Drug Side-Effects Predictions with Random Walk with Restart

Suchanuch Piriyasatit

TBSI, Tsinghua University, China

ABSTRACT

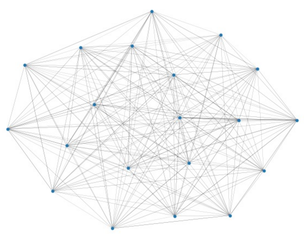
- **Heterogeneous network of drug and side-effects** is constructed based on side-effect similarity, drug chemical structures, and drug protein targets similarity.
- **Random walk with restart** is applied to the heterogeneous network of drug and side-effects to get the **ranking of possible side-effect candidates** of a drug from **the steady-state probability distribution**.

INTRODUCTION

- **Side effects** are considered as disturbances in the form of **molecular intercommunication** such as protein-protein communication or signal pathways.
- Predicting drug side effects then requires **network information** on the drug compounds, protein targets, and the side effects.
- **Random walk** is the state of the art network analysis method that relies on '**guilt by association**' principle: **molecules that are closer in the network tend to have similar properties.**
- **Random walk with restart** introduces the restart probability of walking back to the seed nodes at each timestep to capture the **local neighborhood** of the seed nodes as well as the **network global structure.**

METHOD

Side-Effect Network:



Similarity Matrix:

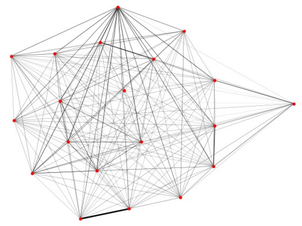
$$S_s(i,j) = \frac{|\Gamma_d(i) \cap \Gamma_d(j)|}{|\Gamma_d(i) \cup \Gamma_d(j)|}$$

$$\bar{S}_s = D_s^{-\frac{1}{2}} S_s D_s^{-\frac{1}{2}}$$

Transition Matrix:

$$M_{ss}(i,j) = p(s_j|s_i) = \frac{(1-\lambda)\bar{S}_s(i,j)}{\sum_k \bar{S}_s(i,k)}$$

Drug Network:



Similarity Matrix:

$$S_d^t(i,j) = \frac{|\Gamma_t(i) \cap \Gamma_t(j)|}{|\Gamma_t(i) \cup \Gamma_t(j)|}$$

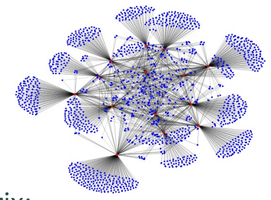
$$S_d^c(i,j) = \text{DiceSimilarity}(i,j)$$

$$S_{dd} = w_d \bar{S}_d^c + (1-w_d)\bar{S}_d^t$$

Transition Matrix:

$$M_{dd}(i,j) = p(d_j|d_i) = \begin{cases} \frac{S_{dd}(i,j)}{\sum_k S_{dd}(i,k)}, & \text{if } \sum_k A_{dd}(i,k) = 0 \\ \frac{(1-\lambda)S_{dd}(i,j)}{\sum_k S_{dd}(i,k)}, & \text{otherwise} \end{cases}$$

Drug-Side Effect Network:



Transition Matrix:

$$M_{sd}(i,j) = p(d_j|s_i) = \begin{cases} \frac{\lambda A_{sd}(j,i)}{\sum_k A_{sd}(k,i)}, & \text{if } \sum_k A_{sd}(k,i) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$M_{ds}(i,j) = p(s_j|d_i) = \begin{cases} \frac{\lambda A_{ds}(i,j)}{\sum_k A_{ds}(i,k)}, & \text{if } \sum_k A_{ds}(i,k) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

Transition matrix of the heterogeneous network:

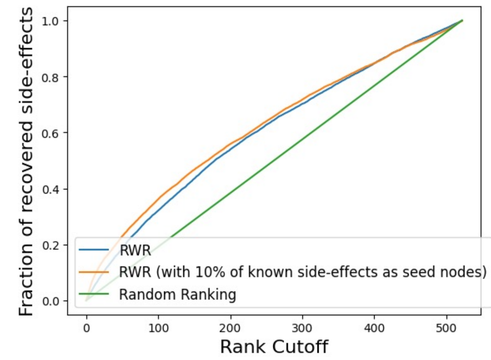
$$M = \begin{bmatrix} M_{ss} & M_{sd} \\ M_{ds} & M_{dd} \end{bmatrix}$$

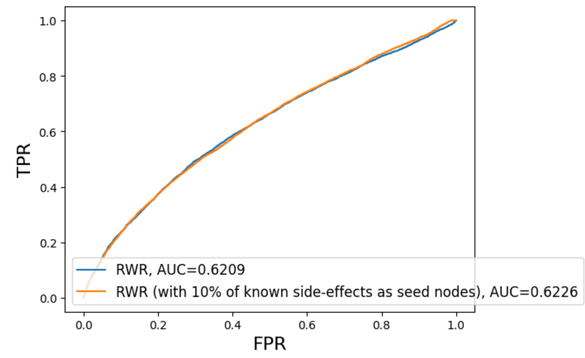
Random Walk with Restart:

Iterative Update: $p_{t+1} = (1-\alpha)M^T p_t + \alpha p_0$

Analytical Solution: $p = \alpha(I - (1-\alpha)M^T)^{-1} p_0$

RESULTS





0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.6162	0.6209	0.6237	0.6254	0.6264	0.6270	0.6274	0.6275	0.6276

AUC of different restart probability values

CONCLUSION

- We integrate multiple network information to construct a heterogeneous network and apply the random walk with restart starting from the query drug.
- The side-effect ranking obtained is better than the random ranking, showing the potential of using biological network data for side-effect prediction.
- Performance could be improved by obtaining more relevant biological network information to better define the transition matrix.

REFERENCES

[1] S. R. Atias N, "An algorithmic framework for predicting side effects of drugs," J Comput Biol, vol. 18, p. 207, 2011.

[2] X. Chen, M. Liu, and G. Yan, "Drug-target interaction prediction by random walk on the heterogeneous network," Molecular bioSystems, vol. 8, pp. 1970–8, 04 2012.