MATERIALS SCIENCE

The phase stability network of all inorganic materials

Vinay I. Hegde[1]*, Muratahan Aykol[2]*†, Scott Kirklin[1]‡, Chris Wolverton[1]†

One of the holy grails of materials science, unlocking structure-property relationships, has largely been pursued via bottom-up investigations of how the arrangement of atoms and interatomic bonding in a material determine its macroscopic behavior. Here, we consider a complementary approach, a top-down study of the organizational structure of networks of materials, based on the interaction between materials themselves. We unravel the complete "phase stability network of all inorganic materials" as a densely connected complex network of 21,000 thermodynamically stable compounds (nodes) interlinked by 41 million tie line (edges) defining their two-phase equilibria, as computed by high-throughput density functional theory. Analyzing the topology of this network of materials has the potential to uncover previously unidentified characteristics inaccessible from traditional atoms-to-materials paradigms. Using the connectivity of nodes in the phase stability network, we derive a rational, data-driven metric for material reactivity, the "nobility index," and quantitatively identify the noblest materials in nature.

INTRODUCTION

Several diverse complex systems are modeled as networks of discrete components linked together: man-made systems such as electrical power grids and the World Wide Web (1, 2), social systems such as friendship and scientific collaborations (3, 4), and natural systems such as metabolism in a cell and food webs (5, 6). Despite substantial variation in the nature of individual components and interconnections, many of these networks show notable similarities in their topology (7, 8), often providing new insights into each respective domain of knowledge. For instance, disparate systems such as the world wide web and metabolic reactions in cellular organisms both have been shown to follow the organizational principles of robust, error-tolerant scale-free networks, with implications for the resilience of the internet and the design of therapeutics (8, 9), respectively.

Recent developments in high-throughput density functional theory (HT-DFT) (10) have resulted in massive computational databases of materials properties (11–15), containing the calculated properties of hundreds of thousands of experimentally reported and hypothetical materials. Such databases have led to new data-driven approaches toward understanding materials. Here, we introduce a previously un-explored paradigm for viewing materials in general, and equilibrium phase diagrams in particular, using the lens of complex network theory. This approach uses the study of similarities and interactions between materials themselves, in notable contrast to the traditional bottom-up approaches toward unlocking structure-property relationships in materials (16, 17).

We use the Open Quantum Materials Database (OQMD) (11, 12), an HT-DFT database containing calculations of nearly all crystallographically ordered, structurally unique materials experimentally observed to date [as collected in the Inorganic Crystal Structure Database (18) repository] and a large number of hypothetical materials constructed using commonly occurring structural prototypes—a total of more than half a million materials—to extract the "universal phase stability network" or the "universal $T = 0$ K phase diagram". We accomplish this by using all the phase data in the OQMD within a convex-hull formalism, and identifying all thermodynamically stable materials and all two-phase equilibria between them. We then represent stable materials as nodes and two-phase equilibria (tie-lines) as edges, thus describing a $T = 0$ K phase diagram as a network encoding thermodynamic stability (illustrated with schematics in Fig. 1).

RESULTS

Overall network connectivity

We find that the phase stability network of all inorganic materials consists of ~21,300 nodes and is remarkably dense with a total of nearly 41 million edges, and extremely well connected with ~3850 edges per node on average ("mean degree" $\langle k \rangle$). This means that every stable inorganic compound can form a stable two-phase equilibrium with 3850 other compounds on average. For comparison, $\langle k \rangle$ for other widely studied networks range from 1.4 (network of email messages) to 113.4 (collaboration network of film actors) (19). The connectance of the materials network, or the fraction of the maximum possible number of edges that are actually present, is 0.18. This is an important statistic for the design of "systems of materials", such as electrodes and electrolytes making up batteries (20), or coating materials separating two reactive components (21), where the longevity of the system relies on stable coexistence of such components. Using a lithium-ion intercalation battery as an example "system of materials", a common approach to tackling electrode degradation is to apply protective coatings on electrode particles. In such a battery, the material in the electrode coating should not react with/be consumed by materials in the electrode as well as those in the electrolyte (22, 23). Thus, the coating-electrode and the coating-electrolyte material pairs must both have tie-lines with each other to stably coexist in the system. In other words, both pairs must be neighboring, connected nodes in the materials network.

The degree distribution in the complete phase stability network, specifically the probability $p(k)$ that a material has a tie-line with other k materials in the network, follows a lognormal form (Fig. 2A and fig. S1). While many widely studied networks are known to have scale-free power-law degree distributions, lognormal distributions are another member of the "heavy-tail" family, are also relatively common, and behave quite similar to power laws (24). Sparsity has been shown to be a necessary condition for the emergence of an exact power law behavior (25), and densification in sparse, scale-free
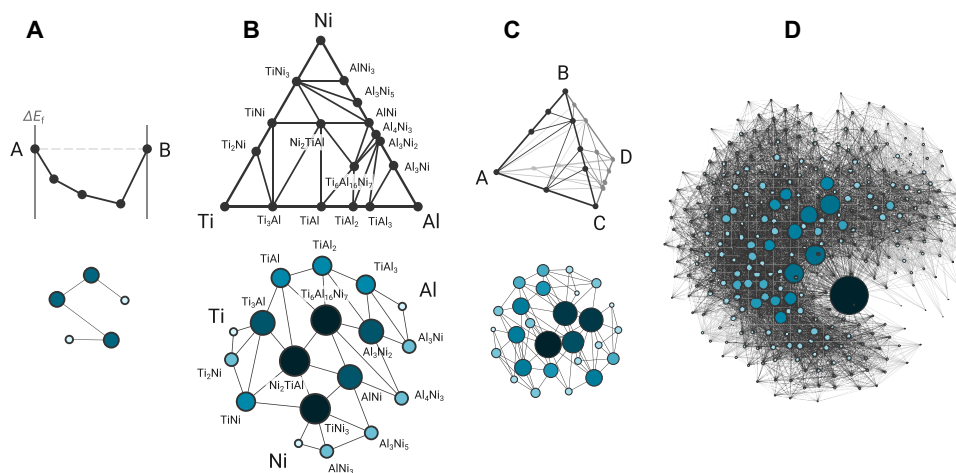
[1]Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208, USA. [2]Toyota Research Institute, Los Altos, CA 94022, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: c-wolverton@northwestern.edu (C.W.); murat.aykol@tri.global (M.A.)
‡Present address: Jump Trading LLC, Chicago, IL 60654, USA.

Fig. 1. Network representation of *T* = 0 K materials phase diagrams. Stable phases and two-phase equilibria (tie-lines) in a phase diagram are represented as nodes and edges, respectively, to create the corresponding network: (**A**) Schematic A-B binary system represented as a typical two-dimensional convex hull of compound formation energies. (**B**) Ti-Ni-Al as an example ternary system, with the *T* = 0 K phase diagram shown as a Gibbs triangle. (**C**) Schematic A-B-C-D quaternary phase diagram shown as a Gibbs tetrahedron. (**D**) The 3*d* transition metal-chalcogen (i.e., 14-dimensional chemical space) materials network. No conventional visual representations of phase diagrams exist at higher than four dimensions. Node sizes shown are proportional to node degree.
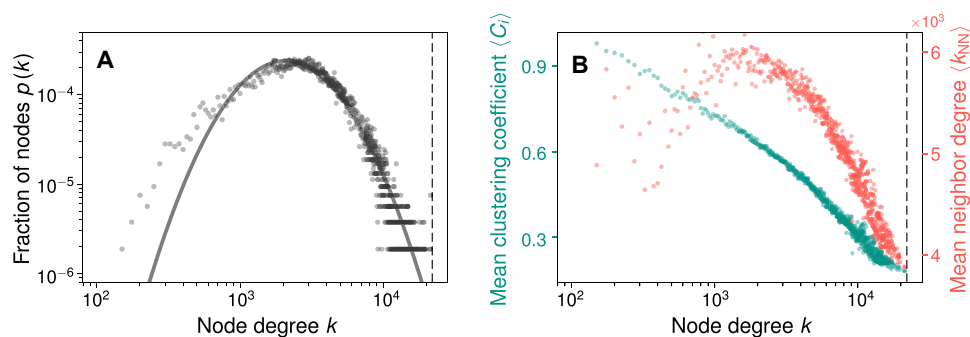
Fig. 2. Overall structure and topology of the materials network. (**A**) The distribution of node degree in the materials network (gray circles) shows a heavy tail; i.e., a sizeable fraction of materials have tie-lines with nearly all other materials. A lognormal fit is shown as a solid gray line. (**B**) The mean local clustering coefficient $\langle \mathcal{C}_i \rangle$ (green) decreases with node degree k, indicating that stable materials form local, highly connected communities. The mean neighbor degree $\langle k_{NN} \rangle$ (red) also decreases with k, implying a weakly dissortative network behavior; i.e., materials with a large number of tie-lines connect with those with fewer tie-lines in the network. In both subplots, the vertical dashed line represents the total number of nodes (stable materials) in the network.

networks leads to distributions that deviate from a power law and become closer to lognormal. Thus, the lognormal behavior of the materials network can be understood to result from its extremely dense connectivity, in contrast to the general sparsity of commonly studied networks.

Network topology

The characteristic path length or mean node-node distance in a network, $\mathcal{L}$, is defined as the number of edges in the shortest path between two nodes, averaged over all pairs of nodes. The longest node-node distance in the network defines its diameter, $\mathcal{L}_{max}$. The characteristic path length of the materials network $\mathcal{L} = 1.8$, and its diameter $\mathcal{L}_{max} = 2$. This remarkably short path length indicates that the materials network has "small-world" characteristics (*1*); i.e., despite its large size, the number of edges that need to be traversed from a given node to any other node is relatively small. The extremely small $\mathcal{L}$ for the materials network can be intuitively understood to be a consequence of the almost complete lack of reactivity of noble gases.

The nonparticipation of noble gases in the formation of compounds (and thus having tie-lines with nearly all materials in the network) places an upper bound of 2 on $\mathcal{L}_{max}$, and since some material pairs already have tie-lines that connect them immediately, the mean path $\mathcal{L}$ is slightly smaller than 2. Even if noble gases are disregarded, the mean path length and diameter of the materials network remain small because of the presence of a few other very highly connected nodes corresponding to extremely stable and nonreactive materials, e.g., binary halides.

Another metric of interest in a real-world network is transitivity or clustering, quantified by its clustering coefficient, $\mathcal{C}$, which is the probability that two nodes connected to the same third node are themselves connected. In other words, given that there exist stable two-phase equilibria A–C and B–C, what is the probability that A and B can stably coexist? Depending on how the averaging is performed, a global ($\mathcal{C}_g$) or mean local ($\bar{\mathcal{C}}_i$) cluster coefficient of a network can be defined (*1*, *19*). For the materials network, the clustering coefficients are $\mathcal{C}_g = 0.41$ and $\bar{\mathcal{C}}_i = 0.55$, comparable to other real-world networks,

and much higher than random networks of the same density. The mean local clustering coefficient of the materials network decreases with increasing node connectivity (Fig. 2B), indicating that stable materials form local highly connected communities in the network, and such behavior often suggests a hierarchical network structure (26). The assortativity coefficient or the Pearson correlation coefficient of degree between pairs of connected nodes in the materials network is −0.13, indicating weakly dissortative mixing behavior. This is also confirmed by the distribution of the mean degree of neighbors of a node of degree k being a decreasing function of k (Fig. 2A). In other words, materials with a high k (i.e., large number of tie-lines) tend to connect with materials with a lower k (i.e., smaller number of tie-lines). This weakly dissortative behavior of the materials network is similar to that observed in most other technological, information, and biological networks and is likely a virtue of such networks being simple graphs (27).

Hierarchy in the materials network

The mean degree or the average number of tie-lines per material $\langle k \rangle$ decreases with the number of components, $\mathcal{N}$ ($\mathcal{N} = 2$ for binary, $\mathcal{N} = 3$ for ternary, etc.; see Fig. 3A), indicating a chemical hierarchy in the materials network. This can be understood to result from an inherent competition for tie-lines that high-$\mathcal{N}$ materials face with low-$\mathcal{N}$ materials in their chemical space, but not vice versa. In other words, ternary compounds $X_a Y_b Z_c$ compete not only with other compounds in the X-Y-Z chemical space but also with binary compounds in the X-Y, Y-Z, and Z-X spaces for tie-lines.

We note that this decrease in $\langle k \rangle$ with $\mathcal{N}$ is distinct from the distribution of number of stable $\mathcal{N}$-ary materials itself (Fig. 3A), which shows a peak at $\mathcal{N} = 3$. Does this peak in the distribution of stable materials imply the existence of infinite, underexplored space for the discovery of previously unknown materials beyond ternaries? The distribution of formation energies of materials as a function of number of components $\mathcal{N}$ (Fig. 3B) reflects the consequence of competition between low- and high-component materials: high-$\mathcal{N}$ compounds appear to need substantially lower formation energies than low-$\mathcal{N}$ ones to become stable. Since there is no obvious underlying reason for the distribution of $T = 0$ K formation energies (with entropic effects neglected) to differ substantially with $\mathcal{N}$, only a few high-$\mathcal{N}$ materials can "survive" as stable phases if the corresponding

lower-$\mathcal{N}$ systems already have several stable phases. This is consistent with the recent reports of a "volcano plot" that emerges for stable inorganic ternary nitrides as a function of energetic competition with their corresponding binary nitrides (28), and an increased probability of phase separation with increasing number of components in a material system (29). Widom (30) further argued that the peak near $\mathcal{N} = 3$ or 4 in such distributions arises from a competition between combinatorial explosion and diminishing volume-to-surface ratio in the composition simplex, as $\mathcal{N}$ increases. Thus, although we do not know of a fundamental law limiting access to thermodynamically stable materials with higher components, a combination of the hierarchy observed in the phase stability network, the distribution of formation energies, and the topology of the convex energy surface all suggest that the scarcity of known high-$\mathcal{N}$ stable materials is not merely a consequence of those chemical spaces being underexplored.

Knowledge extraction: Material nobility index

Since the phase stability network practically encompasses all known inorganic crystalline materials as well as a large number of predicted hypothetical materials, the number of tie-lines emerges as a natural metric of nobility of a crystalline material—it is simply the count of other materials it is determined to have no reactivity against. Thus, while material reactivity or nobility has no standard definitions, a network representation of materials enables us to tackle the chemical nobility of inorganic materials in solid-solid and solid-gas reactions in a completely data-driven fashion, instead of the traditional intuitive or heuristic approaches. Since the number of tie-lines in the materials network is lognormally distributed, we devise a new standard score of material nobility, the "nobility index"

$$ \mathcal{Z}_n = \frac{ln\,(k) - \mu}{\sigma} \tag{1} $$

where k is the node degree or the number of tie-lines a material has and $\mu = 8.06$ and $\sigma = 0.65$ are the mean and standard deviation of the underlying lognormal distribution, respectively. The nobility index is thus agnostic of textbook classifications such as metal, nonmetal, metalloid, ionic, covalent, and so on and works equally well for any given material. Since the tie-lines in the network are as computed
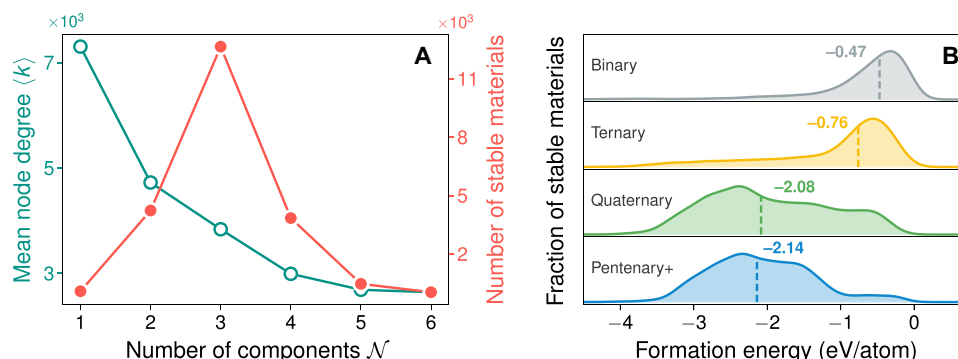
Fig. 3. Hierarchy in the materials network and underlying energetic considerations. (**A**) The mean node degree or average number of tie-lines $\langle k \rangle$ (green, open) decreases as a function of number of components $\mathcal{N}$ (i.e., binary, ternary, and so on), which results from high-$\mathcal{N}$ materials having to compete with low-$\mathcal{N}$ materials for stability. The number of known stable $\mathcal{N}$-ary materials (red) itself actually peaks at $\mathcal{N} = 3$ (ternaries). (**B**) Gaussian kernel density estimates of compound formation energies for all stable materials separated by number of components in the material. Dashed vertical lines indicate the respective median of each distribution. High-$\mathcal{N}$ materials need notably lower formation energies than low-$\mathcal{N}$ materials to become stable, e.g., −2.08 versus −0.47 eV per atom for quaternary and binary materials, respectively.

with DFT, the nobilities of materials predicted herewith are only limited by DFT accuracy in estimating relative stabilities of inorganic materials (*12*, *29*, *31*).

First, we tackle the reactivity or nobility of elements. Noble gases and fluorine form the bounds of the nobility index (Fig. 4), as the noblest and the most reactive, respectively, not only among the elements but in fact among all materials in the network. The most reactive elements following F are P, S, and Cl. Alkali and alkaline earth metals, often considered to be highly reactive metals, are relatively noble in solid-solid and solid-gas reactions, in comparison to early *d*-block or lanthanide elements, which are, along with Al, the most reactive metals. The nobility index increases down a group for metals and increases (decreases) from left to right along a row of the periodic table within the *d*-block (*s*-block). But what is the noblest metal of them all? Ag emerges as the noblest of all elements after noble gases, followed closely by Hg, Os, Re, W, and Cu, all having more than 14,000 tie-lines. Gold, traditionally considered the noblest element (*32*), despite being relatively densely connected with 10,000 tie-lines, is less noble in solid-state reactions. Last, we find that $\mathcal{Z}_n$ is not correlated with other common elemental properties such as electronegativity, atomic radii, melting point, and others (*33*), indicating that the nobility index encodes information not readily captured by those properties (fig. S2).

Beyond elements, what are the noblest inorganic compounds of all? The compounds at the top of the nobility list are IA/IIA-VIIA compounds such as LiF, NaCl, KCl, CsCl, KBr, CsBr, KI, RbI, CaF_2, SrF_2, $CsYbF_3$, $RbYbF_3$, and others, their inertness likely due to stability from strong ionic bonding between their constituents. We exclude rare earth– and actinide-containing compounds from the previous analysis of compound nobility to account for any shortcomings in the DFT description of *f*-block elements and compounds containing them.

DISCUSSION

While some of our findings above are in line with chemical intuition, relative nobilities in certain cases, e.g., silver versus gold, deviate from it. This deviation is, in part, due to the historical context in which these materials have been considered noble or reactive, e.g.,

whether an element oxidizes or corrodes readily in air, reacts with water and/or certain acids, and dissolves in water or electrolytes, and how vigorous such reactions seem. More fundamental approaches to finding descriptors for reactivity go back to electronegativity-related concepts, followed by interrelated theories based on perturbation theory, derivatives of electronic energy such as hardness and softness, and others largely developed for molecules (*34*–*37*). In contrast, the nobility index, $\mathcal{Z}_n$, as derived from the tie-lines in the network of all inorganic materials, represents a general metric emerging directly from bulk thermodynamic data.

High-throughput experimental and computational techniques are leading to an explosive growth in the size of materials databases. Representation and interpretation of the data at a large scale, however, remain a challenge. Here, we show that tools from complex network theory enable us to access otherwise difficult-to-extract information from such large datasets. In other words, the emergence of material reactivity from the collective behavior of all materials in the phase stability network serves as a simple, preliminary example of knowledge extraction out of complex networks of materials. Other similar approaches can be used to discover other hidden knowledge; e.g., analysis of "communities" or "cliques" in the network of all materials can uncover hitherto-unknown relationships between various known materials.

Further, there are various ways our graph theoretic approach to materials data can be used to be immediately applied to materials discovery and design: (i) Direct techniques, e.g., metrics from network theory such as local clustering and similarity, can be used to identify "holes" in the current network—where nodes (i.e., materials) are expected to exist but currently do not. (ii) Indirect techniques, e.g., using the extracted knowledge or quantities derived from the network as input to other approaches such as in materials informatics. For example, using temporal materials discovery information in combination with thermodynamic phase stability networks can help predict synthesizability (*38*). Furthermore, while some of its features resemble other complex networks, the extremely high connectance and the lognormal degree distribution of the presented phase stability network imply that its underlying generative mechanisms may be unique, and developing generative models for such materials networks can have substantial impact on the knowledge discovery of materials in the future.
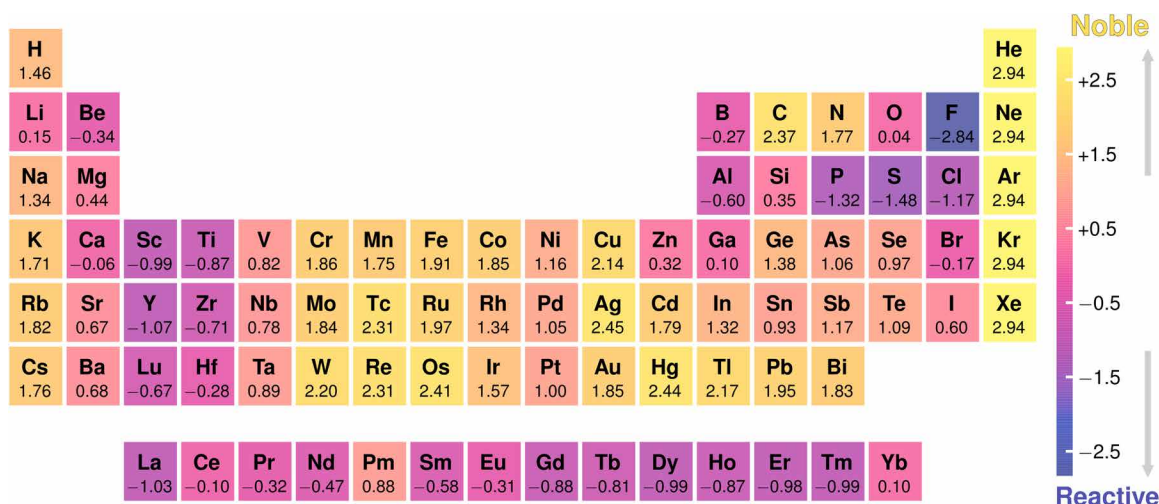
Fig. 4. Nobility index of all elements. The standard score, $\mathcal{Z}_n$, derived in this work using material connectivity in the phase stability network, as a measure of nobility against solid-solid and solid-gas reactions. Nobility increases up the scale. Numerical values of elemental $\mathcal{Z}_n$ are given below the respective symbols.

METHODS

All convex hull constructions were performed using the Qhull library (*39*) as implemented in the qmpy (pypi.org/project/qmpy) package. All network analyses were performed using the graph-tool (*40*) and powerlaw (*41*) packages, and comparison of heavy-tailed distributions was done according to the method of log likelihood ratios as described by Clauset *et al.* (*42*). Details of the divide-and-conquer approach used to tackle the combinatorial explosion in calculating the universal phase diagram, the related exponential increase in the time complexity to construct convex hulls in higher dimensions (*43*), its network representation, and determining the node degree distribution are provided in the Supplementary Materials.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/9/eaay5606/DC1

Section S1. Calculation of the $T = 0$ K universal phase diagram
Section S2. Degree distribution of the network of all materials
Section S3. New information encoded in the nobility index
Table S1. Sample compute times for calculating the existence of a tie-line between two phases.
Fig. S1. Fitting node connectivity data to candidate distributions.
Fig. S2. Comparison of nobility index versus common elemental properties.
Fig. S3. Comparison of number of compounds formed by an element versus its node degree.

REFERENCES AND NOTES

1. D. J. Watts, S. H. Strogatz, Collective dynamics of small-world networks. *Nature* **393**, 440–442 (1998).
2. R. Albert, H. Jeong, A.-L. Barabási, Diameter of the world-wide web. *Nature* **401**, 130–131 (1999).
3. D. J. Watts, P. S. Dodds, M. E. J. Newman, Identity and search in social networks. *Science* **296**, 1302–1305 (2002).
4. M. E. J. Newman, The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 404–409 (2001).
5. R. J. Williams, N. D. Martinez, Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
6. R. Guimerà, L. A. Nunes Amaral, Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
7. A.-L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
8. S. H. Strogatz, Exploring complex networks. *Nature* **410**, 268–276 (2001).
9. H. Jeong, R. Tombor, R. Albert, Z. N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
10. S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
11. J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials design and discovery with high-throughput density functional theory: The Open Quantum Materials Database (OQMD). *JOM* **65**, 1501–1509 (2013).
12. S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, C. Wolverton, The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **1**, 15010 (2015).
13. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
14. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, aflowlib.org: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
15. J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, A. Aspuru-Guzik, The Harvard clean energy project: Large-scale computational screening and design of organic photovoltaics on the world community grid. *J. Phys. Chem. Lett.* **2**, 2241–2251 (2011).
16. M. T. Dove, *Structure and Dynamics: An Atomic View of Materials* (Oxford Univ. Press, 2003).
17. R. Phillips, *Crystals, Defects and Microstructures: Modeling Across Scales* (Cambridge Univ. Press, 2001).
18. A. Belsky, M. Hellenbrandt, V. L. Karen, P. Luksch, New developments in the Inorganic Crystal Structure Database (ICSD): Accessibility in support of materials research and design. *Acta Cryst. B* **58**, 364–369 (2002).
19. M. E. J. Newman, The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
20. S. P. Ong, Y. Mo, W. D. Richards, L. Miara, H. S. Lee, G. Ceder, Phase stability, electrochemical stability and ionic conductivity of the $Li_{10\pm1}MP_2X_{12}$ (M = Ge, Si, Sn, Al or P, and X = O, S or Se) family of superionic conductors. *Energy Environ. Sci.* **6**, 148–156 (2013).
21. M. Aykol, S. Kim, V. I. Hegde, D. Snydacker, Z. Lu, S. Hao, S. Kirklin, D. Morgan, C. Wolverton, High-throughput computational design of cathode coatings for Li-ion batteries. *Nat. Commun.* **7**, 13779 (2016).
22. W. D. Richards, L. J. Miara, Y. Wang, J. C. Kim, G. Ceder, Interface stability in solid-state batteries. *Chem. Mater.* **28**, 266–273 (2016).
23. D. H. Snydacker, V. I. Hegde, C. Wolverton, Electrochemically stable coating materials for Li, Na, and Mg metal anodes in durable high energy batteries. *J. Electrochem. Soc.* **164**, A3582–A3589 (2017).
24. M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions. *Internet Math.* **1**, 226–251 (2003).
25. C. I. Del Genio, T. Gross, K. E. Bassler, All scale-free networks are sparse. *Phys. Rev. Lett.* **107**, 178701 (2011).
26. E. Ravasz, A.-L. Barabási, Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 026112 (2003).
27. S. Maslov, K. Sneppen, A. Zaliznyak, Detection of topological patterns in complex networks: Correlation profile of the internet. *Physica A* **333**, 529–540 (2004).
28. W. Sun, C. J. Bartel, E. Arca, S. R. Bauers, B. Matthews, B. Orvañanos, B.-R. Chen, M. F. Toney, L. T. Schelhas, W. Tumas, J. Tate, A. Zakutayev, S. Lany, A. M. Holder, G. Ceder, A map of the inorganic ternary metal nitrides. *Nat. Mater.* **18**, 732–739 (2019).
29. W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson, G. Ceder, The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, e1600225 (2016).
30. M. Widom, Frequency estimate for multicomponent crystalline compounds. *J. Stat. Phys.* **167**, 726–734 (2017).
31. M. Aykol, S. S. Dwaraknath, W. Sun, K. A. Persson, Thermodynamic limit for synthesis of metastable inorganic materials. *Sci. Adv.* **4**, eaaq0148 (2018).
32. B. Hammer, J. K. Norskov, Why gold is the noblest of all the metals. *Nature* **376**, 238–240 (1995).
33. L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, 16028 (2016).
34. R. G. Parr, L. Szentpály, S. Liu, Electrophilicity index. *J. Am. Chem. Soc.* **121**, 1922–1924 (1999).
35. G. Klopman, Chemical reactivity and the concept of charge- and frontier-controlled reactions. *J. Am. Chem. Soc.* **90**, 223–234 (1968).
36. W. Yang, R. G. Parr, Hardness, softness, and the fukui function in the electronic theory of metals and catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6723–6726 (1985).
37. H. Chermette, Chemical reactivity indexes in density functional theory. *J. Comput. Chem.* **20**, 129–154 (1999).
38. M. Aykol, V. I. Hegde, L. Hung, S. Suram, P. Herring, C. Wolverton, J. S. Hummelshøj, Network analysis of synthesizable materials discovery. *Nat. Commun.* **10**, 2018 (2019).
39. C. B. Barber, D. P. Dobkin, H. Huhdanpaa, The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **22**, 469–483 (1996).
40. T. P. Peixoto, The graph-tool python library. *figshare* (2014).
41. J. Alstott, E. Bullmore, D. Plenz, Powerlaw: A Python package for analysis of heavy-tailed distributions. *PLOS ONE* **9**, e85777 (2014).
42. A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009).
43. H. Sartipizadeh, T. L. Vincent, Computing the approximate convex hull in high dimensions. arXiv:1603.04422 [cs.CG] (8 March 2016).