EXPLORATORY ANALYSIS OF SIMILARITIES BETWEEN COMMON SOCIAL CHOICE RULES

John C. McCabe-Dansted[1] and Arkadii Slinko

Abstract. Nurmi (1987) investigated the relationship between voting rules by determining the frequency that two rules pick the same winner. We use statistical techniques such as clustering and multidimensional scaling to further understand the relationships between rules. We also investigate how the relationships change when elections with Condorcet winners are excluded from the frequency data, and when the homogeneity of the voting population is increased.

Keywords: Voting, Social Choice, Homogeneity, Clusters, Multidimensional Scaling, Condorcet

JEL Classification: D7

1. Introduction

There are many possible voting rules (procedures) and the choice of voting rule does affect the outcome of votes. It has been shown that for various sets of desirable properties, no one voting rule can satisfy them all (Arrow, 1951, 1963; Gibbard, 1973; Satterthwaite, 1975). This clearly makes the choice of voting rule an important one.

There are many feasible voting systems, with much debate as to which is best. In fact 24 existing (and one new) voting rules are considered in this project. The behaviour of voting rules is often not clear by simple inspection. It is therefore desirable to have robust methods for classifying voting rules.

The possible issues that can arise from voting systems may be hard to enumerate, and their importance subjective. Instead we focus on the less subjective measure of similarity (of outcome) between the different voting systems. Nurmi (1987, 1990) was the first to use simple computational techniques to determine similarities between the rules. He examined the frequency that two rules chose the same winner for several pairs of rules. In his investigation the number of voters ranged from five to 301 and the number of alternatives from three to seven. He concluded that dissimilarities between the rules are significant.

Since all comparisons in Nurmi's study were pairwise, the whole picture could not be seen. In 1999 Slinko came up with the idea to use cluster analysis to understand relationships between the rules and through the construction of hierarchical trees to try to represent the whole universe of rules. The first such attempt was undertaken in Leung's thesis (2001) and published in Slinko and Leung (2003).

Their work was extended by Shah (2003), who used multidimensional scaling (see section 2.6) to analyse the differences between the different rules. He found that multidimensional scaling required far more than three dimensions to fully capture the relationships between voting rules. Shah also demonstrated the effectiveness of dendrograms for visualising the results of hierarchical clustering of voting rules. Shah used the average clustering method to generate the hierarchical clusters.

Another attempt to understand similarities of more than two rules was undertaken in Gehrlein and Lepelley (2000) and Merlin et al. (2000), where they calculated the probability that all scoring rules, all voting rules, respectively, elect the same winner.

[1]Corresponding author. The address for correspondence: 86 Patteson Ave, Mission Bay, Auckland 5, New Zealand; Email: turtle@varsity.co.nz; Web: www.dansted.org/turtle

In this paper, we use hierarchical clustering to examine the effect of the homogeneity of the populations on the similarity of outcome of different voting rules. The effect of excluding elections with Condorcet Winners is also considered in order to understand how similar the rules are in resolving the difficulty presented by the absence of a Condorcet winner.[2]

Our method of classification is based on Single Clustering that proved to be relatively stable when tested on our random populations[3]. It is investigated how to visualise this data effectively. As visualisations of data can be misleading, a formal test is developed to determine if changes in the clusters produced are statistically significant. Based on this test, reports are generated that list various statistically significant changes.

2. Method

2.1. Creating a Profile.

We define n as the number of voters, and m as the number candidates. We represent the set of n votes by a n-tuple of linear orders called a profile. We wish to create random profiles a variable homogeneity that we can specify. *Homogeneity is loosely a measure of the tendency of the population to contain a large proportion of voters, who share the same order of candidate preferences.*

Our population creation method is based on the Pólya-Eggenberger Urn Model (Norman and Samuel, 1969).

2.1.1. *Pólya-Eggenberger Urn Model.*

Under the urn model we start with a big urn containing balls, each of a different colour. To generate each random sample, we pull a ball out of the Urn at random and note its colour. After removing each ball, we return the ball that was taken together with a additional balls of the same colour to the urn, for some non-negative integer a.

For generation of random voter profiles, we can replace colours with linear orders.

The parameter a characterises homogeneity, but we wish the value of the parameter of homogeneity to be consistent for different numbers of candidates. Therefore we use b as our parameter of homogeneity, which is defined as $b = \frac{a}{n!}$. The parameter b is consistent, for example, if $b = 1$ the second voter copies the first voter half of the time regardless of the number of candidates.

2.2. Comparing Outcomes.

2.2.1. *Distance between voting rules.*

For a randomly generated set of profiles, e.g. a sample of profiles created using the same homogeneity, we define distance between rules as the frequency that rules fail to pick matching candidates. Where p is a profile from our set of profiles, s is the number of populations in that set and $C(p, rule)$ is the candidate chosen by $rule$ applied to p.

$$Distance(rule1, rule2) = (1/s)\Sigma_p \begin{cases} 0, & C(p, rule1) = C(p, rule2) \\ 1, & otherwise \end{cases}$$

Therefore, the distance, so-defined, is a random variable.

2.3. Hierarchical Clustering.

Hierarchical clustering groups elements together into clusters, and clusters together into super-clusters. An example of a set of hierarchical clusters is the grouping of individual organisms into subspecies, subspecies into species, species into genus and so on. We use the single clustering algorithm to generate hierarchical clusters from the distances defined above.

---

[2]We thank Matthew Jackson for this idea

[3]Both the average clustering and the single clustering method used in this project are documented more completely by Duda et al. (2001).

2.3.1. *Clustering of voting rules.* We found that the "Single" clustering method (Duda et al., 2001) when combined with a margin of error to be the most stable (for voting rules). Rerunning the simulation gives the same set of clusters to within a moderate margin of error, as we discuss further in section 2.4. For convenience, the Single clustering method is outlined here.

Algorithm 2.1. Single clustering method

(1) Start with minimum spanning tree
(2) Remove the longest edge e to form two sub-clusters
(3) The dissimilarity of parent cluster is defined as the length of e
(4) If no edges remain then stop, otherwise return to (2)

2.4. **Clustering Methods and Statistical Significance.** Consider a sub-cluster A and a it's super-cluster B. Define a, b as the dissimilarity within the clusters of A and B respectively. As the length of the edges that the single clustering method removes is monotonically decreasing, $b \geq a$. However random fluctuations may cause the edge whose length corresponds to a to become longer than b. In this case the single clustering method will no longer choose A as a sub-cluster of B. We develop a Z-test with the null hypothesis that $\mu(b) \leq \mu(a)$ and that hence B would not be chosen as a sub-cluster of A if a near infinite number of samples were taken. If the Z-score is greater than 3.5 we may safely conclude that the choice of A as a sub-cluster was not merely due to random fluctuations. Hence we consider such clusters to be "stable" clusters.

2.5. **Measures of Distance between Profile Sets.** Two different measures of distance between profile sets will be explained and discussed in this section: flattened similarity matrix distance and tree distance.

2.5.1. *Flattened Similarity Matrix Distance* (`flat simmat`). For each profile set we may define a similarity matrix M; each element M_{ij} represents the frequency that the rule i picks the same result as rule j. As $M_{ij} = M_{ji}$, each distance is stored twice in the matrix, so we need only consider the values in the lower (or upper) triangle.

The flattened similarity matrix is produced from the similarity matrix distance by replacing each element with its position on a sorted list of all elements in the matrix. For example:

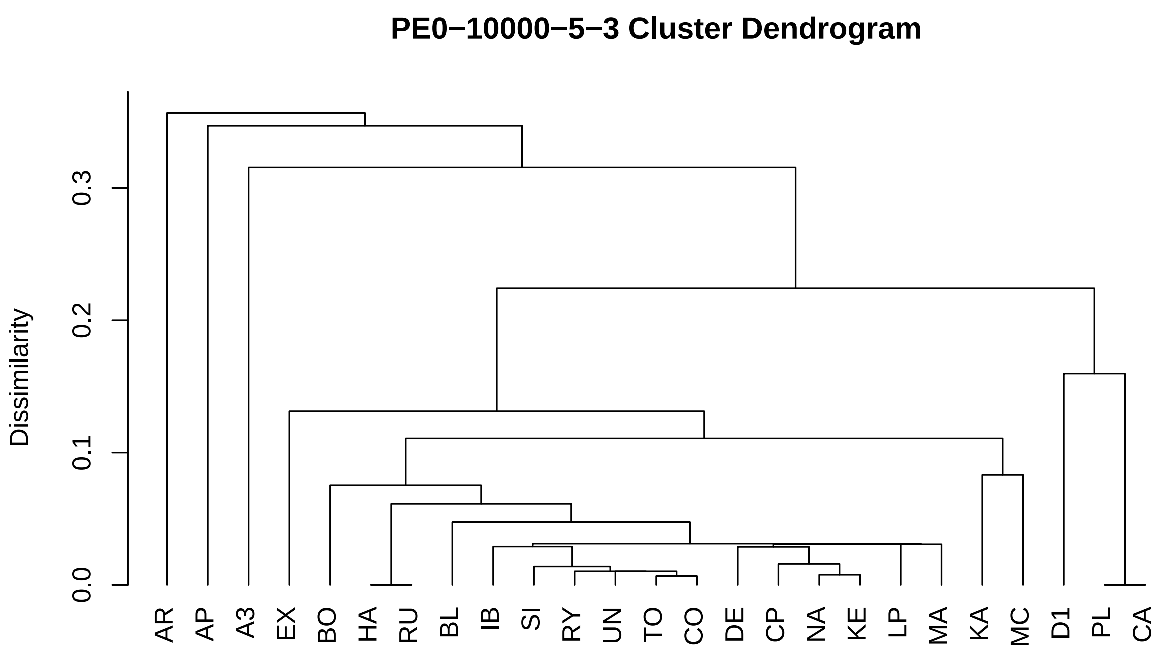
	M_1	M_2	M_3
Similarity M	1 0.03 1 0.02 0.01 1	1 0.01 1 0.02 0.03 1	1 0.1 1 0.4 0.9 1
Flattened M	X 3 X 2 1 X	X 1 X 2 3 X	X 1 X 2 3 X
$\|M_i - M_2\|$	X 2 X 0 2 X		X 0 X 0 0 X
Distance	4		0

As we can see from the table above, M_2 and M_3 have a vastly different magnitude. Since the smallest element of M_2 is the smallest element M_3, second smallest element of M_2 is the second smallest element M_3, and so on, their flattened matrices are the same; the flattened similarity distance between M_2 and M_3 will be 0 regardless of whether the 1-norm or 2-norm is used.

This measure of distance has the following advantages; it is not affected by changes in the overall magnitude of the similarity matrix; comparing the flattened similarity matrix loses less information than comparing the trees or clusters.

2.5.2. *Tree Distance (`tree`).* As the single clustering method is hierarchical, its results may be represented as a tree of clusters as shown in the figure below.

Figure 1. - Example of tree formed from hierarchical clusters (5 Candidates, 3 Voters, $b = 0$)



We may compare distances between trees. We define the distance between trees thus: for each tree form a matrix from number of edges between leafs, call these matrices A, B, the tree distance is the sum of the differences between the matrices. That is, the tree distance is equal to $\Sigma_i \Sigma_j |a_{ij} - b_{ij}|$.

Tree distance has the advantage that it is not affected by changes in the overall magnitude of the similarity matrix, and that its distance represents changes in the topology of the trees used to visualise the clusters. Its main weaknesses are that its results may depend upon the clustering method used, and as most of the information is the similarity matrix is lost in the tree representation, so this measure of distance is more granular than the similarity matrix and flattened similarity matrix.

2.6. **Multidimensional Scaling.** Multidimensional scaling, also known as principle coordinates analysis (Gower, 1966), takes a pairwise set of distances and forms a n dimensional map. The distances on this map are approximately equal to the original distances used as input to the algorithm.

The example used to demonstrate multidimensional scaling in the R statistical package takes a matrix of the distances between the major cities in Europe, and produces a map of the position of the cities. Because only the distances are given, the correct rotation of the map cannot be calculated. Hence the axes of maps produced are arbitrary.

Shah (2003) used multidimensional scaling to analyse the reasons why some rules are similar to others. In this project multidimensional scaling is instead used to visualise the effect that changing the parameters of the population has on the similarity between the voting rules considered. We take the distances defined in section 2.5 and map these distances onto a ordinary two dimensional page.

2.7. **Social Choice Rules Implemented.** We list the rules in the alphabetical order. The complete list of rules is given at the end of the section.

Anti-Plurality rule (AP) - Voters rank all alternatives. The alternative with the fewest number of last place votes is the winner.

The next three rules belong to the class of Approval Voting Rules. These rules take as input not only a profile, but also a threshold separating the approved and disapproved candidates (Bram and Fishburn, 1982). We model this threshold in three different ways and obtain three different rules accordingly:

Approval Uniform(AR) - Each voter approves a random number k of candidates, where k is uniformly chosen from $\{1, 2, \ldots, m-1\}$.

Approval Poisson $\mu = 3$ (A3) - Each voter approves k voters where k is taken from the Poisson distribution with a mean of 3. The motivation consideration of this rule follows: A study by Laslier (to appear) found that if approval voting had been used in France, voters would had approved of on average 3 out of the 16 candidates. For this reason Pritchard and Slinko (2003) suggested that out of all scoring rules[4], 3-approval may best approximate approval voting among scoring rules.

$\lfloor \frac{m}{2} \rfloor$*-Approval* (KA) - Each voter a fixed number of candidates. As shown by Slinko (2002) it possesses certain optimal properties among all scoring rules.

Black's rule (BL) - If a Condorcet winner exists, then the Black winner is the Condorcet winner. If no Condorcet winner exists, then the Black winner is the Borda winner.

Borda rule (BO)

Carey rule (CA) - The following procedure is repeated until only one alternative remains: calculate the plurality score (number of voters that rank this alternative first) for each alternative. Eliminate those alternatives with below average plurality scores. Continue this process until one alternative remains.

Coombs' procedure (CO) - If one alternative is ranked first by more than 50% of voters, then that alternative is the Coombs winner. Otherwise, the following procedure is carried out: eliminate the candidate that receives the largest number of last place ranks. Continue the process until one alternative remains.

Copeland rule (CP) - In this version of the Copeland rule, in case of a tie, second orders Copeland scores are calculated first (Laslier (1997)) and then if they are equal, the preferences of the first voter are consulted.

Dictatorship of the first voter (D1)

Dodgson rule (DE) - The Dodgson score is the minimum number of neighbouring preferences that must be swapped to make the candidate a Condorcet winner. The candidate with the minimum Dodgson score wins (Black, 1958; Tideman, 1987).

Exhaustive procedure (EX) - The following procedure is repeated until only one alternative remains: eliminate the candidate that receives the largest number of last place ranks. We then continue until only one alternative remains. This is similar to Coombs procedure except that the procedure is repeated until all but one of the alternatives are eliminated, rather than stopping when one alternative is ranked first by more than 50% of voters, as is the case with Coombs. The distinction between the two rules is noted in Tideman (1987, 191).

Hare's rule (HA) (also known as Single Transferable Vote or Alternative Vote) - If one alternative is ranked first by more than 50% of voters, then that alternative is the Hare's winner. Otherwise, the following procedure is carried out: eliminate the candidate that receives the fewest number of first place ranks. Continue until one alternative remains.

Inverse Borda rule (IB) - The following procedure is repeated until only one alternative remains: eliminate the candidate with the lowest Borda score. The last alternative remaining is the Inverse Borda winner.

---

[4]For approval voting to be a point scoring rule, voters must approve of a fixed number of candidates.

Kemeny rule (KE) - Kemeni (1959). This rule uses the concept of a distance between relations to find a linear order whose sum of distances to the voters' linear orders is minimised and then takes the top alternative of that linear order. Also known as Slater rule (e.g. Laslier (1997))

Long Path rule (LP) - E.g. (Laslier 1997, 55).

Majoritarian Compromise (MC) - The rule proceeds by first examining the top row (first choice) of each voter's preferences. If an alternative appears in the top row of more than or equal to 50% of voters' preferences, then that alternative is the majoritarian compromise winner. Otherwise, we look at the top two rows of voters' preferences. If any alternatives appear in the top two rows in at least 50% of voters' preferences, then we choose the alternative among this set that appears in the top two rows most often. If no alternative appears in the top two rows of voters' preferences for more than 50% of voters, then we examine the top three rows (first, second, and third choices), and so on (Sertel and Yilmaz, 1999, 620). For an odd number of alternatives the procedure was clarified by Slinko (2002).

Markov rule (MA) - E.g.(Laslier, 1997)

Nanson's procedure (NA) - The following procedure is repeated until only one alternative remains: calculate the Borda score for each alternative. Eliminate those alternatives with below average Borda scores. Continue this process until one alternative remains or until all remaining alternatives have tied Borda scores, in which case the tie should be broken[5].

Plurality rule (PL) - Also known as First Past the Post.

Plurality Runoff procedure (RU)

Simpson's rule (SI) (also known as the minimax/maximin rule) - This rule picks the alternative with the smallest maximum margin of defeat. E.g. Laslier (1997)

Top Cycle rule (TO) - E.g. Laslier (1997)

Uncovered Set rule (UN) - E.g. Laslier (1997)

Raynaud procedure (RY) (also known as the Arrow-Raynaud procedure) – The following process is repeated until only one alternative remains: for each pair of alternatives x and y, determine the number of voters that prefer x over y. For each alternative x, determine the maximum number of voters that prefer x to any other alternative y. Eliminate the alternative with the smallest such maximum. Repeat until only one candidate, which is the Raynaud winner, is left (Lansdowne 1997, 126).

Table 1. - The two letter abbreviations for the Social Choice Rules (Summary)

Abv.	Social Choice Rule	Abv	Social Choice Rule
A3	Approval Poisson $\mu = 3$	IB	Inverse Borda
AR	Approval Uniform	KA	$\lfloor \frac{m}{2} \rfloor$-Approval
AP	Anti-Plurality	KE	Kemeny
BL	Black	LP	Long Path
BO	Borda	MA	Markov
CA	Carey	MC	Majoritarian Compromise
CO	Coombs	NA	Nanson
CP	Copeland	PL	Plurality
D1	Dictatorship	RU	Plurality Runoff
DE	Dodgson Rule	RY	Raynaud
EX	Exhaustive	SI	Simpsons
HA	Hare	TO	Top Cycle
		UN	Uncovered Set

[5]Other definitions of Nanson exist (Richelson 1981, 346) in which Nanson is defined exactly the same as the Inverse Borda rule above – that definition was not used in this paper. The definition we used is the one used by Tideman (1987, 194).

3. Results

3.1. **Effect of Increasing Homogeneity.**

3.1.1. *Comparing Clusters with and without Homogeneity.* Below we see two dendrograms of populations with and without homogeneity.

Figure 2. - Impartial Culture

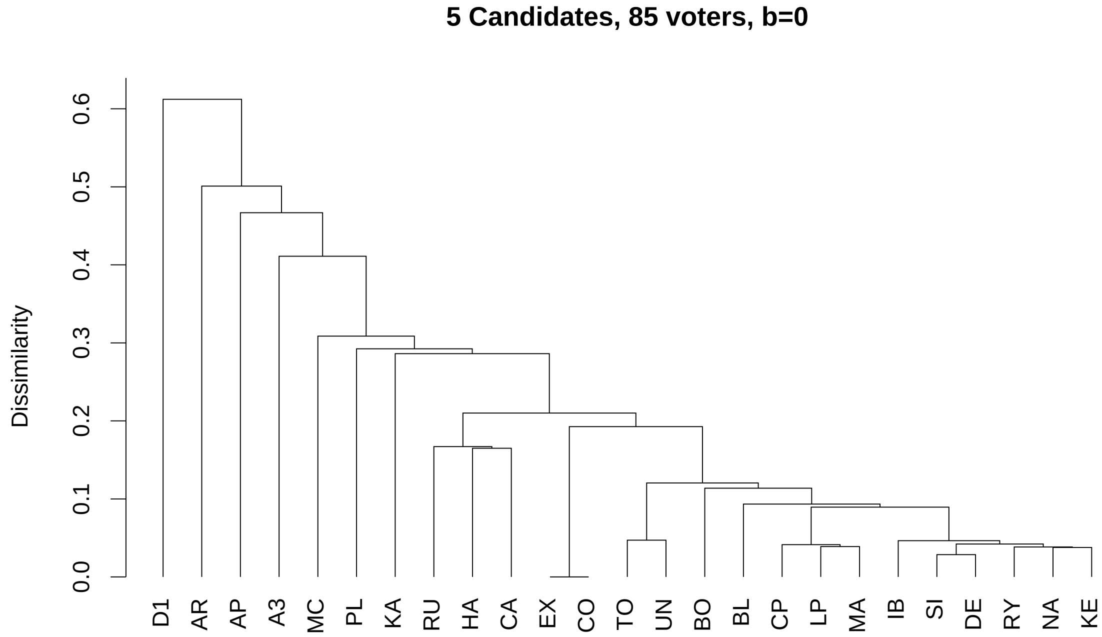
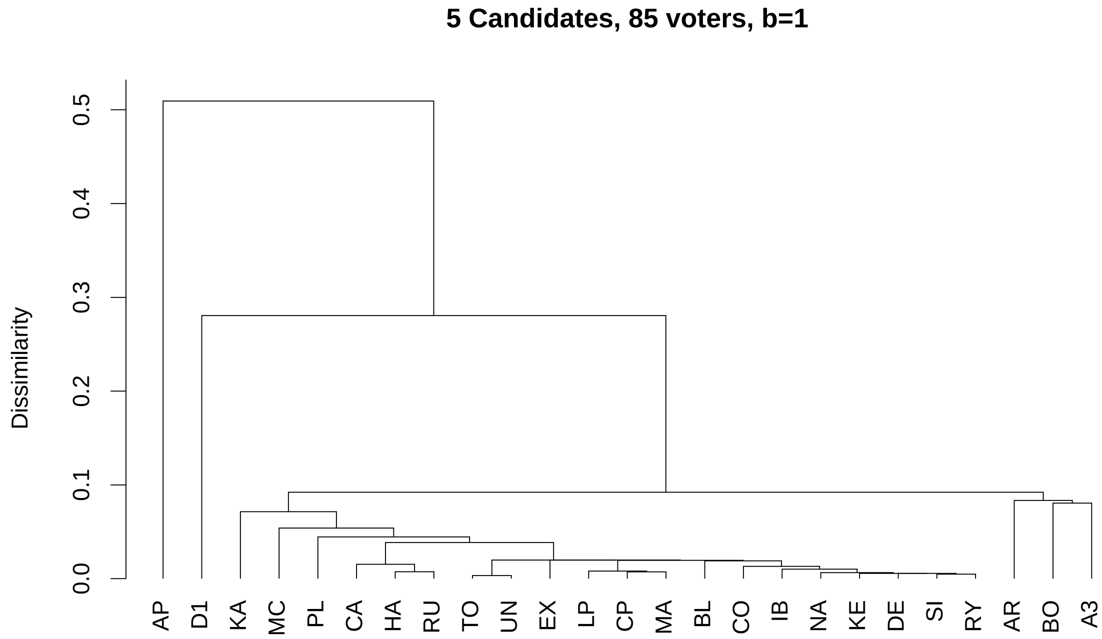


Figure 3. - Homogeneous Culture



3.1.2. *Borda's Cluster.* A new cluster forms in the homogeneous population, containing The Borda's (BO) Rule, Approval with uniform random approvals (AR), Approval with Poisson ($\mu = 3$) random approvals (A3). Borda's Rule is a point scoring rule: each candidate is given a number of points for each vote, depending on where the candidate is ranked on the vote.

Approval voting is similar, giving a point to each candidate the voter approves. Rational voters will always approve their first n preferences, for some n (Brams, Sanver 2003). When there are a large number of voters with the same preferences, and n is uniformly random this will approximate Borda's rule. Since there are five alternatives, the Poisson with mean 3 roughly approximates the uniform distribution on 1,2,3,4,5 and AR also approximates BO. Borda's cluster forms for $b \geq 0.5$.

3.1.3. *Decreasing Similarities.* We would expect that the similarity of the voting rules would increase with b. The more homogeneous the population the more likely a clear winner is to exist, which should be selected by all voting rules. These graphs mostly support that hypothesis, however there are a few exceptions to the rule. The most drastic is the Coombs' (CO) and Exhaustive (EX) rules. Where the voters are impartial these two rules are identical, giving the same result in all 10000 simulations. However when the parameter of homogeneity is increased to one, the rules split into two, and are not even in the same cluster.

Coombs' and Exhaustive rules (CO-EX). Shah (2003) found that for 85 voters the Coombs' and Exhaustive rules are identical, but they diverge as the number of voters decreases. It is unsurprising that these rules are similar. Shah explains that both rules work by successively eliminating candidates with the most last place votes. The only difference between the rules is that Coombs' rule will stop if a candidate has the most first place votes, while the Exhaustive rule will continue until only one candidate remains. Clearly, if no candidate achieves more than half the first place votes until the last step, both rules will choose the same winner. Shah further explains that the less voters there are, the more likely it is that a candidate will gain more than half the first place votes; causing Coombs' rule to choose a different candidate to the Exhaustive rule.

We can explain our result in a similar fashion. An increase in homogeneity will also increase the chance that a candidate gets a majority of the first place vote.

Anti-Plurality and other rules (AP). We can also see that the Anti-Plurality rule remains dissimilar to the other rules once b increases to 1. If a majority votes the same way we would expect that most rules would usually select the majorities first preference. However the Anti-Plurality rule will only select the majorities first place vote if it also has the least last place votes. Since the majority all vote the same way, they will only vote against one of the four other alternatives. There is, therefore, no good reason to believe that the Anti-Plurality rule will pick the majorities first preference, even if the majority is very large. Hence we would expect a moderate increase in homogeneity to result in no substantial increase in similarity between the Anti-Plurality rule and other rules.

A large increase in homogeneity will increase the similarity between Anti-Plurality and the Dictatorship rules. We use the preferences of the first voter to split ties. In a very homogeneous population only a few candidates will be black listed, so the first voter's choice is likely to be chosen; further it is likely that many other voters will have copied the first voter's vote, making the first voter a part of a majority.

The complete list of similarities that decreased when we increased b to 1 is "BL-AP BO-AP BO-BL EX-AP CP-AP IB-AP LP-AP MC-AP DO-AP NA-AP SI-AP TO-AP UN-AP RY-AP CO-AP CO-EX MA-AP DE-AP DQ-AP DQ-DE KE-AP"[6].

We may ask if this list is the same or similar for other assumptions. As table 2 shows, the list is a small subset when we reduce the number of voters to 25. For small b, increasing b will decrease a number of distinct similarities. For larger b, the similarities all increase as expected.

[6]That is, Black's rule (BL) became less similar to Anti-Plurality (AP), Borda's (BO) rule became less similar to AP, and so on. This list was generated by typing "find.sim.decrease(a=5,v=85,N=10000)". using the function find.sim.decrease we developed for the statistical package R.

Table 2. - Dissimilarities that Increase with Homogeneity

#alter	#voter	b	similarities that decrease
5	25	$0 \to 1$	BO-BL CO-EX DQ-DE
5	5	$0 \to 1$	BO-BL DQ-DO (DO-BO DE-BO occur once in three runs)
5	25	$0 \to 0.5$	TO-AP UN-AP RY-AP CO-AP CO-EX MA-AP DE-AP DQ-AP DQ-DE KE-AP BL-AP BO-AP BO-BL EX-AP CP-AP HA-AP IB-AP LP-AP MC-AP DO-AP NA-AP SI-AP
5	25	$0.5 \to 1$	*none increase*
5	25	$0.2 \to 0.5$	BO-BL DQ-DO

In the first row of table 2, the similarities between Borda and Black (BO-BL), between Coombs' and Exhaustive (CO-EX), Dodgson quick approximation and Dodson exact (DQ-DE) decrease.

Black's rule and Borda's rule (BL-BO). Black's rule will pick the Condorcet winner if it exists, choosing the Borda winner otherwise (Aleskerov and Kurbanov, 1999). Thus we would expect the similarity between Black's rule and Borda's rule to decrease as the frequency of that a Condorcet winners exist increases, and increase with Borda's Condorcet efficiency.

Table 3. - Analysis of similarity between Borda and Black

#alter	#voter	b	%Condorcet	BO Condorcet Eff	BL-BO sim
5	25	0	76%	86%	89%
5	25	0.2	84%	84%	87%
5	25	0.5	91%	85%	86%
5	25	1	95%	87%	88%
5	25	2	98%	90%	91%

As shown by the table above we find that for small b, the frequency of Condorcet winners increases rapidly with b, causing the similarity to drop. For larger b, the increase in Borda's Condorcet efficiency will cause the similarity to increase[7] with b.

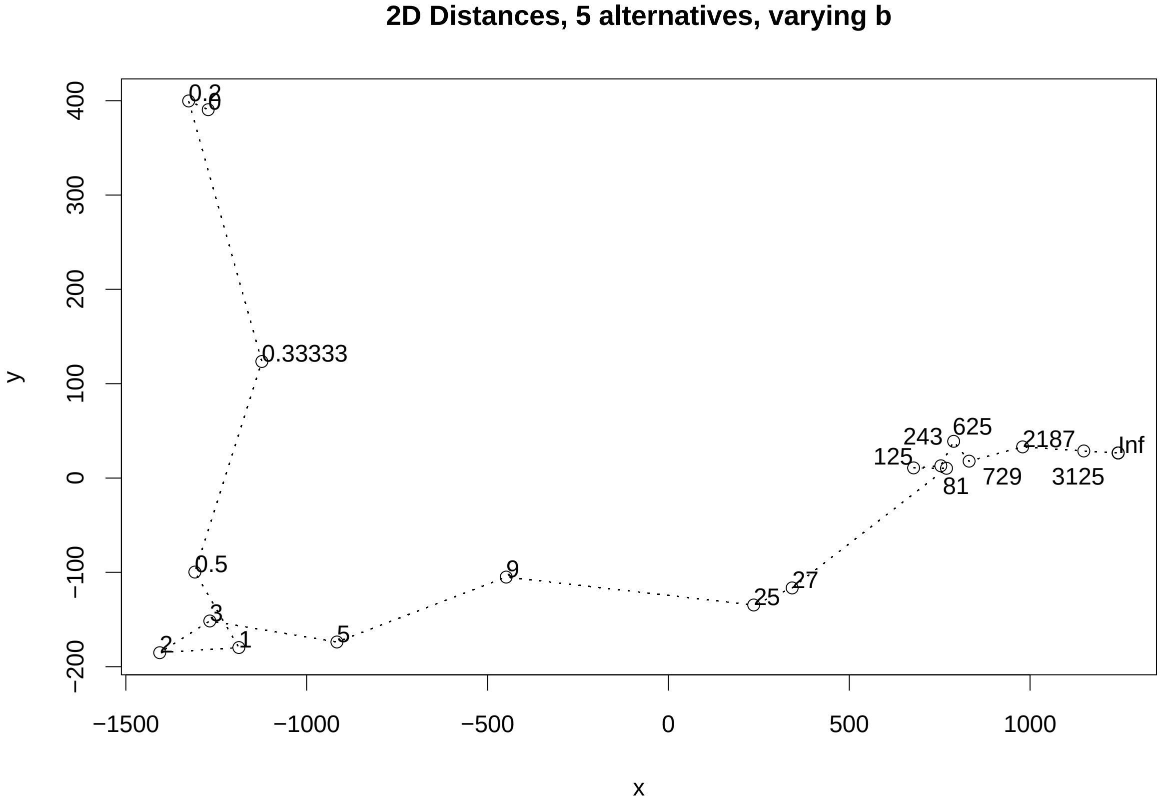
Dodgson Quick approximation and Dodgson's rule (DQ-DE). DQ is an approximation to the Dodgson rule, using the total number of additional victories needed to become a Condorcet winner as the Dodgson score[8]. For example, if an alternative **a** is tied with **b** and loses to **c** by a margin of one, then it needs one additional victory over **c** to become a Condorcet winner. This approximation works on the assumption that if an alternative **a** needs an additional victory over an alternative **b** there will be at least one voter that ranks **b** just over **a**, allowing an additional victory over **b** to be achieved with only one swap. This assumption may not hold with a homogeneous population, so the approximation may diverge from the exact value with an increase in homogeneity.

3.2. **Two Dimensional Map of Changes.** The map of changes in figure 4 on the next page was generated using multidimensional scaling, described in section 2.6. From $b = 0$ to $b = 0.5$ the points move down the graph; from $b = 1$ to $b = 3$ there does not appear to be much movement; the graph changes direction and from $b = 5$ to $b = \infty$ the points move to the right.

[7]Shah (2003) demonstrated a similar result for varying the number of voters.

[8]The DQ approximation to the Dodgson score, is a lower bound on the Dodgson Score. This DQ lower bound was developed in this project as a way to quickly eliminate candidates which could not possibly be Dodgson Winners, improving the computational efficiency of the Dodgson Exact (DE) algorithm.

Figure 4. - Map of Distances, Varying Homogeneity



The x, y Axes are unlabelled as they are arbitrary, see section 2.6
Only the distance between points is meaningful

The downwards movement from 0 to 0.5 may represent the changes in clustering that occur when homogeneity is added. That the points do not move significantly as b goes from 1 to 3 may suggest that no further changes occur in the clustering. Where b is 3, the population is very homogeneous, the second voter will copy the vote of the first voter three-quarters of the time. It is plausible that as the population becomes even more homogeneous, it will become impossible to distinguish between the rules. The change in direction of the graph may represent this different type of change.

This interpretation of the graph suggests the truth of the following statements; increasing b up to 0.5 will form new clusters, but increasing it further will not; There is little difference between the clusters produced for $b = 1, 2, 3$; increasing b beyond 3 will cause clusters to be lost, as they become subsumed by larger clusters.

We shall demonstrate that these statements are mostly true.

Are there new clusters after $b = 0.5$? Each element of Table 4 represents the number of new clusters that form as we increase the parameter b from the value at the top of the column to the value at the left of the row. Increasing b from 0 to 0.2 induces 5 new clusters; increasing b from 0.2 to 1/3 induces 3 new clusters; increasing b from 1/3 to 1/2 or from 1/2 to 1 induces one additional cluster; increasing b from 1 to any of the other values presented does not induce any more clusters. This table was generated with the `newclusters` function developed as part of this project. This function returns only clusters that we can be sure did not occur just by chance (see section 2.4). The entire table of new clusters is generated, rather than just the diagonal values, to prevent gradual changes being lost in the statistical noise.

From table 4, we see that at least one cluster was added after $b = 0.5$, so it is not strictly accurate to say that no new clusters occur after 0.5. However it does appear that most of the new clusters occur before 0.5. Also our dataset fails to demonstrates the existence of any new clusters occurring after $b = 1$. So it appears that few new clusters occur after $b = 0.5$.

Table 4. - Number of New Clusters that Form as b Increases

b	0	1/5	1/3	1/2	1	2	3	5	9	25	27	81	125	243	625	729
1/5	5															
1/3	7	3														
1/2	6	2	1													
1	6	2	2	1												
2	5	2	2	1	0											
3	4	1	1	1	0	0										
5	4	2	2	1	0	0	0									
9	3	1	1	1	0	0	0	0								
25	3	1	1	1	0	0	0	0	0							
27	3	1	1	1	0	0	0	0	0	0						
81	2	0	0	0	0	0	0	0	0	0	0					
125	1	0	0	0	0	0	0	0	0	0	0	0				
243	0	0	0	0	0	0	0	0	0	0	0	0	0			
625	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
729	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Inf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
b	0	1/5	1/3	1/2	1	2	3	5	9	25	27	81	125	243	625	729

Table 5. - Number of Clusters Lost as b Increases

b	0	1/5	1/3	1/2	1	2	3	5	9	25	27	81	125	243	625	729
1/5	5															
1/3	7	1														
1/2	7	2	1													
1	9	4	2	0												
2	10	6	6	3	0											
3	9	6	4	3	0	0										
5	10	7	7	4	1	0	0									
9	9	6	6	4	0	0	0	0								
25	13	10	11	8	5	4	3	1	0							
27	13	10	11	9	5	4	3	1	0	0						
81	13	10	10	9	5	4	3	1	0	0	0					
125	12	10	11	9	6	5	3	2	0	0	0	0				
243	13	10	11	9	5	4	3	1	0	0	0	0	0			
625	13	10	11	9	6	5	3	2	0	0	0	0	0	0		
729	13	10	11	9	7	6	4	3	1	1	1	0	0	0	0	
Inf	13	12	13	11	9	8	6	5	3	3	3	2	1	0	0	0
b	0	1/5	1/3	1/2	1	2	3	5	9	25	27	81	125	243	625	729

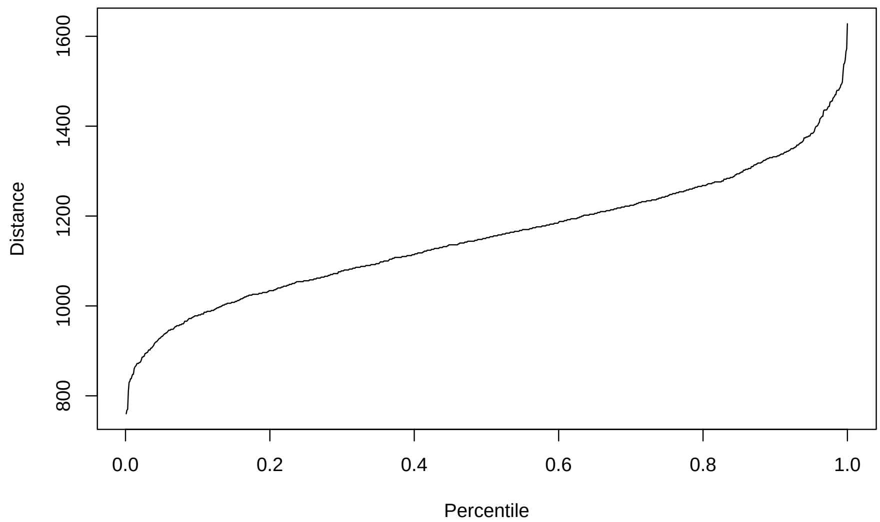
Do any changes in the clusters occur between $b = 1$ and $b = 3$? Each element of the table 5 represents the number of clusters we know exist for the smaller value of b but not the larger value. From this table we can see that our dataset does not demonstrate the loss of any clusters between $b = 1$ and $b = 3$. Similarly from table 4, we can see that no new clusters are known to occur between $b = 1$ and $b = 3$. As we expected, no statistically significant changes occur within this interval.

3.3. **Magnitude of Changes.** Looking at Figure 4, the distance between $b = 0$ and the points $b = 0.5, 1, 2, 3$ appears to be around 600. The exact distances are 596, 620, 620 and 630 respectively. As discussed in section 3.2, new clusters are formed till $b \approx 1$. After $b \approx 3$ it becomes impossible to reliably distinguish some of the rules[9]. We are primarily interested in the region where new clusters are being formed. Whether we take this region as ending at 0.5, 1,2 or 3 we have a distance in the range 596 to 630. This is 300 times the minimum possible change and 17-20%[10] of the maximum distance between two graphs.

3.3.1. *Distance between Random Trees.* We may want to compare a distance to the distance between the two random graphs. A nice property of distances between random trees is that there is a minimum distance between two random trees. As discussed in section 2.5.2, we define the distance between two trees as the distance between the lower triangular matrices representing the number of edges between each of the leaves. This matrix has 300 non-zero entries. The chance that two random matrices are the same in all 300 dimensions is vanishingly small.

We define a random tree as being a tree generated from a random distance matrix. Each element in the distance matrix is independently distributed uniformly between 0 and 1. As discussed in section 2.4, we use the single clustering method to generate the tree from the distance matrix.

Figure 5. - Cumulative Probability Distribution for Distance Between Two Random Trees



Percentile	0	0.05	0.1	0.25	0.5	0.75	0.9	0.95	1
Distance	760	932	980	1056	1152	1244	1332	1384	1628

1000 samples, Min=760, Median=1152, Mean =1153.7, Max=1628

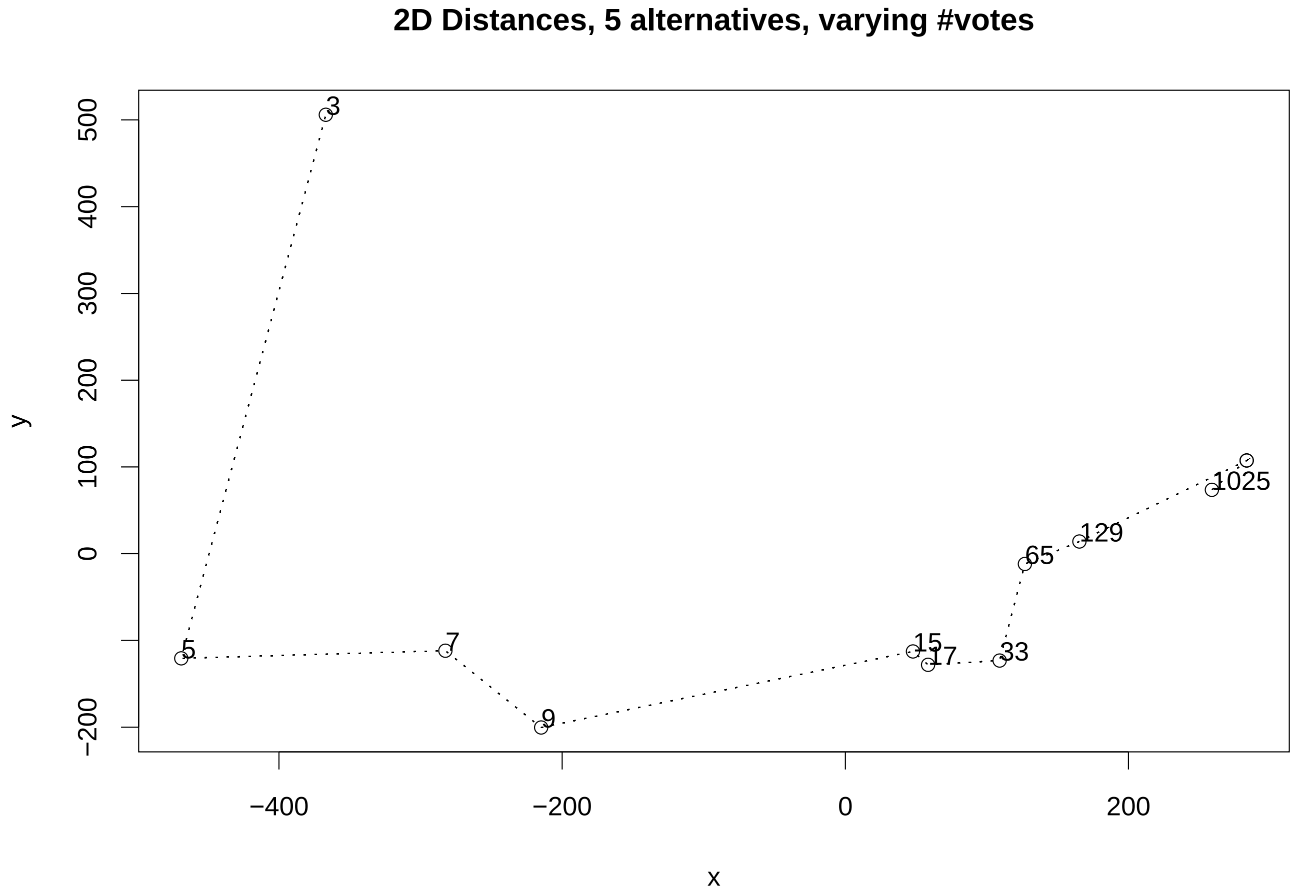
As we can see from figure 5, The distance between two random trees is generally between 932 and 1384.

[9]The only two pairs of rules which gave the same result in all 10000 runs for $b = 5$ were RU-HA and DQ-DO. However the Condorcet Efficient rules picked the same winner more than 99% of the time, making their clustering susceptible to random fluctuations. These fluctuations may introduce an artificial distance between the populations.

[10]The maximum tree distance for a tree with 25 leaves is between 3176 and 3422. The maximum is likely to be 3176 although we do not have a proof of this.

3.3.2. *Magnitude of Distance induced by Change in Number of Voters.* We may wish to compare the magnitude of the effect of varying of parameter of homogeneity b with that of varying the number of voters. As will be discussed below, varying the number of voters between reasonable values has a slightly greater effect than varying the parameter of homogeneity. From table 6, The distance of from 7 to 1025 voters is 620, the same distance caused by varying b from 0 to 1 or 2.

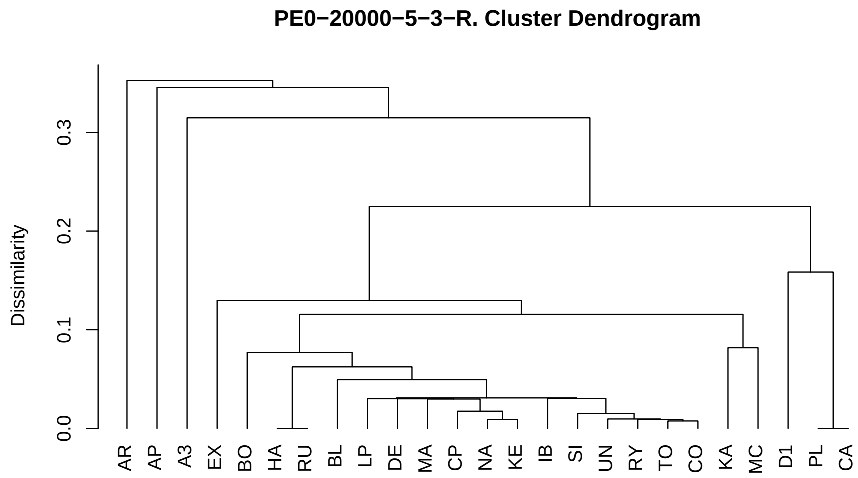
Figure 6. - Map of Tree Distances, Varying number of Voters, $b = 0$



The x, y Axes are unlabelled as they are arbitrary
Only the distance between points is meaningful

We can see from Figure 6 that the tree with 3 voters is substantially different from the other trees. It would be intuitive to believe that this distance is at least in part due to most of the rules giving the same result when there are only three voters; however as shown in figure 7 only two pairs of rules are identical. However the rules in the cluster {CO, CP, DE, IB, KE, LP, MA, NA, RY, SI, TO, UN} join together at approximately the same height, making the sub-clustering very unstable.

Figure 7. - 5 Alternatives, 3 Voters, Impartial Culture ($b = 0$)



As, shown in figure 7 above, three voters is enough to generate interesting elections where most of the rules are distinct. Thus we may consider 3 a reasonable number of voters, when comparing the effect of varying b and the number of voters within reasonable ranges.

Table 6. - Distance Matrix, Varying # Voters, 5 Candidates, impartial culture ($b = 0$)

	3	5	7	9	15	17	33	129	1025
3	0	716	672	744	754	770	794	758	786
5	716	0	406	466	624	632	672	690	774
7	672	406	0	298	482	436	482	514	620
9	744	466	298	0	368	378	430	490	592
15	754	624	482	368	0	136	148	346	376
17	770	632	436	378	136	0	106	348	362
33	794	672	482	430	148	106	0	296	354
129	758	690	514	490	346	348	296	0	172
1025	786	774	620	592	376	362	354	172	0

RUN-1

	3	5	7	9	15
3	0	568	658	738	708
5	568	0	352	428	580
7	658	352	0	250	536
9	738	428	250	0	470
15	708	580	536	470	0

RUN-2

	3	5	7	9	15
3	0	648	662	730	744
5	648	0	378	432	612
7	662	378	0	290	472
9	730	432	290	0	378
15	744	612	472	378	0

RUN-3

	3	5	7	9	15
3	0	728	716	796	770
5	728	0	448	488	646
7	716	448	0	294	444
9	796	488	294	0	366
15	770	646	444	366	0

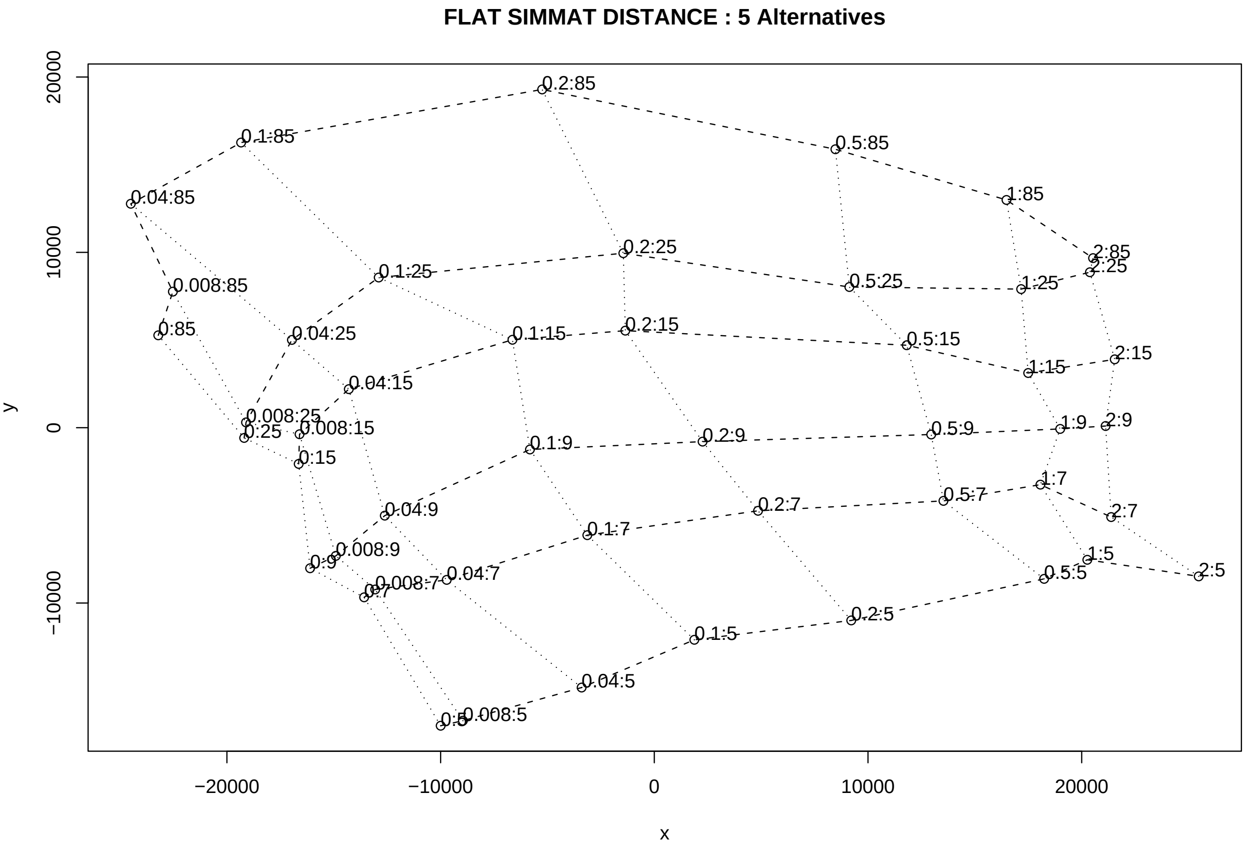
From table 6 above we can see that the minimum distance between the tree where $b = 3$ and the other trees is 672. This suggests that changing the number of voters from 3 has more effect than varying b from 0 to 3. This minimum distance of 672 occurs between $b = 3$ and $b = 7$ rather than the adjacent values $b = 3$ and $b = 5$, as we would expect. However, rerunning the simulation does not always give the same result. In run-1 $b = 5$ is significantly closer to $b = 3$, with a distance of 568.

3.4. **Is Varying the Homogeneity Equivalent to Varying the Number of Voters?**

As we increase the homogeneity of the Pólya-Eggenberger Urn Model, the number of unique votes decrease. We discovered a number of similarities between reducing the number of voters, and introducing homogeneity in the last section. For example, we found that increasing the homogeneity could decrease the similarity between Coombs' and exhaustive rules (CO-EX), which mirrored Shah's (2003) finding that reducing the number of voters could decrease the similarity between these two rules. There are a few things we can do to demonstrate that varying the homogeneity has a substantially different effect to varying the number of voters.

Firstly, look at the following graph. It uses multi-dimensional scaling to represent the flattened similarity matrix distances as closely as possible on a two dimensional surface. The labels on the vertices are of the format $b{:}v$, where b is the parameter of homogeneity and v is the number of voters. If varying b were equivalent to varying v, we would expect that the graph would form a line. However, as we can see the graph forms a clearly defined grid. Changing the homogeneity, is represented by a horizontal change, varying the number of voters results in a primarily vertical change. Thus, though similar, the effect of the parameter of homogeneity is not the same as the effect of varying the number of voters.

Figure 8. - Multidimensional Scaling Plot of Parameter of Homogeneity b v.s. Number of Voters



We may also see that the set of rules that diverge is different for increasing the parameter of homogeneity b and decreasing the number of voters, by comparing the next two tables.

Table 7. - Rules that Diverge as b is Increased from 0 to 1

	AP	KA	AR	BL	BO	EX	CP	HA	IB	LP	MC	DO	NA	PL	RU	SI	TO	UN	RY	CO	CA	D1	MA	DE	DQ	KE	A3
AP	.	.	.	1.	1.	1.	1.	.	1.	1.	1.	1.	1.	.	.	1.	1.	1.	1.	1.	.	.	1.	1.	1.	1.	
KA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AR	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BL	1.	.	.	.	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BO	1.	.	.	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EX	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.	.	.	.	.	.	.	.
CP	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
IB	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LP	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MC	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DO	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NA	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PL	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RU	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SI	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
TO	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
UN	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RY	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CO	1.	.	.	.	.	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MA	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DE	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.	.	.
DQ	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.	.	.	.
KE	1.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

1 indicates that the rules diverge.
Empty space indicates that the rules do not diverge.
This table uses a population with 85 voters and 5 candidates.

Note 3.1. The tables for increasing b to 0.1, 0.2 and 0.5 look much the same, except that anti-plurality diverges from even more rules.

Table 8. - Rules that Diverge as #voters is Decreased from 85 to 25

	AP	KA	AR	BL	BO	EX	CP	HA	IB	LP	MC	DO	NA	PL	RU	SI	TO	UN	RY	CO	CA	D1	MA	DE	DQ	KE	A3
AP				1	1	1	1	1	1	1	1	1	1			1	1	1	1	1	1		1	1	1	1	1
KA																											
AR				1					1	1		1	1			1			1				1	1	1	1	
BL	1		1										1			1			1							1	
BO	1																				1						
EX	1																			1							
CP	1																				1						
HA	1																				1						
IB	1		1							1		1				1			1					1	1	1	
LP	1		1						1				1			1			1		1						
MC	1																										
DO	1		1						1				1			1			1					1	1	1	
NA	1		1	1						1		1				1			1				1	1	1	1	
PL																											
RU																					1						
SI	1		1	1					1	1		1	1						1				1	1	1	1	
TO	1																										
UN	1																										
RY	1		1	1					1	1		1	1			1					1		1	1	1	1	
CO	1					1																					
CA	1				1		1	1		1					1				1				1			1	1
D1																											
MA	1		1										1			1			1		1						
DE	1		1						1			1	1			1			1							1	
DQ	1		1						1			1	1			1			1							1	
KE	1		1	1					1			1	1			1			1		1			1	1		
A3	1																				1						

1 indicates that the rules diverge.
Empty space indicates that the rules do not diverge.
This table uses an impartial population with 5 candidates.

From these two tables we can see that many more rules diverge as we decrease the number of voters, than diverge as we increase b. Increasing b, on the other hand makes anti-plurality diverge from more rules than decreasing the number of voters. Thus we may conclude that varying b does not have the same effect as varying the number of voters.

3.5. **The Effect of Excluding Condorcet Winners.** If we were to choose an electoral system, we may decide that we always want to choose the Condorcet winner when it exists. When deciding which rule to use as a backup, we would only be interested in elections where Condorcet winners do not exist. Thus we may be interested in which clusters form when we only include simulations where no Condorcet winner occurs.

For the following graphs a population with 5 candidates and 85 impartial voters was used.

Figure 9. - Including Condorcet Winners

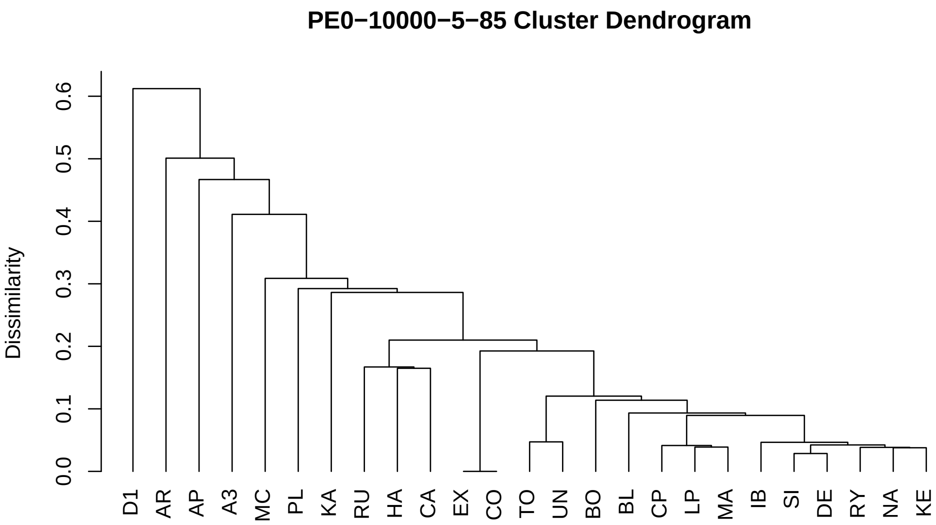
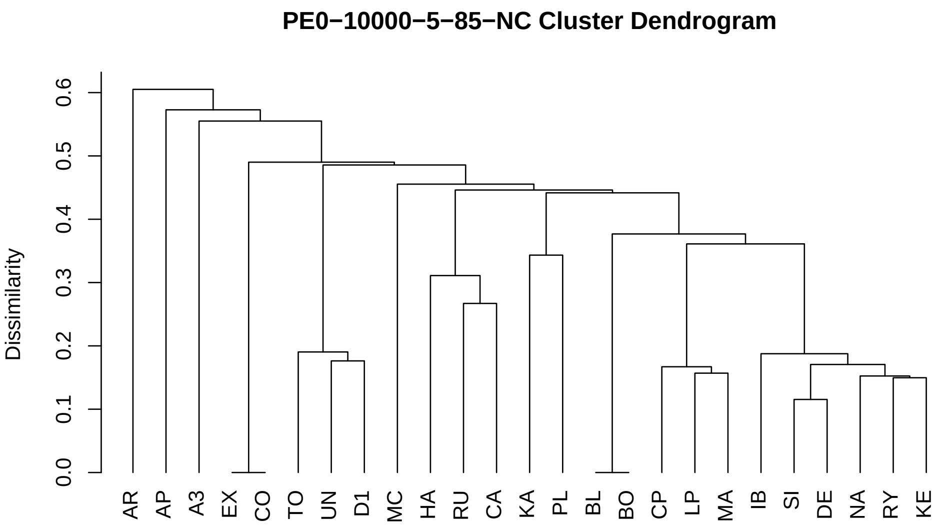


Figure 10. - Excluding Condorcet Winners



There does not appear to be a large difference caused by excluding Condorcet winners. As we would expect, excluding Condorcet winners had no effect on the clustering of Condorcet Efficient rules (see next page), with other Condorcet Efficient rules. The large number of profiles where Condorcet winners exist holds the Condorcet efficient rules close together. When the Condorcet winners are removed the Condorcet efficient rules become further apart, but they all become further apart by the same amount so the clustering between Condorcet efficient rules is unaffected.

Rule	Condorcet Efficiency
A3	0.6086
AP	0.5314
AR	0.5183
BL	1.0000
BO	0.8487
CA	0.8226
CO	0.9055
CP	1.0000
D1	0.2441
DE	1.0000
DO	1.0000
DQ	1.0000
EX	0.9055
HA	0.8872
IB	1.0000
KA	0.7339
KE	1.0000
LP	1.0000
MA	1.0000
MC	0.6959
NA.	1.0000
PL	0.5868
RU	0.8140
RY	0.9983
SI	1.0000
TO	1.0000
UN	1.0000

What is surprising is that excluding elections with Condorcet winners has little effect on most of the clusters containing Condorcet inefficient rules such as anti-plurality (AP) and random approvals (AR).

Only 3 new stable[a] clusters have formed, {D1, TO, UN}, {KA, PL}, and {BL, BO}.

Dictatorship, Top Cycle and Uncovered Set (D1, TO, UN) - The top cycle and uncovered set rules can pick a very large set of tied winners if no Condorcet winner exists. Since we break ties by the preferences of the first citizen, these rules will approximate a dictatorship when no Condorcet winner exists.

K-Approval and Plurality (KA, PL) - The plurality rule chooses the candidate with the most first place preferences, k-approval for 5 candidates will pick the candidate with the most first and second place votes. It is not obvious why removing Condorcet winners would cause these two rules to cluster together.

Black's rule and Borda's Rule (BL, BO) - By definition Black's rule picks the Borda count winner if no Condorcet winner exists, so it is unsurprising that these rules would cluster together when we exclude Condorcet winners.

[a]As defined in section 2.4.

4. Conclusions

The Pólya Eggenberger Urn is an effective model of partially homogeneous voting populations. Introducing homogeneity from the Urn model has some similarities to reducing the number of voters, for example both may cause Coombs' rule and the Exhaustive rule to diverge. There are also a number of differences between introducing homogeneity and reducing the number of voters. For example, introducing moderate homogeneity causes Borda's rule to cluster with the Random Approval rule, but reducing the number of voters does not. For a fixed parameter of homogeneity a the Urn model will not generate a consistent degree of homogeneity. This weakness can be remedied by defining a parameter of homogeneity b, which is adjusted for the number of candidates.

The most interesting range of b is 0 to 0.5. In general the similarity between most rules increases as we increase the homogeneity. Within the range 0 to 0.5, increasing the homogeneity actually reduces the similarity between a number of rules.

Removing elections which resulted in Condorcet winners resulted in little difference in the clustering of the social choice rules. When Condorcet winners were removed only three new clusters formed.

We have found the single clustering method an effective tool for studying social choice rules. Combined with the test described in section 2.4, the single clustering method can be used to quickly find significant differences between the clustering of different populations, differences that we can know are not due to chance. This certainty provides a foundation for the reports developed for this project, such as the number of new clusters that form.

We have also found that multidimensional scaling is an effective tool for visualising the effect that varying the parameters, of a randomly generated population, has on the relationships between the voting rules.

5. Appendix

Sample dendrograms with increasing homogeneity b follow:

Figure 11. - $b = 0$ (Impartial Culture), 85 Voters, 7 Candidates, 10000 samples

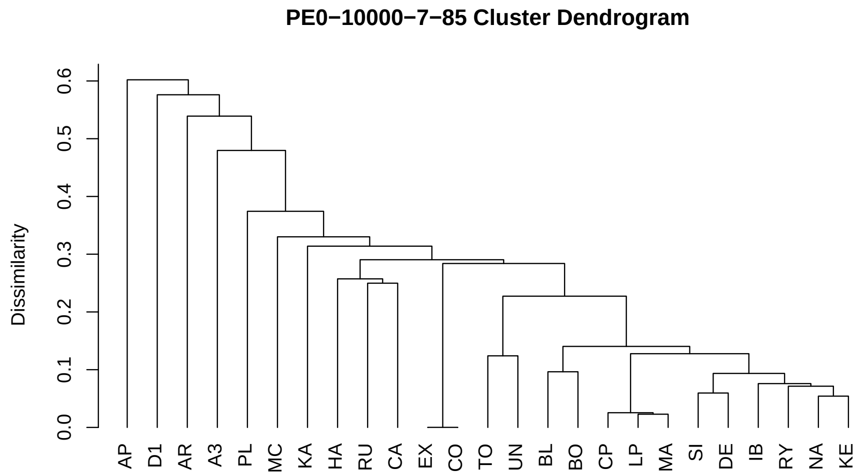


Figure 12. - $b = 0.1$,85 Voters, 7 Candidates, 10000 samples

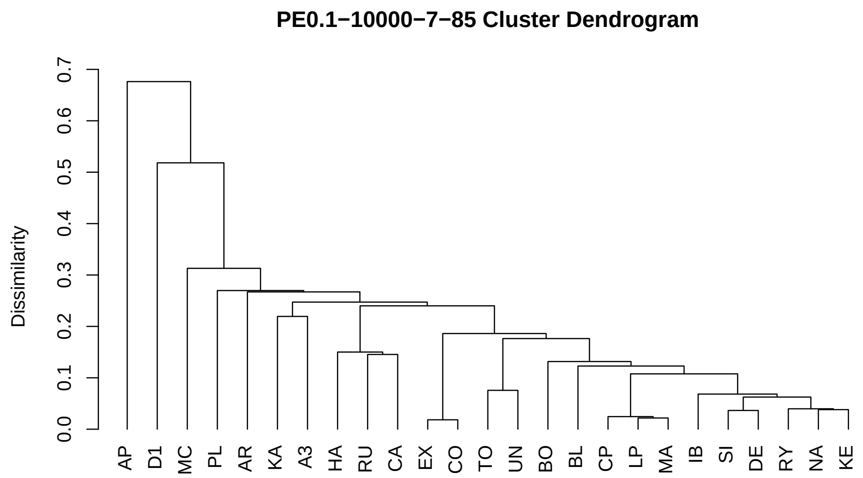


Figure 13. - $b = 0.2$, 85 Voters, 7 Candidates, 10000 samples

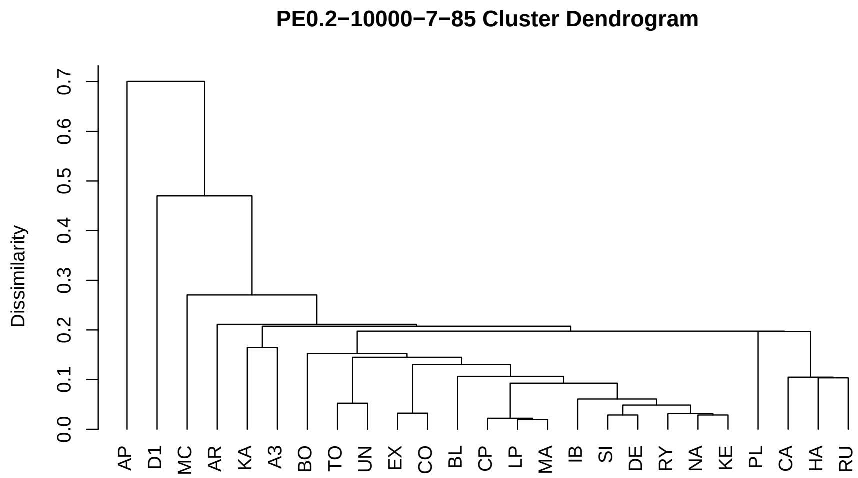


Figure 14. - $b = 0.5$, 85 Voters, 7 Candidates, 10000 samples

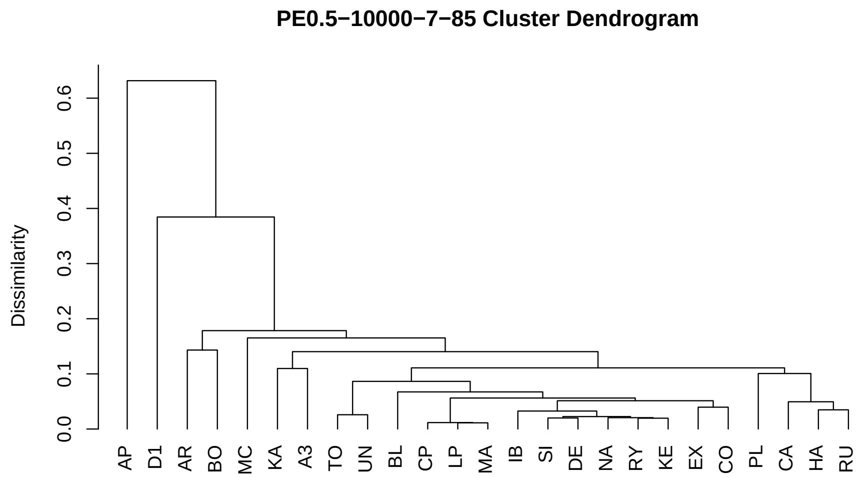
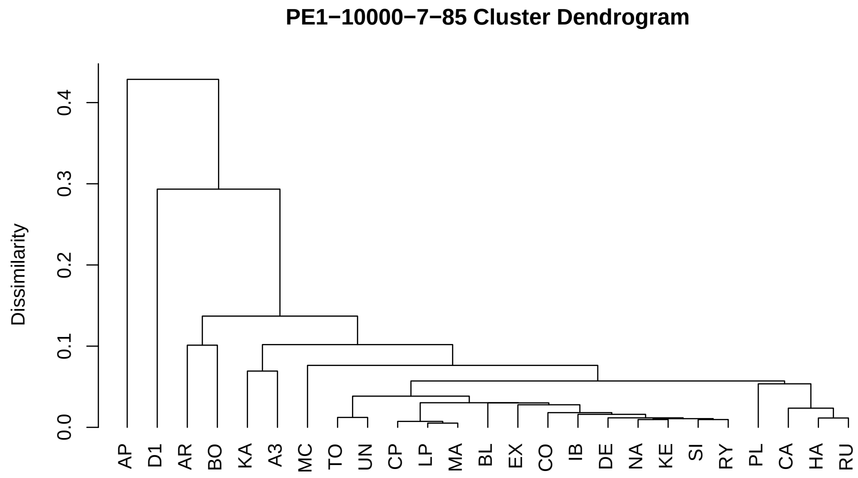


Figure 15. - $b = 1$, 85 Voters, 7 Candidates, 10000 samples



References

Aleskerov, F. and Kurbanov, E. Degree of manipulability of social choice procedures. *Current Trends in Economics*, pages 13–27, 1999. Eds: A. Alkan, C. D. Aliprantis and N. C. Yannelis.

Arrow, K. J. *Social choice and individual values*. Wiley, New York, 1951, 1963.

Black, D. *Theory of commitees and elections*. Cambridge University Press, Cambridge, 1958.

Bram, S. and Fishburn, P. *Approval Voting*. Birkhauser, 1982.

Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*. Wiley, New York, 2nd edition, 2001.

Gehrlein, W. V. and Lepelley, D. The probability that all weighted scoring rules elect the same winner. *Economic Letters*, 66:191–197, 2000.

Gibbard, A. Manipulation of voting schemes: A general result. *Econometrica*, 41, 1973.

Gower, J. C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53:325–328, 1966.

Kemeni, J. Mathematics without numbers. *Daedalus*, 88:577–91, 1959.

Lansdowne, Z. F. Outranking methods for multicriterion decision making: Arrow's and Raynaud's conjecture. *Social Choice and Welfare*, 14:125–128, 1997.

Laslier, J.-F. *Tournament solutions and majority voting*. Springer, Berlin - New York, 1997.

Laslier, J. P. Spatial approval voting, to appear.

Leung, W. *A Statistical Investigation of Social Choice Rules*. Master's thesis, Auckland University, 2001.

Merlin, V., Tataru, M., and Valognes, F. On the probability that all decision rules select the same winner. *Journal of Mathematical Economics*, 33:183–207, 2000.

Norman, L. J. and Samuel, K. *Discrete Distributions*. Houghton Mifflin, Boston, 1969.

Nurmi, H. *Comparing voting systems*. Reidel, 1987.

Nurmi, H. Computer simulation of voting systems. *Contemporary Issues in Decision Making*, 77:1:391–405, 1990. Elsevier Science B. V.

Pritchard, G. and Slinko, A. On the average minimum size of a manipulating coalition, 2003. Available http://www.math.auckland.ac.nz/deptdb/dept_reports/507.pdf.

Richelson, J. T. A comparative analysis of social choice functions IV. *Behavioral Science*, 26:346–353, 1981.

Satterthwaite, M. A. *Strategy-proofness and Arrow's Conditions: Existence and Correspondence*. 1975.

Sertel, M. R. and Yilmaz, B. The majoritarian compromise is majoritarian-optimal and subgame-perfect implementable. *Social Choice and Welfare*, 16:615–627, 1999.

Shah, R. *Statistical Mappings of Social Choice Rules*. Master's thesis, Stanford University, 2003.

Slinko, A. The majoritarian compromise in large societies. *Review of Economics Design*, 7(3):343–350, 2002.

Slinko, A. and Leung, W. Exploratory data analysis of common social choice functions II. *International Conference on the Problems of Control 17-19 June 2003 in Moscow*, 1:224–228, 2003.

Tideman, T. N. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4:185–206, 1987.