A Study of Automatically Acquiring Explanatory Inference Patterns from Corpora of Explanations: Lessons from Elementary Science Exams

Peter Jansen, School of Information, University of Arizona

THIS WORK SUPPORTED BY:

Our long term interest is in building **inference algorithms** capable of **answering questions** and producing **human-readable explanations** by **aggregating information from multiple sources** and knowledge bases.

Our long term interest is in building **inference algorithms** capable of **answering questions** and producing **human-readable explanations** by **aggregating information from multiple sources** and knowledge bases.

Standardized elementary science exam questions require an average of 4 to 6 facts (range 1-16) to answer and explain their reasoning.
(Jansen et al., COLING 2016; Jansen et al., submitted to LREC 2018)

Assembling long chains of facts to answer questions is challenging. Inference algorithms have difficulty aggregating more than 2 pieces of information, and quickly drift off-topic due to "semantic drift".
(Fried et al., TACL 2015; Khashabi et al., IJCAI 2016; Jansen et al., CL 2017)

Some new method of controlling for semantic drift to assemble long inference chains is required for complex explainable QA

This paper investigates whether extracting large "common inference patterns" from corpora of explanations is a viable method.

Main Questions

Are there ***"explanatory patterns"*** in corpora of explanations that are commonly reused?
(i.e. 2 or more facts that are commonly seen together in different explanations)

If explanatory patterns exist, is it possible to ***abstract*** them in some way to make them more general?
(i.e. specific ***kinds of knowledge***, instead of specific ***instances of knowledge***, that tend to be seen together)

How much of explanations for unseen questions is it possible to reconstruct just by ***merging, adapting, or adding to these explanatory patterns?***

(i.e. can we use common explanatory patterns to reduce the need for "raw" information aggregation, and start to build large inferences with 10 or more facts for QA)

Part 1:

Question answering as building explanation graphs

THE BEST AI STILL FLUNKS 8TH GRADE SCIENCE

THEN ONE/WIRED

IN 2012, IBM Watson went to medical school. So said *The New York Times*, announcing that the tech giant's artificially intelligent question-and-answer machine had begun a "stint as a medical student" at the Cleveland Clinic Lerner College of Medicine.

This was just a metaphor. Clinicians were helping IBM train Watson for use in medical research. But as metaphors go, it wasn't a very good one. Three years later, our artificially intelligent machines can't even pass an eighth-grade science test, much less go to medical school.

The top performers successfully answered about 60 percent of the questions. In other words, they flunked.

Aggregation and Inference for Science Exams

Q: Which of the following is an example of an organism taking in nutrients?

(A) A dog burying a bone
(B) A girl eating an apple
(C) An insect crawling on a leaf
(D) A boy planting tomatoes

Science Exam Example Question

Q: Which of the following is an example of an organism taking in nutrients?

(A) A dog burying a bone
(B) A girl eating an apple
(C) An insect crawling on a leaf
(D) A boy planting tomatoes

Rarely will we be able to **retrieve a single passage** in a corpus that directly answers a given question:

"A girl eating an apple is an example of an organism taking in nutrients..."

Aggregation and Inference for Science Exams

Q: Which of the following is an example of an organism taking in nutrients?

(A) A dog burying a bone
(B) A girl eating an apple
(C) An insect crawling on a leaf
(D) A boy planting tomatoes

Girl

"a girl means a human girl"

"humans are living organisms"

Simple Wiktionary

AND

Eating

"eating is when an organism takes in nutrients in the form of food"

4th Grade Study Guide

AND

Apple

"an apple is a kind of fruit"

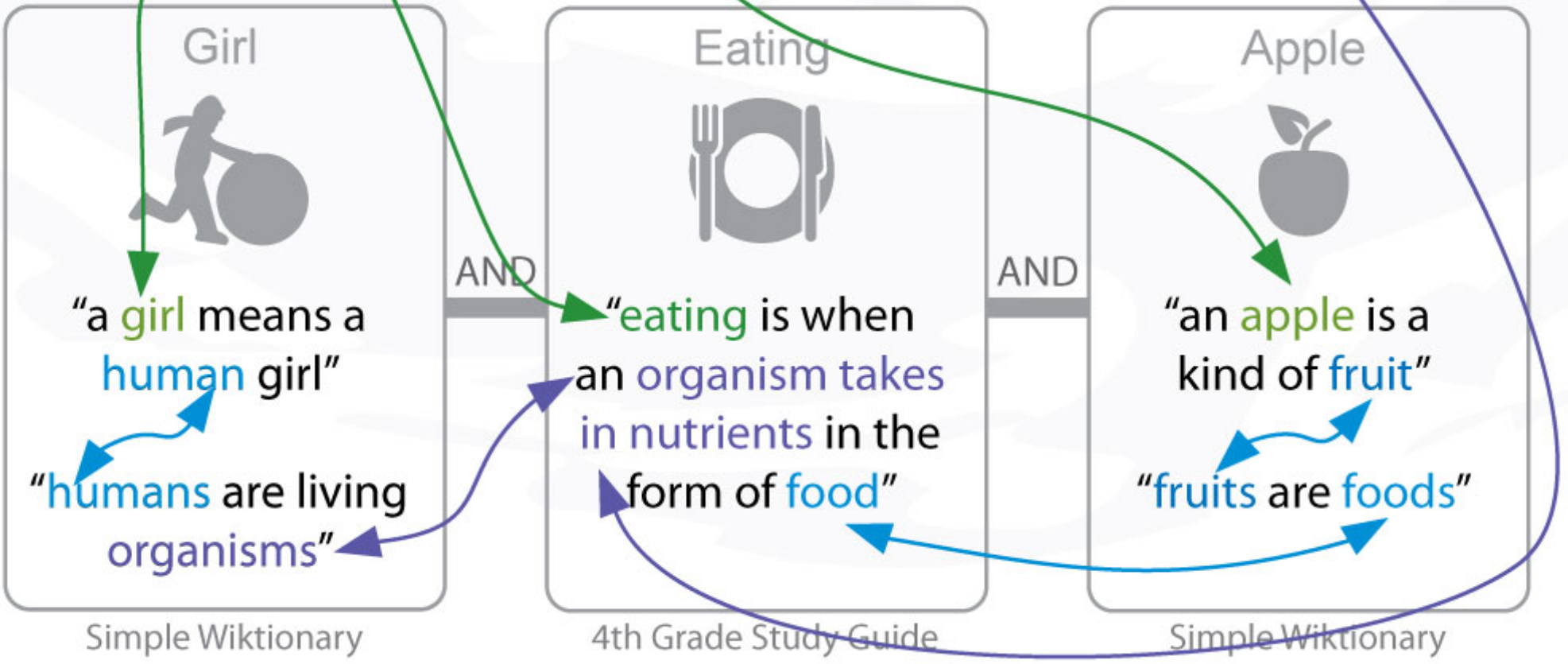
"fruits are foods"

Simple Wiktionary

Aggregation and Inference for Science Exams

Q: Which of the following is an example of an organism taking in nutrients?

(A) A dog burying a bone
(B) A girl eating an apple
(C) An insect crawling on a leaf
(D) A boy planting tomatoes



Explanatory Depth

WorldTree Corpus: 1,680 Standardized Science Questions paired with detailed, manually authored, and lexically-connected explanation graphs.

Question Which characteristic would best help a tree survive the heat of a forest fire?
Answers [A] large leaves [B] shallow roots *[*C] thick bark* [D] thin trunks

Domain Expert (e.g. teacher)

Bark is a protective covering around the trunk and branches of a tree.

Domain Novice (e.g. student)

As an object's thickness increases, it's resistance to damage will also increase.

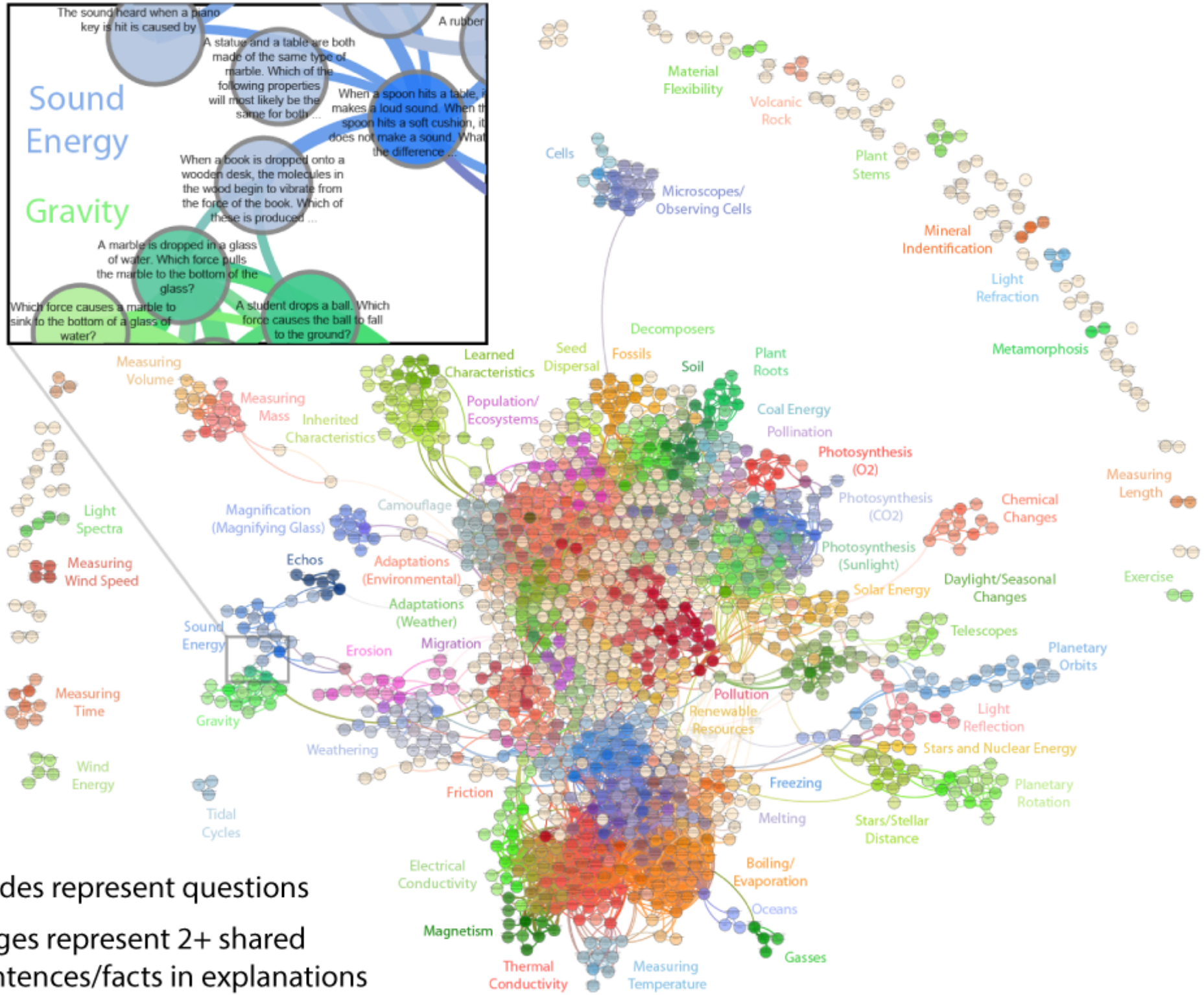
Young Child (e.g. 5-year old)

Protecting something means preventing harm.
Fire causes harm to trees, forests, and other living things.
Thickness is a measure of how thick an object is.
A tree is a kind of living thing.

First Principles

Protecting a living thing has a positive impact on it's survival and health

Target Depth



Nodes represent questions

Edges represent 2+ shared sentences/facts in explanations

Explanations Grounded in Semi-structured Tables

Explanations are represented as one or more rows in 62 semi-structured tables, providing both ***coarse (sentence-level)*** and ***fine-grained (table column)*** *explanation graph* structure.

Question: Which event involves a consumer and a producer in a food chain ?

Process Roles Table

	PROCESS NAME		ACTOR		ROLE		ACTION	PATIENT		PURPOSE
In the	food chain	process, a	green plant	has the role of	producer	which	creates	food	for	consumers
In the	food chain	process, an	animal	has the role of	consumer	which	eats	producers	for	food
In the	food chain	process, a	bacteria	has the role of	decomposer	which	recycles	nutrients		
In the	tree reproduction	process, a	squirrel	has the role of	seed disperser	which	relocates	seeds		

Taxonomy Table

	HYPONYM		SCOPE	HYPERNYM
A	deer	is a kind of		animal
	green	is a kind of		color
	shelter	is a kind of	protective	covering
An	electromagnet	is a kind of	electric	magnet

PartOf Table

	PART		WHOLE
A	leaf	is a part of a	green plant
	roots	are a part of a	plant
	pedals	are a part of a	bicycle
A	cell wall	is a part of a	plant cell

Answer Candidates:

[A] a cat eats a mouse

[*B] a deer eats a leaf

[C] a hawk eats a mouse

[D] a snake eats a rat

Part 2:

Identifying and reusing patterns in corpora of explanations

Approach

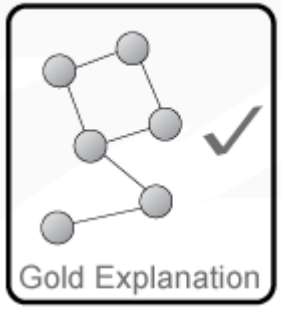
Question

What property of **water** can make a **rock break apart** ?

Answer

freezing and thawing

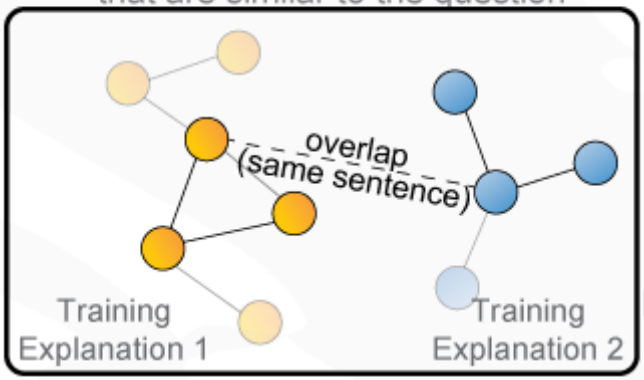
Compare accuracy of assembled explanation with known good explanation in test set.



SEARCH

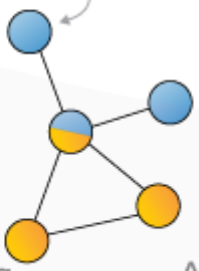
Step 1

Search for explanations in the training set that are similar to the question



one node = one explanation sentence

Step 2 **MERGE**

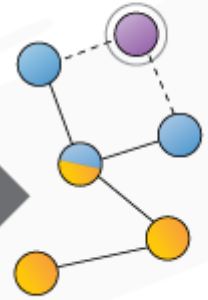


Merge Training Explanations

Step 3 **ADAPT**

Adapt Knowledge to New Question

Step 4 **ADD**



Add New Knowledge

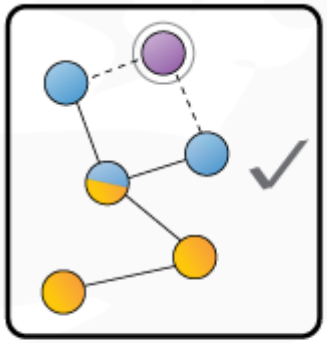
Step 5 **EVALUATE**

Question

What property of water can make a rock break apart ?

Answer

freezing and thawing



Zoom of Explanation Graph Building Process

canyons are formed by flowing water over long periods of time

a canyon is made of rocky material

STEP 4 ADD

cycles of freezing and thawing water cause ice wedging

ice weding is when ice causes rocks to crack by expanding in openings

weathering means breaking down rocks into smaller pieces

STEP 2 MERGE

overlap for merge (same sentence)

weathering is usually caused by water or wind

STEP 3 ADAPT

cracking something may cause that something to break apart

weathering means breaking down rocks into smaller pieces

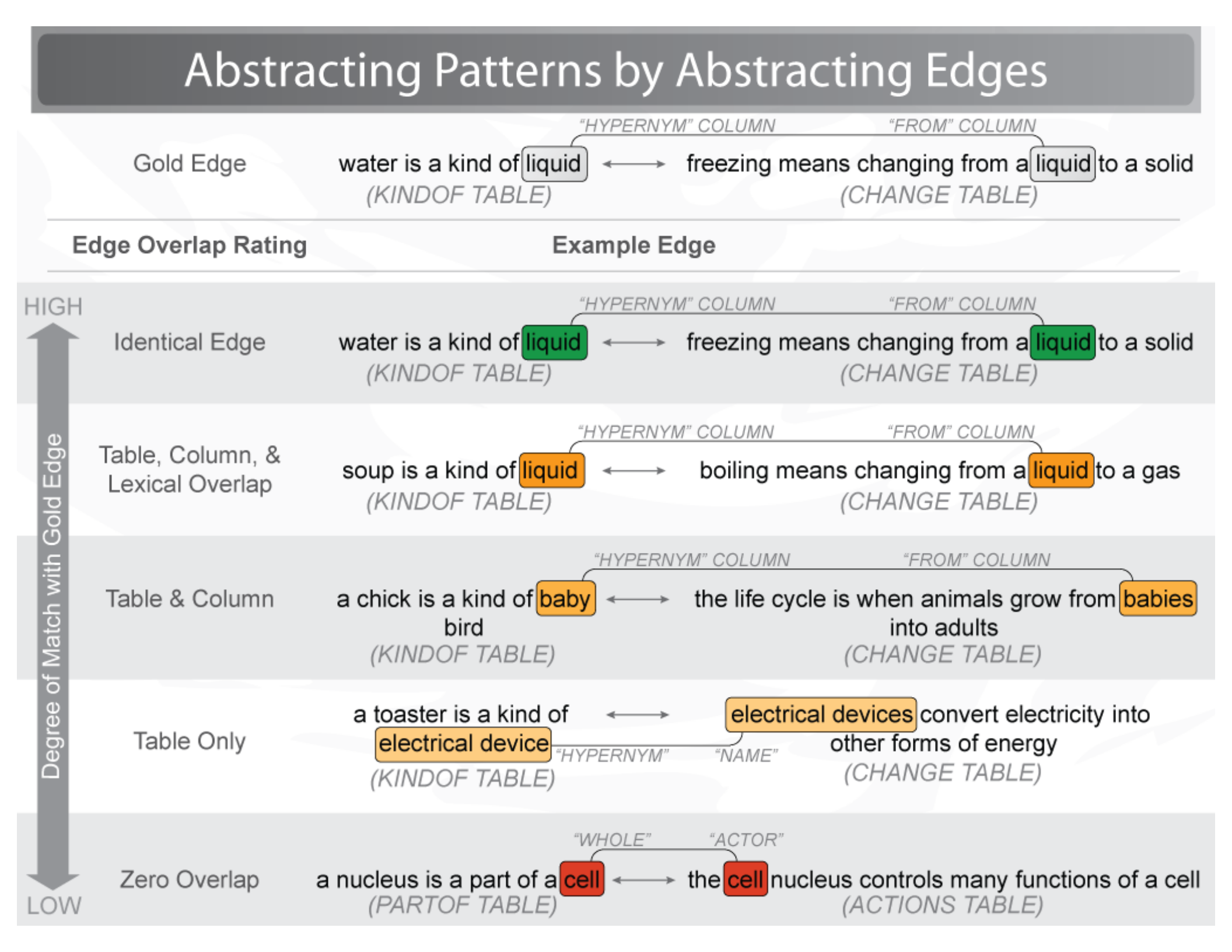
break apart means turn into pieces using force

to cause means to make

generating a force requires expending energy

weathering is a kind of natural process

Abstracting Patterns by Abstracting Edges

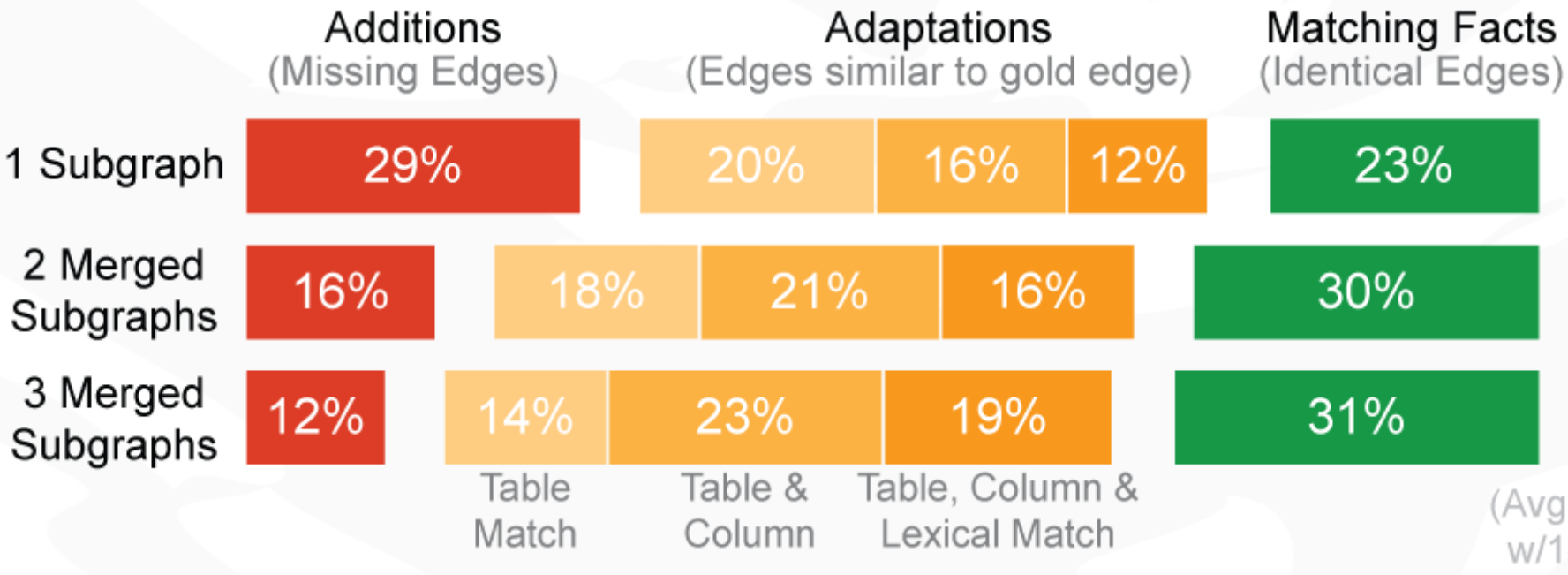


Example Common Edges

Freq.	Abs. Level	Pattern		
11 / 800	Row	water is a kind of **liquid** *(KINDOF TABLE)*	⟷	boiling means changing from a **liquid** to a gas *(CHANGE TABLE)*
5 / 800	Table, Column, & Lexical Overlap	KINDOF TABLE "HYPONYM" COLUMN	"PLANT" ⟷	PARTOF TABLE "WHOLE" COLUMN
		a tree is a kind of **plant**	⟷	a leaf is a part of most **plants**
199 / 800	Table & Column	KINDOF TABLE "HYPONYM" COLUMN	⟷	KINDOF TABLE "HYPERNYM" COLUMN
		a dog is a kind of **animal**	⟷	an **animal** is a kind of organism
80 / 800	Table Only	KINDOF TABLE	⟷	USEDFOR TABLE
		a **magnifying glass** is a kind of tool	⟷	a **magnifying glass** is used to see small objects by making them appear larger

Reconstruction Performance

If we use only "explanatory patterns" from other questions, how close can we come to reconstructing unseen explanations for novel questions?



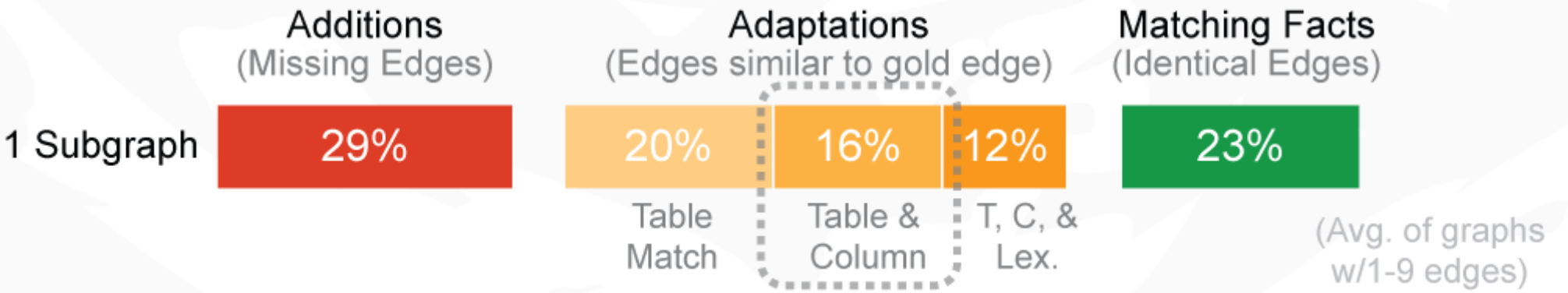
Interpretation: for a hypothetical explanation graph with 9 edges:

3 edges would be identical

4.5 edges would require varying degrees of adaptation

1.5 edges would require "raw" information aggregation

Example Pattern (Short and Highly Abstracted)



Question Wendy wanted to find out if faster wind speeds increased the amount of wind erosion. Which instrument should she use to measure wind speed?

Answers ***[*A] anemometer*** [B] hygrometer [C] rain gauge [D] thermometer

---

Gold Explanation

wind speed is a property of weather *(PROPERTIES table)*
an anemometer is used to measure wind speed *(USERFOR table)*
an anemometer is a kind of instrument *(KINDOF table)*

Similar Explanatory Pattern (from a different, known question)

air pressure is a property of air/the atmosphere *(PROPERTIES table)*
a barometer is used to measure air pressure *(USEDFOR table)*
a barometer is a kind of instrument *(KINDOF table)*

Example common 2-edge pattern at ***"Table & Column"*** level of abstraction

Little Data (Aggregated) is Bigger than the Biggest Data

We mapped out the knowledge required to answer and explain 2,200 3rd to 5th grade standardized elementary science questions.

We found that only 4,000 facts are required to answer and explain these exam questions, as long as they can be aggregated/ combined into short explanations with an average of 6 facts/sentences. (Jansen et al., submitted to LREC 2018)

In contrast, systems with very large knowledge bases (~1,000,000 sentences) are answering ~60% of standardized questions.

Why can these questions be completely answered and explained with 200X less data? *Information Aggregation vs Passage Retrieval*

Information Aggregation: 4,000 facts combined into 6-fact explanations allows for ~4,096,000,000,000,000,000,000 possiblilities

Empirically estimate that 2x10-8, or 133 trillian, would make good explanations (Jansen, submitted to LREC 2018)

133 trillian (little data aggregated) >> 1 million (big data)

Thank You!

Download common explanatory patterns at:
cognitiveai.org/explanations

Computational
Language
Understanding

THIS WORK SUPPORTED BY:

ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE