POWER IN VOTING

JAMES DELLA-GIUSTINA, CRAIG RILEY, MITCHELL MACADAMS, GAVIN ANDRE

1. Introduction

In this paper we aim to quantify the 'power' that a state has for the United States presidential election through electoral college and for the primary process for the Democratic party. We first examine established methods like the Banzhaf and Shapely-Shubick indices for the electoral college and then try to formulate our own power index for the Democratic primary process.

2. Power Index

Here we present an axiomatic approach to defining a power index. We denote the set of players N, the set of winning coalitions $\mathcal{W}$, and a generalized power index as k.

Theorem 1. *A power index k should satisfy:*

1. *Be a function which takes any player i to a non-negative real number: $k(i) \geq 0$ for $\forall i \in N$.*
2. *Satisfy $k(i) = 0$ iff player i is a dummy player.*
3. *k should be determined by the set of winning coalitions $\mathcal{W}$:*
   - *If i and j have symmetric positions in $\mathcal{W}$, then $k(i) = k(j)$.*
4. *$k(i) \geq k(j)$ iff i has a greater number of votes than j.*
5. *(Optional) $\sum_{i=1}^{n} k(i) = 1$.*

2.1. **Shapely-Shubick Power Index.** The *Shapely-Shubick index* hinges on the number of times a player is 'pivotal' in a yes-no voting system. We need to take the total number of players n and find all $n!$ permutations of the orderings of players. Note that the computations can be greatly reduced if we encounter symmetry among any of the players. We then systematically (starting at the left) suppose that every player is voting yes and find which player tips the scale such that their vote forms a winning coalition $S \in \mathcal{W}$. Now that the concept of a player being pivotal has been established, we can define the Shapely-Shubick power index:

Definition 1 ([2])**.** Suppose p is a voter in a yes-no voting system and let N be the set of all voters. Then the *Shapely-Shubick index* of p, denoted SSI(p) is the number given by:

$$\text{SSI}(p) = \frac{\text{number of orderings of } N \text{ for which } p \text{ is pivotal}}{|N|!}$$

2.2. **Banzhaf Power Index.** Another power index we examine is the Banzhaf system, in which we need to also understand the concept of a player being 'critical'. A player p is critical to a winning coalition $S \in \mathcal{W}$ iff $S/\{p\} \notin \mathcal{W}$. We denote the total number of times that a player is critical as η_p. Now that we grasp this idea of being critical, we can therefore define the Banzhaf power index.

Definition 2 ([2])**.** Suppose that p_1 is a player in a yes-no voting system and that the other players are denoted by $p_2, p_3, \ldots, p_n$. Then the *Banzhaf index* of p_1, denoted as $\mathrm{Bl}(p_1)$ is the number given by:

$$\mathrm{Bl}(p_1) = \frac{\eta_{p_1}}{\sum_{i=1}^{n} \eta_{p_i}}$$

3. Electoral College

The electoral college is a voting system used by the U.S. government in presidential elections. Unlike a popular vote, a state is assigned two votes for each of its two senators, with the additional votes coming from the number of congressional seats the state possesses [4]. Congressional seats are distributed to each of the 50 states from decennial population counts [3] with each state receiving at least one seat, and the congressional districts within each state are supposed to have approximately equal population sizes.

As can be seen in Figures 1 and 2, the minimum number of electoral college votes a state has is 3 (Montana, Vermont, DC, etc…), while the largest amount of votes is 55 for California. Altogether we have 538 total electoral college votes, with a minimum quota of 270 needed to win the election for the respective presidential candidates. 48 states (and DC) determine all electoral votes (winner-take-all) based solely off of the number of popular votes for each candidate within the state. Nebraska and Maine deviate from this method [14], and instead appropriate two votes for the winner of the popular vote for the entire state, and then designate 1 vote for the popular vote of each congressional district to the preferred candidate (Maine with 2 districts and Nebraska with 3 districts respectively). Due to this different system, we therefore carefully considered all outcomes possible. For Maine, if we were to consider each congressional district having equal population sizes then you would need to win the popular vote of both districts to win all 4 votes. But this equal population distribution is so extremely unlikely that we instead argue that whomever wins the larger district wins the overall state, and so we split Maine into two parts: one with 3 votes and one with just 1 vote. Similarly for Nebraska, we make a similar simplification and split the state into two parts: one with 4 votes and the other with 1 vote. While some nuance is undoubtedly lost, this was the easiest simplification without having to dive into conditional probabilities and Bayesian statistics. This gave us a total of 53 distinct players ($|N| = 53$) to compute our power indices for.

So how can we determine which state holds the most political power in the electoral college system? Let's apply the Banzhaf and Shapely-Shubick power indices to gauge the political power of each state and then interpret the results.

Because:

$$N = \{\text{Alabama, Alaska,} \ldots \text{Maine-Overall, Maine-1} \ldots \text{Nebraska-Overall, Nebraska-1} \\ \text{Washington D.C.,} \ldots \text{Wisconsin, Wyoming}\} \\ \therefore |N| = 53$$

Then the Banzhaf power index may be possible to compute by hand, but the different orderings required by the Shapely-Shubick power index is 53! and therefore infeasible to do without the help of a computer.

3.1. Methods to Compute Banzhaf & Shapely-Shubick Power Indices.

In order to find the Shapely-Shubick index described in Section 2.1, we need to find all 53! different possible orderings, which is obviously not feasible by hand. To help us compute this index we can therefore use a power index library implemented in Python [7]. Using electoral college distribution data [4], our program is:

```
import powerindices
# For electoral college:
   quota = 270
# Alphabetical order of states and their number of electoral college votes
   weights = [9,3,11,6,55,9,7,3,3,29,16,4,4,20,
      11,6,6,8,8,2,1,1,10,11,16,10,6,10,3,2,1,1,1,
      6,4,14,5,29,15,3,18,7,7,20,4,9,3,11,38,6,3,13,
   12,5,10,3]
# Calculate Power Indices of Shapely Shubick
SSIs = powerindices.compute_ssi(quota,weights)
```

The third and fourth columns of Figures 1 and 2 show the power indices for each state. Just as expected, symmetric players (for example Delaware and Washington D.C. with 4 votes each) have the same power in each respective index.

For the Banzhaf index described in Section 2.2, we needed to find again all possible $S \in \mathcal{W}$ and determine when a player is critical. While this have been less computationally intensive then finding all 53! permutations for the Shapely-Shubick index, it was still infeasible to perform by hand. We therefore employed an online calculator [11] that asked for the quota and weights that we used in our previous code snippet.

We notice that states with a large number of electoral votes like California, New York, Texas, etc.. have larger power indices. California easily holds the most power in the sense that they are 'critical' by the Banzhaf index $\sim$ 11% of the time and similarly they are 'pivotal' $\sim$ 11% of the time. On the other end of the spectrum, we have the small states such as Delaware and D.C. that have slightly more than $\sim$ .005% power in both indices.

4. MRDA Power Index in Primary Voting

To establish our own power index (dubbed MRDA) in the case of primary voting, we first need to explore this multi-faceted and complex process. We chose to focus on the Democratic party and the presidential election of 2020.

In total there are 3,979 delegates distributed across all U.S. states and territories that meet at the Democratic National Convention to determine the party's nominee [15]. Delegates are allocated to states based upon the number of electoral college votes the state possesses and its history of voting for a Democratic nominee. This leads to strongly blue states like California and New York to have a far larger number of delegates than traditionally Republican states like Texas.

Next we need to establish the difference between a state primary and a state caucus, of which both constitute the overall primary [10]. A state primary is similar to a presidential election, registered voters cast a private vote for their candidate of choice for the party. Primaries vary from each state, some allow all registered voters to participate while others require only those registered as Democrats. It should be noted though that a voter is restricted to participate in only one parties' primary, though this may be difficult to enforce [5].

Established Indices for the Electoral College

State	# of Votes	Banzhaf Index	Shapely Shubick Index
Alabama	9	0.01640	0.01638
Alaska	3	0.00545	0.00540
Arizona	11	0.02006	0.02009
Arkansas	6	0.01092	0.01086
California	55	0.11363	0.11030
Colorado	9	0.01640	0.01638
Connecticut	7	0.01274	0.01269
District of Columbia	3	0.00545	0.00540
Delaware	3	0.00545	0.00540
Florida	29	0.05401	0.05484
Georgia	16	0.02929	0.02950
Hawaii	4	0.00727	0.00721
Idaho	4	0.00727	0.00721
Illinois	20	0.03676	0.03715
Indiana	11	0.02006	0.02009
Iowa	6	0.01092	0.01086
Kansas	6	0.01092	0.01086
Kentucky	8	0.01457	0.01453
Louisiana	8	0.01457	0.01453
Maine-Overall	3	0.00545	0.00540
Maine-1 District	1	0.00182	0.00179
Maryland	10	0.01823	0.01823
Massachusetts	11	0.02006	0.02009
Michigan	16	0.02929	0.02950
Minnesota	10	0.01823	0.01823

FIGURE 1. Weights, Banzhaf Power Index, and Shapely-Shubick Index for Alabama through Minnesota.

A state caucus on the other hand is more of an open forum, allowing any registered voter to congregate in some location. Participants are then asked to form groups based on their preferred candidates and a tally is taken. If the number of people for a specific candidate does not meet at least 15%, then their candidate of choice is deemed 'nonviable' and eliminated from the process. At this stage, these individuals can either choose a new candidate to align themselves with or leave the caucus. However, individuals whose candidate is still eligible are essentially 'locked' into their votes and may not change groups. Once a final count has been taken in which every candidate receives above the 15% threshold, the number of votes and percentages are reported to the county convention, which are in turn reported to the 'district' convention and finally to the state's Democratic party.

In either a primary or a caucus, the number of delegates awarded per candidate for each state is proportional to the number of voters who backed said candidate [9].

Established Indices for the Electoral College

State	# of Votes	Banzhaf Index	Shapely Shubick Index
Mississippi	6	0.01092	0.01086
Missouri	10	0.01823	0.01823
Montana	3	0.00545	0.00540
Nebraska-Overall	4	0.00727	0.00721
Nebraska-1 District	1	0.00182	0.00179
Nevada	6	0.01092	0.01086
New Hampshire	4	0.00727	0.00721
New Jersey	14	0.02559	0.02571
New Mexico	5	0.00910	0.00903
New York	29	0.05401	0.05484
North Carolina	15	0.02744	0.02760
North Dakota	3	0.00545	0.00540
Ohio	18	0.03302	0.03331
Oklahoma	7	0.01274	0.01269
Oregon	7	0.01274	0.01269
Pennsylvania	20	0.03676	0.03715
Rhode Island	4	0.00727	0.00721
South Carolina	9	0.01640	0.01638
South Dakota	3	0.00545	0.00540
Tennessee	11	0.02006	0.02009
Texas	38	0.07206	0.07323
Utah	6	0.01092	0.01086
Vermont	3	0.00545	0.00540
Virginia	13	0.02374	0.02383
Washington	12	0.02190	0.02196
West Virginia	5	0.00910	0.00903
Wisconsin	10	0.01823	0.01823
Wyoming	3	0.00545	0.00540

FIGURE 2. Weights, Banzhaf Power Index, and Shapely-Shubick Power Index for Mississippi to Wyoming.

Traditionally, the primary processes start with the Iowa caucus which fell on February 3^{rd} for the 2020 election. Many speculate and say that whomever wins the Iowa caucus will win the party's nominations, this occurring a total of 7 out of 10 times since the 1970's[1]. Next comes New Hampshire on February 11^{th} and Nevada on February 22^{nd}. The next primaries/caucuses occur on March 3^{rd} or 'Super Tuesday' for 14 states and American Samoa. The remainder of the state primaries and caucuses occur anytime between super Tuesday and concluding on August 11^{th} with Connecticut.

Much importance is placed on the results of both the Iowa caucus and super Tuesday, indicating that they will dictate the nominee. Though this has been refuted [13] with clear and empirical evidence, we wanted to include parameters for this 'effect' of both scenarios in our model.

4.1. **Our Model.** In order to create a reasonable measure of political power for the Democratic primaries, a Monte Carlo simulation was chosen. Monte Carlo simulations are more precisely called stochastic computer simulations, which introduces a measure randomness and iterative calculations rather then some deterministic calculation [12]. This simulation method allows us to analyze the system without explicitly constructing equations for every single component, which may be impossible. A random number generator was used to assign a (random) percentage for the main candidate to each state and thereby automatically defining the remaining portion to the other candidate(s).

4.1.1. *Base Case.* Our base case sums the votes each candidate earns in each state based on the random number initially assigned to the main candidate until 1,991 delegates are reached to either the 'main' or 'other' candidates. Once that threshold has been reached and the point at which it has been reach has been determined, that is how the power indices are determined. For example, if that threshold is reached for the main candidate winning the necessary amount of votes, then each states' votes that went to the main candidate are divided by the number of total votes at the point in which the majority threshold (1,991 delegates) are reached. This same process is repeated a certain number of times, in our scenario 100 iterations were performed.

4.1.2. *Positive Iowa Effect.* Another scenario considered was the 'positive Iowa effect'. When it comes to the general election, it is important to note that not all "swing voters" could potentially change their vote to support the other party's candidate with about half of swing voters (16% of all voters) being truly persuadable. In the general election there are considered to be about 11% of voters that identify as swing voters [8]. Because of this, our model assigned a 5% 'bump' to the initial randomized percentage that each candidate would receive from any given state, solely based off the winner of the Iowa caucus. For example if our 'main' candidate won Iowa then we would boost each states vote percentage for the 'main' candidate by 5%. It should be noted that this 'bump' was assigned only to states and territories that voting from Super Tuesday and on. This follows the 2020 Democratic Primaries because the Iowa results underwent two recounts and the results were not released until February 29^{th} [1]. For example suppose our 'main' candidate wins Iowa and California's assigned randomized percentage for the 'main' candidate is 23%. Under this 'positive Iowa effect' California would therefore have 28% of their total delegates assigned to the 'main' candidate.

4.1.3. *Negative Iowa Effect.* Our model also considered what we called the "negative Iowa effect." This is the opposite of what we did above, we assign a -5% "bump" to each states percentage assigned to the candidate who won Iowa once Super Tuesday voting begins.

4.1.4. *Positive Super Tuesday Effect.* The Positive Super Effect was calculated similarly to the Iowa Effect in the sense that a percentage (5%) was assigned to the 'main' candidates to states that voted March 10^{th} and later.

4.1.5. *Negative Super Tuesday Effect.* In contrast to the Positive Super Tuesday Effect, the Negative Super Tuesday Effect assigned a -5% 'bump' to the states that voted March 10^{th} and later.

4.1.6. *Overall: Final Results.* At the end of all subsequent iterations, an average of each state's power index is calculate giving rise to the final power indices using our model.

Tables 3 and 4 show the results of 50 and 100 Monte Carlo simulations for each state based on the order of the primaries/caucus. The composition of different 'cases' for 50 simulations, was 10 from each of the categories positive/negative Iowa effect, positive/negative super Tuesday effect, and 10 cases with no effect at all. Similarly the composition of 100 Monte Carlo simulations was 20 simulations from each of the categories.

It is important to note that our model satisfies all five power index axioms except symmetry. Symmetry is not fundamentally built into our model (although in practice, with enough iterations, it could happen). We note that this is not an issue because of the nature of the primaries in general.

5. Conclusion

Our model predicts that the top five states with the most power are: California, New York, Texas, Florida, and Pennsylvania. This coincides with what was predicted for the general election using the Banzhaf and Shapely-Shubick models. We also note that our model follows each state's power indices within (some small percentage) to what was calculated using the Banzhaf and Shapely-Shubick models. It should be noted that while the Banzhaf and Shapely-Shubick considered the number of electoral college votes, our MRDA index considered the number of primary delegates. Also observe that the number of 'players' in the electoral college is lower then that of the primary process (53 v.s. 57). Because of the unique players only found in one of the games have the lowest power we do not do a direct comparison.

MRDA Power Index for Democratic Primaries

State	Voting Date	# of Delegates	MRDA(50)	MRDA(100)
Iowa	3-Feb	41	0.008240266	0.00860755
New Hampshire	11-Feb	24	0.004864806	0.005009968
Nevada	22-Feb	36	0.007042196	0.007982642
South Carolina	Feb-29	54	0.012131849	0.01213222
Alabama	3-March	52	0.01237927	0.01352207
American Samoa		6	0.001308203	0.001419574
Arkansas		31	0.007912819	0.007916712
California		415	0.110242058	0.116200063
Colorado		67	0.019323116	0.018175265
Maine		24	0.006268083	0.005880509
Massachusetts		91	0.022138376	0.021486596
Minnesota		75	0.01646626	0.017837121
North Carolina		110	0.029748191	0.027152427
Oklahoma		37	0.008248612	0.008919999
Tennessee		64	0.015522213	0.013843222
Texas		228	0.054772508	0.054325303
Utah		29	0.005272269	0.005917347
Vermont		16	0.003232608	0.00368028
Virginia		99	0.024859673	0.024713341
Democrats Abroad	10-March	13	0.003589566	0.003249312
Idaho		20	0.004760408	0.004813418
Michigan		125	0.032979879	0.032358843
Mississippi		36	0.008609933	0.008234897
Missouri		68	0.018818447	0.018790104
North Dakota		14	0.003524872	0.003465818
Washington		89	0.020084517	0.020188066
Northern Marianas	14-March	6	0.001651386	0.001535472
Arizona	17-March	67	0.017561784	0.017588093
Florida		219	0.054093597	0.054450956
Illinois		155	0.034394555	0.036113588
Wisconsin	7-April	84	0.023562098	0.022905364
Alaska	10-April	15	0.003580752	0.003786075

Figure 3. MRDA Power Indices for 50 & 100 Monte Carlo Simulations for February 3^{rd} to April 10^{th}.

MRDA Power Index for Democratic Primaries

State	Voting Date	# of Delegates	MRDA (50)	MRDA (100)
Wisconsin	7-April	84	0.023562098	0.022905364
Alaska	10-April	15	0.003580752	0.003786075
Wyoming	17-April	14	0.003145138	0.003355501
Ohio	28-April	136	0.03465834	0.033469394
Kansas	2-May	39	0.009602212	0.010133968
Nebraska	12-May	29	0.006940825	0.007173716
Oregon	19-May	61	0.016696439	0.0161799
Hawaii	22-May	24	0.006416565	0.005902045
D.C.	2-Jun	20	0.005361166	0.005353619
Indiana		82	0.021603421	0.021539961
Maryland		96	0.026081764	0.025325521
Montana		19	0.005058098	0.005011869
New Mexico		34	0.008418684	0.008158103
Pennsylvania		186	0.046947568	0.045890321
Rhode Island		26	0.006518123	0.006902317
South Dakota		16	0.004560629	0.00426738
Guam	6-Jun	7	0.001822997	0.001815856
US Virgin Islands		7	0.002038386	0.001891112
Georgia	9-Jun	105	0.028274491	0.028233691
West Virginia		28	0.007217587	0.006838353
Kentucky	23-Jun	54	0.012860241	0.012129028
New York		274	0.066081353	0.069269263
Delaware	7-Jul	21	0.005766889	0.005577718
New Jersey		126	0.034807663	0.032129615
Louisiana	11-Jul	54	0.015028407	0.014557296
Puerto Rico	12-Jul	51	0.013003677	0.012515971
Connecticut	11-Aug	60	0.013904168	0.014176268

FIGURE 4. MRDA Power Indices for 50 & 100 Monte Carlo Simulations for April 7^{th} to August 11^{th}.

5 Most 'Powerful' States

State	Banzhaf Index	Shapely-Sh. Index	MRDA (50)	MRDA (100)
California	0.11363	0.11030	0.110242058	0.116200063
New York	0.05401	0.05484	0.066081353	0.069269263
Texas	0.07206	0.07323	0.054772508	0.054325303
Florida.	0.05401	0.05484	0.054093597	0.054450956
Pennsylvania	0.03676	0.03715	0.046947568	0.045890321

FIGURE 5. Five most 'powerful' states and their Shapely-Shubick, Banzhaf, MRDA(50), and MRDA(100) power indices.

5 Most 'Powerful' States

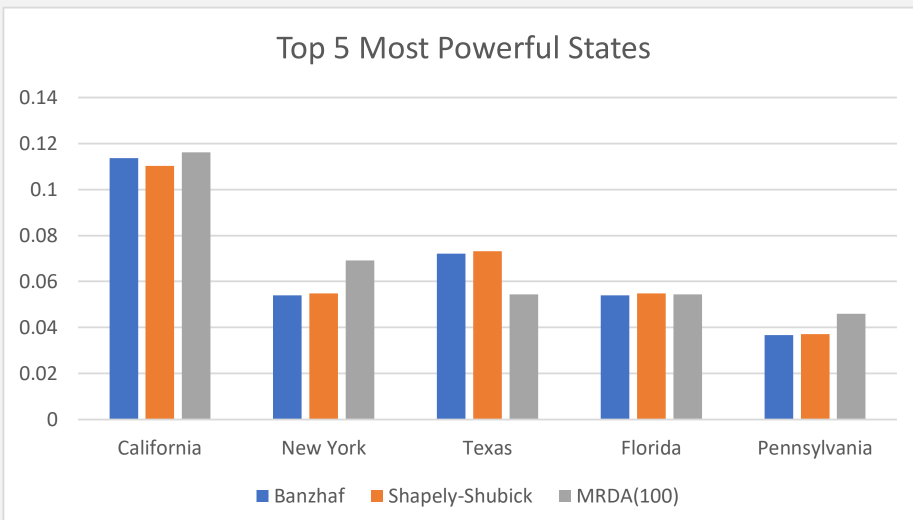

FIGURE 6. Top 5 most powerful states & their power indices according to the three models.

References

[1] *2020 Iowa Caucus Election Results*. URL: https://abcnews.go.com/Politics/2020-iowa-caucus-election-results/story?id=68333529.

[2] Allison M. Pacelli (auth.) Alan D. Taylor. *Mathematics and Politics: Strategy, Voting, Power and Proof*. 2nd ed. Springer-Verlag New York, 2008. ISBN: 9780387776453; 9780387776439; 0387776435; 0387776451.

[3] US Census Bureau. *About Congressional districts*. 2021. URL: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/congressional-dist.html.

[4] *Distribution of electoral votes*. URL: https://www.archives.gov/electoral-college/allocation.

[5] FairVote.org. *Primaries*. URL: https://www.fairvote.org/primaries#open_and_closed_primaries.

[6] Steven B. Horton Frank R. Giordano William P. Fox. *A First Course in Mathematical Modeling*. 5th ed. Cengage Learning, 2013. ISBN: 1-285-05090-8,978-1-285-05090-4.

[7] Frankhuettner. *Frankhuettner/Powerindices: Computation of banzhaf index, Shapley Shubik Index, Coleman Shapley Index*. URL: https://github.com/frankhuettner/powerindices.

[8] Ashley Kirzinger et al. *Data note: A look at swing voters leading up to the 2020 election*. 2019. URL: https://www.kff.org/other/issue-brief/data-note-swing-voters/.

[9] Robert Longley. *How Delegates to the National Presidential Conventions are Chosen*. 2020. URL: https://www.thoughtco.com/how-party-convention-delegates-are-chosen-3320136.

[10] Grace Panetta. *Here's How Democrats Will Elect their Presidential Nominee Over the Next Several Months*. 2020. URL: https://www.businessinsider.com/how-democratic-nomination-delegate-selection-process-works-2020-1.

[11] *Power Indices Calculated by Generating Functions*. URL: https://homepages.warwick.ac.uk/~ecaae/ipgenf.html.

[12] Reuven Y. Rubinstein. *Simulation and the Monte Carlo Method*. 2nd ed. Wiley series in probability and statistics. John Wiley Sons, 1991. ISBN: 9780470139011; 0470139013; 0470177942; 9780470177945.

[13] Craig Allen Smith. "The Iowa Caucuses and Super Tuesday Primaries Reconsidered: How Untenable Hypotheses Enhance the Campaign Melodrama". In: *Presidential Studies Quarterly* 22.3 (1992), pp. 519–529. ISSN: 03604918, 17415705. URL: http://www.jstor.org/stable/27550994.

[14] *Split electoral votes in Maine and Nebraska*. URL: https://www.270towin.com/content/split-electoral-votes-maine-and-nebraska/.

[15] Matt Stevens. *How to win the Democratic nomination, and why it could get complicated*. 2020. URL: https://www.nytimes.com/2020/02/22/us/politics/democratic-primary-dnc-superdelegates.html.