



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και
Υπολογιστών

Συγκριτική Αξιολόγηση Κανόνων Ψηφοφορίας για την Εκλογή Επιτροπών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΛΕΝΗ ΨΑΡΟΥΔΑΚΗ

Επιβλέπων : Δημήτριος Φωτάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2017



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και
Υπολογιστών

Συγκριτική Αξιολόγηση Κανόνων Ψηφοφορίας για την Εκλογή Επιτροπών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΛΕΝΗ ΨΑΡΟΥΔΑΚΗ

Επιβλέπων : Δημήτριος Φωτάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18η Οκτωβρίου 2017.

.....
Δημήτριος Φωτάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

.....
Νικόλαος Παπασπύρου
Αναπληρωτής Καθηγητής Ε.Μ.Π.

.....
Αριστέιδης Παγουρτζής
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2017

.....
Ελένη Ψαρουδάκη

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Ελένη Ψαρουδάκη, 2017.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η υπολογιστική θεωρία της κοινωνικής επιλογής είναι ένας πρόσφατα δημιουργηθείς κλάδος της θεωρητικής επιστήμης των υπολογιστών, που μελετάει την υπολογιστική πολυπλοκότητα σε προβλήματα ψηφοφοριών. Αυτή η περιοχή έχει προσελκύσει το ενδιαφέρον των ερευνητών τα τελευταία χρόνια λόγω των ποικίλων εφαρμογών της σε εκλογές, στην εύρεση των επικρατέστερων υποψηφίων, στα συστήματα προτάσεων κλπ.

Στόχος αυτής της διπλωματικής εργασίας είναι η διεξοδική πειραματική αξιολόγηση των κανόνων ψηφοφορίας για την εκλογή επιτροπών που χρησιμοποιούνται στις προαναφερθείσες εφαρμογές. Για τον λόγο αυτό υλοποιήσαμε τους πιο διάσημους κανόνες εκλογής επιτροπών - SNTV, STV, Bloc, *k*-Borda, Chamberlin-Courant, Monroe - και τους αξιολογήσαμε με γνώμονα ιδιότητες που αποζητώνται στα αποτελέσματα των εφαρμογών (π.χ. ποικιλομορφία, αναλογική αντιπροσωπευτικότητα κλπ.). Σκοπός μας είναι η πλήρης κατανόηση και εύρεση των πλέον κατάλληλων κανόνων για τις παραπάνω εφαρμογές. Για να το πετύχουμε αυτό, εκτελούμε αρχικά ένα πείραμα εκλογών στο δισδιάστατο χώρο, έτσι ώστε να υπάρχει η δυνατότητα οπτικοποίησης τόσο των προτιμήσεων των ψηφοφόρων και όσο και των αποτελεσμάτων των κανόνων.

Επιπλέον, σχεδιάζουμε και υλοποιούμε αποδοτικούς ευρεστικούς αλγορίθμους για τους κανόνες που υπολογιστικά ανήκουν στο NP (Chamberlin-Courant και Monroe), αξιολογούμε τα αποτελέσματα και σχολιάζουμε τις αδυναμίες των αλγορίθμων μας. Στη συνέχεια, αξιολογούμε τους κανόνες και τους αλγορίθμους μας σύμφωνα με την ικανότητα τους να βρίσκουν ευσταθείς επιτροπές. Μελετάμε αν το αποτέλεσμα τους μπορεί να επηρεαστεί από την ελεγχόμενη μεταφορά ή από την προσθήκη ενός ψηφοφόρου. Ερμηνεύουμε τα αποτελέσματά μας σύμφωνα με τις έννοιες της χειραγώγησης, του ελέγχου και της αποκάλυψης πληροφοριών για τον επιπλέον ψηφοφόρο.

Τέλος, σχεδιάζουμε ένα πείραμα για την μελέτη δεδομένων προερχόμενων από πραγματικές ψηφοφορίες. Περιγράφουμε τη λογική που κρύβεται πίσω από τις σχεδιαστικές μας αποφάσεις και πώς η δυσκολία εύρεσης κατάλληλων δεδομένων για την μελέτη κανόνων όπως οι Chamberlin-Courant και Monroe, οδηγεί την έρευνα μας σε διαφορετικά μονοπάτια.

Λέξεις κλειδιά

υπολογιστική θεωρία της κοινωνικής επιλογής, κανόνες εκλογής επιτροπών, ευρεστικοί αλγόριθμοι, ευστάθεια, χειραγώγηση, αποκάλυψη πληροφοριών, επικρατέστεροι υποψήφιοι, ποικιλομορφία, αναλογική αντιπροσωπευτικότητα

Abstract

Computational Social Choice is a new branch of theoretical computer science, that studies the computational complexity of preference aggregation problems, which formed only in early 2000's. This field has gained an enormous attention over the last years due to its various applications in parliamentary elections, portfolio selection, candidate short listing, recommendation systems, etc.

This diploma thesis aims to perform a thorough experimental evaluation of multiwinner voting rules that are used in the aforementioned applications. Therefore we implement the most popular multiwinner voting rules - SNTV, STV, Bloc, k -Borda, Chamberlin-Courant, Monroe - and evaluate them with respect to several properties relevant to these applications objectives (e.g. diversity, proportional representation, etc.). The objective is to fully understand which multiwinner voting rules fit better each of the applications above. Firstly we perform an experiment with elections generated according to the two-dimensional space, in order to appropriately visualize the preferences of the voters and the aggregated outputs of the rules.

Moreover, we design and implement efficient heuristics for multiwinner rules whose outcome is NP-hard computable (e.g. Chamberlin-Courant and Monroe), examine the produced outputs and comment on the heuristics' weaknesses. Afterwards, we evaluate the multiwinner voting rules and our heuristics according to their ability to find stable committees. We study if the outcome can be affected by the controlled movement of a voter or by the addition of a voter. We interpret the results under the concepts of manipulation, control, and disclosure of information.

Finally, we design an experiment for processing real election data, we describe the reasoning behind our design decisions and how the difficulty of finding appropriate data for rules such as Chamberlin- Courant and Monroe, lead us to different paths of future research.

Key words

computational social choice, multiwinner vote rules, heuristics, aggregation rules, stability, manipulation, disclosure of information, diversity, shortlisting, proportional representation

Ευχαριστίες

Η παρούσα διπλωματική εργασία σημαίνει και την ολοκλήρωση ενός σημαντικού κεφαλαίου της ακαδημαϊκής μου πορείας. Στο σημείο αυτό, θα ήθελα να ευχαριστήσω τους ανθρώπους που με βοήθησαν στη διαδρομή μου αυτή.

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κύριο Δημήτριο Φωτάκη, που με καθοδήγησε στην επιλογή ενός ιδιαίτερα ενδιαφέροντος και σύγχρονου θέματος και μου έδωσε την αφορμή να ασχοληθώ σε μεγαλύτερο βάθος με τον τομέα της Υπολογιστικής Θεωρίας της Κοινωνικής Επιλογής. Του οφείλω ένα ιδιαίτερα μεγάλο ευχαριστώ για τη βοήθεια, την εμπιστοσύνη, την ενθάρρυνση, το ενδιαφέρον και τις πολύτιμες συμβουλές σχετικά με τη διπλωματική μου εργασία, καθώς και για το μέλλον μου και τα επόμενα βήματα της ζωής μου. Του οφείλω επίσης ένα μεγάλο ευχαριστώ γιατί αποτέλεσε καθηγητή πρότυπο και χάρη σε εκείνον γνώρισα και αγάπησα τους αλγορίθμους.

Επίσης θα ήθελα να ευχαριστήσω τους αγαπημένους μου φίλους που με συντρόφευαν όλα αυτά τα χρόνια ομορφαίνοντας τη ζωή μου. Γιώργο, Άντα, Δήμητρα, Ελπίδα, Κατερίνα, Στέλλα, Μαρία, Χάρη, Δέσποινα, Ευαγγελία, Κωνσταντίνε, Γιώργο, Γιάννη, Δημήτρη, Γιάννη, Αλέξη, Γιώργο, Κωνσταντίνε και Μαριαλένα, σας ευχαριστώ για όλα όσα περάσαμε μαζί, τις βόλτες, τις συζητήσεις, τα ξενύχτια και τη φιλία σας. Πολλά ευχαριστώ στους φίλους μου Γρηγόρη και Βασίλη που εκτός από τη φιλία τους μου μετέδωσαν την αγάπη για τον προγραμματισμό. Ευχαριστώ ακόμη την ομάδα της βιβλιοθήκης για όσα μου προσέφερε αυτά τα χρόνια.

Τέλος θα ήθελα να ευχαριστήσω μέσα από την καρδιά μου τους γονείς μου, Λευτέρη και Άννα, και την αγαπημένη μου αδερφή, Μαριάνθη, για την αγάπη και τη συνεχή υποστήριξη που μου προσφέρουν τόσα χρόνια.

Ελένη Ψαρουδάκη,

Αθήνα, 18η Οκτωβρίου 2017

Contents / Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Contents / Περιεχόμενα	11
List of Tables / Πίνακες	15
List of Figures / Εικόνες	17
Part I English Text	21
1. Introduction	23
1.1 Motivation	23
1.2 Thesis contribution	24
1.3 Chapter outline	24
2. Background	27
2.1 Introduction	27
2.2 Three Main Types of Multiwinner Elections	28
2.2.1 Shortlisting	28
2.2.2 Selecting a Diverse Committee	28
2.2.3 Proportional Representation	29
2.3 Challenges	29
2.4 Definition of Multiwinner Voting Rules	29
2.4.1 Single Nontransferable Vote Rule (SNTV)	30
2.4.2 Single Transferable Vote Rule (STV)	30
2.4.3 Bloc Rule	32
2.4.4 k-Borda Rule	32
2.4.5 Chamberlin-Courant rule (β -CC)	33
2.4.6 Monroe rule (β -Monroe)	33
3. Design & Implementation of the Experiment over the Two-Dimensional Space	35
3.1 Election	35
3.1.1 Experimental Setup	35

3.1.2	Distribution of Voters & Candidates	36
3.1.3	Interpretation of the position at the two-dimensional plane	36
3.1.4	Assumptions	36
3.2	Results, Comments & Comparison of the Rules	38
3.2.1	Single Nontransferable Vote Rule	38
3.2.2	Single Transferable Vote Rule	40
3.2.3	Bloc Rule	41
3.2.4	k -Borda Rule	42
3.2.5	Chamberlin-Courant Rule	43
3.2.6	Monroe Rule	44
3.2.7	Comparisons of Single Transferable Vote, Chamberlin-Courant & Monroe rules	45
3.2.8	Comparisons over Properties of Rules	48
4.	Heuristics for NP-hard Rules	49
4.1	Chamberlin-Courant Rule	49
4.1.1	Chamberlin-Courant Heuristic	49
4.1.2	Initialization Technique	50
4.1.3	Proposed Algorithm	51
4.1.4	Proposed Algorithm's Weakness	51
4.1.5	Computational Complexity	52
4.2	Monroe Rule	53
4.2.1	Monroe Heuristic Algorithm	53
4.2.2	Different Initialization Techniques	57
4.2.3	Proposed Algorithm	59
4.2.4	Proposed Algorithm's Weakness	59
4.2.5	Computational Complexity	60
5.	Stability	61
5.1	Introduction	61
5.2	Controlled Movement of a Voter	62
5.2.1	Experimental Setup	62
5.2.2	Results	62
5.2.3	Conclusions	67
5.3	New Voter's Addition	68
5.3.1	Experiment	68
5.3.2	Results	68
5.3.3	Conclusions	73
5.4	Summary of Stability Results	75
6.	Design & Implementation of the Experiment over the Multi-Dimensional Space	77
6.1	Election	77
6.1.1	Experimental Setup	77

6.1.2	Representation of the Voters & Candidates	78
6.2	Results	79
7.	Future Work	81
	Μέρος II Ελληνικό Κείμενο	83
1.	Εισαγωγή	85
1.1	Κίνητρο	85
1.2	Συνεισφορά της Εργασίας	86
1.3	Οργάνωση Ελληνικού Κειμένου	86
2.	Θεωρητικό Υπόβαθρο	89
2.1	Εισαγωγή	89
2.2	Κύριοι Τύποι Εκλογών Επιτροπών	90
2.2.1	Εύρεση Επιτροπής με Γνώμονα την Εύρεση των Επικρατέστερων Υποψηφίων (shortlisting)	90
2.2.2	Επιλογή Επιτροπής με Γνώμονα την Ύπαρξη Ποικιλομορφίας (diversity)	90
2.2.3	Επιλογή Επιτροπής με Γνώμονα την Αναλογική Αντιπροσώπηση των Ψηφοφόρων (proportional representation)	91
2.3	Προκλήσεις	91
2.4	Ορισμός των Κανόνων Ψηφοφορίας για την Εκλογή Επιτροπών	92
2.4.1	Κανόνας Single Nontransferable Vote (SNTV)	92
2.4.2	Κανόνας Single Transferable Vote (STV)	92
2.4.3	Κανόνας Bloc	94
2.4.4	Κανόνας k-Borda	95
2.4.5	Κανόνας Chamberlin-Courant (β -CC)	95
2.4.6	Κανόνας Monroe (β -Monroe)	96
3.	Σχεδιασμός & Υλοποίηση του Πειράματος στο Δισδιάστατο Χώρο	97
3.1	Πείραμα	97
3.1.1	Πειραματική Διαδικασία	97
3.1.2	Κατανομή των Ψηφοφόρων & Υποψηφίων στο Χώρο	98
3.1.3	Ερμηνεία της Θέσης στο Δισδιάστατο Χώρο	98
3.1.4	Θεωρήσεις	98
3.2	Αποτελέσματα, Σχολιασμός & Σύγκριση των Κανόνων	100
3.2.1	Κανόνας Single Nontransferable Vote	100
3.2.2	Single Transferable Vote Rule	102
3.2.3	Κανόνας Bloc	103
3.2.4	Κανόνας k-Borda	104
3.2.5	Κανόνας Chamberlin-Courant	105
3.2.6	Κανόνας Monroe	106

3.2.7	Σύγκριση των Κανόνων Single Transferable Vote, Chamberlin-Courant & Monroe	107
3.2.8	Συγκρίσεις των Κανόνων σύμφωνα με τις Ιδιότητες	110
4.	Ευρετικοί Αλγόριθμοι για τους NP-hard Κανόνες	113
4.1	Κανόνας Chamberlin-Courant	113
4.1.1	Ευρετικός Αλγόριθμος για τον Κανόνα Chamberlin-Courant	113
4.1.2	Τεχνική Αρχικοποίησης	114
4.1.3	Προτεινόμενος Ευρετικός Αλγόριθμος	115
4.1.4	Αδυναμία του Προτεινόμενου Αλγορίθμου	115
4.1.5	Υπολογιστική Πολυπλοκότητα	116
4.2	Κανόνας Monroe	117
4.2.1	Ευρετικός Αλγόριθμος για τον Κανόνα Monroe	117
4.2.2	Διαφορετικές Τεχνικές Αρχικοποίησης	120
4.2.3	Προτεινόμενος Αλγόριθμος	122
4.2.4	Αδυναμία του Προτεινόμενου Αλγορίθμου	122
4.2.5	Υπολογιστική Πολυπλοκότητα	122
5.	Ευστάθεια	125
5.1	Εισαγωγή	125
5.2	Ελεγχόμενη Μετακίνηση ενός Ψηφοφόρου	125
5.2.1	Πειραματική Διάταξη	126
5.2.2	Αποτελέσματα	126
5.3	Προσθήκη ενός Νέου Ψηφοφόρου	130
5.3.1	Πειραματική Διάταξη	130
5.3.2	Αποτελέσματα	130
5.4	Σύνοψη των Αποτελεσμάτων	135
6.	Σχεδιασμός & Υλοποίηση του Πειράματος σε Πολυδιάστατο Χώρο	137
6.1	Πείραμα	137
6.1.1	Πειραματική Διαδικασία	137
6.1.2	Αναπαράσταση των Ψηφοφόρων & Υποψηφίων	138
6.2	Αποτελέσματα	139
7.	Μελλοντικές Κατευθύνσεις	141
	Bibliography / Βιβλιογραφία	143

List of Tables / Πίνακες

Part I English Text	23
2.1 The voters preferences over the set of candidates	30
2.2 Results of SNTV rule for the example of table 2.1	30
2.3 Preference ranking of 99995 voters.	31
2.4 STV rule's votes for each candidate in every round	32
2.5 Results of Bloc rule for the example of table 2.1	32
2.6 Results of k -Borda rule for the example of table 2.1	33
3.1 Proportional representation	48
3.2 Diversity	48
5.1 Stability of Multiwinner Vote Rules	75
5.2 Disclosure of information over the additional voter	75
Μέρος II Ελληνικό Κείμενο	85
2.1 Οι προτιμήσεις των ψηφοφόρων στο σύνολο των υποψηφίων	92
2.2 Αποτελέσματα του κανόνα SNTV για το παράδειγμα του πίνακα 2.1	92
2.3 Σειρές κατάταξης των υποψηφίων από τους 99995 ψηφοφόρους.	94
2.4 Οι ψήφοι κάθε υποψηφίου σε κάθε γύρο του κανόνα STV	94
2.5 Αποτελέσματα του κανόνα Bloc για το παράδειγμα του πίνακα 2.1	95
2.6 Αποτελέσματα του κανόνα k -Borda για το παράδειγμα του πίνακα 2.1	95
3.1 Αναλογική αντιπροσωπευτικότητα	110
3.2 Ποικιλομορφία	111
5.1 Ευστάθεια των κανόνων εκλογής επιτροπών	135
5.2 Αποκάλυψη πληροφοριών για τον νέο ψηφοφόρο	135

List of Figures / Εικόνες

Part I English Text	23
2.1 Single Transferable Vote rule	31
3.1 SNTV rule's outcome	38
3.2 SNTV rule outlines the candidates' distribution	39
3.3 SNTV rule's outcome vastly depends on the candidates distribution	39
3.4 Different outcomes for slightly different versions of STV rule	40
3.5 STV rule's outcomes indicate that STV rule guarantees proportional representation	41
3.6 Bloc rules' outcome consist near the edges of small groups	41
3.7 Bloc rule's outcome depends on the density of the candidates' distribution	42
3.8 k -Borda rule's outcome	42
3.9 k -Borda rule chooses candidates located to the voters' center of mass	43
3.10 Chamberlin-Courant rule's outcomes	43
3.11 Chamberlin-Courant rule's outcome, when candidates' do not span enough	44
3.12 Monroe rule's outcomes underlines voters' distribution	44
3.13 Comparisons for uniform distribution	45
3.14 Comparisons for normal distribution	45
3.15 Comparisons for unequal normal distributions	46
3.16 Comparisons for unequal normal distributions	47
4.1 Non-equivalent results for Chamberlin-Courant heuristic	49
4.2 Results for Chamberlin-Courant heuristic before and after k -means++ initialization.	51
4.3 The weakness of the proposed heuristic for Chamberlin-Courant rule	52
4.4 Assignment of candidates with different approaches for Monroe heuristic for the same voters, candidates and initialization points (chosen uniformly at random) . . .	54
4.5 Equivalent results for Monroe heuristic [1]	55
4.6 Equivalent results for Monroe heuristic [2]	55
4.7 Non-equivalent results for Monroe heuristic [3]	56
4.8 Non-equivalent results for Monroe heuristic [4]	56
4.9 Optimal results for Monroe heuristic	57
5.1 Single Nontransferable Vote rule - stable	63
5.2 Single Transferable Vote rule - stable	63
5.3 Bloc rule - stable	64

5.4	<i>k</i> -Borda rule - stable	64
5.5	Chamberlin-Courant heuristic - almost stable	65
5.6	Monroe heuristic - unstable	65
5.7	Movement of a single elected member depends on the candidates' distribution	66
5.8	The stability of Monroe heuristic as the two distributions start to overlap	66
5.9	Single Nontransferable Vote rule - stable	69
5.10	Conclusions over Single Nontransferable Vote rule	69
5.11	Single Transferable Vote rule - unstable	70
5.12	Different results for different versions of Single Transferable Vote rule	70
5.13	Bloc rule - stable	71
5.14	Conclusion over the Bloc rule	71
5.15	<i>k</i> -Borda rule - stable	72
5.16	Chamberlin-Courant heuristic - unstable	73
5.17	Monroe heuristic - unstable	73
 Μέρος II Ελληνικό Κείμενο		85
2.1	Κανόνας Single Transferable Vote	93
3.1	Αποτέλεσμα του κανόνα SNTV	100
3.2	Το αποτέλεσμα του SNTV σχηματίζει το περίγραμμα της κατανομής των υποψηφίων	101
3.3	Το αποτέλεσμα του SNTV επηρεάζεται σημαντικά από την κατανομή των υποψηφίων	101
3.4	Διαφορετικά αποτελέσματα για κοντινές παραλλαγές του κανόνα STV	102
3.5	Ο STV δείχνει να εγγυάται την αναλογική αντιπροσωπευτικότητα	103
3.6	Οι εκλεγμένοι υποψήφιοι από τον κανόνα Bloc σχηματίζουν ομάδες κοντά στις άκρες της κατανομής των υποψηφίων.	103
3.7	Τα αποτελέσματα του κανόνα Bloc εξαρτώνται από την πυκνότητα της κατανομής των υποψηφίων	104
3.8	Αποτελέσματα του κανόνα <i>k</i> -Borda	104
3.9	Ο κανόνας <i>k</i> -Borda επιλέγει υποψήφιους που βρίσκονται κοντά στο κέντρο της μάζας της κατανομής των ψηφοφόρων	105
3.10	Αποτελέσματα του κανόνα Chamberlin-Courant	105
3.11	Το αποτέλεσμα του Chamberlin-Courant, όταν η κατανομή των υποψηφίων δεν καλύπτει το χώρο που καλύπτει η κατανομή των ψηφοφόρων	106
3.12	Το αποτέλεσμα του κανόνα του Monroe σκιαγραφεί την κατανομή των ψηφοφόρων	107
3.13	Συγκρίσεις για ομοιόμορφη κατανομή των ψηφοφόρων	107
3.14	Συγκρίσεις για κανονική κατανομή των ψηφοφόρων	108
3.15	Συγκρίσεις για άνισες κανονικές κατανομές ψηφοφόρων	108
3.16	Συγκρίσεις για άνισες κανονικές κατανομές ψηφοφόρων	109
4.1	Μη ισοδύναμα αποτελέσματα για τον ευρετικό αλγόριθμο του κανόνα Chamberlin-Courant εξαιτίας διαφορετικών αρχικοποιήσεων	113

4.2	Αποτελέσματα για τον ευρετικό αλγόριθμο του Chamberlin-Courant πριν και μετά την αρχικοποίηση του k -means++	115
4.3	Αδυναμία του προτεινόμενου ευρετικού αλγορίθμου του κανόνα Chamberlin-Courant	116
4.4	Προτεινόμενες επιτροπές διαφορετικών προσεγγίσεων του ευρετικού αλγορίθμου του Monroe για τους ίδιους ψηφοφόρους, υποψηφίους και σημεία αρχικοποίησης .	119
4.5	Ισοδύναμα αποτελέσματα του ευρετικού αλγόριθμου του κανόνα [1]	119
4.6	Μη ισοδύναμα αποτελέσματα του ευρετικού αλγορίθμου του κανόνα Monroe [2]	120
4.7	Βέλτιστα αποτελέσματα του ευρετικού αλγορίθμου του κανόνα Monroe	120
5.1	Κανόνας SNTV - ευσταθής	126
5.2	Κανόνας STV - ευσταθής	127
5.3	Κανόνας Bloc - ευσταθής	127
5.4	Κανόνας k -Borda - ευσταθής	127
5.5	Ευρετικός αλγόριθμος Chamberlin-Courant - «σχεδόν» ευσταθής	128
5.6	Ευρετικός αλγόριθμος Monroe - ασταθής	129
5.7	Η ευστάθεια του ευρετικού αλγορίθμου του Monroe καθώς οι κατανομές αρχίζουν να επικαλύπτονται	129
5.8	Κανόνας SNTV - σταθερός	131
5.9	Συμπεράσματα σχετικά με τον κανόνα SNTV	131
5.10	Κανόνας STV - ασταθής	132
5.11	Κανόνας Bloc - ευσταθής	132
5.12	Συμπεράσματα για τον κανόνα Bloc	133
5.13	Κανόνας k -Borda - ευσταθής	133
5.14	Ευρετικός αλγόριθμος για τον κανόνα Chamberlin-Courant - ασταθής	134
5.15	Ευρετικός αλγόριθμος για τον κανόνα Monroe - ασταθής	134

Part I

English Text

Comparative Evaluation of Multiwinner Voting Rules

Chapter 1

Introduction

1.1 Motivation

The multiwinner election is the process of electing a committee, i.e. a subset of the candidates (objects or individuals), in order to represent the voters in a certain way. Everyday business decisions, elections, etc. are trying to aggregate individual and possibly conflicting preference profiles into a collective preference profile. But if there are many different ways to aggregate these preferences, how do we find a committee of representatives that satisfy criteria such as representativity and accountability? How do we ensure that this outcome is not manipulated by a certain person? How will each user benefit from the outcome according to his or her own preferences?

The last few years these questions are troubling researchers of the recently formed field of Computational Social Choice. There have been many theoretical evaluations over the variety of the multiwinner voting rules. Each one of them satisfies different criteria and on many occasions, it is particularly designed for different applications. Choosing a good multiwinner voting rule is not trivial and before choosing one we should ask ourselves what are the needed properties and how the outcome will be used.

Consequently, there has been a classification of the multiwinner election according to the main properties that the outcome must have. These properties are proportional representation (when the composition of the committee should globally reflect the will of the voters), diversity (when the composition of the committee should recognize the variety of voters) and fairness (for applications such as shortlisting, i.e. when similar candidates must be either selected or rejected both). Another property that is vastly studied is this of stability, in many different interpretations e.g. manipulation. Thus, we decided to make a thorough experimental evaluation of multiwinner voting rules in a simpler domain, i.e. two-dimensional space, according to these properties.

Another challenge in the field of Computational Social Choice is some of the voting rules' computational complexity. Many of the rules are NP-hard computable and therefore not applicable to large sets due to time or/and space constraints. Many heuristics and approximation algorithms are proposed and afterward studied under the same theoretical and algorithmic properties. Hence, we created our own heuristics to evaluate for the requested algorithmic properties.

The field of Computational Social Choice is still young and offers many challenges, especially for multiwinner elections. The importance of this field is still being established with already applications in the recommendation systems, for web search engines to decide which pages to display for a given query, etc.

1.2 Thesis contribution

In the context of this thesis, we analyze the most popular multi-winner voting rules. During our experiments in the two-dimensional space, we compared the rules, highlighted their design and feature similarities and differences, and evaluated them with respect to several properties, i.e. proportional representation and diversity. Special focus has been given to the property of stability, as it is essential in many different applications. Within the scope of this work, we limited our research to the two experiments regarding stability, the movement of a voter and the addition of a new one.

Some of the multi-winner voting rules that were studied, are NP-hard to compute i.e., Chamberlin-Courant and Monroe. Instead of running long and massive simulations in order to evaluate these algorithms, we designed and implemented bespoke heuristics focussing on efficiency. These heuristics were later evaluated with respect to the same properties, and their results were thoroughly compared against the baseline with respect to their accuracy. Designing well-defined heuristics and approximation algorithms for NP-hard problems is an active and challenging domain of research, especially for the Monroe rule, due to the challenging and specific requirements.

Finally, we extended our experiments in the multidimensional space, which is more applicable to real life problems. We successfully implemented the voting rules and adapted our heuristics in order to handle real-life data. The discrete multidimensional space created the challenge, due to the lack of relative positions between the candidates and the voters, which we overcame in this work. The validation of the approach was successful over the available dataset. That being said, the lack of appropriate data large scale dataset, proved to be a limitation of the validation process. The data selection in order to fully validate the approach described in this thesis is part of active research and is left as future work.

1.3 Chapter outline

In chapter 2, we present the theoretical background upon which the work is based, clarifying concepts and terminologies introduced later in the dissertation. More specifically, we present the concepts of computational social choice, we discuss the known multiwinner voting rules and describe applications and the desirable properties according to them.

In chapter 3, we describe the experimental procedure and the reasoning behind our design decisions for the experiments in two-dimensional space. In the second part of this chapter, we compare and contrast the multiwinner voting rules according to properties e.g. diversity, and discuss the suitability and expected problems for the use of each voting rule. We also include graphical representations and comments for some of the results of the experimental runs.

In chapter 4 we propose our own heuristics for Chamberlin-Courant and Monroe rules, whose outcome is NP-hard to compute, based on their intrinsic properties. We examine the produced outcomes and decide if our algorithms are good enough, commenting their weaknesses.

In chapter 5, we evaluate the multiwinner voting rules according to the ability to find stable committees. In the first part of the chapter, we study if the outcome can be affected by the movement of one random voter, while in the second part by the addition of a voter.

In chapter 6, we describe the experimental procedure and the reasoning behind our design decisions for processing real election data. In the second part of this chapter, we present the results, commenting how our design decisions affected the outcomes.

In chapter 7, we list planned improvements to our heuristics and propose different approaches that could be implemented and offer performance gains to the election process. Finally, we present possible extension in the study of stability and propose future research paths.

Chapter 2

Background

2.1 Introduction

Every day in our lives we may watch the results of an election or even be part of one as voters or candidates. In our everyday experience, we associate the word elections with politics, but there are many more voting procedures. We may want to elect a representative, find the best day and hour for a meeting or even find the winner of a contest as Eurovision. Each one of the aforementioned examples characterizes a different setting, due to the usage of different voting rules and different objectives. In most of the political systems, we elect the representatives with Single Nontransferable Vote rule, while in Doodle approval voting is used [1]. In Eurovision song contest, a variation of Borda voting rule is applied to select the winner. Sporting activities apply various multi-criteria methods in their championship regulations, e.g. Formula 1 racing uses a variation of the Borda method in order to assign points to the candidates and find the winner [2].

The following examples, even if the nature of the examples differs, the objective remains the same: “find (elect) the winner” and they are called single-winner elections. That being said, there is another family of elections, where the objective is to choose a group of candidates, i.e. a committee. These are called multiwinner elections and they are far more widespread and varied than the single-winner ones. Parliamentary elections, business decisions, portfolio/movie selection, recommendation systems, shortlisting are only some scenarios of multiwinner elections and require rules with different principles.

There are three main types of multiwinner elections [3] based on their objective:

1. shortlisting,
2. selecting a diverse committee, and
3. proportional representation.

In the next sections, we will analyze each category, give examples and afterward comment on the challenges of the multiwinner elections. Lastly, we will present the rules that we used in our experiments.

2.2 Three Main Types of Multiwinner Elections

2.2.1 Shortlisting

Prior to deciding the winner of a contest, the receiver of an award, the selection of the ideal candidate for a job, etc., there is typically a procedure that finds the finalists. The shortlisting rules (also called screening rules [4]) are those rules who can be thought as preliminary selections of candidates for the subsequent ultimate choice of a single candidate. Since only one candidate will be ultimately selected, it must be the best one. The individuals with the highest quality should be picked and any dependencies and interactions of the candidates should not matter (we assume that the voters understand any special requirements, e.g. a level of diversity in order to avoid discriminations in the group of finalists, and that these are reflected in the voters' preference orders) [5], i.e. two similar candidates should either be both elected or eliminated¹. The finalists are then evaluated far more accurately, e.g. with an interview, and the final choice is made; hence this procedure is beyond the scope of shortlisting.

For shortlisting, we could give an example from academia. Consider a situation where a doctoral position is filled at a university. The assigned faculty members rank applicants in order to create a short-list of those to be invited for an interview [3]. Another example could be the procedure of choosing the nominees for an Oscar award. We describe the procedure simplified; there are certain criteria in order to be nominated, i.e. to become a potential nominee and afterward a voting procedure that turns the potential nominee into an official nominee. Finally, the winners are decided with a new voting procedure, as the nominee of each category with the most votes.

2.2.2 Selecting a Diverse Committee

There are many examples of elections where we primely care about the diversity of the elected committee. In these cases the candidates cannot be evaluated independently, i.e. if there are two similar candidates then we may either select one of them or neither (if there are better options), but we should not select them both.

The problem of electing a diverse committee resembles the facility location problem. Consider the task of selecting locations for fire stations in a city. Even though there is a point that minimizes the driving distance for all the points in the city, we do not want to build all the fire stations in that point; we would rather prefer to distribute them more uniformly so that each point of the city is close to a fire station.

We will give an example for selecting a diverse committee from business decisions that must be made. Consider a situation where a store has limited storage space for DVDs. Let assume there is a fixed number of different DVD's that the owners can procure. One of the best strategies is to source a set of DVD's which is as diverse as possible, keeping in mind that each customer should see something appealing to them.

¹ Possible exception could be boundary cases. We will study such cases in the following chapters, as they can affect the outcome.

2.2.3 Proportional Representation

The best-known type of multiwinner elections are parliamentary elections, where an important desideratum is the proportional representation of the views of the society (i.e. the voters). Thus the main objective of proportional representation is to find a committee of k -representatives, each associated with an equally sized constituency of approximately n/k voters (where n is the total number of voters).

An example could be similar to the one used in the previous section. Consider again a situation where a store, with limited storage space for DVD's. Let assume there is a fixed number of different DVD's that they can procure and each one must be bought in an exact number. One of the best strategies is to source the k DVD's that are found appealing in n/k customers, keeping in mind that each customer should find a DVD to buy.

2.3 Challenges

We will focus on two challenges that multiwinner elections lead to [5]. The first point of focus has already been mentioned in this chapter and pertains to the problem of the selection of the appropriate voting rule for a particular application (or generally for an election type). How can a user choose a good voting rule for his or her application? There are several answers to this question. First, there are some voting rules which are designed for certain tasks (e.g. Monroe rule). Secondly, there is an approach of seeking axiomatic properties in order to judge the suitability of a multiwinner vote rule for a particular application[3], e.g. to check whether a rule for finding a proportional committee satisfies Dummett's proportionality [6]. Finally, one can use empirical analysis to compare different rules under particular conditions, including theoretical and experimental evaluations [7], [8], or considering what various rules do on certain simpler domains, where their behavior can be interpreted intuitively [9]. All the approaches above are important and the choice of a voting rule should take all of them into account.

The second point of focus regards our ability to compute the results of multiwinner elections. In the single-winner setting, almost all prominent voting rules are polynomial-time computable, although there are important exceptions, such as the rules of Dodgson [10], Young [11] and Kemeny [12]. For the multiwinner setting, the situation is much more complex. There is a number of polynomial-time computable rules, but many interesting ones are NP-hard, e.g. Chamberlin-Courant [13] and Monroe [14] rules. We can approach this problem either through fast heuristics or approximation algorithms, deterministic or randomized. Such algorithms can be viewed as new, easy to compute, rules, which even sometimes correspond to previously known voting rules. Thus, we study axiomatic properties of the rules defined by such approximation algorithms, just as we do for the original voting rules.

2.4 Definition of Multiwinner Voting Rules

In this Diploma Thesis, we implement the most popular multiwinner voting rules e.g., Single Nontransferable Vote, Single Transferable Vote, Bloc, k -Borda, Monroe, Chamberlin-Courant rules. We present the algorithms and the concepts behind every rule.

2.4.1 Single Nontransferable Vote Rule (SNTV)

This rule is the multiwinner version of *plurality* voting rule. The votes are assigned to candidates by voters' first preference. The k candidates that are elected to the committee, are these with the highest plurality scores.

Example. Let us consider the following simple example. We have a 6 voters and a set of candidates $C = \{a, b, c, d, e\}$. Our objective is to elect a committee with $k = 2$ members.

voter	ballot
v_1	: $b \succ a \succ c \succ d \succ e$
v_2	: $e \succ a \succ b \succ d \succ c$
v_3	: $d \succ a \succ b \succ c \succ e$
v_4	: $c \succ b \succ d \succ e \succ a$
v_5	: $c \succ b \succ e \succ a \succ d$
v_6	: $b \succ d \succ c \succ e \succ a$

Table 2.1: The voters preferences over the set of candidates

For SNTV rule only the first preference matters. As a result v_1 gives one vote to candidate b , v_2 one vote to candidate a , etc. We present the results in the following table and the elected committee is $\{b, c\}$.

	Candidates				
	a	b	c	d	e
Total Votes	0	2	2	1	1

Table 2.2: Results of SNTV rule for the example of table 2.1

2.4.2 Single Transferable Vote Rule (STV)

The multiwinner version of this rule proceeds as follows: at each stage, the votes are assigned to candidates by voters' first preference. If candidates, whose plurality score is at least equal to a computed quota, exist, then they are elected to the committee and all the excess votes are transferred from winners to hopefuls. Otherwise one or more candidates with the lowest plurality scores are dropped from all ballots and their votes are transferred to hopefuls². We repeat this process until all the seats of the committee are full.

In this rule, any candidate, in order to be elected, must have a minimum number of votes - quota. The quota we used is the following:

$$quota = \text{floor}\left(\frac{\text{valid votes to cast}}{\text{seats to fill} + 1}\right) + 1$$

The excess votes are transferred from winners to hopefuls according to the following formula:

$$\left(\frac{\text{votes for second preference}}{\text{total votes of winner}}\right) * \text{surplus votes}$$

² At this point we ignore the procedure of transferring the votes or how we handle the ties while removing candidates. For more information please see section 3.2.

The Single transferable vote method exists in a number of different formulations and has complex mechanisms behind it [15]. The exact algorithm we used is presented in figure 2.1.

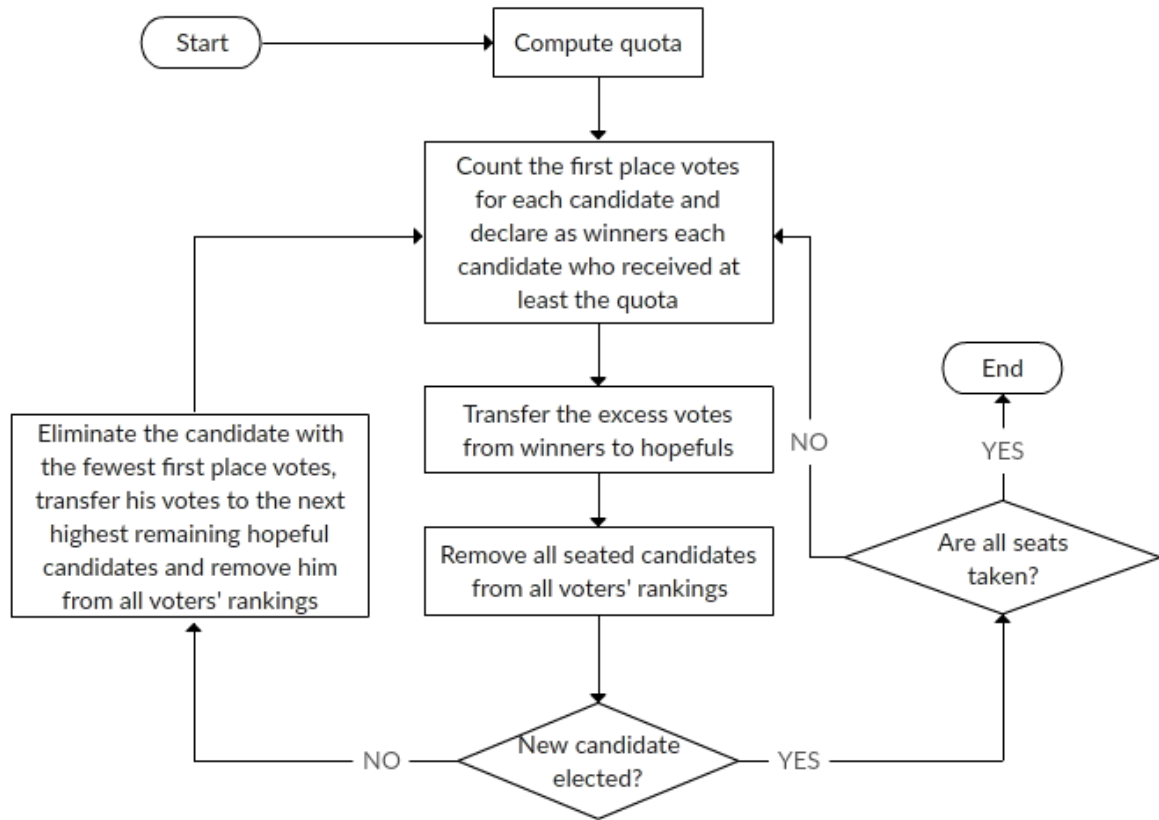


Figure 2.1: Single Transferable Vote rule [16]

When a candidate is eliminated, the votes, gained from election or elimination of other candidates are being transferred to the candidate's next preferences. This process repeats until either a winner is found for every seat or there are as many seats as the remaining candidates.

Example. We will use the example where a set of preferences of 99995 voters are given [16]. Let assume that four seats will be filled. In the table 2.3 first row gives column identification. The second row is the number of voters with each preference. The third row gives first preference of each set of voters. Lower preferences are given in descending order through the tenth row.

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII
4111	3500	100	2722	10490	10511	13936	1595	8000	7639	4000	12051	4095	3153	2422	1020	2150	8500
A	A	A	A	B	B	C	D	D	E	E	F	G	G	G	H	H	H
C	D	D	H	A	C	H	C	G	F	H	B	A	A	A	C	E	F
H	G	G	C	C	H	D	H	E	D	B	H	C	D	H	D	G	E
E	C	C	G	H	F	F	F	H	H	C	G	H	H	E	E	C	G
G	H	H	F	E	G	E	B	B	B	G	E	B	C	F	G	F	A
D	E	B	D	G	A	A	E	F	G	D	C	D	E	B	B	B	C
F	B	F	E	F	E	B	G	C	A	A	A	F	B	D	A	A	D
B	F	E	B	D	D	G	A	A	C	F	D	E	F	C	F	D	B

Table 2.3: Preference ranking of 99995 voters.
Bold indicates preferences considered by STV rule

The quota is 20000. Candidate c_B has 210001 first-place votes, exceeding the quota with residual of 1001 votes. A fraction of $(20000/21001$ or 0.952) of each vote is sufficient to seat candidate c_B . The

remainder of each vote is distributed to candidates c_A and c_C each voter's second preference. Table 2.4 presents the number of votes held by each candidate through the various rounds of the election, elimination, and transfer. A number in bold indicates a candidate's election, while an underlined number indicates a candidate's elimination. The elected committee is $[B, C, G, H]$.

Round	A	B	C	D	E	F	G	H
1	10433	21001	13936	9595	11639	12051	9670	11670
2	10933	-	14437	<u>9595</u>	11639	12051	9670	11670
3	<u>10933</u>	-	16032	-	11639	12051	17670	11670
4	-	-	20643	-	11639	12051	21270	14392
5	-	-	-	-	12116	<u>12051</u>	-	14828
6	-	-	-	-	12116	-	-	27879
7	-	-	-	-	19995	-	-	-

Table 2.4: STV rule's votes for each candidate in every round

2.4.3 Bloc Rule

This rule is the multiwinner version of k -approval voting rule. Voters assign a fixed number of points among the candidates. The k candidates who maximize the number of points are declared winners and are elected to the committee.

Example. Let us consider again the example of table 2.1. Since the elected committee will have two members each voter chooses his favorite two voters and gives a vote in each of them. Voter v_1 gives one vote to candidate b and one to candidate a , voter v_2 gives one vote to candidate e and one to candidate a , etc. The final results are in table 2.5 and the elected committee is $\{a, b\}$.

	Candidates				
	a	b	c	d	e
Total Votes	3	4	2	2	1

Table 2.5: Results of Bloc rule for the example of table 2.1

2.4.4 k-Borda Rule

This rule, proposed by Borda, proceeds as follows: each voter ranks all candidates, and each candidate is given a number of points equal to the number candidates ranked below them.³ The k candidates with the highest Borda score are declared winners.

Example. Let us consider once again the example of table 2.1. Voter v_1 gives four points to candidate b , three points to candidate a , two points to candidate d and finally one point to candidate d , etc. vote to candidate b and one to candidate a , voter v_2 gives four points to candidate e , three points to candidate a , two points to candidate b and finally one point to candidate d , etc. The final results are in the table 2.6 and the elected committee is $\{b, c\}$.

³ This is known as asymmetric Borda score. There are also symmetric Borda scores where the scores of the n alternatives are $\frac{n}{2}, \frac{n-1}{2}, \dots, 0, \dots, -\frac{n-1}{2}, -\frac{n}{2}$. Regardless of using asymmetric or symmetric Borda scores the outcome remains the same.

	Candidates				
	a	b	c	d	e
Total Votes	10	18	13	11	8

Table 2.6: Results of k -Borda rule for the example of table 2.1

2.4.5 Chamberlin-Courant rule (β -CC)

For each voter i and each item c there is a misrepresentation value $\mu_{i,c}$, representing the degree to which item c misrepresents voter i . A positional misrepresentation function makes use of a scoring vector $s = \langle s_1, \dots, s_p \rangle$ such that $s_1 \leq \dots \leq s_p$. In particular, the Borda scoring vector s_B is defined by $s_k = k$ for all k . By $pos_i(c)$ we mean the position of item c in i 's preference ranking (from 1 for the most preferred item to p for the least preferred one). The misrepresentation function induced by s is $\mu_{i,c} = s_{pos_i(c)}$. Intuitively, s_i is the amount of dissatisfaction that a voter derives from being represented by an alternative that she ranks in position i .

An assignment function π maps every voter to an item in the selected subset S . The misrepresentation of voter i under π is $\mu_{i,\pi(i)}$. Once individual misrepresentation has been defined, we need to define the utilitarianism (global misrepresentation is the sum of all individual misrepresentation) of the society when selecting a subset S of items. Formally, the global misrepresentation of assignment π is defined as:

$$\mu_U(\pi) = \sum_{i \leq \pi} \mu_{i,\pi(i)}$$

The Chamberlin-Courant scheme [13] simply outputs the committee of size k that minimizes μ_U . Because there is no constraint on the assignment function, every voter is assigned to her preferred item in the selected subset S . That is, $\pi(i) = \arg \min_{c \in S} \mu_{i,c}$. Then, her misrepresentation when selecting the feasible subset S is equal to $\mu_{i,S} = \min_{c \in S} \mu_{i,c}$. The best committee is then the feasible subset S minimizing $\mu_U(\pi)$ (under utilitarianism) [17].

A simplified definition of the rule could be the following: each voter is assigned to her preferred candidate in the selected committee. The selected committee consists of the candidates that minimize the voters' misrepresentation.

2.4.6 Monroe rule (β -Monroe)

The Monroe scheme [14] additionally from Chamberlin-Courant requires that the assignment π is balanced: each candidate in S must be assigned to at least n/k voters. Formally, the Monroe scheme selects the allocation π minimizing $\mu_U(\pi)$ subject to the constraints $|\pi^{-1}(s)| \geq \lfloor \frac{n}{k} \rfloor$ for all $s \in Range(\pi)$ [17].

A simplified definition for this rule could be the following: each voter is assigned to her preferred candidate in the selected committee and the assignment of the voters to their preferred candidate is balanced. The selected committee consists of the candidates that minimize the voters' misrepresentation.

Chapter 3

Design & Implementation of the Experiment over the Two-Dimensional Space

Trying to gain some better understanding of some known multiwinner voting rules, we perform a thorough experimental evaluation, at first, in elections that are generated within the two-dimensional space. A two-dimensional space experiment [9] not only help us visualize the outcome of the voting rules, but also, through the visualization, check if our intuition according to the rules agrees with our results.

3.1 Election

3.1.1 Experimental Setup

In these experiments, our objective is to elect a committee (a subset of candidates) that represents the voters in a certain way. As described in the previous chapter, in different applications a different mutiwinner voting rule should be used, as different applications require different properties of the outcome. The two-dimensional setup helps us visualize the outcome of the committee in an apparent and easy to comprehend fashion, resulting in the identification of each rule's possible weakness, and the properties each rule satisfies. With these in mind, we can compare the voting rules and determine which rule is better suited for each application, regarding the properties.

For every election's simulation, a set of 1000 voters and 1000 candidates is generated according to a certain distribution. Each voter and candidate is represented as points in the Euclidean Domain. For every voter, there is a preference order of the candidates. A committee of 20 members is elected for every rule. We inspect thoroughly the outcome of each rule, compare them and evaluate the rules in regard to their behavior.

The number of both the voters and candidates is deliberately the same and big enough for a 20 member committee to have a meaning through to the following reasons:

1. The number of candidates should be big enough in order to study the weaknesses of rules such as Single Nontransferable Vote rule.
2. The committee should have enough members to study properties such as proportional representation or diversity.
3. The voter's number should be big enough for rules such as Chamberlin-Courant and Monroe is meaningful to be used.

3.1.2 Distribution of Voters & Candidates

The voters, as well as the candidates, are generated as points in the Euclidean Domain according to some distribution. Thus, each one of them is represented by a pair of numbers (x, y) , the coordinates of the point, that define its exact location on a two-dimensional plane. In the experiment, we aim to understand the behavior of the multiwinner voting rules according to the combination of the distributions. Different distributions are used, e.g., uniform distribution, either as a disc or square, and Gaussian (normal) with different mean and variance, as well as for their combinations.

We studied various cases, including cases where the quota of voters (or candidates) that was generated according to each of the combined distributions were not equal. It is important to note that the distributions might not span the whole Euclidean domain, neither should voters and candidates span the same space. Especially in case of a combination of distributions the least possible overlap was sometimes required to get a clear image of the output. A detailed list of the distributions presented in this diploma thesis follows.

- **Gaussian**

Ideal points are generated using symmetric Gaussian distribution with mean $(0,0)$ and standard deviation 1.

- **Uniform Square**

Ideal points are distributed uniformly on the square $[-3,3] \times [-3,3]$.

- **Uniform Disc**

Ideal points are distributed uniformly on the disc with center $(0,0)$ and radius 3.

- **4-Gaussian**

Ideal points are generated using four symmetric Gaussian distributions with standard deviation 0.5, but with different mean values, namely, $(-1,0)$, $(1,0)$, $(0,-1)$ and $(0,1)$; each mean is used to generate 25% of the points.

- **2-Gaussian unequal**

Ideal points are generated using two symmetric Gaussian distributions. The means and the standard deviations differ; each mean is used to generate a different percentage of the points.

- **3-Gaussian unequal**

Ideal points are generated using three symmetric Gaussian distributions. The means and the standard deviations differ; each mean is used to generate a different percentage of the points.

3.1.3 Interpretation of the position at the two-dimensional plane

Each voter's ranking is generated according to the location of the certain voter and all the candidates on the two-dimensional plane. We consider that the distance of the voter and each candidate represents how the voter the misrepresentation of the voter for this candidate. Smaller the distance from a candidate, bigger the benefit for the voter. Thus, the ranking for a certain voter is generated from the ascending order of the distance from each candidate.

3.1.4 Assumptions

Some of the social choice functions that are used require randomness either in the initialization of some data or in the tie-breaking schema. As a result, it is vital to make some essential assumptions

to be able to completely predict the behavior of the algorithms from the input. It is also necessary to make some assumptions about the way we choose a candidate and fill the committee's positions, in a way that we can «measure» differences between similar inputs.

3.1.4.1 Random Generation & Tie-Breaking Scheme

There are four main approaches for dealing with ties [17]:

1. Use a fixed ordering of the alternatives (or a designed voter) to break all ties.
2. Use a randomized mechanism to break all ties.
3. Deal with set-value outcomes directly.
4. Ignore or suppress the issue (assume no ties exist).

Each of the above aforementioned approaches have a different problem. As a result we should consider the aspects of our experiment in order to choose which is one to use.

As mentioned previously, the locations of both voters and candidates are generated according to some distribution. The series in which each point is generated is random concerning the location of each point on the two-dimensional plane. This does not affect the space of the experiment, notwithstanding it is of high importance as a tie-breaking schema. All the algorithms require sorting of the candidates according to their votes. It becomes clear that a way of predetermining the tie-breaking method is needed.

As the generation of points happens randomly, we can presume that the correlation between each point and its place in the array of the points is also random. We keep locked the series of each point's generation for each simulation independently of the voting rule. This correlation determines the way all the ties are to be handled. In case of a tie, the first generated point will have priority. This way the tie-breaking schema is predetermined, but with the required amount of randomness.

3.1.4.2 Random Initialization

Some of the multiwinner voting rules need a random initialization of the committee. To be able to have comparable outcomes, we must keep locked the voters' initial positions in each simulation independently of the voting rule, as well as the initial random chosen committees. The made initial choice can affect the outcome. After experiments, we decided to use different initialization techniques for every rule, in order to achieve better results. Hence, the initializations were made only one time per voters distribution, regardless of the candidates' distributions.

3.1.4.3 Determinism

The «lock» of each point to a certain place in the array gives us another great advantage. Running the same algorithm as many times with same input, we always produce the same output. Subtracting the randomness in as many aspects as possible, the results may be characterized as deterministic, due to the fact that we always have the same sequence of states.

3.1.4.4 Ties in the Output (Last Place's Ties)

Another necessary assumption regards the permission of ties in the committee's last place. If we choose to ignore or handle them, the outcome will consist of exactly k winners. On the other hand, if we allow them, more than k winners may come up. This choice is of high importance, especially for the study of the stability property, where we must be able to detect real from apparent changes in the outcome.

For example, let assume there is a tie between two candidates in the last position. We could either choose to elect both of them or either one of them. Should we consider the results different? In words of representativity may seem irrelevant but in words of stability, it is not.

We choose the latter choice, that of not allowing any ties in the outcome. The measurement of how stable an outcome is affected by our assumptions. Therefore, we must be very careful with our conclusions, as a bias exist and a single voter manipulability is an inappropriate test[17] if we do not interpret the results taking into consideration this bias. This will be analyzed vastly in chapter 5.

3.2 Results, Comments & Comparison of the Rules

3.2.1 Single Nontransferable Vote Rule

The Single Nontransferable Vote rule is widely used in every aspect of our lives. It is a well-known algorithm for parliamentary elections and other applications where representativity is needed. As described in the previous chapter the Single Nontransferable Vote rule takes into consideration only the first preference of the voters. But how does this affect the results? Can we find any special patterns in the outcomes? And if yes, how can this affect the suitability of this rule over different applications?

In the adjacent example (figure 3.1), where the voters are generated from a Gaussian distribution and the candidates from a Uniform Disc distribution looks like that this rule can achieve representativity.

However, after simulations with different combinations of distributions for voters and candidates, we point out that this property is not guaranteed due to the very nature of the algorithm. More specifically, if we generate voters from a uniform distribution, while the candidates are generated from a Gaussian, we observe that the results do not achieve proportional representation.

In the above examples, the elected committee tends to outline the candidates' distribution. The voters are generated from uniform distributions and, therefore, are uniformly distributed over the two-dimensional space. On the other hand, the candidates are generated from Gaussian distributions and, hence, there are parts of the two-dimensional space with higher density and others with lower.

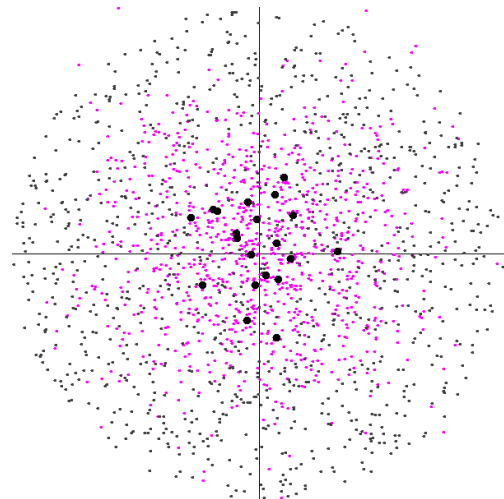


Figure 3.1 SNTV rule's outcome

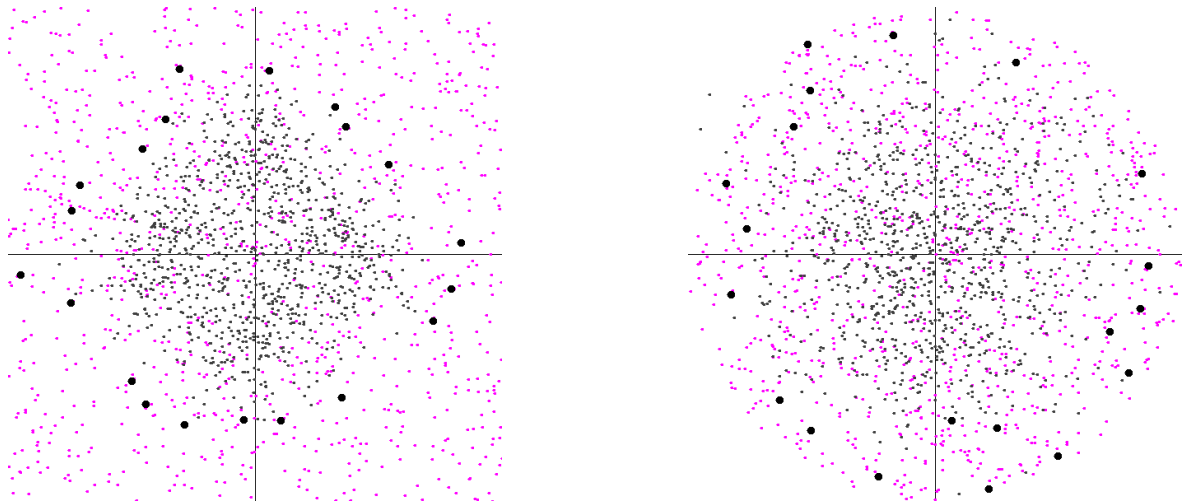


Figure 3.2: SNTV rule outlines the candidates' distribution

It is easy to understand that in the places with high density of candidates, the votes are distributed among them. However, in the places with a low density of candidates, e.g. the outlines of the normal distributions, the votes are distributed among fewer candidates. Given the fact that the voters' distributions span more widely in comparison to the candidates' distribution, there are more voters that prefer the distribution's outline candidates. These two observations can explain this phenomenon, with which we can come across in different cases as well, even in the single-winner elections.

A real-life example can be derived from elections, where the people are divided between two "similar" candidates and as a result none of them is elected, even though they have more votes of the winner cumulative.

Another interesting example is the following. In both cases the voters' distribution is the same and the candidates' distributions are uniform square and uniform disc, respectively. In the first case, all the elected candidates locate in the dense normal distribution in the upper left corner, as expected, and the sparse distribution of the voters seems to be unsatisfied. In the second case, we notice that the sparse distribution is represented, which should happen if we wanted proportional representation. The astonishing observation is that some of the remotest voters are represented, even though the elected candidate can represent only one voter.

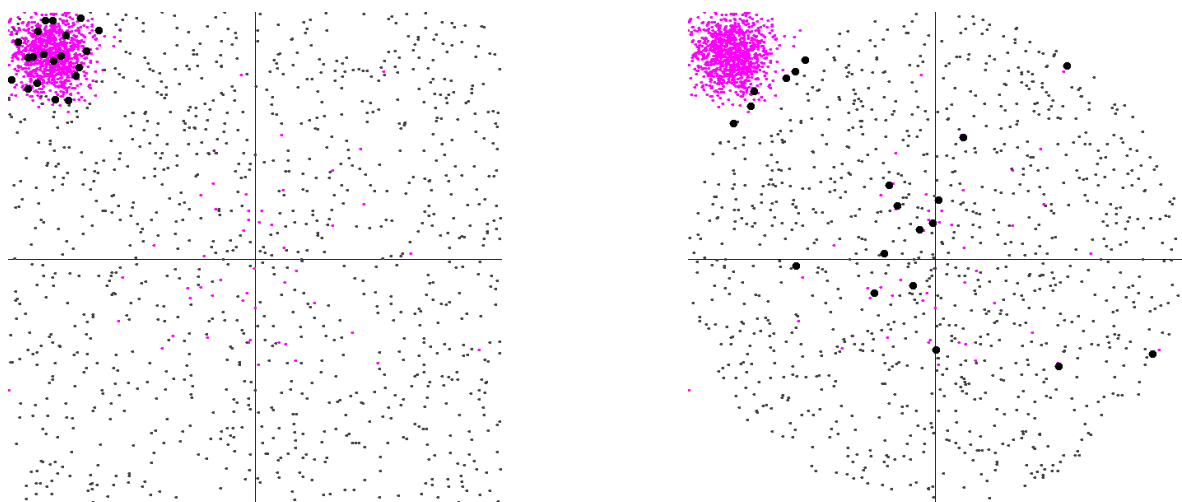


Figure 3.3: SNTV rule's outcome vastly depends on the candidates distribution

This example can point out the importance of the candidates' set. When only a few candidates are suitable for the majority of voters, candidates who represent the slightest amount of voters can be elected.

3.2.2 Single Transferable Vote Rule

The Single Transferable Vote rule has voters rank individual candidates in order of preference. During the count, as candidates are elected or eliminated, surplus or discarded votes that would otherwise be wasted are transferred to other candidates according to the preferences. This system provides approximately proportional representation and enables voters to vote across party lines and to elect independent candidates.

The Single Transferable Vote rule is sensitive to the series of candidates' election and elimination. Even the slightest change to this series affects the outcome. Therefore, even the slightest variation of the rule, even the slightest implementation decision of the algorithm, can affect and as a result alter the outcome.

For example, let assume that we chose to elect only the candidate with the highest first-place votes who received at least the quota, transfer the excess votes and then count the first place votes for each candidate in order to elect the next candidate, instead of electing all the candidates who received the quota at once, the results differ. Another important choice is the identification of the candidate with the fewest first-place votes. In case of a tie, which candidate do we remove first? This choice could be made randomly, but this could result in different outcomes for the exact same input. In chapter 5 we examine the stability of each rule. In order to do so for Single Transferable Vote rule, we have to remove the randomness from it. As a result, we always eliminate the last candidate, according to the sorting, where the ties are broken according to a predetermined way, considering the starting positions.

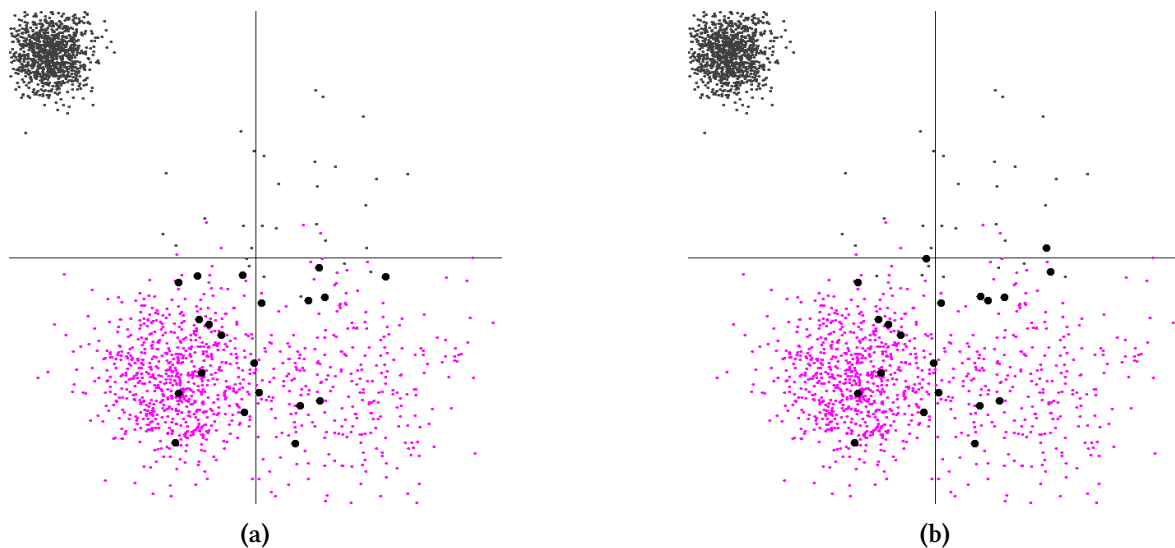


Figure 3.4: Different outcomes for slightly different versions of STV rule

Single Transferable Vote rule is a method of calculating election results that is said to guarantee proportional representation for solid coalitions - sets of voters who share a set of most preferred candidates - under reasonable conditions¹ [16]. This can be explained since Single Transferable Vote

¹ We will examine more thoroughly this assumption in section 3.2.7

minimizes wasted votes, transferring surplus votes to the hopeful candidates.



Figure 3.5: STV rule's outcomes indicate that STV rule guarantees proportional representation

However is a highly problematic rule, due to its dependency on the series that the candidates are elected or eliminated.

3.2.3 Bloc Rule

In Bloc rule, voters are selecting their k favorite candidates. From the outcome we can observe that the elected candidates form little groups, which tend to locate in areas where the candidates' density is low, for example near the edges of candidates' distribution (with the precondition that candidates' distribution spans the area that the voters' distribution spans).

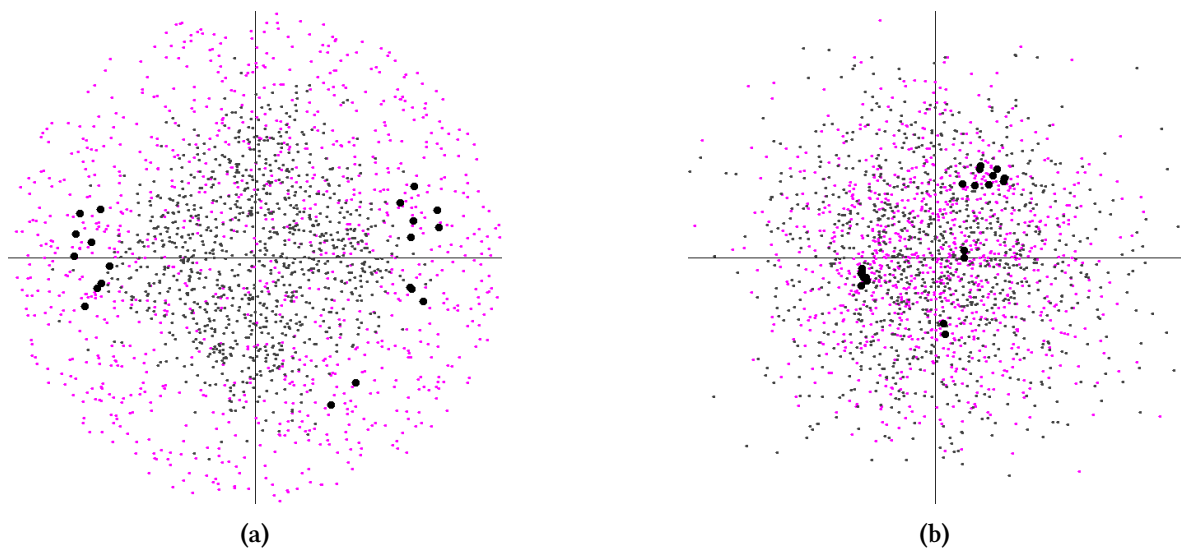


Figure 3.6: Bloc rules' outcome consist near the edges of small groups

As we can see in the above outcomes the elected committee does not seem to represent the majority of the voters. This is a consequence of the nature of the rule. In places where the density of the candidates is high, the votes tend to split among similar candidates and as a result, the candidates are not elected.

The only way for more voters to be represented is to have a sparse distribution of candidates in the locations where voters distribution is dense, as happens in the adjacent example.

But even then the results are not able to be diverse enough and fail to represent the majority of the voters, in the preponderance of the cases.

Hence, this rule is not suitable for application where diversity or representativity is needed, On the other hand, the fairness of the rule cannot be doubted, because all the similar candidates will be either elected or rejected.

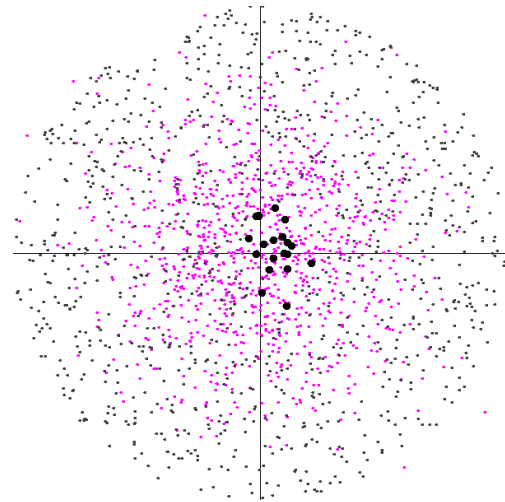


Figure 3.7 Bloc rule's outcome depends on the density of the candidates' distribution

3.2.4 k -Borda Rule

k -Borda rule determines the outcome of the election by giving each candidate, for each ballot, a number of points corresponding to the number of candidates ranked lower. Once all votes have been counted the k candidates with the most points are the winners. The single winner Borda rule is described as consensus based voting system and is used when proportional representation is needed. to find the k most acceptable candidates over the set.

The suitability for this rule for shortlisting is justified by a pattern in the outcomes: the elected committee tends to be located in the center of mass of the voters' set, as long as the candidates' set allows it.

This pattern also justifies this rule's unsuitability for elections: even though the results may be broadly more acceptable, all the elected candidates are located in the same place. As a result, all the candidates are similar and there is not diversity nor proportional representation of the voters' set.

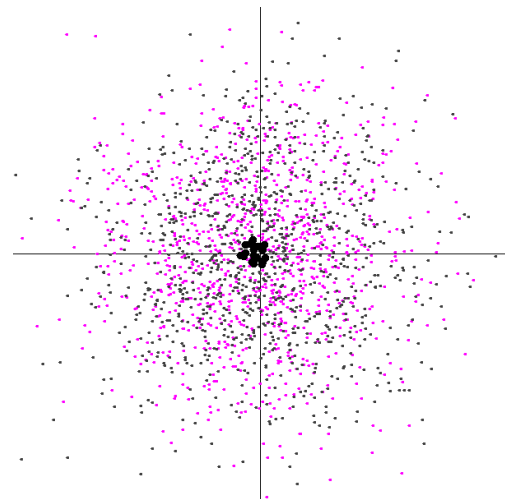


Figure 3.8 k -Borda rule's outcome

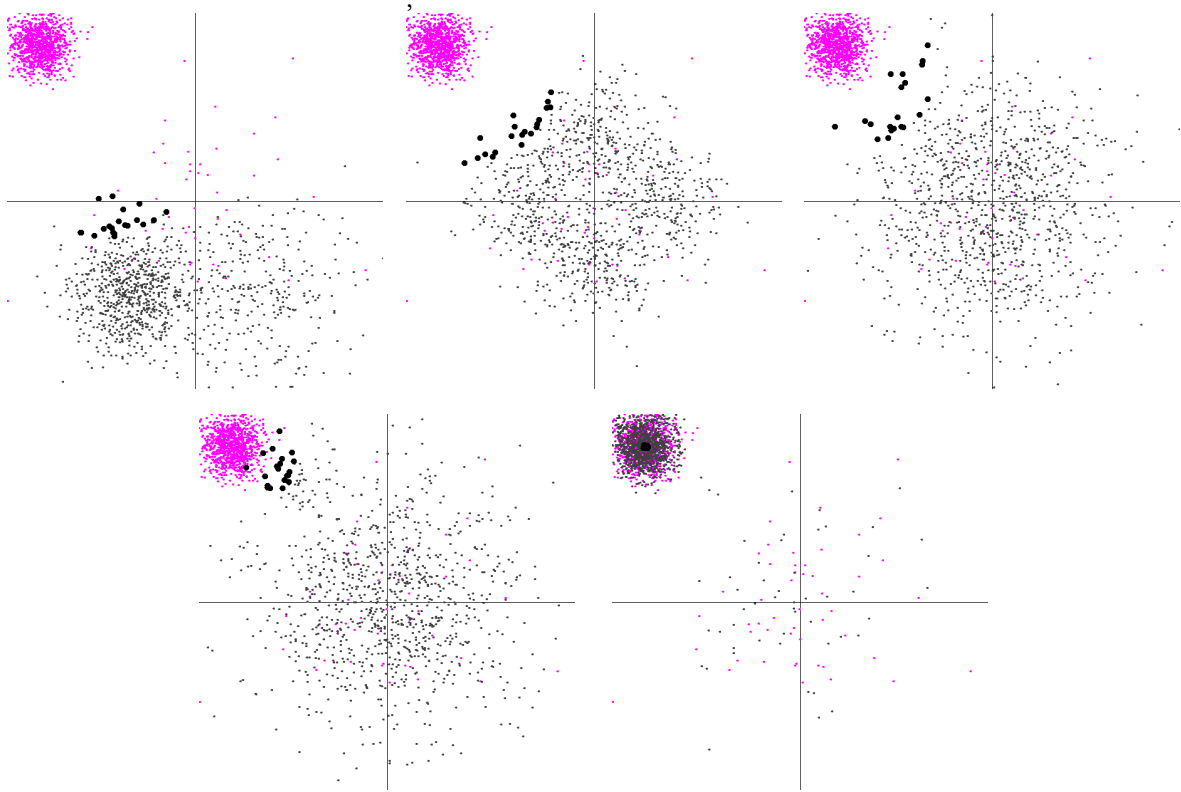


Figure 3.9: k -Borda rule chooses candidates located to the voters' center of mass

3.2.5 Chamberlin-Courant Rule ²

Chamberlin-Courant rule pursues to elect the candidates that minimize the voters' misrepresentation. Thus, candidates are located in the center of each cluster. Due to the fact that there are no restrictions on the population of each cluster, the committee members are spanned in such a way that covers the maximum area in the voters' distribution. On the other hand, due to the same fact, we cannot determine the type of the voters' distribution as the outcome reminds a uniform distribution, as long as the candidates' distribution allows it.

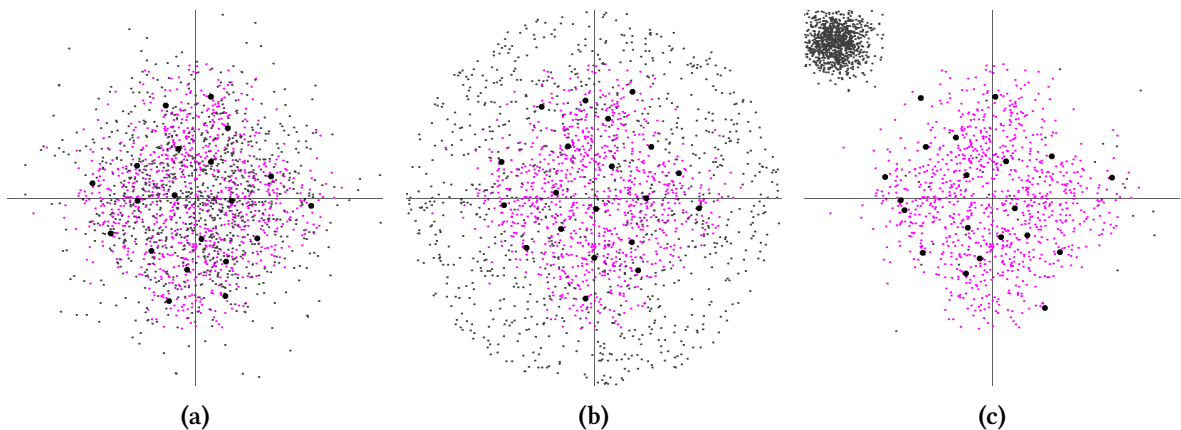


Figure 3.10: Chamberlin-Courant rule's outcomes

Voters follow a 4-Gaussian distribution while the candidates follow (a) Gaussian, (b) Uniform Disc, and (c) 2-Gaussian Unequal distributions

² The results presented in this section correspond to Chamberlin-Courant heuristic, which we analyze in section 4.1 Nonetheless, the observations represent the original rule as well, unless it is noted differently.

In our heuristic, when the candidates' distribution does not span enough to cover the whole voters' distribution, the elected candidates are the ones closer to the centers of the clusters. The NP-hard Chamberlin-Courant algorithm would choose candidates distributed more evenly.

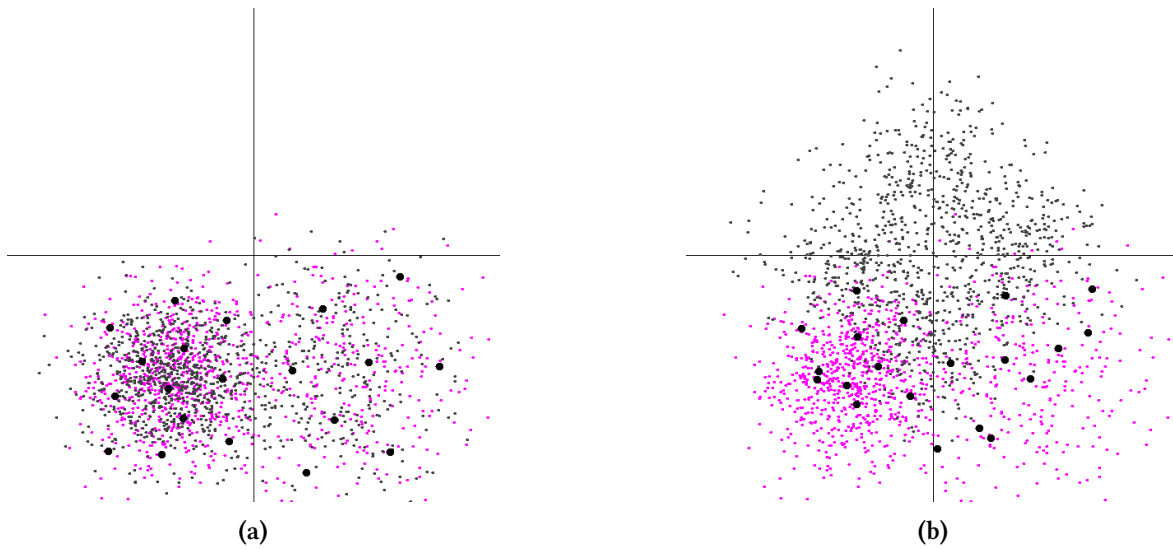


Figure 3.11: Chamberlin-Courant rule's outcome, when candidates' do not span enough
 Voters follow a 2-Gaussian Unequal distribution, while candidates follow (a) 2-Gaussian Unequal,
 and (b) 4-Gaussian distribution

3.2.6 Monroe Rule ³

Monroe rule pursues to elect the candidates that minimize the voters' misrepresentation. Thus, candidates are located in the center of each one of the equally sized clusters. Due to the fact of the restrictions over the population of each cluster, in contrast with the Chamberlin-Courant rule, we are able to define the type of the voters' distribution, as Monroe rule is underlying it.

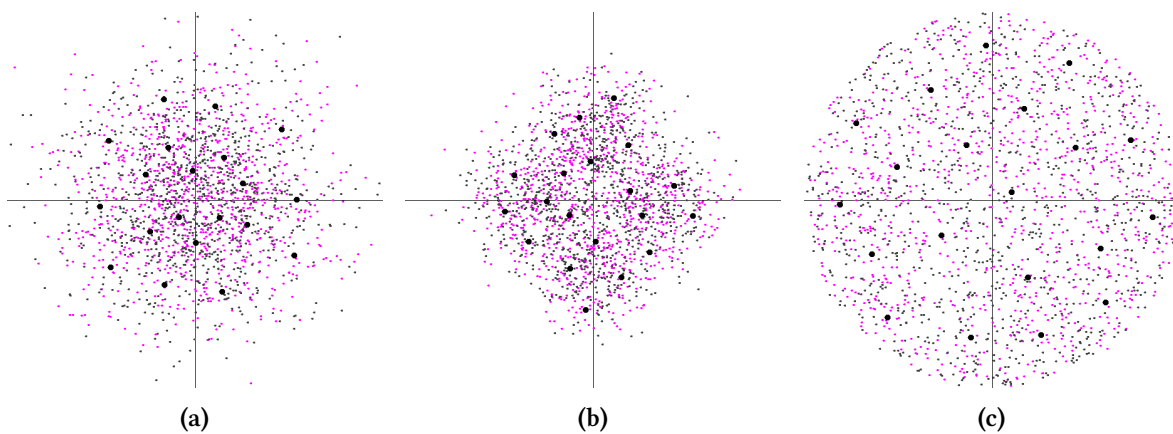


Figure 3.12: Monroe rule's outcomes underlines voters' distribution
 Voters and candidates follow the same distribution: (a) Gaussian, (b) 4-Gaussian, (c) Uniform Disc

However, this can only happen under two conditions: 1. there are enough seats in the committee, and 2. the distributions have enough members to be represented, as in this rule the minorities cannot be represented if they have not the necessary population in order to form their own cluster.

³ The results presented in this section correspond to Monroe heuristic, which we analyze in section 4.2.1 Nonetheless, the observations represent the original rule as well, unless noted differently.

3.2.7 Comparisons of Single Transferable Vote, Chamberlin-Courant & Monroe rules

After the rules' description and some basic comments, we find it appropriate to compare thoroughly the Single Transferable Vote, the Chamberlin-Courant and the Monroe rule, in order to determine when they tend to coincide.

Monroe algorithm resembles Chamberlin-Courant rule, with only difference that in Monroe algorithm in every committee member, there are assigned the same number of voters. As a result, in cases where voters follow a Uniform distribution the results of both Chamberlin-Courant and the Monroe rule are similar and we are not able to distinguish which outcome correspond to each rule.

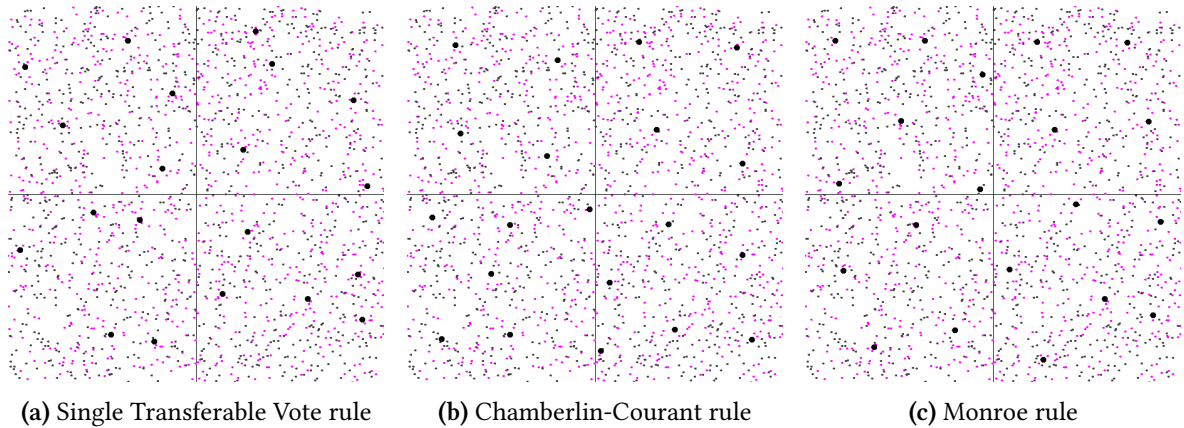


Figure 3.13: Comparisons for uniform distribution

In the above images, we can see that both Chamberlin-Courant and Monroe rules' outcomes, the elected candidates seem to be evenly distributed. However, the Single Transferable Vote rule elected candidates are elected with no regular arrangement across space, due to the elimination series.

On the other hand, when the voters follow a Gaussian distribution, we notice the difference between Chamberlin-Courant and Monroe rule. Monroe rule's outcome underlines the voter's distribution, while Chamberlin-Courant rule's outcome tends to span the whole voter's space in regular arrangement across space.

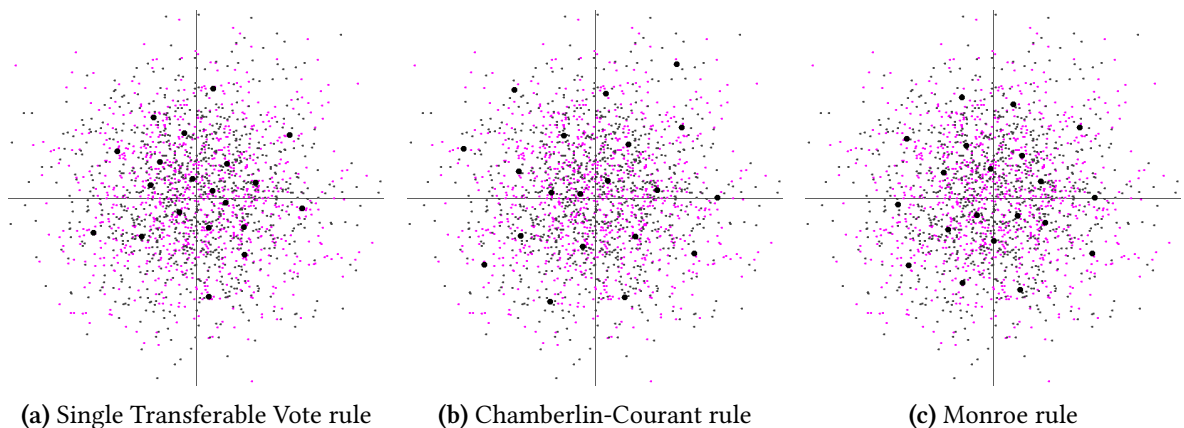


Figure 3.14: Comparisons for normal distribution

Single Transferable Vote rule is said to guarantee proportional representation and in the above example, it is once again proved, even if the candidates are distributed, in a more random way, than in the Monroe rule. Studying simple test cases we conclude that Single Transferable Vote rule's outcomes resemble the Monroe rule's outcomes and this confirms the characterization a rule good for election. But what happens for test cases, where voters consist of two unequal distributions and one distribution is vastly smaller than the other, but big enough to have its own representative in Monroe rule.

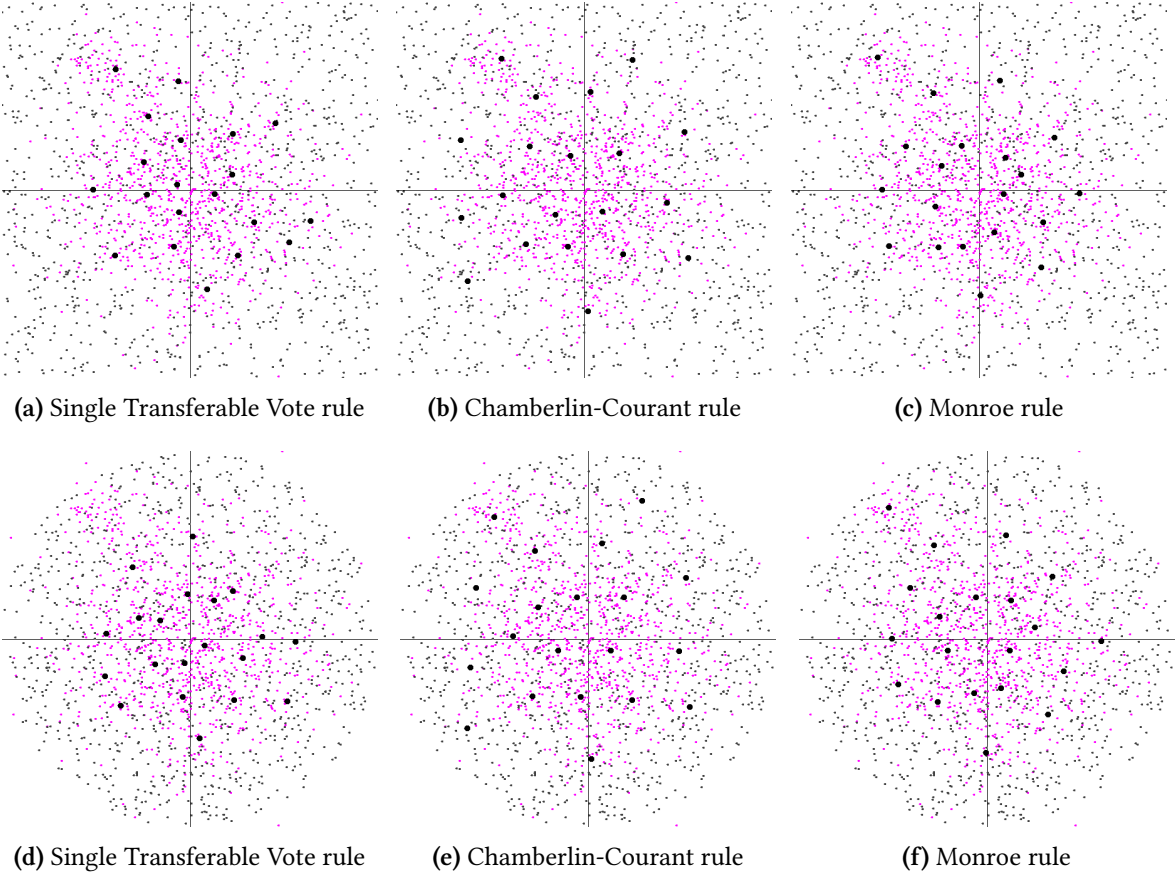


Figure 3.15: Comparisons for unequal normal distributions

For all images the voters are the same. The candidates for images (a), (b) and (c) are the same and follow a uniform square distribution, while for images (d), (e) and (f) a uniform disc.

In the above images, the voters are generated by 95% from a Gaussian distribution with mean (0,0) and standard deviation 1 and by 5% from a Gaussian distribution with mean (-1.2,1.8) and standard deviation 0.25. From this example, it became apparent, that in cases where voters are generated from more than one distributions, even when the density of the distributions is the same, the Single Transferable Vote fails to guarantee a representative for the small distribution, even though there should be one.

Studying a different example, where voters are generated by 95% from a Gaussian distribution with mean (-2.5,2.5) and standard deviation 0.25 and by 5% from a Gaussian distribution with mean (0,0) and standard deviation 1, we also had some interesting observations.

While the candidates' distribution spans the area that the voters' distribution spans, we can see that the Single Transferable Vote may assign one candidate (figure 3.16 (a)) or more (figure 3.16 (d)). It is important that proportionally, only one candidate should be assigned in the small distribution. The ideal representative would be the closest one to the mean of the small distribution. However, in figure 3.16 (a) due to the elimination series of the candidates an extreme candidate is elected. Additionally, in figure 3.16 (d) two candidates near the ideal location of the candidate are elected.

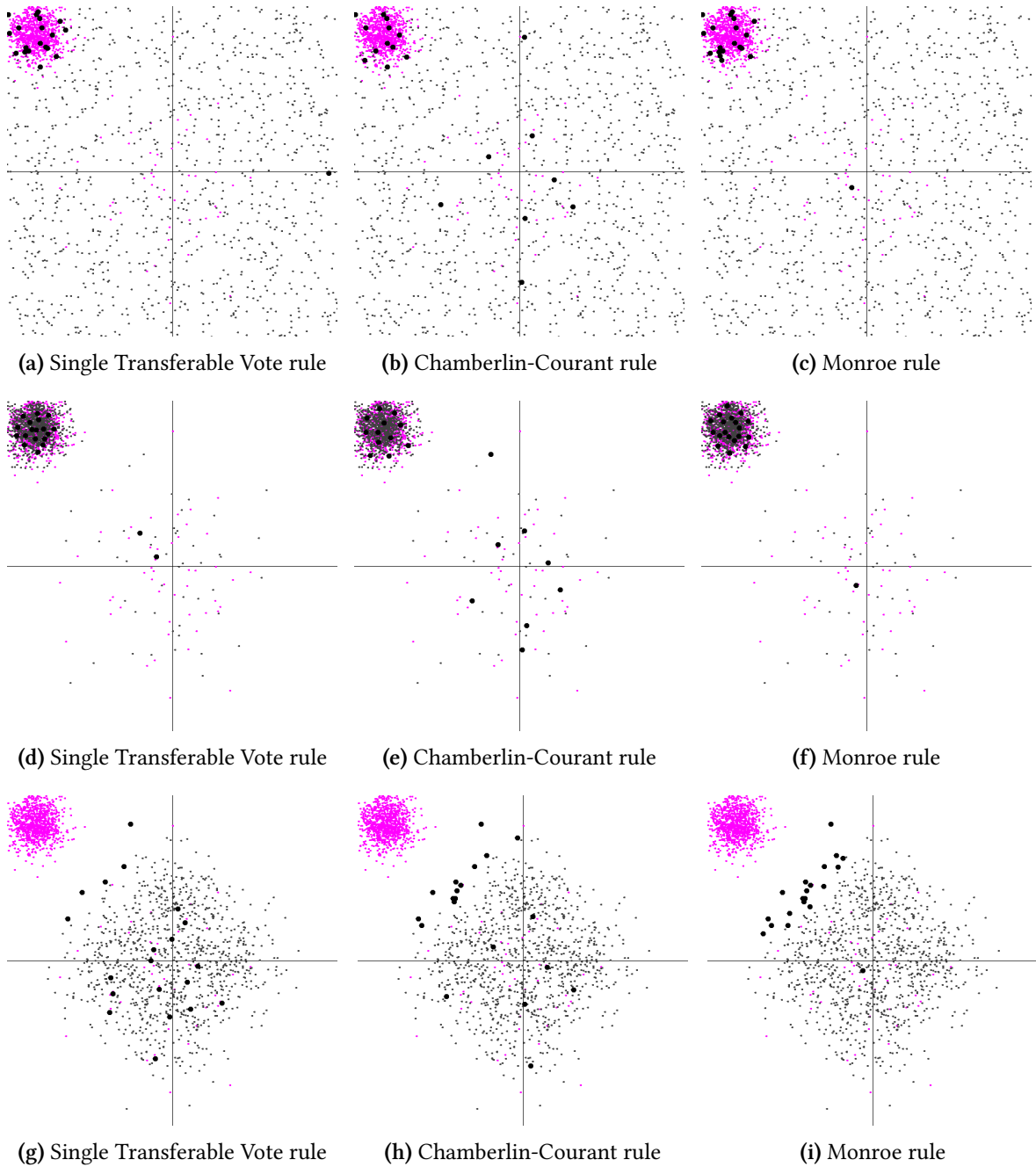


Figure 3.16: Comparisons for unequal normal distributions

For all images the voters are the same. The candidates for images (a), (b) and (c) are the same and follow a uniform square distribution, for images (d), (e) and (f) follow voters' distribution, while for images (g), (h) and (j) follow a 4-Gaussian distribution.

On the other hand, when the candidates' distribution is not spanned enough to cover the area that the voters' distribution span, the outcome of the Single Transferable Vote resembles more the outcomes of Chamberlin-Courant rule than this of Monroe. The majority of the elected candidates are representing a 5% of the voters, which belong to a sparse distribution. This outcome could be even closer to the ideal outcome of the Chamberlin-Courant rule than the outcome of our Chamberlin-Courant rule's heuristic.

Therefore, after our experiments, we have confirmed that Single Transferable Vote rule cannot guarantee proportional representations when minority groups exist. The outcome may benefit the minorities at the expense of the vast majority or even promote some extreme candidates, due to the series of elimination.

3.2.8 Comparisons over Properties of Rules

When making a selection by a voting rule, a desirable requirement is that of proportional representation. Proportional representation stipulates that voters should get representation in a committee or parliament according to the strengths of their numbers. It is widely accepted that proportional representation is the fairest way to reflect the diversity of opinions among the voters. [18]. The evaluation of the rules with respect to proportional representation follows summarized in the following table.

	Proportional Representation		
	Best Case	Average Case	Worst Case
SNTV	✓	✗	✗
STV	✓	✓	✗
Bloc	✗	✗	✗
k -Borda	✓	✗	✗
CC (h)	✓	♣	✗
Monroe (h)	✓	✓	♣

Table 3.1: Proportional representation

✓/ ✗ indicates that the rule achieves/does not achieve proportional representation. ♣ indicates that the rule may or may not achieve proportional representation according to the initialization of the heuristic.

We also evaluated the rules with respect to the diversity of the elected candidates.

	Diversity		
	Best Case	Average Case	Worst Case
SNTV	✓	✓	✗
STV	✓	✓	✗
Bloc	✗	✗	✗
k -Borda	✗	✗	✗
CC (h)	✓	✓	♣
Monroe (h)	✓	♣	✗

Table 3.2: Diversity

✓/ ✗ indicates that the rule achieves/does not achieve diversity. ♣ indicates that the rule may or may not achieve diversity according to the initialization of the heuristic.

Chapter 4

Heuristics for NP-hard Rules

Some of the multiwinner voting rules that we chose to study are NP-hard to compute and as a result, the election of the committee cannot happen in polynomial time. Thus, for the algorithms of Chamberlin-Courant and Monroe, we made our own heuristics and tried to decide whether the outcomes ¹ produced by our heuristics were good enough, according to the expected results.

4.1 Chamberlin-Courant Rule

4.1.1 Chamberlin-Courant Heuristic

In Chamberlin-Courant rule, every voter is assigned to her preferred candidate in the selected committee. The selected committee consists of the candidates that minimize the voters' misrepresentation. For our experiment the misrepresentation is reflected by the distance between the voter and his representative.

Chamberlin-Courant multiwinner voting rule may be received as a cluster analysis problem, similar to k -means. Consequently, the k -means algorithm, which is used to partition the input data set into k partitions (clusters), can be a simple and fast algorithm for this problem, although it offers no approximation guarantees at all [19].

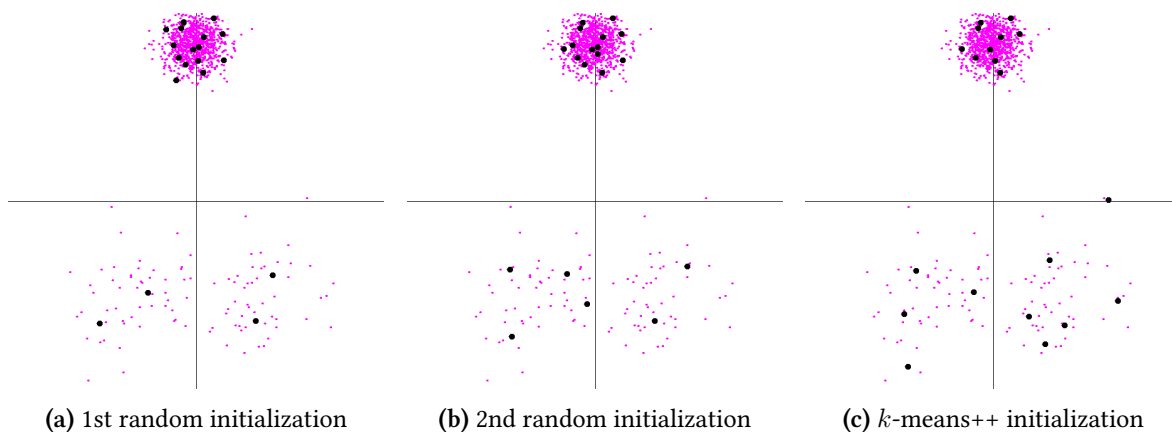


Figure 4.1: Non-equivalent results for Chamberlin-Courant heuristic with different initializations.

¹ We note that in this chapter's images we chose not to visualize the set of candidates in order to have a more clear outcome. The candidates were generated from a Uniform Square distribution. The purple dots represent the voters and the bigger black dots the elected candidates, which are located nearest to the points that our algorithm proposes as winners. As a result, if the distribution of candidates differed, the visualized outcome would differ as well.

We used the following algorithm (algorithm 1), which is an adjustment of k-means algorithm for our problem.

Algorithm 1 Chamberlin-Courant rule's heuristic algorithm

1. Arbitrarily choose an initial k centers $C = \{c_1, c_2, \dots, c_k\}$.
2. For each $i \in \{1, \dots, k\}$, set the cluster C_i to be the set of points in X that are closer to c_i than they are to c_j for all $j \neq i$.
3. For each $i \in \{1, \dots, k\}$, set c_i to be the center of mass of all points in C_i : $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.
4. Repeat Steps 2 and 3 until C no longer changes.

With the following addition²:

5. For each c_i find the closer candidates and declare them winners.
-

It is standard practice to choose the initial centers uniformly at random from X . But the k -means algorithm's results depend on the initialization points and are different. In most cases the outcomes can be perceived as equivalent, measuring the misrepresentation of the voters. Nevertheless, in test cases with no proportional participation of each distribution in the input data or distributions with different density, the outcomes from different initialization points were far from equivalent (figures 4.1 (a) and (b)).

4.1.2 Initialization Technique

Wanting to study this kind of cases, we created specially designed tests in order to see if the results will be improved by choosing the initialized centers in a specific way and not uniformly in random. In particular, let $D(x)$ denote the shortest distance from a data point (voters) to the closest center we have already chosen. The algorithm used was the k -means++ algorithm[19].

Algorithm 2 Chamberlin-Courant rule's heuristic algorithm (k -means++ initialization technique)

- 1a. Take one center c_1 chosen uniformly at random from X .
 - 1b. Take a new center c_i choosing $x \in X$ with probability $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$.
 - 1c. Repeat Step 1b until we have taken k centers altogether.
 - 2-5. Proceed as with the algorithm 1.
-

The results in the symmetrical test-cases were as expected equivalent again, but in our specially designed cases were conspicuously improved (figure 4.1 (c)). The figures 4.2 (a), (b) and (c) reveal the importance of the initialization centers as well.

² In the Chamberlin-Courant rule, we want to find the candidates that minimize the voters' misrepresentation. On the other hand, the k -means is used to partition the data into k clusters. From the k -means or k -means++ algorithm we can find the ideal candidates to be elected. As a result, we must add an extra step in order to make the transition from the ideal to the existing candidates.

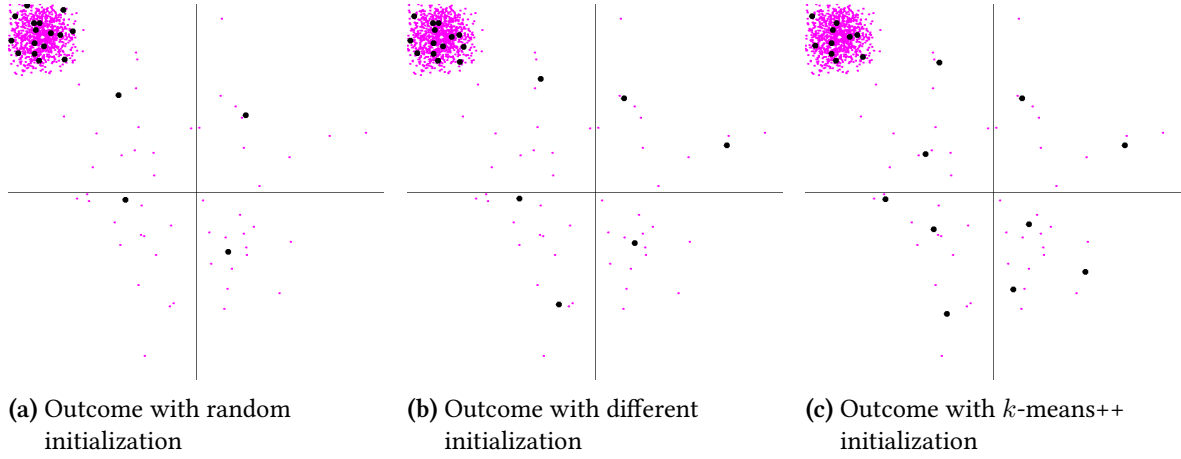


Figure 4.2: Results for Chamberlin-Courant heuristic before and after k -means++ initialization.

4.1.3 Proposed Algorithm

Let $D(x)$ denote the shortest distance from a data point (voters) to the closest center we have already chosen. The algorithm 3 is the proposed heuristic for Chamberlin-Courant rule.

Algorithm 3 Chamberlin-Courant rule's proposed heuristic algorithm

- 1a. Take one center c_1 chosen uniformly at random from X .
 - 1b. Take a new center c_i choosing $x \in X$ with probability $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$.
 - 1c. Repeat Step 1b until we have taken k centers altogether.
 2. For each $i \in \{1, \dots, k\}$, set the cluster C_i to be the set of points in X that are closer to c_i than they are to c_j for all $j \neq i$.
 3. For each $i \in \{1, \dots, k\}$, set c_i to be the center of mass of all points in C_i : $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.
 4. Repeat Steps 2 and 3 until C no longer changes.
 5. For each c_i find the closer candidates and declare them winners.
-

4.1.4 Proposed Algorithm's Weakness

The proposed algorithm finds the ideal clusters regardless of the candidates' distributions and then finds the closest candidates in the proposed cluster centers. If the candidates span the space that voters span, this produces outcomes close to the ideal. But when this is not happening, the outcome can differ from the ideal³. Instead of distributing the candidates in the space that the two distributions overlap, many elected candidates are located in the same place, trying to reach the ideal points.

³ This phenomenon was also remarked in the section 3.2, both in Chamberlin-Courant rule's and Comparisons subsections.

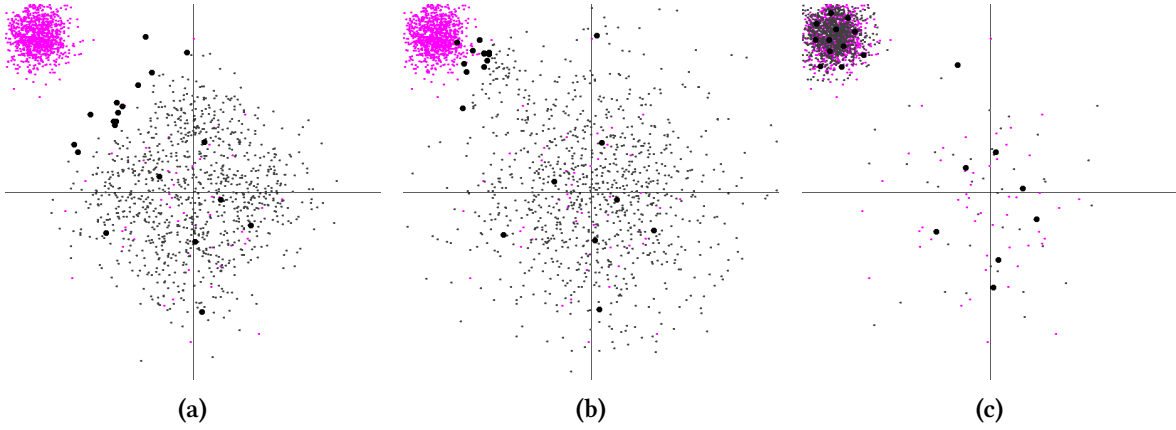


Figure 4.3: The weakness of the proposed heuristic for Chamberlin-Courant rule

4.1.5 Computational Complexity

Since this rule is NP-hard and we use a heuristic based on a clustering algorithm there are no approximation guarantees, since the number of repetitions until the algorithm converges varies.

Notwithstanding, we can compute the computational complexity for every step of the algorithm except the number of repetitions of steps 2 - 4 that the algorithm will need to converge. Let k denote the number of the committee Members, v denote the number of voters and c denote the number of the candidates. In our experiment $k \ll v$ and $v \approx c$.

For steps 1a - 1c (initialization) we have:

$$\mathcal{O}(k \cdot v) = \mathcal{O}(v)$$

For steps 2 - 4 we have:

$$\mathcal{O}(v + k \cdot v + v + k + k^2) = \mathcal{O}((2 + k) \cdot v + k + k^2) = \mathcal{O}(v)$$

Finally for step 5 (elect committee of candidates) we have:

$$\mathcal{O}(k \cdot c) = \mathcal{O}(k \cdot v) = \mathcal{O}(v)$$

4.2 Monroe Rule

4.2.1 Monroe Heuristic Algorithm

Monroe rule, as described in the section 2.4, apart from the Chamberlin-Courant rule requirements, additionally requires that the assignment of each voter to his preferred candidate is balanced.

Due to the similarities of these two algorithms, we decided to base our heuristic for Monroe in the k -means algorithm as well. In order to correct the analogies we tried a k -means based heuristic with a simple twist: we force exactly $\frac{n}{k}$ members in each cluster and no voter is allowed to exist in more than one clusters.

In the elected committee we used a queue of $\{voter, committee\ position, distance\}$ tuples, sorted by distance in order to match the voters to their representatives. Every voter was matched with the representative (center of the cluster) that was nearest to, with the condition of not being already a member of a cluster. The algorithm used follows (algorithm 4).

Algorithm 4 Monroe rule's heuristic algorithm

1. Arbitrarily choose an initial k centers $C = \{c_1, c_2, \dots, c_k\}$.
 2. For each $i \in \{1, \dots, k\}$, set the cluster C_i to be the set of $\frac{n}{k}$ points in X that are closer to c_i than they are to c_j for all $j \neq i$.
 3. For each $i \in \{1, \dots, k\}$, set c_i to be the center of mass of all points in C_i : $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.
 4. Repeat Steps 2 and 3 until C no longer changes.
 5. For each center c_i find the closer candidate and declare them winners.
-

The initial centers are chosen once more uniformly at random from X and the ties are broken arbitrarily but in a consistent way. There are three important questions regarding the proposed algorithm:

1. Is there an optimized initialization technique?
2. What is the most effective way to assign each voter in a cluster?
3. Does the algorithm converge in every single case?

4.2.1.1 Convergence of the proposed algorithm

As far as the convergence of the algorithm, it is not guaranteed. We noticed that the only way of the algorithms non-convergence is to loop infinitely, i.e. a sequence of points chosen for a committee which loops infinitely. This problem derived from the discrete space of our handling of the problem, only $\frac{n}{k}$ voters could be assigned to each cluster and in each loop, one or more voters were changing clusters only to end up in the same cluster started.

There were two options to handle this problem, either we could choose the looping state with the smaller misrepresentation or find a real converge point. We chose to follow the second direction and solve the problem by transferring it to continuous space. Finding in a greedy way the center

of mass of the matching cluster centers of the looped committees and starting the algorithms with these initial centers we could see that the algorithm had already come to converge.

4.2.1.2 Assignment of voters in a cluster

We tried two different approaches that resemble the breadth-first and depth-first search algorithms. Afterwards, we present the two approaches and analyze the advantages and disadvantages of each.

In both approaches, for each cluster center c_i there is a sorted list of the voters in ascending order according to their distance from the c_i . What differs is the synchronization mechanism for enforcing limits of how many voters can a cluster gain at once. The cluster centers can be described as being in a queue, waiting to acquire a lock in order to gain voters.

- **Breadth-first approach**

In this approach, once a cluster center acquires the lock it chooses the nearest voter that has not been assigned already in a cluster, releases the lock and goes back in the end of the queue waiting to get the lock again.

- **Depth-first approach**

In this approach, once a cluster center acquires the lock it chooses the nearest $\frac{n}{k}$ voters that have not been assigned already in a cluster and then releases the lock.

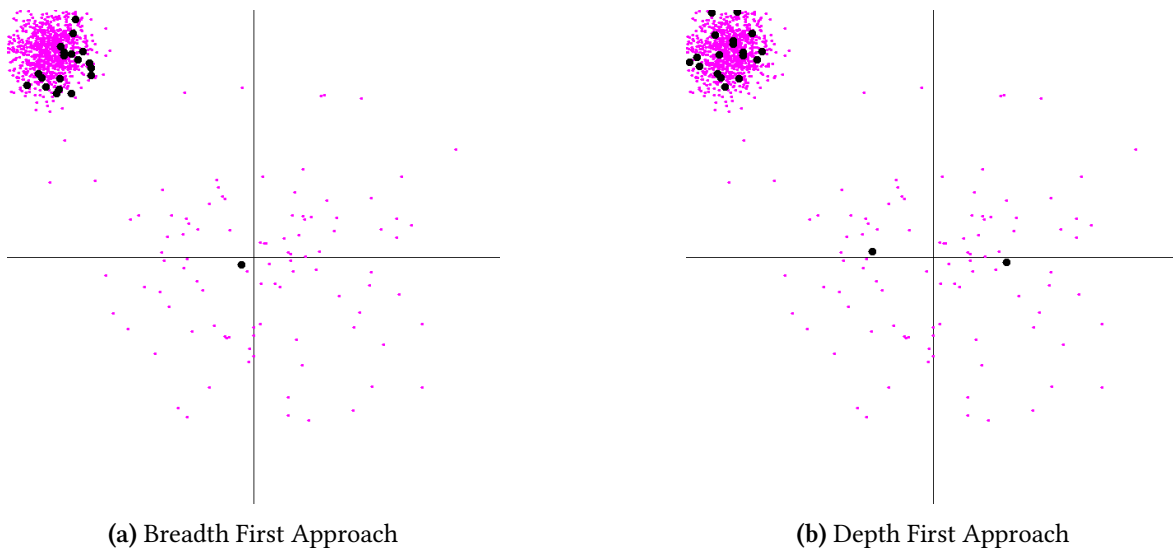


Figure 4.4: Assignment of candidates with different approaches for Monroe heuristic for the same voters, candidates and initialization points (chosen uniformly at random)

10% of the voters is generated from a sparse Symmetric Gaussian Distribution with mean (0,0) and standard deviation 1 and 90% from a Symmetric Gaussian Distribution of a symmetric Gaussian distribution with mean (-2.5,2,5) and standard deviation 0.25

The breadth-first approach works well in most of the inputs. But it has a weak spot for inputs that are a combination of distributions and a small part of the input comes from a sparse distribution with big standard deviation. On the other hand, the depth-first approach, statistically speaking, has better results for inputs as the one described above, given the same initialization points, but has the problem of overfitting. In simple and symmetrical inputs it works poorly.

As we will describe later, the results of the two approaches depend on the initial centers. As a result, we can improve the chance of success of the breadth-first approach if we find the proper initialization points and this is why we propose this approach over the depth-first.

4.2.1.3 Initial k centers

As in the Chamberlin-Courant heuristic, different starting positions produce different outcomes. It is important to identify once again the cases where the results are not equivalent. We noticed that similar test cases used for the Chamberlin-Courant rule can be used to study the importance of the initialization centers. In most of the cases, even in more complex ones without symmetry, the results for different initialization centers were equivalent.

Complex examples of equivalent results. [1] The first test case has the voters' points generated from a symmetric Gaussian distribution with mean $(0,0)$ and standard deviation 1 by 95% and from a symmetric Gaussian distribution with mean $(1.78, 1.7)$ and standard deviation 0.25 by 5%. As expected our heuristic can always distinct the two distributions and assign one representative for the second distribution as it should. For different initialization points, the results are equivalent.

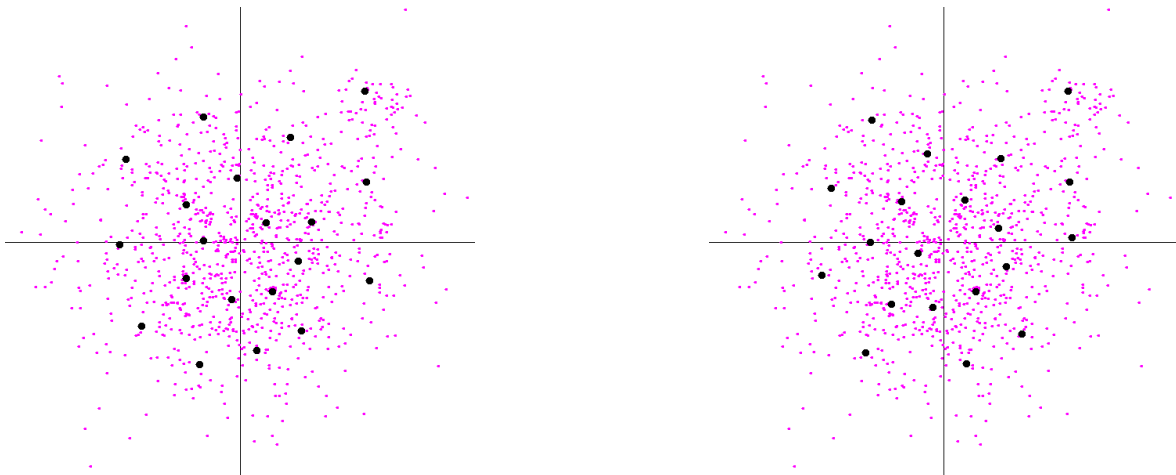


Figure 4.5: Equivalent results for Monroe heuristic [1]

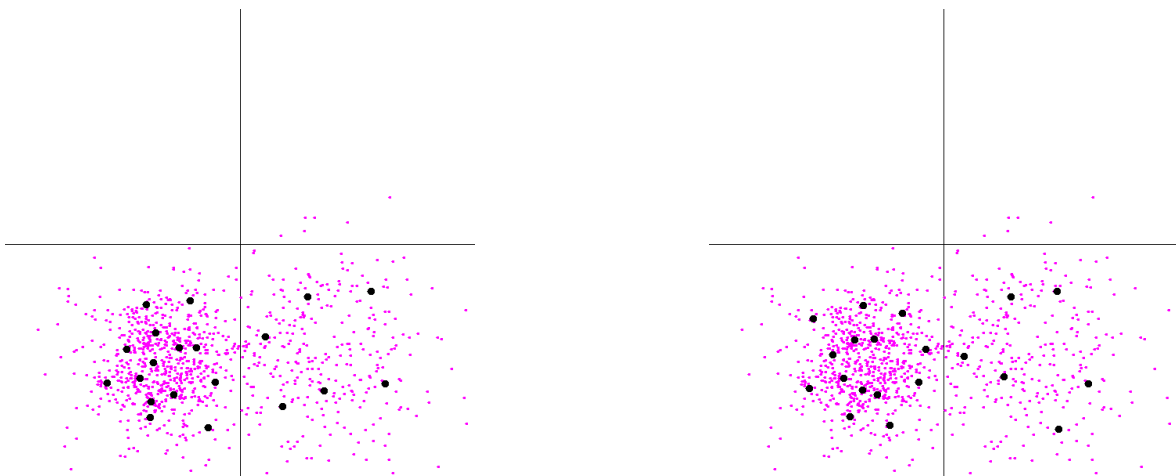


Figure 4.6: Equivalent results for Monroe heuristic [2]

[2] Another complex test case has the voters generated from a symmetric Gaussian distribution

with mean $(1, -1.5)$ and standard deviation 0.75 by 30% and the other 70% from a symmetric Gaussian distribution with mean $(-1, -1.5)$ and standard deviation 0.5.

Complex examples with non-equivalent results. [3] In the following example the voters are generated from a symmetric Gaussian distribution with mean $(0,0)$ and standard deviation 1 by 5% and the other 95% from a symmetric Gaussian distribution with mean $(1.78, 1.7)$ and standard deviation 0.25. We chose to have a sparse distribution with exactly 5% of the points to test if our heuristic could assign a representative for this distribution, as the ratio allows only one cluster. As we can see in the following images different starting positions do make a difference, assigning zero to one cluster centers for this distribution.

[4] A similar test case follows, with the voters being generated from a symmetric Gaussian distribution with mean $(0,0)$ and standard deviation 1 by 10% and the other 90% from a symmetric Gaussian distribution with mean $(2.5, 2.5)$ and standard deviation 0.25. This time the sparse distribution has the 10% of the points, allowing two cluster centers to be assigned. The initialization points may result in zero to two representatives to be assigned in the sparse distribution.

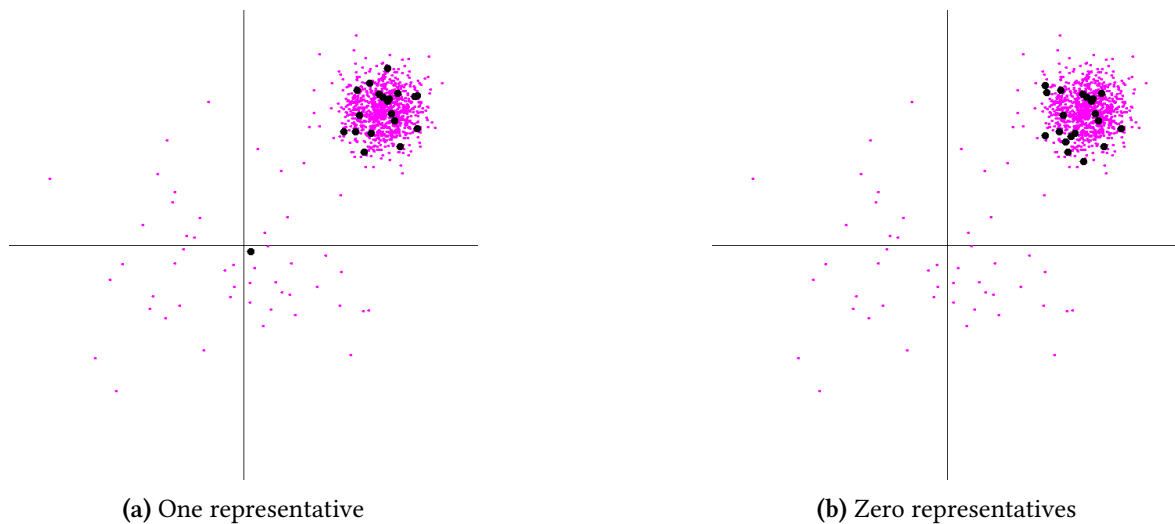


Figure 4.7: Non-equivalent results for Monroe heuristic [3]

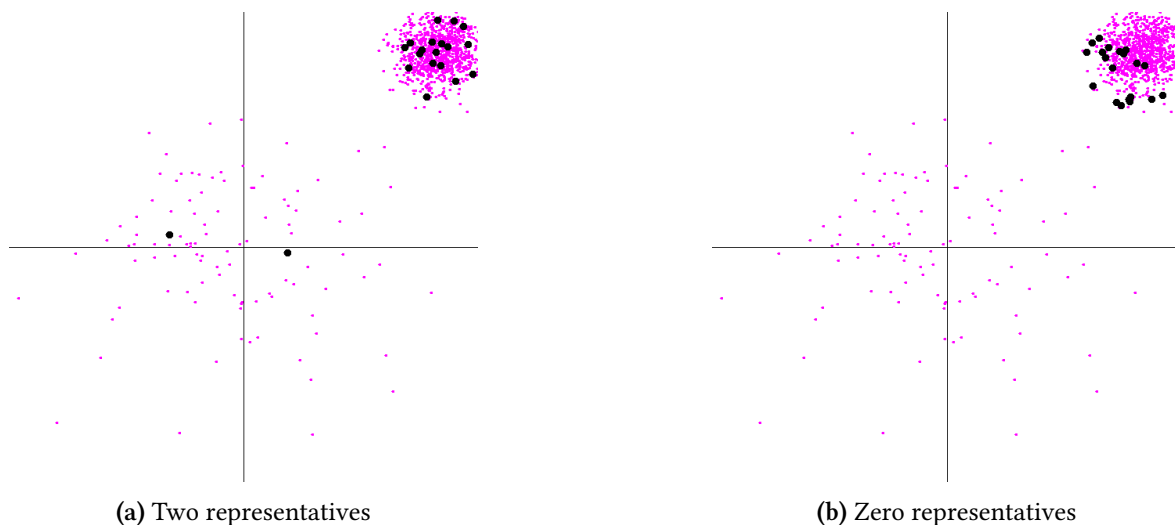


Figure 4.8: Non-equivalent results for Monroe heuristic [4]

4.2.2 Different Initialization Techniques

In order to find a better way to initialize the k centers we focus on the following test case: the voters are generated from a symmetric Gaussian distribution with mean $(0,0)$ and standard deviation 1 by 5% and the other 95% from a symmetric Gaussian distribution with mean $(-2.5, 2.5)$ and standard deviation 0.25. The optimal results for the algorithm would be one representative assigned to the sparse distribution and the remaining nineteen to the other one.

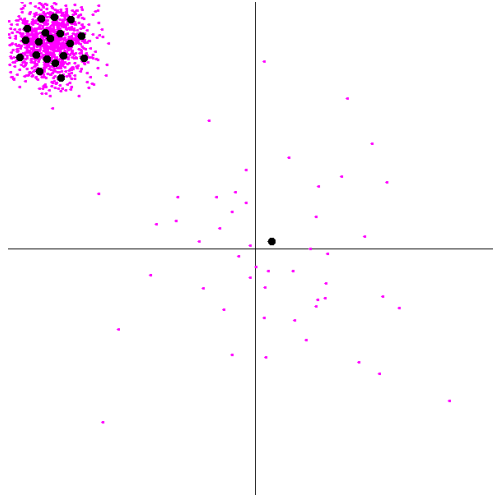


Figure 4.9: Optimal results for Monroe heuristic

At first, the initial centers were chosen uniformly at random. This had about 50% chance of assigning a representative for the sparse distribution. As a result, we needed to improve the success rate and tried the following initialization techniques.

4.2.2.1 k -means++ Initialization Technique

We wanted to confirm that the k -means++ initialization technique, as described in section 4.1, is not proper for this algorithm as well. After many tests, we could see that this initialization technique improved the success rates of the depth-first approach, but without meaning, wherefore it did not make a significant difference in the average-case scenario.

For the breadth-first approach, that we decided to follow, the success rate in this particular example has zeroed. As a result, for our proposed algorithm, we could abandon this technique with safety.

If we only change the $D(x)$ definition to sum of distances from a data point to the centers, we have already chosen and use $D(x)$ instead of $D(x)^2$ the rates go up to 20% which is still not enough and as a result, we can abandon this initialization as well.

4.2.2.2 Adjustment of k -means++ Initialization Technique for Monroe Heuristic

As imagined the aforementioned technique was proper for Chamberlin-Courant heuristic, due to the nature of its objective: to achieve maximum spreading of the initial points. In Monroe's heuristic case the objective is different: to achieve maximum spreading of the initial centers **proportional** to the clusters' number of members.

As a result, we used k -means++ initialization technique with a simple twist. After the selection of a cluster center, the $\frac{n}{k} - 1$ nearest points are removed from the X and they are perceived as the cluster's members. The proposed algorithm can now be described as follows (algorithm 5). With this technique, we were able to achieve 72% success rate in the above test case, but this was still not enough.

Algorithm 5 Monroe rule's heuristic algorithm (adjusted initialization technique)

- 1a. Take one center c_1 chosen uniformly at random from X and remove from X the nearest $\frac{n}{k} - 1$ points.
 - 1b. Take a new center c_i choosing $x \in X$ with probability $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$.
 - 1c. Remove from X the nearest $\frac{n}{k} - 1$ points.
 - 1d. Repeat Steps 1b-1c until we have taken k centers altogether.
 - 2-5. Proceed as with the algorithm 4.
-

4.2.2.3 Better Adjustment of k -means++ Initialization Technique for Monroe Heuristic

Wanting to optimize even more the appropriate initialization technique for this algorithm we added one more step in the aforementioned algorithm. In most of the cases, the chosen centers would not be the "true" representatives for the cluster, as it could be far from the center of mass. As a result we decided that it is important to add another step in the initialization of algorithm 5, in order to improve the success rates (algorithm 6). With this initialization technique the success rates reached 92%.

Algorithm 6 Monroe rule's heuristic algorithm (improved adjusted initialization technique)

- 1a. Take one center c_1 chosen uniformly at random from X , find the center of mass of the nearest $\frac{n}{k}$ points (including c_1) and replace the c_1 with the nearest point of center's mass. Remove the points from X .
 - 1b. Take a new center c_i choosing $x \in X$ with probability $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$.
 - 1c. Find the center of mass of the nearest $\frac{n}{k}$ points (including c_i). Replace c_i , if needed, with the nearest cluster's point of the center of mass.
 - 1d. Remove from X the cluster's members.
 - 1e. Repeat Steps 1b-1d until we have taken k centers altogether.
 - 2-4. Proceed as with the algorithm 4.
-

4.2.2.4 Similar initializations

In this particular test case, the success rates reach 92% with the aforementioned initialization technique. Another similar technique that has equally high success rates is if we use $D(x)$ instead

of $D(x)^2$ and $D(x)$ denotes the sum of the distances from a data point (voter) to the centers we have already chosen following the procedure presented in section 4.2.2.3. High rates can also be achieved either if we only use $D(x)$ instead of $D(x)^2$ or only change the definition of $D(x)$, following the procedure presented in section 4.2.2.2 or 4.2.2.3.

4.2.3 Proposed Algorithm

In order to complete the proposed algorithm for Monroe rule we should add an extra step to deal with the loop detection and incorporate the breadth first approach.

Let $D(x)$ denote the shortest distance from a data point (voter) to the closest center we have already chosen. The proposed heuristic for Monroe rule is the following (algorithm 7).

Algorithm 7 Monroe rule's proposed heuristic algorithm

- 1a. Take one center c_1 chosen uniformly at random from X , find the center of mass of the nearest $\frac{n}{k}$ points (including c_1) and replace the c_1 with the nearest point of center's mass. Remove the points from X .
 - 1b. Take a new center c_i choosing $x \in X$ with probability $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$.
 - 1c. Find the center of mass of the nearest $\frac{n}{k}$ points (including c_i). Replace c_i , if needed, with the nearest cluster's point of the center of mass.
 - 1d. Remove from X the cluster's members.
 - 1e. Repeat Steps 1b-1d until we have taken k centers altogether.
 2. For each $i \in \{1, \dots, k\}$, set the cluster C_i to be the set of $\frac{n}{k}$ points in X that are closer to c_i than they are to c_j for all $j \neq i$, chosen one at the time.
 3. For each $i \in \{1, \dots, k\}$, set c_i to be the center of mass of all points in C_i : $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.
 4. Repeat Steps 2 and 3 until C no longer changes or loop is detected.
 5. In case of loop, find the center of mass of all c_{ij} : $c_i = \frac{1}{j} \sum_{j \in \text{loop}} c_i$.
 6. For each center c_i find the closer candidate and declare them winners.
-

4.2.4 Proposed Algorithm's Weakness

Unlike the Chamberlin-Courant heuristic, Monroe heuristic in cases where the candidates do not span the whole space that voters span has the same behavior as the Monroe rule: The candidates are trying to reach the ideal points, due to the restriction of balanced clusters. If the elected candidates, spanned the space that the two distribution overlap, as in Chamberlin-Courant rule, the misrepresentation would rise as the elected candidates would move away from the voters, they represent.

In this algorithm, the weakness falls in the detection of the minorities and assignment of a representative. In the above sections, a single test case was examined in order to improve our algorithm. Yet if we use similar but different test cases, where the sparse and the dense distributions overlap,

keeping the same proportions (95% and 5% respectively) the success rates differ. The more the two distributions overlap the lower the success rates.

At this point, we should underline that while the means of distributions come closer, the need of distinguishing the distributions is becoming smaller and eventually, it reaches zero when the two means coincide.

4.2.5 Computational Complexity

Since this rule is NP-hard and we use a heuristic based on a clustering algorithm there are no approximation guarantees, since the number of repetitions until the algorithm converges varies.

Notwithstanding, we can compute the computational complexity for every step of the algorithm except the number of repetitions of steps 2 - 4 that the algorithm will need to converge. Let k denote the number of the committee Members, v denote the number of voters and c denote the number of the candidates. In our experiment $k \ll v$ and $v \approx c$.

For steps 1a - 1e (initialization) we have:

$$\mathcal{O}(k \cdot v) = \mathcal{O}(v)$$

For steps 2 - 4 we have:

$$\mathcal{O}(v + k \cdot v + k \cdot v \cdot \log(v) + k + v + k^2) = \mathcal{O}((2 + k + \log(v)) \cdot v + k + k^2) = \mathcal{O}(v \cdot \log(v))$$

For step 5 assuming that a loop was detected and x states are repeated we have:

$$\mathcal{O}(x \cdot k)$$

Finally for step 6 (elect committee of candidates) we have:

$$\mathcal{O}(k \cdot c) = \mathcal{O}(k \cdot v) = \mathcal{O}(v)$$

Chapter 5

Stability

5.1 Introduction

A desirable property of committee election's rules is the stability of the outcome. A need for a stable committee is very simple: can a certain voter manipulate the results and if yes to what extent? Can immorality and new voters added in the election change the outcome?

The stability of the system can be interpreted in a different way; the resistance of the voting rule to manipulation [20], [21]. Over the last 10 years there is a surge of interest regarding the complexity of strategic behavior in multiwinner elections. For some rules, e.g. SNTV and Bloc [22],[23], it is proven that the manipulation is in P, while for others, e.g. STV [24], evidence are given that are computationally resistant to manipulation. Specific attention has also been devoted to the resistance of control [22], [23], [25], [26].

We choose to make a thorough experimental evaluation of the rules we implemented under two scenarios:

1. The controlled movement of a random voter (manipulation by a strategic voter)
2. The addition of a new voter (control by a strategic voter)

Our purpose is to examine if in either case, a rule is practically manipulable/controllable by a strategic voter in our experiment and if yes, under which conditions. This can happen by characterizing the rules stable/unstable in our experiment.

We can also interpret our results in other ways, e.g. in case of either one of the two scenarios caused differences in the outcome can we extract any information over the strategic voter by comparing the outcomes? Is the instability of the outcome based on a strategical behavior and comes as planned or is it in an unexpected way?

5.2 Controlled Movement of a Voter

Trying to measure the stability of a system, we basically want to see if no new players (voters) are removed or added in the game (election), how a movement of a certain voter will affect the outcome. But the movement cannot be unlimited and we must determine the willingness of a voter to move from the voter's real position and the voter's real benefits and misrepresentation.

5.2.1 Experimental Setup

We have, once again, a set of 1000 voters and 1000 candidates generated according to a certain distribution. The voters and candidates are represented as points in the Euclidean Domain. For every voter, there is a preference order of the candidates. A committee of 20 members is elected for every rule. We choose a random voter p which is moved in a random direction inside the limits of the allowed movement and then we elect the 20-membered committee for every rule with the new ranking for voter p . Then we compare the outcomes.

The movement of the voter must be controlled. If the voter could move unlimitedly it would be like we have removed and put a new one in her place. As a result, we decided that the voter p could move as much as p wants given the fact that p 's first preference would not change. This can be mathematically described by the following formula

$$0 < \text{movement} < \frac{\text{distance}(c_2) - \text{distance}(c_1)}{2}$$

wherefore the movement of the voter cannot be bigger than half of the difference of the distances with p 's first two preferred candidates. This was the only restriction that could ensure that the first preference would not change, whatever the direction of the movement.

This restriction must be received with extreme caution. Due to the nature of our experimental setup, there is a bias that we must take into consideration while studying the results. The movement may seem that restricts the voter p only by locking his first preference but also does not let p to change freely his other preferences, since we keep the location of the candidates locked. The changes in his ranking vastly depend on the candidates' distribution.

5.2.2 Results

In this section's images, the voters are represented by purple dots and the candidates by black dots. The elected candidates in the first experiment are represented by big black dots and in the second experiment (after the movement of the voter) are represented by big cyan dots. If the elected committees coincide, the cyan dots seem like having a black outline. If not, each cyan dot without a black outline is associated with a black dot.

5.2.2.1 Single Nontransferable Vote rule

As expected Single Nontransferable Vote rule is completely unaffected by this kind of movement of a certain voter. As this rule can only be affected by the first choice of each voter, any movement that does not change the first preference, cannot affect the outcome.

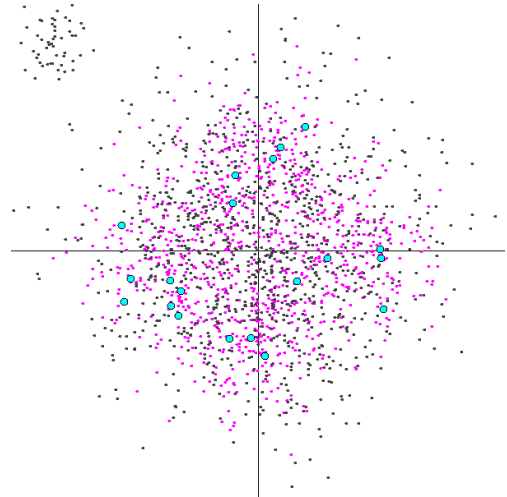


Figure 5.1 Single Nontransferable Vote rule - stable

5.2.2.2 Single Transferable Vote rule

The Single Transferable Vote rule is really sensitive to every movement. However a such a movement, that does not affect the first place votes and does not cause noticeable differences in the rankings, rarely causes differences in the outcomes. As a result, we can say that this rule is stable in this kind of movement.

Nevertheless, if even one change happens in the order the candidates are elected or rejected, the two outcomes would be entirely different.



Figure 5.2: Single Transferable Vote rule - stable

5.2.2.3 Bloc rule

Bloc rule is also unaffected by such a small movement of one voter. This result is also expected, as in this rule each voter approves k candidates. The movement of the voter may change the ranking series of the top k candidates, but if there is only a rearrangement and the top k candidates stay the same, the outcome cannot be affected. Even if some of the top k ranked candidates changed, in order to affect the outcome there should be a tie-breaking or tie-creating situation in the last position after the voter's movement.

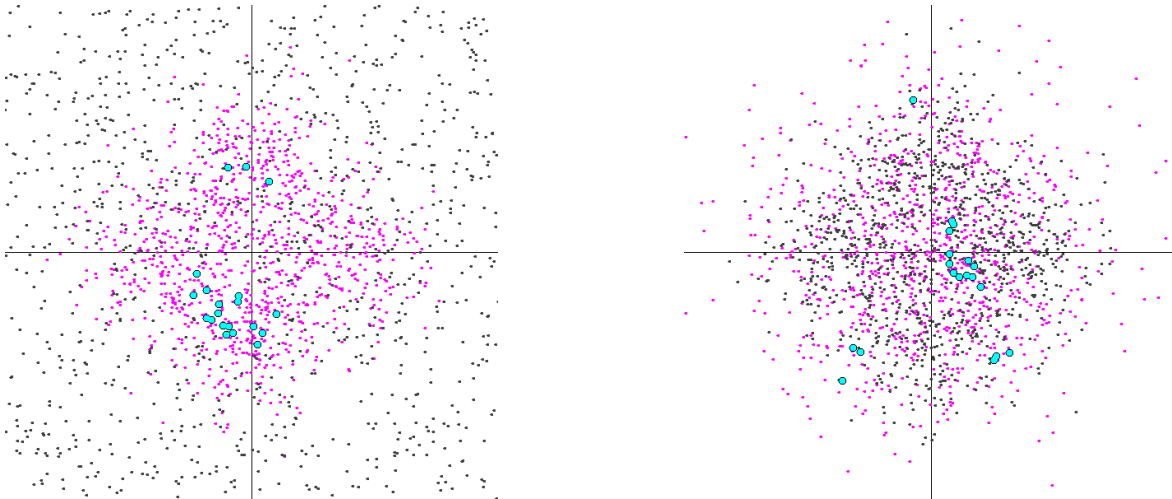


Figure 5.3: Bloc rule - stable

5.2.2.4 k -Borda rule

k -Borda rule takes into consideration the whole ranking of each voter. Consequently, a voter's movement changes the points each candidate is assigned by this voter's rank. Nonetheless, this cannot affect the outcome.

k -Borda rule always declares winners that are closer to the center of mass of the voters. And as expected, one small movement is not enough to change the outcome, due to the fact that the nearest candidates to the center of mass of the voters' set always have the most votes.

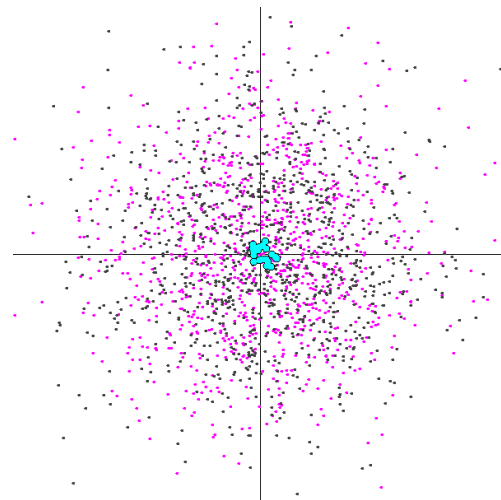


Figure 5.4 k -Borda rule - stable

5.2.2.5 Chamberlin-Courant rule

Naturally, we cannot talk about the stability of the Chamberlin-Courant rule rather than the stability of our heuristic. The results for the Chamberlin-Courant rule indicates that our heuristic is almost stable.

This conclusion comes under a certain condition: in both experiments, the initial place of the committees (initial centers) must be the same. As it was described in section 4.1 our algorithm starting from different initial centers can converge in different points. Therefore, if we ignored this condition the heuristic would appear unstable.

Another important thing to note is the difficulty of more than one cluster center to change. Our heuristic uses k -means in order to find the center of the unequal clusters. After k -means converge and the centers stabilize, our algorithm finds the nearest candidates to the centers of the clusters and returns them as the elected committee. Thus, a way for an elected member to change is a center to move closer to another candidate.

Due to the small movement of only one voter, it is difficult for the clusters to change, but not

impossible, even though this movement tends to zero. There are cases that a movement of a single voter, even though it is so small, can affect the outcome, as the voter moves in a different cluster and as a result some of the points that the algorithm converge differ. Therefore, our algorithm, as a clustering algorithm, cannot be described as perfectly stable since the movement of a point may lead in extreme cases in an entirely different outcome.

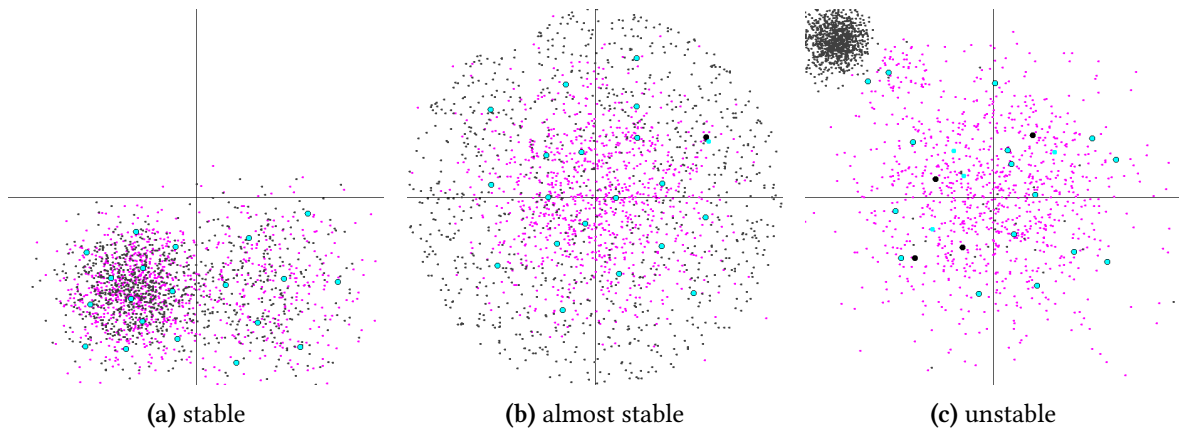


Figure 5.5: Chamberlin-Courant heuristic - almost stable

5.2.2.6 Monroe rule

Monroe rule is NP-hard, as well, and we study the stability of our heuristic, under the same conditions as in Chamberlin-Courant heuristic: the initial centers are the same in both conduction of the experiment. The results indicate that Monroe’s heuristic is unstable under this condition.

As we can see in the results there are cases that the outcome is perfectly stable, but this is not the average case. In the average case, the outcome is unstable. Even such a small movement of only one voter, due to the restriction of the size of the clusters, may result in the movement of the voter in a different cluster which can cause linked movements among the clusters and as a result different convergence points.

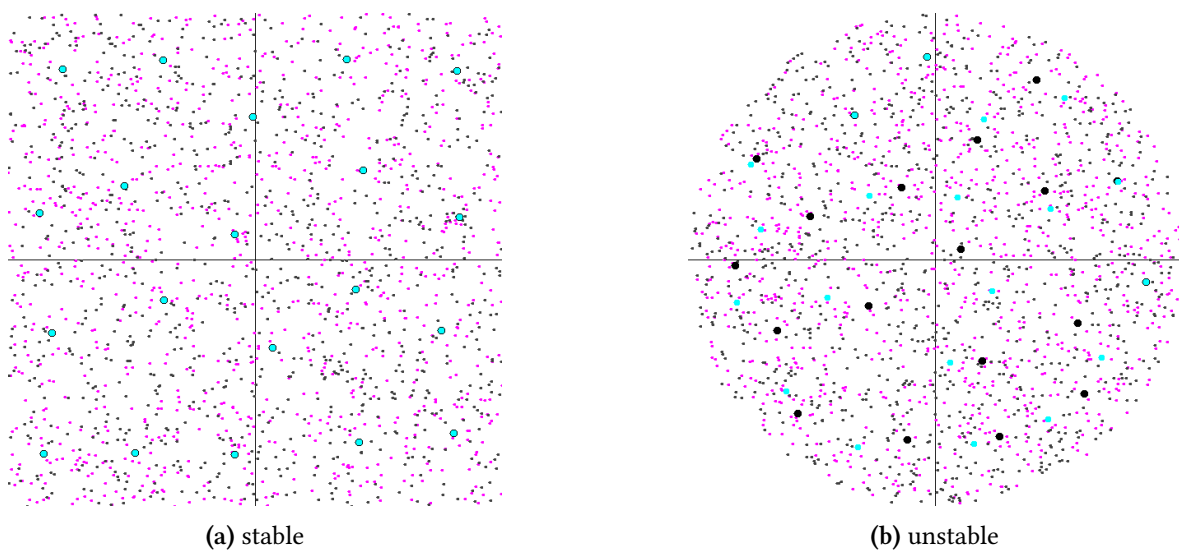


Figure 5.6: Monroe heuristic - unstable

As we described, while we analyzed Chamberlin-Courant heuristic’s stability, due to the movement of the voter, if there are not exchanges of voters among the clusters, there is a small movement

of the cluster’s center, which represents the voter who moved. This movement is proportional to the movement of the voter and may or not result in a movement of the committee. This fully depends on the candidates’ distribution. If a candidate, who locates closer to the new center of the cluster, than the formerly elected candidate, exists we notice a movement of an elected member of the committee. We should also note that the movement of this member also depends on the candidates’ distribution. If a change occurs, the more sparse the distribution, the bigger the change.

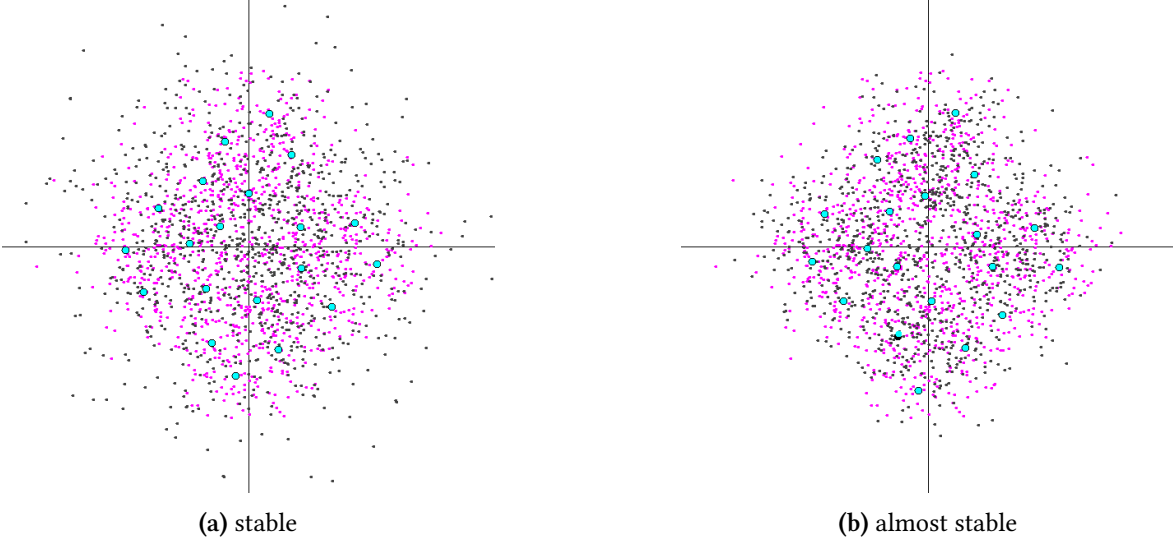


Figure 5.7: Movement of a single elected member depends on the candidates’ distribution

An interesting observation is how the stability or non-stability of the outcome is perceived according to the span of candidates and voters in the two-dimensional space. When the candidates’ distribution do not span the space spanned by the voters’ distribution the results appear to be perfectly stable. While the candidates’ distribution starts to span the space which voters’ distribution spans, the outcome starts to be less stable, up to the point that the two distributions overlap and the outcome is unstable. As a result, we have come to the conclusion that some of the stable results are isolated cases, which exploit the fact that the heuristic does not take into consideration the candidates positions. This observation is the key to our decision to characterize this rule as unstable¹.

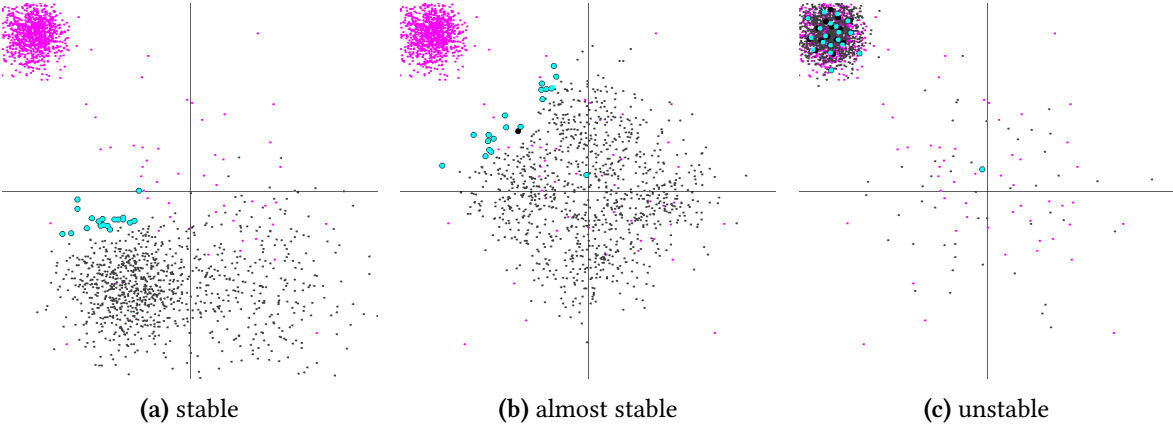


Figure 5.8: The stability of Monroe heuristic as the two distributions start to overlap

¹ In the Chamberlin-Courant heuristic there are outcomes that are not stable enough and do not let us characterize the rule as perfectly stable. However, due to the fact that we could not detect such a pattern as we see in Monroe heuristic, we believe that the stable outcomes are truly stable or almost stable and that the unstable results are isolated cases.

5.2.3 Conclusions

As it became clear, some of the voting rules, e.g. SNTV, Borda, k-Bloc, were completely unaffected by such a small movement that may change some places in the ranking but not the first one. This can be easily explained if we think the nature of these rules.

SNTV can only be affected by the first choice of each voter. Given the fact that this would not change, this rule is insusceptible in this kind of movement. Bloc, on the other hand, can be affected only from the k-first preferences. But the allowed movement does not cause tremendous changes in the first positions and each one of the first k positions is awarded in one point exactly. As a result, a change between the third and fourth place does not affect the outcome. Borda may take into consideration the whole ranking but the result can be taken as fixed as it always coalesces in the center of mass of the voters' distribution as long as the candidates' distribution allows.

An unexpected result was the stability of STV in a small movement like that. Even though STV rule is really sensitive to every movement, no matter how small or big, this movement does not affect the outcomes. However, this conclusion comes with a little skepticism. If this movement results in a change in the order the candidates are elected or rejected, the outcome would be entirely different and the rule would appear unstable.

The Chamberlin-Courant heuristic can be characterized as stable as well. There are some extreme cases where the results may differ, but these constitute the minority. On the other hand, Monroe's heuristic is unstable even with such a small movement. This was not unexpected though. The restrictions of this algorithm, make Monroe heuristic so sensible to make even such a small movement significant enough to change the clusters.

5.3 New Voter's Addition

We also wanted to measure how stable each rule is, according to the addition of a new voter. This case, also covers the movement of a voter in larger extend, thus in this scenario we generate a new voter according to the existing distribution.

5.3.1 Experiment

We have, once again, a set of 1000 voters and 1000 candidates generated according to a certain distribution. The voters and candidates are represented as points in the Euclidean Domain. For every voter, there is a preference order of the candidates. A committee of 20 members is elected for every rule. A new voter p is added in the set of voters and we elect the 20-membered committee for every rule with the new set of 1001 voters. Then we compare the outcomes.

As mentioned earlier, the addition of the new voter happens according to the existing distribution. But each addition may have more or less impact on the outcome due to the randomness in the generation process. As it becomes clear, it is of high importance to find a way to measure the stability of a rule in comparison to the addition, as some additions affect the outcome more.

Regardless of the tendency of the outcome to be affected, we do not expect great changes. We only expect, in most rules, differentiation in the last places of the outcome. As for this expectation, it can be easily explained. The ties are solved by a predetermined way, but if a tie exists in the last places of the committee, there are elected as many candidates as the free positions. As a result in many cases, one candidate that compeer in the last place but is not elected may gain a vote that will help her overcome the tie.

And here comes the big question: Do we really have a change in the committee? Or do we incorrectly think that? Wider research must be done in this area, because if we choose to fill not only the needed positions, there are more questions to ask, e.g., if there is a tie in the second place do we deal with it as only one position is filled or as both the second and third positions are filled? The solution might be to find all the equivalent outcomes and compare the new outcome with each one of them. The pair with fewer differences is the real comparison item.

Another interesting question is if all rules can be judged under the same conditions or if this is unjust and as a result, our conclusions would be biased. Lastly, we would like to see which rules tend to reveal information about the additional voter, by comparing the two elected committees, knowing where the new voter locates.

5.3.2 Results

The images in this section follow the same pattern as above. The voters are represented by purple dots and the candidates by black ones. The additional voter is represented by a big purple dot. The elected candidates in the first experiment are represented by big black dots and in the second experiment (after the addition of the voter) are represented by big cyan dots. If the elected committees coincide, the cyan dots seem like having a black outline. If not, each cyan dot without a black outline is associated with a black dot.

5.3.2.1 Single Nontransferable Vote rule

We expected that the Single Nontransferable Vote rule stable under the addition of one voter and our expectations were confirmed. We also expected, that the outcome could be affected by most one change in the elected committee. As the rule takes into consideration only the first preference of the voters, the addition of only one voter, and as a result of only one vote, can affect the outcome only by promoting one candidate by one rank.

There are two important questions regarding this result. Is there a way to measure the impact of a certain voter? And most importantly can we extract any information regarding the new voter by comparing the two outcomes?

As for the first question, the answer is no. We cannot find a metric to measure the impact on the committee. One voter's addition may result in a small or big change in the candidate who is substituted by the newly elected member.

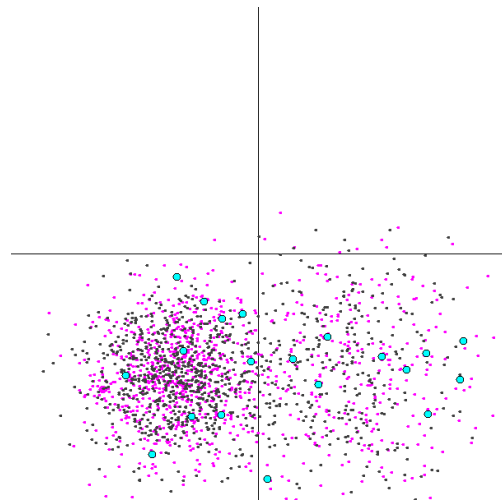


Figure 5.9 Single Nontransferable Vote rule - stable

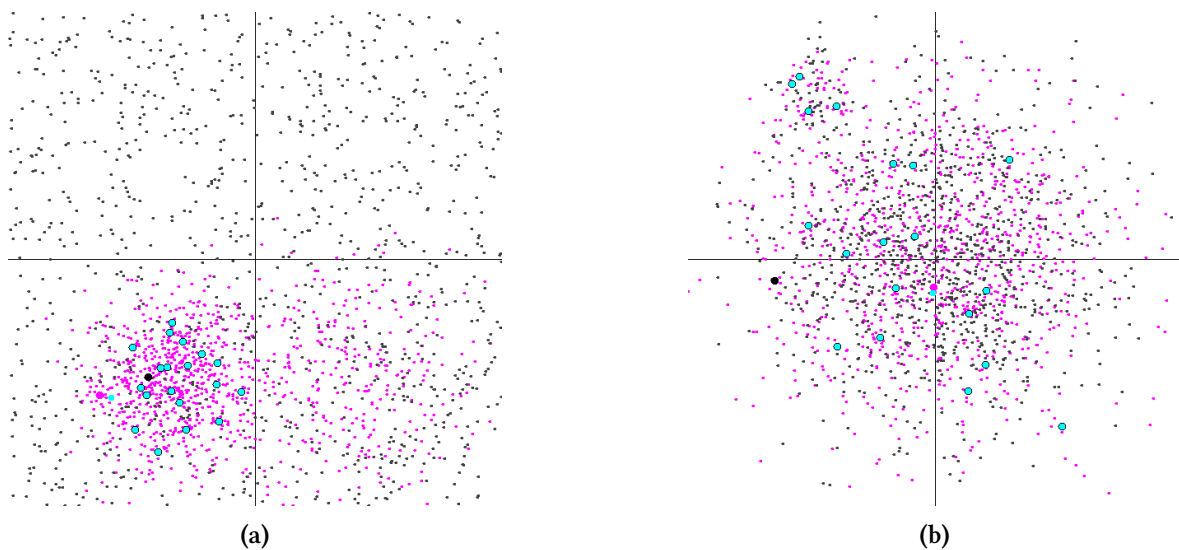


Figure 5.10: Conclusions over Single Nontransferable Vote rule

1. We can extract information for the new voter by comparing the two committees
2. We cannot measure the impact of the voter's addition.

Concerning the second question, if a change occurs in the elected committee after the addition of a voter, we can extract information about the new voter. It is obvious in the above images, that the new candidate (the only cyan dot without outline) locates closer to the new voter (large purple dot) than any other candidate (small black dots). As a result, even if we didn't visualize the new voter, we would be able to guess that the new voter locates very close to the elected candidate.

5.3.2.2 Single Transferable Vote rule

The Single Transferable Vote rule seems unstable under the addition of one voter. This rule, as we mentioned earlier, is really sensitive and even a change in the series of the candidates' election or elimination can change the entire outcome. Even though there were many stable outcomes, we could not characterize the rule as stable, not because of the existence of many unstable results, but due to the very nature of the rule.

Due to the instability the unpredictable results of a single voter's addition, there is no way to gain information about the voter.

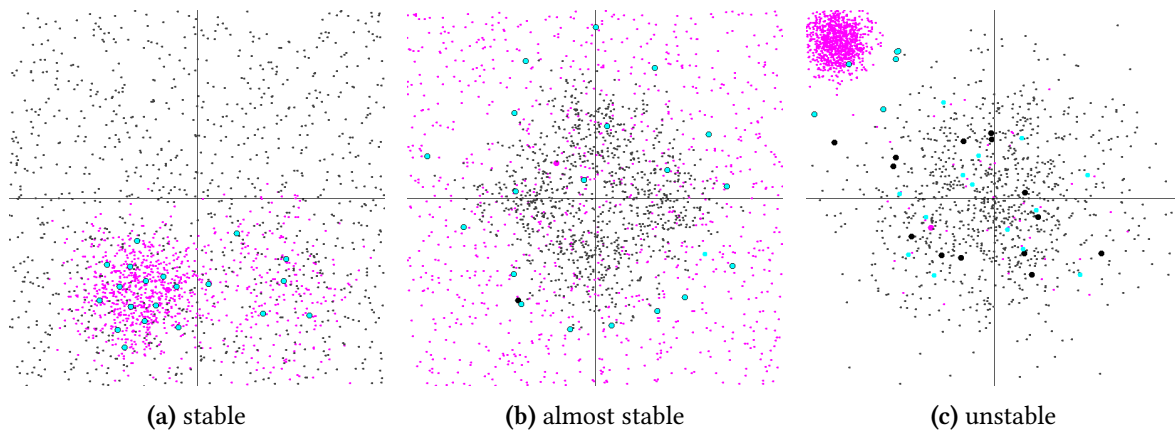


Figure 5.11: Single Transferable Vote rule - unstable
We cannot extract any information for the additional voter

An interesting observation can arise from the comparison of two almost similar versions of the Single Transferable Vote rule. The only thing we changed from the algorithm described in 3.2.2 was the way we transferred the votes of the eliminated candidates. In the first version, we transferred the votes, which the candidate gained from elimination or election of other candidates, evenly, with $\frac{\text{votes to transfer}}{\text{remaining voters}}$ for each voters' next preference. In the second version, all the "transferred" votes are moving to the candidate that the most voters prefer most next.

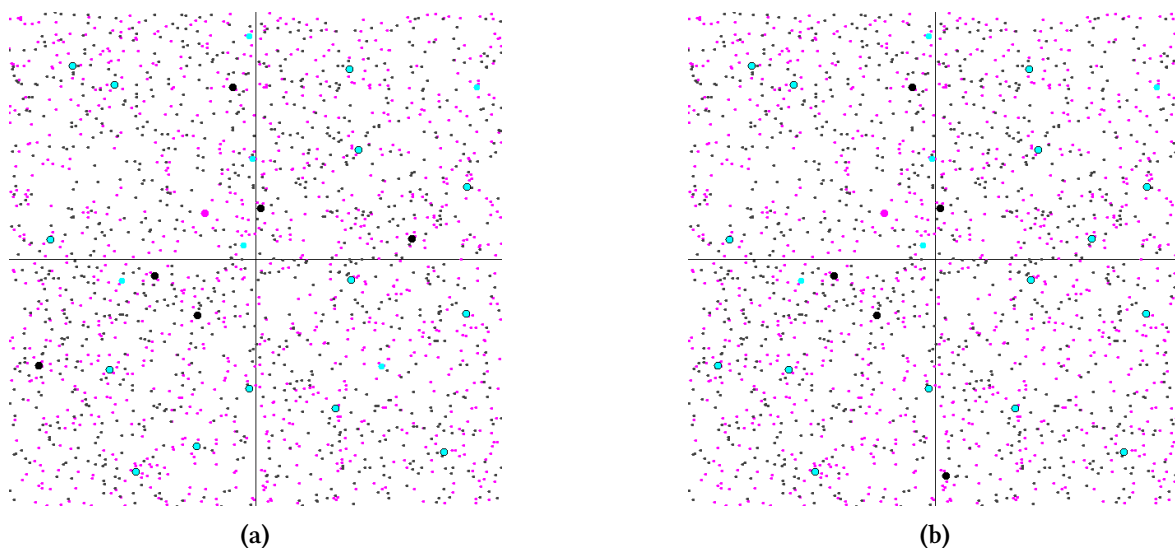


Figure 5.12: Different results for different versions of Single Transferable Vote rule
In both images the same voters and candidates are used

In both algorithms, the results were almost the same in many cases, but there were also differences, e.g. in the same test case one rule may be more stable than the other, etc. However, neither of these versions were stable, and there was no pattern in order to characterize one version more stable than the other.

5.3.2.3 Bloc rule

Bloc rule is stable with the addition of one extra voter. In the majority of the results, the outcomes are perfectly stable. When the two committees did not fully coincide, they differed by only one elected candidate. After a more thorough evaluation, even though the results indicated that only one elected candidate may alter, this only happened when in the $k - 1$ first positions of the committee, the elected candidates have more votes than the last elected candidate. Otherwise, i.e. if l elected candidates are tying in the final l positions, the addition of one voter could potentially change l places in the committee.

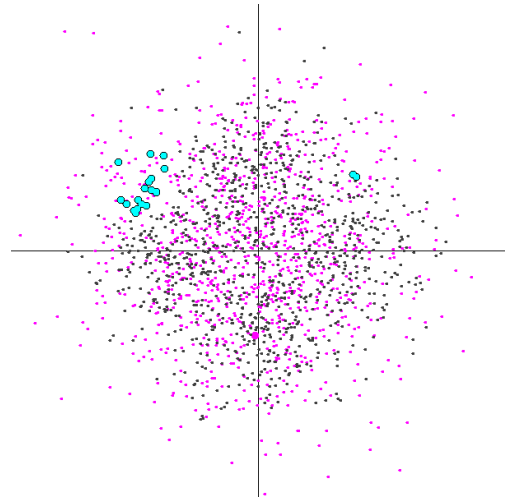


Figure 5.13 Bloc rule - stable

The bigger the impact, the bigger the extraction of information over the additional voter's preferences. From the following images, we can conclude that the newly elected candidate is always closer to the additional voter than the one replaced². If we assume a scenario where more than one candidates differ in the two committees, this could only restrict the area that the additional candidate locates and as a result, we could have more information over the additional voter's preferences.

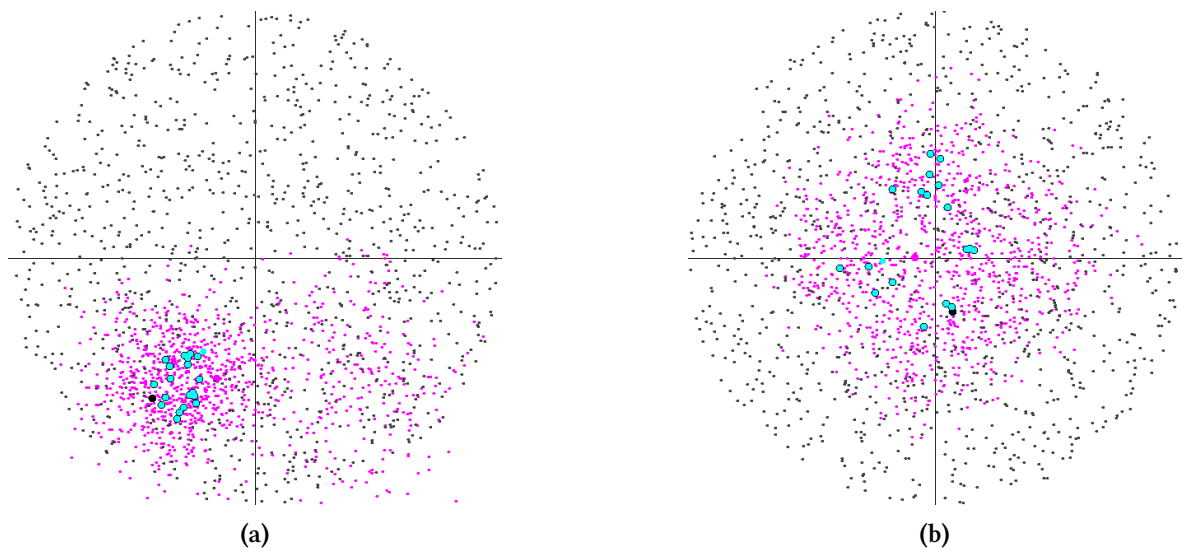


Figure 5.14: Conclusion over the Bloc rule

We can extract information for the additional voter by comparing the two committees.

² The additional voter can distance at most equal the old and newly elected candidates. As a result, we can mathematically describe the part of the space that the new voter locates.

5.3.2.4 k -Borda rule

k -Borda rule is also a stable rule. The committee can be affected but the results reveal that only one elected candidate can alter. However, even though we characterize the rule as stable, we know that the two committees can differ by more than one position. This is an extreme case, though. It would need all the other votes to aggregate in a specific outcome, in order to be possible for the additional voter to change more than one positions of the committee.

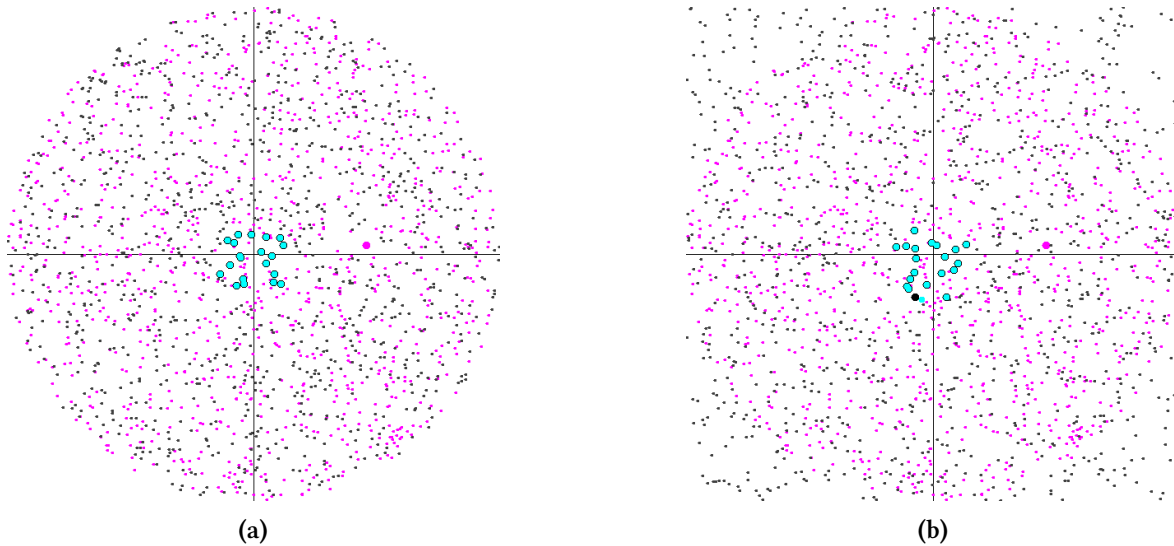


Figure 5.15: k -Borda rule - stable

We can extract some general information for the additional voter, but it is really general, as we only know that the new voter is further away from the replaced candidate than the newly elected. Although we cannot make any estimations over the candidates' exact location, if we do not know any information such as the aggregated ranks for the elected members before and after the addition of the voter.

5.3.2.5 Chamberlin-Courant rule

As mentioned earlier, the Chamberlin-Courant rule is NP-hard and as a result, we are studying the stability of our heuristic. We expect Chamberlin-Courant to be almost stable and the location of the additional voter, in relation with the existing voters and the density of the voters' distribution on this area to affect the outcome.

An important observation is that the outcome that our algorithm suggests will be always affected from the addition of the voter. However, this may not be visible, according to the candidates' distribution. The same proposed outcome may seem stable, almost stable or unstable, due to the different candidates' positions. And even though the outcome is never fully changed, the outcome cannot be characterized as stable, as one or more convergence's points change.

An interesting observation is that our heuristic does not use candidates location, till the algorithm suggest an outcome and finds the closer candidates. As a result, we cannot extract any information for the additional voter, even though we know that something has changed.

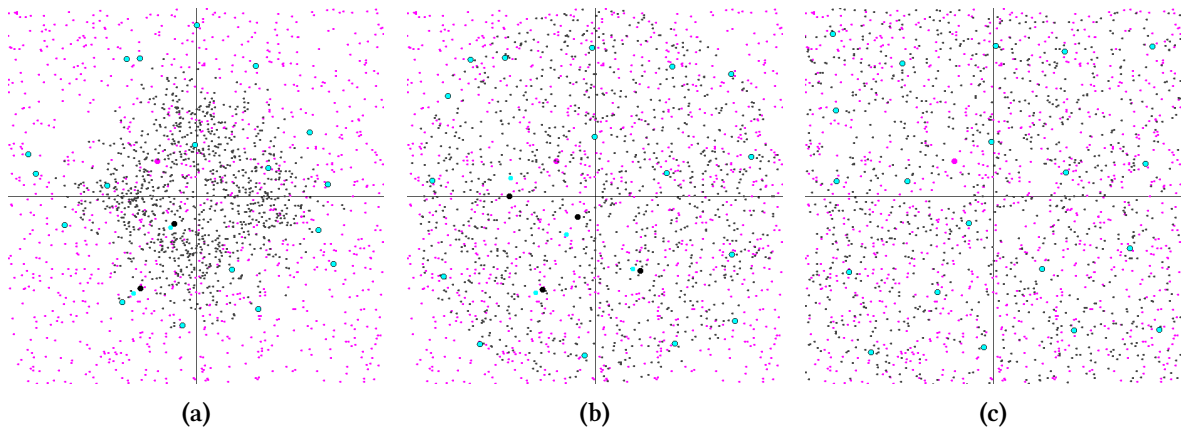


Figure 5.16: Chamberlin-Courant heuristic - unstable

The voters remain the same, while the candidates' distributions differ: (a) 4-Gaussian, (b) Uniform Disc, and (c) Uniform Square

5.3.2.6 Monroe rule

As mentioned earlier, Monroe rule is NP-hard and as a result, we are studying the stability of our heuristic. We expected that Monroe would be completely unstable with the addition of a voter. And the results confirmed it.

However, this may characterize this rule over our conditions, but it may be an unfair characterization due to the nature of the rule. While in the initial experiment there could be no more than 50 candidates per cluster, while an additional voter was added, one cluster should include an extra candidate. This is absolutely certain that would change the results because it would lead in different convergence's points. As a result, this criterion is not a proper one, or even a fair one, for this rule.

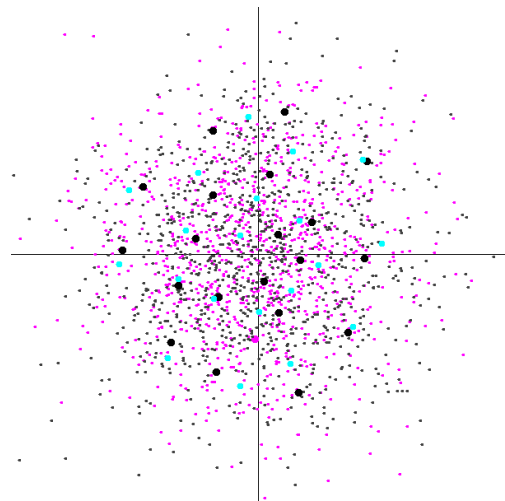


Figure 5.17 Monroe heuristic - unstable

5.3.3 Conclusions

In this chapter, we came to conclusions regarding the stability of the rule's outcomes when a new voter is added, the revelation of information regarding the new voter and the fairness of judging all the rules under the same criteria.

As far for the stability of the rules, we concluded that Single Nontransferable Vote, Bloc, and k -Borda rule can be characterized as stable. Normally, if the outcome is not perfectly stable only one voter can change. In Bloc and k -Borda rule, more can potentially change, but under extremely specific conditions.

The Single Transferable Vote rule is unstable, as the addition of a voter can change the series of candidates' election or elimination. Chamberlin-Courant is unstable as well, as the addition of a

voter can change the convergence points more easily as the movement of a voter. Lastly, Monroe rule is perfectly unstable, as expected. The addition of the voter changes almost every cluster.

As far as the disclosure of information regarding the additional voter, only the stable ones are able to reveal anything about the location of the voter. The more stable the rule the more the information it reveals, when the outcome is not perfectly stable. The rules reveal information in reverse order they can potentially be affected by only one voter, i.e. Single Nontransferable Vote reveals the location of the voter, while k -Borda and Bloc reveal general information, which in case of Bloc can be narrow down the possible locations.

Finally, as far as the fairness of judging all the rules under the same criteria, the only rule that we could claim that was unfairly treated is Monroe heuristic, because the size of clusters change and as a result, the instability of the rule is anticipated. Still the results regarding the stability of Monroe heuristic for the movement of a voter, affirm that Monroe's heuristic is unstable under most conditions and this is only one of them.

5.4 Summary of Stability Results

	Stability					
	Best case		Average case		Worst case	
	Movement of Voter	Additional Voter	Movement of Voter	Additional Voter	Movement of Voter	Additional Voter
SNTV	✓	✓	✓	✓	✓	♣
STV	✓	✓	✓	✗	✗	✗
Bloc	✓	✓	✓	✓	♣	♣
k-Borda	✓	✓	✓	✓	♣	♣
CC	✓	✓	♣	✗	✗	✗
Monroe	✓	✓	✗	✗	✗	✗

Table 5.1: Stability of Multiwinner Vote Rules

✓ indicates that the rule is perfectly stable, while ✗ indicates that the rule is unstable. ♣ indicates that the rule is stable but there can be potentially a small change (under certain conditions).

Regarding the disclosure of information over the additional voter, we can summarize the results in the following table. It is noted that an instability over the outcomes, either small or big, is assumed.

	Rules					
	SNTV	STV	Bloc	k-Borda	CC	Monroe
Disclosure of Information	✓	✗	✓	✓	✗	✗

Table 5.2: Disclosure of information over the additional voter

✓ and ✗ indicate that the rule does or does not disclose any information over the additional voter.

Chapter 6

Design & Implementation of the Experiment over the Multi-Dimensional Space

We also wanted to study examples closer to real life where we only have the voters' rankings. We could create our own test but we choose to use real election sets from "PrefLib: A Library for Preferences"[27]. This is a reference library of preference data and links. They have attempted to unify the formatting of the data as much as possible. They currently have 10 types of data in 3 formats. More details about the formation we choose to study are given later in the chapter.

6.1 Election

6.1.1 Experimental Setup

For every simulation, a certain set of voters and candidates, where provided. All the test cases from PrefLib that were used, came from the SOC (Strict Orders - Complete List) package, meaning that for every voter a preference order of the candidates, with no ties, is given. Our objective is to elect a committee (a subset of candidates) that represents in a certain way the voters. Since both the voters and candidates number differ per experiment, so does the committee members, which range from 3 to 5. We rejected many test cases that had up to 5 candidates or voters. Thus, we ended up with the following test cases.

1	ED-00006-00000003	14	ED-00006-00000034
2	ED-00006-00000004	15	ED-00006-00000035
3	ED-00006-00000007	16	ED-00006-00000036
4	ED-00006-00000008	17	ED-00006-00000037
5	ED-00006-00000011	18	ED-00006-00000044
6	ED-00006-00000012	19	ED-00006-00000046
7	ED-00006-00000018	20	ED-00006-00000048
8	ED-00006-00000021	21	ED-00009-00000001
9	ED-00006-00000022	22	ED-00009-00000002
10	ED-00006-00000028	23	ED-00012-00000001
11	ED-00006-00000029	24	ED-00014-00000001
12	ED-00006-00000032	25	ED-00032-00000002
13	ED-00006-00000033		

In order to study the stability, we copied the test cases twice and edited them, the first time by editing one random voter (without changing the first ranked candidate) and the second time by adding one extra voter.

6.1.2 Representation of the Voters & Candidates

In the previous experiment, where both voters and candidates were represented as points in the Euclidean Domain, we generated the rankings and the voters' misrepresentation from a certain candidate was defined as the distance between the voter and the candidate. The bigger the distance the less the benefit for the voter.

But in this experiment, we do not have two-dimensional coordinates. Instead, we only have the rankings of the voters. Notwithstanding that for the SNTV, STV, k -Borda and Bloc rules, the rankings are enough, since it is the only thing we need in order to assign the votes and find the winners, this is not the case for Chamberlin-Courant and Monroe rules. These rules are trying to minimize the misrepresentation of the voters and as a result, we need to interpret the rankings into misrepresentation with a utilization function, that can be linear, convex or have any other form that may be appropriate at given time.

Moreover, Chamberlin-Courant and Monroe rules are NP-hard. As a result, we used the heuristics presented in chapter 4. Our heuristics use the location of the voters in order to create the proposed clusters and then and only use the candidate's positions in order to find the ones closer to the clusters' centers. In order to be able to use the same algorithms, we interpreted the rankings as locations in a multi-dimensional space. In each experiment there were n dimensions, where n is the number of candidates and each candidate C_i is represented as a vector in basis with the following form:

$$C_0 = [1\ 0\ 0\ \dots\ 0], C_1 = [0\ 1\ 0\ \dots\ 0], \dots, C_{n-1} = [0\ 0\ 0\ \dots\ 1].$$

As a result each voter V_j is represented as:

$$V_j = [c_{0j}\ c_{1j}\ c_{2j}\ \dots\ c_{(n-1)j}],$$

where c_{ij} is the position that voter V_j ranked candidate C_i and $1 \leq c_{ij} \leq n$.

With these in mind, we were able to use the same exact algorithms, with one differentiation. We decided to use L^1 distance metric (Manhattan distance metric) instead of L^2 distance metric (Euclidean distance metric) while calculating the distances since it works better in the multi-dimensional space[28]. This way we have a simple linear utilization function.

We also used an exponential utilization function where each voter V_j is represented as:

$$V_j = [2^{c_{0j}-1}\ 2^{c_{1j}-1}\ 2^{c_{2j}-1}\ \dots\ 2^{c_{(n-1)j}-1}]$$

However, after the experiments, we rejected the exponential utilization function for Chamberlin-Courant heuristic, because the algorithm develops the inability to converge. Further research must be done for the appropriate utilization function.

6.2 Results

The given test cases due to the small number of candidates or/and voters do not let us study properties such as diversity or proportional representation.

In terms of stability, there is no point for studying rules that use only the rankings, because we have already tested them in chapter 5, since the setting is not affected by the representation of the voters in two- and multi-dimensional space. The only thing that the experiment revealed for the SNTV, STV, Bloc and Borda rules is the extent of our conclusions since the weight of each vote changes due to the change of the voters' and candidates' number.

Chamberlin-Courant and Monroe heuristics, on the other hand, are based on clustering algorithms. Therefore the number of voters, candidates and as a result of committee's members render the test cases inappropriate for our study.

An interesting conclusion that came up from the experiments was that in test cases that the rankings were taken from races, almost all the rules had the same outcome. This can be explained by the fact that there is an "objective" ranking (and as a result "objective" winners) and all the voters are trying to approach it. Hence, the small number of the test-cases make us skeptical about our conclusion and further research must be done.

Chapter 7

Future Work

We approached the topic of multiwinner voting rules mainly in the two-dimensional spatial domain. We focused on the heuristics of the NP-hard rules and on the property of stability. This work can be further extended and improved in various ways.

One step of action would be to examine different clustering approaches, e.g. hierarchical agglomerative clustering approach [29], in order to build different successful heuristics or approximation algorithms. This is an active domain of research, especially for Monroe rule due to the built-in restrictions it has.

Secondly, one possible extension could be in the study of stability. Our study was restricted in two basic cases, i.e. the limited movement of one voter and the addition of a voter. This could be expanded in other directions as well, such as the removal of a voter or even a candidate. Due to the interpretation of stability in different ways, e.g. manipulation, control [30], etc. we are interested in all the different cases. Another interesting approach would be to evaluate the scalability of the difficulty of manipulating the results.

Moreover, another interesting approach is based on the challenges of the field. Many voting rules, e.g. k -Borda, Chamberlin-Courant, Monroe, etc. require full preference rankings provided by voters, and with a large number of candidates, such information might be infeasible. As a result, it is difficult to find suitable election samples for studying such rules. Besides, the lack of suitable samples was our problem in our study for multiwinner elections in multi-dimensional space. It is thus natural to study multiwinner voting rules for the case that only partial preference is available. This can also have huge effect to problems where only partial preferences are available but the whole "true" ranking must be revealed (similar to the problem of having noisy votes [31] and try to find the "true" ranking).

Finally, a challenging approach would be to find a way to present the results and to classify the candidates who lost. In most of the cases it is easy to say who won the election, but sometimes more information is needed, e.g. the candidates may wish to know how well they did even if they lost. This can be really challenging for some rules, e.g. Single Transferable Vote rules, due to the complexity of rule. Studies must be made in order to find a way to evaluate the results, rules weaknesses and the success of the candidates. In this approach, the algorithmic properties of the rules (diversity, proportional representation, fairness) can be a really useful and valuable asset.

Μέρος II

Ελληνικό Κείμενο

Συγκριτική Αξιολόγηση Κανόνων Ψηφοφορίας για την
Εκλογή Επιτροπών

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο

Ως εκλογή επιτροπών ορίζεται η διαδικασία εκλογής ενός υποσυνόλου των υποψηφίων (αντικειμένων ή ατόμων) ως αντιπροσώπων. Σε καθημερινές επιχειρηματικές αποφάσεις, στις εκλογές κλπ. στόχος είναι η δημιουργία ενός συλλογικού προφίλ προτίμησης από ένα σύνολο ατομικών και ενδεχομένως αντικρουόμενων προφίλ. Αν όμως η δημιουργία του συλλογικού προφίλ μπορεί να γίνει με πολλούς και διαφορετικούς τρόπους, πώς βρίσκουμε το προφίλ, δηλαδή μια επιτροπή αντιπροσώπων που ικανοποιεί κριτήρια όπως η αντιπροσωπευτικότητα; Πώς μπορούμε να διασφαλίσουμε ότι αυτό το αποτέλεσμα δεν θα είναι χειραγωγήσιμο από ένα συγκεκριμένο άτομο ή ότι ο κάθε χρήστης θα ωφεληθεί από το αποτέλεσμα σύμφωνα με τις δικές του προτιμήσεις;

Τα τελευταία χρόνια, τα ερωτήματα αυτά, απασχολούν τους ερευνητές του πρόσφατα διαμορφωμένου πεδίου της Υπολογιστικής Θεωρίας της Κοινωνικής Επιλογής, το οποίο προσφέρει πολλές προκλήσεις, ειδικά για τις εκλογές επιτροπών. Η σημασία αυτού του πεδίου καθορίζεται ήδη με εφαρμογές σε συστήματα προτάσεων, σε διαδικτυακές μηχανές αναζήτησης για την απόφαση των σελίδων που εμφανίζονται για κάποιο query, κλπ. Έτσι, έχουν γίνει διάφορες θεωρητικές αξιολογήσεις των ποικίλων κανόνων ψηφοφορίας για την εκλογή επιτροπών, καθένας από τους οποίους ικανοποιεί διαφορετικά κριτήρια και σε πολλές περιπτώσεις είναι ιδιαίτερα σχεδιασμένος για συγκεκριμένες εφαρμογές. Η επιλογή ενός καλού κανόνα ψηφοφορίας για την εκλογή επιτροπών δεν είναι τετριμμένη υπόθεση και πριν από αυτή, θα πρέπει να αναρωτηθούμε το είδος της εφαρμογής που θα τον χρησιμοποιήσουμε και αν υπάρχουν απαραίτητες ιδιότητες που θέλουμε η επιτροπή να ικανοποιεί.

Για τον προαναφερθέντα λόγο, έγινε μια ταξινόμηση των κανόνων για την εκλογή επιτροπών, σύμφωνα με τις βασικές ιδιότητες που η επιτροπή θα πρέπει να ικανοποιεί. Αυτές οι ιδιότητες είναι αναλογική αντιπροσώπευση (η σύνθεση της επιτροπής θα πρέπει να αντανακλά σε γενικές γραμμές τη βούληση των ψηφοφόρων), η ποικιλομορφία (η σύνθεση της επιτροπής θα πρέπει να αντανακλά την ποικιλία των ψηφοφόρων) και η δικαιοσύνη (για εφαρμογές εύρεσης των επικρατέστερων υποψηφίων - finalist). Μια ακόμη ιδιότητα που έχει μελετηθεί είναι η αντίσταση στη χειραγώγηση. Έτσι αποφασίσαμε να προβούμε σε μια διεξοδική πειραματική αξιολόγηση των κανόνων ψηφοφορίας για την εκλογή επιτροπών στο δισδιάστατο χώρο, σύμφωνα με αυτές τις ιδιότητες.

Τέλος, σημαντική πρόκληση στον τομέα της Υπολογιστικής Θεωρίας της Κοινωνικής Επιλογής αποτελεί η υπολογιστική πολυπλοκότητα κάποιων κανόνων, οι οποίοι ανήκουν στο NP και επομένως δεν μπορούν να εφαρμοστούν σε μεγάλο αριθμό δεδομένων λόγω περιορισμών χρόνου ή/και χώρου. Πολλοί προσεγγιστικοί και ευρετικοί αλγόριθμοι έχουν προταθεί για τους NP-hard

κανόνες και κατόπιν μελετηθεί σύμφωνα με θεωρητικές και αλγοριθμικές ιδιότητες. Ως εκ τούτου, δημιουργήσαμε τους δικούς μας ευρετικούς αλγορίθμους για να τους αξιολογήσουμε, σύμφωνα με τις ζητούμενες αλγοριθμικές ιδιότητες.

1.2 Συνεισφορά της Εργασίας

Στα πλαίσια αυτής της διπλωματικής εργασίας, αναλύσαμε τους πιο γνωστούς κανόνες ψηφοφορίας. Κατά τη διάρκεια των πειραμάτων μας στο δισδιάστατο χώρο, συγκρίναμε τους κανόνες, τονίσαμε τις ομοιότητες και διαφορές τους ως προς τα χαρακτηριστικά και το σχεδιασμό και τους αξιολογήσαμε σύμφωνα με ιδιότητες όπως η αναλογική αντιπροσωπευτικότητα και η ποικιλομορφία. Ιδιαίτερη βάση δόθηκε στην ιδιότητα της ευστάθειας, καθώς είναι απαραίτητη σε πολλές εφαρμογές. Στα πλαίσια της εργασίας μας, περιορίσαμε τη μελέτη της ευστάθειας στις ακόλουθες δύο περιπτώσεις: τη μετακίνηση ενός ψηφοφόρου και την προσθήκη ενός νέου.

Μερικοί από τους κανόνες που μελετήσαμε, ανήκουν υπολογιστικά στο NP (Chamberlin-Courant και Monroe). Αντί λοιπόν να τρέχουμε μεγάλες και μακροχρόνιες προσομοιώσεις για την αξιολόγηση αυτών των κανόνων, σχεδιάσαμε και μετέπειτα υλοποιήσαμε δικούς μας ευρετικούς αλγορίθμους με έμφαση στην αποτελεσματικότητα. Αυτοί οι ευρετικοί αλγόριθμοι μετέπειτα αξιολογήθηκαν σύμφωνα με την ικανοποίηση των ίδιων ιδιοτήτων και τα αποτελέσματα τους συγκρίθηκαν αναλυτικά με τα αναμενόμενα δίνοντας έμφαση στην ακρίβεια. Ο σχεδιασμός προσεγγιστικών και ευρετικών αλγορίθμων για NP-hard προβλήματα αποτελεί ενεργό και γεμάτο προκλήσεις πεδίο έρευνας, ειδικά για τον κανόνα του Monroe λόγω ιδιαίτερα συγκεκριμένων και ιδιαίτερων απαιτήσεων του.

Τέλος, επεκτείναμε την έρευνα μας στον πολυδιάστατο χώρο, που αντιστοιχεί σε πραγματικά προβλήματα και εφαρμογές. Επιτυχημένα υλοποιήσαμε τους κανόνες ψηφοφορίας και προσαρμόσαμε τους ευρετικούς μας αλγόριθμους ώστε να μπορούν να χρησιμοποιούν πραγματικά δεδομένα. Ο διακριτός πολυδιάστατος χώρος αποτέλεσε πρόκληση, λόγω της έλλειψης σχετικών θέσεων μεταξύ των ψηφοφόρων και υποψηφίων, την οποία ξεπεράσαμε στην εργασία μας. Η επαλήθευση της προσέγγισής μας ήταν επιτυχημένη στα διαθέσιμα δεδομένα. Ωστόσο, η έλλειψη δεδομένων μεγάλης κλίμακας, αποδείχτηκε να είναι περιοριστικός παράγοντας της εργασίας μας. Η συλλογή δεδομένων για την πλήρη επιβεβαίωση και επαλήθευση της προσέγγισης που περιγράφεται σε αυτή την εργασία αποτελεί μέρος ενεργούς έρευνας και σημειώνεται ως μελλοντική έρευνα.

1.3 Οργάνωση Ελληνικού Κειμένου

Στο κεφάλαιο 2, παρουσιάζουμε το θεωρητικό υπόβαθρο στο οποίο βασίζεται η εργασία, διευκρινίζοντας τις έννοιες και τις ορολογίες που εισάγονται αργότερα στην εργασία. Συγκεκριμένα, παρουσιάζουμε έννοιες της υπολογιστικής θεωρίας της κοινωνικής επιλογής, δίνουμε ορισμούς των κανόνων ψηφοφορίας για την εκλογή επιτροπών που χρησιμοποιήσαμε, και περιγράφουμε τις εφαρμογές τους, σε συνδυασμό με τις επιθυμητές ιδιότητες των αποτελεσμάτων.

Στο κεφάλαιο 3, περιγράφουμε την πειραματική διαδικασία και τις σχεδιαστικές αποφάσεις για το πείραμα στο δισδιάστατο χώρο. Στη συνέχεια, συγκρίνουμε και αντιπαραβάλλουμε τους κανόνες ψηφοφορίας εκλογής επιτροπών σύμφωνα με ιδιότητες όπως η ποικιλομορφία και συζητάμε την καταλληλότητα και τα αναμενόμενα προβλήματα από την χρήση κάθε κανόνα ψηφοφορίας.

Παραθέτουμε επίσης γραφικές αναπαραστάσεις και εκτεταμένο σχολιασμό για ορισμένα από τα πειραματικά μας αποτελέσματα.

Στο κεφάλαιο 4 προτείνουμε τους δικούς μας ευρετικούς αλγορίθμους για τους NP-hard κανόνες Chamberlin-Courant και Monroe με βάση τις εγγενείς ιδιότητές των κανόνων. Εξετάζουμε τα παραγόμενα αποτελέσματα και σχολιάζουμε την επιτυχία των αλγορίθμων μας, καθώς και τις αδυναμίες τους.

Στο κεφάλαιο 5, αξιολογούμε τους κανόνες ψηφοφορίας εκλογής επιτροπών σύμφωνα με την ικανότητα τους να βρεθούν σταθερές επιτροπές. Εξετάζουμε αρχικά, εάν το αποτέλεσμα μπορεί να επηρεαστεί από την κίνηση ενός τυχαίου ψηφοφόρου, και στη συνέχεια, από την προσθήκη ενός ψηφοφόρου.

Στο κεφάλαιο 6, περιγράφουμε την πειραματική διαδικασία και τις σχεδιαστικές αποφάσεις του πειράματος για την χρήση πραγματικών δεδομένων από εκλογές.

Στο κεφάλαιο 7, αναφέρουμε βελτιώσεις στους ευρετικούς αλγορίθμους, προτείνουμε διαφορετικές προσεγγίσεις που μπορούν να υλοποιηθούν και να προσφέρουν κέρδη όσον αφορά την απόδοση των αλγορίθμων. Τέλος, παρουσιάζουμε πιθανές προεκτάσεις της μελέτης της ευστάθειας και μελλοντικές κατευθύνσεις σχετικής έρευνας.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Εισαγωγή

Όλοι μας, κάποια στιγμή στη ζωή μας, έχουμε λάβει μέρος σε κάποια εκλογική διαδικασία, είτε ως ψηφοφόροι είτε ως υποψήφιοι. Οι βουλευτικές εκλογές είναι ένα χαρακτηριστικό παράδειγμα και ίσως το πρώτο που περνάει από το μυαλό κάποιου ακούγοντας τον όρο «εκλογές». Ωστόσο υπάρχουν πολλά ακόμα παραδείγματα από την καθημερινή ζωή, όπως η εύρεση της ιδανικής μέρας και ώρας για κάποια συνάντηση ή η εύρεση του νικητή σε κάποιο διαγωνισμό, π.χ. Eurovision. Καθένα από τα παραπάνω παραδείγματα αποτελεί μέλος μιας διαφορετικής κατηγορίας εκλογών, λόγω του διαφορετικού κανόνα ψηφοφορίας που χρησιμοποιείται και των επιθυμητών ιδιοτήτων του αποτελέσματος. Στα περισσότερα πολιτικά συστήματα, εκλέγουμε εκπροσώπους με τον κανόνα Single Nontransferable Vote (απλή αναλογική), ενώ στο Doodle χρησιμοποιείται ο κανόνας approval [1]. Στο διαγωνισμό τραγουδιού της Eurovision, εφαρμόζεται μία παραλλαγή του κανόνα ψηφοφορίας Borda, για την επιλογή του νικητή.

Στα παραπάνω παραδείγματα, ακόμη και αν η φύση τους διαφέρει, ο στόχος παραμένει ο ίδιος: «εύρεση (επιλογή) του νικητή». Οι κανόνες ψηφοφορίας που χρησιμοποιούνται στα παραπάνω παραδείγματα αποτελούν κανόνες ψηφοφορίας για την εύρεση νικητή (Single Winner Voting Rules). Τούτου λεχθέντος, υπάρχει και μια άλλη οικογένεια εκλογών, με στόχο την εκλογή μιας ομάδας υποψηφίων, δηλαδή μιας επιτροπής. Αυτοί είναι οι κανόνες ψηφοφορίας για την εκλογή επιτροπών (Multi Winner Voting Rules) και αντιστοιχούν σε πιο περίπλοκες και ποικίλες καταστάσεις και εφαρμογές. Οι βουλευτικές εκλογές, διάφορες επιχειρηματικές αποφάσεις, δημιουργία portfolio, επιλογή ταινιών, συστήματα συστάσεων, εύρεση finalist είναι μόνο μερικά σενάρια εκλογών για επιτροπές και απαιτούν κανόνες με διαφορετικές αρχές.

Έτσι, οι εκλογές επιτροπών έχουν χωριστεί σε τρεις κύριες κατηγορίες [3] με βάση το στόχο τους:

1. εύρεση επιτροπής με τους επικρατέστερους υποψηφίους (finalist),
2. εύρεση επιτροπής με γνώμονα την ποικιλομορφία,
3. εύρεση επιτροπής με γνώμονα την αναλογική αντιπροσώπευση των ψηφοφόρων.

Στις επόμενες ενότητες, θα αναλύσουμε τις παραπάνω κατηγορίες, δίνοντας παραδείγματα και σχολιάζοντας τις προκλήσεις που υπάρχουν στις εκλογές επιτροπών. Επίσης θα παρουσιάσουμε τους κανόνες που χρησιμοποιήσαμε στα πειράματά μας.

2.2 Κύριοι Τύποι Εκλογών Επιτροπών

2.2.1 Εύρεση Επιτροπής με Γνώμονα την Εύρεση των Επικρατέστερων Υποψηφίων (shortlisting)

Πριν την εύρεση του νικητή ενός διαγωνισμού, του παραλήπτη ενός βραβείου, την επιλογή του ιδανικού υποψήφιου για μία θέση εργασίας κλπ., υπάρχει συνήθως μία διαδικασία προεπιλογής για την εύρεση των επικρατέστερων υποψηφίων (finalist). Οι κανόνες αυτοί (ονομάζονται shortlisting ή screening rules [4]), μπορούν να οριστούν ως εκλογές επιτροπών, που έχουν ως στόχο την επακόλουθη επιλογή ενός υποψηφίου. Δεδομένου ότι μόνο ένας υποψήφιος θα επιλεγεί τελικά, τα άτομα που θα εκλεγούν, θα πρέπει να είναι τα καλύτερα, με την υψηλότερη καταλληλότητα, ανεξάρτητα από αλληλεπιδράσεις μεταξύ των υποψηφίων (υποθέτουμε ότι οι ψηφοφόροι κατανοούν τυχόν ειδικές απαιτήσεις, π.χ. ένα επίπεδο ποικιλομορφίας προκειμένου να αποφευχθούν οι διακρίσεις στην ομάδα των finalist, και ότι αυτές αντικατοπτρίζονται στις προτιμήσεις των ψηφοφόρων) [5], δηλαδή δύο παρόμοιοι υποψήφιοι πρέπει είτε να εκλεγούν είτε να εξαιρεθούν¹. Στη συνέχεια οι finalist αξιολογούνται αυστηρότερα, π.χ. με μια συνέντευξη, και επιλέγεται ο νικητής, διαδικασία που είναι πέρα από την διαδικασία του shortlisting.

Για την κατηγορία της εύρεσης των επικρατέστερων υποψηφίων, δίνουμε ένα παράδειγμα από τον ακαδημαϊκό χώρο. Ας εξετάσουμε μία κατάσταση, όπου μία μεταδιδακτορική θέση ανοίγει σε ένα πανεπιστήμιο. Τα κατάλληλα μέλη από το πανεπιστήμιο κατατάσσουν τους αιτούντες, προκειμένου να δημιουργηθεί ένας κατάλογος (λίστα) των υποψηφίων, οι οποίοι στη συνέχεια θα κληθούν για συνέντευξη [3]. Ένα άλλο παράδειγμα, είναι η διαδικασία της επιλογής των υποψηφίων για κάποιο βραβείο Όσκαρ. Η απλουστευμένη διαδικασία είναι η ακόλουθη: υπάρχουν ορισμένα κριτήρια που κάποιος πιθανός υποψήφιος πρέπει να πληροί και έπειτα από μια διαδικασία ψηφοφορίας επιλέγονται οι πιθανοί υποψήφιοι που τελικά ορίζονται ως επίσημοι υποψήφιοι για το βραβείο. Τέλος, ο νικητής αποφασίζεται, μέσω μίας νέας διαδικασίας ψηφοφορίας, ως υποψήφιος με τους περισσότερους ψήφους, διαδικασία που όπως αναφέρθηκε ξεφεύγει από το shortlisting.

2.2.2 Επιλογή Επιτροπής με Γνώμονα την Ύπαρξη Ποικιλομορφίας (diversity)

Υπάρχουν πολλά παραδείγματα εκλογών, στις οποίες πρωταρχικά αποζητούμε την ποικιλομορφία στην εκλεγμένη επιτροπή. Σε αυτές, οι υποψήφιοι δεν μπορούν να αξιολογηθούν ανεξάρτητα, δηλαδή εάν υπάρχουν δύο παρόμοιοι υποψήφιοι τότε πρέπει είτε να επιλεγεί ένας από αυτούς είτε κανένας (αν υπάρχουν καλύτερες επιλογές), αλλά σε καμία περίπτωση δεν πρέπει να επιλεγούν και οι δύο.

Το πρόβλημα της εκλογής επιτροπής με ποικιλομορφία θυμίζει το Facility Location Problem. Ας εξετάσουμε λοιπόν, το ακόλουθο πρόβλημα. Θέλουμε να τοποθετήσουμε πυροσβεστικούς σταθμούς στην πόλη. Παρόλο που υπάρχει ένα σημείο που ελαχιστοποιεί την απόσταση από το σύνολο των σημείων στην πόλη, δεν θέλουμε να οικοδομήσουμε όλους τους πυροσβεστικούς σταθμούς σε αυτό. Αντίθετα προτιμούμε να τα διανείμουμε ομοιόμορφα, έτσι ώστε κάθε σημείο της πόλης να βρίσκεται κοντά σε έναν πυροσβεστικό σταθμό. Ένα άλλο παράδειγμα από τον τομέα των επιχειρηματικών αποφάσεων, είναι το ακόλουθο. Έστω ότι ο ιδιοκτήτης ενός καταστήματος με περιορισμένο χώρο αποθήκευσης για DVD, πρέπει να προμηθευτεί ένα συγκεκριμένο αριθμό διαφορετικών

¹ Πιθανή εξαίρεση αποτελούν οι συνοριακές περιπτώσεις, τις οποίες θα εξετάσουμε στις επόμενες ενότητες.

ταινιών. Μια από τις καλύτερες στρατηγικές είναι να επιλέξει ένα σύνολο ταινιών με τη μεγαλύτερη δυνατή ποικιλομορφία, έχοντας κατά νου ότι κάθε πελάτης πρέπει να μπορεί να βρει κάτι που του αρέσει.

2.2.3 Επιλογή Επιτροπής με Γνώμονα την Αναλογική Αντιπροσώπευση των Ψηφοφόρων (proportional representation)

Ο πιο γνωστός τύπος εκλογών για τη εύρεση επιτροπών είναι οι βουλευτικές εκλογές, όπου σημαντικό αίτημα είναι η αναλογική αντιπροσώπευση των απόψεων της κοινωνίας (δηλαδή των ψηφοφόρων) και ιδανικά η εύρεση μίας επιτροπής k -αντιπροσώπων, όπου καθένας από αυτούς αντιπροσωπεύει περίπου n/k ψηφοφόρους (όπου n είναι ο συνολικός αριθμός των ψηφοφόρων). Αυτός είναι άλλωστε και ο αυστηρός ορισμός της ιδιότητας αναλογικής αντιπροσωπευτικότητας.

Για την ιδιότητα της αναλογικής αντιπροσωπευτικότητας, δίνουμε ένα παράδειγμα, παρόμοιο με αυτό της ποικιλομορφίας. Ας εξετάσουμε και πάλι την περίπτωση του καταστήματος με τον περιορισμένο χώρο αποθήκευσης DVD. Ας υποθέσουμε ότι υπάρχει ένας συγκεκριμένος αριθμός διαφορετικών ταινιών που οι ιδιοκτήτες μπορούν να προμηθευτούν, και η κάθε διαφορετική ταινία μπορεί να αγοραστεί μόνο σε ομάδες των n/k . Η καλύτερη στρατηγική για να μεγιστοποιήσουμε το κέρδος είναι η προμήθεια των k ταινιών που περίπου n/k πελάτες θα βρουν ελκυστικά, έχοντας κατά νου ότι κάθε πελάτης πρέπει να βρει ένα DVD για να αγοράσει.

2.3 Προκλήσεις

Θα επικεντρωθούμε στις ακόλουθες προκλήσεις που αφορούν τις εκλογές επιτροπών [5]. Η πρώτη αφορά το πρόβλημα της επιλογής του κατάλληλου κανόνα ψηφοφορίας σε μία συγκεκριμένη εφαρμογή (ή γενικότερα για έναν τύπο εκλογών). Πώς μπορεί, λοιπόν, ένας χρήστης να επιλέξει έναν καλό κανόνα ψηφοφορίας για την εφαρμογή του; Υπάρχουν πολλές απαντήσεις στην ερώτηση αυτή. Αρχικά, υπάρχουν ορισμένοι κανόνες ψηφοφορίας οι οποίοι έχουν σχεδιαστεί για να ικανοποιούν ορισμένες ιδιότητες (π.χ. κανόνας Monroe και αναλογική αντιπροσωπευτικότητα). Μία διαφορετική προσέγγιση, αποτελεί η μελέτη αξιωματικών ιδιοτήτων για να κρίνουμε την καταλληλότητα ενός κανόνα ψηφοφορίας για μια συγκεκριμένη εφαρμογή [3] (π.χ. έλεγχο για την ικανοποίηση της ιδιότητας της αναλογικής αντιπροσωπευτικότητας μέσω της ιδιότητας του Dummett [6]). Τέλος, μπορούμε να χρησιμοποιήσουμε εμπειρική ανάλυση για τη σύγκριση διαφορετικών κανόνων υπό συγκεκριμένες συνθήκες, συμπεριλαμβανομένων θεωρητικών και πειραματικών αξιολογήσεων [7], [8], ή μπορούμε να εξετάσουμε την συμπεριφορά των κανόνων σε απλούστερους τομείς, όπου μπορεί η συμπεριφορά τους να ερμηνεύεται διαισθητικά [9]. Όλες οι παραπάνω προσεγγίσεις είναι σημαντικές και η επιλογή ενός κανόνα ψηφοφορίας θα πρέπει να γίνεται λαμβάνοντας τες υπόψιν.

Η δεύτερη αφορά την ικανότητά μας να υπολογίζουμε τα αποτελέσματα των ψηφοφοριών σε πολυωνυμικό χρόνο. Από τους κανόνες για την εύρεση νικητή, σχεδόν όλοι υπολογιστικά ανήκουν στο P, αν και υπάρχουν σημαντικές εξαιρέσεις, όπως οι κανόνες του Dodgson [10], Young [11] και Kemeny [12]. Για την εύρεση επιτροπής, η κατάσταση είναι πιο περίπλοκη. Υπάρχουν πολλοί κανόνες που τα αποτελέσματα υπολογίζονται σε πολυωνυμικό χρόνο, αλλά οι πιο ενδιαφέροντες κανόνες είναι NP-hard, π.χ. Chamberlin-Courant [13] και Monroe [14]. Το πρόβλημα αυτό μπορούμε

να το προσεγγίσουμε είτε με γρήγορους ευρετικούς αλγορίθμους είτε με αλγόριθμους προσέγγισης. Τέτοιοι αλγόριθμοι μπορούν να θεωρηθούν ως νέοι, εύκολοι υπολογιστικά κανόνες. Έτσι μελετάμε τις αξιωματικές ιδιότητες των κανόνων που ορίζονται από τέτοιους αλγόριθμους προσέγγισης, ακριβώς όπως κάνουμε και για τους αρχικούς κανόνες ψηφοφορίας.

2.4 Ορισμός των Κανόνων Ψηφοφορίας για την Εκλογή Επιτροπών

Σε αυτή τη διπλωματική εργασία, εφαρμόζουμε τους πιο δημοφιλείς κανόνες ψηφοφορίας για την εκλογή επιτροπών π.χ., Single Nontransferable Vote, Single Transferable Vote, Bloc, k -Borda, Monroe, Chamberlin–Courant. Στην ενότητα αυτή, παρουσιάζουμε τους αλγορίθμους και τις έννοιες πίσω από κάθε κανόνα.

2.4.1 Κανόνας Single Nontransferable Vote (SNTV)

Ο κανόνας αποτελεί την έκδοση του κανόνα της πλειοψηφίας για την εκλογή επιτροπών. Κάθε ψηφοφόρος δίνει ένα ψήφο στον αγαπημένο του υποψήφιο και τελικά οι k υποψήφιοι με τους περισσότερους ψήφους εκλέγονται.

Παράδειγμα. Ας εξετάσουμε το ακόλουθο απλό παράδειγμα. Έχουμε 6 ψηφοφόρους και ένα σύνολο υποψηφίων $C = \{a, b, c, d, e\}$. Στόχος μας είναι να εκλέξουμε μια επιτροπή με $k = 2$ μέλη.

Ψηφοφόρος	Σειρά Κατάταξης
v_1	: $b \succ a \succ c \succ d \succ e$
v_2	: $e \succ a \succ b \succ d \succ c$
v_3	: $d \succ a \succ b \succ c \succ e$
v_4	: $c \succ b \succ d \succ e \succ a$
v_5	: $c \succ b \succ e \succ a \succ d$
v_6	: $b \succ d \succ c \succ e \succ a$

Πίνακας 2.1: Οι προτιμήσεις των ψηφοφόρων στο σύνολο των υποψηφίων

Για τον κανόνα SNTV μόνο η πρώτη προτίμηση των ψηφοφόρων έχει σημασία. Ως αποτέλεσμα, ο ψηφοφόρος v_1 δίνει ένα ψήφο στον υποψήφιο b , ο v_2 μία ψήφο στον υποψήφιο a , κλπ. Τα αποτελέσματα παρουσιάζονται στον πίνακα 2.2 και τελικά προκύπτει η επιτροπή $\{b, c\}$.

	Υποψήφιοι				
	a	b	c	d	e
Συνολικοί Ψήφοι	0	2	2	1	1

Πίνακας 2.2: Αποτελέσματα του κανόνα SNTV για το παράδειγμα του πίνακα 2.1

2.4.2 Κανόνας Single Transferable Vote (STV)

Ο κανόνας χωρίζεται σε πολλούς γύρους εκτέλεσης. Σε κάθε γύρο, οι ψήφοι αποδίδονται στους υποψηφίους σύμφωνα με την πρώτη προτίμηση των ψηφοφόρων. Εάν υπάρχουν υποψήφιοι, οι οποίοι συγκεντρώνουν τουλάχιστον ίσο αριθμό ψήφων με κάποιο συγκεκριμένο αριθμό, τότε αυτοί εκλέγονται στην επιτροπή και όλοι οι επιπλέον ψήφοι τους μεταφέρονται από τους νικητές² στους

² Ως νικητές ορίζονται οι υποψήφιοι που έχουν εκλεχθεί.

εναπομείναντες³ υποψηφίους. Διαφορετικά, ένας ή περισσότεροι υποψήφιοι με τους λιγότερους ψήφους αφαιρούνται από τη διαδικασία και οι ψήφοι τους μεταφέρονται στους εναπομείναντες υποψηφίους⁴. Η διαδικασία αυτή επαναλαμβάνεται, έως ότου να συμπληρωθούν όλες οι θέσεις της επιτροπής.

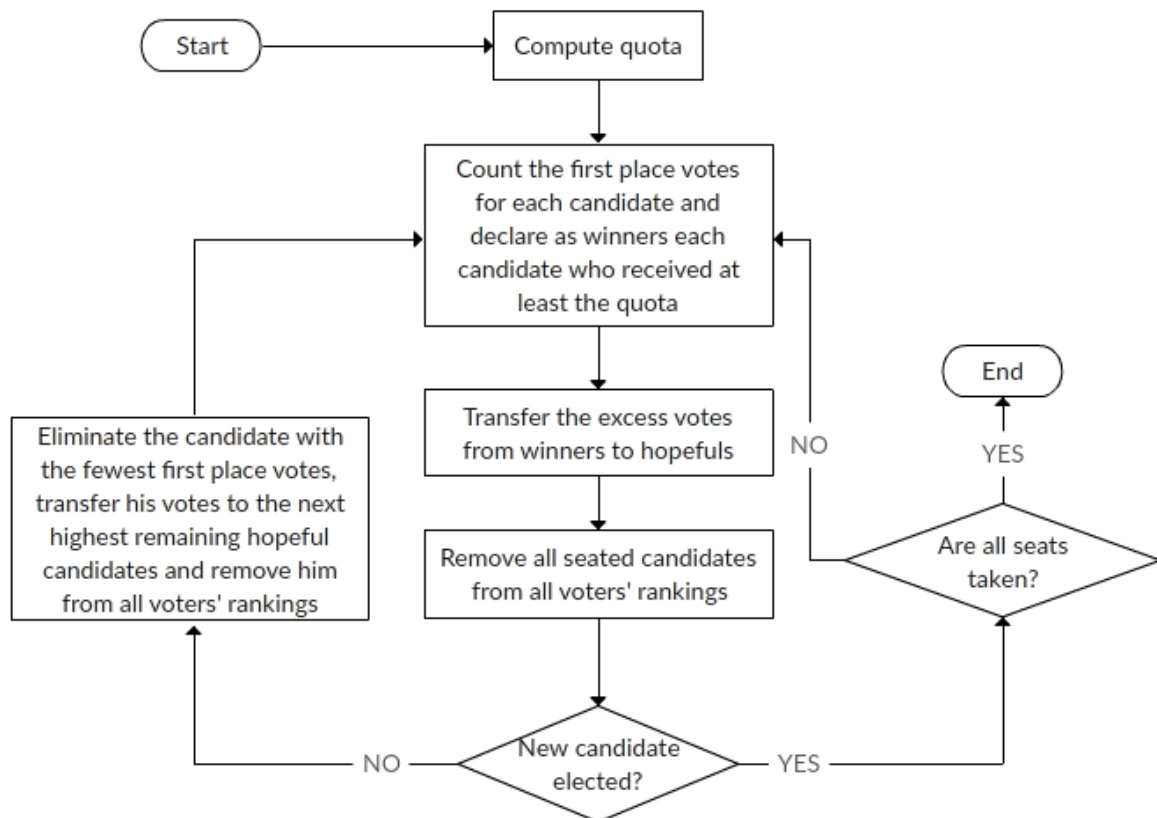
Όπως αναφέρθηκε παραπάνω, οποιοσδήποτε υποψήφιος, προκειμένου να εκλεγεί, πρέπει να έχει ένα ελάχιστο αριθμό ψήφων (quota). Συγκεκριμένα,

$$quota = \text{floor} \left(\frac{\text{valid votes to cast}}{\text{seats to fill} + 1} \right) + 1$$

Οι επιπλέον ψήφοι μεταφέρονται από τους νικητές στους εναπομείναντες σύμφωνα στον ακόλουθο τύπο:

$$\left(\frac{\text{voted for second preference}}{\text{votes of the winner}} \right) * \text{extra votes}$$

Για τον κανόνα STV υπάρχουν διαφορετικές παραλλαγές με περίπλοκους μηχανισμούς [15]. Ο ακριβής αλγόριθμος που χρησιμοποιήσαμε παρουσιάζεται στο σχήμα 2.1.



Εικόνα 2.1: Κανόνας Single Transferable Vote [16]

Όταν ένας υποψήφιος αφαιρείται, οι ψήφοι του (πρώτης προτίμησης ή αυτοί που αποκτήθηκαν από την εκλογή ή την κατάργηση άλλων υποψηφίων) μεταφέρονται στους υποψηφίους που αποτελούν την επόμενη προτίμηση. Η διαδικασία αυτή επαναλαμβάνεται έως ότου να βρεθεί ένας νικητής για κάθε θέση ή ο αριθμός των υποψηφίων να ισούται με τις ελεύθερες θέσεις ή ο συνολικός αριθμός

³ Οι υποψήφιοι που δεν έχουν εκλεγεί ή αφαιρεθεί από τη διαδικασία (hopefuls).

⁴ Σε αυτό το σημείο αγνοούμε τη διαδικασία με την οποία μεταβιβάζονται οι ψήφοι και τον τρόπο χειρισμού των ισοπαλιών στη διαδικασία αφαίρεσης υποψηφίων. Για περισσότερες πληροφορίες, ανατρέξτε στην ενότητα 3.2.

ψηφών να είναι μικρότερος από το *quota*.

Παράδειγμα. Παραθέτουμε ένα παράδειγμα, όπου έχουμε τις προτιμήσεις 99995 ψηφοφόρων [16] και ζητούμενο είναι η εκλογή μίας επιτροπής τεσσάρων μελών. Στον πίνακα 2.3 η πρώτη γραμμή δηλώνει τον δείκτη της στήλης για τη συγκεκριμένη σειρά προτιμήσεων. Στη δεύτερη γραμμή υπάρχει ο αριθμός των ψηφοφόρων με αυτή τη σειρά προτιμήσεων. Στην τρίτη φαίνεται η πρώτη προτίμηση κάθε ομάδας ψηφοφόρων. Οι επόμενες προτιμήσεις βρίσκονται σε φθίνουσα σειρά στις επόμενες γραμμές.

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII
4111	3500	100	2722	10490	10511	13936	1595	8000	7639	4000	12051	4095	3153	2422	1020	2150	8500
A	A	A	A	B	B	C	D	D	E	E	F	G	G	G	H	H	H
C	D	D	H	A	C	H	C	G	F	H	B	A	A	A	C	E	F
H	G	G	C	C	H	D	H	E	D	B	H	C	D	H	D	G	E
E	C	C	G	H	F	F	F	H	H	C	G	H	H	E	E	C	G
G	H	H	F	E	G	E	B	B	B	G	E	B	C	F	G	F	A
D	E	B	D	G	A	A	E	F	G	D	C	D	E	B	B	B	C
F	B	F	E	F	E	B	G	C	A	A	A	F	B	D	A	A	D
B	F	E	B	D	D	G	A	A	C	F	D	E	F	C	F	D	B

Πίνακας 2.3: Σειρές κατάταξης των υποψηφίων από τους 99995 ψηφοφόρους.

Η έντονη γραφή δηλώνει ότι η συγκεκριμένη θέση έχει ληφθεί υπόψιν από τον κανόνα STV

Στο παράδειγμα αυτό ισχύει $quota = 20000$. Ο υποψήφιος c_B έχει 21001 ψήφους πρώτης προτίμησης, υπερβαίνοντας το *quota* κατά 1001 ψήφους. Ένα κλάσμα $20000/21001$ ή 0.952 από κάθε ψήφο, αρκεί για να εκλεχθεί ο υποψήφιος c_B . Το υπόλοιπο ποσοστό του κάθε ψήφου διανέμεται στους υποψηφίους c_A και c_C , καθώς έχουν τη δεύτερη θέση στις προτιμήσεις των ψηφοφόρων που συμμετείχαν στην εκλογή του c_B . Στον πίνακα 2.4 παρουσιάζεται ο αριθμός των ψήφων που έχει κάθε υποψήφιος στον κάθε γύρο των εκλογών. Οι αριθμοί με έντονους χαρακτήρες δηλώνουν την εκλογή του υποψηφίου, ενώ οι υπογραμμισμένοι την κατάργησή του. Τελικά προκύπτει η ακόλουθη επιτροπή $[B, C, G, H]$.

Γύρος	A	B	C	D	E	F	G	H
1	10433	21001	13936	9595	11639	12051	9670	11670
2	10933	-	14437	<u>9595</u>	11639	12051	9670	11670
3	<u>10933</u>	-	16032	-	11639	12051	17670	11670
4	-	-	20643	-	11639	12051	21270	14392
5	-	-	-	-	12116	<u>12051</u>	-	14828
6	-	-	-	-	12116	-	-	27879
7	-	-	-	-	19995	-	-	-

Πίνακας 2.4: Οι ψήφοι κάθε υποψηφίου σε κάθε γύρο του κανόνα STV

2.4.3 Κανόνας Bloc

Στον κανόνα αυτό οι ψηφοφόροι δίνουν ένα ψήφο στους k προτιμότερους υποψηφίους. Τελικά οι k υποψήφιοι με τους περισσότερους ψήφους δηλώνονται νικητές και εκλέγονται στην επιτροπή.

Παράδειγμα. Ας εξετάσουμε και πάλι το παράδειγμα του πίνακα 2.1. Δεδομένου ότι η εκλεγμένη επιτροπή θα έχει δύο μέλη, κάθε ψηφοφόρος επιλέγει τους δύο υποψηφίους που προτιμάει και δίνει μία ψήφο στον καθένα από αυτούς. Δηλαδή, ο ψηφοφόρος v_1 δίνει μία ψήφο στον υποψήφιο b και

μία στον υποψήφιο a , ο ψηφοφόρος v_2 δίνει μία ψήφο στον υποψήφιο e και μία στον υποψήφιο a , κλπ. Τα αποτελέσματα παρουσιάζονται στον πίνακα 2.5 και τελικά προκύπτει η επιτροπή $\{a, b\}$.

	Υποψήφιοι				
	a	b	c	d	e
Συνολικοί Ψήφοι	3	4	2	2	1

Πίνακας 2.5: Αποτελέσματα του κανόνα Bloc για το παράδειγμα του πίνακα 2.1

2.4.4 Κανόννας k-Borda

Στον κανόνα αυτό, που προτάθηκε από τον Borda, κάθε ψηφοφόρος τοποθετεί τους υποψηφίους σε μία λίστα σύμφωνα με φθίνουσα σειρά προτίμησης. Οι υποψήφιοι λαμβάνουν από κάθε ψηφοφόρο, αριθμό ψήφων ίσο με τον αριθμό των υποψηφίων που κατατάχθηκαν κάτω από αυτόν⁵. Οι k υποψήφιοι με το υψηλότερο Borda score είναι οι νικητές.

Παράδειγμα. Ας εξετάσουμε και πάλι το παράδειγμα του πίνακα 2.1. Ο ψηφοφόρος v_1 δίνει τέσσερις ψήφους στον υποψήφιο b , τρεις ψήφους στον υποψήφιο a , δύο ψήφους στον υποψήφιο d και τέλος ένα ψήφο στον υποψήφιο e , κλπ. Τα αποτελέσματα βρίσκονται στον πίνακα 2.6 και η εκλεγμένη επιτροπή είναι $\{b, c\}$.

	Υποψήφιοι				
	a	b	c	d	e
Συνολικοί Ψήφοι	10	18	13	11	8

Πίνακας 2.6: Αποτελέσματα του κανόνα k -Borda για το παράδειγμα του πίνακα 2.1

2.4.5 Κανόννας Chamberlin-Courant (β -CC)

Για κάθε ψηφοφόρο i και κάθε υποψήφιο c υπάρχει μία τιμή μη αντιπροσώπευσης $\mu_{i,c}$, που δηλώνει το βαθμό στον οποίο κάθε υποψήφιος c δεν αντιπροσωπεύει τον ψηφοφόρο i . Μία συνάρτηση μη αντιπροσώπευσης που βασίζεται στη θέση, κάνει χρήση ενός διανύσματος $s = \langle s_1, \dots, s_p \rangle$ έτσι ώστε να ισχύει $s_1 \leq \dots \leq s_p$. Συγκεκριμένα, ένα τέτοιο διάνυσμα είναι αυτό του κανόνα Borda s_B όπου ισχύει $s_k = k$ για κάθε k . Με το $pos_i(c)$ συμβολίζουμε τη θέση που κατέχει ο υποψήφιος c στη σειρά προτίμησης του ψηφοφόρου i (όπου 1 είναι ο πρώτος υποψήφιος στη σειρά προτίμησης και p ο τελευταίος). Η συνάρτηση μη αντιπροσώπευσης που αντιστοιχεί στο s είναι $\mu_{i,c} = s_{pos_i(c)}$. Ενστικτωδώς, το s_i ερμηνεύεται ως ο βαθμός μη ικανοποίησης του ψηφοφόρου αν εκπροσωπηθεί από έναν υποψήφιο που έχει τοποθετήσει στην θέση i .

Έστω μία συνάρτηση αντιστοίχισης π συνδέει κάθε ψηφοφόρο με έναν υποψήφιο στο υποσύνολο S . Η μη αντιπροσώπευση του ψηφοφόρου i σύμφωνα με την π είναι $\mu_{i,\pi(i)}$. Αφότου οριστεί η μη αντιπροσώπευση του κάθε ψηφοφόρου, πρέπει να οριστεί και η ολική μη αντιπροσώπευση της κοινωνίας (utilitarianism), δηλαδή το άθροισμα όλων των ατομικών βαθμών μη αντιπροσώπευσης, όταν επιλεγεί ένα υποσύνολο S των υποψηφίων. Επίσημα, η ολική μη αντιπροσώπευση υπό την

⁵ Αυτό είναι γνωστό ως asymmetric Borda score. Υπάρχει και symmetric Borda score, όπου οι ψήφοι στους n υποψηφίους δίνονται ως $\frac{n}{2}, \frac{n-1}{2}, \dots, 0, \dots, -\frac{n-1}{2}, -\frac{n}{2}$ ανάλογα με τη σειρά κατάταξης. Ανεξάρτητα από ποια από τις δύο συναρτήσεις θα χρησιμοποιήσουμε, το αποτέλεσμα παραμένει το ίδιο.

αντιστοίχιση π ορίζεται ως:

$$\mu_U(\pi) = \sum_{i \leq \pi} \mu_{i, \pi(i)}$$

Ο κανόνας Chamberlin-Courant [13] δίνει ως έξοδο, την επιτροπή μεγέθους k που ελαχιστοποιεί μ_U . Επειδή δεν υπάρχει περιορισμός στην συνάρτηση αντιστοίχισης, κάθε ψηφοφόρος αντιστοιχίζεται απλώς στον προτιμότερο υποψήφιο στο υποσύνολο S . Για αυτό, $\pi(i) = \arg \min_{c \in S} \mu_{i,c}$. Τότε, η μη εκπροσώπηση από την επιλογή του εφικτού υποσυνόλου S ισούται με το $\mu_{i,S} = \min_{c \in S} \mu_{i,c}$. Η καλύτερη επιτροπή είναι λοιπόν το εφικτό υποσύνολο S που ελαχιστοποιεί το $\mu_U(\pi)$ (utilitarianism) [17].

Ένας απλοποιημένος ορισμός για τον κανόνα αυτό είναι ο ακόλουθος: «Κάθε ψηφοφόρος δηλώνει κατά πόσο δεν τον εκπροσωπεί ένας υποψήφιος. Κάθε ψηφοφόρος ορίζεται πως εκπροσωπείται από τον υποψήφιο που προτιμάει περισσότερο σε κάθε επιτροπή. Νικητές ορίζονται οι υποψήφιοι που ελαχιστοποιούν τη μη αντιπροσώπηση των ψηφοφόρων.»

2.4.6 Κανόνας Monroe (β -Monroe)

Στον κανόνα Monroe [14], εκτός από τις απαιτήσεις του κανόνα Chamberlin-Courant, υπάρχει επιπλέον η απαίτηση για την ισοζυγισμένη αντιστοίχιση π του κάθε ψηφοφόρου με τα μέλη της επιτροπής: σε κάθε υποψήφιο στο S πρέπει να αντιστοιχιστούν τουλάχιστον n/k ψηφοφόροι. Επίσημα, ο κανόνας Monroe, επιλέγει την αντιστοίχιση π που ελαχιστοποιεί το $\mu_U(\pi)$ σύμφωνα με τους περιορισμούς $|\pi^{-1}(s)| \geq \lfloor \frac{n}{k} \rfloor$ για όλα τα $s \in \text{Range}(\pi)$ [17].

Ένας απλοποιημένος ορισμός για τον κανόνα είναι ο ακόλουθος: «Κάθε ψηφοφόρος δηλώνει κατά πόσο δεν τον εκπροσωπεί ένας υποψήφιος. Κάθε ψηφοφόρος ορίζεται πως εκπροσωπείται από τον υποψήφιο που προτιμάει περισσότερο σε κάθε δυνατή επιτροπή, με τέτοιο τρόπο που τα μέλη της επιτροπής να εκπροσωπούν ίσο αριθμό ψηφοφόρων. Νικητές ορίζονται οι υποψήφιοι που ελαχιστοποιούν τη μη αντιπροσώπηση των ψηφοφόρων.»

Κεφάλαιο 3

Σχεδιασμός & Υλοποίηση του Πειράματος στο Δισδιάστατο Χώρο

Στην προσπάθειά μας να αποκτήσουμε καλύτερη κατανόηση μερικών από των πιο διάσημων κανόνων εκλογής επιτροπών, εκτελέσαμε μια διεξοδική πειραματική μελέτη, αρχικά στο δισδιάστατο χώρο. Ένα πείραμα στο δισδιάστατο χώρο [9] μας βοηθάει να οπτικοποιήσουμε τα αποτελέσματα των κανόνων ψηφοφορίας, αλλά και γενικότερα τις προτιμήσεις των ψηφοφόρων.

3.1 Πείραμα

3.1.1 Πειραματική Διαδικασία

Σε αυτά τα πειράματα, σκοπός μας είναι να εκλέξουμε μία επιτροπή (ένα υποσύνολο των υποψηφίων), που εκπροσωπεί τους ψηφοφόρους με συγκεκριμένο τρόπο. Όπως περιγράφηκε στην προηγούμενη ενότητα, σε κάθε εφαρμογή θα έπρεπε να χρησιμοποιείται διαφορετικός κανόνας, καθώς ζητούμενο είναι κάθε φορά το αποτέλεσμα να ικανοποιεί διαφορετικές ιδιότητες που εξαρτώνται από το είδος της εφαρμογής. Το δισδιάστατο επίπεδο μας βοηθάει να οπτικοποιήσουμε το αποτέλεσμα με έναν ξεκάθαρο και εύληπτο τρόπο, και έτσι να εντοπίσουμε πιθανές αδυναμίες του κανόνα, καθώς και ιδιότητες που ικανοποιούνται (π.χ. αναλογική αντιπροσωπευτικότητα). Μπορούμε λοιπόν, να συγκρίνουμε τους κανόνες και να αποφασίσουμε ποιος είναι καταλληλότερος για κάθε εφαρμογή, ανάλογα με τις ιδιότητες που εξασφαλίζουν πως ικανοποιεί το αποτέλεσμα.

Για κάθε προσομοίωση εκλογών, ένα σύνολο από 1000 ψηφοφόρους και 1000 υποψήφιους παράγονται, ακολουθώντας συγκεκριμένες κατανομές, ως σημεία στον ευκλείδειο χώρο. Για κάθε ψηφοφόρο υπάρχει μία φθίνουσα σειρά κατάταξης των υποψηφίων σύμφωνα με τη σειρά προτίμησης και τελικά μία επιτροπή 20 ατόμων εκλέγεται από κάθε κανόνα. Στη συνέχεια, συγκρίνουμε τα αποτελέσματα και αξιολογούμε τους κανόνες σύμφωνα με αυτά.

Ο αριθμός των ψηφοφόρων και υποψηφίων είναι στοχευμένα ο ίδιος και αρκετά μεγάλος για τους ακόλουθους λόγους.

1. Ο αριθμός των υποψηφίων πρέπει να είναι αρκετά μεγάλος για να μελετήσουμε τις αδυναμίες των κανόνων.
2. Η επιτροπή πρέπει να έχει αρκετά μέλη για μελετήσουμε ιδιότητες όπως η αναλογική αντιπροσωπευτικότητα ή η ποικιλομορφία. Για να έχει νόημα η επιτροπή 20 ατόμων λοιπόν, απαραίτητος είναι ο αρκετά μεγαλύτερος ο αριθμός των υποψηφίων.
3. Ο αριθμός των ψηφοφόρων πρέπει να είναι αρκετά μεγάλος για να έχει νόημα η μελέτη NP-hard κανόνων όπως ο Chamberlin-Courant και ο Monroe.

3.1.2 Κατανομή των Ψηφοφόρων & Υποψηφίων στο Χώρο

Οι ψηφοφόροι, όπως και οι υποψήφιοι, παράγονται ως σημεία στον ευκλείδειο χώρο σύμφωνα με συγκεκριμένες κατανομές. Καθένας από αυτούς, λοιπόν, αναπαρίσταται από ένα ζεύγος (x, y) , τις συντεταγμένες του σημείου, που ορίζουν την ακριβή του θέση στο δισδιάστατο χώρο. Σε αυτό το πείραμα, στοχεύουμε στην κατανόηση της συμπεριφοράς των κανόνων ανάλογα με τις κατανομές που χρησιμοποιήθηκαν - γκαουσιανή κατανομή (με διαφορετικές τιμές διασποράς και μέσης τιμής), ομοιόμορφη κατανομή, σε μορφή κυκλικού δίσκου ή τετραγώνου, καθώς και συνδυασμοί τους.

Μελετήσαμε ποικίλες περιπτώσεις, συμπεριλαμβανομένων περιπτώσεων όπου οι υποψήφιοι (ή/και οι ψηφοφόροι) παρήχθησαν από συνδυασμό κατανομών σε άνισα ποσοστά. Είναι σημαντικό να σημειωθεί, ότι οι κατανομές μπορεί να μην καλύπτουν όλο το χώρο, ούτε να επικαλύπτονται μεταξύ τους. Συγκεκριμένα χρησιμοποιήθηκαν οι παρακάτω κατανομές.

- **Gaussian (Γκαουσιανή)**

Ιδανικά σημεία παράγονται από μία συμμετρική γκαουσιανή κατανομή με μέση τιμή $(0,0)$ και απόκλιση 1.

- **Uniform Square (Ομοιόμορφη σε μορφή τετραγώνου)**

Ιδανικά σημεία παράγονται ομοιόμορφα στο τετράγωνο $[-3,3] \times [-3,3]$.

- **Uniform Disc (Ομοιόμορφη σε μορφή κυκλικού δίσκου)**

Ιδανικά σημεία παράγονται ομοιόμορφα στον κυκλικό δίσκο με κέντρο το $(0,0)$ και ακτίνα 3.

- **4-Gaussian**

Ιδανικά σημεία παράγονται από τέσσερις συμμετρικές γκαουσιανές κατανομές με απόκλιση 0,5, αλλά με διαφορετικές μέσες τιμές, συγκεκριμένα $(-1,0)$, $(1,0)$, $(0,-1)$ και $(0,1)$. Καθεμία χρησιμοποιείται για την παραγωγή του 25% των σημείων.

- **2-Gaussian unequal**

Ιδανικά σημεία παράγονται από δύο συμμετρικές γκαουσιανές κατανομές, καθεμία με διαφορετική μέση τιμή ή/και απόκλιση. Καθεμία χρησιμοποιείται για την παραγωγή διαφορετικού ποσοστού των σημείων.

- **3-Gaussian unequal**

Ιδανικά σημεία παράγονται από τρεις συμμετρικές γκαουσιανές κατανομές, καθεμία με διαφορετική μέση τιμή ή/και απόκλιση. Καθεμία χρησιμοποιείται για την παραγωγή διαφορετικού ποσοστού των σημείων.

3.1.3 Ερμηνεία της Θέσης στο Δισδιάστατο Χώρο

Η σειρά κατάταξης των υποψηφίων από κάθε ψηφοφόρο προκύπτει από τη θέση τους στο δισδιάστατο χώρο. Συγκεκριμένα θεωρούμε πως η απόσταση ενός ψηφοφόρου από τον κάθε υποψήφιο δείχνει τον βαθμό στον οποίο ο ψηφοφόρος δεν αντιπροσωπεύεται. Όσο μικρότερη η απόσταση, τόσο μεγαλύτερο το όφελος για τον ψηφοφόρο. Έτσι, πρώτος στις προτιμήσεις του ψηφοφόρου κατατάσσεται ο κοντινότερος υποψήφιος κοκ.

3.1.4 Θεωρήσεις

Μερικές από τις συναρτήσεις που χρησιμοποιήσαμε απαιτούν τυχαιότητα στην αρχικοποίηση δεδομένων ή στην λύση των ισοπαλιών. Επομένως, είναι απαραίτητο να γίνουν μερικές θεωρήσεις

για να μπορούμε να προβλέψουμε τη συμπεριφορά των αλγορίθμων για συγκεκριμένα δεδομένα εισόδου. Απαραίτητο είναι επίσης να γίνουν κάποιες θεωρήσεις για την επιλογή του τρόπου με τον οποίο συμπληρώνονται οι θέσεις της επιτροπής, έτσι ώστε να μπορούμε να διακρίνουμε τις πραγματικές διαφορές στις επιτροπές.

3.1.4.1 Λύση Ισοπαλιών

Όλοι οι αλγόριθμοι απαιτούν ταξινόμηση των ψηφοφόρων, συνήθως σύμφωνα με τον αριθμό ψήφων. Είναι ξεκάθαρο λοιπόν πως χρειάζεται ένας μηχανισμός λύσης των ισοπαλιών. Τέσσερις είναι οι βασικές προσεγγίσεις του τρόπου λύσης των ισοπαλιών [17]:

1. Να χρησιμοποιήσουμε μια προαποφασισμένη σειρά για τη λύση των ισοπαλιών.
2. Να χρησιμοποιήσουμε ένα τυχαίο μηχανισμό για τη λύση των ισοπαλιών.
3. Να καταλήξουμε σε ομάδα αποτελεσμάτων.
4. Να αγνοήσουμε τις ισοπαλίες, να θεωρήσουμε πως δεν υπάρχουν.

Καθεμία από τις προαναφερθείσες προσεγγίσεις έχει διαφορετικά προβλήματα. Για να αποφασίσουμε λοιπόν ποια θα χρησιμοποιήσουμε, πρέπει να λάβουμε υπόψη τα χαρακτηριστικά του πειράματος.

Όπως έχει αναφερθεί, οι τοποθεσίες των ψηφοφόρων και υποψηφίων παράγονται σύμφωνα με κάποια κατανομή. Η σειρά με την οποία παράγονται είναι τυχαία σε σχέση με τη θέση του σημείου στο χώρο. Αυτό δεν επηρεάζει το πείραμά μας, μπορεί ωστόσο να χρησιμοποιηθεί για το μηχανισμό λύσης των ισοπαλιών. Κρατάμε λοιπόν την σειρά με την οποία δημιουργήθηκε κάθε σημείο κλειδωμένη σε ένα διάνυσμα, και η σειρά αυτή αποτελεί την προαποφασισμένη σειρά με την οποία λύνουμε τις ισοπαλίες, διατηρώντας παράλληλα την απαραίτητη δόση τυχαιότητας.

3.1.4.2 Τυχαιές Αρχικοποιήσεις

Κάποιοι από τους κανόνες εκλογής επιτροπών απαιτούν τυχαία αρχικοποίηση της επιτροπής, η οποία επηρεάζει τα αποτελέσματα. Για να είναι ωστόσο συγκρίσιμα τα αποτελέσματα, κρατάμε τις αρχικές θέσεις σταθερές ανεξάρτητα από τον κανόνα. Στην πορεία των πειραμάτων, αποφασίσαμε να χρησιμοποιήσουμε διαφορετική αρχικοποίηση για κάθε κανόνα, για να έχουμε καλύτερα αποτελέσματα. Ωστόσο, η αρχικοποίηση για κάθε κανόνα γίνεται μονάχα μία φορά ανά κατανομή ψηφοφόρων, ανεξάρτητα από την κατανομή των υποψηφίων.

3.1.4.3 Ντετερμινισμός

Το «κλειδίωμα» κάθε σημείου σε συγκεκριμένη θέση στα αρχικά διανύσματα προσφέρει ακόμα ένα πλεονέκτημα: όσες φορές κι αν εφαρμόσουμε ένα συγκεκριμένο κανόνα στην ίδια είσοδο θα έχουμε πάντα την ίδια έξοδο (επιτροπή). Αφαιρώντας την τυχαιότητα στο μεγαλύτερο δυνατό βαθμό, οι κανόνες χαρακτηρίζονται ως ντετερμινιστικοί.

3.1.4.4 Ισοπαλίες στο Αποτέλεσμα (Ισοπαλίες της Τελευταίας Θέσης)

Άλλη μία απαραίτητη θεώρηση αφορά τις ισοπαλίες στην τελευταία θέση της επιτροπής. Αν τις αγνοήσουμε ή τις αντιμετωπίσουμε με κάποιο μηχανισμό, η επιτροπή θα αποτελείται από ακριβώς k υποψηφίους. Εναλλακτικά, ο αριθμός των εκλεγμένων υποψηφίων μπορεί να υπερβαίνει τους

k. Αυτή η απόφαση έχει ιδιαίτερη σημασία, ιδιαίτερα για τη μελέτη της ιδιότητας της ευστάθειας, όπου πρέπει να μπορούμε να διακρίνουμε τις πραγματικές διαφορές των επιτροπών.

Ας υποθέσουμε για παράδειγμα, ότι υπάρχει ισοπαλία μεταξύ δύο υποψηφίων για την τελευταία κενή θέση της επιτροπής. Μπορούμε να επιλέξουμε είτε να εκλεγούν και οι δύο είτε μόνο ο ένας τους. Έτσι προκύπτει το ακόλουθο ερώτημα. Πρέπει να θεωρηθούν διαφορετικά αυτά τα αποτελέσματα; Για την έννοια της αντιπροσωπευτικότητας ίσως να είναι αδιάφορο, αλλά για την έννοια της ευστάθειας δεν είναι.

Επιλέξαμε λοιπόν να μην επιτρέπονται οι ισοπαλίες και συγκεκριμένα να λύνονται με προκαθορισμένο τρόπο. Η μέτρηση της ευστάθειας έχει επηρεάσει λοιπόν τις θεωρήσεις του πειράματος, και πρέπει επομένως να είμαστε ιδιαίτερα προσεκτικοί με τα συμπεράσματά μας, καθώς υπάρχει κάποιου είδους «προκατάληψη (bias)» και η χειραγώγηση από έναν υποψήφιο είναι ακατάλληλο τεστ [17] αν δεν ερμηνεύσουμε τα αποτελέσματα λαμβάνοντας υπόψη μας τις «προκαταλήψεις». Περισσότερη ανάλυση σχετικά με την ευστάθεια ακολουθεί στο κεφάλαιο 5.

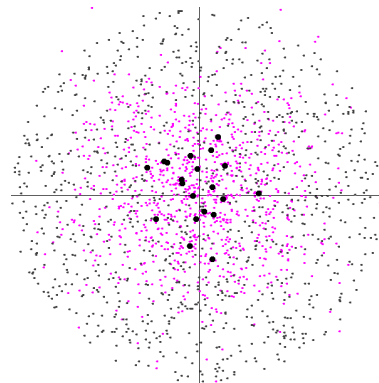
3.2 Αποτελέσματα, Σχολιασμός & Σύγκριση των Κανόνων

3.2.1 Κανόνας Single Nontransferable Vote

Ο κανόνας Single Nontransferable Vote χρησιμοποιείται καθημερινά στις ζωές μας. Είναι ένας πολύ γνωστός αλγόριθμος για την εκλογή επιτροπών, όπου απαραίτητη είναι η αντιπροσωπευτικότητα.

Όπως περιγράφηκε στην ενότητα 2.4 ο SNTV λαμβάνει υπόψιν μόνο την πρώτη προτίμηση των ψηφοφόρων. Πώς όμως αυτό επηρεάζει τα αποτελέσματα; Υπάρχουν ιδιαίτερα μοτίβα που τα χαρακτηρίζουν; και αν ναι, πώς αυτό επηρεάζει την καταλληλότητα του κανόνα για διαφορετικές εφαρμογές;

Στο παράδειγμα της εικόνας 3.1, όπου οι ψηφοφόροι ακολουθούν κανονική κατανομή και οι υποψήφιοι ομοιόμορφη σε μορφή κυκλικού δίσκου φαίνεται ότι ο κανόνας εξασφαλίζει την αντιπροσωπευτικότητα.

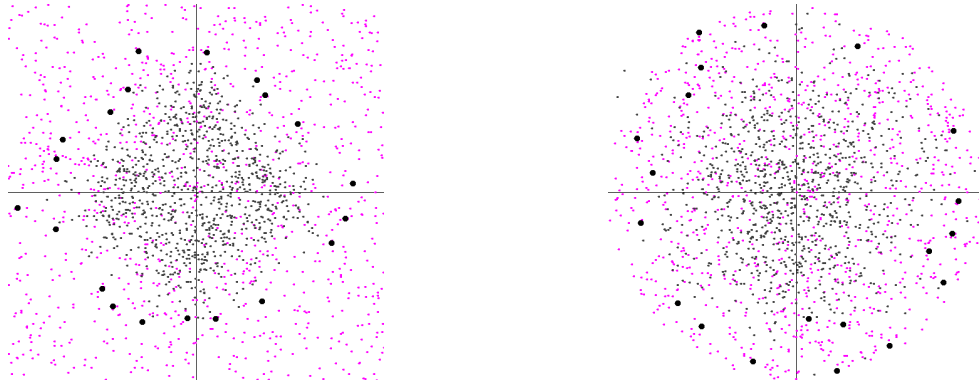


Εικόνα 3.1: Αποτέλεσμα του κανόνα SNTV

Ωστόσο, μετά από πειράματα με διαφορετικούς συνδυασμούς κατανομών, είναι ξεκάθαρο ότι η ιδιότητα δεν είναι εξασφαλισμένη λόγω της φύσης του κανόνα. Συγκεκριμένα, όταν οι ψηφοφόροι ακολουθούν ομοιόμορφη κατανομή, ενώ οι υποψήφιοι κανονική, παρατηρούμε ότι τα αποτελέσματα δεν ικανοποιούν την ιδιότητα της αναλογικής αντιπροσωπευτικότητας.

Στα παραδείγματα των εικόνων 3.2, η εκλεχθείσα επιτροπή τείνει να βρίσκεται στο περίγραμμα της κατανομής των υποψηφίων. Καθώς οι ψηφοφόροι ακολουθούν ομοιόμορφη κατανομή, είναι ομοιόμορφα κατανεμημένοι στο δισδιάστατο επίπεδο. Αντίθετα, οι υποψήφιοι ακολουθούν κανονική κατανομή και επομένως υπάρχουν σημεία του επιπέδου στα οποία υπάρχει μεγαλύτερη συγκέντρωση υποψηφίων από άλλα.

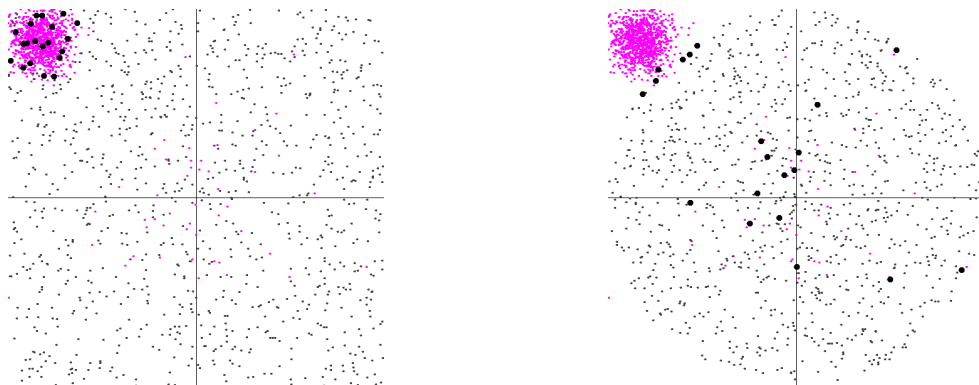
Είναι φυσικό λοιπόν, στα σημεία με μεγαλύτερη πυκνότητα υποψηφίων, οι ψήφοι να μοιρά-



Εικόνα 3.2: Το αποτέλεσμα του SNTV σχηματίζει το περίγραμμα της κατανομής των υποψηφίων

ζονται μεταξύ μεγάλου αριθμού των υποψηφίων, ενώ στα σημεία με μικρότερη πυκνότητα υποψηφίων, π.χ. στα άκρα της κατανομής, οι ψήφοι να μοιράζονται μεταξύ μικρότερου αριθμού υποψηφίων. Αν λάβουμε υπόψη και το γεγονός ότι οι ψηφοφόροι καλύπτουν μεγαλύτερο μέρος του επιπέδου, υπάρχουν περισσότεροι ψηφοφόροι που προτιμούν τους «ακραίους» υποψηφίους. Αυτές οι δυο παρατηρήσεις, εξηγούν αυτό το φαινόμενο, που μπορεί να εμφανιστεί ακόμα και σε εκλογές για την εύρεση ενός μόνο νικητή.

Χαρακτηριστικό παράδειγμα, είναι αυτό των εκλογών, όπου οι ψηφοφόροι υποστηρίζουν δύο «παρόμοιους» υποψηφίους, με αποτέλεσμα να εκλεγεί ένας τρίτος υποψήφιος, παρόλο που αθροιστικά οι δυο τους μπορεί να έχουν μεγαλύτερη υποστήριξη από τον νικητή.



Εικόνα 3.3: Το αποτέλεσμα του SNTV επηρεάζεται σημαντικά από την κατανομή των υποψηφίων

Ένα ακόμη ενδιαφέρον παράδειγμα είναι το αυτό της εικόνας 3.3. Και στις δύο περιπτώσεις έχουμε ακριβώς τους ίδιους ψηφοφόρους, ενώ οι κατανομές των υποψηφίων είναι ομοιόμορφες σε μορφή τετραγώνου και κυκλικού δίσκου αντίστοιχα. Στην πρώτη περίπτωση, όλοι οι υποψήφιοι που εκλέχθηκαν βρίσκονται στην πάνω αριστερή γωνία, όπου βρίσκονται οι περισσότεροι ψηφοφόροι, ενώ οι ψηφοφόροι που ανήκουν στην αραιή κατανομή φαίνεται να μην εκπροσωπούνται (ικανοποιούνται). Αντίθετα, στην δεύτερη περίπτωση, παρατηρούμε ότι οι ψηφοφόροι της αραιής κατανομής εκπροσωπούνται, όπως είναι αναμενόμενο και επιθυμητό. Εντυπωσιακό φαινόμενο είναι ότι υπάρχουν εκπρόσωποι των πιο ακραίων ψηφοφόρων, ακόμα κι αν ένας υποψήφιος εκπροσωπεί μονάχα ένα ψηφοφόρο.

Το παράδειγμα αυτό δείχνει την σημασία του συνόλου των υποψηφίων. Όταν υπάρχουν λίγοι υποψήφιοι που ικανοποιούν το μεγαλύτερο μέρος των ψηφοφόρων, υποψήφιοι που ικανοποιούν

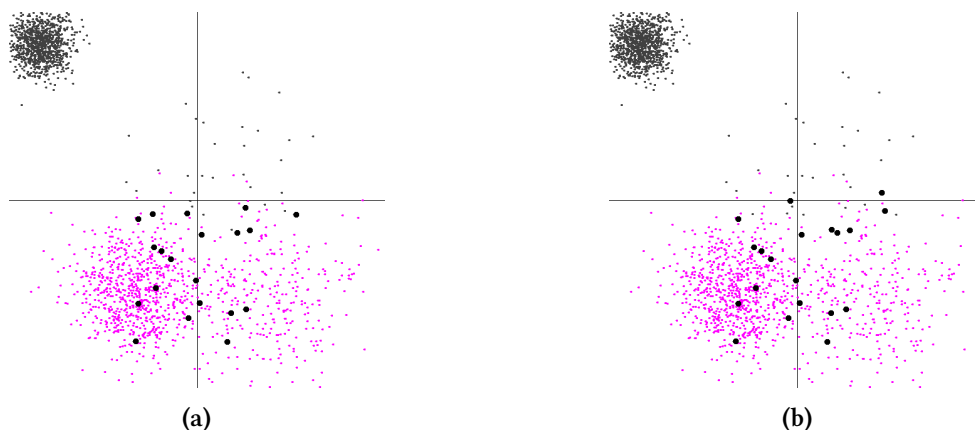
ελάχιστους ψηφοφόρους μπορεί τελικά να εκλεγούν.

3.2.2 Single Transferable Vote Rule

Στον κανόνα Single Transferable Vote, οι ψηφοφόροι τοποθετούν με φθίνουσα σειρά προτίμησης τους υποψήφιους. Κατά την καταμέτρηση των ψήφων, καθώς οι υποψήφιοι εκλέγονται ή απορρίπτονται, οι επιπλέον και οι ψήφοι που εναλλακτικά θα χάνονταν, μεταβιβάζονται στους άλλους υποψήφιους, ανάλογα με τις προτιμήσεις των ψηφοφόρων. Αυτό το σύστημα, παρέχει προσεγγιστικά αναλογική αντιπροσωπευτικότητα και επιτρέπει στους ψηφοφόρους να ψηφίζουν ανεξάρτητα από τα πολιτικά κόμματα και να εκλέγουν ανεξάρτητους υποψήφιους.

Ο κανόνας STV είναι ευαίσθητος προς τη σειρά εκλογής και απόρριψης των υποψηφίων. Ακόμη και η παραμικρή αλλαγή της σειράς αυτής, επηρεάζει το αποτέλεσμα. Έτσι, ακόμα και η πιο μικρή απόφαση στην υλοποίηση του αλγορίθμου, επηρεάζει, και επομένως αλλάζει, το αποτέλεσμα.

Για παράδειγμα, ας υποθέσουμε ότι θέλουμε να εκλέξουμε μόνο τον υποψήφιο με τους περισσότερους ψήφους, που πήρε τουλάχιστον το quota των ψήφων, να μεταβιβάσουμε τους επιπλέον ψήφους και στη συνέχεια να μετρήσουμε ξανά τους ψήφους για να βρούμε τον επόμενο υποψήφιο που θα εκλεγεί ή απορριφθεί, αντί να εκλέξουμε ταυτόχρονα όλους τους υποψήφιους που έλαβαν περισσότερους ψήφους από το quota (εικόνα 3.4 (a) και (b)). Μία επίσης σημαντική επιλογή είναι η επιλογή του υποψηφίου με τις λιγότερες ψήφους. Σε περίπτωση ισοπαλίας ποιος υποψήφιος πρέπει να αφαιρεθεί πρώτος; Αυτή η επιλογή θα μπορούσε να συμβεί με τυχαίο τρόπο. Ωστόσο, αυτό θα οδηγούσε σε διαφορετικά αποτελέσματα, αν τρέχαμε τον αλγόριθμο για ακριβώς την ίδια είσοδο παραπάνω από μία φορές. Στο κεφάλαιο 5 εξετάζουμε την ευστάθεια κάθε κανόνα. Για να μπορέσουμε να μελετήσουμε τον STV λοιπόν, απαραίτητο είναι να περιορίσουμε κατά το δυνατόν την τυχαιότητα του κανόνα. Έτσι, πάντοτε απορρίπτουμε τον τελευταίο υποψήφιο, ανάλογα με την ταξινόμηση που έχει γίνει, όπου οι ισοπαλίες λύνονται με προκαθορισμένο τρόπο¹.

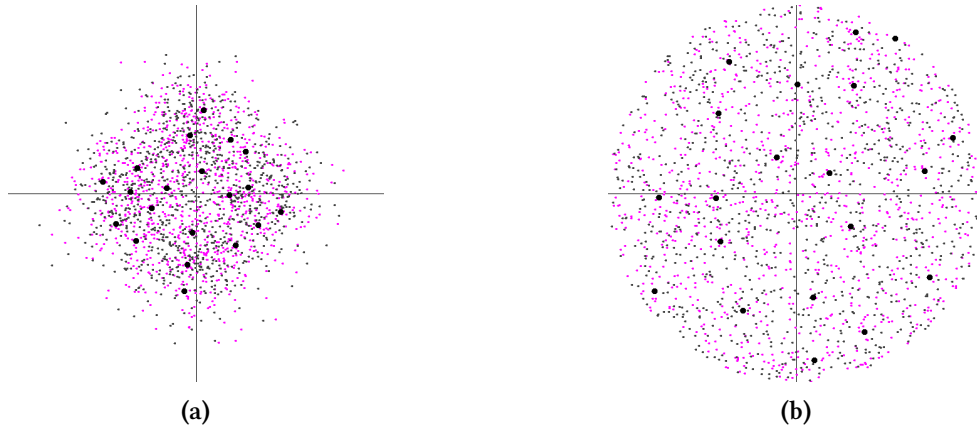


Εικόνα 3.4: Διαφορετικά αποτελέσματα για κοντινές παραλλαγές του κανόνα STV

Ο κανόνας STV είναι μια μέθοδος για τον υπολογισμό των αποτελεσμάτων εκλογών, η οποία λέγεται πως εγγυάται την αναλογική αντιπροσωπευτικότητα για στερεούς συνασπισμούς ψηφοφόρων – ομάδες ψηφοφόρων που μοιράζονται την πρώτη τους προτίμηση - κάτω από λογικές

¹ Στην ταξινόμηση λαμβάνουμε υπόψιν τις θέσεις των υποψηφίων στο αρχικό διάγραμμα, στο οποίο είναι αποθηκευμένοι με τη σειρά που παρήχθησαν.

συνθήκες² [16]. Αυτό εξηγείται καθώς ο κανόνας STV ελαχιστοποιεί τις σπαταλημένες ψήφους, μεταφέροντας τις επιπλέον ψήφους στους εναπομείναντες υποψήφιους.

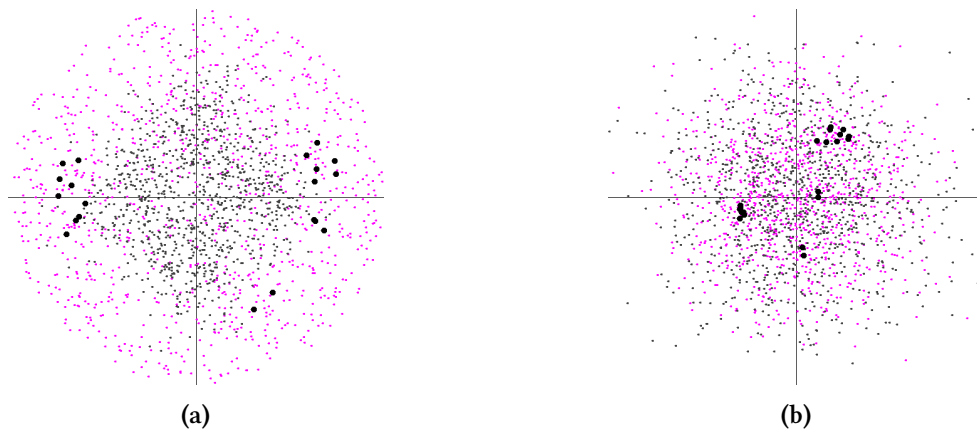


Εικόνα 3.5: Ο STV δείχνει να εγγυάται την αναλογική αντιπροσωπευτικότητα

Ωστόσο, πρόκειται για έναν ιδιαίτερα προβληματικό κανόνα, εξαιτίας της εξάρτησης του από τη σειρά εκλογής ή απόρριψης των υποψηφίων.

3.2.3 Κανόνας Bloc

Στον κανόνα Bloc, οι ψηφοφόροι επιλέγουν τους k αγαπημένους τους υποψήφιους. Στα αποτελέσματα παρατηρούμε ότι οι εκλεγμένες επιτροπές αποτελούνται από μικρές ομάδες υποψηφίων, που συνήθως βρίσκονται σε περιοχές που η πυκνότητα των υποψηφίων είναι μικρή, π.χ. κοντά στις άκρες της κατανομής των υποψηφίων³.



Εικόνα 3.6: Οι εκλεγμένοι υποψήφιοι από τον κανόνα Bloc σχηματίζουν ομάδες κοντά στις άκρες της κατανομής των υποψηφίων.

Όπως βλέπουμε στις εικόνες 3.6 (a) και (b), η εκλεχθείσα επιτροπή, δεν φαίνεται να ικανοποιεί την πλειοψηφία των ψηφοφόρων. Αυτό είναι άμεση συνέπεια της φύσης του κανόνα. Σε περιοχές που η πυκνότητα των υποψηφίων είναι μεγάλη, οι ψήφοι τείνουν να διασπώνται ανάμεσα σε παρόμοιους υποψήφιους και ως αποτέλεσμα οι υποψήφιοι τελικά να μην εκλέγονται.

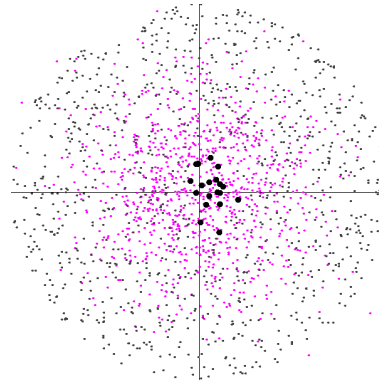
² Η υπόθεση αυτή, εξετάζεται αναλυτικά στην ενότητα 3.2.7

³ Προϋπόθεση αποτελεί η κατανομή των ψηφοφόρων να καλύπτει την περιοχή που καλύπτει η κατανομή των υποψηφίων.

Ο μόνος τρόπος για να εκπροσωπηθούν περισσότεροι ψηφοφόροι, είναι η κατανομή των υποψηφίων να είναι αραιή στις περιοχές που η κατανομή των ψηφοφόρων είναι πυκνή, όπως συμβαίνει στο παράδειγμα της εικόνας 3.7.

Ακόμα και τότε ωστόσο, η επιτροπή δεν εμφανίζει αρκετή ποικιλομορφία και αποτυγχάνει να εκπροσωπήσει την πλειοψηφία των ψηφοφόρων, στην πλειονότητα των περιπτώσεων.

Ο κανόνας αυτός λοιπόν, δεν είναι κατάλληλος για περιπτώσεις που αποζητάται ποικιλομορφία στην επιτροπή και αναλογική αντιπροσωπευτικότητα. Αντίθετα, η δικαιοσύνη του κανόνα, δεν μπορεί να αμφισβητηθεί, καθώς όλοι οι παρόμοιοι υποψήφιοι είτε απορρίπτονται είτε εκλέγονται.



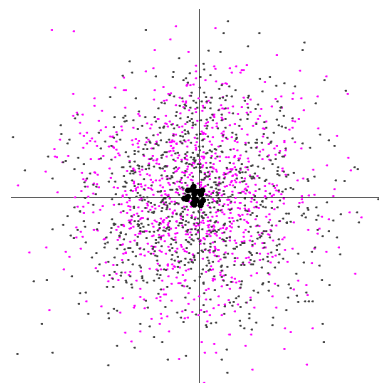
Εικόνα 3.7: Τα αποτελέσματα του κανόνα Bloc εξαρτώνται από την πυκνότητα της κατανομής των υποψηφίων

3.2.4 Κανόνας k -Borda

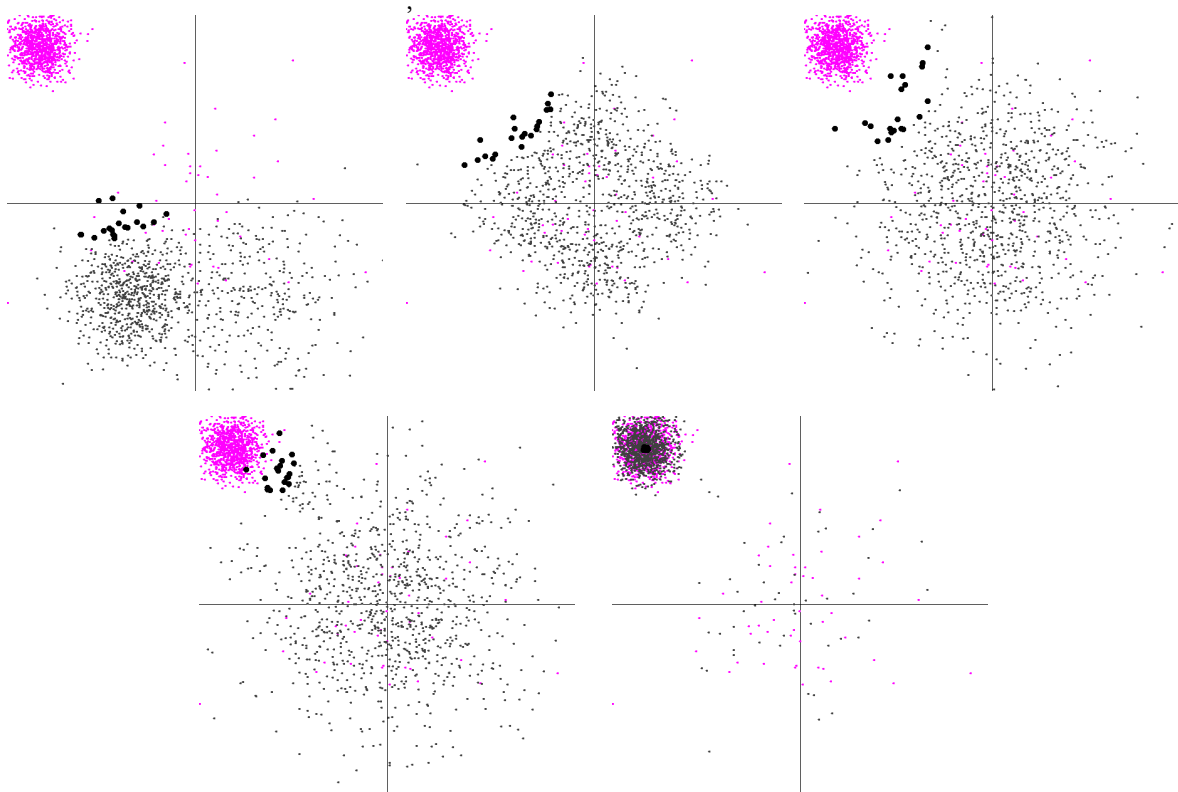
Ο κανόνας k -Borda αποφασίζει το αποτέλεσμα των εκλογών, δίνοντας σε κάθε υποψήφιο, για τη σειρά κατάταξης κάθε ψηφοφόρου, τόσους ψήφους, όσος και ο αριθμός των υποψηφίων που ο ψηφοφόρος κατέταξε κατώτερα. Όταν όλοι οι ψήφοι έχουν καταμετρηθεί, οι k υποψήφιοι με τους περισσότερους ψήφους εκλέγονται. Η εκδοχή του κανόνα του Borda για την εκλογή ενός νικητή μόνο περιγράφεται ως σύστημα ψηφοφορίας που βασίζεται στην επικρατούσα άποψη και χρησιμοποιείται για περιπτώσεις που απαραίτητη είναι η αναλογική αντιπροσωπευτικότητα.

Η καταλληλότητα του κανόνα αυτού για την εύρεση των επικρατέστερων υποψηφίων (ή πιο αποδεκτών) αιτιολογείται από το ακόλουθο μοτίβο: η εκλεχθείσα επιτροπή, τείνει να βρίσκεται στο κέντρο μάζας των ψηφοφόρων όσο η κατανομή των υποψηφίων το επιτρέπει (εικόνα 3.9).

Αυτό το μοτίβο δικαιολογεί και την ακαταλληλότητα του κανόνα αυτού για εκλογές. Παρόλο που η επιτροπή θα ήταν ευρέως περισσότερο αποδεκτή, όλα τα μέλη της θα βρίσκονται στο ίδιο περίπου σημείο. Έτσι, όλοι οι εκλεχθέντες υποψήφιοι θα είναι παρόμοιοι και δεν θα υπάρχει ούτε ποικιλομορφία ούτε αναλογική αντιπροσωπευτικότητα των ψηφοφόρων.



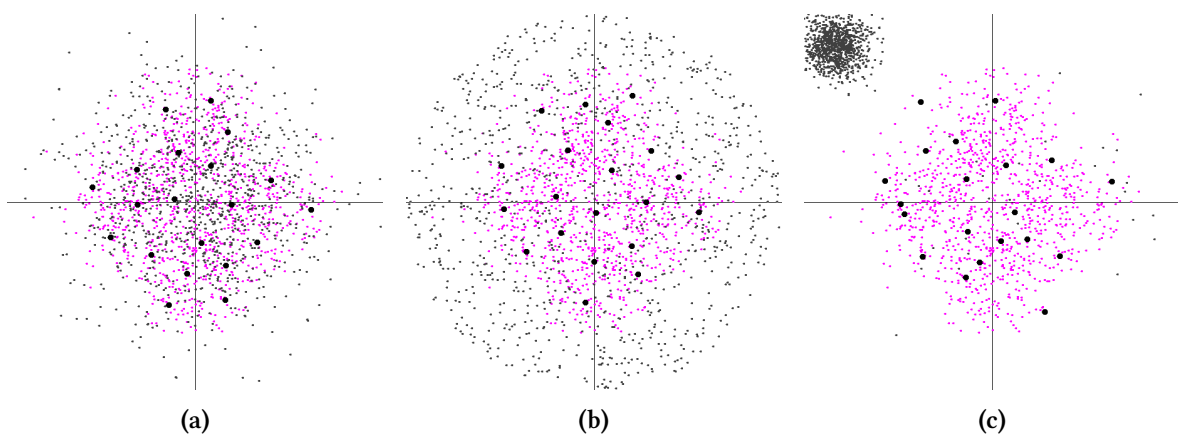
Εικόνα 3.8: Αποτελέσματα του κανόνα k -Borda



Εικόνα 3.9: Ο κανόνας k -Borda επιλέγει υποψήφιους που βρίσκονται κοντά στο κέντρο της μάζας της κατανομής των ψηφοφόρων

3.2.5 Κανόνας Chamberlin-Courant⁴

Ο κανόνας Chamberlin-Courant εκλέγει τους υποψήφιους που ελαχιστοποιούν την μη εκπροσώπηση των ψηφοφόρων, δημιουργώντας ομάδες ψηφοφόρων. Έτσι, οι υποψήφιοι που εκλέγονται, βρίσκονται ιδανικά στο κέντρο κάθε ομάδας. Εφόσον, δεν υπάρχουν περιορισμοί στο μέγεθος κάθε



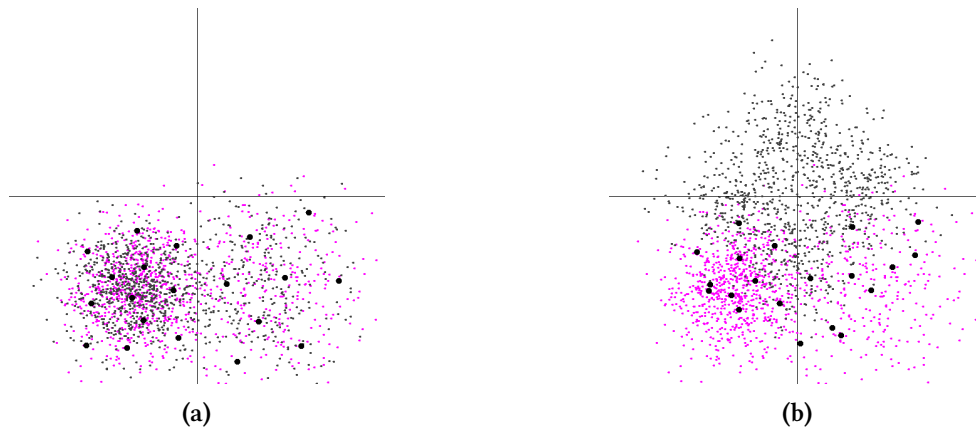
Εικόνα 3.10: Αποτελέσματα του κανόνα Chamberlin-Courant

Οι ψηφοφόροι ακολουθούν κατανομή 4-Gaussian, ενώ οι υποψήφιοι (a) Gaussian, (b) Uniform Disc, and (c) 2-Gaussian Unequal

⁴ Τα αποτελέσματα που παρουσιάζονται σε αυτή την ενότητα είναι του ευρετικού αλγόριθμου του Chamberlin-Courant, τον οποίο αναλύουμε στην ενότητα 4.1 Παρόλα αυτά, οι παρατηρήσεις που παρουσιάζονται αντιπροσωπεύουν και τον πραγματικό κανόνα, εκτός και αν σημειωθεί διαφορετικά.

ομάδας, τα μέλη της επιτροπής καλύπτουν την μέγιστη δυνατή περιοχή που καλύπτει η περιοχή των ψηφοφόρων. Για το λόγο αυτό, δεν μπορούμε να προσδιορίσουμε την κατανομή των ψηφοφόρων, καθώς το αποτέλεσμα θυμίζει ομοιόμορφη κατανομή, εφόσον το επιτρέπει η κατανομή των υποψηφίων.

Στον ευρετικό μας αλγόριθμο, όταν η κατανομή των υποψηφίων δεν καλύπτει ολόκληρη την περιοχή που καλύπτει η κατανομή των ψηφοφόρων, οι εκλεγμένοι υποψήφιοι, είναι αυτοί που βρίσκονται κοντινότερα στο κέντρο των ομάδων. Ο NP-hard Chamberlin-Courant, θα επέλεγε ωστόσο τους ψηφοφόρους περισσότερο ομοιόμορφα.



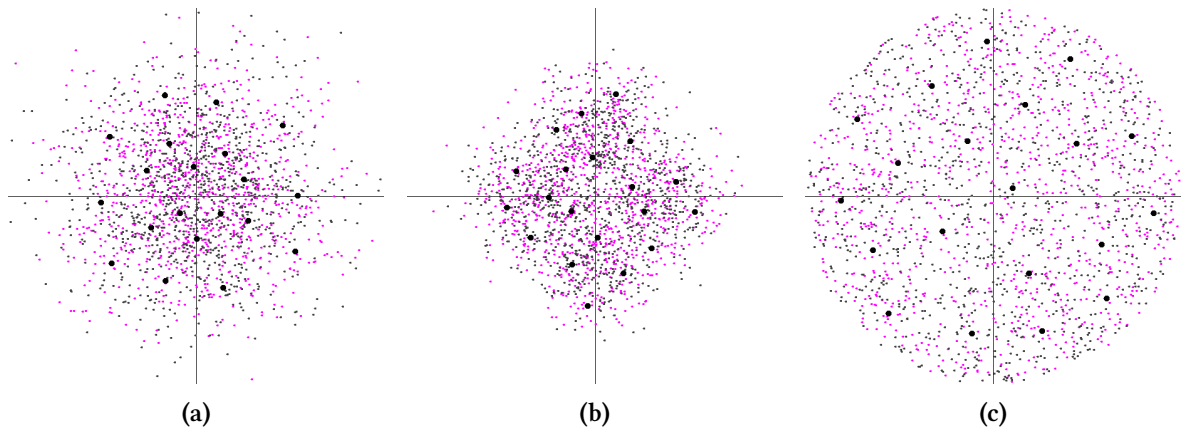
Εικόνα 3.11: Το αποτέλεσμα του Chamberlin-Courant, όταν η κατανομή των υποψηφίων δεν καλύπτει το χώρο που καλύπτει η κατανομή των ψηφοφόρων
Οι ψηφοφόροι ακολουθούν 2-Gaussian Unequal κατανομή, ενώ οι υποψήφιοι ακολουθούν (a) 2-Gaussian Unequal, και (b) 4-Gaussian

3.2.6 Κανόνας Monroe⁵

Ο κανόνας Monroe εκλέγει τους υποψήφιους που ελαχιστοποιούν την μη εκπροσώπηση των ψηφοφόρων, δημιουργώντας ισοπληθείς ομάδες ψηφοφόρων. Έτσι, οι ψηφοφόροι που εκλέγονται, βρίσκονται στο κέντρο κάθε ομάδας. Εξαιτίας του περιορισμού στον αριθμό του πληθυσμού των ομάδων, σε αντίθεση με τον κανόνα του Chamberlin-Courant, εντοπίζουμε το είδος της κατανομής των ψηφοφόρων, καθώς ο κανόνας του Monroe την σκιαγραφεί. Ωστόσο, αυτό μπορεί να συμβεί μόνο κάτω από δύο προϋποθέσεις:

1. η επιτροπή να είναι αρκετά μεγάλη, και
2. οι κατανομές να έχουν αρκετά μέλη για να εκπροσωπηθούν, καθώς σε αυτόν τον κανόνα, οι μειοψηφίες δεν εκπροσωπούνται εάν ο πληθυσμός τους δεν δικαιολογεί τη δημιουργία δικής τους ομάδας.

⁵ Τα αποτελέσματα που παρουσιάζονται σε αυτή την ενότητα είναι του ευρετικού αλγόριθμου του Monroe, που αναλύουμε στην ενότητα 4.2 Παρόλα αυτά, οι παρατηρήσεις αντιπροσωπεύουν και τον πραγματικό κανόνα, εκτός και αν σημειώνεται διαφορετικά.

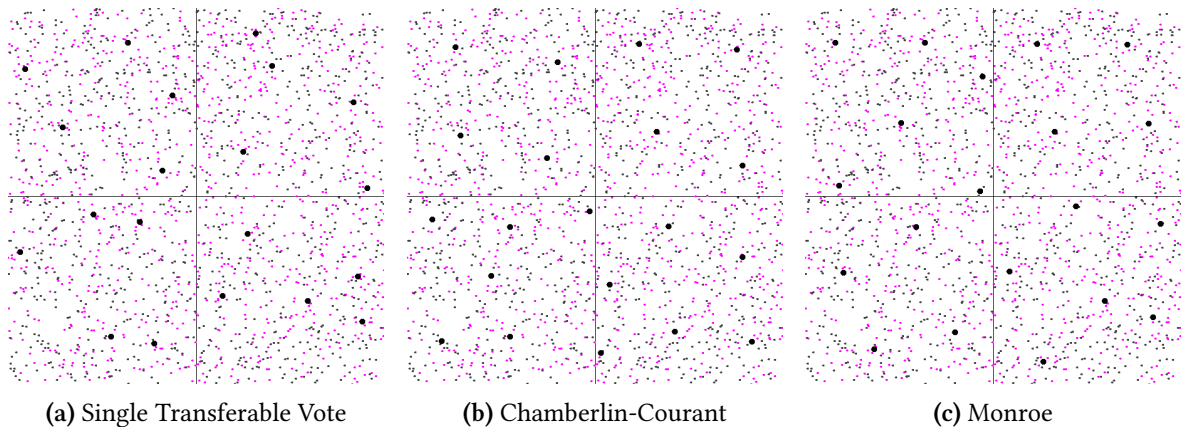


Εικόνα 3.12: Το αποτέλεσμα του κανόνα του Monroe σκιαγραφεί την κατανομή των ψηφοφόρων. Οι υποψήφιοι και ψηφοφόροι ακολουθούν την ίδια κατανομή: (a) Gaussian, (b) 4-Gaussian, (c) Uniform Disc

3.2.7 Σύγκριση των Κανόνων Single Transferable Vote, Chamberlin-Courant & Monroe

Έπειτα από την περιγραφή και σχολιασμό των κανόνων, θεωρούμε απαραίτητο να συγκρίνουμε εκτενώς τους κανόνες Single Transferable Vote, Chamberlin-Courant και Monroe, για να εντοπίσουμε πότε τείνουν να έχουν τα ίδια αποτελέσματα.

Ο κανόνας του Monroe θυμίζει τον Chamberlin-Courant, με μόνη διαφορά ότι στον κανόνα του Monroe, κάθε μέλος της επιτροπής εκπροσωπεί τον ίδιο αριθμό ψηφοφόρων. Έτσι, όταν οι ψηφοφόροι ακολουθούν ομοιόμορφη κατανομή, τα αποτελέσματα και για τους δύο κανόνες είναι παρόμοια και δεν διακρίνουμε ποιο αποτέλεσμα αντιστοιχεί στον κάθε κανόνα.

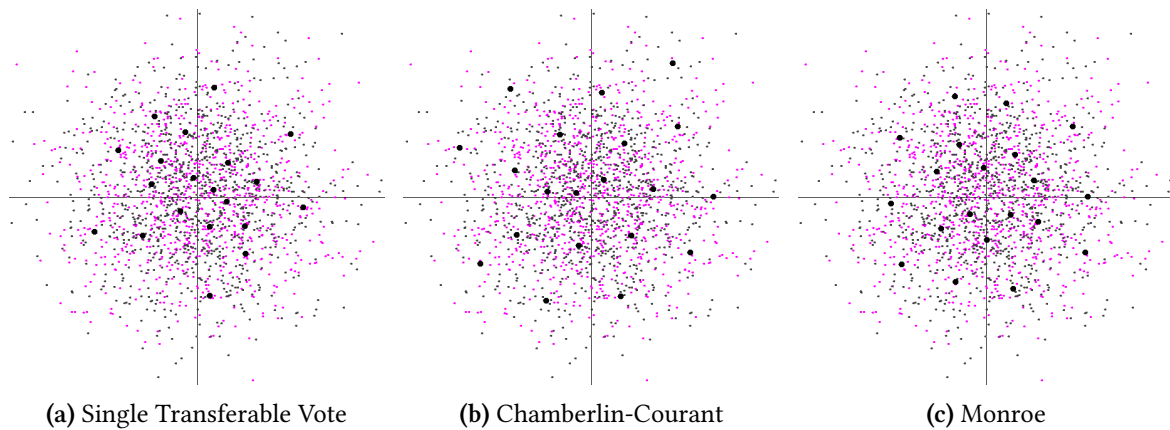


Εικόνα 3.13: Συγκρίσεις για ομοιόμορφη κατανομή των ψηφοφόρων

Στις εικόνες 3.13 (a), (b), (c) βλέπουμε ότι τα μέλη της επιτροπής είναι ομοιόμορφα κατανεμημένα στο χώρο για τους κανόνες Monroe και Chamberlin-Courant. Για τον κανόνα Single Transferable Vote, τα μέλη της επιτροπής βρίσκονται στο χώρο με περισσότερη αταξία, εξαιτίας της σειράς απόρριψης και εκλογής των υποψηφίων.

Αντίθετα, όταν οι ψηφοφόροι ακολουθούν κανονικές κατανομές, ξεκάθαρη είναι η διαφορά μεταξύ των αποτελεσμάτων των κανόνων των Chamberlin-Courant και Monroe. Ο Monroe, τείνει να σκιαγραφήσει την κατανομή (εικόνα 3.14 (c)), ενώ ο Chamberlin-Courant να καλύψει ομοιόμορφα

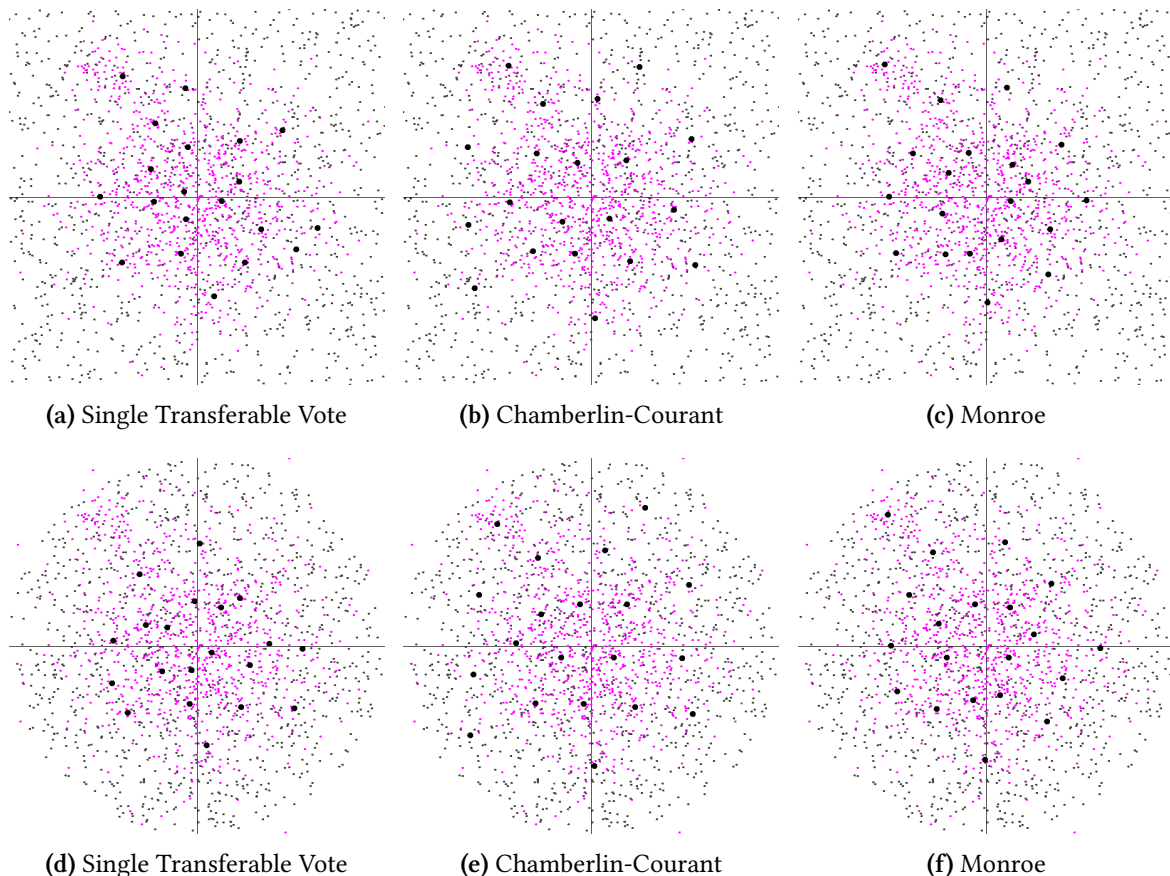
το χώρο που καλύπτει η κατανομή των ψηφοφόρων (εικόνα 3.14 (b)).



Εικόνα 3.14: Συγκρίσεις για κανονική κατανομή των ψηφοφόρων

Ο κανόνας Single Transferable Vote, λέγεται πως εγγυάται την αναλογική αντιπροσωπευτικότητα. Στο παράδειγμα της εικόνας 3.14 (a) φαίνεται πως ισχύει, ακόμα και αν τα μέλη της επιτροπής κατανέμονται με πιο άτακτο τρόπο από ό,τι στον κανόνα του Monroe. Μελετώντας απλά παραδείγματα, καταλήγουμε πως τα αποτελέσματα του κανόνα Single Transferable Vote στη μέση περίπτωση προσεγγίζουν αυτά του Monroe, επιβεβαιώνοντας έτσι τον χαρακτηρισμό του ως καλού κανόνα για εκλογές.

Τι συμβαίνει ωστόσο, σε περιπτώσεις που οι ψηφοφόροι ακολουθούν δύο διαφορετικές κατανομές, σε διαφορετικό ποσοστό, όπου το ένα είναι πολύ μικρότερο από το άλλο, αλλά αρκετά μεγάλο για να έχει εκπρόσωπο στον κανόνα Monroe;

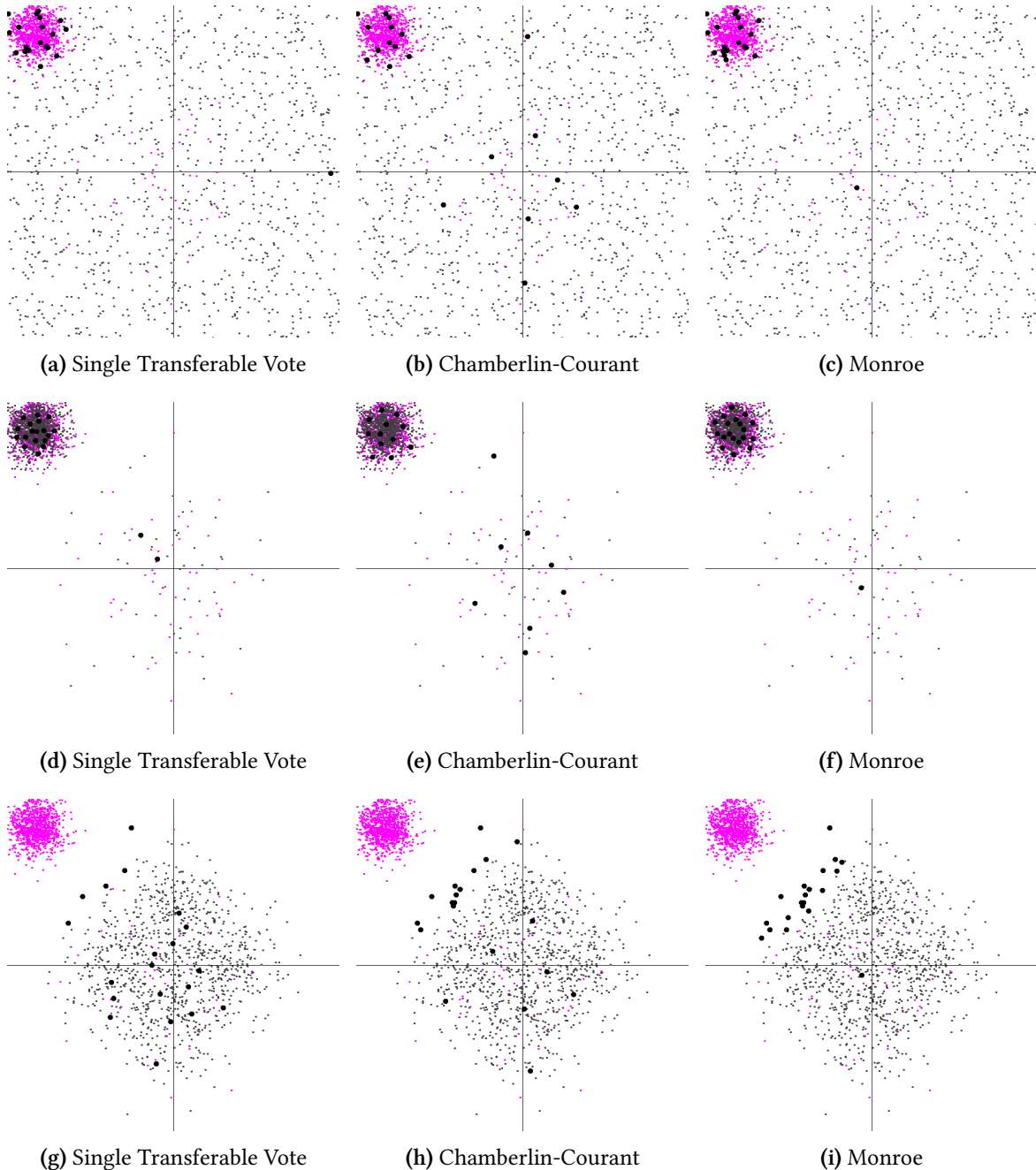


Εικόνα 3.15: Συγκρίσεις για άνισες κανονικές κατανομές ψηφοφόρων

Σε όλες τις εικόνες οι ψηφοφόροι είναι οι ίδιοι. Οι υποψήφιοι των εικόνων (a), (b) και (c) είναι οι ίδιοι και ακολουθούν κατανομή Uniform Square, ενώ για τις εικόνες (d), (e) και (f) Uniform Disc.

Στα παραδείγματα των εικόνων 3.15 οι ψηφοφόροι παράγονται κατά 95% από κανονική κατανομή με μέση τιμή (0,0) και απόκλιση 1, και κατά 5% από κανονική κατανομή με μέση τιμή (-1.2,1.8) και απόκλιση 0.25. Στο παράδειγμα αυτό, γίνεται ξεκάθαρο, ότι σε περιπτώσεις που οι ψηφοφόροι παράγονται από παραπάνω από μία κατανομές, ακόμα και αν η πυκνότητά τους είναι ίδια, ο κανόνας Single Transferable Vote, αποτυγχάνει να εγγυηθεί την αντιπροσώπευση της «μικρής» κατανομής, ακόμα κι όταν πρέπει να υπάρχει.

Μελετήσαμε επίσης την περίπτωση όπου οι ψηφοφόροι παράγονται κατά 95% από κανονική κατανομή με μέση τιμή (-2.5,2.5) και απόκλιση 0.25, και κατά 5% από κανονική κατανομή με μέση τιμή (0,0) και απόκλιση 1.



Εικόνα 3.16: Συγκρίσεις για άνισες κανονικές κατανομές ψηφοφόρων

Σε όλες τις εικόνες οι ψηφοφόροι είναι οι ίδιοι. Οι υποψήφιοι των εικόνων (a), (b) και (c) είναι οι ίδιοι και ακολουθούν κατανομή Uniform Square, στις εικόνες (d), (e) και (f) ακολουθούν την κατανομή των ψηφοφόρων, ενώ στις εικόνες (g), (h) και (j) ακολουθούν κατανομή 4-Gaussian.

Εφόσον η κατανομή των υποψηφίων καλύπτει την περιοχή που οι βρίσκονται οι ψηφοφόροι, μπορούμε να δούμε ότι ο κανόνας Single Transferable Vote μπορεί να αντιστοιχίσει έναν εκπρόσωπο (εικόνα 3.16 (a)) ή περισσότερους (εικόνα 3.16 (d)), ενώ αναλογικά μόνο ένας εκπρόσωπος θα έπρεπε να αντιστοιχιστεί στην μικρή κατανομή. Ο ιδανικός μάλιστα εκπρόσωπος είναι ο υποψήφιος που βρίσκεται κοντινότερα στη μέση τιμή της κατανομής. Παρόλα αυτά, στην εικόνα 3.16 (a) λόγω της σειράς απόρριψης των υποψηφίων εκλέγεται ένας ακραίος υποψήφιος, ενώ στην εικόνα 3.16 (d) εκλέγονται δύο υποψήφιοι κοντά στην τοποθεσία που βρίσκεται ο ιδανικός εκπρόσωπος.

Όταν η κατανομή των υποψηφίων δεν καλύπτει όλο το χώρο που καλύπτει η κατανομή των ψηφοφόρων, το αποτέλεσμα του κανόνα Single Transferable Vote θυμίζει περισσότερο τα αποτελέσματα του Chamberlin-Courant παρά του Monroe. Η πλειοψηφία των εκλεχθέντων υποψηφίων εκπροσωπεί το 5% των ψηφοφόρων, που ανήκουν στην αραιή κατανομή. Το αποτέλεσμα αυτό μάλιστα είναι κοντινότερο στο ιδανικό αποτέλεσμα του Chamberlin-Courant από ό,τι το αποτέλεσμα του ευρετικού μας αλγόριθμου⁶. Επομένως, με τα πειράματά μας επιβεβαιώνεται ότι ο κανόνας Single Transferable Vote δεν μπορεί να εγγυηθεί πλήρως την αναλογική αντιπροσωπευτικότητα όταν υπάρχουν ομάδες μειονοτήτων. Το αποτέλεσμα μπορεί να ευνοεί κάποιες μειονότητες σε βάρος της πλειοψηφίας ή ακόμα και να ευνοήσει ακραίους υποψήφιους, λόγω της σειράς απόρριψης υποψηφίων.

3.2.8 Συγκρίσεις των Κανόνων σύμφωνα με τις Ιδιότητες

Όταν επιλέγουμε έναν κανόνα εκλογής επιτροπών, συνηθισμένο προαπαιτούμενο της επιτροπής είναι να ικανοποιεί την αναλογική αντιπροσωπευτικότητα. Η αναλογική αντιπροσωπευτικότητα προβλέπει ότι οι ψηφοφόροι θα εκπροσωπηθούν σε μία επιτροπή ανάλογα με τα ποσοστά τους. Άλλωστε, είναι ευρέως αποδεκτό ότι η αναλογική αντιπροσωπευτικότητα είναι ο πιο τίμιος τρόπος να αντικατοπτριστούν οι διαφορετικές απόψεις των ψηφοφόρων [18]. Η συνεπτυγμένη αξιολόγηση των κανόνων με γνώμονα το σεβασμό τους στην αναλογική αντιπροσώπευση ακολουθεί στον πίνακα 3.1.

	Αναλογική Αντιπροσωπευτικότητα		
	Καλύτερη Περίπτωση	Μέση Περίπτωση	Χειρότερη Περίπτωση
SNTV	✓	✗	✗
STV	✓	✓	✗
Bloc	✗	✗	✗
<i>k</i> -Borda	✓	✗	✗
CC (h)	✓	♣	✗
Monroe (h)	✓	✓	♣

Πίνακας 3.1: Αναλογική αντιπροσωπευτικότητα

Τα ✓ / ✗ δείχνουν ότι ένας κανόνας πετυχαίνει/δεν πετυχαίνει την αναλογική αντιπροσωπευτικότητα. Το ♣ δείχνει ότι ένας κανόνας μπορεί να επιτύχει ή όχι την αναλογική αντιπροσωπευτικότητα στην επιτροπή λόγω της αρχικοποίησης του ευρετικού αλγόριθμου.

Επιπλέον αξιολογήσαμε τους κανόνες με γνώμονα την ποικιλομορφία στην εκλεχθείσα επιτροπή. Τα αποτελέσματα βρίσκονται συγκεντρωμένα στον πίνακα 3.2.

⁶ Στην ενότητα 4.1 παρουσιάζεται αυτή η αδυναμία του ευρετικού μας αλγόριθμου.

	Ποικιλομορφία		
	Καλύτερη Περίπτωση	Μέση Περίπτωση	Χειρότερη Περίπτωση
SNTV	✓	✓	✗
STV	✓	✓	✗
Bloc	✗	✗	✗
<i>k</i> -Borda	✗	✗	✗
CC (h)	✓	✓	♣
Monroe (h)	✓	♣	✗

Πίνακας 3.2: Ποικιλομορφία

Τα ✓/ ✗ δείχνουν ότι ο κανόνας πετυχαίνει/δεν πετυχαίνει την ποικιλομορφία. Το ♣ δείχνει ότι ένας κανόνας μπορεί να επιτύχει ή όχι την ποικιλομορφία στην επιτροπή λόγω της αρχικοποίησης του ευρετικού αλγόριθμου.

Κεφάλαιο 4

Ευρετικοί Αλγόριθμοι για τους NP-hard Κανόνες

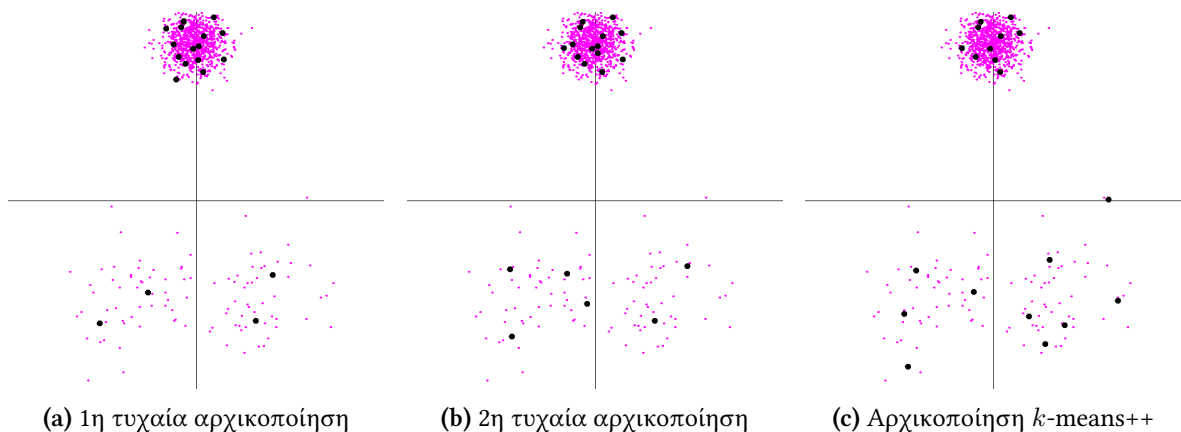
Μερικοί από τους κανόνες που μελετήσαμε ανήκουν υπολογιστικά στο NP και ως αποτέλεσμα δεν μπορούμε να βρούμε την επιτροπή σε πολυωνυμικό χρόνο. Έτσι, για τους κανόνες Chamberlin-Courant και Monroe, δημιουργήσαμε τους δικούς μας ευρετικούς αλγορίθμους και κρίναμε αν τα αποτελέσματα¹ των αλγορίθμων μας ήταν αρκετά κοντινά στα αναμενόμενα.

4.1 Κανόνας Chamberlin-Courant

4.1.1 Ευρετικός Αλγόριθμος για τον Κανόνα Chamberlin-Courant

Στον κανόνα Chamberlin-Courant, κάθε ψηφοφόρος αντιστοιχίζεται στον προτιμητέο του υποψήφιο σε κάθε δυνατή επιτροπή. Η τελική επιτροπή αποτελείται από τους υποψήφιους που ελαχιστοποιούν την μη αντιπροσωπευτικότητα² των ψηφοφόρων.

Ο κανόνας Chamberlin-Courant μπορεί να γίνει αντιληπτός ως ένα πρόβλημα ομαδοποίησης, παρόμοιο με αυτό του k -means. Συνεπώς, ο αλγόριθμος k -means³, κατάλληλα τροποποιημένος, είναι ένας απλός και γρήγορος αλγόριθμος που επιλύει το πρόβλημα, χωρίς ωστόσο να προσφέρει καμία εγγύηση προσέγγισης του πραγματικού αλγορίθμου [19].



Εικόνα 4.1: Μη ισοδύναμα αποτελέσματα για τον ευρετικό αλγόριθμο του κανόνα Chamberlin-Courant εξαιτίας διαφορετικών αρχικοποιήσεων

¹ Σημειώνεται ότι στις περισσότερες εικόνες αυτού του κεφαλαίου, δεν εμφανίζονται οι υποψήφιοι για να είναι πιο ξεκάθαρο το αποτέλεσμα. Τότε, οι υποψήφιοι έχουν παραχθεί σύμφωνα ομοιόμορφη κατανομή. Οι μοβ κουκκίδες αναπαριστούν τους ψηφοφόρους, ενώ οι μαύρες τους εκλεγμένους υποψήφιους, δηλαδή αυτούς που βρίσκονται κοντινότερα στα σημεία που προτείνουν οι αλγόριθμοί μας. Κατά συνέπεια, αν οι κατανομές των υποψηφίων διέφεραν, τα αποτελέσματα θα διέφεραν επίσης.

² Στο πείραμά μας, η μη αντιπροσωπευτικότητα ορίζεται ως την απόσταση του ψηφοφόρου από τον υποψήφιο που τον εκπροσωπεί.

³ Ο αλγόριθμος k -means χρησιμοποιείται για να ομαδοποιήσει τα δεδομένα εισόδου σε k ομάδες (clusters).

Ακολουθεί ο αλγόριθμος που χρησιμοποιήσαμε (αλγόριθμος 1), που αποτελεί προσαρμογή του k -means για τα δεδομένα του προβλήματος.

Αλγόριθμος 1 Ευρετικός αλγόριθμος του κανόνα Chamberlin-Courant

1. Τυχαία διάλεξε ως αρχικά k κέντρα $C = \{c_1, c_2, \dots, c_k\}$.
2. Για κάθε $i \in \{1, \dots, k\}$, θέσε το cluster C_i ως το σύνολο των σημείων στο X που βρίσκονται κοντινότερα στο c_i από ότι στο c_j για κάθε $j \neq i$.
3. Για κάθε $i \in \{1, \dots, k\}$, θέσε c_i το κέντρο της μάζας όλων των σημείων στο C_i : $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.
4. Επανάλαβε τα βήματα 2 και 3 έως ότου το C να μην αλλάζει.

Με την παρακάτω προσθήκη για την εύρεση της επιτροπής⁴:

5. Βρες τους κοντινότερους υποψήφιους στα εκλεγμένα κέντρα των clusters.

Συνήθης τακτική για την αρχική επιλογή των κέντρων από το X , είναι η τυχαία επιλογή k σημείων, όταν όλα τα σημεία μπορούν να επιλεγούν με την ίδια πιθανότητα. Στον αλγόριθμο k -means, τα αποτελέσματα επηρεάζονται από την αρχικοποίηση των κέντρων. Παρόλο που στις περισσότερες περιπτώσεις τα αποτελέσματα θεωρούνται ισοδύναμα, αν μετρήσουμε την μη εκπροσώπηση των ψηφοφόρων, υπάρχουν περιπτώσεις που τα αποτελέσματα ποικίλουν (εικόνες 4.1c (a) και (b)).

4.1.2 Τεχνική Αρχικοποίησης

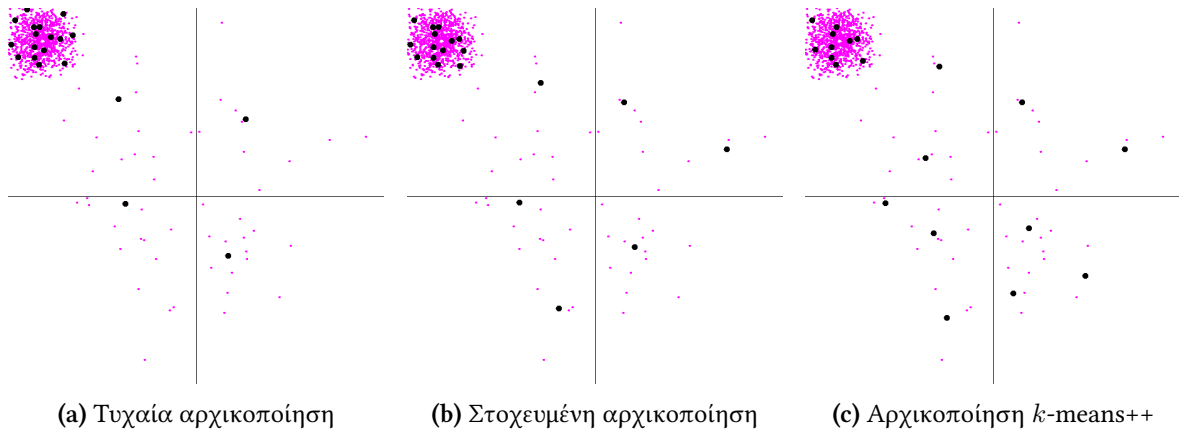
Θέλοντας να μελετήσουμε περιπτώσεις με μη ισοδύναμα αποτελέσματα, δημιουργήσαμε ειδικά σχεδιασμένα τεστ για να ελέγξουμε κατά πόσο μπορούμε να βελτιώσουμε τα αποτελέσματα διαλέγοντας τα αρχικά κέντρα με στοχευμένο τρόπο και όχι τυχαία. Μία από τις αρχικοποιήσεις που δοκιμάσαμε είναι αυτή που χρησιμοποιείται στον αλγόριθμο k -means++ [19] και αποτέλεσε την βέλτιστη αρχικοποίηση, την οποία χρησιμοποιούμε τελικά και στον προτεινόμενο ευρετικό αλγόριθμο. Αυτός περιγράφεται στην συνέχεια (αλγόριθμος 2).

Αλγόριθμος 2 Ευρετικός αλγόριθμος του κανόνα Chamberlin-Courant (αρχικοποίηση k -means++)

- 1a. Πάρε τυχαία ένα κέντρο c_1 από το X .
 - 1b. Πάρε ένα νέο κέντρο c_i διαλέγοντας $x \in X$ με πιθανότητα $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$, όπου ως $D(x)$ ορίζεται η κοντινότερη απόσταση του κάθε σημείου από τα κέντρα που έχουν ήδη εκλεγεί.
 - 1c. Επανάλαβε το βήμα 1b μέχρι να επιλεγούν k κέντρα.
- 2-5. Συνέχισε όπως στον αλγόριθμο 1.

⁴ Στον κανόνα Chamberlin-Courant, στόχος είναι η εύρεση των υποψηφίων που ελαχιστοποιούν την μη αντιπροσώπευση των ψηφοφόρων. Αντίθετα ο αλγόριθμος k -means χρησιμοποιείται για την ομαδοποίηση των δεδομένων σε k ομάδες. Επομένως, από τον αλγόριθμο k -means μπορούμε να βρούμε τους ιδανικούς υποψήφιους για τα δεδομένα μας, και κατά συνέπεια απαραίτητη είναι η προσθήκη ενός βήματος για τη μετάβαση από τους ιδανικούς στους υπάρχοντες υποψήφιους.

Τα αποτελέσματα στα συμμετρικά τεστ ήταν και πάλι ισοδύναμα, ωστόσο στα ειδικά σχεδιασμένα τεστ μας υπήρξε εκθετική βελτίωση (εικόνα 4.1c (c)). Στις εικόνες 4.2 (a),(b) και (c) είναι επίσης φανερή η σημασία της αρχικοποίησης των κέντρων.



Εικόνα 4.2: Αποτελέσματα για τον ευρετικό αλγόριθμο του Chamberlin-Courant πριν και μετά την αρχικοποίηση του k -means++

4.1.3 Προτεινόμενος Ευρετικός Αλγόριθμος

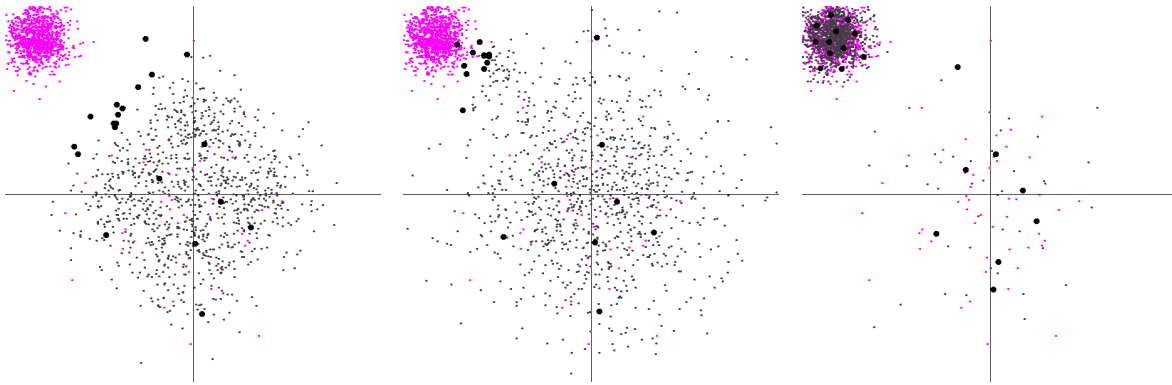
Ο προτεινόμενος ευρετικός αλγόριθμος, ακολουθεί (αλγόριθμος 3).

Αλγόριθμος 3 Προτεινόμενος ευρετικός αλγόριθμος του κανόνα Chamberlin-Courant

- 1a. Πάρε τυχαία ένα κέντρο c_1 από το X .
 - 1b. Πάρε ένα νέο κέντρο c_i διαλέγοντας $x \in X$ με πιθανότητα $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$, όπου ως $D(x)$ ορίζεται η κοντινότερη απόσταση του κάθε σημείου από τα κέντρα που έχουν ήδη εκλεγεί.
 - 1c. Επανάλαβε το βήμα 1b μέχρι να επιλεγούν k κέντρα.
 3. Για κάθε $i \in \{1, \dots, k\}$, θέσε c_i το κέντρο της μάζας όλων των σημείων στο C_i : $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.
 4. Επανάλαβε τα βήματα 2 και 3 έως ότου το C να μην αλλάζει.
 5. Βρες τους κοντινότερους υποψηφίους στα εκλεγμένα κέντρα των clusters.
-

4.1.4 Αδυναμία του Προτεινόμενου Αλγορίθμου

Ο προτεινόμενος αλγόριθμος βρίσκει τα ιδανικά κέντρα, ανεξάρτητα από την κατανομή των υποψηφίων και, στο τέλος μόνο, βρίσκει τους κοντινότερους υποψηφίους στα ιδανικά κέντρα. Στην περίπτωση που οι υποψήφιοι καλύπτουν τον χώρο που βρίσκονται οι ψηφοφόροι, τα αποτελέσματα τείνουν στα ιδανικά. Σε αντίθετη περίπτωση όμως τα αποτελέσματα μπορεί να διαφέρουν σημαντικά, καθώς τα εκλεγμένα μέλη της επιτροπής, αντί να έχουν κατανεμηθεί στην περιοχή που επικαλύπτονται οι δύο κατανομές, τοποθετούνται σε κοντινά σημεία, προσπαθώντας να πλησιάσουν τα ιδανικά σημεία. Αυτό το φαινόμενο γίνεται περισσότερο κατανοητό στις εικόνες 4.3.



Εικόνα 4.3: Αδυναμία του προτεινόμενου ευρετικού αλγορίθμου του κανόνα Chamberlin-Courant

4.1.5 Υπολογιστική Πολυπλοκότητα

Καθώς ο κανόνας Chamberlin-Courant ανήκει υπολογιστικά στο NP και ο ευρετικός μας αλγόριθμος βασίζεται σε ένα αλγόριθμο ομαδοποίησης, δεν υπάρχουν εγγυήσεις για την υπολογιστική του πολυπλοκότητα, καθώς ο αριθμός των επαναλήψεων που θα γίνουν μέχρι τη σύγκλιση του αλγορίθμου ποικίλουν. Ωστόσο, μπορούμε να υπολογίσουμε την υπολογιστική πολυπλοκότητα κάθε βήματος, χωρίς να λάβουμε υπόψιν τον αριθμό των επαναλήψεων. Ας συμβολίσουμε με k τον αριθμό των μελών της επιτροπής, με v τον αριθμό των ψηφοφόρων και με c τον αριθμό των υποψηφίων. Στο πείραμα μας επιπλέον ισχύει ότι $k \ll v$ και $v \approx c$.

Για τα βήματα 1a - 1c (αρχικοποίηση) έχουμε:

$$\mathcal{O}(k \cdot v) = \mathcal{O}(v)$$

Για τα βήματα 2 - 4 έχουμε:

$$\mathcal{O}(v + k \cdot v + v + k + k^2) = \mathcal{O}((2 + k) \cdot v + k + k^2) = \mathcal{O}(v)$$

Τέλος, για το βήμα 5 (εύρεση των μελών της επιτροπής) έχουμε:

$$\mathcal{O}(k \cdot c) = \mathcal{O}(k \cdot v) = \mathcal{O}(v)$$

4.2 Κανόνας Μονroe

4.2.1 Ευρετικός Αλγόριθμος για τον Κανόνα Μονroe

Στον κανόνα Μονroe, όπως περιγράφηκε στην ενότητα 2.4, εκτός από τις απαιτήσεις του κανόνα Chamberlin-Courant, υπάρχει επιπλέον η απαίτηση για την ισοζυγισμένη αντιστοίχιση του κάθε ψηφοφόρου με τα μέλη της επιτροπής.

Λόγω των ομοιοτήτων των δύο κανόνων, αποφασίσαμε να βασίσουμε και τον ευρετικό αλγόριθμο για τον κανόνα Μονroe στον αλγόριθμο k -means. Προκειμένου να διορθώσουμε τις αναλογίες των ομάδων, θέτουμε στον αλγόριθμο τον παρακάτω περιορισμό: κάθε ομάδα πρέπει να έχει ακριβώς $\frac{n}{k}$ μέλη και κανένας ψηφοφόρος δεν μπορεί να ανήκει σε παραπάνω από μία ομάδες.

Για την αντιστοίχιση των ψηφοφόρων με τους αντιπροσώπους τους, χρησιμοποιήσαμε μία ουρά από συνδυασμούς $\{voter, committee\ position, distance\}$, ταξινομημένους σύμφωνα με την απόσταση. Κάθε ψηφοφόρος αντιστοιχίζεται με τον αντιπρόσωπο (κέντρο της ομάδας) που βρίσκεται κοντινότερα, με την προϋπόθεση να μην ανήκει ήδη σε κάποιο cluster και να υπάρχει διαθέσιμη θέση. Ο αλγόριθμος που χρησιμοποιήθηκε ήταν ο ακόλουθος (αλγόριθμος 4).

Αλγόριθμος 4 Ευρετικός αλγόριθμος του κανόνα Μονroe

1. Τυχαία διάλεξε ως αρχικά k κέντρα $C = \{c_1, c_2, \dots, c_k\}$.
2. Για κάθε $i \in \{1, \dots, k\}$, θέσε το cluster C_i ως τα $\frac{n}{k}$ σημεία στο X που βρίσκονται κοντινότερα στο c_i από ότι στο c_j για κάθε $j \neq i$.
3. Για κάθε $i \in \{1, \dots, k\}$, θέσε c_i να είναι το κέντρο της μάζας όλων των σημείων στο C_i :
$$c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x.$$
4. Επανάλαβε τα βήματα 2 και 3 έως ότου το C να μην αλλάζει.
5. Βρες τους κοντινότερους υποψηφίους στα εκλεγμένα κέντρα των clusters.

Τα αρχικά κέντρα επιλέχθηκαν και πάλι τυχαία από το X και οι ισοπαλίες επιλύονταν αυθαίρετα με προεπιλεγμένο τρόπο. Υπάρχουν τρία σημαντικά ερωτήματα γύρω από τον προτεινόμενο αλγόριθμο.

1. Υπάρχει βέλτιστη τεχνική αρχικοποίησης;
2. Ποιος είναι ο αποδοτικότερος τρόπος αντιστοίχισης των ψηφοφόρων στα cluster;
3. Συγκλίνει πάντοτε ο αλγόριθμος;

4.2.1.1 Σύγκλιση του Προτεινόμενου Αλγορίθμου

Όσον αφορά τη σύγκλιση του αλγορίθμου, αυτή δεν είναι εξασφαλισμένη. Παρατηρήσαμε, ωστόσο, ότι ο μόνος τρόπος για να μην συγκλίνει ο αλγόριθμος είναι η ύπαρξη ατέρμονος βρόχου. Αυτό συμβαίνει όταν μία ακολουθία από πιθανές επιτροπές επαναλαμβάνεται επ' άπειρον και οφείλεται στη διακριτή φύση του προβλήματος: ακριβώς $\frac{n}{k}$ ψηφοφόροι πρέπει να αντιστοιχιστούν σε κάθε ομάδα σε κάθε επανάληψη και ένας ή περισσότεροι ψηφοφόροι αλλάζουν ομάδες, μόνο για να καταλήξουν και πάλι στις αρχικές θέσεις μετά από ορισμένες επαναλήψεις.

Υπάρχουν δύο τρόποι για την επίλυση του προβλήματος, είτε να επιλέξουμε την επιτροπή με τη μικρότερη μη αντιπροσώπευση από αυτές που επαναλαμβάνονται είτε να βρούμε το αληθινό σημείο σύγκλισης. Αποφασίσαμε να ακολουθήσουμε τη δεύτερη κατεύθυνση και να λύσουμε το πρόβλημα μεταθέτοντας το στον συνεχή χώρο, αντιστοιχίζοντας με άπληστο τρόπο τα κέντρα των ομάδων από τις διαφορετικές επιτροπές που επαναλαμβάνονται και βρίσκοντας το κέντρο βάρους τους. Η επιτροπή που καταλήγαμε αποτελούσε σημείο σύγκλισης, καθώς αν ξεκινούσαμε τον αλγόριθμο από εκείνα τα σημεία βλέπαμε ότι είχε ήδη συγκλίνει.

4.2.1.2 Αντιστοίχιση των Ψηφοφόρων στις Ομάδες

Δοκιμάσαμε δύο διαφορετικές προσεγγίσεις που θυμίζουν την αναζήτηση κατά πλάτος και κατά βάθος. Στη συνέχεια, παρουσιάζονται οι δύο προσεγγίσεις, καθώς και τα πλεονεκτήματα και μειονεκτήματα της καθεμίας.

Και στις δύο προσεγγίσεις, για κάθε κέντρο των ομάδων c_i υπάρχει μία ταξινομημένη λίστα των ψηφοφόρων σε αύξουσα σειρά σύμφωνα με την απόστασή τους από το c_i . Αυτό που διαφέρει είναι ο μηχανισμός που περιορίζει τον αριθμό των υποψηφίων που αντιστοιχίζονται σε κάθε ομάδα τη φορά. Τα κέντρα των ομάδων μπορούν να περιγραφούν ως μέλη μίας ουράς, που αναμένουν να πάρουν το lock για να επιλέξουν ψηφοφόρους.

- **Προσέγγιση κατά πλάτος (Breadth-first approach)**

Σε αυτή την προσέγγιση, μόλις ένα κέντρο πάρει το lock, διαλέγει τον κοντινότερο ψηφοφόρο που δεν έχει ήδη αντιστοιχιστεί σε κάποια ομάδα, απελευθερώνει το lock και τοποθετείται ξανά στο τέλος της ουράς περιμένοντας να αποκτήσει το lock.

- **Προσέγγιση κατά βάθος (Depth-first approach)**

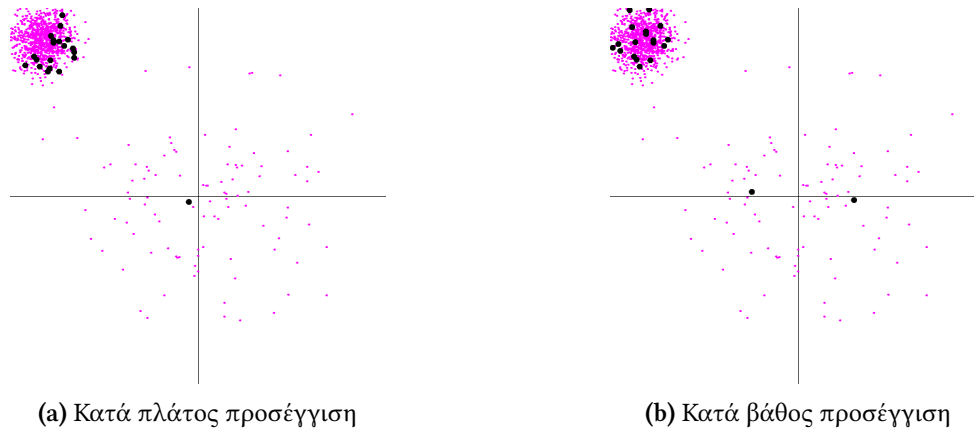
Σε αυτή την προσέγγιση, μόλις ένα κέντρο πάρει το lock, επιλέγει τους κοντινότερους $\frac{n}{k}$ ψηφοφόρους που δεν έχουν αντιστοιχιστεί σε κάποια ομάδα και τότε απελευθερώνει το lock και αφαιρείται από την ουρά.

Η κατά πλάτος προσέγγιση έχει καλά αποτελέσματα για τις περισσότερες εισόδους, ωστόσο παρουσιάζει αδυναμία για εισόδους που αποτελούνται από συνδυασμούς κατανομών, με ένα μικρό μέρος της εισόδου να έχει παραχθεί από αραιή κατανομή με μεγάλη διασπορά. Αντιθέτως, η κατά βάθος προσέγγιση, για ιδιαίτερες εισόδους, όπως αυτή που περιγράφηκε παραπάνω, εμφανίζει στατιστικά καλύτερα αποτελέσματα. Ωστόσο, στην πληθώρα των εισόδων και ειδικά σε απλές και συμμετρικές, αποτυγχάνει.

Τα αποτελέσματα, όποια προσέγγιση και αν χρησιμοποιήσουμε εξαρτώνται από την αρχικοποίηση των κέντρων. Επομένως μπορούμε να βελτιώσουμε τις πιθανότητες επιτυχίας της κατά πλάτος προσέγγισης στις ειδικές εισόδους, αν βρούμε τα κατάλληλα σημεία αρχικοποίησης. Έτσι, διαλέγουμε την κατά πλάτος προσέγγιση.

4.2.1.3 Τεχνική Αρχικοποίησης

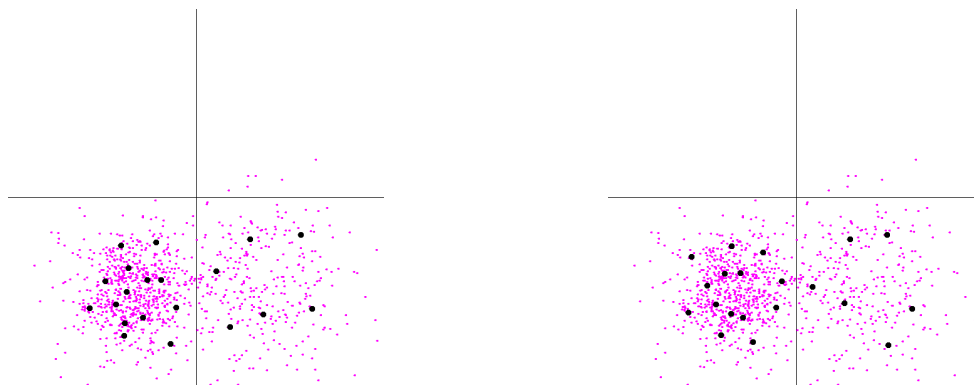
Όπως και στον ευρετικό αλγόριθμο του Chamberlin-Courant, έτσι και στον ευρετικό αλγόριθμο του Monge, διαφορετικά σημεία αρχικοποίησης οδηγούν σε διαφορετικά αποτελέσματα.



Εικόνα 4.4: Προτεινόμενες επιτροπές διαφορετικών προσεγγίσεων του ευρετικού αλγόριθμου του Μονroe για τους ίδιους ψηφοφόρους, υποψηφίους και σημεία αρχικοποίησης 10% από τους ψηφοφόρους παρήχθησαν με συμμετρική γκαουσιανή κατανομή με μέση τιμή (0,0) και απόκλιση 1 και 90% από συμμετρική γκαουσιανή κατανομή με μέση τιμή (-2.5,2,5) και απόκλιση 0.25

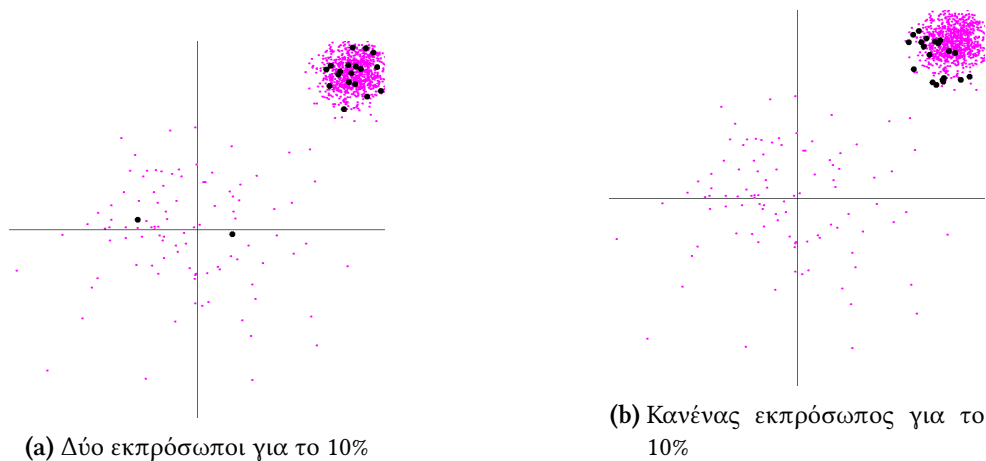
Σημαντικό είναι να εντοπίσουμε τις περιπτώσεις όπου τα αποτελέσματα δεν είναι ισοδύναμα. Παρατηρούμε ότι παρόμοια τεστ με αυτά που χρησιμοποιήθηκαν στον κανόνα Chamberlin-Courant, μπορούν να χρησιμοποιηθούν και για την μελέτη του Μονroe.

[1] **Σύνθετο παράδειγμα με ισοδύναμα αποτελέσματα.** Στο παρακάτω παράδειγμα οι ψηφοφόροι έχουν παραχθεί από μία συμμετρική γκαουσιανή κατανομή με μέση τιμή (1, -1.5) και απόκλιση 0.75 κατά 30% και κατά 70% από μία συμμετρική γκαουσιανή κατανομή με μέση τιμή (-1, -1,5) και απόκλιση 0.5.



Εικόνα 4.5: Ισοδύναμα αποτελέσματα του ευρετικού αλγόριθμου του κανόνα [1]

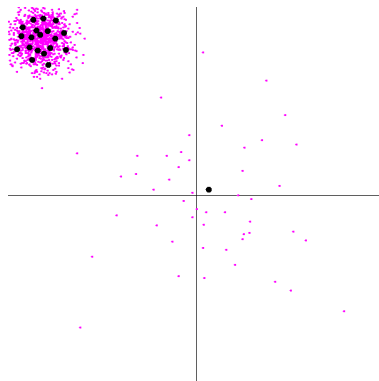
[2] **Σύνθετο παράδειγμα με μη ισοδύναμα αποτελέσματα.** Στο παρακάτω παράδειγμα οι ψηφοφόροι παράγονται από μία συμμετρική γκαουσιανή κατανομή με μέση τιμή (0,0) και απόκλιση 1 κατά 10% και κατά 90% από μία συμμετρική γκαουσιανή κατανομή με μέση τιμή (2.5, 2,5) και απόκλιση 0.25. Επομένως, το 10% των ψηφοφόρων αντιστοιχούν στην αραιή κατανομή, και ως αποτέλεσμα δύο ομάδες (δύο κέντρα) θα έπρεπε να αντιστοιχιστούν σε αυτούς τους ψηφοφόρους. Τα σημεία αρχικοποίησης μπορούν να οδηγήσουν σε 0 έως 2 εκπρόσωπους για αυτή την κατανομή.



Εικόνα 4.6: Μη ισοδύναμα αποτελέσματα του ευρετικού αλγόριθμου του κανόνα Monroe [2]

4.2.2 Διαφορετικές Τεχνικές Αρχικοποίησης

Προκειμένου να βρούμε την ιδανική αρχικοποίηση για τον αλγόριθμο μας επικεντρωθήκαμε στην παρακάτω είσοδο (εικόνα 4.7): οι ψηφοφόροι παράγονται από μία συμμετρική γκαουσιανή κατανομή με μέση τιμή (0,0) και απόκλιση 1 κατά 5% και κατά 95% από μία συμμετρική γκαουσιανή κατανομή με μέση τιμή (-2.5, 2.5) και απόκλιση 0.25. Τα ιδανικά αποτελέσματα αντιστοιχίζουν έναν ακριβώς εκπρόσωπο στην αραιή κατανομή και τους υπόλοιπους 19 στην άλλη.



Εικόνα 4.7: Βέλτιστα αποτελέσματα του ευρετικού αλγόριθμου του κανόνα Monroe

Αρχικά, τα κέντρα επιλέγονταν τυχαία. Αυτό οδηγούσε σε 50% πιθανότητα να εκλεγεί ένας εκπρόσωπος για την αραιή κατανομή. Προσπαθήσαμε λοιπόν να βελτιώσουμε τα ποσοστά επιτυχίας, κάνοντας χρήση των παρακάτω τεχνικών αρχικοποίησης.

4.2.2.1 Τεχνική Αρχικοποίησης k -means++

Αρχικά δοκιμάσαμε την τεχνική αρχικοποίησης του αλγόριθμου k -means++ όπως περιγράφηκε στην ενότητα 4.1, για να επιβεβαιώσουμε πως δεν είναι κατάλληλη για αυτόν τον αλγόριθμο. Μετά από δοκιμές, παρατηρήσαμε ότι βελτιώνονταν τα ποσοστά επιτυχίας για την κατά βάθος προσέγγιση, χωρίς ωστόσο να υπάρχει μεγάλη βελτίωση στη μέση περίπτωση. Αντίθετα, για την προσέγγιση κατά πλάτος, που τελικά αποφασίσαμε να ακολουθήσουμε, τα ποσοστά επιτυχίας πρακτικά μηδενίστηκαν. Επομένως, για τον αλγόριθμό μας, μπορούμε να εγκαταλείψουμε τη συ-

γκεκριμένη τεχνική.

Αν ως $D(x)$ ορίσουμε το άθροισμα των αποστάσεων από ένα σημείο και τα επιλεγμένα κέντρα και χρησιμοποιήσουμε το $D(x)$ αντί του $D(x)^2$ τα ποσοστά ανεβαίνουν στο 20%, δηλαδή αρκετά μικρό ποσοστό και ως αποτέλεσμα εγκαταλείπουμε και αυτή την τεχνική.

4.2.2.2 Προσαρμογή της Αρχικοποίησης του Αλγορίθμου k -means++ για τον Ευρετικό Αλγόριθμο του Μονroe

Όπως είναι λογικό, η αρχικοποίηση του k -means++ ήταν κατάλληλη για τον ευρετικό αλγόριθμο του Chamberlin-Courant, εξαιτίας της φύσης του στόχου της αρχικοποίησης (πρέπει να επιτύχει τη μέγιστη διασπορά των αρχικών κέντρων). Ωστόσο, για τον ευρετικό αλγόριθμο του Μονroe, ο στόχος της αρχικοποίησης διαφέρει, καθώς πρέπει να επιτύχει τη μέγιστη διασπορά των αρχικών κέντρων **αναλογικά** με τον αριθμό των μελών της ομάδας.

Επομένως, τροποποιήσαμε την αρχικοποίηση του k -means++ με τον ακόλουθο τρόπο. Μετά την επιλογή του πρώτου κέντρου, τα $\frac{n}{k} - 1$ κοντινότερα σημεία αφαιρούνται από το X και ορίζονται ως τα μέλη της ομάδας. Ακολουθεί μία περιγραφή του αλγορίθμου με την παραπάνω αρχικοποίηση (αλγόριθμος 5).

Αλγόριθμος 5 Ευρετικός αλγόριθμος του κανόνα Μονroe (τροποποιημένη αρχικοποίηση)

- 1a. Πάρε ένα κέντρο c_1 ομοιόμορφα τυχαία από το X και αφάιρεσε από το X τα κοντινότερα $\frac{n}{k} - 1$ σημεία.
 - 1b. Πάρε ένα νέο κέντρο c_i διαλέγοντας $x \in X$ με πιθανότητα $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$, όπου $D(x)$ η κοντινότερη απόσταση του κάθε σημείου από τα κέντρα που έχουν ήδη εκλεγεί.
 - 1c. Αφαίρεσε από το X τα κοντινότερα $\frac{n}{k} - 1$ σημεία.
 - 1d. Επανάλαβε τα βήματα 1b-1c μέχρι να επιλεγούν k κέντρα.
- 2-5. Συνέχισε όπως στον αλγόριθμο 4

Με την τεχνική αυτή, επιτύχαμε 72% ποσοστά επιτυχίας για τη συγκεκριμένη είσοδο, ωστόσο υπήρχαν ακόμα περιθώρια βελτίωσης.

4.2.2.3 Βελτιωμένη Προσαρμογή της Αρχικοποίησης του Αλγορίθμου k -means++ για τον Ευρετικό Αλγόριθμο του Μονroe

Παρατηρήσαμε ότι στις περισσότερες περιπτώσεις, τα εκλεγμένα κέντρα δεν αποτελούν τους «πραγματικούς» εκπρόσωπους της κάθε ομάδας, καθώς βρίσκονται μακριά από το κέντρο μάζας. Έτσι αποφασίσαμε προσθέσουμε ένα επιπλέον βήμα στον αλγόριθμο 5, για να βελτιώσουμε ακόμα περισσότερο τα αποτελέσματα (βλ. αλγόριθμο 6). Με αυτή την τεχνική αρχικοποίησης τα ποσοστά επιτυχίας έφτασαν το 92%.

Αλγόριθμος 6 Ευρετικός αλγόριθμος του κανόνα Monroe (βελτιωμένη τροποποιημένη αρχικοποίηση)

- 1a. Πάρε ένα κέντρο c_1 ομοιόμορφα τυχαία από το X , βρες το κέντρο μάζας από τα κοντινότερα $\frac{n}{k}$ σημεία (εμπεριέχοντας το c_1) και αντικατέστησε το c_1 με το κοντινότερο σημείο στο κέντρο μάζας. Αφαίρεσε τα σημεία από το X .
 - 1b. Πάρε ένα νέο κέντρο c_i διαλέγοντας $x \in X$ με πιθανότητα $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$, όπου $D(x)$ η κοντινότερη απόσταση του κάθε σημείου από τα κέντρα που έχουν ήδη εκλεγεί.
 - 1c. Βρες το κέντρο μάζας από τα κοντινότερα $\frac{n}{k}$ σημεία (εμπεριέχοντας το c_i). Αντικατέστησε το c_i , αν χρειάζεται, με το κοντινότερο σημείο του cluster με το κέντρο μάζας.
 - 1d. Αφαίρεσε από το X τα μέλη του cluster.
 - 1e. Επανάλαβε τα βήματα 1b-1d μέχρι να επιλεγούν k κέντρα.
- 2-5. Συνέχισε όπως στον αλγόριθμο 4
-

4.2.2.4 Παρόμοιες Τεχνικές Αρχικοποίησης

Στην συγκεκριμένη είσοδο, πετύχαμε ποσοστά επιτυχίας 92% με την αρχικοποίηση που περιγράφεται στον αλγόριθμο 6. Μπορούμε να επιτύχουμε αντίστοιχα ποσοστά με παρόμοιες τεχνικές, για παράδειγμα αν ορίσουμε το $D(x)$ ως το άθροισμα των αποστάσεων του ψηφοφόρου με τα κέντρα που έχουν ήδη εκλεγεί ή/και χρησιμοποιήσουμε το $D(x)$ αντί για $D(x)^2$, ακολουθώντας τα βήματα του αλγορίθμου 6.

4.2.3 Προτεινόμενος Αλγόριθμος

Για να ολοκληρώσουμε τον προτεινόμενο αλγόριθμο προσθέτουμε ένα ακόμα βήμα για την περίπτωση εντοπισμού ατέρμονος βρόχου και ενσωματώνουμε την κατά πλάτος προσέγγιση. Ο αλγόριθμος 7 αποτελεί τον προτεινόμενο αλγόριθμο.

4.2.4 Αδυναμία του Προτεινόμενου Αλγορίθμου

Στον ευρετικό αλγόριθμο του Monro, αδυναμία αποτελεί ο εντοπισμός των μειονοτήτων και η αντιστοίχισή τους σε έναν αντιπρόσωπο. Στις παραπάνω ενότητες, επικεντρωθήκαμε σε ένα τεστ για να βελτιώσουμε τον αλγόριθμό μας. Ωστόσο, σε παρεμφερείς εισόδους, καθώς η αραιή (95%) και πυκνή (5%) κατανομή αρχίζουν και επικαλύπτονται, τα ποσοστά επιτυχίας διαφέρουν και συγκεκριμένα όσο περισσότερο επικαλύπτονται τόσο περισσότερο μειώνονται τα ποσοστά επιτυχίας. Στο σημείο αυτό σημειώνεται ότι καθώς η μέση τιμή της μία κατανομής πλησιάζει την άλλη, η ανάγκη διάκρισης των κατανομών μειώνεται και τελικά μηδενίζεται όταν οι μέσες τιμές ταυτιστούν.

4.2.5 Υπολογιστική Πολυπλοκότητα

Καθώς ο κανόνας Monro ανήκει υπολογιστικά στο NP και ο ευρετικός μας αλγόριθμος βασίζεται σε έναν αλγόριθμο ομαδοποίησης, δεν υπάρχουν εγγυήσεις για την υπολογιστική του πολυ-

Αλγόριθμος 7 Προτεινόμενος ευρετικός αλγόριθμος του κανόνα Monroe

- 1a. Πάρε ένα κέντρο c_1 ομοιόμορφα τυχαία από το X , βρες το κέντρο μάζας από τα κοντινότερα $\frac{n}{k}$ σημεία (εμπεριέχοντας το c_1) και αντικατέστησε το c_1 με το κοντινότερο σημείο στο κέντρο μάζας. Αφαίρεσε τα σημεία από το X .
- 1b. Πάρε ένα νέο κέντρο c_i διαλέγοντας $x \in X$ με πιθανότητα $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$, όπου $D(x)$ η κοντινότερη απόσταση του κάθε σημείου από τα κέντρα που έχουν ήδη εκλεγεί.
- 1c. Βρες το κέντρο μάζας από τα κοντινότερα $\frac{n}{k}$ σημεία (εμπεριέχοντας το c_i). Αντικατέστησε το c_i , αν χρειάζεται, με το κοντινότερο σημείο του cluster με το κέντρο μάζας.
- 1d. Αφαίρεσε από το X τα μέλη του cluster.
- 1e. Επανέλαβε τα βήματα 1b-1d μέχρι να επιλεγούν k κέντρα.
2. Για κάθε $i \in \{1, \dots, k\}$, θέσε το cluster C_i ως τα $\frac{n}{k}$ σημεία στο X που βρίσκονται κοντινότερα στο c_i από ότι στο c_j για κάθε $j \neq i$ σύμφωνα με την κατά πλάτος προσέγγιση.
3. Για κάθε $i \in \{1, \dots, k\}$, θέσε c_i να είναι το κέντρο της μάζας όλων των σημείων στο C_i :
$$c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x.$$
4. Επανάλαβε τα βήματα 2 και 3 έως ότου το C να μην αλλάζει.
5. Σε περίπτωση βρόχου, αντιστοίχισε τα κέντρα με άπληστο τρόπο και βρες το κέντρο μάζας όλων των c_{ij} : $c_i = \frac{1}{j} \sum_{j \in loop} c_{ij}$.
6. Βρες τους κοντινότερους υποψηφίους στα εκλεγμένα κέντρα των clusters.

πλοκότητα, καθώς ο αριθμός των επαναλήψεων που θα γίνουν μέχρι τη σύγκλιση του αλγορίθμου ποικίλουν. Ωστόσο, μπορούμε να υπολογίσουμε την υπολογιστική πολυπλοκότητα κάθε βήματος, χωρίς να λάβουμε υπόψιν τον αριθμό των επαναλήψεων. Ας συμβολίσουμε με k τον αριθμό των μελών της επιτροπής, με v τον αριθμό των ψηφοφόρων και με c τον αριθμό των υποψηφίων. Στο πείραμά μας επιπλέον ισχύει ότι $k \ll v$ και $v \approx c$.

Για τα βήματα 1a - 1e (αρχικοποίηση) έχουμε:

$$\mathcal{O}(k \cdot v) = \mathcal{O}(v)$$

Για τα βήματα 2 - 4 έχουμε:

$$\mathcal{O}(v + k \cdot v + k \cdot v \cdot \log(v) + k + v + k^2) = \mathcal{O}((2 + k + \log(v)) \cdot v + k + k^2) = \mathcal{O}(v \cdot \log(v))$$

Για το βήμα 5 αν θεωρήσουμε πως έχει εντοπιστεί κάποιος βρόχος και x επιτροπές (καταστάσεις) επαναλαμβάνονται, έχουμε:

$$\mathcal{O}(x \cdot k)$$

Τέλος, για το βήμα 6 (εύρεση των μελών της επιτροπής):

$$\mathcal{O}(k \cdot c) = \mathcal{O}(k \cdot v) = \mathcal{O}(v)$$

Κεφάλαιο 5

Ευστάθεια

5.1 Εισαγωγή

Μία επιθυμητή ιδιότητα σε έναν κανόνα εκλογής επιτροπών είναι η ευστάθεια του αποτελέσματος. Η ανάγκης ύπαρξης σταθερής επιτροπής γίνεται εύκολα κατανοητή: μπορεί ένας συγκεκριμένος ψηφοφόρος ή η προσθήκη ψηφοφόρων να επηρεάσει το αποτέλεσμα της ψηφοφορίας; και αν ναι, σε τι βαθμό;

Η ευστάθεια του συστήματος μπορεί να ερμηνευτεί και με έναν διαφορετικό τρόπο, την αντίσταση του κανόνα ψηφοφορίας στην χειραγώγηση [20], [21]. Τα τελευταία 10 χρόνια υπάρχει συσσώρευση ενδιαφέροντος όσον αφορά την πολυπλοκότητα στην στρατηγική συμπεριφορά σε εκλογές επιτροπών. Για μερικούς κανόνες, όπως οι SNTV και Bloc [22],[23], έχει αποδειχθεί ότι η χειραγώγηση βρίσκεται στο P, ενώ για άλλους, όπως ο STV [24], υπάρχουν αποδείξεις ότι υπολογιστικά αντιστέκεται στη χειραγώγηση. Προσοχή έχει δοθεί και στο θέμα της αντίστασης στον έλεγχο [22], [23], [25], [26].

Επιλέξαμε να κάνουμε μία εκτενή πειραματική μελέτη των κανόνων που μελετήσαμε στα ακόλουθα δύο σενάρια:

1. Την ελεγχόμενη μετακίνηση ενός τυχαίου ψηφοφόρου (χειραγώγηση από έναν ψηφοφόρο)
2. Την προσθήκη ενός νέου ψηφοφόρου (έλεγχος από έναν ψηφοφόρο)

Σκοπός μας είναι να μελετήσουμε αν το αποτέλεσμα ενός κανόνα είναι πρακτικά χειραγώγησιμο στις παραπάνω περιπτώσεις, και αν ναι, κάτω από ποιες προϋποθέσεις. Μέσω αυτής της μελέτης χαρακτηρίζουμε τους κανόνες ως ευσταθείς ή ασταθείς σύμφωνα με τους όρους του πειράματός μας. Ερμηνεύουμε επίσης τα αποτελέσματά μας με άλλους τρόπους, για παράδειγμα μελετάμε αν σε περίπτωση αστάθειας μπορούμε να εξάγουμε πληροφορίες για τον ψηφοφόρο που προκάλεσε την αλλαγή, αν η αστάθεια του αποτελέσματος μπορεί να προβλεφθεί και επομένως να θεωρηθεί στρατηγική συμπεριφορά ψηφοφόρου ή αν είναι μη αναμενόμενη, κλπ.

5.2 Ελεγχόμενη Μετακίνηση ενός Ψηφοφόρου

Μετρώντας την ευστάθεια του συστήματος, πρακτικά προσπαθούμε να μελετήσουμε πώς η μετακίνηση ενός συγκεκριμένου ψηφοφόρου μπορεί να επηρεάσει το αποτέλεσμα, αν δεν προστεθεί ή αφαιρεθεί κάποιος παίκτης από την εκλογική διαδικασία. Ωστόσο η κίνηση ενός ψηφοφόρου δεν μπορεί να είναι απεριόριστη και πρέπει να προσδιορίσουμε την προθυμία του ψηφοφόρου να

μετακινηθεί από την πραγματική του θέση και κατά συνέπεια από τα πραγματικά του συμφέροντα και την εκπροσώπησή του.

5.2.1 Πειραματική Διάταξη

Η διάταξη του πειράματος, αποτελείται και πάλι από ένα σύνολο 1000 ψηφοφόρων και 1000 υποψηφίων που έχουν παραχθεί σύμφωνα με κάποια κατανομή ως σημεία στο δισδιάστατο χώρο. Για κάθε ψηφοφόρο υπάρχει η σειρά κατάταξης για όλους τους υποψήφιους και τελικά εκλέγεται μία επιτροπή με 20 μέλη με κάθε κανόνα. Στη συνέχεια διαλέγουμε έναν τυχαίο ψηφοφόρο p και τον μετακινούμε μέσα στα εκάστοτε όρια προς τυχαία κατεύθυνση. Επαναλαμβάνουμε το πείραμα και στη συνέχεια συγκρίνουμε τα αποτελέσματα.

Η κίνηση του τυχαίου ψηφοφόρου p είναι ελεγχόμενη, διότι σε αντίθετη περίπτωση θα ήταν σαν να έχουμε έναν τελείως διαφορετικό υποψήφιο. Επομένως ο p κινείται ελεύθερα προς τυχαία κατεύθυνση μέσα στα ακόλουθα όρια:

$$0 < movement < \frac{distance(c_2) - distance(c_1)}{2},$$

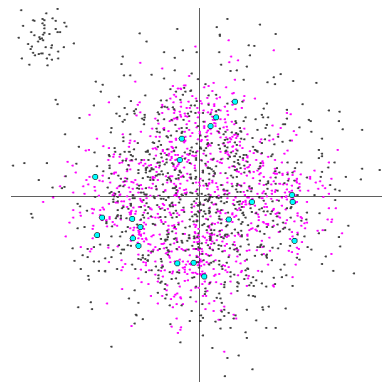
δηλαδή η μετακίνηση δεν μπορεί να ξεπερνά το μισό της διαφοράς των αποστάσεων του p με τις δύο πρώτες προτιμήσεις του, έτσι ώστε να εξασφαλίσουμε ότι η πρώτη του προτίμηση παραμένει σταθερή. Πρέπει ωστόσο να είμαστε προσεκτικοί με αυτόν τον περιορισμό στην ερμηνεία των αποτελεσμάτων μας, καθώς διατηρούμε τις θέσεις των υποψηφίων σταθερές και επομένως η διαφοροποίηση του ψηφοφόρου εξαρτάται από τη θέση των υποψηφίων.

5.2.2 Αποτελέσματα

Στις εικόνες που ακολουθούν οι ψηφοφόροι συμβολίζονται με μοβ κουκκίδες ενώ οι υποψήφιοι με μαύρες. Τα μέλη της επιτροπής του αρχικού πειράματος συμβολίζονται με μεγάλες μαύρες κουκκίδες, ενώ μετά τη μετακίνηση του τυχαίου ψηφοφόρου με γαλάζιες. Αν οι δύο επιτροπές ταυτίζονται όλες οι γαλάζιες κουκκίδες παρουσιάζονται με μαύρο περίγραμμα. Σε αντίθετη περίπτωση, για κάθε γαλάζια κουκκίδα χωρίς περίγραμμα υπάρχει μία αντίστοιχη μαύρη.

5.2.2.1 Κανόνας Single Nontransferable Vote

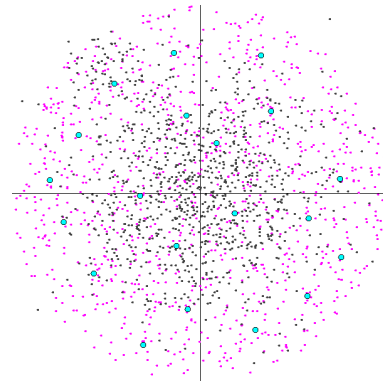
Όπως ήταν αναμενόμενο, το αποτέλεσμα του κανόνα Single Nontransferable Vote δεν επηρεάζεται από μία τέτοια αλλαγή, καθώς ο κανόνας λαμβάνει υπόψη μόνο την πρώτη προτίμηση των ψηφοφόρων.



Εικόνα 5.1: Κανόνας SNTV - ευσταθής

5.2.2.2 Κανόννας Single Transferable Vote

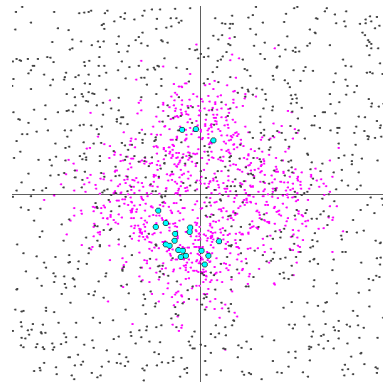
Ο κανόννας Single Transferable Vote rule είναι ευαίσθητος σε οποιαδήποτε αλλαγή. Παρόλα αυτά, μία τέτοια αλλαγή, σπάνια επηρεάζει στο αποτέλεσμα και κατά συνέπεια μπορούμε να χαρακτηρίσουμε τον κανόνα ως ευσταθή. Ωστόσο, αν γινόταν έστω και μία αλλαγή στη σειρά επιλογής ή απόρριψης των υποψηφίων, τα αποτελέσματα μπορεί να διέφεραν ολότελα.



Εικόνα 5.2: Κανόννας STV - ευσταθής

5.2.2.3 Κανόννας Bloc

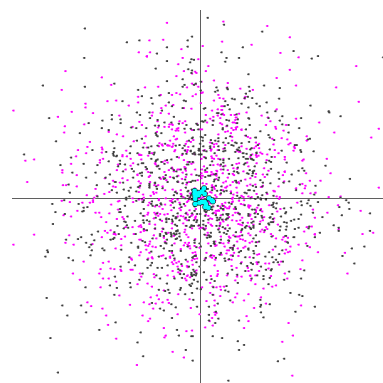
Όπως ήταν αναμενόμενο, τα αποτελέσματα του κανόνα Bloc, φαίνονται επίσης ανεπηρέαστα από τη μικρή κίνηση ενός μόνο ψηφοφόρου. Η κίνηση μπορεί να προκαλέσει αλλαγές στη σειρά κατάταξης των υποψηφίων, ωστόσο σε μεγάλο βαθμό οι αλλαγές αυτές σχετίζονται με αντιμεταθέσεις υποψηφίων και ως κατά συνέπεια οι k προτιμότεροι υποψήφιοι παραμένουν σταθεροί. Ακόμα και αν κάποιος από τους προτιμότερους υποψήφιους αλλάξουν, για να επηρεαστεί το αποτέλεσμα πρέπει να υπάρξει μία δημιουργία ή λύση μιας ισοψηφίας στην «τελευταία» θέση μετά την μετακίνηση του ψηφοφόρου.



Εικόνα 5.3: Κανόννας Bloc - ευσταθής

5.2.2.4 Κανόννας k -Borda

Ο κανόννας k -Borda λαμβάνει υπόψη την πλήρη σειρά κατάταξης. Παρόλο που η μετακίνηση του ψηφοφόρου αλλάζει τη σειρά κατάταξης του ψηφοφόρου, το αποτέλεσμα δεν φαίνεται να επηρεάζεται. Η εκλεχθείσα επιτροπή αποτελείται από τους υποψήφιους που βρίσκονται πιο κοντά στο κέντρο μάζας των ψηφοφόρων. Όπως είναι αναμενόμενο λοιπόν, μια τόσο μικρή μετακίνηση ενός ψηφοφόρου δεν αρκεί για να επηρεάσει το κέντρο μάζας και και ως κατά συνέπεια να επηρεάσει το αποτέλεσμα.



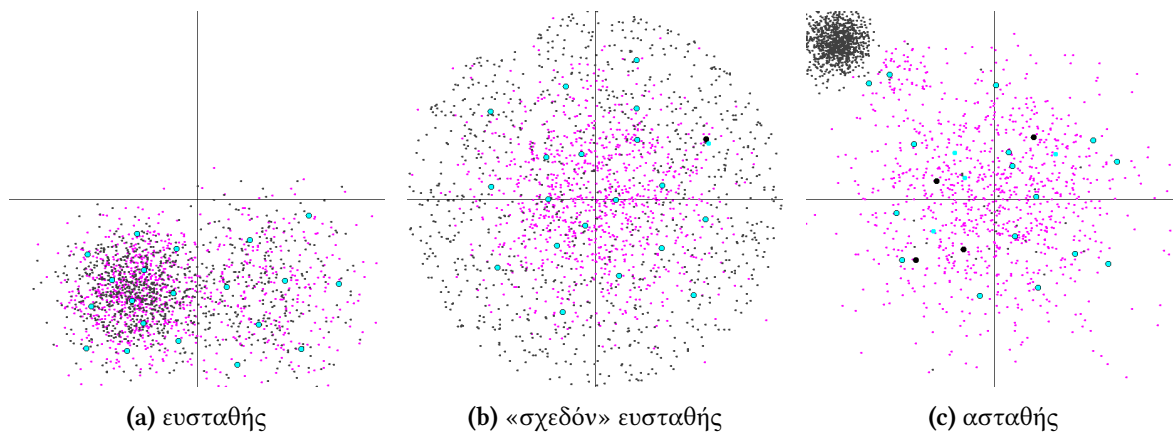
Εικόνα 5.4: Κανόννας k -Borda - ευσταθής

5.2.2.5 Κανόννας Chamberlin-Courant

Όπως είναι φυσικό, δεν μπορούμε να μιλήσουμε για ευστάθεια του κανόνα Chamberlin-Courant, αλλά για την ευστάθεια του ευρετικού μας αλγόριθμου. Τα αποτελέσματα φανερώνουν πως ο αλ-

γόριθμός μας είναι σχεδόν ευσταθής, υπό την παρακάτω προϋπόθεση: τα αρχικά κέντρα και στα δύο πειράματα είναι τα ίδια. Όπως έχει περιγραφεί και στην ενότητα 4.1, τα σημεία σύγκλισης του αλγόριθμου εξαρτώνται από τα σημεία αρχικοποίησης, και συνεπώς αν αγνοούσαμε την παραπάνω προϋπόθεση, ο αλγόριθμος θα κρινόταν ασταθής.

Σημαντικό είναι να τονιστεί η δυσκολία της εμφανούς αλλαγής στην επιτροπή. Καθώς χρησιμοποιούμε τον αλγόριθμο k -means για να βρούμε τα κέντρα των ομάδων των ψηφοφόρων, και μόνο μετά την σύγκλιση βρίσκουμε τους κοντινότερους υποψήφιους σε αυτά, προκειμένου να φανεί η αλλαγή πρέπει κάποιο κέντρο να βρεθεί πιο κοντά σε κάποιον άλλον υποψήφιο, μετά την εκτέλεση του δεύτερου πειράματος. Εξαιτίας δε, της μικρής μετακίνησης μονάχα ενός ψηφοφόρου, είναι δύσκολο να αλλάξουν οι ομάδες (clusters), αν και όχι απίθανο. Σε ορισμένες περιπτώσεις η μετακίνηση του ψηφοφόρου, όσο μικρή και να είναι, μπορεί να οδηγήσει στην μετακίνηση του σε διαφορετική ομάδα και ως κατά συνέπεια σε διαφορετικά σημεία σύγκλισης του αλγόριθμου. Για αυτό και ο αλγόριθμός μας, σαν clustering αλγόριθμος, δεν μπορεί να περιγραφεί ως «τέλεια» ευσταθής, καθώς μία μικρή αλλαγή, σε σπάνιες περιπτώσεις, μπορεί να οδηγήσει σε τελείως διαφορετικό αποτέλεσμα.



Εικόνα 5.5: Ευρετικός αλγόριθμος Chamberlin-Courant - «σχεδόν» ευσταθής

5.2.2.6 Κανόνας Monroe

Και πάλι μελετάμε την ευστάθεια του ευρετικού μας αλγόριθμου, υπό την παρακάτω προϋπόθεση: τα αρχικά κέντρα και στα δύο πειράματα να είναι ίδια. Τα αποτελέσματα δείχνουν πως ο αλγόριθμος μας είναι ασταθής ακόμα και υπό από αυτή την συνθήκη.

Όπως μπορούμε να δούμε στην εικόνα 5.6a υπάρχουν περιπτώσεις που τα αποτελέσματα δείχνουν πως ο αλγόριθμος είναι ευσταθής. Ωστόσο στην μέση περίπτωση ο αλγόριθμος είναι ασταθής (εικόνα 5.6b), διότι ακόμα και μια τόσο μικρή κίνηση ενός μόνο ψηφοφόρου, εξαιτίας του περιορισμού των μελών κάθε ομάδας, μπορεί να οδηγήσει στην μετακίνηση του ψηφοφόρου σε διαφορετική ομάδα, προκαλώντας αλυσιδωτές μεταφορές και ως κατά συνέπεια τελείως διαφορετικά σημεία σύγκλισης.

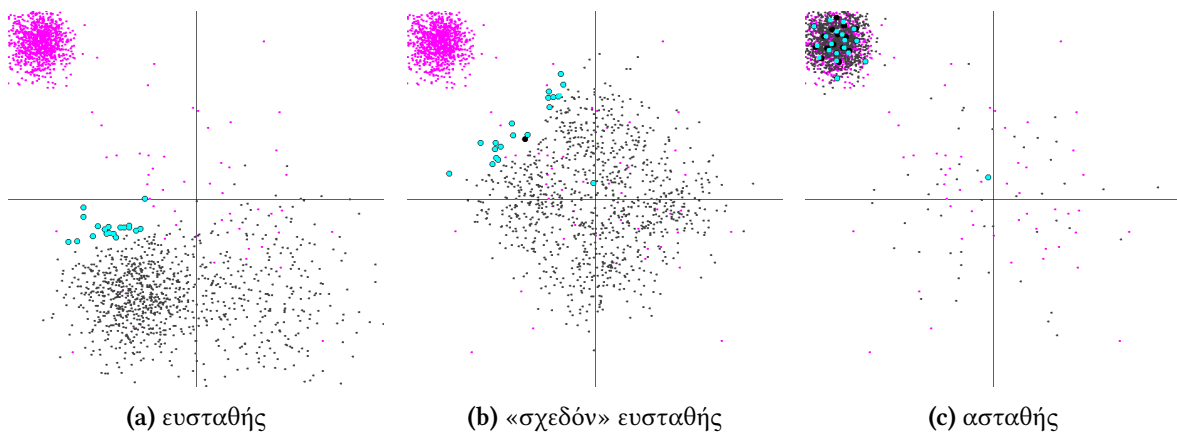
Όπως περιγράψαμε στην ανάλυση της ευστάθειας του ευρετικού αλγόριθμου για τον κανόνα Chamberlin-Courant, αν η μετακίνηση του ψηφοφόρου δεν οδηγήσει σε ανταλλαγές ψηφοφόρων στις ομάδες, προκαλείται μόνο μια μικρή μετατόπιση του κέντρου της ομάδας που ανήκει ο ψηφοφόρος, που μπορεί να οδηγήσει ή να μην οδηγήσει σε αλλαγή της επιτροπής. Αυτό εξαρτάται πλήρως από τους υποψηφίους. Αν ο υποψήφιος που βρίσκεται πιο κοντά στο νέο κέντρο, διαφέρει



Εικόνα 5.6: Ευρετικός αλγόριθμος Monroe - ασταθής

από τον προηγούμενο υποψήφιο που είχε εκλεχθεί, τότε και μόνο τότε παρατηρούμε την αλλαγή. Σημειώνεται επίσης ότι ο βαθμός που η αλλαγή θα είναι ορατή οφείλεται στην ίδια την κατανομή.

Μια ενδιαφέρουσα παρατήρηση λοιπόν είναι πως ο βαθμός που η ευστάθεια ή η αστάθεια του αποτελέσματος γίνεται αντιληπτή, εξαρτάται από τον τρόπο που οι ψηφοφόροι και υποψήφιοι κατανέμονται στον διδιάστατο χώρο.



Εικόνα 5.7: Η ευστάθεια του ευρετικού αλγόριθμου του Monroe καθώς οι κατανομές αρχίζουν να επικαλύπτονται

Όταν η κατανομή των υποψηφίων δεν καλύπτει την περιοχή που καταλαμβάνουν οι ψηφοφόροι ο αλγόριθμος φαίνεται ευσταθής (εικόνα 5.7 (a)). Καθώς ο χώρος που κατανέμονται οι υποψήφιοι τείνει να επικαλύψει τον χώρο που κατανέμονται οι ψηφοφόροι ο αλγόριθμος παρουσιάζεται λιγότερο ευσταθής (εικόνα 5.7 (b)). Όταν οι κατανομές των υποψηφίων και των ψηφοφόρων ταυτίζονται ο αλγόριθμος είναι ασταθής (εικόνα 5.7 (c)). Επομένως καταλήγουμε στο συμπέρασμα, ότι τα ευσταθή αποτελέσματα αποτελούν μεμονωμένες περιπτώσεις, οι οποίες οφείλονται στο γεγονός ότι ο ευρετικός αλγόριθμος δεν λαμβάνει υπόψη τις θέσεις των υποψηφίων για τη δημιουργία των ομάδων. Έτσι λοιπόν χαρακτηρίζουμε τον ευρετικό αλγόριθμο για τον Monroe ως ασταθής¹.

¹ Στον ευρετικό αλγόριθμο του κανόνα Chamberlin-Courant υπάρχουν αποτελέσματα τα οποία δεν είναι αρκετά ευσταθή ώστε να χαρακτηρίσουμε τον αλγόριθμο ως απόλυτα ευσταθή. Ωστόσο, δεδομένου ότι δεν εντοπίσαμε κάποιο μοτίβο όπως στον ευρετικό αλγόριθμο του Monroe, θεωρούμε πως τα ευσταθή ή «σχεδόν» ευσταθή αποτελέσματα είναι η μέση περίπτωση και τα ασταθή μεμονωμένες περιπτώσεις.

5.3 Προσθήκη ενός Νέου Ψηφοφόρου

Επίσης θέλαμε να μελετήσουμε την ευστάθεια των κανόνων, στην περίπτωση της προσθήκης ενός νέου υποψηφίου. Αυτή η περίπτωση μπορεί επίσης να καλύψει τη μη ελεγχόμενη μετακίνηση ενός ψηφοφόρου, καθώς ο νέος ψηφοφόρος παράγεται σύμφωνα με την υπάρχουσα κατανομή.

5.3.1 Πειραματική Διάταξη

Η διάταξη του πειράματος, αποτελείται και πάλι από ένα σύνολο 1000 ψηφοφόρων και 1000 υποψηφίων που έχουν παραχθεί σύμφωνα με κάποια κατανομή ως σημεία στο δισδιάστατο χώρο. Για κάθε ψηφοφόρο υπάρχει η σειρά κατάταξης για όλους τους υποψήφιους και τελικά εκλέγεται μία επιτροπή με 20 μέλη με κάθε κανόνα. Στη συνέχεια προσθέτουμε έναν νέο ψηφοφόρο p και εκτελούμε και πάλι το πείραμα, αυτή τη φορά με 1001 ψηφοφόρους. Τέλος, συγκρίνουμε τα αποτελέσματα.

Η προσθήκη του επιπλέον ψηφοφόρου γίνεται σύμφωνα με την υπάρχουσα κατανομή των ψηφοφόρων. Ωστόσο, η προσθήκη αυτή μπορεί να έχει μεγαλύτερη ή μικρότερη βαρύτητα, που οφείλεται στην τυχαιότητα της διαδικασίας παραγωγής του σημείου. Επομένως, χρήζει ιδιαίτερης σημασίας η εύρεση ενός τρόπου μέτρησης της ευστάθειας σε σύγκριση με τη βαρύτητα της προσθήκης του ψηφοφόρου.

Ανεξάρτητα από την τάση του αποτελέσματος να επηρεαστεί, δεν αναμένουμε μεγάλες αλλαγές στο αποτέλεσμα. Στους περισσότερους κανόνες, αναμένουμε να επηρεαστεί μόνο η τελευταία θέση, εξαιτίας της ύπαρξης ή δημιουργίας ισοπαλιών (για παράδειγμα αν υπάρχουν δύο υποψήφιοι που ισοβαθμούν για την τελευταία θέση της επιτροπής, αλλά μόνο μία ελεύθερη θέση, μετά την προσθήκη του νέου ψηφοφόρου μπορεί ο υποψήφιος που δεν άνηκε στην επιτροπή να κερδίσει την ψήφο που θα εξασφαλίσει τη λύση της ισοπαλίας και τελικά την εκλογή του).

Το παραπάνω οδηγεί σε ένα πολύ σημαντικό ερώτημα: «Υπάρχει πραγματικά αλλαγή στην επιτροπή ή πρόκειται για μία λανθασμένη θεώρηση;». Θα πρέπει να γίνει επιπλέον μελέτη για αυτό το θέμα, καθώς νέα ερωτήματα γεννώνται, π.χ. αν θεωρήσουμε πως όλοι οι υποψήφιοι που ισοβαθμούν στην τελευταία θέση εκλέγονται ανεξαρτήτως των ελεύθερων θέσεων, τότε πώς πρέπει να αντιμετωπίζουμε τις ισοβαθμίες στις προηγούμενες θέσεις κλπ. Λύση για όλα αυτά τα ερωτήματα αποτελεί η εύρεση όλων των ισοδύναμων αποτελεσμάτων και να συγκρίνουμε το νέο αποτέλεσμα με τα προηγούμενα. Τα πραγματικά αποτελέσματα προς σύγκριση είναι αυτά με τις μικρότερες διαφορές.

Ένα ακόμα ενδιαφέρον ερώτημα για μελέτη είναι αν όλοι οι κανόνες μπορούν να κριθούν με τα ίδια κριτήρια ή αν είναι άδικο, και ως αποτέλεσμα τα αποτελέσματα μας βεβιασμένα. Τέλος, ενδιαφέρον παρουσιάζει η μελέτη των κανόνων ως προς το αν μπορούμε να εξάγουμε πληροφορίες για τον νέο ψηφοφόρο, από την σύγκριση των δύο επιτροπών.

5.3.2 Αποτελέσματα

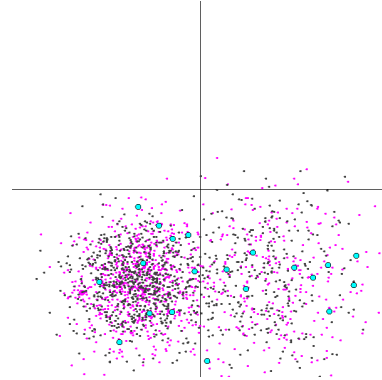
Στις εικόνες που ακολουθούν οι ψηφοφόροι συμβολίζονται με μοβ κουκκίδες, ο νέος ψηφοφόρος με μεγάλη μοβ κουκκίδα ενώ οι υποψήφιοι με μαύρες. Τα μέλη της επιτροπής από το αρχικό πείραμα συμβολίζονται με μεγάλες μαύρες κουκκίδες, ενώ από το δεύτερο πείραμα (δηλαδή μετά τη μετακίνηση του τυχαίου ψηφοφόρου) με γαλάζιες. Αν οι δύο επιτροπές ταυτίζονται όλες οι γα-

λάζιες κουκκίδες παρουσιάζονται με μαύρο περίγραμμα. Σε αντίθετη περίπτωση, για κάθε γαλάζια κουκκίδα χωρίς περίγραμμα υπάρχει μία αντίστοιχη μαύρη.

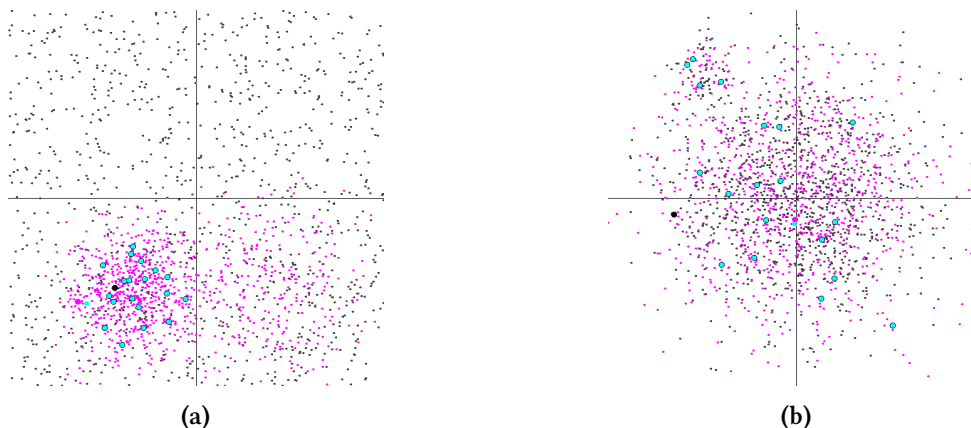
5.3.2.1 Κανόννας Single Nontransferable Vote

Όπως ήταν αναμενόμενο, ο κανόννας SNTV είναι ευσταθής υπό την προσθήκη ενός νέου ψηφοφόρου. Καθώς ο κανόννας λαμβάνει υπόψη μόνο την πρώτη προτίμηση των ψηφοφόρων, η προσθήκη ενός ψηφοφόρου μπορεί να επηρεάσει το αποτέλεσμα μόνο με την αύξηση των ψήφων ενός μόνο υποψηφίου κατά ένα. Έτσι, μπορεί να αλλάξει το πολύ μία θέση. Υπάρχουν δύο σημαντικά συμπεράσματα.

1. Δεν μπορούμε να βρούμε κατάλληλη μετρική για να μετρήσουμε την επίδραση στην επιτροπή. Η προσθήκη ενός ψηφοφόρου μπορεί να οδηγήσει σε μία μικρή ή μεγάλη αλλαγή, συγκρίνοντας τον νεοεκλεχθέντα υποψήφιο με αυτόν που αντικατέστησε.
2. Αν υπάρξει μία αλλαγή στην επιτροπή που εκλέχθηκε μετά την προσθήκη του νέου ψηφοφόρου, μπορούμε να εξάγουμε πληροφορίες για αυτόν, συγκρίνοντας τις δύο επιτροπές. Στις εικόνες 5.9 είναι φανερό, ότι ο νεοεκλεχθείς υποψήφιος (η μοναδική γαλάζια κουκκίδα χωρίς περίγραμμα) βρίσκεται πιο κοντά στον νέο ψηφοφόρο (μεγάλη μοβ κουκκίδα) από οποιοδήποτε άλλον υποψήφιο (μικρές μαύρες κουκκίδες). Επομένως, ακόμα και αν δεν διαχωρίζαμε τον επιπλέον ψηφοφόρο με κουκκίδα διαφορετικού μεγέθους, θα μπορούσαμε να μαντέψουμε την τοποθεσία του, καθώς πρέπει να βρίσκεται πολύ κοντά στον εκλεχθέντα υποψήφιο.



Εικόνα 5.8: Κανόννας SNTV - σταθερός



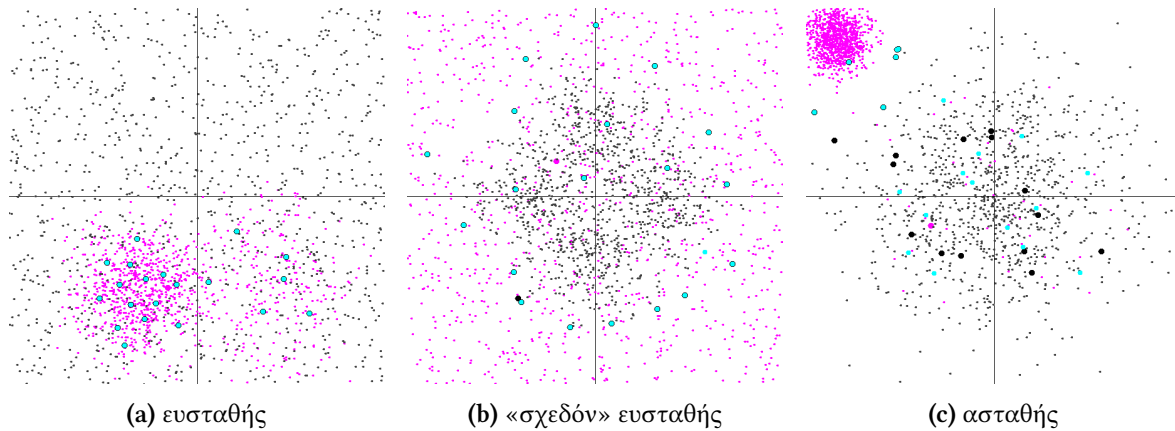
Εικόνα 5.9: Συμπεράσματα σχετικά με τον κανόνα SNTV

1. Δεν μπορούμε να μετρήσουμε την επίδραση της προσθήκης ενός ψηφοφόρου.
2. Μπορούμε να εξάγουμε πληροφορίες από την προσθήκη ενός ψηφοφόρου.

5.3.2.2 Κανόννας Single Transferable Vote

Ο κανόννας Single Transferable Vote είναι ασταθής κάτω από την προσθήκη ενός ψηφοφόρου. Όπως έχει ήδη αναφερθεί, ο κανόννας είναι ευαίσθητος σε οποιαδήποτε αλλαγή. Ακόμη και μία

διαφορά στη σειρά εκλογής ή απόρριψης των υποψηφίων μπορεί να οδηγήσει σε τελείως διαφορετικό αποτέλεσμα. Έτσι, ακόμα και αν σε πολλές περιπτώσεις τα αποτελέσματα ήταν ευσταθή, ο κανόνας χαρακτηρίζεται ως ασταθής λόγω της φύσης του.



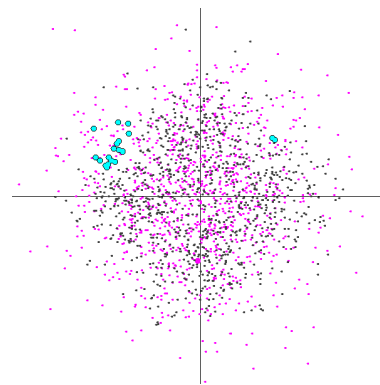
Εικόνα 5.10: Κανόνας STV - ασταθής

Δεν μπορούμε να εξάγουμε πληροφορίες σχετικά με τον πρόσθετο ψηφοφόρο

Εξαιτίας αυτής του της αστάθειας τα αποτελέσματα είναι απρόβλεπτα, μη συγκρίσιμα και ως εκ τούτου δεν μπορούμε να εξάγουμε πληροφορίες για τον επιπλέον ψηφοφόρο.

5.3.2.3 Κανόνας Bloc

Ο κανόνας Bloc είναι ευσταθής κατά την προσθήκη ενός ψηφοφόρου. Στην πλειονότητα των αποτελεσμάτων, μάλιστα, ήταν «τέλεια» ευσταθής. Όταν οι δύο επιτροπές δεν ταυτίζονταν πλήρως, διέφεραν μόνο ως προς έναν υποψήφιο. Μετά από εκτεταμένη ανάλυση, παρόλο που τα αποτελέσματα έδειχναν πως μόνο μία από τις θέσεις της επιτροπής μπορεί να αλλάξει, αυτό συμβαίνει μόνο αν στις $k - 1$ πρώτες θέσεις της επιτροπής οι υποψήφιοι έχουν περισσότερες ψήφους από τον τελευταίο υποψήφιο που εκλέχθηκε. Σε αντίθετη περίπτωση, δηλαδή αν l υποψήφιοι ισοβαθμούν στις τελευταίες l θέσεις της επιτροπής, η προσθήκη ενός ψηφοφόρου θα μπορούσε να αλλάξει και τις l θέσεις στην επιτροπή.



Εικόνα 5.11: Κανόνας Bloc - ευσταθής

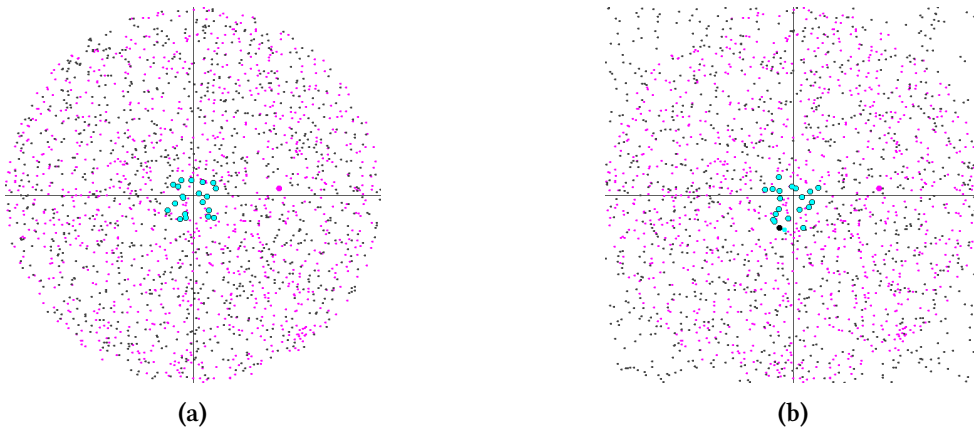
Όσο μεγαλύτερη είναι η επίδραση του ψηφοφόρου τόσες περισσότερες πληροφορίες μπορούμε να εξάγουμε για τον επιπλέον ψηφοφόρο. Από την εικόνα 5.12 μπορούμε να συμπεράνουμε ότι ο νέος υποψήφιος που εκλέχθηκε στην επιτροπή βρίσκεται κοντινότερα στον επιπλέον ψηφοφόρο, παρά σε αυτόν που αντικαταστάθηκε. Αν θεωρήσουμε ότι οι δύο επιτροπές διέφεραν σε παραπάνω από έναν υποψήφιο, η περιοχή στην οποία θα μπορούσε να βρίσκεται ο νέος ψηφοφόρος περιορίζεται περισσότερο και κατά συνέπεια έχουμε περισσότερες πληροφορίες για τις προτιμήσεις του.



Εικόνα 5.12: Συμπεράσματα για τον κανόνα Bloc
Μπορούμε να εξάγουμε πληροφορίες για τον επιπλέον ψηφοφόρο αν συγκρίνουμε τις δύο επιτροπές.

5.3.2.4 Κανόννας k -Borda

Ο κανόννας k -Borda είναι επίσης ευσταθής. Ξέρουμε ότι δυνητικά μπορούν να επηρεαστούν περισσότερες από μία θέσεις της επιτροπής. Ωστόσο τα πειραματικά αποτελέσματα δείχνουν ότι στην πράξη, η αλλαγή περισσότερων από μία θέσεων της επιτροπής, αποτελεί ακραία περίπτωση, καθώς απαιτεί ιδιαίτερα συγκεκριμένα αποτελέσματα πριν την προσθήκη του ψηφοφόρου, έτσι ώστε να μπορεί ο ψηφοφόρος να επηρεάσει περισσότερες θέσεις.

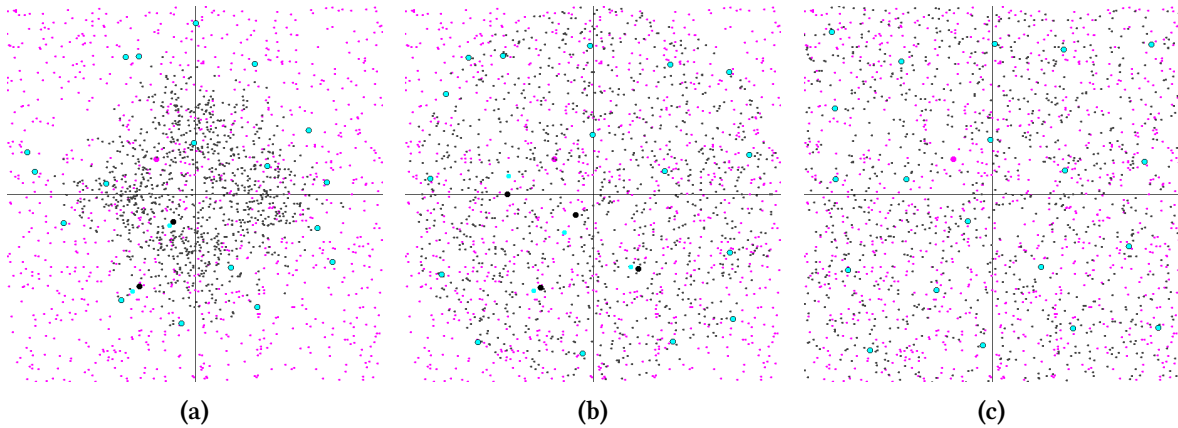


Εικόνα 5.13: Κανόννας k -Borda - ευσταθής

Σχετικά με τον επιπλέον ψηφοφόρο, μπορούμε να εξάγουμε γενικές πληροφορίες, καθώς το μόνο που γνωρίζουμε είναι ότι ο ψηφοφόρος βρίσκεται κοντινότερα (προτιμάει περισσότερο) το νέο μέλος της επιτροπής από το μέλος που αντικαταστάθηκε. Παρόλα αυτά δεν μπορούμε να κάνουμε υποθέσεις για την πραγματική θέση του ψηφοφόρου, αν δεν γνωρίζουμε πληροφορίες για τον συνολικό αριθμό των ψήφων του κάθε υποψήφιου πριν και μετά την προσθήκη του νέου ψηφοφόρου.

5.3.2.5 Κανόννας Chamberlin-Courant

Όπως έχει αναφερθεί, ο κανόννας Chamberlin-Courant είναι NP-hard και επομένως μελετάμε την ευστάθεια του ευρετικού μας αλγορίθμου.



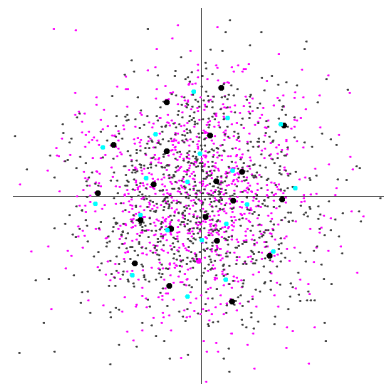
Εικόνα 5.14: Ευρετικός αλγόριθμος για τον κανόνα Chamberlin-Courant - ασταθής. Οι ψηφοφόροι παραμένουν οι ίδιοι, ενώ οι κατανομές των υποψηφίων αλλάζουν: (α) 4-Gaussian, (β) Uniform Disc, and (β) Uniform Square

Σημαντικό σε αυτό το σημείο, είναι να διευκρινίσουμε ότι τα κέντρα που προτείνει ο αλγόριθμος πάντοτε επηρεάζονται από την προσθήκη ενός ψηφοφόρου. Ωστόσο, αυτό μπορεί να μην είναι εμφανές, καθώς αυτό εξαρτάται από τις θέσεις των υποψηφίων (αναζητούμε τους κοντινότερους υποψήφιους στα κέντρα). Ενώ ο αλγόριθμος προτείνει τα ίδια κέντρα, το αποτέλεσμα μπορεί να εμφανίζεται ασταθές, «σχεδόν» ευσταθές ή ευσταθές (εικόνες 5.14 (a), (b), (c) αντίστοιχα). Καθώς λοιπόν τα αποτελέσματα δεν αλλάζουν σε μεγάλο βαθμό, δεν μπορούμε να ορίσουμε το αποτέλεσμα ως ευσταθές καθώς ένα ή περισσότερα σημεία σύγκλισης αλλάζουν. Καταλήγουμε λοιπόν στο χαρακτηρισμό του αλγορίθμου ως ασταθής. Για τον ίδιο λόγο, δεν μπορούμε να εξάγουμε και πληροφορίες για τον επιπλέον ψηφοφόρο, ακόμα κι αν ξέρουμε ότι κάτι έχει αλλάξει.

5.3.2.6 Monroe rule

Όπως έχει αναφερθεί, ο κανόνας Monroe είναι NP-hard και επομένως μελετάμε την ευστάθεια του ευρετικού μας αλγορίθμου. Αναμένουμε ότι ο αλγόριθμος είναι ασταθής υπό την προσθήκη ενός ψηφοφόρου. Τα αποτελέσματα το επιβεβαιώνουν.

Παρόλο που μελετάμε και κατόπιν χαρακτηρίζουμε τον αλγόριθμο υπό τις ίδιες συνθήκες με τους άλλους κανόνες, ο χαρακτηρισμός του φαίνεται άδικος εξαιτίας της φύσης του κανόνα: ενώ στο αρχικό πείραμα σε κάθε μέλος της επιτροπής αντιστοιχούν 50 ψηφοφόροι, μετά την προσθήκη του ψηφοφόρου, σε ένα μέλος της επιτροπής αντιστοιχούν 51 ψηφοφόροι. Αυτό εγγυημένα αλλάζει τα αποτελέσματα, καθώς οδηγεί σε διαφορετικά σημεία σύγκλισης. Επομένως, ο χαρακτηρισμός του αλγορίθμου, που προκύπτει από τη σύγκριση των δύο επιτροπών, είναι άδικος. Ωστόσο, καθώς ο αλγόριθμος είναι ασταθής ακόμα και στην περίπτωση της μετακίνησης ενός ψηφοφόρου, επιβεβαιώνεται ότι ο αλγόριθμος είναι ασταθής στις περισσότερες περιπτώσεις και επομένως δεν χρήζει επιπλέον μελέτης.



Εικόνα 5.15: Ευρετικός αλγόριθμος για τον κανόνα Monroe - ασταθής

5.4 Σύνοψη των Αποτελεσμάτων

Τα αποτελέσματα σχετικά με την ευστάθεια συνοψίζονται στον πίνακα 5.1.

	Ευστάθεια					
	Καλύτερη περίπτωση		Μέση περίπτωση		Χειρότερη περίπτωση	
	Μετακίνηση Ψηφοφόρου	Προσθήκη Ψηφοφόρου	Μετακίνηση Ψηφοφόρου	Προσθήκη Ψηφοφόρου	Μετακίνηση Ψηφοφόρου	Προσθήκη Ψηφοφόρου
SNTV	✓	✓	✓	✓	✓	✖
STV	✓	✓	✓	✗	✗	✗
Bloc	✓	✓	✓	✓	✖	✖
k-Borda	✓	✓	✓	✓	✖	✖
CC	✓	✓	✖	✗	✗	✗
Monroe	✓	✓	✗	✗	✗	✗

Πίνακας 5.1: Ευστάθεια των κανόνων εκλογής επιτροπών

Το ✓ δηλώνει ότι ο κανόνας είναι ευσταθής, ενώ το ✗ δηλώνει ότι ο κανόνας είναι ασταθής. Το ✖ δηλώνει ότι ο κανόνας είναι ευσταθής, αλλά εν δυνάμει μπορεί να υπάρξει μικρή αλλαγή κάτω από συγκεκριμένες προϋποθέσεις.

Σχετικά με την αποκάλυψη πληροφοριών για τον επιπλέον ψηφοφόρο, μπορούμε να συνοψίσουμε τα αποτελέσματα στον πίνακα 5.2. Σημειώνεται ότι απαραίτητη είναι κάποια αλλαγή στα αποτελέσματα για να εξάγουμε πληροφορίες.

	Κανόνες					
	SNTV	STV	Bloc	k-Borda	CC	Monroe
Αποκάλυψη Πληροφοριών	✓	✗	✓	✓	✗	✗

Πίνακας 5.2: Αποκάλυψη πληροφοριών για τον νέο ψηφοφόρο

Το ✓ δηλώνει ότι ο κανόνας αποκαλύπτει πληροφορίες για τον επιπρόσθετο ψηφοφόρο, ενώ το ✗ δηλώνει ότι ο κανόνας δεν αποκαλύπτει πληροφορίες για τον πρόσθετο ψηφοφόρο.

Κεφάλαιο 6

Σχεδιασμός & Υλοποίηση του Πειράματος σε Πολυδιάστατο Χώρο

Ζητούμενο ήταν και η μελέτη πρακτικών εφαρμογών των κανόνων. Θα μπορούσαμε να παράξουμε τις δικές μας εισόδους, ωστόσο προτιμήσαμε την μελέτη πραγματικών δεδομένων, προερχόμενα από την ηλεκτρονική βιβλιοθήκη “PrefLib: A Library for Preferences”[27]. Σε αυτή, έχουν ομαδοποιηθεί τα δεδομένα και μορφολογία τους όσο το δυνατόν περισσότερο¹

6.1 Πείραμα

6.1.1 Πειραματική Διαδικασία

Για κάθε προσομοίωση εκλογών δόθηκε ένα σύνολο ψηφοφόρων και υποψηφίων. Όλα τα δεδομένα που χρησιμοποιήθηκαν από την PrefLib προήλθαν από το πακέτο SOC (Strict Orders - Complete List), που σημαίνει ότι για κάθε ψηφοφόρο, δίνεται η πλήρη σειρά κατάταξης των υποψηφίων, χωρίς ισοπαλίες. Σκοπός μας είναι να εκλέξουμε μία επιτροπή (ένα υποσύνολο των υποψηφίων), που εκπροσωπεί τους υποψήφιους με συγκεκριμένο τρόπο. Καθώς όμως ο αριθμός των υποψηφίων και ψηφοφόρων διαφέρει σε κάθε είσοδο, διαφέρει και το μέγεθος της επιτροπής (3 με 5 άτομα). Για το λόγο αυτό απορρίφθηκαν πολλές εισοδοί, λόγω του μικρού αριθμού των υποψηφίων, και τελικά κρατήσαμε μόνο τις ακόλουθες.

1	ED-00006-00000003	14	ED-00006-00000034
2	ED-00006-00000004	15	ED-00006-00000035
3	ED-00006-00000007	16	ED-00006-00000036
4	ED-00006-00000008	17	ED-00006-00000037
5	ED-00006-00000011	18	ED-00006-00000044
6	ED-00006-00000012	19	ED-00006-00000046
7	ED-00006-00000018	20	ED-00006-00000048
8	ED-00006-00000021	21	ED-00009-00000001
9	ED-00006-00000022	22	ED-00009-00000002
10	ED-00006-00000028	23	ED-00012-00000001
11	ED-00006-00000029	24	ED-00014-00000001
12	ED-00006-00000032	25	ED-00032-00000002
13	ED-00006-00000033		

¹ Αυτή τη στιγμή υπάρχουν 10 τύποι δεδομένων σε 3 διαφορετικές μορφές. Περισσότερες λεπτομέρειες σχετικά με τις επιλογές μας ακολουθούν στις επόμενες ενότητες

Για να μελετήσουμε την ευστάθεια, αντιγράψαμε τις εισόδους και τις επεξεργαστήκαμε με δύο διαφορετικούς τρόπους. Την πρώτη φορά επεξεργαστήκαμε τις προτιμήσεις ενός τυχαίου ψηφοφόρου (χωρίς να αλλάζουμε την πρώτη του προτίμηση) και τη δεύτερη προσθέτοντας έναν ψηφοφόρο.

6.1.2 Αναπαράσταση των Ψηφοφόρων & Υποψηφίων

Στο προηγούμενο πείραμα, όπου και οι ψηφοφόροι και οι υποψήφιοι αναπαρίστανται ως σημεία στο διδιάστατο χώρο, παίρναμε τη σειρά κατάταξης και το βαθμό μη αντιπροσώπευσης των ψηφοφόρων από την απόσταση των ψηφοφόρων - υποψηφίων. Όσο μεγαλύτερη η απόσταση, τόσο μικρότερο κέρδος για τον ψηφοφόρο.

Ωστόσο στο πείραμα αυτό, δεν έχουμε συντεταγμένες. Αντίθετα έχουμε μόνο τη σειρά κατάταξης. Παρόλο που για ορισμένους κανόνες (SNTV, STV, k -Borda και Bloc) η σειρά κατάταξης αρκεί, δεν ισχύει για όλους. Στους κανόνες Chamberlin-Courant και Monroe, στόχος είναι η ελαχιστοποίηση της μη αντιπροσώπευσης των ψηφοφόρων και ως αποτέλεσμα, πρέπει να ερμηνεύσουμε τη σειρά κατάταξης στο βαθμό μη αντιπροσώπευσης. Για να γίνει αυτό, χρησιμοποιούμε μία συνάρτηση αντιστοίχισης, που μπορεί να είναι γραμμική, κυρτή, ή να έχει οποιαδήποτε άλλη κατάλληλη μορφή. Επιπλέον, οι κανόνες Chamberlin-Courant και Monroe ανήκουν υπολογιστικά στο NP. Έτσι, χρησιμοποιήσαμε τους ευρετικούς αλγόριθμους που παρουσιάσαμε στο κεφάλαιο 4. Ωστόσο, οι αλγόριθμοι μας χρησιμοποιούν τη θέση των ψηφοφόρων για να δημιουργήσουν τις ομάδες και μόνο στο τέλος τη θέση των υποψηφίων για να βρουν τους κοντινότερους στα κέντρα των ομάδων, και που τελικά θα αποτελέσουν την επιτροπή. Προκειμένου να χρησιμοποιήσουμε τους ίδιους αλγόριθμους, ερμηνεύσαμε τις σειρές κατάταξης ως θέσεις στον πολυδιάστατο χώρο. Σε κάθε πείραμα υπήρχαν n διαστάσεις, όπου n ο αριθμός των υποψηφίων και ο κάθε υποψήφιος C_i εκπροσωπείται από ένα διάνυσμα βάσης με την ακόλουθη μορφή:

$$C_0 = [1\ 0\ 0\ \dots\ 0], C_1 = [0\ 1\ 0\ \dots\ 0], \dots, C_{n-1} = [0\ 0\ 0\ \dots\ 1].$$

Έτσι, κάθε ψηφοφόρος V_j εκπροσωπείται από το ένα διάνυσμα:

$$V_j = [c_{0j}\ c_{1j}\ c_{2j}\ \dots\ c_{(n-1)j}],$$

όπου c_{ij} η θέση που ο ψηφοφόρος V_j κατέταξε τον υποψήφιο C_i και $1 \leq c_{ij} \leq n$.

Με τον τρόπο αυτό, μπορούσαμε να χρησιμοποιήσουμε τους ίδιους αλγόριθμους, με μια διαφορά. Αποφασίσαμε να χρησιμοποιήσουμε για τη μέτρηση των αποστάσεων νόρμα L^1 (Manhattan distance metric) αντί για νόρμα L^2 (Euclidean distance metric) για τον υπολογισμό των αποστάσεων, καθώς δουλεύει καλύτερα σε πολυδιάστατο χώρο [28].

Η συνάρτηση αντιστοίχισης της σειράς κατάταξης σε συντεταγμένες, που περιγράψαμε παραπάνω, είναι η πιο απλή γραμμική συνάρτηση που θα μπορούσαμε να χρησιμοποιήσουμε. Χρησιμοποιήσαμε και εκθετική συνάρτηση αντιστοίχισης όπου κάθε ψηφοφόρος V_j εκπροσωπείται από το διάνυσμα:

$$V_j = [2^{c_{0j}-1}\ 2^{c_{1j}-1}\ 2^{c_{2j}-1}\ \dots\ 2^{c_{(n-1)j}-1}]$$

Ωστόσο, μετά από πειράματα, απορρίψαμε την εκθετική συνάρτηση για τον ευρετικό αλγόριθμο του Chamberlin-Courant, διότι χάνει την ικανότητα της σύγκλισης. Περαιτέρω έρευνα πρέ-

πει να γίνει για την εύρεση της ιδανικής συνάρτησης αντιστοίχισης.

6.2 Αποτελέσματα

Οι συγκεκριμένες είσοδοι, εξαιτίας του μικρού αριθμού υποψηφίων ή/και ψηφοφόρων δεν μας επέτρεψαν να μελετήσουμε ιδιότητες όπως η αναλογική αντιπροσώπευση. Σχετικά με την ευστάθεια, δεν υπήρχε νόημα να μελετήσουμε τους κανόνες που χρησιμοποιούν μόνο τη σειρά κατάταξης, καθώς έχουν ήδη μελετηθεί στο κεφάλαιο 5, καθώς δεν έχει σημασία αν έχουμε δισδιάστατο ή πολυδιάστατο χώρο. Το μόνο που διέφερε στα πειράματα για τους κανόνες SNTV, STV, Bloc και Borda είναι ο βαθμός που κάθε ψήφος μπορεί να επηρεάσει το αποτέλεσμα, καθώς άλλαζε ο αριθμός των ψηφοφόρων και υποψηφίων.

Αντίθετα, οι ευρετικοί αλγόριθμοι των κανόνων Chamberlin-Courant και Monroe, βασίζονται σε αλγόριθμους ομαδοποίησης. Επομένως, ο αριθμός των ψηφοφόρων, υποψηφίων και ως κατά συνέπεια, των μελών της επιτροπής, καθιστά τις εισόδους ακατάλληλες για την έρευνα μας.

Μια ενδιαφέρουσα παρατήρηση, που προέκυψε από τα πειράματά μας, είναι πως στα πραγματικά δεδομένα όλοι οι κανόνες είχαν σχεδόν τα ίδια αποτελέσματα. Αυτό εξηγείται από το γεγονός ότι υπάρχει μια αντικειμενική αλήθεια, δηλαδή αντικειμενική σειρά κατάταξης (και ως κατά συνέπεια αντικειμενικοί νικητές), καθώς τα δεδομένα προέκυπταν κυρίως από αγώνες. Παρόλα αυτά, ο μικρός αριθμός των εισόδων που εξετάστηκαν, μας καθιστά σκεπτικούς για την συγκεκριμένη παρατήρηση και εκτεταμένη έρευνα θα πρέπει να γίνει προκειμένου να επιβεβαιωθεί η παρατήρηση.

Κεφάλαιο 7

Μελλοντικές Κατευθύνσεις

Προσεγγίσαμε το θέμα των κανόνων ψηφοφορίας για την εκλογή επιτροπών κυρίως στο διδιάστατο επίπεδο. Επικεντρωθήκαμε σε ευρετικούς αλγόριθμους για NP-hard κανόνες και στη μελέτη της ιδιότητας της ευστάθειας. Η δουλειά μας θα μπορούσε να επεκταθεί με ποικίλους τρόπους.

Μια δυνατή επέκταση είναι η μελέτη εφαρμογής διαφορετικών αλγορίθμων ομαδοποίησης, π.χ. hierarchical agglomerative clustering [29], προκειμένου να δημιουργήσουμε διαφορετικούς επιτυχημένους ευρετικούς και προσεγγιστικούς αλγορίθμους. Πρόκειται για έναν ενεργό τομέα έρευνας, ειδικά για τον κανόνα του Monroe, εξαιτίας των ιδιαίτερων περιορισμών του.

Δεύτερον, θα μπορούσε να επεκταθεί η μελέτη της ευστάθειας. Τα πειράματά μας ήταν περιορισμένα σε δύο βασικές περιπτώσεις, την περιορισμένη κίνηση ενός ψηφοφόρου και την προσθήκη ενός ψηφοφόρου. Θα μπορούσαν να μελετηθούν κι άλλες περιπτώσεις, όπως η αφαίρεση ενός ψηφοφόρου ή ακόμα και ενός υποψηφίου. Εξαιτίας των διαφορετικών αποστάσεων που μπορούν να δοθούν στην έννοια της ευστάθειας, π.χ. χειραγώγηση, έλεγχος [30], κ.λ.π. μας ενδιαφέρουν όλες οι δυνατές περιπτώσεις.

Μια ιδιαίτερα ενδιαφέρουσα επέκταση, οφείλεται στις προκλήσεις του τομέα. Πολλοί κανόνες, όπως οι k -Borda, Chamberlin-Courant, Monroe, κ.λ.π. απαιτούν τη γνώση της πλήρους σειράς κατάταξης των υποψηφίων από όλους τους ψηφοφόρους, πράγμα δύσκολο για μεγάλο αριθμό υποψηφίων. Επομένως, είναι δύσκολο να βρούμε κατάλληλα δεδομένα εκλογών για τη μελέτη τέτοιων κανόνων. Άλλωστε, αυτό ήταν και το πρόβλημα μας στο πείραμα στον πολυδιάστατο χώρο. Είναι λοιπόν φυσικό να μελετήσουμε τους κανόνες ψηφοφορίας για την εκλογή επιτροπών στην περίπτωση που υπάρχουν σειρές κατάταξης μη πλήρεις. Τότε, η πραγματική σειρά κατάταξης πρέπει να βρεθεί (παρόμοιο με το πρόβλημα της ύπαρξης ψήφων με θόρυβο [31]).

Τέλος, μια ιδιαίτερα ενδιαφέρουσα προσέγγιση είναι να βρούμε ένα τρόπο να παρουσιάσουμε τα αποτελέσματα και να κατατάξουμε τους ηττημένους υποψήφιους. στις περισσότερες περιπτώσεις είναι εύκολο να εντοπίσει κάποιος το νικητή των εκλογών, αλλά υπάρχουν φορές που απαιτούνται περισσότερες πληροφορίες, π.χ. όταν κάποιος υποψήφιος θέλει να ξέρει πόσο καλά τα πήγε παρόλο που έχασε. Αυτό μπορεί να είναι ιδιαίτερα δύσκολο για ορισμένους κανόνες, όπως για τον Single Transferable Vote rules, εξαιτίας της πολυπλοκότητας του. Εκτεταμένες μελέτες πρέπει να γίνουν, προκειμένου να βρεθεί ένας τρόπος για την αξιολόγηση των αποτελεσμάτων, των αδυναμιών των κανόνων και το βαθμό επιτυχίας των υποψηφίων. Στη μελέτη αυτή, οι αλγοριθμικές ιδιότητες των κανόνων (ποικιλομορφία, αναλογική αντιπροσωπευτικότητα, fairness) αποτελούν έναν χρήσιμο και πολύτιμο βοήθημα [5].

Bibliography / Βιβλιογραφία

- [1] James Zou, Reshef Meir, and David C Parkes. Approval voting behavior in doodle polls. 2014.
- [2] João Carlos Correia Baptista Soares Mello, Luiz Flávio Autran Monteiro Gomes, Eliane Gonçalves Gomes, Maria Helena Campos Soares de Mello, et al. Use of ordinal multi-criteria methods in the analysis of the formula 1 world championship. *Cadernos Ebape. BR*, 3(2):01–08, 2005.
- [3] Edith Elkind, Piotr Faliszewski, Piotr Skowron, and Arkadii Slinko. Properties of multiwinner voting rules. *Social Choice and Welfare*, 48(3):599–632, Mar 2017.
- [4] Salvador Barberà and Danilo Coelho. How to choose a non-controversial list with k names. *Social Choice and Welfare*, 31(1):79–96, 2008.
- [5] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Multiwinner voting: A new challenge for social choice theory. *Trends in Computational Social Choice*, 2017.
- [6] Michael Dummett. *Voting procedures*. Oxford University Press, 1984.
- [7] Ioannis Caragiannis, Swaprava Nath, Ariel D Procaccia, and Nisarg Shah. Subset selection via implicit utilitarian voting. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 151–157. AAAI Press, 2016.
- [8] Mostapha Diss and Ahmed Doghmi. Multi-winner scoring election methods: Condorcet consistency and paradoxes. *Public Choice*, 169(1-2):97–116, 2016.
- [9] Edith Elkind, Piotr Faliszewski, Jean-François Laslier, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. What do multiwinner voting rules do? an experiment over the two-dimensional euclidean domain. In Satinder P. Singh and Shaul Markovitch, editors, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 494–501. AAAI Press, 2017.
- [10] Edith Hemaspaandra, Lane A Hemaspaandra, and Jörg Rothe. Exact analysis of dodgson elections: Lewis carroll’s 1876 voting system is complete for parallel access to np. *Journal of the ACM (JACM)*, 44(6):806–825, 1997.
- [11] Jörg Rothe, Holger Spakowski, and Jörg Vogel. Exact complexity of the winner problem for young elections. *Theory of Computing Systems*, 36(4):375–386, 2003.
- [12] Edith Hemaspaandra, Holger Spakowski, and Jörg Vogel. The complexity of kemeny elections. *Theoretical Computer Science*, 349(3):382–391, 2005.

- [13] John R Chamberlin and Paul N Courant. Representative deliberations and representative decisions: Proportional representation and the borda rule. *American Political Science Review*, 77(3):718–733, 1983.
- [14] Burt L Monroe. Fully proportional representation. *American Political Science Review*, 89(4):925–940, 1995.
- [15] ID Hill, BA Wichmann, and DR Woodall. Algorithm 123. *The Computer Journal*, 30(3):276–281, 1987.
- [16] Chris Geller. Single transferable vote with borda elimination: A new vote counting system. Working paper 2201, Deakin University, Faculty of Business and Law, 2002.
- [17] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia. *Handbook of Computational Social Choice*. Cambridge University Press, New York, NY, USA, 1st edition, 2016.
- [18] Haris Aziz and Barton Lee. Achieving proportional representation via voting. *arXiv preprint arXiv:1708.07580*, 2017.
- [19] David Arthur and Sergei Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07*, pages 1027–1035, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.
- [20] Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.
- [21] Edith Elkind and Helger Lipmaa. *Small Coalitions Cannot Manipulate Voting*, pages 285–297. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [22] Reshef Meir, Ariel D. Procaccia, Jeffrey S. Rosenschein, and Aviv Zohar. Complexity of strategic behavior in multi-winner elections. *J. Artif. Int. Res.*, 33(1):149–178, September 2008.
- [23] Ariel D Procaccia, Jeffrey S Rosenschein, and Aviv Zohar. Multi-winner elections: Complexity of manipulation, control and winner-determination. In *IJCAI*, volume 7, pages 1476–1481, 2007.
- [24] John J Bartholdi and James B Orlin. Single transferable vote resists strategic voting. *Social Choice and Welfare*, 8(4):341–354, 1991.
- [25] Edith Hemaspaandra, Lane A. Hemaspaandra, and Jörg Rothe. Anyone but him: The complexity of precluding an alternative. *Artif. Intell.*, 171(5-6):255–285, April 2007.
- [26] Piotr Faliszewski, Edith Hemaspaandra, Lane A. Hemaspaandra, and Jörg Rothe. Llull and copeland voting computationally resist bribery and constructive control. *J. Artif. Intell. Res.*, 35:275–341, 2009.
- [27] Nicholas Mattei and Toby Walsh. Preflib: A library of preference data [HTTP://PREFLIB.ORG](http://preflib.org). In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT 2013)*, Lecture Notes in Artificial Intelligence. Springer, 2013.

- [28] Charu C. Aggarwal, Alexander Hinneburg, and Daniel A. Keim. On the surprising behavior of distance metrics in high dimensional space. In *Lecture Notes in Computer Science*, pages 420–434. Springer, 2001.
- [29] Daniel Müllner. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*, 2011.
- [30] Edith Hemaspaandra, Lane A. Hemaspaandra, and Jörg Rothe. Hybrid elections broaden complexity-theoretic resistance to control. *Math. Log. Q.*, 55(4):397–424, 2009.
- [31] Ioannis Caragiannis, Ariel D. Procaccia, and Nisarg Shah. When do noisy votes reveal the truth? In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 143–160, New York, NY, USA, 2013. ACM.