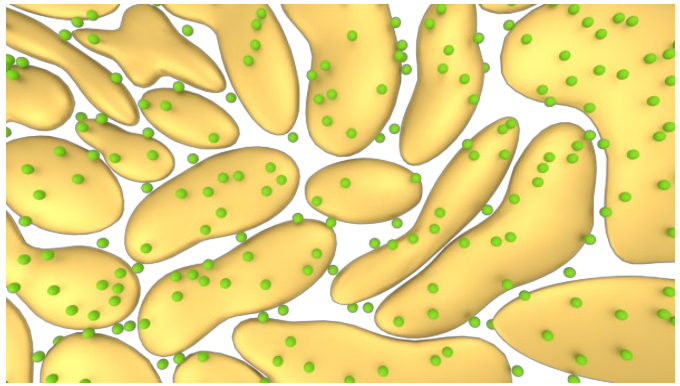
Materials Fingerprinting and Cartography

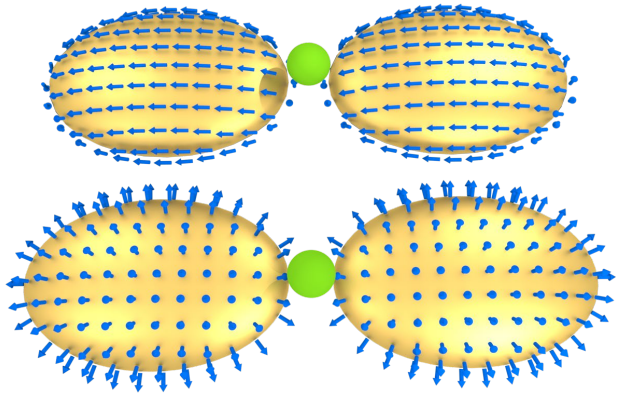
THOMAS A. R. PURCELL

NOVEMBER 5, 2018

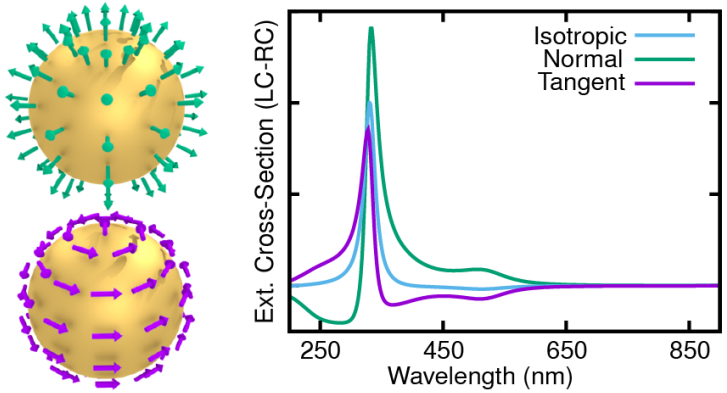
Summary of Ph. D.



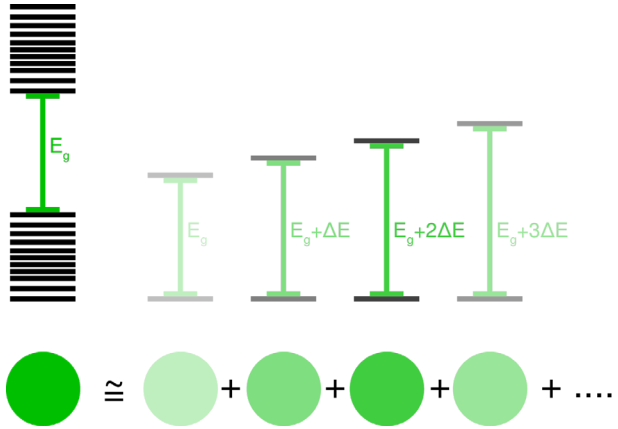
J. Phys. Chem. C. **2016**. *120*. 21837.



J. Phys. Chem. C. **2018**. *122*. 16901.



ACS Photonics. **2018**. *(Accepted)*



J. Chem. Phys. **2018**. *(Submitted)*

Outline

The purpose of materials fingerprinting

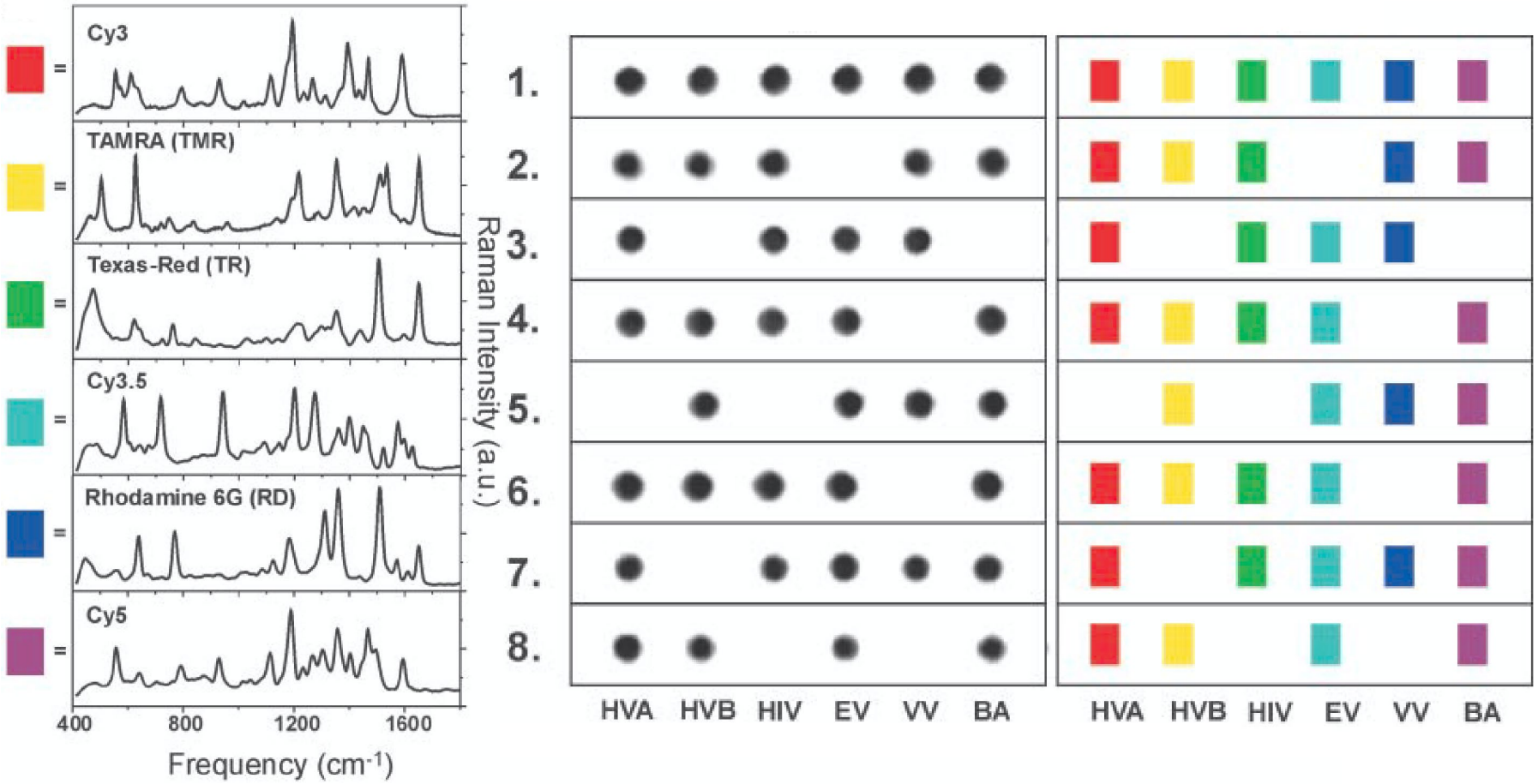
How to implement

Generating material catograms

Developing quantitative materials structure–property relationships with fingerprinting and SiRMS

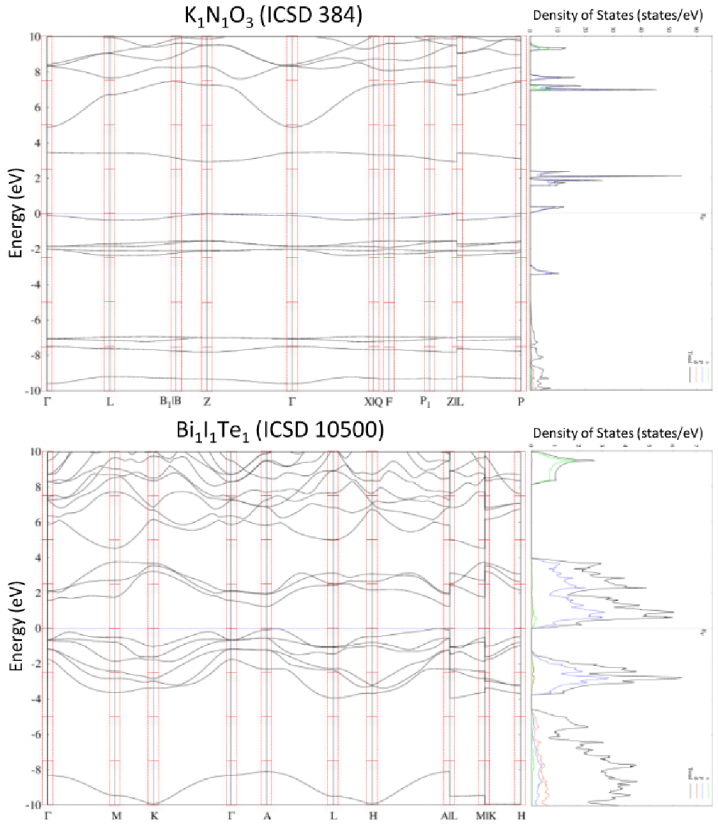
Reference: Isayev, O.; Fourches, D.; Muratov, E. N.; Oses, C.; Rasch, K.; Alexander Tropsha, A.; Curtarolo, S. *Chem. Mater.* **2015**, *27*, 735−743

Molecular Fingerprints

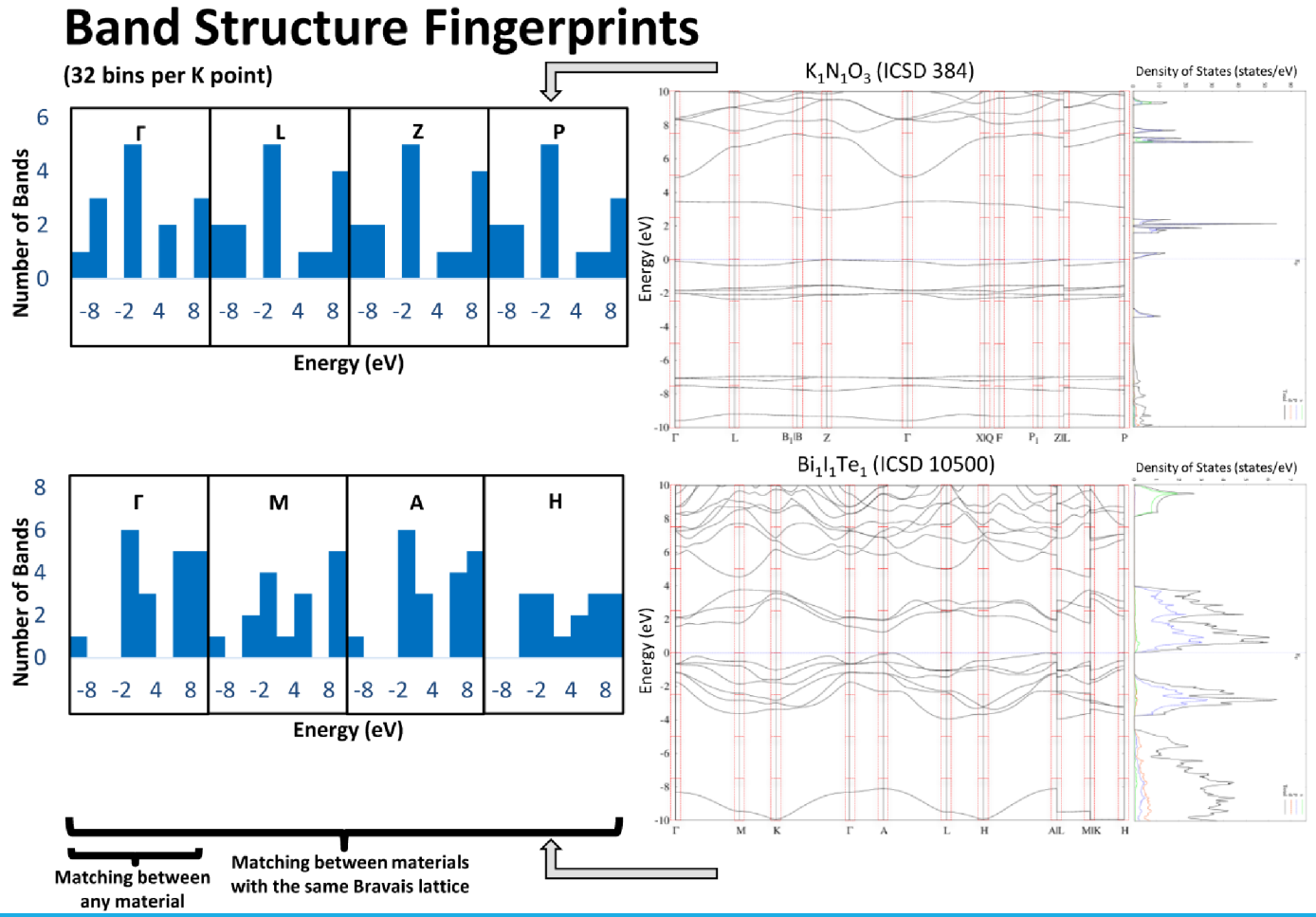


Cao, Y.; Jin, R.; Mirkin, C. *Science*, **2002**, *297*, 1536

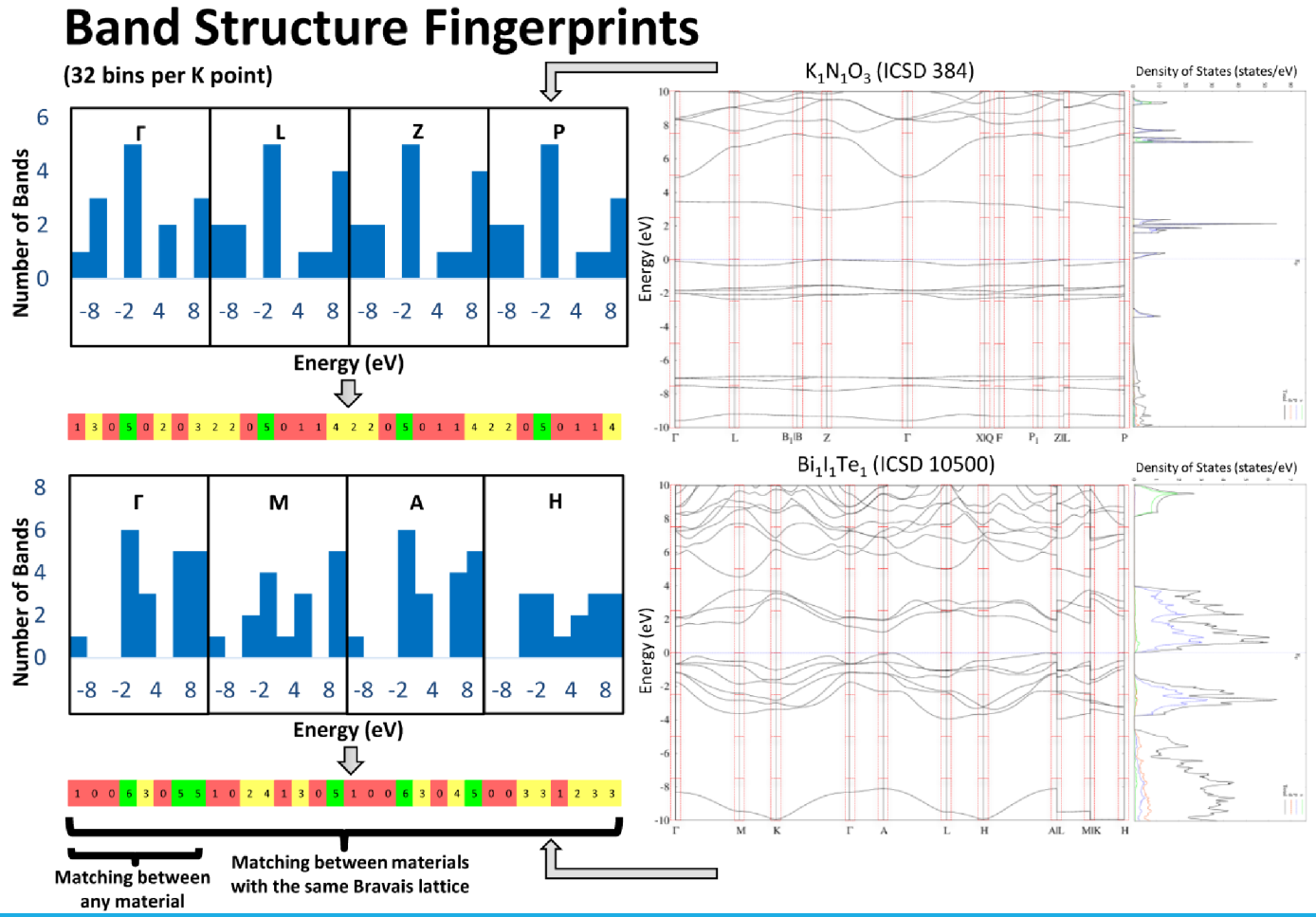
Bandstructures and Density of States are More Complicated than Molecular Spectra



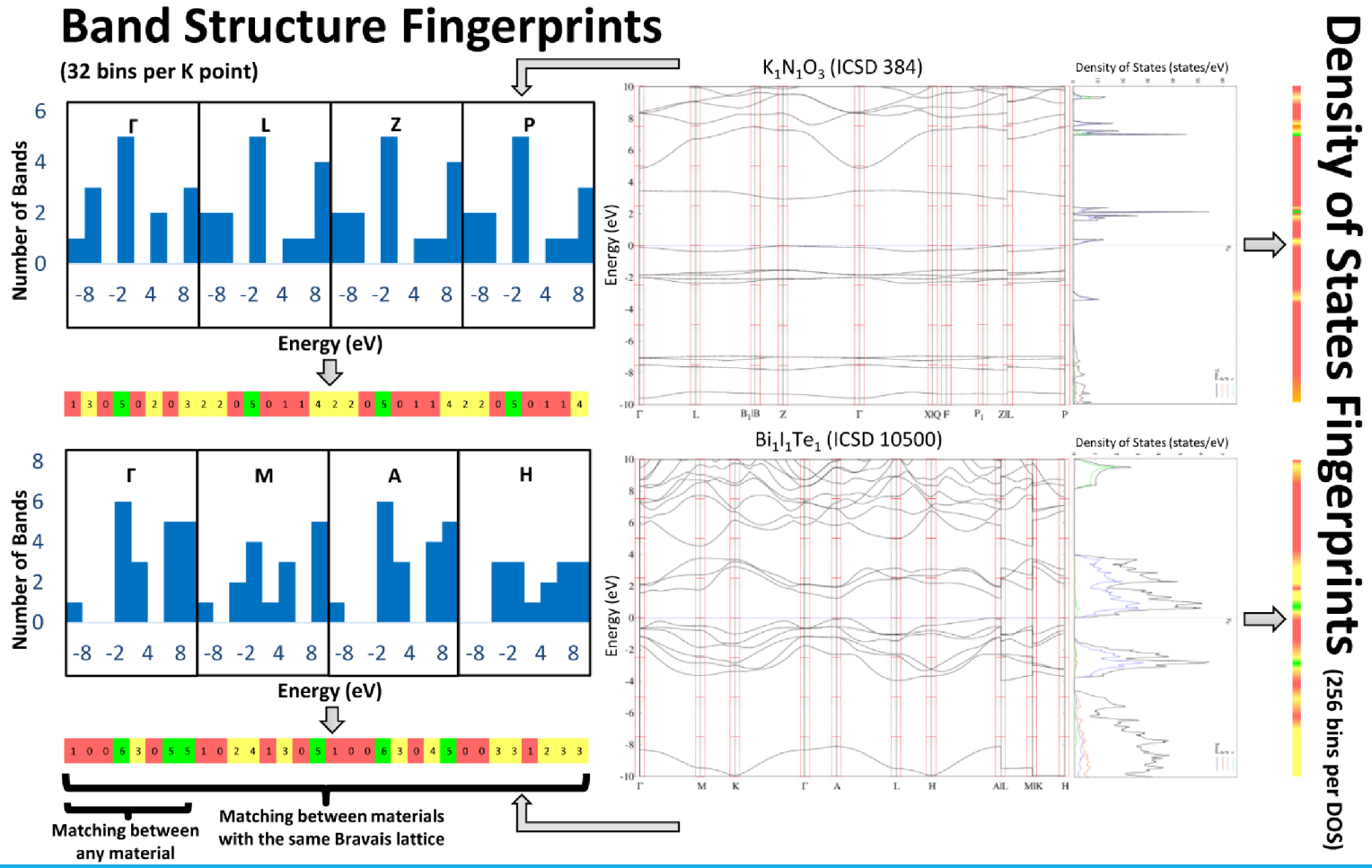
Reducing Data to one Dimension



Vectorizing the Information



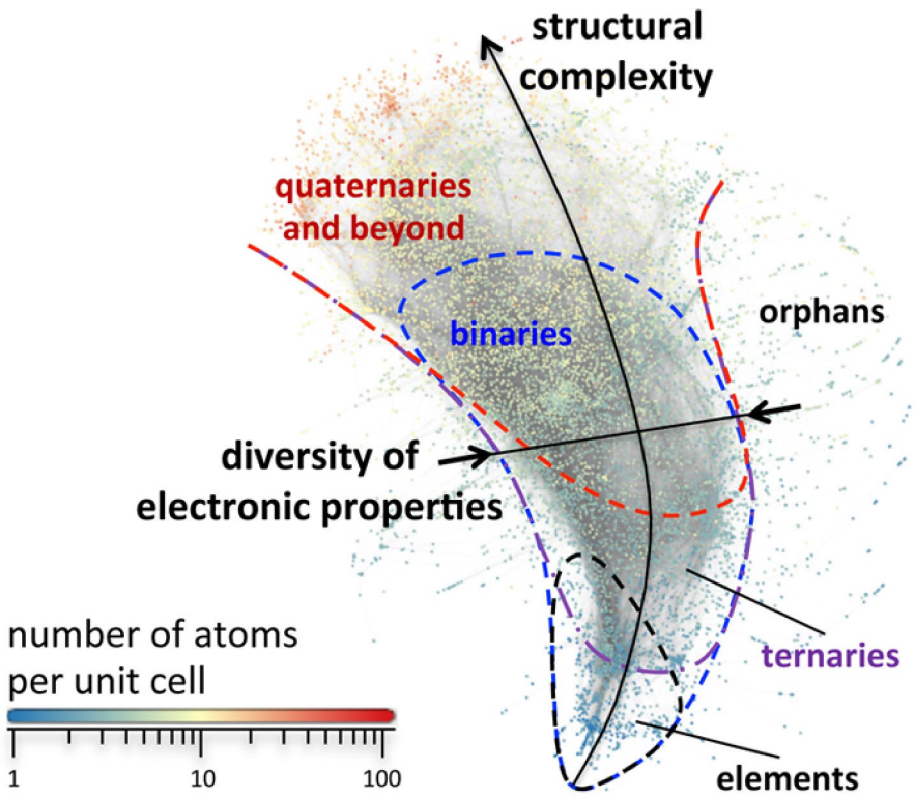
Standardize Density of States



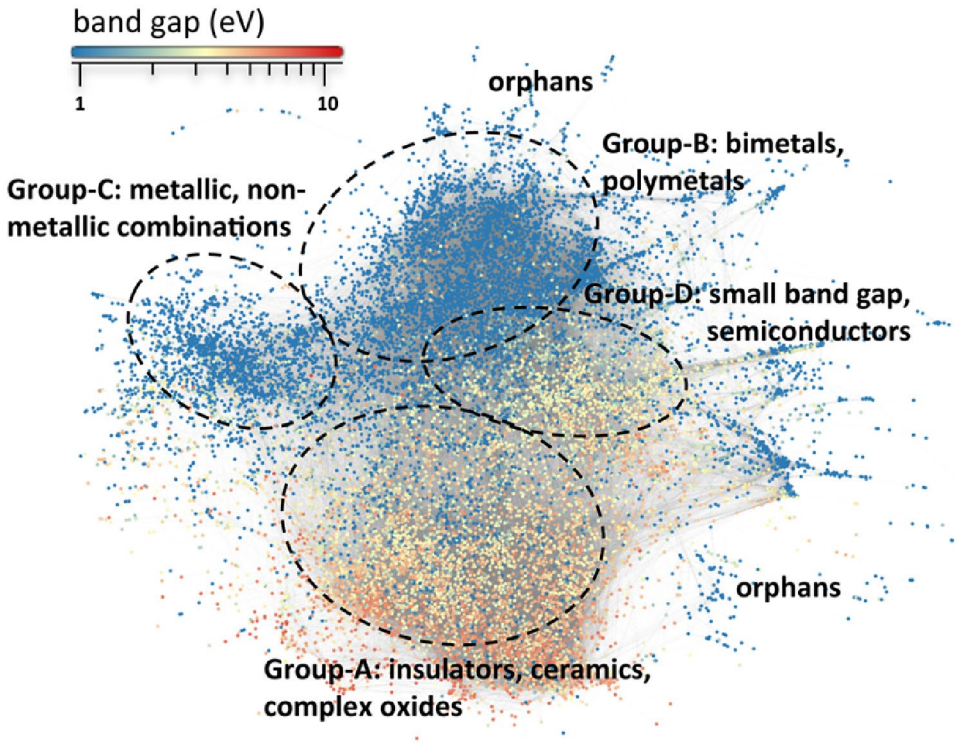
Mapping Materials Based on Similarity

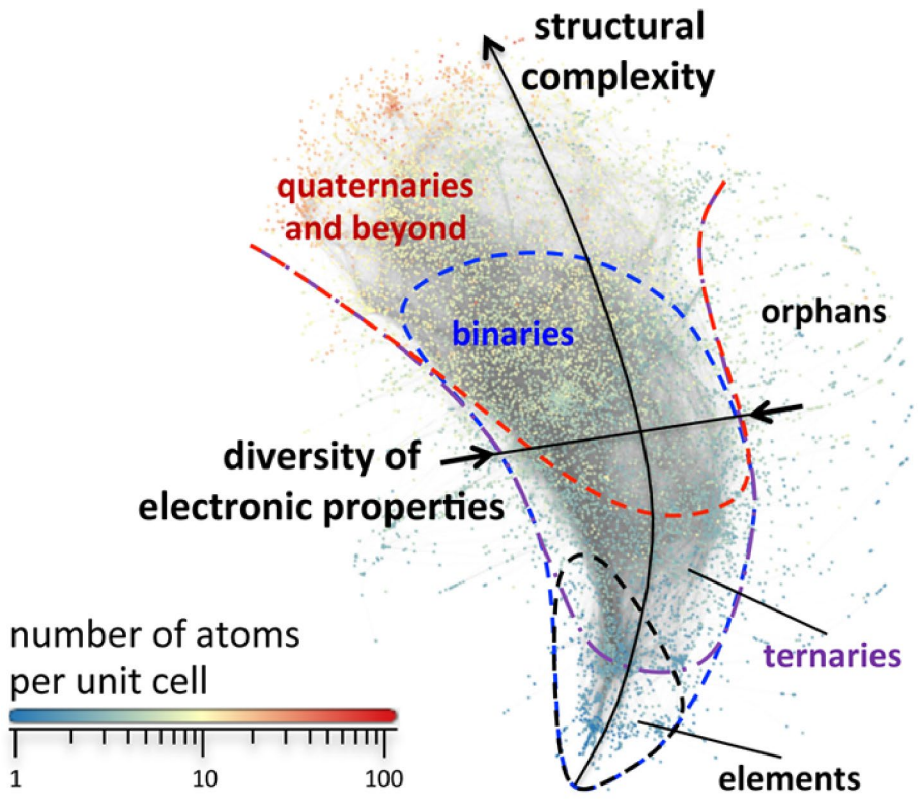
Force directed graph

- Standard Coulomb repulsive force between nodes
- Attractive contribution with a spring constant equal to Tanimoto coefficient between fingerprints
  - $T(a,b) = \frac{N_c}{N_a + N_b - N_c}$
- Threshold for node connection is 0.7
- "The materials networks have been visualized using the Gephi package.[41] The ForceAtlas 2 [42] algorithm, a type of forcedirected layout algorithm, has been used for the graph layout."
  - (41) Bastian, M.; Heymann, S.; Jacomy, M. Int. Conf. Weblogs Social Media 2009, 8, 361–362.
  - (42) Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. PLoS One 2014, 9, e98679

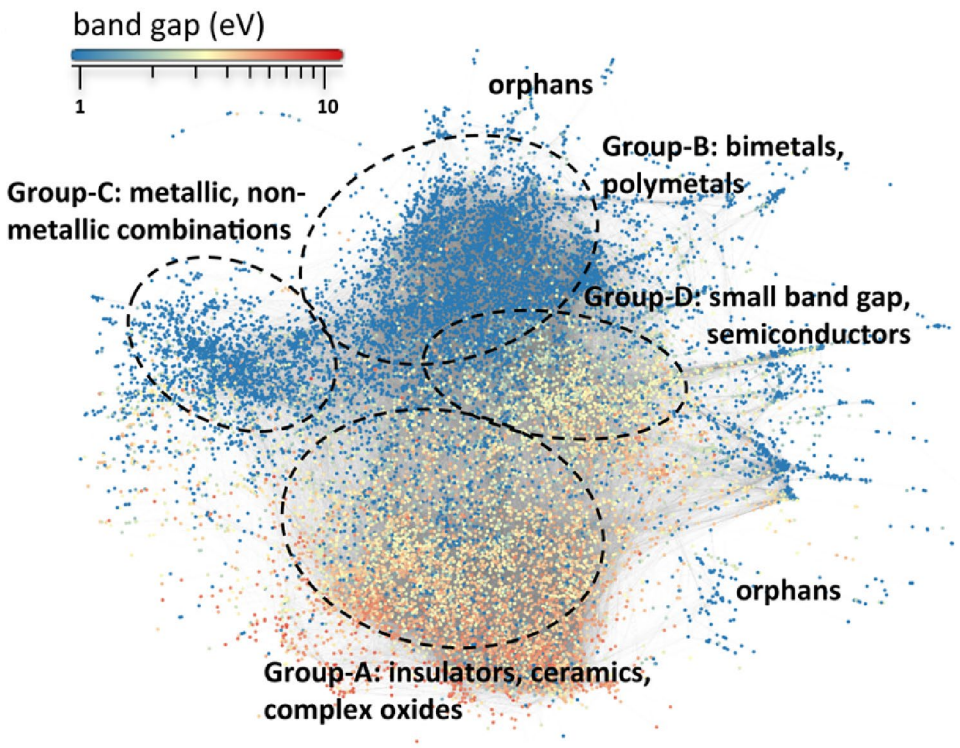


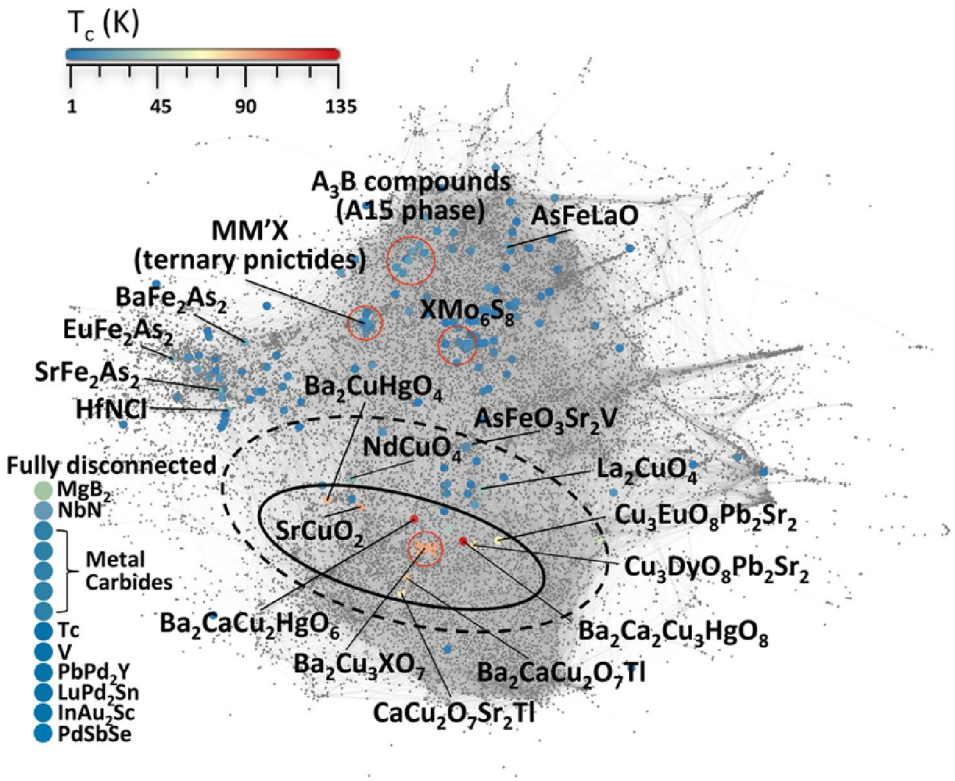
The Bandstructure Fingerprints allow for more seperation



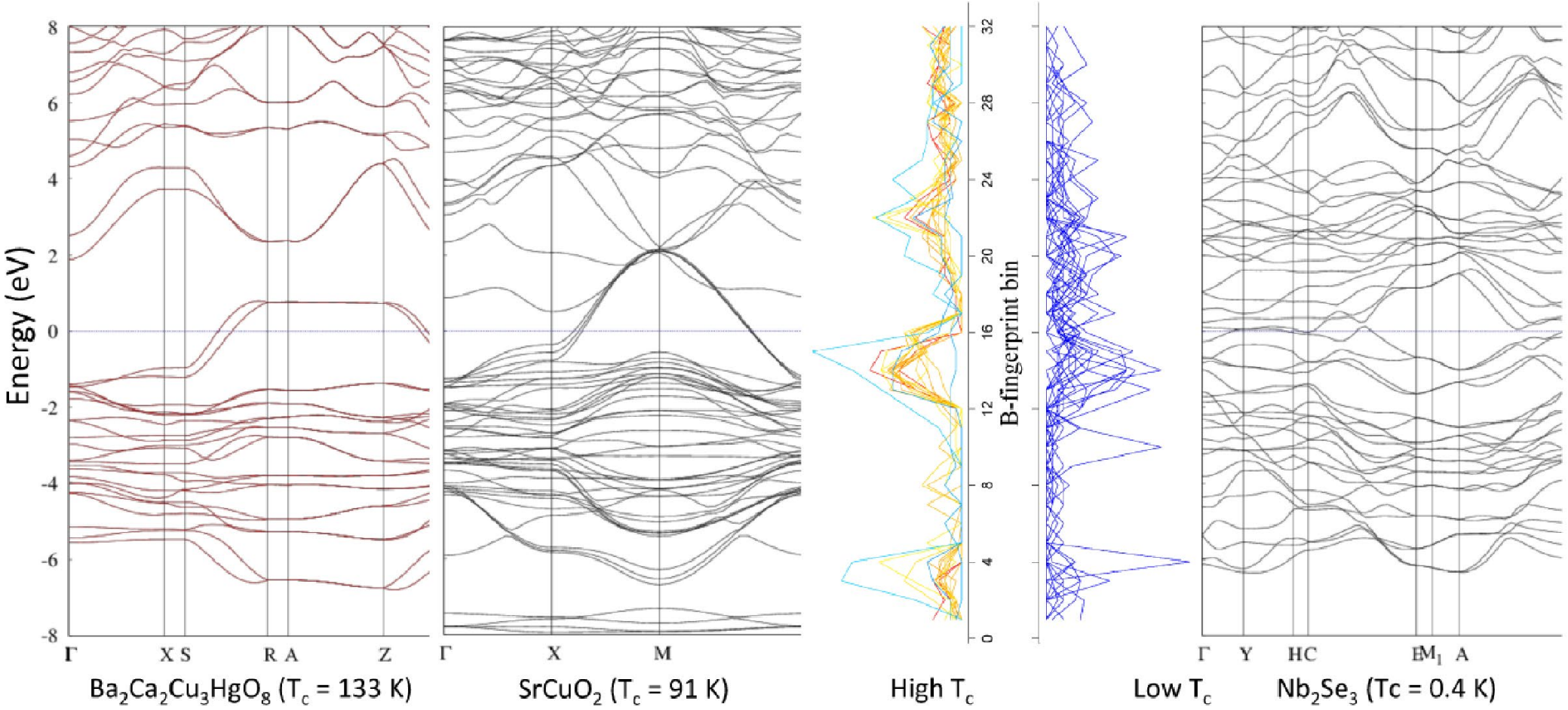


Maps Illustrate regions of High Critical Temperatures for Superconductivity

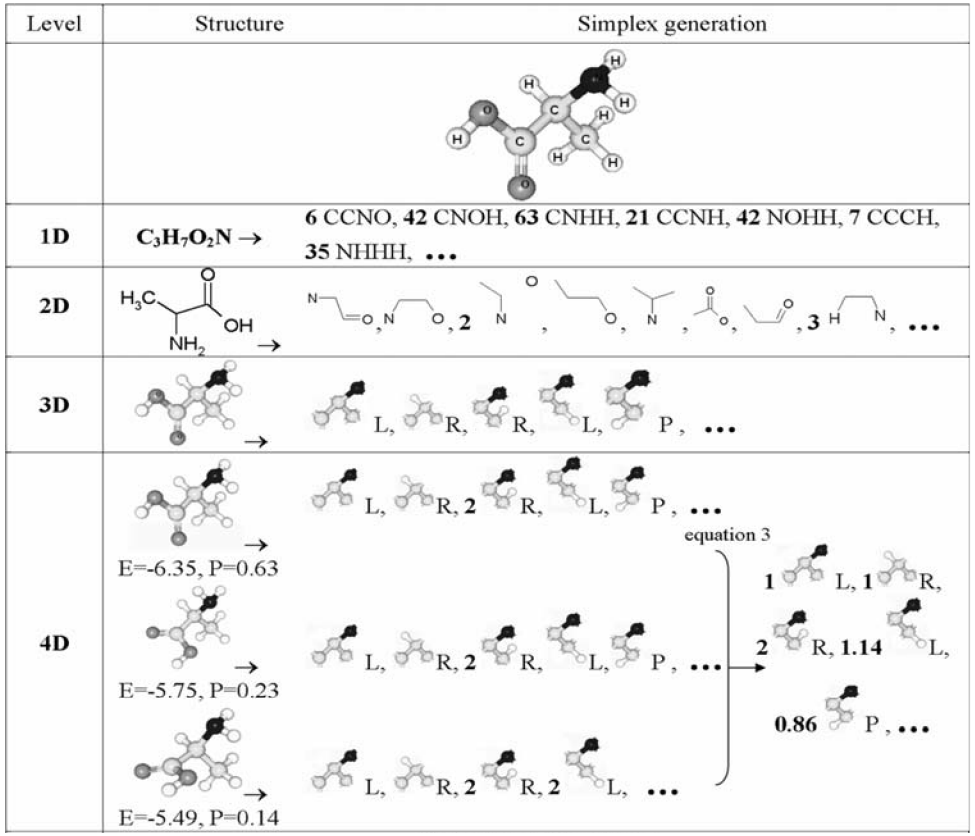


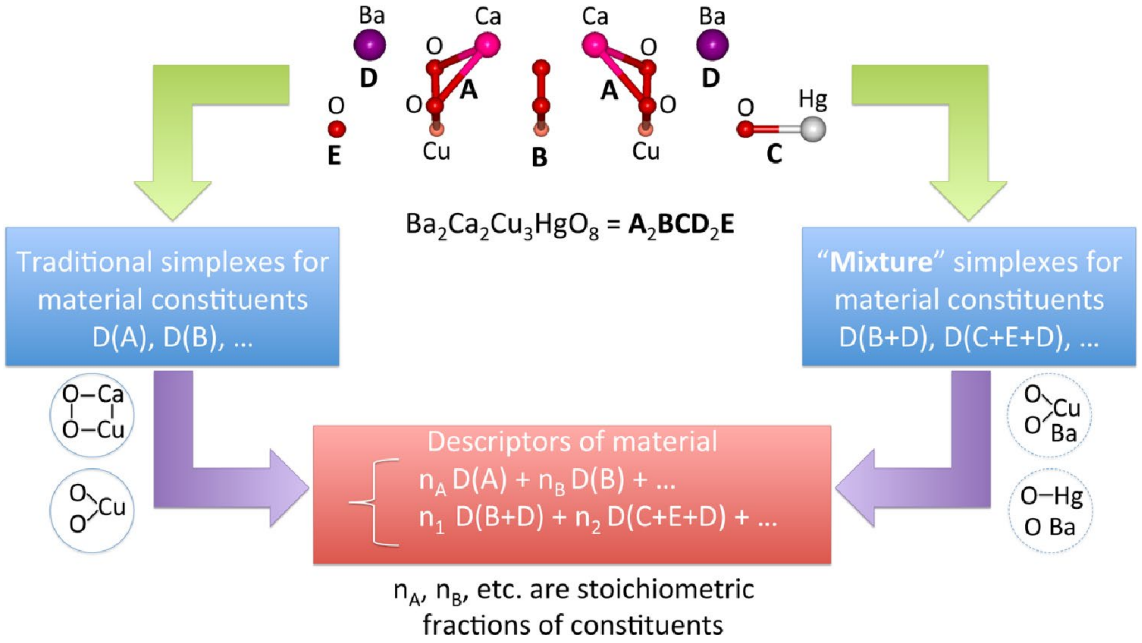


Similarity in Fingerprints



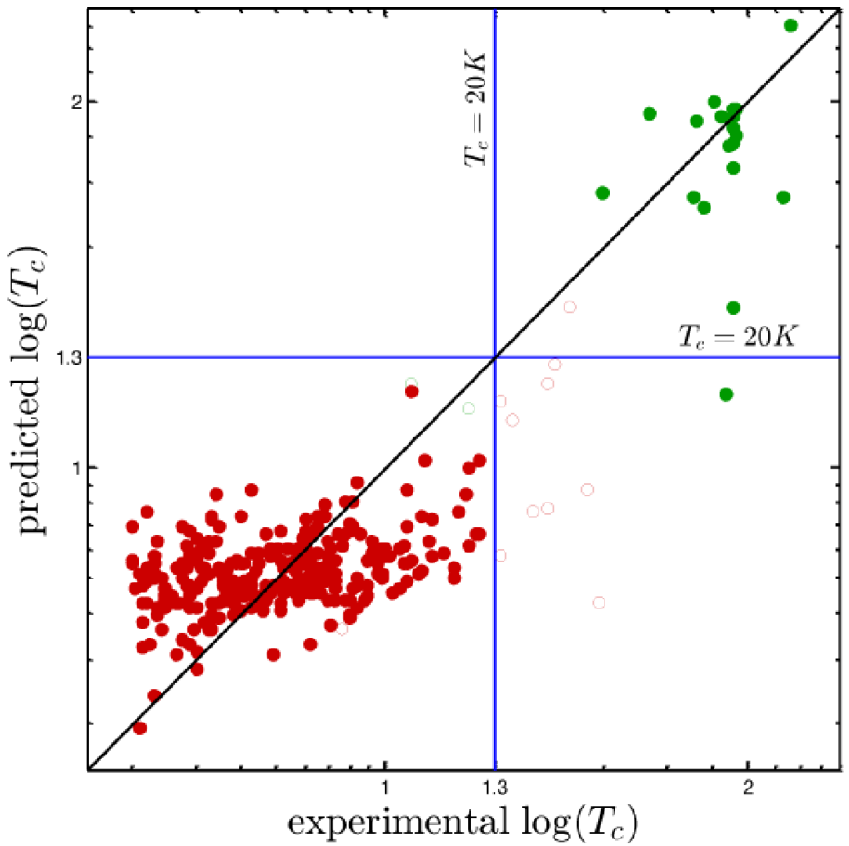
Simplex Representation of Molecular Structure (SiRMS)





Kuz'min, V. E.; Artemenko, A. G. ; Muratov, E.N. *J .Comput. Aided Mol. Des.* **2008,** *22,* 403

Good Classification Model



Data Set

- 295 materials with continuous T_c values ranging from 2 to 133 K
- 464 materials with binary T_c values

Model Generation

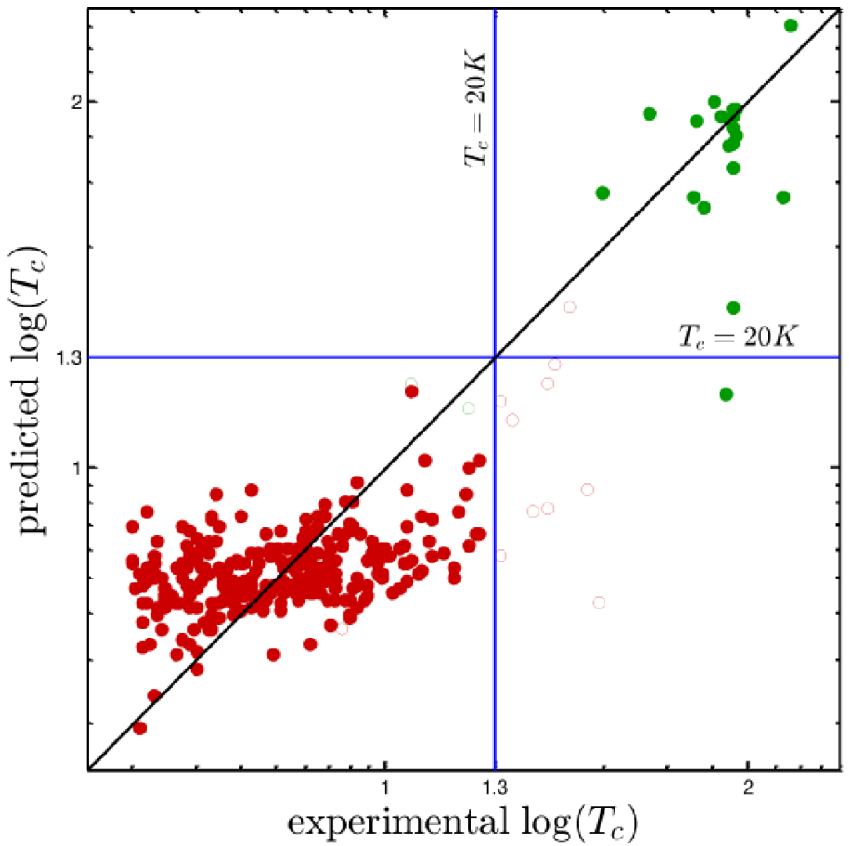
- Random Forest
- Partial Least Squares

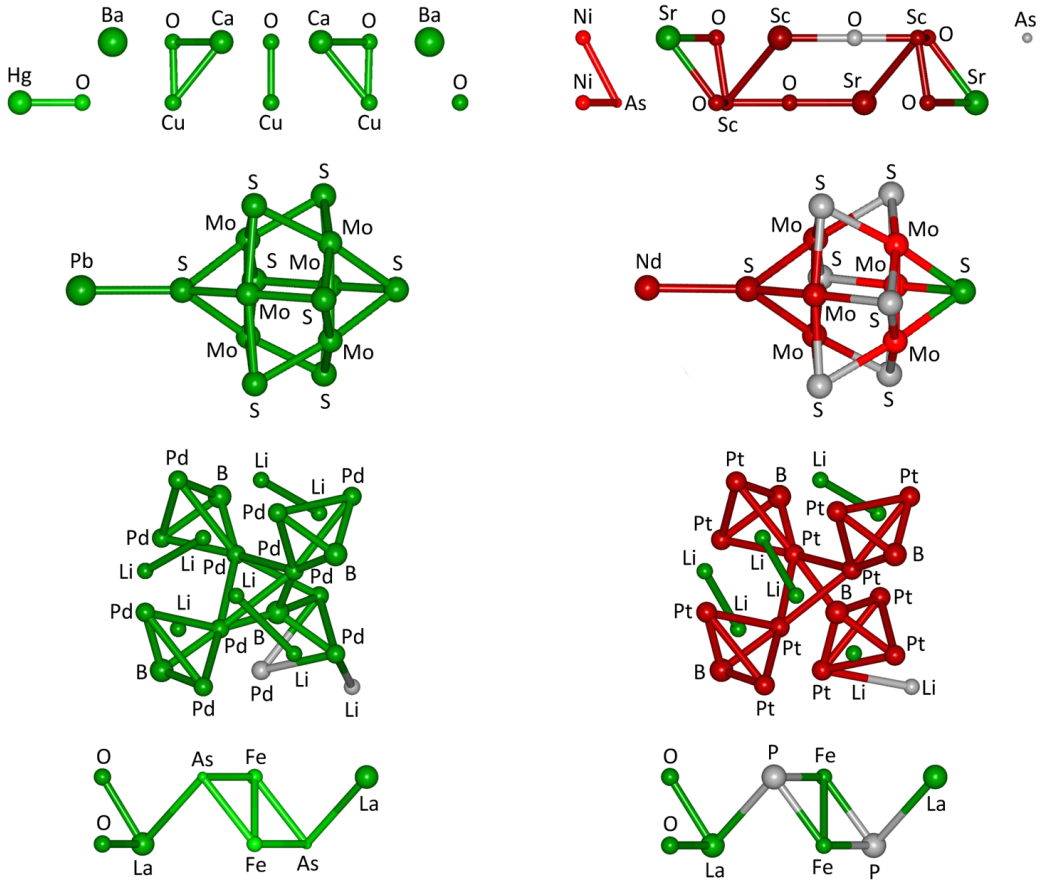
Descriptors

- B- and D-Fingerprints
- Simplex (SiRMS)
- Atomistic

5 fold cross validation

Understanding Structural Impact of Superconductivity





Summary

B- and D- fingerprints provide a good representation of materials

Qualitative cartograms can be developed to map a material space

Quantitative models require more sophisticated descriptors

Cartogram Statistics

	D-fingerprint Network	B-fingerprint Netowrk
Total number of cases	17420	17420
Giant Component	10521 (60.4%)	15535 (89.2%)
Edges	466,000	564,000
Average degree	88.60	72.59
Network diameter	27	23
Power Law	2.745	0.916 (2.04)

Accuracy of the Continuous Model

Model	N	Q^2	RMSE	MAE
RF-SiRMS	295	0.64	0.24	0.18%
PLS-SiRMS	295	0.61	0.25	0.20%
Consensus	295	0.66	0.23	0.18%

Classification Model Accuracy

	No Applicability Domain	With Applicability Domain
Number of Materials	464	464
Number of Predictions	464	451
Accuracy	0.99	0.99
Sensitivity	0.66	0.77
Specificity	1.00	1.00