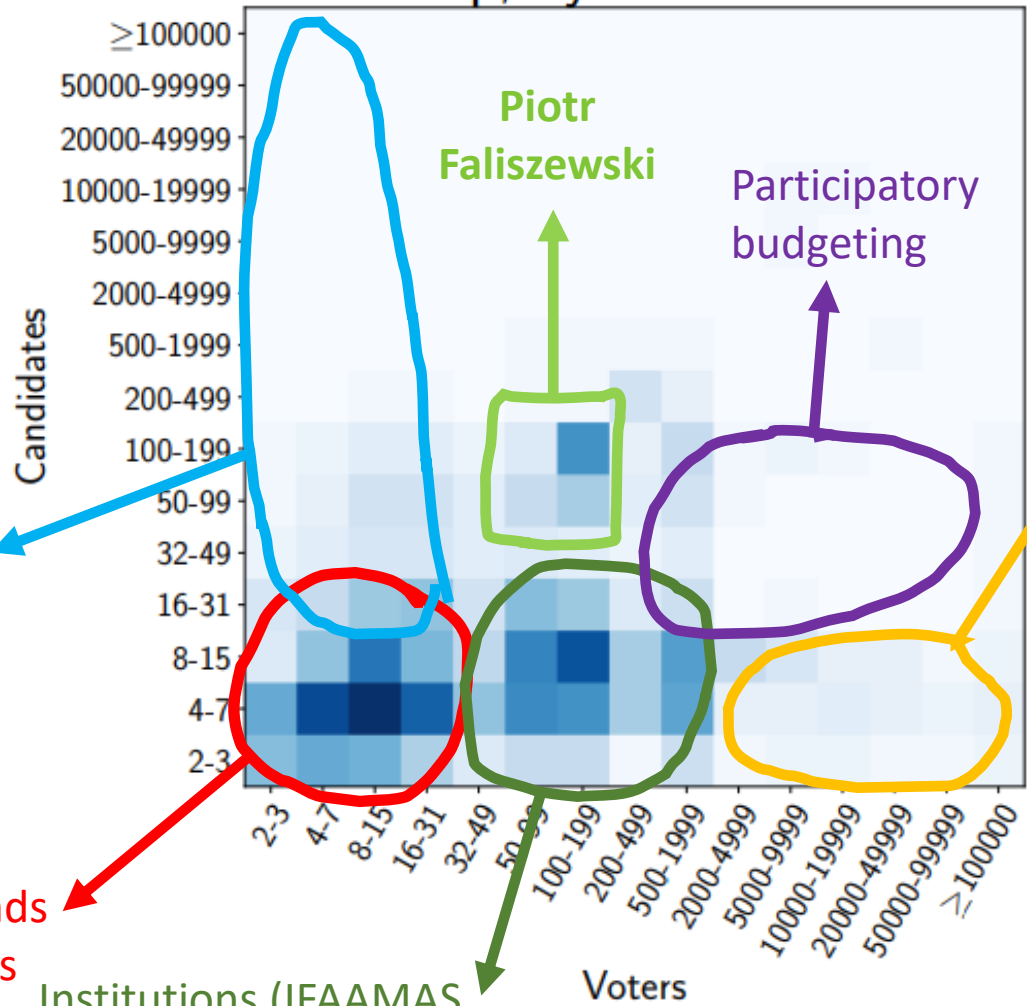
Reasonable numbers of candidates and voters?

Ground-truth search (sporting events, meta-search engines, recommendation systems, etc.)

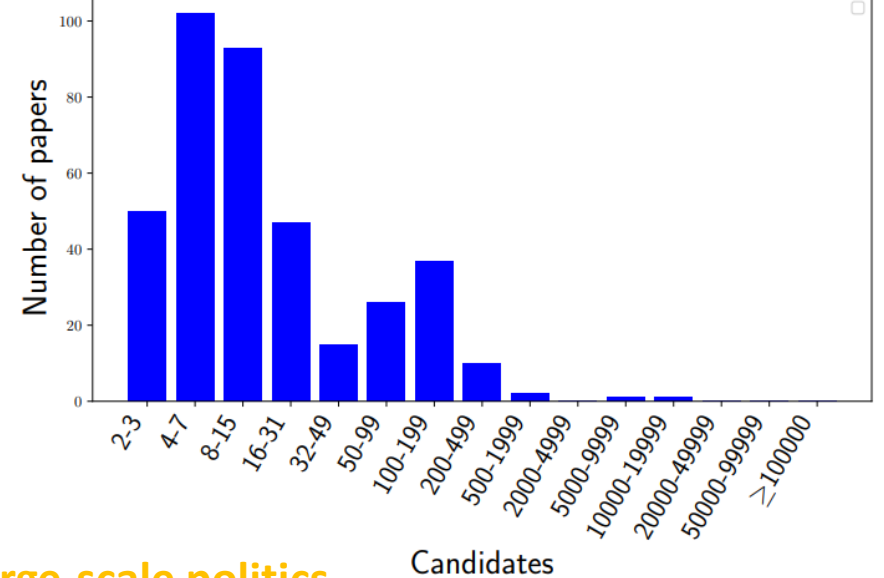
Small committees (e.g., hiring), friends voting on frivolous stuff, „usual life”

Institutions (IFAAMAS board elections, choosing electors at universities, etc.)

Heatmap, Synthetic Elections

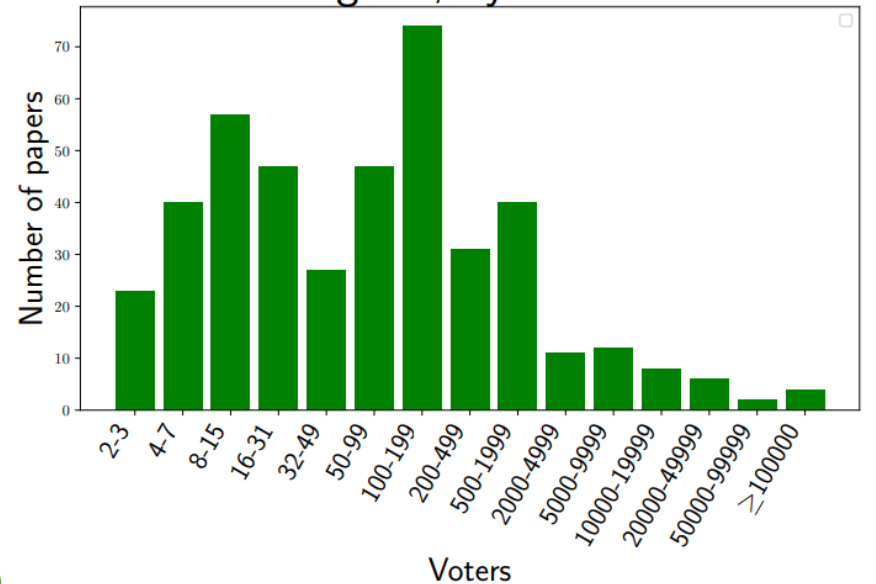


Candidate Histogram, Synthetic Elections



Large-scale politics

Voter Histogram, Synthetic Elections



What Elections to Study?

Structure of the preference orders?

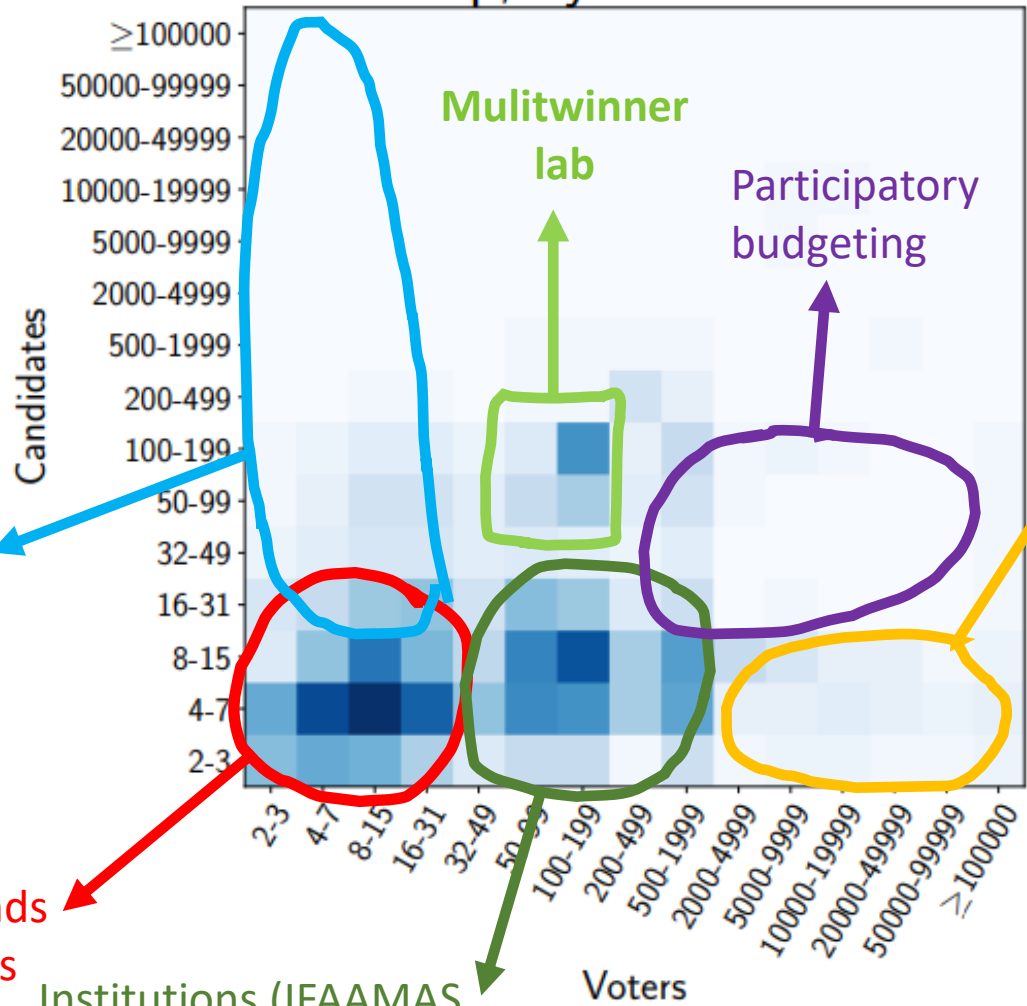
Reasonable numbers of candidates and voters?

Ground-truth search (sporting events, meta-search engines, recommendation systems, etc.)

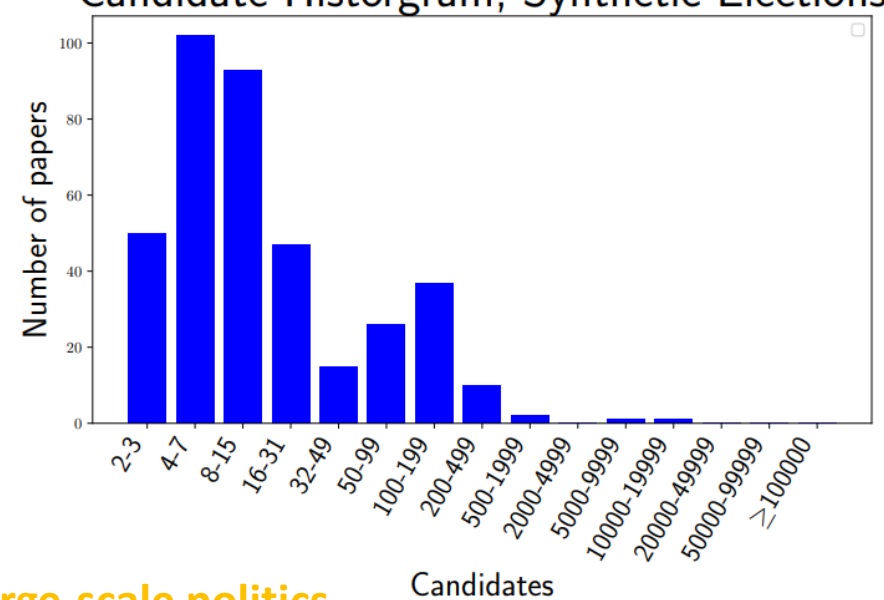
Small committees (e.g., hiring), friends voting on frivolous stuff, „usual life”

Institutions (IFAAMAS board elections, choosing electors at universities, etc.)

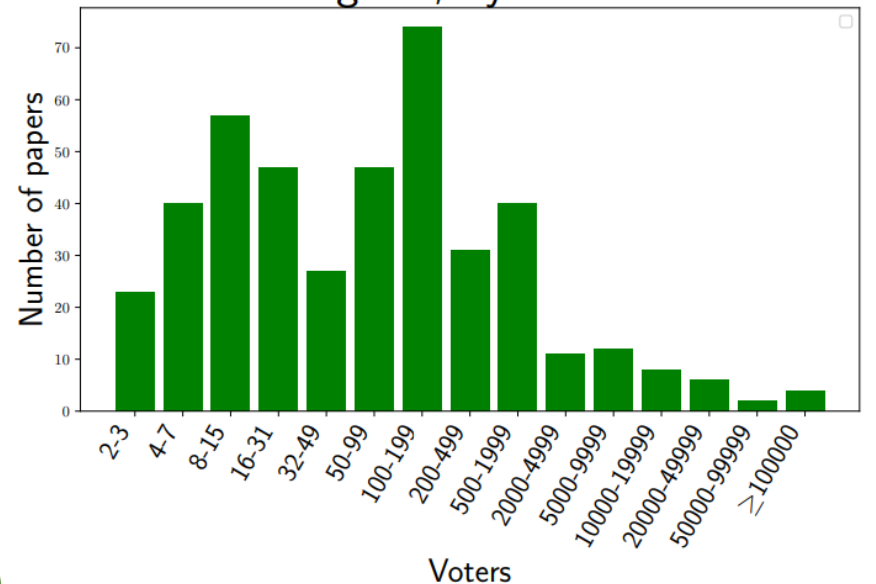
Heatmap, Synthetic Elections



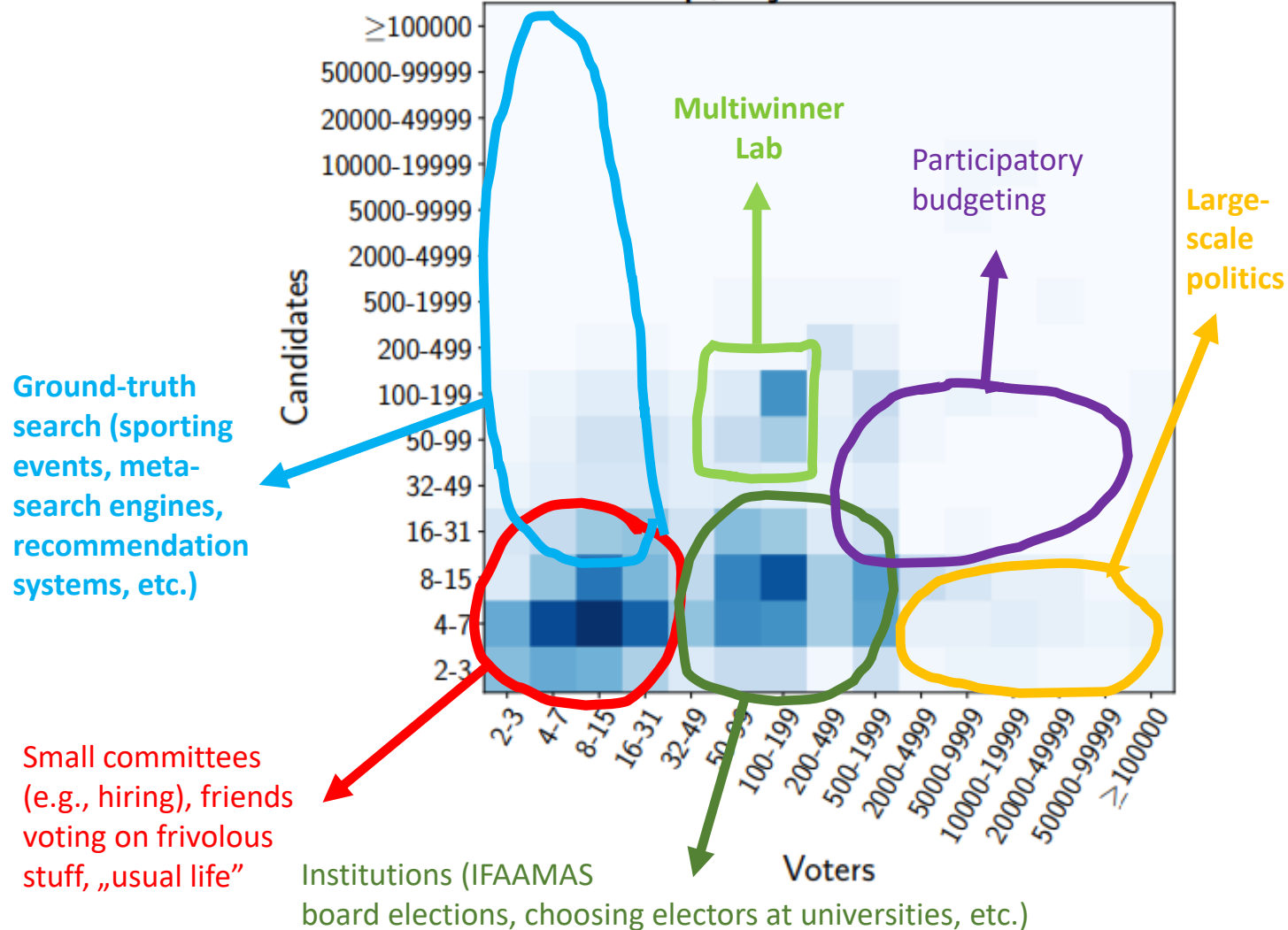
Candidate Histogram, Synthetic Elections



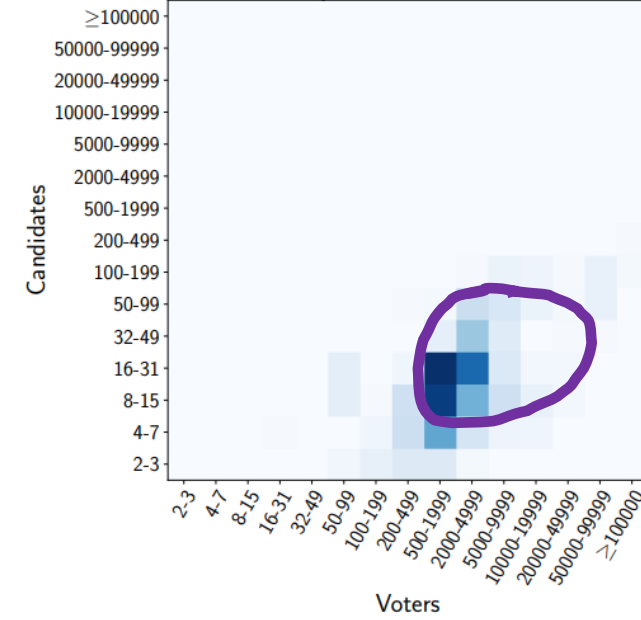
Voter Histogram, Synthetic Elections



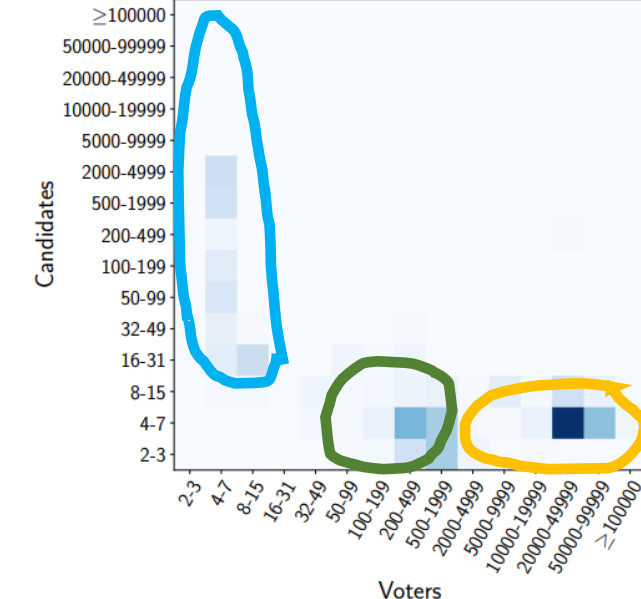
Heatmap, Synthetic Elections



Heatmap, Pabulib Elections



Heatmap, Preflib Elections

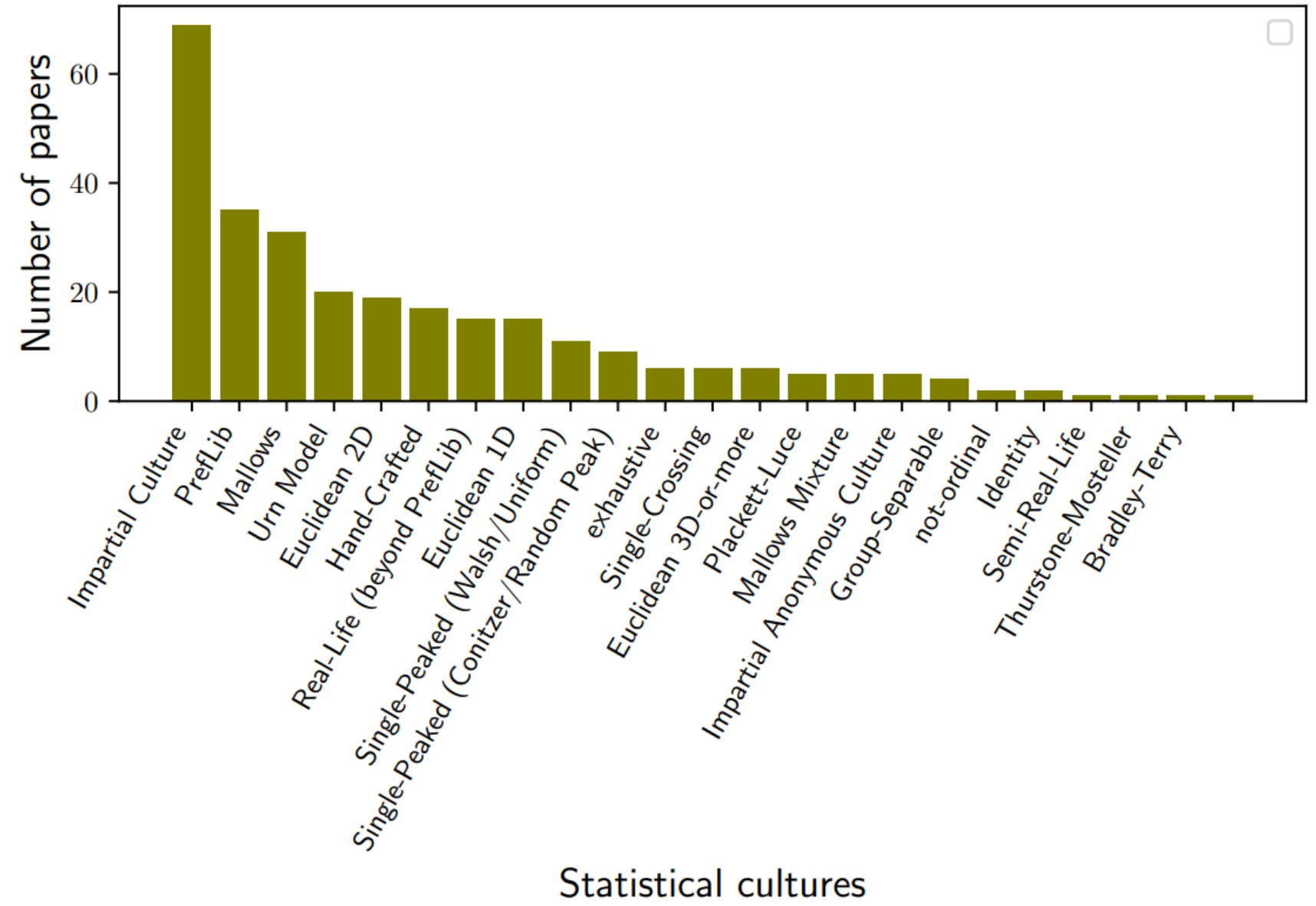


What Elections to Study?

Reasonable numbers of candidates and voters?

Structure of the preference orders?

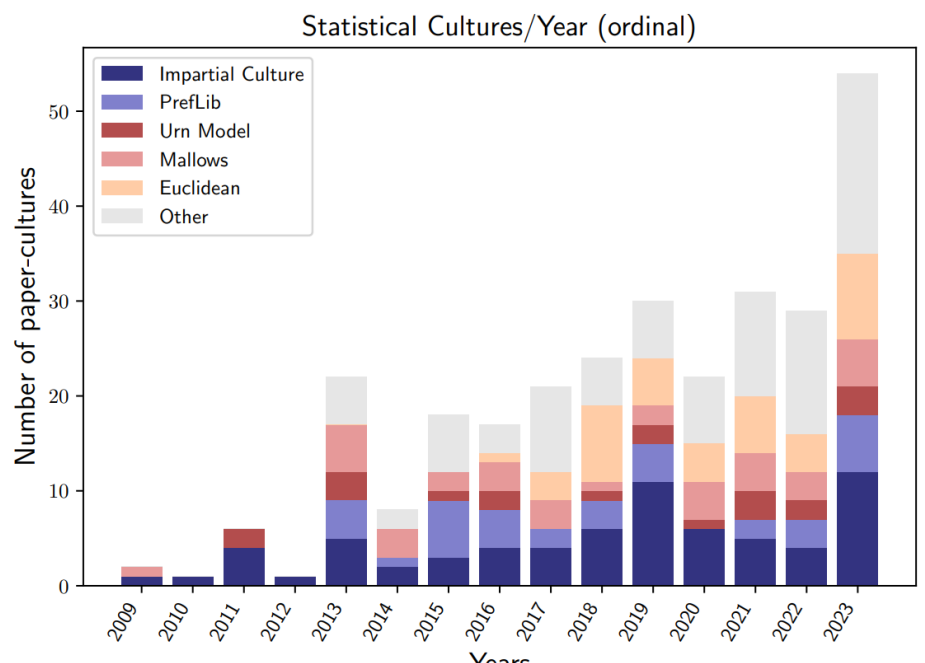
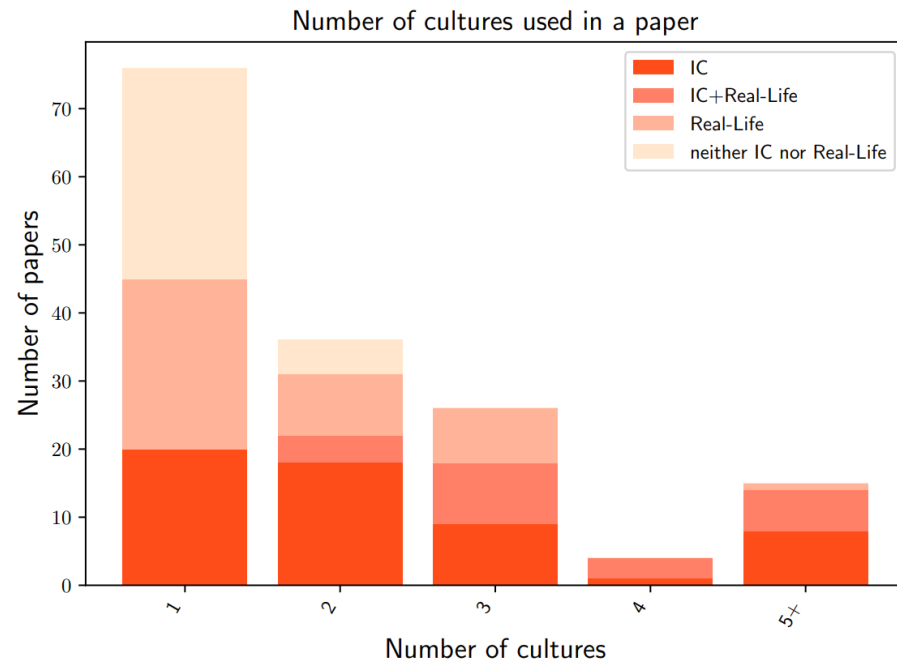
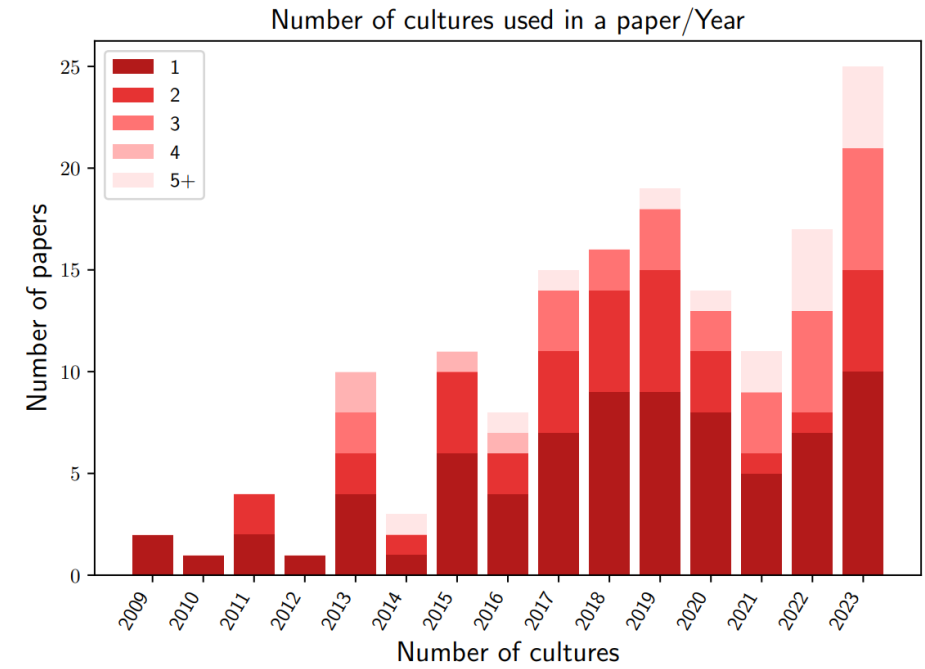
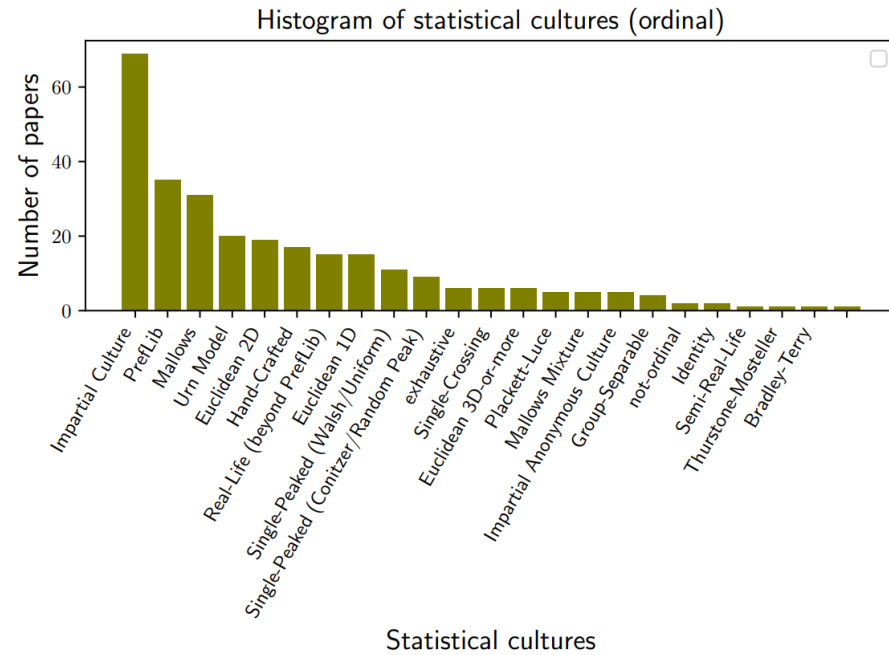
Histogram of statistical cultures (ordinal)

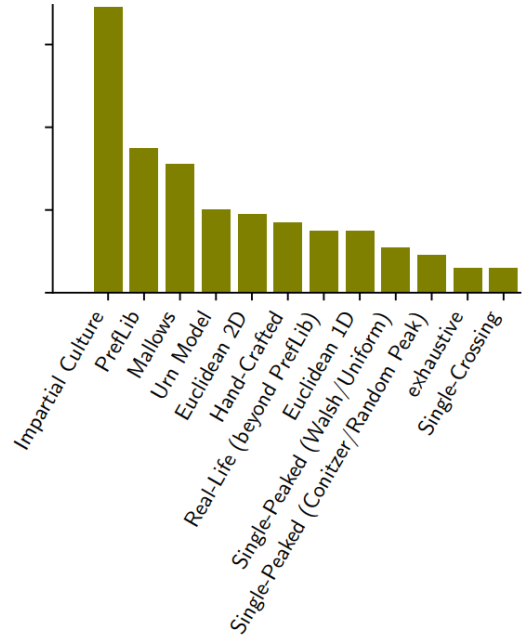


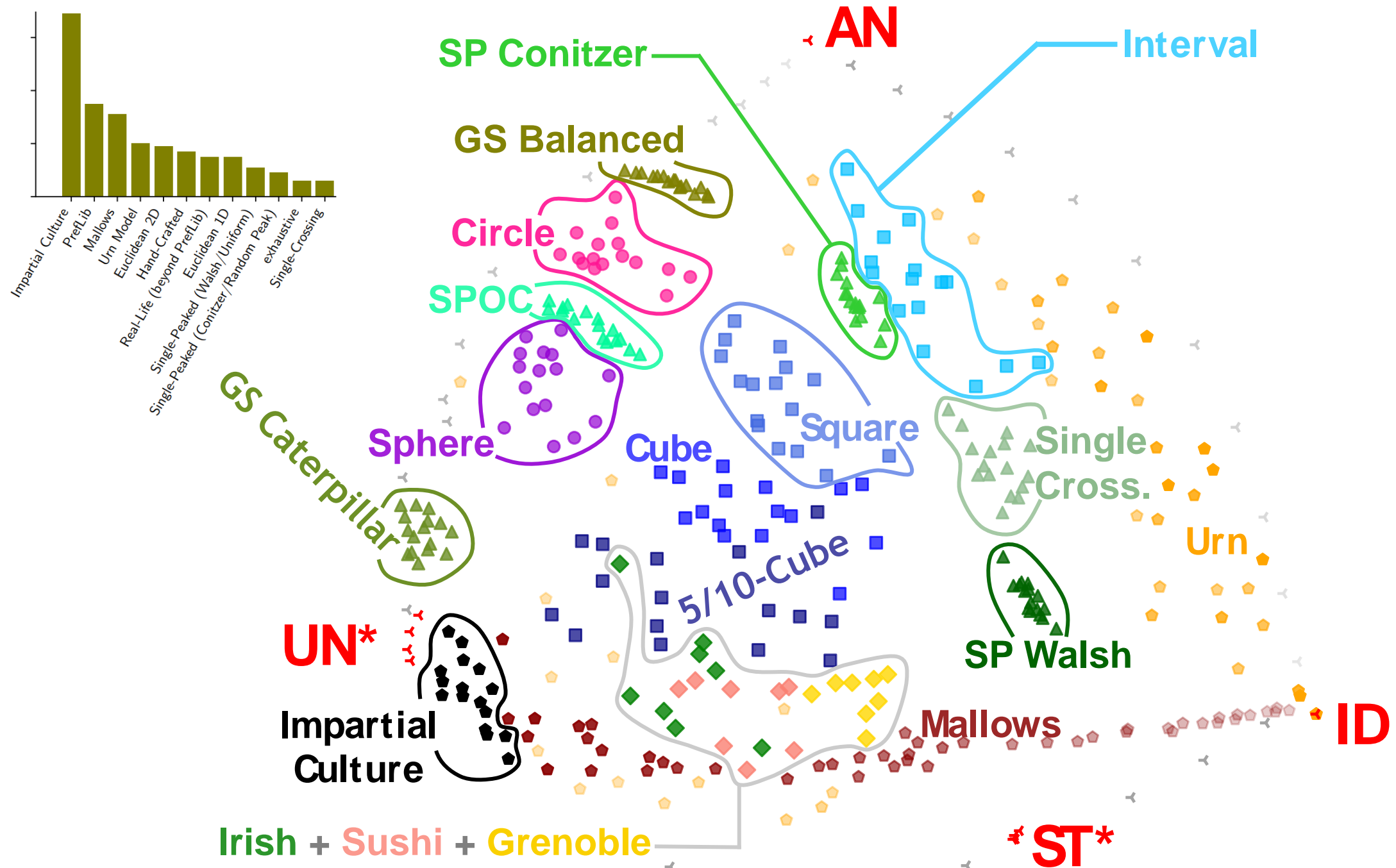
What Elections to Study?

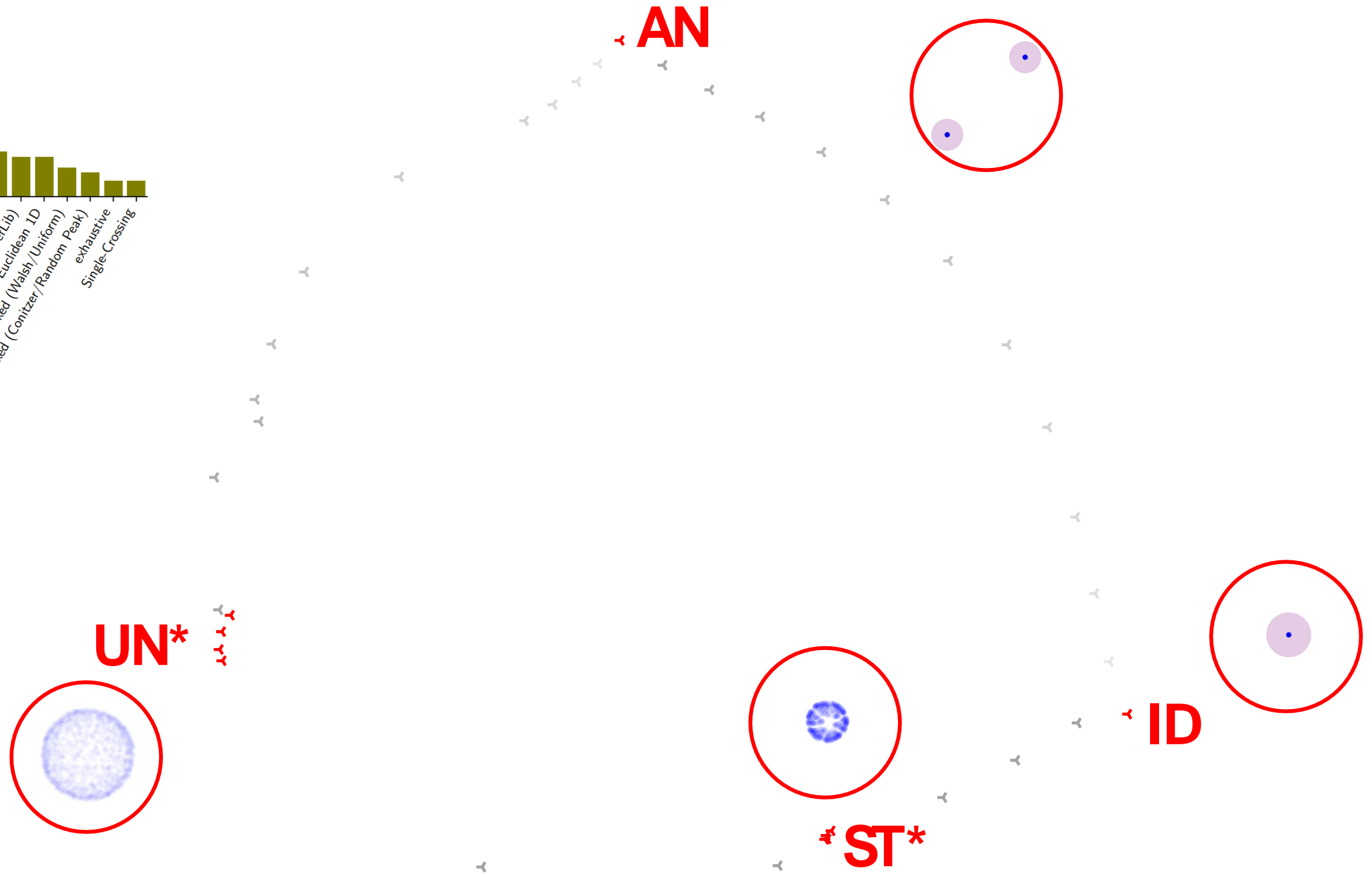
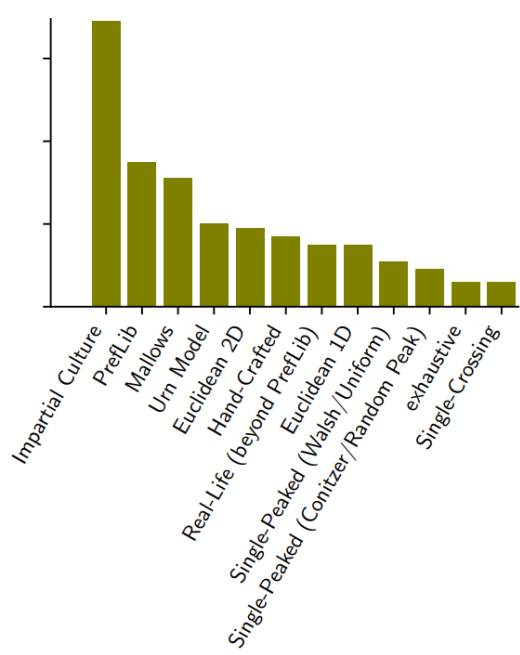
Reasonable numbers of candidates and voters?

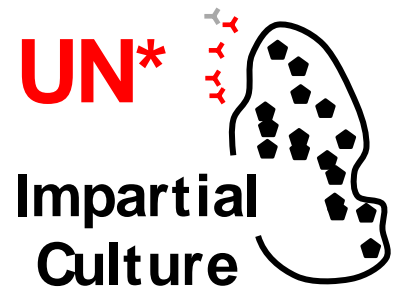
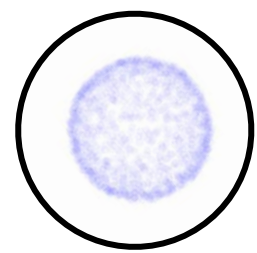
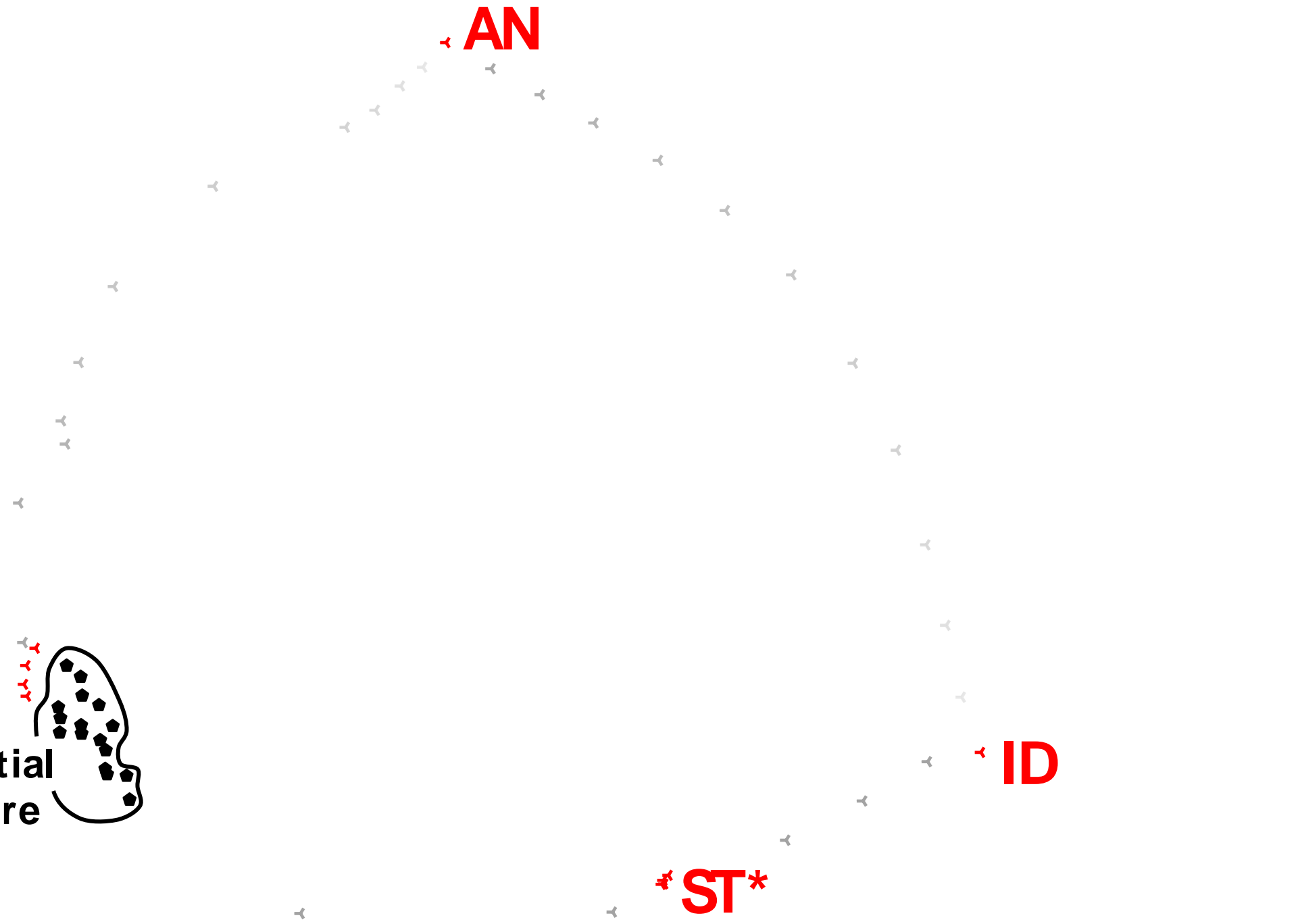
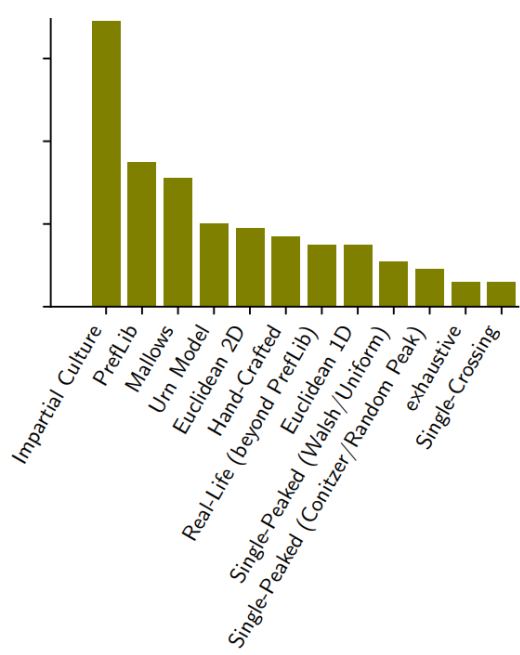
Structure of the preference orders?







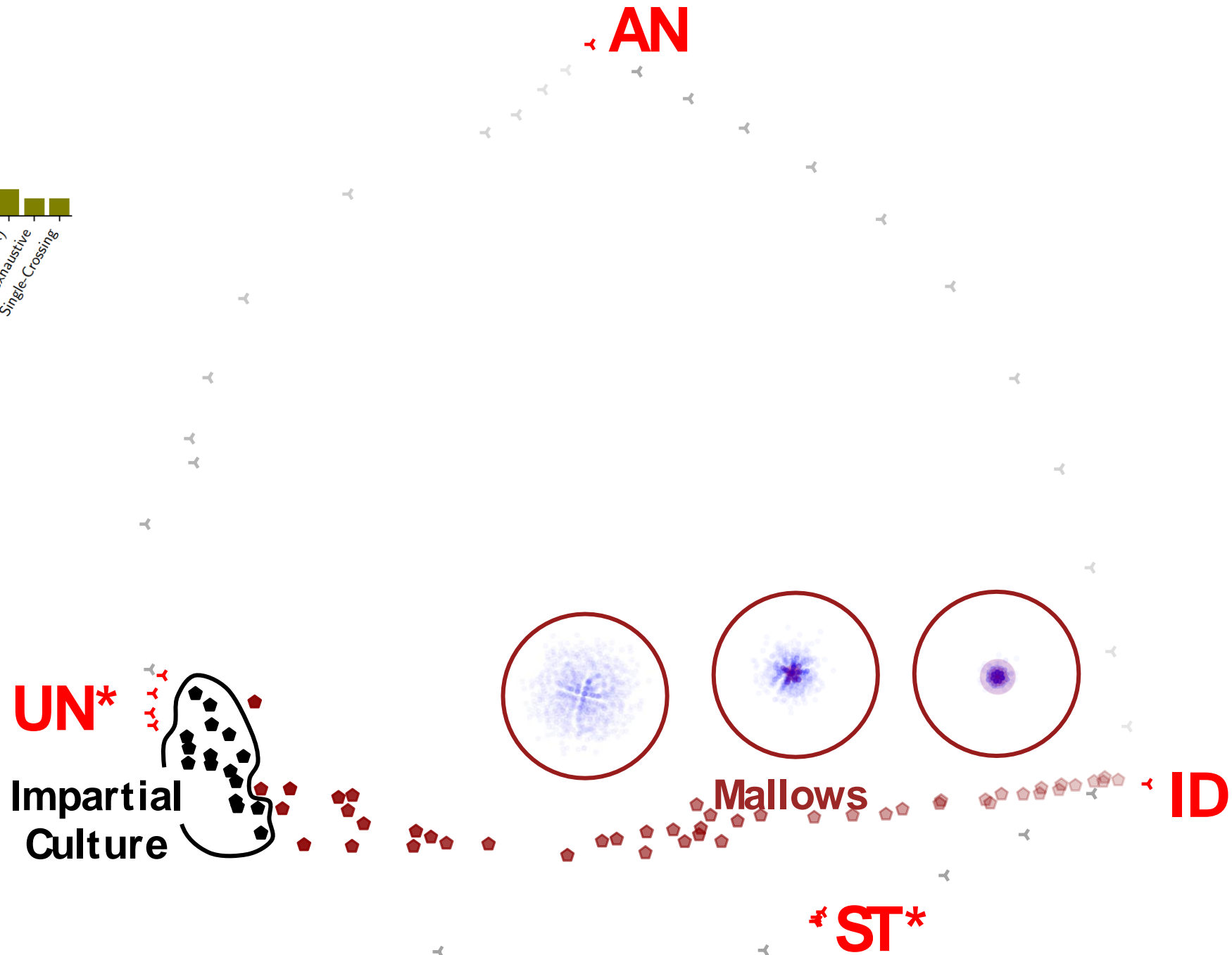
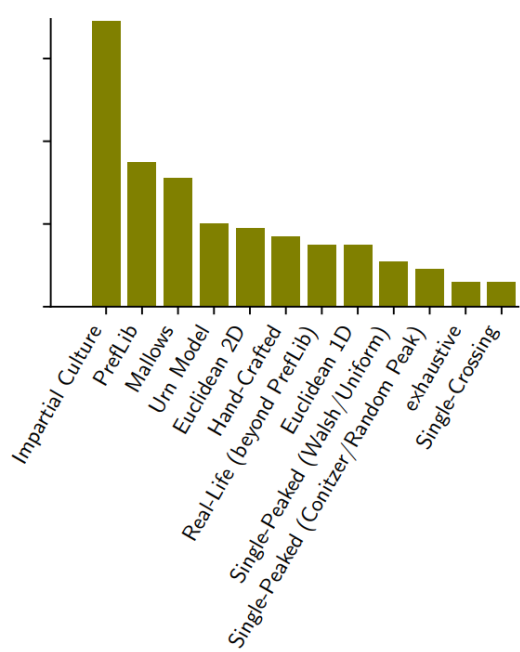


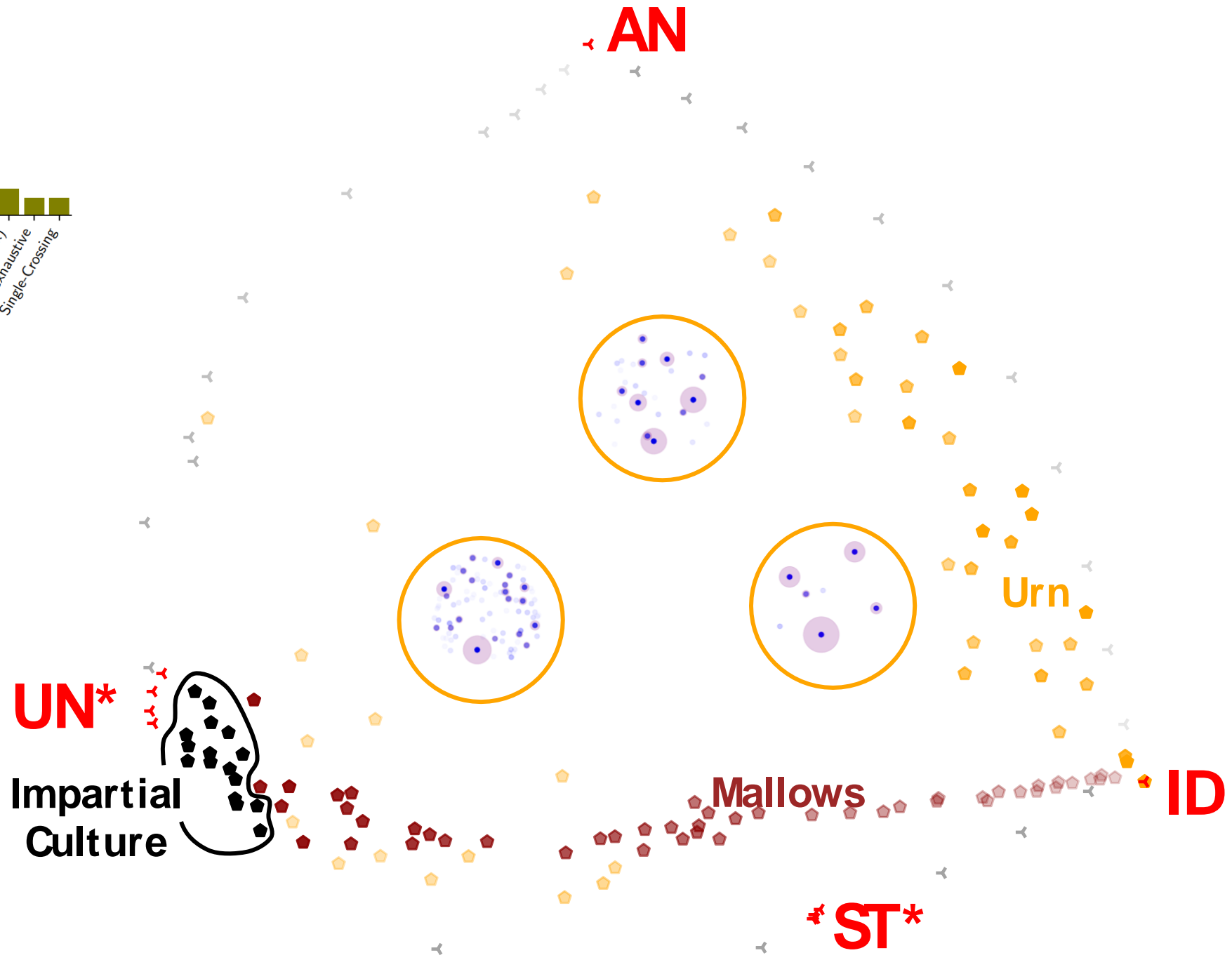
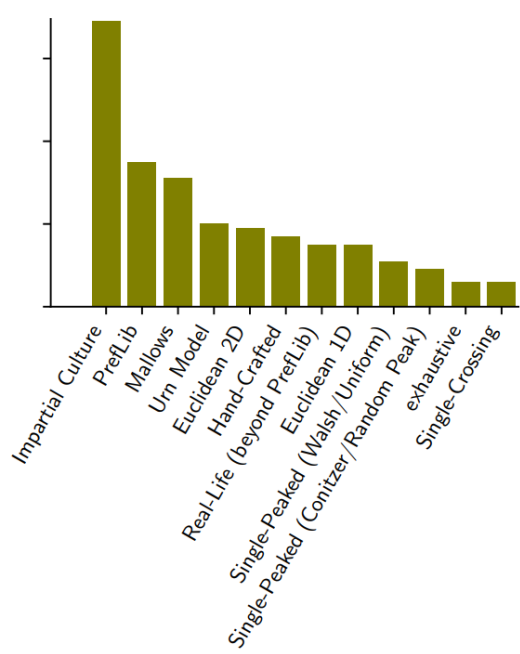


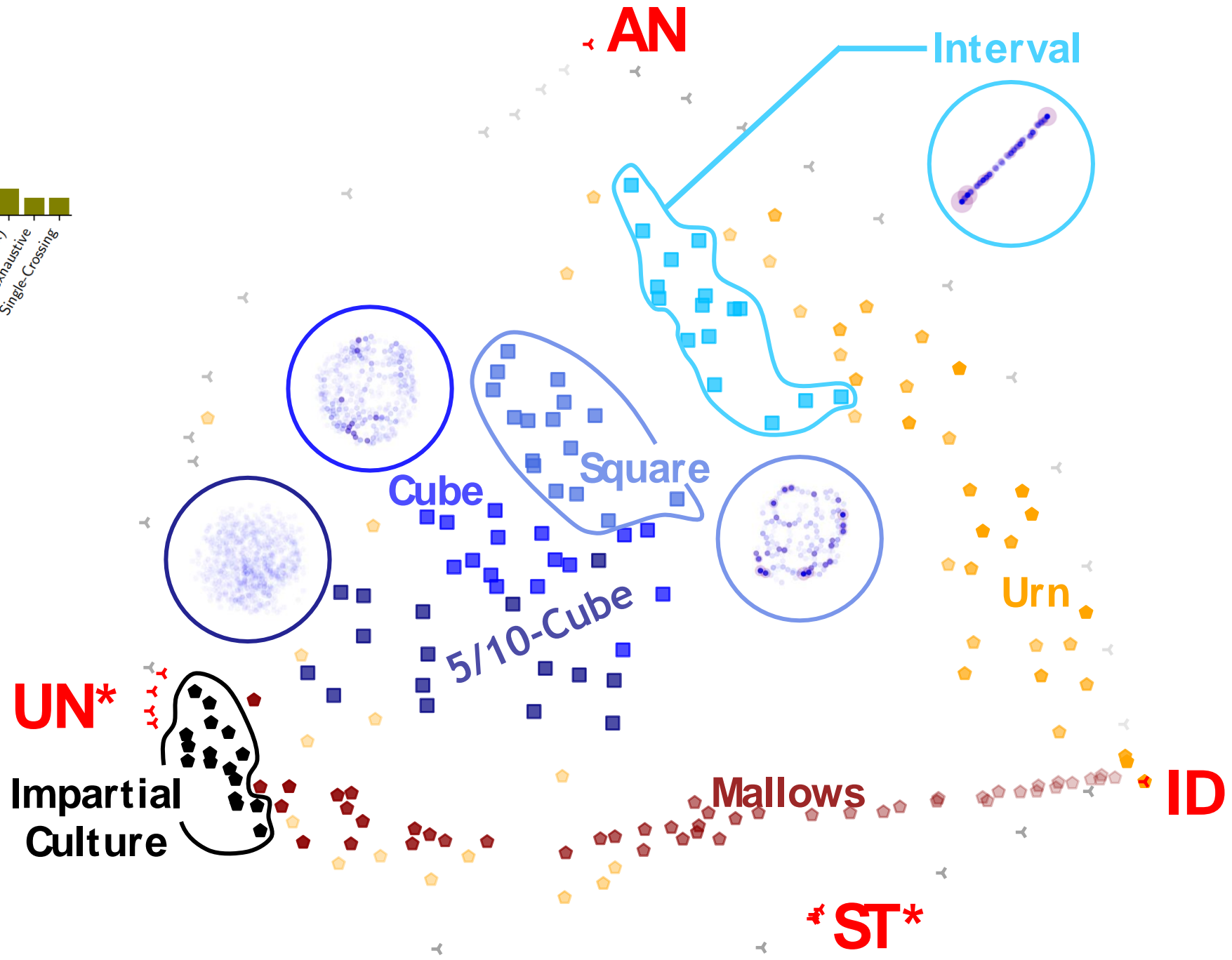
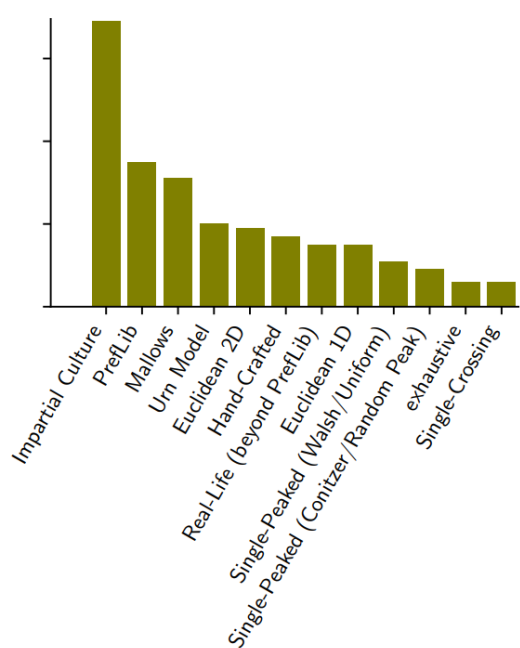
ST

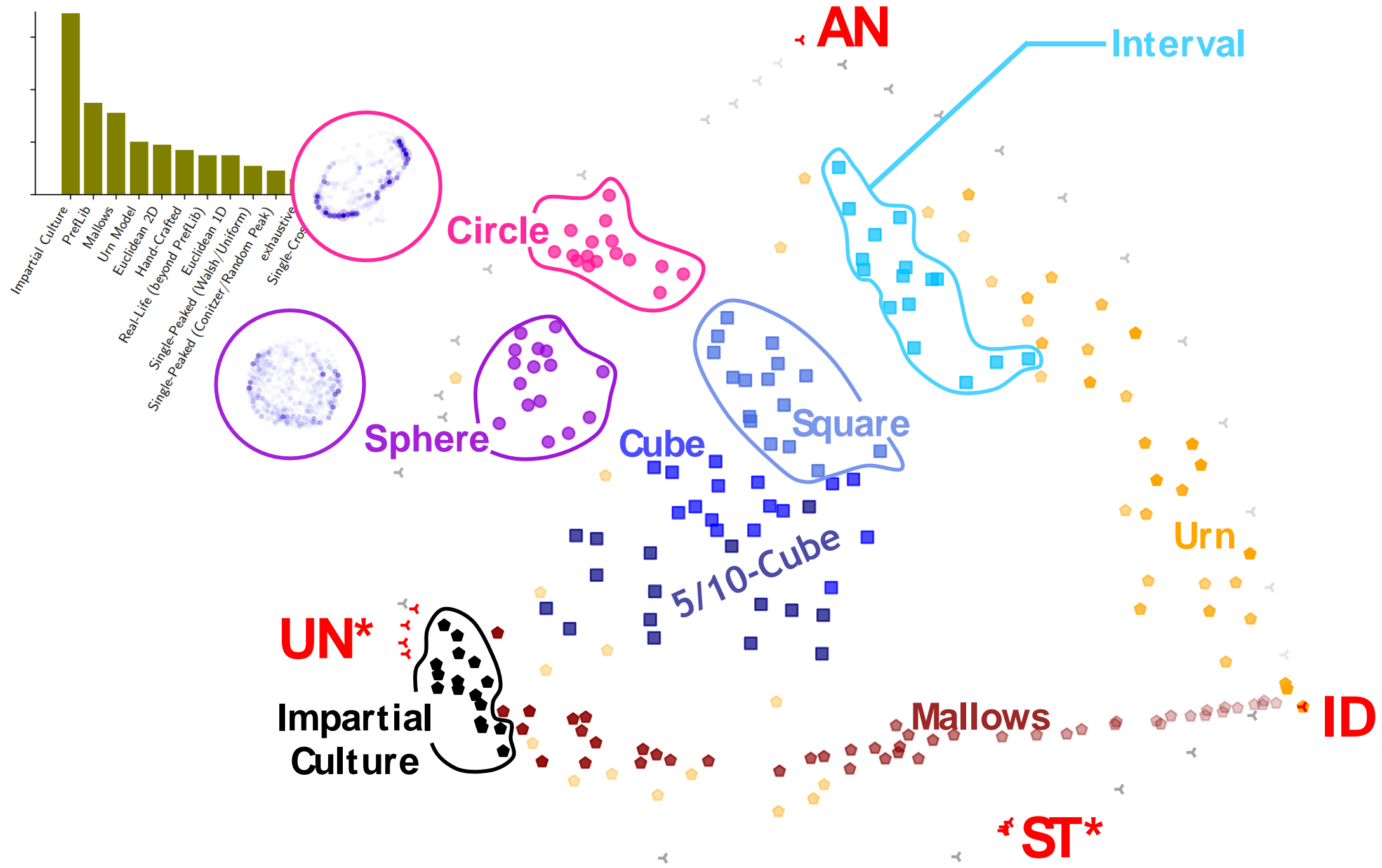
ID

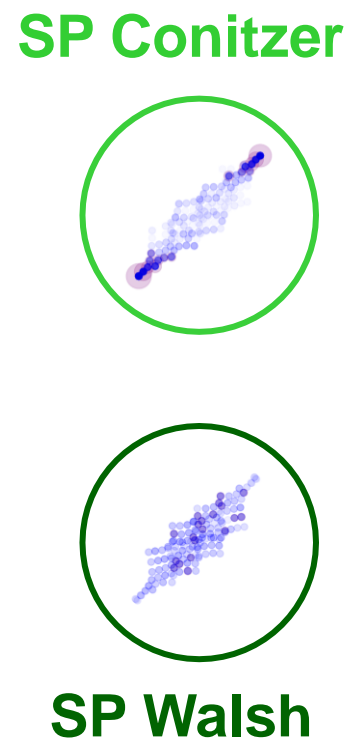
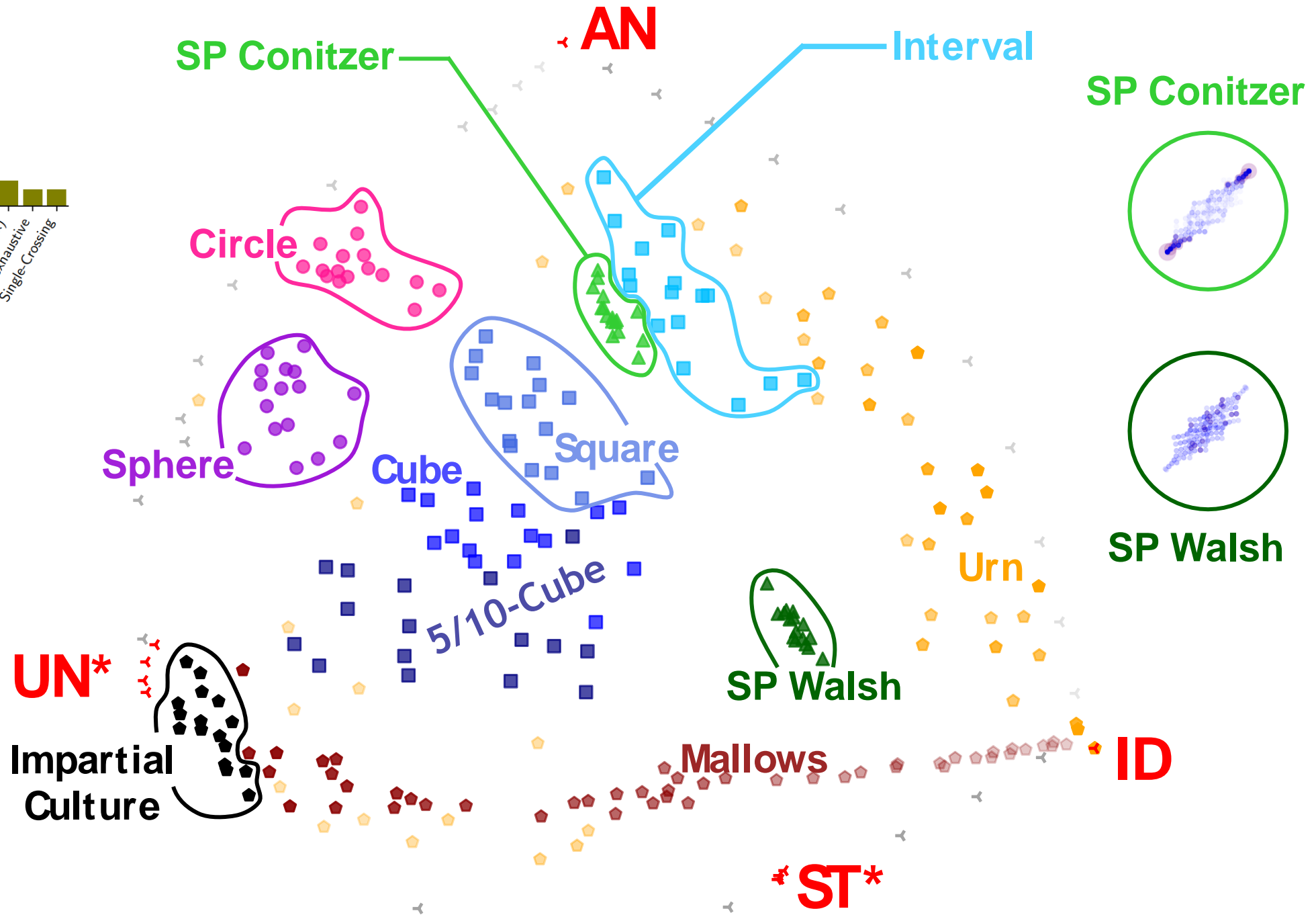
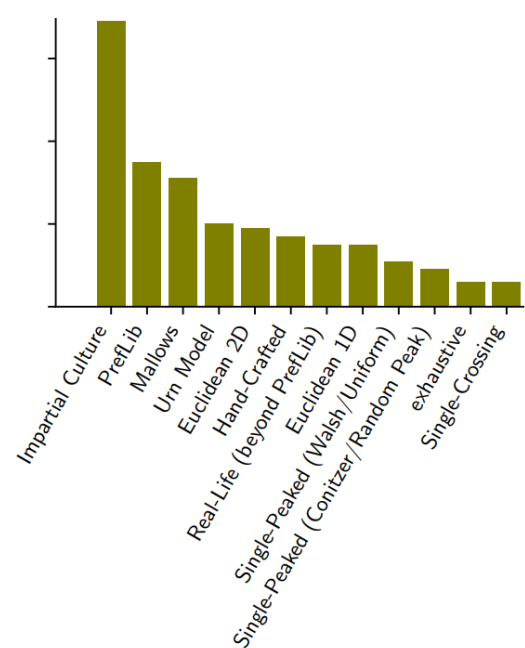
AN

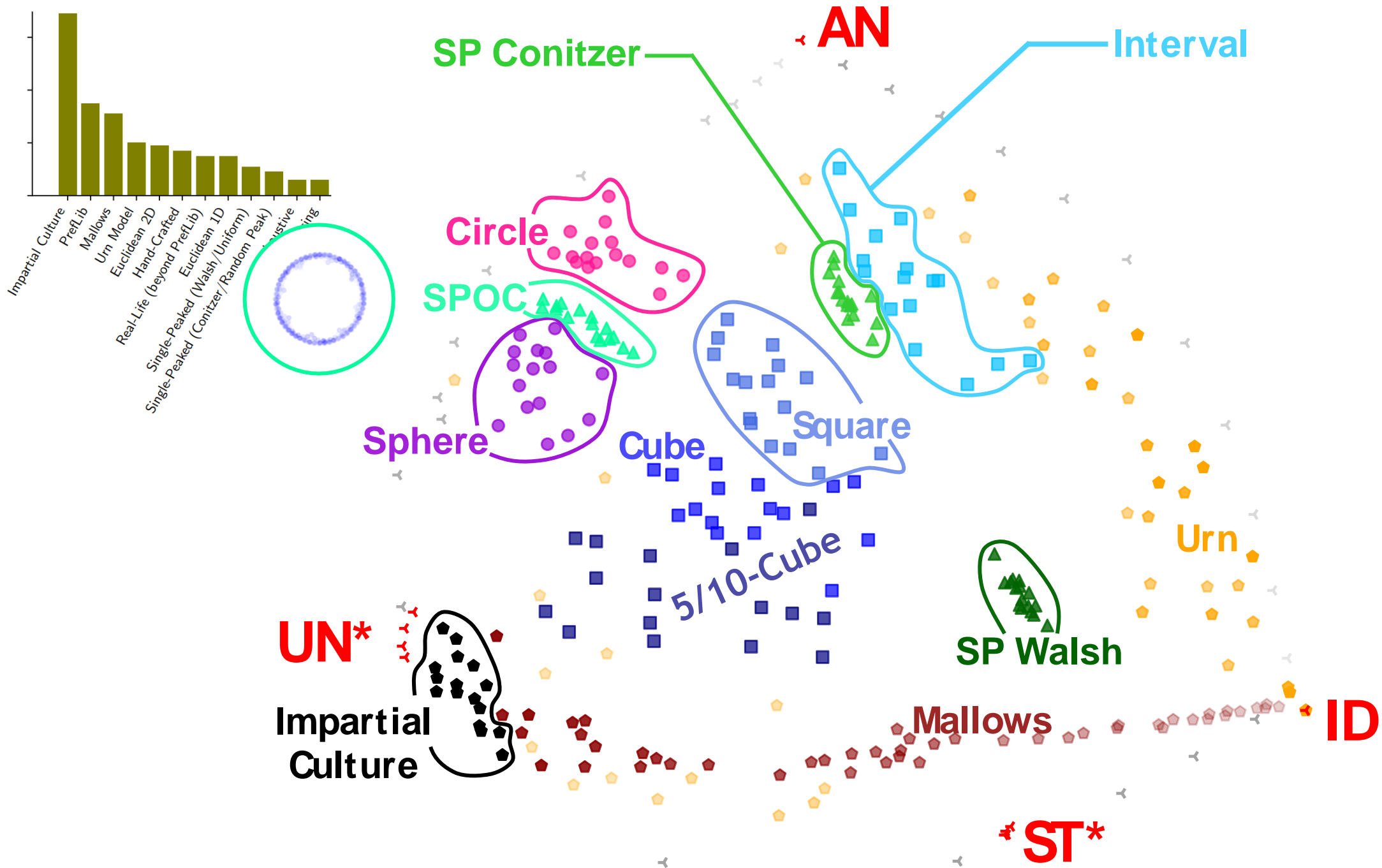


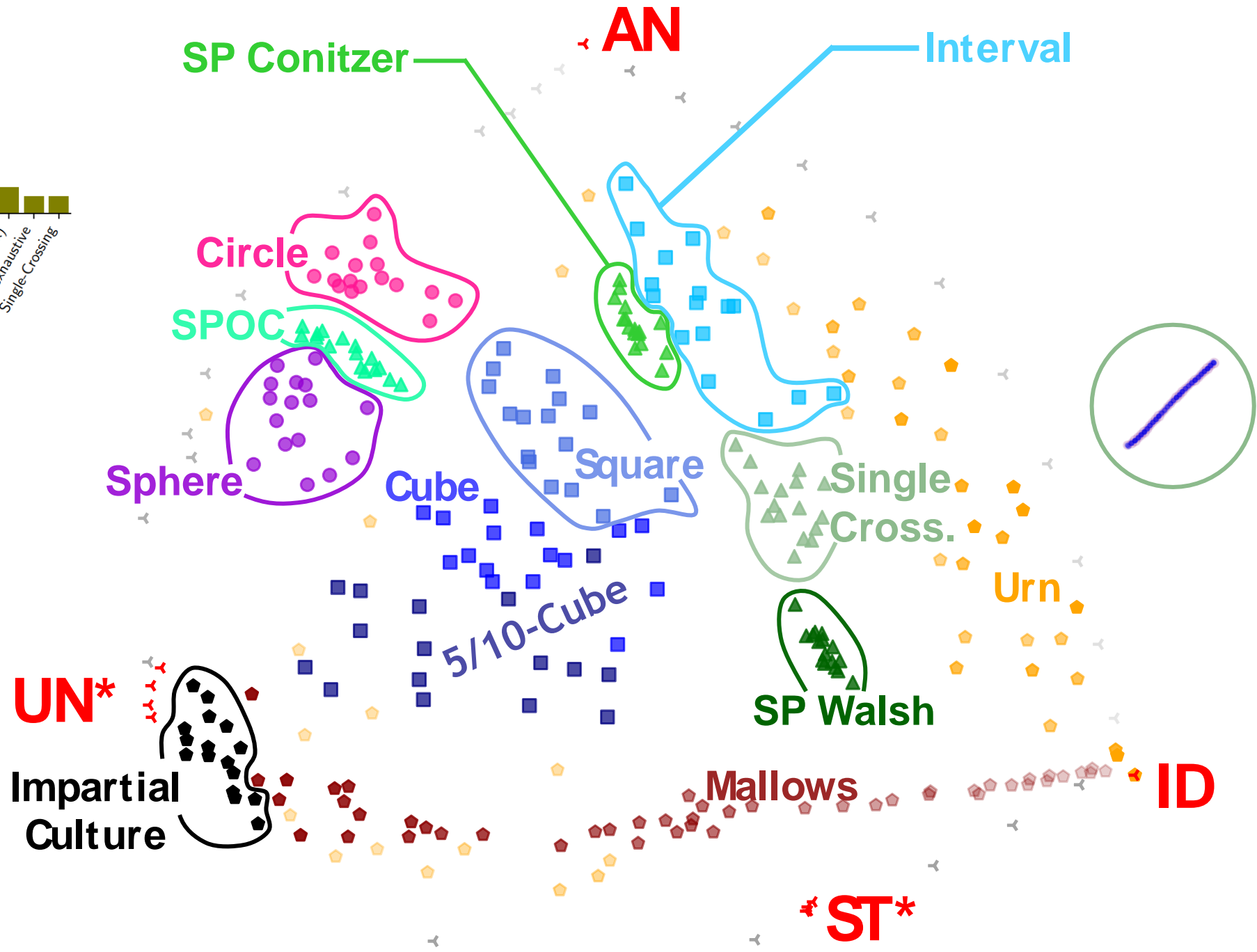
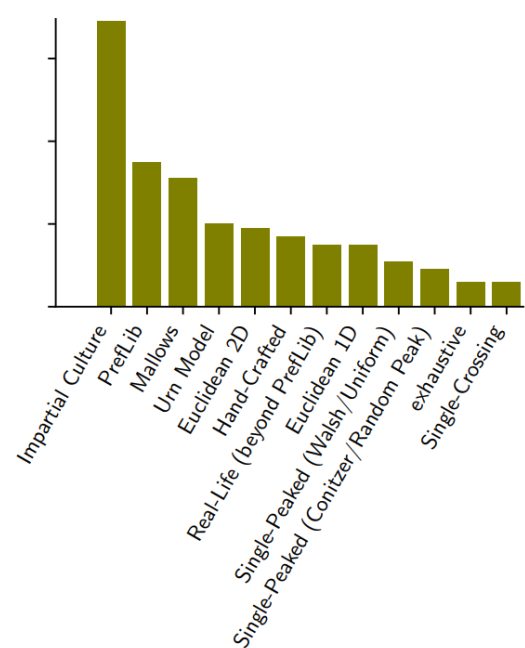


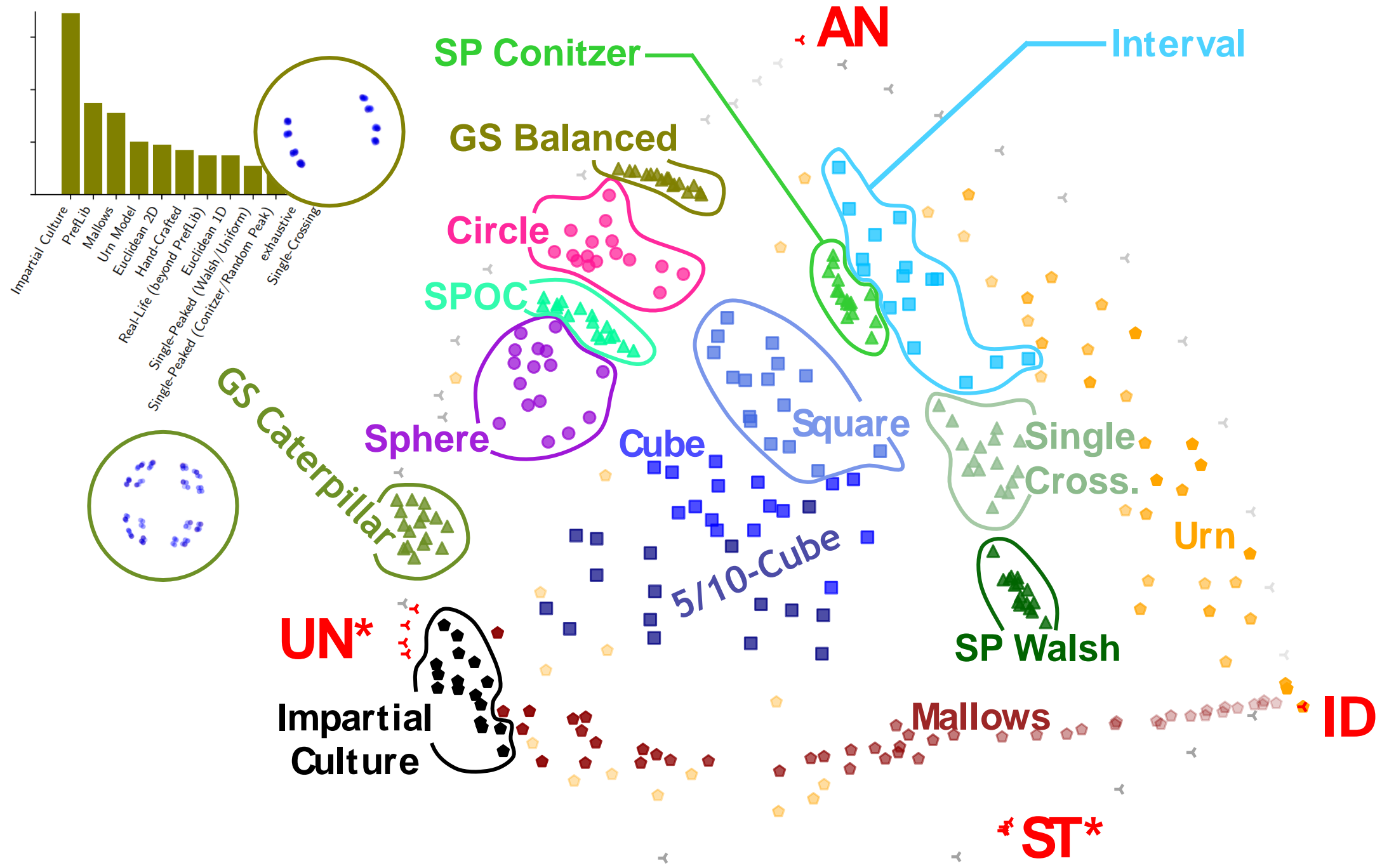


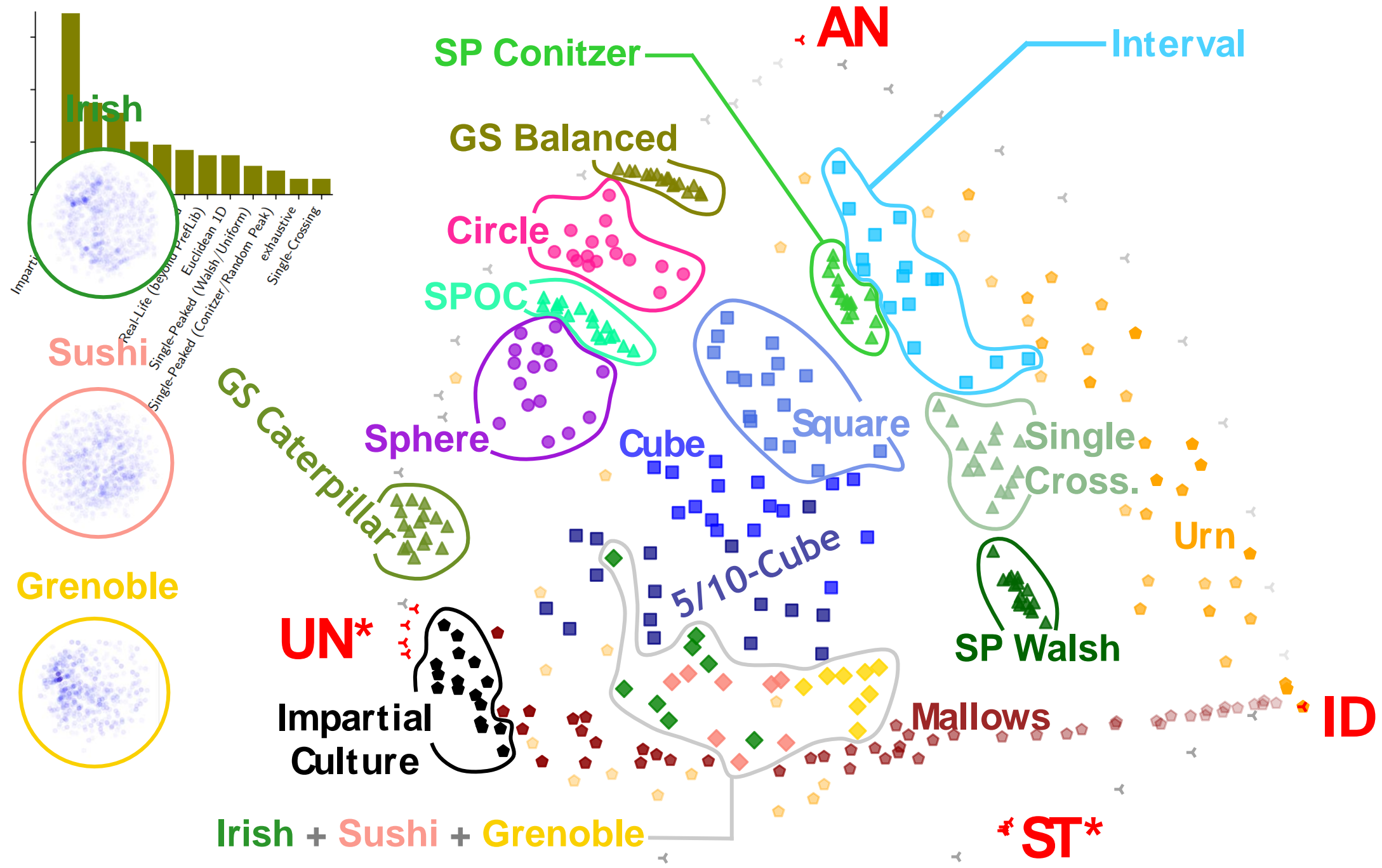


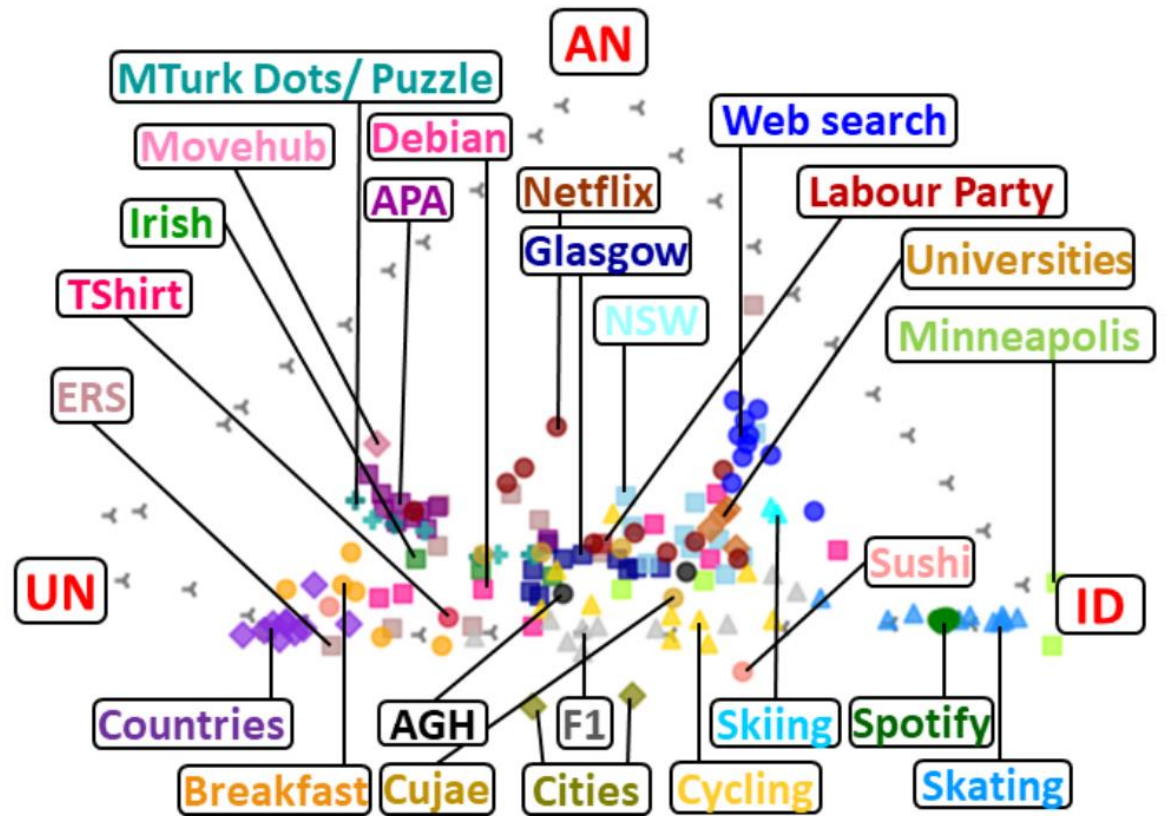
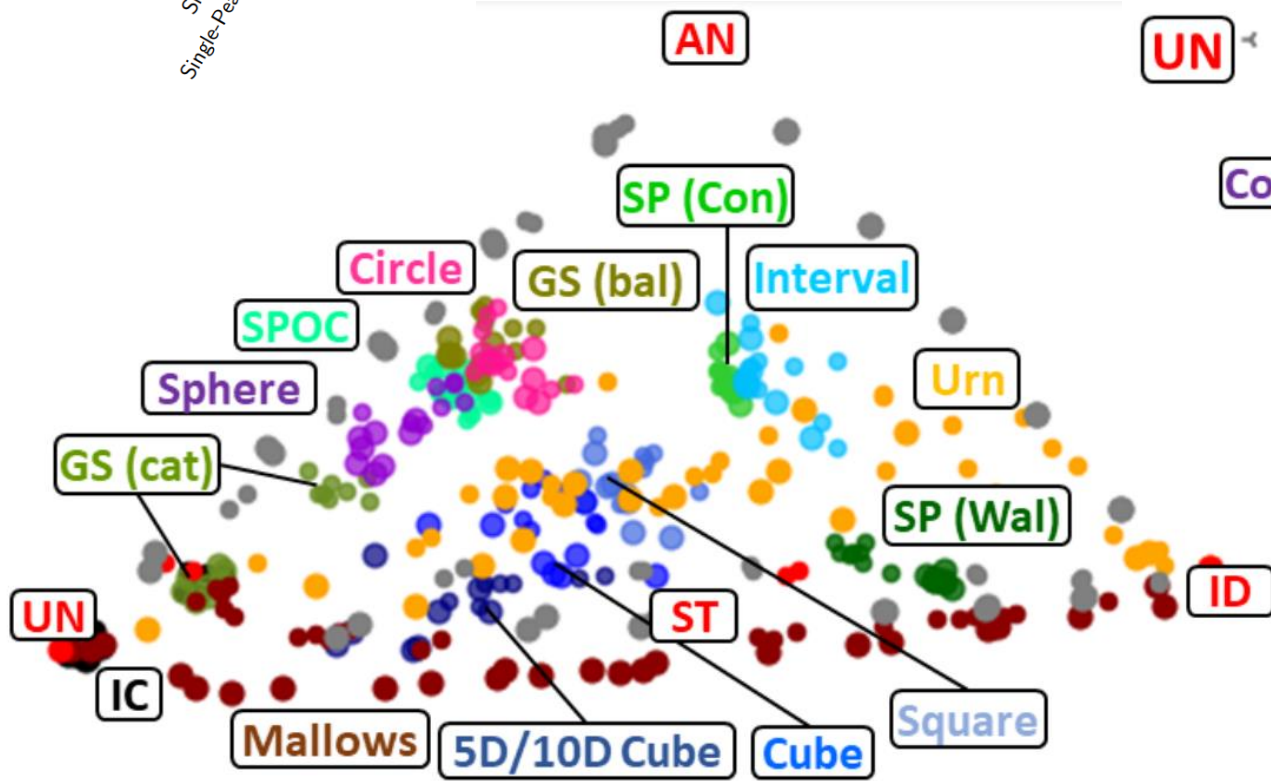
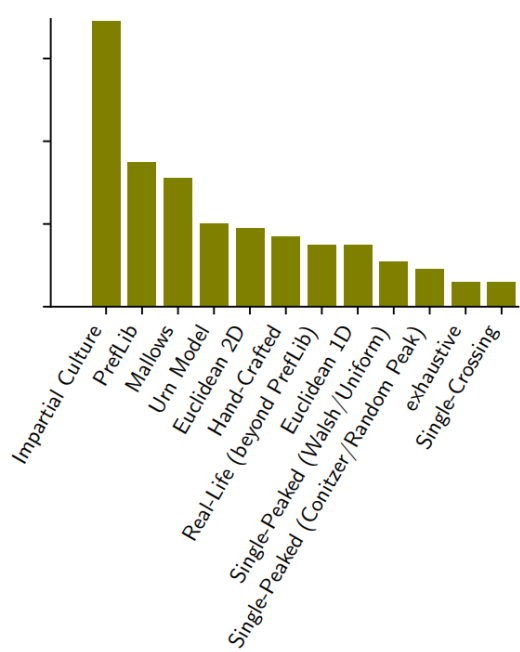












Synthetic and
Preflib Elections

Conclusions

- Experiments have become mainstream
 - Not compulsory though (yet)
- Election sizes
 - Often taken without particularly deep motivations
 - Somehow fit a reasonable picture, though!
- Many papers use just one statistical culture
 - But fewer than one might think!
 - Using multiple data sources is becoming common
 - Using real-life data (Preflib/Pabulib) much more common than one might suspect

Do
More
Experiments!

Please...