SubDyve: Subgraph-Driven Dynamic Propagation for Virtual Screening Enhancement Controlling False Positive

Jungseob Yi

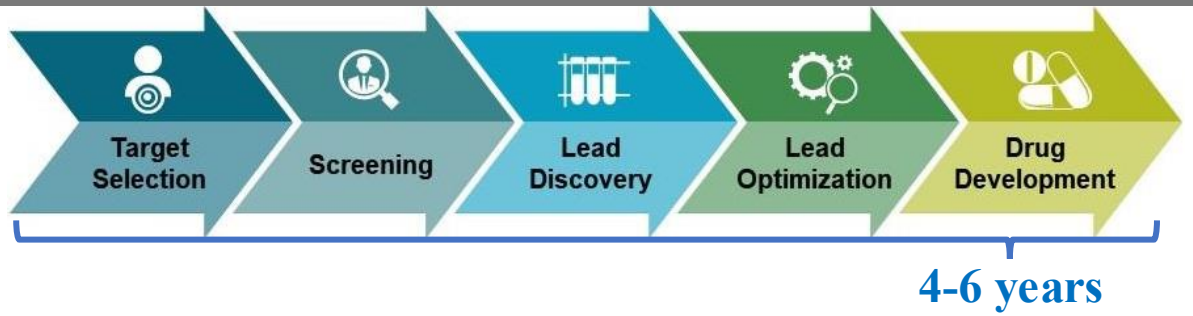
Seoul National University

Ph.D. Student, Interdisciplinary Program in Artificial Intelligence

Bio & Health Informatics Lab.

Contents

Introduction to Virtual Screening



1. Key Challenges in VS

- **Vast search space:** Over 10^{60} candidate compounds make exhaustive search computationally intractable, necessitating VS.
- **Low-label regime:** Screening must begin with only a handful of known active compounds.
- **Core challenge:** Identifying true hits from billions of candidates using minimal labeled data, while ranking them at the top.

2. Limitations of Existing Methods under Low-label Regime

- **Supervised Models:** Require large-scale labeled data; prone to severe overfitting and high false positive rates under limited supervision.
- **Foundation Models:** Score molecules independently, failing to capture inter-molecular dependencies during inference.

3. Structural Limitations of Related Works

- **Network Propagation (NP):**
  - **Limited expressivity:** Relies on pre-defined generic fingerprints (e.g., ECFP), failing to capture activity-determining subgraph patterns.
  - **False positive inflation:** Topological bias over-promotes high-degree nodes, regardless of true bioactivity.
- **Representation-centric models:** Both traditional fingerprints and deep learning models miss **key bioactivity patterns**, treating molecules in isolation.
- **Label-efficient NP:** A promising strategy for low-label settings, yet still fails to overcome generic fingerprint dependency and false positive control.

Research Objectives

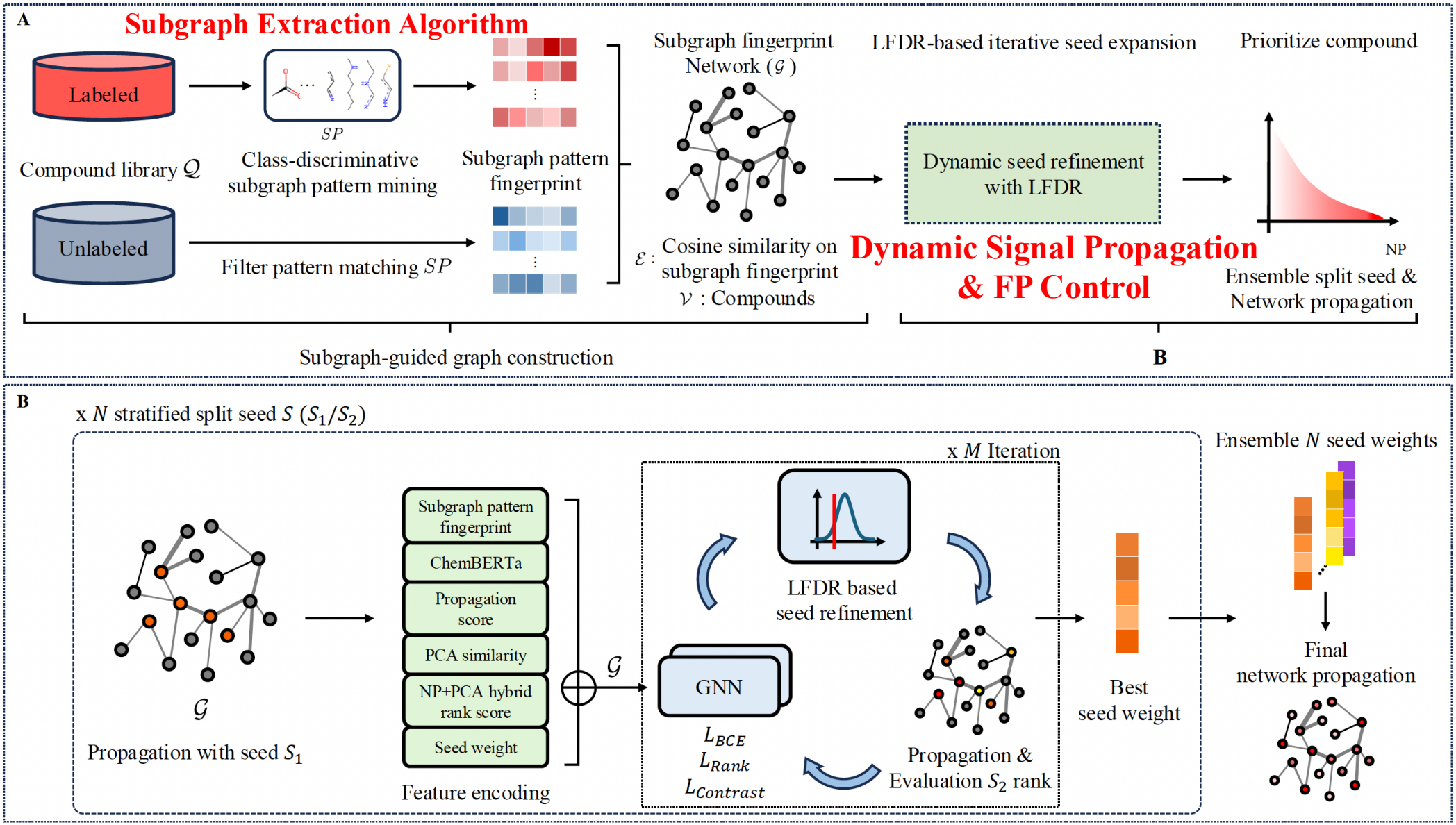
Problem Formulation

✓ Virtual screening as *the problem of ranking a large set of unlabeled candidate* compounds, especially *under a low-label regime*.

Our approach

1. Overcoming the limitations of pre-defined fingerprints (FP)
   → Extract assay-based, class-discriminative subgraph patterns to redefine molecular fingerprints beyond pre-defined descriptors
2. Incorporating inter-molecular relationships
   → Leverage chemical similarity networks and network propagation to incorporate inter-molecular relationships into predictions
3. Controlling graph-induced bias and reducing false positives
   → Enhance prediction reliability and interpretability via node-level ranking and LFDR-based false positive control

SubDyve: Subgraph-Driven Dynamic Propagation for Virtual Screening Enhancement Controlling False Positive



Performance Comparison

Virtual Screening Benchmark Dataset DUD-E

- **Active (ChEMBL)**
- **Decoy (*Similar structure to Active compounds*)**

BEDROC
Degree to which active compounds are enriched at the top of the ranking

$EF_{N\%}$
Fold enrichment of active compounds in the top N% of the ranked list relative to random selection

Ten DUD-E Targets Zero-shot Screening Performance

Protein Target	SubDyve (Ours)		PharmacoMatch(2025)		CDPKit(2024)		DrugCLIP(2023)		MoLFormer(2022)		AutoDock Vina(2021)	
	BEDROC	$EF_{1\%}$	BEDROC	$EF_{1\%}$	BEDROC	$EF_{1\%}$	BEDROC	$EF_{1\%}$	BEDROC	$EF_{1\%}$	BEDROC	$EF_{1\%}$
ACES	**86±2**	**57.0±2.4**	18±1	8.4±1.4	16±2	5.5±1.3	52±2	32.4±1.7	24±2	8.3±0.7	33±1	13.87±0.5
ADA	**83±4**	50.6±5.3	44±4	16.7±4.1	82±3	53.6±4.3	82±3	**60.2±5.3**	72±1	48.3±0.9	7±2	1.05±1.7
ANDR	**72±2**	**37.1±2.1**	33±2	15.8±1.9	26±2	12.6±2.1	64±3	34.3±2.4	9±1	3.0±0.1	34±1	18.41±0.6
EGFR	**86±2**	**60.0±2.3**	11±1	3.1±0.7	26±2	12.2±1.6	40±2	28.7±2.1	75±2	48.1±2.8	14±1	3.68±0.7
FA10	58±2	46.8±1.7	1±1	0.2±0.2	6±1	0.0±0.0	**86±1**	**51.2±1.8**	66±0	36.7±0.4	41±1	15.77±0.8
KIT	44±3	13.8±2.6	4±1	0.0±0.0	9±2	1.1±0.8	10±2	5.2±1.7	**66±1**	**36.8±0.9**	18±2	2.97±1.9
PLK1	**85±3**	**51.7±4.0**	9±2	1.5±1.3	39±3	5.7±2.3	66±4	45.0±4.0	69±0	35.2±4.0	13±1	1.83±0.3
SRC	**61±2**	**35.0±1.8**	27±1	6.0±1.0	28±1	11.1±1.2	16±1	8.1±1.3	48±1	21.5±1.5	13±1	4.00±0.5
THRB	61±2	36.6±2.0	22±1	5.9±1.0	35±2	11.8±1.5	**83±1**	**46.9±1.7**	6±1	1.2±0.1	25±1	4.31±1.0
UROK	37±3	25.6±2.4	4±1	0.6±0.7	55±3	24.5±2.8	**73±3**	**48.1±3.1**	36±2	10.0±1.5	28±1	7.90±0.7
Avg. rank	1.6	1.6	4.5	4.4	3.6	3.7	2.4	2.0	3.0	3.3	4.0	4.0

Performance Comparison

PU-Style Screening Performance

- Target : CDK7
- Dataset
  - Hundreds of CDK7-related compounds from PubChem
  - ~10 million compounds from ZINC

- **Method**
  1. Subgraph pattern mining from CDK7 active compounds
  2. Filtering ZINC and PubChem compounds using mined patterns
  3. Network construction based on structural similarity
- **Experimental Setup**

  10% of CDK7 active compounds are held out;
  remaining actives serve as seeds to evaluate **recovery of masked active compounds**

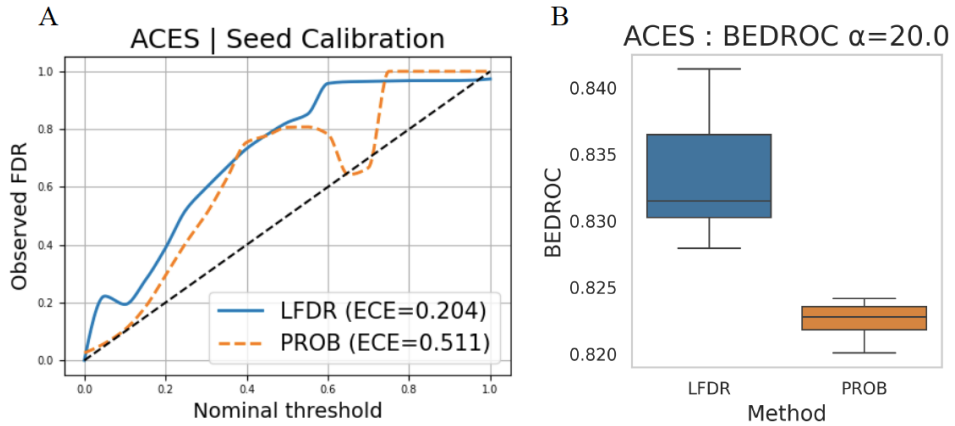
Method	BEDROC (%)	EF				
		0.5%	1%	3%	5%	10%
Deep learning-based						
BIND (Lam et al., 2024)	-	-	-	-	-	0.04 ± 0.08
AutoDock Vina (Eberhardt et al., 2021)	1.0 ± 1.3	-	0.2 ± 0.3	0.6 ± 0.7	1.1 ± 0.6	1.2 ± 0.5
DrugCLIP (Gao et al., 2023)	2.7 ± 1.26	1.63 ± 1.99	1.63 ± 0.81	2.45 ± 1.02	2.53 ± 1.35	2.69 ± 0.62
PSICHIC (Koh et al., 2024)	9.37 ± 3.08	4.07 ± 2.58	6.92 ± 3.30	7.48 ± 2.47	7.02 ± 1.80	5.35 ± 0.94
GRAB (Yoo et al., 2021)	40.68 ± 10.60	44.22 ± 8.35	45.21 ± 5.63	29.78 ± 1.38	18.69 ± 0.47	**10.00 ± 0.00**
Data mining-based						
avalon + NP (Yi et al., 2023)	77.59 ± 1.72	135.76 ± 6.44	87.58 ± 2.9	31.55 ± 0.54	19.67 ± 0.4	9.88 ± 0.16
estate + NP (Yi et al., 2023)	52.44 ± 6.19	94.4 ± 13.68	57.87 ± 7.15	22.71 ± 2.7	15.92 ± 0.85	8.24 ± 0.38
fp4 + NP (Yi et al., 2023)	69.62 ± 3.69	122.76 ± 13.02	75.01 ± 4.21	28.96 ± 1.34	18.36 ± 1.0	9.59 ± 0.29
graph + NP (Yi et al., 2023)	75.86 ± 3.99	126.72 ± 10.05	84.73 ± 3.74	31.68 ± 0.92	19.1 ± 0.47	9.75 ± 0.24
maccs + NP (Yi et al., 2023)	75.44 ± 4.85	135.72 ± 12.7	79.82 ± 4.76	31.0 ± 1.41	18.93 ± 0.66	9.67 ± 0.21
pubchem + NP (Yi et al., 2023)	63.48 ± 5.16	99.17 ± 10.17	69.3 ± 7.08	30.87 ± 1.27	18.77 ± 0.9	9.84 ± 0.15
rdkit + NP (Yi et al., 2023)	79.04 ± 1.96	148.69 ± 4.25	89.24 ± 2.08	31.68 ± 0.92	19.02 ± 0.55	9.55 ± 0.3
standard + NP (Yi et al., 2023)	72.42 ± 3.51	121.97 ± 15.51	84.34 ± 5.56	31.27 ± 0.96	19.01 ± 0.33	9.71 ± 0.24
SubDyve (Ours)	**83.44 ± 1.44**	**155.31 ± 6.38**	**97.59 ± 1.44**	**33.01 ± 0.60**	**19.90 ± 0.18**	**10.00 ± 0.00**
Statistical Significance (p-value)	**	-	**	*	-	-

Ablation Study

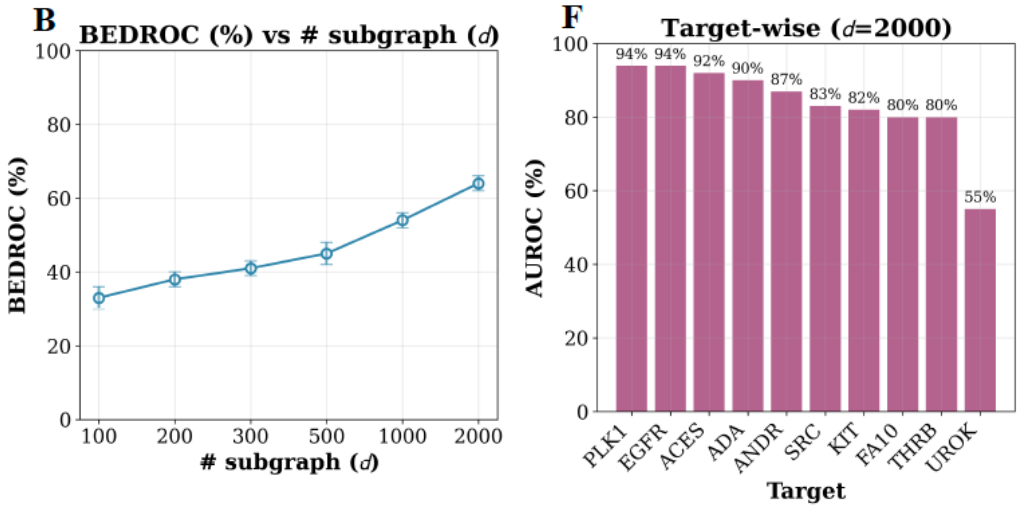
Effects of Subgraph Network & LFDR Seed Refinement

Subgraph	LFDR	BEDROC	$EF_{1\%}$
		79.04 ± 1.96	89.24 ± 2.08
	✓	63.78 ± 11.43	67.22 ± 16.61
✓		78.68 ± 2.87	89.68 ± 3.53
✓	✓	**83.44 ± 1.44****	**97.59 ± 1.44****

FP-pressure via LFDR threshold



Subgraph Pattern Size

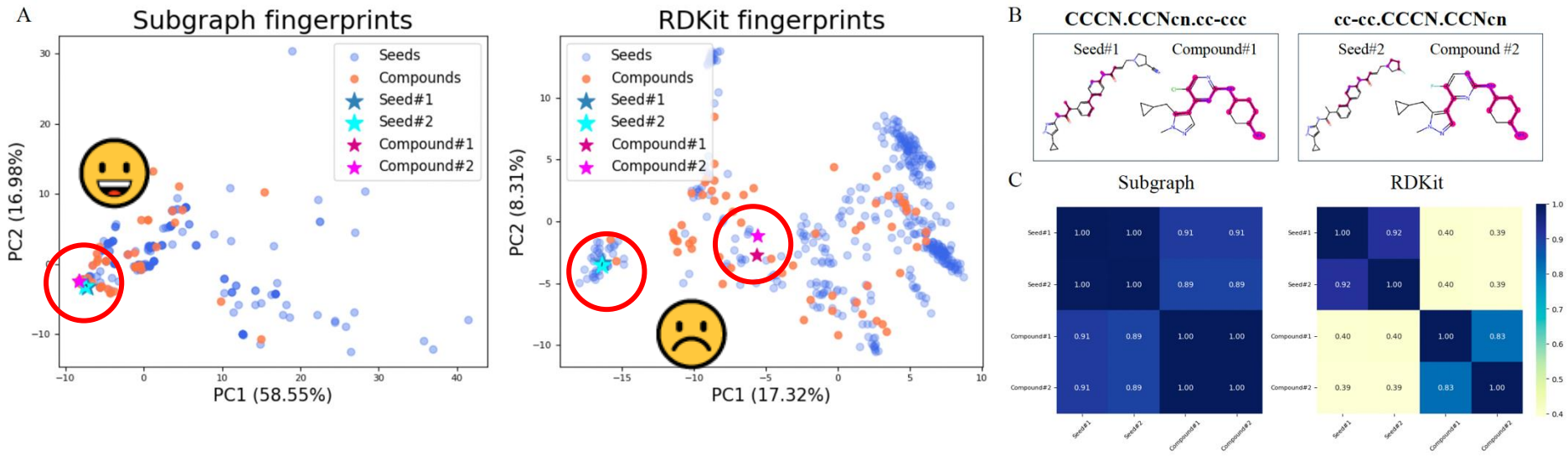


Performance under Varying Seed Set Sizes

No. of Seeds	Method	BEDROC (%)	EF 0.30%	EF 0.50%	EF 1%	EF 3%	EF 5%
50	pubchem + NP	41.13 ± 4.46	44.69 ± 14.09	45.51 ± 7.91	41.97 ± 6.91	25.7 ± 1.99	17.14 ± 1.00
	maccs + NP	47.02 ± 3.83	56.77 ± 15.24	52.81 ± 9.24	50.92 ± 3.15	27.74 ± 2.04	17.05 ± 1.20
	Subgraph + NP	46.33 ± 1.26	37.79 ± 21.22	31.81 ± 12.68	**53.93 ± 4.97**	27.61 ± 1.47	17.27 ± 0.51
	SubDyve	**51.78 ± 3.38**	**69.5 ± 11.81**	**62.53 ± 14.84**	52.66 ± 5.91	**29.48 ± 2.37**	**18.15 ± 0.90**
150	rdkit + NP	50.82 ± 3.79	52.69 ± 6.75	54.62 ± 10.48	54.62 ± 7.24	29.5 ± 1.59	17.79 ± 0.95
	maccs + NP	55.22 ± 4.39	**79.99 ± 15.80**	71.65 ± 13.30	60.69 ± 6.59	30.6 ± 1.29	18.85 ± 0.48
	Subgraph + NP	55.08 ± 1.52	44.39 ± 22.83	61.29 ± 10.07	**67.17 ± 7.24**	30.07 ± 1.38	18.22 ± 0.93
	SubDyve	**59.07 ± 2.25**	74.67 ± 7.46	**73.55 ± 10.51**	66.72 ± 5.29	**32.26 ± 1.04**	**19.73 ± 0.36**
250	fp2 + NP	56.88 ± 5.26	67.45 ± 16.53	75.57 ± 15.28	65.15 ± 8.30	30.19 ± 1.26	18.52 ± 0.61
	avalon + NP	61.29 ± 2.44	97.18 ± 13.25	**86.96 ± 9.16**	68.05 ± 4.42	31.14 ± 0.52	19.51 ± 0.48
	Subgraph + NP	61.96 ± 3.24	41.01 ± 13.89	86.31 ± 11.97	**80.31 ± 4.60**	30.20 ± 1.44	18.49 ± 0.85
	SubDyve	**66.73 ± 2.71**	**97.69 ± 16.55**	85.44 ± 12.82	78.19 ± 3.38	**32.85 ± 0.60**	**19.72 ± 0.36**

Case Study

Substructure Similarity in CDK7-Target Compound Retrieval



A. SubDyve yields highly cohesive structural clusters, whereas RDKit-based retrieval produces scattered embeddings
B. SubDyve identifies functionally relevant substructures overlooked by pre-defined RDKit fingerprints
C. SubDyve maintains higher subgraph-level similarity than RDKit across retrieved compounds

➔ SubDyve demonstrates greater sensitivity to activity-relevant substructures, facilitating the discovery of functionally similar compounds.

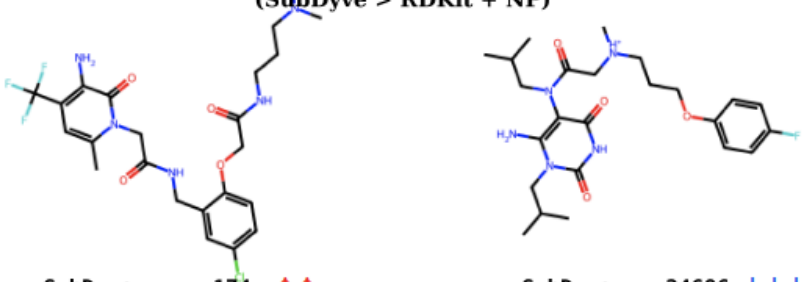
Case Study

Ranking gap analysis for structurally similar pairs

THRB

Pair 1 - Active: CHEMBL44226 **Pair 1 - Decoy: D_C15451008**

Similarity: 0.869 | Rank Diff: +21475 | pval: 0.002
(SubDyve > RDKit + NP)



SubDyve:	174	↑ ↑	SubDyve:	24686	↓ ↓ ↓
RDKit + NP:	291	↑	RDKit + NP:	3328	↓

PLK1

Pair 1 - Active: CHEMBL562735 **Pair 1 - Decoy: D_C40404712**

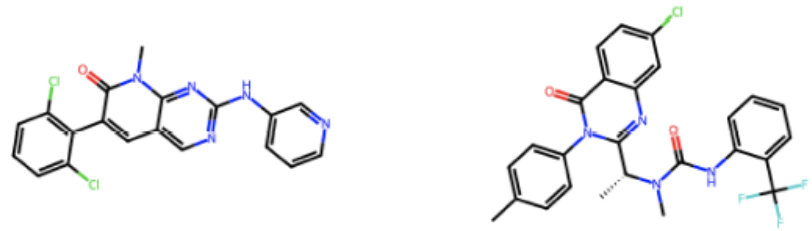
Similarity: 0.846 | Rank Diff: +1132 | pval: 0.002
(SubDyve > RDKit + NP)

SubDyve:	52	↑ ↑	SubDyve:	4598	↓ ↓
RDKit + NP:	82	↑ ↑	RDKit + NP:	3496	↓

SRC

Pair 1 - Active: CHEMBL325589 **Pair 1 - Decoy: D_C40360668**

Similarity: 0.868 | Rank Diff: +370 | pval: 2.66e-11
(SubDyve > RDKit + NP)

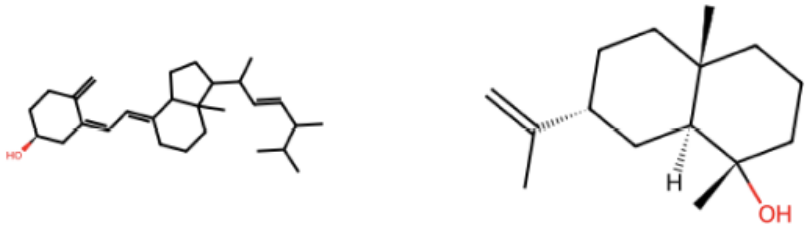


SubDyve:	159	↑ ↑	SubDyve:	16265	↓
RDKit + NP:	447	↑	RDKit + NP:	16183	↓

ANDR

Pair 1 - Active: CHEMBL50241 **Pair 1 - Decoy: D_C43764063**

Similarity: 0.914 | Rank Diff: +572 | pval: 1.43e-04
(SubDyve > RDKit + NP)



SubDyve:	13	↑ ↑ ↑	SubDyve:	2797	↓
RDKit + NP:	309	↑	RDKit + NP:	2521	↓

Time and Memory Cost

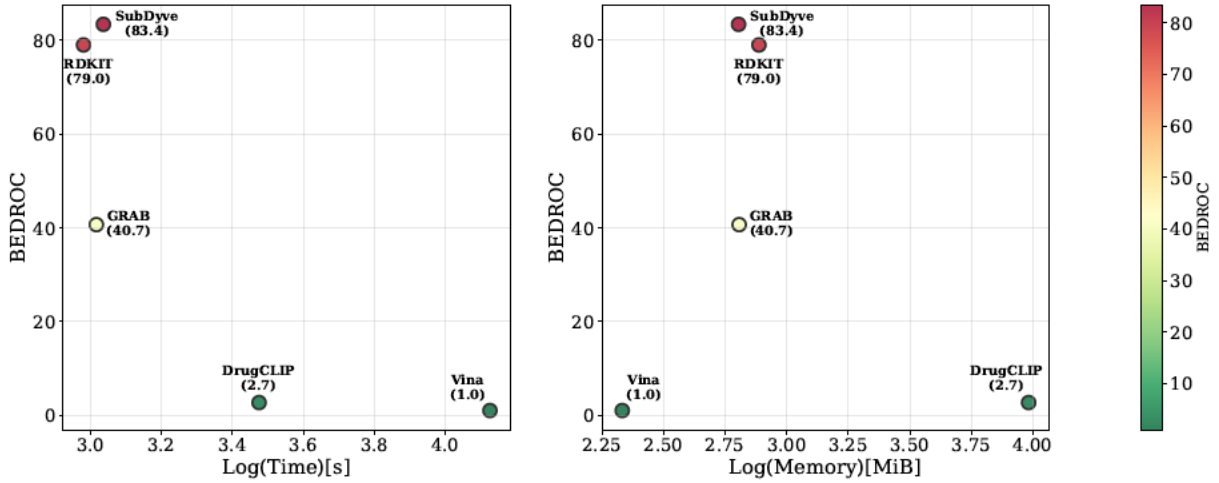
Target	SubDyve Network Building			RDKit
	Subgraph Mining	Network Construction	Total	Preprocessing
Average	10.68	7.20	**17.88**	**9.27**

Subgraph mining incurs
a one-time preprocessing overhead

Model	BEDROC	Preprocessing (sec)	Inference (sec)	Memory Used (MiB)
AutoDock Vina	1.0 ± 1.3	-	13343.37	215.7
DrugCLIP	2.7 ± 1.26	2955.93	29.53	9605.1
GRAB	40.68 ± 10.60	1038.97	0.03	661.1
rdkit + NP	79.04 ± 1.96	956.70	-	771.7
SubDyve	83.44 ± 1.44**	1088.0	-	637.9

Time
÷2.74 vs. DrugCLIP

Memory
÷ 15.06 vs. DrugCLIP



BEDROC
+4.4% vs. RDKit+NP
+42.76% vs. GRAB

Thank you