Integrating Materials and Manufacturing Innovation
https://doi.org/10.1007/s40192-021-00247-y

Supplementary Information for the article "CrabNet for Explainable Deep Learning in Materials Science: Bridging the Gap Between Academia and Industry"

Authors: Anthony Yu-Tung Wang[1], Mahamad Salah Mahmoud[2], Mathias Czasny[1], and Aleksander Gurlo[1]

Affiliations:

(1) Technische Universität Berlin, Institute of Material Science and Technology, Fachgebiet Keramische Werkstoffe / Chair of Advanced Ceramic Materials, Straße des 17. Juni 135, 10623 Berlin, Germany
(2) Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA

Correspondence: Correspondence should be sent to Anthony Yu-Tung Wang at anthony.wang@ceramics.tu-berlin.de

DOI: 10.1007/s40192-021-00247-y

URL of CrabNet GitHub with source code: https://github.com/anthony-wang/CrabNet

DOI of trained CrabNet model weights: https://doi.org/10.5281/zenodo.4633866

Note: all the ZIP files mentioned in this document (**ESM1.zip** to **ESM5.zip**) are available for download from figshare at the following address: https://doi.org/10.6084/m9.figshare.16837999

Supplementary Figures

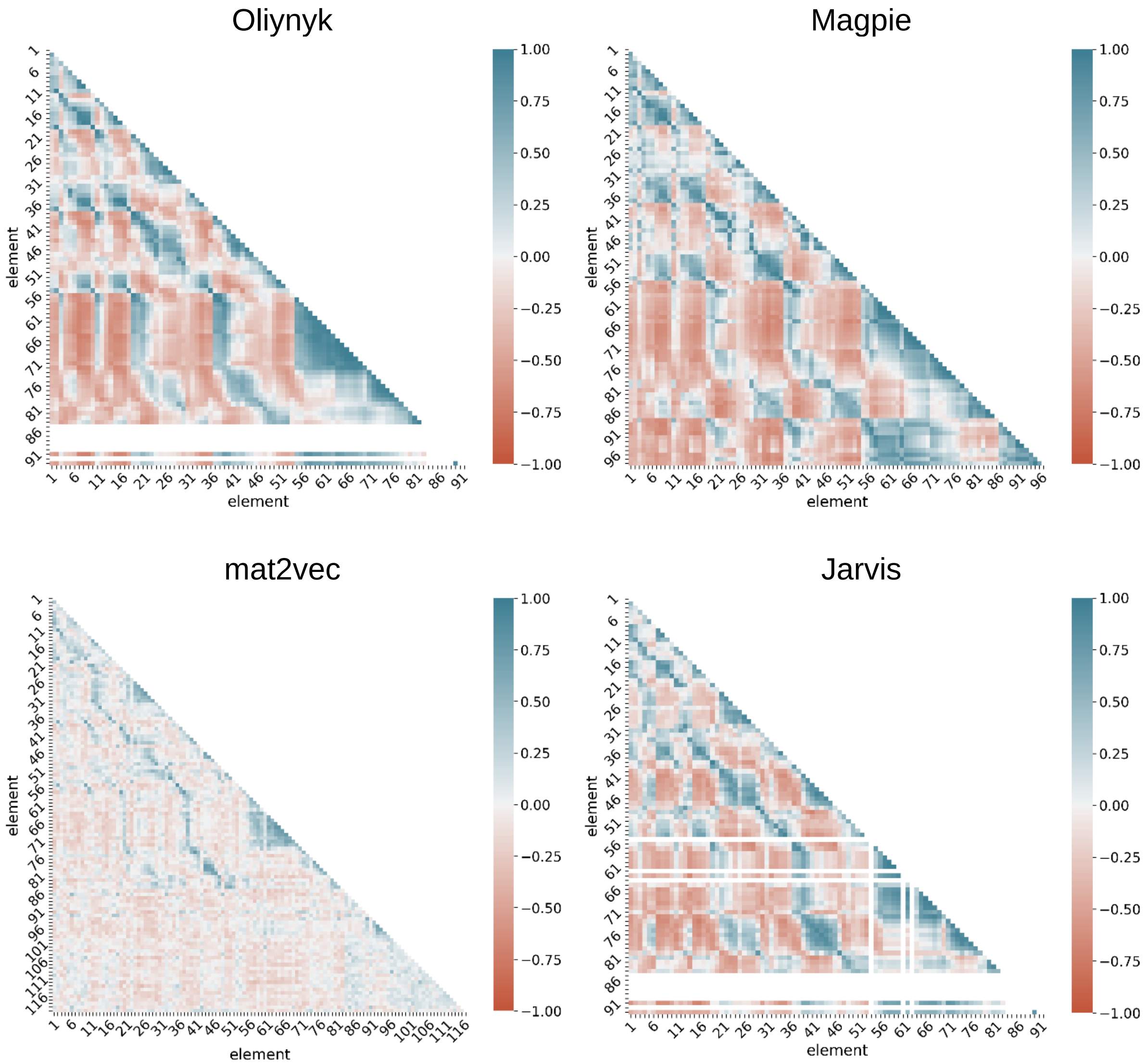

Figure S-1. Uncropped Pearson correlation plots for the element property feature sets Oliynyk, Magpie, mat2vec and Jarvis. The *x*- and *y*-axes are labeled with the atomic numbers. Each cell at coordinate (*x, y*) represents the correlation between the corresponding elements with atomic numbers *x* and *y*. Blue represents a high correlation and red represents a high anticorrelation. Empty rows indicate that no element vector is available.

See the **ESM1.zip** file for the full-resolution correlation plots and for interactive HTML versions of the plots (Internet connection required).

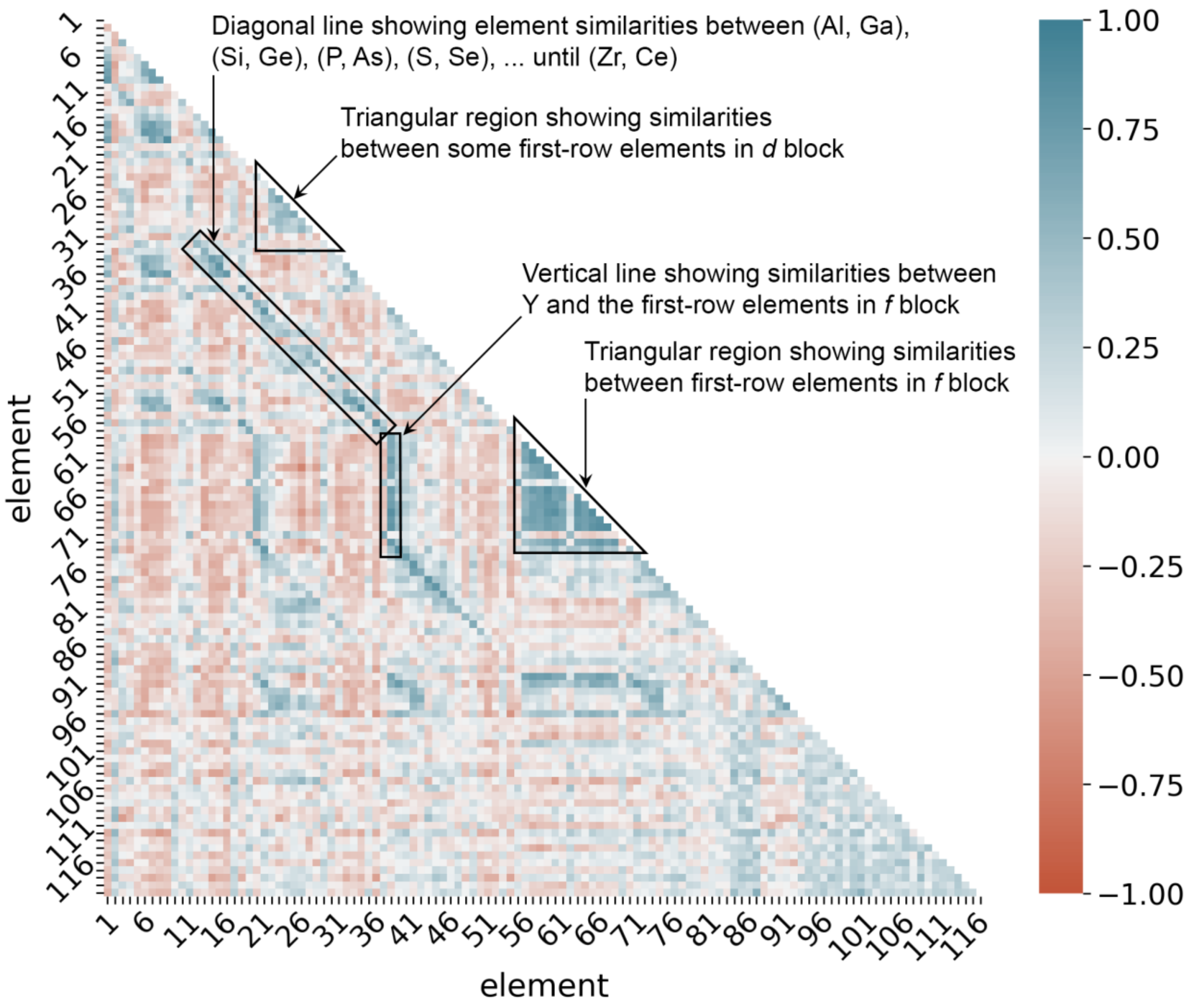

Figure S-2. Annotated heatmap of the Pearson correlation matrix between element vectors extracted from HotCrab for the property OQMD_Bandgap. Here, some regions of interesting similarities between element representations are labeled. These representations are learned entirely from data. The *x*- and *y*-axes are labeled with the atomic numbers. Each cell at coordinate (*x, y*) represents the correlation between the corresponding elements with atomic numbers *x* and *y*. Blue represents a high correlation and red represents a high anticorrelation.

See **ESM2.zip** file for more full-resolution correlation plots (including interactive versions) using element vectors extracted from CrabNet and HotCrab.

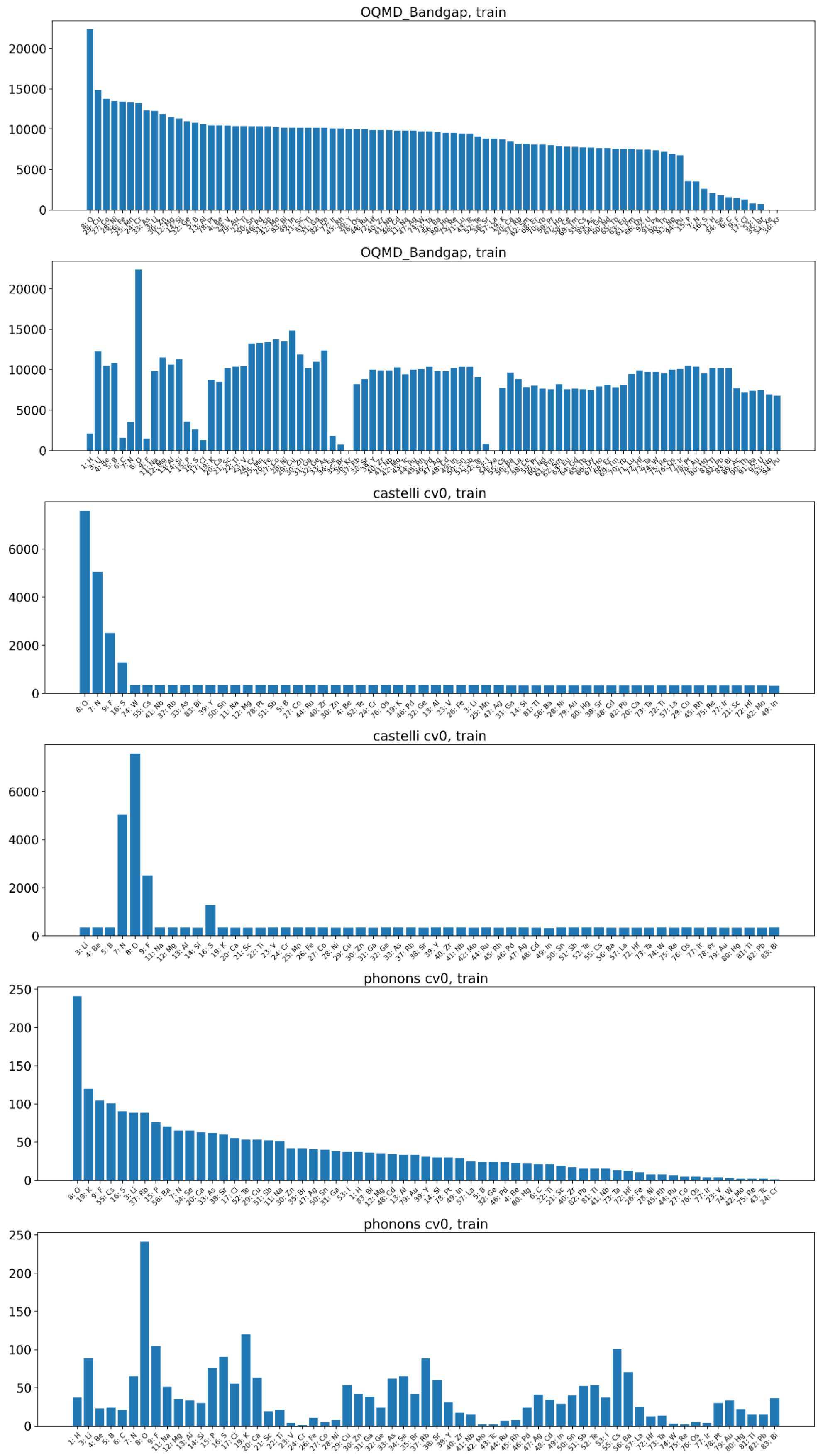

Figure S-3. Example element prevalence plots, sorted by (top) number of occurrences and (bottom) atomic number. See **ESM3.zip** file for more plots.

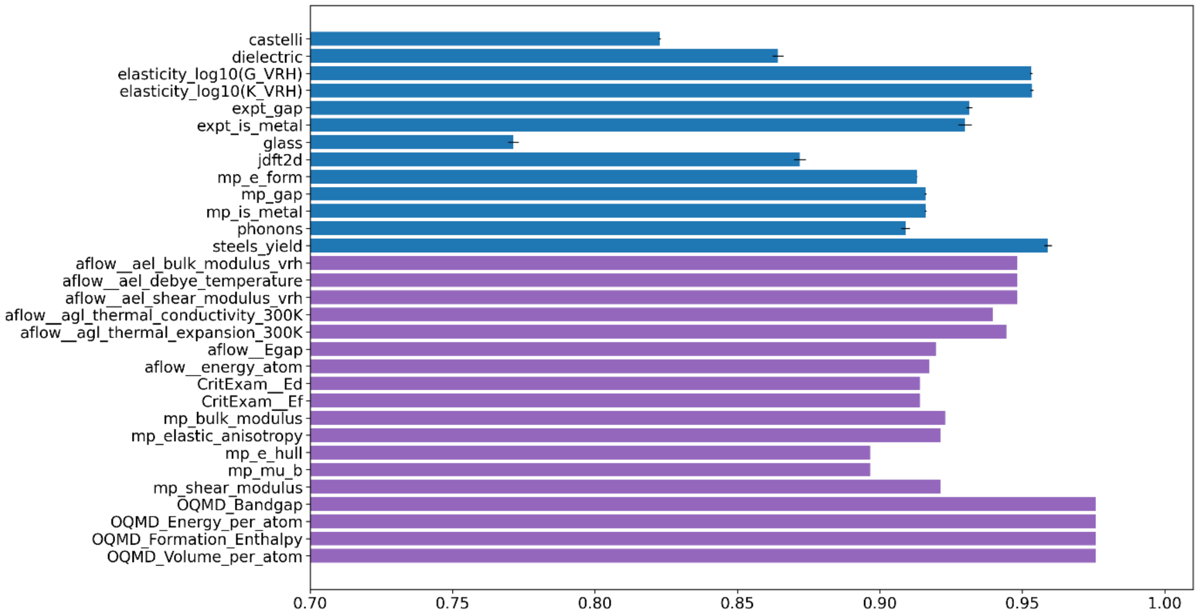

Figure S-4. Shannon equitability indices calculated from the training data splits of the 28 reported datasets in the CrabNet publication. (top) Matbench dataset, (bottom) Extended dataset.

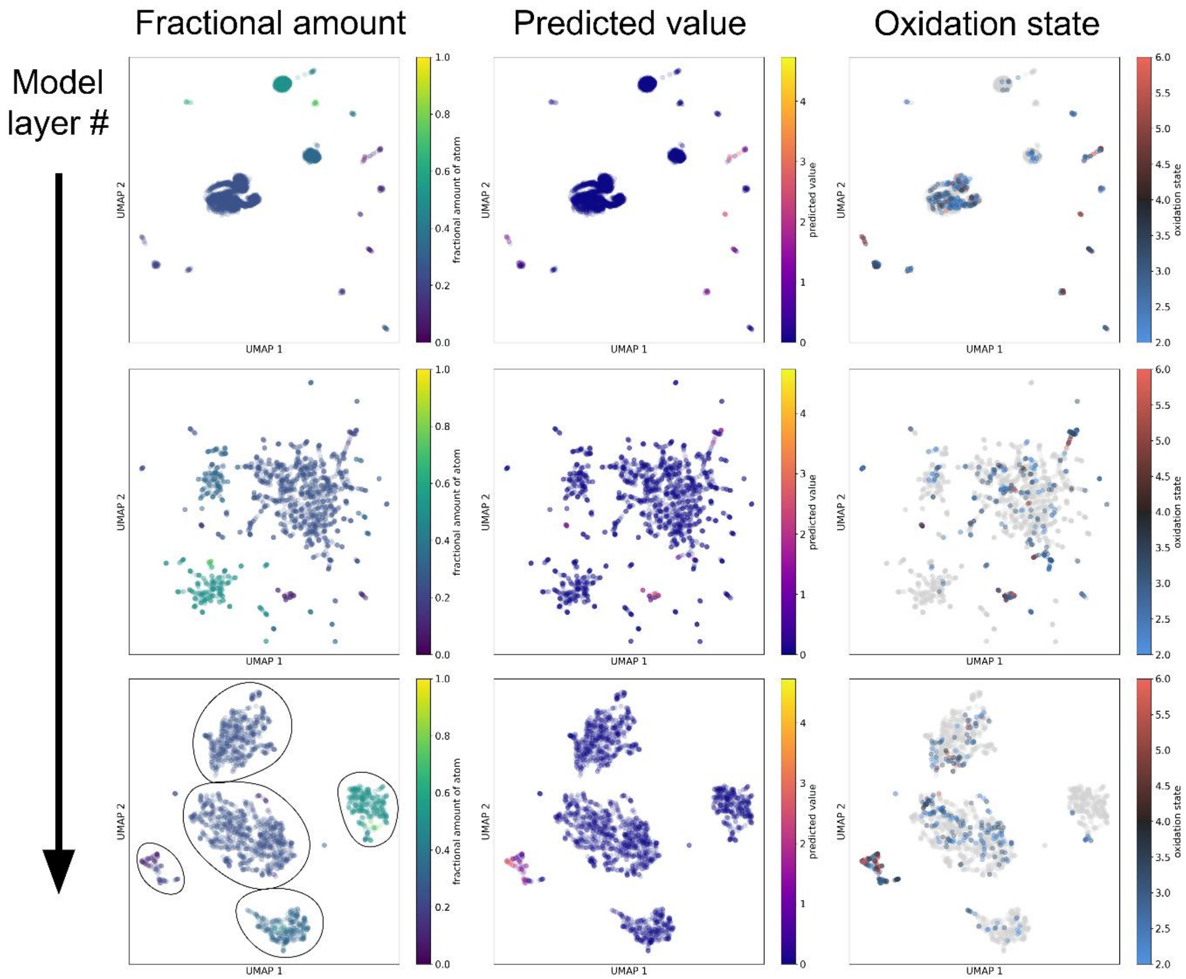

Figure S-5. Vector representations of the chromium element in 2800 different chemical environments and at different layers of the HotCrab model. Each point shows the model-internal representation of the chromium atom, after the information regarding the other atoms in the chemical environment have been introduced via HotCrab through the three attention layers (top row to bottom row). The points are colored by: (left column) the fractional abundance of chromium, (center column) the predicted value of the compound, and (right column) the predicted oxidation state of chromium, where gray points indicate that the oxidation state was unable to be predicted. Five clusters are outlined in the bottom-left plot.

See the **ESM4.zip** file for the individual full-resolution plots and for interactive HTML versions of the plots, for the elements silicon and chromium (Internet connection required).

Supplementary Videos

The attention video files are collected and compressed in a ZIP file (**ESM5.zip**). Please find the files on the figshare website as described on the cover page of this document.