ISBRA 2012 – Dallas 21-23 May

Large Scale Ranking and Repositioning of Drugs with respect to DrugBank Therapeutic Categories

Matteo Re and Giorgio Valentini

Dept. of Computer Science

University of Milan - Italy

Outline

- Drug repositioning
- Large scale ranking of drugs w.r.t. DrugBank therapeutic categories
- *ΨnetPro*: a general framework to construct pharmacological networks
- *Kernelized score functions* for drug ranking in pharmacological networks
- Experiments with 1253 FDA approved drugs
- Conclusions and developments

Drug repositioning

- Small scale (*Kotelnikova et al. 2010, Li et al 2010*)
- Large scale (*Iorio et al 2010, Gottlieb et al 2011*)

Computational tasks related to drug discovery:

- Clustering-based approaches (*Noeske et al 2005, Iorio et al 2010*)
- Prediction of drug-target interactions (*Keiser et al 2009, Yamanishi et al 2010*)
- Prediction of drug-disease association (*Gottlieb et al 2011, Chiang and Butte, 2009*)
- ...

A novel prediction task:

Large scale ranking of drugs w.r.t. DrugBank therapeutic categories

Why DrugBank therapeutic categories?

- Why not diseases? “At present, there is not a comprehensive and systematic representation of known drugs indications that would enable a fine-scale delineation of types of drug-disease relationships” (*Dudley et al 2011*)
- Manually curated using medical literature

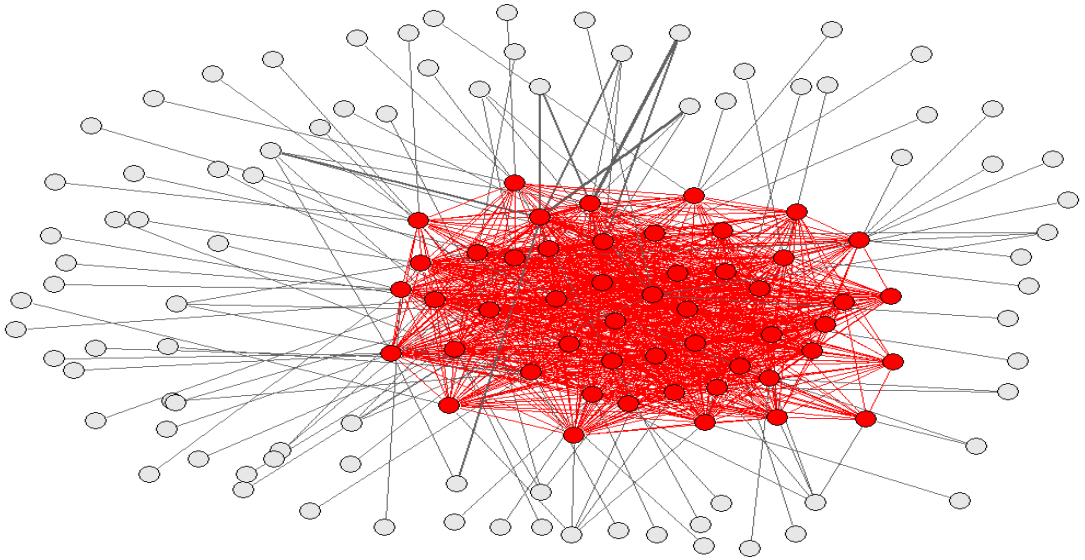
Drug ranking problem

Having :

- A network G=<V,E> connecting a large set of drugs: { Nodes → drugs; Edges → similarities }

- A subset $V_C \subset V$ of drugs belonging to a given therapeutic category C

Rank drugs $v \in V$ w.r.t. to a given therapeutic category C



Drug repositioning in homogeneous pharmacological networks

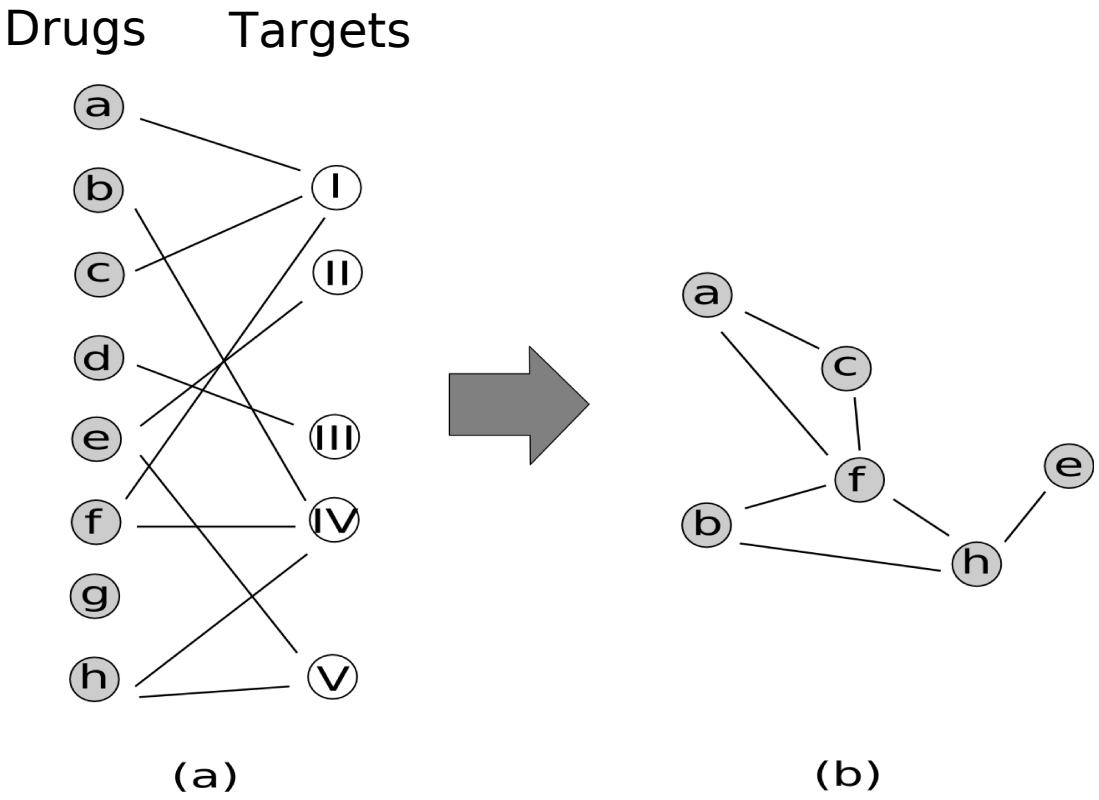
1. *Construction and integration of homogeneous pharmacological networks*

2. *Network-based algorithms to rank drugs*

How to construct meaningful pharmacological networks?

- A direct solution: a pairwise chemical structure similarity network $N_{StructSim}$
- Can we construct other more general pharmacological networks?

ΨNetPro: Pharmacological Space Integration Based on Networks Projections



Bipartite network
(e.g. drug-target)

One-mode pharm. network

Integration of pharmacological spaces

- *Max integration* (union)
- *Min integration* (intersection)
- *Average*
- *Weighted average*
- ...
- *Per edge weighted average*

Per edge weighted average

A set of n pharmacological networks: $G^d = \langle V^d, E^d \rangle, 1 \leq d \leq n$

with weights w_{ij}^d of edges $(v_v, v_j) \in E^d$

The integrated pharmacological network:

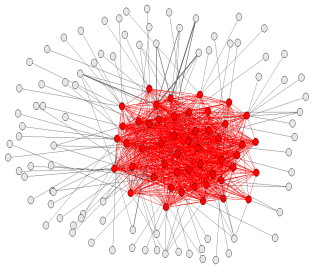
$G = \langle V, E \rangle, V = U_d V^d, E \subseteq U_d E^d$

has weights:

$$w_{ij} = 1/|D(i,j)| \sum_{d \in D(i,j)} w_{ij}^d, \qquad D(i,j) = \{ d \mid v_i \in V^d \wedge v_j \in V^d \}$$

High coverage and no penalization for drugs with a limited number of data sources

Kernelized score functions: an algorithmic scheme for ranking drugs



Drug-drug network

Kernel

Any kernel. E.g.:
- *Linear kernel*
- *Gaussian kernel*
- *Graph kernels*

Score function

Ranked drugs

Average score : $S_{AV}(v, V_C) = \frac{1}{|V_C|} \sum_{x \in V_C} K(v,x)$

kNN score : $S_{kNN}(v, V_C) = \sum_{x \in kNN(v)} K(v,x)$

NN score : $S_{NN}(v, V_C) = max_{x \in V_C} K(v,x)$

Re and Valentini

An example of graph kernel: the Random Walk kernel

- *One-step random walk kernel* (Smola and Kondor, 2003):

$$K = (a-1)\,I + D^{-1/2} W D^{-1/2}$$

Normalized Laplacian of the graph

W : weighted adjacency matrix of the graph
K : Gram matrix with elements $k_{ij} = K(v_i, v_j)$
I : identity matrix
D: diagonal matrix with $d_{ii} = \sum_j w_{ij}$

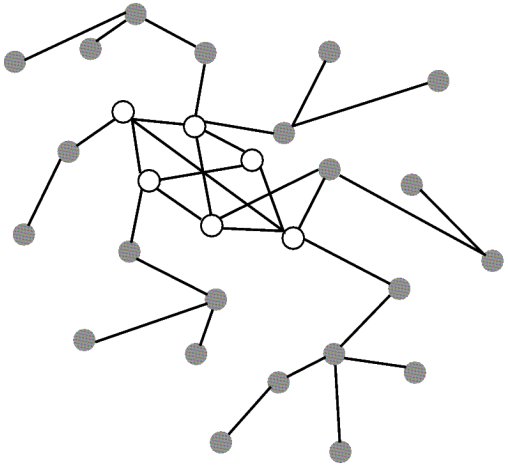
- *q-step random walk kernel:*

$$K_{q-step} = K^q$$

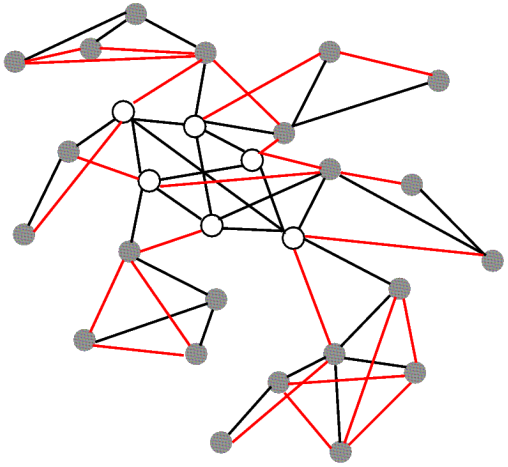
q: number of steps

By setting $q>1$ we can explore also "indirect neighbours" between drugs

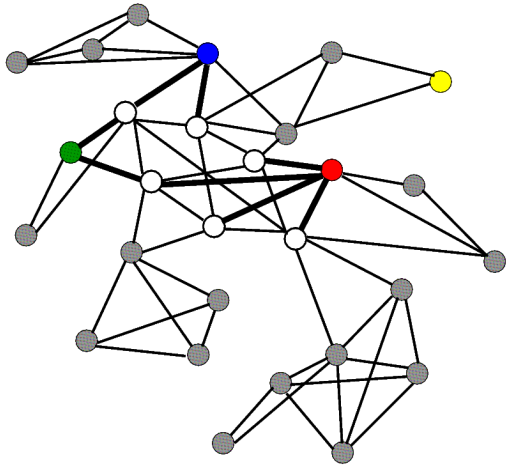
A picture of the ranking method



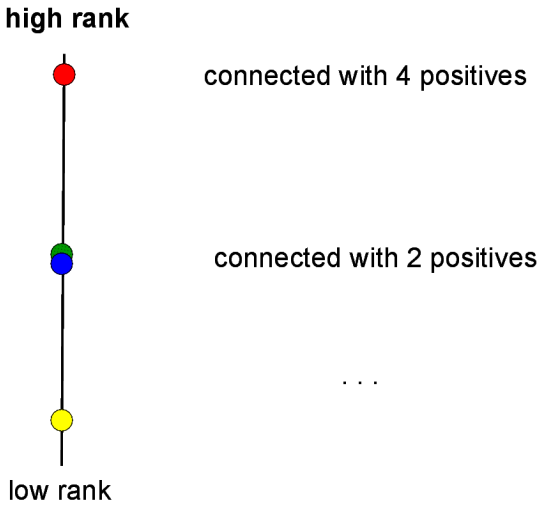
Original network



Random walk kernel



Scoring of unlabeled nodes



Experiments

- 1253 FDA approved drugs
- 51 DrugBank therapeutic classes
- 3 pharmacological networks:
  - $N_{structSim}$: pairwise chemical similarity (*Tanimoto* coefficients)
  - $N_{drugTarget}$: projection from drug-target interactions (from *DrugBank 3.0*)
  - $N_{drugChem}$: projection from chemical interactions (from *STITCH 2.0*)
- Binarization and Graph Laplacian normalization

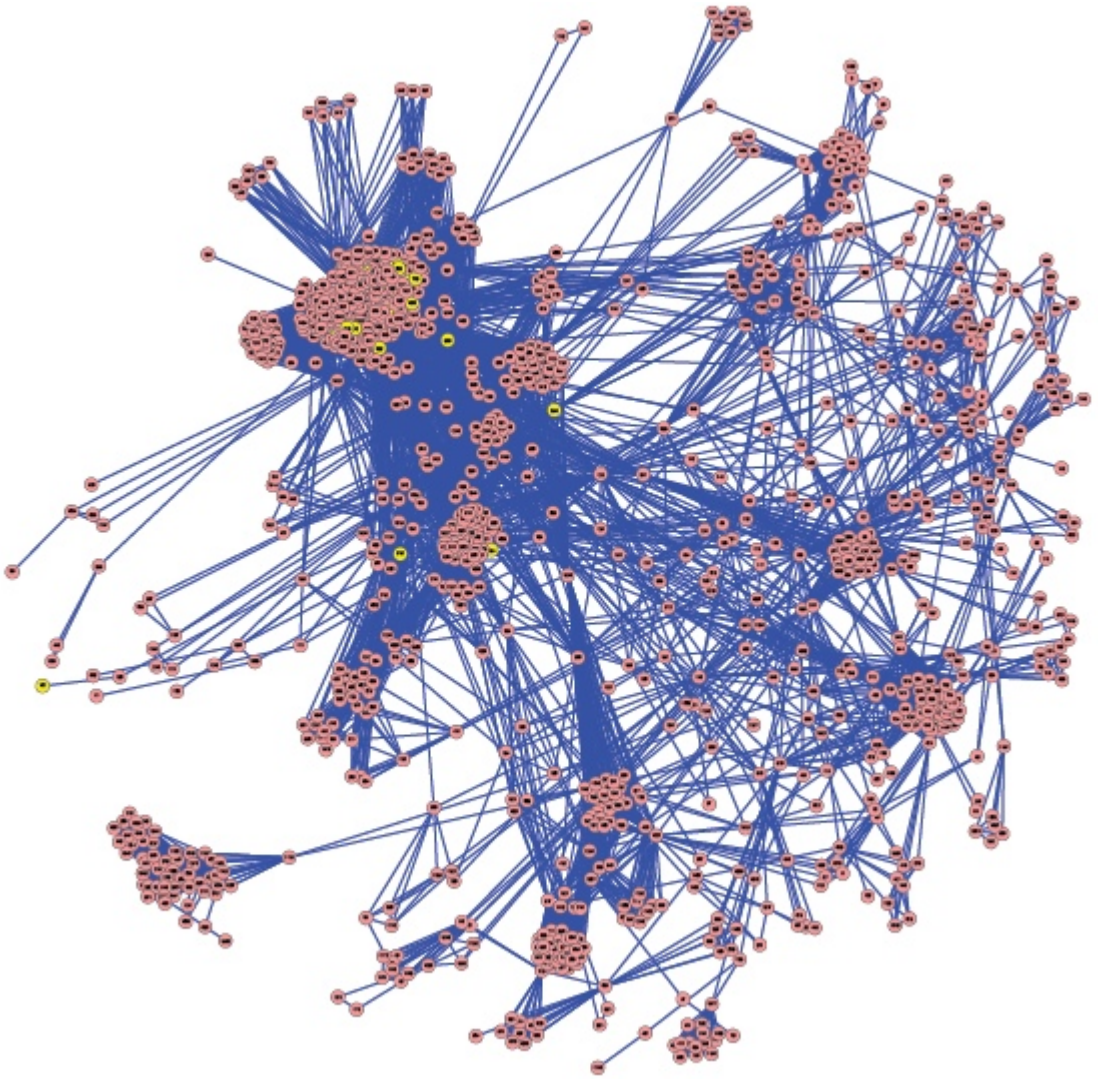
Progressive integration through "per edge" weighted average

$N_{structSim} \rightarrow W_1$ (1253 nodes, 13010 edges)

$N_{structSim} + N_{drugTarget} \rightarrow W_2$ (1253, 43827)

$N_{structSim} + N_{drugTarget} + N_{drugChem} \rightarrow W_3$ (1253, 96711)

A view of the integrated pharmacological network with Cytoscape



Results: AUC

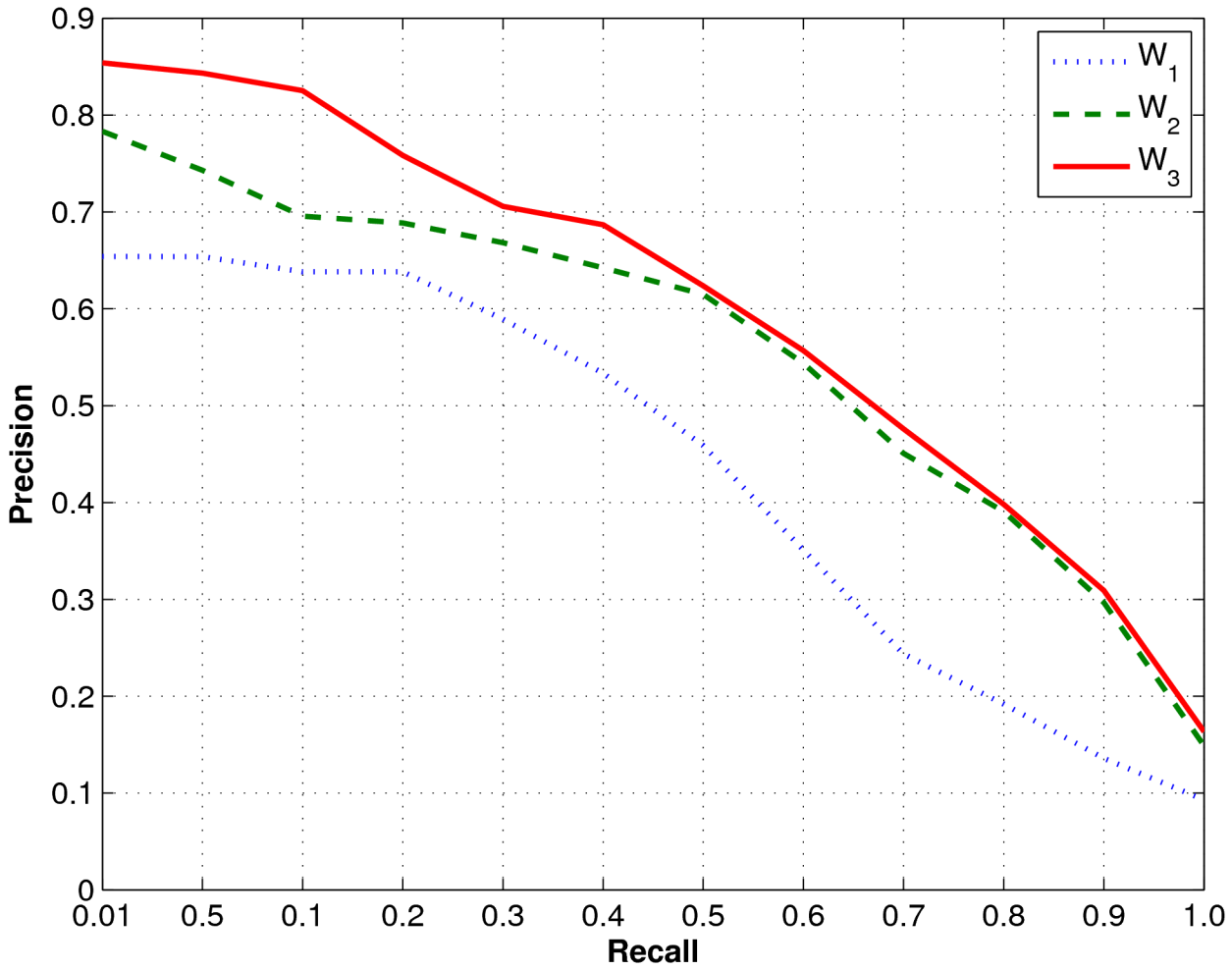
Kernelized score functions with random walk kernels compared with *Random Walk* (RW) and *Random Walk with Restart* (RWR) algorithms:

- 5-fold CV
- AUC results averaged across 51 DrugBank therapeutic classes:

	RW	RWR	S_{AV}	S_{NN}	S_{kNN}
W_1	0.6846	0.8037	0.8262	0.8074	0.8277
W_2	0.5780	0.9171	0.9232	0.9066	0.9230
W_3	0.5334	0.9258	0.9312	0.9129	0.9299

- $W_1 \rightarrow W_2 \rightarrow W_3$: AUC increments are statistically significant (Wilcoxon rank sum test, α=0.01)
- RW fails
- S_{AV} and S_{kNN} significantly better than the other methods (Wilcoxon rank sum test, α=0.01)

Results: precision at fixed recall



S_{kNN}: precision a fixed recall levels.

Time complexity

	RW	RWR	S_{AV}	S_{NN}	S_{kNN}
time (sec.)	13840	645	5	5	12

5-fold CV repeated 10 times for 51 therapeutical categories

- No model learning is required (transductive method)
- Score computation complexity : $O(|V|\ |V_c|)$

Approximately linear when $|V_c| << |V|$

Preliminary analysis of top ranked "false positives"

- *"Anti HIV agents"*: first top ranked FP is *Darunavir* (annotated in DrugBank as "*HIV Protease Inhibitor*")

- *"GABA modulators"*: *Adinazolam* and other 4 top ranked "false positives" are benzodiazepines, known to modulate the effect of GABA (*Hanson et al*, 2008)

Conclusions

- *ΨNetPro*: a general framework for the construction and integration of pharmacological spaces based on networks projections
- *Kernelized score functions*: an algorithmic scheme for ranking drugs in pharmacological networks
- Cross-validated results show that our proposed methods are able to recover *DrugBank therapeutic categories* and to potentially reuse existing drugs for novel therapeutic indications

Developments and research perspectives

1. Integration of projected one-mode pharmacological networks from different two-mode networks: e.g. annotated side-effects (*SIDER*), curated pathway DB (*Reactome*), gene expression signature repositories (*Connectivity Map*)
2. Novel algorithms from the proposed algorithmic scheme:
- novel distance measures and score functions
- design of novel kernels well suited to the topology of the drug-drug networks
3. Low complexity of the algorithm: applicability to thousands of investigational compounds (not only FDA approved drugs)
4. Experimenting with different variants of network projections and integration
5. Systematic analysis of top ranked "false positive" drugs extended to all the therapeutic categories, or using other taxonomies (supported by text mining and text disambiguation techniques?)

Thank you for your attention!

Matteo Re

Giorgio Valentini