

Введение в GNN

Руководитель Центра Компетенций Data Science

Зелинский Никита

План лекции

- О1.** Обзор кейсов "в ширину"
- О2.** Перспективные направления исследований
- О3.** Современный стек в графах
- О4.** Базовые понятия и приемы в GNN
- О5.** Где брать актуальную информацию: полезные ресурсы, курсы и персоналии

План семинара

- О1.** Детальный пайплайн подготовки данных для обучения GNN
- О2.** GraphSAGE
- О3.** Обучение GNN на задаче бинарной классификации

Дисклеймеры

О1. Много ссылок и материалов

Дисклеймеры

- O1.** Много ссылок и материалов
- O2.** Не все статьи я помню детально

Дисклеймеры

- О1.** Много ссылок и материалов
- О2.** Не все статьи я помню детально
- О3.** Не все статьи я читал

Дисклеймеры

- О1.** Много ссылок и материалов
- О2.** Не все статьи я помню детально
- О3.** Не все статьи я читал
- О4.** Выводы и информация актуальна на начало ноября 2023,
но мир меняется очень быстро

Дисклеймеры

- О1.** Много ссылок и материалов
- О2.** Не все статьи я помню детально
- О3.** Не все статьи я читал
- О4.** Выводы и информация актуальна на начало ноября 2023,
но мир меняется очень быстро
- О5.** Впереди много интересного и максимальная отдача на пересечении
областей знаний – LLM точно не заменит собой весь AI/DS

Кейсы

01

1. RecSys



<https://recsys.acm.org/recsys23/session-6/>

Session-based Recommendation with Graph Neural Networks (2018)
<https://arxiv.org/pdf/1811.00855v4.pdf>



2023 OTTO – Multi-Objective Recommender System

Build a recommender system based on real-world e-commerce sessions
(30 000\$ prize, 2 574 teams)

<https://www.kaggle.com/competitions/otto-recommender-system>

7-е место

<https://www.kaggle.com/competitions/otto-recommendersystem/discussion/383769>

1. RecSys



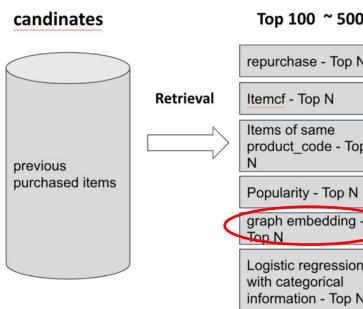
2022 H&M Personalized Fashion Recommendations

Provide product recommendations based on previous purchases
(50 000\$ prize, 2 952 teams)

<https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations>

1-е место

<https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations/discussion/324070>



One question is what's the improvement of graph embedding in your recalling and ranking? That's a part which I wanna try but didn't have enough time.

Thanks for your sharing.

↪ Reply



senkin13

Posted a year ago · 1st in this Competition

TOPIC AUTHOR

about 0.002 up

↪ Reply



sirius

Posted a year ago · 3rd in this Competition

Wow, that's really powerful!

1. RecSys



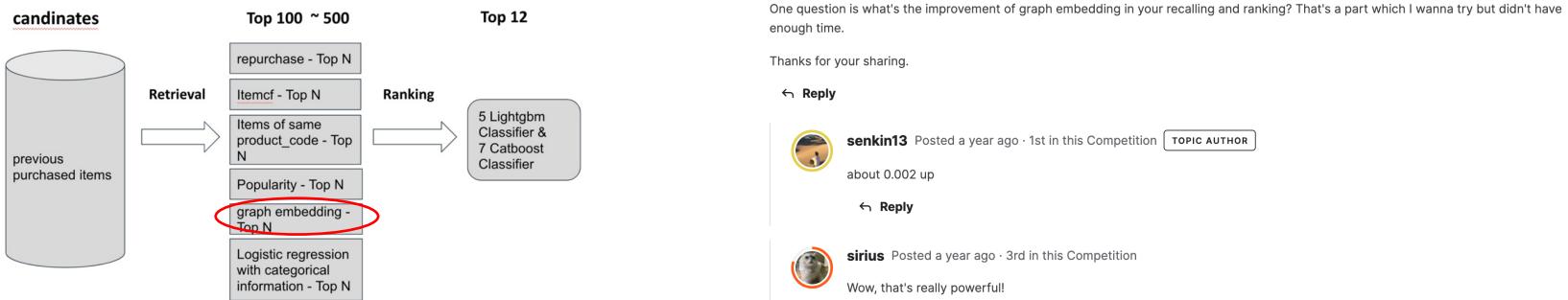
2022 H&M Personalized Fashion Recommendations

Provide product recommendations based on previous purchases
(50 000\$ prize, 2 952 teams)

<https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations>

1-е место

<https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations/discussion/324070>



<https://arxiv.org/abs/2311.00423v1>

LLMRec: Large Language Models with Graph Augmentation for Recommendation. LLMRec - это новый фреймворк и датасет, улучшающий рекомендательные системы путем применения простых, но эффективных стратегий дополнения графов на основе LLM.
– Статья принята на WSDM'24

2. Supply chains management

Цель: купить дешевле, вовремя
(когда у других дефицит), снизить
риски дефицита у себя
(сможет ли цепь выполнить заказ?)



Насколько такие цепи прозрачны?

Контракты конфиденциальны:

- Кто, кому, сколько, по какой цене?
- А в будущем?

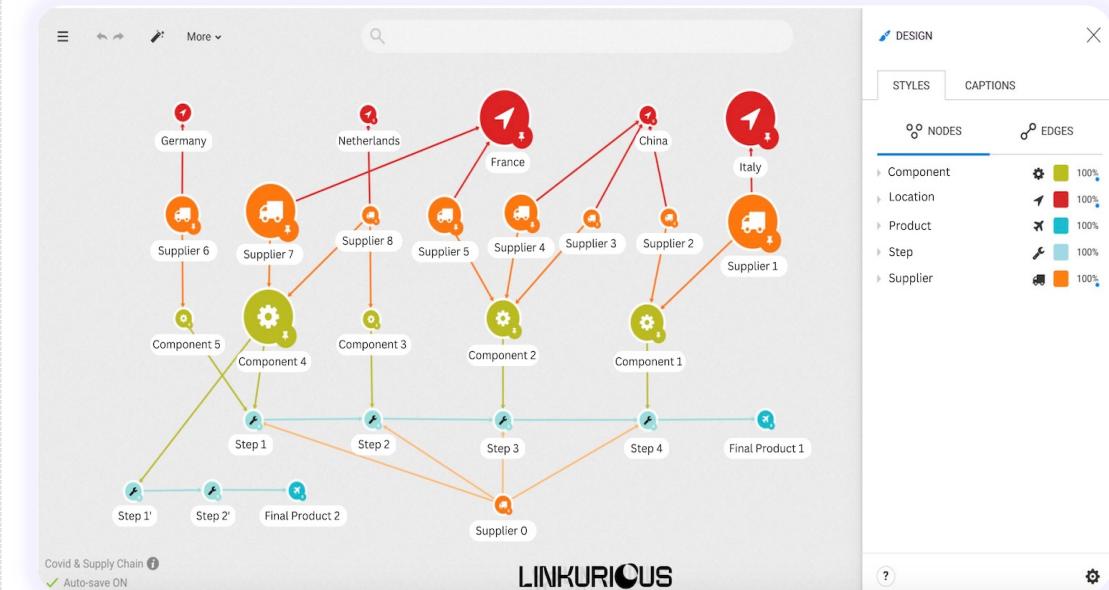
Кто ключевой связующий поставщик?
(Nexus supplier)



Supply Chain Link Prediction on Uncertain Knowledge Graph
https://kdd.org/exploration_files/b124-AI4Manufacturing_paper5.pdf

Чуть далее: извлечение цепочки поставок из новостей
Extracting supply chain maps from news articles using deep neural networks (2020)
<https://api.repository.cam.ac.uk/server/api/core/bitstreams/1494b97c-0c4b-4a7ba562-10e1b0d11586/content>

Кусочек модельной цепи поставок



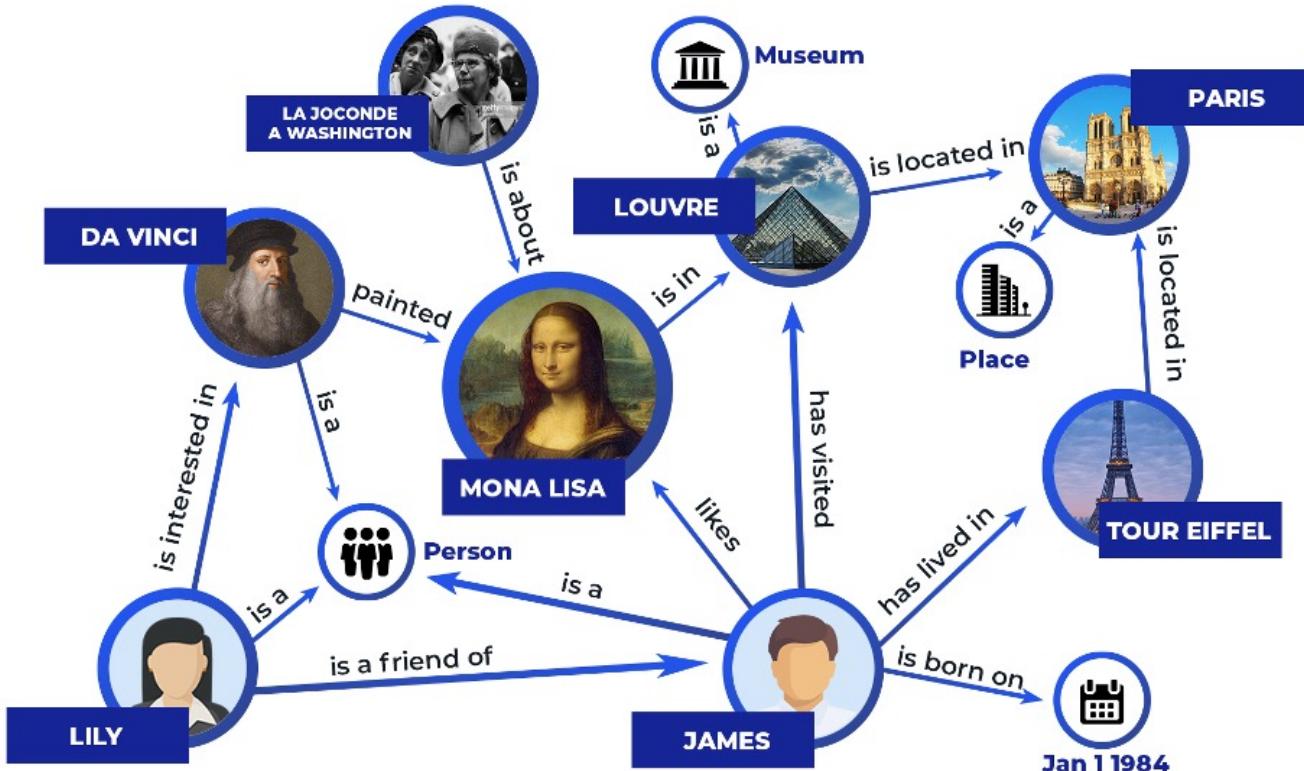
<https://linkurious.com/blog/supply-chain-data-analysis/>



Cambridge University

<https://dl.acm.org/doi/abs/10.1145/3575637.3575655>
<https://api.repository.cam.ac.uk/server/api/core/bitstreams/92a7b226-2f41-4703-ad15-69c53ec7f26a/content>
<https://github.com/grandintegrator/Link-Prediction-Supply-Chains>

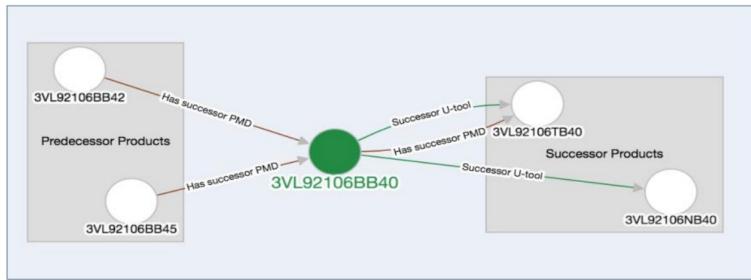
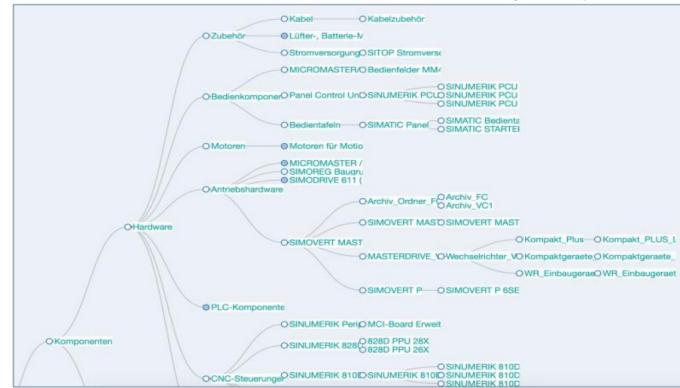
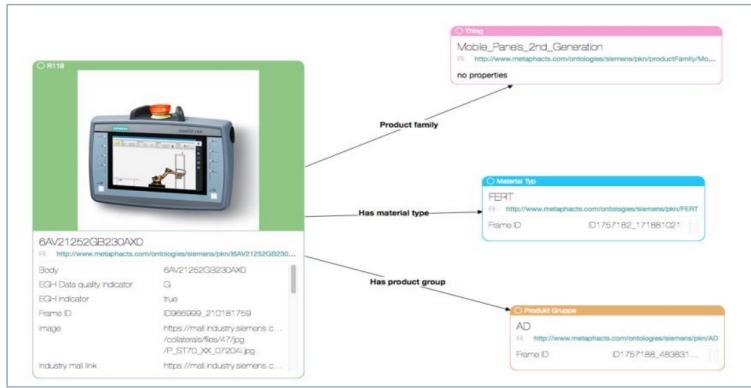
3. Knowledge graph



3. Knowledge graph (Siemens, 2017)

Graph Visualization and Exploration

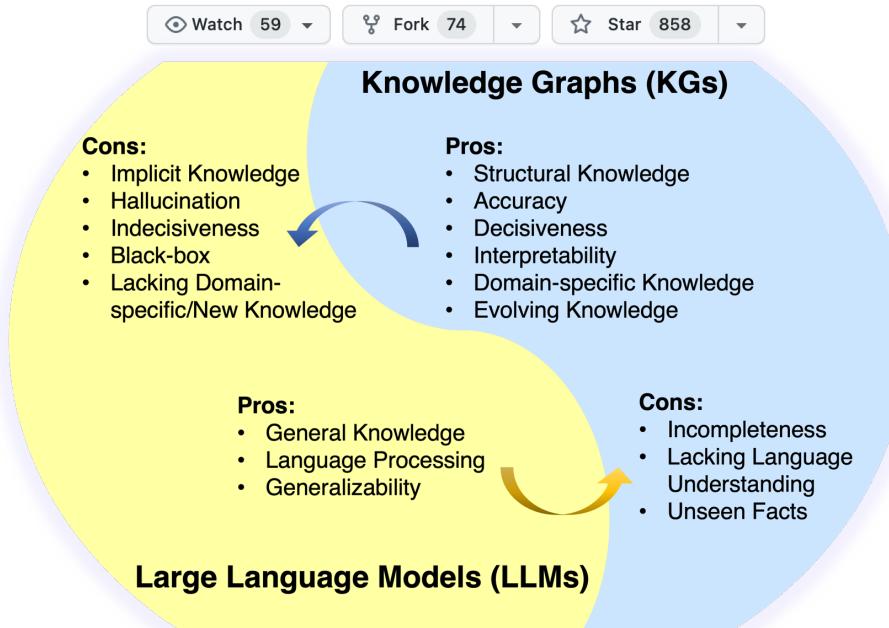
SIEMENS
Ingenuity for life



Variety of graph structures:

- Product metadata and relationships
- Successor, predecessor network
- Taxonomic information

3. Robust reasoning over Knowledge Graph with LLM (2023)



<https://github.com/RManLuo/Awesome-LLM-KG>

Towards foundation models for knowledge graph reasoning

<https://arxiv.org/pdf/2310.04562.pdf>

<https://towardsdatascience.com/ultra-foundation-models-for-knowledge-graph-reasoning-9f8f4a0d7f09>



NEURAL INFORMATION
PROCESSING SYSTEMS

Do Temporal Knowledge Graph Embedding Models Learn or Memorize

<https://openreview.net/forum?id=UMokRwWfLW>

**Правила показали около-SOTA результаты, причина – лики при обучении
КГ-эмбеддингов**

4. Document understanding (link prediction)

<https://paperswithcode.com/sota/entity-linking-on-funsd?p=doc2graph-a-task-agnostic-document>

Filter: [untagged](#) [Edit Leaderboard](#)

Rank	Model	F1	Paper	Code	Result	Year	Tags
1	TPP (LayoutMask)	79.23	Reading Order Matters: Information Extraction from Visually-rich Documents by Token Path Prediction	🔗	2023	13 октября	
2	SINGU_GROUP	70.51	DGCN Based Solution for Entity Linking on Visual Rich Document	🔗	2022		
3	SERA	65.96	Entity Relation Extraction as Dependency Parsing in Visually Rich Documents	🔗	2021		
4	Doc2Graph	53.36	Doc2Graph: a Task Agnostic Document Understanding Framework based on Graph Neural Networks	🔗	2022		

Doc2Graph: a Task Agnostic Document Understanding Framework based on Graph Neural Networks (2022)

<https://arxiv.org/abs/2208.11168>

<https://github.com/andreagemelli/doc2graph>

5. Fraud detection – GTAN – SOTA на датасетах Amazon-Fraud и Yelp-Fraud

Rank	Model	AUC-ROC	Averaged Precision	Paper	Code	Result	Year	Tags
1	GTAN	97.50	89.26	Semi-supervised Credit Card Fraud Detection via Attribute-driven Graph Representation			2023	
2	RLC-GNN	97.48		RLC-GNN: An Improved Deep Architecture for Spatial-Based Graph Neural Network with Application to Fraud Detection			2021	
3	RioGNN	96.19		Reinforced Neighborhood Selection Guided Multi-Relational Graph Neural Networks			2021	
4	PC-GNN	95.86	85.49	Pick and Choose: A GNN-based Imbalanced Learning Approach for Fraud Detection			2021	
5	CARE-GNN	89.73	82.19	Enhancing Graph Neural Network-based Fraud Detectors against Camouflaged Fraudsters			2020	

<https://paperswithcode.com/sota/fraud-detection-on-amazon-fraud>

<https://arxiv.org/pdf/2307.05633.pdf>

<https://arxiv.org/pdf/2205.13084.pdf>

Кейсы в финансах

6. Probability of Default



Financial Default Prediction via Motif-preserving Graph Neural Network with Curriculum Learning (2023)

<https://dl.acm.org/doi/abs/10.1145/3580305.3599351>

7. AML (anti-money laundering)

Finding Money Launderers Using Heterogeneous Graph Neural Networks (2023)

<https://arxiv.org/abs/2307.13499>

8. Portfolio management

Graph Neural Networks for Asset Management (2021)

<https://research-center.amundi.com/files/nuxeo/dl/52877ebd-db7c-4ddd-95dc-b77f6b5bcf52>

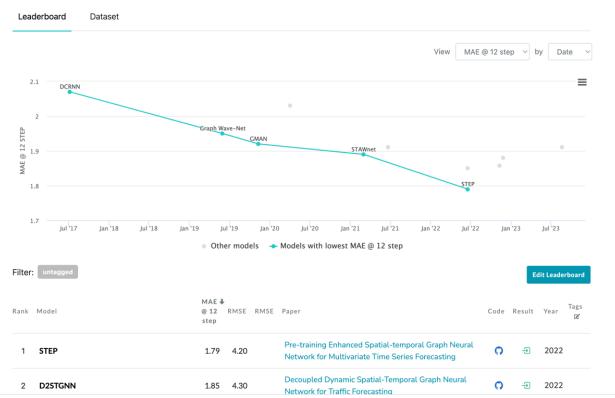
Multi-Level Graph Neural Network for Information Fusion in Learning Stock Market Dynamics (2023)

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4423354

9. ETA (estimated time of arrival) and traffic prediction -- SOTA



Traffic Prediction on PEMS-BAY



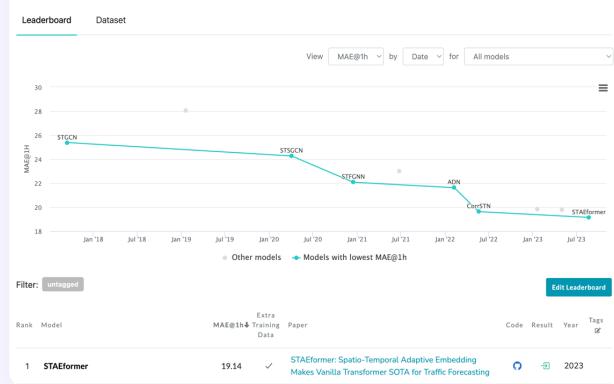
<https://paperswithcode.com/sota/traffic-prediction-on-pems-bay>

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

<https://arxiv.org/pdf/2108.11482.pdf>

<https://www.youtube.com/watch?v=zpDdvI95igc>

Traffic Prediction on PeMS07



<https://paperswithcode.com/sota/traffic-prediction-on-pems07>

Estimating package arrival time via heterogeneous hypergraph neural network (2023)

<https://www.sciencedirect.com/science/article/abs/pii/S095741742302242X>

Multi-attention graph neural networks for city-wide bus travel time estimation using limited data (2022)

<https://www.sciencedirect.com/science/article/abs/pii/S0957417422004717>

10. Causal Learning

Relating Graph Neural Networks to Structural Causal Models (2021) (Graph VAE)

<https://arxiv.org/pdf/2109.04173.pdf>

<https://www.youtube.com/watch?v=XC-Bfg3dO0I>



CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting (2022)

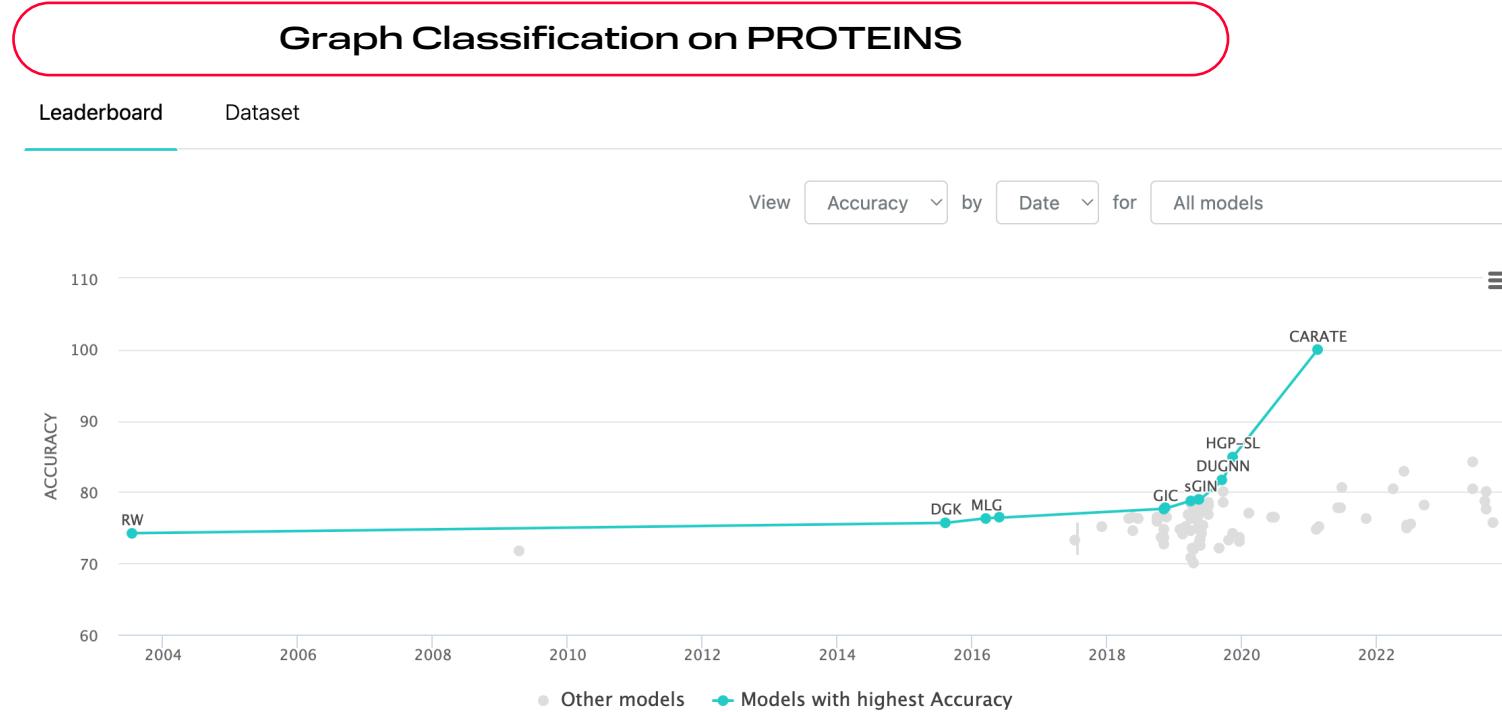
<https://ojs.aaai.org/index.php/AAAI/article/view/21479>

Hierarchical Graph Neural Networks for Causal Discovery and Root Cause Localization (2023)

<https://arxiv.org/pdf/2302.01987.pdf>

<https://www.youtube.com/watch?v=AStrI3FhMWg>

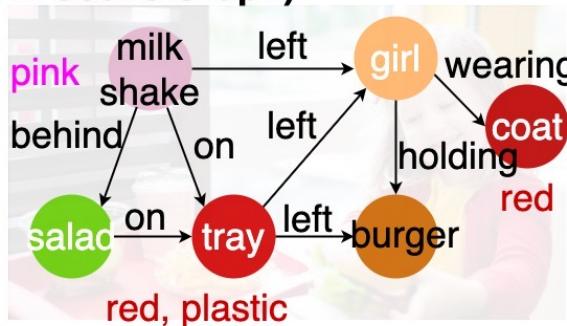
11. Molecular embeddings (88 моделей и все с GNN) – SOTA



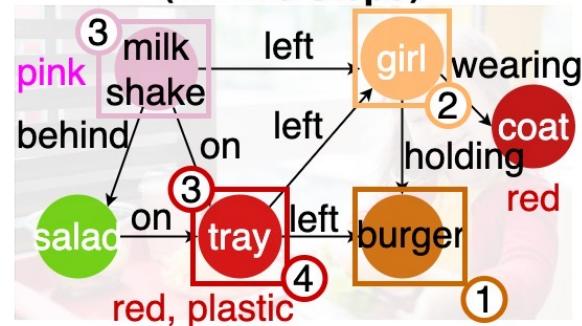
12. Visual query answering (2019 - now)



**Input: Image
(Represented as A Scene Graph)**



**Step 1: Scene Graph Reasoning
(4 Time Steps)**



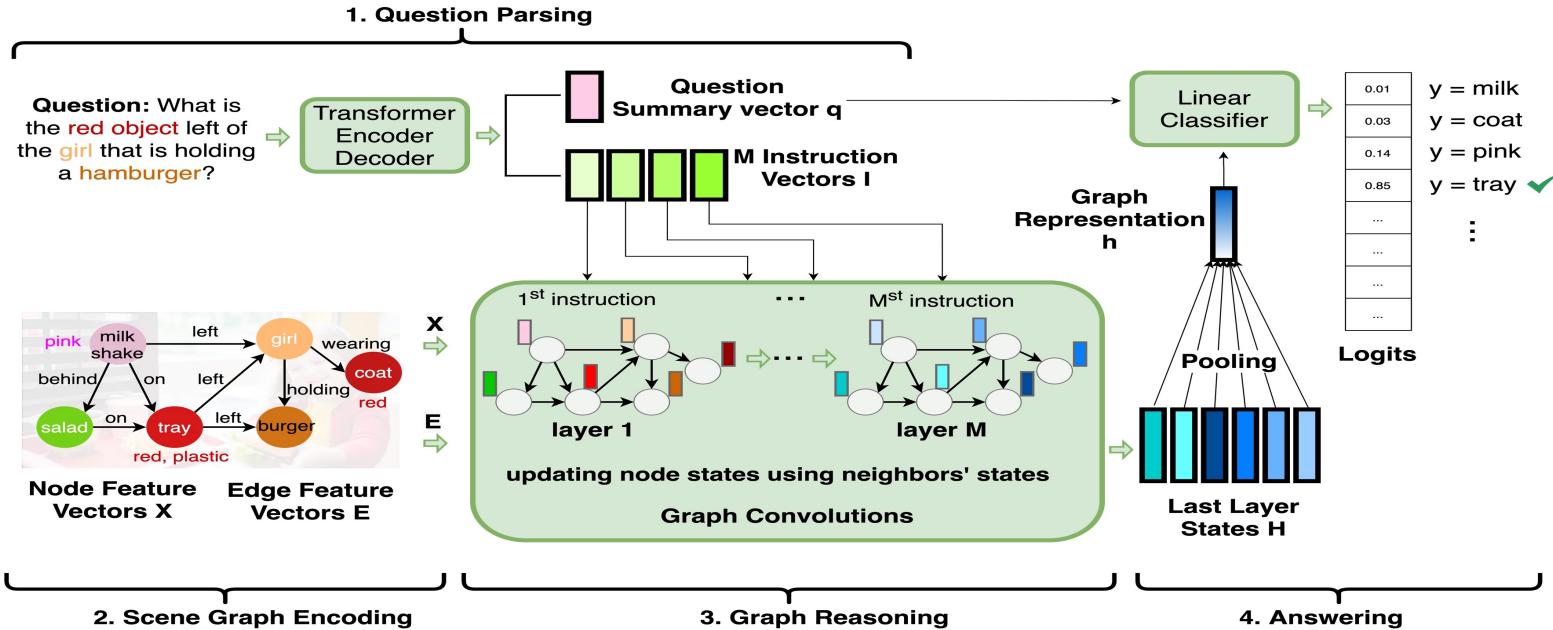
Input: Question

What is the **red object** left of the **girl** that is holding a **hamburger**?

Step 2: Answer Prediction

Answer: Tray
(Graph Classification Problem)

12. Visual query answering



13. Прогноз погоды – SOTA

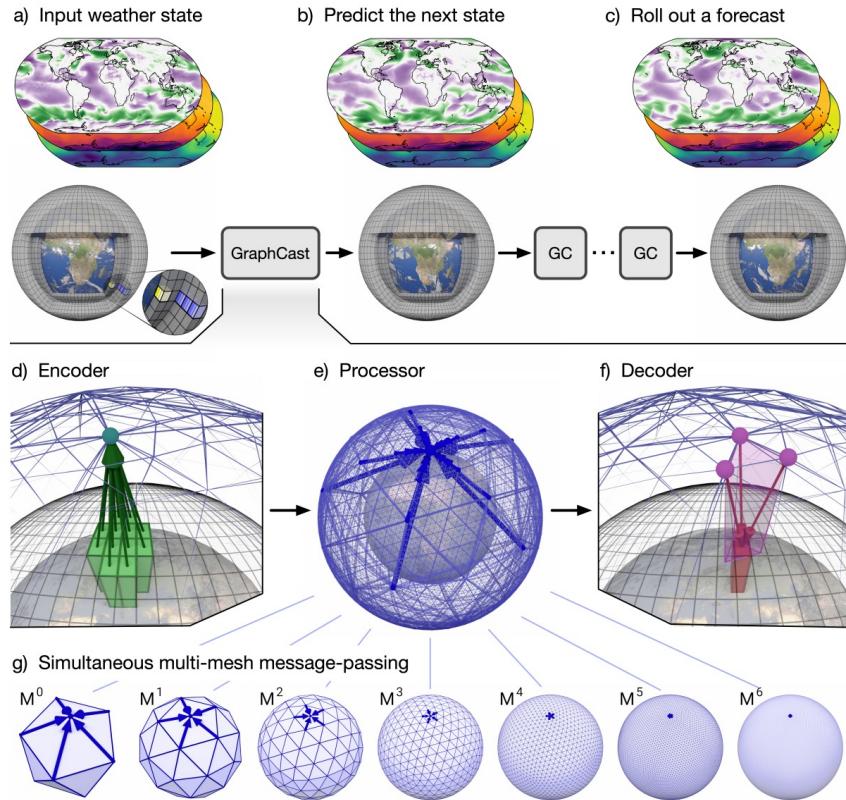


GraphCast: Learning skillful medium-range global weather forecasting (24.12.2022)

<https://arxiv.org/pdf/2212.12794v2.pdf>

<https://github.com/google-deepmind/graphcast>

<https://paperswithcode.com/paper/graphcast-learning-skillful-medium-range>



14. Социальные сети и СМИ – SOTA

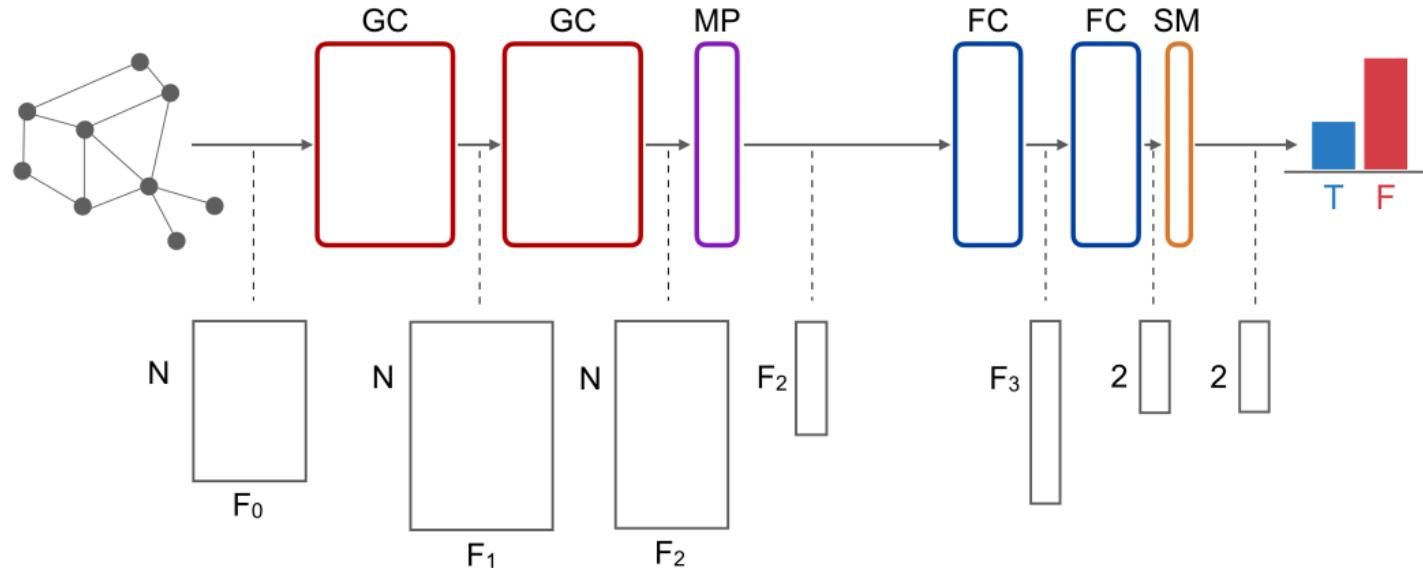


Figure 5: The architecture of our neural network model. Top row: GC = Graph Convolution, MP = Mean Pooling, FC = Fully Connected, SM = SoftMax layer. Bottom row: input/output tensors received/produced by each layer.

Fake News Detection on Social Media using Geometric Deep Learning (2019)

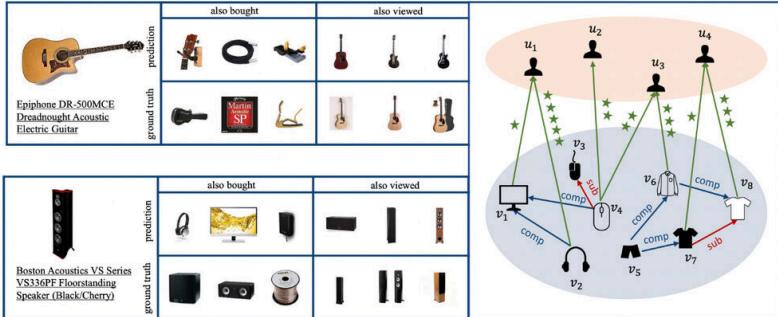
<https://arxiv.org/pdf/1902.06673.pdf>

Подборка кейсов (2022) от
[Nikita Iserson](#)

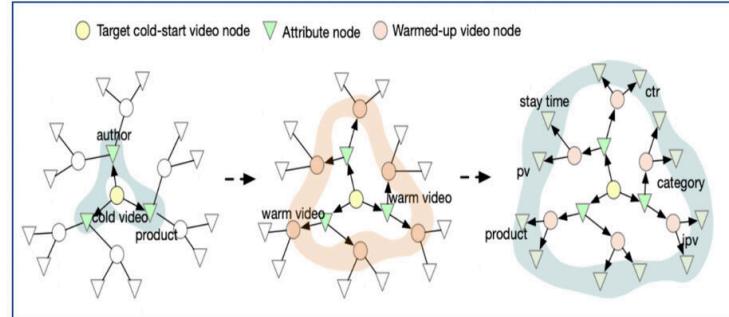
1ni. E-com

Graph Neural Networks for E-Commerce

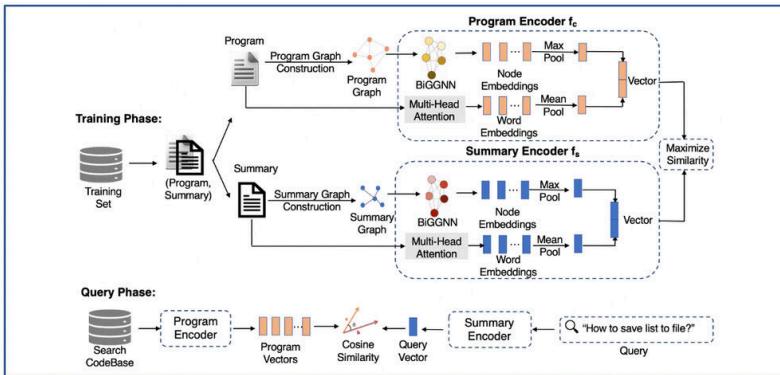
Product Matching, Item Relationship Prediction



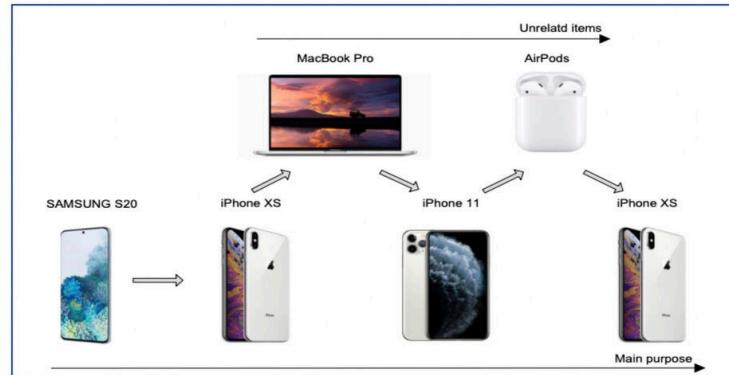
New Product / New Shop Forecasting / Strict Cold Start



Improving Search Relevance and Information Retrieval



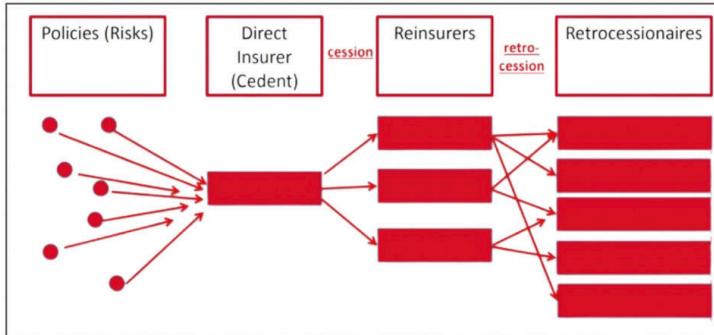
Session-Based Recommendation and Click Models



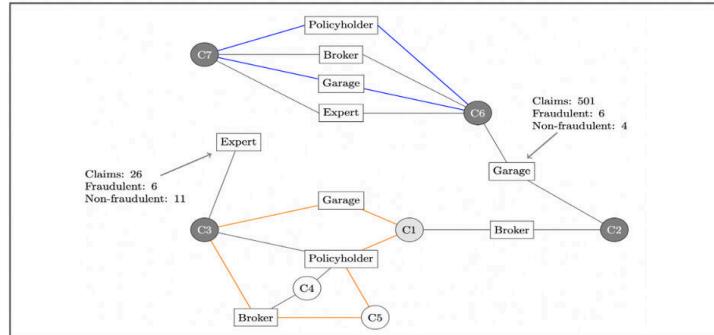
2ni. Страхование

Graph Neural Networks for Insurance

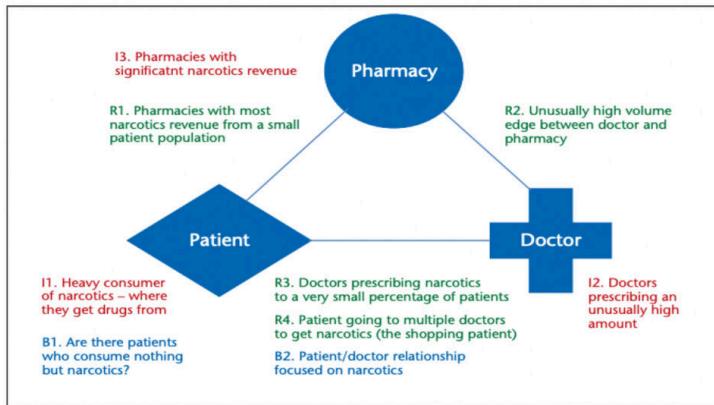
Modeling Risk, Reinsurance Networks, Cascading Losses



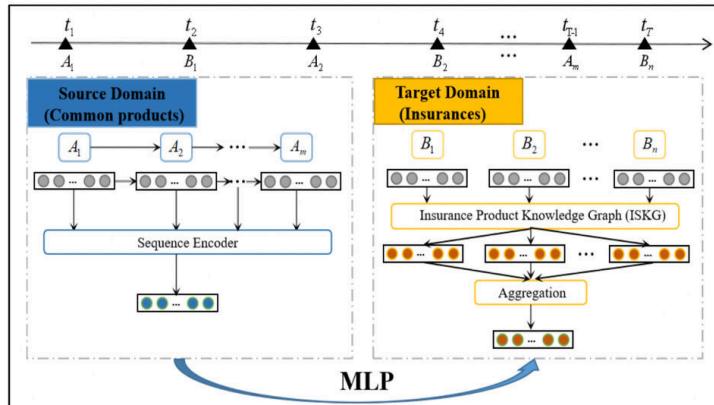
Fraud Detection in Motor Insurance Sector



Healthcare & Medicare (Treatment, Actuarial, Fraud & Abuse)



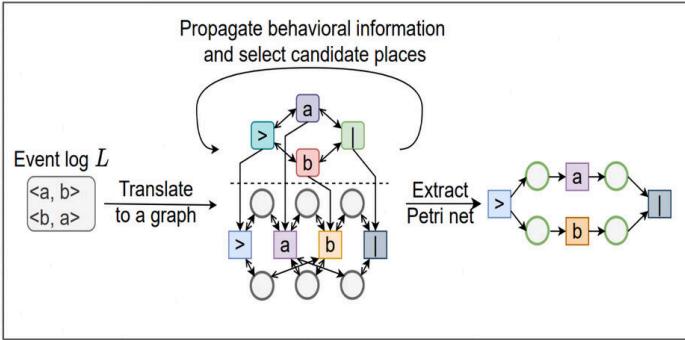
Recommender Systems and CyberInsurance Ratemaking



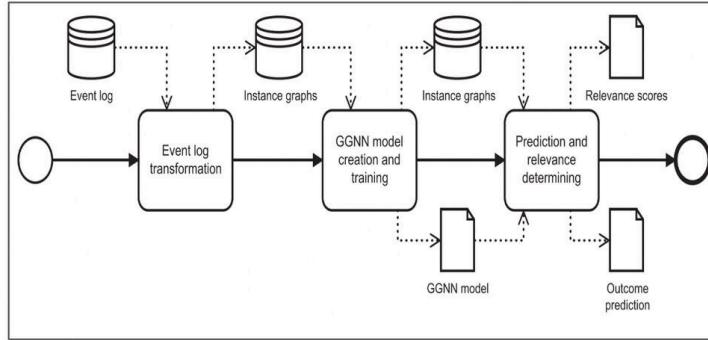
3ni. Process Mining

Graph Neural Networks for Process Analytics

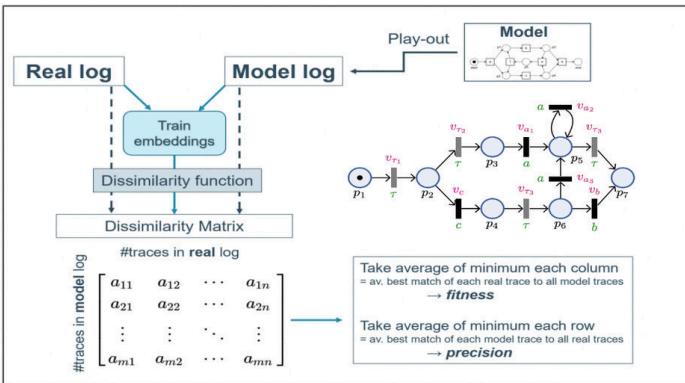
Process Discovery



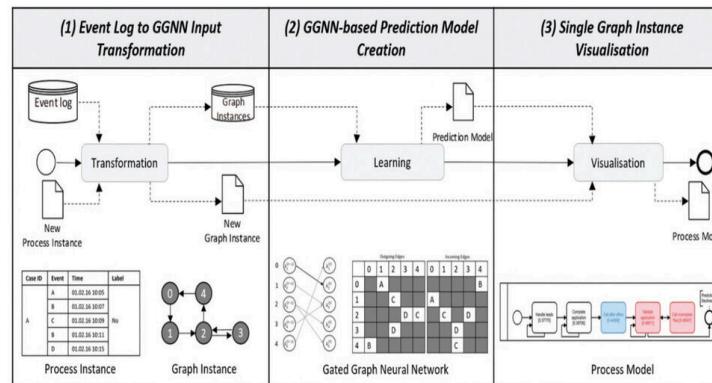
Process Monitoring (Activity - KPI Relevance)



Conformance Checking & Trace Embedding

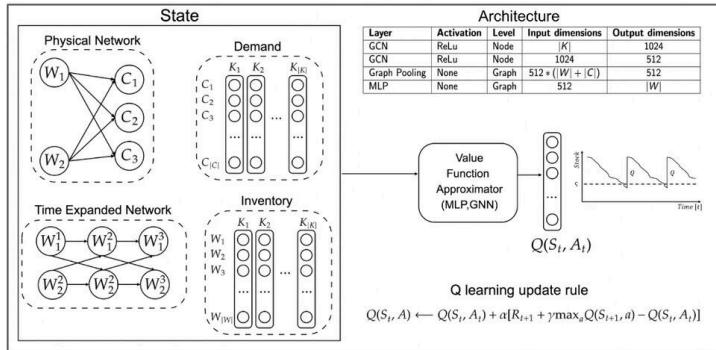


Explainable Process Prediction

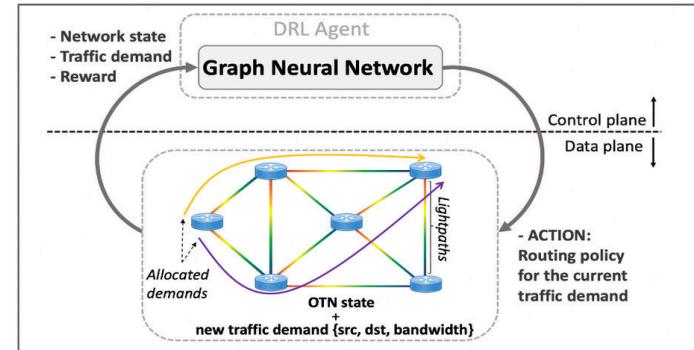


Graph Neural Networks with Reinforcement Learning

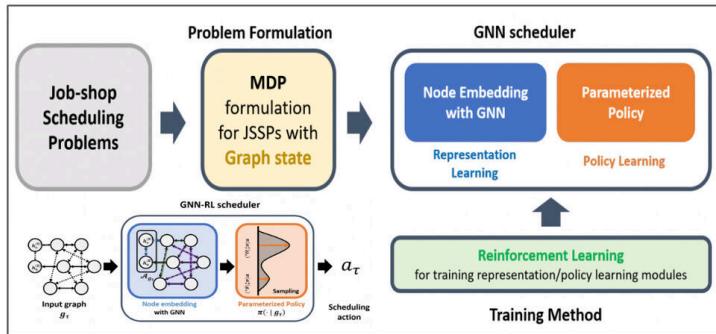
Multi-Echelon Inventory Optimization



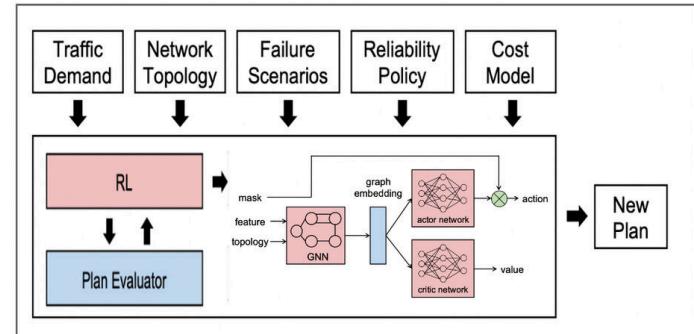
Logistics, Routing and Transportation



Manufacturing Control and Scheduling



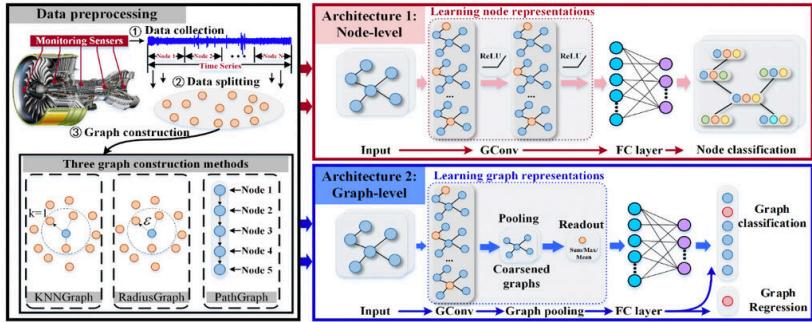
Dynamic Network Control and Planning



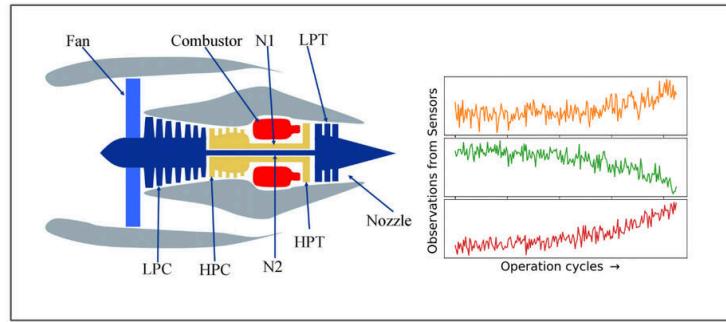
5ni. Производство

Graph Neural Networks for Equipment Health Monitoring

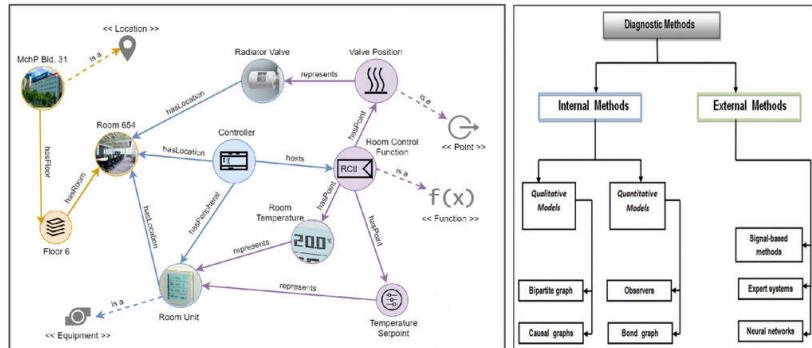
Prognostics & Health Management, Fault Detection



Remaining Useful Lifetime Estimation, Anomaly Detection



Semantic and Causal Condition Monitoring



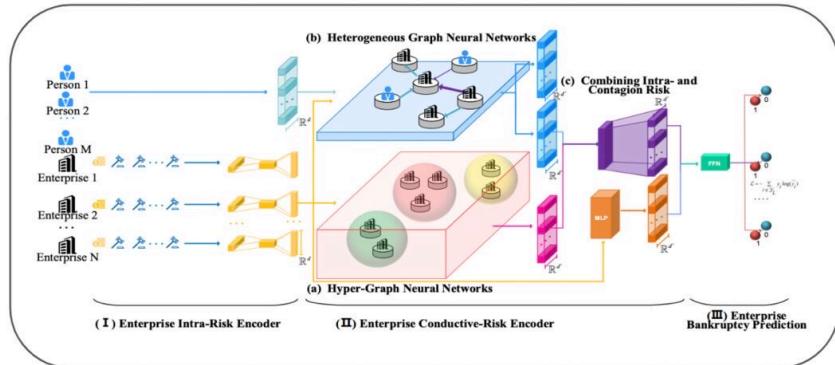
Structural Health Monitoring



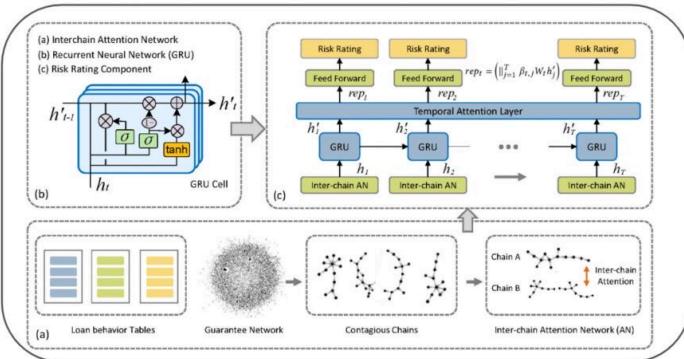
6ни. Финансы

Graph Machine Learning for Financial Networks

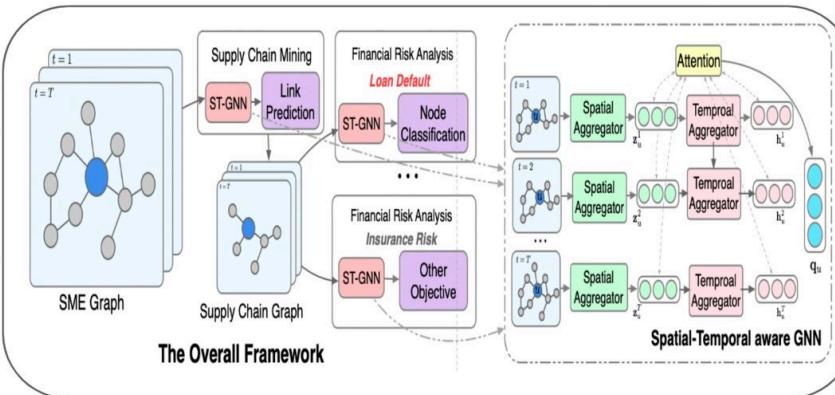
Bankruptcy, Stock, Financial Event Prediction



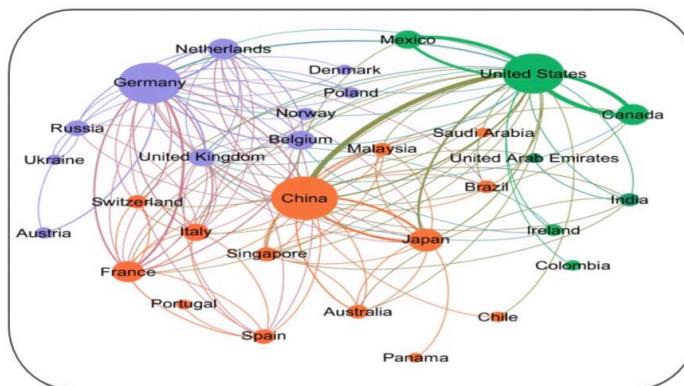
System Risk & Stress Testing



Supply Chain Risk Assessment



International Trade Modeling - Gravity Models



Перспективные направления исследований

02

2. Перспективные направления исследований



- Temporal graphs and Causality
- Uncertainty estimation
- Time-series prediction
- Explainability
- Large-scale (Kumo – 50 млрд вершин, etc.)
- Physics-Inspired Graph Neural Networks

Physics-Inspired Graph Neural Networks (2023), M. Bronstein

https://www.youtube.com/watch?v=bb_JKLS8h4A

Evaluating explainability for graph neural networks (2023)

<https://www.nature.com/articles/s41597-023-01974-x>

Современный стек в графах

ОЗ

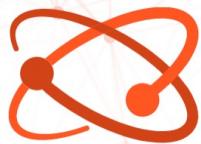
3. Современный стек в графах



PyTorch
geometric

https://github.com/pyg-team/pytorch_geometric

<https://pytorch-geometric.readthedocs.io/en/latest/>



Torch
Spatiotemporal

<https://torch-spatiotemporal.readthedocs.io/en/latest/>

<https://github.com/TorchSpatiotemporal/tsl>

Для распределенного обучения

graph-learn

Alibaba.com

<https://github.com/alibaba/graphlearn-for-pytorch>

3. 23 ноября 2023



<https://github.com/MobileTeleSystems/CoolGraph>

- Легкий старт с GNN
- Автоподбор почти всего, в тч параметров предобработки данных, размера батча, архитектур, параметров слоев и функций активации
- Поддерживает фичи вершин, ребер, группы вершин с разными наборами фичей
- Эксперименты трекаются в MLFlow
- Можно использовать для быстрого получения эмбеддинга

3. последний релиз 15 августа 2023



<https://github.com/dmlc/dgl>



⚠ Important Announcement: Deprecation of MXNet and TensorFlow Support in DGL 2.0

Questions



frozenbugs ⚡

Aug 23

Aug 23

❗ Important DGL Update: After DGL 2.0 (ETA: 2023Q4), we'll be prioritizing PyTorch backend due to resource constraints, saying goodbye to MXNet and TensorFlow support. Don't worry though – if you still need them, you can use older DGL versions. We're here to help you transition smoothly. Thanks for your understanding and ongoing support!

1 / 2

Aug 23

Базовые операции и приемы

04

Pipeline



01

02

03



Индексация,
семплинг,
трансформации
фичей,
маскирование и
пр.
– на семинаре

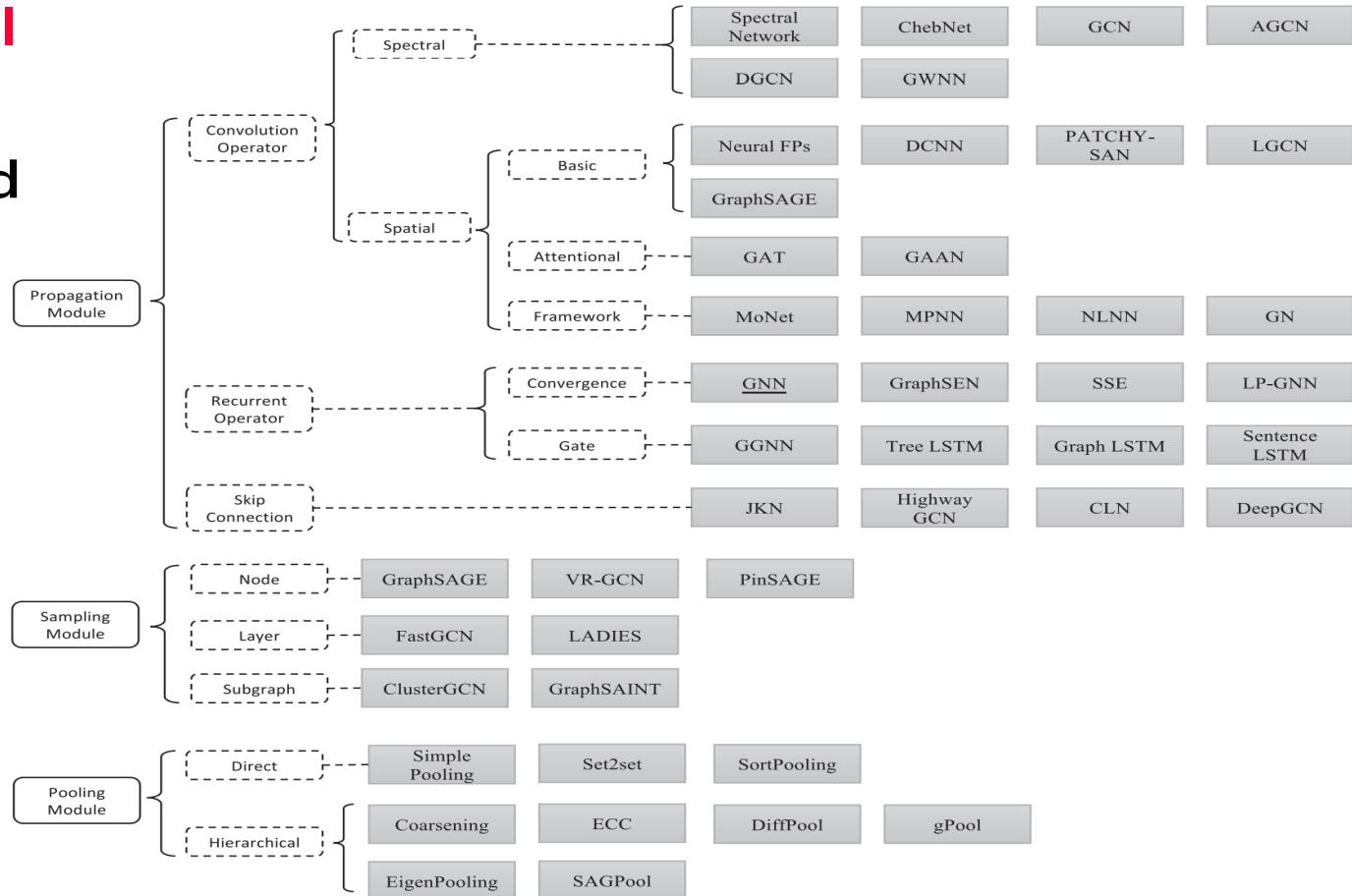


Батчи и
нормализация
– на семинаре



Message
Passing

Graph neural networks: a review of methods and applications (2020)



14. Социальные сети и СМИ

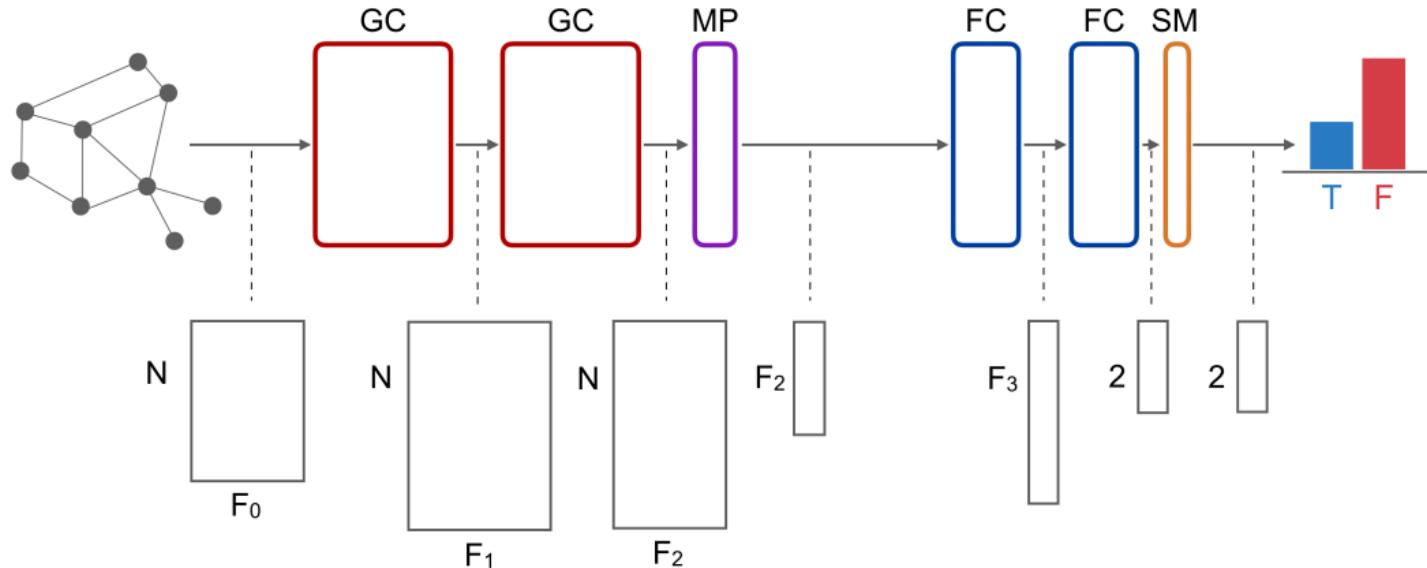


Figure 5: The architecture of our neural network model. Top row: GC = Graph Convolution, MP = Mean Pooling, FC = Fully Connected, SM = SoftMax layer. Bottom row: input/output tensors received/produced by each layer.

Fake News Detection on Social Media using Geometric Deep Learning (2019)

<https://arxiv.org/pdf/1902.06673.pdf>

Графовая свертка GraphConv

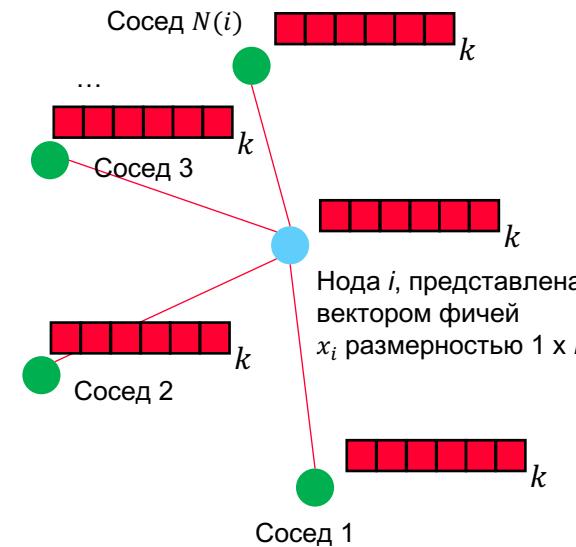


Нода i , представлена
вектором фичей
 x_i размерностью $1 \times k$

https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.conv.GraphConv.html

<https://arxiv.org/pdf/1704.01212.pdf>
<https://arxiv.org/pdf/1810.02244.pdf>
<https://arxiv.org/pdf/1709.05584.pdf>

Графовая свертка GraphConv



[https://pytorch-
geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.conv.GraphConv.html](https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.conv.GraphConv.html)

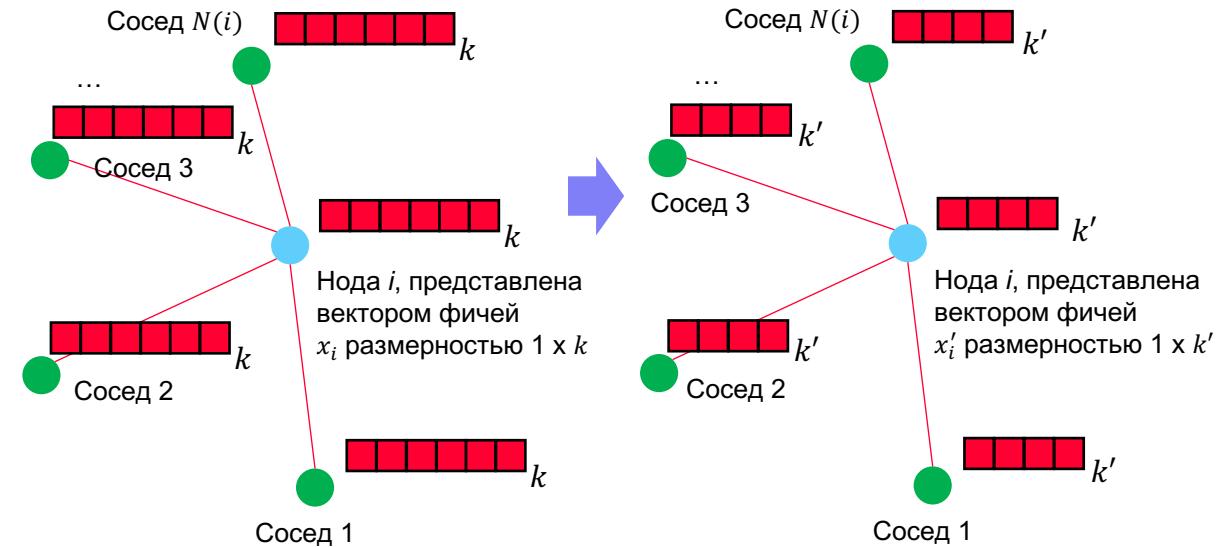
<https://arxiv.org/pdf/1704.01212.pdf>
<https://arxiv.org/pdf/1810.02244.pdf>
<https://arxiv.org/pdf/1709.05584.pdf>

Графовая свертка GraphConv

$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} Wx_j$$

$$x'_i = \boxed{Hx_i} + \sum_{j \in [1, N(i)]} \boxed{Wx_j}$$

H и W имеют размерность $k \times k'$:
 $(1 \times k) * (k \times k') = 1 \times k'$



https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.conv.GraphConv.html

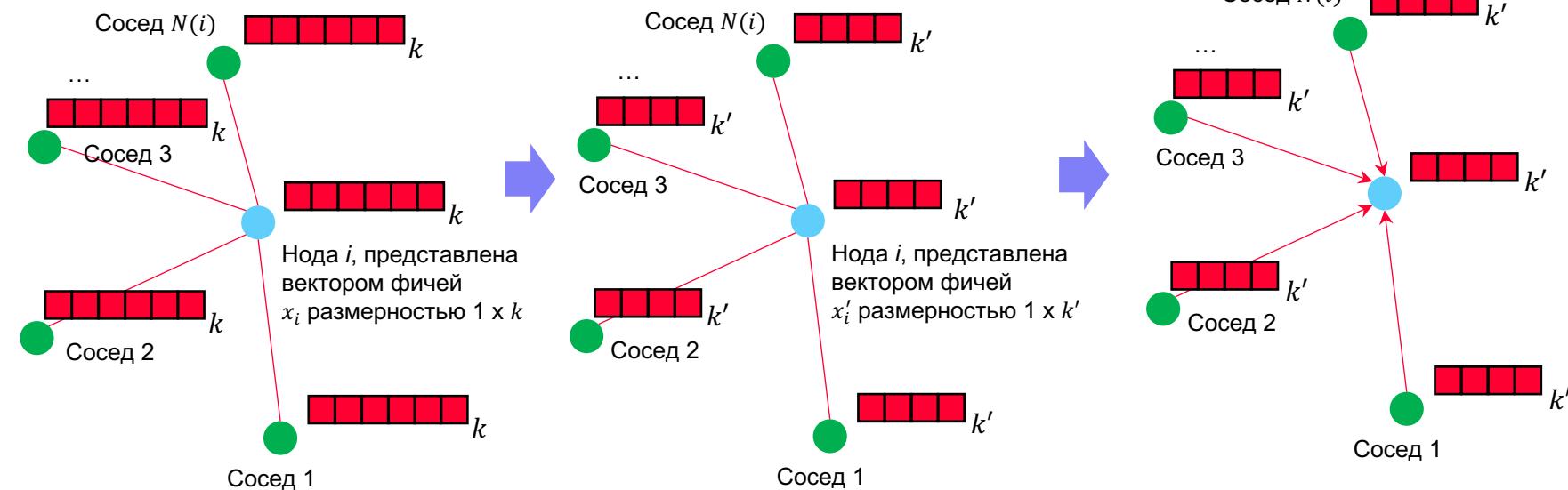
<https://arxiv.org/pdf/1704.01212.pdf>
<https://arxiv.org/pdf/1810.02244.pdf>
<https://arxiv.org/pdf/1709.05584.pdf>

Графовая свертка GraphConv

$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} Wx_j$$

$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} Wx_j$$

W и H имеют размерность $k \times k'$:
 $(1 \times k) * (k \times k') = 1 \times k'$



https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.conv.GraphConv.html

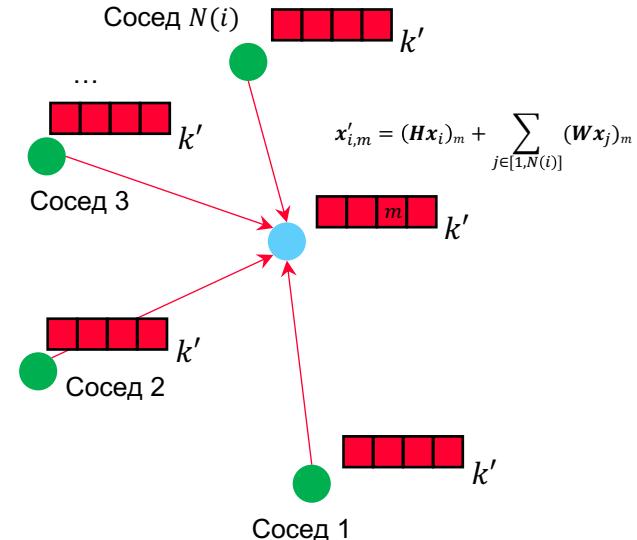
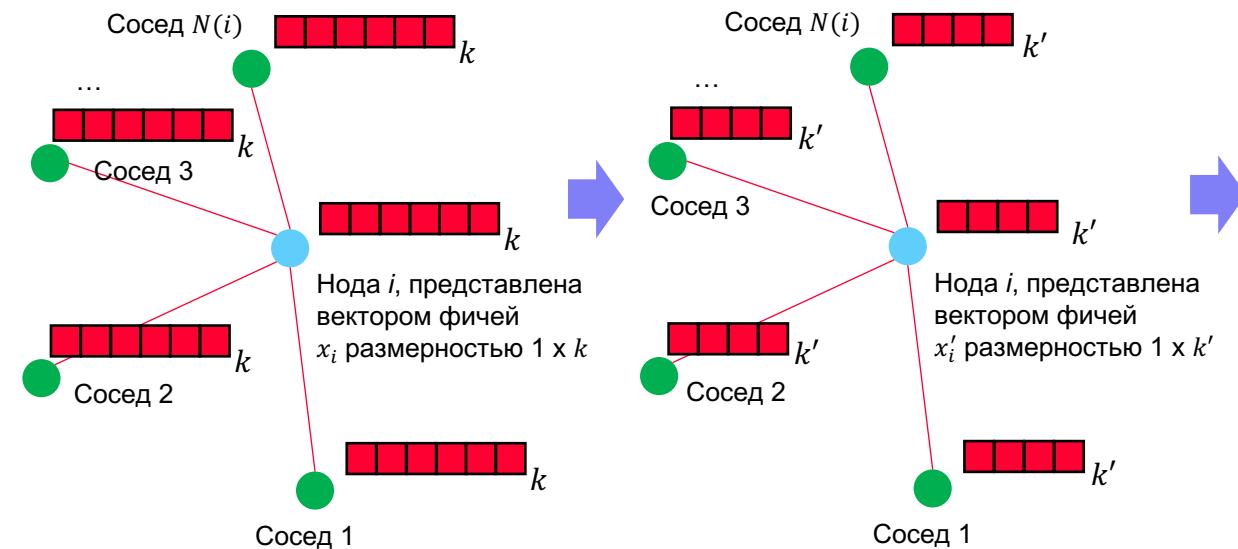
<https://arxiv.org/pdf/1704.01212.pdf>
<https://arxiv.org/pdf/1810.02244.pdf>
<https://arxiv.org/pdf/1709.05584.pdf>

Графовая свертка GraphConv

$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} Wx_j$$

$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} Wx_j$$

W и H имеют размерность $k \times k'$:
 $(1 \times k) * (k \times k') = 1 \times k'$



https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.conv.GraphConv.html

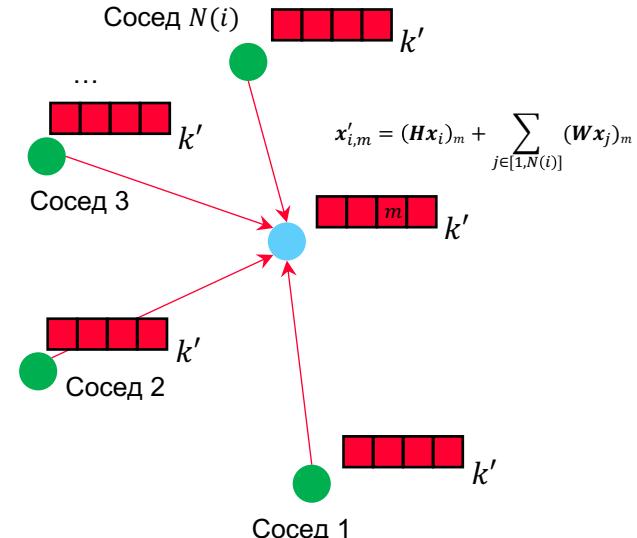
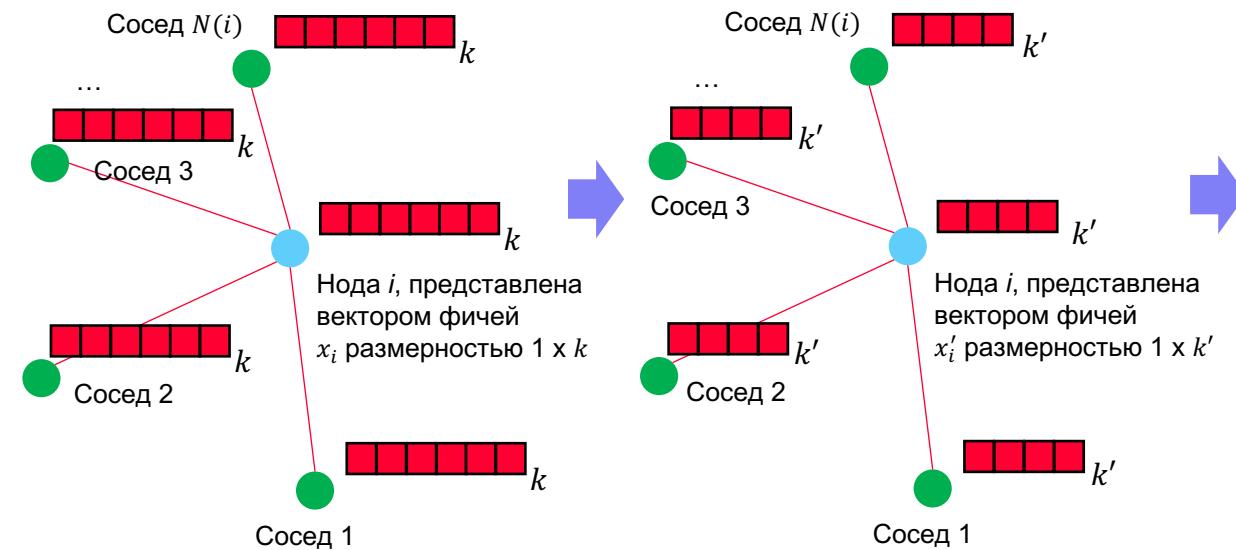
<https://arxiv.org/pdf/1704.01212.pdf>
<https://arxiv.org/pdf/1810.02244.pdf>
<https://arxiv.org/pdf/1709.05584.pdf>

Графовая свертка GraphConv

$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} Wx_j$$

$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} Wx_j$$

W и H имеют размерность $k \times k'$:
 $(1 \times k) * (k \times k') = 1 \times k'$



https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.conv.GraphConv.html

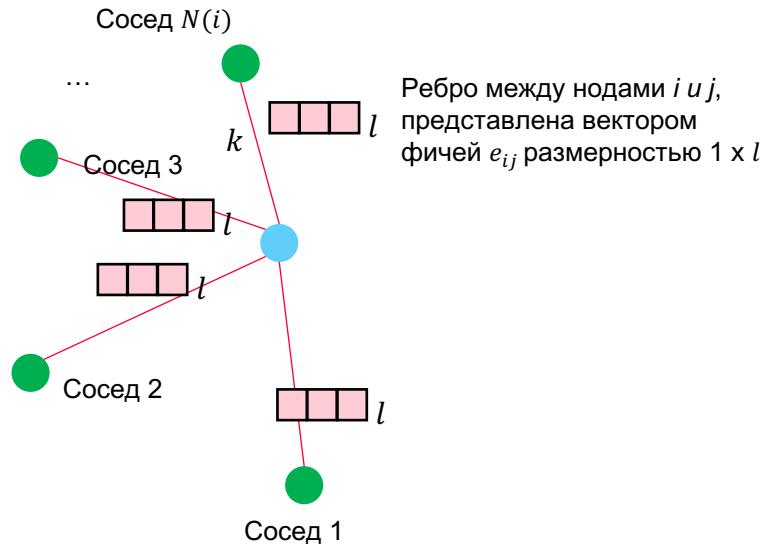
Вместо суммы – другие агрегирующие функции

<https://arxiv.org/pdf/1704.01212.pdf>
<https://arxiv.org/pdf/1810.02244.pdf>
<https://arxiv.org/pdf/1709.05584.pdf>

Графовая свертка NNConv



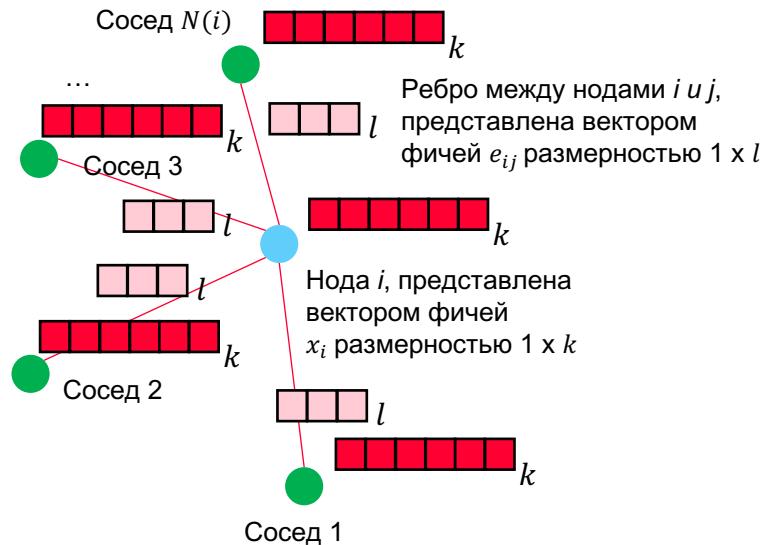
$$\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i + \sum_{j \in [1, N(i)]} \mathbf{x}_j \text{NN}(e_{ij})$$



Графовая свертка NNConv



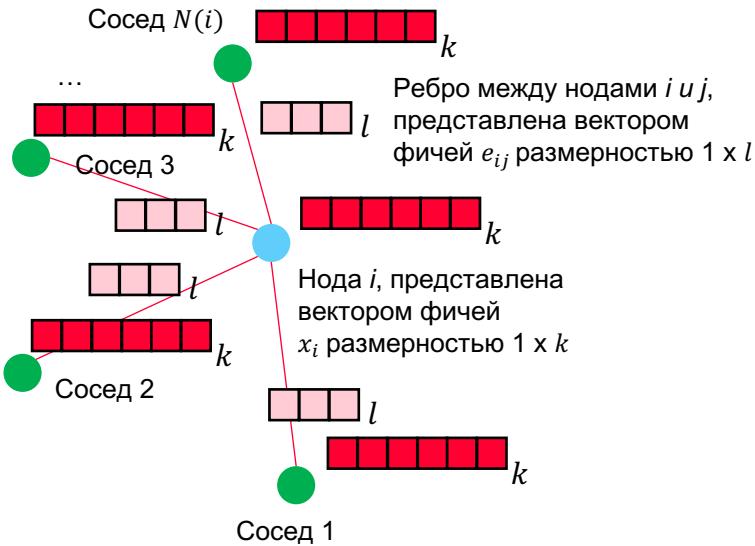
$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} x_j NN(e_{ij})$$



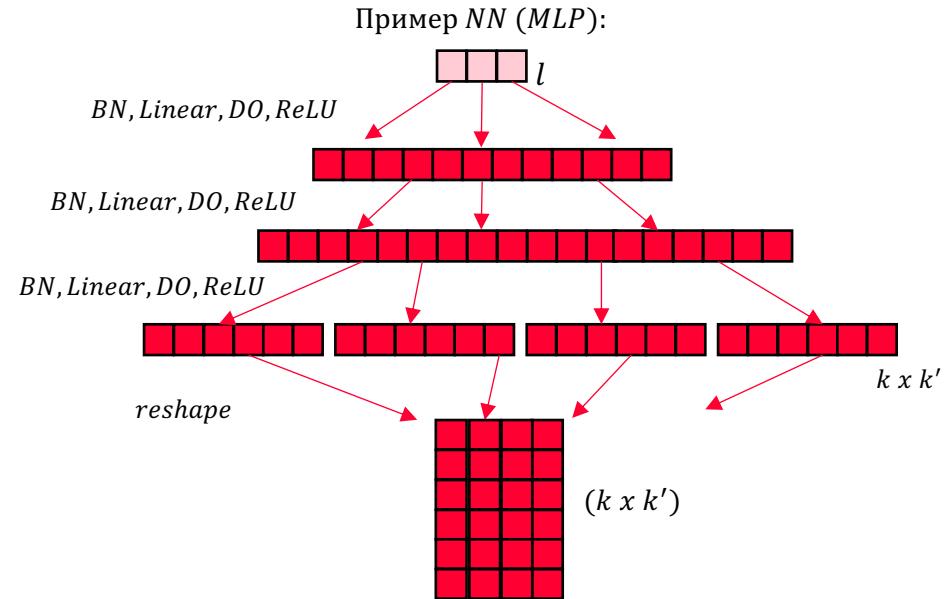
Графовая свертка NNConv



$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} x_j NN(e_{ij})$$



H и выходной слой NN имеют размерность $k \times k'$:
 $(1 \times k)^* (k \times k') = 1 \times k'$
Входной слой NN имеет размерность l
(NN – часто MLP)

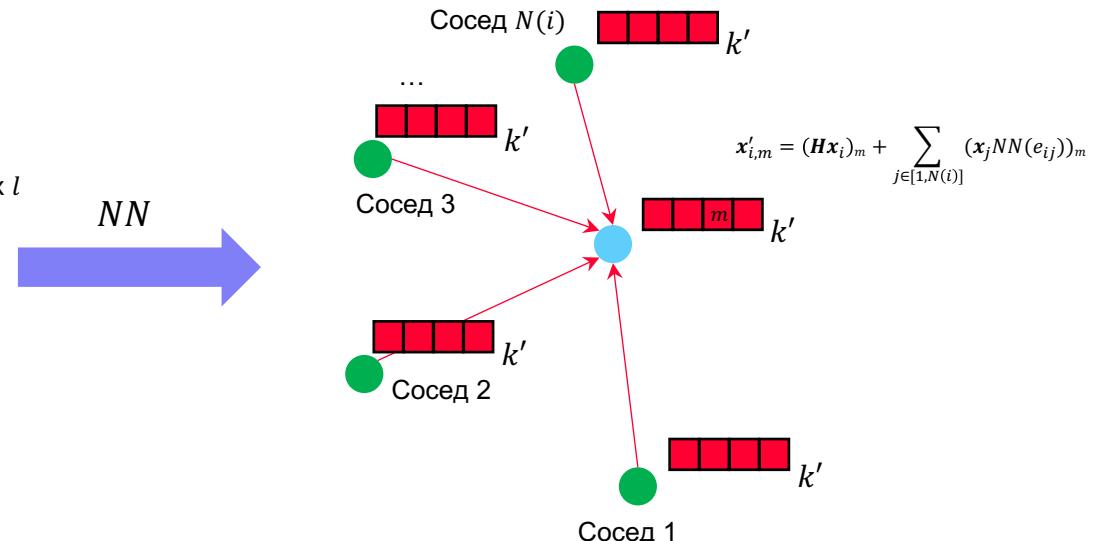
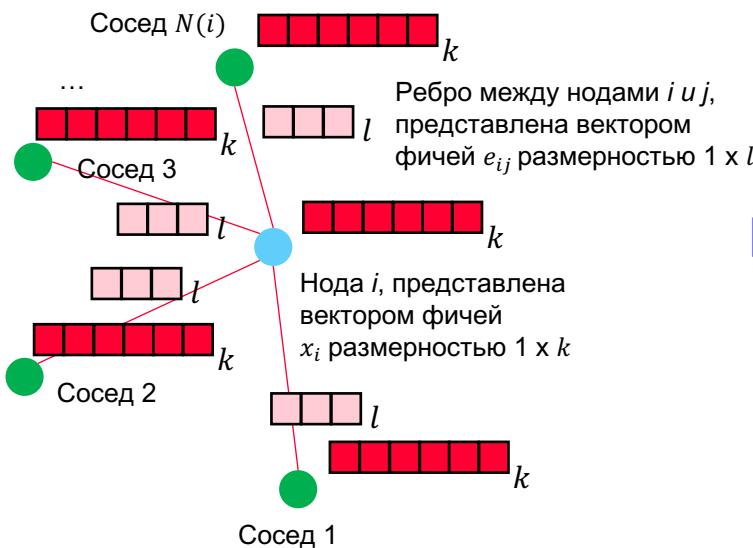


Графовая свертка NNConv

k'

$$x'_i = Hx_i + \sum_{j \in [1, N(i)]} x_j NN(e_{ij})$$

H и выходной слой NN имеют размерность $k \times k'$:
 $(1 \times k) * (k \times k') = 1 \times k'$
Входной слой NN имеет размерность l
(NN – часто MLP)



Где брать
актуальную
информацию:
полезные ресурсы и
персоналии

05

Персоналии



Проф. Михаил Бронштейн
DeepMind Professor of AI,
University of Oxford

<https://www.cs.ox.ac.uk/people/michael.bronstein/>



Prof. Jure Leskovec

<http://snap.stanford.edu/people.html>



Михаил Галкин

<https://migalkin.github.io/>

Курсы по GNN



Penn, Graph Neural Networks

<https://gnn.seas.upenn.edu/lectures/>



Weights and biases, ML with Graphs

<https://wandb.ai/syllogismos/machine-learning-with-graphs/reports/1-Introduction-Structure-of-Graphs--VmlldzozNzU1NDU>



Stanford / Fall 2023, CS224W: Machine Learning with Graphs

<http://web.stanford.edu/class/cs224w/>

Семинары и конференции



Temporal Graph Learning Workshop @ NeurIPS 2023

<https://sites.google.com/view/tglworkshop-2023/home>

New Frontiers in Graph Learning, Workshop @ NeurIPS 2022

<https://nips.cc/virtual/2022/workshop/49963>



STANFORD

Stanford Graph Learning Workshop

<https://snap.stanford.edu/graphlearning-workshop-2023/>



RecSys, графовая секция

<https://recsys.acm.org/recsys23/session-6/>



1st Workshop on Scene Graphs and Graph Representation Learning

https://openaccess.thecvf.com/ICCV2023_workshops/SG2RL



3 семинара по обучению на графах

<https://kdd.org/kdd2023/workshops/>

Ресурсы и датасеты

Почитать



<https://dl.acm.org/doi/10.1145/3580305.3599207>



Papers With Code

<https://paperswithcode.com/>



Robust reasoning over Knowledge Graph with LLM
<https://github.com/RManLuo/Awesome-LLM-KG>



[tg: @graphML](#)

Датасеты



<https://docs.dgl.ai/api/python/dgl.data.html#fraud-dataset>



<https://pytorch-geometric.readthedocs.io/en/latest/modules/datasets.html>