Assessing the importance of species in perturbed food webs through graph embedding

Dalla Riva, Giulio V. Minh, Tang Priebe, Carey E.

January 5, 2015

Abstract

Food-web ecologists have developed and adopted a variety of species importance measures. Here, we introduce and explore a novel measure of this importance, the *embedding influence*, which is based on the food-webs Random Dot Product Graph model and is robust to small mispecification in the food-web fine wiring. The embedding influence enable us to predict the overall structural effect of the removal of a species from a food web, e.g., following an extinction event. We show that the distribution of the embedding influence among the species in a food web exhibit an evolutionary signature.

KEYWORDS: Food Webs, Adjacency-Spectral Graph Embedding, Embedding influence, Node Centrality, Network Stability, Trophic Niche Dimensionality.

1 Introduction

A food web is the ecological network of who-eats-whom in a community of species: the nodes are species and the arrows between nodes indicate the existence of a trophic relationship, e.g., a predation interaction. A food web can be formalized as a random graph model realization, so that the study of ecosystems can benefit from tools developed in the broad area of complex-network theory.

Biodiversity [31] has a major relevance for the planet, as recognized by the International Union for Conservation of Nature and the United Nations [15]. The increasing human environmental footprint is increasing the extinction risks of single species and directly affecting the structure of food webs [**?**, 3, 7, 19, 20]. The two processes, far from being independent, amplify one the other [24]. This will imply severe economical and social consequences [6].

Moreover, budget limitations focus conservation efforts on a selected number of species. It is no surprise, then, that the identification of the species having a crucial role in ecological networks—those species whose extinction or functional loss will most affect the ecosystem—is an important and difficult task for ecologists. This task is made more difficult from the complex and eventually chaotic dynamical behaviour of food webs [**?**, 18, 25].

In this setting, ecologists have originally developed or adopted from other fields various approach to measure the influence of a species in a food web [10,17] which span most of the centrality measures based on network structure [4, 5, 12–14]. All these measures, however, rely on a rigidly deterministic approach to food webs, albeit the increasing evidence that ecological networks are inherently stocastic [**?**, **?**].

In a plea for groups, [1] proposed to focus on the functional grouping of species in food web, as recently applied to the Serengeti food web [2].

Random Dot Product Graph A natural framework for network partitioning is the stochastic blockmodel graph theory, originally developed for undirected social network analysis [16] and subsequently generalised to directed graphs [30]. Under this model, each of the n species of a food web are assigned to one of K blocks whose within and between linking probability are given by the model parameters. We do not observe the assignment of the species; rather, we observe the realized adjacency matrix. It has been proved that a consistent block estimator based on the spectral partitioning of the normalized Laplacian of the adjacency matrix exist [22], that is the proportion of nodes assigned to the wrong group by the estimator converge in probability to zero as the size of the graph grows to infinity. We will present the definitions in the context of *Random Dot Product* graphs [11], which can be seen as a particular example of stochastic blockmodels.

We will consider the generalization of the latter result to the spectral partitioning of adjacency matrices [28]. Here we propose a measure, called the *embedding influence*, which express the difference between the original

food web and the food web derived by the loss, or non-observation, of one of the species. The method doesn't require the knowledge of the number of the blocks but just an upper bound on the rank of the communication probability matrix. The computation of the embedding influence of each node via simple brute force is feasible for food webs of usual size.

In summary, the proposed method consists in computing the estimated parameter matrix for the original observed adjacency matrix, as shown in [11], removing the species i from the graph embedding and computing the distance of the reduced matrix to the estimated parameter matrix for the modified adjacency matrix obtained by removing from the original food web the species i.

We will apply it to two compilations of the Serengeti food web, namely [2] and [9], the Caribbean marine food web in its original compilation [21] and to the food web we derived from the latter clustering the species in their families.

Finally we will discuss a possible interpretation of the results in terms of species loss, wherein the influence represents how much a food web change its geometry after the lost species extinction, and in terms of non observation of a one species due to possible sampling error during the food web compilation (see [23, Section 11.2]), wherein the influence represents how much a statistical description of the food web may be different.

2 Method

2.1 Graphs and embedding

A food web shows the "resource $\rightarrow$ consumer" relationship between species in a certain ecosystem. We will also use, in a similar fashion, the "prey $\rightarrow$ predator" jargon. Formally, a food web is a directed graph $G = (V, E)$ which vertices V are the species of the ecosystem and the links $e \in E$ are ordered couples between two species v_i, v_j, such that a link $e = (v_i, v_j)$ exists if and only if v_i is a resource of v_j, that is v_j feeds on v_i (or, in other words v_i is a prey of v_j and v_j preys upon v_i). In general food webs may exhibit loops, i.e. species that prey upon themselves; symmetrical linkings where both (v_i, v_j) and (v_j, v_i) are observed in E, i.e. two species mutually preying one on the other; and closed path, although these are rare cases and are omitted for simulation purposes. More generally, the abundance of *motifs* (small sized

subgraphs [27]) in empirical observed graphs is consistently different from the one which could be expected by most of the random graph null models [26].

The adjacency matrix of a graph G is the matrix $A_G \in \{0,1\}^{n\times n}$ where the ith-jth element $A_G^{(i,j)}$ is 1 if v_i preys upon v_j and is 0 otherwise.

2.2 Random dot product graphs

Consider a food web $G = (V, E)$, defined as above, with n species, that is $|V(G)| = n$.

In the random dot product model to each species is assigned a *latent vector* of dimension d and the probability that an edge $v_i \to v_j$ is given by the dot product of the latent vectors of v_i and v_j

Let $R = [R_1, R_2, \ldots, R_n]^T$ and $L = [L_1, L_2, \ldots, L_n]^T$ be $n \times d$ random matrices, where $R_i, L_i \in \mathbb{R}^d$ for all i are respectively the latent vector of species v_i as a predator and as a prey (we could say that the predation relationship are driven by d traits). The latent vectors R_i, L_i are such that $\mathbb{P}(\langle L_i, R_j\rangle \in [0,1]) = 1$ for all i, j.

Conditioned on L and R the entry of the random adjacency matrix A_G are independent and $A_{Gi,j}$ is a Bernoulli random variable of parameter $\langle L_i, R_j\rangle$. Hence, rembembering that $A_{Gi,j}$ takes value in $\{0,1\}$, we have that the probability of observing a certain graph G (which adjacency matrix is $A_{Gi,j}$) is:

$$\mathbb{P}(A_G|L,R) = \mathbb{P}(A_{Gi,j}|L_{i,j}, R_{i,j})$$

$$\mathbb{P}(A_G|L,R) = \prod_{i\neq j} \langle L_i, R_j\rangle^{A_{Gi,j}} (1 - \langle L_i, R_j\rangle)^{1-A_{Gi,j}}$$

A rank d embedding of a matrix $M \in \mathbb{R}^{n\times n}$ is the given of a pair of $\mathbb{R}^{n\times d}$ matrices $\hat{L}_A$ and $\hat{R}_A$ such that the couple $(\hat{L}_A, \hat{R}_A)$ minimizes the Frobenius norm of the difference $M - LR^T$ between all the matrices L and R in $\mathbb{R}^{n\times d}$.

In particular, when a food web is a realization of a random dot product model and A_G is its observed adjacency matrix, $\hat{L}_{A_G}$ and $\hat{R}_{A_G}$ are an estimate of the latent vectors of the species in the random graph.

Consider a singular value decomposition of A_G, that is $A_G = LSR^T$. Hence L and R are real, orthogonal $n \times n$ matrices and S is a $n \times n$ diagonal matrix which non increasingly ordered entries $\Sigma_1, \Sigma_2, \ldots, \Sigma_n$ are the singular values of A_G. Let R' be the $n \times d$ matrix given by the first d columns of R, L' be the $n \times d$ matrix given by the first d columns of L and $\sqrt{S'}$ be the diagonal matrix defined by the square root of the d greatest singular values

There may be another couple of relevant papers published in Ecology: still searching them.

of A_G. Define $\hat{L}$ as $L'\sqrt{S'}$ and $\hat{R}$ as $R'\sqrt{S'}$. Hence the pair of matrices $\hat{L}$ and $\hat{R}$ are an embedding of A_G.

2.3 Embedding influence

Let's consider the parameter matrix X of the food web G, so that $G \sim \text{rdpg}(X)$ or, in other words, $X = LR^T$ where $L, R \in \mathbb{R}^{n \times d}$ are respectively the latent vector of the species in G as preys and as predators.

Consider a is suitable rank d as fixed (we will see that, in our result, the embedding influence measure is relatively robust to the choice of d) and define $\hat{X}$ the rank d adjacency spectral estimate of X computed on A_G.

Consider the graph $G_{-i} = G \setminus i$, obtained removing the ith row - ith column from the adjacency matrix A_G, which adjacency matrix will be denoted A_{G-i}. Define $\hat{X}_{G-i}$ as the rank d adjacency spectral estimate of the parameter matrix computed on A_{G-i}.

In a similar way, define $\hat{X}^{-i}$ as the set of embedded points obtained deleting the ith point from the embedding $\hat{X}$ computed on A_G.

This introduction does not result clear to an ecologist - i.e., what's the meaning of $\hat{X}_{G-i}$ and $\hat{X}^{-i}$?

Definition 2.1. Embedding influence. The rank d **embedding influence** of a species v_i is the size of the difference between the original rank d adjacency spectral estimate parameter matrix estimated and the rank d adjacency spectral estimate parameter matrix computed on A_{G-i}, that is

$$\text{influence}(v_i) := s(i) := ||\hat{X}_{G-i} - \hat{X}^{-i}||_F$$

3 Results

We will say that a species is in the k *stronger influenceing* species set if its embedding influence is equal to or greater than the kth last element of the increasingly ordered sequence of embedding stresses. In particular the *strongest influenceing* is the species with highest embedding influence.

We computed the embedding influence sequence, identifying the stronger influenceing species, for three different food webs: the Serengeti food webs as published by [2] and [9] and the Caribbean marine ecosystem food web as published by [21].

3.1 Serengeti food webs

Baskerville et al. The Serengeti food web compiled by [2] is populated by 161 species and 592 predation relationship.

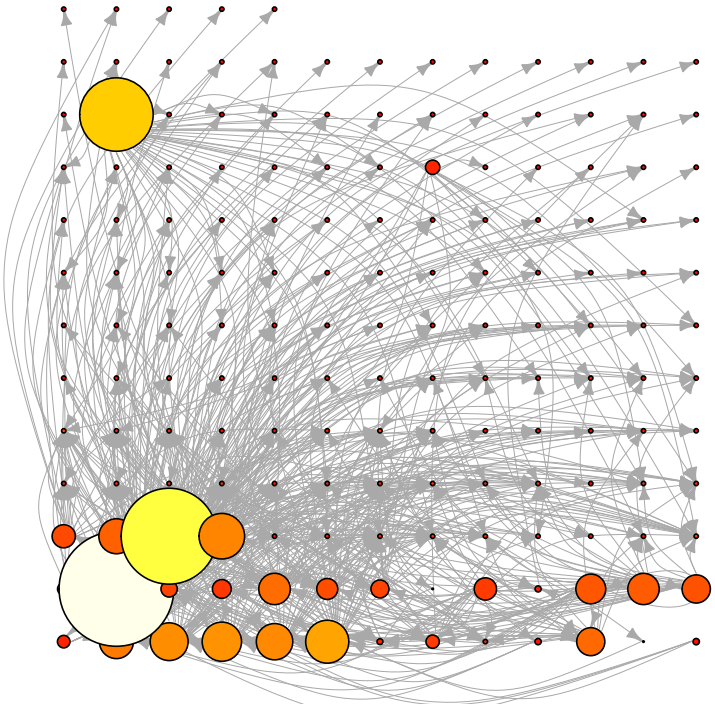

Figure 1: The Serengeti food (*Baskerville et al.*). Node sizes show out-degree / in-degree ratio; node colours show node degree (as heat map).

Of the species studied by Baskerville et al. 129 are plants, 23 herbivores and 9 carnivores. Most of the links, 507, are between herbivores species and plants, while 85 are between animal species. The connectance of the graph is equal to 0.023 is relatively small, compared to other food webs, and that was explained by the authors as a consequence of the higher taxonomic resolution of the plants species. The diameter of the food web as a directed graph is 2 (is 6 as a undirected graph).

Baskerville et al. computed a stochastic block partition of the food web, in a Bayesian framework where data and parameters are uncertain, where groups could have low or high within-connectance and between-connectance.

The consensus partition of the groups identified 14 groups: two of carnivores, four of herbivores and eight of plants.

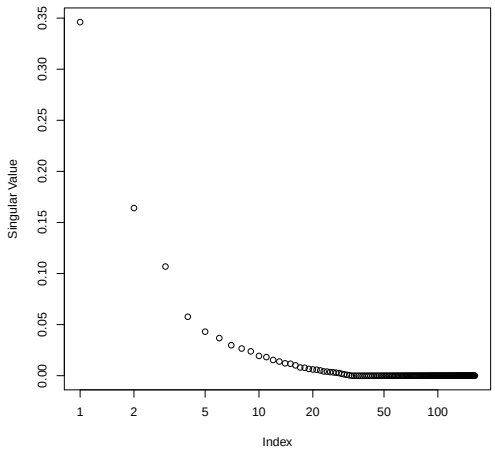

Figure 2: Singular values of the adjacency matrix for the Baskerville's Serengeti food web.

From the sequence of the singular value of the Serengeti adjacency matrix, shown in figure 3.1, we estimated that the optimal ranking dimension for the food web graph is between $d = 2$ and $d = 4$

An ecologist may rise concerns how we choose the right ranking dimension.
My two cents: the choice can be based on the comparison between the SVD sequence of the food web analysed and a suitable neutral model random network. Here a *suitable neutral model* is not a food wev theoretical model but a network where each node plays exactly the same role, has the same centrality and so on – e.g., a toroidal lattice, an Erdös-Renyi model, ... We consider those singular values that has a greater gap than what expected by the neutral model.
Do you think it may be a sensible approach?

In ranking dimension 2, the embedding influence has mean equal to 0.02677, median equal to 0.00164 and variance equal to 0.01665. The 3 stronger influenceing species are Procavia Capensis (the rock hyrax which embedding influence is equal to 1.3254, that is around 50 times the mean), Heterohyrax

Figure 3: The three strongest embedding influence species (from the left *Procavia Capensis*, *Heterohyrax Brucei*, *Loxodonta Africana*) in the Serengeti National Park based on the Baskerville et al. food web.

Brucei (the yellow-spotted rock hyrax which embedding influence is equal to 0.88876) and Loxodonta Africana (the African bush elephant which embedding influence is equal to 0.31937). The two hyrax constitute one of the 14 groups computed by [2], namely the group 5. The African elephant is in the group 6 with other 4 species. Both the hyrax are classified as *least concern* by the IUCN, while the African elephant is considered vulnerable, although its population its increasing.

The rank 2 embedding influence of the species in the Serengeti food web is shown in figure 3.1.

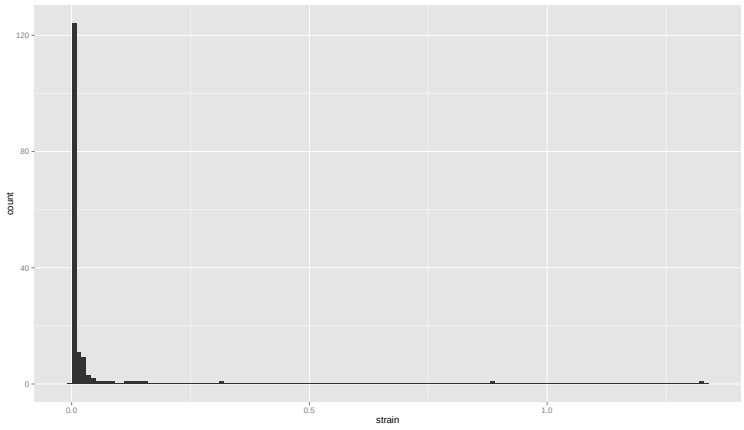

Figure 4: Rank 2 embedding influence for the Serengeti food web.

Choosing a different embedding dimension we found consistency in the stronger influenceing species. For $d = 3$ the 3 stronger influenceing species were the same as for $d = 2$, while for $d = 4$ the 4 stronger influenceing species

were (with decreasing influence) Loxodonta Africana, Procavia Capensis, Giraffa Camelopardalis (the giraffe, also in group 6 as the elephant) and Heterohyrax Brucei.

de Visser et al. The Serengeti food web compiled by [9] is populated by 86 species and 547 predation relationship. The connectance of the graph is equal to 0.07396, more than three time the connectance found in the Baskerville et al. compilation. It is worth to say that the two food web are complementary and the overlap is little. The diameter of the food web both as a directed and as an undirected graph is 4.

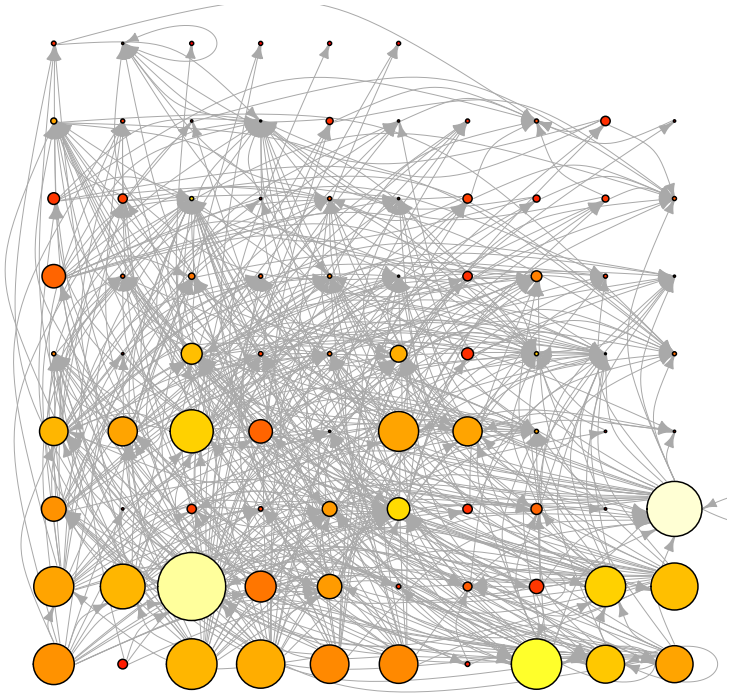

Figure 5: The Serengeti food (*de Visser et al.*). Node sizes show out-degree / in-degree ratio; node colours show node degree (as heat map).

From the sequence of the singular value of the Serengeti adjacency matrix we estimated that the optimal ranking dimension for the food web graph is, again, between $d = 2$ and $d = 4$. In ranking dimension 2, the embedding influence takes values between 0.00009 and 0.33118, with mean equal to

0.05867, median equal to 0.02959 and variance equal to 0.00497. The rank 2 strongest influenceing species is the Heterohyrax Brucei. In the following position we find two nature resource, named by de Visser et al. "Grass and herbs" and "Fruits and Nectar". Again, we find consistent identification of the strongest influenceing species varying the embedding ranking, although it appears to be less strong than in the Baskerville et al. compilation.

The 10 stronger embedding influenceer for the Serengeti (*de Visser et al.*) food web is given in table 3.1.

rank 2	rank 3	rank 4
Heterohyrax Brucei	Resource Grass	Heterohyrax Brucei
Resource Grains	Resource Crops	Dictyoptera
Resource Fruits	Resource Grains	Pelomys Fallax (creek rat)
Thripidae	Resource Fruits	Resource Grains
Micropteropus pusillus	Heterohyrax Brucei	Steatomys pratensis
Diptera	Micropteropus pusillus	Apalis Flavida
Heteroptera	Hodotermes sp.	Resource Fruits
Lycaon pictus	Heteroptera	Resource Crops
Hodotermes sp.	Diptera	Orb-weaver Spiders
Madoqua kirkii	Helogale Parvula	Resource Grass

I removed the part on the Caribbean marine foodweb as it does not add any insight. Ready to ask if reviewers requires more examples.

3.2 Node network indexes and embedding influence

In order to compare the embedding influence of a node with other known measures, we computed for each food web and each of the species present in the food web degree, out-degree, betweenness, closeness centrality, page rank and eigenvector centrality. Figure 6 shows a scatter plot of some of these measure against rank 2 embedding influence for the Serengeti (*Baskerville et al.*) food web.

We fitted these measures with the embedding influence for ranking dimension 2, 3 and 4. Figure 7 show the best linear regression models, for different model dimensions, for rank dimension 2, along with the coefficient of determination R^2, for the Serengeti Baskerville food web. In the appendix similar result are shown for the other food webs.

We didn't found any consistent signal across the various food webs analysed. Hence we argue that, although none of the previously stated measure

There is an everyday longer and longer list for *centralities* measures: ready to add or re-

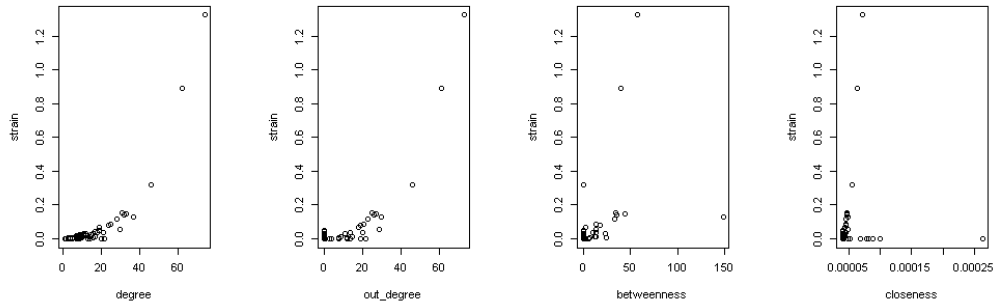

Figure 6: Rank 2 embedding influence against (from the left) degree, out-degree, betweenness and closeness for the Serengeti (*Baskerville et al.*) food web.

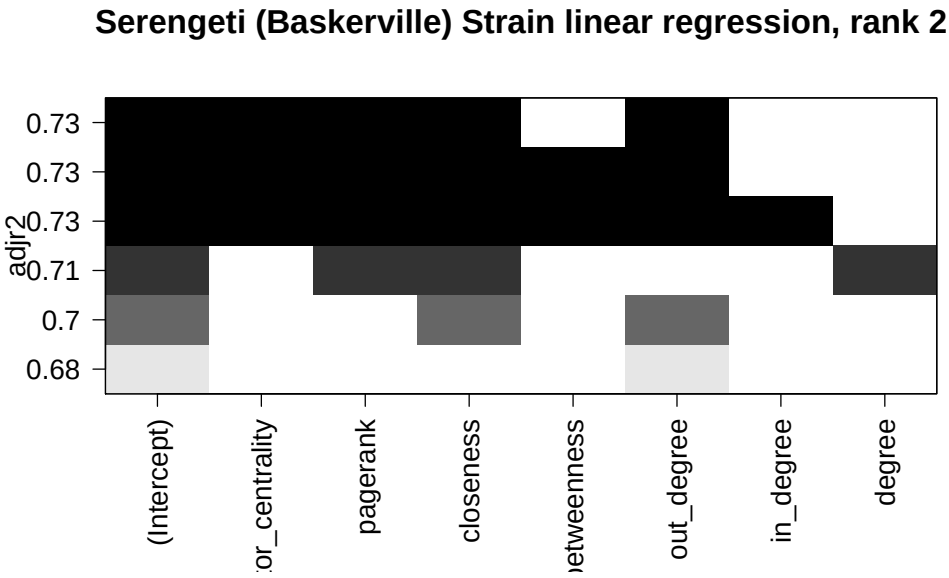

Figure 7: The best linear models for the Serengeti (*Baskerville et al.* food web embedding influence, computed on the proposed measures for rank dimension 2.

is a good substitute, some measures, i.e., the out degree, the degree and the closeness centrality, play a role in defining the embedding influence of a node.

4 Discussion

4.1 Motivations

The motivation to consider a species reduced food web graph G_{-i} and explore the embedding influence sequence $\overline{s(i)}$ arises in at least two real world case.

On the one hand the extinction of the species v_i: in this scenario we are interested in knowing how much the ecosystem has to adapt in order to support the species loss. On the other hand we know that the compilation of food webs is subject to errors and, hence, we can be interested in knowing how much a different picture we would have drawn if one of the species we know is there were not recorded.

4.2 Further analysis

Random generated graph embedding influence Although the degree of a node, and in second order its closeness centrality, appears as the best predictor within the proposed set of measures, the relationship between the more studied measures and the embedding influence remain to be explored and clarified. None of the existing measure appears to be equivalent to the embedding influence.

We already observed, in the empirical case proposed, the behaviour of the embedding influence compared to other network node statistics, often used to estimate the relative importance of a node in the graph.

In order to explore further a possible relationship, we computed the embedding influence sequence on different families of random network models, namely Erdös-Renyi, Watts-Strogatz's Small-World and a random dot product mode with random vectors. We generated random network with a size and connectance comparable to that of the food webs previous introduced.

As for the empirical observed food webs, we fitted the previously stated node relative importance measures to the embedding influence.

References

[1] Stefano Allesina and Mercedes Pascual. Food web models: a plea for groups. *Ecology letters*, 12(7):652–662, 2009.

[2] Edward B Baskerville, Andy P Dobson, Trevor Bedford, Stefano Allesina, T Michael Anderson, and Mercedes Pascual. Spatial guilds in the serengeti food web revealed by a bayesian group model. *PLoS computational biology*, 7(12):e1002321, 2011.

[3] Céline Bellard, Cleo Bertelsmeier, Paul Leadley, Wilfried Thuiller, and Franck Courchamp. Impacts of climate change on the future of biodiversity. *Ecology letters*, 15(4):365–377, 2012.

[4] Phillip Bonacich. Power and centrality: A family of measures. *American journal of sociology*, pages 1170–1182, 1987.

[5] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.

[6] Bradley J Cardinale, J Emmett Duffy, Andrew Gonzalez, David U Hooper, Charles Perrings, Patrick Venail, Anita Narwani, Georgina M Mace, David Tilman, David A Wardle, et al. Biodiversity loss and its impact on humanity. *Nature*, 486(7401):59–67, 2012.

[7] Isabelle M Côté and Emily S Darling. Rethinking ecosystem resilience in the face of climate change. *PLoS biology*, 8(7):e1000438, 2010.

[8] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*, page 1695, 2006.

[9] Sara N de Visser, Bernd P Freymann, and Han Olff. The serengeti food web: empirical quantification and analysis of topological changes under increasing human impact. *Journal of Animal Ecology*, 80(2):484–494, 2011.

[10] Ernesto Estrada. Characterization of topological keystone species: local, global and "meso-scale" centralities in food webs. *Ecological Complexity*, 4(1):48–57, 2007.

[11] Donniell E Fishkind, Daniel L Sussman, Minh Tang, Joshua T Vogelstein, and Carey E Priebe. Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown. *SIAM Journal on Matrix Analysis and Applications*, 34(1):23–39, 2013.

[12] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[13] Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1979.

[14] Frank Harary. Status and contrastatus. *Sociometry*, 22(1):23–43, 1959.

[15] David L. Hawksworth. *Biodiversity: measurement and estimation*, volume 345. Springer, 1995.

[16] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[17] Ana M Martín González, Bo Dalsgaard, and Jens M Olesen. Centrality measures and the importance of generalist species in pollination networks. *Ecological Complexity*, 7(1):36–43, 2010.

[18] Robert M. May. *Stability and Complexity in Model Ecosystems*. Princeton University Press, 1973.

[19] Sonja B. Otto, Björn C. Rall, and Ulrich Brose. Allometric degree distributions facilitate food-web stability. *Nature*, 450(7173):1226–1229, 2007.

[20] Stuart L Pimm. The complexity and stability of ecosystems. *Nature*, 307(5949):321–326, 1984.

[21] Enrico L Rezende, Eva M Albert, Miguel A Fortuna, and Jordi Bascompte. Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecology Letters*, 12(8):779–788, 2009.

[22] Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

[23] Axel G. Rossberg. *Food Webs and Biodiversity*. John Wiley & Sons, 2013.

[24] Torbjörn Säterberg, Stefan Sellman, and Bo Ebenman. High frequency of functional extinctions in ecological networks. *Nature*, 2013.

[25] Ricard V Sole and M^{a} Montoya. Complexity and fragility in ecological networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1480):2039–2045, 2001.

[26] Daniel B. Stouffer, Juan Camacho, Wenxin Jiang, and Luís A. Nunes Amaral. Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences*, 274(1621):1931–1940, 2007.

[27] Ddaniel B. Stouffer, Juan Camacho, R. Guimera, C. A. Ng, and Luís A. Nunes Amaral. Quantitative patterns in the structure of model and empirical food webs. *Ecology*, 86(5):1301–1311, 2005.

[28] Daniel L Sussman, Minh Tang, Donniell E Fishkind, and Carey E Priebe. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, 107(499):1119–1128, 2012.

[29] RDevelopmentCore Team. R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria, 2007. Technical report, ISBN 3-900051-07-0, 2012.

[30] Yuchung J Wang and George Y Wong. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19, 1987.

[31] Bruce A Wilcox. In situ conservation of genetic resources: determinants of minimum area requirements. 1984.

A Stochastic blockmodels graph

We introduced the embedding influence for random dot product food webs, but it's possible to give a definition of it in the more general framework of stochastic blockmodels graphs in a very similar way, assigning to every species in the same block the same latent vector.

The *stochastic blockmodel graph* is defined by the number of blocks K, the *block probability vector* $\rho \in (0,1]^K$ and the *communication probability matrix* $X \in [0,1]^{K\times K}$. The parameter ρ satisfies $\sum_{k=1}^{K} \rho_k = 1$ while X is *identifiable*, that is for each $i, j \leq K$, $i \neq j$, the ith row and the jth row of X are different and the ith column and the jth column of X are different.

Each of the n species is grouped in one of the blocks labelled $1, 2, \ldots, K$. Thus, it's defined a *block membership function* $\beta : V(G) \to \{1, 2, \ldots, K\}$. The probability for a species v_i to be grouped in block k is given by ρ_k. Conditioned on the block membership function β, the probability that there exist an edge (v_i, v_j) is given by $X_{\beta(v_i),\beta(v_j)}$ and is independent from the other couples of species. The model can be generalised to the case where there are S probability community matrices.

Consider the adjacency matrix A_G observed for the foodweb G as a particular outcome of a stochastic blockmodel graph adjacency matrix of which we don't know the parameters nor the block membership function.

The rank d singular value decomposition of the matrix $A_G \in \{0,1\}^{n^2}$ is the given of the six matrices

$$\{U, U_d, S, S_d, V, V_d\}$$

with $U, V \in \mathbb{R}^{n\times d}$, $U_d, V_d \in \mathbb{R}^{n\times(n-d)}$), $S \in \mathbb{R}^{d\times d}$, $S_d \in \mathbb{R}^{(n-d)\times(n-d)}$, such that

$$A_G = [U|U_d]\,(S \oplus S_d)\,[V|V_d]$$

where $[U|U_d]$, which columns are called the left-singular vectors of A_G, and $[V|V_d]$, which columns are called the right-singular vectors of A_G, are real orthogonal matrices, while $\Sigma = S \oplus S_d$ is the diagonal matrix of the singular values of A_G, i.e. a diagonal matrix with descending entries $\Sigma^{1,1} \geq \Sigma^{2,2} \geq \cdots \geq \Sigma^{n,n}$.

Define $\sqrt{S} \in \mathbb{R}^{d\times d}$ as the diagonal matrix with entries are the square roots of the diagonal entries of S. Then we define L and R as $U\sqrt{S}$ and $V\sqrt{S}$ respectively.

We can, now, cluster the rows of L or R or $[L|R]$ into at most d clusters under the minimum least squares criterion (but other algorithm could be used): if the rows of X are pairwise different, compute

$$\mathbb{R}^{n\times d} \ni \hat{X} := \min_{M\in\mathbb{R}^{n\times d}} ||M - L||_F$$

where the matrices M have at most K distinct rows; if the columns of X are pairwise different, compute

$$\mathbb{R}^{n\times d} \ni \hat{X} := \min_{M\in\mathbb{R}^{n\times d}} ||M - R||_F$$

where the matrices M have at most K distinct rows; otherwise compute

$$\mathbb{R}^{n\times 2d} \ni \hat{X} := \min_{M\in\mathbb{R}^{n\times 2d}} ||M - [L|R]\,||_F$$

where the matrices M have at most K distinct rows. The clusters so computed give a consistent estimates for the true graph blocks.

We are now able to define the embedding influence as the distance between $\hat{X}_{G-i}$ and $\hat{X}^{-i}$.

B Food webs network analysis

We computed degree, in-degree, out-degree, betweenness, closeness centrality, page rank and eigenvector centrality using R 3.0.3 [29] and the packages *igraph* [8], *shapes*, *ppls*, *ggplot2*, *car*, *leaps* (**REFS!!!**).

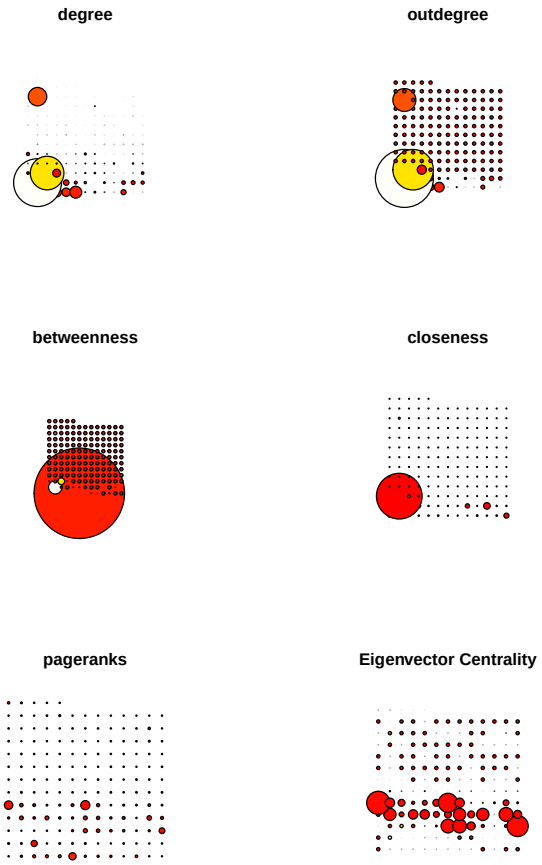

Figure 8: This plots offer a graphic representation of the rank 2 embedding influence of nodes, as a colour in the heat map, compared to other nodes measures, as size, for the Serengeti (Baskerville et al.) food web.