## Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Duration</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 PST</td>
<td>1 hr 50 mins</td>
<td><strong>Part I - Review of CSKGs</strong></td>
</tr>
<tr>
<td>15 min</td>
<td></td>
<td>Introduction to commonsense knowledge (slides) - Pedro</td>
</tr>
<tr>
<td>25 min</td>
<td></td>
<td>Review of top-down commonsense knowledge graphs (slides) - Mayank</td>
</tr>
<tr>
<td>70 min</td>
<td></td>
<td>Review of bottom-up commonsense knowledge graphs (slides+demo) - Mayank, Filip, Pedro</td>
</tr>
<tr>
<td>10 min</td>
<td></td>
<td><strong>Break</strong></td>
</tr>
<tr>
<td>10:00 PST</td>
<td>45 min</td>
<td><strong>Part II - Integration and analysis</strong></td>
</tr>
<tr>
<td>35 min</td>
<td></td>
<td>Consolidating commonsense graphs (slides) - Filip</td>
</tr>
<tr>
<td>10 min</td>
<td></td>
<td>Consolidating commonsense graphs (demo) - Pedro</td>
</tr>
<tr>
<td>10 min</td>
<td></td>
<td><strong>Break</strong></td>
</tr>
<tr>
<td>10:55 PST</td>
<td>1 hr 05 mins</td>
<td><strong>Part III - Downstream use of CSKGs</strong></td>
</tr>
<tr>
<td>50 min</td>
<td></td>
<td>Answering questions with CSKGs (slides+demo) - Filip</td>
</tr>
<tr>
<td>15 min</td>
<td></td>
<td>Wrap-up (slides) - Mayank</td>
</tr>
</tbody>
</table>
Answering Questions with CSKGs

Filip Ilievski
Commonsense Knowledge Graphs
The Commonsense Knowledge Graph (CSKG)

7 sources

2.3M nodes
6M edges

Semantic Parsing

Construct semantic representations of question and answers
Grounding Questions

Link semantic parses to KG
Grounding Answers

Link semantic parses to KG
Reasoning

Find and rank connections for question/answer pairs

Connection subgraph is an explanation
Motivating Example

When boiling butter, when it's ready, you can...

- pour it on a plate
- pour it into a jar
Motivating Example

When boiling butter, when it’s ready, you can...

- pour it on a plate
- pour it into a jar

Required Common Sense:
- Things that boil are liquid (when they’re ready)
- Liquids can be poured
- Butter can be a liquid
- Jars hold liquids
- Plates (typically) do not contain liquids

Required Linguistic Understanding:
- The antecedent of ‘it’ is ‘butter’
Semantic Parsing: Text to Meaning Representation

When boiling butter, when it’s ready, you can...
..pour it on a plate

When boiling butter, when it’s ready, you can...
...pour it into a jar
Semantic Parsing: Text to Meaning Representation

Three steps:

1. **Semantic Role Labeling**
   - Graphical encoding of dependencies between subjects/verbs in a sentence.

2. **Coreference Resolution**
   - Link mentions of entity within and across sentences.

3. **Named Entity Recognition**
   - Map fine-grained entities (e.g., “John”) to common entities (e.g., “Person”).
   - Better generalization
Semantic Role Labeling
Labels predicates (verbs) and their associated arguments.

John is waiting for his car to be finished.
Coreference Resolution

Links mentions of a single entity in a sentence or across sentences.

John is waiting for his car to be finished.
Named Entity Recognition

Marks each node if it is a named entity along with the entity type.

John is waiting for his car to be finished.

Diagram:

- John (PERSON)
- waiting (subject)
- for his car to be finished
- his car (object)
- finished
Semantic Parse: Question/Context

Which answer choice is better?

Q. When boiling butter, when it is ready, you can...

Ans 1  pour it on a plate.

Ans 2  pour it into a jar
Semantic Parse: Answer 1

When boiling butter, when it is ready, you can pour it on a plate.
When boiling butter, when it is ready, you can pour it into a jar.
Shortcomings and Future Directions

Many different ways to parse a sentence/sentences

- Semantic role labeling focuses on predicates, but ignores things like prepositional phrases.
- Can incorporate dependency parsing, abstract meaning representations (AMR), etc.

Future Work: Explore other meaning representations, including logic
Linking to Commonsense KG

When boiling butter, when it's ready, you can...

..pour it on a plate

Score possible reasoning in CSKG

score(q, a_i)
When boiling butter, when it’s ready, you can...

...pour it into a jar
Linking to Commonsense KG

When boiling butter, when it’s ready, you can...

Which reasoning is better?

score(q, a₁) \land score(q, a₂)

...pour it into a jar
Linking to CSKG

So that reasoning can take place!
The boy loved telling scary stories.

The boy ——— /c/en/boy
loved ——— /c/en/loved
telling ——— /c/en/telling
scary stories ——— /c/en/horror_stories

Generalizes to concepts (not just lexical)
Approach

- Embed words and phrases
  - Tokenization/concept matching
  - “Natural language processing” or “Natural”, “language”, “processing”?
  - Use embeddings
  - ConceptNet Numberbatch [Speer et al., AAAI 2017]
  - BERT [Devlin et al., 2018]
  - Node representation = function of word embeddings

- Compute alignment between text and KG embeddings
  - Cosine/L2 distance
Examples

“amused” -> amused (0.0), amusedness (0.04), amusedly (0.12), ...

“Tina, a teenager” -> teenager (0.0), tina (0.0), subteen (0.01), ...

“With how popular her mother is” -> mother (0.0), with (0.0), is (0.0), ...

“Scary stories” -> stories (0.0), scary (0.0), scarisome (0.02), ...
Examples

“amused” -> amused (0.0), amusedness (0.04), amusedly (0.12), ...

“Tina, a teenager” -> teenager (0.0), tina (0.0), subteen (0.01), ...

“With how popular her mother is” -> mother (0.0), with (0.0), is (0.0), ...

“Scary stories” -> stories (0.0), scary (0.0), scarisome (0.02), ...

- Potentially better links: scary_story, horror_story
  - horror_story appears in the top-5 using original averaging method
Challenges


- Average embedding is closer to his. Car is not linked.
- Alternatives:
  - Link each word. Simple, but not compositional.
  - Link root of dependency parse. Discards even more information.

Polysemous words: “Doggo is good boy” vs. “Toilet paper is a scarce good”

- Can perform word sense disambiguation/link to WordNet nodes instead.
  - Better to handle at linking or graph reasoning step?

Evaluation
Fidelity vs. Utility Trade-off

- Exact matches may exist, but are not always useful
- Incorporate node degree?
Neuro-symbolic Reasoning Approaches
Neuro-Symbolic Reasoning Approaches

Knowledge enhances language models

Language models fill in knowledge gaps
Neuro-Symbolic Reasoning
Approaches

Knowledge enhances language models

Language models fill in knowledge gaps

Kaixin Ma, Filip Ilievski, Jon Francis, Yonatan Bisk, Eric Nyberg, Alessandro Oltramari. In prep.
Structured evidence in CSKGs

Q: Bob the lizard lives in a warm place with lots of water. Where does he probably live?

A: tropical rainforest
Structured evidence in CSKGs

Q: Bob the lizard lives in a warm place with lots of water. Where does he probably live?

A: tropical rainforest

AtLocation
(ConceptNet)
Structured evidence in CSKGs

Q: Bob the *lizard* lives in a warm *place* with lots of water. Where does he probably live?

A: *tropical rainforest*

- AtLocation (ConceptNet)
- HasInstance (FrameNet-ConceptNet)
Structured evidence in CSKGs

Q: Bob the lizard lives in a warm place with lots of water. Where does he probably live?

A: tropical rainforest
HyKAS (based on Ma et al. 2019)

Q: Bob the lizard lives in a warm place with lots of water. Where does he probably live?
A: Tropical rainforest

Q: /c/en/lizard, vg:water, …
A: /c/en/tropical, …

(lizard, AtLocation, tropical rainforest)
(place, HasInstance, tropical)
(water, MayHaveProperty, tropical)

Lizards can be located in tropical rainforests. Tropical is a kind of a place. Water can be tropical.
Bob the lizard lives in a warm place with lots of water. Where does he probably live?

**A:** Tropical rainforest

Lizards can be located in tropical rainforests. Tropical is a kind of a place. Water can be tropical.
HyKAS (based on Ma et al. 2019)

Q: Bob the lizard lives in a warm place with lots of water. Where does he probably live?
A: Tropical rainforest

Lizards can be located in tropical rainforests. Tropical is a kind of a place. Water can be tropical.

[CLS] Bob the lizard lives in a warm place with lots of water. Where does he probably live? [SEP] tropical rainforest [SEP]

[CLS] Bob the lizard lives in a warm place with lots of water. Where does he probably live? [SEP] mountain [SEP]

[CLS] Bob the lizard lives in a warm place with lots of water. Where does he probably live? [SEP] desert [SEP]
'No-knowledge’ baseline is strong

### CommonSense QA

<table>
<thead>
<tr>
<th>Train+inference Knowledge</th>
<th>Dev acc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>76.7</td>
</tr>
<tr>
<td>ATOMIC</td>
<td>77.1</td>
</tr>
<tr>
<td>ConceptNet</td>
<td>80.1</td>
</tr>
<tr>
<td>CSKG</td>
<td>79.5</td>
</tr>
<tr>
<td>CSKG -symmetric -overlapping</td>
<td>79.7</td>
</tr>
<tr>
<td>CSKG in a separate OCN</td>
<td>80.1</td>
</tr>
<tr>
<td>ConceptNet (2-hop)</td>
<td>80.5</td>
</tr>
</tbody>
</table>

### SocialIQA

<table>
<thead>
<tr>
<th>Train knowledge</th>
<th>Inference Knowledge</th>
<th>Dev acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>78.7</td>
</tr>
<tr>
<td>ATOMIC</td>
<td>ATOMIC</td>
<td>79.04</td>
</tr>
<tr>
<td>CSKG</td>
<td>CSKG</td>
<td>78.56</td>
</tr>
<tr>
<td>CSKG</td>
<td>-</td>
<td>77.22</td>
</tr>
<tr>
<td>CSKG -Visual Genome</td>
<td></td>
<td>78.4</td>
</tr>
<tr>
<td>CSKG</td>
<td>ConceptNet</td>
<td>78.61</td>
</tr>
<tr>
<td>CSKG</td>
<td>Visual Genome</td>
<td>78.04</td>
</tr>
<tr>
<td>CSKG</td>
<td>ConceptNet+Visual Genome</td>
<td>78.81</td>
</tr>
<tr>
<td>CSKG</td>
<td>-RelatedTo</td>
<td>78.4</td>
</tr>
<tr>
<td>CSKG</td>
<td>-Synonym-Antonym</td>
<td>78.66</td>
</tr>
</tbody>
</table>
Adding knowledge helps

### CommonSense QA

<table>
<thead>
<tr>
<th>Train+inference Knowledge</th>
<th>Dev acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>76.7</td>
</tr>
<tr>
<td>ATOMIC</td>
<td>77.1</td>
</tr>
<tr>
<td>ConceptNet</td>
<td>80.1</td>
</tr>
<tr>
<td>CSKG</td>
<td>79.5</td>
</tr>
<tr>
<td>CSKG -symmetric -overlapping</td>
<td>79.7</td>
</tr>
<tr>
<td>CSKG in a separate OCN</td>
<td>80.1</td>
</tr>
<tr>
<td>ConceptNet (2-hop)</td>
<td>80.5</td>
</tr>
</tbody>
</table>

### SocialIQA

<table>
<thead>
<tr>
<th>Train knowledge</th>
<th>Inference Knowledge</th>
<th>Dev acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>78.7</td>
</tr>
<tr>
<td>ATOMIC</td>
<td>ATOMIC</td>
<td>79.04</td>
</tr>
<tr>
<td>CSKG</td>
<td>CSKG</td>
<td>78.56</td>
</tr>
<tr>
<td>CSKG</td>
<td>-</td>
<td>77.22</td>
</tr>
<tr>
<td>CSKG</td>
<td>-Visual Genome</td>
<td>78.4</td>
</tr>
<tr>
<td>CSKG</td>
<td>ConceptNet</td>
<td>78.61</td>
</tr>
<tr>
<td>CSKG</td>
<td>Visual Genome</td>
<td>78.04</td>
</tr>
<tr>
<td>CSKG</td>
<td>ConceptNet+Visual Genome</td>
<td>78.81</td>
</tr>
<tr>
<td>CSKG</td>
<td>-RelatedTo</td>
<td>78.4</td>
</tr>
<tr>
<td>CSKG</td>
<td>-Synonym-Antonym</td>
<td>78.66</td>
</tr>
</tbody>
</table>
Different knowledge helps different problems

### CommonSense QA

<table>
<thead>
<tr>
<th>Train+inference Knowledge</th>
<th>Dev acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>76.7</td>
</tr>
<tr>
<td>ATOMIC</td>
<td>77.1</td>
</tr>
<tr>
<td>ConceptNet</td>
<td>80.1</td>
</tr>
<tr>
<td>CSKG</td>
<td>79.5</td>
</tr>
<tr>
<td>CSKG -symmetric -overlapping</td>
<td>79.7</td>
</tr>
<tr>
<td>CSKG in a separate OCN</td>
<td>80.1</td>
</tr>
<tr>
<td>ConceptNet (2-hop)</td>
<td>80.5</td>
</tr>
</tbody>
</table>

### SocialIQA

<table>
<thead>
<tr>
<th>Train knowledge</th>
<th>Inference Knowledge</th>
<th>Dev acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>78.7</td>
</tr>
<tr>
<td>ATOMIC</td>
<td>ATOMIC</td>
<td>79.04</td>
</tr>
<tr>
<td>CSKG</td>
<td>CSKG</td>
<td>78.56</td>
</tr>
<tr>
<td>CSKG</td>
<td>-</td>
<td>77.22</td>
</tr>
<tr>
<td>CSKG</td>
<td>-Visual Genome</td>
<td>78.4</td>
</tr>
<tr>
<td>CSKG</td>
<td>ConceptNet</td>
<td>78.61</td>
</tr>
<tr>
<td>CSKG</td>
<td>Visual Genome</td>
<td>78.04</td>
</tr>
<tr>
<td>CSKG</td>
<td>ConceptNet+Visual Genome</td>
<td>78.81</td>
</tr>
<tr>
<td>CSKG</td>
<td>-RelatedTo</td>
<td>78.4</td>
</tr>
<tr>
<td>CSKG</td>
<td>-Synonym-Antonym</td>
<td>78.66</td>
</tr>
</tbody>
</table>
More knowledge is not always better
Enhancing CSKGs with Language Models

EMNLP Findings 2020
Neuro-Symbolic Reasoning Approaches

Knowledge enhances language models

Language models fill in knowledge gaps
Retrieving KG facts does not suffice

Challenges
- KG incompleteness
- Introducing irrelevant facts

Q: In what geological feature will you find fungus growing?
A: shower stall  B: toenails  C: basement  D: forest  E: cave
Retrieving KG facts does not suffice

Challenges
- KG incompleteness
- Introducing irrelevant facts

Solution
- Learn a path generator to connect entities mentioned in context with novel multi-hop knowledge paths
A KG-augmented QA Framework

- **Context Module**
  - Encode question and answer choices as unstructured evidence

- **Knowledge Module**
  - Encode knowledge facts (paths) as structured evidence

- **Reasoning Module**
  - Score a question-choice pair based on un/structured evidence
Path Generator for Connecting Dots

Goal: generate a multi-hop path between two entities

1. **Path Sampling** with KG random walk
2. **Training** by fine-tuning GPT-2
3. **Inference** by using greedy decoding
Generating knowledge paths is better than merely *retrieving* them

<table>
<thead>
<tr>
<th>Methods</th>
<th>RoBERTa-large</th>
<th>AristoRoBERTa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine-tuned LMs (w/o KG)</td>
<td>64.80 (±2.37)</td>
<td>78.40 (±1.64)</td>
</tr>
<tr>
<td>+ RN</td>
<td>65.20 (±1.18)</td>
<td>75.35 (±1.39)</td>
</tr>
<tr>
<td>+ RGCN</td>
<td>62.45 (±1.57)</td>
<td>74.60 (±2.53)</td>
</tr>
<tr>
<td>+ GconAtten</td>
<td>64.75 (±1.48)</td>
<td>71.80 (±1.21)</td>
</tr>
<tr>
<td>+ Link Prediction</td>
<td>66.30 (±0.48)</td>
<td>77.25 (±1.11)</td>
</tr>
<tr>
<td>+ PG-Local</td>
<td>70.05 (±1.33)</td>
<td>79.80 (±1.45)</td>
</tr>
<tr>
<td>+ PG-Global</td>
<td>68.40 (±0.31)</td>
<td>80.05 (±0.68)</td>
</tr>
<tr>
<td>+ PG-Full</td>
<td><strong>71.20 (±0.96)</strong></td>
<td>79.15 (±0.78)</td>
</tr>
</tbody>
</table>

Test Accuracy on OpenBookQA
Consistent improvements with less training data

Test Accuracy on CommonsenseQA and OpenBookQA with different amount of training data.
Interpretability with “real” structured paths

Q1: Where would you find magazines along side many other printed works?
PG-Global (2-hop): {magazine, IsA, book, AtLocation, bookstore}
PG-Scratch: {magazine, IsA, magazine, AtLocation, bookstore}

Q2: If you want harmony, what is something you should try to do with the world?
PG-Global (2-hop): {harmony, MotivatedByGoal, make better world, HasPrerequisite, make peace}
PG-Scratch: {harmony, UsedFor, committing perjury, Causes, make peace}

Q3: Janet was watching the film because she liked what?
PG-Global (1-hop): {film, CausesDesire, being entertained}
PG-Scratch: {film, HasContext, being entertained}
Role of knowledge

Existing benchmarks

New benchmarks?

Few shot, zero shot
**Agenda**

<table>
<thead>
<tr>
<th>Time</th>
<th>Duration</th>
<th>Session</th>
<th>Presenters</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 PST</td>
<td>1 hr 50 mins</td>
<td><strong>Part I - Review of CSKGs</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 min</td>
<td>Introduction to commonsense knowledge (slides) - Pedro</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 min</td>
<td>Review of top-down commonsense knowledge graphs (slides) - Mayank</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70 min</td>
<td>Review of bottom-up commonsense knowledge graphs (slides+demo) - Mayank, Filip, Pedro</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 min</td>
<td><strong>Break</strong></td>
<td></td>
</tr>
<tr>
<td>10:00 PST</td>
<td>45 min</td>
<td><strong>Part II - Integration and analysis</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35 min</td>
<td>Consolidating commonsense graphs (slides) - Filip</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 min</td>
<td>Consolidating commonsense graphs (demo) - Pedro</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 min</td>
<td><strong>Break</strong></td>
<td></td>
</tr>
<tr>
<td>10:55 PST</td>
<td>1 hr 05 mins</td>
<td><strong>Part III - Downstream use of CSKGs</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50 min</td>
<td>Answering questions with CSKGs (slides+demo) - Filip</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 min</td>
<td>Wrap-up (slides) - Mayank</td>
<td></td>
</tr>
</tbody>
</table>
Wrap-up
What Is Common Sense?

Common sense is sound practical judgement concerning everyday matters,
or a basic ability to perceive, understand, and judge that is shared by ("common to") nearly all people.

Wikipedia
Essential for humans to live and interact with each other in a reasonable and safe way.

Essential for AI to understand human needs and actions better.

For example, it’s ok to keep the closet door open, but it’s not ok to keep the fridge door open, as the food inside might go bad.
A Common Sense Task

**Input:** a set of common concepts

**Output:** a sentence using these concepts

- A dog leaps to catch a thrown frisbee. [Humans]
- The dog catches the frisbee when the boy throws it.
- A man throws away his dog's favorite frisbee expecting him to catch it in the air.

GPT2: A dog throws a frisbee at a football player. [Machines]
UniLM: Two dogs are throwing frisbees at each other.
BART: A dog throws a frisbee and a dog catches it.
T5: dog catches a frisbee and throws it to a dog

https://inklab.usc.edu/CommonGen/
Role Of Knowledge

- **dog**
  - capable of
    - **play**
    - **frisbee**
  - **play**
    - **game**
  - **throw**
  - has property
    - **fun for dog**
    - **located at**
      - **park**
  - **catch**
    - **frisbee**
    - **type of**
      - **antonym**
  - **related to**
    - **game**
  - **located at**
    - **park**
  - **created by**
    - **play**
    - **game**
  - **wants to**
    - **play**
    - **frisbee**
  - **used for**
    - **frisbee**
    - **synonym**
    - **flying disk**
    - **subclass of**
      - **disc**
Common Sense Knowledge Graphs

- **Cyc** [Lenat et al., 1984]
- **Open Mind Common Sense** [Minski, Singh, Havasi, 1999]
- **ConceptNet** [Liu, Singh, 2004]
- **WebChild** [Tandon et al., 2014]
- **WebChild 2.0** [Tandon et al., 2017]
- **OpenCyc 4.0** [Lenat 2012]
- **NELL** [Carlson et al., 2010]
- **NELL** [Mitchell et al., 2015]
- **Wikidata** [Vrandečić, 2012]
- **COMET** [Bosselut et al., 2019]
- **Atomic** [Sap et al., 2019]
- **ConceNet 5.5** [Speer et al., 2017]

*Image credit: USC Viterbi School of Engineering*
Dimensions Of Common Sense Knowledge

**Representation**
- symbolic
- natural language
- neural

**Creation method**
- expert input
- crowdsourcing
- information extraction, machine learning

**Knowledge type**
- entities and actions
- inferential/rules

**Topic**
- general
- social
Why is top-down knowledge necessary?

“In Artificial intelligence, commonsense knowledge is the set of background information that an individual is intended to know or assume and the ability to use it when appropriate.”

**Argument:** This knowledge cannot be acquired simply through text (or in an otherwise ‘inductive’ fashion)
Taxonomy of 30 representational areas
Example of a ‘top-down’ CSKG: Cyc
Limitations of top-down CSKGs

Many of the same issues that other top-down systems (including, famously, expert systems) have, such as brittleness, expense of acquisition...

Even if it were possible, we can never get away from language models completely
The many faces of ConceptNet
ATOMIC: 880,000 triples for AI systems to reason about causes and effects of everyday situations
Commonsense Knowledge in Wikidata

shower part of bathroom

reading uses written work

queen follows jack

political opposition opposite of government
Wikidata-CS Is Small But Novel

ConceptNet

3.4M edges

Wikidata-CS

102K edges

2.4K edges
Commonsense Knowledge Sources

• ConceptNet
  – Information about everyday objects, actions, states and relationships among them, extensive links to WordNet
  – Incomplete coverage, “related-to” accounts for 75% of statements

• ATOMIC
  – Pre- and post-states for events and their participants, physical and mental aspects covered
  – Only 25% of nodes have links to ConceptNet, difficult to combine with other resources

• WordNet
  – Meanings of words & relationships to other words, high coverage, many resources have links to WordNet, example sentences
  – No description of the properties of objects or roles in verbs, only is-a and part-of relations

• VerbNet, FrameNet
  – Defines participants/roles for a large number of situations/frames, links to verbs, syntactic forms and example sentences
  – No semantic typing of roles, many roles are very abstract (e.g., Agent), lacks info about state changes, or pre-post conditions

• Visual Genome
  – “Visual” commonsense, many possible attributes, relationships/actions among objects, linked to WordNet, many edges for a KG
  – No abstraction mechanism to understand prevalence of relations

• Wikidata
  – Comprehensive descriptions of objects, both specific (named entities) and generic (nouns)
  – Sparse information about events and states, much knowledge is on instance-level and abstraction is non-trivial
Consolidation Hypothesis

*Integrating multiple knowledge sources in CSKG is beneficial for downstream reasoning tasks.*
Principles for a modular and useful CSKG

P1. Embrace heterogeneity of nodes
objects, classes, words, actions, frames, states

P2. Reuse edge types across resources
/r/HasProperty from ConceptNet applicable for attributes in Visual Genome

P3. Leverage external links
many sources map to WordNet

P4. Generate high-quality probabilistic links
many facts not explicitly stated

P5. Enable access to labels
text labels and aliases are the key, in particular for NLP use cases
The Commonsense Knowledge Graph (CSKG)

7 sources
2.3M nodes
6M edges

Neuro-Symbolic Reasoning
Approaches

Knowledge enhances language models

Language models fill in knowledge gaps
A KG-augmented QA Framework

- **Context Module**
  - Encode question and answer choices as unstructured evidence
- **Knowledge Module**
  - Encode knowledge facts (paths) as structured evidence
- **Reasoning Module**
  - Score a question-choice pair based on un/structured evidence
Our final takeaways

- Commonsense (CS) reasoning is a difficult general AI problem that has come of age
  - Ironically, exposed both the strengths and limitations of neural networks, including language representation learning
  - We hypothesize that a neuro-symbolic approach is necessary for CS reasoning
- CS knowledge, appropriately contextualized, is critical for robust CS reasoning and QA
- Much progress has been achieved in integrating multiple sources into a single CSKG, but many open challenges remain


