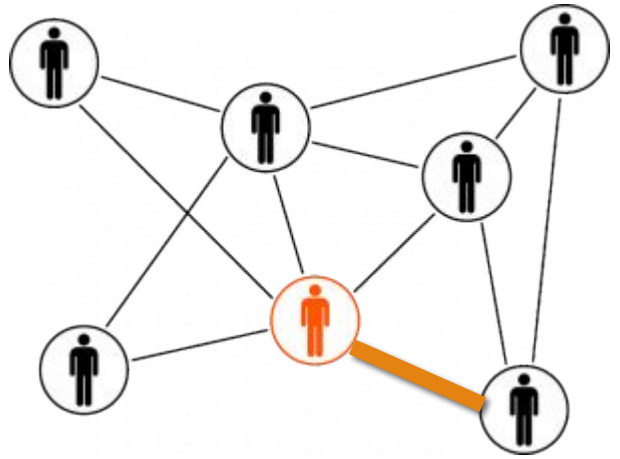
Link prediction

- One of the most popular topics in complex networks
  - Use cases:
    - Finding relationship between proteins
    - Social network interaction
    - Co-authorship (papers)
    - Identifying fraudulent transactions (bank)
    - Drug-drug interaction prediction
    - ..



Drug-drug interaction prediction (DDI)

Shtar, Guy, Lior Rokach, and Bracha Shapira. "Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures." *PloS one* 14.8 (2019): e0219796.
Shtar G, Greenstein-Messica A, Mazuz E, Rokach L, Shapira B. Predicting drug characteristics using biomedical text embedding. BMC bioinformatics. 2022 Dec;23(1):1-7.
2 more extensions under review

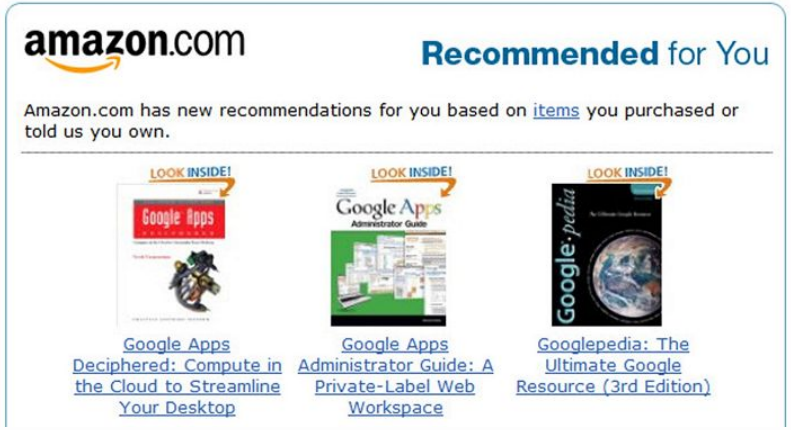
Problem formulation

Predicting unknown drug-drug interactions
Binary

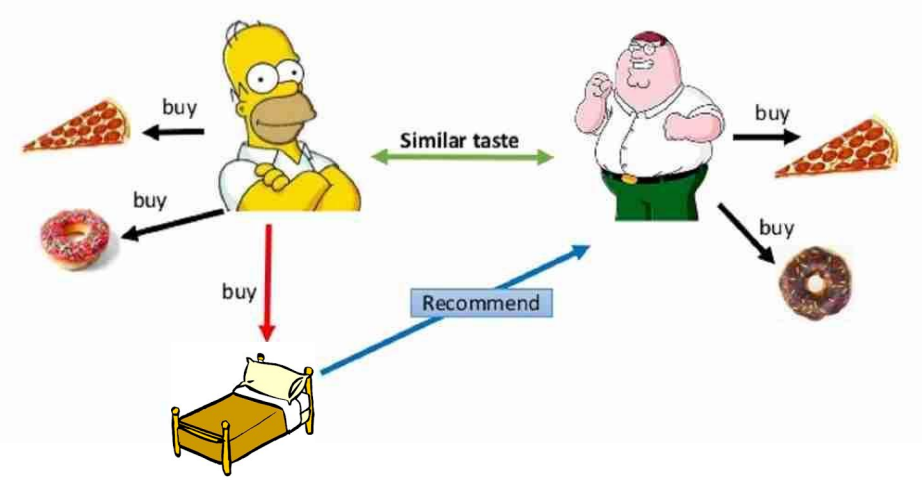
DRUG	INTERACTION
Abacavir	Acetaminophen may decrease the excretion rate of Abacavir which could result in a higher serum level.
Abametapir	The serum concentration of Acetaminophen can be increased when it is combined with Abametapir.
Abatacept	The metabolism of Acetaminophen can be increased when combined with Abatacept.
Abemaciclib	The metabolism of Abemaciclib can be increased when combined with Acetaminophen.

Drug-drug interaction prediction

How do on-line stores recommend products



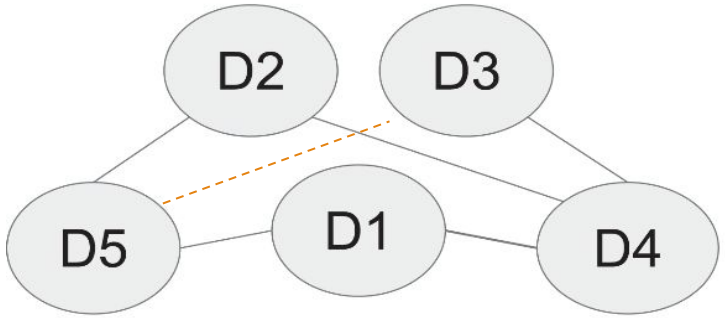
Best performing methods do not use information about the product.
Collaborative filtering.



Drug-drug interaction prediction

As a network

D3 and D5 interact?



Nodes represent drugs, and edges represent interactions

Link prediction similarity measures

Common neighbors: $S^{CN}(u,v) = |\Gamma(v) \cap \Gamma(u)|.$

node neighbours

$$\Gamma(u) = \{v \,|\, e_{uv} \in E\}$$

Jaccard index: $S^{Jaccard}(u,v) = \frac{|\Gamma(v) \cap \Gamma(u)|}{|\Gamma(v) \cup \Gamma(u)|}.$

Adamic Adar: $S^{AA}(u,v) = \sum_{w \in \Gamma(v) \cap \Gamma(u)} \frac{1}{\log |\Gamma(w)|}.$

Improvement – averaged each measure by considering the value of the neighbors. example:

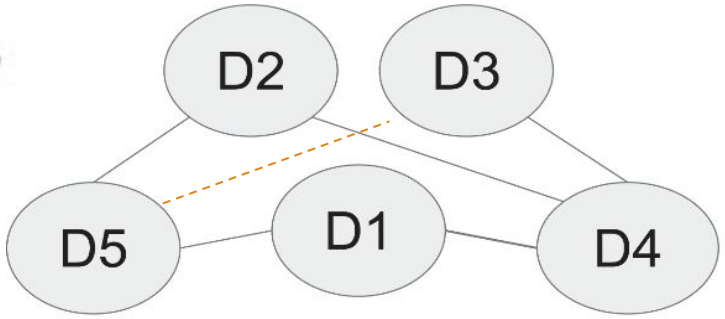
$$\widetilde{S}^{ACN}(u,v) = \frac{1}{|\Gamma(v)|} \sum_{w \in \Gamma(v)} S^{CN}(w,u).$$

Example:

Common neighbors: $S^{CN}(u,v) = |\Gamma(v) \cap \Gamma(u)|.$

S(D5,D4) = |Γ(*D5*) ∩ Γ(*D4*)| = {D2,D1} ∩ {D1,D2,D3} = 2

Other nodes has highest similarity with D5?



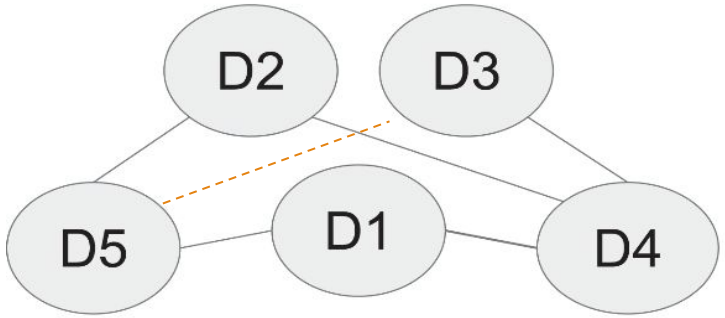
Nodes represent drugs, and edges represent interactions

Example:

$$S^{Jaccard}(u,v) = \frac{|\Gamma(v) \cap \Gamma(u)|}{|\Gamma(v) \cup \Gamma(u)|}.$$

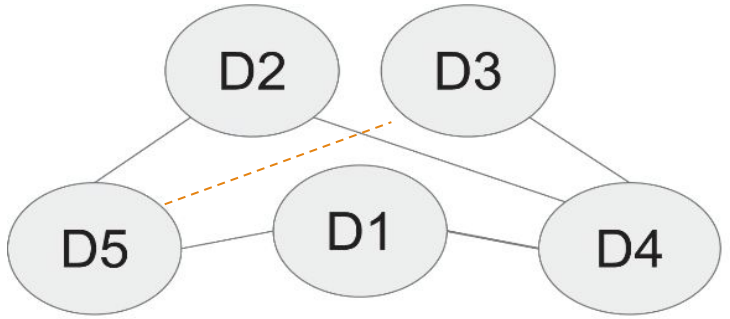
|Γ(*D5*) ∩ Γ(*D4*)| = {D2,D1} ∩ {D1,D2,D3} = 2

What is the Jaccard here?



Nodes represent drugs, and edges represent interactions

Matrix factorization



Nodes represent drugs, and edges represent interactions

Adjacency matrix

Drug	D1	D2	D3	D4	D5
D1	0	0	0	1	1
D2	0	0	0	1	1
D3	0	0	0	1	0
D4	1	1	1	0	0
D5	1	1	0	0	0

Rows matrix

0	-0.9
0	-0.9
0	-0.54
-1.1	0
-1.1	0

Columns matrix

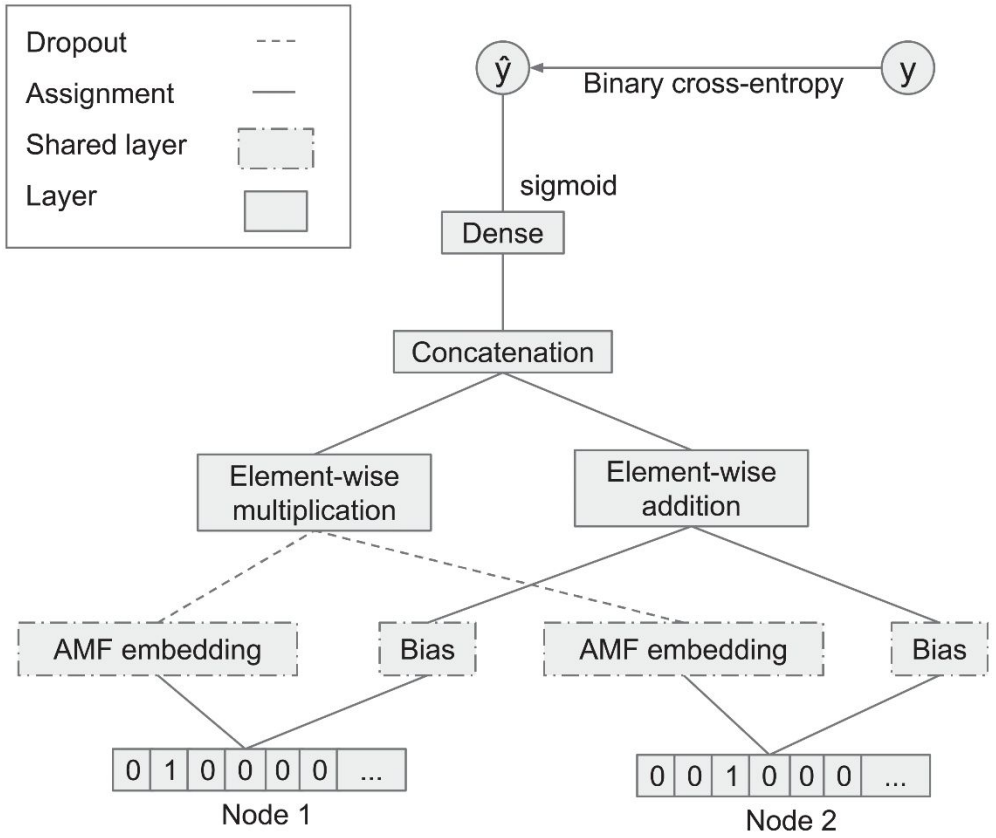
-0.9	-0.9	-0.9	0	0
0	0	0	-0.12	-0.9

Prediction

Drug	D1	D2	D3	D4	D5
D1	0	0	0	1.1	0.93
D2	0	0	0	1.1	0.93
D3	0	0	0	0.81	**0.74**
D4	1	1	0.81	0	0
D5	0.93	0.93	**0.74**	0	0

Methods

Our method, adjacency matrix factorization:



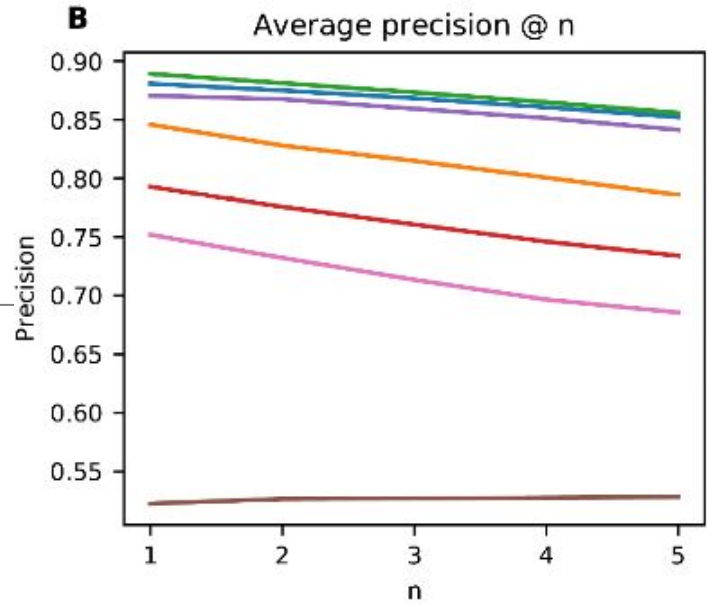
Results Holdout

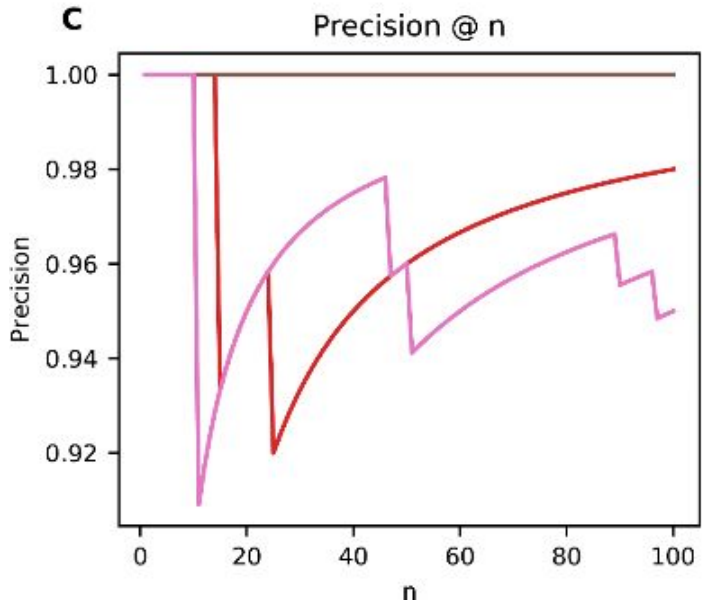
Holdout = randomly sampling 20% of interactions for testing set

Algorithm	AUROC curve	AUPR curve
XGBoost Ensemble from all	0.991	0.960
AMF	0.990	0.950
Vilar et al. [8]	0.952	0.784
Average common neighbors	0.938	0.738
Average Jaccard coefficient	0.967	0.840
Adamic/Adar	0.933	0.728

Results Holdout

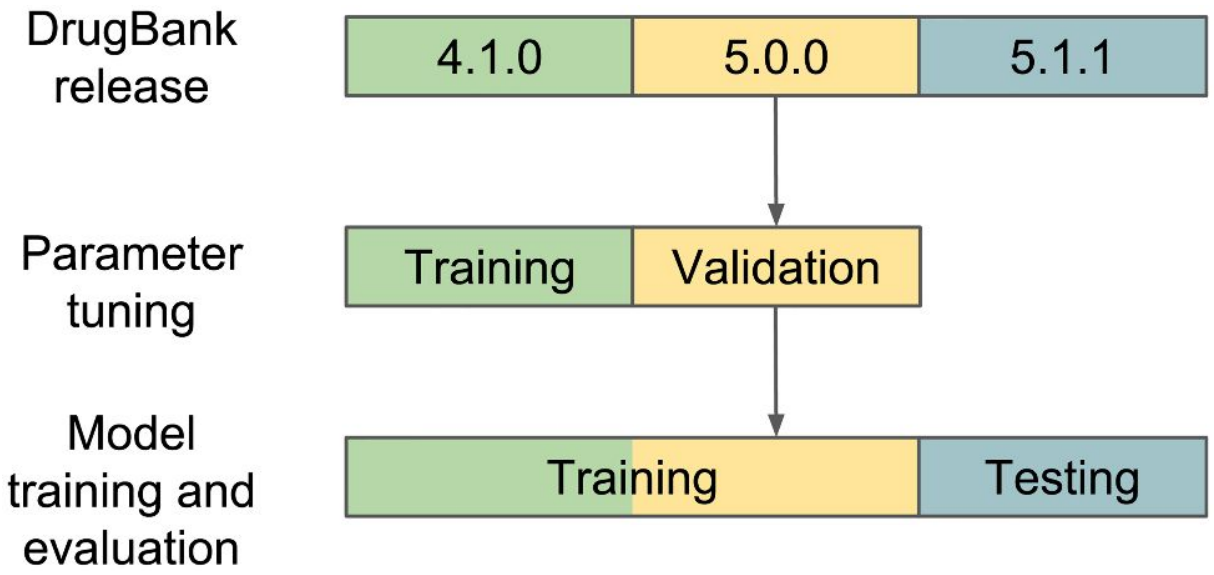
?





Evaluation

Retrospective

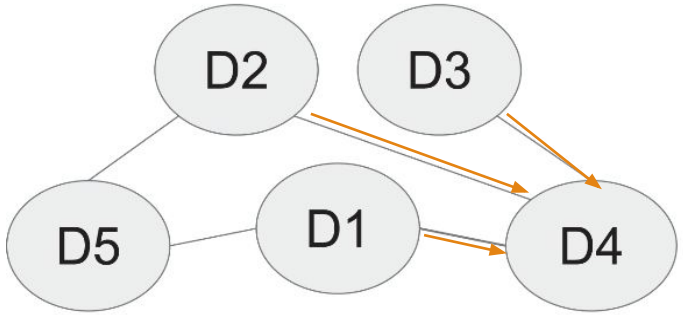


Results Retrospective

Algorithm	AUROC curve	AUPR curve
XGBoost	0.814	0.425
[redacted]		
AMF	0.748	0.304
Vilar et al. [8]	0.787	0.38
Average common neighbors	0.802	0.385
Average Jaccard coefficient	0.804	0.370
Adamic/Adar	0.791	0.388

AMFP

After AMF, use the following algorithm to propagate node embeddings (latent factors) over the network.

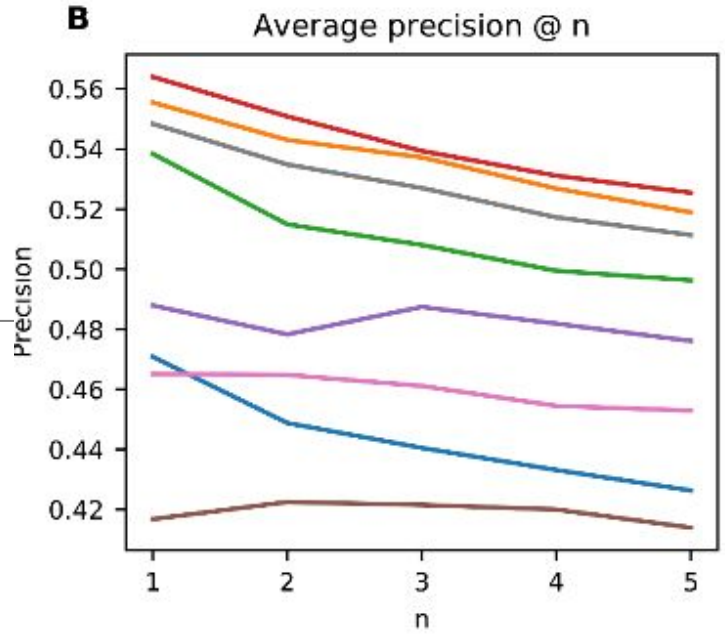


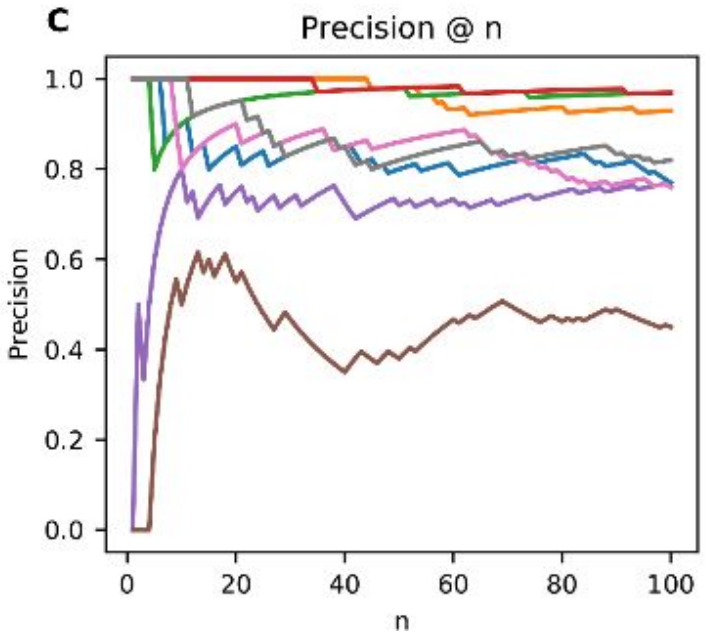
$$P'_{D4} = \alpha \cdot \frac{1}{3} \cdot (P_{D3} + P_{D1} + P_{D2}) + (1 - \alpha) \cdot P_{D4}$$

Results Retrospective

Algorithm	AUROC curve	AUPR curve
XGBoost	0.814	0.425
AMFP	0.807	0.417
AMF	0.748	0.304
Vilar et al. [8]	0.787	0.38
Average common neighbors	0.802	0.385
Average Jaccard coefficient	0.804	0.370
Adamic/Adar	0.791	0.388

Results retrospective





And what about additional data?

Will other data improve the model?

What is the limitation of this method?

Substructure
Target
Transporter
Enzyme
Pathway
Indication
Label
Off label
CN
AA
RA
Katz
ACT
RWR

Method	Similarity	AUROC curve	AUPR curve
AMF (my work)		**0.9591**	**0.8108**
XGB		0.9588	0.8017
Neighbor recommender	Substructure	0.9362	0.7593
	Target	0.8197	0.3642
	Transporter	0.7143	0.3288
	Enzyme	0.7562	0.3774
Zhang et a.	Pathway	0.812	0.5714
	Indication	0.9119	0.5992
	Label	0.9359	0.7537
	Off label	0.9397	0.768
	CN	0.9411	0.7671
	AA	0.9414	0.7676
	RA	0.9432	0.7704
	Katz	0.9373	0.7352
	ACT	0.9044	0.7239
	RWR	0.9409	0.7666
Matrix perturbation method		0.9484	0.7818
Weighted average ensemble		0.9507	0.7955
Ensemble classifier (L1)		0.9571	0.8073
Ensemble classifier (L2)		0.9562	0.806

Results

All possible predictions were ranked

First predictions:

- Curcumin **and** Primidone
- Ceritinib **and** Fluvoxamine
- Rifapentine **and** Fluvoxamine
- Curcumin **and** Pentobarbital
- Curcumin **and** Rifapentine
- Lumacaftor **and** Fluvoxamine

Reference

[Vilar et al.] Vilar S, Uriarte E, Santana L, Tatonetti NP, Friedman C. Detection of Drug-Drug Interactions by Modeling Interaction Profile Fingerprints. PLOS ONE. 2013;8(3):1–11.

[Zhang et al.] Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. BMC Bioinformatics. 2017;18(1):18. pmid:28056782