NetLSD: Hearing the Shape of a Graph

Anton Tsitsulin [1] Davide Mottin [1] Panagiotis Karras [2]
Alex Bronstein [3] Emmanuel Müller [1]

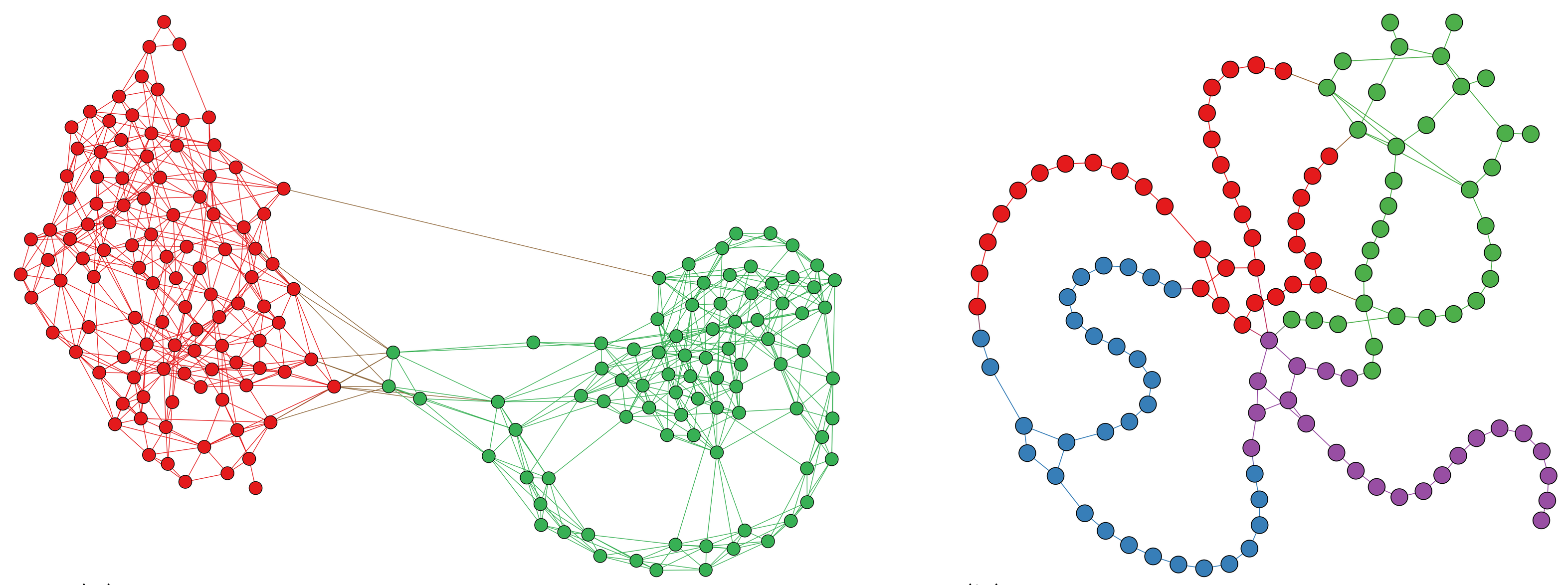
[1] Hasso Plattner Institute [2] Aarhus University [3] Technion

Expressive graph comparison is hard

Four key properties for expressive graph similarity:

- **Permutation invariance**: reordering nodes does not change the similarity
- **Scale-adaptivity**: structure is captured on **both** local and global scales
- **Size-invariance**: structure of the graph **may not** depend on its size
- **Scalability**: able to deal with **both** many and big graphs

Scalability is possible with a suitable **representation (descriptor)!** to **analyze** (e.g., classification, clustering) large graph collections.



(a) Protein network. $|V| = 190, |E| = 744, d = 9$. (b) Enzyme. $|V| = 125, |E| = 141, d = 32$.

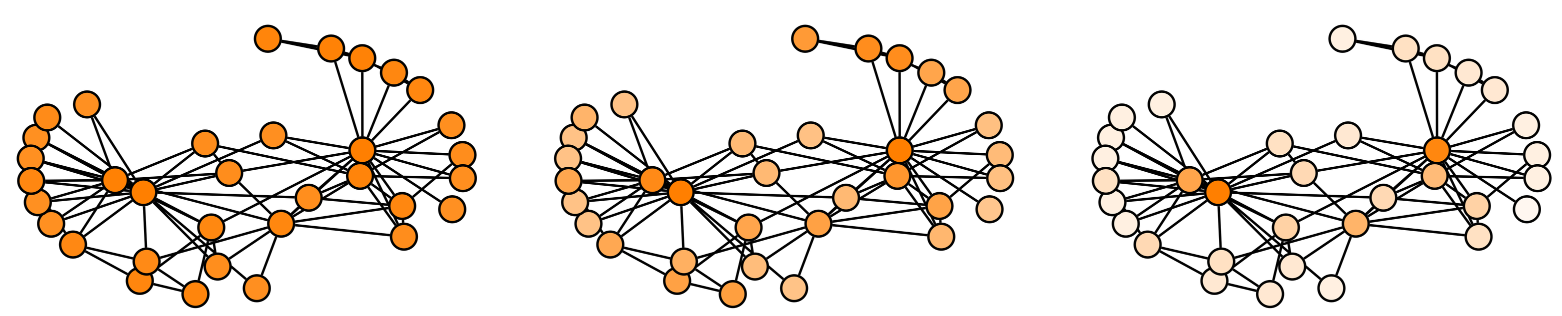
Figure 1: How can we compare these two graphs?

Graph structure at different scales

In different applications, graphs are analyzed at **different scales**:

- **Local interactions**, e. g. molecular bonds in computational biology
- **Medium-scale structure**, e. g. core-periphery in economic networks
- **Global connectivity**, e. g. community structure in social networks

We argue that **scale is a continuum**, as in Kronecker graphs.



(a) Small t (b) Medium t (c) Large t

Figure 2: Heat distribution (diagonal of H_t) at different scales on the Karate club graph.

Code & data

code	`github.com/xgfs/netlsd`
usage	`pip install netlsd`
contact	`anton.tsitsulin@hpi.de`

Rushing to dinner? Read this!

NetLSD is a graph descriptor that allows to compare graphs:

- **Fast:** in $\mathcal{O}(1)$, with $\mathcal{O}(m)$ precomputation;
- **On multiple scales:** capturing both local and global information;
- **Of different sizes:** it can (optionally) disregard the size of the graphs

We take a **geometric approach** to graphs. We start with the optimal transport of the **heat kernel** and adapt a powerful **lower bound** first introduced for manifolds.

We propose novel **evaluation tasks**, and show that our approach achieves **state-of-the-art** for classification.

Heat kernel to the rescue

Heat kernel can be defined in terms of the Laplacian matrix $\mathcal{L} = I - D^{-1/2} A D^{-1/2}$ that has eigenvalue decomposition $\mathcal{L} = \Phi \Lambda \Phi^\top$. Then, the heat equation is

$$\frac{\partial u_t}{\partial t} = -\mathcal{L} u_t,$$

Solution to the heat equation is given bythe **heat kernel** matrix:

$$H_t = e^{-t\mathcal{L}} = \Phi e^{-t\Lambda} \Phi^\top = \sum_{j=1}^{n} e^{-t\lambda_j} \phi_j \, \phi_j^\top,$$

Heat kernel matrix involves pairs of nodes, so we use its **trace**:

$$h_t = \mathrm{tr}(H_t) = \sum_j e^{-t\lambda_j},$$

where timescale t encodes an **explicit notion of scale**. We sample t logarithmically, and compare h_t with L_2 distance. h_t is a family of low-pass filters, we can also use a band-pass filter such as **wave kernel trace**:

$$w_t = \mathrm{tr}(W_t) = \sum_j e^{-it\lambda_j}$$

Million-node graphs? Not an issue anymore

Computing h_t requires the eigenvalues of a graph. Full eigendecomposition takes $\mathcal{O}(n^3)$: slow for large graphs. We can employ any spectrum estimation method, but we propose **two speedup techniques**:

- **Taylor expansion** for the matrix exponential, as first two terms can be computed in $\mathcal{O}(m)$, third can be computed with counting triangles
- **Spectrum interpolation** for the middle part of the spectrum, as we can compute lower and upper parts quickly. Our interpolation has geometric justification, the *Weyl`s law*.

We classify large-scale graph collections with up to a **million** nodes. NetLSD is the *first* method that allows expressive comparison of such graphs.

Theory: computational geometry

Definition. Mémoli [1] suggests a spectral definition of Gromov-Wasserstein distance between Riemannian manifolds. Matching a pair of points (x, x') on manifold $\mathcal{M}$ to a pair of points (y, y') on manifold $\mathcal{N}$ at scale t costs

$$\Gamma(x, y, x', y', t) = |H_t^{\mathcal{M}}(x, x') - H_t^{\mathcal{N}}(y, y')|.$$

The distance between $\mathcal{M}$ and $\mathcal{N}$ is defined as the infimal measure coupling

$$d(\mathcal{M}, \mathcal{N}) = \inf_{\mu} \sup_{t>0} e^{-2(t+t^{-1})} \|\Gamma\|_{L^2(\mu \times \mu)},$$

where the infimum is sought over all measures on $\mathcal{M} \times \mathcal{N}$ marginalizing to the standard measures on $\mathcal{M}$ and $\mathcal{N}$.

Theorem. [1] Spectral Gromov-Wasserstein distance is lower bounded by

$$d(\mathcal{M}, \mathcal{N}) \geq \sup_{t>0} e^{-2(t+t^{-1})} |h_t^{\mathcal{M}} - h_t^{\mathcal{N}}|.$$

We adopt this result *mutatis mutandis* to graphs, substituting the Laplace-Beltrami operator of the manifold with the normalized graph Laplacian.

Experiments

NetLSD is both versatile and expressive. Table 1 shows that **only** NetLSD captures nuances of graph community structure, while Table 2 shows that it captures **natural properties** of real graphs.

	$n \sim \mathcal{P}(\lambda)$				
Method	64	128	256	512	1024
$h(G)/h(\bar{K})$	54.53	62.27	70.83	**76.45**	**78.40**
$w(G)/w(K)$	55.51	**63.85**	**72.12**	**77.59**	**79.39**
FGSD	55.44	54.99	53.86	52.74	50.92
NetSimile	**59.55**	56.57	59.41	66.23	60.58

Table 1: Accuracy in detecting graphs with communities.

Method	MUTAG	PROTEINS	ENZYMES	COLLAB	IMDB-M
$h(G)$	**86.47**	64.89	31.99	68.00	40.51
$w(G)$	83.35	**66.80**	40.41	**75.77**	**42.66**
FGSD	84.90	65.30	**41.58**	67.37	39.71
NetSimile	84.09	62.45	33.23	73.96	41.14

Table 2: Accuracy of a 1-NN classifier.

References

[1] Facundo Mémoli.
A Spectral Notion of Gromov-Wasserstein Distance and Related Methods.
Applied and Computational Harmonic Analysis, 30(3):363--401, 2011.