

# Affinity2Vec: Drug-Target Binding Affinity Prediction through Representation Learning, Graph Mining, and Machine Learning

## SUPPLEMENTARY MATERIAL

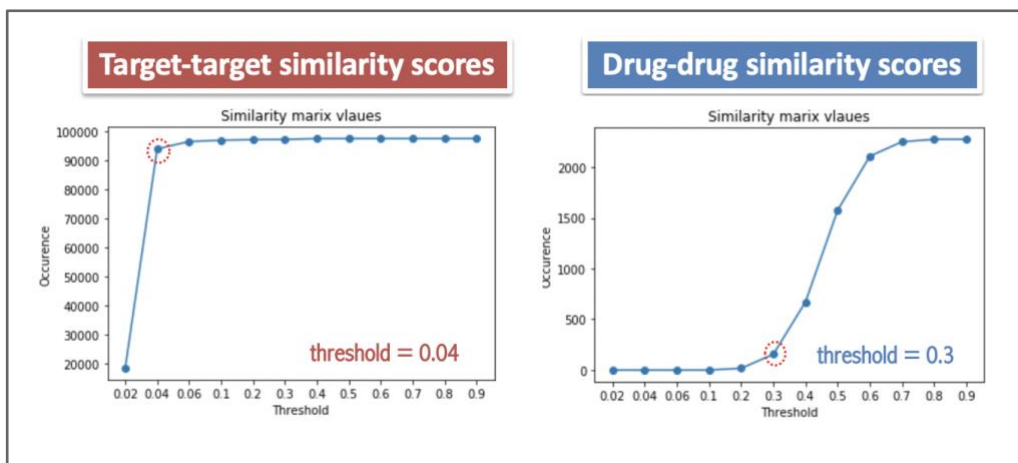
Maha A. Thafar<sup>1,2</sup>, Mona Alshahrani<sup>3</sup>, Somayah Albaradei<sup>4</sup>, Takashi Gojobori<sup>1</sup>, Magbubah Essack<sup>1\*</sup>, Xin Gao<sup>1\*</sup>

\*To whom correspondence should be addressed.

### Supplementary Figure with explanation

**Supplementary Figure S1:** An illustration of setting the thresholds for drug similarity and target similarity for the Davis dataset.

We did an empirical analysis for the drug-drug similarity graph and target-target similarity graph to specify different threshold values for each graph separately, as explained and illustrated in supplementary figure S1 in the supplementary material file. In particular, for the Davis dataset as an example, the threshold for drug similarity is set to 0.3 while the threshold of target similarity is set to 0.04. We plotted all occurrences of similarity below several thresholds ranging from 0.02 to 0.9 as shown in the figure below, and we assign the threshold value when there is a big changing in the curve.



### Supplementary Table

**Supplementary Table S1:** Best-obtained performance of several Affinity2Vec models using different features for Davis & KIBA Datasets in term of rm2 and AUPR - **Bold font with underline indicates the best results, and bold font alone shows the second-best results.**

*The abbreviation in the tables: Dr: Drugs, Pr: Proteins, aaseq: amino-acid sequence, D: Davis dataset, and K: KIBA dataset. G1 consists of (DTBA training part, Ddsim1 2D chemical structures similarity, TTsim1: targets' sequence*

alignment similarity using normalized Smith-waterman scores), G2 consists of (DTBA training part, DDisim2 drugs' SMILES embeddings cosine similarity, TTSim2: targets' sequence embeddings cosine similarity)

Model name	Features Type	Davis dataset		KIBA dataset	
		rm2	AUPR	rm2	AUPR
<b>Affinity2Vec Pscore</b>	Meta-path scores of G1	0.679	0.72	0.615	0.746
	Meta-path scores of G2	0.53	0.61	<b><u>0.783</u></b>	<b><u>0.926</u></b>
	Meta-path scores of G1 & G2	0.677	0.718	<b>0.767</b>	<b>0.909</b>
<b>Affinity2Vec Embed</b>	Dr SMILES embeddings FV + Pr aaseq embeddings FV	0.57	0.633	0.58	0.68
<b>Affinity2Vec Hybrid</b>	Meta-path scores of G1 + Dr SMILES embeddings FV + Pr aaseq embeddings FV	<b><u>0.693</u></b>	<b><u>0.734</u></b>	0.686	0.80
	Meta-path scores of G2 + Dr SMILES embeddings FV + Pr aaseq embeddings FV	0.58	0.64	<b>0.765</b>	<b>0.91</b>
	Meta-path scores of G1 & G2+ Dr SMILES embeddings FV + Pr aaseq embeddings FV	<b>0.69</b>	<b>0.731</b>	<b>0.766</b>	<b>0.902</b>