Network-based approach for drug repurposing using drug signature and disease phenotype

Ariane ALIX
Master MVA École Normale Supérieure Paris-Saclay

Summary

- Introduction
- The data
- Principle of the method
- Implementation
- Results

Introduction

- Drug repurposing consists of the investigation of already used drugs, to see if they can be used for treating other diseases.

- Interesting because it is a fast and cost-effective strategy for drug discovery and launching.

- Drug–disease associations prediction method based on a recommendation system (bipartite graph) where the studied features are the drug signatures and diseases phenotypes.

The data

Drug signatures

and disease phenotypes

Gene expression profiles

- **Disease phenotype:**
  A list of genes that are up or down-regulated when compared to healthy controls.
  48 diseases and 11,393 genes.

- **Drug signature**:
  Gene expression response within cells when they are stimulated with a drug.
  617 drugs and 12,717 genes.

	GSK-3-inhibitor-II	idelalisib	Salermide
RNF14	0.013539515	0.002370315	1.60E-05
UBE2Q1	-0.003986923	0.002493697	-0.005188018
RNF17	-0.013635419	-0.007685559	0.000911763
RNF10	-0.01313385	0.022245417	0.005333798
RNF11	0.005126541	0.000776593	0.012249535
C6ORF123	-0.002572781	-0.005962765	-0.003951608
RNF13	0.006363582	-0.000873812	0.001708042

Drug-disease associations table

Table referencing some approved drug-disease associations.

- 7325 matches of 1543 distincts drugs and 1465 distinct diseases.
- 48/48 diseases of the phenotypes tables are in the association table.
- 45/617 of the drugs in the signatures are in the association table.

In total: 2115 distinct drugs and 1465 distinct diseases.

Drug	Disease	Status
Cetuximab	Malignant tumor of colon	Approved
Dornase alfa	Cystic Fibrosis	Approved
Denileukin diftitox	Lymphoma, T-Cell, Cutaneous	Approved
Etanercept	Rheumatoid Arthritis	Approved
Leuprolide	Endometriosis	Approved
Peginterferon alfa-2a	Hepatitis C, Chronic	Approved
Alteplase	Acute myocardial infarction	Approved
Sermorelin	Pituitary dwarfism	Approved
Urokinase	Pulmonary Embolism	Approved

Principle of the method

Similarity metrics and recommendation systems

Statistically Significant Connectivity Map (ssCMap)

Connectivity between two drugs or two diseases with a signature of N genes?

The i^{th} most important gene effect will be assigned the rank (N−i+ 1) if it is upregulated and −(N−i+ 1) if it is donwregulated.

ssCMap connectivity score with R_1 and R_2 the ranks :

$$C(R_1, R_2) = \frac{\sum_{i=1}^{N} R_1(g_i) R_2(g_i)}{\sum_{i=1}^{N} (N - i + 1)^2}$$

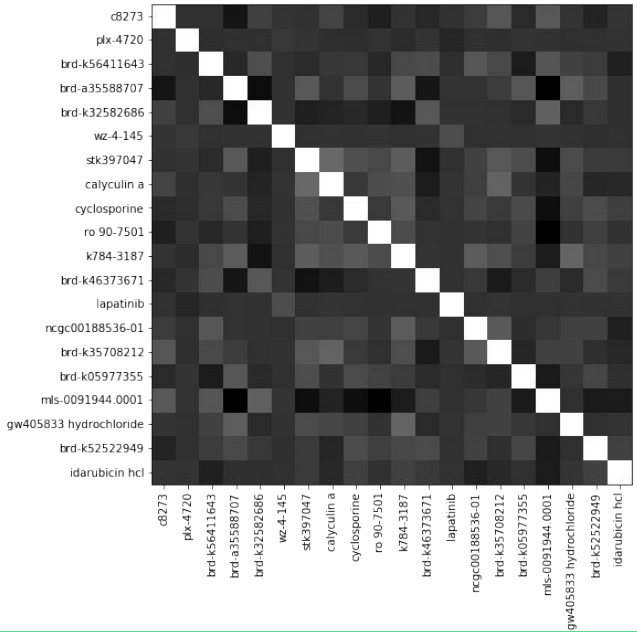
Recommendation systems

Implementation

Computation of the matrices, choice of parameters

S^{drug} and $S^{disease}$ the drug-drug and disease-disease similarity matrices

Computed with the previous formula + normalization between 0 and 1. Extract:



S^{dd} the disease-disease similarity matrix based on the interactions with similar drugs in the network

If two diseases and *i* and *j* are linked by many highly similar drugs then s_{ij}^{dd} will be high. It is computed with:

$$s_{ij}^{\mathrm{dd}} = \frac{\sum_{k=1}^{N_{\mathrm{drug}}} \sum_{l=1}^{N_{\mathrm{drug}}} \left(a_{il} a_{jk} s_{lk}^{\mathrm{drug}} \right)}{\sum_{k=1}^{N_{\mathrm{drug}}} \sum_{l=1}^{N_{\mathrm{drug}}} \left(a_{il} a_{jk} \right)}$$

Final disease-disease similarity matrix S

Combination of S^{dd} and $S^{disease}$:

$$s_{ij} = \alpha s_{ij}^{\text{disease}} + (1 - \alpha) s_{ij}^{\text{dd}}$$

α is a parameter to choose to get the best results.
To do that, we computed S for different values of α, with cross-validation.

Weight matrix W

Matrix of weights (corresponding to objection-projection of the bipartite graph) such that the reco can be computed with *R = W* x *A*.

$$w_{ij} = \frac{s_{ij}}{k(t_i)^{1-\lambda} k(t_j)^{\lambda}} \sum_{l=1}^{N_{\text{drug}}} \frac{a_{il} a_{jl}}{k(d_l)}$$

λ is a parameter to choose to get the best results.
To do that, we computed W for different values of λ, with cross-validation.

Cross-validation

- 10 folds
- Each time removing 20% of the associations from the A table
- Testing each fold with 81 W matrices with combinations of parameters λ and α
- Total: 810 W matrices
- Best recovery of missing associations with λ=0.1 and α=0.1

Results

Examples of interesting results

ind_id	abdominal actinomycosis	previous reco
benzylpenicillin	1.000000	1
phenoxymethylpenicillin	0.845747	0
procaine benzylpenicillin	0.727375	0
cefixime	0.540433	0
cefradine	0.537274	0
cefprozil	0.529992	0
cefaclor	0.529436	0
cyclacillin	0.354836	0
cephalexin	0.328189	0
antipyrine	0.303486	0

ind_id	addison disease	previous reco
fludrocortisone	0.131250	1
hydrocortisone	0.097084	1
cortisone acetate	0.097005	1
prednisone	0.085866	1
dalfampridine	0.081788	0
glatiramer acetate	0.081788	0
betamethasone	0.079831	1
dexamethasone	0.079320	1
triamcinolone	0.074891	1
prednisolone	0.072578	1

ind_id	alcohol withdrawal delirium	previous reco
chlordiazepoxide	1.000000	1
oxazepam	1.000000	1
diazepam	1.000000	1
clorazepate	0.554800	0
etizolam	0.442906	0
halazepam	0.442906	0
chlormezanone	0.442906	0
ethyl loflazepate	0.442906	0
meprobamate	0.442906	0
trifluoperazine	0.442906	0

ind_id	hypersomatotropic gigantism	previous reco
lanreotide	1.000000	1
pegvisomant	0.772644	0
lisuride	0.381525	0
bromocriptine	0.211878	0
octreotide	0.132879	0
abacavir	0.000000	0
olsalazine	0.000000	0
oritavancin	0.000000	0
orciprenaline	0.000000	0
oprelvekin	0.000000	0