IDENTIFYING COMMON BACKBONES OF INTERACTIONS UNDERLYING FOOD WEBS VIA NON-DETERMINISTIC ALIGNMENTS

Yifan Xu, *Carlos Taveras*, *Lydia Beaudrot*†, *César A. Uribe*

*Rice University, Houston, TX 77005, USA
†Michigan State University, East Lansing, MI 48824, USA

ABSTRACT

Climate change reshapes food webs by altering species distributions and interactions, making it essential to identify structural backbones of interactions that persist across ecosystems. Deterministic alignment methods are computationally slow and restricted to one-to-one correspondences. We introduce a scalable, non-deterministic alignment framework inspired by optimal transport that captures overlapping species roles via many-to-many mappings. Framed via motif-role profiles as a Gromov–Wasserstein transport problem, our method is both efficient and interpretable. We apply the proposed method to a large continental-scale data set of 129 mammal food webs in Sub-Saharan Africa. Pairwise alignments are identified, and we uncover robust backbones with greater connectivity and transitivity than those expected under null models. The proposed approach provides a formal, reproducible tool for forecasting ecosystem reorganization and conservation efforts.

Index Terms— Graph alignment, ecological networks, optimal transport

1. INTRODUCTION

Food webs are directed networks that model predator-prey interactions and map the flow of energy through ecosystems. *Climate change alters these networks* by reshaping species distributions, modifying interaction strengths, and, in some cases, dismantling trophic pathways altogether [1]. Such perturbations hinder predictions of ecosystem responses, as changes in interaction structure can cascade to affect biodiversity and ecosystem services [2]. This makes food web analysis indispensable for understanding the stability, functioning, and resilience of ecological communities under environmental stressors [2–4]. In this context, identifying structural patterns that persist across food webs, referred to as the **backbone of interactions**, becomes crucial [5]. These *backbones* highlight recurrent trophic links that endure despite climate-driven variability, revealing ecological regularities that constrain community assembly. By isolating these persistent features, food web analysis not only deepens our theoretical understanding of ecosystem organization but also provides a framework for forecasting how ecosystems might reorganize under future climate scenarios.

Backbones of interactions are typically identified through comparative analyses of ecological networks across communities, ecosystems, or large datasets. Traditionally, such comparisons have focused on network-level descriptors, e.g., connectance, modularity, and trophic-level distributions, that capture broad structural features of food webs [6–8]. While informative, these descriptors provide only coarse-grained descriptions and often fail to capture species-level similarities or functional roles. To overcome this limitation, recent approaches emphasize direct species alignments based on their structural positions within networks. A pioneering effort by [5] introduced a deterministic network alignment framework that maps species one-to-one across food webs, identifying analogous roles through motif-role similarity [9–13]. Despite its fine-grained perspective, this deterministic strategy has drawbacks that hinder its wide-scale applicability. Specifically, it scales poorly as food web size increases and enforces the restrictive assumption of strict one-to-one role correspondence, potentially missing cases where ecological roles overlap or only partially align [14]. Moreover, the reliance on annealing-based optimization adds technical complexity, making alignments difficult to reproduce and computationally expensive for statistically significant results, limiting applicability to large-scale food webs.

We introduce a **non-deterministic** alignment framework, coupled with an optimization-based method, to identify the backbones of interactions in food webs. Analogous to the relaxation introduced by Kantorovich to the classic Monge's Optimal Transport (OT) problem [15, 16], we relax the deterministic pairings into many-to-many alignments. In summary, our contributions are as follows:

1. We provide a first mathematically rigorous formulation of ecological network alignment based on network motifs.
2. We propose a computationally efficient, scalable, non-deterministic alignment framework and algorithm.
3. We introduce the notion of backbone of interactions for non-deterministic alignments and show they satisfy the criteria provided by [5].

2. COMPUTING NON-DETERMINISTIC MOTIF-BASED NETWORK ALIGNMENTS

Aligning two ecological networks amounts to identifying correspondences between their species based on the ecological roles they play within their respective networks. Formally, our task is to find an *alignment* (defined below) that minimizes a cost functional quantifying the aggregated dissimilarity between species across two food webs.

Definition 2.1. *Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ denote two directed networks where $m = |V_1|$ and $n = |V_2|$. Moreover, let $\mu \in \Delta^{m-1}, \nu \in \Delta^{n-1}$ be two distributions representing species importance. A **non-deterministic alignment** between G_1, G_2 is a matrix $T \in \mathcal{C}_1(\mu) \cap \mathcal{C}_2(\nu)$, where*

$$\mathcal{C}_1(\mu) := \{T \in [0,1]^{m \times n} \mid T\mathbb{1}_n \preceq \mu\}, \quad and$$

$$\mathcal{C}_2(\nu) := \{T \in [0,1]^{m \times n} \mid T^\top \mathbb{1}_m \preceq \nu\}, \quad (1)$$

with T_{ij} representing the alignment between species i and j.

Next, we build the components to quantify the quality of a non-deterministic alignment. Specifically, we will provide a novel formalization of the *motif-based* alignment approach introduced in [5] and will uncover its geometric structure by showing that the task is analogous to a Gromov-Wasserstein [17] transport plan computation.

Network motifs are small subnetworks that enumerate all possible interaction patterns among species. They are widely used to capture *local structural information* in ecological networks, including food webs, mutualistic systems, and multitrophic interaction networks [9–13]. In ecological applications, motif structure has been linked to interpretable functional roles (e.g., basal producers, intermediate consumers, top predators) and to properties such as stability, robustness, and energy flow. Motif-role profiles quantify structural dissimilarity in species' ecological roles across networks by recording how often each species occupies distinct topological positions [14].

Given a species s in a food web G, its *motif-role profile* is a vector $m(s) \in \mathbb{R}^p$ whose i-th entry counts how many times s appears in role i across all occurrences of directed k-node motifs in G. For example, there are 13 non-isomorphic 3-node motifs, which yield $p = 30$ unique roles [18]. Given two species i, j from possibly different webs, their dissimilarity is computed as $d(m(i), m(j))$, for some discrepancy $d : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}$, e.g., $d(x, y) := 1 - \rho(x, y)$ where $\rho(x, y)$ represents the Pearson correlation coefficient between x and y. In the context of network alignment between G_1 and G_2, we compute the dissimilarities between every pair of species from the two networks, yielding the cost matrix $C_{ij} = d(m(i), m(j))$ for the alignment formulations.

Our objective is to find an optimal non-deterministic alignment between graphs G_1 and G_2 by solving the following optimization problem

$$\min_{T \in [0,1]^{m \times n}} \alpha \langle T, A_1(C \odot T)A_2 \rangle + (1-\alpha)\langle C, T \rangle - \epsilon \langle T, \mu\nu^\top \rangle$$

$$\text{s.t. } T \in \mathcal{C}_1(\mu) \cap \mathcal{C}_2(\nu). \quad (2)$$

where $\alpha \in [0, 1]$ is a tradeoff parameter between first-order and second-order alignments, $\epsilon > 0$ self-alignment penalty, and A_1, A_2 are the adjacency matrices of the underlying undirected graphs of G_1, G_2, respectively.

Similar to the Fused Gromov-Wasserstein approach introduced in [19, 20], $\alpha = 0$ emphasizes direct role matching similar to the Wasserstein distance, while $\alpha = 1$ emphasizes neighbor-role consistency similar to the Gromov-Wasserstein approach. Moreover, $\epsilon > 0$ controls when no alignment should be preferred over a weak alignment.

We propose a KL proximal point method with a Dykstra subroutine to enforce constraints and compute approximately optimal non-deterministic alignments as described in the following iterations:

$$\hat{T}^{(k)} = P_{\mathcal{C}_1}\left(T^{(k)} \odot \exp(-\gamma^{-1}Q^{(k)})\right),$$

$$T^{(k+1)} = P_{\mathcal{C}_2}\left(\hat{T}^{(k)} \odot \exp(-\gamma^{-1}Q'^{(k)})\right), \quad (3)$$

$$Q^{(k)} := \alpha A_1(C \odot T^{(k)})A_2 + \frac{1}{2}(1-\alpha)C - \frac{1}{2}\epsilon\mu\nu^\top,$$

$$Q'^{(k)} := \alpha C \odot (A_1\hat{T}^{(k)}A_2) + \frac{1}{2}(1-\alpha)C - \frac{1}{2}\epsilon\mu\nu^\top.$$

Assumption 1. *The accumulative asymmetrical error (AAE) of the iterates generated by Eq. (3) is bounded, i.e.,*

$$\sum_{k=0}^{\infty}\left(D_h(\pi^{(k+1)}, w^{(k)}) - D_h(w^{(k)}, \pi^{(k+1)})\right) < \infty,$$

where $D_h(x, y)$ is the Bregman divergence between x and y associated with a strictly convex function h.

Assumption 1 has been shown to hold for h quadratic, and has been empirically verified for h being the relative entropy [21, Figure 3].

Proposition 1. *Let G_1, G_2 be two graphs with adjacency matrices A_1, A_2, C a cost matrix, $\alpha \in [0, 1]$, and $\epsilon > 0$, and $T^{(0)} = \frac{1}{mn}\mathbb{1}_m\mathbb{1}_n^\top$. Let Assumption 1 hold. Then, the iterates generated by Eq. (3) converge to a stationary point of (2)*

Deterministic alignments arise as a special case with $T_{ij} \in \{0, 1\}$, and μ and ν are uniform. Our next theoretical result formalizes the heuristic developed in [5] for the alignment computation and shows its geometric structure as an optimization problem with a Gromov-Wasserstein-like cost functional.

Proposition 2. *Let G_1, G_2 be two graphs with adjacency matrices A_1, A_2 respectively, C a cost matrix, then an alignment*

$$\hat{T}^* = \arg\min_{T \in \{0,1\}^{m \times n}} \langle T, A_1(C \odot T)A_2 \rangle + \langle T, \Xi(T) \rangle,$$

$$\text{s.t. } T \in \mathcal{C}_1(\mathbb{1}_m) \cap \mathcal{C}_2(\mathbb{1}_n), \quad where \quad (4)$$

$$\Xi(T) = (1-\epsilon)\left(\max\left(A_1\mathbb{1}_m\mathbb{1}_n^\top, \mathbb{1}_m\mathbb{1}_n^\top A_2\right) - A_1TA_2\right).$$

is optimal if and only if it minimizes [5, Eq. (3)].

We examine the various properties of deterministic and non-deterministic alignments in [22, Appendix A.2]. Moreover, algorithmic details and proofs for Proposition 1 and 2 can be found in [22, Appendices A.1, B.1., B.2.].

3. NON-DETERMINISTIC BACKBONES

Consider a set of networks $\mathcal{D} = \{G_1, G_2, \cdots, G_N\}$, the pairwise non-deterministic alignments between them $\{T^{ij}\}_{i,j=1}^N$ and the pairwise dissimilarity matrices between them $\{C^{ij}\}_{i,j=1}^N$. Moreover, let $[N]$ denote the set $\{1, \cdots, N\}$ for all positive integers up to N.

Definition 3.1 (Non-deterministic role similarity score). *Given a graph G_i for some $i \in [N]$ and an ordering of its species, the role similarity score of species k in V_i is defined as*

$$\sum_{p=1}^N \sum_{q=1}^{|V_p|} (1 - C^{ip}_{jq}) T^{ip}_{jq}.$$

A species' role similarity score is high when: 1) a small amount of its mass is self-aligned, and 2) the alignment incident to this species generates low direct-matching cost. Under this definition, a species with a high role similarity score is said to be **well-aligned** with respect to the dataset $\mathcal{D}$.

Definition 3.2 (Top-k backbone of interactions). *Given a graph G_i and the role similarity scores of its species, for an integer $k \geq 2$, we define the top-k backbone of G_i, denoted as $B_i = (V_{B_i}, E_{B_i})$, to be the subgraph induced by the k species with the highest role similarity scores.*

Definition 3.3 (Non-deterministic transitivity score). *For each species in V_{B_i}, which is the j-th species in V_i, its non-deterministic transitivity score is:*

$$Transitivity(j) = \frac{\displaystyle\sum_{\substack{p,q\in[N] \\ p<q;p,q\neq i}} \sum_{j_p\in V_p} \sum_{j_q\in V_q} T^{ip}_{jj_p} \cdot T^{iq}_{jj_q} \cdot T^{pq}_{j_p j_q}}{\displaystyle\sum_{\substack{p,q\in[N] \\ p<q;p,q\neq i}} \sum_{j_p\in V_P} \sum_{j_q\in V_q} T^{ip}_{jj_p} \cdot T^{iq}_{jj_q}}.$$

The transitivity score of a food web is the average among its backbone species.

Intuitively, for each species j in the backbone B_i, we examine its alignment to species j_p in G_p and j_q in G_q, across all pairs of networks G_p, G_q not equal to G_i. The numerator accumulates these alignment strengths weighted by how strongly j_p and j_q are themselves aligned (as measured by $T^{pq}_{j_p j_q}$), representing how well the species alignments form a transitive triangle. The denominator serves as a normalization factor, ensuring the score lies between 0 and 1, by aggregating all alignment-strength products $T^{ip}_{jj_p} \cdot T^{iq}_{jj_q}$ regardless of whether j_p and j_q align.
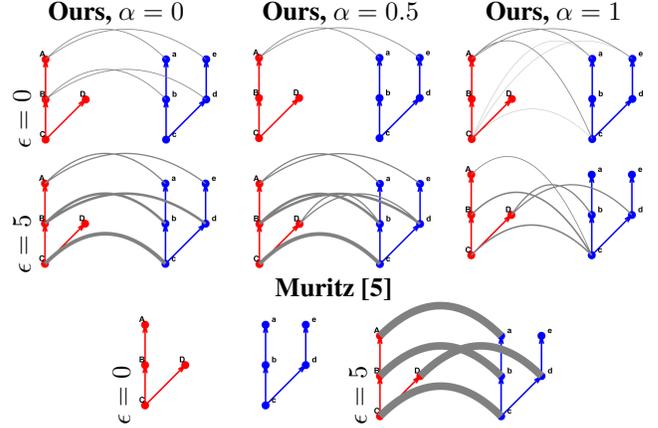

Fig. 1: Optimal alignment between two toy networks with various alignment methods. Thicker lines indicate stronger alignment. Multiple deterministic alignments exist, so an arbitrary one is picked for both values of ϵ.

4. NUMERICAL ANALYSIS ON SUB-SAHARAN AFRICA MAMMAL FOOD WEBS

We first evaluate our method on a pair of small synthetic networks, setting the species significance distributions μ and ν to be uniform. Figure 1 compares deterministic and non-deterministic alignments under varying values of the tradeoff parameter α and the self-alignment penalty ϵ. Species **A** in the red network shares an identical role profile to both species **a** and **e** in the blue network; likewise, Species **B** matches both species **b** and **d**. Deterministic alignment captures only a single pairing per species, while non-deterministic alignment identifies *all* valid pairings and assigns each an alignment strength proportional to role similarity. This highlights the ability of non-deterministic alignment to detect structural redundancies.

We next applied our method to a real-world ecological dataset, performing pairwise alignments between 129 mammal food webs across Sub-Saharan Africa, using the dataset compiled by [20], see [22, Appendix A.3] for additional details. Non-deterministic alignments completed in 1.00 ± 0.20 seconds on average, substantially faster than deterministic alignments, which required 7.92 ± 2.31 seconds. Figure 2 illustrates the alignment between the Cliff of Bandiagara and Namaqua networks. Both deterministic and non-deterministic approaches recover expected matches, such as *Caracal caracal* in both networks and *Leptailurus serval* (Bandiagara) with *Felis silvestris* (Namaqua). However, the non-deterministic framework reveals additional ecological role redundancies by relaxing the one-to-one constraint. For instance, while the deterministic alignment maps *Genetta genetta* (Namaqua) solely to *Papio anubis* (Bandiagara), the non-deterministic alignment uncovers that *Herpestes ichneumon* (Bandiagara) also occupies a similar ecological role. This example illustrates a general limitation of deterministic methods: by enforcing strict one-to-one pairings, they obscure many ecologically valid correspon-
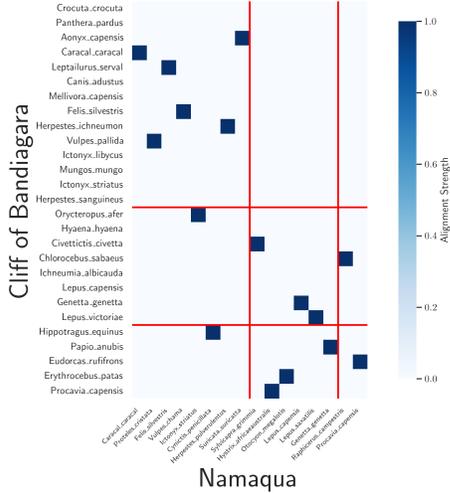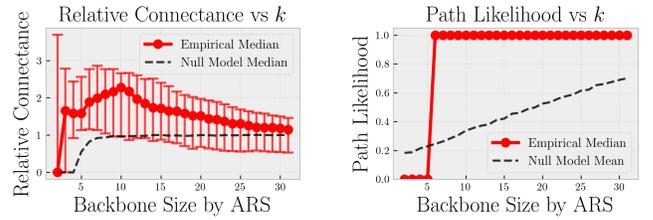
(a) Deterministic alignment (Muritz [5])
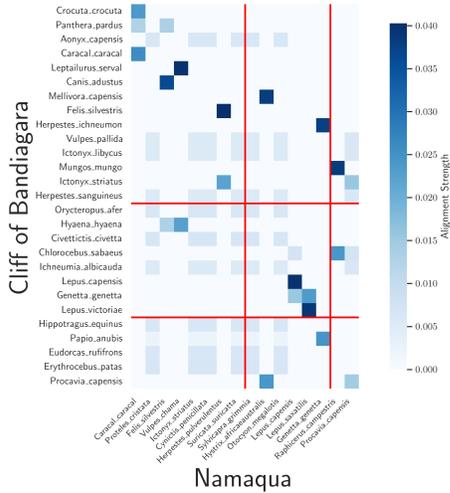
(b) Non-deterministic alignment ($\alpha = 0.5$)

Fig. 2: Alignment heatmaps between the species in Cliff of Bandiagara and the species in Namaqua. Values in the heatmap represent the strength of alignment between species.

dences and arbitrarily discard alternatives when species roles partially overlap. Because backbones are derived from these alignments, such restrictions propagate to backbone identification and can yield incomplete or even misleading backbones. In contrast, our non-deterministic framework systematically captures all valid correspondences within a single reproducible solution, thereby producing backbones that more faithfully represent the redundancies and structural regularities present in ecological networks.

We extract the backbones of interactions, sets of interactions that persist across ecological networks, using our pairwise alignments. We followed the backbone definition in [5] when applicable, and made extensions to non-deterministic alignments when necessary; see [22, Appendix A.4] for details.

Figure 3 summarizes three key properties of our top-k back-

(a) Relative connectance, the ratio of backbone connectance to that of the full food web.

(b) Path likelihood, the fraction of backbones that induce a connected subnetwork.
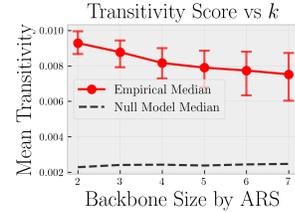
(c) Transitivity, the non-deterministic alignment transitivity score.

Fig. 3: Structural properties of our top-k backbones as a function of k.

bones: relative connectance, path likelihood, and transitivity. Each plot compares these properties against the corresponding averages (or medians) under a null model, in which k species are randomly selected 50 times (for relative connectance and path likelihood) and 25 times (for transitivity score), without replacement, for each food web. Across all measures, the backbones revealed by our non-deterministic alignments consistently exhibited both higher connectance and transitivity than the null model, suggesting that their connectivity and transitivity are statistically significant. These findings show that our method yields backbones that consistently meet the backbone criteria proposed in [5], while offering scalability and reproducibility.

5. CONCLUSIONS AND FUTURE WORK

We introduced a non-deterministic alignment framework for identifying the backbone of interactions in ecological networks. Our method relaxes one-to-one role correspondences, allowing the capture of overlapping and redundant ecological roles. Our experiments primarily considered uniform distributions of the species importance parameters, μ and ν. Future work should incorporate ecologically meaningful weights such as degree, biomass, or abundance. Parameter-free backbone constructions can be developed, such as forward backbones. Moreover, other criteria for species should be explored, such as the average entropy of their alignment distributions, which would highlight species with consistently unique roles across networks.

7. REFERENCES

[1] M. Lurgi, B. C. Lopez, and J. M. Montoya, "Climate change impacts on body size and food web structure on mountain ecosystems," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1605, pp. 3050–3057, 2012.

[2] T. H. Oliver, M. S. Heard, N. J. Isaac, D. B. Roy, D. Procter, F. Eigenbrod, R. Freckleton, A. Hector, C. D. L. Orme, O. L. Petchey, *et al.*, "Biodiversity and resilience of ecosystem functions," *Trends in ecology & evolution*, vol. 30, no. 11, pp. 673–684, 2015.

[3] E. Thébault and C. Fontaine, "Stability of ecological communities and the architecture of mutualistic and trophic networks," *Science*, vol. 329, no. 5993, pp. 853–856, 2010.

[4] A. D. Barnes, M. Jochum, J. S. Lefcheck, N. Eisenhauer, C. Scherber, M. I. O'Connor, P. de Ruiter, and U. Brose, "Energy flux: the link between multitrophic biodiversity and ecosystem functioning," *Trends in ecology & evolution*, vol. 33, no. 3, pp. 186–197, 2018.

[5] B. Bramon Mora, D. Gravel, L. J. Gilarranz, T. Poisot, and D. B. Stouffer, "Identifying a common backbone of interactions underlying food webs from different ecosystems," *Nature Communications*, vol. 9, no. 1, p. 2603, 2018.

[6] J. A. Dunne, R. J. Williams, and N. D. Martinez, "Network structure and biodiversity loss in food webs: robustness increases with connectance," *Ecology letters*, vol. 5, no. 4, pp. 558–567, 2002.

[7] R. J. Williams and N. D. Martinez, "Simple rules yield complex food webs," *Nature*, vol. 404, no. 6774, pp. 180–183, 2000.

[8] D. B. Stouffer, E. L. Rezende, and L. A. N. Amaral, "The role of body mass in diet contiguity and food-web structure," *Journal of Animal Ecology*, vol. 80, no. 3, pp. 632–639, 2011.

[9] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.

[10] B. Baiser, R. Elhesha, and T. Kahveci, "Motifs in the assembly of food web networks," *Oikos*, vol. 125, no. 4, pp. 480–491, 2016.

[11] S. Cenci, C. Song, and S. Saavedra, "Rethinking the importance of the structure of ecological networks under an environment-dependent framework," *Ecology and evolution*, vol. 8, no. 14, pp. 6852–6859, 2018.

[12] P. V. Paulau, C. Feenders, and B. Blasius, "Motif analysis in directed ordered networks and applications to food webs," *Scientific reports*, vol. 5, no. 1, p. 11926, 2015.

[13] J. Tavella, F. M. Windsor, D. C. Rother, D. M. Evans, P. R. Guimaraes Jr, T. P. Palacios, M. Lois, and M. Devoto, "Using motifs in ecological networks to identify the role of plants in crop margins for multiple agriculture functions," *Agriculture, Ecosystems & Environment*, vol. 331, p. 107912, 2022.

[14] D. B. Stouffer, M. Sales-Pardo, M. I. Sirer, and J. Bascompte, "Evolutionary conservation of species' roles in food webs," *Science*, vol. 335, no. 6075, pp. 1489–1492, 2012.

[15] C. Villani *et al.*, *Optimal transport: old and new*, vol. 338. Springer, 2008.

[16] L. V. Kantorovich, "On the translocation of masses," in *Dokl. Akad. Nauk. USSR (NS)*, vol. 37, pp. 199–201, 1942.

[17] F. Mémoli, "Gromov-wasserstein distances and the metric approach to object matching," *Foundations of Computational Mathematics*, vol. 11, pp. 417–487, Aug. 2011.

[18] B. B. Mora, A. R. Cirtwill, and D. B. Stouffer, "pymfinder: a tool for the motif analysis of binary and quantitative complex networks," *BioRxiv*, p. 364703, 2018.

[19] T. Vayer, L. Chapel, R. Flamary, R. Tavenard, and N. Courty, "Fused gromov-wasserstein distance for structured objects," *Algorithms*, vol. 13, no. 9, 2020.

[20] K. M. Hung, L. Beaudrot, A. E. Finneran, A. G. Zalles, and C. A. Uribe, "Quantifying functionally equivalent species and ecological network dissimilarity with optimal transport distances," *Methods in Ecology and Evolution*, vol. n/a, no. n/a, 2025.

[21] J. Li, J. Tang, L. Kong, H. Liu, J. Li, A. M.-C. So, and J. Blanchet, "A convergent single-loop algorithm for relaxation of gromov-wasserstein in graph data," *arXiv preprint arXiv:2303.06595*, 2023.

[22] Y. Xu, C. Taveras, L. Beaudrot, and C. A. Uribe, "Identifying common backbones of interactions underlying food webs via non-deterministic alignments." Manuscript, 2026. Available at: `https://cauribe.rice.edu/xu_et_al_icassp2026_appendix/`.

A. SUPPLEMENTARY MATERIALS

A.1. The optimization algorithm for non-deterministic alignment

Problem setup and splitting: Here, we use the split variables π and w to derive Algorithm 1; in the main text, both are represented by the single alignment variable T.
Recall the non-deterministic alignment problem (Problem 2) with inequality marginals $\mathcal{C}_1 := \{ \pi \in [0,1]^{m \times n} : \pi \mathbb{1}_n \preceq \mu \}$ and $\mathcal{C}_2 := \{ w \in [0,1]^{m \times n} : w^\top \mathbb{1}_m \preceq \nu \}$. Similar to the splitting idea presented in [21], we rewrite cost function as follows:

$$g(\pi, w) := \alpha \langle \pi, A_1(C \odot w)A_2 \rangle + (1-\alpha)\langle C, \frac{1}{2}(\pi + w) \rangle - \epsilon \langle \frac{1}{2}(\pi + w), \mu\nu^\top \rangle$$

so Problem 2 can be rewritten as:

$$\min_{\pi = w} g(\pi, w) + \mathbb{I}_{\mathcal{C}_1}(\pi) + \mathbb{I}_{\mathcal{C}_2}(w) \quad (5)$$

where $\mathbb{I}_{\mathcal{C}}$ denotes the indicator function for the constraint set $\mathcal{C}$. Then, we can penalize the equality constraint in the cost function and construct the penalized cost function:

$$\min_{\pi = w} g(\pi, w) + \mathbb{I}_{\mathcal{C}_1}(\pi) + \mathbb{I}_{\mathcal{C}_2}(w) + \gamma D_h(\pi, w)$$
$$D_h(X, Y) := \sum_{ij} \left[x_{ij} \log \frac{x_{ij}}{y_{ij}} - x_{ij} + y_{ij} \right] \quad (6)$$

where D_h is the Bregman divergence induced by the negative entropy $h(X) = \sum_{ij} x_{ij} \log x_{ij}$, and $\gamma > 0$.
KL-Bregman Alternating Projected Gradient [21] : At iteration k, we alternate between linearizing g in each block and solve a KL–proximal step subproblem:

$$\pi^{k+1} = \arg\min_{\pi \in \mathcal{C}_1} \left\langle Q^{(k)}, \pi \right\rangle + \gamma D_h(\pi, w^k)$$
$$w^{k+1} = \arg\min_{w \in \mathcal{C}_2} \left\langle Q^{'(k)}, w \right\rangle + \gamma D_h(w, \pi^{k+1}), \quad (7)$$

where the cost surrogates $Q^{(k)}$ and $Q^{'(k)}$ are defined as:

$$Q^{(k)} = \alpha A_1(C \odot w^k)A_2 + \frac{1}{2}(1-\alpha)C - \frac{1}{2}\epsilon\mu\nu^\top$$
$$Q^{'(k)} = \alpha\, C \odot (A_1\pi^{k+1}A_2) + \frac{1}{2}(1-\alpha)C - \frac{1}{2}\epsilon\,\mu\nu^\top.$$

Each subproblem has the same form: minimize a linear term plus a Bregman divergence to the previous iteration, under nonnegativity and a **single** set of marginal constraints.
Dual of the KL–prox subproblem and optimizer: We write the generic subproblem as

$$\min_{X \in \mathcal{C}'} \langle Q, X \rangle + \gamma D_h(X, Y) \quad (8)$$

where $\mathcal{C}'$ is either $\mathcal{C}_1$ or $\mathcal{C}_2$. First, consider (8) with row constraints $\mathcal{C}_1 = \{X \in [0,1]^{m \times n} : X\mathbb{1} \preceq \mu\}$. Introducing dual variables $\lambda \in \mathbb{R}^m_{\geq 0}$, the Lagrangian is

$$\mathcal{L}(X, \lambda) = \langle Q, X \rangle + \gamma D_h(X, Y) + \langle \lambda, X\mathbb{1} - \mu \rangle \quad (9)$$

Minimizing w.r.t. X entrywise yields, for any $x_{ij} > 0$:

$$0 = \frac{\partial \mathcal{L}}{\partial x_{ij}} = Q_{ij} + \gamma \log \frac{x_{ij}}{y_{ij}} + \lambda_i$$
$$\implies x_{ij} = y_{ij} \exp\left(-\frac{Q_{ij} + \lambda_i}{\gamma} \right)$$

If we denote $\hat{X}_{ij} = y_{ij} \exp\left(-\frac{Q_{ij}}{\gamma} \right)$ as the unconstrained optimizer, then the primal optimizer can be written as:

$$X^\star_{ij} = y_{ij} \exp\left(-\frac{Q_{ij}}{\gamma} \right) \exp\left(-\frac{\lambda_i^\star}{\gamma} \right) = \hat{X}_{ij} \exp\left(-\frac{\lambda_i^\star}{\gamma} \right), \quad (10)$$

with complementary slackness $\lambda_i^\star((X^\star\mathbb{1})_i - \mu_i) = 0$. An analogous derivation with column constraints $\mathcal{C}_2 = \{X \geq 0, X^\top\mathbb{1} \preceq \nu\}$ introduces $\eta \in \mathbb{R}^n_{\geq 0}$ and yields

$$X^\star_{ij} = y_{ij} \exp\left(-\frac{Q'_{ij}}{\gamma} \right) \exp\left(-\frac{\eta_j^\star}{\gamma} \right), \quad \eta_j^\star\left((X^{\star\top}\mathbb{1})_j - \nu_j\right) = 0. \quad (11)$$

Enforcing constraints with Dykstra-style projections: The structure of the Lagrangian optimizer shows that enforcing $\pi\mathbb{1} \preceq \mu$ (or $w^\top\mathbb{1} \preceq \nu$) is equivalent to scaling each row i (or column j) by a factor $\exp\left(-\frac{\lambda_i}{\gamma} \right)$ (or $\exp\left(-\frac{\eta_i}{\gamma} \right)$). Complementary slackness implies that:

- if the current row sum $(\hat{X}\mathbb{1})_i \leq \mu_i$ (or column sum $(\hat{X}^\top\mathbb{1})_j \leq \nu_j$), then $\lambda_i^* = 0$ (or $\eta_i^* = 0$) and we leave the row unchanged,

- if $(\hat{X}\mathbb{1})_i \geq \mu_i$ (or $(\hat{X}^\top\mathbb{1})_j \geq \nu_j$), then $\lambda_i^* > 0$ (or $\eta_i^* > 0$) and the row is uniformly scaled until it satisfies the constraint.

Therefore, the Dykstra projection steps onto the constraint sets are simply

$$P_{\mathcal{C}_1} = \text{diag}\left(\min\left(\frac{\mu}{\hat{X}\mathbb{1}}, \mathbb{1} \right) \right) \hat{X}$$
$$P_{\mathcal{C}_2} = \hat{X}\text{diag}\left(\min\left(\frac{\nu}{\hat{X}^\top\mathbb{1}}, \mathbb{1} \right) \right)$$

Summary: Now, we are ready to present Algorithm 1. In short, we

1. linearize cost surrogate around one operator: $Q^{(k)} \leftarrow \alpha A_1(C \odot T^{(k-1)})A_2 + \frac{1}{2}(1-\alpha)C - \frac{1}{2}\epsilon\nu\mu^\top$

2. do a multiplicative update with step size γ: $\hat{T} \leftarrow T^{(k-1)} \odot \exp\left(-\frac{Q^{(k)}}{\gamma} \right)$

3. project iterate back into $\mathcal{C}_1$: row_factor $\leftarrow \min(\frac{\mu}{\hat{T}\mathbb{1}_n}, \mathbb{1}_m)$ and $\hat{T} \leftarrow \mathrm{diag}(\text{row\_factor})\hat{T}$

4. repeat for the other operator and the column constraints, until a stopping criteria is met.

Algorithm 1: The KL Proximal Point Algorithm with Dykstra Subroutine

---

Input : - the adjacency matrices for the underlying undirected graphs A_1, A_2
- the dissimilarity matrix C
- the tradeoff parameter $\alpha \in [0, 1]$
- the self-alignment penalty $\epsilon > 0$
- the step size $\gamma > 0$
- the normalized species significance distributions μ, ν
- some additional stopping criteria for main loop (step tolerance, etc.)

Output : T, the non-deterministic alignment between G_1 and G_2

$T^{(0)} \leftarrow \frac{1}{mn}\mathbb{1}_m\mathbb{1}_n^\top$;
foreach $k = 1, 2, \cdots, max\_iter$ **do**
    $Q^{(k)} \leftarrow$
    $\alpha A_1(C \odot T^{(k-1)})A_2 + \frac{1}{2}(1-\alpha)C - \frac{1}{2}\epsilon\nu\mu^\top$;
    $\hat{T} \leftarrow T^{(k-1)} \odot \exp\left(-\frac{Q^{(k)}}{\gamma}\right)$;
    row_factor $\leftarrow \min(\frac{\mu}{\hat{T}\mathbb{1}_n}, \mathbb{1}_m)$;
    $\hat{T} \leftarrow \mathrm{diag}(\text{row\_factor})\hat{T}$;
    $Q^{(k)\prime} \leftarrow \alpha C \odot (A_1\hat{T}A_2) + \frac{1}{2}(1-\alpha)C - \frac{1}{2}\epsilon\nu\mu^\top$;
    $\hat{T} \leftarrow \hat{T} \odot \exp(-\frac{Q^{(k)\prime}}{\gamma})$;
    col_factor $\leftarrow \min(\frac{\nu}{\hat{T}^\top\mathbb{1}_m}, \mathbb{1}_n)$;
    $T^{(k)} \leftarrow \hat{T}\mathrm{diag}(\text{col\_factor})$;
    Check additional stopping criteria for the main loop.;
return $T^{(k)}$;

A.2. Proof of Proposition 1

We first write the split form of 2 as seen in Appendix A.1:

$$\min_{\pi=w} g(\pi, w) + \mathbb{I}_{\mathcal{C}_1}(\pi) + \mathbb{I}_{\mathcal{C}_2}(w), \quad (12)$$

$$\text{where } \mathcal{C}_1 = \{X \in [0, 1]^{m \times n} : X\mathbb{1} \preceq \mu\},$$
$$\mathcal{C}_2 = \{X \in [0, 1]^{m \times n} : X^\top\mathbb{1} \preceq \nu\}.$$

Let $h(X) = \sum_{ij} x_{ij} \log x_{ij}$ be the negative entropy function and D_h be its Bregman divergence. Let $\{(\pi^{(k)}, w^{(k)})\}_{k \geq 0}$ be the sequence generated by alternatively solving the subproblems in 7. We define the potential function:

$$F_\gamma(\pi, w) = g(\pi, w) + \mathbb{I}_{\mathcal{C}_1}(\pi) + \mathbb{I}_{\mathcal{C}_2}(w) + \gamma D_h(\pi, w). \quad (13)$$

Lemma 1. *The following properties hold:*

1. *$f(\pi, w) = \alpha\langle\pi, A_1(C \odot w)A_2\rangle$ is bilinear;*

2. *$\mathcal{C}_1$ and $\mathcal{C}_2$ are closed, convex polytopes;*

3. *the potential function F_γ is coercive.*

Proof of Lemma 1. The three statements are proved in parallel:

1. This follows immediately due to the linearity of inner products, matrix multiplication, and the Hadamard product.

2. **Closedness:** Both inequality constraints $X\mathbb{1} \preceq \mu$ (resp. $X^\top\mathbb{1} \preceq \nu$) can be represented as an intersection of closed half-spaces. Since $[0, 1]^{m \times n}$ is a product of closed intervals which is also closed, we have that the intersection of closed sets is closed; hence, $\mathcal{C}_1$ and $\mathcal{C}_2$ are closed.
   Convexity: Both the unit box $[0, 1]^{m \times n}$ and the linear inequalities are convex, and hence their intersection is convex.
   Polytope: Both C_1 and C_2 are bounded by the unit box $[0, 1]^{m \times n}$. In addition, they are both given by finitely many linear inequalities, so they are polyhedra. Since they are both bounded polyhedra, by definition, they are polytopes.

3. Since both C_1 and C_2 are bounded, $\mathcal{C}_1 \times \mathcal{C}_2$ is bounded. Hence, whenever $\|(\pi, w)\| \to \infty$, we must have that $(\pi, w) \notin \mathcal{C}_1 \times \mathcal{C}_2$, which implies either $\pi \notin \mathcal{C}_1$ or $w \notin \mathcal{C}_2$. In the former case, we have that $\mathbb{I}_{\mathcal{C}_1}(\pi) \to \infty$, and in the latter case we have that $\mathbb{I}_{\mathcal{C}_2}(w) \to \infty$, both of which implies $F_\gamma(\pi, w) \to \infty$, as desired.

$\square$

Proof of Proposition 1. Since the accumulative asymmetrical error is bounded and the statements in Lemma 1 hold, then by [21, Theorem 3.6], every limit point of the sequence $\{(\pi^{(k)}, w^{(k)})\}_{k \geq 0}$ generated by 7 belongs to the fixed point set of the Bregman Alternating Projected Gradient. Hence, Algorithm 1 generates a sequence $\{T^{(k)}\}_{\geq 0}$ that converges to a stationary point of problem 12, and therefore a local minimizer of 2. $\square$

Notes on the accumulative asymmetrical error (AAE): The AAE is defined as:

$$\sum_{k=0}^{\infty} \left(D_h(\pi^{(k+1)}, w^{(k)}) - D_h(w^{(k)}, \pi^{(k+1)})\right)$$

Although it is difficult to theoretically show that the AAE is bounded, previous work has shown that it is often the case in practice [21, Figure 3]. We observe that the AAE is also often bounded in practice.

A.3. Proof of Proposition 2

Before we commence with the proof, we provide definitions for deterministic alignments and the corresponding optimal deterministic alignment problem.

Deterministic alignment: Deterministic alignment (as defined in [5]) can be formulated as a stricter special case to our alignment model, where $T \in \{0,1\}^{m \times n}$, and $\mu = \mathbb{1}_m$ and $\nu = \mathbb{1}_n$:

Definition A.1 (Feasible Deterministic Alignment). *Given two networks G_1, G_2, a **deterministic alignment** is a binary matrix $T \in \{0,1\}^{m \times n} \cap C_1(\mathbb{1}_m) \cap C_2(\mathbb{1}_n)$. $T_{ij} = 1$ indicates that the alignment identifies the ecological role of species i and j to be similar, and $T_{ij} = 0$ otherwise.*

Definition A.2 (Optimal Deterministic Alignment). *Under the same assumptions in Def. 2.1 and with a choice of the self-penalty matrix $\Xi(T)$, an **optimal deterministic alignment** is a solution of*

$$\min_{T \in \{0,1\}^{m \times n}} \alpha \langle T, A_1(C \odot T)A_2 \rangle + (1-\alpha)\langle C, T \rangle + \langle T, \Xi(T) \rangle$$
(14)

$$s.t. \ \ T \in C_1(\mathbb{1}_m) \cap C_2(\mathbb{1}_n)$$

Proof of Proposition 2. Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ have underlying undirected adjacency matrices A_1, A_2, and let $C \in \mathbb{R}^{m \times n}$ be the cost matrix (with discrepancy $1 - \rho(i,j)$). In our deterministic model (Def. A.1, A.2), a feasible $T \in \{0,1\}^{m \times n} \cap C_1(\mathbb{1}_m) \cap C_2(\mathbb{1}_n)$ encodes a deterministic matching between V_1 and V_2. Define the alignment set

$$\lambda(T) = \{(i,j) \in V_1 \times V_2 : T_{ij} = 1\} \cup$$
$$\{(i, \emptyset) \in V_1 \times \emptyset : \sum_j T_{ij} = 0\} \cup$$
$$\{(\emptyset, j) \in \emptyset \times V_2 : \sum_i T_{ij} = 0\}$$

Conversely, any alignment λ as defined in [5] induces such a feasible T.

We prove that when $\alpha = 1$ and

$$\Xi(T) = (1-\epsilon)\left(\max\left(A_1 \mathbb{1}_m \mathbb{1}_n^\top, \mathbb{1}_m \mathbb{1}_n^\top A_2\right) - A_1 T A_2\right),$$
(15)

the cost function in 4 is equal to [5, Eq. (3)], so that their minimizers coincide. First, with $\alpha = 1$, the cross-alignment terms of 4 collapses to

$$\langle T, A_1(C \odot T)A_2 \rangle = \sum_{i,j} T_{ij} \sum_{\hat{i} \in \mathcal{N}_{G_1}(i)} \sum_{\hat{j} \in \mathcal{N}_{G_2}(j)} C_{\hat{i}\hat{j}} T_{\hat{i}\hat{j}},$$

where $\mathcal{N}_G(i)$ denote the set of 1-hop neighbors of i in G. For each aligned pair of species $x = (i,j) \in \lambda(T)$, the inner double sum ranges over the set of neighbor pairings $\lambda_x := \{(\alpha, \beta) \in \lambda(T) : \alpha \in \mathcal{N}_{G_1}(i), \beta \in \mathcal{N}_{G_2}(j)\}$. Thus

$$\langle T, A_1(C \odot T)A_2 \rangle = \sum_{x \in \lambda(T)} \sum_{(\alpha,\beta) \in \lambda_x} C_{\alpha\beta}$$
$$= \sum_{x \in \lambda(T)} \sum_{(\alpha,\beta) \in \lambda_x} 1 - \rho(\alpha, \beta) \quad (16)$$

which is exactly the cross-alignment term in [5, Eq. (3)].

Next, for some alignment $x = (i,j) \in \lambda(T)$, write the "matched neighbor count" as

$$(A_1 T A_2)_{ij} = |\{(\alpha, \beta) \in \lambda_x\}|.$$

Then, the penalty matrix in 15 gives, entrywise,

$$\Xi_{ij}(T) = (1-\epsilon)\left(\max\left((A_1 \mathbb{1}_m)_i, (A_2 \mathbb{1}_n)_j\right) - (A_1 T A_2)_{ij}\right).$$

If $T_{ij} = 1$, then the term $\max\left((A_1 \mathbb{1}_m)_i, (A_2 \mathbb{1}_n)_j\right) - (A_1 T A_2)_{ij}$ represents "the number of unmatched neighbors on the higher-degree side of the alignment," i.e., $\max(k_x^\alpha, k_x^\beta)$ in [5, Alignment algorithm section of Supplementary Methods]. Therefore,

$$\langle T, \Xi(T) \rangle = \sum_{x \in \lambda(T)} (1-\epsilon)\max(k_x^\alpha, k_x^\beta) = \sum_{x \in \lambda(T)} \xi_x. \ (17)$$

Combining Equations 16 and 17, we have

$$\langle T, A_1(C \odot T)A_2 \rangle + \langle T, \Xi(T) \rangle$$
$$= \sum_{x \in \lambda(T)} \sum_{(\alpha,\beta) \in \lambda_x} 1 - \rho(\alpha, \beta) + \sum_{x \in \lambda(T)} \xi_x$$
$$= \sum_{x \in \lambda(T)} \left(\sum_{(\alpha,\beta) \in \lambda_x} (1 - \rho(\alpha, \beta)) + \xi_x\right),$$

as desired. □