HELMHOLTZ
MUNICH

AIH Institute of AI for Health

UC SANTA CRUZ | BE Baskin Engineering

UC SANTA BARBARA
Bren School of Environmental
Science & Management

Decarbonization using Data's Shape

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

Agenda
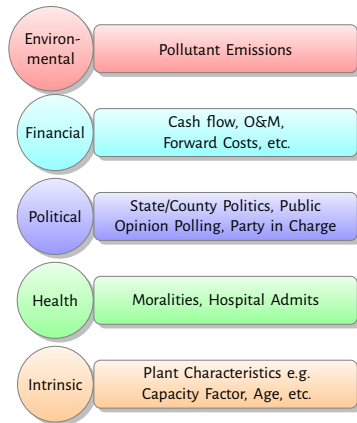
Coal Plant Retirement

- Significant 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.
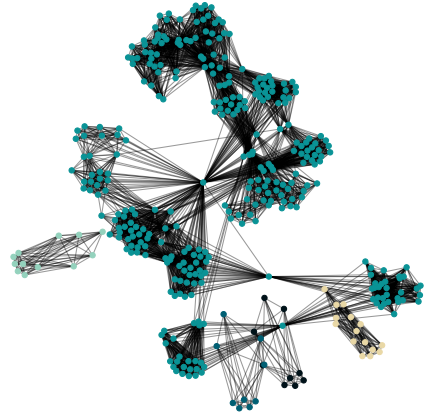
Data Collection

Sources

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

Environ-mental	Pollutant Emissions
Financial	Cash flow, O&M, Forward Costs, etc.
Political	State/County Politics, Public Opinion Polling, Party in Charge
Health	Moralities, Hospital Admits
Intrinsic	Plant Characteristics e.g. Capacity Factor, Age, etc.

Krv Analytics

Topological
Hyperparameter
Evalutation
Mapping
Algorithm

An example Graph Model of US Coal Plants.

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

Krv
Analytics

The Curse of Dimensionality

The Curse of Dimensionality

- As the dimension increases, the *number of data points needed to guarantee reliable results grows exponentially*.

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 to break down.

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 to break down.
- *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**."

"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

Krv
Analytics

Our Solution

Our Solution

We combine methods from:

Our Solution

We combine methods from:

1. Graph Learning

Our Solution

We combine methods from:

1. Graph Learning
2. Topology

Our Solution

We combine methods from:

1. Graph Learning
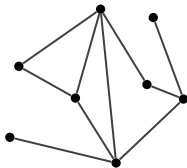2. Topology

Latest Paper [3]:

Graph Learning
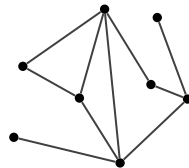
Graph Learning

Fake News Detection (2019)
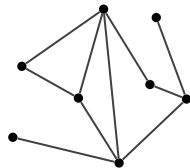
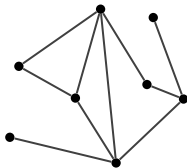# Graph Learning

Fake News Detection (2019)

- Bronstein and co-authors use *graph learning* to identify fake news with exceptional accuracy ($\sim 93\%$).

Graph Learning

Fake News Detection (2019)

- Bronstein and co-authors use *graph learning* to identify fake news with exceptional accuracy (~ 93%).
- Their company **Fabula AI** was acquired by Twitter in 2020 to fight the spread of misinformation.

Graph Learning
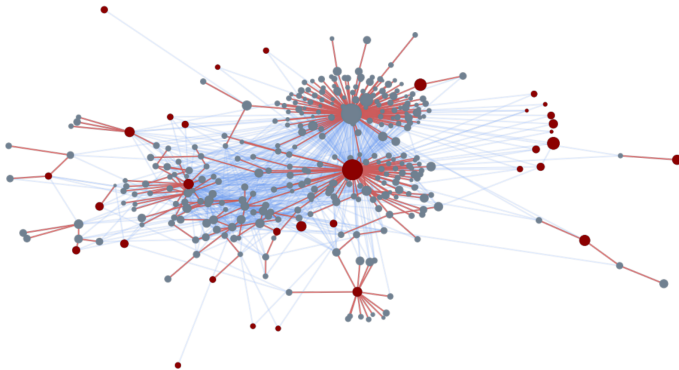
News Spreading across Twitter

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.

Open Questions

Open Questions

- How do we generate *informative* graph models of our data?

Open Questions

- How do we generate *informative* graph models of our data?
- What is the right tool for dealing with even *sparser* data?

Open Questions

- 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?

Open Questions

- 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!

Why Topological Data Analysis (TDA)?

Why Topological Data Analysis (TDA)?

1 Captures structure at multiple scales

Why Topological Data Analysis (TDA)?

1 Captures structure at multiple scales

2 Works with *sparse* data

Why Topological Data Analysis (TDA)?

1. Captures structure at multiple scales
2. Works with *sparse* data
3. *No* labels required

Why Topological Data Analysis (TDA)?

1. Captures structure at multiple scales
2. Works with *sparse* data
3. *No* labels required
4. Avoids inductive biases

Why Topological Data Analysis (TDA)?

1. Captures structure at multiple scales
2. Works with *sparse* data
3. *No* labels required
4. Avoids inductive biases
5. Transparent and interpretable algorithm

THEMA

THEMA

- To suit the needs of our collaborators, we use THEMA to perform *unsupervised classification* of US Coal Plants.

THEMA

- To suit the needs of our collaborators, we use THEMA to perform *unsupervised classification* of US Coal Plants.
- Our models provide a road map for strategizing plant decomission and retirement.

THEMA

- To suit the needs of our collaborators, we use THEMA to perform *unsupervised classification* of US Coal Plants.
- Our models provide a road map for strategizing plant decomission and retirement.
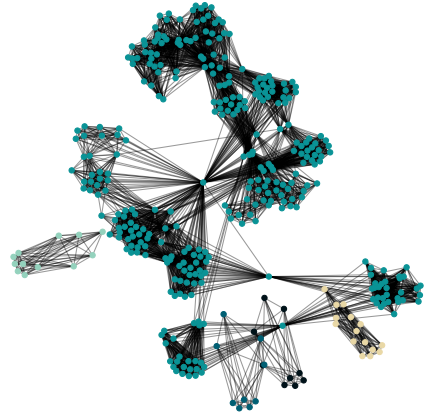
An example Graph Model of US Coal Plants.

How does it work?

How does it work?

1 Search for *shape* and *symmetries* in your data at multiple scales.

Krv Analytics

How does it work?

1. Search for *shape* and *symmetries* in your data at multiple scales.
2. Generate *graph models* that capture this shape.

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.

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*.

How does this address the Coal Plant Retirement Problem?

Krv
rv
Analytics

Coal Plant Analysis
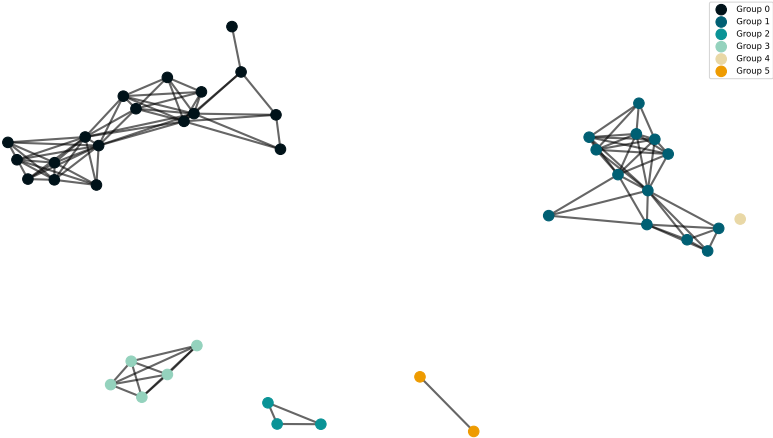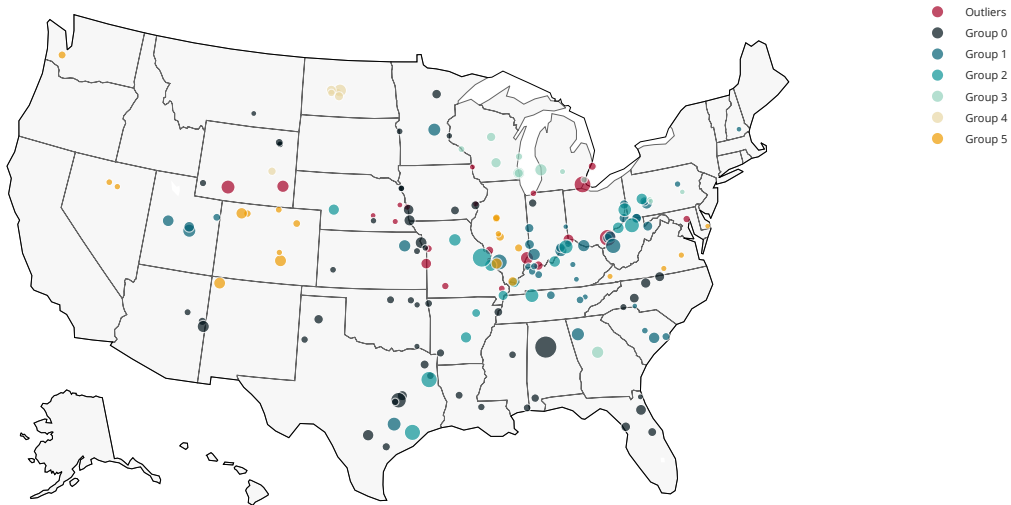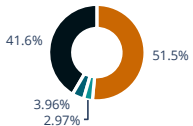
Outliers
Group 0
Group 1
Group 2
Group 3
Group 4
Group 5

Decarbonization Demo

Krv
Analytics

Group 0: 55 Plants — 51.5%, 41.6%, 3.96%, 2.97%

Group 1: 43 Plants — 46%, 42%, 12%
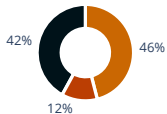
Group 2: 17 Plants — 100%

Group 3: 14 Plants — 100%

Group 4: 6 Plants — 100%

Group 5: 21 Plants — 58%, 42%

Outliers: 22 Plants — 100%

Legend:
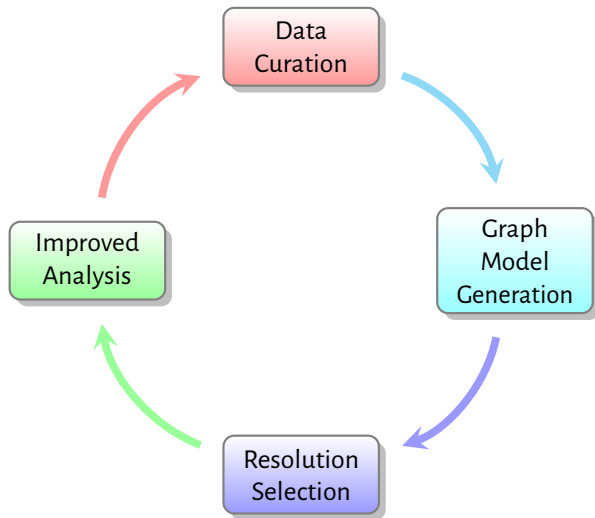- Total Coal Cost Going Forward ($/MWh)
- Hospital Admits, All Respiratory DAC
- Emissions Control Retrofit Costs
- Plant Coal Generation (%)
- Plant Retirement Status

Nameplate Capacity

Annual Coal Net Generation (MWh)

Annual SO2 Emissions (tons)

Total Coal Cost Going Forward ($/MWh)

Ave. Cash Flow since 2012

Age

Mortality (high estimate)

Hospital Admits, All Respiratory DAC

State Coal Generation Percent (%)

Decarbonization Demo
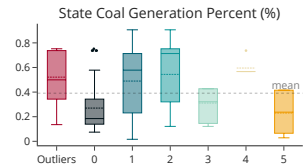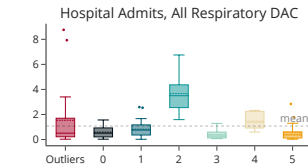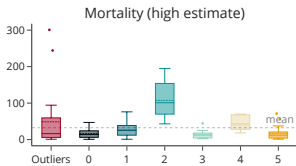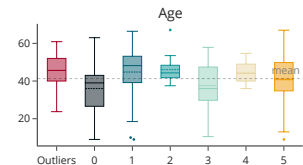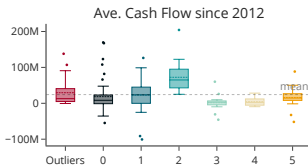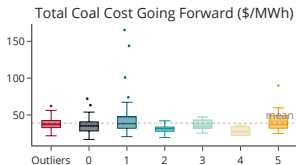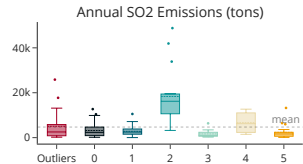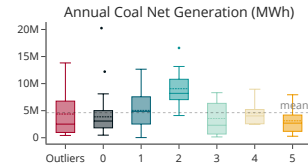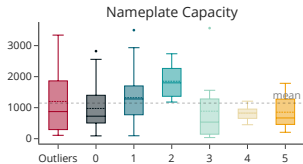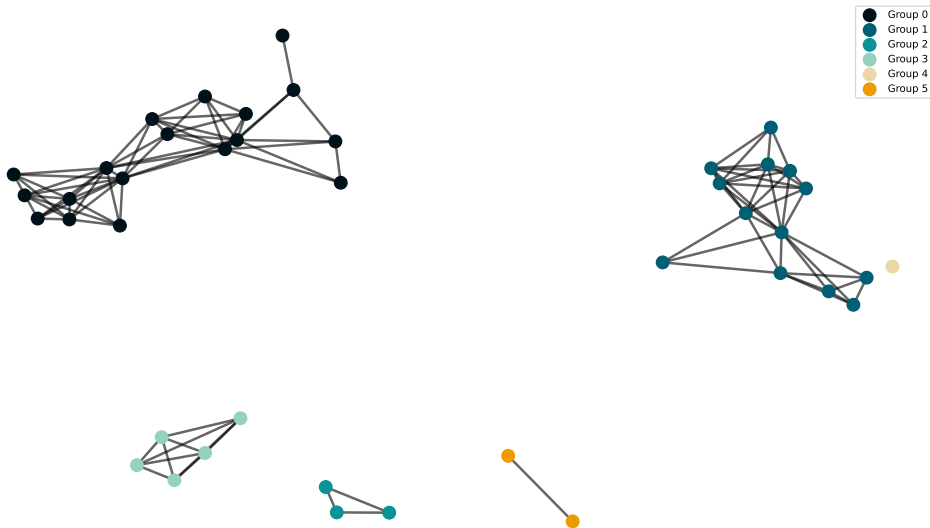
Lets take a step back...

Krv
Analytics

Our Service

Looking Toward Industry

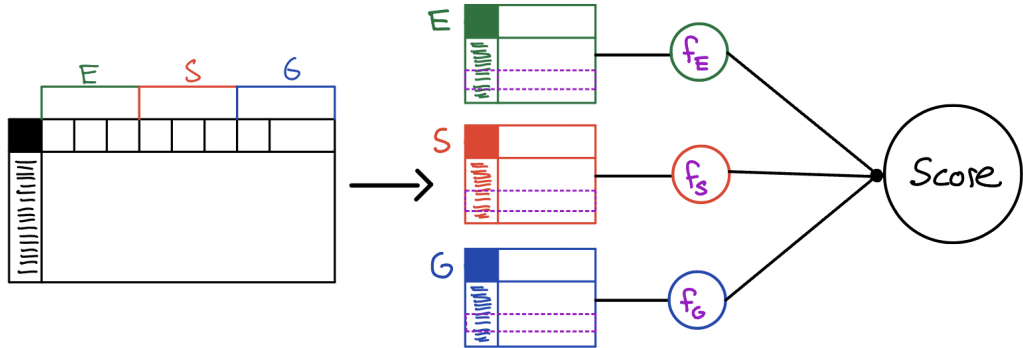*What other fields could benefit from our approach?*

- Healthcare
- Finance Sector

Looking Toward Industry

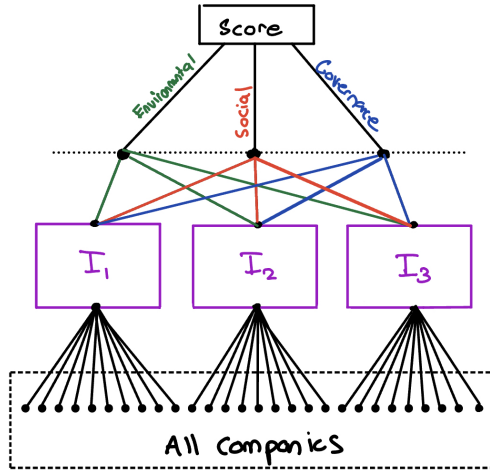*What other fields could benefit from our approach?*

- Healthcare ⟼ Patients
- Finance Sector ⟼ Companies

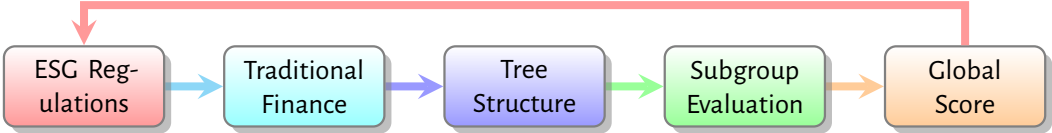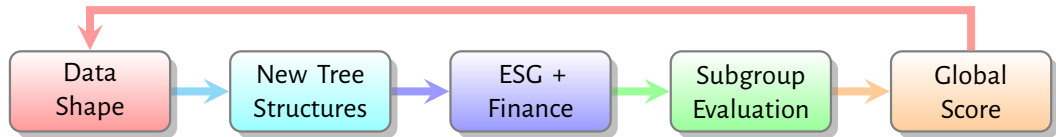# ESG Evaluation

Localizing Scores

Current Framework

A New Take

What can THEMA do for you?

Krv
Analytics

References

[1]   M. M. Bronstein et al., *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges*, 2021, arXiv: `2104.13478 [cs.LG]`.

[2]   F. Monti et al., *Fake News Detection on Social Media using Geometric Deep Learning*, 2019, arXiv: `1902.06673 [cs.SI]`.

[3]   J. Southern et al., *Curvature Filtrations for Graph Generative Model Evaluation*, 2023, arXiv: `2301.12906 [cs.LG]`.

Our Code:

Krv Analytics