Engineering 5 (2019) 397–405

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Deep Matter & Energy—Review

Data-Driven Discovery in Mineralogy: Recent Advances in Data Resources, Analysis, and Visualization

Robert M. Hazen [a,*], Robert T. Downs [b], Ahmed Eleish [c], Peter Fox [c], Olivier C. Gagné [a], Joshua J. Golden [b], Edward S. Grew [d], Daniel R. Hummer [e], Grethe Hystad [f], Sergey V. Krivovichev [g], Congrui Li [c], Chao Liu [a], Xiaogang Ma [h], Shaunna M. Morrison [a], Feifei Pan [c], Alexander J. Pires [b], Anirudh Prabhu [c], Jolyon Ralph [i], Simone E. Runyon [a,j], Hao Zhong [c]

[a] Geophysical Laboratory, Carnegie Institution for Science, Washington, DC 20015, USA
[b] Department of Geosciences, The University of Arizona, Tucson, AZ 85721-0077, USA
[c] Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
[d] School of Earth and Climate Sciences, University of Maine, Orono, ME 04469, USA
[e] Department of Geology, Southern Illinois University, Carbondale, IL 62901, USA
[f] Mathematics, Statistics, and Computer Science, Purdue University Northwest, Hammond, IN 46323-2094, USA
[g] Kola Science Centre of the Russian Academy of Sciences, Apatity, Murmansk Region 184209, Russia
[h] Department of Computer Science, University of Idaho, Moscow, ID 83844-1010, USA
[i] Mindat.org, Mitcham CR4 4FD, UK
[j] Department of Geology and Geophysics, University of Wyoming, Laramie, WY 82071-2000, USA

ARTICLE INFO

Article history:
Received 15 November 2018
Revised 18 February 2019
Accepted 13 March 2019
Available online 2 May 2019

Keywords:
Mineral evolution
Mineral ecology
Skyline diagrams
Network analysis
Cluster analysis
Chord diagrams
Klee diagrams

ABSTRACT

Large and growing data resources on the diversity, distribution, and properties of minerals are ushering in a new era of data-driven discovery in mineralogy. The most comprehensive international mineral database is the IMA database, which includes information on more than 5400 approved mineral species and their properties, and the mindat.org data source, which contains more than 1 million species/locality data on minerals found at more than 300 000 localities. Analysis and visualization of these data with diverse techniques—including chord diagrams, cluster diagrams, Klee diagrams, skyline diagrams, and varied methods of network analysis—are leading to a greater understanding of the co-evolving geosphere and biosphere. New data-driven approaches include mineral evolution, mineral ecology, and mineral network analysis—methods that collectively consider the distribution and diversity of minerals through space and time. These strategies are fostering a deeper understanding of mineral co-occurrences and, for the first time, facilitating predictions of mineral species that occur on Earth but have yet to be discovered and described.

1. Introduction

The discovery, description, and development of Earth's mineral wealth have long been central pursuits of the Earth sciences. For much of that history, the discoveries of new mineral resources and novel mineral species have been based as much on chance finds as on empirical guidelines. The old adage, "Gold is where you find it," has applied to most natural resources, but data-driven discovery is now changing that mantra. In this contribution, we review the nature of large and growing mineralogical data resources and describe some of the analytical and visualization methods that are being applied to understand the diversity and distribution of minerals in space and time.

Recent studies fall under three broad headings. *Mineral evolution* is the investigation of Earth's changing near-surface mineralogy over 4.5 billion years of history—studies that reveal the striking co-evolution of the geosphere and biosphere and the increasing diversity and complexity of mineral species driven by the chemical differentiation of Earth [1–27]. *Mineral ecology*, a complementary pursuit, investigates the diversity and spatial distribution of Earth's minerals, including consideration of the

* Corresponding author.
E-mail address: rhazen@ciw.edu (R.M. Hazen).

https://doi.org/10.1016/j.eng.2019.03.006
2095-8099/

unusual distribution of rare minerals on Earth [28–39]. Finally, *mineral network analysis* provides a powerful means to analyze and visualize the complex distributions of minerals and their properties through space and time [40]. Taken together, these approaches have the potential to change our view of the evolving mineralogy of Earth and other terrestrial worlds.

2. Mineral data resources

Data-driven discovery relies on comprehensive and reliable tabulations of mineral species, their properties, and their distributions in space and time. The official list of mineral species approved by the International Mineralogical Association (IMA) is documented by the IMA database,[†] which is maintained at the Department of Geosciences, The University of Arizona [41]. In addition to recording more than 5400 mineral species, the RRUFF data resource compiles data on crystal structures, compositions, Raman spectra, and other physical properties. Mineral evolution studies require data on mineral ages, localities, and context—data that is compiled at the Mineral Evolution Database.[‡] More than 185 000 individual locality/age data for minerals are available through this rapidly expanding, open-access resource.

The largest data resource on the global distribution of minerals is mindat.org,[††] an international, crowd-sourced effort led by Jolyon Ralph and the Hudson Institute of Mineralogy. The mindat.org data source has recorded more than 1.1 million mineral/locality data from approximately 300 000 localities worldwide—data that are essential in the analysis and visualization of mineral diversity and distribution relationships.

The essential resources of the IMA database and the mindat.org data source are amplified by a number of other data compilations, most notably the petrological and geochemical resources under the umbrella of the Interdisciplinary Earth Data Alliance (IEDA[‡‡]), including EarthChem[†††] (e.g., Ref. [42]).

An ongoing challenge in developing these critical data resources is the vast amount of "dark data"—that is, information on mineral compositions, localities, and other data that is available only through hard-copy publications, proprietary corporate documents (notably companies in the natural resources industry), or privately held research records. Data-driven discovery cannot reach its full potential until a culture of data sharing is fully embraced by the Earth science community, with the implementation of "FAIR" (i.e., findable, accessible, interoperable, and reusable) data practice [43].

Given the rich and growing open-access mineralogical data resources, opportunities for applying a range of powerful analytical and visualization methods beckon [44,45]. In this article, we review a few of these methods as they relate to the fields of mineral evolution, mineral ecology, and mineral network analysis.

3. Mineral evolution

Mineral evolution is the study of the changing near-surface mineralogy of Earth and other terrestrial worlds through deep time [5,19]. Our detailed understanding of Earth's 4.5-billion-year history of mineralogical change, coupled with a growing understanding of the mineralogy of other solar system bodies [46,47], reveals that a planet's mineralogy evolves through a sequence of stages, each the result of new physical, chemical, and (in the case of Earth) biological modes of mineral paragenesis.

The greater than 185 000 individual locality/age for minerals tabulated in the mineral evolution database, though far short of recording all available mineral/age information, is sufficiently extensive to reveal striking patterns in Earth's evolving mineralogy. Three first-order trends stand out.

The first trend in the temporal distribution of minerals is a marked episodicity that reflects the supercontinent cycle of the past 3 billion years [8,12]. We find that Earth has preserved pulses of mineralization during five purported episodes of the convergence and assembly of sometime isolated landmasses into single supercontinents (Fig. 1) [39]. The convergence of continents and consequent orogenic events not only induce mineralization; these mineralizing events are also more likely to be preserved in the cores of the resulting mountain ranges. More detailed investigation of these trends reveals additional subtleties, for example in the unique tectonic and geochemical setting of the assembly of Rodinia at ~1.3 to 0.9 Ga [27].

The second significant temporal trend in Earth's evolving mineralogy is an observed increase in the average oxidation state of transition metals [20,48]. Thus, for example, the minerals of manganese display a systematic increase in redox state over the past 500 million years, with other fluctuations occurring earlier in Earth's history (Fig. 2). Similar trends have been observed for all of the redox-sensitive, first-row transition metals (Fig. 3[‡‡‡]), as well as for uranium [6] and rhenium [20].

The third trend in the evolution of the mineral world is its increasing structural and chemical complexity with the flow of geological time (Fig. 4) [5,11,26]. Numerical estimates of complexity using information-based measures have facilitated the analysis of quantitative correlations between chemical and structural complexities of minerals for a total of 4962 datasets on the chemical compositions and 3989 datasets on the crystal structures of minerals [23,26]. This analysis demonstrates that there is an overall trend of increasing structural complexity with increasing chemical complexity. Moreover, analysis of mean chemical and structural complexities for mineral groups occurring in different geological periods [5,15] has demonstrated that both are gradually increasing in the course of mineral evolution. By analogy with biological evolution [49], the increasing mineral complexity follows an overall

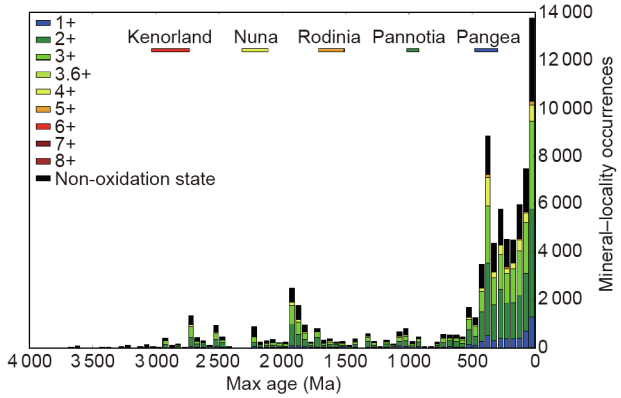

Fig. 1. First-row transition metal mineral–locality occurrences by max age (minerals listed once with highest oxidation state from any first-row transistion elements in formula). Our record of Earth's minerals through time typically reveals pulses of mineralization that are associated with the supercontinent cycle. In this graph of approximately 60 000 mineral/age data for minerals incorporating first-row transition metals, pulses of mineralization are associated with the supercontinents Kenorland, Nuna, Rodinia, Pannotia, and Pangea. Note that mineralization associated with Rodinian assembly at ~1.3–0.9 Ga is less distinct than the peaks with other supercontinents, as a consequence of its unique tectonic setting [39]. 1+–8+ refer to different oxidation states.

[†] https://rruff.info/ima/.
[‡] https://rruff.info/evolution.
[††] https://www.mindat.org.
[‡‡] https://www.iedadata.org/.
[†††] https://www.EarthChem.org.

[‡‡‡] See https://dtdi.carnegiescience.edu for an animated version.

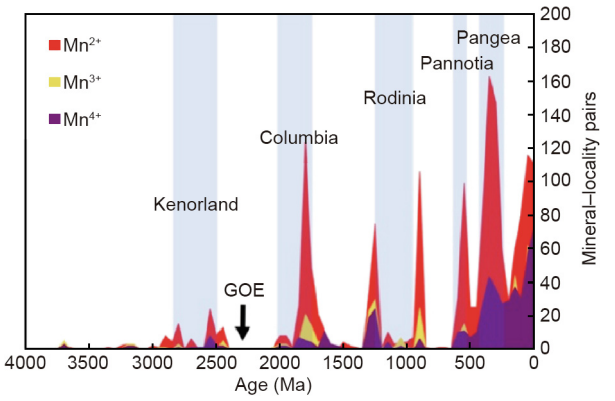

Fig. 2. Changes in Earth's near-surface oxidation state, the consequence of the evolution of oxygenic photosynthesis, are reflected in the changing ratios of manganese in the II (Mn^{2+}), III (Mn^{3+}), and IV (Mn^{4+}) oxidation states. The average oxidation state of manganese increases, most notably during the past 500 million years. GOE: Great Oxidation Event.

passive trend: More complex minerals form with the passage of geological time, yet the simpler ones are not replaced (see also Ref. [35]). The observed correlations suggest that, at a first approximation, chemical differentiation is a major force driving the increase of complexity of minerals throughout Earth's history. New levels of complexity and diversification observed in mineral evolution are achieved through local concentrations of particular rare elements and the creation of new geochemical environments.

4. Mineral ecology

Mineral ecology considers the diversity and spatial distribution of minerals, in much the same way as studies of biological ecosystems document distributions of living species. Earth's minerals are distributed according to a "large number of rare events" (LNRE) frequency spectrum, which is common to both biological ecosystems and the distribution of words in a book [29,31,37]. In each instance, a few species or words are extremely common, but most species or words are rare.

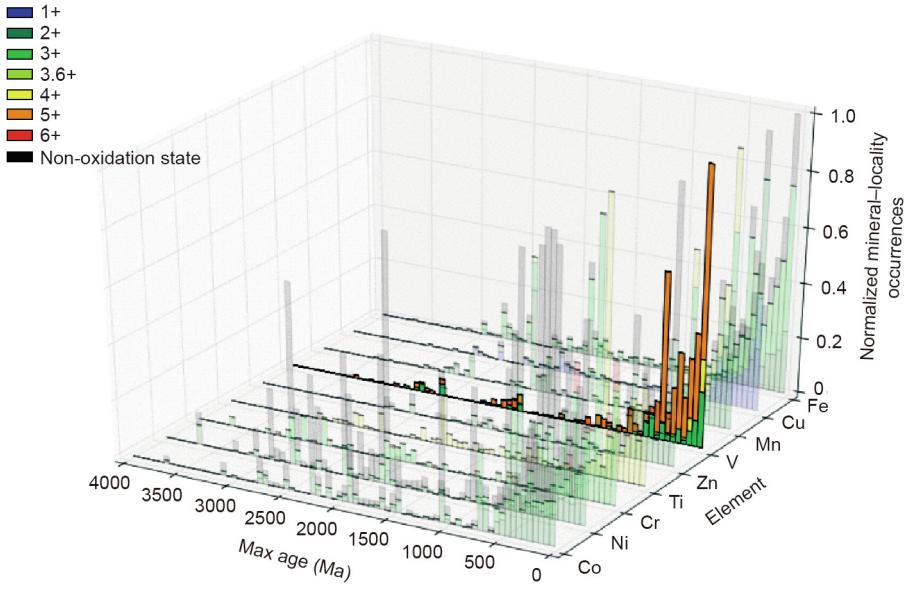

Fig. 3. Normalized mineral–locality occurrences by max age for different elements. A "skyline diagram" of minerals containing first-row transition elements reveals systematic trends associated with the supercontinent cycle and Earth's changing atmospheric composition.

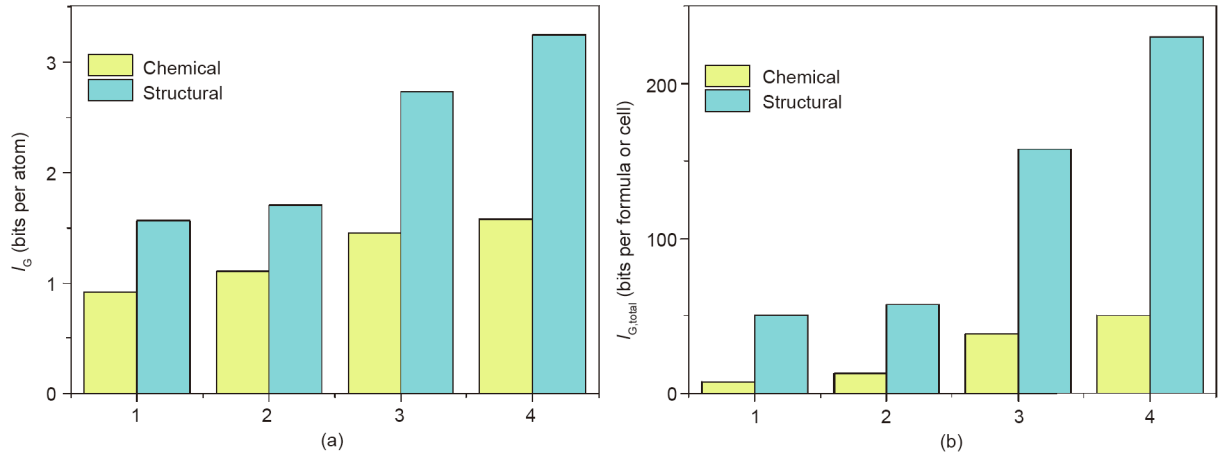

Fig. 4. Mean chemical and structural information-based complexities for minerals occurring in different eras of mineral evolution (1 = 12 "ur-minerals" [5]; 2 = 60 minerals of chondritic meteorites [5]; 3 = 420 minerals of the Hadean epoch [11]; 4 = all minerals of the post-Hadean era) calculated for a total of 4962 datasets on the chemical compositions and 3989 datasets on the crystal structures of minerals [26]. (a) Shannon information per atom (I_G); (b) Shannon information per unit cell or formula unit ($I_{G,total}$).

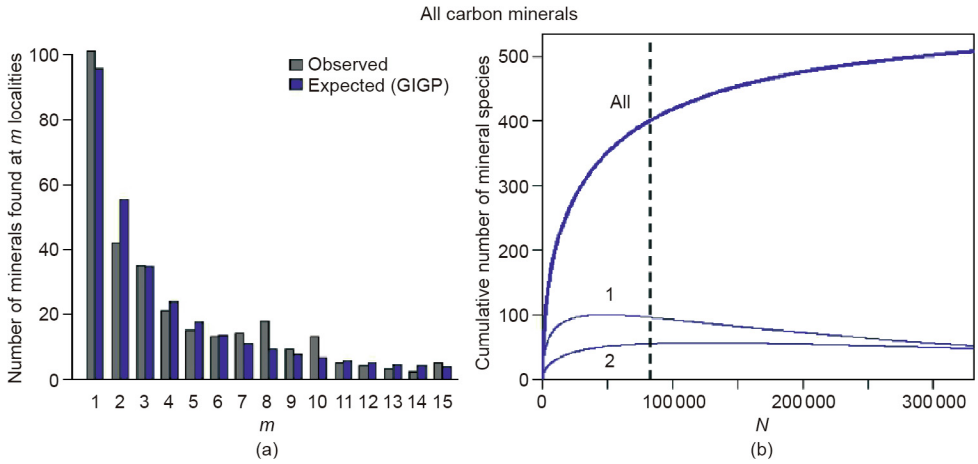

Fig. 5. (a) The frequency spectrum for carbon-bearing minerals reveals that most minerals are rare. The horizontal axis records the exact number of localities (m) at which a carbon-bearing mineral species is found. The vertical axis indicates how many mineral species occur at exactly that number of localities. Grey bars are the observed values, while blue bars indicate the modeled values. Of the 403 documented carbon-bearing minerals in 2016, more than 100 are known from only one locality, while 40 have been described from exactly two localities. (b) This "large number of rare events" distribution facilitates calculation of an accumulation curve (upper blue curve), shown here on a graph of the number of observed mineral/locality data (N, X axis) versus the estimated number of different mineral species (Y axis). Extrapolation of this curve to the right suggests that an additional 145 carbon-bearing minerals await discovery and description [33]. The vertical dashed line indicates the number of mineral/locality data (82 922) and known species (403) as of 2016. Curves 1 and 2 represent the evolving numbers of different mineral species identified from exactly one or two localities, respectively—values that change systematically as more mineral/locality data accumulate. Note that these curves go through a maximum value; the number of minerals known from only one locality is now declining as more mineral/locality data are reported.

Our detailed understanding of distributions of common and rare mineral species is made possible by the mineral/locality data in mindat.org. These data facilitate the calculation of "accumulation curves," which reveal estimates of the numbers of "missing" minerals—those types that occur on Earth but have yet to be discovered and described [28,32]. For example, in a detailed study of the more than 400 carbon-bearing minerals, Hazen et al. [33] predicted that an additional ~145 carbon-bearing minerals await discovery (Fig. 5) [33]. In addition, they listed several hundred candidates for these missing minerals, noting that most would be hydrous carbonates, with a special emphasis on calcium- and sodium-bearing phases that may have been overlooked because they are relatively nondescript—typically white or grey in color and poorly crystallized [32]. This work inspired the Carbon Mineral Challenge,[†] an international project supported by the Deep Carbon Observatory[‡] to find as many of the missing carbon-bearing minerals as possible. As of 20 May 2019, at least 30 new carbon-bearing species had been discovered, described, and approved by the IMA.

5. Mineral co-occurrence and network analysis

One of the most important challenges of mineralogy is to understand the diversity and distribution of minerals in the context of coexisting assemblages of minerals—a problem that requires considering hundreds of species simultaneously. The large and growing mindat.org data resource, coupled with a variety of analytical and visualization methods, is revolutionizing our ability to document these complex multidimensional systems.

5.1. Chord diagrams

The first step in any analysis of mineral coexistence is to construct a data object with each mineral species as a separate field. In the simple case of a pairwise mineral co-occurrence matrix, each matrix element represents the number of times that two minerals occur together. These data can be represented by a variety of techniques. Chord diagrams array a group of related mineral species as arcs of a circle, with curved lines connecting coexisting species (Fig. 6). Widely employed in gene analysis, such chord diagrams can also prove useful in mineralogy by illustrating numerous pairwise occurrences in a single visual representation. Chord diagrams can be explored in interactive displays, with embedded metadata on numbers of occurrences, as well as details on localities and other coexisting species.

5.2. Klee diagrams

Klee diagrams (sometimes referred to as "heat maps"; Fig. 7) also represent the frequency with which pairs of objects—such as minerals or their essential chemical elements—coexist, and thus are a complementary visualization tool to the chord diagram

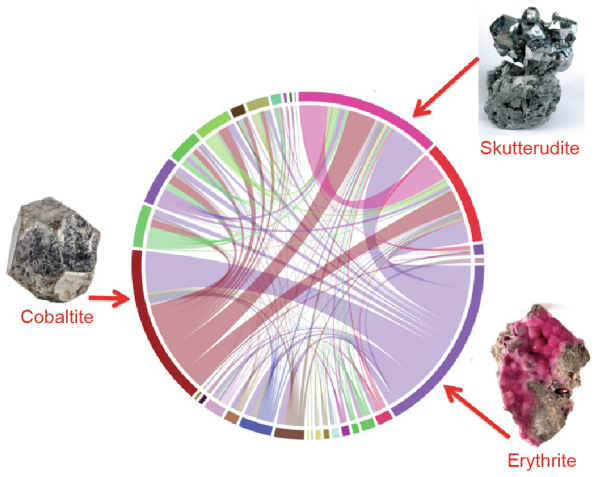

Fig. 6. A chord diagram of the 43 most common cobalt-bearing minerals reveals coexisting pairs of minerals. This rendering reveals that the secondary mineral erythrite ($Co_3(AsO_4)_2{\cdot}8H_2O$) is the most abundant cobalt mineral, and that it is most commonly associated with the two most common primary cobalt ore minerals, cobaltite (CoAsS) and skutterudite ($CoAs_{3-x}$).

[†] https://mineralchallenge.net/.
[‡] https://deepcarbon.net/.

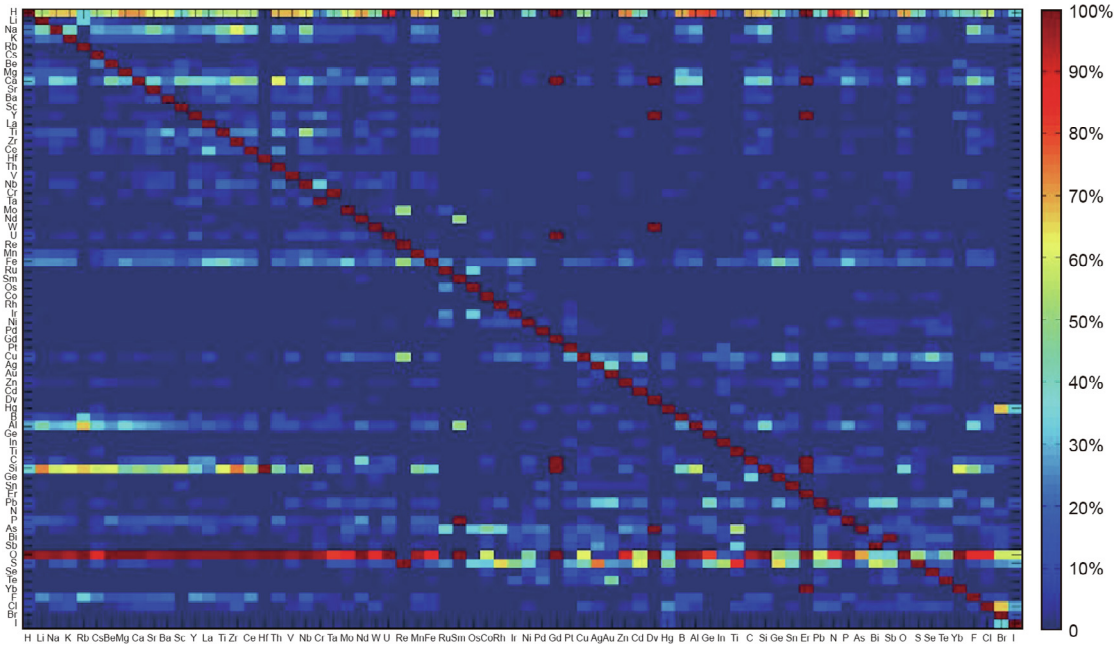

Fig. 7. Klee diagrams (sometimes referred to as "heat maps") represent the frequency with which pairs of minerals, elements, or other objects coexist. This rendering displays a 72 × 72 matrix of coexisting chemical elements in minerals, in which each matrix element represents the fraction of minerals with element *X* that also incorporates element *Y*. This matrix is not symmetrical; for example, all minerals containing beryllium also incorporate oxygen, but only a small fraction of oxygen-bearing minerals incorporate beryllium.

shown in Fig. 6. This method facilitates rapid analysis of coexisting pairs of minerals or elements; however, it is often desirable to understand the associations of more than two objects at a time. Accordingly, Ma et al. [50] have explored the use of interactive three-dimensional Klee diagrams to understand coexisting elements in minerals (Fig. 8). In spite of their potential for quickly revealing occurrence trends among thousands of mineral pairs, Klee diagrams have not yet been widely applied to mineral coexistence relationships.

5.3. Network analysis

Network analysis is an especially useful tool for exploring complex interrelationships among numerous mineral species [40]. The use of network graphs to elucidate connections in the contexts of social groups [51–54], technological networks [55–58], and biological systems [59–62] are well known. Each network consists of vertices (or nodes), some of which are connected to each other by edges (or links). Distances between nodes, and hence the length of links, are determined by the degree of association of the two nodes; shortest distances represent the strongest links. Vertices and edges can be sized, shaped, and colored to indicate additional attributes of the system.

Networks of coexisting minerals provide vivid examples of network graphs.[†] In Fig. 9 [40], individual nodes represent mineral species. The nodes are sized to represent the relative number of localities of each species, while node colors can represent compositional, structural, paragenetic, or other information. These highly interactive visual displays represent projections from multidimensional space into two- or three-dimensional space, in order to show the connections from each mineral node to all other co-occurring mineral nodes. In general, for a well-connected network of *N* different mineral species, the rendering is a projection from *N* – 1 dimensions. In many instances, a three-dimensional rendering provides important additional information, even though the projection may be from much higher dimensions.

Network graphs not only represent local properties—such as all of a given mineral's coexisting species—but they also reveal global trends not easily discerned from the data alone, such as clustering by chemistry or paragenetic mode, the degree of a network's interconnectedness, and otherwise hidden compositional and temporal trends. A distinct advantage of network statistical analysis is the opportunity for network metrics that characterize global and local statistical properties of the network [63,64]. Metrics, including density, centrality, and diameter, facilitate the comparison of related networks, such as those representing minerals incorporating different chemical elements or a time series for minerals of a given element [40].

5.4. Bipartite network graphs

Several rendering options exist for network graphs. Of special importance to mineralogy are bipartite graphs [65], which can display two distinct types of nodes, such as representing both mineral species and their localities (Fig. 10). A striking feature of mineral bipartite networks—one not reported from such graphs of other natural or artificial systems—is the distribution of locality nodes in a "U-shape" (or "vase shape" in three dimensions), with fewer very common minerals inside the U (or vase), and many more rare minerals decorating the periphery (Fig. 10). This distribution is a visual representation of an LNRE spectrum, with relatively few very common minerals and numerous rare species.

6. The future

Data-driven discovery in mineralogy is still in its infancy. Open-access mineral data resources need to be expanded by at least an order of magnitude, with a special effort made to recover dark data that will otherwise be lost. New analytical and visualization methods, some tailored specifically to mineralogy, must be created and

[†] See https://dtdi.carnegiescience.edu for interactive examples.

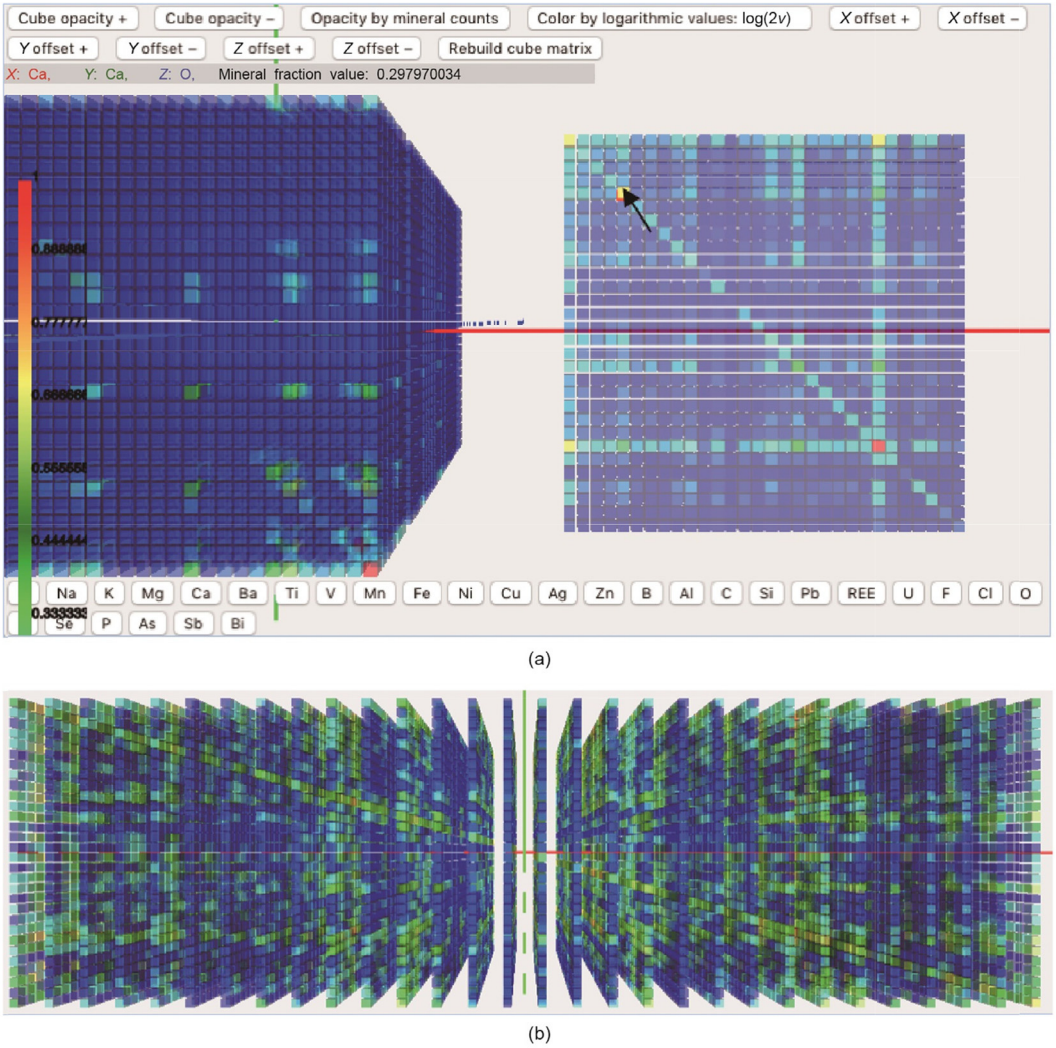

Fig. 8. A three-dimensional interactive Klee diagram facilitates the exploration of triplets of coexisting minerals or elements. This example from Ref. [50] records the frequency of co-occurrence of triplets of chemical elements in minerals. (a) The cube-shaped rendering is difficult to interpret, but any planar slice of the cube can be viewed independently; (b) alternatively, the cube can be rendered in an "exploded" version to allow users to see the "inside" of the cube. The red line indicates the centerline of the 3D diagram. The arrow points to one of many "hot spots," in this case Ca + Ca + O, where the combination of elements is more commonly found in minerals than would be predicted based on crustal abundances. REE: rare earth elements.

implemented. In addition, opportunities will emerge to apply these methods to other terrestrial planets and moons, as data from Mars, the moon, and other worlds are gathered.

A critical need is to merge a variety of databases with a deep-time component and correlate their various data fields. Efforts are underway to correlate mineralogical databases to other deep-time databases, such as geochemical, paleontological, and protein databases, in order to gain a more holistic picture of the co-evolution of Earth's geosphere and biosphere.[†] These studies have the potential to reveal how Earth's changing near-surface mineralogy and geochemistry have influenced the biochemistry of organisms, and how life, in turn, created new mineral species and geochemical niches.

6.1. Affinity analysis

Perhaps the most exciting prospect is the targeted discovery of new mineral occurrences, including new economically valuable resources, employing the methods of affinity analysis. A taste of what is to come was provided by a recent prediction by Jolyon Ralph of mindat.org (personal communications, May 2018). Using only pairwise mineral correlations, Ralph predicted that the uncommon mineral wulfenite ($PbMoO_4$) should occur at Cookes Peak, New Mexico, a lead–zinc–silver mining district with more than 75 reported mineral species but lacking reports of wulfenite. Subsequent scrutiny by local mineral collectors revealed pockets of this beautiful, but previously overlooked, mineral.

Affinity analysis (also known as "market-basket analysis" when applied to product recommendations by online companies), employs a similar approach but with multidimensional positive and negative co-occurrence information [66–68]. Initial trials of affinity analysis to minerals will expand search algorithms beyond pairwise coexistence data to larger combinations of characteristic mineral species. In the near future, we hope to interrogate mindat.org to compile lists of "missing" minerals with their probabilities of occurrence at known localities—a testable approach to the development of predictive mineralogy.

An aspiration of our program is to search for mineral and other natural resources by applying affinity analysis to expansive data

[†] For example, https://dtdi.carnegiescience.edu.

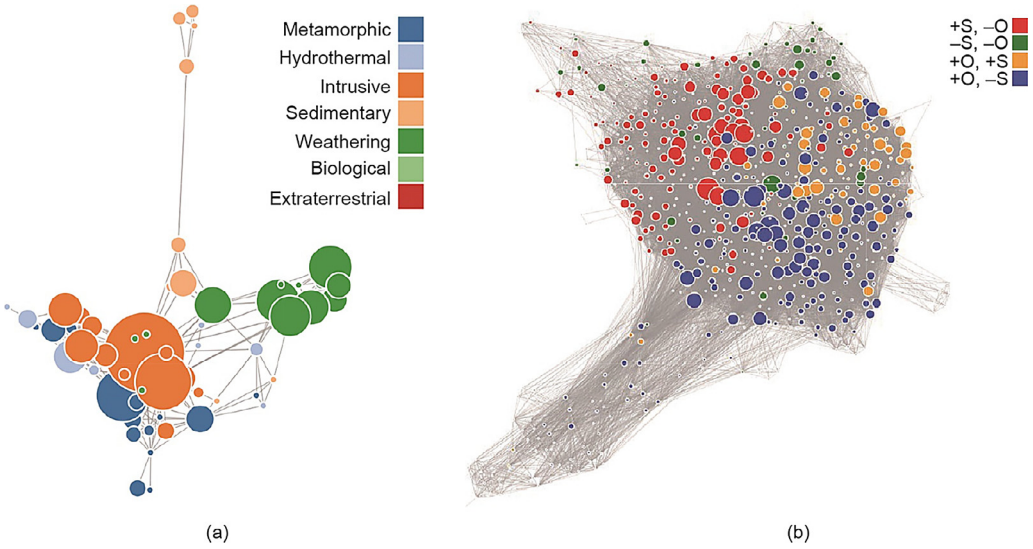

Fig. 9. Network graphs of mineral species. (a) 58 chromium-bearing minerals: nodes are sized according to mineral frequency of occurrence, and colored according to mode of formation (see inset). This low-density network shows strong clustering based on paragenetic mode. (b) 664 copper-bearing minerals: nodes are sized according to mineral frequency of occurrence; nodes are colored according to the presence or absence of S or O (see inset; after Ref. [40]).

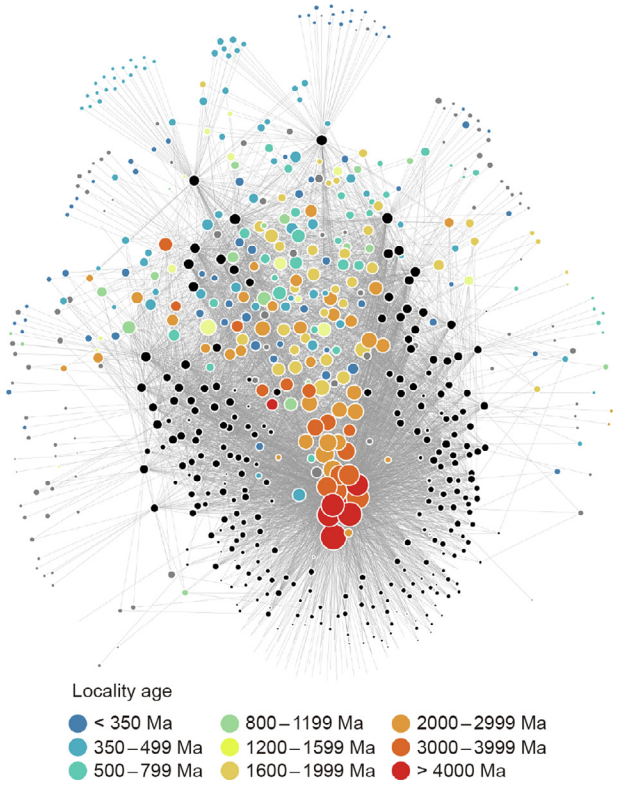

Fig. 10. Bipartite network of 403 carbon-bearing mineral species. Colored circles represent carbon mineral species, with circle sizes representing relative frequency of occurrence and colors (see inset scale) corresponding to the age of earliest known occurrences of those minerals. Black circles represent regional localities, with sizes corresponding to the relative numbers of different carbon-bearing minerals found at those localities. The network rendering reveals important information regarding the diversity and distribution of carbon minerals through space and time. In particular, the "U-shaped" distribution of black locality nodes, with a few very common carbon minerals "inside" and many more rare carbon minerals "outside," is an alternative visual representation of the LNRE distribution illustrated in Fig. 5. Note that most of the common minerals are more ancient, whereas most of the rare minerals are more recent. See also http://dtdi.carnegiescience.edu/node/4557 for an interactive version.

resources that include numerous fields related to mineral occurrences, their chemical compositions (including trace-element and isotopic data), and physical properties, as well as the physical, chemical, and biological environmental context of those mineralized occurrences. We anticipate that recommender systems will play a key role in the next generation of natural resource exploration.

6.2. Crystal chemical systematics

Data-driven efforts by Gagné and Hawthorne [69–72] and Gagné [73] have recently provided a baseline statistical knowledge of the bonding behavior of atoms in oxide, oxysalt, and nitride crystals. This congregated knowledge, soon to be expanded to sulfide and sulfosalt minerals, allows prediction of the most likely composition of "missing minerals" in a much more precise way by combining knowledge of the ideal bond valences of a crystal structure [74] and the ability of the ions to adopt predicted bonding requirements. This influx of organized bonding data further allows the derivation of high-quality bond-valence parameters (e.g., Ref. [75]), which are useful in the context of mineral evolution to better infer the oxidation state of redox-sensitive transition metals in studying Earth's changing near-surface environments.

7. Conclusions

Data-driven discovery in mineralogy represents one aspect of the dynamic "open data movement" that has the potential to change the pace of scientific discovery [76,77]. Progress will depend on parallel advances in building comprehensive data resources, developing and implementing advanced analytical and visualization methods, and applying these capabilities to outstanding mineralogical problems. In some cases, these data science methods will accelerate hypothesis-driven science by enhancing our understanding of the diversity and distribution of minerals—findings that we know we don't know. Even more exciting is the prospect that multidimensional analysis of mineral systems will lead to the abductive discovery of new and unexpected insights—discoveries of what we didn't know we didn't know.

Acknowledgements

We are grateful to Ho-Kwang Mao and the organizers of this special issue for the opportunity to share our results. This publication is a contribution to the Deep Carbon Observatory.

Studies of mineral evolution and mineral ecology are supported by grants from the Alfred P. Sloan Foundation (G-2016-7065), the W. M. Keck Foundation (grant entitled "Co-Evolution of the Geosphere and Biosphere"), the John Templeton Foundation (60645), the NASA Astrobiology Institute (1-NAI8_2-0007), a private foundation, and the Carnegie Institution for Science. Sergey V. Krivovichev acknowledges support from the Russian Science Foundation (19-17-00038).

Compliance with ethics guidelines

Robert M. Hazen, Robert T. Downs, Ahmed Eleish, Peter Fox, Olivier C. Gagné, Joshua J. Golden, Edward S. Grew, Daniel R. Hummer, Grethe Hystad, Sergey V. Krivovichev, Congrui Li, Chao Liu, Xiaogang Ma, Shaunna M. Morrison, Feifei Pan, Alexander J. Pires, Anirudh Prabhu, Jolyon Ralph, Simone E. Runyon, and Hao Zhong declare that they have no conflict of interest or financial conflicts to disclose.

References

[1] Gastil G. The distribution of mineral dates in time and space. Am J Sci 1960;258(1):1–35.
[2] Nash JT, Granger HC, Adams SS. Geology and concepts of genesis of important types of uranium deposits. Econ Geol 1981:63–116.
[3] Zhabin AG. Is there evolution of mineral speciation on Earth? Dokl Earth Sci Sect 1981;247:142–4.
[4] Yushkin NP. Evolutionary ideas in modern mineralogy. Zap Vses Mineral Obshch 1982;116(4):432–42. Russian.
[5] Hazen RM, Papineau D, Bleeker W, Downs RT, Ferry J, McCoy T, et al. Mineral evolution. Am Mineral 2008;93(11–12):1693–720.
[6] Hazen RM, Ewing RJ, Sverjensky DA. Evolution of uranium and thorium minerals. Am Mineral 2009;94(10):1293–311.
[7] Hazen RM, Bekker A, Bish DL, Bleeker W, Downs RT, Farquhar J, et al. Needs and opportunities in mineral evolution research. Am Mineral 2011;96(7):953–63.
[8] Hazen RM, Golden JJ, Downs RT, Hysted G, Grew ES, Azzolini D, et al. Mercury (Hg) mineral evolution: a mineralogical record of supercontinent assembly, changing ocean geochemistry, and the emerging terrestrial biosphere. Am Mineral 2012;97(7):1013–42.
[9] Hazen RM, Papineau D. Mineralogical co-evolution of the geosphere and biosphere. In: Knoll AH, Canfield DE, Konhauser KO, editors. Fundamentals of geobiology. Oxford: Wiley-Blackwell; 2012. p. 333–50.
[10] Hazen RM, Jones AP, Kah L, Sverjensky DA. Carbon mineral evolution. In: Hazen RM, Jones AP, Baross J, editors. Carbon in Earth. Washington, DC: Mineralogical Society of America; 2013. p. 79–107.
[11] Hazen RM, Sverjensky DA, Azzolini D, Bish DL, Elmore S, Hinnov L, et al. Clay mineral evolution. Am Mineral 2013;98(11–12):2007–29.
[12] Hazen RM, Liu XM, Downs RT, Golden JJ, Pires AJ, Grew ES, et al. Mineral evolution: episodic metallogenesis, the supercontinent cycle, and the coevolving geosphere and biosphere. Soc Econ Geolog Special Pub 2014;18:1–15.
[13] Hazen RM, Grew ES, Origlieri M, Downs RT. On the mineralogy of the "Anthropocene Epoch". Am Mineral 2017;102(3):595–611.
[14] Hazen RM. Evolution of minerals. Sci Am 2010;302(3):58–65.
[15] Hazen RM. Paleomineralogy of the Hadean Eon: a preliminary list. Am J Sci 2013;313(9):807–43.
[16] Hazen RM. Mineral evolution, the Great Oxidation Event, and the rise of colorful minerals. Mineralog Record 2015;46(805–816):34.
[17] Hazen RM. An evolutionary system of mineralogy: proposal for a classification based on natural kind clustering. Am Mineral. In press.
[18] Hazen RM, Eldredge N. Themes and variations in complex systems. Elements 2010;6(1):43–6.
[19] Hazen RM, Ferry JM. Mineral evolution: mineralogy in the fourth dimension. Elements 2010;6(1):9–12.
[20] Golden J, McMillan M, Downs RT, Hystad G, Stein HJ, Zimmerman A, et al. Rhenium variations in molybdenite (MoS_2): evidence for progressive subsurface oxidation. Earth Planet Sci Lett 2013;366:1–5.
[21] Grew ES, Hazen RM. Evolution of the minerals of beryllium. Stein 2013:4–19.
[22] Grew ES, Hazen RM. Beryllium mineral evolution. Am Mineral 2014;99(5–6):999–1021.
[23] Krivovichev SV. Structural complexity of minerals: information storage and processing in the mineral world. Mineral Mag 2013;77(3):275–326.
[24] Krivovichev SV. Structural complexity of minerals and mineral parageneses: information and its evolution in the mineral world. In: Armbruster T, Danisi RM, editors. Highlights in mineralogical crystallography. Berlin/Boston: de Gruyter; 2015. p. 31–74.
[25] Grew ES, Dymek RF, De Hoog JCM, Harley SL, Boak JM, Hazen RM, et al. Boron isotopes in tourmaline from the 3.7–3.8 Ga Isua Belt, Greenland: sources for boron in Eoarchean continental crust and seawater. Geochim Cosmochim Acta 2015;163:156–77.
[26] Krivovichev SV, Krivovichev VG, Hazen RM. Structural and chemical complexity of minerals: correlations and time evolution. Eur J Mineral 2018;30(2):231–6.
[27] Liu C, Knoll AH, Hazen RM. Geochemical and mineralogical evidence that Rodinian assembly was unique. Nat Commun 2017;8(1):1950.
[28] Hystad G, Downs RT, Hazen RM. Mineral species frequency distribution conforms to a large number of rare events model: prediction of Earth's missing minerals. Math Geosci 2015;47(6):647–61.
[29] Hystad G, Downs RT, Grew ES, Hazen RM. Statistical analysis of mineral diversity and distribution: Earth's mineralogy is unique. Earth Planet Sci Lett 2015;426:154–7.
[30] Hystad G, Downs RT, Hazen RM, Golden JJ. Relative abundances for the mineral species on Earth: a statistical measure to characterize Earth-like planets based on Earth's mineralogy. Math Geosci 2017;49(2):179–94.
[31] Hazen RM, Grew ES, Downs RT, Golden J, Hystad G. Mineral ecology: chance and necessity in the mineral diversity of terrestrial planets. Can Mineral 2015;53(2):295–323.
[32] Hazen RM, Hystad G, Downs RT, Golden J, Pires A, Grew ES. Earth's "missing" minerals. Am Mineral 2015;100(10):2344–7.
[33] Hazen RM, Hummer DR, Hystad G, Downs RT, Golden JJ. Carbon mineral ecology: predicting the undiscovered minerals of carbon. Am Mineral 2016;101(4):889–906.
[34] Hazen RM, Hystad G, Golden JJ, Hummer DR, Liu C, Downs RT, et al. Cobalt mineral ecology. Am Mineral 2017;102(1):108–16.
[35] Grew ES, Krivovichev SV, Hazen RM, Hystad G. Evolution of structural complexity in boron minerals. Can Mineral 2016;54(1):125–43.
[36] Grew ES, Hystad G, Hazen RM, Krivovichev SV, Gorelova LA. How many boron minerals occur in Earth's upper crust? Am Mineral 2017;102(8):1573–87.
[37] Hazen RM, Ausubel J. On the nature and significance of rarity in mineralogy. Am Mineral 2016;101(6):1245–51.
[38] Liu C, Hystad G, Golden JJ, Hummer DR, Downs RT, Morrison SM, et al. Chromium mineral ecology. Am Mineral 2017;102(3):612–9.
[39] Liu C, Eleish A, Hystad G, Golden JJ, Downs RT, Morrison SM, et al. Analysis and visualization of vanadium mineral diversity and distribution. Am Mineral 2018;103(7):1080–6.
[40] Morrison SM, Liu C, Eleish A, Prabhu A, Li C, Ralph J, et al. Network analysis of mineralogical systems. Am Mineral 2017;102(8):1588–96.
[41] Downs RT. The RRUFF project: an integrated study of the chemistry, crystallography, Raman and infrared spectroscopy of minerals. In: Proceedings of the 19th General Meeting of the International Mineralogical Association; 2006 July 23–28; Kobe, Japan; 2006.
[42] Lehnert KA, Walker D, Sarbas B. EarthChem: a geochemistry data network. Geochim Cosmochim Acta 2007;71:A559.
[43] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018.
[44] Fox P, Hendler J. Changing the equation on scientific data visualization. Science 2011;331(6018):705–8.
[45] Hazen RM. Data-driven abductive discovery in mineralogy. Am Mineral 2014;99(11–12):2165–70.
[46] Papike JJ, editor. Planetary materials. Chantilly: Mineralogical Society of America; 1998.
[47] Morrison SM, Downs RT, Blake DF, Vaniman DT, Ming DW, Rampe EB, et al. Crystal chemistry of martian minerals from Bradbury Landing through Naukluft Plateau, Gale crater, Mars. Am Mineral 2018;103(6):857–71.
[48] Liu XM, Kah LC, Knoll AH, Cui H, Kaufman AJ, Shahar A, et al. Tracing Earth's CO_2 evolution using Zn/Fe ratios in marine carbonate. Geochem Perspect Lett 2016;2:24–34.
[49] Carroll SB. Chance and necessity: the evolution of morphological complexity and diversity. Nature 2001;409(6823):1102–9.
[50] Ma X, Hummer D, Golden JJ, Fox PA, Hazen RM, Morrison SM, et al. Using visualized exploratory data analysis to facilitate collaboration and hypothesis generation in cross-disciplinary research. ISPRS Int J Geoinf 2017;6(11):368.
[51] Otte E, Rousseau R. Social network analysis: a powerful strategy, also for the information sciences. J Inf Sci 2002;28(6):441–53.
[52] Abraham A, Hassanien AE, Snasel V, editors. Computational social network analysis: trends, tools and research advances. New York: Springer; 2010.
[53] Pinheiro CAR. Social network analysis in telecommunications. Hoboken: Wiley; 2011.
[54] Kadushin C. Understanding social networks. New York: Oxford University Press; 2012.
[55] Hwang N, Houghtalen R. Fundamentals of hydraulic engineering systems. Upper Saddle River: Prentice Hall; 1996.
[56] Guimerà R, Mossa S, Turtschi A, Amaral LAN. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. Proc Natl Acad Sci USA 2005;102(22):7794–9.

[57] Dong W, Pentland A. A network analysis of road traffic with vehicle tracking data. In: Proceedings of the American Association of Artificial Intelligence, Spring Symposium, Human Behavior Modeling; 2009 Mar 23–25; Palo Alto, CA, USA; 2009. p. 7–12.
[58] Pagani GA, Aiello M. The power grid as a complex network: a survey. Phys A 2013;392(11):2688–700.
[59] Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, et al. Network analysis of protein structures identifies functional residues. J Mol Biol 2004;344(4):1135–46.
[60] Banda-R K, Delgado-Salinas A, Dexter KG, Linares-Palomino R, Oliveira-Filho A, Prado D, et al. Plant diversity patterns in neotropical dry forests and their conservation implications. Science 2016;353(6306):1383–7.
[61] Corel E, Lopez P, Méheust R, Bapteste E. Network-thinking: graphs to analyze microbial complexity and evolution. Trends Microbiol 2016;24(3):224–37.
[62] Muscente AD, Prabhu A, Zhong H, Eleish A, Meyer MB, Fox P, et al. Quantifying ecological impacts of mass extinctions with network analysis of fossil communities. Proc Natl Acad Sci USA 2018;115(20):5217–22.
[63] Kolaczyk ED. Statistical analysis of network data. New York: Springer; 2009.
[64] Newman MEJ. Networks: an introduction. New York: Oxford University Press; 2013.
[65] Asratian AS, Denley TMJ, Häggkvist R. Bipartite graphs and their applications. New York: Cambridge University Press; 1998.
[66] Adomavicius G, Tuzhilin A. Context-aware recommender systems. In: Ricci F, Rokach L, Shapira B, Kantor PB, editors. Recommender systems handbook. Boston: Springer; 2011. p. 217–53.
[67] Ricci F, Rokach L, Shapira B. Introduction to recommender systems handbook. In: Ricci F, Rokach L, Shapira B, Kantor PB, editors. Recommender systems handbook. Boston: Springer; 2011. p. 1–35.
[68] Panniello U, Tuzhilin A, Gorgoglione M. Comparing context-aware recommender systems in terms of accuracy and diversity. User Model User-adapt Interact 2014;24(1–2):35–65.
[69] Gagné OC, Hawthorne FC. Bond-length distributions for ions bonded to oxygen: alkali and alkaline-earth metals. Acta Crystallogr B Struct Sci Cryst Eng Mater 2016;72(Pt 4):602–25.
[70] Gagné OC, Hawthorne FC. Bond-length distributions for ions bonded to oxygen: results for the non-metals and discussion of lone-pair stereoactivity and the polymerization of PO_4. Acta Crystallogr B 2018;74:79–96.
[71] Gagné OC, Hawthorne FC. Bond-length distributions for ions bonded to oxygen: metalloids and post-transition metals. Acta Crystallogr B 2018;74:63–78.
[72] Gagné OC, Hawthorne FC. Bond-length distributions for ions bonded to oxygen: results for the transition metals and discussion of d^0 cations and the Jahn-Teller effect. Acta Cryst B 2018;74(Pt 1):79–96.
[73] Gagné OC. Bond-length distributions for ions bonded to oxygen: results for the lanthanides and actinides and discussion of the f-block contraction. Acta Crystallogr B 2018;74:49–62.
[74] Gagné OC, Mercier PHJ, Hawthorne FC. *A priori* bond-valence and bond-length calculations in rock-forming minerals. Acta Crystallogr B 2018;74:470–82.
[75] Gagné OC, Hawthorne FC. Comprehensive derivation of bond-valence parameters for ion pairs involving oxygen. Acta Crystallogr B Struct Sci Cryst Eng Mater 2015;71(Pt 5):562–78.
[76] Schutt R, O'Neil C. Doing data science: straight talk from the frontline. New York: O'Reilly; 2013.
[77] Kitchin R. The data revolution: big data, open data, data infrastructures & their consequences. London: Sage; 2014.