nature energy

Supplementary information https://doi.org/10.1038/s41560-025-01871-0

Strategies to accelerate US coal power phase-out using contextual retirement vulnerabilities

In the format provided by the authors and unedited

Contents

A Supplementary Notes

A.1 Data

Please see Tables for references on the specific datapoints used and their sources (Supplemental Table 1), generator-level data aggregation to produce our plant-level dataset (Supplemental Table 2), additional information on missing data (Supplemental Table 3), and variables used in graph construction vs. those only used for post-analysis (Supplemental Table 1).

A.1.1 Description

Our study employs a comprehensive dataset focused on plant-level data, complemented by some aggregated generator-level statistics. Encompassing a total of 198 coal plants, the dataset excludes combined heat and power (CHP) coal plants and those retired by 2022 to ensure relevance to our analysis (see Supplemental Table 4 for information on the percent of coal plant capacity dropped by only including select plant sectors). Additionally, many CHP plants were missing a large percentage of data, so we felt it prudent, as they also make up such a small percent of capacity, to remove them from the analysis. Each plant's age is computed as of 2023, considering both chronological and generator-weighted age, resulting in an age distribution ranging from 10 to 69 years old, with a non-weighted mean age of 42.3 years old and a weighted mean age of 41.9 years old. While retirements commonly occur at the generator level, we made the decision to perform the analysis at the plant level for reasons related to data availability. About 80% of the 62 variables are plant-level; there are only 13 generator level variables. See Supplemental Table 2 for the full list of these generator variables. Most generator-level variables are related to timing and status of retirement, capacity, and capacity factor (primarily technical criteria). Most financial data are only available at the plant-level.

Our model includes total and levelized pollutant values to comprehensively evaluate the environmental impact of coal plants. The use of both metrics enhances the robustness of our results. Contextual information such as plant location, grid power mix, and general facility details are also incorporated to provide a comprehensive backdrop to each individual coal plant. As well as including plant emissions, impact of coal plants on the population is evaluated based on population density within 3 miles of a plant and data from pollution transport model used to estimate coal plant health impacts provided by the Sierra Club. This dual approach ensures a comprehensive understanding of potential effects on nearby communities and the regions/people actually affected by pollutants. Additionally, community-level impact evaluation is supplemented through the inclusion of demographc-index metrics on the communities within 3 miles of coal plants, enabling us to understand local impacts on disadvantaged and low-income communities, as well as coal plant emission levels and their health impacts.

To gauge economic efficiency, we utilize marginal cost of energy estimates (MCOE) represented by the *Forward Cost of Coal*. MCOE serves as a key metric in determining a plant's economic competitiveness in meeting current demand. Complementing MCOE, cashflow data from RMI is employed to assess the long-term financial viability and potential profitability of coal plants. Both 10-year and 1-year cashflows are analyzed to provide a nuanced understanding of how the evolving energy landscape impacts overall plant profitability. We include data from a financial analysis comparing coal plants to nearby renewable Levelized Costs of Electricity (LCOEs). This includes the percent difference and identification of the cheaper renewable resource, offering insights into the comparative economics of coal and renewables throughout the US. In consideration of environmental/financial implications, we incorporate the emissions control retrifit expenditures, alongside their corresponding levelized costs. This broader perspective allows us to assess the true operating costs and social implications of electricity generation from coal.

We consider not only state governor party affiliation, but also the prevailing party in the state legislature, providing a multifaceted view of the political climate. Additionally, the dataset includes state-level coal debt securitization policy status, shedding light on the financial mechanisms in play. In tandem with political considerations, we incorporate county-level public opinion data, capturing sentiments on coal plant favorability and attitudes towards regulating coal power emissions. This local-level perspective allows for a more granular understanding of how communities perceive and interact with coal plants. The change in public opinion since 2018 is encorporated to discern evolving sentiments and attitudes towards coal power within the counties these plants are located in.

While utility-level clean energy target information is excluded from the initial dataset processing due to missing data, it is considered in post-analysis. This is because the majority of this data is non-categorical and non-numerical, so imputation of missing data is avoided here to maintain the integrity of our results. That said, a binary metric of 'Utility Climate Target' distilled from this data was included in graph construction. Plant ownership information is included in the analysis as it represents an important piece of information to contextualize plant financials and age. While a non-negligable portion of plants were missing ownership type, the categorical nature of the data allowed for imputation while maintaining integrity in graph-construction, as well as without making assumptions where data did not exist.

A.1.2 Imputation Strategy and Distribution Preservation

Our imputation method involves sampling from the distribution of each feature to generate missing values. For continuous variables, we sample from a normal distribution, and for categorical variables, we use a uniform distribution. To ensure the imputation process doesn't alter the original data distribution, we measured the Earth Mover's Distance (EMD) between the original and imputed feature distributions. The results, presented in Supplemental Table 5, demonstrate that the imputation process preserves the statistical properties, with minimal EMD values indicating no significant distortion.

Representational Variability and Imputation Consistency. We employed Presto [1], a topology-based score, to assess the representational variability between multiple imputations. Additionally, we calculated average correlations (Pearson, Spearman, and Distance) between imputed column spaces, indicating high correlation and consistency. The analysis, shown in Supplemental Table 6, revealed extremely low representational variation between imputations, confirming the high consistency of imputed embeddings. This low variability supports our decision to use 10 imputations, as increasing the number beyond 10 did not significantly impact the final graph representation. Thus, 10 imputations provide a robust representation of the data without unnecessary computational burden.

The imputation strategy, coupled with the validation of its effectiveness, ensures that our analysis handles missing data accurately and reliably, preserving the integrity of the original data distribution.

A.2 Benchmarking Against Recent Retirements

Overview: Benchmarking our method against plants retired from 2020-2023 demonstrated strong performance: 90% (47/52) of analyzed plants aligned with the pre-defined retirement archetypes. This high degree of alignment, including the consistency of real-world shutdown drivers with group profiles, validates the robustness and practical relevance of our approach.

We tested the robustness of our group-based framework by analyzing coal plants that retired between 2020 and 2023. While our primary analysis focuses on plants with announced future retirements, this retrospective check assesses whether recently closed plants align with the characteristics of groups identified as more vulnerable, thus providing a grounded validation against real-world outcomes.

We incorporate the retired plants into the existing network structure (Figure 2) via a routine that we call *target matching*. This approach preserves the original graph topology, thus serving as a validation for the unsupervised analyses derived in the main text. The goal of target matching is to inject each new plant into the graph based on a set of standardized features using a random-walk-inspired routine based on vector similarities. In practice this computes the probability of each new plant belonging to a given group, giving an estimate of how well each plant matches the archetypes extracted in our original model. Full assignment outcomes and match classifications are reported in Supplemental Table 10, with explanations for the classification criteria detailed in Supplemental Table 9.

A.2.1 Target Matching Routine

To match a new plant into one of our existing groups, we simulate a series of "random walks" across the network of already-retired plants. At each step of a walk, the plant moves from one node to a neighboring one, with closer matches (based on plant characteristics) being more likely choices. Occasionally, if none of the local options are a good fit, the walk "jumps" directly to the best overall match with a 25% chance. We repeat this process many times, and in the end, the new plant is assigned to the group that it most frequently landed in during these walks. We compute transition probabilities from neighbor distances $\{d_i\}_{i=1}^{K}$ by first forming:

$$w_i = \begin{cases} \exp\left(-s\, d_i\right), & s > 0, \\ d_i, & s = 0, \end{cases} \qquad p_i = \frac{w_i}{\sum_{j=1}^{K} w_j} \qquad \left(\text{or } p_i = 1/K \text{ if } \sum_j w_j = 0\right).$$

Given a target feature vector $\mathbf{x}$ and node features $\{\mathbf{f}_i\}$, we set

$$D_i = \|\mathbf{x} - \mathbf{f}_i\|_2, \qquad \bar{G} = \frac{1}{|C|} \sum_{c \in C} D_c.$$

We then run n independent random walks of length m on graph $G = (V, E)$:

1. Initialize $v_0 \sim \text{Uniform}(V)$.
2. For $t = 1, \ldots, m$:
   - Let $N_t = \{v_{t-1}\} \cup \{j : (v_{t-1}, j) \in E\}$.
   - Compute $\{p_j\}$ from $\{D_j\}_{j \in N_t}$.
   - Sample $v_t \sim p_j$.

- If $\max_{j \in N_t} D_j > \bar{G}$ and $u < \frac{1}{4}$ with $u \sim U[0,1]$, jump to $v_t = \arg\min_{i \in V} D_i$.

3. Record v_m as one sample.

We ran the random walk algorithm with $n = 200$ samples and $m = 3000$ steps per walk for each coal plant in our dataset with $s = 2$.

A.2.2 Target Matching Data

Because our framework integrates static snapshots of political, demographic, public-opinion, and other contextual data drawn from multiple sources (see Supplemental Information A.1), we limit this extension of the analysis to coal plants that retired between 2020 and 2023. This time window reflects a balance between recency and data completeness: it captures the most recent wave of coal plant retirements while also ensuring that we can reliably link each plant to covariates measured close to its shutdown—though not every variable is present for every plant, depending on when it went offline and when its data were collected.

In this additional analysis, we use the same data handling, processing steps, and variables as described in the main methods and Supplemental Information A.1.

A.2.3 Target Matching Results

While most plants land in a single group nearly 100% of the time—indicating a strong and confident match—some are split across two or more groups (Supplemental Table 10). This means the model sees similarities with multiple group profiles, suggesting the plant sits near the boundary between archetypes or shares traits with more than one. These probabilities reflect both feature similarity and how the plant connects into the structure of the existing graph.

We assessed whether each plant's real-world retirement drivers aligned with those characterizing its assigned group. Overall, **62%** of plants matched exclusively to a single group, while the remaining **36%** had partial membership across multiple groups—the vast majority of these had over **90%** alignment with one group, indicating a clear primary fit. In total, **92%** met a bar of alignment (*Partial Match* or stronger), reflecting strong overall correspondence between plant-level and group-level retirement drivers.

A.3 Note on the Evolving Landscape of Coal Plant Retirements

Coal plant retirements have transitioned from being predominantly organic and market-driven to deliberate actions influenced by environmental priorities. Historically, retirements occurred as plants became financially unsustainable due to shifting market dynamics, aging infrastructure, or technological obsolescence. These retirements, while informative about operational and procedural challenges, were not tied to explicit environmental or climate goals.

Today, coal retirements are shaped by a different set of forces. The global push to address climate change and reduce greenhouse gas emissions has made coal phaseouts a strategic necessity. International climate agreements such as the Paris Agreement, coupled with heightened public awareness and stricter regulatory frameworks, now drive retirement decisions. This shift reflects a growing recognition of the outsized role coal-fired power plays in global carbon emissions and climate challenges. Several factors differentiate the current coal retirement landscape from the past:

1. *Climate Objectives as Central Drivers.* Modern retirements are heavily influenced by external factors such as climate agreements and emission reduction mandates, which were largely absent in earlier transitions.
2. *Complex Decision-Making.* Decisions to retire coal plants now involve multiple dimensions, including economic viability, technological advances, regulatory compliance, and social equity considerations.
3. *Stakeholder Involvement.* Modern coal retirements increasingly prioritize community impacts and equity concerns, elevating the importance of a just transition for workers and affected regions.
4. *New Economic Lifelines for Coal Plants.* Emerging factors, such as cryptocurrency operations, have introduced unexpected dynamics. By purchasing power through power purchase agreements (PPAs) or adding grid demand, cryptocurrency mining has temporarily extended the lifespans of struggling coal plants, underscoring the need to analyze retirement trends through a contemporary lens.

This evolving landscape underscores the timeliness of our dataset and methodological approach. By capturing the unique factors driving modern coal retirements, our work offers a comprehensive view of this critical energy transition. Unlike past trends, the current retirement context demands nuanced analysis that integrates environmental, economic, technological, and social considerations.

A.4 Supplementary Analysis

This section presents additional insights into the coal fleet derived from `THEMA`, offering analyses that were not included in the main text.

A.4.1 Detailed Vulnerability Descriptions in Groups 0, 3, and 4

This section provides additional details and insights to complement the analysis presented in the results section. It offers a deeper exploration of the factors influencing coal plant vulnerabilities and retirement pressures, further elucidating the trends and dynamics driving the outcomes discussed in the main text.

Group 0: Fuel Blend Plants

Plants in Group 0 struggle with aging infrastructure and regulatory pressures [2, 3] Table 1. Most plants are in Republican-majority states, yet fewer than half have Republican governors. Figure 5a highlights a division among plants nearing retirement (blue sink nodes), splitting them into two distinct sets of eight plants each with announced retirements (henceforth, Left Block and Middle Block). Non-retiring plants contained within sink nodes are considered to be in high proximity to retirement as they share the highest degree of similarity encoded within the graph structure.

Low Hanging Fruit Summary:

- **Target Non-Retiring Left Block Plants:** Focus advocacy on plants in the left graph section not scheduled for retirement, as their similarities to retiring plants make them ripe for retirement efforts.
- **Emphasize Environmental Justice:** Highlight the significant CO_2 emissions and adverse effects of aging plants in disadvantaged areas.

- **Leverage Compliance Violations:** Explore possibility of coal ash pond violations to expedite closures within Group 0.

Analysis. The Middle Block includes eight planned retirements (7.7 GW capacity, 19 MT CO_2 emissions annually) and five non-retiring plants (4.5 GW, 7.6 MT CO_2). Among the retirees, 25% (2/8) are closing due to coal ash pond non-compliance [4], and 37.5% (3/8) are being phased out due to clean energy mandates tied to aging infrastructure [5, 6]. The remaining plants are converting to natural gas [7]. Seven of eight planned retirements are in Republican states, where public support for coal remains high. Of plants in the Middle Block not planning to retire, 60% (3/5) also burn petroleum byproducts such as petroleum coke, with all exhibiting extremely low utilization rates (average capacity factor below 20%). Given this low usage, the retirement of these plants is expected to have a relatively minimal impact on overall coal reliance, but some may even serve as standby capacity, which reduces their retirement priority.

The Left Block comprises eight planned retirements (7.8 GW capacity, 25 MT CO_2) and five non-retiring plants (5.3 GW, 18 MT CO_2). All plants are in Republican-controlled states, predominantly situated in more disadvantaged communities with higher racial diversity, averaging 55% on the EPA's EJSCREEN demographic index (vs. plants in the Middle Block have demographic index scores consistently below 35%). Half (4/8) of the planned retirements are transitioning to natural gas due to regulatory compliance issues [7, 8]. Despite a median capacity factor under 40%, non-retiring plants emitted 18 MT CO_2 in 2022, indicating a significant environmental impact. The non-retiring plants score even higher (57.5%) than those with plans to retire (51.6%) on the EPA's EJSCREEN demographic index.

In contrast, the grey Low Proximity coldspot (far from retirement) on the right of Figure 5a highlights Pennsylvania coal plants with an 80% utilization rate. PJM's history of retaining coal plants due to resource adequacy issues [6] suggests that increased utilization makes it harder to retire these plants. Because of their similarity to the rest of the group, these younger (average age 30 years) and smaller (average size 160 MW) plants are likely subject to similar regulatory pressures due to inefficient generators, non-compliant coal ash disposal, and clean energy targets in the region.

Group 4: Expensive Plants

In Group 4, planned retirements arise from high forward costs [9, 10, 11, 12], regulatory updates [13, 14, 15], and utility/state-level climate targets. Planned retirement hotspots within Group 4 are divided into three sub-regions (Figure 5b) *Right Block*, *Middle Block*, and *Left Block*.

Low Hanging Fruit Summary:

- **Target High-Cost, High-Emission Left Cohort Plants:** Advocate for retiring plants in densely populated areas to emphasize health and economic benefits.
- **Expedite Middle Cohort Retirements:** Pressure regulatory bodies to accelerate retirements of non-compliant large plants through enhanced oversight.
- **Capitalize on Natural Gas Penetration:** Highlight reduced capacity value of the Far from Retirement plants, emphasizing how the higher penetration of natural gas in the energy mix of these plant's states/grid regions facilitates easier decommissioning.

Analysis. The Left Block has two planned retirements (1.6 GW capacity, 5.8 Mt CO_2) and one non-retiring plant (53 MW). 10 non-retiring plants (3.7 GW capacity, $\sim$16 Mt CO_2 annually) are situated in the orange nodes on the left of the graph, in *Mid Proximity* to retirement. All plants have low capacity factors (∼30%) and higher marginal costs, situated in densely populated areas (median of 25,000 people within 3 miles versus the group median of 2,000). The higher percentage of wind generation (24%) in these regions decreases coal capacity values. This higher wind generation helps to combat the heavy net-load placed on the grid in regions with a higher solar penetration [16], effectively reducing the capacity value of coal and making these plants easier to retire.

Five plants in the Middle Cohort (4.5 GW, 19.6 Mt CO_2) are retiring, while only two small plants remain. The retiring plants exemplify the group's profile, being costly and uneconomical without significant investment for compliance upgrades [14]. Non-retiring plants, located in remote states, represent a negligible portion of total capacity (less than 0.5 GW).

The Right Cohort includes three scheduled retirements (4.6 GW, 15 Mt CO_2) and two non-retiring plants (4 GW, 14.2 Mt CO_2). The two non-retiring plants, Kyger Creek in Ohio and John E. Amos in West Virginia, averaging over 1.7 GW, are over 50 years old and operate with marginal costs 20% higher than the average renewable generation costs in their regions. Despite their higher operational costs and substantial environmental impacts, these large plants pose a challenge for retirement due to their size and the political climate of the region.

The two light grey nodes at the bottom of the graph signify plants furthest from retirement, with an average age of 32 years. Although younger, they face similar cost and emissions pressures as the rest of the group, situated in natural gas-heavy states (46% of resource mix).

Group 3: High Health Impact Plants

In the case of Group 3, a majority of nodes are close to retirement (Figure 6a). To understand the specific drivers of retirement within this group, we use community label propagation [17] to categorize coal plants based on their retirement patterns. This method assigns labels to each plant, indicating their membership in distinct subsets. While Djikstra's algorithm helps gauge proximity to retirement, additional analysis is needed when most plants are near retirement. By identifying common retirement attributes within specific subgroups, we can classify their retirement drivers in more detail.

Within the scope of Group 3's high rates of hospitalization linked to plant pollution, SO_2, PM, and NO_x emissions, as well as other criteria air pollutant emissions, we determine differences between the subgroupings of plants present in the graph structure (see Figure 6). Subgroups 0 and 1 are significantly closer to retirement, with 35% and 38% of plants in high proximity, respectively.

Low Hanging Fruit Summary:

- **Target SubGroup 1 Non-Planned Retirement Plants:** Despite lower emissions, these plants contribute significantly to pollution and operate in disadvantaged communities, making their retirement crucial.
- **Highlight Health Impacts:** Regulatory and economic pressures, alongside pollution levels, facilitate retirement discussions for the majority of plants in proximity to retirement.
- **Enhanced Financial Instruments for SubGroup 2 Plants:** IPP-owned plants in SubGroup 2 may, regardless of their high emissions, present a signficant challange to retire early due to ownership structures.

Analysis. SubGroup 0 comprises the highest-polluting coal plants in Group 3, exhibiting PM emission rates 20% above the fleet average. This subgroup anticipates 7 planned retirements totaling 15.8 GW and 48.3 Mt CO2/yr, primarily due to Clean Air Act violations [18, 19, 20, 21] and cost competitiveness issues against renewables [22, 23]. Non-planned retirements have a higher median capacity factor of 60%, compared to 44% for retiring plants, and are located in states with Democratic governors, which may facilitate expedited retirements. SubGroup 1 features plants under economic pressure, with 4 out of 5 planned retirements in Texas, retiring mainly due to cost inefficiencies in modern energy markets [24, 25] and expensive environmental retrofits [26]. These plants have PM emissions approximately one-quarter of those in SubGroup 0, operating in disadvantaged communities (over 50% national percentile for people of color vs. 15% for SubGroups 0 and 2), making them favorable for early retirement. Subgroup 2, despite severe economic decline—evidenced by a median cash flow dropping to -$6.5 million in 2020 from an 8-year average of nearly $100 million—shows only one planned retirement plant with a 69% capacity factor, lacking other plants in proximity to retirement. Non-planned retirement plants in this subgroup are primarily IPP merchant coal plants, which are challenging to retire due to operators' motivations to recover losses [27]. Thus, this subgroup of high health impact plants requires innovative financing and incentives to prevent prolonged coal generation.

A.4.2 Detailed Vulnerability Descriptions in Groups 1, 2, 5, 6, and 7

This section provides an examination of plant groups *not addressed* in detail in the results section, emphasizing the specific reasons driving planned retirements and highlighting key factors that make other non-retiring plants vulnerable to similar pressures. The analysis spans five distinct plant groups, providing insights into regulatory, economic, and environmental drivers of retirement decisions.

Each group of coal plants faces unique vulnerabilities that make non-retiring plants susceptible to future closure. Whether driven by regulatory pressure, economic challenges, or public health concerns, the factors outlined below provide policymakers with clear pathways for accelerating the retirement of coal plants. By targeting specific economic and environmental levers, policymakers can align coal plant retirements with broader climate and public health goals, while also ensuring a just transition for workers and communities.

Group 1: Heavily Retrofitted but Economically Vulnerable Plants. Non-retiring plants in Group 1 face significant economic challenges due to the high costs of maintaining retrofitted technology, combined with increasing competition from cheaper renewable energy sources. These factors make them likely candidates for market-based retirement incentives such as carbon pricing or financial support for transitioning to cleaner energy. Additionally, many of these plants are located near disadvantaged communities, where the environmental and health impacts of continued coal operation are particularly severe. This provides an opportunity for policymakers to implement "just transition" policies, which could include retraining programs and investments in renewable energy infrastructure to replace the economic role of coal plants in these areas. Moreover, as state and federal carbon reduction targets become stricter, the regulatory pressure on these plants will only increase, eventually making continued operations financially untenable.

Group 2: Democratic Plants. The planned retirements in this group are largely influenced by state-level emissions standards, clean energy mandates, and utility carbon reduction goals. For example, Washington's TransAlta plant is closing by 2025 due to state mandates, while Xcel Energy's Hayden and Craig plants in Colorado are retiring in response to the utility's 100% carbon-free target by 2050. Environmental groups have also played a significant role in advancing plant closures, such as in the case of the Baldwin plant. Non-retiring plants in Group 2 are located in states with little political and public support for coal power, making them highly susceptible to future regulatory pressure. As state-level emissions reduction mandates and renewable energy targets (such as Renewable Portfolio Standards) become more aggressive, these plants will face increasing difficulties in maintaining compliance. The regulatory environment in these Democratic-majority states favors clean energy transitions, creating a policy framework that is likely to accelerate the retirement of remaining coal plants. In addition to regulatory pressure, public opinion and advocacy from environmental groups play a crucial role in pushing for early retirement timelines. In regions with high emissions, non-retiring plants may face growing scrutiny from both regulators and the public, further driving the momentum for closure.

Group 5: Young Plants. Although Group 5 plants have no planned retirements due to the relatively young age (average of 32 years) and high profitability, they are not immune to future retirement pressures. While they currently face less immediate pressure, policymakers can take proactive steps by offering financial incentives for transitioning to renewable energy, which may prompt plant owners to consider early closure. Additionally, many of these plants are located in grids that are increasingly reliant on wind energy, which will further diminish the role of coal in the energy mix.

Group 6: Retiring Plants in Anti-Coal Regions. Retirements in this group are driven by strong public opposition to coal, utility carbon reduction commitments, and state climate targets. For example, the Erickson plant in Michigan is retiring due to public pressure and the Lansing utility's net-zero target. The Sherco plant in Minnesota is closing as part of Xcel Energy's transition to 100% carbon-free electricity by 2050. Non-retiring plants in Group 6 operate in regions where public opposition to coal is strong, creating a political environment that favors clean energy transitions. This makes these plants particularly vulnerable to future regulatory action, as policymakers are under increasing pressure to implement stricter emissions standards and accelerate plant closures. In addition to public opposition, these plants face increasing economic difficulties as they compete with cheaper natural gas and renewable energy sources. The combination of public pressure and the economic challenges of maintaining coal operations creates a clear pathway for future retirements. Policymakers can support this transition by offering economic incentives for early closure, as well as programs to support displaced workers and local economies.

Group 7: Air Quality Offenders. Despite high emissions (SO2, NOx, Mercury), no planned retirements have been announced in this group. These plants continue to operate, benefiting from strong pro-coal political support, but they remain vulnerable to future regulatory action due to their significant environmental impact. Stricter enforcement of EPA standards, as well as state-level emissions regulations, is likely to increase the cost of compliance for these plants, making continued operation financially unsustainable. Additionally, these plants contribute significantly to poor air quality and public health problems, making them targets for environmental justice campaigns and public health advocacy. While they benefit from strong pro-coal sentiment in the regions where they operate, the growing evidence of health risks associated with their emissions may lead to political shifts, especially as federal emissions penalties increase.

A.5 Our Algorithm: `THEMA`

The custom Topological Hyperparameter Evaluation and Mapping Algorithm (`THEMA`) is a cornerstone of our analysis pipeline for studying coal plant retirements. `THEMA` is an open-source Python package developed by Krv Analytics, freely available for use and modification. To access the THEMA code and documentation, visit our GitHub repository: https://github.com/Krv-Analytics/Thema. The repository includes the latest version of the code, installation instructions, user guides, etc.

Purpose and Unique Capabilities. THEMA was specifically designed to address challenges in analyzing high-dimensional, complex datasets, such as those encountered in coal plant retirement studies. In these contexts, traditional methods often struggle to tune parameters and identify meaningful data representations that lead to actionable insights. THEMA leverages advanced topological data analysis (TDA) techniques and hyperparameter optimization to uncover hidden patterns and relationships in multifaceted datasets. This capability is particularly relevant to coal plant retirement research, where the interplay of economic, environmental, technological, and social factors creates a high-dimensional decision landscape. By using TDA, THEMA maps complex relationships into lower-dimensional representations without losing critical information. This makes it easier to identify actionable insights and prioritize interventions that align with policy goals, stakeholder interests, and sustainability objectives.

Broader Applications. THEMA is not limited to coal plant retirements. Its versatility makes it valuable for researchers and practitioners tackling problems in energy and climate analysis, where complexity and dimensionality pose significant challenges. By integrating THEMA into their workflows, governments, policymakers, and analysts can better navigate complex datasets and develop informed strategies to address pressing energy transitions and climate goals. We encourage the community to explore THEMA and adapt it to their specific needs, fostering collaboration and innovation in data-driven decision-making for sustainable futures.

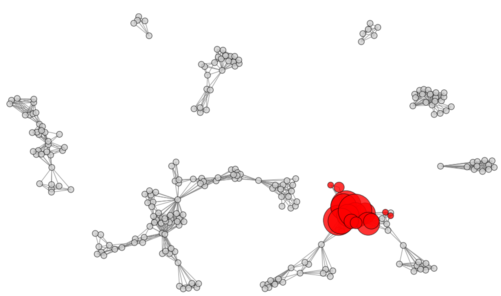
Algorithm 1 `THEMA`: Topological Hyperparameter Evaluation and Mapping Algorithm.

Require: X: original dataset
Require: Ψ: a set of embedding algorithms for X, with hyperparameters α for each $\psi \in \Psi$: $\psi_\alpha(X) \subset \mathbb{R}^D$.
Require: $\mathcal{R}$: a set of dimensionality reduction algorithms, with hyperparameters β for each $R \in \mathcal{R}$: $R_\beta \circ \psi_\alpha(X) \subset \mathbb{R}^d$, where $d \ll D$.
Require: $\mathcal{C}$: a set of graph construction algorithms, with hyperparameters γ for each $C \in \mathcal{C}$;
each $C_\gamma(E)$ constructs a graph from embedding $E \subset \mathbb{R}^d$.
Require: $\mathcal{F}$: a set of boolean filters, where each $F \in \mathcal{F}$ is an indicator function on graphs.
Require: dist_G: a semi-metric function over graphs, such that $\mathrm{dist}_G(G_1, G_2) \in \mathbb{R}_{\geq 0}$.
Require: $\mathcal{O}$: a routine for total ordering (sorting) an input set of graphs.
Require: $\mathcal{P}$: a partitioning method on pairwise distance matrix $D \in \mathbb{R}^{N \times N}$, outputting a partition $\{P_1, \ldots, P_k\}$.
Require: $\vec{n_\Psi} \in \mathbb{N}^{|\Psi|}$, $\vec{n_R} \in \mathbb{N}^{|\mathcal{R}|}$, $\vec{n_C} \in \mathbb{N}^{|\mathcal{C}|}$: sampling sizes for each $\psi \in \Psi$, $R \in \mathcal{R}$, and $C \in \mathcal{C}$.

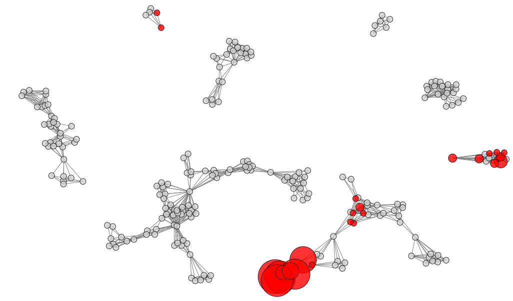
```
function THEMA(X, Ψ, R, C, F, dist_G, O, P, n⃗_Ψ, n⃗_R, n⃗_C)
    E ← ∅                                              ▷ Initialize set of low-dimensional embeddings.
    G ← ∅                                              ▷ Initialize set of graphs.
    M ← ∅                                              ▷ Initialize set of final selected models.

                                                       ▷ Generate Embeddings.
    for ψ ∈ Ψ do
        H_ψ ← hyperparameter space for ψ               ▷ Realize ψ's hyperparam. combos as a vector space.
        for j = 1 to n^i_ψ do                          ▷ n_ψ ∈ n⃗_Ψ is number of samples from H_ψ.
            α_j ← random sample from H_ψ
            V ← ψ_{α_j}(X)                             ▷ Embed dataset X into R^D.
            for R ∈ R do
                H_R ← hyperparameter space for R       ▷ Realize R's hyperparam. combos as vector space.
                for l = 1 to n_R do                    ▷ n_R ∈ n⃗_R is number of samples from H_R.
                    β_l ← random sample from H_R
                    E ← R_{β_l}(V)                     ▷ Map to low dim. representation.
                    AddItem(E, E)
                end for
            end for
        end for
    end for
                                                       ▷ Construct Graph Models.
    for C ∈ C do
        H_C ← hyperparameter space for C               ▷ Realize C's hyperparam. combos as vector space.
        for q = 1 to n_C do                            ▷ n_C ∈ n⃗_C is number of samples from H_C.
            γ_q ← random sample from H_C
            for E ∈ E. do
                G ← C_{γ_q}(E)                         ▷ Build graph representation of E.
                AddItem(G, G)
            end for
        end for
    end for
                                                       ▷ Select Pertinent Models.

    G_F ← {G ∈ G : F(G) = 1, ∀F ∈ F}                   ▷ Apply graph filters.
    N ← |G_F|                                          ▷ Number of filtered graphs.
    D ← 0 ∈ R^{N×N}                                    ▷ Initialize distance matrix.
    for i = 1 to N do
        for j = 1 to N do
            D[i, j] ← dist_G(G^i_F, G^j_F)             ▷ Compute pairwise distances for G_F.
        end for
    end for
    for P ∈ P(D) do
        P_sorted ← O(P)                                ▷ Sort partitioned graphs.
        G_top ← Top element in P_sorted                ▷ Select top graph in partition.
        AddItem(M, G_top)
    end for
    return M                                           ▷ Return selection.
end function
```



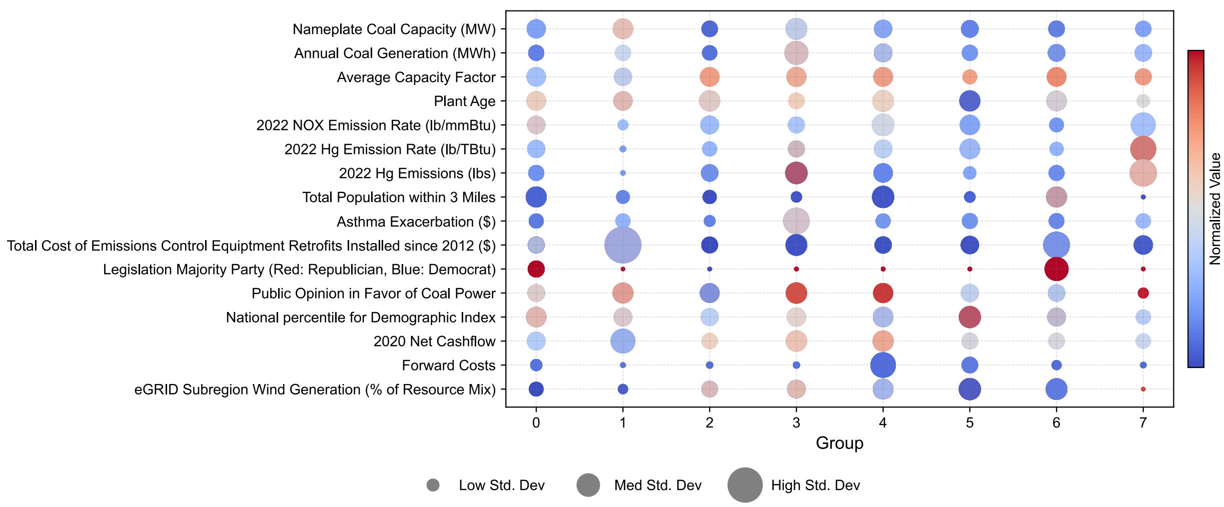
(a) Target Matching for Morgantown Generating Coal Plant: 100% Match into Group 0



(b) Target Matching for Asheville Coal Plant: 92.5% Match into Group 0, 5% into Group 5, 2.5% into Group 1

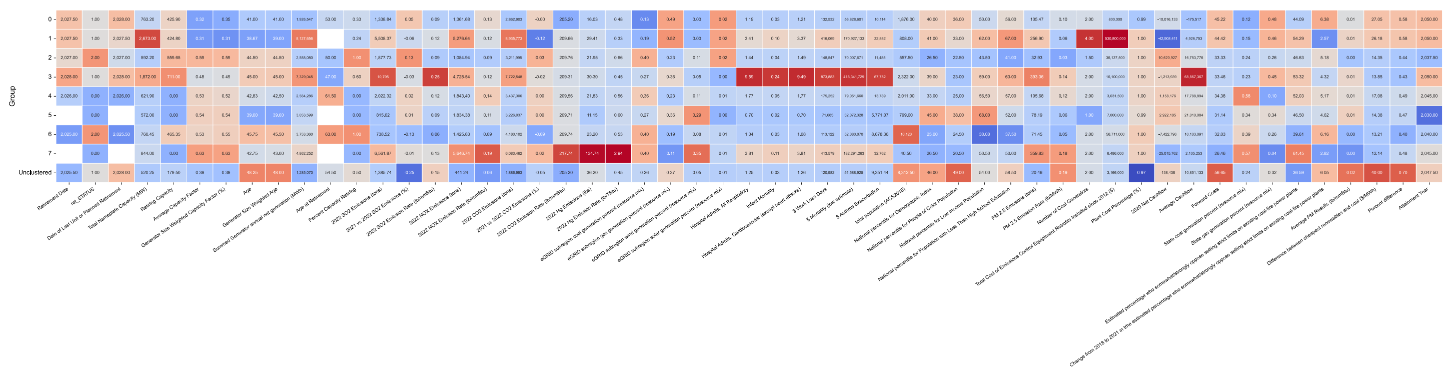
Supplemental Figure 1: Target Matching Example

Visualizing the target matching process for adding a single coal plant on the network. Red nodes/size show the location and frequency of where random walks end, indicating network areas most connected to the target plant and informing its group assignment. We performed the ‘target matching’ process 200 separate times for *each* plant, where each of those 200 walks takes 3000 steps away from the starting node.



Supplemental Figure 2: Median Group Characteristics Dotplot

The dotplot compares median group characteristics. Each dot represents a feature within a group, where the color gradient reflects the normalized value of the feature (as indicated by the color bar on the right). Dot size is proportional to the standard deviation of the feature within each group, illustrating variability.



Supplemental Figure 3: Detailed Group Characteristics Heatmap

This heatmap displays the group's median value for every column listed.

Note that the unclustered group has large standard deviations in most columns, as they are *not* similar. This has been included as a reminder that a portion of coal capacity remains 'Unclustered' and outside the scope of this analysis.

B Supplementary Figures

C Supplementary Tables

Variable	Source	Used in Graph Construction
EIA Plant Code	EIA eGrid Dataset - 2021	
Plant Name	EIA eGrid Dataset - 2021	
Retirement Date	Sierra Club Retired and Announced Coal	
EIA Utility Code	EIA eGrid Dataset - 2021	
Retirement Status	Sierra Club Retired and Announced Coal	X
Date of Last Unit or Planned Retirement	Sierra Club Retired and Announced Coal	
Total Nameplate Capacity (MW)	EIA eGrid Dataset - 2021	X
Retiring Capacity	EIA eGrid Dataset - 2021 & Sierra Club Retired and Announced Coal	X
Average Capacity Factor	EIA eGrid Dataset - 2021	X
Generator Size Weighted Capacity Factor (%)	EIA eGrid Dataset - 2021	X
Age	EIA eGrid Dataset - 2021	X
Generator Size Weighted Age	EIA eGrid Dataset - 2021	X
Summed Generator annual net generation (MWh)	EIA eGrid Dataset - 2021	X
Age at Retirement	Sierra Club Retired and Announced Coal	
Percent Capacity Retiring	EIA eGrid Dataset - 2021 & Sierra Club Retired and Announced Coal	X
State	EIA eGrid Dataset - 2021	
2022 SO2 Emissions (tons)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
2021 vs 2022 SO2 Emissions (%)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
2022 SO2 Emission Rate (lb/mmBtu)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
2022 NOX Emissions (tons)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
2022 NOX Emission Rate (lb/mmBtu)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
2022 CO2 Emissions (tons)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
2021 vs 2022 CO2 Emissions (%)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
2022 CO2 Emission Rate (lb/mmBtu)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
2022 Hg Emissions (lbs)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
2022 Hg Emission Rate (lb/TBtu)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
Facility has one or more low-emitting EGUs (LEE) units that do not report hourly emissions	EPA MATS Data Consolidated - 2021	X
eGRID subregion name	EIA eGrid Dataset - 2021	
eGRID subregion coal generation percent (resource mix)	EIA eGrid Dataset - 2021	X
eGRID subregion gas generation percent (resource mix)	EIA eGrid Dataset - 2021	X
eGRID subregion wind generation percent (resource mix)	EIA eGrid Dataset - 2021	X
eGRID subregion solar generation percent (resource mix)	EIA eGrid Dataset - 2021	X
Hospital Admits, All Respiratory	Coal Plant Health Impacts 2019 - COBRA Pollution Transport Model Results, Provided by the Clean Air Task Force	X
Infant Mortality	Coal Plant Health Impacts 2019 - COBRA Pollution Transport Model Results, Provided by the Clean Air Task Force	X
Hospital Admits, Cardiovascular (except heart attacks)	Coal Plant Health Impacts 2019 - COBRA Pollution Transport Model Results, Provided by the Clean Air Task Force	X
$ Work Loss Days	Coal Plant Health Impacts 2019 - COBRA Pollution Transport Model Results, Provided by the Clean Air Task Force	X
$ Mortality (low estimate)	Coal Plant Health Impacts 2019 - COBRA Pollution Transport Model Results, Provided by the Clean Air Task Force	X
$ Asthma Exacerbation	Coal Plant Health Impacts 2019 - COBRA Pollution Transport Model Results, Provided by the Clean Air Task Force	X
total population (ACS2018)	EPA Power Plants and Neighboring Communities	X
National percentile for Demographic Index (within 3 miles of plant)	EPA Power Plants and Neighboring Communities	X
National percentile for People of Color Population (within 3 miles of plant)	EPA Power Plants and Neighboring Communities	X
National percentile for Low Income Population (within 3 miles of plant)	EPA Power Plants and Neighboring Communities	X
National percentile for Population with Less Than High School Education (within 3 miles of plant)	EPA Power Plants and Neighboring Communities	X
PM 2.5 Emissions (tons)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
PM 2.5 Emission Rate (lb/MWh)	EPA's Clean Air Markets Division 2021 and 2022 Annual Power Plant Emissions	X
Average PM Results (lb/mmBtu)	EPA MATS Data Consolidated - 2021	X
Mapped Fuel Type	EPA Power Plants and Neighboring Communities	X
Coal Fuels	EIA eGrid Dataset - 2021	X
Number of Coal Generators	EIA eGrid Dataset - 2021	X
Non-Coal FUELS	EIA eGrid Dataset - 2021	X
Total Cost of Emissions Control Equipment Retrofits Installed since 2012 ($)	EIA Form 860 Data, Schedule 6A, Emissions Control Equipment	X
Levelized Cost of Emissions Control Equipment Retrofits Installed since 2012 ($/MW)	Calculated	X
Plant Coal Percentage (%)	EIA eGrid Dataset - 2021	X
2020 Net Cashflow	RMI Uneconomic Dispatch Dataset	X
Average Cashflow (since 2012)	RMI Uneconomic Dispatch Dataset	X
Forward Costs	Energy Innovation Coal Cost Crossover 3.0 Dataset	X
Coal Debt Securitization Policy	RMI State Utility Policies Data	X
State coal generation percent (resource mix)	EIA eGrid Dataset - 2021	X
State gas generation percent (resource mix)	EIA eGrid Dataset - 2021	X
LAT	EIA eGrid Dataset - 2021	X
LON	EIA eGrid Dataset - 2021	X
Governor Party	RMI State Utility Policies	X
Legislation Majority Party	RMI State Utility Policies	X
Estimated percentage who somewhat/strongly oppose setting strict limits on existing coal-fire power plants	Yale Climate Opinion Map Aggregated Dataset - 2008-2021	X
Change from 2018 to 2021 in the estimated percentage who somewhat/strongly oppose setting strict limits on existing coal-fire power plants	Yale Climate Opinion Map Aggregated Dataset - 2008-2021	X
Cheapest Renewable Resource	Energy Innovation Coal Cost Crossover 3.0 Dataset	X
Difference between cheapest renewables and coal ($/MWh)	Energy Innovation Coal Cost Crossover 3.0 Dataset	X
Renewables or Coal	Energy Innovation Coal Cost Crossover 3.0 Dataset	
Percent difference between cheapest renewables and coal (%)	Energy Innovation Coal Cost Crossover 3.0 Dataset	X
Sector	EIA eGrid Dataset - 2021	
Utility Name	EIA eGrid Dataset - 2021	
Ownership Type	RMI Coal Asset Valuation Tool & SEPA Utility Carbon-Reduction Tracker Dataset	X
Final Target Summary	SEPA Utility Carbon-Reduction Tracker	
Attainment Year	SEPA Utility Carbon-Reduction Tracker	
Interim Target(s) Summary	SEPA Utility Carbon-Reduction Tracker	
Mandatory or Voluntary Target	SEPA Utility Carbon-Reduction Tracker	
Target Scope	SEPA Utility Carbon-Reduction Tracker	
Utility Climate Target (yes or no)	Calculated from Utility Target Data	X
FIPS Code	Calculated from Lat/Lon	

Supplemental Table 1: Dataset & Data Sources Breakdown

Variable	Aggregation Method	Comments
Retirement Date	First occurrence (earliest retirement date)	11.1% of plants (23.2% of plants with retirement plans) only have partial retirement plans, reflected in the row below
Date of Last Unit or Planned Retirement	Maximum value (latest retirement date)	
Retirement Status	Maximum value	0: No Retirement Plans, 1: Planned Partial Retirement (e.x. 2 of 3 coal generators retiring), 2: Full Retirement Planned
Total Nameplate Capacity (MW)	Sum of nameplate capacities of all coal generators at plant	
Retiring Capacity	Sum of generator retiring capacities (MW)	
Capacity Factor	Mean (average) capacity factor across generators	
Generator Size Weighted Capacity Factor (%)	Capacity-weighted average capacity factor	
Age	Mean (average) coal generator age	
Generator Size Weighted Age	Capacity-weighted average age	
Summed Generator annual net generation (MWh)	Sum of generator annual net generation	
Age at Retirement	Maximum age at retirement (oldest unit retiring)	
Percent Capacity Retiring	Percentage of plant coal capacity retiring	Sum of retiring coal generators at plant (MW) ÷ Sum of all coal generators at plant (MW)
Total Cost of Emissions Control Equipment Retrofits Installed since 2012 ($)	Sum of boiler-level costs	Includes control technologies for: particulate matter (PM), sulfur dioxide (SO_2), nitrogen oxides (NO_x), mercury, and acid gases.

Supplemental Table 2: Aggregation of Generator-Level Variables for Plant-Level Dataset

Variable	Plants Missing Data (count)	Capacity (MW) Missing (%)
2022 SO2 Emissions (tons)	21	4.77%
2021 vs 2022 SO2 Emissions (%)	21	4.77%
2022 SO2 Emission Rate (lb/mmBtu)	21	4.77%
2022 NOX Emissions (tons)	21	4.77%
2022 NOX Emission Rate (lb/mmBtu)	21	4.77%
2022 CO2 Emissions (tons)	22	4.82%
2021 vs 2022 CO2 Emissions (%)	22	4.82%
2022 CO2 Emission Rate (lb/mmBtu)	22	4.82%
2022 Hg Emissions (lbs)	52	11.49%
2022 Hg Emission Rate (lb/TBtu)	52	11.49%
Hospital Admits, All Respiratory	16	2.23%
Infant Mortality	16	2.23%
Hospital Admits, Cardiovascular	16	2.23%
$ Work Loss Days	16	2.23%
$ Mortality (low estimate)	16	2.23%
$ Asthma Exacerbation	16	2.23%
PM 2.5 Emssions (tons)	1	0.46%
PM 2.5 Emission Rate (lb/MWh)	1	0.02%
Average PM Results (lb/mmBtu)	45	15.57%
2020 Net Cashflow	28	5.69%
Average Cashflow	28	5.69%
Forward Costs	15	2.46%
Cheapest Renewable Resource	15	2.46%
Difference between cheapest renewables and coal...	15	2.46%
Ownership Type	52	25.08%

Supplemental Table 3: Random-Sample Imputed Values Breakdown

Plant-level sector	Number of Plants	Coal Capacity (% of total Nameplate)
Commercial CHP	4	0.04%
Electric Utility	154	77.59%
IPP CHP	10	0.96%
IPP Non-CHP	44	20.68%
Industrial CHP	29	0.72%

Supplemental Table 4: Distribution of Coal Plants by Sector

Column	Normal Raw	Normal Clean	KS Statistic	KS P-Value	MW Statistic	MW P-Value	Comparative EMD	Same Distribution	Relative EMD Mean	Relative EMD Std	Z-Score Comparative EMD vs. Relative EMDs
$ Asthma Exacerbation	False	False	0.005495	1.0	16623.5	0.951541	1.628420e-16	False	0.194959	0.071196	-2.738362
$ Mortality (low estimate)	False	False	0.005495	1.0	16644.0	0.935287	7.114034e-17	False	0.174316	0.101231	-1.721958
$ Work Loss Days	False	False	0.005495	1.0	16516.5	0.964241	2.244656e-17	False	0.228390	0.170088	-1.342780
2020 Net Cashflow	False	False	0.000000	1.0	14450.0	1.000000	0.000000e+00	False	0.374198	0.092317	-4.053421
2021 vs 2022 CO2 Emissions (%)	False	False	0.005682	1.0	15484.0	0.997074	6.406649e-20	False	0.322982	0.041298	-7.820745
2021 vs 2022 SO2 Emissions (%)	False	False	0.011299	1.0	15675.0	0.991712	1.048638e-16	False	0.286232	0.069565	-4.114585
2022 CO2 Emission Rate (lb/mmBtu)	False	False	0.011364	1.0	15400.0	0.926965	5.604832e-16	False	0.567940	0.041616	-13.646987
2022 CO2 Emissions (tons)	False	False	0.005682	1.0	15422.0	0.945294	6.649609e-17	False	0.139444	NaN	NaN
2022 Hg Emission Rate (lb/TBtu)	False	False	0.006849	1.0	10726.0	0.925456	1.977109e-16	False	0.262883	0.118284	-2.222470
2022 Hg Emissions (lbs)	False	False	0.006849	1.0	10692.0	0.962963	9.093820e-17	False	0.296178	0.119888	-2.470444
2022 NOX Emission Rate (lb/mmBtu)	False	False	0.005650	1.0	15752.5	0.927581	2.481929e-16	False	0.157037	0.070810	-2.217724
2022 NOX Emissions (tons)	False	False	0.005650	1.0	15579.0	0.929644	1.453909e-16	False	0.193151	0.087582	-2.205381
2022 SO2 Emission Rate (lb/mmBtu)	False	False	0.005650	1.0	15751.5	0.928406	1.459961e-16	False	0.259629	0.140161	-1.852361
2022 SO2 Emissions (tons)	False	False	0.005650	1.0	15622.5	0.965616	2.519759e-17	False	0.167600	0.071663	-2.338723
Average Cashflow	False	False	0.005882	1.0	14430.5	0.983273	5.755200e-18	False	0.186043	0.020138	-9.238616
Average PM Results (lb/mmBtu)	False	False	0.013072	1.0	11665.0	0.959719	7.478781e-17	False	0.188646	0.072879	-2.588481
Difference between cheapest renewables and coal...	False	False	0.005464	1.0	16657.0	0.931494	4.644858e-17	False	0.567693	0.086285	-6.579262
Forward Costs	False	False	0.005464	1.0	16708.5	0.972017	4.565735e-17	False	0.520293	0.173745	-2.994572
Hospital Admits, All Respiratory	False	False	0.005495	1.0	16558.5	0.997615	8.922865e-17	False	0.242167	0.164818	-1.469297
Hospital Admits, Cardiovascular (except heart a...	False	False	0.005495	1.0	16616.5	0.957096	1.500250e-16	False	0.210366	0.171851	-1.224114
Infant Mortality	False	False	0.016484	1.0	16626.5	0.949161	1.000039e-16	False	0.220232	0.159612	-1.379795
PM 2.5 Emssions (tons)	False	False	0.005076	1.0	19419.5	0.989764	4.024206e-18	False	0.221425	0.002720	-81.407361

Supplemental Table 5: Imputation Effects on Original Distribution

Metric	Mean Value	Standard Deviation	Interpretation
Average Pearson Correlation	0.9737	-	Average correlation between column distributions of different imputations
Average Spearman Correlation	0.9610	-	Average rank correlation between column distributions of different imputations
Average Distance Correlation (dcor)	0.9705	-	Average distance correlation between column distributions of different imputations
Presto Representational Variability	0.0092	-	Measures the representational variance between imputations
Average Distance Between Imputations	0.0024	0.0028	Distance between imputations across different features
Average Presto Distance Between Raw and Imputed	0.0944	0.0013	Larger difference between raw dataset and imputed versions compared to between imputations

Supplemental Table 6: Summary of correlation and representational variability results.

Pearson, Spearman, and Distance Correlations measure the average correlation between column distributions of different imputations. Presto represents the variance between imputations, while the Presto distance reflects the larger difference between the raw dataset and the imputed versions compared to between imputations.

Category	Number of Plants
Group 0	9
Group 3	1
Group 6	1
Group 4	1
Unclustered	1

Supplemental Table 7: Plants that Burn Less Than 50% Coal

13 (6.53%) of plants burn under 50% coal, and 93.47% burn over 50% coal. Note that of these 13 plants, only **6** are *not* classified as coal plants based on the EIA's "Mapped Fuel Type" in the PPNC Dataset.

Group Mapped Fuel Type	0	1	2	3	4	5	6	7	Unclustered
Coal	36	3	14	29	74	11	13	6	6
Gas	1	0	0	0	0	0	1	0	2
Oil	2	0	0	0	0	0	0	0	0

Supplemental Table 8: Plants "Mapped Fuel Type" Classification from EIA PPNC Data

Plant Name	**Group Match** (Group Num: Match Score)	Match to Archetype?	Match Explaation
Dolet Hills	4: 0.94, 5: 0.06	Perfect Match	Plant's high operating costs and utility cost reduction efforts directly match Group 4's primary driver of high operating costs.
Martin Drake / / South Plant	7: 1.0	Close Match	Plant is retiring due to aging and transition to cleaner energy, but were previously in violation of federal air quality regulations and, as such, match the 'Air Quality Offenders' Group 7 archetype.
E D Edwards	6: 1.0	Partial Match	Plant closed due to legal settlement from environmental violations. This partially aligns with Group 6's 'Political Opposition' (manifested as regulatory and legal action) and indirectly with 'State-driven climate goals' which often underpin such environmental standards.
Lowman Energy Center	0: 1.0	Perfect Match	Plant's aging infrastructure and coal ash pond violation directly match Group 0's primary drivers. The fuel blend note reinforces this.
Waukegan	4: 1.0	Perfect Match	Plant's poor financial results and economic pressures directly align with Group 4's high operating costs, and the utility transition matches secondary drivers.
Escalante	4: 1.0	Perfect Match	Plant's economic in-viability and competition from renewables align perfectly with Group 4's archetype of high operating costs making plants expensive and uncompetitive.
Will County	5: 0.07, 6: 0.93	Perfect Match	Plant's retirement driven by state climate goals, renewable mandates, and economic difficulties directly aligns with Group 6's primary and secondary drivers.
Oklaunion	3: 0.95, 5: 0.05	Partial Match	Plant drivers are economic in-viability and aging/retrofit costs. Group 3 focuses on public health and environmental regulations. While high retrofit costs match the environment regulations focus and secondary economic driver of group 3, the primary stated drivers for the plant do not strongly align with Group 3's primary focus on health impacts or direct regulatory pressure.
Joppa Steam	2: 0.915, 6: 0.085	Partial Match	Plant drivers are economic and aging. Group 2's primary focus is political/regulatory pressure. While economic issues can be a secondary consequence in Group 2, the plant's stated reasons don't highlight the primary political/regulatory drivers of the archetype.
Bridgeport Station	0: 1.0	Perfect Match	Utility net zero target and replacement with natural gas are both aligned with Group 0's archetypes.
Crawfordsville Power Plant	6: 1.0	Partial Match	Plant's retirement is due to aging infrastructure. While Group 6 includes 'aging coal plants no longer cost-competitive' as a secondary driver, the primary drivers are political opposition and broader economic difficulties, which are not explicitly stated for the plant. Aging can contribute, but isn't the core of Group 6.
C D McIntosh Jr	0: 0.98, 1: 0.02	Close Match	Plant's aging infrastructure and increasing operational costs align well with one of Group 0's primary drivers. The fuel blend note is also consistent. Lacks explicit mention of regulatory non-compliance for a perfect match.
Lansing	2: 0.095, 5: 0.07, 4: 0.835	Close Match	Plant's retirement due to utility's clean energy transition plan and emissions reduction aligns well with Group 4's secondary drivers. While not explicitly stating 'high operating costs' as the primary reason, these transition plans often target such plants.
R Gallagher	2: 0.085, 4: 0.915	Close Match	Plant's retirement due to an 'Economic Decision' strongly suggests alignment with Group 4's core theme of high costs, making it expensive to operate. Lacks specific details for a perfect match.
Elmer Smith	4: 0.95, 5: 0.05	Perfect Match	The explicit statements that the retirement is 'purely economical' and due to 'faltering economics' perfectly match Group 4's primary driver of high operating costs.
Morgantown Generating Plant	0: 1.0	Perfect Match	Plant drivers of aging infrastructure, regulatory non-compliance, and coal ash pond violation are a perfect match for Group 0's primary and secondary drivers.
Paradise	4: 1.0	Perfect Match	Plant's retirement due to environmental regulations, economic factors, and emissions reduction aligns perfectly with Group 4's primary (high operating/retrofit costs due to regulations) and secondary drivers (utility/state targets).
Dan E Karn	0: 1.0	Close Match	Plant's aging infrastructure and alignment with renewable targets match Group 0's primary driver of aging and secondary drivers related to clean energy transitions. The fuel blend note is consistent. Lacks explicit mention of regulatory non-compliance for a perfect match.
Kenneth C Coleman	3: 1.0	Weak Match	Plant drivers are aging and economic factors. Group 3's primary focus is public health and environmental regulations. While economic difficulties can be a secondary factor in Group 3 if linked to compliance, the primary drivers stated for the plant do not align with Group 3's main archetype.
St Clair	2: 1.0	Perfect Match	Plant's retirement due to failure to meet air quality standards (regulatory pressure), state climate goals, and renewable mandates directly aligns with Group 2's primary and secondary drivers.
Chalk Point LLC	0: 1.0	Perfect Match	Plant's aging infrastructure and regulatory non-compliance (evidenced by water pollution violations) perfectly match Group 0's primary drivers.
Trenton Channel	0: 1.0	Partial Match	Plant's 'Utility Net Zero' goal aligns with Group 0 secondary drivers.
Dickerson	2: 0.905, 0: 0.095	Close Match	Plant's 'high compliance costs' strongly suggest retirement due to regulatory pressure making it uneconomic, aligning well with Group 2's primary theme of political/regulatory pressure and its secondary driver of becoming uneconomic due to standards.
Meramec	6: 0.13, 4: 0.78, 2: 0.09	Perfect Match	Plant's retirement due to environmental regulations, economic factors, and emissions reduction aligns perfectly with Group 4's primary (high operating/retrofit costs due to regulations) and secondary drivers (utility/state targets).
River Rouge	4: 0.96, 5: 0.04	Close Match	Plant's retirement due to utility's clean energy transition plan and emissions reduction aligns well with Group 4's secondary drivers. These transition plans often target expensive plants, fitting the Group 4 archetype.
San Juan	0: 1.0	Close Match	Plant's coal ash pond violation (regulatory non-compliance) and environmental justice concerns align very well with Group 0's primary and secondary drivers. Lacks explicit 'aging infrastructure' for a perfect score.
J B Sims	2: 1.0	Weak Match	Plant's driver is aging infrastructure and economic invibility. Group 2's primary focus is political/regulatory pressure. However, while this plant was not specifically retired for environmental reasons, it is in Michigan and its retirement furthered GHG reduction goals and helps to meet RPS standards. Therefore, it is a weak match but still meets Group 2's driver of: strong state-level emissions or cleanenergy mandates
Dunkirk Generating Plant	0: 1.0	Close Match	Plant's aging infrastructure and increasing operational costs align with a primary driver of Group 0. The attempted (but costly) NG conversion is also thematically relevant to 'Fuel Blend Plants'. Lacks explicit regulatory non-compliance.
Eckert Station	3: 1.0	Partial Match	Plant's retirement is due to a utility's clean energy transition and emissions reduction. Group 3's core is public health concerns and environmental regulations. While emissions reduction benefits health, the plant's stated drivers don't specifically highlight existing high health impacts or direct regulatory pressure due to health as the *primary* cause.
R M Heskett	6: 1.0	Close Match	Plant's aging infrastructure, economic factors, and high maintenance costs align well with Group 6's 'Economic Difficulties' primary driver and its 'aging plants no longer cost-competitive' secondary driver. Lacks explicit mention of 'political opposition.'
Hoot Lake	4: 0.97, 5: 0.03	Close Match	The plant's extreme age (100 years) strongly implies it would face high operating costs and significant retrofit requirements to continue, aligning it closely with the 'Expensive Plants' archetype of Group 4.
Asbury	3: 1.0	Partial Match	Plant's high operating and retrofit costs could align with Group 3 if these costs are driven by environmental regulations related to health impacts. It matches the 'economic difficulties due to compliance' secondary driver, but lacks explicit mention of primary health concerns or health-driven regulations.
Avon Lake	4: 1.0	Perfect Match	The explicit statement about economic reasons ('Over the past few years, the plant...') directly matches Group 4's primary driver of high operating costs making the plant expensive.
Asheville	5: 0.05, 0: 0.925, 1: 0.025	Close Match	Plant's 'Economic In-viability' is the driver. Group 0's primary drivers are Aging Infrastructure & Regulatory Non-Compliance. That said, although not an explicit driver, the plant had coal ash problems and radiated ground water. This plant gets a Close Match due to these coal ash violations and being a fuel-blend plant, but not a perfect match due the mismatch in primary drivers.
W H Sammis	4: 1.0	Close Match	Plant's retirement due to 'Utility-driven clean energy transitions' aligns well with Group 4's secondary drivers, as such transitions often prioritize retiring expensive coal plants.
Conesville	2: 0.995, 0: 0.005	Close Match	Plant's 'Regulatory Non-Compliance' aligns with Group 2's 'Regulatory Pressure,' and 'Economic In-viability' matches a secondary outcome. Aging can contribute. The focus on regulatory issues leading to closure fits Group 2 well.
Bull Run	3: 0.965, 1: 0.035	Perfect Match	Plant's retirement due to 'Environmental regulations and economic factors,' and to 'reduce emissions,' aligns perfectly with Group 3's primary driver of environmental regulations (often linked to health) and its secondary driver of economic difficulties from compliance.
Genoa	5: 1.0	No Match	Plant retired at 52, so not a 'young plant'. Although costs were a factor, it is not specifically stated that this is 'renewable cost competition'. As such, the plant does not match the group's archetype.
Chesterfield Power Station	0: 1.0	Perfect Match	Plant's aging infrastructure, state net-zero targets, and coal ash expenses directly match Group 0's primary and secondary drivers. The planned gas conversion also aligns.
Somerset Operating Co LLC	0: 1.0	Close Match	Plant's aging infrastructure and associated economic factors/high maintenance costs align well with the 'Aging Infrastructure' aspect of Group 0's primary drivers. Lacks the explicit 'Regulatory Non-Compliance' component.
Pleasants Power Station	4: 1.0	Close Match	The retirement driver is a near-percect match to Group 4's drivers of High operational costs and economic pressures. However, this plant gets a Close Match and not a Perfect one as it is in WV and the profitability is not directly linked to renewable competition.
Lewis & Clark	4: 0.96, 5: 0.04	Perfect Match	The economic reason 'Low-cost power available on the market' clearly implies the plant is comparatively expensive, aligning perfectly with Group 4's primary driver of high operating costs.
W H Zimmer	3: 1.0	Partial Match	Plant's drivers are market dynamics, economic factors, and emissions reduction. Group 3 focuses on public health and environmental regulations. While emissions reduction is relevant, and economics can be a secondary factor, the stated primary drivers don't highlight specific health impacts or regulatory pressure due to health as the core reason.
Boardman	2: 1.0	Perfect Match	The explicit driver of 'environmental regulations' directly matches Group 2's primary driver of Political & Regulatory Pressure for Clean Energy Transitions.
Wansley	0: 1.0	Close Match	Plant's utility-driven clean energy transition alligns well with Group 0's primary drivers, but lacks explicit mention of regulatory non-compliance, additionally stat (GA) has no clean energy targets. The plant was also subject to a natural gas conversion.
Indiantown Cogeneration LP	0: 0.975, 1: 0.025	Close Match	Plant's aging infrastructure and associated economic factors/high maintenance costs align well with the 'Aging Infrastructure' aspect of Group 0's primary drivers. Lacks the explicit 'Regulatory Non-Compliance' component.
A B Brown Generating Station	4: 1.0	Perfect Match	Plant's retirement due to environmental regulations, economic factors, and emissions reduction aligns perfectly with Group 4's primary (high operating/retrofit costs due to regulations) and secondary drivers (utility/state targets).
R D Green	5: 0.035, 0: 0.965	Perfect Match	Plant's stated drivers of Coal Ash Violations and a natural gas conversion are a perfect match for Group 0's primary drivers.
Birchwood Power	0: 1.0	Weak Match	Plant's driver is 'inability to compete' (economic). Group 0's primary drivers are Aging Infrastructure & Regulatory Non-Compliance. While economic factors can be related, they are not the stated primary drivers for this group archetype, making it a weak match.
H W Pirkey Power Plant	4: 0.965, 5: 0.035	Perfect Match	Retiring to 'comply with new environmental regulations' directly implies that the cost of compliance (retrofits or ongoing) makes the plant expensive, aligning perfectly with Group 4's primary driver concerning environmental retrofit/compliance costs.
Chalk Point Steam	4: 1.0	Perfect Match	Plant's retirement due to being 'uneconomic and regulations' with 'unfavorable economics' is a perfect match for Group 4's primary drivers of high operating costs and environmental retrofit/compliance requirements.
Cheswick Power Plant	3: 1.0	Weak Match	Plant drivers are 'Not cost competitive' and 'General Anti-Coal Sentiment.' Group 3's primary focus is public health and environmental regulations. While there could be indirect links (e.g., sentiment driven by health, or costs increased by regulations), the stated drivers don't directly align with Group 3's core archetype.
Taconite Harbor Energy Center	6: 0.93, 5: 0.07	Perfect Match	Plant's retirement driven by emissions reductions, state climate goals, renewable mandates, and being uneconomic directly aligns with Group 6's primary (Political Opposition, Economic Difficulties) and secondary drivers.

Supplemental Table 9: Plant-Level Match Explanations

This table explains the rationale behind each plant's "Match to Archetype?" classification. It provides qualitative reasoning for why each plant did or did not align with its group's retirement drivers. Each plant's retirement drivers are listed in Supplemental Table 10.

Plant Name	**Group Match** (Group Num: Match Score)	Match to Archetype?	Driver 1	Driver 2
Dolet Hills	4: 0.94, 5: 0.06	Perfect Match	Utility Cost reduction effort	High operating costs
Martin Drake // South Plant	7: 1.0	Close Match	Aging infrastructure and transition to cleaner energy	Sulfur-dioxide emissions violated federal standards.
E D Edwards	6: 1.0	Close Match	Environmental violations leading to legal settlement	Settlement required plant closure by 2022
Lowman Energy Center	0: 1.0	Perfect Match	Aging infrastructure	Coal Ash Pond Violation
Waukegan	4: 1.0	Perfect Match	"Poor financial results"	Economic pressures + utility transition away from coal
Escalante	4: 1.0	Perfect Match	Economic In-viability & Renewable Competion	Fuel Conversion (to hydrogen)
Will County	5: 0.07, 6: 0.93	Perfect Match	State-driven climate goals + renewable mandates	Economic difficulties – bankrupt in 2012
Oklaunion	3: 0.95, 5: 0.05	Partial Match	Economic In-viability & Renewable Competion	Aging infrastructure – high retrofit costs
Joppa Steam	2: 0.915, 6: 0.085	Partial Match	Economic unviability and aging infrastructure	High maintenance costs and low efficiency
Bridgeport Station	0: 1.0	Partial Match	Utility net zero target	Environmental justice concerns in disadvantaged communities
Crawfordsville Power Plant	6: 1.0	Partial Match	Aging infrastructure	NaN
C D McIntosh Jr	0: 0.98, 1: 0.02	Close Match	Aging infrastructure – Increasing costs of operation	NaN
Lansing	2: 0.095, 5: 0.07, 4: 0.835	Close Match	Utility's clean energy transition plan	Retirement of coal plants to reduce emissions
R Gallagher	2: 0.085, 4: 0.915	Close Match	Economic Deciscion	NaN
Elmer Smith	4: 0.95, 5: 0.05	Perfect Match	"decision to retire Unit 1 is purely economical."	tens of millions of dollars of new investments will be needed to keep the plant running, retail rates will increase by 20 percent by 2018 and 80 percent by 2025
Morgantown Generating Plant	0: 1.0	Perfect Match	Aging Infrastructure & Regulatory Non-Compliance	Coal Ash Pond Violation
Paradise	4: 1.0	Perfect Match	Environmental regulations and economic factors	Retirement of coal units to reduce emissions
Dan E Karn	0: 1.0	Close Match	Aging Infrastructure (60+)	Renewable Targets by 2040 + agreement with Michigan State
Kenneth C Coleman	3: 1.0	Weak Match	Aging infrastructure and economic factors	High maintenance costs and low efficiency
St Clair	2: 1.0	Perfect Match	Economics + failure to meet federal air quality standards	State-driven climate goals + renewable mandates
Chalk Point LLC	0: 1.0	Perfect Match	Aging Infrastructure & Regulatory Non-Compliance	Water pollution violations
Trenton Channel	0: 1.0	Partial Match	Utility Net Zero	General Anti-Coal Sentiment
Dickerson	2: 0.905, 0: 0.095	Close Match	high compliance costs	NaN
Meramec	6: 0.13, 4: 0.78, 2: 0.09	Perfect Match	Environmental regulations and economic factors	Retirement of coal units to reduce emissions
River Rouge	4: 0.96, 5: 0.04	Close Match	Utility's clean energy transition plan	Retirement of coal plants to reduce emissions
San Juan	0: 1.0	Close Match	Coal Ash Pond Violation	Environmental justice concerns in disadvantaged communities
J B Sims	2: 1.0	Weak Match	age, high operating costs, GHG reductions	Needed expensive upgrades
Dunkirk Generating Plant	0: 1.0	Close Match	Aging infrastructure – Increasing costs of operation	Attempted NG conversion – too costly
Eckert Station	3: 1.0	Partial Match	Utility's clean energy transition plan	Retirement of coal plants to reduce emissions
R M Heskett	6: 1.0	Close Match	Aging infrastructure and economic factors	High maintenance costs and low efficiency
Hoot Lake	4: 0.97, 5: 0.03	Close Match	aging infrastructure (100 years!)	NaN
Asbury	3: 1.0	Partial Match	High operating costs	High retrofit costs required
Avon Lake	4: 1.0	Perfect Match	economic: "Over the past few years, the plant's capacity factor diminished as the economic realities of the decline in coal played out."	NaN
Asheville	5: 0.05, 0: 0.925, 1: 0.025	Close Match	Economic In-viability	Coal Ash Violations - not mentioned directly as a driver
W H Sammis	4: 1.0	Close Match	Utility-driven clean energy transitions	NaN
Conesville	2: 0.995, 0: 0.005	Close Match	Aging Infrastructure & Regulatory Non-Compliance	Economic In-viability
Bull Run	3: 0.965, 1: 0.035	Perfect Match	Environmental regulations and economic factors	Retirement of coal units to reduce emissions
Genoa	5: 1.0	No Match	Fuel shortages, low plant capacity	According to Dairyland, it is cheaper to buy power from the local grid than to run the power plant.
Chesterfield Power Station	0: 1.0	Perfect Match	Aging Infrastructure	State net-zero targets + Coal Ash Expenses
Somerset Operating Co LLC	0: 1.0	Weak Match	Aging infrastructure and economic factors	High maintenance costs and low efficiency
Pleasants Power Station	4: 1.0	Close Match	high operating costs, and the changing energy landscape	NaN
Lewis & Clark	4: 0.96, 5: 0.04	Perfect Match	economic: Low-cost power available on the market, due to low-cost natural gas and increasing wind resources, as well as rising costs to operate the coal-fired facilities	NaN
W H Zimmer	3: 1.0	Partial Match	Market dynamics and economic factors	Retirement of coal units to reduce emissions
Boardman	2: 1.0	Perfect Match	environmental regulations: more than a half billion dollars in pollution controls needed to keep operational"	NaN
Wansley	0: 1.0	Close Match	utility-driven clean energy transition	Natural gas conversion
Indiantown Cogeneration LP	0: 0.975, 1: 0.025	Close Match	Aging infrastructure and economic factors	High maintenance costs and low efficiency
A B Brown Generating Station	4: 1.0	Perfect Match	Environmental regulations and economic factors	Retirement of coal units to reduce emissions
R D Green	5: 0.035, 0: 0.965	Perfect Match	Coal Ash Violation	Natural Gas Conversion
Birchwood Power	0: 1.0	Weak Match	noneconomic - "inability to compete with lower-cost electricity produced by natural gas and renewable energy..."	NaN
H W Pirkey Power Plant	4: 0.965, 5: 0.035	Perfect Match	"comply with new environmental regulations"	NaN
Chalk Point Steam	4: 1.0	Perfect Match	uneconomic and regulations - "unfavorable economic conditions and increased costs associated with environmental compliance"	NaN
Cheswick Power Plant	3: 1.0	Weak Match	Not cost competitive	General Anti-Coal Sentiment
Taconite Harbor Energy Center	6: 0.93, 5: 0.07	Perfect Match	Emissions reductions	State-driven climate goals + renewable mandates

Supplemental Table 10: Plant-Level Alignment with Retirement Archetypes
Each retiring coal plant is compared against the primary and secondary drivers of its assigned group. Match classifications reflect how well real-life retirement reasons align with group-level archetypes, based on publicly available documentation. See Supplemental Table 9 for explanations of the "Match to Archetype?" categorization.

Match Type	Count	Fraction	Percentage
Perfect Match	22	22/52	42.31%
Close Match	19	19/52	36.54%
Partial Match	7	7/52	13.46%
Weak Match	3	3/52	5.77%
No Match	1	1/52	1.92%

Supplemental Table 11: Match Count and Percentages for Plants
A breakdown of the match status, from our retired plant *Target Marching Analysis*, for how well these plants allign with our group-based retirement archetypes.

D Supplementary References

[1] Jeremy Wayland, Corinna Coupette, and Bastian Rieck. "Mapping the Multiverse of Latent Representations". In: *Forty-first International Conference on Machine Learning*. 2024.

[2] Darrell Proctor. *Southern Will Close More than Half of Coal Fleet*. en-US. Nov. 2021.

[3] *Cleantech Edge: Five is the new zero for energy transition debt*. en-us.

[4] *A Louisiana utility hopes to get more time to close its 4 coal ash ponds under a new rollback | Business News | nola.com*.

[5] *'Served its purpose': Duke's coal power plants will retire*.

[6] Casey Roberts. *PJM Thwarts Maryland's Coal-Free Ambitions While Costing Marylanders Millions | Sierra Club*. en.

[7] Joe Sylvester jsylvester@thedanvillenews.com. *Talen moving toward converting Montour Plant to gas*. en. Feb. 2022.

[8] "E.C. Gaston Steam Plant". In: *Ashtracker* (Oct. 2022).

[9] *Operating costs at NIPSCO coal plants eyed for retirement run above market*. en-us.

[10] *Indiana's Hoosier Energy to retire its 1,070 MW coal plant by 2023*. en-US.

[11] *NIPSCO Announces Timeline for Coal-Fired Power Plant*. en-US.

[12] Sarah Bowman. *AES Indiana plans to leave coal power behind by 2025, parent company says*. en-US.

[13] Paul Monies. *Proposed settlement would retire coal units at Oologah power plant*. en-US.

[14] *Legislators breathe a little life into coal power plant due to be retired*.

[15] *TVA agrees to dig up 12 million tons of coal ash at Gallatin plant*.

[16] Jennie Jorgenson et al. *Comparing Capacity Credit Calculations for Wind: A Case Study in Texas*. en. Tech. rep. NREL/TP-5C00-80486, 1823456, MainId:42689. Sept. 2021, NREL/TP–5C00–80486, 1823456, MainId:42689. DOI: 10.2172/1823456.

[17] Gennaro Cordasco and Luisa Gargano. *Community Detection via Semi-Synchronous Label Propagation Algorithms*. arXiv:1103.4550 [physics]. Mar. 2011. DOI: 10.1504/..045103.

[18] Kyle Davidson. *DTE reaches settlement on energy plan, agrees to retire coal plants early • Michigan Advance*. en-US. July 2023.

[19] *In settlement, power plants to shut by '30 | Arkansas Democrat Gazette*. en. Mar. 2021.

[20] Jeff Della Rosa. *Judge approves Entergy Arkansas, Sierra Club agreement to retire coal, natural gas power plants*. en-US. Mar. 2021.

[21] *AEP sets retirement date for massive Rockport coal unit in Indiana*. en-us.

[22] Post et al. *DTE Electric agrees to speed Michigan coal plant retirements, renewable and energy storage build-out*. en-US.

[23] Post et al. *Xcel to retire Texas coal-fired power plant early, speeding up companywide exit from coal to 2030*. en-US.

[24] Leo Bertucci | lbertucci@vicad.com. *Coleto Creek Power Plant avoids new EPA rule because of impending closure*. en. Mar. 2023.

[25] By Diego Mendoza-Moyers. *Fate of CPS' last coal plant in sight*. en. Apr. 2022.

[26] Mark Rosenberg | mrosenberg@vicad.com. *Coleto Creek Power Plant shutting down by 2027*. en. Dec. 2020.

[27] *Vistra to retire 6.8 GW coal, blaming 'irreparably dysfunctional MISO market'*. en-US.