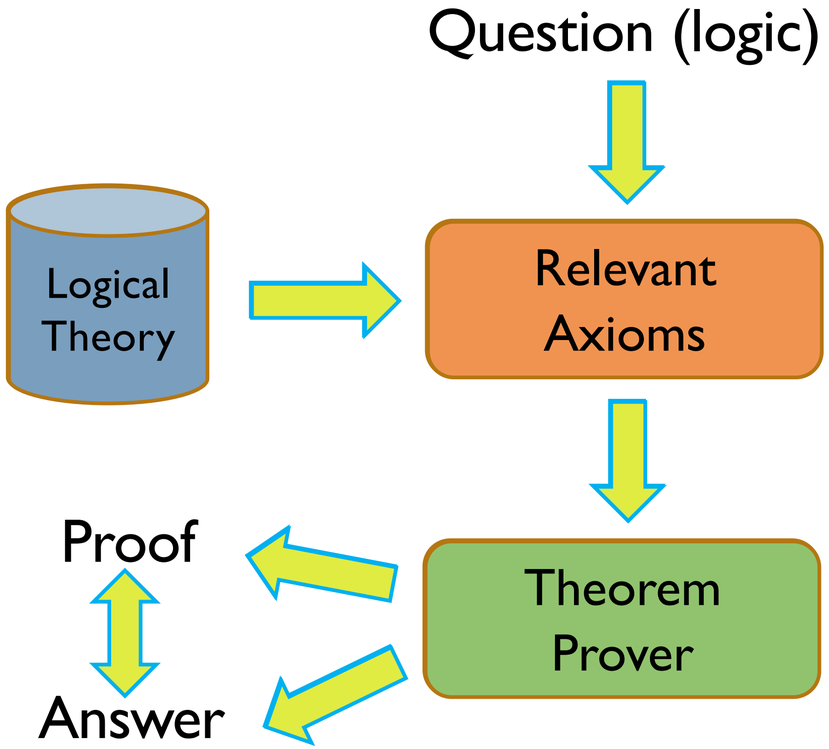
Probabilistic Graph Reasoning for Natural Proof Generation

Changzhi Sun
Xinbo Zhang
Jiangjie Chen
Chun Gan
Yuanbin Wu
Jiaze Chen
Hao Zhou
Lei Li

Reasoning Over Formal Representation

- Pros
  - Interpretable
  - Easy combine human knowledge
- Cons
  - Knowledge acquisition bottleneck
  - Brittleness
    - when confront with unusual or atypical cases



Formal Reasoning

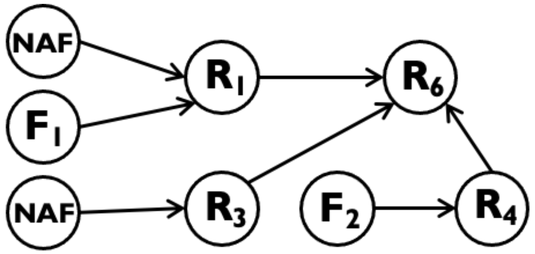
[Saha+ 2020, Musen&Lei 1988]

Reasoning over Natural Language

- **Input**: a set of facts and rules and a question expressed in natural language.
- **Output**:predict the answer and provide proof to prove or disprove the question.
- **Proof:**
  - Node: fact, rule or NAF
  - Edge: logical deduction
- **Potential advantages**
  - Write theories in natural language
  - Have the machine apply general knowledge

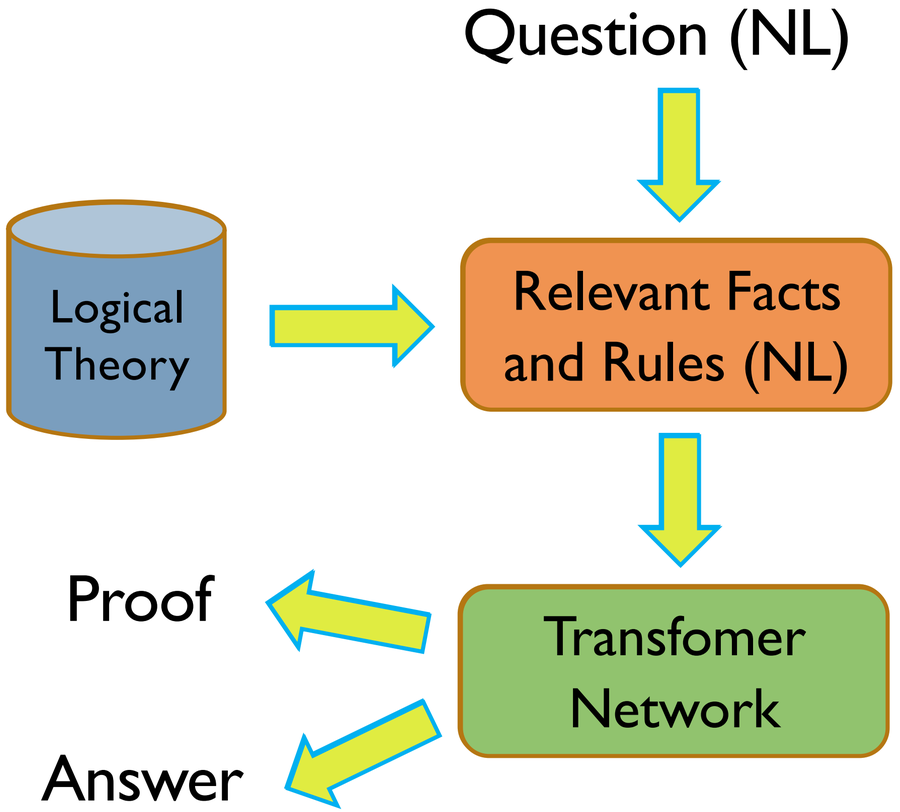
[Clark+ 2020, Saha+ 2020]

Facts :
F_1: The circuit includes the battery.
F_2: The wire is metal.
F_3: The circuit includes the bell.
Rules :
R_1: If the circuit includes the battery and the battery is not flat then the circuit is powered.
R_2: If the circuit includes the switch and the switch is on then the circuit is complete.
R_3: If the circuit does not have the switch then the circuit is complete.
R_4: If the wire is metal then the wire is conducting.
R_5: If the wire is plastic then the wire is not conducting.
R_6: If the circuit is powered and the circuit is complete and the wire is conducting then the current runs through the circuit.
Question : The current runs through the circuit.
Answer : True
Proof :



Existing Solution

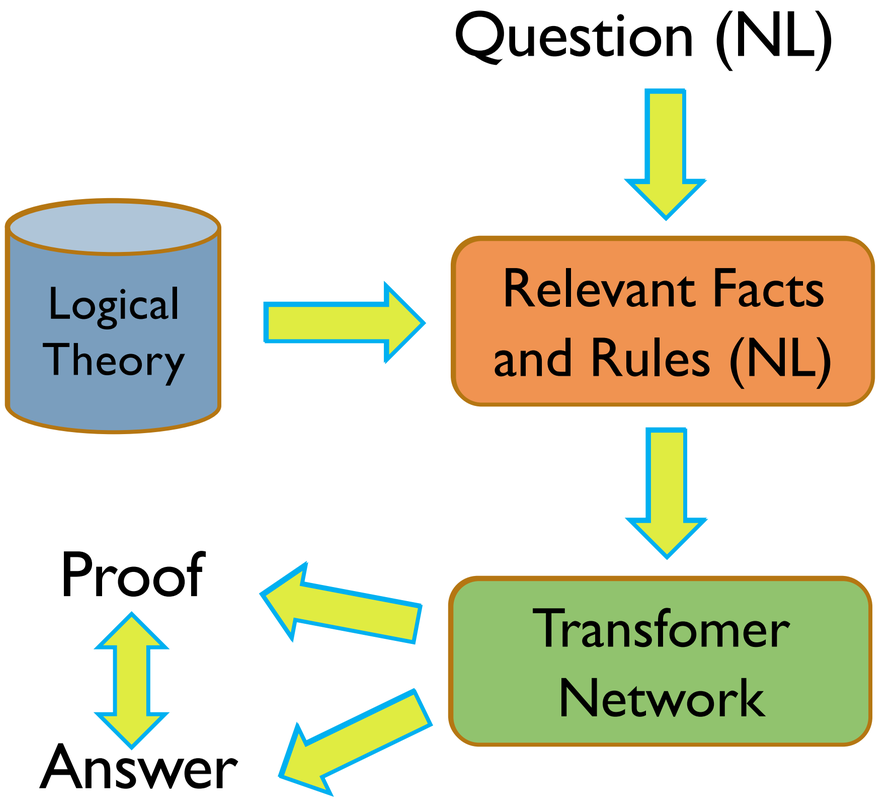
- **PRover**
  - Three sub-task, multi-task learning
    - Question answering
    - Node prediction
    - Edge prediction
  - Nodes, edges and answer are independent on each other



Our Solution

$$P(Y = y) \propto \Phi^A(a) \prod_i \Phi_i^V(v_i, a) \prod_{i,j} \Phi_{ij}^E(v_i, v_j, e_{ij}, a)$$

- **PRobr**
  - Undirected graphical model
  - Nodes, edges and answer are dependent on each other
  - Learning by variational approximation



Results

- Results
  - Fully supervised setting: **PRobr > PRover**
  - Few-shot & zero shot setting: **PRobr >> PRover**
    - Comparable proof performance
    - Very high QA performance (+ 10%-30%)

Fully supervised

D	Cnt	QA			PA		FA	
		RT	PV	PB	PV	PB	PV	PB
0	6299	100	100	**100**	98.4	**98.4**	98.4	**98.4**
1	4434	98.4	99.0	**99.9**	93.2	**94.3**	93.1	**94.3**
2	2915	98.4	98.8	**99.9**	84.8	**86.1**	84.8	**86.1**
3	2396	98.8	99.1	**100**	80.5	**82**	80.5	**82**
4	2134	99.2	98.8	**100**	72.5	**76.1**	72.4	**76.1**
5	2003	99.8	99.3	**100**	65.1	**72.2**	65.1	**72.2**
All	20192	99.2	99.3	**99.9**	87.1	**88.8**	87.1	**88.8**

Few-shot

Train Data		QA		PA		FA	
		PV	PB	PV	PB	PV	PB
	100%	99.3	**99.9**	87.1	**88.8**	87.1	**88.8**
RC	10%	94.5	**99.9**	**63.6**	60.4	**63.3**	60.4
	5%	80.6	**99.7**	34.0	**44.2**	32.1	**44.2**
	1%	70.2	**88.2**	20.0	**21.6**	15.1	**20.3**
RQ	30k	97.8	**99.9**	72.5	**86.8**	72.4	**86.8**
	10k	87.1	**99.9**	44.0	**72.4**	42.7	**72.3**
	1k	51.3	**82.1**	**28.0**	21.1	15.0	**18.4**

Zero-shot

Test	Cnt	QA			PA		FA	
		RT	PV	PB	PV	PB	PV	PB
B1	40	97.5	95.0	**100.0**	92.5	**100.0**	92.5	**100.0**
B2	40	100	95.0	**100.0**	95.0	**100.0**	95.0	**100.0**
E1	162	96.9	100	**100.0**	95.1	**97.5**	95.1	**97.5**
E2	180	98.3	100	**100.0**	91.7	**93.3**	91.7	**93.3**
E3	624	91.8	89.7	**98.2**	72.3	**79.3**	71.8	**79.3**
E4	4224	76.7	84.8	**95.6**	**80.6**	77.7	**80.6**	77.7
All	5270	80.1	86.5	**96.3**	**80.7**	79.3	**80.5**	79.3

Paper & Code