Submitted to *American Mineralogist*, 07 February 2017, for consideration as an "Outlooks" contribution.

# AM6104: Revision #1 (April 6, 2017)
# Network analysis of mineralogical systems

SHAUNNA M. MORRISON[1], CHAO LIU[1], AHMED ELEISH[1,2], ANIRUDH PRABHU[2], CONGRUI LI[2], JOLYON RALPH[3], ROBERT T. DOWNS[4], JOSHUA J. GOLDEN[4], PETER FOX[2], DANIEL R. HUMMER[1,5], MICHAEL B. MEYER[1], AND ROBERT M. HAZEN[1,*]

[1]Geophysical Laboratory, Carnegie Institution for Science, 5251 Broad Branch Road NW, Washington, D.C. 20015, U. S. A.

[2]Tetherless World Constellation, Department of Earth and Environmental Sciences, Rensselaer Polytechnic Institute, Troy, New York 12180, U. S. A.

[3]Mindat.org, 128 Mullards Close, Mitcham, Surrey, CR4 4FD, United Kingdom

[4]Department of Geosciences, University of Arizona, 1040 East 4th Street, Tucson, Arizona 85721-0077, U. S. A.

[5]Department of Geology, Southern Illinois University, Carbondale, Illinois 62901, U. S. A.

## ABSTRACT

A fundamental goal of mineralogy and petrology is the deep understanding of mineral phase relationships and the consequent spatial and temporal patterns of mineral coexistence in rocks, ore bodies, sediments, meteorites, and other natural polycrystalline materials. The multi-dimensional chemical complexity of such mineral assemblages has traditionally led to experimental and theoretical consideration of 2-, 3-, or *n*-component

systems that represent simplified approximations of natural systems. Network analysis provides a dynamic, quantitative, and predictive visualization framework for employing "big data" to explore complex and otherwise hidden higher-dimensional patterns of diversity and distribution in such mineral systems. We introduce and explore applications of mineral network analysis, in which mineral species are represented by nodes, while coexistence of minerals is indicated by lines between nodes. This approach provides a dynamic visualization platform for higher-dimensional analysis of phase relationships, because topologies of equilibrium phase assemblages and pathways of mineral reaction series are embedded within the networks. Mineral networks also facilitate quantitative comparison of lithologies from different planets and moons, the analysis of coexistence patterns simultaneously among hundreds of mineral species and their localities, the exploration of varied paragenetic modes of mineral groups, and investigation of changing patterns of mineral occurrence through deep time. Mineral network analysis, furthermore, represents an effective visual approach to teaching and learning in mineralogy and petrology.

**Keywords:** network analysis, igneous petrology, mineral evolution, copper, chromium, phase equilibria, Bowen's reaction series, visualization, big data, cluster analysis, multi-dimensional scaling, education

* E-mail: rhazen@ciw.edu

## INTRODUCTION

Network analysis encompasses a powerful array of mathematical and visualization methods that have found numerous applications in the presentation and interpretation of "big data" in varied fields of technology and science (Kolaczyk 2009; Newman 2013). Technological networks include the physical infrastructures of power grids (Pagani and Aiello 2013), roads (Dong and Pentland 2009), and water supply systems (Hwang and Houghtalen 1996; Geem 2010), as well as communications infrastructure (Pinheiro 2011), commercial distribution networks (Guimerá et al. 2005), and the Internet and other information networks (Otte and Rousseau 2002). In the familiar realm of social interactions, networks are used to quantify and visualize data in such diverse topics as the spread of disease, the links among Facebook "friends," the structure of terrorist organizations, and connections among research collaborators (Otte and Rousseau 2002; Abraham et al. 2010; Scott and Carrington 2011; Kadushin 2012). Network analysis has been applied in biology to the study of ecosystem diversity (Banda et al. 2016), food webs (Martinez 1992; Dunne et al. 2008), neural networks (Müller et al. 1995), biochemical pathways (Costanzo et al. 2016), proteomics and protein-protein interactions (Amital et al. 2004; Harel et al. 2015; Ueza et al. 2016; Leuenberger et al. 2017), paleogeography (Sidor et al. 2013; Dunhill et al. 2016; Huang et al. 2016), and evolution (Vilhena et al. 2013; Cheng et al. 2014; Corel et al. 2016). In each of these network applications and more, the modeling, graphing, and analysis of data reveals previously unrecognized patterns and behaviors in complex systems.

Qualitative network-like representations of minerals have been presented previously (e.g., Christy et al. 2016). However, in spite of its utility and widespread application,

quantitative network analysis does not appear to have been applied to mineralogical problems. Here we introduce and apply network analysis to topics in mineralogy and petrology—fields that are especially amenable to this approach because they consider systems of numerous mineral species that co-exist in myriad combinations in varied deposits. In particular, we demonstrate that network analysis of equilibrium mineral assemblages has the potential to elucidate phase relationships in complex multi-dimensional composition space, while revealing previously hidden trends in spatial and temporal aspects of mineral diversity and distribution.

In this contribution we consider varied network representations of three contrasting mineral systems: (1) common rock-forming minerals in intrusive igneous rocks; (2) terrestrial minerals containing the element chromium; and (3) minerals containing the element copper. These subsets of the more than 5200 mineral species approved by the International Mineralogical Association's Commission on New Minerals and Mineral Names (IMA-CNMMN) exemplify the potential of network analysis to address fundamental questions in mineralogy and petrology.

## EXAMPLES OF MINERAL NETWORKS

Minerals, whether in rocks, sediments, meteorites, or ore deposits, exist as assemblages of coexisting species. Here we introduce mineral networks as a strategy to represent and analyze the large and growing data resources related to these assemblages with a variety of mathematical and graphical models—network "renderings" that are available through open access sources. In each case mineral networks employ nodes (also known as vertices), each corresponding to a mineral species. Some node pairs are

connected by links (also known as edges), which indicate that those two minerals are found together at the same location or deposit. Variations in the ways that nodes and links are represented highlight different aspects of network relationships, as illustrated in the following examples.

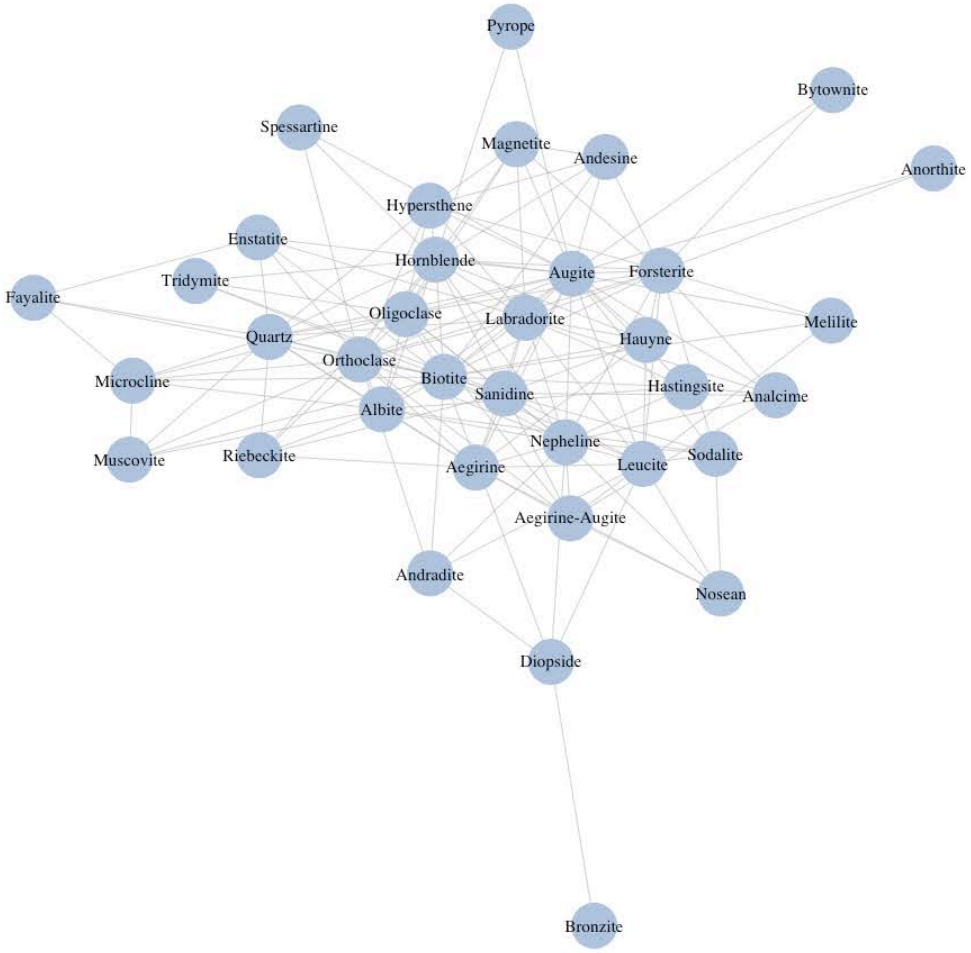
*Fruchterman-Reingold Force-Directed Networks:* Figure 1A illustrates a simplified Fruchterman-Reingold force-directed network (Fruchterman and Reingold 1991; Csardi and Nepusz 2006), representing 36 major rock-forming minerals that occur in holocrystalline intrusive igneous rocks, as described in Alfred Harker's classic *Petrology for Students* (Harker 1964). Mineralogical descriptions of 77 igneous rocks, each with 1 to 6 major minerals (see Supplemental Information 1), provide the input data for this visualization.

The Fruchterman–Reingold force-directed graph algorithm is based on two main principles: (1) vertices connected by an edge should be drawn near each other and (2) vertices generally should not be drawn too close to each other. These criteria resemble those of molecular or planetary simulations where bodies exert both attractive and repulsive forces on one another. This method attempts to balance the energy of the system through iterative displacement of the vertices by calculating the effect of attractive forces on each vertex, then calculating the effect of repulsive forces, and finally limiting the total displacement with a temperature parameter. In this rendering, we have no control over the length of the edges; edge length is determined by the final positions of vertices as the system reaches equilibrium, however, highly connected groups of nodes will tend to form clusters.

In Figure 1, we created a simplified Fruchterman-Reingold force directed network using the igraph package in R. We imported tabulated data on coexisting rock-forming minerals into R as a data frame, which was then converted into a matrix object to enable visualization using the igraph package. The igraph software enables a high level of customization based on different network metrics. If "auto.layout" is used, then the package finds the best-suited algorithm based on the nodes and the number of links between the nodes. After some preliminary analysis, we found the best-suited algorithm to be the Fruchterman-Reingold force-directed network with self-loops removed.

Note that many of the mineral names employed by Harker do not correspond to approved IMA-CNMMN species. In some instances, such as "biotite," "hornblende," and "tourmaline," the names once commonly employed by optical petrologists have been replaced by several related species (i.e., annite, fluorannite, siderophyllite, and tetraferriannite for "biotite"). In the case of plagioclase feldspar, on the other hand, Harker distinguishes six compositional variants—albite, oligoclase, andesine, labradorite, bytownite, and anorthite—as opposed to the two end-member species albite and anorthite recognized as valid species by the IMA-CNMMN.

133 **A**




134

**B**

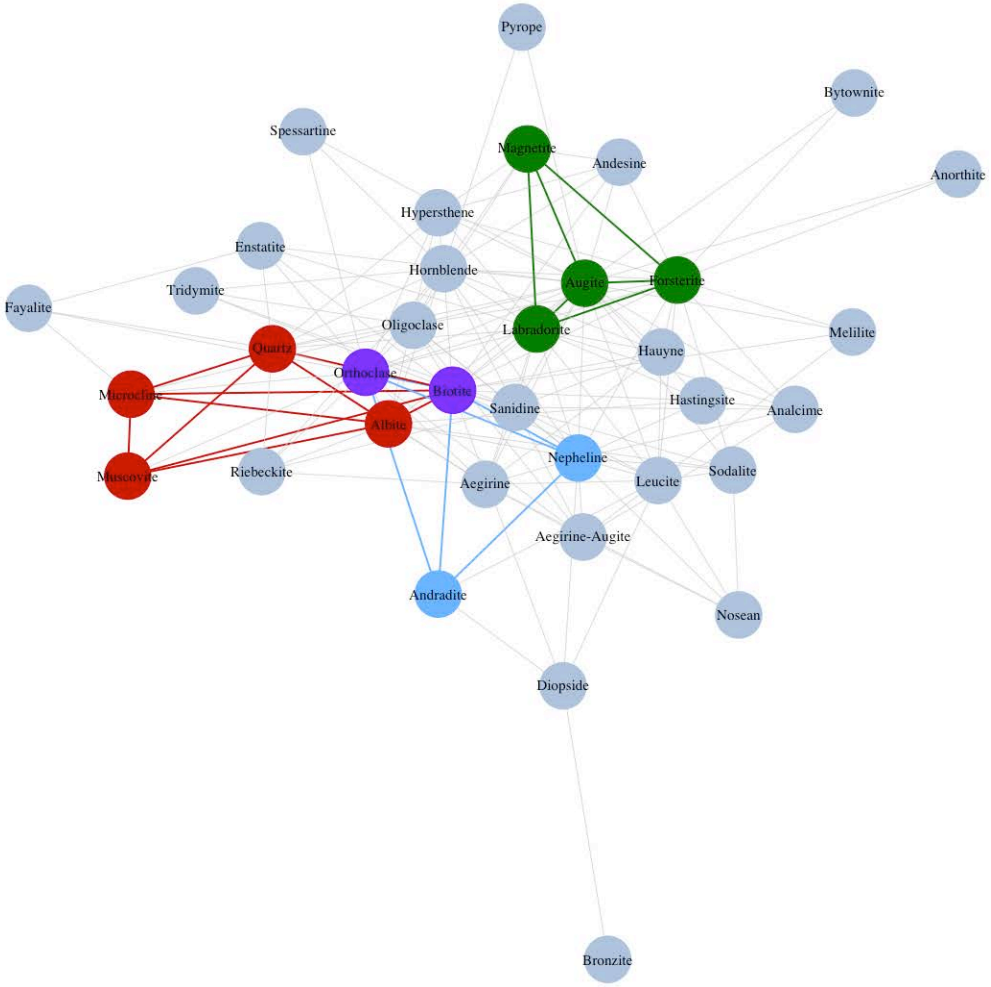



**Figure 1.** (A) A Fruchterman-Reingold force-directed network diagram of 36 rock-forming minerals in holocrystalline intrusive igneous rocks. Each circular node represents a rock-forming mineral and each link indicates pairs of coexisting minerals in one or more rocks, as recorded in Harker (1964). (B) Different types of igneous rocks appear as closely linked clusters, or "cliques," in this diagram.

A consequence of these graphical procedures is that each igneous rock type, such as granite, olivine basalt, or nepheline syenite, is embedded as a localized, fully interconnected subset of nodes, or "clique," in this network (Figure 1B). For example, the clique for minerals commonly found in granite includes quartz, muscovite, biotite,

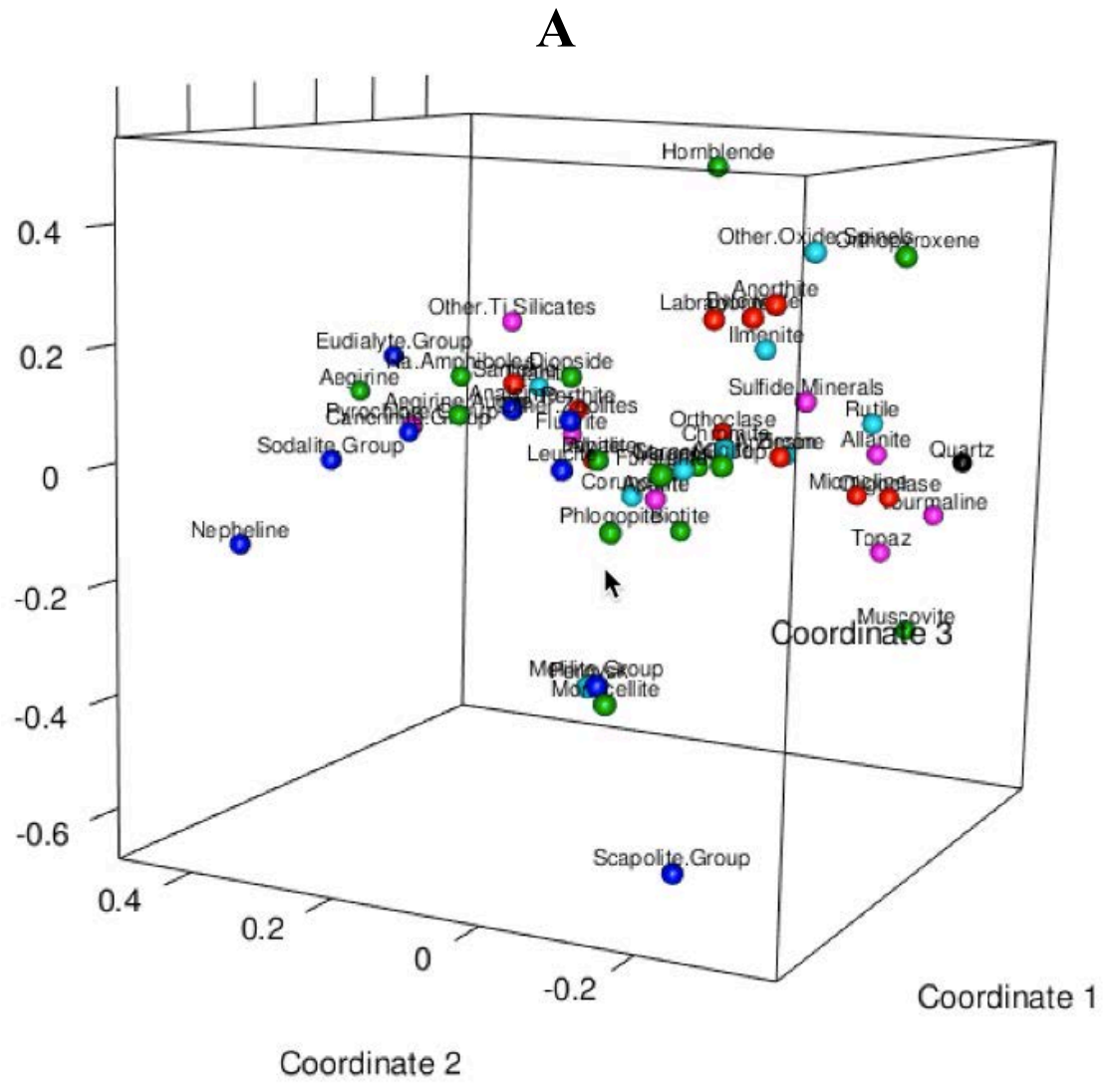
148 orthoclase, albite, oligoclase, microcline, hornblende, and riebeckite, whereas olivine
149 basalt contains the clique of labradorite, augite, forsterite, and magnetite. Each of the 77
150 holocrystalline igneous rocks described by Harker (1964) is similarly embedded in this
151 network. Thus, this visualization in a sense represents the sweep of igneous petrology in a
152 single diagram—a result that hints at the large amount of multi-dimensional information
153 embedded in network representations, while also suggesting a visual opportunity for
154 teaching and learning about rocks and minerals.
155
156 *Multi-Dimensional Scaling and Mineral Phase Topologies:* A major research
157 objective of mineralogy and petrology for more than a century has been the elucidation of
158 mineral reaction series and phase equilibria (e.g., Bowen 1928; Yoder 1976). We
159 postulate that, because mineral networks are based on observed assemblages of
160 coexisting minerals, they must embed information on phase topologies and thus have the
161 potential to reveal phase relationships in systems not yet studied experimentally.
162 To illustrate this potential we compiled coexisting mineral data on varied intrusive
163 igneous rocks from *A Descriptive Petrography of the Igneous Rocks* by Albert Johannsen
164 (1932-1938). The relatively small number of primary rock-forming minerals in intrusive
165 igneous rocks, coupled with the likelihood that these minerals formed under equilibrium
166 conditions and are not subject to the complications of metamorphism, diagenesis, and
167 other alteration processes, make these minerals an excellent test case for network
168 analysis. We consolidated the lists of minerals in Johannsen's multi-volume treatment of
169 729 crystalline igneous rocks into coexistence data for the 51 primary rock-forming

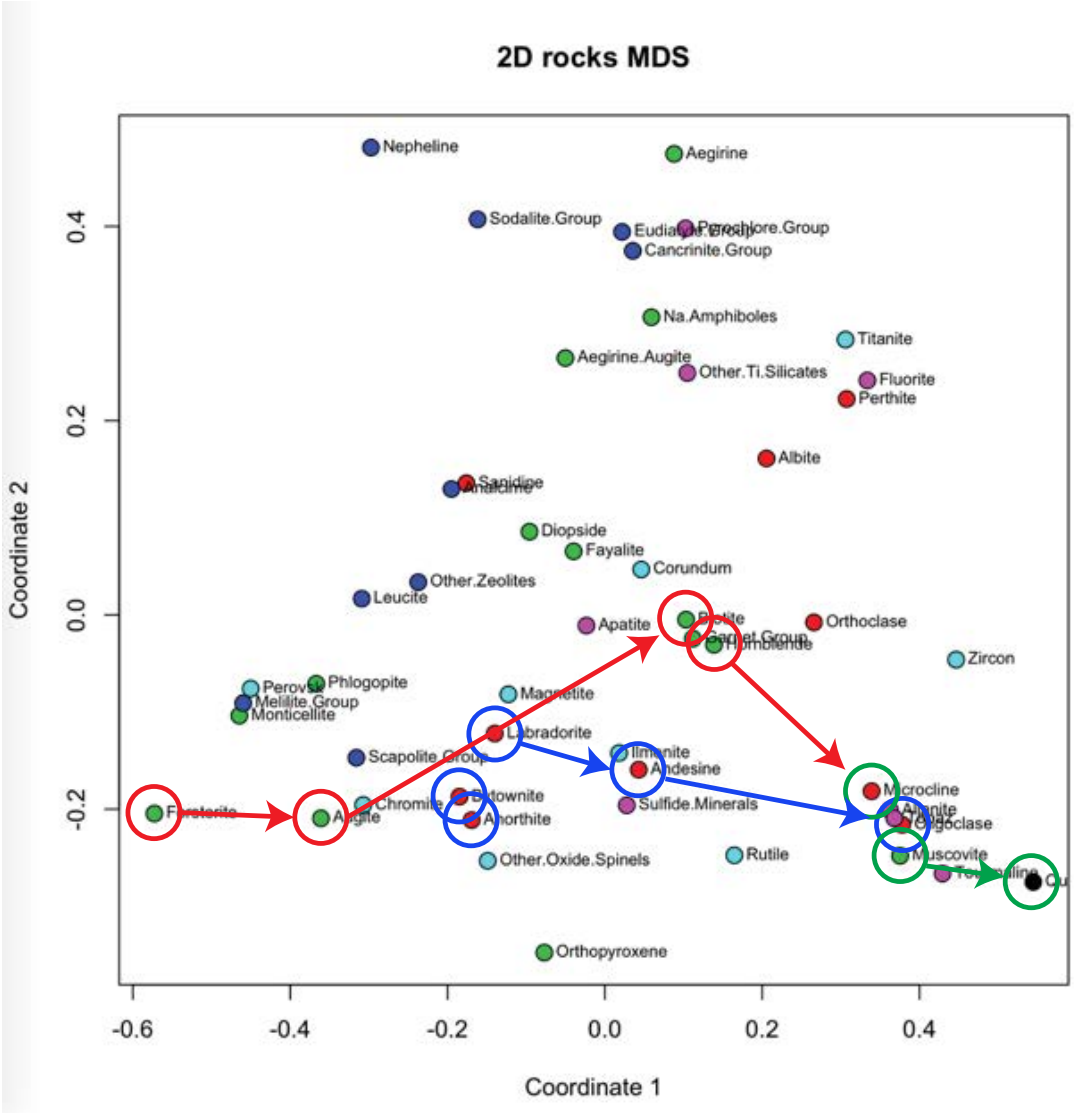
minerals (Supplemental Information 2). We used a variety of mineral network renderings to study the patterns of coexisting phases in these rocks.

We initially employed multi-dimensional scaling (MDS) in both three- and two-dimensional renderings (https://github.com/lic10/DTDI-DataAnalysis; Figures 2 and 3). MDS is an approach to visualizing the similarities between points of a high-dimensional dataset in a lower-dimensional space. The similarities between the data points are represented as distances between the projected points in the lower-dimensional space, where the objective of the scaling is to determine the coordinates of these projected points while preserving the distances as well as possible. In our case, the data points are mineral species, and the distances between points are inversely related to the degree of coexistence of the two minerals. We created the MDS diagrams in Figures 2 and 3 using the "cmdscale" command of the "stats" package in R (see https://github.com/lic10/DTDI-DataAnalysis). We loaded the Johannsen igneous rock dataset (1932-1938) into R as a data frame, and generated a second data frame as a symmetric 51×51 mineral matrix in which the value recorded at matrix element *ij* represents the calculated distance, $d_{ij}$ between nodes *i* and *j*. Distances were projected on both 2- and 3-dimensional spaces. We used the "rgl" package in R (Adler et al. 2016) to generate the 3D plot. In general, a network containing $N$ nodes requires a representation in $(N - 1)$ dimensions to satisfy exactly all $d_{ij}$. Consequently, MDS diagrams of fewer than $(N - 1)$ dimensions employ distance least-squares analysis to distribute nodes as a projection from higher-dimensional space.

Familiar aspects of igneous mineral phase relationships are embedded in the multi-dimensional scaling diagram for igneous minerals. For example, Bowen (1928) proposed

a mineral reaction series for igneous rocks in which Mg-Fe minerals tend to crystallize in a mafic cooling sequence (olivine → pyroxene → hornblende → biotite), whereas plagioclase feldspars transition from more calcium-rich to more sodium-rich varieties. At lower temperatures, late-stage minerals display a trend from alkali feldspar to muscovite to quartz. These mineral crystallization trends are mimicked from left-to-right in the MDS diagram, as illustrated in Figure 2B.





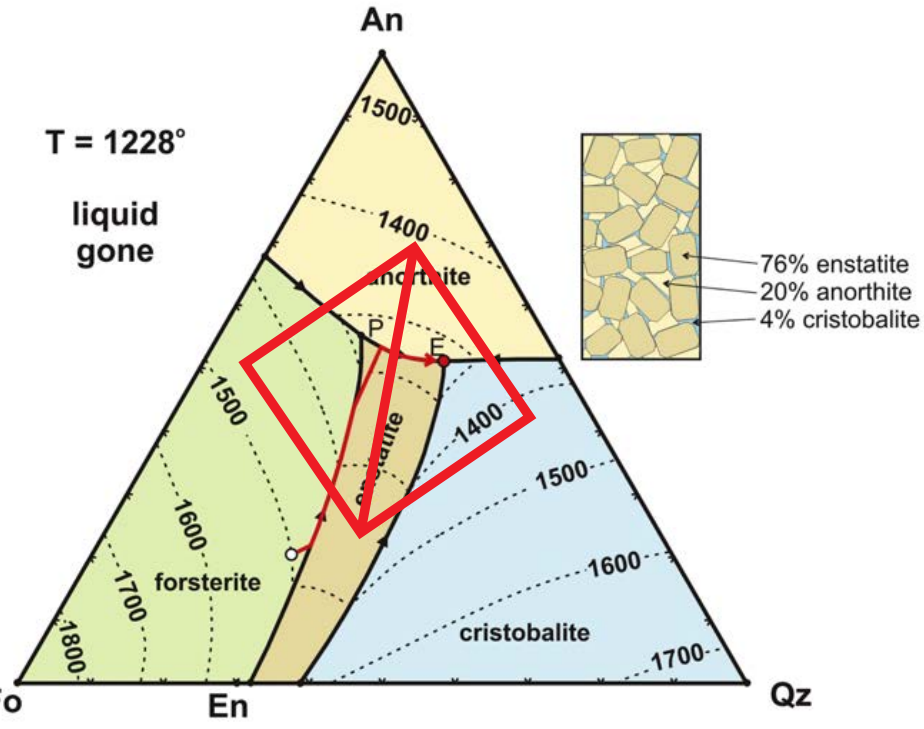

**B**

**Figure 2.** Multi-dimensional scaling diagrams of 51 rock-forming minerals in 729 igneous rocks tabulated by Johannsen (1932-1938). A. Three-dimensional rendering (see Supplemental Information 3 for animation). B. Two-dimensional rendering with minerals from Bowen's reaction series (Bowen 1928) circled and connected with arrows [mafic trend (circled in red), plagioclase series (circled in blue), and late-stage trend (circled in green)].

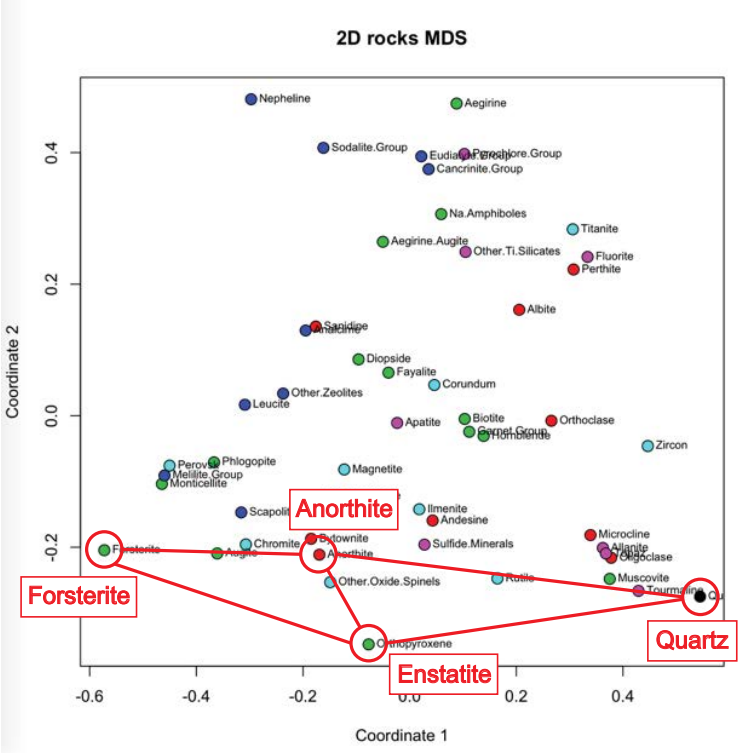
In addition, the topology of phase connections in mineral network diagrams mirrors their phase relationships. For example, the "AFQ" ternary phase diagram for the system anorthite ($CaAl_2Si_2O_8$)—forsterite ($Mg_2SiO_4$)—silica ($SiO_2$) illustrates that quartz may coexist with anorthite and an intermediate mineral enstatite ($MgSiO_3$), but not with

213 forsterite (Figure 3A). The topology of this phase diagram is also embedded in the MDS

214 diagram (Figure 3B).




215 **A** **B**




216

Figure 3. The topologies of phase diagrams, such as the anorthite-forsterite-quartz ternary solidus diagram (A), are mirrored in the topologies of mineral network diagrams (B). Ternary diagram after Anderson (1915).

The phase relationships of igneous rocks have been well documented through decades of studies in experimental petrology and thermochemical modeling, so the examples in Figures 2 and 3 illustrate the necessary conformity of network diagrams to established phase relationships. However, numerous other mineralogical systems have not been investigated in this detail. Much work remains to be done, but we postulate that mineral network analysis of coexisting species in other complex natural chemical systems holds the prospect of revealing unknown phase relationships through multi-dimensional analysis. In such analyses, care must be taken to ensure that connections between the mineral nodes actually represent equilibrium phase assemblages. In situations such as intrusive igneous rocks and cogenetic hydrothermal ore minerals, equilibrium formation is a safe assumption, and linked nodes will represent adjacent stability fields on the relevant phase diagram. However, greater care must be exercised when dealing with assemblages including secondary minerals such as oxidative weathering products, diagenesis, metamorphism, etc.

*Cluster Analysis and Paragenetic Modes:* A valuable attribute of network diagrams is that the node representations can incorporate additional dimensions of information through their size, color, shape, and patterning. In Figures 4 and 5B, we scaled node diameters and inter-node distances for Cr mineral species in the following way: If two minerals A and B occur at $a$ and $b$ localities, respectively, and they co-occur at $c$ localities, then the node diameters of A and B are $\log_2(a)$ and $\log_2(b)$, and the distance of

the link connecting A and B is [1 – $c$/min($a$,$b$)], where min($a$,$b$) is the smaller of $a$ and $b$. If $A$ and $B$ always occur together then we assign a minimum distance of 0.1.

Cluster analysis employs mineral network data to identify subsets of closely related species—an approach that can reveal previously unrecognized relationships among species. For example, we performed cluster analysis on the 30 most common terrestrial Cr minerals. We included minerals that satisfy three criteria: (1) Cr occupies more than 50% of at least one symmetrically distinct crystal lattice site; (2) the mineral occurs at three or more localities; and (3) the mineral co-occurs with other Cr minerals at two or more localities. In Figure 4 we applied the Walktrap Algorithm (Pons and Latapy 2005) of the igraph package in R to mineral coexistence data in mindat.org to detect clusters of closely related Cr minerals. This approach is based on the analysis of random walks among links. Random walks are more likely to stay within a single cluster because there are more links within a cluster than links leading to different clusters. When we employ this algorithm to perform 5-step random walks on the Cr mineral graph, the minerals separate naturally into four clusters, each of which can be associated with a different paragenetic mode. The largest of the four clusters includes 17 $Cr^{3+}$ minerals, all of which are high-temperature igneous, metamorphic, and hydrothermal species (group 1). Three additional clusters falling peripherally to this central cluster include all $Cr^{6+}$ minerals, seven of which (group 2) form from low-temperature, oxidized hydrothermal fluids leaching Cr-rich igneous rocks. The remaining six $Cr^{6+}$ minerals, which lie above the central cluster, are sedimentary species found in soils (group 3) and in desert environments (group 4). Cluster analysis is consistent with the observation that chromium in terrestrial $Cr^{6+}$ minerals is probably sourced from $Cr^{3+}$ reservoirs, either through

hydrothermal leaching or oxidative weathering (e.g., Liu et al. 2017). We conclude that cluster analysis holds promise for revealing patterns of diagenesis and distribution in a variety of mineral systems.

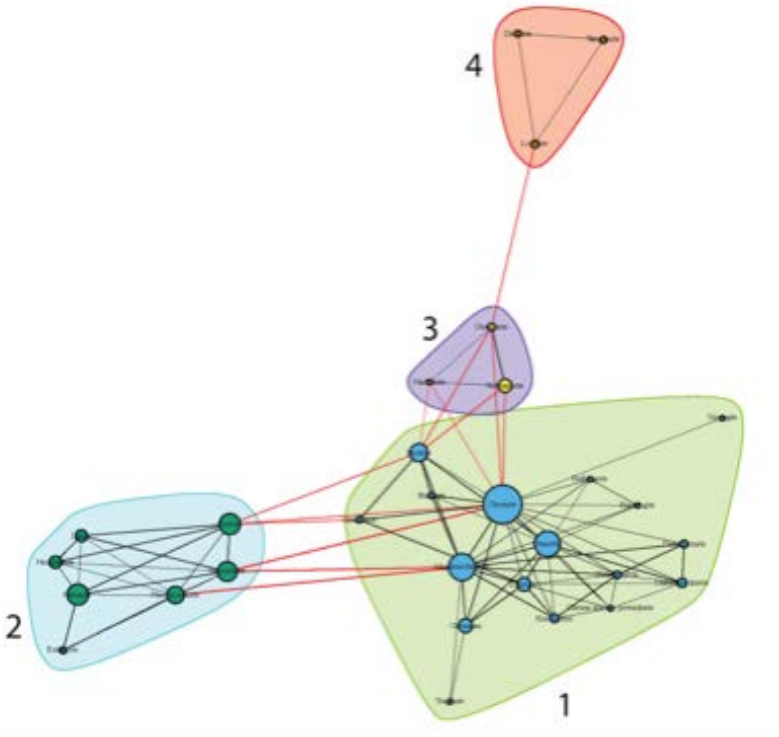



Figure 4. Cluster analysis of 30 common chromium-bearing minerals reveals segregation into 4 groups. The central cluster (group 1) includes 17 $Cr^{3+}$ species formed through igneous, metamorphic, or hydrothermal processes. The left-hand cluster (group 2) includes 7 $Cr^{6+}$ species formed through hydrothermal alteration, whereas the two smaller clusters (groups 3 and 4) include chromate minerals precipitated in soils and desert environments. Black lines indicate coexistence of minerals within a cluster, and red lines indicate coexistence between minerals of neighboring clusters.

*Force-Directed Mineral Graphs:* An important potential contribution of mineral network analysis lies in the simultaneous visualization and study of relationships among scores or hundreds of minerals that are related by composition, age, tectonic setting, deposit type, or numerous other variables. Force-directed graphs (Figure 5), which represent the distribution of nodes as a dynamic network with balanced spring-

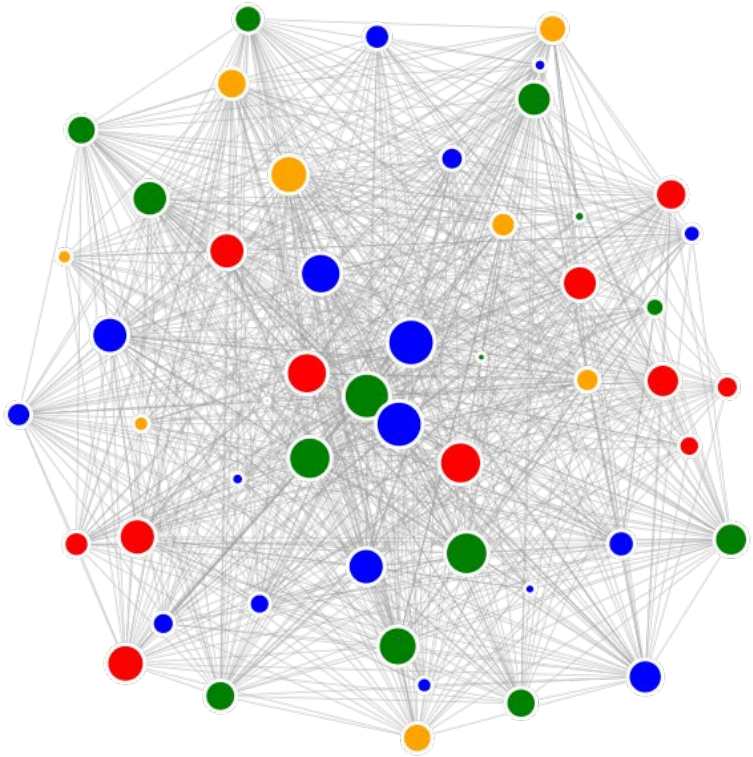
like interactions among nodes, are particularly useful in this regard. We generate these graphs by algorithms that run through a number of iterations, displacing the nodes according to fictive attractive and repulsive forces that they exert on each other, until a layout is found that minimizes the "energy" of the system and possibly satisfies other constraints such as drawing connected nodes at certain separations. These methods are implemented in highly customizable modules in multiple programming languages, such as Javascript and R, making it possible to render the graphs through a number of interfaces including web browsers.

In Figure 5, we created the browser-based force-directed graphs using the D3 4.0 d3-force module, which simulates physical forces using velocity Verlet integration (Verlet 1967) and implements the Barnes–Hut approximation (Barnes and Hut 1986) for performing *n*-body simulations, similar to those of molecular or planetary systems. For each of the three graphs we compiled a symmetric matrix whose non-diagonal elements represent the number of localities where two minerals coexist and whose diagonal elements represent the total number of localities at which each mineral is found. As a preliminary step we imported these data into R as data frames and converted into two lists, one with nodes representing all the minerals in the dataset, and the other with links representing coexistence relationships between the nodes. We created the list of nodes by extracting the row or column names of the data frame, each of which represents a mineral, and we produced the list of links by iterating over the upper or lower triangle of the matrix and copying the row name, column name, and computing a coexistence metric between the two minerals. We added additional fields to the nodes list, such as mineral compositions, the number of localities at which the

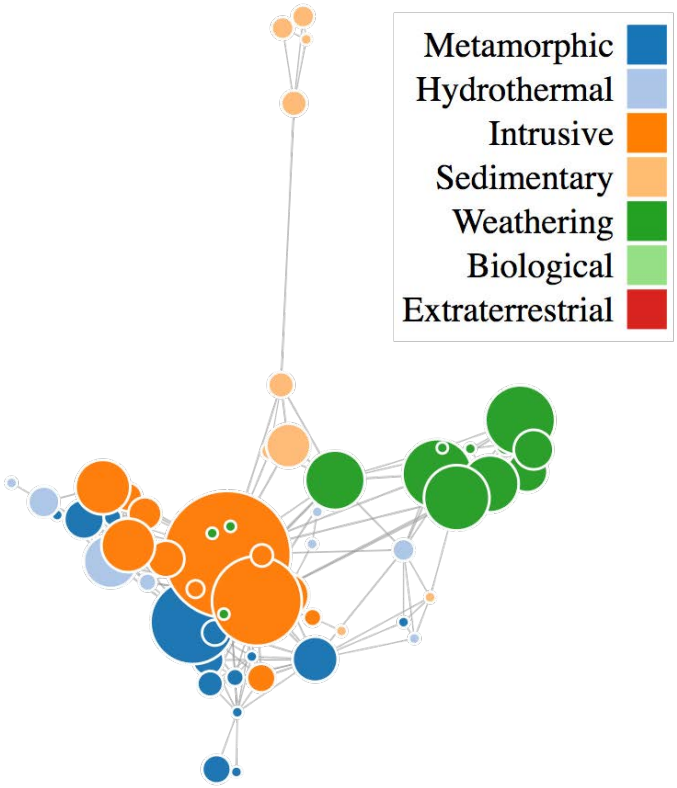
mineral occurs, and/or structural classification of the mineral.

We combined these two lists and converted them into a JSON (Javascript Object Notation) file, which is stored along with a web page written in HTML (Hypertext Markup Language) and Javascript that uses functions from the D3 4.0 library. The data file is read from the file system and rendered when the page is opened in a web-browser. Our Javascript code generates the layout by performing a many body (*n*-body) simulation and constraining edge lengths to values that equal the coexistence metric multiplied by a constant to make the connections more apparent. We set node sizes to the binary logarithm of the abundance value of a mineral in the cases of Cu and igneous rocks diagrams, and the actual abundance values in the Cr diagram. Node colors in Figure 5 variously indicate the structural classification of the minerals (igneous network), paragenetic mode (Cr network), and composition (Cu network).

The mineral network diagrams in this study require data on coexisting minerals in individual rocks or from individual localities. We manually generated spreadsheets of coexisting minerals in igneous rocks from text and tables in Harker (1964) and Johannsen (1932-1938) as presented in Supplemental Information 1 and 2. We used a PERL script to construct spreadsheets of coexisting chromium and copper minerals, which are generated automatically from data on coexisting species from localities recorded in the crowd-sourced mineral website mindat.org. We define Cr- or Cu-minerals as those reported in the official IMA list of minerals at rruff.info/ima. For each pair of coexisting minerals we generated a file that contains all localities at which those two minerals occur. A second program reads the assembled files to obtain the number of localities at which each pair occurs and outputs these counts in matrix form.




330 **A**




331 **B**

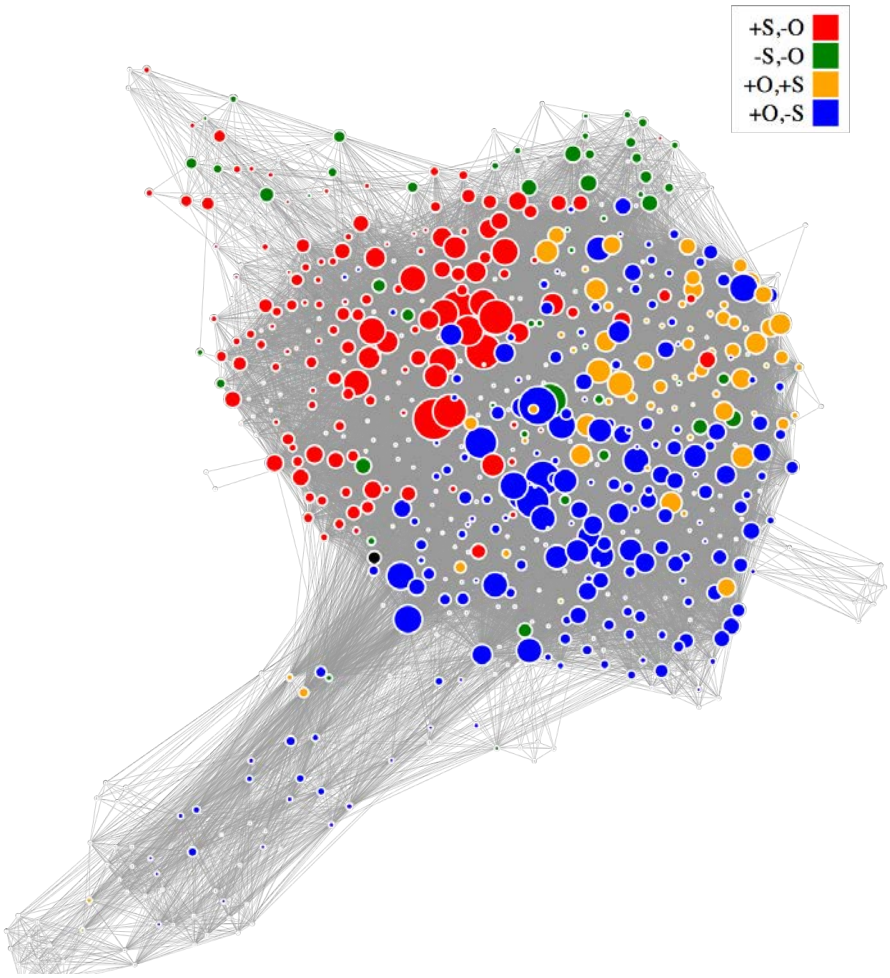



Figure 5. Force-directed network graphs of minerals. A. 51 rock-forming igneous minerals sorted by structural groups. B. 58 chromium minerals sorted by paragenetic mode. C. 664 copper minerals sorted by composition. See Supplemental Information 4, 5, and 6, respectively, for animations of these three dynamic graphs.

An important feature of browser-based force-directed graphs is that they can be manipulated with a mouse—individual nodes can be "pulled aside," thus deforming the network and illustrating the number and nature of links to other nodes (see movies in Supplemental Information 4, 5, and 6). Figure 5 presents static images of three contrasting force-directed graphs: (1) 51 common rock-forming igneous minerals; (2) 58 terrestrial minerals of chromium; and (3) 664 minerals of copper.

345 In Figure 5A, which represents connections among 51 igneous minerals, the node
346 colors indicate broad compositional groups (see figure for key). Note that while colors
347 are largely mixed, the red (quartz and feldspar minerals) and orange (feldspathoids and
348 zeolite mineral) nodes tend to concentrate near the lower and upper halves of the
349 network, respectively—a feature that reflects the natural avoidance of quartz and
350 feldspathoids. Node colors in Figure 5B for chromium minerals correspond to
351 paragenetic modes; note the strong clustering of nodes by color—an observation that
352 parallels the cluster analysis in Figure 4. Node colors in Figure 5C for copper minerals
353 indicate mineral compositions separated according to the presence or absence of sulfur or
354 oxygen. Strong segregation by color reveals clustering according to these compositional
355 variables for sulfides, sulfates, and oxygen-bearing species.
356
357 *Network metrics:* An important attribute of networks is the ability to compare and
358 contrast their topological characteristics through the use of many quantitative network
359 metrics (e.g., Newman 2013; Table 1). For example, a network's edge density $D$, defined
360 as the ratio of the number of observed links to the maximum possible number of links,
361 quantifies the extent to which a network is interconnected. For a network with $N$ nodes
362 and $L$ links:
363 $$D = 2L/[N(N-1)]. \qquad \text{[Equation1]}$$
364 $D$ can vary from 0 in a network with no links to 1 for a fully connected network. For
365 mineral networks, 0 means every mineral occurs by itself, whereas 1 means every
366 mineral co-occurs with every other mineral.

Freeman network centralization or degree centralization, *FNC*, is one of several measures of how many nodes play central roles in the network. In a network of *N* nodes, degree centralization for each node *i* is the number of links to other nodes, or node degree, deg(*i*). Freeman network centralization is defined as:

$$FNC = \sum_{i=1}^{N} \frac{\deg_{max} - \deg(i)}{(N-1)(N-2)}, \qquad \text{[Equation 2]}$$

in which $\deg_{max}$ is the maximum degree node. *FNC* can vary from 0 to 1; in a mineral network, low centralization indicates that minerals are uniformly interconnected, whereas high centralization indicates that only one or a few minerals are highly connected.

Transitivity, *T*, is defined by the ratio of the number of loops of length three and the number of paths of length two in a network. In mineral networks, 0 means that minerals co-occur only as pairs and 1 means that each mineral co-occurs with at least two others.

Diameter, *d*, of a network with *N* nodes is defined as the maximum value of the shortest path (i.e., "degree of separation") between any two nodes in the network, as determined by the number of edges and the average edge length between the two nodes.

Mean distance, *MD*, of a network with *N* nodes indicates the average path length, calculated from the shortest paths between all possible pairs of nodes. In a mineral network, *MD* represents the average separation between mineral pairs.

The three force-directed mineral networks illustrated in Figure 5 differ significantly in their network metrics. The igneous mineral network (Figure 5A) is relatively dense with high transitivity (*D* = 0.64; *T* = 0.77), while the network is decentralized (*FNC* = 0.34), and compact (*d* = 1.1; *MD* = 1.36). Two minerals, biotite and magnetite, have links to all other minerals; thus, manipulating the nodes for biotite and magnetite (see movie in Supplemental Information 4) results in a rapid return to the initial equilibrium network

configuration with those nodes appearing near the center of the network. Manipulation of quartz (near the bottom of the network) and nepheline (near the top), by contrast, illustrates the avoidance of those two minerals, which do not co-occur in igneous rocks. We postulate that the relatively high density and low diameter of this network are manifestations of high-temperature equilibrium associated with intrusive igneous rocks, for which a relatively few common rock-forming minerals occur in several lithologies.

**Table 1.** Network metrics for force-directed graphs (see Figure 5)

| Mineral System | Density | Centralization | Transitivity | Diameter | Mean Distance |
|---|---|---|---|---|---|
| Igneous minerals | 0.64 | 0.34 | 0.77 | 1.1 | 1.36 |
| Cr minerals | 0.05 | 0.33 | 0.44 | 6.0 | 2.65 |
| Cu minerals | 0.12 | 0.68 | 0.48 | 4.0 | 1.93 |

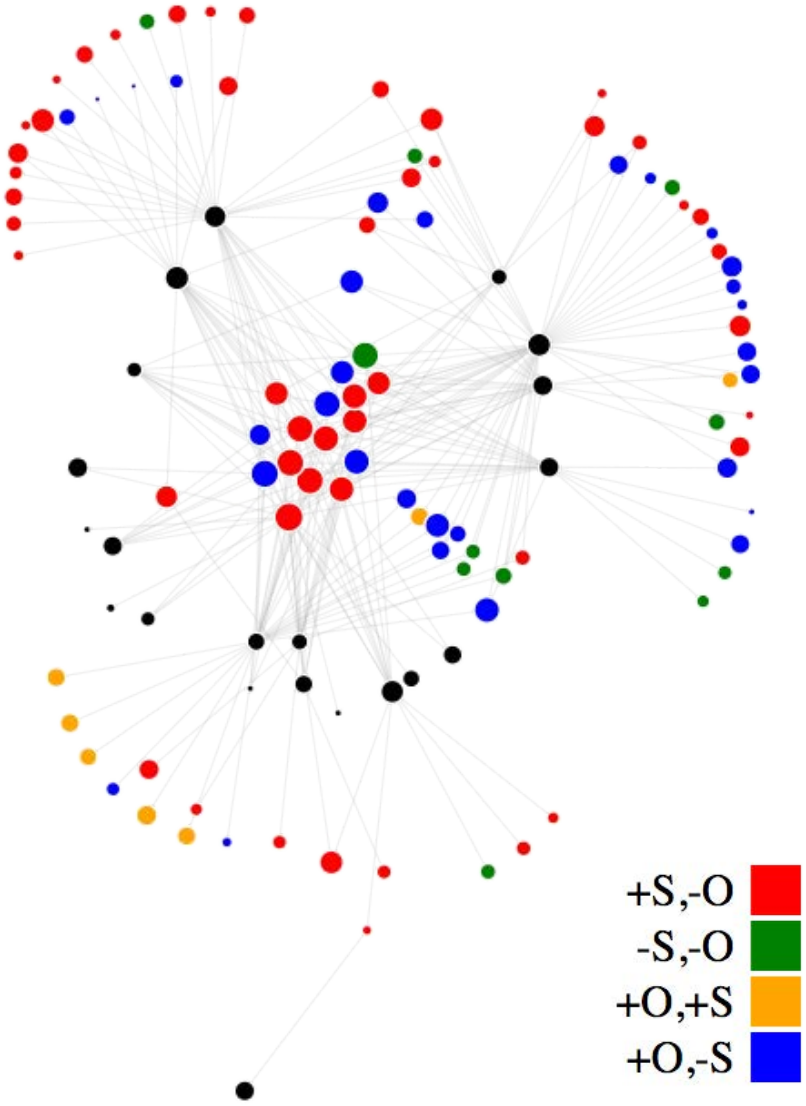
The network for 58 terrestrial chromium minerals (Figure 5B) contrasts with that for igneous minerals in that it possesses much lower density and transitivity ($D$ = 0.05; $T$ = 0.48), and greater diameter and mean distance ($d$ = 6; $MD$ = 2.65). These values are consistent with the cluster analysis (Figure 4), which revealed four groups of minerals that are largely separate from each other. A striking feature of this Cr mineral network is the segregation of nodes by colors, which represent paragenetic modes (see figure caption). As revealed by cluster analysis, chromium minerals occurring through weathering, formed during metamorphism, found in sediments, or crystallized through igneous processes tend not to co-occur and thus appear as somewhat isolated clusters in Figure 5B. On the other hand, hydrothermal Cr minerals are much more interconnected with phases formed through other paragenetic processes. Such complex relationships

among 58 minerals become obvious through manipulation of a force-directed graph (see movie in Supplemental Information 5) and exemplify the wealth of information contained in these network diagrams.

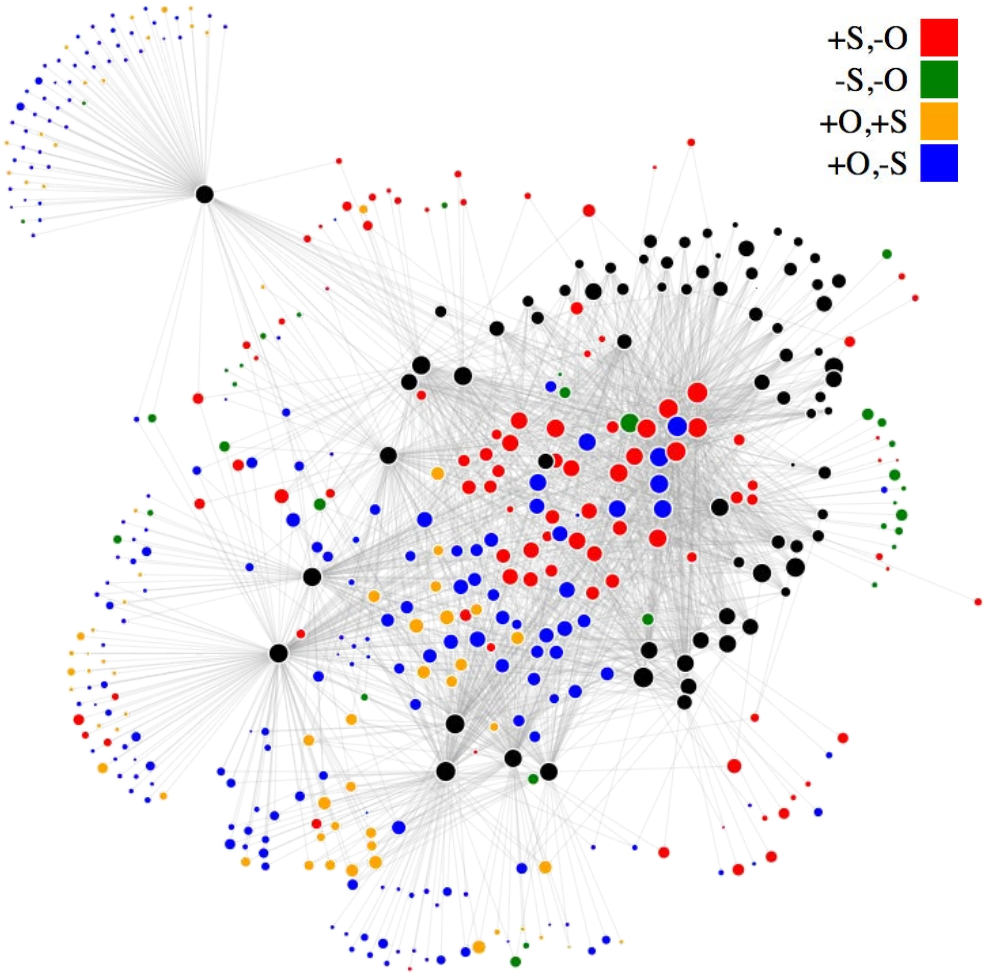
Copper minerals (Figure 5C) provide a third, contrasting example of a mineral network with relatively low density and transitivity ($D$ = 0.12; $T$ = 0.44), but high centrality ($FNC$ = 0.68), and intermediate diameter and mean distance ($d$ = 4; $MD$ = 1.93). Aspects of the co-existence of copper minerals are revealed in Figure 5C, which is colored according to the presence or absence of the two principal anions, O and S. A strong degree of segregation is seen for sulfides (red), sulfates (orange), and minerals with O but not S (blue). By contrast, copper minerals with neither O nor S (green) are much more widely distributed, as they are found associated with a variety of other copper minerals. Manipulations of the nodes for the two most interconnected copper minerals, chalcopyrite and malachite, reveal connections to all regions of the graph and result in significant distortion of the entire network (Supplemental Information 6). Manipulation of the node for native copper, on the other hand, shows greater connectivity to oxides and sulfates than to sulfides—an insight not readily obvious from tables of coexisting mineral species (and a finding that will be explored in more detail in a forthcoming study). The ability to view and interrogate simultaneously and dynamically the relationships among hundreds of mineral species underscores the power of force-directed mineral network visualizations.

*Bipartite Networks:* Bipartite networks incorporate two types of nodes and thus reveal information complementary to the previous examples (e.g., Asratian et al. 1998). Of

special interest in mineralogy are network diagrams that include nodes for both mineral species and their localities, with links connecting localities to mineral species found at those localities. In Figures 6A and 6B we present bipartite networks for copper minerals for two contrasting geological time intervals, from the Archean Eon (4.0 to 2.5 Ga) and the Cenozoic Era (66 Ma to present), respectively. We color mineral nodes according to mineral compositions with respect to the presence or absence of oxygen and sulfur, as in Figure 5C. Locality nodes, which in this case represent countries or broad geographic regions, appear in black.




445 **A**




**B**

Figure 6. Bipartite networks for copper minerals from the Archean Eon (A) and Cenozoic Era (B) reveal distinctive patterns of mineral diversity and distribution through space and time. Black nodes represent localities, whereas colored nodes represent mineral species linked to those localities. The distinctive pattern of an "O" or "U"-shape arrangement of localities with relatively few common minerals in the center area and a greater number of rare minerals in peripheral positions conforms to a Large Number of Rare Events frequency distribution (Hazen et al. 2015; Hystad 2015). Note also the increase in mineral diversity, as well as the evolution of mineral compositions, from A to B.

As with the previously demonstrated force-directed mineral networks, we employed mineral/locality data to produce the bipartite graphs. We imported data into R where two sets of nodes were extracted, one containing mineral species and the other containing the localities where these minerals occur. We combined the two sets of nodes into one list

and added an attribute to each item in the list, determining its type as either mineral or locality. We then generated a list of links from the data such that each link connects a mineral species to a locality. Following the same procedure as with the force-directed graphs created using the D3 4.0 library, we combined the data structures representing the nodes and links into a JSON file linked to an HTML page such that the diagrams can be rendered and manipulated in a web browser.

These striking bipartite networks provide simultaneous visual representations of data on the diversity and abundances of mineral species, as well as their geographical distributions, compositional characteristics, and geological ages. As such, these diagrams demonstrate the potential of network analysis to explore simultaneously numerous parameters related to mineral diversity and distribution and thus to reveal previously unrecognized aspects of mineral evolution and mineral ecology. Insights from these visualizations include:

- In both networks the nodes of the force-directed graph self-organize into a distinctive pattern with black locality nodes forming an "O" or "U"-shape arrangement. The commoner minerals, those found at numerous localities, appear as colored nodes near the center of these diagrams, whereas a significantly greater number of rare minerals that occur at only one or two localities plot as colored nodes in clusters and "fans" of minerals arranged around the periphery of the diagram. This unanticipated elegant geometry is the visual manifestation of a Large Number of Rare Events (LNRE) frequency distribution that characterizes Earth's near-surface mineralogy (Hazen et al. 2015; Hystad et al. 2015).

- The Archean bipartite network (Figure 6A), with 97 Cu minerals from 45 broad geographical localities, reveals that sulfide minerals dominated copper mineralogy prior to the Great Oxidation Event (e.g., Hazen et al. 2008; Canfield 2014; Lyons et al. 2014). Sulfides represent 17 (74%) of the 23 common Archean copper minerals located "inside" the ring of black locality nodes and 32 (50%) of the 64 rare minerals located around the periphery. Note also the relative paucity of sulfate minerals—only 7 species (7%), all of them rare, out of 97 Archean copper minerals.
- The Cenozoic bipartite network for copper minerals contrasts with that of the Archean Eon in a number of respects. The significant increase in the number of identified mineral species (colored nodes), from 97 to almost 400, is to be expected when comparing Earth's recent mineralogy with the scant record of rocks more than 2.5 billion years old. However, there are also striking and previously unrecognized differences in the distributions of mineral compositions from these two geological intervals. Sulfide minerals (red nodes) continue to make up a significant fraction of the most common species located near the center of the diagram. Of the approximately 100 mineral nodes located within the "U" of black locality nodes, more than 40 are sulfide minerals. Furthermore, most of these phases are concentrated at the "bottom" of the "U"—a position representing the most widely distributed copper minerals. Of the remaining common Cu phases in the central region, most are carbonate, phosphate, and other minerals that contain oxygen but not sulfur (blue nodes concentrated in the "upper" region inside the "U"), perhaps reflecting the oxygenation by photosynthesis of Earth's

atmosphere and oceans, and the corresponding generation of novel oxidized copper mineral species.

- Peripheral (i.e., rare) copper minerals from the Cenozoic Era differ markedly in composition from those of the Archean Eon. Sulfide minerals account for only about 50 (<20%) of the more than 280 rare species, whereas at least 210 (~75%) oxygen-bearing minerals, 60 of them sulfates, decorate the diagram in sprays and clusters of phases known from only one or two geographic regions.

These intriguing insights regarding copper mineral evolution and ecology have been hidden among large data tables of more than 600 species from more than 10,000 localities, representing more than 100,000 individual mineral-locality data (rruff.inof/ima; mindat.org). Research now in progress will investigate these intriguing trends for copper mineral evolution and ecology in greater detail, while searching for patterns that might point to the prediction of new copper minerals and ore deposits.

## CONCLUSIONS

Network analysis provides mineralogists and petrologists with a dynamic, multi-dimensional, quantitative visualization approach to explore complex and otherwise hidden patterns of diversity and distribution in systems of numerous minerals—information that heretofore has been buried in large and growing mineral data resources. Open access data repositories now document more than 5200 mineral species (rruff.info/ima), from 275,000 localities, incorporating approximately one million mineral/locality data (mindat.org). It is thus possible to employ mineral network visualizations to quantitatively investigate patterns of coexistence, phase relationships,

528 reaction pathways, network metrics, frequency distributions, and deep-time evolution of
529 virtually any mineral group.

530 We suggest that further investigation of mineral networks will reveal previously
531 hidden patterns of species coexistence and clustering based, for example, on structure
532 type, chemistry, age, solubility, hardness and other mechanical properties, redox state,
533 depth and temperature of formation, year and method of mineral discovery, and
534 paragenetic mode. Mineral metadata, furthermore, permit exploration of mineral subsets
535 through filtering by geographic region, tectonic setting, co-occurrence with varied
536 biozones, economic resources, environmental characteristics, and other key parameters.
537 In addition, networks are now being generated for minerals on Mars, the Moon, and
538 Vesta (as represented by "HED" achondrite meteorites) with the motivation to compare
539 and contrast mineral evolution and ecology of different planets and moons.

540 Given the inherent beauty and richness of these visualization tools, it is perhaps easy
541 to become distracted from the varied, multi-dimensional, and as yet unexplored aspects of
542 mineralogy that networks promise to illuminate. We look to a future when the
543 consolidated network of all 5200 mineral species, distributed among hundreds of
544 thousands of localities, will offer an unparalleled open access research tool. We conclude
545 that mineral network analysis, by combining the potential of big data mineralogy with a
546 dynamic and accessible visual aesthetic, represents a powerful new method to explore
547 fundamental problems in mineralogy and petrology.

548

549

550

## ACKNOWLEDGEMENTS

We thank Alex Pires for help in database development and Sophie Kolankowski, Benno Lee, Marshall X. Ma, Han Wang, Stephan Zednik, and Hao Zhong for assistance in the development of varied visualization methods. Craig Schiffries, John Hughes, and an anonymous reviewer provided thoughtful comments and suggestions. This work was supported by the W. M. Keck Foundation's Deep-Time Data Infrastructure project, with additional support by the Deep Carbon Observatory, the Alfred P. Sloan Foundation, a private foundation, the Carnegie Institution for Science, and NASA NNX11AP82A, Mars Science Laboratory Investigations. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration.

## REFERENCES

Abraham, A., Hassanien, A.-E., and Snasel, V. [Editors] (2010) Computational Social Network Analysis: Trends, Tools and Research Advances. Springer, New York.

Adler, D, et al. (2016). rgl: 3D Visualization Using OpenGL. R package version 0.95.1441. http://CRAN.R-project.org/package=rgl

Amital, G., Shemesh, A., Sitbon, E., Shklar, M., Netanely, D., Venger, I., and Pietrokovski, S. (2004) Network analysis of protein structures identifies functional residues. Journal of Molecular Biology, 344, 1135-1146.

Anderson, O. (1915) The system anorthite-forsterite-silica. American Journal of Science, 39, 407-454.

Asratian, A.S., Denley, T.M.J., and Häggkvist, R. (1998) Bipartite Graphs and their Applications. Cambridge University Press, New York.

Banda-R, K. et al. (2016) Plant diversity patterns in neotropical dry forests and their conservation implications. Science, 353, 1383-1387.

Barnes, J. and Hut, P. (1986) A hierarchical O(N log N) force-calculation algorithm. Nature, 324, 446-449.

Bowen, N.L. (1928) The Evolution of the Igneous Rocks. Princeton University Press, New Jersey.

Canfield, D. (2014) Oxygen: A Four-Billion Year History. Princeton University Press, New Jersey.

Cheng, S., Karker, S., Bapteste, E., Yee, N., Falkowski, P., and Bhattacharya, D. (2014) Sequence similarity network reveals the imprints of major diversification events in the evolution of microbial life. Frontiers in Ecology and Evolution, 2, 72. Doi:

10.3389/fevo.2014.00072

Christy, A.G., Mills, S.J., Kampf, A.R., Houseley, R.M., Thorne, B., and Marty, J. (2016) The relationship between mineral composition, crystal structure and paragenetic sequence: the case of secondary Te mineralization at the Bird Nest druft, Otto Mountain, California, USA. Mineralogical Magazine, 80, 291-310.

Corel, E., Lopez, P., Méheust, R., and Bapteste, E. (2016) Network-thinking: Graphs to analyze microbial complexity and evolution. Trends in Microbiology, 24, 224-237. DOI: 10.1016/j.tim.2015.12.003

Costanzo, M. et al. (2016) A global genetic interaction network maps a wiring diagram of cellular function. Science, 353, 1381.

Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. InterJournal, Complex Systems, 1695, 1-9.

Dong, W. and Pentland, A. (2009) A network analysis of road traffic with vehicle tracking data. In Proceedings of the American Association of Artificial Intelligence, Spring Symposium, Human Behavior Modeling, Palo Alto CA, pp.7-12.

Dunhill, A.M., Bestwick, J., Narey, H., and Sciberras, J. (2016) Dinosaur biogeographical structure and Mesozoic continental fragmentation: A network–based approach. Journal of Biogeography. doi: 10.1111/jbi.12766 (in press).

Dunne, J.A., Williams, R.J., Martinez, N.D., Wood, R.A., and Erwin, D.H. (2008) Compilation and network analysis of Cambrian Food webs, PLoS Biology, 6, 693-708.

Fruchterman, T.M.J. and Reingold, E.M. (1991) Graph drawing by force-directed placement. Software: Practice and Experience, 21, 1129-1164.

609 Geem, Z.W. (2010) Optimal cost design of water distribution networks using harmony
610 search. Engineering Optimization 38, 259-277.
611 Guimerá, R., Mossa, S., Turschi, A., and Amaral, L.A.N. (2005) The worldwide air
612 transportation metwork: anomalous centrality, community structure, and cities' global
613 roles. Proceedings of the National Academy of Sciences USA, 102, 7794-7799.
614 Harel, A., Karkar, S., Cheng, S., Falkowski, P.G., and Bhattacharya, D. (2015)
615 Deciphering primordial cyanobacterial genome functions from protein network
616 analysis. Current Biology, 25, 628-634.
617 Harker, A. (1964) Harker's Petrology for Students, 8th edition, revised. Cambridge
618 University Press, New York.
619 Hazen, R.M., Papineau, D., Bleeker, W., Downs, R.T., M., J., McCoy, T.L., Sverjensky,
620 D.A., and Yang, H. (2008) Mineral evolution. American Mineralogist, 93, 1693-1720.
621 Hazen, R.M., Grew, E.S., Downs, R.T., Golden, J., and Hystad, G. (2015) Mineral
622 ecology: Chance and necessity in the mineral diversity of terrestrial planets. Canadian
623 Mineralogist, 53, 295-323. DOI: 10.3749/canmin.1400086.
624 Huang, B., Zhan, R.-B., and Wang, G.-X. (2016) Recovery brachiopod associations from
625 the lower Silurian of South China and their paleoecological implications. Canadian
626 Journal of Earth Science, 53, 674-679.
627 Hwang, N. and Houghtalen, R. (1996) Fundamentals of Hydraulic Engineering Systems.
628 Prentice Hall, Upper Saddle River, New Jersey.
629 Hystad, G., Downs, R.T., and Hazen, R.M. (2015) Mineral frequency distribution data
630 conform to a LNRE model: Prediction of Earth's "missing" minerals. Mathematical
631 Geosciences, 47, 647-661.

632 Johannsen, A. (1932-1938) A Descriptive Petrography of the Igneous Rocks: 4 Volumes.
633 University of Chicago Press, Illinois.
634 Kadushin, C. (2012) Understanding Social Networks. Oxford University Press, New
635 York.
636 Kolaczyk, E.D. (2009) Statistical Analysis of Network Data. Springer, New York.
637 Leuenberger, P., Ganscha, S., Kahraman, A., Cappellitti, V., Boersema, P.J., von Mering,
638 C., Claassen, M., and Picotti, P. (2017) Cell-wide analysis of protein thermal
639 unfolding reveals determinants of thermostability. Science, 355, 812.
640 Liu, C., Hystad, G., Golden, J.J., Hummer, D.R., Downs, R.T., Morrison, S.M., Grew,
641 E.S., and Hazen, R.M. (2017) Chromium mineral ecology. American Mineralogist, in
642 press.
643 Lyons, T.W., Peinhard, C.T., and Planavsky, N.J. (2014) The rise of oxygen in Earth's
644 early ocean and atmosphere. Nature, 506, 307-314.
645 Martinez, N.D. (1992) Constant connectance in community food webs. American
646 Naturalist, 139, 1208-1212.
647 Müller, B., Reinhardt, J., and Strickland, M.T. (1995) Neural Networks: An Introduction.
648 2nd Edition. Springer, New York.
649 Newman, M.E.J. (2013) Networks: An Introduction. Oxford University Press, New York.
650 Otte, E. and Rousseau, R. (2002) Social network analysis: a powerful strategy, also for
651 the information sciences. Journal of Information Science, 28, 441-453.
652 Pagani, G.A. and Aiello, M. (2013) The power grid as a complex network: A survey.
653 Physica A, 392, 2688-2700.
654 Pinheiro, C.A.R. (2011) Social Network Analysis in Telecommunications. Wiley,

655 Hoboken, New Jersey.
656 Pons, P, and Latapy, M. (2005) Computing communities in large networks using random
657 walks. International Symposium on Computer and Information Sciences. Springer,
658 New York.
659 Scott, J. and Carrington, P.J. (2011) The SAGE Handbook of Social Network Analysis.
660 SAGE: Los Angeles, California.
661 Sidor, C.A., Vilhena, D.A., Angielczyk, K.D., Huttenlocker, A.K., Nesbitt, S.J., Peecook,
662 B.R., Steyer, J.S., Smith, R.M.H., and Tsuji, L.A. (2013) Provincialization of
663 terrestrial faunas following the end–Permian mass extinction. Proceedings of the
664 National Academy of Sciences USA, 110, 8129–8133. doi: 10.1073/pnas.1302323110.
665 Uezu, A., Kanak, D.J., Bradshaw, T.W.A., Soderblom, E.J., Catavero, C.M., Burette,
666 A.C., Weinberg, R.J., and Soderling, S.H. (2016) Identification of an elaborate
667 complex mediating postsynaptic inhibition. Science, 353, 1123-1128.
668 Verlet, L. (1967) Computer "Experiments" on Classical Fluids. I. Thermodynamical
669 Properties of Lennard−Jones Molecules. Physical Review, 159, 98-103.
670 Vilhena, D.A., Harris, E.B., Bergstrom, C.T., Maliska, M.E., Ward, P.D., Sidor, C.A.,
671 Strömberg, C.A.E., and Wilson, G.P. (2013) Bivalve network reveals latitudinal
672 selectivity gradient at the end–Cretaceous mass extinction. Science Reports, 3. doi:
673 10.1038/srep01790.
674 Yoder, H.S. Jr (1976) Generation of Basaltic Magma. National Academy of Sciences
675 Washington DC.
676

677 **Supplemental Information**

678

679 **Supplemental Information Contains:**

680 **1.** Excel file of 36 minerals from 76 intrusive igneous rocks, as extracted from Harker
681 (1964).

682 **2.** Excel file of 51 minerals from 725 intrusive igneous rocks, as extracted from
683 Johannsen (1932-1938).

684 **3.** Movie of the rotation of the 3-dimensional multi-dimensional scaling diagram of rock-
685 forming minerals in intrusive igneous rocks.

686 **4.** Movie of the manipulation of nodes for biotite, magnetite, quartz, and nepheline in the
687 force-directed graph of intrusive igneous minerals.

688 **5.** Movie of the manipulation of nodes for chromite and crocoite in the force-directed
689 graph of terrestrial chromium minerals.

690 **6.** Movie of the manipulation of nodes for chalcopyrite, malachite, and copper in the
691 force-directed graph of the 242 most common copper minerals (≥ 50 localities).