Fast graph kernel with optical random features

Hashem Ghanem   Nicolas Keriven   Nicolas Tremblay
GIPSA-lab, CNRS, UGA, Grenoble INP, France

2366

1: Summary

The **graphlet kernel** [1] is a classical method in graph classification. It however suffers from a high computation cost due to the isomorphism test it includes.

We propose to leverage kernel random features [3] within the graphlet framework, and establish a theoretical link with the MMD metric. If this method can still be prohibitively costly for usual random features, we then incorporate optical random features [2] that can be computed in *constant time*.

3: Efficiency of $GSA-\varphi$ with kernel RF *w.r.t.* the MMD metric.

- $\kappa(\mathbf{x},\mathbf{x}') = \mathbb{E}_{\mathbf{w}\sim p}\ \xi_{\mathbf{w}}(\mathbf{x})^*\xi_{\mathbf{w}}(\mathbf{x}')$ (kernels with RFs).

$$\varphi_{RF}(\mathbf{x}) = \tfrac{1}{\sqrt{m}}(\xi_{\mathbf{w}_j}(\mathbf{x}))_{j=1}^m$$

- $\kappa(\mathbf{x},\mathbf{x}') \approx \varphi_{RF}(\mathbf{x})^*\varphi_{RF}(\mathbf{x}')$
- $\kappa_{GS}(\mathbf{x},\mathbf{x}') = \exp^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}} \Rightarrow \varphi_{GS}(\mathbf{x}) = \frac{\sqrt{2}}{\sqrt{m}}\cos(\mathbf{W}^T\mathbf{x}+b)$

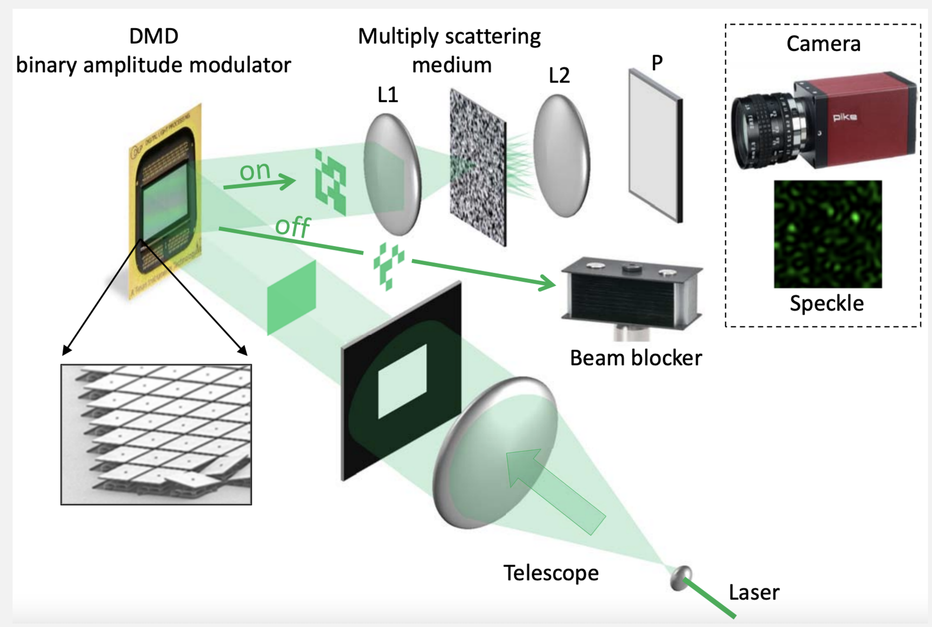
Theorem Let $\mathcal{G}$ and $\mathcal{G}'$ be two graphs. Assume that $|\xi_{\mathbf{w}}(F)| \leqslant 1$. Then, for all $\delta > 0$, with probability at least $1-\delta$:

$$\left|\|\mathbf{z}_{\mathcal{G}} - \mathbf{z}_{\mathcal{G}'}\|^2 - MMD(\mathbf{f}_{\mathcal{G}}, \mathbf{f}_{\mathcal{G}'})^2\right| \lesssim \sqrt{\log(1/\delta)}(m^{-1}+s^{-1})$$

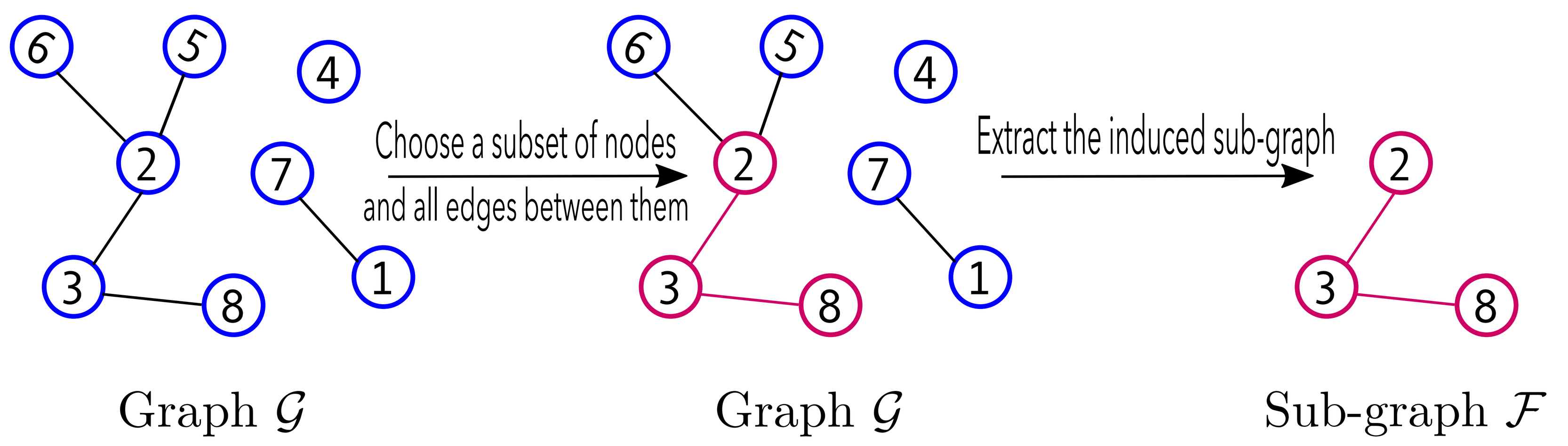
- MMD is a true metric between graphlet histograms
- $m \approx s$ **is still quite big for traditional RFs**

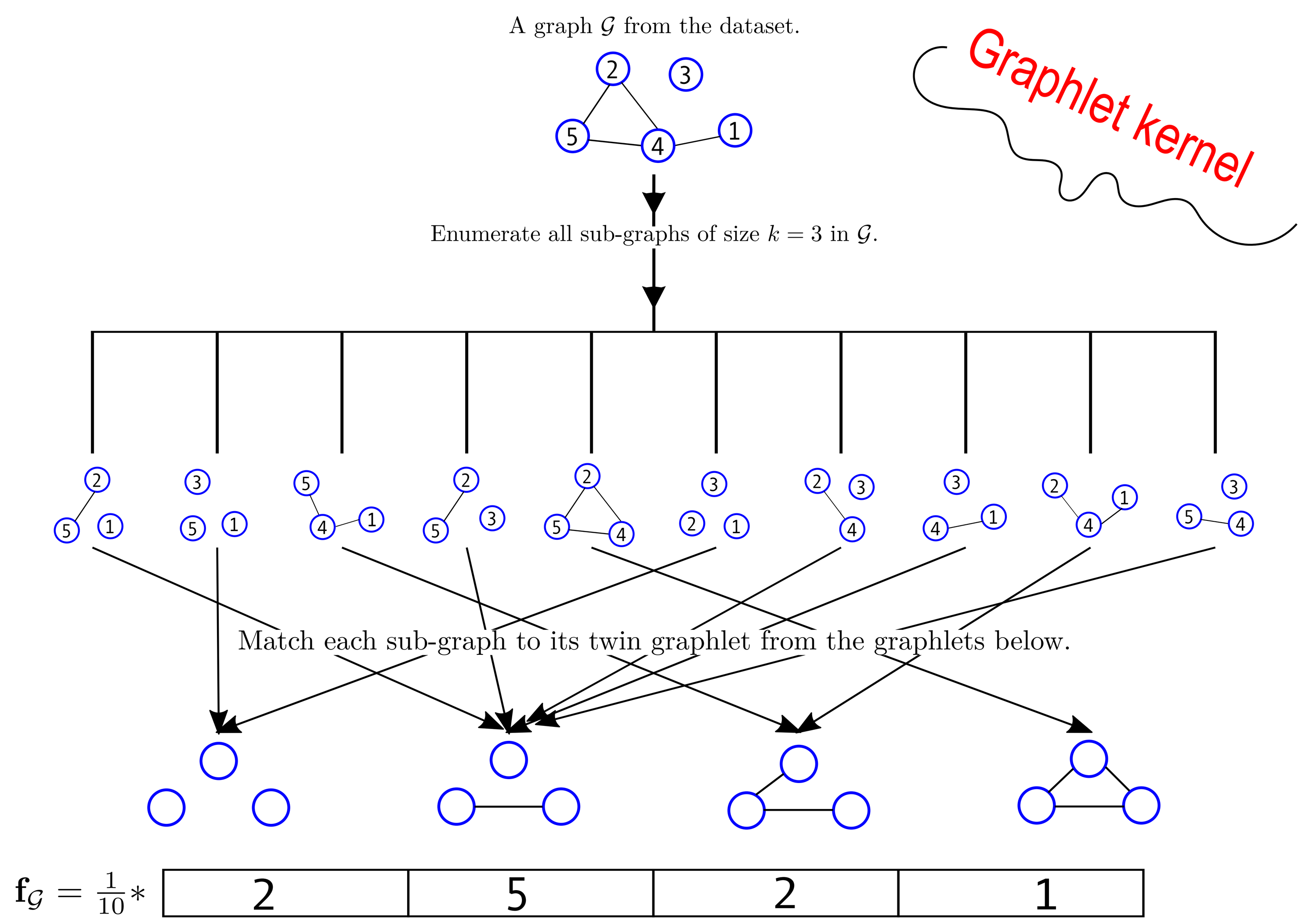
4: Incorporating Optical random features

- $\varphi_{OPU}(\mathbf{x}) = |\mathbf{W}\mathbf{x}+\mathbf{b}|^2;\ \mathbf{W}\in\mathbb{R}^{m\times d}, \mathbf{x}\in\mathbb{R}^d$
- $\varphi_{OPU}(\mathbf{x}_1)^T\varphi_{OPU}(\mathbf{x}_2) \approx \kappa_{OPU}(\mathbf{x}_1,\mathbf{x}_2)$
- Optical Processing Units (OPUs) evaluate φ_{OPU} in $\mathcal{O}(1)$ in both input/output dimensions (d, m).



2: From the graphlet kernel to $GSA-\varphi$

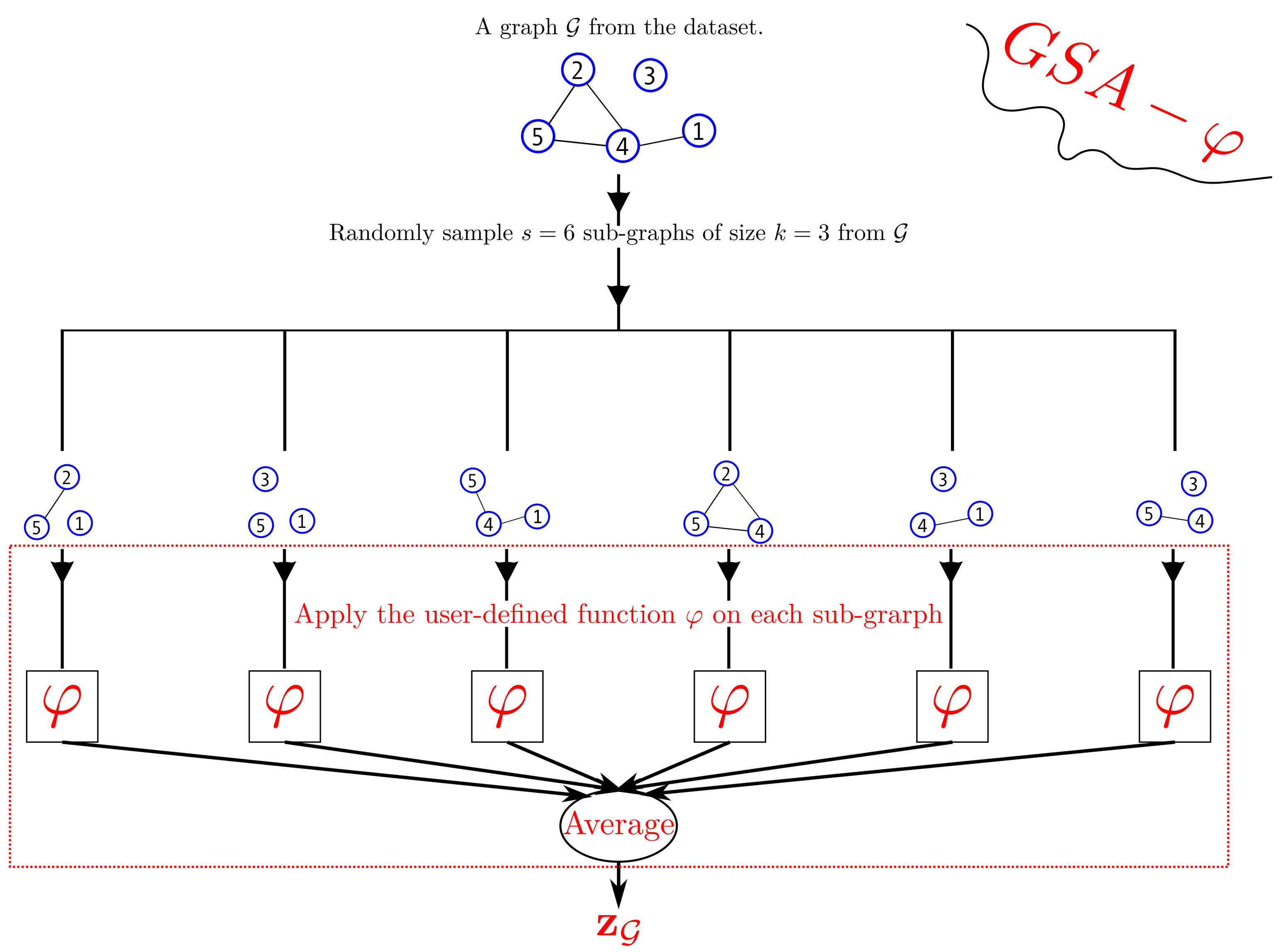




The prob. mass function $\mathbf{f}_{\mathcal{G}}$ is the representation vector for $\mathcal{G}$.

- Exp. cost: $O(\binom{v}{k}N_kC_k^{\cong})$
  - $N_kC_k^{\cong}$: cost of graphlet matching, doubly exponential in k
- Sampling: $O(sN_kC_k^{\cong})$
  - s: number of randomly sampled subgraphs
  - can use different samplers: uniform, random walk...
- Still exponential.

- Replace *graphlet matching* with a generic map φ : {size-k subgraphs} $\mapsto \mathbb{R}^m$



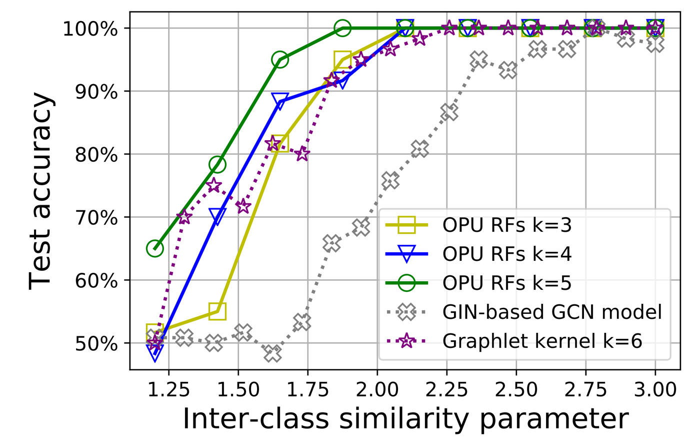
$\mathbf{z}_{\mathcal{G}}$ represents $\mathcal{G}$.

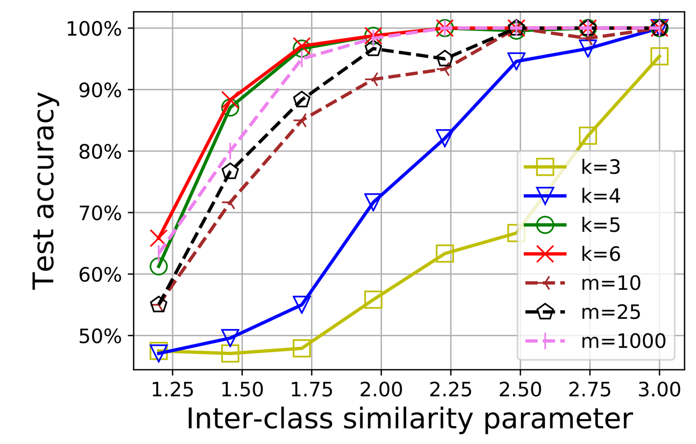
- **Gen. cost:** $O(sC_\varphi)$
- Graphlet: $C_\varphi = N_kC_k^{\cong}$
- Kernel RF [3]: $C_\varphi = mk^2$
- Optical RF [2]: $C_\varphi = 1$

5: Experiments

$GSA-\varphi_{OPU}$ Vs. graphlet kernel and GNNs

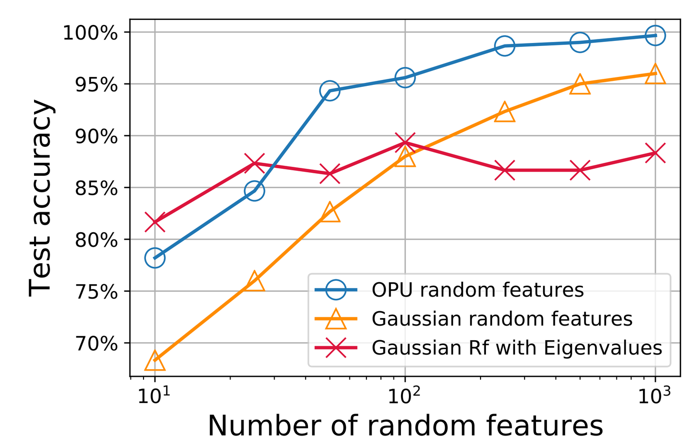
- Dataset: 300 graphs based on the stochastic block model.
- Lft: $GSA-\varphi_{OPU}$ with uniform sampling.
- Rgt: $GSA-\varphi_{OPU}$ with random walks, graphlet kernel, and GIN model.
- If not mentioned: $s=2000$, $m=5000$.

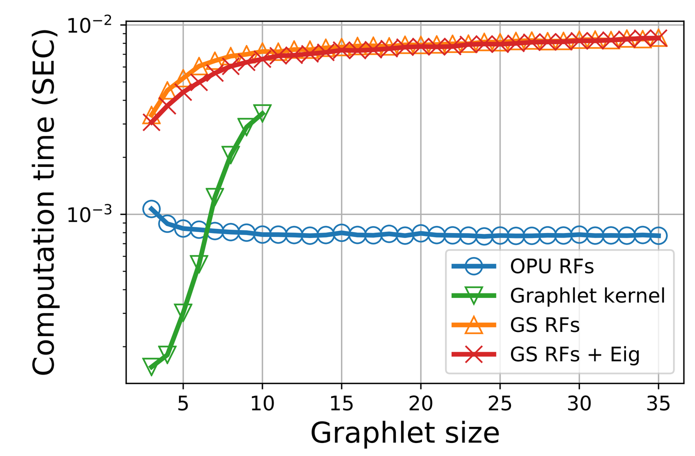




$GSA-\varphi$ with different φ_{RF} + comp. cost

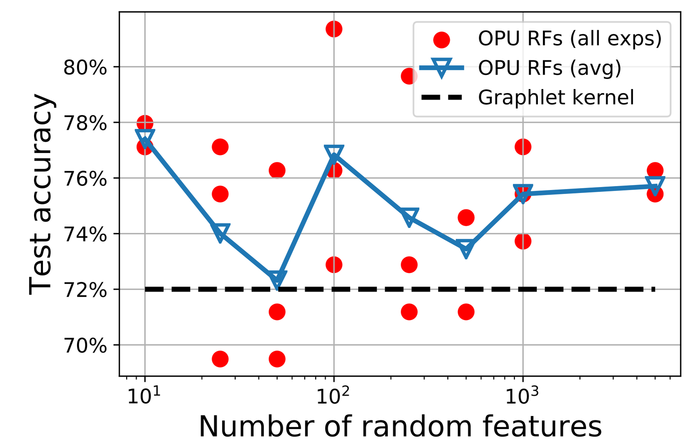
- $s=2000$, $m=5000$.

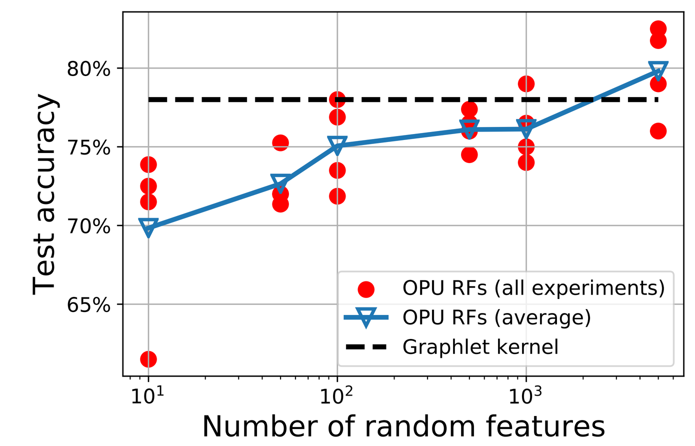




Results on real world datasets.

- Lft: D&D dataset, rgt: Reddit-Binary. ($s=2000$, $m=5000$).





[1] Saade et al. **Random projections through multiple optical scattering: Approximating kernels at the speed of light**. *ICASSP*, 2016.
[2] Shervashidze et al. **Efficient graphlet kernels for large graph comparison**. *International Conference on Artificial Intelligence and Statistics*, 2009.
[3] Rahimi et al. **Random features for large-scale kernel machines**. *NIPS*, 2007.

hashem.ghanem@grenoble-inp.fr, nicolas.keriven@grenoble-inp.fr, nicolas.tremblay@grenoble-inp.fr

This work was supported by LabEx PERSYVAL (ANR-11-LABX-0025-01), ANR GraVa (ANR-18-CE40-0005), (ANR-15-IDEX-02), and (ANR-19-P3IA-0003).