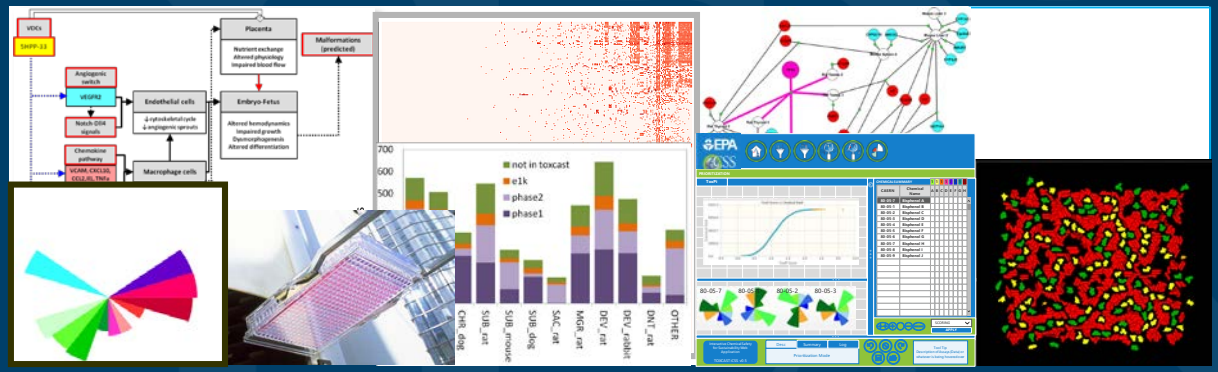
Case study of read-across predictions using a Generalized Read-Across (GenRA) Approach



George Helman[1,2], Grace Patlewicz[2], Imran Shah[2]

[1]Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, Tennessee, USA
[2]National Center for Computational Toxicology, US EPA, RTP, NC, USA

Outline

- Review GenRA approach
- Incorporate new physicochemical similarity context using two different approaches: one which is similar to current methods and one which is novel
- Evaluate performance of the two approaches and compare
- Case study of a read-across using GenRA with both approaches for addressing physchem.

Current Category Workflow in GenRA

Decision context

screening level assessment of hazard based on toxicity effects from ToxRef

Similarity contexts	GenRA
Structure similarity	√
Physicochemical	Subject of this study
Bioactivity e.g. ToxCast	√
Reactivity	-
Metabolic	-
Toxicokinetic	-

and p value metrics

Analogue evaluation

valuate consistency and concordance f experimental data of the source nalogues across the endpoint or etween endpoints using the data atrix

Read-across

Similarity weighted average – many to one read-across

Physchem Similarity Context

- Important context of similarity in read-across
- Models “bioavailability”
- Properties selected: Lipinski Rule of 5 (LogP, MW, # HB donors/acceptors)
- Two approaches investigated as a means to identify source analogs and evaluate their predictive performance:

Approach 1: “Filter”

Subcategorize from a set of analogs identified based on structural similarity

Common approach

Approach 2: “Search Expansion”

“Frontload” both structure and physchem into analog identification

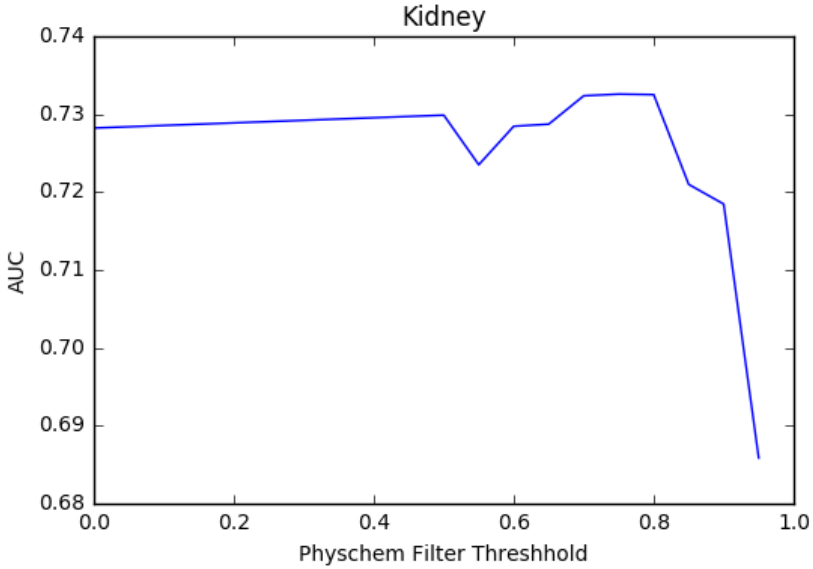
Novel approach

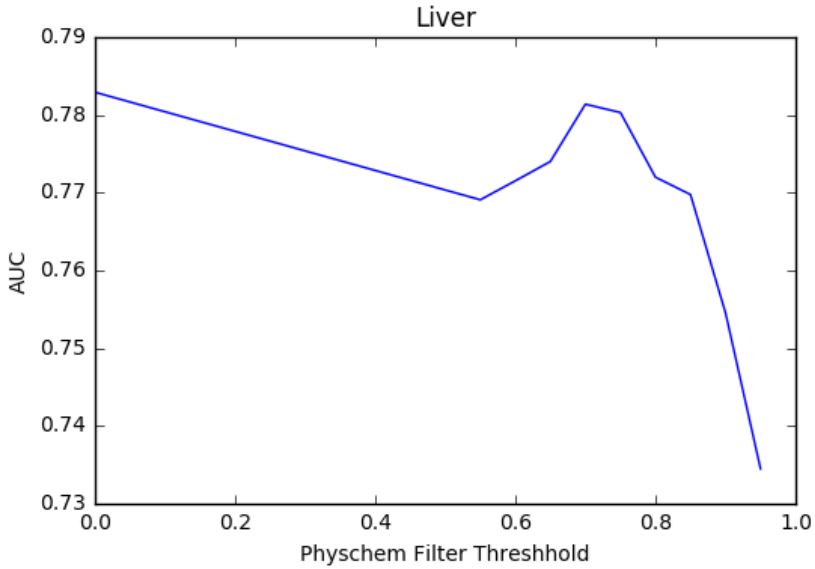
Approach 1: Filter

- Similarity search using Jaccard distance of Morgan chemical fingerprints to find source analogs. (Default of 10 nearest neighbors (k=10))
- Calculate physchem similarity between target and source analogs using a generalized Jaccard similarity metric
- Reduce neighborhood based on the physchem similarity threshold

Approach 1: Results

This approach does not perform as well as the GenRA baseline for the entire dataset, nor does it significantly improve any target organ predictions.



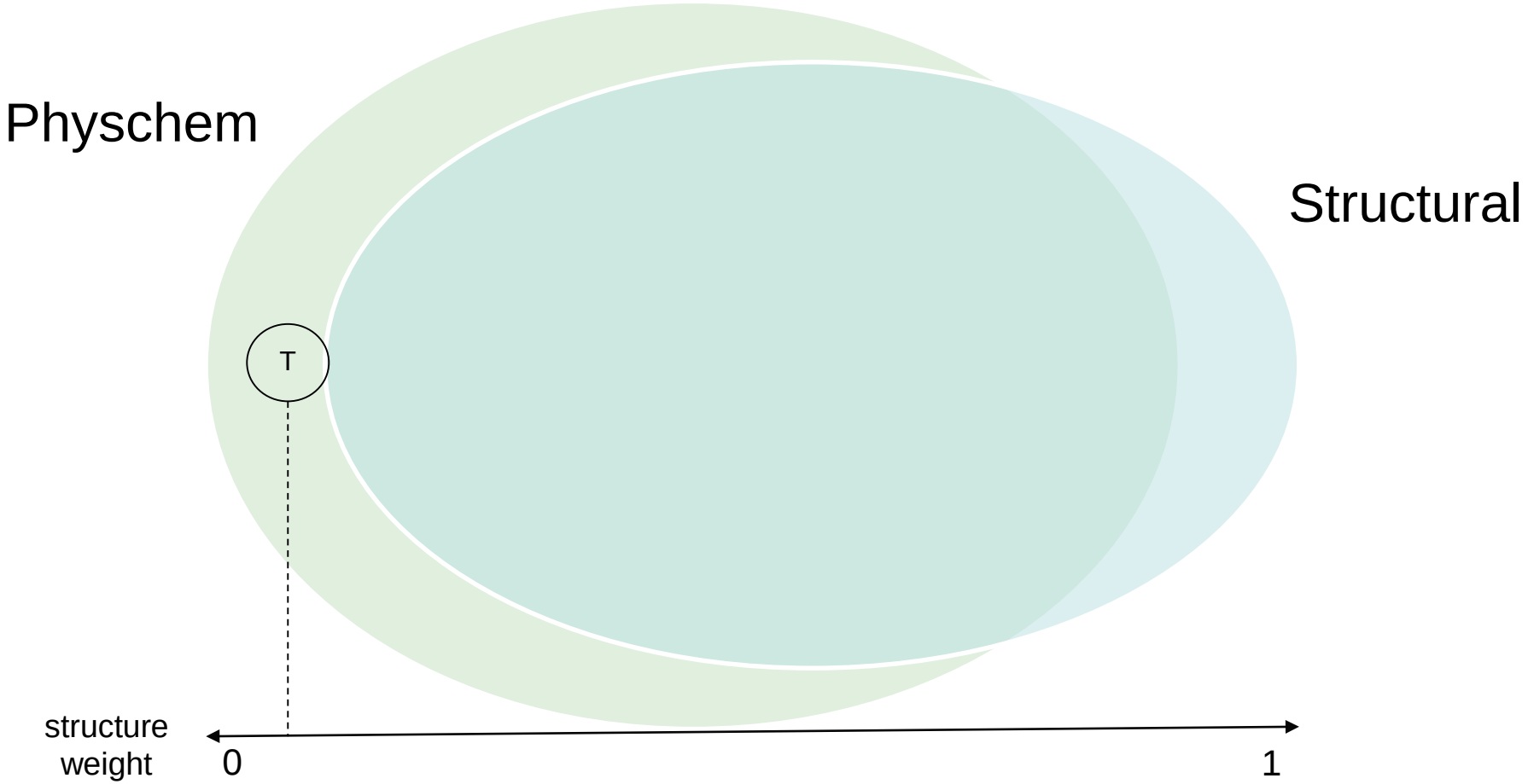


Target organ predictions aggregated over study type to organ level

Approach 2: Search Expansion

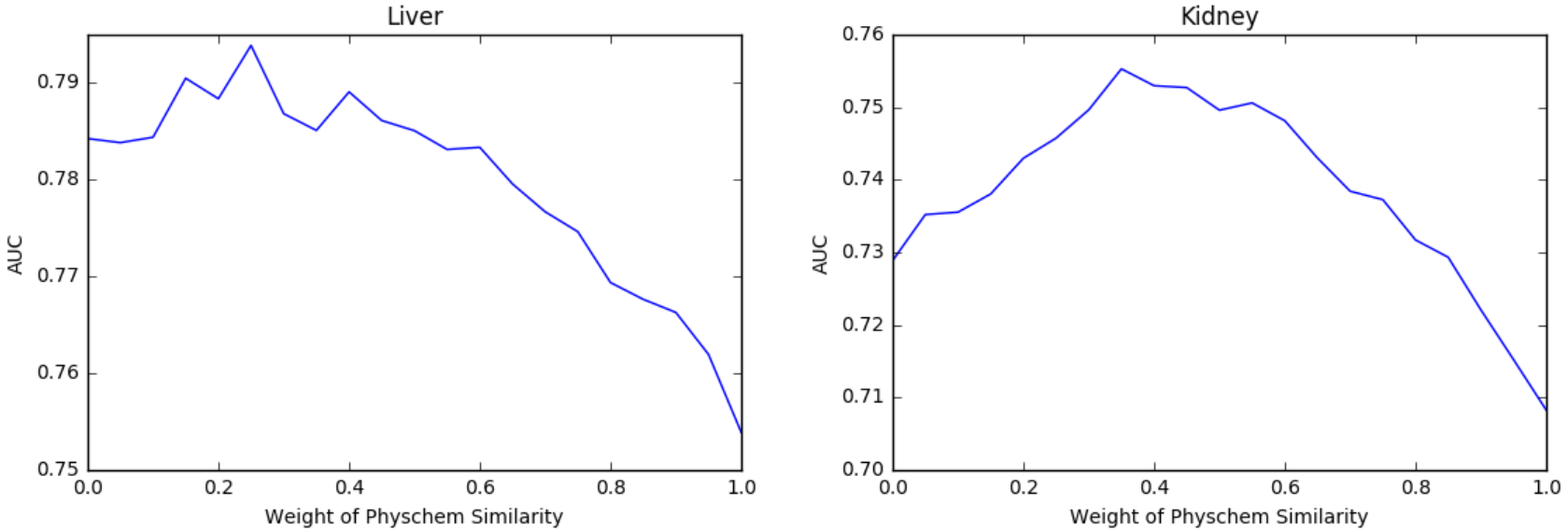
- Similarity search using weighted sum of Morgan chemical fingerprint distance and physchem distance to find source analogs (default of 10 nearest neighbors (k=10))
  - $Sim(A,B)=Jac(FP_A,FP_B)W_1 + Jac(PC_A, PC_B)W_2$
- Search through weight values to find the best performing combination i.e. what is the “sweet spot” of weights to achieve the best read-across prediction

Neighborhoods



Approach 2: Results

This approach shows a small improvement over baseline for entire dataset, but large improvement in certain organs.

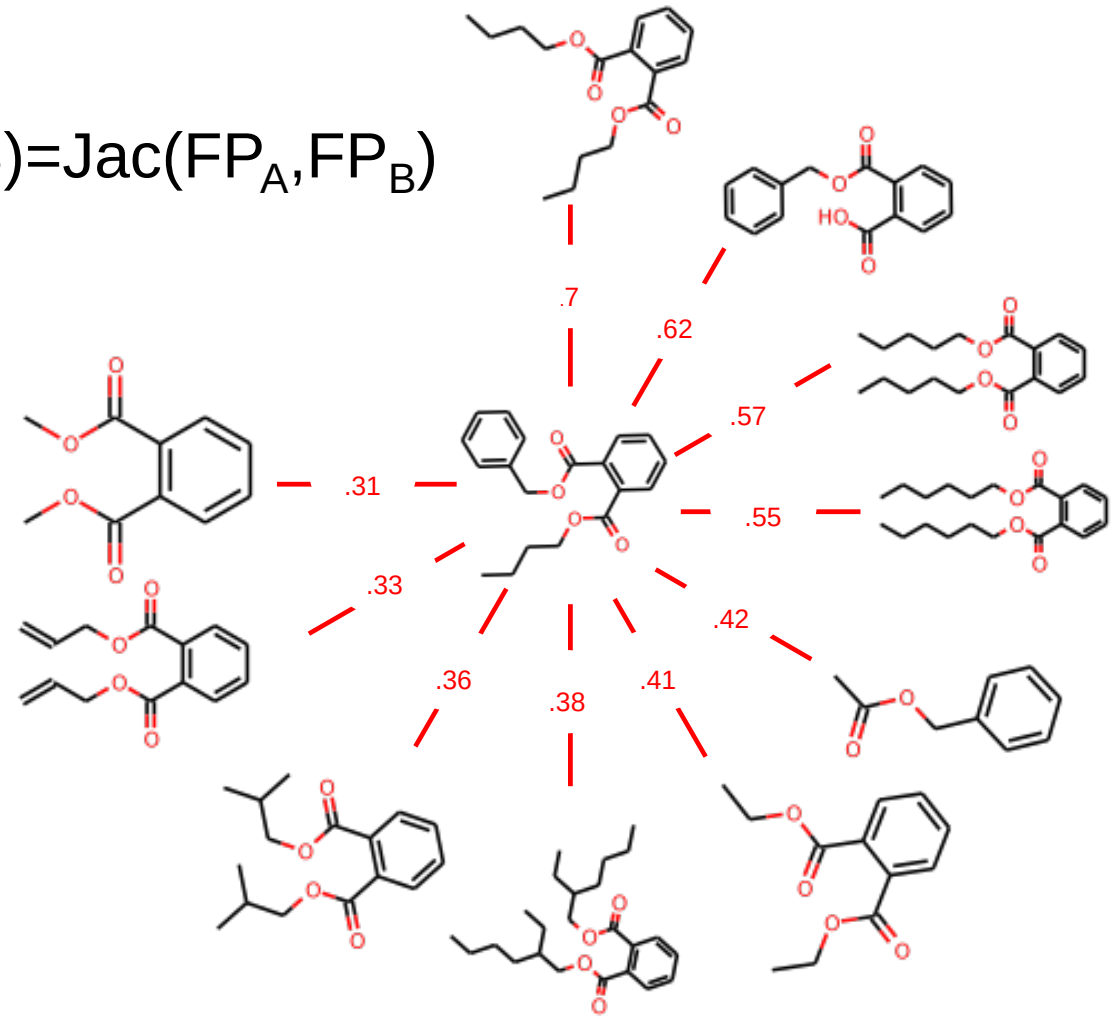


Target organ predictions aggregated over study type to organ level

Target organ predictions that were significantly improved: Intestine Large, Intestine Small, Mammary Gland, Pancreas, Ureter, Urinary Bladder

Case Study: Butyl Benzyl Phthalate

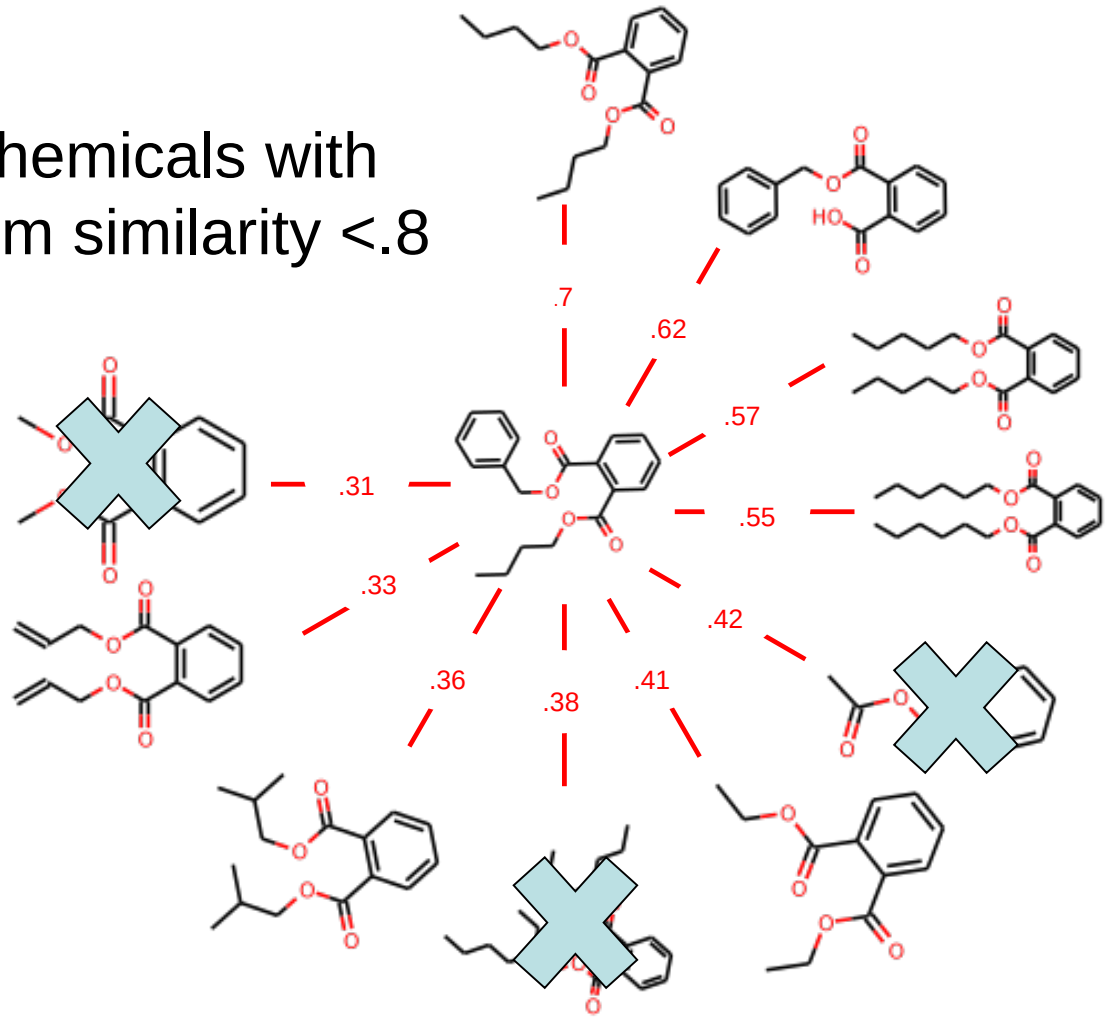
$Sim(A,B)=Jac(FP_A,FP_B)$



Case Study: Butyl Benzyl Phthalate

Approach 1: Filter

Reject chemicals with physchem similarity <.8



Case Study: Butyl Benzyl Phthalate

Approach 1: Filter Results

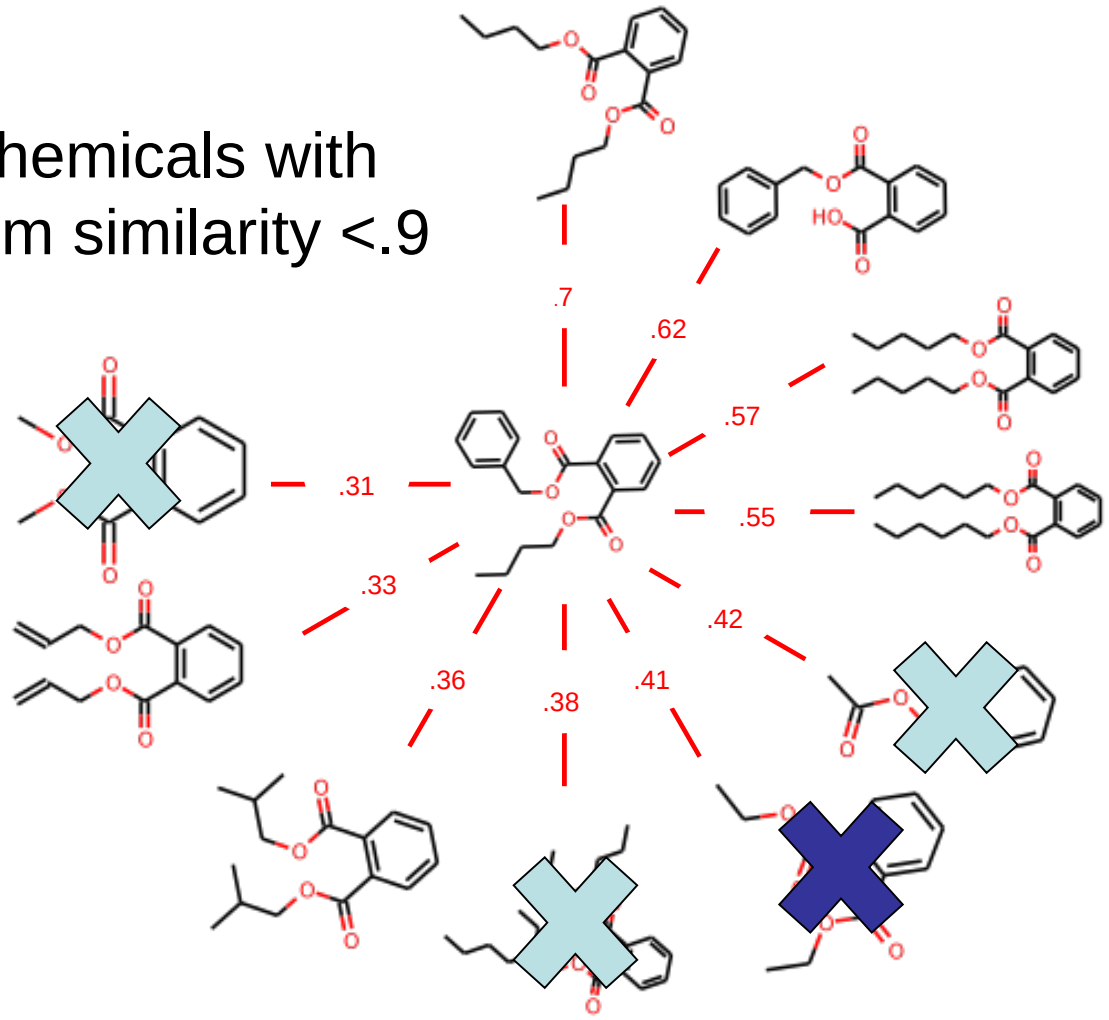
Endpoint	Original Neighborhood Prediction	Filtered Neighborhood Prediction
Body Weight	.78	.55
Clinical Chemistry	.27	.55
Food Consumption	0	0
Hematology	0	0
Kidney	.27	.55
Liver	1	1
Mortality	.27	.55
Pancreas	.27	0
Prostate	0	0
Skin	.27	.55
Spleen	0	0
Tissue NOS	0	0
Urinary Bladder	0	0

- Chronic studies
- All positive effects
- Higher prediction indicates more and stronger positive neighbors
- Filtering flips incorrect predictions for 4 endpoints.

Case Study: Butyl Benzyl Phthalate

Approach 1: Strict Filter

Reject chemicals with physchem similarity <.9



Case Study: Butyl Benzyl Phthalate

Approach 1: Strict Filter Results

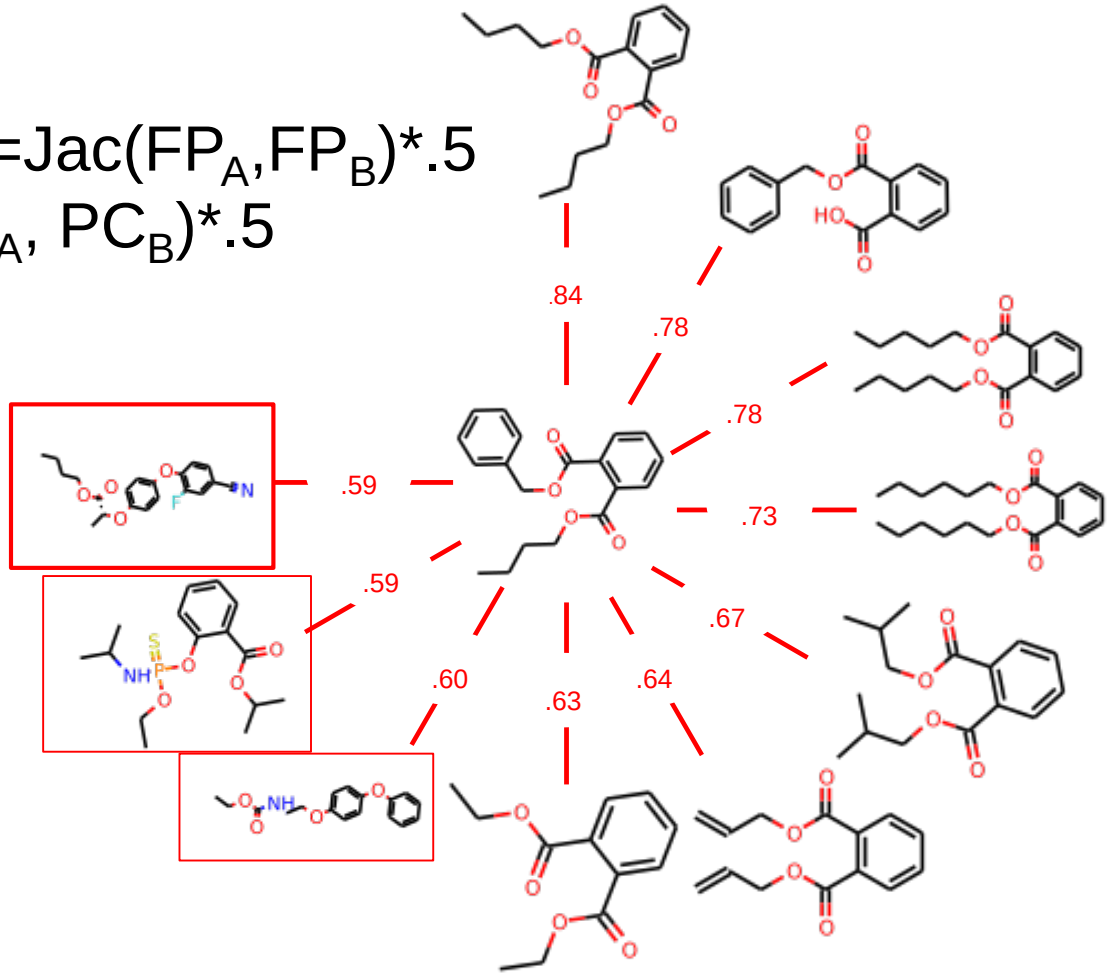
Endpoint	Original Neighborhood Prediction	Filtered Neighborhood Prediction
Body Weight	.78	0
Clinical Chemistry	.27	0
Food Consumption	0	0
Hematology	0	0
Kidney	.27	0
Liver	1	1
Mortality	.27	0
Pancreas	.27	0
Prostate	0	0
Skin	.27	0
Spleen	0	0
Tissue NOS	0	0
Urinary Bladder	0	0

- Performs worse than unfiltered neighborhood
- Filtering too strictly can remove very important neighbors

Case Study: Butyl Benzyl Phthalate Search

$$Sim(A,B)=Jac(FP_A,FP_B)*.5 + Jac(PC_A, PC_B)*.5$$

New Analogs identified to add to the overall neighborhood



Case Study: Butyl Benzyl Phthalate

Search Results

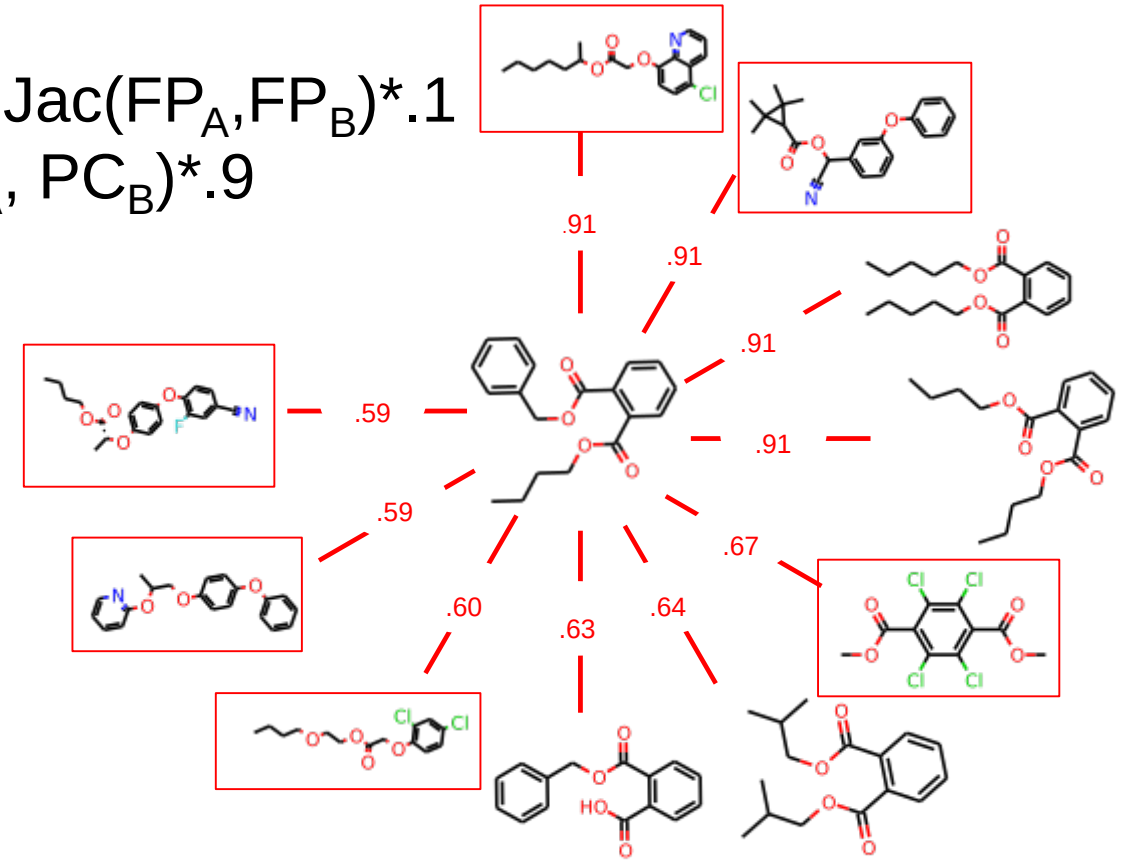
Endpoint	Structure Prediction	Structure + Pchem Prediction
Body Weight	.78	.79
Clinical Chemistry	.27	.60
Food Consumption	0	.20
Hematology	0	.20
Kidney	.27	.60
Liver	1	.80
Mortality	.27	.40
Pancreas	.27	0
Prostate	0	0
Skin	.27	.21
Spleen	0	.20
Tissue NOS	0	0
Urinary Bladder	0	0

- Adding phys-chem to similarity search flips incorrect predictions for 2 endpoints
- Improves many others
- Only 2 endpoints decrease

Case Study: Butyl Benzyl Phthalate

High-weight Search

$$Sim(A,B)=Jac(FP_A,FP_B)*.1 + Jac(PC_A, PC_B)*.9$$



Case Study: Butyl Benzyl Phthalate

High-weight Search Results

Endpoint	Structure Prediction	Structure + Pchem Prediction
Body Weight	.78	1
Clinical Chemistry	.27	.80
Food Consumption	0	.60
Hematology	0	.80
Kidney	.27	.80
Liver	1	1
Mortality	.27	.80
Pancreas	.27	0
Prostate	0	.2
Skin	.27	.2
Spleen	0	.40
Tissue NOS	0	0
Urinary Bladder	0	0

- Greater weight on physchem flips incorrect predictions for 6 endpoints
- Improves prediction for Spleen
- Only 2 endpoints decrease

Conclusion

- Systematic investigation performed to evaluate whether there is an improvement to GenRA (baseline) read-across predictions when accounting for physchem similarity
- Used 2 approaches to determine whether filtering on the basis of physchem from an existing set of structural analogues or whether accounting for physchem during a search resulted in better read-across performance
- Filtering neighborhoods based on physchem is sensitive to the strictness of the filter and can change predictions drastically. It often does not perform as well as the unfiltered neighborhood.
- Adding physchem to a similarity search can result in analogues that would otherwise not be considered, which often improves read-across performance over and above a purely structural search.
- Future work:
  - investigate other contexts of similarity such as metabolism, reactivity
  - Implement insights into the GenRA tool