Fast Graph Kernel with Optical Random Features

Hashem GHANEM

Nicolas KERIVEN   Nicolas TREMBLAY

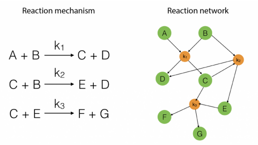
IEEE ICASSP 2021

June 06/2021

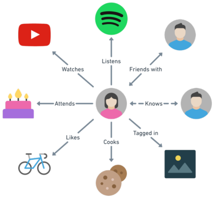
Outline

Graphs

- Model: (objects→ nodes) , (relations → edges).

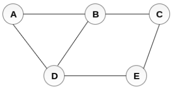


(a) Chemical reactions



(b) Social networks

- Adjacency matrix to represent a graph of v nodes:
  - $\mathbf{A} \in \mathbb{R}^{v \times v}$: $a_{i,j} = 1$ if nodes (i, j) have edge, 0 otherwise.



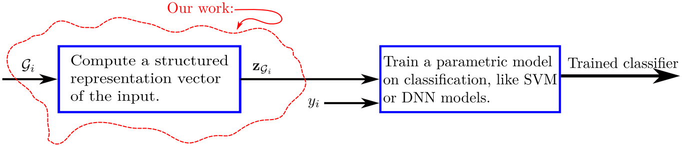
Undirected Graph

	A	B	C	D	E
A	0	1	0	1	0
B	1	0	1	1	0
C	0	1	0	0	1
D	1	1	0	0	1
E	0	0	1	1	0

Adjacency Matrix

Graph classification & applications

- Supervised classification:
  - Pre-labeled dataset: $(\{\mathcal{G}_1, \ldots, \mathcal{G}_n\}, \{y_1, \ldots, y_n\})$.
  - Each graph $\mathcal{G}_i$ belongs to class y_i.
  - Task: a classification algorithm that, given in input a new graph, output the class to which it belongs.
- Graphs have different sizes (#nodes), so a classifier has 2 blocks:



- Applications in real world: Marketing, Banking, Biology.
- Example in biology:
  - amino acids are linked to form a protein.
  - enzymes are specific type of proteins
  - predict whether a protein is an enzyme or not.

State-of-the-art methods

- Graph kernels based algorithms:
  - Fixed graph representation is computed.
  - **Graphlet kernel** is based on counting subgraphs.

Our Contribution: inspired by the graphlet kernel, we propose a fast and efficient graph classification framework, which leverages optical random features.

Outline

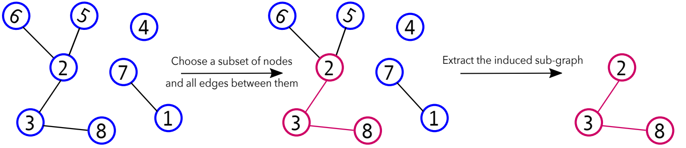
Necessary tools and definitions

- Graphlet kernel needs an integer parameter k to be fixed.

- Also, the set of all different graphs of size k.
    - these graphs are called *graphlets*.

Graphlets for $k = 3$.

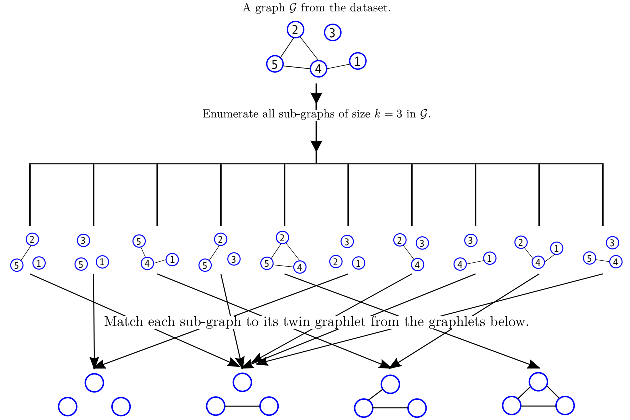
- Definition: to get a sub-graph $\mathcal{F}$ from a graph $\mathcal{G}$:



Graph $\mathcal{G}$ Graph $\mathcal{G}$ Sub-graph $\mathcal{F}$

Graphlet kernel illustration

Goal: compute representation vectors for graphs, for example:

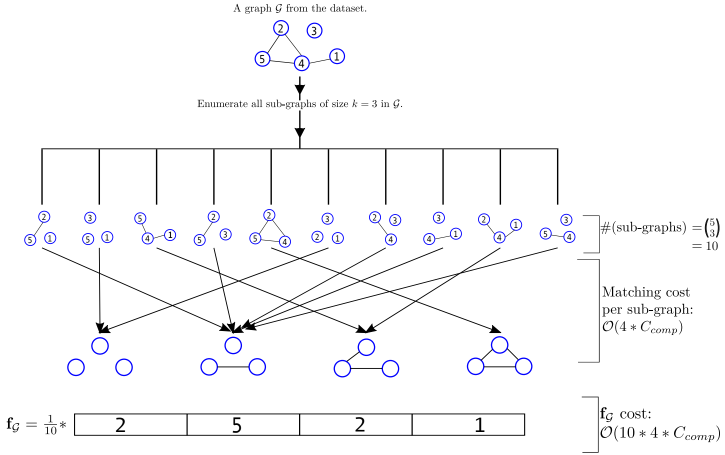


$\mathbf{f}_{\mathcal{G}} = \frac{1}{10} *$

2	5	2	1

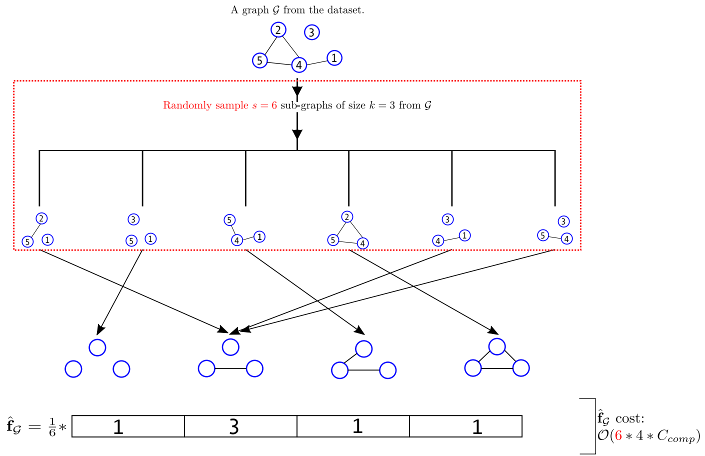
The probability mass function $\mathbf{f}_{\mathcal{G}}$ is the representation vector of $\mathcal{G}$.

Graphlet kernel **cost**



$\mathbf{f}_{\mathcal{G}}$ cost exponential in k: $C_{gk} = \mathcal{O}(\#subgraphs * \#graphlets * C_{comp})$

Acceleration: estimate $\mathbf{f}_{\mathcal{G}}$ with s subgraphs



$\hat{\mathbf{f}}_{\mathcal{G}}$ cost is lower but still exponential in k:

$$C_{gk+gs} = \mathcal{O}(s * \#graphlets * C_{comp})$$

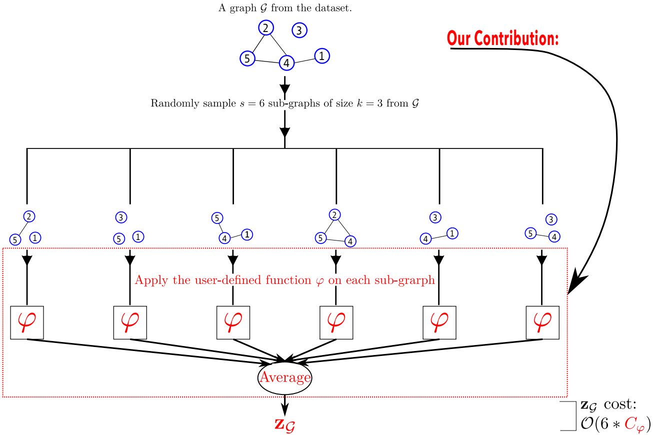
$\Rightarrow$ must deal with the matching stage for a lower cost.

Outline

Proposed classification framework $GSA - \varphi$

Replace subgraph-to-graphlet matching with a fast user-defined map:

$$\varphi : \{\text{size-}k \text{ subgraphs}\} \mapsto \mathbb{R}^{m} \quad , m \text{ is user-chosen.}$$



$\mathbf{z}_{\mathcal{G}}$ is the representation vector of $\mathcal{G}$, with cost: $C_{GSA-\varphi} = \mathcal{O}(s * C_{\varphi})$.

$GSA - \varphi$ with kernel random features

- Kernels with random features are functions that:
$$\kappa(\mathbf{x}, \mathbf{x}') = \mathbb{E}_{\mathbf{w} \sim p} \; \xi_{\mathbf{w}}(\mathbf{x})^* \xi_{\mathbf{w}}(\mathbf{x}')$$

- Defining : $\varphi_{RF}(\mathbf{x}) = \frac{1}{\sqrt{m}}(\xi_{\mathbf{w}_j}(\mathbf{x}))_{j=1}^m$
we can write: $\kappa(\mathbf{x}, \mathbf{x}') \approx \varphi_{RF}(\mathbf{x})^* \varphi_{RF}(\mathbf{x}')$

- Example: Gaussian kernel $\kappa_G(\mathbf{x}, \mathbf{x}') = \exp^{-\frac{\left\|\mathbf{x}-\mathbf{x}'\right\|^2}{2\sigma^2}}$ corresponds to:
$$\varphi(\mathbf{x}) = \frac{\sqrt{2}}{\sqrt{m}} cos(\mathbf{W}^T \mathbf{x} + b), \;\; \mathbf{x} \in \mathbb{R}^d, \mathbf{W} \in \mathbb{R}^{m \times d}$$

- $\mathbf{x}$ is a subgraph adjacency matrix $\Rightarrow \varphi_{RF}(\mathbf{x})$ costs: $\mathcal{O}(mk^2)$

- Computation cost of $GSA - \varphi_{RF} = \mathcal{O}(smk^2)$

Concentration analysis

- Recall: ($\mathcal{G}+$ subgraph sampling) $\rightarrow \mathbf{f}_\mathcal{G}$: discrete prob. dist. of graphlets.
- Recall: $\mathbf{z}_\mathcal{G}$ is the graph representation of $GSA - \varphi$.
- The Euclidean metric $\|\mathbf{z}_\mathcal{G} - \mathbf{z}_{\mathcal{G}'}\|^2$ converges to the **MMD metric**: $MMD(\mathbf{f}_\mathcal{G}, \mathbf{f}_{\mathcal{G}'})^2$
- MMD is a true metric on distributions for many schemes of kernel random features

Theorem

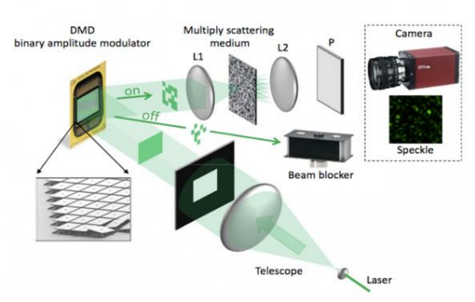
Let $\mathcal{G}$ and $\mathcal{G}'$ be two graphs. Assume that $|\xi_\mathbf{w}(F)| \leq 1$. Then, for all $\delta > 0$, with probability at least $1 - \delta$:

$$\left| \|\mathbf{z}_\mathcal{G} - \mathbf{z}_{\mathcal{G}'}\|^2 - MMD(\mathbf{f}_\mathcal{G}, \mathbf{f}_{\mathcal{G}'})^2 \right| \leq$$

$$\frac{4\sqrt{\log(6/\delta)}}{\sqrt{m}} + \frac{8\left(1 + \sqrt{2\log(3/\delta)}\right)}{\sqrt{s}}$$

$GSA - \varphi$ with optical random features

- Model: $\varphi_{OPU}(\mathbf{x}) = |\mathbf{W}\mathbf{x} + \mathbf{b}|^2$; $\mathbf{W} \in \mathbb{R}^{m \times d}, \mathbf{b} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^d$
- When $m \mapsto \infty$ then: $\varphi_{OPU}(\mathbf{x_1})^T \varphi_{OPU}(\mathbf{x_2}) \approx \kappa_{OPU}(\mathbf{x_1}, \mathbf{x_2})$



OPU's Experimental setup [Saade et al., 2016].

- $C_{\varphi_{OPU}} = \mathcal{O}(1)$ in the input and output dimensions.
- The computation cost $C_{GSA-\varphi_{OPU}} = \mathcal{O}(s)$

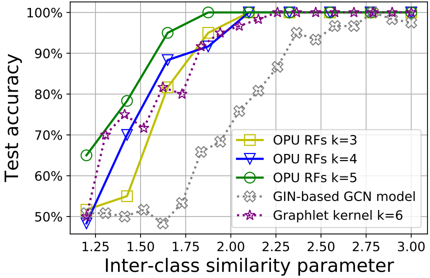
Comparison between computation costs

Graphlet kernel		$O\big(s * (\#graphlets) * C_{comp}\big)$
GSA-φ with:	φ_{Gs}	$O(smk^2)$
	φ_{Gs+eig}	$O(s(mk + k^3))$
	φ_{OPU}	$O(s)$

Outline

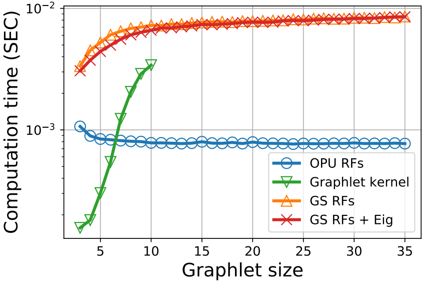
$GSA - \varphi_{OPU}$ Vs. graphlet kernel and GCN models

- Dataset:
  - 300 2-classes labeled graphs.
  - generated based on SBM model.
  - inter-class similarity r controls the problem difficulty.



Fixed parameters: $s = 2000$, $m = 5000$

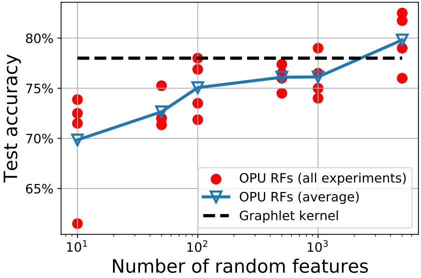
k-graphlet Kernel for $k = 3$



Fixed parameters: $r = 1.1$, $s = 2000$, $m = 5000$ and $\sigma = 0.1$.

$GSA - \varphi_{OPU}$ on the Reddit dataset

- $k = 7$, $s = 4000$.
- For every m, the experiment is repeated 4 times (red dots), then averaged (blue curve).



Conclusion and future work

Conclusion:

1. GSA-φ with RF maps separates well between graphs.
2. GSA-φ with optical RF maps is faster than traditional graphlet kernel.
3. GSA-φ with optical RF maps performs better than the graphlet kernel, and better than a particular graph convolutional networks on graph classification.

Future work:

1. Combine our algorithm with GCNs when we have node features.
2. Further analysis of the MMD metric properties on particular graph models.

References I

Hashem Ghanem, Nicolas Keriven, and Nicolas Tremblay. Fast graph kernel with optical random features. *ICASSP*, 2021.

Alaa Saade, Francesco Caltagirone, Igor Carron, Laurent Daudet, Angélique Drémeau, Sylvain Gigan, and Florent Krzakala. Random projections through multiple optical scattering: Approximating kernels at the speed of light. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6215–6219. IEEE, 2016.

Nino Shervashidze, SVN Vishwanathan, Tobias Petri, Kurt Mehlhorn, and Karsten Borgwardt. Efficient graphlet kernels for large graph comparison. In *Artificial Intelligence and Statistics*, pages 488–495, 2009.

Johannes Kobler, Uwe Schöning, and Jacobo Torán. *The graph isomorphism problem: its structural complexity*. Springer Science & Business Media, 2012.

Aman Sinha and John C Duchi. Learning kernels with random features. In *Advances in Neural Information Processing Systems*, pages 1298–1306, 2016.

References II

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=ryGs6iA5Km`.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in neural information processing systems*, pages 1177–1184, 2008.