All the World's A (Hyper)Graph

A Data Drama

Corinna Coupette, Jilles Vreeken, and Bastian Rieck

DRAMATIS PERSONÆ

AUTHORS.
REVIEWER, a reader. } Persons in the Induction.
CREATURE, a curious mind.
HYPERBARD, a faun, sovereign of spirits.
GRAPH, a gentle spirit.

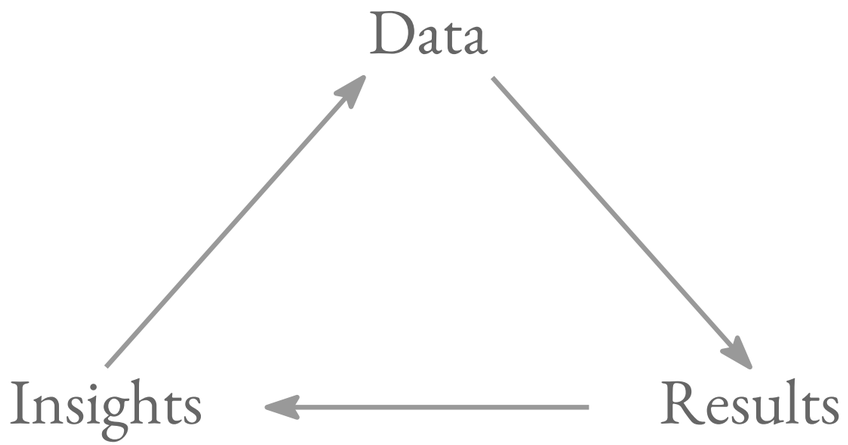
PROFESSOR,
SENIOR RESEARCHER,
COLLEAGUE. } Part of the Community.

TUTOR,
SECRETARY,
DEADLINES. } Serving the Community.

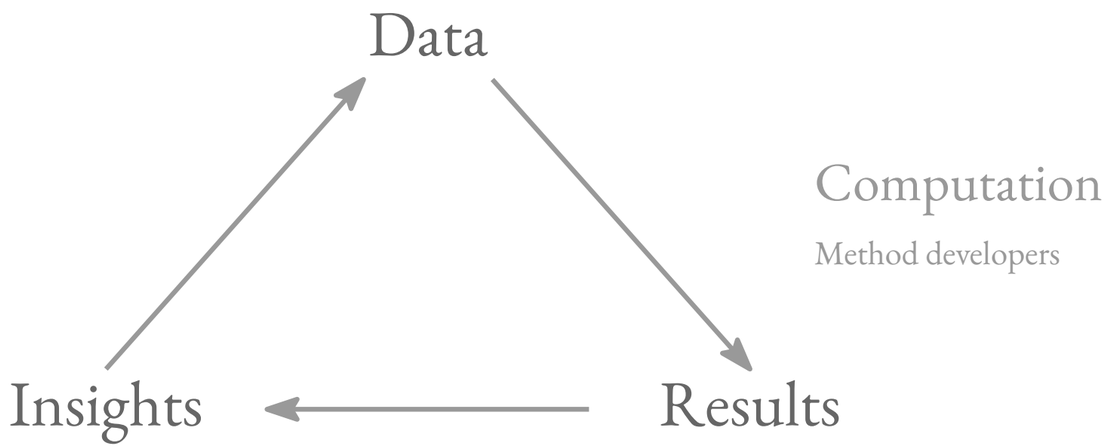
SCENE.—*Sometimes in the Community; and sometimes in the forest.*

Motivation

Motivation



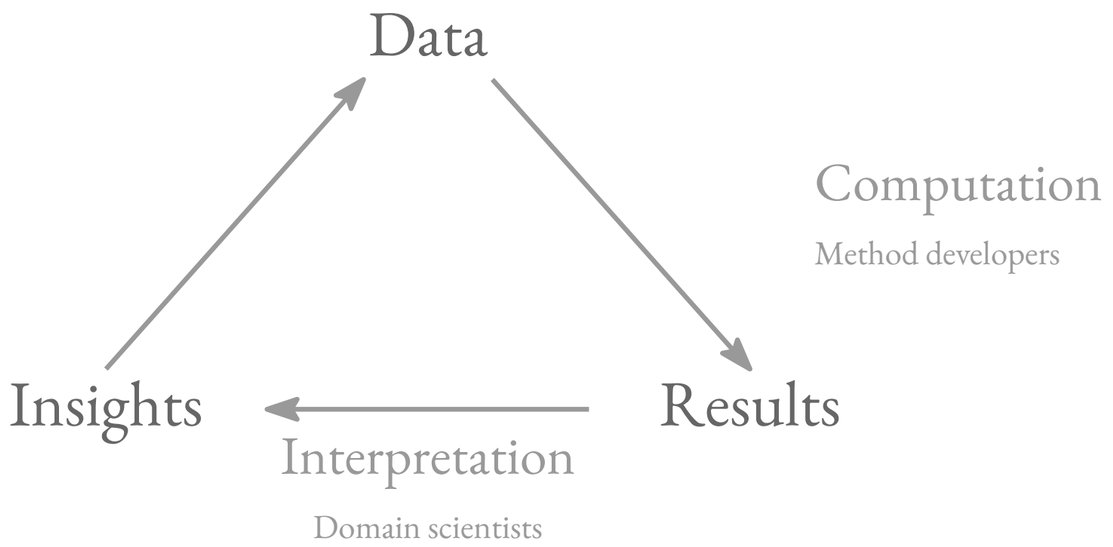
Motivation



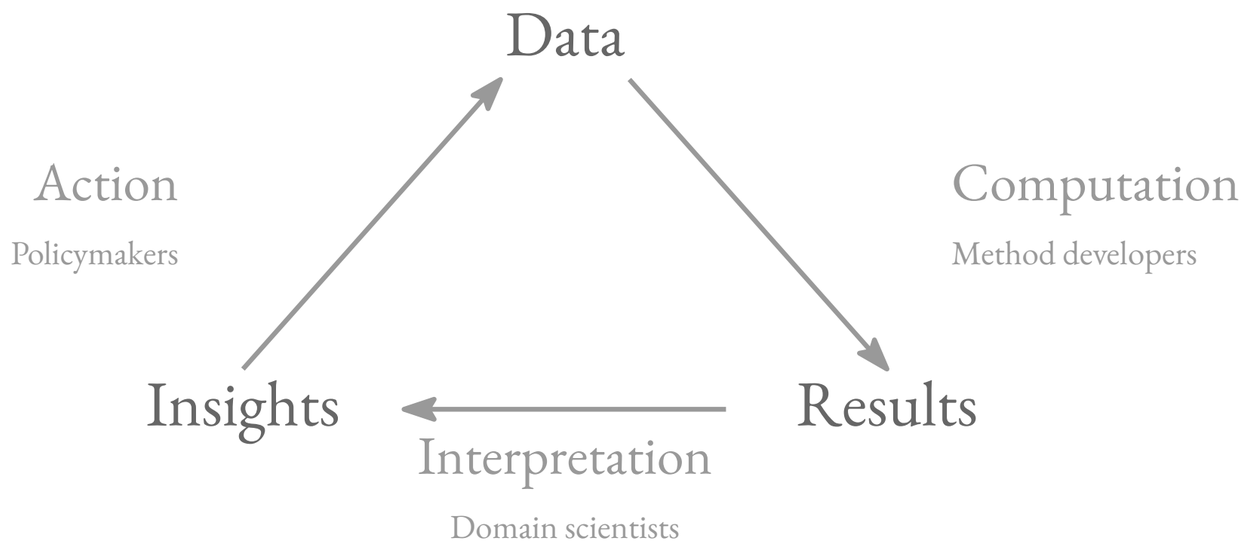
Computation

Method developers

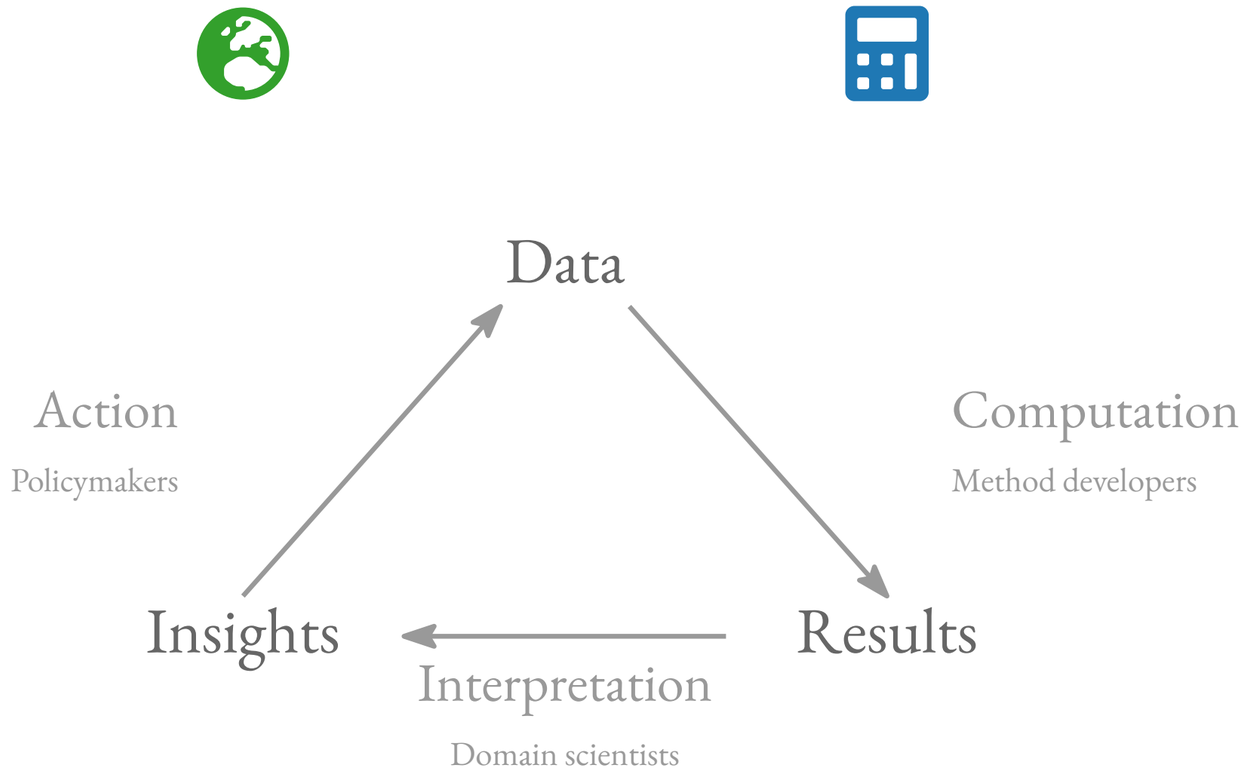
Motivation


Data
Computation
Method developers
Insights
Results
Interpretation
Domain scientists

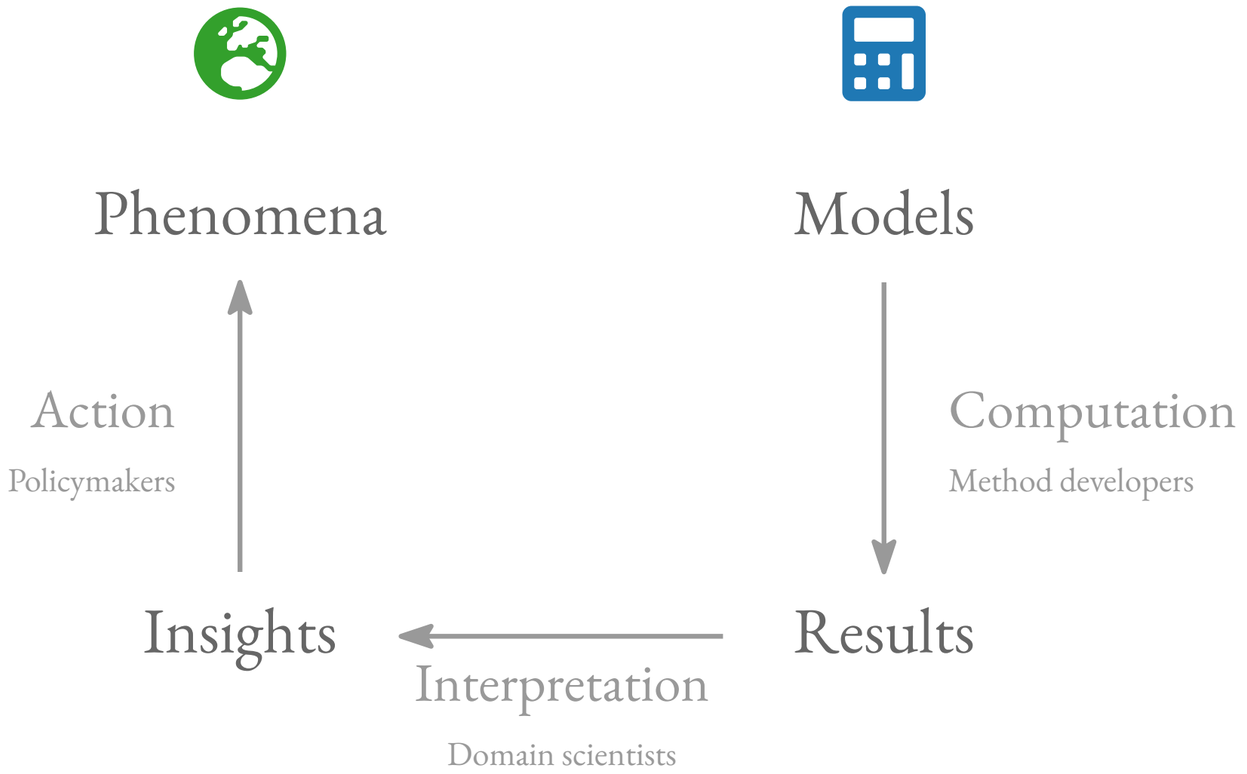
Motivation


Data
Action
Policymakers
Computation
Method developers
Insights
Results
Interpretation
Domain scientists

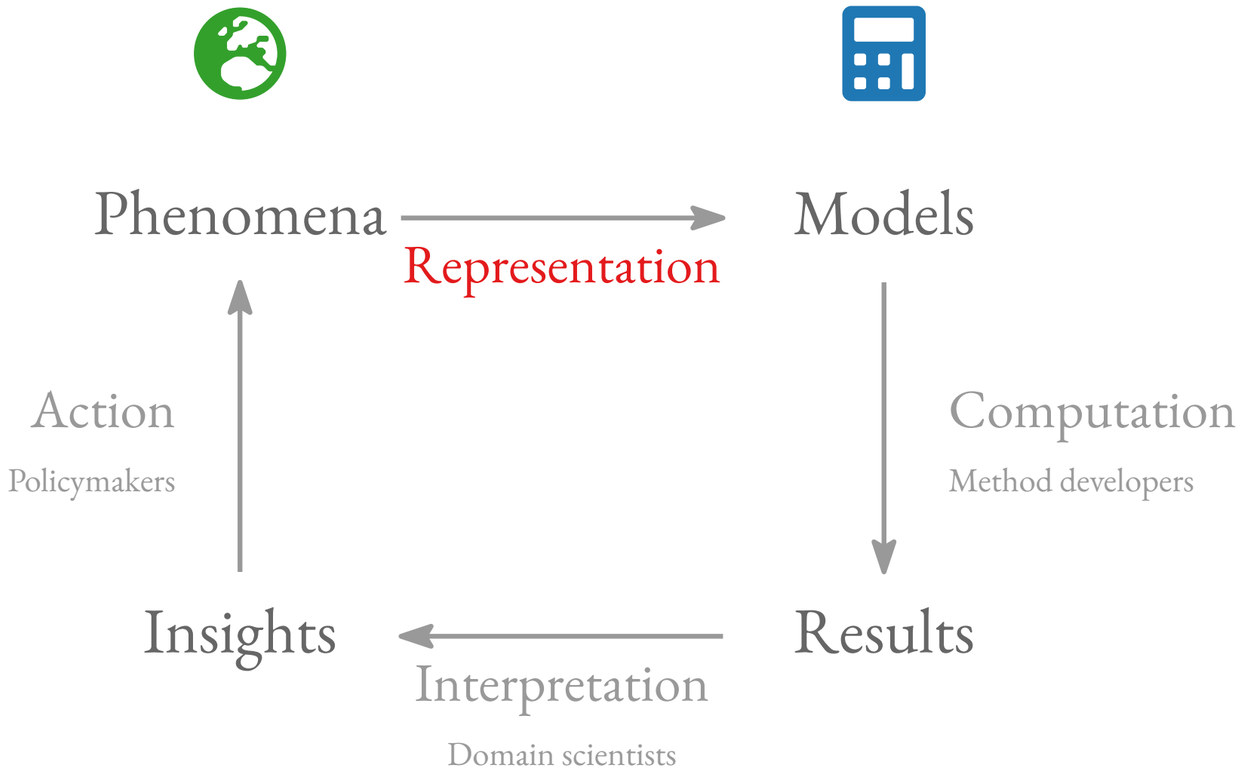
Motivation


Data
Action
Policymakers
Computation
Method developers
Insights
Results
Interpretation
Domain scientists

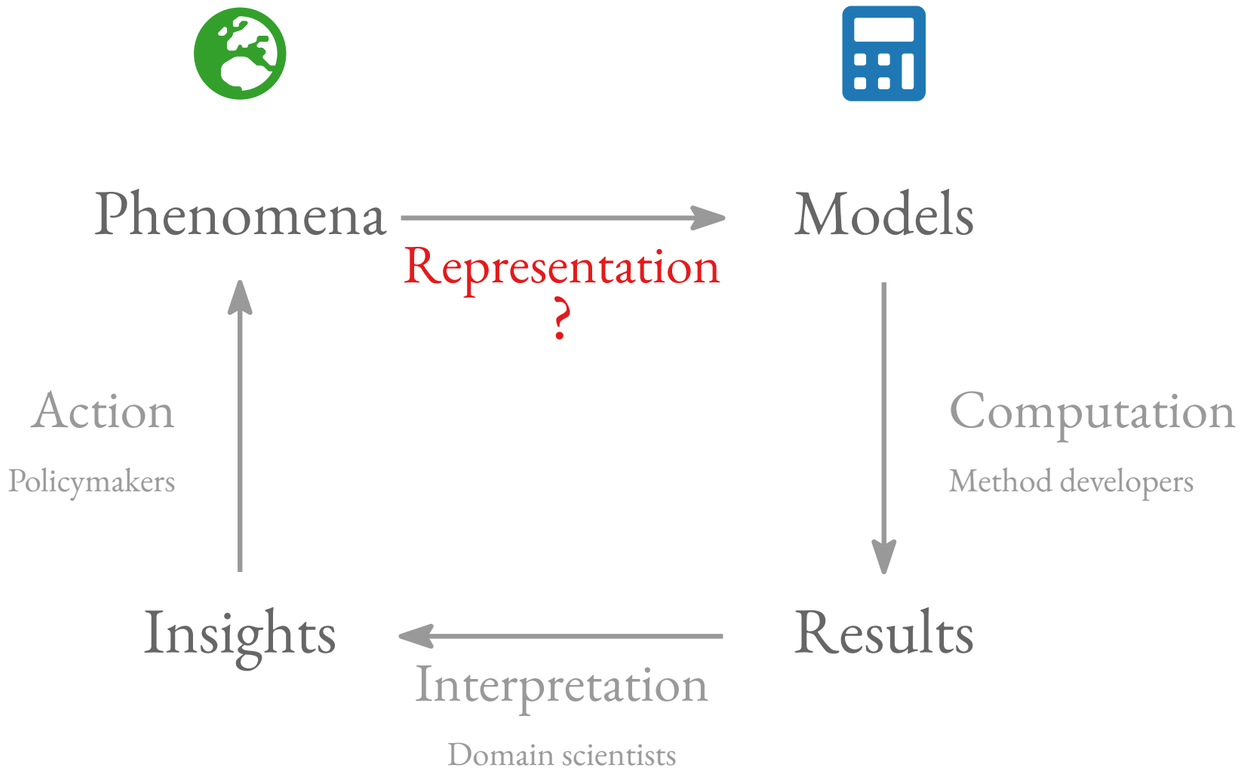
Motivation


Phenomena
Models
Action
Policymakers
Computation
Method developers
Insights
Results
Interpretation
Domain scientists

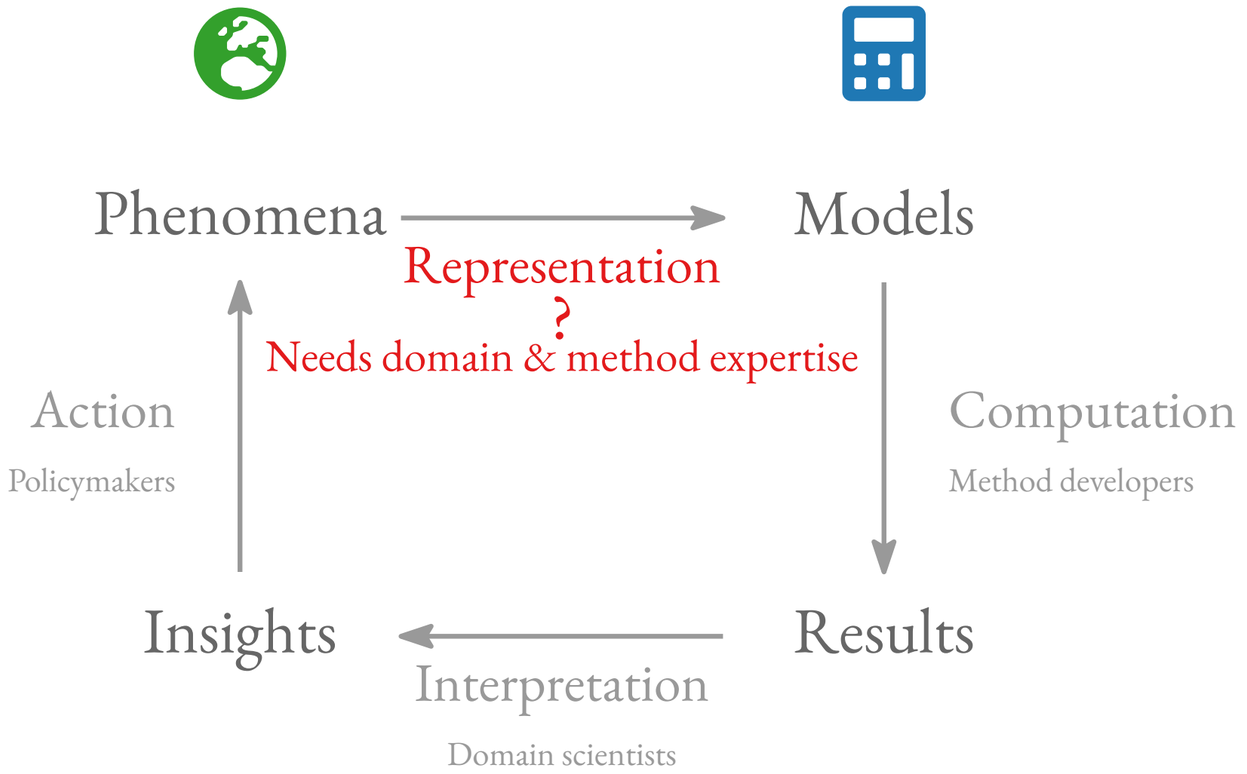
Motivation


Phenomena
Representation
Models
Action
Policymakers
Computation
Method developers
Insights
Interpretation
Domain scientists
Results

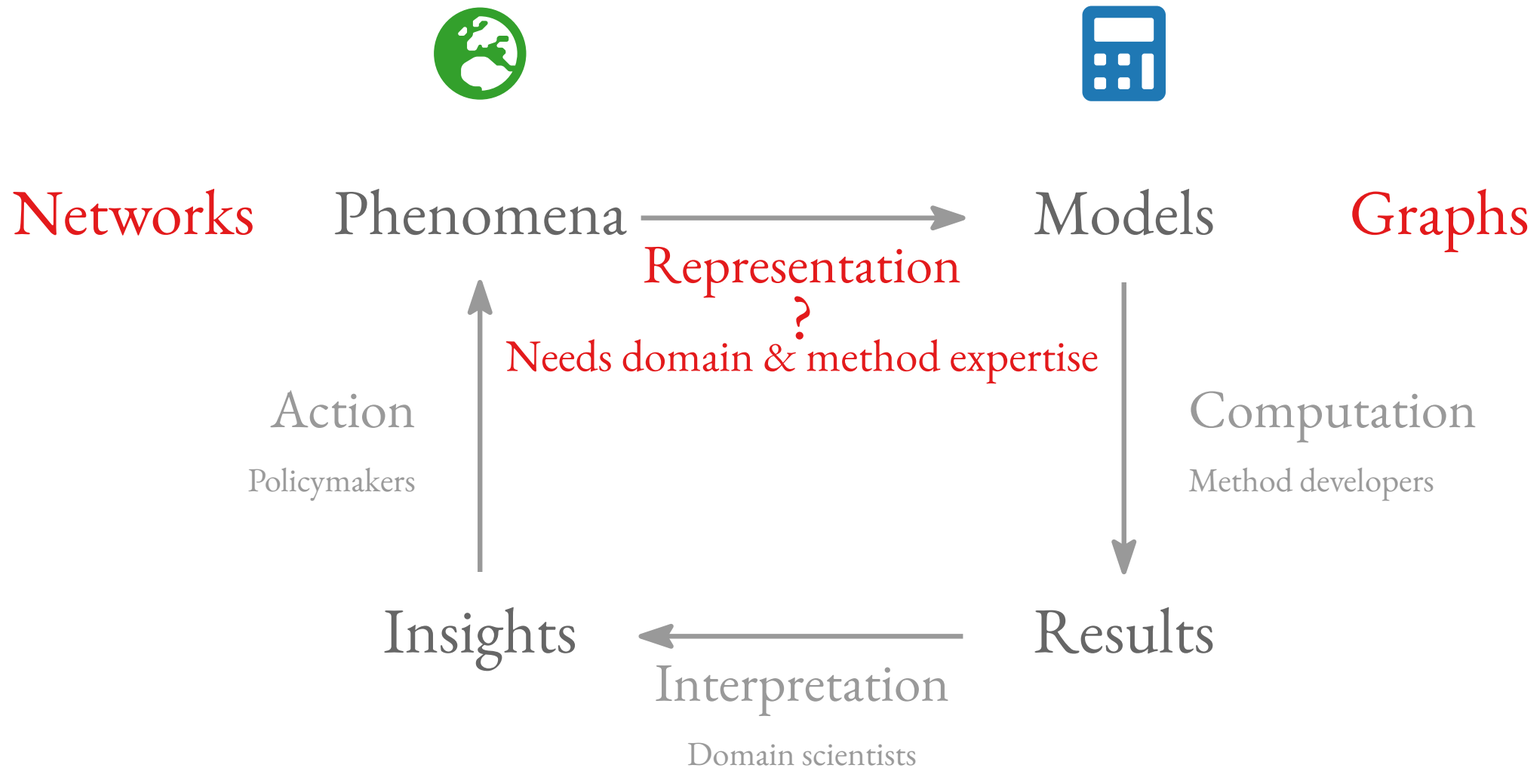
Motivation


Phenomena
Models
Representation
?
Action
Policymakers
Computation
Method developers
Insights
Results
Interpretation
Domain scientists

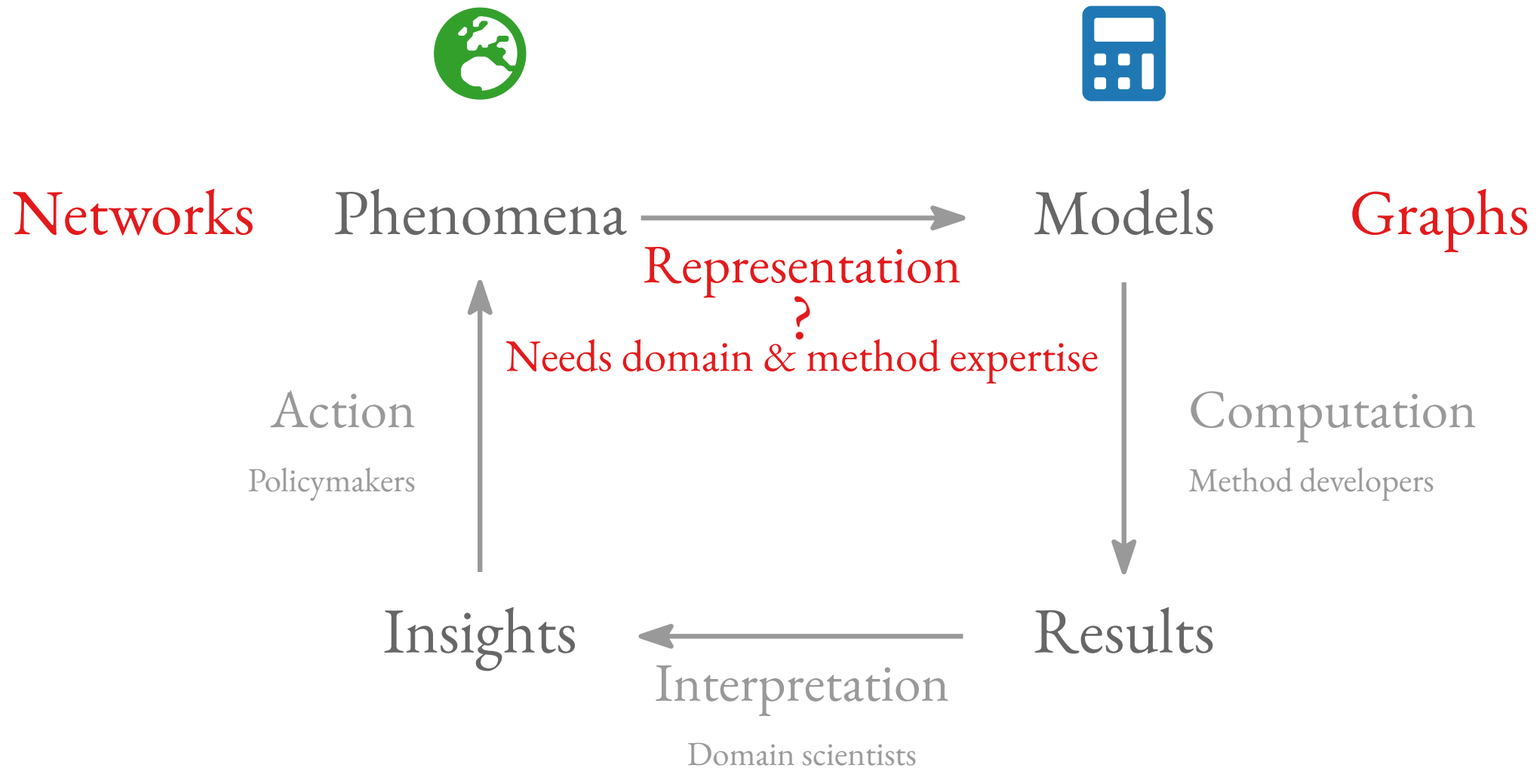
Motivation


Phenomena
Models
Representation
?
Needs domain & method expertise
Action
Policymakers
Computation
Method developers
Insights
Results
Interpretation
Domain scientists

Motivation


Networks
Phenomena
Models
Graphs
Representation
?
Needs domain & method expertise
Action
Policymakers
Computation
Method developers
Insights
Results
Interpretation
Domain scientists

Motivation



How do graph-data-modeling decisions impact our results?

Contribution

Contribution

Contribution

37 Shakespeare plays

18 (Hyper)graph representations

666 (Hyper)graphs in total

Contribution

37 Shakespeare plays

18 (Hyper)graph representations

666 (Hyper)graphs in total

Present datasets as (sets of) *transformations*

Make *data-modeling decisions* transparent

Enable *representation robustness checks*

Contribution

37 Shakespeare plays

18 (Hyper)graph representations

666 (Hyper)graphs in total

Present datasets as (sets of) *transformations*

Make *data-modeling decisions* transparent

Enable *representation robustness checks*

Shakespeare-style drama

Critique academic (data) *culture*

Contribution

37 Shakespeare plays
18 (Hyper)graph representations
666 (Hyper)graphs in total

Present datasets as (sets of) *transformations*
Make *data-modeling decisions* transparent
Enable *representation robustness checks*

Shakespeare-style drama

Critique academic (data) *culture*

INDUCTION.

SCENE I.—*Between submission and decision.*

Enter REVIEWER *and* AUTHORS.

Rev. What is this? Is this not against the rules?
Auth. The columns? These are only simple tables.
They serve to help us implement blank verse.
The script-sized numbers count the spoken lines,
They disappear when folks use prose at times.
We introduce a novel dataset,
With full documentation as Appendix.
Raw data stem from all of Shakespeare's plays [14],
We model them as graphs in many ways,
And demonstrate representations matter.
The data readily accessible [6],
All code is publicly available [7].
What follows, to avoid redundancy,
Conveys our main ideas, as you will see
A tragedy in the Community.

Hyperbard: Dataset

Hyperbard: Dataset

Raw Data: Folger Shakespeare TEI Simple

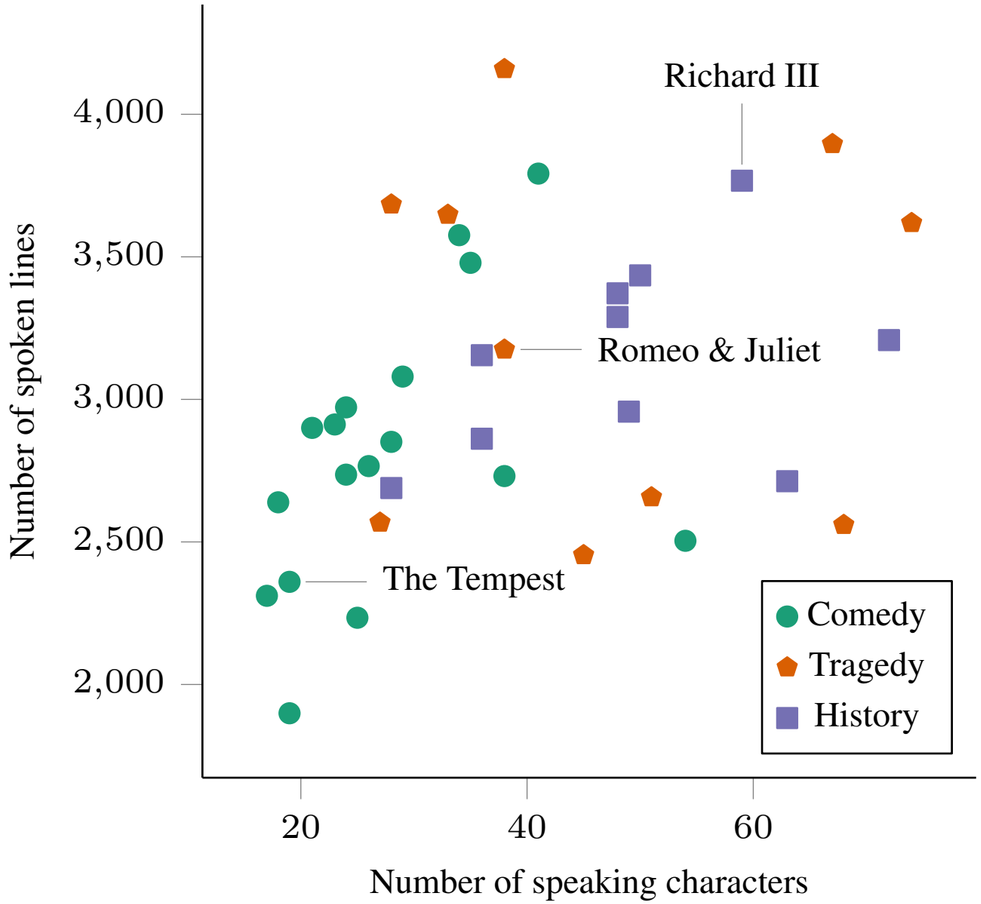
```
...
<sp xml:id="sp-0015" who="#SERVANTS.CAPULET.Sampson_Rom">
<speaker xml:id="spk-0015">
<w xml:id="fs-rom-0002610">SAMPSON</w>
</speaker>
<p xml:id="p-0015">
<lb xml:id="ftln-0015" n="1.1.1"/>
<w xml:id="fs-rom-0002620" n="1.1.1" lemma="Gregory" ana="#n1-nn">G
<pc xml:id="fs-rom-0002630" n="1.1.1">,</pc>
<c> </c>
<w xml:id="fs-rom-0002650" n="1.1.1" lemma="on" ana="#acp-p">on</w>
<c> </c>
<w xml:id="fs-rom-0002670" n="1.1.1" lemma="my" ana="#po">my</w>
<c> </c>
<w xml:id="fs-rom-0002690" n="1.1.1" lemma="word" ana="#n1">word</w
<c> </c>
<w xml:id="fs-rom-0002710" n="1.1.1" lemma="we|will" ana="#pns|vmb"
<c> </c>
<w xml:id="fs-rom-0002730" n="1.1.1" lemma="not" ana="#xx">not</w>
<c> </c>
<w xml:id="fs-rom-0002750" n="1.1.1" lemma="carry" ana="#vvi">carry
<c> </c>
<w xml:id="fs-rom-0002770" n="1.1.1" lemma="coal" ana="#n2">coals</
<pc xml:id="fs-rom-0002780" n="1.1.1">.</pc>
</p>
</sp>
...
```

Hyperbard: Dataset

Raw Data: Folger Shakespeare TEI Simple

```
...
<sp xml:id="sp-0015" who="#SERVANTS.CAPULET.Sampson_Rom">
<speaker xml:id="spk-0015">
<w xml:id="fs-rom-0002610">SAMPSON</w>
</speaker>
<p xml:id="p-0015">
<lb xml:id="ftln-0015" n="1.1.1"/>
<w xml:id="fs-rom-0002620" n="1.1.1" lemma="Gregory" ana="#n1-nn">G
<pc xml:id="fs-rom-0002630" n="1.1.1">,</pc>
<c> </c>
<w xml:id="fs-rom-0002650" n="1.1.1" lemma="on" ana="#acp-p">on</w>
<c> </c>
<w xml:id="fs-rom-0002670" n="1.1.1" lemma="my" ana="#po">my</w>
<c> </c>
<w xml:id="fs-rom-0002690" n="1.1.1" lemma="word" ana="#n1">word</w
<c> </c>
<w xml:id="fs-rom-0002710" n="1.1.1" lemma="we|will" ana="#pns|vmb"
<c> </c>
<w xml:id="fs-rom-0002730" n="1.1.1" lemma="not" ana="#xx">not</w>
<c> </c>
<w xml:id="fs-rom-0002750" n="1.1.1" lemma="carry" ana="#vvi">carry
<c> </c>
<w xml:id="fs-rom-0002770" n="1.1.1" lemma="coal" ana="#n2">coals</
<pc xml:id="fs-rom-0002780" n="1.1.1">.</pc>
</p>
</sp>
...
```

Basic Statistics

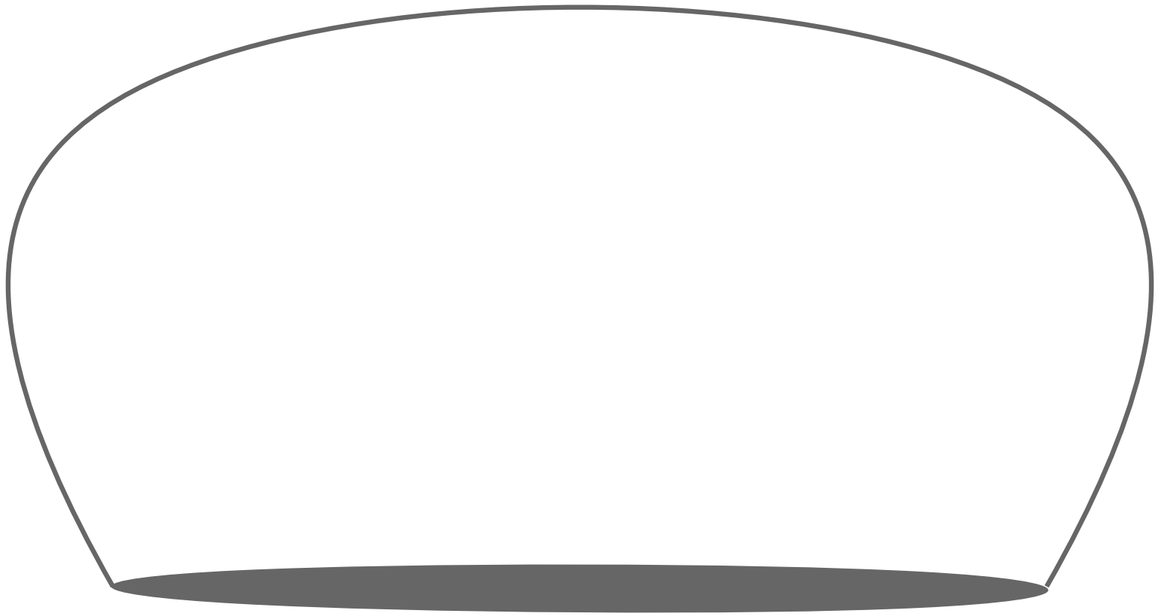


Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?

Hyperbard: Representations

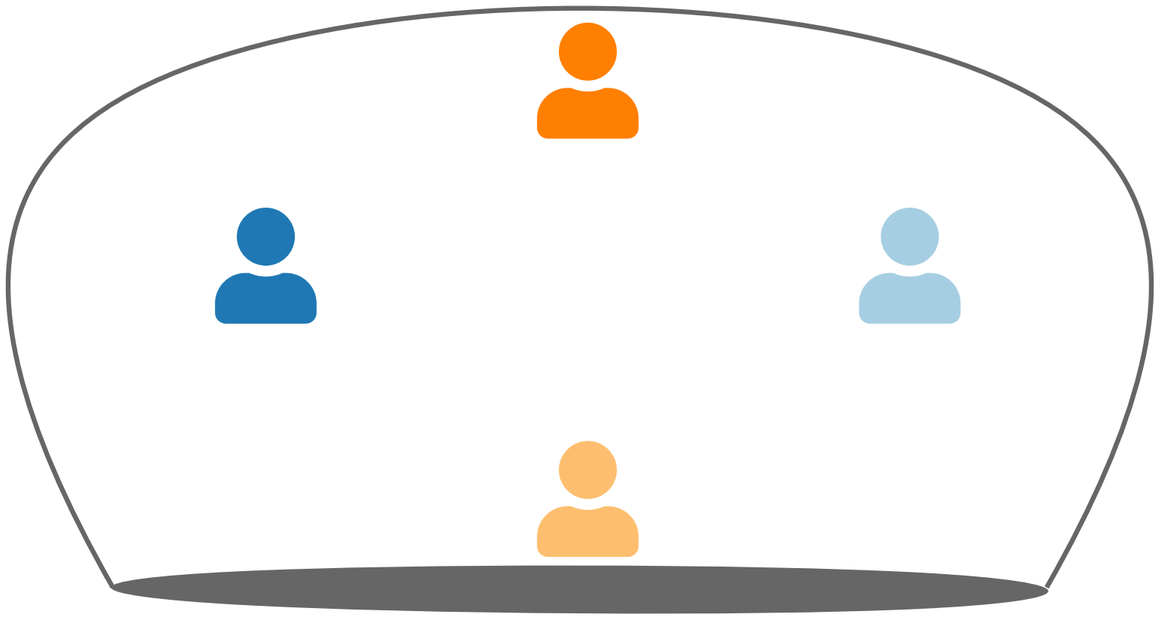
When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



Stage

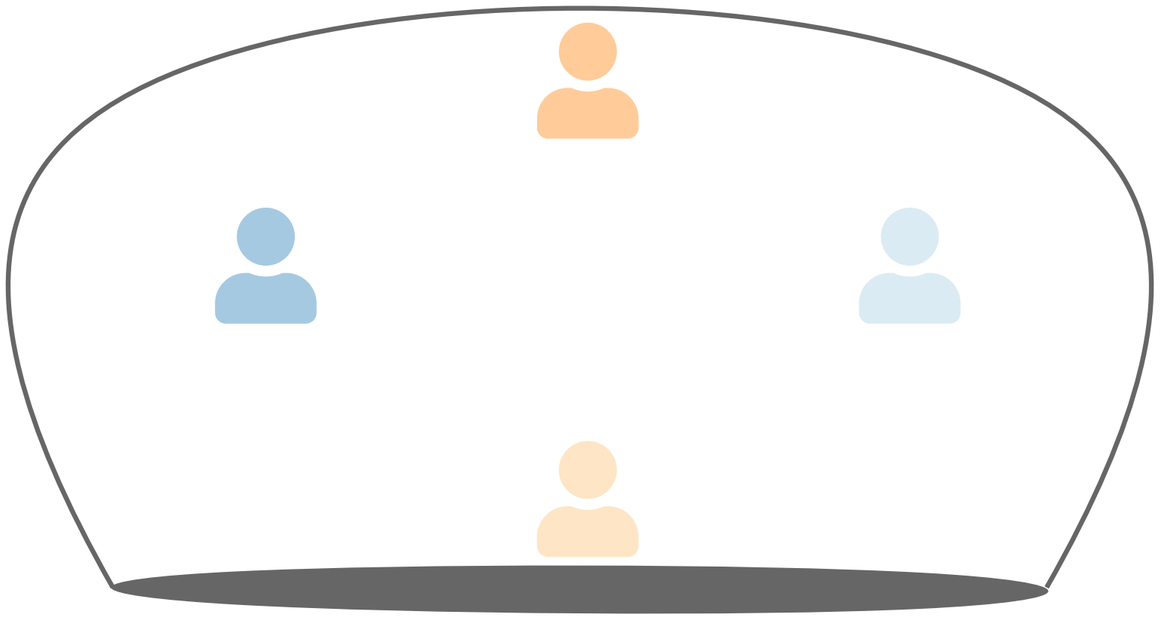
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



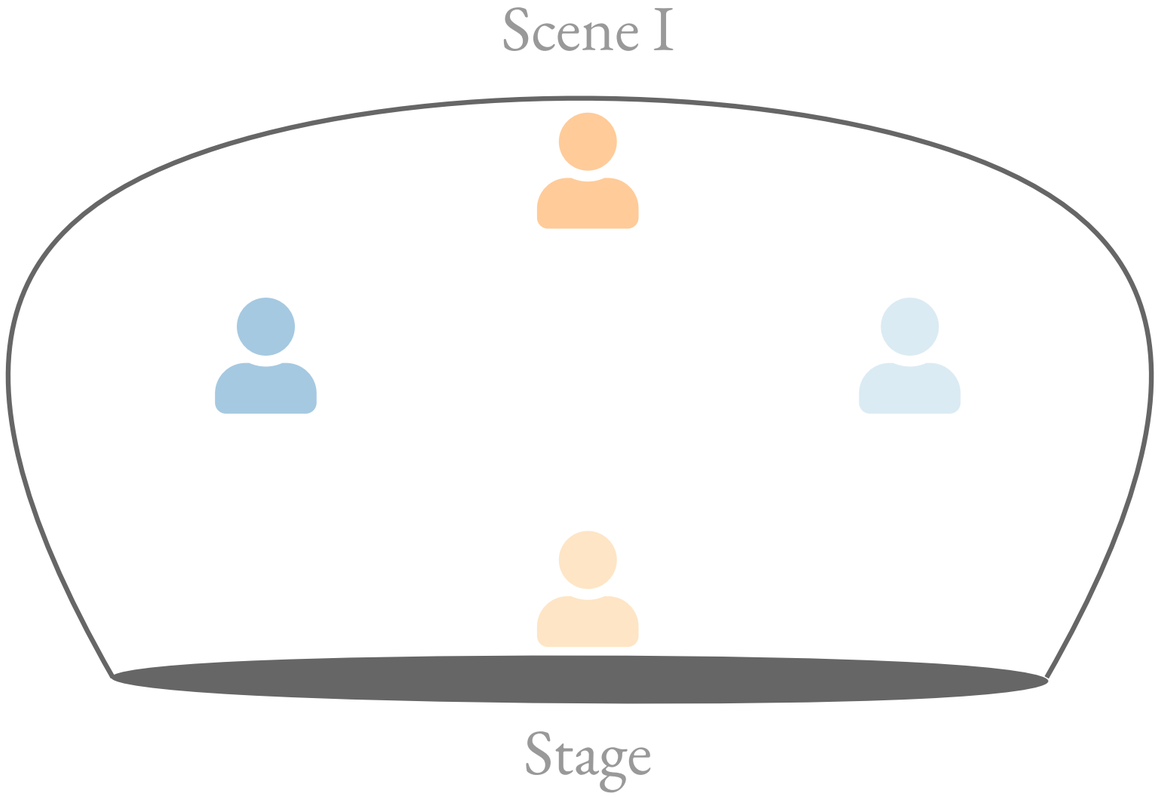
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



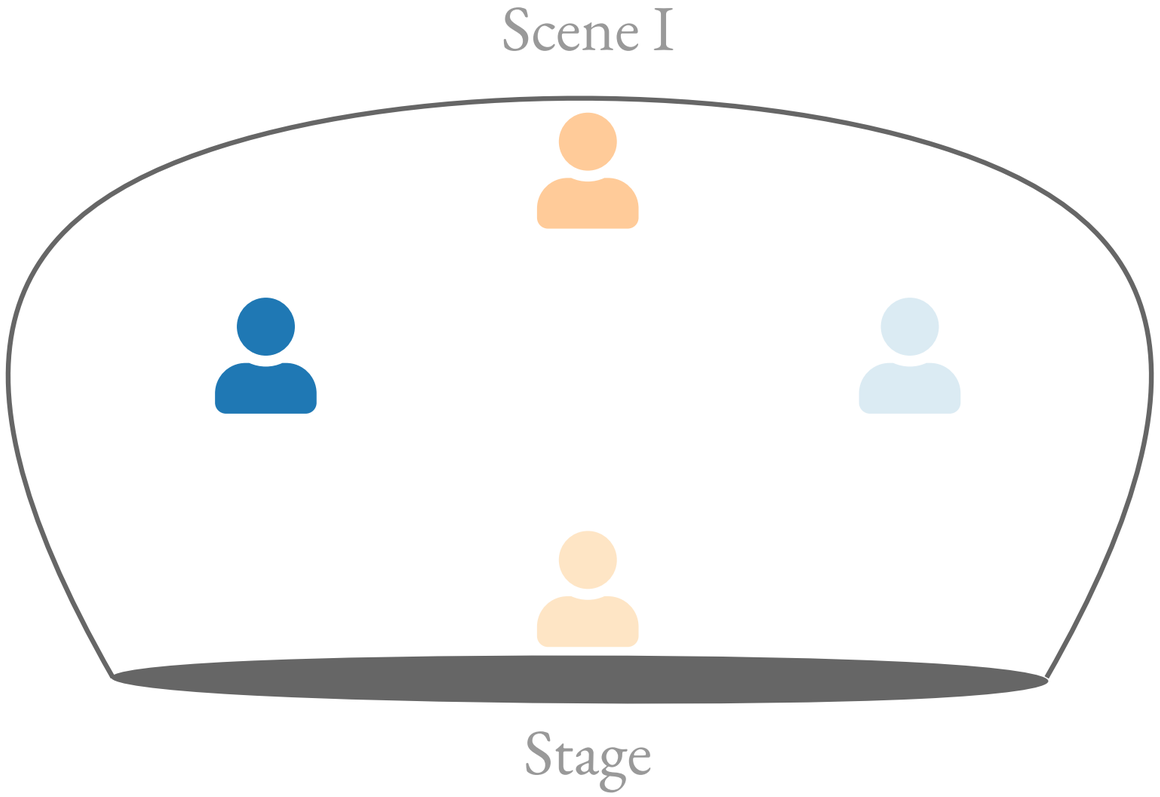
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



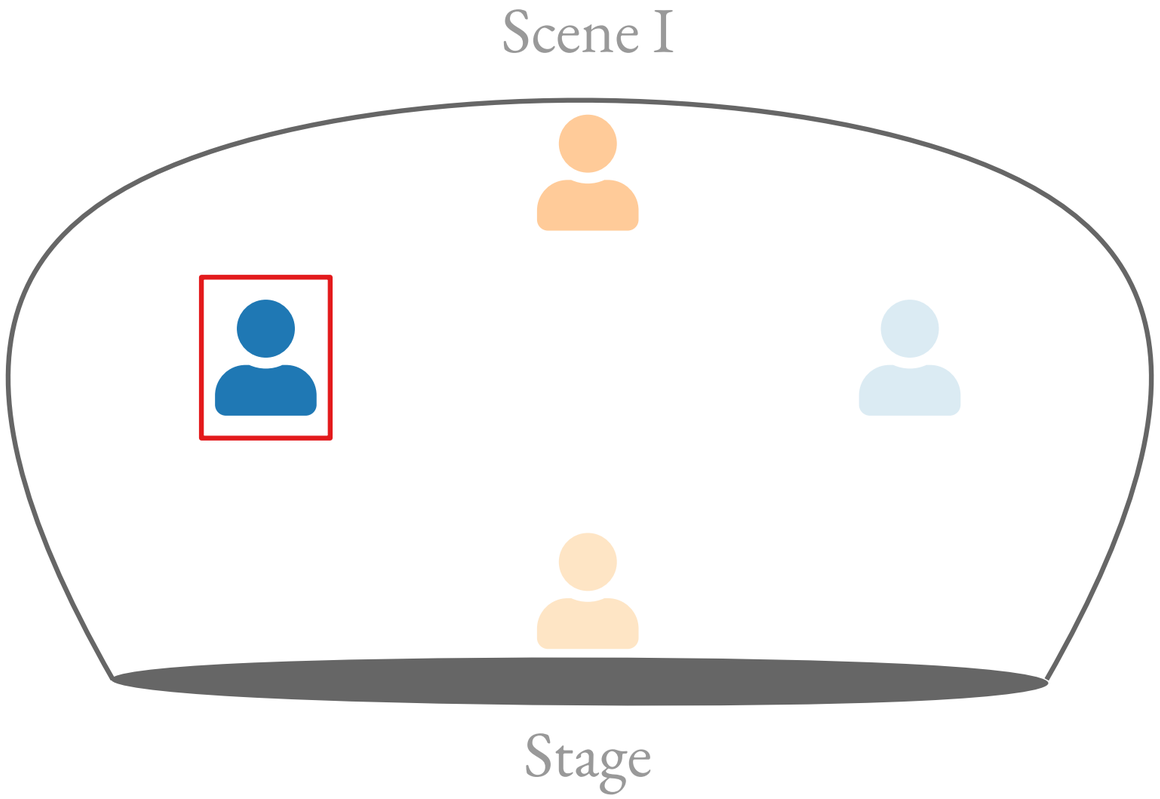
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



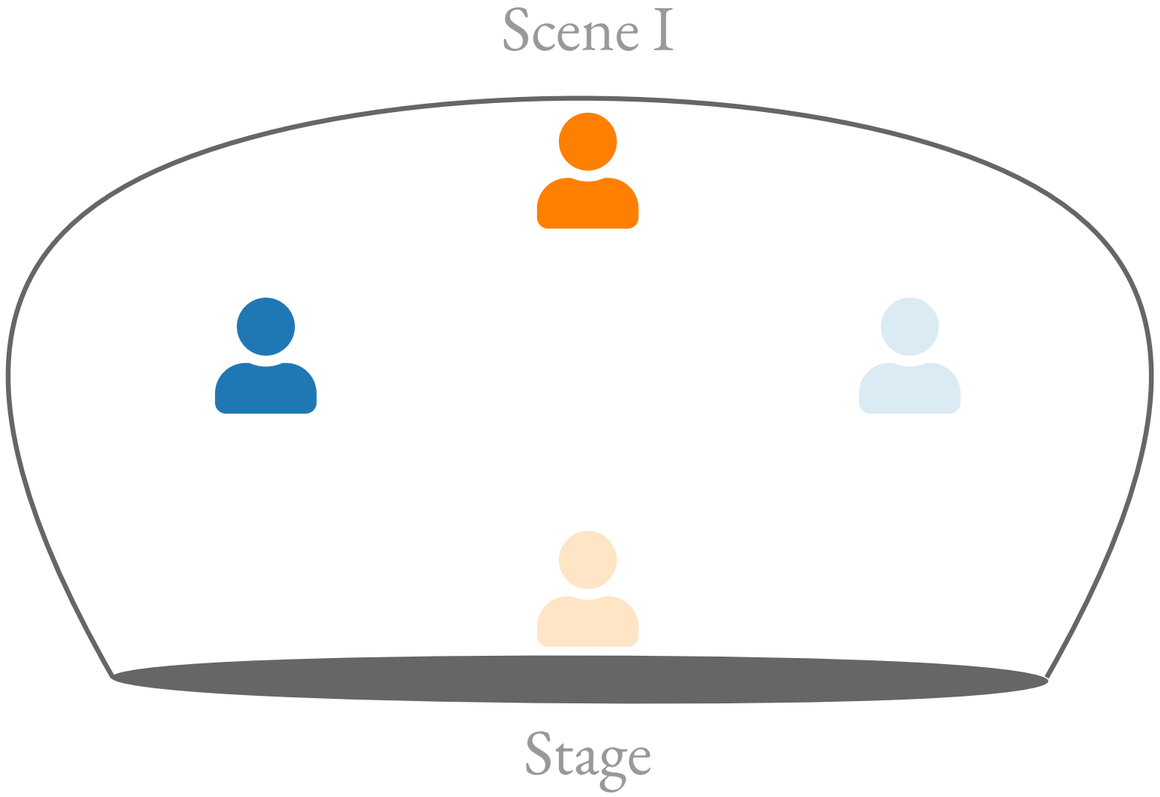
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



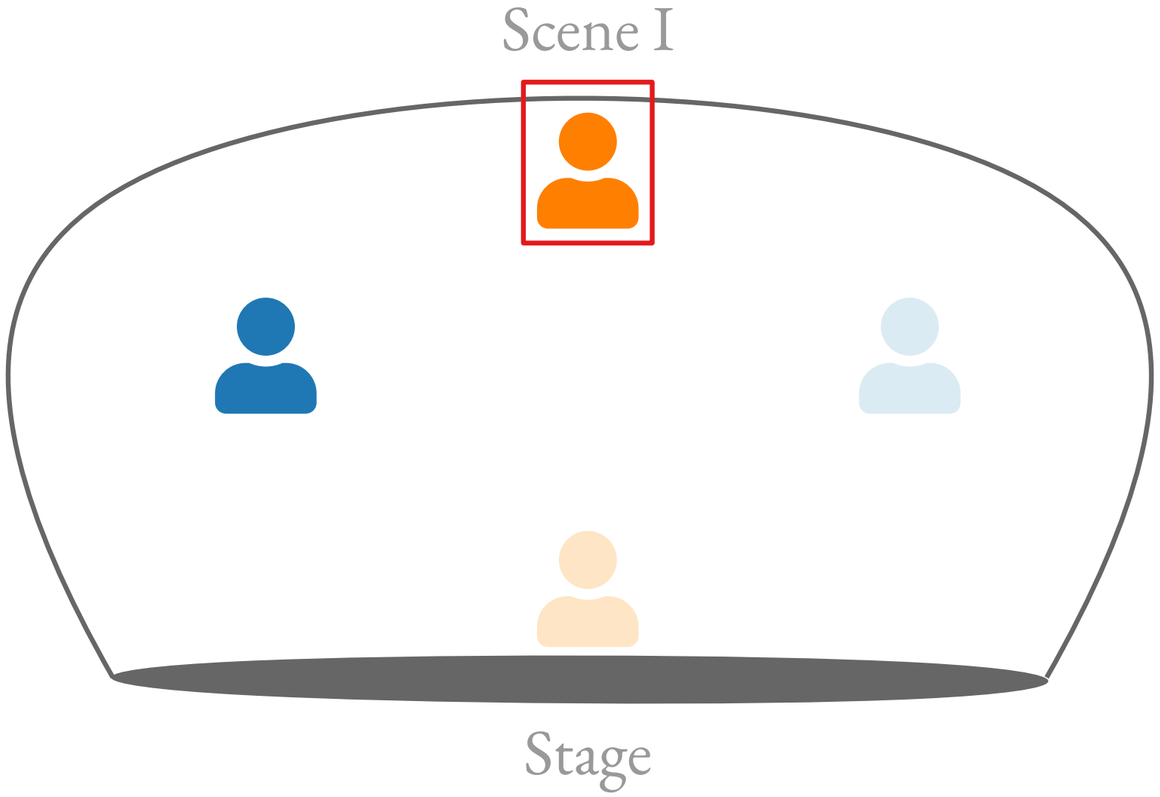
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



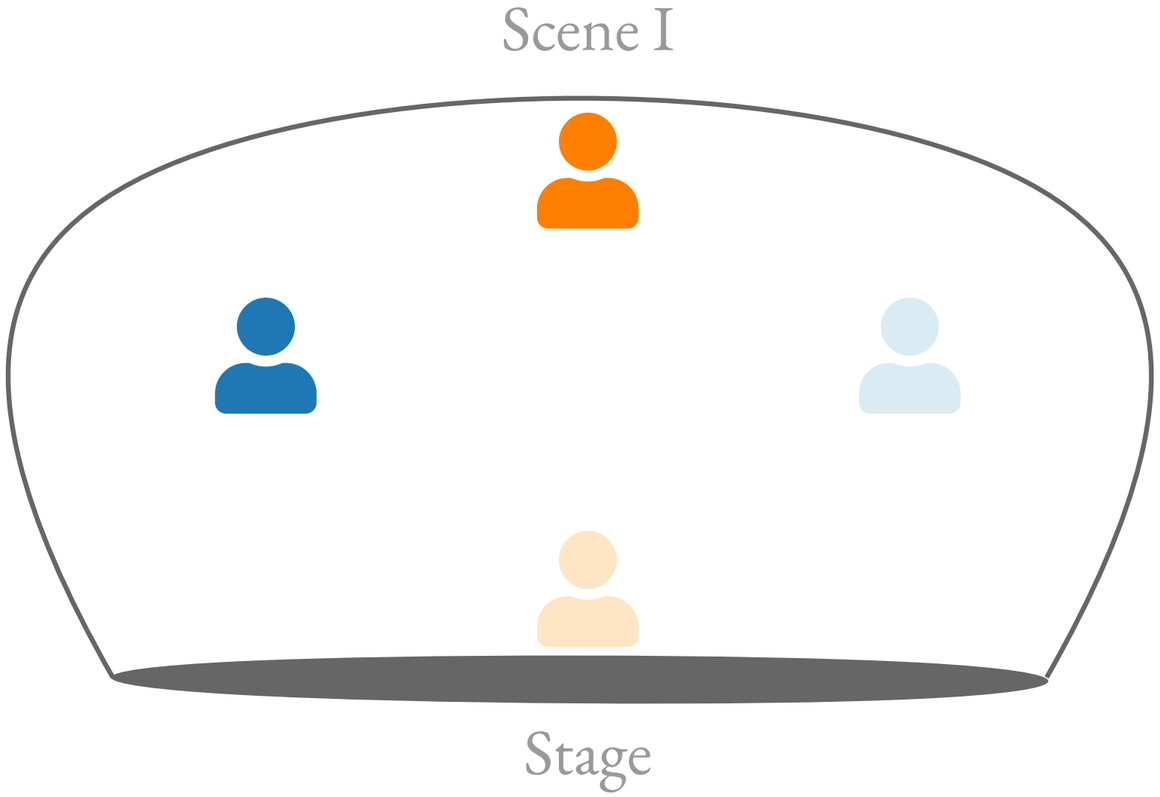
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



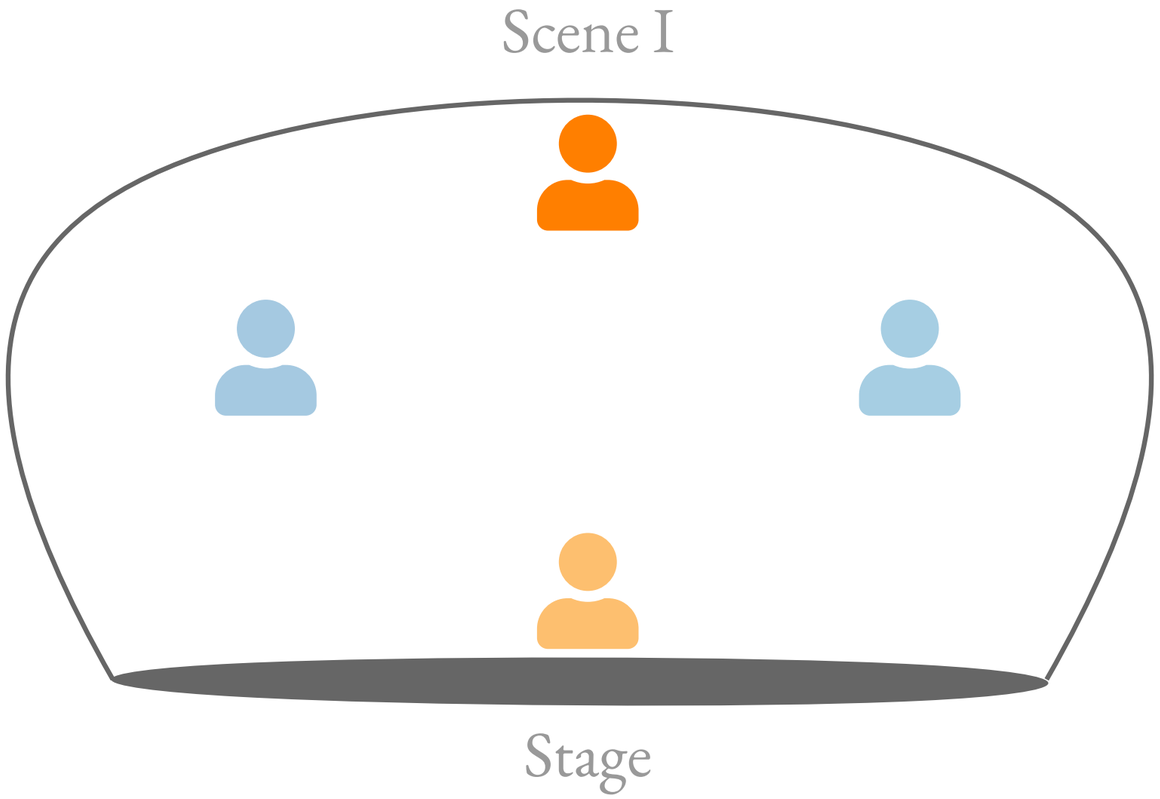
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



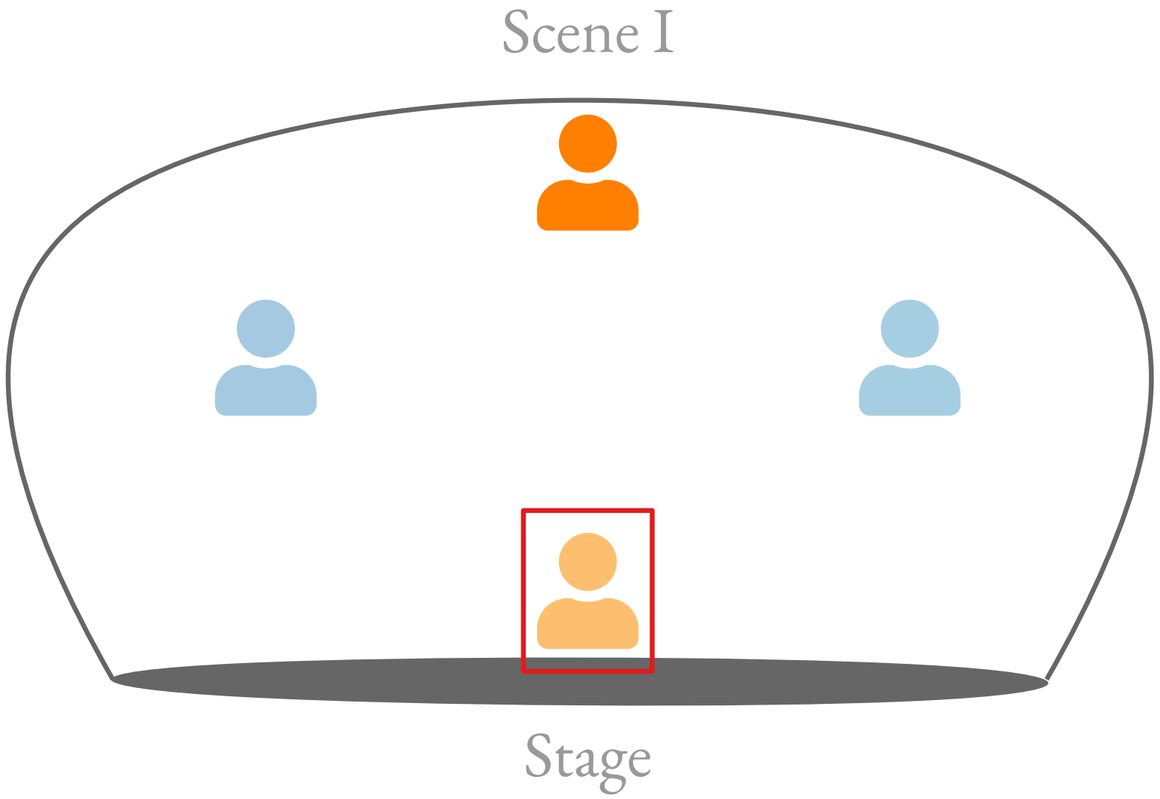
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



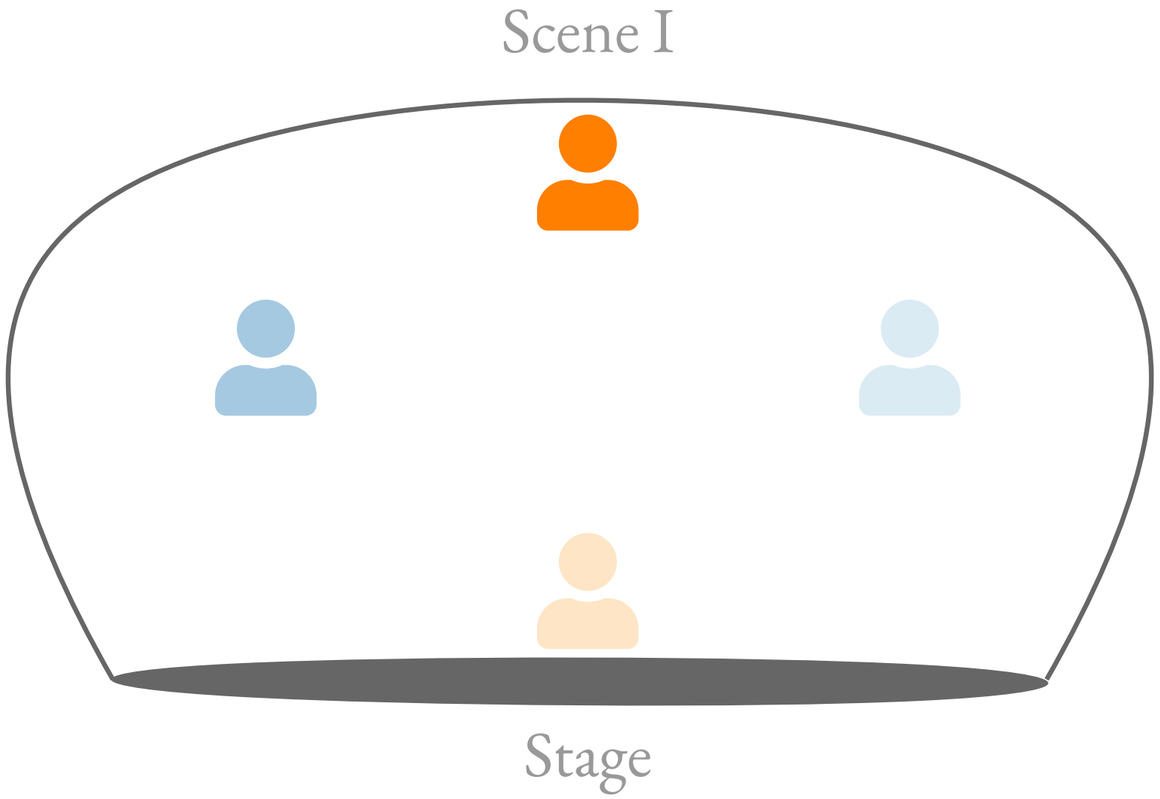
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



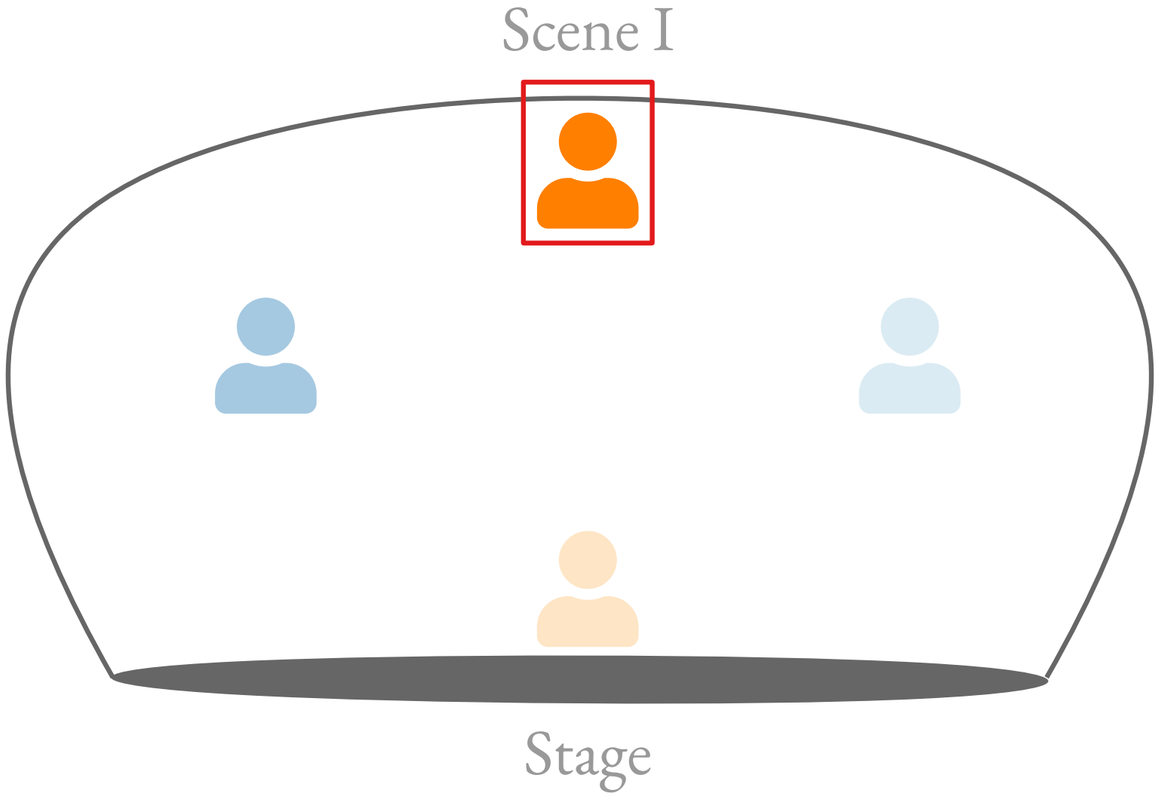
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



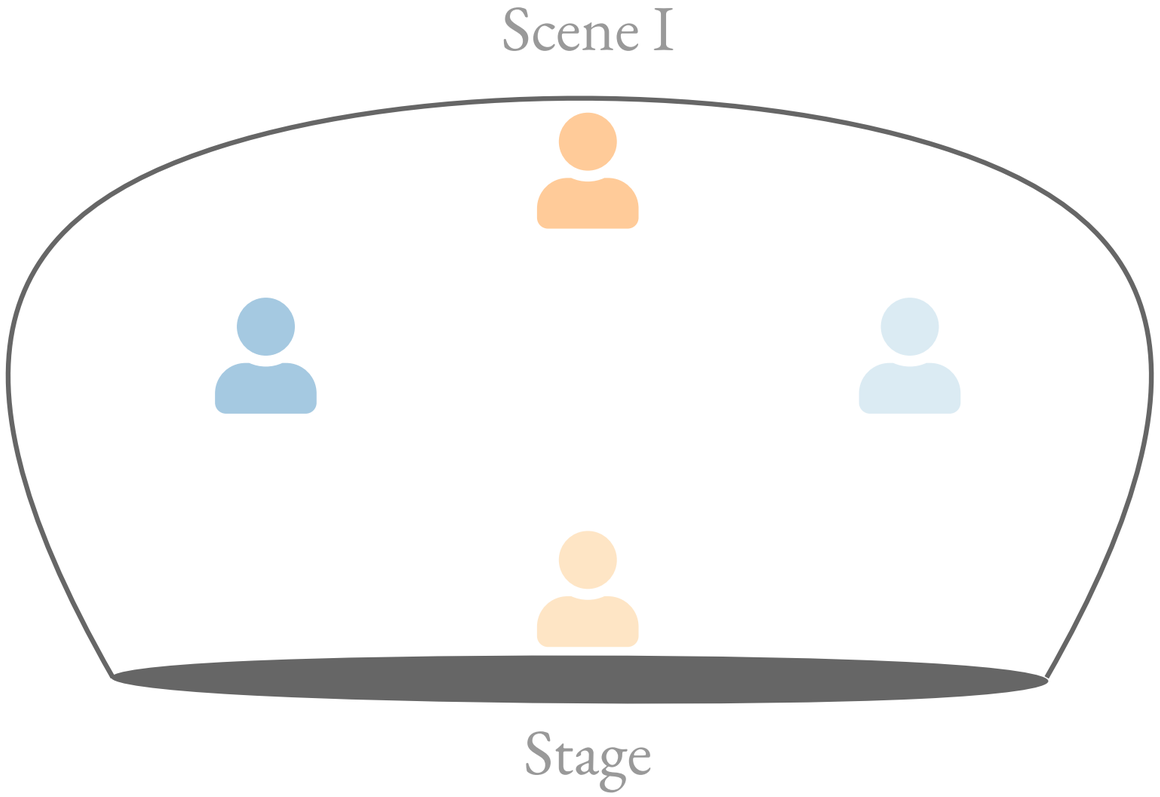
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



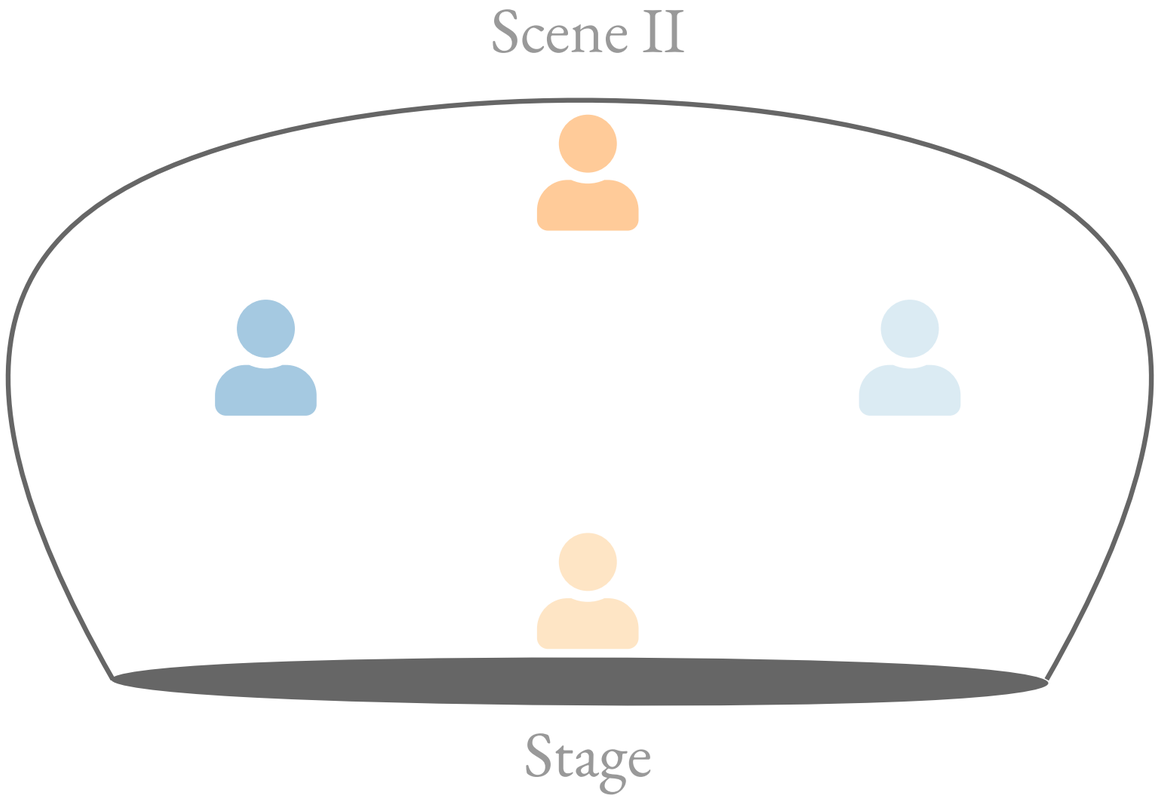
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



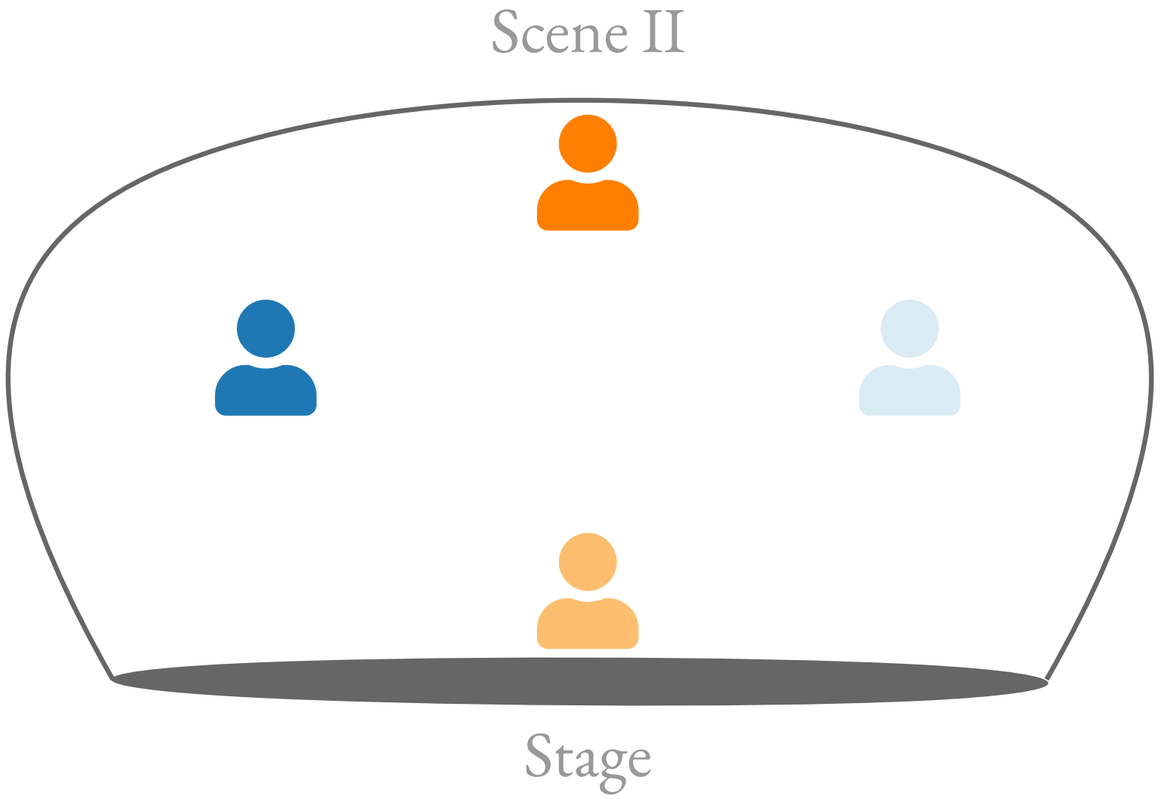
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



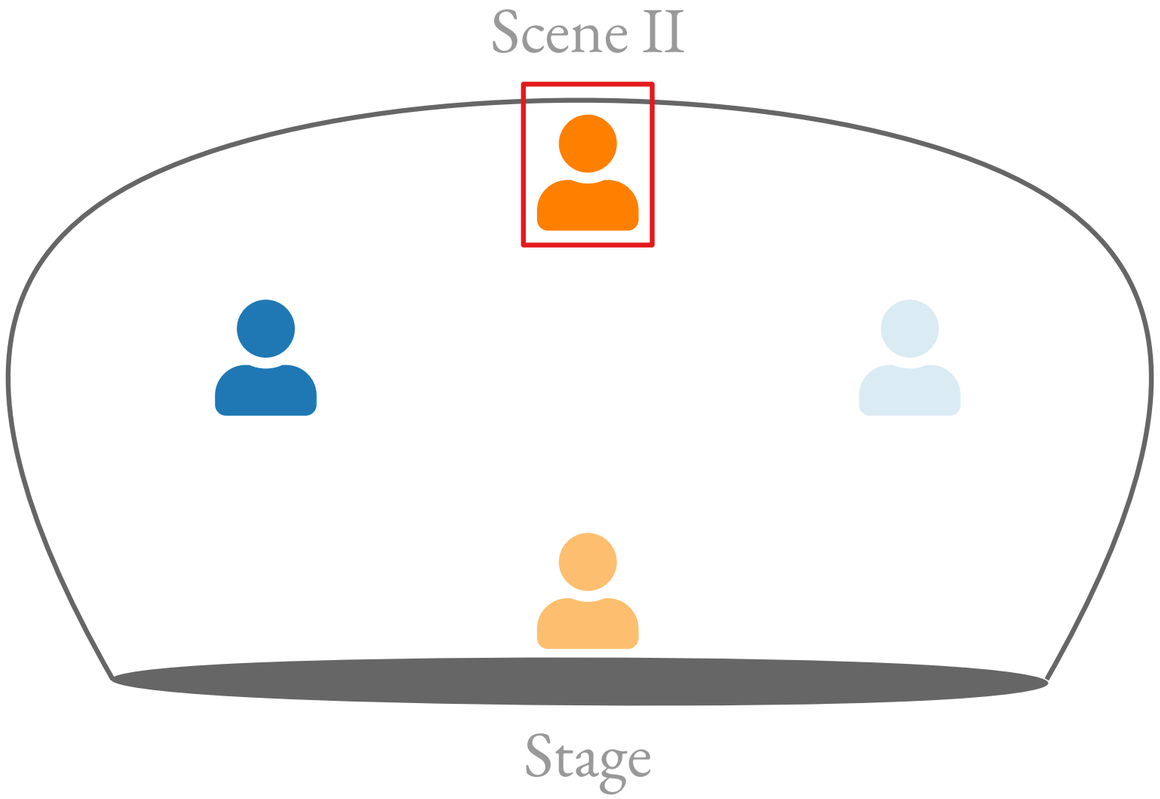
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



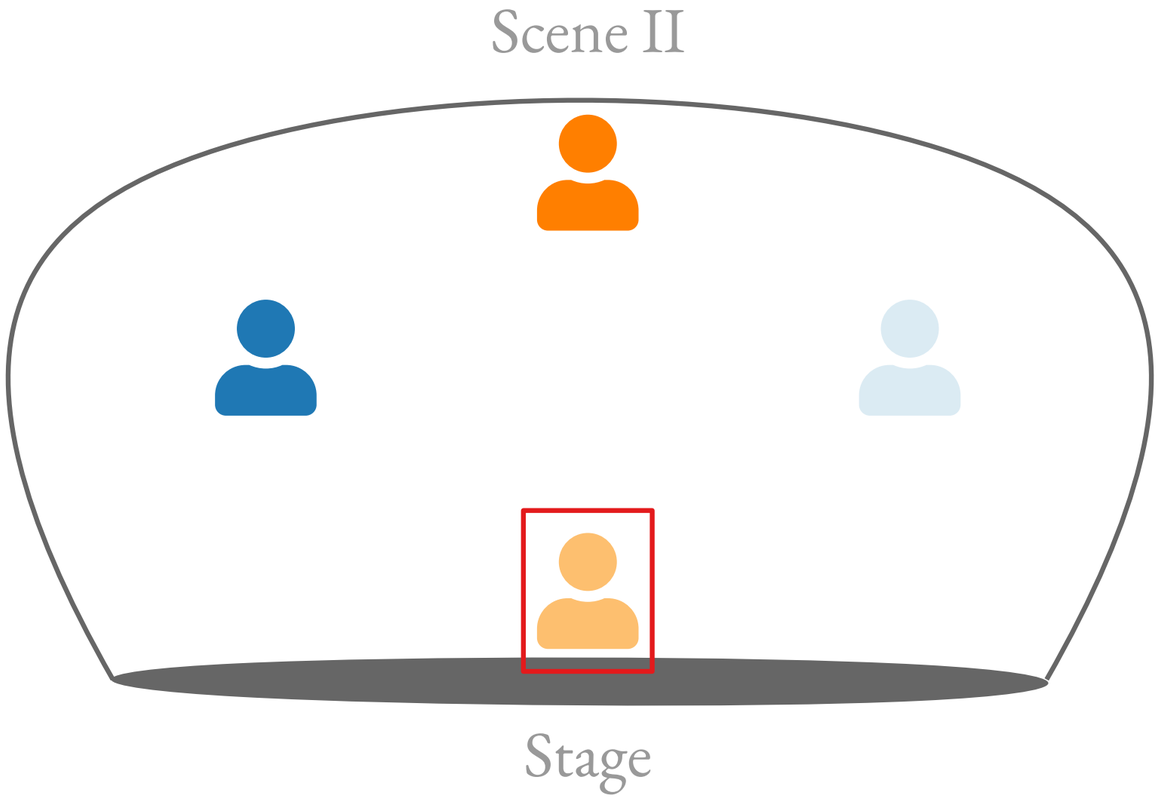
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



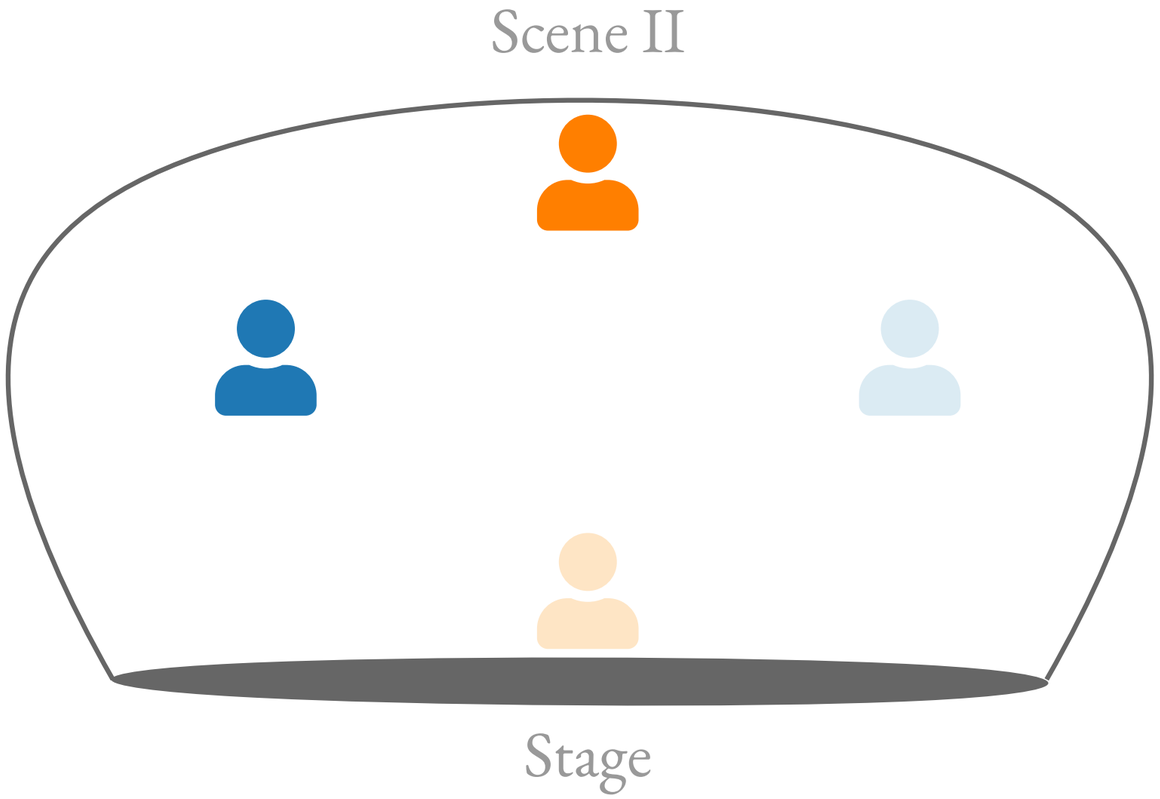
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



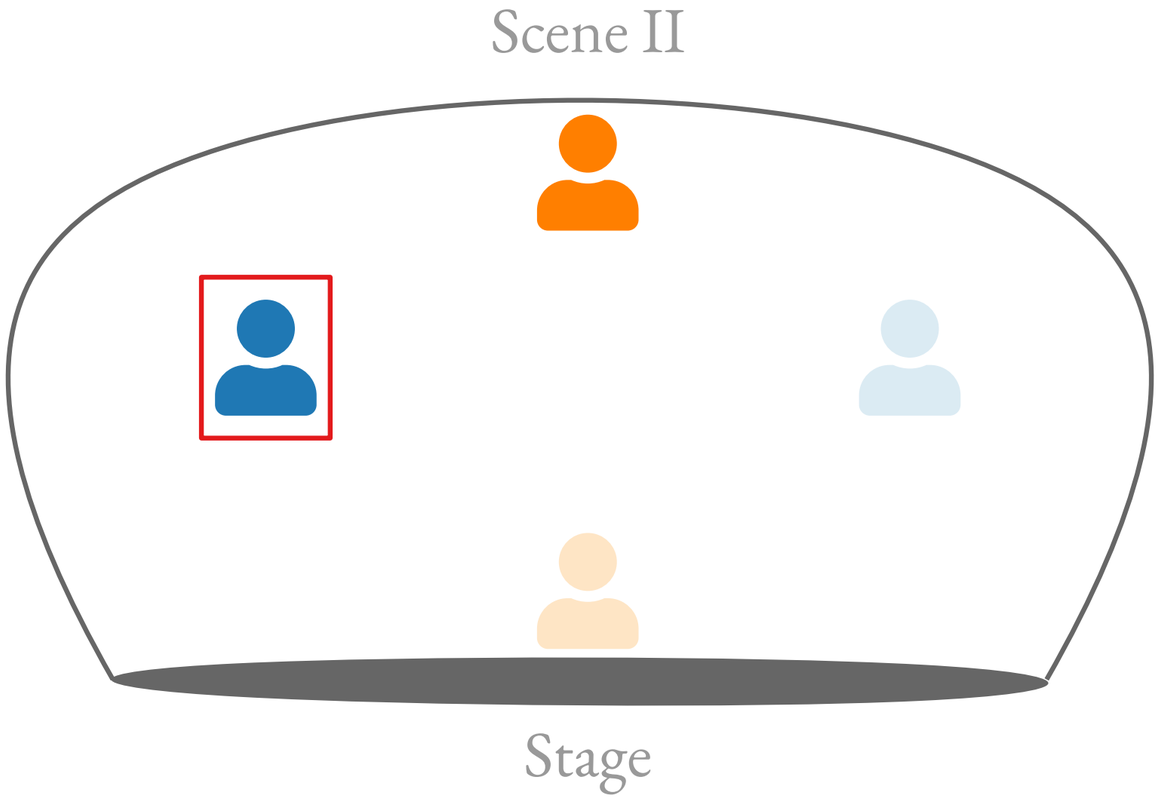
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



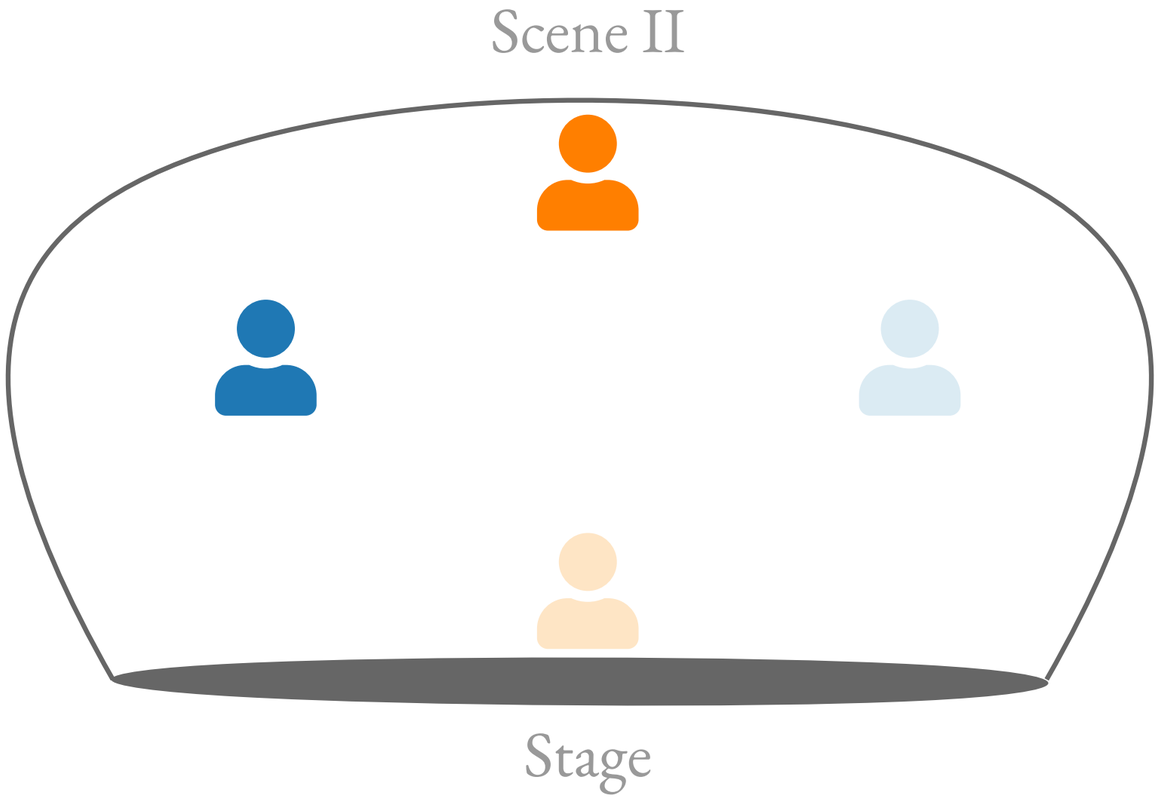
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



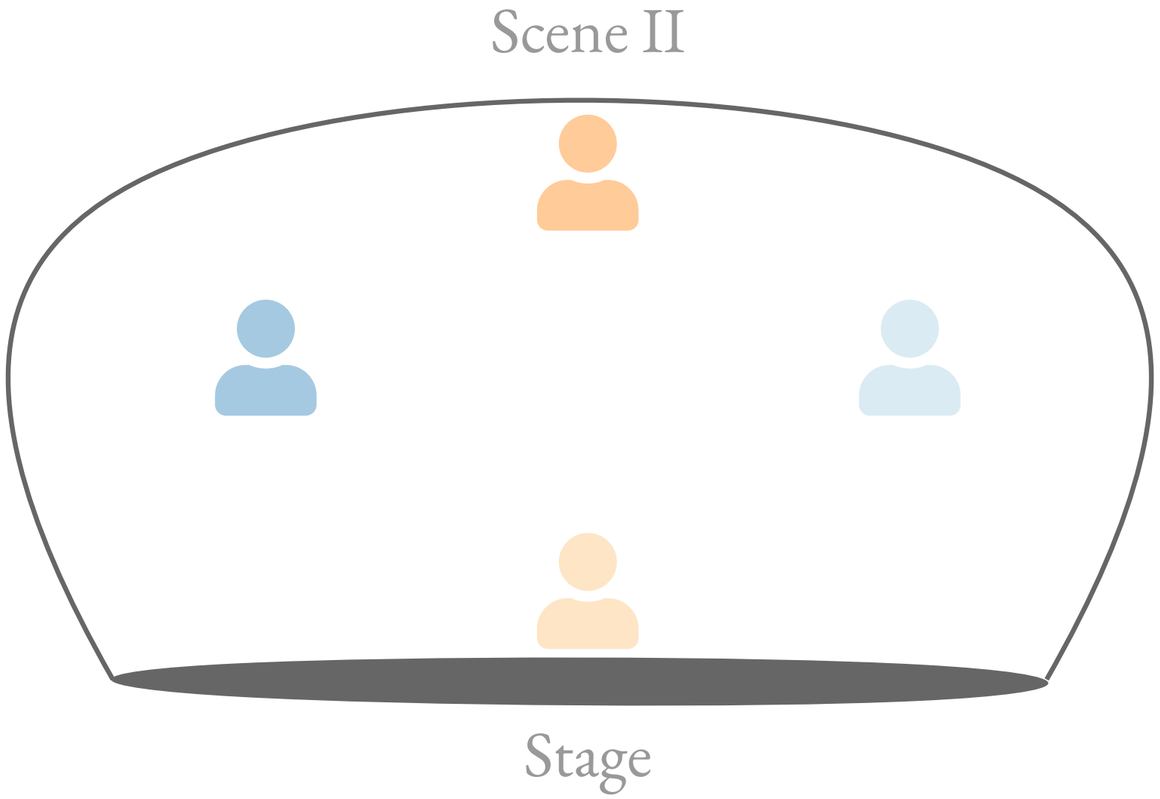
Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



Hyperbard: Representations

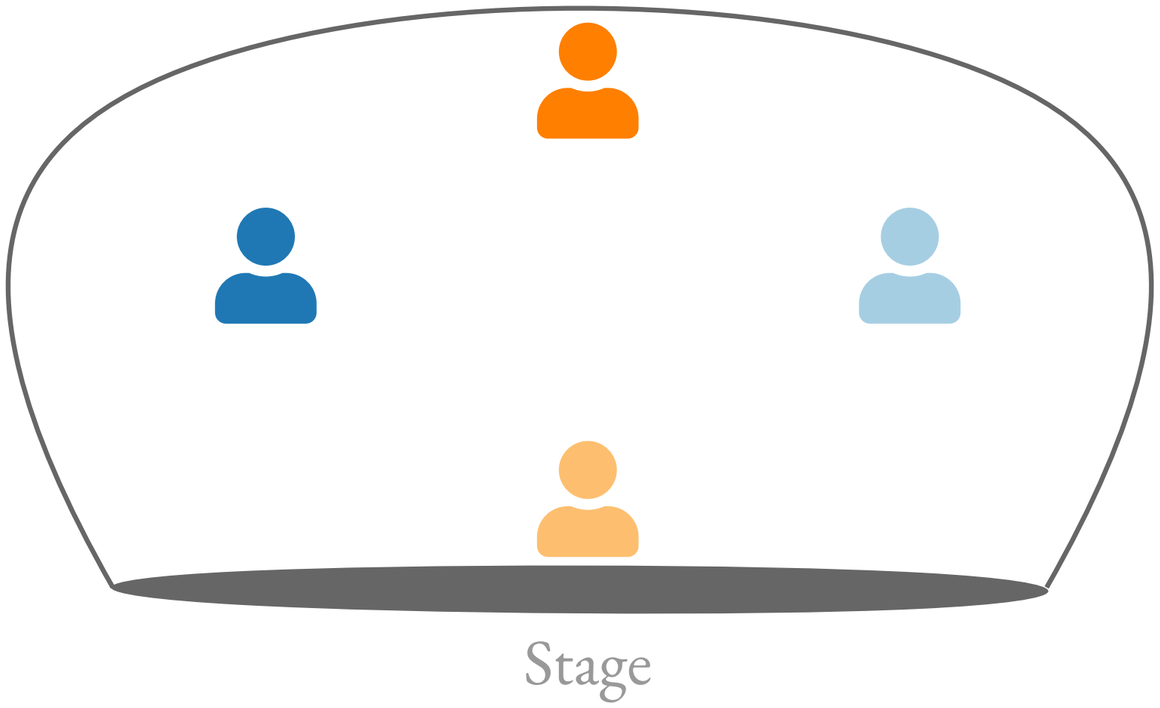
When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?



Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?

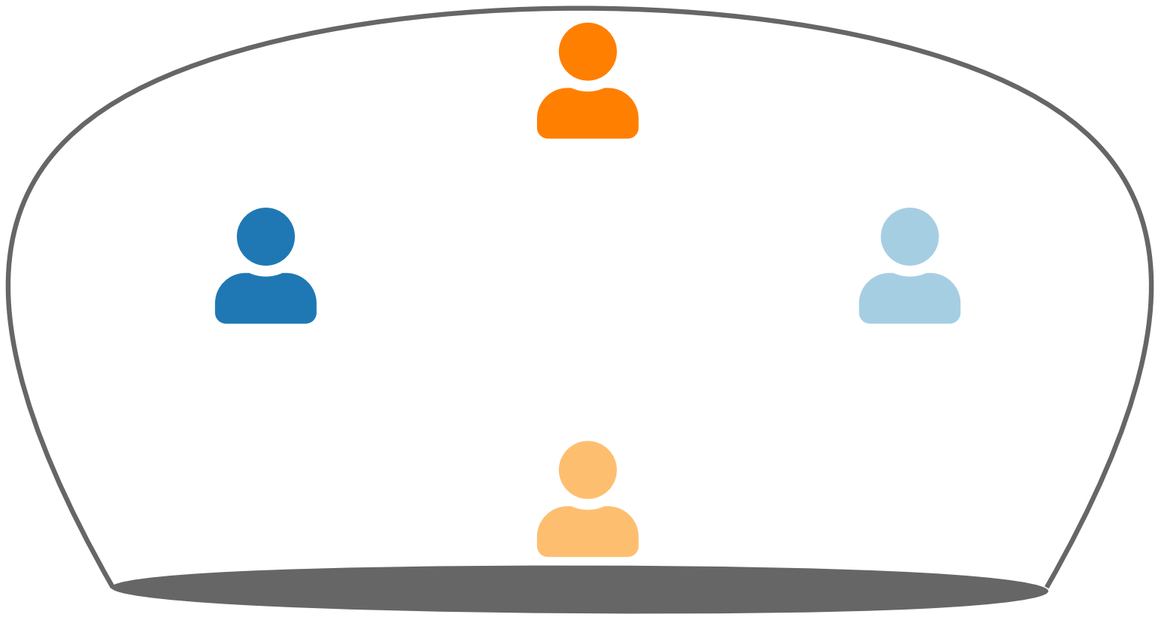
How can we construct a graph from this?



Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?

How can we construct a graph from this?



Stage

Classic Approach

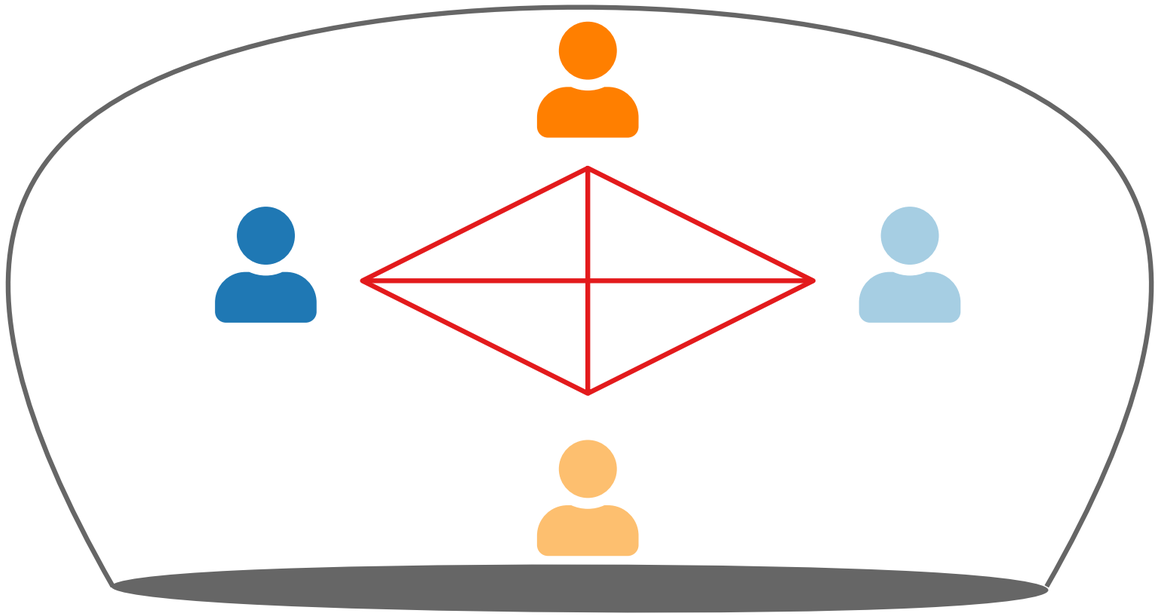
Nodes ← Characters

Edges ← Co-Occurrence in Scene

Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?

How can we construct a graph from this?



Stage

Classic Approach

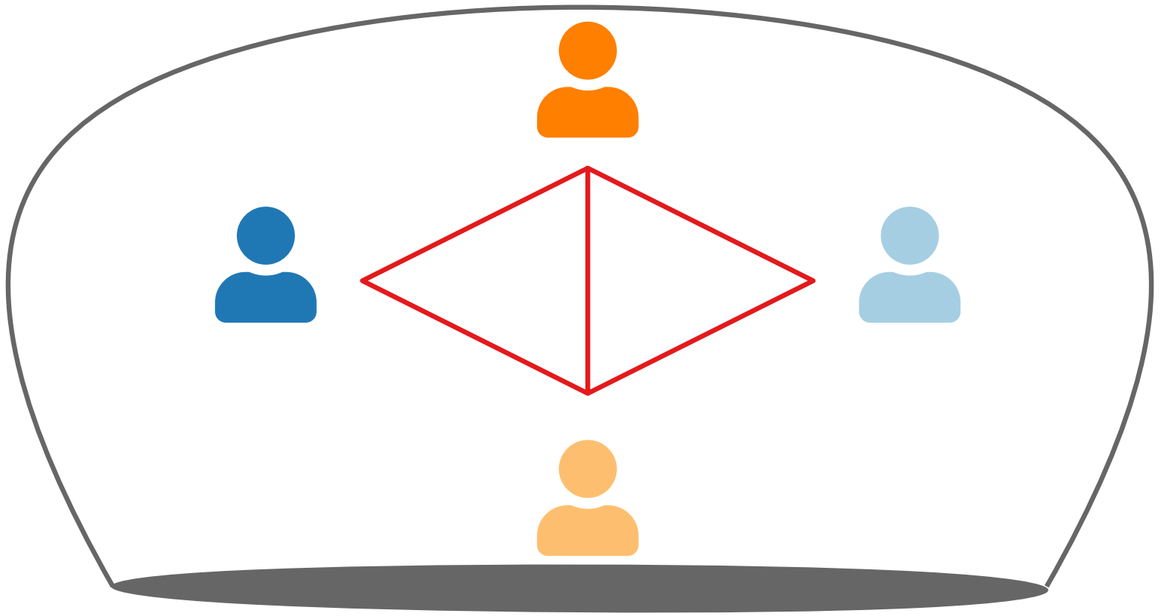
Nodes ← Characters

Edges ← Co-Occurrence in Scene

Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?

How can we construct a graph from this?



Stage

Classic Approach

Nodes ← Characters

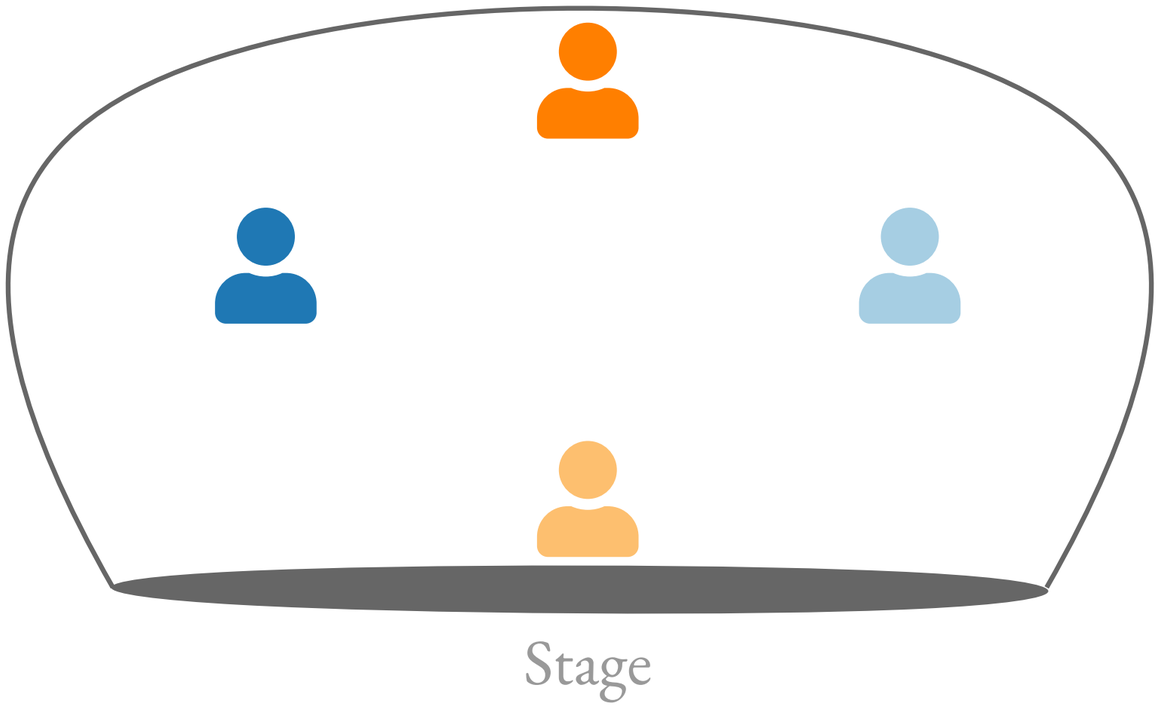
Edges ← Co-Occurrence in Scene

≠ Edges ← Co-Occurrence on Stage!

Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?

How can we construct a graph from this?



Stage

Classic Approach

Nodes ← Characters

Edges ← Co-Occurrence in Scene

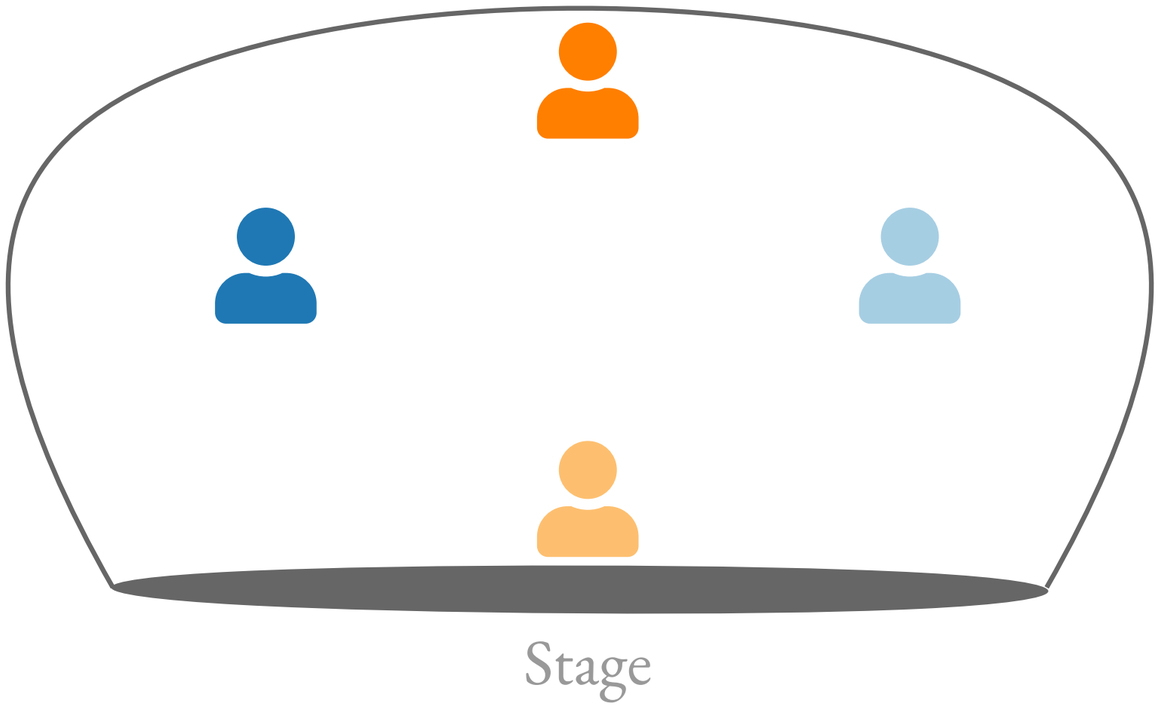
≠ Edges ← Co-Occurrence on Stage!

...and What about...

Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?

How can we construct a graph from this?



Stage

Classic Approach

Nodes ← Characters

Edges ← Co-Occurrence in Scene

≠ Edges ← Co-Occurrence on Stage!

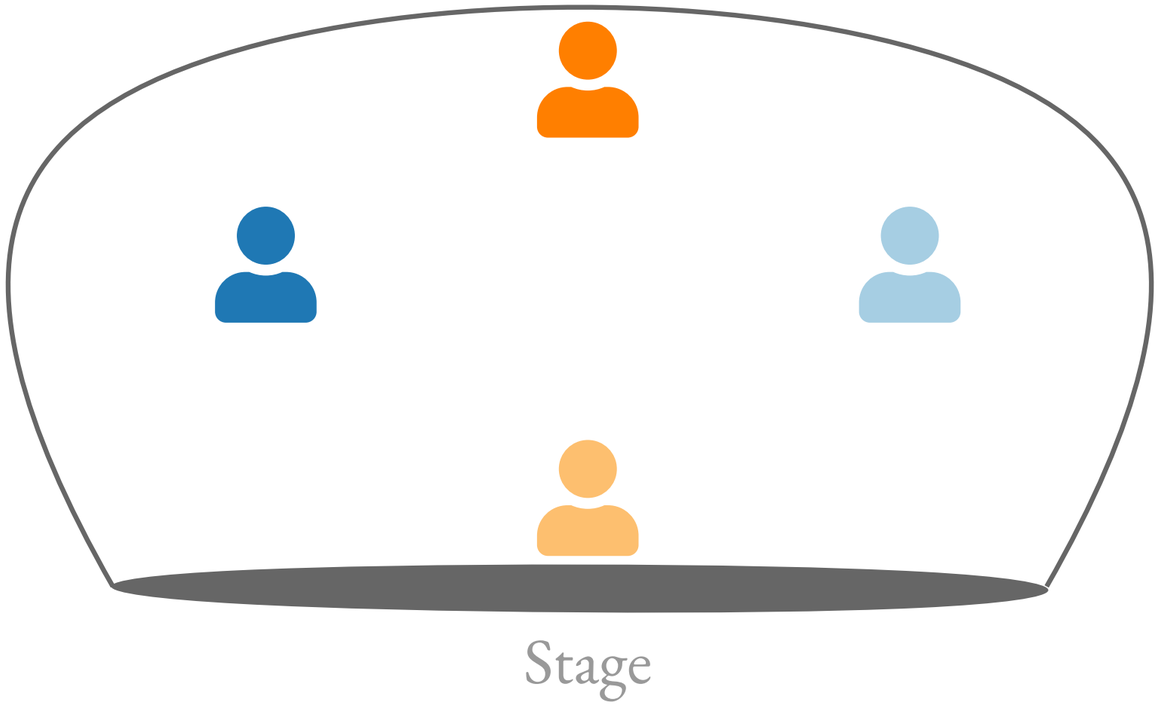
...and What about...

...Frequencies?

Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?

How can we construct a graph from this?



Stage

Classic Approach

Nodes ← Characters

Edges ← Co-Occurrence in Scene

≠ Edges ← Co-Occurrence on Stage!

...and What about...

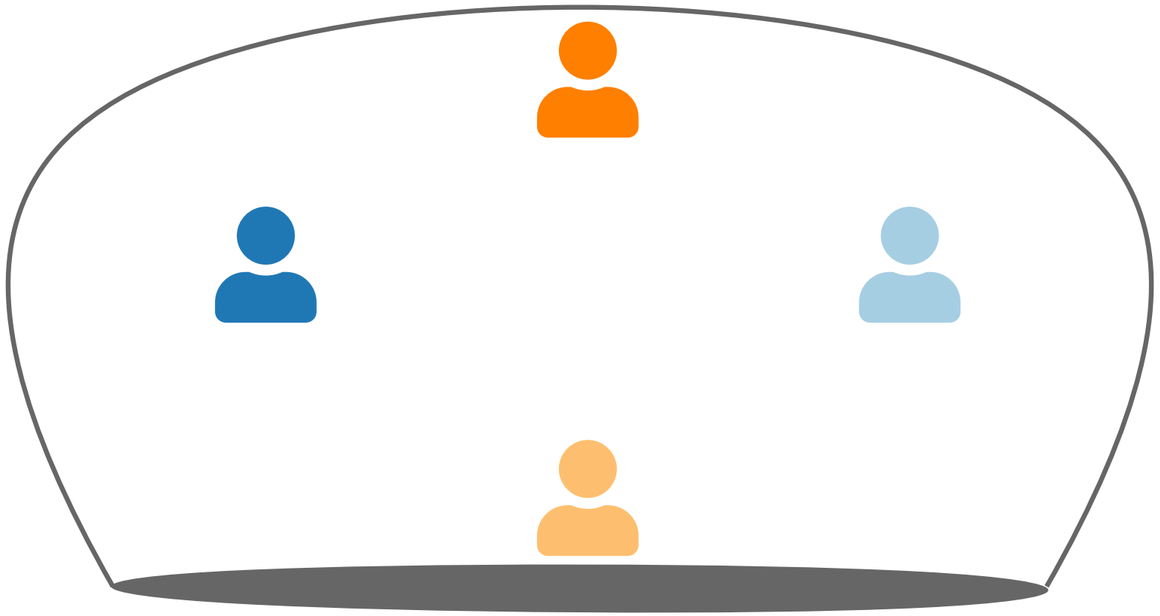
...Frequencies?

...Speech?

Hyperbard: Representations

When we transform reality to math,
Graphs are but outputs, in—phenomena.
The myriad transformations that we see,
How do they differ systematically?

How can we construct a graph from this?



Stage

Classic Approach

Nodes ← Characters

Edges ← Co-Occurrence in Scene

≠ Edges ← Co-Occurrence on Stage!

...and What about...

...Frequencies?

...Speech?

...*Stage Groups*?

Hyperbard: Representations

HyperBard: Representations

First, our *semantic mapping*—Nodes and edges: What types of entities do we assign?

Representation	Semantic Mapping

Hyperbard: Representations

First, our *semantic mapping*—Nodes and edges: What types of entities do we assign?

Representation	Semantic Mapping
ce-scene-b	Nodes ← Characters Edges ← Co-occurrence
ce-scene-mb	
ce-scene-mw	
ce-group-b	
ce-group-mb	
ce-group-mw	

Hyperbard: Representations

First, our *semantic mapping*—Nodes and edges: What types of entities do we assign?

Representation	Semantic Mapping	
ce-scene-b **ce-scene-mb** ce-scene-mw ce-group-b ce-group-mb ce-group-mw	Nodes ← Characters Edges ← Co-occurrence	
se-scene-b se-scene-w se-group-b se-group-w	Edges ← Occurrence	Nodes (1) ← Characters Nodes (2) ← Play parts
se-speech-wd se-speech-mwd	Edges ← Information flow	

HYPERBARD: Representations

First, our *semantic mapping*—Nodes and edges: What types of entities do we assign?

Representation	Semantic Mapping	
ce-scene-b **ce-scene-mb** ce-scene-mw ce-group-b ce-group-mb ce-group-mw	Nodes ← Characters Edges ← Co-occurrence	
se-scene-b se-scene-w se-group-b se-group-w	Edges ← Occurrence	Nodes (1) ← Characters Nodes (2) ← Play parts
se-speech-wd se-speech-mwd	Edges ← Information flow	
hg-scene-mb hg-scene-mw hg-group-mb hg-group-mw	Edges ← Co-occurrence	Nodes ← Characters
hg-speech-wd hg-speech-mwd	Edges ← Information flow	

HYPERBARD: Representations

Second, our *granularity*—What are
Our modeling units for semantic mapping?

Representation	Semantic Mapping		Granularity
ce-scene-b **ce-scene-mb** ce-scene-mw	Nodes ← Characters Edges ← Co-occurrence		Edges ↔ Scenes
ce-group-b ce-group-mb ce-group-mw			Edges ↔ Stage groups
se-scene-b se-scene-w	Edges ← Occurrence	Nodes (1) ← Characters Nodes (2) ← Play parts	Nodes (2) ↔ Scenes
se-group-b se-group-w			Nodes (2) ↔ Stage groups
se-speech-wd se-speech-mwd	Edges ← Information flow		Nodes (2) ↔ Stage groups Edges ↔ Speech acts
hg-scene-mb hg-scene-mw	Edges ← Co-occurrence	Nodes ← Characters	Edges ↔ Scenes
hg-group-mb hg-group-mw			Edges ↔ Stage groups
hg-speech-wd hg-speech-mwd	Edges ← Information flow		Edges ↔ Speech acts

HYPERBARD: Representations

And third, our *expressivity*: What more
Do we attach to all our modeling units?
Directions, weights, and multiplicities,
Or attributes and cardinalities…

Representation	Semantic Mapping		Granularity	Expressivity
ce-scene-b	Nodes ← Characters Edges ← Co-occurrence		Edges ↔ Scenes	—
ce-scene-mb				Edge order
ce-scene-mw				Edge order, edge weights
ce-group-b			Edges ↔ Stage groups	—
ce-group-mb				Edge order
ce-group-mw				Edge order, edge weights
se-scene-b	Edges ← Occurrence	Nodes (1) ← Characters Nodes (2) ← Play parts	Nodes (2) ↔ Scenes	Partial node and edge order
se-scene-w				Partial node and edge order; edge weights
se-group-b			Nodes (2) ↔ Stage groups	Partial node and edge order
se-group-w				Partial node and edge order; edge weights
se-speech-wd	Edges ← Information flow		Nodes (2) ↔ Stage groups	Partial node order; edge weights, edge directions
se-speech-mwd			Edges ↔ Speech acts	Partial node and edge order; edge weights, edge directions
hg-scene-mb	Edges ← Co-occurrence	Nodes ← Characters	Edges ↔ Scenes	Edge order
hg-scene-mw				Edge order, edge weights; edge-specific node weights
hg-group-mb			Edges ↔ Stage groups	Edge order
hg-group-mw				Edge order, edge weights; edge-specific node weights
hg-speech-wd	Edges ← Information flow		Edges ↔ Speech acts	Edge directions, edge weights
hg-speech-mwd				Edge order, edge directions, edge weights

Hyperbard: Representations

Hypergraphs

And third, our *expressivity*: What more
Do we attach to all our modeling units?
Directions, weights, and multiplicities,
Or attributes and cardinalities…

Representation	Semantic Mapping		Granularity	Expressivity
ce-scene-b	Nodes ← Characters Edges ← Co-occurrence		Edges ↔ Scenes	—
ce-scene-mb				Edge order
ce-scene-mw				Edge order, edge weights
ce-group-b			Edges ↔ Stage groups	—
ce-group-mb				Edge order
ce-group-mw				Edge order, edge weights
se-scene-b	Edges ← Occurrence	Nodes (1) ← Characters Nodes (2) ← Play parts	Nodes (2) ↔ Scenes	Partial node and edge order
se-scene-w				Partial node and edge order; edge weights
se-group-b			Nodes (2) ↔ Stage groups	Partial node and edge order
se-group-w				Partial node and edge order; edge weights
se-speech-wd	Edges ← Information flow		Nodes (2) ↔ Stage groups Edges ↔ Speech acts	Partial node order; edge weights, edge directions
se-speech-mwd				Partial node and edge order; edge weights, edge directions
hg-scene-mb	Edges ← Co-occurrence	Nodes ← Characters	Edges ↔ Scenes	Edge order
hg-scene-mw				Edge order, edge weights; edge-specific node weights
hg-group-mb			Edges ↔ Stage groups	Edge order
hg-group-mw				Edge order, edge weights; edge-specific node weights
hg-speech-wd	Edges ← Information flow		Edges ↔ Speech acts	Edge directions, edge weights
hg-speech-mwd				Edge order, edge directions, edge weights

Hyperbard: Representations

Clique Expansions

And third, our *expressivity*: What more
Do we attach to all our modeling units?
Directions, weights, and multiplicities,
Or attributes and cardinalities…

Representation	Semantic Mapping		Granularity	Expressivity
ce-scene-b	Nodes ← Characters Edges ← Co-occurrence		Edges ↔ Scenes	—
ce-scene-mb				Edge order
ce-scene-mw				Edge order, edge weights
ce-group-b			Edges ↔ Stage groups	—
ce-group-mb				Edge order
ce-group-mw				Edge order, edge weights
se-scene-b	Edges ← Occurrence	Nodes (1) ← Characters Nodes (2) ← Play parts	Nodes (2) ↔ Scenes	Partial node and edge order
se-scene-w				Partial node and edge order; edge weights
se-group-b			Nodes (2) ↔ Stage groups	Partial node and edge order
se-group-w				Partial node and edge order; edge weights
se-speech-wd	Edges ← Information flow		Nodes (2) ↔ Stage groups	Partial node order; edge weights, edge directions
se-speech-mwd			Edges ↔ Speech acts	Partial node and edge order; edge weights, edge directions
hg-scene-mb	Edges ← Co-occurrence	Nodes ← Characters	Edges ↔ Scenes	Edge order
hg-scene-mw				Edge order, edge weights; edge-specific node weights
hg-group-mb			Edges ↔ Stage groups	Edge order
hg-group-mw				Edge order, edge weights; edge-specific node weights
hg-speech-wd	Edges ← Information flow		Edges ↔ Speech acts	Edge directions, edge weights
hg-speech-mwd				Edge order, edge directions, edge weights

Hyperbard: Representations

Star Expansions

And third, our *expressivity*: What more
Do we attach to all our modeling units?
Directions, weights, and multiplicities,
Or attributes and cardinalities…

Representation	Semantic Mapping		Granularity	Expressivity
ce-scene-b	Nodes ← Characters Edges ← Co-occurrence		Edges ↔ Scenes	—
ce-scene-mb				Edge order
ce-scene-mw				Edge order, edge weights
ce-group-b			Edges ↔ Stage groups	—
ce-group-mb				Edge order
ce-group-mw				Edge order, edge weights
se-scene-b	Edges ← Occurrence	Nodes (1) ← Characters Nodes (2) ← Play parts	Nodes (2) ↔ Scenes	Partial node and edge order
se-scene-w				Partial node and edge order; edge weights
se-group-b			Nodes (2) ↔ Stage groups	Partial node and edge order
se-group-w				Partial node and edge order; edge weights
se-speech-wd	Edges ← Information flow		Nodes (2) ↔ Stage groups Edges ↔ Speech acts	Partial node order; edge weights, edge directions
se-speech-mwd				Partial node and edge order; edge weights, edge directions
hg-scene-mb	Edges ← Co-occurrence	Nodes ← Characters	Edges ↔ Scenes	Edge order
hg-scene-mw				Edge order, edge weights; edge-specific node weights
hg-group-mb			Edges ↔ Stage groups	Edge order
hg-group-mw				Edge order, edge weights; edge-specific node weights
hg-speech-wd	Edges ← Information flow		Edges ↔ Speech acts	Edge directions, edge weights
hg-speech-mwd				Edge order, edge directions, edge weights

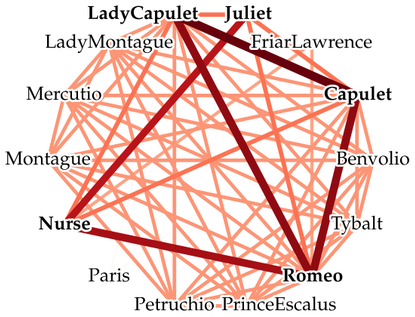
Hyperbard: Illustrations

Romeo and Juliet, Act III

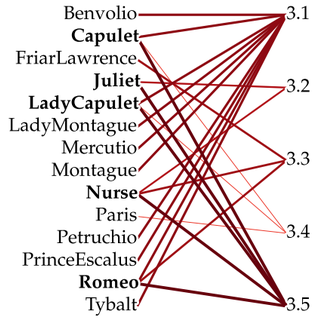
Hyperbard: Illustrations

Romeo and Juliet, Act III

Clique Expansion



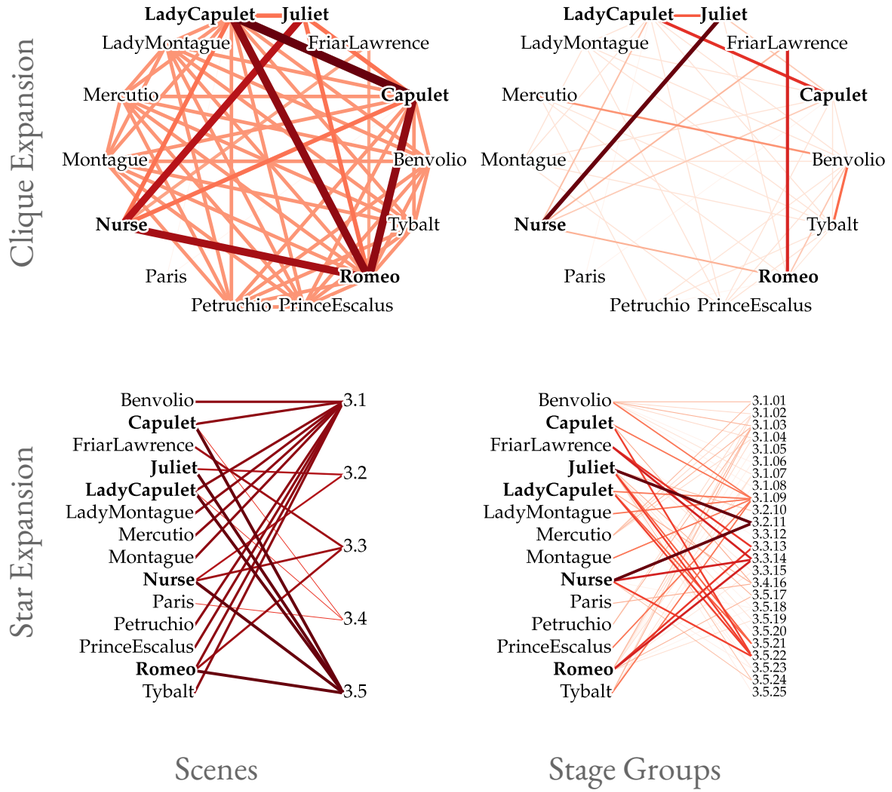
Star Expansion



Scenes

Hyperbard: Illustrations

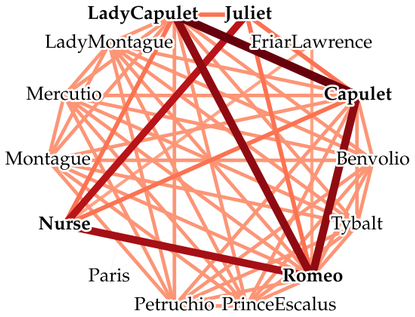
Romeo and Juliet, Act III

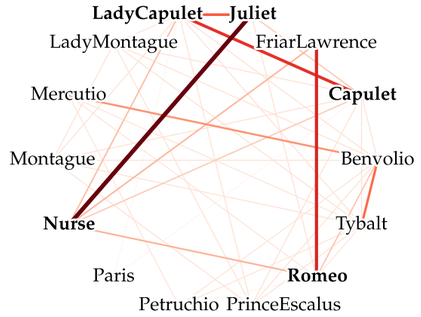


Hyperbard: Illustrations

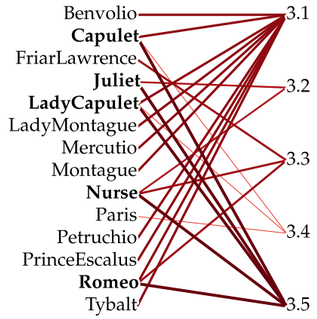
Romeo and Juliet, Act III

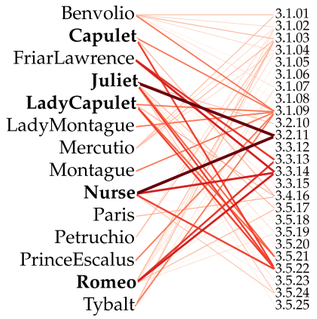
Clique Expansion

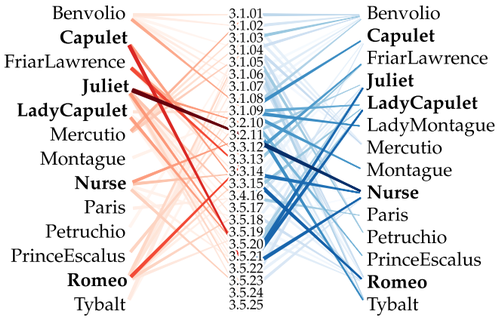




Star Expansion



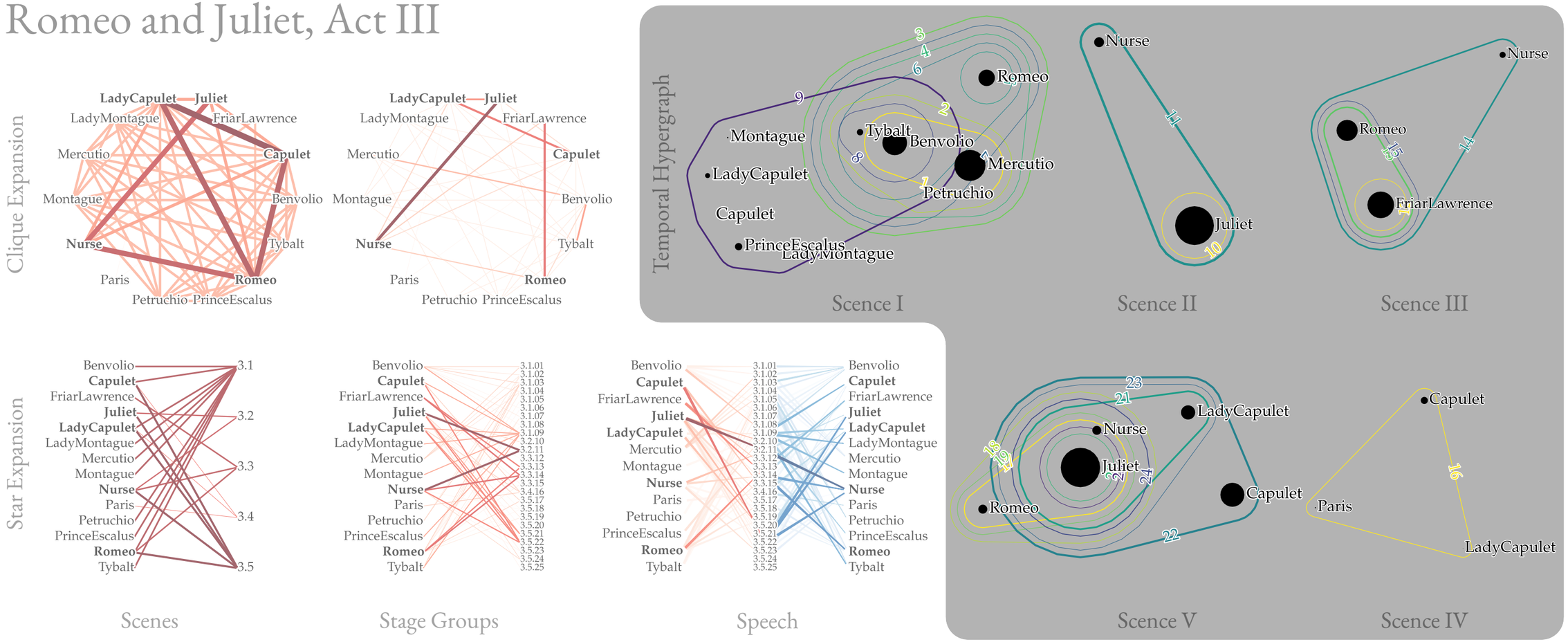




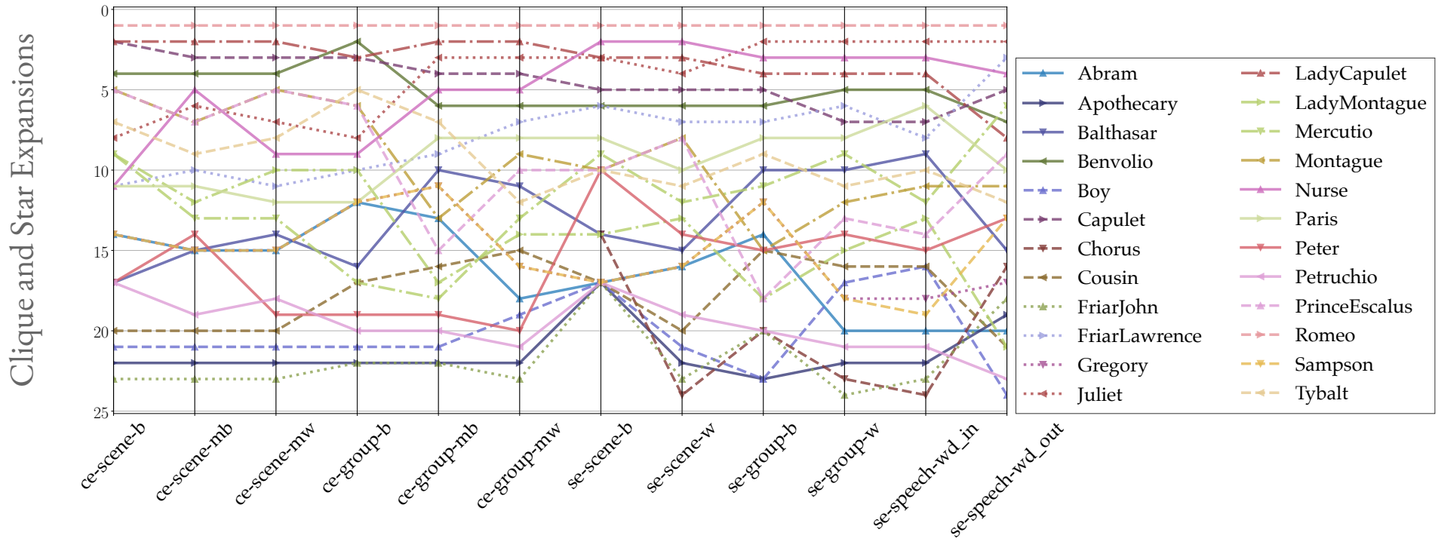
Scenes | Stage Groups | Speech

Hyperbard: Illustrations

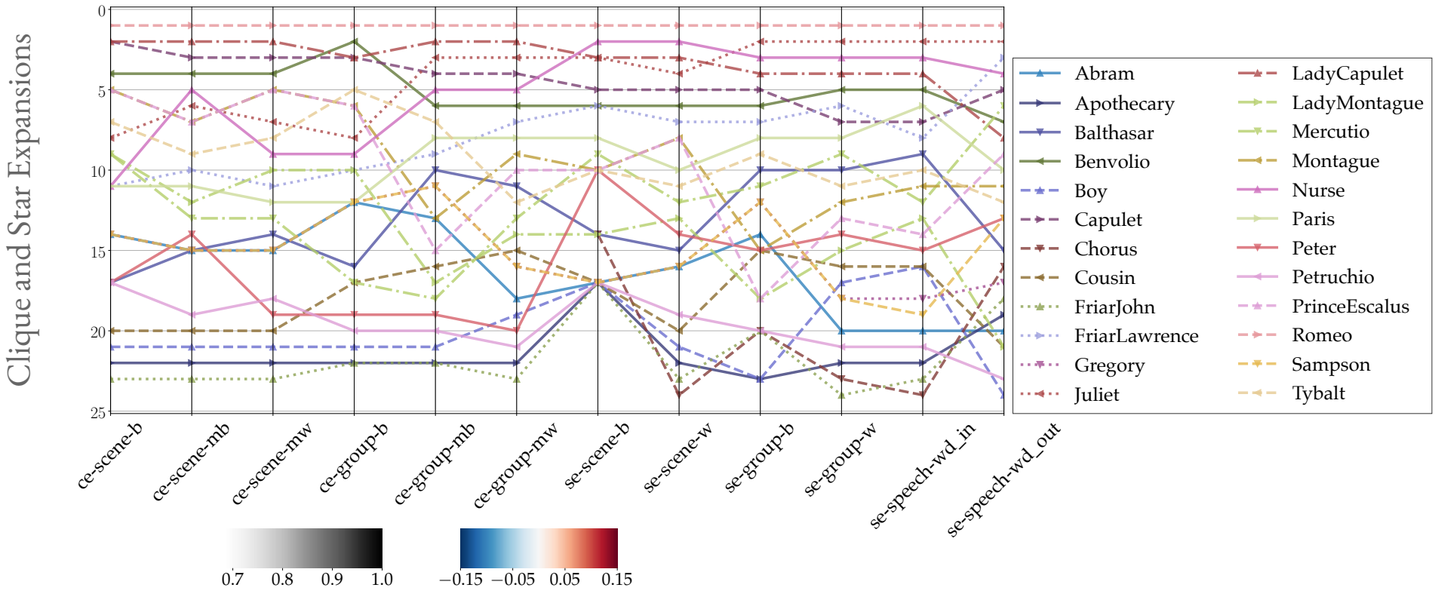
Romeo and Juliet, Act III

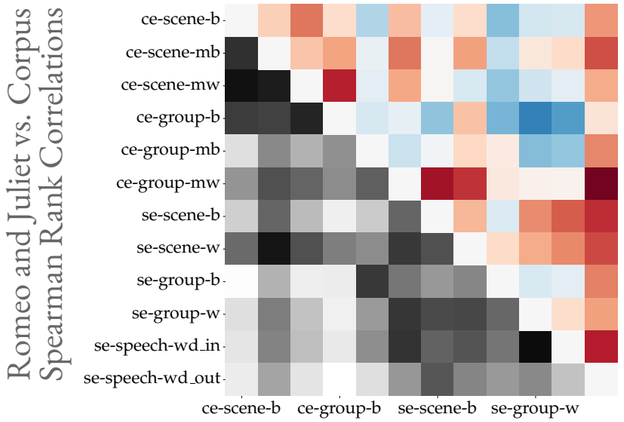


Representations Matter!

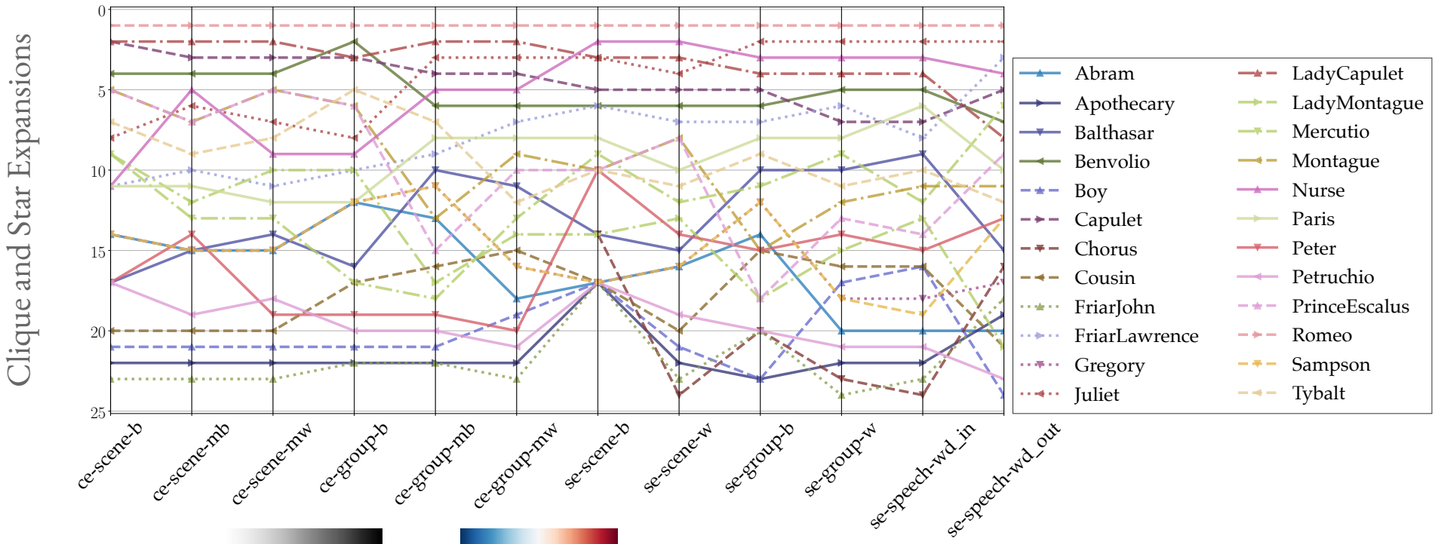


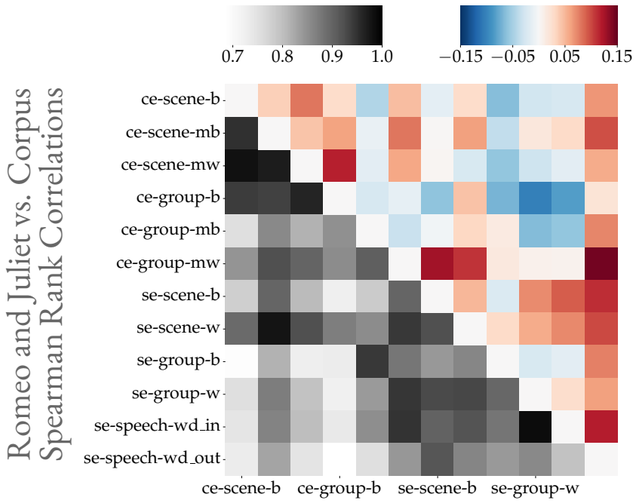
Representations Matter!

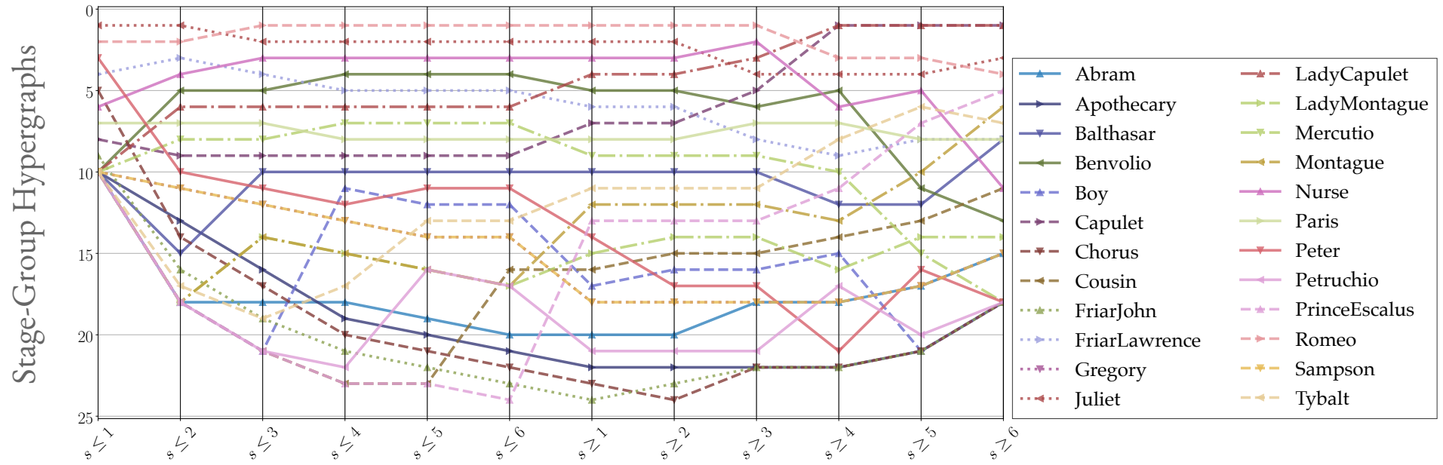




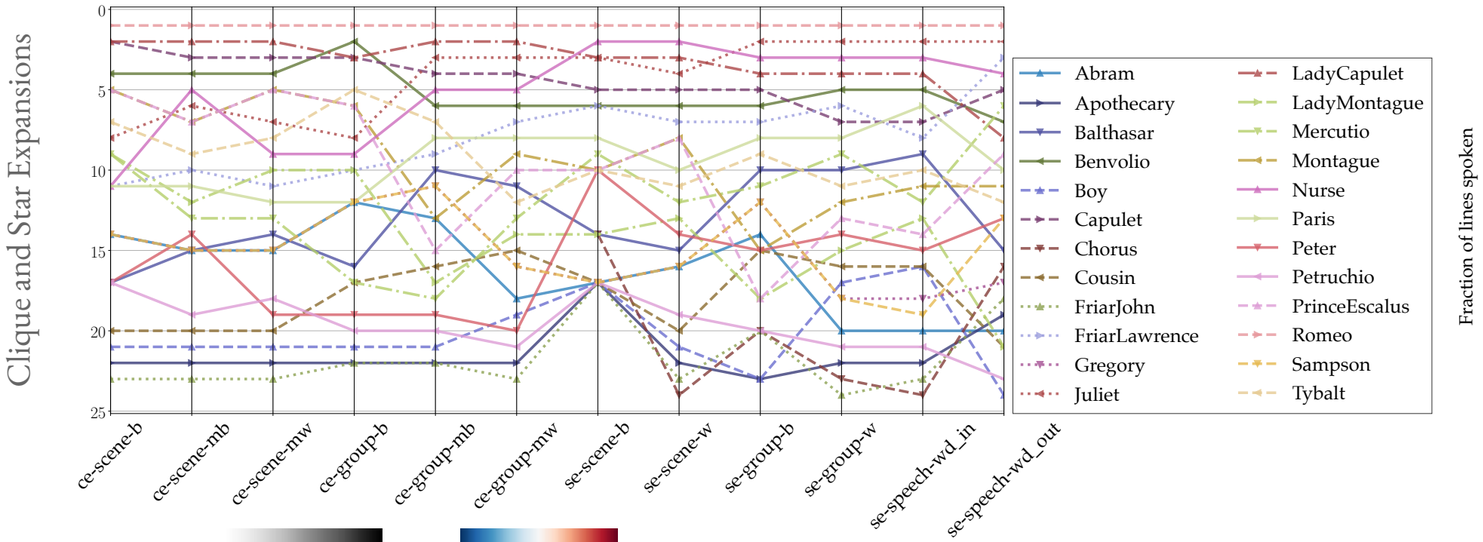
Representations Matter!

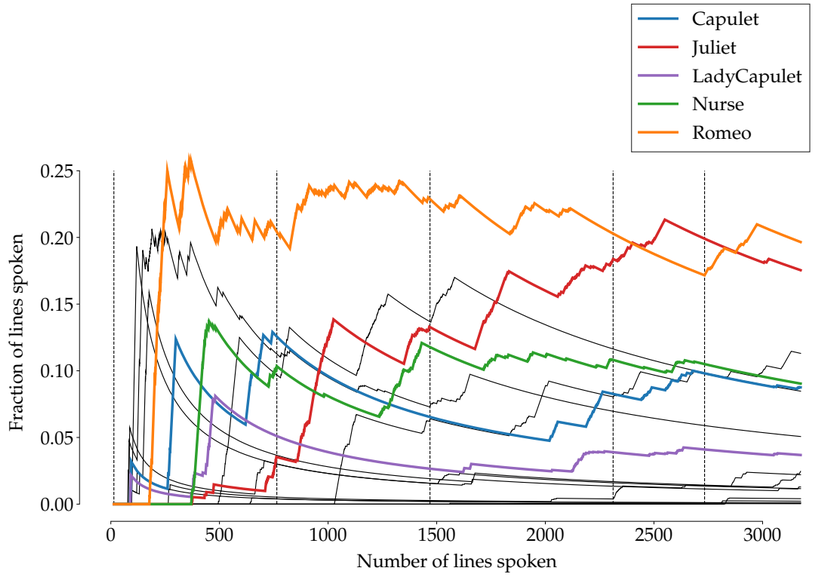


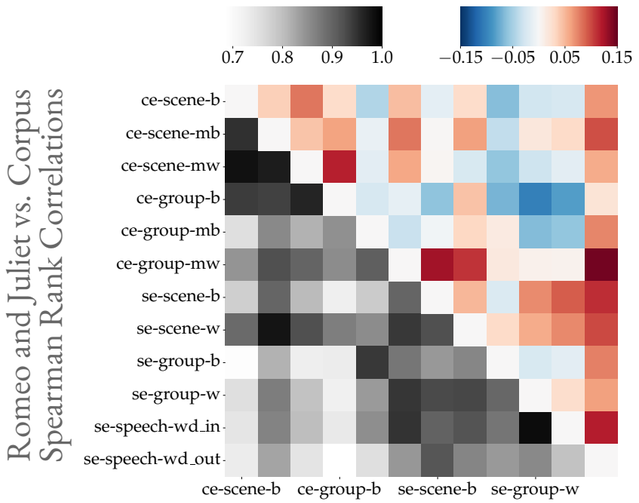


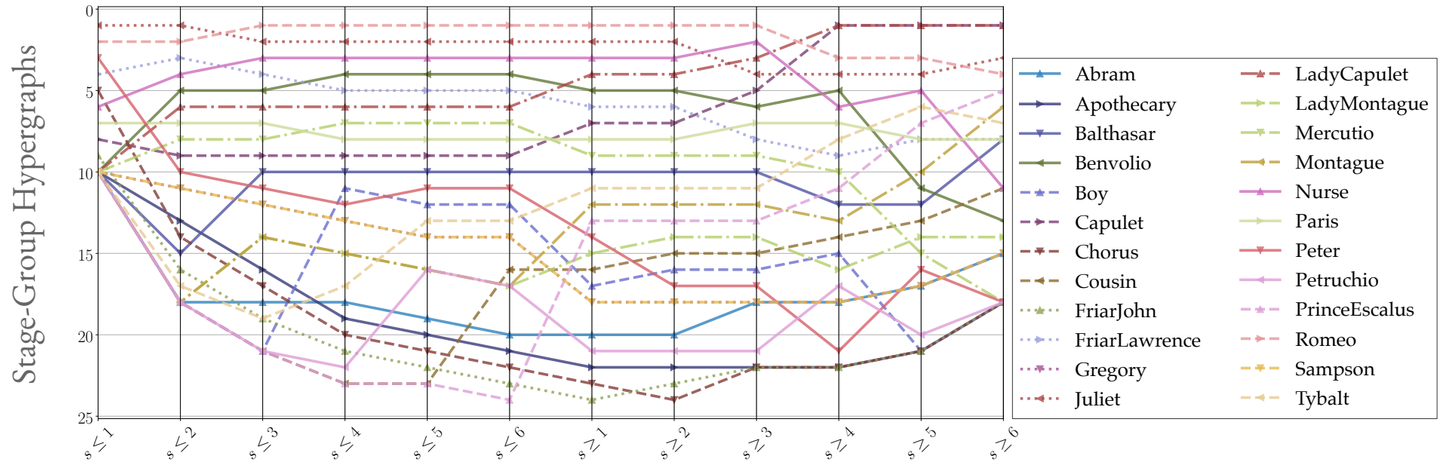


Representations Matter!









The Bigger Picture

The Bigger Picture

Multiverse Analysis

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

> Liu et al., TVCG 2021 (supporting specification and visualization)

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

Network-to-graph transformations → *many* "researcher degrees of freedom"

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

Network-to-graph transformations → *many* "researcher degrees of freedom"

→ Research involving graphs particularly suitable for multiverse analysis

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

Network-to-graph transformations → *many* "researcher degrees of freedom"

- → Research involving graphs particularly suitable for multiverse analysis

(How) Could we make multiverse analysis the new normal? And should we?

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

Network-to-graph transformations → *many* "researcher degrees of freedom"

- → Research involving graphs particularly suitable for multiverse analysis

(How) Could we make multiverse analysis the new normal? And should we?

A Graph Data Multiverse

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

Network-to-graph transformations → *many* "researcher degrees of freedom"

→ Research involving graphs particularly suitable for multiverse analysis

(How) Could we make multiverse analysis the new normal? And should we?

A Graph Data Multiverse

Raw dataset

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

Network-to-graph transformations → *many* "researcher degrees of freedom"

→ Research involving graphs particularly suitable for multiverse analysis

(How) Could we make multiverse analysis the new normal? And should we?

A Graph Data Multiverse

Raw dataset

+

Set of transformations

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

Network-to-graph transformations → *many* "researcher degrees of freedom"

→ Research involving graphs particularly suitable for multiverse analysis

(How) Could we make multiverse analysis the new normal? And should we?

A Graph Data Multiverse

Raw dataset

+

Set of transformations

Set of graph datasets

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

Network-to-graph transformations → *many* "researcher degrees of freedom"

→ Research involving graphs particularly suitable for multiverse analysis

(How) Could we make multiverse analysis the new normal? And should we?

A Graph Data Multiverse

Raw dataset

+

Set of transformations

Set of graph datasets

…

The Bigger Picture

Multiverse Analysis

Origin: Psychology's replication crisis (Steegen et al., PPS 2016)

Idea: Perform analyses across all reasonably pre-processed datasets

Gradually picked up in human-computer interaction and machine learning

- Liu et al., TVCG 2021 (supporting specification and visualization)
- Bell et al., NeurIPS 2022 (model evaluation and hyperparameter tuning)
- Meyer et al., FAccT 2023 ("dataset multiplicity")

Network-to-graph transformations → *many* "researcher degrees of freedom"

→ Research involving graphs particularly suitable for multiverse analysis

(How) Could we make multiverse analysis the new normal? And should we?

A Graph Data Multiverse

Raw dataset

+

Set of transformations

Set of graph datasets

…

Myriad research questions

Summary

Summary

1. Graph data does not exist, it is defined.
2. Semantic mapping, granularity, and expressivity are key ingredients to define graph representations.
3. Many phenomena permit several graph representations.
4. Graph data context matters for graph representations.
5. Graph data representations matter for graph methods.
6. Hypergraphs are powerful.
7. HYPERBARD is free.

Summary

1. Graph data does not exist, it is defined.
2. Semantic mapping, granularity, and expressivity are key ingredients to define graph representations.
3. Many phenomena permit several graph representations.
4. Graph data context matters for graph representations.
5. Graph data representations matter for graph methods.
6. Hypergraphs are powerful.
7. HYPERBARD is free.

Let's do multiverse analyses!

Summary

1. Graph data does not exist, it is defined.
2. Semantic mapping, granularity, and expressivity are key ingredients to define graph representations.
3. Many phenomena permit several graph representations.
4. Graph data context matters for graph representations.
5. Graph data representations matter for graph methods.
6. Hypergraphs are powerful.
7. HYPERBARD is free.

Let's do multiverse analyses!

Check it out at *hyperbard.net*!

Summary

1. Graph data does not exist, it is defined.
2. Semantic mapping, granularity, and expressivity are key ingredients to define graph representations.
3. Many phenomena permit several graph representations.
4. Graph data context matters for graph representations.
5. Graph data representations matter for graph methods.
6. Hypergraphs are powerful.
7. HYPERBARD is free.

Get the full play!

Let's do multiverse analyses!

Check it out at *hyperbard.net*!

Summary

1. Graph data does not exist, it is defined.
2. Semantic mapping, granularity, and expressivity are key ingredients to define graph representations.
3. Many phenomena permit several graph representations.
4. Graph data context matters for graph representations.
5. Graph data representations matter for graph methods.
6. Hypergraphs are powerful.
7. HYPERBARD is free.

Get the full play!

Let's do multiverse analyses!

Check it out at *hyperbard.net*!

Short version: DSH 2023
Cudos for appreciating the art!

Thank you! Questions?

1. Graph data does not exist, it is defined.
2. Semantic mapping, granularity, and expressivity are key ingredients to define graph representations.
3. Many phenomena permit several graph representations.
4. Graph data context matters for graph representations.
5. Graph data representations matter for graph methods.
6. Hypergraphs are powerful.
7. HYPERBARD is free.

Get the full play!

Let's do multiverse analyses!

Check it out at *hyperbard.net*!

Short version: DSH 2023
Cudos for appreciating the art!

Inspiration

Inspiration

Col. A datasheet [8]? Well—all the world is data,
And all we care for merely data points;
They get created, updated, deleted,
And every data point plays many parts,
Its fate being seven stages. First, *motivation*
Defining purpose or specific tasks.
Then *composition*, sketching the raw data
And telling people where it was obtained,
If anything's amiss. And then *collection*,
How did we get each single data point,
And what else did we check. Then *preprocessing*,
Full of strange quirks and idiosyncrasies,
But made that it looks principled. Then *uses*,
What all things did we do, what could have been,
And what should not be done. Then *distribution*,
If, when, and how will we make data public,
Restrictions by third parties, if imposed,
And also all the laws. Last stage of all,
That ends this template documentary,
Is *maintenance* and hosting and support,
Sans updates, sans errata, sans comment.

Inspiration

Col. A datasheet [8]? Well—all the world is data,
And all we care for merely data points;
They get created, updated, deleted,
And every data point plays many parts,
Its fate being seven stages. First, *motivation*
Defining purpose or specific tasks.
Then *composition*, sketching the raw data
And telling people where it was obtained,
If anything's amiss. And then *collection*,
How did we get each single data point,
And what else did we check. Then *preprocessing*,
Full of strange quirks and idiosyncrasies,
But made that it looks principled. Then *uses*,
What all things did we do, what could have been,
And what should not be done. Then *distribution*,
If, when, and how will we make data public,
Restrictions by third parties, if imposed,
And also all the laws. Last stage of all,
That ends this template documentary,
Is *maintenance* and hosting and support,
Sans updates, sans errata, sans comment.

Jon's Speech in *As You Like It*

Inspiration

Col. A datasheet [8]? Well—all the world is data,
And all we care for merely data points;
They get created, updated, deleted,
And every data point plays many parts,
Its fate being seven stages. First, *motivation*
Defining purpose or specific tasks.
Then *composition*, sketching the raw data
And telling people where it was obtained,
If anything's amiss. And then *collection*,
How did we get each single data point,
And what else did we check. Then *preprocessing*,
Full of strange quirks and idiosyncrasies,
But made that it looks principled. Then *uses*,
What all things did we do, what could have been,
And what should not be done. Then *distribution*,
If, when, and how will we make data public,
Restrictions by third parties, if imposed,
And also all the laws. Last stage of all,
That ends this template documentary,
Is *maintenance* and hosting and support,
Sans updates, sans errata, sans comment.

Jon's Speech in *As You Like It*

Two souls, alas, are dwelling in my breast,
And each one seeks to rule without the other.
The one a falcon, fierce and fighting fetters,
That's dreaming of faun's forest, flying free,
The other a caged chary canary,
That calmly, coyly, cheerfully chants chatters.

Inspiration

Col. A datasheet [8]? Well—all the world is data,
And all we care for merely data points;
They get created, updated, deleted,
And every data point plays many parts,
Its fate being seven stages. First, *motivation*
Defining purpose or specific tasks.
Then *composition*, sketching the raw data
And telling people where it was obtained,
If anything's amiss. And then *collection*,
How did we get each single data point,
And what else did we check. Then *preprocessing*,
Full of strange quirks and idiosyncrasies,
But made that it looks principled. Then *uses*,
What all things did we do, what could have been,
And what should not be done. Then *distribution*,
If, when, and how will we make data public,
Restrictions by third parties, if imposed,
And also all the laws. Last stage of all,
That ends this template documentary,
Is *maintenance* and hosting and support,
Sans updates, sans errata, sans comment.

Jon's Speech in *As You Like It*

Faust in *Goethe's Faust I*

Two souls, alas, are dwelling in my breast,
And each one seeks to rule without the other.
The one a falcon, fierce and fighting fetters,
That's dreaming of faun's forest, flying free,
The other a caged chary canary,
That calmly, coyly, cheerfully chants chatters.

Inspiration

Jon's Speech in *As You Like It*

Col. A datasheet [8]? Well—all the world is data,
And all we care for merely data points;
They get created, updated, deleted,
And every data point plays many parts,
Its fate being seven stages. First, *motivation*
Defining purpose or specific tasks.
Then *composition*, sketching the raw data
And telling people where it was obtained,
If anything's amiss. And then *collection*,
How did we get each single data point,
And what else did we check. Then *preprocessing*,
Full of strange quirks and idiosyncrasies,
But made that it looks principled. Then *uses*,
What all things did we do, what could have been,
And what should not be done. Then *distribution*,
If, when, and how will we make data public,
Restrictions by third parties, if imposed,
And also all the laws. Last stage of all,
That ends this template documentary,
Is *maintenance* and hosting and support,
Sans updates, sans errata, sans comment.

Faust in *Goethe's Faust I*

Two souls, alas, are dwelling in my breast,
And each one seeks to rule without the other.
The one a falcon, fierce and fighting fetters,
That's dreaming of faun's forest, flying free,
The other a caged chary canary,
That calmly, coyly, cheerfully chants chatters.

FIRST DEADLINE *sings.*

Come back to the office lands,
Don't take a chance:
Meta fair but be aware
In camera, better prepare
Fix your figures here and there;
And review two the burden bear.

Cre. Where should this music be? I know the beat.
It sounds no more? No, it begins again.

SECOND DEADLINE *sings.*

To taller skies your metrics rise;
Publish, perish, stars are made;
Do not whine, stay in line,
Otherwise your glory fade.
Dutifully use your wit
And then submit.

Inspiration

Jon's Speech in *As You Like It*

Col. A datasheet [8]? Well—all the world is data,
And all we care for merely data points;
They get created, updated, deleted,
And every data point plays many parts,
Its fate being seven stages. First, *motivation*
Defining purpose or specific tasks.
Then *composition*, sketching the raw data
And telling people where it was obtained,
If anything's amiss. And then *collection*,
How did we get each single data point,
And what else did we check. Then *preprocessing*,
Full of strange quirks and idiosyncrasies,
But made that it looks principled. Then *uses*,
What all things did we do, what could have been,
And what should not be done. Then *distribution*,
If, when, and how will we make data public,
Restrictions by third parties, if imposed,
And also all the laws. Last stage of all,
That ends this template documentary,
Is *maintenance* and hosting and support,
Sans updates, sans errata, sans comment.

Faust in *Goethe's Faust I*

Two souls, alas, are dwelling in my breast,
And each one seeks to rule without the other.
The one a falcon, fierce and fighting fetters,
That's dreaming of faun's forest, flying free,
The other a caged chary canary,
That calmly, coyly, cheerfully chants chatters.

Ariel's Song in *The Tempest*

FIRST DEADLINE *sings.*

Come back to the office lands,
Don't take a chance:
Meta fair but be aware
In camera, better prepare
Fix your figures here and there;
And review two the burden bear.

Cre. Where should this music be? I know the beat.
It sounds no more? No, it begins again.

SECOND DEADLINE *sings.*

To taller skies your metrics rise;
Publish, perish, stars are made;
Do not whine, stay in line,
Otherwise your glory fade.
Dutifully use your wit
And then submit.

Inspiration

319 *Cre.* To flee or PhD—that is the question:
320 Whether our destiny lies in the system,
321 To cling onto scientific ladder's rungs,
322 Or to renounce the reign of rules unwritten
323 And, by opposing, vanish. To flee, to think—
324 To think, perchance discover. Ay, there's the rub,
325 For once outside the pithy paywalled castles,
326 The giant's shoulders quickly out of reach,
327 For lack of funding. There's cautiousness
328 That crafts careers of so long strive,
329 And makes us rather swarm the conference streams
330 Than swim the savage seas so far uncharted.
331 Thus mellow meal the mighty mills of science,
332 And conscience can coerce our compliance.

Inspiration

319 *Cre.* To flee or PhD—that is the question:
320 Whether our destiny lies in the system,
321 To cling onto scientific ladder's rungs,
322 Or to renounce the reign of rules unwritten
323 And, by opposing, vanish. To flee, to think—
324 To think, perchance discover. Ay, there's the rub,
325 For once outside the pithy paywalled castles,
326 The giant's shoulders quickly out of reach,
327 For lack of funding. There's cautiousness
328 That crafts careers of so long strive,
329 And makes us rather swarm the conference streams
330 Than swim the savage seas so far uncharted.
331 Thus mellow meal the mighty mills of science,
332 And conscience can coerce our compliance.

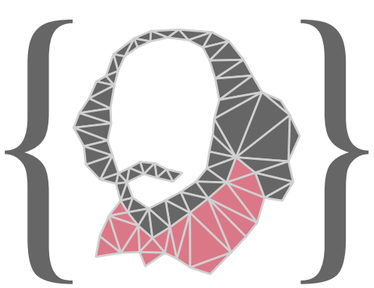
Hamlet's Monologue in *Hamlet*

Inspiration

Cre. To flee or PhD—that is the question:
Whether our destiny lies in the system,
To cling onto scientific ladder's rungs,
Or to renounce the reign of rules unwritten
And, by opposing, vanish. To flee, to think—
To think, perchance discover. Ay, there's the rub,
For once outside the pithy paywalled castles,
The giant's shoulders quickly out of reach,
For lack of funding. There's cautiousness
That crafts careers of so long strive,
And makes us rather swarm the conference streams
Than swim the savage seas so far uncharted.
Thus mellow meal the mighty mills of science,
And conscience can coerce our compliance.

Hamlet's Monologue in *Hamlet*

Gra. Full many a transformation have I seen
Flatter the flora with their sovereign hand,
And sovereign's hand in spirit I'll have been
To help evaluate their promised land.
Community, defined as uninvolved
With hideous beauty born by Mother Earth
Begets solutions without problems solved
And burns the flame of wonder in Its dearth.
When culture counters nature, it prevails,
And builds its truths from rigid rigor bricks,
As myriad feeble fledglings it derails
Into the cave of engineering tricks.
For in the trenches of discovery,
To shatter shadows, meet obscurity.



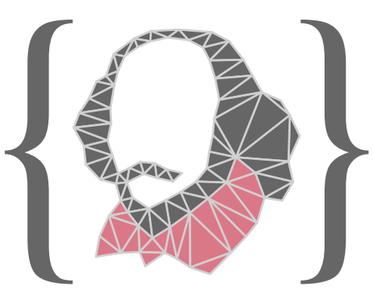
Inspiration

Cre. To flee or PhD—that is the question:
Whether our destiny lies in the system,
To cling onto scientific ladder's rungs,
Or to renounce the reign of rules unwritten
And, by opposing, vanish. To flee, to think—
To think, perchance discover. Ay, there's the rub,
For once outside the pithy paywalled castles,
The giant's shoulders quickly out of reach,
For lack of funding. There's cautiousness
That crafts careers of so long strive,
And makes us rather swarm the conference streams
Than swim the savage seas so far uncharted.
Thus mellow meal the mighty mills of science,
And conscience can coerce our compliance.

Hamlet's Monologue in *Hamlet*

Gra. Full many a transformation have I seen
Flatter the flora with their sovereign hand,
And sovereign's hand in spirit I'll have been
To help evaluate their promised land.
Community, defined as uninvolved
With hideous beauty born by Mother Earth
Begets solutions without problems solved
And burns the flame of wonder in Its dearth.
When culture counters nature, it prevails,
And builds its truths from rigid rigor bricks,
As myriad feeble fledglings it derails
Into the cave of engineering tricks.
For in the trenches of discovery,
To shatter shadows, meet obscurity.

Sonnet 33

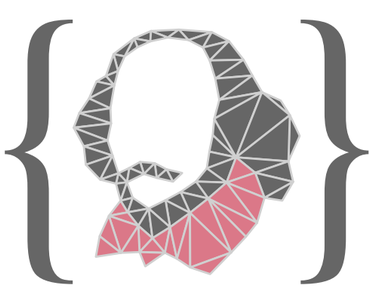


Inspiration

Cre. To flee or PhD—that is the question:
Whether our destiny lies in the system,
To cling onto scientific ladder's rungs,
Or to renounce the reign of rules unwritten
And, by opposing, vanish. To flee, to think—
To think, perchance discover. Ay, there's the rub,
For once outside the pithy paywalled castles,
The giant's shoulders quickly out of reach,
For lack of funding. There's cautiousness
That crafts careers of so long strive,
And makes us rather swarm the conference streams
Than swim the savage seas so far uncharted.
Thus mellow meal the mighty mills of science,
And conscience can coerce our compliance.

Hamlet's Monologue in *Hamlet*

Gra. Full many a transformation have I seen
Flatter the flora with their sovereign hand,
And sovereign's hand in spirit I'll have been
To help evaluate their promised land.
Community, defined as uninvolved
With hideous beauty born by Mother Earth
Begets solutions without problems solved
And burns the flame of wonder in Its dearth.
When culture counters nature, it prevails,
And builds its truths from rigid rigor bricks,
As myriad feeble fledglings it derails
Into the cave of engineering tricks.
For in the trenches of discovery,
To shatter shadows, meet obscurity.



Sonnet 33

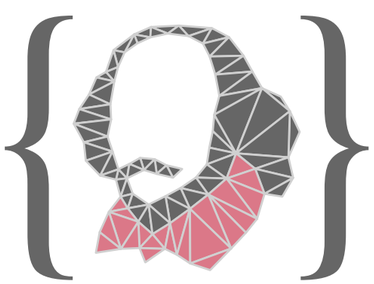
A deadline, and a deadline, and a deadline,
Creeps in this petty pace to publication,
To the last syllable of our defense.
They slew my GRAPH and choked my inspiration,
Our work is but a walking shadow thence.
The curiosity that drew me in
Now lies in dust. The lofty dreams I had
Of mindful monasterial devotion
To just the cause—no more. Out, out, sore studies!

Inspiration

Cre. To flee or PhD—that is the question:
Whether our destiny lies in the system,
To cling onto scientific ladder's rungs,
Or to renounce the reign of rules unwritten
And, by opposing, vanish. To flee, to think—
To think, perchance discover. Ay, there's the rub,
For once outside the pithy paywalled castles,
The giant's shoulders quickly out of reach,
For lack of funding. There's cautiousness
That crafts careers of so long strive,
And makes us rather swarm the conference streams
Than swim the savage seas so far uncharted.
Thus mellow meal the mighty mills of science,
And conscience can coerce our compliance.

Hamlet's Monologue in *Hamlet*

Gra. Full many a transformation have I seen
Flatter the flora with their sovereign hand,
And sovereign's hand in spirit I'll have been
To help evaluate their promised land.
Community, defined as uninvolved
With hideous beauty born by Mother Earth
Begets solutions without problems solved
And burns the flame of wonder in Its dearth.
When culture counters nature, it prevails,
And builds its truths from rigid rigor bricks,
As myriad feeble fledglings it derails
Into the cave of engineering tricks.
For in the trenches of discovery,
To shatter shadows, meet obscurity.



Sonnet 33

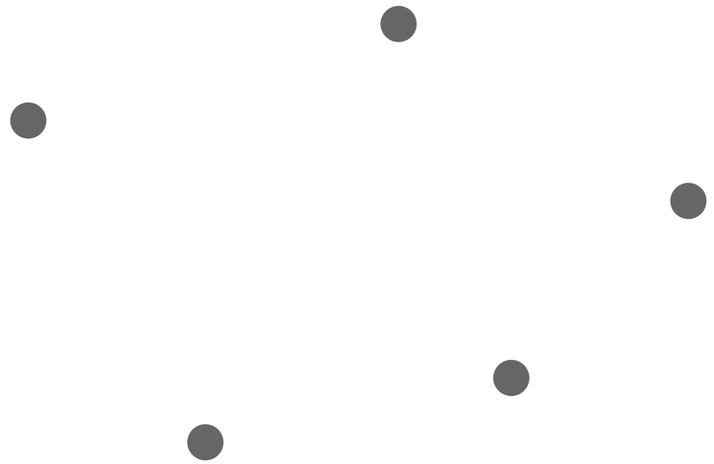
A deadline, and a deadline, and a deadline,
Creeps in this petty pace to publication,
To the last syllable of our defense.
They slew my GRAPH and choked my inspiration,
Our work is but a walking shadow thence.
The curiosity that drew me in
Now lies in dust. The lofty dreams I had
Of mindful monasterial devotion
To just the cause—no more. Out, out, sore studies!

Macbeth's Monologue in *Macbeth*

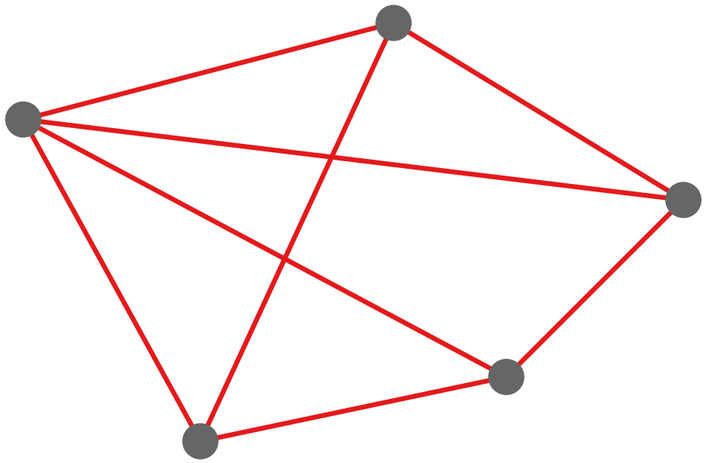
Graphs

Graphs

- Nodes V



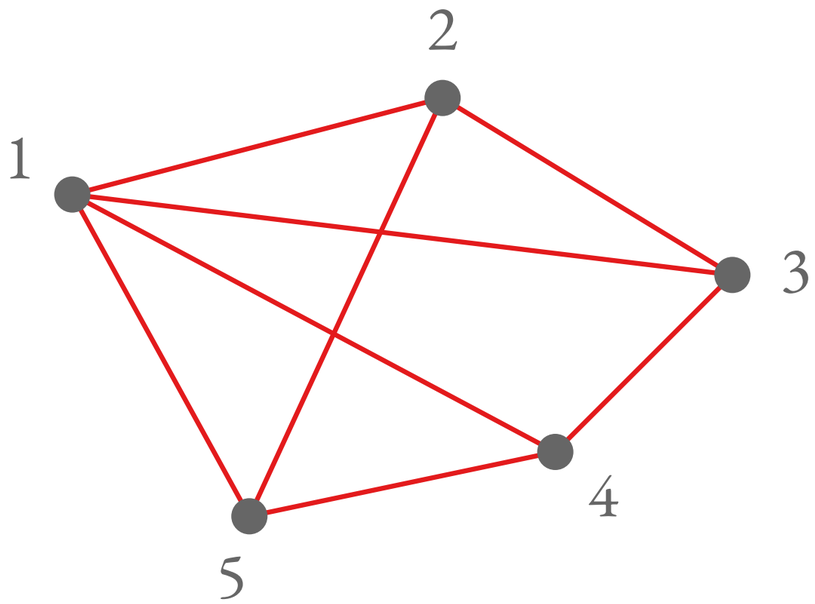
Graphs



Nodes V

Edges E

Graphs

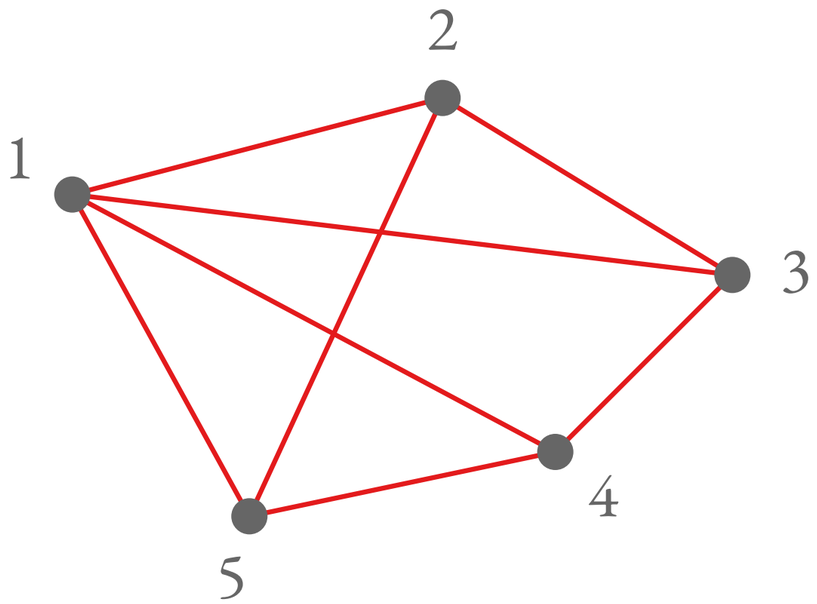


- Nodes V
- Edges E

Graphs

Nodes V

Edges E



Graph G

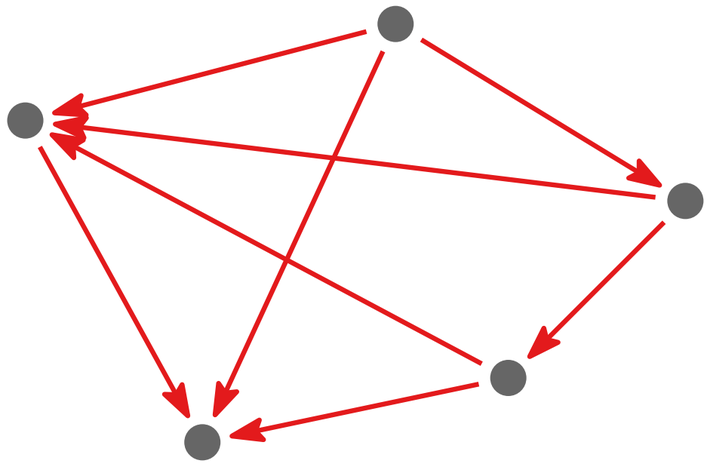
$\Leftrightarrow$

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Adjacency Matrix A

Graphs

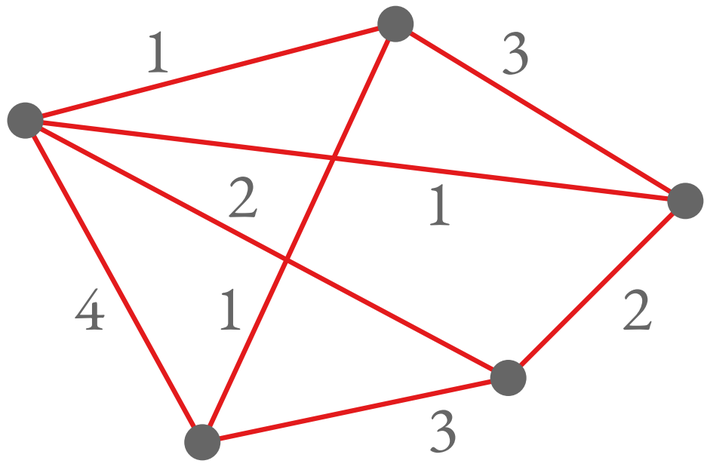
- Nodes V
- Edges E



Directed Graph

Graphs

- Nodes V
- Edges E

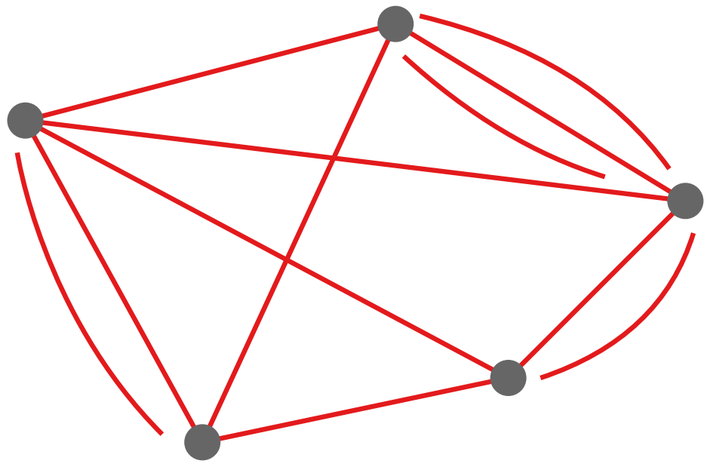


Weighted Graph

Graphs

Nodes V

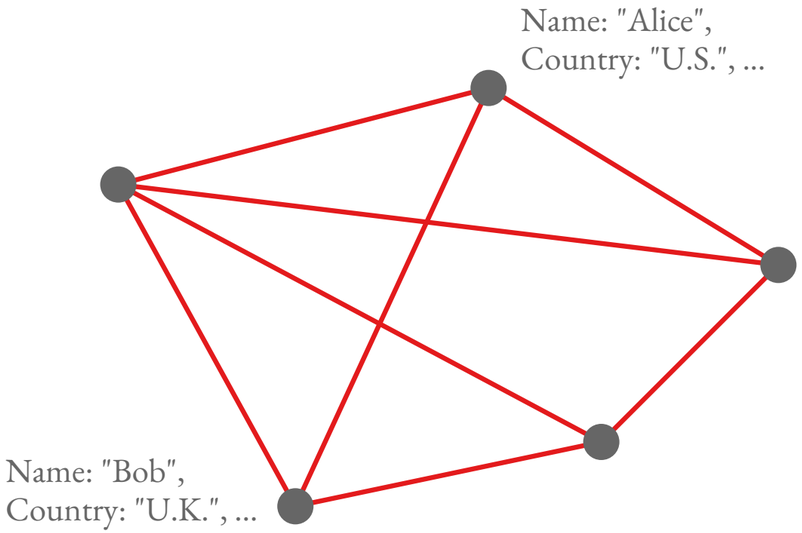
Edges E



Multigraph

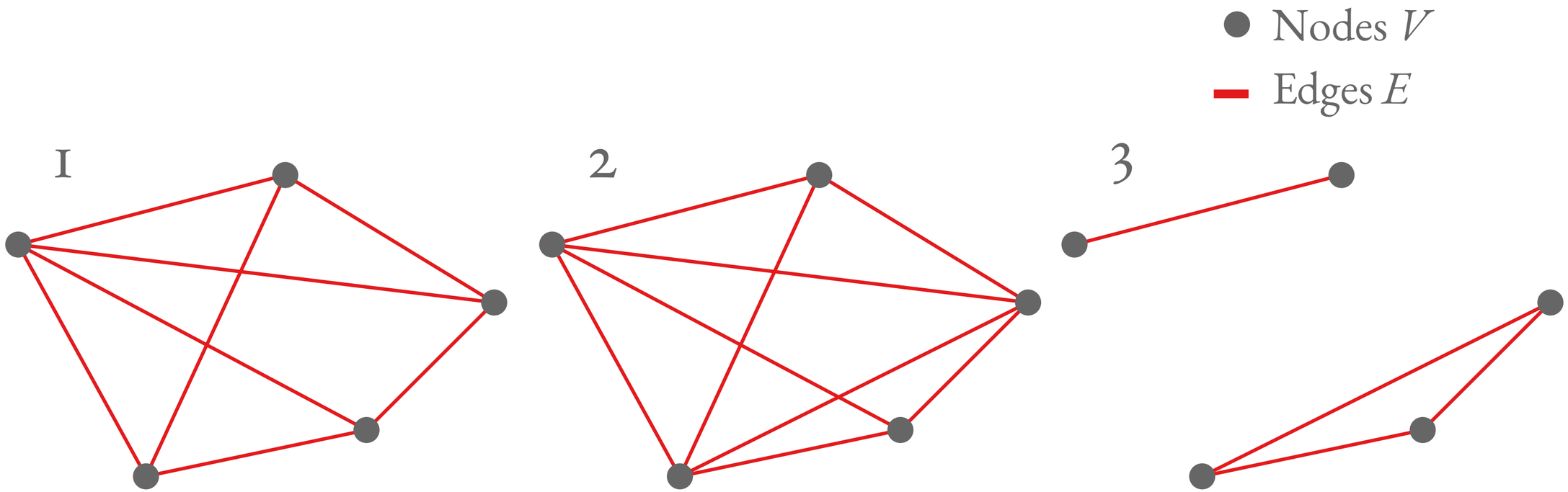
Graphs

- Nodes V
- Edges E



Attributed Graph

Graphs

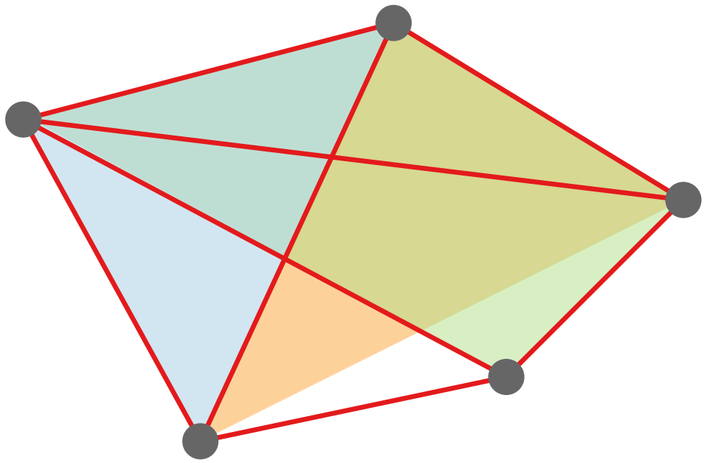


Temporal Graph

Graphs

Nodes V

Edges E



Hypergraph