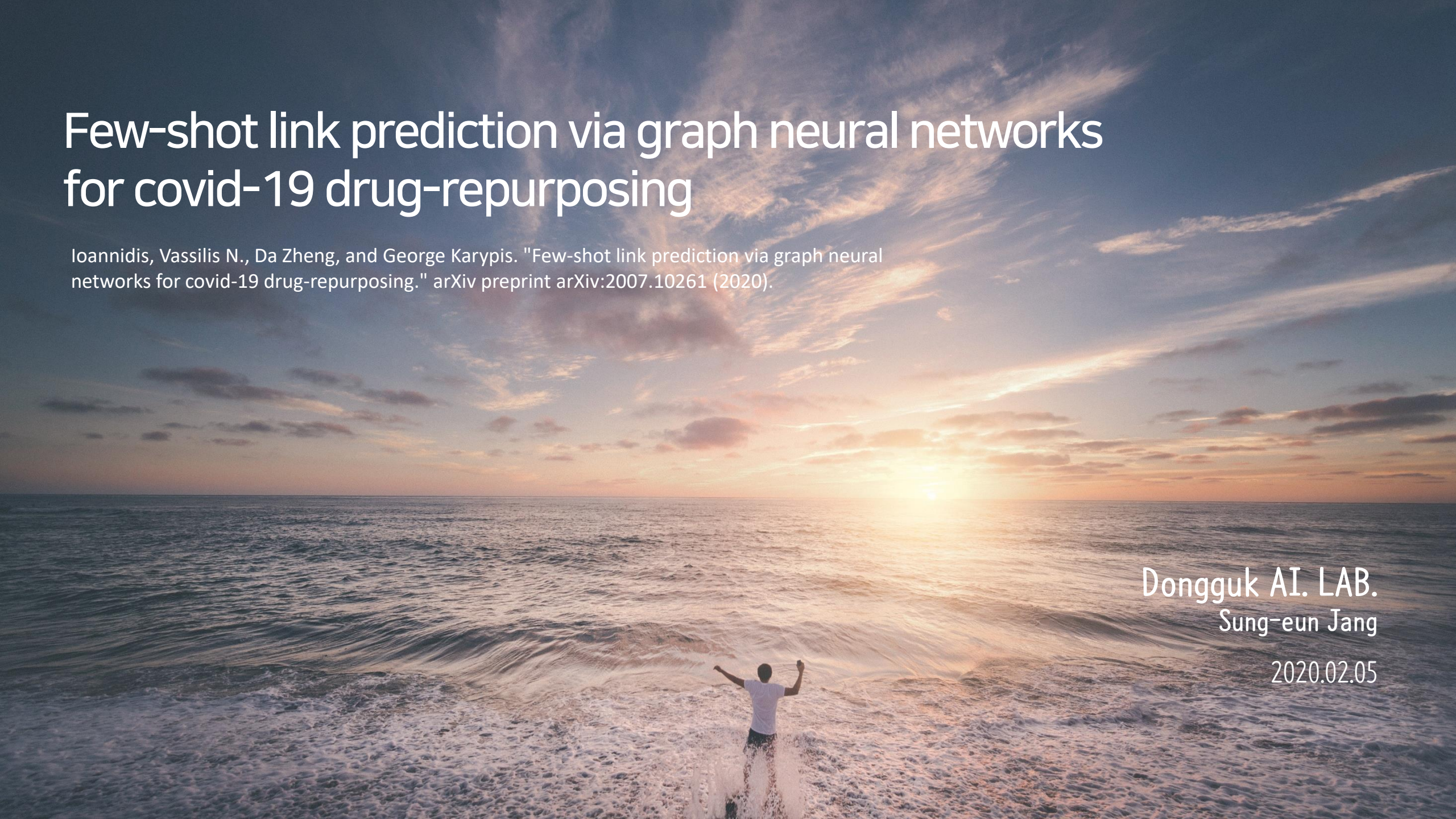
Few-shot link prediction via graph neural networks for covid-19 drug-repurposing

Ioannidis, Vassilis N., Da Zheng, and George Karypis. "Few-shot link prediction via graph neural networks for covid-19 drug-repurposing." arXiv preprint arXiv:2007.10261 (2020).



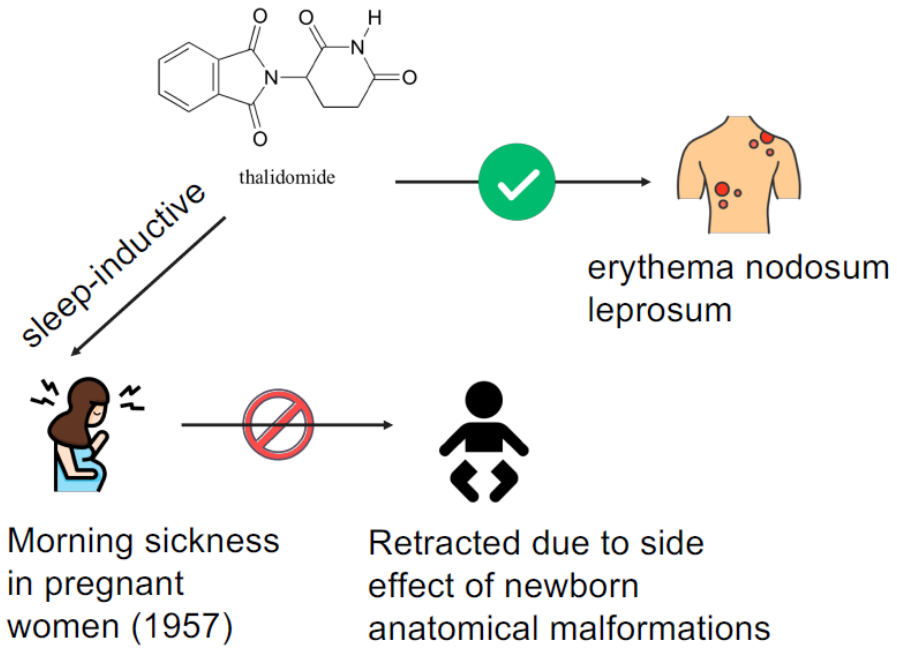
Dongguk AI. LAB.
Sung-eun Jang

2020.02.05

Introduction

Drug repurposing

- Drug repurposing (drug repositioning/reprofiling/re-tasking)



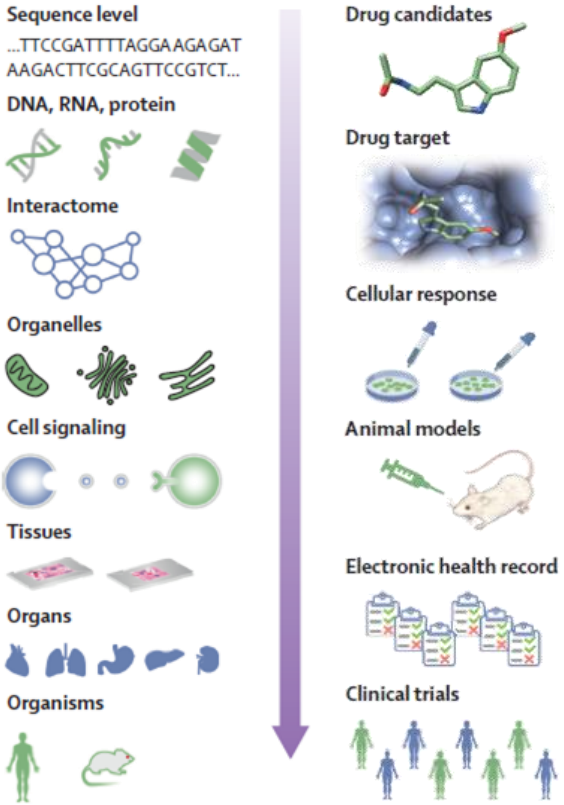
A drug discovery strategy from existing drugs

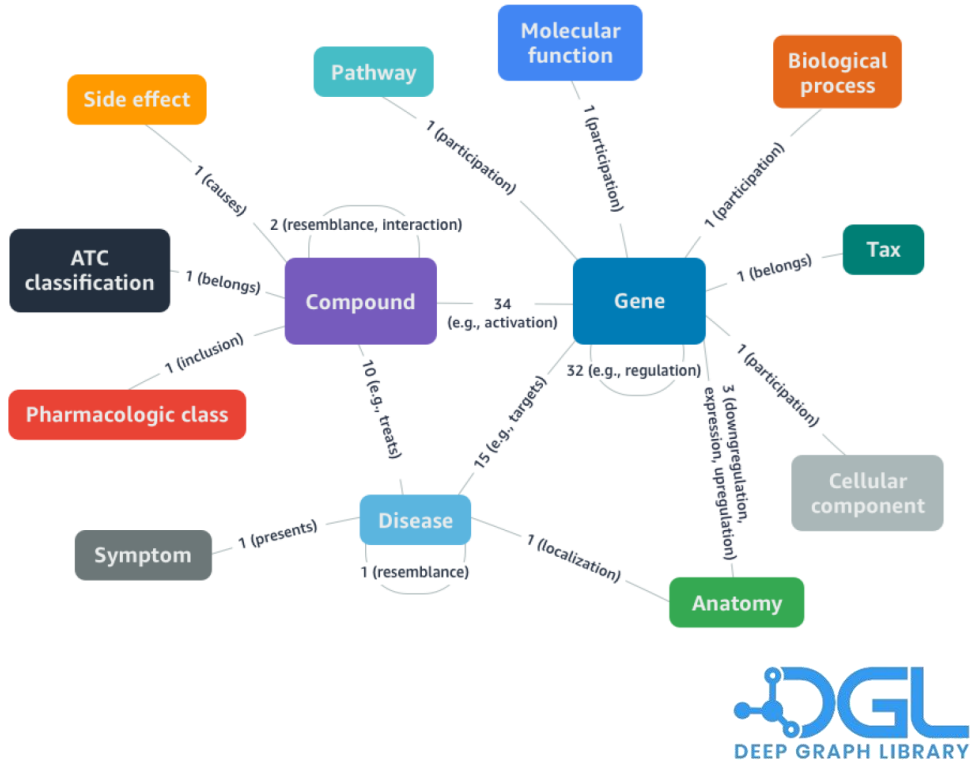
Significantly shortens the time and reduces the cost compared to de novo drug discovery

Introduction

Drug repurposing knowledge graph (DRKG)

- Drug Repurposing Knowledge Graph (DRKG)



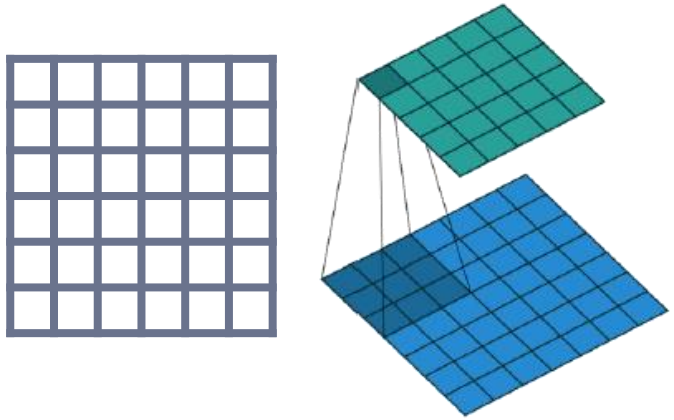


Introduction

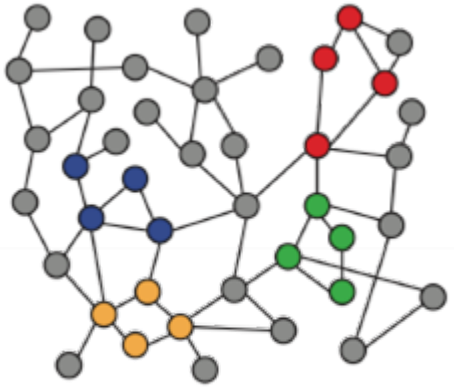
Graph Convolutional Networks

- Graph

image



graph

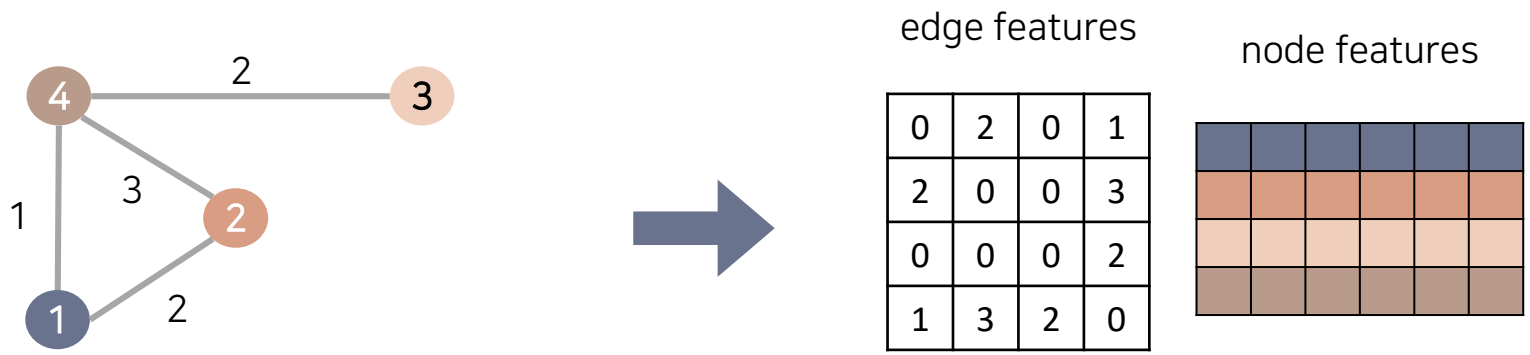


- Arbitrary size
- Complex topological structure
- No fixed node ordering
- Dynamic and multimodal features

Introduction

Graph Convolutional Networks

- Graph



A collection of objects where some pairs of **objects(nodes)** are connected by **links(edges).**

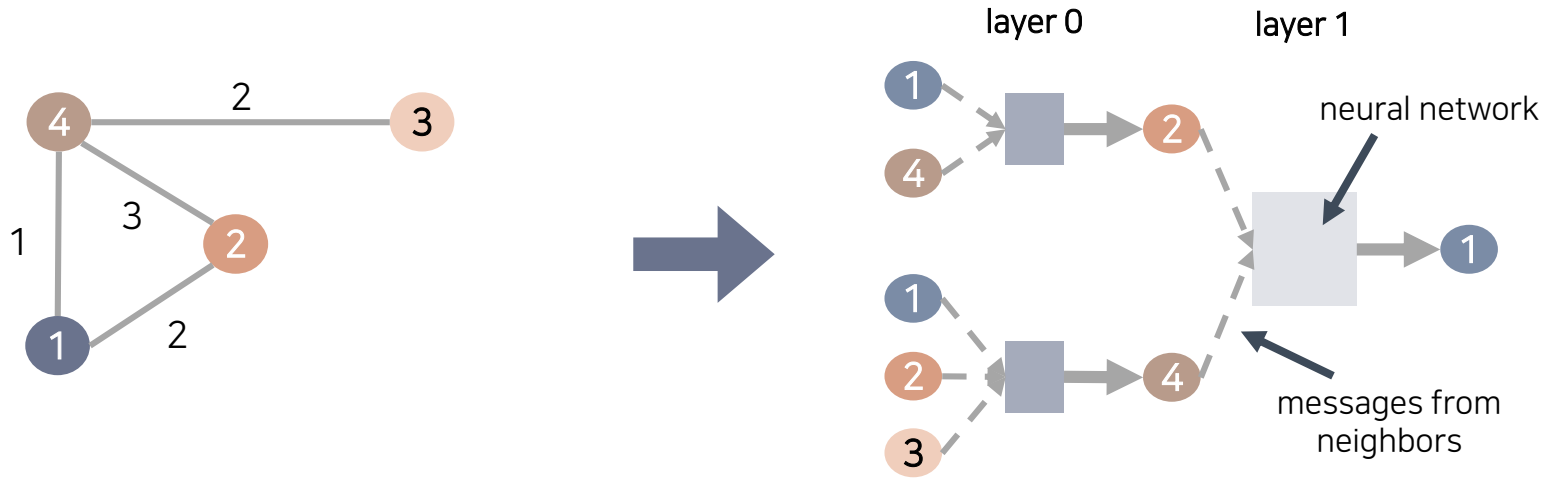
General language for describing **interactions between components** in complex systems.

➔ By explicitly modeling relationships in complex domain(knowledge, text, image…), can achieve better performance in prediction.

Introduction

Graph Convolutional Networks

- Graph Convolution



Connectionist models using **generalized convolution** based propagation rules

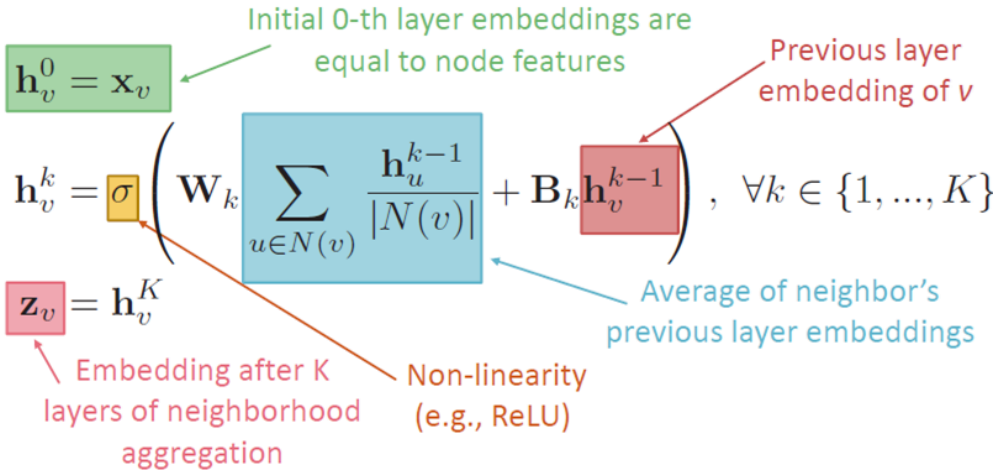
Capture the dependence of graphs via **message passing** between the nodes of graphs.

➔ Represent node information from its neighborhood with arbitrary depth

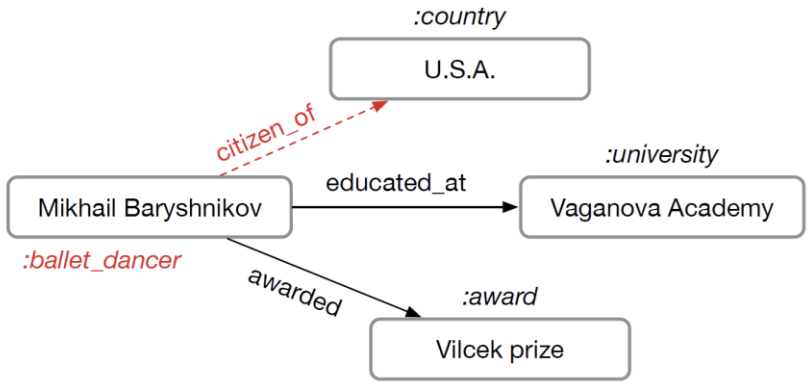
Introduction

Graph Convolutional Networks

- Graph Convolutional Networks



Graph Convolutional Network



$$h_i^{(l+1)} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{|\mathcal{N}_i^r|} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}\right)$$

Relational Graph Convolutional Network

Introduction

Graph Convolutional Networks

- Link prediction in RGCN

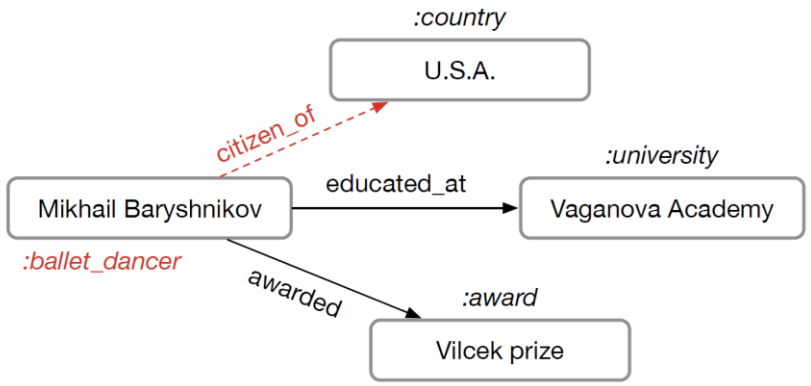
$$\mathbf{h}_r := \mathrm{diag}(\mathbf{r})$$

$$f(n_t, r, n_{t'}) = h_{n_t}^{\mathrm{T}} h_r h_{n_{t'}}$$

$$\mathcal{L} := \log(1 + \exp(-y \times \mathbf{h}_{n_t}^{\top} \mathbf{h}_r \mathbf{h}_{n_{t'}}))$$

Vulnerable when only few training edges are available for a certain relation type

The small number of edges will challenge the learning of the embedding vector r for the rare relation



$$h_i^{(l+1)} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{|\mathcal{N}_i^r|} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}\right)$$

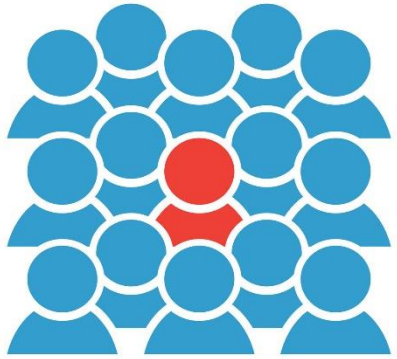
Relational Graph Convolutional Network

Introduction

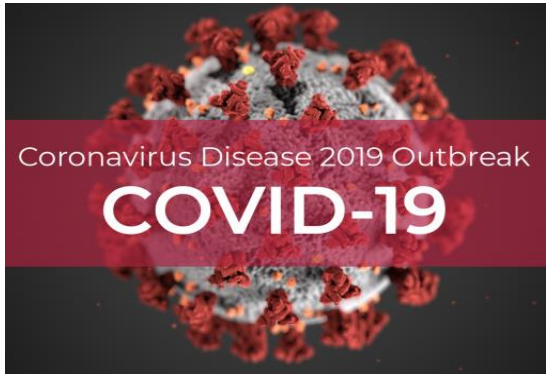
Graph Convolutional Networks

- Limitation

Rare disease



Novel disease



Links to be predicted could belong to rare types such as the case in novel diseases

GCNs are ill-equipped in learning rare link types

➔ motivates the task of **few-shot link prediction**

Method

Proposed method

- Inductive RGCN

$$\mathbf{h}_r := \text{diag}(\mathbf{r}) \quad \Rightarrow \quad \mathbf{h}_r := \frac{1}{|\mathcal{E}_r|} \sum_{(n_t, n_{t'}) \in \mathcal{E}_r} \sigma(\mathbf{W}_1 \sigma(\mathbf{W}_2(\mathbf{h}_{n_t} \| \mathbf{h}_{n_{t'}}))$$

Relation embedding is calculated as a nonlinear function of the node embedding for all node pairs participating to a certain relation type r

This allows the I-RGCN to learn relation embeddings in an inductive fashion.

➔ Enables the model to generalize to rare or even unseen relations.

Experiments

Dataset

- Relational Graph Neural Networks

IMDB

	Nodes	Edges	Relation types
IMDb	movie : 4,278 director : 2,081 actor : 5,257	17,106	12

DBLP

	Nodes	Edges	Relation types
DBLP	author : 4,057 paper : 14,328 term : 7,723 venue : 20	119,783	12

DRKG

8104 candidate drugs

442 Covid-19 related genes

107 relation type

17 node type

Experiments

Result

- Relational Graph Neural Networks

Table 2. Experiment results (%) on the IMDb dataset for k-shot link prediction.

	MRR				Hit 1				Hit 10			
K	ComplEx	RotatE	RGCN	I-RGCN	ComplEx	RotatE	RGCN	I-RGCN	ComplEx	RotatE	RGCN	I-RGCN
10	6.88	11.80	1.32	**33.56**	1.75	6.74	0.13	**25.32**	14.25	18.35	1.07	**53.55**
50	8.48	12.56	15.26	**53.24**	3.34	7.76	7.42	**45.14**	15.70	19.12	28.88	**69.32**
100	8.61	12.57	18.78	**53.63**	3.44	7.86	9.59	**40.27**	15.71	18.40	36.84	**77.38**
1000	68.37	70.09	95.23	**96.06**	65.48	67.52	91.72	**93.56**	72.23	73.50	99.72	**99.81**

Table 3. Experiment results (%) on the DBLP dataset for k-shot link prediction.

	MRR				Hit 1				Hit 10			
K	ComplEx	RotatE	RGCN	I-RGCN	ComplEx	RotatE	RGCN	I-RGCN	ComplEx	RotatE	RGCN	I-RGCN
10	5.97	7.41	7.42	**27.95**	1.37	2.13	2.95	**13.78**	11.81	15.36	12.51	**60.51**
50	6.32	7.87	17.25	**83.42**	1.55	2.26	10.95	**72.54**	12.59	16.64	26.84	**96.82**
100	7.24	10.66	32.45	**90.00**	1.99	04.43	23.46	**85.27**	14.96	20.92	49.85	**97.61**
1000	36.56	46.51	91.34	**96.82**	30.62	39.83	86.43	**94.41**	46.45	59.27	98.59	**99.81**

Experiments

Result

- Link prediction

Table 4. Experiment results (%) on the IMDb dataset for link prediction.

Metrics	MRR				Hit 1				Hit 10			
Training links	ComplEx	RotatE	RGCN	I-RGCN	ComplEx	RotatE	RGCN	I-RGCN	ComplEx	RotatE	RGCN	I-RGCN
95%	94.15	93.97	93.38	**95.12**	**93.75**	93.48	89.31	92.75	94.74	94.56	**99.29**	98.24
90%	88.87	88.89	89.30	**93.98**	88.25	87.87	83.45	**91.52**	89.74	90.35	97.66	**98.12**
80%	78.55	78.90	83.46	**90.03**	77.45	76.83	76.59	**86.70**	79.96	81.94	94.72	**95.93**
70%	69.59	69.56	82.73	**87.00**	67.98	66.73	76.09	**82.95**	71.89	73.49	93.76	**94.07**
60%	60.40	60.90	78.16	**81.53**	58.49	57.60	70.63	**77.01**	62.92	65.57	**91.14**	90.27

Experiments

Result

- Drug inhibits gene scores

Use corona-virus related diseases* as target diseases and 442 Covid-19 related genes from the relations

Predict links among gene entities associated with the target disease and drug entities

Check hit rate of the overlap among the top 100 predicted drugs and the drugs used in clinical trials per gene

Several of the widely used drugs in clinical trials appear high on the predicted list, and that I-RGCN shows a higher hit rate

*SARS, MERS and SARS-COV2 (their functionality is similar)

Table 5. Drug inhibits gene scores for Covid-19. Note that a random classifier will result to approximately 5.3 per drug. This suggests that the reported predictions are significantly better than random.

I-RGCN		RGCN	
Drug name	# hits	Drug name	# hits
Dexamethasone	240	Chloroquine	69
Ribavirin	142	Colchicine	41
Colchicine	128	Tetrandrine	40
Chloroquine	115	Oseltamivir	37
Methylprednisolone	86	Azithromycin	36
Tofacitinib	75	Tofacitinib	33
Thalidomide	70	Ribavirin	32
Losartan	64	Methylprednisolone	30
Hydroxychloroquine	48	Deferoxamine	30
Oseltamivir	46	Thalidomide	25
Deferoxamine	34	Dexamethasone	24
Ruxolitinib	23	Bevacizumab	21
Azithromycin	23	Hydroxychloroquine	19
Nivolumab	11	Losartan	19
Tradipitant	11	Ruxolitinib	13
Bevacizumab	10	Eculizumab	12
Eculizumab	7	Tocilizumab	11
Baricitinib	6	Anakinra	11
Sarilumab	6	Sarilumab	8
Tetrandrine	6	Nivolumab	6

Conclusion

- I-RGCN consistently outperforms baseline models in the IMDB and DBLP datasets for few-shot link prediction
- Also formulate the Covid-19 drug-repurposing task as a link prediction over the DRKG
- I-RGCN successfully identifies a subset of clinical trial drugs for Covid-19 and can be used to assist researchers and prioritize existing drugs for further investigation in the Covid-19 treatment.

References

- Ioannidis, Vassilis N., Da Zheng, and George Karypis. "Few-shot link prediction via graph neural networks for covid-19 drug-repurposing." arXiv preprint arXiv:2007.10261 (2020).
- Zhou, Yadi, et al. "Artificial intelligence in COVID-19 drug repurposing." The Lancet Digital Health (2020).
- Tayarani-N, Mohammad-H. "Applications of artificial intelligence in battling against Covid-19: a literature review." Chaos, Solitons & Fractals (2020): 110338.
- Schlichtkrull, Michael, et al. "Modeling relational data with graph convolutional networks." European semantic web conference. Springer, Cham, 2018.
- Chen, Mingyang, et al. "Meta relational learning for few-shot link prediction in knowledge graphs." arXiv preprint arXiv:1909.01515 (2019).
- https://github.com/gnn4dr/DRKG
- https://www.dgl.ai/news/2020/06/09/covid.html


Thank you