communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-07940-0

Network-informed analysis of a multivariate trait-space reveals optimal trait selection

Quan Pan [1] ✉, **Marijn Bauters** [1], **Marc Peaucelle** [1,2], **David Ellsworth** [3], **Jens Kattge** [4,5] &
Hans Verbeeck [1]

Trait-based analyses have shown great potential to advance our understanding of terrestrial ecosystem processes and functions. However, challenges remain in adequately synthesising a multidimensional and covarying trait space. Reducing the number of studied traits while identifying the most informative ones is increasingly recognized as a priority in functional ecology. Here, we develop a trait reduction procedure based on network analysis of a global dataset comprising 27 traits in three steps. We first construct all possible reduced networks and identify optimal reduced networks that capture the structure of the full 27-trait network. Then we apply the constraints on trait consistency to identified optimal reduced networks and establish consistent network series across ecoregions. We find the best performing networks that capture the three main dimensions of the full network (hydrological safety, leaf economic strategy, and plant reproduction and competition) and the global variance of network metrics. Finally, we find a parsimonious representation of trait covariation strategies is achieved by a 10-trait network which preserves 60% of all the original information while costing only 20.1% of the full suite of traits. Our results show the network reduction approach can improve our understanding on the main plant strategies and facilitate the future trait-based research.

Plant functional traits are morphological, chemical, physiological, and phenological characteristics measured at the individual level. They represent the ecological strategies of plants in multiple dimensions responding to abiotic and biotic environmental constraints[1–3]. As such, trait-based approaches have been proven useful for exploring the relationship between biodiversity and ecosystem functioning, and for improving the representation of diverse plant strategies in next-generation vegetation models[4]. There is a multitude of traits to choose from and costs associated with measuring them, so using the most informative traits is a priority for trait-based research[5,6], especially when the number of trait measurements is limited by fieldwork time and budget. However, trait selection is often done in an unexamined and partial way and thereby it is unclear whether the chosen suite of traits is optimally reflecting ecosystem functioning[5,7].

According to the response-effect trait framework, a certain set of traits, which represents multiple dimensions of plant functional strategies in responding to the environment, have the potential to predict a substantial portion of variation in plant function and ecological process[3,8,9]. For example, the leaf economic spectrum[1], i.e. trade-off between mass investment in leaf structure and the potential rate of resource return[1,10], has been widely used to represent plant strategies and functional diversity in both vegetation modelling and field research[11,12]. However, in addition to the leaf economic spectrum[1], more trade-offs exist in the trait space, linked to various life history strategies, such as nutrient acquisition or environmental adaptation strategies[13,14]. Relying on only one set of strongly correlated traits potentially eliminates important dimensions while overemphasising others. This leads to interpretation biases when the relationship between leaf economic traits and productivity is modulated by other constraints, such as water availability[15], and different trait combinations can lead to varying conclusions, as the predictive capacity of traits for ecosystem functioning depends on both the number of included traits and their linkages to ecological mechanisms[6,7]. As such, the full dimension of trait space and understanding of plant adaptation strategies are needed for capturing functional trade-offs in the ecosystem[16].

---

[1]Q-ForestLab, Laboratory of Quantitative Forest Ecosystem Science, Department of Environment, Ghent University, Gent, 9000, Belgium. [2]INRAE, Université de Bordeaux, Villenave-d'Ornon, Bordeaux, 33140, France. [3]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, 2751 New South Wales, Australia. [4]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Leipzig, 04103, Germany. [5]Max Planck Institute for Biogeochemistry, Jena, 07745, Germany. ✉e-mail: quan.pan@ugent.be

With the development of large trait databases and improved gap-filling methods[17,18], multivariate trait analyses have been extensively applied to identify the most informative traits for a parsimonious representation of the whole trait space while reducing redundant information[19–21]. As the main analysis tool, principal component analysis (PCA) has been widely used to quantify the structure of trait space and identify the main functional dimensions in it[1,3,20]. Aiming to summarize as much trait variation as possible onto the first few axes, PCA may overemphasize certain dimensions while underrepresenting others when traits representing different dimensions are used unevenly[19,20], and potentially leading to a biased interpretation of plant trait dimension and strategies. For example, Weigelt et al. [22]. showed that different conclusions regarding the coupling between aboveground and belowground traits might be interpreted in reverse when size-related traits and seed mass were included in the analysis. In contrast, network analysis aims to transform various systems with complex interactions into networks based on graph theory[23]. It has been widely used in various fields such as social science, economics and epidemiology[24,25]. It can also be used in plant trait research by providing network metrics for describing the pattern of trait-trait correlations novelly[26,27]. Built on pairs of trait-trait relationships, the plant trait network provides an improved resolution of dimensions in the trait space, where intensively correlated traits can be aggregated to a module while independent traits might occupy a distinct module[28]. Therefore, network analysis will have the potential to identify a limited set of key traits from the multivariate trait space to represent the full trait space unbiasedly and inform the optimal trait selection.

Here, we propose a network reduction procedure to identify the most parsimonious and cost-efficient sets of functional traits that represent the functional complexity of plants (Fig. 1). The most parsimonious set of traits can be identified by systematically removing traits from the full network and calculating the structural dissimilarity between the reduced networks and the full network (Table 1). From the most parsimonious networks we can directly identify the most cost-efficient sets of traits accounting for the availability of trait measurements representing measurement costs (Fig. 1, arrow from panel a to d). However, we show that this simple approach can easily be modified to account for aspects that are relevant to trait ecology. First, the parsimonious sets of traits should ideally be part of a series of networks sequentially built by excluding traits one by one from the full dataset because only then are the results robust and practical for applications. Second, ideally, the final set of traits should reflect the correlation structure of traits at the global scale and grasp the changes of the full network across ecoregions (Fig. 1b). To address these additional aspects, we suggest a three-step approach, which we applied to the example of 27 gap-filled traits

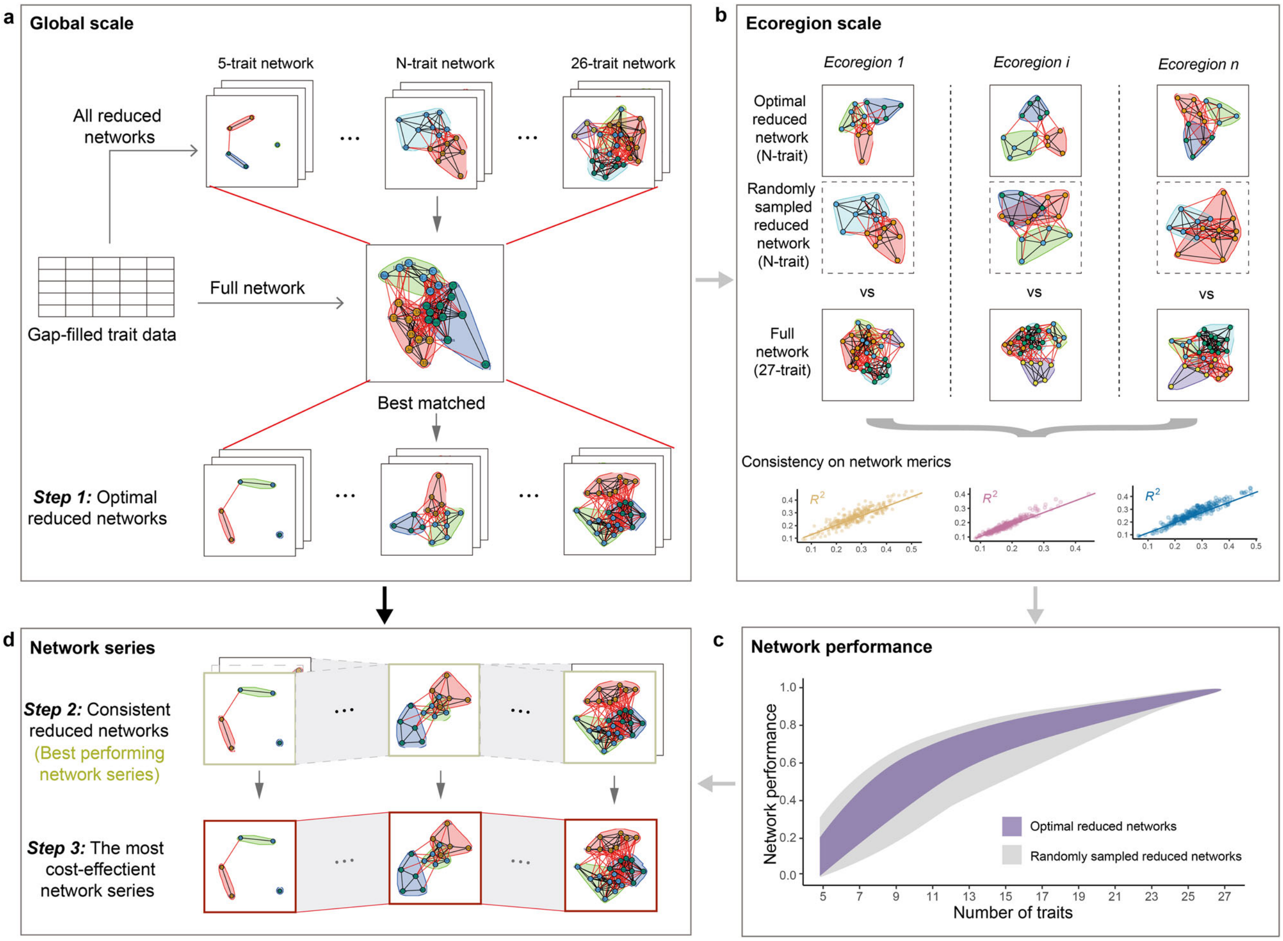

Fig. 1 | Overview of network reduction procedure. a On the global scale, we built all possible reduced networks across size from 5 to 26 traits and calculated their dissimilarity to the full network. The best-matched networks in structure were selected as optimal reduced networks, from which we identified the network series and consistent reduced networks. **b** On the ecoregion scale (i.e. for each of the 262 ecoregions, each reduced network was constructed separately), we assessed the consistency between the full network and both optimal reduced networks and random sampled networks based on their network metrics (density: the ratio of the number of actual links between traits to all possible connections; connectivity: the ratio of links to all possible links in the network in account of the strength of the links; modularity: the complexity of networks by assessing the strength of traits clustering into distinct modules). **c** Then the performance of reduced networks was calculated based on the consistency of these three metrics. **d** Finally, the optimal reduced networks identified in panel a were used to construct the network series (black arrow from **a** to **d**); based on the network performance assessed via panel b&c (grey arrows) and a cost-benefit analysis, the best performing network series and the most cost-efficient network series were identified. All the terminologies in the framework are explained in Table 1.

Table 1 | Terminologies in the network reduction procedures and their explanations

Terminology	Explanation
The full network	The maximum network consisting of the 27 selected traits as nodes, with their interrelationships represented as edges.
Reduced networks	Network consisting of the remaining nodes and their edges after excluding a specific set of traits from the full network.
Optimal reduced networks	Among the reduced networks, those with the highest structural similarity to the full network are identified as the optimal reduced networks.
Network series	A series of reduced networks with the size (N) ranging from max to min; in the network series, every N-trait network includes all the traits present in the N-1-trait network.
Consistent reduced networks	The reduced networks in network series.
Best performing network series	A specific set of consistent reduced networks which have the best network performance on average.
The most cost-efficient network series	A specific set of consistent reduced networks which have the highest performance per unit of measurement cost on average.

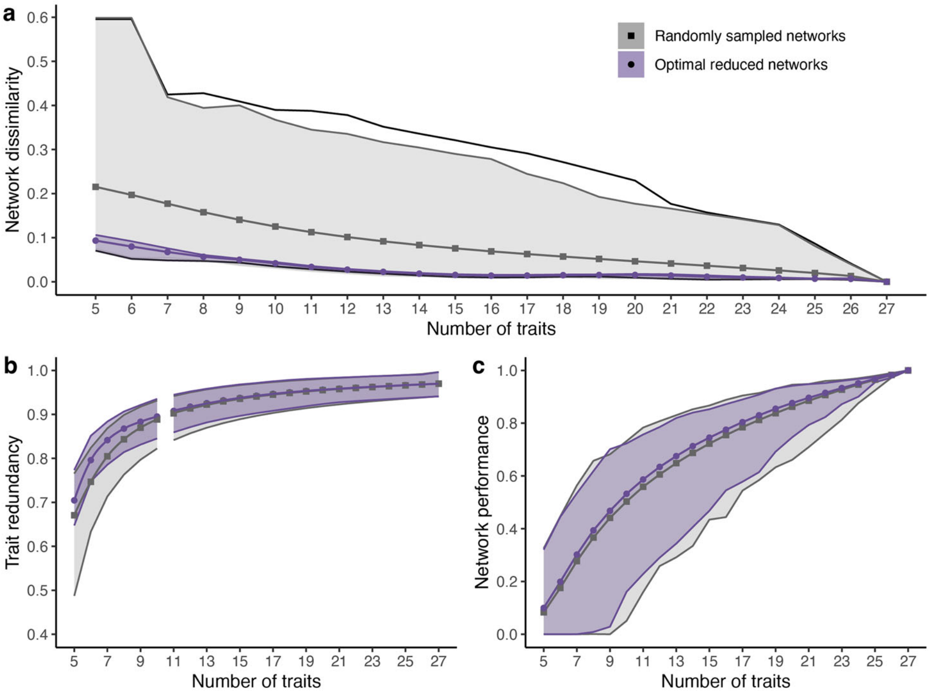

Fig. 2 | The dissimilarity between optimal reduced networks of different sizes to the full network, and the trait redundancy and performance of optimal reduced networks. a. The dissimilarity between reduced networks to the full network. The coloured areas show the envelopes of the values of dissimilarity to the full network from randomly sampled networks (grey) and optimal reduced networks, i.e. the top 5% matched networks (purple) of different sizes, calculated with the approach from Jiang et al.,(2021); the upper black boundary line indicates the maximum dissimilarity of reduced networks in different sizes; the lines in the middle indicate the averaged network dissimilarity for randomly sampled networks (grey) and optimal reduced networks (purple). **b**. The redundancy of traits in reduced networks. The areas show the spread of average redundancy of every trait in randomly sampled reduced networks (grey) and optimal reduced networks (purple) in different sizes. The lines indicate the averaged redundancy of traits. **c**. The performance of optimal reduced networks and randomly selected networks. The coloured areas show the distribution of the performance of randomly sampled reduced networks (grey) and optimal reduced networks (purple), respectively. The lines in the middle of these areas indicate the averaged performance of randomly sampled networks (grey) and optimal reduced networks (purple).

selected from the global plant trait database TRY[17] and visualized in Fig.1. Step 1) We first screened all possible reduced networks with a number of traits varying from 5 to 26 and calculated the structure dissimilarity between the reduced network and the full network. Networks that best matched the full trait network in structure were defined as "optimal reduced networks" (Fig. 1a). We then assessed the performance of these optimal reduced networks based on their capacity to grasp the changes of the full network across ecoregions (Fig. 1b, c). We also sampled a set of randomly reduced networks to assess the improvement of optimal reduced networks. Step 2) From these optimal reduced networks, we constructed a set of network series and identified the best performing network series that can best represent the full trait network for both the global and ecoregion scales (Fig. 1d). Step 3) We estimated the cost of trait measurements based on data availability in the TRY[17] database and performed a cost-benefit analysis to identify the most cost-efficient network (Fig. 1d).

Results

Optimal reduced networks and trait redundancy

In general, network dissimilarity (the difference between the reduced network and the full network in structure) decreases with network size (Fig. 2a). The weighted dissimilarity (WD) shows that at least one reduced network was able to capture the structure of the full network, even in the minimal case with five traits. When the number of traits increased from 5 to 10, the trait redundancy (defined with Eq. 5) of the optimal network increased by 27% from 0.7 to 0.9 on average and was 3.4% higher than that of the randomly sampled networks (Fig. 2b). Trait redundancy continued to increase slowly for both the optimal and randomly sampled networks to reach values from 0.91 to 0.97 when 26 traits were included. This translated into a 2.9% higher average performance of the optimal reduced networks in capturing the shifts of metrics of the full network across ecoregions compared to randomly sampled reduced networks, with minimum performance values being improved by 11.1% (Fig. 2c, Supplementary Table 2). Although both optimal and randomly sampled reduced networks exhibit substantial redundancy in plant leaf traits, followed by stem traits, with seed traits showing the least redundancy in trait space (Supplementary Fig. 1), the optimal reduced networks outperform randomly sampled reduced networks in capturing the shifts of all network metrics across ecoregions, especially for network connectivity, which increased by 18.5% in the minimum (Supplementary Fig. 2, Supplementary Table 2).

Consistent reduced network series and best performing network series

From the optimal reduced networks, 3979 consistent reduced networks were found, translating into the creation of 3,133,589 potential network series. As the size of the network increased from 5 to 27 traits, the dissimilarity of all these consistent reduced networks accounted for 91.8% of the distribution of optimal reduced networks (Fig. 3a). The minimum performance of the consistent reduced networks increased by 14% compared to optimal networks (Fig. 3b, Supplementary Table 3), and the R^2 in predicting network modularity improved by 17.1% on average (Supplementary Table 3, Supplementary Fig. 3).

The three distinctive dimensions in the full 27-trait space (Fig. 4a–c), mainly reflect plant functional dimensions in hydrological safety strategy (dominated by plant height, stem conduit density and specific root length), leaf economic strategy (dominated by leaf C:N, leaf nitrogen content and specific leaf area) and plant reproduction and competition (dominated by seed mass, stem vessel element length and wood fibre length). The best 10-trait reduced network (weighted dissimilarity (WD) = 0.0437) managed to capture this structural complexity of the full trait network, as the traits of the best 10-trait network are also distributed in the three different modules (Fig. 4d): specific leaf area, leaf carbon nitrogen ratio and leaf nitrogen isotope ratio are part of leaf economic module; plant height, stem conduit density, stem diameter and specific root length are involved in the hydraulic safety module; wood fibre length seed germination rate, and stem vessel element length are residing in plant reproduction and competition module. Similar to the full trait space, these three modules also comprise the three main axes in the reduced trait space, which explains 33%, 17% and 16% of the variance of the trait space (Fig.4e, f).

The most cost-efficient network series

The performance of both cost-efficient network series and best performing network series improved gradually with the network size, accompanied by a corresponding increase in the cost of network traits measurement (Fig. 5). However, the most cost-efficient consistent networks achieved a comparable performance with the best performing consistent networks with 44.2% cost reduction on average. As such, with a network of 7 traits, network performance reached 41.6%, with the same predicting capability as the best performing network series but with only 15.2% measuring cost of the full network. With 10 traits included, the network series achieved nearly 60% performance at just 20.1% measuring cost of the full network, whereas the network of the same size in the best performing network series accounted for 45.5% of expenditure (Supplementary Table 4). The largest marginal benefit was obtained for a network with 18 traits when seed mass was added, with an increased performance of 84.3% (Supplementary Fig. 4).

Discussion

Our results proved that a limited number of key traits could well grasp the overall functional complexity of trait space. The best network series provides a ranking of plant traits (SLA, L15N, LCNR, WDL, SRL, SCDen, SD, SCElen, SdGE, Height, LPC.m, DispL, Lwid, Llen, SdM, SCDia, LDMC, RD, Lthick, LA, LNC.m, SdL, SSD, LCC.m, LFM, SdNum, LWC). It prioritizes traits that best represent the full trait network structure and capture shifts in the network metrics across ecoregions. For example, compared to the six commonly used traits and plant global spectrum[3,7], the six-trait network in the best network series also presented the three main dimensions and exhibited better performance in capturing the full complexity of the original dataset (Supplementary Fig. 5, Supplementary Fig. 6). Specifically, the following six traits corresponded to the leaf economic strategy represented by specific leaf area and leaf carbon/nitrogen ratio[1,3], the leaf nitrogen isotope signature and specific root length potentially capturing the resources acquisition strategy and efficiency of plants[29], the stem conduit density representing a key aspect of plant hydraulic safety[30], and the wood fiber length as a proxy of the plant growth and competition strategies, especially at the sapling stage[31]. With all the dimensions represented, these traits exhibited better performance in capturing the full complexity of the original dataset (Supplementary Fig. 6).

The key set of traits identified through the network reduction procedure potentially enhances our understanding of the interactions of plant functional strategies. In the best performing 10-trait reduced network (Fig. 6), traits related to reproduction and growth are presented as an independent module in both of woody and non-woody plant trait networks. While the core trait in the leaf economic spectrum[1], specific leaf area, was more strongly associated with nutrient acquisition traits in woody plants, whereas specific leaf area was more associated with plant size and hydraulic traits in non-woody plants (Fig. 6). The shifting of specific leaf area in different network modules suggests that, as an indicator of fast-slow strategies, it is also associated with other distinct functions in woody and non-woody plants. The fast growth strategy of wood plants is showing a larger reliance on their capacity on nutrient acquisition[32,33], while in non-woody plants, the growth rates are more linked to plant size and hydraulic safety strategies rather than nutrient acquisition[34,35]. Therefore, our findings demonstrated that woody plants have evolved diverse strategies focusing on size and hydraulic safety[36], whereas non-woody plants have primarily diversified toward nutrient acquisition strategies[33]. Additionally, the trait

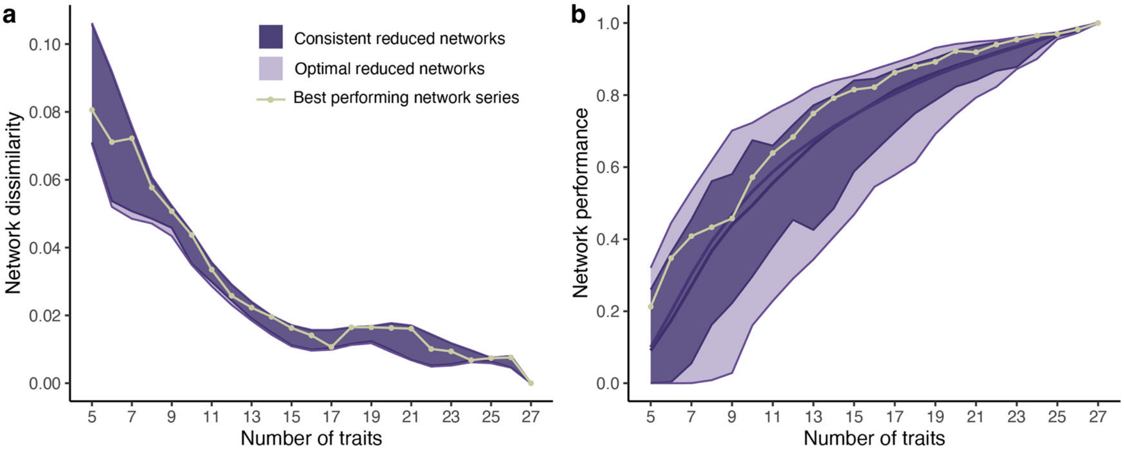

Fig. 3 | The dissimilarity and network performance of the consistent reduced networks and the best-performing network series. **a** The dissimilarity between consistent reduced networks to the full network. The dark purple area shows the distribution of the dissimilarity between consistent reduced network to full networks while the light purple areas show the dissimilarity of optimal reduced network, and the beige line represents the dissimilarity of the optimal network series. **b** The coloured regions show the distribution of network performance of optimal reduced networks (light purple) and consistent reduced networks (dark purple) along the sizes. The beige line in the dark purple area shows the performance of the networks in the best-performing network series.

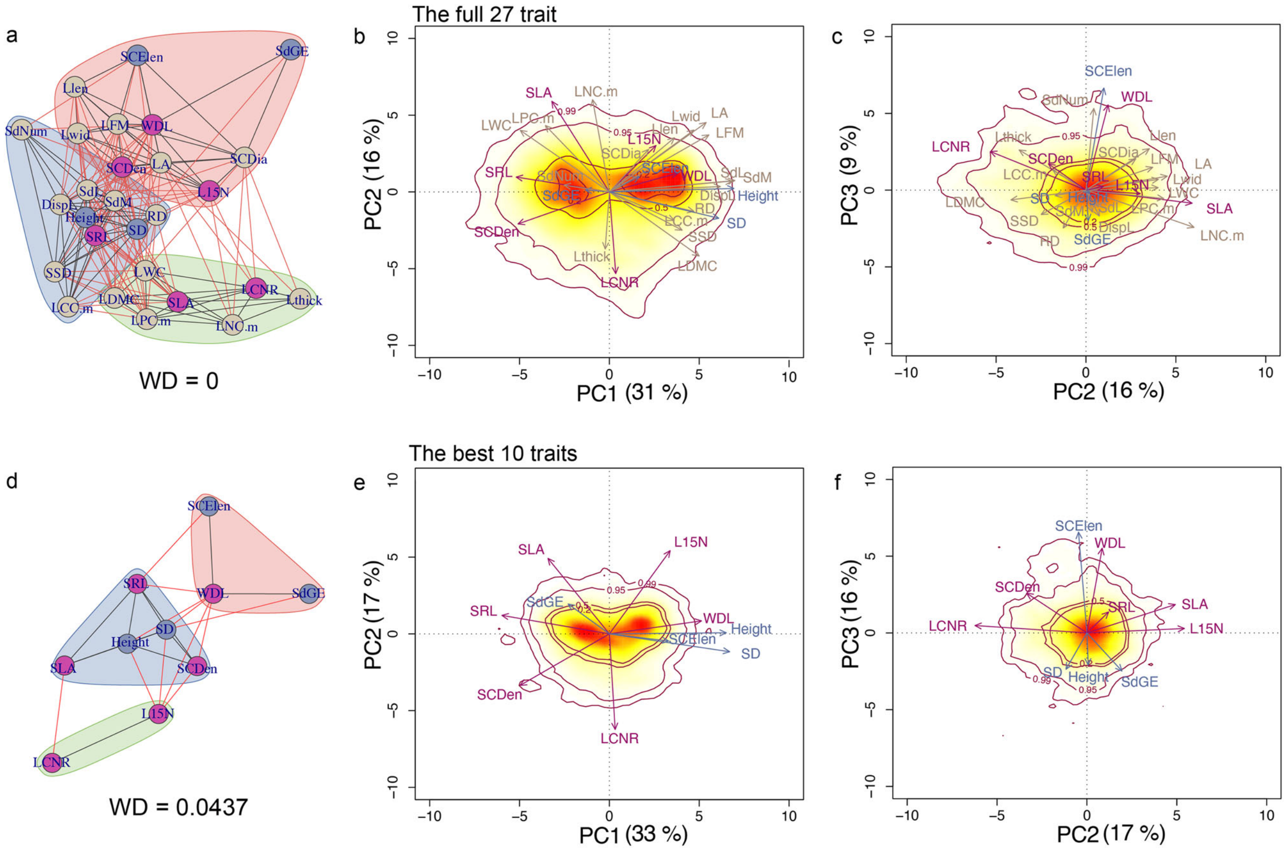

Fig. 4 | The module composition and PCA of best performing networks. a,d show the module composition of the full trait network and reduced 10-trait network in the best performing network series; the colours of modules show the leaf economic module (green), hydraulic safety module (blue) and plant reproduction and competition module (red) in the network independently. WD indicates the structure dissimilarity (weighted dissimilarity) of the reduced network to the full network. **b,c,e,f** show the results of PCA in revealing the dimensions in the full trait space and the best reduced trait space. The colours of traits and PCA arrows indicate the newly added traits in different size of networks. Note: Specific leaf area (SLA), Leaf carbon content (LCCm), Leaf nitrogen content (LNCm), Leaf phosphorus content (LPCm), Leaf thickness (Lthick), leaf dry matter content (LDMC), Leaf nitrogen (N) isotope signature (L15N), Leaf length (Llen), Leaf width (Lwid), Leaf carbon/nitrogen ratio (LCNR), Leaf fresh mass (LFM), leaf area (LA), Leaf water content (LWC), Plant height (Height), Stem diameter (SD), Root rooting depth (RD), Specific root length (SRL), Seed dry mass (SdM), Seed length (SdL), Seed germination rate (SdGE), Seed number per reproduction unit (SdNum), Dispersal unit length (DispL), Stem specific density (SSD), Stem conduit density (SCDen), Stem conduit diameter (SCDia), Stem vessel element length (SCElen), Wood fibre length (WDL).

reduction approach allows a concise representation of a multivariate trait space. For example, instead of including more traits (such as leaf area, leaf width and leaf fresh mass) that represent similar functional dimensions as L15N, the reduced 10-trait network attributes more weight to L15N in the first two main axes and retains a similar three-module structure and comparable functional axes in the PCA to the full 27-trait network (Fig. 4).

Compared to specific ecophysiological processes, the representation of plant diversity in Earth system models is still simply represented by a small number of plant functional types (PFTs), which are defined based on simple categorical traits like plant growth form (tree, shrub, herb) and leaf lifeform (evergreen vs deciduous)[37]. It was proved to fail to encompass the diverse plant strategies in the processes of plant growth, reproduction and survival[38]. Challenges remain in defining a subset of traits that best captures multidimensional variation in plant function and processes[8,39]. The ranking of plant traits is valuable for plant trait selection in trait-based ecology[22]. Our best-performing networks can be potentially used to inform the key traits that should be considered as the important functional dimensions of plants[40], or re-evaluating the choice of traits that are currently used to define PFTs in vegetation models[41].

Maximising trait measurements is a critical imperative for researchers seeking to capture the maximal trait complexity, as a restricted set of traits may lead to the omission of critical plant trade-off dimensions. Here, our most cost-efficient network series is highly informative for practitioners or empiricists having to decide which traits to measure, which can potentially maximise the complexity captured, and minimise redundancy and trait measuring costs. For example, the reduced network, consisting of 10 cost-efficient traits (Supplementary Fig. 7), also captured the three main dimensions of the full trait network and was as performant as the 10-trait network in the best performing network series while being easier to measure (Fig. 5). This suggests that leaf water content and plant height can serve as proxies for plant hydraulic safety strategy[42], offering a low-cost alternative to specific root length. Similarly, the application of leaf area, leaf dry matter content, and leaf nitrogen content enables the reduced trait network to capture the leaf economic strategy of plants with less expense[43].

Network analysis offers a comprehensive method to analyse functional complexity and shows three advantages for multivariate trait analysis in this study. First, as He et al.[28] pointed out, traits that interact strongly can often be combined into one module in the trait networks, which potentially relate to specific aspects of plant function. In our analysis, we found, within the intricate 27 trait space, plant hydraulic safety strategy was not solely dependent on stem hydraulic traits but intricately linked to multifaceted trait strategies spanning multiple plant organs (Fig. 4). e.g., height and specific stem density were potentially related to plant hydraulic safety boundaries[42]; while seed mass and dispersal distance were important to plants in the populations migrating in response to extreme drought[44]. Next, the network structure was less sensitive to the orthogonal traits, which might change the interpretation of functional dimensions of trait space when they were weakly correlated with other traits[22]. It is noteworthy that traits

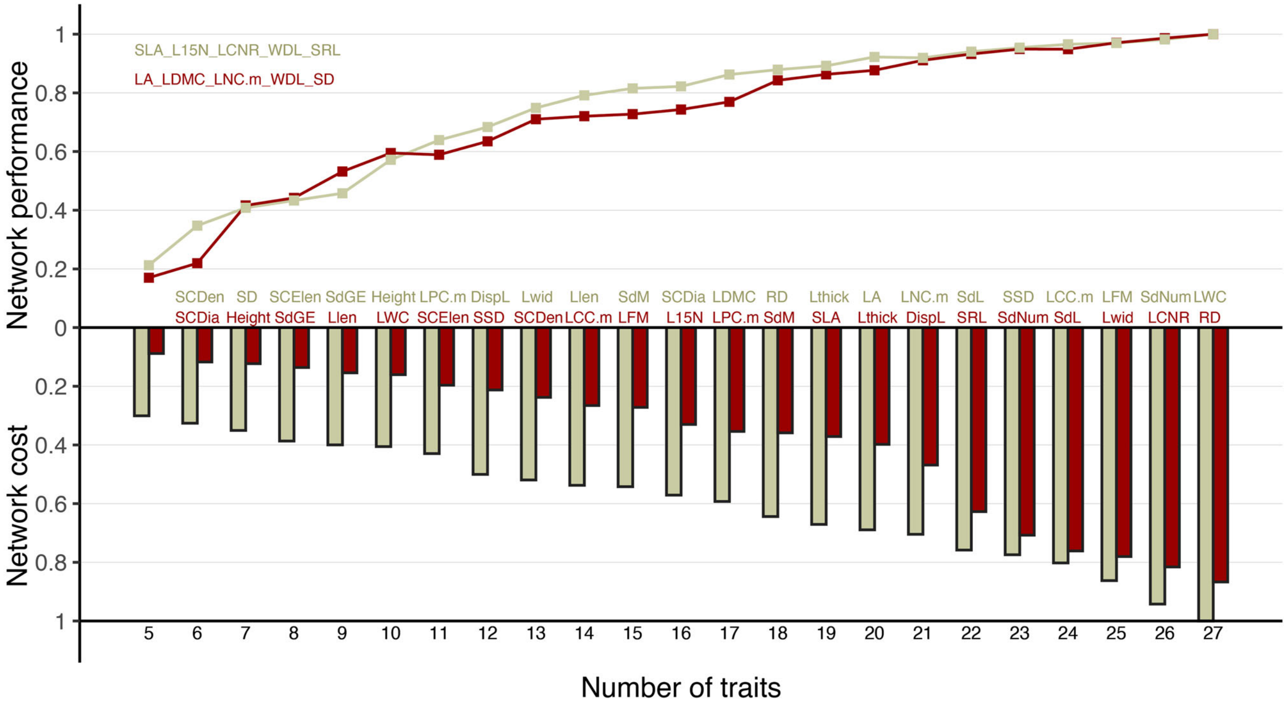

Fig. 5 | The cost and performance of the best performing network series and the most cost-efficient network series. The beige line shows the increasing performance of networks in the best performing network series along the size, and the brown line shows the performance of networks in the most cost-efficient network series. The bar plot represents the cost of networks in the best performing network (beige) series and the most cost-efficient network series (red). The coloured text in the left top shows the initial trait combination (5 traits) of best performing network series (beige) and the most cost-efficient network series (red), and the traits over the 0-axis are the new traits added in the N + 1 trait network. Note: Specific leaf area (SLA), Leaf carbon content (LCCm), Leaf nitrogen content (LNCm), Leaf phosphorus content (LPCm), Leaf thickness (Lthick), leaf dry matter content (LDMC), Leaf nitrogen (N) isotope signature (L15N), Leaf length (Llen), Leaf width (Lwid), Leaf carbon/nitrogen ratio (LCNR), Leaf fresh mass (LFM), leaf area (LA), Leaf water content (LWC), Plant height (Height), Stem diameter (SD), Root rooting depth (RD), Specific root length (SRL), Seed dry mass (SdM), Seed length (SdL), Seed germination rate (SdGE), Seed number per reproduction unit (SdNum), Dispersal unit length (DispL), Stem specific density (SSD), Stem conduit density (SCDen), Stem conduit diameter (SCDia), Stem vessel element length (SCElen), Wood fibre length (WDL).

embodying distinct trade-offs consistently occupy distinct positions in our best network series, such as seed germination rate (Fig. 4). Furthermore, the metrics provided by network analysis enabled the identification of key traits, which capture network structure properly. The key traits may either play central roles, acting as bridges connecting other traits in different dimensions[34,45], or show distinctive dimensions of plant strategies[40,46]. In contrast, highly redundant traits are less important in the full trait space as they can be properly represented with a set of other traits in the same module, given the high redundancy of specific stem density, leaf width, leaf carbon and phosphorus content (Supplementary Fig. 1). Additionally, the sensitivity analysis shows network dissimilarity applied in the network reduction was robust to the threshold (|r|) for network construction (Supplementary Fig. 8).

Although our results show the ability of the trait reduction procedure in selecting key traits from the multivariate trait space, there are two major concerns related to the initial composition of the trait space. Despite the fact that we expect the 27 traits included in this analysis to reveal the prime plant functional trade-offs that are at play globally, some important traits and trait space dimensions might be missing from the analysis[47]. Thus, increasing efforts for assessing and including more plant traits are still needed for further plant trait networks research. The second one lies in the potential redundancy of the initial trait space when strong correlations exist among a group of traits. Even though a great proportion of trait redundancy was reduced by using a smaller number of traits, these strongly correlated patterns might be retained by the best-reduced trait space. Therefore, traits that exhibit mathematical relationships or can be inferred by each other should be excluded from the initial dataset[14]. Additionally, as trait costs vary by species lifeform, the most cost-efficient reduced trait networks may differ across studies. Future research may need to customize trait costs to identify the most efficient networks for their specific objectives.

In conclusion, we demonstrate the network reduction approach, which provide a renewed perspective for assessing trait-trait relationships. It could be a valuable tool for plant trait selection in ecosystem research. In particular, our results show that the initial 27 traits could be reduced to a subset of traits that can optimally represent the full trait space while reducing trait redundancy. Identified by its network performance and cost-efficiency, the best trait combinations potentially enhance our understanding of the trade-off strategies of terrestrial plants and will practically facilitate researchers in trait selection and application.

Methods

Trait information

Gap-filled trait data. We used a global plant trait dataset with 27 gap-filled traits derived from the TRY database[17]. These traits represent some of the most measured and available traits in functional ecology, including leaf, stem, root, seed and whole plant traits. The plant strategies represented by these traits were briefly clarified based on their associated ecological processes (Supplementary Table 1). They potentially reveal the prime plant functional trade-offs that are at play globally[17]. The missing data were imputed with a Bayesian hierarchical probabilistic matrix factorization (BHPMF)[17,18]. This method has been proven to be reliable in exploring trait-environmental relationships[12,20]. To link trait data with local regions and environmental factors, ecoregions that contain geographically distinct assemblages of species were used[48]. All traits were aggregated into species median value (for plant size traits like plant height, stem diameter and rooting depth, 95th percentile value was used) in every ecoregion. In total, 368,940 records were integrated into 36,628

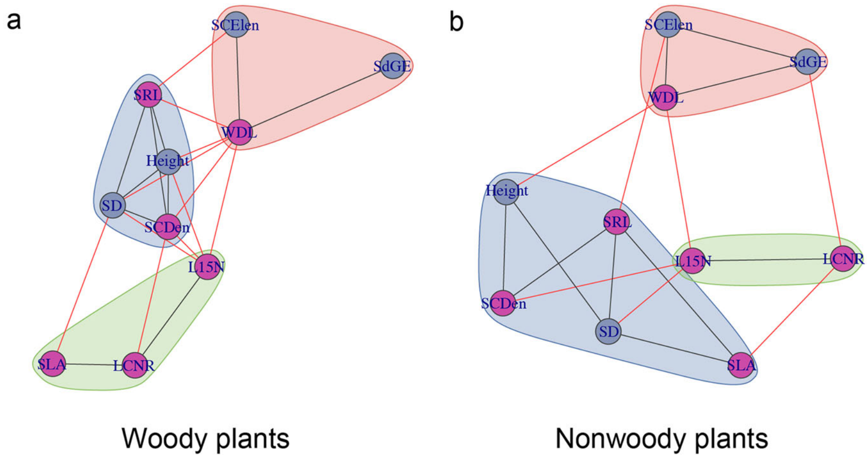

Fig. 6 | Network module compositions of the best performing 10-trait reduced network in woody and non-woody plants. The colours of modules and traits show the leaf economic module (green), hydraulic safety module (blue) and plant reproduction and competition module (red) in the woody plant trait network (**a**) and nonwoody plant network (**b**). The colours of traits in networks indicate the newly added traits in different size of networks.

species-level records in 483 ecoregions according to their georeference in TRY database[17]. Traits that did not follow a normal distribution were log-transformed to approximate normality.

Network analysis

Network building. To construct the networks, traits were considered as nodes (or vertices) and trait-trait relationships were represented by edges connecting each pair of nodes[28]. Calculated with the 'stats' package in R[49], the absolute value of pairwise Pearson correlation coefficients (| r |) was used to represent the distance of the edges between a pair of traits, and a larger Pearson |r| indicates a shorter distance between traits in the network. In the network reduction procedure, recognizing that the number tested trait correlation greatly exceeds the number of observed species in some ecoregions, we avoid spurious correlations by setting a fixed threshold ($p >= 0.05$ or $|r| < 0.2$)[14,45], without applying correlation correction[50]. Although this may introduce false positives in some network edges, it ensures that the constructed networks remain reliable in ecoregions with a small number of observed species. Finally, network graph objects were generated and analysed with the package 'igraph' in R[51].

Network metrics. Network analysis provides various metrics to illustrate network similarity and trait-trait interaction patterns, enabling the identification of key traits in trait space that can capture network structure properly. The variances of these network metrics reflect the interaction strategies of different functions of plants[52,53]. Here, we used a combination of three main network metrics-the Edge Density, Connectivity and Modularity-based on their ecological significance identified in previous studies[45,53,54], to evaluate trait interactions and the structure of a trait space[45,54]. "Edge Density" and "Connectivity" measure the degree to which traits work together throughout the entire network. Edge Density is the ratio of the number of actual links between traits to all possible connections. Connectivity here is defined as the ratio of links to all possible links in the network, weighted by the strength of the links[26,27]. The "Modularity" quantifies the complexity of networks by assessing the strength of traits clustering into distinct modules. In a network, high modularity indicates extensive links among traits within modules but sparse links among traits between modules. The identified network modules potentially correspond to specific aspects of plant function, aligning with concept of plant functional spectra[28,53]. Modularity maximisation is recognized as one of the most prominent algorithms for modularity detection in network science[55]. Despite it may lead to overfitting issues where spurious communities can be identified[56], the modularity maximization algorithm can effectively summarize the key modules within the trait network and the main functional dimensions of plants[28], and sensitive to detect the modularity changes even for the simplest graph perturbation[57]. Here, the network module composition was detected with the R function 'cluster_optimal' in 'igraph'[51], which calculated the maximal network modularity by transforming the maximisation of the modularity of all possible partitions into an integer linear programming (ILP) problem.

Trait redundancy. Trait redundancy is the overlap of trait information. Quantifying trait redundancy helps investigate the degree to which dimensions are overweight. Here, the three main network metrics (edge density, connectivity and modularity) were used to represent the information of the trait in context of trait space, and the trait redundancy was calculated based on the network sensitivity, referring the relative changes of network metrics to the removal of traits which represents the extra information added by this trait. We let edge density and connectivity share the half weight, as they both indicate the strength of connections among traits in the network based on their definitions (4), and the redundancy of a trait was calculated using the following Eq. (5).

$$\Delta Density_{G_N(t)} = Density_{G_N} \tag{1}$$

$$\Delta Connectivity_{G_N(t)} = Connectivity_{G_N} - Connectivity_{G_N(t)} \tag{2}$$

$$\Delta Modularity_{G_N(t)} = Modularity_{G_N} - Modularity_{G_N(t)} \tag{3}$$

$$Sensitivity_{G_N(t)} = \frac{\sqrt{\left(0.5 \times \Delta Density^2_{G_N(t)} + 0.5 \times \Delta Connectivity^2_{G_N(t)} + \Delta Modularity^2_{G_N(t)}\right)}}{\sqrt{\left(0.5 \times Density^2_{G_N} + 0.5 \times Connectivity^2_{G_N} + Modularity^2_{G_N}\right)}} \tag{4}$$

$$Redundancy_{G_N(t)} = 1 - sensitivity_{G_N(t)} \tag{5}$$

Where G_N is the N-trait reduced network, and the $G_N(t)$ is the N-trait reduced network with trait t removed; the $Density_{G(N)}$, $Connectivity_{G(N)}$ and $Modularity_{G(N)}$ are the values of edge density, connectivity and modularity of the N-trait reduced network $G(N)$; the $Density_{G_N(t)}$, $Connectivity_{G_N(t)}$ and $Modularity_{G_N(t)}$ are the values of edge density, connectivity and modularity of the trait t removed reduced network $G_N(t)$.

Network dissimilarity. The dissimilarity quantifies the structure difference between reduced networks and the full network. Here, it was calculated with WD-measure (Eq. (6)) which was initially defined by Schieber et al.[58], and has been extended to weighted networks[59]. The WD-measure quantifies network structure difference through the Jensen–Shannon divergence[59], which provides a symmetric and bounded metric (ranging from 0 to ln2) for comparing the similarity of the distribution of nodes connectivity, nodes heterogeneity and nodes centrality between networks consistently, enabling to capture network dissimilarities at both global and local scale[58].

$$WD = w_1\sqrt{JSD\left(u_{G_N}, u_{G_M}\right)/\log 2} + w_2\left|\sqrt{WNND(G_N)} - \sqrt{WNND(G_M)}\right| + \tfrac{w_3}{2}\left(\sqrt{JSD\left(P_{\alpha G_N}, P_{\alpha G_M}\right)/\log 2} + \sqrt{JSD\left(P_{\alpha G_N^c}, P_{\alpha G_M^c}\right)/\log 2}\right) \tag{6}$$

where $w_1, w_2,$ and w_3 are the weights of each term with $w_1 + w_2 + w_3 = 1$. Same with a previous study[58], we selected $w_1 = 0.45$, $w_2 = 0.45$ and $w_3 = 0.1$.JSD is the Jensen–Shannon divergence and G_N^c and G_M^c indicate the complement of G_N and G_M, which consist of all the nodes presented in G_N^c and G_M^c and the edges not presented in G_N and G_M. $JSD\left(u_{G_N}, u_{G_M}\right)$ measures the difference between the networks' averaged node-distance distributions, and $JSD\left(P_{\alpha G_N}, P_{\alpha G_M}\right)$ measures the difference between the α-centrality values of the networks. The $WNND$ is defined as:

$$WNND(G) = \frac{JSD\left(P_1^\omega, P_2^\omega, \ldots, P_N^\omega\right)}{\log(m+1)} \tag{7}$$

$$JSD\left(P_1^\omega, P_2^\omega, \ldots, P_N^\omega\right) = \frac{1}{N}\sum_{i,j} p_i^\omega(j)\log\left(\frac{p_i^\omega(j)}{\mu_j^\omega}\right) \tag{8}$$

where m is the number of columns of the discontinued weighted distance probability matrix P_ω. $P_i^\omega = \left\{p_i^\omega(j)\right\}$ is the discontinued weighted distance distribution in node i. The average of the neighbours proportion μ_j^ω can be obtained through the distance probability matrix P_ω as:

$$\mu_j^\omega = \sum_{i=1}^{N} p_i^\omega(j)/N \tag{9}$$

where $p_i^\omega(j)$ is the fraction of nodes connected to node i at discontinued weighted distance j, transformed with the methods in Jiang et al. [59].

Network performance. The performance measure F_{G_N} is designed based on the consistency of the network metrics of the reduced network G_N and the full network G_{27} for each ecoregion. We used the coefficient of determination (R^2) of the linear regression between the metrics of reduced networks and the full net for computing network consistency. The linear regression was fitted with the 'lm' function in R[49]. Ecoregions with less than 20 species were discarded[20], and a total of 262 ecoregions were included for this analysis. We assigned equal weight to edge density and connectivity, as both metrics similarly represent the strength of connections among traits within the network.

$$F_{G_N} = \sqrt{\sqrt{R_d^2 \times R_c^2} \times R_m^2} \tag{10}$$

where the R_d^2, R_c^2 and R_m^2 are the coefficients of determination of the linear regression on edge density, connectivity and modularity between the reduced network and the full network in 262 ecoregions.

Network module composition and Principal component analysis. To identify the network module composition and dominant axes of trait space, we used the optimal module clustering and principal component analysis (PCA). This was conducted on traits in different sizes of reduced networks using the 'igraph' and 'stats' packages in R[49,51].

Identification of network series

Step 1: optimal reduced networks. The optimal reduced networks are supposed to exhibit a similar structure as the full network. The number of potential reduced networks varies with the number of traits used to construct that network, ranging from a few to thousands of possible combinations. We thus limited the number of optimal reduced networks by imposing thresholds. First, for networks of 5 to 6 traits, and of 21 to 25 traits, we extracted the top 5% networks that best matched the full network based on dissimilarity. For trait networks of 7 to 20 traits, where the top 5% would still have rendered a lot of different combinations, we capped the number of networks with 30.000 combinations - again based on the best ranking networks in terms of dissimilarity. Finally, for the trait network of 26 traits - with 27 possible combinations, we took the 10 best-matched (based on lowest dissimilarity) networks. Simultaneously, we randomly sampled 100,000 reduced networks in different sizes as a contrast to optimal reduced networks. For both optimal reduced networks and randomly sampled reduced networks, we calculated the trait redundancy and network performance, and then compared the difference in trait redundancy and examined whether these optimal reduced networks are performing better than other networks.

Step 2: Consistent reduced network series and best-performing network series. In optimal reduced networks described in Step 1, the best performing network of N traits could contain a completely different set of traits than the best-performing network of N + 1 traits. In practice, there is a need for consistent sets of traits, ordered from high-priority traits to highly redundant traits. Indeed, the number of traits one can sample or include will be constrained by research budgets and project timeframes. As such, there is a need for consistent reduced trait networks that perform optimally regarding a full trait network, but where traits can be added or removed following the maximum added value of these respective traits. Consequently, a reduced trait network of N traits would be optimally performing regarding the full network, while the trait network of N + 1 traits would do so as well but contain N common traits and one added trait only. From optimal reduced networks, the network series can be identified starting from the 5-trait network, and the next trait 'i' can be added to the current N-trait network only if the network composed of these N + 1 traits is also included in the optimal reduced networks. Finally, the best-performing network series are identified based on the averaged performance of the reduced network in capturing the change of the three main network structure metrics in the full trait network across ecoregion.

Step 3: The most cost-efficient network series. The optimal network series identified in Step 2 extract the key traits that best represent the interaction pattern of the full trait space while reducing trait redundancy. However, in order to inform trait selection for observational ecosystem science, we also need to account for the ease of acquisition of functional traits. In other words, a reduced network of eight traits might marginally outperform another network of eight different traits in terms of trait space representation, but if the former set is much more difficult to measure, practitioners might be better off with the latter trait set. Thus, there is a need to determine the cost-effectiveness of trait measurements. To this goal, we introduced a method accounting for an acquisition cost to each trait. The cost of measuring each trait was defined here based on data availability in the database (version 5, Supplementary Table 1) with the assumption that the cost of measuring a trait is inversely proportional to the number of its observations in the database.

The cost of measuring each trait was defined here based on data availability in the database (version 5, Supplementary Table 1) with the assumption that the cost of measuring a trait is inversely proportional to the number of its observations in the database. Then, the cost-efficiency (CE_{G_N}) of the reduced network was calculated with the Eq. (13).

$$Weight_t = \frac{Max(count(T_{1\sim 27}))}{count(T_t)} \tag{11}$$

$$Cost_t = \frac{Weight_t}{\sum_1^{27} Weight_i} \tag{12}$$

$$CE_{G_N} = \frac{F_{G(N)}}{Cost_{G_N}} \tag{13}$$

where the $count(T_t)$ is the number of species measured trait t; F_{G_N} is the performance of network G_N; $Cost_{G_N}$ is the total cost of all traits in network G_N. The calculation of specific leaf area requires the measurements of leaf area and leaf mass which are much less reported in TRY. This reporting bias may lead to overestimating of the measuring cost to these traits. Thus, we manually assigned the same measurement cost to leaf area, leaf fresh mass, leaf water content, and leaf dry mass content as that of specific leaf area. Finally, the most cost-efficient network series was identified based on a cost-efficiency analysis.

Marginal benefit. To identify the best network in the most cost-efficient network series, we calculated the marginal benefit of adding trait t MB_t following Eq. (14).

$$MB_t = \frac{F_{G(N)} - F_{G(N,-t)}}{Cost_t} \tag{14}$$

where F_{G_N} is the performance of network G_N, $F_{G_{(N,-t)}}$ is the performance of network $G_{(N,-t)}$, $G_{(N,-t)}$ is the network G_N with trait t removed; $Cost_t$ is the cost of trait t, evaluated according to criteria described in Step 3 above.

Statistics and reproducibility

In this study, a global gap-filled trait dataset was used, covering 262 ecoregions and 27 traits. Trait-trait correlation was estimated with Pearson correlation coefficients through the 'stats' package in R. Network construction and network metrics calculation were conducted using the igraph package (version 1.2.6). The calculation of network dissimilarity and network performance was conducted on the HPC of Ghent University. All analyses were conducted using publicly available R scripts to ensure reproducibility (R version 4.1.2).

Data availability

The initial trait and environment data primarily used in this study were obtained from TRY.

Code availability

R code and data for analysis and visualizing have been deposited on Figshare[60].

Received: 28 May 2024; Accepted: 17 March 2025;

References

1. Wright, I. J. et al. The worldwide leaf economics spectrum. *Nature* **428**, 821–827 (2004).
2. Violle, C. et al. Let the concept of trait be functional!. *Oikos* **116**, 882–892 (2007).
3. Díaz, S. et al. The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).
4. Zakharova, L., Meyer, K. M. & Seifan, M. Trait-based modelling in ecology: A review of two decades of research. *Ecol. Modell.* **407**, 108703 (2019).
5. Liu, C., Li, Y., Yan, P. & He, N. How to improve the predictions of plant functional traits on ecosystem functioning?. *Front. Plant Sci.* **12**, 622260 (2021).
6. Hagan, J. G., Henn, J. J. & Osterman, W. H. A. Plant traits alone are good predictors of ecosystem properties when used carefully. *Nat. Ecol. Evol*. https://doi.org/10.1038/s41559-022-01920-x (2023).
7. van der Plas, F. et al. Plant traits alone are poor predictors of ecosystem properties and long-term ecosystem functioning. *Nat. Ecol. Evol.* **4**, 1602–1611 (2020).
8. Funk, J. L. et al. Revisiting the Holy Grail: using plant functional traits to understand ecological processes. *Biol. Rev. Camb. Philos. Soc.* **92**, 1156–1173 (2017).
9. Lavorel, S. & Garnier, E. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Funct. Ecol.* **16**, 545–556 (2002).
10. Reich, P. B. Key canopy traits drive forest productivity. *Proc. Biol. Sci.* **279**, 2128–2134 (2012).
11. Sakschewski, B. et al. Leaf and stem economics spectra drive diversity of functional plant traits in a dynamic global vegetation model. *Glob. Chang. Biol.* **21**, 2711–2725 (2015).
12. Bruelheide, H. et al. Global trait-environment relationships of plant communities. *Nat. Ecol. Evol.* **2**, 1906–1917 (2018).
13. Walker, A. P., McCormack, M. L., Messier, J., Myers-Smith, I. H. & Wullschleger, S. D. Trait covariance: the functional warp of plant diversity?. *N. Phytol.* **216**, 976–980 (2017).
14. Poorter, H., Lambers, H. & Evans, J. R. Trait correlation networks: a whole-plant perspective on the recently criticized leaf economic spectrum. *N. Phytol.* **201**, 378–382 (2014).
15. Staples, T. L., Dwyer, J. M., England, J. R. & Mayfield, M. M. Productivity does not correlate with species and functional diversity in Australian reforestation plantings across a wide climate gradient. *Global Ecol. Biogeogr*. https://doi.org/10.1111/geb.12962 (2019).
16. Laughlin, D. C. & Messier, J. Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends Ecol. Evol.* **30**, 487–496 (2015).
17. Kattge, J. et al. TRY plant trait database - enhanced coverage and open access. *Glob. Chang. Biol.* **26**, 119–188 (2020).
18. Schrodt, F. et al. BHPMF - a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Glob. Ecol. Biogeogr.* **24**, 1510–1521 (2015).
19. Maynard, D. S. et al. Global relationships in tree functional traits. *Nat. Commun.* **13**, 3185 (2022).
20. Joswig, J. S. et al. Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nat. Ecol. Evol.* **6**, 36–50 (2022).
21. Mouillot, D. et al. The dimensionality and structure of species trait spaces. *Ecol. Lett.* **24**, 1988–2009 (2021).
22. Weigelt, A. et al. The importance of trait selection in ecology. *Nature* **618**, E29–E30 (2023).
23. Newman, M. The Structure and Function of Complex Networks. *SIAM Rev.* **45**, 167–256 (2003).
24. Nishi, A., Shirado, H., Rand, D. G. & Christakis, N. A. Inequality and visibility of wealth in experimental social networks. *Nature* **526**, 426–429 (2015).
25. Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl Acad. Sci. USA* **106**, 21484–21489 (2009).
26. Xie, J., Wang, Z. & Li, Y. Stomatal opening ratio mediates trait coordinating network adaptation to environmental gradients. *N. Phytol.* **235**, 907–922 (2022).
27. Felipe-Lucia, M. R. et al. Land-use intensity alters networks between biodiversity, ecosystem functions, and services. *Proc. Natl Acad. Sci. USA* **117**, 28140–28149 (2020).
28. He, N. et al. Plant trait networks: improved resolution of the dimensionality of adaptation. *Trends Ecol. Evol.* **35**, 908–918 (2020).
29. Craine, J. M. et al. Ecological interpretations of nitrogen isotope ratios of terrestrial plants and soils. *Plant Soil* **396**, 1–26 (2015).
30. Feild, T. S. & Brodribb, T. J. Hydraulic tuning of vein cell microstructure in the evolution of angiosperm venation networks. *N. Phytol.* **199**, 720–726 (2013).
31. González-Melo, A. Wood anatomical traits mediate life-history variations at the sapling, but not at the adult stage. *Trees* **36**, 1337–1347 (2022).
32. Rennenberg, H. et al. Nitrogen balance in forest soils: nutritional limitation of plants under climate change stresses. *Plant Biol. (Stuttg.)* **11**, **Suppl 1**, 4–23 (2009).

33. Xia, J. & Wan, S. Global response patterns of terrestrial plant species to nitrogen addition. *N. Phytol.* **179**, 428–439 (2008).
34. Kleyer, M. et al. Trait correlation network analysis identifies biomass allocation traits and stem specific length as hub traits in herbaceous perennial plants. *J. Ecol.* **107**, 829–842 (2019).
35. Wellstein, C. et al. Effects of extreme drought on specific leaf area of grassland species: A meta-analysis of experimental studies in temperate and sub-Mediterranean systems. *Glob. Chang. Biol.* **23**, 2473–2481 (2017).
36. Pivovaroff, A. L. et al. Multiple strategies for drought survival among woody plant species. *Funct. Ecol.* **30**, 517–526 (2016).
37. Sitch, S. et al. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Glob. Change Biol.* **9**, 161–185 (2003).
38. Scheiter, S., Langan, L. & Higgins, S. I. Next-generation dynamic global vegetation models: learning from community ecology. *N. Phytol.* **198**, 957–969 (2013).
39. Keller, A. et al. Ten (mostly) simple rules to future-proof trait data in ecological and evolutionary sciences. *Methods Ecol. Evol*. https://doi.org/10.1111/2041-210X.14033 (2022).
40. Rüger, N. et al. Demographic trade-offs predict tropical forest dynamics. *Science* **368**, 165–168 (2020).
41. Anderegg, L. D. L. et al. Representing plant diversity in land models: An evolutionary approach to make "Functional Types" more functional. *Glob. Chang. Biol.* **28**, 2541–2554 (2022).
42. Liu, H. et al. Hydraulic traits are coordinated with maximum plant height at the global scale. *Sci. Adv.* **5**, eaav1332 (2019).
43. Polley, H. W., Collins, H. P. & Fay, P. A. Community leaf dry matter content predicts plant production in simple and diverse grassland. *Ecosphere* **13**, (2022).
44. Zhu, K., Woodall, C. W. & Clark, J. S. Failure to migrate: lack of tree range expansion in response to climate change. *Glob. Chang. Biol.* **18**, 1042–1052 (2012).
45. Li, Y. et al. Leaf trait network architecture shifts with species-richness and climate across forests at continental scale. *Ecol. Lett.* **25**, 1442–1457 (2022).
46. Larson, J. E., Anacker, B. L., Wanous, S. & Funk, J. L. Ecological strategies begin at germination: Traits, plasticity and survival in the first 4 days of plant life. *Funct. Ecol.* **34**, 968–979 (2020).
47. Walker, T. W. N. et al. Leaf metabolic traits reveal hidden dimensions of plant form and function. *Sci. Adv.* **9**, eadi4029 (2023).
48. Olson, D. M. et al. Terrestrial ecoregions of the world: A new map of life on earth. *Bioscience* **51**, 933 (2001).
49. R Core Team. R: The R Project for Statistical Computing. https://www.r-project.org/ (2023).
50. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.: Ser. B (Methodol.)* **57**, 289–300 (1995).
51. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. (2006).
52. Li, Y. et al. Leaf trait networks based on global data: representing variation and adaptation in plants. *Front. Plant Sci.* **12**, 710530 (2021).
53. Liu, C., Li, Y. & He, N. Differential adaptation of lianas and trees in wet and dry forests revealed by trait correlation networks. *Ecol. Indic.* **135**, 108564 (2022).
54. Flores-Moreno, H. et al. Robustness of trait connections across environmental gradients and growth forms. *Global Ecol. Biogeogr.* https://doi.org/10.1111/geb.12996 (2019).
55. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl Acad. Sci. USA* **103**, 8577–8582 (2006).
56. Peixoto, T. P. *Descriptive vs. Inferential Community Detection in Networks: Pitfalls, Myths and Half-Truths*. https://doi.org/10.1017/9781009118897 (Cambridge University Press, 2023).
57. Brandes, U. et al. On Modularity Clustering. *IEEE Trans. Knowl. Data Eng.* **20**, 172–188 (2008).
58. Schieber, T. A. et al. Quantification of network structural dissimilarities. *Nat. Commun.* **8**, 13928 (2017).
59. Jiang, Y., Li, M., Fan, Y. & Di, Z. Characterizing dissimilarity of weighted networks. *Sci. Rep.* **11**, 5768 (2021).
60. Pan, Q. Network-informed analysis of a multivariate trait-space reveals optimal trait selection. https://doi.org/10.6084/m9.figshare.25968190.v1.

Acknowledgements

Quan Pan is supported by the China Scholarship Council (CSC) [No. 202104910073]; Marijn Bauters is supported through a senior postdoctoral fellowship funded by the Flemish Science Foundation FWO. Marc Peaucelle would like to acknowledge the financial support from the H2020 Marie Skłodowska-Curie Actions (LEAF-2-TBM grant no. 891369). And we would like to thank all PIs contributing to the TRY database, whose efforts allowed this analysis. Finally, we thank the four reviewers for their constructive comments and suggestions on earlier drafts of this manuscript.

Author contributions

M.B., Q.P., M.P., H.V. and D.E. conceived and designed the study. J.K. provided the gap-filled plant trait data. Q.P. performed data analyses and wrote the original draft. All authors participated in data interpretation, edited the draft and agreed to the published version of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-07940-0.

Correspondence and requests for materials should be addressed to Quan Pan.

Peer review information *Communications Biology* thanks Jianming Deng and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editors: Aylin Bircan, David Favero. A peer review file is available.