Towards Ecological Network Analysis with Optimal Transport

Kai M. Hung[1], Alex G. Zalles[2], Ann E. Finneran[3], Lydia Beaudrot[3], César A. Uribe[4]

[1]Department of Computer Science, [2]Department of Computational Applied Mathematics and Operations Research, [3]Department of Biosciences, [4]Department of Electrical and Computer Engineering

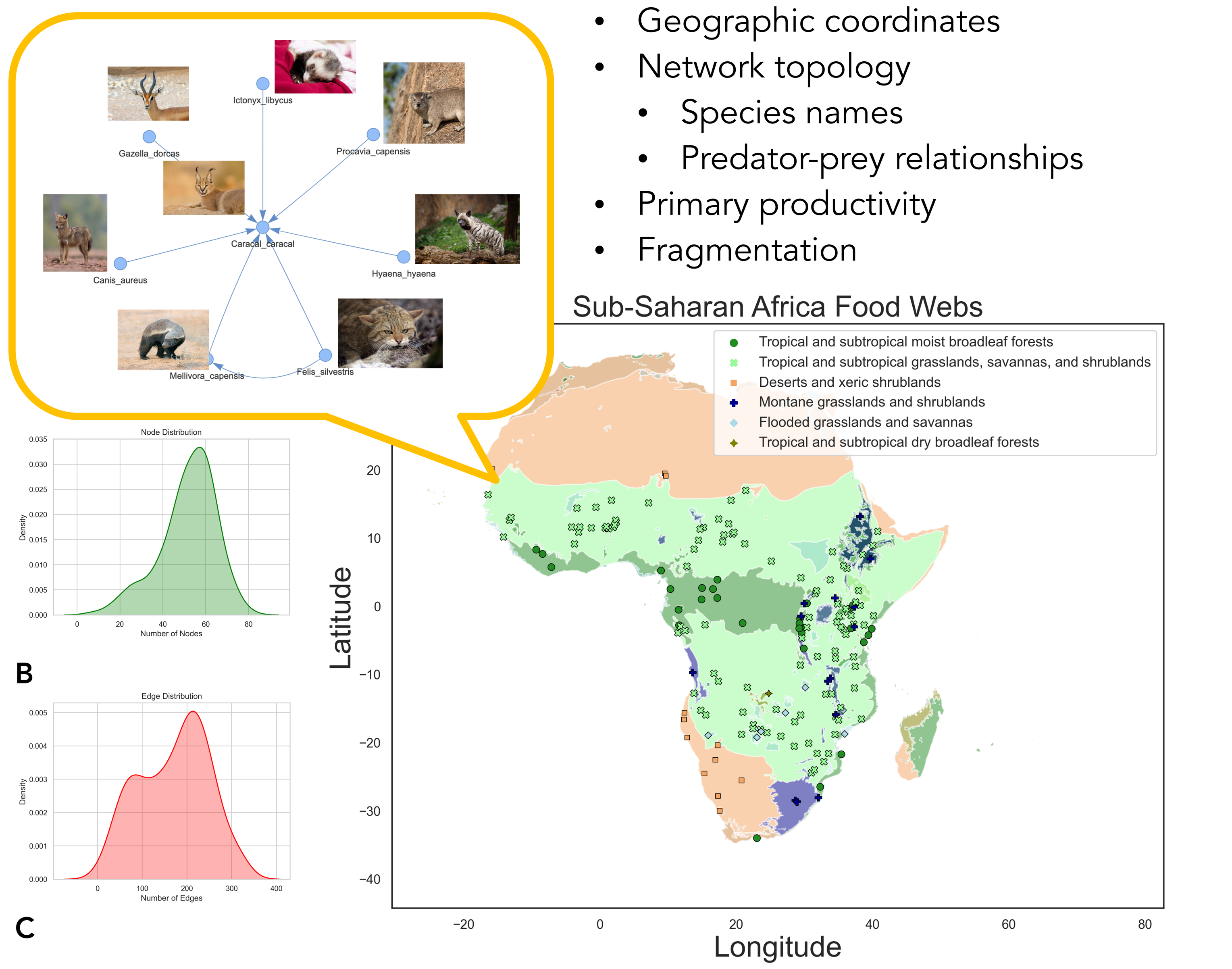
Background

- Predator-prey interactions in **food webs** play a critical role in maintaining biodiversity.
- Climate change causes species to shift their geographic ranges to track preferred preys.
- Understanding **changes among food webs** can thus inform local impact of climate change.

Data

Our group curated 170 mammal food web networks of Sub-Saharan Africa from literature. Attributes include but are not limited to:

- Geographic coordinates
- Network topology
  - Species names
  - Predator-prey relationships
- Primary productivity
- Fragmentation



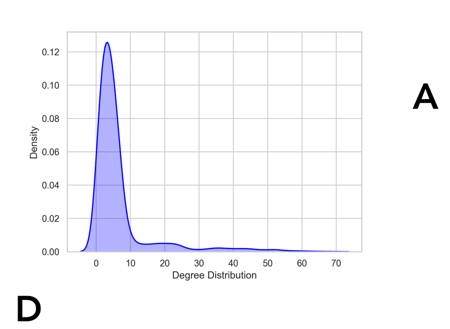
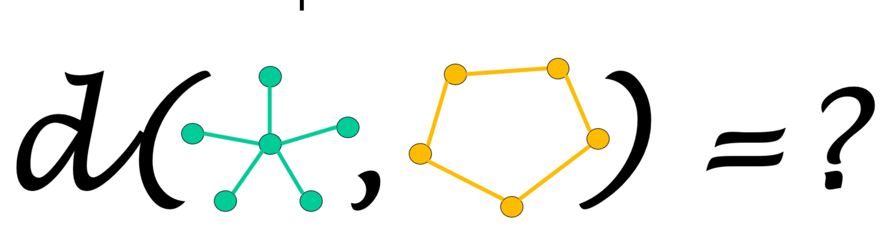


Figure 1: (A) Geographic distribution of the 170 mammal food webs colored by biome type. (B) The node distribution across all food web networks. (C) The edge distribution across all networks. (D) The degree distribution across all networks.

Problem of Interest

A Topological Data Analysis Toolkit for Food Webs

But… how do we compare networks?



Existing methods (ex. graph kernels, graph neural networks) are either computationally intensive, difficult to interpret, applicable only to a singular network, or do not apply to networks with different sizes.

Gromov-Wasserstein Distance

Definition 1: Given a graph $G = (V, E)$, we define its **measure network** as $M_G = (p_G, w_G)$ where p_G is the node distribution and $w_G : V \times V \rightarrow \mathbb{R}$ is a bounded edge weight function.

Definition 2: Given two measure networks M_{G_1} and M_{G_2}, we define the **discrete network Gromov-Wasserstein (GW) distance** by

$$d_{gw}(G_1, G_2) := \min_{\substack{T \in \mathbb{R}_{\geq 0}^{n_1 \times n_2},\\ T\mathbf{1} = p_{G_1},\\ T^\top \mathbf{1} = p_{G_2}}} \left\{ \sum_{u,u' \in V_1} \sum_{v,v' \in V_2} |w_{G_1}(u,u') - w_{G_2}(v,v')| T_{u,v} T_{u',v'} \right\}$$

where $T \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$ is an optimal transport (OT) plan between M_{G_1} and M_{G_2}.

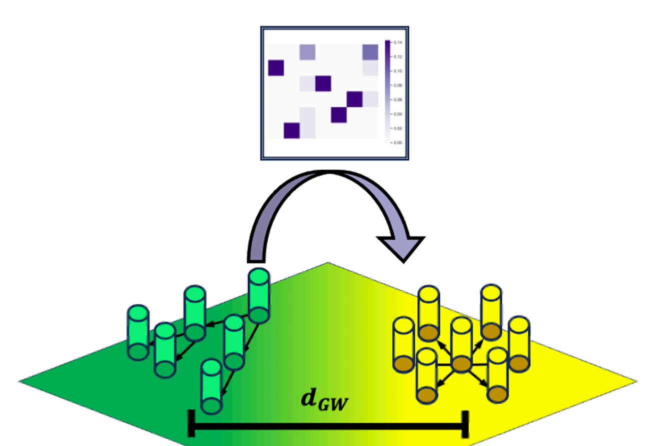


Figure 2: Conceptual illustration for the space characterized by d_{gw}.

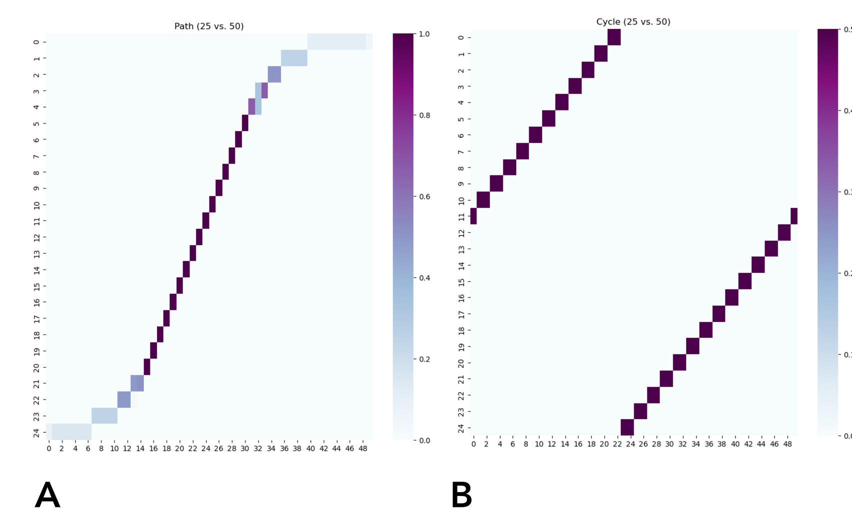


Figure 3: The optimal transport maps between (A) a 25-path and 50-path graph, and (B) a 25-cycle and 50-cycle graph.

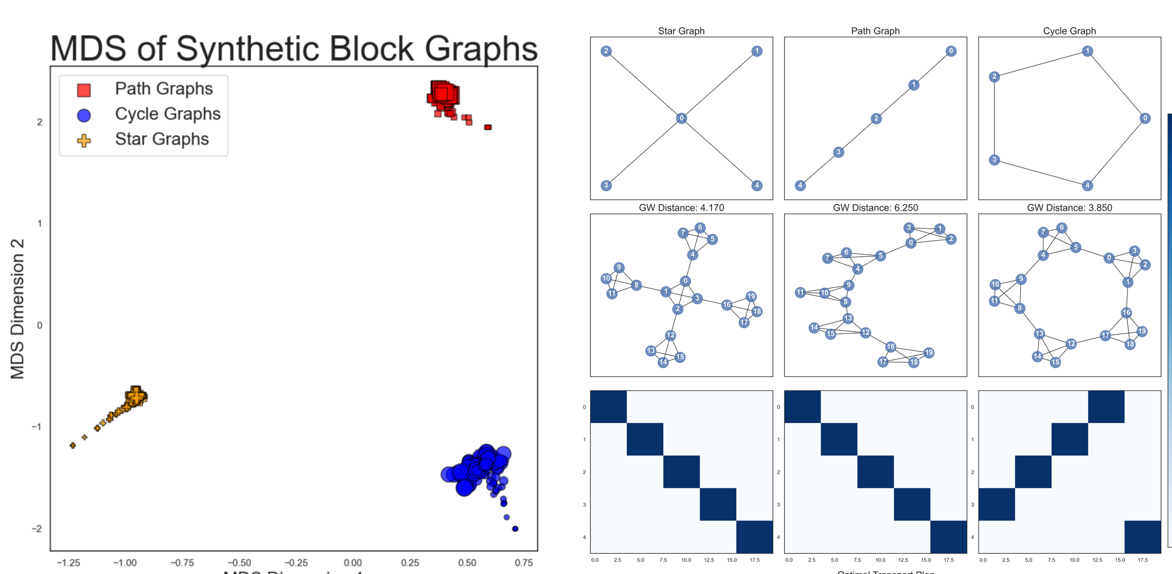


Figure 4: Numerical experiments for GW distance between block graphs of different sizes. (Left) MDS visualization. (Right) OT plans between select graphs.

Gromov-Wasserstein Barycenter

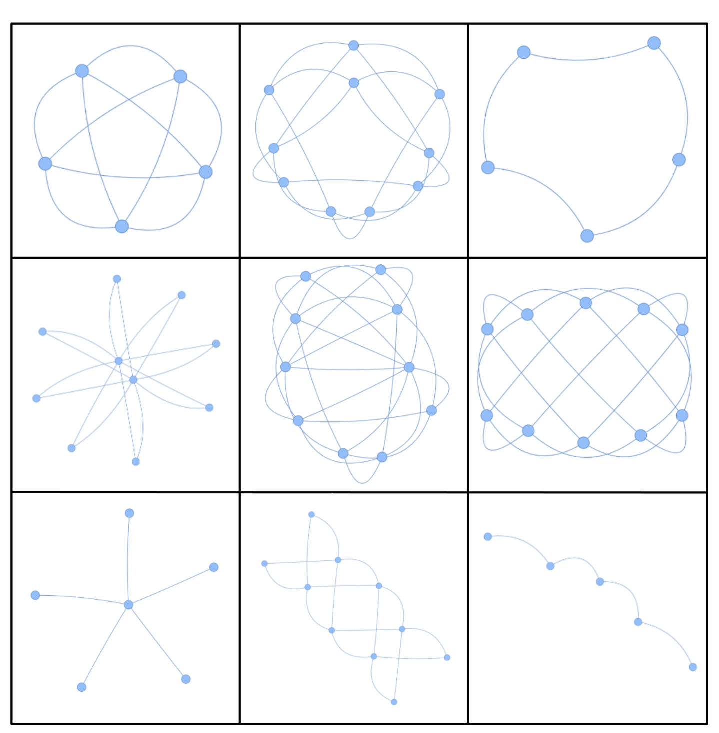


Figure 5: GW barycenter of complete graph (top-left), cycle graph (top-right), star graph (bottom-left), and path graph (bottom-right).

Definition 3: Let $\{C_s, p_s\}_{s=1}^{S}$ be S pairs of cost matrices C_s and node distribution p_s. The **Gromov-Wasserstein Barycenter**[2] $\mathscr{B}$ of $\{C_s, p_s\}_{s=1}^{S}$ is defined as

$$\mathscr{B} := \min_{C \in \mathbb{R}^{n \times n}} \sum_s \lambda_s d_{gw}(C, C_s, p, p_s)$$

where λ_s is the weight of (C_s, p_s).

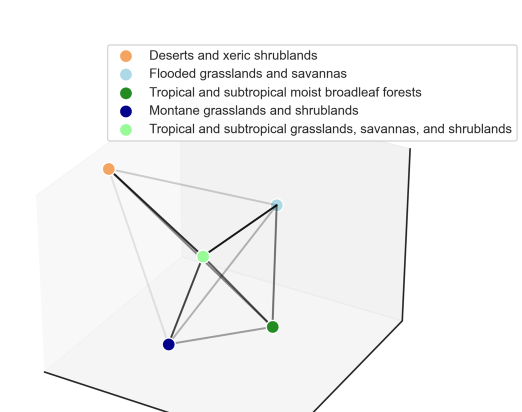


Figure 6: Distance between Barycenters of different biomes. Edge strength scaled with d_{gw}.

Gromov-Wasserstein on Food Webs

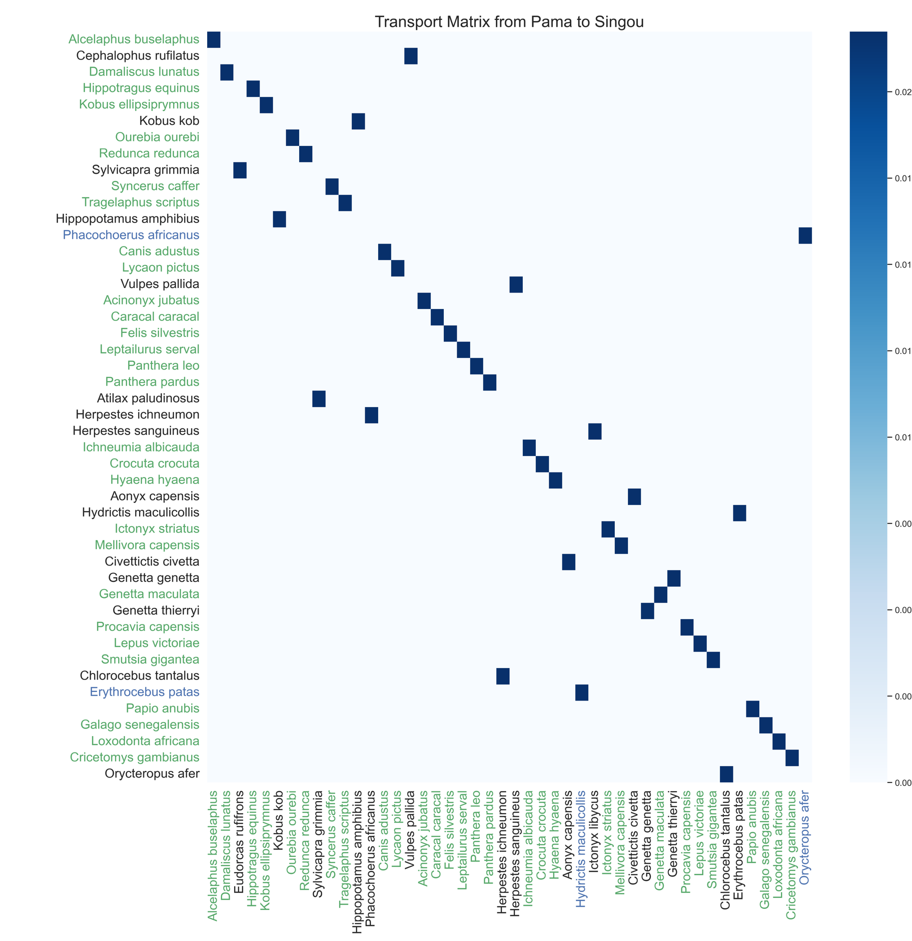


Figure 7: OT plan between mammal communities at Pama and Singou in Sub-Saharan Africa. Green highlights denote mass transported between shared species. Blue highlights surprising basal species mappings.

Site 1	Site 2	d_{gw}
Pama	Singou	1414.01
Singou	Tamou	1414.02
Pama	Tamou	1886.60
Air and Tenere	Ouadi Rime Ouadi Achim	1886.60
Mana Pools	Vwaza Marsh	2595.25
Lolldaiga	Mago	2772.26
Banhine	Mudumu	2865.89
Mudumu	Okavango Delta	3726.56
Kourtiagou	Madjoari	3773.17

Table 1: Top 9 most similar food webs under d_{gw}.

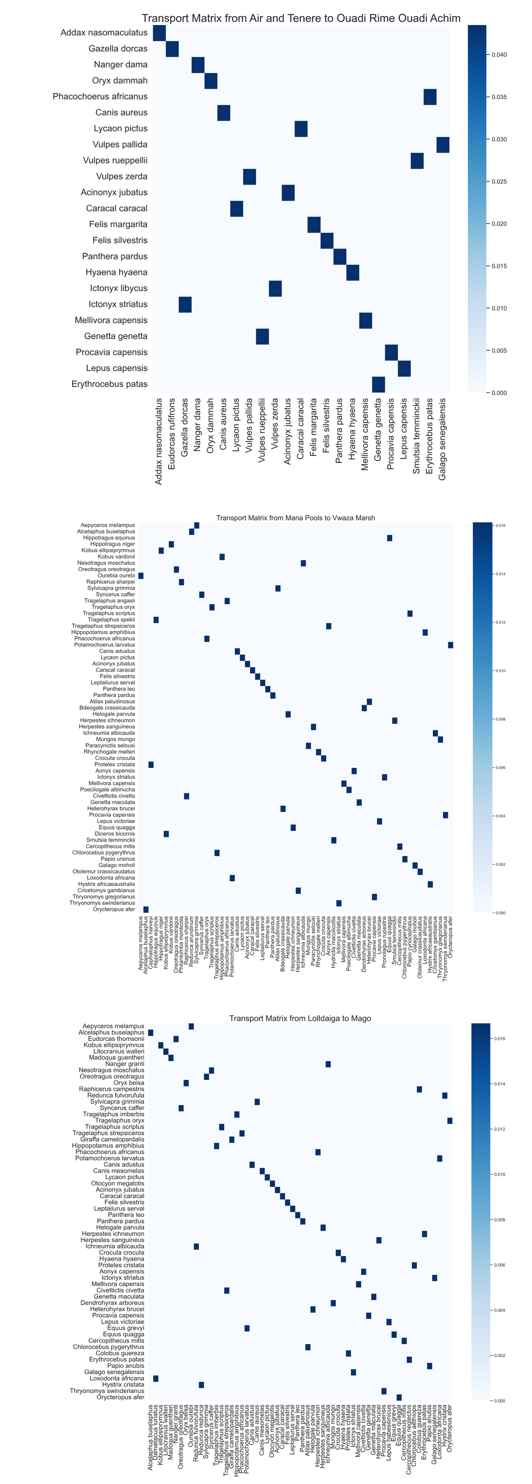


Figure 8: OT Plans for the top 3 similar webs from outside of the tropical and subtropical grasslands, savanna, and shrublands biome.

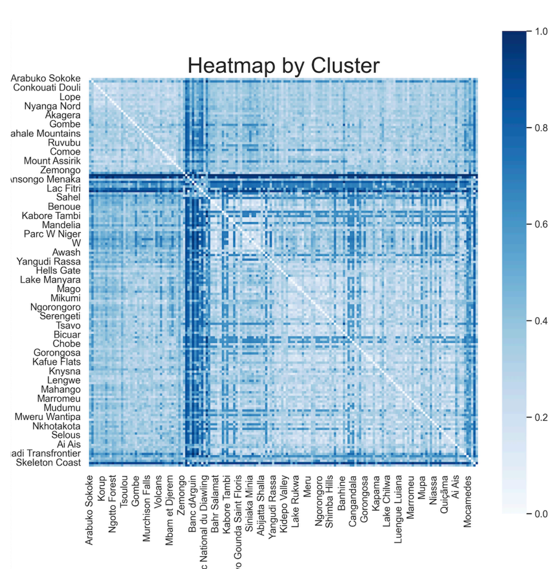


Figure 9: Heatmap of pairwise d_{gw} ordered by Sorensen-Dice cluster groups.

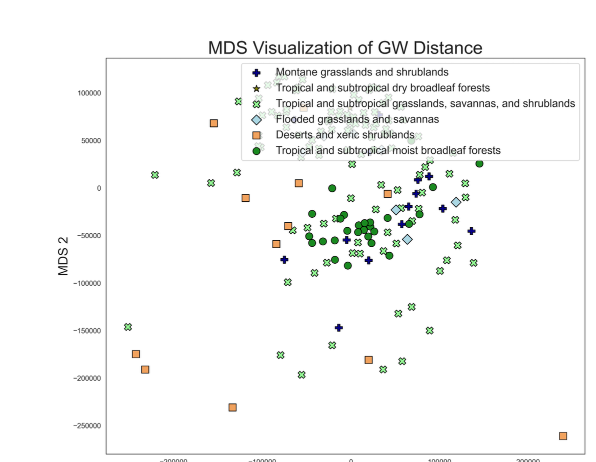


Figure 10: MDS over food webs.

Future Works

- Incorporate features beyond network topology via the **Fused Gromov-Wasserstein distance**.

References

1. Samir Chowdhury and Facundo Mémoli. The Gromov-Wasserstein distance between networks and stable network invariants. 2019
2. Peyré, G., Cuturi, M., and Solomon, J. Gromov-Wasserstein averaging of kernel and distance matrices. In Proceedings of the 33rd International Conference on Machine Learning, 2016.

For more information, you can reach me at kai.hung@rice.edu or visit my website via the QR code. This work is supported by NSF Grant No. 2213568. .