

Supplementary Material

1 Analysis of some details of LRSSL algorithm

1.1 Algorithm complexity

Here we discuss the algorithm complexity of LRSSL. Suppose the number of the drugs in the training dataset is n , the number of drug feature profiles is m , the largest dimension of drug features is d , the number of diseases is c . The complexity for computing graph Laplacian matrix L is $\mathcal{O}(mn^2d)$. For each iteration, the optimization of projection matrix G has the complexity of $\mathcal{O}(m^2n^2d + m^2n^2c + md^2c)$, the complexity for learning F is $\mathcal{O}(mdc + n^2c)$. As in the training stage of our study, d is much larger than m , n and c , the most complex part in each iteration is about $\mathcal{O}(d^2)$.

1.2 Convergence of algorithm

In this subsection, we prove the convergence of the optimization algorithm of LRSSL. When F is fixed, the objective function becomes the following problem:

$$\begin{aligned} \min_{G_p} & Tr \left(\sum_{p=1}^m G_p^T X_p (\mu I - \mu^2 P^T) X_p^T G_p \right. \\ & \left. - 2\mu Y^T P^T \sum_{p=1}^m X_p^T G_p - \mu^2 \sum_{p=1}^m \sum_{q \neq p}^m G_p^T X_p P^T X_q^T G_q \right) \\ & + \lambda \sum_{p=1}^m \|e_{1 \times d_p} G_p\|_F^2, \quad s.t. \quad G_p \geq 0 \end{aligned} \quad (1.1)$$

Let

$$A_p = X_p (\mu I - \mu^2 P^T) X_p^T + \lambda e_{1 \times d_p}^T e_{1 \times d_p} \quad (1.2)$$

$$B_p = \mu X_p P Y + \mu^2 \sum_{q \neq p}^m X_p P^T X_q^T G_q \quad (1.3)$$

The objective function (1.1) becomes:

$$\min_{G_p} \sum_{p=1}^m Tr(G_p^T A_p G_p) - 2Tr(G_p^T B_p) \quad s.t. \quad G_p \geq 0 \quad (1.4)$$

G_p is updated by:

$$G_p(i, j) \leftarrow G_p(i, j) \sqrt{\frac{(A_p^- G_p)(i, j) + (B_p^+)(i, j)}{(A_p^+ G_p)(i, j) + (B_p^-)(i, j)}}, \quad (1.5)$$

To prove the convergence, we have to show that the update rule of G_p will monotonically decrease the objective function (1.4) when F is fixed. The auxiliary function approach is employed and the prove process is similar to [2].

Theorem 1.1. *Let*

$$H(G) = \text{Tr}(G^T AG) - 2\text{Tr}(G^T B) \quad (1.6)$$

Then the following function $h(G, \tilde{G})$

$$\begin{aligned} h(G, \tilde{G}) = & - \sum_{i,j} 2(B^+)(i, j)\tilde{G}(i, j)\left(1 + \log \frac{G(i, j)}{\tilde{G}(i, j)}\right) + \sum_{i,j} (B^-)(i, j)\frac{G(i, j)^2 + \tilde{G}(i, j)^2}{\tilde{G}(i, j)} \\ & + \sum_{i,j} \frac{(A^+\tilde{G})(i, j)G(i, j)^2}{\tilde{G}(i, j)} - \sum_{i,j,k} (A^-)(j, k)\tilde{G}(i, j)\tilde{G}(i, k)\left(1 + \log \frac{G(i, j)G(i, k)}{\tilde{G}(i, j)\tilde{G}(i, k)}\right) \end{aligned} \quad (1.7)$$

is an auxiliary function for $H(G)$; that is, $H(G) \leq h(G, \tilde{G})$ and $H(G) = h(G, G)$. $h(G, \tilde{G})$ is convex and its global minimum is

$$G(i, j) = \underset{G}{\operatorname{argmin}} h(G, \tilde{G}) = \tilde{G} \sqrt{\frac{(A^-\tilde{G})(i, j) + (B^+)(i, j)}{(A^+\tilde{G})(i, j) + (B^-)(i, j)}} \quad (1.8)$$

Proof. The function $H(G)$ is

$$H(G) = \text{Tr}(-2G^T B^+ + 2G^T B^- + G^T A^+ G - G^T A^- G) \quad (1.9)$$

We have

$$\text{Tr}(G^T A^+ G) \leq \sum_{i,j} \frac{(A^+\tilde{G})(i, j)G(i, j)^2}{\tilde{G}(i, j)} \quad (1.10)$$

Because $a \leq (a^2 + b^2)/2b, a, b > 0$, then

$$\text{Tr}(G^T B^-) = \sum_{i,j} G(i, j)(B^-)(i, j) \leq \sum_{i,j} (B^-)(i, j)\frac{G(i, j)^2 + \tilde{G}(i, j)^2}{2\tilde{G}(i, j)} \quad (1.11)$$

As $x \geq 1 + \log(x)$ for $x > 0$, we obtain

$$\frac{G(i, j)}{\tilde{G}(i, j)} \geq 1 + \log \frac{G(i, j)}{\tilde{G}(i, j)} \quad (1.12)$$

and

$$\frac{G(i, j)G(i, k)}{\tilde{G}(i, j)\tilde{G}(i, k)} \geq 1 + \log \frac{G(i, j)G(i, k)}{\tilde{G}(i, j)\tilde{G}(i, k)} \quad (1.13)$$

then

$$\text{Tr}(G^T B^+) = \sum_{i,j} (B^+)(i, j)G(i, j) \geq \sum_{i,j} (B^+)(i, j)\tilde{G}(i, j)\left(1 + \log \frac{G(i, j)}{\tilde{G}(i, j)}\right) \quad (1.14)$$

and

$$\text{Tr}(GA^- G^T) \geq \sum_{i,j,k} (A^-)(j, k)\tilde{G}(i, j)\tilde{G}(i, k)\left(1 + \log \frac{G(i, j)G(i, k)}{\tilde{G}(i, j)\tilde{G}(i, k)}\right) \quad (1.15)$$

By summing over (1.10), (1.11), (1.14) and (1.15), we get $H(G) \leq h(G, \tilde{G})$ and $H(G) = h(G, G)$.

According to [?], the Hessian matrix of $h(G, \tilde{G})$ is a diagonal matrix with positive diagonal values. Thus, $h(G, \tilde{G})$ is a convex function and the global minimum is obtained by setting the derivative of $h(G, \tilde{G})$ to zeros, which is (1.8). \square

Theorem 1.2. *The objective function of (1.4) is monotonically decreased by updating G_p as (1.5).*

Proof. According to the definition of auxiliary function, we can obtain:

$$H(G_p^{(t)}) = h(G_p^{(t)}, G_p^{(t)}) \geq h(G_p^{(t+1)}, G_p^{(t)}) \geq H(G_p^{(t+1)}) \quad (1.16)$$

Thus $H(G_p^{(t)})$ is monotonic decreasing. \square

From theorem 1.1 and theorem 1.2, the convergence of the proposed method is proved. We also verify the convergence of LRSSL by experiment. Figure S1 shows that the algorithm is rapidly converged after a few of iterations. We also examine the influence of the initiation of G_p , and find there is no significant effect for the performance of the method(Figure S8) but the convergent rate become slower when G_p is initiated from large interval(Figure S9).

2 Implementation details of the compared methods

L1LOG and L1SVM are trained using LIBLINEAR[3]. The parameter C in L1LOG is set to 5000 and the parameter C in L1SVM is set to 8000, which are selected by cross validation. The negative samples in these two methods are randomly selected from the unlabelled data in the training set, and the size of the negative samples is the same as the positive samples in the training set. For PreDR, Chemical similarity, target similarity and side-effect similarity of drugs are calculated in the same way as [4]. SVM in PreDR is trained using LIBSVM[1]. The parameters of SVM are selected by grid search, the parameter C is set to $\sqrt{2}$, the parameter γ is set to 4. The randomly selected negative samples are twice the size of the training positives. For TL-HGBI, the drug-drug similarity is calculated based on the Tanimoto score of the fingerprints of drugs. The target-target similarity is calculated using the Smith-Waterman algorithm and normalized. The parameters of TL-HGBI are optimized by grid search approach, the decay factor is set to 0.1 and the similarity threshold is set to 0.4.

3 Methods for the measurement of correlations between the feature weights of LRSSL and the frequencies of disease related features

To explore the potential meanings of the weights in G_p systematically, we firstly get the chemicals and genes from CTD database which are directly related to the diseases, and the drugs and targets that are used to build our model are excluded. Then the chemical substructures, protein domains and gene ontology terms of these disease related chemicals and genes are extracted, these features are defined as disease related features from database and the frequencies of appearance of the features are calculated. After that we measure the spearman's rank correlations between the weights in our method and the frequencies of disease related features respectively. We also calculate the correlations between the weights and the frequencies of features from the randomly selected chemicals and genes (the number of random chemicals and genes linked to each disease is the same as the disease related chemicals and genes) and the correlations between random weights (from uniform distribution on the interval [0,1]) and the frequencies of features of the disease related chemicals and genes. The results are shown in Figure S6.

References

- [1] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May 2011.
- [2] C. H. Q. Ding, T. Li, and M. I. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):45–55, January 2010.
- [3] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *J. Mach. Learn. Res.*, 9:1871–1874, June 2008.
- [4] Yongcui Wang, Shilong Chen, Naiyang Deng, and Yong Wang. Drug repositioning by kernel-based integration of molecular structure, molecular activity, and phenotype data. *PloS One*, 8(11):e78518, 2013.

Table S1: Prediction Performance of LRSSL using single source of drug features

Metric	Chemical Substructure Features			Target Domain Features	Target Annotation Features	Without Disease Similarity
	PubChem fingerprints(881 bit)	CDK standard fingerprints(1024 bit)	MACCS fingerprints(166bit)			
AUC	0.8360 ± 0.0027	0.8357 ± 0.0024	0.8221 ± 0.0027	0.8937 ± 0.0036	0.9055 ± 0.0027	0.9067 ± 0.0022
MAP	0.1549 ± 0.0018	0.1534 ± 0.0018	0.1424 ± 0.0019	0.3021 ± 0.0050	0.3039 ± 0.0043	0.3495 ± 0.0045
HLU	24.08 ± 0.24	24.10 ± 0.27	21.94 ± 0.16	39.11 ± 0.50	39.93 ± 0.39	43.97 ± 0.28
Precision at top-20	0.0635 ± 0.0008	0.0638 ± 0.0010	0.0601 ± 0.0008	0.1022 ± 0.0009	0.1028 ± 0.0009	0.1121 ± 0.0010

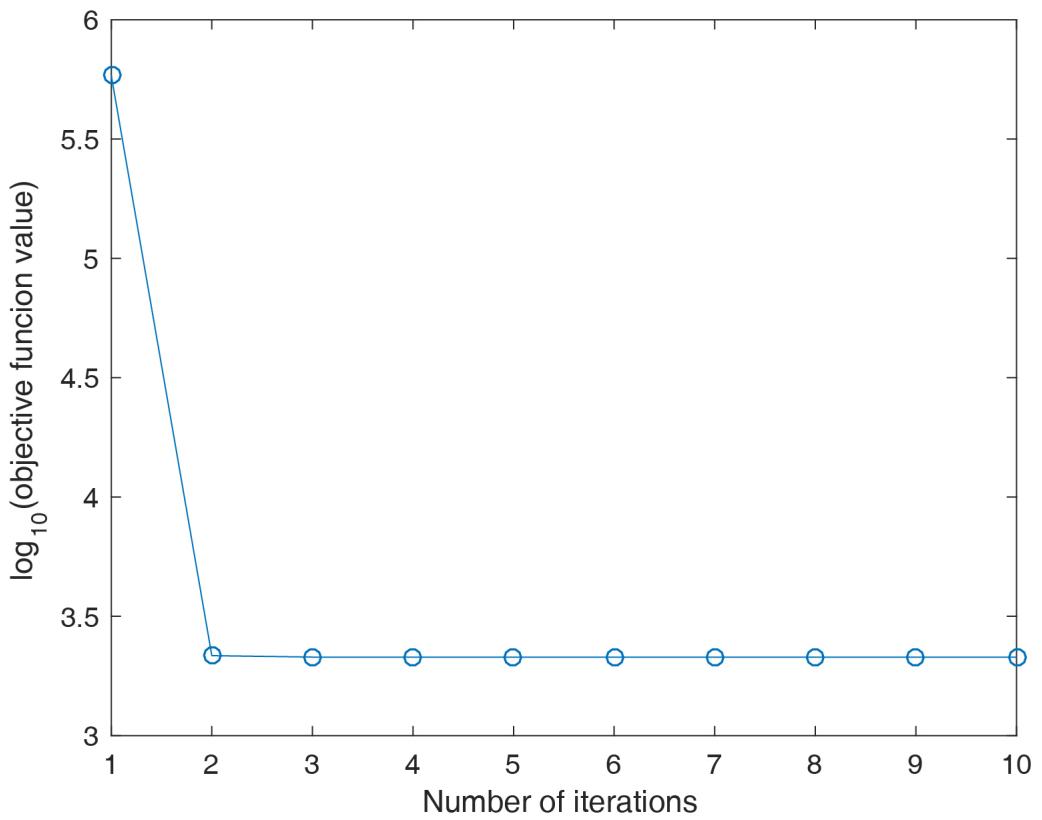


Figure S1: Convergence curve of the objective function value during iteration.

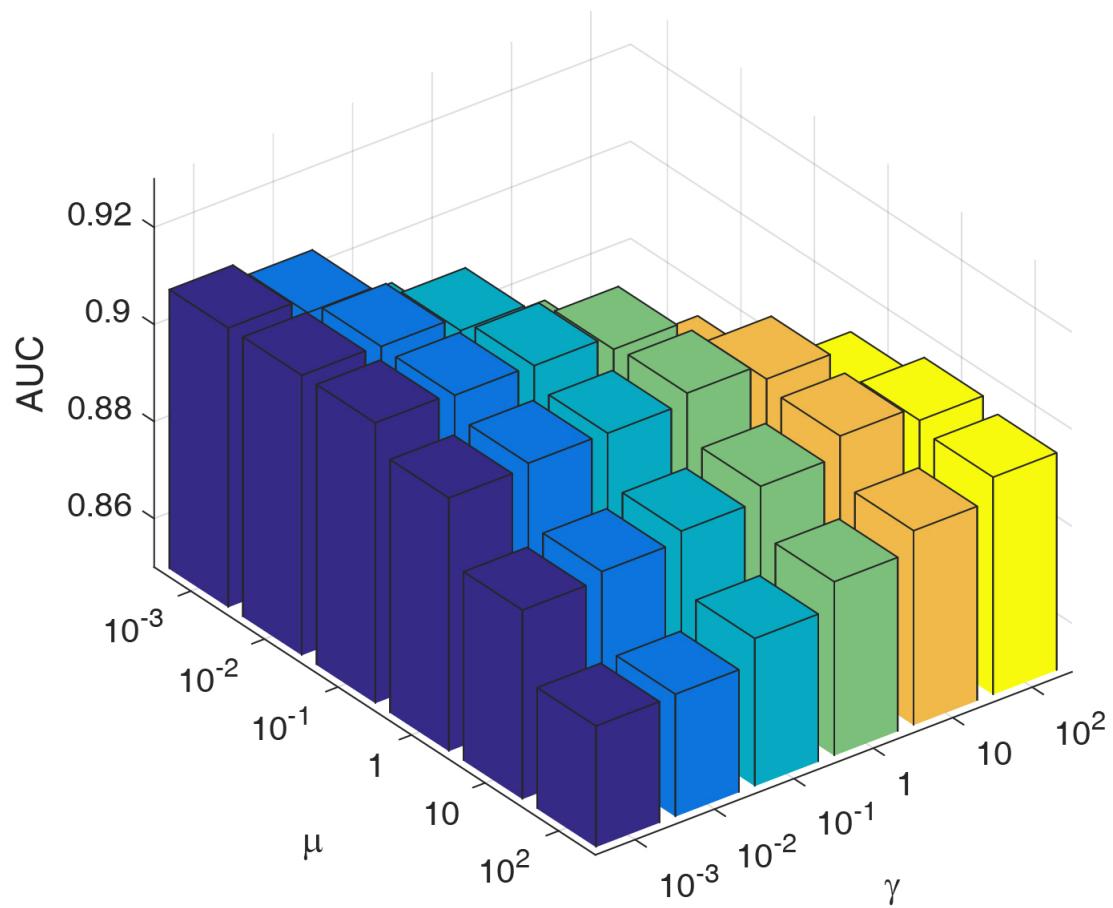


Figure S2: Drug indication predication performance varies with different values of μ and λ .

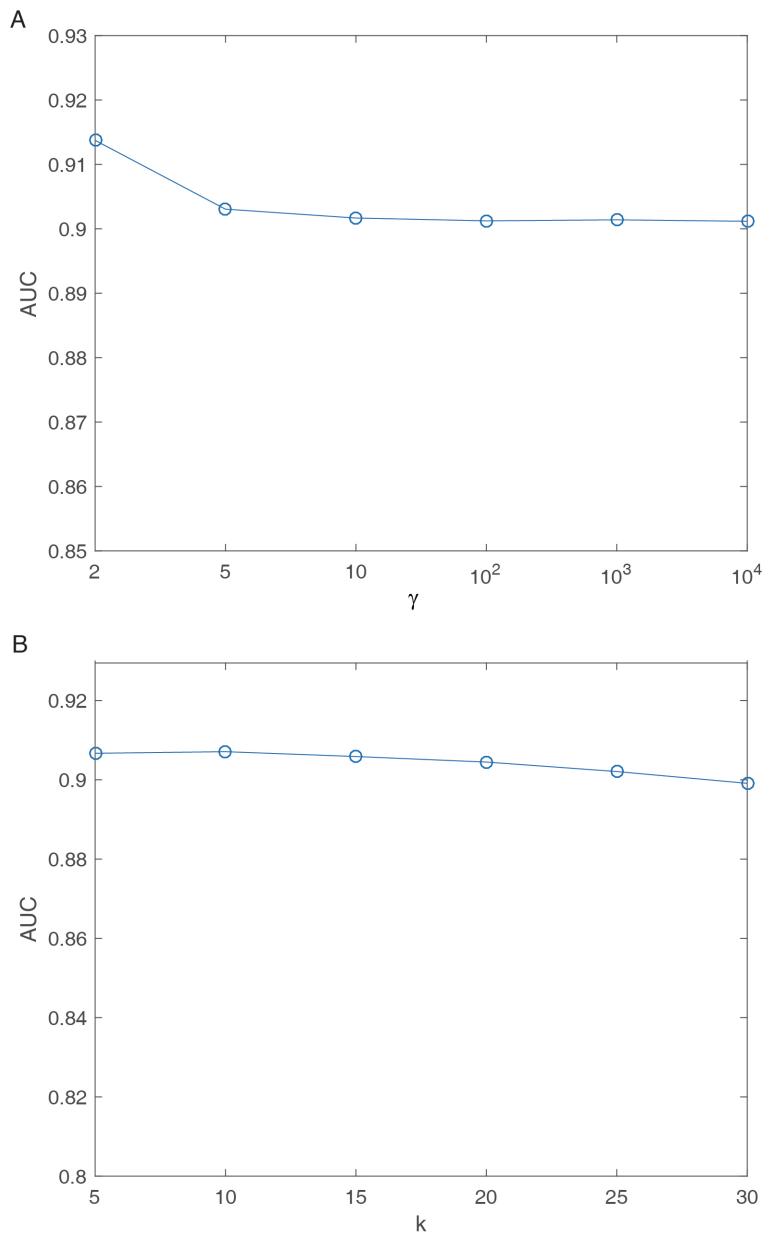


Figure S3: The impact of parameter γ and k on drug indication predication performance.

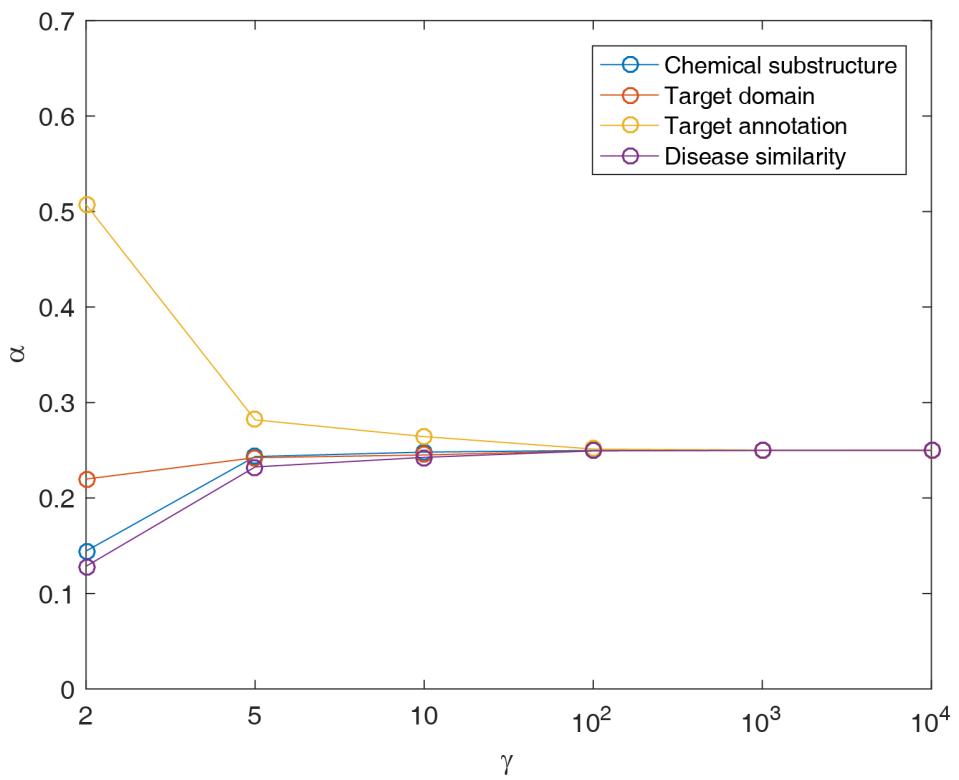


Figure S4: The impact of parameter γ on the values of α .

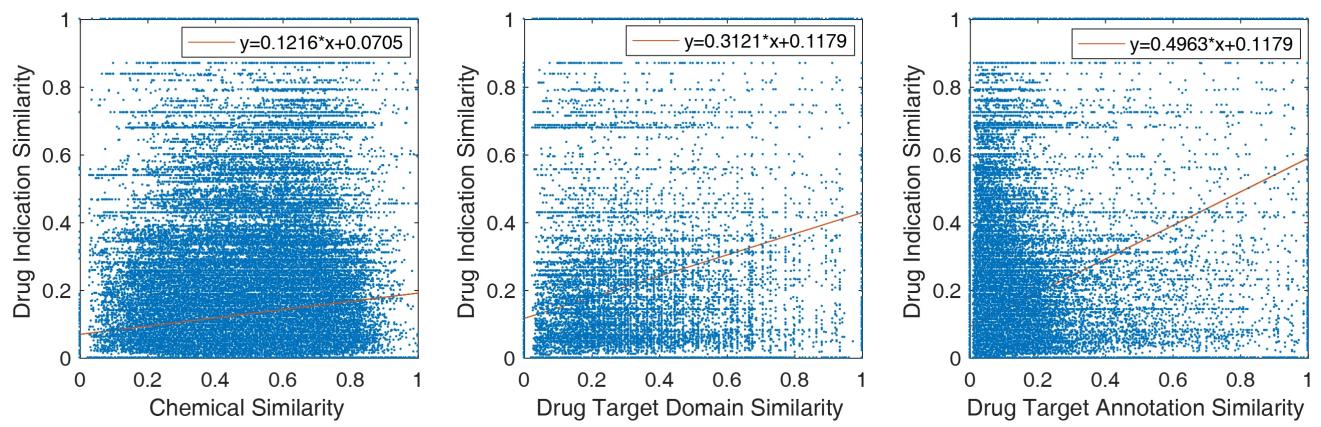


Figure S5: Regression of the drug feature similarity against the drug indication similarity.

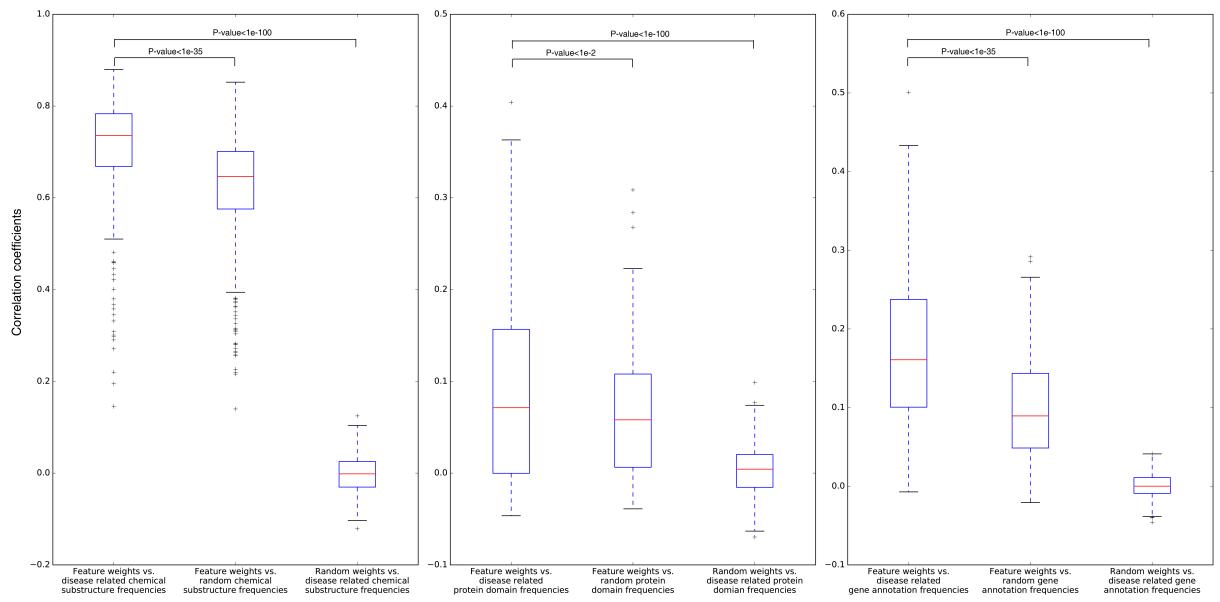


Figure S6: Distribution of correlation coefficients between feature weights and disease related feature frequencies. Left panel: drug chemical substructure; middle panel: target domain; right panel: target annotation. In all three panel, the correlation coefficients between feature weights of LRSSL and disease related feature frequencies are significantly larger than the other two groups of random control(rank sum test, $P\text{-value} < 0.01$)

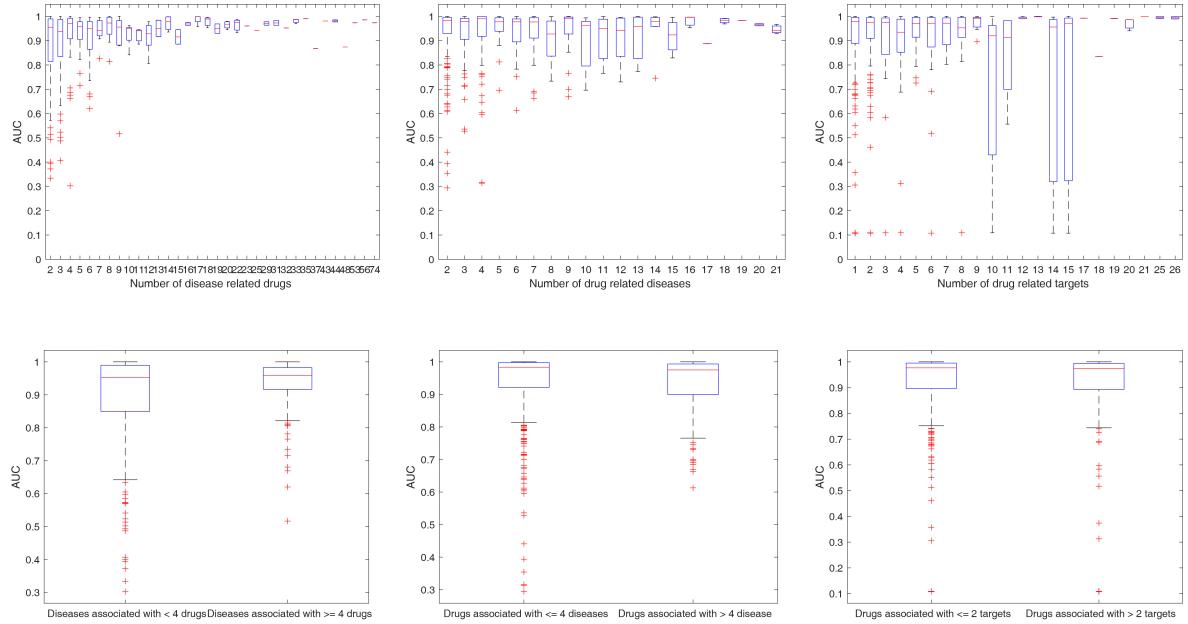


Figure S7: The relationships between the performance of LRSSL and the number of known drug-disease or drug-target associations. In the lower panels, the AUC scores are grouped based on the medians of the number of known associations, and no significant difference is found (rank sum test, P-value > 0.05)

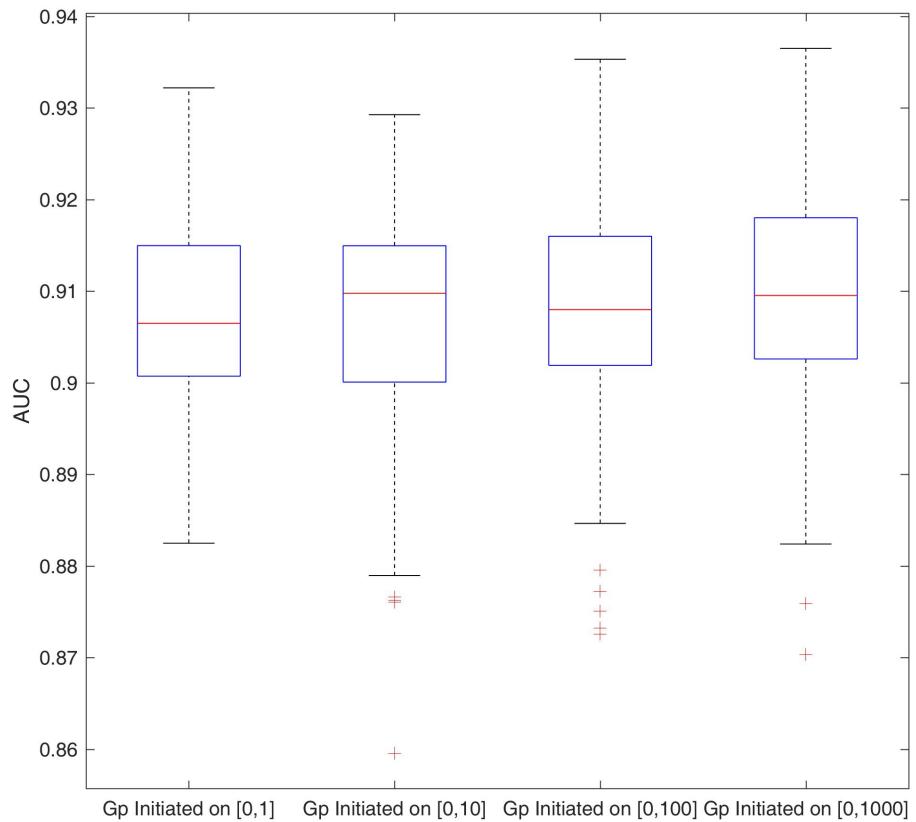


Figure S8: The relationships between the performance of LRSSL and the initiation intervals of G_p . For each interval, five-fold cross-validation is repeated for 20 times. There is no statistically significant difference in AUC values(Kruskal-Wallis Test, P-value=0.4107)

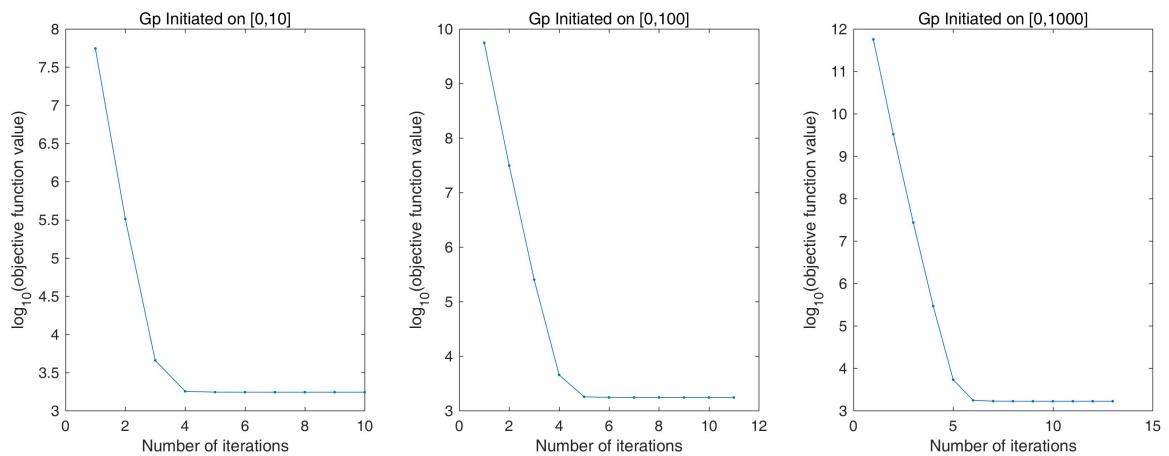


Figure S9: The relationships between the converge rate of LRSSL and the initiation intervals of G_p .