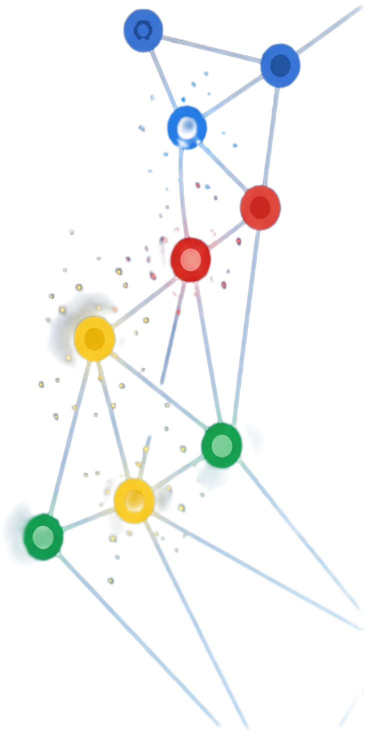


Graph Neural Network at Google

Mathieu Guillame-Bert

Fast Machine Learning for Science, 2025


Always have been
Wait, those are GNNs?
TRANSFORMER
CNN
imgflip.com

A word of introduction

Mathieu
Guillame-Bert

Michael
Galkin

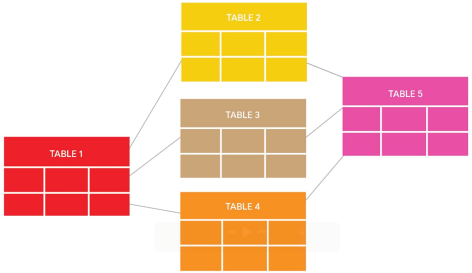
Agenda

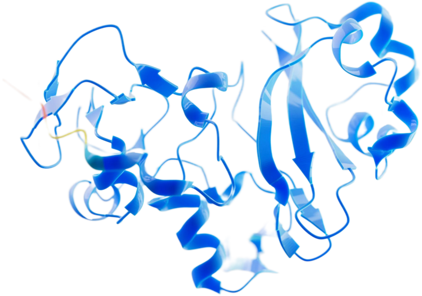
1. What are GNNs?
2. Examples of GNN Applications
3. Scaling GNNs
4. Future of GNNs

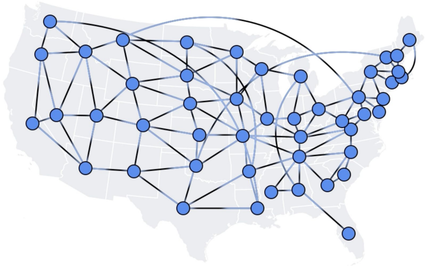
What are GNNs?

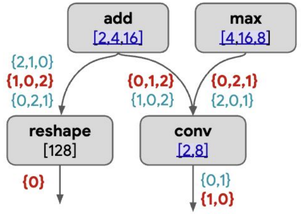
Relational Data: The Native Language of the Real World

- Real-world data is often inherently relational.
- Examples
  - **Transaction logs:** User interactions with products, websites, videos, topics.
  - **Protein study:** How different atoms, proteins, or amino acids behave together.
  - **Transportation logistic:** Spatio-temporal descriptions for routing and supply chains.
  - **Monitoring:** How different parts of a system (e.g., machines in a data center) behave and interact.







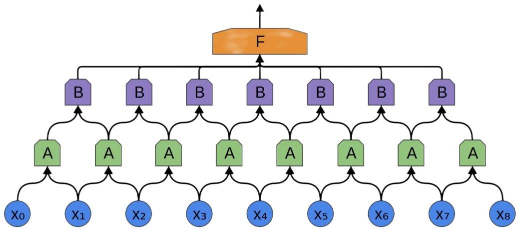


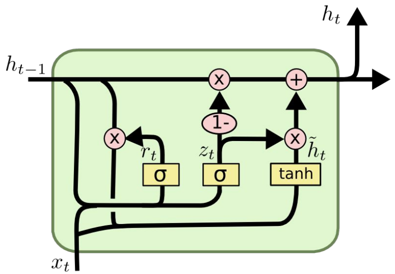
Fitting relational data into ML models (1/2)

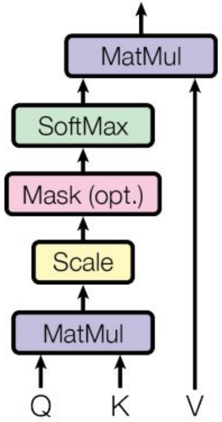
- Early ML techniques focused on tabular data i.e. **fixed-size**, independent data points **without relations** or **invariants**.
  - e.g. Linear model, decision tree, svm, (A)NN, kNN, Naive Baye.
- Data type commonly with invariants: Image, text, time-series, video, graphs, first order logic program.
- Not taking invariants in account gives poor results
  - e.g. The position of the cat does not matter in an image cat detector.
- How to consume data with variable size / invariants / relations?
  - Hand-crafted transformation to tabular data
  - Augment / duplicate dataset with invariant transformations
  - Enforce invariants pattern in learning algorithm / model e.g., a decision tree with pattern matching splits, logic programming, **NN architecture**.

Fitting relational data into ML models (2/2)

- **Neural network architecture design**: Structure a NN to embed useful **biases** or **invariances** so the model.
- **Example**
  - **Convolution Neural Network (CNN):** Image, time-series, text
  - **Long Short-Term Memory (LSTM):** Text, time-series
  - **Graph Neural Networks (GNN):** Graph
  - **Transformers:** Text, image, graph, sets, <everything>



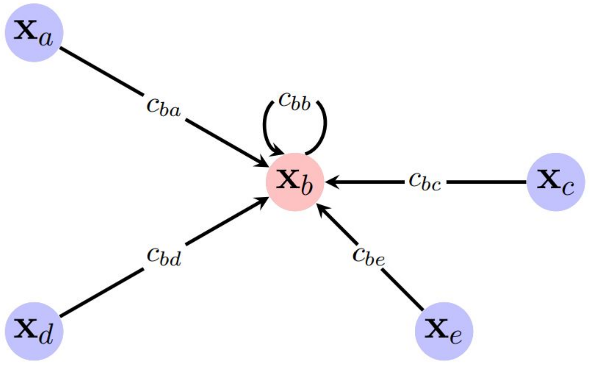




- [A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects]
- [Colah's blog]

GNN core idea

- The input data is a **graph** (nodes & edges).
- Nodes have a state **xi**; initialized as **input features**.
- Nodes **send** messages to their neighbors.
- Node **update** their state by applying a FF-NN on their previous state and received messages.
- After a few sends/updates, the **node state** is pulled and used as a fixed-size embedding.
    - e.g. for supervised learning, distance learning, generation, anomaly detection.



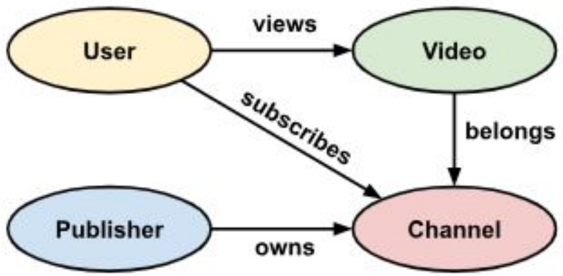
The node Xb's state is updated using its previous state and its neighbor states.

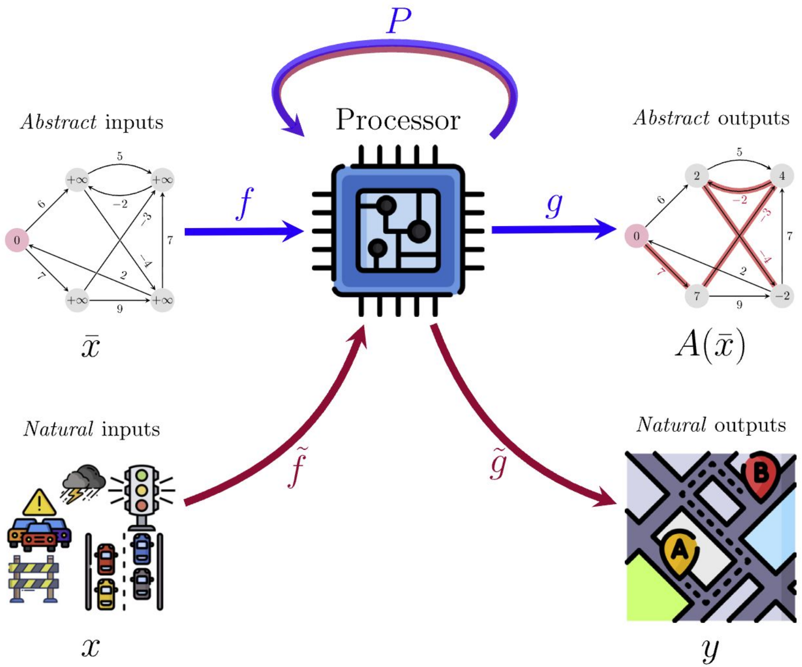
$$\mathbf{h}_u = \phi\left(\mathbf{x}_u, \bigoplus_{v \in \mathcal{N}_u} \psi(\mathbf{x}_u, \mathbf{x}_v)\right)$$

The classical message passing equation. Φ and ψ are neural networks.

The different flavors of GNNs: Heterogeneous GNN (1/3)

- Graphs with multiple node types and edge types
  - Make it easy to feed natively rich real world relational data.
  - Especially popular with recommender systems and fraud detection.
- Encoder - Processor - Decoder as a typical GNN architecture

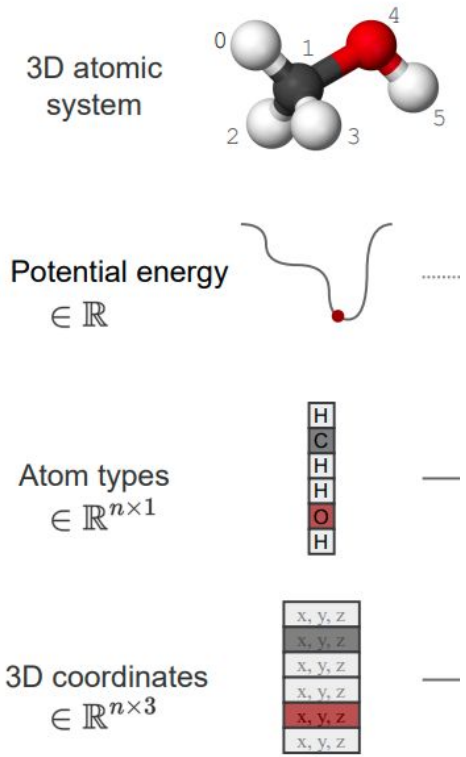


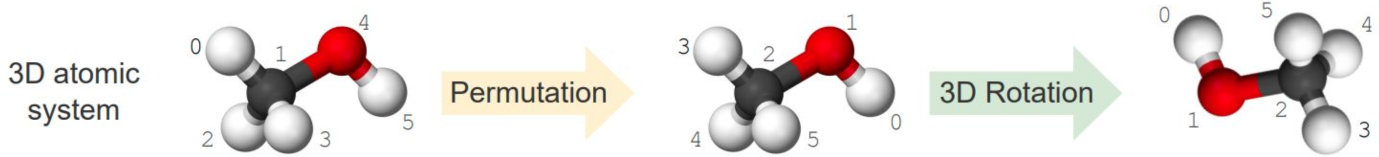


Veličković, Blundell. Neural Algorithmic Reasoning. Patterns 2021

The different flavors of GNNs: Geometric GNNs (2/3)

- The graphs represent 3D structures
  - e.g., Molecules, crystals, proteins, point clouds, meshes.
- Node features: XYZ coordinates, etc.
- Edge features: Distance, orientation, etc.
- GNNs have to capture **invariances** and **equivariances** after common transformations (translation, rotation, etc)
  - Energy of a system is invariant to rotations - model via atom distances
  - Forces are equivariant to rotations - model via spherical harmonics and tensor products

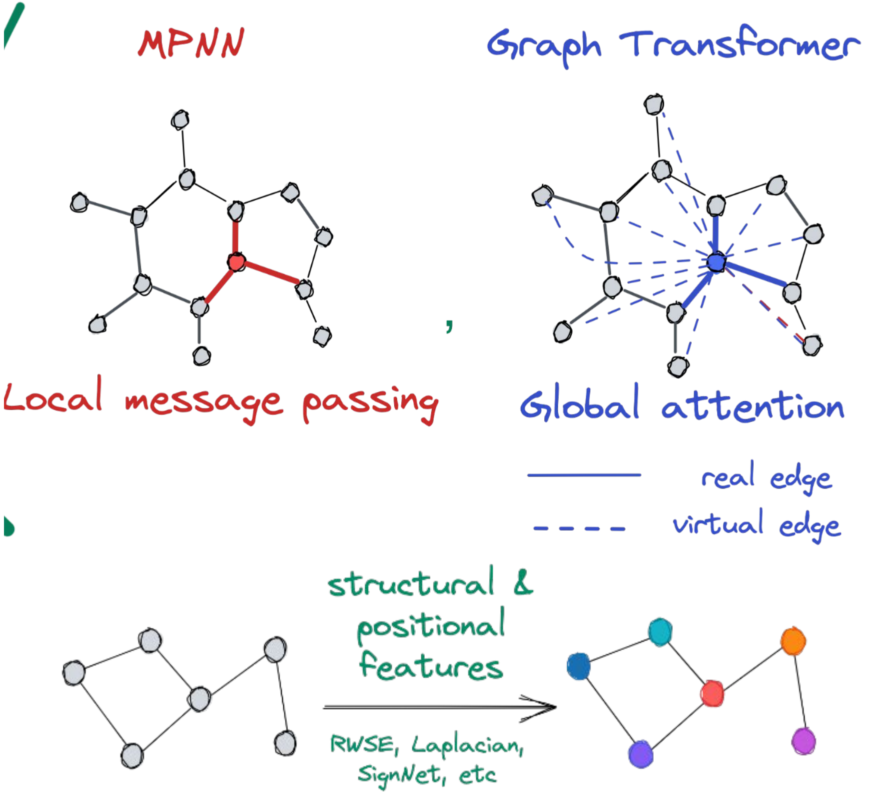




Duval, Mathis, Joshi, Schmidt et al. [A Hitchhiker's Guide to Geometric GNNs for 3D Atomic Systems](), 2023

The different flavors of GNNs: Graph Transformers (3/3)

- Each node "attend" to each other node in a fully connected graph.
- Effectively mitigate oversquashing, oversmoothing, better in long-range tasks
- Since input is a set now, nodes need **positional encodings** (obtained from the adjacency matrix, akin to RoPE in LLMs)

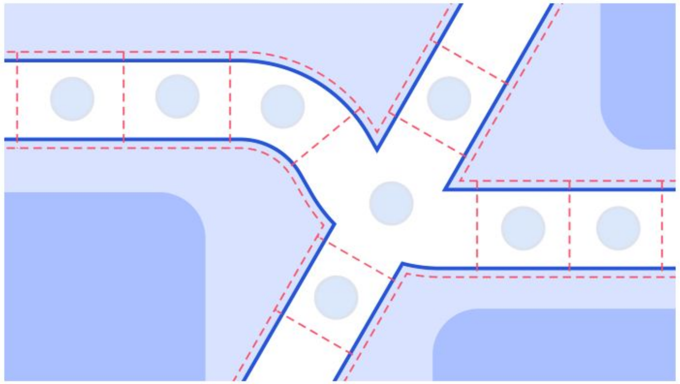


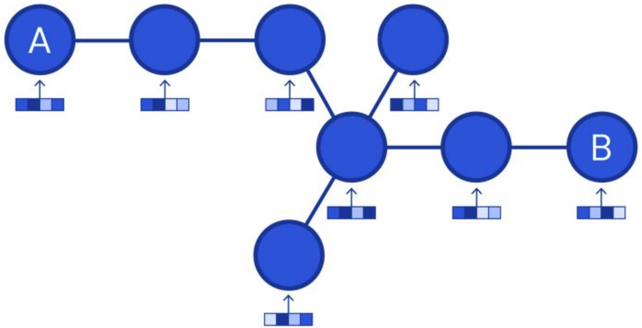
- Recipe for a General, Powerful, Scalable Graph Transformer, NeurIPS 2022

Examples of GNN Applications

Google Maps ETA Prediction

- **Goal:** Predict time of arrival (ETA) for a given route and starting time.
- **Data:** Road network + real-time and historical traffic data.
- **Modeling:**
  - **Route Segmentation:** Routes are divided into super-segments (approx. 80m), which are further split into smaller segments.
  - **Graph Creation:** Each super-segment forms a graph.
  - **Node Features:** Nodes include descriptions (e.g., highway status) and real-time + historical traffic information (time-series).
  - Per super-segment/segment learnable embeddings.
  - A GNN is trained to predict the traversal time for each super-segment.

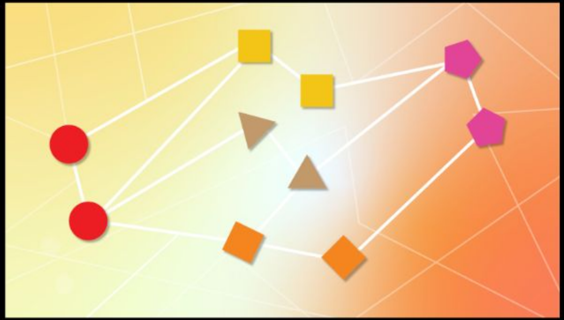




- [ETA Prediction with Graph Neural Networks in Google Maps (2021)]

Home > Blog >

Graph foundation models for relational data

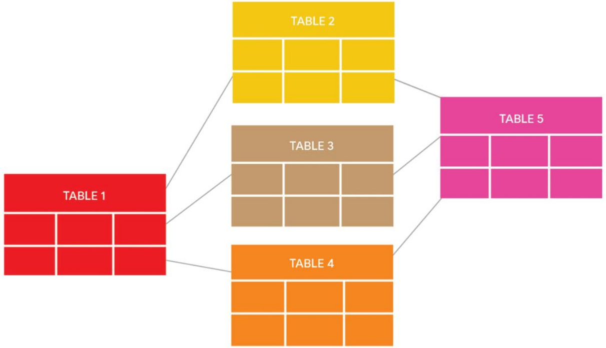


July 10, 2025 ·

Michael Galkin, Research Scientist, Google Research, and Pramod Doguparty, Software Engineer, Google Ads

Treating relational data as connected graphs enables generalist representations over very different graphs - culminating in **Graph Foundation Models (GFMs)**

Brings 3-40x gains compared to best tabular baselines (eg, GBDTs)

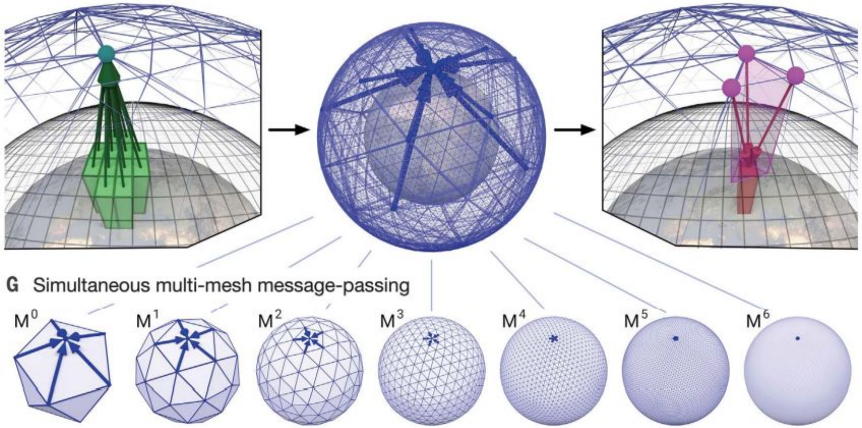


Graph Foundation Model

https://research.google/blog/graph-foundation-models-for-relational-data/

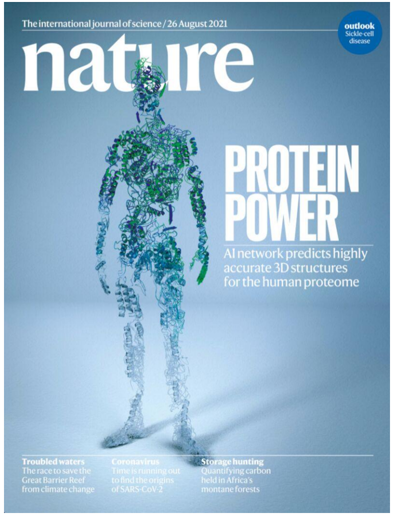
Weather Forecasting

- **Goal:** 10 days weather forecast
- **Data:** 39y of weather data (temp., wind, pleasure, precipitation, humidity)
- **Modeling:**
    - Grid over earth 721x1440.
    - Graph structure connecting neighbor cells + up/down mesh resolution (6x).
    - GNN takes current and -6h state, and predict +6h state.
    - Applied as an autoregressive model.

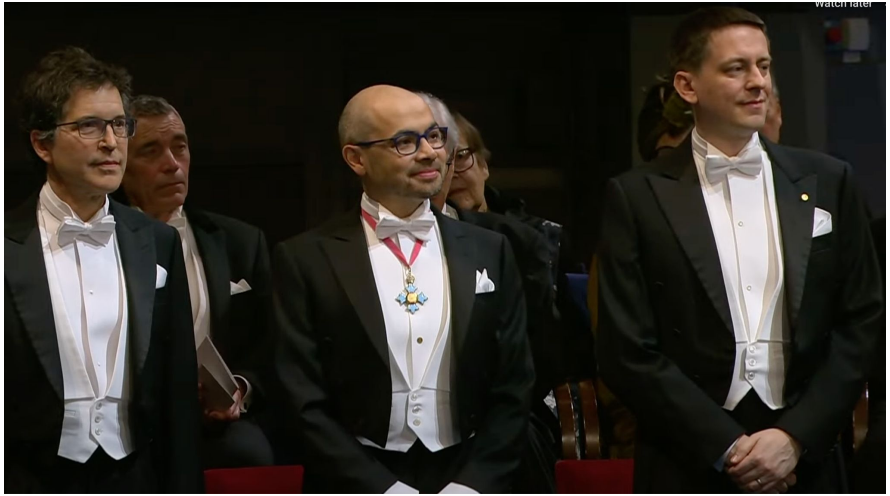


- [Learning skillful medium-range global weather forecasting (2021)]

Computational Chemistry



- **AlphaFold** - a poster child of Geometric Deep Learning
  - Amino acid sequence → protein structure
  - Immense advancement in computational biology
  - Nobel Prize 2024 in Chemistry



- **ProteinMPNN**
  - Protein structure → amino acid sequence.
  - Each a amino acid is a node.
  - Edge represent chemical bonds and spatial proximity.

Scaling GNNs

Very Large Graphs (1/3)

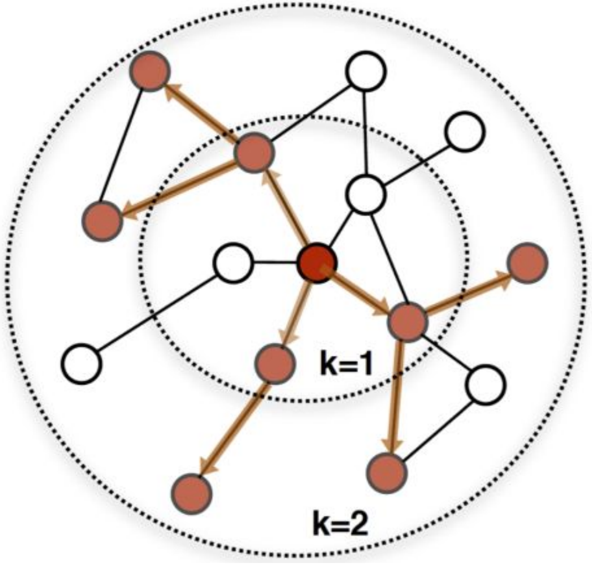
- **Small Graphs:** When data **naturally** divides into **small graphs**, GNN are trained directly (possibly with mini-batching and/or distributed training).

- **Large Graphs:** However, data sometimes includes **large graphs** (or is a single massive graph).
    - **Youtube** has XXXM daily active users, XB views per days, XXB of videos.
    - **Pinterest** graph has XB nodes and XXB edges.
    - **Alibaba** graph has XB nodes and XXB edges.
    - **Github** as XXXM users and XXXB files.

- **Core idea:** We need to split the graph data...without hurting training

Very Large Graphs (2/3)

- **Idea:** Sample the sub-graph around the node of interest
- **Example:** Recursive node-wise Sampling (GraphSAGE).
  - Scalable & easy to implement using popular distributed computation primities (e.g., managers/workers, mapreduce).
  - Few hops (e.g., 4 hops with 16 branches) is often enough.
  - Can easily be extended e.g. weighted sampling, temporal sampling, masked sampling.
  - Popular in the tech industry.
- Layer-wise sampling / subgraph sampling are less common.

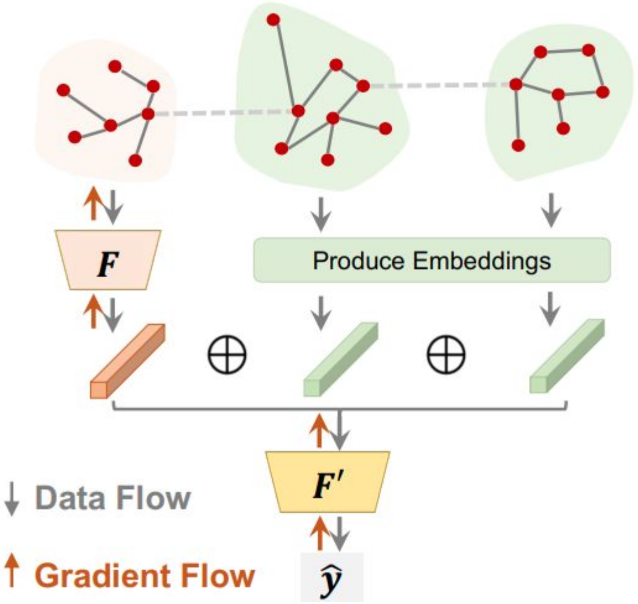


- [Inductive Representation Learning on Large Graphs (2017)]

Very Large Graphs (3/3)

- **Idea:** Split the graph over multiple machines.

- **Example:** Graph partitioning + divide-and-conquer computation (GST).
  - Partition the graph using METIS.
  - Run the forward pass on segments independently (distributed).
  - Run the backward pass on a subset of segments.
  - Aggregate the segment representation.



- Learning Large Graph Property Prediction via Graph Segment Training (2023)

Specialized hardware

- **GNN computation**
  - **Irregular memory access**: The output connections of the i-th node change from one graph (or graph sample) to another.
  - **Variable length data**: Different graph have different number of nodes and edges. Different nodes have a variable number of neighbors.

- Modern **hardware** (GPU/TPU, CPU at some extend) and **tooling** is optimized for **fixed size** data with **sequential** (or predictable) memory access patterns.

- GNN computation is slow (if you count flops).

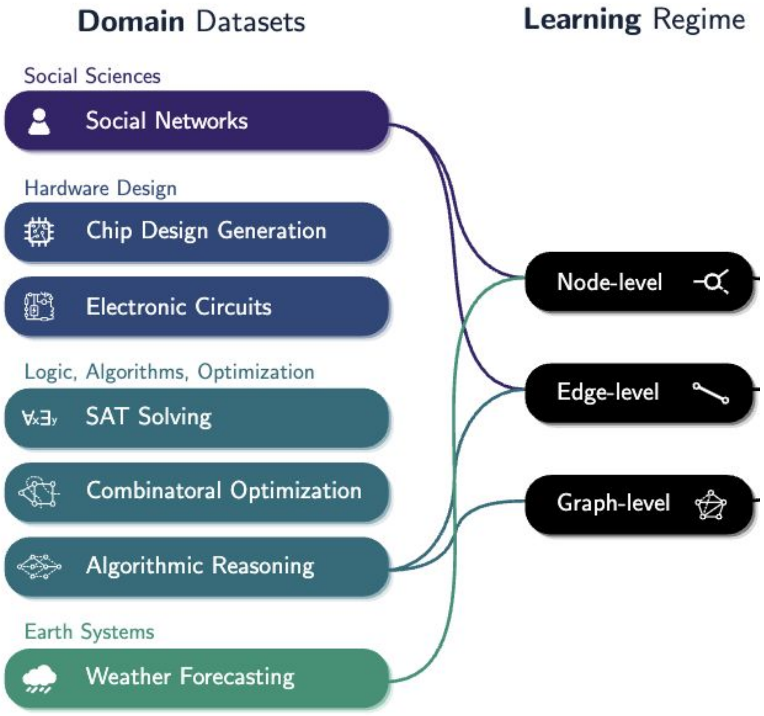
Efficiency

- GNN library have low level specialized kernels
  - PyG and GDL have specialized Cuda C++ implementation.
  - Nvidia is providing special primitives (cuEquivariance).
- Python kernel languages (e.g., Triton for GPU, Pallas for GPU/TPU) offer good perspective, but they are not yet competitive.
- GNN on FPGA / ASIC is actively research--but we have little experience with it...

Future of GNNs and conclusion

Future of GNNs

- Lot of small graphs
  - Molecules / proteins / crystals
- Combinatorial optimization problems
  - Turn ILP problems into GNN problems
- Circuit / chip design
  - Generative graph modeling task
- Foundational graph models
  - Generalist GNNs (for relational data), etc



Thank you for your attention