Using Machine Learning, Data Mining, and Experimental Data to Design Ultrastable Metal-Organic Frameworks

Aditya Nandy[1, 2]**, Shuwen Yue**[2]**, Changhwan Oh**[2,3]**, and Heather J. Kulik**[1, 2]

nandy@mit.edu and hjkulik@mit.edu

1. Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139
2. Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139
3. Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

MIT Portugal
2022 Annual Conference

Introduction

Metal-organic frameworks (MOFs) are hybrid inorganic-organic single site materials that can be used for gas separations, gas storage, and catalysis. They are of tremendous importance for climate change due to their ability to be molecular sieves for CO_2 capture and conversion. Due to their modular nature, significant efforts to construct combinatorial MOF libraries have led to the rise of hypothetical MOF databases. Although this modular nature enables MOFs to be useful for design purposes, they suffer from serious stability issues that limit their practical use. A MOF may have promising properties, but be unstable under thermal conditions. This has been a 20-year challenge that remains unsolved. We utilize existing methods of MOF construction and combine them with information regarding MOF stability to build a new and improved database of ultrastable MOFs for thermal applications.

Methods

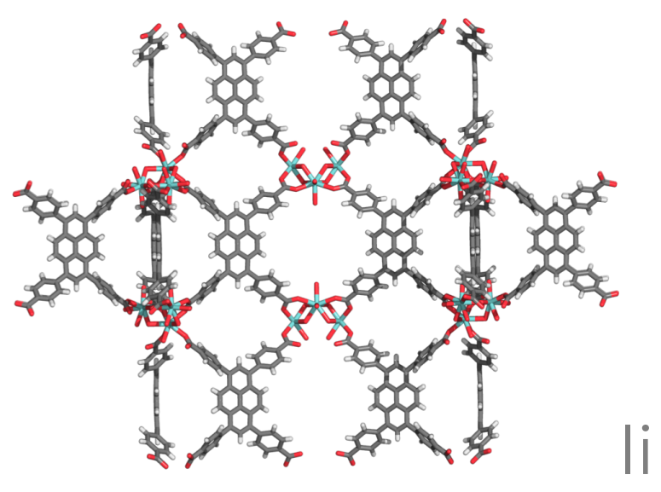
We extended revised autocorrelations (RACs) that worked to fingerprint transition metal complexes[1] for machine learning (ML) to MOFs.[2] RACs have different starts / scopes that were adjusted for MOF connectivities.

$$P_d = \sum_i \sum_j P_i P_j \delta(d_{ij}, d)$$

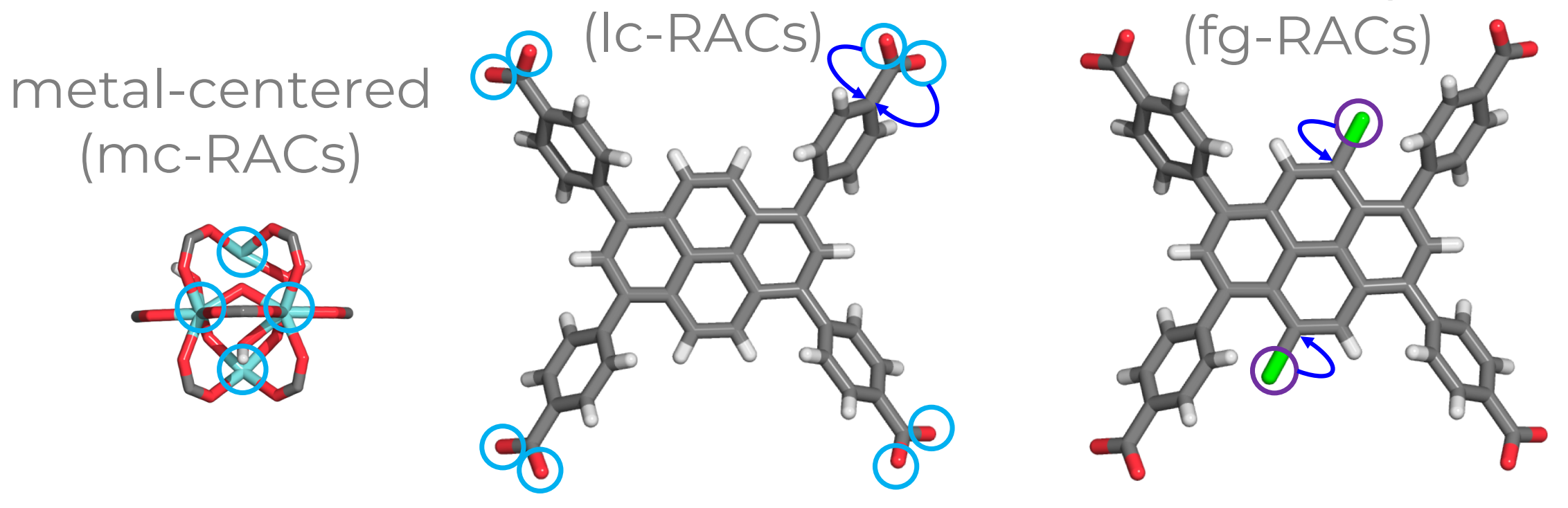
$$P'_d = \sum_i \sum_j (P_i - P_j) \delta(d_{ij}, d)$$

$$\underbrace{\hphantom{P'_d = \sum_i \sum_j (P_i - P_j) \delta(d_{ij}, d)}}$$

Products and differences of atomic quantities on molecular graph
P = Z, T, I, S, χ

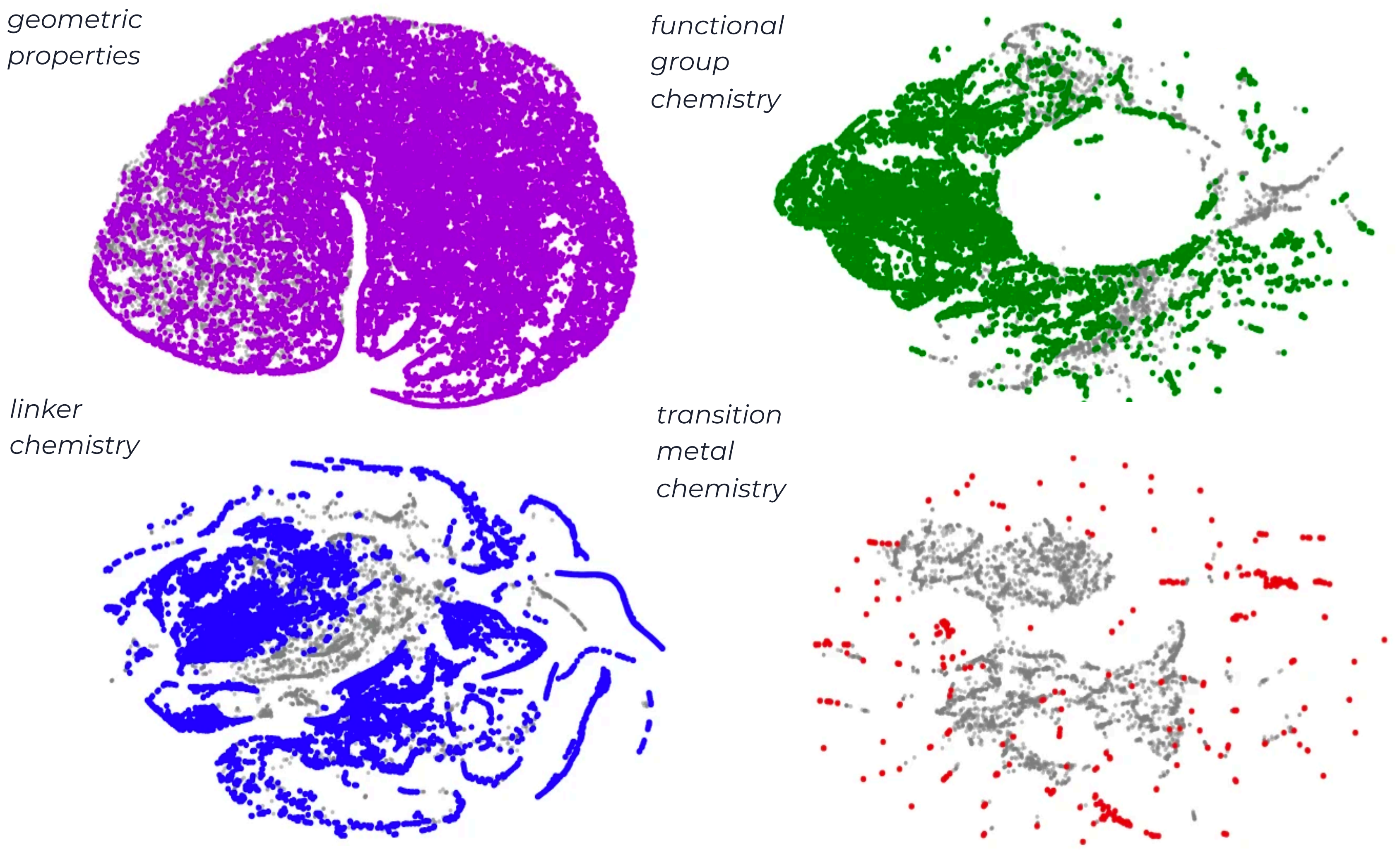


to fingerprint a MOF like NU-1000, we first split the MOF into its constituent pieces and then use different scope RACs on the molecular graph



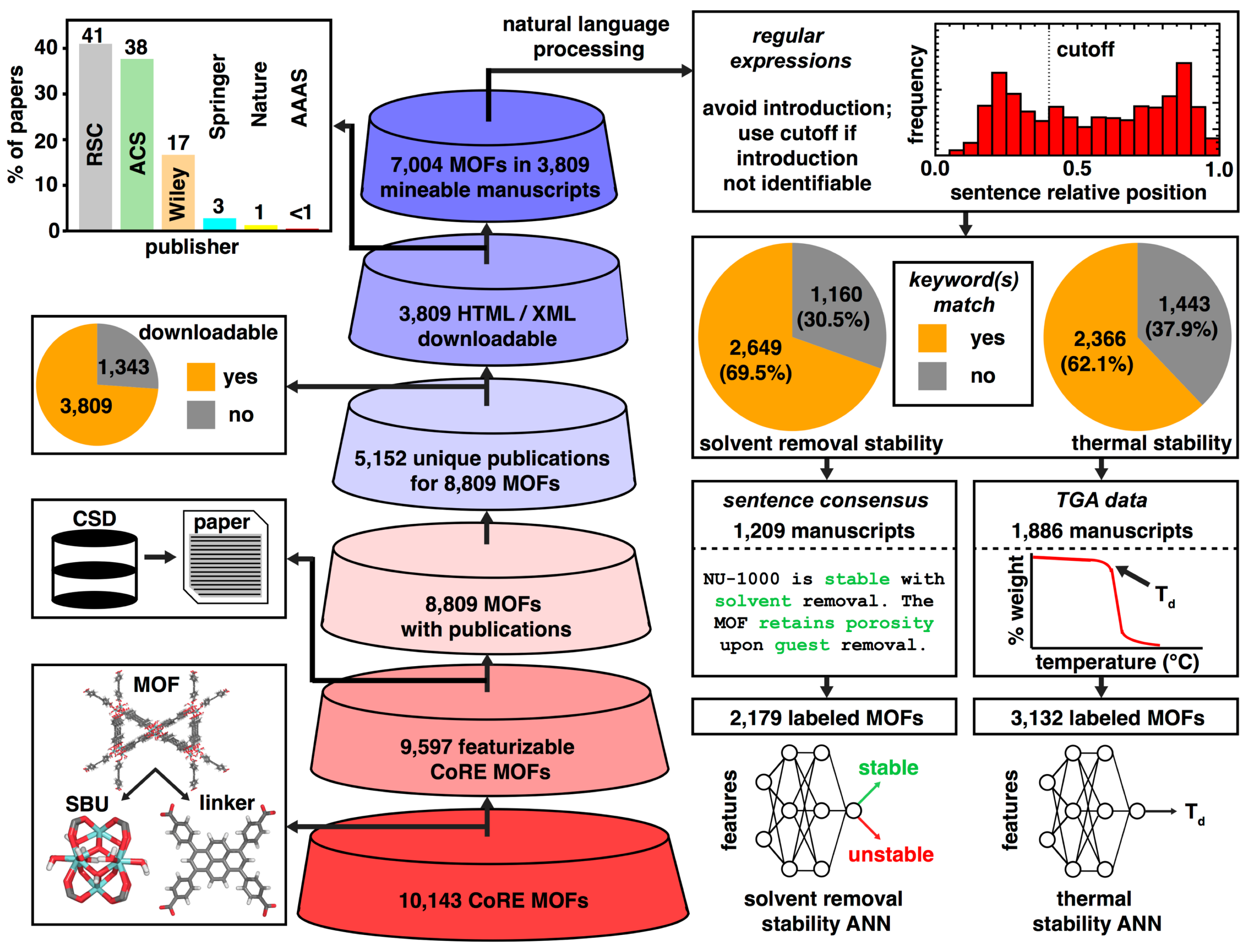
Hypothetical MOF Diversity

RACs enable us to analyze existing hypothetical MOFs (colored) and compare them to the full design space of experimental and hypothetical MOFs (gray). We find that hypothetical MOFs lack the transition metal chemistry present in the experimental space, meaning that we need to use experimental data to span a property space for model predictions. This produces an open challenge: how do we get experimental data in large quantities?



2D projections are t-SNE plots in subsets of MOF RACs

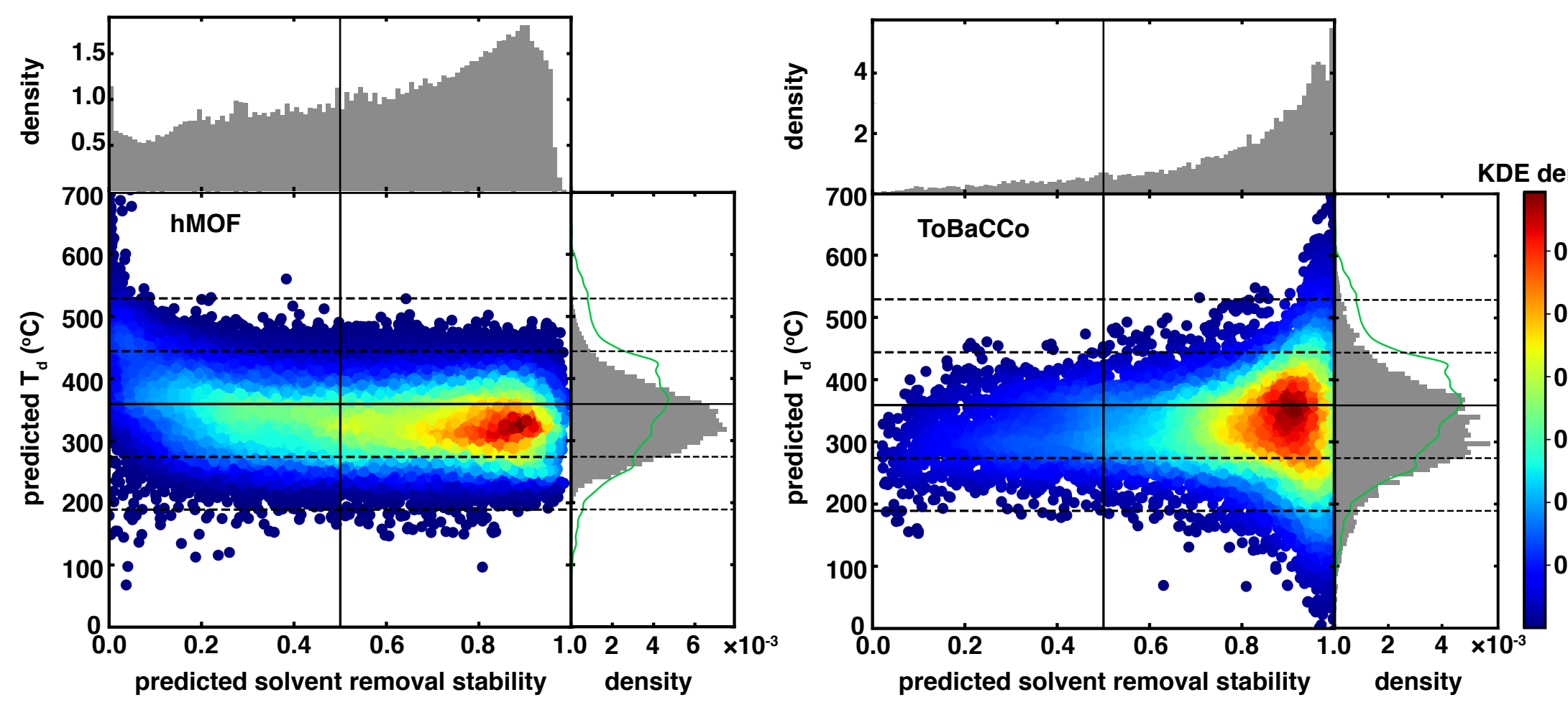
Mining Literature Data



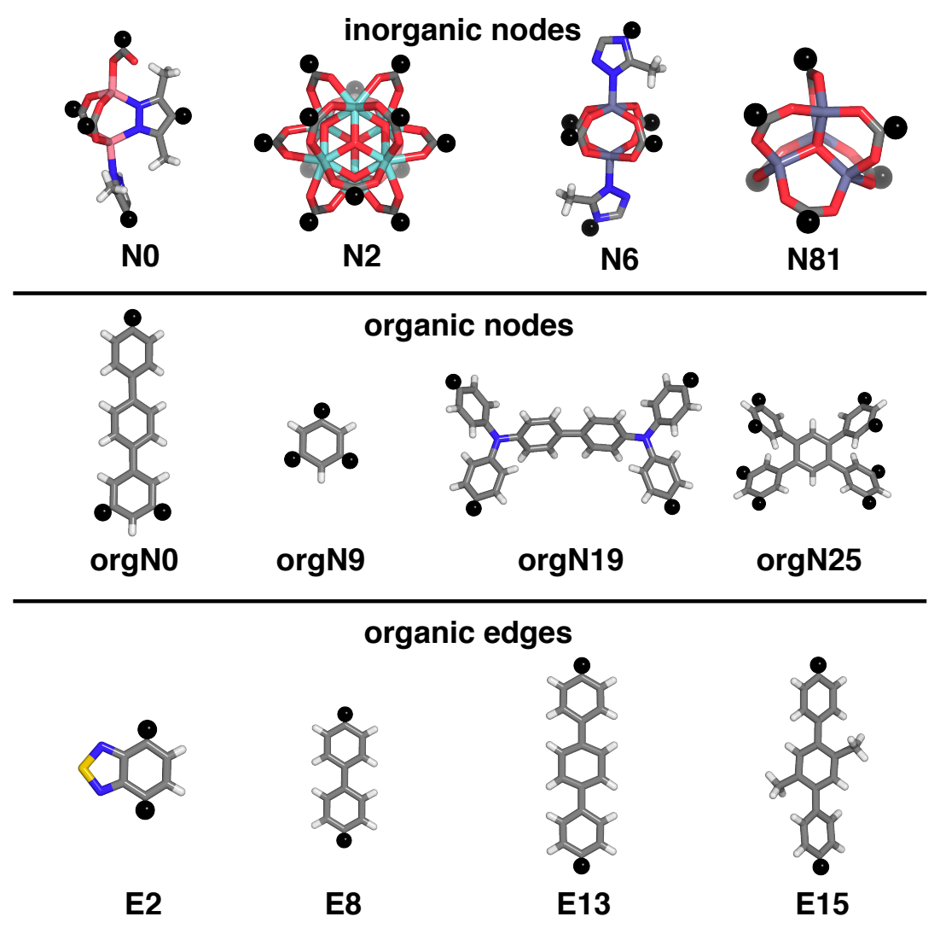
The extant MOF literature provides an opportunity[3] for us to mine information[4] on MOF stability to then train artificial neural networks (ANNs) on experimental data, mapping RACs to stability labels. We can use these ANNs to screen new MOF databases.

Analyzing Existing Databases

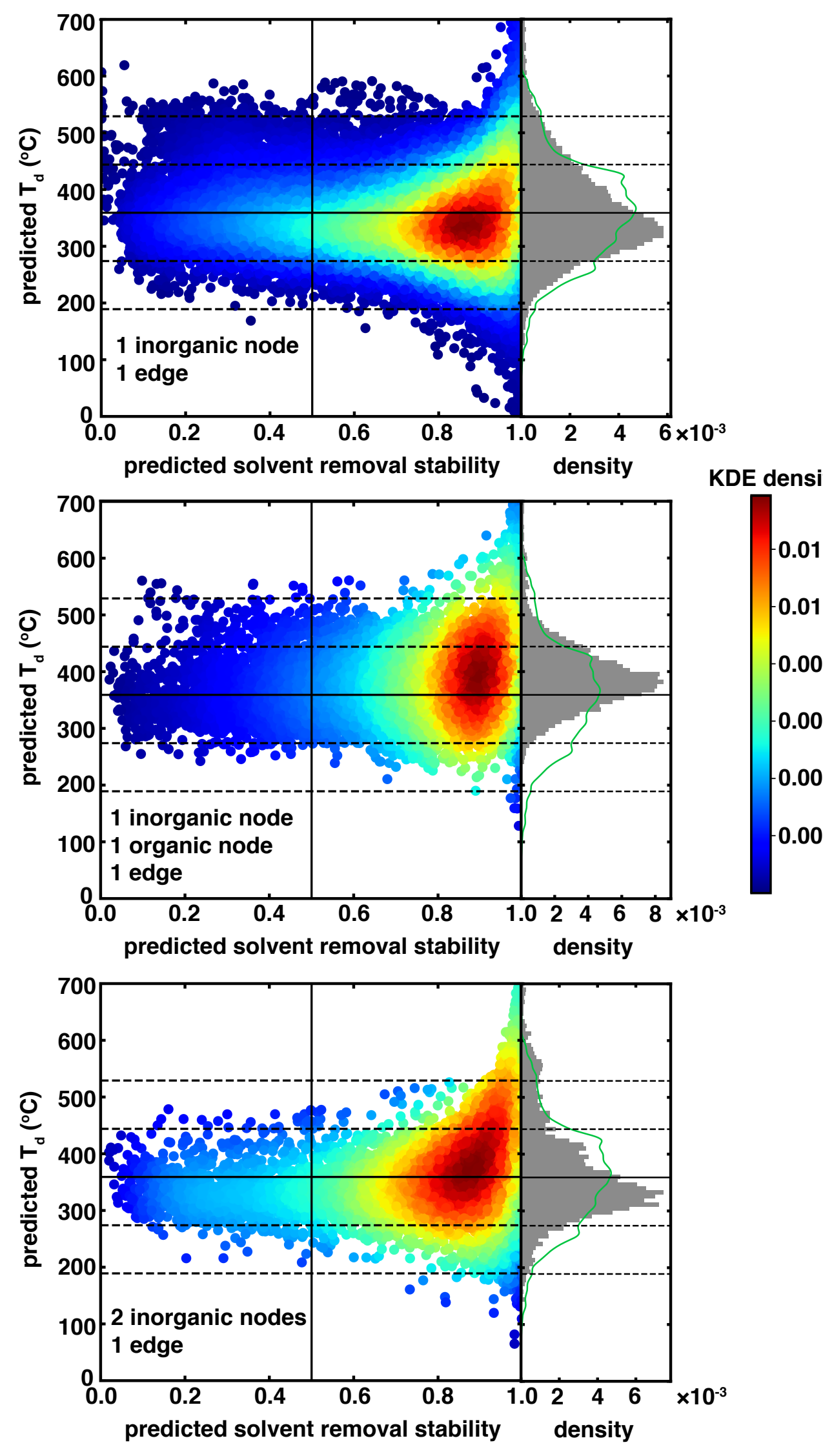
We use our ANNs to evaluate the stabilities of hypothetical MOF databases and find that these databases do not contain MOFs that are particularly stable.



Designing Stable MOFs



We can use existing experimental databases to mine realistic inorganic and organic building blocks that we can recombine into novel but realistic MOFs.



After mining the building blocks, we can use existing tools to recombine these building blocks into ~100,000 new MOFs. Our previously trained ML models enable us to make informed predictions on these new MOFs, which demonstrate an order of magnitude enrichment in MOFs that are likely to be stable. We can then investigate these MOFs for gas separation and storage.

Conclusions

1. ML models can learn experimental MOF stability and inform the design of new MOFs that are likely to be stable
2. Mining building blocks from existing experimental MOFs and recombining them highlights the types of MOFs that should be sought for design of new gas separation or capture materials.

Co-funded by:

under the Seed Project: Engineering Metal-Organic Frameworks for Stability in Gas Storage Applications
MIT Portugal Program: Climate Science and Climate Change

References

[1]Janet, J.P. and Kulik, H.J. *J. Phys. Chem. A.* **2017**, *121*, 8939-8954.
[2]Moosavi, S.M.; Nandy, A.; Smit, B.; Kulik, H.J. *et al. Nat. Commun.* **2020**, *11*, 4068.
[3]Nandy, A.; Duan, C.; and Kulik, H.J. *J. Am. Chem. Soc.* **2021**, *143*, 17535-17547.
[4]Nandy, A.; Terrones, G.; Kulik, H.J. *et al. Sci. Data.* **2022**, *9*, 74.

visit **mofsimplify.mit.edu** to give us feedback!