Ranking influential nodes in networks from aggregate local information

Silvia Bartolucci,[1,2] Fabio Caccioli,[1,3,4] Francesco Caravelli,[5] and Pierpaolo Vivo[6,*]

[1]*Department of Computer Science, University College London, 66–72 Gower Street, London WC1E 6EA, United Kingdom*
[2]*Centre for Financial Technology, Imperial College Business School, South Kensington, London SW7 2AZ, United Kingdom*
[3]*Systemic Risk Centre, London School of Economics and Political Sciences, London WC2A 2AE, United Kingdom*
[4]*London Mathematical Laboratory, London WC 8RH, United Kingdom*
[5]*Theoretical Division (T4), Condensed Matter & Complex Systems, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[6]*Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom*

Many complex systems exhibit a natural hierarchy in which elements can be ranked according to a notion of "influence". While the complete and accurate knowledge of the interactions between constituents is ordinarily required for the computation of nodes' influence, using a low-rank approximation we show that—in a variety of contexts—local and aggregate information about the neighborhoods of nodes is enough to reliably estimate how influential they are without the need to infer or reconstruct the whole map of interactions. Our framework is successful in approximating with high accuracy different incarnations of influence in systems as diverse as the WWW PageRank, trophic levels of ecosystems, upstreamness of industrial sectors in complex economies, and centrality measures of social networks, as long as the underlying network is not exceedingly sparse. We also discuss the implications of this "emerging locality" on the approximate calculation of nonlinear network observables.

I. INTRODUCTION

Understanding how complex systems are internally organized and function collectively is a central question in many areas of science, with applications as diverse as hubs detection in epidemiology [1,2] and the understanding of protein-protein interaction networks in biological applications [3] to the study of systemic risk in financial networks [4,5]. One of the most relevant issues is how constituents—for example, nodes in a complex network—can be efficiently *ranked* according to some meaningful *local* or *global* metric [6]. The simplest local observable is the degree k_i of node i, which is a useful, albeit simplistic, way to assess how "important" a node is based only on the number and strength of its immediate contacts. Global measures, such as different types of *centrality*, take a more holistic view of a node in the context of how fast information can travel through it from and towards the outer regions of the network. The downside is that they are usually more difficult to compute accurately and are very sensitive to minute details of how the network is organized [7]. Moreover, there is now quite some evidence that many centrality measures are typically correlated with the local degree [8–12] despite their premise of global character.

*Corresponding author: pierpaolo.vivo@kcl.ac.uk

In the context of quintessentially global measures, a prominent way to classify and rank nodes is based on an umbrella notion that we will refer to as *influence*, which can be defined self-consistently across the whole network. The idea is that a node i's influence $\mathcal{I}_i$ should be slightly larger—on average—than that of the nodes "influenced" by i, i.e., the nodes j in its neighborhood ∂i. In formulas,

$$\mathcal{I}_i = 1 + \langle \mathcal{I}_{j \in \partial i} \rangle . \tag{1}$$

In theoretical ecology, for instance, species forming a complex ecosystem can be assigned a natural "influence" (*trophic*) level in the food-chain hierarchy, precisely computed using (1): apex predators are high-up in the food chain because they either eat top-level predators themselves or because they consume a large number of medium- or low-level species. The notion of influence as defined in (1) appears, however, in a much broader range of incarnations: the PageRank algorithm used by Google [13] classifies web pages as very relevant if they are linked to by other very relevant addresses or by very many other pages. In economics, industrial sectors of different countries can be ranked based on their position and importance in the production chain using so-called *upstreamness* metrics, which measure the distance of a product from final consumption [14–17]. The Katz centrality $\mathcal{K}_i$ of node i, which is a weighted sum of paths traveling through i [18], can be itself cast in the form of (1), as well as the mean first passage time of a walker from a source to a target node [19–21]. These examples support the notion that the "influence" framework is in principle a rather general and intuitive way to address the ranking problem on networks from a global perspective.

Influence is indeed an intrinsically nonlocal quantity, which depends on the full set of interactions in the network. Rearranging slightly (1), we may rewrite the influence vector $\mathcal{I}$ of the N nodes as

$$\mathcal{I} = (\mathbb{1} - A)^{-1} \mathbf{1}, \tag{2}$$

where $\mathbb{1}$ is the identity matrix, $\mathbf{1}$ is a vector of all ones, and A is the *interaction matrix*: Depending on the circumstances, it may simply be the weighted adjacency matrix of the network (denoted by $\tilde{A}$ in the following) or a simple function of it that encodes pairwise interactions between constituents. The matrix $G(A) = (\mathbb{1} - A)^{-1}$ appearing in (2) is referred to as the *resolvent* of A. In many interesting examples—where some form of dissipation or external contribution is involved—the matrix A is *non-negative* and *substochastic* ($0 \leqslant \sum_j A_{ij} \leqslant 1$): Each entry may represent the fraction of calories a species contributes to the diet of another in an ecosystem or the fraction of money an industrial sector injects into the economic cycle.

The seemingly harmless (2) has, however, three important drawbacks: (i) It requires the inversion of a possibly large and ill-conditioned matrix, which makes a numerical approach computationally intensive and prone to inaccuracies [22]; (ii) the nonlinear relation between the influence and the interaction matrix A makes it difficult to infer the functional dependence of the former on network parameters (e.g., the mean degree) from the knowledge of the latter, unless the matrix possesses symmetries or a specific structure; and (iii) it requires the complete and accurate knowledge of *all* the pairwise interactions strengths in the network (encoded in A), which may be not realistically achieved in many real-life settings.

For instance, reconstructing the food web interaction tables is a notoriously difficult task [23], which can be accomplished with some acceptable accuracy only for very small systems, comprising not more than a handful of species [24]. In economic studies, the compilation of input-output tables of developed countries—unraveling the complex web of interconnections between industrial and financial sectors—is routinely plagued with all sorts of sampling and surveying errors [25,26]. Calculating the ranking of constituents (e.g., webpages) in large systems is computationally expensive, although the problem can be solved quite efficiently using recursive iterations: The classical PageRank algorithm is, however, difficult to directly implement and optimize in distributed computation frameworks such as Spark [27], therefore fostering the quest for faster approximate routes [28]. Having a tool to efficiently and accurately rank constituents of a system *we know very little about* would therefore be very useful and of broad applicability.

In this paper, we indeed show that—in a variety of important contexts—the influence attributes of a node i in a complex network can be determined with high accuracy using only *local* and *aggregate* information. In particular, the detailed knowledge of the complete state of the network is often irrelevant. This *emerging locality* of network influence stems from the observation that a non-negative and substochastic interaction matrix A often displays a "large" Perron-Frobenius eigenvalue, which is well separated from the bulk of all the others. When this happens, A can be faithfully approximated by a rank-1 matrix $\hat{A}$, which retains only some local information stored in A.

As long as the network of interactions is "not too sparse", this approximation dramatically reduces the amount of detailed information needed to rank constituents as fast as possible, without compromising accuracy. Moreover, it makes it possible to accurately approximate all sorts of nonlinear functions of the interaction matrix, as we demonstrate below in the case of exponential and power-law centrality measures.

II. RESULTS

A. Rank-1 approximation for influence

Consider the resolvent $G(A) = (\mathbb{1} - A)^{-1}$ of the interaction matrix A of a system of interest and imagine that a detailed knowledge of all the entries of A is not available. We are only able to accurately estimate $2N$ constants $\mathbf{r} = (r_1, \ldots, r_N)$ and $\mathbf{c} = (c_1, \ldots, c_N)$, namely the *sums* of the N rows and columns of A. This scenario is extremely common: In economics, often only aggregate information is available about sectorial outputs, or interbank exposure [29], and the same happens in the study of ecosystems where often only the overall "energy balance" of each species can be accurately estimated [30].

In this scenario, we may construct a simple rank-1 approximation $\hat{A}$ for the matrix A as follows:

$$\hat{A} = \frac{1}{N} \mathbf{g} \mathbf{q}^T = \begin{pmatrix} \frac{g_1 q_1}{N} & \cdots & \frac{g_1 q_N}{N} \\ \vdots & \ddots & \vdots \\ \frac{g_N q_1}{N} & \cdots & \frac{g_N q_N}{N} \end{pmatrix}, \tag{3}$$

where the entries of the column vectors $\mathbf{g} = (g_1, \ldots, g_N)$ and $\mathbf{q} = (q_1, \ldots, q_N)$ can be determined by solving a set of equations imposing the constraint that A and $\hat{A}$ share the same row and column sums

$$r_i = \sum_j A_{ij} \equiv \frac{\sum_k q_k}{N} g_i = \bar{q}\, g_i, \tag{4}$$

$$c_j = \sum_i A_{ij} \equiv \frac{\sum_k g_k}{N} q_i = \bar{g}\, q_j. \tag{5}$$

This yields eventually the unique matrix (see Supplemental Material [31])

$$\hat{A} = \frac{1}{mN} \mathbf{r} \mathbf{c}^T \tag{6}$$

with $m = \frac{1}{N} \sum_{ij} A_{ij} = \frac{1}{N} \sum_j c_j = \frac{1}{N} \sum_i r_i$. The rank-1 matrix $\hat{A}$ in (6) is nothing but the maximum entropy reconstructed matrix (see, e.g., [32,33]) subject to the row and column constraints in (4) and (5); see the Supplemental Material [31]. In the case when A is the unweighted adjacency matrix of a network, replacing A with its rank-1 counterpart $\hat{A}$ corresponds to the *annealed network* approximation [34,35]. Also, the same result could be interpreted as a first-order expansion of the standard fitness model [36,37]. Note that $\hat{A}$ as constructed here would not necessarily be the optimal (e.g., in the sense of Frobenius norm) rank-1 Eckart-Young-Mirsky approximant $\bar{A}$ [38,39] of the original matrix A. However, the construction of $\bar{A}$ requires the full and complete knowledge of A, which of course would make the whole enterprise pointless.

If the only information we have is about row sums, then the corresponding rank-1 matrix is

$$\hat{A} = \begin{pmatrix} \frac{r_1}{N} & \cdots & \frac{r_1}{N} \\ \vdots & \ddots & \vdots \\ \frac{r_N}{N} & \cdots & \frac{r_N}{N} \end{pmatrix}. \qquad (7)$$

Approximating the resolvent—or Green's function—with a low-rank matrix to speed up computations is a customary tool in hard-core condensed matter [40,41] and fluid dynamics [42], while being far less common in the context of complexity studies (see, however, Ref. [43] for a thorough study of the "low-rank" hypothesis in complex systems). It is indeed well known [44] that matrix inversion scales as N^3 in the worst-case scenario of a structure-less dense matrix. This implies that for large matrices, the inversion in (2) is typically hard.

Clearly, $\hat{A}$ has a single nonzero, real, and positive eigenvalue $\hat{\lambda}_1 = \frac{1}{mN} \sum_j r_j c_j$ due to the Perron-Frobenius theorem and $N-1$ zero eigenvalues, and therefore we may expect that this approximation will work better the larger the "spectral gap" of the original matrix A. The spectral gap is defined as $\lambda_1 - \max\{|\lambda_2|, \ldots, |\lambda_{N-1}|\}$, with λ_1 real and <1 being the Perron-Frobenius eigenvalue. Since the spectral gap is usually larger when a graph is denser [45–47], we expect that the approximate formula may not be suitable for "too-sparse" graphs. This expectation is in agreement with earlier findings in the context of mean first passage time of random walks on networks [20] and will be extensively tested below.

In physical terms, the matrix $\hat{A}$ is a "perfectly balanced" version of A, where the interactions are spread out as evenly as possible among the constituents: Under $\hat{A}$, every species would consume the same amount of all the others, and every industrial sector would rely on the same fraction of goods produced by all the others to function.

Armed with this rank-1 approximation, we may now proceed to evaluate the approximate resolvent,

$$G(\hat{A}) = (\mathbb{1} - \hat{A})^{-1} = \mathbb{1} + \frac{\hat{A}}{1 - \frac{1}{mN} \sum_j r_j c_j}, \qquad (8)$$

using the Sherman-Morrison formula [48] for the inverse of a rank-1 matrix, from which it follows that the influence of the ith node is approximated by

$$\mathcal{I}_i \approx 1 + \omega(\hat{\lambda}_1) r_i, \qquad (9)$$

where the coefficient $\omega(\hat{\lambda}_1) = (1 - \hat{\lambda}_1)^{-1} = (1 - \frac{1}{mN} \sum_j r_j c_j)^{-1}$ [or $(1 - \frac{1}{N} \sum_j r_j)^{-1}$ in the case of row-constraints only, a formula that we used in Ref. [49] to study production networks, using techniques from disordered systems] is a simple yet nontrivial function of the Perron-Frobenius eigenvalue of $\hat{A}$. Within our rank-1 approximation, the influence of node i is fully determined by the interplay of (i) *local* and *aggregate* information, namely the sum of incoming weights into node i and (ii) a suitable average of incoming weights across *all* nodes in the network. In cases where the interaction matrix A is proportional to the adjacency matrix, e.g., in the case of Katz centrality, the $\{r_i\}$ and $\{c_i\}$ are obviously proportional to the out-strength and in-strength of node i, respectively. The out-strength $k_i^{(\text{out})}$ of node i is equal to $\sum_j \tilde{A}_{ij}$, whereas the in-strength $k_i^{(\text{in})}$ is equal
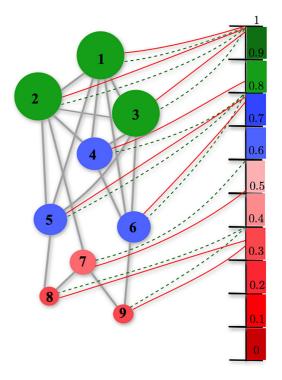

FIG. 1. Illustrative example of the relationship between approximate and exact Katz centrality of nodes. Comparison between the ranking of influential nodes based on Katz centrality [Eq. (12); red solid lines] and our approximate formula [Eq. (9) with only row constraints; green dashed lines] for a small, undirected, and dense network of $N = 9$ nodes (edges are gray solid lines). The color of each node reflects the Katz centrality ranking (normalized to 1), while the node size is proportional to its degree. The matrix $A = \alpha \tilde{A} = (\eta/k_{\max})\tilde{A}$ ($\eta = 0.46$) is proportional to the adjacency matrix $\tilde{A}$ of the network, normalized by its largest degree $k_{\max}$. Both the full and approximate centrality measures are normalized to a maximum value of 1 and compared on the same scale (on the right). As we can see, the rank-1 formula provides a very good approximation to the exact value of the Katz centrality of each node and correctly reflects the "degree centrality" of each node (represented in the sketch by the node size).

to $\sum_j \tilde{A}_{ji}$. If the network is unweighted, then the strengths reduce to the out-degree and in-degree, respectively. Our formula (9) therefore explicitly shows how *global* quantities (in this case the Katz centrality of node i) and *local* ones (in this case its weighted degree) can be strongly correlated (being in fact proportional to each other), a fact that has been previously noted in several network instances [10,33,50–53] (see Fig. 1 for a sketch of the mutual relation between Katz centrality, our approximate formula, and degree centrality).

We further corroborate this observation by studying a class of network transformations (Maslov-Sneppen [54,55]) that preserve the degree sequence while drastically altering the topology, and we show that indeed the Katz centrality of nodes is preserved in the not-too-sparse regime.

It is interesting to compare the rank-1 approximation $G(\hat{A})$ for the resolvent $G(A)$ in Eq. (8) with an alternative approach, which is possibly the simplest approximation or reconstruction scheme one could adopt. The resolvent $G(A)$ can be

formally expanded in a power series as

$$G(A) = \mathbb{1} + A + A^2 + \dots , \qquad (10)$$

which admits an interesting interpretation, e.g., in the economical context: The first contribution accounts for the direct increase in output of all industrial sectors that is necessary to meet an increase in final demand. The second contribution accounts for the increase in output that is needed to meet the increment in input required by all sectors to meet the increase in final demand. This chain of kth order effects is encoded in the kth term of the expansion and unravels the technological interdependence of the productive system within an economy. The crudest approximation to the resolvent consists in truncating the above expansion at a given finite order. As we show in the Supplemental Material [31], however, these approximations do not reproduce the influence rankings as accurately as our approximate formula Eq. (9).

In the next section, we spell out in more detail a few commonly used incarnations of the influence relation (2), before putting the approximate formula (9) to the test on synthetic and empirical networks.

B. Notable examples of influence

In social networks, the influence of nodes can be assessed via the *Katz centrality* [18], which takes into account the total number of walks between a pair of agents. The Katz centrality computes the relative influence of a node within a network by measuring the number of its immediate neighbors (first degree nodes) and also all other nodes in the network that connect to the node under consideration through these immediate neighbors. Connections made with distant neighbors are, however, penalized by an attenuation factor $\alpha < 1$. In formulas, the Katz centrality of the ith node of a network of size N described by the $\{0, 1\}$-adjacency matrix $\tilde{A}$ is given by

$$\mathcal{K}_i = \sum_{\ell=1}^{\infty} \sum_{j=1}^{N} \alpha^\ell (\tilde{A}^\ell)_{ji}. \qquad (11)$$

It measures the total number of paths of given length reaching i from any other node, where paths of length ℓ are given a discounted weight by a factor α^ℓ. Equation (11) can be cast in the "influence" form (2) as

$$\mathcal{K}(A) = (\mathbb{1} - A)^{-1}\mathbf{1} - \mathbf{1}, \qquad (12)$$

with $A = \alpha\tilde{A}$.

Similarly, the PageRank algorithm used by Google [13] returns rankings $\mathcal{R}$ of N websites based on the assumption that more "important" websites are likely to receive more links from other websites. Given a web environment composed of N sites, one forms the rescaled adjacency matrix $\tilde{A}$ is such a way that $\tilde{A}_{ij}$ is the ratio between number of links pointing from page j to page i to the total number of outbound links of page j. Then one defines the *damping factor d* as the probability that an imaginary surfer who is randomly clicking on links will keep doing so at any given step (typically, $d \simeq 0.85$). Given these ingredients, the ranking of pages follows the formula

$$\mathcal{R}(A) = \frac{1-d}{N}(\mathbb{1} - A)^{-1}\mathbf{1}, \qquad (13)$$

where $A = d\tilde{A}$.

Influence measures can also be introduced in the context of hierarchies in complex ecosystems. In an ecosystem composed of N species, one introduces the interaction matrix S_{ij} that represents the fraction of biomass transferred from the species j to the species i, often described as the fraction of j in the "diet" of i. The *trophic levels* of species are the relative positions that they occupy in the ecosystem [23,56–60], i.e., apex predators have a larger trophic level than phytoplankton. The trophic level of species i is defined as

$$\mathcal{T}_i = 1 + \sum_j S_{ij}\mathcal{T}_j, \qquad (14)$$

or, if expressed in the typical "influence" form,

$$\boldsymbol{\mathcal{T}}(A) = (\mathbb{1} - A)^{-1}\mathbf{1}, \qquad (15)$$

with $A \equiv S$. This definition indeed implies that $\mathcal{T}_i = 1$ (the lowest possible value) is reserved to species i such that $A_{ij} = 0$ for all j, i.e., those who occupy the lowest level of the food chain. Trophic levels have been studied in ecology to analyze structural properties of ecosystems [57–65], and more recently the same concept has been applied to investigate and classify financial strategies [66] and to quantify amplification of shocks in financial exposure networks [67], as well as functional and dynamical properties of nodes in networks [68,69], including the effect of global directionality in directed networks [70].

Finally, in the context of complex economies and so-called input-output models [71], measures of influence are used to determine the role and importance of industrial sectors within an economy [49]. We can define the so-called *upstreamness* [14–16] as follows:

$$\mathcal{L}(A) = (\mathbb{1} - A)^{-1}\mathbf{1}, \qquad (16)$$

where the matrix A_{ij} is the dollar amount of sector i's output that is needed to produce one dollar of j's output. The upstreamness of an industry, or economic sector, is used to assess their relative position in the value production chain [16] and their relation to economic growth [17].

We are now ready to put the approximate, *local* formula (9) (and its row-constraints-only counterpart) for each node's influence to the test: In the next section, we assess how well the formulas perform on network synthetic data (in the context of Katz centrality), as well as on different incarnations of the "influence" framework in the context of economics, social networks, relevance of web pages, and ecology, as described above.

C. Test of the formula on data

In this section, we will test how the rank-1 formula with single and double constraints (9) performs on synthetic network data and on six sources of empirical data, considering different incarnations of the influence ranking. Tests include computing the Katz centrality on synthetic Erdős-Rényi (ER) and scale-free (SF) networks, measuring the so-called *upstreamness*, i.e., intersectorial dependencies in complex economies, *Katz centrality* of users of social networks, relevance of webpages using *PageRank*, and *trophic levels* of species within an ecosystem.
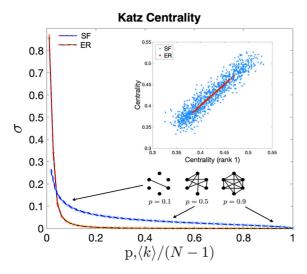
FIG. 2. Katz centrality of Erdős-Rényi and scale-free networks. We plot the σ metric [defined in (17)] as a function of the sparsity parameter p for ER and the equivalent $\langle k \rangle/(N-1)$ for SF networks ($\gamma = 2.2$) of size $N = 1000$, where $\langle k \rangle$ is the mean degree. The curves correspond to the average σ over $M = 3600$ instances. The spread over such realizations is comparable to the thickness of the curves, while the dots mark the values at which the curves have been calculated. In the inset, we provide instead a scatter plot of numerical vs approximate centrality measures for $p = 0.5$.

FIG. 3. Maslov-Sneppen dynamics. Black crosses: Average value of the derivative of the Katz centrality ranking ($\alpha = 0.3$). The average is over 100 initially drawn ER networks with $N = 100$, and the total number of Maslov-Sneppen steps ($T = 100$), and it is plotted as a function of the edge probability p. The red area is the standard deviation of $\partial_t \mathcal{K}_i^{(\ell)}$ across time or instances.

1. Katz centrality of synthetic networks

We first test the rank-1 formula on synthetic data. We consider two types of networks in Fig. 2: (i) directed ER random graphs, with p being the probability of having an edge between any two nodes of the graph and (ii) SF graphs with a power-law degree distribution.

We use $\alpha = 0.3/\bar{d}$, where $\bar{d}$ is the maximum eigenvalue. For Erdős-Rényi graphs, $p = \langle k \rangle/(N-1)$, with $\langle k \rangle$ the mean degree, which can be also used as a measure of sparsity for SF graphs. A scale-free graph is defined by the degree distribution $P(k) \propto k^{-\gamma}$ (here we use $\gamma = 2.2$) and can be constructed using the Bianconi-Barabási algorithm [72,73]. We define the following metric:

$$\sigma = \left\langle \left| \frac{\mathcal{K}_i^{(\mathrm{num})}}{\mathcal{K}_i^{(\mathrm{approx})}} - 1 \right| \right\rangle, \qquad (17)$$

where $\mathcal{K}_i^{(\mathrm{num})}$ is the Katz centrality of the ith node of the mth instance ($m = 1, \ldots, M$), evaluated by numerical inversion (12), $\mathcal{K}_i^{(\mathrm{approx})}$ is the approximate Katz centrality of the same node specializing the rank-1 formula (9), and the average $\langle \cdot \rangle$ is performed over the M instances and the N nodes. In Fig. 2, we plot σ as a function of p for ER and SF graphs averaged over $M = 3600$ instances, where a lower σ clearly indicates a better agreement between the exact and approximate formulas.

We see that for ER graphs, the agreement between the rank-1 approximation of the Katz ranking and the actual value evaluated via matrix inversion is almost perfect, unless the graph is very sparse. For scale-free graphs the agreement is still good, but the noise is again larger for sparser networks. This is in agreement with a similar observation in the context
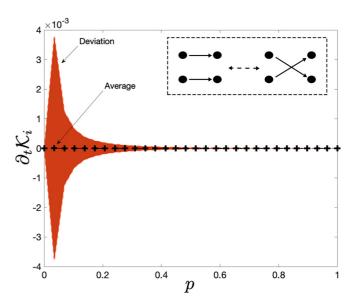
of mean first passage time for random walks on networks [20] that the accuracy of the approximate formula (9) depends on the spectral gap of the interaction matrix, which makes it not suitable for "too sparse" networks.

We now test the fact that the Katz centrality of a node in not-too-sparse networks should be largely determined by the node's strength (or degree) by considering a class of simple degree-preserving transformations that dramatically alter the topology of a network. The Maslov-Sneppen rule [54,55], shown in Fig. 3, works as follows: Given two pairs of nodes i and j and i' and j' of a directed graph, such that the links $i \to j$ and $i' \to j'$ do exist, one performs the rewiring $i \to j'$ and $i' \to j$. Clearly, both the out-strengths of i and i' are unchanged, and so are the in-strength of j and j'. We generate $M = 100$ Erdős-Rényi graphs of size $N = 100$ with edge probability p and perform 100 dynamical (time) steps t. At each time step, we select two pairs of edges at random, and we rewire them, as depicted in the inset of Fig. 3. For each network ℓ in the sample, we calculate the Katz centrality $\mathcal{K}_i^{(\ell)}(t)$ ($\alpha = 0.3$) of each node i. Then, as a function of p and for every sample ℓ and every node, we calculate the discrete time derivative $\partial_t \mathcal{K}_i^{(\ell)}(t) = \mathcal{K}_i^{(\ell)}(t) - \mathcal{K}_i^{(\ell)}(t-1)$. Then, we calculate $\langle \partial_t \mathcal{K}_i^{(\ell)}(t) \rangle = \frac{1}{NM(T-1)} \sum_{t=1}^{T-1} \sum_{i=1}^{N} \sum_{\ell=1}^{M} \partial_t \mathcal{K}_i^{(\ell)}(t)$, which is the average change in the centrality value across nodes, time, and samples. This average clearly depends only on the edge probability p. In Fig. 3, we see that the average derivative is close to zero across the full range of p, while the fluctuations—assessed by the standard deviation across instances and time—also become negligible after $p \approx 0.3$. This confirms that the Katz centrality is strongly correlated to the degree sequence of the network and even more so the denser the network.
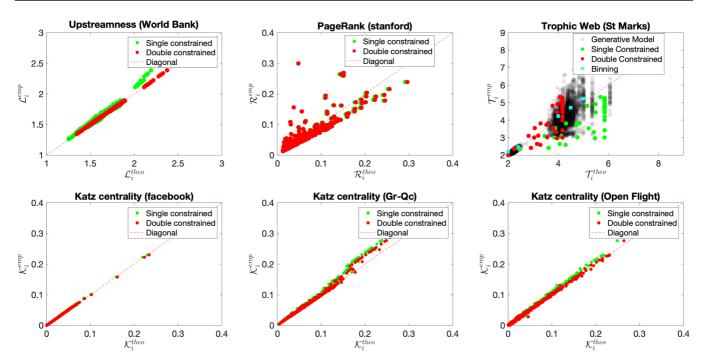
FIG. 4. Test of the approximate formula for influence on real data. Scatter plots of different incarnations of the *empirical* influence [computed from the data interaction matrices using Eq. (2)] vs our approximate formulas for respectively the single- (green dots) and doubly constrained (red dots) [Eq. (9)] models. In the top left figure we compute the upstreamness of industrial sectors in complex economies for the World Bank dataset; in the top center figure, the PageRank of webpages for the 9914-nodes Stanford dataset; and in the top right figure we compute the trophic levels of species in the St. Marks ecosystem. In the bottom row, from left to right, we compute in order the Katz centrality of the Facebook users dataset, the arXiv Gr-Qc collaboration network, and the OpenFlights dataset (details on the datasets are provided in the main text).

2. Katz centrality of social network data

The Katz centrality (11) is also calculated on the widely used Ego-Facebook dataset [74], comprising the matrix of interactions of 4039 users (nodes) and 176 468 friendships (edges). We chose the value of the damping factor $\alpha = 2.87 \times 10^{-4}$, which is the largest admissible value for which (i) the series in (11) converges and (ii) the matrix $\alpha\tilde{A}$ is substochastic. We also evaluated the Katz centrality on the arXiv GR-QC collaboration network, composed of 5242 nodes and 14 490 edges, representing coauthorship on general relativity and quantum cosmology papers [75]. We also consider the OpenFlights network dataset [75,76], a directed graph composed of 2939 nodes (representing airports) and 30 501 edges (weighted by the number of air routes connecting them). Each point in the lower panels of Fig. 4 represents the centrality of one of the nodes. We compare the empirical Katz centrality against the theoretical rank-1 approximation with single (green dots) and double [red dots, see (9)] constraints.

3. Economic dataset

The upstreamness (16) is computed using the data from the National Input-Ouptut Tables (NIOT) of the World Input Output Database (2013 release) [77]. The NIOT dataset represents the flow of money between 35 industrial sectors for 39 world economies, and it includes the years 1995–2011. The full list of countries and economic sectors as well as a scheme of the structure of the input-output tables can be found in Refs. [49,77].

The matrix A from which we calculate the upstreamness is computed from the full input-output table of each country normalizing the entries of the I-O table by the total output (the row sum of the matrix plus the demand of external entities) [49]. Each point in the upper left panel of Fig. 4 represents the average upstreamness (over all sectors) of a given country in a given year [the plot thus includes 39 (countries) × 17 (years) = 663 data points].

4. PageRank

The PageRank (13) is evaluated on the *wb-cs.Stanford* dataset, a collection of 9914 indexed webpages of the World Wide Web [78,79] dating from 2001, for a value of the damping parameter equal to $d = 0.3$ (upper middle panel, Fig. 4).

5. Trophic levels

The trophic levels [defined in (15)] are computed for the 51 species that are part of the St. Marks food web [80] (upper right panel, Fig. 4). Since on average the size of ecosystems does not normally exceed ~200 species, to match the empirical data with the rank-1 approximation we have also constructed the corresponding stochastic niche model in order to generate and reproduce the typical trophic diameter of the food web, as it is customary in this type of analysis [24,81–83]. We used a "cascade," or niche model, extracted from the empirical food web A, not only reproducing the mean and the variance of the empirical data but rather estimating the full distribution [84]. Each black dot in the upper right panel of

Fig. [4] corresponds to one of the $T_k(A_s)$ values ($k = 1, \ldots, 51$ and $s = 1, \ldots, 100$) generated by 100 instances of the random model with $N = 51$ generative species each—plotted against the corresponding theoretical value estimated using [(9)]. While there are sample to sample fluctuations, binning (purple dots in Fig. [4], upper right panel) reveals that the average trophic levels nicely follow the theoretical approximations. More generally, we observe an excellent agreement between our approximate formula—which requires neither matrix inversion nor the full details of the pairwise interaction matrix—and the empirical data in all cases, especially using the doubly constrained version of the influence, Eq. [(9)]. For all datasets, we measured the Pearson's and Spearman's correlation coefficients between the empirical values and those estimated using the approximate formula. We observe correlation coefficients well above 95% for all cases (in fact, above 98% in most instances) with the exception of the St. Marks food web, for which we observe correlations around 0.6 (Spearman's coefficient) and 0.8 (Pearson's coefficient). This further supports the claim that the approximate formula provides quite good estimates of nodes influence in a variety of contexts.

D. Effect of sparsity on the accuracy of the local approximation

In this section, we perform numerical experiments on synthetic data to assess the performance of the local approximation for the influence [Eq. [(9)] specialized to the case of row sums only].

We start from a random $N \times N$ dense matrix $A_N(t = 0)$ (typically, $N = 1000$) with exponentially distributed entries (representing a *fully connected* and weighted initial network) and collect the row sums $\{r_1, \ldots, r_N\}$. Then, at each discrete time $t = 1, 2, \ldots$, we pick one nonzero entry $[(A_N)_{k\ell}]$ at random and set it to zero (unless it was the last nonzero entry of its row). Then we redistribute the nullified $(A_N)_{k\ell}$ entry evenly to all other nonzero entries of its row k to preserve each r_k along the process.

This way we progressively "sparsify"—in a controlled way—the matrix $A_N(t = 0)$ while retaining its initial row sums unchanged. The process ends when only one nonzero entry per row (clearly equal to r_k in row k) survives.

For each matrix $A_N(t)$ produced with this procedure, we compute the "real" influence values $\mathcal{I}(t) = [\mathbb{1} - A_N(t)]^{-1}\mathbf{1}$, which we compare with the approximate values $\hat{\mathcal{I}}$ from Eq. [(9)]. We repeat the above steps over several realizations of the matrix $A_N(t = 0)$ and the sparsification trajectory, and we compute over time as a measure of the estimation error the quantity,

$$\sigma(t) = \left\langle \left| \frac{\mathcal{I}_i(t)}{\hat{\mathcal{I}}_i} - 1 \right| \right\rangle, \tag{18}$$

where the average $\langle \cdot \rangle$ is taken over nodes and realizations.

In the numerical experiment we consider in this section, one instance of $A_N(t = 0)$ is generated as follows: We first create a matrix W of size $N \times (N + 1)$, where all entries are drawn from an exponential distribution of average 1. We then set $A_N(t = 0)_{ij} = W_{ij} / \sum_{k=1}^{N+1} W_{ik}$, with i and j both ranging between 1 and N. In this way, we obtain an $N \times N$ full substochastic matrix (we did set diagonal entries to zero as
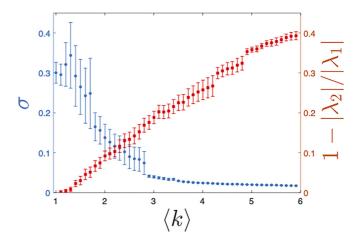

FIG. 5. Sparsification procedure. Left axis: Estimation error associated with the sparsification procedure as a function of the average degree of the network obtained through the sparsification procedure. Averages are computing over 1000 realizations for networks of size $N = 1000$. Right axis: Spectral gap as a function of the average degree computed over 30 realizations of networks of size $N = 1000$. Bars refer to standard errors. The larger the average degree, the smaller the estimation error and the larger the spectral gap.

we do not consider the case of a node in the network pointing to itself).

We report in Fig. [5] results obtained averaging over 1000 instances of networks of size $N = 1000$. In the x axis of the figure, we record the average degree of the network during the different steps of the sparsification procedure, so that time increases moving from the right to the left along the x axis. We observe that the estimation error increases over time as the network becomes sparser, but that it remains fairly low throughout the sparsification procedure—the error remains roughly below 5% up until the initially fully connected network has been reduced to a network of average degree $\langle k \rangle \approx 3$.

In Fig. [5], we also report the spectral gap associated with the networks obtained during the sparsification procedure. As a measure of spectral gap, we consider the quantity $g = 1 - |\lambda_2|/|\lambda_1|$, where λ_1 and λ_2 are the eigenvalues with largest and second largest modulus respectively. Interestingly, and as expected, we find that the behavior of σ is in direct correspondence with the decrease over time of the spectral gap, which follows from the graph becoming progressively sparser, compatibly with the findings in Ref. [20].

These findings confirm that our approximate formula [(9)] provides a very good approximation to the true influence values up to an already-sparse regime characterized by a large percentage of links removed.

E. Beyond influence

The rank-1 approximation for the resolvent allows us to get analytical access to a number of other network observables that can be written as functions $F(A)$ of the interaction matrix A. The key identity is the Dunford-Taylor formula [85], with which we can write a matrix function as the following

integral:

$$F(A) = \frac{1}{2\pi i} \oint_\Gamma dz (z\mathbb{1} - A)^{-1} F(z), \qquad (19)$$

assuming that F is analytic and the domain Γ encircles the spectrum of A.

Approximating A with $\hat{A}$ as in (3) and replacing the resolvent $G(A)$ with $G(\hat{A})$ as defined in (8), we obtain the following approximation (see Supplemental Material [31] for the derivation):

$$F(A) \approx F(0)\mathbb{1} + \frac{F(\xi) - F(0)}{mN\xi} \boldsymbol{rc}^T, \qquad (20)$$

where $\xi = \sum_i r_i c_i / (mN)$.

For instance, many graph invariants, such as topological indices, can be written in the form of a matrix exponential. A nonlocal quantity of interest is the so-called *graph communicability* [86] given by

$$c_{ij} = (e^{\beta\tilde{A}})_{ij}, \qquad (21)$$

where $e^{\beta\tilde{A}} = \sum_{k=0}^{\infty} \beta^k \tilde{A}^k / k!$ is the exponential of the adjacency matrix. This measure accounts for all possible channels of communication between node i and j, giving more weight to the shorter paths connecting them on the network represented by $\tilde{A}$. The communicability of a node i is given by $c_i = \sum_j c_{ij}$, and the total communicability of the graph is defined as the *Estrada index* $EE(\tilde{A}) = \sum_i c_i$. Using (20) we can also provide a natural explanation for the observed correlation between the strengths of nodes (or simply their degrees for unweighted graphs) and the communicability of the network, observed in Ref. [87]. In particular, we can approximate the communicability matrix as follows:

$$e^{\beta\tilde{A}} \approx \mathbb{1} + \frac{\exp\left[\beta \frac{\boldsymbol{k}^{(\text{out})} \cdot \boldsymbol{k}^{(\text{in})}}{\sum_j k_j^{(\text{out})}}\right] - 1}{\boldsymbol{k}^{(\text{out})} \cdot \boldsymbol{k}^{(\text{in})}} \boldsymbol{k}^{(\text{out})} \boldsymbol{k}^{(\text{in})T}, \qquad (22)$$

where $\boldsymbol{k}^{(\text{out})}$ and $\boldsymbol{k}^{(\text{in})}$ are the column vectors of out- and in-strengths of the network, and $\cdot$ denotes the dot product. For uncorrelated networks, the average value of the communicability over an ensemble of networks can therefore be computed—at least in principle—using

$$\langle f(\boldsymbol{k}^{(\text{in})}, \boldsymbol{k}^{(\text{out})}) \rangle = \sum_{\boldsymbol{m},\boldsymbol{n}} f(\boldsymbol{m}, \boldsymbol{n}) \prod_{j=1}^{N} p(m_j, n_j), \qquad (23)$$

where $p(k^{(\text{in})}, k^{(\text{out})})$ is the joint distribution of in- and out-strengths of the nodes.

In order to see how well our approximate formula performs as a function of the connectivity of the graph, we define the metric [analog of (17)]

$$\sigma = \left\langle \left| \frac{c_i^{(\text{num})}}{c_i^{(\text{approx})}} - 1 \right| \right\rangle, \qquad (24)$$

where $c_i^{(\text{num})}$ is the communicability of the ith node of the mth instance ($m = 1, \ldots, M$) of an Erdős-Rényi and a scale-free graph of size $N = 1000$, evaluated using (21), while $c_i^{(\text{approx})}$ is the approximate communicability of the same node obtained from (22). The average $\langle \cdot \rangle$ is performed over the nodes of an instance. In Fig. 6, we plot σ as a function of the edge probability p, where a lower σ clearly indicates a better agreement
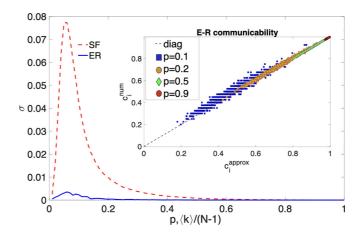

FIG. 6. Communicability of Erdős-Rényi and scale-free networks. Plot of the σ metric for the communicability of Erdős-Rényi (24) and scale-free graphs as a function of the edge probability p and $\langle k \rangle / (N-1)$ for $\gamma = 2.2$ and $\beta = 0.3$. Inset: Communicability c_i of the ith node of Erdős-Rényi network computed using our approximate formula (22) against $c_i^{(\text{num})}$ obtained from Eq. (21) for different edge probabilities p ($N = 1000$, $\beta = 0.3$).

between the exact and approximate formulas. As expected, the denser the graph the better the agreement with our rank-1 approximation. In the inset of Fig. 6 we show a scatter plot of the communicability of the ith node computed using our approximate formula (22) against empirical data for different edge probabilities p.

These results can be easily generalized to further incarnations of exponential measures. These include for example the generalized Estrada index defined as $GE(\tilde{A}) = \text{Tr}(e^{\beta\tilde{A}}) = \sum_i \lambda_i(e^{\beta\tilde{A}})$, where Tr is the matrix trace, originally introduced to measure the degree of folding of a protein and later extended to characterize topological properties of graphs [86,88–92].

Evaluating these types of observable with high accuracy requires nontrivial computational resources for large graphs, as they normally scale as N^3 with the number of nodes [90,91]. In addition, there are not many exact or even approximate formulas in the case of dense graphs, exception made for very specific topologies for which the graph eigenvalues are exactly known. Results for dense *weighted* graphs are even scarcer. In this context, accurate formulas derived from a simple rank-1 approximation may be potentially quite useful.

Our rank-1 approximation using the Dunford-Taylor identity (20) can also be applied to further nonlinear observables in the context of centrality measures. A straightforward extension is for example the application to centrality measures of the form $F(A) = A^k$. Using (20) we obtain $\langle A^k \rangle_{ij} \approx \frac{\xi^{k-1}}{mN}(\boldsymbol{rc}^T)_{ij}$. A notable example for this class of centrality measure is the so-called diffusion centrality [93,94], which is defined in terms of a weighted sum of matrix powers and used in simple models of information diffusion to measure the expected number of times a node receives a given piece of information.

III. DISCUSSION

Constituents of an interacting system organized in a network structure may contribute very differently to the functional tasks the organization is set to achieve as a whole. It is therefore very important to devise meaningful ways to rank them according to useful metrics. Inherently *local* rankings (based, e.g., on the degree of each node) are easy and quick to compile but only provide limited information. *Global* rankings (e.g., centrality measures) take instead into account the full state of the network and provide in principle a more complete picture but are often cumbersome to compute and highly sensitive to the accuracy with which the full set of pairwise interactions are known.

We showed that a large class of nominally *global* rankings given by the "influence" levels $\mathcal{I}_i$ [see (2)] may be accurately estimated using only *local* aggregate information about incoming weights and may not require the full knowledge—or even an accurate reconstruction—of the whole set of pairwise interactions. The accuracy of the approximation depends on the existence of a "large" spectral gap in the interaction matrix: We found that this condition is not too hard to materialize in empirical data sets (as we show in Fig. 4), at least away from the "high sparsity" regime.

Our findings follow from the observation that the matrix $\hat{A}$ that best approximates a given interaction matrix A (in the maximum entropy sense), and shares with A the row and column sums, is rank-1. This provides an immediate route—via the Sherman-Morrison formula [48]—to approximate the resolvent matrix $G(A) = (\mathbb{1} - A)^{-1}$, which appears in the general definition of "influence" [see Eq. (2)] and thus in all its particular incarnations described above. The *emerging locality* that stems from this approximation—and its empirical validity across fields and disparate applications—is particularly appealing for various reasons:

(i) In ecology and economy—among many other fields—a complete and accurate control over the interactions between species and productive sectors is plagued with a wealth of theoretical and practical problems. Our approach offers an avenue to forego the full knowledge of the interaction matrix and determine the set of influence levels with high accuracy using only limited and local information about the activity of each player.

(ii) Centrality measures are known to correlate with the degree of the corresponding nodes. At least in the context of "spectrally gapped" networks, our approach makes this connection clear and transparent, providing an exact expression and clear interpretation for the proportionality constant and allowing for more straightforward analytical treatments.

(iii) As the resolvent $G(A)$ is the building block of any analytic matrix function via the Dunford-Taylor formula (19), the rank-1 approximation $G(\hat{A})$ for the resolvent actually carries over to a number of nonlinear network observables beyond the influence. We have presented examples for exponential and power-law centrality measures and provided evidence that our rank-1 approximation can be very accurate, and more analytically tractable than its original version, at least in the not-too-sparse regime.

A few comments on our framework are in order:

(i) The theoretical results and empirical validation presented here are limited to quantities (such as the Katz centrality) that can be cast in the "influence" form of Eq. (2) and do not necessarily extend beyond this class.

(ii) Our approximate formulas based on a rank-1 approximation are less and less accurate the sparser the network. This, indeed, may appear to be a severe limitation in practice, as many empirical networks are known or believed to be very sparse. As a matter of fact, there is a growing interest in the study of dense networks (both empirical and synthetic), with many interesting applications [95,96], for which our framework is perfectly applicable. Moreover, our experiments with sparser networks demonstrate that our approximate formula remains valid for "moderately" sparse graphs, and—even when it does not capture each of the N values of the influence very accurately—it still provides a meaningful relative ranking of constituents in a fast, efficient, and easily interpretable way.

(iii) If we only have at our disposal the sums of rows and/or columns of the interaction matrix (as we assumed throughout this paper), it is of course impossible to know *a priori* what its spectral gap (or any other meaningful measure of sparsity) is—therefore the expected accuracy of our formula cannot be easily predicted. It is therefore fair to say that the optimal setting in which to employ our formula is when there is some information about the sparsity level of the network under consideration, in addition to the row and column sums information. This limitation is, however, unavoidable and implicit in the observation that many reconstructed networks—with wildly different features—may be compatible with a given aggregate instance, and therefore different reconstruction schemes may perform with different levels of accuracy [29] without any obvious *a priori* criterion to tell them apart.

(iv) Our rank-1 approximation is obtained from "one-point" information about the sums of rows (or of rows and columns), and no information about correlations between such rows or rows and columns sum is injected in it. It is therefore plausible that our approximation may work better on empirical or synthetic networks with "small" degree-degree correlations. Testing this hypothesis goes beyond the scope of this paper, and we leave this as a question for further investigations.

In future studies, it will be important to (i) characterize in a more quantitative way the accuracy of the approximation as a function of the spectral gap, (ii) study more thoroughly the effect of graph heterogeneity on the quality of the approximation, and (iii) explore rank-k (with $k > 1$) systematic approximations for the resolvent.

All raw data analyzed are available via the public repositories cited in the bibliography. The code used to produce the figures and the derived data is available on request from the corresponding author.

The authors declare no competing interests.

[1] D. Bucur and P. Holme, Beyond ranking nodes: Predicting epidemic outbreak sizes by network centralities, PLoS Comput. Biol. **16**, e1008052 (2020).

[2] A. Barrat, M. Barthèlemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, UK, 2009).

[3] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Lethality and centrality in protein networks, Nature (Lond.) **411**, 41 (2001).

[4] S. Battiston, M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli, Debtrank: Too central to fail? Financial networks, the FED and systemic risk, Sci. Rep. **2**, 541 (2012).

[5] M. Bardoscia, P. Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, T. Squartini, and G. Caldarelli, The physics of financial networks, Nat. Rev. Phys. **3**, 490 (2021).

[6] D. Chen, L. Lü, M. S. Shang, Y. C. Zhang, and T. Zhou, Identifying influential nodes in complex networks, Physica A **391**, 1777 (2012).

[7] G. Ghoshal and A.-L. Barabási, Ranking stability and super-stable nodes in complex networks, Nat. Commun. **2**, 394 (2011).

[8] M. E. J. Newman, A measure of betweenness centrality based on random walks, Soc. Netw. **27**, 39 (2005).

[9] S. Oldham, B. Fulcher, L. Parkes, A. Arnatknevičiūtė, C. Suo, and A. Fornito, Consistency and differences between centrality measures across distinct classes of networks, PLoS ONE **14**, e0220061 (2019).

[10] T. W. Valente, K. Coronges, C. Lakon, and E. Costenbader, How correlated are network centrality measures? Connections (Toronto, Ont.) **28**, 16 (2008).

[11] C. Li, Q. Li, P. Van Mieghem, H. E. Stanley, and H. Wang, Correlation between centrality metrics and their application to the opinion model, Eur. Phys. J. B **88**, 65 (2015).

[12] S. Rajeh and H. Cherifi, Ranking influential nodes in complex networks with community structure, PLoS ONE **17**, e0273610 (2022).

[13] S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, Comput. Netw. ISDN Syst. **30**, 107 (1998).

[14] P. Antràs and D. Chor, Organizing the global value chain, Econometrica **81**, 2127 (2012).

[15] P. Antràs, D. Chor, T. Fally, and R. Hillberry, Measuring the upstreamness of production and trade flows, Am. Econ. Rev.: Pap. Proc. **102**, 412 (2012).

[16] T. Fally, Production staging: Measurement and facts, https://www2.gwu.edu/~iiep/assets/docs/fally_productionstaging.pdf.

[17] J. McNerney, C. Savoie, F. Caravelli, V. M. Carvalho, and J. D. Farmer, How production networks amplify economic growth, Proc. Natl. Acad. Sci. USA **119**, e2106031118 (2022).

[18] L. Katz, A new status index derived from sociometric analysis, Psychometrika **18**, 39 (1953).

[19] N. Masuda, M. A. Porter, and R. Lambiotte, Random walks and diffusion on networks, Phys. Rep. **716-717**, 1 (2017).

[20] S. Bartolucci, F. Caccioli, F. Caravelli, and P. Vivo, "Spectrally gapped" random walks on networks: A mean first passage time formula, SciPost Phys. **11**, 088 (2021).

[21] A. Bassolas and V. Nicosia, First-passage times to quantify and compare structural correlations and heterogeneity in complex systems, Commun. Phys. **4**, 76 (2021).

[22] P. S. Dwyer and F. V. Waugh, On errors in matrix inversion, J. Am. Stat. Assoc. **48**, 289 (1953).

[23] J. A. Dunne, Food webs, in *Computational Complexity*, edited by R. Meyers (Springer, New York, 2012).

[24] R. J. Williams and N. D. Martinez, Simple rules yield complex food webs, Nature (Lond.) **404**, 180 (2000).

[25] P. Kop Jansen, Analysis of multipliers in stochastic input-output models, Region. Sci. Urb. Econ. **24**, 55 (1994).

[26] P. Kop Jansen and T. Ten Raa, The choice of model in the construction of Input-Output coefficients matrices, Int. Econ. Rev. **31**, 213 (1990).

[27] Z. Han and Y. Zhang, Spark: A big data processing platform based on memory computing, in *Proceedings of the 7th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP'15)* (IEEE, Los Alamitos, 2015), pp. 172–176.

[28] I. Mitliagkas, M. Borokhovich, A. G. Dimakis, and C. Caramanis, FrogWild! fast pageRank approximations on graph engines, Proc. VLDB Endow. **8**, 874 (2015).

[29] K. Anand *et al.*, The missing links: A global study on uncovering financial network structures from partial data, J. Financ. Stabil. **35**, 107 (2018).

[30] E. L. Berlow *et al.*, Interaction strengths in food webs: issues and opportunities, J. Anim. Ecol. **73**, 585 (2004).

[31] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.033123 for further technical details and derivations, including uniqueness of rank-1 approximation, Maximum Entropy approach, and generalized non-linear formula.

[32] T. Squartini, G. Caldarelli, G. Cimini, A. Gabrielli, and D. Garlaschelli, Reconstruction methods for networks: The case of economic and financial systems, Phys. Rep. **757**, 1 (2018).

[33] G. Cimini, T. Squartini, D. Garlaschelli, and A. Gabrielli, Systemic risk analysis on reconstructed economic and financial networks, Sci. Rep. **5**, 15758 (2015).

[34] G. Bianconi, Mean field solution of the Ising model on a Barabási–Albert network, Phys. Lett. A **303**, 166 (2002).

[35] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Critical phenomena in complex networks, Rev. Mod. Phys. **80**, 1275 (2008).

[36] J. Park and M. E. J. Newman, The statistical mechanics of networks, Phys. Rev. E **70**, 066117 (2004).

[37] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Scale-Free Networks from Varying Vertex Intrinsic Fitness, Phys. Rev. Lett. **89**, 258702 (2002).

[38] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen, Math. Annal. **63**, 433 (1907).

[39] C. Eckart and G. Young, The approximation of one matrix by another of lower rank, Psychometrika **1**, 211 (1936).

[40] L. Zeng, Y. He, M. Povolotskyi, X. Liu, G. Klimeck, and T. Kubis, Low rank approximation method for efficient Green's function calculation of dissipative quantum transport, J. Appl. Phys. **113**, 213707 (2013).

[41] R. Yokota, H. Ibeid, and D. Keyes, Fast multipole method as a matrix-free hierarchical low-rank approximation, in *International Workshop on Eigenvalue Problems: Algorithms, Software and Applications in Petascale Computing* (Springer, Berlin, 2015), pp. 267–286.

[42] L. Lestandi, Numerical study of low rank approximation methods for mechanics data and its analysis, J. Sci. Comput. **87**, 14 (2021).

[43] V. Thibeault, A. Allard, and P. Desrosiers, The low-rank hypothesis of complex systems, arXiv:2208.04848.

[44] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Is matrix inversion an N^3 process? in *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, UK, 1992), pp. 95–98.

[45] D. M. Cvetković, P. Rowlinson, and S. Simić, *An Introduction to the Theory of Graph Spectra*, Vol. 75 (Cambridge University Press, Cambridge, UK, 2010).

[46] P. Van Mieghem, *Graph Spectra for Complex Networks* (Cambridge University Press, Cambridge, UK, 2010).

[47] V. A. R. Susca, P. Vivo, and R. Kühn, Second largest eigenpair statistics for sparse graphs, J. Phys. A: Math. Theor. **54**, 015004 (2021).

[48] J. Sherman and W. J. Morrison, Adjustment of an inverse matrix corresponding to a change in one element of a given matrix, Ann. Math. Stat. **21**, 124 (1950).

[49] S. Bartolucci, F. Caccioli, F. Caravelli, and P. Vivo, Inversion-free Leontief inverse: Statistical regularities in input-output analysis from partial information, arXiv:2009.06350.

[50] T. S. Evans and B. Chen, Linking the network centrality measures closeness and degree, Commun. Phys. **5**, 172 (2022).

[51] C. Lozares, P. López-Roldán, M. Bolibar, and D. Muntanyola, The structure of global centrality measures, Int. J. Soc. Res. Methodol. **18**, 209 (2015).

[52] S. Fortunato, M. Boguñá, A. Flammini, and F. Menczer, Approximating PageRank from In-Degree, in *Algorithms and Models for the Web-Graph (WAW'06)*, Lecture Notes in Computer Science, Vol. 4936, edited by W. Aiello, A. Broder, J. Janssen, and E. Milios (Springer, Berlin, 2008).

[53] M. Cui, M. S. Mariani, and M. Medo, Algorithmic bias amplification via temporal effects: The case of PageRank in evolving networks, Commun. Nonlin. Sci. Numer. Simul. **104**, 106029 (2022).

[54] K. Sneppen and S. Maslov, Specificity and stability in topology of protein networks, Science **296**, 910 (2002).

[55] S. Maslov, S. Krishna, T. Y. Pang, and K. Sneppen, Toolbox model of evolution of prokaryotic metabolic networks and their regulation, Proc. Natl. Acad. Sci. USA **106**, 9743 (2009).

[56] B. Hannon, The structure of ecosystems, J. Theor. Biol. **41**, 535 (1973).

[57] J. T. Finn, Measures of ecosystem structure and function derived from analysis of flows, J. Theor. Biol. **56**, 363 (1976).

[58] S. Levine, Several measures of trophic structure applicable to complex food webs, J. Theor. Biol. **83**, 195 (1980).

[59] S. Allesina and M. Pascual, Googling food webs: Can an eigenvector measure species' importance for coextinctions? PLOS Comput. Biol. **5**, e1000494 (2009).

[60] M. Scotti, C. Bondavalli, A. Bodini, and S. Allesina, Using trophic hierarchy to understand food web structure, Oikos **118**, 1695 (2009).

[61] M. Scotti and F. Jordán, Relationships between centrality indices and trophic levels in food webs, Commun. Ecol. **11**, 59 (2010).

[62] P. W. Price, C. E. Bouton, P. Gross, B. A. McPheron, J. N. Thompson, and A. E. Weis, Interactions among three trophic levels: Influence of plants on interactions between insect herbivores and natural enemies, Annu. Rev. Ecol. Syst. **11**, 41 (1980).

[63] S. Johnson, V. Domínguez-García, L. Donetti, and M. A. Munoz, Trophic coherence determines food-web stability, Proc. Nat. Ac. Sci. U.S.A. **111**, 50 (2014).

[64] B. Saint-Béat, D. Baird, H. Asmus, R. Asmus, C. Bacher, S. R. Pacella, G. A. Johnson, V. David, A. F. Vézina, and N. Niquil, Trophic networks: How do theories link ecosystem structure and functioning to stability properties? A review, Ecol. Indicat. **52**, 458 (2015).

[65] C. Borrvall *et al.*, Biodiversity lessens the risk of cascading extinction in model food webs, Ecol. Lett. **3**, 131 (2000).

[66] M. P. Scholl, A. Calinescu, and J. D. Farmer, How market ecology explains market malfunction, Proc. Natl. Acad. Sci. USA **118**, e2015574118 (2021).

[67] B. Sansom, S. Johnson, and R. S. MacKay, Trophic incoherence drives systemic risk in financial exposure networks (No. 39), Rebuilding Macroeconomics Working Paper (2021), https://www.rebuildingmacroeconomics.ac.uk/economies-stable.

[68] R. S. MacKay, S. Johnson, and B. Sansom, How directed is a directed network? R. Soc. Open Sci. **7**, 201138 (2020).

[69] C. Shuaib *et al.*, Trophic analysis of a historical network reveals temporal information, Appl. Netw. Sci. **7**, 31 (2022).

[70] N. Rodgers, P. Tino, and S. Johnson, Influence and influenceability: Global directionality in directed complex networks, arXiv:2210.12081.

[71] W. Leontief, *Input-output Economics* (Oxford University Press, Oxford, 1986).

[72] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science **286**, 509 (1999).

[73] G. Bianconi and A.-L. Barabási, Competition and multiscaling in evolving networks, Europhys. Lett. **54**, 436 (2001).

[74] J. McAuley and J. Leskovec, Learning to discover social circles in ego networks, in *Advances in Neural Information Processing Systems*, edited by F. Pereira, C. J. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., 2012), Vol. 25.

[75] R. Rossi and N. Ahmed, The network data repository with interactive graph analytics and visualization, in *Proceedings of the 29th AAAI Conference on Artificial Intelligence* (2015), Vol. 29.

[76] J. Leskovec, J. Kleinberg, and C. Faloutsos, Graph evolution: Densification and shrinking diameters, ACM Trans. Knowl. Discov. Data **1**, 2-es. (2007).

[77] M. P. Timmer, E. Dietzenbacher, B. Los, R. Stehrer, and G. J. de Vries, An illustrated user guide to the world input–output database: The case of global automotive production, Rev. Int. Econ. **23**, 575 (2015).

[78] A. D. Gleich, L. Zhukov, and P. Berkhin, *Fast Parallel PageRank: A Linear System Approach*. Yahoo! Research Technical Report YRL-2004-038 (2004).

[79] J. Hirai, S. Raghavan, H. Garcia-Molina, and A. Paepcke, Webbase: A repository of web pages, Comput. Netw. **33**, 277 (2000).

[80] D. Baird, J. Luczkovich, and R. R. Christian, Assessment of spatial and temporal variability in ecosystem attributes of the St Marks National Wildlife Refuge, Apalachee Bay, Florida, Estuar. Coast. Shelf Sci. **47**, 329 (1998).

[81] J. E. Cohen, F. Briand, and M. C. Newman, *Community Food Webs: Data and Theory*, *Biomathematics*, Vol. 20 (Springer, Berlin, 1990).

[82] S. L. Pimm, J. H. Lawton, and J. E. Cohen, Food web patterns and their consequences, Nature (Lond.) **350**, 669 (1991).

[83] D. Tilman, Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly, Proc. Natl. Acad. Sci. USA **101**, 10854 (2004).

[84] R. E. Ulanowicz, *Growth and Development: Ecosystems Phenomenology* (Springer Science & Business Media, New York, 2012).

[85] N. J. Higham, *Functions of Matrices: Theory and Computation* (SIAM, Philadelphia, PA, 2008).

[86] E. Estrada, *The Structure of Complex Networks* (Oxford University Press, Oxford, 2011).

[87] E. Estrada and N. Hatano, Communicability in complex networks, Phys. Rev. E **77**, 036111 (2008).

[88] E. Estrada, Characterization of 3D molecular structure, Chem. Phys. Lett. **319**, 713 (2000).

[89] M. E. J. Newman, The structure and function of complex networks, SIAM Rev. **45**, 167 (2003).

[90] E. Estrada, N. Hatano, and M. Benzi, The physics of communicability in complex networks, Phys. Rep. **514**, 89 (2012).

[91] M. Benzi and C. Klymko, Total communicability as a centrality measure, J. Complex Netw. **1**, 124 (2013).

[92] M. Aprahamian, D. J. Higham, and N. J. Higham, Matching exponential-based and resolvent-based centrality measures, J. Complex Netw. **4**, 157 (2016).

[93] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, The diffusion of microfinance, Science **341**, 1236498 (2013).

[94] F. Bloch, M. O. Jackson, and P. Tebaldi, Centrality measures in networks, arXiv:1608.05845.

[95] O. T. Courtney and G. Bianconi, Dense power-law networks and simplicial complexes, Phys. Rev. E **97**, 052303 (2018).

[96] F. Ma, X. Wang, P. Wang, and X. Luo, Dense networks with scale-free feature, Phys. Rev. E **101**, 052317 (2020).