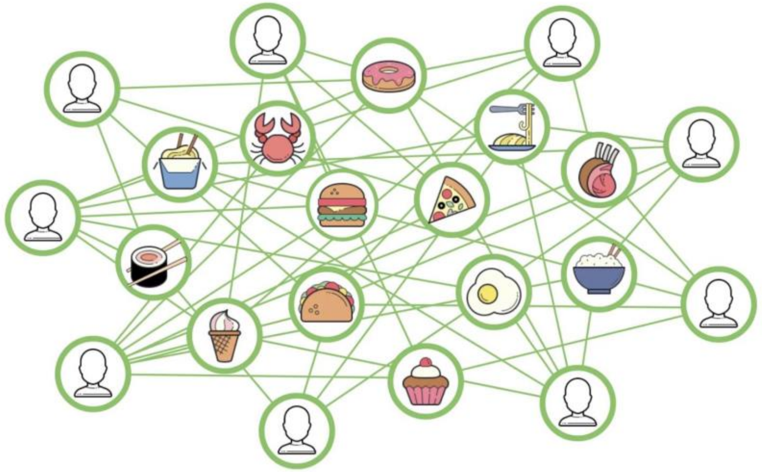
# Graph Learning Methods in Session-based Recommendations and Legal Case Retrieval

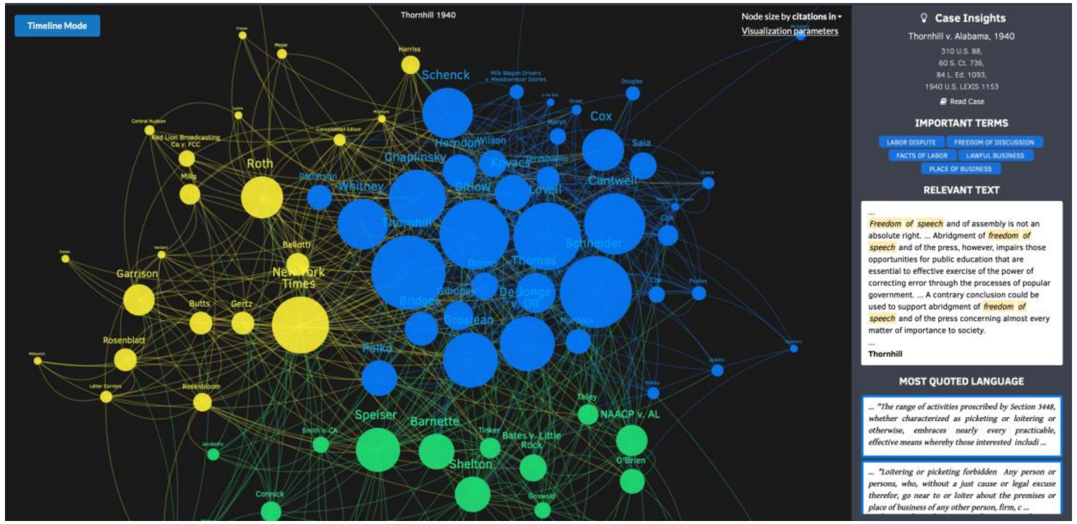
Ruihong Qiu
r.qiu@uq.edu.au
The University of Queensland

# Background

# Structural Information in IR

- Interaction-level structure
  - User-item interaction in recommender systems;
  - legal case reference, etc.

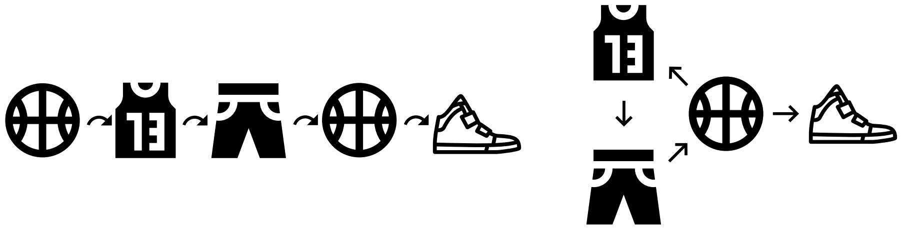




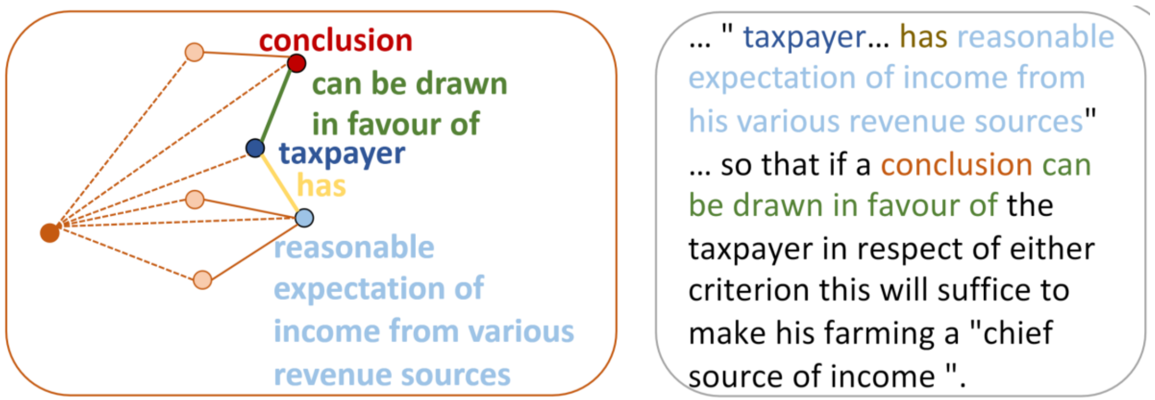

https://kumo.ai/ns-newsarticle-using-graph-learning-for-personalization-how-gnns-solve-inherent-structural-issues-with-recommender-systems
https://medium.com/interacta/visualizing-american-case-law-5bd0243a7097

# Structural Information in IR

- Content-level structure
  - User history; legal case text structure etc.



(a) (b)




Rethinking the item order in session-based recommendation with graph neural networks
CaseGNN: Graph Neural Networks for Legal Case Retrieval with Text-Attributed Graphs

# Challenge

# An Informative Graph?

- The inductive bias in a graph:
  - Node:
    - Interaction-level: user, item, document, case
    - Content-level: item, word, sentence
  - Edge (structure):
    - Interaction-level: interaction, semantics relationship
    - Content-level: time order, knowledge relation

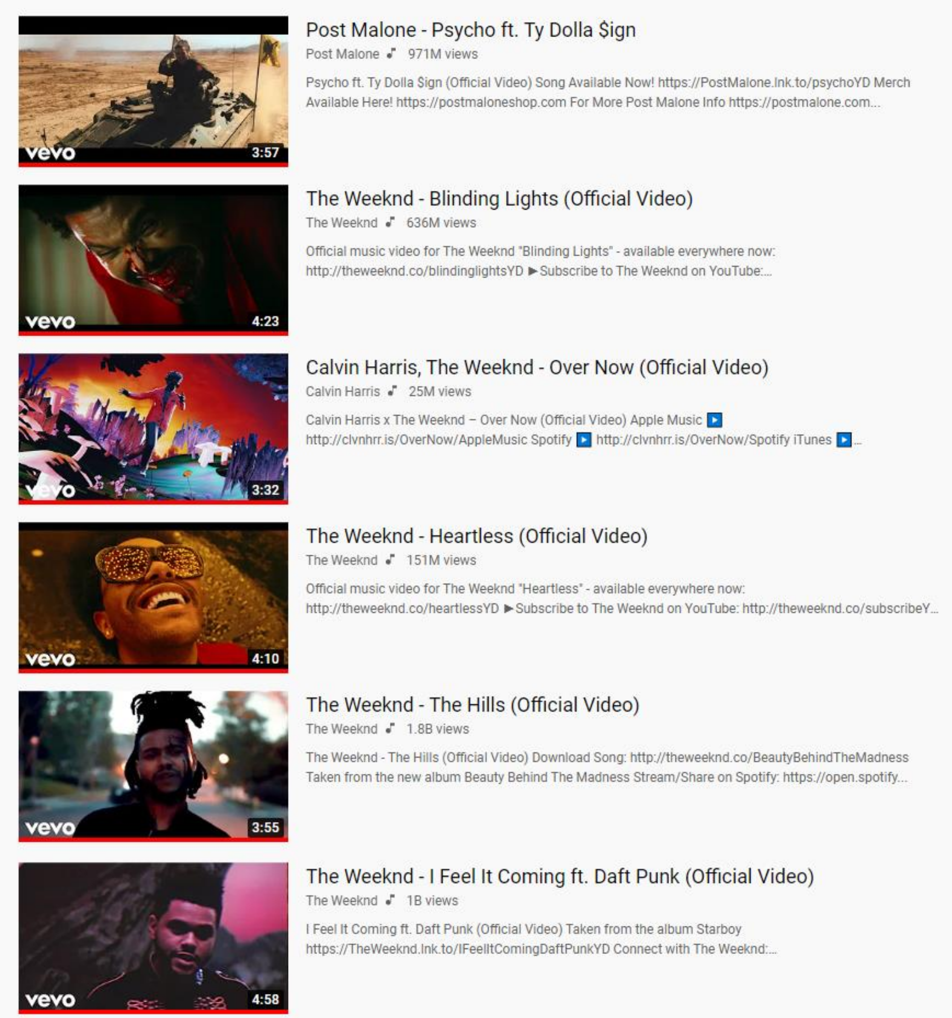
# A Powerful Graph Learning Model?

- A general graph neural networks (GNN):
  - GCN, GAT, GraphSAGE, GIN, SGC, etc.
- Specific design:
  - Suitable for interaction data, item data, text data, etc.

[Kipf et al. 2016, Veličković et al. 2017, Hamilton et al. 2017, Xu et al. 2018, Wu et al. 2019]

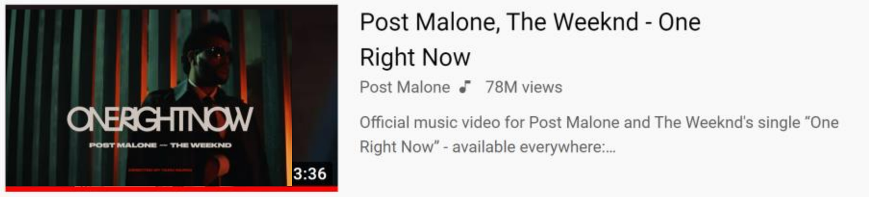
# Example I: Modelling User Session with Graph

# My Own User Session




- From bottom to top indicates from old to new.

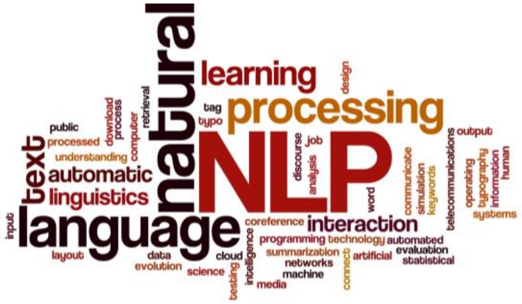
- What's next?

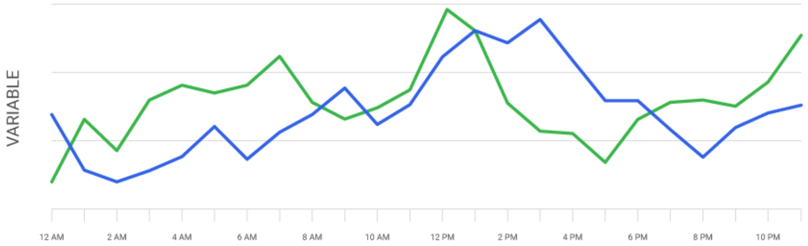



Screenshots from my own account

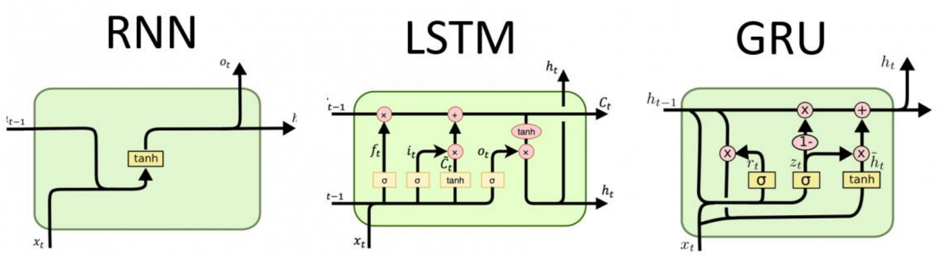
# Modelling General Sequential Patterns

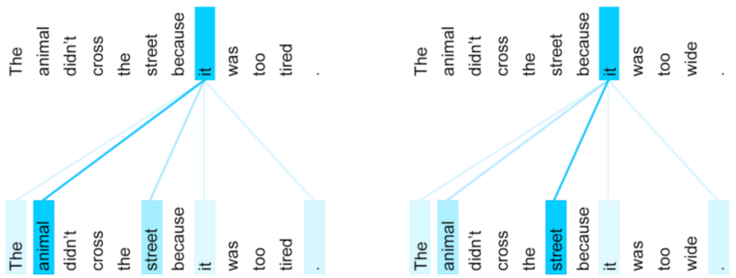
- General sequences: **sentence** (semantic), **time series** (real number, periodic or trended)




- General tools: recurrent model, attention model …

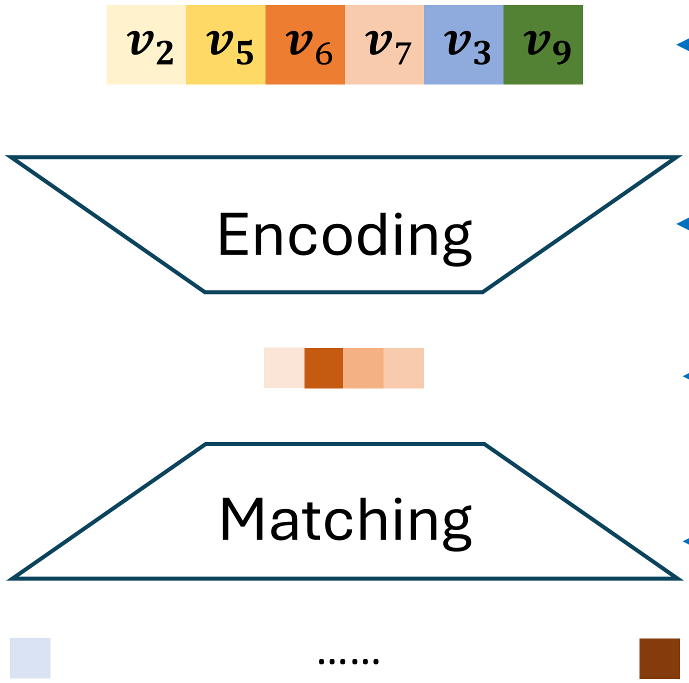



https://medium.com/digital-assistant-academy/the-importance-of-natural-language-understanding-in-voice-technology-a8e603654015
https://www.scylladb.com/glossary/time-series-data/
Development of PPTNet a Neural Network for the Rapid Prototyping of Pulsed Plasma Thrusters
https://www.tensorflow.org/text/tutorials/transformer

# Direct Modelling of User Sessions

- Directly using **existing** models: GRU4Rec, SASRec, BERT4Rec …




RNN, GRU, LSTM, Attention …

Session representations

Ranking prediction over all items, dot product …

[Hidasi et al. 2015, Kang et al. 2018, Sun et al. 2019] 

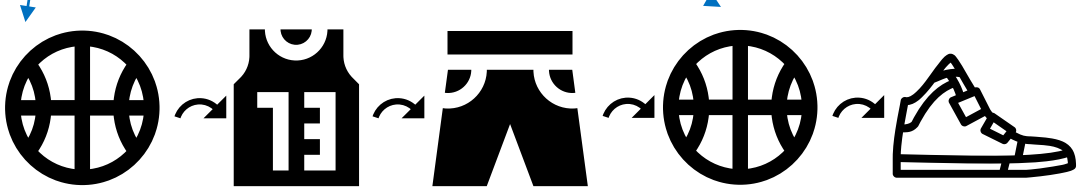
# Issues in Modelling User Sessions

- The user session (shopping or watching history) is **different** from general sequences.
  - No **grammar semantics**;
  - No **periodic feature**;
  - No **continuous trending** etc.

- Need to **identify** the **difference** and **develop proper models** to learn sequential patterns in user sessions.

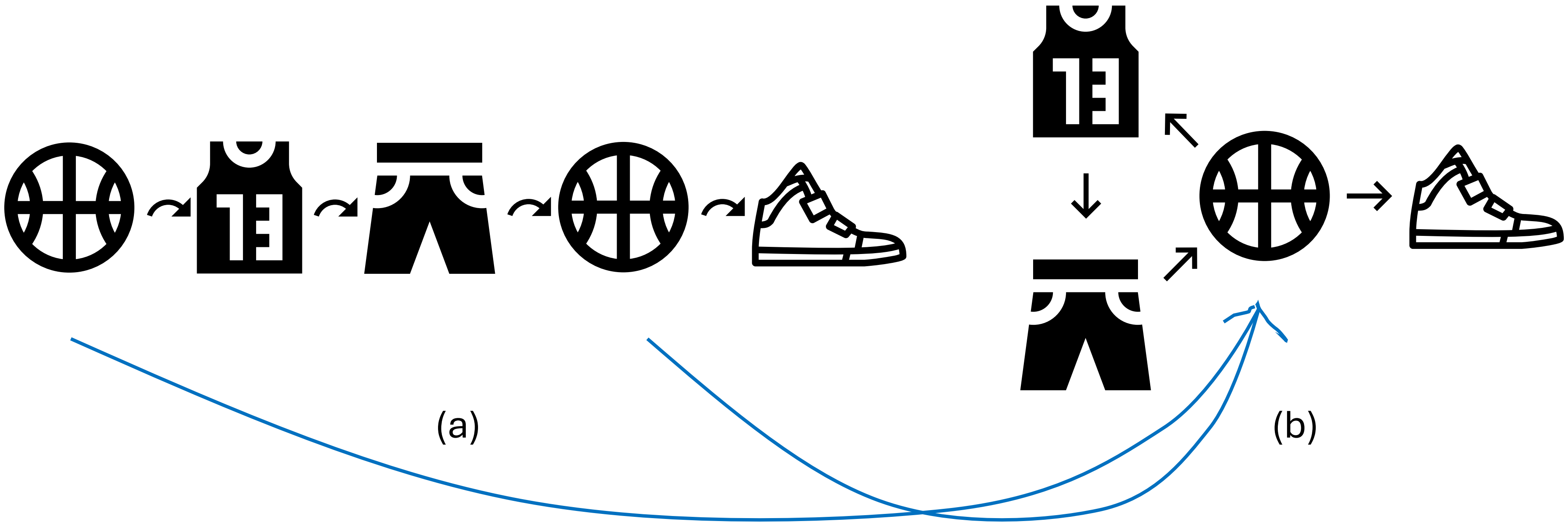
# Structural Information

- A sequence of anonymous user history within a short time period.
- Items have **reappearance** in the sequence, such as **re**-click products; **re**-listen to songs; **re**-watch videos.
- They are **PIVOTAL**.



# Converting to Graph

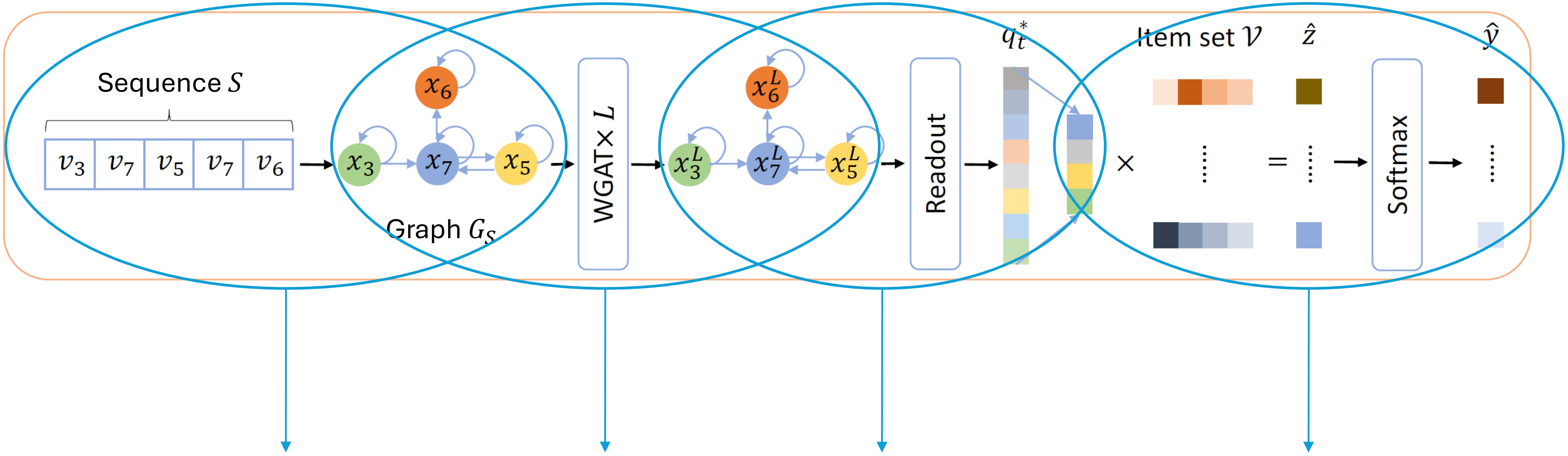
- Not like 1D sequence, **graph** has the topological structure with **nodes** and **edges**.
- **Convert** a sequence into a graph.
- **Pivotal** items become **pivots**.



(a)

(b)

# FGNN Model




Graph construction

Weighted Graph Attention Layer

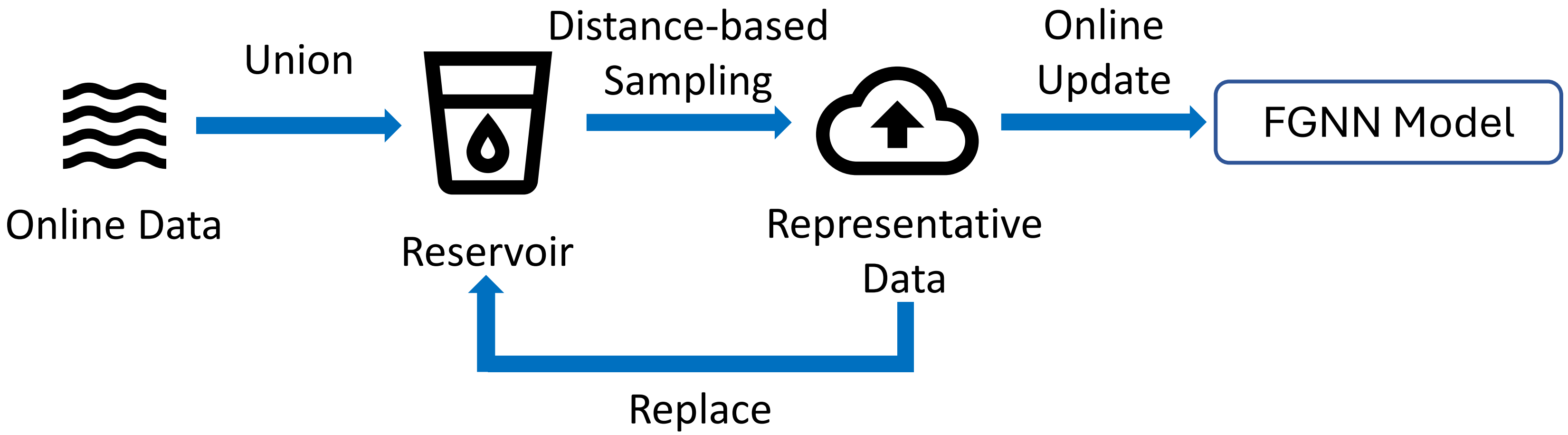
Set2Set readout

Matching

[Qiu et al. CIKM 2019]

# Streaming Scenario: GAG Model

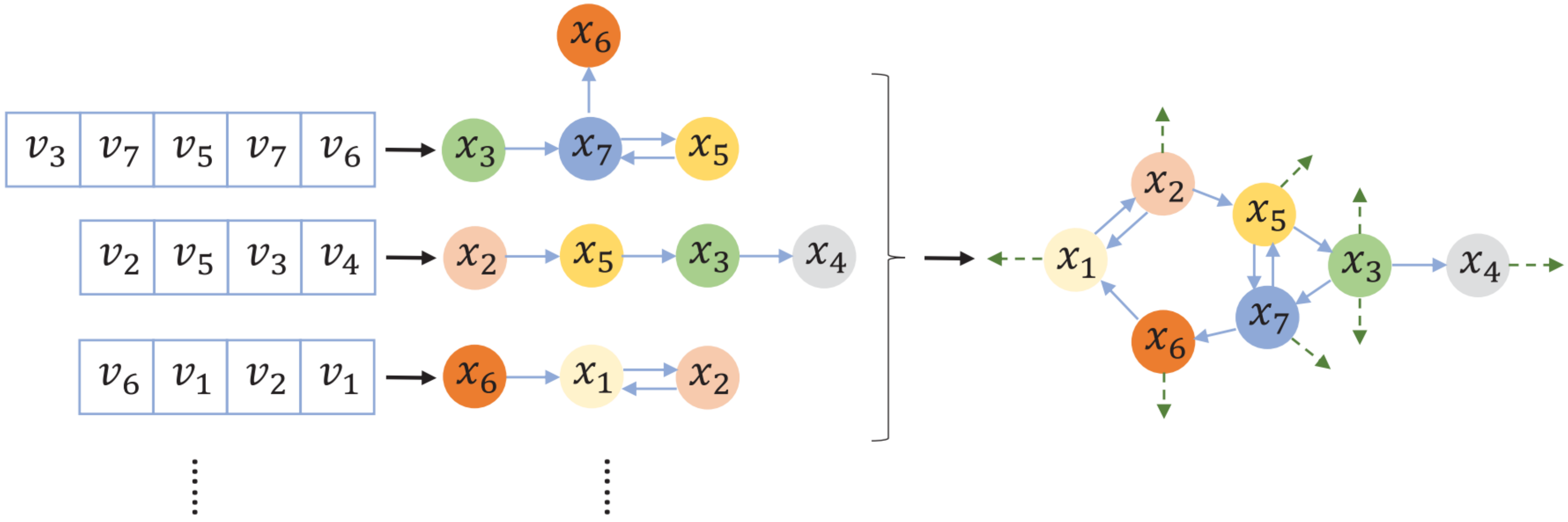
- Model trained on offline data
- Needs to update with online data




[Qiu et al. SIGIR 2020, Liu, Qiu et al. ICDM 2023]

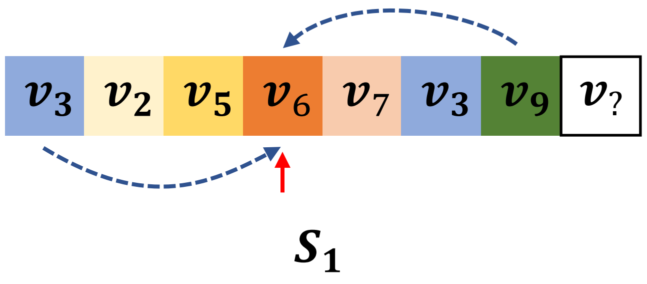
# Cross-sequence Scenario

- Multiple sequence could contain same items
- Link multiple sequence together



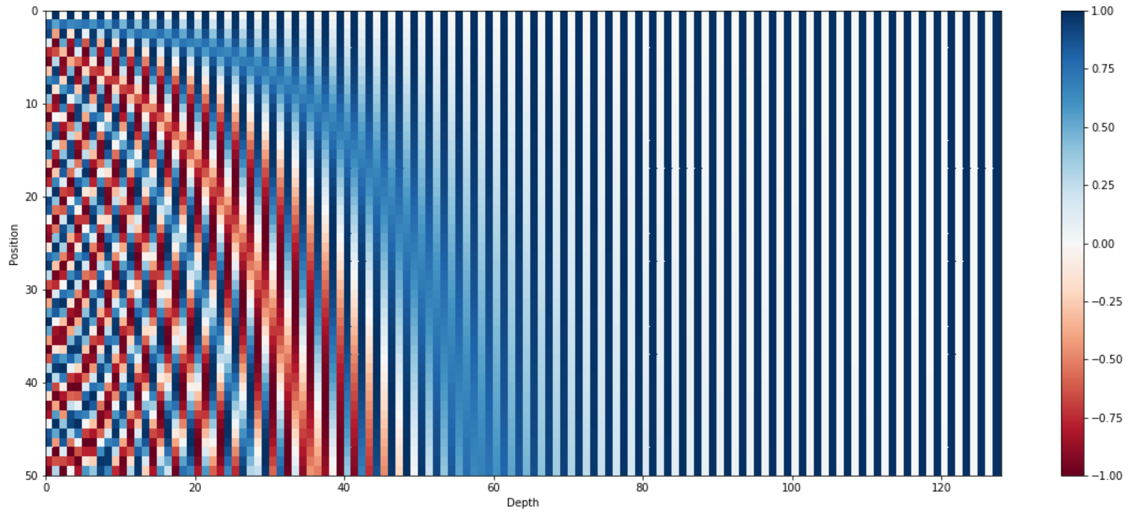
# Positional Information

- An interaction at **different positions** will carry **different meanings** for user preference.
- Within a sequence, an **early** interaction would indicate the **initial** intention; A **later** interaction would indicate the **latest** intention.



# Position in General Sequence

- Attention mechanism has a **positional encoding** (PE) (counting from left to right).
- Based on **sin/cos** function. Each position corresponds to one row.

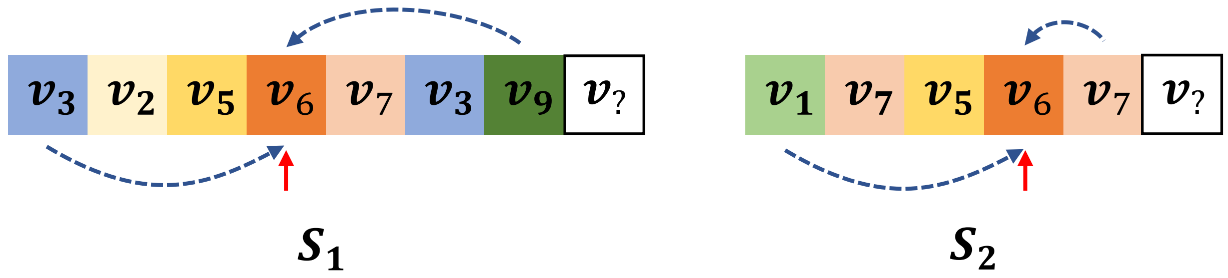



https://kazemnejad.com/blog/transformer_architecture_positional_encoding/

[Vaswani et al. 2017]

# Problems of PE In Interaction Sequence

- Original PE only tells **how far** away from the **beginning** (forward-awareness).
- **Problematic** example in following sessions for $v_6$:




- Need to know how far away from **both** the **beginning** (forward-awareness) and the **ending** (backward-awareness).

# Dual Positional Encoding

- PE for interaction sequence needs to be both **forward-aware** and **backward-aware**
- Dual Positional Encoding (**PDE**):

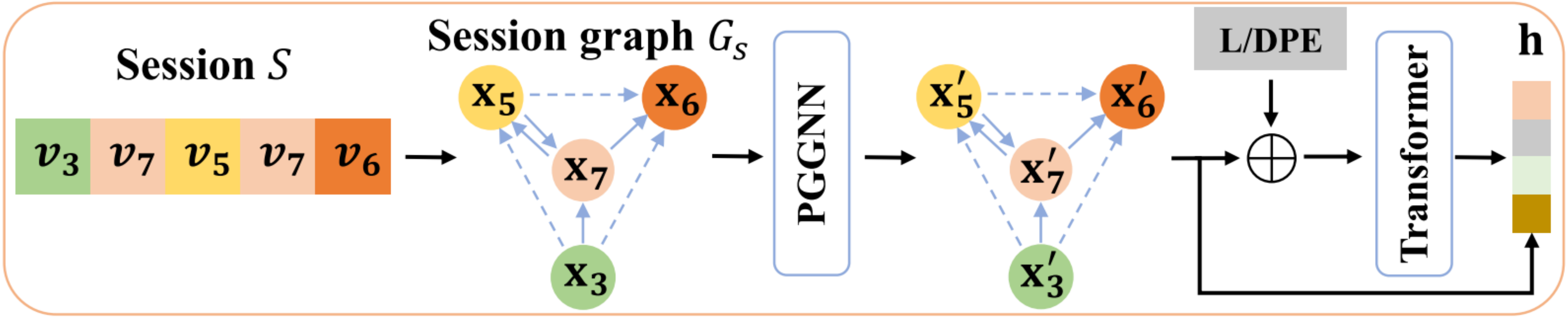
$$
\begin{aligned}
\boldsymbol{p}^{l}_{pos,2i} &= \sin(pos/f(i)),\\
\boldsymbol{p}^{l}_{pos,2i+1} &= \cos(pos/f(i)),\\
\boldsymbol{p}^{l}_{pos,2i+d/2} &= \sin((l-pos-1)/f(i)),\\
\boldsymbol{p}^{l}_{pos,2i+1+d/2} &= \cos((l-pos-1)/f(i)),
\end{aligned}
$$

**Theorem 4.3.1.** *Dual positional encoding can represent the positional information of SBRS because it is both forward-aware and backward-aware.*

- Also a Learnable Dual Positional Encoding (**LDPE**)

[Qiu et al. TOIS 2021]

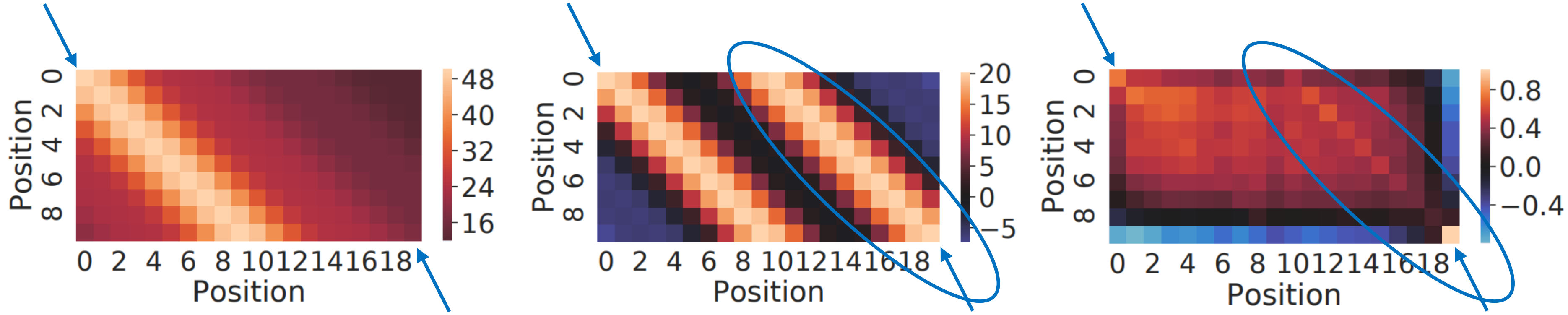
# PosRec Model




[Qiu et al. TOIS 2021] 

# Visualisation of PEs

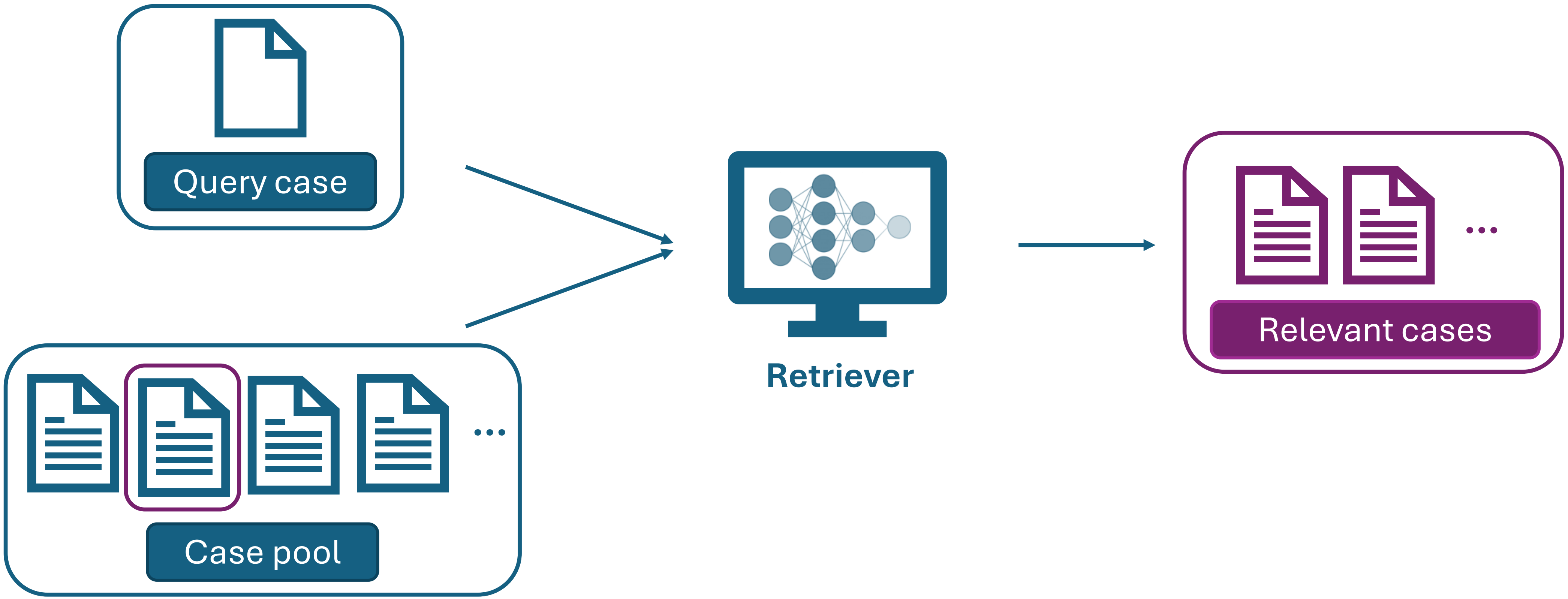
- One sequence has 10 items, and the other one has 20.
- **Forward-awareness** for the beginning position.
- Only Dual and Learnable Dual PE have **backward-awareness**.
- Learnable one can identify **different levels of importance**.




(a) General (b) Dual (c) Learnable Dual

# Example II: Legal Case Retrieval with Graph

# Legal Case Retrieval Workflow



# Legal Cases

**Lafond v. Muskeg Lake Cree Nation** (2008), 330 F.T.R. 60 (FC) → **Parties**: plaintiff & defendant

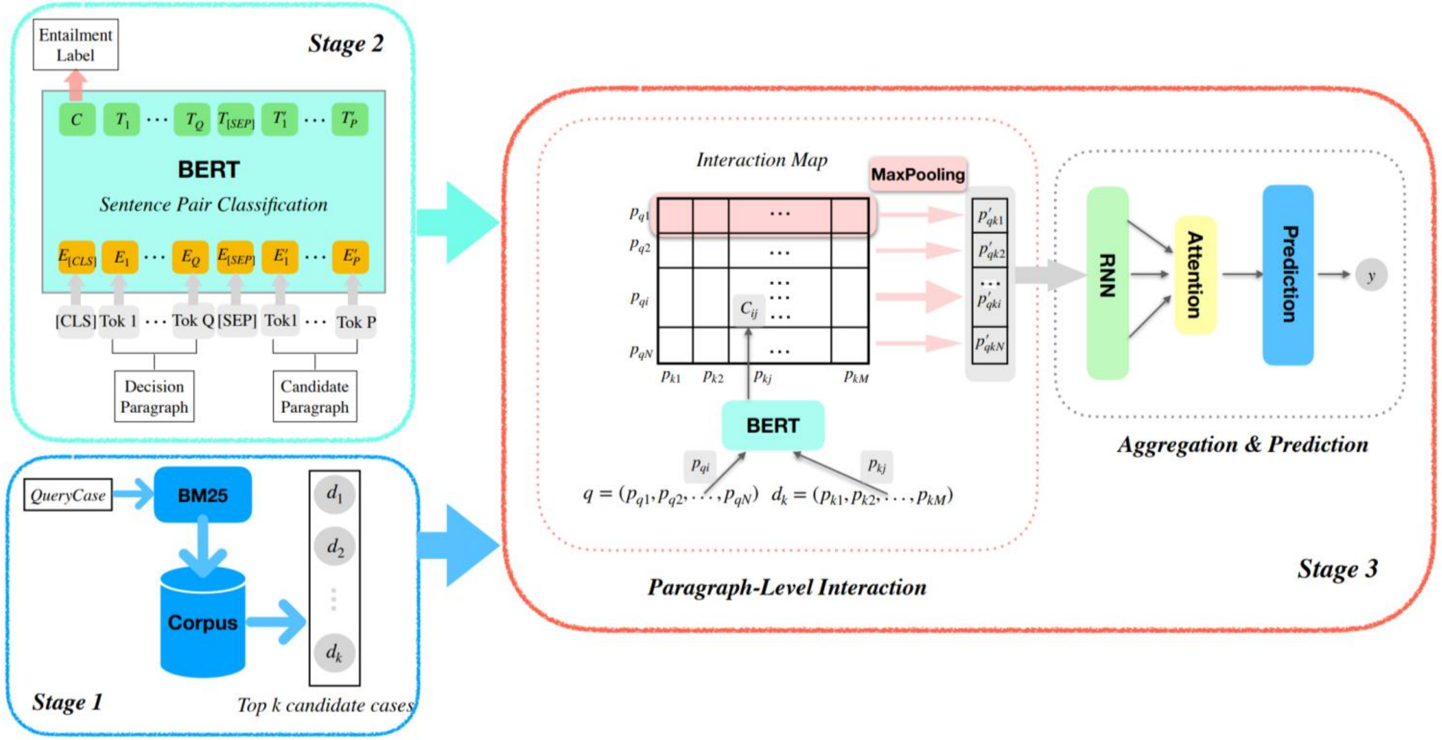
**Summary**:
Lafond was elected as a councillor to the Muskeg Lake Cree Nation Band Council. After … → **Case Summaries**

In the recent decision of **FRAGMENT_SUPPRESSED**, the Federal Court of Appeal: … → **Citation**

For these reasons, the application for judicial review of Chief Ledoux's **decision** will be allowed. → **Judgment**

# Related Work in Legal Case Retrieval

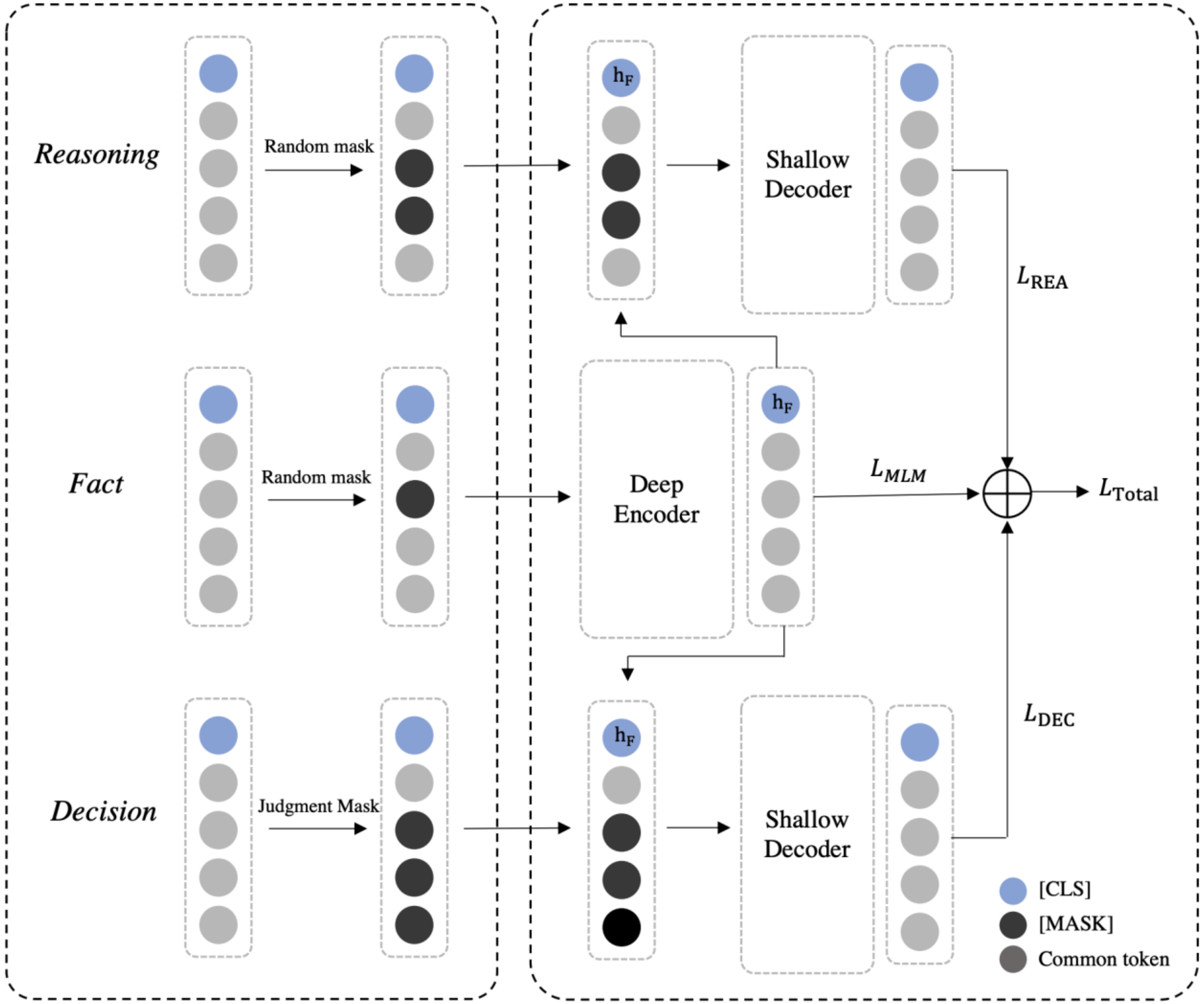
- Bert-based model
- BERT-PLI
- Encode paragraphs with BERT
- Paragraph-level interaction




[Shao et al. 2020] 

# Related Work in Legal Case Retrieval

- Bert-based model
- SAILER
- Generative pretraining




[Li et al. 2023]

# Related Methods' Characteristics

- **Pros**
  - Better accuracy with semantics by **legal corpus pre-training**
  - Dividing case text for **lengthy** problem

- **Cons**
  - Case text dividing -> loss of legal **context** information & case **global** view

# Challenges

- Legal structural information:
  - **High-order interactions** of elements in a case: parties, crime activities and evidences
- Lengthy legal text limitation:

| Datasets | LeCaRD | COLIEE2023 |
|---|---|---|
| Language | Chinese | English |
| Avg. length/case | 8,275 | 5,566 |
| Largest length of cases | 99,163 | 61,965 |
| Avg. relevant cases/query | 10.33 | 2.69 |

# Solution

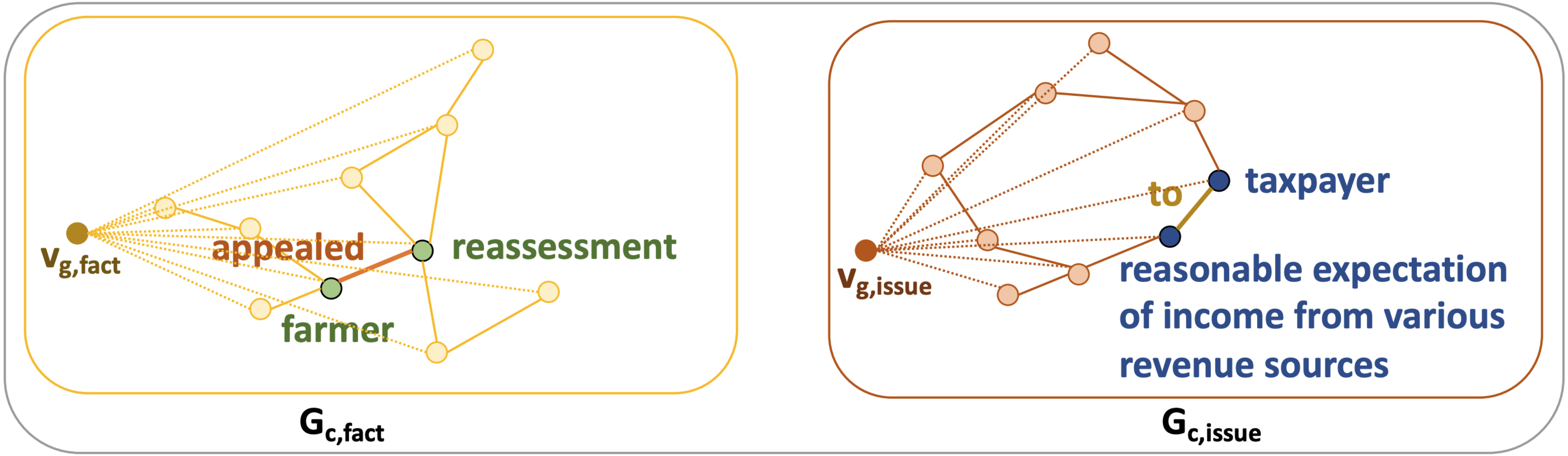
- Graph data is an effective data structure to incorporate the **abundant structural information in legal cases**.

- **Transform** a legal **case** into a Text-Attributed Case **Graph** (TACG).

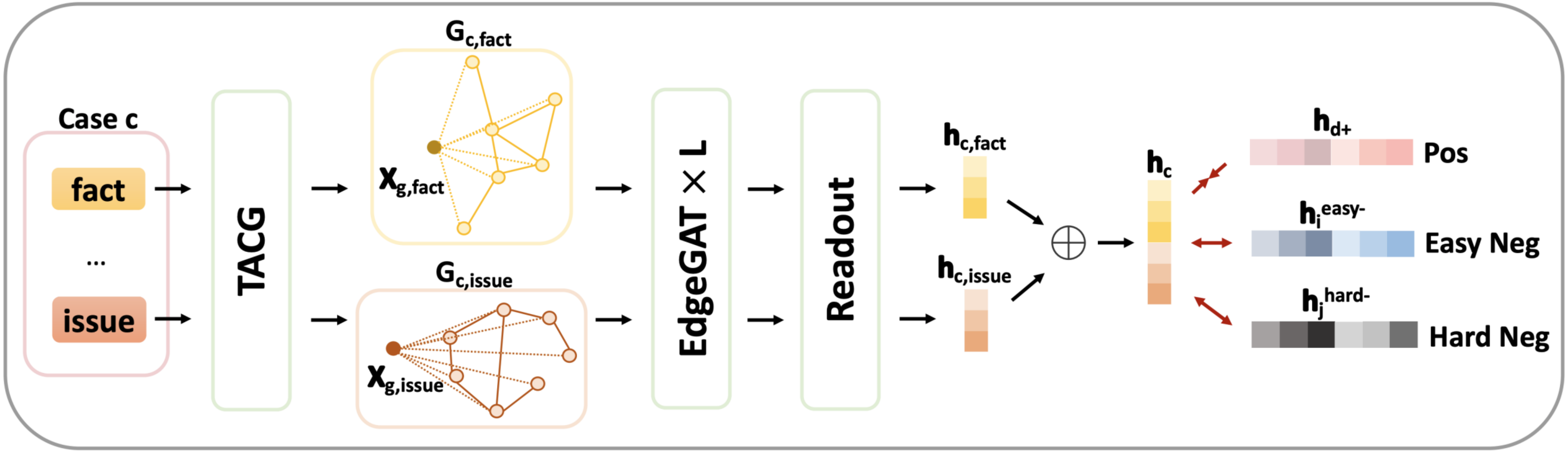
- An Edge Graph Attention Layer (EdgeGAT) and a readout function are proposed to obtain a graph level case representation.

# TACG

- Extract the **entities** and the **relations** using Information Extraction
- Separate legal **fact** and legal **issue**
- Create text-attributed **case graph**, with a **virtual** node



# CaseGNN Framework




[Tang, Qiu et al. ECIR 2024] 

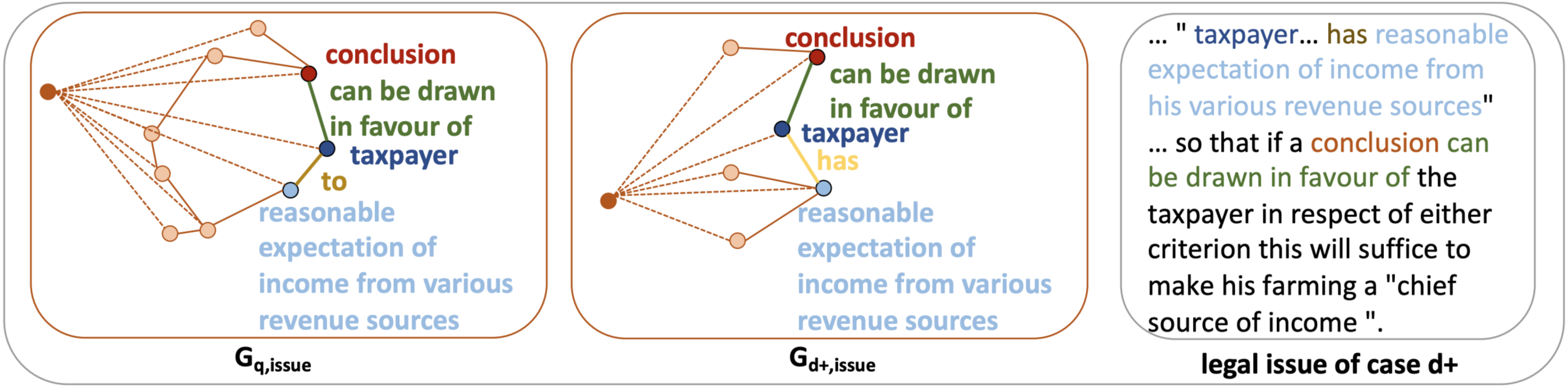
# Overall Performance

| Methods | COLIEE2022 | | | | | | | COLIEE2023 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 | P@5 | R@5 | Mi-F1 | Ma-F1 | MRR@5 | MAP | NDCG@5 |
| **One-stage** | | | | | | | | | | | | | | |
| BM25 | 17.9 | 21.2 | 19.4 | 21.4 | 23.6 | 25.4 | 33.6 | 16.5 | 30.6 | 21.4 | 22.2 | 23.1 | 20.4 | 23.7 |
| LEGAL-BERT | 4.47 | 5.30 | 4.85 | 5.38 | 7.42 | 7.47 | 10.9 | 4.64 | 8.61 | 6.03 | 6.03 | 11.4 | 11.3 | 13.6 |
| MonoT5 | 0.71 | 0.65 | 0.60 | 0.79 | 1.39 | 1.41 | 1.73 | 0.38 | 0.70 | 0.49 | 0.47 | 1.17 | 1.33 | 0.61 |
| SAILER | 16.6 | 15.2 | 14.0 | 16.8 | 17.2 | 18.5 | 25.1 | 12.8 | 23.7 | 16.6 | 17.0 | 25.9 | 25.3 | 29.3 |
| PromptCase | 17.1 | 20.3 | 18.5 | 20.5 | 35.1 | 33.9 | 38.7 | 16.0 | 29.7 | 20.8 | 21.5 | 32.7 | 32.0 | 36.2 |
| CaseGNN (Ours) | **35.5**±0.2 | **42.1**±0.2 | **38.4**±0.3 | **42.4**±0.1 | **66.8**±0.8 | **64.4**±0.9 | **69.3**±0.8 | **17.7**±0.7 | **32.8**±0.7 | **23.0**±0.5 | **23.6**±0.5 | **38.9**±1.1 | **37.7**±0.8 | **42.8**±0.7 |
| **Two-stage** | | | | | | | | | | | | | | |
| SAILER | **23.8** | 25.7 | 24.7 | 25.2 | 43.9 | 42.7 | 48.4 | 19.6 | 32.6 | 24.5 | 23.5 | 37.3 | 36.1 | 40.8 |
| PromptCase | 23.5 | 25.3 | 24.4 | **30.3** | 41.2 | 39.6 | 45.1 | **21.8** | 36.3 | **27.2** | 26.5 | 39.9 | 38.7 | 44.0 |
| CaseGNN (Ours) | 22.9±0.1 | **27.2**±0.1 | **24.9**±0.1 | 27.0±0.1 | **54.9**±0.4 | **54.0**±0.5 | **57.3**±0.6 | 20.2±0.2 | **37.6**±0.5 | 26.3±0.3 | **27.3**±0.2 | **45.8**±0.9 | **44.4**±0.8 | **49.6**±0.8 |

- CaseGNN is better in most situations.

# CaseGNN Case Study

- Successful retrieval by CaseGNN but not by PromptCase.




- Original **text**: entities and relationships are **far** from each other. Language models are **not good at long dependency**.
- **TACG**: brings multiple **entities together**.

# Takeaway and Conclusions

# Takeaway and Conclusions

- There are **abundant structural information** in different IR scenarios
- The **graph construction determines** how good a graph-based method can be
- The graph learning module is effective. But a **dedicated design** can further improve performance.

Thank You!

Ruihong Qiu
r.qiu@uq.edu.au