Deep mining heterogeneous networks of biomedical linked data to predict novel drug–target associations

赵连敏

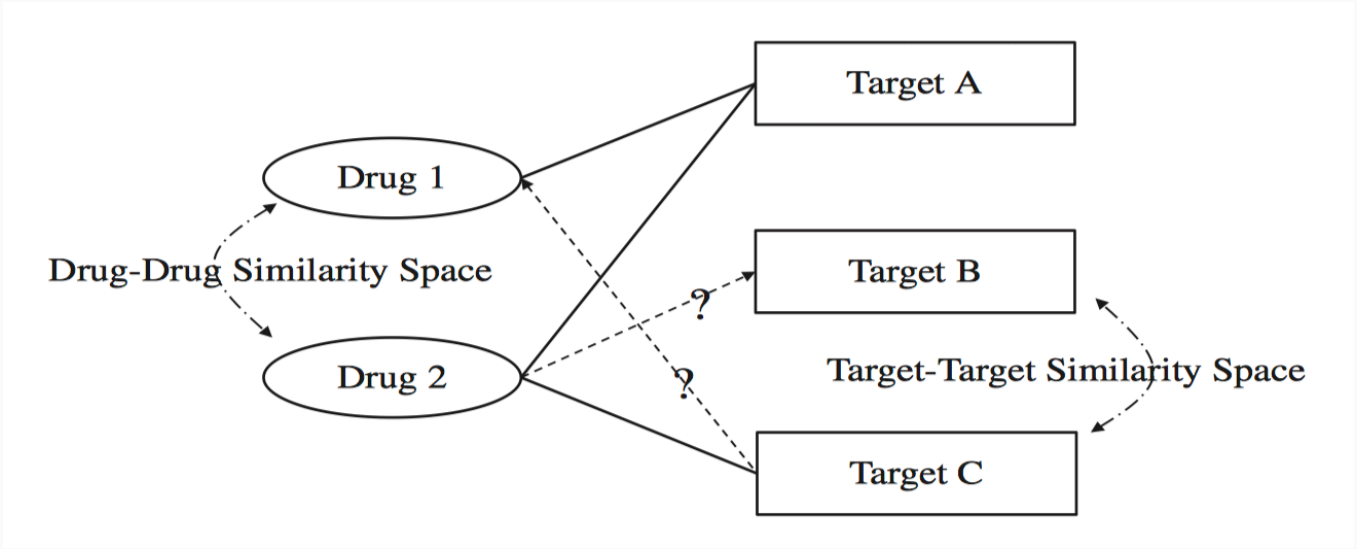
2018 年 8 月 22 日

厦门大学信息科学与技术学院

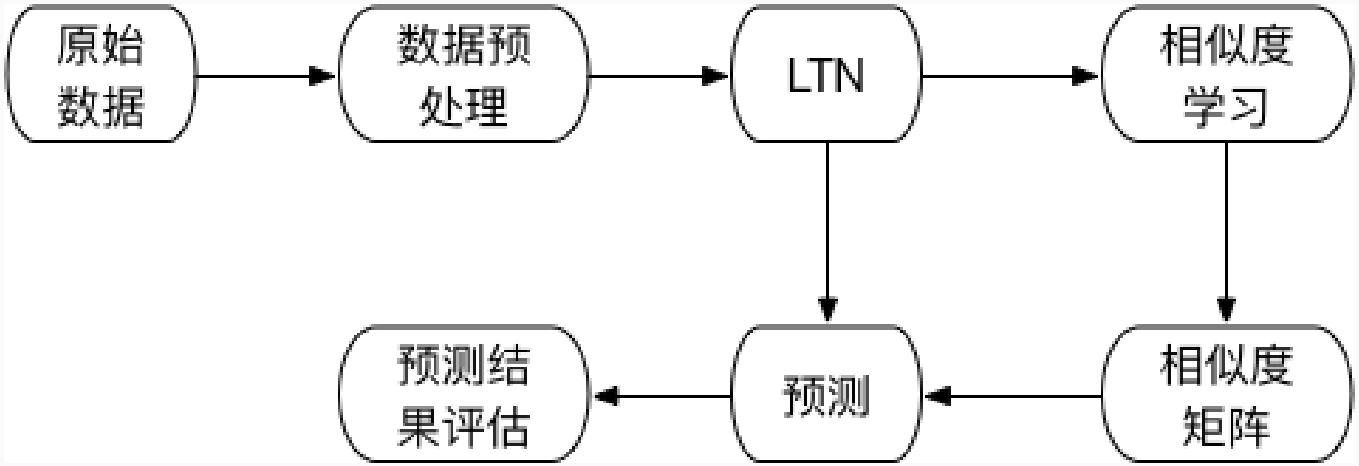
摘要

- 课题来源:
  - 基因异构网络有着丰富的信息，但尚未被用于用于预测药物特征——靶向关系预测
  - 深度学习在大型网络处理方面有着巨大的优势
- 研究成果:
  - 其性能优于其余基于拓扑结构的相似性比对方法
  - 该方法处理三元信息优于二元信息
  - 使用网络拓扑结构度量优于使用药物分子结构或基因序列信息

绪论


Target A
Drug 1
Drug-Drug Similarity Space
Target B
?
Drug 2
?
Target-Target Similarity Space
Target C

Materials and methods



相似性计算算法详解

1. 随机选取图节点
2. 从该节点生成随机序列（RandomWalk）
3. 由此序列优化迭代表示矩阵（SkipGram）
4. 相似度计算

DeepWalk

优化目标：

$$P(\{v_{i-w}, ..., v_{i+w}\}/v_i|\Phi(v_i)) = \prod_{j=i-w, j\neq i}^{i+w} P(v_j|\Phi(v_i))$$

针对其中每个节点，可以转化为

$$P(v_j|\Phi(v_i)) = \prod_{l=1}^{\lceil log|V| \rceil} \frac{1}{1 + e^{(-\Phi(v_i)\cdot\Psi(b_l))}}$$

相似度：

$$sim(u, v) = \frac{u \cdot v}{|u||v|}$$

Association discovering

DBSI:

$$confidence_{DBSI}(d_i, t_j) = \frac{\sum_{l=1, l\neq i}^{n} sim(d_i, d_l) a_{l,j}}{\sum_{l=1, l\neq i}^{n} sim(d_i, d_l)}$$

TBSI:

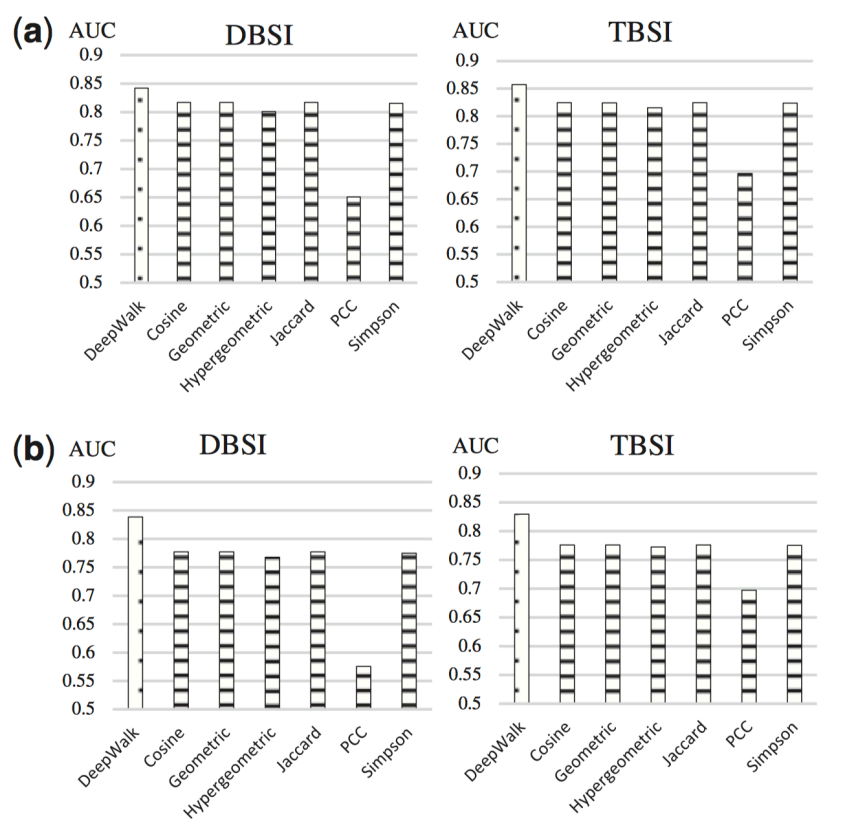
$$confidence_{TBSI}(d_i, t_j) = \frac{\sum_{l=1, l\neq j}^{m} sim(t_j, t_l) a_{i,l}}{\sum_{l=1, l\neq i}^{m} sim(t_j, t_l)}$$

标准化处理:

$$nomarlizedConfidence_{DBSI}(d_i, t_j) = \frac{confidence_{DBSI}(d_i, t_j) - Max(d_i, \cdot)}{Max(d_i, \cdot) - Min(d_i, \cdot)}$$

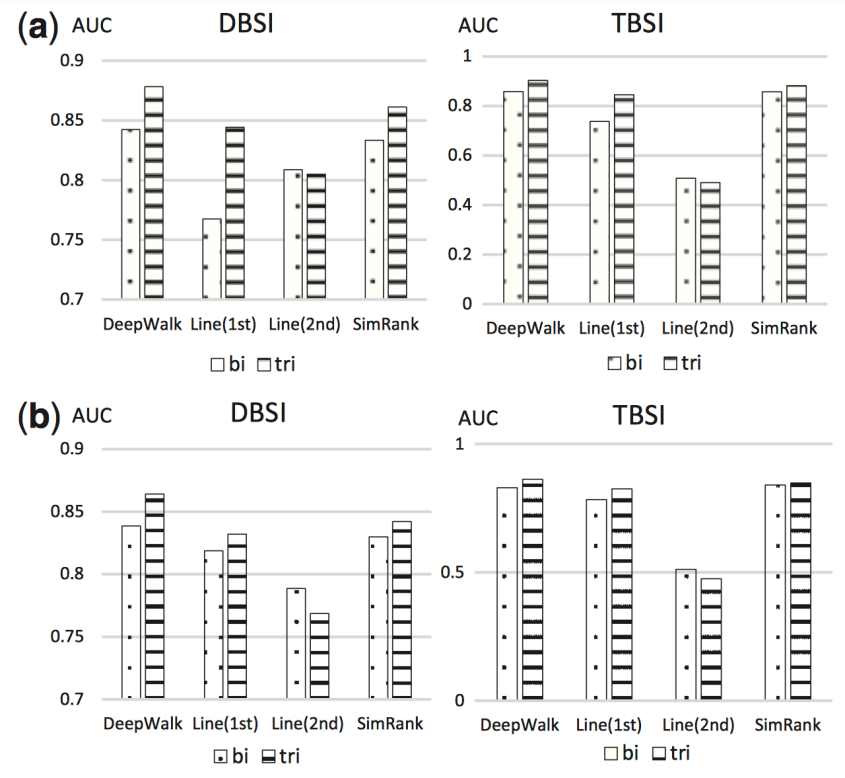
$$nomarlizedConfidence_{TBSI}(d_i, t_j) = \frac{confidence_{TBSI}(d_i, t_j) - Max(\cdot, t_j)}{Max(\cdot, t_j) - Min(\cdot, t_j)}$$

结果分析



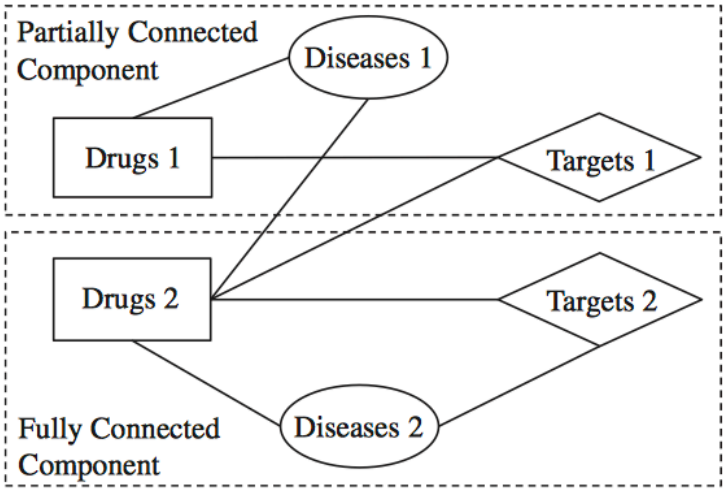
二元信息网络使用交叉验证与外部信息的 AUC 比对

结果分析



二元与三元信息网络使用交叉验证与外部信息的 AUC 比对

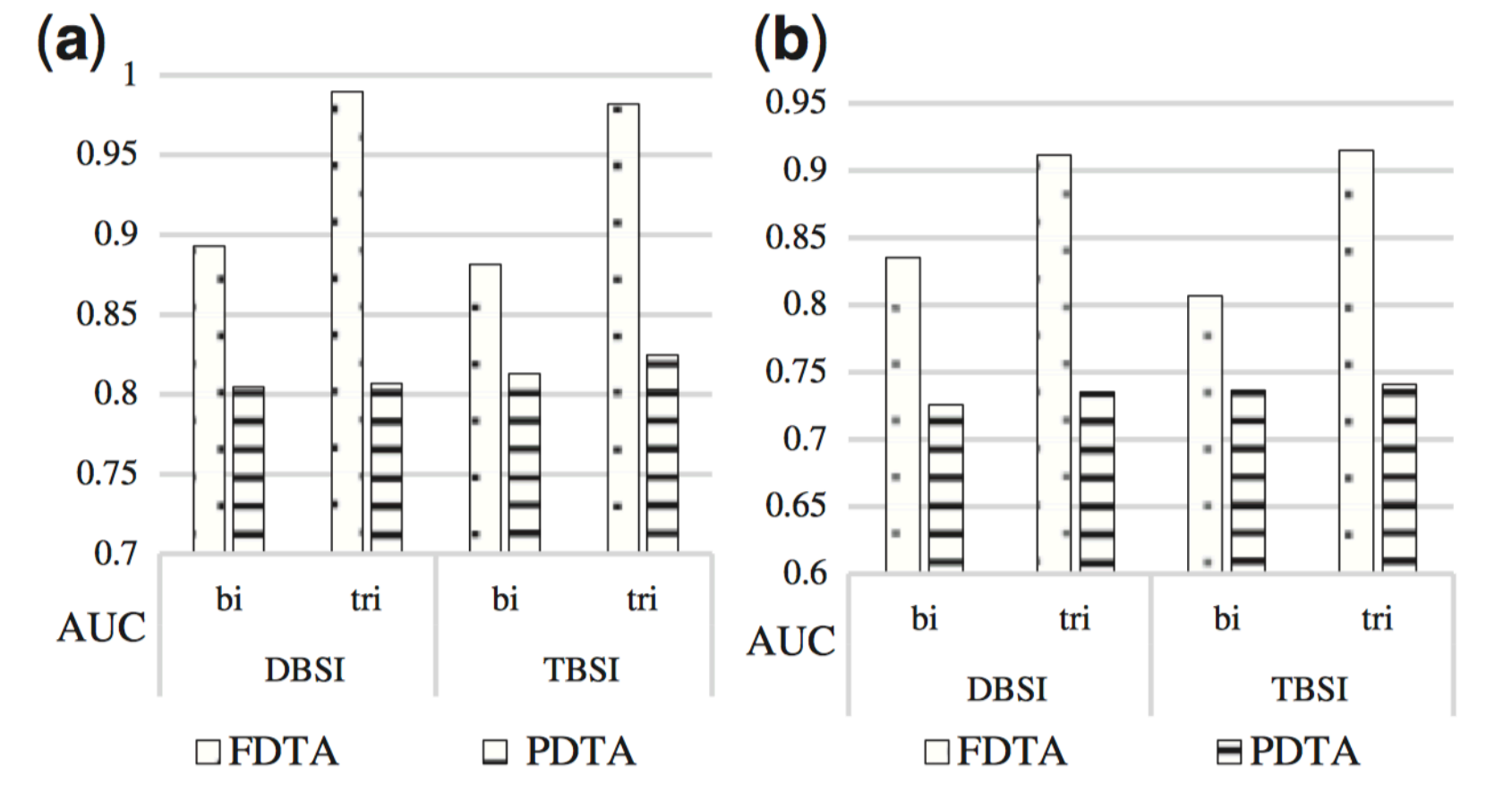
结果分析



		Partially Connected Component	Fully Connected Component	Total
Vertices	Drugs	3067	1341	4408
	Targets	4040	513	4553
	Diseases	1390	62	1452
Edges	Drug-Target	4844	2154	6998
	Drug-Disease	63	7969	8032
	Disease-Target	0	1684	1684

区分 FDTA 与 PDTA

结果分析



DeepWalk 方法数据汇总

结果分析

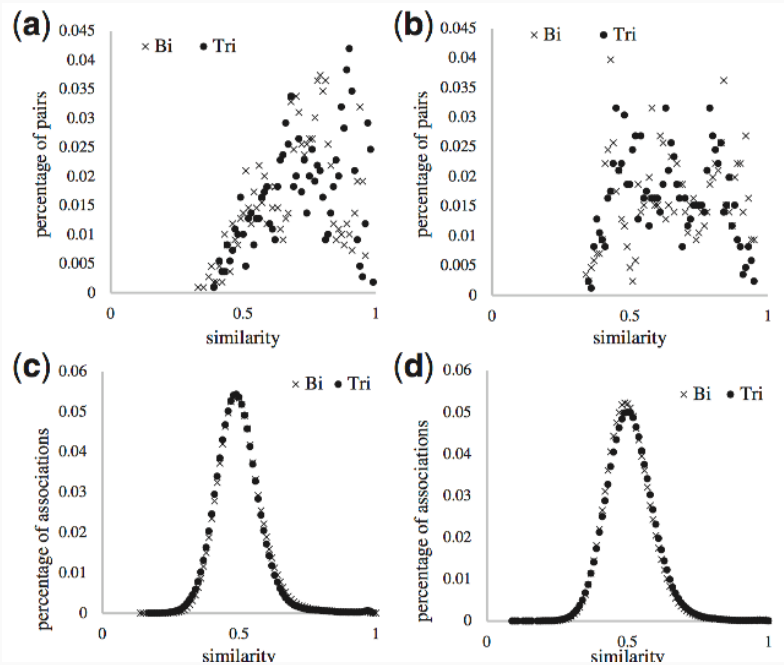


图 1: DBSI

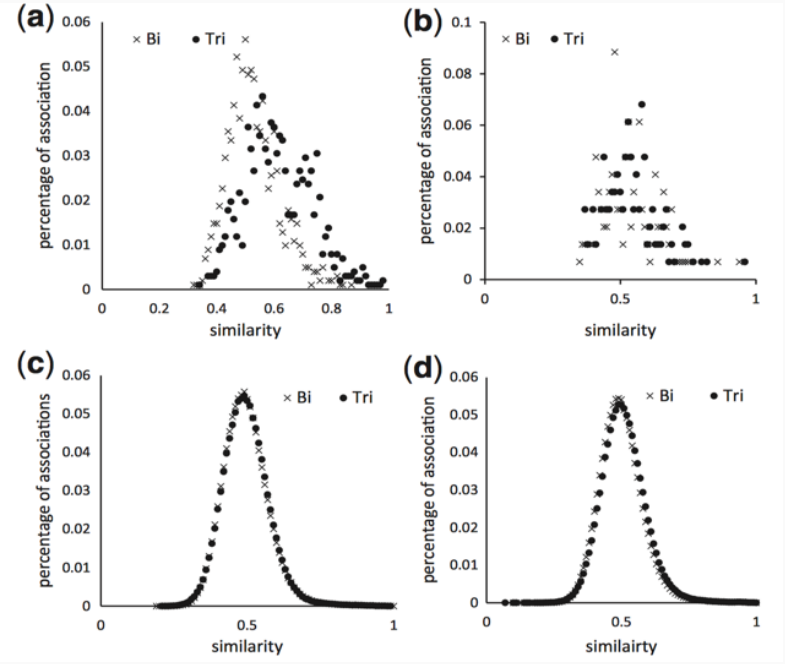
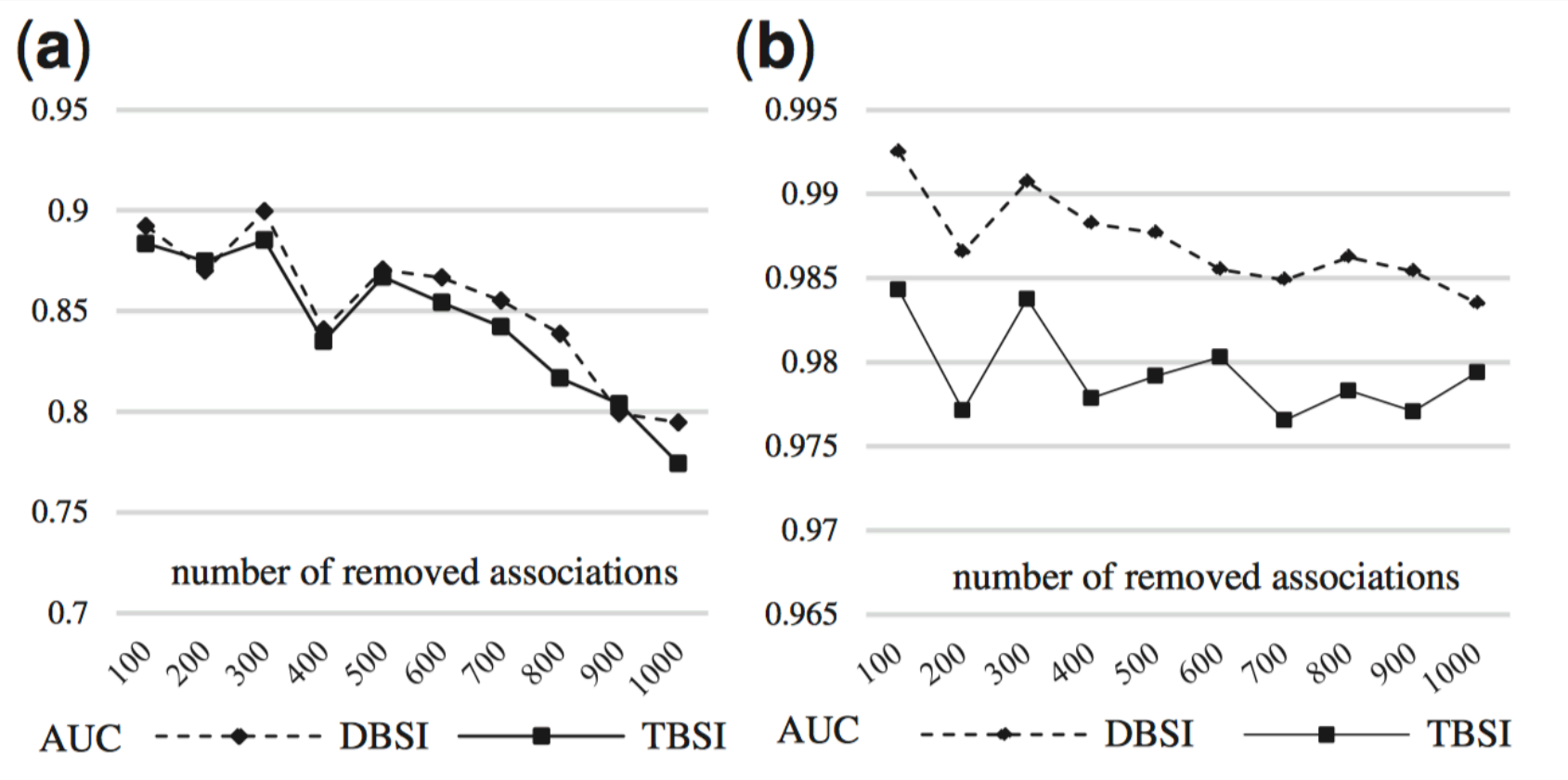


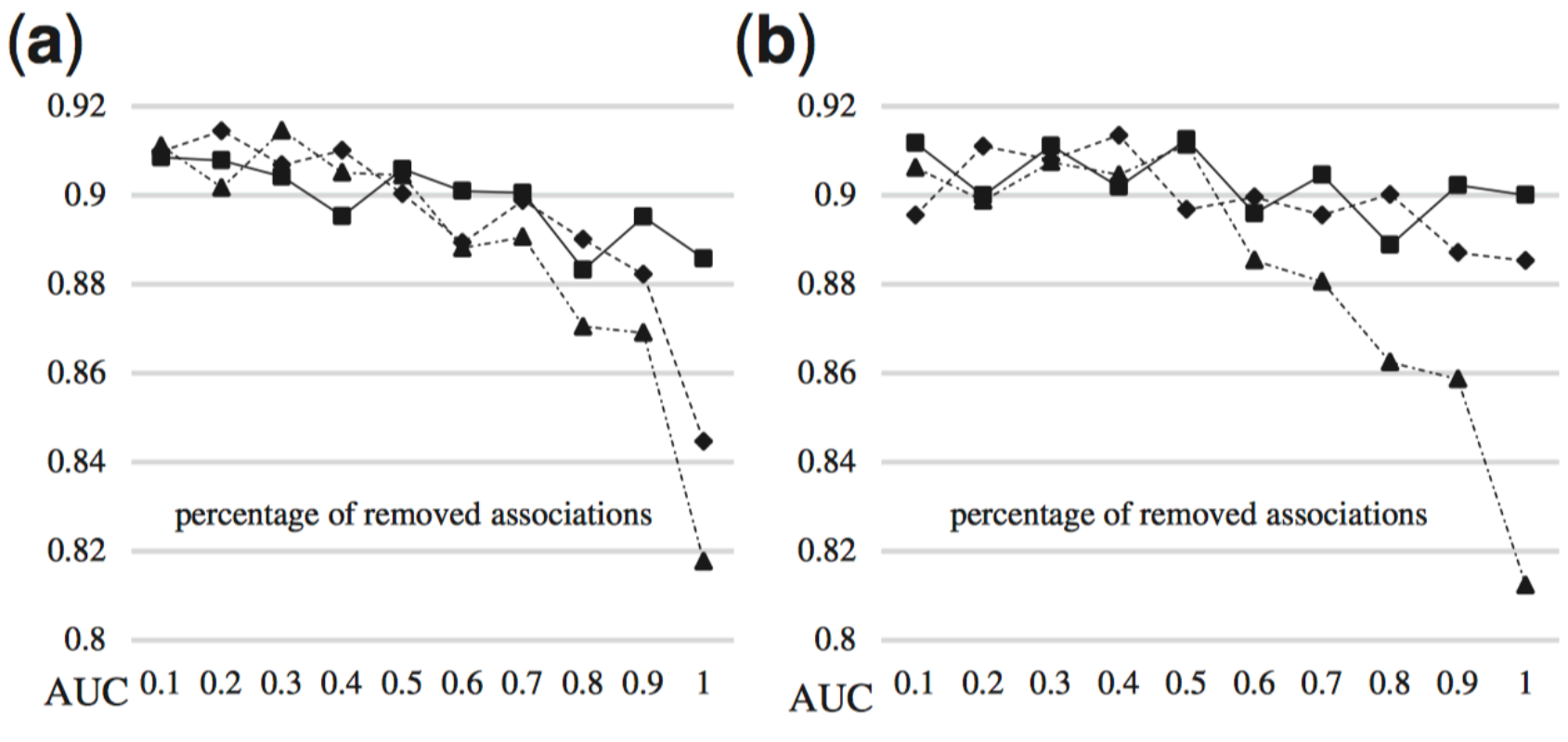
图 2: TBSI

结果分析



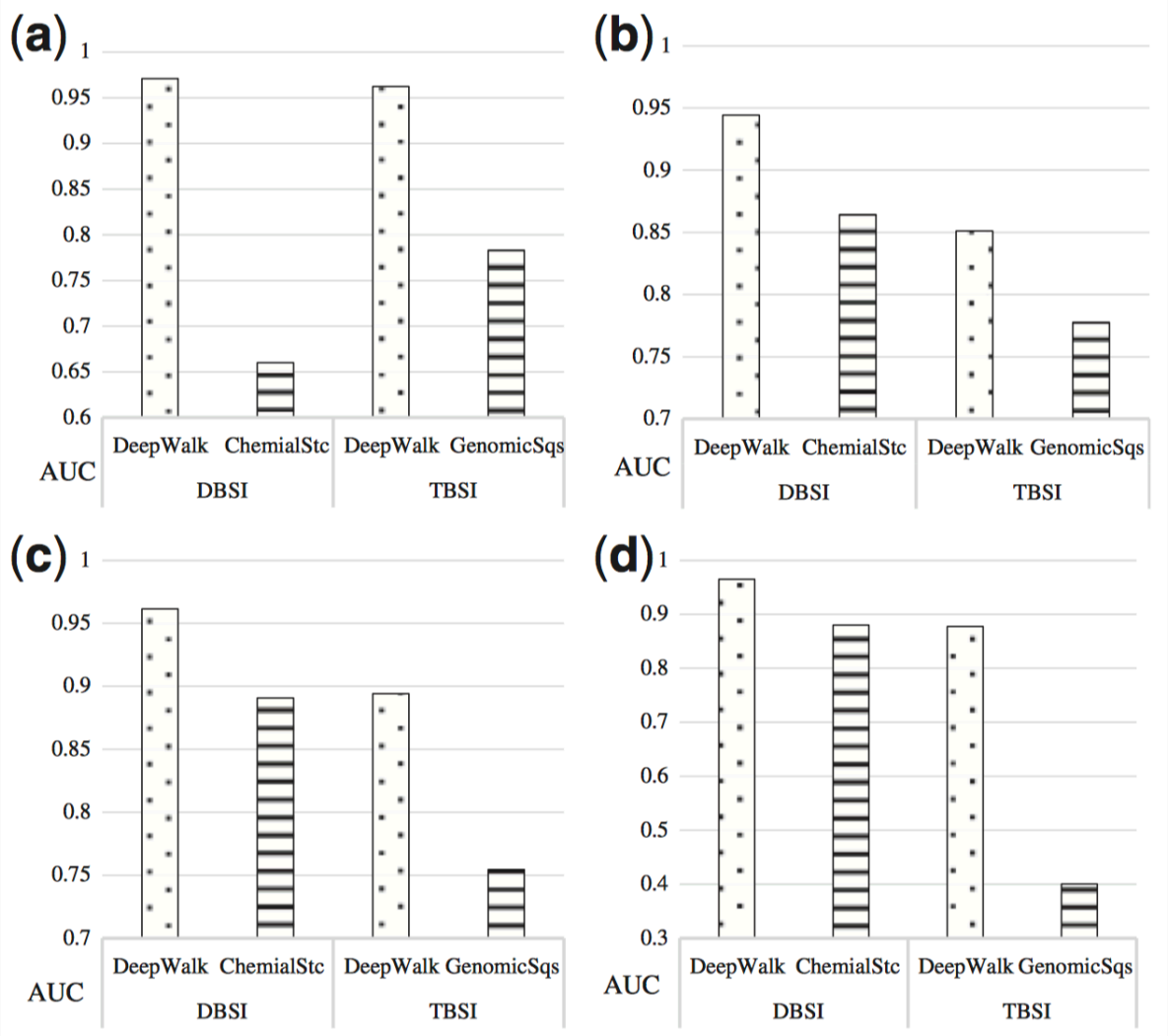
随机去除数据点的影响

结果分析



取出不同联系的影响

结果分析



DeepWalk 与传统方法比较

结果讨论

- 不足之处：
  - 无法产生新的发现
  - 很难再扩展考虑别的信息
- 展望：将更多的信息纳入异构网络