Applications of Link Prediction in Biology and Medicine

HauTen Lee

School of System Science
Beijing Normal University

Dec 7th, 2020

ACKNOWLEDGEMENT

The presenter would like to emphasize the theory and applications of network in a biological or medical context instead of the disciplines themselves.

TOC

Tasks

- At Micro Level, Predict Possible Interactions in a/an
  - Protein-Protein Interaction (PPI) Network
  - RNA-Protein Network[1]
  - Gene Regulatory Network (GRN)[2]
- At Macro Level, Predict
  - Future diseases on the base of a patient's current health status (aka comorbidity network)[3]
  - Side effects in drug-drug interaction network
  - Interactions among diseases, medicines and proteins

---

[1] Huan Hu et al. "LPI-ETSLP: lncRNA–protein interaction prediction using eigenvalue transformation-based semi-supervised link prediction". In: *Molecular BioSystems* (2017).

[2] Turki Turki and Jason TL Wang. "A new approach to link prediction in gene regulatory networks". In: *ICIDEAL*. 2015.

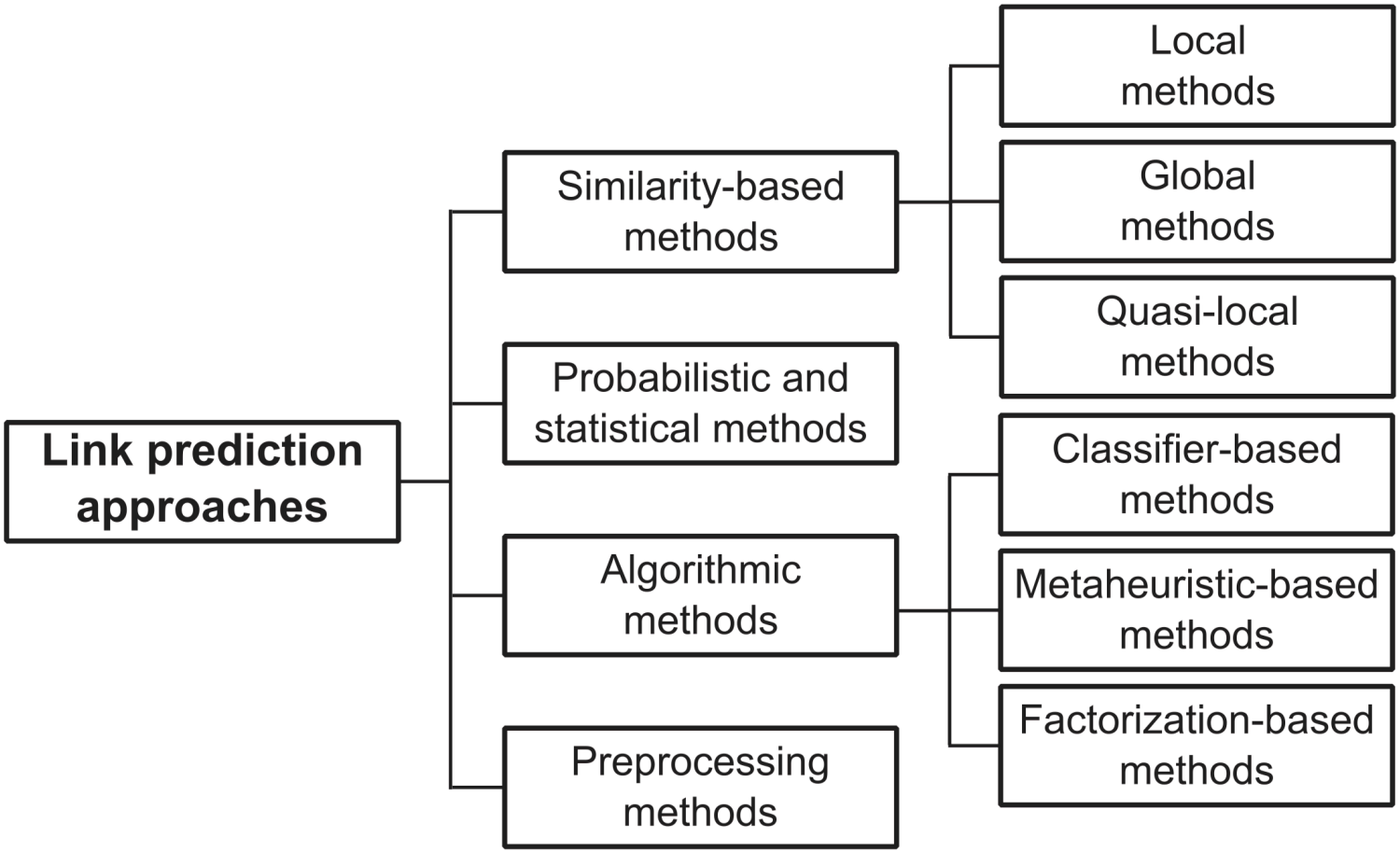
[3] Francesco Folino and Clara Pizzuti. "Link prediction approaches for disease networks". In: *ICIT in Bio-and Medical Informatics*. 2012.

Tasks

How come?

Practical experiments are time-consuming, or the knowledge is rare and usually not observed.

Link Prediciton Techniques[4]



[4] Víctor Martínez et al. "A survey of link prediction in complex networks". In: *ACM computing surveys (CSUR)* (2016).

Node-Neighborhood-Based Approaches

- Common Neighbors (CN)

$$CN(x, y) = \Gamma(x) \cap \Gamma(y)$$

- Jaccard (JC)

$$JC(x, y) = \frac{\Gamma(x) \cap \Gamma(y)}{\Gamma(x) \cup \Gamma(y)}$$

- Resource Allocation (RA)
- ...

Another Viewpoint

In 2010, Yong-Yeol Ahn et al[5] reconsidered communities in a network as a group of links rather than nodes which have identical characteristics.

[5] Yong-Yeol Ahn et al. "Link communities reveal multiscale complexity in networks". In: *nature* (2010).

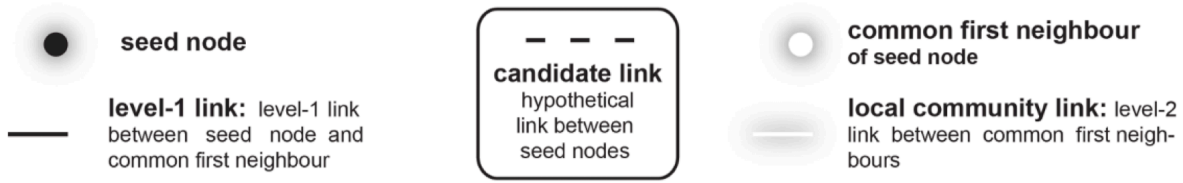
A Local Community Approach to Link Prediction[6]

CAR-Index and Extensions

Explicative example: how CAR-Index predicts the likelihood of candidate links

CN-Index = # common first neighbours

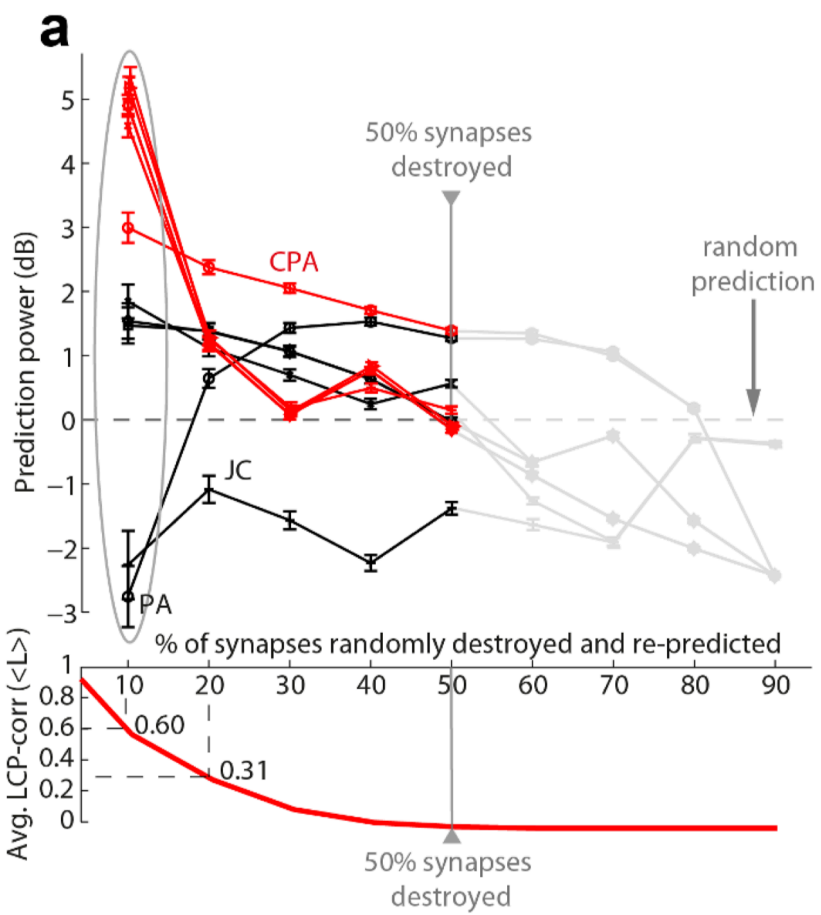
CAR-Index = (# common first neighbours) x (# local community links)

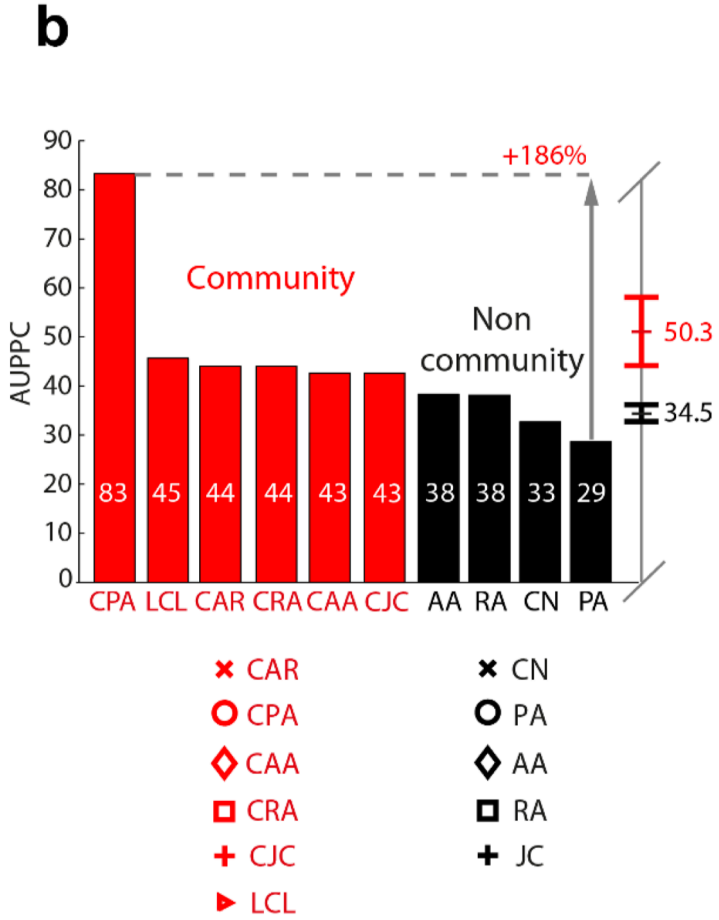


[6] Carlo Vittorio Cannistraci et al. "From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks". In: *Scientific reports* (2013).

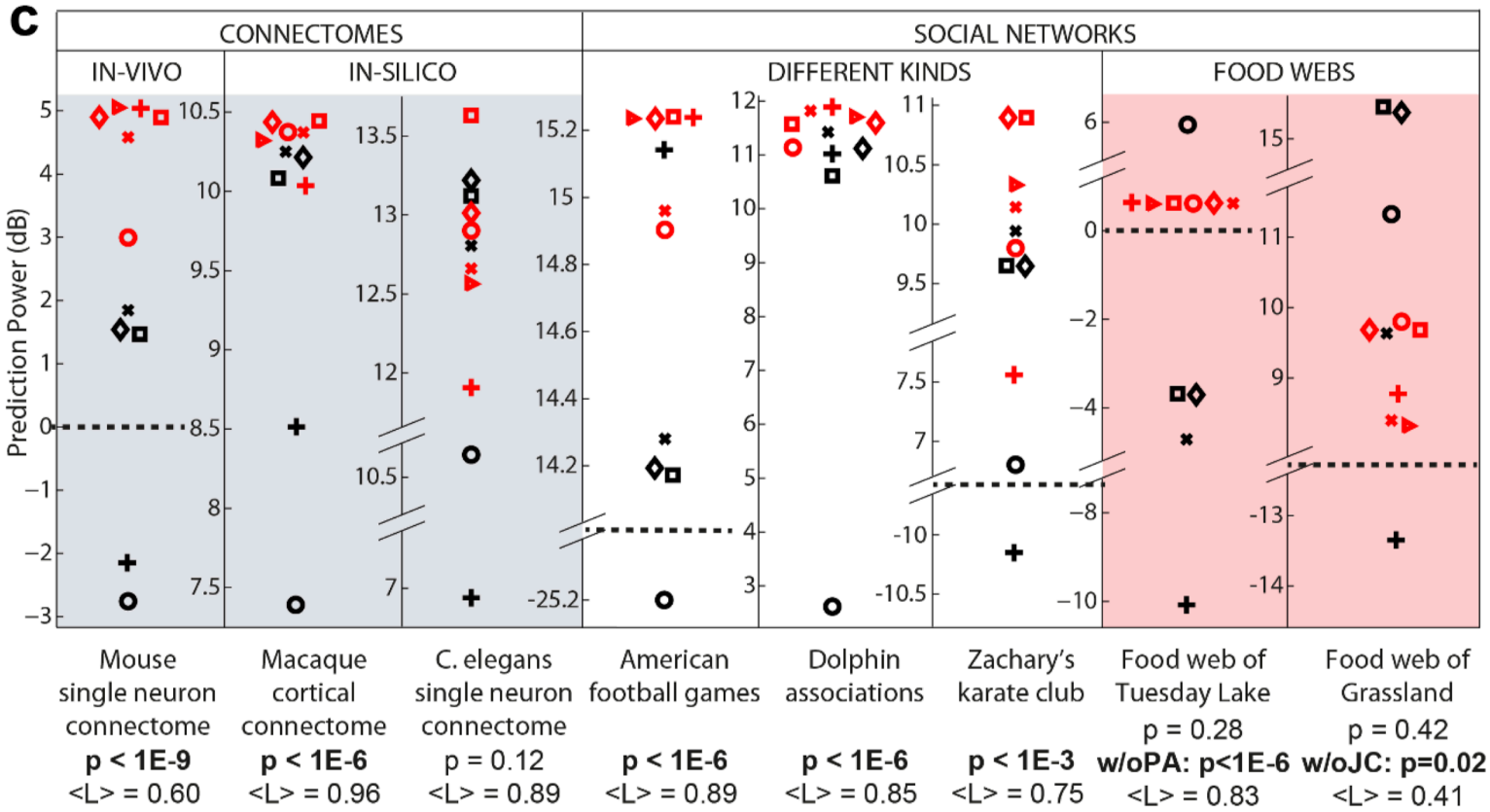
A Local Community Approach to Link Prediction

$$\text{Prediction Power(dB)} = 10 \cdot \log_{10} \frac{\overline{\text{Precision}}_{\text{Prediction Technique}}}{\overline{\text{Precision}}_{\text{Random Predictor}}}$$

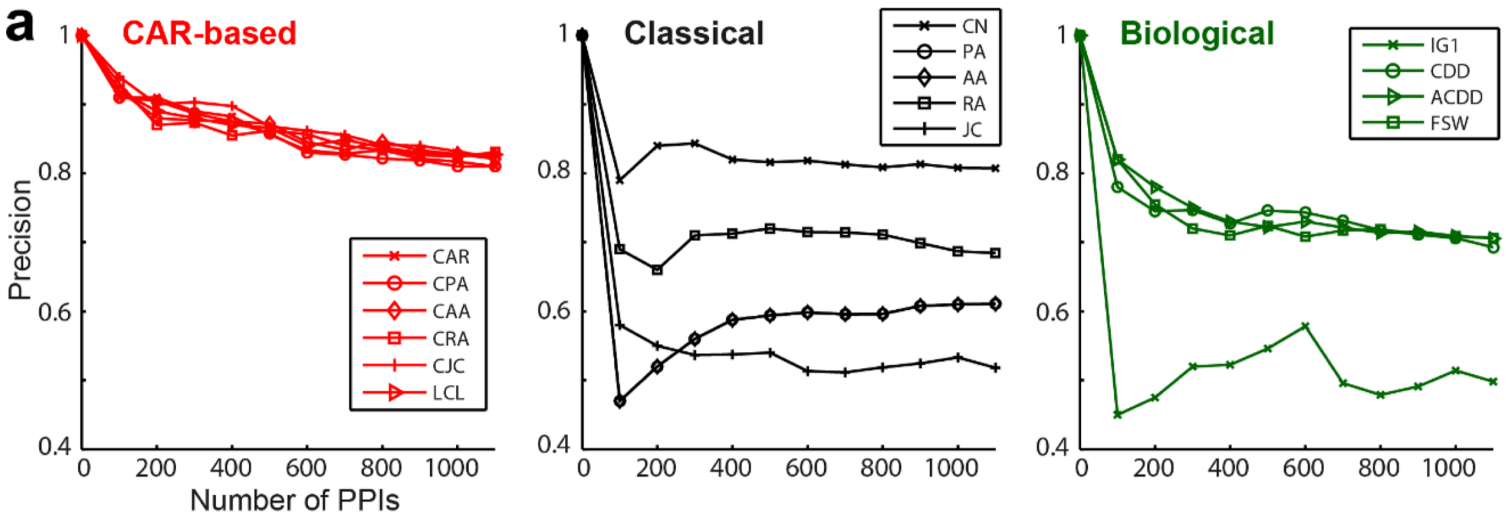




A Local Community Approach to Link Prediction



A Local Community Approach to Link Prediction



From the Perspective of Classification[7]

Link prediction in a drug-drug interaction(DDI) network,

Argument

Whether there is a link or not?
This is a 0/1 question.
Shall apply classifiers.

[7] Andrej Kastrin et al. "Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning". In: *PloS one* (2018).

From the Perspective of Classification

In training set,
Unsupervised Random Forest & k-nearest-neighbor, Supervised Gradient Boosting Machine outperformed Supervised Support Vector Machine & Classification Tree

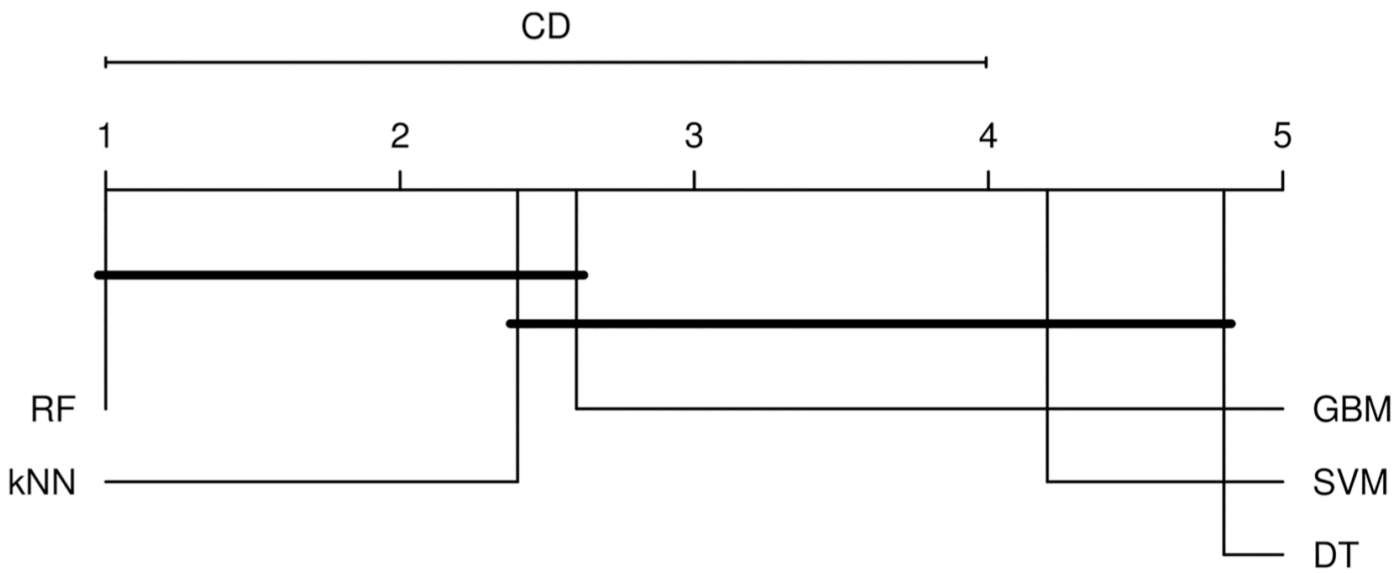


Fig 1. Critical difference (CD) plot for training data. Plot shows the pairwise differences in performance among classifiers. The horizontal scale shows the average rank of each classifier, with smaller ranks indicating better performance. Classifiers connected by a dark line had statistically identical performance at the $p = 0.05$ level.

From the Perspective of Classification

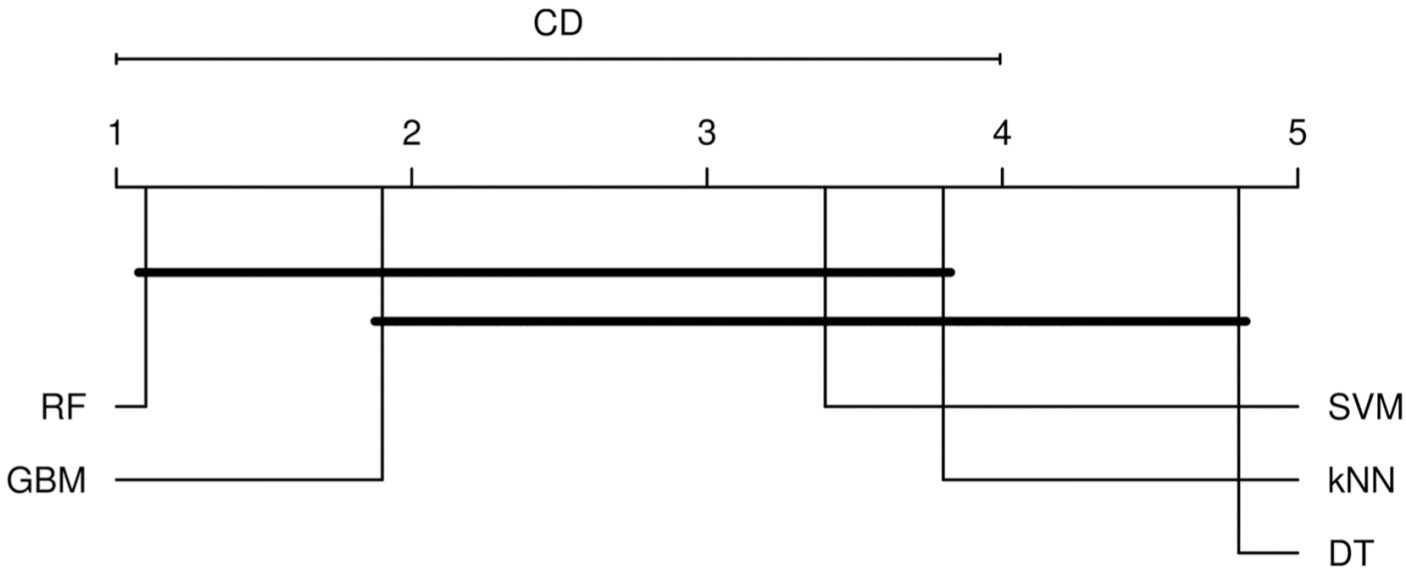
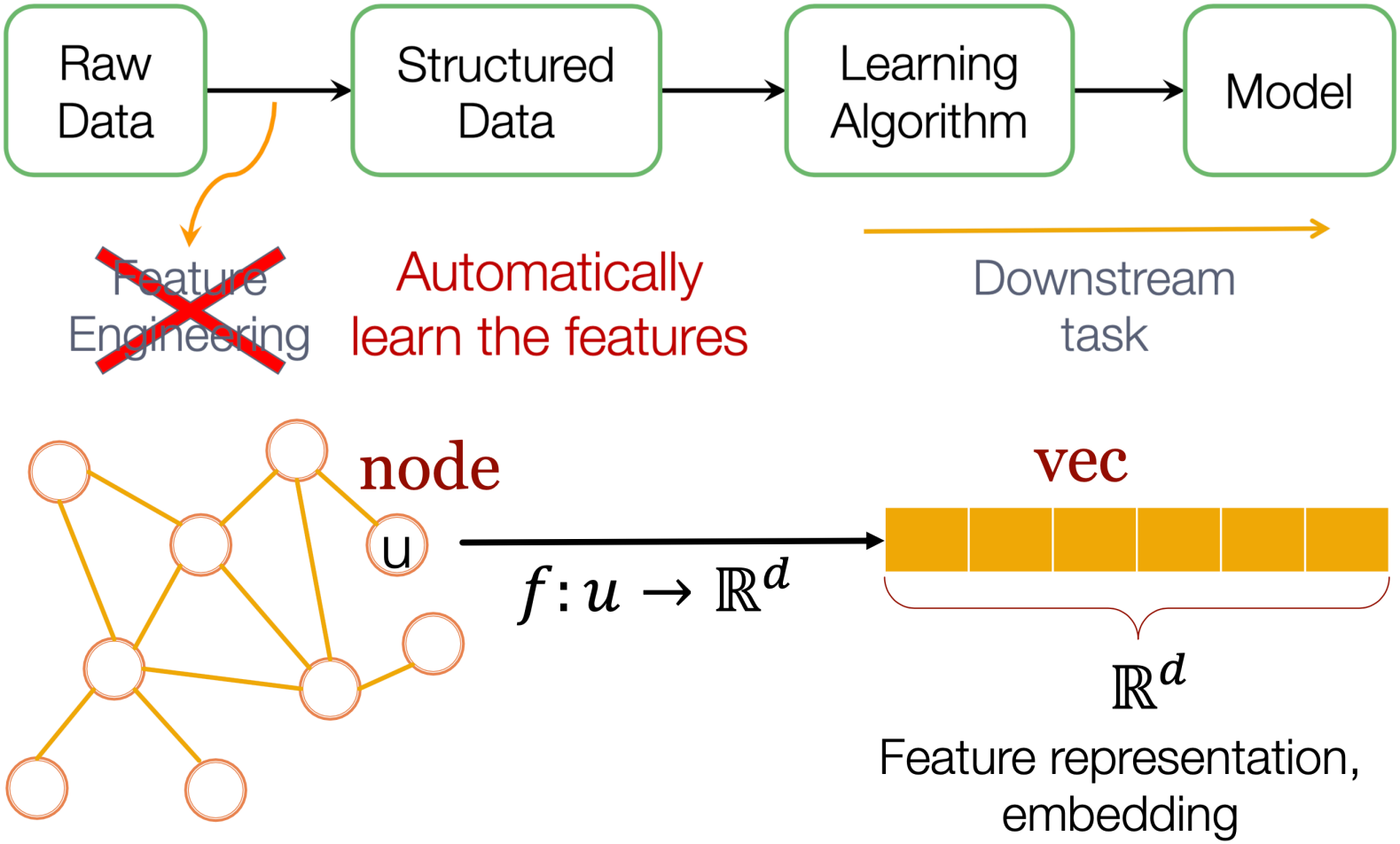


Fig 2. Critical difference (CD) plot for test data. Plot shows the pairwise differences in performance among classifiers. The horizontal scale shows the average rank of each classifier, with smaller ranks indicating better performance. Classifiers connected by a dark line had statistically identical performance at the $p = 0.05$ level.

TOC

Graph Neural Network[8] & Embedding



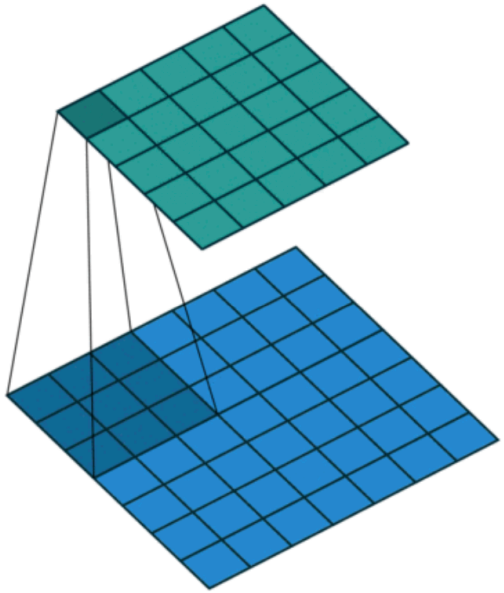
[8] Jure Leskovec and Michele Catasta. *Stanford CS224W*. http://web.stanford.edu/class/cs224w/.

GNN: Orientation

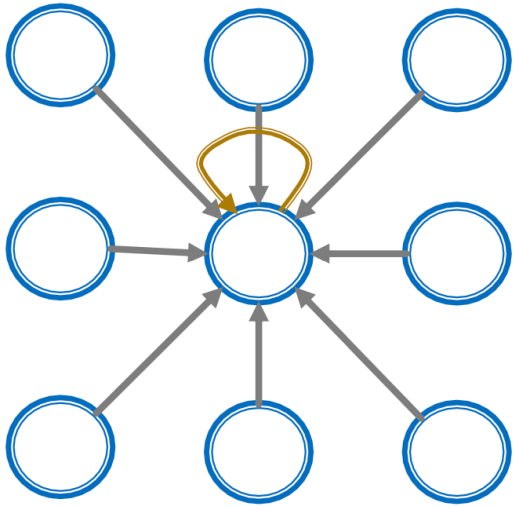
- Node Ordering Invariant
- Applicable to Graphs of Different Sizes
- Computationally Inexpensive

GNN: Intuition

Regular Convolutional Filter, Now We Generalize to Graphs

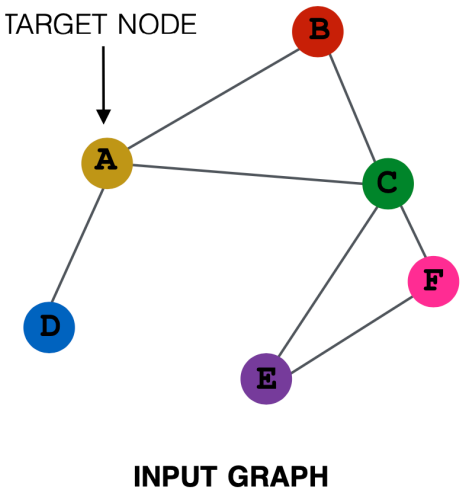


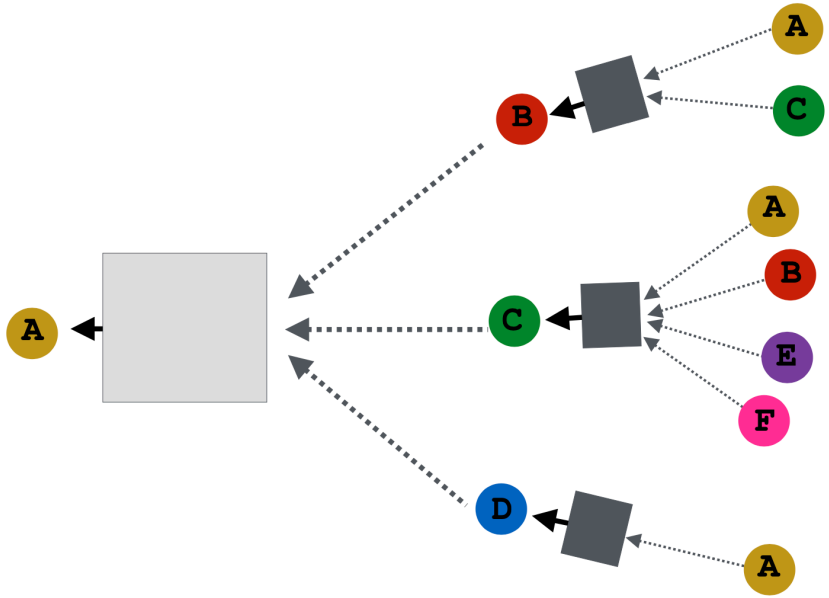
Image



Graph

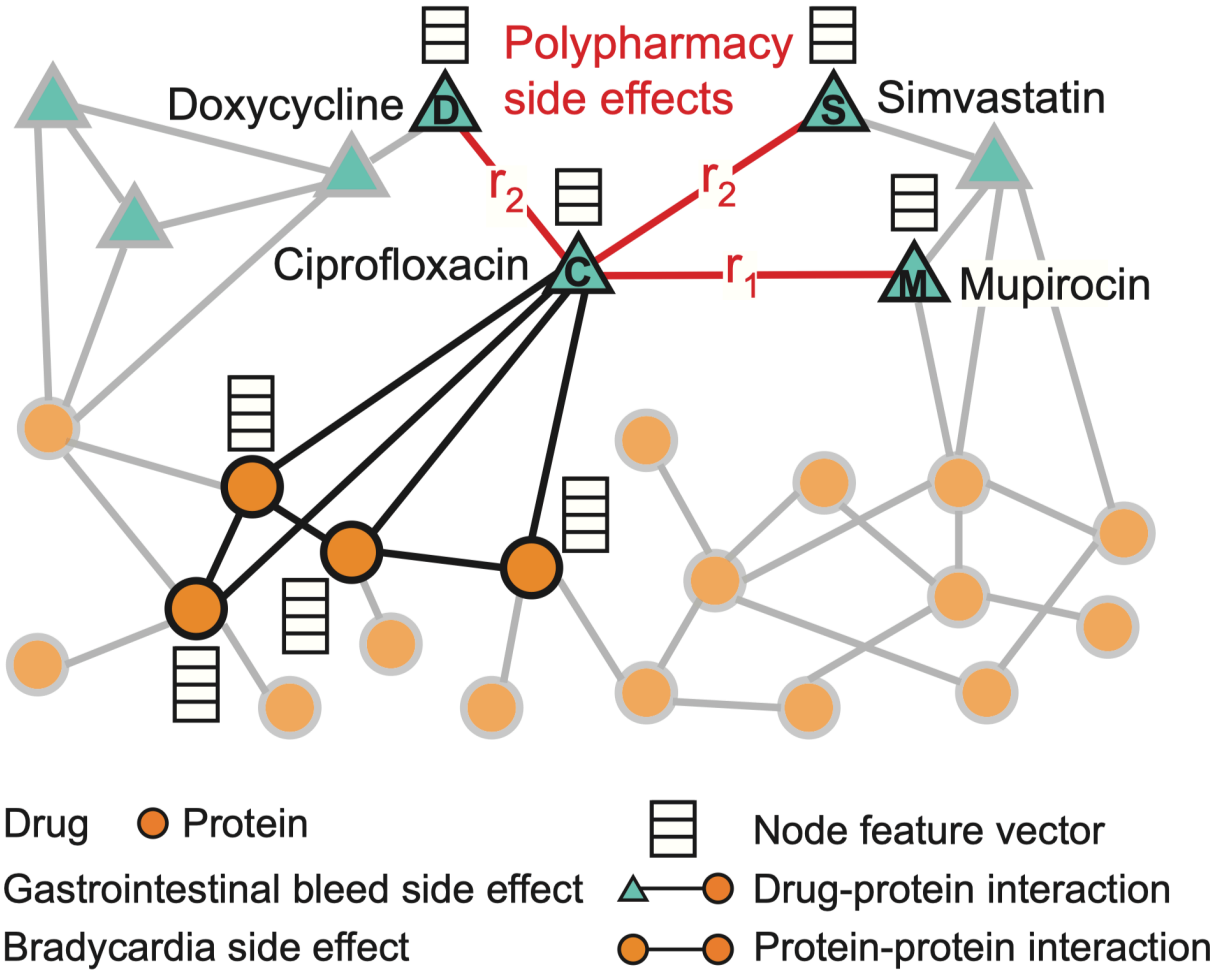
GCNN: Aggregation





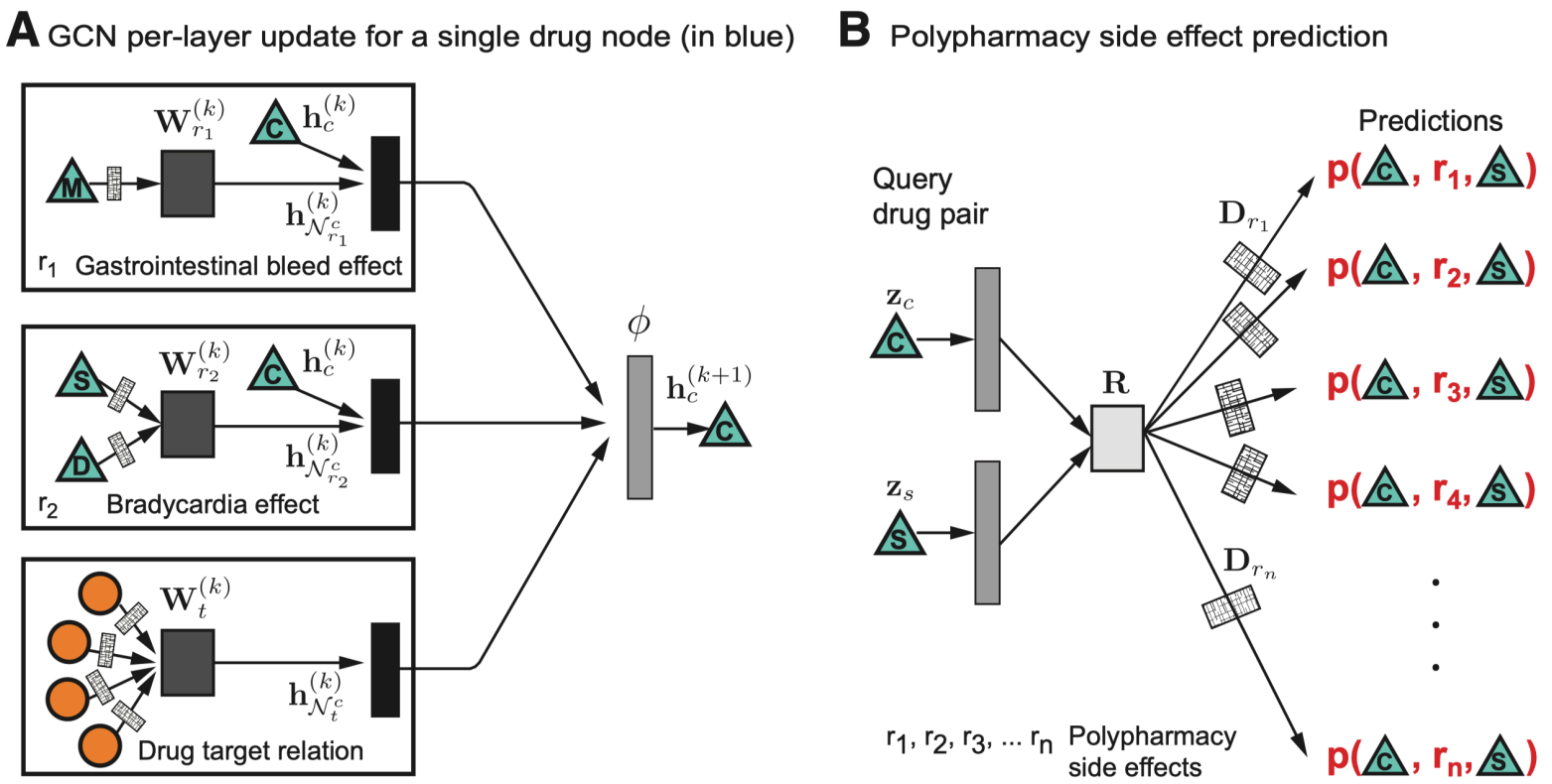
Decagon: Polypharmacy & Side Effects[9]

Types of Nodes, Types of Edges



[9] Marinka Zitnik et al. "Modeling polypharmacy side effects with graph convolutional networks". In: *Bioinformatics* (2018).

Decagon

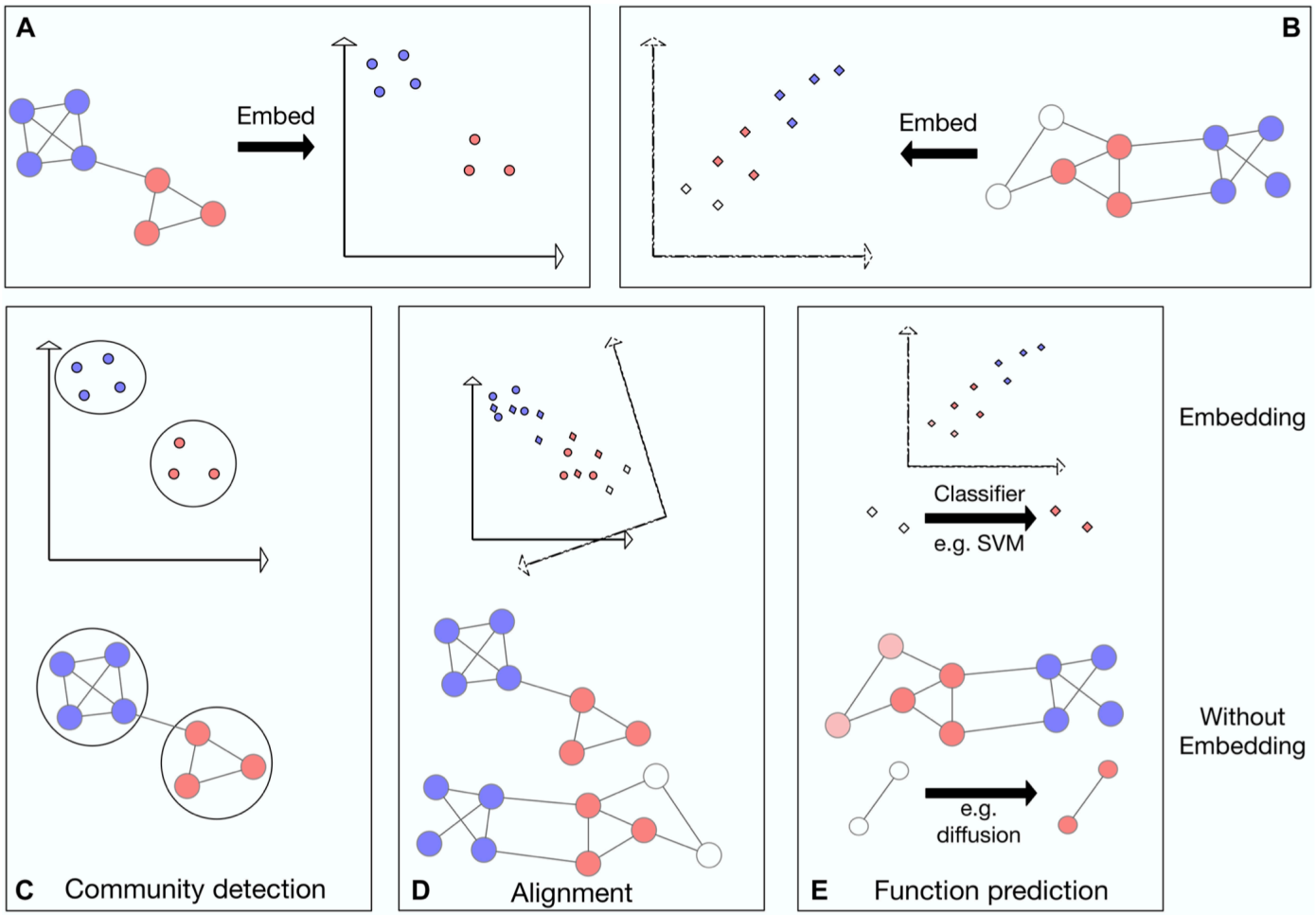

A GCN per-layer update for a single drug node (in blue)
B Polypharmacy side effect prediction
r1 Gastrointestinal bleed effect
r2 Bradycardia effect
Drug target relation
Query drug pair
R
Predictions
r1, r2, r3, ... rn Polypharmacy side effects

Decagon: Results

Approach	AUROC	AUPRC
Decagon	0.872	0.832
RESCAL tensor factorization	0.693	0.613
DEDICOM tensor factorization	0.705	0.637
DeepWalk neural embeddings	0.761	0.737
Concatenated drug features	0.793	0.764

Polypharmacy effect r	Drug i	Drug j	Evidence
Sarcoma	Pyrimethamine	Aliskiren	Stage *et al.* (2015)
Breast disorder	Tolcapone	Pyrimethamine	Bicker *et al.* (2017)
Renal tubular acidosis	Omeprazole	Amoxicillin	Russo *et al.* (2016)
Muscle inflammation	Atorvastatin	Amlodipine	Banakh *et al.* (2017)
Breast inflammation	Aliskiren	Tioconazole	Parving *et al.* (2012)

Embedding: Schematic Representing[10]



[10] Walter Nelson et al. "To embed or not: network embedding as a paradigm in computational biology". In: *Frontiers in genetics* (2019).

Embedding vs. Graph-Based

Fucntion Prediction	GeneMANIA[11]	Mashup[12]
AUPR		
MF	0.327	**0.372**
BP	0.213	**0.222**
CC	**0.514**	0.487

Conclusion

In the series of experiments, sometimes graph-based methods outperformed network tools embedding. However, sometimes the results are reversed.

[11] Sara Mostafavi et al. "GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function". In: *Genome biology* (2008).

[12] Hyunghoon Cho et al. "Compact integration of multi-network topology for functional analysis of genes". In: *Cell systems* (2016).

TOC

Future Research Directions[13]

- Integrating domain knowledge of biological network to proposed algorithms
- Interpreting biological networks evolution using link prediction algorithms
- ...

[13] Sadegh Sulaimany et al. "Link prediction potentials for biological networks". In: *International Journal of Data Mining and Bioinformatics* (2018).

Promotion

Powered by LaTeX & awesomeBNUbEAMer.

Get Started

awesomeBNUbEAMer is accessible at

- htlee6.github.io/awesomeBNUbEAMer

- Overleaf, an online editor