

# Exclusion Zones of Instant Runoff Voting

Kiran Tomlinson<sup>1</sup>, Johan Ugander<sup>2</sup>, Jon Kleinberg<sup>3</sup>

<sup>1</sup>Microsoft Research

<sup>2</sup>Yale University

<sup>3</sup>Cornell University

kitomlinson@microsoft.com, johan.ugander@yale.edu, kleinberg@cornell.edu

## Abstract

Recent research on instant runoff voting (IRV) shows that it exhibits a striking property over unimodal one-dimensional preferences: there is an *exclusion zone* around the median voter such that the winner must come from the exclusion zone, unless no such candidate exists. Thus, IRV cannot elect an extreme candidate in this setting as long as a sufficiently moderate candidate runs. In this work, we examine the mathematical structure of exclusion zones as a broad phenomenon in more general preference spaces. We prove that with voters uniformly distributed over any  $d$ -dimensional hyperrectangle (for  $d > 1$ ), IRV has no such exclusion zone. However, we also show that IRV exclusion zones are not solely a one-dimensional phenomenon. For irregular higher-dimensional preference spaces with fewer symmetries than hyperrectangles, IRV can have nontrivial exclusion zones. As a further exploration, we study IRV exclusion zones with graph-based preferences, where nodes represent voters who prefer candidates closer to them in the graph. Here, we show that IRV exclusion zones present a surprising computational challenge: even checking whether a given set of positions is an IRV exclusion zone is NP-hard. We develop an efficient randomized approximation algorithm for checking and finding exclusion zones in graphs. Finally, we report on computational experiments with exclusion zones: (i) performing an exhaustive computer search of small graphs and trees, we find nontrivial IRV exclusion zones in most small graphs; and (ii) applying our approximation algorithm to a collection of larger real-world school friendship networks, we find that about 60% of these networks have probable nontrivial IRV exclusion zones. While our focus is on IRV, the properties of exclusion zones we establish provide new methods for analyzing voting systems in metric spaces more generally.

**Code** — [github.com/tomlinsonk/irv-exclusion-zones](https://github.com/tomlinsonk/irv-exclusion-zones)

**Extended version** — [arxiv.org/abs/2502.16719](https://arxiv.org/abs/2502.16719)

## Introduction

An important principle in voting theory is that different voting systems can implicitly favor different regions of the underlying ideological space, for example by benefitting more moderate or more extreme positions. The classic result in this vein is the Median Voter Theorem (Black 1948), which

states that when preferences are single-peaked, voting systems satisfying the Condorcet criterion elect the candidate closest to the median voter. That is, Condorcet methods strongly favor moderate candidates in one-dimensional preference spaces. However, this result does not provide us with tools to analyze non-Condorcet methods or voting systems under higher-dimensional preferences. In those cases, prior work has often turned to simulation to explore the typical behavior of various voting systems (Elkind et al. 2017; McGann, Grofman, and Koetzle 2002; Merrill III 1984).

Recent work on instant runoff voting (IRV), which is not Condorcet, provides a new approach to theoretically characterizing the candidate positions favored by a voting system. Tomlinson, Ugander, and Kleinberg (2024) proved that IRV with symmetrically-distributed one-dimensional preferences exhibits a striking property: it can only elect a candidate in an interval around the median voter, termed the *exclusion zone*. More generally, an exclusion zone for a given voting system and voter distribution is a set  $S$  of candidate positions with the property that if any candidate belongs to  $S$ , then the winner of the election must belong to  $S$ . One of the main results of Tomlinson, Ugander, and Kleinberg (2024) is that with voters uniformly distributed over  $[0, 1]$ , the interval  $[1/6, 5/6]$  is an exclusion zone of IRV. Thus, regarding the more central points of the interval as the more moderate positions, IRV cannot elect an extreme candidate over a sufficiently moderate one.

However, this prior work left open the question of whether exclusion zones were specific to the structure of the unit interval, or whether they represented a more general principle. In this work we address this question, extending the analysis of IRV exclusion zones to voters uniformly distributed over metric spaces beyond the unit interval, and fleshing out an understanding of exclusion zones as a general phenomenon of voting systems. We resolve the natural open question about whether the one-dimensional IRV exclusion zone generalizes to higher dimensions, showing for both  $L_1$  and  $L_2$  distance metrics that uniform voters over any  $d$ -dimensional hyperrectangle with  $d > 1$  produce no (nontrivial) IRV exclusion zones. While this strong nonexistence result seems to suggest that IRV exclusion zones are a purely one-dimensional phenomenon, we show this is not the case via a counterexample, an irregular higher-dimensional preference space with a nontrivial IRV exclusion zone. In or-

der to prove our main result for hyperrectangles, we derive general properties of exclusion zones that reveal their structure. We show that the exclusion zones of any voting system in any metric preference space form a nesting from the trivial exclusion zone (containing all positions) down to a unique *minimal exclusion zone*. These properties of exclusion zones allow us to develop a constructive proof recipe for the nonexistence of nontrivial exclusion zones, which we term the Condorcet Chain Lemma. Informally, this lemma states that if we can construct a sequence of elections such that (1) the first election is lost by a Condorcet winner, (2) the last election is won by a Condorcet loser, and (3) the winner of each election loses the next one, then there is no nontrivial exclusion zone. Our proof for hyperrectangles constructs such a sequence of elections.

In addition to higher-dimensional spatial preferences, we also consider a setting motivated by friendship- or allegiance-based preferences: voting on graphs, where nodes represent voters (some of which run for election) who prefer candidates closer to them in the graph. (Note that one-dimensional voting is closely related to voting on path graphs.) In this setting, we consider the algorithmic problem of identifying IRV exclusion zones in general graphs, which we show is computationally hard. We find that even checking whether a given set of nodes is an IRV exclusion zone is co-NP-complete, and finding a graph’s minimal IRV exclusion zone is at least as hard.

On the positive side, we are able to obtain tractability results by introducing a notion of approximate exclusion zones; roughly speaking, a set of nodes  $S$  is a  $(1 - \epsilon)$ -*approximate exclusion zone* if a random set of candidates containing at least one element of  $S$  has its winner from  $S$  with probability at least  $1 - \epsilon$ . Using this definition, we are able to give a randomized approximation algorithm with the following two-sided guarantee: it outputs a set of nodes that is both guaranteed to be a subset of the true minimal IRV exclusion zone, and also with high probability to be an approximate exclusion zone. (For this algorithm, we must also design an efficient subroutine for checking if a set of nodes is an approximate exclusion zone.) This algorithm is useful—both theoretically and in practice—as a certifier that a graph has no nontrivial exclusion zone: its underlying guarantee tells us that if it outputs the full node set, then it has provided such a certificate. We also show that the problem of checking IRV exclusion zones is fixed-parameter tractable in the number of nodes outside the candidate set, so large IRV exclusion zones can be efficiently checked.

While identifying graph IRV exclusion zones is hard in general, we derive the minimal IRV exclusion zones for several graph families, including distance-regular graphs (van Dam, Koolen, and Tanaka 2016), paths, perfect binary trees, and bistars. As with continuous metric spaces, this analysis is aided by the Condorcet Chain Lemma. We demonstrate the usefulness of our approximation algorithm on a collection of 56 real-world friendship networks (Paluck, Shepherd, and Aronow 2016) that are much too large for exact computation. We find that 33/56 networks (59%) likely have approximate IRV exclusion zones, and certify that the other 23 have no nontrivial IRV exclusion zones. The approximate

exclusion zones we find tend to be large, comprising an average of 95% of the network. Nodes that are excluded are usually on the fringe of the network, although sometimes entire communities are excluded. Through an exhaustive computer search of small graphs and trees, we find that most connected graphs have nontrivial IRV exclusion zones.

Overall, our work highlights exclusion zones as rich mathematical objects illuminating the behavior of voting systems with metric voter preferences. We focus on IRV, partly because it is the predominant single-winner alternative to plurality voting—used for instance in Australia, Ireland, and Maine, as well as cities including San Francisco and New York City (for primaries)—and partly to resolve open questions from prior work introducing exclusion zones. However, our study of exclusion zone properties applies more broadly, situating them as a general phenomenon for voting, and setting the stage for future work on exclusion zones of other voting systems.

## Related Work

Our work builds directly on the concept of exclusion zones introduced by Tomlinson, Ugander, and Kleinberg (2024), who studied them for IRV and plurality with one-dimensional Euclidean preferences. The concepts of Condorcet winning sets,  $\theta$ -winning sets, dominating sets, and undominated sets (Elkind, Lang, and Saffidine 2015; Bloks 2018; Charikar et al. 2025; Charikar, Ramakrishnan, and Wang 2025; Nguyen, Song, and Lin 2025) are related in spirit to exclusion zones, in that they describe sets of candidates preferred by voters to candidates outside the set. However, there are a few important distinctions. First, exclusion zones are sets of possible candidate positions across a specified family of profiles (the voting theoretic term for a collection of voter preferences over a set of candidates) rather than candidates in a specific profile. Second, exclusion zones are defined by the outcome of the election under a given voting system rather than pairwise preferences of voters.

Another notable approach to studying voting systems in metric spaces comes from the literature on utilitarian metric distortion (Anshelevich et al. 2018). In this framework, voters and candidates have unknown positions in a metric space and the *distortion* of a voting system is the worst-case ratio (over positions consistent with expressed voter rankings) between the total distance from voters to the elected candidate and the minimal total distance to a candidate. No voting system can have metric distortion better than 3 (Anshelevich, Bhardwaj, and Postl 2015) and voting systems achieving this bound are known (Gkatzelis, Halpern, and Shah 2020; Kizilkaya and Kempe 2022). The distortions of other voting rules are also known, including for Borda count, plurality, IRV, and Copeland (Anshelevich et al. 2018; Anagnostides, Fotakis, and Patsilinakos 2022). Distortion is a valuable tool for comparing voting systems, but answers a different question than our work. We ask what regions of a space are favored by a voting system over all possible candidate sets rather than how bad an outcome can be in a worst case over voter and candidate positions.

Relating to our study of voting on graphs, graph-based preferences have a long history in the facility location liter-

ature, where facilities can be viewed as candidates and customers can be viewed as voters. A Condorcet node is then a facility preferred by more than half of customers to any other facility (Wendell and McKelvey 1981; Bandelt 1985; Hansen, Thisse, and Wendell 1986). Graph-based preferences also occasionally appear in the social choice literature as a special case of metric preferences (Skowron and Elkind 2017). One working paper uses graph-based voting to explain the success of the Medici family in medieval Florence (Telek 2016). From a different angle, graph-distance voting has recently been proposed as a node centrality measure (Brandes, Laubmann, and Rothe 2022; Skibski 2023).

## Voting Preliminaries

We begin by defining terms for our analysis. An *election* consists of a set of candidates  $C$ , voters  $V$  with preferences over  $C$ , and a voting rule  $r$  specifying how a winner should be elected from  $C$  based on the voter preferences. We focus on metric voter preferences (also called spatial preferences), where voters and candidates have positions in a bounded metric space  $(M, d)$  with set  $M$  and metric  $d$ . Voters prefer candidates closer to them under  $d$ . To model a large voting population, we let  $V$  be a continuous and positive distribution over  $M$  and measure votes with real-valued shares (similar results would also hold for discrete sets of voters). In this work, we always have  $V$  uniformly distributed over  $M$ , for simplicity rather than necessity. The tuple  $(M, d, V, r)$  defines an *election setting*, which in combination with a candidate configuration  $C \subset M$  fully specifies an election.

The simplest voting rule is *plurality*, where the candidate most preferred by the greatest number of voters is elected (with ties broken in some pre-specified way). We focus on a prominent alternative to plurality, *instant runoff voting* (IRV), also known as ranked-choice voting, single transferable vote (STV), or the Hare method. Under IRV, we repeatedly compute plurality vote shares for each candidate and eliminate the candidate with the fewest votes, until only the winner remains. In practice, this procedure is accomplished by asking voters to rank the candidates in order of preference, so that these “runoffs” can be computed “instantly.”

We now define a slight variation of classic Condorcet winners and losers, but using *positions* in the metric space rather than candidates in a particular profile. Given  $(M, d, V)$ , a *weak Condorcet position* is a candidate position  $x \in M$  such that for any other  $y \in M$ ,  $x$  is preferred to  $y$  by at least half of the voters (such a position may not exist).<sup>1</sup> Intuitively, a candidate at a Condorcet position is Condorcet winner against any possible set of opponents from  $M$ . Similarly, a *weak anti-Condorcet position* is a position  $y \in M$  such that for any other  $x \in M$ ,  $x$  is preferred to  $y$  by at least half of the voters. For example, with uniformly distributed voters over  $[0, 1]$ ,  $1/2$  is a Condorcet position, while 0 and 1 are weak anti-Condorcet positions. Any Condorcet voting rule elects a candidate at a Condorcet position whenever one is present; neither IRV nor plurality are Condorcet.

<sup>1</sup>The set of weak Condorcet positions is known as the *core* (Plott 1967; Schofield 1983).

## Exclusion Zones and Their Properties

We now define our object of focus: exclusion zones, introduced by Tomlinson, Ugander, and Kleinberg (2024). Intuitively, an exclusion zone of an election setting is a set  $S$  of candidate positions such that any election with at least one candidate from  $S$  is won by a candidate in  $S$ . This indicates which positions are favored by the voting rule.

**Definition 1.** A nonempty set  $S \subseteq M$  is an *exclusion zone* of an election setting  $(M, d, V, r)$  if the winner under  $r$  with voters  $V$  is guaranteed (for all tiebreaking orders) to be in  $S$  for any candidate configuration  $C$  with  $C \cap S \neq \emptyset$ .

When  $M = [0, 1]$ ,  $d$  is the Euclidean metric (i.e., voters have *1-Euclidean preferences* (Elkind, Lackner, and Peters 2025)), and the voters  $V$  are uniform over  $[0, 1]$ ,  $S = [1/6, 5/6]$  is an exclusion zone of IRV (Tomlinson, Ugander, and Kleinberg 2024). IRV also has exclusion zones for other symmetric voter distributions over  $[0, 1]$  (Tomlinson, Ugander, and Kleinberg 2024). However, this earlier work did not further explore the concept of exclusion zones, which we now show exhibit some elegant structure. We begin by showing that exclusion zones are nested inside  $M$ .

**Proposition 1.** Let  $S \neq T$  be two exclusion zones of  $(M, d, V, r)$ . Either  $S \subset T$  or  $T \subset S$ .

*Proof.* Suppose not, i.e.,  $S \not\subset T$  and  $T \not\subset S$ . Then there is some  $s \in S$  such that  $s \notin T$  and some  $t \in T$  such that  $t \notin S$ . Consider the election with candidates  $C = \{s, t\}$ . If  $s$  is not guaranteed to win, then  $S$  is not an exclusion zone. If  $t$  is not guaranteed to win, then  $T$  is not an exclusion zone. So either  $S$  or  $T$  is not an exclusion zone, a contradiction.  $\square$

In the 1-Euclidean uniform voters case, Tomlinson, Ugander, and Kleinberg (2024) showed that any interval  $[c, 1 - c]$  for  $c \in [0, 1/6]$  is an exclusion zone of IRV—these zones are indeed nested—and that no interval smaller than  $[1/6, 5/6]$  is an exclusion zone. Here, we formalize the notion of the smallest exclusion zone of a voting system.

**Definition 2.** An exclusion zone  $S$  is *minimal* if no proper subset of  $S$  is an exclusion zone.

In this language,  $[1/6, 5/6]$  is the minimal IRV exclusion zone for uniform voters over  $[0, 1]$ . Proposition 1 reveals that there cannot be more than one minimal exclusion zone, and we can also show one always exists. See the extended version for all proofs omitted for space (Tomlinson, Ugander, and Kleinberg 2025).

**Proposition 2.** Let  $S$  be the set of all exclusion zones of  $(M, d, V, r)$ . The unique minimal exclusion zone is given by  $S^* = \bigcap_{S \in \mathcal{S}} S$ .

Next, we say that exclusion zone  $S = M$  is the *trivial* exclusion zone, which always satisfies the definition. If  $S$  is a proper subset of  $M$ , we call  $S$  *nontrivial*. We thus have that the set of exclusion zones forms a nested chain from the trivial exclusion zone  $M$  down to the minimal exclusion zone  $S^*$ . In some cases,  $S^* = M$ , in which case there is no nontrivial exclusion zone. As a warmup, we now provide some examples of voting systems and their exclusion zones.

**Proposition 3.** *Let  $(M, d)$  be a bounded interval in one dimension with the Euclidean metric. For any Condorcet voting rule  $r$  and any voter distribution  $V$  over  $M$ ,  $\{\text{median}(V)\}^2$  is the minimal exclusion zone of  $(M, d, V, r)$ .*

In this Condorcet setting, we can actually characterize the entire space of exclusion zones, although the statements are simpler if we restrict  $M$  to  $[0, 1]$  and restrict  $V$  to be a symmetric distribution about  $1/2$ . To do this, we first develop some useful facts about exclusion zones, which will also prove useful for our main results in the next section.

**Proposition 4.** *Given an election setting  $(M, d, V, r)$  such that  $r$  picks the majority winner in pairwise contests,*

- (a) *Any weak Condorcet position is in the minimal exclusion zone.*
- (b) *For any exclusion zone  $S$ , if there is some candidate configuration including  $x \in S$  where  $y$  wins under some tiebreaking of  $r$ , then  $y \in S$ .*
- (c) *If the minimal exclusion zone contains a weak anti-Condorcet position, then the minimal exclusion zone is trivial.*

Using Proposition 4(b), we can now characterize the space of all exclusion zones for Condorcet methods with symmetric 1-Euclidean voters.

**Proposition 5.** *Let  $M = [0, 1]$ ,  $d$  be the  $L_2$  metric, and  $V$  be a distribution over  $[0, 1]$  symmetric about  $1/2$ . For any Condorcet voting rule  $r$ , the exclusion zones of  $(M, d, V, r)$  are exactly  $\mathcal{S} = \{\{1/2\}\} \cup \bigcup_{c \in [0, 1/2)} \{[c, 1 - c], (c, 1 - c)\}$ .*

As another example, consider plurality voting with uniform voters over  $[0, 1]$ . It is known that in this setting, no proper subset of  $(0, 1)$  is an exclusion zone (Tomlinson, Ugander, and Kleinberg 2024). We complete this characterization by showing that  $(0, 1)$  is an exclusion zone.<sup>3</sup>

**Proposition 6.** *The minimal exclusion zone of plurality with uniform voters over  $[0, 1]$  is  $(0, 1)$ .*

In several instances, we will encounter election settings where the minimal exclusion zone is trivial (equivalently, there are no nontrivial exclusion zones), including our main result for hyperrectangles. Proving that a voting system has no nontrivial exclusion zones appears to require proving a broad statement of nonexistence. However, we show that the structure of exclusion zones explored above enables a constructive proof recipe that we term the Condorcet Chain Lemma. Specifically, the facts from Proposition 4 allow us to chain together a sequence of candidate configurations, starting from a weak Condorcet position and ending at a weak anti-Condorcet position, implying that there are no nontrivial exclusion zones.

<sup>2</sup>Since  $V$  has positive density over a bounded interval  $[a, b]$ ,  $\text{median}(V)$  is uniquely defined as the point  $m$  where  $\int_a^m f_V(x) dx = \int_m^b f_V(x) dx = 1/2$ .

<sup>3</sup>As long as no duplicate candidate positions are allowed; if we do allow duplicates, the minimal exclusion zone of plurality is trivial. Throughout the paper, we assume the candidate set has no duplicates, although for IRV this does not affect exclusion zones.

**Lemma 1** (Condorcet Chain Lemma). *Given an election setting  $(M, d, V, r)$  such that  $r$  picks the majority winner in pairwise contests, if there are configurations  $C_1, \dots, C_n$  with candidates  $w_1 \in C_1, \dots, w_n \in C_n$  such that:*

1.  *$C_1$  contains a weak Condorcet position, but a different candidate  $w_1$  wins (possibly after tiebreaking),*
2. *each  $C_{i+1}$  includes  $w_i$ , but some other candidate  $w_{i+1}$  wins (possibly after tiebreaking),*
3.  *$w_n$  is a weak anti-Condorcet position,*

*then  $(M, d, V, r)$  has no nontrivial exclusion zones.*

*Proof.* By Proposition 4(a), the minimal exclusion zone  $S^*$  contains all weak Condorcet positions. But  $w_1$  defeats a weak Condorcet winner, so  $w_1 \in S$  by Proposition 4(b). Repeated applications of Proposition 4(b) along the chain of configurations (i.e., induction) show that each  $w_i \in S^*$ , so  $w_n \in S^*$ . By Proposition 4(c), this means  $S^* = M$ , as  $w_n$  is a weak anti-Condorcet position, so there are no nontrivial exclusion zones.  $\square$

## IRV Exclusion Zones in Higher Dimensions

Now that we have established some useful general properties of exclusion zones, we turn to our first main results. While Tomlinson, Ugander, and Kleinberg (2024) characterized the exclusion zones of IRV in one-dimension, that work left a significant open question: does IRV have nontrivial exclusion zones with higher-dimensional preferences? The proof technique in one dimension fails to generalize, leaving the answer uncertain. Here, we resolve this open question: using the Condorcet Chain Lemma, we show that IRV has no nontrivial exclusion zones with uniformly distributed voters over hyperrectangles of dimension two or greater (with both  $L_1$  and  $L_2$  distance metrics). However, we also show that there exist higher-dimensional preference structures that induce nontrivial IRV exclusion zones. The example we construct generates an IRV exclusion zone by breaking the strong symmetry of hyperrectangles, which our proof suggests is the cause of their lack of IRV exclusion zones.

To apply the Condorcet Chain Lemma to higher-dimensional Euclidean preferences, we start by finding (anti-)Condorcet positions in hyperrectangles. We say that voters have  $L_p$  preferences, or are  $L_p$  voters, in a  $d$ -dimensional space if they vote according to the  $L_p$  metric, with distance function  $d_p(x, y) = \left(\sum_{i=1}^d |x_i - y_i|^p\right)^{1/p}$ .

**Lemma 2.** *With uniform  $L_p$  voters over a  $d$ -dimensional hyperrectangle  $[0, w_1] \times \dots \times [0, w_d]$ , the center  $c = (w_1/2, \dots, w_d/2)$  is a Condorcet position and the corners are weak anti-Condorcet positions for any  $d \geq 1$  and  $p \geq 1$ .*

So, if we can find a sequence of configurations where the winner of each configuration loses in the next and linking the center and the corners, then that hyperrectangle has no nontrivial exclusion zones by the Condorcet Chain Lemma. This is impossible in one dimension (as this would contradict the known exclusion zone), but we find such a sequence for any  $d \geq 2$ . As a warmup, we show what one such sequence looks like, for the special case of a square with uniform  $L_2$  voters.

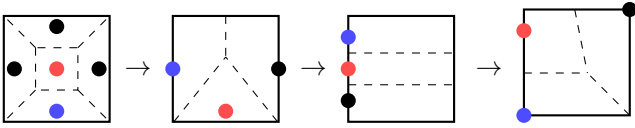


Figure 1: A visual sketch of our proof of Proposition 7, showing a sequence of elections satisfying the Condorcet Chain Lemma, proving that it has no nontrivial IRV exclusion zones with uniform  $L_2$  voters. In each configuration, the red candidate is eliminated first and the blue candidate is a possible winner of the resulting tiebreak. Critically, the red candidate was a possible winner of the previous configuration. The first configuration includes the center, the Condorcet position, and the winner of the last configuration is a corner, a weak anti-Condorcet position. The positions and vote shares are given in the proof in the extended version.

**Proposition 7.** *The square with uniform  $L_2$  voters has no nontrivial IRV exclusion zone.*

As a visual sketch of proof, Figure 1 shows a sequence of configurations that satisfies the requirements of the Condorcet Chain Lemma, where the red candidate loses at each step, but was a blue possible winner in the previous step.

Moving on from the square, we now provide a much more general result, although the sequence of configurations requires many more steps. We begin by showing that it suffices to construct such a sequence for two-dimensional rectangles—we can essentially decompose a hyperrectangle into a sequence of rectangles, one for each dimension, and make progress towards a corner one dimension at a time.

**Lemma 3.** *Let  $p \geq 1$ . If there is a sequence of configurations satisfying Lemma 1 for any rectangle with uniform  $L_p$  voters, then every  $d$ -dimensional hyperrectangle for  $d \geq 2$  has no nontrivial exclusion zones with uniform  $L_p$  voters.*

By providing such sequences for arbitrary rectangles with uniform  $L_1$  and  $L_2$  voters, we obtain our main result for higher dimensional preferences.

**Theorem 1.** *Every  $d$ -dimensional hyperrectangle ( $d \geq 2$ ) has no nontrivial exclusion zones with uniform  $L_1$  or  $L_2$  voters.*

This strong result seems to suggest that nontrivial exclusion zones for IRV are a purely one-dimensional phenomenon. However, we find that this is not the case. Indeed, a second dimension does make it much easier for non-central candidates to combine and squeeze out a more central candidate, but this needs to be paired with the strong symmetry of hyperrectangles with uniform voters in order to make the minimal exclusion zone trivial. By breaking this symmetry, we can find higher-dimensional preference spaces with nontrivial IRV exclusion zones.

**Theorem 2.** *Consider the shape  $F$  (see Figure 2) formed by a rectangle of height  $1/10$  and width 8 (having its lower left corner at the origin) with two right triangles of side lengths  $(2, 2, \sqrt{8})$  and  $(1, 1, \sqrt{2})$  placed on the top left and bottom right of the rectangle. The set  $S = \{(x, y) \in F \mid x - y \leq 6\}$  is an IRV exclusion zone with uniform  $L_1$  voters over  $F$ .*



Figure 2: The shape  $F$  from Theorem 2, which has a nontrivial IRV exclusion zone with uniform  $L_1$  voters. The shaded set  $S$  is a nontrivial IRV exclusion zone.

The preference space  $F$  in Theorem 2 has a natural interpretation: there are two relevant policy dimensions, one in which voters are strongly polarized towards opposite extremes (the  $x$  dimension) and one in which voters are mostly moderate (the  $y$  dimension), but tend to have opposite leanings on opposite sides of the  $x$  dimension. In this setting, IRV can never elect a candidate on the smaller extreme side. The same idea can be generalized to higher dimensions. However, the construction does not seem to work for  $L_2$  preferences; it remains an open question whether there are connected higher-dimensional preference spaces with nontrivial IRV exclusion zones for  $L_2$  voters (if we allow disconnected spaces, consider uniform voters over a large polytope and a small polytope that are very far apart: only candidates in the large polytope can win).

## Voting With Graph-Based Preferences

We now turn our attention to a different metric space: unweighted graphs. In this setting, voters and candidates are nodes in a graph, with preferences determined by path distance. We assume every node in the graph represents a single voter and some subset of these nodes run as candidates. Note that distance ties are common in unweighted graphs; we say that each node has vote share 1 that it evenly distributes among all closest candidates.<sup>4</sup> Formally, we define an election with graph-based preferences as follows.

**Definition 3.** An election on a graph  $G = (V, E)$  has election setting  $(V, d_G, \text{Uniform}(V), r)$ , where  $d_G$  is the path distance metric of  $G$  and voters are uniform over  $V$ . The candidates are positioned at nodes in the graph, with  $C \subseteq V$ .

Such graph-based preferences are used in the facility location literature (Wendell and McKelvey 1981; Bandelt 1985; Hansen, Thisse, and Wendell 1986), and in our voting setting can model friendship- or allegiance-based voting (Telek 2016). For instance, consider a class president election. If every student votes for the candidate they are closest friends with (measured by path distance in the friendship graph), then their preferences are given by the graph metric.

<sup>4</sup>This approach to resolving indifference among voters has recently been called Split-IRV to contrast it with Approval-IRV, where each tied candidate receives one full approval vote (Delemazure and Peters 2024). We use Split-IRV rather than Approval-IRV as it more closely parallels the continuous metric space case. Split-IRV has also been used in real-world elections (Mollison 2023), although there are good theoretical reasons to prefer Approval-IRV (Delemazure and Peters 2024).

## Finding IRV Exclusion Zones in General Graphs

Given our graph-based preference setting, we begin by considering a general algorithmic problem: can we identify the IRV exclusion zones of a given graph?

**Definition 4.** Given a graph  $G = (V, E)$  and a set of nodes  $S \subseteq V$ , IRV-EXCLUSION is the decision problem asking whether  $S$  is an IRV exclusion zone of  $G$ . MIN-IRV-EXCLUSION is the optimization problem whose solution is the minimal IRV exclusion zone of  $G$ .

We show that exclusion zones are very difficult to identify in general—even checking whether a given set is an IRV exclusion zone is computationally hard.

**Theorem 3.** IRV-EXCLUSION is co-NP-complete and MIN-IRV-EXCLUSION is NP-hard.

Despite the hardness of identifying exclusion zones, we show that we can identify *approximate* exclusion zones, which we define to be node sets that behave like exclusion zones most of the time.

**Definition 5.** A set of nodes  $S$  is a  $(1 - \epsilon)$ -approximate exclusion zone for voting rule  $r$  if, drawing a uniformly random candidate configuration  $C \subseteq V$  with  $C \cap S \neq \emptyset$ , the winner under  $r$  is in  $S$  w.p. at least  $1 - \epsilon$ .

We show that by keeping track of pairwise wins and losses in sufficiently many sampled elections, we can identify approximate exclusion zones with high probability.

**Theorem 4.** Let  $G$  be a graph with  $n$  nodes and  $m$  edges, and pick any desired  $\epsilon, \delta \in (0, 1)$ . There is a randomized algorithm returning a set  $S$  in time  $O((n^3 + n^2m) \log(1/\delta)/\epsilon^2)$  such that:

1.  $S$  is a subset of the minimal IRV exclusion zone of  $G$ , and
2.  $S$  is a  $(1 - \epsilon)$ -approximate IRV exclusion zone with probability at least  $1 - \delta$ .

We can also show that the computational hardness arises specifically from checking whether small node sets are exclusion zones. Intuitively, for small node sets, there are exponentially many candidate configurations (in the number of nodes outside the set) which could be counterexamples. Formally, we show that IRV-EXCLUSION is fixed-parameter tractable for the parameter  $|V \setminus S|$ .

**Theorem 5.** Let  $G$  be a graph with  $n$  nodes and  $m$  edges. For  $|S| = n - c$ , there is an algorithm for IRV-EXCLUSION with runtime  $O(2^c n(n + m))$ .

For MIN-IRV-EXCLUSION, we can test progressively larger node sets with the algorithm from Theorem 5 (see the extended version for details). We did this to find the minimal IRV exclusion zones of all connected graphs on 3–7 nodes and all trees on 3–15 nodes. Some examples are shown in Figure 3. In the extended version, we list the number of graphs and trees with nontrivial and 2-node IRV exclusion zones. (No graph can have a one-node exclusion zone, since any one node can be the first eliminated by a tiebreak when there is a candidate at every node.) We find that the vast majority of small graphs and trees have nontrivial exclusion zones, indicating that it is common for small graphs to have sets of nodes that are easily excluded under IRV.

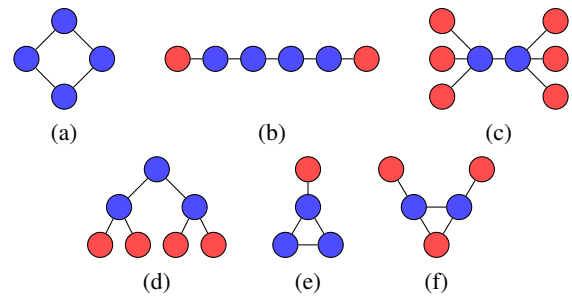


Figure 3: Some graphs with their minimal IRV exclusion zones in blue and excluded nodes in red: (a) the 4-cycle, (b) the 6-path, (c) the 6-leaf bistar, (d) the height-2 perfect binary tree, (e) the smallest connected cyclic graph with a nontrivial IRV exclusion zone, and (f) the smallest (in nodes, then in edges) connected graph whose minimal IRV exclusion zone does not consist of all non-leaf nodes.

## Graph Families With Known IRV Exclusion Zones

While finding exact IRV exclusion zones in graphs is hard in general, we can identify them in some families of graphs. First, in any graph where every pairwise contest is a tie (i.e., every node is a weak Condorcet position), the minimal exclusion zone is trivial by Proposition 4. Such graphs include complete graphs, cycles, and all other distance-regular graphs (van Dam, Koolen, and Tanaka 2016). We also show that paths, bistars, and even-height perfect binary trees have nontrivial IRV exclusion zones. The bistar graph, consisting of two star graphs of equal size whose centers are joined by an edge, is the simplest example of a graph with the smallest possible IRV exclusion zone: two nodes.

**Proposition 8.** The minimal IRV exclusion zone of a bistar graph consists of its two hub nodes.

The result of Tomlinson, Ugander, and Kleinberg (2024) on uniform 1-Euclidean preferences extends naturally to path graphs, although with some additional messiness caused by discretizing the interval. Recall that a path is a graph with edges  $\{1, 2\}, \{2, 3\}, \dots, \{n - 1, n\}$ .

**Proposition 9.** The minimal IRV exclusion zone of the path on  $n$  nodes is  $S = \{\lceil n/6 + 1/2 \rceil, \dots, n - \lceil n/6 + 1/2 \rceil + 1\}$ .

Next, we consider a case showing how exclusion zones can behave in unexpected ways: perfect binary trees. We show that perfect binary trees have nontrivial IRV exclusion zones if and only if they have even (and nonzero) height. This occurs because the root has an advantage in even-height trees, but can be tied by a set of leaves in odd-height trees.

**Theorem 6.** Perfect binary trees with odd height have no nontrivial IRV exclusion zones. However, for perfect binary trees with even height  $h > 0$ , the minimal IRV exclusion zone is the set of internal nodes.

## IRV Exclusion Zones in Real-World Graphs

We now ask whether real-world social networks, especially ones where we might expect distance-based preferences, have nontrivial IRV exclusion zones. To this end, we use a

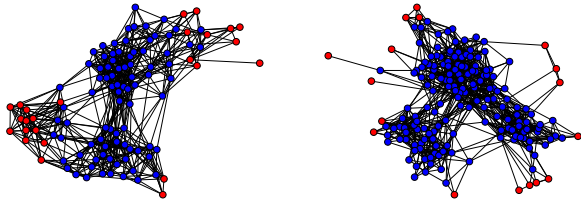


Figure 4: Two of the school social networks from Paluck, Shepherd, and Aronow (2020) (ID 5 on the left and 50 on the right) with probable 0.99-approximate IRV exclusion zones in blue and excluded nodes in red. Nodes that cannot win under IRV can include fringe nodes or entire communities.

collection of 56 social networks from public middle schools in New Jersey (Paluck, Shepherd, and Aronow 2016, 2020), where edges represent pairs of students who spend time with each other. If IRV were used for a class president election under graph-based preferences, for instance, any exclusion zone in the graphs would tell us which nodes have a chance of winning and which could not win any election against a member of the exclusion zone. The graphs in this dataset have 110–844 nodes, making them much too large for the exact algorithm we applied in the exhaustive search of small graphs. Instead, we apply our approximation algorithm from Theorem 4 with parameters  $\epsilon = \delta = 0.01$  (see the extended version for additional experiment details). We find that 23 of the 56 schools verifiably have no nontrivial IRV exclusion zones, owing to the fact that the algorithm returned the full node set (and any set returned by the algorithm must be a subset of the minimal exclusion zone). The remaining 33 school graphs (59%) returned nontrivial probable approximate exclusion zones. Even when they are nontrivial, these (probable, approximate) exclusion zones tend to be large, comprising 92% of the graph on average. See Figure 4 for visualizations of two of the smaller schools, with the node sets returned by the approximation algorithm in blue.

In combination with the exhaustive search of small graphs and trees, these experiments on real-world networks demonstrate that nontrivial IRV exclusion zones frequently exist in practice. However, for real-world networks, they tend to be quite large, revealing only a small fraction of nodes that cannot win under IRV. Our approximation algorithm also allows us to explore IRV exclusion zones in graphs orders of magnitude larger than the exact algorithm.

## Discussion

Exclusion zones are a relatively new addition to the landscape of voting theory, and their structure is still not well understood. Following initial results on one-dimensional IRV and plurality, we have built up a much more general understanding of exclusion zones, answered some of the questions left open by prior work, and provided new results about the complexity of computing exclusion zones.

In particular, we have characterized structural properties and resolved an open question about exclusion zones of IRV in higher-dimensional preference spaces, showing that

with uniform voters over any  $d$ -dimensional hyperrectangle with  $d > 1$ , there are no nontrivial IRV exclusion zones. With graph-based preferences, we showed that IRV exclusion zones are computationally hard to identify, but also provided an efficient approximation algorithm. Through computational experiments, we have also discovered that nontrivial exclusion zones are abundant in small graphs, and likely also present in many larger graphs arising in social network data. While our focus has been on IRV, the ideas we develop around exclusion zones, such as the nesting of exclusion zones and the Condorcet Chain Lemma, apply to any voting system over any metric space. This suggests that exclusion zones may aid in understanding other voting systems, a promising direction for future work.

The major open questions from our work concern which voting rules have nontrivial exclusion zones in which preference spaces. In the 1-Euclidean domain, we already understand the exclusion zones of IRV, plurality, and Condorcet methods. Do other non-Condorcet methods, like approval, Borda, Coombs, and top-two runoff, have nontrivial exclusion zones with 1-Euclidean preferences? In higher dimensions, are there election sequences satisfying the Condorcet Chain Lemma for any of these other non-Condorcet methods? Whenever there are weak Condorcet positions, the minimal exclusion zone of any Condorcet method is the set of such positions, but there may still be something interesting to say about the particular nesting of all exclusion zones with Condorcet methods. For instance, in one dimension, the set of exclusion zones of Condorcet methods captures the essence of the Median Voter Theorem. When there is a Condorcet winning position in higher-dimensional preference spaces (as with uniform voters over hyperrectangles), is there something we can say about the set of exclusion zones of Condorcet voting rules?

In terms of improvements in proof techniques, the Condorcet Chain Lemma can only show that the minimal exclusion zone is trivial in preference spaces where there are both Condorcet and anti-Condorcet positions. In cases with no Condorcet positions, is there another way to certify triviality? Note that there do exist preference spaces with no Condorcet positions and a trivial minimal exclusion zone: in work addressing different questions than exclusion zones, Skibski (2023, Figure 2) shows a 12-node graph with no Condorcet position, and we verified using our optimized exhaustive search that it has no nontrivial IRV exclusion zone. In the opposite case, when there are nontrivial exclusion zones, our main proof approach with IRV has exploited the structure of the voting system: to argue that  $S$  is an exclusion zone, we show that the last candidate in  $S$  cannot be eliminated by candidates outside of  $S$ . This takes advantage of the fact that IRV eliminates candidates one at a time, so if the election began with at least one candidate in  $S$ , then at some point only one will remain in  $S$ . However, as we have seen in one dimension, it is possible to argue through other means that voting systems like plurality and Condorcet methods have nontrivial exclusion zones, so we are optimistic that exclusion zones will prove useful beyond IRV in the ongoing endeavor to characterize voting systems.

## Acknowledgments

This work was supported in part by ARO MURI, a Simons Collaboration grant, a grant from the MacArthur Foundation, a Vannevar Bush Faculty Fellowship, AFOSR grant FA9550-19-1-0183, and NSF CAREER Award #2143176. Thanks to Kate Donahue, Jason Gaitonde, Raunak Kumar, Sloan Nietert, Katherine Van Koeving, and the attendees of the AMS Special Session on Mathematics of Decisions, Elections, and Games at the 2025 Joint Mathematics Meetings for helpful discussions and feedback.

## References

- Anagnostides, I.; Fotakis, D.; and Patsilina, P. 2022. Dimensionality and coordination in voting: The distortion of STV. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 4776–4784.
- Anshelevich, E.; Bhardwaj, O.; Elkind, E.; Postl, J.; and Skowron, P. 2018. Approximating optimal social choice under metric preferences. *Artificial Intelligence*, 264: 27–51.
- Anshelevich, E.; Bhardwaj, O.; and Postl, J. 2015. Approximating optimal social choice under metric preferences. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 777–783.
- Bandelt, H.-J. 1985. Networks with Condorcet solutions. *European Journal of Operational Research*, 20(3): 314–326.
- Black, D. 1948. On the rationale of group decision-making. *Journal of Political Economy*, 56(1): 23–34.
- Bloks, S. 2018. *Condorcet Winning Sets*. Ph.D. thesis, London School of Economics and Political Science. Accessed 2/16/2025.
- Brandes, U.; Laßmann, C.; and Rothe, J. 2022. Voting for centrality. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, 1554–1556.
- Charikar, M.; Lassota, A.; Ramakrishnan, P.; Vetta, A.; and Wang, K. 2025. Six candidates suffice to win a voter majority. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, 1590–1601.
- Charikar, M.; Ramakrishnan, P.; and Wang, K. 2025. Approximately Dominating Sets in Elections. arXiv:2504.20372.
- Delemazure, T.; and Peters, D. 2024. Generalizing Instant Runoff Voting to Allow Indifferences. In *Twenty-Fifth ACM Conference on Economics and Computation (EC'24)*.
- Elkind, E.; Faliszewski, P.; Laslier, J.-F.; Skowron, P.; Slinko, A.; and Talmon, N. 2017. What do multiwinner voting rules do? An experiment over the two-dimensional Euclidean domain. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.
- Elkind, E.; Lackner, M.; and Peters, D. 2025. Preference Restrictions in Computational Social Choice: A Survey. arXiv:2205.09092.
- Elkind, E.; Lang, J.; and Saffidine, A. 2015. Condorcet winning sets. *Social Choice and Welfare*, 44(3): 493–517.
- Gkatzelis, V.; Halpern, D.; and Shah, N. 2020. Resolving the optimal metric distortion conjecture. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, 1427–1438. IEEE.
- Hansen, P.; Thisse, J.-F.; and Wendell, R. E. 1986. Equivalence of solutions to network location problems. *Mathematics of Operations Research*, 11(4): 672–678.
- Kizilkaya, F. E.; and Kempe, D. 2022. PluralityVeto: A Simple Voting Rule Achieving Optimal Metric Distortion. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, 349–355.
- McGann, A. J.; Grofman, B.; and Koetzle, W. 2002. Why party leaders are more extreme than their members: Modeling sequential elimination elections in the US House of Representatives. *Public Choice*, 113(3): 337–356.
- Merrill III, S. 1984. A comparison of efficiency of multi-candidate electoral systems. *American Journal of Political Science*, 23–48.
- Mollison, D. 2023. Fair votes in practice. arXiv:2303.15310.
- Nguyen, T.; Song, H.; and Lin, Y.-S. 2025. A few good choices. arXiv:2506.22133.
- Paluck, E. L.; Shepherd, H.; and Aronow, P. M. 2016. Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3): 566–571.
- Paluck, E. L.; Shepherd, H. R.; and Aronow, P. 2020. Dataset for: Changing Climates of Conflict: A Social Network Experiment in 56 Schools, New Jersey, 2012-2013.
- Plott, C. R. 1967. A notion of equilibrium and its possibility under majority rule. *The American Economic Review*, 57(4): 787–806.
- Schofield, N. 1983. Generic instability of majority rule. *The Review of Economic Studies*, 50(4): 695–705.
- Skibski, O. 2023. Closeness centrality via the Condorcet principle. *Social Networks*, 74: 13–18.
- Skowron, P.; and Elkind, E. 2017. Social choice under metric preferences: Scoring rules and STV. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.
- Telek, A. 2016. Power networks: A network approach to voting theory.
- Tomlinson, K.; Ugander, J.; and Kleinberg, J. 2024. The Moderating Effect of Instant Runoff Voting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 9909–9917.
- Tomlinson, K.; Ugander, J.; and Kleinberg, J. 2025. Exclusion Zones of Instant Runoff Voting. arXiv:2502.16719.
- van Dam, E. R.; Koolen, J. H.; and Tanaka, H. 2016. Distance-Regular Graphs. *The Electronic Journal of Combinatorics*, #DS22.
- Wendell, R. E.; and McKelvey, R. D. 1981. New perspectives in competitive location theory. *European Journal of Operational Research*, 6(2): 174–182.