

Predicting Adverse Drug Effects: A Heterogeneous Graph Convolution Network with a Multi-layer Perceptron Approachs

Yintzer (Jerry) Shih

National Chung Hsing University

Taichung, Taiwan



In collaboration with YS Chen (Dept. of Nephrology, Taichung Tzu Chi Hospital)

CS Chien (NCHU) & CS Tsai (NCHU)

About me

POSITION:

- Professor of Applied Math, NCHU
- Dean, College of Science, NCHU
- Director, Big Data Center, NCHU
- Director, AI & Data Science MS Program, NCHU

EDUCATION: Ph.D. in 1998
University of Maryland, College Park

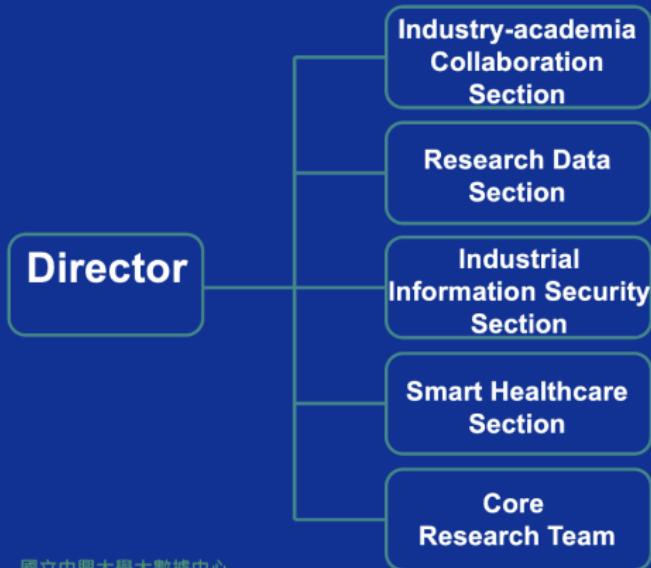
RESEARCH INTERESTS:

Numerical Modeling, Imaging Process,
Data Analysis, Scientific Computing

EXPERIENCE

- Director, Alumni Center, NCHU
- Department Chair of Applied Mathematics, NCHU
Director of Graduate Institute of Statistics, NCHU
- Senior Software Engineer
Aviation Weather Center
NCEP/NOAA, Kansas City, MO
- Senior Software Analyst
Medical Optical Imaging,
Charlotte, NC

The Big Data Center of NCHU



Purpose

- Enhance interdisciplinary teaching research
- Innovation of data analysis technology
- Further industry-academia cooperation
- Talent cultivation

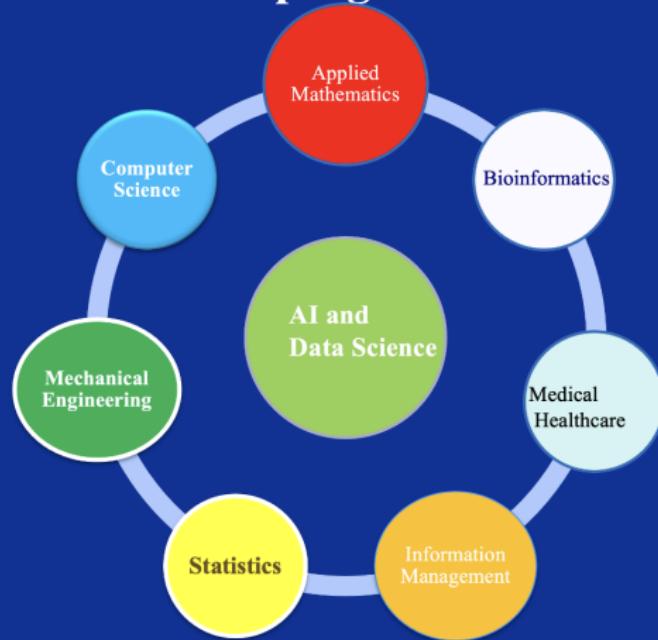
Focus

- Smart Agriculture
- Smart Healthcare
- Information Security



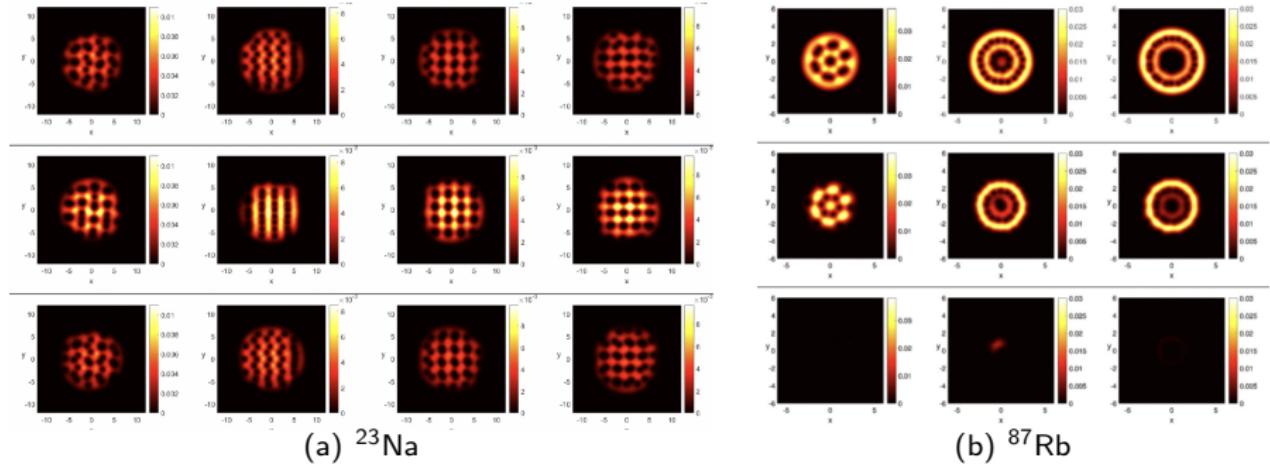
國立中興大學大數據中心
Big Data Center, NCHU

Research topics in AI master program



Recent work in BEC

A visualization the physical phenomena of rotating spin-1 Bose–Einstein condensates (BEC)[1].



1. Sriburadet, Shih*, Jeng, Hsueh, and Chien, A numerical scheme for the ground state of rotating spin-1 Bose-Einstein condensates, *Scientific Reports*, 11(2021) 22801.



A note on stochastic polynomial chaos expansions for uncertain volatility and Asian option pricing[☆]

Y.-T. Lin^a, Y.-T. Shih^{b,*}, C.-S. Chien^b, Q. Sheng^c

^aInstitute of Mathematics, Academia Sinica, Taipei 10617, Taiwan

^bDepartment of Applied Mathematics, National Chung Hsing University, Taichung 40227, Taiwan

^cDepartment of Mathematics, Baylor University, Waco, TX 76798-7328, USA

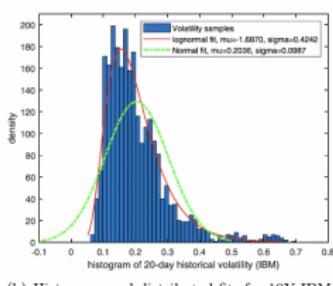
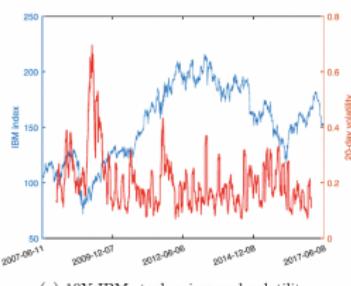


Table 4

The computational times and the maximum-norm errors by gPC and aPC for Experiment with those from the MCS.

N	5	6	7
Speed	Elapsed time (sec.)	gPC aPC	67.75 77.32
	MCS : gPC (elapsed time)	90.04 97.04	120.99 129.35
Error	Mean	gPC aPC	9.13E5 8.24E5
	Variance	gPC aPC	1.27E5 7.44E6
		5020:1	387:1

Outline

- Objective
- Related works
- Methodology
- Experiments & Results
- Conclusion



BMJ Yale

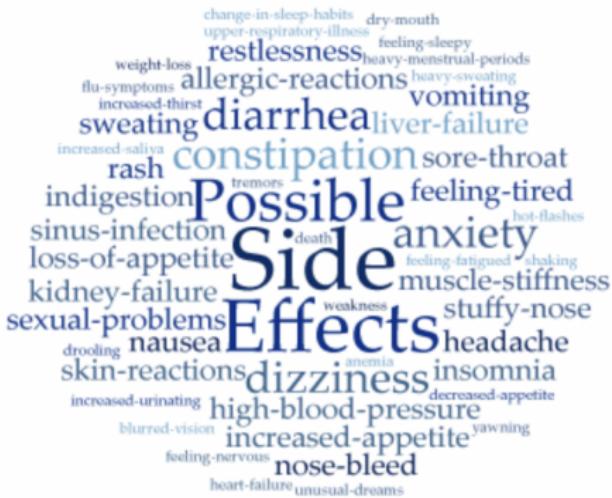
Predicting Adverse Drug Effects:A Heterogeneous Graph Convolution Network with a Multi-layer Perceptron Approach

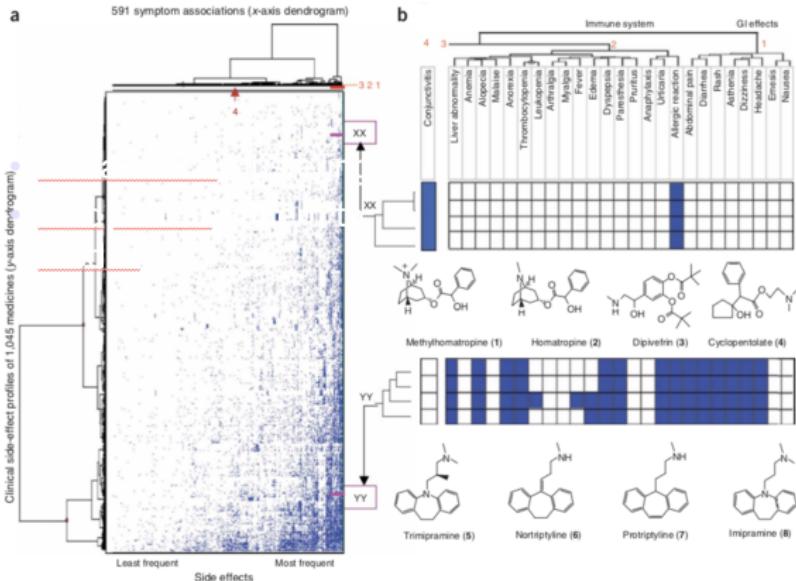
Y.-H. Chen, Y.-T. Shih, C.-S. Chien, C.-S. Tsai
doi: <https://doi.org/10.1101/2022.03.22.22272749>

Motivation

- Unexpected responses to drugs beyond their anticipated therapeutic effects
[Edwards and Aronson \(2000\) Lancet](#)
- Affect quality of life and even death, not alone economic loss
[Scheiber et al. \(2009\) J. Chem. Inf. Model](#)
- Large number of hospital admissions, 42% of which could be prevented
[Gurwitz \(2003\) JAMA](#)
- Drugs approved by FDA are sometimes recalled because of side effects
[Veeran et al. \(2013\) J. Pharm. Health Serv. Res.](#)

Google trends about Side Effects (Adverse Reactions): fda.gov/drugs/





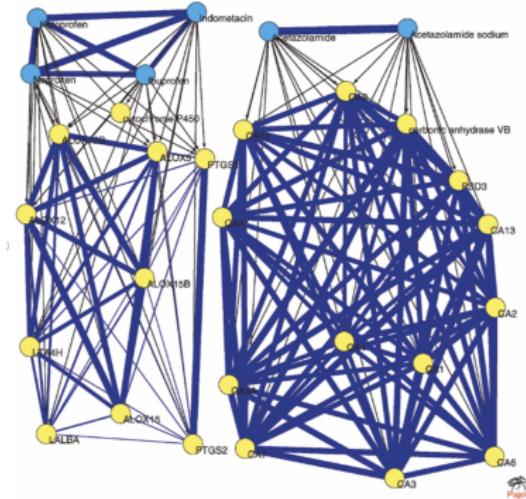
Fliri et al. [1] described the translation of adverse effect data derived from drug labels into effect spectra and provided a hierarchical clustering finding the similar pattern to the side effects.

1. Fliri. et al., Analysis of drug induced-effect patterns to link structure and side effects of medicine, *Nature Chem Bio*, 2005.

Previous studies in side effects

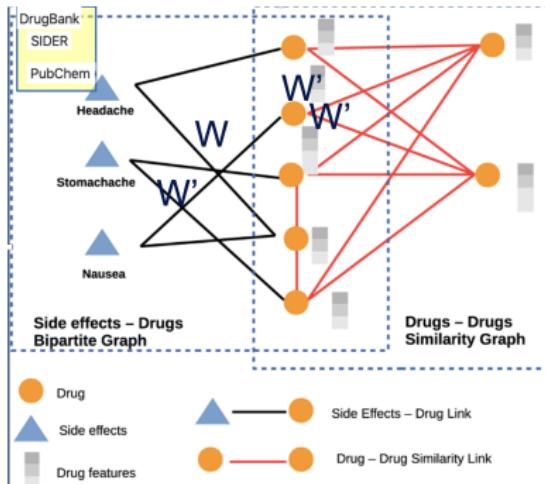
1. Observe the side effects of drugs that appear in the cell or animal studies
(1) Bailey *et al.*, ATLA 2014, (2) Gmez-Lechn *et al.*, Expert Opin. Drug Metab. Toxicol. 2014
2. Datasets integration, including protein function, genomic profiles
Chiu *et al.*, BMC Med. Genomics, 2019; Timilsina *et al.*, Sci. Rep., 2019
3. Graph neural network applied in disease prediction and classification
Zhou, *et al.*, J. Am. Heart Assoc., 2020

Similarity and networks



- Drug (blue)-Protein (yellow) networks
- Thickness of lines: degree of similarity
- Similar drugs often target on the similar proteins
- Predict missing link with random walk (previously not found)

Chen et al. (2012) Drug-target interaction prediction by random walk on the heterogeneous network



[1] Created a semantic similarity network of drugs by using textual embedding, and predict new side effect in side effect-drug network by NMF via a diffusion-based method.

1. Timilsina et al. (2019) Discovering links between side effects and drugs using a diffusion based method, Sci. Rep.

Datasets

The database resources:

- 1 The Side Effect Resource (SIDER) <http://sideeffects.embl.de/>
- 2 OFF label drug SIDE effectS (OFFSIDES) <https://tatonettilab.org/offsides/>
- 3 US FDA Adverse Event Reporting System (FAERS) <https://open.fda.gov/data/faers/>

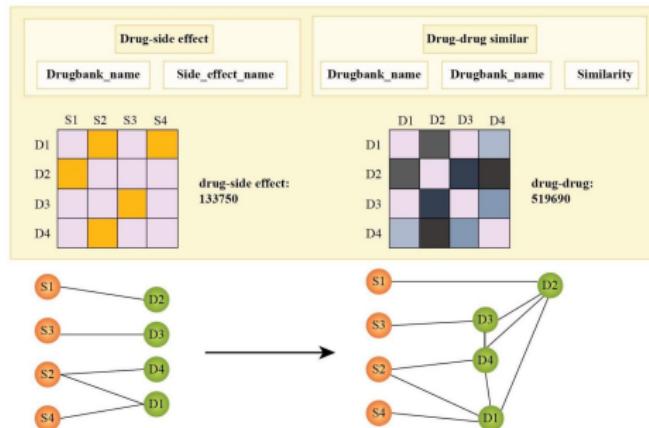
Dataset	# of drugs	# of ADE	# of pairs
SIDER	1020	5599	133,750
OFFSIDES	2738	14544	3,206,558
FAERS	4245	17671	3,766,382

The date on the download version: Sider (2018.11), Offsides (2019.11), Fares (2019.09)

Similarity and Networks

Preprocess the datasets by the similarity scores derived from the **word2vec** package [1].

- The word representations are derived from the large corpus of biomedical and general-domain texts.
- The process of combining SIDER dataset is illustrated in



1. Mikolov, et al. (2013). "Efficient Estimation of Word Representations in Vector Space". [arXiv:1301.3781](https://arxiv.org/abs/1301.3781)

Visualization the nodes: drugs and side effects

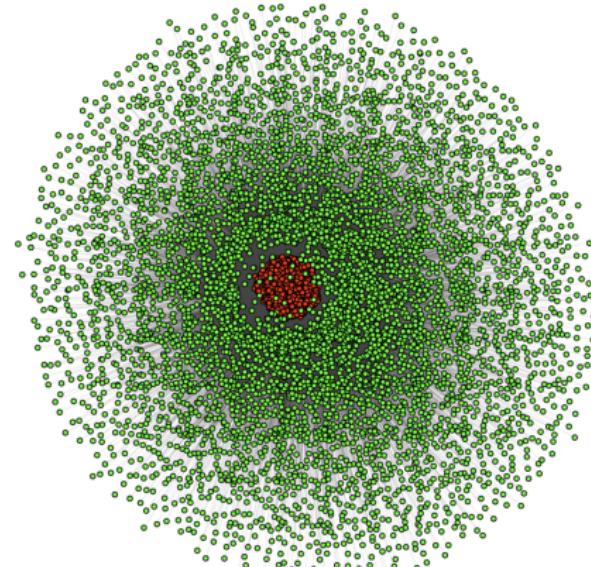


Figure: Red: drug nodes vs. Green: side effect nodes

Link Prediction

1 Similarity-based link prediction in social networks

- Local approach: Jaccard index, Preferential attachment (PA) index, Adamic Adar (AA) index, Resource allocation (RA) index
- Global approach: Katz, Personalized PageRank (PPR)

2 Matrix factorization: NMF-based methods

3 Machine learning base: GNN

Non-negative matrix factorization (NMF)

- Let the bipartite matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{m \times n}$
the elements a_{ij} represent the links of side effects and drugs
- The NMF decomposes \mathbf{A} into two low rank, latent feature matrices $\mathbf{W} = (w_{ij}) \in \mathbb{R}^{m \times k}$ and $\mathbf{H} = (h_{jl}) \in \mathbb{R}^{k \times n}$, s.t.

$$\mathbf{A} \approx \mathbf{WH}, \text{ for small } k > 0.$$

The MNF solves the minimization problem of for all $\mathbf{W}, \mathbf{H} \geq 0$

$$\min_k \|\mathbf{A} - \mathbf{WH}\|_F,$$

and the recovered elements a_{ij} contain the potential links.

NMF with heat diffusion (NFMHD) [1]

- Consider an undirected network graph $G = (\mathcal{V}, \mathcal{E})$, $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$, $\mathcal{E} = \{\overrightarrow{v_i v_j}\}$, and the probability space $\mathcal{P} = \{\omega_{ij}\}$.
- Let $\mathbf{f}(t)$ be the link vector connecting the nodes v_i and v_j at time t , and the discrete approximation for the heat diffusion flow is

$$\mathbf{f}(t) = e^{\alpha t \mathbf{D}} \mathbf{f}(0), \quad \text{for node } v_j$$

α the heat coef., $\mathbf{D} = (d_{ij})$ the diffusion matrix

$$d_{ij} = \begin{cases} -\frac{\tau_i}{m_i} \sum_k \omega_{ik}, & j = i \\ \frac{\omega_{ji}}{m_j}, & \overrightarrow{v_j v_i} \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \quad i, j = 1, 2, \dots, n$$

the degrees $m_j \geq 1$, and τ_j is the flag index, $0 \leq \tau_j \leq 1$.

- The diffusion matrix is derived from the drug-drug similarity matrix.

1. Timilsina et al. (2019) Discovering links between side effects and drugs using a diffusion based method, Sci.

Rep.

Let Γ_u the set of neighbors of node u .

- **Adamic Adar (AA)**: the probability of two linking nodes v_i and v_j by

$$\text{AA}(v_i, v_j) = \sum_{u \in \Gamma_{v_i} \cap \Gamma_{v_j}} \frac{1}{\log |\Gamma_u|}.$$

- **Resource allocation (RA)**: predict the links by $\text{RA}(v_i, v_j) = \sum_{u \in \Gamma_{v_i} \cap \Gamma_{v_j}} \frac{1}{|\Gamma_u|}$.
- **Katz**: for the probability of existing links depends on the number of paths,

$$\text{Katz}(v_i, v_j) = \sum_{l=1}^{\infty} \beta^l \cdot [\text{paths}_{v_i, v_j}^{<l>}],$$

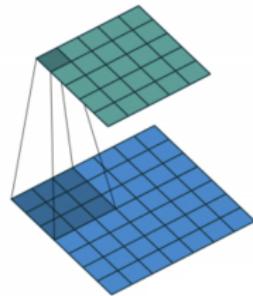
β the attenuation factor, $\text{paths}_{v_i, v_j}^{<l>}$ the degree of connection by the walk length l .

- **Personalized PageRank (PPR)**: likelihood of random walk or random neighbor.

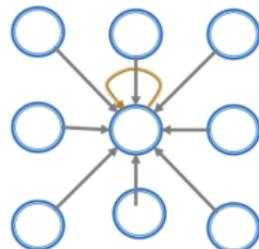
The GCN

What is a GCN?

- 1 GNNs are Neural Networks that operate on graph-structured data.
- 2 To capture correlation inside the sample & across the sample.
- 3 GCNs are the generalized version of CNNs where the numbers of nodes connections vary and the nodes are unordered.



Image



Graph

The GCN

The drug information structure is of network type, such as protein-protein, drug-target, or drug-genome interaction, so the biomedical data is particularly suitable for the network prediction methods [1,2].

1. Mizutani et al, (2012) Relating drugprotein interaction network with drug side effects, Bioinformatics.
2. Turanli et al. (2018) A network-based cancer drug discovery: from integrated multi-omics approaches to precision medicine, Curr. Pharm. Des.

The GCN

The spatial-based method of GCN is similar to the convolutional neural network (CNN), and the features of the links to given node v can be aggregated and updated, and for $k = 1, 2, \dots, K$

$$\mathbf{h}_v^k = \sigma(\mathbf{W}_k \sum_{u \in \mathcal{N}(v)} \frac{\mathbf{h}_u^{k-1}}{|\mathcal{N}(v)|} + \mathbf{B}_k \mathbf{h}_v^{k-1}),$$

where \mathbf{h}_v^k the feature of the link, weighting matrices \mathbf{W}_k and \mathbf{B}_k for feature transformation and message passing, and σ is a differentiable aggregator function.

The embedding generation of GCN

The GCN propagate information between different layers and search depths of data.

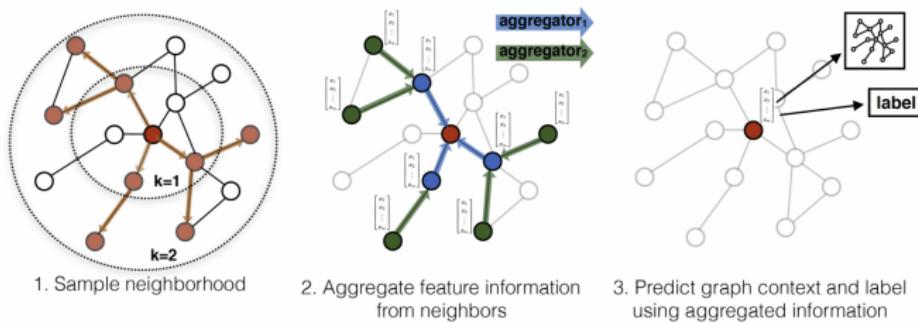


Figure: Visual illustration of the GCN sample and aggregate approach from (2018) Hamilton, Ying, and Leskovec, Inductive Representation Learning on Large Graphs, arXiv.

The GCNMLP

The GCNMLP can be generalized as

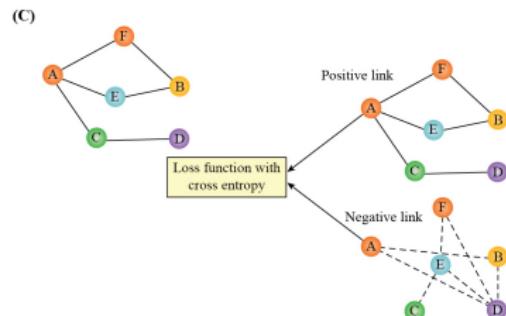
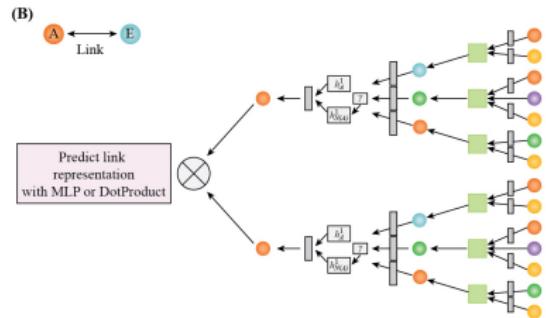
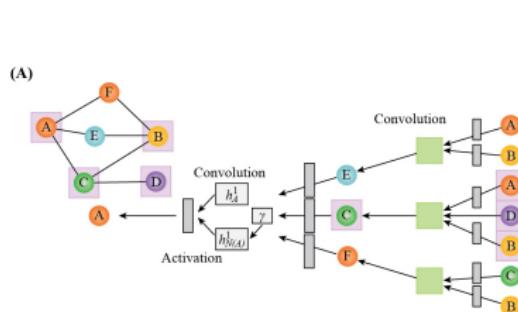
$$\mathbf{h}_v^k = \text{MLP} \left\{ \text{concat}(\mathbf{h}_v^{k-1}, \sum_{u \in \mathcal{N}(v)} \frac{\mathbf{h}_u^{k-1}}{|\mathcal{N}(v)|} + \max_{u \in \mathcal{N}(v)} \mathbf{h}_u^{k-1}) \right\} \quad (1)$$

MLP: function of a multi-layer perceptron

'concat' : the concatenate function for the node features.

The GCNMLP predicts the link between two adjacent nodes with a probability from the node's features.

The architecture of GCNMLP



Theoretical analysis

Conjecture.

Let $\mathbf{h}_v^0 \in \mathbf{U}$ denote an initial feature inputs on graph $G = (\mathcal{V}, \mathcal{E})$, \mathbf{U} is a compact subset of \mathbf{R}^d . The sequence $\{\mathbf{h}_v^k\}$ is generated by the algorithm of the GCNMLP. Then there exist a positive constant ϵ and $K > 0$ so that

$$\|\mathbf{h}_v^k - \mathbf{h}_v^{k-1}\|_2 \leq \epsilon \text{ for } k > K.$$

Proof. See [1] for details.

1. Hamilton, Ying, and Leskovec (2018) Inductive Representation Learning on Large Graphs, arXiv.

Database and setup

- Three datasets SIDER, OFFSIDES, and FAERS for drugs link prediction.
- The dataset was shuffled before being split into 90% for training and 10% for testing 10-fold cross-validation procedure to test our predictive model.
- All experiments were run 10 times including the mean results and the standard deviations. We divide the data into training data and validation data.
- Select dim # = 200 in word2vec; PC with Intel core i5 6-core, 48GB RAM; Python 3.6

Evaluation scales

The four outcomes can be formulated in a confusion matrix: True positive (TP), False negative (FN), False positive (FP), and True negative (TN).

- Precision = $\frac{TP}{TP+FP}$, Recall = $\frac{TP}{TP+FN}$, F1 = $\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
- AUROC (the area under Receiver Operating Characteristic curve), ROC curve is created by plotting the TP rate vs. the FP rate.
- AUPR (the area under the Precision-Recall Curve).

Different dropout rates for SIDER

A trend of deteriorating performance of GCNMLP when the dropout rate was increasing.

Dropout	Precision	Recall	F1 score	AUROC	AUPR
0%	0.865±0.003	0.865±0.003	0.865±0.003	0.939±0.002	0.941±0.002
10%	0.862±0.003	0.858±0.003	0.858±0.003	0.934±0.002	0.932±0.003
30%	0.857±0.003	0.853±0.003	0.853±0.003	0.926±0.002	0.926±0.002
50%	0.848±0.003	0.843±0.003	0.842±0.003	0.917±0.002	0.918±0.003

Different sizes of the training set

- The training/testing sets were split from 90%-10% to 50%-50%, and show the stability of the GCNMLP.

Training	Precision	Recall	F1 score	AUROC	AUPR
90%	0.865±0.003	0.865±0.003	0.865±0.003	0.939±0.002	0.941±0.002
80%	0.864±0.004	0.860±0.004	0.860±0.004	0.937±0.002	0.937±0.004
50%	0.864±0.000	0.860±0.000	0.860±0.000	0.937±0.000	0.938±0.0000

Comparing different hop numbers for GCNMLP

- The GCNMLP performs the best with the hop (aggregating) number being 3.
- The performance of the GCNMLP are almost the same on all the evaluation scales except the AUPR.
- But the deviation of the latter is greater than that of the former.

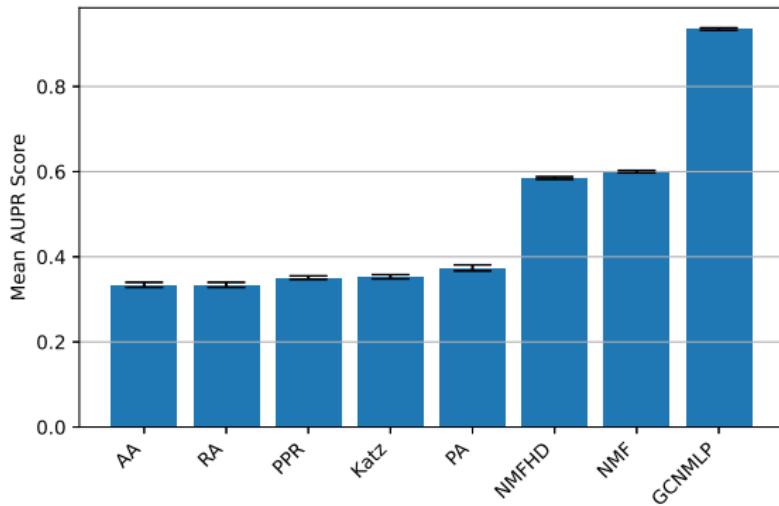
hop #	Precision	Recall	F1 score	AUROC	AUPR
$K = 2$	0.863 ± 0.003	0.859 ± 0.003	0.859 ± 0.003	0.935 ± 0.003	0.935 ± 0.003
$K = 3$	0.865 ± 0.003	0.865 ± 0.003	0.865 ± 0.003	0.939 ± 0.002	0.941 ± 0.002
$K = 4$	0.864 ± 0.002	0.864 ± 0.002	0.864 ± 0.002	0.938 ± 0.002	0.939 ± 0.003
$K = 5$	0.864 ± 0.003	0.864 ± 0.003	0.864 ± 0.003	0.938 ± 0.003	0.940 ± 0.004

Comparing the performance of various algorithms

- GCNMLP performs best among these methods on Recall, F1 score, and AUPR.
- The NMF-based methods are superior to the heuristic network link-prediction methods on these three items.
- NMF outperforms the other methods on AUROC, and the NMFHD is superior to the other methods on Precision.

Model	Precision	Recall	F1 score	AUROC	AUPR
GCNMLP	0.865±0.003	0.865±0.003	0.865±0.003	0.939±0.002	0.941±0.00
NMF	0.906±0.003	0.659±0.001	0.726±0.002	0.946±0.001	0.600±0.00
NMFHD	0.930±0.002	0.623±0.002	0.689±0.002	0.943±0.001	0.585±0.00
AA	0.840±0.008	0.525±0.003	0.541±0.006	0.909±0.001	0.334±0.00
RA	0.490±0.006	0.500±0.000	0.494±0.000	0.909±0.001	0.334±0.00
PPR	0.836±0.007	0.542±0.002	0.570±0.003	0.908±0.001	0.351±0.00
Katz	0.778±0.005	0.587±0.004	0.629±0.005	0.909±0.002	0.353±0.00

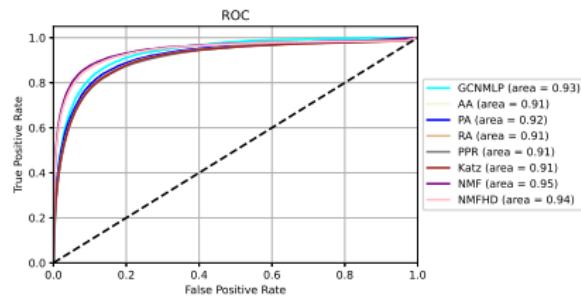
AUPR score of various algorithms



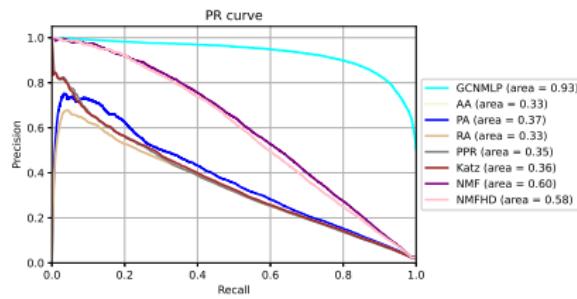
NMF & NMFHD proposed by Timilsina et al. Discovering links between side effects and drugs using a diffusion based method. Sci. Rep. 2019.

The ROC and PR curves

- The NMF performs best on the ROC curve among all methods.
- The GCNMLP is superior to all the other methods on the AUPR.
- Both heuristic network link-prediction and NMF-based methods are inferior to the GCNMLP.



(d)



(e)

The GCNMLP and various ML methods

The GCNMLP vs. various ML methods.

Model	AUROC	AUPR
GCNMLP	0.939 ± 0.002	0.941 ± 0.002
collaborative filtering	0.901 ± 0.002	0.609 ± 0.003
logistic regression	0.898 ± 0.058	0.836 ± 0.126
k-nearest neighbor classifier	0.882 ± 0.090	0.764 ± 0.161
support vector machine	0.898 ± 0.048	0.836 ± 0.106
random forest	0.889 ± 0.069	0.799 ± 0.125
gradient boosting classifier	0.906 ± 0.043	0.843 ± 0.080

New predictions not seen in the dataset SIDER

Vancomycin			Amlodipine		
Side effect	Prob.	Lit.	Side effect	Prob.	Lit.
Decreased appetite	1		Diarrhoea	1	✓
Paraesthesia	1	✓	Somnolence	1	
Dyspepsia	1		Feeling abnormal	1	
Musculoskeletal discomfort	1		Decreased appetite	0.999999	
Anorexia	1		Paraesthesia	0.999999	
Hyperhidrosis	1		Dyspepsia	0.999999	✓
Diarrhoea	1		Anorexia	0.999999	
Somnolence	1		Musculoskeletal discomfort	0.999999	✓
Feeling abnormal	1		Hyperhidrosis	0.999999	
Leukopenia	0.999999	✓	Convulsion	0.999999	

New predictions not seen in the dataset SIDER, continued

Cisplatin			Glimperide		
Side effect	Prob.	Lit.	Side effect	Prob.	Lit.
Diarrhoea	1		Shock	0.999999	
Somnolence	0.999999	✓	Dry mouth	0.999998	
Decreased appetite	0.999999		Leukopenia	0.999998	✓
Feeling abnormal	0.999999		Confusional state	0.999998	
Dyspepsia	0.999999		Oedema	0.999998	
Paraesthesia	0.999999		Angioedema	0.999997	✓
Musculoskeletal discomfort	0.999999		Discomfort	0.999996	
Anorexia	0.999999	✓	Agitation	0.999996	
Hyperhidrosis	0.999999		Vision blurred	0.999996	
Convulsion	0.999999	✓	Hypertension	0.999996	

The performance of the GCNMLP

The GCNMLP performs best with respect to all the evaluation scales on FAERS, and the results on OFFSIDES are better than their counterparts on SIDER.

Dataset	AUROC	AUPR	Precision	Recall	F1 score
SIDER	0.939 ± 0.002	0.941 ± 0.002	0.865 ± 0.003	0.865 ± 0.003	0.865 ± 0.003
OFFSIDES	0.964 ± 0.009	0.955 ± 0.013	0.902 ± 0.011	0.903 ± 0.012	0.902 ± 0.012
FAERS	0.973 ± 0.002	0.972 ± 0.002	0.913 ± 0.001	0.913 ± 0.001	0.913 ± 0.001

Conclusion

- We proposed the GCNMLP to predict drug side effects on the datasets SIDER, OFFSIDES, and FAERS.
- The GCNMLP outperforms traditional methods, such as matrix factorization methods, network link-prediction methods and machine learning methods.
- From our model, some predicted side effects can be validated in the literatures, and reveals potential link information.
- Our study suggests that the GCNMLP might be used as a model for exploring drug side effects, and provide for better personalized medicine.
- For our future study, extend the database including the drug-protein or pharmacogenetics info and perform the permutation test for the invariant or equivariant nature of graphs.

Thank you for your attention!



Chen, Shih, Chien & Tsai

Predicting Adverse Drug Effects: A Heterogeneous Graph Convolution Network with a Multi-layer Perceptron Approach

MedRxiv, doi: <https://doi.org/10.1101/2022.03.22.22272749>.



Edwards & Aronson.

Adverse drug reactions: definitions, diagnosis, and management.
Lancet. 2000;356(9237):1255-9.



Scheiber, Chen, Milik, Sukuru, et al.

Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis.
Journal of chemical information and modeling. 2009;49(2):308–317.



Gurwitz, Field, Harrold, et al.

Incidence and Preventability of Adverse Drug Events Among Older Persons in the Ambulatory Setting
JAMA. 2003;289(9):1107-1116.