AWS Life Science Hackathon

10/1 – 10/3

Team 9:
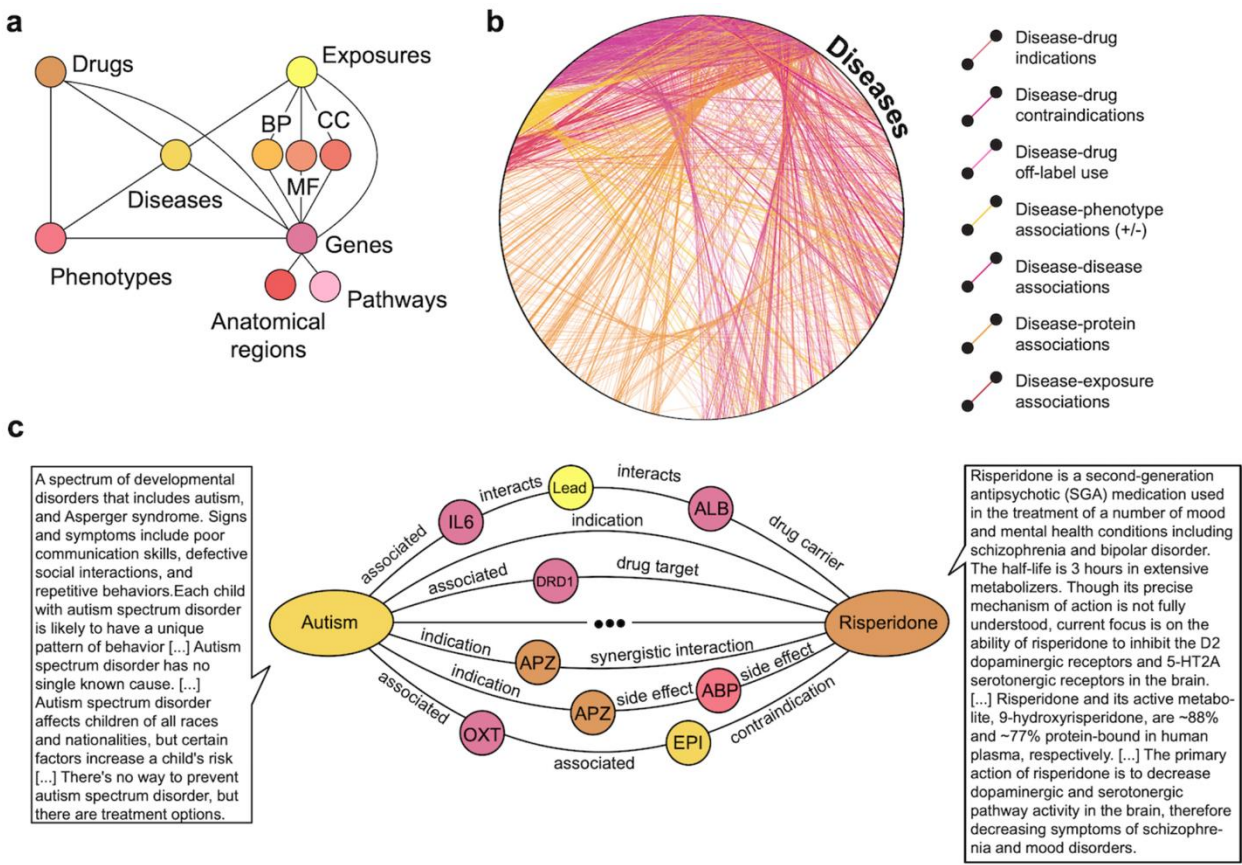
Aymen Maqsood Mulbagal

Ben Stear

JP Courneya

Mina P. Peyton
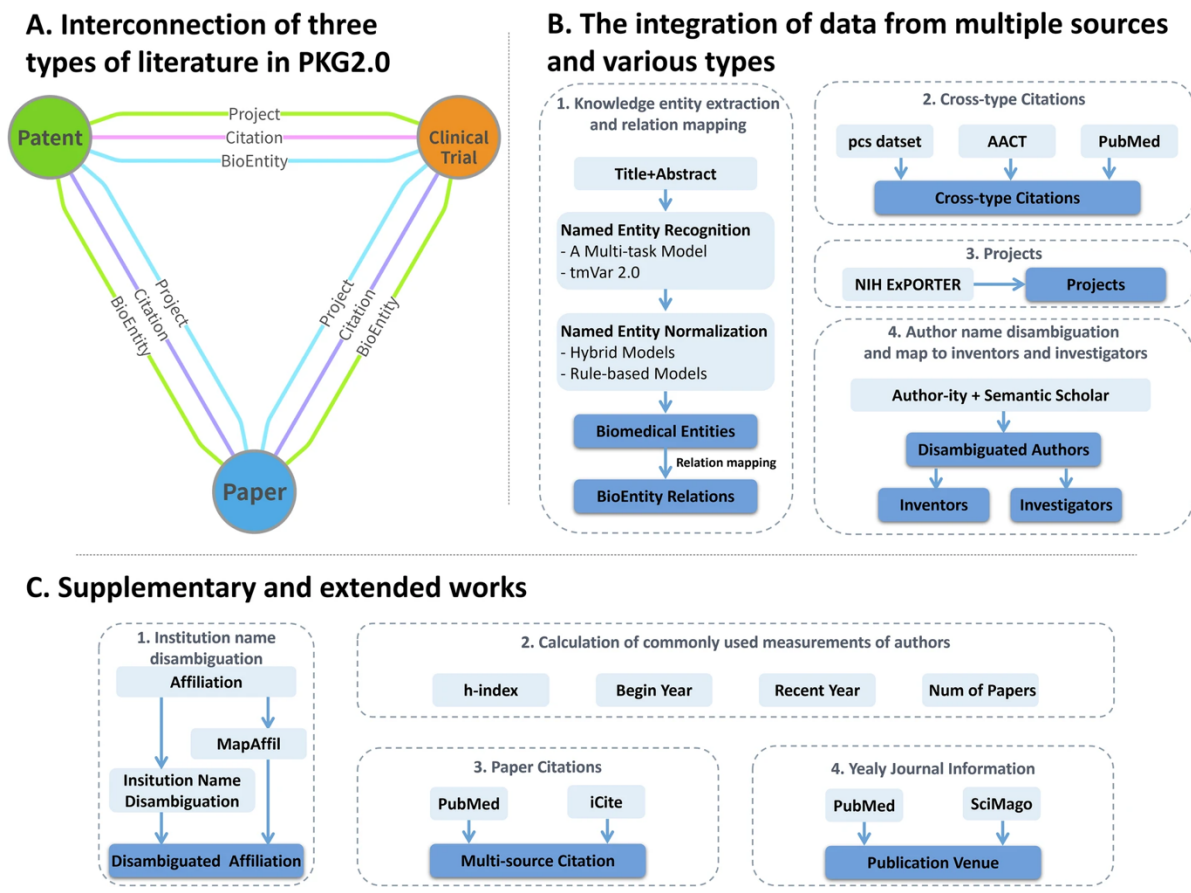
Victor Felix

Arun Bondali

Michelle Holko

Precision Medicine Knowledge Graph (PrimeKG)

PrimeKG:

- integrates 20 resources:
  - 17,080 diseases with 4,050,249 relationships representing ten major biological scales
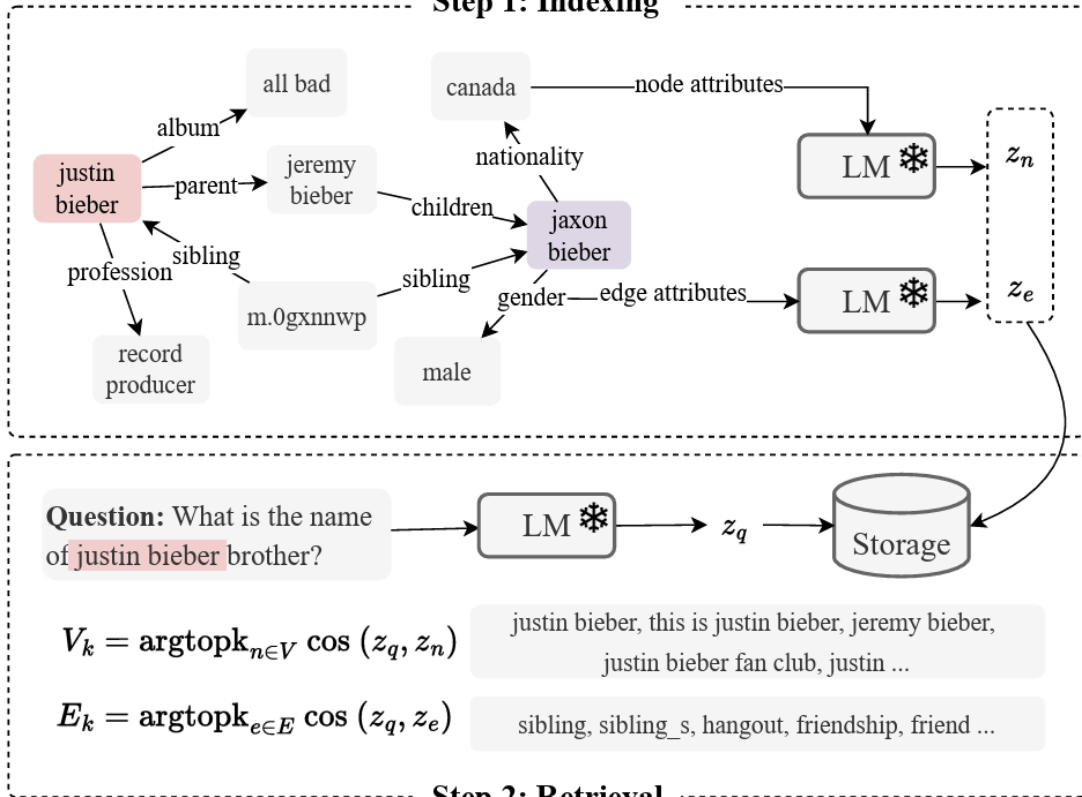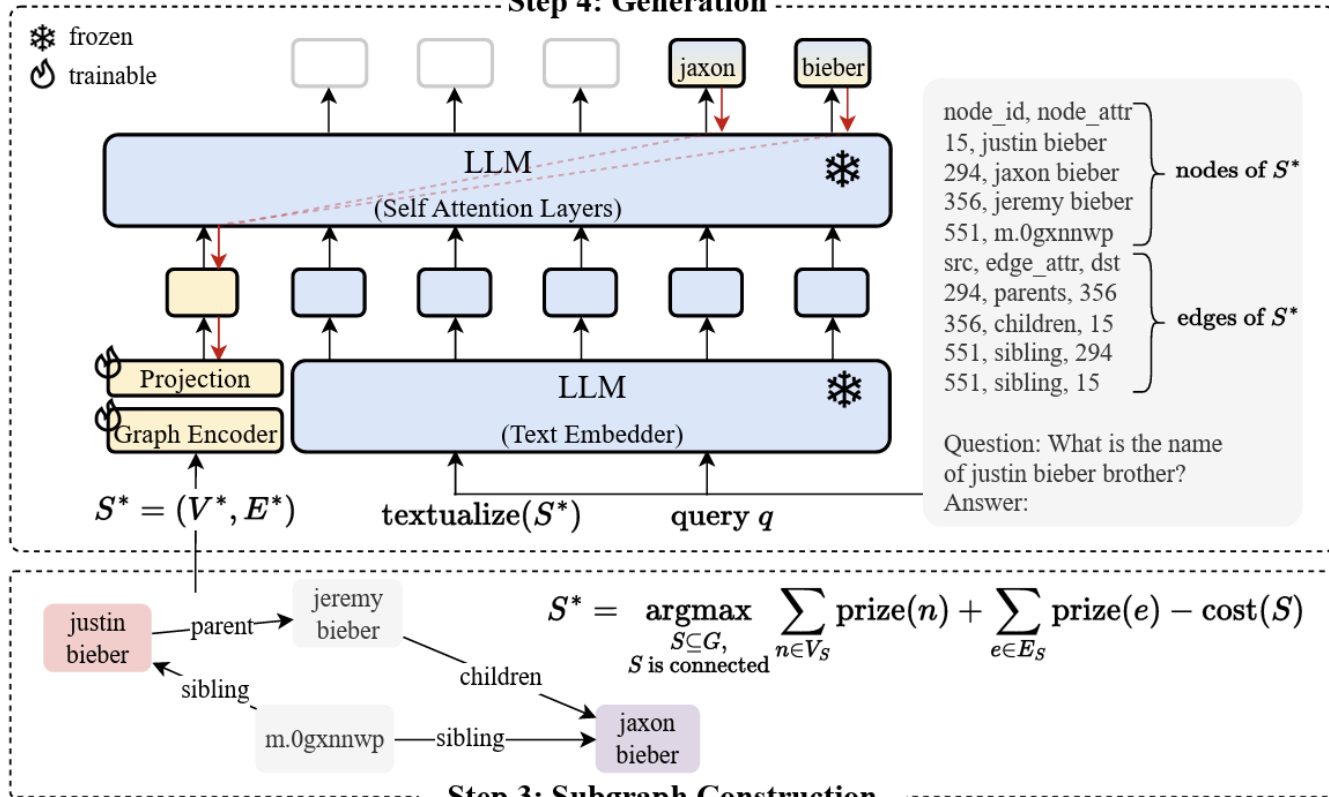
Pubmed Knowledge Graph (PubKG)

PubKG:

- integrates papers, patents, and clinical trials as core data
  - connections between papers, patents, and clinical trials through bioentities, projects, authors, and citations

Chandak et. al., Scientific Data, 2023; Xu et. al., Scientific Data, 2025

G-Retriever

Step 1: Indexing

Step 2: Retrieval

$$V_k = \text{argtopk}_{n \in V} \cos(z_q, z_n)$$

justin bieber, this is justin bieber, jeremy bieber, justin bieber fan club, justin ...

$$E_k = \text{argtopk}_{e \in E} \cos(z_q, z_e)$$

sibling, sibling_s, hangout, friendship, friend ...

Step 3: Subgraph Construction

$$S^* = \underset{\substack{S \subseteq G, \\ S \text{ is connected}}}{\text{argmax}} \sum_{n \in V_S} \text{prize}(n) + \sum_{e \in E_S} \text{prize}(e) - \text{cost}(S)$$

Step 4: Generation

node_id, node_attr
15, justin bieber
294, jaxon bieber   } nodes of S^*
356, jeremy bieber
551, m.0gxnnwp

src, edge_attr, dst
294, parents, 356
356, children, 15   } edges of S^*
551, sibling, 294
551, sibling, 15

Question: What is the name of justin bieber brother? Answer:

$S^* = (V^*, E^*)$ textualize(S^*)   query q

He et. al., arXiv, 2024

GraphRAG (GNN + LLM)

GNN+LLM Graph RAG (GNN Feeds LLM)

Given a query at inference time...

Knowledge Graph (KG)

Retrieve Subgraph Context

Retrieved Node/Edge Features

GNN Encoder

Natural Language Query

Textified Graph

Concatenate Embeddings

LLM Decoder

Natural Language Response

LLM Encoder
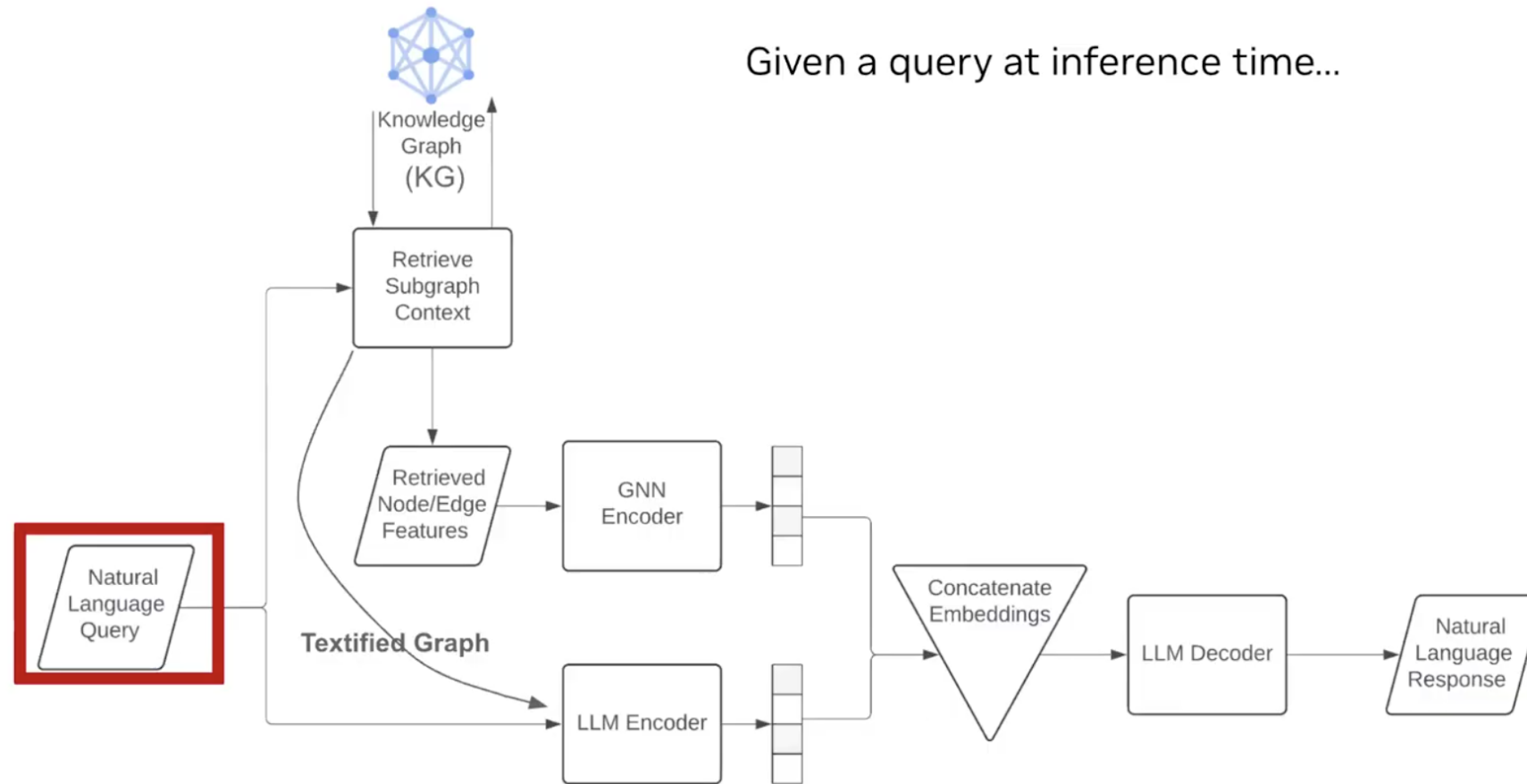
*RAG = Retrieval Augmented Generation

*G-retriever: https://arxiv.org/abs/2402.07630

Workflow Overview