

Likelihood of the Existence of Average Justified Representation

Qishen Han
Rutgers University
hnickc2017@gmail.com

Biaoshuai Tao
Shanghai Jiao Tong University
bstao@sjtu.edu.cn

Lirong Xia
Rutgers University
xialirong@gmail.com

Chengkai Zhang
Rutgers University
chengkai.zhang@rutgers.edu

Houyu Zhou
UNSW Sydney
houyu.zhou@unsw.edu.au

Abstract

We study the approval-based multi-winner election problem where n voters jointly decide a committee of k winners from m candidates. We focus on the axiom *average justified representation* (AJR) proposed by Fernández, Elkind, Lackner, García, Arias-Fisteus, Basanta-Val, and Skowron (2017). AJR postulates that every group of voters with a common preference should be sufficiently represented in that their average satisfaction should be no less than their Hare quota. Formally, for every group of $\lceil \ell \cdot \frac{n}{k} \rceil$ voters with ℓ common approved candidates, the average number of approved winners for this group should be at least ℓ . It is well-known that a winning committee satisfying AJR is not guaranteed to exist for all multi-winner election instances. In this paper, we study the likelihood of the existence of AJR under the Erdős-Rényi model. We consider the Erdős-Rényi model parameterized by $p \in [0, 1]$ that samples multi-winner election instances from the distribution where each voter approves each candidate with probability p (and the events that voters approve candidates are independent), and we provide a clean and complete characterization of the existence of AJR committees in the case where m is a constant and n tends to infinity. We show that there are two phase transition points p_1 and p_2 (with $p_1 \leq p_2$) for the parameter p such that: 1) when $p < p_1$ or $p > p_2$, an AJR committee exists with probability $1 - o(1)$, 2) when $p_1 < p < p_2$, an AJR committee exists with probability $o(1)$, and 3) when $p = p_1$ or $p = p_2$, the probability that an AJR committee exists is bounded away from both 0 and 1.

1 Introduction

In an approval-based multi-winner election, n voters jointly decide a committee of k winners from m candidates by casting approval ballots (i.e., each voter's ballot contains a binary string of length m indicating whether or not she approves each candidate). The set of k winners is called the winning committee. Approval-based multi-winner election has been extensively studied in the past literature (see the recent book by Lackner and Skowron [16] for a comprehensive survey). Fairness in multi-winner approval voting is crucial for democratic governance, ensuring that diverse voter preferences are adequately reflected in elected committees. Among those fairness notions, *proportional representation* is arguably the most important consideration, and it postulates that every party of the voters should receive a number of seats proportional to its population. For example, to proportionally represent a party of $n/2$ voters, at least $k/2$ candidates approved by them should be selected as winners. For elections where voters are not pre-specified into parties/groups,

recent work has been considering *justified representation* (JR) and its variants [1] which generalize the concept of proportional representation. Along this line of fairness axioms, the notion of “party” is naturally generalized to the notion of *cohesive group*—a group of voters with similar preferences reflected by their approval ballots, and the winning committee should ensure that every cohesive group that justifies ℓ Hare quotas should somehow have a satisfaction level of ℓ in some sense. Here, the Hare quota [21], first proposed by Alexander Hamilton for use in United States congressional apportionment, is defined as n/k —the minimum number of voters in a group that justifies one seat in the winning committee.

Formally, we say that a set of voters V is an ℓ -cohesive group if $|V| \geq \ell \cdot \frac{n}{k}$ and voters in V approve at least ℓ candidates in common. To formulate what it means by saying an ℓ -cohesive group has a satisfaction level of ℓ , different criteria have been proposed that yield different variants of JR. Naturally, the satisfaction of a single voter can be defined as the number of candidates in the winning committee that she approves. It is then natural to define the satisfaction of an ℓ -cohesive group to be the average satisfaction of the voters in this group. As an ℓ -cohesive group deserves ℓ quotas in the spirit of proportional/justified representation, we would like to require that **every ℓ -cohesive group has the average satisfaction at least ℓ** . This is exactly the notion of *average justified representation* (AJR) proposed by Fernández et al. [11]. The formal mathematical definition of AJR is deferred to Definition 2.1.

While AJR is a natural axiom aligning with the concept of proportional/justified representation, a winning committee satisfying AJR may not exist, as demonstrated in the example in Fig. 1. Subsequent work by Aziz et al. [2], Skowron [23] further explored this notion, showing that the well-known *Proportional Approval Voting* (PAV) rule guarantees an average satisfaction of $\ell - 1$ for every ℓ -cohesive group,¹ one below the proportional ideal of ℓ .

In addition, those worst-case scenarios where AJR committees fail to exist may be rare in practice. Indeed, empirical evidence suggests that AJR might often hold: Brill et al. [7, 9] studies a stronger representation axiom (requiring every voter in an ℓ -cohesive group to individually approve at least ℓ winners). Their experiments, across multiple probabilistic models, show that for large enough committee sizes k , there exist winning committees satisfying this stronger condition (that implies AJR). This disparity between worst-case impossibility result and empirical results motivates a theoretical investigation of AJR under probabilistic assumptions.

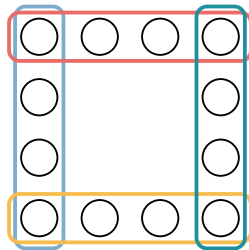


Figure 1: An example where AJR committees do not exist. In this example, each black circle stands for a voter. The colored rectangles stand for candidates. If a circle is in a rectangle, it means that this voter approves this candidate. In this example, $n = 12$ and $m = 4$, and we are to select a winning committee of size three (i.e., $k = 3$). In this case, there are four 1-cohesive groups, each corresponding to one side of the large square. Selecting any three of these four candidates will lead to some group’s average satisfaction being only 0.5, which is less than the requirement 1.

¹More precisely, PAV guarantees an average satisfaction of $\ell - 1 + o(1/k)$. Moreover, for every constant $c > 0$, no rule can guarantee average satisfaction at least $(\ell - 1 + c)$ in general. Hence, we refer to $\ell - 1$ as its average satisfaction for simplicity.

We initiate the study of the likelihood of AJR by using a simple, well-established model in multi-winner voting: *the Erdős–Rényi bipartite random graph* where each voter approves each candidate independently with probability p (see, e.g., Peters et al. [19], Szufa et al. [24], Brill et al. [7], Elkind et al. [10]). This model captures a neutral “base case” where voter preferences lack inherent structure—a natural starting point for theoretical analysis. We study the following problem in this paper.

Under what values of p does an AJR committee exist with high probability? Conversely, when is an AJR committee unlikely to exist?

Motivated by that the number of candidates m is typically small while the number of voters n can be very large (e.g., a political election in a country), we assume m is a constant (and so is the size of the winning committee k) and study the above problem by analyzing the probabilities asymptotically with $n \rightarrow \infty$. Under the Erdős–Rényi model, AJR holds trivially at the extremes: when $p = 1$, all committees satisfy AJR; when $p = 0$, no cohesive group exists, so every winning committee automatically satisfies AJR. Naturally, we may expect that AJR committees exist with probability $1 - o(1)$ for small enough p (in which case no cohesive group exists with high probability) and large enough p (in which case, with high probability, every cohesive group has a high enough satisfaction for an arbitrary winning committee). The core challenge lies in the intermediate regime: is there any intermediate regime of p where AJR committees are unlikely to exist? Does the likelihood for the existence of AJR committees change smoothly as p changes? Or, are there any phase transition points? If so, how many?

On Other Justified Representation Variants. Many variants along the line of justified representation that are weaker than AJR have been proposed, with the most notable examples of *JR* [1], *extended JR* (EJR) [2], and *proportional JR* (PJR) [11, 6, 8]. For EJR, it is required that every ℓ -cohesive group contains *one* voter with a satisfaction of at least ℓ . JR is an even weaker notion than EJR with the same requirement, except that only 1-cohesive groups are concerned. PJR is also a weaker notion than EJR that requires each ℓ -cohesive group must have a *total* satisfaction of at least ℓ (or equivalently, an average satisfaction of $\ell/|V|$ for every ℓ -cohesive group V). All these three axioms are guaranteed to be satisfiable, and, for each of them, a winning committee satisfying the axiom can be found in polynomial time [1, 3, 20]. It is easy to see that EJR (and thus JR) is substantially weaker than AJR: it only ensures the satisfaction of *one* voter in each ℓ -cohesive group (instead of the *average* satisfaction of the whole group) being at least ℓ . PJR, on the other hand, places requirements with respect to average satisfactions, but it is even weaker than EJR, as mentioned above. In Example 1.1, we demonstrate that the requirement postulated by EJR (and thus PJR and JR) can be very lenient that admits both natural winning committees and not-so-natural committees, while the requirement of AJR refines the set of valid winning committees by only keeping those natural ones.

Example 1.1. Consider a voting instance with $n = k^2$ and $m = 2k$ where k is the size of the winning committee. The voters are indexed by $\{v_{ij}\}_{i=1,\dots,k;j=1,\dots,k}$. There are k “row” candidates $r_1, \dots, r_k$ and k “column” candidates $c_1, \dots, c_k$. Each voter v_{ij} approves exactly two candidates: r_i and c_j .

In this instance, each of the $2k$ candidates “induces” a 1-cohesive group, i.e., each “row” and each “column” of $\{v_{ij}\}_{i=1,\dots,k;j=1,\dots,k}$ form a 1-cohesive group. It is easy to see that any set of k candidates gives an EJR (and thus PJR and JR) committee. In this case, the satisfactions for some of the voters can be 0. On the other hand, there are only two AJR committees: $\{r_1, \dots, r_k\}$ and $\{c_1, \dots, c_k\}$. In any of the two AJR committees, every voter’s satisfaction is 1.

In addition to those JR variants mentioned above that are weaker than AJR, other JR variants that are incomparable to AJR have also been proposed. For example, Peters et al. [20] introduced full JR (FJR), which considers not only cohesive groups but also near-cohesive ones. Building on this idea, Kalayci et al. [15] extended PJR to that setting and proposed full PJR (FPJR). Brill and Peters [5] introduced EJR+ and PJR+, which offer an alternative weakening of cohesiveness. The relationship between these notions is given in Fig. 2.

All these variants of JR mentioned above are *yes-or-no* measurements, i.e., a winning committee either satisfies the requirement or not. Previous literature has also proposed many *quantitative* measurements on the fairness of a winning committee, such as *proportionality degree* [2, 23, 13] and *EJR degree* [25]. Roughly speaking, given a winning committee, its proportionality degree is the average satisfaction for the cohesive group with the least average satisfaction (a more formal definition is in the next paragraph), and its EJR degree is the minimum (taking over all cohesive groups) number of voters in an ℓ -cohesive group whose satisfactions are at least ℓ . The notion of EJR degree is a quantitative generalization of EJR: instead of just *requiring* at least *one* voter in an ℓ -cohesive group with satisfaction at least ℓ , this notion further investigates if “one” is the best number possible and tries to *optimize* this number.

The notion of proportionality degree, on the other hand, is more aligned with the notion of AJR by putting the average satisfaction as an objective to be maximized. More formally, given a function $f : \mathbb{Z}^+ \rightarrow \mathbb{R}_{\geq 0}$, the proportionality degree of a winning committee is f if the average satisfaction of each ℓ -cohesive group is at least $f(\ell)$. In terms of this, AJR exactly requires a proportionality degree of $f(\ell) = \ell$, which, as we have seen in Fig. 1, may not always be satisfiable. As we have also mentioned before, the PAV rule can guarantee a proportionality degree of $f(\ell) = \ell - 1$. Nevertheless, we would also like to remark that the particular choice $f(\ell) = \ell$ as the requirement in the definition of AJR has its special meaning: it faithfully reflects the idea of proportional representation as ℓ is exactly the Hare quota for an ℓ -cohesive group.

Core Stability. Along the line of defining fairness based on proportional representation, another criterion is *core stability* [12, 22] which has also been studied widely [14, 18, 17, 27]. In the context of multi-winner election, a winning committee W of size k is *core stable* if there is no *blocking coalition*, where a subset V of voters is said to form a blocking coalition if there is another winning committee W' of size $\lfloor k \frac{|V|}{n} \rfloor$ such that every voter in V has a strictly higher satisfaction for W' compared to W . Consider a group of voters V with size $|V| = \ell \cdot \frac{n}{k}$. Instead of imposing a constraint ℓ on the satisfaction of V like in those JR notions, core stability requires that there is no way to select ℓ candidates such that every voter in V is strictly happier compared to the current winning committee W . Notice also that there is no requirement that those voters in V must approve ℓ common candidates (i.e., be an ℓ -cohesive group). Unlike the case with AJR that is known to be unsatisfiable for some instances, it is an open problem if a core stable committee always exists.

AJR and core stability do not imply each other (see Fig. 2 where we have also included core stability). The example in Fig. 1 shows that a core stable committee may not satisfy AJR: it is easy to verify that every set of 3 candidates forms a core stable winning committee, while we have seen that no AJR committee exists in this instance. To see that an AJR committee may not be core stable, consider an instance with $n = 8$, $k = 4$, and candidates $\{x_1, x_2, x_3, x_4, y, z\}$. Voters 1, 2, and 3 approve $\{x_1, y\}$, voters 4, 5, and 6 approve $\{x_1, z\}$, and voters 7 and 8 approve $\{x_2, x_3, x_4\}$. Committee $W = \{x_1, x_2, x_3, x_4\}$ satisfies AJR. However, it does not satisfy core stability, as the set of voters $\{1, 2, 3, 4, 5, 6\}$ can deviate to $W' = \{x_1, y, z\}$.

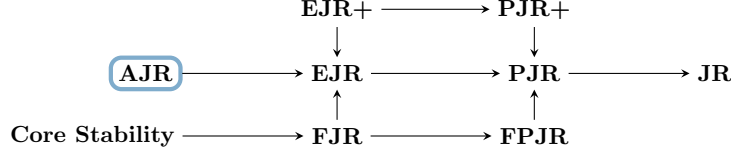


Figure 2: Diagram of relationships among different variants of JR. Arrows indicate implications.

2 Overview of Contributions

For an integer m , we use $[m]$ to denote the set $\{1, 2, \dots, m\}$.

Approval-Based Multi-Winner Election. There are n voters who cast approval votes among a group of candidates $M = [m]$. We use i to denote a generic candidate and v to denote a generic voter. Each voter can approve an arbitrary number of candidates. For a voter v , we use $A(v)$ to denote the set of candidates approved by v , i.e., approval ballot of voter v . The goal is to elect a winner *committee* of size k , denoted by W .

Definition 2.1 (Average Justified Representation (AJR)). *A winning committee W of size $|W| = k$ provides AJR if for every $\ell \in [k]$ there does not exist a group of voters V satisfying*

1. *Size Constraint:* $|V| \geq \ell \cdot \frac{n}{k}$;
2. *Cohesiveness Constraint:* $|\bigcap_{v \in V} A(v)| \geq \ell$;
3. *Underrepresented Constraint:* $\sum_{v \in V} |A(v) \cap W| < \ell \cdot |V|$.

For each ℓ , we call a set/group of voters V an ℓ -cohesive group towards L if $|V| \geq \ell \cdot \frac{n}{k}$, $L \subseteq |\bigcap_{v \in V} A(v)|$, and $|L| = \ell$ (corresponding to the first two constraints in the definition above). We call an ℓ -cohesive group V *underrepresented* if it satisfies Constraint (3) (in addition to satisfying (1) and (2)). Let $U_v(W) = |A(v) \cap W|$ be the *utility* of voter v achieved by the winning committee W , $U(V, W) = \sum_{v \in V} |A(v) \cap W|$ be the total utility of V achieved by W , and $u(U, V) = \frac{1}{|V|} U(V, W)$ be the average utility of V achieved by W . Then, an ℓ -cohesive group V is underrepresented if the average utility of V is less than ℓ .

Erdős–Rényi Bipartite Model. The approval ballot $A(v)$ of each voter v is generated i.i.d.. A voter v has a probability of $p \in [0, 1]$ to approve each candidate $i \in M$. The events of v approving different candidates are independent. Therefore, the probability that v 's approval ballot is $A(v)$ is $p^{|A(v)|} \cdot (1 - p)^{m - |A(v)|}$. We use π to denote the distribution of $A(v)$.

2.1 Our Results

We fully characterize the existence of an AJR committee under the Erdős–Rényi bipartite model. This paper presents the following result.

Theorem 2.2. *For any constant m , k , and p with $m > k \geq 2$ and $p \in [0, 1]$, and for $n \rightarrow \infty$, the likelihood that there exists a committee W that provides AJR has the following trichotomy. Let $p_1^* = \frac{1}{k}$ and p_2^* be the maximum $x \in [0, 1]$ such that*

$$k(2x(1 - x)^k + kx^2(1 - x)^{k-1}) = 1. \quad (1)$$

Then the likelihood that an AJR committee exists is

1. $1 - o(1)$ if $p < p_1^*$ or $p > p_2^*$;
2. $o(1)$ if $p_1^* < p < p_2^*$;
3. $\Theta(1)$ and $1 - \Theta(1)$ (i.e., bounded away from 0 and 1) if $p = p_1^*$ or $p = p_2^*$.

As the first remark, the case with $k = 1$, which is excluded from Theorem 2.2, is trivial: an AJR committee always exists. To see this, in the case $k = 1$, the only possible cohesive group is the set of all voters. If they approve a common candidate, then selecting this candidate to be the winning committee gives an AJR committee. Otherwise, there is no cohesive group at all, in which case an arbitrary winning committee automatically achieves the AJR requirement.

Theorem 2.2 captures all the remaining cases. The case for $k = 2$ is also special, Equation (1) only has one solution $x = \frac{1}{2}$, in which case $p_1^* = p_2^*$. In this case, case (2) of the theorem never applies, and case (3) applies only when $p = \frac{1}{2}$. For general cases $k \geq 3$, the following proposition shows that $p_2^* > p_1^* = \frac{1}{k}$ and p_2^* is of the same order of $\frac{1}{k}$. Notice that the upper bound $\frac{5}{k}$ is not tight, but it shows that $p_2^* \rightarrow 0$ as $k \rightarrow \infty$.

Proposition 2.3. *For $k \geq 3$, we have $\frac{1}{k} < p_2^* < \min\{1, \frac{5}{k}\}$ for p_2^* defined in Theorem 2.2.*

The proof of this proposition uses some observations in the proof of Theorem 2.2, and is deferred to Sect. 4.3.

Therefore, for $k \geq 3$, Theorem 2.2 identifies two phase-transition points for p , namely, p_1^* and p_2^* . For $p < p_1^*$ and $p > p_2^*$, an AJR committee exists with high probability. For $p_1^* < p < p_2^*$, no AJR committee exists with high probability. At the two phase-transition points, the probability for the existence of an AJR committee is bounded away from 0 and 1. It does not come as a surprise that an AJR committee exists with high probability for p being either too small or too large. Intuitively, for very small p , it is likely that there is no cohesive group at all, so every winning committee automatically satisfies AJR; for very large p , every candidate is approved by a large fraction of voters, and the AJR condition is satisfied for every committee of k candidates. It is much less intuitively clear what the story is for p being neither very small nor very large, and our Theorem 2.2 states that the AJR committee is unlikely to exist for p falling into a *single* intermediate interval. In addition, the length of this intermediate interval decreases to 0 (see Proposition 2.3) as $k \rightarrow \infty$.

In Fig. 3, we plot the values of p_1^* and p_2^* for $k = 2, \dots, 10$.

2.2 Comparisons with Other Work

Our result theoretically verifies the empirical findings in Brill et al. [7, 9]: for large k , most values of p would guarantee the existence (with high probability) of AJR committees. Moreover, our result provides a fine-grained characterization of the existence of AJR committees by precisely describing the relationship between p and the likelihood of AJR committees.

Empirical studies for EJR, PJR, and JR (discussed in Sect. 1) have also been performed, with the main focus on the *number* of valid committees (since we know valid committees always exist for these notions). For EJR and PJR, under the Erdős–Rényi model, Bredereck et al. [4] shows that the number of valid committees goes down and then up as p increases, and, for JR, Elkind et al. [10] shows that this number is constantly $\binom{m}{k}$ (i.e., all committees are valid) for $n \rightarrow \infty$. Our theoretical result is similar in that all committees satisfy AJR for small or large p . The difference is that, for values of p that are neither too small nor too large, Bredereck et al. [4] shows that the number of EJR committees goes down, while ours shows that no AJR committee exists at all. This makes a distinction between AJR and EJR/PJR/JR. Notice also that both Bredereck et al. [4] and

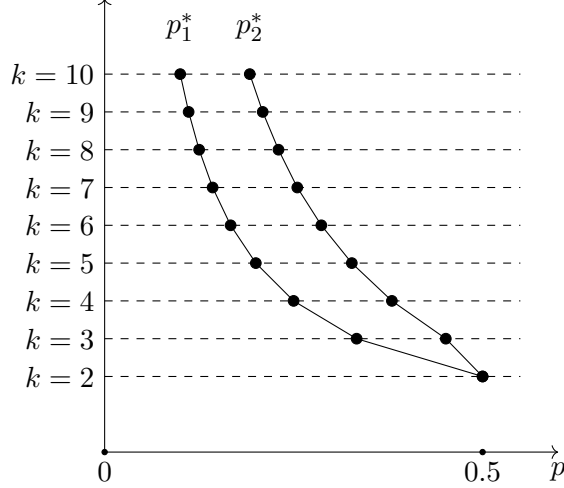


Figure 3: The values of p_1^* and p_2^* for $k = 2, \dots, 10$.

Elkind et al. [10] also consider the setting where fewer than k candidates are selected and whether those JR notions are still satisfiable, which is not the focus of this paper.

Theoretically, Xia [27] studies the similar problem on the likelihood of existence of committees satisfying these axioms. Our result for AJR makes a sharp contrast with Xia’s result on core stability. In Xia’s setting, each candidate c_j is independently approved with probability p_j , and Xia finds that a committee satisfying core stability is very likely to exist regardless of the choice of p_j . In our work, we consider a more specialized model in which each candidate is approved independently with the same probability p . Despite that our model is the symmetric special case, we show that there is a regime where AJR does not exist with high probability, suggesting that AJR may be harder to satisfy than CS. For example, for some natural parameters such as $m = 10$, $k = 4$, and $p = 1/3$, for sufficiently large n , almost every committee satisfies core stability, while none of them satisfies AJR.

2.3 Structure of Our Paper

The remaining part of our paper is organized as follows. In Sect. 3, we provide an intermediate characterization (Lemma 3.1) that characterizes the phase-transition points for p for the existence of AJR committees. The characterization is cumbersome but more intuitive for analysis based on Hoeffding’s inequality. In addition, the characterization itself does not have a closed form that identifies the number of phase-transition points. In Sect. 4, we investigate the characterization obtained in Sect. 3 and show how it implies the clean characterization in our main result in Theorem 2.2. The corner case where p is exactly at one of the phase-transition points is analyzed in a completely different way. In Sect. 5, we discuss the polyhedron approach proposed by Xia [26] and show how it can be applied to analyze the corner case when p is at a phase-transition point. Finally, we conclude our paper in Sect. 6.

3 An Intermediate Characterization

In this section, we provide an intermediate characterization (Lemma 3.1) on the likelihood of AJR. This characterization will eventually lead to Theorem 2.2, which is discussed in Sect. 4.

To introduce the intermediate characterization, we begin by describing some very high-level ideas. Firstly, the likelihood of the existence of an AJR committee is $1 - o(1)$ for p being sufficiently small or sufficiently large. For small p , with high probability, there is no cohesive group at all, so every winning committee W automatically satisfies AJR. For large p , every winning committee W is likely to be approved by many voters, making the average utility for every cohesive group large, in which case AJR is also satisfied.

Identifying the threshold for the existence of cohesive groups. It is easy to find the threshold for p below which no cohesive group exists. Consider a set L of ℓ candidates. Each voter approves all the candidates in L with probability p^ℓ . Thus, the expected number of voters approving all candidates in L is np^ℓ . When $p < \sqrt[\ell]{\frac{\ell}{k}}$, we have $np^\ell < \ell \cdot \frac{n}{k}$. By Hoeffding's inequality, the probability that there exists a group of $\ell \cdot \frac{n}{k}$ voters that approve every candidate in L is $\Theta(\exp(-n))$. By applying a union bound over all those $\binom{m}{\ell}$ sets of ℓ candidates and noticing that $\binom{m}{\ell}$ is a constant, the probability an ℓ -cohesive group exists is $o(1)$. On the other hand, when $p > \sqrt[\ell]{\frac{\ell}{k}}$, we have $np^\ell > \ell \cdot \frac{n}{k}$. By Hoeffding's inequality, the number of voters approving all candidates in L is at least $\ell \cdot \frac{n}{k}$ with probability $1 - \Theta(\exp(-n))$, in which case ℓ -cohesive groups exist with high probability. Therefore, the value $\sqrt[\ell]{\frac{\ell}{k}}$ is the threshold for the existence of ℓ -cohesive groups. Notice that this value minimizes at $\ell = 1$. The threshold for the existence of cohesive groups (with all ℓ 's) is $\frac{1}{k}$. This is exactly where the value p_1^* in Theorem 2.2 comes from.

Analysis for the case where p is above the threshold $\sqrt[\ell]{\frac{\ell}{k}}$. Following our analysis in the previous paragraph, when $p \geq \sqrt[\ell]{\frac{\ell}{k}}$, for every set L of ℓ candidates, there is at least $\Theta(1)$ probability that an ℓ -cohesive group towards L exists. To characterize how large the value p must be to guarantee the existence of an AJR committee, let us consider the worst-case scenario where $W \cap L = \emptyset$, i.e., consider every ℓ -cohesive group where none of its common approved candidates is selected in W . Intuitively, this is the “worst-case” as such a cohesive group is least satisfied and most likely to violate the AJR condition. The analysis with $W \cap L \neq \emptyset$ requires some extra analysis, and we skip it at this moment for the purpose of a more intuitive discussion. In the following, we fix $W = \{1, \dots, k\}$ and $L = \{k+1, \dots, k+\ell\}$.

Among those voters who approve L , every $\ell \cdot \frac{n}{k}$ of them form an ℓ -cohesive group. To make sure AJR is satisfied, we need that those $\ell \cdot \frac{n}{k}$ voters in the group with *minimum utilities* have an average utility of at least ℓ . To form such a group, we iteratively “pick” voters with minimum utilities among those voters who approve L . We first pick those voters with utility 0. These voters approve all candidates in L and no candidate in W . The expected number of this type of voters is $np^\ell(1-p)^k$. Next, we pick those voters with utility 1. These are the voters who approve all candidates in L and exactly one candidate in W . The expected number of them is $np^\ell \cdot \binom{k}{1} p(1-p)^{k-1} = \binom{k}{1} p^{\ell+1} (1-p)^{k-1} \cdot n$. In general, the expected number of voters who approve all candidates in L and exactly t candidates in W is $\binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n$. We keep this kind of greedy iterative selections until the total number of voters selected meets $\ell \cdot \frac{n}{k}$. This process will stop at some t_ℓ where the number of voters in the ℓ -cohesive group with utilities at most t_ℓ exceeds $\ell \cdot \frac{n}{k}$. More precisely, t_ℓ is the smallest integer such that

$$n_\ell := \sum_{t=0}^{t_\ell} \binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n \geq \ell \cdot \frac{n}{k}. \quad (2)$$

Here, n_ℓ is the expected number of voters in the cohesive group with utilities at most t_ℓ . Finally, we need to remove $n_\ell - \ell \cdot \frac{n}{k}$ voters from those voters with utility t_ℓ to make the number of voters exactly $\ell \cdot \frac{n}{k}$. This completes the selection of $\ell \cdot \frac{n}{k}$ voters in the ℓ -cohesive group with minimum utilities. Fig. 4 illustrates the selection and the notations t_ℓ and n_ℓ .

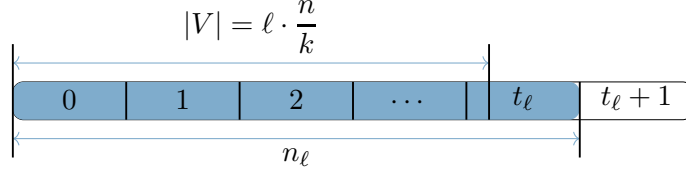


Figure 4: Illustration of construction group V . Numbers in the chunk represent the number of approved candidates in W for each chunk of voters.

Notice that the above analysis and the voters' distribution shown in Fig. 4 are all with respect to *expected* numbers of voters. However, given that the maximum value of ℓ (which is k) is a constant, we have a constant number of voter chunks in Fig. 4. By a combination of Hoeffding's inequality and union bounds, the actual voter distribution follows the description of Fig. 4 with high probability.

Lastly, we compute the expected average utility of those $\ell \cdot \frac{n}{k}$ selected voters. As mentioned before, the expected number of voters who approve all candidates in L and exactly t candidates in W is $\binom{k}{t} p^{\ell+t} (1-p)^{k-t} n$, which gives the expected total utility of $t \cdot \binom{k}{t} p^{\ell+t} (1-p)^{k-t} n$. The last set of voters with utility exactly t_ℓ is "truncated" such that $n_\ell - \ell \cdot \frac{n}{k}$ voters with utility t_ℓ are removed. Putting together, the expected average utility for those $\ell \cdot \frac{n}{k}$ agents approving L with minimum utilities is given by

$$u_\ell = \frac{k}{\ell \cdot n} \cdot \left(\sum_{t=0}^{t_\ell} t \cdot \binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n - t_\ell \cdot (n_\ell - \ell \cdot n/k) \right).$$

If we have a large enough p such that $u_\ell > \ell$ holds for all $\ell \in [k]$, then the least satisfied $\ell \cdot \frac{n}{k}$ voters who approve L meets the AJR requirement with high probability. Precisely, this probability is $1 - \Theta(\exp(-n))$ by Hoeffding's inequality, as $u_\ell > \ell$ implies that the *total* utility of these $\ell \cdot \frac{n}{k}$ voters is greater than $\ell \cdot \ell \cdot \frac{n}{k}$ by $\Theta(n)$. By taking a union bound over all sets of L , we conclude that $W = \{1, \dots, k\}$ is an AJR committee with high probability.

On the other hand, if p is not large enough such that $u_\ell < \ell$ for some ℓ , then, with high probability, those $\ell \cdot \frac{n}{k}$ voters approving L with minimum utilities, which form an ℓ -cohesive group, fail to meet the AJR requirements. By taking a union bound over the choice of W , we conclude that an AJR committee does not exist with high probability.

Putting the above intuitions together, we have the following intermediate characterization for the likelihood of an AJR committee.

Lemma 3.1. *For any constant m, k , and p , the likelihood that there exists an AJR committee W has the following trichotomy. Let $\mathcal{L} = \{\ell \mid 1 \leq \ell \leq \min\{k, m-k\}, p^\ell \geq \frac{\ell}{k}\}$. Let t_ℓ be the smallest positive integer such that $n_\ell = \sum_{t=0}^{t_\ell} \binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n \geq \ell \cdot \frac{n}{k}$. Let*

$$u_\ell = \frac{k}{\ell \cdot n} \cdot \left(\sum_{t=0}^{t_\ell} t \cdot \binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n - t_\ell \cdot (n_\ell - \ell \cdot n/k) \right). \quad (3)$$

Then,

1. If $p < \frac{1}{k}$ (which means $\mathcal{L} = \emptyset$), $p = 1$, or $u_\ell > \ell$ for every $\ell \in \mathcal{L}$, then the likelihood for the existence of an AJR committee is $1 - o(1)$;
2. If there exists an $\ell \in \mathcal{L}$ such that $u_\ell < \ell$, then the likelihood for the existence of an AJR committee is $o(1)$;
3. Otherwise, when $u_\ell \geq \ell$ for every $\ell \in \mathcal{L}$ and there exists an $\ell \in \mathcal{L}$ such that $u_\ell = \ell$, the likelihood for the existence of an AJR committee is $\Theta(1)$ and $1 - \Theta(1)$.

The proof for the first two parts of Lemma 3.1 follows the same high-level ideas described before Lemma 3.1. We defer the proof to Appendix A. The third part with $u_\ell = \ell$ is analyzed by applying the techniques of *polyhedron approach* in Xia [26], and it is handled in Sect. 5.

In Sect. 4, we further analyze the relationship between u_ℓ and ℓ and show how to obtain the much cleaner characterization in Theorem 2.2.

Only 1-Cohesive Groups Matter. During the proof of Theorem 2.2 in the coming section, we discover an interesting phenomenon stated in Proposition 4.2: for every $\ell \geq 2$, if p is large enough such that ℓ -cohesive groups exist, i.e., $p^\ell \geq \frac{\ell}{k}$, then we automatically have $u_\ell > \ell$. This means that the pre-conditions for (2) and (3) in Lemma 3.1 can only hold for $\ell = 1$. For $\ell \geq 2$, the value of p that is large enough for the existence of an ℓ -cohesive group is also large enough for this ℓ -cohesive group to satisfy AJR for any given winning committee of k candidates. Therefore, AJR can only fail for those 1-cohesive groups under the Erdős–Rényi bipartite model. We are not sure if the phenomenon that 1-cohesive groups are the hardest to satisfy holds under other random models, and it is an interesting future direction to see how broadly this observation can be applied.

4 Proof of Theorem 2.2

In this section, we take a closer look to the characterization in Lemma 3.1 and reveal some relationships between p , u_ℓ , and ℓ , which enable us to conclude Theorem 2.2.

One challenge for obtaining Theorem 2.2 from Lemma 3.1 is that the definitions of t_ℓ and n_ℓ , which are given in Equation (2), are complicated and difficult to analyze. This makes the characterization of u_ℓ (which depends on t_ℓ and n_ℓ) difficult. To overcome this, we begin by viewing u_ℓ in a slightly different way: instead of letting the summation to up to the particular value t_ℓ defined in Lemma 3.1, we consider a general “stopping value” T instead of t_ℓ . That is, we define

$$\begin{aligned} U(T) &= \frac{k}{\ell} \left(\sum_{t=0}^T t \binom{k}{t} p^{\ell+t} (1-p)^{k-t} - T \left(\sum_{t=0}^T \binom{k}{t} p^{\ell+t} (1-p)^{k-t} - \frac{\ell}{k} \right) \right) \\ &= T - \frac{k}{\ell} \left(\sum_{t=0}^T (T-t) \binom{k}{t} p^{\ell+t} (1-p)^{k-t} \right). \end{aligned} \tag{4}$$

Notice that, compared with Equation (3) in Lemma 3.1, we have substituted n_ℓ to the equation, replaced t_ℓ by a general input value T , and canceled n .

Recall that t_ℓ is the minimum utility value such that the expected number of voters who approve L with utilities at most t_ℓ exceeds $\ell \cdot \frac{n}{k}$. When $T < t_\ell$, we have not selected enough number of voters to reach $\ell \cdot \frac{n}{k}$. In this case, $\sum_{t=0}^T \binom{k}{t} p^{\ell+t} (1-p)^{k-t} - \frac{\ell}{k} < 0$, and $U(T)$ treats those $(\ell \cdot \frac{n}{k} - \sum_{t=0}^T \binom{k}{t} p^{\ell+t} (1-p)^{k-t} \cdot n)$ extra voters as they receive utility T . Since in reality their utilities are between T and t_ℓ , $U(T)$ has underestimated u_ℓ . On the other hand, when $T > t_\ell$, we have selected more voters than needed. In this case, $\sum_{t=0}^T \binom{k}{t} p^{\ell+t} (1-p)^{k-t} - \frac{\ell}{k} > 0$, and $U(T)$

excludes these $(\sum_{t=0}^T \binom{k}{t} p^{\ell+t} (1-p)^{k-t} - \frac{\ell}{k})n$ voters as if they receive utility T . Since in reality these agents' utilities is between t_ℓ and T , $U(T)$ again underestimates u_ℓ . Therefore, we have the following proposition. The full proof is in Appendix B.1.

Proposition 4.1. *For any k, ℓ , and p , we have $u_\ell = \max_{1 \leq T \leq k} U(T)$.*

With this new interpretation of u_ℓ , we are ready to prove Theorem 2.2. Our proof consists of two parts.

In the first part, we will show the following proposition.

Proposition 4.2. *For each $\ell \geq 2$, if $p^\ell \geq \frac{\ell}{k}$, then $u_\ell > \ell$.*

This proposition implies that, for each $\ell \geq 2$, either with high probability no ℓ -cohesive group exists (in the case $p^\ell < \frac{\ell}{k}$) or with high probability the AJR condition is met for every ℓ -cohesive group ($u_\ell > \ell$ as suggested by the proposition). As a result, we do not need to worry about ℓ -cohesive groups for $\ell \geq 2$ at all. This proposition also implies, in Lemma 3.1, the pre-condition for (2) and (3) can only happen with $\ell = 1$.

In the second part, we study the case with $\ell = 1$ and prove the following proposition.

Proposition 4.3. *When $\ell = 1$ and $1 > p > \frac{1}{k}$,*

1. $U(2) < 1$ implies $u_1 < 1$;
2. $U(2) = 1$ implies $u_1 = 1$;
3. $U(2) > 1$ implies $u_1 > 1$.

In addition, the equation $U(2) = 1$ has exactly one root p^ in $(\frac{1}{k}, 1)$, and*

1. $U(2) < 1$ when $\frac{1}{k} < p < p^*$, and
2. $U(2) > 1$ when $p > p^*$.

When $\ell = 1$, $p = \frac{1}{k}$, we have $1 = U(k) > U(k-1) > \dots > U(1)$.

First of all, notice that we always have $\ell = 1 \in \mathcal{L}$ (see Lemma 3.1 for definition of $\mathcal{L}$) when $1 > p \geq \frac{1}{k}$. Therefore, in the proposition above, $u_1 > 1$, $u_1 < 1$, and $u_1 = 1$ precisely correspond to (1), (2), and (3) in Lemma 3.1, respectively.

The first part of the proposition implies that the second function $U(2)$ in the family $\{U(T)\}_{T=1,\dots,k}$ characterizes the relationship between u_ℓ and ℓ (which is just u_1 and 1 since Proposition 4.2 tells us we do not need to look at $\ell \geq 2$). There is a complicated reason why $U(2)$ (instead of $U(T)$ for other values of T) is so special here. We will provide intuitions and formal proofs later. The second part of the proposition gives a precise characterization of $U(2)$. By noticing that Equation (1) in Theorem 2.2 is exactly the equation $U(2) = 1$, we conclude Theorem 2.2, except for the only corner case $p = \frac{1}{k}$. The last part discusses the corner case $p = \frac{1}{k}$. By Proposition 4.1 and Proposition 4.2, it implies $u_\ell \geq \ell$ always holds and $u_\ell = \ell$ for $\ell = 1$. Lemma 3.1 says that the likelihood of an AJR committee is $\Theta(1)$ and $1 - \Theta(1)$, which agrees with the description of Theorem 2.2.

Now, it remains to prove the two propositions. Before this, we first argue that we can assume $\ell < k$ without loss of generality. For $\ell = k$, according to Lemma 3.1, the only case we need to take such an ℓ into consideration is when $p = 1$ (as this is the only p making $p^\ell \geq \frac{\ell}{k}$). In this case, it is obvious that AJR committees exist with probability 1, which also agrees with our characterization in Theorem 2.2, as Proposition 4.3 implies that p_2^* in Theorem 2.2 is strictly less than 1. In the remaining part of this section, we will assume $\ell < k$.

4.1 Proof of Proposition 4.2

For $\ell \geq 2$, we will show that all ℓ -cohesive groups always satisfy the AJR condition.

Since we have assumed $k > \ell$, we have $\ell + 1 \in \{1, \dots, k\}$ and $u_\ell = \max_{1 \leq T \leq k} U(T) \geq U(\ell + 1)$. It then suffices to show that $U(\ell + 1) > \ell$ for all $\ell \geq 2$, that is,

$$U(\ell + 1) = \ell + 1 - \frac{k}{\ell} \left(\sum_{t=0}^{\ell} (\ell + 1 - t) \binom{k}{t} p^{\ell+t} (1-p)^{k-t} \right) > \ell.$$

Rearranging the inequality, we obtain

$$\frac{kp^\ell}{\ell} \left(\sum_{t=0}^{\ell} (\ell + 1 - t) \binom{k}{t} p^t (1-p)^{k-t} \right) < 1. \quad (5)$$

We perform the following calculations on the left-hand side of (5).

$$\begin{aligned} & \frac{kp^\ell}{\ell} \left(\sum_{t=0}^{\ell} (\ell + 1 - t) \binom{k}{t} p^t (1-p)^{k-t} \right) \\ &= \frac{kp^\ell}{\ell} \left(\sum_{t=0}^{\ell} \frac{k \cdots (k-t+1)}{\ell \cdots (\ell-t+2)} \cdot (\ell + 1 - t) \frac{\ell \cdots (\ell-t+2)}{t!} p^t (1-p)^{k-t} \right) \\ &\leq \frac{kp^\ell}{\ell} \left(\sum_{t=0}^{\ell} \binom{k}{\ell} \cdot \binom{\ell}{t} p^t (1-p)^{k-t} \right) \\ &\quad \left(\text{since } \binom{k}{\ell} \geq \frac{k \cdots (k-t+1)}{\ell \cdots (\ell-t+2)} \text{ (see explanations below) and } \binom{\ell}{t} = (\ell + 1 - t) \frac{\ell \cdots (\ell-t+2)}{t!} \right) \\ &= \frac{kp^\ell}{\ell} \binom{k}{\ell} \cdot (1-p)^{k-\ell} \quad \left(\sum_{t=0}^{\ell} \binom{\ell}{t} p^t (1-p)^{\ell-t} = (p + (1-p))^\ell = 1 \right) \end{aligned}$$

For the middle inequality, $\binom{k}{\ell} \geq \frac{k \cdots (k-t+1)}{\ell \cdots (\ell-t+2)}$ holds because $\binom{k}{\ell} / \frac{k \cdots (k-t+1)}{\ell \cdots (\ell-t+2)} = \frac{1}{\ell-t+1} \binom{k-t}{\ell-t}$ is decreasing in k and substituting the minimum value of k , which is $k = \ell + 1$, gives the ratio 1.

Let $f(p) = \frac{kp^\ell}{\ell} \binom{k}{\ell} \cdot (1-p)^{k-\ell}$. We compute its derivative

$$f'(p) = \frac{k}{\ell} \binom{k}{\ell} p^{\ell-1} (1-p)^{k-\ell-1} (\ell - kp).$$

We have $f'(p) < 0$ since $p \geq \sqrt[\ell]{\frac{\ell}{k}} > \frac{\ell}{k}$. Thus, we have

$$f(p) \leq f\left(\sqrt[\ell]{\frac{\ell}{k}}\right) = \binom{k}{\ell} \left(1 - \sqrt[\ell]{\frac{\ell}{k}}\right)^{k-\ell}.$$

Now, it remains to show the following proposition.

Proposition 4.4. *For any integers ℓ and k with $2 \leq \ell < k$,*

$$\binom{k}{\ell} \left(1 - \sqrt[\ell]{\frac{\ell}{k}}\right)^{k-\ell} < 1.$$

The proof of Proposition 4.4 is fully technical and involves some numerical checks by computer programs for small enough k and ℓ . See Appendix B.2.

4.2 Proof of Proposition 4.3

Recall that when $\ell = 1$,

$$U(T) = T - k \left(\sum_{t=0}^T (T-t) \binom{k}{t} p^{1+t} (1-p)^{k-t} \right) \quad (6)$$

and

$$u_1 = \max_{1 \leq T \leq k} U(T). \quad (7)$$

We will prove the three parts of the propositions in the reverse order. The last part of the proposition can be proved by simple calculations. The proof for the second part depends on the result in the last part, and it can be easily obtained by computing the derivative of $U(2)$ with respect to p . The proof for the first part is the most technically challenging one, and it depends on the results for the second and third parts. We will provide some intuitions and then the formal proof after we prove the third and the second parts.

We first prove the last part of the proposition.

Proof (last part of Proposition 4.3). When $p = \frac{1}{k}$, we have

$$\begin{aligned} U(k) &= k - k \left(\sum_{t=0}^k (k-t) \binom{k}{t} p^{1+t} (1-p)^{k-t} \right) \\ &= k - k \left(\sum_{t=0}^{k-1} (k-t) \binom{k}{t} p^{1+t} (1-p)^{k-t} \right) \quad (\text{the last term is 0}) \\ &= k - k^2 p (1-p) \sum_{t=0}^{k-1} \binom{k-1}{t} p^t (1-p)^{k-1-t} \\ &= k - k^2 p (1-p) (p + (1-p))^{k-1} = k - k^2 p (1-p) = 1, \end{aligned}$$

and

$$\begin{aligned} U(T+1) - U(T) &= 1 - kp \left(\sum_{t=0}^T \binom{k}{t} p^t (1-p)^{k-t} \right) \\ &\quad (\text{the last term in the summation of } U(T+1) \text{ is 0}) \\ &= 1 - \left(\sum_{t=0}^T \binom{k}{t} p^t (1-p)^{k-t} \right) \quad (kp = 1) \\ &> 1 - \left(\sum_{t=0}^k \binom{k}{t} p^t (1-p)^{k-t} \right) = 0. \end{aligned}$$

Thus, $1 = U(k) > U(k-1) > \dots > U(2) > U(1)$ and $u_1 = U(k) = 1$. □

We next prove the second part of the proposition.

Proof (second part of Proposition 4.3). We have

$$U(2) = 2 - k(2p(1-p)^k + kp^2(1-p)^{k-1}).$$

Direct calculations reveal

$$\frac{\partial U(2)}{\partial p} = k(1-p)^{k-2} (k(k-1)p^2 - 2(1-p)^2), \quad (8)$$

so, with $p < 1$, $\frac{\partial U(2)}{\partial p} = 0$ implies

$$k(k-1)p^2 - 2(1-p)^2 = 0,$$

which further implies

$$p_0 = \frac{1}{1 + \sqrt{\frac{k(k-1)}{2}}},$$

and, in addition, $U(2)$ is decreasing for $p \in (0, p_0)$ and is increasing for $p \in (p_0, 1)$. By the last part of Proposition 4.3 (which has just been proved), we have $U(2) < 1$ at $p = \frac{1}{k}$. On the other hand, $U(2) = 2 > 1$ at $p = 1$. Since $\frac{\partial U(2)}{\partial p}$ changes signs only once, from negative to positive, on the interval $(0, 1)$, we conclude the second part of the proposition. $\square$

Now it remains to show the first part of the proposition.

First of all, notice that $U(1) < 1$ holds for $\frac{1}{k} < p < 1$. Thus, we do not need to consider $U(1)$: if there exists $T > 1$ such that $U(T) \geq 1$ at a particular value of p , then $u_1 = \max_{1 \leq T \leq k} U(T)$ cannot be $U(1)$; if $U(T) < 1$ for all $T > 1$ at a particular value of p , then $u_1 < 1$ even when u_1 takes $U(1)$.

We next consider the easy case with $k = 2$. In this case, the family only contains $U(1)$ and $U(2)$, and it suffices to look at $U(2)$ alone. Since we have seen $U(1) < 1$ for $\frac{1}{k} < p < 1$, the first part of Proposition 4.3 holds trivially (if $U(2) < 1$, then $u_1 = \max\{U(1), U(2)\} < 1$ given $U(1) < 1$; for the remaining two cases with $U(2) = 1$ or $U(2) > 1$, we have $u_1 = \max\{U(1), U(2)\} = U(2)$ given $U(1) < 1$). In the remaining part of this section, we will assume $k \geq 3$.

We first characterize the function family $\{U(T)\}$. Fig. 5 illustrates the curves of $U(T)$ for

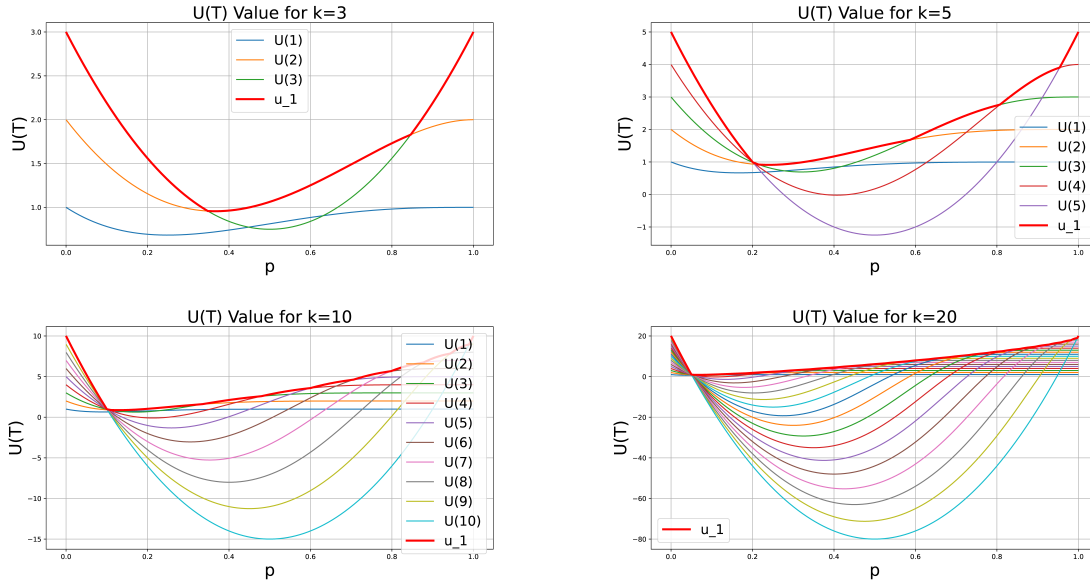


Figure 5: Values of $U(T)$ with different k

different values of k . Recall that we want to find out for which values of p we have $u_1 < 1$, and

$u_1 = \max_{1 \leq T \leq k} U(T)$. The bold red curve in the figure represents u_1 , which takes the maximum over all $U(T)$'s, and we can see that the curve is below $y = 1$ for a small region between $p_1^* = \frac{1}{k}$ and some p_2^* . To formally conclude Proposition 4.3, we first state some observations from the pictures and show that how these observations can prove the first part of Proposition 4.3. We will then formally prove these observations.

Starting from $p = \frac{1}{k}$, we examine the behavior of u_1 from the picture as p increases. The third part of the Proposition 4.3 (which has proved already) indicates that $u_1 = U(k) = 1$, and when p slightly increases by a little bit, all values of $U(T)$ are below 1. This is formally proved in Claim 4.5, and this shows that the left-hand side threshold for $u_1 < 1$ is exactly at $p_1^* = \frac{1}{k}$. Next, we characterize the right-hand side threshold p_2^* . We first observe from the picture that $U(2)$ reaches its minimum first after ignoring $U(1)$ (proved in Claim 4.7). This implies u_1 is still below 1 before $U(2)$ reaches its minimum. In addition, $U(2)$ is the maximum among all $U(T)$'s at the time $U(2)$ reaches its minimum (Claim 4.8 and Claim 4.6), and $U(2)$ continues to be the maximum until we reach a point where $U(2) = U(3)$. We will show in Claim 4.9 that $U(2) > 1$ (and so $u_1 > 1$) at the time we reach $U(2) = U(3)$. Finally, when p continues to increase, the value of $U(2)$ is increasing, so we have $u_1 \geq U(2) > 1$. To summarize, $U(2)$ characterizes the relationship between u_1 and 1:

- between $\frac{1}{k}$ and the point where $U(2)$ takes its minimum, we know all $U(T)$'s are below 1, although $U(2)$ may not be the maximum over all $U(T)$'s;
- between the point $U(2)$ takes its minimum and the point $U(2) = U(3)$, we will show that $U(2)$ is the maximum over all $U(T)$'s, so $u_1 = U(2)$; since we know $U(2) > 1$ by the time $U(2) = U(3)$, the intersection of $U(2)$ and $y = 1$ happens in this regime;
- between the point $U(2) = U(3)$ and $p = 1$, we see that $U(2) > 1$, so $u_1 \geq u(2) > 1$ even if $U(2)$ is no longer the maximum over all $U(T)$'s.

Now we begin to prove the proposition formally. Although these observations can be easily seen from the pictures, the proofs for many of them are rather tricky.

Claim 4.5. *When $p = \frac{1}{k}$, we have $\frac{\partial U(T)}{\partial p} < 0$ for all $2 \leq T \leq k$; that is, each $U(T)$ is decreasing at $p = \frac{1}{k}$.*

Proof. By writing $U(T) = T - kp \left(\sum_{t=0}^T (T-t) \binom{k}{t} p^t (1-p)^{k-t} \right)$ and differentiating by the two parts, the term p outside and the summation, we can reach the following expression by some calculations involving rewriting the summations and the combinatorial numbers. The detailed calculations are available in Appendix B.3.

$$\frac{\partial U(T)}{\partial p} = -k \left(\sum_{t=0}^{T-1} \binom{k}{t} p^t (1-p)^{k-t-1} [(T-t)(1-p) - (k-t)p] \right)$$

We will show that each term in the summation is non-positive and at least one of them is negative. Since each $\binom{k}{t} p^t (1-p)^{k-t-1}$ is positive at $p = \frac{1}{k}$, it suffices to show that $(T-t)(1-p) - (k-t)p$ is non-negative for all t and positive for at least one t .

The value of $(T-t)(1-p) - (k-t)p$ is $p((T-t)(k-1) - k + t)$ since $p = \frac{1}{k}$. If $t = 0$, we have $(T-t)(k-1) - k + t = T(k-1) - k \geq k-2 > 0$. If $t \geq 1$, we have $(T-t)(k-1) - k + t \geq k-1-k+t \geq 0$. $\square$

Claim 4.6. *For all $1 < T < k$, if $U(T) < U(T-1)$ then $U(T+1) < U(T)$.*

Proof. By noticing the last term in the summation of $U(T)$ is 0, we have

$$\begin{aligned} 0 > U(T) - U(T-1) &= 1 - kp \left(\sum_{t=0}^{T-1} \binom{k}{t} p^t (1-p)^{k-t} \right) \\ &> 1 - kp \left(\sum_{t=0}^T \binom{k}{t} p^t (1-p)^{k-t} \right) = U(T+1) - U(T), \end{aligned}$$

which implies the claim. $\square$

Claim 4.7. $\frac{\partial U(T)}{\partial p} < 0$ implies $\frac{\partial U(T+1)}{\partial p} < 0$.

Proof. Let $C_t = -k \binom{k}{t} p^t (1-p)^{k-t-1} < 0$ for each $t = 0, 1, \dots, k$. We have

$$\frac{\partial U(T)}{\partial p} = -k \sum_{t=0}^T (T-t) \binom{k}{t} ((1+t)p^t (1-p)^{k-t} - (k-t)p^{1+t} (1-p)^{k-t-1}) = \sum_{t=0}^T (T-t) C_t [1+t-(1+k)p] < 0, \quad (9)$$

and

$$\frac{\partial U(T+1)}{\partial p} - \frac{\partial U(T)}{\partial p} = \sum_{t=0}^T C_t [1+t-(1+k)p]. \quad (10)$$

We discuss the following cases:

- $1 \geq (1+k)p$. Since $C_t < 0$, every term in the summation (10) is negative. We are done.
- $1+T-(1+k)p < 0$, i.e., every term in the summation (9) is positive. This is a contradiction to our assumption.
- Otherwise, there exists $0 \leq t' < T$, such that $1+t'-(1+k)p \leq 0$ and $1+(t'+1)-(1+k)p > 0$.

$$\begin{aligned} & \frac{\partial U(T+1)}{\partial p} - \frac{\partial U(T)}{\partial p} \\ &= \sum_{t=0}^T C_t [1+t-(1+k)p] \\ &= \frac{1}{T-t'} \left(\sum_{t=0}^{t'} (T-t') C_t [1+t-(1+k)p] + \sum_{t=t'+1}^T (T-t') C_t [1+t-(1+k)p] \right) \quad (11) \\ &< \frac{1}{T-t'} \left(\sum_{t=0}^{t'} (T-t) C_t [1+t-(1+k)p] + \sum_{t=t'+1}^T (T-t) C_t [1+t-(1+k)p] \right) \\ &= \frac{1}{T-t'} \frac{\partial U(T)}{\partial p} < 0 \end{aligned}$$

Thus, $\frac{\partial U(T+1)}{\partial p} < \frac{\partial U(T)}{\partial p} < 0$. $\square$

Claim 4.8. For $k \geq 3$, when $\frac{\partial U(2)}{\partial p} = 0$, we have $U(3) - U(2) < 0$.

Proof (sketch). We have

$$U(2) = 2 - k(2p(1-p)^k + kp^2(1-p)^{k-1})$$

and

$$U(3) = 3 - k \left(3p(1-p)^k + 2kp^2(1-p)^{k-1} + \frac{1}{2}k(k-1)p^3(1-p)^{k-2} \right).$$

We have seen in the proof for the second part of Proposition 4.3 that $\frac{\partial U(2)}{\partial p} = 0$ implies

$$k(k-1)p^2 - 2(1-p)^2 = 0, \quad (12)$$

which further implies

$$p = \frac{1}{1 + \sqrt{\frac{k(k-1)}{2}}}. \quad (13)$$

On the other hand,

$$\begin{aligned} U(3) - U(2) &= 1 - k^2p^2(1-p)^{k-2} \left(\frac{(1-p)^2}{kp} + 1 - p + \frac{1}{2}(k-1)p \right) \\ &= 1 - k^2p^2(1-p)^{k-2} \left(\frac{k(k-1)p^2}{2kp} + 1 - p + \frac{1}{2}(k-1)p \right) \quad (\text{Substituting (12)}) \\ &= 1 - k^2p^2(1-p)^{k-2}(1 + (k-2)p). \end{aligned}$$

Let $\phi(k, p) = k^2p^2(1-p)^{k-2}(1 + (k-2)p)$. It remains to show that $\phi(k, p) > 1$ for $k \geq 3$ and p given by (13).

We begin by providing a sketch to prove this. Notice that (13) implies $p \approx \frac{\sqrt{2}}{k}$. For sufficiently large k , by omitting lower order terms and substituting $p \approx \frac{\sqrt{2}}{k}$, we have

$$\phi(k, p) \approx k^2 \left(\frac{\sqrt{2}}{k} \right)^2 \left(1 - \frac{\sqrt{2}}{k} \right)^k \left(1 + k \cdot \frac{\sqrt{2}}{k} \right) \approx 2 \cdot \left(\frac{1}{e} \right)^{\sqrt{2}} (1 + \sqrt{2}) \approx 1.1738 > 1.$$

Therefore, $\phi(k, p) > 1$ holds for sufficiently large k .

For a formal proof, we need to work out the constants in those approximate equalities and find out a large enough k for which $\phi(k, p)$ holds. We see that the inequality holds for $k > 1000$. Then the inequality for $k \leq 1000$ is checked by a computer program by enumerating all possible values of k . The details are available in Appendix B.4. $\square$

Claim 4.9. For $k \geq 3$, when $\frac{\partial U(2)}{\partial p} > 0$ and $U(3) = U(2)$, we have $U(2) > 1$.

Proof. By (8), $\frac{\partial U(2)}{\partial p} > 0$ implies

$$k(k-1)p^2 > 2(1-p)^2. \quad (14)$$

On the other hand, simplifying $U(2) = U(3)$ gives

$$\frac{1}{2}k^2(k-1)p^3(1-p)^{k-2} = 1 - kp(1-p)^k - k^2p^2(1-p)^{k-1}. \quad (15)$$

Finally,

$$U(2) = 2 - k(2p(1-p)^k + kp^2(1-p)^{k-1})$$

$$\begin{aligned}
&= \left(1 - kp(1-p)^k\right) + \left(1 - kp(1-p)^k - k^2p^2(1-p)^{k-1}\right) \\
&= \left(1 - kp(1-p)^k\right) + \frac{1}{2}k^2(k-1)p^3(1-p)^{k-2} \quad (\text{by (15)}) \\
&= 1 + kp(1-p)^{k-2} \left(\frac{1}{2}k(k-1)p^2 - (1-p)^2\right) \\
&> 1, \quad (\text{by (14)})
\end{aligned}$$

which concludes the claim. $\square$

Now we are ready to prove the first part Proposition 4.3.

Proof (first part of Proposition 4.3). To prove the first part of the proposition, we will show that there exist p_1 and p_2 with $\frac{1}{k} < p_1 < p_2 < 1$ such that

1. for $p \in (\frac{1}{k}, p_1)$, $U(T) < 1$ for all $T = 2, \dots, k$,
2. for $p \in [p_1, p_2)$, $u_1 = U(2)$, and
3. for $p \in [p_2, 1]$, $U(2) > 1$.

This will be sufficient for proving the first part of Proposition 4.3 due to Proposition 4.1.

We will set p_1 to the value of p where $\frac{\partial U(2)}{\partial p} = 0$, which has been given in (13). Notice that $p_1 = \frac{1}{1 + \sqrt{\frac{k(k-1)}{2}}} > \frac{1}{k}$. We will set p_2 to be the value of p where $U(2) = U(3)$, and we will see later that $p_2 > p_1$.

For (1), by Claim 4.5 and the last part of the proposition, we have $U(T) < 1$ for $T = 2, \dots, k$ in $(\frac{1}{k}, \frac{1}{k} + \varepsilon)$ for sufficiently small $\varepsilon > 0$. On the other hand, we have seen in the proof of the second part of Proposition 4.3 that $\frac{\partial U(2)}{\partial p} < 0$ for $p \in (\frac{1}{k}, p_1)$, and Claim 4.7 implies $U(T)$ is decreasing on $(\frac{1}{k}, p_1)$ for every T . Putting together, we have $U(T) < 1$ for every T and $p \in (\frac{1}{k}, p_1)$.

For (2), due to Claim 4.6, it suffices to show that $U(2) > U(3)$ for $p \in [p_1, p_2)$. Let

$$F(p) = U(3) - U(2) = 1 - k \left(p(1-p)^k + kp^2(1-p)^{k-1} + \frac{1}{2}k(k-1)p^3(1-p)^{k-2} \right)$$

and

$$\begin{aligned}
F'(p) &= \frac{\partial(U(3) - U(2))}{\partial p} \\
&= -k(1-p)^{k-3} \left[(1-p)^3 + kp(1-p)^2 + \frac{1}{2}k(k-1)p^2(1-p) - \frac{1}{2}k(k-1)(k-2)p^3 \right].
\end{aligned}$$

Let

$$f(p) = (1-p)^3 + kp(1-p)^2 + \frac{1}{2}k(k-1)p^2(1-p) - \frac{1}{2}k(k-1)(k-2)p^3.$$

One can easily check that, when $k = 3$, the equation $f(p) = 0$ has only one solution. For general $k \geq 4$, the function $f(p)$ is a cubic function whose derivative changes signs at most twice. Since $f(0) > 0$, $f(1) < 0$, and $f'(0) = k-3 > 0$, it is clear that $f(p) = 0$ has only one solution. Thus, $F(p)$ is monotonically decreasing and then monotonically increasing on $(0, 1)$. Since $F(0) = F(1) = 1 > 0$ and $F(p_1) < 0$ by Claim 4.8, there exists exactly one point $p_2 \in (p_1, 1)$ such that $F(p_2) = 0$. Therefore, when $p \in [p_1, p_2)$, $F(p) < 0$, that is, $U(3) < U(2)$.

To see (3), Claim 4.9 tells that $U(2) > 1$ when $p = p_2$. Since $U(2)$ is increasing on $(p_1, 1)$ and $p_2 > p_1$, we conclude $U(2) > 1$ for $p \in [p_2, 1]$. $\square$

4.3 Proof of Proposition 2.3

With the above analysis, we can now easily prove Proposition 2.3 in Sect. 2.1.

Let $\phi(x) = k(2x(1-x)^k + kx^2(1-x)^{k-1})$ be the left-hand side of Equation (1). For $\ell = 1$, $\phi(p)$ is just $2 - U(2)$. In the proof of the second part of Proposition 4.3, we have seen that $U(2)$ is increasing and then decreasing on the interval $(0, 1)$. In addition, the third part of Proposition 4.3 implies $U(2) < 1$ at $p = \frac{1}{k}$. Therefore, $\phi(x)$ is increasing and then decreasing on $(0, 1)$, and $\phi(\frac{1}{k}) > 1$. Noticing that $\phi(1) = 0$, there is a unique solution in $(\frac{1}{k}, 1)$ for the equation $\phi(x) = 1$, and this solution is exactly p_2^* . In addition, any x with $\phi(x) < 1$ gives an upper bound to p_2^* . Taking $x = \frac{5}{k}$ as it is in the proposition, we have

$$\begin{aligned} \phi\left(\frac{5}{k}\right) &= 10\left(1 - \frac{5}{k}\right)^k + 25\left(1 - \frac{5}{k}\right)^{k-1} \\ &< 10e^{-5} + 25e^{-5\frac{k-1}{k}} && \text{(by the inequality } 1+x < e^x \text{ for } x \neq 0) \\ &\leq 10e^{-5} + 25e^{-5 \times \frac{2}{3}} && \text{(since } k \geq 3) \\ &< 1. && (10e^{-5} + 25e^{-5 \times \frac{2}{3}} \text{ is approximately } 0.959) \end{aligned}$$

5 Proof of Lemma 3.1 at Phase-Transition Points

The third part of Lemma 3.1 with $u_\ell = \ell$ is analyzed by applying the techniques of *polyhedron approach* in Xia [26]. The techniques are reviewed in Sect. 5.1. Finally, Propositions 5.4 and 5.5 characterize the $u_\ell = \ell$ case and show that the likelihood of AJR committee's existence and non-existence are both $\Theta(1)$ respectively.

5.1 Techniques of Polyhedron Approach

Definition 5.1 (Poisson multivariate variables (PMVs) [26]). *Given any $q, n \in \mathbb{Z}^+$ and any vector $\vec{\pi} = (\pi_1, \dots, \pi_n)$ of n distributions over $[q]$, an (n, q) -PMV, denoted by $\vec{X}_{\vec{\pi}}$, is the histogram of n independent random variables $Y_1, Y_2, \dots, Y_n$, where Y_i follows distribution π_i .*

A polyhedron $\mathcal{H} \subseteq \mathbb{R}^q$ is characterized by a matrix A and a vector $\vec{b}$, i.e., $\mathcal{H} := \{\vec{x} \in \mathbb{R}^q : A\vec{x} \leq \vec{b}\}$. Let $\dim(\mathcal{H})$ be the dimension of the polyhedron $\mathcal{H}$. We say that $\mathcal{H} \subseteq \mathbb{R}^q$ is full-dimensional if $\dim(\mathcal{H}) = q$. Let $\mathcal{H}_{\leq 0} = \{\vec{x} \in \mathbb{R}^q : A\vec{x} \leq \vec{0}\}$ be the characteristic cone of $\mathcal{H}$. Let $\mathcal{H}_n^{\mathbb{Z}}$ be the set of all integral points whose L_1 norm is n in the polyhedron $\mathcal{H}$.

Given a set Π of distributions over $[q]$, we can view each $\pi \in \Pi$ as a point in $\{(x_1, \dots, x_q) \in \mathbb{R}^q \mid \sum_{i=1}^q x_i = 1\}$, so Π is a subset of $\mathbb{R}^q$. We say Π is closed if Π is a closed set in $\mathbb{R}^q$. We further use $\text{CH}(\Pi)$ to denote the convex hull of Π . We say a distribution π is strictly positive if there exists a constant $\varepsilon > 0$ such that the probability of any $i \in [q]$ under π is at least ε . A distribution set Π is strictly positive if any distribution $\pi \in \Pi$ is strictly positive.

Given $q \in \mathbb{Z}^+$, a polyhedron $\mathcal{H} \subseteq \mathbb{R}^q$, and a set Π of distributions over $[q]$, we are interested in $\sup_{\vec{\pi} \in \Pi^n} \Pr[\vec{X}_{\vec{\pi}} \in \mathcal{H}]$ and $\inf_{\vec{\pi} \in \Pi^n} \Pr[\vec{X}_{\vec{\pi}} \in \mathcal{H}]$.

Theorem 5.2 (Smooth Likelihood of PMV-in-polyhedron [26]). *Given any $q \in \mathbb{Z}^+$, any closed and strictly positive Π over $[q]$, and any polyhedron $\mathcal{H}$ characterized by an integer matrix A , for any $n \in \mathbb{Z}^+$,*

$$\sup_{\vec{\pi} \in \Pi^n} \Pr(\vec{X}_{\vec{\pi}} \in \mathcal{H}) = \begin{cases} 0 & \text{if } \mathcal{H}_n^{\mathbb{Z}} = \emptyset \\ \exp(-\Theta(n)) & \text{if } \mathcal{H}_n^{\mathbb{Z}} \neq \emptyset \text{ and } \mathcal{H}_{\leq 0} \cap \text{CH}(\Pi) = \emptyset \\ \Theta\left(\sqrt{n}^{\dim(\mathcal{H}_{\leq 0})-q}\right) & \text{otherwise} \end{cases}$$

$$\inf_{\vec{\pi} \in \Pi^n} \Pr(\vec{X}_{\vec{\pi}} \in \mathcal{H}) = \begin{cases} 0 & \text{if } \mathcal{H}_n^{\mathbb{Z}} = \emptyset \\ \exp(-\Theta(n)) & \text{if } \mathcal{H}_n^{\mathbb{Z}} \neq \emptyset \text{ and } \text{CH}(\Pi) \not\subseteq \mathcal{H}_{\leq 0} \\ \Theta\left(\sqrt{n}^{\dim(\mathcal{H}_{\leq 0})-q}\right) & \text{otherwise} \end{cases}$$

When applying the polyhedral approach to our problem, we set $q = 2^m$, where $[q]$ represents the set of all subsets of M , corresponding to the approval ballots of voters. Each Y_i is the random variable corresponding to voter i 's approval ballot. Let π_i be the corresponding distribution of Y_i , i.e., for each $q' \in [q] = [2^m]$ that represents a subset $S_{q'}$ of candidates, π_i assigns probability $p^{|S_{q'}|}(1-p)^{m-|S_{q'}|}$ to q' under the Erdős-Rényi bipartite model with parameter p . We clearly have $\pi_1 = \dots = \pi_n$, and we denote this common distribution by π . Let $\vec{X}_{\vec{\pi}}$ be the corresponding $(n, 2^m)$ -PMV where $\vec{\pi} = \{\pi_1, \dots, \pi_n\}$. We have the following corollary for i.i.d. distributions to characterize the probability that $\vec{X}_{\vec{\pi}}$ is in some polyhedral $\mathcal{H}$.

Corollary 5.3. *Consider any $q \in \mathbb{Z}^+$, any strictly positive distribution π over $[q]$, and any polyhedron $\mathcal{H}$. For any $n \in \mathbb{Z}^+$, letting $\vec{X}_{\vec{\pi}}$ be the (n, q) -PMV with $\vec{\pi} = (\pi_1, \dots, \pi_n)$ and $\pi_1 = \dots = \pi_n = \pi$, we have*

$$\Pr(\vec{X}_{\vec{\pi}} \in \mathcal{H}) = \begin{cases} 0 & \text{if } \mathcal{H}_n^{\mathbb{Z}} = \emptyset \\ \exp(-\Theta(n)) & \text{if } \mathcal{H}_n^{\mathbb{Z}} \neq \emptyset \text{ and } \pi \notin \mathcal{H}_{\leq 0} \\ \Theta\left(\sqrt{n}^{\dim(\mathcal{H}_{\leq 0})-q}\right) & \text{otherwise} \end{cases}$$

Proof. We apply Theorem 5.2 with $\Pi = \{\pi\}$. The set Π is finite, so it is closed. In addition, $\text{CH}(\Pi) = \{\pi\}$. Thus, the conditions $\mathcal{H}_{\leq 0} \cap \text{CH}(\Pi) = \emptyset$ and $\text{CH}(\Pi) \not\subseteq \mathcal{H}_{\leq 0}$ for $\sup_{\vec{\pi} \in \Pi^n} \Pr(\vec{X}_{\vec{\pi}} \in \mathcal{H})$ and $\inf_{\vec{\pi} \in \Pi^n} \Pr(\vec{X}_{\vec{\pi}} \in \mathcal{H})$ respectively become the same, which is $\pi \notin \mathcal{H}_{\leq 0}$. The corollary follows. $\square$

To conclude Lemma 3.1 for the Case $u_\ell = \ell$, we will prove the following two propositions in the next two sub-sections using polyhedron approach.

Proposition 5.4. *For any constants m, k, p and letting $\mathcal{L}$ be as defined in Lemma 3.1, if $u_\ell \geq \ell$ for every $\ell \in \mathcal{L}$ and there exists $\ell \in \mathcal{L}$ such that $u_\ell = \ell$, then the probability that every committee W fails to satisfy the AJR condition on some ℓ -cohesive group is $\Theta(1)$.*

Proposition 5.5. *For any constant m, k, p , if $u_\ell \geq \ell$ for every $1 \leq \ell \leq k$ such that $p^\ell \geq \frac{\ell}{k}$, then the likelihood that an AJR committee exists is $\Theta(1)$.*

Both propositions share a similar proof idea. First, we construct a polyhedron $\mathcal{H}$ that characterizes a sub-event of “an AJR committee (does not) exist” respectively. Let $\vec{X}_{\vec{\pi}}$ be the PMV of n votes following distribution π . We show that $\Pr(\vec{X}_{\vec{\pi}} \in \mathcal{H}) = \Theta(1)$ by applying Corollary 5.3. Since both sub-events have a non-vanishing probability $\Theta(1)$, both “an AJR committee exists” and “an AJR committee does not exist” has probability $\Theta(1)$, and an AJR committee exists with probability $\Theta(1)$ and $1 - \Theta(1)$.

To show $\Pr(\vec{X}_{\vec{\pi}} \in \mathcal{H}) = \Theta(1)$ we prove three claims. (1) $\pi \in \mathcal{H}_{\leq 0}$, (2) $\mathcal{H}_n^{\mathbb{Z}} \neq \emptyset$, and (3) $\dim(\mathcal{H}_{\leq 0}) = q$. Claim (1) and (2) ensure that Corollary 5.3 falls into the third case, and (3) ensures that the third case implies $\Theta(1)$.

5.2 Proof of Proposition 5.4

We use the polyhedron approach to characterize a sub-event of “for every committee, an underrepresented ℓ -cohesive group exists”. For this proof, we fix an ℓ such that $u_\ell = \ell$. Note that u_ℓ and t_ℓ are determined solely on ℓ .

Recall that M is the set of candidates. Let $\mathcal{M}_k$ be the set of all k -subset of M and $\mathcal{M}_\ell$ be the set of all ℓ -subset of M . Let $\phi : \mathcal{M}_k \rightarrow \mathcal{M}_\ell$ be a mapping that maps every winning committee W to an ℓ -set $L_W \subseteq M \setminus W$. The existence of ϕ is guaranteed by $\ell \leq m - k$.

For each voter i , let Y_i be the random variable on 2^M denoting i 's approval ballot. For each $S \subseteq M$, let $x_S = |\{i \mid Y_i = S\}|$ be the number of voters whose approval ballot is exactly S . Note that $\sum_{S \subseteq M} x_S = n$. Let π be the distribution on 2^M in which each candidate is approved independently with probability p . Let $\vec{X}$ be the histogram of n independent variables $(Y_1, Y_2, \dots, Y_n)$, where each Y_i follows π independently. Consider the polyhedron $\mathcal{H} \subseteq \mathbb{R}^{2^m}$ defined by the following three sets of constraints. Note that t_ℓ is a constant given m, k, ℓ , and p . Note also that each variable x_S below is relaxed so that fractional values are allowed.

$$\begin{aligned} \forall W \in \mathcal{M}_k, \sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x_S &\geq \ell/k \cdot \sum_{S \subseteq M} x_S. \\ \forall W \in \mathcal{M}_k, \sum_{t=0}^{t_\ell-1} \sum_{L_W \subseteq S, |S \cap W|=t} x_S &\leq \ell/k \cdot \sum_{S \subseteq M} x_S. \\ \forall W \in \mathcal{M}_k, \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} t \cdot x_S - \left(\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x_S - \ell/k \cdot \sum_{S \subseteq M} x_S \right) \cdot t_\ell \\ &\leq \ell^2/k \cdot \sum_{S \subseteq M} x_S - t_\ell - \frac{1}{k}. \end{aligned}$$

For each winning committee W , the three constraints respectively say the followings:

1. there exists a group with at least $\ell \cdot n/k$ voters, all of which approve all candidates in L_W as well as at most t_ℓ candidates in W ;
2. the number of voters that approve all candidates in L_W as well as at most $t_\ell - 1$ candidates in W is at most $\ell \cdot n/k$; and
3. the total utility of $\ell \cdot n/k$ voters with the minimum utility (fractional voters allowed) is strictly lower than $\ell^2 \cdot n/k - t_\ell$. After taking the ceiling function, the total utility for those $\lceil \ell \cdot n/k \rceil$ voters with the minimum utilities is strictly less than $\ell^2 \cdot n/k$ (since $\lceil \ell \cdot n/k \rceil - \ell \cdot n/k < 1$ and t_ℓ is an upper bound on the utility of a single voter in this group). Notice that this implies the average utility for those $\lceil \ell \cdot n/k \rceil$ voters is strictly less than ℓ , failing the AJR condition.

It is not hard to verify that $\mathcal{H}$ is a sub-event of “every winning committee W fails the AJR condition on some ℓ -cohesive group”. Therefore, it remains to show $\Pr[\vec{X} \in \mathcal{H}] = \Theta(1)$. We will show the following three claims, which imply $\Pr[\vec{X} \in \mathcal{H}] = \Theta(1)$ by the third case of Corollary 5.3.

Claim 5.6. $\pi \in \mathcal{H}_{\leq 0}$.

Proof. The characteristic cone $\mathcal{H}_{\leq 0}$ can be written in the following form:

$$\forall W \in \mathcal{M}_k, \sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x_S \geq \ell/k \cdot \sum_{S \subseteq M} x_S.$$

$$\begin{aligned}
\forall W \in \mathcal{M}_k, \sum_{t=0}^{t_\ell-1} \sum_{L_W \subseteq S, |S \cap W|=t} x_S &\leq \ell/k \cdot \sum_{S \subseteq M} x_S. \\
\forall W \in \mathcal{M}_k, \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} t \cdot x_S - \left(\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x_S - \ell/k \cdot \sum_{S \subseteq M} x_S \right) \cdot t_\ell \\
&\leq \ell^2/k \cdot \sum_{S \subseteq M} x_S.
\end{aligned}$$

As π is a distribution on 2^m , we can represent it as a vector on $\mathbb{R}^{2^m}$, where π_S is the probability that the approval set of an agent is exactly S .

We then show that π satisfies the three constraints above for any $W \in \mathcal{M}_k$. By definition, t_ℓ is the largest number of winners a voter in the underrepresented ℓ -cohesive group may approve in expectation, where every Y_i follows π . That is (Equation (2)), t_ℓ is the smallest integer such that

$$n_\ell := \sum_{t=0}^{t_\ell} \binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n \geq \ell \cdot \frac{n}{k}.$$

On the other hand, by the definition of π , $\sum_{L_W \subseteq S, |S \cap W|=t} \pi_S = \binom{k}{t} p^{\ell+t} (1-p)^{k-t}$, and $\sum_{S \subseteq M} \pi_S = 1$. Therefore, the first two constraints hold. When π is assigned in the third constraint, the left-hand side becomes $\ell/k \cdot u_\ell$, which is ℓ^2/k due to our choice of ℓ with $u_\ell = \ell$, while the right-hand side is exactly ℓ^2/k . Therefore, the third constraint holds. $\square$

Claim 5.7. *For all sufficiently large n , $\mathcal{H}$ has a inner point $\vec{X}_n$ such that $\|\vec{X}_n\|_1 = n$, and all entries in $\vec{X}_n$ are integers. That is, $\mathcal{H}_n^{\mathbb{Z}} \neq \emptyset$.*

Proof. Our construction of $\vec{X}_n$ consists of two steps. We first construct $\vec{X}'_n$, which is an inner point of $\mathcal{H}$ and has $\|\vec{X}'_n\|_1 = n$. Then we round $\vec{X}'_n$ to get the integer point $\vec{X}_n$.

$\vec{X}'_n$ is as follows. For each $t = 0, 1, 2, \dots, m$ and each S with $|S| = t$,

- If $t = \ell$, $x'_S = p^t \cdot (1-p)^{m-t} \cdot n + t_\ell \cdot 2^m$.
- If $t = 0$, $x'_S = p^t \cdot (1-p)^{m-t} \cdot n - t_\ell \cdot 2^m \cdot \binom{m}{\ell}$.
- For all other t , $x'_S = p^t \cdot (1-p)^{m-t} \cdot n$.

Since m is a constant and $t_\ell \leq k \leq m$, we can guarantee that $x'_S \geq 0$ for all sufficiently large n . Note that $u_\ell \geq \ell$ directly implies that $t_\ell \geq \ell$.

$\vec{X}'_n$ is modified from the expectation of each x_S under π . For each set S of ℓ candidates, $t_\ell \cdot 2^m$ voters are switched from approving nobody to approving candidates in S only. Since we do not change the total number of voters, $\|\vec{X}'_n\|_1 = n$ holds.

Now we show that $\vec{X}'_n$ is an inner point of $\mathcal{H}$. After the modification, the left-hand side for all three constraints decreases, while the right-hand side remains unchanged. Therefore, $\vec{X}'_n$ strictly satisfies all constraints and is an inner point.

1. For constraint 1, $t_\ell \cdot 2^m$ more voters are approving each L_W compared to the expectation. Therefore, $\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x'_S \geq \ell/k \cdot \sum_{S \subseteq M} x_S + t_\ell \cdot 2^m > \ell/k \cdot \sum_{S \subseteq M} x_S$.
2. For constraint 2, by the definition of t_ℓ , we have $\sum_{t=0}^{t_\ell-1} \binom{k}{t} p^{\ell+t} \cdot (1-p)^{k-t} < \ell/k$. Therefore,

$$\sum_{t=0}^{t_\ell-1} \sum_{L_W \subseteq S, |S \cap W|=t} x'_S = \sum_{t=0}^{t_\ell-1} \binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n + t_\ell \cdot 2^m$$

$$\begin{aligned}
&= \ell/k \cdot n - \Theta(n) + t_\ell \cdot 2^m \\
&< \ell/k \cdot \sum_{S \subseteq M} x_S.
\end{aligned}$$

The first equality comes from that x'_S has $t_\ell \cdot 2^m$ more voters approving each L_W compared to the expectation. The last strict inequality holds for all sufficiently large n .

3. For constraint 3, for each group L_W , we compare the voters that approve L_W between the expectation and $\vec{X}'_n$. There are $t_\ell \cdot 2^m$ more voters that do not approve any candidates in W in $\vec{X}'_n$, while for each $1 \leq t \leq k$, the number of voters that approve t candidates in W is unchanged. Therefore, when the $\ell \cdot n/k$ (fractional) voters with the minimum utility are counted, $t_\ell \cdot 2^m$ voters with utility 0 substitute voters with the utility strictly larger than 0 compared to the expectation, with at least one voter with utility t_ℓ . Therefore, the new total utility is at most $\ell^2 \cdot n/k - t_\ell - t_\ell \cdot 2^m + 1 < \ell^2/k \cdot \sum_{S \subseteq M} x'_S - t_\ell - \frac{1}{k}$.

Now we round the fractional instance $\vec{X}'_n$ into an integer instance $\vec{X}_n$. Following the spirit of the intermediate value theorem, it is guaranteed that there exists a rounding such that $\|\vec{X}_n\|_1 = n$ and for every S , $x'_S - 1 \leq x_S \leq x'_S + 1$. Therefore,

$$\begin{aligned}
\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} |x_S - x'_S| &\leq \sum_{t=0}^k \sum_{L_W \subseteq S, |S \cap W|=t} 1 \\
&= \sum_{L_W \subseteq S} 1 \\
&= 2^{m-\ell},
\end{aligned}$$

which is a constant in our setting. Then we show that $\vec{X}_n$ is also an inner point of $\mathcal{H}$.

1. For constraint 1, the left-hand side difference between x'_S and x_S is at most $2^{m-\ell}$, while x'_S exceeds the threshold for at least $t_\ell \cdot 2^m$.

$$\begin{aligned}
\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x_S &\geq \sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x'_S - \sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} |x'_S - x_S| \\
&\geq \ell/k \cdot \sum_{S \subseteq M} x_S + t_\ell \cdot 2^m - 2^{m-\ell} \\
&> \ell/k \cdot \sum_{S \subseteq M} x_S.
\end{aligned}$$

2. For constraint 2, the left-hand side difference between x'_S and x_S is at most $2^{m-\ell}$, while the gap between the left-hand side on x'_S and the right-hand side is $\Theta(n)$. Therefore, for all sufficiently large n , (2) is satisfied, and the inequality is strict.
3. For constraint 3,

$$\sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} t \cdot x_S - \left(\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x_S - \ell \cdot n/k \right) \cdot t_\ell$$

$$\begin{aligned}
&= - \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} (t_\ell - t) \cdot x_S + t_\ell \cdot \ell \cdot n/k \\
&= - \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} (t_\ell - t) \cdot (x'_S - (x'_S - x_S)) + t_\ell \cdot \ell \cdot n/k \\
&\leq - \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} (t_\ell - t) \cdot x_S + t_\ell \cdot \ell \cdot n/k + (t_\ell - 1) \cdot 2^{m-\ell} \\
&= \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} t \cdot x'_S - \left(\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x'_S - \ell \cdot n/k \right) \cdot t_\ell + (t_\ell - 1) \cdot 2^{m-\ell} \\
&\leq \ell^2 \cdot n/k - t_\ell \cdot 2^m + 1 + (t_\ell - 1) \cdot 2^{m-\ell} \\
&< \ell^2/k \cdot \sum_{S \subseteq M} x_S - t_\ell - \frac{1}{k}.
\end{aligned}$$

The first inequality comes from that $t \geq 1$ so that $t_\ell - t \leq t_\ell - 1$ and $\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} |x_S - x'_S| \leq 2^{m-\ell}$. The second inequality comes from the satisfaction of (3) of x'_S .

Therefore, we have shown that $\vec{X}_n$ is an inner point of $\mathcal{H}$ for all sufficiently large n . $\square$

Claim 5.8. $\mathcal{H}_{\leq 0}$ is full-dimensional, i.e., $\dim(\mathcal{H}_{\leq 0}) = 2^m$.

Proof. By using a similar analysis, we can show that $\vec{X}_n$ is an inner point of $\mathcal{H}_{\leq 0}$. Therefore, $\mathcal{H}_{\leq 0}$ is full-dimensional. $\square$

With all three claims holding, we apply Corollary 5.3 and show that $\Pr[\vec{X} \in \mathcal{H}] = \Theta(1)$.

5.3 Proof of Proposition 5.5.

The proof of Proposition 5.5 resembles those of Proposition A.2 and 5.4. In this proof we fix a winning committee W . By using a similar analysis in Proposition A.2, we can show that the probability that W fails to provide AJR on any ℓ -cohesive groups with $p < \sqrt[\ell]{\ell/k}$ and the probability that W fails on any L with $L \cap W \neq \emptyset$ is $o(1)$. By Lemma A.3 and $t_\ell \geq 1$, we know that the minimum expected average utility among all ℓ -cohesive groups towards L with $L \cap W \neq \emptyset$ is strictly larger than ℓ ($u_{\ell,h}^* > u_{\ell-h} + h$, where $u_{\ell-h} \geq \ell - h$), which implies $o(1)$ in the likelihood of failing to provide AJR.

Then it remains to show that the likelihood that W fails on some candidate set L with $L \cap W = \emptyset$ is $1 - \Theta(1)$. We prove this through the polyhedron approach. Specifically, the polyhedron $\mathcal{H}'$ is characterized as follows. For each $\ell \in \min(k, m - k)$ with $p \geq \sqrt[\ell]{\ell/k}$, and each ℓ -candidate group L such that $L \cap W = \emptyset$, there are the following three constraints. In total, there will be at most $3 \sum_{\ell=1}^k \binom{m}{\ell} \leq 3 \cdot k \cdot \binom{m}{k}$ constraints.

$$\begin{aligned}
\sum_{t=0}^{t_\ell} \sum_{L \subseteq S, |S \cap W|=t} x_S &\geq \ell/k \cdot \sum_{S \subseteq M} x_S. \\
\sum_{t=0}^{t_\ell-1} \sum_{L \subseteq S, |S \cap W|=t} x_S &\leq \ell/k \cdot \sum_{S \subseteq M} x_S.
\end{aligned}$$

$$\begin{aligned} \sum_{t=1}^{t_\ell} \sum_{L \subseteq S, |S \cap W|=t} t \cdot x_S - \left(\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x_S - \ell/k \cdot \sum_{S \subseteq M} x_S \right) \cdot t_\ell \\ \geq \ell^2/k \cdot \sum_{S \subseteq M} x_S \end{aligned}$$

$\mathcal{H}'$ characterizes the following event: for every $\ell \in \min\{k, m-k\}$ and every group L ,

1. there exists a group with at least $\ell \cdot n/k$ voters, all of whom approve all candidates in L as well as at most t_ℓ candidates in W ;
2. the number of voters that approve all candidates in L as well as at most $t_\ell - 1$ candidates in W is at most $\ell \cdot n/k$; and
3. the total utility of $\ell \cdot n/k$ voters with the minimum utility (fractional voters accepted) is at least $\ell^2 \cdot n/k$. In this way, for the $\lceil \ell \cdot n/k \rceil$ voters with the minimum utility, the total expected utility will be at least $\ell^2 \cdot n/k + t_\ell \cdot (\lceil \ell \cdot n/k \rceil - \ell \cdot n/k)$. Given that $u_\ell \geq \ell$, there must be $t_\ell \geq \ell$, since t_ℓ is the largest expected single-voter utility in the group. Therefore, the total expected utility of $\lceil \ell \cdot n/k \rceil$ voters with the minimum utility will be at least $\ell \cdot \lceil \ell \cdot n/k \rceil$, which implies the average utility being at least ℓ .

Note that in the polyhedron, t_ℓ for different ℓ may be different.

Therefore, $\mathcal{H}'$ is a subevent of “ W does not fail to provide AJR on any non-overlapping set of ℓ -candidates with W and $p \geq \sqrt[\ell]{\ell/k}$ ”. Therefore, showing that $\Pr[\vec{X} \in \mathcal{H}'] = \Theta(1)$ implies that the probability W fails on some non-overlapping candidate set L is $1 - \Theta(1)$. Then applying the union bound on all three cases where W fails AJR, we have the probability that W fails AJR is at most $1 - \Theta(1) + o(1) + o(1) = 1 - \Theta(1)$, which implies our conclusion.

Now we prove the claims that apply Corollary 5.3. This resembles the proof of Proposition 5.4.

Claim 5.9. $\pi \in \mathcal{H}'_{\leq 0}$.

Proof. Note that $\mathcal{H}'$ does not contain any constant term. Therefore, $\mathcal{H}'_{\leq 0} = \mathcal{H}'$. The proof then resembles that of Claim 5.6. By definition, t_ℓ is the largest number of winners a voter in the underrepresented ℓ -cohesive group may approve in expectation where every y_j follows π . Therefore, constraints 1 and 2 hold. When π is assigned in constraint 3, the left-hand side becomes $\ell/k \cdot u_\ell$, which will not exceed ℓ^2/k , while the right-hand side is exactly ℓ^2/k . Therefore, constraint 3 holds. $\square$

Claim 5.10. For all sufficiently large n , $\mathcal{H}'$ has a inner point $\vec{X}_n$ such that $\|\vec{X}_n\|_1 = n$, and all entries in $\vec{X}_n$ are integers.

Proof. Our construction of $\vec{X}_n$ consists of two steps. We first construct $\vec{X}'_n$, which is a inner point of $\mathcal{H}'$ and has $\|\vec{X}'_n\|_1 = n$. Then we round $\vec{X}'_n$ to get the integer point $\vec{X}_n$.

$\vec{X}'_n$ is constructed by a two-step modification from the expectation. In the beginning, for every $S \subseteq M$, $x'_S = p^t \cdot (1-p)^{m-t} \cdot n$, where $t = |S|$. Then we do the following modification:

1. For $S = \emptyset$, $x'_S \leftarrow x'_S - k \cdot (2^m - 1)$; for all other S , $x'_S \leftarrow x'_S + k$. For each $S \neq \emptyset$, k voters approving nobody turn to approve S .
2. For all S with $|S \cap W| = 0$, $x'_S \leftarrow x'_S - k^3 \cdot 2^m$; for all S with $|S \cap W| = 1$, $x'_S \leftarrow x'_S + k^2 \cdot 2^m$. For each S with $S \cap W = \emptyset$ and each candidate $j \in W$, $k^2 \cdot 2^m$ voters approving exactly S additionally approve j .

Since both m and k are constants and $t_\ell \leq k$, we can guarantee that $x'_S \geq 0$ for all sufficiently large n . We do not change the total number of voters, so $\|\vec{X}'_n\|_1 = n$ holds. Now we show that $\vec{X}'_n$ is an inner point of $\mathcal{H}'$ by showing that for each L , all three constraints are satisfied.

1. For constraint 1, modification 1 brings k more voters approving at least all candidates in L , compared to the expectation, and modification 2 only transfers voters between different $S \supseteq L$ and does not affect the number of voters approving all candidates in L . Therefore, we have

$$\sum_{t=0}^{t_\ell} \sum_{L \subseteq S, |S \cap W|=t} x'_S \geq \ell/k \cdot \sum_{S \subseteq M} x_S + k \cdot \sum_{t=0}^{t_\ell} \sum_{L \subseteq S, |S \cap W|=t} 1 > \ell/k \cdot \sum_{S \subseteq M} x_S.$$

2. For constraint 2, by the definition of t_ℓ , there is $\sum_{t=0}^{t_\ell-1} \binom{k}{t} p^{\ell+t} \cdot (1-p)^{k-t} < \ell/k$. On the other hand, modification 1 brings at most $k \cdot (2^m - 1)$ more votes approving L . Therefore,

$$\begin{aligned} \sum_{t=0}^{t_\ell-1} \sum_{L \subseteq S, |S|=t+\ell} x'_S &= \sum_{t=0}^{t_\ell-1} \binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n + k \cdot (2^m - 1) \\ &= \ell/k \cdot n - \Theta(n) + k \cdot (2^m - 1) \\ &< \ell/k \cdot \sum_{S \subseteq M} x_S. \end{aligned}$$

The last strict inequality holds for all sufficiently large n .

3. For constraint 3, for each L , we compare the voters that approve L between the expectation and $\vec{X}'_n$. Modification 1 brings t_ℓ more voters approving L compared to the expectation. The total number of the increased voters will not exceed $k \cdot (2^m - 1)$. Modification 2 switches $k^3 \cdot 2^m$ voters with utility from 0 to 1. Therefore, the effect of the modification to the group V of ℓ/k voters with the minimum utilities is as follows. After modification 1, at most $k \cdot (2^m - 1)$ voters in V with utility t_ℓ are changed by voters with a lower utility, leading to a decrease of the total utility of at most $k^2 \cdot (2^m - 1)$. Then, after modification 2, $k^3 \cdot 2^m$ voters with utility 0 are switched to utility 1, implying an increase to the expected utility of $k^3 \cdot 2^m$. Therefore, the new expected utility after the modifications is at least $\ell^2 \cdot n/k + (k-1) \cdot k^2 \cdot 2^m > \ell^2/k \cdot \sum_{S \subseteq M} x_S$.

Now we round the fractional instance $\vec{X}'_n$ into integer instance $\vec{X}_n$. Following the spirit of the intermediate value theorem, there exists a rounding such that $\|\vec{X}_n\|_1 = n$ and for every S $x'_S - 1 \leq x_S \leq x'_S + 1$. Therefore, for each ℓ , $\sum_{t=0}^{t_\ell} \sum_{i \in S, |S \cap W|=t} |x_S - x'_S| \leq 2^{m-\ell}$, which is a constant in our setting. Then we show that $\vec{X}_n$ is also an inner point of $\mathcal{H}'$.

1. For constraint 1, the left-hand side difference between x'_S and x_S is at most $\sum_{t=0}^{t_\ell} \sum_{i \in S, |S \cap W|=t} 1$, while x'_S exceeds the threshold of $k \cdot \sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} 1$.

$$\begin{aligned} \sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x_S &\geq \sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x'_S - \sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S|=t+\ell} |x'_S - x_S| \\ &\geq \ell/k \cdot \sum_{S \subseteq M} x_S + (k-1) \cdot \sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} 1 \\ &> \ell/k \cdot \sum_{S \subseteq M} x_S. \end{aligned}$$

2. For constraint 2, the difference of the left-hand side between x'_S and x_S is at most $2^{m-\ell}$, while the gap between the left-hand side with respect to x'_S and the right-hand side is $\Theta(n)$. Therefore, for all sufficiently large n , constraint 2 is satisfied, and the inequality is strict.
3. For constraint 3,

$$\begin{aligned}
& \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} t \cdot x_S - \left(\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x_S - \ell \cdot n/k \right) \cdot t_\ell \\
&= - \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} (t_\ell - t) \cdot x_S + t_\ell \cdot \ell \cdot n/k \\
&= - \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} (t_\ell - t) \cdot (x'_S - (x'_S - x_S)) + t_\ell \cdot \ell \cdot n/k \\
&\geq - \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} (t_\ell - t) \cdot x_S + t_\ell \cdot \ell \cdot n/k - (t_\ell - 1) \cdot 2^{m-\ell} \\
&= \sum_{t=1}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} t \cdot x'_S - \left(\sum_{t=0}^{t_\ell} \sum_{L_W \subseteq S, |S \cap W|=t} x'_S - \ell \cdot n/k \right) \cdot t_\ell - (t_\ell - 1) \cdot 2^{m-\ell} \\
&\geq \ell^2 \cdot n/k + (k-1) \cdot k^2 \cdot 2^m - (t_\ell - 1) \cdot 2^{m-\ell} \\
&> \ell^2/k \cdot \sum_{S \subseteq M} x_S.
\end{aligned}$$

The first inequality comes from that $t \geq 1$ so that $t_\ell - t \leq t_\ell - 1$ and $\sum_{t=0}^{t_\ell} \sum_{i \in S, |S|=t+\ell} |x_S - x'_S| \leq 2^{m-\ell}$. The second inequality comes from the satisfaction of constraint 3 of x'_S .

Therefore, we have shown that $\vec{X}_n$ is an inner point of $\mathcal{H}'$ for all sufficiently large n . $\square$

Claim 5.11. $\mathcal{H}'_{\leq 0}$ is full-dimensional. $\dim(\mathcal{H}'_{\leq 0}) = 2^m$.

Proof. Since $\mathcal{H}' = \mathcal{H}'_{\leq 0}$, $\vec{X}_n$ is an inner point of $\mathcal{H}'_{\leq 0}$. Therefore, $\mathcal{H}'_{\leq 0}$ is full-dimensional. $\square$

With all three claims, we apply Corollary 5.3 and show that $\Pr[\vec{X}_\pi \in \mathcal{H}] = \Theta(1)$.

Consequently, the likelihood that W fails AJR on ℓ -candidate set L with $p < \sqrt[\ell]{\frac{\ell}{k}}$ is $o(1)$, on L with $L \cap W \neq \emptyset$ is $o(1)$, and on all the other L is $1 - \Theta(1)$. Therefore, the likelihood of W providing AJR is $\Theta(1)$, which completes the proof.

6 Conclusion and Future Directions

In this paper, we give a complete characterization of the substantial likelihood of an AJR committee when instances are sampled from the Erdős-Rényi bipartite model. Our results not only theoretically verify the previous empirical observations, such as Brill et al. [7, 9], but also give more detailed structural insights.

One natural future direction is to see if our results extend to other random models that are more general and/or more practical. The Erdős-Rényi bipartite model, albeit mathematically simple and natural, can be inaccurate when describing the real-world election instances, as the

events that voters approve candidates can be dependent and with different probabilities. Given that analyzing the seemingly simplest Erdős-Rényi model is already quite technically involved, it is expected that studying the likelihood of AJR committees’ existence on other more general/realistic models is even more challenging. Nevertheless, our paper moves a first step towards this goal, and our observations and techniques can be potentially useful for future studies along this direction. Additionally, exploring the existence of winning committees satisfying other ideal properties, such as core stability, might provide valuable insights into tackling the challenge in the classical multi-winner approval voting, since it is still an open problem whether the core stability can always be satisfied in multi-winner approval voting.

In addition, this work can be extended along the direction of proportionality degree. Instead of focusing on the particular choice $f(\ell) = \ell$ under the AJR definition, we can study the distribution of the proportionality degree, and this is interesting even under the simplest Erdős-Rényi model. For example, how does the average satisfaction for an ℓ -cohesive group increase above $f(\ell) = \ell$ as p increases in $(p_2^*, 1]$ before uniformly equals k at $p = 1$? Also, given that PAV guarantees $f(\ell) = \ell - 1$, is this tight under the Erdős-Rényi model? In particular, between p_1^* and p_2^* where we have shown $f(\ell) = \ell$ is unlikely to hold, are there some values of p where the minimum average satisfaction is concentrated around $\ell - \frac{1}{2}$? For one step towards the last question, our Proposition 4.2 and Lemma 3.1 imply that a cohesive witness to the violation of $f(\ell) = \ell$ can only be a 1-cohesive group (for all values of p). Thus, we only need to analyze 1-cohesive groups for this question with $\ell = 1$.

We have also mentioned after Lemma 3.1 and in Proposition 4.2 that 1-cohesive groups are the only barriers that prevent the existence of AJR committees. It is interesting to see how broadly this observation can be applied. Is it a feature that is unique for Erdős-Rényi model?

Finally, our results can be further refined in the following aspects. We can consider the extension to the general setting where m and k are not necessarily constants. Our results can also be refined by studying the “rates” in the probabilities at the two transition points. For example, how fast does the probability change when p is approaching each of the two transition points?

Acknowledgments

The research of Biaoshuai Tao is supported by the National Natural Science Foundation of China (No. 62472271). The research of Lirong Xia is supported by NSF 2450124, 2517733, and 2518373.

The authors gratefully thank the anonymous reviewers from SODA’26, FOCS’25, EC’25, and WINE’24 for their valuable suggestions.

References

- [1] Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48 (2):461–485, 2017.
- [2] Haris Aziz, Edith Elkind, Shenwei Huang, Martin Lackner, Luis Sánchez Fernández, and Piotr Skowron. On the complexity of extended and proportional justified representation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI*, pages 902–909. AAAI Press, 2018.
- [3] Haris Aziz, Edith Elkind, Shenwei Huang, Martin Lackner, Luis Sánchez-Fernández, and

- Piotr Skowron. On the complexity of extended and proportional justified representation. In *Proceedings of the AAAI Conference on Artificial Intelligence, AAAI*, volume 32, 2018.
- [4] Robert Brederick, Piotr Faliszewski, Andrzej Kaczmarczyk, and Rolf Niedermeier. An experimental view on committees providing justified representation. In *IJCAI*, pages 109–115, 2019.
 - [5] Markus Brill and Jannik Peters. Robust and verifiable proportionality axioms for multiwinner voting. In *Proceedings of the 24th ACM Conference on Economics and Computation, EC*, page 301. ACM, 2023.
 - [6] Markus Brill, Rupert Freeman, Svante Janson, and Martin Lackner. Phragmén’s voting methods and justified representation. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI*, pages 406–413. AAAI Press, 2017.
 - [7] Markus Brill, Jonas Israel, Evi Micha, and Jannik Peters. Individual representation in approval-based committee voting. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI*, pages 4892–4899. AAAI Press, 2022.
 - [8] Markus Brill, Rupert Freeman, Svante Janson, and Martin Lackner. Phragmén’s voting methods and justified representation. *Mathematical Programming*, 203(1):47–76, 2024.
 - [9] Markus Brill, Jonas Israel, Evi Micha, and Jannik Peters. Individual representation in approval-based committee voting. *Social Choice and Welfare*, 64(1):69–96, 2025.
 - [10] Edith Elkind, Piotr Faliszewski, Ayumi Igarashi, Pasin Manurangsi, Ulrike Schmidt-Kraepelin, and Warut Suksompong. Justifying groups in multiwinner approval voting. *Theoretical Computer Science*, 969:114039, 2023.
 - [11] Luis Sánchez Fernández, Edith Elkind, Martin Lackner, Norberto Fernández García, Jesús Arias-Fisteus, Pablo Basanta-Val, and Piotr Skowron. Proportional justified representation. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI*, pages 670–676. AAAI Press, 2017.
 - [12] Donald B Gillies. Solutions to general non-zero-sum games. *Contributions to the Theory of Games*, 4(40):47–85, 1959.
 - [13] Lukasz Janeczko and Piotr Faliszewski. The complexity of proportionality degree in committee elections. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI*, pages 5092–5099. AAAI Press, 2022.
 - [14] Zhihao Jiang, Kamesh Munagala, and Kangning Wang. Approximately stable committee selection. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC*, pages 463–472. ACM, 2020.
 - [15] Yusuf Hakan Kalayci, Jiasen Liu, and David Kempe. Full proportional justified representation. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems, AAMAS*, page 1070–1078, 2025.
 - [16] Martin Lackner and Piotr Skowron. *Multi-Winner Voting with Approval Preferences*. Springer Briefs in Intelligent Systems. Springer, 2023.

- [17] Ivan-Aleksandar Mavrov, Kamesh Munagala, and Yiheng Shen. Fair multiwinner elections with allocation constraints. In *Proceedings of the 24th ACM Conference on Economics and Computation, EC*, pages 964–990. ACM, 2023.
- [18] Kamesh Munagala, Yiheng Shen, Kangning Wang, and Zhiyi Wang. Approximate core for committee selection via multilinear extension and market clearing. In *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 2229–2252. SIAM, 2022.
- [19] Dominik Peters, Grzegorz Pierczynski, Nisarg Shah, and Piotr Skowron. Market-based explanations of collective decisions. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI*, pages 5656–5663. AAAI Press, 2021.
- [20] Dominik Peters, Grzegorz Pierczyński, and Piotr Skowron. Proportional participatory budgeting with additive utilities. *Advances in Neural Information Processing Systems*, 34: 12726–12737, 2021.
- [21] Friedrich Pukelsheim. Quota methods of apportionment: Divide and rank. In *Proportional Representation: Apportionment Methods and Their Applications*, pages 95–105. Springer, 2017.
- [22] Lloyd S Shapley and Martin Shubik. On market games. *Journal of Economic Theory*, 1(1): 9–25, 1969.
- [23] Piotr Skowron. Proportionality degree of multiwinner rules. In *Proceedings of the 22nd ACM Conference on Economics and Computation, EC*, pages 820–840. ACM, 2021.
- [24] Stanislaw Szufa, Piotr Faliszewski, Lukasz Janeczko, Martin Lackner, Arkadii Slinko, Krzysztof Sornat, and Nimrod Talmon. How to sample approval elections? In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI*, pages 496–502. ijcai.org, 2022.
- [25] Biaoshuai Tao, Chengkai Zhang, and Houyu Zhou. The degree of (extended) justified representation and its optimization. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems, AAMAS*, pages 2024–2032, 2025.
- [26] Lirong Xia. How likely are large elections tied? In *Proceedings of the 22nd ACM Conference on Economics and Computation, EC*, pages 884–885. ACM, 2021.
- [27] Lirong Xia. A linear theory of multi-winner voting. *CoRR*, abs/2503.03082, 2025. URL <https://doi.org/10.48550/arXiv.2503.03082>.

A Proofs for the First Two Parts of Lemma 3.1

A.1 Proof for the First Part: the Case $u_\ell > \ell$

We first deal with the easy cases where $p < \frac{1}{k}$ and where $p = 1$. For the former, we formalize the analysis in the paragraph *Identifying the threshold for the existence of cohesive groups* before Lemma 3.1 into the following lemma.

Lemma A.1. *For every $k, \ell \leq k, p < \sqrt[\ell]{\frac{\ell}{k}}$, and n , the probability that there exists an ℓ -size candidate set L and a group of voter V such that V is an ℓ -cohesive group towards L is $o(1)$.*

Proof. For a fixed ℓ and a fixed group L of ℓ candidates, the expected number of agents approving all candidates in L is np^ℓ , while a cohesive group towards L exists if and only if at least $\ell \cdot \frac{n}{k}$ approve all candidates in L . When $p < \sqrt[\ell]{\frac{\ell}{k}}$, we have $np^\ell < \ell \cdot \frac{n}{k}$. Therefore, we could apply the Hoeffding inequality as follows. Let Z_i be the random variable where $Z_i = 1$ if and only if agent i approves all candidates in L .

$$\begin{aligned} \Pr[\text{A cohesive group for } L \text{ exists}] &= \Pr\left[\sum_{i=1}^n Z_i \geq \ell \cdot \frac{n}{k}\right] \\ &= \Pr\left[\sum_{i=1}^n Z_i - \mathbb{E}\left[\sum_{i=1}^n Z_i\right] \geq \ell \cdot \frac{n}{k} - np^\ell\right] \\ &\leq \exp\left(-2 \left(\frac{\ell}{k} - p^\ell\right)^2 n\right) \\ &= \exp(-\Theta(n)). \end{aligned}$$

Then, by applying a union bound on all $\binom{m}{\ell}$ of ℓ -size candidate sets (where m is a constant), the probability that an ℓ -cohesive group exists is at most $\binom{m}{\ell} \cdot \exp(-\Theta(n)) = o(1)$. $\square$

Then notice that $\sqrt[\ell]{\frac{\ell}{k}}$ is minimized at $k = 1$. Therefore, when $p < \frac{1}{k}, p < \sqrt[\ell]{\frac{\ell}{k}}$ holds for any $\ell = 1, 2, \dots, k$. In this case, probability $1 - o(1)$, no cohesive group exists, in which case any winning committee automatically satisfies AJR.

For the latter case where $p = 1$, any winning committee gives each voter utility k , in which case AJR is also satisfied.

To conclude the first part of the lemma, we will prove the following proposition.

Proposition A.2. *For any constant m, k, p , if $u_\ell > \ell$ for every $\ell \leq \min\{k, m - k\}$ such that $p^\ell \geq \frac{\ell}{k}$, then the likelihood that an AJR committee exists is $1 - o(1)$.*

Proof. We fix the winning committee $W = \{1, 2, \dots, k\}$, and we will prove that W satisfies AJR with high probability. The event that W fails AJR can be divided into three scenarios: (1) when $p < \sqrt[\ell]{\frac{\ell}{k}}$ (in which case $p^\ell < \frac{\ell}{k}$ and we are not guaranteed $u_\ell > \ell$), an ℓ -cohesive group is underrepresented, (2) when $p \geq \sqrt[\ell]{\frac{\ell}{k}}$, an ℓ -cohesive group towards L with $L \cap W = \emptyset$ is underrepresented, and (3) when $p \geq \sqrt[\ell]{\frac{\ell}{k}}$, an ℓ -cohesive group towards L with $L \cap W \neq \emptyset$ is underrepresented. We show that each scenario has a likelihood $o(1)$ to occur. Therefore, according to the union bound, the probability that W fails to provide AJR is $o(1)$.

Case 1. Lemma A.1 directly implies that the probability that such an ℓ -cohesive group exists when $p < \sqrt[\ell]{\frac{\ell}{k}}$ is $o(1)$.

Case 2. For each $t = 0, 1, \dots, k$, let V_t be the group of voters who approve every candidate in L as well as exactly t candidates in W . Since $L \cap W = \emptyset$, the expectation of $|V_t|$ is $\binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n$. For each t and any constant $\varepsilon > 0$, by applying the Hoeffding's inequality, the probability that $||V_t| - \mathbb{E}[|V_t|]| \geq \varepsilon n$ is at most $\Theta(\exp(-n))$. That is, the size of V_t concentrates to its expectation with high probability. Applying the union bound on all $t = 0, 1, \dots, k$, the probability that there exists a t such that $|V_t|$ deviates from its expectation by at least εn is also at most $\Theta(\exp(-n))$.

Conditioned on that all $|V_t|$ concentrates to their expectations respectively, we can lower bound the average utility of a group V with high probability. A group V with the minimum average utility can be generated by greedily fetching all voters in V_0 , then $V_1, V_2, \dots$ until it has at least $\ell \cdot \frac{n}{k}$ voters. For a sufficiently small constant ε (which does not depend on n), since each $|V_t|$ does not deviate from expectation for more than εn , the average utility of V will be at least $u_\ell - \frac{k}{\ell \cdot n} \cdot k \cdot \varepsilon n \cdot k = u_\ell - \frac{k^3}{\ell} \cdot \varepsilon$ (at most k groups of V_t , each group has at most $\varepsilon \cdot n$ more voters than expectation, substituting voters with an utility of at most k). Note that in case 2, we have $u_\ell > \ell$. This is because (1) $p^\ell \geq \frac{\ell}{k}$, (2) $\ell \leq k$ by definition, and (3) $L \cap W = \emptyset$ implies that $\ell \leq m - k$, which together guarantees that $u_\ell > \ell$ by the statement of Proposition A.2. Therefore, by taking an $\varepsilon < (u_\ell - \ell) \cdot \frac{\ell}{k^3}$, the probability that there exists an ℓ -cohesive group V towards L such that the average utility of V is strictly less than ℓ is bounded by the likelihood where some $|V_t|$ deviates from the expectation more than $\varepsilon \cdot n$, leading to $\Theta(\exp(-n))$.

Case 3. For ℓ -cohesive groups towards L with $L \cap W \neq \emptyset$, the following lemma guarantees that the average utilities of those groups are strictly larger than ℓ with high probability.

Lemma A.3. *For any constant m, k, p , for every $1 \leq h < \ell \leq k$ such that $p^\ell \geq \frac{\ell}{k}$, we have $u_{\ell,h}^* > u_{\ell-h} + h$ where $u_{\ell,h}^*$ is the minimum expected average utility among all ℓ -cohesive groups towards L such that $|L \cap W| = h$.*

Proof. Consider the ℓ -cohesive group V_1 with $\ell \cdot n/k$ voters. The expected number of voters that approve L (with $|L \cap W| = h$) as well as exactly t winners in W is $\binom{k-h}{t-h} \cdot p^{\ell+t-h} \cdot (1-p)^{k-t} \cdot n$. Let t_ℓ^* be the smallest integer such that

$$n_\ell^* := \sum_{t_\ell^*=0}^{t_\ell} \binom{k-h}{t-h} \cdot p^{\ell+t-h} \cdot (1-p)^{k-t} \cdot n \geq \ell \cdot \frac{n}{k}.$$

Then we have

$$u_{\ell,h}^* = \frac{k}{\ell \cdot n} \cdot \left(\sum_{t=h}^{t_\ell^*} t \cdot \binom{k-h}{t-h} \cdot p^{\ell+t-h} \cdot (1-p)^{k-t} \cdot n - t_\ell^* \cdot (n_\ell^* - \ell \cdot n/k) \right).$$

On the other hand, $u_{\ell-h}$ is the minimum expected average utility among all $(\ell-h)$ -cohesive group supporting some L' such that $L' \cap W = \emptyset$,

$$u_{\ell-h} = \frac{k}{(\ell-h) \cdot n} \cdot \left(\sum_{t=0}^{t_{\ell-h}} t \cdot \binom{k}{t} \cdot p^{\ell+t-h} \cdot (1-p)^{k-t} \cdot n - t_{\ell-h} \cdot (n_{\ell-h} - (\ell-h) \cdot n/k) \right).$$

Next, we construct u'_ℓ satisfying $u_{\ell-h}^* + h < u'_\ell < u_\ell^*$. Let t'_ℓ be the smallest integer such that

$$n'_\ell = \sum_{t=h}^{t'_\ell} \binom{k}{t-h} \cdot p^{\ell+t-2h} \cdot (1-p)^{k-t-h} \cdot n \geq \frac{\ell \cdot n}{k}.$$

First, we show that

$$u'_\ell = \frac{k}{\ell \cdot n} \cdot \left(\sum_{t=h}^{t'_\ell} t \cdot \binom{k}{t-h} \cdot p^{\ell+t-2h} \cdot (1-p)^{k-t-h} \cdot n - t'_\ell \cdot (n'_\ell - \ell \cdot n/k) \right) < u_{\ell,h}.$$

To verify this, u'_ℓ can be viewed as the minimum expected average utility of an ℓ -cohesive group V_3 such that the expected number of voters that approve t winners is $\binom{k}{t-h} \cdot p^{\ell+t-2h} \cdot (1-p)^{k-t-h} \cdot n$. For each $h \leq t \leq t'_\ell$, the number of voters that approve t winners in V_3 is more than that in V_1 . In addition, both $u_{\ell,h}$ and u'_ℓ consider $\ell \cdot \frac{n}{k}$ voters. Hence, we have $u'_\ell < u_{\ell,h}$.

Then, we show that $u'_\ell > u_{\ell-h}^* + h$. Note that u'_ℓ can be reformulated as

$$\begin{aligned} u'_\ell &= \frac{k}{\ell \cdot n} \cdot \left(\sum_{t=0}^{t'_\ell-h} (t+h) \cdot \binom{k}{t} \cdot p^{\ell+t-h} \cdot (1-p)^{k-t} \cdot n - t'_\ell \cdot (n'_\ell - \ell \cdot n/k) \right) \\ &= \frac{k}{\ell \cdot n} \cdot \left(\sum_{t=0}^{t'_\ell-h} t \cdot \binom{k}{t} \cdot p^{\ell+t-h} \cdot (1-p)^{k-t} \cdot n - (t'_\ell - h) \cdot (n'_\ell - \ell \cdot n/k) \right) + h, \end{aligned}$$

where the first term can be seen as the average utility of $\ell \frac{n}{k}$ voters with lower utilities in which there are $\binom{k}{t} \cdot p^{\ell+t-h} \cdot (1-p)^{k-t} \cdot n$ voters approving t winners, which is same as V_2 . However, $u_{\ell-h}$ only considers $(\ell-h) \cdot \frac{n}{k}$ voters with lower utilities, which implies that $u_{\ell,h}^* > u_{\ell-h} + h$. $\square$

Consider a set L of ℓ candidates with $|L \cap W| = h$. If $h = \ell$, every voter approving L will have the utility of at least ℓ , and then AJR is satisfied. When $h < \ell$, we show that the likelihood that AJR fails on an ℓ -cohesive group approving L will be at most the likelihood of AJR fails on an $(\ell-h)$ -cohesive group approving $L \setminus W$, which implies $o(1)$. From $|L \cap W| = h$, since $p \geq \sqrt{\frac{\ell}{k}} > \sqrt{\frac{\ell-h}{k}}$, we have that the minimum expected average utility $u_{\ell-h}$ among all $(\ell-h)$ -cohesive groups towards $L \setminus W$ is strictly larger than $\ell-h$ (by the condition in Proposition A.2 and noticing $\ell-h \leq m-k$). Then by Lemma A.3, the minimal expected average utility of an ℓ -cohesive group towards L is strictly larger than ℓ . Next, following a similar analysis with Case 2, we have that the likelihood that AJR fails on an ℓ -cohesive group towards L is at most $o(1)$.

Now we apply the union bound on all the cases. Since m, k, ℓ, p are either constants or bounded by constants, the likelihood that AJR fails is $o(1)$. $\square$

A.2 Proof for the Second Part: the Case $u_\ell < \ell$

For the second part of the lemma, we will prove the following proposition.

Proposition A.4. *For any constant m, k, p , if there exists ℓ with $\ell \leq \min\{k, m-k\}$ such that $p^\ell \geq \frac{\ell}{k}$ and $u_\ell < \ell$, then the likelihood that an AJR committee exists is $o(1)$.*

Proof. When $p = \sqrt{\frac{\ell}{k}}$ for some ℓ , u_ℓ is guaranteed to be at least ℓ , which is proved in the lemma below.

Lemma A.5. *For any k, ℓ , when $p = \sqrt{\frac{\ell}{k}}$, we have $u_\ell \geq \ell$.*

Proof. When $\ell = k$, we have $p = 1$, and the average utility for any subset of voters is exactly $k = \ell$, implying that $u_\ell = \ell$. We assume $\ell < k$ from now on. Since $\sum_{t=0}^{t_\ell} \binom{k}{t} \cdot p^t \cdot (1-p)^{k-t} \leq 1$ and the equality is met if and only if $t_\ell = k$, we have that $n_\ell = \sum_{t=0}^{t_\ell} \binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n \geq p^\ell \cdot n = \ell \cdot \frac{n}{k}$.

if and only if $t_\ell = k$. By our definitions of t_ℓ and n_ℓ in Equation (2), we have $t_\ell = k$ and $n_\ell = \ell \cdot \frac{n}{k}$, and

$$\begin{aligned}
u_\ell &= \frac{k}{\ell \cdot n} \cdot \left(\sum_{t=0}^k t \cdot \binom{k}{t} \cdot p^{\ell+t} \cdot (1-p)^{k-t} \cdot n \right) \\
&= \frac{k}{\ell} \cdot p^\ell \left(\sum_{t=1}^k t \cdot \binom{k}{t} \cdot p^t \cdot (1-p)^{k-t} \right) && \text{(the first term is 0)} \\
&= \frac{k}{\ell} \cdot p^\ell \left(\sum_{t=1}^k k \cdot \binom{k-1}{t-1} p^t (1-p)^{k-t} \right) && \text{(rewrite the combinatorial number)} \\
&= \frac{k^2}{\ell} \cdot p^{\ell+1} \left(\sum_{t=0}^{k-1} \binom{k-1}{t} p^t (1-p)^{k-1-t} \right) \\
&= \sqrt[\ell]{\ell k^{\ell-1}} && \text{(since } p = \sqrt[\ell]{\frac{\ell}{k}} \text{ and the summation is } (p + (1-p))^{k-1} = 1) \\
&\geq \ell, && \text{(since } k > \ell \text{ and } \ell \geq 1)
\end{aligned}$$

which concludes the lemma. $\square$

To conclude Proposition A.4, we only need to consider the case where $p^\ell > \frac{\ell}{k}$. For each winning committee W , we fix an ℓ -candidate set L_W such that $L_W \cap W = \emptyset$. Then we show that W is likely to fail AJR on an ℓ -cohesive group towards L_W . Finally, we apply the union bound on all W and show that the likelihood that every W fails AJR on L_W is also $1 - o(1)$.

This proof is similar to Case 2 in the proof of Proposition A.2. For each $t = 0, 1, \dots, k$, let V_t be the group of voters who approve all candidates in L as well as exactly t candidates in W . For some sufficiently small ε , the probability where there exists a $t \in [t_\ell]$ such that $|V_t|$ deviates from its expectation by at least $\varepsilon \cdot n$ is at most $\Theta(\exp(-n))$.

Conditioned on that no $|V_t|$ deviates much, we consider the average utility of the group by picking voters from $V_0, V_1, \dots$ until it has at least $\ell \cdot \frac{n}{k}$ voters. For sufficiently small ε , since each $|V_t|$ does not deviate by more than εn , the average utility of V will be at most $u_\ell + \frac{k}{\ell \cdot n} \cdot k \cdot \varepsilon n \cdot k = u_\ell + \frac{k^3}{\ell} \cdot \varepsilon$ (at most k groups of V_t , each group has at most $\varepsilon \cdot n$ less voters that are substituted by voters with a utility of at most k). Therefore, for sufficiently small ε , the probability that there exists an ℓ -cohesive group V towards L_W with an average utility strictly less than ℓ is at least $1 - o(1)$. Thus, the probability that an arbitrary given winning committee W satisfies AJR is $o(1)$. By a union bound on all $\binom{m}{k}$ possible winning committees (and notice that $\binom{m}{k}$ is a constant), the probability that there exists an AJR committee is $o(1)$. $\square$

B Missing Proofs in Sect. 4

B.1 Proof of Proposition 4.1

Note that $u_\ell = U(t_\ell)$. Let $C_t = \binom{k}{t} p^{\ell+t} (1-p)^{k-t}$. For all $T < t_\ell$,

$$\begin{aligned}
U(T) &= \frac{k}{\ell} \left(\sum_{t=0}^T t C_t - T \left(\sum_{t=0}^T C_t - \frac{\ell}{k} \right) \right) \\
&= \frac{k}{\ell} \left(\sum_{t=0}^T t C_t + \sum_{t=T+1}^{t_\ell} t C_t - \sum_{t=T+1}^{t_\ell} t C_t - T \left(\sum_{t=0}^T C_t - \frac{\ell}{k} \right) \right)
\end{aligned}$$

$$\begin{aligned}
&< \frac{k}{\ell} \left(\sum_{t=0}^T tC_t + \sum_{t=T+1}^{t_\ell} tC_t - T \sum_{t=T+1}^{t_\ell} C_t - T \left(\sum_{t=0}^T C_t - \frac{\ell}{k} \right) \right) \\
&= U(t_\ell).
\end{aligned}$$

For all $T > t_\ell$,

$$\begin{aligned}
U(T) &= \frac{k}{\ell} \left(\sum_{t=0}^T tC_t - T \left(\sum_{t=0}^T C_t - \frac{\ell}{k} \right) \right) \\
&= \frac{k}{\ell} \left(\sum_{t=0}^T tC_t - \sum_{t=t_\ell+1}^T tC_t + \sum_{t=t_\ell+1}^T tC_t - T \left(\sum_{t=0}^T C_t - \frac{\ell}{k} \right) \right) \\
&< \frac{k}{\ell} \left(\sum_{t=0}^T tC_t - \sum_{t=t_\ell+1}^T tC_t + T \sum_{t=t_\ell+1}^T C_t - T \left(\sum_{t=0}^T C_t - \frac{\ell}{k} \right) \right) \\
&= U(t_\ell).
\end{aligned}$$

We conclude the proposition.

B.2 Proof of Proposition 4.4

The proof consists of four parts:

1. we first prove that the inequality holds for small k , with $k \leq \frac{\sqrt{5}+1}{2}\ell$,
2. next, we prove that the inequality holds for large k , with $k \geq 29\ell$,
3. for the middle regime $\frac{\sqrt{5}+1}{2}\ell \leq k \leq 29\ell$, we show that the inequality holds for sufficiently large ℓ , with $\ell \geq 3947$,
4. finally, the remaining values of (k, ℓ) with $\ell < 3947$ and $\frac{\sqrt{5}+1}{2}\ell \leq k \leq 29\ell$ are checked by computer programs.

Proposition B.1. $\binom{k}{\ell} \left(1 - \sqrt[\ell]{\frac{\ell}{k}} \right)^{k-\ell} < 1$ holds when $k \leq \frac{\sqrt{5}+1}{2}\ell$.

Proof. Since $\ln(1-x) < -x$ for $x \in (0, 1)$, substituting $x = \frac{1}{\ell} \ln \frac{k}{\ell}$ into the expression, we have

$$\ln \left(1 - \frac{1}{\ell} \ln \frac{k}{\ell} \right) < -\frac{1}{\ell} \ln \frac{k}{\ell},$$

where the inequality holds when $k \leq \ell e^\ell$. Exponentiating both sides, we have

$$1 - \frac{1}{\ell} \ln \frac{k}{\ell} < e^{-\frac{1}{\ell} \ln \frac{k}{\ell}} = \sqrt[\ell]{\frac{\ell}{k}}.$$

Combining with $\binom{k}{\ell} \leq k^{k-\ell}$, we further have

$$\binom{k}{\ell} \left(1 - \sqrt[\ell]{\frac{\ell}{k}} \right)^{k-\ell} < k^{k-\ell} \cdot \left(\frac{1}{\ell} \ln \frac{k}{\ell} \right)^{k-\ell}.$$

Since $\ln(1+x) < x$ for $x > 0$, substituting $x = \frac{k}{\ell} - 1$ into the expression, we have

$$k^{k-\ell} \cdot \left(\frac{1}{\ell} \ln \frac{k}{\ell} \right)^{k-\ell} < k^{k-\ell} \cdot \left(\frac{1}{\ell} \cdot \frac{k-\ell}{\ell} \right)^{k-\ell} = \left(\frac{k(k-\ell)}{\ell^2} \right)^{k-\ell}.$$

When $k \leq \frac{\sqrt{5}+1}{2}\ell$, we have

$$\frac{k(k-\ell)}{\ell^2} \leq \frac{\sqrt{5}+1}{2} \left(\frac{\sqrt{5}+1}{2} - 1 \right) = 1,$$

which completes the proof. $\square$

Proposition B.2. $\binom{k}{\ell} \left(1 - \sqrt[\ell]{\frac{\ell}{k}} \right)^{k-\ell} < 1$ holds when $k \geq 29\ell$.

Proof. Since $\ln(1-x) < -x$ for $x \in (0, 1)$, substituting $x = \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}}$ into the expression, we have

$$\ln \left(1 - \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}} \right) < - \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}}.$$

Multiplying both sides by $(k-\ell)$ and exponentiating both sides, we further have

$$\left(1 - \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}} \right)^{k-\ell} < \exp \left(- (k-\ell) \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}} \right).$$

Since $\binom{k}{\ell} \leq \left(\frac{ek}{\ell} \right)^\ell$, we only need to show

$$\left(\frac{ek}{\ell} \right)^\ell \cdot \exp \left(- (k-\ell) \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}} \right) \leq 1,$$

which is equivalent to

$$\ell(1 + \ln k - \ln \ell) - k \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}} + \ell \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}} \leq 0.$$

Since $\ell \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}} < \ell$ and let $t = \frac{k}{\ell}$, we have

$$\ell(1 + \ln k - \ln \ell) - k \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}} + \ell \left(\frac{\ell}{k} \right)^{\frac{1}{\ell}} < 2\ell + \ell \ln t - \ell \cdot t^{1-\frac{1}{\ell}} \leq (2 + \ln t - t^{\frac{1}{2}})\ell,$$

and it suffices to show $(2 + \ln t - t^{\frac{1}{2}})\ell < 0$.

Since $(2 + \ln t - t^{\frac{1}{2}})$ is monotonically decreasing with respect to $t \geq 4$ and $2 + \ln 29 - 29^{\frac{1}{2}} < 0$, we have $(2 + \ln t - t^{\frac{1}{2}})\ell < 0$ when $t \geq 29$, i.e., $k \geq 29\ell$. $\square$

Proposition B.3. $\binom{k}{\ell} \left(1 - \sqrt[\ell]{\frac{\ell}{k}} \right)^{k-\ell} < 1$ holds when $\frac{\sqrt{5}+1}{2}\ell < k < 29\ell$ and $\ell \geq 3947$.

Proof. First, we can use a similar analysis as Proposition B.1 and Proposition B.2, we have $\binom{k}{\ell} \leq \left(\frac{ek}{\ell}\right)^\ell$ and $\left(1 - \sqrt[\ell]{\frac{\ell}{k}}\right)^{k-\ell} < \left(\frac{1}{\ell} \ln \frac{k}{\ell}\right)^{k-\ell}$. Thus,

$$\binom{k}{\ell} \left(1 - \sqrt[\ell]{\frac{\ell}{k}}\right)^{k-\ell} < \left(\frac{ek}{\ell}\right)^\ell \cdot \left(\frac{1}{\ell} \ln \frac{k}{\ell}\right)^{k-\ell} = \left(\left(\frac{ek}{\ell}\right) \cdot \left(\frac{1}{\ell} \ln \frac{k}{\ell}\right)^{\frac{k}{\ell}-1}\right)^\ell.$$

Letting $t = \frac{k}{\ell}$, we only need to show

$$et \left(\frac{1}{\ell} \ln t\right)^{t-1} < 1.$$

Since $\ell \geq 3947$ and $t \leq 29$, we have $\frac{1}{\ell} \ln t < 1$. Hence,

$$et \left(\frac{1}{\ell} \ln t\right)^{t-1} < 29e \left(\frac{1}{3947} \ln 29\right)^{\frac{\sqrt{5}-1}{2}} < 1.$$

□

For the remaining case, we can use numerical analysis to show that $\binom{k}{\ell} \left(1 - \sqrt[\ell]{\frac{\ell}{k}}\right)^{k-\ell} < 1$ holds for every $\ell < 3947$ and $\frac{\sqrt{5}+1}{2}\ell < k < 29\ell$. The codes (in C++) for this checking are presented below.

```

1  #include<iostream>
2  #include<algorithm>
3  #include<cmath>
4
5  using namespace std;
6
7  int main()
8  {
9      for (int ell = 2; ell < 3948; ell++)
10     {
11         for (int k = (1.5 * ell); k < ell * 29 + 1; k++)
12         {
13             double result = 1;
14             for (int i = 0; i < k - ell; i++)
15             {
16                 // one term in the combinatorial number
17                 result *= 1.0 * (k - i) / (i + 1);
18                 // one term for (1 - sqrt[ell]{ell / k})^{k - ell}
19                 result *= 1 - pow(1.0 * ell / k , 1.0 / ell);
20             }
21             if (result >= 1)
22             {
23                 cout << "Error: " << ell << " " << k << " " << result << endl;
24                 return 0;
25             }
26         }
27         cout << ell << endl;
28     }

```

```

29   return 0;
30 }

```

B.3 Detailed Calculations of $\frac{\partial U(T)}{\partial p}$ in Claim 4.5

$$\begin{aligned}
\frac{\partial U(T)}{\partial p} &= -k \left(\sum_{t=0}^T (T-t) \binom{k}{t} p^t (1-p)^{k-t} \right) \\
&\quad - k \left(\sum_{t=1}^T (T-t) k \binom{k-1}{t-1} p^t (1-p)^{k-t} - \sum_{t=0}^T (T-t) k \binom{k-1}{t} p^{1+t} (1-p)^{k-t-1} \right) \\
&\quad \text{(differentiate and rewrite the combinatorial numbers)} \\
&= -k \left(\sum_{t=0}^T (T-t) \binom{k}{t} p^t (1-p)^{k-t} \right) \\
&\quad - k \left(\sum_{t=0}^{T-1} (T+1-t) k \binom{k-1}{t} p^{t+1} (1-p)^{k-t-1} - \sum_{t=0}^T (T-t) k \binom{k-1}{t} p^{1+t} (1-p)^{k-t-1} \right) \\
&\quad \text{(rewrite the second summation from 0 to } T-1 \text{)} \\
&= -k \left(\sum_{t=0}^{T-1} (T-t) \binom{k}{t} p^t (1-p)^{k-t} - k \sum_{t=0}^{T-1} \binom{k-1}{t} p^{1+t} (1-p)^{k-t-1} \right) \\
&\quad \text{(combine the 2nd and 3rd summations; noting that the last terms in the 1st and 3rd summations are 0)} \\
&= -k \left(\sum_{t=0}^{T-1} (T-t) \binom{k}{t} p^t (1-p)^{k-t} - \sum_{t=0}^{T-1} (k-t) \binom{k}{t} p^{1+t} (1-p)^{k-t-1} \right) \\
&= -k \left(\sum_{t=0}^{T-1} \binom{k}{t} p^t (1-p)^{k-t-1} [(T-t)(1-p) - (k-t)p] \right).
\end{aligned}$$

B.4 The Remaining Part for the Proof of Claim 4.8

The formal proof for

$$\phi(k, p) = k^2 p^2 (1-p)^{k-2} (1 + (k-2)p) > 1$$

goes into two parts. For the first part, we have used a computer program to check that $\phi(k, p) > 1$ holds for all $k = 3, 4, \dots, 1000$ and p given by (13). The codes are attached to the end of this proof. Now, it remains to consider the second part with $k > 1000$.

To characterize the approximation $p \approx \frac{\sqrt{2}}{k}$, we compute

$$\frac{p}{\frac{\sqrt{2}}{k}} = \frac{1}{\frac{\sqrt{2}}{k} + \sqrt{1 - \frac{1}{k}}} \in \left(\frac{1}{\frac{\sqrt{2}}{k} + 1}, \frac{1}{\sqrt{1 - \frac{1}{k}}} \right) \subseteq \left(\frac{1}{\frac{\sqrt{2}}{1000} + 1}, \frac{1}{\sqrt{1 - \frac{1}{1000}}} \right) \subseteq (0.998, 1.001).$$

Therefore,

$$0.998 \cdot \frac{\sqrt{2}}{k} \leq p \leq 1.001 \cdot \frac{\sqrt{2}}{k}.$$

Substituting this into ϕ , we have

$$\phi(k, p) = k^2 p^2 (1-p)^{k-2} (1 + (k-2)p)$$

$$\begin{aligned}
&\geq k^2 \left(0.998 \cdot \frac{\sqrt{2}}{k}\right)^2 \left(1 - 1.001 \cdot \frac{\sqrt{2}}{k}\right)^{k-2} \left(1 + (k-2) \cdot 0.998 \cdot \frac{\sqrt{2}}{k}\right) \\
&> \left(0.998 \cdot \sqrt{2}\right)^2 \left(1 - 1.001 \cdot \frac{\sqrt{2}}{k}\right)^k \left(1 + 0.998 \cdot \sqrt{2} - 2 \cdot 0.998 \cdot \frac{\sqrt{2}}{k}\right) \\
&\hspace{15em} \text{(since } 1 - 1.001 \cdot \frac{\sqrt{2}}{k} < 1) \\
&> \left(0.998 \cdot \sqrt{2}\right)^2 \left(1 - 1.001 \cdot \frac{\sqrt{2}}{1000}\right)^{1000} \left(1 + 0.998 \cdot \sqrt{2} - 2 \cdot 0.998 \cdot \frac{\sqrt{2}}{1000}\right) \\
&\hspace{10em} \text{(since } k > 1000 \text{ and the function } f(k) = (1 - 1/k)^k \text{ is increasing in } k) \\
&> 1.16, \hspace{15em} \text{(computed by a calculator)}
\end{aligned}$$

which concludes the claim.

The codes (in C++) below check that $\phi(k, p) > 1$ holds for all $k = 3, 4, \dots, 1000$ and p given by (13).

```

1  #include<iostream>
2  #include<algorithm>
3  #include<cmath>
4
5  using namespace std;
6
7  int main()
8  {
9      for (int k = 3; k <= 1000; k++)
10     {
11         double p = 1.0 / (1 + pow(k * (k - 1) / 2, 0.5));
12         double E = k * k * p * p * pow(1 - p, k - 2) * (1 + (k - 2) * p);
13         cout << E << endl;
14         if (E <= 1)
15         {
16             cout << "Error: " << k << " " << E << endl;
17             return 0;
18         }
19     }
20     return 0;
21 }
```