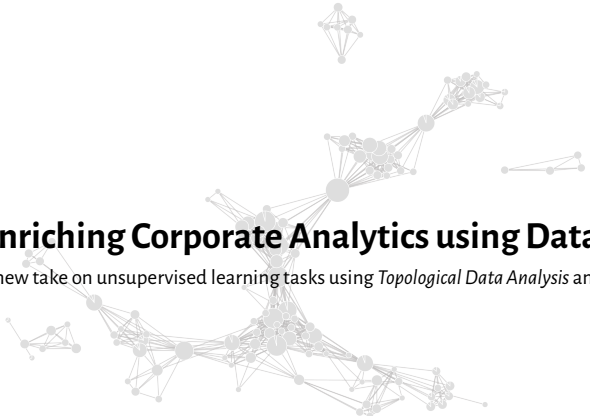


Enriching Corporate Analytics using Data's Shape

A new take on unsupervised learning tasks using *Topological Data Analysis* and *Graph Learning*.

Today's Talk

1. Our work on Coal Phaseout
2. Quick intro to Graph Learning and Topology
3. Overview of our method
4. Other applications of interest

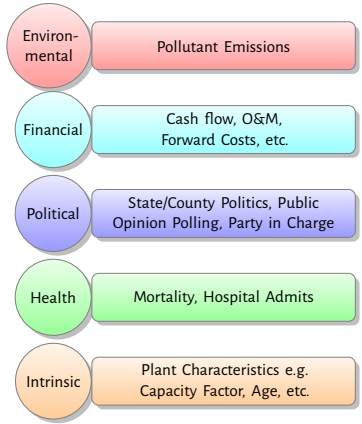
The Coal Problem

- Significant environmental, financial, and public health benefits associated with phasing out coal.
- *Should be easy, right?*
- Coal phaseout is complex and multifaceted. Few historical examples of coal plants that have been labeled as either ”good” or ”bad” to retire.
- Challenging to train machine learning models due to this lack of labeled data.

Data Collection

Data Sources

- US EIA
- US EPA
- Clean Air Task Force
- Yale Program on Climate Change Communications
- Energy Innovation
- Rocky Mountain Institute
- Sierra Club



*Why can't **standard** analysis tools address the Coal Problem?*

The Curse of Dimensionality

- As the dimension increases, the number of data points needed to guarantee reliable results grows exponentially.
- Sparse data causes standard statistical and machine learning techniques break down.
- In the absence of a rich, well-labeled training set, many deep learning frameworks fail to perform well.
- *How can we extract insights from datasets and problems that suffer from the curse of dimensionality?*

"While simple arguments reveal the impossibility of learning from generic high-dimensional data as a result of the curse of dimensionality, there is hope for physically-structured data, where we can employ two fundamental principles: ***symmetry*** *and* ***scale separation****."*

– Micheal Bronstein [1]

DeepMind Professor of AI, Oxford University
(former) Head of Graph Learning Research, Twitter

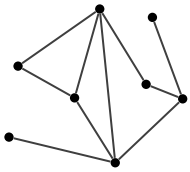
Our Solution

We combine methods from:

1. Graph Learning
2. Topology

Curvature Filtrations for Graph Generative Model Evaluation (2023) [3].
Authors: Joshua Southern, Jeremy Wayland, Michael Bronstein, Bastian Rieck.

What is a Graph?



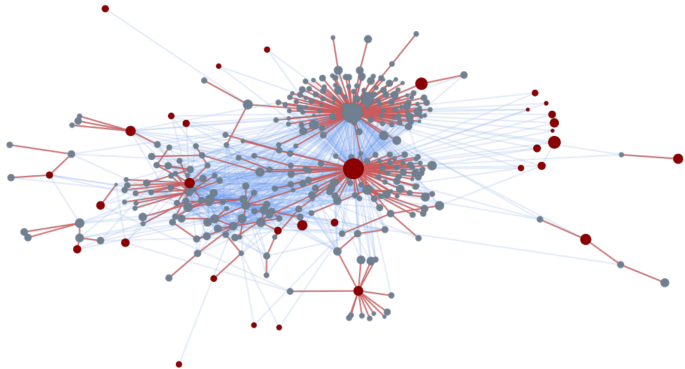
A graph $G = (N, E)$.

Graph Learning

Fake News Detection (2019)

- Bronstein and co-authors use *graph learning* to identify fake news with exceptional accuracy ($\sim 93\%$).
- They were also able to extract meaningful insights about Twitter's users and assign users a *credibility* rating.
- Their company **Fabula AI** was acquired by Twitter in 2020 to fight the spread of misinformation.

Graph Learning



News Spreading across Twitter

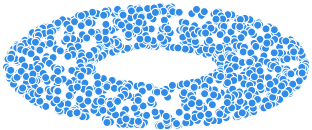
A single news story spreading on a subset of the Twitter social network, modeled as a graph. [2].
Light blue edges are social connections between users. Red nodes are users who tweeted the *url* directly. Red edges represent the spread of the *url* through the network.

Graph Learning

- How do we generate *informative* graph models of our data?
- What is the right tool for dealing with even *sparser* data?
- What happens in the case that we do *not* have well-defined labels?

Topology has exactly the properties we are looking for to address these questions!

What is Topology?



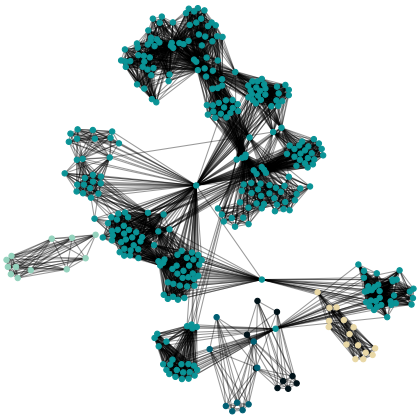
Reality is often messy…

Why Topological Data Analysis (TDA)?

1. Works with sparse data
2. Does *not* require labeled data
3. Captures structure of data at multiple scales
4. Avoids inductive biases
5. Interpretable and Transparent
6. Known Failure Modes

Coal Mapper

- To suit the needs of our collaborators, we developed a *topological clustering algorithm*, that we call the **Coal Mapper**.

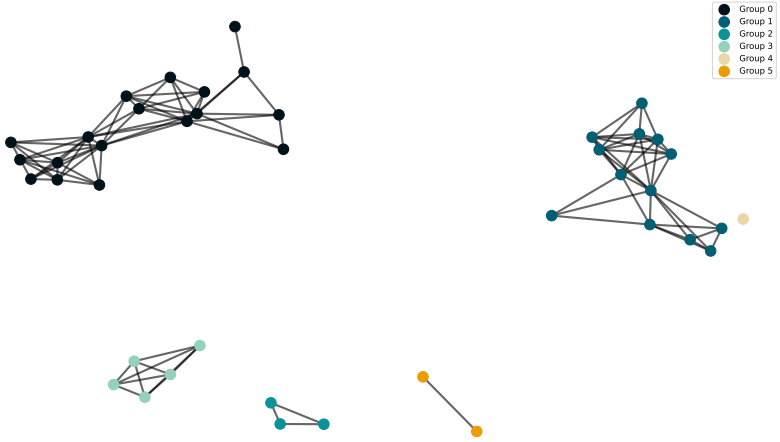


An example Graph Model of US Coal Plants

How does it work?

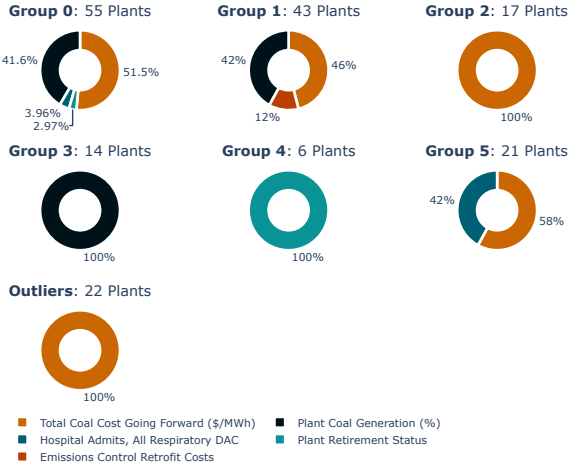
1. Search for *shape* and *symmetries* in your data at multiple scales.
2. Generate *graph models* that capture this shape.
3. Identify important *scales* and select informative models.
4. Extract informative *groupings*.

Addressing the Coal Problem

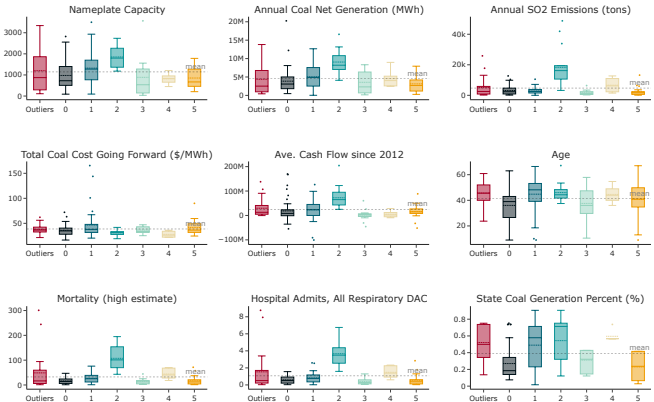


How do we extract insights from our Graph Models?

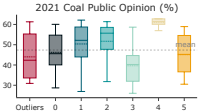
Understanding Group Composition

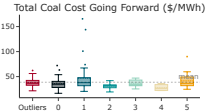


Analyzing Group Key Features

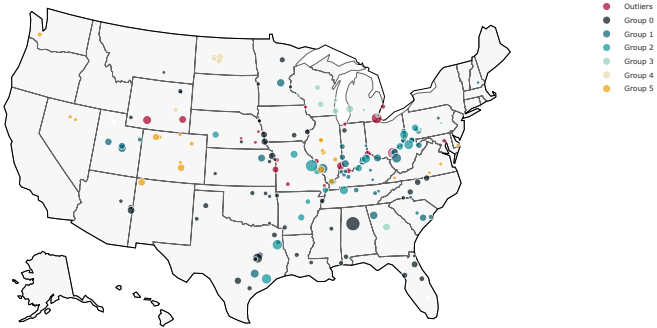


Analyzing Group Key Features

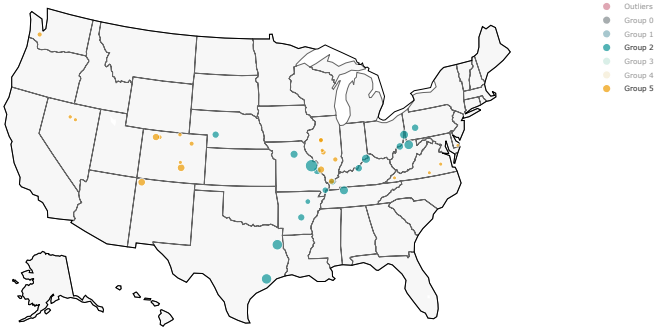




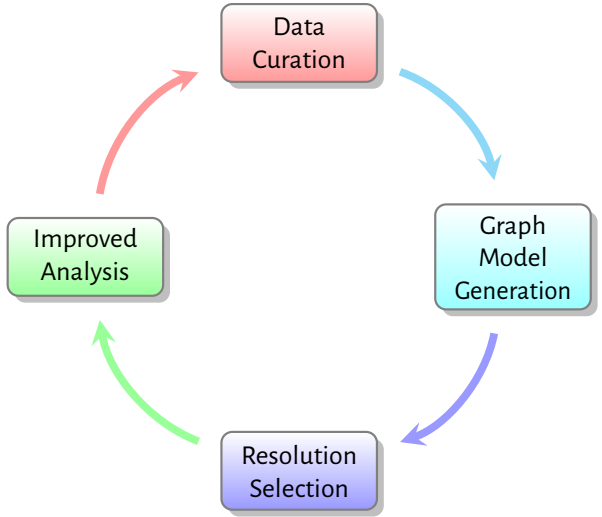
Geographic Layout



Geographic Layout



Our Codebase



Interfacing with Existing Frameworks

1. Can be integrated into standard prediction architectures
2. Complementary to Deep Learning and Large Language Models

Looking Toward Industry

What other fields could benefit from our approach?

- Finance Sector → Companies
- Healthcare → Patients
- Environmental Policy → Coral Reef Destruction

What can we do for you?