Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Εργαστήριο Λογικής και Επιστήμης Υπολογιστών
(Co.Re.Lab)

Αξιολόγηση Κανόνων Συνάθροισης Ιεραρχικών Μετρικών Προτιμήσεων ως προς τον Λόγο Παραμόρφωσης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΕΩΡΓΙΟΣ Α.ΧΙΟΝΑΣ

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

Αθήνα, Ιανουάριος 2022

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Εργαστήριο Λογικής και Επιστήμης Υπολογιστών
(Co.Re.Lab)

Αξιολόγηση Κανόνων Συνάθροισης Ιεραρχικών Μετρικών Προτιμήσεων ως προς τον Λόγο Παραμόρφωσης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΕΩΡΓΙΟΣ Α.ΧΙΟΝΑΣ

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

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 7η Ιανουαρίου 2022.

....	....	....
Δημήτριος Φωτάκης Αναπλ. Καθηγητής Ε.Μ.Π.	Αριστείδης Παγουρτζής Καθηγητής Ε.Μ.Π.	Ευάγγελος Μαρκάκης Αναπλ. Καθηγητής Ο.Π.Α.

Αθήνα, Ιανουάριος 2022

....

Γεώργιος Α.Χιονάς

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

Περίληψη

Η παρούσα διπλωματική εργασία καταπιάνεται με την ανάλυση της αποδοτικότητας διαφόρων κανόνων ψηφοφορίας όταν αυτοί λαμβάνουν περιορισμένη πληροφορία όσον αφορά τις προτιμήσεις των ψηφοφόρων. Σε μια διαδικασία εκλογής, οι ψηφοφόροι, ως επί το πλείστον, υποβάλλουν τις προτιμήσεις τους με ένα τακτικό τρόπο, είτε ψηφίζοντας τον προτιμότερο για αυτούς υποψήφιο είτε καταθέτοντας μια κατάταξη των υποψηφίων με βάση τις προτιμήσεις τους. Θεωρώντας πως οι προτιμήσεις των ψηφοφόρων μπορούν να ποσοτικοποιηθούν, οι προαναφερθέντες τρόποι συλλογής των προτιμήσεων συνιστούν μια σύμπτυξη της συνολικής πληροφορίας. Για το λόγο αυτό συχνά οι μηχανισμοί εκλογής δεν έχουν τη δυνατότητα να διακρίνουν και να εκλέξουν το βέλτιστο για την κοινωνία αποτέλεσμα. Η μετρική της παραμόρφωσης χρησιμοποιείται για να ποσοτικοποιήσει την επιλογή του κανόνα ψηφοφορίας σε σχέση με το βέλτιστο αποτέλεσμα, δοθέντος περιορισμένης πληροφορίας. Στην συγκεκριμένη εργασία θα αναλύσουμε την παραμόρφωση που επιφέρουν γνωστοί μηχανισμοί ψηφοφορίας. Κυρίως θα ασχοληθούμε με την περίπτωση όπου ψηφοφόροι και υποψήφιοι αποτελούν σημεία ενός μετρικού χώρου και εν προκειμένω οι προτιμήσεις των ψηφοφόρων αντιστοιχούν στην εγγύτητα τους από τους υποψηφίους. Στην περίπτωση όπου το ζητούμενο είναι η εκλογή μιας επιτροπής υποψηφίων, απαιτείται η επιλογή των υποψηφίων να συνιστά ένα σύνολο όσο το δυνατό αντιπροσωπευτικότερο προς την κοινωνία. Στο πλαίσιο του μετρικού χώρου, στην περίπτωση όπου ζητείται η εκλογή μιας επιτροπής υποψηφίων οι κατατάξεις των ψηφοφόρων δεν αποτελούν επαρκή πληροφορία ούτως ώστε η παραμόρφωση να είναι πεπερασμένη. Για το λόγο αυτό εξετάζουμε ένα νέο μοντέλο εξαγωγής των προτιμήσεων των ψηφοφόρων προκειμένου να λάβουμε καλύτερα αποτελέσματα σε ότι αφορά την συνολική παραμόρφωση.

Λέξεις κλειδιά

υπολογιστική θεωρία της κοινωνική επιλογής, κανόνες ψηφοφορίας, κοινωνική ευημερία, παραμόρφωση, αναλογική εκπροσώπηση, προσεγγιστικοί μηχανισμοί.

Abstract

The purpose of this diploma dissertation is to examine the efficiency of voting rules when they receive limited information regarding voters' preferences. Normally, an election is resolved based on each voter's linear ordering of candidates or in other words their ordinal preferences. We use the common assumption that voters hold a private cardinal valuation for each candidate. Thus the embedding of cardinal preferences into ordinal ones entails a degree of distortion. In this thesis, we will mostly examine the setting where agents represent points in a metric space and subsequently each voter prefers candidates that are closer to him to the ones that are further. We will study the multiwinner problem where it is required to elect a subset of candidates as winners such that each voter feels represented, or in the setting of the metric space the clusters of voters formed are as compact as possible. In this setting the ordinal preferences of voters are not enough to ensure bounded distortion. We examine a new rule of eliciting voters' preference and analyse the incurred distortion.

Key words

computational social choice , social choice rule, social welfare, distortion, proportional representation, approximation mechanisms.

Ευχαριστίες

Η εκπόνηση της διπλωματικής μου εργασίας σηματοδοτεί και επίσημα το τέλος των προπτυχιακών μου σπουδών στην σχολή των Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών. Σε αυτό το σημείο θα ήθελα να ευχαριστήσω όσους ανθρώπους ξεχώρισα αυτά τα χρόνια. Καταρχάς, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου, κ. Δημήτρη Φωτάκη για τον χρόνο που μου αφιέρωσε, για τη βοήθεια και τις συμβουλές που μου έδωσε. Επίσης, ευχαριστώ τον διδακτορικό φοιτητή Παναγιώτη Πατσιλινάκο για τη βοήθεια που μου προσέφερε κατά τη διάρκεια της εκπόνησης της διπλωματικής εργασίας. Θα ήθελα να ευχαριστήσω τους στενούς μου φίλους για τις όμορφες στιγμές που περάσαμε αυτά τα χρόνια. Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου, τον πατέρα μου, την μητέρα και την αδερφή μου για την στήριξη τους όλα τα χρόνια.

Γεώργιος Α.Χιονάς,

Αθήνα, 7η Ιανουαρίου 2022

Περιεχόμενα

Κεφάλαιο 1

Εκτεταμένη Ελληνική Περίληψη

Η διπλωματική εργασία έχει αποδοθεί στα αγγλικά για λόγους προσβασιμότητας. Στο συγκεκριμένο κεφάλαιο θα γίνει μια εκτεταμένη περίληψη της εργασίας στην ελληνική γλώσσα, ορίζοντας όλες τις βασικές έννοιες, ορισμούς και θεωρήματα, χωρίς όμως τις τεχνικές λεπτομέρειες των αντίστοιχων αποδείξεων. Η δομή του αυτού του κεφαλαίου είναι αντίστοιχη με αυτή του κύριου μέρους της διπλωματικής που αποδίδεται στην αγγλική γλώσσα.

Εισαγωγή

Το πλαίσιο της κοινωνικής επιλογής απαρτίζεται από παίκτες/άτομα και ενδεχόμενες εκβάσεις. Ο κύριος στόχος της θεωρίας της κοινωνικής επιλογής είναι η ανάλυση των ατομικών προτιμήσεων πάνω στις πιθανές εκβάσεις με σκοπό την εξαγωγή της έκβασης, η οποία ικανοποιεί τους περισσότερους παίκτες. Οι άνθρωποι στην καθημερινότητα καλούνται να καταθέσουν τις προτιμήσεις τους σε ερωτήματα και αντίστοιχα ένας μηχανισμός καλείται να βρει το βέλτιστο αποτέλεσμα. Τα παραδείγματα είναι πολλά. Μερικά από αυτά είναι η εκλογή ενός ή περισσότερων υποψηφίων με βάση τις προτιμήσεις των ψηφοφόρων ή επιλογή τοποθεσίας στην οποία θα χτιστεί μια δημόσια υπηρεσία ούτως ώστε να εξυπηρετούνται όλοι οι πολίτες. Τα τελευταία 15 χρόνια, τα θέματα της κοινωνικής επιλογής έχουν μελετηθεί από αρκετούς επιστήμονες της θεωρίας των υπολογιστών καθώς αυτά μπορούν να μοντελοποιηθούν ως προβλήματα βελτιστοποίησης. Συγκεκριμένα τα προβλήματα αυτά μελετώνται από τη σκοπιά του πως ο μηχανισμός λαμβάνει τις προτιμήσεις των παικτών και πως διαχειρίζεται αυτήν την πληροφορία ώστε να εξάγει το βέλτιστο αποτέλεσμα. Για να ορίσουμε το βέλτιστο αποτέλεσμα, χρειάζεται αφενός να ποσοτικοποιήσουμε τις προτιμήσεις των ατόμων πάνω στα πιθανά αποτελέσματα και αφετέρου να ορίσουμε την αντικειμενική συνάρτηση που επιζητούμε να μεγιστοποιήσουμε ή να ελαχιστοποιήσουμε. Σε ορισμένα προβλήματα αυτή η προσέγγιση είναι εύλογη, όπως στην περίπτωση που οι προτιμήσεις αντιστοιχούν σε αποστάσεις. Σε άλλα προβλήματα, η προσέγγιση αυτή δεν είναι είναι τόσο φυσική, καθώς είναι δύσκολο να ποσοτικοποιηθεί η προτίμηση ενός ατόμου σε σχέση με ένα υποψήφιο. Έχοντας όλες τις ποσοτικοποιημένες προτιμήσεις των παικτών, είναι εύκολο να βρεθεί η βέλτιστη έκβαση, ανάλογα με την αντικειμενική συνάρτηση βελτιστοποίησης. Μια ενδεχόμενη χαλάρωση στον τρόπο με τον οποίο οι παίκτες καλούνται να καταθέσουν τις προτιμήσεις τους είναι η κατάταξη των πιθανών εκβάσεων με βάση τις προτιμήσεις τους. Η συγκεκριμένη προσέγγιση αποτελεί σίγουρα έναν πιο φυσικό τρόπο για τον άνθρωπο καθώς είναι ευκολότερο να δηλώσει ότι προτιμάει τον υποψήφιο A από τον υποψήφιο B παρά να ποσοτικοποιήσει τις αντίστοιχες προτιμήσεις. Το ζήτημα είναι ότι η συμπύκνωση της πληροφορίας με την μορφή των κατατάξεων και εν γένει με οποιαδήποτε άλλη μορφή, καθιστά συχνά οποιονδήποτε μηχανισμό αδύνατο να εξάγει το βέλτιστο αποτέλεσμα. Για να μετρήσουμε την επίδραση της απώλειας της πληροφορίας στην επιλογή του βέλτιστου αποτελέσματος, χρησιμοποιείται η μετρική της παραμόρφωσης. Η συγκεκριμένη μετρική ανήκει στην οικογένεια του λόγου προσέγγισης χειρότερης περίπτωσης. Η έννοια αυτή χρησιμοποιείται σε διάφορους τομείς της επιστήμης των υπολογιστών και ορίζεται αναλόγως. Μερικά παραδείγματα είναι το approximation ratio στους προσεγγιστικούς αλγόριθμους και το price of anarchy στη θεωρία των παιγνίων. Η περισσότερη βιβλιογραφία πάνω στην μετρική της παραμόρφωσης έχει αφιερωθεί στο πλαίσιο όπου απαιτείται η επιλογή ενός νικητή, έχοντας ως είσοδο τις προτιμήσεις των ψηφοφόρων σε μορφή κατατάξεων. Οι πραγματικές-ποσοτικοποιημένες προτιμήσεις των ψηφοφόρων δεν είναι γνωστές στον εκάστοτε μηχανισμό. Αρχικά το πρόβλημα μελετήθηκε για την περίπτωση όπου οι πραγματικές προτιμήσεις αντιστοιχούν στην εκτίμηση που έχει ένας ψηφοφόρος για ένα υποψήφιο. Οι εκτιμήσεις είναι αυθαίρετες τιμές με το μόνο περιορισμό να είναι ότι για κάθε ψηφοφόρο το άθροισμα όλων των εκτιμήσεων πρέπει να είναι σταθερό. Η αντικειμενική συνάρτηση που θα εξετάσουμε είναι η επιλογή του υποψηφίου που μεγιστοποιεί το άθροισμα των εκτιμήσεων των ψηφοφόρων, η λεγόμενη κοινωνική ευημερία. Σε αυτό το πλαίσιο έχουν δοθεί κάτω όρια για τη παραμόρφωση που μπορεί να πετύχει ένας μηχανισμός και επίσης έχουν δοθεί μηχανισμοί που το επιτυγχάνουν. Θα μελετήσουμε επίσης το πρόβλημα όπου απαιτείται η επιλογή k υποψηφίων, όπου πάλι οι ψηφοφόροι διατηρούν μια εκτίμηση για τον κάθε υποψήφιο. Θα δείξουμε ότι το πρόβλημα των k υποψηφίων αποτελεί γενίκευση του προβλήματος του ενός υποψηφίου, το οποίο δεν είναι προφανές εξ αρχής. Το πρόβλημα επίσης έχει μελετηθεί εκτενώς, στο πλαίσιο όπου οι ψηφοφόροι και οι υποψήφιου ανήκουν σε ένα μετρικό χώρο. Αντίστοιχα έχουν δοθεί κάτω όρια για την παραμόρφωση. Στο πλαίσιο αυτό έχουν διατυπωθεί πολλοί μηχανισμοί που προσεγγίζουν το κάτω όριο και μηχανισμός που το πετυχαίνει. Στο συγκεκριμένο πλαίσιο, το πρόβλημα της επιλογής k υποψηφίων δεν αποτελεί γενίκευση της επιλογής ενός υποψηφίου και εν γένει ένας μηχανισμός χρειάζεται περισσότερη πληροφορία από τους ψηφοφόρους ώστε να είναι δυνατό να επιτευχθεί πεπερασμένη παραμόρφωση.

Συνεισφορά: Σε αυτήν την πτυχιακή εργασία εξετάζουμε το πρόβλημα της επιλογής k υποψηφίων, όταν οι ψηφοφόροι και οι υποψήφιοι ανήκουν σε ένα μετρικό χώρο. Αρχικά δείχνουμε ότι δεν είναι δυνατό για κανένα κανόνα ψηφοφορίας να πετύχει πεπερασμένη παραμόρφωση, αν έχουμε ως είσοδο μόνο τις κατατάξεις των ψηφοφόρων πάνω στους υποψήφιους. Προτείνουμε ένα πιο αποτελεσματικό τρόπο με τον οποίο εξάγουμε τις προτιμήσεις των ψηφοφόρων. Απαιτούμε από τον κάθε ψηφοφόρο να μας δώσει προσεγγιστικά την απόσταση του από τους t προτιμότερους υποψηφίους με βάση εκείνον. Με βάση αυτό το μοντέλο και με παράμετρο των αριθμό t δείχνουμε τα όρια της παραμόρφωσης

Εκτιμήσεις

Στην ενότητα αυτή θα εξετάσουμε το πλαίσιο στο οποίο οι πραγματικές προτιμήσεις των ψηφοφόρων αντιστοιχούν στην εκτίμηση που τρέφουν για κάθε υποψήφιο. Θεωρούμε ένα σύνολο V από n ψηφοφόρους και ένα σύνολο C από m υποψήφιους. Κάθε ψηφοφόρος i διατηρεί την πραγματική του εκτίμηση $v_i : C \to R_{\geq 0}$. Οι πραγματικές εκτιμήσεις όλων των ψηφοφόρων ορίζουν ένα διάνυσμα $\boldsymbol{v} = \{v_1, v_2, \ldots, v_n\}$. Η κοινωνική ευημερία που επιφέρει η επιλογή ενός υποψηφίου c με βάση το διάνυσμα $\boldsymbol{v}$ των πραγματικών εκτιμήσεων ορίζεται ως

$$SW(\boldsymbol{v}, c) = \sum_{i \in V} v_i(c)$$

Επίσης θεωρούμε ότι ο μηχανισμός έχει πρόσβαση στις προτιμήσεις των ψηφοφόρων μέσω του διανύσματος $\boldsymbol{I}$. Για ένα ψηφοφόρο i το I_i περιέχει λιγότερη πληροφορία σε σχέση το v_i που περιέχει ολόκληρη την πληροφορία. Η πιο συνήθης περίπτωση είναι όταν το I_i αντιστοιχεί στην κατάταξη των υποψηφίων με βάση τις προτιμήσεις του ψηφοφόρου i. Με βάση τα παραπάνω έχουμε τον εξής ορισμό.

Ορισμός 1.1. *Η παραμόρφωση(distortion) για ένα μηχανισμό f που εξάγει την συμπυκνωμένη* *πληροφορία I_i από κάθε ψηφοφόρο i και επιλέγει ως νικητή τον υποψήφιο $f(V, \boldsymbol{I})$ είναι:*

$$dist((f, \boldsymbol{I})) = sup_{\boldsymbol{v} \rhd \boldsymbol{I}} \frac{max_{j \in C} SW(j|\boldsymbol{v})}{SW(f(V, \boldsymbol{I})|\boldsymbol{v})}$$

Στην περισσότερη βιβλιογραφία, σε αυτό το πλαίσιο ακολουθείται η σύμβαση ότι το άθροισμα των εκτιμήσεων ενός ψηφοφόρου για όλους τους υποψήφιους όλους είναι σταθερό. Θα ακολουθήσουμε κι εμείς αυτή τη σύμβαση. Με βάση αυτή, οι εκτιμήσεις μπορούν να κοινωνικοποιηθούν έτσι ώστε να αθροίζουν στο 1.

Θεώρημα 1.1. *Για το πρόβλημα της επιλογής ενός υποψηφίου ως νικητή, στο πλαίσιο των κανονικοποιημένων εκτιμήσεων κάθε μηχανισμός έχει παραμόρφωση $\Omega(m^2)$*

Θεώρημα 1.2. *Για το πρόβλημα της επιλογής ενός υποψηφίου ως νικητή, στο πλαίσιο των κανονικοποιημένων εκτιμήσεων ο κάνονας Plurality πετυχαίνει παραμόρφωση $\mathcal{O}(m^2)$.*

Ο κανόνας Plurality επιλέγει τον υποψήφιο ο οποίος έχει αναδειχθεί πρώτος στις περισσότερες κατατάξεις των ψηφοφόρων. Ουσιαστικά, μας αρκεί μόνο η πρώτη ψήφος κάθε ψηφοφόρου για την εφαρμογή του κανόνα Plurality, ο οποίος πετυχαίνει τη καλύτερη δυνατή παραμόρφωση.

Επιλογή k υποψηφίων ως νικητές

Το πρόβλημα της επιλογής k υποψηφίων ως νικητές συναντάται σε αντίστοιχα προβλήματα στην πραγματικότητα. Το πιο χαρακτηριστικό είναι οι εκλογές, όπου απαιτείται η δημιουργία μιας επιτροπής k υποψηφίων, οι οποίοι θα αντιπροσωπεύουν το σύνολο σε όσο το δυνατόν μεγαλύτερο βαθμό. Για το λόγο αυτό ορίζουμε την εκτίμηση ενός ψηφοφόρου i για ένα σύνολο X υποψηφίων μεγέθους k ως εξής.

$$v_i(X) = \max_{c \in X} v_i(c)$$

Δηλαδή η εκτίμηση του ψηφοφόρου για ένα υποσύνολο υποψηφίων προκύπτει αποκλειστικά από τον υποψήφιο του υποσυνόλου που εκτιμά περισσότερο. Σε ότι αφορά την παραμόρφωση θεωρούμε πάλι ότι η είσοδος που δίνεται στο μηχανισμό είναι οι κατατάξεις των ψηφοφόρων πάνω στους υποψήφιους. Προκύπτουν τα ακόλουθα αποτελέσματά.

Θεώρημα 1.3. *Για το πρόβλημα της επιλογής k υποψηφίων ως νικητές, ισχύουν τα παρακάτω κάτω όρια για την παραμόρφωση που μπορεί να πετύχει κάθε μηχανισμός f*

- *Για $k \leq m/6$ $dist(f) \geq 1 + m \cdot \frac{m-3k}{6k}$*
- *Για $k \leq m/2$ $dist(f) \geq 1 + m$*
- *Για $k \geq m/2$ $dist(f) \geq 1 + m \cdot \frac{m-k}{k}$*

Θεώρημα 1.4. *Για το πρόβλημα της επιλογής k υποψηφίων ως νικητές ο κανόνας k-Plurality πετυχαίνει παραμόρφωση $\mathcal{O}(\frac{m^2}{k})$*

Ο κανόνας k-Plurality επιλέγει τους k υποψήφιους που έχουν αναδειχθεί πρώτοι στις κατατάξεις των ψηφοφόρων τις περισσότερες φορές.

Παρατηρώντας τα κάτω όρια της παραμόρφωσης, γίνεται αντιληπτό ότι όσο ο αριθμός k των νικητών αυξάνεται, τα όρια "χαλαρώνουν", γεγονός που υποδεικνύει ότι με την ίδια πληροφορία μπορούμε να επιτύχουμε καλύτερα αποτελέσματα. Με άλλα λογία, το πρόβλημα γίνεται ευκολότερο όσο αυξάνεται ο αριθμός των νικητών. Τέλος, όπως αναφέρθηκε, στην περίπτωση του ενός υποψηφίου ως νικητή αλλά και στην περίπτωση του συνόλου υποψηφίων ως νικητές, μπορούμε να επιτύχουμε βέλτιστη παραμόρφωση αξιοποιώντας μόνο την πρώτη επιλογή κάθε ψηφοφόρου. Υπό αυτή την έννοια η περαιτέρω πληροφορία που παρέχουν οι κατατάξεις είναι περιττή. Ουσιαστικά, μπορούμε να συμπεράνουμε ότι σε αυτό το πλαίσιο, οι κατατάξεις των ψηφοφόρων δεν αποτελούν τον πιο αποδοτικό τρόπο, σε ότι αφορά το μέγεθος της πληροφορίας σε σχέση με την παραμόρφωση που μπορεί να επιτευχθεί. Ένα επόμενο βήμα, το οποίο έχει πραγματοποιηθεί είναι η συστηματικότερη μελέτη του ισοζυγίου μεταξύ του μεγέθους της πληροφορίας που δίνει ο κάθε ψηφοφόρος και των ορίων της παραμόρφωσης που μπορεί να επιτύχει κάθε μηχανισμός δεδομένης αυτής της πληροφορίας.

Επιλογή ενός υποψηφίου ως νικητή σε μετρικό χώρο

Σε αυτό το πλαίσιο θεωρούμε ότι οι ψηφοφόροι και υποψήφιοι αναπαριστούν σημεία σε ένα μετρικό χώρο. Σε αυτήν την περίπτωση οι προτιμήσεις των ψηφοφόρων αντιστοιχούν στις αποστάσεις τους από τους υποψηφίους, δηλαδή ένας ψηφοφόρος προτιμά κάθε υποψήφιο που απέχει μικρή απόσταση σε σχέση με αυτούς που βρίσκονται πιο μακριά. Επομένως αναζητούμε τον υποψήφιο-σημείο στο χώρο για τον οποίο το άθροισμα όλων των αποστάσεων μεταξύ αυτού και των ψηφοφόρων είναι ελάχιστο. Από εδώ και στο εξής, θα χρησιμοποιούμε τη συνάρτηση d, όπου είναι μια αρνητική συνάρτηση που δίνει την απόσταση μεταξύ δύο σημείων στο χώρο.΄ Η ένωση του συνόλου V των ψηφοφόρων και του συνόλου C των υποψηφίων μαζί με την συνάρτηση d ορίζουν το μετρικό χώρο. Για 3 σημεία x, y, z που ανήκουν σε ένα μετρικό χώρο ισχύει η τριγωνική ανισότητα, δηλαδή $d(x, y) \leq d(x, z) + d(z, y)$.
Η αντικειμενική συνάρτηση που επιζητούμε να ελαχιστοποιήσουμε σε αυτήν την περίπτωση είναι το συνολικό κοινωνικό κόστος, το οποίο για ένα υποψήφιο x ορίζεται ως εξής.

$$SC(x, V|d) = \sum_{i \in V} d(i, x)$$

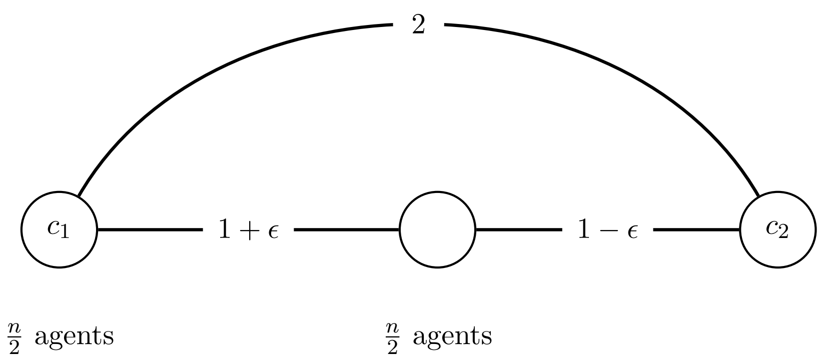
Ορισμός 1.2. *Για το πρόβλημα της επιλογής ενός υποψηφίου ως νικητή σε μετρικό χώρο, η παραμόρφωση(distortion) για ένα μηχανισμό f που έχει ως είσοδο τη συμπυκνωμένη πληροφορία I για τις προτιμήσεις των ψηφοφόρων και επιλέγει ως νικητή τον υποψήφιο $w = f(V, \boldsymbol{I})$ ορίζεται ως:*

$$dist(f, I) = \frac{SC(f(V, \boldsymbol{I}), V|d)}{\max_{x \in C} SC(x, V|d)}$$

Αξίζει να σημειωθεί πως ο περιορισμός της τριγωνικής ανισότητας μειώνει το βαθμό αυθαιρεσίας των προτιμήσεων των ψηφοφόρων, σε σχέση με το προηγούμενο πλαίσιο που εξετάστηκε που και για το λόγο αυτό αναμένουμε τα κάτω όρια της παραμόρφωσης να είναι χαμηλότερα.

Σε αυτήν την ενότητα θεωρούμε ότι η συμπυκνωμένη πληροφορία I που λαμβάνει κάθε μηχανισμός είναι οι κατατάξεις σ των ψηφοφόρων. Θεωρούμε σε κάθε περίπτωση ότι οι κατατάξεις είναι συνεπείς με τις πραγματικές εκτιμήσεις τους δηλαδή αν για ένα ψηφοφόρο i και δυο υποψήφιους w, x ισχύει ότι $d(i, w) \leq d(i, x)$ αυτό αποτυπώνεται και στην κατάταξη του δηλαδή $w \succeq_i x$.

Θεώρημα 1.5. *Η παραμόρφωση ενός μηχανισμού f που λαμβάνει ως είσοδο τις κατατάξεις σ των ψηφοφόρων είναι μεγαλύτερη ή ίση του 3.*



Σχήμα 1.1: Lower bound

Θεωρούμε ότι έχουμε 2 υποψήφιους c_1, c_2, όπου οι μισοί ψηφοφόροι προτιμούν τον c_1 ενώ οι υπόλοιποι τον c_2. Κανένας μηχανισμός δεν μπορεί να ξεχωρίσει ποιο είναι ο βέλτιστος, όποτε χωρίς βλάβη της γενικότητας θεωρούμε ότι επιλέγεται ο υποψήφιος c_2 ως νικητής. Κατασκευάζουμε ένα παράδειγμα έτσι ώστε οι μισοί ψηφοφόροι να βρίσκονται στο ίδιο σημείο με τον υποψήφιο c_1 και οι υπόλοιποι να βρίσκονται στο ίδιο σημείο και να απέχουν απόσταση $1+\epsilon$ από τον c_1 και απόσταση $1-\epsilon$ από τον c_2, όπως φαίνεται και στο Σχήμα 1.1. Με βάση τις πραγματικές αποστάσεις ο υποψήφιος που ελαχιστοποιεί το κοινωνικό κόστος είναι ο c_1. Τα αντίστοιχα κόστη για τους 2 υποψηφίους είναι $SC(c_2, d) = n/2 \cdot 2 + n/2 \cdot (1-\epsilon)$ και $SC(c_1, d) = n/2 \cdot (1+\epsilon)$. Οπότε η παραμόρφωση τείνει στο 3 καθώς το ϵ τείνει στο 0.

Στη συνέχεια θα παρουσιάσουμε τα άνω όρια της παραμόρφωσης που πετυχαίνουν μερικοί μηχανισμοί. Για την απόδειξη των άνω ορίων της παραμόρφωσης το επόμενο λήμμα είναι χρήσιμο.

Λήμμα 1.1. *Για δύο υποψήφιους w, x ορίζουμε το σύνολο wx ως το σύνολο των ψηφοφόρων που προτιμούν τον w σε σχέση με τον x. Λέμε ότι ο υποψήφιος w νικάει τον x αν $|wx| \geq \frac{n}{2}$.*

Ορισμός 1.3. *Ένας υποψήφιος w ονομάζεται Condorcet winner αν για κάθε άλλο υποψήφιο x ισχύει ότι $|wx| \geq \frac{m}{2}$.*

Θεώρημα 1.6. *Για δύο υποψήφιους w, x ισχύει $\frac{SC(w,d)}{SC(x,d)} \leq \frac{2n}{|wx|} - 1$.*

Η παραπάνω ανισότητα είναι σημαντική για να φράξουμε τη παραμόρφωση ενός μηχανισμού που επιλέγει τον υποψήφιο w, αν μπορούμε να βρούμε ένα κάτω όριο για την ποσότητα $|wx|$, για κάθε άλλο υποψήφιο x

Με βάση το Θεώρημα 1.6, αν υπάρχει Condorcet winner τότε η επιλογή του συγκεκριμένου υποψηφίου ως νικητή έχει distortion το πολύ 3. Πάρα ταύτα δεν υπάρχει πάντα Condorcet winner.

Πολλοί από τους γνωστούς κανόνες για την επιλογή ενός υποψηφίου ως νικητή ανήκουν στην οικογένεια των positional scoring rules. Σε αυτούς τους κανόνες η κατάταξη σ_i κάθε ψηφοφόρου i αντιστοιχίζεται σε ένα διάνυσμα $\boldsymbol{s}^i = \{s_1, s_2, .., s_m\}$ με $s_1 \geq s_2 \geq \cdots \geq s_m$, δηλαδή κάθε υποψήφιος λαμβάνει ένα βαθμό από κάθε ψηφοφόρο με βάση την κατάταξη του και ο νικητής είναι αυτός που συγκεντρώνει αθροιστικά τη μεγαλύτερη βαθμολογία. Τέτοιοι μηχανισμοί, για παράδειγμα είναι οι κανόνες *Plurality* και *Borda* και τα αντίστοιχα διανύσματα είναι:

- Plurality: $s = \{1, 0, \ldots, 0\}$
- Borda: $s = \{m-1, m-2, \ldots, 0\}$

Θεώρημα 1.7. *Οι κανόνες Plurality και Borda έχουν distortion το πολύ $2m - 1$.*

Για να πετύχουμε σταθερό distortion χρειάζεται να εξετάσουμε τις σχέσεις των υποψηφίων αν ζεύγη με βάση τις κατατάξεις των ψηφοφόρων. Για το λόγο αυτό εισάγουμε την έννοια των tournament graphs, όπου σε αυτούς τους γράφους κάθε κόμβος αντιστοιχεί σε έναν υποψήφιο. Ορίζουμε δύο τέτοια είδη γράφων.

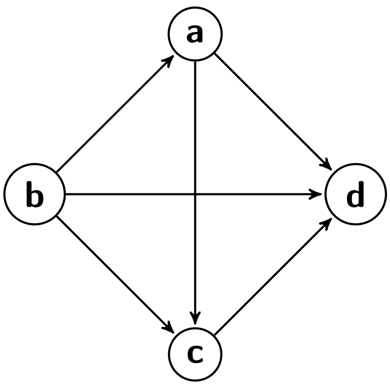
Ορισμός 1.4. *Το tournament graph είναι ένας κατευθυνόμενος, όπου κάθε κόμβος αντιστοιχεί σε ένα υποψήφιο και υπάρχει ακμή από τον κόμβο a στον κόμβο b αν ισχύει ότι* $|ab| \geq \frac{n}{2}$

Ορισμός 1.5. *Το weighted tournament graph είναι ένας πλήρες κατευθυνόμενος γράφος με βάρη, στον οποίο η ακμή από τον κόμβο a στον κόμβο b έχει βάρος* $|ab|$ *και αντίστοιχα η ακμή από τον κόμβο b στον κόμβο a έχει βάρος* $|ba|$.

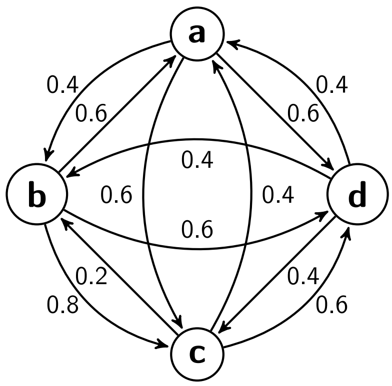
Στο παρακάτω παράδειγμα 5 ψηφοφόροι καταθέτουν τις κατατάξεις για τους 4 υποψηφίους a, b, c, d και με βάση αυτές προκύπτουν οι γράφοι ου ορίστηκαν παραπάνω.

- σ_1: $a \succ b \succ c \succ d$
- σ_2: $a \succ c \succ b \succ d$
- σ_3: $d \succ b \succ c \succ a$
- σ_4: $b \succ a \succ c \succ d$
- σ_5: $d \succ b \succ c \succ a$

Οι γράφοι που προκύπτουν φαίνονται στο Σχήμα 1.2



(a) Tournament graph



(b) Weighted tournament graph

Σχήμα 1.2: Tournament graphs

Ορισμός 1.6 (Uncovered set)**.** *Ένας υποψήφιος w ανήκει στο uncovered set αν για κάθε άλλο υποψήφιο x ισχύει ότι* $|wx| \geq \frac{n}{2}$ *ή υπάρχει ένας άλλος υποψήφιος y τέτοιος ώστε* $|wy| \geq \frac{n}{2}$ *και* $|yx| \geq \frac{n}{2}$.

Μπορούμε να παρατηρήσουμε πως ο ορισμός του uncovered set απότελεί μια χαλάρωση σε σχέση με τον ορισμό των Condorcet winners. Το σημαντικό σημείο είναι ότι υπάρχει πάντα ένας υποψήφιος που ανήκει σε αυτό το σύνολο. Επίσης αξίζει να σημειωθεί η συσχέτιση του tournament graph με τους υποψήφιους που ανήκουν στο uncovered set. Κάθε υποψήφιος που ανήκει στο uncovered set, για τον αντίστοιχο κόμβο του γραφήματος ισχύει ότι μπορεί να μεταβεί σε οποιονδήποτε άλλο κόμβο σε το πολύ δυο βήματα.

Θεώρημα 1.8. *Η παραμόρφωση του μηχανισμού ο οποίος επιλογή έναν υποψήφιο που ανήκει στο uncovered set είναι το πολύ 5.*

Οι ορισμοί των tournament graph και uncovered set, σχετίζονται με τις προτιμήσεις της πλειοψηφίας των ψηφοφόρων για κάθε ζεύγος υποψηφίων. Το επόμενο βήμα για να αναζητήσουμε μηχανισμούς με πιο μικρή παραμόρφωση είναι να εξετάσουμε για κάθε ζεύγος ψηφοφόρων τις προτιμήσεις τους για κάθε ζεύγος υποψηφίων.

Ορισμός 1.7 (Matching Uncovered Set). *Το Matching Uncovered Set είναι ένα σύνολο υποψηφίων w, τέτοιοι ώστε για κάθε άλλο υποψήφιο $x \in C \setminus \{w\}$ υπάρχει ένα τέλειο ταίριασμα στο διμερές γράφημα $G(w,x) = (V, V, E_{w,x})$, στο οποίο υπάρχει ακμή $(i,j) \in E_{w,x}$ αν υπάρχει ένας υποψήφιος y τέτοιος ώστε $w \succeq_i y$ και $y \succeq_j x$.*

Διευκρινίζουμε ότι ισχύει η σχέση $w \succeq_i w$ για κάθε υποψήφιο w και ψηφοφόρο i.

Θεώρημα 1.9. *Η επιλογή ενός υποψήφιου που ανήκει στο matching uncovered set επιφέρει distortion το πολύ 3.*

Μπορούμε να ορίσουμε ένα ορίσουμε ένα διαφορετικό διμερές γράφημα, στο οποίο η αναζήτηση ενός τέλειου ταιριάσματος είναι ισοδύναμη με το παραπάνω.

Ορισμός 1.8 (Integral domination graph). *Το Integral domination graph ενός υποψήφιου w είναι το διμερές γράφημα $G_w = (V, V, E_w)$, στο οποίο υπάρχει η ακμή $(i,j) \in E_w$ αν για τον ψηφοφόρο i ισχύει ότι $w \succeq_i top(j)$, όπου $top(j)$ είναι προτιμότερος υποψήφιος του j.*

Θεώρημα 1.10. *Η επιλογή ενός υποψήφιου του οποίου το Integral domination graph δέχεται τέλειο ταίριασμα, επιφέρει distortion το πολύ* 3.

Επιλογή k υποψηφίων ως νικητές σε μετρικό χώρο

Σε αυτήν την ενότητα θα μελετήσουμε το πρόβλημα της επιλογής k υποψηφίων ως νικητές, όταν οι ψηφοφόροι και υποψήφιοι ανήκουν σε ένα μετρικό χώρο. Επιδιώκουμε πάλι το σύνολο των υποψηφίων που θα εκλεγεί να είναι αντιπροσωπευτικό προς όλους. Για το λόγο αυτό, η εκτίμηση ενός ψηφοφόρου για ένα σύνολο k υποψηφίων προκύπτει αποκλειστικά από τον προτιμότερο εξ αυτών, δηλαδή τον κοντινότερο υποψήφιο που ανήκει στο εκάστοτε σύνολο. Τυπικά, η απόσταση που έχει ένας ψηφοφόρος i από ένα σύνολο X υποψηφίων μεγέθους k ορίζεται ως $d(i, X) = \min_{c \in X} d(i, c)$ και κατ' επέκταση το συνολικό κοινωνικό κόστος που προκύπτει από την επιλογή του συνόλου X υποψηφίων ορίζεται ως

$$SC(C, V) = \sum_{i \in V} \min_{c \in X} d(i, c)$$

Αν ήταν γνωστές όλες οι αποστάσεις το πρόβλημα θα αναγόταν για την επίλυση του προβλήματος *metric k-Medians*. Συγκεκριμένα στο πρόβλημα αυτό χρειάζεται να επιλέξουμε k κέντρα(υποψήφιους) και να αντιστοιχίσουμε τα υπόλοιπα σημεία(ψηφοφόρους) σε αυτά έτσι ώστε τα συμπλέγματα που θα δημιουργηθούν να είναι όσο το δυνατόν πιο συμπαγή. Το πρόβλημα *metric k-Medians* ανήκει στην κατηγορία των NP δύσκολων προβλημάτων. Για το λόγο αυτό, έχουν δοθεί αρκετοί προσεγγιστικοί αλγόριθμοι με σταθερό λόγο προσέγγισης που επιλύουν το πρόβλημα σε πολυωνυμικό χρόνο. Οι περισσότεροι από αυτούς τους αλγορίθμους χρησιμοποιούν τεχνικές προσέγγισης που βασίζονται στο γραμμικό προγραμματισμό.

Εμείς θα μελετήσουμε το distortion που δημιουργείται στο συγκεκριμένο πρόβλημα, όταν οι κανόνες ψηφοφορίας λαμβάνουν περιορισμένη πληροφορία.

Αρχικά οποιοσδήποτε μηχανισμός εκλογής k υποψηφίων σε μετρικό χώρο, ο οποίος λαμβάνει ως είσοδο τις κατατάξεις των ψηφοφόρων μπορεί να πετύχει τα εξής αποτελέσματα σε συνάρτηση με τον αριθμό k των νικητών:

- $k = 2$ τότε $dist = \Omega(n)$
- $k > 2$ τότε $dist = \infty$

Παρατηρούμε λοιπόν, πως οι κατατάξεις δεν αποτελούν επαρκή πληροφορία για να επιτευχθεί πεπερασμένη παραμόρφωση. Διαισθητικά η παραμόρφωση απειρίζεται όταν στην ανάθεση που πραγματοποιεί ο μηχανισμός, με την πληροφορία που διαθέτει, ένας η περισσότεροι ψηφοφόροι διανύουν απόσταση που μπορεί να είναι αυθαίρετα μεγαλύτερη σε σχέση με τις αντίστοιχες αποστάσεις στη βέλτιστη ανάθεση. Με άλλα λόγια, μετά την επιλογή του συνόλου των υποψηφίων W από τον κανόνα ψηφοφορίας δημιουργούνται k συμπλέγματα από ψηφοφόρους, όπου το καθένα αντιστοιχεί σε έναν υποψήφιο. Αν κάποιος από τους ψηφοφόρους διανύει αυθαίρετα μεγαλύτερη απόσταση από αυτή που θα διένυε αν είχε ανατεθεί στον βέλτιστο υποψήφιο, τότε η παραμόρφωση είναι άπειρη. Για το λόγο αυτό, δεν μπορούμε να γενικεύσουμε τις τεχνικές και τους μηχανισμούς που χρησιμοποιήσαμε στην περίπτωση της εκλογής ενός υποψηφίου ως νικητή σε μετρικό χώρο, καθώς σε εκείνη την περίπτωση δημιουργούνταν μόνο ένα σύμπλεγμα. Αξίζει να σημειωθεί ότι η στην περίπτωση της εκλογής ενός υποψηφίου ως νικητή, η τυχαία επιλογή ενός υποψηφίου ο οποίος κατατάσσεται πρώτος από τουλάχιστον ένα ψηφοφόρο, επιφέρει παραμόρφωση $\mathcal{O}(n)$.

Συγκεκριμένα, για το πρόβλημα της επιλογής k υποψηφίων, απαιτείται ποσοτική πληροφορία για τις αποστάσεις των ψηφοφόρων από τους υποψηφίους, ούτως ώστε να επιτευχθεί πεπερασμένη παραμόρφωση. Προτείνουμε λοιπόν έναν πιο αποδοτικό τρόπο να εξάγουμε πληροφορία για τις προτιμήσεις των ψηφοφόρων. Ζητάμε από κάθε ψηφοφόρο να δώσει μια εκτίμηση της απόστασης του από τους t κοντινότερος υποψηφίους του. Ορίζουμε δηλαδή συγκεκριμένα διαστήματα και ζητάμε από κάθε ψηφοφόρο να επιλέξει σε ποιο από αυτά τα διαστήματα ανήκει η πραγματική του απόσταση από τους t προτιμότερους υποψηφίους του. Αξίζει να σημειωθεί ότι όσο μεγαλύτερη είναι η απόσταση του ψηφοφόρου από ένα υποψήφιο που πρόκειται να καταθέσει, το αντίστοιχο διάστημα είναι μεγαλύτερο, καθώς είναι δύσκολο να γίνει μια τέτοια προσέγγιση με ακρίβεια. Για να ορίσουμε τα προαναφερθέντα διαστήματα θεωρούμε ότι είναι γνώστη η ελάχιστη απόσταση μεταξύ δύο υποψηφίων και την ορίζουμε ως d_{min}. Το διάστημα i ορίζεται:

$$b_i = \left\{ \begin{array}{lll} (d_{i-1}, d_i) & \text{όπου } d_i = \gamma \cdot d_{i-1} & , i \geq 1 \\ (0, d_0) & \text{όπου } d_0 = \frac{d_{min} - \epsilon}{2} & , i = 0 \end{array} \right\}$$

Από τον παραπάνω ορισμό, παρατηρούμε ότι τα μεγέθη των διαστημάτων αυξάνονται εκθετικά με το γ, οπού το γ είναι μια σταθερά. Σε αυτό το πλαίσιο μελετάμε την παραμόρφωση που μπορεί να επιτύχει κάθε μηχανισμός παραμετροποιώντας τον αριθμό t των υποψηφίων που απαιτείται να καταθέσει κάθε ψηφοφόρος. Όπως αναφέρθηκε το πρόβλημα metric k-Medians είναι NP δύσκολο. Στα επόμενα θεωρήματα, χρησιμοποιούμε έναν από τους υπάρχοντες προσεγγιστικούς αλγορίθμους με λόγο προσέγγισης β ως μαύρο κουτί. Η συνολική παραμόρφωση θα εξαρτάται από το β. Αν δεν ενδιαφερόμαστε για το υπολογιστικό κόστος τότε αντικαθιστούμε $\beta = 1$.

Θεώρημα 1.11. *Αν κάθε ψηφοφόρος καταθέτει τα διαστήματα στα οποία ανήκουν όλοι οι υποψήφιοι, δηλαδή $t = m$, τότε υπάρχει μηχανισμός που πετυχαίνει παραμόρφωση το πολύ $\beta \cdot \gamma$.*

Σε αυτήν την περίπτωση, ουσιαστικά, ο μηχανισμός διαθέτει μια εκτίμηση για την απόσταση μεταξύ κάθε ψηφοφόρου i και υποψήφιου j. Η εκτίμηση αυτή αποκλίνει το πολύ γ φορές από την πραγματική. Οπότε αν θεωρήσουμε ότι η κάθε απόσταση αυτές είναι ίση με το άνω όριο του αντίστοιχου διαστήματος και λύσουμε το πρόβλημα metric k-Medians με αυτές τις αποστάσεις χρησιμοποιώντας ένα προσεγγιστικό αλγόριθμο με λόγο β προκύπτει παραμόρφωση το πολύ $\beta \cdot \gamma$.

Θεώρημα 1.12. *Αν κάθε ψηφοφόρος καταθέτει τα διαστήματα στα οποία ανήκουν οι $t > \frac{m}{2}$ προτιμότεροι υποψήφιοι, τότε υπάρχει μηχανισμός που πετυχαίνει σταθερή παραμόρφωση.*

Το σημαντικό σημείο στην περίπτωση όπου ο αριθμός των υποψηφίων που δίνει κάθε ψηφοφόρος είναι $t > \frac{m}{2}$ είναι ότι κάθε ζεύγος ψηφοφόρων έχει καταθέσει τουλάχιστον ένα κοινό υποψήφιο.

Συνεπώς, παρά την μείωση του αριθμού t, η απόσταση μεταξύ κάθε ψηφοφόρου i και υποψήφιου j μπορεί να προσεγγιστεί και στη συνέχεια με αυτές τις προσεγγιστικές αποστάσεις λύνουμε το πρόβλημα το πρόβλημα *metric k-Medians*. Συγκεκριμένα οι αποστάσεις που θα χρησιμοποιήσουμε για την επίλυση του *metric k-Medians* ορίζονται ως εξής:

$$d_{ext}(i,c) = \left\{ \begin{array}{ll} d_i & \text{αν } c \in b_i \\ \min\limits_{\substack{j \in V \\ b \in b_i \cap b_j}} \left(d_{ext}(i,b) + d_{ext}(j,c) + d_{ext}(j,b)\right) & \text{αν } c \notin b_i \end{array} \right\}$$

Θεώρημα 1.13. *Αν κάθε ψηφοφόρος καταθέτει τα διαστήματα στα οποία ανήκουν οι $t = \frac{m}{2}$ υποψήφιοι, τότε κάθε κανόνας ψηφοφορίας έχει παραμόρφωση $\Omega(n)$ και υπάρχει μηχανισμός που πετυχαίνει παραμόρφωση $\mathcal{O}(n)$.*

Το κάτω όριο προκύπτει από το γεγονός ότι για $t = \frac{m}{2}$ υπάρχει περίπτωση οι ψηφοφόροι να χωριστούν σε δυο ομάδες όπου κάθε ομάδα δίνει τους ίδιους $t = \frac{m}{2}$ υποψηφίους. Συνεπώς οι αποστάσεις μεταξύ των ψηφοφόρων της μίας ομάδας και των υποψηφίων που έχουν δοθεί από την άλλη ομάδα ψηφοφόρων είναι άγνωστες και δεν μπορούν να προσεγγιστούν.

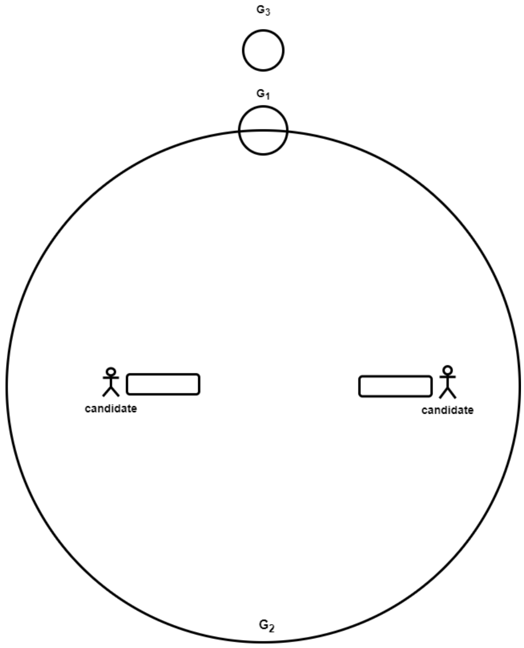
Ο μηχανισμός που πετυχαίνει παραμόρφωση $\mathcal{O}(n)$ λειτουργεί ως εξής:

- Έστω ότι δημιουργούνται δύο ομάδες ψηφοφόρων, όπου η κάθε ομάδα ψηφίζει τους ίδιους $t = \frac{m}{2}$ υποψηφίους. Το πρόβλημα χωρίζεται σε δύο υποπροβλήματα P_1, P_2. Λύνουμε το κάθε πρόβλημα ξεχωριστά για αριθμό νικητών $k_1 = 1, \ldots, k-1$ και $k_2 = 1, \ldots, k-1$ και κρατάμε την λύση που ελαχιστοποιεί το κοινωνικό κόστος τέτοια ώστε $k_1 + k_2 = k$.

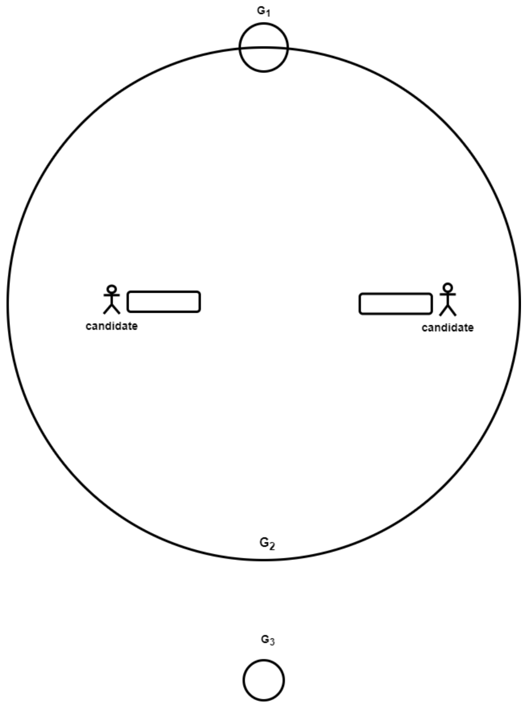
- Διαφορετικά λύνουμε το πρόβλημα με βάση το Θεώρημα 1.12

Πόρισμα 1.1. *Με βάση τα παραπάνω θεωρήματα γίνεται κατανοητό πως για να επιτύχει ένας μηχανισμός σταθερή παραμόρφωση με την πληροφορία που διαθέτει, πρέπει να μπορεί να ανακατασκευάσει προσεγγιστικά το μετρικό χώρο πάνω στον οποίο βρίσκονται οι ψηφοφόροι και οι υποψήφιοι.*

Θεώρημα 1.14. *Αν κάθε ψηφοφόρος καταθέτει τα διαστήματα στα οποία ανήκουν οι $t < \frac{m}{2}$ υποψήφιοι, τότε κάθε κανόνας ψηφοφορίας έχει άπειρη παραμόρφωση.*



(a) Πρώτο παράδειγμα



(b) Δεύτερο παράδειγμα

Σχήμα 1.3: Δυο παραδείγματα τα οποία οποιοσδήποτε μηχανισμός δεν μπορεί να τα διαχωρίσει

Θεωρούμε ότι ο αριθμός των ζητούμενων υποψηφίων είναι $t = \frac{m}{2} - 1$ και m είναι άρτιος. Ο αριθμός των νικητών είναι 3. Με βάση το Σχήμα 1.3 θεωρούμε ότι οι δύο μικροί κύκλοι G_1, G_3 αναπαριστούν δύο ομάδες ψηφοφόρων, όπου η κάθε μία έχει ψηφίσει τους ίδιους t υποψήφιους. Ο μεγάλος κύκλος G_2 αναπαριστά την τρίτη ομάδα ψηφοφόρων που δίνουν τους ίδιους υποψήφιους, όπου αυτοί είναι οι 2 εναπομείναντες υποψήφιοι $\left(m - 2 \cdot \left(\frac{t}{2} - 1\right)\right)$ και οι υπόλοιποι ανήκουν στο σύνολο των υποψηφίων που έχουν ψηφίσει οι ψηφοφόροι της ομάδας G_1. Η ομάδα G_2 χωρίζεται σε υποομάδες ψηφοφόρων, όπου κάθε μία απέχει αυθαίρετα μικρή απόσταση από ένα εκ των δύο εναπομένων υποψηφίων. Τα δύο παραδείγματα που φαίνονται στο σχήμα 1.3 δεν μπορούν να διαχωριστούν από κανένα κανόνα ψηφοφορίας. Στο πρώτο παράδειγμα η βέλτιστη λύση, αντιστοιχίζει τις ομάδες G_1, G_2 σε ένα σύμπλεγμα και κάθε μία από τις υποομάδες της G_2 σε ένα διαφορετικό σύμπλεγμα. Στο δεύτερο παράδειγμα, η βέλτιστη λύση αντιστοιχίζει τις δύο υποομάδες της G_2 σε ένα σύμπλεγμα και τις ομάδες G_1, G_3 σε δύο διαφορετικά συμπλέγματα. Οπότε, χωρίς βλάβη της γενικότητας, θεωρούμε ότι ο κανόνας ψηφοφορίας δίνει την ανάθεση με βάση το δεύτερο παράδειγμα. Οι αποστάσεις τότε που θα διανύσουν οι ψηφοφόροι των δύο υποομάδων στο πρώτο παράδειγμα είναι αυθαίρετα μεγαλύτερες από αυτές τις βέλτιστης λύσης και για το λόγο αυτό η παραμόρφωση είναι άπειρη.

Με βάση τον νέο τρόπο συλλογής των προτιμήσεων των ψηφοφόρων παρατηρούμε πως ένας μηχανισμός για να πετύχει πεπερασμένη παραμόρφωση χρειάζεται από κάθε ψηφοφόρο τις προσεγγιστικές αποστάσεις του από τουλάχιστον τους μισούς υποψηφίους. Αξίζει να σημειωθεί ότι το όριο των απαιτούμενων υποψήφιων για σταθερή παραμόρφωση δεν θα άλλαζε αν οι ψηφοφόροι κατέθεταν τις ακριβείς αποστάσεις για ίδιο αριθμό υποψηφίων. Χρησιμοποιώντας δηλαδή αυτό το μοντέλο απαιτείται αρκετή πληροφορία από κάθε ψηφοφόρο ούτως ώστε να επιτευχθεί πεπερασμένη παραμόρφωση. Αυτό συμβαίνει διότι η απόσταση ενός ψηφοφόρου από τους t προτιμότερους υποψηφίους, εν γένει, δεν μας εγγυάται ότι μπορούμε να άνω φράξουμε τις αποστάσεις με τους υπόλοιπους υποψηφίους.

Συμπεράσματα

Στις παραπάνω ενότητες μελετήσαμε διάφορους κανόνες εκλογής. Για να κρίνουμε πόσο αποτελεσματικά λειτουργούν οι κανόνες, δεδομένης της περιορισμένης πληροφορίας που δέχονται από τους ψηφοφόρους, χρησιμοποιήσαμε την μετρική της παραμόρφωσης. Συγκεκριμένα μελετήσαμε δυο διαφορετικά πλαίσια. Στο πρώτο, οι προτιμήσεις των ψηφοφόρων ως προς τους υποψηφίους εκφράστηκαν ως κανονικοποιημένες εκτιμήσεις που αθροίζουν στο 1 για κάθε ψηφοφόρο. Εξετάσαμε τις περιπτώσεις όπου χρειάζεται να εκλεγεί είτε ένας υποψήφιος ως νικητής είτε ένα σύνολο υποψηφίων και αποδείξαμε ότι οι κανόνες που εφαρμόστηκαν στην πρώτη περίπτωση μπορούν να γενικευτούν και στην περίπτωση εκλογής ενός συνόλου υποψηφίων λαμβάνοντας μάλιστα καλύτερα αποτελέσματα όσο αυξάνεται το μέγεθος του συνόλου των νικητών. Το δεύτερο πλαίσιο που εξετάσαμε ήταν αυτό στο οποίο ψηφοφόροι και υποψήφιοι αντιστοιχούν σε σημεία ενός μετρικού χώρου. Σε αυτήν την περίπτωση, οι προτιμήσεις των ψηφοφόρων ως προς τους υποψηφίους εκφράζονται ως αποστάσεις από αυτούς. Για τους κανόνες που μελετήσαμε, στόχος ήταν η ελαχιστοποίηση του κοινωνικού κόστους. Στην περίπτωση εκλογής ενός υποψηφίου ως νικητή, οι κατατάξεις των ψηφοφόρων ήταν αρκετές ώστε να επιτευχθεί χαμηλή παραμόρφωση και συγκεκριμένα δόθηκε ο αλγόριθμος που πετυχαίνει το κατώτατο όριο παραμόρφωσης, δηλαδή 3. Όσον αφορά το πρόβλημα της επιλογής ενός συνόλου υποψηφίων αποδείξαμε ότι οι κατατάξεις δεν αποτελούν αρκετή πληροφορία για να επιτευχθεί πεπερασμένη παραμόρφωση. Για το λόγο αυτό αναλύσαμε ένα πιο αποτελεσματικό τρόπο συλλογής των προτιμήσεων των ψηφοφόρων. Πάρα ταύτα για να επιτευχθεί πεπερασμένη παραμόρφωση απαιτείται αρκετή πληροφορία από κάθε ψηφοφόρο, καθώς ο ζητούμενος αριθμός υποψηφίων που πρόκειται να ψηφίσει ενδέχεται να βρίσκονται κοντά του και συνεπώς είναι αδύνατο να ανακτηθεί η θέση του στο χώρο σε σχέση με τους υπόλοιπους υποψήφιους. Αξίζει να σημειωθεί ότι το συγκεκριμένο πρόβλημα, από άποψη παραμόρφωσης δεν έχει μελετηθεί αρκετά και γι' αυτό αποτελεί ανοιχτό ερώτημα η εύρεση αποδοτικότερων μηχανισμών ώστε η πληροφορία που απαιτείται από κάθε ψηφοφόρο να είναι όσο το δυνατό λιγότερη. Το μοντέλο που προτείναμε, σέβεται ένα από τα βασικότερα αξιώματα της θεωρίας της κοινωνικής επιλογής, την ανωνυμία, που σημαίνει πως ο κανόνας ψηφοφορίας συμπεριφέρεται στον κάθε ψηφοφόρο ως ίσο. Πάρα ταύτα, πιστεύουμε ότι αξίζει να μελετηθεί το ισοζύγιο μεταξύ παραμόρφωσης και της συνολικής επικοινωνίας σε μηχανισμούς όπου εκτελούνται σε δύο στάδια: στο πρώτο θα εξάγεται ίση πληροφορία από κάθε ψηφοφόρο και στο δεύτερο θα αναγνωρίζονται ορισμένοι ψηφοφόροι ως αντιπρόσωποι και θα εξάγεται περαιτέρω πληροφορία από αυτούς. Επίσης, πρέπει να επισημαίνουμε ότι μελετήσαμε διάφορους κανόνες εκλογής ως προς την βελτιστοποίηση μιας συγκεκριμένης αντικειμενικής συνάρτησης, της ελαχιστοποίησης του συνολικού κοινωνικού κόστους (αντίστοιχα μεγιστοποίηση της κοινωνικής ευημερίας). Η παραμόρφωση των κανόνων εκλογής εξετάζεται και πάνω σε άλλες αντικειμενικές συναρτήσεις, όπως για παράδειγμα, η εκλογή ενός υποψηφίου έτσι ώστε να ελαχιστοποιείται η μέγιστη απόσταση που διανύει ένας ψηφοφόρος. Τέλος, αξίζει να σημειωθεί πως σε αυτή την διπλωματική εργασία μελετήσαμε μόνο ντερμινιστικούς μηχανισμούς. Στη βιβλιογραφία, έχουν αναπτυχθεί και τυχαιοποιημένοι μηχανισμοί, όπου λόγω της τυχαιότητας τα αντίστοιχα όρια της παραμόρφωσης είναι μικρότερα.

Κείμενο στα αγγλικά

Chapter 1

Introduction

In social choice theory, the primary objective is to analyse individual preferences in order to reach a decision that reflects on their collective opinion. Collective decision making problems come in different ways. They include problems of how to divide resources fairly, how to match people based on their preferences and generally how to aggregate individuals' opinions. We can think of many instances in real life in which individuals express their preferences over an issue and a mechanism aggregates their preferences in order to reach to an outcome that aligns best with the social opinion. Namely, such instances may be choosing an electoral candidate or a committee of candidates, a public policy or choosing the most suitable places to build public facilities. The most usual example, however is voting.

The roots of social choice go way back to ancient times and the first organized committees where individuals have to make decisions to serve their common cause. In the Middle Age, the Catalan philosopher Ramon LLull proposed that the outcomes of voting rules should be based on the pairwise contests between pairs of candidates. Yet, the first mathematical observations came during the period of Enlightenment, mainly thanks to the works of Jean-Charles de Borda and Marquis de Condorcet. The former proposed a new voting rule, widely known until today as Borda rule. The latter though, argued against his rule and highlighted a critical problem, called Condorcet paradox.

Moving forward in time, in the middle of the twentieth century, the work of Kennal Arrow generalised the problem observed by Condorcet. Specifically, Arrow proved that there is not any voting rule that satisfies simultaneously three keys properties: non-dictatorship, Pareto efficiency and Independence of Irrelevant Alternatives(IIA). Non-dictatorship states that the voting rule must not mimic the preferences of a single voter, the dictator. Pareto efficiency states that if each individual prefers alternative a to b, which is denoted by $a \succ b$, then the outcome of the voting rule must obey this societal preference order. Lastly, IIA states that the relative ranking between two alternatives a, b output by the voting rule must not relate to the individuals' preferences regarding a third alternative c. In the light of this classic impossibility result [1] as well as Gibbard and Satterthwaite impossibility result [2],[3] for axiomatic approaches to social choice, a new field of research emerged, the computational social choice(COMSOC). This is a field that combines computer science and voting theory. The arrival of computer science in the field of social choice led researchers to reexamine the old established problems from scratch and subsequently led to new fruitful questions. Research in computational social choice can be split in two directions. First, through the lens of computer scientists, social choice rules can be viewed as approximation algorithms and thus, one of the main fields of the research of computational social choice is to apply computational techniques in order to provide better analysis of social choice mechanisms. Second, researchers study the application of social choice theory to computational environments. For instance, it has been proposed that social choice theory can provide tools in joint decision making where software agents are heterogeneous and probably selfish. The context of this thesis belongs to the first field mentioned. We will study social choice rules under utilitarian view. In order to define formally the problem, we need to quantify voters' preferences over candidates. We will examine two different frameworks. First, we will adopt the classic theory of Von Neumann and Morgenstern [4] where individual preferences are captured through a utility function, which assigns numerical (or cardinal) values to each alternative. In this framework one of the most common objectives, is to find the outcome, i.e. select an alternative that maximizes the sum of voters' valuations. The second framework that we will examine is the one where agents are located in a metric space and thus voters' preferences are translated to distances from candidates. Here the distances correspond to *spatial preferences*, which mean that the metric space can be viewed as an ideological space in which a more preferred candidate would be closer to a voter. The spatial model of preferences has received an extensive amount of literature in the social choice theory [5], [6], [7], [8],[9], [10]. The problem in each case would be trivial if a mechanism acquired each voter's precise utility for each candidate. Though, it seems impractical to ask each agent to define precisely what her valuation is for each alternative. Furthermore, behavioral economists have argued that the agents cannot obtain the full information concerning their utility, or that obtaining this information requires a high cognitive cost. Additionally, sending more information also casts a higher burden on the mechanism. With that being said, each mechanism should elicit voters' preferences in a more conceivable way. For instance, an ordinal ranking over the prospective alternatives is, undoubtedly, a more natural and easy way to gather voters' preferences. Procaccia and Rosenschein proposed the framework of implicit utilitarian voting, whereby voters expressed ranked preferences over alternatives are seen as proxy for their underlying numerical utility functions. The implicit loss of information due to the embedding of cardinal preferences into ordinal render the mechanisms that elicit voters preferences in this way more difficult to distinguish the socially optimal outcome. Procaccia and Rosenschein [11] introduce the notion of distortion in order to quantify the drop in efficiency due to the loss of information. More formally distortion is a measure for the best worst-case approximation of the objective function that can be achieved given the available information. The notion of approximation due to some kind of limitations has been studied in other fields of computer science as well. Such notions are the approximation ratio used in approximation algorithms [12] in order to approximate NP-hard problems in polynomial time, the Price of Anarchy [13] used in game theory in order to measures loss of efficiency in worst case due to selfish behavior and the approximation ratio used in designing truthful mechanisms. It is important to note that in this thesis we assume that voters are non strategic, which means that the information they provide to the voting rule aligns with their true preferences. This is done, in order to keep with the line of work on analyzing the distortion of social choice functions, and avoid issues of game-theoretic modeling and equilibrium existence or selection, see [14]. We are going to examine both the single winner setting and the multiwinner setting. In the latter setting, the goal is to select a fixed number k of candidates as winners so as each individual would feel represented. Specifically, the metric multiwinner setting can be viewed as the well studied problem of facility location [15], [16], [17], where voters correspond to customers and candidates correspond to feasible locations for a new facility to open.

1.1 Organization

The first chapters of this thesis are dedicated to exploring the literature of distortion.

In Chapter 2 we give the formal definition of distortion and examine the single winner setting when agents have normalized valuations. We highlight the importance of normalized valuations in the framework of distortion and generally in the computational social choice theory. We review the lower bounds of distortion when the voting rule is given the ordinal preferences of voters. Furthermore, we examine the multiwinner setting with normalized valuations, where a subset of candidates with a fixed size has to be chosen. In this setting we consider that a voter's utility for a set of candidates derives only from her favorite alternative that belongs in the set. This definition captures the notion of proportional representation, which is usually a required property. We introduce *regret*, another measure of loss due to limited information and give the relation with distortion. We review the lower bounds of distortion, given the ordinal preferences of voters. Based on the lower bounds we conclude that in the multiwinner setting with normalized valuations, the problem become easier as the number of winners k increases.

In Chapter 3 we briefly review a slightly different work on distortion. In this context the information given is not a (partially) ranking of voters but is measured in bits, or in other words, measured in computational complexity. Specifically we examine the tradeoff between distortion and communication complexity both in single winner and multiwinner settings with normalized valuations. The literature of communication complexity lies slightly out of our the scope of our thesis, yet the results are intuitive.

In Chapter 4 we examine the single winner setting when agents and candidates are in a metric space. Due to the restrictions of the metric space the lower bound of distortion is constant, when voters give their ordinal preferences. We develop tools in order to prove upper bounds on many well known voting rules. Moreover we review the notion of majority graphs which are formed based on the ordinal preferences of voters. The notion is critical in order to construct mechanisms that achieve constant upper bounds. Lastly, we examine a slightly different framework where apart from the ordinal preferences of voters, the positions of alternatives are given too. In this framework there is a mechanism that achieves tight distortion using only the top choice of each voter.

1.2 Contribution

In Chapter 5 we present our results in the multiwinner setting when the agents lie in a metric space. We show that it is impossible to achieve bounded distortion given only the ordinal preferences. Therefore, mechanisms need more information, cardinal information. We propose a new framework in which voters provide an approximation of their real distances for a given number of their most preferred candidates. We investigate how the distortion increases as the number candidates asked per voter decays.

1.3 Social Choice Rules and Distortion

As stated, the objective of social choice theory is to analyse the individual preferences in order to conclude to an outcome that is best for the society. To do so, individuals submit their preferences in a fixed way and a social choice rule/mechanism aggregate their preferences and output an outcome that is considered socially optimal. A socially optimal outcome can be interpreted in a few ways. The most common ones are.

- **Utilitarian view**: Aims to maximize the sum of utilities
- **Egalitarian view**: Aims to maximize the minimum utility, thus maximizing fairness.

In this thesis we are going to examine the efficiency of the voting rules through the lens of *Utilitarian view*. In most cases we will assume that voters submit their ordinal preferences, which is a ranking of candidates based on their preferences. As a warm-up, we present some of the most popular social choice rules that elicit the ordinal preferences of voters and output a candidate as a winner.

Plurality Rule: The candidate who is ranked first by the most voters is selected as the winner

Borda Rule: For each vote, each candidate receives a number of points corresponding to the number of candidates ranked lower than him. For each ballot, the candidate ranked last receives 0 points, the next lowest candidate receives 1 point and so one. The candidate that gathers the highest total score, i.e. the highest borda count is selected as the winner.

Harmonic Rule: The voter awards the first ranked candidate 1 point, the second $\frac{1}{2}$ points, the third $\frac{1}{3}$ and so on. The candidate that receives the highest total score is selected as the winner.

Veto Rule: The voter awards each candidate one point except of one candidate. The candidate that receives the highest total score is selected as the winner.

k-Approval: The voter awards k candidates with 1 point and with 0 the others. The candidate that receives the highest total score is selected as the winner.

The above voting rules belong to the family of positional scoring rule where the ranking of each voter is translated to a scoring vector and the candidate with the highest total score is selected as winner. The next rule is the Single Transferable Vote, a multi round voting rule that is used in government elections in a few countries, such as Australia, India and Ireland.

Single Transferable Vote(STV)[18]: STV is a multi round voting rule that works as follows. In each round the candidate with the lowest plurality votes(votes in which a candidate is ranked first) is eliminated and these votes are transferred to the next most preferred candidate. After $m - 1$ rounds only one candidate is left and is chosen as the winner.

The following two rules choose the winner based on pairwise majority contests. The winner of the pairwise majority contest between two candidates a, b is the one such that if we restricted the voting to those two candidates, she would be preferred by the majority of voters.

Copeland Rule: The candidate who wins the largest number of pairwise majority contests is chosen as the winner.

Ranked Pairs: Initially, the rule computes the number of wins of each candidate against each candidate in their pairwise contest. Afterwards, the pairs are sort based on the number of wins in a decreasing order. A graph G is constructed as follows. For each pair candidates (x, y) in the sorted list, add the directed edge (x, y) if it will not create a circle. The winner is the source node o the resulting acyclic graph.

In the following chapters we will examine other more sophisticated social rules as well. As mentioned before, a social choice rule can be viewed as an approximation algorithm that tries to choose the best possible candidate. In order to measure how well a social choice rule works, given some fixed input, we will use the metric of distortion. We will establish lower and upper bounds. The lower bound of distortion, given a fixed input, such as the ordinal rankings, is independent of the social choice rule used. The lower bound derives only from the limitations of the respective input and none of the voting rules cannot achieve better distortion than the lower bound. The upper bound of the distortion of each mechanism will be the measure in order to review how well the mechanism performs given the respective information.

Chapter 2

Voters Preferences Expressed as Normalized Valuations

2.1 Single Winner Problem

In this framework, we assume that the preferences of voters are expressed as cardinal valuations. Formally, let V be a set of n voters and C be a set of m candidates. Each voter has a private valuation $v_i : C \to R_{\geq 0}$. Given the valuation profile of voters $\boldsymbol{v} = \{v_1, v_2, .., v_n\}$ the social welfare for an alternative $c \in C$ equals to:

$$SW(\boldsymbol{v}, c) = \sum_{i \in V} v_i(c)$$

Given the valuation profile of voters $\boldsymbol{v}$, it is trivial to select the candidate that maximizes the social welfare.
A social choice rule though can only have access to the preference profile $\boldsymbol{I}$. The preference profile $\boldsymbol{I}$ contains limited information compared to the valuation profile $\boldsymbol{v}$. The most common case is when the preference profile $\boldsymbol{I}$ corresponds to the ordinal preferences of the voters, that is a ranking in which each voter sorts the candidates based on her preferences. In this case, we will use the binary symbol $\succ$, to denote the preference of a voter between two candidates. For example for a voter i and two candidates x, y, the relation $x \succ_i y$ denotes that voter i prefers candidate x to candidate y.
We assume that each preference profile $\boldsymbol{I}$ is consistent with their real valuation profile $\boldsymbol{v}$. In other words, for a voter i and two candidates w, x if $v_i(w) \geq v_i(x)$, which means voter i values more candidate w than candidate x, this relation is apparent in her ordinal preferences and thus $w \succeq_i x$.

Definition 2.1. *The distortion of a social choice rule f that elicits the limited information I_i from each voter $i \in V$ and selects the alternative $f(V, \boldsymbol{I})$ as winner is defined as:*

$$dist((f, \boldsymbol{I})) = sup_{\boldsymbol{v} \rhd \boldsymbol{I}} \frac{max_{j \in C} SW(j|\boldsymbol{v})}{SW(f(V, \boldsymbol{I})|\boldsymbol{v})}$$

We note that distortion is a multiplicative measure of loss.

2.1.1 Impossibility Results with unrestricted valuations

In most cases the valuations of agents are studied under specific restrictions. The most common is the assumption that there exists a constant $K \in N$ such that for each voter i it holds that $\sum_{j \in C} v_i(j) = K$. Subsequently, the valuations can be normalized such that for each voter i it holds that $\sum_{j \in C} v_i(j) = 1$. The latter assumption has been termed as *unit-sum*. In the framework of distortion the normalized valuations are mandatory in order to bound the distortion. To justify this statement, consider an instance where the set of voters is $V = \{1, 2, 3\}$ and the set of candidates is $C = \{c_1, c_2\}$. The voting rule has only access to the ordinal preferences and these are given above.

- $c_1 \succ_1 c_2$
- $c_2 \succ_2 c_1$
- $c_2 \succ_3 c_1$

We fix the voters valuations as:

- $v_1(c_1) = c, v_1(c_2) = 0$, for $c > 0$
- $v_2(c_1) = 0, v_2(c_2) = 1$
- $v_3(c_1) = 0, v_3(c_2) = 1$

Therefore any voting rule, given only the ordinal preferences will choose candidate 2 as winner. Based on the voters valuations the social welfare of candidate c_1 is $SW(c_1|v) = c$ and $SW(c_2|v) = 2$ respectively for c_2. If $c > 2$ the optimal candidate is c_1 and thus the distortion is

$$dist(f) = \frac{SW(c_1, v)}{SW(c_2, v)} = \frac{c}{2}$$

and may be unbounded as c increases. The assumption of restricted valuations can also ensure that every voter has equal influence in determining the socially optimal outcome. The use of normalized valuations has been further analysed by Aziz [19]

2.1.2 Lower Bounds and Upper Bounds

Given the voters ordinal preferences, one of the easiest ways to aggregate their preferences is to select the candidate that is ranked first by the most candidates. This is called the Plurality rule, as mentioned in the previous section. We will first prove the upper bound of the distortion of Plurality rule.

Theorem 2.1. *The distortion of Plurality rule is $\mathcal{O}(m^2)$, where m is the number of voters.*

Before we proceed to the proof we highlight two useful bounds for the voters valuations that will help us establishing the lower bound of distortion. Let $\sigma_i(x)$ be the position voter i ranks candidate x.

- **Lower bound**: If $\sigma_i(x) = 1$, then $v_i(x) \geq \frac{1}{m}$
- **Upper bound**: $v_i(x) \leq \frac{1}{\sigma_i(x)}$

Proof Suppose that w is the candidate chosen as winner by the Plurality rule. By the pigeonhole principle it holds that at least $\frac{n}{m}$ voters ranked w first. For those voters, since the sum of valuations add up to 1, their valuation for candidate w is at least $\frac{1}{m}$. Hence, the social cost of w is at least $\frac{n}{m^2}$. The social cost for any candidate c is at most n, i.e. the edge case where every voter i has valuation $v_i(c) = 1$, and $v_i(c') = 0$, for $c' \neq c$. Therefore the worst case distortion of Plurality rule is $\frac{n}{\frac{n}{m^2}} = m^2$ □

Interestingly, we will see that this simple rule achieves the lowest possible distortion, given the ordinal preferences. Note, that we only use the most preferred candidate of each voter rather than the whole ranking. We recall, that the only restriction a social choice rule can capitalise, is that for each voter the valuations of all candidates add up to 1. Apart from that, the valuations can be totally abstract. In order to make this clear, we consider now the distortion of Borda rule.

Theorem 2.2. *The distortion of Borda rule is unbounded.*

Proof Consider an instance where two candidates $x, w \in C$ gather the highest score, i.e. the highest borda count. For candidate w all the voters ranked her second. For candidate x half of the voters ranked her first and the others ranked him third, and thus both candidates gather equal borda count. Without loss of generality, we assume that candidate w is chosen as the winner. We can fix the valuation profile of voters such that each voter's valuation for candidate w is 0 and 1 for their top choice and subsequently the candidate that maximizes social welfare is x. Hence, the distortion in this case is unbounded, since the social welfare of candidate w is 0. □

In other words, a voting rule that may choose a candidate that is not ranked first by none of the voters has unbounded distortion.

We will now see that the Plurality rule achieves tight distortion, by proving that the respective lower bound is $\Omega(m^2)$

Theorem 2.3. *Given the ordinal preferences, every social choice rule f has distortion $\Omega(m^2)$.*

Proof Let $C = \{c_1, c_2, \ldots, c_m\}$ be the set of candidates. We partition C in two sets: $X = \{c_1, \ldots, c_{m-2}\}$ and $Y = C \setminus X = \{c_{m-1}, c_m\}$. We construct the ranking profile as follow. Each of the $m-2$ candidates in set X is ranked first by $\frac{n}{m-2}$ voters and each of the 2 candidates in set Y is ranked second by $\frac{n}{2}$ voters. Any voting rule has to choose either one of the $m-2$ candidates in set X or one of two candidates in set Y and there is no way to distinguish a candidate in each set. Without loss of generality, suppose that the voting rule selects a candidate c_1 from set X or c_m from set Y. We now fix the valuation profile. For each voter that ranks c_1 first, set her valuations equal to $\frac{1}{m}$ for each candidate. For the remaining voters, for those who ranked candidate c_m second, set their valuations equal to 1 for their top candidate and 0 for the others. Lastly, for those who ranked c_{m-1} second set their valuations equal to $\frac{1}{2}$, for their top two candidates and 0 for the others. Thus for candidates c_1, c_m the incurred social welfare is $\frac{n}{m\cdot(m-2)}$. The social welfare for candidate c_{m-1} is $\frac{n}{4}$, which is the optimal. Therefore, the distortion is $\frac{m\cdot(m-2)}{4}$, yielding a lower bound of $\Omega(m^2)$. □

We conclude that if the voting rule is deterministic and the input is the ordinal rankings of voters, the problem is resolved since the Plurality rule achieves tight distortion. It is worth-noting though that, given the ordinal rankings, the only useful information that the voting rule gained for each voter regarding her cardinal valuations was a lower bound of $\frac{1}{m}$ for her valuation for her most preferred candidate. We are going to examine voting rules that elicit voters' preferences more efficiently in Chapter 3.

2.2 Multi Winner Problem (k-Voting)

In the multi winner elections the objective is to select a subset of k candidates as winners that maximizes the social welfare. Much of the work on multiwinner elections has focused on the question of proportional or diverse representation: How can we choose a winning committee where every voter feels represented. Multiwinner elections clearly arise in choosing a winning parliament in representative democracies. The most common framework that will be examined in this Chapter is again the implicit utilitarian voting, i.e. the case where the voters give their ordinal preferences. Many multiwinner rules rely on ideas from single winner rules. We provide some of them below:

- **Single Transferable Vote**: A multistage voting rule based on plurality scores. If there is a candidate w whose plurality score is at least $q = \frac{|V|}{k+1} + 1$ then do the following:
    - include w to selected subset

- delete the q votes in which w was ranked first
- remove w from all the remaining votes
- if the plurality score for each candidate is less than q remove the candidate with the lowest plurality score

- **Single Nontransferable Vote(k-Plurality Rule)**: Return the k candidates with the highest plurality score.
- **Bloc**: Return the k candidates with the highest approval votes.
- **k-Borda Rule**: Return the k candidates with the highest Borda score.
- **Chamberlin-Courant Rule [20]**: Given a committee X with k candidates, a voter is represented by the candidate that he ranks the highest among candidates from X. If $x(v)$ denotes the candidate that is representing voter v, the optimal committee under the Chamberlin-Courant voting rule seeks to minimize the sum or the maximum value of $(\sigma_v(x(v)))$, taken over all voters v (where $\sigma_v(c)$ denotes the ranking of the candidate c in the vote v).

In multiwinner voting the utility of voter for a subset of voters is not as straightforward to define as in the single winner setting. In order to capture the notion of proportional representation in the definition of voter's utility we assume that each voter has unit demand for a set of candidates [21]. This means that her utility for a set of candidates derives only from her favorite candidate in the set. In order to see the importance of unit demand valuations in proportional representation, suppose, for contrast, that each voter's utility for a set of candidate derives from all the candidates in the set. Consider an instance with 5 voters and the objective is to select a committee of 5 candidates. Based on the latter definition, if 3 voters have the same valuations for the candidates, the optimal set may include candidates that are preferred only by 3 out of 5 voters and thus leaving the remaining voters unrepresented.

We can now define formally the utility a voter i has for a subset S of candidates:

$$v_i(S) = \max_{X \in S} v_i(X)$$

Therefore the social welfare for a set S of k candidates and a valuation profile $\boldsymbol{v}$ is

$$SW(S, \boldsymbol{v}) = \sum_{i \in V} \max_{X \in S} u_i(X)$$

We assume again that voters have unit-sum valuations.

2.2.1 Lower and Upper Bounds

The distortion in the multiwinner elections, given only the ordinal preferences of voters was studied by Caragiannis et al [22]. Some notions in single winner framework can be naturally generalized for the multiwinner setting and will be useful in this analysis.

Definition 2.2. *For a preference profile $\boldsymbol{\sigma}$, that is the rankings of candidates, given by the voters and a set of candidates $S \subset C$, we denote*
$\sigma_i(S) = \min_{x \in S} \sigma_i(x)$, the candidate in the set that voter i ranks the highest ranking. Moreover for a voter i we can define the relation between two set of candidates X, Y as $X \succ_i Y$ which means that from the candidates in these two sets, voter i prefers a candidate in set X

Definition 2.3. *The plurality score of a set $S \subseteq C$ is the number of votes in which alternatives in S are ranked first, i.e:*

$$plu(S, \sigma) = \sum_{x \in S} plu(x, \sigma)$$

In their work, they established certain lower bounds and showed that the *k-plurality rule* achieves tight distortion. In order to prove those bounds we need to define another measure of loss, the regret. In the next definitions, we will use A_k to denote all the subsets of candidates of size k.
For a multiwinner voting rule f which chooses the subset X given the preference profile σ the regret is defined:

$$regret(f) = \frac{1}{n} \sup_{v \triangleright \sigma} \left(\max_{S \in A_k} SW(A, v) - SW(X, v) \right)$$

Regret is an additive measure of loss, whereas distortion is a multiplicative. The following lemma demonstrates how distortion and regret are related.

Lemma 2.1. *For a voting rule f and a preference profile σ that select the subset of candidates X the regret of f is given by*

$$reg(f) = \max_{S \in A_k} \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{\mathbb{1}[S \succ_i X]}{\sigma_i(S)}$$

and the distortion is given by

$$dist(f) = 1 + m \cdot \frac{n \cdot reg(f)}{plu(X, \sigma)}$$

Proof We give the sketch of how these formulas are proved. As you may observe, the proof will be similar to the one in the single winner setting. We will use again the two useful bounds established in the previous Chapter. As stated, both distortion and regret are measures based on worst case analysis. We want to fix a valuation profile $\boldsymbol{v^*}$ such that the chosen subset X has the least possible social welfare under $\boldsymbol{v^*}$. For each voter i compare the chosen set X and the arbitrary set S. If $S \succ_i X$ we want to maximize the difference between $v_i(S)$ and $v_i(X)$ and we do so by setting $v_i(X) = 0$ and $v_i(S) = \frac{1}{\sigma_i(S)}$ which is the highest possible valuation. Remind that $\sigma_i(S)$ is the highest ranked candidate in subset S based on voter i.
If $X \succ_i S$ we want to minimize the difference between $v_i(S)$ and $v_i(X)$ which means set the two valuations equal and want to be as small as possible in order to maximize the distortion. This is achieved by setting both utilities zero if $\sigma_i(X) > 1$ and by setting both valuations $\frac{1}{m}$ if $\sigma_i(X) = 1$, which is the lowest possible valuation for a top candidate. □

The above formula of distortion formalizes the importance of plurality score for each candidate in the framework of distortion. It is easy to observe that a voting rule which selects a candidate x with $plu(x, \boldsymbol{\sigma}) = 0$ has unbounded distortion.
By using Lemma 2.1 and following similar reasoning when we prove the lower bound in the single winner setting, we establish the following lower bounds for every deterministic voting rule f, depending on k

- For $k \leq m/6$ $dist(f) \geq 1 + m \cdot \frac{m-3k}{6k}$
- For $k \leq m/2$ $dist(f) \geq 1 + m$
- For $k \geq m/2$ $dist(f) \geq 1 + m \cdot \frac{m-k}{k}$

As it was mentioned above, the voting rule f that selects the k candidates with the highest plurality score achieves tight distortion.

Theorem 2.4. *The k-Plurality rule achieves distortion of at most* $1 + m \cdot \left(\frac{m}{k} - 1\right)$

Proof We consider that the rule selects the subset X of candidates. We know that the sum of plurality scores of all the candidates equals n. By the pigeonhole principle the sum of the k highest plurality scores is at least $k \cdot n/m$. Moreover for every $S \in A_k \setminus X$ the number of voters i for whom it holds $S \succ_i X$ is at most $n - plu(X, \boldsymbol{\sigma})$. Therefore by Lemma 2.1 the distortion of this rule is at most:

$$1 + m \cdot \frac{n \cdot reg(f)}{plu(X, \sigma)} \leq 1 + m \cdot \frac{n - plu(X, \sigma)}{plu(X, \sigma)} = 1 + m \cdot \left(\frac{n}{plu(X, \sigma)} - 1\right)$$
$$\leq 1 + m \cdot \left(\frac{m}{k} - 1\right)$$

□

Given only the ordinal preferences of voters, the problem has been settled down as we reviewed that plurality rule achieves tight distortion, by just using the top choice of each voter. It is worth-noting that the lower bounds of distortion, given the ordinal preferences, decay as k increases. Less formally, the problem becomes easier as k increases. This is not obvious apriori because by increasing the value of k, this allows us returning larger sets that achieve higher social welfare, but it also raises the optimal social welfare against which a voting rule needs to compete. As we will see in the next Chapter, the observation that distortion decays as the number or winners increases, can be generalised with any given input.

Chapter 3

Communication-Distortion Tradeoff with normalized valuations

So far, in the normalized valuations setting, we examine rules that elicit information from voters in a specific way which is the ordinal rankings. In this case the bits of information required from each voter is $\Theta(m \cdot logm)$. This holds as the possible different responses that each voter can submit are $m!$. Asymptotically, by taking the logarithm of this quantity we get $\Theta(\log m!) = \Theta(m \log m)$ bits of information. This is the communication complexity of those voting rules. For further information about the communication complexity in voting rules see [23]. Obviously, it is desirable for a voting rule to have low communication complexity and low distortion. Typically, though by eliciting more information from the voters enables voting rules to achieve lower distortion. Mandal et al. [24] took a novel approach in analyzing voting rules. They considered them as a combination of two rules:

- **Elicitation rule**: Asks voters to answer a query based on their preference
- **Aggregation rule**: Outputs a set of candidates as winners based on the information collected in the elicitation rule

Working on this framework they examine the frontier of the tradeoff between distortion and communication complexity and they provide upper and lower bounds on the communication complexity required in order to achieve at most distortion d. They implement both deterministic and randomized rules. A randomized voting rule assigns a probability distribution over all feasible outcomes. For example in the single winner setting, the voting rule assigns a probability distribution over all candidates and in the multiwinner setting, with k winners, assigns a probability distribution over all possible sets of k candidates. In this case the distortion is the worst case ratio between the optimal social welfare and the expected social welfare of the subset selected by the voting rule. In this section though, we are going to review only the deterministic rules. The two rules that we are going to examine rely on an input format called *threshold format* whereby each voter is asked to submit whether her utility for each candidate is above or below a given threshold. This idea was introduced by Benadè et al. [25]. As we will see, the next algorithms generalize this idea by constructing more than one *threshold formats*.

First we review a voting rule called $PREFTHRESHOLD_{t,l}$ for the single winner setting, which uses deterministic elicitation and deterministic aggregation. The idea is that each voter is asked to submit approximately the valuations for her top t candidates. In order to approximate voters' valuations the interval $[0, 1]$ is split into l exponentially spaced buckets. Therefore each voter submits in which bucket l belong her valuation from each of her top t candidates. The aggregation rule outputs the candidates with the highest estimated welfare, which is derived from the above approximations in the elicitation step. That is, the rule sums up the upper bounds of the bucket in which each valuation belongs. The formal algorithm is given above.

Algorithm 1: $PREFTHRESHOLD_{t,l}$

Elicitation Rule:

Partition the interval $[0, 1]$ in $l + 1$ buckets, where $B_0 = (0, \frac{1}{m^2}]$ and $B_p = (m^{-1+(p-1/l)}, m^{-1+(p/l)}]$, for $p = 1, \dots, l$

Ask each voter i to pick the set S_i^t of the t most preferred candidates and identify the bucket B_p each valuation $v_i(a)$ belong for $a \in S_i^t$

Aggregation Rule:

For each p let U_p denote the upper bound the respective bucket B_p

For each voter i and candidate a define $\hat{v}_i(a) = U_p$ if $v_i(a) \in B_p$ and $\hat{v}_i(a) = 0$ otherwise

For each candidate a define $\hat{sw}(a) = \sum_{i \in V} \hat{v}_i(a)$

Output the candidate x such that $w \in \arg\max_{a \in C} \hat{sw}(a)$

The next theorem provides the bounds of the voting rule $f = PREFTHRESHOLD_{t,l}$

Theorem 3.1.
For $1 < t \leq m$, the communication complexity and distortion are:

$$C(f) = \Theta\left(t \cdot log\left(\frac{m(l+1)}{t}\right)\right), \quad dist(f) = \mathcal{O}\left(m^{1+2/l}/t\right)$$

For $t = 1$ it holds

$$C(f) = \log(ml), \quad dist(f) = \mathcal{O}\left(m^{1+1/l}\right)$$

Proof The desired communication complexity derives from the fact that a voter can give $\binom{m}{t} \cdot (l+1)^t$ different responses for $t > 1$ and $m \cdot l$ for $t = 1$. And asymptotically by taking the logarithm we conclude that:

- $C(f) = \Theta\left(\log\left[\binom{m}{t} \cdot (l+1)^t\right]\right) = \Theta\left(t \log\left(\frac{m(l+1)}{t}\right)\right)$, for $t > 1$
- $C(f) = \log(ml)$, for $t = 1$

□

As expected, if t or l increases the communication complexity increases and the distortion drops. The above parameterization provides us the ability to make some remarks:

- For $t = 1$ and $l = 2$, we get distortion of $\mathcal{O}(m\sqrt{m})$ with communication complexity $\Theta(\log m + 1)$
- For $t = m$, $l = \log m$, we get constant distortion with communication complexity $\mathcal{O}(m \log\log m)$

We recall that by eliciting the ordinal preferences of voters which translates to $\Theta(m \log m)$ bits, we got a lower bound of distortion at $\Omega(m^2)$ for the single winner. However we proved that the plurality rule achieves the optimal distortion given the ordinal rankings, which in fact can be achieved by providing only the first vote which translated in $\log m$ bits of information. Thus, with one more bit of information *PREFTHRESHOLD* gives subquadratic distortion of $\mathcal{O}(m\sqrt{m})$. We can infer that the ordinal preferences are not the optimal way to elicit information from the voters.

In their follow up work Mandal et al. [26] gave a similar algorithm for the multiwinner voting setting with deterministic elicitation and randomized aggregation. Again the idea is to ask voters more efficiently about their preferences rather than giving their ordinal preferences.

We provide again the idea of the algorithm. The interval of $[0,1]$ which is the range within the valuations for each candidate belong, is partitioned into $\log m$ exponentially value-sized buckets. Furthermore, the set of integers $[m] \cap \{0\}$ is partitioned into $\log m$ exponentially quantity-buckets. For every combination of each value-sized bucket and each quantity-sized bucket the voters are asked to give the set of candidates for whom the valuations belong to the respective value bucket, if the cardinality of this set belongs to the quantity bucket. In order to keep communication complexity low, it is not necessary to ask voters to submit sets with high cardinality, because in these sets the valuations for each candidate are approximately the same. Thus, the rule learns approximately the distribution of each voter valuation. The aggregation rule forms a number of possible subsets of k candidates based on the voters preferences and picks one of them randomly. The formal algorithm is given below.

Algorithm 2: k-Selection

Elicitation Rule:

Set $t = \frac{t}{144 \log^4 m}$

Partition the interval $[0,1]$ into $2 \log m + 1$ buckets: $B_0 = [0, \frac{1}{m^2})$ and $B_j = [\frac{2^{j-1}}{m^2}, \frac{2^j}{m^2})$ for $j = 1, \ldots, 2 \log m$

Partition the interval $[m] \cup \{0\}$ into $\log m + 1$ buckets: $C_0 = \{0\}$, $C_j = \{2^{j-1}, \ldots, 2^j\}$ for $j = 1, \ldots, \log m$

Set $q_s = \log \frac{m}{tk}$

For each pair (p,q) such that $p \in \{1, \ldots, 2 \log m\}$ and $q \in \{1, \ldots, q_s\}$

  Each voter i calculates set S_i^{pq} of candidates a such that $v_i(a) \in B_p$

  Send S_i^{pq} if $|S_i^{pq}| \in C_q$.

Aggregation Rule:

For each pair (p,q) such that $p \in \{1, \ldots, 2 \log m\}$ and $q \in \{1, \ldots, q_s\}$

  IF $q > q_s$ select $\hat{A}_{pq}$ uniformly at random from $\{S \subseteq C : |S| = k\}$

  ELSE Obtain S_i^{pq} from each voter i.

    Choose $\hat{A}_{pq} \in \arg \max_{S \subseteq C : |S| = k} \{S_i^{pq} \cap S \neq \emptyset\}$

Output uniformly at random one of $(1 + 2 \log m) \cdot (1 + \log m)$ subsets of $\hat{A}_{pq}$

Theorem 3.2. *For $d \geq 144 \log^4 m$ there is a voting rule with deterministic elicitation with $\mathcal{O}(d)$ distortion and communication complexity $\mathcal{O}\left(\frac{m}{kd} \log^6 m\right)$*

Proof Suppose we have a valuation profile V^p for each $p \in \{0, 1, .., 2 \log m\}$ that corresponds to the respective upper bound of bucket B_p. We can find the optimal sets of k candidates A_p^* for each p.

Since $\sum_{p=1}^{2 \log m} SW(A_p^*, V_p) \geq SW(A^*, v)$, by the pigeonhole principle there exists p $\in \{0, 1, ..2 \log m\}$ such that

$$SW(A_p^*, V_p^*) \geq \frac{1}{1 + 2 \log m} \cdot SW(A^*, v)$$

. Following the same reasoning we can define the valuation profile V^{pq}, where for each voter i it holds

$$V_i^{pq} = \left\{ \begin{array}{lll} V_i^0 & if |\{a : V_i^0(a) = \frac{1}{m^2}\}| \in C_q & and \quad p = 0 \\ V_i^p & if |\{a : V_i^p(a) = \frac{2^p}{m^2}\}| \in C_q & and \quad p \geq 1 \\ 0 & otherwise & \end{array} \right\}$$

It holds that $\sum_{q=0}^{\log m} SW(A_{pq}^*, V^{pq}) \geq SW(A_p^*, V^p)$. By pigeonhole principle there exists $q \in \{0, 1, .., \log m\}$ such that

$$SW(A_{pq}^*, V^{pq}) \geq \frac{1}{1 + \log m} SW(A_p^*, V^p)$$

Combining the two results we get that there is a pair (p', q') such that

$$SW(A^*_{p'q'}, V^{p'q'}) \geq \frac{SW(A^*, v)}{(1+\log m)\cdot(1+2\log m)} \geq \frac{SW(A^*, v)}{6\log^2 m}$$

The next step is to bound the expected value of the social welfare of $\hat{A}_{pq}$ selected by the mechanism with the optimal subset A^*_{pq} with respect to valuation profile V^{pq}. For every voter i and candidate a it holds that $v_i(a) \geq \frac{V_i^{pq}}{2} - \frac{1}{2m^2}$. This inequality is relaxed in order to be true for $p = 0$ as well. Hence for any subset $S \subset A$ such that $|S| = k$ we get

$$SW(S, u) = \sum_{i\in V} \max_{a\in S} v_i(a) \geq \sum_{i\in V} \frac{V_i^{pq}}{2} - \frac{n}{2m^2} = \frac{1}{2} sw(S, V^{pq}) - \frac{n}{2m^2}$$

.

It is important to note that when $q < q_s$ the subset $\hat{A}_{pq}$ selected by the mechanism is equal to the respective A^*_{pq}.
Hence the only thing left to bound is the social welfare of the random subset $\hat{S}$ selected when $q > q_s$. In this case it holds that $\mathbb{E}(sw(\hat{S}, V^{pq})) \geq \frac{1}{2t} \cdot sw(A^*_{pq}, V^{pq})$
Thus, irrespective of the value of q we can bound the expected value of the social welfare of $\hat{A}_{pq}$ selected by the mechanism with respect to the original valutation profile v

$$\mathbb{E}(sw(\hat{A}_{pq}, v)) \geq \frac{1}{2} sw(\hat{A}_{pq}, V^{pq}) - \frac{n}{2m^2} \geq \frac{1}{4t} \cdot sw(\hat{A}_{pq}, V^{pq}) - \frac{n}{2m^2} \geq \frac{sw(A^*, v)}{24t\log^2 m} - \frac{n}{2m^2}$$

Lastly, by dividing with the optimal social welfare we get:

$$\frac{\mathbb{E}(sw(\hat{A}_{pq}, v))}{sw(A^*, v)} \geq \frac{1}{24t\log^2 m} - \frac{n}{2m^2 \cdot sw(A^*, v)} \geq \frac{1}{24t\log^2 m} - \frac{1}{2m} \geq \frac{1}{48t\log^2 m}$$

The second inequality holds due to the fact that the social welfare is a non decreasing function with respect to the number of winners and for the single winner there is at least one candidate a such that $sw(a, v) \geq \frac{n}{m}$ and consequently $sw(A^*, v) \geq \frac{n}{m}$. The last inequality holds by the definition of $t = \frac{d}{144\log^4 m} \leq \frac{m}{24\log^2 m}$. Lastly, remind that the mechanism picks one of the $\hat{A}_{pq}$ subset uniformly at random. Thus, the probability that the pair (p', q') is selected is $\frac{1}{(1+2\log m)\cdot(1+\log m)} \geq \frac{1}{6\log^2 m}$. Hence, we conclude that the distortion is at most $d = 144\log^4 m$

For the communication complexity, each voter submits $(1 + 2\log m)\cdot(1 + \log m) \leq 6\log^2 m$ sets. The size of each set is at most $\frac{m}{tk} = \frac{144m\log^4 m}{dk}$. Hence the total communication from each voters is at most $\mathcal{O}(\frac{m}{dk}\log^6 m)$ bits. □

The authors gave respective lower bounds for both problems. These are presented below:

- Let f be a single winner voting rule which achieves distortion $dist(f) \leq d$. Then $C(f) = \Omega(\frac{m}{d})$
- Let f be a multiwinner voting rule which achieves distortion $dist(f) \leq d$. Then $C(f) = \Omega(\frac{m}{kd})$

These lower bounds are proved using tools from the multi-part communication complexity literature. Specifically the authors reduce the voting problem to the multi-party set disjointness problem. We are not going into details as communication complexity is out of our scope. For further reading about some of the techniques used communication complexity theory we refer the reader to [27]. Based on the latter lower bounds, the two rules that we review achieve tight distortion up to some logarithmic factors. Another important observation is that the lower and upper bounds decrease with k. This formalises the statement that as number of winner increases the problem become easier, which means that the voting rules need lesser information.

Chapter 4

Single Winner Problem in a Metric Space

In this chapter we will examine the setting where agents and candidates lie in a metric space. In this setting, the underlying assumption is, that the closer a candidate is to a voter, the more similar their positions are on a specific subject. We assume again that voters provide their ordinal preferences. Thus voters rank candidates by increasing distances from them. We begin with the simplest metric space, the line ($\mathbb{R}$). A natural way to interpret this setting is as left wing and right wing line, where voters and candidates are located on the line. The structure of line led scientists to establish two significant properties about the formation of candidates and voters. First, the group of voters are said to have single peaked preferences over the set of candidates. We define formally the single preference profile.

Definition 4.1. *We say that a preference profile is single peaked if there exists an ordering σ, denoted by $\succ$, over the set C of candidates such that the ordinal preference of every voter i denoted by $\succ_i$ has the following structure. Let $top(i)$ be the most preferred candidate of i. For every pair of candidates x, y such that $x \succ_i y$ in i's ordinal preference, we have either $top(i) \succ x \succ y$ or $y \succ x \succ top(i)$ in the ordering σ.*

The above definition indicates that the further a candidate is located from voter i's top choice, the less is preferred by her and that there is an ordering of candidates consistent with the ordinal preferences of voters. This specific class of preference relations implies a number of fruitful properties and has been studied extensively by political and computer scientists [28], [29], [30], [31], [32], [33], [34], [35].
The second established property states that there is an ordering over the set V of voters as well. This property is called single crossing [36].

Definition 4.2. *We say that a preference profile belong to the single crossing domain, if it admits a permutation of voters such that for any pair of candidates x, y there is an index $j(x, y)$ such that either all voters i with $i < j(x, y)$ prefer candidate x to y and all voters i with $i > j(x, y)$ or vice versa.*

In other words, preferences are single-crossing if there exists a linear ordering of voters such that for any pair of candidates there is a single spot, along this ordering, where the voters switch from preferring one candidate to the other one. Scientists have analysed the multiple aspects of the majority rule under single crossing preferences [37],[38]. Furthermore, single crossing preferences relate to the field of income redistribution [39], coalition formation [40], local public goods [41], [42] and with the choice of constitutional voting rule [43].

Since the setting of line may seem as a simplistic way to interpret political views, we can consider the political compass in 4.1, where candidates and voters correspond to points in the plane, in which the horizontal axis measures their *economic views* and the vertical axis measures their *social views*.

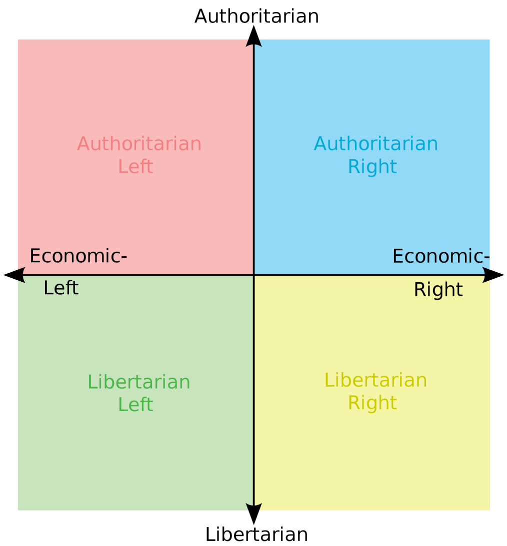


Figure 4.1: Political compass where the dimensions correspond to economic and social views

Going a step beyond, political scientists have analysed the so called *Downsian proximity model* [29]. One way to interpret this model, is as a high dimensional metric space where each dimension corresponds to a different issue. And thus, the more issues a voter aligns with a candidate, the more close to her he will be. As far as the distortion of each mechanism is concerned, the lower and upper bounds that we will examine hold for every metric space.

As stated, in these cases, voters preferences are defined as distances over the candidates and we aim to select a candidate that minimizes the total social cost, i.e. the sum of distances of all agents from the chosen alternative. In this framework we expect to get tighter bounds since the restrictions of the metric space reduce the level of arbitrariness in the voters' preferences.

4.1 Model

Let V be a set of n voters, C be a set of m candidates and we define the set of all agents $A = V \cup C$. We define also $d : A \times A \rightarrow R_{\geq 0}$, a non-negative function that measures the distance between points that correspond to voters and candidates. Specifically, d satisfies the above properties:

- identity of indiscernibles: $d(x, x) = 0$
- symmetry: $d(x, y) = d(y, x)$
- triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$

The pair (A, d) defines the metric space.
The social cost for a candidate $c \in C$ is defined as

$$SC(c, V|d) = \sum_{i \in V} d(i, c)$$

For a social choice rule f, that is given the ordinal ranking profile of voters $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ and selects alternative $f(V, \sigma) = w \in C$ as winner, the distortion is defined as:

$$dist(f) = \max \frac{SC(f(V, \sigma), V|d)}{\min_{x \in C} SC(x, V|d)}$$

With slight abuse of notation we will often omit the metric space d in our definitions, by writing $SC(w, V)$. We define also some useful notation:
For candidates x, y $xy = \{i \in V : x \succ_i y\}$ is the set of voters that prefer x to y. Similarly, for candidates x, y, z we define the set $xyz = \{i \in V : x \succ_i y \succ_i z\}$.
Moreover, we say that candidate x (weakly) defeats candidate y if $|wx| \geq \frac{n}{2}$

4.2 Lower Bounds

As stated, in the following sections we assume that the information given by each voter i to mechanisms is a preference ranking σ_i over the candidates in C where voter i ranks candidates in a non-decreasing order of their distance from her. We assume that d is consistent with the preference ranking σ_i if and only if for voter i and for every candidates c, c' where $d(i, c) \leq d(i, c')$ follows that $c \succ_i c'$.

Theorem 4.1. *Given the voters' ordinal preferences there is no (determinstic) social choice rule that has distortion less than 3.*

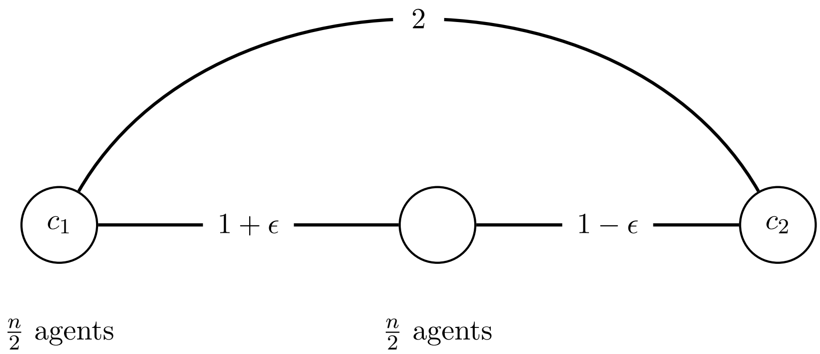


Figure 4.2: Lower bound

Proof Consider the case where there are only two candidates c_1, c_2. Half of the voters prefer c_1 and the other half of them prefer c_2. Without loss of generality we assume that a social choice rule chooses c_2 as the winner.
We can design the following instance as illustrated in Figure 4.2. The voters who prefer c_1 are collocated with him, therefore for them it holds that $d(i, c_1) = 0$ and also $d(i, c_2) = 2$.
For the other half of the voters who prefer c_2, fix the distances for them as: $d(i, c_2) = 1-\epsilon$ and $d(i, c_1) = 1+\epsilon$. The total social cost incurred by selecting c_2 is $SC(c_2, V) = 2 \cdot \frac{n}{2} + (1-\epsilon) \cdot \frac{n}{2}$.
Whereas the optimal candidate is c_1 and the social cost in this case is $SC(c_1, V) = (1+\epsilon) \cdot \frac{n}{2}$
Therefore the distortion equals $\frac{SC(c_2,V)}{SC(c_1,V)} = \frac{2 \cdot \frac{n}{2} + (1-\epsilon) \cdot \frac{n}{2}}{(1+\epsilon) \cdot \frac{n}{2}}$ which approaches 3 as $\epsilon \to 0$ □

Thus, no matter how efficient a deterministic voting rule will be, there will always be an instance in which the voting rule will select a candidate that yields social cost at least 3 times greater than the optimal candidate.

4.3 Lower Bounds on Positional Scoring Rules

As mentioned in the first chapter, many of the well known voting rules belong to the family of positional scoring rules. A positional scoring rule maps the voter's ordinal preference to a

scoring vector $\boldsymbol{s} = \{s_1, s_2, .., s_m\}$, where $s_1 \geq s_2 \geq ... \geq s_m$. If a voter i ranks candidate c at position j then he receives s_j points from her. The winner is the alternative that gathers the most total points.
We present the scoring vectors for some positional scoring rules:

- Plurality : $\boldsymbol{s} = \{1, 0, .., 0\}$
- Borda : $\boldsymbol{s} = \{m - 1, m - 2, .., 0\}$
- Harmonic : $\boldsymbol{s} = \{1, 1/2, .., 1/m\}$
- Veto : $\boldsymbol{s} = \{1, 1, .., 1, 0\}$
- k-Approval : $\boldsymbol{s} = \{1, 1, .., 1, 0, ..., 0\}$

For the positional scoring rules we consider two cases in order to prove lower bounds on distortion

- $s_1 = s_2$. In this case the worst-case distortion is unbounded. Consider the instance where the candidates are placed on the real line and the position of each candidate $c_j \in C$ is j. All the voters all collocated with candidate c_1. Therefore, for $j > 1$ the distance between candidate c_j and each voter i is $d(i, c_j) = j - 1$. The winners based on this positional scoring rule are candidates c_1 and c_2 aggregating total score of $n \cdot s_1 = n \cdot s_2$. The optimal candidate is c_1 yielding social cost of 0, whereas the rule may choose candidate c_2. The distortion in this case is unbounded.

- $s_1 > s_2$. For every positional scoring rule f with this restriction, there exists a voting preference profile such that $dist(f) \geq 1 + 2\sqrt{\ln m - 1}$. For the technical proof of this proposition we refer the reader to original paper [44].

The first bullet states that each positional scoring rule must respect each voter's top choice, in order to achieve bounded distortion. As we can observe, the lower bound of distortion of $1 + 2\sqrt{\ln m - 1}$ for positional scoring rules scales with the number m of candidates and diverges from the general lower bound of 3 as m increases. This result highlights the problem of positional scoring rules in the context of distortion. Voters embed their cardinal preferences into ordinal and the position scoring rules remap the ordinal preferences into cardinal ones with a fix way. This remapping may differ compared to the voters' original cardinal preferences.

Lemma 4.1. *The scoring rules $f = \{Veto, k\text{-}Approval\}$ belong to the first case. Therefore the distortion of these rules is unbounded.*

4.4 Upper Bounds on Popular Voting Rules

4.4.1 Linear Upper Bounds

First we will show the upper bounds that have been proven on the above positional scoring rules. In order to get the upper bounds on the distortion of Plurality and Borda rules, a critical lemma is used.

Lemma 4.2. *For every pair of alternatives x, w it holds that:*
$\frac{SC(w,V)}{SC(x,V)} \leq \frac{2n}{|wx|} - 1$*, where $|wx|$ is the number of voters that prefer w to x.*

Proof In order to upper bound the social cost of candidate w we split her social cost in voters that belong to wx and xw. Remind that set wx consists of voters that prefer w to x.

$$\frac{SC(w,V)}{SC(x,V)} = \frac{\sum_{i\in wx} d(i,w) + \sum_{i\in xw} d(i,w)}{\sum_{i\in V} d(i,x)}$$

$$\leq \frac{\sum_{i\in wx} d(i,x) + \sum_{i\in xw} (d(i,x) + d(x,w))}{\sum_{i\in V} d(i,x)} = \frac{\sum_{i\in V} d(i,x) + \sum_{i\in xw} d(w,x)}{\sum_{i\in V} d(i,x)}$$

$$= 1 + \frac{|xw| \cdot d(w,x)}{\sum_{i\in V} d(i,x)} = 1 + \frac{(n - |wx|) \cdot d(w,x)}{\sum_{i\in V} d(i,x)}$$

Lastly, we have to lower bound the social cost of candidate x by $\sum_{i\in V} d(i,x) \geq \sum_{i\in wx} d(i,x)$ and we also know by the triangle inequality that for each voter $i \in wx$ it holds that $d(i,x) \geq \frac{1}{2} d(x,w)$.
Summing for all voters in wx we get $\sum_{i\in V} d(i,x) \geq \frac{1}{2} \sum_{i\in wx} d(x,w) = \frac{1}{2} \cdot |wx| \cdot d(x,w)$

All together

$$\frac{SC(w,V)}{SC(x,V)} \leq 1 + \frac{2(n - |wx|)}{|wx|} = \frac{2n}{|wx|} - 1$$

□

The above lemma is useful in order to derive upper bounds on distortion of voting rules if for the chosen candidate w and for every other candidate x we can lower bound the quantity $|wx|$. Thus, we get the following upper bounds for both Plurality and Borda rules.

Theorem 4.2. *For the voting rules $f = \{Plurality, Borda\}$ it holds that:* $dist(f) \leq 2m - 1$

Proof In both cases by the pigeonhole principle it holds that $|wx| \geq \frac{n}{m}$. Therefore by appplying lemma 4.2 we get

$$\frac{SC(w,d)}{SC(x,d)} \leq \frac{2n}{\frac{n}{m}} - 1 = 2m - 1$$

□

4.4.2 Sublinear Upper Bounds

Theorem 4.3. *The distortion of the Harmonic rule is asymptotically bounded by $\mathcal{O}(\frac{m}{\ln m})$, which almost matches the lower bound of distortion of positional scoring rules.*

Theorem 4.4. *The worst-case distortion of the STV rule is asymptotically bound by $\mathcal{O}(\ln m)$*

For the technical proof of the former bound we refer the reader to the original paper [45].

4.4.3 Constant Upper Bounds

Next, we review voting rules that achieve constant distortion. To derive constant bounds, the notion of tournament graphs is used [46]. These graphs are constructed by examining the pairwise relations between any two candidates $a, b \in C$. Each node corresponds to a candidate. We define the following graphs.

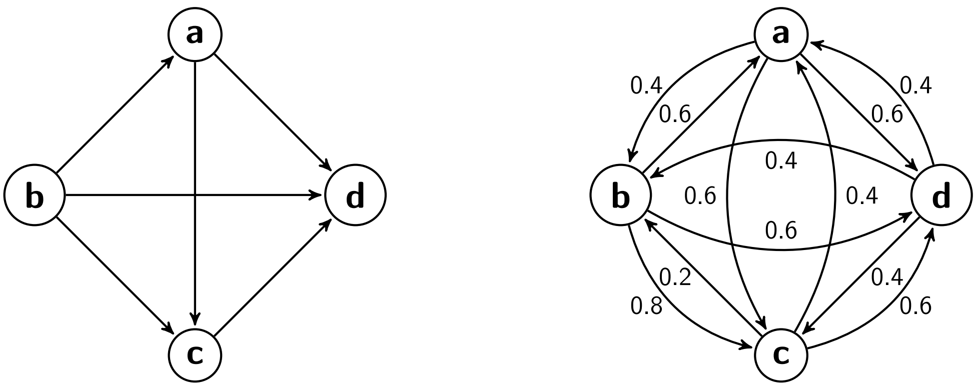
- The **tournament graph**: A directed edge from node a to node b is drawn if $|ab| \geq \frac{n}{2}$

- The **weighted tournament graph**: It can be considered as generalisation of the tournament graph. For each pair of nodes a, b two weighted directed edges are drawn. A directed edge from node a to node b is drawn with weight $\frac{|ab|}{n}$ and a directed edge from node b to node a is drawn with weight $1 - \frac{|ab|}{n}$

To illustrate these graphs consider the following example where 5 voters give their ordinal preferences on 4 candidates.

- σ_1: $a \succ b \succ c \succ d$
- σ_2: $a \succ c \succ b \succ d$
- σ_3: $d \succ b \succ c \succ a$
- σ_4: $b \succ a \succ c \succ d$
- σ_5: $d \succ b \succ c \succ a$

The incurred graphs are given in Figure 4.3



(a) Tournament graph (b) Weighted tournament graph

Figure 4.3: Tournament graphs

By observing the tournament graph, one can say that a node that has many outgoing edges seems as a rational choice as winner since he defeats many other candidates. In particular, if there is a source node in the tournament graph, it means that this candidate defeats every other candidate. The candidate that corresponds to a source node is called Condorcet winner and by Lemma 4.2 the distortion incurred by selecting such a candidate is at most 3, which is the best possible. However, it is not guaranteed that there will always exist such a candidate.

Definition 4.3. *The uncovered set is a set of candidates $a \in C$ such that for any candidate $b \neq a$ we have:*

- *either a weakly defeats b, i.e. $|ab| \geq \frac{n}{2}$*
- *or there is another candidate $c \notin \{a, b\}$ such that candidate a weakly defeats c and c weakly defeats b, i.e. $|ac| \geq \frac{n}{2}$ and $|cb| \geq \frac{n}{2}$*

Intuitively, this set of candidates constitutes a relaxation from the Condorcet winners. Every candidate w that belongs to the uncovered set seems as a compelling choice, considering that for every other candidate x either the majority of voters prefers w , or if it is not the case, there is another alternative y who is preferred by the majority of voters compared to x and yet the majority prefers w to y. The following lemma strengthens the importance of the uncovered set.

Lemma 4.3. *The uncovered set is always non-empty.*

Proof In order to verify this, consider the tournament graph of candidates of an arbitrary instance. A candidate x belongs to the uncovered set, based on the above definition, if and only if the corresponding node x in the tournament graph can reach every other node in one or two steps. Let x be the node of the tournament graph with the greatest out-degree. Suppose there is a node y that doesn't satisfy the above two conditions, that is, there is not an edge from x to y and there is not any other node w such that there is an edge from x to w and an edge from w to y. Therefore node y has greater out-degree from x which is a contradiction. □

Next, we will prove that the choice of a candidate that belongs to the uncovered set yields constant distortion.

Theorem 4.5. *Let w be a candidate that belongs to the uncovered set and x be the optimal candidate. Then $SC(w, d) \leq 5 \cdot SC(x, d)$*

In order to prove the distortion we will use the following lemmas.

Lemma 4.4. *Consider a vector $v \in \mathbb{R}^m$ where $v_1 \geq v_2 \geq ... \geq v_m$. If for all $k \in [m]$ it holds that $\sum_{i=1}^{k} a_i \geq \sum_{i=1}^{k} b_i$ then*

$$\sum_{i=1}^{m} v_i a_i \geq \sum_{i=1}^{m} v_i b_i$$

Lemma 4.5. *For a pair of candidates x, w if*

$$\sum_{i \in V} d(i, x) \geq \frac{1}{\gamma} \sum_{i \in xw} \min_{w \succ_i z} d(x, z)$$

for some $\gamma > 1$ then

$$SC(w, V) \leq (1 + \gamma) \cdot SC(x, V)$$

We will now prove the distortion of the upper bound on the distortion of picking a candidate from the uncovered set.

Proof We suppose that w is the chosen candidate and x is the optimal. If $|wx| \geq \frac{n}{2}$ from Lemma 4.2 we get immediately that $SC(w, d) \leq 3SC(x, d)$. Suppose this is not the case. Thus based on the definition of the uncovered set there exist a candidate y such that $|wy| \geq \frac{n}{2}$ and $|yx| \geq \frac{n}{2}$. We examine the cases where $d(x, y) \geq d(y, w)$ and $d(x, y) < d(y, w)$ separately.

- $d(x, y) \geq d(y, w)$
  In order to lower bound the social cost of candidate x we will use the fact that for each voter $i \in yx$ it holds $d(x, y) \leq d(i, x) + d(i, y) \leq 2d(i, x)$ and since we examine the case where $d(y, w) \leq d(x, y)$, it follows that $d(i, x) \geq \frac{1}{2} d(y, w)$. Thus

$$SC(x, d) = \sum_{i \in V} d(i, x) \geq \sum_{i \in yx} d(i, x) \geq \frac{1}{2} \sum_{i \in yx} d(x, w)$$

$$= \frac{1}{2} |yx| \cdot d(x, w) \geq \frac{n}{4} d(x, w) \geq \frac{1}{4} |xw| \cdot d(x, w)$$

$$\geq \frac{1}{4} \sum_{i \in xw} \min_{w \succeq_i z} d(x, z)$$

  And by applying Lemma 4.5 we get

$$\sum_{i \in V} d(i, w) \leq 5 \cdot \sum_{i \in V} d(i, w)$$

- $d(x, y) < d(y, w)$
  In this case we will use the sets wx, xwy, ywx to get to the desired distortion.

$$|wx| + |xwy| = |ywx| + |wyx| + |wxy| + |xwy|$$
$$\geq |wy| \geq |yw| \geq |yxw| + |xyw| \geq \frac{1}{2} \cdot |yxw| + \frac{1}{2} \cdot |xyw|$$

and

$$|wx| + |yxw| = |ywx| + |wyx| + |wxy| + |yxw|$$
$$\geq |yx| \geq \frac{n}{2} \geq \frac{1}{2} \cdot (|yxw| + |xyw| + |xwy|)$$

We are going to fix two vectors α, β and apply Lemma 4.4. We set:

$$\alpha_1 = |wx| + |xwy| \quad \alpha_2 = |yxw| \quad \alpha_3 = -|xwy|$$
$$\beta_1 = \frac{1}{2}|yxw| + \frac{1}{2}|xyw| \quad \beta_2 = \frac{1}{2}|xwy| \quad \beta_3 = 0$$
$$v_1 = d(x, w) \quad v_2 = d(x, y) \quad v_3 = d(x, y)$$

It holds that $\alpha_1 \geq \beta_1$ and by the last inequality we get $\alpha_1 + \alpha_2 + \alpha_3 \geq \beta_1 + \beta_2 + \beta_3$ and since $\alpha_2 \leq 0$ and $\beta_2 = 0$ we get $\alpha_1 + \alpha_2 \geq \beta_1 + \beta_2$. Thus we can apply Lemma 4.4 and we get

$$\sum_{i=1}^{3} \alpha_i v_i = (|wx| + |xwy|) \cdot d(x, w) + (|yxw| - |xyw|) \cdot d(x, y)$$
$$\geq \sum_{i=1}^{3} \beta_i v_i = \frac{1}{2}\left(|yxw| + |xyw|\right) \cdot d(x, w) + \frac{1}{2}|xwy| \cdot d(x, y)$$

Now we can derive the desired bound

$$\sum_{i \in V} d(i, x) \geq \sum_{i \in wx} d(i, x) + \sum_{i \in xwy} d(i, x) + \sum_{i \in yxw} d(i, x) \tag{4.1}$$
$$\geq \frac{1}{2}|wx| \cdot d(x, w) + |xwy| \cdot \left(\frac{d(x, w) - d(x, y)}{2}\right) + \frac{1}{2}|yxw| \cdot d(x, y) \tag{4.2}$$
$$= \frac{1}{2}d(x, w) \cdot (|wx| + |xwy|) + \frac{1}{2}d(x, y) \cdot (|yxw - |xwy|) \tag{4.3}$$
$$\geq \frac{1}{4} \cdot (|yxw| + |xyw|) \cdot d(x, w) + \frac{1}{4} \cdot |xwy| \cdot d(x, y) \tag{4.4}$$
$$\geq \frac{1}{4} \cdot \sum_{i \in xw} \min_{w \succeq_i z} d(x, z) \tag{4.5}$$

We will now explain why each inequality holds. The first inequality holds as $wx \cup xwy \cup yxw \subseteq V$. The third inequality holds since we apply Lemma 4.4. The last inequality holds because $xw = yxw \cup xyw \cup xwy$, $w \succeq_i w \ \forall i \in V$ and $w \succeq_i y \ \forall i \in xwy$.
Finally, we can apply Lemma 4.5 for $\gamma = 4$ and thus we get

$$\sum_{i \in V} \leq 5 \cdot \sum_{i \in V} d(i, x)$$

Therefore, in each case we prove that by selecting a candidate from the uncovered set, the distortion is at most 5.

□

Definition 4.4 (Copeland Rule [47]). *Copeland rule outputs the set of candidates with the maximum Copeland score. The Copeland score for candidate w is defined as the number of candidates x that defeats*
(i.e. $|x \in C : |wx| \geq \frac{n}{2}|$).

Lemma 4.6. *The alternatives that Copeland rule outputs is a subset of the uncovered set.*

Proof Let w be the candidate that Copeland rule outputs and let $\{c_1, c_2, .., c_k\}$ be the candidates that w weakly defeats. Suppose there is a candidate x that none of the $\{w, c_1, c_2, .., c_k\}$ defeats him. Therefore he defeats all of them which is a contradiction because he would have highest Copeland score than w

□

Subsequently, Copeland rule yields distortion of at most 5.

4.5 Optimal Metric Distortion

In this section we further review mechanisms that achieve improved constant distortion. First we examine the work of Munagala et al. [48]. They made two generalizations in the context of uncovered sets which was defined in 4.4.3
They define the λ weighted uncovered sets.

Definition 4.5. *Let $\lambda \in [0.5, 1]$ be a constant. The λ weighted uncovered set is the set of candidates $a \in C$ such that for any candidate $b \neq a$ we have:*

- *either $|ab| \geq (1 - \lambda)n$*
- *or there is another candidate $c \notin \{a, b\}$ such that $|ac| \geq (1 - \lambda)n$ and $|cb| \geq \lambda n$*

It is easy to see that for $\lambda = 0.5$ we get the uncovered set.
They proved that that λ weighted uncovered sets are always non-empty. More importantly they proved that by setting $\lambda = \frac{\sqrt{5}-1}{2} \approx 0.618$ the candidates that belong to this weighted uncovered set have worst case distortion of $\sqrt{5} + 2 \approx 4.236$.
The proof of their distortion analysis is based on some modifications on the proof of the distortion of uncovered set.

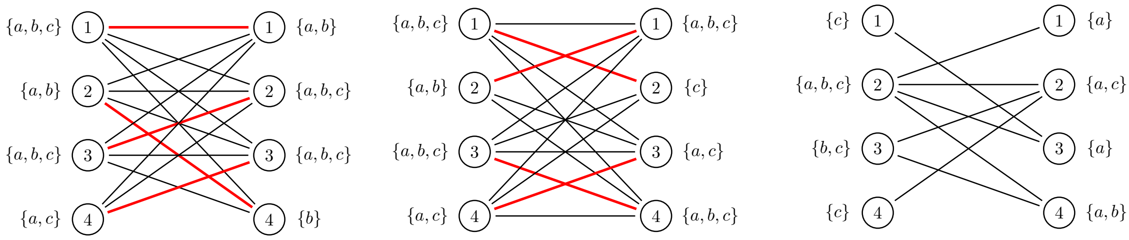
The second critical point of their work was the definition of matching uncovered set.

Definition 4.6. *The matching uncovered set is a set of candidates $a \in C$ such that for every candidate $b \in C \setminus \{a\}$ there exists a perfect matching in the bipartite graph*
$G(a, b) = (V, V, E_{a,b})$, where $(i, j) \in E_{a,b}$ if there exists a candidate $x \in C$ such that voter i weakly prefers a to x and voter j weakly prefers candidate x to b.

Remind that for a voter i and each candidate c, it holds that voter i weakly prefers candidate c to c.

Consider the following ordinal preference profile of 4 voters over 3 candidates.

- σ_1: $a \succ b \succ c$
- σ_2: $c \succ a \succ b$
- σ_3: $a \succ c \succ b$
- σ_4: $b \succ a \succ c$



(a) Graph $G(a,b)$ (b) Graph $G(a,c)$ (c) Graph $G(c,a)$

Figure 4.4: Graphs defined in 4.6. There exists an edge (i,j) if the intersection of the respective subsets of candidates is non empty.

Based on the above ordinal preferences graphs $G(a,b), G(a,c)$ and $G(c,a)$ have been constructed. Candidate a belongs to the matching uncovered set, since graphs $G(a,b)$ and $G(a,c)$ admit a perfect matching. However candidate c does not belong to the matching uncovered set since graph $G(c,a)$ does not admit a perfect matching.

Theorem 4.6. *If there is a candidate w that belongs to the matching uncovered set, the distortion of this candidate is at most 3.*

The worst-case distortion of picking a candidate from the matching uncovered set is better than choosing one from the uncovered set. In a high level, this can be justified because the relation a candidate w, that belongs to the matching uncovered set, has with the rest of the candidates is stronger than the respective relation when w belongs to uncovered set. More specifically, in the latter case the condition for those candidates must apply for the majority of voters, whereas in the former case the condition must apply for every pair of voters. Thus, the matching uncovered set is a subset of the uncovered set.

In order to see this difference, consider the ordinal preference profile over three candidates x, y, z, where :

- $\frac{n}{2} - 1$ voters rank them as $y \succ x \succ w$
- $\frac{n}{2} - 1$ voters rank them as $x \succ w \succ y$
- The remaining 2 voters rank them as $w \succ y \succ x$

In this case, candidate w defeats y, y defeats x and x defeats w and thus all of them belong to the uncovered set. However, only candidate y belong to the matching uncovered set.

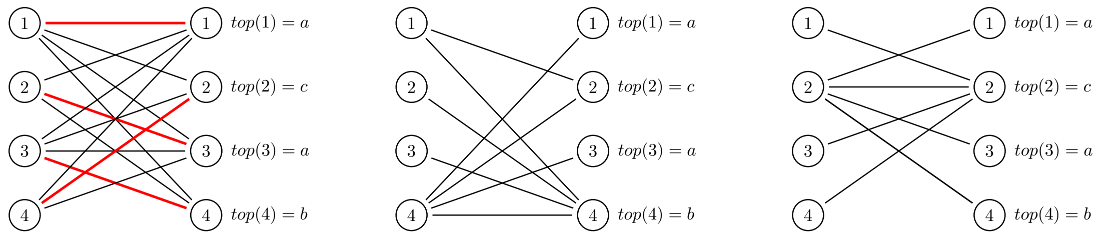
Gatzelis et al. [49] devise their work by building on this work. They define the integral domination graph $G(a)$ for every candidate $a \in C$, which is directly correlated with the bipartite graph $G(a,b)$ defined above.

Definition 4.7. *The integral domination graph of candidate $A \in C$ is a bipartite graph $G_a = (V, V, E_a)$ where $(i,j) \in E_a$ if voter i weakly prefers candidate a to the top choice of voter j, i.e. $a \succ_i top(j)$.*

In order to see how these graphs are constructed consider again the ordinal preference profile given above, when we define the matching uncovered set.

- σ_1: $a \succ b \succ c$
- σ_2: $c \succ a \succ b$
- σ_3: $a \succ c \succ b$

- σ_4: $b \succ a \succ c$



(a) $G(a)$ (b) $G(b)$ (c) $G(c)$

Figure 4.5: Integral domination graph of each candidate. $G(a)$ admits a perfect matching.

They proposed a rule called *PLURALITYMATCHING* that selects the candidate whose integral domination graph admits a perfect matching. The importance of this rule is twofold; there is always such a candidate and the distortion of selecting this candidate is at most 3, which is the lowest possible. The heavy lifting of this work was to prove the existence of such a candidate. Below we only give the proof of the distortion of this rule.

Theorem 4.7. *Selecting a candidate whose integral domination graph admits a perfect matching yields distortion at most 3 in general metric spaces*

Proof Suppose that candidate w is a candidate whose integral domination graph admits a perfect matching and x is the optimal candidate. We will upper bound the social cost of candidate w as follows:

$$\begin{aligned}SC(w,V) &= \sum_{i\in V} d(i,w) \leq \sum_{i\in V} d(i,top(M(i))) \leq \sum_{i\in V}\{d(i,x)+d(x,top(M(i))\}\\ &= SC(x,V)+\sum_{i\in V} d(x,top(M(i)) = SC(x,V)+\sum_{i\in V} d(x,top(i))\\ &\leq SC(x,V)+\sum_{i\in V}\{d(z,i)+d(i,top(i))\} = 2SC(x,V)+\sum_{i\in V} d(i,top(i))\\ &\leq 2SC(x,V)+\sum_{i\in V} d(i,x) = 3\cdot SC(x,V)\end{aligned}$$

The first inequality holds due to the structure of integral domination graph. The second due to triangle inequality. Moreover it holds that $\sum_{i\in V} d(x,top(M(i)) = \sum_{i\in N} d\,(x,top(i))$ because M is a perfect matching. And the last inequality holds as candidate $top(i)$ is the closest candidate to voter i and hence $d\,(x,top(i)) \leq d(i,x)$. □

Intuitively, by selecting such a candidate w we can say that for each voter j who has another candidate ranked first denoted as $top(j)$, there is at least one voter i who prefers candidate a to candidate $top(j)$. Therefore the selected candidate a is placed approximately at the middle of the voters in the metric space.

Therefore, as far as deterministic rules are concerned, given the ordinal preferences, this rule resolves the single winner problem in a metric space. At this point, we think it is interesting to introduce a generalisation of metric spaces that has been proposed by Anshelevich et al. [50].

Definition 4.8 (α-decisive metric space)**.** *We say a metric space is* a-decisive *if for each voter i whose top choice is candidate $top(i)$, it holds that $d(i, top(i)) \leq \alpha \cdot d(i, x)$, for every other candidate x and $\alpha \in [0, 1]$*

The notion of a-decisiveness captures how strong the voter's preference for her top choice is compared to other candidates. General metric spaces are $1 - decisive$. The metric space is $0 - decisive$ if each voter is also a candidate which occurs in settings such as a committee where the members have to choose one of them to represent them. In particular, though, a-decisiveness is important in our analysis as many of the worst case examples occur when many voters are indifferent between their top candidate and the optimal one. The proof of the lower bound 4.2 relies on this statement. For general a-decisive metric spaces, the lower bound of distortion that has been established is $2 + \alpha - 2(1 - \alpha)/\lfloor m \rfloor_{even}$. [49]. Note that this bound approaches $2 + \alpha$ as the number of candidates m increases. The *PLURALITYRULE* achieves tight distortion as well in general a-decisive metric spaces as $m \to \infty$. This is true if we replace the last inequality of the proof as $d(x, top(i)) \leq \alpha \cdot d(i, x)$ and subsequently this results in distortion of $2 + \alpha$.

4.6 Optimal Metric Distortion Given Candidates Position

In this section we examine a slightly different framework proposed by Anshelevich and Zhu [51].
The voters and candidates are located in a metric space. The ordinal preferences of voters are given and we assume that the exact locations of candidates in the metric space are known. With this additional assumption they propose a simple mechanism that achieve distortion at most 3. Notice that the lower bound of distortion remains at 3 even when the locations of candidates are given.
The mechanism consists of two steps:

- For every voter i generate a projected voter $\tilde{i}$ in the location of his favorite candidate
- Solve the problem optimally in the projected instance $(\tilde{V}, C, d)$, since all the distances are known and output the winner

Theorem 4.8. *The above voting rule achieves distortion at most 3 for the total social cost*

Proof Let w be the candidate selected by this voting rule and x be the optimal candidate that minimizes the total social cost. We get

$$\frac{SC(w, \tilde{V})}{SC(x, \tilde{V})} = \frac{\sum_{i \in \tilde{V}} d(i, x)}{\sum_{i \in \tilde{V}} d(i, x)} \leq \frac{\sum_{i \in V} d(\tilde{i}, w)}{\sum_{i \in V} d(\tilde{i}, x)} \leq 1$$

For the original instance we have:

$$\frac{SC(w, V)}{SC(x, V)} = \frac{\sum_{i \in V} d(i, x)}{\sum_{i \in V} d(i, x)} \leq \frac{\sum_{i \in V} d(i, \tilde{i}) + \sum_{i \in V} d(\tilde{i}, w)}{\sum_{i \in V} d(i, x)} = \frac{\sum_{i \in V} d(i, \tilde{i})}{\sum_{i \in V} d(i, x)} + \frac{\sum_{i \in V} d(\tilde{i}, w)}{\sum_{i \in V} d(i, x)}$$

The latter inequality results from triangle inequality. In order to complete the proof two observations have to be made.

$$\frac{\sum_{i \in V} d(i, \tilde{i})}{\sum_{i \in V} d(i, x)} \leq 1$$

and

$$\frac{\sum_{i \in V} d(\tilde{i}, w)}{\sum_{i \in V} d(i, x)} \leq \frac{\sum_{i \in V} d(\tilde{i}, w)}{\sum_{i \in V} \frac{d(\tilde{i}, x)}{2}} = 2 \cdot \frac{\sum_{i \in V} d(\tilde{i}, w)}{\sum_{i \in V} d(\tilde{i}, x)} \leq 3$$

The first holds because $\tilde{i}$ is the top choice of i. For the second observation we use lemma . Therefore combining these inequalities we get

$$\frac{SC(w,V)}{SC(x,V)} \leq 3$$

□

Less formally, we can say that candidates that are ranked first by some voter act as proxies and every proxy has strength equal to number of voters that ranked her first.
It is worth noting that in order to prove the upper bound of the distortion of this mechanism only the top choice of each voter was used rather than the whole ordinal preference profile. Hence, given the distances between candidates and the top choice of each voter is enough to make an accurate approximation of how the voters and candidates are distributed in the metric space.

Chapter 5

Multiwinner voting in a metric space

In this chapter we examine the multiwinner voting(k-winner) problem in which agents lie in a metric space. In this case, in order to capture again the notion of proportional representation we assume that the voters have unit-demand utilities, which means that their utility for a set of candidates derives only from his favorite candidate among them, i.e. the closest one.

Formally, the distance a voter i has from a set X of candidates is defined as:

$$d(i, X) = \min_{c \in X} d(i, c)$$

Consequently, the total social cost of a subset X of candidates for the set of voters V in a metric d is computed as:

$$SC(X, V|d) = \sum_{i \in V} \min_{c \in X} d(i, c)$$

In other words, a multiwinner voting rule is seeking to select k candidates such that the clusters of voters incurred, are the most compact. Specifically, if the distances were known, the problem would translate in solving an instance of the *metric k-median problem*, which is the problem of finding k centers such that the clusters formed by them are the most compact. The *metric k-median problem* can be viewed as a special case of a wider family of problems, the metric facility location. Consider we have a set of agents V and a set of facilities F distributed in a metric space. The problem is to decide which facilities should serve and assign each voter to the closest open facility so as to minimize a certain objective function. Facility assignment problems may have different constraints. Two such common constraints are

- Each facility F_i has a capacity cap_i, which is the maximum number of agents that can be assigned to it.
- Each facility F_i has a facility cost c_i, which is the opening cost of a the facility.

The k-median problem differs from the facility location problem in two aspects - there is no cost for opening facilities and there is an upper bound, k, on the number of facilities that can be opened. We are interested in the case where the number of facilities/candidates chosen is exactly k.
The *k-median problem* (so as facility location) belong to NP-hard problems. There exist a number of approximation algorithms in order to solve the problem in polynomial time. The first approximation algorithm for the metric k-median problem was given in the work of Bartal [52],[53], in which he shows how to approximate any finite metric space by a tree space. Moreover Tardos et al. [54], designed the first constant factor approximation algorithm for the metric k-median problem by giving an algorithm that finds a solution of value at most $6\frac{2}{3}$ times larger than the optimal. Many of the algorithms proposed are based on linear

programming. For sake of completeness we give the integer problems of incapacitate facility location and metric k-median.

$$
\begin{array}{lll}
\text{minimize} & \sum_{i \in V, j \in F} d_{ij} x_{ij} + \sum_{j \in F} f_j y_j & \\
\text{subject to} & \sum_{i \in V, j \in F} x_{ij} \geq 1, & \forall i \in V \\
 & y_j - x_{ij} \geq 0, & i \in V, j \in F \\
 & y_j \in \{0, 1\}, & j \in F \\
 & x_{ij} \in \{0, 1\}, & i \in V, j \in F
\end{array}
$$

Figure 5.1: Incapacitate facility location Linear Problem

The above program is an integer program for incapacitate facility location, i.e. the problem in which each facility has unbounded capacity. In this program, y_j is an indicator variable and denotes if facility j is open and f_j is the cost of opening facility j. x_{ij} is an indicator variable denoting whether agent i is assigned to facility j and d_{ij} is their respective distance.

$$
\begin{array}{lll}
\text{minimize} & \sum_{i \in V, j \in F} d_{ij} x_{ij} & \\
\text{subject to} & \sum_{j \in F} x_{ij} \geq 1, & \forall i \in V \\
 & y_j - x_{ij} \geq 0, & i \in V, j \in C \\
 & \sum_{j \in C} y_j = k & \\
 & y_j \in \{0, 1\}, & j \in C \\
 & x_{ij} \in \{0, 1\}, & i \in V, j \in C
\end{array}
$$

Figure 5.2: Metric k-Median Linear Problem

As mentioned, the k-median problem can be interpreted as an instance of the k-winner problem. Thus, in the above program y_j is an indicator variable denoting whether candidate j is chosen. x_{ij} is an indicator variable denoting whether voter i is represented by candidate j and d_{ij} is their respective distance.

In the framework of distortion in k-winner setting, we will often use one of the known approximation algorithms as a black box, in order to extract the value of distortion. As a warm up, first, we examine the distortion when voters provide their ordinal preferences.

5.1 Lower Bounds of Distortion Given the Ordinal Preferences

In the previous sections we study the case where the voters give their ordinal preferences regarding the single voting problem. Firstly, we assume again that the only information given are the voters' ordinal preferences. We derive the following lower bounds.

Theorem 5.1. *For the metric k-winner problem given the ordinal preferences the following lower bounds hold:*

- *If $k = 2$, for any voting rule f the distortion is $\Omega(n)$*
- *If $k > 2$, for any voting rule f the distortion is unbounded*

Proof

- For $k = 2$. Suppose there are three groups of voters v_1, v_2, v_3, i.e $V = v_1 \cup v_2 \cup v_3$ and three candidates x, y, z and the number of winners is $k = 3$. The ordinal preferences given to the voting rule are:

  · $v_1 : x \succ y \succ z$

  · $v_2 : y \succ x \succ z$

  · $v_3 : z \succ y \succ x$

  Consider the instance where each group of voters is collocated with a different candidate, i.e. $d(v_1, x) = d(v_2, y) = d(v_3, z) = 0$. We fix the number of voters in each group as $|v_1| = |v_2| = n/2$ and $|v_3| = \mathcal{O}(1)$. If the voting rule did not select candidate z then we could fix the distances such that $d(v_1, v_2) << d(v_2, v_3)$ and thus the distortion would be unbounded. Therefore, suppose without loss of generality, that the given voting rule picks the set
  $W = \{y, z\}$. We fix the distances such that $d(v_1, v_2) = d(v_2, v_3) = l$. Hence, the optimal solution is $X = \{x, y\}$. The distortion incurred in this case is

$$dist(f, \sigma) = \frac{SC(W, V)}{SC(X, V)} = \frac{l \cdot n/2}{l} = \Omega(n)$$

- For $k > 2$ Suppose there are four groups of voters v_1, v_2, v_3, v_4, i.e. $V = v_1 \cup v_2 \cup v_3 \cup v_4$ and four candidates x, y, z, w. The ordinal preferences given to the voting rule are:

  · $v_1 : x \succ y \succ z \succ w$

  · $v_2 : y \succ x \succ z \succ w$

  · $v_3 : z \succ w \succ y \succ x$

  · $v_4 : w \succ z \succ y \succ x$

  Consider again that each group of voters is collocated with a different candidate, i.e. $d(v_1, x) = d(v_2, y) = d(v_3, z) = d(v_4, w) = 0$. We fix the number of voters in each group as $|v_1| = |v_2| = |v_3| = |v_4| = n/4$. Without loss of generality, suppose the given voting rule picks the subset $W = \{x, y, z\}$. In this case, we can fix an instance where $d(v_1, v_2) << d(v_3, v_4)$ and hence the optimal subset is $X = \{x, z, w\}$. Therefore the distortion will be unbounded.

□

Based on the above lower bounds, we deduce that the voters' ordinal preferences are not enough in order to achieve bounded distortion. We need a certain amount of cardinal information. Intuitively, given this information, none of the voting rules can form the clusters correctly and subsequently there will always be instances in which voters will pay a distance arbitrarily larger than the one in the optimal assignment. This is not the case in the metric single winner setting since there is only one cluster and thus picking the top candidate of a random voter i yields distortion of $\mathcal{O}(n)$. This can be generalized in multiwinner setting. If the voting rule forms the clusters of voters correctly but fails to choose the optimal candidate for each cluster the incurred distortion is $\mathcal{O}(n)$.

5.2 Ordinal Preferences given the distances between facilities

With that being said, a multiwinner voting rule f needs cardinal information in order to achieve bounded approximation. Before we present our results, we review again the framework proposed by Anshelevich et al. [51]. In this framework the distances between each pair of candidates are known and voters give their ordinal rankings. As mentioned in the respective section of the single winner setting, providing this information, a mechanism can approximate how the voters and candidates are distributed in the metric space. Thus the same algorithm yields bounded distortion in the multiwinner setting as well. In order to solve the *k-winner* problem given all the distances between candidates, an approximation algorithm, with approximation ratio β is used as a black box. The mechanism consists of two steps:

- For every voter i generate a projected voter $\tilde{i}$ in the location of his favorite candidate
- Solve the problem in the projected instance $(\tilde{V}, C, d)$ using an approximation algorithm with approximation ratio β and output this assignment.

Theorem 5.2. *The above mechanism yields distortion of at most* $2\beta + 1$

Proof We consider as usual W the subset of candidates that the mechanism outputs and X the optimal subset of candidates. We want to prove that $SC(W, V|d) \leq (2\beta+1) \cdot SC(W, X|d)$. In order to prove that we define the following quantities.
$s_i = \min_{c \in W} d(i, c)$, $\quad t_i = d(i, top(i))$, $\quad b_i = d(top(i), \arg\min_{c \in W} d(i, c))$
$s_i^* = \min_{c \in X} d(i, c)$, $\quad b_i^* = d(top(i), \arg\min_{c \in X} d(i, c))$
By the triangle inequality it holds

$$SC(W, V|d) = \sum_{i \in V} s_i \leq \sum_{i \in V} (b_i + t_i) = \sum_{i \in V} b_i + \sum_{i \in V} t_i$$

Since the total social cost for subset W is a β approximation for the projected instance it follows that

$$SC(W, \tilde{V}|d) = \sum_{i \in V} b_i \leq \beta \cdot SC(X, \tilde{V}|d) = \beta \cdot \sum_{i \in V} b_i^* \leq 2\beta \sum_{i \in V} s_i^*$$

The last inequality holds because for each voter i we have $t_i \leq s_i^*$ and by the triangle inequality we get $b_i^* \leq t_i + s_i^* \leq 2s_i^*$.
Combining the last two inequalities with the above bounds we derive

$$SC(W, V|d) \leq \sum_{i \in V} b_i + \sum_{i \in V} t_i \leq 2\beta \sum_{i \in V} s_i^* + \sum_{i \in V} t_i \tag{5.1}$$

$$\leq 2\beta \sum_{i \in V} s_i^* + \sum_{i \in V} s_i^* = (2\beta + 1) \cdot SC(W, X|d) \tag{5.2}$$

□

Thus, if we are not interested in achieving polynomial running time, by setting $\beta = 1$ we get distortion of at most 3. This bound is tight.

5.3 Bucket Preferences

In this section we present our results. We propose a new framework where each voter is asked to give an approximation of his distance from his t most preferred candidates. The

idea is to use multiple threshold approvals, as explained in Chapter 3. We highlight that this query model, apart from eliciting information from voters more efficiently, provides a more conceivable way of querying the voters about their preferences rather than asking for their ordinal preferences. It is also important to note that as the distance between a voter i and candidate c increases, the respective approximation that we ask them to submit is relaxed and thus the required cognitive effort is reduced. By eliciting information from each voter in this way, we get a partial ranking of candidates combined with an approximation of their distances. We will define the above approximations and examine how the number t of candidates submitted influence the distortion. We assume that the minimum distance d_{min} between two candidates is known.
We define the bucket x as:

$$b^x = \begin{cases} [d_{x-1}, d_x) & \text{where } d_x = \gamma \cdot d_{x-1} & , x \geq 1 \\ [0, d_0) & \text{where } d_0 = \frac{d_{min}-\epsilon}{2} & , x = 0 \end{cases} \Bigg\}$$

For a voter i and candidate c we denote that $c \in b_i^x$ if $d_{x-1} \leq d(i,c) < d_x$.
The size of the buckets increases exponentially with γ, which is a constant. The voters are asked to submit in which bucket the distance from each of their top t candidates belong.

It is mandatory to know the minimum distance between two candidates in order to define the smallest bucket d_0. By defining the first bucket as above, we ensure that a voter may submit one candidate at most in this bucket.

For each voter i that assigns candidate x in bucket $b_k = (d_{k-1}, d_k)$, we define the extended distance
$d_{ext}(i,x) = d_k$, i.e. equal to the upper bound of the respective bucket.

5.4 Number of queries - Distortion Tradeoff

In this section, we will examine how the number t of candidates submitted by each voter influences distortion.
We consider the case where $t = m$. That is, ask each voter to submit in which bucket his distance from every candidate belongs. Thus, the voting rule elicits all the extended distances between voters and candidates.

Theorem 5.3. *If the number of queries per voter is $t = m$, there is a multiwinner voting rule that achieves distortion at most $\beta \cdot \gamma$.*

Algorithm 3: Multiwinner rule given full bucket profile

Procedure *k-Median Solver(V,C,d)*:
  return A subset W of size k using an approximation algorithm
end
Multiwinner Rule
  Input V, C and bucket profile $\boldsymbol{b} = \{b_1, \ldots, b_n\}$
  For each voter i and candidate c such that $c \in b_i^x = (d_{x-1}, d_x)$
    Define $d_{ext}(i,c) = d_x$
  return *k-Median Solver(V,C,d_{ext})*
end

Proof For a voter i and a candidate c it holds that $d(i,c) \leq d_{ext}(i,c) \leq \gamma \cdot d(i,c)$. Suppose that the voting rule for the selection of the subset of k candidates uses an approximation algorithm A that solves the k-median problem using the distances d_{ext} and achieves approximation ratio

β. Define W the set of candidates that this algorithm outputs, Y the optimal set given the distances d_{ext} and X the optimal set of candidates for the original problem. Thus it holds:

$$SC(W,V|d) \leq SC(W,V,|d_{ext}) \leq \beta \cdot SC(Y,V|d_{ext}) \leq \beta \cdot SC(X,V|d_{ext}) \leq \beta \cdot \gamma \cdot SC(X,V|d)$$

□

Theorem 5.4. *If the number of queries per voter is $t > \frac{m}{2}$, there is a multiwinner voting rule that achieves constant distortion for the total social cost.*

Algorithm 4:

Procedure *k-Median Solver(V,C,d)***:**
  return A subset W of size k using an approximation algorithm
end
Multiwinner Rule
  Input V, C, $\boldsymbol{b} = \{b_1, \ldots, b_n\}$
  Ask each voter i in which bucket b_i belong the distance $d(i,c)$ for each of his top t preferred candidates c.
  For each voter i and candidate c define the extended distance
  if $c \in b_i^x = (d_{x-1}, d_x)$ **then**
    $d_{ext}(i,c) = d_x$
  else
    $d_{ext}(i,c) = \min_{\substack{j \in V \\ b \in b_i \cap b_j}} \{d_{ext}(i,b) + d_{ext}(j,b) + d_{ext}(j,c)\}$
  end
  return *k-Median Solver(V,C,d_{ext})*
end

For each voter i and candidate c define $d_{ext}(i,c)$ as:

$$d_{ext}(i,c) = \left\{ \begin{array}{ll} d_i & \text{if } c \in b_i \\ \min_{\substack{j \in V \\ b \in b_i \cap b_j}} \left(d_{ext}(i,b) + d_{ext}(j,c) + d_{ext}(j,b)\right) & \text{if } c \notin b_i \end{array} \right\}$$

Proof In order to use the *k-Median Solver* we need to define the distances between each voter i and candidate c. Since each voter submits only his top t candidates we define the remaining distances as written above. For a voter i and a candidate c that does not belong to his top t most preferred candidates, we define their distances through another voter j that has submitted at least one common candidate with voter i. The above definition of $d_{ext}(i,c)$ for a voter i and candidate c that does not belong to the submitted candidates is an overestimation of the real distance and it is derived by application the triangle inequality twice.
i.e., $d(i,c) \leq d(i,b) + d(b,c) \leq d(i,b) + d(j,b) + d(j,c)$
A single such estimation may be arbitrarily bigger than the real distance, as the distance $d(b,c)$ can be arbitrarily smaller that distances $d(j,b)$ and $d(j,c)$. However, if we compare pairwise the social cost of candidate c incurred by this estimation with the original one, then we get a constant approximation. Consider two voters i,j and a candidate c. The incurred social cost is:

- $SC_{ext}(c,i,j) = d_{ext}(i,c) + d_{ext}(j,c) \leq \gamma SC(c,i,j)$, if $c \in b_i \cap b_j$
- $SC_{ext}(c,i,j) = d_{ext}(i,b) + d_{ext}(j,b) + 2d_{ext}(j,c) \leq 3\gamma\left(d(i,c) + d(j,c)\right) = 3\gamma SC(c,i,j)$, if $c \in b_j \setminus b_i$ and $b \in b_i \cap b_j$

In order to see why the last bullet holds, consider the edge case in Figure 5.3 where there are two voters i, j and two candidates b, c. Specifically, $b \in b_i^x = (d_{x-1}, d_x)$, $c \notin b_i$ and $b, c \in b_j^y = (d_{y-1}, d_y)$. We fix the distances as follow. $d(b,c) = \epsilon \to 0$, $d(i,b) = d_{x-1}$ and $d(j,b) = d(j,c) = d_{y-1}$. Hence the extended distances that the mechanism can estimate are $d_{ext}(i,b) = d_x$ and
$d_{ext}(j,b) = d_{ext}(j,c) = d_y$. Following the definition of $SC_{ext}(c,i,j)$ for candidate j we get: $SC_{ext}(c,i,j) = d_{ext}(i,b) + d_{ext}(j,b) + 2d_{ext}(j,c) = d_x + 3d_y$. Thus

$$\frac{SC_{ext}(c,i,j)}{SC(c,i,j)} = \frac{d_x + 3d_y}{d_{x-1} + d_{y-1}} = \frac{\gamma \cdot (d_{x-1} + 3d_{y-1})}{d_{x-1} + d_{y-1}} \leq 3 \cdot \gamma$$

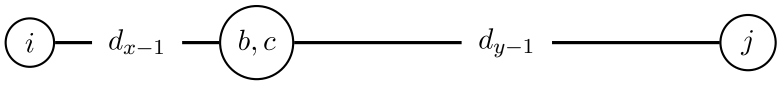


Figure 5.3: Voter i submits candidate b and voter j submits candidates both b, c. $d(b,c) \to 0$

Lemma 5.1. *Going a step further, we can estimate the social cost of a candidate c for voter v_0 that submits c and $i = \{1,2,3\}$ voters, namely v_1, v_2, v_3 that do not submit c.*

- $i = 1$: $SC_{ext}(c, v_0, v_1) \leq 3\gamma SC(c, v_0, v_1)$
- $i = 2$: $SC_{ext}(c, v_0, v_1, v_2) \leq 5\gamma SC(c, v_0, v_1, v_2)$
- $i = 3$: $SC_{ext}(c, v_0, v_1, v_2, v_3) \leq 7\gamma SC(c, v_0, v_1, v_2, v_3)$

Let X be the optimal subset of winners and cl^* the optimal clusters of voters incurred by this selection. We want to prove that given the distances d_{ext}, as defined above, we can estimate the social cost of each optimal cluster by searching exhaustively all the possible solutions.

Consider an optimal cluster $cl \subset V$, where $n' = |cl|$ and x the respective optimal candidate assigned to this cluster. For a candidate $w \in C$ we define:

$$B_w = \frac{\sum_{i \in cl} \mathbb{1}\{w \in b_i\}}{n'}$$

which is the percentage of voters in the cluster cl that submitted candidate w in their buckets.

Remind that we ask voters to submit the t-most preferred voters, thus it holds that $\sum_{c \in C} B_c = t$. By the pigeonhole principle there is a candidate w such that $B_w \geq \frac{t}{m}$. Since $t > m/2$ it follows that $B_w > \frac{1}{2}$, which means that there is at least one candidate that is submitted by more than half of the voters in cluster cl.

We consider the following cases.

- If for the optimal candidate x in cluster cl it holds that $B_x \geq \frac{1}{2}$, then we can estimate the social cost of candidate x, by a constant factor of at most $3 \cdot \gamma$ by Lemma 5.1
- In the contrary case, assume that for the optimal candidate x it holds that $B_x < \frac{1}{2}$.
  If $t \geq \frac{3m}{4}$, since $B_w \geq \frac{t}{m} > \frac{3}{4}$, we can infer that $|wx| \geq \frac{n'}{4}$. Hence by Lemma 4.2 we get $SC(w, cl) \leq 7 \cdot SC(x, cl)$.
  Now, suppose that $\frac{m}{2} < t < \frac{3m}{4}$. If $B_x < \frac{1}{4}$ then we can infer again that $|wx| \geq \frac{n'}{4}$ and by Lemma 4.2 we get $SC(w, cl) \leq 7 \cdot SC(x, cl)$.
  Lastly, if $\frac{1}{4} \leq B_x < \frac{1}{2}$, it follows that at least $\frac{n'}{4}$ voters of the cluster have submitted candidate x and hence we can estimate this social cost by a constant of 7γ by Lemma 5.1

We prove that if each voter provide the buckets in which his $t > \frac{m}{2}$ most preferred candidates belong, there is a voting rule that can produce a solution with constant distortion δ Thus if the voting rule uses an approximation algorithm with constant approximation ratio β given the distances d_{ext} and return the output of the approximation algorithm, the total distortion is at most $\beta \cdot \delta$

□

Up until now, we examined cases where the number of queries per voter was greater than $\frac{m}{2}$. Hence every pair of voters submitted at least one common candidate and that help us establish an upper bound for each distance between voter i and candidate c.

Theorem 5.5. *If the number of queries per voter is $t = \frac{m}{2}$ the distortion of every multiwinner voting rule is $\Omega(n)$*

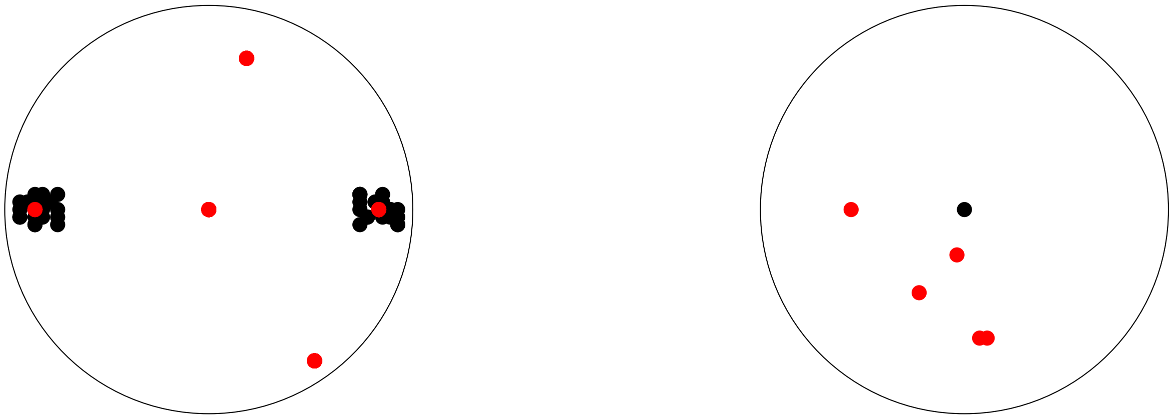


Figure 5.4: Two groups of voters are formed. The black dots correspond to the voters and the red to candidates.

Proof Consider the case illustrated in Figure 5.4. In this instance $n-1$ voters belong to group G_1 and the remaining voter i form another group, G_2. Each group of voters submit the same $\frac{m}{2}$ candidates. As it is shown, inside group G_1 two clusters are formed. The diameter of group G_1 is $d(G_1) = 2x$. The optimal solution is to choose two candidates $X = \{c_1, c_2\}$ from group G_1 and assign voter i to c_2, the closest one. Let $d(i, c_2) = 2x$. Thus the social cost incurred is $SC(X, V, d) = 2x$. A voting rule, given only the results of the queries must pick one candidate from each group. Let $W = \{w_1, w_2\}$ be the chosen candidates. Suppose w_1 is the candidate that minimizes the social cost of group G_1 yielding a cost of $(n-1) \cdot x$ and let w_2 be the top candidate of voter i with distance $d(i, w_2) = 2x - \epsilon$. Thus the total social cost is $SC(W, V, d) = (n-1) \cdot x + 2x - \epsilon$. The distortion in this case is

$$dist(f) = \frac{SC(W,V,d)}{SC(X,V,d)} = \frac{(n-1) \cdot x + 2x - \epsilon}{2x} = n - 1$$

The last lower bound indicates that in order to achieve constant distortion in the multiwinner setting a mechanism must elicit implicitly or not an approximation of the distance between each pair of candidates. This is the reason why the mechanism that we examined in Section 5.2 achieves constant distortion, by knowing apriori the distances between any two candidates.

□

Before we prove the distortion for the other values of t, we give the following useful lemma

Lemma 5.2. *Consider an instance, the optimal subset of candidates X and the optimal clusters cl^* induced by this selection. For each optimal cluster cl_i^*, if we replace the respective* *optimal candidate x_i with a candidate w, that is the top choice for some voter in the cluster cl_i^*, the total social cost incurred is an $\mathcal{O}(n)$ approximation.*

Proof This lemma is essentially a generalization of Lemma 4.2. Since $|wx_i| > 0$, by Lemma 4.2 it holds that $\frac{SC(w,cl_i^*)}{SC(x_i,cl_i^*)} \leq 2n - 1 = \mathcal{O}(n)$ □

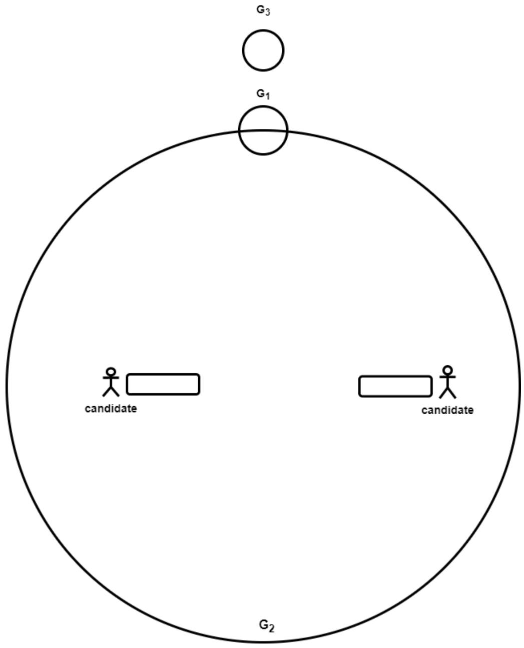
Definition 5.1 (Group of Voters)**.** *We say that voters i, j belong to group G_m if they submit the same candidates, probably in different order.*

Moreover by $C(G_m)$ we will denote the subset of candidates submitted by group G_m.
We define also the diameter $D(G_m) = \max_{i \in Gm, c \in C(G_m)} \{d(i,c)\}$, which is the largest distance between a voter in group G_m and a candidate submitted by them. Next, we will examine cases where the number of queries is $t \leq \frac{m}{2}$. In this cases, after querying the voters, there may be formed groups of voters G_1, G_2 such that $C(G_1) \cap C(G_2) = \varnothing$. This means that no voting rule can approximate the distances between voters from group G_1 and candidates $C(G_2)$. It is worth-noting that each pair of groups G_m, G_n such that $C(G_m) \cap C(G_n) = \varnothing$ is well separated from the other, in the sense that if the problem is limited to the submetric space that contains only voters from $G_m \cup G_n$ and candidates from $C(G_m) \cup C(G_n)$, it is possible to produce a solution that yields distortion $\mathcal{O}(n)$. We will now prove these observations formally.

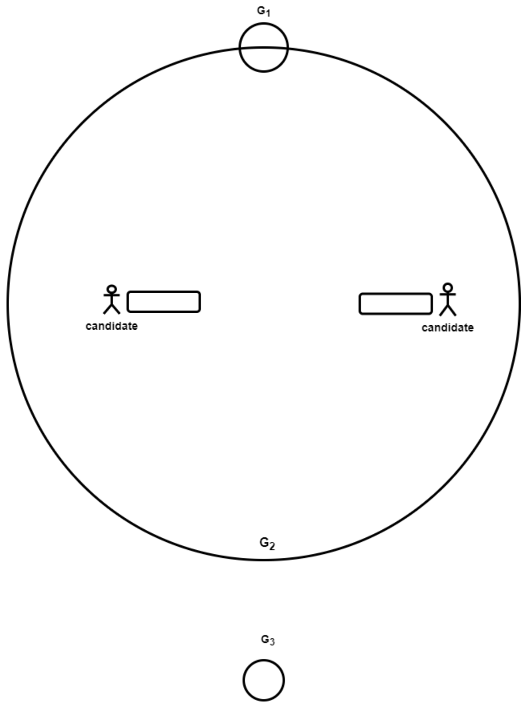
Theorem 5.6. *If the number of queries per voter is $t = \frac{m}{2}$ there is a multiwinner voting rule that achieves distortion $\mathcal{O}(n)$*

Proof We consider two cases after querying the voters. First, suppose that the distance between each voter i and candidate $c \notin b^i$ (which means that does not belong in the top t preferred candidates) can be estimated through another voter as stated in Theorem 5.4. In this case we apply Algorithm 5.4 that returns a solution of constant distortion. In contrast, suppose now that two groups of voters G_1, G_2 are formed such that $C(G_1) \cap C(G_2) = \varnothing$. Notice that this is the worst case. We can solve the subproblems for each group G_1, G_2 for $k_1 = 1, \dots, k-1$ and $k_2 = 1, \dots, k-1$ respectively, using the Algorithm 5.3 since for each group all the distances are known. The mechanism outputs the solution that minimizes the total social cost such that $k_1 + k_2 = k$. Notice that if $k_1 > 1$ and $k_2 > 1$ the distortion is constant. This is true, because the social cost is a non-increasing convex function with the respect to the number of winners. Now suppose that this mechanism returns a subset of winners such that $k_1 = k-1$ and $k_2 = 1$ and in the optimal solution all the winners belong to group G_1. The distance that each voter from group G_2 has to go though would be larger than any of the distances payed by the assignment of the mechanism. Thus the worst case distortion is $\mathcal{O}(n)$. □

Theorem 5.7. *If the number of queries per voter is $t < \frac{m}{2}$ the distortion of every multiwinner rule is unbounded.*



(a) First instance



(b) Second instance

Figure 5.5: Two instances that cannot be distinguished by any voting rule, given the top t candidates. Each circle correspond to a group of voters. The rectangles inside group G_2 correspond to two cluster of voters and each of them is close to different candidate.

Proof Consider an instance where the number of winners is $k = 3$ and the number of queries per voter is $t = \frac{m}{2} - 1$ and m is even. After querying the voters three groups of voters are formed G_1, G_2, G_3. As defined above, each group of voters submits the same t candidates. We fix the distances such that $R(G_1) << R(G_2)$ and $R(G_3) << R(G_2)$ and $C(G_1) \cap C(G_3) = \emptyset$. In other words, G_1 and G_3 form two small clusters with no common candidates and the diameter of group G_2 is arbitrarily bigger than the others. Thus the distance between groups G_1 and G_3 cannot be estimated. The voters that belong to group G_2 submit the remaining 2 candidates from $C \setminus (C(G_1) \cup C(G_3))$ and $t-2$ candidates from $C(G_1)$. Moreover group G_2 can be be split into two clusters of voters such that each of them is collocated with one of the two candidates from $C \setminus (C(G_1) \cup C(G_3))$. We can now fix two instances as illustrated in Figure 5.5 that each voting rule considers the same. In the first instance the optimal solution is to map each group of voters into a cluster, whereas in second instance the optimal solution is to assign voters from groups G_1G_2 as one cluster and split voters from group G_3 into two clusters. In each case if the voting rule assigns the clusters incorrectly the total social can be arbitrarily bigger than the optimal and thus the distortion is unbounded.

□

We examined the distortion incurred in our query model, depending on the number of approximate distances submitted by each voter. We proved that in order to achieve constant distortion, each voter had to submit the bucket in which each of his $t > \frac{m}{2}$ most preferred candidates belong. If the number of queries per voter is $t = \frac{m}{2}$ there is a voting rule that achieves distortion $\mathcal{O}(n)$ which is tight and if the number of queries per voter is $t < \frac{m}{2}$ the distortion is unbounded. It is worth-noting that the same impossibility results would have emerged, had each voter submitted the exact distances from their top t candidates. Obviously, in order to achieve bounded distortion the required number of queries, elicited in this way, is large. As explained in previous section, in metric spaces, in order for a voting rule to achieve bounded distortion, it must be possible to estimate how voters and candidates are distributed in the space. In general, the distance of the t most preferred candidates gives a local point of view and does not help us understand the voter's position in the metric space, as these t candidates can be arbitrarily close to him in comparison with the other agents. We highlight that in this thesis we stick to a model that respects the axiom of anonymity which states that each voter should be treated alike and thus our query model elicits the same information from each one. A next interesting step, that violates anonymity, would be to examine the tradeoff between the number of queries and distortion when the voting rule queries targeted voters that identifies as critical after an initial query model that elicits the same information from each voter.

Chapter 6

Concluding Remarks and Related Work

In this thesis we examine extensively the framework of distortion in two different settings. Firstly, we examine the setting where voters' preferences over candidates are expressed as numerical valuations. Given the ordinal preferences, we review the lower bounds of distortion and mechanisms that achieve tight distortion. We conclude that, in this setting, as the number of winners increases the problem becomes easier since given the same information, the bounds of distortion decay. However we only study deterministic mechanisms. In the literature there is work dedicated in the distortion of randomized mechanisms Boutilier et al. [55], Caragiannis et al. [24], Caragiannis et al. [22]. Distortion has also been studied under different objectives. An interesting example is the participatory budgeting which was studied by Benadè et al. [25]. Participatory budgeting can be thought as an extension of the multiwinner setting. In this case we are given a budget, each alternative has an associated cost, and the goal is to choose a subset of alternatives so as to maximize the social welfare of the agents while ensuring that the total cost of the chosen alternatives does not exceed the budget.

Secondly, we examine the distortion when voters and candidates lie in a metric space. The distortion in the single winner setting, given the ordinal preferences is well understood. Specifically we know that any deterministic algorithm has distortion at least 3, which is also the upper bound of distortion of the mechanism established in the work of Gatzelis et al. [49]. We review as well that, given only the top choice of each voter the Plurality rule is the optimal determinstic algorithm, achieving distortion of at most $2m - 1$. Anagnostides et al. [56] examine the decay in distortion when voters provide incomplete rankings, i.e the k-top preferences. As stated above, in this thesis we only examine deterministic mechanisms. Anshelevich et al. [50] studied the distortion of randomized social choice mechanisms when agents lie in a metric space. The respective lower bound in the single winner setting is 2. Moreover, besides the objective of the total social cost the distortion has been studied under the objective of minimizing the median cost which is the $\lceil \frac{n}{2} + 1 \rceil$- largest cost for a chosen candidate. This objective captures the notion of fairness.

Our innovative results concern the distortion in the metric multiwinner voting. As mentioned, if the whole distance profile was given, the metric multiwinner voting is equivalent to *metric k-Median Problem.* Since the ordinal rankings are not enough to achieve bounded distortion, we propose a new framework that reduces the cognitive ability required from voters, as they are asked to give only an approximation of their distance from their most preferred candidates. We examined the tradeoff between distortion and the number of such queries/approximations per voter and prove that in order to achieve bounded distortion this number should be bigger than $\frac{m}{2}$. A recommended idea would be to reduce the number of such queries per voter in a first step and ask in a second step targeted voters to answer a few more queries.

Bibliography

[1] Kenneth Joseph Arrow. *Social Choice and Individual Values.* New York, NY, USA: Wiley: New York, 1951.

[2] Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.

[3] Mark Allen Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

[4] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, 1944.

[5] James M. Enelow and Melvin J. Hinich. *The Spatial Theory of Voting.* Number 9780521275156 in Cambridge Books. Cambridge University Press, 1984.

[6] Norman Schofield. The spatial model of politics. *The Spatial Model of Politics*, pages 1–241, 12 2007.

[7] Otto A. Davis, Melvin J. Hinich, and Peter C. Ordeshook. An expository development of a mathematical model of the electoral process. *American Political Science Review*, 64(2):426–448, 1970.

[8] Kenneth Arrow. *Advances in the Spatial Theory of Voting.* Cambridge University Press, 1990.

[9] Richard D. McKelvey, Peter C. Ordeshook, and Kenneth Arrow. *A Decade of Experimental Research on Spatial Models of Elections and Committees*, page 99–144. Cambridge University Press, 1990.

[10] Samuel Iii Merrill and Bernard Grofman. A unified theory of voting: Directional and proximity spatial models. 1999.

[11] Ariel D. Procaccia and Jeffrey S. Rosenschein. The distortion of cardinal preferences in voting. In Matthias Klusch, Michael Rovatsos, and Terry R. Payne, editors, *Cooperative Information Agents X, 10th International Workshop, CIA 2006, Edinburgh, UK, September 11-13, 2006, Proceedings*, volume 4149 of *Lecture Notes in Computer Science*, pages 317–331. Springer, 2006.

[12] Vijay V. Vazirani. *Approximation Algorithms.* Springer Publishing Company, Incorporated, 2010.

[13] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In Christoph Meinel and Sophie Tison, editors, *STACS 99*, pages 404–413, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.

[14] Michal Feldman, Amos Fiat, and Iddan Golomb. On voting and facility location, 2015.

[15] M. Charikar and S. Guha. Improved combinatorial algorithms for the facility location and k-median problems. In *40th Annual Symposium on Foundations of Computer Science (Cat. No.99CB37039)*, pages 378–388, 1999.

[16] Kamal Jain, Mohammad Mahdian, Evangelos Markakis, Amin Saberi, and Vijay V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing lp. *J. ACM*, 50(6):795–824, nov 2003.

[17] Vincent Cohen-Addad, Marcin Pilipczuk, and Michał Pilipczuk. A polynomial-time approximation scheme for facility location on planar graphs, 2019.

[18] Chris Geller. Single transferable vote with borda elimination: a new vote counting system. Working papers, Deakin University, Department of Economics, 2002.

[19] Haris Aziz. Justifications of welfare guarantees under normalized utilities. *SIGecom Exch.*, 17(2):71–75, January 2020.

[20] John R. Chamberlin and Paul N. Courant. Representative deliberations and representative decisions: Proportional representation and the borda rule. *The American Political Science Review*, 77(3):718–733, 1983.

[21] Tjalling C. Koopmans and Martin Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25(1):53–76, 1957.

[22] 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*, IJCAI'16, page 151–157. AAAI Press, 2016.

[23] Vincent Conitzer and Tuomas Sandholm. Communication complexity of common voting rules. In *Proceedings of the 6th ACM Conference on Electronic Commerce*, EC '05, page 78–87, New York, NY, USA, 2005. Association for Computing Machinery.

[24] Debmalya Mandal, Ariel D Procaccia, Nisarg Shah, and David Woodruff. Efficient and thrifty voting by any means necessary. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[25] Gerdus Benade, Swaprava Nath, Ariel D. Procaccia, and Nisarg Shah. Preference elicitation for participatory budgeting. *Manag. Sci.*, 67:2813–2827, 2017.

[26] Debmalya Mandal, Nisarg Shah, and David P. Woodruff. Optimal communication-distortion tradeoff in voting. In *Proceedings of the 21st ACM Conference on Economics and Computation*, EC '20, page 795–813, New York, NY, USA, 2020. Association for Computing Machinery.

[27] Eyal Kushilevitz and Noam Nisan. *Communication Complexity.* Cambridge University Press, 1996.

[28] Duncan Black. On the rationale of group decision-making. *Journal of Political Economy*, 56(1):23–34, 1948.

[29] Anthony Downs. An economic theory of political action in a democracy. *Journal of Political Economy*, 65(2):135–150, 1957.

[30] Duncan Black. *The theory of committees and elections.* University Press, Cambridge, 1958.

[31] H. Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

[32] Barbera Salvador, Gul Faruk, and Stacchetti Ennio. Generalized Median Voter Schemes and Committees. *Journal of Economic Theory*, 61(2):262–289, December 1993.

[33] Diana Richards, Whitman A. Richards, and Brendan D. McKay. Collective Choice and Mutual Knowledge Structures. Research in Economics 98-04-032e, Santa Fe Institute, April 1998.

[34] Salvador Barberà. An introduction to strategy-proof social choice functions. *Social Choice and Welfare*, 18(4):619–653, 2001.

[35] Ken ichi Inada. The simple majority decision rule. *Econometrica*, 37(3):490–506, 1969.

[36] Joshua Gans and Michael Smart. Majority voting with single-crossing preferences. *Journal of Public Economics*, 59(2):219–237, 1996.

[37] Jean-Michel Grandmont. Intermediate preferences and the majority rule. *Econometrica*, 46(2):317–30, 1978.

[38] P. Rothstein. Order restricted preferences and majority rule. *Social Choice and Welfare*, 7(4):331–342, 1990.

[39] Allan H. Meltzer and Scott F. Richard. A rational theory of the size of government. *Journal of Political Economy*, 89(5):914–927, 1981.

[40] Gabrielle Demange. Intermediate preferences and stable coalition structures. *Journal of Mathematical Economics*, 23(1):45–58, 1994.

[41] Frank Westhoff. Existence of equilibria in economies with a local public good. *Journal of Economic Theory*, 14(1):84–112, February 1977.

[42] Dennis Epple and Glenn J. Platt. Equilibrium and local redistribution in an urban economy when households differ in both preferences and incomes. *Journal of Urban Economics*, 43(1):23–51, 1998.

[43] Salvador Barberà and Matthew Jackson. Choosing how to choose: Self-stable majority rules and constitutions. *The Quarterly Journal of Economics*, 119(3):1011–1048, 2004.

[44] Elliot Anshelevich, Onkar Bhardwaj, and John Postl. Approximating optimal social choice under metric preferences. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, page 777–783. AAAI Press, 2015.

[45] Piotr Skowron and Edith Elkind. Social choice under metric preferences: Scoring rules and stv. In *AAAI*, 2017.

[46] H. Moulin. Choosing from a tournament. *Social Choice and Welfare*, 3(4):271–291, 1986.

[47] Arthur Herbert Copeland. A "reasonable" social welfare function, seminar on mathematics in social sciences. 1951.

[48] Kamesh Munagala and Kangning Wang. Improved metric distortion for deterministic social choice rules. *Proceedings of the 2019 ACM Conference on Economics and Computation*, Jun 2019.

[49] Vasilis Gkatzelis, Daniel Halpern, and Nisarg Shah. Resolving the optimal metric distortion conjecture, 2020.

[50] Elliot Anshelevich and John Postl. Randomized social choice functions under metric preferences, 2016.

[51] Elliot Anshelevich and Wennan Zhu. Ordinal approximation for social choice, matching, and facility location problems given candidate positions, 2018.

[52] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 184–193, 1996.

[53] Yair Bartal. On approximating arbitrary metrices by tree metrics. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, page 161–168, New York, NY, USA, 1998. Association for Computing Machinery.

[54] Moses Charikar, Sudipto Guha, Éva Tardos, and David B. Shmoys. A constant-factor approximation algorithm for the k-median problem. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, STOC '99, page 1–10, New York, NY, USA, 1999. Association for Computing Machinery.

[55] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D. Procaccia, and Or Sheffet. Optimal social choice functions: A utilitarian view. EC '12, page 197–214, New York, NY, USA, 2012. Association for Computing Machinery.

[56] Ioannis Anagnostides, Dimitris Fotakis, and Panagiotis Patsilinakos. Metric-distortion bounds under limited information, 2021.