# Ensemble Analysis of Ethnic and Partisan Fairness in
Cal 3 vs. California

Indiana Huey

I. Introduction

The Cal 3 Initiative moved to split California into three separate states: colloquially NorCal, Cal, and SoCal. According to the initiative's primary advocate, venture capitalist Tim Draper, Cal 3 would "empower regional communities to make better, fairer and more sensible decisions for their citizens," helping to solve "California's most pressing issues, including the state's failing school systems … highest-in-the-nation taxes, deteriorating infrastructure and strained government" [1]. The initiative was due to be on the ballot in 2018. However, the proposition was removed following the California Supreme Court case *Planning and Conservation League v. Padilla* in which it was ruled as lacking the authority to revise the state constitution [2]. Nevertheless, Cal 3 garnered a non-trivial amount of support with 17 percent of likely voters supporting the idea [1].

California Three States Initiative (Proposed)

Source: Ballotpedia
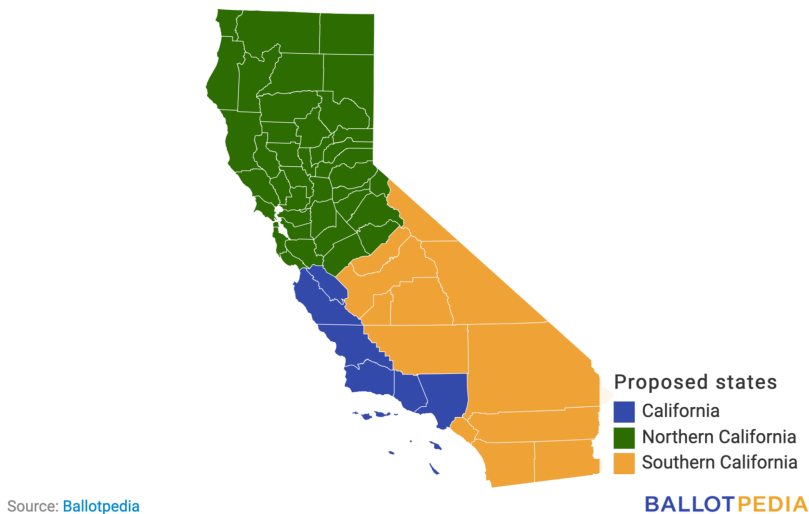
This paper examines the effects of Cal 3 on the political redistricting process—i.e. the drawing of congressional districts, which determines the amount of political representation different groups possess in the U.S. House of Representatives. Specifically, this paper presents ensemble analyses of Cal 3—i.e. aggregated ensembles of NorCal, Cal, and SoCal—in comparison to the same of California, assessing the differences in ethnic and partisan fairness. The analyses indicate that Cal 3 would result in minor decreases to the percentage of districts that are majority-Hispanic or -Latino, to the Republican seat share, and to the efficiency of Republican votes, reducing ethnic and partisan fairness.

II. Method

As ethnic data was recorded on the level of census tracts while partisan data was recorded on the level of precincts, two separate sets of ensembles were generated in order to examine both ethnic and partisan effects. In all cases, each ensemble contained 50,000 congressional redistricting plans, generated via the Recombination algorithm by DeFord et al. Ensembles of Cal 3 were generated by aggregating separately generated ensembles of NorCal, Cal, and SoCal. For instance, recording the number of majority-Hispanic or- Latino districts in a plan of Cal 3 involved summing the number of such districts across NorCal, Cal, and SoCal. The number of districts for each of NorCal, Cal, and SoCal was determined by multiplying the number of total districts in California by the proportion California's population contained within each of NorCal, Cal, and SoCal. With California containing 52 congressional districts, NorCal, Cal, and SoCal were allocated 18, 16, and 18 districts, respectively.

A. Ethnic Data

Using census reports from 2020 and shape files from 2022 courtesy of the U.S. Census Bureau, ensembles containing ethnic data—specifically, the total number of counted people and number of counted people who identify as Hispanic or Latino in each region—were generated on the regional granularity of census tracts. In order to find confidence in the convergence of the ensembles of California, three ensembles were generated from three random initial states. Showing the distributions of the number of cut edges across districts and where the dashed lines are the mean values, figure 1 demonstrates that the ensembles were in fact initialized from different states. Showing the same, figures 2 and 3 suggest that the ensembles converged or were close to convergence after 50,000 steps as there is significant overlap between the distributions and there is little change between 25,000 steps and 50,000 steps.
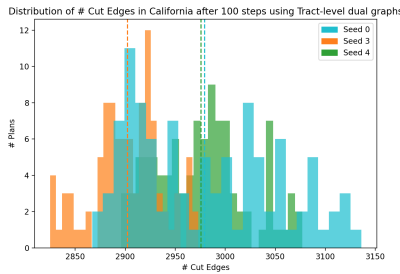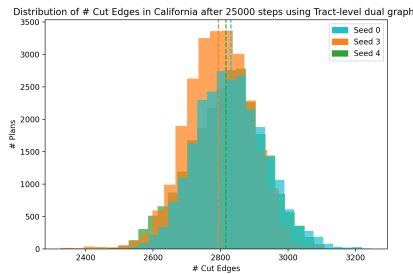

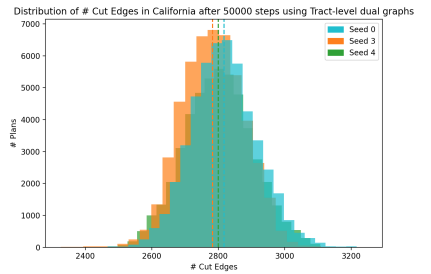Fig. 1.　　　　　　　　　　　Fig. 2.　　　　　　　　　　　Fig. 3.

In order to find confidence in the convergence of ensembles of Cal 3, the same was done using two ensembles. Fewer ensembles for evidence of convergence were generated for Cal 3 following the notion that ensembles of Cal 3 are likely to converge after roughly the same number of steps required for ensembles of California to converge as Cal 3 is comprised of subparts of California and especially as each step in an ensemble of Cal 3 involves three iterations of Recombination as opposed to one as ensembles of Cal 3 are aggregates of the

separate ensembles of NorCal, Cal, and SoCal. Figures 5 and 6 suggest that the ensembles converged or were close to convergence after 50,000 steps following the same reasoning as for the ensembles of California.
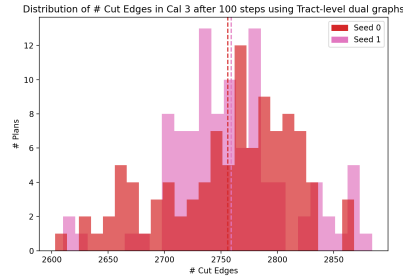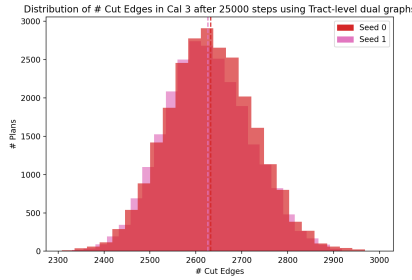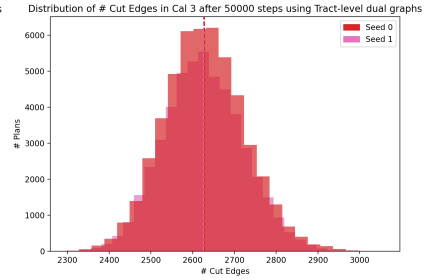

Fig. 4.

Fig. 5.

Fig. 6.

B. Partisan Data

Using voting data and shape files from 2016 courtesy of the Los Angeles Times Data Viz team, ensembles containing partisan data—specifically, the number of votes casted for each candidate in the 2016 presidential election in each region—were generated on the regional granulatory of precincts. In order to find confidence in the convergence of the ensembles of California, ensembles were generated by the same process as before. However, in this case only two ensembles were generated instead of three for computational conservation. Figures 8 and 9 suggest convergence following the same reasoning as before.
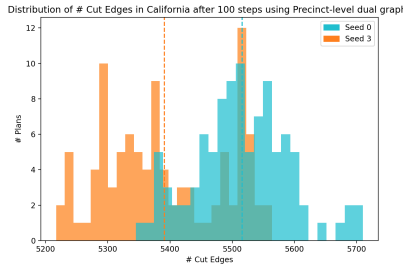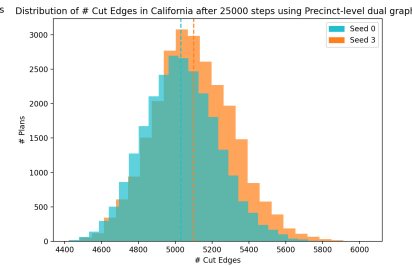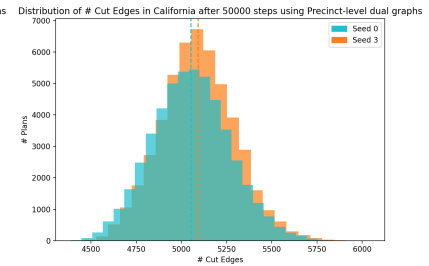

Fig. 7.

Fig. 8.

Fig. 9.

Following the same notion as before that convergence of California ensembles suggests convergence of Cal 3 ensembles after the same number of steps, only one ensemble of Cal 3 was generated for computational conservation and its convergence was assumed.

C. Limitations of Partisan Data

Voting data was only available on the precinct level, while population data was only available on the census tract level. With no key between the datasets by which to join them—except for into the county level, which is too large for redistricting in California and especially Cal 3—voting data was only accessible void of population data, specifically population count per region. Consequently, the ensembles containing partisan data base population count on vote count, such that the regional population counts rely on the likely false assumption that voter turnout is even across regions. As a result, the regional population counts

used during generation of these ensembles likely systemically underrepresent some regions more than others, possibly skewing the distribution of plans from which the ensembles draw.

On a positive note, the allocation of the number of districts for NorCal, Cal, and SoCal was computed to be 18, 15, and 18 using vote count as a proxy for population count. The similarity of this allocation to the allocation of 18, 16, and 18 which used more reliable census data indicates that voter turnout is roughly even across NorCal, Cal, and SoCal. However, this does not necessarily guarantee the same across precincts, which is still important for the representativity of the ensembles.

Lastly, in addition to being six years ago, the 2016 presidential race could be considered to have been particularly abnormal such that the numbers of votes casted for Trump and Clinton might not be representative of the numbers of Republican and Democrat voters in general.
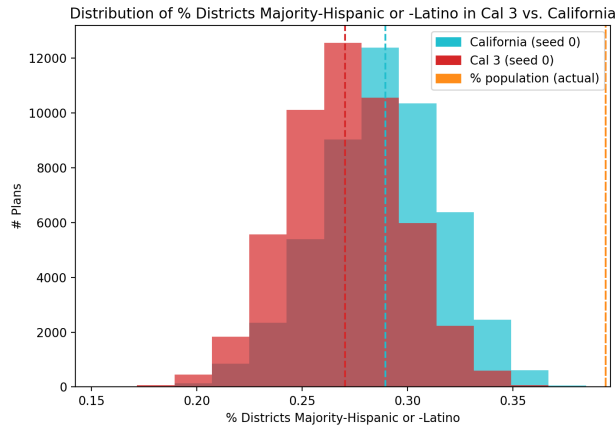
III. Results

Distribution of % Districts Majority-Hispanic or -Latino in Cal 3 vs. California

Fig. 10

Distribution of Republican Seat Share in Cal 3 vs. California

Fig. 11
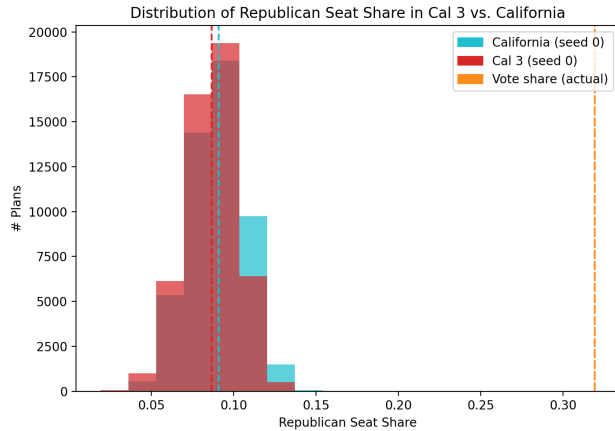
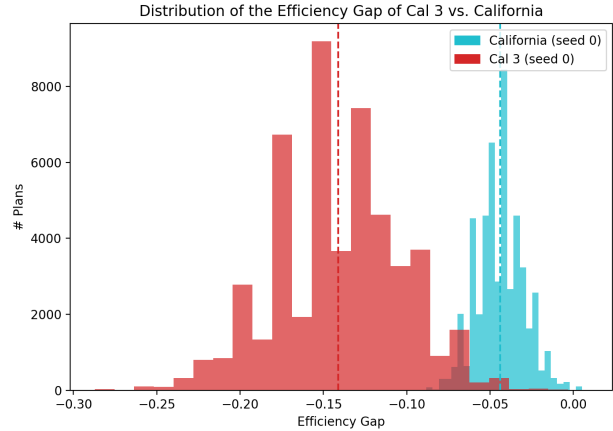Distribution of the Efficiency Gap of Cal 3 vs. California

Fig. 12

IV. Discussion & Conclusion

The results suggest that Cal 3 would slightly reduce ethnic and partisan fairness. Figure 10 indicates that the percentage of districts that are majority-Hispanic or -Latino in Cal 3 would likely be lower than that in California, reducing Hispanic and Latino proportionality as the

percentage of the counted population that identifies as Hispanic or Latino is greater than the percentage of districts that are majority-Hispanic or Latino in the ensemble distributions of both Cal 3 and California. Figure 11 indicates the same for the percentage of districts that are majority-Republican—i.e. Republican seat share—with the distribution of Cal 3 being further from proportionality than that of California. Figure 12 possibly explains this decrease in Republican seat share, showing that the plans in Cal 3 have lower efficiency gaps than most plans in California and indicating that Republican votes would be used less efficiently in Cal 3 than in California.

By each metric, the difference between Cal 3 and California is minor. For instance, the difference between the mean values of the distributions of the percentage of districts that are majority-Hispanic or Latino of Cal 3 and California is no more than 5 percent, and the same for Republican seat share is even less. Nevertheless, these ensemble analyses suggest that Cal 3 would slightly reduce both ethnic and partisan fairness, decreasing proportionality for the Hispanic and Latino population and for Republican voters and decreasing the already low efficiency of Republican votes.

Resources

1. Please access the implementation here:
   drive.google.com/drive/folders/1C1S3Jw2S9EnPniPwVvL4anmhjDi-IaYR?usp=sharing
2. Ethnic and population data from the U.S. Census Bureau is found here: www.census.gov
3. Partisan data from the Los Angeles Times Data Viz team is found here:
   github.com/datadesk/california-2016-election-precinct-maps

References

[1] A. Brinklow, "Everything you need to know about Three Californias," *Curbed San Francisco*, 2018,
    sf.curbed.com/2018/6/14/17464134/three-californias-tim-draper-ballot-iniative.
[2] "California Proposition 9, Three States Initiative (2018)," *Ballotpedia*,
    ballotpedia.org/California_Proposition_9,_Three_States_Initiative_(2018).