MaterialNet: A web-based graph explorer for materials science data

Roni Choudhury[1], Muratahan Aykol[2], Samuel Gratzl[1], Joseph Montoya[2], and Jens Hummelshøj[2]

1 Kitware Inc. **2** Toyota Research Institute

Summary

Materials science research deals primarily with understanding the relationship between the structure and properties of materials. With recent advances in computational power and automation of simulation techniques, material structure and property databases have emerged (Curtarolo et al., 2012; Jain et al., 2013; Kirklin et al., 2015), allowing a more data-driven approach to carrying out materials research. Recent studies have demonstrated that representing these databases as material networks can enable extraction of new materials knowledge (Hegde, Aykol, Kirklin, & Wolverton, 2018; Isayev et al., 2015) or help tackle challenges like predictive synthesis (Aykol, Hegde, et al., 2019) that require relational information between materials. Materials databases have become very popular because they enable their users to do rapid prototyping by searching near globally for figures of merit for their target application. However, both scientists and engineers have little in the way of visualization of aggregates from these databases, that is, intuitive layouts that help understand which materials are related and how they are related. The need for a tool that does this is particularly crucial in materials science because properties like phase stability and crystal structure similarity are themselves functions of a material dataset, rather than of individual materials.

In these new approaches, materials can be represented with a graph structure that has *nodes* standing in for materials, and *links* between them encoding the appropriate relationships of interest, such as thermodynamic co-existence, chemical similarity or co-occurrence in text, to name a few. **MaterialNet** is an open-source web application designed to explore the topology of such material networks, while also displaying information about each material, highlighting its immediate neighborhood within the graph, and offering several auxiliary tools to help drill down into the details contained within the dataset. Such graph-encoded datasets for a collection of materials can become large and complex: for example, the materials stability network exemplified in this paper contains on the order of 20,000 materials, with on the order of 200,000 links between them (Aykol, Hegde, et al., 2019). Large graphs of this sort demand *interactive visualization*, empowering materials researchers to explore the data, a key user requirement highlighted by field experts (Aykol, Hummelshøj, et al., 2019). To the best of our knowledge, there exists no other interactive visualization tool for materials networks. MaterialNet provides interactive "maps" of the materials space exposed in large material databases, helping researchers navigate this space with a particular research task in mind, as showcased in Figure 1. A live deployment of MaterialNet can be accessed at http://maps.matr.io/.
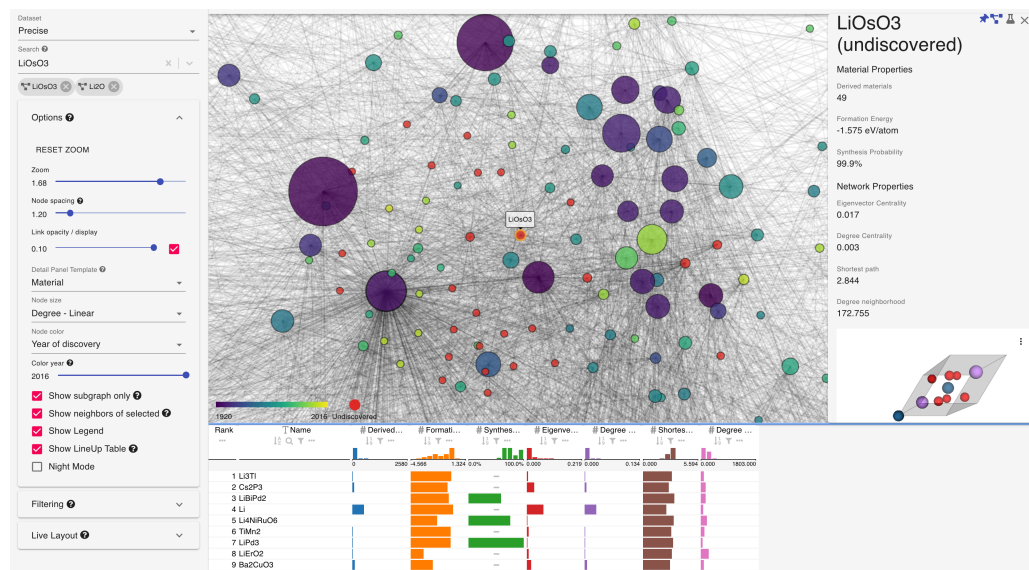
Figure 1: MaterialNet displaying the local network environment of a target material. The main material selected, LiOsO3, is hypothetical (i.e., a computational discovery) and is therefore labeled "undiscovered". A well-known material, Li2O, that exists in the vicinity of this target material is also added to the displayed subgraph. Node sizes represent the node degree (i.e. the number of links) for each material.

User Interface Features and Architecture

Interactive Node-Link Diagram

MaterialNet uses GeoJS (McKenzie, Chaudhary, Harris, & Beezley, 2014), an open-source mapping library developed by Kitware, to provide a highly interactive, attractive rendering of a node-link diagram representing the graph. Nodes in the graph are displayed with a color, encoding a data value carried on it (e.g., year of discovery, formation energy), and with a specific size, encoding some other data value (e.g., degree of connectivity within the graph). Layout software such as Gephi provides an initial configuration of the nodes in the graph, and MaterialNet has its own dynamic, force-based layout mode as well. The GeoJS node-link diagram is equipped with standard mouse gestures to enable interactive exploration: for example, by clicking on a node, that material's data appears in a panel on-screen, and its one-hop neighborhood is highlighted in the graph view. Such nodes may also be "pinned", leaving them visible in a list while other actions are performed. They can also be made part of a chemical subspace, which selects all materials containing one of several chemical elements. Additionally, MaterialNet can dynamically reconfigure the layout of these subgraphs, using a force-based simulation to bring related nodes closer together, providing a more organic sense of continuity and structure.

Searching, Filtering, and Chemical Subspaces

Effective, contextual search and filtering of materials (e.g., by using chemical formulas, constraining to chemical spaces, filtering based on material properties, historical material data) are essential for this visualization tool to be useful for the researchers. For this purpose, MaterialNet includes a simple search bar that can be used to find materials by their chemical makeup. Once a material of interest is found, it can be "pinned" to keep it in a special, persistent group of materials, allowing for exploration of its graph neighborhood, of the materials

spanned by its chemical subspace, or of materials found via fresh searches. MaterialNet allows filtering the available materials by setting ranges on data values such as formation energy. These methods enable the researcher to winnow out unrelated materials while building up a specific group of interest, allowing for a focus on the smaller subgraph spanned by those interesting materials only.

Chemical Detail Display

While interacting with the network, the user may often want to access a detailed view of data for specific materials. When focusing on a single material, MaterialNet displays its associated data in a table, along with an interactive 3D view of its crystal structure (using Kitware's VTK.js toolkit (Jourdain, Martin, & Schroeder, 2017)), and hyperlinks to external materials databases such as the OQMD.

Tabular Data View with LineUp

MaterialNet incorporates LineUp (Gratzl, Lex, Gehlenborg, Pfister, & Streit, 2013), an open-source tabular data visualization system developed in a collaboration between Kitware and Harvard University. LineUp displays all of the materials in the currently selected subgraph and allows interactive sorting of those materials using custom weighting functions. This tabulation system also allows the researcher to select materials directly from its table view, focusing them in the graph view. Combining the two visualization modes and linking them in this way enables researchers to use different modes of analysis to find what they are looking for. MaterialNet also includes prominent documentation activated by question mark icons situated throughout the interface. Any new user can quickly learn the system simply by trying to use it, reading these embedded help panels as they do so.

Future Work

In the future, we aim to explore several directions to increase the reach and value of MaterialNet:

- MaterialNet has been architected to support different types of data. The current deployment at http://maps.matr.io/ includes the materials stability network (Aykol, Hegde, et al., 2019), a text co-occurrence network extracted from MatScholar (Tshitoyan et al., 2019; Weston et al., 2019), and a materials similarity network (Ward et al., 2017), but the tool can be easily extended to display any other type of material network as well.
- A more powerful and flexible search mode, featuring a domain-specific language tailored to searching materials databases, will extend the researcher's ability to find materials with very specific properties or ranges of properties.
- New visualization modes (e.g., related to the behavior of materials in a chemical subspace) and integrations with existing visualization modes and tools will provide researchers with a choice of powerful ways to view the data, or subsets of it.

References

Aykol, M., Hegde, V. I., Hung, L., Suram, S., Herring, P., Wolverton, C., & Hummelshøj, J. S. (2019). Network analysis of synthesizable materials discovery. *Nature Communications*, *10*(1), 2018. doi:10.1038/s41467-019-10030-5

Aykol, M., Hummelshøj, J. S., Anapolsky, A., Aoyagi, K., Bazant, M. Z., Bligaard, T., Braatz, R. D., et al. (2019). The Materials Research Platform: Defining the requirements from user stories. *Matter*, *1*(6), 1433–1438. doi:10.1016/j.matt.2019.10.024

Curtarolo, S., Setyawan, W., Wang, S., Xue, J., Yang, K., Taylor, R. H., Nelson, L. J., et al. (2012). AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science*, *58*, 227–235. doi:10.1016/j.commatsci.2012.02.002

Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H., & Streit, M. (2013). LineUp: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '13)*, *19*(12). doi:10.1109/TVCG.2013.173

Hegde, V. I., Aykol, M., Kirklin, S., & Wolverton, C. (2018). The phase diagram of all inorganic materials. Retrieved from https://arxiv.org/abs/1808.10869

Isayev, O., Fourches, D., Muratov, E. N., Oses, C., Rasch, K., Tropsha, A., & Curtarolo, S. (2015). Materials Cartography: Representing and mining materials space using structural and electronic fingerprints. *Chemistry of Materials*, *27*(3). doi:10.1021/cm503507h

Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., et al. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, *1*(1). doi:10.1063/1.4812323

Jourdain, S., Martin, K., & Schroeder, W. (2017, October). vtk.js: The Visualization Toolkit on the web. https://blog.kitware.com/vtk-js-the-visualization-toolkit-on-the-web/.

Kirklin, S., Saal, J. E., Meredig, B., Thompson, A., Doak, J. W., Aykol, M., Rühl, S., et al. (2015). The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies. *npj Computational Materials*, *1*(1), 15010. doi:10.1038/npjcompumats.2015.10

McKenzie, S., Chaudhary, A., Harris, C., & Beezley, J. (2014, October). Introducing GeoJS. https://blog.kitware.com/introducing-geojs/.

Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K. A., et al. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*, *571*, 95–98. doi:10.1038/s41586-019-1335-8

Ward, L., Liu, R., Krishna, A., Hegde, V. I., Agrawal, A., Choudhary, A., & Wolverton, C. (2017). Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B*, *96*(2), 024104. doi:10.1103/PhysRevB.96.024104

Weston, L., Tshitoyan, V., Dagdelen, J., Kononova, O., Trewartha, A., Persson, K. A., Ceder, G., et al. (2019). Named entity recognition and normalization applied to large-scale information extraction from the materials science literature. *Journal of Chemical Information and Modeling*, *59*(9), 3692–3702. doi:10.1021/acs.jcim.9b00470