Extracting Trends from Incomplete Ordinal Preferences

MSc Thesis *(Afstudeerscriptie)*

written by

Vahid M. Hashemi
(born September 13th, 1980 in Mashhad, Iran)

under the supervision of **Dr Ulle Endriss**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:
September 20, 2012

Members of the Thesis Committee:
Prof Jan van Eijck
Dr Ulle Endriss
Umberto Grandi
Prof Benedikt Löwe

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

In conventional preference aggregation, usually the result of aggregating multiple preferences is a single preference. Thus, the preference of the majority is seen as the aggregated preference of the whole society. But often, members of a society have many different tastes and these tastes are partly recognizable from their preferences. There are many applications in which having more than one preference as the aggregation of the input preferences is more useful (or necessary). Another aspect of such applications is that the agents may not have a preference ordering consisting of all candidates (this is not actually feasible when there is a very large number of candidates, e.g., all music or movies on a huge database), and they would only be required to submit partial preferences on the set of candidates.

In this thesis we suggest a new approach in computational social choice to have more than one preference as the aggregated preference of the society. We present a model to extract these aggregated preferences (we call them trends) from incomplete ordinal preferences. Furthermore, we introduce a number of axiomatic properties to evaluate this new model and use them to investigate the properties of our proposed methods for this model.

Acknowledgments

I was very lucky that my supervisor was Dr. Ulle Endriss. First he gave me the freedom to work on a topic that I liked. Despite all the inconvenience happened during this work, he did not stop helping. Without his patience and encouragement alongside of his kindness and flexibility, this thesis would not have finished. Ulle is a great supervisor and person. I am very grateful to him.

I would like to thank many friend who supported me, but cannot name them all here. I am grateful to all of them, but I want to especially mention Stéphane, Inés, and Maryam.

And my dear Sara, thanks for everything.

Contents

Chapter 1

Introduction

1.1 Motivation

The aggregating of preferences in a society is one of the main concerns in the social choice context. But in a society with diverse tendencies, it is not very clear what the collective choice is and how we can obtain it.

Nowadays in many applications (especially in the world of the Internet, online voting, and huge amounts of data) there are many instances in which the number of candidates are very high and also the number of voters is large. Ranking web pages, videos, songs, movies, books, ... are some common examples of such voting or ranking systems.

For example, think of an online system for ranking/rating a general database of movies or music. Despite the diversity of both movies or music and people all around the world and all the various types of productions and different tastes of individuals, the classical methods of aggregating preferences would mix all of these and produce a single ranking as the social aggregated preference. If they want to produce more than one ranking, they usually do not really use a different social preference aggregation method, but use a preprocessing phase to do the task. The most common solution of this problem, in classical methods, is categorizing items with respect to their genres, language, geographical and chronological parameters, etc., and then for each of these categories, producing a single ranking. But this would not solve the problem properly. In this approach, first the categories have to be defined and items have to be assigned to them (manually or automatically), and this is not an easy or clear procedure.

When this is done manually, which is still the case for most real-world applications (like attributed genres for movie and music), besides being a manual procedure, there are other problems as well. Assigning an (or a set of) attribute(s) to an entity is not always clear-cut. There are usually marginal cases where the appropriate attribute is disputed. Another problem is that when the set of alternatives (or individuals in some cases) is partitioned, the inter-relations of the different parts are lost.

For automated categorizing, there has been lots of research on clustering and segmentation [15], but the issues mentioned for manual categorizing still remain.

Another related body of research is recommender systems [16, 22, 26]. Specifically, collaborative filtering deals with finding users with similar tastes for each user [25, 12]. The assumption in this approach is that if two agents have a similar preference on a set of alternatives, then they probably also have similar interests on other alternatives. Based on this assumption, the system recommends new items to the user, using the choices of the other matched user(s) [11].

Although, this approach addresses part of the problem we raised in the beginning, there are some significant differences. Recommender systems and collaborative filtering are mostly designed as a user-based system, i.e., they aim to find a (number of) relevant choice(s) for each user based on the previous choices of the user. But in the social choice context, we usually look for collective choices or preferences.

Another issue is that in collaborative filtering, the methods are mostly based on a kind of partitioning (of the users based on their taste, or sometimes of the items, based on a similar set of corresponding users). This would still have the disadvantage we mentioned before, that is, that it only considers similar users (items) for each user (item). This is also the case, when methods do not really partition but use the data of the neighbors of each entity and again do not consider the useful information outside of its vicinity. Another important point is, for example, if there are two users which have the same taste on half of their choices but the rest of their choices are completely different, they are not considered similar in most cases and their strong agreement on half of their choices are totally ignored.

There are also some works on the relation of collaborative filtering and social choice theory [18].

If the number of candidates (alternatives) is large, it is not possible to get the complete preference from each voter. So we will only get a preference on a subset of alternatives for each voter (most likely on the most important ones for him, but not necessarily). This issue is addressed for combinatorial domains [6], but not for the general case.

So, in summary, these kinds of systems cause some particular issues, mainly because of the diversity of the voters' preferences and the high number of alternatives; which is not well explored in the preference aggregation and computational social choice context.

One of these important issues that has not been investigated very well, or not considered at all (as far as we know), in the social choice context is: When there are many voters and many alternatives, it is natural to imagine some groups of voters in which the voters in each group have roughly similar preferences (at least on a part of their preferences) and different from the rest of the society's preferences (group of voters with similar taste).

So maybe it would be more rational and useful to report more than one aggregated preference (maybe with a score for each) as the result, in contrast to the classic methods that just concentrate all the preferences in one social preference.

For this issue, in our model, we have a preference (or scores) on a subset of alternatives for each voter. The subset that each voter chooses to report his opinion on depends on something like:

- the order of the alternatives that are the most important for him (not necessarily the most preferable ones, like in bipolar preferences [2]);
- alternatives he knows or remembers at that time (because of the huge amount of candidates, it is not possible for voters to know or remember or even check all the alternatives);
- there are no clear differences between some alternatives for a voter so he does not report anything about them;
- some limitations have been imposed on the ballots due to voting rules, implementation constraints, communication restrictions, etc . . . .

As the output, we extract a number of aggregated preferences (trends), maybe with a score for each that shows the reliability or popularity of that result. The number of trends in the result can be determined by a constant number (report that many trends with highest scores) or by a constant threshold (report all trends with higher score than the threshold) or just all of the trends with their scores.

We have developed a model that attempts to address these issues. The model could be defined in more general settings, as we explain below, but here we will present it in a setting that is closer to classic preference aggregation systems.

We can assume different models for systems of obtaining trends as a result of aggregating different preferences.

There are different possible ways of expressing preferences by voters and also for aggregation of these preferences as the result. In the first step, there are two general ways for expressing preferences:

- Describing the preference by ordering candidates with respect to each other (Ranked voting systems / Ordinal).
- Assigning each candidate a score that indicates the amount it is favored (Rated voting systems / Cardinal).

It is clear that every cardinal preference profile can easily be converted into an ordinal one, by just ordering the candidates by their scores. In this thesis, as in most conventional preference aggregation methods, we only deal with ordinal preferences (both for the input and output).

Some of the most common variants of ordinary preference aggregation systems, based on the assumptions about expressing input preferences, are:

- All voters report their total order on the complete list of candidates and the result is in the same format, i.e., a total order of all candidates. (This may be considered as the most desirable way of preference aggregating if it would be practically feasible.)

- There is a particular number, say k, $k < m$ (number of candidates), that each voter should report a total order on the exact (at most) k best candidates in his/her view. In this case the rest of the candidates in each reported preference are assumed to be less preferable for that voter and we do not have any more information about the voter's preference for them.

- It is also possible to relax the assumption of reporting the order on the most preferable candidates of each voter, so voters can report their order on the exact (at most) k candidates from any part of their virtually total order of all the candidates (e.g., the $k-1$ best and the single worst candidates). The result usually has the same format (order of k most aggregated preferable candidates) or a total order of all the candidates.

- There is no restriction on the number of candidates that are compared by each voter. This means that each voter can report his/her preference among as many candidates as he/she wishes. Therefore, the amount of data acquired in this case is very dependent on the individual voters' profiles. So it is attempted to obtain as complete an order on the candidates as possible.

The first case, is the most prevailing one in the social choice context. Other than complete preferences, most common variants of the systems which are studied in the field are partial order preferences and incomplete total orders. In most studies, the incomplete total orders are considered to be as the second case, i.e., they assume that the unmentioned alternatives are the least preferable ones and with this assumption they usually build up an order on the complete set of alternatives (this is often a partial order since the rest of the alternatives are supposed to be indifferent among themselves).

In our proposed model, we consider the general case in which the individuals' preferences are total orders on arbitrary subsets of alternatives.

So the system gets an incomplete ordinal preference from each voter and gives a number of incomplete orders of candidates as trends. We call this a *Trend Aggregation Procedure* (TAP).

To simplify the data from the input preferences and decrease the computational complexity of the algorithm, we may convert this data to its atomic parts (e.g., a set of weights for candidates or a set of weighted pairwise comparisons). In the TAPs studied in this paper, we only deal with the conversion to a set of weighted pairwise comparisons, and call this conversion the *decomposition* process. The reverse process, i.e., converting this data into a preference on a subset of candidates, is called a *composition* process.

Although there are some works on aggregation of preferences with partial orders (for example [20]), it seems that the concepts of having more than one aggregated preference as the result and considering the relation of the alternatives from input preferences are new. Finding possible and necessary winners [19] can be seen as one of the closest concepts to our idea, since it proposes more than one winner as the result. But they are only alternative winners and not aggregated preferences. Maybe we can see it as a very special case of our method, i.e., we can design some settings for the model which results in the top candidates of the output trends being possible or necessary winners.

There are also some works on aggregating preferences in multi-issue domains like [8, 14] which are somewhat related to our work since they consider multiple issues (attributes) for the alternatives; but these attributes are fixed and predefined in their model and the model's result is still a single aggregation.

1.2 Toward the Model

Here, when we talk about the "relation" between candidates, we roughly refer to the implicit relation expressed by comparing a set of candidates in a voter's preference. For example, when an individual expresses his preference on a number of alternatives, we can assume that the alternatives have one or more aspects in common that enabled the voter to compare them with each other. This is also the case when we put a subset of alternatives in an output preference (trend).

In conventional preference aggregation, the goal is to find an order for the set of inputs which are preferences of different individuals that are usually uniform (e.g., a total order on the set of candidates, as assumed in the seminal work of Arrow [1]).

Also, relations between candidates in the view of the voters are not taken into account in conventional methods of preference aggregation. This may come from the assumption that each voter reports a complete order, so the relation does not have a meaning in such a case; or when we want to obtain just one total preference, then the relation is not very important. These facts make it easy to work with the input preferences, since we do not need to care about the relation between the candidates, beyond pairwise orders, in each reported preference. Thus, in most cases, we can just decompose each preference (a total order) to its primitive pairwise orders and aggregate these pairwise orders using the desired method.

But when we have a large number of both candidates and voters with very different properties and attitudes, the concept of "relation" becomes crucial, especially when the number of candidates is so large that no voter knows all the candidates or wants/is able to compare all the candidates with each other. This will be more important when we want to find trends in the preferences of voters and report more than one aggregated preference as the result. Therefore we cannot use previous methods anymore, i.e., we can not ignore the relations and decompose input preferences freely. On the other hand, the order of candidates is important as well, so we cannot use partitioning and clustering methods. Therefor one of the main concerns here is to find a way for aggregating preferences while considering their relations.

In our model, we want to obtain a number of trends rather than one social preference. So in addition to specifying a number of orders on the set of alternatives, it is important for them to have a kind of reasonable relation. To explain this, first notice that when we have, for example, an order on the set of $\{A, B, C\}$ like $A \succ B \succ C$, it, in fact, means that $A \succ B$, $B \succ C$ and $A \succ C$. But in our model these two different ways of expressing the order between these alternatives have an important difference. In the first one, in addition to specifying an order on them, it determines that there is some sort of relation between these three alternatives. However in the second case where there are three independent comparisons between the three alternatives, they can potentially come from three different voters. Thus, despite the fact that these three together are equivalent to the first order from the ordering aspect, we can not consider them as related alternatives as we can in the first case. So we try to distinguish these two different expressions in our method.

Let us use an example to explain our idea and give a better understanding of the kinds of problems we are facing here.

Example 1.1. *Consider a system with 15 alternatives (for example, movies) $a, b, c, d, e, f, g, h, q, r, s, t, x, y, z$; and 10 voters $v_1, ..., v_{10}$. Let the voters' ballots be as follows:*

$v_1 : a \succ b \succ c \succ d \succ h,$
$v_2 : a \succ g \succ c \succ d,$
$v_3 : b \succ c \succ d \succ e \succ f,$
$v_4 : c \succ a \succ d \succ g \succ f,$
$v_5 : f \succ g \succ h,$
$v_6 : q \succ r \succ s \succ t,$
$v_7 : r \succ q \succ t,$
$v_8 : t \succ s \succ q \succ r,$
$v_9 : x \succ y \succ z,$
$v_{10} : a \succ q \succ x,$

An interpretation of this preference profile might be as follows:

There is a restriction on the maximum number of alternatives that each candidate can report its order on; this number is 5. v_1 knows and remembers all the candidates. It prefers a the most and then b, c and d are the most preferred alternatives and h is the worst alternative; e, f, g are less important and the rest are totally indifferent for this voter. So it prefers to report h at the end of its order rather than one of e, f or g.

v_2 just knows a, g, c, d and has a clear preference on them as reported. The rest have a mixture of these reasons to report their preferences on the set of alternatives as they did.

Another observation on this preference profile is that we can consider some sort of relations on the members of these three sets of alternatives: $A = \{a, b, c, d, e, f, g, h\}$, $B = \{q, r, s, t\}$ and $C = \{x, y, z\}$. We can see that except for v_{10}, all other voters' preferences are in a subset of one of these sets. We can interpret this as alternatives having three different general genres: A, B, C. Each voter $v_1, ..., v_9$ is a fan of one of these genres or has comparisons just on members of one genre; and voter v_{10} compares one alternative from each genre (e.g., the best of each genre for it).

Now to get a better understanding of our proposed problem, assume these different situations:

- *There are 100 voters exactly like each of $v_1, ..., v_5$ and 10 for each of the rest of the types:*

  In this situation almost all conventional methods of aggregating preferences (which have a single aggregated preference as the result), will result in a preference in which all the alternatives of genre A are on the top. It would be more problematic if the method only reports preferences on a subset of alternatives that it considers as the best. In this case, the result will probably consist of a subset of alternatives in genre A, and there is no information about the other genres, even their best candidates.

  So the result does not represent a part of society (fans of genres B and C) or is useless for them.

- *There are 100 voters of each type except for v_{10} that has just 10 voters:*

  This case will result in a social preference which is a mixture of the different genres. So it is somehow useless again since it has mixed the alternatives of different genres and the genres are not recognizable anymore.

- *There are 100 voters of type v_{10} and 10 for each of the rest of the types.*

  Here we will have the same problem and the aggregated preference will consist of just the best of each genre.

1.3 Intuition

Here we will present a few simple situations to illustrate the concept and give some insight into the problem and then use them to intuitively obtain a number of properties that should hold in our method.

For each aggregated preference (trend), we can assume a value which represents one or a mixture of the concepts of the strength, popularity or reliability of that trend as an aggregated preference. We denote this value (score) of a trend P by $v(P)$.

When we compare two scenarios of voting with two sets of voters' preferences, the desired values for some of the trends are compared. Note that, theoretically, any complete order of any subset of candidates can be seen as an aggregated preference (even with a value of zero, though we usually do not consider them as aggregated preferences). Here, in most cases, we do not mind which ones should be reported as aggregated preferences for the result, and just compare the values of some specific aggregated preferences.

In a very specific case in which the set of individuals consists of groups of voters with the exact same preferences, the concept of a trend is clearer. In particular, if these preferences are on the disjoint subsets of alternatives with the same size, there should not be any debate on the selection of trends (e.g., a society in which each individual's preference is one of the three types of v_5, v_7 or v_9 in the previous example). The preferences with different sizes might cause a dispute about considering priority for preferences with greater length. But with this assumption, the only parameter that makes a priority for a preference would be the size of the group of voters which the preference represents.

Any other situation can raise difficulty, from both conceptual and computational aspects. As we mentioned, even if there are disjoint preferences with different lengths, the question is do we want to give advantage to longer preferences, and if yes, in what sense? For example, if we have 32 voters of type v_1, 35 voter of type v_7 and 33 of type v_9, is a trend of type v_1 is a better representative of the community or v_7? That is, we have $v(v_7) \geq v(v_9)$. Now, where should we put $v(v_1)$ in this inequality?

If we drop the restriction of disjoint preferences, not only does the concept of a trend get complicated, but we will have difficulties in the computational part of the problem. In the previous cases, the process was easy. There were a few disjoint preferences and we only needed to count the number of occurrences of each preference as an individuals' votes. But if this is not the case, that is, the diversity of the preferences are high and so the individuals' preferences are not limited to a few preferences, it is almost impossible or meaningless to count identical preferences.

So, to be able to aggregate data from different preferences, we need to decompose the preferences into smaller entities that make it possible to do calculations. In general, in conventional methods, there are three main approaches for doing calculations on the preferences (we will see instances for each of these cases in the next chapter):

- Assign a score to each alternative in each preference and then do the calculation on these scores. This is usually called positional scoring.

- Compute the results of the pairwise elections for each pair of alternatives from the preferences; then use these results for the rest of the calculations.

- Define a sort of distance between two preferences and then do the calculations based on this distance and using the complete preferences.

In general, methods of the first type (positional scoring methods) are computationally easier and faster than the other methods. But they usually have drawbacks with regards to theoretical properties. This is due to the fact that these methods ignore the relative relations of the alternatives and project the information of each preference for each candidate to only one score.

The second type of methods, the most common ones, consist of a variety of different methods. Depending on how the data from the pairwise elections are being used, these methods can be computationally easy or hard. These methods preserve more information from each preference than the positional scoring methods. But, while they hold the relative preference of each pair, they ignore the position of the alternatives (and also relative position for each pair) in the preferences, which is partly considered in the positional scoring methods.

The third class of methods preserve all the information of the preferences since they do not transform the preferences to other forms of data. Although this is a great advantage, its drawback is that it makes these methods computationally difficult. Another problem, which is the case for most of the methods of all of these three classes, but is more serious for the third type, is that these methods strongly rely on uniform (complete) orders. That is, they are hardly applicable for preferences with different lengths.

Thus, as a result, we need to decompose preferences to be able to perform calculations (specially when the preferences are incomplete orders). This is also crucial from the computational complexity point of view. But, at the same time, we want to preserve information as much as possible so we can preserve a trace of the relations between the alternatives.

Now, we explain our idea on a couple of basic situations. First, let us take another look at the case of a preference on the set of $\{A, B, C\}$. As we said, we want to have a distinction between a preference $p : A \succ B \succ C$, and the three preferences $q_1 : A \succ B$, $q_2 : B \succ C$, and $q_3 : A \succ C$ together. We show the score of aggregated preferences in a society with only one input of p with v_p, and for the case of three preferences q_1, q_2 and q_3 we use v_q. From what we said about the concepts of trends and relations, we have:

$$v_p(A \succ B \succ C) \geq v_q(A \succ B \succ C).$$

This is because of the assumption that, in our model, these two different ways of expressing the order between these alternatives have an important difference. Also, since the alternatives in the first case have more correlation, the preference would have more support for a trend on them. On the other hand, we do not want to prevent the three independent preferences from building a trend when they posses all the parts of the trend; which is in fact the main point of our model: to gather the scattered pieces of information among the different input preferences. So, we only want to assign a lower score in the second case.

We use the following observation to make this distinction. In the first case, since the three alternatives are related (occur in the same preference), we have transitivity; i.e., from $A \succ B$ and $B \succ C$, we can imply that $A \succ C$. The implicit assumption in transitivity is, when one prefers A to B and B to C, it also prefers A to C at least as much as it preferred A to B (or B to C). This is an important piece of information which is lost in the second case, in which all the three pairs have the same status. Considering this fact, it is more plausible that A is more preferable than C for the voter of the first case rather than the one in the second case who said $A \succ C$. Note that it is not clear that A is less preferable than C for the second voter just because there is no other candidate in its preference profile and it could be even the opposite. But the point is, from these data, it is more plausible to be so. That is:

$$v_p(A \succ C) \geq v_q(A \succ C)$$

The arguable question in our model for this example is: what can we say about the score of $A \succ B$ in each of these two scenarios and their comparison? On one hand we can say that since the voter in the first scenario has specified its preference on more candidates, its ordering is more reliable; but on the other hand, one may say that since the voter in the second scenario only compares A and B, it means that just these two candidates were important for it and so it is more precise (or they could be potentially the best and worst choices for that voter, while in the first case there is at least one more alternative worse than B).

So the question of choosing between these interpretations mostly depends on our intuition and the properties and applications of our problem in practice, unless one of these entails some unwanted properties or has a conflict with other primitive properties. For now, we will not consider neither of these interpretations and treat them equally. That is, in this example we have:

$$v_p(A \succ B) = v_q(A \succ B)$$

On the basis of these observations, we propose our method. First, like in many conventional preference aggregation methods, we decompose the input preferences into pairwise comparisons. However, unlike for conventional methods, in the process of decomposition we also consider the relative position of the two alternatives. That is, instead of assigning a constant value to each occurrence of a pair in a preference, we also consider the distance of the two alternatives in each input preference and assign a higher value to the more distant pairs. This value, we can say, represents the “amount” or “intensity” of a preference for those two alternatives in the view of the voter.

Note that, e.g., in the preference $A \succ B \succ C$, it is reasonable to say that A is preferred over C more than it is preferred over B, and also more than B is preferred over C. However, it is not always true to say that A is preferred to B exactly twice as much as the other two preferences.

From this viewpoint, we can also think that it is plausible for A to be more preferable to C in $A \succ B \succ C$ than when the reported preference is $A \succ C \succ B$, but not necessarily.

Hence, we do not necessarily use the distance of a pair itself as the value to assign to the pair; but use a non-decreasing function on this distance, which can be a better approximation for this concept of intensity of the preference (the distance itself is a special case).

At this stage, we can use these values to aggregate the information of all input preferences. For each ordered pair (i.e., one for $A \succ B$ and one for $B \succ A$), we add up the values of their pairwise comparisons from all input preferences.

Now the problem is how to combine these fragments of data in a way to obtain relevant total orders on the subsets of candidates, i.e., aggregated preferences. The main difference from the classical preference aggregation methods which decompose the input preferences is that here we want to obtain more than one collective preference and these collective preferences could be on a subset of the alternatives rather than on all of them. So, the classical methods could hardly work here. In this phase, we merge pairwise orders to get chains of ordered preferences while trying to retrieve the relations of the candidates that we preserved in the decomposition phase.

To do this, let us take a look at the simple case of $A \succ B \succ C$ again. Since we decomposed it into the three pairwise comparisons $A \succ B$, $B \succ C$ and $A \succ C$, it is natural to do the reverse procedure and merge these three pairs into a total order on them. Notice that, here we do not have transitivity, that is, from $A \succ B$ and $B \succ C$ we can not conclude that $A \succ C$; although it may be the case for each individual's preference. So we may not have as much supportive data for $A \succ C$ as we have for $A \succ B$ and $B \succ C$, and also, in most cases, we do not have the same supportive data for the last two.

Therefore, for these three candidates, and more particularly for these three pairwise comparisons, we have three independent values coming from the summation of their corresponding weights. Let us call these three values a, b, and c corresponding to $A \succ B$, $B \succ C$ and $A \succ C$ respectively. Now we want to assign a value to the aggregated preference of $A \succ B \succ C$, based on these values. The most confident thing that we can do is to consider each of these three with the value of $d = min\{a, b, c\}$ instead of their own value. Then we have three segments of the aggregated preference (trend) of $A \succ B \succ C$ with the same supportive value and so we can say that there is such a trend in the preferences of the voters and assign this value (or a corresponding value) to this trend.

We can also think of other alternatives for assigning a value to such a trend. For example, maybe the summation or multiplication of the values of the basic pairwise comparisons of each trend could be a good alternative in some cases. In addition, since we will have different trends with different lengths in our model, perhaps considering the length of the trends in computing the value can be beneficial or even necessary in these two cases. So we may use the average of these values (arithmetic or geometric mean) as the score of each trend.

Furthermore, since we use the distances of pairwise comparisons in the decomposition phase, it is more reasonable to combine them using the reverse of the same procedure we used to decompose them. For example, consider again $A \succ B$, $B \succ C$, and $A \succ C$, with the corresponding values a, b, and c, respectively. We should divide c by the value that we would use if we wanted to assign a score to the pairwise comparison of $A \succ C$ in preference $A \succ B \succ C$. So, for each possible trend, we use (the minimum or average of) these values (for all pairwise comparisons of that trend) to assign a score to that trend. We report all of these trends with their corresponding scores or select a number of them to represent the most important trends in the society.

1.3.1 Chapter overview

We will discuss the details of this method in this thesis. In this thesis, our focus is on the theoretical aspect of this problem. So, we only define trends and methods for assigning scores to them formally; and we will not present a particular algorithm for producing the trends.

In the next chapter, we will have a short review of classic social choice theory. We focus mostly on the different methods of voting and preference aggregation to see what the common methods in these fields are. We also present some of the most important axiomatic properties (criteria) which are introduced to evaluate different aggregation methods. We will generalize some of these axioms later to evaluate our proposed model.

In Chapter 3, we formally introduce our model. First, we present a general framework for trend aggregation procedures. Then we explain our method formally. We will also present three more methods which satisfy a particular property. Finally, we suggest a number of filters to refine the huge number of the output trends of the model.

In Chapter 4, we investigate the properties of the model. We present a number of axiomatic properties, which are either the generalization of classic axioms in our framework or totally new axioms based on the new aspects of this framework. We choose a number of major variants of the methods that we proposed, and for each of the properties, we will check in which methods the properties hold and in which methods they do not.

In the last chapter, we summarize what we have done. We will also have a discussion of our results. As we said, we will not present any algorithm in this thesis; so we will not have any result about the computational complexity of the problem. However, we will have a brief discussion about the size of the problem and the potential complexity of any trend-producing algorithm. At the end, we will present a number of suggestions for possible future work.

Chapter 2

Voting and Preference Aggregation Procedures

In this chapter we will introduce (classic) voting and preference aggregation procedures. It will consists of their theoretical aspects such as their properties. We will particularly use some of these properties later and try to investigate whether they or their modified versions hold in our model.

Most of the content of this chapter can be found in most introductory books on social choice and voting theory. A few references are [3, 4, 5].

2.1 Voting Systems

A *voting system* consists of a set of individuals or *voters*, a set of *alternatives* or *candidates* and a *voting procedure* which describes how to vote and how to count the votes to aggregate the individuals' preferences and obtain a *collective decision* or *social choice*.

Voters might be people of a society or virtual agents such as websites in the World Wide Web.

Candidates can also represent different entities, e.g., the set of candidates could be a subset of the voters (ordinary elections), or a disjoint set of individuals of the same type or of a completely different type (e.g., a referendum).

In some contexts, "voting systems" may refer to a general set of systems that also contains preference aggregation. In the context of voting theory, however, we usually consider a distinction between systems that only aim to choose one or more candidates, which are called *voting systems*, and *preference aggregation systems*, which look for rank of all or a subset of the candidates. The goal of a voting system might be to elect one or more candidates.

Considering the above, there are two remaining questions: "How to vote?" and "How to count the votes?". A voting procedure defines the format of an acceptable vote or *ballot* that are called *admissible ballots* or *valid ballots*; and a method for *aggregating* these ballots to obtain a winner or the winners.

2.1.1 Notation and Basic Definitions

Here we only deal with finite sets of voters and candidates. Let $\mathcal{V} = \{v_1, ..., v_n\}$ be the set of voters and $\mathcal{X} = \{c_1, ..., c_m\}$ the set of alternatives (candidates). We may refer to voters just by their index, i.e., i for voter v_i. In this case the set of voters will be denoted by $\mathcal{N} = \{1, 2, ..., n\}$.

Each voter can cast a valid ballot. The set of valid ballots is determined by the voting procedure. If the ballot of voter i is d_i, for all voters $i = 1, ..., n$, the n-tuple $d = (d_1, d_2, ..., d_n)$ is the *ballot response profile.*

Assume each voter i has a preference that is a *weak order* $\succsim_i$ on $\mathcal{X}$; $\succsim_i$ is *transitive* and *complete*, i.e.:

- $\forall x, y, z \in \mathcal{X} : x \succsim_i y \ \wedge \ y \succsim_i z \Rightarrow x \succsim_i z$ (Transitive)
- $\forall x, y \in \mathcal{X} : x \succsim_i y \ \vee \ y \succsim_i x$ (Complete)

$x \succsim_i y$ means voter i prefers x to y or is indifferent between these two. If i prefers x to y, it is denoted by $x \succ_i y$ and called a *strict preference.* Indifference between x and y for voter i is denoted by $x \sim_i y$.

Since the indifference relation is reflexive, symmetric, and transitive, it is an equivalence relation. So $\sim_i$ partitions $\mathcal{X}$ into r indifference classes $\mathcal{X}_1, ..., \mathcal{X}_r$. For each x and y in different classes, exactly one of these is true: $x \succ_i y$ or $y \succ_i x$.

If for all $x, y \in \mathcal{X}$: $x \sim_i y \Leftrightarrow x = y$, then $\succsim_i$ is equivalent to $\succ_i$ and is called a *linear order* or *strict ranking* and $x_1 \succ_i x_2 \succ_i ... \succ_i x_m$ is abbreviated as $x_1 x_2 ... x_m$.

In most cases the assumption is that the voters' preferences are a linear order. From now on, we assume that the preferences of all voters are linear orders of all the candidates and the voters cast the exact same order as their ballots. So, we use the terms of preference and ballot for a voter interchangeably.

We denote the set of strict linear orders on set of $\mathcal{X}$ as:

$$\mathcal{L}(\mathcal{X}) \ := \ \{\succ | \succ \text{ is a strict linear order on } \mathcal{X}\}$$

Since all the preferences are strict, there is no more indifference relations and so we only use the strict preference relations ($\succ$).

The set of preferences of all voters (an n-tuple $v = (\succ_1, \succ_2, ..., \succ_n)$) of linear orders on $\mathcal{X}$, one for each voter) is a *voter preference profile.*

The set of all possible ballot response profiles for the set of voters $\mathcal{X}$ is $\mathcal{L}(\mathcal{X})^n$, and each of its members is a ballot response profile $d = (d_1, d_2, ..., d_n)$, $d_i \in \mathcal{L}(\mathcal{X})$ for each voter.

If i's ballot (d_i) prefers x to y, it is denoted by xd_iy. If xd_iy for all d_i in $d = (d_1, ..., d_n) \in \mathcal{L}(\mathcal{X})^n$ (i.e., x is preferred to y in all ballots); we say that x *dominates* y, or xdy.

Let $d = (d_1, ..., d_i, ..., d_n)$ and d'_i be an admissible ballot; then $(d_{-i}, d'_i) = (d_1, ..., d'_i, ..., d_n)$ is a ballot response profile, obtained by replacing d_i by d'_i in d.

Let d_i and d'_i be two admissible ballots for a voter i, and let x be an arbitrary candidate. We say that x has got advantage or is raised in voter i's ballot (d'_i) compared to d_i and denote it as $d'_i >_x d_i$ (d'_i is better for x than d_i), if the position of x in d'_i is better (higher) than in d_i.

If $d = (d_1, ..., d_n)$ and $d' = (d'_1, ..., d'_n)$ are two ballot response profiles, and $d'_i >_x d_i$ for all $i = 1, ..., n$, we write $d' >_x d$.

If $d = (d_1, ..., d_n)$ is a ballot response profile and $A \subseteq \mathcal{X}$ is a subset of voters, we use $d|_A$ for the ballot response profile of voters in A. So if $|A| = j$, $d|_A$ is a j-tuple.

2.1.2 Social Choice Functions

A *voting rule* or *social choice function* (SCF) is a function that for each input profile produces an outcome as the winner(s) of the election or collective choice of the society. To be precise, in the literature there is often a distinction between voting rules and social choice functions. That is, in SCFs an extra set is defined as the feasible set of alternatives for the possible winners of the procedure. In this respect, voting rules are special cases of SCFs.

Here we restrict ourselves to the voting rule definition. The aim of an election can either be finding a single candidate (the winner) or a set of k candidates. In both cases, the procedure may calculate the (set of) winners directly or first produce a ranking on the alternatives and then select the top (k) candidate(s) as the winner(s).

Although for most voting procedures there exist tie and a complete voting system should have a solution for such situations, here we ignore the tie-breaking issue and assume that the procedures only produce a set of exact number of candidates as desired. Such a procedure is called *resolute* or *decisive*, and procedures with ties are called irresolute.

Definition 2.1 (Social choice function)**.** *Let n be the number of voters and $\mathcal{X}$ the set of candidates. A* social choice function *is a mapping from the set of all possible profiles (linear orders) to the family of non-empty subsets of $\mathcal{X}$,i.e.,*

$$F : \mathcal{L}(\mathcal{X})^n \to 2^{\mathcal{X}}/\{\emptyset\}.$$

Sometimes, in the literature, a distinction considered between the terms "voting procedure" and "voting rule", in which voting rules are specifically the resolute voting procedures with a single winner. In this sense, the term *voting correspondence* is used for the irresolute procedures. However, we may use the terms voting rule and voting procedure interchangeably in the absence of tie.

When k is the number of desired winners of the procedure, the function F is called a *choose-k voting procedure.* For example procedures which aim to elect one candidate are choose-1 rules.

Definition 2.2 (Choose-k voting procedure)**.** *A choose-k voting procedure is a social choice function in which the outcome is a set of k winning alternatives, that is,*

$$F : \mathcal{L}(\mathcal{X})^n \to \{X \subseteq \mathcal{X} : |X| = k\}.$$

Here, since we ignore ties, we defined F to have exactly k members (rather than $\geq k$ in the more general case). In case of dealing with tie-breaking, we may let F be such that $|F(d)| \geq k$. In particular, the formal definition of choose-1 voting procedures (voting rules) is as follows:

Definition 2.3 (Voting rule)**.** *A choose-1 voting rule is a social choice function in which the outcome is a single winning alternative, that is,*

$$F : \mathcal{L}(\mathcal{X})^n \to \mathcal{X}.$$

2.2 Voting Procedures

In this section, we introduce a few common voting procedures. In general, voting procedures can be categorized based on the type of their voters' ballots. In this regard, there are two main types of voting procedures. The *nonranked voting* systems and the *ranked voting* systems. In nonranked voting systems, voters express their opinions by only specifying which candidates they vote for. In this sense, each voter can indicate two levels of distinction among the alternatives, namely, voted and unvoted alternatives. In ranked voting systems, voters are capable of expressing their preferences with more than two levels of support for the candidates. The ballots of these types are usually linear orders (strict rankings) on a subset of the alternatives (in a more general setting, the ballots could be partial orders). The ballots can be restricted by a specific number as the maximum length of ballot (number of alternatives in the input rankings) or the maximum number of levels of distinctions a vote can express.

Another major type of ballots are scoring ballots. In scoring ballots, the voters can express their preferences by scores they assign to the alternatives. Strict rankings can be considered as a special case of scoring ballots.

However, as we have fixed it before, here we consider that all the ballots are linear orders on the set of candidates.

Although, as mentioned before, for almost all of the coming procedures there exist tie and a complete voting system should have a solution for such situations, here we ignore the tie-breaking issue.

Also, notice that some of these procedures may not completely conform with the formal definition of the voting rules we presented before or the assumption we have about the preferences of the voters.

Most of these procedures have been designed to elect one candidate, but many of them also work for multi-winner elections.

2.2.1 Plurality Rule

The simplest and most obvious type of social choice function is the *Plurality rule*, in which each voter submits the name of only one of the candidates and the winner is the candidate receiving most votes.

The number of votes each candidate receives (number of votes with the candidate on the top), is called the *plurality score*. This is the most common method for elections in practice. For elections with only two candidates, it is the unique reasonable and flawless procedure. Although it is an appropriate procedure for two candidates, in general cases it receives some criticism:

- It completely ignores the preferences on all candidates behind the most favorite one for each voter.

- The fact that the voters have no way to express their opinion for more than one candidate causes the votes to be divided between similar candidates.

- This can also causes a voter to vote for another candidate, rather than its most preferred candidate, when it thinks its favorite candidate has little chance to win.

2.2.2 Majority rule

The *majority rule* can be considered as a special case of the plurality rule. In the majority rule, the winner is a candidate who receives a majority of votes. When the needed majority is "more than half of the voters", it is called simple majority. If blank votes are not allowed, the majority rule and the plurality rule are equivalent for the case of two candidates.

2.2.3 Plurality with Run-Off

This is similar to the plurality rule, but if there is no candidate receiving at least a specific ratio of the total votes (usually half of the votes), the winner will be elected in a second round. The two candidates with most votes in the first round will go to the second round and the winner will be chosen by plurality (majority) rule between them.

2.2.4 Anti-plurality Rule

The *anti-plurality rule* is like the plurality rule, with the difference that each voter, instead of voting only for the most desirable candidate, votes for all candidates except the least desirable one. This method is also called *veto rule*.

2.2.5 Approval Voting

In *approval voting* each voter can vote for as many candidates as it wants, and the winner is the candidate who gets the most votes. So a ballot may consist of any subset of $\mathcal{X}$ with size $0, 1, \ldots$ or $m-1$ (size m, i.e., voting for $\mathcal{X}$, is practically like voting for none). In other words, each voter indicates the candidates it approves and the candidate who is approved by the most voters will be elected.

Approval voting is theoretically popular and extensively analyzed due to its simplicity and properties, and is fairly popular in practice, specially in some professional societies, such as the American Mathematical Society (AMS), for electing their committees.

A couple of important points about approval voting:

- In contrast to plurality voting, there is no need to strategically not vote for the favorite candidate even when it has a very low chance of winning.

- Despite its simplicity (in terms of both understanding the rules for voters and also communicating and expressing them), it has the capacity of getting more information from voters than procedures like plurality voting (of course not the complete preference ordering).

2.2.6 k-Approval Rule

The special case of approval voting (or the extension of the plurality rule) to elect one (or k) candidate(s) is called *k-Approval*: each voter votes for exactly (at most) k candidates and the candidate receiving the most votes is the winner.

2.2.7 Borda Rule

The Borda rule was proposed by Jean-Charles de Borda in 1770 [9]. It receives the full preferences of the voters, so each voter submits a complete ranking of all m candidates. Then, for each ranked list of voters, the Borda rule assigns points $0, 1, \ldots, m-1$ to the m candidates in each ranking, respectively from the last member of the list to the first one. So the most preferred candidate of each voter receives $m-1$ from that ballot and the least preferred one gets 0. The sum of these points from all voters' ballots for each candidate is called its *Borda count*, and the candidate with the highest Borda count wins.

2.2.8 Positional Scoring Rules

A *positional scoring rule* is defined by a *scoring vector* $s = \langle s_1, ..., s_m \rangle$ with $s_1 \geq s_2 \geq ... \geq s_m$. Each voter's ballot is a full ranking of all candidates. Candidates receive points according to their position in each ranking. For each voter's ballot, the candidate in the ith position gets s_i points; and the sum of these points for each candidate is its score. The candidate with the highest score wins.

Positional scoring is the natural generalization of the Borda rule. The Borda rule is a positional scoring with $s_i = m - i$ (i.e., scoring vector of $\langle m-1, m-2, ..., 1, 0 \rangle$).

Also, the plurality rule can be seen as a very special case of the positional scoring rule where all scores are 0 except for the first one, and $s_1 = 1$ (i.e., scoring vector: $\langle 1, 0, ..., 0 \rangle$).

Likewise, the veto rule is a positional scoring rule such that all scores are 1 except for the last one, which has score 0 (i.e., scoring vector: $\langle 1, ..., 1, 0 \rangle$).

2.2.9 Single Transferable vote

The *Single Transferable vote* (STV) is a voting rule that works as follows: each voter has a preference on the candidates, and there is a quota attributed to the election. In each round, the plurality scores of the remaining candidates are calculated. If any of the candidates reaches the quota, that candidate is the winner. Otherwise, the candidate with the least plurality score is eliminated and the process is repeated. In the next round, the votes of this eliminated candidate transfer to the remainder of the candidates.

If there is no quota attributed to a single-winner election, for m voters, the rule consists of $m - 1$ rounds and the winner is the only remaining candidate in the last round.

Usually, STV is considered as a voting rule for multi-winner elections; and the special case of single-winner elections is called *instant-runoff voting* (IRV) or *alternative vote* (AV).

2.2.10 Condorcet Method

A *Condorcet candidate* (*Condorcet winner*) of an election is the candidate who wins every pairwise majority comparison against the other candidates [7]. Such a candidate does not always exist. When an election method selects the Condorcet candidate whenever it exists, the method is called a *Condorcet method*. So, Condorcet methods are a class of different voting rules, a few of which we will explain separately.

2.2.11 Copeland Rule

In the *Copeland rule*, for each candidate the number of wins and loses in pairwise majority contests is calculated and the difference of these two numbers for each candidate is considered as its score, which is called the *Copeland score*. The candidate with the highest score is the winner.

This voting procedure does not necessarily need the complete preference ordering of voters. For example, the input ballots could be an ordering on a subset of candidates for each voter.

Note that if there is no tie in pairwise comparisons, then the candidate with the highest score always has the highest number of wins. But when there is a tie in pairwise comparisons, the candidate with the highest number of wins might not be the winner.

It is easy to see that the Copeland rule elects the Condorcet candidate when it exists.

2.2.12 Minimax Rule

The *minimax* rule which is also known as the *Simpson-Kramer method* is a Condorcet method. In this method, like in the Copeland method, pairwise elections are considered. But, unlike the Copeland method, which only counts the win/lose results of the pairwise elections, here the magnitude of the win/lose are also taken into the account. That is, for each pairwise election, the difference of the votes in favor of one candidate and the votes against that candidate are counted. Now, for each candidate, the score of the worst result is considered and the winner is the candidate with the minimum value for this score.

2.2.13 Bucklin Rule

The *Bucklin rule* can be considered as an extension of the plurality rule. First, like plurality, the number of votes each candidate receives from the first choice of voters is counted. If any candidate receives the majority of votes, the candidate is the winner. Otherwise, the procedure continues by counting two best choices of each voter as their votes. Now, if any candidate gets the majority of votes (i.e., it is the first or second choice for more than half of the voters), that candidate wins. If not, the procedure continues with the top three choices, and so on.

In any level, if more than one candidate pass the threshold, usually the candidate with highest votes in that round is considered as the winner.

2.2.14 Dodgson Rule

In 1876 [10], Dodgson (better known as Lewis Carroll) proposed a ranked voting rule which is a Condorcet method. The main idea of *Dodgson's method* is finding the candidate which needs the least changes in the votes to become the Condorcet candidate. The number of needed changes in this method is calculated as follows (note that as it is a ranked voting rule, the assumption is that the ballots are in the form of a linear order). For each candidate, the minimum number of exchanges of adjacent candidates in the voters' ballots that makes the candidate a Condorcet winner is considered as the distance of that candidate from being the Condorcet winner. The candidate with the lowest distance in this regard is the winner of the Dodgson rule.

2.2.15 Kemeny Rule

The *Kemeny rule* [13] was originally designed for preference aggregation, but it can also be used as a voting rule. The Kemeny method, as a preference aggregation rule, finds a collective preference for a set of preferences. The method defines a distance (the *Kemeny distance*) between two preferences (orderings) as the number of pairs where their relative ranking in two orderings are different. With this definition of the distance between two preferences, the Kemeny distance of an ordering and the profile of n preferences is defined as the sum of the distance of the ordering from each voter's preference in the profile. Now, the method chooses the ranking, among all possible rankings, which has the minimum distance from the profile. This ranking is the collective preference of the Kemeny method, which is called the *Kemeny consensus*.

When the Kemeny method is used as a voting rule, the top candidate of a Kemeny consensus is considered as a winner of the election and is called the Kemeny winner.

2.2.16 Young Rule

The main idea of *Young's rule* [28] is similar to the Dodgson rule, but, instead of the number of exchanges, the distance of a candidate from being the Condorcet winner is defined as the minimum number of voters whose removal makes the candidate a Condorcet winner and the candidate with the lowest distance is the winner.

2.3 Axiomatic Properties

In general, there is no "complete" or "proper" voting system. In fact, for more than two candidates, there are many cases where it is not clear what should be selected as the social choice. In fact, Arrow showed [1] that there is no voting rule for more than two candidates which satisfies a few simple intuitive criteria. As a result, it is usually not easy to say that voting system A is better than B. Also, the appropriateness of a voting system depends on the context and area it will be applied in, so different voting systems may be suitable for different applications. So, in order to be able to compare or evaluate different voting systems, many intuitive or useful properties have been proposed. Here we will review some of these criteria.

2.3.1 Anonymity

A voting procedure is *anonymous* if all voters are treated equally and the names of the voters do not matter, i.e., exchanging the ballots of any two voters does not affect the election's outcome.

Definition 2.4. *Let ρ be a permutation on the set $\{1, ..., n\}$. If $d = (d_1, ..., d_n)$ is in $\mathcal{L}(\mathcal{X})^n$, and ρd, defined as $\rho d = (d_{\rho(1)}, ..., d_{\rho(n)})$, then ρd is also in $\mathcal{L}(\mathcal{X})^n$. F is anonymous if for all $d \in \mathcal{L}(\mathcal{X})^n$ and all permutations ρ on $\{1, ..., n\}$:*

$$F(\rho d) = F(d).$$

Most voting procedures that we will deal with are anonymous (with the no tie-breaking assumption mentioned before). Examples of voting procedures failing anonymity are when there is a specific voter (e.g., a chairperson) who is the tie-breaker or if the voters' ballots have different weights.

2.3.2 Neutrality

A voting procedure is *neutral* if all candidates are treated equally and the names of the candidates do not matter; i.e., exchanging the name of any two candidates x and y in the ballots of all voters changes the outcome of the election accordingly (If x was elected at first, now y should be elected, and vice versa; and if some other candidate different from x and y was elected, it should still be elected).

Definition 2.5. *Let σ be a permutation on the set of candidates. If for $d = (d_1, ..., d_n) \in \mathcal{L}(\mathcal{X})$, $\sigma(d)$ is defined as $\sigma(d) = (\sigma(d_1), ..., \sigma(d_n))$, and $\sigma(Y) = \{\sigma(y) : y \in Y\}$ for every $Y \subseteq \mathcal{X}$; then $\sigma(d) \in \mathcal{L}(\mathcal{X})^n$ and $\sigma(Y) \subseteq \mathcal{X}$. F is neutral if for all $d \in \mathcal{L}(\mathcal{X})^n$ and all permutations σ on $\mathcal{X}$:*

$$F(\sigma(d)) = \sigma(F(d)).$$

Like anonymity, neutrality holds for most voting procedures we will discuss. Examples of failing neutrality is when in the case of tie, the incumbent candidate or the first candidate in some deterministic order (e.g., in lexicographic order) get advantage; or even more, if defeating the incumbent candidate needs more votes than the incumbent needs itself.

2.3.3 Pareto Optimality

In the simple case of choose-1 voting procedures, the *Pareto principle* can be defined as follows: A voting procedure satisfies the (weak) Pareto principle, if for each two candidates x and y, if x is preferred over y unanimously, then y cannot win the election.

In general a voting procedure is *Pareto optimal* or satisfies the Pareto principle, if when a dominated candidate is in the choice set, then every candidate that dominates it is also in the choice set.

Definition 2.6. *F is Pareto optimal, if for all $x, y \in \mathcal{X}$ and $d \in \mathcal{L}(\mathcal{X})^n$:*

$$xdy \ \wedge \ y \in F(d) \ \Rightarrow \ x \in F(d).$$

2.3.4 Monotonicity

A voting procedure is *monotonic* if raising a winner candidate's place in a voter's ballot never causes it to not win anymore.

Definition 2.7. *F is monotonic if for any candidate x and ballot response profile d, $\forall i : 1, ..., n, d'_i \in \mathcal{L}(\mathcal{X})^n$:*

$$x \in F(d) : d'_i >_x d_i \Rightarrow x \in F(d_{-i}, d'_i).$$

2.3.5 Condorcet Criterion

The *Condorcet winner* is a candidate that beats every other candidate in a pairwise comparison. There is no guarantee to have a Condorcet winner in all cases, but if it exists, it will be unique.

A voting procedure satisfies the *Condorcet principle* if it elects the Condorcet winner whenever it exists.

Similarly, the *Condorcet loser* is a candidate that is beaten by every other candidate in a pairwise comparison. Like the Condorcet winner, the Condorcet loser is unique if it exists.

Sometimes, alongside checking if a voting procedure is electing the Condorcet winner, it is also checked that the procedure doesn't elect the Condorcet loser.

2.3.6 Consistency

A voting procedure is *consistent* if when the set of voters is divided arbitrarily into two or more subsets and the procedure is run on each of them separately, if all of these separate elections have the same winner then an election on the entire set of candidates entails that same result too.

Definition 2.8. *F is consistent if for any partition of V to r subsets $V_1, V_2, ..., V_r$:*
$F(d|_{V_1}) = F(d|_{V_2}) = ... = F(d|_{V_r}) \Rightarrow F(d) = F(d|_{V_1})$.

2.3.7 Homogeneity

The *homogeneity* criterion says that replicating all votes (voters) uniformly does not affect the result of the election. That is, if a candidate wins in an election, the same candidate must win if for each voter in the previous situation there are k voters with a similar vote. Most voting rules satisfy homogeneity.

2.3.8 Non-Imposition

One of the very natural properties for a voting rule to be "fair" is that all candidates must have the chance of being a winner. That is, for any candidate, there must be a profile that yields to that candidate winning. This criterion is called *non-imposition*. Almost all of the (non-dictatorial) conventional voting methods satisfy non-imposition.

2.3.9 Participation

One of the well-known paradoxes in social choice theory is the *no show paradox*. The no show paradox happens when a favorable choice of a voter is already in the outcome of the election without his vote, but by adding that vote, the outcome becomes less favorable for the voter. In this regard, the *participation* criterion has been defined. A voting procedure has the participation property, if the no show paradox does not happen for that procedure. The practical result of this criterion is that when a voting rule satisfies participation then voting honestly is always better than not voting.

Voting methods such as plurality, approval and the Borda rule satisfy the participation criterion. All Condorcet methods fail to satisfy participation [17].

2.3.10 Reversal symmetry

The *reversal symmetry* property has been proposed by Saari [23] as a criterion for single-winner methods. It says that if all the votes (preferences) are reversed, then the result should be reversed. In particular, in a single-winner election, the strict winner should not still be the winner if the votes are reversed. So, if reversal symmetry does not hold in a single-winner election, it means that there exists a candidate that is both the socially best and worst candidate simultaneously.

Most voting methods satisfy the reversal symmetry property and this is usually easy to check. However, many of the voting rules with more than one round such as instant-runoff fail this criterion. Other voting rules that fail the criterion are methods which use only the top preference of each voter, like plurality. Also, it is easy to see that when a method does not satisfy the Condorcet loser criterion, it may fail reversal symmetry too.

2.3.11 Independence of Irrelevant Alternatives

One of the main criteria that Arrow considered in his famous impossibility theorem is *Independence of Irrelevant Alternatives* (IIA) [1]. The criterion requires that if the relative situation of two candidates are the same in two profiles, their relative situation in the social outcome should be the same. That is, if two candidates have the same relative rankings in the two profiles, and one of them wins under the first profile, the other candidates should not win under the second profile.

Although it seems to be an easy criterion for voting methods and an intuitively necessary requirement for any ordinary voting rule, it is a strong requirement and is violated in most voting methods.

2.4 Preference Aggregation

In the previous sections the focus was on voting systems which result in one candidate or a subset of the candidates as the winner. Systems that not only elect k alternatives as the winners, but also obtain a collective ranking of them are called *preference aggregation systems.*

The function that aggregates preferences to a collective preference is called a *social welfare function* (SWF). It is also called a *social preference function*, or in short *preference function* (PF).

Most voting systems can be considered as a special case of preference functions. That is, the winner(s) of a voting system is the best candidate(s) (top choice(s) in the social ranking) in its corresponding preference function.

Consequently, most of the basic voting procedures can easily be extended to preference functions and their axioms can be generalized for a preference function.

We will introduce the modified version of some of these procedures and properties below.

2.4.1 Notation and Basic Definitions

In the context of social welfare functions, we mostly use the terms agents for voters and alternatives for candidates; although we may use both terms interchangeably. Most of the definitions and notations are the same or similar to their counterparts in voting systems and so we do not repeat all of them here.

Let $\mathcal{V} = \{v_1, ..., v_n\}$ be the set of agents (voters) and let $\mathcal{X} = \{c_1, ..., c_m\}$ be the set of alternatives (candidates). There might be restrictions on the way that voters can express their preferences. For example, forcing reported preferences to be strict, or reporting preference only on a number of candidate rather than all of them. But here, as before, we assume that voters' preferences are linear orders. So, the set of admissible ballots for the set of alternatives $\mathcal{X}$ is $\mathcal{L}(\mathcal{X})$.

The n-tuple $d = (d_1, d_2, ..., d_n)$ which is $d_i \in \mathcal{L}(\mathcal{X})$ for all voters $i = 1, ..., n$, is the ballot response profile.

Definition 2.9 (Social welfare function)**.** *For the set of alternatives $\mathcal{X}$ and n voters, a social welfare function aggregates n preferences (linear order) into a collective preference (linear order). That is,*

$$F : \mathcal{L}(\mathcal{X})^n \rightarrow \mathcal{L}(\mathcal{X}).$$

2.4.2 Preference Aggregation Procedures

In this section we introduce some basic preference aggregation procedures. In fact most of these procedures are just an extension of their corresponding voting methods. There are three main approaches for the extension of a voting rule to a preference aggregation method.

- If the voting method first finds a specific linear order of the alternatives and then gives the top alternative of that ordering as the winner, we simply use that ordering as the social preference.
- If the voting method assigns a score to all alternatives and then selects the one with highest (lowest) score as the winner, we can use these scores to rank the rest of the alternatives.
- In other cases, we can put the winner of the voting system as the top member of the social ranking and then run the same voting rule on the rest of the alternatives and put the new winner as the second member of the ranking and repeat this procedure to produce a complete ranking of the alternatives.

Since voting rules have been primarily designed for real-world voting, some voting systems have a non-ranked input rather than ranking on the alternatives which is expected for preference aggregation procedures. This shows that we can extract social rankings even from non-ranked input. (Note that non-ranked ballots are also actually a sort of ranking, e.g., a single vote for an alternative v_1 is equivalent to the weak order: $v_1 \succ v_2 \sim ... \sim v_n$.)

Likewise, most of the axiomatic properties for voting systems can be adapted for preference aggregation. Hence we do not repeat them here.

Chapter 3

The Model

In this chapter we introduce our model in a formal manner. Most of the following concepts and definitions could be defined more generally, that is in a way that covers preferences represented by other means, like cardinal preferences (utility functions). However, as is usual in abstract social choice theory, we just consider ordinal preferences.

3.1 Notation and Basic Definitions

Let $\mathcal{N} = \{1, 2, ..., n\}$ be a finite set of n *voters* (*individuals*) and set $\mathcal{X} = \{c_1, c_2, ..., c_m\}$ a finite set of *candidates* (*alternatives*).

A *strict linear order* $\succ$ on a set S is a binary relation on S that is irreflexive, transitive, and complete. Each voter expresses a preference by providing a strict linear order on a subset of $\mathcal{X}$. A profile, $\mathcal{R}$, is the set of such preferences of all the voters. Our goal is to produce, for each input profile, a set of (weighted) trends which are aggregations of voter preferences. Trends are also strict linear orders on subsets of $\mathcal{X}$.

For the set of alternatives $\mathcal{X}$, we denote the set of strict linear orders on subsets of $\mathcal{X}$ as $\mathcal{L}(\mathcal{X})$:

$$\mathcal{L}(\mathcal{X}) \ := \ \{(S, \succ) \mid S \subseteq \mathcal{X} \text{ and } \succ \text{ is a strict linear order on } S\}$$

We usually denote the members of $\mathcal{L}(\mathcal{X})$ as tuples $(c_{\pi_1}, ..., c_{\pi_k})$, which means $c_{\pi_1} \succ ... \succ c_{\pi_k}$.

Definition 3.1 (Preferences)**.** *A preference is an element of* $\mathcal{L}(\mathcal{X})$.

If a voter wants to provide more than one linear order, we can add virtual voters for the extra preferences and assign each of the preferences to one virtual voter. So, we can assume that each voter has only one linear order as its preference.

We denote the preference of voter i as p_i. We also write $a \succ_i b$ for a and b in $\mathcal{X}$ when a comes before b in p_i, i.e., when voter i prefers a to b. We use $S(p)$ for the set of candidates (a subset of $\mathcal{X}$) compared in the preference p and $l(p)$ for the number of these candidates or, in other words, the length of the preference p. So, for $p = (c_{\pi_1}, ..., c_{\pi_k})$, $S(p) = \{c_{\pi_1}, ..., c_{\pi_k}\} \subseteq \mathcal{X}$ and $l(p) = |S(p)| = k$.

In our model, an outcome is a set of preferences instead of just one aggregated preference. We call each of these preferences a *trend*. So, a trend is just a preference, which is part of the outcome and in fact is an aggregated preference obtained from input preferences. We often also assign a value to each trend which determines its reliability or popularity or any other concept that explains its strength. In this case we call it a *weighted trend*.

Definition 3.2 (Trends). *A trend is an aggregated preference, i.e., an element of $\mathcal{L}(\mathcal{X})$. A weighted trend is an element of $\mathcal{L}(\mathcal{X}) \times \mathbb{R}$.*

We usually write trends in the same way as preferences but with capital letters, e.g., P_i. When it causes no confusion, we may also refer to a weighted trend as simply a "trend". If P is a weighted trend, $v(P)$ is its weight.

So the output of this model (a set of aggregated trends) is expressed as $\boldsymbol{P} = \{P_1, P_2, ..., P_l\}$ if trends are not weighted, and for weighted trends: $\boldsymbol{P_w} = \{(P_1, v(P_1)), ..., (P_l, v(P_l))\}$. In general, it is not necessary for the sets of candidates of different trends in $\boldsymbol{P}$ to be disjoint; they can even be equal as long as their orders are not the same.

Let p_1 and p_2 be two preferences. We say p_1 is a *subpreference* of p_2 (and p_2 is a *superpreference* of p_1), if p_1 can be obtained from p_2 by just dropping some candidates from it.

Definition 3.3 (Subpreference). *p_1 is a subpreference of p_2 which we write as $p_1 \sqsubseteq p_2$, iff:*

$$\forall c_1, c_2 \in S(p_1), c_1 \succ_{p_1} c_2 \Rightarrow c_1 \succ_{p_2} c_2.$$

If $p_1 \sqsubseteq p_2$ and $p_1 \neq p_2$ we call p_1 a proper subpreference *of p_2 and write $p_1 \sqsubset p_2$. We also can say that p_1 can be deduced from p_2 and write it as, $p_2 \models p_1$.*

By the definition we have: $p_1 \sqsubseteq p_2 \Rightarrow S(p_1) \subseteq S(p_2)$.

For weighted preferences, we say a preference p_1 is a *weighted subpreference* of p_2 if it does not have a greater weight than p_2 in addition to being a subpreference of it.

Definition 3.4 (Weighted subpreference). *Let p_1 and p_2 be two weighted preferences. p_1 is a weighted subpreference of p_2 which we write as $p_1 \sqsubseteq_w p_2$, iff: $p_1 \sqsubseteq p_2$ and $v(p_1) \leq v(p_2)$.*

$p_1 \sqsubset_w p_2$ if $p_1 \sqsubseteq_w p_2$ but $p_1 \neq p_2$ or $v(p_1) \neq v(p_2)$.

Again we can say that p_1 can be deduced from p_2 and write it as, $p_2 \models p_1$.

3.2 Formal Framework

A Trend Aggregation Procedure (TAP) is a procedure which takes the preferences of voters and extracts a number of preferences, which are the social trends, from these inputs. In the most general setting, the input preferences could be expressed in many different ways (ordinal or cardinal, total or partial, etc.), and for the general framework of TAPs we need not impose any restriction on them. Similarly, trends can be in different formats (ordinal or cardinal, weighted or non-weighted, etc.). However, for the remainder of this thesis, we shall assume that preferences and trends are linear order, as defined above.

Definition 3.5 (Trend aggregation procedure)**.** *A trend aggregation procedure (TAP) is a function $\boldsymbol{F}$ mapping any profile of preferences to a set of of trends:*

$$\boldsymbol{F} : \mathcal{L}(\mathcal{X})^n \to 2^{\mathcal{L}(\mathcal{X})}$$

If the system reports weighted trends we call it a Weighted Trend Aggregation Procedure (WTAP).

Definition 3.6 (Weighted trend aggregation procedure)**.** *A WTAP is a TAP $\boldsymbol{F}$ with the addition of a weight for each trend. That is,*

$$\boldsymbol{F} : \mathcal{L}(\mathcal{X})^n \to \mathbb{R}^{\mathcal{L}(\mathcal{X})}$$

Another possible extension of TAPs is adding a total order on the set of the output trends instead of assigning a score to each of them. It is clear that each WTAP can be interpreted as this type of trend aggregating system.

When we deal with a number of total orders as input preferences, usually we need to break these lengthy orders into basic data. This is mainly due to computational aspects, but it is not only for implementation purposes. Even in theory, most procedures defined for aggregating preferences do not work with the whole orders in the process of aggregation, since it is not easy to define operations on the long orders. So they transform inputs to a number of smaller pieces of data in the first step (of course with the cost of losing some data in most cases). For example, in the Borda rule or any other scoring rule, the whole order of each voter's preference reduces to a score for each candidate in that order; or in Copeland's method, first each input preference is reduced to a set of pairwise comparisons between the candidates.

We call a method that uses this scheme for aggregating preferences a *two-phase trend aggregation procedure*, which means the procedure consists of a phase —we call it *decomposition*— of converting preferences to a set of basic relations and a phase of computing aggregated preference(s) from these data (*composition*).

In any two-phase procedure, we call the set of converted data which is the output of the decomposition phase and the input for the composition phase, the *intermediary set*. This is the set of all relations that are admissible to be both the output of the first phase and the input for the second phase. Usually there is an *intermediary domain* set, $\mathcal{I}(\mathcal{X})$, which characterizes the form of the members of an intermediatory set. In this case, the intermediatory set is a (weighted) powerset of an intermediatory domain, e.g., $2^{\mathcal{I}(\mathcal{X})}$ or $\mathbb{R}^{\mathcal{I}(\mathcal{X})}$ when it is weighted.

Definition 3.7 (Two-phase trend aggregation procedure)**.** *A trend aggregation procedure is two-phase if it results from consecutive runs of a composition phase after a decomposition in which the output of the decomposition phase is the input of the composition phase. That is, if $\mathcal{I}(\mathcal{X})$ is an intermediary domain set, $\boldsymbol{D}$ and $\boldsymbol{C}$ are the two functions that stand for decomposition and composition such that:*

$$\boldsymbol{D} : \mathcal{L}(\mathcal{X})^n \longrightarrow \mathbb{R}^{\mathcal{I}(\mathcal{X})};$$

and

$$\boldsymbol{C} : \mathbb{R}^{\mathcal{I}(\mathcal{X})} \longrightarrow \mathbb{R}^{\mathcal{L}(\mathcal{X})}.$$

The intermediary relations are usually the pairwise comparisons or a function which assigns a score to every candidate in each input preference. If we limit the intermediary relations to score-assigning functions, we call it a *0-degree two-phase procedure*; and if it is limited to only pairwise comparisons of the candidates, we call it *1-degree two-phase procedure.*

For example, among the conventional voting systems, we can consider the positional scoring methods like the Borda count as 0-degree two-phase procedures and the majority ranking methods like the Condorcet methods as 1-degree two-phase procedures.

Definition 3.8. *A 0-degree two-phase procedure is a two-phase procedure where its intermediary domain set $\mathcal{I}$ is: $\mathcal{I}(\mathcal{X}) = \mathcal{X}$. A 1-degree two-phase procedure is a two-phase procedure where its intermediary domain set $\mathcal{I}$ is: $\mathcal{I}(\mathcal{X}) = \mathcal{X}^2$*

In this thesis we mostly focus on "*1-degree two-phase WTAPs*".

3.3 Methods

Up to here we introduced a general framework for our model, particularly a family of TAPs which we call 1-degree two-phase weighted trend aggregation procedures. As we have already defined, a TAP of this type consists of two phases, namely decomposition and composition. Now to complete the details of our proposed method, we present our proposal for this part of the model. We propose a general class of methods for each of the decomposition and composition phases.

3.3.1 Decomposition

When we talk about relations here, we mean the implicit relation each voter conveys by reporting a preference on a subset of candidates. We assume there was some sort of relation among the members of each of these partial preferences which caused them to occur in a single preference.

We need a sort of decomposition method that makes it possible to add up data, but also preserves the relations to some extent.

Since we are dealing with 1-degree TAPs our only option to handle this is choosing reasonable pairwise comparisons and assigning appropriate weights to them. So each input preference with length k (i.e., a total order on a subset of candidates consisting of k members), should be decomposed into the $\binom{k}{2}$ pairwise comparisons it consists of, as in most conventional preference aggregation methods based on pairwise comparisons. Although the method that we will use to compose these fragments of data is pivotal to retain the relations, reasonably assigned weights can also play an important role in this regard.

In fact, as we will see later, if we want to consider all possible trends in our method as output trends, the weights play no role in shaping the trends, they only affect the weights of trends. However, if we want to select a subset of trends for output, then these weights would have a determining role in deciding which trends should be reported, and of course, with what weight. So, if we do not want to consider weights and are not selecting a subset of the trends we can assign a constant weight (simply 1) to all pairwise comparisons.

To explain the idea of our proposal, let us start with a basic case. When we have $A \succ B \succ C$, if we can talk about something like the amount of preference, it is reasonable to say that the amount of preferring A to C is more than preferring it to B and also B to C; but it is not always true to say this amount is exactly equal to sum of the values of the other two. From this viewpoint, we can also think that it is plausible for A to be more preferable to C in $A \succ B \succ C$ than when the reported preference is a single comparison between them (i.e., $A \succ C$). It may even be so in the case of, for example, $A \succ C \succ B$, but not necessarily.

So it seems reasonable to assign higher weights to the pairwise comparisons of more distant pairs in the preferences in the decomposition phase. To this end, we use an increasing (or non-decreasing) function that assigns a value to the pairs based on their distance in the preference they are contained in.

Definition 3.9. *The distance of two members c_i and c_j of a preference p, which we denote it as $d_p(c_i, c_j)$, is the difference in their place in p. That is,*

$$d_p(c_i, c_j) = |\{x \in S(p) | c_i \succ_p x \succ_p c_j\}| + 1.$$

So, in brief, for a reported preference on k candidates ($k \leq m$, m is the number of total candidates):

$$c_1 \succ c_2 \succ \cdots \succ c_i \succ \cdots \succ c_j \succ \cdots \succ c_k;$$

we can assign a weight to the pairwise comparison of $c_i \succ c_j$ by using $g(l)$, where $l = d_p(c_i, c_j)$ and g is an increasing (or non-decreasing) function.

We may impose other properties on the function g in addition to monotonicity to obtain extra properties in the results, but for the general case we just consider g to be a non-decreasing function from $\mathbb{Z}$ to $\mathbb{R}$ in which $g(x) = 0$ for all $x \leq 0$.

Now, after assigning weights to all of these pairwise comparisons from all of the input preferences, we can add them up for each pairwise comparison.

The maximum number of possible pairwise comparisons is two times the number of candidate pairs (one for each side, i.e., one for preferring A to B and one for preferring B to A). So this would be $2\binom{m}{2} = m(m-1)$, in which m is the total number of candidates. Comparing this number to the number of all possible preferences, which is $\sum_{i=2}^{m} m!/(m-i)! > m!$, shows that there is a great advantage in accumulating the whole data in less than m^2 values instead of more than $m!$ ones.

If we denote the assigned value for the comparison $c_i \succ c_j$ from the preference p as $v_p(c_i \succ c_j)$ and the summation of these values from all input preferences as $v(c_i \succ c_j)$, we have:

$$v_p(c_i \succ c_j) = g(d_p(c_i, c_j))$$

and

$$v(c_i \succ c_j) = \sum_{r=1}^{n} v_{p_r}(c_i \succ c_j)$$

We call this class of decomposition procedures *monotonic decompositions.*

Definition 3.10 (Monotonic decomposition)**.** *A decomposition procedure*

$$\boldsymbol{D} : \mathcal{L}(\mathcal{X})^n \longrightarrow \mathbb{R}^{\mathcal{I}(\mathcal{X})}$$

is monotonic if $\mathcal{I}(\mathcal{X}) = \{(x, y) | x, y \in \mathcal{X}, x \neq y\}$ *and it uses a non-decreasing function* g*,*

$$g : \mathbb{Z} \longrightarrow \mathbb{R}\ , \quad g(x) = 0 \textit{ for } x \leq 0$$

to accumulate a value for each pairwise comparison of candidates c_i *and* c_j *(*$i, j \leq m$ *and* n *is the number of voters) as:*

$$v(c_i \succ c_j) = \sum_{r=1}^{n} g(d_{p_r}(c_i, c_j)).$$

3.3.2 Composition

Now the problem is how to combine these fragments of data in a way to obtain some total orders on the subsets of candidates, i.e., aggregated preferences. In this phase we should try to merge pairwise orders to get some chain of orders and alongside it retrieve relations of the candidates that we tried to preserve in the decomposition phase.

To do this, let us take a look at the simple case of $A \succ B \succ C$ again. As we decomposed it into the three pairwise comparisons of $A \succ B$, $B \succ C$ and $A \succ C$, it is natural to do the reverse procedure and merge these three pairs to a total order on them. Notice that here we do not have transitivity. That is, from $A \succ B$ and $B \succ C$ we cannot conclude that $A \succ C$; although it may be the case for each individual's preference. So we may not have as much supportive data for the $A \succ C$ as we do have for $A \succ B$ and $B \succ C$, and also in most cases we do not have the same supportive data for the last two.

Therefore, for these three candidates and especially for these three pairwise comparisons we have three independent values coming from the summation of their corresponding weights. Let us call these three values a, b and c corresponding to $A \succ B$, $B \succ C$ and $A \succ C$ respectively. Now we want to assign a value to the aggregated preference of $A \succ B \succ C$, based on these values. The most confident thing that we can say is to consider each of these three with the value of $d = \min\{a, b, c\}$ instead of their own value, and then we have three fractions of the aggregated preference (trend) of $A \succ B \succ C$ with the same supportive value and so we can say that there is such a trend in the preferences of the voters and assign this value (or some corresponding value) to this trend.

We can also think about some other alternatives for assigning a value to such a trend. For example, maybe the summation or multiplication of the values of basic comparisons of each trend would be good choices in some cases. In addition, since we will have different trends with different lengths in our model, considering the length of the trends is good or maybe essential in such cases. So we may use the average of these values (arithmetic or geometric mean) for the reliability of each trend.

Furthermore, when we use the distances of pairwise comparisons in the decomposition phase, it might be more reasonable to use a sort of reverse of the same procedure we used to combine them. For example, consider again $A \succ B$, $B \succ C$ and $A \succ C$ with the corresponding values a, b and c, respectively. We should use the values of $a/g(1)$, $b/g(1)$ and $c/g(2)$ instead of the initial values, in which g is a non-decreasing function similar (or maybe identical) to the one used in the decomposition phase. Hence, we can assign $d = \min\{a/g(1), b/g(1), c/g(2)\}$ for the reliability of the trend $A \succ B \succ C$. In the general case, if a is the summation of weights for $c_i \succ c_j$, then for computing the reliability of trend

$$P : c_1 \succ c_2 \succ \cdots \succ c_i \succ \cdots \succ c_j \succ \cdots \succ c_k,$$

(let $l = d_P(c_i, c_j)$ and g is a non-decreasing function), we would use the value of $a/g(l)$ instead of a. Similar to the simple method, here we can (or should) use the average of these values too.

So in brief the combination method would be: For a trend like

$$P : c_1 \succ c_2 \succ \cdots \succ c_i \succ \cdots \succ c_j \succ \cdots \succ c_k,$$

if we have $v(c_i \succ c_j) = a$ (aggregated from all input preferences), the support of this value for trend P is:

$$v'_P(c_i \succ c_j) = a/g(l);$$

and if $\boldsymbol{T}$ is min or a kind of average (arithmetic or geometric mean) function, then the support for trend P is calculated by:

$$v(P) = \boldsymbol{T}_{i<j\leq k}\{v'_P(c_i \succ c_j)\}.$$

Definition 3.11 (Power mean)**.** *A* power mean *(generalized mean) with exponent ρ (ρ is a non-zero real number) of the positive real numbers $x_1, ..., x_n$ is:*

$$M_\rho(x_1, ..., x_n) = (\frac{1}{n}\sum_{i=1}^{n} x_i^\rho)^{1/\rho};$$

and its asymptotic values for $\rho = 0$, $+\infty$ or $-\infty$ are defined as:

$$M_0(x_1, ..., x_n) = \sqrt[n]{\prod_{i=1}^{n} x_i},$$

$$M_\infty(x_1, ..., x_n) = \max(x_1, ..., x_n),$$

$$M_{-\infty}(x_1, ..., x_n) = \min(x_1, ..., x_n).$$

As we can see, M_1 is the arithmetic mean and M_0 is the geometric mean. The most promising choices for the function $\boldsymbol{T}$ in this model are the arithmetic mean(M_1), geometric mean(M_0), min($M_{-\infty}$) and max(M_∞) functions.

Definition 3.12 (Monotonic composition)**.** *A composition procedure*

$$\boldsymbol{C} : \mathbb{R}^{\mathcal{I}(\mathcal{X})} \longrightarrow \mathbb{R}^{\mathcal{L}(\mathcal{X})}.$$

is monotonic if $\mathcal{I}(\mathcal{X}) = \{(x, y)|x, y \in \mathcal{X}, x \neq y\}$ and it uses a non-decreasing function g,

$$g : \mathbb{Z} \longrightarrow \mathbb{R}\ , \quad g(x) = 0 \text{ for } x \leq 0,$$

and a power mean function $\boldsymbol{T}$ to assign to each trend $P \in 2^{\mathcal{L}(\mathcal{X})}$ a weight as:

$$v(P) = \boldsymbol{T}_{i<j\leq k}\left\{\frac{v(c_i \succ c_j)}{g(d_P(c_i, c_j))}\right\}$$

Now we have completed the definitions of all the necessary parts of a trend extractation procedure and can define our method formally.

Definition 3.13 (Monotonic 1-degree two-phase WTAPs)**.** *A monotonic 1-degree two-phase WTAP, $\boldsymbol{F}^g_{\boldsymbol{T}}$ (or $\boldsymbol{F}^g_\rho$), is a trend aggregation procedure with composition and decomposition phases that are monotonic with an identical non-decreasing function g and which uses a power mean function $\boldsymbol{T}$ (or M_ρ) for composition.*

Since we only deal with monotonic 1-degree two-phase WTAPs, we may simply call them TAPs. In the next section, we will introduce some slightly different versions of TAPs. We will also discuss some possible ways to refine this potentially enormous number of trends to a more reasonable number of selected ones. In the next chapter, we will investigate the properties of these models in general and for a few special cases, each with a specific monotonic function or mean function.

3.4 Normalization

The value that has been assigned to the trends can be normalized proportionally in order to fall into a specific range, e.g., $[0,1]$. However, here by normalization we do not mean this trivial calculation.

To explain this normalization let us take another look at the method we had, which we call Method 0:

3.4.1 Method 0

Here we only put together the definitions of the different parts of the method that we have seen in the previous section.

Definition 3.14 (Method 0)**.** *The Method 0 TAP is a monotonic 1-degree two-phase TAP which uses*

$$v_p(c_i \succ c_j) = g(d_p(c_i, c_j))$$

in the decomposition phase to assign a value to the pairwise comparison of c_i and c_j in preference p, and

$$v(c_i \succ c_j) = \sum_{r=1}^{n} v_{p_r}(c_i \succ c_j)$$

is used for the aggregated value of the pairwise comparison from all input preferences. Also,

$$v(P) = \boldsymbol{T}_{i<j\leq k}\left\{\frac{v(c_i \succ c_j)}{g(d_P(c_i, c_j))}\right\}$$

is used in the composition phase to calculate the weight of any trend P.

Now suppose the basic case of one voter with preference $p : A \succ B \succ C$. All the subpreferences of this preference are potential output trends. With use of the proposed method, the weights of these trends would be:

$$v(P_1 = A \succ B) = \boldsymbol{T}(g(1)/g(1)) = \boldsymbol{T}(1) = 1$$
$$v(P_2 = A \succ C) = \boldsymbol{T}(g(2)/g(1)) = \boldsymbol{T}(g(2)) = g(2) \geq 1$$
$$v(P_3 = A \succ B \succ C) = \boldsymbol{T}(g(1)/g(1), g(1)/g(1), g(2)/g(2)) = \boldsymbol{T}(1,1,1) = 1$$

So $v(A \succ C) \geq v(A \succ B \succ C) = v(A \succ B)$.

It may seem counter-intuitive to assign a value to $A \succ C$ which is higher than the weight of P_3, while P_3 is exactly the same singular input preference. The contradictory aspect of this result is more clear in the case where selecting only one of these trends is needed. In this case, P_2 will be chosen due to its greater weight despite the fact that P_3 has the same information as the input and P_2 is just a subpreference of the same input preference.

On the other hand, this can be justified by interpreting it as the greater weight of P_2 representing more confidence in it and making the output result more robust.

So, whether this kind of result is acceptable or not depends on our definition and interpretation of trends' weights and the application. In this regard, we will present some more methods which handle this issue, each of which has its own interpretation and properties.

We can avoid choosing a preference like P_2 over P_3 in such a case by imposing restrictions on the selection procedure. We will discuss this kind of procedures in the next section, but in most cases the selection procedure on these weights does not lead to an acceptable result and therefore we need other methods which produce different weights.

Here we present three alternative methods for the initial method. Each of these can be seen as a modification of the original method.

First let us define a generalization of the above example to clarify what property we are seeking.

Definition 3.15 (Singular election)**.** *An election with only one voter is a* singular election *and the preference of this unique voter is the* singular vote *or* singular preference*.*

Definition 3.16 (Informativeness)**.** *A TAP is called* informative *if for any singular election with singular vote p, the weight of p as a trend is maximal; that is,*

$\forall P_i \in \boldsymbol{P} : v(P_i) \leq v(p)$.

3.4.2 Method 1

The issue explained above was caused by the fact that for a particular input preference p, its subpreference q gets the same values for its pairwise comparisons in the decomposition phase. However, in the composition formula, due to the lesser distance between its member candidates compared to the corresponding distance in the initial preference p, it gets greater weight. This means that there is an advantage for the shorter trends in the original formula. So a basic solution is to add a normalization factor to compensate for this "unfair" advantage of shorter trends.

We are looking for a normalization factor that makes the weight of any subpreference of singular preference p at most $v(p)$. So, we define a factor based on the length of the trends, which is applied to the raw values of trends' weights from Method 0, i.e., each value will be divided by the corresponding factor. Since the raw weight of trend p in this case is 1, if it remains the same, all the other weights should be downgraded under 1 to satisfy the informativeness property. Hence for each length, the factor should be at least the maximum possible weight of trends with that length from Method 0.

Consider a singular election with m candidates and the singular preference of $p = c_1 \succ c_2 \succ \cdots \succ c_m$; we calculate the normalization factor μ as follows:

$$\mu(k) = \max_{l(P)=k} \left\{ \boldsymbol{T}_{c_i \succ c_j \in P} \left\{ \frac{v_p(c_i \succ c_j)}{g(d_P(c_i, c_j))} \right\} \right\}.$$

Since $v_p(c_i \succ c_j) = g(d_p(c_i, c_j))$ and $d_p(c_i, c_j) = j - i$, we can replace $v_p(c_i \succ c_j)$ in the formula with $g(j-i)$. This shows that $\mu(k)$ only depends on the value of k (and the number of alternatives, m) and the definition of the functions g and $\boldsymbol{T}$; and not on any part of the input. So, in general, if g and $\boldsymbol{T}$ are fixed, we denote the normalization factor for preferences of length k when there are m alternatives as $\mu_m(k)$.

Definition 3.17 (Method 1)**.** *If $v(P)$ is the weight of trend P in Method 0, its weight in Method 1 is $v_\mu(P)$ which is calculated as: $v_\mu(P) = v(P)/\mu(l(P))$, in which*

$$\mu_m(k) = \max_{1 \le \sigma_1 < \cdots < \sigma_k \le m} \left\{ \boldsymbol{T}_{i<j\le k} \left\{ \frac{g(\sigma_j - \sigma_i)}{g(j-i)} \right\} \right\}.$$

From the definition, we have $\mu_m(2) = g(m-1)/g(1) = g(m-1)$ and $\mu_m(m) = 1$.

In this method we do not need to change any part of the TAP except for adding the normalization factor (μ) to the composition formula or in other words, applying the normalization factor to the results of that TAP.

3.4.3 Method 2

In the original method we use the absolute distance of the pairs to calculate the value of the pairwise comparisons, regardless of the length of the whole preference.

For example, the pairwise comparison $A \succ C$ gets the same value, $g(2)$, from both preferences $p_1 : A \succ B \succ C$ and $p_2 : A \succ B \succ C \succ \cdots \succ Z$. But from another point of view, we can assume that C is possibly less preferred to A in the second case. This is because, in the first case the longest distance in p_1 is also 2, the same as the distance of A and C in that preference; while the distance between A and C is very small compared to the length of the p_2.

So here we consider the length of each preference as the scale for measuring the distance rather than the absolute distance. That is, in each preference the maximum value is assigned to the pairwise comparison of the most distant pair in that preference and the rest of the pairs get the proportional value with respect to their relative distance in the preference.

For instance, we can adopt 1 for the value of the pairwise comparison of the first and last candidates of each preference. Then for other pairs the value would be the ratio of their absolute distance to the maximum distance of that preference; i.e., if d is the absolute distance of two alternatives in preference p and $l = l(p)$, we assign $f(d, l) = d/(l - 1)$ for the value of this pairwise comparison in p.

We can generalize this formula by using the function g: $f(d, l) = g(d)/g(l - 1)$. Now we can replace this function f for the g in the decomposition and composition formulas.

Definition 3.18 (Method 2)**.** *A TAP of Method 2 is a monotonic 1-degree two-phase TAP with decomposition phase:*

$$\tilde{v}_p(c_i \succ c_j) = f(d_p(c_i, c_j), l(p))$$

$$\tilde{v}(c_i \succ c_j) = \sum_{r=1}^{n} \tilde{v}_{p_r}(c_i \succ c_j)$$

and composition formula:

$$\tilde{v}(P) = \boldsymbol{T}_{i<j\leq k} \left\{ \frac{v(c_i \succ c_j)}{f(d_P(c_i, c_j), l(P))} \right\};$$

where $f(d, l) = g(d)/g(l - 1)$.

3.4.4 Method 3

In the previous methods, all the values from different input preferences are added for each pairwise comparison. This makes them 1-degree two-phase procedures. When a pairwise comparison $(A \succ B)$ occurs in an input preference with distance d, we can say that this $A \succ B$ supports any pairwise comparison of A and B with distance d or less in trends; thus we do not assign more support for a trend because of more distance in the input.

So if we can manage to separate this data for different distances, we can make the procedure informative. This needs more data than an intermediary set of a

1-degree WTAP has, yet it still deals with just pairwise comparisons. Indeed, for each pairwise comparison, it has a vector of values for each distance from 1 to $m-1$ instead of only one value. This makes the size of the intermediary set m times larger. We call this kind of procedures *extended 1-degree* two-phase procedures.

The idea is that if there is enough support for any part of a trend from an input preference, this preference would contribute one unit (not more) to the confidence of that part of the trend; and if it is not enough for full support it will contribute proportionally to its share (like in the initial method). Here by 'enough support for a pair', we mean that they have at least the same distance in the input as they have in the output trend.

We show the number of occurrences of any pairwise comparison with a specific distance t in the input preferences by $v_t(c_i \succ c_j)$:

$$v_t(c_i \succ c_j) = \#\{p_r | d_{p_r}(c_i, c_j) = t\}$$

These v_ts will replace the v in the decomposition phase, and the composition part would use this formula for the weight of a trend P:

$$v_{\#}(P) = \boldsymbol{T}_{i<j\leq k} \left\{ \frac{\sum_{r=1}^{d_P(c_i,c_j)-1} v_r(c_i \succ c_j) \cdot g(r)}{g(d_P(c_i,c_j))} + \sum_{r=d_P(c_i,c_j)}^{m-1} v_r(c_i \succ c_j) \right\}$$

Definition 3.19 (Method 3). *Method 3 is a monotonic extended 1-degree two-phase TAP with decomposition phase:* $v_t(c_i \succ c_j) = |\{p_r | d_{p_r}(c_i, c_j) = t\}|$ *and composition formula:*

$$v_{\#}(P) = \boldsymbol{T}_{i<j\leq k} \left\{ \frac{\sum_{r=1}^{m-1} v_r(c_i \succ c_j) \cdot \min(g(r), g(d_P(c_i,c_j)))}{g(d_P(c_i,c_j))} \right\}$$

3.4.5 Examples

Here we present an example to show the calculation procedures of these methods in detail.

Example 3.1. *Consider a singular election with the vote* $p : A \succ B \succ C$.

We will calculate, for each of the four score-assigning methods, the weights of these three trends:

$P_1 : A \succ B$,

$P_2 : A \succ C$,

$P_3 : A \succ B \succ C$.

First, we calculate the aggregated weights (for Method 0 and Method 1) of all pairwise comparisons:

$v(A \succ B) = v(B \succ C) = g(1)$,

$v(A \succ C) = g(2)$.
Now, for these two methods, we have:
Method 0:
$v(P_1) = \boldsymbol{T}(g(1)/g(1)) = \boldsymbol{T}(1) = 1$,
$v(P_2) = \boldsymbol{T}(g(2)/g(1)) = \boldsymbol{T}(g(2)) = g(2) \geq 1$, *and*
$v(P_3) = \boldsymbol{T}(g(1)/g(1), g(1)/g(1), g(2)/g(2)) = \boldsymbol{T}(1, 1, 1) = 1$.

Method 1:
The value of μ *when* $m = 3$ *is:* $\mu_3(2) = g(m-1) = g(2)$, $\mu_3(3) = 1$;
and the weights of the trends are:
$v_\mu(P_1) = 1/\mu(2) = 1/g(2) \leq 1$,
$v_\mu(P_2) = g(2)/\mu(2) = g(2)/g(2) = 1$, *and*
$v_\mu(P_3) = 1/\mu(3) = 1$.

Method 2:
Aggregated weights of all pairwise comparisons in Method 2 are calculated as:
$\tilde{v}(A \succ B) = v(B \succ C) = 1/2$ *and*
$\tilde{v}(A \succ C) = 2/2 = 1$;
and the weights of the trends are:
$\tilde{v}(P_1) = \boldsymbol{T}(0.5/1) = 1/2$,
$\tilde{v}(P_2) = \boldsymbol{T}(1/1) = 1$, *and*
$\tilde{v}(P_3) = \boldsymbol{T}(0.5/0.5, 0.5/0.5, 1) = 1$.

Method 3:
For Method 3, the number of occurrences of each pair with each distance is:
$v_1(A \succ B) = 1$,
$v_1(B \succ C) = 1$, *and*
$v_2(A \succ C) = 1$;
and the weight of the trends are:
$v_\#(P_1) = \boldsymbol{T}(1.g(1)/g(1)) = \boldsymbol{T}(1) = 1$,
$v_\#(P_2) = \boldsymbol{T}(1.\min(g(2), g(1))/g(1)) = \boldsymbol{T}(1) = 1$, *and*
$v_\#(P_3) = \boldsymbol{T}(g(1)/g(1), g(1)/g(1), g(2)/g(2)) = \boldsymbol{T}(1, 1, 1) = 1$.

We can summarize these results in the following table (we can see that all of the three new methods satisfy informativeness):

3.5 Refinement

Our proposed method produces a huge number of trends as result, i.e., all possible trends (trends which have support for all of their parts). To make the output tractable and presentable, we need to prune them and select only a few trends. The general idea is that longer trends are better since they contain more data

	P_1	P_2	P_3
Method 0	1	≥ 1	1
Method 1	≤ 1	1	1
Method 2	< 1	1	1
Method 3	1	1	1

Table 3.1: Example 3.1

and are more useful. On the other hand, trends with higher weights are stronger and enjoy the support of more voters. So there is a trade-off between choosing an informative trend with a possibly lower weight or a shorter but more popular one. Since there is no concrete choice, we will present a number of options which lean to one side or provide some sort of compromise in between. First we need some definitions.

Given a set of prefrences, a *maximal* preference is a preference which is not a subpreference of any other member of the set. Since the subpreference relation is a partial order, there can be more than one maximal element in a set of preferences. We call the set of all maximal members of a set P, the *maximal set* and denote it as $\mathcal{M}(P)$.

Definition 3.20 (Maximal set)**.** *The maximal set of a set P is the set consisting of all the maximal members of P.*

$$\mathcal{M}(P) = \{p \in P | \nexists q \in P : p \sqsubset q\}$$

Like unweighted preferences, a *weighted maximal* preferences is a preference that is not a weighted subpreference of any other member of the set. We call the set of all weighted maximal members of a set P, the *weighted maximal set* and denote it as $\mathcal{M}_w(P)$.

Definition 3.21 (Weighted maximal set)**.** *The weighted maximal set of a set P is the set consisting of all the weighted maximal members of P.*

$$\mathcal{M}_w(P) = \{p \in P | \nexists q \in P : p \sqsubset_w q\}.$$

By the definition, for any set of weighted preferences P: $\mathcal{M}(P) \subseteq \mathcal{M}_w(P)$. This is because there could be subpreferences of a maximal preference in $\mathcal{M}(P)$ with greater weights, so they would be members of $\mathcal{M}_w(P)$.

As an example, consider a chain of preferences $p_2 \sqsubseteq p_3 \sqsubseteq \cdots \sqsubseteq p_m$ (p_i with length i), in which $v(p_2) > v(p_3) > \cdots > v(p_m)$. In such a case, for the set $S = \{p_2, p_3, \ldots, p_m\}$; $\mathcal{M}_w(S) = S$ while $\mathcal{M}(S) = \{p_m\}$.

We may use a threshold on either the length or the weight of preferences to put aside weaker preferences in that respect. We denote the set of preferences in P with the weight of at least α as $\mathcal{T}_w^\alpha(P)$, and the set of preferences in P with the length of at least k as $\mathcal{T}_l^k(P)$,

Definition 3.22. *For a set of preferences P, a real number $\alpha \geq 0$ and an integer $k \geq 0$:*
$\mathcal{T}_w^\alpha(P) = \{p \in P | v(p) \geq \alpha\}$
$\mathcal{T}_l^k(P) = \{p \in P | l(p) \geq k\}$

The set of top trends with regard to their weight or length can be seen as the special case of these functions when $\alpha = \max\{v(p)|p \in P\}$ and $k = \max\{l(p)|p \in P\}$ respectively. We denote these two sets of preferences, those with the highest weight and those with the highest length, by the same notation without the threshold value, i.e., $\mathcal{T}_w(P)$ and $\mathcal{T}_l(P)$ respectively.

Definition 3.23 (Top sets)**.** *For set of preferences P:*
$\mathcal{T}_w(P) = \{p \in P | \nexists q \in P : v(p) < v(q)\}$
$\mathcal{T}_l(P) = \{p \in P | \nexists q \in P : l(p) < l(q)\}$

Replacing a set of trends by its (weighted) maximal set is the most confident option to decrease the size of the output without loosing any useful data, since when we have a preference, there is not much more data in its subpreferences. The maximal set would shrink the size of a set significantly in most cases, although for the weighted maximal set it depends very much on the weights and the method we have chosen to assign weights. If the method is not informative, the weighted maximal set would not be that much smaller than the original set in most cases.

If we want to make the size of the set smaller or filter shorter or weaker trends, we may use a threshold on the length or weight or both. Again, we can replace this filtered set by its maximal set. In general, we can use any mixture of these operations to create different pruned subsets with different properties. As we will se, the order does not matter in most cases, but in some, changing the order would produce different results.

For all of the main four operations ($\mathcal{M}$, $\mathcal{M}_w$, $\mathcal{T}_w^\alpha$, $\mathcal{T}_l^k$; and also for the top sets), we have $F(F(P)) = F(P)$. We also have $\mathcal{M}_w(\mathcal{M}(P)) = \mathcal{M}(\mathcal{M}_w(P)) = \mathcal{M}(P)$. Furthermore, we have $\mathcal{T}_l(\mathcal{M}(P)) = \mathcal{T}_l(\mathcal{M}_w(P)) = \mathcal{M}(\mathcal{T}_l(P)) = \mathcal{M}_w(\mathcal{T}_l(P)) = \mathcal{T}_l(P)$. The list of all variations is as follows (first we have the main operations themselves):

1. $\mathcal{M}(P)$

2. $\mathcal{M}_w(P)$

3. $\mathcal{T}_w^\alpha(P)$

4. $\mathcal{T}_l^k(P)$

5. $\mathcal{T}_w(P)$

6. $\mathcal{T}_l(P)$

7. $\mathcal{T}_w^\alpha(\mathcal{T}_l^k(P)) = \mathcal{T}_l^k(\mathcal{T}_w^\alpha(P)) = \mathcal{T}_w^\alpha(P) \cap \mathcal{T}_l^k(P)$

8. $\mathcal{M}(\mathcal{T}_l^k(P) = \mathcal{T}_l^k(\mathcal{M}(P))$

9. $\mathcal{M}_w(\mathcal{T}_l^k(P) = \mathcal{T}_l^k(\mathcal{M}_w(P))$

10. $\mathcal{T}_w^\alpha(\mathcal{M}(P))$

11. $\mathcal{M}(\mathcal{T}_w^\alpha(P))$

12. $\mathcal{T}_w^\alpha(\mathcal{M}_w(P)) = \mathcal{M}_w(\mathcal{T}_w^\alpha(P)) = \mathcal{T}_w^\alpha(P) \cap \mathcal{M}_w(P)$

13. $\mathcal{T}_l(\mathcal{T}_w(P))$

14. $\mathcal{T}_w(\mathcal{T}_l(P))$

15. $\mathcal{M}(\mathcal{T}_w(P)) = \mathcal{M}_w(\mathcal{T}_w(P))$

16. $\mathcal{T}_w(\mathcal{M}(P)) = \mathcal{T}_w(\mathcal{M}_w(P))$

Note that we have: $\mathcal{T}_w^\alpha(\mathcal{M}(P)) \subseteq \mathcal{M}(\mathcal{T}_w^\alpha(P))$.

Also $\mathcal{T}_w(\mathcal{T}_l(P)) \neq \mathcal{T}_l(\mathcal{T}_w(P))$, although $\mathcal{T}_w^\alpha(\mathcal{T}_l^k(P)) = \mathcal{T}_l^k(\mathcal{T}_w^\alpha(P))$, since the implicit threshold values of the two sides are different. $\mathcal{T}_w(\mathcal{T}_l(P))$ is the set of preferences with the highest weight among the longest preferences of P, while $\mathcal{T}_l(\mathcal{T}_w(P))$ is the set of longest preferences among the preferences with highest weight in P.

Example 3.2. *Let $P = \{p_1, p_2, \ldots, p_9\}$ in which:*
$p_1 : (4)\ A \succ B \succ C \succ D \succ E \succ F$
$p_2 : (6)\ A \succ B \succ D \succ F$
$p_3 : (7)\ A \succ C \succ E$
$p_4 : (8)\ A \succ E$
$p_5 : (7)\ A \succ F$
$p_6 : (3)\ X \succ Y$
$p_7 : (5)\ X \succ Y \succ Z$
$p_8 : (6)\ R \succ S$
$p_9 : (4)\ A \succ B \succ C$

and the numbers in the parentheses are the weights of the preferences. We have:

1. $\mathcal{M}(P) = \{p_1, p_7, p_8\}$

2. $\mathcal{M}_w(P) = \{p_1, p_2, p_3, p_4, p_5, p_7, p_8\}$

3. $\mathcal{T}_w^5(P) = \{p_2, p_3, p_4, p_5, p_7, p_8\}$

4. $\mathcal{T}_l^3(P) = \{p_1, p_2, p_3, p_7, p_9\}$

5. $\mathcal{T}_w(P) = \{p_4\}$

6. $\mathcal{T}_l(P) = \{p_1\}$

7. $\mathcal{T}_w^\alpha(\mathcal{T}_l^k(P)) = \mathcal{T}_l^k(\mathcal{T}_w^\alpha(P)) = \mathcal{T}_w^\alpha(P) \cap \mathcal{T}_l^k(P) = \{p_2, p_3, p_7\}$

8. $\mathcal{M}(\mathcal{T}_l^3(P) = \mathcal{T}_l^3(\mathcal{M}(P)) = \{p_1, p_7\}$

9. $\mathcal{M}_w(\mathcal{T}_l^3(P) = \mathcal{T}_l^3(\mathcal{M}_w(P)) = \{p_1, p_2, p_3, p_7\}$

10. $\mathcal{T}_w^5(\mathcal{M}(P)) = \{p_7, p_8\}$

11. $\mathcal{M}(\mathcal{T}_w^5(P)) = \{p_2, p_3, p_7, p_8\}$

12. $\mathcal{T}_w^5(\mathcal{M}_w(P)) = \mathcal{M}_w(\mathcal{T}_w^5(P)) = \mathcal{T}_w^5(P) \cap \mathcal{M}_w(P) = \{p_2, p_3, p_4, p_5, p_7, p_8\}$

13. $\mathcal{T}_l(\mathcal{T}_w(P)) = \{p_4\}$

14. $\mathcal{T}_w(\mathcal{T}_l(P)) = \{p_1\}$

15. $\mathcal{M}(\mathcal{T}_w(P)) = \mathcal{M}_w(\mathcal{T}_w(P)) = \{p_4\}$

16. $\mathcal{T}_w(\mathcal{M}(P)) = \mathcal{T}_w(\mathcal{M}_w(P)) = \{p_8\}$

	$v_p(c_i \succ c_j)$	$v(c_i \succ c_j)$	$v'_P(c_i \succ c_j)$	$v(P)$
Method 0: v	$g(d_p(c_i, c_j))$	$\sum_{r=1}^{n} v_{p_r}(c_i \succ c_j)$	$\frac{v(c_i \succ c_j)}{g(d_P(c_i, c_j))}$	$\boldsymbol{T}_{i<j\leq k}\left\{v'_P(c_i \succ c_j)\right\}$
Method 1: v_μ	$g(d_p(c_i, c_j))$	$\sum_{r=1}^{n} v_{\mu_{p_r}}(c_i \succ c_j)$	$\frac{v_\mu(c_i \succ c_j)}{g(d_P(c_i, c_j))}/\mu(l(P))$	$\boldsymbol{T}_{i<j\leq k}\left\{v'_{\mu P}(c_i \succ c_j)\right\}$
Method 2: $\tilde{v}$	$f(d_p(c_i, c_j), l(p))$	$\sum_{r=1}^{n} \tilde{v}_{p_r}(c_i \succ c_j)$	$\frac{\tilde{v}(c_i \succ c_j)}{f(d_P(c_i, c_j), l(P))}$	$\boldsymbol{T}_{i<j\leq k}\left\{\tilde{v}'_P(c_i \succ c_j)\right\}$
Method 3: $v_\#$	$v_t(c_i \succ c_j) = \#\{p_r \mid d_{p_r}(c_i, c_j) = t\}$		$\frac{\sum_{r=1}^{m-1} v_r(c_i \succ c_j).\min(g(r), g(d_P(c_i, c_j)))}{g(d_P(c_i, c_j))}$	$\boldsymbol{T}_{i<j\leq k}\left\{v'_{\# P}(c_i \succ c_j)\right\}$

$$\boldsymbol{T} : M_\rho(x_1, ..., x_n) = (\frac{1}{n}\sum_{i=1}^{n} x_i^\rho)^{1/\rho};$$

$$M_0(x_1, ..., x_n) = \sqrt[n]{\prod_{i=1}^{n} x_i}\ ,$$

$$M_\infty(x_1, ..., x_n) = \max(x_1, ..., x_n),$$

$$M_{-\infty}(x_1, ..., x_n) = \min(x_1, ..., x_n).$$

$$\mu_m(k) = \max_{1\leq\sigma_1<\cdots<\sigma_k\leq m}\left\{\boldsymbol{T}_{i<j\leq k}\left\{\frac{g(\sigma_j - \sigma_i)}{g(j-i)}\right\}\right\}.$$

$$f(d, l) = g(d)/g(l-1)$$

Chapter 4

Properties

In this chapter, we introduce a number of axiomatic properties for assessing our model and its various methods. Most of these properties are generalizations of classic properties in social choice and voting theory; others are somewhat new, due to the new parameters that the model has.

For the classic properties, considering the more dimensions the model has, one can find different ways of generalizing or modifying an axiomatic property, in different directions or aspects. Here we present the generalizations which seem most natural and promising, and for a few of them, more than one modification is presented. For each of these properties, we will try to find out whether our model satisfies that property or not.

Since we have introduced different methods of assigning weights and many combinations of filters for refinement; the number of all possible TAPs, even in their general settings (any monotonic function g or f, and any generalized mean function), is too large to evaluate a property for all cases (at least in this thesis). However, since TAPs' behaviors vary for different filters, we will try to cover the basic ones (namely: $\mathcal{M}$, $\mathcal{M}_w$, $\mathcal{T}_l^k$, $\mathcal{T}_l$, $\mathcal{T}_w^\alpha$, $\mathcal{T}_w$). But, as we will see, for some properties the whole problem gets too complicated and we leave their evaluation for some cases for future work.

Although we proposed four weight-assigning methods, here we investigate the properties of the first three of them. We put aside Method 3 because it differs from the rest in that it is not really a 1-degree two-phase procedure.

First, we need some more notation and definitions to facilitate the presentation of the properties.

4.1 Notation

In the last chapter, the implicit assumption was that the input profile is given. If we want to specify that a function or variable corresponds to a particular profile $\mathcal{R}$, we put $\mathcal{R}$ as a superscript of that function or variable.

For instance, $\boldsymbol{P}^{\mathcal{R}}$ denotes the output trends for input profile $\mathcal{R}$; and $v^{\mathcal{R}}(P)$ is the weight of trend P produced by the input profile $\mathcal{R}$.

Similarly, we denote the weight of a trend P in the particular trend set $\boldsymbol{P}$ by $v_{\boldsymbol{P}}(P)$.

For a preference p, we have defined $S(p)$ as the set of all alternatives in p. We use the same notation for the set of all alternatives in profile $\mathcal{R}$ (or output trends $\boldsymbol{P}$), that is: $S(\mathcal{R}) = \bigcup_{p\in\mathcal{R}} S(p)$

We also use $s^{\mathcal{R}}(p)$ to show the set of all voters in $\mathcal{R}$ for which p is a subpreference of their preference (here $\mathcal{R}$ is not necessarily the input profile and it can be any other set of preferences like $\boldsymbol{P}$. In that case $s^{\boldsymbol{P}}(p)$ is the set of all preferences in $\boldsymbol{P}$ such that p is their subpreference). This is mostly used for the case where the set of voters prefer x to y: $s^{\mathcal{R}}(x \succ y)$. We might use this with only an alternative instead of a preference p, $s^{\mathcal{R}}(x)$, for the set of voters who have x in their preference.

Each preference of length k contains $\binom{k}{2}$ pairwise comparisons. We denote the set of all of these pairwise comparisons in a preference p with $C(p)$; i.e., $C(p) = \{(x, y) | x \succ_p y\}$. Similarly, for an input profile $\mathcal{R}$ (or set of output trends $\boldsymbol{P}$), we denote the set of all pairwise comparisons with $C(\mathcal{R})$ (or $C(\boldsymbol{P})$); i.e., $C(\mathcal{R}) = \bigcup_{p\in\mathcal{R}} C(p)$.

We call the first (best) alternative of a preference its *top*; that is: for a preference p, $a = top(p) \iff \forall x \in p, x \neq a : a \succ_p x$.

For any k, $top_k(p)$ is the preference of k first (best) alternatives in p. That is, if $p = p_1 \succ \cdots \succ p_k \succ \cdots \succ p_l$ then $top_k(p) = q$ with $q = p_1 \succ \cdots \succ p_k$. We call q the *top-k* subpreference of p.

We may restrict the preferences in $s(p)$ to only preferences for which p is their top subpreference (or in case of $s(x)$, x is their top member). We denote this as $s^{\mathcal{R}}_{top}(p)$ which means the set of voters whose vote in $\mathcal{R}$ has p as top subpreference.

We previously defined a profile as an n-tuple of preferences. If the voters are not important but only the votes, we may consider it as a multiset of preferences. In that case, we can also write the input profile as a set of weighted votes, in which the weights are the number of votes of that kind. More formally, $\mathcal{R} = \{(p_1, v_1), (p_2, v_2), \ldots, (p_k, v_k)\}$, in which there are v_i votes of type p_i and $\sum_{i=1}^{k} v_i = n$. This latter way of representing an input profile conforms to the way we represent output trends.

If we have two disjoint sets of voters with profiles $\mathcal{R}_1$ and $\mathcal{R}_2$, we can merge them into a single profile $\mathcal{R}$, written as $\mathcal{R} = \mathcal{R}_1 \oplus \mathcal{R}_2$. Formally, if $\mathcal{R}_1 = (p_1, \ldots, p_{n_1})$ and $\mathcal{R}_2 = (q_1, \ldots, q_{n_2})$, then $\mathcal{R}_1 \oplus \mathcal{R}_2 = (p_1, \ldots, p_{n_1}, q_1, \ldots, q_{n_2})$. If the preferences are weighted (as in the input profile of weighted votes or output trends) we do as follows: If the preferences in $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are from the set of preferences $\{p_1, \ldots, p_k\}$ and each of these preferences occurred at least in one of the two preference sets, i.e., $\boldsymbol{P}_1 = \{(p_{\sigma_1}, v_1(p_{\sigma_1})), \ldots, (p_{\sigma_{k_1}}, v_1(p_{\sigma_{k_1}}))\}$ and $\boldsymbol{P}_2 = \{(p_{\sigma'_1}, v_2(p_{\sigma'_1})), \ldots, (p_{\sigma'_{k_2}}, v_2(p_{\sigma'_{k_2}}))\}$, in which σ and σ' are permutations on integers 1 to k; then $\boldsymbol{P}_1 \oplus \boldsymbol{P}_2 = \{(p_1, v_1(p_1) + v_2(p_1)), \ldots, (p_k, v_1(p_k) + v_2(p_k))\}$.

For a set of weighted preferences $\boldsymbol{P_w}$, $\boldsymbol{P} = U(\boldsymbol{P_w})$ gives the set of the preferences without their weights.

As we have explained in the previous chapter, we may filter preferences in the result of any TAP with different methods in the refinement phase. We denote the results of this phase with calligraphic letters, so if $\boldsymbol{P}$ ($\boldsymbol{P_w}$) is the set of all (weighted) trends before refinement, the end product of the model after the refinement phase is $\boldsymbol{\mathcal{P}}$ ($\boldsymbol{\mathcal{P}_w}$).

Respectively, we denote the function of a TAP without (before) the refinement phase by $\boldsymbol{F}$; and the complete function, including refinement, by $\boldsymbol{\mathcal{F}}$.

For the three main methods we have introduced, if we want to specify what weight calculating method the TAP is using, we use the following notation: $\boldsymbol{F}_v$ (Method 0), $\boldsymbol{F}_{v_\mu}$ (Method 1), $\boldsymbol{F}_{\tilde{v}}$ (Method 2).

A TAP is characterized completely with its weight-assigning method and its refinement filter. For example, $\mathcal{M}(\boldsymbol{F}_{v_{\min}}^{g})$ is a TAP of Method 0 which uses the function g and min and its results are refined by $\mathcal{M}$, that is, its output consists of only maximal trends. When functions g or $\boldsymbol{T}$ are fixed or not important in a particular case, we may specify the TAP with only the type of method it uses and the refinement filter. For example, $\mathcal{M}(\boldsymbol{F}_{v_\mu})$ can be any TAP which uses Method 1 and is refined by $\mathcal{M}$.

Most of the properties that we will investigate are centered around the existence and shape of the trends rather than their weights. So since $\mathcal{M}$, $\mathcal{T}_l$, and $\mathcal{T}_l^k$ are independent from the weights of the trends, we usually do not specify the weight-assigning method of their corresponding TAPs unless it is necessary (e.g., when the weights play a crucial role in the property). Therefore, in these cases, we treat these refinement filters as a TAP; e.g., if we say $\mathcal{M}$ satisfies a property, it means that all the three TAPs $\mathcal{M}(\boldsymbol{F}_v)$, $\mathcal{M}(\boldsymbol{F}_{v_\mu})$ and $\mathcal{M}(\boldsymbol{F}_{\tilde{v}})$ satisfy that property.

This is also the case for the TAPs without refinement. Similarly, when weights do not matter for a property, we use $\boldsymbol{F}$ as the representative for all three TAPs: $\boldsymbol{F}_v$, $\boldsymbol{F}_{v_\mu}$, and $\boldsymbol{F}_{\tilde{v}}$.

In general, for each property, we try to check all the TAPs of any of the three weight-assigning methods with any of the basic refinement filters (and also without any refinement). These are $3 \cdot 7 = 21$ different classes of TAPs (each can have different functions for g and $\boldsymbol{T}$). In this chapter, we only deal with these TAPs. So, when we say a property holds for all TAPs, we mean this set of TAPs. When the property is not sensitive to the weights of the trends, the number of these TAPs can be reduced to 13, namely: $\boldsymbol{F}$, $\mathcal{M}$, $\mathcal{T}_l^k$, $\mathcal{T}_l$, $\mathcal{M}_w(\boldsymbol{F}_v)$, $\mathcal{T}_w^\alpha(\boldsymbol{F}_v)$, $\mathcal{T}_w(\boldsymbol{F}_v)$, $\mathcal{M}_w(\boldsymbol{F}_{v_\mu})$, $\mathcal{T}_w^\alpha(\boldsymbol{F}_{v_\mu})$, $\mathcal{T}_w(\boldsymbol{F}_{v_\mu})$, $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$, $\mathcal{T}_w^\alpha(\boldsymbol{F}_{\tilde{v}})$, and $\mathcal{T}_w(\boldsymbol{F}_{\tilde{v}})$. To avoid repeating this long list of TAPs, we may call them *major TAPs*.

4.2 Preliminary Propositions

First we present a few basic propositions which we will use in the rest of the chapter frequently. They also may provide an intuition for the behavior of the basic elements of the model which can lead to understanding why some standard properties do not work here, despite what it might seem like at first glance. In all the formulas in this section, $\boldsymbol{F}$ is a TAP without a refinement phase.

Proposition 4.1. *For any TAP $\boldsymbol{F}$ we have:*
$\forall \mathcal{R} : \mathcal{R} \subseteq \boldsymbol{F}(\mathcal{R})$,
$\forall \mathcal{R} : C(\mathcal{R}) = C(\boldsymbol{F}(\mathcal{R})) = C(\mathcal{M}(\boldsymbol{F}(\mathcal{R}))) = C(\mathcal{M}_w(\boldsymbol{F}(\mathcal{R})))$.

Proof. Each input preference is a potential trend itself. The model produces all possible trends, so all input preferences will be part of the set of output trends (before the refinement phase). Hence: $\mathcal{R} \subseteq \boldsymbol{F}(\mathcal{R})$.

For the first part, $\mathcal{R} \subseteq \boldsymbol{F}(\mathcal{R})$ implies $C(\mathcal{R}) \subseteq C(\boldsymbol{F}(\mathcal{R}))$; and since the procedure does not create new comparisons for pairs, $C(\mathcal{R}) = C(\boldsymbol{F}(\mathcal{R}))$.

Similarly, $\mathcal{M}(\boldsymbol{F}(\mathcal{R})) \subseteq \boldsymbol{F}(\mathcal{R})$ implies $C(\mathcal{M}(\boldsymbol{F}(\mathcal{R}))) \subseteq C(\boldsymbol{F}(\mathcal{R}))$. $\mathcal{M}(\boldsymbol{F}(\mathcal{R}))$ eliminates only trends which are subpreferences of other trends in $\mathcal{M}(\boldsymbol{F}(\mathcal{R}))$. So all pairwise comparisons of each of the eliminated trends exist in at least one of the trends in $\mathcal{M}(\boldsymbol{F}(\mathcal{R}))$, this means that $\mathcal{M}$ does not shrink the set of pairwise comparisons; hence $C(\boldsymbol{F}(\mathcal{R})) = C(\mathcal{M}(\boldsymbol{F}(\mathcal{R})))$.

Using the same argument, we have $C(\boldsymbol{F}(\mathcal{R})) = C(\mathcal{M}_w(\boldsymbol{F}(\mathcal{R})))$. So: $C(\mathcal{R}) = C(\boldsymbol{F}(\mathcal{R})) = C(\mathcal{M}(\boldsymbol{F}(\mathcal{R}))) = C(\mathcal{M}_w(\boldsymbol{F}(\mathcal{R})))$. □

Proposition 4.2. *For any TAP $\boldsymbol{F}$ and any preference P:*
$\forall \mathcal{R} : P \in \boldsymbol{F}(\mathcal{R}) \iff C(P) \subseteq C(\boldsymbol{F}(\mathcal{R}))$,
$\forall \mathcal{R}_1, \mathcal{R}_2 : C(\mathcal{R}_1) = C(\mathcal{R}_2) \implies U(\boldsymbol{F}(\mathcal{R}_1)) = U(\boldsymbol{F}(\mathcal{R}_2))$.

Proof. The left-to-right direction of the first equivalence is trivial. For the opposition direction: $C(P) \subseteq C(\boldsymbol{F}(\mathcal{R})) = C(\mathcal{R})$, i.e., all the pairwise comparisons of P have support in $\mathcal{R}$ and so by definition $P \in \boldsymbol{F}(\mathcal{R})$.

Since $C(\boldsymbol{F}(\mathcal{R})) = C(\mathcal{R})$, we substitute $C(\mathcal{R})$ in the first equivalence and from $C(\mathcal{R}_1) = C(\mathcal{R}_2)$ we have: $\forall P \in \boldsymbol{F}(\mathcal{R}_1) \iff C(P) \subseteq C(\mathcal{R}_1) \iff C(P) \subseteq C(\mathcal{R}_2) \iff P \in \boldsymbol{F}(\mathcal{R}_2)$. So, if $C(\mathcal{R}_1) = C(\mathcal{R}_2)$, then $U(\boldsymbol{F}(\mathcal{R}_1)) = U(\boldsymbol{F}(\mathcal{R}_2))$. □

Proposition 4.3. *For any TAP $\boldsymbol{F}$ and input profiles $\mathcal{R}_1$ and $\mathcal{R}_2$:*
$\exists \mathcal{R} : \mathcal{R}_2 = \mathcal{R}_1 \oplus \mathcal{R} \implies C(\mathcal{R}_1) \subseteq C(\mathcal{R}_2)$,
$C(\mathcal{R}_1) \subseteq C(\mathcal{R}_2) \implies U(\boldsymbol{F}(\mathcal{R}_1)) \subseteq U(\boldsymbol{F}(\mathcal{R}_2))$.

Proof. $\mathcal{R}_2$ has all the preferences in $\mathcal{R}_1$, so it has all the pairwise comparisons of $\mathcal{R}_1$, i.e., $C(\mathcal{R}_1) \subseteq C(\mathcal{R}_2)$. If $\mathcal{R}_2$ has all the pairwise comparisons of $\mathcal{R}_1$, it produces all the trends that $\mathcal{R}_1$ can produce; and possibly some additional trends, so, $U(\boldsymbol{F}(\mathcal{R}_1)) \subseteq U(\boldsymbol{F}(\mathcal{R}_2))$. □

Proposition 4.4. *For any two trend sets* $\boldsymbol{P}_1$ *and* $\boldsymbol{P}_2$*, if* $U(\boldsymbol{P}_1) = U(\boldsymbol{P}_2)$ *then* $\mathcal{M}(\boldsymbol{P}_1) = \mathcal{M}(\boldsymbol{P}_2)$*,* $\mathcal{T}_l^k(\boldsymbol{P}_1) = \mathcal{T}_l^k(\boldsymbol{P}_2)$ *and* $\mathcal{T}_l(\boldsymbol{P}_1) = \mathcal{T}_l(\boldsymbol{P}_2)$*. Additionally if* $\forall P \in \boldsymbol{P}_1 : v_{\boldsymbol{P}_1}(P) \leq v_{\boldsymbol{P}_2}(P)$ *then* $\mathcal{T}_w^\alpha(\boldsymbol{P}_1) \subseteq \mathcal{T}_w^\alpha(\boldsymbol{P}_2)$*. Such a relation does not hold for* $\mathcal{T}_w$ *and* $\mathcal{M}_w$*.*

Proof. Since the set of trends are the same (without considering the weights), their length and the maximality relations remain intact. So their maximal set and length threshold (and top length) set will be the same; i.e., $\mathcal{M}(\boldsymbol{P}_1) = \mathcal{M}(\boldsymbol{P}_2)$, $\mathcal{T}_l(\boldsymbol{P}_1) = \mathcal{T}_l(\boldsymbol{P}_2)$ and $\mathcal{T}_l^k(\boldsymbol{P}_1) = \mathcal{T}_l^k(\boldsymbol{P}_2)$. Since the weights are not lower in $\boldsymbol{P}_2$, the threshold on weights would not remove any member of $\mathcal{T}_w^\alpha(\boldsymbol{P}_1)$, and higher weights may push new trends above the threshold; so: $\mathcal{T}_w^\alpha(\boldsymbol{P}_1) \subseteq \mathcal{T}_w^\alpha(\boldsymbol{P}_2)$.

The trends with maximal weight can be different in the two sets and so the top weight sets can consist of completely different trends. Thus, there is no such a relation for top weight sets ($\mathcal{T}_w$).

For $\mathcal{M}_w$, assume a simple case where $\boldsymbol{P}_1 = \{P, Q\}$ such that $P \sqsubset Q$. Now, consider a case in which $v_{\boldsymbol{P}_1}(P) > v_{\boldsymbol{P}_1}(Q)$, so $\mathcal{M}_w(\boldsymbol{P}_1) = \{P, Q\}$. Increasing the value of Q in $\boldsymbol{P}_2$ can cause a situation in which: $v_{\boldsymbol{P}_2}(P) \leq v_{\boldsymbol{P}_2}(Q)$, i.e., $P \sqsubset_w Q$ and so $\mathcal{M}_w(\boldsymbol{P}_2) = \{Q\}$. In other words, rather surprisingly, increasing the values of trends may result in elimination of some of them.

The reverse is also possible, i.e., increasing the weights can add new trends to $\mathcal{M}_w(\boldsymbol{P}_2)$. For this, similarly consider the case in which $v_{\boldsymbol{P}_1}(P) \leq v_{\boldsymbol{P}_1}(Q)$, i.e., $P \sqsubset_w Q$ and so $\mathcal{M}_w(\boldsymbol{P}_1) = \{Q\}$. Now, if $v_{\boldsymbol{P}_2}(P) > v_{\boldsymbol{P}_2}(Q)$, P will also become a member of $\mathcal{M}_w(\boldsymbol{P}_2)$; i.e., $\mathcal{M}_w(\boldsymbol{P}_2) = \{P, Q\}$. □

Proposition 4.5. *For any two trend sets* $\boldsymbol{P}_1$ *and* $\boldsymbol{P}_2$*, if* $U(\boldsymbol{P}_1) \subseteq U(\boldsymbol{P}_2)$ *then* $\mathcal{T}_l^k(\boldsymbol{P}_1) \subseteq \mathcal{T}_l^k(\boldsymbol{P}_2)$*. Additionally if* $\forall P \in \boldsymbol{P}_1 : v_{\boldsymbol{P}_1}(P) \leq v_{\boldsymbol{P}_2}(P)$ *then* $\mathcal{T}_w^\alpha(\boldsymbol{P}_1) \subseteq \mathcal{T}_w^\alpha(\boldsymbol{P}_2)$*. Such a relation does not hold for* $\mathcal{M}$*,* $\mathcal{M}_w$ *and* $\mathcal{T}_w$*.*

Proof. For $\mathcal{T}_l^k$ and $\mathcal{T}_w^\alpha$, we have the same argument as in the previous proposition; but since we may have new trends, the equality for $\mathcal{T}_l^k$ to be replaced by subset relation.

For $\mathcal{M}$, a new trend could be a new maximal trend that adds to the number of trends in $\mathcal{M}(\boldsymbol{P})$, or it can be a subprefrence of an existing trend that does not change the set. But, in the first case, it is also possible that some maximal trends become subpreferences of this new trend; and so the new maximal set loses some of its members.

For $\mathcal{M}_w$, we can use the same argument, or deduce directly from the previous proposition (also for $\mathcal{T}_w$). □

4.3 Axiomatic Properties

Here we present several axiomatic properties (or evaluation criteria) for our model.

4.3.1 Anonymity

In the proposed TAP model, since the voters do not play any role in the model and we just deal with their reported preferences (obviously equally), the procedure is anonymous. That is, changing the name of the voters does not affect the procedure's result.

4.3.2 Neutrality

Like anonymity, neutrality also holds for our model, since the alternatives are treated equally in the procedure. So if we change the name of the candidates, the output trends will be the same if we change the candidates' names accordingly.

4.3.3 Homogeneity

Classic homogeneity says that replicating all voters uniformly does not affect the election outcome. A generalization of it would be: replicating each voter with the same number of similar voters does not change the resulting trends. The possible changes in the trends' weights can be fixed with proper normalization.

Theorem 4.1. *For any TAP, substituting each input preference with k identical preferences does not affect the resulting trends and their ordering. The new trends' weights will be k times their corresponding trends.*

Proof. If we replace a preference P with k identical Ps, the corresponding binary comparisons and their weights will remain the same and will just be repeated k times. Because we add up these weights in the algorithm, the resulting weights for each binary comparison at the end of the decomposition phase would be k times the initial case (without k copies). Since this does not add or remove any binary comparisons, it does not change the trend structure. So the resulting trends will remain the same, only all weights will be k times the initial case. Note that in case of $\mathcal{T}_w^\alpha$, we assume that a normalization factor (of $1/k$) will be applied at the end (or the threshold value will be multiplied by k); otherwise $\mathcal{T}_w^\alpha(\boldsymbol{F}(\mathcal{R})) \subseteq \mathcal{T}_w^\alpha(\boldsymbol{F}(k\mathcal{R}))$, since the new weights may push new trends above the threshold. □

4.3.4 Unanimity

Since in most cases of classical voting (or preference aggregation) systems there is only one outcome (either a winner or an aggregated preference), unanimity is usually defined as: if there is a consensus among all voters, that would be the system's outcome. Also, as in most settings, Pareto optimality implies unanimity; in the literature they are considered equivalent in many cases. However, due to the more complex structure, we define them differently here. Also, since in general, the output consists of more than one trend (with different weights); it is not really clear what the winner is, and even in the consensus case there could be more than one trend in the output.

So, we define two kinds of unanimity properties. Unanimity, which means the consensus preference will be one of the output trends; and unique unanimity, which means the consensus preference will be the only output trend. Note that uniqueness needs a refinement on the set of trends.

Definition 4.1 (Unanimity)**.** *A TAP is unanimous if when all voters have the same preference, that preference is part of the outcome:*

$$\mathcal{R} = (p, \ldots, p) \implies p \in \boldsymbol{\mathcal{F}}(\mathcal{R}). \tag{4.1}$$

We say a TAP has the unique unanimity *property if the consensus preference is the only trend of the output:*

$$\mathcal{R} = (p, \ldots, p) \implies U(\boldsymbol{\mathcal{F}}(\mathcal{R})) = \{p\}. \tag{4.2}$$

By definition, unanimity is weaker than unique unanimity.

Theorem 4.2. *The TAP's: $\mathcal{M}(\boldsymbol{F})$, $\mathcal{T}_l$, $\mathcal{M}_w(\boldsymbol{F}_{v_\mu})$, $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$ are uniquely unanimous; $\boldsymbol{F}$, $\mathcal{T}_l^k$, $\mathcal{M}_w(\boldsymbol{F}_v)$, $\mathcal{T}_w^\alpha(\boldsymbol{F}_{v_\mu})$, $\mathcal{T}_w^\alpha(\boldsymbol{F}_{\tilde{v}})$ are unanimous and $\mathcal{T}_w^\alpha(\boldsymbol{F}_v)$ is not unanimous.*

Proof. An election with unanimous votes of k voters is equivalent to a singular election in which the vote is replicated k times. Therefore, using homogeneity, we consider a singular election and the results will be the same.

If there is only one voter with preference p then $U(\boldsymbol{F}(\mathcal{R})) = \{q | q \sqsubseteq p\}$. So, $p \in \boldsymbol{F}(\mathcal{R})$, but it is not the only member of $\boldsymbol{F}(\mathcal{R})$; i.e., $\boldsymbol{F}$ is unanimous.

The maximal set and top length set of $\boldsymbol{F}(\mathcal{R})$ are both $\{p\}$ and so $\mathcal{M}$ and $\mathcal{T}_l$ are uniquely unanimous.

p is a member of $\mathcal{M}_w(\boldsymbol{F}(\mathcal{R}))$, so $\mathcal{M}_w$ is unanimous. Now to check if it is unique or not, we need to consider the scoring method. If the method is informative, the weight of p is maximal; and so p is the unique member of $\mathcal{M}_w(\boldsymbol{F}(\mathcal{R}))$, otherwise there could be other trends in the set. Hence, $\mathcal{M}_w(\boldsymbol{F}_{v_\mu})$ and $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$ are uniquely unanimous and $\mathcal{M}_w(\boldsymbol{F}_v)$ is unanimous.

For the weight threshold, in Method 0, p has the minimum weight. So for any threshold higher than its weight, p is not a member of $\mathcal{T}_w^\alpha(\boldsymbol{F}_v(\mathcal{R}))$ and therefore $\mathcal{T}_w^\alpha(\boldsymbol{F}_v)$ is not unanimous. On the other hand, since Method 1 and Method 2 are informative, p is the member of any weight threshold set. Because in these two methods other trends could gain the same maximum weight as p has, the weight threshold (and top weight set) will not guaranty the uniqueness of p. Hence, $\mathcal{T}_w^\alpha(\boldsymbol{F}_{v_\mu})$ and $\mathcal{T}_w^\alpha(\boldsymbol{F}_{\tilde{v}})$ are unanimous. □

4.3.5 Groundedness

We can see a common aspect in this model and *ontology merging* in the sense that in both, the input and output are partial and the framework itself does not rule out appearance of an alternative (formula) in the outcome without any direct support in the input.

An axiom, called *groundedness*, has been introduced in this regard [21]. Here we use a similar idea to define groundedness for our model.

Definition 4.2 (Groundedness). *We call a TAP grounded if no alternative appears in the output trends unless it occurs in at least one input preference.*

$$\forall \mathcal{R} : S(\boldsymbol{\mathcal{F}}(\mathcal{R})) \subseteq S(\mathcal{R}) \tag{4.3}$$

Note that, although our framework does not impose groundedness, our proposed methods do in all of the presented cases. These methods only produce a trend when all of its pairwise comparisons have occurred in the input preferences and so the same holds for the individuals.

Besides the methods that have been presented, there are possible scenarios within the model's framework which violate the groundedness property. For example, if we consider the alternatives that have not been compared in each voter's preference as the voters least preferred choices and then apply the rest of the procedure as before, then we may end up with alternatives in output trends which have not been in any of the input preferences.

4.3.6 Existence

Here we deal with partial preferences both in the input and in the output, so the existence of a particular alternative in the output in different situations is noteworthy. For example, an existence axiom could be stated as: If an alternative occurs in at least one of the input preferences, it should appear in at least one of the output trends; i.e., for all x:

$$\exists p_i : x \in S(p_i)) \implies \exists P_i : x \in S(P_i).$$

However, since this formulation is almost equivalent to another axiom that we will introduce later (in Section 4.3.8), we do not define this as a separate axiom. Here we use a universal quantifier for the premise and define two versions; one requires the existence of the alternative in at least one trend and the other the existence in all output trends.

Therefore we define the *weak existence* property as follows:

Definition 4.3 (Weak existence). *We say a TAP has the weak existence property if, when there is an alternative x which occurs in all input preferences, then x appears in at least one output trend.*

$$\forall x[\forall p_i : x \in S(p_i) \implies x \in S(\boldsymbol{\mathcal{P}})] \tag{4.4}$$

Accordingly, the *strong existence* property (which we may call just existence) is defined as follows:

Definition 4.4 (Strong existence)**.** *We say a TAP has the strong existence property if, when there is an alternative x which occurs in all input preferences, then x appears in all output trends.*

$$\forall x[\forall p_i : x \in S(p_i) \implies \forall P_i \in \boldsymbol{\mathcal{P}} : x \in S(P_i)] \tag{4.5}$$

Lemma 4.1. *If there is an alternative x which occurs in all input preferences, then for any output trend in $\boldsymbol{P}$, x appears in the trend or in a superpreference of that trend in $\boldsymbol{P}$. That is, if $\boldsymbol{P} = \boldsymbol{F}(\mathcal{R})$,*
$\forall x[\forall p_i : x \in S(p_i) \implies \forall P \in \boldsymbol{P} : \exists P' \in \boldsymbol{P} : P \sqsubseteq P' \wedge x \in S(P')]$

Proof. We show that for any $P \in \boldsymbol{P}$ such that $x \notin P$, there is a $P' \in \boldsymbol{P}$ with $P \sqsubset P'$ and $x \in P'$.

Let $P : c_1 \succ ... \succ c_k$ be a trend without x. If $\forall i = 1, \ldots, k : \exists p_j : x \succ_{p_j} c_i$ then $P' : x \succ c_1 \succ ... \succ c_k$ is a trend. If not, there is at least one c_i that x never beats it in any input preference. Now, suppose c_t is the last alternative of this kind in P; i.e., $v(x \succ c_t) = 0$ and for $i = t+1, \ldots, k : v(x \succ c_i) > 0$. This means that in all preferences containing x and c_t, $c_t \succ x$; and since x appears in all input preferences, whenever we have c_t in a preference, x comes somewhere after that. Considering the place of c_t in P, for each $i = 1, \ldots, t-1$, there is at least one preference in which: $c_i \succ c_t$; and since $c_t \succ x$ for every preference with c_t, we have $c_i \succ x$ for all $i = 1, \ldots, t-1$. Hence, $P' : c_1 \succ ...c_t \succ x \succ \cdots \succ c_k$, is also a trend. □

Corollary 4.1. *If a candidate exists in all input preferences, it will exist in all maximal output trends and top length trends. That is,*
$\forall x[\forall p_i : x \in S(p_i) \implies \forall P_i \in \mathcal{M}(\boldsymbol{P}) : x \in S(P_i)]$,
$\forall x[\forall p_i : x \in S(p_i) \implies \forall P_i \in \mathcal{T}_l(\boldsymbol{P}) : x \in S(P_i)]$.

Proof. From the lemma, each trend in $\boldsymbol{P}$ is a subpreference of a trend that contains x. So, all maximal trends contain x. Also, since all top length trends are maximal by definition, all trends in the top length set also contain x. □

Theorem 4.3. *All major TAPs, except for weight thresholds, are weakly existent. $\mathcal{M}$ and $\mathcal{T}_l$ are strongly existent.*

Proof. The previous corollary implies that $\mathcal{M}$ and $\mathcal{T}_l$ are strongly existent. Since $\mathcal{M}$ and $\mathcal{T}_l$ are subsets of $\mathcal{M}_w$ and $\mathcal{T}_l^k$ respectively, $\mathcal{M}_w$ and $\mathcal{T}_l^k$ are weakly existent. Likewise, $\boldsymbol{F}$ is weakly existent; and it is not strongly existent. The length threshold set ($\mathcal{T}_l^k$) is not strongly existent, because of the inclusion of extra trends in case of any threshold other than the maximum length.

When the weights play a role in filtering the output set, it is a bit more complicated. For each weight-assigning method, it depends on the functions g

(or f) and $\boldsymbol{T}$. Here, we will not explore all of these possible situations and will only present some significant cases.

For Method 0, consider the singular election with input preference $p : A \succ C \succ B$. We have (with any mean function), $v(A \succ B) = g(2)$, while $v(p) = 1$. So, when $g(2) > g(1) = 1$ (as it is almost always the case), this means that $A \succ B$ as an output trend is a weighted maximal trend for the output set of $\boldsymbol{F}_v$; i.e. $\mathcal{M}_w(\boldsymbol{F}_v)$ does not satisfy strong existence. In addition, any threshold greater than the weight of trend p would eliminate it and so $\mathcal{T}_w^\alpha(\boldsymbol{F}_v)$ is not even weakly existent.

Now, for the other two methods, first we examine the case of $\boldsymbol{T} = \min$, as it is the main choice for $\boldsymbol{T}$ and is the most promising one. Since these two methods are informative, the previous example will not work for them.

For Method 2, consider the following scenario. The input preferences are p_1 and p_2 such that $p_1 : A \succ B \succ C$ and $p_2 : A \succ C \succ B$. Now, we calculate the scores of three trends p_1, p_2 and $p : A \succ B$. The aggregated weight of pairwise comparisons are:

$v(A \succ B) = v(A \succ C) = g(1)/g(2) + g(2)/g(2) = 1 + 1/g(2)$, and

$v(B \succ C) = v(C \succ B) = g(1)/g(2) = 1/g(2)$.

Therefore, the weights of the three trends are:

$v(p) = 1 + 1/g(2)$,

$v(p_1) = \boldsymbol{T}\{v(A \succ B)/(1/g(2)), v(B \succ C)/(1/g(2)), v(A \succ C)/1\} = \boldsymbol{T}\{g(2) + 1, 1, 1 + 1/g(2)\}$, and by symmetry, $v(p_2) = v(p_1)$.

Now, if $\boldsymbol{T} = \min$ then $v(p_1) = 1$ and so $v(p) > v(p_1) = v(p_2)$. Thus, like Method 0, when $\boldsymbol{T} = \min$, $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$ is not strongly existent and $\mathcal{T}_w^\alpha(\boldsymbol{F}_{\tilde{v}})$ is not existent at all.

Note that for this example, depending on the function g, a range of choices for the power mean violate existence. For example, if $g(2) < \sqrt{2}$, all the power mean functions with the exponent of less than 1 (including M_1 which is the arithmetic mean) violate existence.

We can also see that if $\boldsymbol{T} = \max$, for any preference without x, there is a superpreference of it with x which has a greater weight. That is, when $\boldsymbol{T} = \max$, $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$ satisfies strong existence and $\mathcal{T}_w^\alpha(\boldsymbol{F}_{\tilde{v}})$ is weakly existent. □

4.3.7 Pareto Optimality

In classical voting theory, usually two kinds of Pareto optimality are defined. In most settings, the expected outcome of the two versions are the same; and the weaker has stronger requirements. As mentioned before, in the literature, Pareto optimality is sometimes considered equivalent with unanimity. In general, roughly speaking, the Pareto optimality property says: if all members of a society have "almost" the same opinion about two alternatives, the social outcome should do the same. The stricter definition of "almost", the stronger the axiom. For example, a weak version may require that if all voters prefer A to B, so does the society, while a strong version only requires that no one prefers B to A. Note that, in most classical settings, the conclusion part is the same for both versions because there is one social preference and usually this social preference is not partial.

Here we use a similar idea to generalize the concept of Pareto optimality with two different approaches. Note that, here the outcome consists of more than one trend and these trends are partial preferences. So we can have different versions in the premise and conclusion parts. Therefore, we cannot say that one axiom is mathematically weaker than the other and can by logically implied by the other axiom.

First, we define a property in our model which we call *existential Pareto optimality.*

Definition 4.5 (Existential Pareto optimality)**.** *A TAP is existentially Pareto optimal, if when some voters prefer an alternative to another, then there are some trends in which the former is preferred to the latter.*

$$\forall x, y[\exists p_i : x \succ_{p_i} y \implies \exists P_i : x \succ_{P_i} y] \tag{4.6}$$

The other version, which we call *universal Pareto optimality*, is obtained by replacing existential quantifiers in the axiom by universal quantifiers.

Definition 4.6 (Universal Pareto optimality)**.** *A TAP is universally Pareto optimal, if when all voters prefer an alternative to another, all output trends also prefer the former to the latter.*

$$\forall x, y[\forall p_i : x \succ_{p_i} y \implies \forall P_i : x \succ_{P_i} y] \tag{4.7}$$

Theorem 4.4. *The TAPs: $\boldsymbol{F}$, $\mathcal{M}(\boldsymbol{F})$, $\mathcal{M}_w(\boldsymbol{F}_v)$, $\mathcal{M}_w(\boldsymbol{F}_{v_\mu})$ and $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$ are existentially Pareto optimal. No TAP with a threshold filter is existentially Pareto optimal.*

Proof. If $\exists p_i : x \succ_{p_i} y$ then there is at least one trend P in $\boldsymbol{F}(\mathcal{R})$ such that $x \succ_P y$. This trend or one of its superpreferences will be in the trends' maximal set and weighted maximal sets by definition. So, $\boldsymbol{F}$, $\mathcal{M}(\boldsymbol{F})$, $\mathcal{M}_w(\boldsymbol{F}_v)$, $\mathcal{M}_w(\boldsymbol{F}_{v_\mu})$ and $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$ are existentially Pareto optimal.

For $\mathcal{T}_l^k$, consider a case in which the only preference containing x and y is $x \succ y$, and all other preferences are on alternatives other than these two. So, the only possible trend that contains x and y would be $x \succ y$ and this will be eliminated by any threshold on the length of the trends.

We can use a similar argument for the weight threshold in any of the three weight assigning methods. □

Like strong existence, we do not know whether all TAPs satisfy universal Pareto optimality. But we have this:

Theorem 4.5. *Universal Pareto optimality is strictly weaker than strong existence.*

Proof. Suppose that in all input preferences $x \succ y$. Then from strong existence, x and y exist in all output trends. Since $x \succ y$ is in all preferences (and so there is no support for $y \succ x$), we have $x \succ y$ in all output trends. That is, strong existence implies the universal Pareto optimality property.

Consider the case in which the input preferences are $x \succ y \succ z$ and $y \succ z \succ x$, and the output trend is $y \succ z$. This satisfies universal Pareto optimality, but violates strong existence. So, universal Pareto optimality is strictly weaker than strong existence. □

4.3.8 Covering

The existence property is about a specific alternative with a particular, rather unlikely, situation in the input profile. We also saw (in Section 4.3.6) that the existence property could be defined in another way, we define it here alongside another property (covering). It is reasonable to seek procedures that have all the compared alternatives of the input in their output.

Definition 4.7 (Covering). *We call a TAP* covering *if the output trends consist of exactly all the alternatives occurred in the input profile.*

$$\forall \mathcal{R} : S(\mathcal{R}) = S(\boldsymbol{P}) \tag{4.8}$$

The definition is equivalent to:

$$\forall x[x \in S(\mathcal{R}) \iff x \in S(\boldsymbol{P})] \tag{4.9}$$

We saw a possible formulation for the existence axiom:

$$\forall x[\exists p_i : x \in S(p_i) \implies \exists P_i : x \in S(P_i)] \tag{4.10}$$

This existence axiom is equivalent to the right-to-left direction of the Equation 4.9. The left-to-right direction is equivalent to the groundedness property.

So, if this existence axiom holds for a grounded TAP, the TAP satisfies the covering property. As we have seen, all TAPs that we are dealing with are grounded; so we just need to check the existence axiom.

Theorem 4.6. *The covering property holds for* $\boldsymbol{F}$ *and all the maximal filters. It does not, however, hold for length or weight threshold filters.*

Proof. We need to check the existence axiom. If x is in one input preference, that preference is a trend itself and that trend or its superpreference will remain in the set of maximal trends. So covering holds for maximal filters.

If x occurred in just one input preference (e.g., $x \succ y$), this will be the only trend representing x in the output. This could be eliminated easily by length or weight threshold filters. So covering does not hold for these thresholds. □

4.3.9 Partitioning

In this model, the size of the set of alternatives is very crucial from the computational aspect. Because the complexity of the problem is fairly high, it is very important to be able to find a way to reduce the size of the problem or divide the problem to a number of smaller subproblems. In this regard, we present the partitioning property. This is a rather trivial but definite solution if it is applicable. More detailed partitioning of the problem might result in a situation where only approximate solutions are possible. We define partitioning when we want to produce the aggregated trends for two (or more) sets of disjoint alternatives.

Definition 4.8 (Partitioning)**.** *We say a TAP has the partitioning property if when there are two sets of disjoint alternatives, computing the trend set of their corresponding input profiles separately and then merging the resulting trend sets or first merging the input profiles and then computing the trends, result in the same outcome.*

$$S(\mathcal{R}_1) \cap S(\mathcal{R}_2) = \emptyset \implies \boldsymbol{\mathcal{F}}(\mathcal{R}_1 \oplus \mathcal{R}_2) = \boldsymbol{\mathcal{F}}(\mathcal{R}_1) \oplus \boldsymbol{\mathcal{F}}(\mathcal{R}_2) \tag{4.11}$$

Obviously, this can be generalized for more than two disjoint sets:

$$\bigcap_{i=1}^{k} S(\mathcal{R}_k) = \emptyset \implies \boldsymbol{\mathcal{F}}(\mathcal{R}_1 \oplus \cdots \oplus \mathcal{R}_k) = \boldsymbol{\mathcal{F}}(\mathcal{R}_1) \oplus \cdots \oplus \boldsymbol{\mathcal{F}}(\mathcal{R}_k) \tag{4.12}$$

Theorem 4.7. *The TAPs: $\boldsymbol{F}$, $\mathcal{M}$, $\mathcal{M}_w(\boldsymbol{F}_v)$, $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$, $\mathcal{T}_l^k$, $\mathcal{T}_w(\boldsymbol{F}_v)$ and $\mathcal{T}_w(\boldsymbol{F}_{\tilde{v}})$ satisfy partitioning.*

Proof. Since the sets of alternatives are disjoint, putting them together will not add new pairwise comparisons. So, the trends remain intact. There also will not be any changes in their weights, except in Method 1 (due to the change in the value of μ). Hence, the set of all possible trends would be the union of the output trends of the two disjoint profiles. Since the (weighted) maximal relation does not change, $\mathcal{M}$, $\mathcal{M}_w(\boldsymbol{F}_v)$ and $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$ also satisfy the partitioning property. This is also the case for the threshold filters, because if a trend in any of the former profiles passed the threshold, it will still pass. So, $\mathcal{T}_l^k$, $\mathcal{T}_w^\alpha(\boldsymbol{F}_v)$ and $\mathcal{T}_w^\alpha(\boldsymbol{F}_{\tilde{v}})$ satisfy partitioning.

Note that both threshold filters satisfy partitioning when there is a specific value for the threshold. In case of top sets, since the maximum value (length or weight) depends on the corresponding set, the partitioning property is not satisfied. That is, if the maximum length or weight in the two sets are different, we cannot consider them as special cases of the threshold function, because there are in fact two different threshold functions in the equivalence. However we can fix this by adding an extra round of filtering; i.e, if $\mathcal{T}$ is a top set function (either top length or weight) and $\boldsymbol{F}$ is $\boldsymbol{F}_v$ or $\boldsymbol{F}_{\tilde{v}}$, we have:

$$S(\mathcal{R}_1) \cap S(\mathcal{R}_2) = \emptyset \implies \mathcal{T}(\boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2)) = \mathcal{T}(\mathcal{T}(\boldsymbol{F}(\mathcal{R}_1)) \oplus \mathcal{T}(\boldsymbol{F}(\mathcal{R}_2))) \tag{4.13}$$

Furthermore, for Method 1, we can also have partitioning with a minor change in the formula. Here, the problem is that the value of the μ depends on the total number of the alternatives in the profile. Consider the case where there is a subset of alternatives such that each of the input preferences is either totaly inside or outside of this set. Then if we calculate the value of the μ based on the size of this subset instead of the whole set, the weights remain the same in both cases. Note that, this change results in a different weight for each output trend. These new weights would even change the relative weights of the trends. For example, if the profile consists of two preferences, $p_1 : A \succ B \succ C$ and $p_2 : D \succ E$; the weight of the trend $D \succ E$ will be greater than the weight of the trend $A \succ B$, while they would have the same weight in the original method. □

In practice, using this property, can reduce the computation time of the trend aggregation significantly (if it is applicable). For any implementation of our method, we can first run a linear algorithm to partition the set of alternatives based on the input preferences, then apply the main algorithm on each part (and its corresponding preferences) separately.

4.3.10 Informativeness

We previously defined informativeness, in a very special case of singular elections, to characterize the weight-assigning methods. Here we define it as an axiomatic property for elections of any size. Note that, since here we define informativeness for the complete TAPs (i.e., with refinement phase), it can be the case that a TAP with an uninformative weight-assigning method satisfies the informativeness property.

Definition 4.9 (Isolated preference). *We call an input preference isolated if its set of alternatives is disjoint from the rest of the preferences in the input profile. That is, p is isolated if: $S(p) \cap S(\bigcup_{i \in \mathcal{X}, p_i \neq p} p_i) = \emptyset$.*

Definition 4.10 (Informativeness). *We call a TAP informative, if any isolated preference of the input profile is an output trend and has a greater weight than any of its subpreferences in the output trend set.*

$$\forall p \in \mathcal{R}[S(p) \cap S(\bigcup_{i \in \mathcal{X}, p_i \neq p} p_i) = \emptyset \implies \forall q \in \boldsymbol{\mathcal{P}}, q \sqsubseteq p : v(q) \leq v(p)] \tag{4.14}$$

Note that when p is not in the output trend set, then $v(p) = 0$ by definition.

We introduced Method 2 as a solution to develop informative TAPs. However, it is important to observe that not all the variants of Method 2 are informative. The informativeness of Method 2 depends on the functions f and $\boldsymbol{T}$. Here we will show that when $\boldsymbol{T} = \min$, the weight-assigning method is informative. But, this is not necessarily the case for all the other settings of Method 2.

Lemma 4.2. *In a singular election with input preference p, for Method 2 with $\boldsymbol{T} = \min$, p has a greater weight than all of its subpreferences. That is, for TAP $\boldsymbol{F}_{\tilde{v}}$ which is using the* min *function, we have:*
$\mathcal{R} = \{p\} \implies \forall q \sqsubseteq p : v(q) \leq v(p).$

Proof. Let $p : c_1 \succ \cdots \succ c_m$, and $q : c'_1 \succ \cdots \succ c'_k$ be an arbitrary subpreference of p. By definition, when p is the only input its weight as an output trend is one unit, i.e., $v(p) = 1$. So, we need to show that the weight of no trend would exceed 1.

If $c_1 \neq c'_1$ or $c_m \neq c'_k$, it means that $d_p(c'_1, c'_k) < m - 1$. Since g is non-decreasing, $v_p(c'_1, c'_k) = g(d_p(c'_1, c'_k))/g(m-1) < 1$. So, $v'_q(c'_1, c'_k) = v(c'_1, c'_k)/f(k-1, k-1) = v_p(c'_1, c'_k) < 1$. Since $\boldsymbol{T} = \min$, we have $v(q) \leq v'_q(c'_1, c'_k) < 1$.

Now, if the first and last alternatives of both p and q are the same, then $v'_q(c'_1, c'_k) = f(m-1, m-1)/f(k-1, k-1) = 1$. Because $\boldsymbol{T} = \min$, we have $v(q) \leq v'_q(c'_1, c'_k) = 1$. We can see that the equality only occurs when all alternatives in q are proportionally in the same place in p as they are in q. That is, only when $m - 1 = t(k - 1)$ and $c'_i = c_{t(i-1)+1}$. □

Theorem 4.8. *$\mathcal{M}(\boldsymbol{F})$ and $\mathcal{T}_l(\boldsymbol{F})$ are informative for all three weight-assigning methods. $\boldsymbol{F}$, $\mathcal{M}_w(\boldsymbol{F})$, $\mathcal{T}_l^k(\boldsymbol{F})$, $\mathcal{T}_w^\alpha(\boldsymbol{F})$ and $\mathcal{T}_l(\boldsymbol{F})$ are informative when their weight-assigning method is informative.*

Proof. If a TAP satisfies partitioning, the problem is reduced to a singular election. In case of Method 1, the order of the weights in each partition remains unchanged; so it also can be reduced to a singular election. This is also the case for the top length and weight sets. So, all of them are informative when their weight-assigning method is informative. That is, Method 1 and the settings of Method 2 which are informative (e.g., when $\boldsymbol{T} = \min$).

In addition, for $\mathcal{M}$ and $\mathcal{T}_l$, since they eliminate the subpreferences of each trend in their outcome; they are informative in all cases even when their weight-assigning method is not informative. □

4.3.11 Inclusiveness

We can see trend aggregation from two different points of view. First, as a system that finds the most popular trends in the society. This can be seen as the social trends that are usually high-weighted but with short length. Another point of view is to report a comprehensive output, so even the marginal members of the society can find trends in the outcome that are affected by their opinions. This approach is essential when we want to use trend aggregation as a recommendation system or decision support system. In this approach, a voter can find trends that are aggregated from the society while encompassing his opinion and having new suggestions for him. Considering this approach, we define the *inclusiveness* property.

Definition 4.11 (Inclusiveness)**.** *We say a TAP is inclusive, if for any input preference, that preference or a superpreference of it is a member of the set of output trends.*

$$p \in \mathcal{R} \implies \exists q \in \boldsymbol{\mathcal{P}} : p \sqsubseteq q \tag{4.15}$$

Theorem 4.9. *$\boldsymbol{F}$ and all (weighted) maximal TAPs are inclusive. TAPs with any kind of threshold are not inclusive.*

Proof. $\boldsymbol{F}$ reports all the input preferences as trends in the output, so it is inclusive. Maximal sets retain all preferences or their superpreferences, so they are all inclusive.

In TAPs with thresholds, there can always be a profile in which the length or weight of an input preference in the outcome is very low, because there was not enough support in the input for that trend or its superpreferences. In particular, we can consider an isolated preference in a profile which produces both longer trends and trends with higher weights. So, all TAPs with a threshold phase violate inclusiveness. □

4.3.12 Idempotence

In this model, the input and output have the same format (we can have the input in the format of weighted votes). So, it would be interesting to investigate what will happen if the output of the procedure is considered as a new input. A desirable property is that this consecutive running of the procedure produces the same result as if it runs once. This is called idempotence. Here, we define this property for our model formally. However, since we do not have definitive results for most cases (TAPs), we will not have any theoretical results for this property.

Definition 4.12 (Idempotence)**.** *We call a TAP* idempotent *if applying it to the output of any profile produces the same output. Note that we do not consider the weights of trends in the equation here.*

$$\forall \mathcal{R} : U(\boldsymbol{\mathcal{F}}(\boldsymbol{\mathcal{F}}(\mathcal{R}))) = U(\boldsymbol{\mathcal{F}}(\mathcal{R})) \tag{4.16}$$

From Proposition 4.1, the information in the output of the TAPs $\boldsymbol{F}$ and $\mathcal{M}$ are the same as their input profiles. So, from Proposition 4.2, these outputs produce the same results as the input. That is, $\boldsymbol{F}$ and $\mathcal{M}$ are idempotent.

Note that Proposition 4.1 also holds for $\mathcal{M}_w$, but we need to check that the procedure does not change the weights of the trends in a way that a previously weighted maximal trend will not be maximal anymore or vice versa.

For $\mathcal{T}_l^k$, we can easily see that since all the input preferences are potential trends (Proposition 4.1) and the output trends of $\mathcal{T}_l^k$ have a length of at least k, all the output trends of $\mathcal{T}_l^k$ will be in the output of the second run. Because there is no more support in the output of $\mathcal{T}_l^k$, it will not produce extra trends (Proposition 4.3). So, $\mathcal{T}_l^k$ is also idempotent.

4.3.13 Non-Imposition

Classical non-imposition requires that any social winner (preference) is possible. Sometimes the possibility of having any ranking as output is called *strong non-imposition*.

For our model, the equivalent of the strong non-imposition property would be that all possible (feasible) sets of trends are achievable. The problem is, unlike the classical methods, all the possible sets of preferences are not valid outputs (for example, in the output of any TAP without refinement, for each trend in the output set, all of its subpreferences should be in the output too). So, it is more reasonable to consider just feasible sets of trends. But, the characterization of the feasible outcome in each case is complicated itself. So, we leave out investigating the strong non-imposition property for our model and only present a general formulation of it for TAPs:

Definition 4.13 (Strong non-imposition)**.** *A TAP satisfies strong non-imposition property if for all feasible output trend sets, there is an input profile which produces that output. That is, if $\boldsymbol{\mathcal{P}}^*$ is the set of all acceptable trend sets as the output of the corresponding TAP, then:*

$$\forall \boldsymbol{\mathcal{P}} \in \boldsymbol{\mathcal{P}}^* : \exists \mathcal{R} : U(\boldsymbol{\mathcal{F}}(\mathcal{R})) = \boldsymbol{\mathcal{P}}. \tag{4.17}$$

Furthermore, the weighted version is:

$$\forall \boldsymbol{\mathcal{P}}_{\boldsymbol{w}} \in \boldsymbol{\mathcal{P}}^* : \exists \mathcal{R} : \boldsymbol{\mathcal{F}}(\mathcal{R}) = \boldsymbol{\mathcal{P}}_{\boldsymbol{w}}. \tag{4.18}$$

Now, we introduce a weaker version of non-imposition. Instead of all combinations of trends in the output, we only consider each trend individually and check if every admissible trend is achievable as a member of the output of an input profile. We call this, weak non-imposition or in short non-imposition.

Definition 4.14 (Non-imposition)**.** *A TAP has the non-imposition property, if for any possible trend, there is a profile which produces an output containing that trend.*

$$\forall P \in \mathcal{L}(\mathcal{X}) : \exists \mathcal{R} : P \in U(\boldsymbol{\mathcal{F}}(\mathcal{R})). \tag{4.19}$$

Theorem 4.10. *All major TAPs are non-imposed, except for the top weight set $\mathcal{T}_w(\boldsymbol{F}_v)$.*

Proof. Except for $\mathcal{T}_w^{\alpha}(\boldsymbol{F}_v)$, the rest are easy. For any desired trend P, the profile $\mathcal{R} = \{P\}$ produces trend P as one of the output trends. $\boldsymbol{F}$ produces P and all of its subpreferences. Since P is the maximal preference, the outcome for $\mathcal{M}$ is $\{P\}$ itself. This is also the case for $\mathcal{M}_w$ if the method is informative; i.e., $\mathcal{M}_w(\boldsymbol{F}_{v_\mu})$ and $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$. In case of $\mathcal{M}_w(\boldsymbol{F}_v)$, P is still a maximal preference and a member of the output trend set, although it is not the only member. P is also a member of $\mathcal{T}_l^k$ and the only member of $\mathcal{T}_l$.

In informative methods, $\boldsymbol{F}_{v_\mu}$ and $\boldsymbol{F}_{\tilde{v}}$, P has the maximum weight and so is a member of the output trend set for $\mathcal{T}_w^\alpha(\boldsymbol{F}_{v_\mu})$ and $\mathcal{T}_w^\alpha(\boldsymbol{F}_{\tilde{v}})$; and it is the only member of the output for $\mathcal{T}_w(\boldsymbol{F}_{v_\mu})$ and $\mathcal{T}_w(\boldsymbol{F}_{\tilde{v}})$.

For $\mathcal{T}_w^\alpha(\boldsymbol{F}_v)$, the input profile $\{P\}$ does not work anymore. Because it has a lesser value than all of its subpreferences. We can fix this by adding a few more copies of P to the input profile and an extra preference with a disjoint set of alternatives. This will increase the proportional value of trend P to help it pass the threshold. But it will not work for the top weight set, because its subpreferences always have greater weight and thus P cannot be in the outcome of $\mathcal{T}_w(\boldsymbol{F}_v)$. □

4.3.14 Consistency

A reasonable property for a voting system is, when there are two sets of disjoint voters which have a common social choice or preference, the collective result of the union of these two sets should be the same. This is called the consistency [27] criterion. It is also called reinforcement [17] or separability [24] in the literature.

In our model, we can use a similar idea and define different versions of it. The equivalent of the classic consistency could be expressed as:

$$\boldsymbol{\mathcal{F}}(\mathcal{R}_1) \cap \boldsymbol{\mathcal{F}}(\mathcal{R}_2) \neq \emptyset \implies \boldsymbol{\mathcal{F}}(\mathcal{R}_1) \cap \boldsymbol{\mathcal{F}}(\mathcal{R}_2) = \boldsymbol{\mathcal{F}}(\mathcal{R}_1 \oplus \mathcal{R}_2) \tag{4.20}$$

This is a strong requirement for a TAP, because the two input profiles could produce totally new trends together in most cases. Therefore, we replace the equality with the subset relation to provide a more promising version. We consider this as the main consistency property in our model.

Definition 4.15 (Consistency). *We call a TAP consistent if for any two input profiles, the common trends of their separate outputs are in the set of output trends of the merger of the two profiles.*

$$\boldsymbol{\mathcal{F}}(\mathcal{R}_1) \cap \boldsymbol{\mathcal{F}}(\mathcal{R}_2) \subseteq \boldsymbol{\mathcal{F}}(\mathcal{R}_1 \oplus \mathcal{R}_2) \tag{4.21}$$

Another possible axiom related to the concept of consistency is:

$$U(\boldsymbol{\mathcal{F}}(\mathcal{R}_1)) = U(\boldsymbol{\mathcal{F}}(\mathcal{R}_2)) = \boldsymbol{\mathcal{P}} \implies U(\boldsymbol{\mathcal{F}}(\mathcal{R}_1 \oplus \mathcal{R}_2)) = \boldsymbol{\mathcal{P}} \tag{4.22}$$

Here, we investigate the main version which has been defined in definition 4.15.

Theorem 4.11. *The TAP $\boldsymbol{F}$ and all the threshold TAPs are consistent. The (weighted) maximal TAPs and top length set are not consistent.*

Proof. Proposition 4.3 implies that $\boldsymbol{F}(\mathcal{R}_1) \subseteq \boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2)$ and the same for $\boldsymbol{F}(\mathcal{R}_2)$. So, $\boldsymbol{F}(\mathcal{R}_1) \cap \boldsymbol{F}(\mathcal{R}_2) \subseteq \boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2)$ and $\boldsymbol{F}$ is consistent.

From Proposition 4.5, since $\boldsymbol{F}(\mathcal{R}_1) \subseteq \boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2)$, we have $\mathcal{T}_l^k(\boldsymbol{F}(\mathcal{R}_1)) \subseteq \mathcal{T}_l^k(\boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2))$. We have the same for $\mathcal{R}_2$, so $\mathcal{T}_l^k(\boldsymbol{F}(\mathcal{R}_1)) \cap \mathcal{T}_l^k(\boldsymbol{F}(\mathcal{R}_1)) \subseteq \mathcal{T}_l^k(\boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2))$.

Because $\mathcal{R}_1 \oplus \mathcal{R}_2$ has more input preferences than $\mathcal{R}_1$ or $\mathcal{R}_2$ seperately, the weights of trends for the input profile $\mathcal{R}_1 \oplus \mathcal{R}_2$ are not less than the weight of their corresponding trends for input profiles $\mathcal{R}_1$ or $\mathcal{R}_2$. Hence, from Proposition 4.5, we have $\mathcal{T}_w^\alpha(\boldsymbol{F}(\mathcal{R}_1)) \subseteq \mathcal{T}_w^\alpha(\boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2))$ and the same for $\mathcal{R}_2$. So, $\mathcal{T}_w^\alpha(\boldsymbol{F}(\mathcal{R}_1)) \cap \mathcal{T}_w^\alpha(\boldsymbol{F}(\mathcal{R}_1)) \subseteq \mathcal{T}_w^\alpha(\boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2))$.

We could also use the argument that the threshold functions are consistent because if a trend is in the outcome of both profiles, it means that the trend passes the length or weight threshold in both cases. Adding extra preferences to the input does not decrease the lengths or weights of the trends, so they would pass the threshold in case of the union of the profiles. In fact, for $\mathcal{M}_w(\boldsymbol{F}_v)$ and $\mathcal{M}_w(\boldsymbol{F}_{\tilde{v}})$, we can easily see that they can pass thresholds as high as two times the original one. That is, $\mathcal{T}_w^\alpha(\boldsymbol{F}(\mathcal{R}_1)) \cap \mathcal{T}_w^\alpha(\boldsymbol{F}(\mathcal{R}_2)) \subseteq \mathcal{T}_w^{2\alpha}(\boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2))$

For $\mathcal{M}$, $\mathcal{M}_w$ and $\mathcal{T}_l$, consider the following example: $\mathcal{R}_1 = \{A \succ B, B \succ C\}$, $\mathcal{R}_2 = \{A \succ B, A \succ C\}$.

The set of output trends for $\boldsymbol{F}$ is the input profile itself for both cases and their maximal sets remain the same. Now, for $\mathcal{R}_1 \oplus \mathcal{R}_2$, we have: $\mathcal{R}_1 \oplus \mathcal{R}_2 = \{A \succ B, B \succ, A \succ C\}$ and $\boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2) = \{A \succ B, B \succ, A \succ C, A \succ B \succ C\}$. In this case, $\mathcal{M}(\boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2)) = \{A \succ B \succ C\}$; while $\mathcal{R}_1 \cap \mathcal{R}_2 = \{A \succ B\}$. So, $\mathcal{M}$ is not consistent.

For this example, $\mathcal{T}_l$ has the exact same effect as $\mathcal{M}$, so the top set $\mathcal{T}_l$ is also not consistent.

The same argument can be used for $\mathcal{M}_w$, regardless of its weight-assigning method, to show that $\mathcal{M}_w$ of any kind is not consistent.

For $\mathcal{T}_w$, consider the same example, but with multiple input preferences as follows: $\mathcal{R}_1 = \{(A \succ B, 2), (B \succ C, 3)\}$, $\mathcal{R}_2 = \{(A \succ B, 2), (A \succ C, 3)\}$. In this case the top weight trend of the outputs are $B \succ C$ and $A \succ C$ respectively, and their intersection is the empty set; but the top weight set of $\mathcal{R}_1 \oplus \mathcal{R}_2$ is $A \succ B$. So, $\mathcal{T}_w$ is not consistent. However, if the two outputs have common top weight trends, we can see that these common trends would be the top weight trends of $\boldsymbol{F}(\mathcal{R}_1 \oplus \mathcal{R}_2)$. Consequently, if we restrict the consistency property for situations in which the output of the two profiles have common trends; then $\mathcal{T}_w$ is consistent. □

4.3.15 Participation

In standard voting systems, the *no show paradox* is a well-studied issue. The no show paradox happens when a favorable choice of a voter is already in the outcome of the election without his vote; but by adding that vote, the outcome becomes less favorable for the voter. The participation criterion has been defined to address such situations. A voting procedure has the participation property, if the no show paradox does not happen for that procedure.

In our model, the concept of a favorable outcome for a voter other than his exact preference is not very clear. So, here we consider the case in which a preference has been chosen as an output trend; and we investigate what will happen if that preference is added to the input profile.

Definition 4.16 (Participation)**.** *A TAP has the participation property if for any preference that is already an output trend, adding that trend to the set of input preferences will not make the situation of the trend in the output worse.*

$$P \in \boldsymbol{\mathcal{F}}(\mathcal{R}) \implies P \in \boldsymbol{\mathcal{F}}(\mathcal{R}+P) \wedge (v^{\mathcal{R}+P}(P) \geq v^{\mathcal{R}}(P)) \tag{4.23}$$

Lemma 4.3. *If $P \sqsubseteq Q$, there is a positive ϵ such that for all alternatives x and y in Q: ${v'}_Q^{\mathcal{R}+P}(x,y) \leq {v'}_Q^{\mathcal{R}}(x,y)+\epsilon$; and the equality holds when $P=Q$, i.e., ${v'}_P^{\mathcal{R}+P}(x,y) = {v'}_P^{\mathcal{R}}(x,y)+\epsilon$. For Method 0, $\epsilon=1$, for Method 1, $\epsilon = 1/\mu(l(P))$; and for Method 2, $\epsilon = \frac{g(l(Q)-1)}{g(l(P)-1)}$, which is equal to 1 when $P=Q$.*

Proof. In Method 0 we have:

$${v'}_Q^{\mathcal{R}+P}(x,y) = \frac{\sum_{p_i \in \mathcal{R}+P} v_{p_i}(x,y)}{g(d_Q(x,y))} = \frac{\sum_{p_i \in \mathcal{R}} v_{p_i}(x,y) + v_P(x,y)}{g(d_Q(x,y))} = {v'}_Q^{\mathcal{R}} + \frac{g(d_P(x,y))}{g(d_Q(x,y))}.$$

Since $P \sqsubset Q$, we have $d_P(x,y) \leq d_Q(x,y)$ and at least for one pair the inequality is strict. So, because g is non-decreasing, $\frac{g(d_P(x,y))}{g(d_Q(x,y))} \leq 1$.

Hence, if $P \sqsubset Q$: ${v'}_Q^{\mathcal{R}+P}(x,y) \leq {v'}_Q^{\mathcal{R}}(x,y)+1$;
and if $P=Q$: ${v'}_P^{\mathcal{R}+P}(x,y) = {v'}_P^{\mathcal{R}}(x,y)+1$;

In Method 1, the definition is ${v_\mu}'_P(x,y) = {v'}_P(x,y)/\mu(l(P))$. So, with a similar argument, we have:

$${v_\mu}'^{\mathcal{R}+P}_Q(x,y) = \frac{\sum_{p_i \in \mathcal{R}} v_{p_i}(x,y) + v_P(x,y)}{g(d_Q(x,y)).\mu(l(Q))} = {v_\mu}'^{\mathcal{R}}_Q + \frac{g(d_P(x,y))}{g(d_Q(x,y)).\mu(l(Q))}.$$

So, if $\epsilon = 1/\mu(l(Q))$, in case of $P=Q$ we have: ${v_\mu}'^{\mathcal{R}+P}_P(x,y) = {v'}_P^{\mathcal{R}}(x,y)+\epsilon$;
and if $P \sqsubset Q$: ${v_\mu}'^{\mathcal{R}+P}_Q(x,y) \leq {v'}_Q^{\mathcal{R}}(x,y)+\epsilon$.

Similarly, for Method 2:

$$\tilde{v}'^{\mathcal{R}+P}_Q(x,y) = \frac{\sum_{p_i \in \mathcal{R}} v_{p_i}(x,y) + v_P(x,y)}{f(d_Q(x,y),l(Q))} = \tilde{v}'^{\mathcal{R}}_Q + \frac{f(d_P(x,y),l(P))}{f(d_Q(x,y),l(Q))}.$$

Since $g(d_P(x,y)) \leq g(d_Q(x,y))$, we have:

$$\frac{f(d_P(x,y),l(P))}{f(d_Q(x,y),l(Q))} = \frac{g(d_P(x,y))/g(l(P)-1)}{g(d_Q(x,y))/g(l(Q)-1} \leq \frac{g(l(Q)-1)}{g(l(P)-1)}.$$

Hence, if $P \sqsubset Q$: $\tilde{v}'^{\mathcal{R}+P}_{Q}(x,y) \leq \tilde{v}'^{\mathcal{R}}_{Q}(x,y) + \frac{g(l(Q)-1)}{g(l(P)-1)}$;
and if $P = Q$: $\tilde{v}'^{\mathcal{R}+P}_{P}(x,y) = \tilde{v}'^{\mathcal{R}}_{P}(x,y) + 1$. □

Theorem 4.12. *The TAPs $\boldsymbol{F}$, $\mathcal{M}$, $\mathcal{T}_l^k$, $\mathcal{T}_w^\alpha(\boldsymbol{F}_v)$, $\mathcal{T}_w^\alpha(\boldsymbol{F}_{v_\mu})$ and $\mathcal{T}_w^\alpha(\boldsymbol{F}_{\tilde{v}})$ satisfy the participation property. There are also some cases in which participation holds for TAPs with a weighted maximal set.*

Proof. We need to show that the weight of P is greater when we add it to the input profile and it will still be a member of the output trend set.

First, we prove that $v^{\mathcal{R}+P}(P) \geq v^{\mathcal{R}}(P)$. Using the previous lemma, we have:
$v^{\mathcal{R}+P}(P) = \boldsymbol{T}_{i<j\leq l(P)}\{v'^{\mathcal{R}+P}_{P}(c_i,c_j)\} = \boldsymbol{T}_{i<j\leq l(P)}\{v'^{\mathcal{R}}_{P}(c_i,c_j) + \epsilon\} \geq \boldsymbol{T}_{i<j\leq l(P)}\{v'^{\mathcal{R}}_{P}(c_i,c_j)\} = v^{\mathcal{R}}(P)$.

Now, we need to check whether P is in the output or not. By adding a new preference to the input, no trend will disappear from the output of $\boldsymbol{F}$, so it satisfies participation.

Also, since all the pairwise comparisons from this new preference already existed (they shaped P), there will be no new trend. But after the refinement, we may lose some trends. Since there is no change in the shape of trends, maximal set and threshold on the lengths (and also top length set) do not change the output. Also, since the weight of Q is not less, it will not be harmed from a threshold on the weights. Hence, participation holds for $\mathcal{M}$, $\mathcal{T}_l$, $\mathcal{T}_l^k$ and $\mathcal{T}_w^\alpha$.

Now, the only concern is about the weighted maximal set. We should check that adding new preferences will not make P a weighted subpreference of any other trend like Q. This could happen if $P \sqsubseteq Q$ and $v^{\mathcal{R}}(P) > v^{\mathcal{R}}(Q)$ but $v^{\mathcal{R}+P}(P) \leq v^{\mathcal{R}+P}(Q)$. Now, we show that this will not happen in some specific cases; that is, if $v^{\mathcal{R}}(P) > v^{\mathcal{R}}(Q)$ then we also have $v^{\mathcal{R}+P}(P) > v^{\mathcal{R}+P}(Q)$:

In Method 0 we have:
$v'_P(x,y) = v(x,y)/g(d_P(x,y))$ and $v'_Q(x,y) = v(x,y)/g(d_Q(x,y))$. So, when $P \sqsubseteq Q$, $v'_P(x,y) \geq v'_Q(x,y)$.

In case of $\boldsymbol{T} = \min$:
$v^{\mathcal{R}}(P) > v^{\mathcal{R}}(Q) \implies \min_{x,y\in S(P)}\{v'_P(x,y)\} > \min_{x,y\in S(Q)}\{v'_Q(x,y)\}$.

Now, we can distinguish two cases for the pair (x,y) in which $v'_Q(x,y) = \min_{x,y\in S(Q)}\{v'_Q(x,y)\} = v^{\mathcal{R}}(Q)$:

If $P \not\models (x \succ y)$, adding P to the input profile does not increase the value of $v'_Q(x,y)$. So, $v^{\mathcal{R}+P}(Q) = v^{\mathcal{R}}(Q)$. Hence, if $v^{\mathcal{R}}(P) > v^{\mathcal{R}}(Q)$, we have: $v^{\mathcal{R}+P}(P) \geq v^{\mathcal{R}}(P) > v^{\mathcal{R}}(Q) = v^{\mathcal{R}+P}(Q) \implies v^{\mathcal{R}+P}(P) > v^{\mathcal{R}+P}(Q)$.

If $P \models (x \succ y)$, adding P to the input profile increases both the value of $v'_P(x,y)$ and $v'_Q(x,y)$. From the lemma we have:
$v'^{\mathcal{R}+P}_{Q}(x,y) \leq v'^{\mathcal{R}}_{Q}(x,y) + 1 \implies v^{\mathcal{R}+P}(Q) = \min_{x,y\in S(Q)}\{v'^{\mathcal{R}+P}_{Q}(x,y)\} \leq \min_{x,y\in S(Q)}\{v'^{\mathcal{R}}_{Q}(x,y)+1\} = v^{\mathcal{R}}(Q)+1 \implies v^{\mathcal{R}+P}(Q) \leq v^{\mathcal{R}}(Q)+1$, and similarly:

$v'^{\mathcal{R}+P}_P(x,y) = v'^{\mathcal{R}}_P(x,y) + 1 \implies v^{\mathcal{R}+P}(P) = v^{\mathcal{R}}(P) + 1.$
So, if $v^{\mathcal{R}}(P) > v^{\mathcal{R}}(Q)$:
$v^{\mathcal{R}+P}(P) = v^{\mathcal{R}}(P) + 1 > v^{\mathcal{R}}(Q) + 1 \geq v^{\mathcal{R}+P}(Q) \implies v^{\mathcal{R}+P}(P) > v^{\mathcal{R}+P}(Q).$

Hence, if $T = \min$, a TAP with weighted maximal set in Method 0 satisfies participation. □

4.3.16 Reversal Symmetry

The generalization of the reversal symmetry property is easy.

If $p = c_1 \succ c_2 \succ \cdots \succ c_k$ is a preference, we show its reverse as $\overline{p}$ which is $\overline{p} = c_k \succ \cdots \succ c_2 \succ c_1$. Accordingly, for a profile $\mathcal{R} = (p_1, p_2, \ldots, p_n)$, its revers is $\overline{\mathcal{R}} = (\overline{p_1}, \overline{p_2}, \ldots, \overline{p_n})$

Definition 4.17 (Reversal symmetry)**.** *A TAP has the reversal symmetry property if inverting all the input preferences results in the inversion of all output trends.*

$$\boldsymbol{\mathcal{F}}(\overline{\mathcal{R}}) = \overline{\boldsymbol{\mathcal{F}}(\mathcal{R})} \tag{4.24}$$

Theorem 4.13. *For any TAP $\boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{F}}(\overline{\mathcal{R}}) = \overline{\boldsymbol{\mathcal{F}}(\mathcal{R})}$.*

Proof. For any profile $\mathcal{R} = (p_1, p_2, \ldots, p_n)$, we have:

$\forall c_i, c_j \in \mathcal{X} : v_{\mathcal{R}}(c_i, c_j) = v_{\overline{\mathcal{R}}}(c_j, c_i)$. So, for any trend $P \in \boldsymbol{P}$ there is a trend $\overline{P} \in \overline{\boldsymbol{P}}$; and since their pairwise values are identical for both P and $\overline{P}$, their corresponding weights would be the same in either of the four different methods. So, the set of all trends for these two cases are equivalent, except that the trends are the reverse of their correspondents.

From the definition, $p \sqsubset q \iff \overline{p} \sqsubset \overline{q}$; and so: $\mathcal{M}(\overline{\boldsymbol{P}}) = \overline{\mathcal{M}(\boldsymbol{P})}$. Hence, applying any of the refinement methods would have the same effect on both sets, since they are only dealing with the weights, lengths, and maximal sets.

So , $\boldsymbol{\mathcal{F}}(\overline{\mathcal{R}}) = \overline{\boldsymbol{\mathcal{F}}(\mathcal{R})}$ □

4.4 Summary

We have introduced many axiomatic properties in this chapter. We have evaluated different TAPs with each of these properties separately. Here, we will summarize the results we have seen and give an overall evaluation of the different TAPs and our model in general.

First, let us point out an important issue about the properties that have been presented here. Although we have introduced a large number of properties, one can see that there is no sign of some of the very common and important axioms of social choice theory, namely independence of irrelevant alternatives and monotonicity. The reason is that the formulation of these axioms in our framework is rather complicated. There could be many interpretations and formulations for these axioms and most of them do not seem promising. Also, most of these formulations are complicated, and are not easy to understand and work with. So, we can hardly find interesting results for them.

For example, consider monotonicity. A possible interpretation of monotonicity can be explained as follows:

If an alternative x is moved to a higher rank in an input preference and the rest of the input remains unchanged, then the number of output trends with x as their top alternative will not be less and the weight of trends that already had x as their top will not be lower.

We can define a similar property for moving x lower and resulting changes in trends with x on their bottom.

This can be formulates as follows:

$$\forall x : (\forall y \neq x : s^{\mathcal{R}}(x \succ y) \subseteq s^{\mathcal{R}'}(x \succ y) \wedge (\forall y, z \neq x : s^{\mathcal{R}}(y \succ z) = s^{\mathcal{R}'}(y \succ z) \implies |\{P \in \boldsymbol{P} | x = top(P)\}| \leq |\{P \in \boldsymbol{P}' | x = top(P)\}| \ \wedge \ \forall P \in \boldsymbol{P} : v^{\mathcal{R}}(P) \geq v^{\mathcal{R}'}(P)) \tag{4.25}$$

Besides this complicated formula, another problem is that in the framework of trend aggregation, a small change in one input preference (like adding or removing an alternative or swapping two alternatives) can affect the outcome to a large extent.

Because of these reasons, we leave investigation of these axioms (and many more axioms) for future work.

4.4.1 Evaluation of the Results

As it might be expected, the maximal set filter has the best performance among TAPs. From the theoretical point of view, the maximal set (specially the weighted maximal set) is the most reasonable filter because it does not lose any useful information and it only removes the redundant parts of the output. On the other hand, practically, it may not reduce the size of the output as much as we want and we do not have any control on the size of its outcome (in contrast to threshold filters).

The maximal set filter satisfies more properties than the weighted maximal set because the weights can result in unexpected changes in the outcome. This can be seen more clearly in the weight threshold filters and results in losing many axiomatic properties for these TAPs. For a few of the properties, we can see that informative methods have better results than the original (uninformative) method.

Overall, we can say that there is a trade-off between having a theoretically "good" trend aggregation procedure or a practically useful one. Maybe, by combining, these filters we can achieve TAPs that are in between, but we are not sure and it needs more investigation.This is also important because by combining the filters we can have control on another trade-off in trend aggregation: the tradeoff between the length and weights of the desirable trends.

We summarized all the results for the axiomatic properties in below, in Table 4.4.1. Because in most cases the results for a threshold set and its corresponding top set are the same, we combine them in a single column in the table. If the property holds only for the top set, we designate it by mentioning the name of the top set directly in that column. If it is only violated by the top set, then the name of the tops set comes with a negative sign in that cell.

As we mentioned at the beginning of the chapter, we use $\boldsymbol{F}$, $\mathcal{M}$ and $\mathcal{T}_l$ for all of the weight-assigning methods; for each of them we have only one column. In the only case (informativeness) that their results depend on the weight-assigning method, we use the character i to show that the TAP satisfies the informativeness axiom for the informative weight-assigning methods. We also use min once in the table for the case that the property holds when the $\boldsymbol{T}$ function of the TAP is min. In the cases that we do not have definitive results, we use a question mark.

				v		v_μ		$\tilde{v}$	
	$\boldsymbol{F}$	$\mathcal{M}$	$\mathcal{T}_l^k$	$\mathcal{M}_w$	$\mathcal{T}_w^\alpha$	$\mathcal{M}_w$	$\mathcal{T}_w^\alpha$	$\mathcal{M}_w$	$\mathcal{T}_w^\alpha$
Anonymity	✓	✓	✓	✓	✓	✓	✓	✓	✓
Neutrality	✓	✓	✓	✓	✓	✓	✓	✓	✓
Homogeneity	✓	✓	✓	✓	✓	✓	✓	✓	✓
Groundedness	✓	✓	✓	✓	✓	✓	✓	✓	✓
Unanimity	✓	✓	✓	✓	X	✓	✓	✓	✓
Unique Unanimity	X	✓	$\mathcal{T}_l$	X	X	✓	X	✓	X
Non-imposition	✓	✓	✓	✓	✓(-$\mathcal{T}_w$)	✓	✓	✓	✓
Covering	✓	✓	X	✓	X	✓	X	✓	X
Idempotence	✓	✓	✓	?	?	?	?	?	?
Informativeness	i	✓	i/$\mathcal{T}_l$	X	X	✓	✓	✓	✓
Inclusiveness	✓	✓	X	✓	X	✓	X	✓	X
EPO	✓	✓	X	✓	X	✓	X	✓	X
W. Existence	✓	✓	✓	✓	X	✓	X	✓	X
Existence	X	✓	$\mathcal{T}_l$	X	X	?	X	?	X
UPO	X	✓	$\mathcal{T}_l$	X	X	?	X	?	X
Partitioning	✓	✓	✓	✓	✓	?	?	✓	✓
Consistency	✓	X	✓(-$\mathcal{T}_l$)	X	$\mathcal{T}_w^{2\alpha}$	X	✓?	X	$\mathcal{T}_w^{2\alpha}$
Participation	✓	✓	✓	min	✓	?	✓	?	✓
Reversal Symmetry	✓	✓	✓	✓	✓	✓	✓	✓	✓

Chapter 5

Conclusion

5.1 Summary

In this thesis we have tried to bring forth the new concept of having multiple aggregated preferences for a preference aggregation system and extracting trends from partial input preferences. In this regard, we have introduced a new concept in the social choice context, that we called a trend, and formulated the first steps toward a theory for it. We tried to model the problem and present a solution with the two-phase methods. We also suggested some operations (filters) to refine the resulted trends.

The main theoretical results we have in this thesis are the axiomatic properties that we defined (extended from the classic axioms or newly defined). We had an extensive investigation about these axioms and the selected major TAPs. This helps us have an evaluation of our proposed methods and gives some support for the reliability of the model.

5.2 Computational Complexity

In this thesis, the focus was on the axiomatic aspects of the model. Here we will have a brief discussion on the computational aspect of the problem.

If there are m alternatives, the maximum number of possible trends is

$$\sum_{i=1}^{m} \frac{m!}{(m-i)!} = m! \sum_{i=1}^{m} \frac{1}{(m-i)!} = m! \sum_{i=1}^{m} \frac{1}{m!} \approx m!e \approx 2.7m!$$

This will occur when for any two alternatives x and y, both $x \succ y$ and $y \succ x$ have support in the input preferences.

If a TAP is supposed to produce all possible trends, it means that there is a potential situation that the number of output trends is $2.7m!$. So, even without considering the complexity of producing the trends, only reporting them has the complexity order of $\Omega(m!)$.. Therefore, in general, the procedure may not seems very promising. On the other hand, in practice the number of all possible trends for most cases is much lower than this.

When we are looking for a limited number of trends (based on their weight, length, or maximality), then there is a chance of having tractable algorithms.

The most basic algorithm is to produce all possible trends (or even uncompleted trends), calculating their weights and then running them through the refinement phase. But if we can develop an algorithm which produces only a limited number of trends, this would be a great advantage for trend aggregation. For example, consider the case that the $\boldsymbol{T}$ function used in the TAP is min and we are looking for the trends with a weight higher than α. In this case, after the decomposition phase and calculating the aggregated weights of pairwise comparisons, we can easily omit the relations between pairs with lower aggregated weight than α and then produce all possible trends, which is the desirable outcome. This could result in a major drop in the computation time of the algorithm in most cases, but the theoretical complexity would still be the same in this case (because of the possible worst cases).

There might be a chance of better theoretical results for complexity of approximation algorithms, which is an interesting problem to work on later.

In practice, only a few (constant) number of the 'best' trends are desirable and useful. Developing effective algorithms for these cases is another interesting problem, in particular if we can find faster algorithms.

5.3 Future Work

One of the first priorities for future work is designing an algorithm to produce the trends in a reasonable time. Then the implementation of this algorithm and having some experimental results can help us see the practical reliability and effectiveness of the model. If we can develop an algorithm which produce the most desirable trends directly, it would be a big achievement. Otherwise, considering the axiomatic properties of the different refinement methods, we might be able to develop faster algorithm for special cases. Another option is using approximation algorithms.

Another direction for the future work is on the axiomatic properties. As we have seen, for a number of axioms we do not have concrete results in all cases. This is the first step. But there are also a couple of important properties which we do not have enough results for or even some that are such that a generalization of them for our framework is very problematic. There is also a chance that we can find more general results for a set of primitive operations we can define on the preferences, such as adding an alternative to a preference, deleting one from a preference or swapping two alternatives in a preferences. We can also investigate the behavior of the functions $\boldsymbol{T}$ and g and see what the best choices for them are from axiomatic or practical point of view.

Another aspect that we can work on is other versions of the model, like systems with cardinal preferences or allowing indifference in the input preferences (or also in output trends). Using interactive systems is another extension of the model. In that case, we can iterate the procedure of finding trends and in each iteration users provide more precise preferences based on the suggestion they receive from the model.

We also can use the method presented in this thesis as a voting system, which chooses the top candidate of the "best" (or the majority of) output trend as the social choice. It can also use top candidates of other trends as possible winners. The properties of such a voting method is an interesting issue for further investigations.

Also, in general, our model can produce many similar trends. For example, two trends with only one swap between two of their alternatives. As an example, consider these two trends:

$A \succ B \succ \cdots \succ I \succ J \succ \cdots \succ Z$

and

$A \succ B \succ \cdots \succ J \succ I \succ \cdots \succ Z.$

Maybe it would be more useful if we could cluster such trends in a way like:

$A \succ B \succ \cdots \succ I \sim J \succ \cdots \succ Z$

This is not an easy procedure and even the representation of these similar trends is a potential interesting problem. We can also think about the input preferences with different levels of preferences. For example, a voter may be able to express that by the input preference of $A \succ B$, it means A and B are among the best choices for the voter or A is the best and B is the one of the worst choices for the voter. Developing a language to represent these clusters and also the levels of preference is a possible topic for future work.

We can also consider classical preference aggregation methods as a tool for trend aggregation and then compare the outcome of those methods with ours.

Appendix A

Notation	Meaning
$\mathcal{N}$	Set of all (n) voters
$\mathcal{X}$	Set of all (m) alternatives
$\mathcal{R}$	Profile (input preferences)
$\boldsymbol{P}$	Set of all trends, without weights (output)
$\boldsymbol{P_w}$	Set of all trends, with weights (output)
$\boldsymbol{\mathcal{P}}$	Set of all trends after refinement, without weights (output)
$\boldsymbol{\mathcal{P}_w}$	Set of all trends after refinement, with weights (output)
P	Trend (member of $\boldsymbol{P}$ or $\boldsymbol{\mathcal{P}}$)
P_w	Weighted trend (member of $\boldsymbol{P_w}$ or $\boldsymbol{\mathcal{P}_w}$)
$S(P)$	Set of all alternatives occurred in P
$S(\mathcal{R}), S(\boldsymbol{P})$	Set of all alternatives occurred in any preference in $\mathcal{R}$ ($\boldsymbol{P}$)
$U(\boldsymbol{P_w})$	Set of trends in $\boldsymbol{P_w}$ without their weights
$\mathcal{M}(X)$	Set of maximal members of X
$\mathcal{M}_w(X)$	Set of weighted maximal members of X
$\mathcal{T}_l^k(X)$	Set of members of X with length of at least k
$\mathcal{T}_l(X)$	Set of the longest members of X
$\mathcal{T}_w^\alpha(X)$	Set of members of X with the weight of at least α
$\mathcal{T}_w(X)$	Set of members of X with maximal weight.
$l(p)$	Length of preference p
$v^{\mathcal{R}}(P)$	Weight of trend P for profile $\mathcal{R}$
$v_P'^{\mathcal{R}}(x, y)$, $v_P'^{\mathcal{R}}(x \succ y)$	The value that $x \succ y$ supports for trend P
$v_{\boldsymbol{P}}(P)$	Weight of trend P in the trend set $\boldsymbol{P}$
$s^{\mathcal{R}}(p)$	Set of all voters (in profile $\mathcal{R}$) that p is the subpreference of it
$C(\mathcal{R})$	Set of all pairwise comparisons in $\mathcal{R}$
$\boldsymbol{F}$	Aggregation procedure without the refinement phase
$\boldsymbol{\mathcal{F}}$	Aggregation procedure with the refinement phase

Table A.1: Table of Notation

Bibliography

[1] K. J. Arrow. *Social Choice and Individual Values.* John Wiley and Sons, New York, 2nd edition, 1963.

[2] S. Bistarelli, M. S. Pini, F. Rossi, and K. B. Venable. Bipolar preference problems: framework, properties and solving techniques. In *Proceedings of the 11th annual ERCIM international conference on Recent advances in constraints*, CSCLP'06, pages 78–92, Berlin, Heidelberg, 2007. Springer-Verlag.

[3] S. J. Brams and P. C. Fishburn. Voting procedures. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Handbook of Social Choice and Welfare*, volume 1, chapter 4, pages 173–236. Elsevier, 2002.

[4] F. Brandt, V. Conitzer, and U. Endriss. Computational social choice. In G. Weiss, editor, *Multiagent Systems.* MIT Press, Forthcoming.

[5] Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. A short introduction to computational social choice. In *Proceedings of the 33rd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM-2007)*, volume 4362 of *LNCS*, pages 51–69. Springer-Verlag, January 2007.

[6] Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. Preference handling in combinatorial domains: From ai to social choice. *AI Magazine, Special Issue on Preferences*, 29(4):37–46, 2008.

[7] N. Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* Imprimerie royale, Paris, France, 1785.

[8] V. Conitzer and T. Sandholm. Computing shapley values, manipulating value division schemes, and checking core membership in multi-issue domains. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI'04, pages 219–225. AAAI Press, 2004.

[9] J. C. de Borda. *Memoire sur les Elections au Scrutin.* Histoire de l'Academie Royale des Sciences, Paris, 1781.

[10] C. L. Dodgson. *A Method for Taking Votes on More than Two Issues.* Clarendon Press, 1876.

[11] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Commun. ACM*, 35(12):61–70, Dec. 1992.

[12] H. Kautz, B. Selman, and M. Shah. Referral web: combining social networks and collaborative filtering. *Commun. ACM*, 40(3):63–65, Mar. 1997.

[13] J. G. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.

[14] J. Lang and L. Xia. Sequential composition of voting rules in multi-issue domains. *Mathematical Social Sciences*, 57(3):304–324, May 2009.

[15] T. Lu and C. Boutilier. Budgeted social choice: From consensus to personalized decision making. In *IJCAI*, pages 280–286, 2011.

[16] P. Melville and V. Sindhwani. Recommender systems. In *Encyclopedia of Machine Learning*, pages 829–838. 2010.

[17] H. Moulin. Condorcet's principle implies the no show paradox. *Journal of Economic Theory*, 45(1):53–64, June 1988.

[18] D. M. Pennock, E. Horvitz, and L. C. Giles. Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 729–734. AAAI Press / The MIT Press, 2000.

[19] M. S. Pini, F. Rossi, K. B. Venable, and T. Walsh. Computing possible and necessary winners from incomplete partially-ordered preferences. In *Proceedings of the 2006 conference on ECAI 2006: 17th European Conference on Artificial Intelligence August 29 – September 1, 2006, Riva del Garda, Italy*, pages 767–768, Amsterdam, The Netherlands, 2006. IOS Press.

[20] M. S. Pini, F. Rossi, K. B. Venable, and T. Walsh. Aggregating partially ordered preferences. *J. Log. Comput.*, 19(3):475–502, 2009.

[21] D. Porello and U. Endriss. Ontology merging as social choice. In *Proceedings of the 12th International Workshop on Computational Logic in Multiagent Systems (CLIMA-2011)*, volume 6814 of *LNAI*, pages 157–170. Springer-Verlag, July 2011.

[22] P. Resnick and H. R. Varian. Recommender systems. *Commun. ACM*, 40(3):56–58, Mar. 1997.

[23] D. Saari. *Geometry of voting*. Number 3 in Studies in Economic Theory. Springer, Berlin, 1994.

[24] J. H. Smith. Aggregation of preferences with variable electorate. *Econometrica*, 41(6):1027–41, November 1973.

[25] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Adv. in Artif. Intell.*, 2009:4:2–4:2, Jan. 2009.

[26] L. Terveen and W. Hill. Beyond recommender systems: Helping people help each other. In *HCI in the New Millennium*, pages 487–509. Addison-Wesley, 2001.

[27] H. P. Young. Social choice scoring functions. *SIAM Journal on Applied Mathematics*, 28:824–838, 1975.

[28] H. P. Young. Extending condorcet's rule. *Journal of Economic Theory*, 16(2):335–353, December 1977.