

<https://doi.org/10.1038/s42005-025-02073-6>

Ranking nodes in bipartite systems with a non-linear iterative map

Andrea Mazzolini¹ ✉, Michele Caselle² & Matteo Osella²

Ranking nodes in networks according to a defined measure of importance is an extensively studied task, with applications in ecology, economic trade networks, and social networks. This paper introduces a method based on a non-linear iterative map to evaluate node relevance in bipartite networks. By tuning a single parameter γ , the method captures different concepts of node importance, including established measures like degree centrality, eigenvector centrality and the fitness-complexity ranking. The algorithm's flexibility allows for efficient ranking optimization tailored to specific tasks, outperforming state-of-the-art algorithms. We apply this method to ecological mutualistic networks, where ranking quality can be assessed by the extinction area - the rate at which the system collapses when species are removed in a certain order. The map with the optimal γ value surpasses existing ranking methods on this task. Additionally, our method excels in evaluating nestedness, another crucial structural property of ecological systems, requiring specific node rankings. Finally, we explore theoretical aspects of the map, revealing a phase transition at a critical γ dependent on the data structure that can be characterized analytically for random networks. Near the critical point, the map exhibits unique features and a distinctive “triangular” packing pattern of the incidence matrix.

How to quantify the “importance” of a node in a graph is an extensively studied question in network theory. Several definitions of node centrality have been proposed, each exploiting different aspects of the network structure¹. The simplest one is the well-known *degree centrality*, in which the importance of a node is simply its degree². The natural extensions of this idea are centrality measures, which take into account also the importance of the node's neighbors. The *eigenvector centrality*³ and the *Katz centrality*⁴ are two examples. The popular *PageRank centrality*⁵ is also based on a similar concept. In this case, the importance of a node is a linear function of the importance of its neighbors, but their contribution is normalized by their out-degree. While these centrality measures are the most relevant for this paper, there are plenty of other examples in network theory^{6–10}.

Bipartite networks are a particular class of networks in which links only connect nodes from two distinct sets. This structure describes a vast number of complex systems, specifically component systems. In component systems, each realization is an assembly of basic building blocks, such as books composed of words or genomes composed of genes, which naturally translates into a bipartite network of connections between basic components and realizations^{11–14}. This analogy has led to a fruitful exchange of methods between network theory and statistical data analysis in different domains^{15–17}.

While extensions of node centrality measures to bipartite networks have been proposed, the problem of node ranking in these ubiquitous structures is still an active area of research. A possible naive approach is to project the links of the bipartite network on one of the two node sets, and then directly use standard tools of network analysis on the resulting unipartite structure^{18–21}. However, the projection clearly hinders a fundamental property of the network with possible relevant consequences on the analysis results^{22–25}. Few methods that explicitly keep into account the bipartite structure have been proposed, such as extensions of the eigenvector centrality^{26,27}, or methods based on the statistics of random walkers on the network²⁸. Finally, a statistical approach based on a non-linear map, the so-called fitness-complexity map, was introduced in the context of economics, and then generalized and applied in other fields^{29–34}.

The present paper proposes a general and flexible method to rank the importance of nodes in bipartite networks based on a non-linear map inspired by the fitness-complexity concept. The non-linearity of the map depends on a single parameter γ . The parameter value sets what structural features are relevant for defining the node importance and thus the final node ranking. Indeed, tuning γ we can interpolate between the eigenvector centrality, the degree centrality, and the original fitness-complexity map.

¹Laboratoire de physique de l'École normale supérieure (PSL University), CNRS, Paris, France. ²Department of Physics and INFN, University of Turin, Turin, Italy.

✉ e-mail: andrea.mazzolini.90@gmail.com

Depending on the system under analysis and the specific task, different choices of γ can be optimal.

Ecological networks provide an illustrative example in which different algorithms, and the resulting rankings, can be quantitatively compared^{33,35}. In fact, the so-called extinction area can be used to quantify how relevant a node ranking is in complex ecological networks. This quantity measures how fast the network collapses when nodes (e.g., species) are removed from the system in a specific order. Therefore, the ranking of species maximizing the extinction area can inform preservation strategies and interventions in endangered ecosystems. The fitness-complexity map was reported to outperform previously existing algorithms when applied to this class of mutualistic systems³³. We will show that our generalized non-linear map can provide even better results by choosing the appropriate value of γ .

In the same context of ecological systems, nestedness is a long-studied structural property^{35–37}. A nested ecological system is essentially composed by few *generalist* species that tend to interact with almost all the other species, and by several *specialists* that preferentially interact with generalists, rather than with other specialists. The identification of nested structures and the generalist species at their core can help guide the design of conservation and preservation efforts, e.g., refs. 38,39. Therefore, in addition to the theoretical interest in understanding the evolution of ecological networks, accurate nestedness measures also have practical importance. Several popular measures of nestedness require finding the ordering of rows and columns of the incidence matrix that maximizes its “triangular shape.” We will show that again our algorithm improves the performance of the fitness-complexity map, which, in turn, has been shown to perform better than previous state-of-the-art algorithms⁴⁰.

Having established the practical usefulness of this flexible non-linear map, we will focus on its general mathematical properties and on the role of the parameter γ . More specifically, we will show that the dynamical system defined by the map undergoes a phase transition in its convergence properties for a critical value γ_c that depends on structural network properties. We will explore in detail the close connection between the phase transition, the maximization of the extinction area, and the re-ordering of the incidence matrix in a triangular form, with all the ones located in the upper-left corner and the bottom-right corner only composed by zeros. This connection is also instrumental to define the parameter region where the value of γ maximizing the extinction area can be found.

Methods

Datasets

The ecological interaction matrices used in this work were downloaded from Web of Life (<http://www.web-of-life.es/>), an open access database with a large number of published species interaction networks. The full list of matrices used in the result section is reported in the Supplementary Table 1. We primarily focus on large systems, i.e., those with a product of the number of rows and columns of at least 500, which are, in general, harder to treat with a full exploration of possible rankings or the application of computationally expensive genetic algorithms. In particular, the illustrative example discussed in the result section is a plant-pollinator interaction matrix based on the work of C. Robertson who collected in 1929 the list of interactions between plants and insects in Carlinville, Illinois. This dataset was validated and updated in ref. 41, extending the network to 1044 animal species visiting 456 plant species.

A general and flexible measure of node importance for bipartite systems

This section presents the non-linear map and how to use it to measure the node importance for bipartite networks. We consider unweighted undirected bipartite networks, with an $N \times M$ incidence matrix A , with $A_{ij} \in \{0, 1\}$ and $A_{ij} = 1$ if node i is connected to node j on the opposite node set (Fig. 1a, b). We denote with \mathbf{x} the scores of the nodes in the first node set of the network, and with \mathbf{y} the scores relative to the second one. The scores are iteratively updated by the map, which starts from vectors of ones as

initial conditions, i.e., $x_i^{(0)} = 1$ and $y_j^{(0)} = 1$, and is defined by the equations

$$\begin{cases} x_i^{(t)} = \frac{\tilde{x}_i^{(t)}}{\langle \tilde{x}^{(t)} \rangle}, \text{ where } \tilde{x}_i^{(t)} = \sum_{j=1}^M A_{ij} \left(y_j^{(t-1)} \right)^\gamma, \\ y_j^{(t)} = \frac{\tilde{y}_j^{(t)}}{\langle \tilde{y}^{(t)} \rangle}, \text{ where } \tilde{y}_j^{(t)} = \sum_{i=1}^N A_{ij} \left(x_i^{(t-1)} \right)^\gamma. \end{cases} \quad (1)$$

The angular brackets denote the vector average $\langle \mathbf{x} \rangle = \sum_{i=1}^N x_i / N$. The map, possibly, converges to stationary values, which represent the importance measure of the nodes. In general, there is no guarantee that a stationary condition exists. However, we could define a ranking of the nodes from the map behavior at large t for all the empirical systems considered. A detailed discussion about the map convergence is postponed to the section about the phase transition.

The numerical approximations that have to be applied to reliably simulate the map in the presence of possible score divergences, and a pseudocode are discussed in detail in the Supplementary Note 1. The algorithm, and some of its applications, can be downloaded from the public repository <https://github.com/amazzoli/xymap.git>.

The map depends on the free parameter γ which can be tuned depending on the idea of node importance suggested by the system in analysis. In particular, the generalized map recovers known measures of importance for specific values of γ , as detailed in the following.

Degree centrality, $\gamma = 0$. The simplest case is the degree centrality, which is defined as the number of neighbors of a node, and can be obtained for $\gamma = 0$. In this case, the stationary solution is reached at the first map iteration, and the scores are simply the node degree divided by the average degree. Figure 1c shows the degree ranking of the toy example illustrated in Fig. 1a.

Singular vector centrality, $\gamma = 1$. For $\gamma = 1$ the importance of a node is proportional to the importance of the connected nodes in the other node set. In other words, a node is important if it has a lot of important neighbors. The illustrative example in Fig. 1d shows that this notion of importance leads to a different ranking with respect to the case of $\gamma = 0$ (i.e., degree centrality). In particular, the second squared node now surpasses the first one in the ranking because, even if it has less connections, it is connected to the most important circular node in the opposite side of the network.

This is the foundational idea of the eigenvector centrality in classical unipartite networks³. According to this measure, the importance c_i of a node i is

$$c_i = \lambda^{-1} \sum_{j=1}^N B_{ij} c_j,$$

where B is the $N \times N$ adjacency matrix, which we assume to be symmetric and irreducible, i.e., the network is undirected and connected. The equation above, when written in matrix form, is satisfied by each eigenvector of B with corresponding eigenvalue λ : $B\mathbf{c} = \lambda\mathbf{c}$. To resolve the ambiguity of having multiple possible choices for the scores, the scores are required to be all positive. This constrains the choice to the eigenvector corresponding to the largest eigenvalue, since the Perron–Frobenius theorem proves that it is the only eigenvector with all positive components⁴². Therefore, the leading eigenvector of B defines the so-called eigenvector centrality. This definition of score is arbitrary to within a multiplicative constant. However, one is usually interested in the node ranking, which is not affected by this degree of freedom. An important property of the eigenvector centrality is that it can be computed in a dynamical way, in the similar spirit of the map given by Eq. (1). Starting from a vector of ones as initial condition, the recursive relation $\mathbf{c}^{(t+1)} = \lambda^{-1} B\mathbf{c}^{(t)}$, converges to the leading eigenvector, in the limit of large t .

This measure can be extended to bipartite networks²⁷. In this case, we can consider the rectangular incidence matrix A and substitute the concept

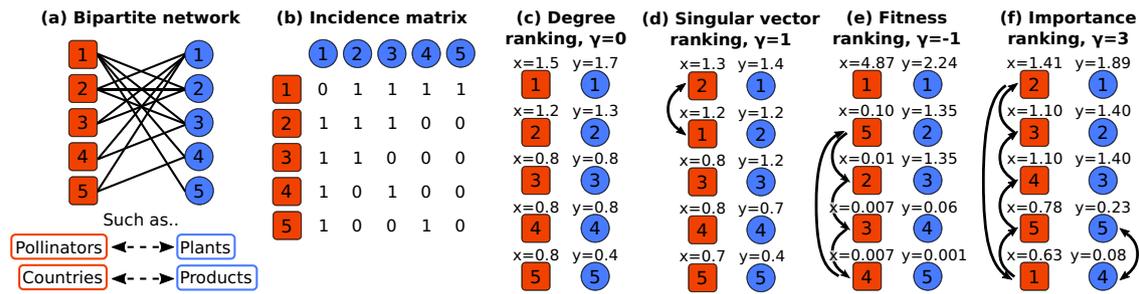


Fig. 1 | Toy example of a bipartite system and its node ranking. The network is represented as a non directed, unweighted, bipartite network, **a** where a set of squared nodes, e.g., countries or pollinator animals, interact with a disjoint set of circular nodes, e.g., products or plants. Such a network is associated with the

incidence matrix **b** in which the element is equal to 1 if there is a link between the two nodes. The panels from **c**–**f** show the scores and the rankings that Eq. (1) provides for four values of γ and are discussed in the main text. The black arrows between nodes are changes in the ranking with respect to the degree ranking.

of eigenvectors with left/right singular vectors, and eigenvalues with singular values. Supplementary Note 2 shows the proof that the iteration of the Eq. (1) for $\gamma = 1$ leads to a vector of scores x which is proportional to the leading left singular vector u_1 of A (corresponding to the largest singular value σ_1), while y is proportional to the leading right singular vector v_1 :

$$\begin{aligned} x &\propto u_1 & \text{where } A^T u_1 &= \sigma_1 v_1 \\ y &\propto v_1 & A v_1 &= \sigma_1 u_1 \end{aligned} \quad (2)$$

Again, the Perron–Frobenius theorem guarantees that the leading singular vectors are positive, and therefore consistent with the unique definition of a positive score.

It is useful to describe one possible example of application of this score system: rating the desirability of products which have been used by a set of consumers²⁷. This can be described by a consumer-product bipartite system which has links whether a consumer has tried a product. Clearly, the more a product is consumed (high degree) the more important it is, but one can also speculate that an “experienced” consumer increases further the importance of the products that it has tried. The evaluation of consumer experience can be made by the amount of the tried products and if those tried products are considered very desirable. Another important example is the measure of economic complexity introduced in²⁶ (with a small difference in the normalization factor of the map).

Fitness-complexity map, $\gamma = -1$. The choice of the exponent $\gamma = -1$ recovers the fitness-complexity map introduced in the context of economics^{29,30,32}. The score x_i corresponds to the fitness, while the score y_j is the inverse of the complexity. Therefore, the fitness-complexity map is recovered through the substitution $\tilde{q}_j = \tilde{y}_j^{-1}$ for every j , and defining the complexity as $q_j = \tilde{q}_j / \langle \tilde{q} \rangle$.

In this regime, the map attributes large importance to nodes with many connections, but, differently from the previous case of $\gamma = 1$, the connections with low-importance neighbors have a larger weight. This effect is shown by the ranking in Fig. 1e, where the squared node 5 increases its position in the ranking with respect to the degree sorting, and surpasses the higher-degree node 2. Indeed, node 5 is connected to the circular node 4, which has the second lowest importance, while the squared nodes 2–4, are not.

The *fitness-complexity* map was introduced to quantify the non-monetary competitiveness of a country on the basis of its exported products. The basic idea is that the fitness (i.e., the x score) of a country is proportional to the sum of the complexities of its exported products, which in our notation is the sum of the inverse of the “simplicities,” given by the y scores. The simplicity of a product is large if many countries can export that product, in particular if these countries have low fitness. On the contrary, a product is complex (i.e., low y) if only few countries with high fitness can produce it.

The range of application of the fitness-complexity map extends to other bipartite networks. For example, it has been used to analyze the contribution of different countries to a set of scientific topics⁴³. A different field of

application are mutualistic ecological networks. In the case of the interactions between plants and pollinators, the “fitness” of an animal pollinator seems to be significantly related to its ecological importance within the ecosystem, while the “complexity” of a plant quantifies its vulnerability to system perturbations³³. This ecological example will be our main illustrative application.

Behavior at large exponents. It is instructive to characterize the map behavior for large exponents. In the case of $\gamma \gg 1$, the first score of Eq. (1) can be approximated as:

$$x_i^{(t)} \propto \frac{\max_{j \in J(i)} \left(y_j^{(t-1)} \right)^{\gamma}}{\max_{j \in \{1, \dots, N\}} \left(y_j^{(t-1)} \right)^{\gamma}}, \quad (3)$$

where the maximum at numerator comes from the approximation of $\tilde{x}^{(t)}$, and it is evaluated among the set $J(i)$ of neighbors of i . The maximum at denominator is instead the leading term of $\langle \tilde{x}^{(t)} \rangle$, and considers the largest y -score among all the nodes of the second node set. A similar expression can then be found for the update of the y scores, whose values depend only on the maximal x of their neighbors. This expression implies that only the nodes connected with the maximal y -score have non-negligible x -score and, therefore, will dominate the ranking. In this limit, the degree of the node does not enter explicitly in the computation. This can be seen in Fig. 1f, which displays a ranking calculated for $\gamma = 3$. The squared node 1, which is the most connected among the squared nodes, becomes the last one in the ranking, since it lacks the connection with the most important circular node, i.e., the first one. The opposite limit of $\gamma \rightarrow -\infty$ leads exactly to the same result of Eq. (3), but with a minimum instead of a maximum. Therefore, the x -score is determined by the connection with the minimal y -score.

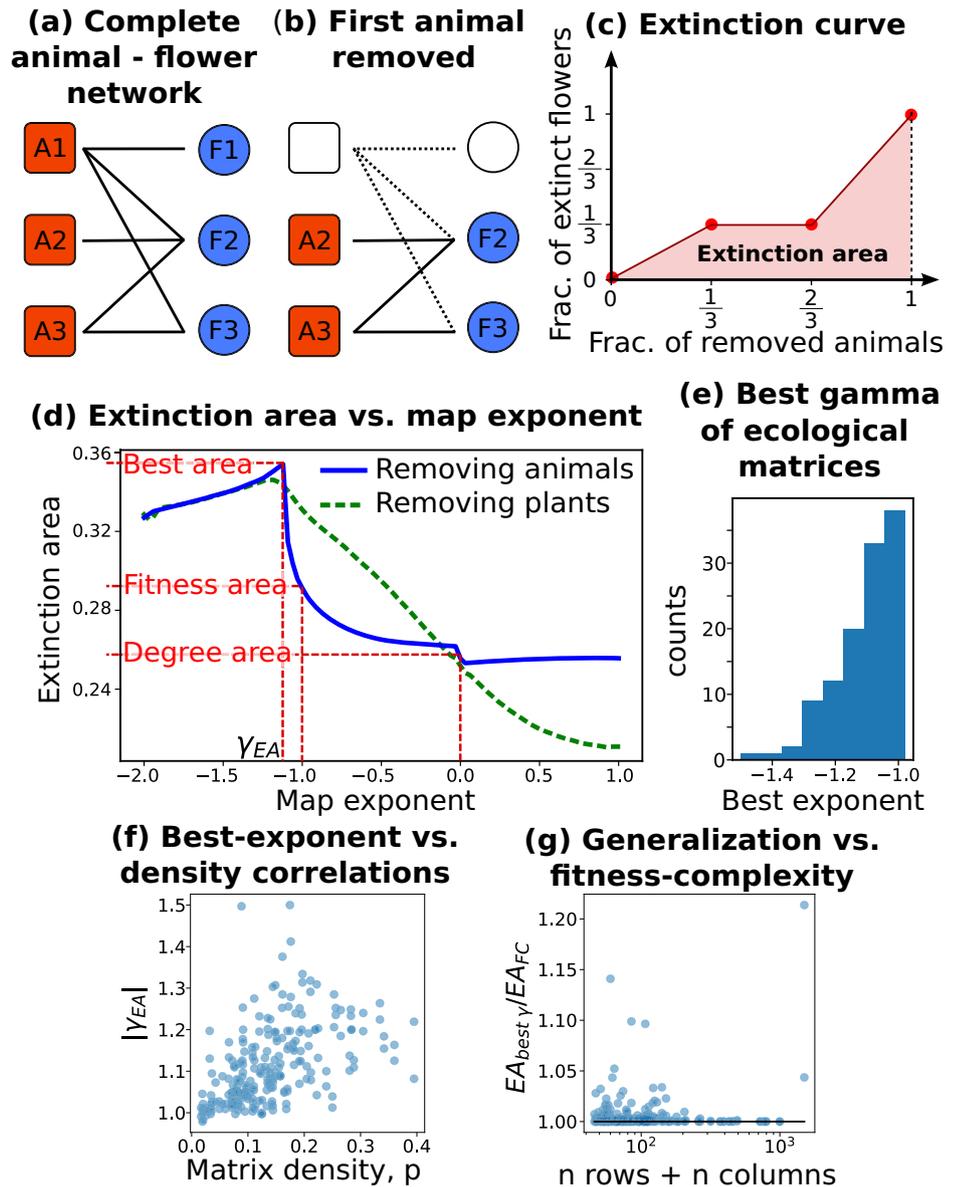
General intuition behind the map behavior. Putting all these considerations together one can conclude that the absolute value of γ sets the balance between two properties in determining the importance of a node. The first is the degree of the node which, for $\gamma = 0$, is the only relevant feature. The second property is the connection with important nodes in the other node set in the limit of large $|\gamma|$, only the maximum (or minimum) score among the connected neighbors contributes to the importance. The sign of γ determines if the importance is given by having high-score neighbors (positive γ) or low-score ones (negative γ).

Results

Looking for the specie ranking which maximizes the extinction area

We can now evaluate the ranking generated by the map on ecological systems of mutualistic interactions between species, such as plants and pollinators. We use the extinction area to quantitatively evaluate the goodness of the rankings^{33,44}. The computation of this quantity is illustrated in Fig. 2a–c. Given an animal ranking (e.g., (A_1, A_2, A_3)), we can first remove

Fig. 2 | Computing the extinction area in ecological networks. a–c show a toy example of the extinction area computation. Starting from the complete network (a), animals (matrix rows) are removed according to a certain ranking. In (b) the animal A_1 has been deleted leading to the extinction of the plant F_1 which remains without links. The extinction area is the integral of the curve (c) which shows the fraction of extinct plants as a function of the fraction of removed animals. **d** is the extinction area computed at different map exponents γ of the mutualistic system Robertson 1929. The blue continuous line correspond to animal removal, where the animal ranking is provided by the x -score at the given γ . The green dash-dot line is computed using the opposite procedure: the plants are progressively removed following the y ranking. **e** shows the number of matrices whose extinction area by animal removal is maximized at given exponents, γ_{EA} , among the 116 mutualistic networks of the Web-of-Life database, see ?? and table S1, having size (product between the number of rows and columns) larger than 500. **f** shows the correlation between matrix density and map exponent that maximizes the extinction area - Spearman correlation $\rho = 0.61$, $p = 10^{-24}$. **g** reports the ratio between the extinction areas obtained with the fitness-complexity map with our generalization. We consider both rows and columns removal procedures.



the first animal (A_1) and all its links. As a consequence, a certain number of plants remain without links and thus goes extinct (F_1 in the example).

By progressively removing all the animals according to the ranking, we can calculate the fraction of extinct plants as a function of the fraction of removed animals (Fig. 2c). The extinction area is defined as the integral of this curve, and quantifies the speed of the ecosystem collapse for the given ranking.

A previous work³³ demonstrated that the fitness-complexity map, that we recover for $\gamma = -1$, provides an animal ordering that leads to a larger extinction area than other algorithms. Figure 2d shows that, in a specific case, our map can reach even larger values of extinction area for other values of γ . In particular, there is a maximum at $\gamma_{EA} \simeq -1.12$. The improvement with respect to the fitness-complexity map is even larger than the improvement that the fitness-complexity map has compared to the simple degree ranking ($\gamma = 0$).

The value of γ that maximizes the area, γ_{EA} , is dataset dependent. We notice that in all the cases we studied γ_{EA} was strictly smaller than -1 and typically within the range $\gamma_{EA} \sim [-1.4, -1]$. We report in Fig. 2e the distribution of the best parameters for all the mutualistic networks discussed in “Dataset” section. We shall further discuss the issue of finding the best

gamma and, in particular, the origin of the $\gamma_{EA} < -1$ found in result section about the phase transition.

Given the variability in the exponents that maximize the extinction area across different ecosystems, we can test whether simple statistical features of the interaction matrices correlate with these values. The most significant (anti)correlation we observe is with matrix density, i.e., the fraction p of actual links among all possible ones. In fact, Fig. 2f suggests that very sparse matrices are optimally ranked by exponents close to one, whereas denser matrices require smaller exponents. The corresponding analysis for uniform random matrices, reported in Supplementary Fig. 8, confirms that, on average, smaller exponents are needed for denser matrices. For random matrices, we also observe a clear dependence on system size (i.e., the number of species) at a given interaction density. The possible reasons behind these dependencies are further discussed later.

For several systems, the exponent γ_{EA} that maximizes the extinction area is close to -1 (Fig. 2e), where our map becomes equivalent to the fitness-complexity approach. Indeed, for these interaction networks, our map rankings provide little to no improvement in the extinction area. However, Fig. 2g highlights several notable exceptions where our map leads to significantly larger extinction areas –by up to 20%—as observed for the

ecological system used for Fig. 2d. In these cases, the species ranking obtained with γ_{EA} can be significantly different from the ranking corresponding to $\gamma = -1$. Supplementary Fig. 9 analyzes the ranking variations using γ_{EA} or $\gamma = -1$ for the specific network analyzed in Fig. 9d. The rank is well conserved only for the lowest-ranked species. However, in many cases, species shift by as many as 100 positions (out of a total of 1044 species), with some changing by up to 500 positions. The top two species remain the same in both rankings, but the top ten differs dramatically. Species ranked in the top ten of the fitness-complexity ranking drop to positions 200 or 300 in the best-gamma ranking, and vice versa. Consistently with the interpretation discussed in the method section, since γ_{EA} is relatively smaller and far from $\gamma = 0$, species degree plays a less significant role in determining the ranking. The color pattern of Supplementary Fig. 9 shows that the increased importance of “low-importance” connected species in the other set in the ranking. In this example, using the fitness complexity map all top 10 species have a large degree (around 100), while in the γ_{EA} ranking only two species have large degree, five have an intermediate degree (~ 50) and three have a small degree (~ 10).

One further question is how close the obtained extinction area is to the global maximum. We can explore a large population of rankings by employing a genetic algorithm that searches for the best extinction area. Supplementary Note 3 shows that, for different ensembles of randomly generated matrices, the average best extinction area found by our map always coincides or is better than the genetic-algorithm result. It is worth noticing that the iterative map is typically 100 times faster than a genetic algorithm, as also discussed in Supplementary Note 3. In the same section, we analyze the extinction area obtained by different algorithms when applied to the small ecological matrices of the Web of Life database, for which we can run a genetic algorithm in a reasonable amount of time. Also in this case our map find the same extinction area of the genetic algorithm, but for those very small ecosystems the ranking difference with respect to the fitness-complexity map is negligible.

Finally, we investigated possible differences in progressively removing rows or columns. As panel A shows, there can be differences in the EA value at fixed exponent, but the maximum is roughly at the same value. Supplementary Fig. 10 indeed shows that the optimization procedures by removing rows or columns lead to almost the same maximum for each matrix. There is more variability in the exponents at which the maximum is attained using the two alternative removal procedures. The exponents often do not coincide precisely, but are however strongly correlated (Supplementary Fig. 10).

Looking for the species ranking that maximizes the matrix nestedness

The classical way to measure the nestedness of a network depends on the reordering of rows and columns⁴⁵. Given a row and column order, one can compute a nestedness *temperature* and the final temperature of the network is the minimal one among all the possible orders. Therefore, this measure relies on algorithms that have to explore the factorially growing space of possible rankings of rows and columns.

The fitness-complexity map has proven to be highly effective for this purpose, identifying rankings that, in several cases, outperform state-of-the-art search algorithms such as BINMATNEST^{40,46}. This task can be tackled with our map by varying the parameter γ and choosing the value minimizing the nestedness temperature. We adopted the temperature definition of ref. 46 and⁴⁰, whose implementation can be found in the repository associated to the manuscript <https://github.com/amazzoli/xymap.git>. The temperatures that we find are typically better than the ones at $\gamma = -1$ (Fig. 3a). In particular, we find lower temperatures also for larger matrices, where the fitness-complexity map was performing slightly worse than the BINMATNEST, as reported in Fig. 3 of ref. 40. Furthermore, our map, with an optimized gamma value, systematically results in a lower temperature compared to BINMATNEST (Fig. 3b) or a more recent algorithm that addresses the nestedness maximization problem as an optimization problem using statistical physics techniques⁴⁷ (Fig. 3c).

An absolute value of measured nestedness does not have a particular meaning per se. In ecology, one can compare the nestedness of different ecological communities or the empirical value against a suitable null model to determine its statistical significance³⁵. The Z-score is a simple measure that can be used to assess significance of the difference in nestedness temperature between a real network and the average over an ensemble of random networks, expressed in units of standard deviations. We consider a random null model that conserves the degree of rows and columns on average and compute the average temperature and its standard deviation over an ensemble of 30 random matrices for each of the 116 empirical ecological networks. We can then compare the Z-scores obtained by measuring the temperature with our method, with the fitness-complexity map (Fig. 3d) and with BINMATNEST (Fig. 3e). The variations in the computed Z-scores can go up to 0.5 units of standard deviations. Despite not being a drastic change, it can still have consequences in specific cases. For example, the network labeled as “M PL 046” in Web of Life database, move from a $Z = -1.6$ computed with BINMATNEST to a $Z = -2$, possibly crossing a significance threshold.

The best parameters for minimizing the nestedness temperature, γ_{nest} take values in the range $[-1.5, 0]$. Figure 3f shows the scatter plot of this value and the one that maximizes the extinction area for the same matrix. One would naively expect that these two tasks are somehow similar, since they both try to “triangularize” the incidence matrix. However, the exponents for the two tasks take values in different regions of the parameter space, i.e., before and after $\gamma = -1$. Moreover, the values are not correlated (Spearman $\rho = 0.01$, $p = 0.9$).

Finally, an alternative metric of nestedness is often used in the ecological literature with the acronym NODF, which stands for nestedness metric based on overlap and decreasing fill (NODF)⁴⁸. While this measure still depends on the row and column order, its dependency is much weaker. In particular, our algorithm cannot find better rankings to maximize this metric than the simple degree ordering, which our map recovers for $\gamma = 0$ (Supplementary Fig. 11). The fitness-complexity map ($\gamma = -1$) and the two other methods tested in Fig. 3b, c even lead to a slight decrease in this particular metric (Supplementary Fig. 11).

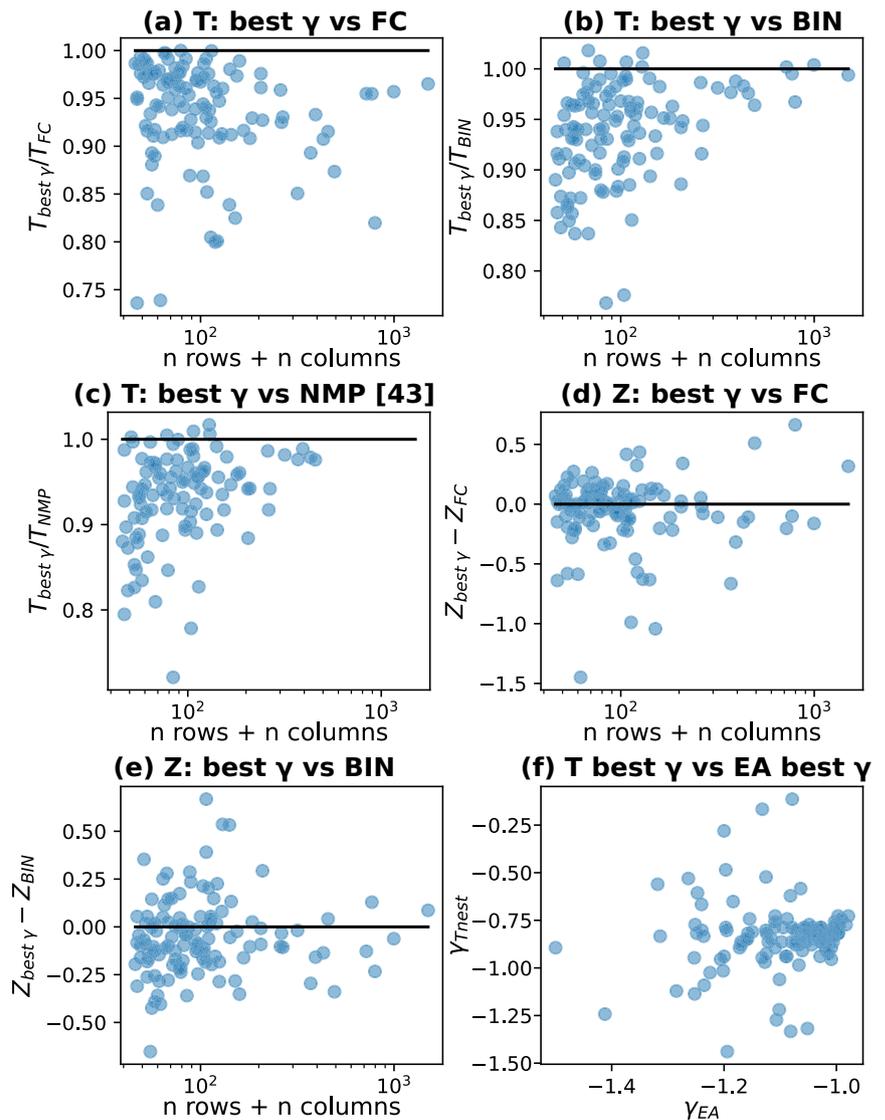
A phase transition suggests the value of γ maximizing the extinction area

This section examines on the stationary behavior of the general map described by Eq. (1) for different values of γ . Previous work^{49,50} provided some analytical insights on map convergence properties as a function of the input matrix for the specific case of $\gamma = -1$ and for a generalization of the fitness-complexity map similar, but not equivalent, to our map (its comparison with our map is postponed to the discussion).

As the number of iterations of the map increases, the scores x_i and y_j tend to an asymptotic behavior which strongly depends on the value of γ . For special choices of the adjacency matrix and of γ , this limiting behavior may even be an oscillatory type. However, the generic situation is that the scores tend to well defined and finite fixed points. Note that the divergence to infinite values is not allowed since the trajectories are normalized at each time step, limiting the maximum value of the score to the number of nodes in the corresponding layer. We observe that the fixed points are all strictly positive for positive parameters ($\gamma > 0$), as well as for negative and small parameters, i.e., $\gamma < 0$ and $|\gamma| \ll 1$. An example is reported in Fig. 4a1. On the other hand, for small enough values of γ , some of the scores can tend to zero as in the examples of Fig. 4a2–4. This latter phenomenology could create an ambiguity in the ranking definition. However, not all the nodes’ scores tend to zero with the same speed, and we can thus rank first the nodes that have a slower score decay, as better explained in Supplementary Note 1.

Interestingly, how the fraction of zero-scores changes as a function of γ strongly resembles a phase transition. This is displayed in Fig. 4a for the particular case of the Robertson 1929 matrix. For exponents close to zero, the score is proportional to the node degree, implying that all the fixed points are positive: $f = 1$. Also when the exponent is decreased from zero to values

Fig. 3 | Nestedness in ecological networks. a ratio of the nestedness temperature between the fitness-complexity map, $\gamma = -1$, T_{FC} and the best temperature found by our algorithm, $T_{best\gamma}$. On the x-axis we show the matrix size as the sum of the number of rows and columns. Each point is an ecological matrix of section “Dataset”. **b** temperature ratio using the ordering of BINMATNEST, from the R package vegan, nestedtemp method. **c** temperature ratio using the ordering obtained from the procedure discussed in ref. 47. We used the algorithm of the repository associated to the paper with an hyperparameter $\beta = 50$. Notice that the algorithm failed for around 10% of the matrices, for which we do not report any point. **d** difference in the temperature Z-score for best-gamma ranking and the fitness complexity ranking. The underlying null model preserves on average the degree of rows and columns. **e** difference in the temperature Z-score for best-gamma ranking and the BINMATNEST ranking. **f** for each matrix we plot the value of γ that maximizes the extinction area, γ -axis, and the one that maximized the nestedness temperature, x -axis.



near -1 , for instance at $\gamma = -0.75$ in Fig. 4a1, all the scores reach positive values. Approaching $\gamma = -1$, a small fraction of trajectories begins decaying to zero, as in Fig. 4a2. Moving from -1 toward a “critical” exponent, the fraction of positive trajectories decreases, as shown in Fig. 4a3, for $\gamma = -1.1$. Finally, after a discontinuous transition, there is a “condensation phenomenon” with most trajectories converging to zero with few exceptions, as in Fig. 4a4 for $\gamma = -1.3$. As it typically happens in phase-transition phenomena, the time required for convergence increases approaching the transition. This abrupt transition from a macroscopic number of positive scores to a fraction of order $1/N$ is present across all the different empirical cases we have considered (see Supplementary Note 4).

Moreover, the behavior of the order parameter as a function of γ in the case of random interaction matrices A_{ij} shows that the jump becomes steeper with the matrix size. This is exactly the expected phenomenology for a phase transition in the thermodynamic limit, i.e., in the large matrix size limit. This allows to define a critical value γ_c which in the thermodynamic limit separates the $f \sim 1$ phase (for $\gamma > \gamma_c$) from the $f \sim 0$ phase (for $\gamma < \gamma_c$). In the empirical cases, this critical value is rounded out due to the finite size of the interaction matrix, as it can be seen in Fig. 4. However, despite finite size effects, the critical γ region is narrow and the transition in the order parameter is still evident.

Even if the underlying dynamics is different, this phase transition strongly resembles the condensation phenomena discussed for instance in

ref. 51 or in ref. 52. The score of one (or few) node takes a macroscopic value, while the other scores are pushed to zero due to the normalization constraint. For this reason, we denote the $f = 0$ phase for $\gamma < \gamma_c$ as the “condensed phase” and the γ_c as the “condensation point”.

Unfortunately, there is no obvious way to predict the value of γ_c for a given empirical interaction matrix. However, the calculation of Supplementary Note 5 derives an upper bound in the limit of large matrices:

$$\gamma_c \leq -1 \tag{4}$$

While we have begun analyzing the relationship between the convergence transition and various properties of the input networks, there remains significant room for further investigation. For the fitness-complexity map and for a different generalization, a previous work⁴⁹ proposed an ansatz for predicting the fraction of scores converging to positive values depending on the final reordering of the incidence-matrix rows and columns. Testing this intuition and its implications in the context of our map is a possible research direction. An alternative approach is to simplify the class of input matrices in the analysis, as done in ref. 50 for studying fitness-complexity convergence using perfectly nested matrices. In a similar vein, in the next section, we consider uniform random matrices, for which we derive a scaling relationship between the transition exponent and certain matrix properties.

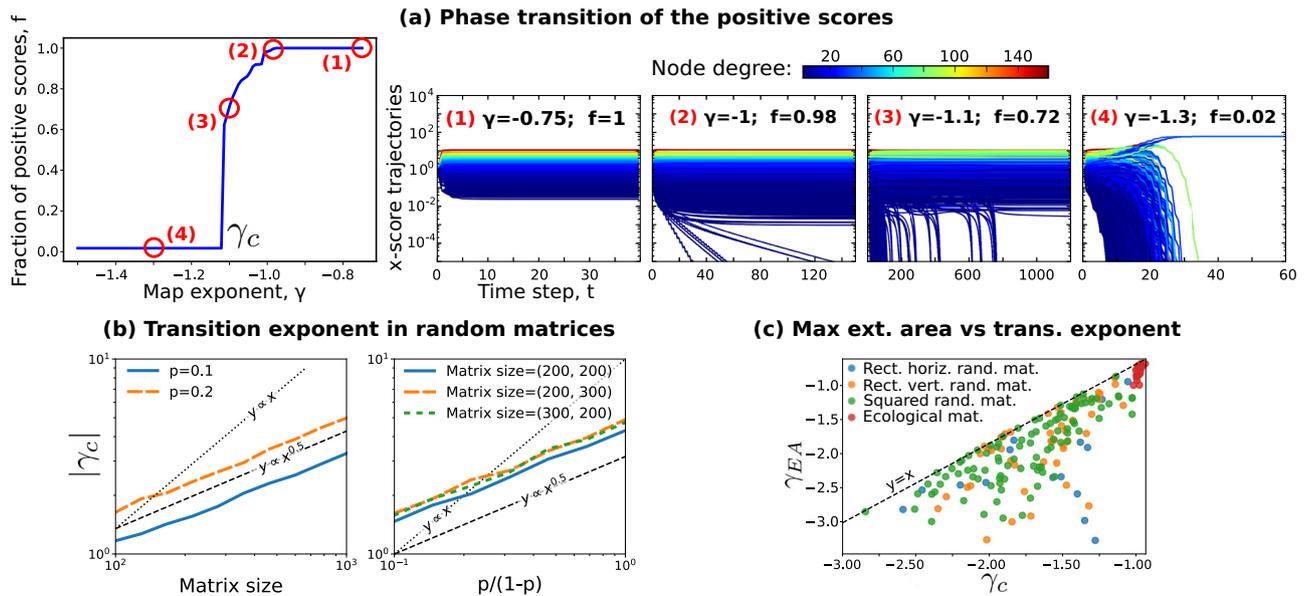


Fig. 4 | Map phase transition. **a** shows the transition of the fraction of scores which converge to positive values of the plant-pollinator matrix Robertson 1929 (see Sec. “Datasets”). The four red circles above the blue line of the plot on the left are associated to the four plots on the right, which show the x -score trajectories for different exponents. The line colors refer to the node degree. **b** displays the absolute value of the transition exponent, γ_c , for the x -score as a function of the size and the density in uniform random matrices. Different line colors and styles represent different matrix densities on the left and different matrix sizes on the right. Each point is

computed as an average over five matrices generated with the same parameters. **c** shows the scatter plot of the exponent which maximizes the extinction area, γ_{EA} and the exponent of the phase transition, γ_c for uniform random matrices and the 20 largest ecological matrices (red points). The other three colors are different shapes of uniform random matrices: horizontal $N = s/1.5$, $M = s^*1.5$ in blue, vertical $N = s^*1.5$, $M = s/1.5$ in orange, and squared $N = M = s$ in green. The considered matrices have been generated with different sizes s , between 50 and 200, and different densities between 0.2 and 0.4.

Phase transition in uniform random matrices. In the case of uniform random matrices, the scaling of the critical map-exponent with the parameters of the incidence matrix can be understood with the following computation. Let us choose the matrix elements to be Bernoulli random variables, such that $A_{ij} = 1$ with probability p and $A_{ij} = 0$ with $1 - p$, where the matrix has size (N, M) . The map starts from initial conditions of ones, $x_i^{(0)} = 1$ and $y_j^{(0)} = 1$, and, after one step, it leads to sums of independent and identically distributed random variables: $\tilde{x}_i^{(1)} = \sum_{j=1}^M A_{ij}$. Using the central limit theorem one can obtain the following expressions:

$$\begin{cases} x_i^{(1)} = 1 + \sqrt{\frac{1-p}{Mp}}\eta \\ y_j^{(1)} = 1 + \sqrt{\frac{1-p}{Np}}\eta, \end{cases} \quad (5)$$

where η is the standard normal random variable ($\langle \eta \rangle = 0$ and $\text{Var}(\eta) = 1$). At the next step, the non-normalized score reads:

$$\tilde{x}_i^{(2)} = \sum_i^M A_{ij} \left(1 + \sqrt{\frac{1-p}{Np}}\eta \right)^y.$$

If $|\gamma|$ is small, the score can be approximated at first order with the following small quantity: $|\gamma| \sqrt{(1-p)/(\min(N, M)p)} \ll 1$. This leads to a summation of variables, both for $x^{(2)}$ and $y^{(2)}$, which have the same first and second moment of the Bernoulli matrix element. As a consequence, by applying again the central limit theorem, the score at the second step has the same distribution of the score at the first step (Eq. (5)), implying that the stationary solution has been reached. The scores follow a Gaussian distribution, and are all positive. Thus for these values of γ the system is in the $f = 1$ phase. We know that for large enough negative values of γ this solution is certainly unstable and the system is in the condensed phase. If we assume that the onset of the condensed phase coincides with the values of γ for which the above approximation is not valid any more, i.e., $|\gamma| \sqrt{(1-p)/(\min(N, M)p)} = O(1)$, we can derive the following scaling

relationships for the critical exponent:

$$\gamma_c \sim -\sqrt{\frac{\min(N, M)p}{1-p}}. \quad (6)$$

Those predictions are verified in Fig. 4b.

The connection between condensation and the maximal extinction area. The condensation of a few scores to positive values seems to be related to the good performance in finding the extinction area. This is shown by Fig. 4c, where the transition exponent, computed according to the procedure described in Supplementary Note 4, is plotted as a function of the exponent that maximizes the extinction area γ_{EA} , which is always located in the condensation phase. Combining this empirical observation with Eq. (4), we can derive the following bound:

$$\gamma_{EA} \leq \gamma_c \leq -1, \quad (7)$$

which can explain the distribution of exponents maximizing the extinction area of different ecosystems in Fig. 2e. Moreover, the bound induces a correlation between the two exponents, leading to similar dependencies on matrix size and density, at least for uniform random matrices. Supplementary Fig. 8 illustrates the scaling behavior of the exponent that maximizes the extinction area with respect to matrix size and density. Indeed, these scalings resemble those of the critical exponent in Fig. 4b. However, this connection between the two exponents does not fully account for certain patterns we observed in empirical ecological networks. For example, Fig. 2f reports a negative correlation between γ_{EA} and matrix density, whereas γ_c shows no significant correlation with density ($\rho = 0.06$, $p = 0.5$). This suggests the presence of more complex relations in presence of structured non-random matrices between the phase transition, the matrix properties, and the system stability captured by the maximal extinction area.

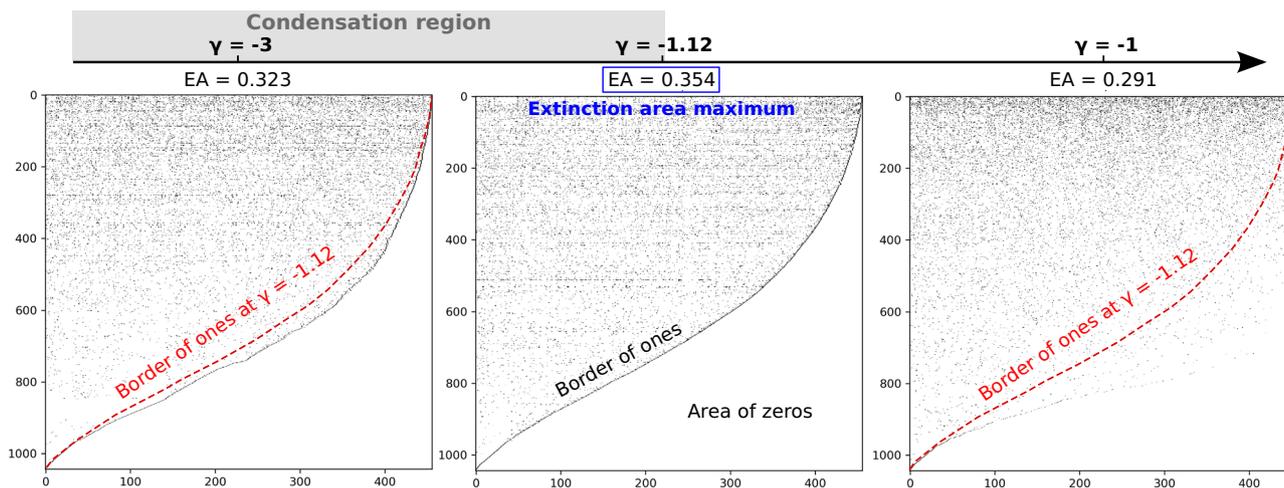


Fig. 5 | Incidence matrix reordered with the map ranking. Examples of matrices with rows and columns sorted according to the x and y ranking for three different map exponents, $\gamma = -3, -1.12, -1$, of the same mutualistic system (Robertson 1292 matrix). A black dot corresponds to the entry 1, while a white dot to the entry 0.

$\gamma \approx -1.12$ is the value which maximizes the extinction area, as shown in Fig. 2d. The exponents less or equal than -1.12 are in the condensation phase, see Fig. 4a, and present the typical border of ones that separates a right-bottom area of zeros from the rest of the matrix.

The condensation implies a triangular shape of the incidence matrix

The shape of the incidence matrix, when ordered according to the map ranking, shows a clear geometric pattern in the condensation phase. Figure 5 shows the matrix corresponding to an ecological example (Robertson 1929), with the ones represented in black and the zero entries in white. Three different rankings are displayed corresponding to three different value of the exponent γ . In the condensed phase, a large area of contiguous zeros appears in the bottom right corner. The area is well separated by the top left area by a continuous border of ones. The calculation of Supplementary Note 5 shows how both these features can be derived from the map (Eq. (1)) with the only assumptions of having the system in the $f=0$ condensed phase, and the solution to be stable.

There is a connection between this packing phenomenology and the extinction area. Supplementary Note 6 proves that the extinction area obtained by removing the rows is a linear function of the area of consecutive zeros from the matrix bottom. Specifically, if we define z_j as the number of consecutive zeros from the matrix bottom of the column j , and $Z = \sum_1^M z_j$ the total area of zeros, the extinction area obtained removing rows, E , reads:

$$E = \frac{1}{N} \left(\frac{Z}{M} + 1 \right), \tag{8}$$

where N and M are the number of rows and columns respectively. A similar relation can be derived also for the transposed quantities. In other words, the extinction area obtained by removing columns (removing nodes from the second set and counting how many nodes of the first one get extinct) is linearly dependent on the area of zeros from the matrix right. These relations connect the extinction area with the visual triangular pattern of the adjacency matrix, and we expect that the ordering with the largest area of zeros (second matrix of Fig. 5) would also have the largest extinction area.

The results of the last two sections suggest that the behavior of the map in the condensed phase drives the ranking of nodes in a way that it “triangularizes” the adjacency matrix, leading to a large area of zeros of the bottom right corner, which also implies a large (almost maximal) extinction area. It is important to stress that what we are able to prove analytically is the correspondence between the area of zeros in the right lower part of the matrix and the extinction area. Moreover, we can prove that in the condensation phase (i.e., for large enough negative values of γ), the matrix takes a triangular shape with a continuous border of ones. However, we cannot prove that close to the transition the map provide an optimal packing of zeros corresponding to the higher extinction area. As a general empirical

fact, in all the studied cases the best extinction area is always reached in the condensation phase and close to the transition point, and therefore for $\gamma < -1$.

Discussion

Bipartite networks are the natural description for various complex systems in technological^{12,53}, social^{54,55}, economical^{26,30}, linguistic¹⁵, and biological^{11,56,57} domains. We introduce a tool to rank nodes in bipartite networks according to different concepts of importance. The scores used for ranking are defined through a non-linear map connecting the two sets of nodes of the network. The map depends on the network incidence matrix and on an exponent γ which dictates the non-linearity, and implicitly defines node importance. By selecting γ , users can tailor the analysis to the specific task. For example, the sign of γ defines if a node is important when connected to high-score neighbors or low-score neighbors. We have shown how specific γ values ($0, \pm 1, \pm\infty$) correspond to well known definitions of node importance. The interpretation of other values of γ is not trivial, but qualitatively the γ value informs about the importance attributed to the scores of connected nodes. More generally, given a metric to measure how good a ranking is, we can look for the γ value that maximizes it. For instance, we examined the problem of identifying the maximal extinction area in ecological interaction matrices, thus ranking the species according to their importance in the ecosystem stability. Analogously, we analyzed when our map can identify species rankings that minimize the nestedness temperature. These two tasks are central for understanding the structure of complex ecosystems and to inform preservation strategies by identifying core species for the system stability. These tasks have been previously effectively addressed using the fitness-complexity map, which correspond to $\gamma = -1$ in our framework^{33,40}. Our results indicate that the optimal γ values range from $[-1.4, -1]$ for the extinction area, and from $[-1, 0]$ for nestedness. The optimal values vary based on the specific task and dataset, making them challenging to predict in advance. However, theory can help in limiting the range of parameters that should be explored. For example, we could analytically set the upper bound at $\gamma_{EA} = -1$ for the extinction area, Eq. (7). Our theoretical argument is based on the presence of a dynamical phase transition for large enough negative values of γ . This transition separates the score space into two phases. For $\gamma > \gamma_c$ all scores remain finite and positive, whereas for $\gamma < \gamma_c$ a “condensed” phase emerges where only one or few scores dominate, pushing the others towards zero through normalization. We first established an upper bound for this transition as $\gamma_c \leq 1$, and then we postulated (and check numerically, Fig. 4) that the extinction area is maximized in the condensed phase, $\gamma_{EA} \leq \gamma_c$. This picture is coherent with the

visual inspection of the incidence matrix ordered according to the ranking in the condensed phase. We proved that the extinction area is proportional to the area of consecutive zeros from the bottom-right of the matrix and, indeed, the matrix takes a clear triangular shape, with a border of ones separating a “large” bottom-right area of zeros.

While we have characterized some relationships between incidence matrix properties, the critical exponent, and the maximization of the extinction area, our understanding remains incomplete, particularly for complex empirical systems. The specific exponents γ_c and γ_{EA} are likely influenced by how structural properties and correlations interact in shaping the system’s robustness. Our flexible framework seems able to capture these relationships by weighting node importance in different ways, depending on the system in analysis. Moreover, the presence of a critical behavior raises the possibility that the model could capture complex non-local structures in its proximity. However, interpreting the final exponent values we identify is not trivial.

The main limitation of our approach with respect to simpler algorithms is that there is an additional hyper-parameter γ that have to be optimized depending on the task. This introduces a trade-off between computational cost and the precision at which the user wants to identify the ranking that, for instance, maximizes the extinction area. For small enough systems, directly setting $\gamma = 1$ is typically a safe choice to maximize the extinction area, but for minimizing the nestendness temperature different exponents can be needed. For large ecosystems, the optimal γ is often different from -1 in both tasks, and thus the cost-precision trade-off has to be considered.

Although we focused on ecological systems, it would be valuable to explore the applicability of our generalized map in other fields. For instance, several results have been obtained using the fitness-complexity algorithm (i.e., the $\gamma = -1$ case) in economic systems^{29–32,58}. In particular, there is an interesting ongoing debate on the meaning and the evaluation of the “complexity” of exported products and on the “fitness” of the exporter nations^{59–61}. Different economic definitions would ideally correspond to different values of γ and can thus lead to different rankings of nations.

Besides its practical applications, our non-linear map represents a relatively simple model with a rich phenomenology and an emergent “dynamical” phase transition that could be explored in more depth with statistical physics tools. For example, understanding the deep connection between the condensation phase transition and the maximum of the extinction area, shown by Fig. 4c, is still an open issue. These two features seem closely connected by the triangular shape of the incidence matrix. However the actual proof that the condensation is necessary for the maximal extinction area (or the maximal that our mapping can provide) is still missing. The possible connection between the condensation we observe and the well studied condensation transition in network theory is also an open problem⁵². Moreover, the recently discovered analogy between the fitness-complexity map and the Sinkhorn–Knopp algorithm⁶² suggests a possible alternative direction to frame and understand the properties of our non-linear map.

Our framework represents only one possible generalization of ranking algorithms based on non-linear maps as the fitness-complexity map²⁹. A previous attempt along this line³⁴ generalized the map in a non-symmetric manner and can actually be easily included in our framework by re-defining the exponents for the x and y node types, i.e., $\gamma_x = -\beta$ and $\gamma_y = -1/\beta$. However, our symmetric choice in Eq. (1) leads, for example, to larger extinction areas as analyzed in more detail in Supplementary Note 7. The behavior of our map for $\gamma \rightarrow -\infty$ discussed in the method section is also reminiscent of the “Minimal Extremal Metric” introduced in ref. 50, where the score of the second node set is given by the minimal score of the first set. However, this extremal metric is obtained as the limit for large γ of the different non-symmetric map discussed above³⁴. In particular, differently from our case, the score on one set of nodes is still computed as in the fitness-complexity map.

A more general extension of our map, which we plan to investigate, involves a two-parameter version where the non-linearity differs between the two node sets, i.e., $\gamma_x \neq \gamma_y$. In this case, the γ_x γ_y plane could be explored to select the best combination of exponents given the task at hand. This

generalization would allow the introduction of different principles of node importance on the two sets with the consequent coupled rankings. The statistical properties of this more general map in terms of convergence and phase transitions could present an even richer phenomenology.

Finally, our approach can also be extended to study node centrality on large multiplex networks^{63,64}. Multiplex networks are playing an increasingly important role in modern big data analyses. Many of these networks could be represented as multi-partite networks and could thus be approached with an extension of our map. Moreover, multiplex networks implicitly have a natural bipartite scheme with the nodes on one side and the layers of the multiplex on the other side. Weighted links connect a node with a layer with a weight proportional to the number (and centrality) of neighboring nodes. Following⁶⁴, we may rank the importance of nodes by increasing the centrality of nodes that receive links from highly influential layers, and symmetrically enhance the influence of a layer if it contains central nodes. This idea naturally leads to an iterative map, but extended to include a weighted adjacency matrix. An additional non-trivial extension of our approach could be developed for temporal networks⁶⁵, adding yet another layer of dynamics to the non-linear map.

Data availability

The values of γ that maximize the extinction area and minimize the nestendness temperature in the ecological networks are reported in the Supplementary Table 1. The corresponding ranking of species can be obtained by using the code in the repository <https://github.com/amazzoli/xymap>.

Code availability

The node scores at the basis of our results can be computed with the software in the repository <https://github.com/amazzoli/xymap>. It contains also the code to generate the panels in Figs. 2d, 4a and 5. The code to generate all the other figures are available upon request.

Received: 31 July 2024; Accepted: 27 March 2025;

Published online: 10 April 2025

References

- Newman, M. *Networks* (Oxford University Press, 2018).
- Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1**, 215–239 (1978).
- Bonacich, P. Power and centrality: a family of measures. *Am. J. Sociol.* **92**, 1170–1182 (1987).
- Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43 (1953).
- Brin, S. & Page, L. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **30**, 107–117 (1998).
- Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41 (1977).
- Kleinberg, J. M. Authoritative sources in a hyperlinked environment. *J. ACM* **46**, 604–632 (1999).
- Newman, M. E. A measure of betweenness centrality based on random walks. *Soc. Netw.* **27**, 39–54 (2005).
- Estrada, E. & Rodriguez-Velazquez, J. A. Subgraph centrality in complex networks. *Phys. Rev. E* **71**, 056103 (2005).
- Borgatti, S. P. & Everett, M. G. A graph-theoretic perspective on centrality. *Soc. Netw.* **28**, 466–484 (2006).
- Mazzolini, A., Gherardi, M., Caselle, M., Lagomarsino, M. C. & Osella, M. Statistics of shared components in complex component systems. *Phys. Rev. X* **8**, 021023 (2018).
- Mazzolini, A. et al. Zipf and heaps laws from dependency structures in component systems. *Phys. Rev. E* **98**, 012315 (2018).
- Mazzolini, A., Colliva, A., Caselle, M. & Osella, M. Heaps’ law, statistics of shared components, and temporal patterns from a sample-space-reducing process. *Phys. Rev. E* **98**, 052139 (2018).
- Lazzardi, S. et al. Emergent statistical laws in single-cell transcriptomic data. *Phys. Rev. E* **107**, 044403 (2023).

15. Gerlach, M., Peixoto, T. P. & Altmann, E. G. A network approach to topic models. *Sci. Adv.* **4**, eaaq1360 (2018).
16. Valle, F., Osella, M. & Caselle, M. A topic modeling analysis of tcga breast and lung cancer transcriptomic data. *Cancers* **12**, 3799 (2020).
17. Valle, F., Osella, M. & Caselle, M. Multiomics topic modeling for breast cancer classification. *Cancers* **14**, 1150 (2022).
18. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440 (1998).
19. Newman, M. E. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci USA* **98**, 404–409 (2001).
20. Cancho, R. F. I. & Solé, R. V. The small world of human language. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **268**, 2261–2265 (2001).
21. Everett, M. G. Centrality and the dual-projection approach for two-mode social network data. *Methodol. Innov.* **9**, 2059799116630662 (2016).
22. Guillaume, J.-L. & Latapy, M. Bipartite structure of all complex networks. *Inf. Process. Lett.* **90**, 215–221 (2004).
23. Guillaume, J.-L. & Latapy, M. Bipartite graphs as models of complex networks. *Phys. A Stat. Mech. Appl.* **371**, 795–813 (2006).
24. Latapy, M., Magnien, C. & Del Vecchio, N. Basic notions for the analysis of large two-mode networks. *Soc. Netw.* **30**, 31–48 (2008).
25. Pavlopoulos, G. A. et al. Bipartite graphs in systems biology and medicine: a survey of methods and applications. *Gigascience* **7**, giy014 (2018).
26. Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proc. Natl. Acad. Sci. USA* **106**, 10570–10575 (2009).
27. Daugulis, P. A note on a generalization of eigenvector centrality for bipartite graphs and applications. *Networks* **59**, 261–264 (2012).
28. Yildirim, M. A. & Coscia, M. Using random walks to generate associations between objects. *PLOS ONE* **9**, 104813 (2014).
29. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. & Pietronero, L. A new metrics for countries' fitness and products' complexity. *Sci. Rep.* **2**, 723 (2012).
30. Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G. & Pietronero, L. Measuring the intangibles: a metrics for the economic complexity of countries and products. *PLoS ONE* **8**, e70726 (2013).
31. Tacchella, A., Mazzilli, D. & Pietronero, L. A dynamical systems approach to gross domestic product forecasting. *Nat. Phys.* **14**, 861–865 (2018).
32. Cristelli, M., Tacchella, A. & Pietronero, L. The heterogeneous dynamics of economic complexity. *PLoS ONE* **10**, e0117174 (2015).
33. Domínguez-García, V. & Muñoz, M. A. Ranking species in mutualistic networks. *Sci. Rep.* **5**, 8182 (2015).
34. Mariani, M. S., Vidmer, A., Medo, M. & Zhang, Y.-C. Measuring economic complexity of countries and products: which metric to use? *Eur. Phys. J. B* **88**, 1–9 (2015).
35. Mariani, M. S., Ren, Z.-M., Bascompte, J. & Tessone, C. J. Nestedness in complex networks: observation, emergence, and implications. *Phys. Rep.* **813**, 1–90 (2019).
36. Patterson, B. D. & Atmar, W. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biol. J. Linn. Soc.* **28**, 65–82 (1986).
37. Ulrich, W., Almeida-Neto, M. & Gotelli, N. J. A consumer's guide to nestedness analysis. *Oikos* **118**, 3–17 (2009).
38. Cutler, A. H. Nested biotas and biological conservation: metrics, mechanisms, and meaning of nestedness. *Landsc. Urban Plan.* **28**, 73–82 (1994).
39. Pocock, M. J., Evans, D. M. & Memmott, J. The robustness and restoration of a network of ecological networks. *Science* **335**, 973–977 (2012).
40. Lin, J.-H., Tessone, C. J. & Mariani, M. S. Nestedness maximization in complex networks through the fitness-complexity algorithm. *Entropy* **20**, 768 (2018).
41. Marlin, J. C. & LaBerge, W. E. The native bee fauna of carlinville, illinois, revisited after 75 years: a case for persistence. *Conserv. Ecol.* **5**, 9 (2001).
42. Hogben, L. *Handbook of Linear Algebra* (Chapman and Hall/CRC, 2013).
43. Cimini, G., Gabrielli, A. & Labini, F. S. The scientific competitiveness of nations. *PLoS ONE* **9**, e113470 (2014).
44. Allesina, S. & Pascual, M. Googling food webs: can an eigenvector measure species' importance for coextinctions? *PLoS Comput. Biol.* **5**, e1000494 (2009).
45. Atmar, W. & Patterson, B. D. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* **96**, 373–382 (1993).
46. Rodríguez-Gironés, M. A. & Santamaría, L. A new algorithm to calculate the nestedness temperature of presence-absence matrices. *J. Biogeogr.* **33**, 924–935 (2006).
47. Mariani, M. S., Mazzilli, D., Patelli, A., Sels, D. & Morone, F. Ranking species in complex ecosystems through nestedness maximization. *Commun. Phys.* **7**, 102 (2024).
48. Almeida-Neto, M., Guimaraes, P., Guimaraes Jr, P. R., Loyola, R. D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**, 1227–1239 (2008).
49. Pugliese, E., Zaccaria, A. & Pietronero, L. On the convergence of the fitness-complexity algorithm. *Eur. Phys. J. Spec. Top.* **225**, 1893–1911 (2016).
50. Wu, R.-J., Shi, G.-Y., Zhang, Y.-C. & Mariani, M. S. The mathematics of non-linear metrics for nested networks. *Phys. A Stat. Mech. Appl.* **460**, 254–269 (2016).
51. Godrèche, C. Condensation for random variables conditioned by the value of their sum. *J. Stat. Mech. Theory Exp.* **2019**, 063207 (2019).
52. Bianconi, G. & Barabasi, A. Bose-Einstein condensation in complex networks. *Phys. Rev. Lett.* **86**, 5632–5635 (2001).
53. Pang, T. Y. & Maslov, S. Universal distribution of component frequencies in biological and technological systems. *Proc. Natl. Acad. Sci. USA* **110**, 6235–6239 (2013).
54. Vasques Filho, D. & O'Neale, D. R. Transitivity and degree assortativity explained: the bipartite structure of social networks. *Phys. Rev. E* **101**, 052305 (2020).
55. Koskinen, J. & Edling, C. Modelling the evolution of a bipartite network-peer referral in interlocking directorates. *Soc. Netw.* **34**, 309–322 (2012).
56. Corel, E. et al. Bipartite network analysis of gene sharings in the microbial world. *Mol. Biol. Evol.* **35**, 899–913 (2018).
57. Valle, F., Caselle, M. & Osella, M. Exploring the latent space of transcriptomic data with topic modeling. *bioRxiv* 2024–10 (2024).
58. Servedio, V. D., Buttà, P., Mazzilli, D., Tacchella, A. & Pietronero, L. A new and stable estimation method of country economic fitness and product complexity. *Entropy* **20**, 783 (2018).
59. Morrison, G. et al. On economic complexity and the fitness of nations. *Sci. Rep.* **7**, 15332 (2017).
60. Sciarra, C., Chiarotti, G., Ridolfi, L. & Laio, F. Reconciling contrasting views on economic complexity. *Nat. Commun.* **11**, 3352 (2020).
61. Balland, P.-A. et al. Reprint of the new paradigm of economic complexity. *Res. Policy* **51**, 104568 (2022).
62. Mazzilli, D., Mariani, M. S., Morone, F. & Patelli, A. Equivalence between the fitness-complexity and the sinkhorn-knopp algorithms. *J. Phys. Complex.* **5**, 015010 (2024).
63. Battiston, F., Nicosia, V. & Latora, V. The new challenges of multiplex networks: measures and models. *Eur. Phys. J. Spec. Top.* **226**, 401–416 (2017).
64. Rahmede, C., Iacovacci, J., Arenas, A. & Bianconi, G. Centralities of nodes and influences of layers in large multiplex networks. *J. Complex Netw.* **6**, 733–752 (2018).
65. Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97–125 (2012).

Acknowledgements

This work has been partially supported by the CRT Foundation, within the framework of the Ordinary Call for Proposal 2022, First Round, for the project “GENPHYS: Statistical Physics for Genomic Data Mining”. We also acknowledge financial support from the Italian “Ministero dell’Università della Ricerca”, PRIN 2022 - COD. 2022PY8MHN - GeCoS: Genomic Component Systems.

Author contributions

A.M., M.C., and M.O designed the project and conducted the theoretical study, and A.M. conducted the data analysis and performed the numerical simulations.

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/s42005-025-02073-6>.

Correspondence and requests for materials should be addressed to Andrea Mazzolini.

Peer review information *Communications Physics* thanks Virginia Domínguez-García and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025