

THE WORLD IN KNOWLEDGE GRAPHS

Key Questions

What can we learn from textual data?

What are expressive data representations?

Why do LLMs enhance the analysis of text?

Different kinds of research



Qualitative

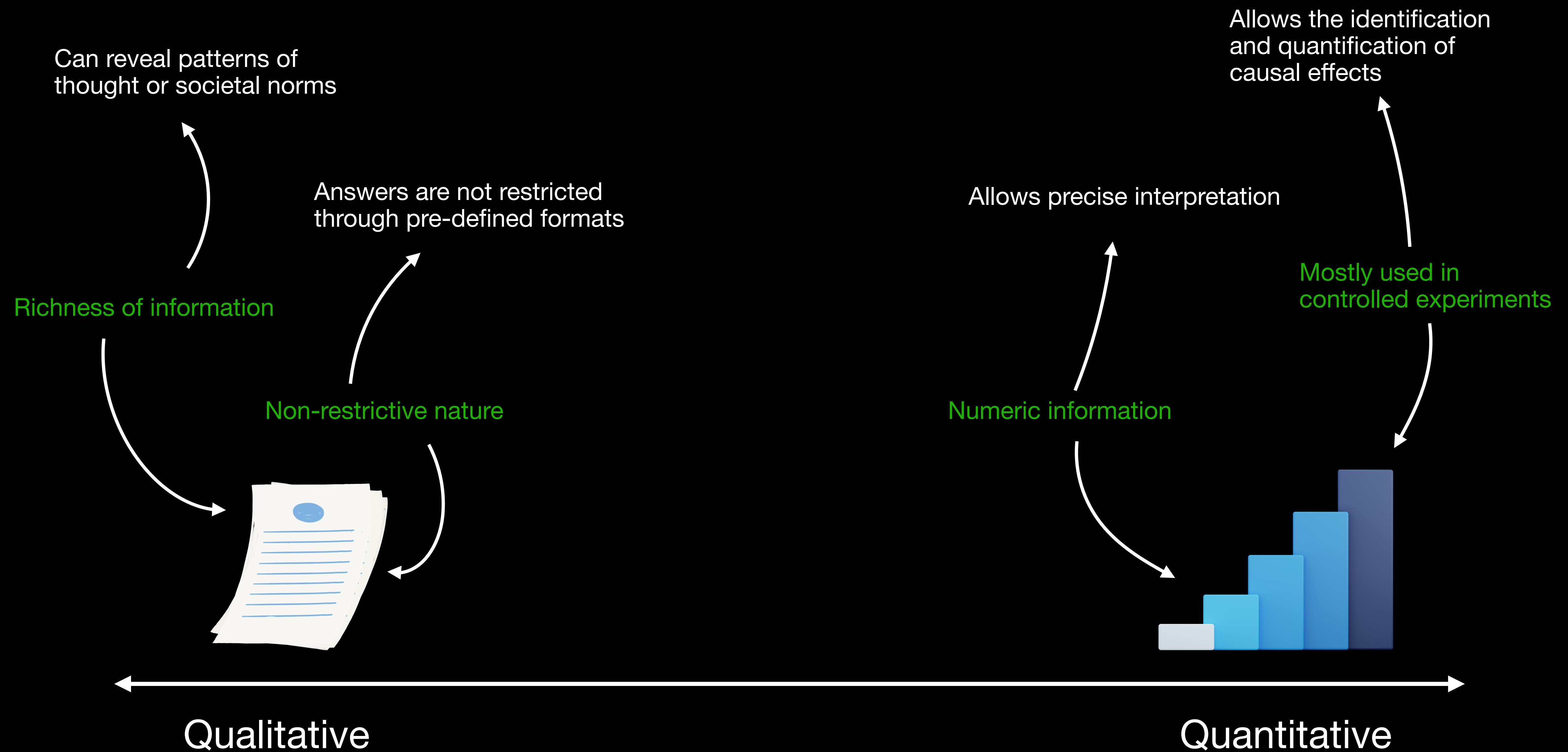
Quantitative

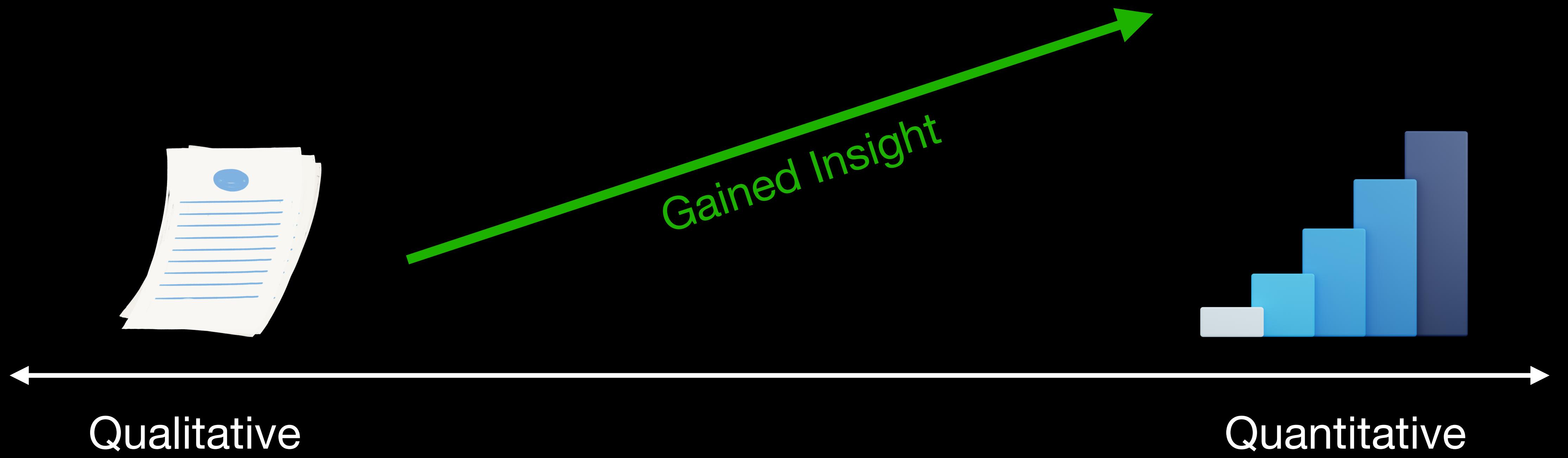
The primary goal is knowledge acquisition

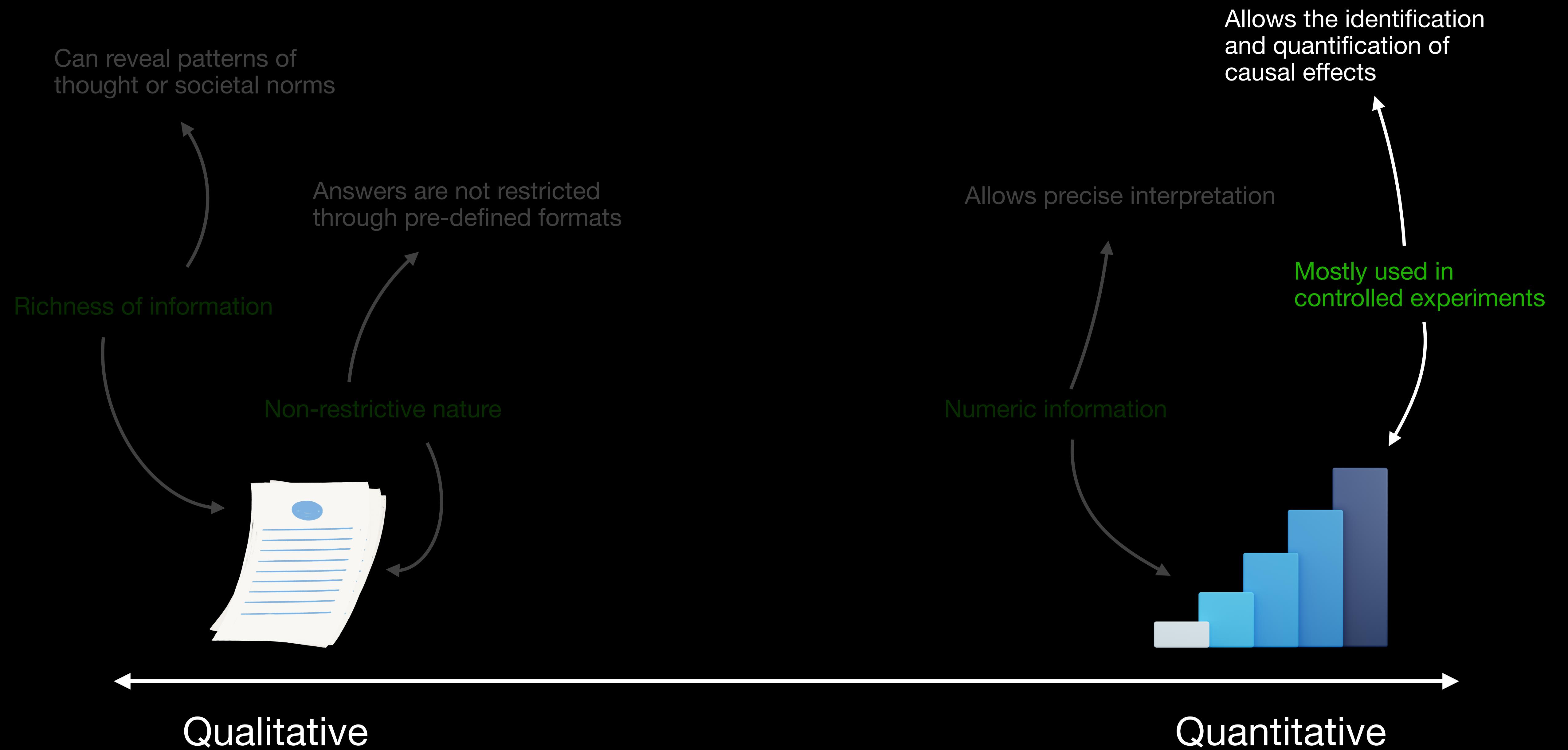


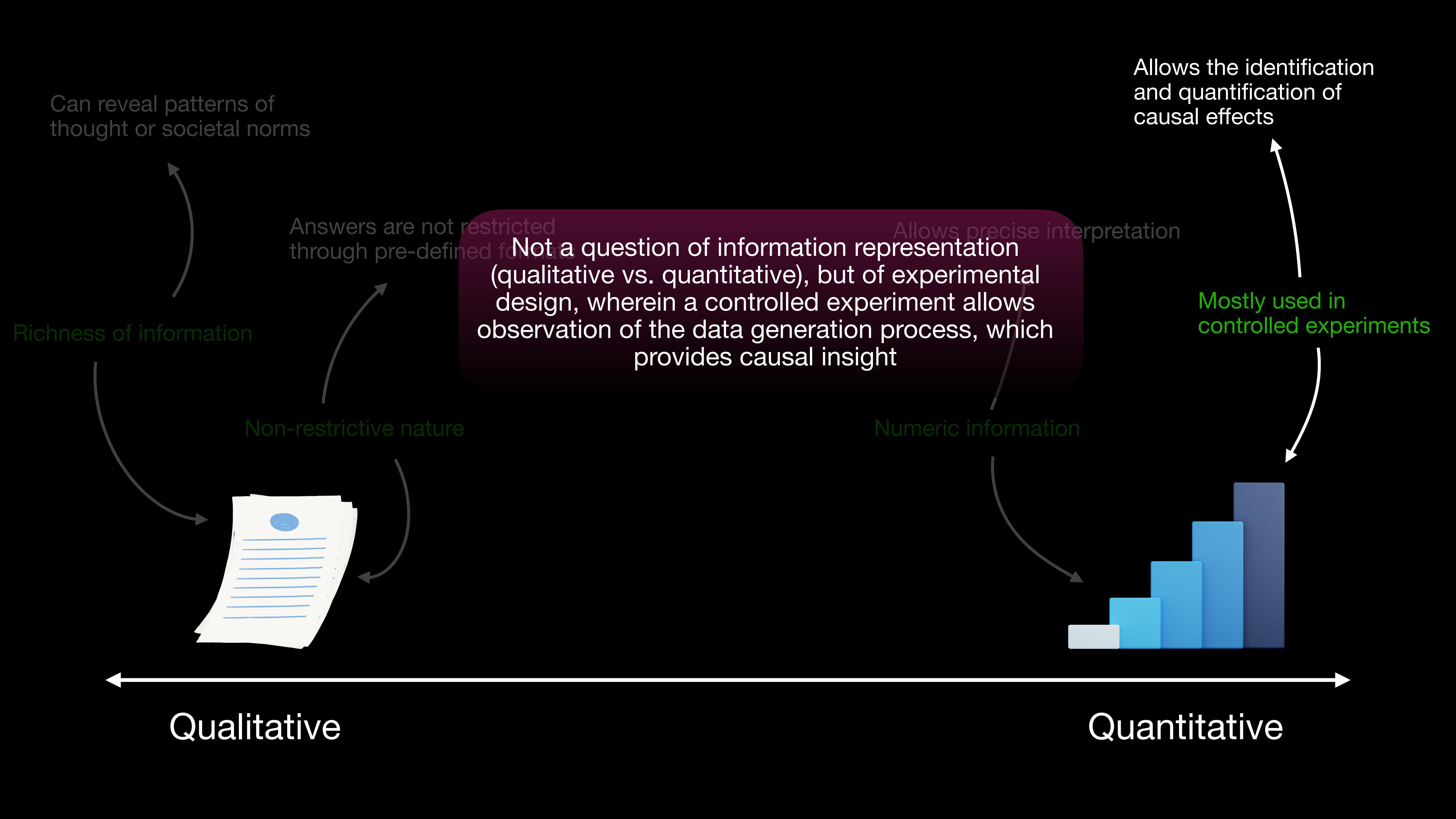
Qualitative

Quantitative









Qualitative and *quantitative* data differ in their **representation**



Different **representations** open up different levels of analysis



Causal insight does not depend on the representation but knowing or inferring the **data generation process**



*Extracting meaning from text
requires converting the **chaotic**
to **expressive representations***

Extracting meaning from text
requires converting the chaotic
to expressive representations*

Extracting interpretable (and quantifiable) information

Text as Data

Working with qualitative data at scale

*... so what is an
expressive representation
of the information about the world*

Ladder of Understanding Text

Ladder of Understanding Text

"While many environmentalists praise Senator Miller for her ambitious climate policies, highlighting her commitment to reducing carbon emissions by investing in renewable energy that highly supports emission decline, critics argue that her plans are unrealistic and could harm the economy, citing the potential loss of jobs in traditional energy sectors and the high costs of transitioning to green technology. Further, they argue that nuclear energy also leads to emission decline. In response, Miller asserts that the long-term benefits of a sustainable environment outweigh the short-term economic challenges, arguing that job growth in renewable energy industries and cost savings from reduced health impacts of pollution will offset the initial expenses. She further states that nuclear energy undermines renewable solutions."

Sentiment Analysis

"While many environmentalists praise Senator Miller for her ambitious climate policies, highlighting her commitment to reducing carbon emissions by investing in renewable energy that highly supports emission decline, critics argue that her plans are unrealistic and could harm the economy, citing the potential loss of jobs in traditional energy sectors and the high costs of transitioning to green technology. Further, they argue that nuclear energy also leads to emission decline. In response, Miller asserts that the long-term benefits of a sustainable environment outweigh the short-term economic challenges, arguing that job growth in renewable energy industries and cost savings from reduced health impacts of pollution will offset the initial expenses. She further states that nuclear energy undermines renewable solutions."

Sentiment Analysis

Positive

"While many environmentalists praise Senator Miller for her ambitious climate policies, highlighting her commitment to reducing carbon emissions by investing in renewable energy that highly supports emission decline, critics argue that her plans are unrealistic and could harm the economy, citing the potential loss of jobs in traditional energy sectors and the high costs of transitioning to green technology. Further, they argue that nuclear energy also leads to emission decline. In response, Miller asserts that the long-term benefits of a sustainable environment outweigh the short-term economic challenges, arguing that job growth in renewable energy industries and cost savings from reduced health impacts of pollution will offset the initial expenses. She further states that nuclear energy undermines renewable solutions."

Positive

Negative

Negative

Topic Modeling

"While many environmentalists praise Senator Miller for her ambitious climate policies, highlighting her commitment to reducing carbon emissions by investing in renewable energy that highly supports emission decline, critics argue that her plans are unrealistic and could harm the economy, citing the potential loss of jobs in traditional energy sectors and the high costs of transitioning to green technology. Further, they argue that nuclear energy also leads to emission decline. In response, Miller asserts that the long-term benefits of a sustainable environment outweigh the short-term economic challenges, arguing that job growth in renewable energy industries and cost savings from reduced health impacts of pollution will offset the initial expenses. She further states that nuclear energy undermines renewable solutions."

Topic Modeling

Topic 1: Climate policies and their goals

Topic 2: Economic impact of the policies

Topic 3: Job market implications

Topic 4: Costs of green technology transition

Topic 5: Long-term benefits of sustainability

Topic 6: Health impacts of pollution

Argument Mining

"While many environmentalists praise Senator Miller for her ambitious climate policies, highlighting her commitment to reducing carbon emissions by investing in renewable energy that highly supports emission decline, critics argue that her plans are unrealistic and could harm the economy, citing the potential loss of jobs in traditional energy sectors and the high costs of transitioning to green technology. Further, they argue that nuclear energy also leads to emission decline. In response, Miller asserts that the long-term benefits of a sustainable environment outweigh the short-term economic challenges, arguing that job growth in renewable energy industries and cost savings from reduced health impacts of pollution will offset the initial expenses. She further states that nuclear energy undermines renewable solutions."

Argument Mining

Main Claim (Pro): Senator Miller's climate policies are beneficial

Premise 1: Renewable engird highly reduces emissions

Premise 2: Investing in renewable energy

Counterclaim 1 (Con): Her plans are unrealistic and could harm the economy

Premise 1: Potential loss of jobs in traditional energy sectors

Premise 2: High costs of transitioning to green technology

Counterclaim 2 (Con): Nuclear energy is good for the environment

Premise 1: Nuclear energy leads to emission decline

Rebuttal Claim: The long-term benefits of renewable engird outweigh the short-term economic challenges

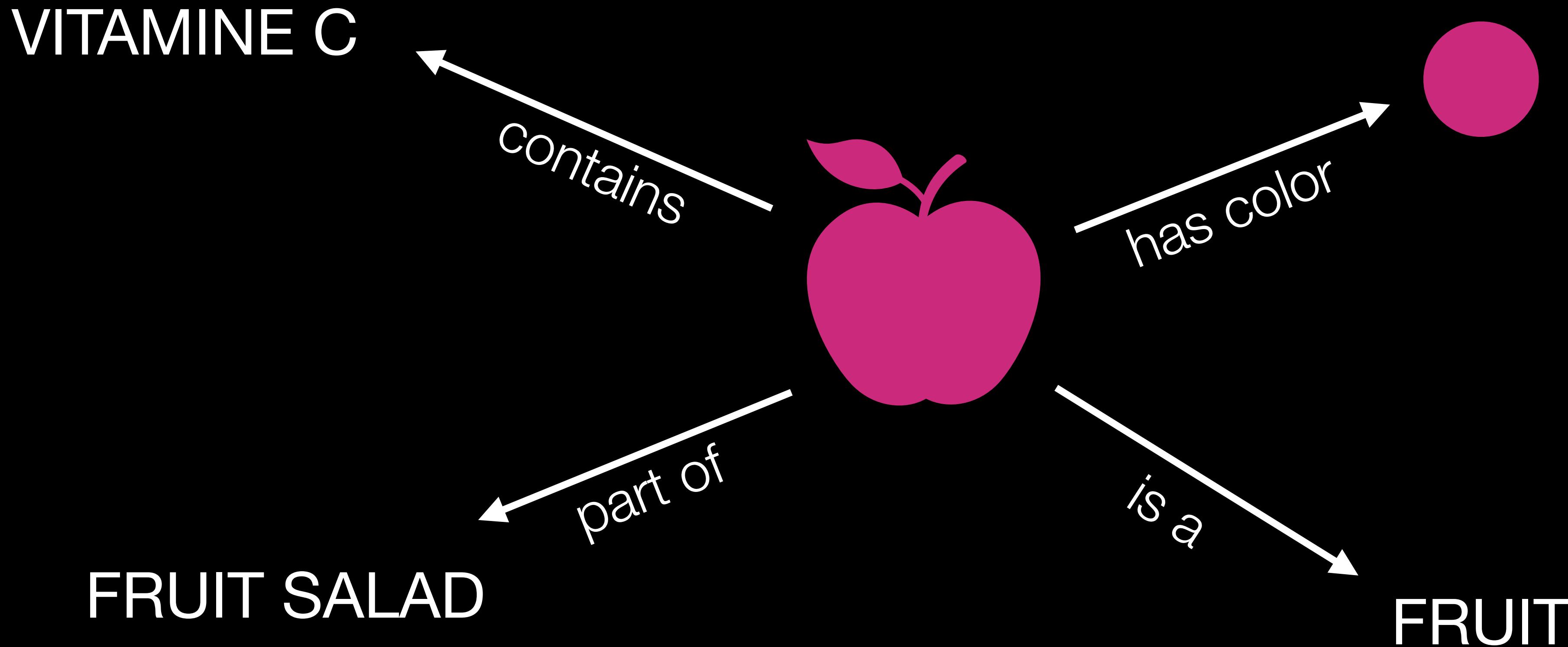
Premise 1: Job growth in renewable energy industries

Premise 2: Cost savings from reduced health impacts of pollution

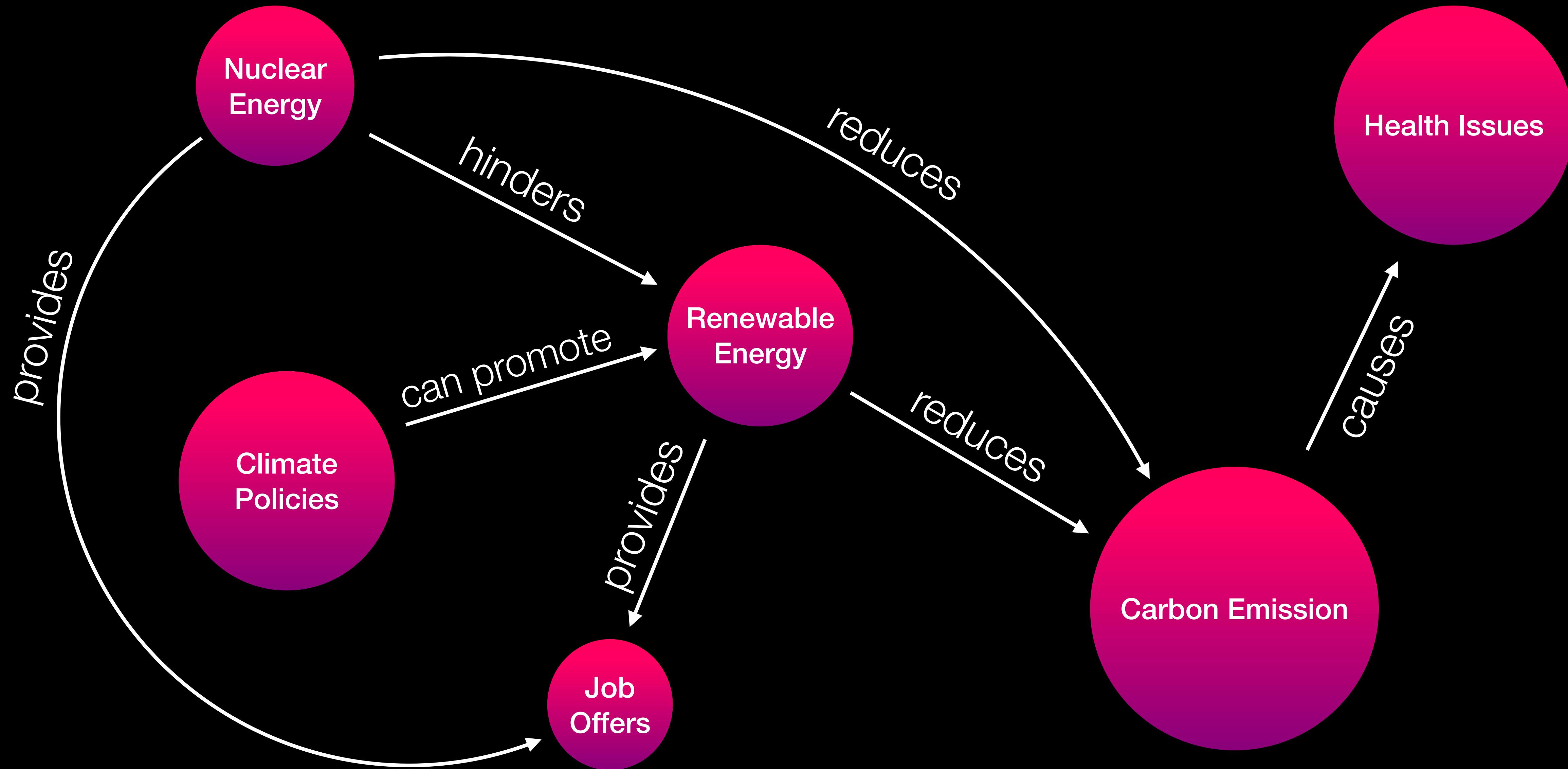
Rebuttal Claim 2: Nuclear energy hinders renewable solutions

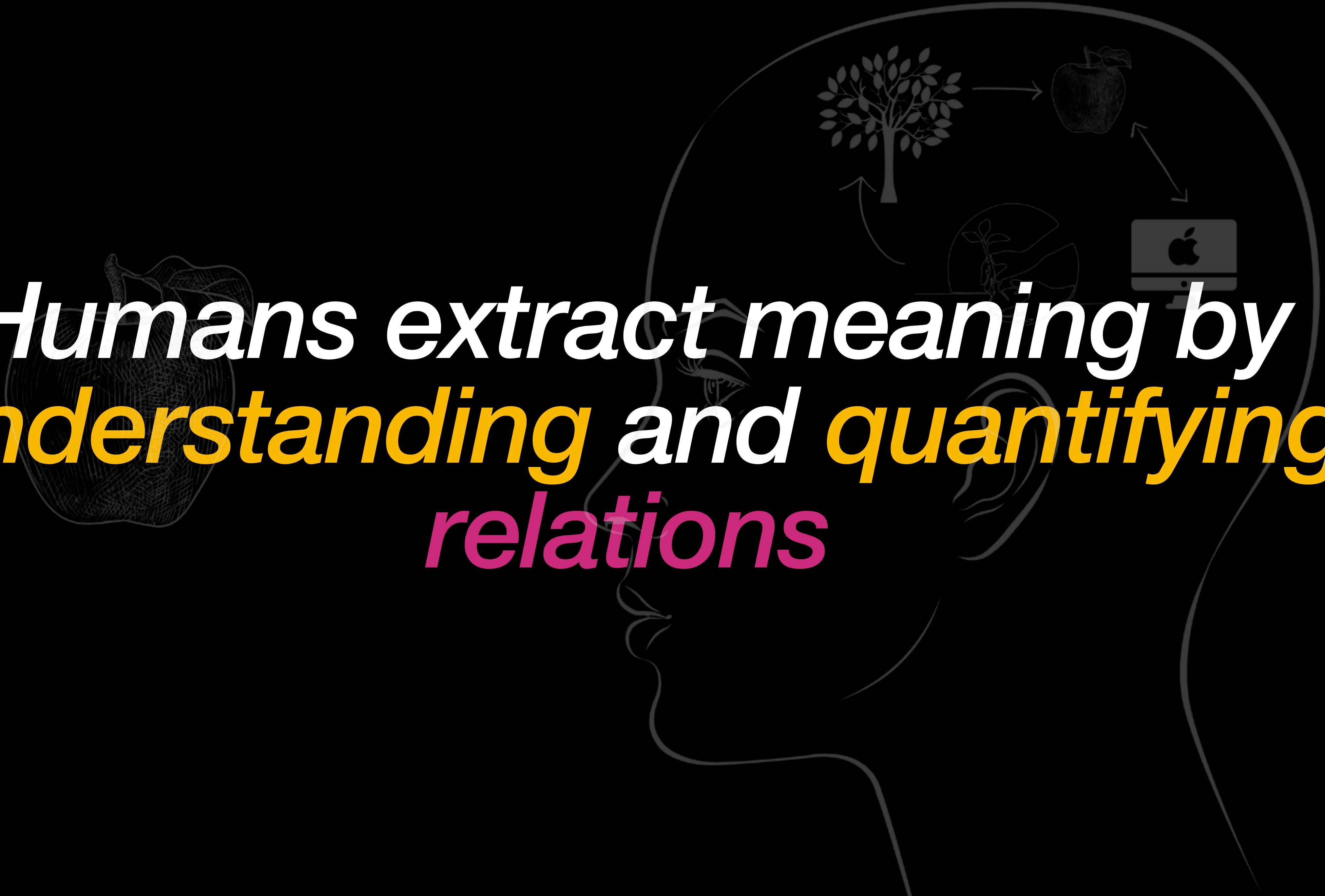
*... what is even more
expressive?*

Looking at how the world is organized ...

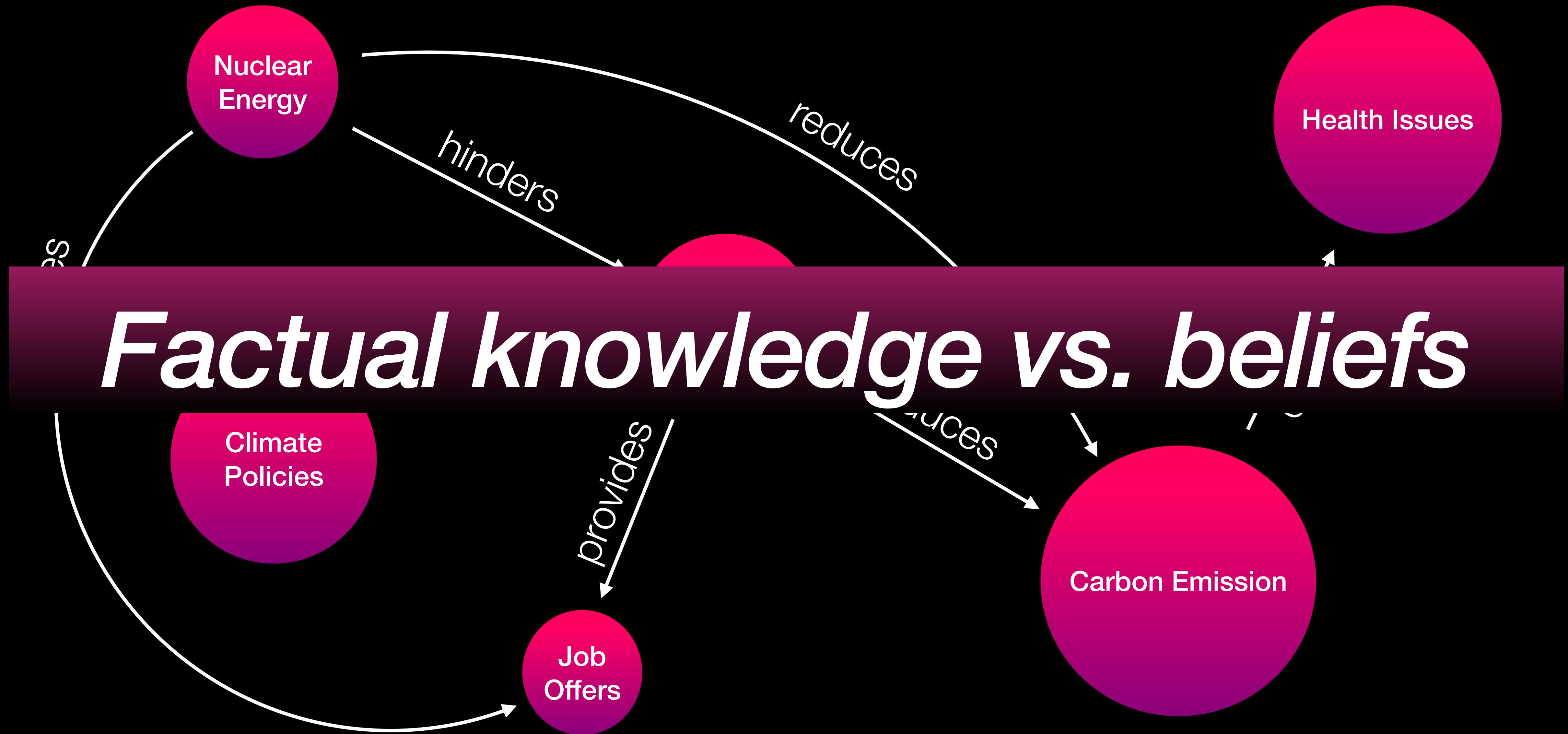


*The world can be represented as concepts (nodes)
and their relations (edges)*

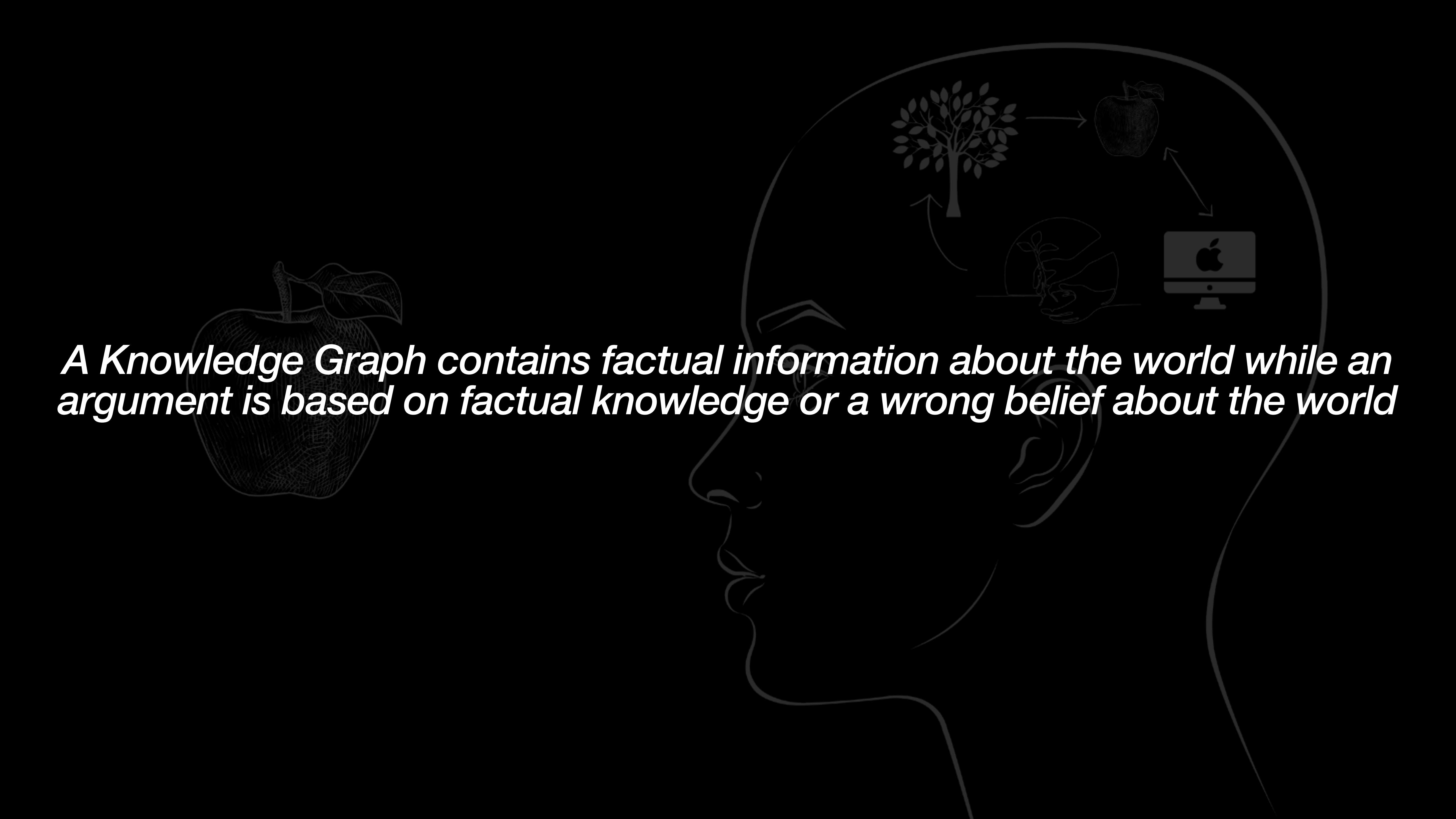




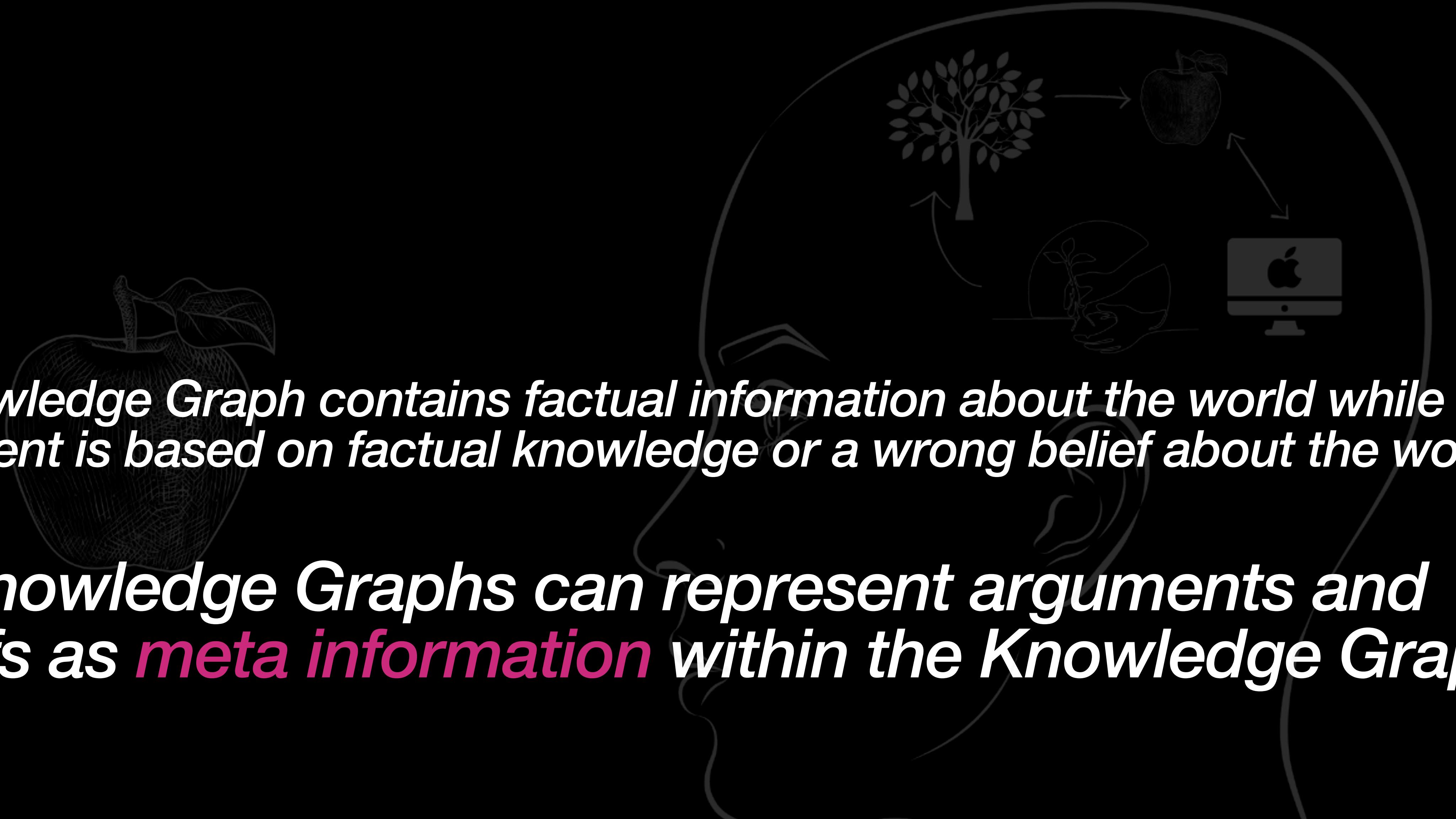
*Humans extract meaning by
understanding and quantifying
relations*



*Graphs allow modeling the **perception** of the world and the **actuality** of how concepts (entities) relate to each other*

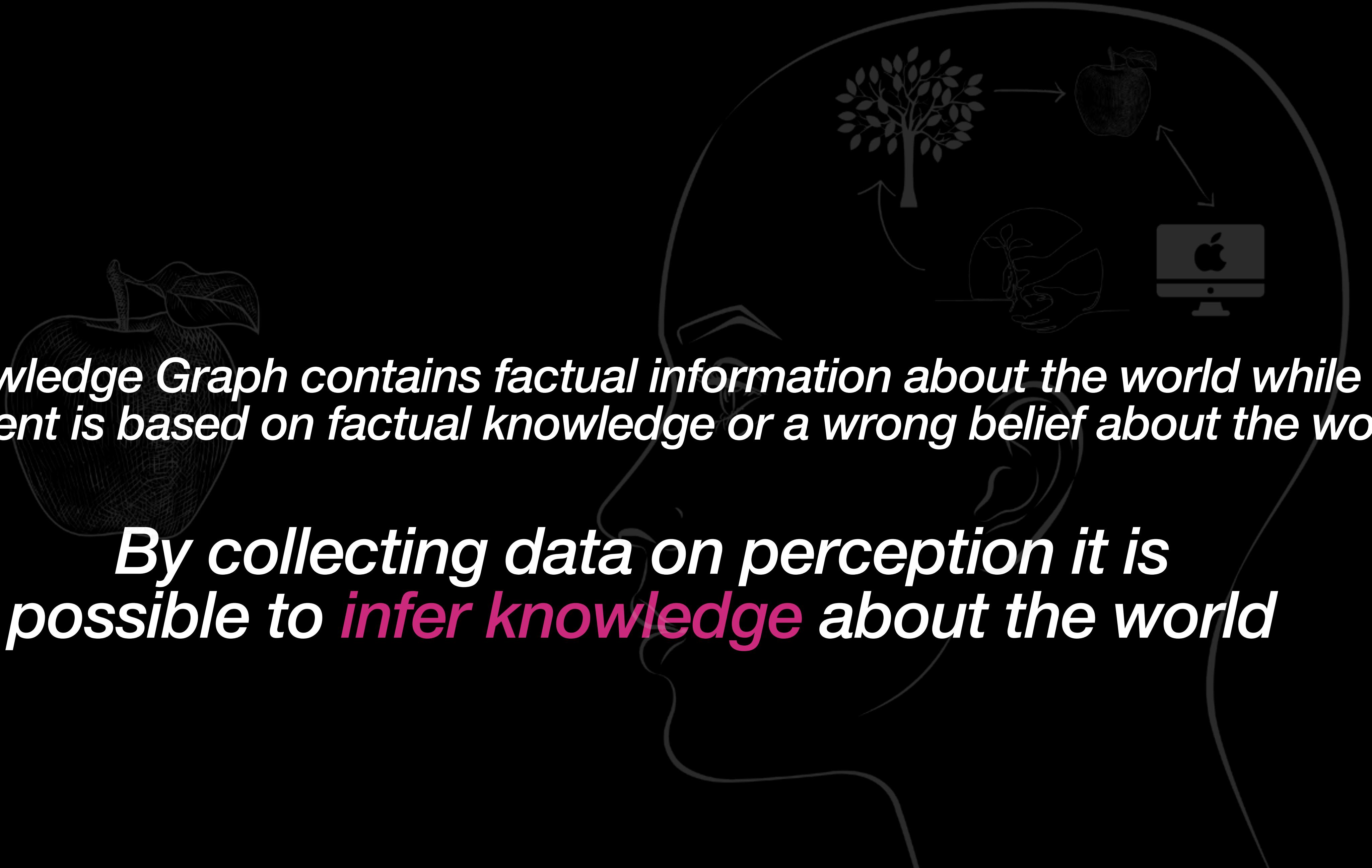


A Knowledge Graph contains factual information about the world while an argument is based on factual knowledge or a wrong belief about the world



A Knowledge Graph contains factual information about the world while an argument is based on factual knowledge or a wrong belief about the world

*Knowledge Graphs can represent arguments and beliefs as **meta information** within the Knowledge Graph*



A Knowledge Graph contains factual information about the world while an argument is based on factual knowledge or a wrong belief about the world

By collecting data on perception it is possible to infer knowledge about the world

Deep Dive into Knowledge Graphs

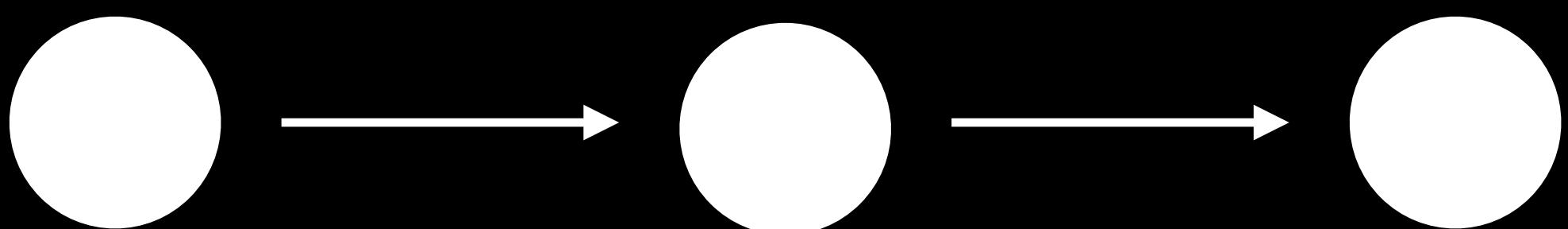


What is a Knowledge Graph

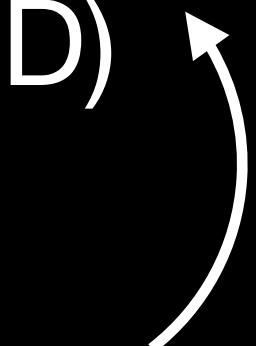
In knowledge representation and reasoning, a **knowledge graph** is a knowledge base that uses a graph-structured data model or topology to represent and operate on data. Knowledge graphs are used to store **interlinked descriptions of entities** – objects, events, situations or abstract concepts – while also encoding the free-form semantics or relationships underlying these entities.

What is NOT a Knowledge Graph

Directed Acyclic Graphs (DAG)



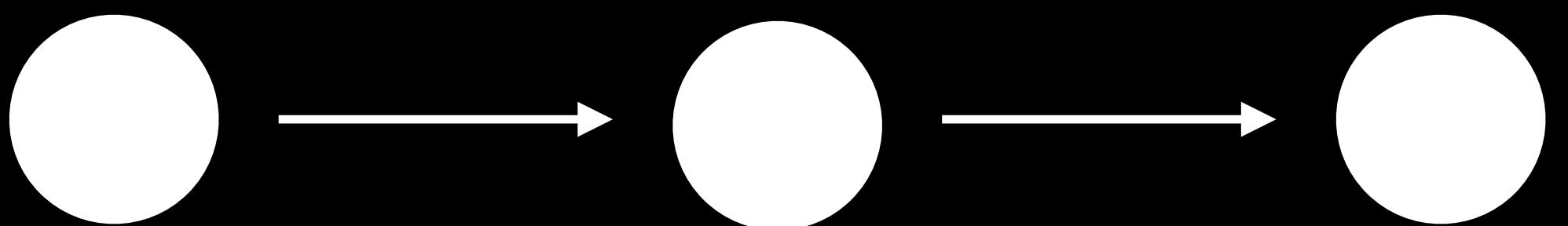
A key feature of a KG is that entity descriptions should be interlinked to one another. The definition of one entity includes another entity. This linking is how the graph forms. (e.g. A is B. B is C. C has D. A has D)



It might be sufficient for an application to just have a string ‘Italy’ associated with the string ‘GDP’ and a number ‘1.95 trillion’ without needing to define what countries are or what the ‘Gross Domestic Product’ of a country is. It’s the connections and the graph that make the KG, not the language used to represent the data.

What is NOT a Knowledge Graph

Directed Acyclic Graphs (DAG)



A key feature of a KG is that entity descriptions should be interlinked to one another. The definition of one entity includes another entity. This linking is how the graph forms. (e.g. A is B. B is C. C has D. A has D)

Knowledge bases without formal structure and semantics, e.g. Q&A “knowledge base” about a software product, also do not represent a KG

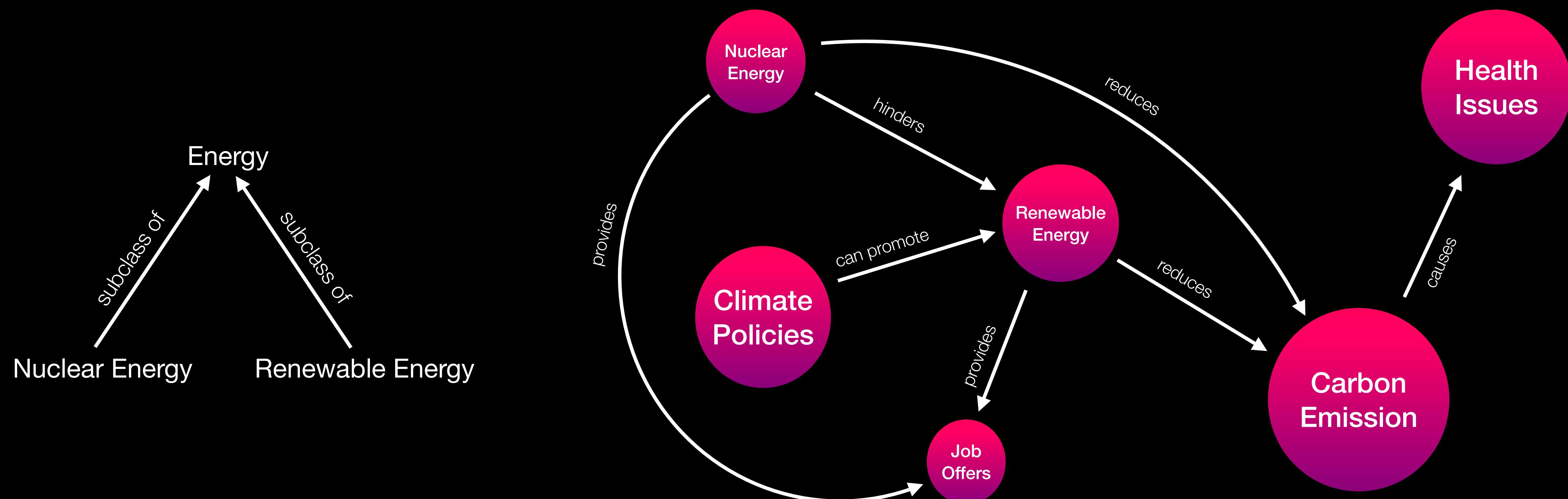
Knowledge Graph allow inference through formally defined semantics

What are formalized semantics?

What are Ontologies?

Web Ontology Language (OWL)

The W3C Web Ontology Language (OWL) is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things



Web Ontology Language (OWL)

:Person a owl:Class .

:Woman a owl:Class ; rdfs:subClassOf :Person . :

Parent a owl:Class ; rdfs:subClassOf :Person .

:Mother a owl:Class ;

rdfs:subClassOf [a owl:Class ;

owl:intersectionOf (:Woman :Parent)] .

Allows reasoning

:Person a owl:Class .

:hasParent a owl:ObjectProperty ; rdfs:domain :Person ; rdfs:range :Person .

:hasChild a owl:ObjectProperty ; owl:inverseOf :hasParent .

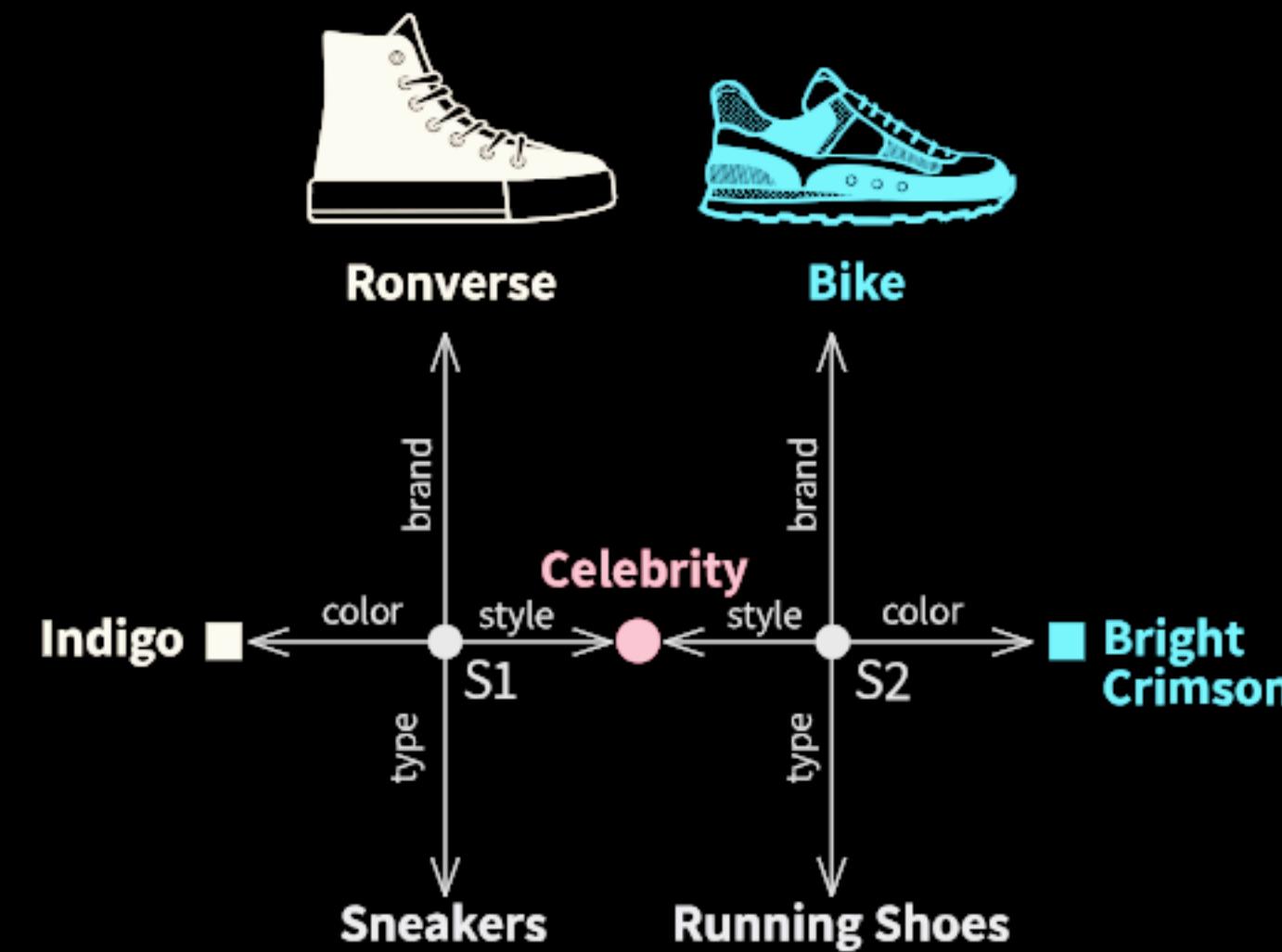
:hasGrandparent a owl:ObjectProperty ; owl:propertyChainAxiom (:hasParent :hasParent) .

:Alice a :Person . :Bob a :Person . :Carol a :Person .

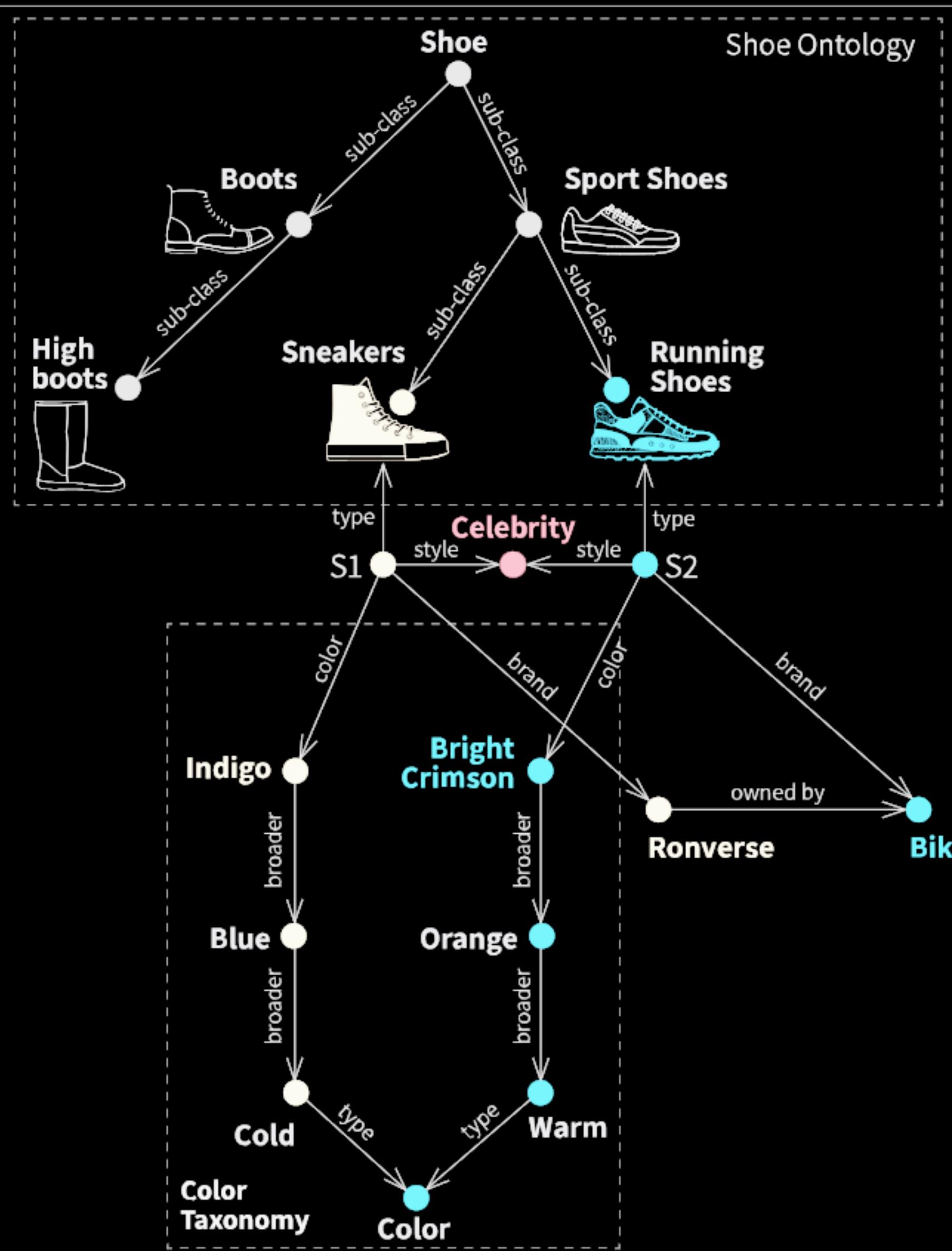
:Alice :hasParent :Bob . :Bob :hasParent :Carol .

Carol is the grandparent of Alice

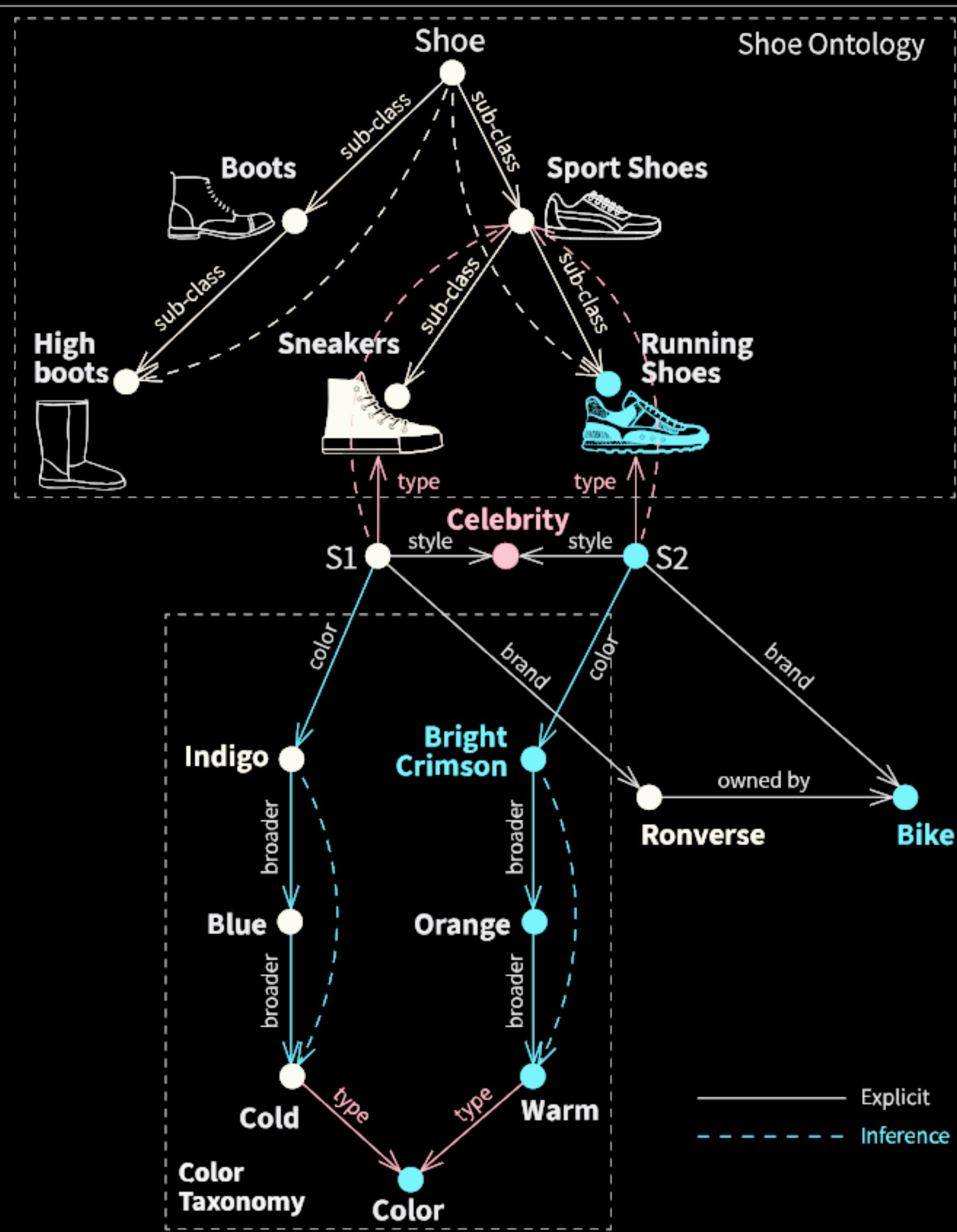
Plain Graph



Knowledge Graph



Knowledge Graph with Inference

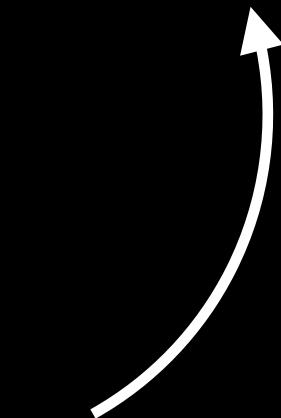


Explicit
Inference

Ontologies allow precise definitions of
the **semantics** of concepts

Knowledge Graph store semantically defined concepts with other data (strength of a relation, temporal data...) in **one knowledge representation**

Knowledge Graph store semantically defined concepts with other data (strength of a relation, temporal data...) in **one knowledge representation**



We can define a knowledge graph as a graph $G = (E, R)$ with entities $e \in E = E_1 \cup E_2 \cup \dots \cup E_n$ coming from formal structures E_i like ontologies and relations R .

The main purpose of ontologies (semantic web)
is to **precisely define information** and make
them **interoperable**

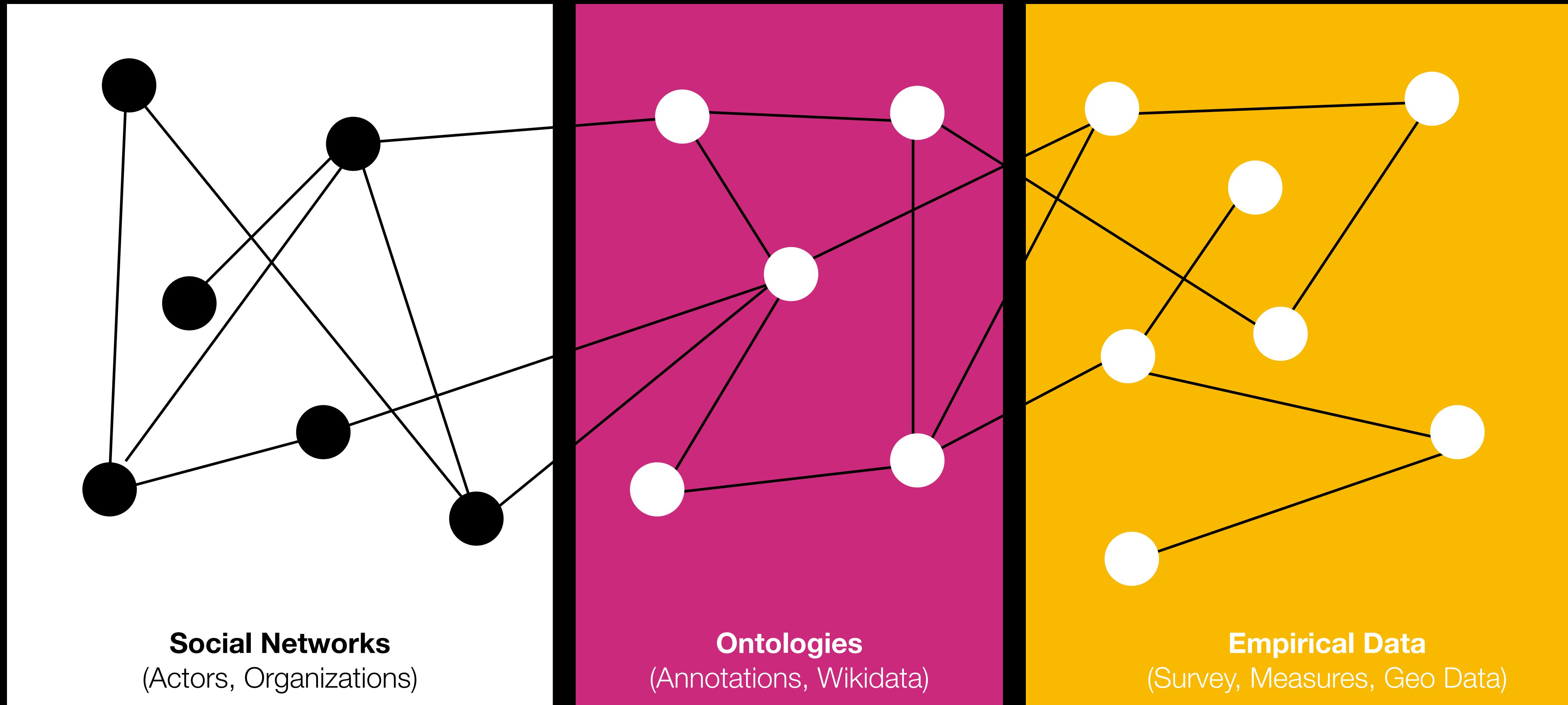
The main purpose of ontologies (semantic web)
is to **precisely define information** and make
them **interoperable**



It is thus a crucial concept on the way to store and retrieve linked data,
to generate knowledge and wisdom, to search within data, information
and knowledge and a prerequisite for a detailed data analysis

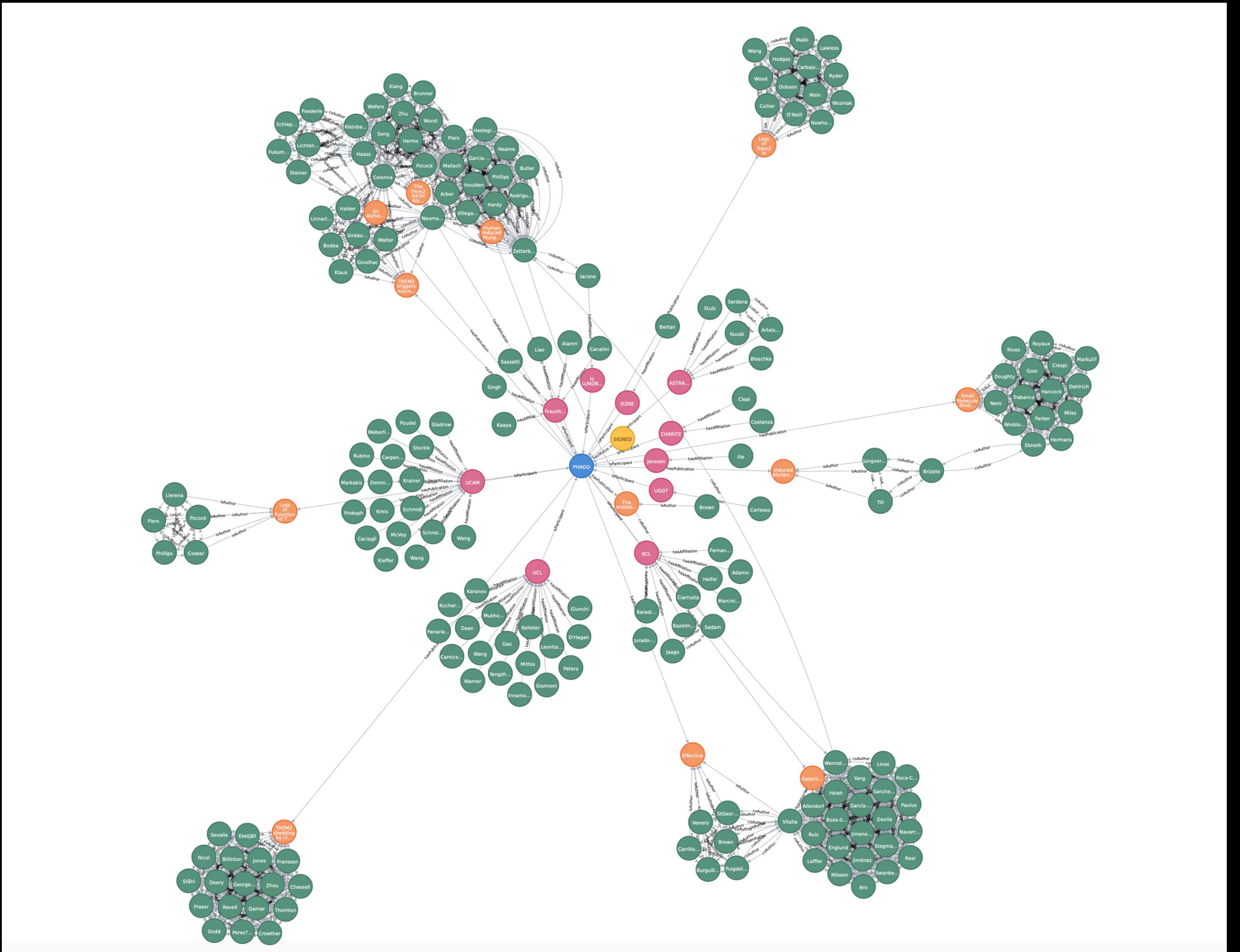
The **context of data** is needed to generate
the knowledge necessary for further analysis

Social Network Analysis combined with KGs



Social Network Analysis combined with KGS

Which people might be a good choice for collaborating on a particular topic?

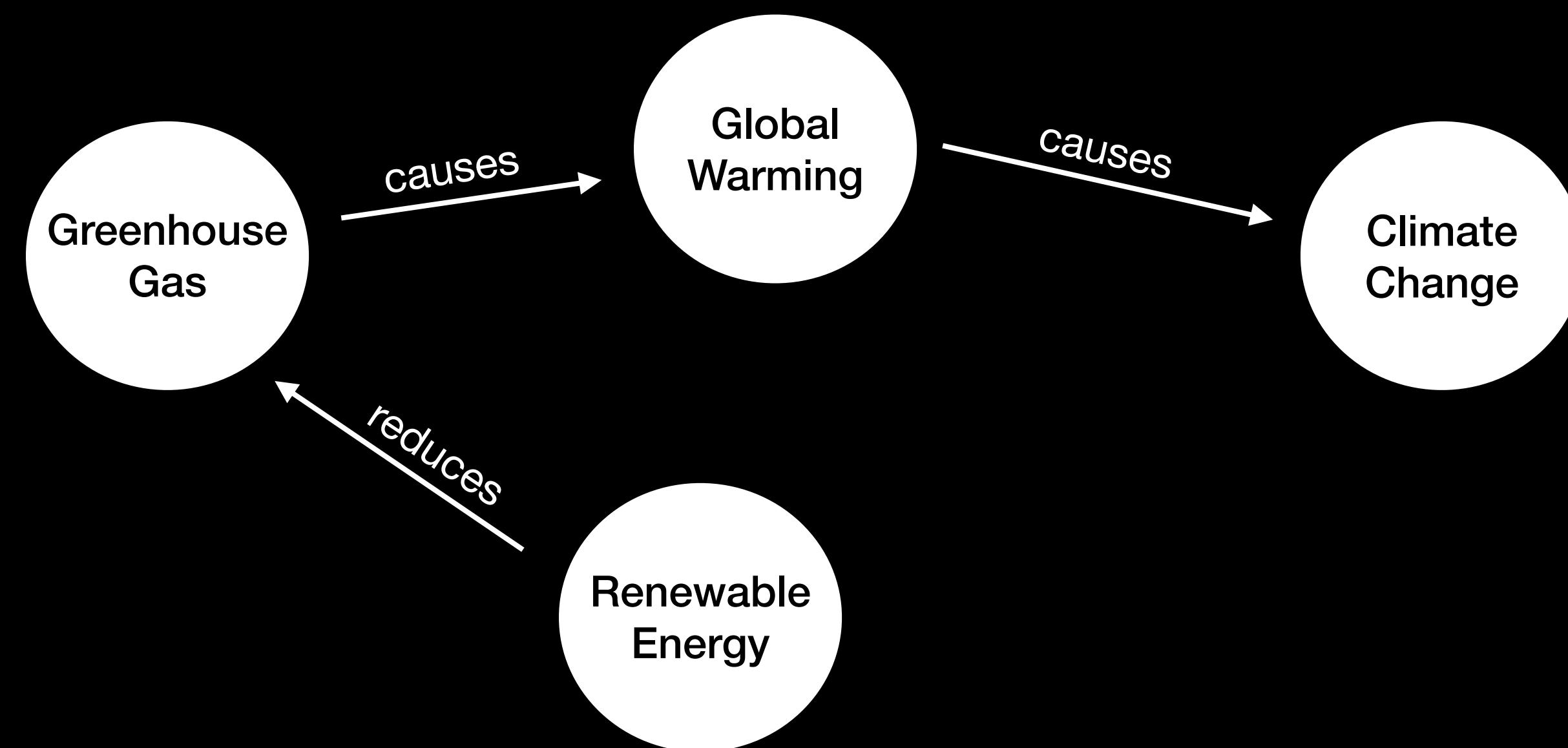


Offers all kinds
of computational
methods

Argument Mining combined with KGs

Premise: Renewable energy reduces greenhouse gas emissions.

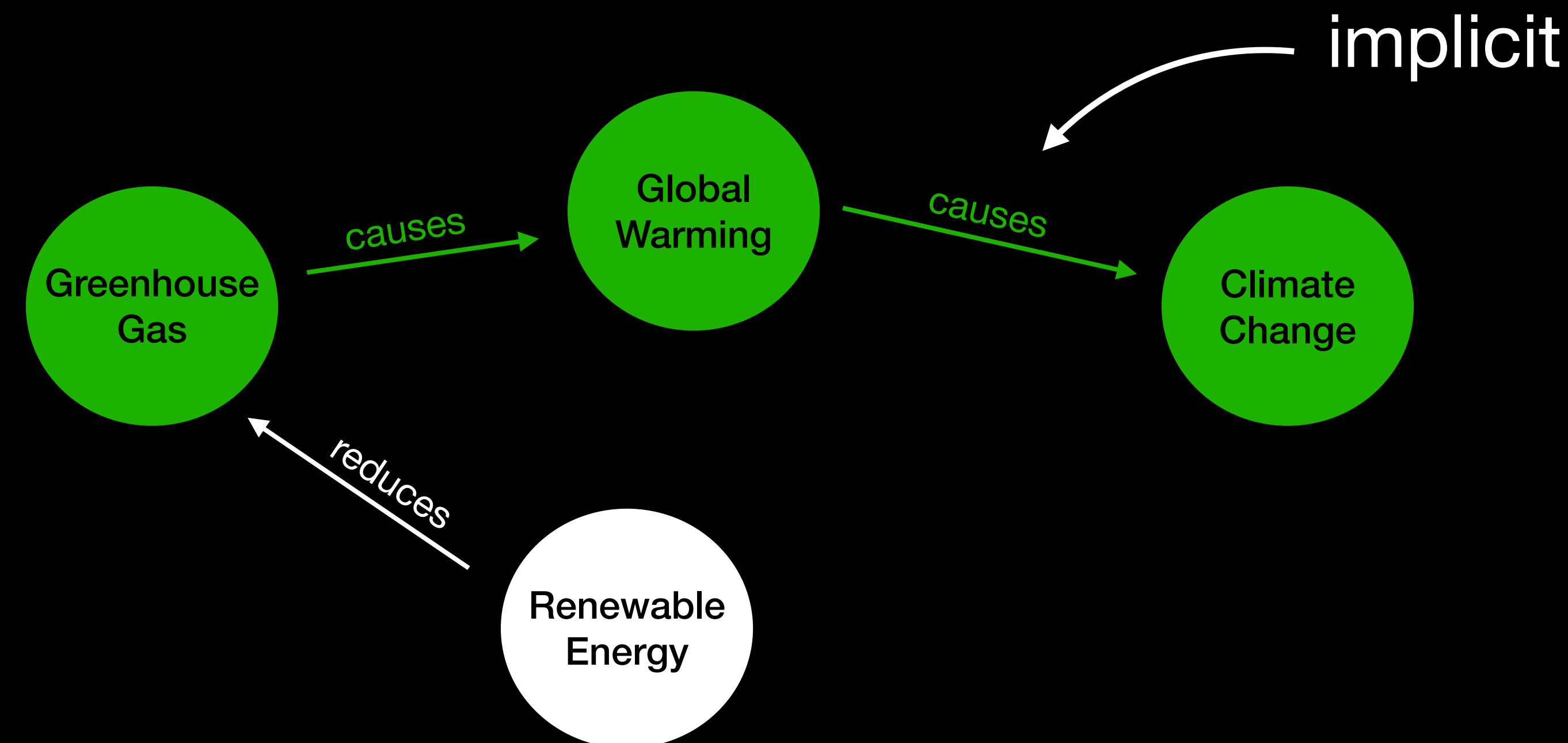
Conclusion: Implementing renewable energy policies will mitigate climate change and improve public health.



Argument Mining combined with KGs

Premise: Renewable energy reduces greenhouse gas emissions.

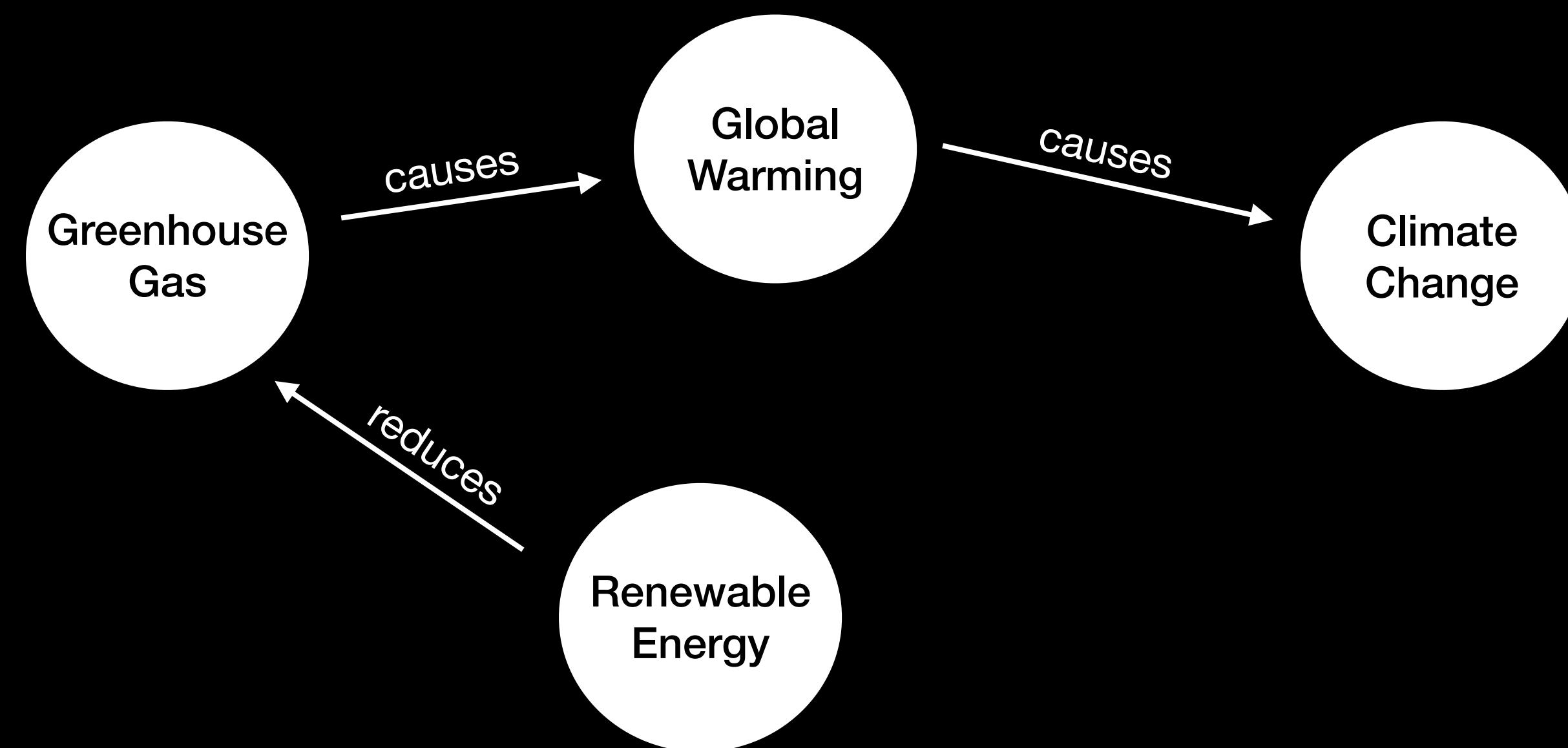
Conclusion: Implementing renewable energy policies will mitigate climate change and improve public health.



Argument Mining combined with KGs

Premise: Renewable energy reduces greenhouse gas emissions.

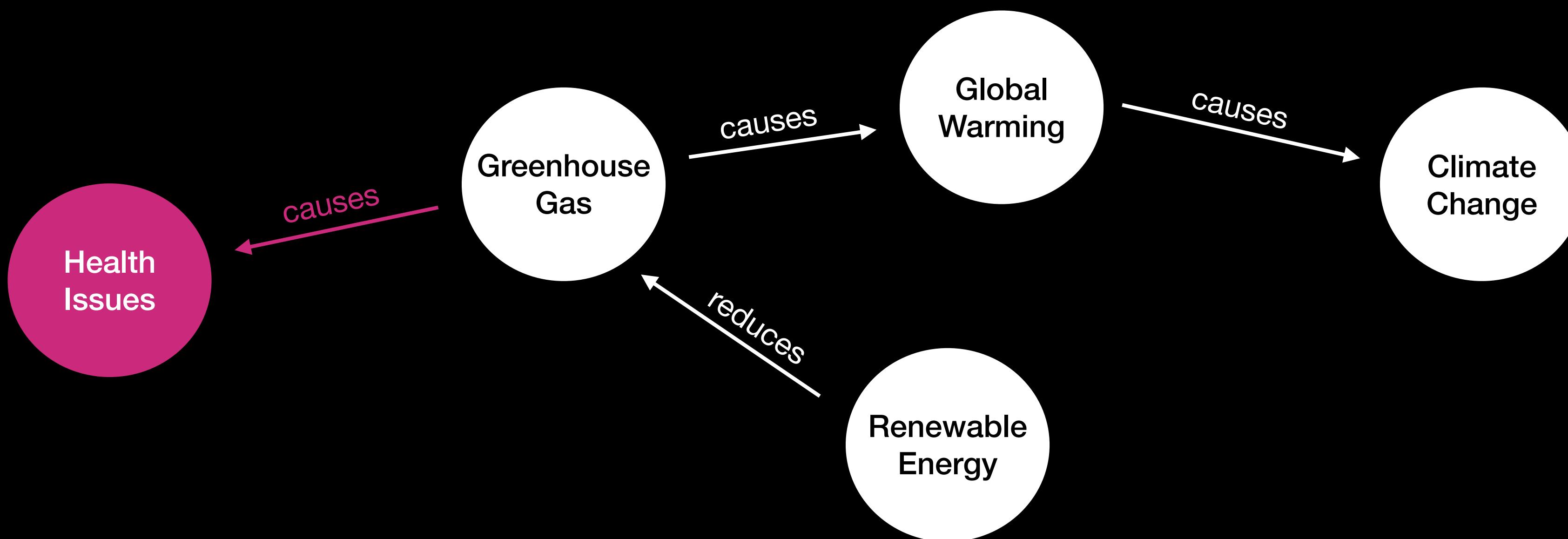
Conclusion: Implementing renewable energy policies will mitigate climate change and **improve public health**.



Argument Mining combined with KGs

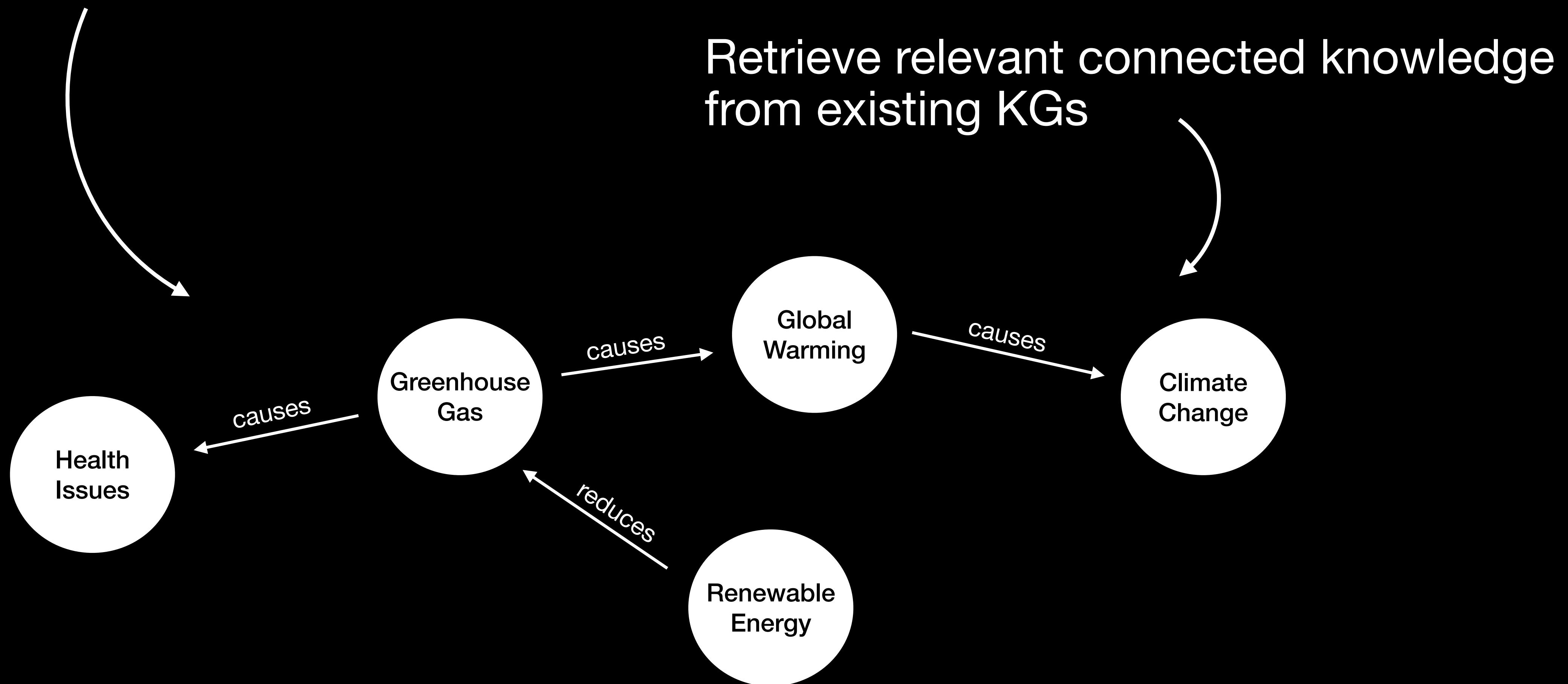
Premise: Renewable energy reduces greenhouse gas emissions

Conclusion: Implementing renewable energy policies will mitigate climate change and **improve public health.**



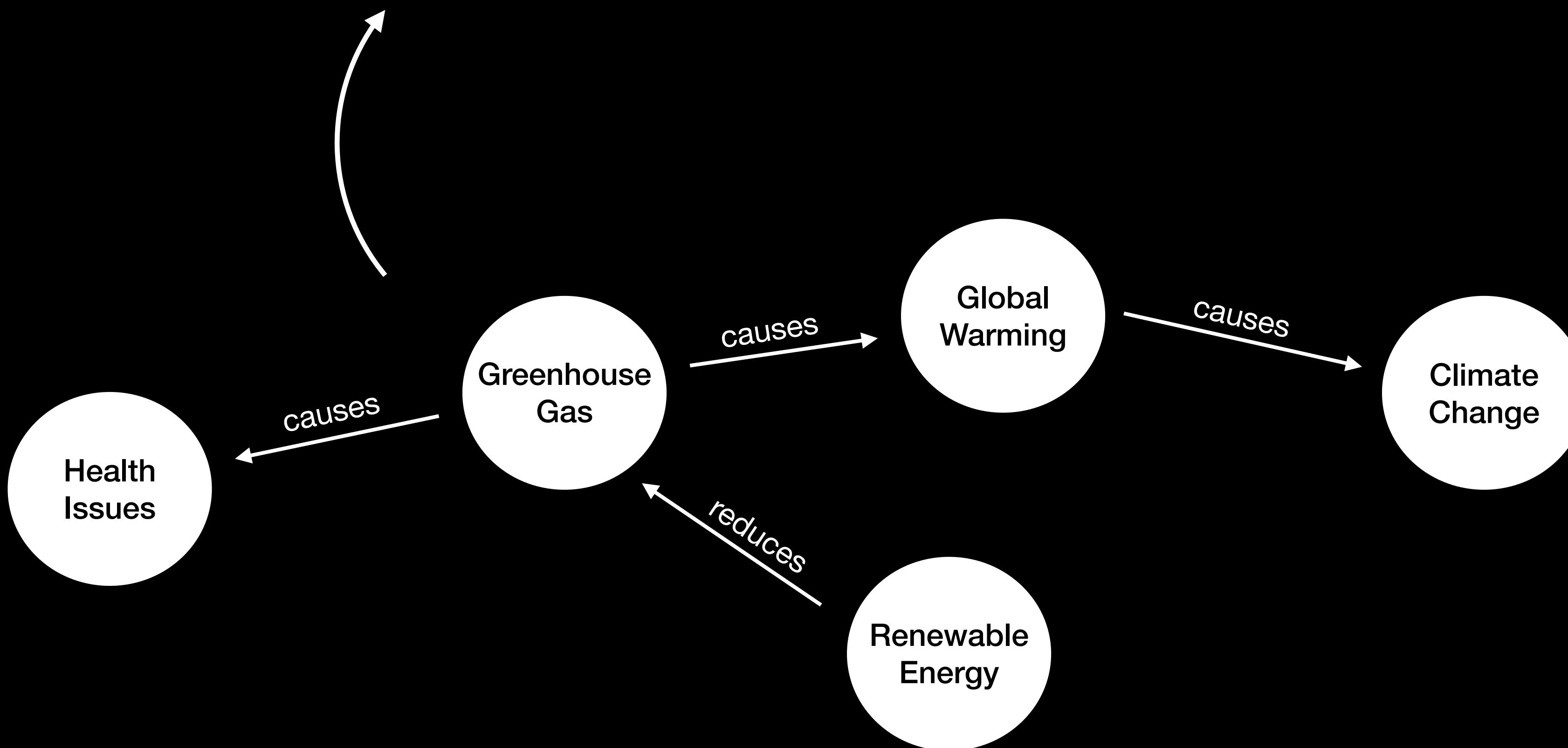
Argument Mining combined with KGs

Using the parametric knowledge of LLMs



Argument Mining combined with KGs

Predict validity or novelty of arguments



Steps after defining the ontology

Extracting and classifying information from text

ORGANISATION

LOCATION

DATE

PERSON

WEAPON

The ISIS ORG has claimed responsibility for a suicide bomb blast in the

Tunisian LOC capital earlier this week DATE, the militant group ORG's

Amaq news agency ORG said on Thursday DATE. A militant PER wearing

an explosives belt WEAPON blew himself up in Tunis LOC

NER and RE Methods

Rule-Based Approaches

Machine Learning Approaches (Conditional Random Fields, Support Vector Machines,...
e.g., using CRFs to learn from features like word shapes and POS tags to identify entities)

Deep Learning Approaches (special ML): Most advanced use Transformer-based
models like BERT that leverage self-attention mechanism to capture contextual information



Transition to Generative models that do not require extensive training and can
provide flexibility with regard to understanding text

Extracting causal relations from text

Explicit vs. Implicit

Explicit links: *Hence, therefore, because of, since, as a result...*

Explicit causative verbs: *Breaks, destroys, reduces...*

Explicit conditional expressions: *If..., then...*

Implicit: *"He derives great joy and happiness from cycling"*

Something which decoder models (like BERT) are not able to do easily

Extracting causal relations from text

Explicit vs. Implicit

Explicit links: *Hence, therefore, because of, since, as a result...*

Explicit causative verbs: *Breaks, destroys, reduces...*

Explicit conditional expressions: *If..., then...*

Implicit: “*He derives great joy and happiness from cycling*”

Intra-sentential vs Inter-sentential

Entity and Relation Disambiguation



needed for quantitative insights

Key Questions

What can we learn from textual data?

Textual data contains information about the sentiment, topics, argumentative structure, and relational structure of concepts, all of which provide different levels of understanding about aspects of the world

What are expressive data representations?

While each of these data analysis techniques provides research insights, knowledge graphs resemble how humans perceive the world and can represent the contextual information of data

Why do LLMs enhance the analysis of text?

LLMs have been used for years to extract knowledge from text, while generative models in particular increase the flexibility and depth of information extraction.

Key Takeaway

(Knowledge) Graphs model the world closely as it is.

This enables **two key things**:

They provide an intuitive way to capture and **represent perception**.

They are able to model the highly interconnected nature of the world, making them one of the **most expressive data representation** formats.

THANK YOU

info.bersch@gmail.com

clarabersch.org

[linkedin.com/in/clara-bersch/](https://www.linkedin.com/in/clara-bersch/)