IMPERIAL

Topological Graph Kernels from Tropical Geometry

The Geometric Realization of AATRN (Applied Algebraic Topology Research Network)
IMSI Chicago

Y. Cao & AM (2025). Metric Graph Kernels via the Tropical Torelli Map, arXiv:2505.12129.

Anthea Monod
18-22 August 2025

From Graphs to Metric Graphs

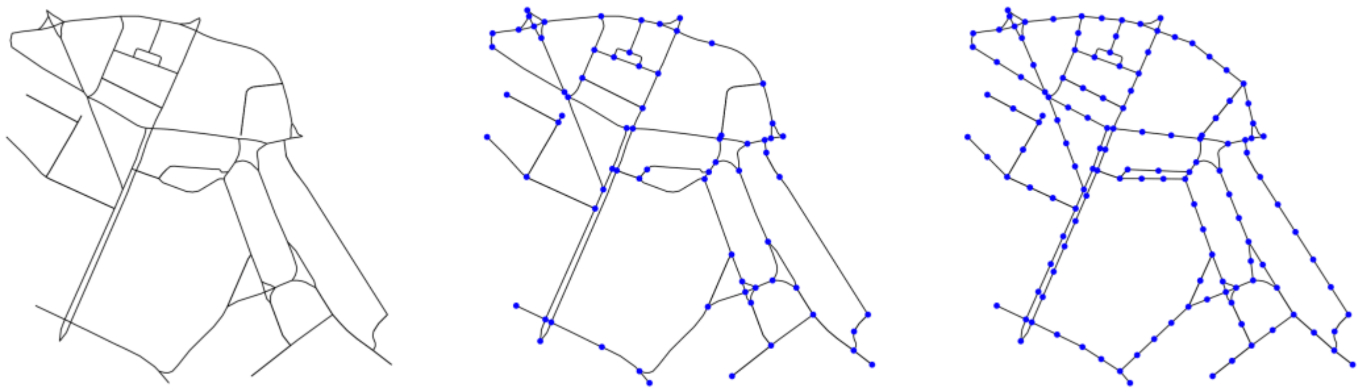
- A graph is made up of discrete sets of nodes and edges
- A *metric graph* is a 1-dimensional metric space that can be realized as the underlying space of a graph
  $\longrightarrow$ Metric graphs are geometric realizations of graphs with a length function on their edges!

Edge Subdivisions and Refinements

- For a graph $G = (V, E)$ and $\ell: E \to \mathbb{R}_+$ a length function, an *edge subdivision* of $e = [u, v] \in E$ is an operation on G:
  1. Add a new node w to V
  2. Replace e by two edges $e' = [u, w]$ and $e'' = [w, v]$ such that $\ell(e) = \ell(e') + \ell(e'')$
- G' is a *refinement* $G' \geq G$ if G' can be obtained from G by a sequence of edge subdivisions

Notice that refinement is different from subgraph inclusion!

- If G' is a refinement, then there exists an injection $V(G) \to V(G')$ and a surjection $E(G') \to E(G)$
- If G is a subgraph of G', then $V(G) \to V(G')$ and $E(G') \to E(G)$ are both inclusions



Road network near the Victoria tube station in London

Graph Kernels and Metric Graph Kernels

Kernels are a central class of similarity function in machine learning and data analysis
They capture nonlinear relationships by embedding data (graphs) into high dimensional feature spaces

Let $\mathcal{X}$ be a set and $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric function. Typically,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j k(x_i, x_j) \geq 0$$

For any positive definite kernel k, there exists a unique Hilbert space (RKHS) $\mathcal{H}$ and feature map $\phi: \mathcal{X} \to \mathcal{H}$ such that

$$k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$$

A popular way to construct kernels is to use distance functions $d(x, y)$ on $\mathcal{X}$
Schoenberg's Theorem (Sejdinovic et al., 2013): If $d(x, y)$ is conditionally negative type

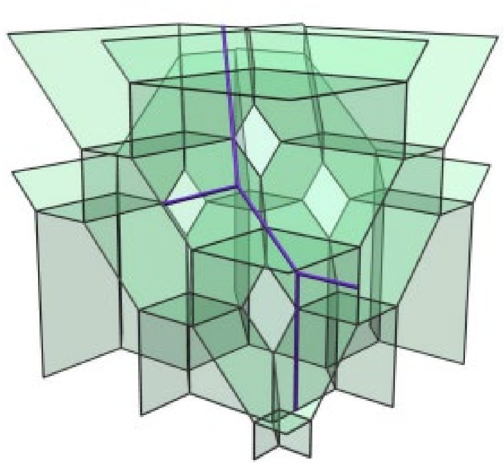
$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j d(x_i, x_j) \leq 0$$

Then the RBF-defined kernel $k(x, y) = \exp(-\gamma d^2(x, y))$ is positive definite for any scaling parameter $\gamma > 0$.

Definition (Cao & AM, 2025): A graph kernel k is called a *metric graph kernel* if $k(G_1', G_2') = k(G_1, G_2)$ for any refinements $G_1' \geq G_1$ and $G_2' \geq G_2$.
$\longrightarrow$ AFAIK, all existing graph kernels are based on nodes, edges, and subgraphs so they fail to be metric graph kernels!

Algebraic and Tropical Geometry

An algebraic variety [Picture This Maths]



A tropical variety [Böhm et al., 2017]

Algebraic Geometry

Locus of zeros of polynomial systems

Tropical Geometry

Boundaries between linear parts of tropical polynomials

Torelli Maps

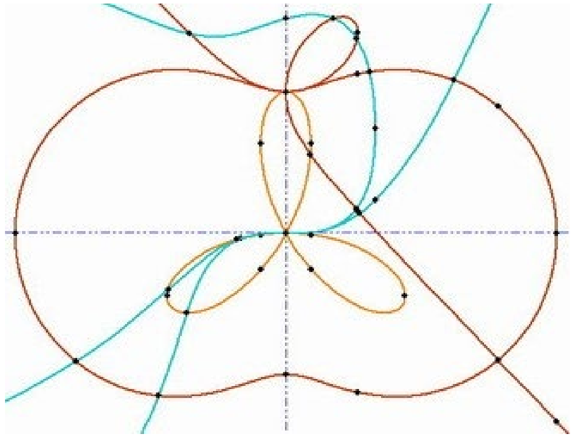
In classical algebraic geometry over the complex numbers:

Every smooth projective algebraic curve X of genus g encodes rich geometric information in Jac$(X) = \mathbb{C}^g/L$

The Torelli map sends each curve X to its Jacobian Jac(X); it is injective!

We also have Torelli maps in tropical geometry!

- Tropical curves are not smooth curves but rather *metric graphs* Γ
- Jac$(\Gamma) = \mathbb{R}^g/L \longrightarrow$ To compute stuff, we have to choose a vector space basis for $\mathbb{R}^n$ and a lattice basis for L (more on this later)



A complex algebraic curve [Peter Beelen]

The Tropical Torelli Map for Weighted Graphs: Sending a Graph to a Matrix

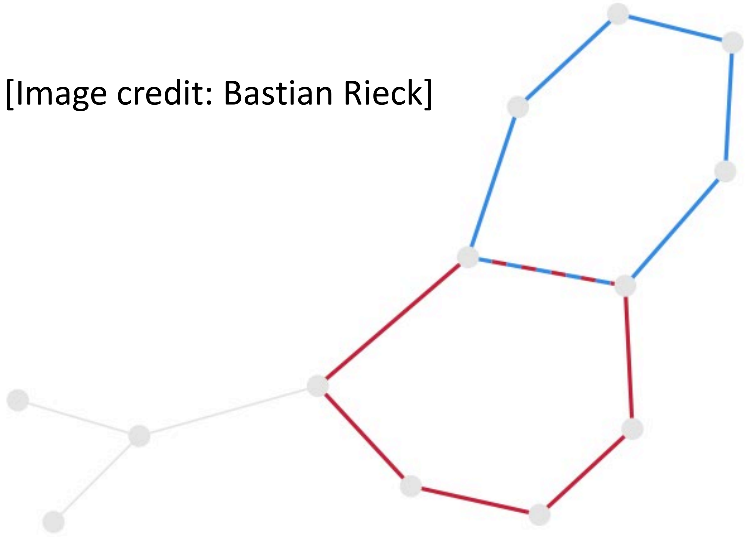
- Let $G = (V, E)$ be a graph with a positive weight function $\ell: E \to \mathbb{R}_+$
- Identify each edge e as $[0, \ell(e)]$ and glue all intervals to get $|G|$
- $|G|$ is in 1:1 correspondence with abstract tropical curves in tropical geometry (Chan, 2021; Cao & AM, 2025a)
- The classical tropical Torelli map sends any metric graph to a flat torus
- For a weighted graph with *generic length function*, compute a unique SPD matrix Q to represent the flat torus
- **Theorem** (Cao & AM, 2025b): $Q(G) = Q(G')$ for any refinement G' of G
- $Q(G)$ contains the intrinsic geometric and topological information on the underlying metric graph $|G|$

Mapping a Graph to a Matrix Using 1-Cycles

Graph Homology

- Let $G = (V, E)$ be a connected graph with n nodes and m edges
- Let $C_0(G; \mathbb{R})$ be the n-dim'l vector space spanned by V;
  $C_1(G; \mathbb{R})$ be the m-dim'l vector space spanned by E
- The boundary map $\partial: C_1(G; \mathbb{R}) \to C_0(G; \mathbb{R})$ is a linear map, $\partial([u, v]) = v - u$
- The 1-homology group is $H_1(G; \mathbb{R}) = \ker(\partial)$
- Its dimension $g = \dim\big(H_1(G; \mathbb{R})\big) = m - n + 1$ is called the *genus* of G;
  any element of $\sigma \in H_1(G; \mathbb{R})$ is called a *1-cycle*

[Image credit: Bastian Rieck]



Representing the Graph as a Matrix of 1-Cycles

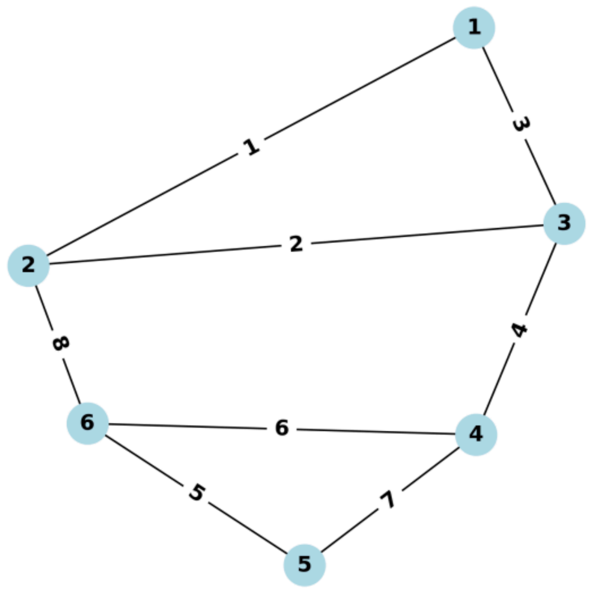
- Define an inner product Q_G on $C_1(G; \mathbb{R})$ by $Q_G\big(e_i, e_j\big) = \delta_{ij}\sqrt{\ell(e_i)\ell(e_j)}$
- Notice that under Q_G, an element $e_i \in E$ has norm $\sqrt{\ell(e_i)}$ when viewed as a 1-chain in $C_1(G; \mathbb{R})$ rather than its length $\ell(e_i)$ when viewed as an edge in G (Ji, 2012)
- The inner product is compatible with edge subdivision: For e subdivided into e', e''

$$\|e' + e''\|_Q^2 = \ell(e') + \ell(e'') = \ell(e) = \|e\|_Q^2$$

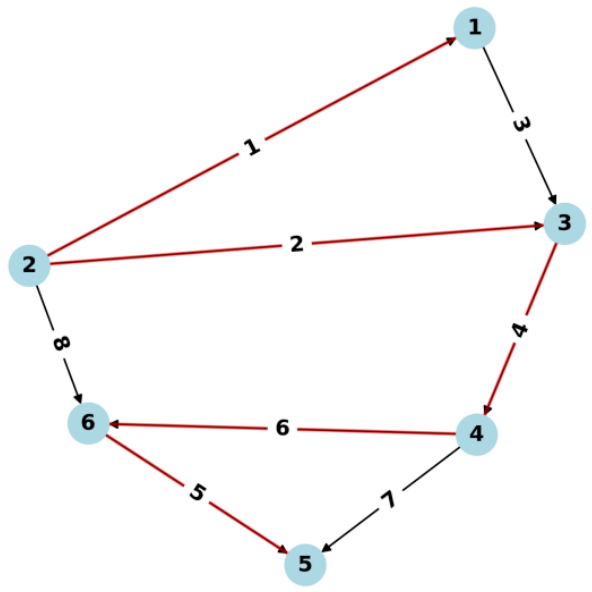
- Then fix a 1-cycle basis $\sigma_1, \dots, \sigma_g$ for $H_1(G; \mathbb{R})$ and the inner product Q_G is represented by a matrix Q

Definition (Cao & AM, 2025b): Let $\mathcal{M}_G$ be the space of weighted graphs of genus g and $SPD(g)$ be the space of $g \times g$ SPD matrices. The *tropical Torelli map* is given by $\mathcal{T}: \mathcal{M}_G \to SPD(g), G \mapsto Q(G)$

Torelli Algorithm



(a) Example graph



(b) Orientation and minimal spanning tree

1. Compute a minimal spanning tree T
2. Fix an orientation, each edge *not* in T determines a 1-cycle:

$$\sigma_1 = e_3 - e_2 + e_1;\ \sigma_2 = e_7 - e_5 - e_6;\ \sigma_3 = e_8 - e_6 - e_4 - e_2$$

3. Write down the cycle–edge incidence matrix $M = \begin{bmatrix} 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 \end{bmatrix}$

4. Compute $Q = MLM^T = \begin{bmatrix} 6 & 0 & 2 \\ 0 & 18 & 6 \\ 2 & 6 & 20 \end{bmatrix}$

Uniqueness & Computational Complexity

The output becomes unique when we assume genericity of length functions:
Definition: Let $G = (V, E)$ be a connected graph. A length function $\ell: E \to \mathbb{R}_+$ is *generic* if G has a unique minimal spanning tree and the lengths of edges are distinct.

If G is equipped with a generic length function, it admits a canonical orientation

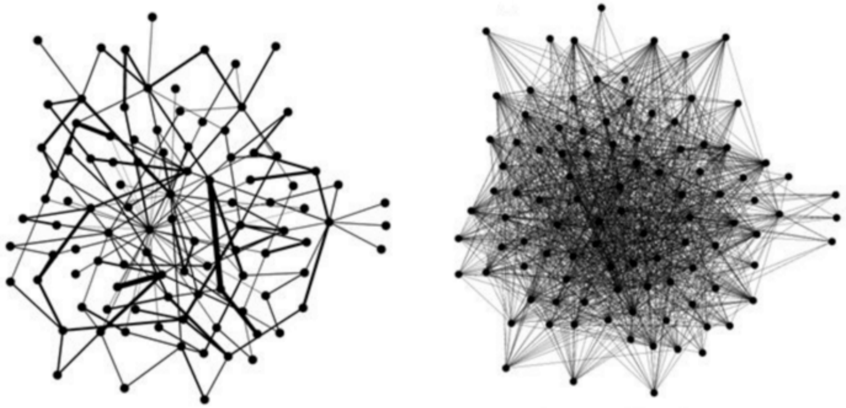
Theorem (Cao & AM, 2025b): Let $G = (V, E)$ be a connected graph with a generic length function ℓ. Under the canonical orientation, Q is unique and invariant under any refinement of G.

Complexity of the Tropical Torelli Map: $O(gn(g + \log n))$

- Compute the reduced cycle–edge incidence matrix: $O(gn \log n)$
- Compute Q: $O(g^2 n)$

Three Sparsity Classes Based on g

1. Sparse graphs: $m \asymp n + c \Longrightarrow g = O(1)$ then the complexity is $O(n \log n)$
2. Semi-sparse graphs: $m \asymp cn \Longrightarrow g = O(n)$ then the complexity is $O(n^3)$
3. Dense graphs: $m \asymp n^{1+c} \Longrightarrow g = O(n^{1+c})$ then the complexity is $O(n^{3+2c})$



[Image credit: Emory Oxford College]

Tropical Metric Graph Kernels

- Q is a symmetric positive definite matrix (SPD)
- To compare graphs with different genus, we enlarge the embedding space to positive semi-definite matrices (PSD):
  - Pad Q by zero if $\dim(Q) < g_0$
  - Extract a submatrix by randomly choosing g_0 rows and columns if $\dim(Q) > g_0$

The Tropical Torelli–Euclidean Kernel:

$$k_{TTE}(G_1, G_2) = \exp(-\gamma \|Q(G_1) - Q(G_2)\|_F^2)$$

The Tropical Torelli–Wasserstein Kernel:

$$k_{TTW}(G_1, G_2) = \exp(-\gamma W^2(Q(G_1), Q(G_2)))$$

where $W^2(Q(G_1), Q(G_2)) = \mathrm{Tr}(Q(G_1)) + \mathrm{Tr}(Q(G_2)) - 2\mathrm{Tr}\left(\sqrt{\sqrt{Q(G_1)}Q(G_2)\sqrt{Q(G_1)}}\right)$

This derives from the inner product form $\left(\mathbb{R}^g/\sqrt{Q}\,, \|\cdot\|_2\right)$ because remember, Jac$(\Gamma) = \mathbb{R}^g \,/\, L$

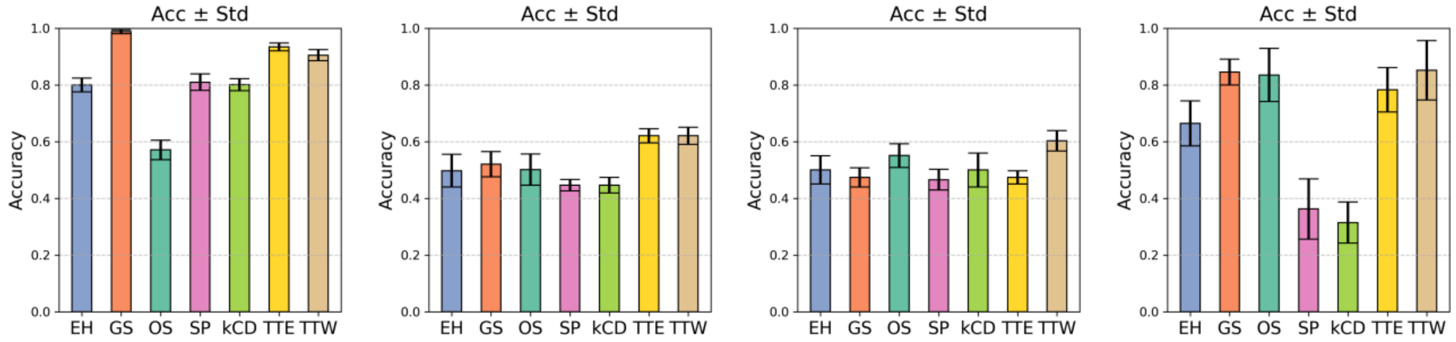
Computational Complexity: For N matrices Q_i of dimension g_0, the time complexity for the full kernel matrix is:

- TTE: $O(N^2 g_0^2)$
- TTW: $O(N^2 g_0^3)$

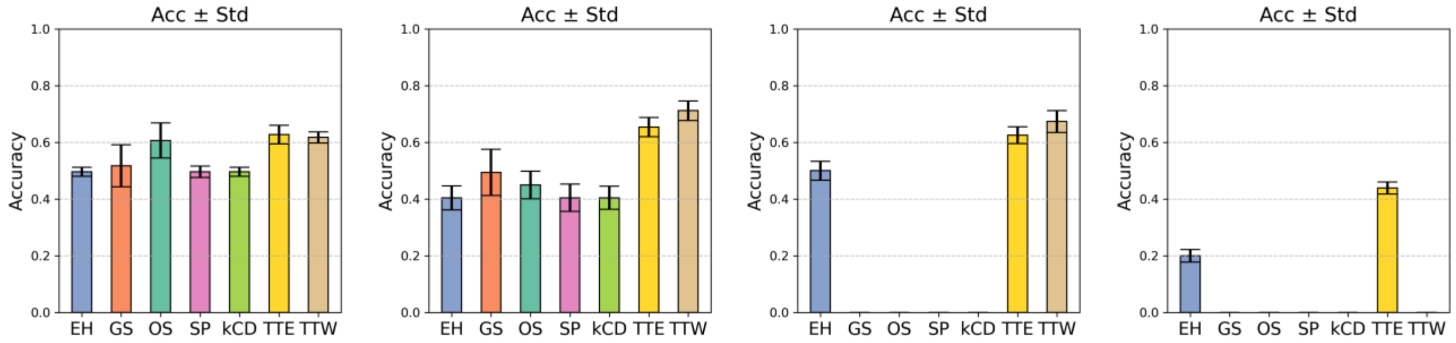
Experiments: Benchmarking

Benchmarked against 23 graph datasets and compared to 5 other graph kernels in classification tasks:

- Edge–Histogram kernel
- Graphlet sampling kernel
- ODD–STh kernel
- Shortest Path kernel
- k-Core Decomposition kernel



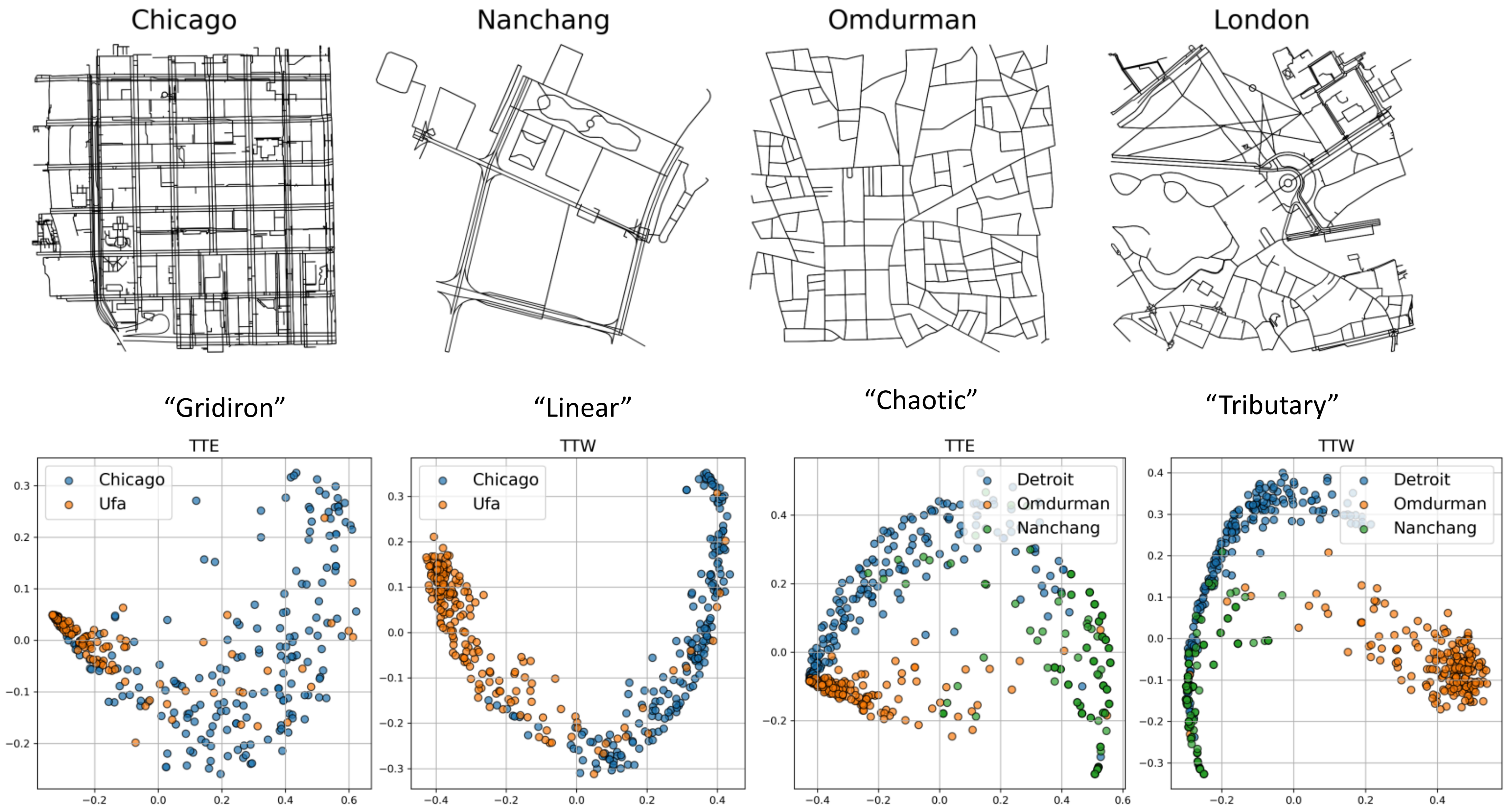
(a) AIDS (b) FRANKENSTEIN (c) IMDB-BINARY (d) MUTAG



(e) NCI109 (f) PROTEIN (g) REDDIT-BINARY (h) REDDIT-5K

Name	$\log(\bar{g}/\bar{n})$	Sparsity	GS	OS	SP	kCD	TTE	TTW
AIDS	-2.26	S	9.59	44.30	106.51	231.80	**2.53**	175.45
BZR	-2.29	S	8.57	6.49	3.63	11.81	**0.73**	9.70
MSRC-9	0.36	SS	111.21	5.99	2.78	12.86	**1.85**	31.72
MSRC-21	0.45	SS	1004.74	109.10	68.17	257.27	**25.97**	1497.16
BZR-MD	2.26	D	2328.25	0.97	**0.91**	63.16	19.21	1410.91
ER-MD	2.31	D	5503.19	**1.48**	2.78	119.88	31.80	5681.93

An Application to Urban Road Network Classification



URN Classification Results & Runtime

Name	$\bar{n}$ (*std.*)	$\bar{g}$ (*std.*)	OS Acc (*std.*)	OS Time	SP Acc (*std.*)	SP Time	TTE Acc (*std.*)	TTE Time	TTW Acc (*std.*)	TTW Time
`2-1S`	120.40 (*66.73*)	54.26 (*34.47*)	84.00 (*12.21*)	188.70	92.25 (*5.30*)	21.01	87.50 (*5.24*)	14.96	**93.50** (*4.90*)	114.63
`2-1M`	362.35 (*203.81*)	187.96 (*123.03*)	M	M	M	M	67.10 (*7.48*)	105.43	**89.75** (*3.61*)	1357.85
`2-2S`	96.40 (*77.20*)	38.84 (*36.89*)	90.50 (*1.98*)	252.25	**94.67** (*3.23*)	39.06	91.50 (*3.11*)	17.82	92.67 (*3.43*)	166.45
`2-2M`	289.60 (*239.37*)	123.00 (*121.43*)	M	M	M	M	92.67 (*3.82*)	119.28	**94.50** (*2.99*)	1553.34
`3-S`	55.02 (*51.09*)	22.31 (*24.06*)	62.50 (*15.69*)	99.02	84.01 (*3.43*)	13.27	87.67 (*4.03*)	7.25	**93.17** (*2.83*)	79.36
`3-M`	250.49 (*206.92*)	131.61 (*128.22*)	M	M	M	M	75.60 (*2.80*)	143.54	**84.13** (*2.63*)	2750.13
`4-S`	81.82 (*74.89*)	37.68 (*38.39*)	**84.38** (*3.80*)	365.67	83.00 (*4.58*)	67.85	81.12 (*4.12*)	22.07	81.00 (*4.36*)	287.45
`4-M`	283.36 (*238.34*)	133.03 (*125.64*)	M	M	M	M	68.12 (*3.80*)	155.03	**77.25** (*3.70*)	3187.99

Summary

- Using tropical geometry, we defined new kernels for metric graphs
- The kernels are based on the 1-homology group of a graph
- They are invariant to edge-subdivisions so we can compare graphs that represent different underlying spaces
- First connection between tropical and information geometry
- They are fast to compute
- They outperform other kernels in the absence of node/edge labels on both synthetic and real data

Limitations

- They don’t incorporate label or attribute information
- They don’t work on trees! (Maybe add an extra “sink” node to each graph that connects to the rest of the nodes?)

References

- Brannetti, S., Melo, M. and Viviani, F., 2011. On the tropical Torelli map. *Advances in Mathematics*, *226*(3), pp.2546-2586.
- Cao, Y. and AM, 2025. Computing the Tropical Abel–Jacobi Transform and Tropical Distances for Metric Graphs. *arXiv:2504.11619*.
- Chan, M., 2021. Moduli spaces of curves: classical and tropical. *Notices of the American Mathematical Society*, *68*(10), pp.1-1713.
- Miranda, R., 1995. *Algebraic curves and Riemann surfaces* (Vol. 5). American Mathematical Soc.
- Thanwerdas, Y. and Pennec, X., 2023. O(n)-invariant Riemannian metrics on SPD matrices. *Linear Algebra and its Applications*, *661*, pp.163-201.

IMPERIAL

Engineering and Physical Sciences Research Council

Y. Cao & AM (2025). Metric Graph Kernels via the Tropical Torelli Map, arXiv:2505.12129.

Thank You!

IMSI Chicago

Inés and Arne are here and you should check out their posters and talk to them:

Anthea Monod
18-22/8/2025