Graph Filtration Kernels

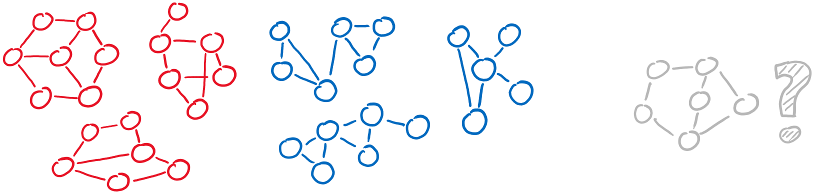
Till Schulz, Pascal Welke, Stefan Wrobel



UNIVERSITÄT BONN
Fraunhofer IAIS

Graph Classification

The task of graph classification is among the most common machine learning tasks:



One of the most successful graph classification methods rely on graph kernels.

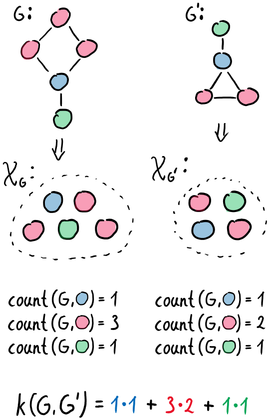
Traditional Kernels

Most traditional graph kernels define graph similarity by comparing counts of mutual features:

$$k(G, G') = \sum_{f \in \mathcal{F}} \mathrm{count}(G, f) \cdot \mathrm{count}(G', f)$$

with

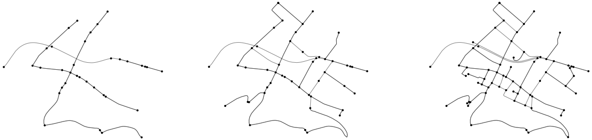
- feature domain $\mathcal{F}$, and
- $\mathrm{count}(G, f)$ denoting the frequency of f in G.



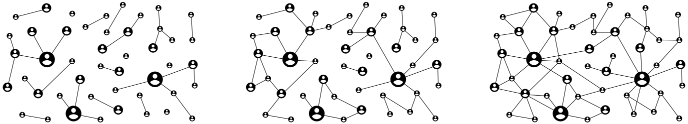
Graph at Different Scales

Often, graphs can be viewed at different resolutions:

- Street maps: Roads may be of different relevance.



- Social networks: Friendships may be of different significance.



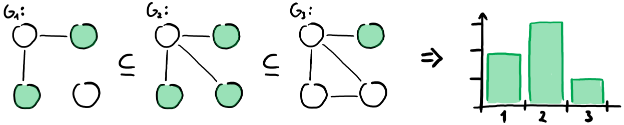
Graph Filtrations

Graph *filtrations* are nested subgraph sequences

$$G_1 \subseteq G_2 \subseteq \ldots \subseteq G_k = G$$

which view graph G at different resolutions.

This concept let's us generate *feature distribution histograms*:



UNIVERSITÄT BONN Fraunhofer IAIS

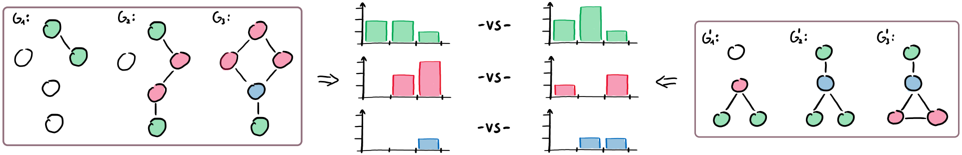
Graph Filtrations Kernels: Idea

Traditional graph kernels:

Compare feature counts.

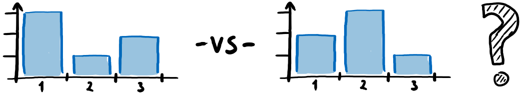
Graph filtration kernels:

Compare feature distributions over filtrations.



Filtration Histogram Distance

Q: How are such filtration histograms being compared?



A natural distance measure on distributions is the optimal transport distance.

- *Informally*: It is the minimum effort necessary to turn one histogram into another.

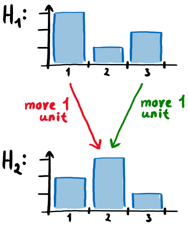
Filtration Histogram Distance: Ground Distance

Ground distance: Defines the cost for moving mass from one point to another.

Comparing filtration histograms requires a 1-dim. ground distance since we compare feature occurrences in a sequence:

$$d(\alpha_i, \alpha_j) = |\alpha_i - \alpha_j|$$

where values $\alpha_i \in \mathbb{R}_{\geq 0}$ are associated with each histogram index.



Costs:

$d(\alpha_1, \alpha_2) = |1-2| \times 1$

$d(\alpha_3, \alpha_2) = |3-2| \times 1$

Optimal transport distance:

$W_d(H_1, H_2) = |1-2| \times 1 + |3-2| \times 1 = 2$

The filtration histogram distance gives rise to proper base kernels:

$$K_{\circ}(G,G') = exp\left(-\gamma \, W_d\left(\, , \, \right)\right)$$

More generally:

$$\kappa_f(G,G') = exp(-\gamma\ \mathcal{W}_d(H_f(G), H_f(G')))$$

where $\gamma \geq 0$ and $H_f(G)$ is the filtration histogram w.r.t. feature f.

The kernel $\kappa_f(G,G')$ compares G and G' w.r.t. to a single feature f.

The Graph Filtration Kernel

The Graph Filtration Kernel is a linear combination of base kernels κ_f:

$$K_{\text{Filt}}^{\mathcal{F}} = \sum_{f \in \mathcal{F}} \beta\beta' \kappa_f(G, G')$$

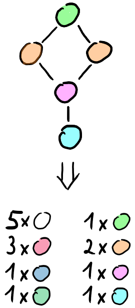
Details:

- Computing the optimal transport distance between histograms requires equal mass.
- Thus, a mass-normalization is necessary.
- This, however, removes frequency information.
- To "reverse" this, $\kappa_f(G, G')$ is weighted by the original histogram masses $\beta = ||H_f(G)||_1$ and $\beta' = ||H_f(G')||_1$.

Graph Filtration Kernels work for any kind of feature.

In the following, we consider a specific type of feature: Weisfeiler-Lehman labels.

- Iterative node relabeling by compressing each node's label and that of its neighbors.



Weisfeiler-Lehman Filtration Kernel: Complexity

The Weisfeiler-Lehman Filtration Kernel has linear complexity:

Theorem

The Weisfeiler-Lehman filtration kernel $K_{Filt}^{\mathcal{F}_{WL}}(G, G')$ *on graphs* G, G' *can be computed in time* $O(hkm)$*, where*

- h *is the number of performed iterations,*
- k *is the filtration length,*
- *and* m *denotes the number of edges.*

Weisfeiler-Lehman Filtration Kernel: Expressivity

Tracking Weisfeiler-Lehman labels over filtrations increases expressivity:

Theorem (simplified)

There exist filtrations such that the Weisfeiler-Lehman Filtration kernel is complete, i.e., it can distinguish all non-isomorphic graphs.

UNIVERSITÄT BONN

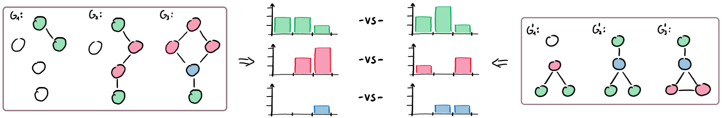
Fraunhofer
IAIS

Weisfeiler-Lehman Filtration Kernel: Evaluation

1. The Weisfeiler-Lehman Filtration Kernel outperforms state-of-the-art methods on several benchmark datasets.
2. Short filtrations are often sufficient, i.e. considering $G_1 \subseteq G_2 \subseteq \ldots \subseteq G_k$ for small values k leads to good predictive performances.
   - Kernel runtime complexity increases by only a small linear factor.
3. Experiments on synthetic datasets empirically confirm the theoretical results on the kernel expressivity.

Conclusion

- Graph Filtration Kernels compare graphs on different resolutions:



- We introduced a kernel instance: The Weisfeiler-Lehman Filtration Kernel
  - has linear complexity, and
  - yields complete kernels.

- The Weisfeiler-Lehman Filtration Kernel leads to significant performance increases for several real-world benchmark datasets.