KGC

The Knowledge Graph Conference

# Causal Knowledge Graph (KG)-Infused Prescriptive Analytics

2022 Knowledge Graph Conference, May 6th

**Dr. Victor Z. Chen, CFA**
Director of Data Analytics and Insights, Experimental Design and Causal Inference
**Fidelity Investments**

# Agenda

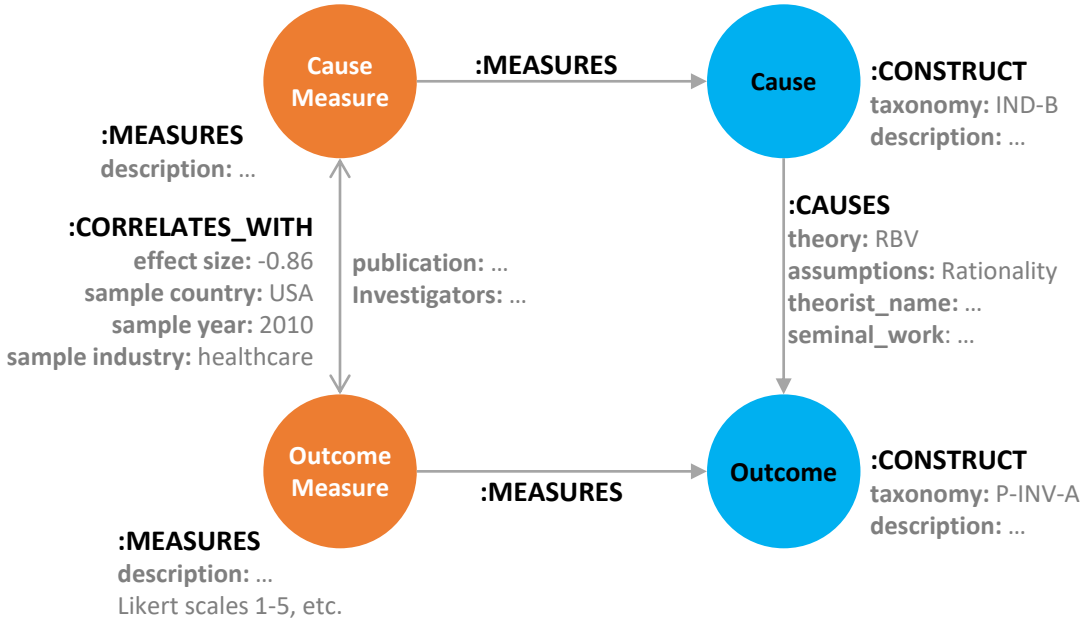
- 1. What is causal knowledge graph (KG)?
- 2. How to construct a causal KG?
- 3. How to infuse a causal KG into data analytics?

# 1. What is a Causal KG?

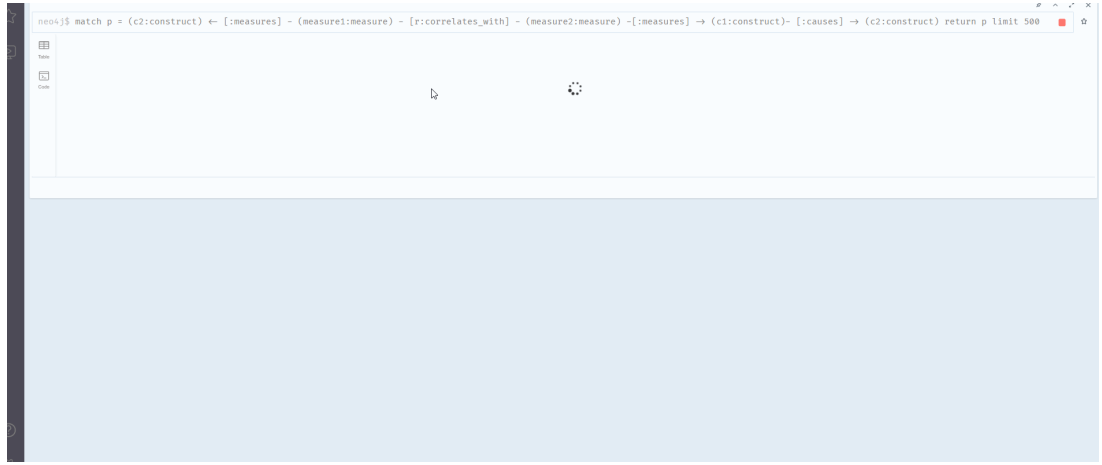
A unified and cumulative representation of knowledge

- Entities (e.g., constructs and variables)
- Directed Links:
  - Ontological relationships (e.g., taxonomy, semantics)
  - Causal logic (e.g., common sense, theoretical hypothesis, expert assumptions)
  - Prior empirical evidence (e.g., experiments and Bayesian meta-analysis)

A causal knowledge unit

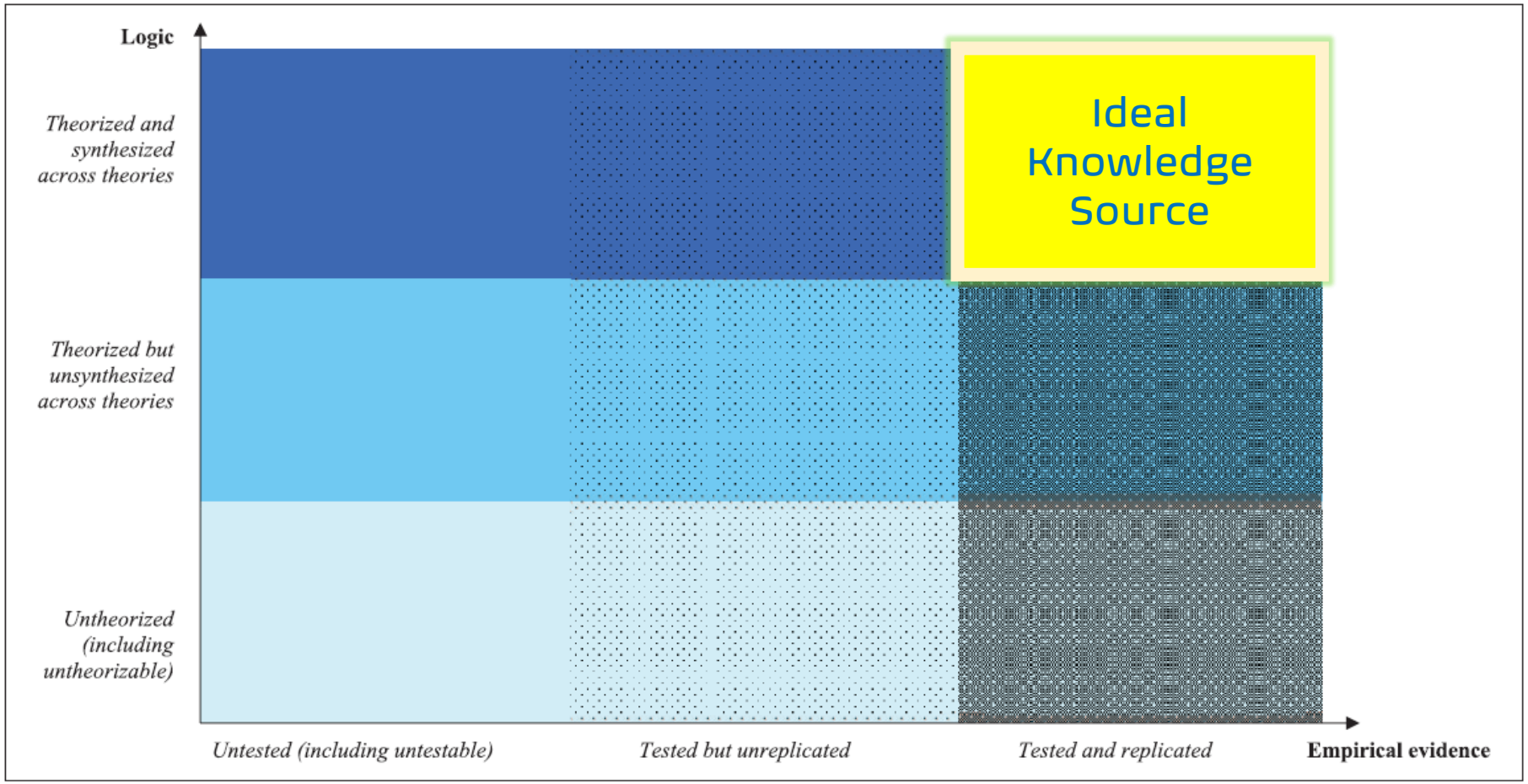



A causal KG database



Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

# 2. How to construct a Causal KG?




Source: Chen and Hitt (2021)

Sourcing

Marketing

A higher sales commission increased revenue from household users

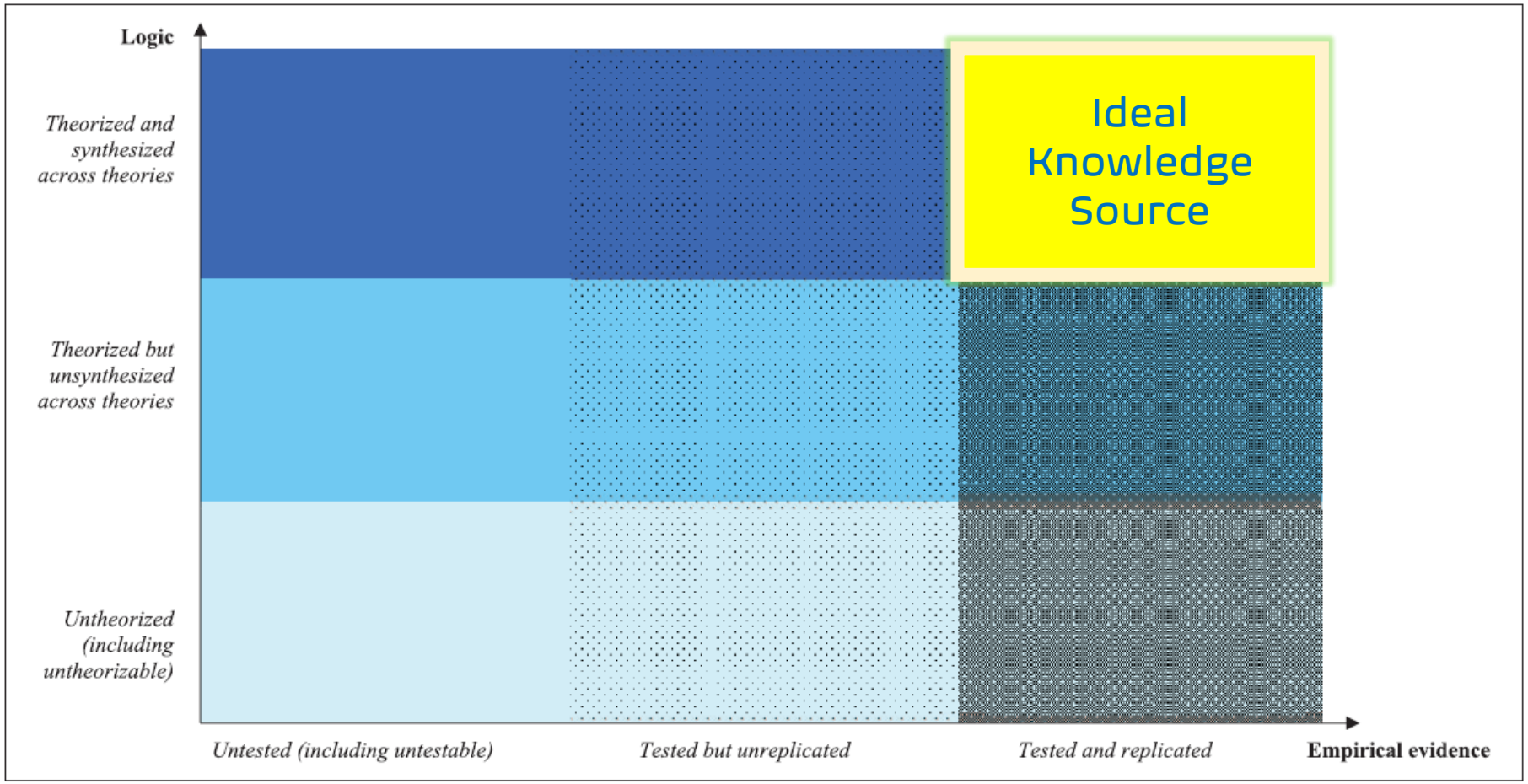
HRM

An entrepreneurial team leader allows more flexible incentives such as increasing sales commission

Finance

Increases of revenue from household users made the majority and still growing share in recent five years' profits

# 2. How to construct a Causal KG?




Source: Chen and Hitt (2021)

Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

## Sourcing

### Marketing

A higher sales commission increased revenue from household users

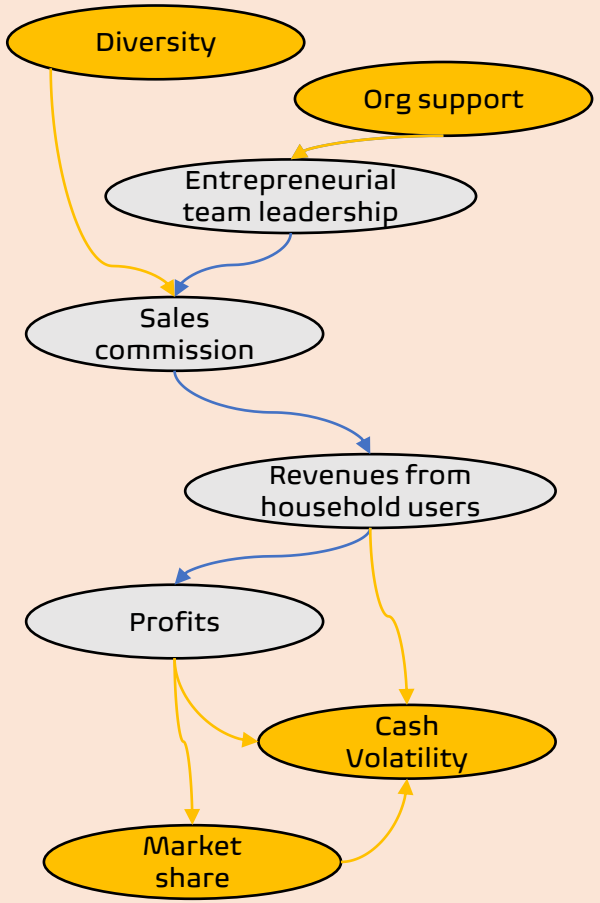
### HRM

An entrepreneurial team leader allows more flexible incentives such as increasing sales commission

### Finance

Increases of revenue from household users made the majority and still growing share in recent five years' profits

## Graphing




## Synthesis

## Analysis

Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

## Sourcing

### Marketing

A higher sales commission increased revenue from household users

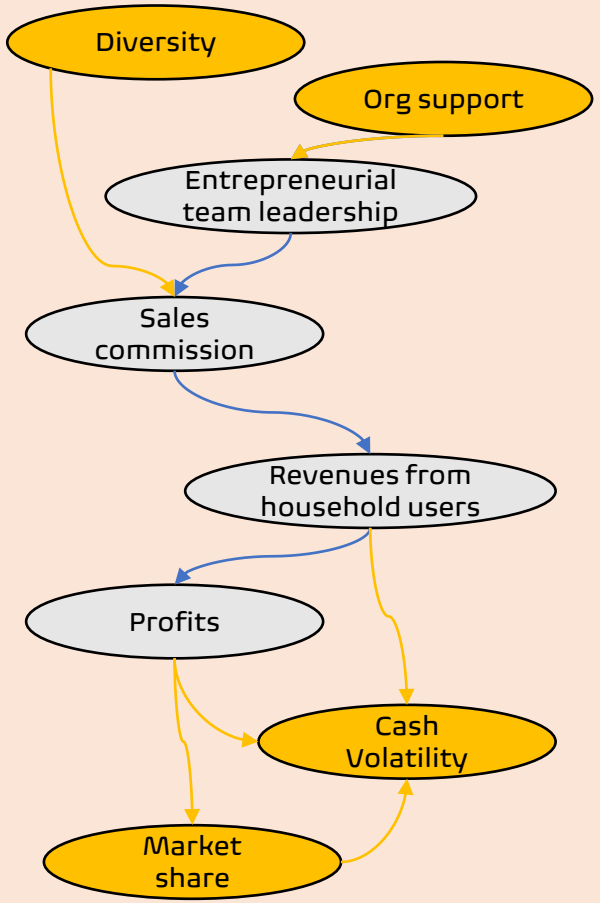
### HRM

An entrepreneurial team leader allows more flexible incentives such as increasing sales commission

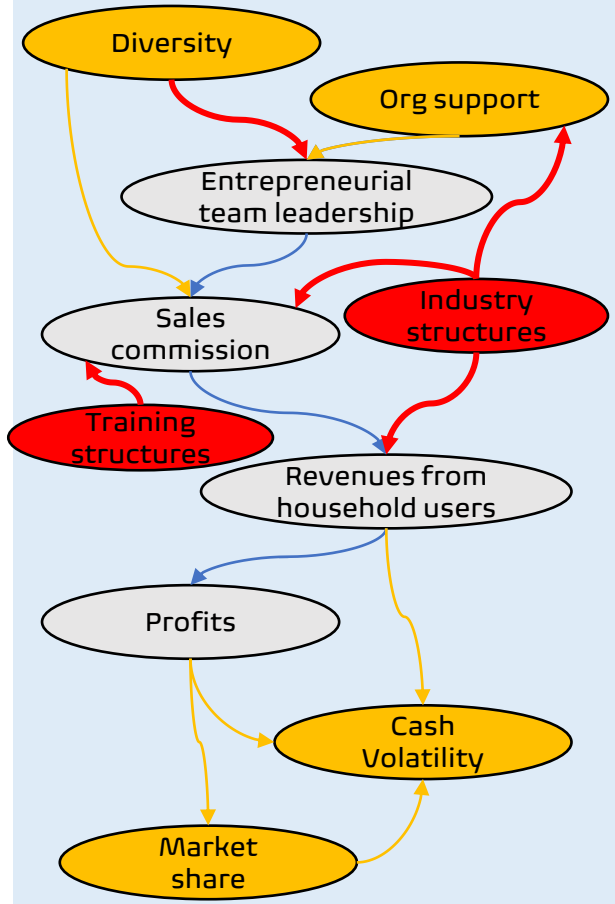
### Finance

Increases of revenue from household users made the majority and still growing share in recent five years' profits

## Graphing




## Synthesis




## Analysis

Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

## Sourcing

**Marketing**

A higher sales commission increased revenue from household users

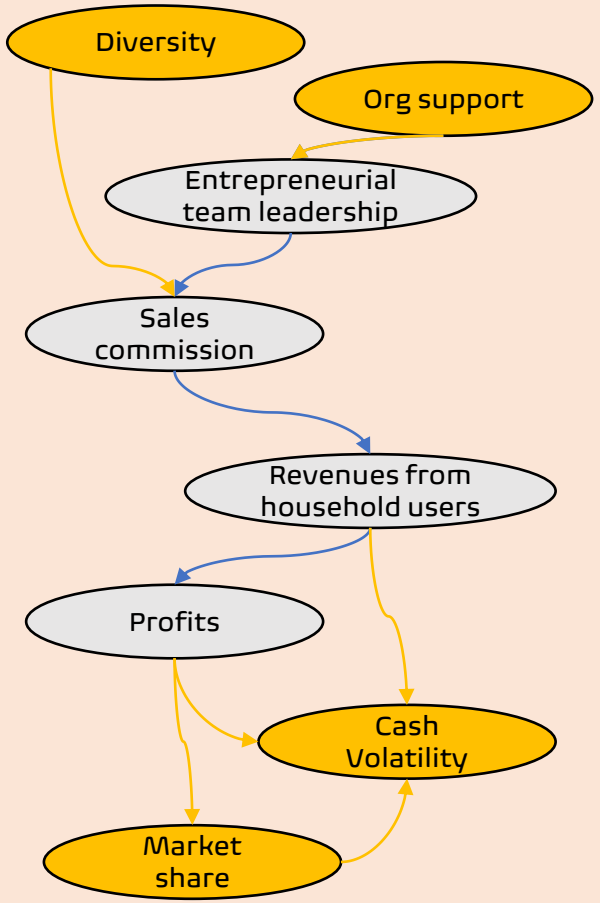
**HRM**

An entrepreneurial team leader allows more flexible incentives such as increasing sales commission

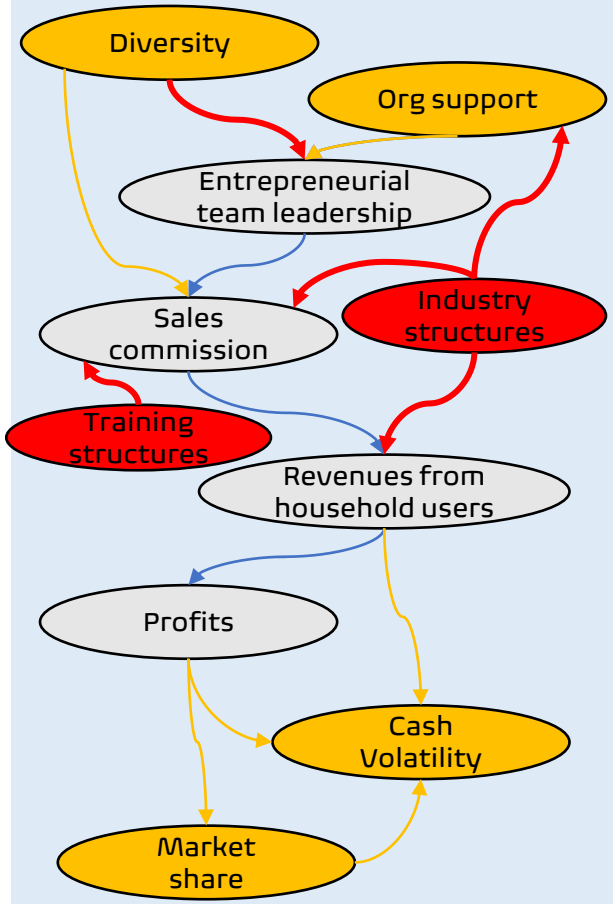
**Finance**

Increases of revenue from household users made the majority and still growing share in recent five years' profits

## Graphing

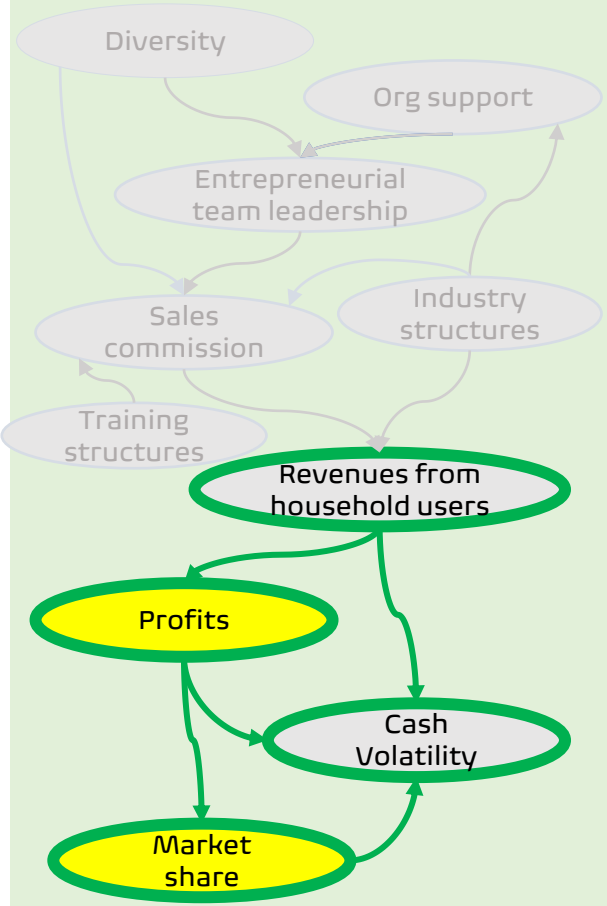



## Synthesis




## Analysis

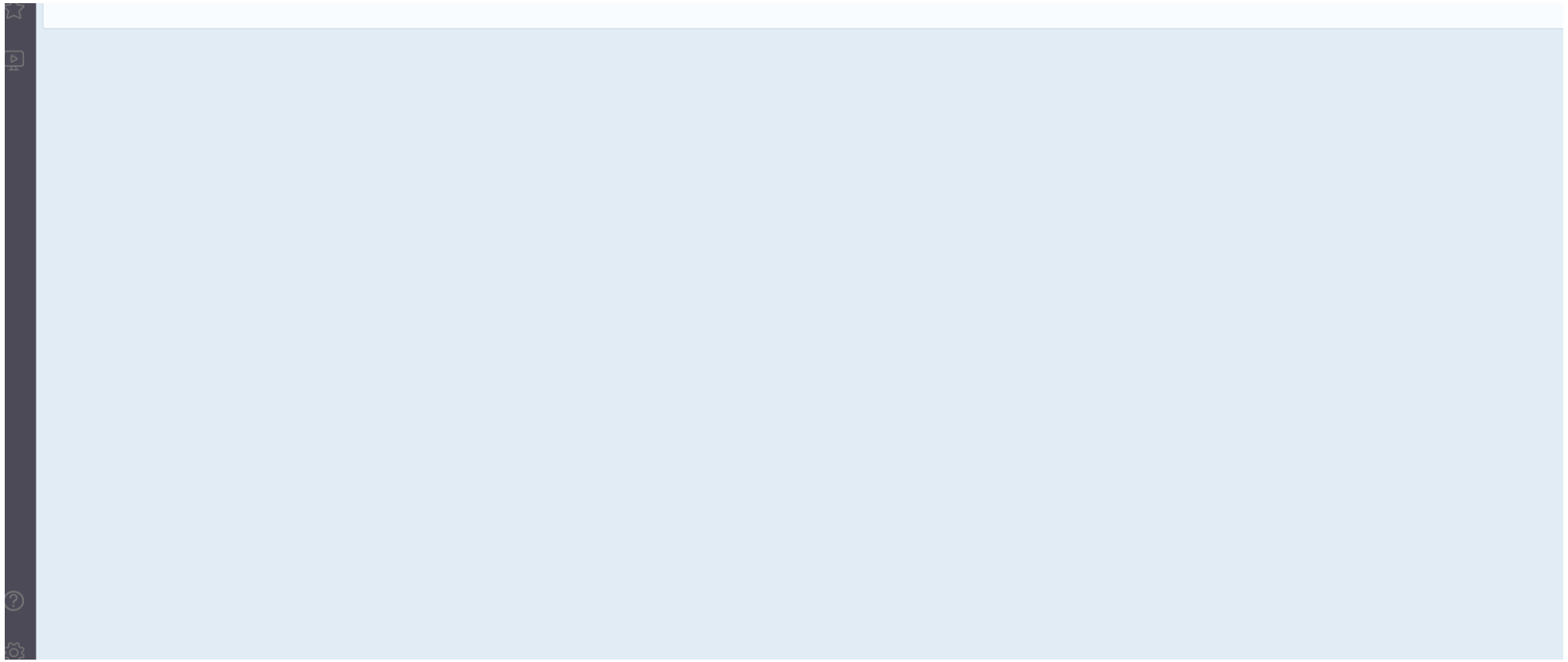
Causal Graph-based Data Science




Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

# 2. How to construct a Causal KG?

**Step 1. Constructing and loading a unified taxonomy**



# 2. How to construct a Causal KG?

## Step 2. Adding theoretical/causal assertions

```
neo4j$ match p = (predictor) - [c:causes] → (outcome) return predictor.name, outcome.name, c.theory
```

| | predictor.name | outcome.name |
|---|---|---|
| 1 | "firm behavior on the community" | "stock market-based performance" |
| 2 | "firm behavior on the community" | "survey-based financial performance" |
| 3 | "firm behavior on the community" | "market growth-based performance" |
| 4 | "firm behavior on the community" | "accounting-based performance" |
| 6 | "firm behavior on the environment" | "accounting-based performance" |
| 8 | "firm behavior on the environment" | "market growth-based performance" |

Started streaming 54 records in less than 1 ms and completed after 3 ms.

```
neo4j$ match p = (predictor) - [c:causes] → (outcome) return p
```


Overview

**Node labels**

* (17) construct (17)

**Relationship Types**

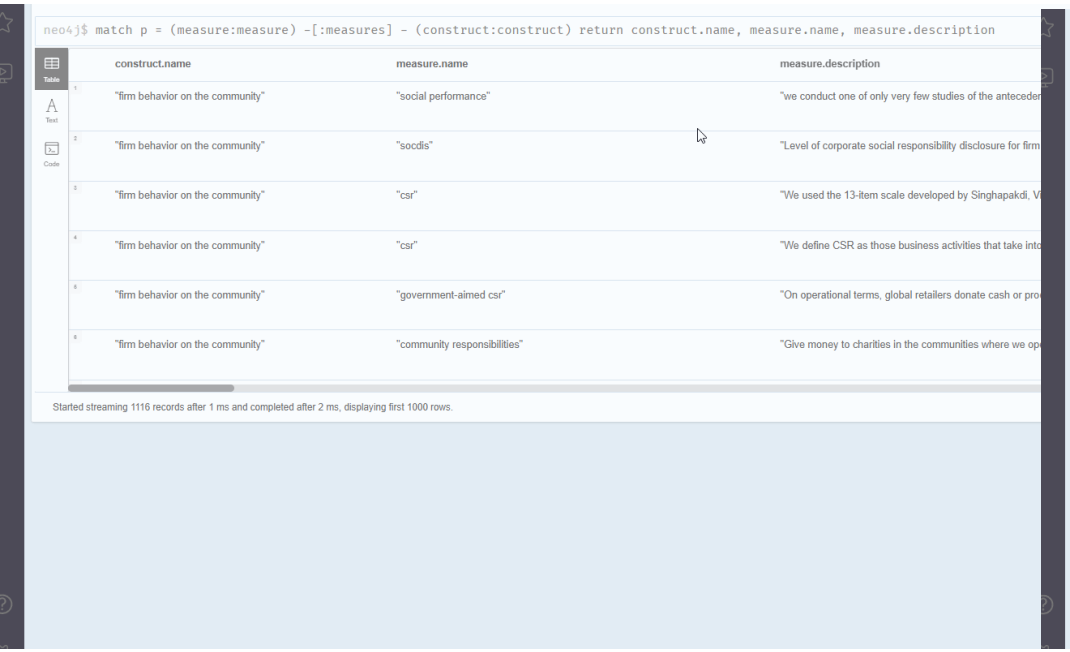
* (54) causes (54)

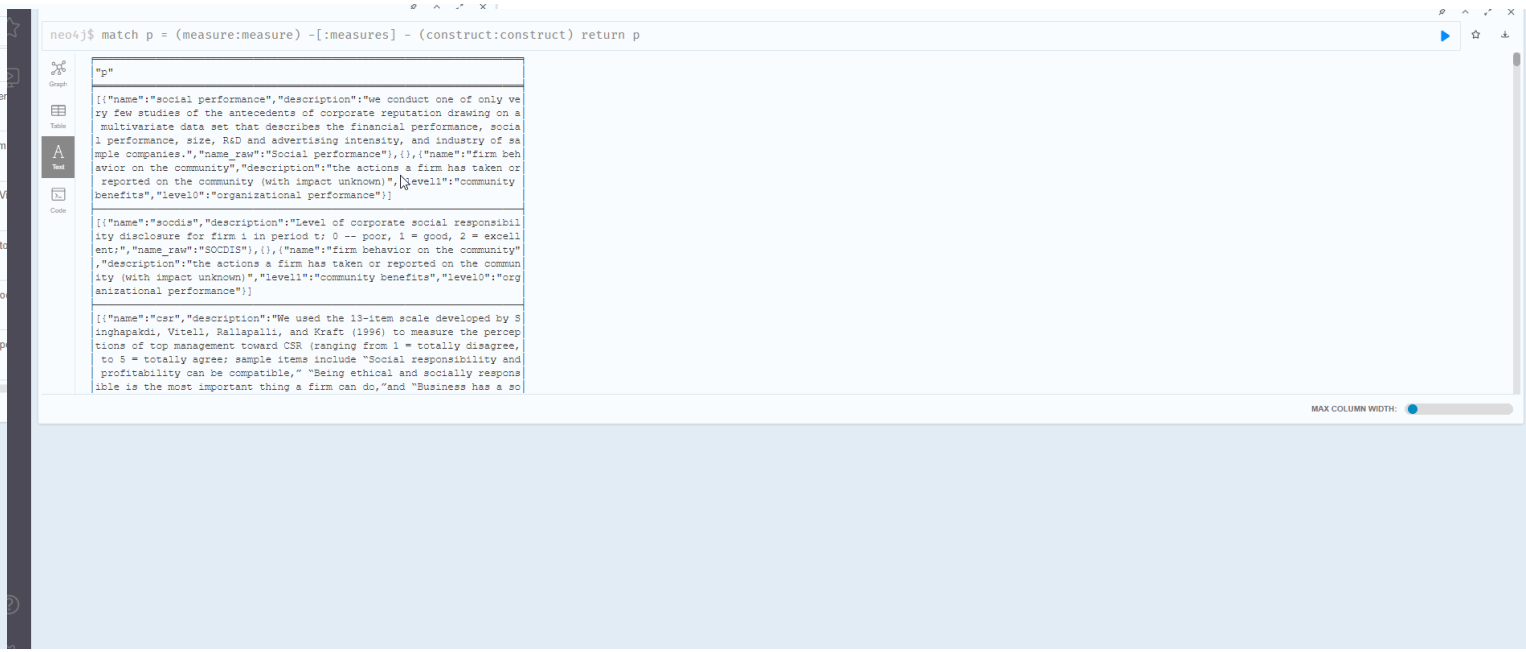
Displaying 17 nodes, 54 relationships.

Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

# 2. How to construct a Causal KG?

## Step 3. Linking constructs with measures/variables

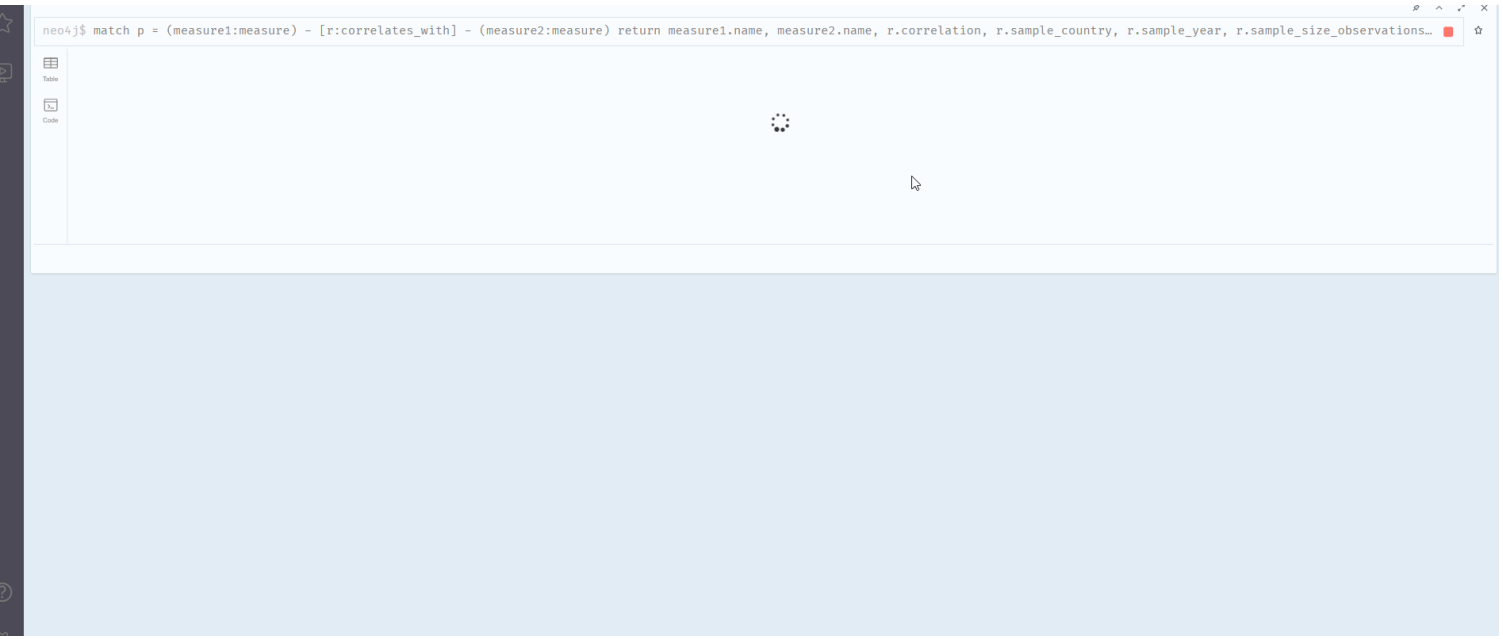


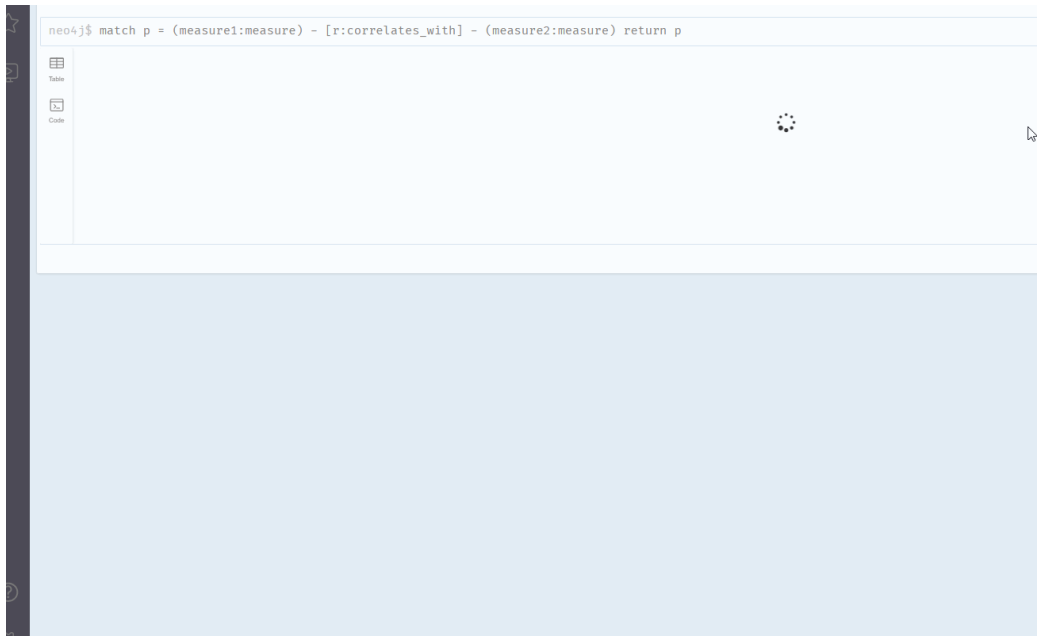




Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

# 2. How to construct a Causal KG?

## Step 4. Adding statistical evidence and sample meta data

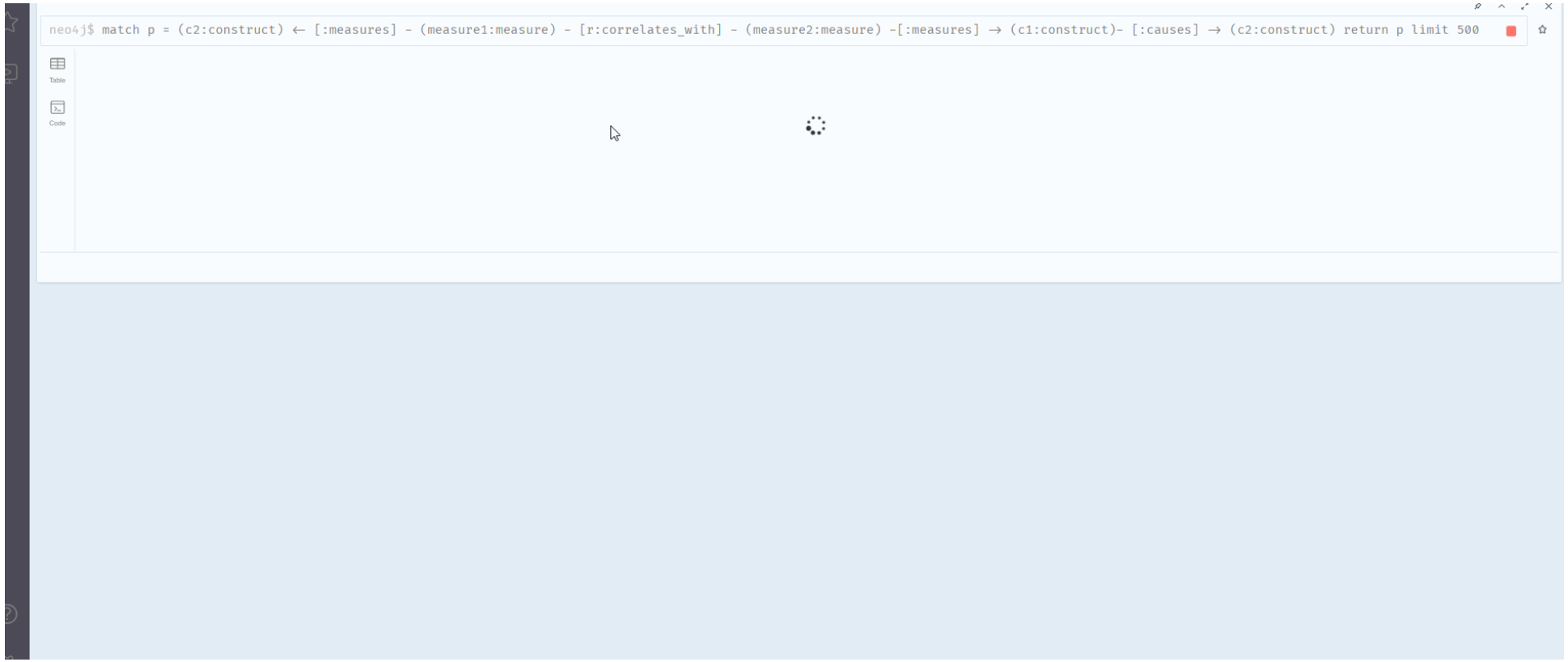


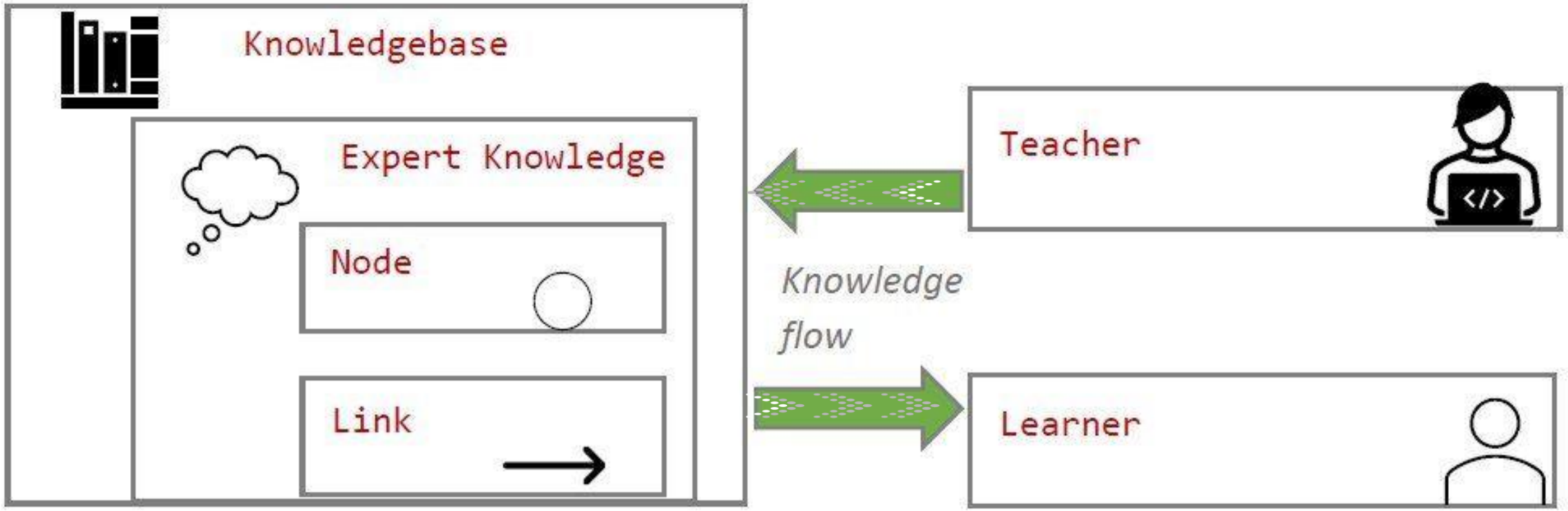

Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

# 2. How to construct a Causal KG?

**Here you go! A complete causal knowledge graph**

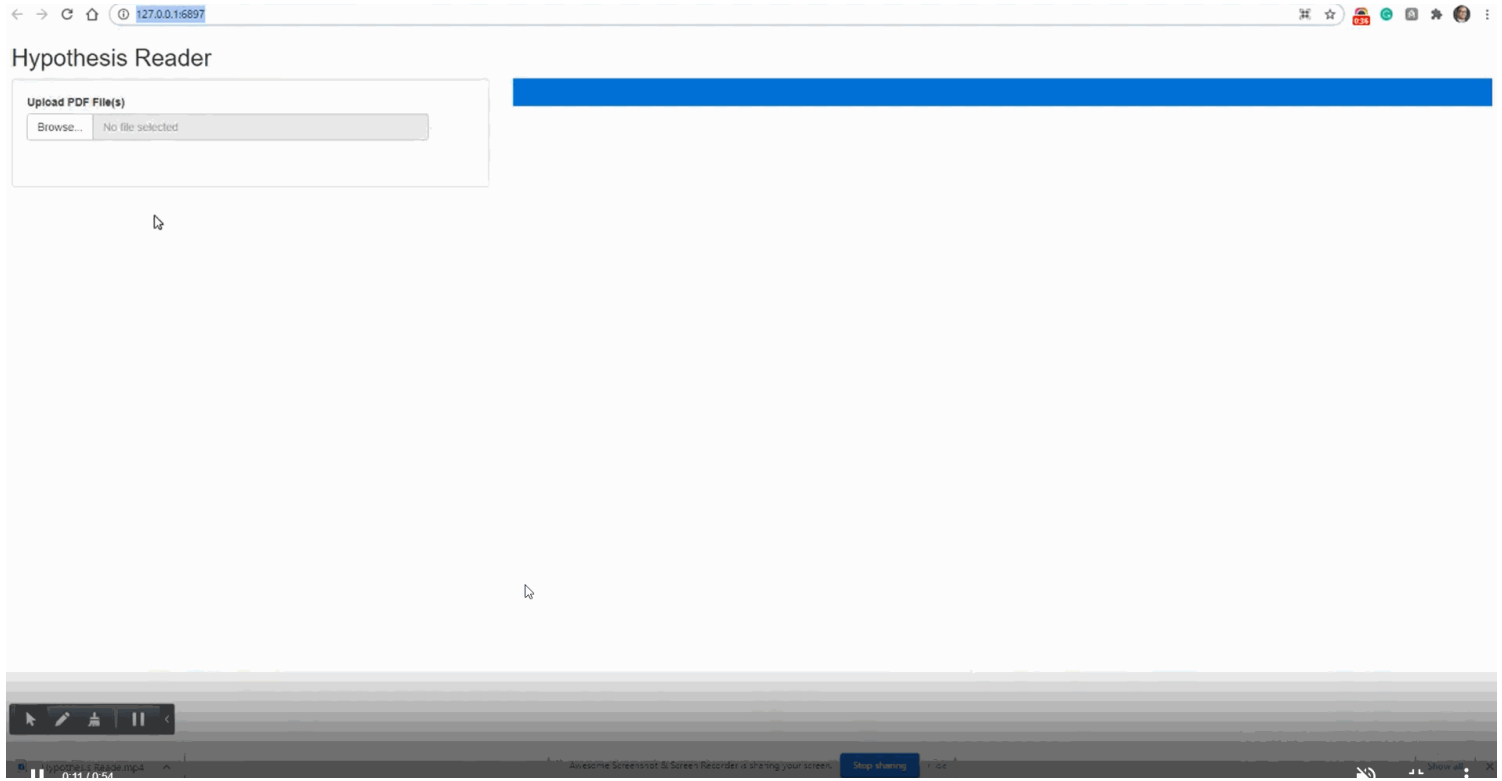


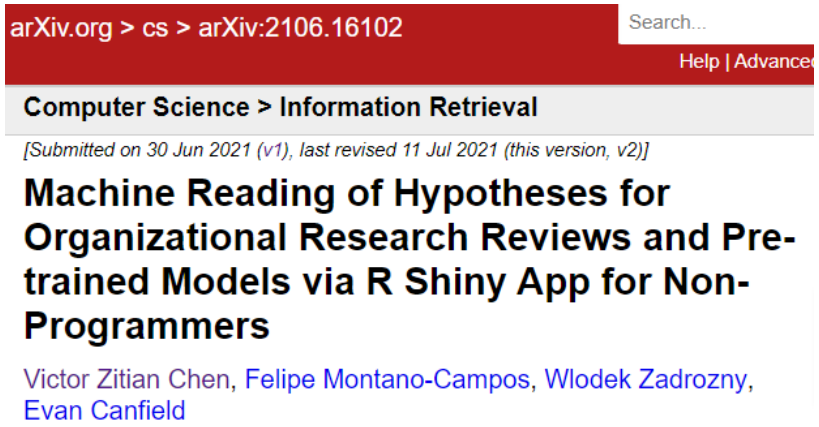
# 2. How to construct a Causal KG? An Expert Community Approach



# 2. How to construct a Causal KG? A Machine Learning Approach

## 1. Machine Reading of Hypotheses



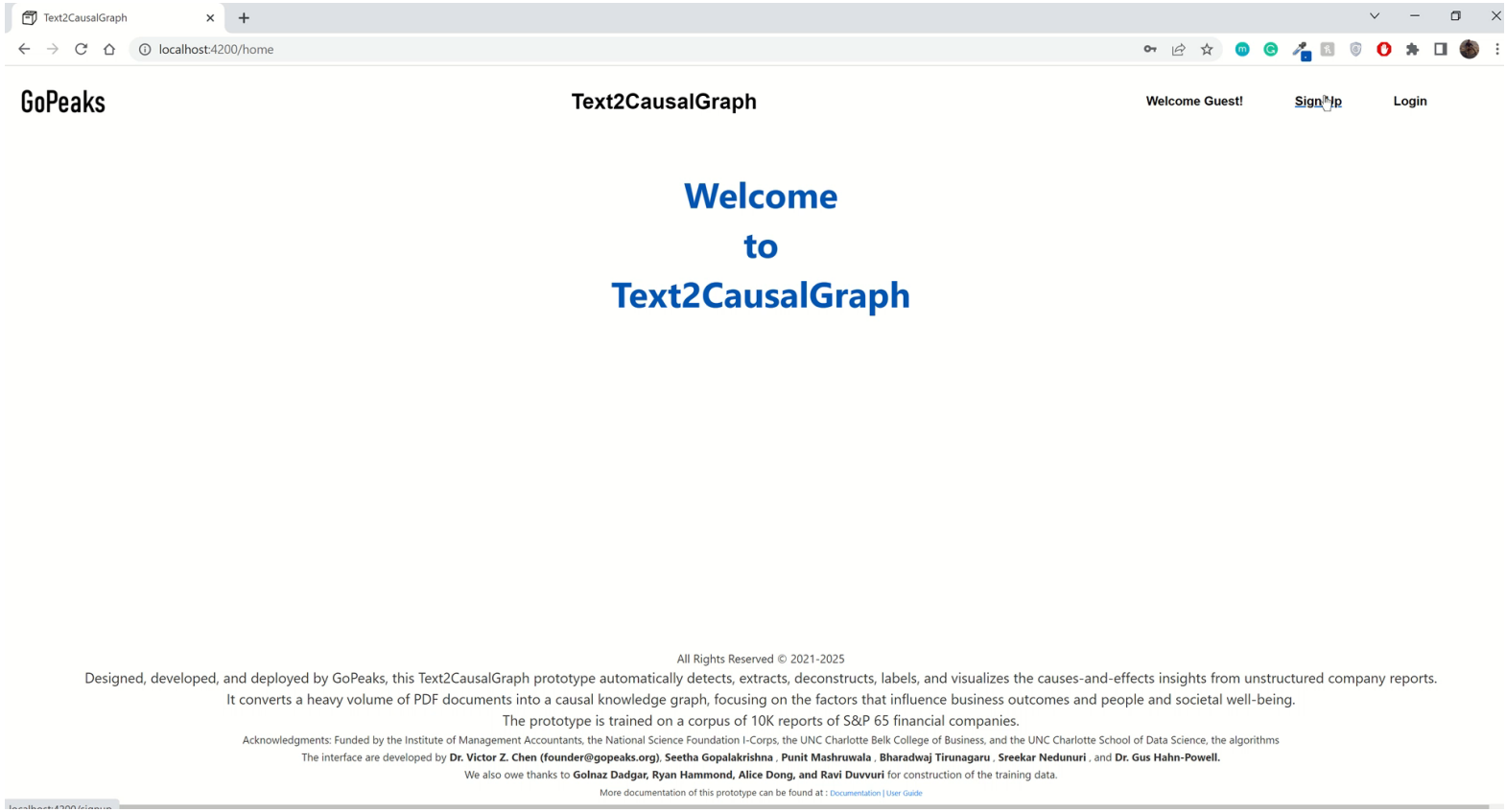
# 2. How to construct a Causal KG? A Machine Learning Approach

## 2. Entity Classification

| | Model | P | R | F1 |
|---|---|---|---|---|
| Engineered features | XGBoost | 0.74 | 0.71 | 0.72 |
| | Multinomial naïve Bayes | 0.80 | 0.81 | 0.80 |
| | SVM | 0.81 | 0.81 | 0.81 |
| | Logistic regression | 0.81 | 0.81 | 0.81 |
| Latent features | FastText | 0.84 | 0.79 | 0.81 |
| | Fine-tuned BERT | 0.59 | 0.51 | 0.40 |
| | SpaCy embeddings with LR | 0.74 | 0.75 | 0.75 |
| | Elmo embeddings with LR | 0.78 | 0.78 | 0.78 |

# 2. How to construct a Causal KG? A Machine Learning Approach

## 3. Converting Text into KG Database

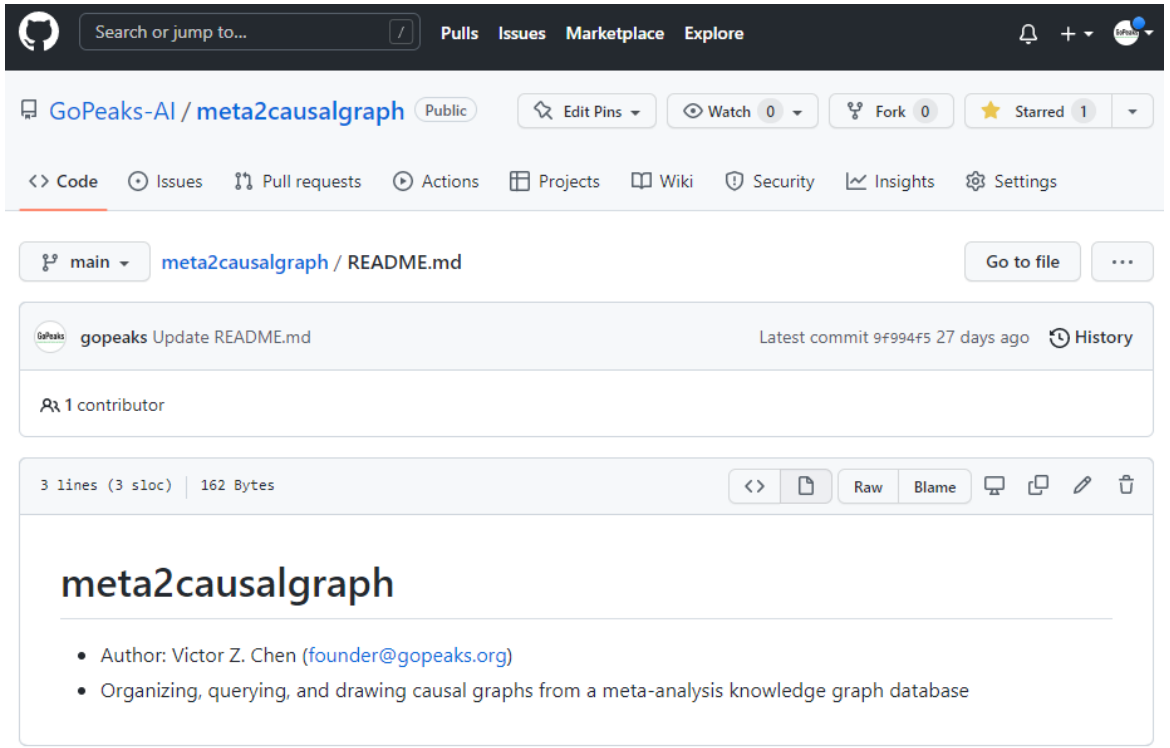



Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

# 2. How to construct a Causal KG?

A demonstration using Python and Neo4j

https://github.com/GoPeaks-AI/meta2causalgraph

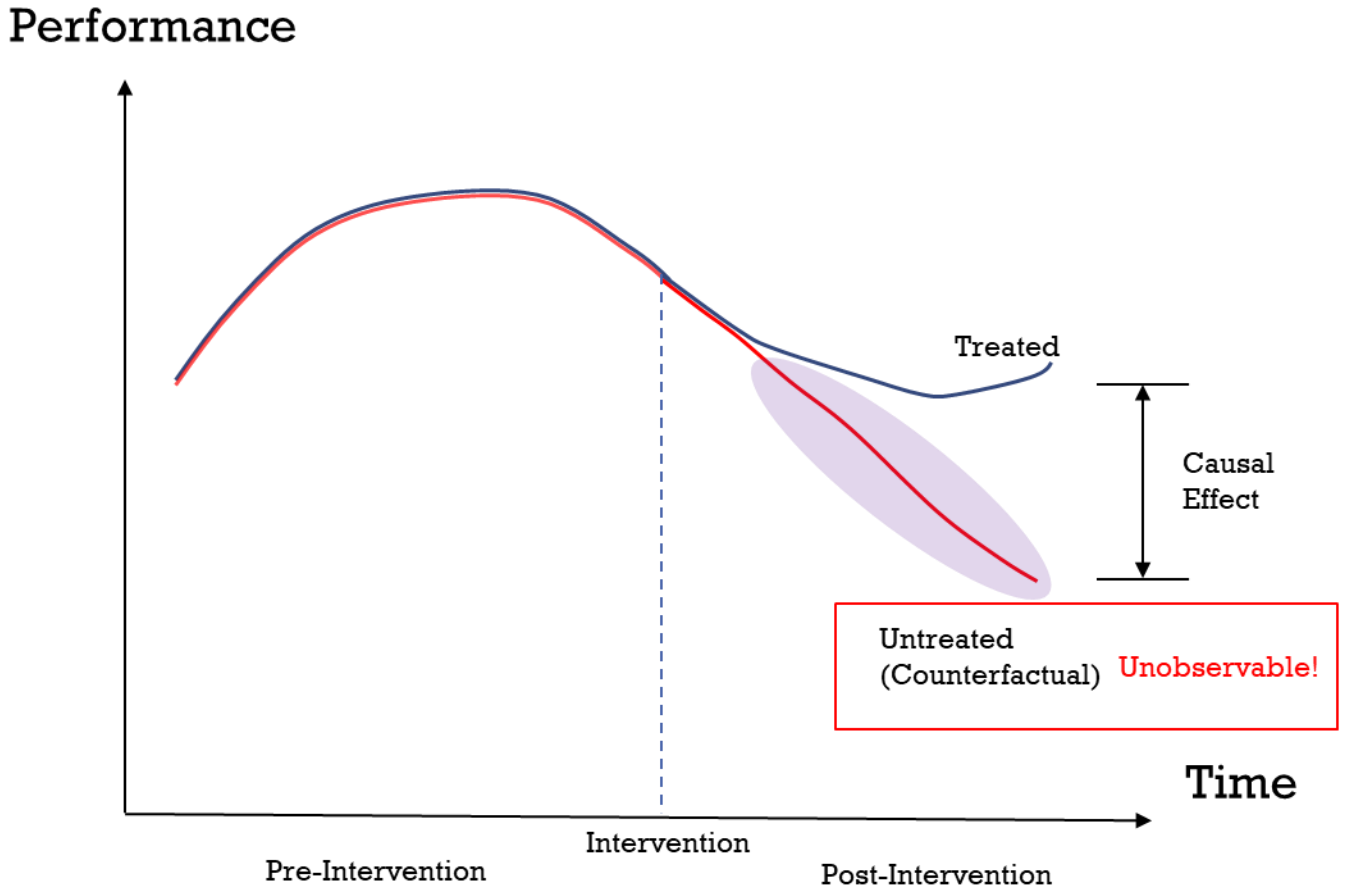



Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

# 3. How to infuse Causal KG into prescriptive analytics?

The essence of causal analysis:

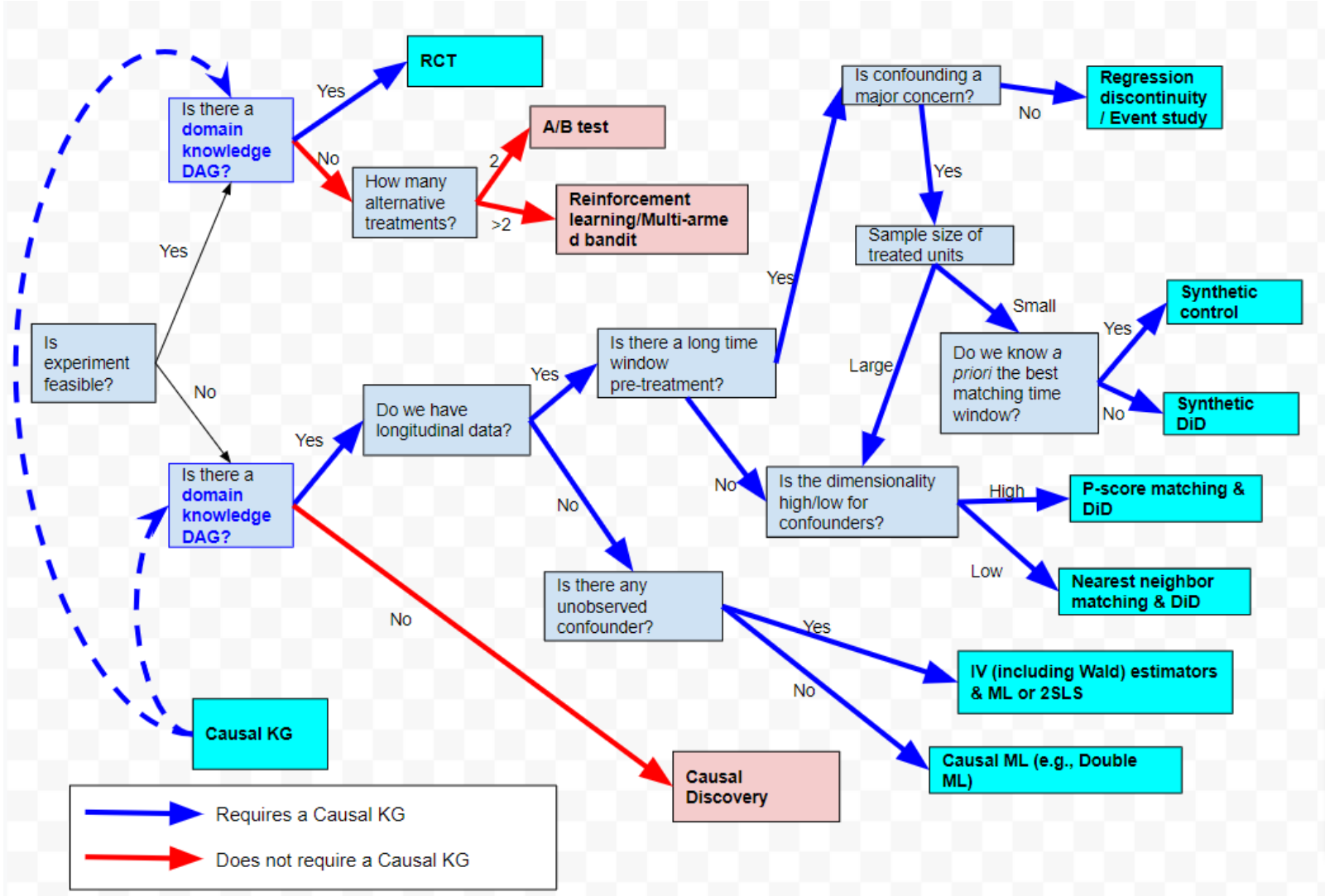
- Self-counterfactual comparison
  - Balancing: Similar distribution in covariates
  - Debiasing: Removing the confounding



# 3. How to infuse Causal KG into prescriptive analytics?

Causal KG may enable and improve many causal analytics methods by:

- Identifying what covariates need to be balanced
- Identifying what confounding biases there are in observational data

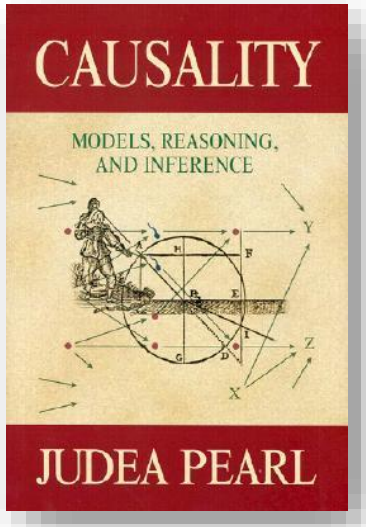


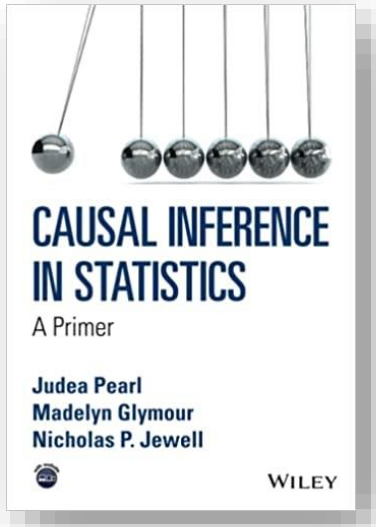

Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

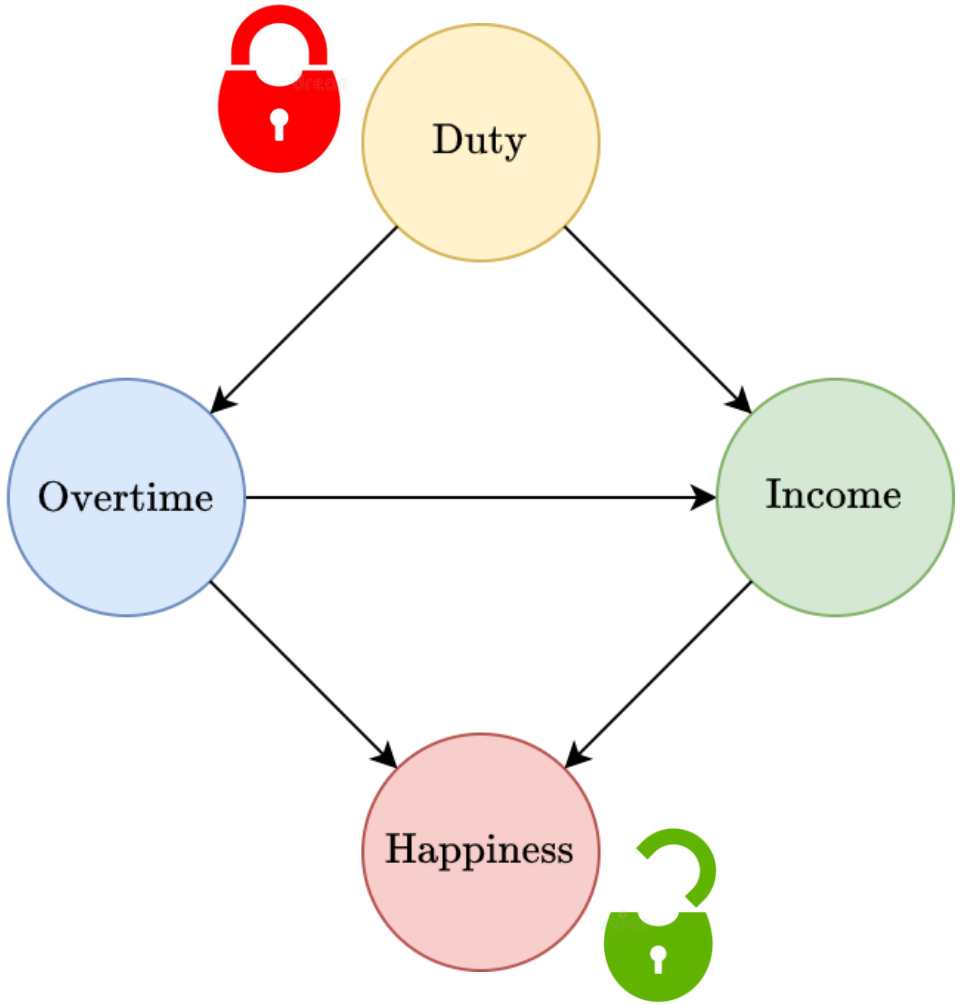
# 3. How to infuse Causal KG into prescriptive analytics?

Theory: Do Calculus

e.g., back-door closing, front-door opening, no unobserved confounders, etc.



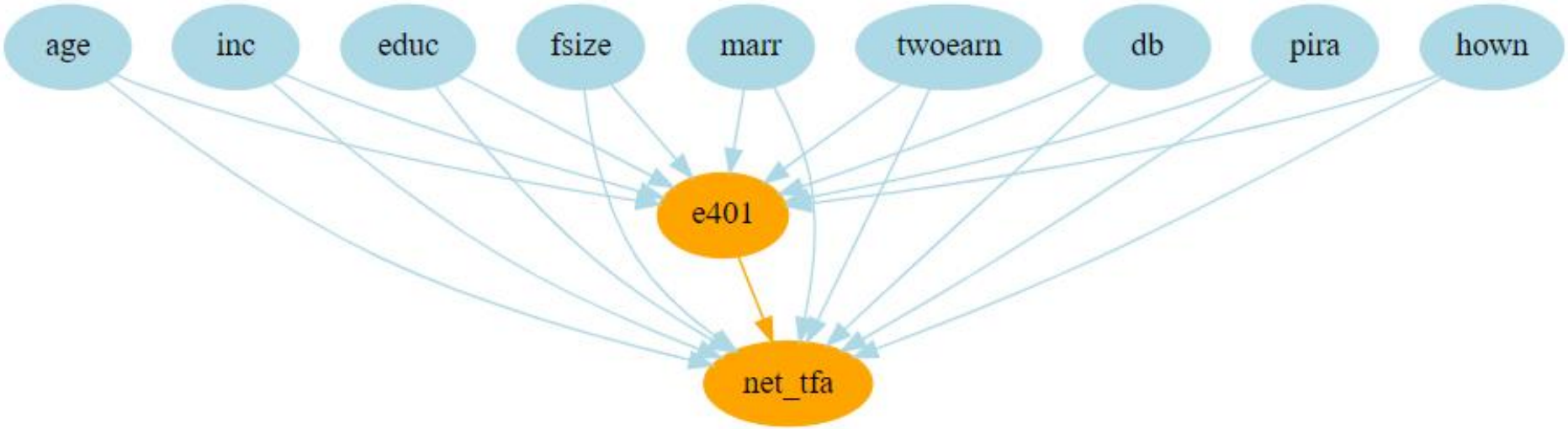




Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.

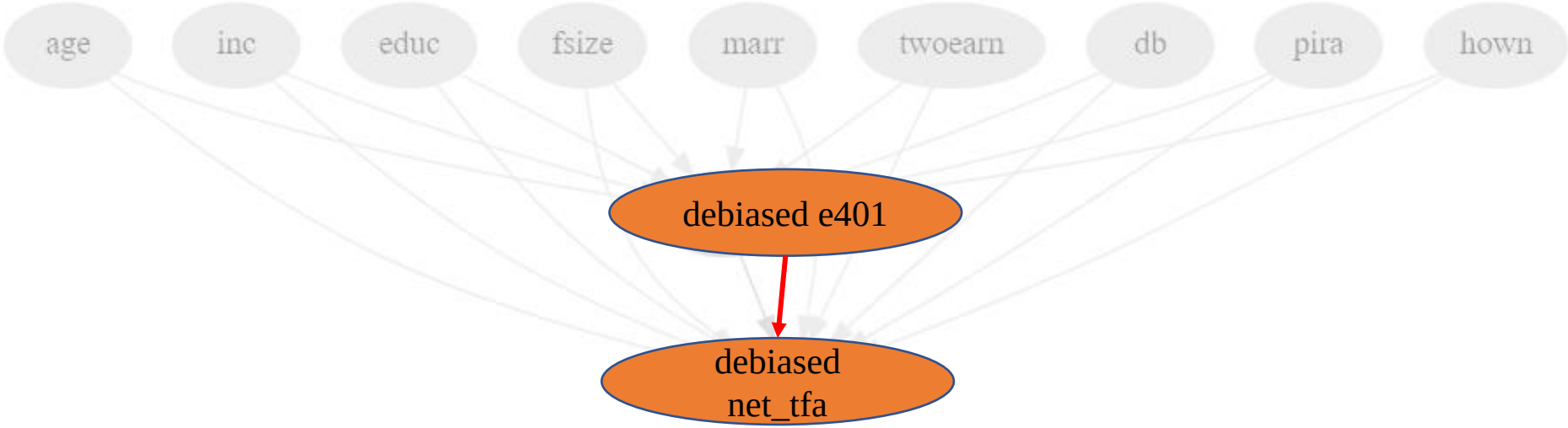
# 3. How to infuse Causal KG into prescriptive analytics?

A demonstration using Jupyter Notebook and Neo4j



# 3. How to infuse Causal KG into prescriptive analytics?

## A demonstration using Jupyter Notebook and Neo4j



# Want to learn more?

GoPeaks
A SHARED BRAIN FOR PEOPLE AND SOCIETY

GoPeaks.org

GitHub

github.com/GoPeaks-AI

towards
data science

@victorzitianchen

Disclaimer: This talk does not represent the view of Fidelity Investments. It is based on my previous research at GoPeaks, funded by the National Science Foundation and the Institute of Management Accountants. Please visit www.GoPeaks.org for more information and readings.