

图神经网络与认知推理

张鹏

2019年12月

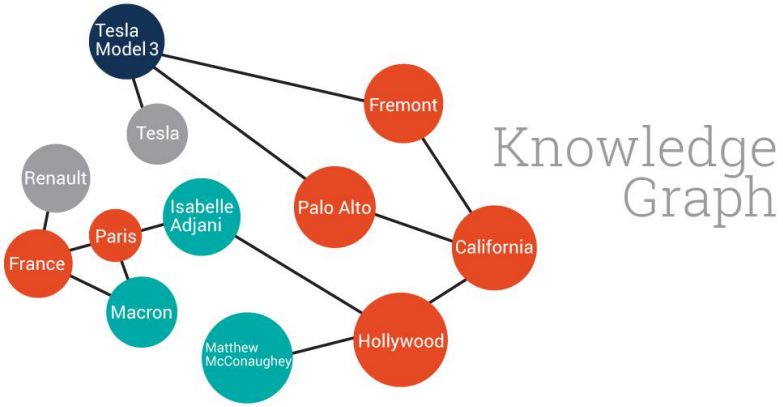
北京智谱华章科技有限公司
Beijing Knowledge Atlas Technology Co., Ltd

知识图谱

- “Knowledge graph”由Google于2012年提出
- 知识工程，专家系统
- CYC: 世界上历史最长的AI项目 (1985)



Edward Feigenbaum
Father of KB
Turing Award



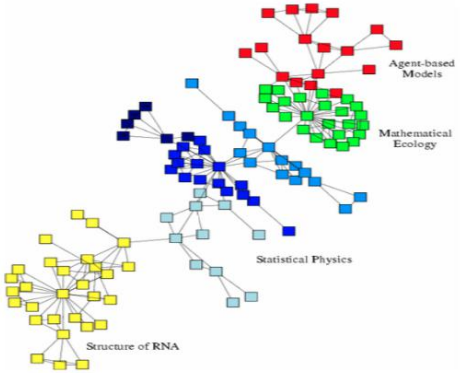


Tim Berners Lee
Father of WWW
Turing Award

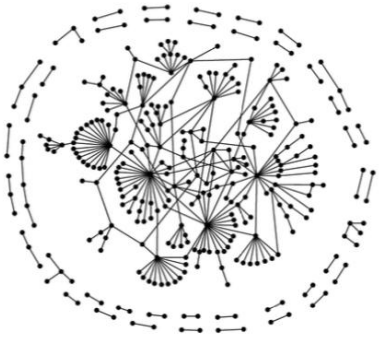
数据中的网络[1]



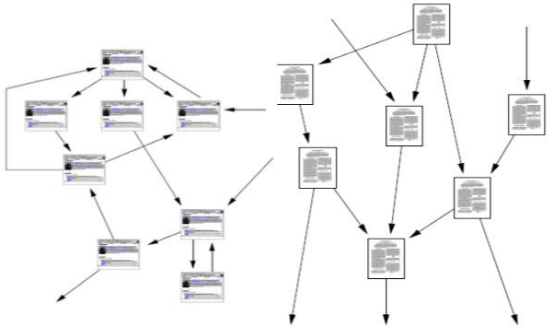
Social networks



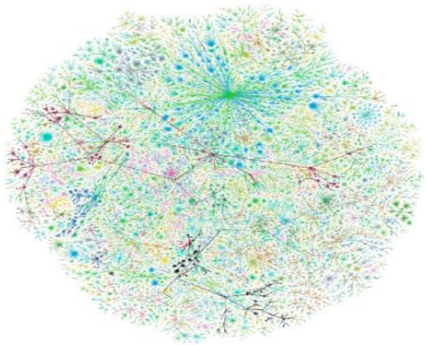
Economic networks



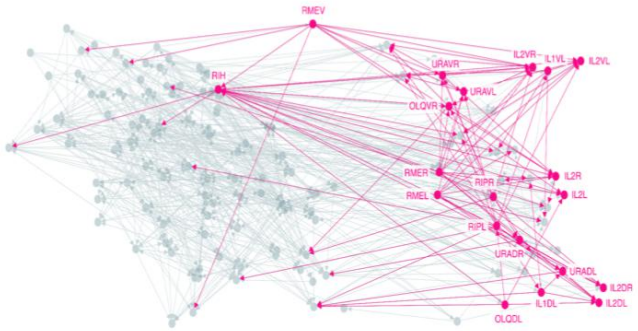
Biomedical networks



Information networks:
Web & citations



Internet



Networks of neurons

1. Example from Jure's slides

一个挑战性问题：多跳知识问答

多跳问答

Question: Who is the director of the 2003 film which has scenes in it filmed at the Quality Cafe in Los Angeles?

Quality Café

The Quality Cafe is a now-defunct diner in Los Angeles, California. The restaurant has appeared as a location featured in a number of Hollywood films, including Old School, Gone in 60 Seconds, ...

Los Angeles

Los Angeles is the most populous city in California, the second most populous city in the United States, after New York City, and the third most populous city in North America.

Alessandro Moschitti

Alessandro Moschitti is a professor of the CS Department of the University of Trento, Italy. He is currently a Principal Research Scientist of the Qatar Computing Research Institute (QCRI)



WIKIPEDIA
The Free Encyclopedia

Old School

Old School is a 2003 American comedy film released by Dream Works Pictures and The Montecito Picture Company and directed by Todd Phillips.

Todd Phillips

Todd Phillips is an American director, producer, screenwriter, and actor. He is best known for writing and directing films, including Road Trip (2000), Old School (2003), Starsky & Hutch (2004), and The Hangover Trilogy.

Tsinghua University

Tsinghua University is a major research university in Beijing and dedicated to academic excellence and global development. Tsinghua is perennially ranked as one of the top academic institutions in China, Asia, and worldwide...

挑战 1: 短视检索问题

Question: Who is the director of the 2003 film which has scenes in it filmed at the Quality Cafe in Los Angeles?

Document A

Old School is a 2003 American comedy film released by DreamWorks Pictures and The Montecito Picture Company…

Document B

Many directors shoot scenes in Hollywood, Los Angeles, which is notable as the home of the U.S. film industry.

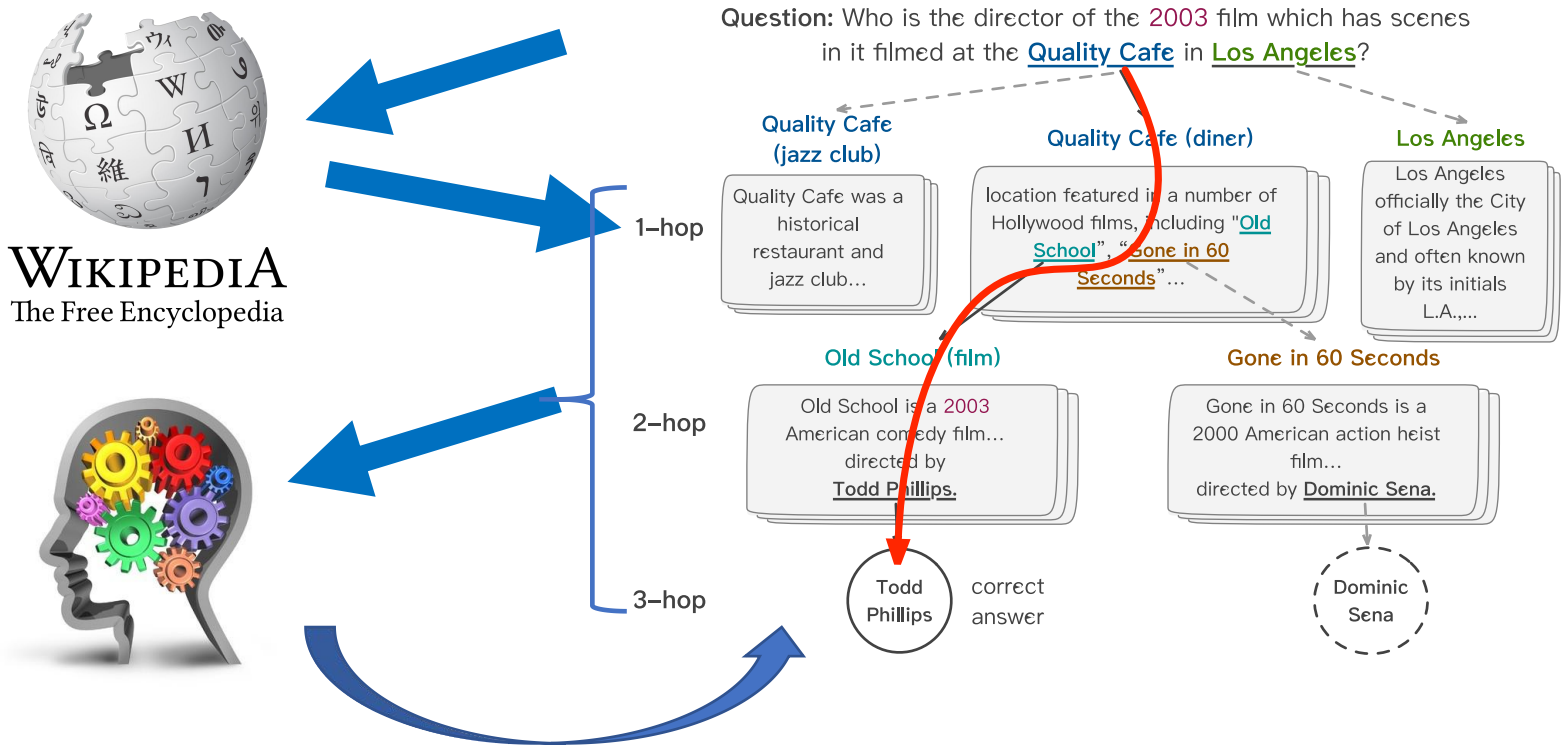
对于多跳问答的问题，问题和答案间拥有多少相似词汇已经不能作为答案检索的重要依据。

挑战 2: 可解释性

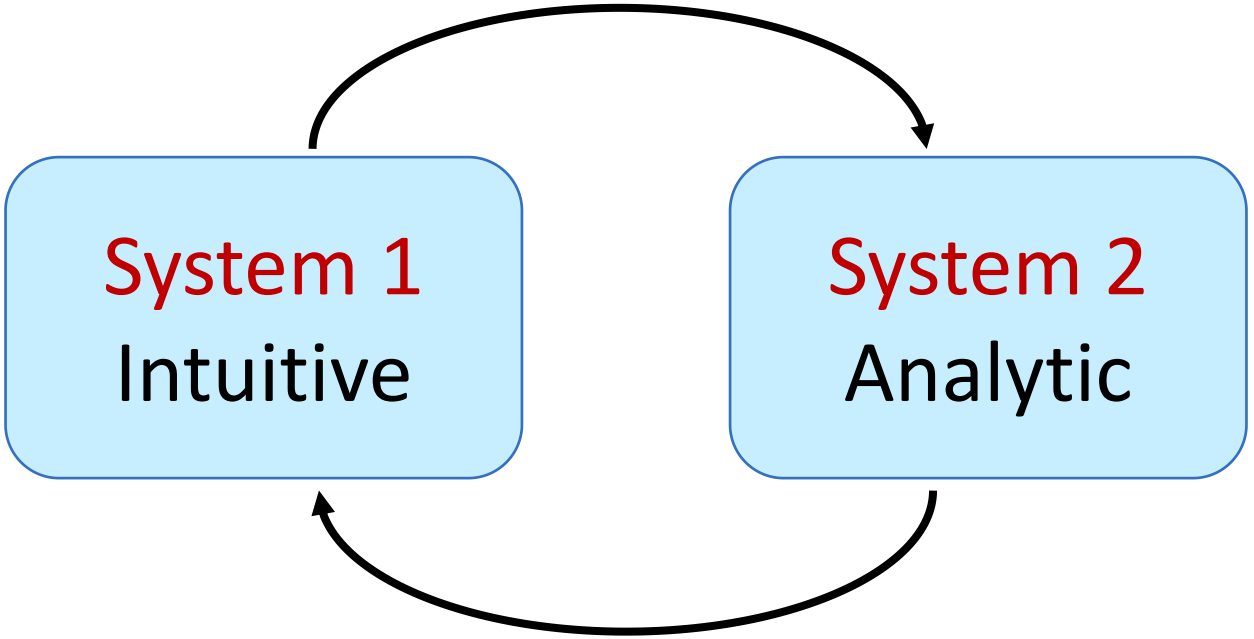
- 大部分阅读理解方法都只能看做**黑盒**:
  - 输入: 问题和文档
  - 输出: 答案文本块 (在文档中的起止位置)

- 如何让用户可以验证答案的对错:
  - 推理路径或者子图
  - 每个推理节点上的支撑事实
  - 用于对比的其他可能答案和推理路径

认知图谱: 表示, 推理和决策



认知科学的理论

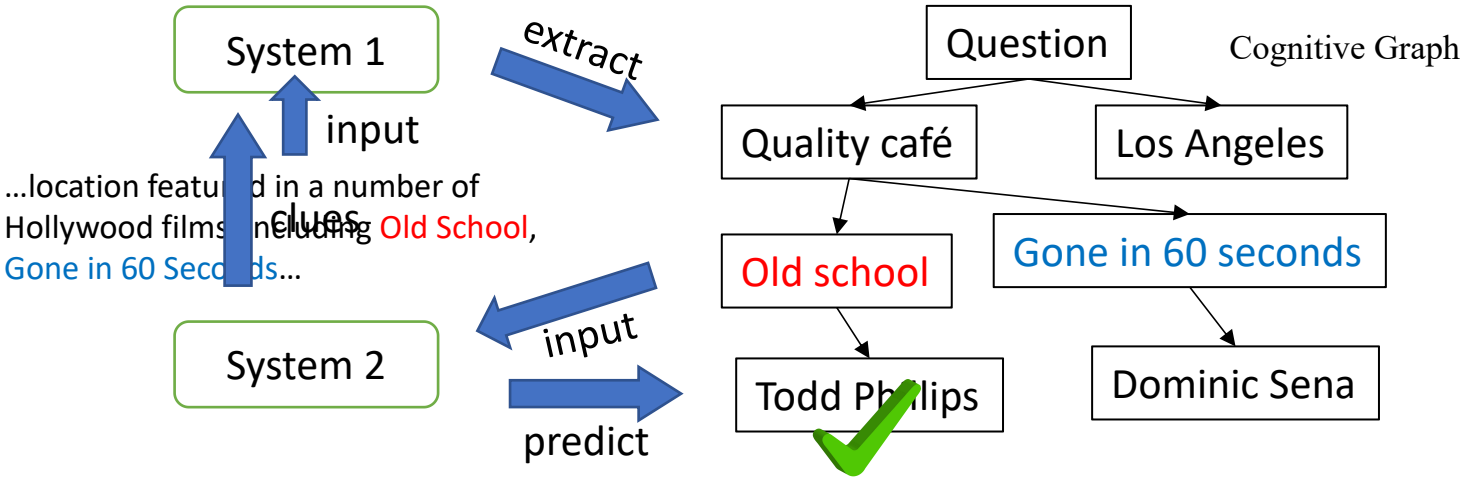


双通道处理理论

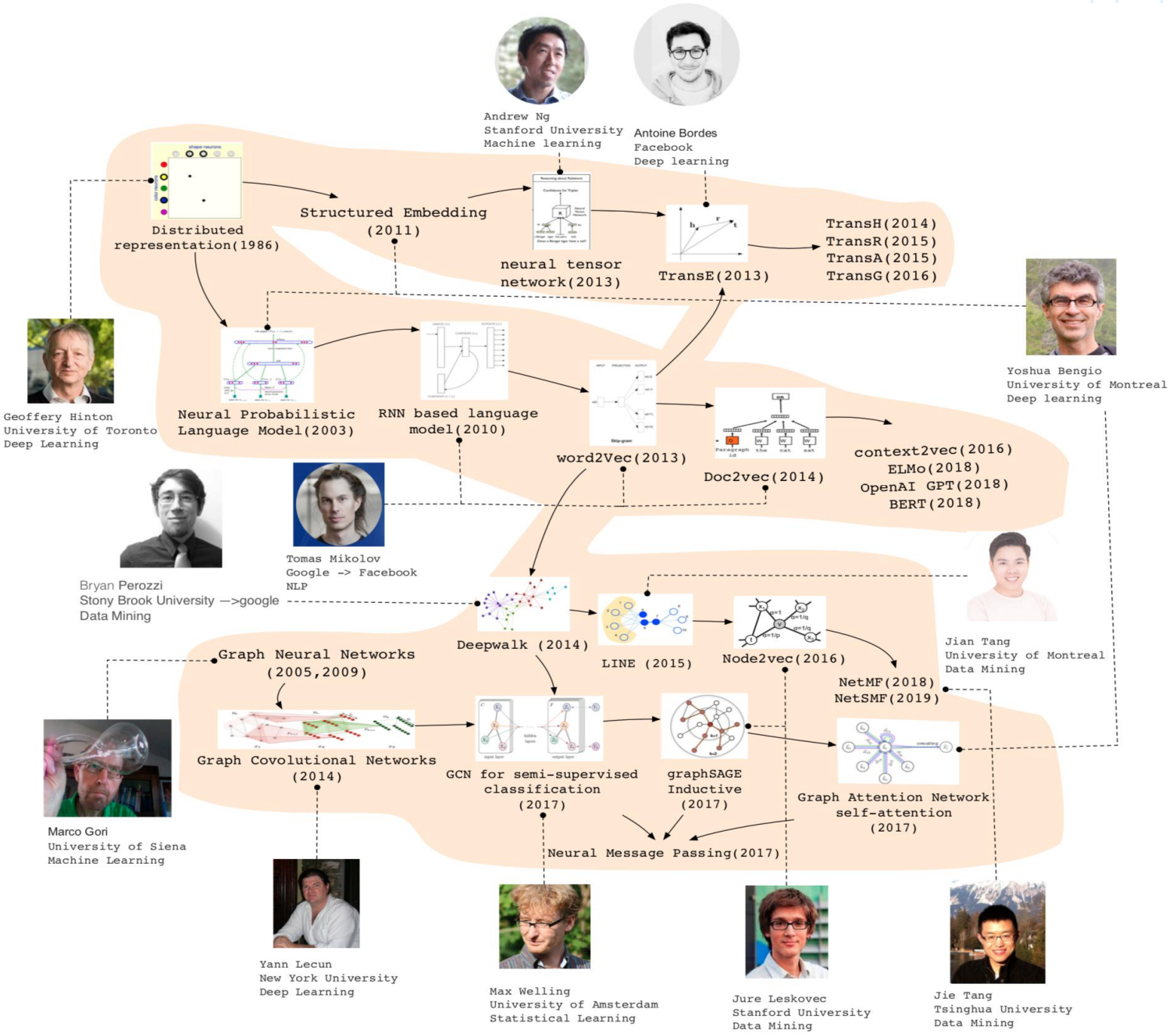
1. Jonathan St BT Evans. 1984. Heuristic and analytic processes in reasoning. British Journal of Psychology, 75(4):451–468.

CogQA: 基于认知图谱的问答

- 基于认知理论的**迭代**式框架
- System 1
  - 抽取实体构建认知图谱
  - 生成每个节点的语义向量
- System 2
  - 基于语义向量和图进行推理
  - 向System 1反馈线索用于抽取下一跳实体



问题1: 如何表示图中的一个实体?


Andrew Ng
Stanford University
Machine learning
Antoine Bordes
Facebook
Deep learning
Distributed representation(1986)
Structured Embedding (2011)
neural tensor network(2013)
TransE(2013)
TransH(2014)
TransR(2015)
TransA(2015)
TransG(2016)
Geoffery Hinton
University of Toronto
Deep Learning
Neural Probabilistic Language Model(2003)
RNN based language model(2010)
word2Vec(2013)
Doc2vec(2014)
context2vec(2016)
ELMo(2018)
OpenAI GPT(2018)
BERT(2018)
Yoshua Bengio
University of Montreal
Deep learning
Bryan Perozzi
Stony Brook University —>google
Data Mining
Tomas Mikolov
Google -> Facebook
NLP
Deepwalk (2014)
LINE (2015)
Node2vec(2016)
Jian Tang
University of Montreal
Data Mining
Graph Neural Networks (2005,2009)
NetMF(2018)
NetSMF(2019)
Graph Covolutional Networks (2014)
GCN for semi-supervised classification (2017)
graphSAGE Inductive (2017)
Graph Attention Network self-attention (2017)
Marco Gori
University of Siena
Machine Learning
Neural Message Passing(2017)
Yann Lecun
New York University
Deep Learning
Max Welling
University of Amsterdam
Statistical Learning
Jure Leskovec
Stanford University
Data Mining
Jie Tang
Tsinghua University
Data Mining

网络嵌入方法的矩阵形式统一

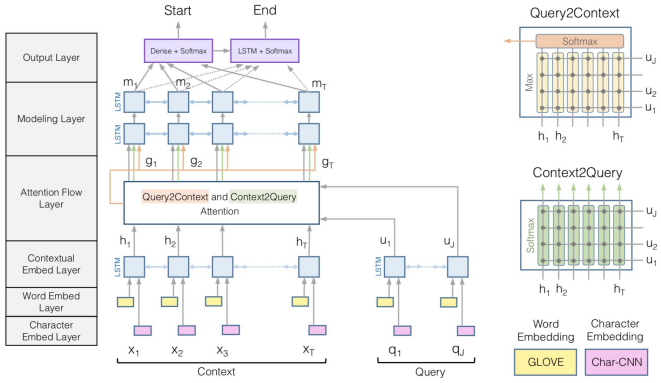
Algorithm	Closed Matrix Form
DeepWalk	$\log\left(\mathrm{vol}(G)\left(\frac{1}{T}\sum_{r=1}^{T}(\boldsymbol{D}^{-1}\boldsymbol{A})^r\right)\boldsymbol{D}^{-1}\right)-\log b$
LINE	$\log\left(\mathrm{vol}(G)\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\right)-\log b$
PTE	$\log\left(\begin{bmatrix}\alpha\,\mathrm{vol}(G_{\mathrm{ww}})(\boldsymbol{D}_{\mathrm{row}}^{\mathrm{ww}})^{-1}\boldsymbol{A}_{\mathrm{ww}}(\boldsymbol{D}_{\mathrm{col}}^{\mathrm{ww}})^{-1}\\ \beta\,\mathrm{vol}(G_{\mathrm{dw}})(\boldsymbol{D}_{\mathrm{row}}^{\mathrm{dw}})^{-1}\boldsymbol{A}_{\mathrm{dw}}(\boldsymbol{D}_{\mathrm{col}}^{\mathrm{dw}})^{-1}\\ \gamma\,\mathrm{vol}(G_{\mathrm{lw}})(\boldsymbol{D}_{\mathrm{row}}^{\mathrm{lw}})^{-1}\boldsymbol{A}_{\mathrm{lw}}(\boldsymbol{D}_{\mathrm{col}}^{\mathrm{lw}})^{-1}\end{bmatrix}\right)-\log b$
node2vec	$\log\left(\frac{\frac{1}{2T}\sum_{r=1}^{T}\left(\sum_u \boldsymbol{X}_{w,u}\underline{\boldsymbol{P}}^r_{c,w,u}+\sum_u \boldsymbol{X}_{c,u}\underline{\boldsymbol{P}}^r_{w,c,u}\right)}{(\sum_u \boldsymbol{X}_{w,u})(\sum_u \boldsymbol{X}_{c,u})}\right)-\log b$

1. Qiu et al. Network embedding as matrix factorization: unifying deepwalk, line, pte, and node2vec. *WSDM'18.* **The most cited paper in WSDM'18 as of May 2019**

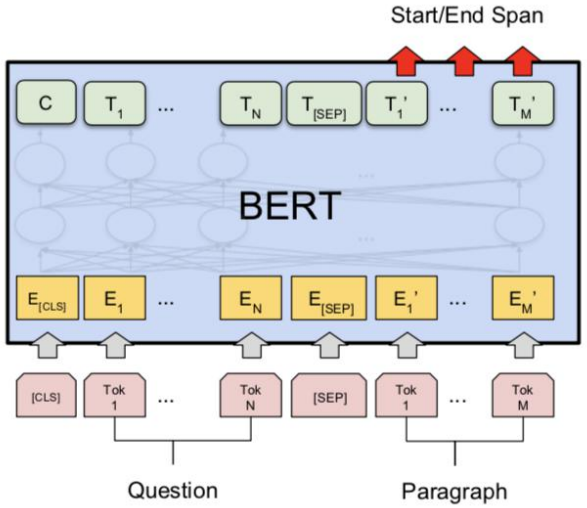
最新的进展: BIDAF, BERT, XLNet

- 目标：理解整个文档，而不仅仅是局部片段

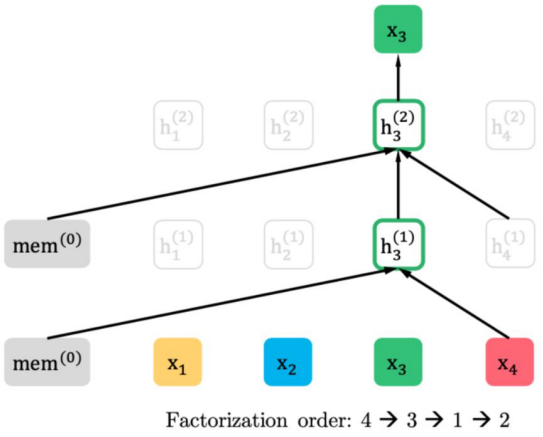
BiDAF



BERT



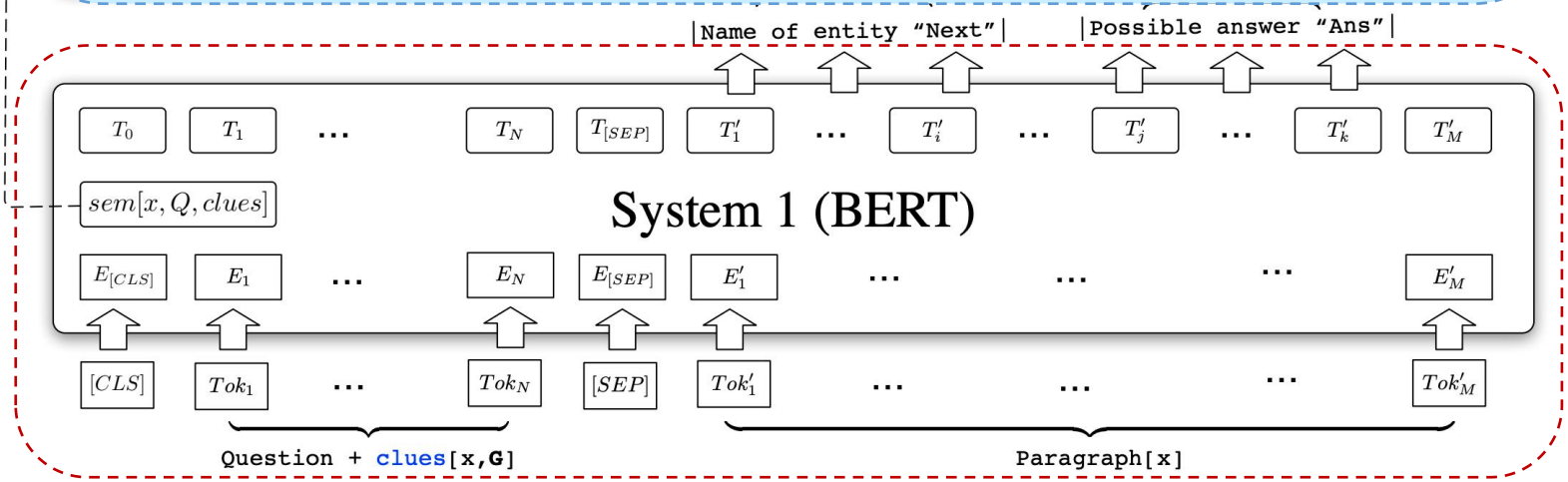
XLNet



认知图谱: 深度学习 + 双通道处理理论

?

System 2:
显式推理
和决策

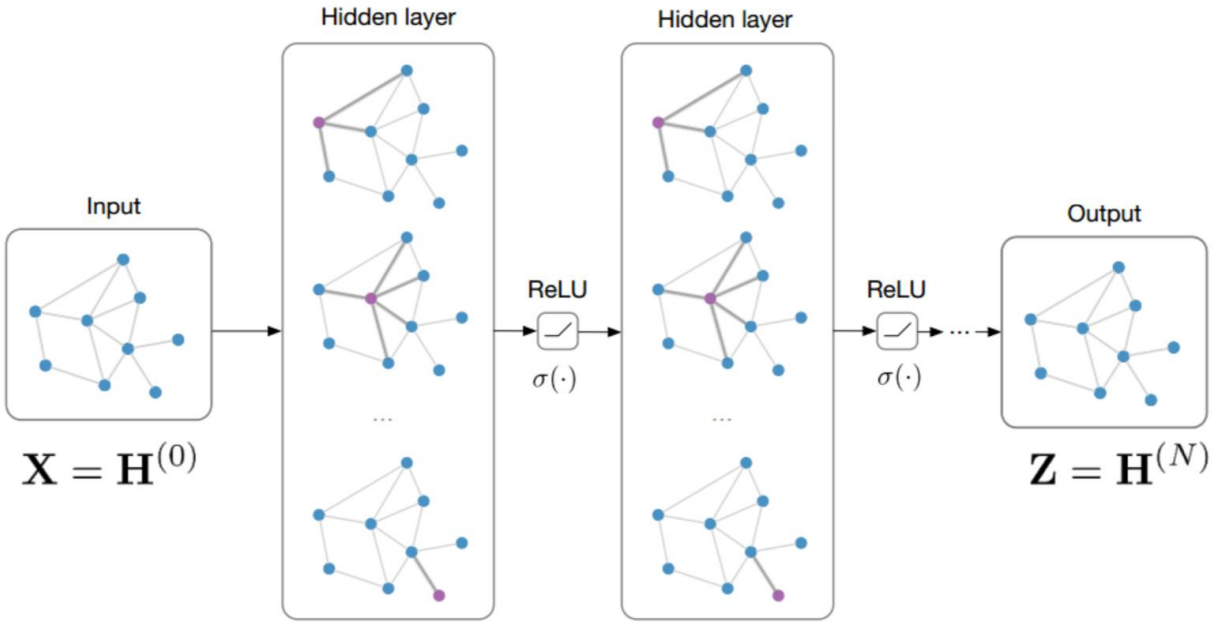


System 1:
隐式知识
扩展

问题2: 认知图谱中如何决策?

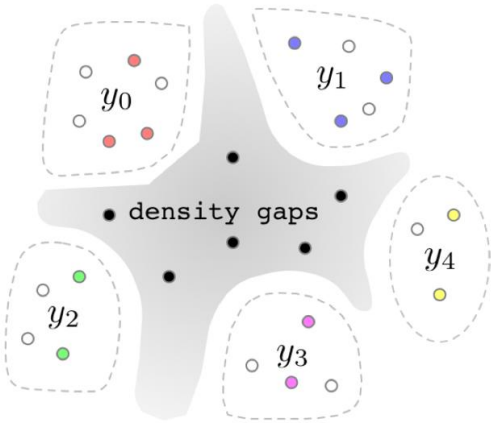
GNN/GCN Model Architecture

- **Input**: preprocess adjacency matrix $\hat{\mathbf{A}}$ and feature matrix $\mathbf{X} \in R^{N \times E}$



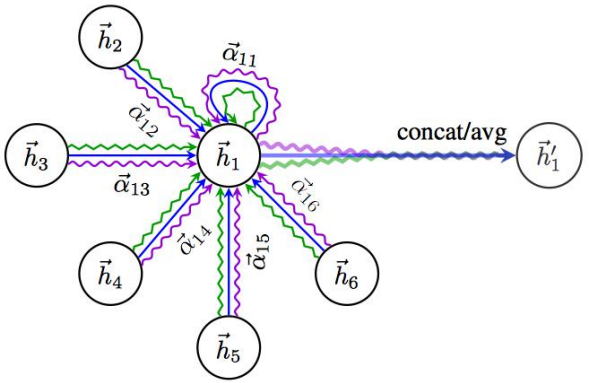
- where $\hat{\mathbf{A}} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$, $\tilde{A} = A + I_N$ is the adjacency matrix of graph G with added self-connections and $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$

Recent GCN Research



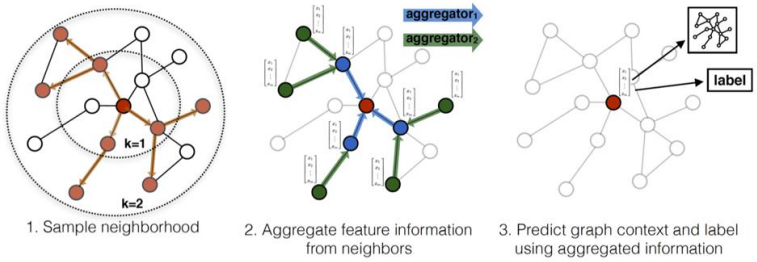
2018 GraphSGAN

FastGCN 2018

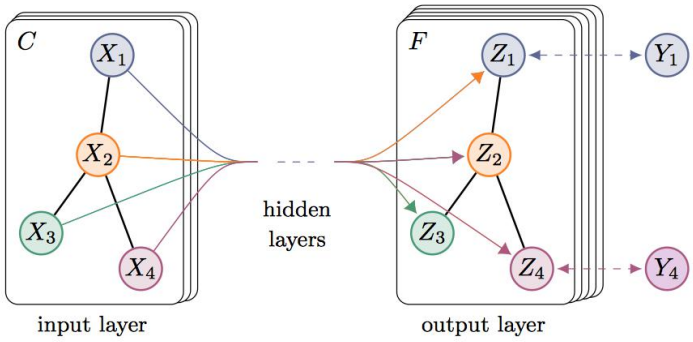


2018 GAT

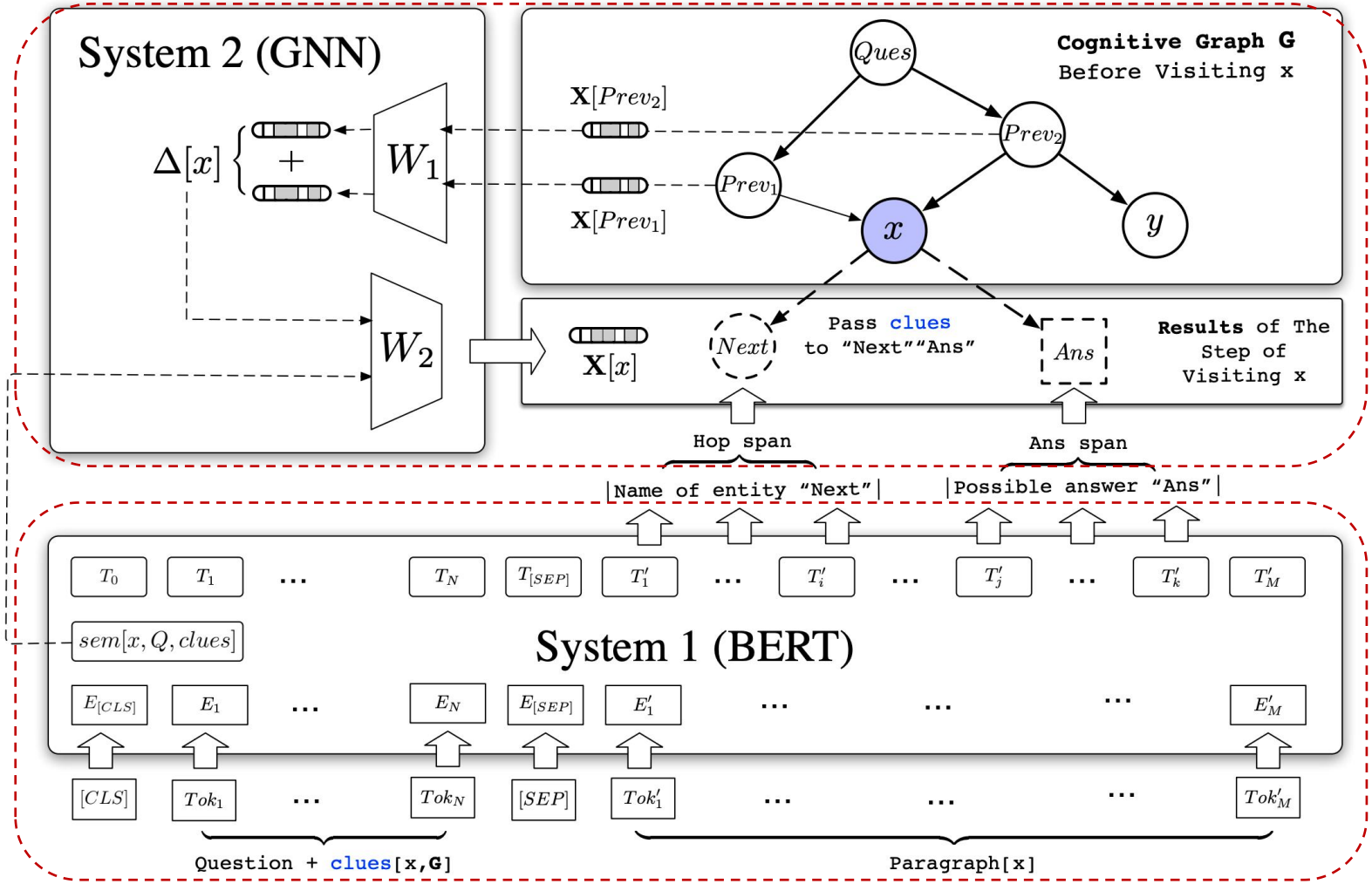
GraphSAGE 2017



2017 GCN



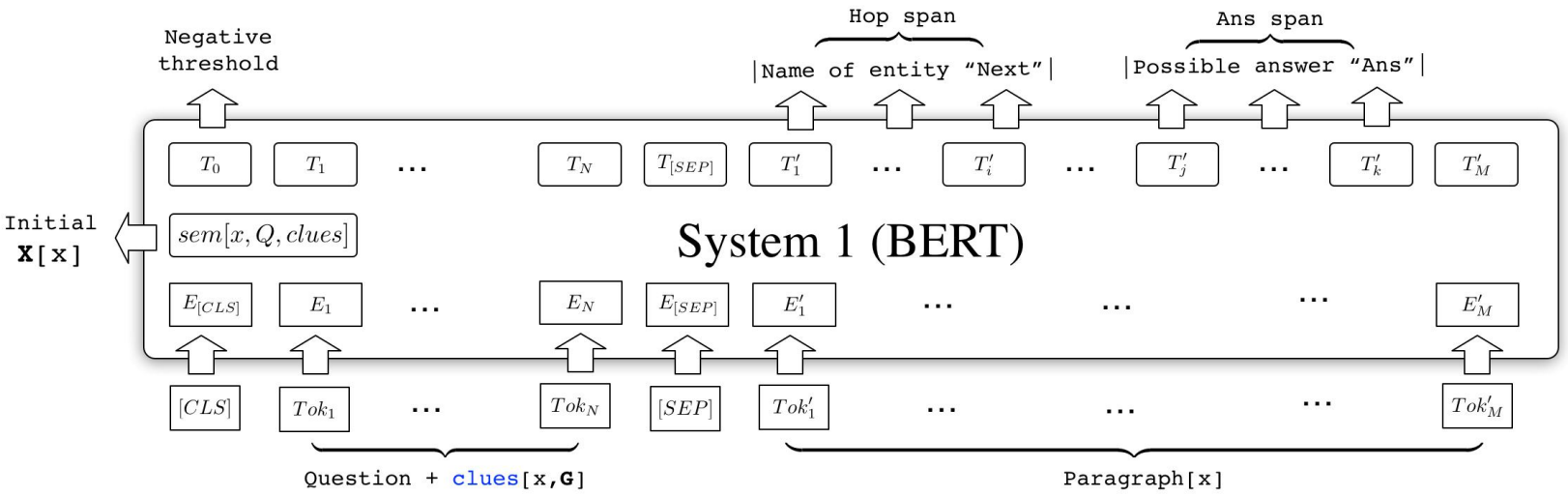
认知图谱: 深度学习 + 双通道处理理论



System 2:
显式推理
和决策

System 1:
隐式知识
扩展

System 1: 基于BERT的实现



$$[CLS]\ Question\ [SEP]\ clues[x, \mathcal{G}]\ [SEP] \quad Para[x]$$

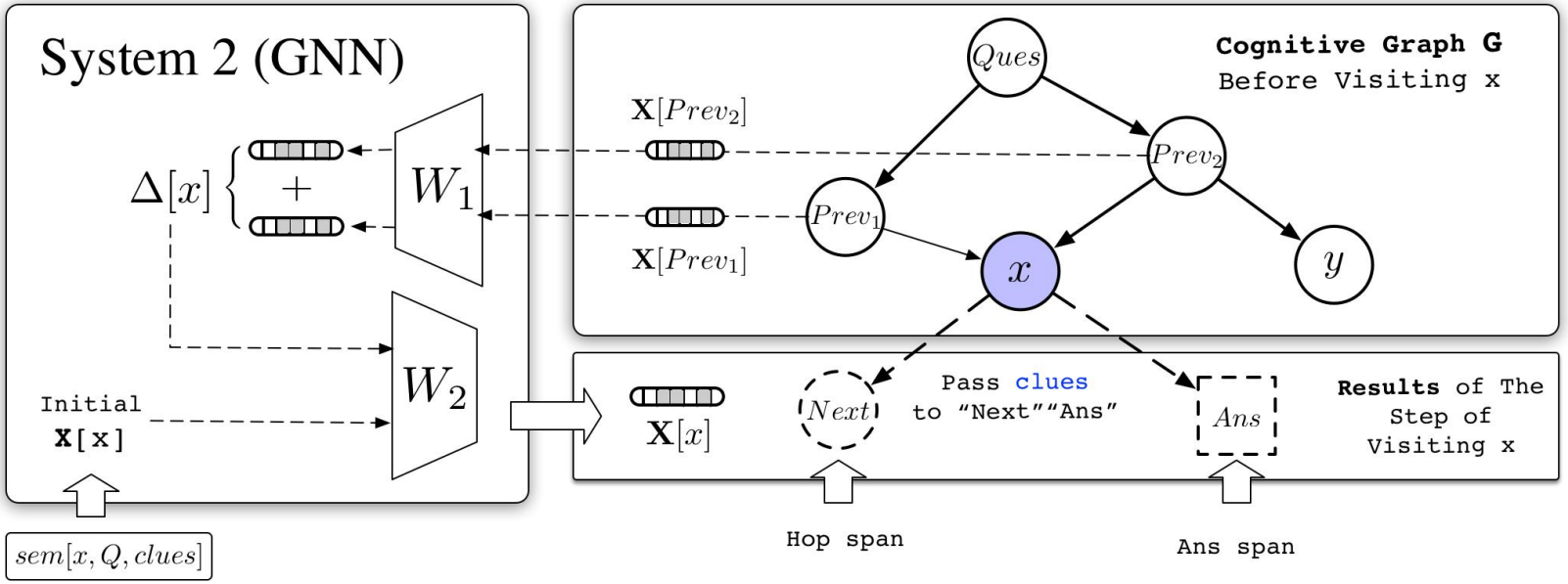
Sentence A　　Sentence B

$$P_{ans}^{start}[i] = \frac{e^{\mathbf{S}_{ans} \cdot \mathbf{T}_i}}{\sum_j e^{\mathbf{S}_{ans} \cdot \mathbf{T}_j}}$$

$$end_k = \underset{start_k \le j \le start_k + maxL}{\arg\max} P_{ans}^{end}[j]$$

- 抽取前k个下跳实体，并预测答案候选
- 生成实体的语义向量

System 2: 基于GNN的实现



At each step, hidden representations $\mathbf{X}$ for nodes are updated according to the propagation rules:

$$\Delta = \sigma((AD^{-1})^T \sigma(\mathbf{X}W_1))$$
$$\mathbf{X}' = \sigma(\mathbf{X}W_2 + \Delta)$$

Predictor F is a two-layer MLP, which predicts the final answer based on hidden representations $\mathbf{X}$:

$$answer = \underset{\text{answer node } x}{\arg\max}\ \mathcal{F}(\mathbf{X}[x])$$

CogQA算法

Algorithm 1: Cognitive Graph QA

Input:
System 1 model $\mathcal{S}_1$, System 2 model $\mathcal{S}_2$,
Question Q, Predictor $\mathcal{F}$,Wiki Database $\mathcal{W}$

Initialize cognitive graph $\mathcal{G}$ with entities mentioned in Q and mark them *frontier nodes*
repeat
  pop a node x from frontier nodes
  collect $clues[x, \mathcal{G}]$ from predecessor nodes of x
    // eg. $clues$ can be sentences where x is mentioned
  fetch $para[x]$ in $\mathcal{W}$ if any
  generate $sem[x, Q, clues]$ with $\mathcal{S}_1$ // initial $\mathbf{X}[x]$
  if *x is a hop node* **then**
    find hop and answer spans in $para[x]$ with $\mathcal{S}_1$
    for *y in hop spans* **do**
      if $y \notin \mathcal{G}$ *and* $y \in \mathcal{W}$ **then**
        create a new hop node for y
      if $y \in \mathcal{G}$ *and* $edge(x, y) \notin \mathcal{G}$ **then**
        add edge (x, y) to $\mathcal{G}$
        mark node y as a frontier node
    end
    for *y in answer spans* **do**
      add new answer node y and edge (x, y) to $\mathcal{G}$
    end
  end
  update hidden representation $\mathbf{X}$ with $\mathcal{S}_2$
until *there is no frontier node in $\mathcal{G}$ or $\mathcal{G}$ is large enough*;
Return $\arg\max_{\text{answer node } x} \mathcal{F}(\mathbf{X}[x])$

- 预处理
  - 用模糊匹配标记下跳实体和答案候选
  - 构建初始图谱
- 训练任务 #1 (System 1):
  - 下跳位置和答案候选抽取
- 训练任务 #2 (System 1 & 2):
  - 答案节点预测

性能测试

- HotpotQA （类似SQuAD的一个问答测试数据集）
- CogQA 在约3个月的时间内保持排名第一

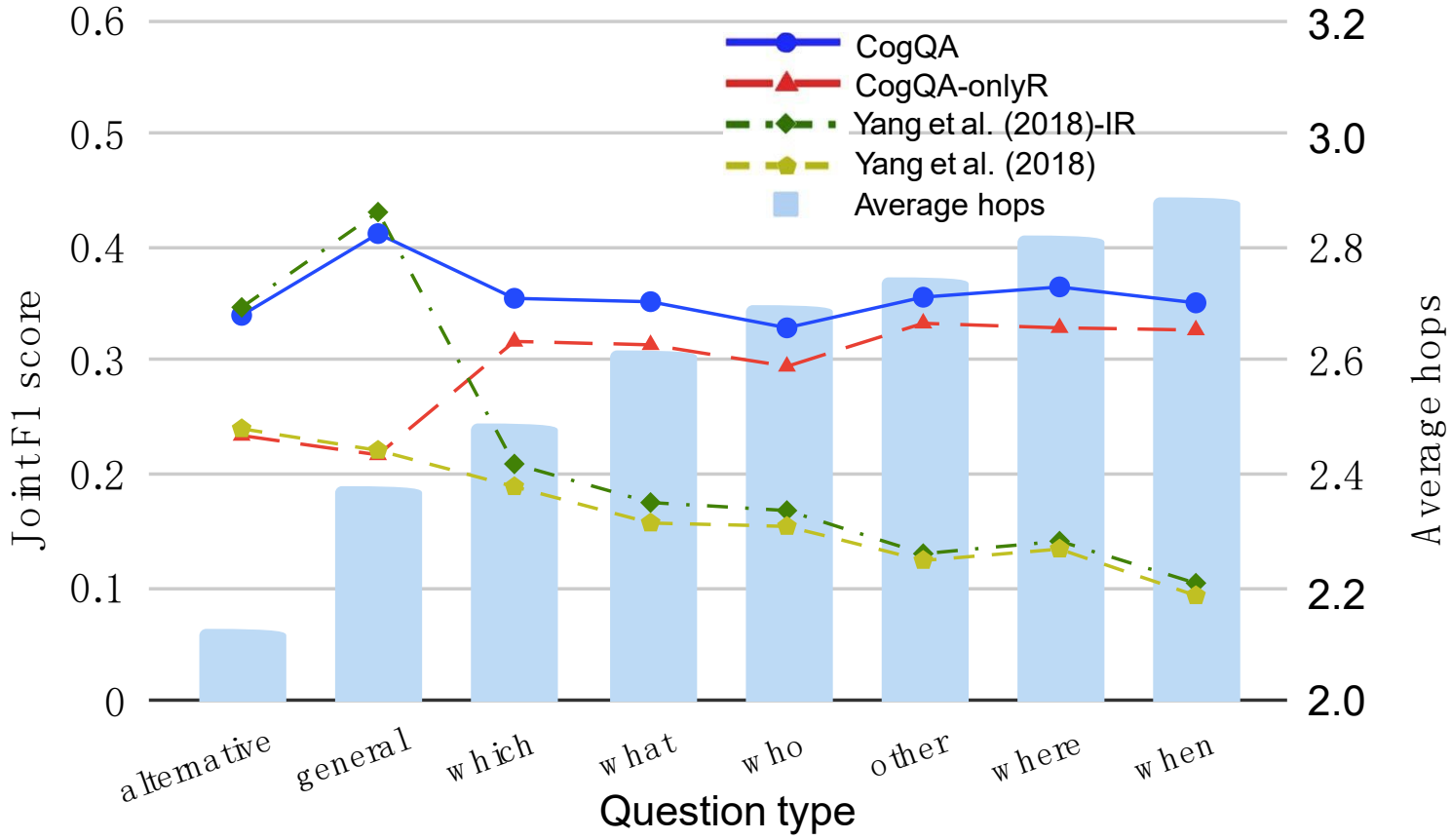
	Model	Ans				Sup				Joint			
		EM	F_1	Prec	Recall	EM	F_1	Prec	Recall	EM	F_1	Prec	Recall
Dev	Yang et al. (2018)	23.9	32.9	34.9	33.9	5.1	40.9	47.2	40.8	2.5	17.2	20.4	17.8
	Yang et al. (2018)-IR	24.6	34.0	35.7	34.8	10.9	49.3	52.5	52.1	5.2	21.1	22.7	23.2
	BERT	22.7	31.6	33.4	31.9	6.5	42.4	54.6	38.7	3.1	17.8	24.3	16.2
	CogQA-sys1	33.6	45.0	47.6	45.4	**23.7**	58.3	**67.3**	56.2	12.3	32.5	39.0	31.8
	CogQA-onlyR	34.6	46.2	48.8	46.7	14.7	48.2	56.4	47.7	8.3	29.9	36.2	30.1
	CogQA-onlyQ	30.7	40.4	42.9	40.7	23.4	49.9	56.5	48.5	**12.4**	30.1	35.2	29.9
	CogQA	**37.6**	**49.4**	**52.2**	**49.9**	23.1	**58.5**	64.3	**59.7**	12.2	**35.3**	**40.3**	**36.5**
Test	Yang et al. (2018)	24.0	32.9	-	-	3.86	37.7	-	-	1.9	16.2	-	-
	QFE	28.7	38.1	-	-	14.2	44.4	-	-	8.7	23.1	-	-
	DecompRC	30.0	40.7	-	-	N/A	N/A	-	-	N/A	N/A	-	-
	MultiQA	30.7	40.2	-	-	N/A	N/A	-	-	N/A	N/A	-	-
	GRN	27.3	36.5	-	-	12.2	48.8	-	-	7.4	23.6	-	-
	CogQA	**37.1**	**48.9**	-	-	**22.8**	**57.7**	-	-	**12.4**	**34.9**	-	-

Table 1: Results on HotpotQA (fullwiki setting). The test set is not public. The maintainer of HotpotQA only offers EM and F_1 for every submission. N/A means the model cannot find supporting facts.

** https://github.com/THUDM/CogQA

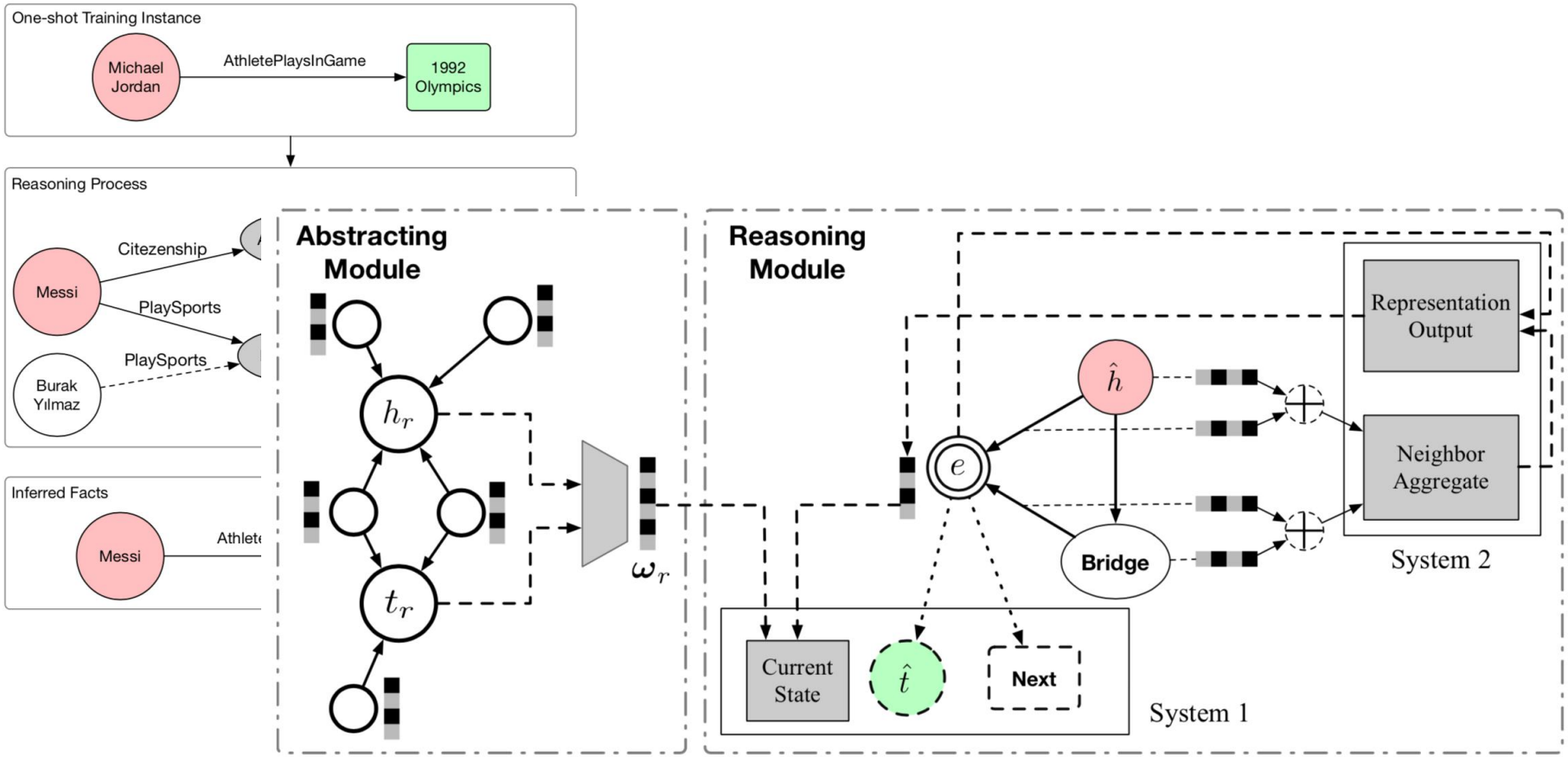
推理的力量

CogQA 在**多跳问题**上表现尤其优异



其他应用

- 用认知图谱的方法**补全知识图谱**



智谱: 认知图谱



Edward Feigenbaum
Turing Award Winner

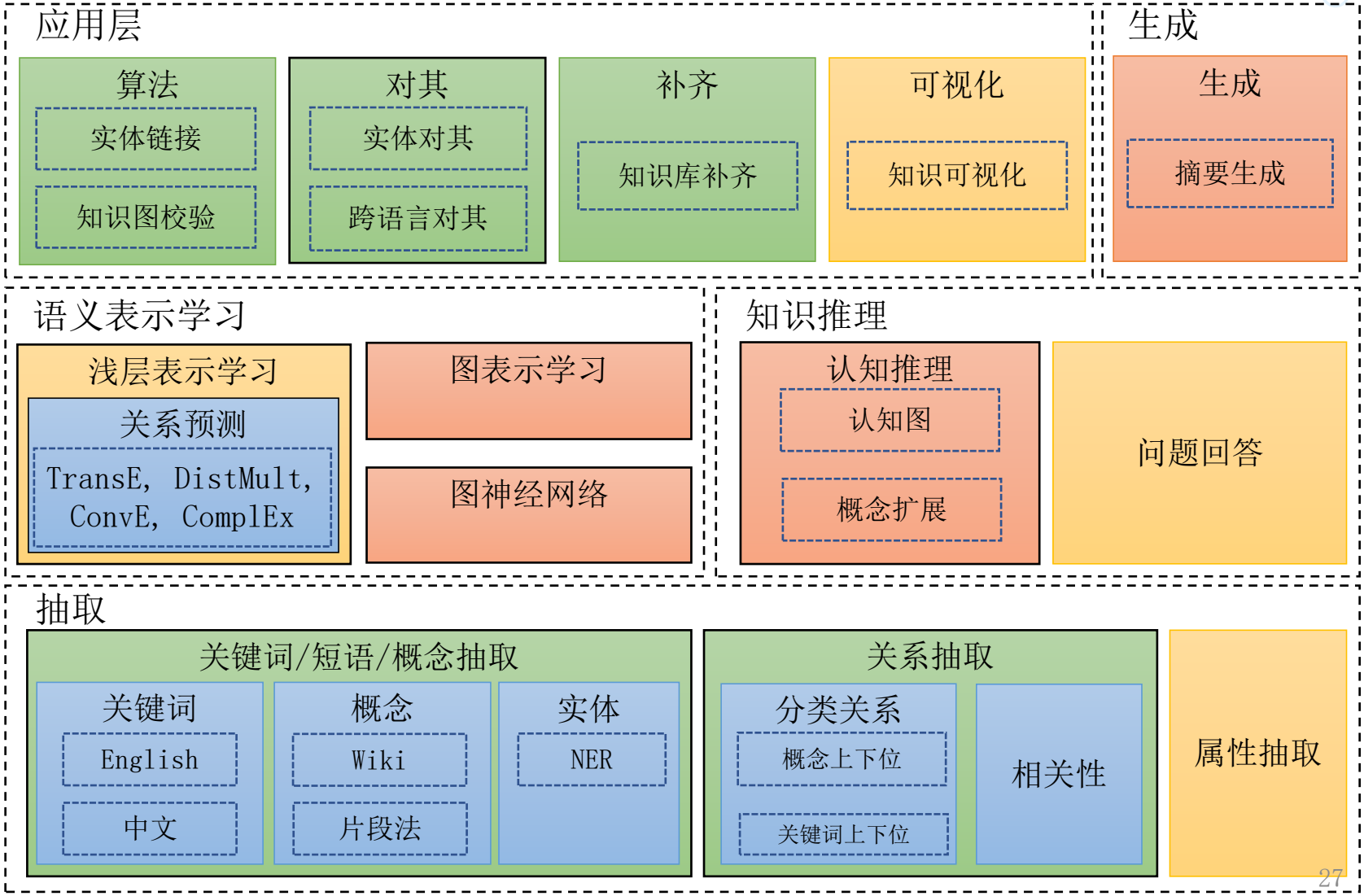
亿级常识知识图谱



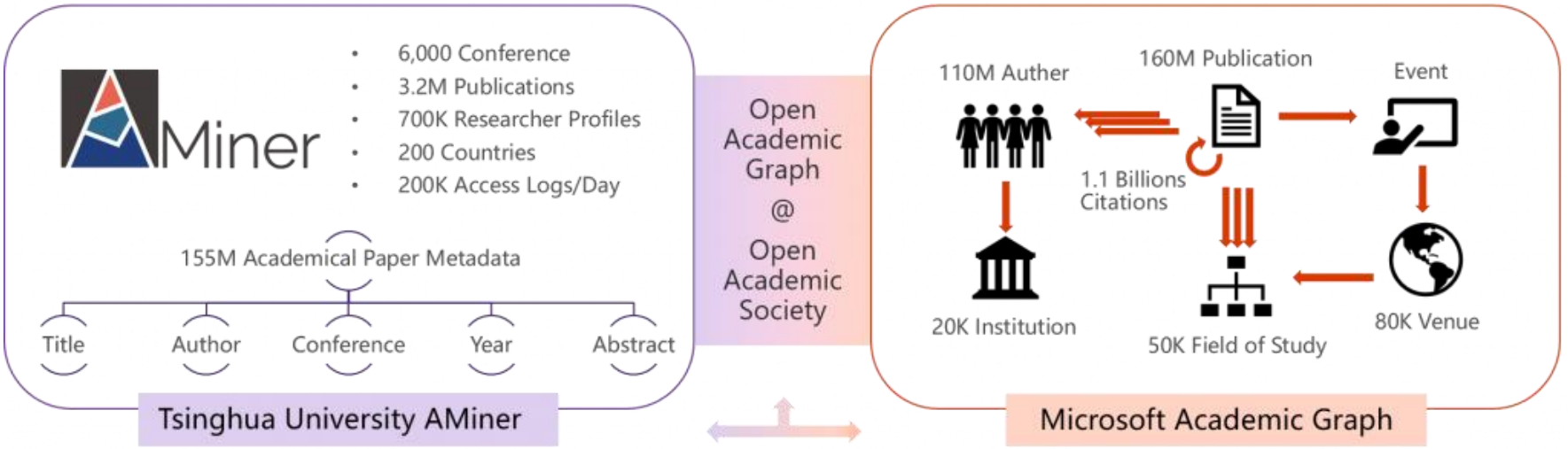
Tim Berners Lee
Turing Award Winner

Big Data ⇨ Knowledge ⇨ Intelligence

“智谱”工具集


应用层
算法
实体链接
知识图校验
对其
实体对其
跨语言对其
补齐
知识库补齐
可视化
知识可视化
生成
生成
摘要生成
语义表示学习
浅层表示学习
关系预测
TransE, DistMult, ConvE, ComplEx
图表示学习
图神经网络
知识推理
认知推理
认知图
概念扩展
问题回答
抽取
关键词/短语/概念抽取
关键词
English
中文
概念
Wiki
片段法
实体
NER
关系抽取
分类关系
概念上下位
关键词上下位
相关性
属性抽取

案例：大规模异构学术图谱的融合



Linking *large-scale heterogeneous* academic graphs

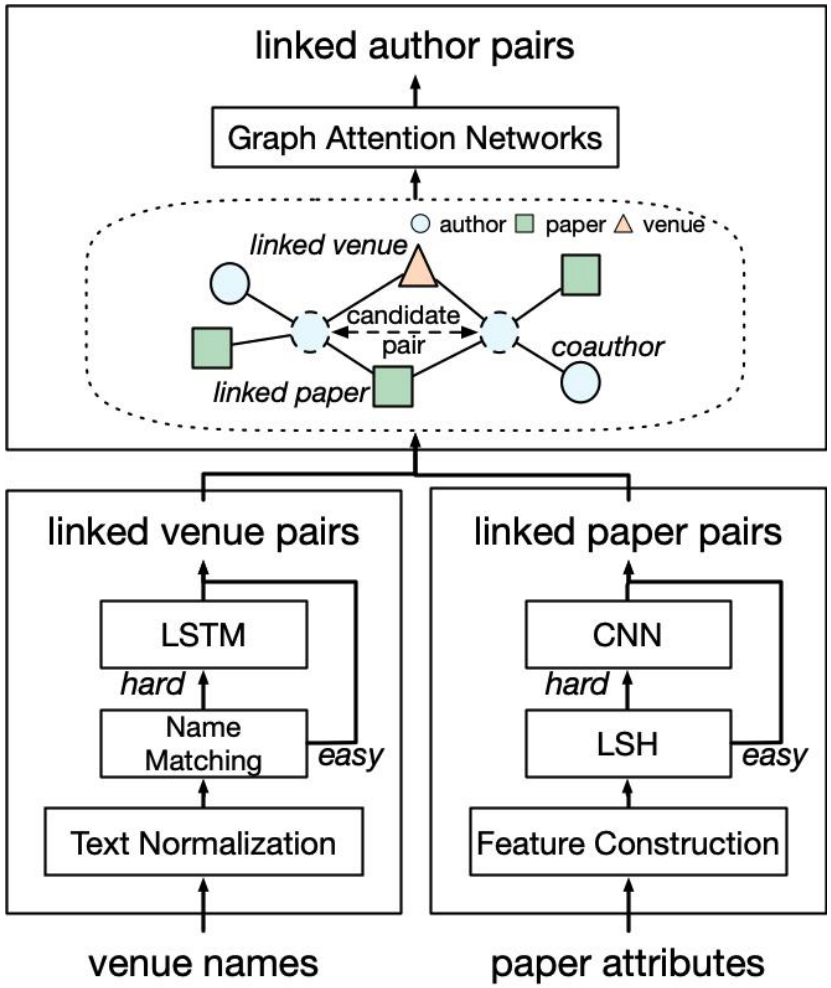
1. F. Zhang, X. Liu, J. Tang, Y. Dong, P. Yao, J. Zhang, X. Gu, Y. Wang, B. Shao, R. Li and K. Wang. OAG: Toward Linking Large-scale Heterogeneous Entity Graphs. KDD'19.

方案 -- LinKG

作者链接模块

期刊/会议链接模块

论文链接模块



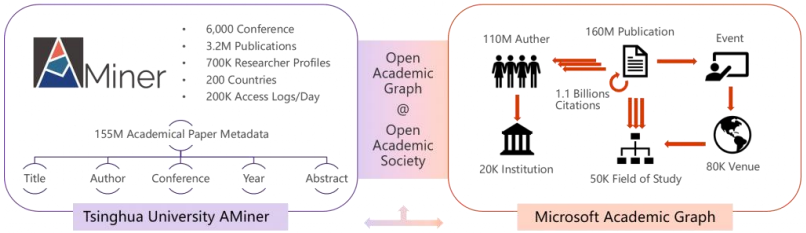
实验效果

Table 1: Results of linking heterogeneous entity graphs. "–" indicates the method does not support the entity linking.

Methods		Keyword	SVM	Dedupe	COSNET	MEgo2Vec	LinKG$_C$	LinKG$_L$	LinKG
Venue	Prec.	80.15	81.69	84.25	–	–	84.67	**91.16**	**91.16**
	Rec.	83.76	83.45	80.92			85.81	**87.58**	**87.58**
	F1	81.91	82.56	82.55			85.23	**89.33**	**89.33**
Paper	Prec.	**99.57**	87.73	99.30	–	–	96.60	89.85	96.70
	Rec.	25.31	85.42	87.09			94.55	89.71	**94.70**
	F1	40.36	86.56	92.80			95.57	89.78	**95.69**
Author	Prec.	44.48	84.70	50.65	91.73	91.03	81.30	84.92	**95.37**
	Rec.	80.63	92.22	85.46	85.33	90.82	84.95	**94.75**	93.48
	F1	57.33	88.30	63.60	88.42	90.92	83.09	89.57	**94.42**
Overall	Prec.	80.22	86.54	82.26	91.73	91.03	91.06	88.19	**96.11**
	Rec.	45.71	87.70	86.38	85.33	90.82	91.05	91.38	**94.11**
	F1	58.23	87.12	84.27	88.42	90.92	91.05	89.76	**95.10**

OAG: Open Academic Graph

https://www.openacademic.ai/oag/



Data set	#Pairs/Venues	Date
Linking relations	29,841	2018.12
AMiner venues	69,397	2018.07
MAG venues	52,678	2018.11

Table 1: statistics of OAG venue data

Data set	#Pairs/Papers	Date
Linking relations	91,137,597	2018.12
AMiner papers	172,209,563	2019.01
MAG papers	208,915,369	2018.11

Table 2: statistics of OAG paper data

Data set	#Pairs/Authors	Date
Linking relations	1,717,680	2019.01
AMiner authors	113,171,945	2018.07
MAG authors	253,144,301	2018.11

Open Academic Graph

Open Academic Graph (OAG) is a large knowledge graph unifying two billion-scale academic graphs: Microsoft Academic Graph (MAG) and AMiner. In mid 2017, we published OAG v1, which contains 166,192,182 papers from MAG and 154,771,162 papers from AMiner (see below) and generated 64,639,608 linking (matching) relations between the two graphs. This time, in OAG v2, author, venue and newer publication data and the corresponding matchings are available.

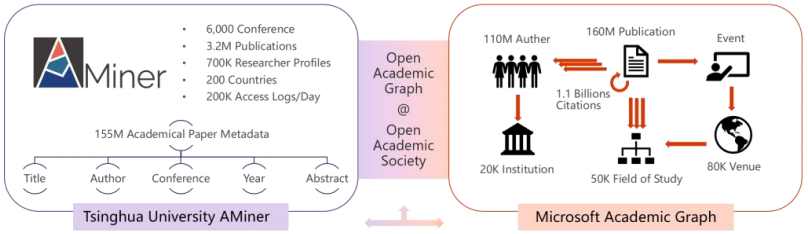
Overview of OAG v2

The statistics of OAG v2 is listed as the three tables below. The two large graphs are both evolving and we take MAG November 2018 snapshot and AMiner July 2018 or January 2019 snapshot for this version.

[1] F. Zhang, X. Liu, J. Tang, Y. Dong, P. Yao, J. Zhang, X. Gu, Y. Wang, B. Shao, R. Li and K. Wang. OAG: Toward Linking Large-scale Heterogeneous Entity Graphs. KDD'19.

应用

- 数据集成
- 图挖掘
  - 合作和引用
- 文本挖掘
  - 标题和摘要
- 科学学研究



Data set	#Pairs/Venues	Date
Linking relations	29,841	2018.12
AMiner venues	69,397	2018.07
MAG venues	52,678	2018.11

Table 1: statistics of OAG venue data

Data set	#Pairs/Papers	Date
Linking relations	91,137,597	2018.12
AMiner papers	172,209,563	2019.01
MAG papers	208,915,369	2018.11

Table 2: statistics of OAG paper data

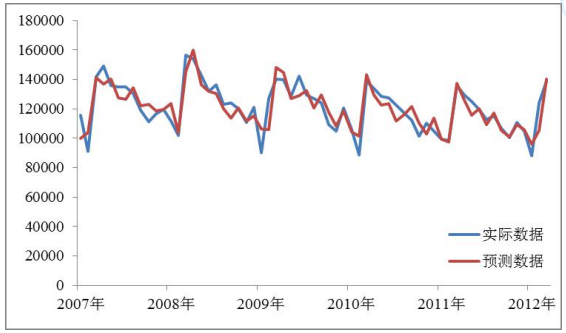
Data set	#Pairs/Authors	Date
Linking relations	1,717,680	2019.01
AMiner authors	113,171,945	2018.07
MAG authors	253,144,301	2018.11

https://www.openacademic.ai/oag/

认知图谱赋能API

科技趋势预测





实时数据采集系统

面向科研人员、期刊、科研管理部门的智能服务

知识获取

科技信息网络

网络媒体

语义

话题分布

关联分析

模式

行为模式

传播模式

预测

关系预测

趋势预测

知识

知识推送

专家智库

AI总线

亿级常识知识图谱

❑国家部委











❑企业合作











❑战略合作

- 清华大学人工智能研究院
- 北京智源研究院
- 中国工程院知识中心

❑国际合作

- John Hopcroft, Professor, 图灵, Cornell University
- Jon Kleinberg, Professor, 院士(美), Cornell University
- Dame Wendy Hall, Professor, University of Southampton
- Jiawei Han, Professor, ACM Fellow, UIUC
- Philip Yu, Professor, ACM Fellow, UIC
- Qiang Yang, Professor, ACM Fellow, HKUST
- Christos Faloutsos, ACM Fellow, CMU
- Michalis Vazirgiannis, Professor, Ecole Polytechnique

谢 谢！