A The proof of Proposition 1

Proof. Because $\mathcal{K}$ is a Hilbert-Schmidt kernel, we have $\iint_{X\times X} \left|\mathcal{K}(\boldsymbol{x},\boldsymbol{y})\right|^2 \mathrm{d}\mu(\boldsymbol{x})\mathrm{d}\mu(\boldsymbol{y}) < \infty$. Then,

$$\begin{aligned}\iint_{X\times X} \left|\mathcal{H}^{(q)}(\boldsymbol{x},\boldsymbol{y})\right|^2 \mathrm{d}\mu(\boldsymbol{x})\mathrm{d}\mu(\boldsymbol{y}) &= \int_X\int_X \left|\frac{\mathcal{K}(\boldsymbol{x},\boldsymbol{y})+\mathcal{K}(\boldsymbol{y},\boldsymbol{x})}{2}\right|^2 \mathrm{d}\mu(\boldsymbol{x})\mathrm{d}\mu(\boldsymbol{y}) \\ &\leq \int_X\int_X \frac{1}{2}\big(\left|\mathcal{K}(\boldsymbol{x},\boldsymbol{y})\right|^2 + \left|\mathcal{K}(\boldsymbol{y},\boldsymbol{x})\right|^2\big)\mathrm{d}\mu(\boldsymbol{x})\mathrm{d}\mu(\boldsymbol{y}) \\ &= \int_X\int_X \left|\mathcal{K}(\boldsymbol{x},\boldsymbol{y})\right|^2 \mathrm{d}\mu(\boldsymbol{x})\mathrm{d}\mu(\boldsymbol{y}) < \infty\,.\end{aligned}$$

It holds that the symbol " $\leq$" in the third row of the above equation because of the Cauchy-Schwarz inequality. The Cauchy-Schwarz inequality states that $\forall a_i, b_i \in \mathbb{R}$, $i = 1,\cdots,k$, then $\big(\sum_{i=1}^k a_i b_i\big)^2 \leq \big(\sum_{i=1}^k a_i^2\big)\big(\sum_{i=1}^k b_i^2\big)$. Then, we have $\left|\frac{\mathcal{K}(\boldsymbol{x},\boldsymbol{y})}{2} + \frac{\mathcal{K}(\boldsymbol{y},\boldsymbol{x})}{2}\right|^2 \leq (\frac{1}{4}+\frac{1}{4})\cdot\big(\left|\mathcal{K}(\boldsymbol{x},\boldsymbol{y})\right|^2 + \left|\mathcal{K}(\boldsymbol{y},\boldsymbol{x})\right|^2\big)$, and measure μ is non-negative. Thus, symbol " $\leq$" in the third row holds. We finish the proof. $\square$

B The proof of Proposition 2

Proof. The associated kernel $\mu(\boldsymbol{x},\boldsymbol{y},q)$ of the operator (5) is a Hermitian kernel, i.e., $\mu(\boldsymbol{x},\boldsymbol{y},q) = \overline{\mu(\boldsymbol{y},\boldsymbol{x},q)}$, then the eigenvalues of (5) are real and the spectral radius is equal to the operator norm $R(\rho) = \|T^{(q)}\|$ [36]. Let $f \in L^2(X,\mu)$, we have

$$\begin{aligned}\langle T^{(q)} f, f\rangle &= \int_X\int_X \rho(\boldsymbol{x},\boldsymbol{y},q) f(\boldsymbol{y})\overline{f(\boldsymbol{x})}\mathrm{d}\mu(\boldsymbol{x})\mathrm{d}\mu(\boldsymbol{y}) \\ &= \int_X\int_X \mathcal{H}^{(q)}(\boldsymbol{x},\boldsymbol{y})\frac{f(\boldsymbol{y})}{\sqrt{m(\boldsymbol{y})}}\frac{\overline{f(\boldsymbol{x})}}{\sqrt{m(\boldsymbol{x})}}\mathrm{d}\mu(\boldsymbol{x})\mathrm{d}\mu(\boldsymbol{y})\,.\end{aligned}$$

If we apply the Cauchy-Schwarz inequality as follows,

$$\begin{aligned}\left|\int_X \mathcal{H}^{(q)}(\boldsymbol{x},\boldsymbol{y})\frac{f(\boldsymbol{y})}{\sqrt{m(\boldsymbol{y})}}\mathrm{d}\mu(\boldsymbol{y})\right| &\leq \bigg(\int_X \left|\mathcal{H}^{(q)}(\boldsymbol{x},\boldsymbol{y})\right|\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}}\bigg(\int_X \left|\mathcal{H}^{(q)}(\boldsymbol{x},\boldsymbol{y})^{\frac{1}{2}}\frac{f(\boldsymbol{y})}{\sqrt{m(\boldsymbol{y})}}\right|^2\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}} \\ &\leq \bigg(\int_X \left|\mathcal{H}^{(q)}(\boldsymbol{x},\boldsymbol{y})\right|\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}}\bigg(\int_X \left|\mathcal{H}^{(q)}(\boldsymbol{x},\boldsymbol{y})\right|\frac{|f(\boldsymbol{y})|^2}{m(\boldsymbol{y})}\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}} \\ &= \bigg(\int_X S(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}}\bigg(\int_X S(\boldsymbol{x},\boldsymbol{y})\frac{|f(\boldsymbol{y})|^2}{m(\boldsymbol{y})}\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}},\end{aligned}$$

where $S(\boldsymbol{x},\boldsymbol{y}) = \left|\mathcal{H}^{(q)}(\boldsymbol{x},\boldsymbol{y})\right| = \frac{1}{2}\left|\mathcal{K}(\boldsymbol{x},\boldsymbol{y})+\mathcal{K}(\boldsymbol{y},\boldsymbol{x})\right|$. The asymmetric kernel $\mathcal{K}$ is non-negative, thus $S(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{2}\big(\mathcal{K}(\boldsymbol{x},\boldsymbol{y})+\mathcal{K}(\boldsymbol{y},\boldsymbol{x})\big)$. Then we have

$$\begin{aligned}\left|\int_X \mathcal{H}^{(q)}(\boldsymbol{x},\boldsymbol{y})\frac{f(\boldsymbol{y})}{\sqrt{m(\boldsymbol{y})}}\mathrm{d}\mu(\boldsymbol{y})\right| &\leq \bigg(\int_X S(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}}\bigg(\int_X S(\boldsymbol{x},\boldsymbol{y})\frac{|f(\boldsymbol{y})|^2}{m(\boldsymbol{y})}\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}} \\ &= \sqrt{m(\boldsymbol{x})}\bigg(\int_X S(\boldsymbol{x},\boldsymbol{y})\frac{|f(\boldsymbol{y})|^2}{m(\boldsymbol{y})}\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}}.\end{aligned}$$

Consequently,

$$\langle T^{(q)} f, f\rangle \leq \int_X \left|f(\boldsymbol{x})\right|\bigg(\int_X S(\boldsymbol{x},\boldsymbol{y})\frac{|f(\boldsymbol{y})|^2}{m(\boldsymbol{y})}\mathrm{d}\mu(\boldsymbol{y})\bigg)^{\frac{1}{2}}\mathrm{d}\mu(\boldsymbol{x}).$$

We can apply the Cauchy-Schwarz inequality again,

$$\begin{aligned}
\langle T^{(q)} f, f\rangle &\leq \|f\| \left(\int_X \int_X S(\boldsymbol{x},\boldsymbol{y}) \frac{|f(\boldsymbol{y})|^2}{m(\boldsymbol{y})} \mathrm{d}\mu(\boldsymbol{y}) \mathrm{d}\mu(\boldsymbol{x}) \right)^{\frac{1}{2}} \\
&= \|f\| \left(\int_X \frac{|f(\boldsymbol{y})|^2}{m(\boldsymbol{y})} \left(\int_X S(\boldsymbol{x},\boldsymbol{y}) \mathrm{d}\mu(\boldsymbol{x}) \right) \mathrm{d}\mu(\boldsymbol{y}) \right)^{\frac{1}{2}} \\
&= \|f\| \left(\int_X \frac{|f(\boldsymbol{y})|^2}{m(\boldsymbol{y})} m(\boldsymbol{y}) \mathrm{d}\mu(\boldsymbol{y}) \right)^{\frac{1}{2}} = \|f\|^2 .
\end{aligned}$$

It can be noticed that the operator norm $\|T^{(q)}\| = 1$. Therefore, the spectral radius $R(\rho) = 1$. We finish the proof.

□

C Selection of the scaling parameter q.

As illustrated in Fig. F1, we suggest choosing a period that encompasses the range of the skew-symmetric component, i.e., $\overline{a}$ should be less than π, where $\overline{a}$ is defined as follows,

$$\overline{a} := \sup_{\boldsymbol{x},\boldsymbol{y} \in X} |\mathcal{K}(\boldsymbol{x},\boldsymbol{y}) - \mathcal{K}(\boldsymbol{y},\boldsymbol{x})| ,$$

and the period of the phase function is $T = \frac{2\pi}{2\pi q} = 1/q$. Thus, we have

$$T = \frac{1}{q} > 2\overline{a} \quad \Rightarrow \quad q < \frac{1}{2\overline{a}} .$$

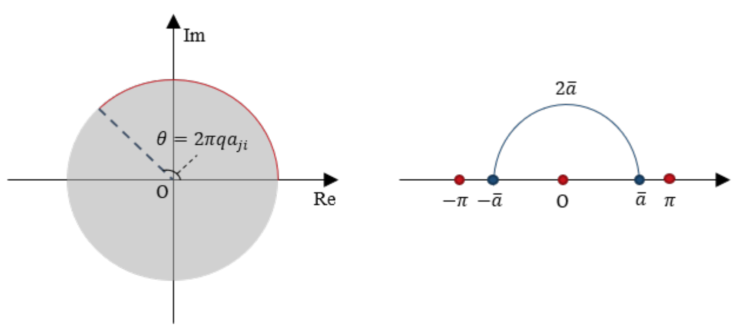


Figure F1: A simple illustration for selecting q.

D Algorithm

To apply MagDM, we require a dataset and an asymmetric kernel or an asymmetric Gram matrix, as well as the scaling parameter q and the desired accuracy. Algorithm 1 outlines the MagDM procedure.

Algorithm 1 MagDM for asymmetric kernels

Input: The Gram matrix $\boldsymbol{K}$ of dataset X endowed with an asymmetric kernel $\mathcal{K}$, the scaling parameter q and a preset accuracy δ.
Output: The diffusion map $\psi^{t,(q)}$ of X.
1: Calculate the Hermitian Gram matrix $\boldsymbol{H}$ of the asymmetric Gram matrix $\boldsymbol{K}$ by (3) and (4).
2: Calculate the t-powers kernel matrix $\boldsymbol{H}^t$.
3: Run eigen-decomposition of $\boldsymbol{H}^t$ and denote its eigen-system as $\{\lambda_n^{(q)}, \phi_n^{(q)}\}$.
4: $s(\delta, t) \leftarrow \max\{n \in \mathbb{N} : |\lambda_n^{(q)}| > \delta|\lambda_1^{(q)}|\}$.
5: Return the diffusion map $\psi^{t,(q)}$ by (8).

Limitations. Researchers should note that the MagDM method proposed in this paper has some limitations. One such limitation is its dependence on the choice of asymmetric kernel functions, which can impact its performance. Additionally, MagDM may be computationally expensive for large datasets, as it requires $\mathcal{O}(N^2)$ memory and $\mathcal{O}(N^3)$ computational complexity to derive the spectral decomposition. However, this limitation can be addressed through the use of out-of-sample extensions, which are discussed in Appendix E.

E Out-of-sample extensions.

Out-of-sample extensions are useful in many applications where low-dimensional embeddings computed on the original dataset are extended to new data. The Nyström extension is a well-known technique used in the machine learning community to approximate the Gram matrix by a low-rank embedding. However, the out-of-sample extension of duffision maps for asymmetric kernels has not been studied before. Here, we present the corresponding Nyström-based extension for out-of-sample cases. As discussed earlier, the integral operator (5) is compact and self-adjoint, whose spectral decomposition is $\{\lambda_n^{(q)}, \phi_n^{(q)}\}$. If $\lambda_n^{(q)} \neq 0$, the following identity holds for $\boldsymbol{x} \in X$:

$$\phi_n^{(q)}(\boldsymbol{x}) = \frac{T^{(q)}}{\lambda_n^{(q)}} \phi_n^{(q)}(\boldsymbol{x}) = \int_X \frac{\rho(\boldsymbol{x}, \boldsymbol{y}, q)}{\lambda_n^{(q)}} \phi_n^{(q)}(\boldsymbol{y}) \mathrm{d}\mu(\boldsymbol{y}) .$$

The Nyström extension extends the equation above to new data Z such that $X \subseteq Z$ as follows,

$$\phi_n^{(q)}(\boldsymbol{z}) = \int_X \frac{\rho(\boldsymbol{z}, \boldsymbol{y}, q)}{\lambda_n^{(q)}} \phi_n^{(q)}(\boldsymbol{y}) \mathrm{d}\mu(\boldsymbol{y}) , \tag{a1}$$

where $\boldsymbol{z} \in Z$ and $\phi_n^{(q)}(\boldsymbol{z}) = \sum_{\boldsymbol{y} \in X} \frac{\rho(\boldsymbol{z}, \boldsymbol{y}, q)}{N\lambda_n^{(q)}} \phi_n^{(q)}(\boldsymbol{y})$ is the empirical form of (a1) for X. This allows the eigenfunctions to be extended for new data, enabling the extension of MagDM (8) as follows,

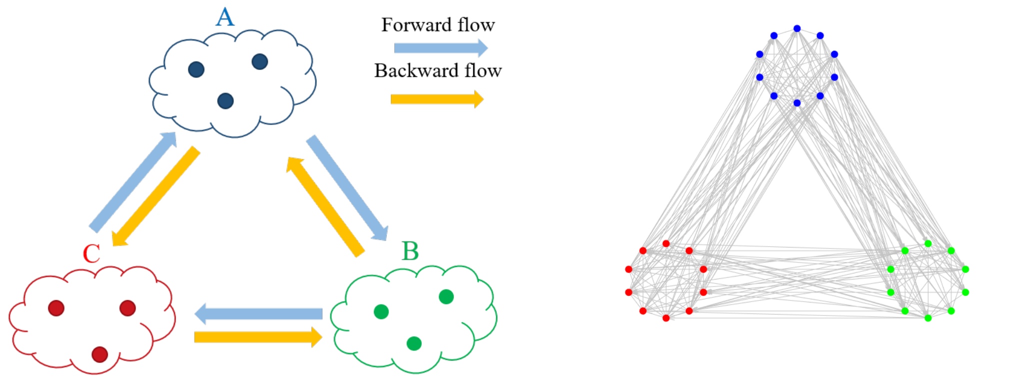
$$\psi^{t,(q)}(\boldsymbol{z}) = \sum_{\boldsymbol{y} \in X} \left[\frac{\rho(\boldsymbol{z}, \boldsymbol{y}, q)}{N} \phi_1^{(q)}(\boldsymbol{y}), \quad \frac{\rho(\boldsymbol{z}, \boldsymbol{y}, q)}{N} \phi_2^{(q)}(\boldsymbol{y}), \quad \cdots , \quad \frac{\rho(\boldsymbol{z}, \boldsymbol{y}, q)}{N} \phi_{s(\delta,t)}^{(q)}(\boldsymbol{y}) \right]^\top .$$

F Descriptions and figures of datasets

In this section, we visualize the datasets using either expert knowledge or a force-directed layout. We hope that these visualizations will help readers gain a better understanding of the data.

F.1 The first artificial network

Fig. F2(a) provides the running flow of the first artificial network and Fig. F2(b) shows an example of the directed graphs generated with $P = 0$, where the asymmetric adjacency connection can be considered as an asymmetric kernel.

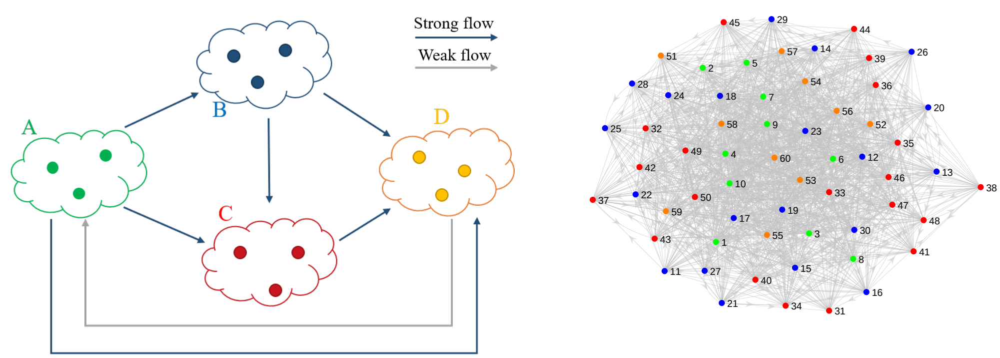


(a) The running flow of three groups. (b) Graph using the expert knowledge positions.

Figure F2: An illustration of the first artificial network. (a) The running flow of three groups A, B and C. The directed/asymmetric information is nested in the running flow. (b) An instance of a directed graph generated by the running flow with backward flow probability $P = 0$.

F.2 The second artificial network

The running flow of the second artificial network comprises four groups (A, B, C, and D). The structure of the flow is apparent, with groups A and D serving as out-come and in-come nodes, respectively, while groups B and C function as communicators. Groups B and C are two dense sets containing 20 nodes and Group A and D are a pair of datasets whose interconnections are much more than Group B and C.



(a) The running flow of four groups. (b) Graph using a force-directed layout.

Figure F3: An illustration of the second artificial network. (a) The running flow of three groups A, B, C and D. The directed/asymmetric information is nested in the running flow. (b) An instance of a directed graph generated by the running flow with Groups A and D playing a particular role (green and orange nodes) and Groups B and C playing a role of a communicator (red and blue nodes).

F.3 The Möbius strip

The Möbius strip dataset is a set of 300 points randomly distributed along the Möbius strip. The parametric form of the Möbius strip is defined by,

$$x(u, v) = \left(1 + \frac{v}{2} \cos \frac{u}{2}\right) \cos u, \quad y(u, v) = \left(1 + \frac{v}{2} \cos \frac{u}{2}\right) \sin u, \quad z(u, v) = \frac{v}{2} \sin \frac{u}{2},$$

where $0 \leq u \leq 2\pi$ and $-0.5 \leq v \leq 0.5$. The dataset is with a color drift in the counterclockwise direction on the x-y plane in Fig. F4.

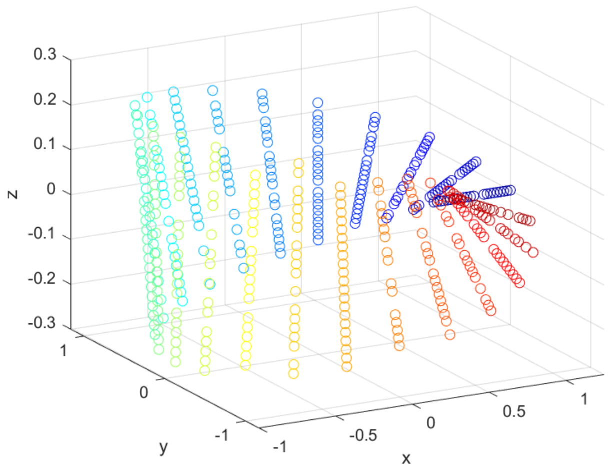
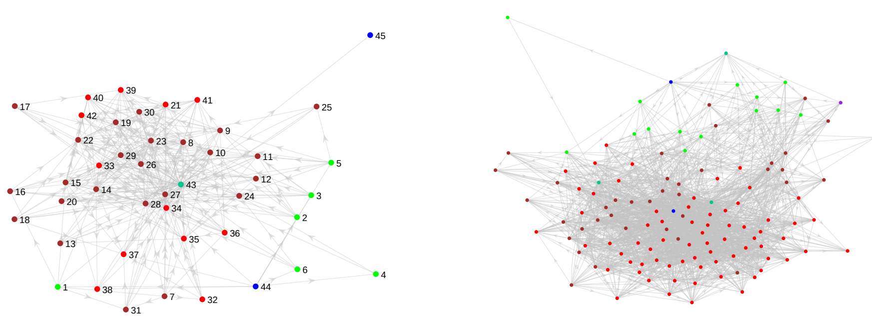


Figure F4: Dataset with 300 random points in the Möbius strip.

F.4 Two trophic networks

We have chosen two specific trophic networks: the Mondego [32] and Florida [33] networks, which are part of the Pajek datasets. These networks have recorded the trophic exchanges at Mondego estuary and Florida bay during the wet season, respectively. Based on the roles of the nodes in these ecosystems, we have classified them into different categories, as shown in Appendix F5. The green nodes, such as 2um Spherical Phytoplankt and Phytoplankton, are producers that generate their own food through photosynthesis or chemosynthesis. The brown and red nodes are low-level consumers like littorina and high-level consumers like bonefish and crocodiles that feed on other organisms for energy. Additionally, the purple node in the Florida network represents decomposers that break down dead or decaying organic matter. Finally, the blue and turquoise blue nodes correspond to the input, output, and organic matter of the ecosystem, respectively.



(a) Mondego.　　(b) Florida.

Figure F5: The trophic networks using force-direct layout. Nodes among this network are classified into several categories, The green nodes are producers that generate their own food through photosynthesis or chemosynthesis. The brown and red nodes are low-level consumers and high-level consumers that feed on other organisms for energy. Additionally, the purple node in the Florida network represents decomposers that break down dead or decaying organic matter. The blue and turquoise blue nodes correspond to the input, output, and organic matter of the ecosystem.