Journal of Data Science 0 (0), 1–18
July 2026

DOI: 10.6339/26-JDS1243
Computing in Data Science

The Marked Edge Walk: A Novel MCMC Algorithm for Sampling of Graph Partitions

Atticus McWhorter[1,*] and Daryl DeFord[2]

[1]*Department of Mathematics, Dartmouth College, USA*
[2]*Department of Mathematics and Statistics, Vassar College, USA*

Abstract

Novel Markov Chain Monte Carlo (MCMC) methods have enabled the generation of large ensembles of redistricting plans modeled as a graph partitioning problem. However, existing algorithms such as Reversible Recombination (RevReCom) and Metropolized Forest Recombination (MFR) have strong preferences for distributions related to the spanning tree measure. In this paper we introduce the Marked Edge Walk (MEW), a novel Markov chain proposal for sampling from the space of graph partitions. The walk operates on the space of spanning trees with marked edges, allowing for calculable transition probabilities for use in the Metropolis-Hastings algorithm. Empirical results on real-world dual graphs show convergence under a broad class of target distributions less constrained by spanning tree counts, including policy-based distributions, such as competitiveness on New Hampshire that are independent of spanning trees, and compactness and partisan symmetry distributions on New Hampshire and Texas that, while related to spanning trees, can now be properly targeted with a smaller degree of spanning tree bias, which represents an advancement in flexible ensemble generation.

Keywords *computational redistricting; graph partitioning; Markov Chain Monte Carlo; Metropolis Hastings; spanning trees*

1 Introduction

Recent advances in computational capabilities and data availability have greatly increased legislators' abilities to optimize political redistricting plans. In these algorithms, US states are partitioned into geographic units which are then encoded as planar graphs, transforming the problem to one about studying connected graph partitions (DeFord et al., 2021). Quantitative analysis of redistricting plans in this setting has often been carried out by generating large 'ensembles' of comparator plans using random walks to explore the space of balanced graph partitions (Clelland et al., 2020). In these efforts to develop a baseline of 'typical' plans, researchers have largely turned to Markov Chain Monte Carlo (MCMC) (DeFord and Duchin, 2019; Clelland et al., 2022; Caldera et al., 2020; DeFord et al., 2020; Chikina et al., 2017; Herschlag et al., 2020; Fifield et al., 2020a) or other Sequential Monte Carlo (SMC) methods (McCartan and Imai, 2023; O'Sullivan et al., 2026) to generate large collections of plans based on parameters drawn from legislative guidance or historical plans.

Before discussing specific algorithms, it is worth motivating why uniform sampling from the space of redistricting plans is not our goal. As discussed in Najt et al. (2021), an unconstrained

*Corresponding author. Email: atticus.w.mcwhorter.gr@dartmouth.edu.

© 2026 The Author(s). Published by the School of Statistics and the Center for Applied Statistics, Renmin University of China. Open access article under the CC BY license.
Received November 13, 2025; Accepted August 24, 2026

uniform distribution is not a useful baseline for policy analysis because most plans in the space are extremely non-compact, and the uniform distribution assigns equal weight to highly irregular plans. One approach to counteract this might be to restrict the state space with binary constraints on relevant criteria, for example by excluding all partitions that are significantly imbalanced or non-compact. This runs the risk of disconnecting the state space and also does not preferentially sample from plans that score better on relevant criteria, which is a desirable feature in applied work in this area. This type of approach has been used effectively when combined with some of the other methods described below (Herschlag et al., 2020).

Determining the distributions over graph partitions that can be sampled from efficiently and effectively is an active area of research. There are generic hardness results for uniform sampling of connected, balanced partitions on planar graphs (Najt et al., 2019), and a developing literature on better bounds for uniform sampling grid-like graphs (Frieze and Pegden, 2023), subexponential algorithms (Cai et al., 2024), and the corresponding enumeration problem (Donnay and Kahle, 2025). Beyond the uniform distribution, there has also been a significant amount of recent theoretical work on sampling from the spanning tree distribution which is described in more detail below (Cannon et al., 2022, 2024, 2025; Charikar et al., 2023). Given that the uniform distribution is often an undesirable target and that the governing legislation is not described in terms of the combinatorial spanning tree measurements that many of the current methods are most effective at sampling from, this motivates our interest in developing algorithms that can target a broader class of distributions.

A number of different algorithms are available to researchers who hope to generate an ensemble of redistricting plans, including several with public open-source implementations. As a developing field however, there is still room for improvement in methodology. A natural approach is to generalize proposals from statistical physics models, making small updates to the partition labels at each step (Chikina et al., 2017, 2020; Fifield et al., 2020a). The Recombination (ReCom) algorithm (DeFord et al., 2021), which iteratively merges pairs of adjacent districts and then bipartitions the merged pair by sampling a spanning tree and cutting an edge, has gained significant popularity both in academic work (Caldera et al., 2020; DeFord et al., 2020; DeFord and Duchin, 2019; Clelland et al., 2022; McWhorter and DeFord, 2026) and in litigation, due to its favorable mixing properties, and studies using ReCom have been used to argue the (un)constitutionality of districting plans in courts as high as the U.S. Supreme Court (Mathematicians, Law Professors, and Students, 2019). However, the invariant distribution of ReCom is unknown (Cannon et al., 2022).

The Reversible Recombination (RevReCom) algorithm (Cannon et al., 2022) tailors ReCom to satisfy the detailed balance condition and target the spanning tree distribution, a distribution which weights each graph partition according to the product of the number of spanning trees of each part. Great attention has been devoted to the spanning tree distribution (Cannon et al., 2022; Autry et al., 2023, 2021; Charikar et al., 2023; Clelland et al., 2021). Procaccia and Tucker-Foltz (2022) showed that the spanning tree distribution is inversely exponentially related to the total boundary lengths of the districts, and Cannon et al. (2024) developed an algorithm that can sample from the distribution in polynomial time.

Despite great attention to the spanning tree distribution, there is still interest in sampling from more general distributions, particularly those that incorporate policy considerations from a given state. Metropolized Forest Recombination (MFR) (Autry et al., 2023) is designed to do exactly that, utilizing the powerful tools of the Metropolis-Hastings algorithm to target any desired invariant distribution. To achieve this, the authors 'lift' to an expanded state space of spanning forests that allows for easier calculation of the forward and reverse transition proba-

bilities. However, it is still an open question what distributions can effectively be sampled using this method. As with ReCom, this approach has been successfully applied in both academic contexts (Zhao et al., 2022) as well as litigation, and it has also been extended to the multiscale setting (Autry et al., 2021; Chuang et al., 2025).

One advantage of the MFR approach presented in Autry et al. (2023) is that it makes explicit the contribution of the spanning tree density to the target distribution with a parameter that interpolates the relative impact of this term. We discuss this in more detail in Section 4 below, but a key motivation for our work is targeting distributions that are not dominated by this term. This motivation is common to other recently proposed sampling methods that take a similar approach to the one we develop here, in walking on the lifted space of trees in an attempt to reduce the reliance on the spanning tree distribution. In particular, the Cycle Walk (DeFord et al., 2025) and Balanced Up Down (BUD) Walk (Akitaya et al., 2026) fall into this category. Section 3.1 of Akitaya et al. (2026) describes the relationships and motivations between these proposals and the one we present here.

In this paper, we present the marked edge walk (MEW), an algorithm for sampling from the space of balanced graph partitions under a tuneable distribution. We demonstrate convergence on the uniform distribution on Cheshire County, competitiveness, compactness, and multivariate distributions on New Hampshire, and a multivariate distribution on Texas. We note that when targeting a Gaussian distribution on cut edges, MEW implicitly targets a lognormal distribution on spanning trees. While this remains dependent on spanning trees, it still represents an advancement over prior methods, which only partially account for the spanning tree degeneracy factor.

2 Procedure

The space of interest of MEW is the space of balanced graph partitions of a fixed graph $G = (V, E)$. In our application, we choose G to have a finite set of vertices V and undirected edges E. A graph partition $\xi : V \to \{1, 2, \ldots, d\}$ sorts each vertex into a part. However, instead of representing each state as a graph partition, as DeFord et al. (2021), we instead lift to the space of spanning trees with marked edges. More formally, the state space of MEW is

$$\mathcal{X} = \{(T, M) : T \text{ is a spanning tree of } G, M \subseteq T\}. \tag{1}$$

We will enforce that the forest $F = T \setminus M$ is balanced, where a forest F is balanced if each connected component D_i satisfies $|D_i| = n/d \pm \epsilon$ (node balance) or $\sum_{v \in D_i} w(v) = W/d \pm \epsilon$ (population balance) for vertex weights $w : V \to \mathbb{R}_{\geqslant 0}$ and total weight $W = \sum_{v \in V} w(v)$. In this work, we enforce population balance by rejecting proposals that yield unbalanced forests, similar to the approach in DeFord et al. (2021). Another approach can be found in Autry et al. (2023, 2021), where the authors incorporate balance conditions in the target distribution, allowing for both hard and soft balance constraints.

From a state $x \in \mathcal{X}$, the MEW transition consists of two parts:

1. **Cycle basis step:** Choose an edge $e_+ \in E \setminus T$ according to some distribution ν_1. Let C be the unique cycle in $T \cup \{e_+\}$. Choose an edge $e_- \in C$ according to some distribution ν_2 and subject to the constraint that $e_- \notin M$ (marked edges cannot be removed), and set $T' = (T \cup \{e_+\}) \setminus \{e_-\}$.
2. **Marked edge step:** Choose a marked edge $m \in M$ according to some distribution μ_1. Choose an endpoint $u \in m$ according to some distribution μ_2. Choose a neighbor $v \in N_{T'}(u)$ according to some distribution μ_3, and let $m' = \{u, v\}$. Set $M' = (M \cup \{m'\}) \setminus \{m\}$.

The choice to select a neighbor of an endpoint of the current marked edge, rather than an arbitrary edge from T', is deliberate. By moving the marked edge locally in the tree, the resulting partition is more likely to remain balanced, reducing the rate of rejections due to population imbalance. For the current study and the following calculations, we choose $\nu_1, \nu_2, \mu_1, \mu_2$, and μ_3 to be uniform, but one could easily incorporate other distributions to, for example, avoid repeated selection of high-degree vertices, incentivize the selection of plans with low isoperimetric scores, or disincentivize the splitting of counties or towns as in Clelland et al. (2022); McWhorter and DeFord (2026). A schematic of the procedure as chosen in our study is shown in Figure 1.

An alternate formulation of MEW could take a probabilistic choice between the cycle basis step and the marked edge step. In this variant, with probability p, one would take a cycle basis step, and with probability $1-p$, a marked edge step. This approach has the benefit of symmetric transitions, adding computational simplicity and yielding the theoretical result that, if the walk is irreducible, the unique stationary distribution is the uniform distribution on $\mathcal{X}$. The cycle basis step can be viewed as an up-down step as in the recently proposed BUD Walk (Akitaya et al., 2026) or a 1-cycle step in the cycle walk (DeFord et al., 2025).

3 Transition Probabilities

While the probabilistic variant targets the uniform distribution, for redistricting applications, we often want to target alternative distributions $p(x)$. To achieve this, we employ MEW (performing both a cycle basis step and a marked edge step sequentially) as a proposal in the Metropolis-Hastings algorithm. The algorithm generates samples from a target distribution $p(x)$ by rejection sampling according to an acceptance probability

$$a = \min\left\{1, \frac{p(x')}{p(x)}\frac{P(x|x')}{P(x'|x)}\right\}, \tag{2}$$

where $P(x'|x)$ and $P(x|x')$ are the forward and backward transition probabilities of MEW.

Given two states $x = (T, M)$ and $x' = (T', M')$ (as defined in Section 2), the transition probability $P(x'|x)$ decomposes naturally according to the two components of the tuple (T, M). In this section, we present the transition probabilities under uniform selection distributions $\nu_1, \nu_2, \mu_1, \mu_2$, and μ_3. The full transition probability for general selection distributions is tractable but lengthy and is therefore omitted here.

The tree component of the transition probability depends on the length of the cycle C formed by $T \cup \{e_+\}$, and the constraint that marked edges cannot be removed. The forward transition probability $P(T'|T, M) = (|E \setminus T| \cdot |C \setminus M|)^{-1}$, that is, the probability of choosing the correct edges e_+ and e_-. Taking the ratio,

$$\frac{P(T|T', M')}{P(T'|T, M)} = \mathbb{1}(m' \neq e_+) \cdot \frac{|C \setminus M|}{|C \setminus M'|}, \tag{3}$$

where $\mathbb{1}(X)$ is the indicator function that returns 1 if X is true and 0 otherwise. This term is needed because if the moved marked edge $m' = e_+$, then $P(T|T') = 0$ (since we cannot remove a marked edge).

The marked edge transition probability depends on the vertex degrees in the spanning tree and whether the marked edge configuration changes. Let $\deg_G(u)$ refer to the degree of vertex u in the graph G. Then the forward transition probability is $P(M'|T', M) = 1/|M| \cdot 1/2 \cdot (1/\deg_{T'}(u) + \mathbb{1}(m = m')/\deg_{T'}(v))$, that is, the probability of choosing the correct marked edge m, endpoint u,

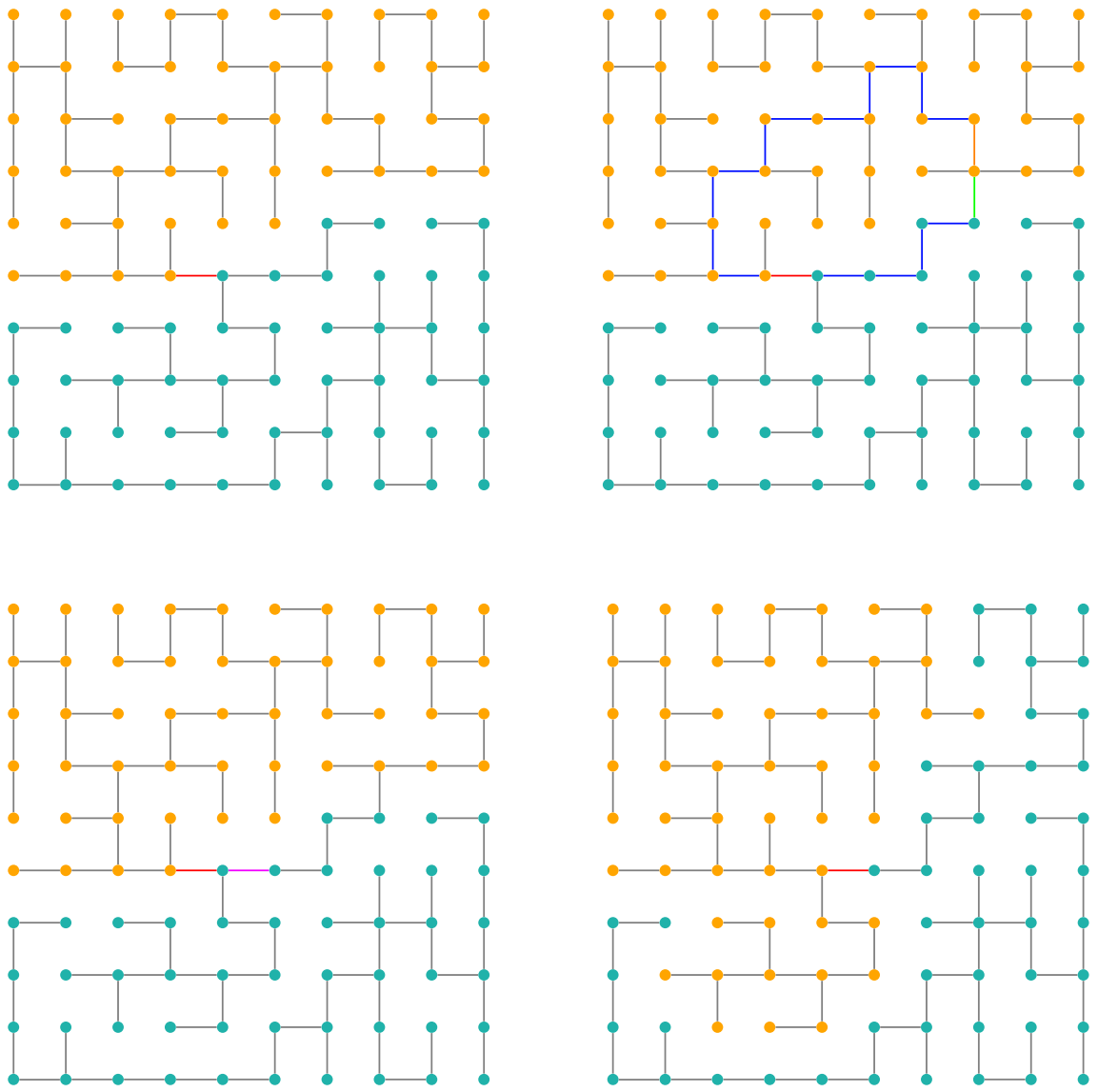


Figure 1: Illustration of MEW's two-step procedure. (Top-left) Initial state with spanning tree (black edges) and marked edges (red) that create a balanced partition when removed. (Top-right) Cycle basis step: adding edge e_+ (green) creates cycle C (blue), then removing edge e_- (orange) yields new tree T'. (Bottom-left) Marked edge step: the right vertex of the marked edge (red) is chosen, and its right neighbor is chosen, yielding a new marked edge (pink). (Bottom-right) These two steps create a new partition, completing one step of the MEW procedure.

and neighbor v. The indicator term accounts for the possibility that when $m = m'$ (the marked edge doesn't move), either endpoint could be selected. Taking the ratio,

$$\frac{P(M|T, M')}{P(M'|T', M)} = \frac{\deg_{T'}(u)}{\deg_T(u)} \cdot \left[\frac{\deg_T(u) + \deg_T(v)}{\deg_{T'}(u) + \deg_{T'}(v)} \cdot \frac{\deg_{T'}(v)}{\deg_T(v)} \right]^{\mathbb{1}(m=m')}. \tag{4}$$

Putting it all together,

$$\frac{P(x|x')}{P(x'|x)} = \mathbb{1}(m' \neq e_+) \cdot \frac{|C \setminus M|}{|C \setminus M'|} \cdot \frac{\deg_{T'}(u)}{\deg_T(u)} \cdot \left[\frac{\deg_T(u) + \deg_T(v)}{\deg_{T'}(u) + \deg_{T'}(v)} \cdot \frac{\deg_{T'}(v)}{\deg_T(v)} \right]^{\mathbb{1}(m=m')}. \quad (5)$$

In most cases, this expression will simplify to unity because most cycles do not intersect the changing marked edge, and the cycle basis walk generally preserves the degree of the marked edges. A simple example where this symmetry is broken is given in Appendix A.

4 Target Distribution

Recall that $\mathcal{X} = \{x = (T, M) : T \text{ is a spanning tree of } G, M \subseteq T\}$ is the lifted space. We define

$$P_d(G) = \{\xi : V \to \{1, 2, \ldots, d\} \mid \forall i, \xi_i \text{ is connected}\}, \quad (6)$$

where ξ_i represents the induced subgraph of G on vertex set $\xi^{-1}(i)$. That is, $P_d(G)$ is the space of d-partitions of G where each part is connected. Since MEW targets a distribution over $\mathcal{X}$, we choose a distribution $p(x)$ over $\mathcal{X}$ that induces a desired distribution $\pi(\xi)$ for $\xi \in P_d(G)$.

To achieve this, we choose a target distribution of the form

$$p(x) \propto \frac{\exp(J(\xi(x)))}{\tau(\xi(x))}, \quad (7)$$

where $\xi(x)$ denotes the connected partition $\xi \in P_d(G)$ corresponding to the state $x \in \mathcal{X}$. The degeneracy factor $\tau : P_d(G) \to \mathbb{Z}_+$ counts the number of states in $\mathcal{X}$ that map to the same partition in $P_d(G)$. The energy function $J : P_d(G) \to \mathbb{R}$ may be user-specified to incorporate various objective measures on $P_d(G)$. In this work, we include in every experiment a population constraint of 5% that is wrapped into the energy function following Autry et al. (2023).

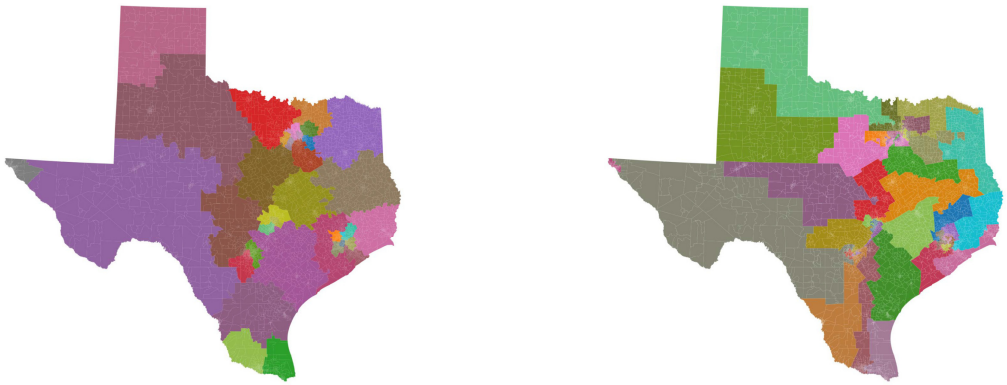
In this framework, multiple lifted states $x \in \mathcal{X}$ can correspond to the same partition $\xi \in P_d(G)$. The degeneracy factor

$$\tau(\xi) = \Pi_{i=1}^{d} t(\xi_i) \cdot t(Q), \quad (8)$$

accounts for this, where $t(\xi_i)$ counts the number of spanning trees in component ξ_i, and $Q = G \backslash \xi$ is the quotient graph.

It is important to note that while the degeneracy factor $\tau(\xi)$ corrects for the multiplicity of lifted states mapping to the same partition, it does not account for the distribution of $P_d(G)$ itself. Let $\pi(\xi) \propto f(E(\xi))$ be the target distribution, where f is a desired function of an observable E. The induced marginal distribution over E is $P(E) = f(E) \cdot \Omega(E)$, where $\Omega(E) = \sum_{\xi \in P_d(G)} \mathbb{1}(E(\xi) = E)$ is the multiplicity of partitions with observable $E(\xi) = E$. Thus, we only recover $P(E) = f(E)$ when $\Omega(E)$ is (approximately) constant. DeFord et al. (2021) showed that there are far more partitions with higher cut edge counts than those with fewer cut edges, suggesting $\Omega(E)$ is not constant in this case. Therefore, when targeting a distribution based on cut edges as in Sections 5.4 and 5.6, we may expect to see a bias towards less compact partitions.

This choice of target distribution allows us to sample in the lifted space $\mathcal{X}$ under a target distribution over $P_d(G)$. One reason this is so important is that, especially on large dual graphs, the spanning tree count of partitions is so astronomically large that it dwarfs other constraints that we might try to build into an energy function. For example, in Figure 2, we show two possible redistricting plans of Texas's congressional districts. The map on the left is drawn from the



(a) (b)

Figure 2: Comparison of a sampled Texas redistricting plan (a) and the enacted 2020 plan (b). The spanning tree products differ by approximately 10^{466}, demonstrating the extreme and impractical nature of spanning tree-based probability weighting in real-world applications.

uniform distribution on $\mathcal{X}$, corresponding to the spanning tree distribution, and the map on the right is the 2020 enacted plan. The sampled plan has a degeneracy factor of $\tau(\xi_{\text{sampled}}) \approx 10^{5160}$, and the enacted plan has a degeneracy factor of $\tau(\xi_{\text{enacted}}) \approx 10^{4694}$. Thus, when sampling from the spanning tree distribution, the sampled plan is about 10^{466} times more likely to be sampled than the enacted plan. To put the number 10^{466} into perspective: there are an estimated 10^{80} atoms in the observable universe. So, the sampled plan is more likely than the enacted plan by a factor that is astronomically larger than the number of atoms in the universe. This demonstrates that the spanning tree distribution gives enormous weight to highly compact plans, whereas enacted plans tend to have far more cut edges. This motivates the need for an algorithm that can shift the sampling distribution towards plans with a target number of cut edges, or more generally, encode policy-based criteria into the sampling process.

Previous studies discount the degeneracy factor (Autry et al., 2023, 2021), taking a target of the form $p(x) \propto \exp(J(\xi(x)))/\tau(\xi(x))^{\gamma}$, where $0 \leqslant \gamma < 1$. This only partially accounts for the spanning tree bias with the goal of achieving better mixing and convergence results. However, this partial correction still leaves the sampler heavily biased towards plans with large spanning tree counts. MEW's ability to maneuver locally in $\mathcal{X}$ allows us to use the full degeneracy factor ($\gamma = 1$), properly accounting for all configurations. This allows us to target a broader class of policy-based distributions over $P_d(G)$ without the overwhelming bias from spanning tree counts, allowing us to focus on other criteria such as compactness, competitiveness, or partisan symmetry measures, even when these criteria remain implicitly related to spanning trees.

5 Results

We evaluate the efficacy of MEW through numerical experiments on three real-world graphs with varying energy functions J. First, we target the uniform distribution over partitions of the dual graph of Cheshire County, an example that permits comparison to an enumerated baseline. Second, we target the spanning tree distribution on the dual graph of New Hampshire, and since ReCom draws independent samples from this distribution when $d = 2$, we can use it as

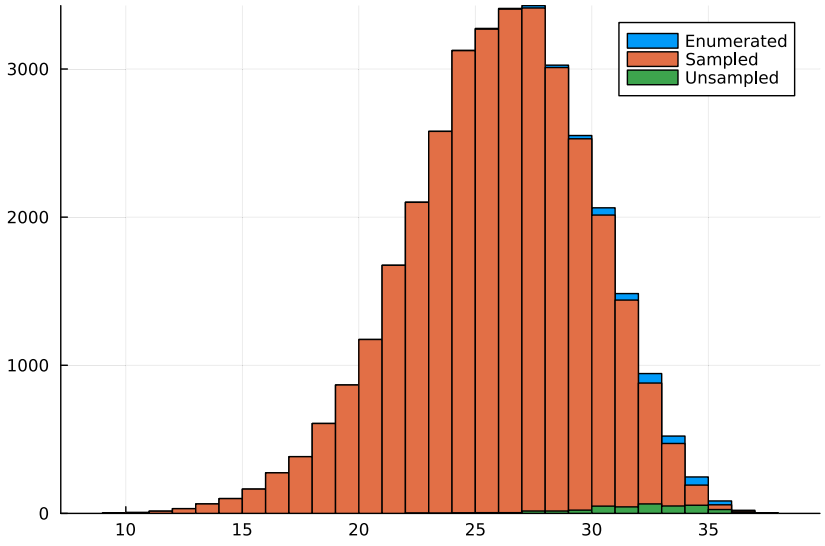


Figure 3: Comparison of cut edges for enumerated, sampled, and unsampled partitions in Cheshire County's dual graph. After running MEW for 1 million steps, 99% of the 34,225 balanced 2-partitions were sampled. Unsampled partitions (green) tend to have higher cut edges than sampled partitions (orange).

a baseline. Third, we target competitiveness, compactness, and multivariate distributions over the dual graph of New Hampshire. Finally, we target a multivariate distribution on the dual graph of Texas's voting precincts to validate MEW's scalability to large dual graphs with many districts. Our results demonstrate strong evidence of mixing in all cases.

5.1 Cheshire County

The dual graph of Cheshire County, New Hampshire, has 27 vertices, 63 edges, and 34,225 balanced 2-partitions. By enumerating all the partitions in Cheshire County using the enumpart algorithm introduced in Fifield et al. (2020b), we can directly compare them with an ensemble generated by MEW. When we run MEW for 1 million steps targeting the uniform distribution on partitions ($J(\xi(x)) = \text{const.}$), we successfully sample 99% of the partitions. The cut edge counts of the sampled, unsampled, and enumerated partitions are shown in Figure 3. The unsampled partitions have more cut edges than the sampled plans. This is a similar result to the observed behavior of MFR (Autry et al., 2023). After 10 million runs, MEW successfully samples all of the enumerated partitions, and the gap disappears.

What we observe here are the two opposing forces of MEW. One force is the force towards more compact plans. Since the cycle basis walk targets the uniform distribution on spanning trees, partitions that have higher spanning tree counts will be favored by the proposal step. Thus, we will systematically under-sample less compact plans. The opposite force pushes the walk towards less compact plans. The degeneracy factor τ in the target distribution encodes a higher acceptance probability for plans with lower spanning tree counts. When these two forces are in balance, the chain is successful.

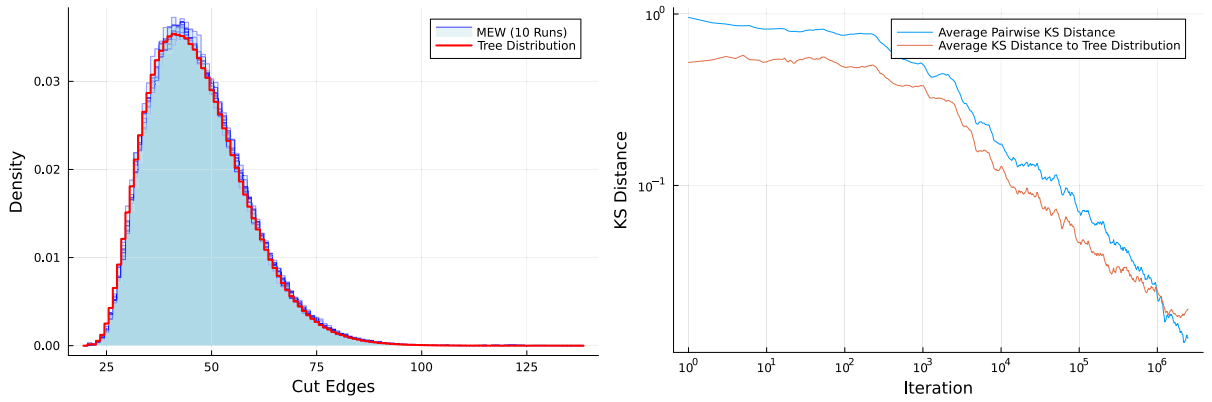


Figure 4: (a) Comparison of MEW (blue) to a recombination baseline (red) when sampling with a target of the spanning tree distribution. (b) Decay of the average pairwise KS distance (blue) and the average Kolmogorov-Smirnov (KS) distance to the spanning tree distribution (red). After many iterations, MEW successfully samples from the spanning tree distribution, as evidenced by its convergence to the baseline.

5.2 The Spanning Tree Distribution

We investigate the spanning tree distribution as a target distribution for two primary reasons. Since much attention has been devoted to its study, researchers may want to use MEW to target the distribution or related distributions. Secondly, since many algorithms are available to target the distribution, we can use it as another baseline to assess convergence. To this end, we employ ReCom, which, in the $d = 2$ case, draws independent samples from the spanning tree distribution; each partition has probability proportional to the product of the spanning tree product of each part and the spanning tree product of the quotient graph. Notice that this is exactly the uniform distribution on $\mathcal{X}$, thus we set

$$J(\xi(x)) = \ln(\tau(\xi(x))). \tag{9}$$

We note that this differs slightly from the definition of the spanning tree distribution given earlier, which omits the quotient graph term. However, this term is negligible in practice and does not materially affect the distribution.

As we can see in Figure 4, MEW successfully samples from this distribution. The average pairwise Kolmogorov-Smirnov (KS) distance between the chains decays to nearly zero (0.0124 after 2.5 million steps), and the average KS distance from the chains to the baseline also decays to nearly zero (0.0188 after 2.5 million steps). The figure also reveals a limitation of sampling from the spanning tree distribution and related distributions. When sampling less compact partitions to target realistic plans (the enacted plan has 72 cut edges) or to investigate the effects of highly noncompact districts (targeting plans with 100 or more cut edges), we would see long mixing times and high rejection rates due to the distribution's naturally compact structure.

5.3 Competitiveness Distribution

New Hampshire has two congressional districts, and previous investigations show that the seats are occupied by either two Democrats or one Democrat and one Republican (Palmer et al., 2024).

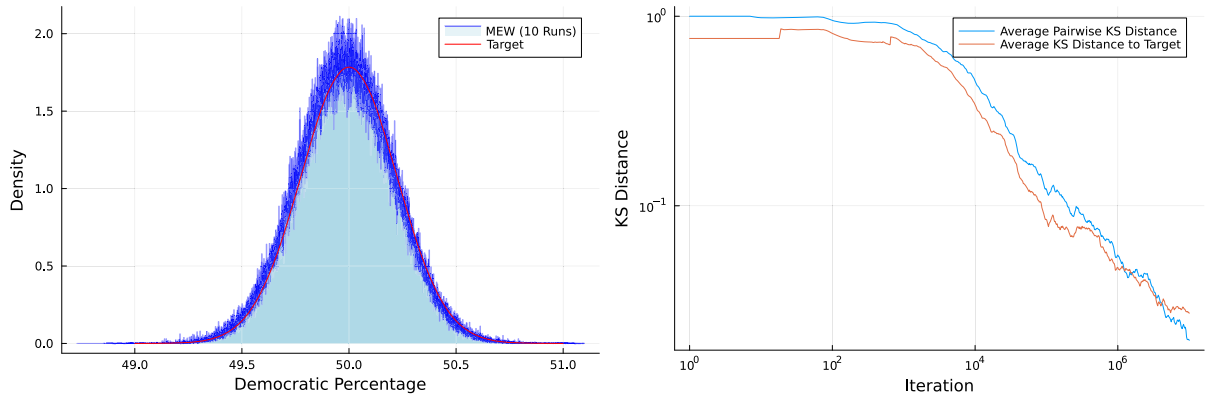


Figure 5: (a) Comparison of the sampled distribution (blue) with the target distribution (red), showing approximate agreement. (b) Average pairwise KS distance (blue) and average KS distance to the target distribution (red) vs iteration, showing decay to nearly zero after 10 million steps. These diagnostics suggest MEW is effectively sampling from the target distribution.

To ensure competitiveness in the second district, we construct an ensemble of redistricting plans with a tight race in the second district. We encode this in the energy function as

$$J(\xi(x)) = -10 \cdot (p - 0.5)^2, \tag{10}$$

where p represents the share of Democratic votes in the second district. Then the induced energy function on $P_2(G)$ is a normal distribution centered on 50% with variance 1/20.

In Figure 5 (a), we plot the resultant distribution and can see that it roughly agrees with the target distribution. In Figure 5 (b), we plot the average pairwise KS distance and the average KS distance to the target distribution, showing that both decay to nearly zero after 10 million steps (0.019 and 0.027 respectively). Given these heuristics, we are confident that MEW is successfully sampling from this target distribution.

This is the first example in our experiments of targeting a Gaussian distribution. This is distinct from prior approaches, which typically use exponential energy functions that push the sampler towards more competitive or compact plans. By contrast, targeting a Gaussian centers the sampling window around a target value, such as 50% Democratic vote share, and the resulting ensemble closely approximates the target when $\Omega(E)$ is approximately constant in the region of interest. This allows for a new ability to evaluate proposed legislative changes: by shifting the target distribution, one could assess the resulting partisan outcomes before any maps are drawn.

5.4 Compactness Distribution

When sampling from the uniform distribution over graph partitions, Najt et al. (2019) showed that most plans in the space are extremely non-compact. To create a meaningful and realistic ensemble, we target compact plans. The number of cut edges in a partition has become a popular metric for compactness in mathematical settings due to its low complexity and its relation to the border length of spanning forests (Clelland et al., 2021; Procaccia and Tucker-Foltz, 2022). To target realistic, compact plans, we assume an energy function

$$J(\xi(x)) = -0.1 \cdot (c - 72)^2, \tag{11}$$

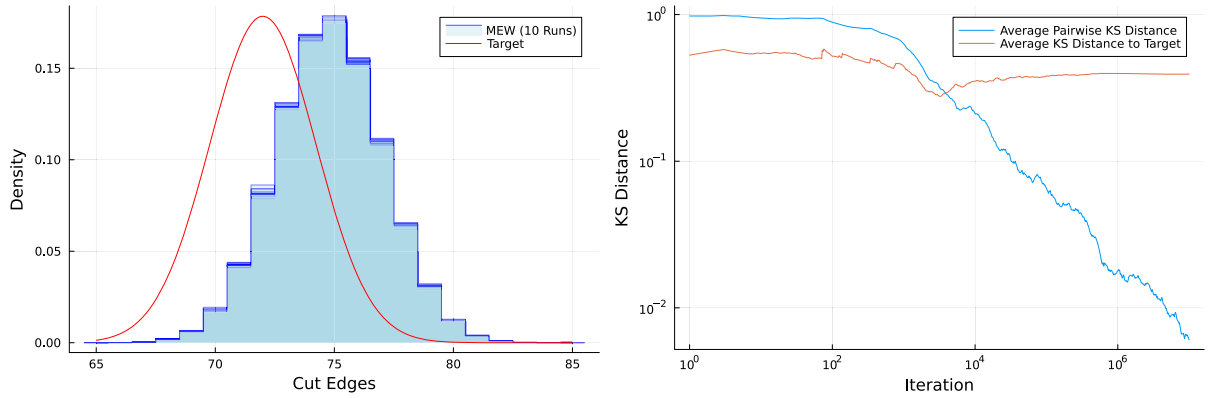


Figure 6: (a) Comparison of the sampled distribution (blue) against the target distribution (red). The empirical distribution is shifted from the target across all 10 chains. (b) Average pairwise KS distance (blue) and average KS distance to the target distribution (red). The KS distance to the target plateaus whereas the pairwise KS distance decays to nearly zero, evidence of convergence to the shifted distribution.

where c represents the number of cut edges. As before, this gives us a Gaussian centered on 72 with a variance of 5. The value 72 corresponds to the number of cut edges in the enacted plan.

With this target distribution, we do not have the same tight agreement as before. In Figure 6 (a), we see that the empirical distribution is shifted relative to the target. However, this shifted distribution shows great evidence of convergence. In Figure 6 (b), we see that the pairwise KS distance decays to nearly zero. After 10 million steps, the average pairwise KS distance is 0.0061.

Following DeFord et al. (2021), we hypothesize that $P_2(G)$ is exponentially distributed with respect to cut edges in the region of interest for redistricting. This exponential distribution imposes a shift on the sampled distribution. In Appendix B, we conduct an experiment with a variable constructed specifically to have an exponential distribution and observe a similar shift behavior.

5.5 Multivariate Distribution

In practical applications, researchers will need to encode various constraints into their energy functions. To investigate the behavior of the chain when sampling from a multivariate distribution, we may simply take a linear combination of the competitiveness and compactness energy functions. This yields a multivariate Gaussian with zero correlation. This energy function has four tunable parameters; the means control the center of the distribution, and the weights control the spread in each of the principal axes. A general energy function of this form is

$$J(\xi(x)) = -\beta_1(p - \mu_p)^2 - \beta_2(c - \mu_c)^2, \tag{12}$$

where β_1 and β_2 are the weights on vote percentage and cut edges, respectively, p represents the Democratic party's vote share in the second district, and c is the number of cut edges.

In Figure 7, we plot the results of a convergence analysis designed to determine the parameter range within which MEW converges. In this analysis, 10 chains of 100 thousand steps are run at each parameter combination, and we plot the average pairwise (2-dimensional) KS distance in a heatmap. In Figure 7 (a), we fix $\beta_1 = 10$ and $\beta_2 = 0.1$, and search a grid of target mean

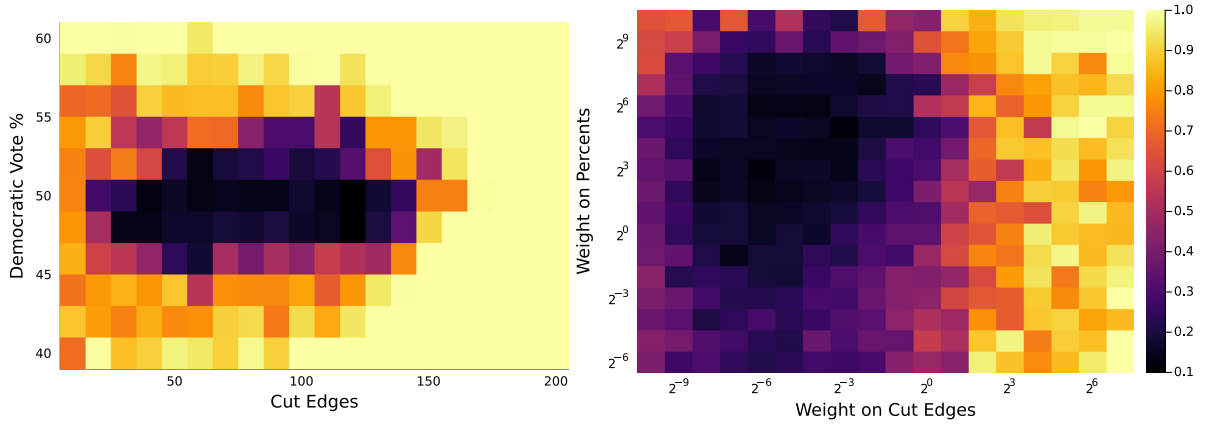


Figure 7: (a) Average pairwise KS distances as a function of target distribution means. The black region indicates the chain's convergence zone after 100 thousand steps, centered on the practical region of interest (72 cut edges, 50% Democratic vote share). Poor convergence occurs outside this zone. (b) Average pairwise KS distances as a function of target weights, showing a broader convergence region. Convergence degrades sharply for cut edge weights $\beta_2 > 1$.

pairs. The black region represents a conservative estimate of the convergence region of the chain (since each chain only ran for 100 thousand steps). The convergence region covers the practical zone of interest, centered on 72 cut edges and 50% Democratic vote, and the worst convergence is observed well outside of this zone.

In Figure 7 (b), we fix $\mu_p = 0.5$ and $\mu_c = 72$ and search a grid of target weight pairs. The convergence region is much larger than before, and we have a similar result to above; the convergence region contains the zone of practical interest. Rarely would researchers assign $\beta_2 > 1$, as this targets a very narrow distribution. Convergence falls off rapidly for cut edge weights larger than 1; any plan that differs from the target mean of cut edges will do so by several standard deviations. In the percentage axis, the convergence region is much wider.

5.6 Texas

Many applications will be to states with much larger dual graphs and multiple districts. To test MEW's performance on such examples, we run the chain on the dual graph of Texas, which contains 8,933 vertices and 24,514 edges (compared to New Hampshire's 320 vertices and 854 edges). Again, we take a target distribution of a bivariate Gaussian on compactness and political metrics.

For compactness, we use the number of cut edges, and in the place of a competitiveness metric, we choose the Mean-Median score, a partisan symmetry metric. The score compares a party's average vote share to its median vote share across districts, with a positive score indicating an advantage for that party. While this metric has noted limitations (DeFord and Veomett, 2026; Ratliff et al., 2025), as do all single-metric approaches, the ability to target explicit distributions on single metrics enables researchers to explore metric-specific implications in redistricting analysis.

With these two metrics, we target an energy function of the same form as Equation (12), but now with p representing the mean-median score, $\mu_p = 0$ (the mean-median score prescription for a perfectly 'fair' plan), $\beta_1 = 100{,}000$, $\mu_c = 3{,}346$ (the enacted plan value), and $\beta_2 = 0.01$.

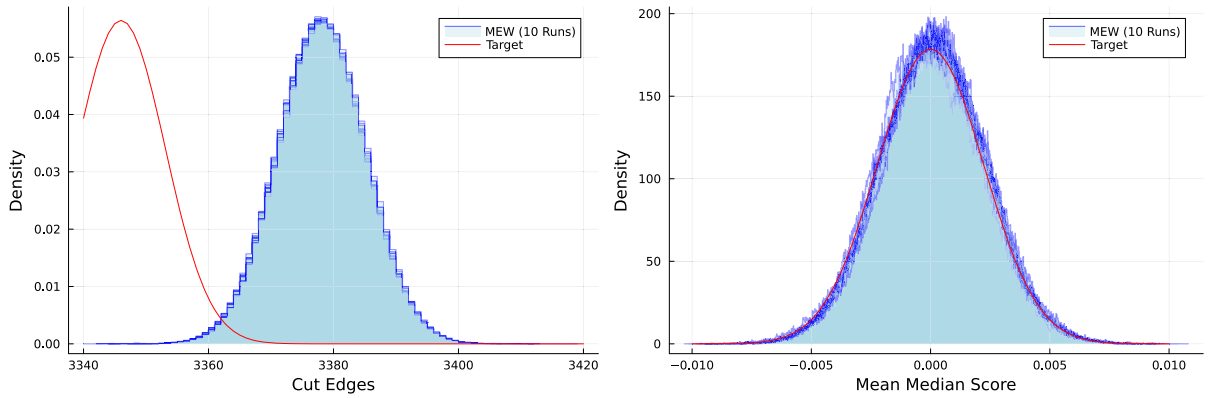


Figure 8: Marginal distributions for Texas. (a) Cut edges distribution displaying the characteristic shift observed in compactness metrics, centered at the enacted plan value of 3,346 cut edges. (b) Mean-Median score distribution showing convergence around the target value of 0.

Under this target distribution, we see good evidence of convergence. Over ten chains of 4 million steps each, the average pairwise (2-dimensional) KS distance between chains is 0.0285. In Figure 8 we plot the marginal distributions. As in Section 5.4, we observe a shift in the sampled marginal distribution of cut edges.

6 Conclusion and Discussion

In this paper, we introduce the marked edge walk (MEW), a novel MCMC algorithm for sampling from the space of balanced graph partitions under a tuneable target distribution. The procedure combines a step of the cycle basis walk on spanning trees with a marked edge transition, and this balance of global and local steps allows for good exploration of the state space while maintaining easily calculable transition probabilities. Unlike previous approaches such as the Metropolized flip walk or Metropolized Forest ReCom (MFR), MEW mixes well under target distributions unrelated to the spanning tree distribution.

When targeting the uniform distribution on the dual graph of Cheshire County, the spanning tree distribution, and a competitiveness distribution on the dual graph of New Hampshire, the chain exhibits fast mixing evidenced by rapidly decaying pairwise KS distance and KS distance to the target distribution.

When targeting a compactness distribution based on cut edges, the empirical distribution is shifted from the target. However, the chains maintain strong convergence properties, with pairwise KS distances decaying to nearly zero. We hypothesize that this shift is due to the underlying exponential structure of $P_d(G)$. This hypothesis is supported by the observation of a similar shift observed on an observable with a known exponential distribution. MEW also converges well to a multivariate target distribution on the dual graph of New Hampshire and Texas, displaying a wide region of convergence and scalability to large graphs with many districts.

One limitation of MEW is computational efficiency. Each acceptance ratio includes calculating the spanning tree count of the districts, which due to Kirchhoff's matrix-tree theorem is straightforward while computationally taxing. New Hampshire's dual graph contains only 320 vertices, while applications to other states (such as Texas with 8,933 vertices) see greatly increased run-times. The MFR algorithm includes memoization of this calculation to increase time

efficiency. Additionally, the observed shift in the ensemble when targeting a compactness distribution suggests MEW favors plans that are slightly less compact than the target distribution. Future work on the algorithm could explore ways to correct for the bias, possibly by adjusting the target distribution or proposal step. Because we hypothesize that the shift is a result of the underlying structure of $P_d(G)$, research into quantifying this structure for different graph types would allow us to correct for the shift in the target distribution.

The ability to target a broader class of distributions opens possibilities for prospective redistricting analysis. The ensemble method has primarily been used to evaluate existing or proposed maps against a distribution of alternatives. However, MEW's capacity to sample from specified target distributions, particularly non-optimization distributions, allows for a new use case: evaluating the expected consequences of proposed redistricting criteria before maps are drawn. For example, policymakers considering the impact of population deviation strictness on partisan outcomes could target different population-based distributions and observe how the resulting ensembles differ in partisan outcomes. While such applications would require careful consideration of the biases we observe (such as the exponential tilt in Section 5.4), they represent a potential shift from retrospective evaluation to prospective policy analysis.

And lastly, an important research direction is incorporating MEW in optimization schemes. Its calculable transition probabilities and favorable mixing make it a great candidate for tempering schemes like those used in Autry et al. (2021); Chuang et al. (2025). Because MEW can target a broader class of distributions, it could optimize for objectives that have been difficult to address with prior proposals.

Supplementary Material

The data and code needed to replicate these findings can be found at https://github.com/amcwhorter/MEW/. All computations were performed on doob, a 96-core 3.6GHz AMD Epyc system managed by the Dartmouth College Department of Mathematics.
The supplementary material for this article is available on the official website of the Journal of Data Science at https://jds-online.org/. It may include additional results, appendices, tables, figures, code, and data.

A Asymmetric Transition Probability

Here, we include an example transition that breaks symmetry from a state $x = (T, M)$ (Figure A1, left) to $x' = (T', M')$ (Figure A1, right). In this case, the forward transition probability, $P(M'|M, T') \propto 1/\deg_{T'}(u) = 1/3$. In the reverse transition, $P(M|M', T) \propto 1/\deg_{T}(u) = 1/2$. Since the neighborhood of u changes between T and T', the forward and reverse transition probabilities are not the same.

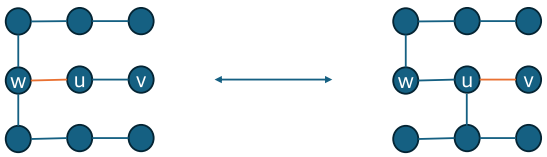


Figure A1: Example states $x = (T, M), x' = (T', M')$ that break symmetry.

B Exponential Shift Experiment

Because of the unknown structure of $P_d(G)$, there is no guarantee that the sampled distribution of an observable will show perfect agreement with the target distribution. In this example, we hypothesize that the underlying distribution of $P_2(G)$ is exponentially distributed with respect to cut edges, as discussed in DeFord et al. (2021).

To investigate this hypothesis, we first consider a toy model. Consider a Markov chain that draws independent samples from an exponential distribution $g(x) = \lambda \cdot \exp(-\lambda x)$ and uses Metropolis-Hastings rejection to target a Gaussian distribution of the form $p(x) \propto \exp(-\beta(x-\mu)^2)$. The acceptance ratio would include the ratio of the two distributions, which simplifies to a shifted Gaussian with mean $\hat{\mu} = \mu + \lambda/2\beta$ and variance $\hat{\sigma}^2 = 1/2\beta$. If we do not take into account the effect of the underlying exponential distribution, we are effectively targeting an exponentially tilted distribution with mean $\hat{\mu} = \mu - \lambda/2\beta$ and variance $\hat{\sigma}^2 = 1/2\beta$. If our hypothesis were true, when sampling from an underlying exponential distribution with parameter λ, we expect a shift $\partial\mu = \hat{\mu} - \mu = -\lambda/2\beta$.

We then construct an observable with a known exponential distribution. We assign each node $n \in V$ a weight w_n drawn independently and uniformly from $[0, 1]$. Let $\mathbf{s}$ denote the vector of total weights summed across parts. For sufficiently large parts, the central limit theorem ensures each component $\mathbf{s}_i \sim \mathrm{N}(|\xi_i|/2, |\xi_i|/12)$, where $|\xi_i|$ denotes the number of nodes in ξ_i.

Without loss of generality, we focus on $\mathbf{s}_1$. To construct an exponentially distributed variable $Z \sim \mathrm{Exp}(\lambda)$, we apply a probability integral transform. First, we standardize $\mathbf{s}_1$ and apply its CDF.

$$U = \Phi\left(\frac{\mathbf{s}_1 - |\xi_1|/2}{\sqrt{|\xi_1|/12}}\right), \tag{13}$$

where Φ is the standard normal CDF. Since $U \sim \mathrm{Uniform}(0, 1)$, we then apply the inverse transform

$$Z = -\frac{1}{\lambda}\ln(1-U) \sim \mathrm{Exp}(\lambda). \tag{14}$$

With the exponentially distributed Z as our variable of interest, we run 10 chains of 1 million steps each, targeting a normal distribution with mean $\mu = 2$ and variance $\sigma^2 = 1/2\beta$. Recall that our toy model predicts that we will converge to a Gaussian with mean $\hat{\mu} = \mu - \lambda/2\beta$ and variance $\hat{\sigma}^2 = 1/2\beta$. First, fixing $\lambda = 1$ and varying β, we observe a shift $\partial\mu \approx -0.57(\beta^{-1}) + 0.01$ with correlation coefficient $r = -0.97$. Additionally, the sampled variance satisfies $\hat{\sigma}^2 \approx 0.49\beta^{-1}$ with correlation coefficient $r = 0.999$. When varying the exponential parameter (with fixed $\beta = 32$), we find $\partial\mu \approx -0.87(\lambda/\beta) + 0.01$ with correlation coefficient $r = -0.97$ while λ is uncorrelated with the sampled variance.

Overall, this experiment demonstrates that MEW exhibits systematic shifts when sampling over an exponentially distributed observable, and the behavior closely resembles that predicted by rejection sampling of independent draws from an exponential distribution. While the empirical coefficients deviate slightly from theoretical predictions, this supports the hypothesis that the underlying distribution of $P_d(G)$ with respect to cut edges is approximately exponential.

Acknowledgement

The authors thank the editor, associate editor, and the Journal of Data Science for considering our manuscript for publication. We also thank Kristopher Tapp and Greg Herschlag for helpful conversations.

Funding

This work was funded by the Dartmouth Fellowship to Atticus McWhorter and by the Alfred P. Sloan Foundation (G-2021-16778) and the National Science Foundation (DMS-1928930) which supported Daryl DeFord during his visit to SLMath in 2023 where this collaboration began.

References

Akitaya HA, Cannon S, Herschlag G, Schoenbach G, Tapp K, Tucker-Foltz J (2026). The balanced up-down walk. arXiv preprint: https://arxiv.org/abs/2602.11993.

Autry E, Carter D, Herschlag GJ, Hunter Z, Mattingly JC (2023). Metropolized forest recombination for Monte Carlo sampling of graph partitions. *SIAM Journal on Applied Mathematics*, 83(4): 1366–1391. https://doi.org/10.1137/21M1418010

Autry EA, Carter D, Herschlag GJ, Hunter Z, Mattingly JC (2021). Metropolized multiscale forest recombination for redistricting. *Multiscale Modeling & Simulation*, 19(4): 1885–1914. https://doi.org/10.1137/21M1406854

Cai JY, Kruse J, Mayer K, Szabo DP (2024). A uniformly random solution to algorithmic redistricting. arXiv preprint: https://arxiv.org/abs/2402.13868.

Caldera S, DeFord D, Duchin M, Gutekunst SC, Nix C (2020). Mathematics of nested districts: The case of Alaska. *Statistics and Public Policy*, 7(1): 39–51. https://doi.org/10.1080/2330443X.2020.1774452

Cannon S, Duchin M, Randall D, Rule P (2022). Spanning trees and redistricting: new methods for sampling and validation. arXiv preprint: https://arxiv.org/abs/2210.01401.

Cannon S, Pegden W, Tucker-Foltz J (2024). Sampling balanced forests of grids in polynomial time. In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC 2024)*, 1676–1687. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3618260.3649699

Cannon S, Pegden W, Tucker-Foltz J (2025). Sampling tree-weighted partitions without sampling trees. arXiv preprint: https://arxiv.org/abs/2508.11130.

Charikar M, Liu P, Liu T, Vuong TD (2023). On the complexity of sampling redistricting plans. arXiv preprint: https://arxiv.org/abs/2206.04883.

Chikina M, Frieze A, Mattingly JC, Pegden W (2020). Separating effect from significance in Markov chain tests. *Statistics and Public Policy*, 7(1): 101–114. https://doi.org/10.1080/2330443X.2020.1806763

Chikina M, Frieze A, Pegden W (2017). Assessing significance in a Markov chain without mixing. *Proceedings of the National Academy of Sciences*, 114(11): 2860–2864. https://doi.org/10.1073/pnas.1617540114

Chuang G, Herschlag G, Mattingly JC (2025). Multiscale parallel tempering for fast sampling on redistricting plans. *Multiscale Modeling & Simulation*, 23(4): 1515–1550. https://doi.org/10.1137/24M1635806

Clelland J, Colgate H, DeFord D, Malmskog B, Sancier-Barbosa F (2022). Colorado in context: Congressional redistricting and competing fairness criteria in Colorado. *Journal of Computational Social Science*, 5(1): 189–226. https://doi.org/10.1007/s42001-021-00119-7

Clelland J, DeFord D, Duchin M (2020). Aftermath: The ensemble approach to political redistricting. *Math Horizons*, 27(3): 34–35. https://doi.org/10.1080/10724117.2019.1680187

Clelland JN, Bossenbroek N, Heckmaster T, Nelson A, Rock P, VanAusdall J (2021). Compactness statistics for spanning tree recombination. arXiv preprint: https://arxiv.org/abs/2103.02699.

DeFord D, Duchin M (2019). Redistricting reform in Virginia: Districting criteria in context. *Virginia Policy Review*, 12(2): 120–146.

DeFord D, Duchin M, Solomon J (2020). A computational approach to measuring vote elasticity and competitiveness. *Statistics and Public Policy*, 7(1): 69–86. https://doi.org/10.1080/2330443X.2020.1777915

DeFord D, Duchin M, Solomon J (2021). Recombination: A family of Markov chains for redistricting. *Harvard Data Science Review*, 3(1). https://doi.org/10.1162/99608f92.eb30390f

DeFord D, Veomett E (2026). Bounds and bugs: The limits of symmetry metrics to detect partisan gerrymandering. *Election Law Journal: Rules, Politics, and Policy*, 25(2): 127–153. https://doi.org/10.1089/elj.2024.0038

DeFord DR, Herschlag G, Mattingly JC (2025). A cycle walk for sampling measures on spanning forests for redistricting. arXiv preprint: https://arxiv.org/abs/2509.08629.

Donnay C, Kahle M (2025). Asymptotics of redistricting the $n \times n$ grid. *The American Mathematical Monthly*, 132(9): 856–866. https://doi.org/10.1080/00029890.2025.2542711

Fifield B, Higgins M, Imai K, Tarr A (2020a). Automated redistricting simulation using Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 29(4): 715–728. https://doi.org/10.1080/10618600.2020.1739532

Fifield B, Imai K, Kawahara J, Kenny CT (2020b). The essential role of empirical validation in legislative redistricting simulation. *Statistics and Public Policy*, 7(1): 52–68. https://doi.org/10.1080/2330443X.2020.1791773

Frieze A, Pegden W (2023). Subexponential mixing for partition chains on grid-like graphs. In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 3317–3329. SIAM, Philadelphia, PA. https://doi.org/10.1137/1.9781611977554.ch127

Herschlag G, Kang HS, Luo J, Graves CV, Bangia S, Ravier R, Mattingly JC (2020). Quantifying gerrymandering in North Carolina. *Statistics and Public Policy*, 7(1): 30–38. https://doi.org/10.1080/2330443X.2020.1796400

Mathematicians, Law Professors, and Students (amici curiae) (2019). Amicus brief of mathematicians, law professors, and students in support of the appellees and affirmance. Brief for amici curiae, *Rucho v. Common Cause*, Nos. 18-422, 18-726, Supreme Court of the United States, filed March 8, 2019. Counsel: G.-U. E. Charles, D. Gupta (counsel of record), and D. Wilf-Townsend. https://www.supremecourt.gov/DocketPDF/18/18-726/91448/20190308182045139_18-422%2018-726%20Amicus%20BOM%20Mathematicians.pdf.

McCartan C, Imai K (2023). Sequential Monte Carlo for sampling balanced and compact redistricting plans. *The Annals of Applied Statistics*, 17(4): 3300–3323. https://doi.org/10.1214/23-AOAS1763

McWhorter A, DeFord D (2026). Free elections in the free state: Ensemble analysis of redistricting in New Hampshire. *Journal of Computational Social Science*, 9(2): 35. https://doi.org/10.1007/s42001-026-00466-3

Najt E, DeFord D, Solomon J (2019). Complexity and geometry of sampling connected graph partitions. arXiv preprint: https://arxiv.org/abs/1908.08881.

Najt E, DeFord D, Solomon J (2021). Empirical sampling of connected graph partitions for redistricting. *Physical Review E*, 104(6): 064130. https://doi.org/10.1103/PhysRevE.104.064130

O'Sullivan P, Imai K, McCartan C (2026). Generalized sequential Monte Carlo sampling for

redistricting simulation. arXiv preprint: https://arxiv.org/abs/2603.22188.

Palmer M, Schneer B, DeLuca K (2024). A partisan solution to partisan gerrymandering: The define–combine procedure. *Political Analysis*, 32(3): 295–310. https://doi.org/10.1017/pan.2023.39

Procaccia AD, Tucker-Foltz J (2022). Compact redistricting plans have many spanning trees. In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 3754–3771. SIAM, Philadelphia, PA.

Ratliff T, Somersille S, Veomett E (2025). Don't trust a single gerrymandering metric. *La Matematica*, 4(3): 764–809. https://doi.org/10.1007/s44007-025-00172-y

Zhao Z, Hettle C, Gupta S, Mattingly JC, Randall D, Herschlag GJ (2022). Mathematically quantifying non-responsiveness of the 2021 Georgia congressional districting plan. In: *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, Article 15, 1–11. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3551624.3555300