



Cite this: DOI: 10.1039/d6dd00006a

# A graph-based approach to selection of feasible compositions for compositionally graded alloy design

Mikayla Obrist,<sup>ab</sup> James Hanagan,<sup>b</sup> Marshall Allen,<sup>c</sup> Bernard Gaskey,<sup>a</sup> Richard Malak<sup>c</sup> and Raymundo Arróyave<sup>\*bcd</sup>

Compositionally graded alloys (CGAs) offer a promising solution for engineering components that must perform under spatially varying conditions and requirements. However, selecting the appropriate compositions along a gradient is challenging because of potential chemical incompatibility, mechanical mismatch, or thermal expansion differences, which can compromise manufacturability or part performance. This study expands on recent advances in graph-based computational approaches for CGA design by addressing the question of how to work with a design space containing multiple isolated regions of feasible compositions. Using the quinary Nb–Cr–V–W–Zr system, the thermal and mechanical properties were computed using a combination of CALPHAD simulations, rule-of-mixtures models, and empirical estimates. The resulting data were embedded into a labeled property graph (LPG) structure and filtered using constraints that reduced the potential for solidification-related defects, promoted phase stability across the gradient, and ensured processability *via* additive manufacturing. A set of constraints produced four isolated subgraphs containing groups of connected compositions that were further evaluated using key property metrics to highlight their respective strengths and weaknesses in the context of CGA design and manufacturing. Overall, this work establishes methods for subgraph analysis and selection to provide guidance on narrowing design spaces down to one traversable group of compositions that aligns with performance goals and manufacturing constraints for any CGA design scenario.

Received 8th January 2026

Accepted 27th May 2026

DOI: 10.1039/d6dd00006a

rsc.li/digitaldiscovery

## 1 Introduction

Monolithic alloys have long served as the foundation of engineering materials, offering uniform and reliable properties throughout a component. While this approach has been widely adopted for its simplicity and established performance, it limits the ability to tailor local properties in components that experience spatially varying service conditions. Modern components are often subjected to complex thermal gradients, mechanical loading profiles, or corrosive environments that vary spatially within the same structure. As engineering demands become more complex, applications increasingly require materials with spatially tailored properties that cannot be achieved using monolithic alloys.

The need for spatially optimized material performance has motivated the development of compositionally graded alloys (CGAs).<sup>1</sup> CGAs are materials characterized by a gradual and deliberate variation in the chemical composition across a single component. This graded nature enables a material to exhibit different microstructures and properties locally without introducing sharp interfaces that typically cause mechanical or thermal incompatibilities. As a result, CGAs allow the continuous tailoring of material behavior within a single part, enabling each region to be optimized for a specific function. For example, by controlling the spatial distribution of alloying elements, CGAs can be designed to combine high core strength, high thermal resistance in external regions, and improved surface corrosion resistance within a single functionally graded component (*e.g.*, turbine blades).<sup>1</sup> Engineered compositionally graded alloys thus provide a deterministic framework for the co-design of chemistry, microstructure, and performance in continuous metallic systems.

This versatility makes CGAs attractive for demanding industrial applications such as aerospace, where materials must withstand extreme temperatures, high mechanical stresses, and chemically aggressive environments.<sup>2</sup> Recent advances in additive manufacturing (AM) have demonstrated the ability to

<sup>a</sup>Los Alamos National Laboratory, Los Alamos, NM 87545, USA<sup>b</sup>Department of Materials Science and Engineering, Texas A&M University, College Station, TX 77843, USA. E-mail: rarroyave@tamu.edu<sup>j</sup>J. Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, TX 77843, USA<sup>d</sup>Wm Michael Barnes '64 Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843, USA

fabricate graded alloys with smoothly varying compositions, enabling controlled transitions in microstructure and properties.<sup>3–5</sup> These advances highlight the potential of CGAs to reconcile competing performance requirements across a range of industries. For example, CGAs enable combinations of low density, high mechanical strength, and thermal stability, while also providing design pathways to tailor performance for specific applications.<sup>6</sup>

Beyond aerospace, CGAs have shown promise in sectors such as automotive and biomedical engineering. In these areas, graded materials are used to balance competing requirements. For example, they enable a balance between wear resistance, ductility, and weight reduction in automotive components, or enhance both biocompatibility and durability in medical implants.<sup>7,8</sup> These applications highlight the ability of CGAs to balance multiple performance demands within a single component, a challenge that traditional homogeneous alloys struggle to meet.

Despite their advantages, designing and manufacturing CGAs presents significant challenges. The interplay among mechanical, thermal, and chemical properties is complex, making it difficult to predict and optimize performance. Changes that improve one property may simultaneously compromise another; for example, strengthening an alloy often reduces its ductility.<sup>9</sup> Compatibility between alloy systems is critical, as mismatches in properties such as thermal expansion can induce residual stresses, cracking, or interfacial delamination. Recent studies further demonstrate that compositional gradients strongly affect the evolution of microstructure and properties during processing and service.<sup>10,11</sup>

Given this growing complexity, efficiently exploring alloy systems to design CGAs has become increasingly important. Traditional path-planning algorithms, such as Rapidly-Exploring Random Trees (RRT\*), have been applied in previous CGA studies to identify feasible compositional transitions.<sup>3,4,12,13</sup> While much more effective than previous approaches, these algorithms are computationally expensive and difficult to work with when the material design space is overconstrained.

RRT\* incrementally constructs a search tree to explore high-dimensional spaces; however, when the feasible region is heavily restricted by thermodynamic, mechanical, or processability constraints, its reliance on probabilistic sampling and local rewiring leads to substantial computational overhead. Furthermore, limited interpretability of the design space in high dimensions can result in failed path-planning runs due to an overconstrained design space where the obstacle region completely separates the two endpoints. This scenario results in time-consuming constraint relaxation, retraining of constraint classification models, and path-planning reruns to continue toward a valid CGA design. As the design process considers more material properties, processing variables, and thermodynamic constraints, more efficient computational frameworks are required to enable real-time exploration of alloy systems, specifically in the context of designing CGA composition paths.<sup>14,15</sup>

This motivates the need for more efficient, scalable frameworks that can handle constraint-driven CGA design environments. Prior to the development of more advanced graph-based frameworks, researchers relied on relational databases or Resource Description Framework (RDF)-based representations,<sup>16</sup> but these approaches struggled with performance and scalability in dense, high-dimensional datasets.<sup>17</sup> To address these limitations, researchers have increasingly explored labeled property graphs (LPGs). LPGs are graph-based structures composed of nodes and edges, both of which can store key-value pairs to directly represent descriptive attributes such as composition values, processing parameters, or performance metrics.<sup>18</sup> This results in a more compact, information-rich structure that supports efficient querying of complex relationships. Allen *et al.* demonstrated the utility of LPGs in metal additive manufacturing by introducing a schema that captures interdependencies among compositions, processing conditions, and performance outcomes.<sup>1</sup> Their implementation showcased how LPGs can scale with increasing system complexity while maintaining computational efficiency, establishing a foundation for this study. Building on this work, we expand the use of LPGs by introducing subgraph analysis techniques for identifying and selecting regions of feasible compositions that are suitable for CGA design.

While the implementation by Allen *et al.*<sup>1</sup> demonstrated how LPGs can organize and traverse complex compositional data, the graph model also enables analytical capabilities that extend well beyond basic data representation. Property-based filtering can isolate compositions that satisfy specific criteria; path traversal can identify viable compositional gradients; and subgraph extraction can reveal clusters of related compositions that remain connected after filtering. The LPG framework also supports schema validation to ensure data consistency and enables flexible querying for exploratory analysis.<sup>18</sup> By only maintaining the connecting edges between compositions that satisfy all imposed constraints, the resulting graph explicitly defines feasible paths, enabling the use of efficient traversal algorithms and avoiding costly *post hoc* validation steps. This approach stands in sharp contrast to search-based methods such as RRT\*, offering a scalable alternative for accelerated design.

To support the exploration of compositional trends and material behavior in CGA design, this study integrates a suite of computational tools grounded in the CALculation of PHase Diagrams (CALPHAD) framework. Thermodynamic simulations using Thermo-Calc yield phase-stability data, solidus and liquidus temperatures, and equilibrium phase fractions, whereas the rule of mixtures provides rapid estimates of bulk descriptors such as the Pugh ratio. Mechanical property models further estimate key performance metrics, including creep resistance, yield strength, and the Kou hot-cracking criterion, which support the assessment of alloy viability across the design space. To interpret this high-dimensional dataset, visualization techniques are essential. Vela *et al.*<sup>19</sup> outlined best practices for visualizing high-entropy and compositionally complex alloy spaces, emphasizing the importance of methods that reveal composition–property relationships. Following this approach,



Uniform Manifold Approximation and Projection (UMAP) was employed for dimensionality reduction, enabling visualization of clusters within the five-dimensional alloy space. Because nonlinear projections can distort local relationships, Kernel Density Estimation (KDE) plots were applied alongside UMAP to validate clustering patterns and ensure the fidelity of observed trends.

These concepts and tools form a foundation for navigating the complex design space of compositionally graded alloys. As engineering applications increasingly require tailored, location-specific material performance, CGAs offer a pathway toward multifunctional components, but their potential depends on equally sophisticated data modeling and exploration techniques. When constraints are applied to the alloy design space, multiple feasible subgraphs can emerge, each corresponding to a distinct region of viable compositions. Identifying and analyzing these subgraphs is essential for exploring trade-offs, narrowing design options, and guiding processing decisions. The approach combines labeled property graphs (LPGs) with dimensionality reduction and density-based visualization to identify, interpret, and validate clusters of viable compositions under processing and performance constraints. Using the quinary Nb–Cr–V–W–Zr system as a representative case, we demonstrate how constraint filtering and subgraph analysis can reveal composition paths suitable for additive manufacturing and high-temperature structural applications. The remainder of this paper describes the generation of the computational dataset, the property modeling and graph construction methods, and the subgraph analysis workflow used to evaluate composition–property relationships in support of CGA design.

## 2 Methods

### 2.1 Dataset description

The dataset used in this study was generated through a combination of computational thermodynamic simulations and analytical modeling techniques for the quinary Nb–Cr–V–W–Zr system. These elements were selected based on their relevance to high-entropy alloy (HEA) research and potential for structural applications under extreme thermal and chemical environments.<sup>20</sup> This system was previously investigated by Allen *et al.*,<sup>21</sup> who examined its industrial relevance based on favorable properties identified through CGA analysis. The present simulations map phase stability and property trends across the full compositional space, with emphasis on body-centered cubic (BCC) phase stability and relevant material properties. BCC phase stability is critical because BCC-structured HEAs possess superior high-temperature strength, making them strong candidates for structural use under extreme conditions.<sup>22</sup> Individual contributions of each element to targeted properties were considered during element selection, including Nb for BCC stabilization, V for ductility and solid-solution strengthening, W for high melting point and creep resistance, Cr for oxidation resistance, and Zr for corrosion resistance.<sup>20</sup> The resulting dataset defines a high-dimensional compositional landscape for subsequent graph-based analysis. Compositions within this space were generated by discretizing the Nb–Cr–V–

W–Zr compositional domain into a structured grid of atomic fractions, enabling systematic sampling of the design space for subsequent thermodynamic evaluation. The following sections detail the generation of thermodynamic and mechanical property data for this alloy system.

The graph-based framework employed in this work is not inherently limited to quinary systems and can be extended to higher-dimensional alloy design spaces with additional elements. Each dimension may also represent a predefined alloy (*i.e.*, a multi-component constituent), rather than being restricted to a single element, with the same framework applied. In such cases, nodes represent compositions in higher-dimensional compositional spaces, and edges continue to encode compositional proximity and feasibility under imposed constraints. The primary challenge associated with increasing the number of elements is the combinatorial growth of the design space, which leads to larger graphs and increased computational cost for property evaluation and graph construction. However, this limitation is computational rather than conceptual and can be mitigated through other strategies. As such, the approach remains applicable to more complex multi-component alloy systems, provided appropriate sampling and computational resources are employed.

### 2.2 Material property generation

A targeted set of thermodynamic and mechanical properties was computed to enable detailed subgraph analysis. Property calculations used CALPHAD-based thermodynamic modeling and analytical formulations developed for high-temperature alloys. The computed properties included liquidus and solidus temperatures, equilibrium and Scheil BCC phase fractions, coefficient of thermal expansion (CTE), density, and yield strength evaluated at 1000 °C, Nabarro–Herring creep rate, Pugh ratio,<sup>23</sup> and the Kou hot-cracking criterion.<sup>24</sup> All properties were evaluated consistently across the compositional space to produce a comprehensive dataset used for graph construction, constraint filtering, and visualization.

**2.2.1 Thermo-Calc simulations.** Thermodynamic simulations were performed using Thermo-Calc's TC–Python API, which provides access to the CALPHAD engine through a Python interface. Automated batch processing evaluated thermodynamic properties for approximately 10 000 unique compositions within the CGA design space. Simulations used the TCHEA6 thermodynamic database, developed for high-entropy alloy systems.<sup>25</sup> Key outputs included liquidus and solidus temperatures, coefficient of thermal expansion (CTE), density, equilibrium and Scheil BCC phase fractions, and the Kou hot-cracking criterion.

Liquidus and solidus temperatures were obtained from Thermo-Calc equilibrium calculations, and the solidification range was defined as their difference. Equilibrium phase fractions, including BCC phase fraction, were computed using the Equilibrium Single Point Calculator over a temperature range from 25 °C to 2750 °C, with values extracted at the relevant evaluation temperature for each composition. The coefficient of thermal expansion (CTE) and density at 1000 °C were also



obtained from equilibrium calculations under the same thermodynamic conditions.

Scheil–Gulliver simulations were performed to approximate non-equilibrium solidification behavior, assuming no diffusion in the solid and complete mixing in the liquid. The Scheil BCC phase fraction was obtained from the final solidified state of the Scheil curve for each composition. The Kou hot-cracking criterion was evaluated directly from the Scheil solidification results by extracting the temperature–solid fraction relationship near the end of solidification and computing the derivative defined below.<sup>26</sup> The Kou criterion, derived from the Scheil solidification curve, is defined as

$$\left| \frac{dT}{d(f_s)^{1/2}} \right|,$$

the derivative of temperature with respect to the square root of the solid fraction near the end of solidification.<sup>24,27</sup> These CALPHAD outputs formed the basis of the CGA database, with each entry representing a unique composition and its associated thermodynamic properties.

While raw Thermo-Calc outputs and proprietary databases cannot be redistributed due to licensing restrictions, the simulation procedures and property definitions described here are sufficient to reproduce the thermodynamic calculations using equivalent Thermo-Calc configurations. In addition, the Zenodo repository includes a metadata file describing each variable in the dataset, including definitions and the methods used for property evaluation, to facilitate independent interpretation and reuse of the data.

**2.2.2 Pugh ratio.** The Pugh ratio was selected as an approximate indicator of ductility and is defined as the ratio of the bulk to shear modulus,  $B/G$ , of an alloy.<sup>23</sup> It captures the balance between an alloy's tendencies to fracture or plastically deform. Bulk and shear moduli were estimated using rule-of-mixtures calculations based on the elastic moduli of the pure elements. This metric has been shown to correlate with fracture deformation behavior in compression tests.<sup>28</sup> As a first-order approximation, the Pugh ratio provides a computationally efficient descriptor of ductility that can be readily incorporated into high-throughput alloy design workflows.

**2.2.3 Yield strength model.** Yield strength was estimated using the Maresca–Curtin strengthening model,<sup>29</sup> which provides a reliable lower-bound estimate of the tensile strength for single-phase body-centered cubic (BCC) alloys.<sup>30,31</sup> The model expresses the temperature – and strain rate – dependent shear strength,  $\tau_y$ , as

$$\tau_y(T, \dot{\epsilon}) = \tau_{y0} \exp \left[ -\frac{1}{0.55} \left( \frac{kT}{\Delta E_b} \ln \frac{\dot{\epsilon}_0}{\dot{\epsilon}} \right)^{0.91} \right], \quad (1)$$

where  $\tau_{y0}$  is the zero-temperature shear stress,  $k$  is Boltzmann's constant,  $T$  the absolute temperature, and  $\Delta E_b$  the energy barrier for dislocation motion. The applied strain rate,  $\dot{\epsilon}$ , was set to  $10^{-3} \text{ s}^{-1}$ , and the reference strain rate,  $\dot{\epsilon}_0$ , to  $10^4 \text{ s}^{-1}$ . Further details regarding parameter selection are provided in ref. 32. The yield strength,  $\sigma_y$ , was then obtained as

$$\sigma_y(T, \dot{\epsilon}) = M\tau_y(T, \dot{\epsilon}), \quad (2)$$

where a Taylor factor of  $M = 3$  (ref. 29) was applied for the BCC compositions examined in this study. The Maresca–Curtin model assumes homogeneous single-phase microstructures and relies on material-specific fitting parameters reported in previous studies.<sup>21,29</sup> Despite these assumptions, it provides a computationally tractable approach for estimating yield strength across large compositional spaces. The calculated yield strengths were incorporated into the labeled property graph database to support subgraph analysis.

**2.2.4 Creep model.** Creep behavior was estimated using the Nabarro–Herring model, assuming lattice diffusion as the dominant creep mechanism. The strain rate,  $\dot{\epsilon}_{\text{NH}}$ , is given by eqn (3), where  $A_n$  is a dimensionless constant,  $D$  the intrinsic diffusivity,  $b$  the Burgers vector,  $\sigma$  the applied stress,  $d$  the grain size,  $k$  Boltzmann's constant, and  $T$  the absolute temperature.<sup>33</sup> Although the dominant creep mechanism may vary with composition, stress, and temperature, this model provides a useful first-order, mechanistically grounded baseline for assessing diffusional creep resistance.<sup>34</sup>

$$\dot{\epsilon}_{\text{NH}} = A_n \frac{Db^3\sigma}{d^2kT} \quad (3)$$

For this study,  $A_n$  was set to 50,  $\sigma$  to 200 MPa, and the grain size to 200  $\mu\text{m}$ . These parameters were held constant across the alloy design space to enable rank-order comparison of diffusional creep resistance. In future CGA design efforts, these predictions can be refined as experimental data become available. Once a subgraph is selected for CGA design, higher fidelity but lower throughput methods can be applied for property prediction within that region. The diffusivity,  $D$ , was obtained from Thermo-Calc's MOBHEA3 database for each element in the composition of interest.<sup>35</sup> Both the diffusing and gradient elements were defined as the element of interest, and the reference element was chosen as the atom with the largest radius, or the second largest if the diffusing species already occupied that role. The Burgers vector,  $b$ , was estimated from the rule-of-mixtures average of the elemental atomic volumes.

All property models used in this study are based on established thermodynamic, empirical, or analytical formulations that have been validated in prior literature. These models are used here to enable consistent, high-throughput evaluation of compositional trends across the design space.

## 2.3 Dataset preparation and integration

Thermodynamic and mechanical datasets were merged into a unified format capturing compositional, phase, and mechanical property data. Data processing was performed using Python's Pandas library to enable efficient manipulation and cleaning.<sup>36</sup> The unified dataset contained tens of thousands of alloy compositions, structured to facilitate subsequent analyses. This dataset was then transformed into a graph-based representation using NetworkX.<sup>37</sup> Nodes represented individual alloy compositions with associated property attributes, and



edges connected compositions with compositional similarity. Exploratory visualization of property distributions was performed using Matplotlib and Seaborn. Uniform Manifold Approximation and Projection (UMAP) was subsequently applied to reduce the high-dimensional dataset to two dimensions, enabling visual identification of compositional trends.

All data processing and analysis were performed using Python within a Jupyter Notebook environment. The primary libraries used include NumPy and Pandas for data handling, NetworkX for graph construction and analysis, and Matplotlib and Seaborn for visualization. Additional utilities such as PIL were used for image handling where needed. The dataset was loaded from a structured Excel file, and graph connectivity was defined using precomputed node and neighbor arrays stored in NumPy (.npy) format, which encode compositional grid points and their adjacency relationships in the design space. These arrays were used to construct the labeled property graph.

Constraint-based filtering and subgraph extraction were performed using NetworkX graph operations. All visualizations of compositional trends, subgraphs, and property distributions were generated using Matplotlib and Seaborn. The full implementation, including code, dependencies, and a reproducible workflow, is provided in a public GitHub repository, while the dataset is hosted in a Zenodo repository.

#### 2.4 Design space filtering and constraints

In the graph-based representation of the alloy design space, the unconstrained composition set was modeled as a grid of nodes, each connected to adjacent nodes by edges. When constraints were applied, edges associated with nodes that failed to meet the defined criteria were removed. The resulting graph contained connected groups of nodes representing compositions that satisfied all constraints, forming subgraphs of feasible compositions. Depending on the constraint type, feasible compositions could occur within a single continuous subgraph or be divided among several disconnected subgraphs. These subgraphs inform CGA design by ensuring that the selected gradient endpoints belong to the same connected subgraph, allowing traversal without crossing infeasible regions of the alloy space.

To identify alloy compositions with favorable solidification behavior and stable microstructures, three filtering criteria were applied: solidification range, equilibrium BCC phase fraction, and Scheil BCC phase fraction. The solidification range—defined as the temperature difference between the liquidus and solidus—was constrained to  $\leq 50$  °C. This constraint favored compositions with narrow freezing intervals, reducing susceptibility to microsegregation, porosity, and hot-cracking—defects that are particularly detrimental during rapid solidification in additive manufacturing. A second filter required an equilibrium BCC phase fraction of  $\geq 0.999$  at the target temperature. This ensured compatibility with the Maresca–Curtin strength model,<sup>29</sup> yielding conservative and reliable mechanical predictions. Nearly single-phase BCC structures simplify modeling and improve mechanical stability, especially at elevated temperatures.

To capture solidification behavior representative of manufacturing conditions, Scheil–Gulliver simulations were performed. Compositions were retained only when their Scheil BCC phase fraction was  $\geq 0.999$ . These candidates were further screened to ensure an equilibrium BCC phase fraction of  $\geq 0.999$  between 1273 °C and their respective Scheil-predicted solidus temperatures. This dual-filter approach ensured that selected alloys maintained BCC-dominant microstructures under both equilibrium and rapid solidification conditions, enhancing their suitability for practical processing and application.

#### 2.5 Visualization analysis

After constraint-based filtering, Uniform Manifold Approximation and Projection (UMAP) was used to visualize compositional relationships in two dimensions. UMAP is a nonlinear dimensionality reduction technique, particularly suited for high-dimensional datasets such as compositional design spaces. For this analysis, input features consisted of normalized elemental compositions, ensuring that each alloy's chemical makeup was represented proportionally and comparably within the manifold structure. UMAP embeddings enabled the identification of local and global patterns in composition space, including clusters of similar alloys, continuous gradients of property variation, and isolated subspaces with unique characteristics. This was particularly useful for tracking how constraints shaped the topography of the design space and for visually revealing regions with potential for further exploration.

UMAP was used as a qualitative visualization tool with the objective to capture general structural trends in the compositional design space (*e.g.*, clustering and relative positioning of alloys). It is also noted that UMAP embeddings are sensitive to the dimensionality and discretization of the composition space, requiring retuning for different alloys systems. For this reason, more recent work has explored affine projection methods as a more consistent alternative for compositional visualization, as described by Vela *et al.*<sup>19</sup>

To further validate the significance of the visual clusters observed in UMAP projections, Kernel Density Estimation (KDE) plots were generated in parallel with the UMAP space. KDE is a statistical method used to estimate the probability density function of a dataset, allowing for the detection of high density regions where feasible compositions concentrate. By integrating both KDE plots with UMAPs, it was possible to assess whether visually identified clusters were true or artifacts from the dimensionality reduction process. Together, the combined use of UMAP and KDE offered a powerful framework for visual exploration of the compositional design space.

**2.5.1 UMAPs of pure elements.** Individual UMAP plots were generated to analyze how each of the five elements, Nb, Cr, V, W, and Zr, is distributed within the reduced two-dimensional compositional space (Fig. 1). These visualizations revealed a pattern unique to each element, highlighting regions of concentration and compositional gradients. By establishing this geometric baseline understanding of elemental relationships and distributions, it became possible to identify and



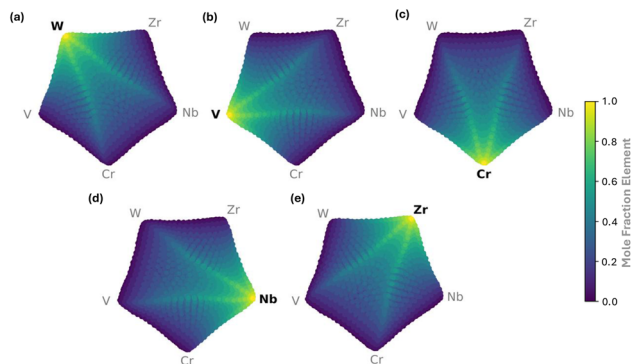


Fig. 1 Individual UMAPs showing atomic fraction distributions of Cr, Nb, V, W, and Zr within the compositional space.

interpret how subsequent constraints influence the compositional structure, ultimately allowing for targeted alloy selection within the refined space. These plots provide a clear view of how each element is spread across the design space and serve as a reference when applying filters or constraints later in the analysis.

## 3 Results

### 3.1 Compositional filtering strategy

Exploring an unconstrained, high-dimensional compositional space enables the discovery of new alloy chemistries. However, this same design freedom introduces challenges in rendering large datasets interpretable and extracting meaningful trends. Without targeted filtering, the design space contains numerous compositions that may be theoretically interesting but are redundant, impractical to manufacture, or unlikely to exhibit desirable properties. Many regions include alloys prone to poor solidification behavior, formation of deleterious phases, or unfavorable mechanical properties, underscoring the need for rapid identification and filtering. To make exploration tractable, filters must be selected to prioritize manufacturability and structural integrity across the compositional space. The filtering strategy employed in this work isolates compositions with favorable solidification characteristics and stable microstructures suitable for high-temperature service and additive manufacturing. It is also important to note that uncertainties in CALPHAD predictions and empirical property models may influence the precise location of constraint boundaries; therefore, conservative threshold values were selected to reduce the inclusion of compositions near feasibility limits and improve confidence in the resulting subgraphs.

### 3.2 Solidification range

The solidification range, or the difference between the equilibrium liquidus and solidus temperatures, plays a key role in determining how uniformly an alloy solidifies. Narrowing this range can aid in reducing solidification-related defects such as hot-cracking and microsegregation and improving microstructural consistency. In this work, a threshold of  $\geq 50$  °C was

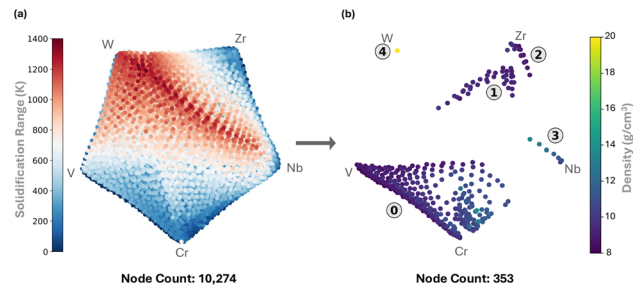


Fig. 2 UMAP projections of the compositional design space under an equilibrium solidification range constraint. (a) All nodes prior to filtering with color bar denoting solidification range of nodes. (b) Nodes remaining after applying a solidification range filter leaving nodes with a solidification range of  $\geq 50$  °C. These remaining nodes were colored by their respective density to highlight specific property variation across the filtered space.

selected consistent with prior CGA design studies, where similar limits have been used as manufacturability constraints to mitigate hot-cracking susceptibility and chemical inhomogeneity during additive manufacturing during additive manufacturing.<sup>3</sup>

The entire database was first displayed in a UMAP (Fig. 2a) which shows groups of compositions with small solidification ranges and others as high as 1400 °C. For the first filtering constraint, the nodes that had a solidification range greater than or equal to 50 °C were removed to focus on the most promising subset of nodes. This reduced the 10 274 total nodes from the database down to 352 alloy compositions that satisfied the criterion. A lower solidification range reduces the likelihood of solidification-related defects such as hot-cracking or other microstructural inconsistencies which are key characteristics in additive manufacturing or production of alloys in industry. The filtering revealed five individual subgraphs made up of closely related families of compositions. Each of the five subgraphs is displayed in a UMAP (Fig. 2b) with the color scale now representing the equilibrium density of the node composition evaluated at 1000 °C. This temperature was chosen for its representation of the extreme environments often encountered in industrial applications. Density was chosen as an arbitrary property that is often prioritized in alloy design, but any property of interest can be selected depending on the design scenario.

Of the five subgraphs identified, four were considered valid; the fifth consisted of a single isolated node. The separated, traversable subgraphs allowed detailed analysis of the most relevant compositions. Subgraph 0 contained the largest number of traversable nodes (293). This substantial node count enabled multiple traversal paths and composition gradient options depending on application needs. For example, a gradient within subgraph 0 can follow a path from high Cr to high V (path 1 in Fig. 3a). An alternative, lower density path (orange) connects a V-rich region to a more compositionally balanced alloy. Multiple traversal routes are possible within subgraph 0 provided the path remains confined to these 293 feasible nodes. Including nodes outside this subgraph would



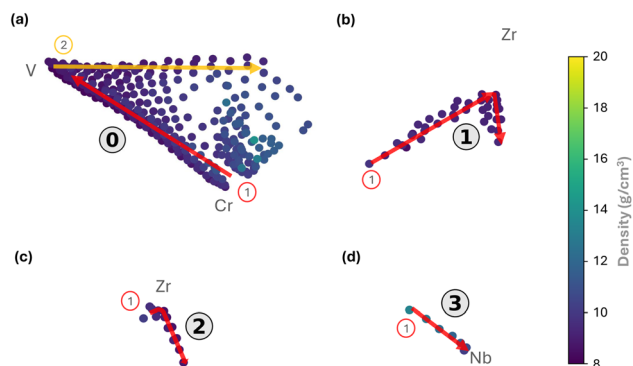


Fig. 3 Subgraphs (a–d) extracted from the filtered design space with solidification ranges below 50 °C. Each UMAP subgraph represents a distinct group of compositionally connected nodes. Highlighted red and orange paths indicate example traversal routes through the subgraphs, illustrating potential compositional transitions for CGA design.

require crossing infeasible compositions. Combining filtering with density-coded UMAPs offers an effective means to visualize the composition space and support subgraph selection for graded material design. Aside from subgraph 0, the remaining subgraphs contained fewer nodes and correspondingly fewer candidate FGM paths. Subgraph 1 contained 37 nodes, subgraph 2 contained 14, and subgraph 3 contained 8. Representative paths from the smaller subgraphs are shown in Fig. 3b–d in red.

### 3.3 Equilibrium BCC phase fraction constraint

Solidification range is not the only constraint that can be applied to this system and to CGA design as a whole. Another

important screening parameter to consider when choosing an alloy composition is the system's equilibrium phase fractions. Typically, the BCC phase being present and dominant is necessary for HEAs/CGAs because it offers high strength, good high-temperature stability, and favorable manufacturability. Alloys of a multi-element system like this may form multiple phases at equilibrium conditions, including undesirable phases. In high-temperature applications, a stable BCC structure with no formation of potentially deleterious secondary phases is an important indication of a promising alloy composition.

The full dataset was used to generate a new set of UMAPs (Fig. 4a–d) each displaying all 10 274 nodes with the color bar representing their BCC phase fraction. The dark purple nodes represent alloys that are nearly fully BCC, while the white and green nodes correspond to alloys in which the BCC phase is not predominant. To more closely mimic a variety of real-world applications, multiple temperatures were queried. Most high-temperature industrial parts are expected to service a wide range of temperatures over their lifetimes, so BCC phase stability across a broad temperature range was established as a filtering criterion.

The equilibrium BCC filtration took the original database and removed all edges between nodes with less than 0.999 BCC mole fraction. This ensures almost entirely single-phase BCC stability at a chosen temperature. Twelve temperatures were chosen incrementally from 25 °C to 2750 °C and evaluated to show their BCC phase fraction present. Of these twelve temperatures, representative examples were chosen to plot filtered UMAPs (Fig. 4e–h). The UMAPs are plotted with coloring once again based on the density at 1000 °C.

Each of these filtered UMAPs produced only a single subgraph, meaning all the nodes that passed the filtering

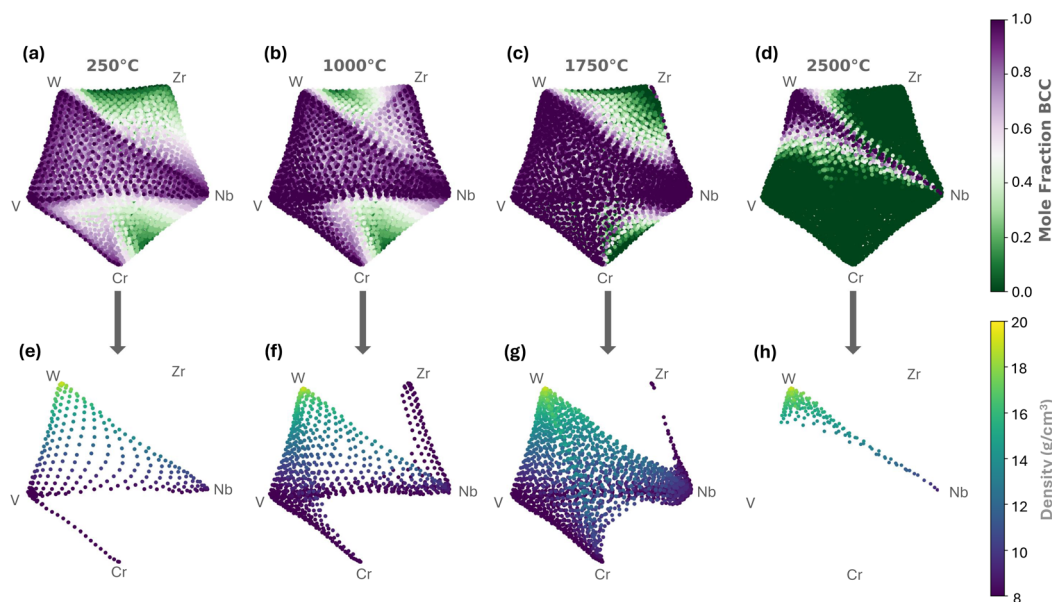


Fig. 4 UMAP projections (a–h) of the design space before and after BCC phase filtering at 250 °C, 1000 °C, 1750 °C, and 2500 °C. (a–d) show UMAPs with nodes colored by BCC phase fraction. (e–h) show nodes remaining after applying a BCC filtering constraint (BCC fraction  $\geq 0.999$ ), colored by density to highlight property distribution after filtering.



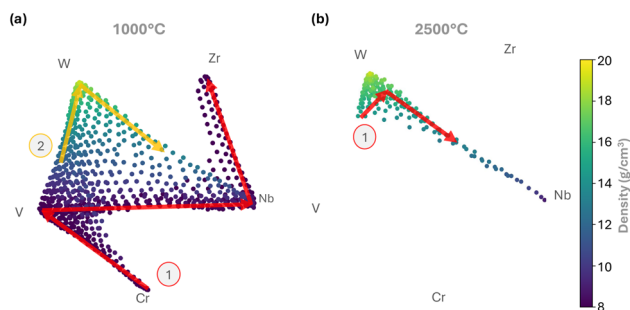


Fig. 5 UMAP projections of the filtered design space (BCC fraction  $\geq 0.999$ ) at (a) 1000 °C and (b) 2500 °C, with nodes colored by density. In (a), two path traversal options (red and orange) highlight possible CGA design routes. In (b), a single red path illustrates a potential design path at elevated temperature.

criterion were linked through compositional similarity. With only a single subgraph at each temperature, there is much more flexibility to choose compositional pathways than there was when filtering by solidification range. For example, at 1000 °C, one can choose a path starting from a predominantly Cr region to a predominantly V region, moving into a high Nb concentration, and finishing with a high Zr concentration (Fig. 5a – red path). This path was chosen to demonstrate a CGA choice with multiple composition steps all following the low density region of the subgraph. The orange path in Fig. 5a represents a shorter step traversal within the same temperature subgraph. Conversely, a shorter two-step path can be chosen in a higher density region at 2500 °C to identify alloys that can withstand particularly extreme temperatures. This path starts at a mid-range composition of V and W before moving into a high W composition to a middle region with a balance of W, Zr, and Nb. This is depicted by the red path in Fig. 5b.

The temperature selected for equilibrium BCC filtering plays a significant role in choosing an appropriate composition set for CGA design. Fig. 6 is a graphical representation of how node count increases and decreases with varying degrees of temperature. There is an apparent peak of BCC nodes greater than

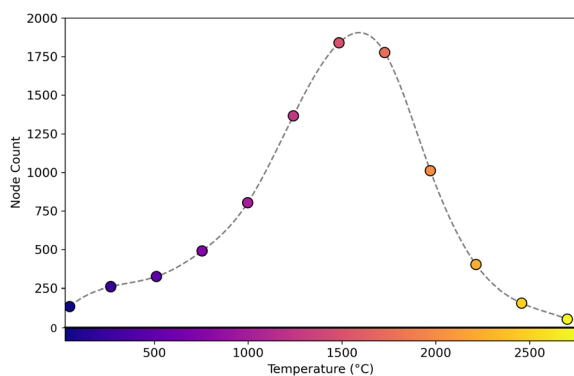


Fig. 6 Node count versus temperature for BCC phase filtering threshold (BCC fraction  $\geq 0.999$ ). The plot shows the number of retained nodes changing with temperature. The node count peaks at 1700 °C then decreases, highlighting the effect of temperature on BCC phase fraction.

0.999 phase fraction at about 1700 °C. This is caused by temperature-driven solid-solution effects. As temperature rises, entropy plays a greater role in phase stability, favoring high-entropy phases like the disordered BCC solid-solution. At very high temperatures, however, the number of stable compositions drops as melting begins for a portion of the alloy space. To design alloys for very high service temperatures up to 2500 °C, one must consider that the choice of composition nodes significantly decreases from 9490 nodes (at 1750 °C) to 5029. Alternatively, if the desired service temperature is limited to below 1000 °C, a different reduced set of nodes is available.

### 3.4 Scheil BCC phase fraction constraint

**3.4.1 Solidification behavior.** Equilibrium phase fraction is a relevant property to assess the suitability of alloys for prolonged exposure to elevated temperatures, but manufacturing processes including additive manufacturing feature rapid heating, cooling, and solidification that can form alloy microstructures out of equilibrium. Scheil solidification simulations are a common tool to model these processes. In Scheil simulations, the solid-liquid interface remains at local equilibrium and the liquid is assumed to be well mixed, but solid-phase diffusion is taken to be negligible, so segregation of elements in the solid is possible. Scheil simulations can be used to produce more realistic estimates of phase fractions that result from non-equilibrium manufacturing processes.

To maintain consistency with the previous equilibrium phase fraction filtering, any node at each temperature step that had less than 0.999 mole fraction BCC was eliminated from the database. From this filtration, a UMAP was produced to visually display the results of all nodes that passed filtration (Fig. 7a). At first glance, the projection seems to show that all the nodes have remained in one large subgraph zigzagging through the space. However, the Scheil filtering actually separated the nodes into four individual traversable subgraphs (Fig. 7b). This is a crucial step in the querying process that demonstrates the importance of the methods discussed in this work for identification and selection of subgraphs for CGA design. This particular constraint underscores the point that UMAPs alone cannot

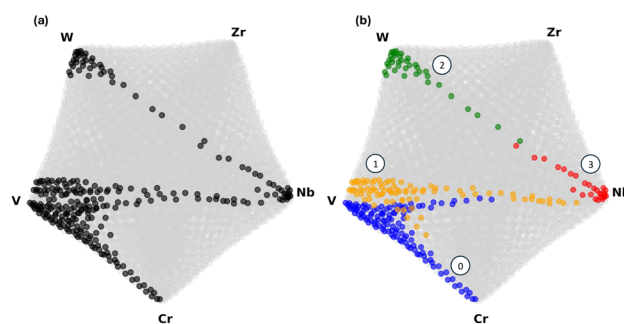


Fig. 7 UMAP projections of the design space after applying a Scheil solidification-based BCC phase fraction constraint (BCC  $\geq 0.999$ ). (a) shows nodes that remain after filtering at each temperature step using Scheil simulation results. (b) displays four resulting individual subgraphs, each colored uniquely.



always discern between disconnected regions of the feasible space and that graph-based representations are necessary to fully identify individual and disconnected subgraphs. Without this step, the infeasible regions between subgraphs that cannot be traversed would not be known and, if chosen, would possibly cause secondary phases to form during manufacturing.

Four KDE plots—one for each subgraph—were generated with the node count density of the element on the y-axis and their atomic fraction on the x. Subgraph 0, comprised of 118 nodes, has minimal W in the majority of nodes. The three subsequent elements, Nb, Cr, and V, are all spread well into the higher atomic fractions meaning some alloy compositions in subgraph 0 contain high amounts of V, Cr, and Nb in that order. Subgraph 1, depicted in orange in the UMAP (Fig. 7a), is V-rich similar to subgraph 0. The next greatest constituent in subgraph 1 would be Nb, compared to Cr in subgraph 0. Subgraph 2 is W-rich, with Nb and to a lesser extent, Cr and V are also present. Lastly, subgraph 3, which is shown by the red nodes in the UMAP, is predominantly Nb-rich with small amounts of Zr and W. Analyzing the subgraphs, based on the relationship between the UMAPs and subplots, we observe that the distribution of UMAP points is consistent with the statistical representation in the KDE plots (Fig. 8).

Understanding the compositional breakdown for each subgraph is a starting point to design a tailored CGA for a specific application, but how do we know if these alloys are useful for part design? The Scheil-filtered subgraphs can be further broken down by important material property parameters of each subgraph. This workflow allows a staged approach where an alloy can be designed by first filtering absolute requirements and subsequently prioritizing one or more properties of interest while selecting a subgraph for CGA design.

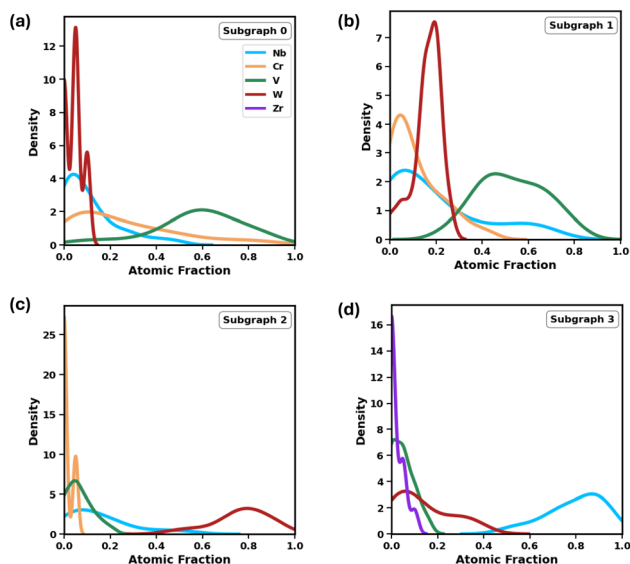


Fig. 8 KDE plots (a–d) showing the elemental composition distributions for each of the four subgraphs produced after Scheil BCC phase fraction filtering ( $BCC \geq 0.999$ ). Each plot illustrates the relative spread of components of a subgraph in relation to node density.

To further assess solidification behavior under non-equilibrium conditions, the Kou criterion, a derivative of Scheil solidification simulations,<sup>24</sup> was applied. This metric provides an estimate of an alloy's susceptibility to hot-cracking during solidification. A higher Kou value generally indicates greater susceptibility to cracking, while lower values suggest more favorable solidification behavior.

Each of the four subgraphs was taken from the large UMAP (Fig. 7b) and isolated into individual UMAPs (Fig. 9a–d). The color bars to the right of the UMAPs display Kou hot-cracking criterion values with the darker colors denoting alloys more prone to cracking upon solidification. Paired with the UMAPs is a KDE plot which displays density distributions comparing the subgraph property value distributions (Fig. 9e).

Subgraph 0 exhibits the lowest Kou criterion values, suggesting it contains compositions with the least susceptibility to solidification cracking. In contrast, subgraphs 2 and 3 display elevated values, indicating a higher likelihood of crack formation during solidification. Within each subgraph, however, there are traversable regions that maintain low Kou values,

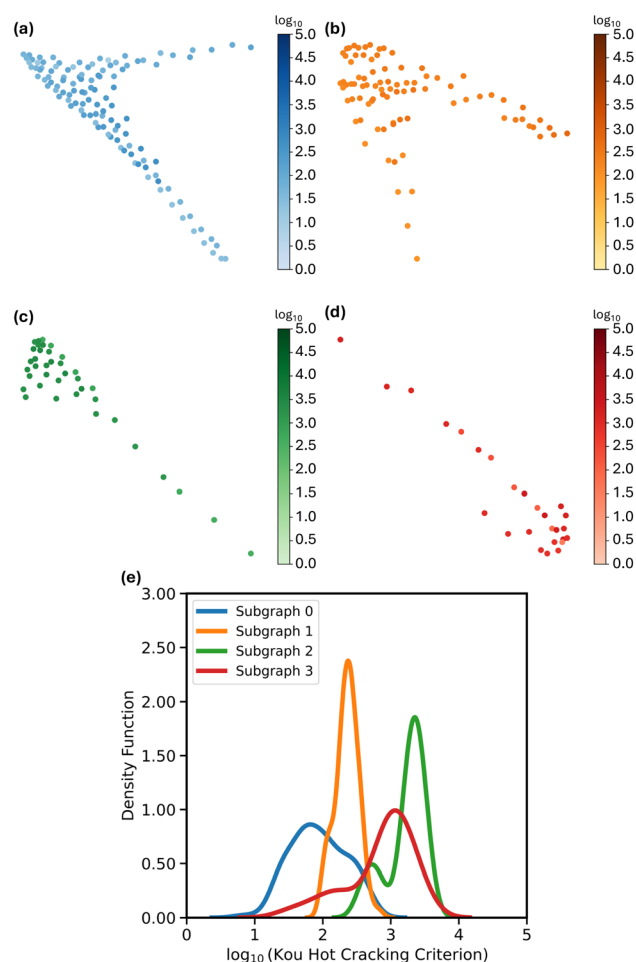


Fig. 9 Figures (a–d) show UMAP projections of the four subgraphs resulting from the Scheil BCC phase fraction constraint ( $BCC \geq 0.999$ ), with nodes colored by their Kou hot-cracking criterion values. (e) demonstrates corresponding KDE plot of the property value distribution across the subgraphs in relation to node density.



which is an important consideration when designing a CGA to minimize defect risk during processing. A subgraph with a wide range of Kou values is not necessarily undesirable, however. A wider range simply requires careful path planning to ensure the Kou criterion is incorporated into the path cost so that low risk regions are prioritized.

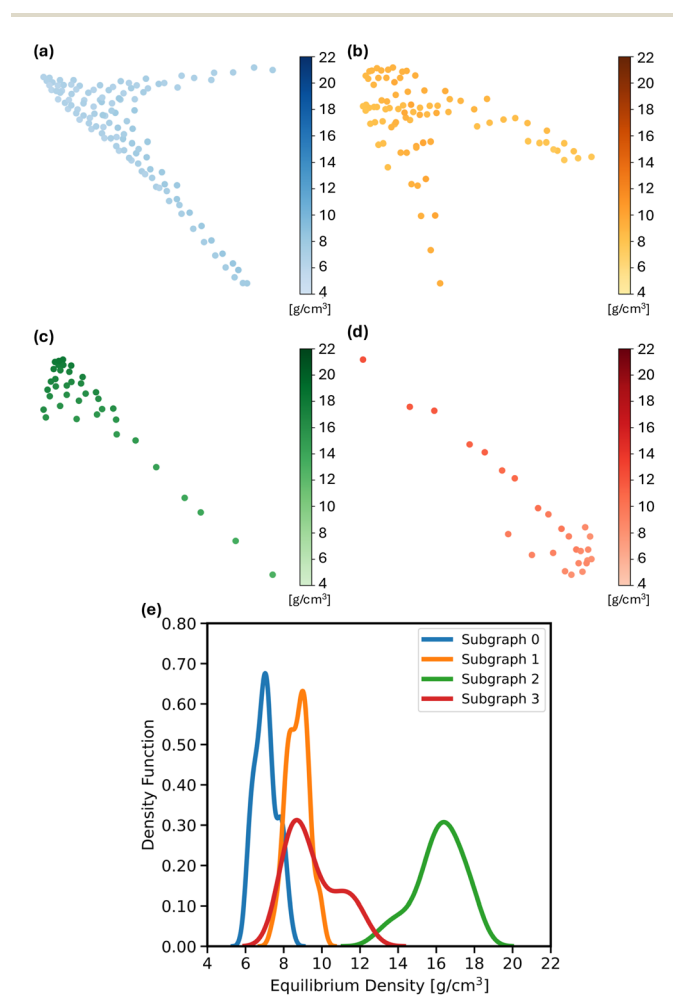
**3.4.2 Equilibrium density.** In addition to Kou hot-cracking criterion, density was evaluated across the composition space to assess its implications for weight and performance. Density remains a key parameter in applications where weight reduction is critical or where discrepancies in density could introduce stress.

Subgraph 0 (Fig. 10a) presents a narrow spread and the lowest density values, suggesting a promising candidate for a lightweight application. Subgraph 1 (Fig. 10b) has a similar distribution centered on density values around  $8.5 \text{ g cm}^{-3}$ . Subgraphs 2 and 3 (Fig. 10c and d) have a wider range of densities, which could be advantageous in applications where the part needs to have a graded weight distribution. Graded density profiles, such as those in subgraphs 2 and 3, could

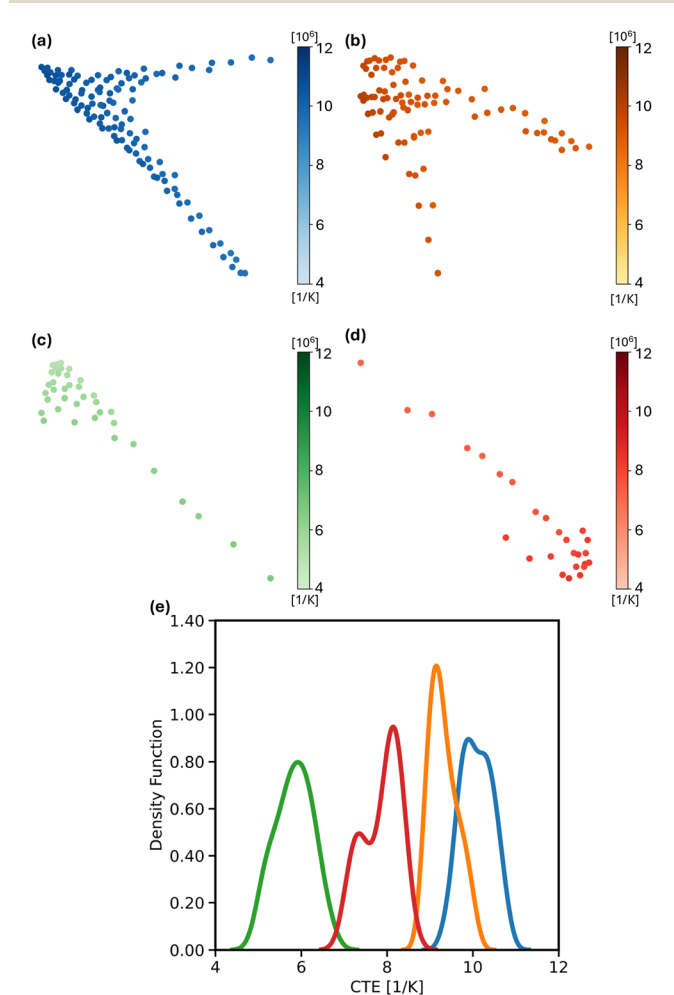
provide lightweight sections that reduce overall part mass while incorporating denser regions that deliver location-specific strength where needed.

**3.4.3 Coefficient of thermal expansion.** In a system with a variety of materials, CTE must be closely matched to ensure compatibility, so it is often a factor in alloy selection. CTE variability is important in CGA design because a larger discrepancy between CTE values can lead to internal stresses and dimensional instability during thermal cycling.

The UMAPs and KDE plot show a step in CTE values between each of the four subgraphs. Each subgraph shows a similar dispersion between CTE values, but the distributions are shifted to center around different values (Fig. 11a). The relatively narrow distribution for each subgraph suggests that internal CTE mismatch would not be a significant design concern for these compositions, and CTE would be more relevant for compatibility with other subgraphs in a system. Both density and CTE are important alloy selection criteria because they determine weight and dimensional stability for a final part.



**Fig. 10** Figures (a–d) show UMAP projections of the four subgraphs resulting from the Scheil BCC phase fraction constraint ( $\text{BCC} \geq 0.999$ ), with nodes colored by their equilibrium density values. Figure (e) demonstrates corresponding KDE plot of the property value distribution across the subgraphs in relation to node density.



**Fig. 11** Figures (a–d) show UMAP projections of the four subgraphs resulting from the Scheil BCC phase fraction constraint ( $\text{BCC} \geq 0.999$ ), with nodes colored by their coefficient of thermal expansion values. (e) demonstrates corresponding KDE plot of the property value distribution across the subgraphs in relation to node density.



Meanwhile, the framework here additionally supports the visualization of properties that play an important role in screening compositions based on part performance rather than manufacturability and compatibility.

**3.4.4 Mechanical property analysis.** Yield strength (YS) is a vital benchmark for determining the upper limit of safety and effectiveness of a material which is important in material selection. In engineering design, yield strength is one of the key material parameters. It is immediately apparent from the KDE plot in Fig. 12 that subgraph 3 (Fig. 12c) shows the lowest YS values with a peak at around 100 MPa. The lower YS values are displayed in a lighter red color in subgraph 3's UMAP with increasing values towards the center of the composition space reaching up to 460 MPa. This behavior is consistent with the Maresca–Curtin model's focus on solid-solution strengthening and the fact that increasing complexity in composition is known to contribute to increased strength through this mechanism.<sup>29</sup>

The two highest strength subgraphs, 1 and 2 (Fig. 12a and b), share a similar distribution with a much wider shape than the previous two subgraphs. Since they are such wide distributions,

however, path selection choice plays a much more prominent role. This variability could be an advantage or a disadvantage depending on the application. For example, in the design process it may be beneficial to retain high yield strength in one region while deliberately relaxing it in another to allow for the optimization of other properties such as lower density in that region. The broad distributions of subgraphs 1 and 2 allow for more flexibility in tailoring local properties along a composition path.

Creep rate is a material parameter that indicates how quickly a material deforms over time under a constant load and temperature. Depending on the rate, it can be the limiting factor for part lifetime at elevated temperatures determining when a part will fail or no longer meet geometric requirements. In almost all structural applications, the lowest possible creep rate is ideal. The UMAPs (Fig. 13a–d) show less of a gradual color gradient, but sharper contrasts of property node color throughout. This is mirrored in the KDE plot as well—each subgraph shows a bimodal distribution. Stark differences in

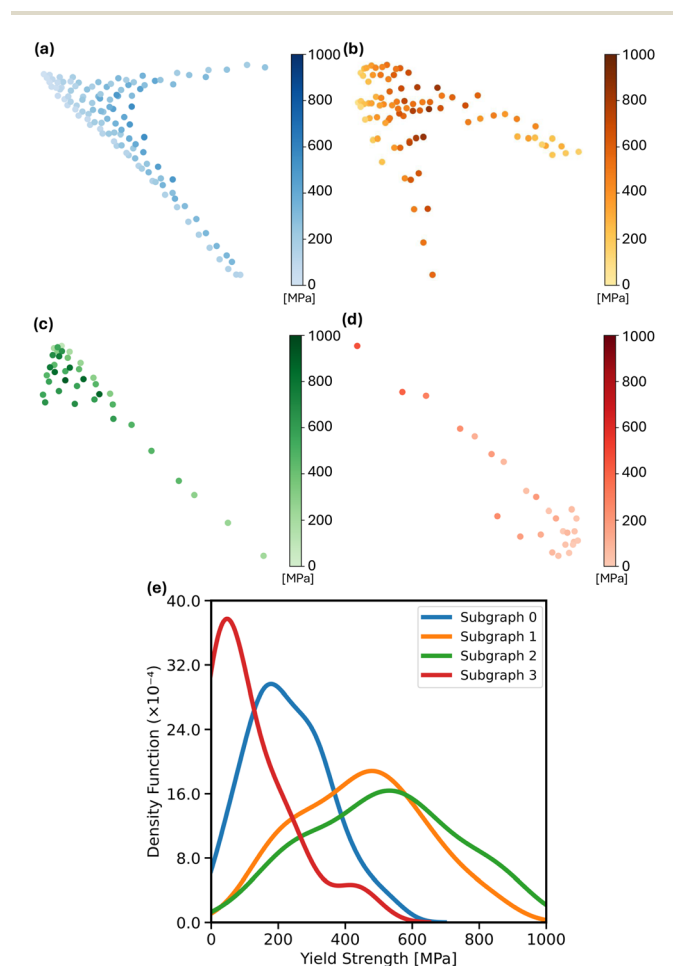


Fig. 12 (a–d) show UMAP projections of the four subgraphs resulting from the Scheil BCC phase fraction constraint ( $BCC \geq 0.999$ ), with nodes colored by their yield strength values. (e) demonstrates corresponding KDE plot of the property value distribution across the subgraphs in relation to node density.

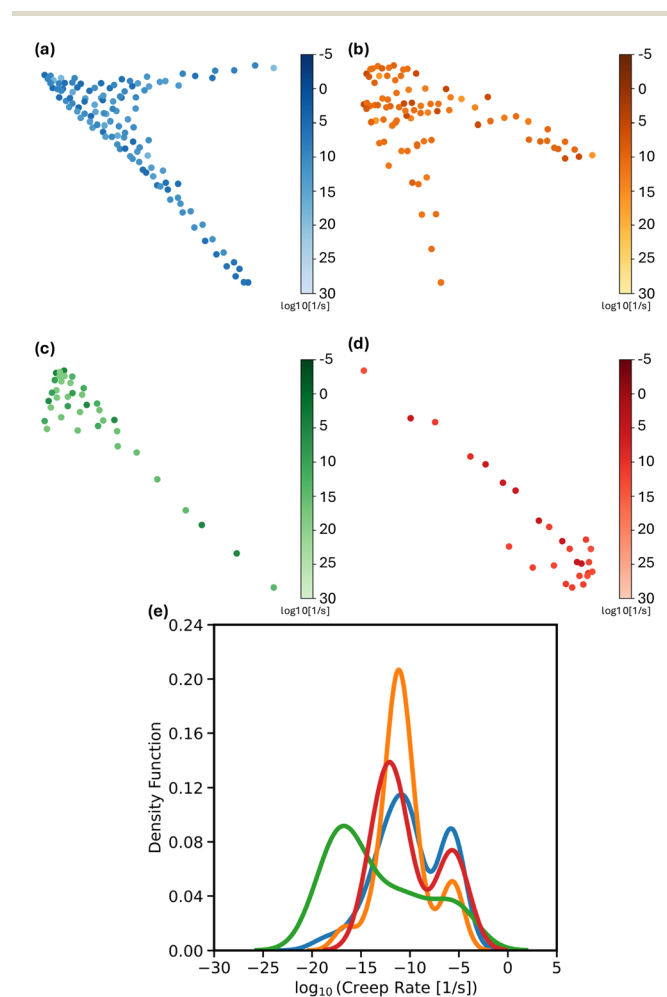


Fig. 13 (a–d) show UMAP projections of the four subgraphs resulting from the Scheil BCC phase fraction constraint ( $BCC \geq 0.999$ ), with nodes colored by their Nabarro–Herring creep values. (e) demonstrates corresponding KDE plot of the property value distribution across the subgraphs in relation to node density.



creep values among themselves can produce neighboring regions with sharply different properties. Having one region of high creep rate so close to a region of low creep rate will lead to mechanical service instability and potential internal stress accumulation.

For the most part, all the subgraphs contain a similar range of creep rates, meaning that selection of an appropriate subgraph on its own may not mitigate risks of a CGA with undesirable creep rates. Path planning and proper traversal through the nodes of the subgraphs could be one way to combat this issue. Subgraph 0 (Fig. 13a) has the highest node count and therefore the greatest opportunity of the four to avoid unwanted instabilities. The paths may be short, but the number of steps that can be taken is not limited to just one or two direction changes and since each node is feasible within this subgraph, paths of similar creep rates can be found and combined. Overall, subgraph 2 (Fig. 13c) has the widest spread and the largest maximum creep rate, which is undesirable. However, its peaks are the least prominent, meaning the gradient between nodes may be more gradual and easier to traverse. A lower creep rate is favorable when designing an alloy system, and in this case, subgraph 2 exhibits the lowest overall values. This makes it the most promising candidate in terms of creep resistance, despite its wider spread and higher maximum values.

The last material property analyzed for this dataset was Pugh ratio. Pugh ratio, as detailed in the methods, is the ratio of a material's bulk to shear modulus. This ratio tells us whether the alloy will be ductile or brittle based on the value. The higher the Pugh ratio, generally the more ductile the alloy which is particularly desirable when designing to avoid cracking or failure during service. Similar to the previous properties, a wide spread will indicate a larger discrepancy in values between the nodes in the subgraph. A wide range of Pugh ratio values may not necessarily be a bad thing. For example, a part being designed can be chosen to maximize ductility in one region and allow a less ductile region elsewhere in favor of optimizing a more important property in that location.

Subgraph 2 (Fig. 14c) stands out with its sharp and narrow distribution centered on a small range of values. This W-rich subgraph displays a mid-to-low range Pugh ratio, centered around 2.25, alluding to a potentially brittle alloy. The Pugh ratio has a threshold value of 1.75—anything above this value tends to be ductile and numbers below are typically more brittle.<sup>38</sup> There are minimal node values in subgraph 2 that circle this 1.75 ductile–brittle threshold, and during path traversal, values lower than this should be avoided if this subgraph is selected for CGA design. This precedent should be applied to each of the four subgraphs except subgraph 3 which never reaches a Pugh ratio value below 2.2. Subgraph 3 objectively has the best Pugh ratio values since it has the highest density of nodes above the ductile–brittle threshold which promotes more opportunities in choosing a ductile region during CGA design.

**3.4.5 Property trade-off analysis.** UMAPs and KDE plots provide valuable visual insight into the behavior of each subgraph. With many competing properties, trade-offs are inevitable, and prioritizing one or two key attributes is often

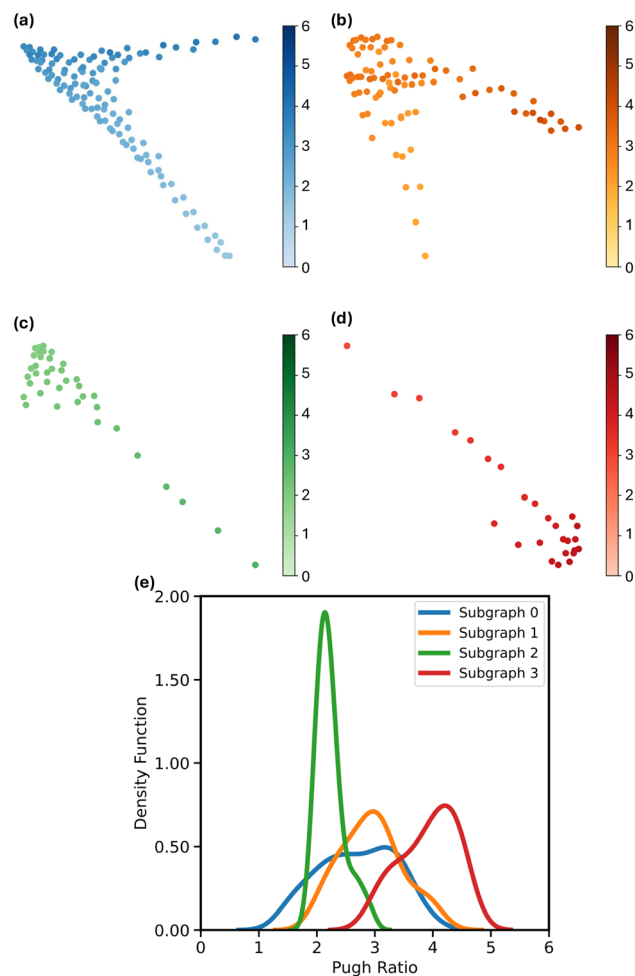


Fig. 14 (a–d) show UMAP projections of the four subgraphs resulting from the Scheil BCC phase fraction constraint ( $BCC \geq 0.999$ ), with nodes colored by Pugh ratio values. (e) demonstrates corresponding KDE plot of the property value distribution across the subgraphs in relation to node density.

more practical than attempting to maximize every metric. Radar plots offer a clear way to compare how each subgraph ranks across multiple properties, helping to guide targeted design decisions. While these plots do not indicate which properties exist simultaneously for specific compositions, they indicate the individual property extrema that could be achieved by a CGA designed within a given subgraph. For example, the composition with the highest yield strength in a given subgraph may not possess the lowest density, but since those properties are present in the same subgraph, a CGA could be designed with those two property extrema as endpoints. Meanwhile, the distributions communicated by the radar plots indicate what properties could be expected along the compositional gradient connecting such endpoints.

Understanding how each subgraph ranks in terms of property performance is important, but so is considering the property value variability within each alloy. A wide spread of values, as mentioned earlier, can indicate either flexibility in subgraph traversal or potential property mismatches between neighboring regions. Fig. 15 represents a box-and-whisker analysis



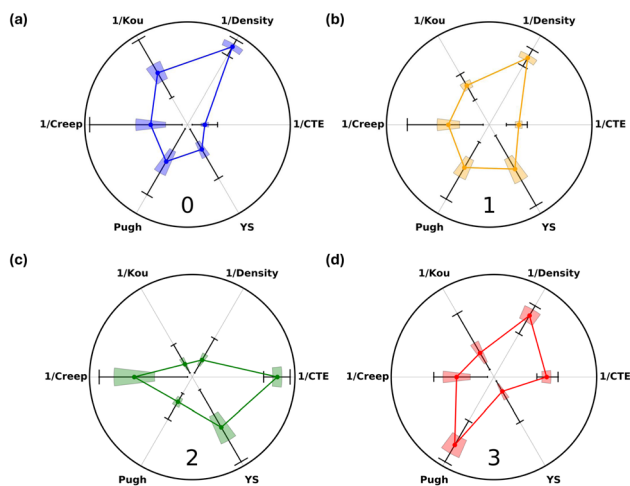


Fig. 15 Radar plots (a–d) show box-and-whisker analyses of six properties across the four subgraphs after applying Scheil BCC phase constraint ( $BCC \geq 0.999$ ). Black lines mark min–max values, shaded boxes show the interquartile range, and points indicate the median. Density, CTE, Kou criterion, and creep are inverted so higher radial values reflect better performance consistently across all properties. Wider ranges denote greater property variability within each subgraph.

for each of the six properties and how they are presented in each of the four subgraphs. The minimum and maximum values of each property are shown at the ends of each whisker. This demonstrates the range of values that could be encountered when selecting a subgraph. The larger ranges display more property value variability, which could affect the outcome of the graded alloy when designing a composition profile. The shaded box on each radial line represents the interquartile range of values with Q1 closest to the center of the circle and Q3 furthest out. The point in the middle of the shaded box represents Q2, or the median value for that property. Values for density, CTE, creep, and Kou criterion were inverted similarly so that greater radial distances correspond to improved property performance. Lower density is typically preferred in a high-temperature application for lightweight purposes. A lower CTE means the part expands less with temperature, improving performance. A lower Kou criterion indicates a reduced risk of solidification cracking and creep was inverted similarly because a lower creep means the material will deform more slowly over time, which is preferable. All property values were normalized using min–max scaling, where each value was transformed by subtracting the global minimum and dividing by the overall global range (max–min), ensuring all properties were mapped to a 0–1 scale.

If subgraph 0 were selected for CGA design, it would offer low density, low hot-cracking susceptibility, high CTE, and moderate mechanical properties (YS, Creep, Pugh). The radar plot shows consistently low density across most nodes, which is favorable for weight-sensitive applications. CTE values, on the other hand, tend to be high for this subgraph. A high CTE is generally a drawback since higher expansion can lead to dimensional instability or thermal mismatch in service. That being said, the spread in CTE values is somewhat narrow, potentially minimizing the risk of mismatch within any CGA designed using this subgraph. The Kou hot-cracking criterion

falls near the middle of the dataset, with a wide enough spread to include both favorable and less favorable options depending on the selected path. Creep rates are higher in many nodes, indicating compositions more susceptible to deforming quickly over time and temperature. The wide range also suggests that some compositions may perform better than others, and careful path planning is vital. The Pugh ratio spans a wide range, reflecting relative differences in ductility and brittleness compared with other subgraphs rather than absolute classification. While a Pugh ratio of about 1.75 is often cited as a ductile–brittle threshold, values here are interpreted in a comparative sense. Lastly, yield strength is generally low, which is typically undesirable for structural applications of high-entropy alloys, but it shows little variation, suggesting consistency through compositions. These trends suggest subgraph 0 is not ideal for strength-critical applications but may still perform reliably in scenarios where other factors like low density take priority. With thoughtful path selection, it could be a strong option for CGA designs focused on lightweight structures as long as high CTE and limited strength are accounted for.

Subgraph 1 shows low-to-moderate density, moderate Kou, creep, Pugh, and YS values, and high CTE. The radar plot indicates tight density values, making this subgraph a fairly reliable option for weight-sensitive applications, similar to subgraph 0. CTE remains high across nodes, which is generally a drawback due to the risk of thermal instability, although it presents minimal variability in its values, diminishing the risk of thermal mismatch between any two compositions. The Kou hot-cracking criterion is slightly elevated with little variation, suggesting a consistent but higher risk of cracking during solidification. Creep resistance sits near the middle of the dataset but spans a wide range, meaning some paths may offer better or worse long-term performance than others. The Pugh ratio shows a similar pattern where most values are moderate, but enough variation exists to include both ductile and brittle behavior. Yield strength is also centered around a mid-range value but has one of the widest spreads of the four subgraphs, offering flexibility but requiring careful path planning to avoid weak regions. Overall, subgraph 1 offers balanced properties with potential for optimization of strength and ductility. However, its high CTE and cracking risk make it less ideal for thermally demanding environments or sensitive manufacturing conditions. With the right priorities and path choices, it remains a viable candidate for a CGA design where thermal stability is less critical.

Subgraph 2 presents a more constrained but high-strength profile, with high density, elevated cracking risk, and mixed mechanical performance. The radar plot shows consistently high density values with limited spread, making this subgraph less suitable for applications where weight reduction is a priority. CTE values are generally low, which is favorable for maintaining dimensional stability under thermal cycling, but the Kou hot-cracking criterion is high with little variation, suggesting a consistent and elevated risk of cracking during solidification that could pose manufacturing challenges. Creep resistance falls near the middle of the dataset but shows a broad



range across nodes, meaning long-term performance under load and temperature could vary significantly depending on the specific compositional gradient path. Yield strength also sits in the moderate range but spans a large distribution, offering flexibility to tune strength levels through path planning. In contrast, the Pugh ratio is low across most compositions with minimal spread, indicating consistently brittle behavior throughout the subgraph. Generally, subgraph 2 is better suited for applications where mechanical strength takes priority over weight or ductility. Its consistently low CTE and potential for high yield strength make it appealing for thermally stable, strength-driven CGA designs. However, elevated cracking risk, brittleness, and high density must be considered when selecting paths within this subgraph.

Subgraph 3 displays a mixed profile, notably with a high Pugh ratio, high Kou criterion, and low yield strength. The radar plot shows low-to-moderate density values with limited variation, making it a reasonably consistent option for reducing weight. CTE is centered around moderate values with very little spread. The Kou hot-cracking criterion, however, is notably high and highly variable. This suggests that while some compositions may perform acceptably during solidification, others are likely to be more prone to cracking, so careful selection is critical. Creep resistance trends slightly higher than average, which could be a concern for applications requiring long-term thermal stability. Its moderate variability does allow for path selection to avoid poor-performing compositions. The Pugh ratio is the highest among the four subgraphs, containing compositions that are all above the ductile–brittle threshold, which would aid in design with minimal risk of encountering a brittle composition region. Yield strength, on the other hand, is lower across the board, with a small spread, making this subgraph a poor candidate for strength-critical applications. Overall, subgraph 3 offers some promising features, including decent weight control, higher ductility, and moderate thermal compatibility. However, its lower strength and elevated hot-cracking risk limit its broader applicability.

The Scheil-filtered subgraphs were evaluated across key material properties to assess their viability for CGA design, revealing distinct trade-offs between mechanical performance, thermal behavior, and processability. Subgraph 0 emerged as a balanced option with consistently low density and moderate cracking susceptibility, although its higher CTE and low yield strength limit its suitability for strength-critical or thermally sensitive applications. Subgraphs 1, 2, and 3 each offered distinct advantages but with greater variability or drawbacks in other areas. The varied selection of subgraphs in this constrained design space provides an excellent case study to demonstrate the factors that should be considered when deciding between subgraphs for any given CGA design scenario. Ultimately, selecting a subgraph and traversal path depends on application-specific priorities, with careful planning needed to balance property gradients and avoid regions of poor performance. If resources allow, multiple subgraphs could be selected and a set-based approach may be taken to design and compare competing CGAs. The methods presented here provide

a structured way to evaluate these trade-offs and support subgraph selection as an early step toward informed CGA design.

Taken together, the results of this study demonstrate that the proposed framework can be used to identify compositionally connected regions of the design space that satisfy manufacturability and phase stability constraints. By restricting candidate compositions to a single traversable subgraph, it becomes possible to design continuous composition pathways for functionally graded alloys that are predicted to remain fully BCC during both equilibrium and non-equilibrium solidification or exhibit a narrow solidification range. Within such a subgraph, additional properties including mechanical performance, thermal behavior, and density can be evaluated and balanced to meet application-specific requirements. In this sense, the framework provides a practical pathway from high-dimensional computational screening to the design of physically realizable compositionally graded materials.

## 4 Conclusion

This study applies a structured, data-driven framework to analyze subgraphs within high-dimensional alloy design spaces as a precursor to compositionally graded alloy design. By integrating thermodynamic modeling, mechanical property estimation, and graph-based data structures, the framework supports efficient exploration of subgraph options suitable for CGA design and additive manufacturing. At the core of this approach is the use of labeled property graphs (LPGs), which store, connect, and query thousands of alloy compositions and their associated properties utilizing an assigned node and edge property–relation approach. Thermodynamic data for phase fractions along with the Kou hot-cracking criterion, density, and CTE were generated using CALPHAD-based simulations. Additional bulk properties such as yield strength, creep resistance, and Pugh ratio were estimated using rule-of-mixtures models and empirical formulas. These values were compiled into a comprehensive dataframe that formed the basis for subgraph filtering and visualization. The Nb–Cr–V–W–Zr system used in this demonstration was the quinary design space of focus chosen from a prior study on CGA design applied to turbine blades.<sup>21</sup> It was selected here for its relevance to high-entropy alloy (HEA) research and potential for high-temperature structural use. Various application-relevant filtering constraints were then applied to this system and evaluated to illustrate their impact on CGA design considerations.

Three filtering constraints were applied to the dataset to reduce the raw compositional space to more practical and manufacturable subsets. These included a solidification range filter ( $\leq 50$  °C), an equilibrium BCC phase fraction filter ( $\geq 0.999$ ), and a Scheil-based BCC phase filter (also  $\geq 0.999$ ) to reflect non-equilibrium solidification conditions. Each filter reduced the number of viable nodes and revealed how different constraints shaped the design space. In particular, the Scheil BCC constraint produced an interesting separation into four distinct and traversable subgraphs, each represented by connected nodes that could be explored without crossing through



infeasible regions. Once separated, these four Scheil-filtered subgraphs were analyzed using a set of material properties relevant to CGA performance and processability: the Kou hot-cracking criterion, equilibrium density, coefficient of thermal expansion (CTE), yield strength, creep rate, and Pugh ratio. This provided a detailed view of how each subgraph performed across multiple metrics and helped identify trade-offs between competing properties.

Component 0 from the Scheil BCC phase fraction constraint emerged as the most well-balanced option overall, offering consistently low density and moderate hot-cracking susceptibility, with relatively stable mechanical properties such as creep and Pugh ratio across its nodes. However, it also exhibited drawbacks, including generally low yield strength and a higher than average CTE. The other components, while seemingly less promising, still had distinct advantages. Component 1 displayed excellent hot-cracking resistance and favorable CTE behavior, but had higher density. Component 2 offered high yield strength and advantageous thermal properties, though it came with greater variability in creep resistance and density. Component 3 provided low density and good creep resistance, but suffered from weaker mechanical strength. Each of these subgraphs presented viable design opportunities, especially when aligned with specific application needs and careful path planning to manage property gradients.

In summary, this work establishes methods for subgraph analysis and selection in high-dimensional alloy spaces. By combining property-driven filtering, graph-based subgraph extraction, and visual analysis tools like UMAP and KDE,<sup>19</sup> this framework enables more informed decision making during early-stage design. It moves beyond traditional search algorithms by narrowing design possibilities into targeted, traversable groups of compositions that align with both performance goals and manufacturing constraints for CGAs. As demands for tailored materials continue to grow, approaches like this will be essential for the informed development of optimized graded alloy systems.

An additional opportunity for future work lies in integrating the proposed graph-based framework with closed-loop or autonomous experimentation platforms. In such settings, experimentally measured properties could be used to update node attributes and constraint evaluations in real time, enabling dynamic refinement of feasible subgraphs and compositional pathways. This type of integration would allow the design space representation to evolve as new data becomes available, supporting adaptive and data-driven exploration of complex alloy systems.

## Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data for this article, including alloy property predictions are available via Zenodo at <https://doi.org/10.5281/zenodo.19473312>.

[zenodo.19473312](https://doi.org/10.5281/zenodo.19473312). Due to commercial licensing terms, raw Thermo-Calc outputs have been omitted from this data. Code capable of reproducing the figures in this work is also available at [https://github.com/ArroyaveLab/CGA\\_Subgraph\\_Analysis](https://github.com/ArroyaveLab/CGA_Subgraph_Analysis).

## Acknowledgements

The authors wish to acknowledge the support from NSF through Grant No. NSF-DMREF-2119103 and 2323611. We also acknowledge the computational resources provided by TAMU High Performance Research Computing. MO also acknowledges the O-REU program held at Texas A&M University under the direction of Prof. Michael Demkowicz. Los Alamos National Laboratory is operated on behalf of the U.S. Department of Energy under contract (89233218CNA000001). MO and BG would like to acknowledge financial support from the Laboratory Directed Research and Development program.

## Notes and references

- 1 M. Allen, R. Arróyave and R. Malak, *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2024, p. V02BT02A011.
- 2 P. Collins, R. Banerjee, S. Banerjee and H. Fraser, *Mater. Sci. Eng. A*, 2003, **352**, 118–128.
- 3 J. Hanagan, N. Person, D. Salas, M. Allen, W. Xu, D. Lewis, C. Acemi, B. Butler, J. D. Paramore, G. M. Pharr, *et al.*, *Mater. Des.*, 2026, 115547.
- 4 O. Eliseeva, T. Kirk, P. Samimi, R. Malak, R. Arróyave, A. Elwany and I. Karaman, *Mater. Des.*, 2019, **182**, 107975.
- 5 L. D. Bobbio, B. Bocklund, E. Simsek, R. T. Ott, M. J. Kramer, Z.-K. Liu and A. M. Beese, *Add. Manufact.*, 2022, **51**, 102649.
- 6 M. J. Abere, H. Choi, L. Van Bastian, L. Jauregui, T. F. Babuska, M. A. Rodriguez, F. W. DelRio, S. R. Whetten and A. B. Kustas, *JOM*, 2024, **76**, 4273–4284.
- 7 R. Mahamood and E. Akinlabi, in *Functionally Graded Materials*, Springer, Cham, 2017.
- 8 E. Alabort, Y. Tang, D. Barba and R. Reed, *Acta Mater.*, 2022, **229**, 117749.
- 9 T. M. Pollock and A. Van der Ven, *MRS Bull.*, 2019, **44**, 238–246.
- 10 X. Yu, J. Xue, Q. Shen, Z. Zheng, N. Ou, W. Wu and L. Jin, *Mater. Chem. Phys.*, 2023, **307**, 128121.
- 11 Y. Su, B. Chen, C. Tan, X. Song and J. Feng, *J. Mater. Process. Technol.*, 2020, **283**, 116702.
- 12 T. Kirk, R. Malak and R. Arroyave, *J. Mechan. Des.*, 2021, **143**, 031704.
- 13 T. Kirk, E. Galvan, R. Malak and R. Arroyave, *J. Mechan. Des.*, 2018, **140**, 111410.
- 14 S. Karaman and E. Frazzoli, *Int. J. Robot. Res.*, 2011, **30**, 846–894.
- 15 O. Adiyatov and H. A. Varol, *2013 IEEE International Conference on Mechatronics and Automation*, 2013.
- 16 R. Duke, V. Bhat and C. Risko, *Chem. Sci.*, 2022, **13**, 13646–13656.



- 17 Z. Ma, M. A. M. Capretz and L. Yan, *Knowl. Eng. Rev.*, 2016, **31**, 391–413.
- 18 C. Sharma and R. Sinha, *Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT '19)*, 2019, pp. 71–80.
- 19 B. Vela, T. Hastings, M. Allen and R. Arróyave, *Digital Discovery*, 2025, **4**, 181–194.
- 20 L. Zhuo, Y. Xie and B. Chen, *J. Mater. Res. Technol.*, 2024, **33**, 1097–1129.
- 21 M. D. Allen, V. Attari, B. Vela, J. Hanagan, R. Malak and R. Arróyave, *arXiv*, 2024, preprint, arXiv:2412.03674, DOI: [10.48550/arXiv.2412.03674](https://doi.org/10.48550/arXiv.2412.03674).
- 22 F. Liu, P. K. Liaw and Y. Zhang, *Metals*, 2022, **12**, 501.
- 23 S. Pugh, *Lond. Edinb. Dubl. Phil. Mag.*, 1954, **45**, 823–843.
- 24 S. Kou, *Acta Mater.*, 2015, **88**, 366–374.
- 25 T.-C. Software, Thermodynamic Database for High Entropy Alloys (TCHEA) Version 6, 2024, <https://web.archive.org/web/20230127050814/https://thermocalc.com/products/databases/high-entropy-alloys/>.
- 26 Z. Yang, H. Sun, Z.-K. Liu and A. M. Beese, *Add. Manufact.*, 2023, **73**, 103672.
- 27 T. Soysal and S. Kou, *Acta Mater.*, 2018, **143**, 181–197.
- 28 P. Singh, B. Vela, G. Ouyang, N. Argibay, J. Cui, R. Arroyave and D. D. Johnson, *Acta Mater.*, 2023, **257**, 119104.
- 29 F. Maresca and W. A. Curtin, *Acta Mater.*, 2020, **182**, 235–249.
- 30 C. Baruffi, F. Maresca and W. Curtin, *MRS Commun.*, 2022, **12**, 1111–1118.
- 31 B. Vela, D. Khatamsaz, C. Acemi, I. Karaman and R. Arróyave, *Acta Mater.*, 2023, **261**, 119351.
- 32 C. Acemi, B. Vela, E. Norris, W. Trehern, K. C. Atli, C. Cleek, R. Arróyave and I. Karaman, *Acta Mater.*, 2024, **281**, 120379.
- 33 A. K. Mukherjee, in *Plastic Deformation of Materials: Treatise on Materials Science and Technology*, ed. R. J. Arsenault, Academic Press, 1975, vol. 6, p. 174.
- 34 T. G. L. David and M. Owen, *Mater. Sci. Eng. A*, 1996, **216**, 20–29.
- 35 T.-C. Software, Mobility Database for High Entropy Alloys (MOBHEA) Version 3, 2024, <https://web.archive.org/web/20230127050814/https://thermocalc.com/products/databases/high-entropy-alloys/>.
- 36 Pandas Documentation, User Guide: Scaling to large datasets, [https://pandas.pydata.org/docs/user\\_guide/scale.html](https://pandas.pydata.org/docs/user_guide/scale.html).
- 37 NetworkX Developers, *NetworkX — Network Analysis in Python*, <https://networkx.org/>.
- 38 O. Senkov, G. Wilks, D. Miracle, C. Chuang and P. Liaw, *Intermetallics*, 2010, **18**, 1758–1765.

