FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Assignment of bachelor's thesis

Title:	Maps of elections
Student:	Jitka Mertlová
Supervisor:	doc. RNDr. Dušan Knop, Ph.D.
Study program:	Informatics
Branch / specialization:	Knowledge Engineering
Department:	Department of Applied Mathematics
Validity:	until the end of summer semester 2024/2025

Instructions

When we hear the term "voting system", usually our mind conjures up an image of some sort of election – presidential elections, parliamentary ones, there are many to pick from. The topic of computational social choice is, however, much broader. We pick between different alternatives every day; from the kind of sushi we want for lunch to which thermometers in a room we decide to trust when they all give different measurements. This phenomenon is not limited to humans either – a group of autonomous agents for example may also need to decide how to use up their resources in order to complete a set of tasks. All these situations can be modeled as elections in the form of datasets containing votes and these datasets then can be used for various purposes. Frequently, we also want to generate synthetic datasets from some statistical distributions – the urn model, the Mallows model, or the impartial culture model, just to name a few very common ones.

When working with these datasets, it is important to think about how the space of all possible elections looks like. For example, we may test an approximation algorithm and report its accuracy and computation time. However, it is possible we inadvertently tested it only on a small subsection of the election space and our algorithm achieves significantly worse results in real-life use as a result. Another similar example would be testing data for a machine learning model. If the training data does not contain any elections from a particular subspace, then the model will poorly classify such elections in the test data. To counter this, we first need a way of conceptualizing the election space, which is a complex task. Distance metrics between election datasets and their projection onto a plane called “maps of elections” by Szufa et al. [1] aim to achieve precisely that.

The goals of this thesis are as follows:
1) Describe different voting systems, their strengths, and their weaknesses.
2) Familiarize yourself with the topic of maps of elections and describe the current state of the art.
3) Design and implement a distance metric between elections which can deal with elections that have a differing number of candidates. Test this metric on synthetic datasets and optionally on real-life datasets (for example from preflib.org). Discuss maps generated using this metric.
4) Optionally, explore the possibilities of a distance metric based on election features and discuss the results.

Bibliography:
[1] SZUFA, Stanisław; FALISZEWSKI, Piotr; SKOWRON, Piotr; SLINKO, Arkadii; TALMON, Nimrod. Drawing a Map of Elections in the Space of Statistical Cultures. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2020, pp. 1341–1349. AAMAS'20. ISBN 9781450375184.

Bachelor's thesis

MAPS OF ELECTIONS

Jitka Mertlová

Faculty of Information Technology
Department of Applied Mathematics
Supervisor: doc. RNDr. Dušan Knop, Ph.D.
January 11, 2024

Citation of this thesis: Mertlová Jitka. *Maps of Elections*. Bachelor's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2024.

Contents

List of Figures

List of Tables

List of code listings

There are many people who accompanied me on the journey of this thesis and on the journey of my studies and to whom I would like to thank from the bottom of my heart. My most sincere thanks and appreciation goes to my thesis supervisor Dušan Knop and to my IJCAI co-authors Piotr Faliszewski, Tomasz Wąs, Stanisław Szufa and Pierre Nunn. No student could ask for better supervisor(s) or better colleagues. It was (and still is) the greatest privilege to work with you.

I would like to say thanks to all members of Dvanáctka*—our long distance chats, study sessions and memes got me through the covid semesters and beyond. I would like to say thanks to Soptík, David, Honza, Tonda, Jirka, and Franta, for teaching me how to code and hunt bugs. I would like to say thanks to everyone in room A-1246. For everything.*

My deepest thank you goes to my beloved Martin, my confidant, my partner and my biggest supporter, who encouraged me to keep working, cheered me up when I felt sad, comforted me when I felt stressed, and above all, who always believed in me. I could not have done this without you.

Z celého srdce bych chtěla poděkovat mamince, tatínkovi, Honzíkovi, Marťovi, babičce Marcele, tetě Jítě a Artíkovi. A také těm, kteří už tu se mnou být nemohou—babičce Jarušce a dědovi Pepimu. Bez vaší lásky a podpory by tahle práce nikdy nevznikla.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis. I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as a school work under the provisions of Article 60 (1) of the Act.

In Prague on January 11, 2024 ....

Abstract

The *maps of elections* framework is a methodology for visualizing and analyzing election datasets. So far, this framework was restricted to collections of election datasets that have equal numbers of candidates. In this thesis, we extend this framework to be able to deal with datasets containing different numbers of candidates. We do this in two ways: we extend an already existing *positionwise* distance, and we also develop the idea and implementation of a feature distance. We test the new distances on synthetic data generated using the `Mapel` package.

Keywords computational social choice, voting systems, maps of elections, generative models, distance metrics, elections of different sizes

Abstrakt

Maps of elections je framework, který umožňuje vizualizovat a analyzovat sady volebních datasetů. Dosud bylo možné pracovat pouze s datasety, ve kterých měly všechny volby stejný počet kandidátů. V této práci framework rozšiřujeme o možnost současného zpracování datasetů s různými počty kandidátů. Toho dosahujeme jednak rozšířením již existující *positionwise* metriky, a dále zavedením a implementací tzv. *feature* metriky. Tyto metriky testujeme pomocí syntetických dat generovaných balíčkem `Mapel`.

Klíčová slova výpočetní sociální volba, volební systémy, mapy voleb, generativní modely, metriky vzdálenosti, volby rozdílných velikostí

Abbreviations

AI	Artificial Intelligence
EMD	Earth Mover's Distance
FR	Fruchterman-Reingold Embedding
IC	Impartial Culture
IAC	Impartial Anonymous Culture
IIA	Independence of Irrelevant Alternatives
KK	Kamada-Kawai Embedding
MDS	Multidimensional Scaling Embedding
MOV	Margin of Victory
PCC	Pearson Correlation Coefficient
SAT	Boolean Satisfiability Problem
SC	Single Crossing
SP	Single Peaked
SPOC	Single Peaked On a Circle
STV	Single Transferable Vote
WD	Wasserstein Distance

Chapter 1

Introduction

…in which we learn what this thesis is (and is definitely not) about.

All voting is a sort of gaming, like checkers or backgammon, with a slight moral tinge to it, a playing with right and wrong, with moral questions; and betting naturally accompanies it. The character of the voters is not staked. I cast my vote, perchance, as I think right; but I am not vitally concerned that right should prevail. I am willing to leave it to the majority.
— Henry David Thoreau [1]

When people hear “voting” or “election”, the sort of topic referred to by the opening quote usually pops into their minds. The general election, politicians and their campaigns, ballots and voting polls. Normative judgements about the character of each candidate and their promises, critical choices behind the voting screen …and then the usually straightforward computation of the winner or allocation of seats. However, that is by no means all there is to it! If it were so, then voting rules would be a dull topic for a computer science thesis, would they not? Fortunately for us, the world of elections is much richer (and therefore much more intriguing) to study. To illustrate this claim, let us imagine a few distinct real-life scenarios that can be modelled as an election:

- A few friends wish to find a consensus on which restaurant to eat lunch in.
- A championship in the sport of your choosing needs a system of matches and scoring that would allow for determination of a winner, while also giving its participants a feel of fair play.
- A school canteen manager has a list of meals that can be prepared from the ingredients in stock and a weekly schedule that offers five different meals per day. He needs to figure out which dishes to serve on which days of the week so that as many children as possible are willing to eat at least one of the offered meals each day. To make matters worse, since children can be famously bad at making decisions, it would be optimal if each child fancied only one or two dishes on the daily menu rather than four or (God forbid!) all five.
- A mine security engineer is getting different measurements from carbon monoxide sensors and needs to decide which sensors to trust.
- A neural network used in machine learning algorithms needs to decide between contradictory predictions of neurons in the outer layer while completing a multi-class classification task.

Rothe [2] describes computational social choice as a relatively recent branch of classical social choice theory which concerns itself (as the name suggests) mainly with its computational aspects.

Since our technological capabilities have grown significantly over the past few decades, the research interest in computational experiments grew with it. When running such experiments, we usually work with a collection of datasets. Each dataset represents one particular election instance—it contains all the votes cast in that election.

In order to then interpret the experimental results, it is important to think about how the space of all possible elections looks like. For example, we may test an approximation algorithm and report its accuracy and computation time. However, it is possible we inadvertently tested it only on a small subsection of the election space and our algorithm achieves significantly worse results in real-life use as a result. Another similar example would be testing data for a machine learning model. If the training data does not contain any elections from a particular subspace, then the model will poorly classify such elections in the test data. To counter this, we first need a way of conceptualizing the election space, which is a complex task. Distance metrics between election datasets and their projection onto a plane called *maps of elections* by Szufa et al. [3] and Boehmer et al. [4] aim to achieve precisely that.

The thesis has the following aims:

- Describe different voting systems, their strengths, and their weaknesses.
- Familiarize yourself with the topic of maps of elections and describe the current state of the art.
- Design and implement a distance metric between elections which can deal with elections that have a differing number of candidates. Test this metric on synthetic datasets and optionally on real-life datasets (for example from preflib.org [5]). Discuss maps generated using this metric.
- Optionally, explore the possibilities of a distance metric based on election features and discuss the results.

The thesis is structured as follows. The first chapter is dedicated to some of the basic building blocks of computational social choice—elections, voting and voting systems. The second chapter describes all the components that are required to make maps of elections—statistical cultures for synthetic data generation, distance metrics and embedding algorithms. Finally, the third chapter discusses the various distance metrics we propose—their definitions, implementations and experimental results.

1.1 Mapel Package

`Mapel`[1] is an ecosystem of Python packages whose development was started by Szufa et al. [3]. It focuses on topics of computational social choice, mainly on maps of elections [6], *stable marriages* [7] and *stable roommates* [8] problems. It has been in continuous development since 2020 and it is used mainly to visualize and analyze election datasets. I have worked with its subsection called `mapel-elections` [6] and a majority of the code created for this thesis was made with the intention to be integrated into `Mapel`. As of the submission of this thesis, I have contributed to the package with 3 successful pull requests—a table of my commits can be found in Appendix. The rest of my code will be gradually integrated into `Mapel` in the near future.

1.2 Acknowledgements

This thesis is part of a larger project called *Distances Between Top-Truncated Elections of Different Sizes*. A paper of this name is scheduled to be submitted to the IJCAI 2024 conference

[1] `https://github.com/szufix/mapel`

in mid January 2024. My co-authors are Piotr Faliszewski, Tomasz Wąs, Stanisław Szufa and Pierre Nunn. In this thesis, I present my own work done for this project. Since (unlike IJCAI conference) there is no strictly enforced page limit, I take the opportunity to discuss all my material at length, including the "blind alleys", less interesting results, and my implementation in code. To make my experiments runnable, I have included the full `Mapel` package including my code and the code of my co-authors which I build upon. A description of which parts of code were done by me and how to run the experiments can be found in Appendix.

In researching this topic, I was supported by FIT CTU's summer program for undergraduate researchers called *Výlet 2023*.

1.3 Preliminaries

Let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}_+$ be the set of non-negative real numbers, and $\mathbb{N}$ be the set of natural numbers[2]. For a positive integer m, we write $[m]$ to denote the set of integers from 1 to m. Given two sets A and B such that $|A| = |B|$, we denote the set of all bijections from A to B as $\Phi(A, B)$.

An *undirected graph* is a pair $G_u = (V, E)$, where V is a set of vertices and E is a collection of edges between these vertices, such that for every distinct pair of vertices u and v, the edge between them is denoted $\{u, v\}$. A *directed graph* is a pair $G_d = (V, E)$, where V is a set of vertices and E is a collection of edges between these vertices, such that for every distinct pair of vertices u and v, the edge between them is denoted either (u, v) or (v, u) depending on its orientation[3]. If every edge has a real number r corresponding to it as its weight, the graph is called *edge weighted*.

A *partition* of a set V is a collection of sets $V_1, \ldots, V_n$ such that $V_1 \cup V_2 \cup \ldots \cup V_n = V$ and for all $i, j \in [n], i \neq j$ it holds that $V_i \cap V_j = \emptyset$.

Let $x = (x_1, \ldots, x_t)$ and $y = (y_1, \ldots, y_t)$ be two vectors from $\mathbb{R}_+^t$ whose entries sum up to 1. The *Earth mover's distance*[4] between x and y, denoted $\mathrm{EMD}(x, y)$, is defined as the lowest total cost of operations that transform vector x into vector y. Each operation is of the form "subtract δ from position i and add δ to position j" and costs $\delta \cdot |i - j|$. Such an operation is legal if the current value at position i is at least δ [10].

We adopt the following distance and similarity definitions from the work of Deza et al. [11]. We write $\cos(x, y)$ to denote the *cosine similarity* between vectors x and y. Given that $x \cdot y$ denotes the dot product between these vectors and $\|x\|$ denotes a norm of vector x, we define the cosine similarity as such:

$$\cos(x, y) = \frac{x \cdot y}{\|x\|\|y\|}$$

We write $\ell_1(x, y)$ to denote the ℓ_1 *distance* (also called the Manhattan distance) between vectors x and y. It is defined as follows:

$$\ell_1(x, y) = \sum_{i=1}^{t} |x_i - y_i|$$

We write $\ell_2(x, y)$ to denote the ℓ_2 *distance* (also called the Euclidean distance) between vectors x and y. We define it as:

$$\ell_2(x, y) = \sum_{i=1}^{t} (x_i - y_i)^2$$

Let $R \subseteq C^2$ be a binary relation over a set C.

[2] We do not consider 0 to be a natural number.

[3] Given the edge (u, v), v is called the *head* vertex and u is called the *tail* vertex of this edge.

[4] The earth mover's distance was defined by Rubner et al. [9] as a distance between two distributions. It has wide use in pattern recognition, image processing, and many more.

- R is *connected* if it holds that $\forall a, b \in C, a \neq b \colon aRb \vee bRa$.
- R is *transitive* if it holds that $\forall a, b, c \in C \colon aRb \wedge bRc \Rightarrow aRc$.
- R is *asymmetric* if it holds that $\forall a, b \in C \colon aRb \Rightarrow \neg(bRa)$.
- R is *antisymmetric* if it holds that $\forall a, b \in C \colon (aRb \wedge bRa) \Rightarrow a = b$.
- R is *reflexive* if it holds that $\forall a \in C \colon aRa$.
- R is *irreflexive* if it holds that $\forall a \in C \colon \neg(aRa)$.

Binary relation R over a set C is called a *partial order* if it is reflexive, antisymmetric and transitive. The pair (C, R) is then called a partially ordered set. Two elements $a, b \in C$ are *comparable* if it holds either aRb or bRa. We say that R is a *total order* over C if every pair of elements form C is comparable.

Given two vectors $x = (x_1, \ldots, x_t)$ and $y = (y_1, \ldots, y_t)$, we denote the arithmetic average of each one's elements as $\bar{x}$ and $\bar{y}$, respectively. Their Pearson Correlation Coefficient (PCC) is defined as follows [3]:

$$\mathrm{PCC}(x, y) = \frac{\sum_{i=1}^{t}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{t}(x_i - \bar{x})^2 \sum_{i=1}^{t}(y_i - \bar{y})}}$$

Chapter 2

Elections and Voting Rules

...in which we describe the building blocks that elections and voting rules are made of.

This chapter is structured as follows. First, we provide basic terminology used in the field of computational social choice. Second, an overview of the most interesting voting rules, their classification, and their properties are presented. Finally, the use of voting rules in AI applications and the ramifications they are guaranteed to have are discussed. The structure of this chapter and the definitions provided are based on Rothe [2]. The notation is adopted from Szufa et al. [3] and Boehmer et al. [4].

On an intuitive level, an election may be conceptualized as a set of decision-making agents, *the voters*, expressing their preferences over a set of alternatives, *the candidates*. Formally [3]:

▶ **Definition 2.1** (Election)**.** *An* election *E is an ordered pair (C, V), where $C = \{c_1, \dots, c_m\}$ is the set of* candidates *and $V = \{v_1, \dots, v_n\}$ is the set of* voters*. Let ζ be a bijection between voter names and votes, which means that each voter casts exactly one vote. Both the voter and their vote are denoted by v_i, with their distinction clarified when necessary*[1]*.*

This thesis always considers elections whose candidates set and voters set are not empty. It is worth noting that the literature sometimes uses the term "alternative" as a synonym to "candidate", the term "ballot" as a synonym to "vote" and the term "election profile" or simply "profile" as a synonym to the "set of votes" (with respect to a given set of candidates). What is a vote may change depending on the voting system used. This thesis deals with voting rules that require one of the following types of votes:

- Each voter submits a full ranking of all the candidates from most to least desirable one, called an *ordinal preference.*[2] An election consisting of these votes is called an *ordinal election.*

  ▶ **Definition 2.2** (Preference Relation $\prec$)**.** *The* preference relation *$\prec$ is a relation over candidates that is connected, transitive, and asymmetric.*[3] *Let a and b be candidates. The term $a \prec b$ denotes that candidate a is preferred to candidate b and $v\colon a \prec b$ assigns this preference to voter v.*

  Additionally, the term $a \prec all$ denotes that candidate a is preferred to all other candidates; again, in the case of $v\colon a \prec all$, this preference is assigned to voter v. Analogously, $all \prec a$

[1] In this thesis, we usually use them interchangeably.

[2] In the literature, the term *preference profile* or *preference order* is sometimes used instead of "ordinal preference".

[3] Rothe [2] ascribes these properties to the symbol for a strict linear order (>), as he uses it throughout his book. Since we use the preference relation $\prec$ instead, we took the liberty to extend Rothe's definition to it.

indicates that all other candidates are preferred to candidate a and in the case of $v\colon all \prec a$, this preference is assigned to voter v.The set of all total orders over C is denoted by $\mathcal{L}(C)$.

- Each vote consists of two sets of candidates—the approved and the disapproved ones. There is no further preference ordering within these two sets. Voters are not allowed to express indifference about any candidate.
- A combined approach, each vote consists of both a preference order and sets of approved and disapproved candidates. Depending on the rule, voters may be required to rank all candidates or only approved ones in their ordinal preference.

Note that the votes need not be defined in a way that disallows the voter to express indifference about any of the candidates or to rank two candidates the same.[4] In fact, there is a considerable body of literature on the topic of incomplete preferences and the properties and challenges such elections have when we apply voting rules on them (see, e.g., [12] and references therein). However, this caveat is beyond the scope of this thesis.

Rothe [2] proposes to formalize voting rules and social welfare as functions that take an election as its argument:

▶ **Definition 2.3** (Voting Rule)**.** *A* voting rule *is a function that maps the election to a set of its winners:*

$$f\colon \{(C,V) \mid (C,V) \textit{is an election}\} \to 2^C.$$

The function's domain is the space of all possible elections, its range is the powerset of the set of candidates. In principle, it is possible that $f(C,V)=\emptyset$.

▶ **Definition 2.4** (Social Welfare Function)**.** *A* social welfare function *maps the election to a complete ranking of all candidates:*

$$f\colon \{(C,V) \mid (C,V) \textit{is an election}\} \to \mathcal{L}(C).$$

The term $f(C,V)\colon a \prec b$ signifies that given the election (C,V) and candidates a,b, the social choice function f produces an ordering that prefers candidate a to candidate b.

Instead of a "voting rule", the term "voting system" is often used as a synonym. Formally, a voting rule only provides us with a set of winners without any further ordering of the losing candidates, whereas a social welfare function outputs a preference list over all of the candidates in C. However, some voting rules already return a complete linear ordering of the candidates along with the winners (e.g., Kemeny), or they can be easily extended to do so, usually based on the score they allocate to each candidate (e.g., all scoring protocols). For ease of notation, voting rules are often treated similarly to the social welfare function in this thesis. When important, the distinction between them is always clarified.

2.1 Basic Voting Rules

This section describes some basic voting rules. A more thorough overview of voting rules is provided in Section 2.3. We start with probably the best known voting rule among the public, the *majority voting rule*. Formally:

▶ **Definition 2.5** (Majority Voting Rule)**.** *The* majority winner *is the candidate who occupies the top position*[5] *in more than half the votes. In the case of an even number of voters, a weaker*

[4] In case of voters being allowed to be indifferent about candidates, for example, the preference relation $\prec$ cannot be required to be asymetric, because such a requirement also implies irreflexivity. Clearly, irreflexivity cannot be satisfied when voters may be indifferent about some of the candidates [2].

[5] *Top position* means that the candidate in question is preferred to all other candidates, i.e., $a \prec all$ for $a \in C$. Analogously, the bottom position means that all other candidates are preferred to the candidate in question.

rule called the simple majority may be used. A candidate is a simple majority winner *if it is in the top position in at least half of the votes.*

Clearly, since the total pool of top positions is limited to $|V|$, the majority winner is always unique if it exists. If it does exist, it is also a simple majority winner (and vice versa if there is an odd number of voters). If there is an even number of voters, a candidate is a simple majority winner, but not a majority winner, if and only if it is ranked in the top position by exactly half of the voters. There may be none, one, or two simple majority winners.[6]

A different approach to selecting the winner of an election are the *pairwise comparisons*, which the *Condorcet voting rule* uses. In essence, each candidate enters a head-to-head contest with all other candidates. Formally:

► **Definition 2.6** (Condorcet Voting Rule)**.** *The* Condorcet winner *is the candidate that is preferred to every other candidate in a pairwise comparison by more than half of the voters. A* weak Condorcet winner *is a candidate that is preferred to every other candidate by at least half of the voters.*

Naturally, every Condorcet winner is also a weak Condorcet winner but not vice versa. Since the Condorcet winner needs to beat *every* other candidate in a pairwise comparison, it is always unique if it exists. Like the majority winner, often the Condorcet winner in fact does not exist. This situation is then called the *Condorcet paradox*. In every election, there is none or one Condorcet winner and none, one or more weak Condorcet winners (see Table 2.1). Analogously, the candidate who is defeated by every other alternative in their head-to-head contest with a majority is called the *Condorcet loser*.

Table 2.1 The profile on the left produces one weak Condorcet winner (candidate a). The profile on the right produces three weak Condorcet winners (candidates b, c and d).

Number of Votes	Preference Order
3	$a \prec c \prec d \prec b$
2	$c \prec b \prec a \prec d$
1	$d \prec b \prec a \prec c$

Number of Votes	Preference Order
3	$b \prec c \prec d \prec a$
3	$c \prec b \prec d \prec a$
3	$d \prec b \prec c \prec a$
3	$d \prec c \prec b \prec a$

A useful election representation to determine the Condorcet winner is the *majority graph*. Formally:

► **Definition 2.7** (Majority Graph)**.** *Let $E = (C, V)$ be an election. Then the* majority graph G_E *is an edge weighted directed graph, such that its set of vertices is the set of candidates C and for every two candidates a, b, (a, b) is an edge in this graph if and only if candidate a is preferred over candidate b by at least half of the voters. The weight of this edge is equal to the difference between the number of voters who voted $a \prec b$ and the number of voters who voted $b \prec a$.*

To determine the Condorcet winner, one may run the Kosaraju's algorithm[7] on the majority graph and then find the sink[8] component. If there is only one vertex present in this component, then the corresponding vertex is the Condorcet winner. If there are multiple vertices in the sink component, then this component is called a *Condorcet top cycle* and it means that there is no Condorcet winner at all.

In a sense, majority and Condorcet voting represent a very different view of making a group decision; thus both giving rise to distinct families of voting rules. In majority voting, the position

[6] The only case when two simple majority winners emerge is when one candidate is ranked first in exactly one half of the votes and another candidate is ranked first in the other half.

[7] The Kosaraju's algorithm, invented in 1978 by Sambasiva Rao Kosaraju, finds the strongly connected components in an oriented graph [13].

[8] The sink component is a strongly connected component in an oriented graph that has no edges leaving it [13].

of the candidates in the votes in their entirety matters the most. This gives rise to the family of scoring protocols discussed in Subsection 2.3.1.

In contrast, the family of pairwise comparisons based on Condorcet voting examines candidates in head-to-head contests with each other. Then it aggregates the results of these contests to determine the winner, again discussed in detail in Subsection 2.3.2.

Although both these rules are fundamental enough to examine whether other voting rules comply with the winners these two would have chosen, perhaps their biggest flaw is that in practice they often do not select *any* winner at all. One would argue that especially in such uncertain cases we would profit the most from having a robust voting rule to find a societal consensus. Thus, both majority voting and Condorcet voting cannot be considered the ultimate voting rules, but merely a stepping stone.

The last rule to be discussed in this section is the *Kemeny voting rule*:

▶ **Definition 2.8** (Kemeny Voting Rule). *Let $E = (C, V)$ be an election, $u, v \in V$ and $a, b \in C$. The distance between candidates a and b with respect to u and v is denoted $d_{u,v}(a, b)$:*

- $d_{u,v}(a, b) = 0 \iff$ *both u and v agree either on $a \prec b$ or $b \prec a$.*
- $d_{u,v}(a, b) = 2 \iff (u : a \prec b \land v : b \prec a)$ *or vice versa.*
- $d_{u,v}(a, b) = 1 \iff (u : a \prec b \lor u : b \prec a)$ *and v is ambivalent about the two candidates, or vice versa.*[9]

The distance between $u, v \in V$ is denoted as follows:

$$dist(u, v) = \sum_{\{a,b\} \in \binom{C}{2}} d_{u,v}(a, b)$$

Then the Kemeny score of a preference order $r \in \mathcal{L}(C)$ can be computed:

$$KemenyScore_{(C,V)}(r) = \sum_{v \in V} dist(r, v)$$

A preference order with the smallest Kemeny score is selected as the winning ordering and its top candidate is a Kemeny winner.

Clearly, the Kemeny voting rule belongs to the family of pairwise comparisons. It is also a social welfare function rather than a simple voting rule, since it outputs a complete ranking on all candidates.

2.2 Properties of Voting Rules

In order to give evidence-based judgments about quality of voting rules, we first define the specific properties that the examined voting rules may or may not have.[10]

[9] This case cannot happen with votes being strict preference orders and it is mentioned only for the sake of completeness.

[10] Such classification of voting rules needs to be done with a bit of a disclaimer, though. As Blais [14] correctly pointed out, when evaluating the quality of a voting rule, there is always some form of normative judgement on what the desiderata should be. This effect is negligible for the purposes of this thesis, but it is worth to bear in mind.

2.2.1 Anonymity and Nondictatorship

A voting rule is considered *anonymous* when all voters are treated equally, i.e., the rule only considers the set of votes themselves, not the names of the voters. If we fix the set of votes and make every voter cast someone else's vote by changing the naming bijection δ, the voting rule selects the same winners as before. This property is closely linked with another desirable feature of voting rules—*nondictatorship*, i.e., the absence of a dictator. To formally describe this concept, we first turn to Arrow's [15] definition of a *dictatorial social welfare function*:

▶ **Definition 2.9** (Dictatorial Social Welfare Function). *A social welfare function f is said to be* dictatorial *if there exists a voter $v \in V$ such that $\forall a, b \in C$, it holds that if $v\colon a \prec b$, then $f\colon a \prec b$, regardless of the other voters. Voter v is called the* dictator*.*

Rothe [2] transferred this notion to voting rules:

▶ **Definition 2.10** (Dictatorial Voting Rule). *A voting rule f is* dictatorial *if there is a voter v such that for every election (C, V) where $v \in V$, the preference order of v aligns with $f(C, V)$, without regard for preferences of other voters. The voter v is then called the* dictator*.*

Consequently, a voting rule is nondictatorial if none of the voters is the dictator. Majority, Condorcet, and Kemeny voting rules are clearly all nondictatorial.

2.2.2 Neutrality and Citizens' Sovereignty

Neutrality is in many ways a property analogous to anonymity, with the difference that it requires equal treatment of all candidates, not voters. If we fix the set of votes in an election and permutate the candidate names, winners of the new election will be the candidates who got assigned names of the previous election winners. Again, the only thing taken into consideration by the voting rule is the set of votes, not which specific candidate bears which specific name. A trivial example of a nonneutral voting rule would be the following $f\colon \{(C, V) \mid a \in C\} \to a$. The *citizens' sovereignty* criterion requires that for every candidate, there exists a set of votes such that the election rule selects that candidate as a winner.

▶ **Theorem 2.11.** *Every neutral voting rule that does not always return an empty set as the set of winners also fulfills the citizens' sovereignty criterion.*

Proof. Since there exists a set of votes that elects a winner and candidate names are irrelevant because of neutrality, we simply need to permutate the candidate names to obtain a winning profile for any other candidate. ◀

Majority, Condorcet, and Kemeny voting rules are all neutral and citizen sovereign.

2.2.3 Condorcet and Majority Criteria

A (simple) majority-consistent voting rule always selects the (simple) majority winner whenever it exists, i.e., it respects the (simple) majority winner. Similarly, a (weak) Condorcet-consistent voting rule always selects the (weak) Condorcet winner whenever there is one, i.e., it respects the (weak) Condorcet winner.

The Condorcet voting rule respects the Condorcet winner, but not the majority winner. Conversely, the majority voting rule respects the majority winner, but not the Condorcet winner. The Kemeny voting rule is Condorcet consistent, but not majority consistent.

2.2.4 Consistency and Pareto Consistency

Let $E = (C, V)$ be an election and f a voting rule. The rule f is considered *Pareto-consistent* if $\forall a, b \in C, a \prec b$ in all votes, then also $f(E)\colon a \prec b$.

Let $V_1, V_2, \ldots, V_p$ be a partition of V. Let $E_1, \ldots, E_p$ be elections such that $E_1 = (C, V_1)$, $E_2 = (C, V_2), \ldots, E_p = (C, V_p)$. The rule f is considered *consistent* if, for any partition choice, it holds:

$$\forall a \in C\colon f(E_1) = a \wedge f(E_2) = a \wedge \ldots \wedge f(E_p) = a \implies f(E) = a$$

All voting rules discussed in this thesis are Pareto consistent but not all are consistent. Majority voting rule is consistent, while Condorcet voting rule is not.

When it comes to Kemeny voting rule, the situation gets a bit tricky because Kemeny behaves like a social welfare function. It is the only voting rule that is neutral, consistent, and compliant with the Condorcet criterion at the same time, which was proven by Young and Levenbrick [16]. This, however, only holds for the definition of consistency in which we require that the Kemeny rule produces the same preference order on all partitions, not only the same winning candidate. If the winning preference order in any of the partitions differs from the rest and only the winning candidate is the same, there is no guarantee that this candidate will win the entire election. This situation is shown in Table 2.2. This election profile was taken from the Kemeny Wikipedia page [17].

Table 2.2 The profile P_1 (on the left) and the profile P_2 (on the right) form a partition of P, i.e., $P_1 \cap P_2 = \emptyset$ and $P_1 \cup P_2 = P$. P_1 produces a as a Kemeny winner with the winning preference order being $a \prec b \prec c$. P_2 produces also a as a Kemeny winner, but the winning preference order is $a \prec c \prec b$. Even though both P_1 and P_2 select the same winning candidate, their output preference order differs, therefore no claims can be made about whether $Kemeny(P) = a$ or not based on consistency. Indeed, P produces b as a Kemeny winner, with the output preference order being $b \prec a \prec c$, which differs from the results of the partition.

Number of Votes	Preference Order
7	$a \prec b \prec c$
6	$b \prec c \prec a$
3	$c \prec a \prec b$

Number of Votes	Preference Order
8	$a \prec c \prec b$
7	$b \prec a \prec c$
7	$c \prec b \prec a$

2.2.5 Monotonicity and Homogentity

Let $E = (C, V)$ be an election, $v \in V$ be a voter, and f be a voting rule. Let $E' = (C, V')$ be an election such that $V' = V \setminus \{v\} \cup \{u\}$, where the voter u is created from voter v by copying its preferences and then moving some candidate $a \in C$ to a position closer to the top of the vote than it previously occupied[11]. Other candidates remain in the same positions relative to each other. The rule f is *monotonous* if it holds:

$$\forall a \in C\colon f(E) = a \implies f(E') = a.$$

A more restrictive requirement would be the *strong monotonicity*. It is defined analogously to monotonicity, but relative positions of all candidates apart from a do not need to remain the same in the new election. For each vote, it only needs to hold that the candidates who were ranked behind a in the original election are still ranked behind a in the new election. Every rule that is strongly monotonous is also monotonous.

Let $E'' = (C, V'')$ be an election such that V'' is created from V by multiplying the number of occurrences of each vote by the same factor k, i.e., $k|V| = |V''|$. A rule f is *homogeneous* if it

[11] For example, given $C = \{a, b, c, d\}$, vote $b \prec c \prec a \prec d$ could become $b \prec a \prec c \prec d$.

holds:

$$\forall a \in C \colon f(E) = a \implies f(E'') = \{a\}$$

2.2.6 Participation and Twin Paradox

Let $E = (C, V)$ and $E' = (C, V')$ be elections such that $V' = V \cup \{u\}$ and $u \notin V$. A voting rule f satisfies the *participation* criterion if it holds that $\forall a, b \in C$, if $f(E) \colon a \prec b$ and $v \colon a \prec b$, then $f(E') \colon a \prec b$. In other words, when a voter casts their vote, it will always either make the resulting preference order more aligned with the voter's preferences or it will not change the resulting preference order at all. However, it cannot make the resulting preference order less favorable for the voter. The situation when any voter achieves a better result for themselves by not casting their vote is known as the *no-show paradox* (this term was introduced by Brams and Fishburn [18]).

A voter w is called a twin of voter v if their votes are identical[12]. Let $E'' = (C, V'')$ be an election such that $V'' = V \cup \{w\}$ and $w \notin V$. Let voter w be a twin of voter $v \in V$. Voting rule f suffers from the twin paradox if, for any pair of candidates $a, b \in C$, it holds that $f(E) = a$ and $f(E'') = b$ and $w \colon a \prec b$.

If a voting rule suffers from the twin paradox, it cannot satisfy the participation criterion, since the twin voters would have been better off not casting their ballots. A voting rule that does not display the twin paradox is said to be *twins welcome* compliant.

Felsenthal and Nurmi [19] correctly point out that there are many definitions of properties similar to the participation criterion described here and the corresponding paradoxes; however, their thorough discussion is beyond the scope of this thesis.

2.2.7 Independence of Irrelevant Alternatives and Independence of Clones

Let there be an election $E = (C, V)$ and a voting rule f. Let $a, b \in C$ be a pair of candidates. Election $E' = (C', V')$ is created as such: $C' = C \cup \{c\}$, $c \notin C$. The votes from V are copied into V' and the candidate c is inserted into each vote in V' at a random position. This means that the relative position of all candidates apart from $\{c\}$ with respect to each other is the same in both elections. The *independence of irrelevant alternatives* (IIA) requires that if $f(E) \colon a \prec b$ then also $f(E') \colon a \prec b$.

No preference-based voting system presented in this thesis satisfies IIA[13]. When it comes to approval-based voting, while intuitively less likely to violate the IIA based on the notion that approval or disapproval of a candidate has little to do with other candidates present in the election, Ohtsubo and Watanabe [20] showed that this assumption deserves a considerable amount of scrutiny. They studied how the so-called *contrast effect* alters voter judgement of candidates based on their mental comparison with a new candidate, thus making approval voting systematically non compliant with IIA[14].

The *clone* of candidate a is any candidate c that is right next to a in every vote. It is important to note that since a and c are practically identical, a does not beat its clone in every vote—the

[12] In accordance with intuition, identical votes do not differ in preference ordering of any pair of candidates and their approval and disapproval sets are the same (if applicable).

[13] In fact, no preference based voting system with $|C| \geq 3$ can be simultaneously nondictatorial, Pareto consistent and IIA. The author of this impossibility theorem is Kenneth J. Arrow, and thus the theorem bears his name. [15]

[14] An example of the contrast effect could be when Annemarie is trying to decide what she wants for lunch. At first, she finds an apple and a banana in her backpack and decides that she will have the apple and save the banana for dinner. Then her friend Clara comes over and gives her a piece of banana bread. Annemarie decides to save the slice for later because she does not want to eat sweets for lunch; however, now that she is already thinking about bananas, she has started to prefer the banana over the apple. So in the end, she eats the banana and saves the apple for later. Therefore, the introduction of a new candidate (banana bread) made her change the judgement from *apple* $\prec$ *banana* to *banana* $\prec$ *apple* [20].

ordering $a \prec c$ is just as likely as the ordering $c \prec a$. The only thing that is guaranteed is that there is no other candidate between these two. Let there be an election $E = (C, V), a \in C$, a voting rule f, and $f(E) = \{a\}$. Rule f is independent of clones if $\forall b \in C, b \neq a, b$ cannot be made a winner by adding clones to E.

Independence of clones is tricky to assess for the majority voting rule. By adding a clone of the majority winner, it is possible to either make both the original winner and the clone weak majority winners, make the clone the majority winner or keep the original majority winner. No other candidate, however, can benefit from addition of a clone than the clone candidate itself. Neither Condorcet [21] nor Kemeny voting rules are independent of clones.

2.2.8 Resoluteness

Another highly desirable property is *resoluteness*. A voting rule is considered resolute if for every election, there exists always exactly one winner; this single winning candidate is then called the *unique winner* of the election. All voting rules discussed in this thesis are not resolute unless they implement some sort of tie-breaking within its structure[15]. To illustrate this, consider the trivial case of two candidates, and all voters' preferences split into two groups of equal size, with one preferring the first candidate and the other one the other candidate. There is no reasonable voting rule which is both neutral and nondictatorial that produces a unique winner of this election without tie-breaking of some sort. It may be part of the voting rule itself, therefore appearing to have no need for additional tie-breaking, but at that point, it is just a difference in nomenclature. Majority, Condorcet, and Kemeny voting rules are not resolute.

2.2.9 Strategy-proofness

The notion of a voting rule being resistant to manipulation is expressed by the property of *strategy-proofness*. If a voting rule is not strategy-proof, it is susceptible to *manipulation*. Casting an untruthful vote (i. e., a vote that does not align with inner preferences of the voter) or adding a new candidate for the purpose of changing the winners of the election are the two most significant ways to influence voting rules discussed in this thesis. If an election is strategy-proof, then it naturally follows that it also fulfills the participation criterion—if no strategic voting is possible, then every voter achieves the best result for themselves if they attend the election and cast their truthful vote. Unfortunately, all voting rules described in this thesis do not satisfy the strategy-proofness requirement[16]. However, it is useful to know how "sure" the winner's victory was, with the idea that more secure victory means less room to manipulate the election. To this end, the corresponding metric is defined as follows:

▶ **Definition 2.12** (Margin of Victory)**.** *With respect to a voting rule f and an election (C, V), the* margin of victory *(MOV) is the smallest number of votes, which, when changed*[17]*, would result in a different set of election winners.*

---

[15]If $|C| \geq 3$, there is no preference-based voting rule that simultaneously satisfies nondictatorship, citizens' sovereignty, strong monotonicity, and resoluteness. This finding is called the Muller and Satterthwaite impossibility theorem [2, 22].

[16]Moreover, if $|C| \geq 3$, there is no preference-based voting rule that simultaneously satisfies strategy-proofness, nondictatorship, resoluteness, and citizens' sovereignty. This fact is often called the Gibbard and Satterthwaite impossibility theorem [2, 23, 24].

[17]To change a vote means to generate any other preference order over the candidates and replace the original vote with it, e.g., $a \prec b \prec c \prec d$ can be changed to $a \prec d \prec c \prec b$ or $b \prec a \prec d \prec c$ or any other preference order over $\{a, b, c, d\}$.

2.3 Examples of Voting Rules

The following subsections provide a thorough overview of various voting rules, their formal definitions, and characteristic properties. For each voting rule, the corresponding function is denoted by its name, i.e., majority$(E) = \{a\}$ means that the majority voting rule selected candidate a as the winner of the election E.

2.3.1 Scoring Protocols

All voting rules in this subsection depend on the so-called *scoring vector*, denoted α, which determines how many points a candidate gets for its ranking in the vote. The score of each candidate is then the sum of points the candidate obtained over all the votes. The candidate who obtains the highest total score is then a winner of the election. The following voting rules differ only in which scoring vector they use:

▶ **Definition 2.13** (k-Approval)**.** *For a fixed $1 \leq k \leq m - 1$, the first k elements in vector α are equal to 1 and the remaining elements are equal to 0. Such α is then the* k-approval *scoring vector. The 1-approval is more commonly called the* plurality voting rule*, and (m-1)-approval is called the* veto*.*

▶ **Definition 2.14** (Borda Count)**.** *The vector α is defined as follows:* $\alpha = (m-1, m-2, \ldots, 1, 0)$.

The computational complexity of both k-approval and Borda count is in the P[18] class, since summing the partial scores over each vote and computing the result is easily done in polynomial time. Plurality voting complies with majority criterion, while Borda does not.

The term *Borda paradox* refers to a situation when a Condorcet loser is elected by some voting rule as the winner. It is a bit of misnomer, though, as it never occurs in the Borda voting rule. However, it occurs in plurality voting, as illustrated in Table 2.3.

Table 2.3 Voting profile in which candidate a is a plurality winner and a Condorcet loser.

Number of Votes	Preference Order
4	$a \prec b \prec c$
3	$b \prec c \prec a$
3	$c \prec b \prec a$

No scoring protocol can be independent of clones — whenever we introduce a clone of the election winner, it is going to place above the previous winner in some of the votes, essentially robbing the previous winner of some portion of points. Because of that, it can allow another candidate to take over. An example of such manipulation in plurality voting is shown in Table 2.4.

Table 2.4 The election profile on the left produces candidate a as a plurality winner. When a clone of a denoted c is added, the election profile on the right produces b as a plurality winner.

Number of Votes	Preference Order
4	$a \prec b$
3	$b \prec a$

Number of Votes	Preference Order
2	$a \prec c \prec b$
2	$c \prec a \prec b$
3	$b \prec a \prec c$

[18]In complexity theory, a decision problem is in P class if it is solvable in polynomial time. [13] An equivalent definition is that it can be solved using a deterministic Turing machine in polynomial time. In contrast, a decision problem is said to be NP-hard if it is solvable using nondeterministic Turing machine in polynomial time. A problem is NP-complete if it is NP hard and it itself is in the NP class. [2]

In contrast, all scoring protocols naturally obey monotonicity for the identical reason — by moving a candidate up in some of the votes, he can only gain points and his opponents can only lose points.

2.3.2 Pairwise Comparisons

The core idea of the voting systems discussed in this subsection is pairwise comparisons introduced by the Condorcet voting rule in Section 2.1. Unlike this rule however, the following rules always produce a set of winners, even in cases where no Condorcet winner exists.

▶ **Definition 2.15** (Copeland$^\alpha$ Voting Rule)**.** *Let $\alpha \in [0,1]$ be a parameter. For each pairwise comparison, the candidate that receives the majority of votes gains 1 point. In case of a tie, both candidates gain α points. The candidate who obtains the highest total score is then a winner of the election.*

Dodgson and Young voting rules are quite similar to the Kemeny voting rule (introduced in Section 2.1) in the way they approach the election. If there exists a Condorcet winner, it is elected. If not, then the candidate who is "closest" to becoming a Condorcet winner is elected instead, with the definition of a distance metric being different for each rule.

▶ **Definition 2.16** (Dodgson Voting Rule)**.** *The* Dodgson score *of a candidate is the minimum number of swaps between adjacent candidates in the votes so that the candidate becomes the Condorcet winner. A candidate who obtains the lowest total score is then a winner of the election.*

Given an election $E = (C, V)$ and a candidate c, it is $P_{||}^{NP}$-hard[19] problem to decide whether $\text{Dodgson}(E) = \{c\}$, which was proven by Hemaspaandra, Hemaspaandra, and Rothe. [2, 25] The Dodgson voting rule is neither monotonous, nor consistent, or homogeneous.

▶ **Definition 2.17** (Young Voting Rule)**.** *The* Young score *for each candidate is equal to the minimum number of votes that need to be deleted from V to allow that candidate to become the Condorcet winner. A candidate with the lowest Young score wins the election.*

Despite being monotonous, the Young voting rule is neither homogeneous nor consistent. It is also susceptible to the twin paradox, as shown in Table 2.5. The example profiles were created based on Felsenthal [26]. Given an election $E = (C, V)$ and a candidate c, it is $P_{||}^{NP}$-complete problem to decide whether $\text{Young}(E) = \{c\}$.

▶ **Definition 2.18** (Maximin Voting Rule)**.** *Let $E = (C, V)$ be an election and a and b be a pair of candidates. The number of voters that have the preference $a \prec b$ is denoted by $N(a, b)$. The* maximin score *of a is computed as follows:*

$$\text{MaximinScore}(a) = \min_{a \neq b} N(a, b)$$

The candidate with the highest maximin score wins the election.[20]

[19] A decision problem is $P_{||}^{NP}$-hard if it can be decided in polynomial time using a deterministic Turing machine that has access to an NP oracle and asks it the queries in parallel. Therefore, one query cannot depend on the oracle's answer to another query. An example of NP oracle implementation is a SAT solver that is given a set of formulas and decides whether they are satisfiable or not. A problem is $P_{||}^{NP}$-complete if it is $P_{||}^{NP}$-hard and it is in $P_{||}^{NP}$ class.

[20] The maximin voting rule is also frequently called the Simpson voting rule.

Table 2.5 The profile on the left produces candidate e as a Young winner. When 9 clones of the vote $e \prec d \prec a \prec b \prec c$ are added, the profile on the right produces candidate d as a Young winner, even though the added voters have the preference of $e \prec d$.

Number of Votes	Preference Order
11	$b \prec a \prec d \prec e \prec c$
10	$e \prec c \prec b \prec d \prec a$
10	$a \prec c \prec d \prec b \prec e$
2	$e \prec c \prec d \prec b \prec a$
2	$e \prec d \prec c \prec b \prec a$
2	$c \prec b \prec a \prec d \prec e$
1	$d \prec c \prec b \prec a \prec e$
1	$a \prec b \prec d \prec e \prec c$
1	$e \prec d \prec a \prec b \prec c$

Number of Votes	Preference Order
11	$b \prec a \prec d \prec e \prec c$
10	$e \prec c \prec b \prec d \prec a$
10	$a \prec c \prec d \prec b \prec e$
2	$e \prec c \prec d \prec b \prec a$
2	$e \prec d \prec c \prec b \prec a$
2	$c \prec b \prec a \prec d \prec e$
1	$d \prec c \prec b \prec a \prec e$
1	$a \prec b \prec d \prec e \prec c$
10	$e \prec d \prec a \prec b \prec c$

2.3.3 Approval Voting

Approval-based voting was described in great detail by Brams and Fishburn [18]. They assert that it increases the voter turnout and other empirically desirable traits.

▶ **Definition 2.19** (Approval Voting Rule)**.** *The approval score of each candidate is the number of voters who placed the said candidate in their set of approved candidates. The candidate with the highest score wins the election.*

2.3.4 Voting Systems with Multiple Stages

Some voting rules require multiple stages to select a winner. Due to their iterative nature, their definitions usually strongly resemble an algorithm.

▶ **Definition 2.20** (STV)**.** *In the first round, a check is made to see whether there exists a majority winner. If there is one, it is selected as the STV winner and the process stops. If not, then the candidate with the least number of top positions is removed from all votes. Then another round is held to see whether a majority winner exists. The rule continues to iterate until the set of winners ceases to be empty, at which point it stops and returns the winners.*

A rather interesting property of STV is that while in theory, of course, prone to strategic voting, Bartoldi and Orlin [27] described that it is *computationally* resistant to manipulation, regardless of whether the manipulator learns all the other votes. Thus, while STV is not monotone (see Table 2.6), Bartoldi and Orlin suggest that this flaw is more or less "covered up" by the inherent computational resistance to manipulation. This claim was experimentally tested by Gallagher [28], who concluded that of 1,326 examined cases of real-life data, violation of monotonicity occurred in approximately 1.5 % of them.

Another theoretical (and undesirable) property of STV that rarely affects the rule's behavior on real-life datasets is the no-show paradox, which was experimentally shown by Mohsin [29]. Mohsin showed this result also for Copeland and maximin voting rules.

▶ **Definition 2.21** (Bucklin Voting rule)**.** *The rule proceeds in stages with a scoring vector α_i in the i-th round. The scoring vector α_i has 1 in the first i positions and 0 in all the remaining positions.*

Table 2.6 The election profile on the left produces candidate a as an STV winner in the first round. When two $b \prec c \prec a$ votes are changed to $a \prec b \prec c$ by improving position of candidate a, the profile on the right produces c as an STV winner.

Number of Votes	Preference Order
15	$a \prec c \prec b$
3	$a \prec b \prec c$
16	$b \prec c \prec a$
15	$c \prec a \prec b$

Number of Votes	Preference Order
15	$a \prec c \prec b$
5	$a \prec b \prec c$
14	$b \prec c \prec a$
15	$c \prec a \prec b$

Chapter 3

Maps of elections

...in which we learn how to project elections onto a planar surface, what such projections mean and how they may be useful.

When working with election datasets, one might wish to classify them based on their inherent properties, which would then give them a way to describe the space of all possible elections. An algorithm that generates a uniform distribution of synthetic data covering the whole election space would be of great use for, e.g., benchmarking approximation algorithms or studying the generative models for synthetic data themselves, among many other uses. With this final goal in mind, a possible approach is to gather a large number of datasets, define a distance metric between two elections, compute this distance for each election pair, and then embed these distances into a 2D space, making a planar projection. The resulting plot is called a *map of elections* and was proposed by Szufa et al. [3] and Boehmer et al. [10]; it offers valuable information on how the election space may be conceptualized.

This chapter describes all the major components for making a map of elections, especially various kinds of distance metrics with their strengths and weaknesses and generative models for synthetic data. The notation and definitions used in this chapter were adopted from Szufa et al. [3] and Boehmer et al. [10].

3.1 Distance Metrics

A *distance metric* is defined as follows:

▶ **Definition 3.1** (Distance)**.** *Let X be a set. We say that a function $\mathrm{d}\colon X \times X \to \mathbb{R}$ is a distance over X if $\forall x, y, z \in X$ all of the following holds:*

non-negativity: $d(x, y) \geq 0$,

identity of indiscernibles: $d(x, y) = 0 \iff x = y$,

symmetry: $d(x, y) = d(y, x)$,

triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

If a metric does satisfy all the criteria apart from the identity one, it is called a *pseudodistance* instead. Faliszewski et al. [30] relax the identity condition so that only $d(x, x) = 0$ is required, and keep the distance nomenclature for the sake of simplicity. This thesis adopts a similar approach.

3.1.1 Distances Between Voters

In order to talk about distances between elections, we first need to define distances between voters [30]:

▶ **Definition 3.2** (Discrete Distance). *Let $E = (C, V)$ be an election. For voters $u, v \in V$, the* discrete distance $\mathrm{d_{disc}}(u, v)$ *is 0 if $u = v$ and 1 otherwise.*

▶ **Definition 3.3** (Swap Distance). *Let $E = (C, V)$ be an election. For voters $u, v \in V$, the* swap distance $\mathrm{d_{swap}}(u, v)$ *is the number of swaps of adjacent candidates that must be performed to transform u into v.*

For a candidate $c \in C$, his position in a vote $v \in V$ is denoted as $\mathrm{pos_v}(c)$. If $v\colon c \prec all$, then $\mathrm{pos_v}(c) = 1$ and if $v\colon all \prec c$, then $\mathrm{pos_v}(c) = |C|$.

▶ **Definition 3.4** (Spearman distance). *Let $E = (C, V)$ be an election. For voters $u, v \in V$, the* Spearman *distance* $\mathrm{d_{Spear}}(u, v)$ *is equal to* $\sum_{c \in C} |\,\mathrm{pos_v}(c) - \mathrm{pos_u}(c)|$.

3.1.2 Isomorphic Distances

Szufa et al. [3] describe the idea behind the isomorphic distance of two elections as finding of a mapping between candidates and voters of the elections such that the sum of distances across all individual matched pairs of voters is minimal. Either the swap distance or the Spearman distance can be used for computing the distance between votes, giving rise to the isomorphic distances ID-$\mathrm{d_{swap}}$ and ID-$\mathrm{d_{Spear}}$, formally defined as follows:

▶ **Definition 3.5** (Swap Isomorphic Distance and Spearman Isomorphic Distance). *Let there be two elections $E_1 = (C_1, V_1)$ and $E_2 = (C_2, V_2)$ such that $|C_1| = |C_2|$ and $|V_1| = |V_2|$. Let $\delta\colon C_1 \longrightarrow C_2$ and $\sigma\colon V_1 \longrightarrow V_2$ be two bijections on the candidates' and the voters' names, respectively.*[1] *For $D \in \{\mathrm{d_{swap}}, \mathrm{d_{Spear}}\}$ we define*[2]*:*

$$\mathrm{d_D^{\delta,\sigma}\text{-}ID(E_1, E_2)} = \sum_{v_1 \in V_1} d_D(\delta(v_1), \sigma(v_1)).$$

The isomorphic distance $\mathrm{d_D\text{-}ID(E_1, E_2)}$ *denotes the minimum* $\mathrm{d_D^{\delta,\sigma}\text{-}ID(E_1, E_2)}$ *taken over all δ and σ.*

While these distances are very appealing on an intuitive level, their computational complexity renders them unusable for our purposes. Moreover, they require both the number of voters and candidates to be equal across all elections. This poses a significant problem when applying these metrics to a collection of real-life datasets, in which this is rarely the case. The solution to this is twofold. In the case of the same number of candidates and differing numbers of votes, we will denote ℓ as the minimum number of votes across all the elections studied. We can then delete at random votes from elections for which $|V| > \ell$ until for each election, we have $|V| = \ell$. In the case of differing numbers of candidates, an analogous approach of deleting a candidate selected at random from all the votes until each election has the same number of candidates may be deployed. Although rather easy to execute, the deletion approach is highly undeterministic and causes a great deal of data distortion and misinterpretation, making it unusable when comparing multiple elections with largely differing number of candidates. The *positionwise* and *pairwise* distances introduced in the following subsection, however, offer an approach to compare datasets with differing number of votes.

[1] Szufa [3] stresses that the bijection is not over votes but voter names, meaning that voters with the same preference order in V_1 can be mapped to different voters in V_2.

[2] Note that we overload our notation in the following way: if $v = (c_1, c_2, \ldots, c_m)$, then we write $\delta(v)$ to denote $(\delta(c_1), \delta(c_2), \ldots, \delta(c_m))$.

3.1.3 Pairwise and Positionwise Distances

Given an election $E = (C, V)$, position $i \in [|C|]$, and candidates a and b, the fraction of voters in V who prefer a over b is denoted as $M_E(a, b)$ and the fraction of voters in V that rank a on the i-th position is denoted as $\psi_E(a, i)$. Then a *candidate distribution vector*, denoted $\psi_E(a)$, is defined as follows:

$$\psi_E(a) = (\psi_E(a, 1), \psi_E(a, 2), \ldots, \psi_E(a, n)).$$

Szufa et al. [3] define positionwise and pairwise distances as follows:

▶ **Definition 3.6** (Positionwise Distance)**.** *Let $E_1 = (C_1, V_1)$ and $E_2 = (C_2, V_2)$ be two elections such that $|C_1| = |C_2|$. Let $\delta\colon C_1 \to C_2$ be a bijection. We define*

$$\delta\text{-POS}(E_1, E_2) = \sum_{c \in C_1} \text{EMD}(\psi_{\text{E}_1}(c), \psi_{\text{E}_2}(\delta(c))).$$

The positionwise distance *between elections E_1 and E_2, denoted* $\text{POS}(E_1, E_2)$*, is the minimum of the* $\delta\text{-POS}(E_1, E_2)$ *values, taken over δ.*

▶ **Definition 3.7** (Pairwise Distance)**.** *Let $E_1 = (C_1, V_1)$ and $E_2 = (C_2, V_2)$ be two elections such that $|C_1| = |C_2|$. Let $\delta\colon C_1 \to C_2$ be a bijection. We define*

$$\delta\text{-PAIR}(E_1, E_2) = \sum_{(c,d) \in C_1 \times C_2} \left| M_{E_1}(c, d) - M_{E_2}(\delta(c), \delta(d)) \right|.$$

The pairwise distance *between elections E_1 and E_2, denoted* $\text{PAIR}(E_1, E_2)$*, is the minimum value of the* $\delta\text{-PAIR}(\text{E}_1, \text{E}_2)$ *values, taken over δ.*

Again, note that while the number of voters need not be the same across all elections for these distances, it is still required that the number of candidates remains the same across all elections. While the pairwise distance is still NP hard to compute, the positionwise distance, on the other hand, can be computed in polynomial time, which is why Szufa et al. [3] and Boehmer et al. [4] focus on it. It was also used by Faliszewski et al. [31], Boehmer et al. [32] and Boehmer et al. [33], among others.

The `Mapel` package computes the positionwise distance by reducing the election instances to their corresponding *frequency matrices*. Instead of trying every possible mapping δ to find the minimal one, it uses methods from `Scipy` Python package based on linear programming [34].

▶ **Definition 3.8** (Frequency Matrix)**.** *Let $E = (C, V)$ be an election, $C = \{c_1, \ldots, c_m\}$ be a set of candidates and $V = \{v_1, \ldots, v_n\}$. Given $i \in [m]$, let $\psi_E(c_i)$ be the candidate distribution vector for candidate c_i. The* frequency matrix *corresponding to this election, denoted F_E, is an $m \times m$ matrix such that it has vector $\psi_E(c_i)$ in its ith column:*

$$F_E = [\psi_E(c_1), \ldots, \psi_E(c_m)]$$

Each element of the frequency matrix has a nonnegative value and each of its column's elements sum up to 1.

Computation of the positionwise distance between two elections then boils down to finding the matching between columns of their frequency matrices such that the sum of Earth mover's distances between matched columns is minimal. [3, 4].

Since both positionwise and pairwise distances try all possible candidate matchings and pick the optimal one, they are both neutral and anonymous—this is an important (and favourable) feature of these distances, since they can be used to draw maps of synthetic datasets, where the candidate names carry no inherent meaning, alongside real-life datasets.

3.2 Generative Models

When generating synthetic data, we need to define a set of rules that the generative model will use. Such rules are called a *statistical culture* and usually take the form of an algorithm that prescribes how to generate vote v_{i+1} given already generated votes $v_0, \ldots, v_i$. In this thesis, we largely use the same statistical cultures as Szufa et al. [3] and Boehmer et al. [4] used to generate synthetic data.

3.2.1 Impartial Culture and Impartial Anonymous Culture

In the *Impartial Culture* (IC) model, all votes are sampled uniformly at random, which means that when we construct the election vote by vote, every possible vote has exactly the same probability of being added to the election. An extension of this model is the *Impartial Anonymous Culture* (IAC), in which every possible voting situation (i.e., the numbers of votes with a given preference profile) appears with the same probability. This means that when constructing an election under IAC, the votes that had already been added to the election in previous iterations are accounted for when computing the probabilities of the next added vote. Subsequently, IAC generates the election profiles uniformly at random.

3.2.2 Urn Model

In the *urn model*, the parameter α is introduced to describe the correlation between votes. In the beginning, the urn contains one copy of each possible vote, i.e., $m!$ votes. In each iteration, a random vote is taken from the urn, its copy is included in the election, whereupon it is returned together with $\alpha m!$ of its copies to the urn. If $\alpha = 0$, then the urn model behaves identically to IC. If $\alpha = 1/m!$, the urn model behaves identically to IAC. In general, the higher the α parameter, the more correlated the votes in an election are. If $\alpha = \infty$, then a unanimous election is produced. Szufa et al. [3] point out that out of all the statistical cultures studied in their paper, the urn model produces elections with largest distances between them, which results in this culture being the most spread out in their maps. We denote the urn model parameterized by α as $\mathrm{Urn}(\alpha)$.

3.2.3 Extreme Points

Boehmer et al. [10] introduced the concept of "extreme points" on the map of elections, called a *compass*. Given that m is the number of candidates, these extreme points are special elections with the following properties:

Identity compass point corresponds to election in which all voters agree on placement of all the candidates—all the votes are perfect copies of each other. The frequency matrix of such an election is called the *identity matrix*. It is denoted as $\mathrm{ID_m}$ and it is an $m \times m$ matrix with ones on the diagonal and zeros everywhere else.

Uniformity compass point corresponds to election in which each candidate takes each position equally often[3]. The *uniformity matrix* is denoted as $\mathrm{UN_m}$ and each of its elements is equal to $1/m$.

Antagonism compass point corresponds to election in which half of the voters have opposite preferences orders to the other half. The *antagonism matrix* is denoted as $\mathrm{AN_m}$. Given that $\mathrm{rID_m}$ denotes the identity matrix with reversed order of columns, the antagonism matrix is defined as follows:

$$\mathrm{AN_m} = \frac{1}{2}\,\mathrm{ID_m} + \frac{1}{2}\,\mathrm{rID_m}$$

[3] This sounds awfully lot like the IC, right?

Stratification compass point corresponds to election in which all voters agree on one half of the candidates being preferable to the other half, but on nothing else (i.e., there is no agreement on preferences within the preferable and nonpreferable group). Assume that m is an even number. Then the *stratification matrix* is defined as follows:

$$\mathrm{ST_m} = \begin{bmatrix} \mathrm{UN_{m/2}} & 0 \\ 0 & \mathrm{UN_{m/2}} \end{bmatrix}$$

Convex combinations of compass point matrices form *paths between extreme points*. Since there are four compass points, the corresponding paths are called *ANUN*, *ANID*, *STID*, *STUN*.

The `Mapel` package implements also an approximation of these elections, since it sometimes happens that there is no election which would correspond to the defined extreme matrix. The approximate compass elections are sampled from real elections which resemble the extreme points as closely as possible.

3.2.4 Mallows Models

Mallows model has two parameters—a number $\phi \in [0, 1]$ and a central vote v. The model then generates votes at random such that every vote u has the probability $\phi^{\mathrm{d_{swap}}(v,u)}$ to be generated. If $\phi = 0$, then the only vote with non-negative probability of being generated is one that has zero swap distance from the central vote[4]. The only such vote is v itself, which means the election contains only copies of vote v and is unanimous. If $\phi = 1$, then the probability of each vote is the same no matter the exponent[5], meaning the model behaves identically to IC. We denote the Mallows model parameterized by ϕ and a central vote v drawn uniformly at random as Mallows(ϕ).[6]

Boehmer et al. [4] introduced a normalised version of Mallows parameterized by rel-$\phi \in [0, 1]$ rather than ϕ and their work in this field was continued by Faliszewski et al. [35], who introduced parametrization by norm-ϕ instead of rel-ϕ. Merits of these changes include a more evenly distributed Mallows elections when drawn on a map. For further details on rel-ϕ and norm-ϕ definitions and implementations, see e.g., Boehmer et al. [36] and their references.

3.2.5 Single-Peaked Models

To discuss these models, we first need to define the *single-peakedness* (SP) property [3]:

▶ **Definition 3.9** (Single-Peaked Vote). *Let $C = \{c_1, \ldots, c_m\}$ be a set of candidates and let $c_1 \lhd c_2 \lhd \cdots \lhd c_m$ be a strict, total order over C, referred to as the* societal axis*. Let vote v be a preference order over C. We say that v is single peaked with respect to $\lhd$ if for each $\ell \in [m]$, the set of ℓ top ranked candidates according to v forms an interval within $\lhd$. We say that v is* single peaked on a circle *if for each $\ell \in [m]$, the set of ℓ top ranked candidates either forms an interval within $\lhd$ or a complement of an interval.*

▶ **Definition 3.10** (Single-Peaked Election). *Let $C = \{c_1, \ldots, c_m\}$ be a set of candidates and let $c_1 \lhd c_2 \lhd \cdots \lhd c_m$ be a societal axis. An election is single peaked if there exists a $\lhd$ such that each voter's preference order is single-peaked with respect to it. An election is single peaked on a circle if there exists a $\lhd$ such that each voter's preference order is single peaked on a circle with respect to it.*

[4] Indeed, $0^0 = 1$, but $0^k = 0$ for every k such that $k \neq 0$.
[5] Again, $1^k = 1$ for every k such that $k \geq 1$.
[6] A slight variation on this is the *Bi-Mallows* model parameterized in the same way as standard Mallows. The difference is that half of the votes is reversed after being drawn, i.e., vote $a \prec b \prec c$ becomes $c \prec b \prec a$.

The `Mapel` package implements two cultures that produce single peaked elections: a culture based on an approach of Walsh [37] and a culture based on an approach of Conitzer [38]. Surprisingly, Szufa et al. [3] discovered that these two models each produce an election cluster with large distance between the two. Because of that, we assume it is highly likely that these two models do not cover the whole space of SP elections.

3.2.6 Single-Crossing Model

Another interesting property that an election may have is being *single crossing* (SC).

▶ **Definition 3.11** (Single-Crossing Election). *Let $E = (C, V)$ be an election. E is called single crossing (SC) if it is possible to order the voters in such a way that for every pair of candidates $a, b \in C$, the set of voters with $a \prec b$ preference either forms a prefix or suffix of this order.*

▶ **Definition 3.12** (Single-Crossing Domain). *Let $\mathcal{D}$ be a set of preference orders. $\mathcal{D}$ is called a* single crossing domain *if every election containing only votes from this domain is single crossing.*

The `Mapel` package implements SC model by generating the whole SC domain and then draws the required number of votes from it uniformly at random. The process of generating the domain, however, is not uniform.

3.2.7 Euclidean Models

The Euclidean elections are defined as follows:

▶ **Definition 3.13** (t-Euclidean Election). *Let t be a positive integer. An election $E = (C, V)$ is t-*Euclidean *if it is possible to associate each candidate and each voter with their ideal point in the t-dimensional Euclidean space $\mathbb{R}^t$ in such a way that the following holds: For each voter v and a pair of candidates a, b, v: $a \prec b$ if and only if v's point is closer to the point of a than to the point of b.*

The implementation in `Mapel` package chooses the ideal points for candidates and voters and then derives voters preference orders to fit them. Given $t \in \mathbb{N}$, the ideal points are generated uniformly at random either from $[-1, 1]^t$ (the t-dimensional hypercube model, denoted tD-Hypercube) or from a hypersphere with center at $(1, \ldots, 1)$ and radius 1 (the t-dimensional hypersphere model, denoted tD-Hypersphere). The most notable tD-Hypercube models are *Interval* (1D-Hypercube), *Square* (2D-Hypercube) and *Cube* (3D-Hypercube). Most notable tD-Hypersphere models are *Circle* (2D-Hypersphere) and *Sphere* (2D-Hypersphere).

3.3 Space Projections

When the distance between each election pair is computed, the final task lies in computation of a suitable embedding. The `Mapel` package implements its embeddings based on several algorithms for graph drawing—elections are treated as vertices and their distances as weighted edges. The `Mapel` embeddings notable for our purposes are the following:

Fruchterman-Reingold (FR) embedding uses the force-directed placement algorithm introduced by Fruchterman and Reingold [39].

Kamada-Kawai (KK) embedding is based on an algorithm for drawing undirected graphs by Kamada and Kawaii [40]. It treats graph vertices as particles connected by springs. The degree of imbalance of the system is then expressed as the total energy of the springs and the algorithm then tries to minimize this energy.

Multidimensional scaling (MDS) embedding is based on[7] standard multidimensional scaling methods, which take a set of distances between points as an input and output their configuration in Euclidean space [41].

Although embedding algorithms add another significant layer of complexity to the studied problem, their thorough discussion is beyond the scope of this thesis.

[7] As the name suggests.

Chapter 4

Elections of Different Sizes

...in which we show the metrics we have defined and the maps we have drawn.

All of the distances described in the previous chapter share a common weakness—they require that all elections have the same number of candidates[1]. This is very restrictive when trying to create a comprehensive map of a large number of real-life datasets. In this thesis, we propose two different approaches to solve this issue. The first one is to extend the existing positionwise distance, which is described in Section 4.3. The second one is to create a framework based on election features, as described in Section 4.4.

4.1 Metodology and datasets

Before we discuss any of our experiments and their resuts, we need to describe the datasets we run our experiments on and the methodology of our experiments. Each `Mapel` *experiment* contains a set of election instances on which it runs the distance computations and then embeds the results. Election instances can be either created directly from votes or generated from various statistical cultures described in Chapter 3. When generated from a statistical culture, we can define a *family size*, which specifies how many election instances will be generated from the statistical culture.

We decided to generate the following distinct datasets:

- *Standard 10×100 dataset* (see Table 4.1) contains only elections of the same size—every election has 10 candidates and 100 voters. Each model has a family size of 15, with the exception of the extremes, where the family size is 1. This dataset was made to closely resemble the dataset of Szufa et al. [3], so that the resulting figures can be compared.
- *Standard mixed dataset* extends the standard 10×100 dataset with elections of different sizes—every culture from the 10x100 dataset is present with 10, 12 and 15 candidates and a family size of 7. The only exception are again the extremes with family size of 1. Again, each election has 100 voters.
- *Mallows mixed dataset* (see Table 4.2) focuses on Mallows elections with larger variance in the choice of the parameter value. The family size is 30. Apart from Mallows elections, it also contains IC as a reference point with family size of 20 and UN, ID and AN compass points with family size of 1. Compass point elections have 100 voters, the rest of the elections have 500 voters.

[1] We say that they require that all elections have the same *size*.

Table 4.1 Statistical cultures in standard 10x100 dataset.

Model Group	Model and Parameters
IC	IC
Urn	Urn(0.01)
Urn	Urn(0.02)
Urn	Urn(0.05)
Urn	Urn(0.1)
Urn	Urn(0.2)
Urn	Urn(0.5)
Mallows	Norm-Mallows(0.001)
Mallows	Norm-Mallows(0.01)
Mallows	Norm-Mallows(0.1)
Mallows	Norm-Mallows(0.5)
Mallows	Norm-Mallows(0.75)
Mallows	Norm-Mallows(0.99)
Mallows	Bi-Mallows(0.001)
Mallows	Bi-Mallows(0.01)
Mallows	Bi-Mallows(0.1)
SP	Walsh
SP	Conitzer
SP	SPOC
SC	SC
tD-Hypercube	Interval
tD-Hypercube	Square
tD-Hypercube	Cube
tD-Hypercube	4D H-Cube
tD-Hypercube	10D H-Cube
tD-Hypercube	20D H-Cube
tD-Hypersphere	Circle
tD-Hypersphere	Sphere
tD-Hypersphere	5D H-Sphere
Extremes	ID
Extremes	UN
Extremes	AN

Table 4.2 Statistical cultures in Mallows mixed dataset.

Model Group	Model and Parameters
IC	IC
Mallows	Norm-Mallows(0.001)
Mallows	Norm-Mallows(0.01)
Mallows	Norm-Mallows(0.05)
Mallows	Norm-Mallows(0.1)
Mallows	Norm-Mallows(0.25)
Mallows	Norm-Mallows(0.5)
Mallows	Norm-Mallows(0.75)
Mallows	Norm-Mallows(0.99)
Mallows	Norm-Mallows(0.999)

A detailed guide on how to use the to the Mapel package was done by Stanislaw Szufa [42]. In our experiments, we use the *offline experiment* class in order to store our election instances and computation results. Our experiment structure is shown in Code listing 4.1.[2]

Code listing 4.1 `Mapel` experiment structure.

```
import mapel.elections as mapel

def experiment_1():
    positionwise_name = 'emd-positionwise'
    my_distance_name = 'example_distance'

    # Load offline experiment
    experiment = mapel.prepare_offline_ordinal_experiment
                      (experiment_id="example_1")
    experiment.prepare_elections()

    # Compute custom distance and print map
    experiment.add_distance(my_distance_name, my_distance_function)
    experiment.compute_distances(distance_id=my_distance_name)
    example_embedding='kk'
    experiment.embed_2d(embedding_id=example_embedding)
    experiment.print_map_2d(legend=False)

    # Compute correlation if applicable
    experiment.compute_distances(distance_id=positionwise_name)
    print("Test correlation between: "
                + my_distance_name + " and "
                + positionwise_name + ": \n")
    experiment.print_correlation_between_distances
            (distance_id_1=positionwise_name,
             distance_id_2=my_distance_name)

    # Compute a custom feature
    feature_name = 'example_feature'
    experiment.add_feature(feature_name, my_feature_function)
    experiment.compute_feature(feature_name)

    # Load a feature from csv file
    experiment.features[feature_name] = experiment.import_feature
                                                   (feature_name)

    # Prepare feature vectors for feature distace computation
     experiment.prepare_feature_vectors(['d', 'a', 'p'])

if __name__ == "__main__":
    experiment_1()
```

[2]Note that we usually run the experiment in parts rather than in bulk as shown here.

4.2 Reference maps

In this section, we introduce maps made by Szufa et al. [3], Boehmer et al. [4] and Faliszewski et al. [31], which we use as a comparison to judge the merits of each of our new metrics.

Figure 4.1 shows the "original" map of elections made by Szufa et al. [3], which Boehmer et al. [4] builds upon by adding the compass to it and testing where real-life elections are placed (see Figure 4.2). Figure 4.3 shows maps done by Faliszewski et al. [31], which visualize three extreme points (UN, ID, AN) and their relationship to agreement, polarization and diversity indices (defined in Subsection 4.4.1).[3] From these maps, we derive the criteria that our maps should meet:

- Cluster elections based on their statistical cultures.
- Place the UN, ID and AN extreme points so that the rest of the elections are placed roughly between them.
- Place IC and Mallows(0.99) elections close to the UN extreme point.
- Place Mallows(0.001) and Mallows(0.01) close to the ID extreme point.
- Place Mallows elections as a path between UN and ID extreme points, such that lower parameter values are closer to ID and higher parameter values are closer to UN.
- Place the Cube, 4D H-Cube, 10D H-Cube and 20D H-Cube relatively close to each other.

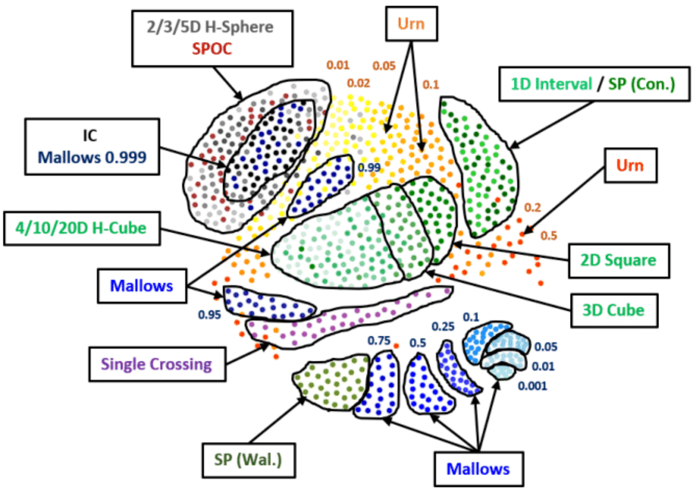
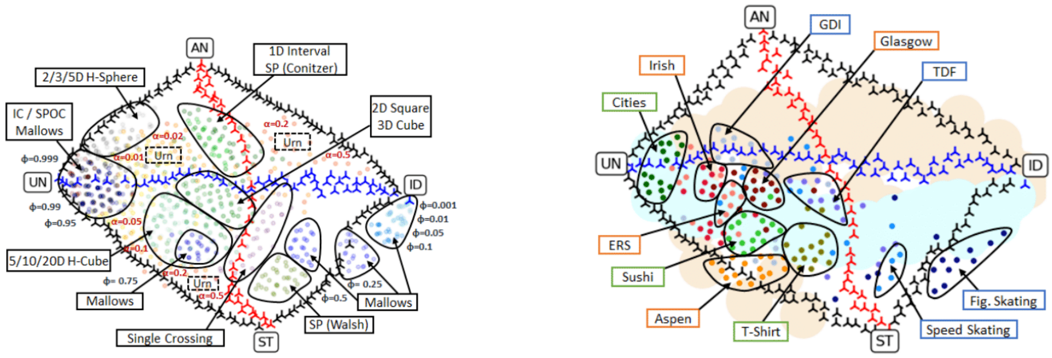


Figure 4.1 A map of synthetic datasets with 10 candidates and 100 votes each. We call it the *original map* for short.

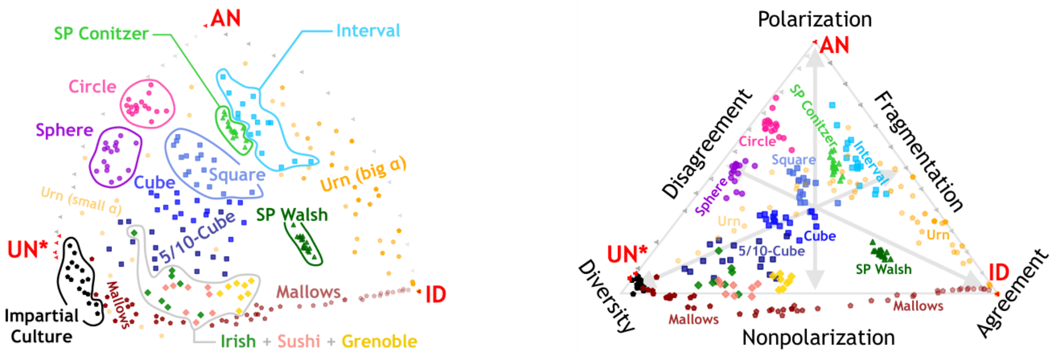
4.3 Extended Positionwise Distance: Positionwise Infinity

The idea behind positionwise infinity distance is simple: given two elections E_1 and E_2 with r and s candidates respectively, we conceptually extend frequency matrices of both these elections

[3] For the reader's convenience, we decided to put all reference maps in one place so that they can be easily searched for when we compare them to our own maps. We realise that it is unfortunate to introduce the diversity maps before defining the diversity, agreement and polarization indices, but we consider it to be the lesser of evils.



■ **Figure 4.2** The compass and the dataset of Szufa et al. [3] (left) and the compass and various real-life elections (right). We call it the *compass maps* for short.



■ **Figure 4.3** A map of elections obtained using isomorphic swap distance and MDS embedding (left) and an affine transformation of this plot where x/y coordinates of the elections are their agreement/diversity indices (right). We call it the *diversity maps* for short.

to infinity and then compute the distance between them. Since infinite matrices are rather hard to work with in terms of computation, in our code implementation we settle for extending each matrix by the factor of $\mathrm{lcm}(r, s)/r$ and $\mathrm{lcm}(r, s)/s$ respectively instead. We also decided to use the *Wasserstein distance* (WD) instead of EMD for our positionwise infinity distance[4].

Formally:

▶ **Definition 4.1** (Stepwise Function). *Given a vector $a = (a_1, \ldots, a_m) \in \mathbb{R}^m_+$, we identify it with a* stepwise function *$a\colon [0, 1] \to \mathbb{R}_+$, such that for each $i \in [m]$ and each x in $(\frac{i-1}{m}, \frac{i}{m}]$, we have:*

$$a(x) = (m \cdot a_i)/(a_1 + \cdots + a_m)$$

For the sake of completeness, we set $a(0) = 0$. For each $x \in [0, 1]$ we set $A(x) = \int_0^x a(y)dy$. The normalizing factor m is chosen so that $A(1) = 1$[5].

▶ **Definition 4.2** (Wasserstein Distance). *For two vectors $a \in \mathbb{R}^s$ and $b \in \mathbb{R}^t$[6], we define their* Wasserstein distance *(WD), denoted $W(a, b)$, as the Wasserstein distance of their corresponding*

[4] The Wasserstein distance is closely related to the Earth mover's distance and its choice is discussed in the *Distances Between Top-Truncated Elections of Different Sizes* IJCAI paper. We consider this discussion beyond the scope of this thesis.

[5] Because of that, one may interpret $a(x)$ as a probability density function and $A(x)$ as its cumulative distribution function.

[6] It is possible that $s \neq t$.

functions:

$$W(a,b) = \int_0^1 |A(x) - B(x)| dx$$

As $a(x)$ and $b(x)$ are stepwise functions and $A(x)$ and $B(x)$ are piecewise linear functions, $W(a,b)$ can be computed in polynomial time, assuming we can perform arithmetic operations in polynomial time.

We redefine the standard positionwise distance so that it uses WD instead of EMD:

▶ **Definition 4.3** (Wasserstein Positionwise Distance)**.** *Let $E_a = (C_a, V_a)$ and $E_b = (C_b, V_b)$ be two elections such that $|C_a| = |C_b|$, $C_a = \{c_{a_1}, \dots, c_{a_m}\}$ and $C_b = \{c_{b_1}, \dots, c_{b_m}\}$. Let $F_{E_a} = [\psi_{E_a}(c_{a_1}), \dots, \psi_{E_a}(c_{a_m})]$ and $F_{E_b} = [\psi_{E_b}(c_{b_1}), \dots, \psi_{E_b}(c_{b_m})]$ be their respective frequency matrices. The Wasserstein positionwise distance between matrices F_{E_a} and F_{E_b} is defined as follows:*

$$\text{W-POS}(F_{E_a}, F_{E_b}) = \min_{\sigma \in \Pi([m],[m])} \frac{1}{m} \sum_{i \in [m]} W(\psi_{E_a}(c_{a_i}), \psi_{E_b}(c_{b_{\sigma(i)}}))$$

The positionwise distance between E_a and E_b is then equal to the positionwise distance between their frequency matrices:

$$\text{W-POS}(E_a, E_b) = \text{W-POS}(F_{E_a}, F_{E_b})$$

For a frequency matrix $F_E = [\psi_E(c_1), \dots, \psi_E(c_m)]$ and a positive integer k, we write $\text{str}_{\text{mk}}(F_E)$ to denote the matrix obtained from F_E by copying each of its component vectors k times, so that for each $i \in [mk]$, the i-th vector of $\text{str}_{\text{mk}}(F_E)$ is $\psi_E(c_{\lceil i/k \rceil})$.

▶ **Definition 4.4** (Positionwise Infinity Distance)**.** *Let E_a and E_b be two elections with possibly differing numbers of candidates and voters. Let $F_{E_a} = [\psi_{E_a}(c_{a_1}), \dots, \psi_{E_a}(c_{a_r})]$ and $F_{E_b} = [\psi_{E_b}(c_{b_1}), \dots, \psi_{E_b}(c_{b_s})]$ be their respective frequency matrices, such that* $\text{lcm}(r,s) = m$*. Then the* positionwise infinity *distance between E_a and E_b, denoted* $\text{POS-INFTY}(E_1, E_2)$*, is defined as follows:*

$$\text{POS-INFTY}(E_1, E_2) = \text{W-POS}(\text{str}_\text{m}(F_{E_a}), \text{str}_\text{m}(F_{E_b}))$$

In our implementation, we normalize this distance by the number of candidates. Figures 4.4 and 4.5 show maps the positionwise infinity distance produced given the standard and the standard mixed datasets, respectively[7]. When we compare the two figures with each other, we see that the positionwise infinity distance successfuly clusters elections of the same culture but different size, which is precisely what we desire.Both figures resemble the original map and compass maps with their positioning of cultures relative to each other. IC is correctly paced near the UN extreme, Mallows (0.001) and Mallows(0.01) merge with ID extreme point and so on. Interestingly, there are nonzero distances between extreme point elections of different sizes. Even though worrisome at first glance, this result is natural - since we use only approximations of extreme points, there has to be a verz small, but nonyzero distance between them. The same effect can me observed for Mallows (0.001) and Mallows(0.01), since their extremely low parameter values make them very close to identity extreme point in practice. It also seems that some clustering based on election size emerges the closer we get to ID extreme point. A reason for this currently remains unknown to us.

The PCC between the standard EMD-positionwise distace by Szufa et al. [3] and our positionwise infinity distance is 0.999, which is a very strong positive correlation. This value was computed based on distance vectors from the standard dataset—since EMD-positionwise distance can work only with elections of the same sizes.

[7] All positionwise infinity maps use the KK embedding.

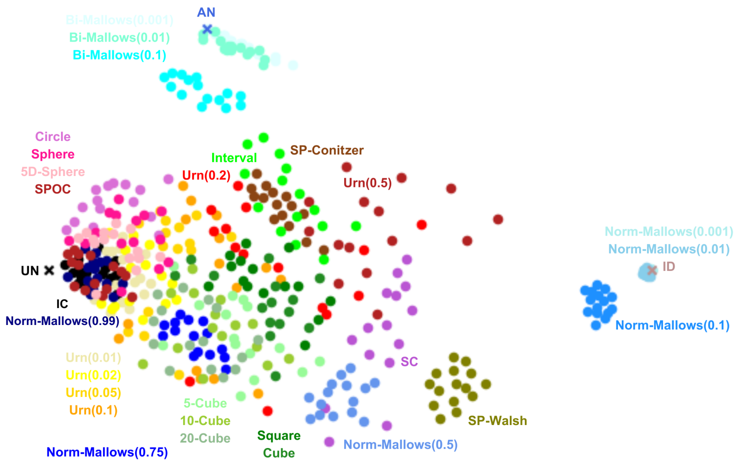


Figure 4.4 The positionwise infinity distance and the standard dataset.

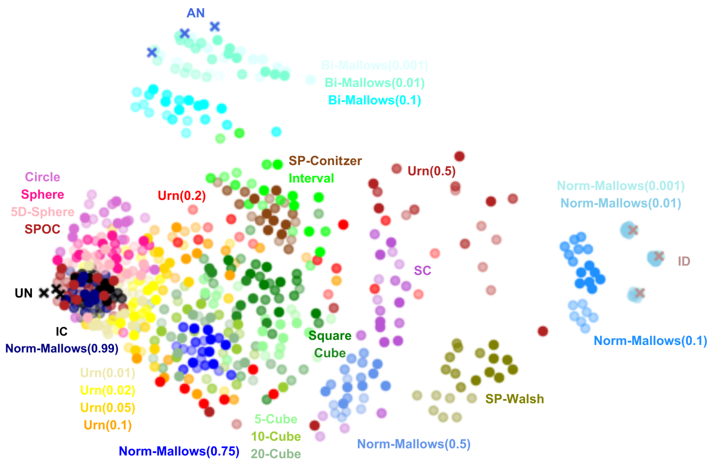


Figure 4.5 The positionwise infinity distance and the standard mixed dataset.

4.4 Feature Distance

Let $\Lambda = \{\lambda_1, \dots, \lambda_k\}$ be a set of functions that accept an election as their argument and output a value from $[0, 1]$. Each of these functions can be called an *election feature*. Given an election E, we call $(\lambda_1(E), \dots, \lambda_k(E))$ the *feature vector* of E and denote it as Λ_E. The feature distance d_Λ between two elections E_1 and E_2 is then the distance between their corresponding feature vectors Λ_{E_1} and Λ_{E_2}. Neither EMD nor WD can be used in this case, however, because there is no inherent meaning to the ordering of the election features in the feature vector. In their stead, we use the ℓ_1 and ℓ_2 distances, giving rise to ℓ_1-d_Λ and ℓ_2-d_Λ feature distances[8].

All of the election features presented in the subsections below are normalized between 0 and 1 when used for the feature distance computation. All feature distance maps use the MDS embedding.

[8]Sadly, cosine similarity fails both triangle inequality and identity of indiscernibles. Therefore, it is not a distance metric and cannot be used for our experiments alongside the ℓ_1 and ℓ_2 distances.

4.4.1 DAP Distance

The *DAP distance* takes advantage of *agreement*, *diversity* and *polarization* indeces introduced by Faliszewski et al. [31].

▶ **Definition 4.5** (Agreement Index)**.** *Let $E = (C, V)$ be an election. Its* agreement index*, denoted A_E, is computed as follows:*

$$A_E = \frac{\sum_{\{a,b\} \subseteq C} |M_E(a,b) - M_E(b,a)|}{\binom{|C|}{2}}$$

In order to define diversity and polarization, we first need to define the k-Kemeny ranking:

▶ **Definition 4.6** (k-Kemeny Ranking)**.** k-Kemeny rankings *of election $E = (C, V)$ are the elements of a set $\mathcal{X} = \{\chi_1, \dots, \chi_k\}$ of k linear orders over C that minimize $\sum_{v \in V} \min_{i \in [k]} \mathrm{swap}(v, \chi_i)$.*

▶ **Definition 4.7** (Diversity Index)**.** *Let $E = (C, V)$ be an election. The* diversity index*, denoted D_E, is defined as:*

$$D_E = \frac{\sum_{k \in [|V|]} \kappa_k(E)/k}{|V| \cdot \binom{|C|}{2}}$$

▶ **Definition 4.8** (Polarization Index)**.** *Let $E = (C, V)$ be an election. The* polarization index*, denoted P_E, is computed as follows:*

$$P_E = \frac{2(\kappa_1(E) - \kappa_2(E))}{|V| \cdot \binom{|C|}{2}}$$

▶ **Definition 4.9** (DAP Feature vector)**.** *Given an election E, the* DAP feature vector *is defined as:*

$$\Lambda_{\mathrm{DAP_E}} = (D_E, A_E, P_E)$$

Since the DAP distance between two election is equal to the distance between their DAP feature vectors, we distinguish ℓ_1-DAP and ℓ_2-DAP feature distances. Both are anonymous and neutral.

Figures 4.6 and 4.7 show maps the ℓ_1-DAP distance produced given the standard and the standard mixed datasets respectively. Similarly, figure 4.8 shows maps the ℓ_2-DAP distance produced given these datasets.

We note that there is no significant difference between the ℓ_1 and ℓ_2 distance metrics—both produce maps of triangular shape with extreme points in their vertices and place each statistical culture in the similar spot on the map with respect to other cultures. We therefore conclude that in terms of the DAP feature distance, the choice between ℓ_1 and ℓ_2 is not important. Because of this, we further describe only ℓ_1-DAP distance applied on the standard mixed dataset (Figure 4.7). This map bears a striking similarity to our reference diversity maps[9]—placement of all the cultures is almost identical, perhaps with the difference that the triangle shape is slightly stretched to the right bottom.

[9] Figure 4.3.

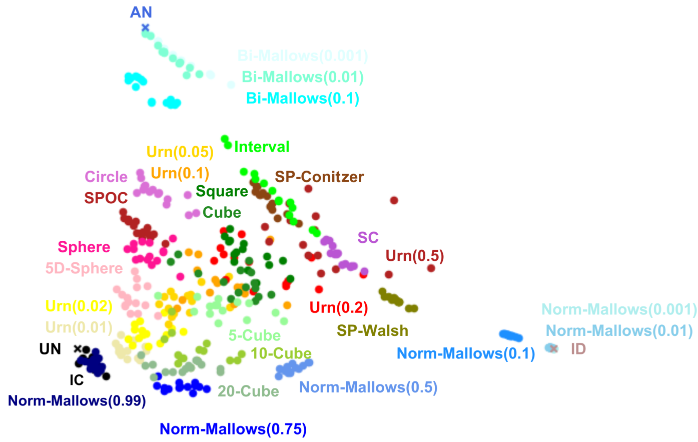


Figure 4.6 The ℓ_1-DAP distance applied on the standard dataset.

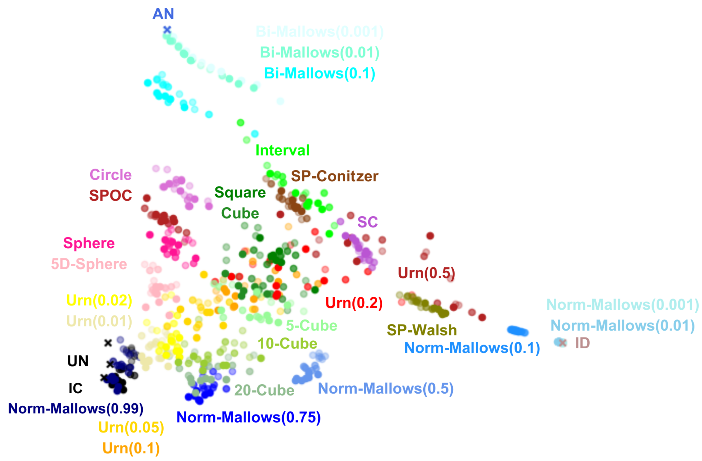


Figure 4.7 The ℓ_1-DAP distance applied on the standard mixed dataset.

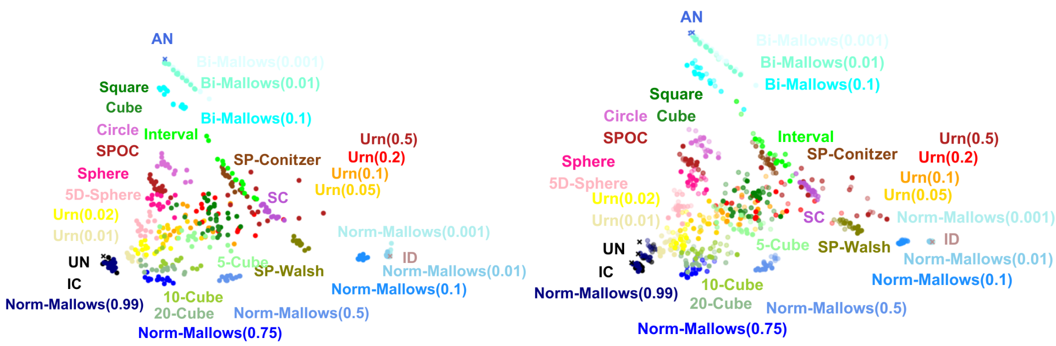


Figure 4.8 The ℓ_2-DAP distance applied on the standard dataset (left) and on the standard mixed dataset (right).

4.4.2 APE Distance

In this thesis, we introduce a new election feature called the *entropy index*. It is defined as follows:

▶ **Definition 4.10** (Entropy Index). *Let $E = (C, V)$ be an election. The* entropy index *of its frequency matrix F_E, denoted H_E, is defined as follows:*

$$H_E = -\sum_{i=1}^{|C|}\sum_{j=1}^{|C|} F_{E_{i,j}} \cdot \log_2(F_{E_{i,j}})$$

Naturally, for a given number of candidates the largest possible entropy is that of F_{UN} matrix.

▶ **Definition 4.11** (APE Feature Vector). *Given an election E, the* APE feature vector *is defined as:*

$$\Lambda_{\mathrm{APE_E}} = (A_E, P_E, H_E)$$

Since the APE distance between two election is equal to the distance between their APE feature vectors, we again distinguish ℓ_1-APE and ℓ_2-APE feature distances.

Figures 4.9 and 4.10 show maps the ℓ_1-APE distance produced given the standard and the standard mixed datasets. Similarly, figure 4.11 shows maps the ℓ_2-APE distance produced given these datasets.

As with the DAP distance, we note that there is still no major difference between the ℓ_1-APE and ℓ_2-APE distance metrics, so we further describe only maps made with ℓ_1-APE distance. The maps still have roughly triangular form, even though it is more deformed into arc-like shape in the bottom left. Extreme points are placed in the appropriate triangle vertices. While still roughly similar to the reference diversity maps, the urn elections especially seem to be more stacked to the bottom left part of the map.

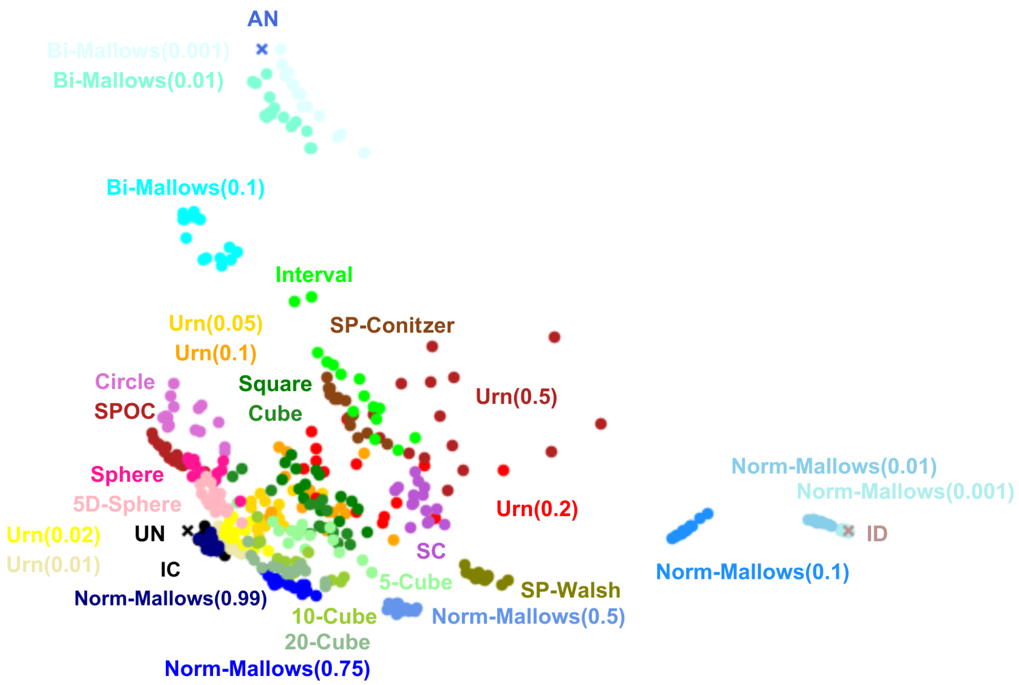


Figure 4.9 The ℓ_1-APE distance applied on the standard dataset.

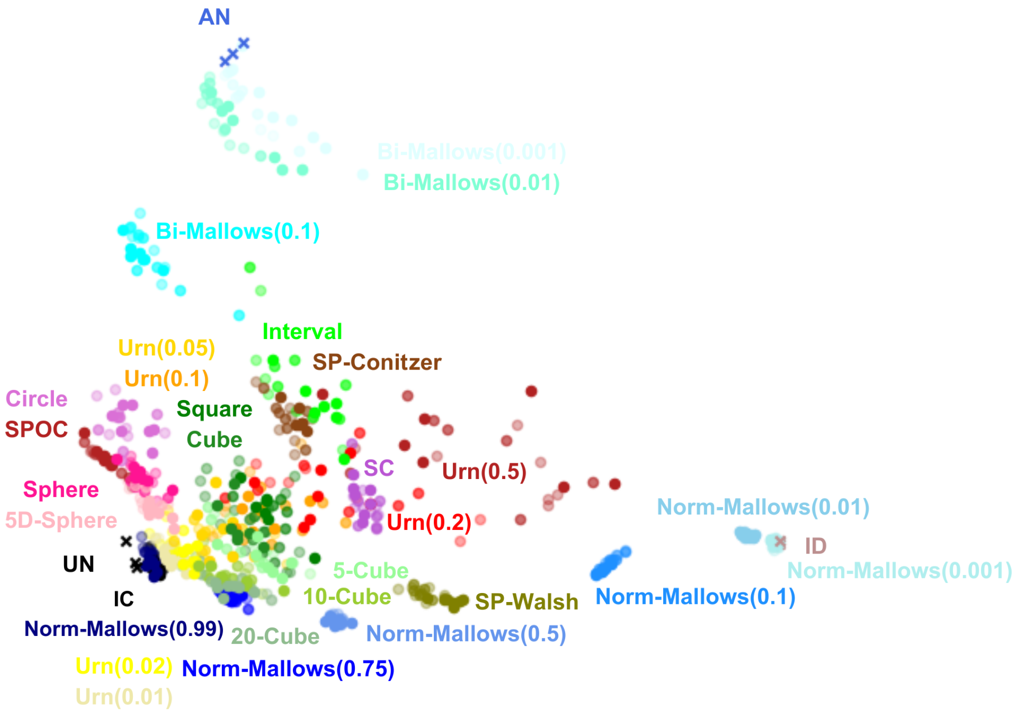


Figure 4.10 The ℓ_1-APE distance applied on the standard mixed dataset.

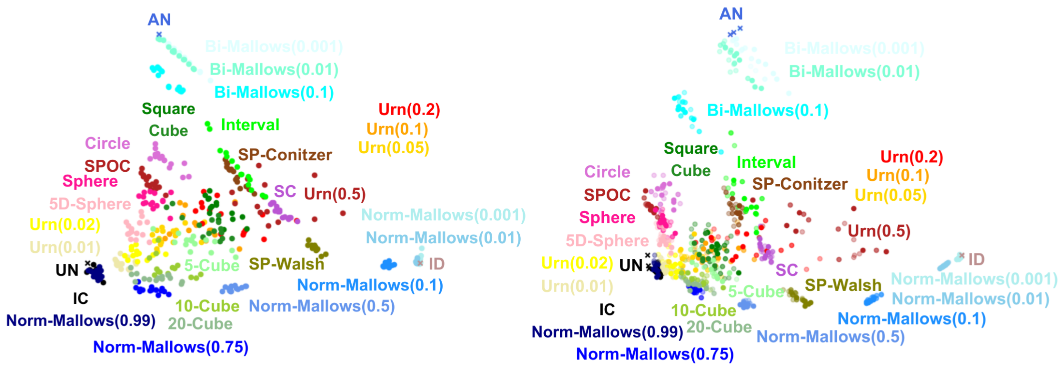


Figure 4.11 The ℓ_2-APE distance applied on the standard dataset (left) and on the standard mixed dataset (right).

4.4.3 Compass Distance

The *compass* distance is based on distances from the extreme points:

▶ **Definition 4.12** (Compass feature vector)**.** *Let E be an election with m candidates. Its* compass feature vector *is defined as follows:*

$$\Lambda_{COMPASS_E} = (\,\text{POS-INFTY}(E, \text{UN}_m), \text{POS-INFTY}(E, \text{AN}_m), \text{POS-INFTY}(E, \text{ID}_m), \text{POS-INFTY}(E, \text{ST}_m))$$

The compass feature distance is anonymous and neutral and again we distinguish between

ℓ_1-COMPASS and ℓ_2-COMPASS distances. Figure 4.12 shows maps the ℓ_1-COMPASS distance produced given the standard and the standard mixed datasets. Similarly, Figure 4.13 shows maps the ℓ_2-COMPASS distance produced given these datasets. All these maps manage to separate different statistical cultures, which is a good sign. Cultures in the standard mixed dataset form clusters based on the statistical culture, not based on the number of candidates, which is also promising. Unlike the original and the compass maps, however, these maps appear to be rather "one dimensional"—squashed in terms of the y axis. In addition to this, they also split the Bi-Mallows cultures in half, which is not desirable. However, since we verified that all Bi-Mallows elections in the experiment have a relatively small value of the compass distance between them[10], this seems to be a problem of the embedding rather than the distance per se. For the time being, we do not examine this distance any further, but we would like to go back to it in our future works.

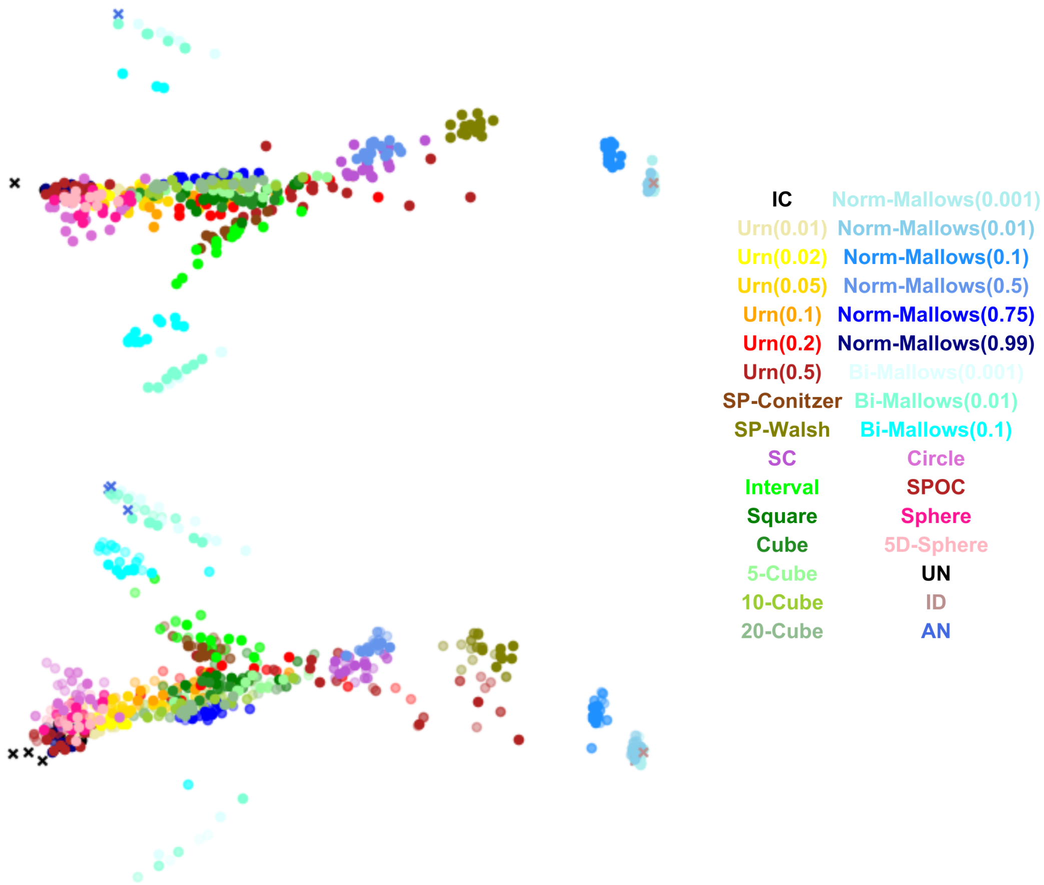


Figure 4.12 The ℓ_1-COMPASS distance with the standard dataset (top) and the standard mixed dataset (bottom).

[10] We manually tested this and the average distance between a pair of Bi-Mallows elections is roughly six times smaller than the average distance between a pair of elections in which one is Bi-Mallows and the other one is IC.

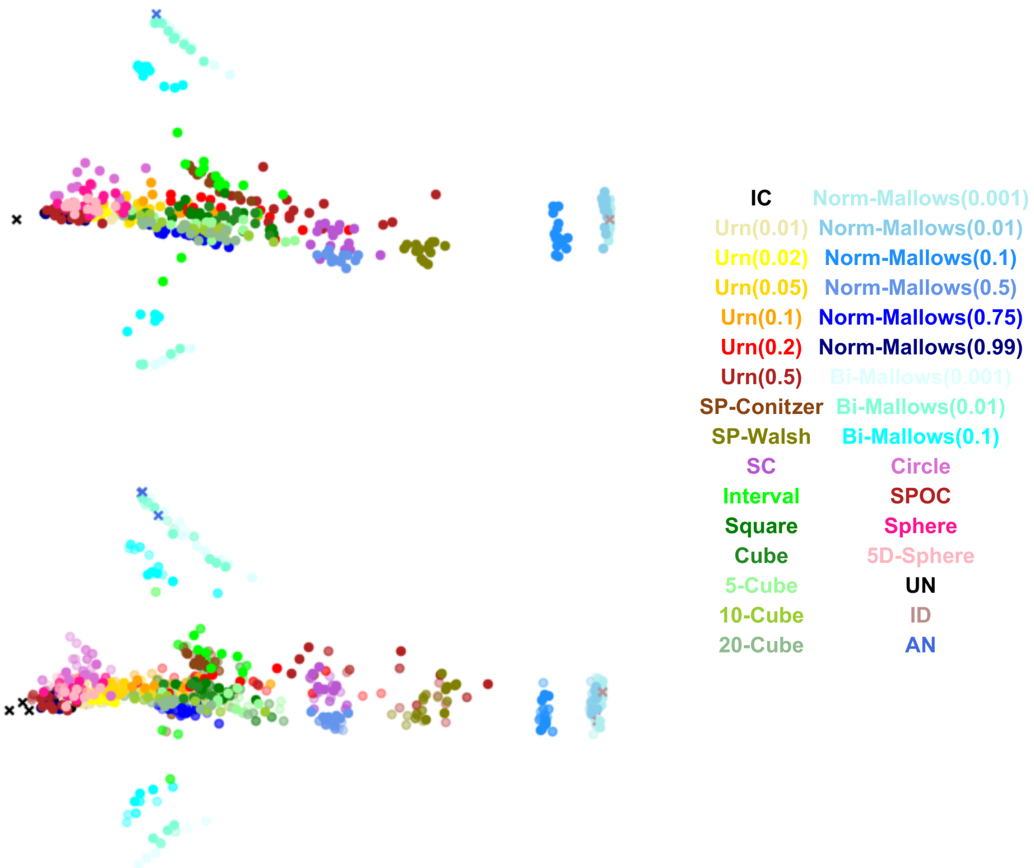


Figure 4.13 The ℓ_2-COMPASS distance with the standard dataset (top) and the standard mixed dataset (bottom).

4.4.4 MOV Distance

The *MOV feature distance* is based on the concept of margin of victory. Given an election E, we denote $\text{MOV}_X(E)$ to be its margin of victory given the voting rule X.

▶ **Definition 4.13** (MOV feature vector)**.** *Let E be an election with m candidates. Its* MOV feature vector *is defined as follows:*

$$\Lambda_{MOV_E} = (\text{MOV}_{\text{PLURALITY}}(E), \text{MOV}_{\text{VETO}}(E), \text{MOV}_{\text{1-APPROVAL}}(E), \ldots, \text{MOV}_{\text{m-APPROVAL}}(E), \text{MOV}_{\text{BORDA}}(E))$$

We based our implementation of MOV on algorithms described by Xia [43]. The MOV feature distance is again anonymous and neutral and we distinguish between ℓ_1-MOV and ℓ_2-MOV distances. Figure 4.14 shows the standard dataset, the ℓ_1-MOV distance (top) and the ℓ_2-MOV distance (bottom). Since these maps fail the basic criterion to separate statistical cultures into clusters, we conclude the MOV distance is much less promising than other feature distances we propose and thus we do not examine it any further.

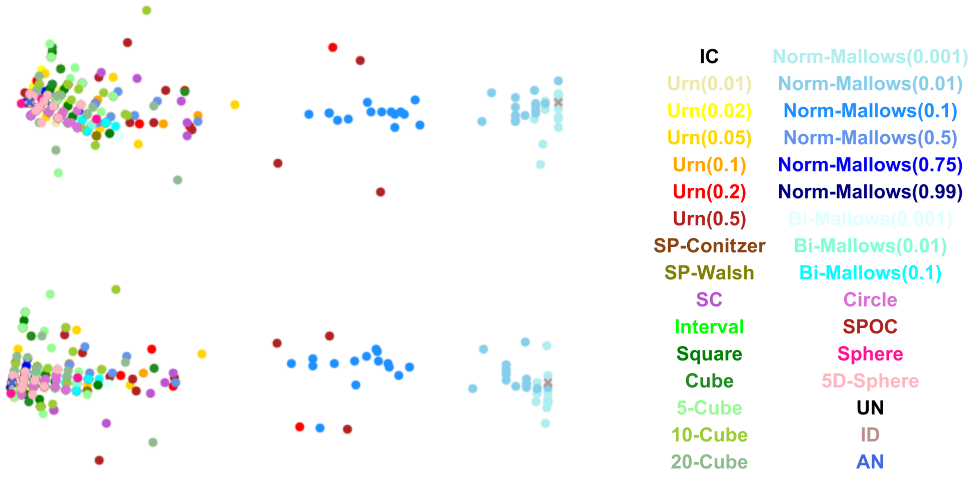


Figure 4.14 Standard dataset with the ℓ_1-MOV distance (top) and with the ℓ_2-MOV distance (bottom).

4.5 Mallows mixed dataset

We tested the following distance metrics on the mallows mixed datasets: positionwise infinity, ℓ_1-DAP, ℓ_2-DAP, ℓ_1-APE, and ℓ_2-APE. Surprisingly, all of these distances (both the positionwise infinity and feature metrics[11]) produced strikingly similar maps—three extreme points as vertices of a triangle and a path of Mallows elections from UN to ID extreme points. The map for ℓ_1-DAP is shown below as Figure 4.15.

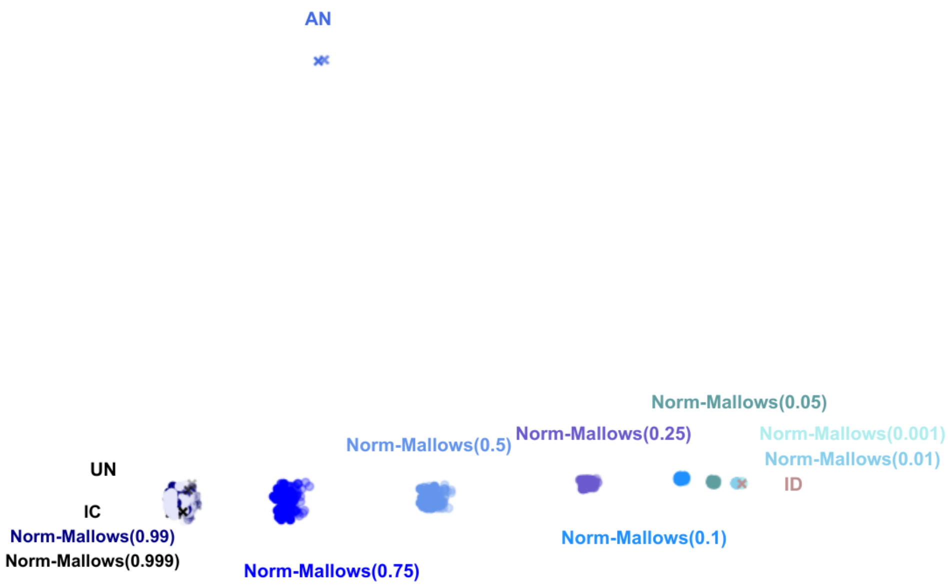


Figure 4.15 Mallows mixed dataset with the ℓ_1-DAP distance.

[11] Note that for this dataset, we used approximated agreement, polarization and diversity indices to speed up their computation. For the purposes of our distance metric, the approximation error should be negligible.

Chapter 5
Conclusion

…in which we sum up all the work done in this thesis, discuss what could have been done better, what could have been done worse, and how the results in this thesis may be used in the future.

In Chapter 2 we described some of the building blocks of computational social choice: elections, voting systems and their approaches to making a collective decision. We discussed the properties of the introduced voting systems and when possible, showed the interesting ones on election profiles.

In Chapter 3 we described the maps of elections framework—the distance metrics between voters and elections, generative models for synthetic data and embedding algorithms.

In Chapter 4 we presented our distance metrics for elections of different sizes and we discussed the experimental results these distances yielded. Based on similarity between Figures 4.4 and 4.5 and Figure 4.2, as well as on the value of PCC, we consider our positionwise infinity distance to be a good extension of the "old" EMD-positionwise distance. When it comes to the feature distances, we consider ℓ_1-DAP, ℓ_2-DAP, ℓ_1-APE, and ℓ_1-APE all to be very good distance metrics as discussed. The DAP distances produced a more triangle-like shape, which is more in line with our expectations based on the reference diversity map (Figure 4.3), but it is not self-evident that the triangle shape is necessarily superior. The compass metric had some favourable traits, but we suppose it will require nontrivial fine-tuning. Sadly, we conclude that the MOV features we tested are not informative enough to be used as a feature vector.

Our work could be continued in the following ways:

- Fine-tune the compass distance and consider its other possible variants, e.g., a feature vector which would contain positionwise distances from selected point on paths between the extreme points.
- In this thesis, we used only projections on a 2D plane, which is the state of the art to the best of our knowledge. In the future, we would like to test projections onto a 3D plane (probably a sphere).
- We ended up not testing our metrics on real-life election data. We suspect that they would follow the same pattern as they did in the reference compass map (see the left map in Figure 4.2), but it would be worthwhile to check this assumption.
- It would be worthwhile to develop a framework for judging the quality of an election map. So far, we form the judgment based on human perception of how the map fulfils our preconceived desiderata. However, it could be beneficial to formalize this in form of some quality metric.

Appendix A
Code

My commits to the `Mapel` Github repository [44] are shown in Table A.1.

Table A.1 List of my `Mapel` commits.

Date	Hash Identifier
2023-10-02	d7cf20b93725e560ec9e685fb91dd73399a17e35
2023-10-24	6abc2c4a3155b8d63c28434cb984b7caa4640fcc
2023-11-09	5a9cc3f44b64eb8dc25c7e211c5031de6b721ad4

The majority of my code was made with the intention to be integrated into `Mapel`. However, major changes were done in the `Mapel` package while I was working on this thesis. These changes sadly break backwards compatibility. My implementation of MOV and compass distances as well as the `utilities.py` script for generating the map.csv file are all dependent on the old `Mapel` version, making the `compass.py`, `mov_featurewise.py` and `utilities.py` files not runnable (or produce unexpected results). Since the MOV distance is not useful, it will be dropped, and the compass distance and utilities script will be rewritten to comply with current form of *Mapel* package in the future. The positionwise infinity implementation does not depend on the mapel version and the DAP and APE distances were made to be runnable in the new `Mapel` version[1]. Table A.2 describes important files and folders in the attachment file. In order to run (most of) the code, it is necessary to install the current `Mapel` package from provided source code in the `mapel` folder. For guidance on how to use the `Mapel` package, consult the manual written by Szufa [42].

[1] Which of course makes them not runnable in the old one. There is no winning sometimes.

Table A.2 Description of my source files.

File and Its Location	Description
attachment/mapel/	The current Mapel package source code.
attachment/experiments/	The folder with experiment data.
attachment/mapel/mapel-elections/src/ mapel/elections/objects/ ElectionExperiment.py	Mapel file with added functionality to support my distances.
attachment/mapel/mapel-elections/src/ mapel/elections/objects/ ElectionFeatures.py	My file containing supporting code for MOV distance and feature distance.
attachment/compass.py	My implementation of the compass distance.
attachment/compute_features.py	Script for computing election features.
attachment/diversity_truncated.py	Fast computation of approximate indices, implemented by Tomasz Wąs.
attachment/entropy.py	Fast computation of entropy index, implemented by Stanisław Szufa.
attachment/features_vector.py	My implementation of the feature distance.
attachment/mov_featurewise.py	My implementation of compass distance.
attachment/ positionwise_different_sizes.py	My implementation of the positionwise infinity distance.
attachment/ positionwise_diffsize_experiments.py	Driver functions for positionwise infinity distance.
attachment/utilities.py	Utility script to generate a map.csv file.

Bibliography

1. AZQUOTES.COM. *Henry David Thoreau Quote* [online]. [visited on 2024-12-24]. Available from: `https://www.azquotes.com/quote/1091498`.
2. ROTHE, Jörg (ed.). *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016 [visited on 2023-02-16]. Springer Texts in Business and Economics. ISBN 978-3-662-47903-2. Available from DOI: `10.1007/978-3-662-47904-9`.
3. SZUFA, Stanisław; FALISZEWSKI, Piotr; SKOWRON, Piotr; SLINKO, Arkadii; TALMON, Nimrod. Drawing a Map of Elections in the Space of Statistical Cultures. In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2020, pp. 1341–1349. AAMAS '20. ISBN 9781450375184.
4. BOEHMER, Niclas; BREDERECK, Robert; FALISZEWSKI, Piotr; NIEDERMEIER, Rolf; SZUFA, Stanisław. Putting a Compass on the Map of Elections. In: ZHOU, Zhi-Hua (ed.). *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. International Joint Conferences on Artificial Intelligence Organization, 2021, pp. 59–65. Available from DOI: `10.24963/ijcai.2021/9`. Main Track.
5. MATTEI, Nicholas; WALSH, Toby. Preflib: A library for preferences http://www. preflib. org. In: *Algorithmic Decision Theory: Third International Conference, ADT 2013, Bruxelles, Belgium, November 12-14, 2013, Proceedings 3*. Springer, 2013, pp. 259–270.
6. SZUFA, Stanisław. *Mapel-elections* [online]. Version 2.0.25 [visited on 2024-01-01]. Available from: `https://pypi.org/project/mapel-elections/`.
7. SZUFA, Stanisław. *Mapel-marriages* [online]. Version 2.0.25 [visited on 2024-01-01]. Available from: `https://pypi.org/project/mapel-marriages/`.
8. SZUFA, Stanisław. *Mapel-roommates* [online]. Version 2.0.25 [visited on 2024-01-01]. Available from: `https://pypi.org/project/mapel-roommates/`.
9. RUBNER, Yossi; TOMASI, Carlo; GUIBAS, Leonidas J. A metric for distributions with applications to image databases. In: *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*. IEEE, 1998, pp. 59–66.
10. BOEHMER, Niclas; BREDERECK, Robert; FALISZEWSKI, Piotr; NIEDERMEIER, Rolf; SZUFA, Stanisław. *Putting a Compass on the Map of Elections* [online]. 2021. [visited on 2023-07-31]. Available from DOI: `https://doi.org/10.48550/arXiv.2105.07815`.
11. DEZA, Elena; DEZA, Michel Marie. *Encyclopedia of distances*. Springer, 2009. Available also from: `https://www.uco.es/users/ma1fegan/Comunes/asignaturas/vision/Encyclopedia-of-distances-2009.pdf`.

12. KONCZAK, Kathrin; LANG, Jérôme. Voting procedures with incomplete preferences. In: *Proc. IJCAI-05 Multidisciplinary Workshop on Advances in Preference Handling*. Citeseer, 2005, vol. 20. Available also from: `http://www.eecs.harvard.edu/cs286r/courses/fall11/papers/KL05.pdf`.
13. MAREŠ, Martin; VALLA, Tomáš. *Průvodce labyrintem algoritmů*. CZ.NIC, 2017. ISBN 978-80-88168-19-5.
14. BLAIS, André. The debate over electoral systems. *International Political Science Review*. 1991, vol. 12, no. 3, pp. 239–260.
15. ARROW, Kenneth J. *Social choice and individual values*. Vol. 12. Yale university press, 2012. ISBN 978-0-300-17931-6.
16. YOUNG, H Peyton; LEVENGLICK, Arthur. A consistent extension of Condorcet's election principle. *SIAM Journal on applied Mathematics*. 1978, vol. 35, no. 2, pp. 285–300.
17. WIKIPEDIA. *Consistency criterion — Wikipedia, The Free Encyclopedia* [online]. 2023. [visited on 2023-06-03]. Available from: `https://en.wikipedia.org/wiki/Consistency_criterion#Kemeny-Young_method`.
18. BRAMS, Steven J.; FISHBURN, Peter C. Approval Voting. *American Political Science Review*. 1978, vol. 72, no. 3, pp. 831–847. Available from DOI: `10.2307/1955105`.
19. FELSENTHAL, Dan S; NURMI, Hannu. Two types of participation failure under nine voting methods in variable electorates. *Public Choice*. 2016, vol. 168, pp. 115–135.
20. OHTSUBO, Yohsuke; WATANABE, Yoriko. Contrast effects and approval voting: An illustration of a systematic violation of the independence of irrelevant alternatives condition. *Political Psychology*. 2003, vol. 24, no. 3, pp. 549–559.
21. TIDEMAN, T Nicolaus. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*. 1987, vol. 4, pp. 185–206.
22. MULLER, Eitan; SATTERTHWAITE, Mark A. The equivalence of strong positive association and strategy-proofness. *Journal of Economic Theory*. 1977, vol. 14, no. 2, pp. 412–418.
23. GIBBARD, Allan. Manipulation of voting schemes: a general result. *Econometrica: journal of the Econometric Society*. 1973, pp. 587–601.
24. SATTERTHWAITE, Mark Allen. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory*. 1975, vol. 10, no. 2, pp. 187–217.
25. HEMASPAANDRA, Edith; HEMASPAANDRA, Lane A; ROTHE, Jörg. Exact analysis of Dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. *Journal of the ACM (JACM)*. 1997, vol. 44, no. 6, pp. 806–825.
26. FELSENTHAL, Dan S. Review of paradoxes afflicting various voting procedures where one out of m candidates (m 2) must be elected [online]. 2010 [visited on 2023-06-03]. Available from: `https://www.lse.ac.uk/cpnss/assets/documents/voting-power-and-procedures/workshops/2010/duBaffy2010-Felsenthal.pdf`.
27. BARTHOLDI III, John J; ORLIN, James B. Single transferable vote resists strategic voting. *Social Choice and Welfare*. 1991, vol. 8, no. 4, pp. 341–354. Available also from: `https://courses.cs.duke.edu/fall06/cps296.2/stv_hard.pdf`.
28. GALLAGHER, Michael. Monotonicity and non-monotonicity at PR-STV elections. In: *Annual Conference of the Elections, Public Opinion and Parties (EPOP) specialist group, University of Lancaster*. 2013, pp. 13–15.

29. MOHSIN, Farhad; HAN, Qishen; RUAN, Sikai; CHEN, Pin-Yu; ROSSI, Francesca; XIA, Lirong. Computational Complexity of Verifying the Group No-show Paradox. In: *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems* [online]. 2023, pp. 2877–2879 [visited on 2023-07-13]. Available from: `https://www.southampton.ac.uk/~eg/AAMAS2023/pdfs/p2877.pdf`.

30. FALISZEWSKI, Piotr; SKOWRON, Piotr; SLINKO, Arkadii; SZUFA, Stanislaw; TALMON, Nimrod. Isomorphic Distances Among Elections. In: FERNAU, Henning (ed.). *Computer Science - Theory and Applications - 15th International Computer Science Symposium in Russia, CSR 2020, Yekaterinburg, Russia, June 29 - July 3, 2020, Proceedings*. Springer, 2020, vol. 12159, pp. 64–78. Lecture Notes in Computer Science. Available from DOI: `10.1007/978-3-030-50026-9\_5`.

31. FALISZEWSKI, Piotr; KACZMARCZYK, Andrzej; SORNAT, Krzysztof; SZUFA, Stanislaw; WAS, Tomasz. Diversity, Agreement, and Polarization in Elections. *CoRR*. 2023, vol. abs/2305.09780. Available from DOI: `10.48550/arXiv.2305.09780`.

32. BOEHMER, Niclas; FALISZEWSKI, Piotr; NIEDERMEIER, Rolf; SZUFA, Stanisław; WĄS, Tomasz. Understanding Distance Measures Among Elections. In: RAEDT, Lud De (ed.). *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*. International Joint Conferences on Artificial Intelligence Organization, 2022, pp. 102–108. Available from DOI: `10.24963/ijcai.2022/15`. Main Track.

33. BOEHMER, Niclas; BREDERECK, Robert; ELKIND, Edith; FALISZEWSKI, Piotr; SZUFA, Stanisław. Expected Frequency Matrices of Elections: Computation, Geometry, and Preference Learning. *Advances in Neural Information Processing Systems*. 2022, vol. 35, pp. 31347–31359.

34. *SciPy v1.11.4 Manual* [online]. [N.d.]. [visited on 2024-01-07]. Available from: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.linear_sum_assignment.html`.

35. FALISZEWSKI, Piotr; SORNAT, Krzysztof; SZUFA, Stanisław. The complexity of subelection isomorphism problems. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2022, vol. 36, pp. 4991–4998. No. 5.

36. BOEHMER, Niclas; FALISZEWSKI, Piotr; KRAICZY, Sonja. Properties of the mallows model depending on the number of alternatives: a warning for an experimentalist. In: *International Conference on Machine Learning*. PMLR, 2023, pp. 2689–2711.

37. WALSH, Toby. Generating single peaked votes. *arXiv preprint arXiv:1503.02766*. 2015. Available also from: `https://arxiv.org/pdf/1503.02766.pdf`.

38. CONITZER, Vincent. Eliciting single-peaked preferences using comparison queries. In: *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. 2007, pp. 1–8. Available from DOI: `10.1145/1329125.1329204`.

39. FRUCHTERMAN, Thomas MJ; REINGOLD, Edward M. Graph drawing by force-directed placement. *Software: Practice and experience*. 1991, vol. 21, no. 11, pp. 1129–1164.

40. KAMADA, Tomihisa; KAWAI, Satoru, et al. An algorithm for drawing general undirected graphs. *Information processing letters*. 1989, vol. 31, no. 1, pp. 7–15. Available also from: `https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b8d3bca50ccc573c5cb99f7d201e8acce6618f04`.

41. MEAD, Al. Review of the development of multidimensional scaling methods. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1992, vol. 41, no. 1, pp. 27–39.

42. SZUFA, Stanisław. *Mapel Tutorial* [online]. 2023. [visited on 2023-08-01]. Available from: https://file.notion.so/f/s/e7bcf5dd-864e-4003-9c8b-38ae6d719287/Untitled.pdf?id=8cf2bea8-3240-42c9-924b-6c86e1db52c4&table=block&spaceId=7402432b-2656-4896-a448-256d81a35ee4&expirationTimestamp=1690876800000&signature=VrPAotQgmM9lEo0Ly8BOO359qFhLxJoiFkiUoVSKzZo.

43. XIA, Lirong. Computing the margin of victory for various voting rules. In: *Proceedings of the 13th ACM conference on electronic commerce*. 2012, pp. 982–999.

44. SZUFA, Stanisław. *Mapel* [https://github.com/szufix/mapel]. GitHub, 2020 [visited on 2024-01-01]. Available from: https://github.com/szufix/mapel.