

# Pattern Mining for Anomaly Detection in Graphs: Application to Fraud in Public Procurement

Lucas Potin<sup>1</sup>, Rosa Figueiredo<sup>1</sup>, Vincent Labatut<sup>1</sup>, Christine Largeron<sup>2</sup>  
1 : LIA – Laboratoire Informatique d'Avignon / 2 : LabHC – Laboratoire Hubert Curien

## Context

DeCoMaP [1] project – Fraud Detection in Public Procurement

1. **Collection** and **analysis** of data related to French public procurement → done in [2]
2. Develop automatic methods leveraging relational information for **fraud** detection → present work [3]

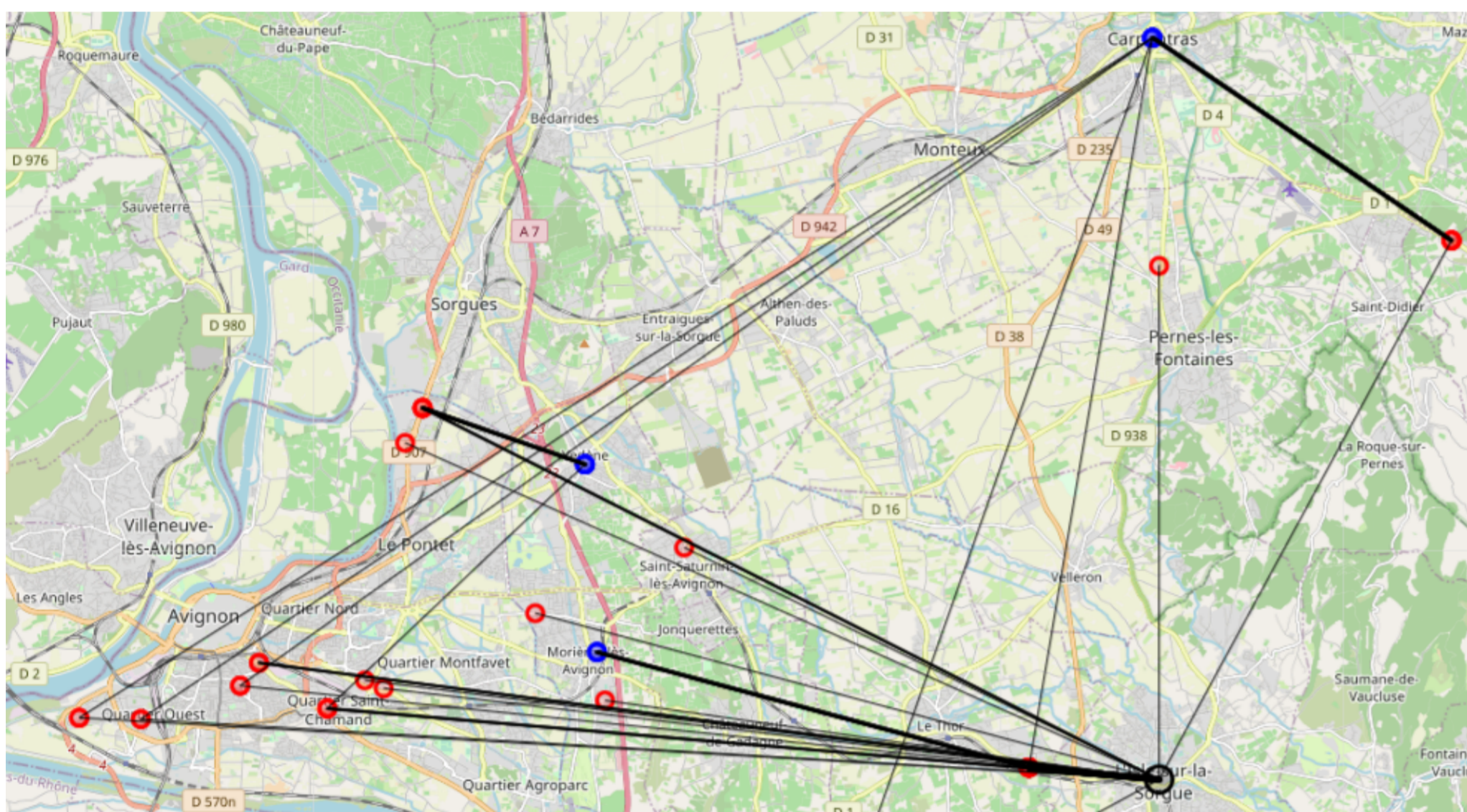
## Public Procurement

14% of European Union GDP in 2022

Raw data available at **European** level in the **Tenders Electronic Daily (TED)**

Use of standard indicators (**red flags**) on contracts attributes to determine bad practices

**Publicly** available database containing consolidated French TED data, including **identification numbers** [2]



Example of a graph (main municipality in black, other municipalities in blue, buyers in red)

## Graph Extraction

Construction of a **set** of graphs

Graphs refers to contracts in the same **activity** domain and **geographic** location

One graph per **municipality**

Nodes are **agents**, edges are **contracts**

Each graph is labeled according to the **number of red flags** among its contracts

Methods	MUTAG	NCI1	D&D	PTC	FOPPA
PANG	<b>0.88 (0.04)</b>	0.79 (0.02)	0.77 (0.03)	0.60 (0.13)	<b>0.95 (0.01)</b>
CORK	0.66 (0.08)	0.78 (0.02)	0.73 (0.03)	0.54 (0.06)	0.63 (0.05)
WL	0.86 (0.06)	<b>0.83 (0.01)</b>	<b>0.82 (0.01)</b>	0.67 (0.06)	0.90 (0.05)
WL_OA	0.86 (0.06)	0.81 (0.03)	0.77 (0.03)	0.55 (0.11)	0.90 (0.05)
Graph2Vec	0.84 (0.07)	0.82 (0.01)	0.72 (0.03)	<b>0.61 (0.11)</b>	0.91 (0.04)
DGCNN	0.86 (0.04)	0.74 (0.01)	0.79 (0.01)	0.58 (0.05)	0.89 (0.01)

F-Score ( $\pm$  sd) for the Anomalous class

Representation Size $s$	Anomalous Class	Normal Class
10	0.66 (0.05)	0.73 (0.05)
50	0.74 (0.05)	0.77 (0.04)
100	0.81 (0.05)	0.83 (0.04)
150	0.88 (0.03)	0.88 (0.03)
(all) 15,793	<b>0.93 (0.02)</b>	<b>0.93 (0.02)</b>

F-Score ( $\pm$  sd) depending on parameter  $s$ , the size of  $\mathcal{P}_s$ , on FOPPA with PANG

## References

- [1] DeCoMaP website.  
<https://decomap.univ-avignon.fr/>.
- [2] L. Potin, V. Labatut, C. Largeron, and P. H. Morand.  
FOPPA: an open database of french public procurement award notices from 2010–2020.  
*Scientific Data*, 10:303, 2023.
- [3] L. Potin, R. Figueiredo, V. Labatut, and C. Largeron.  
Pattern mining for anomaly detection in graphs: Application to fraud in public procurement.  
*In European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2023*, 2023.

## Proposed Method

Consider the task as a graph classification problem

1. Represent set of public procurement notices as attributed graphs
2. Derive pattern-based representation of graphs
3. Apply standard classifier

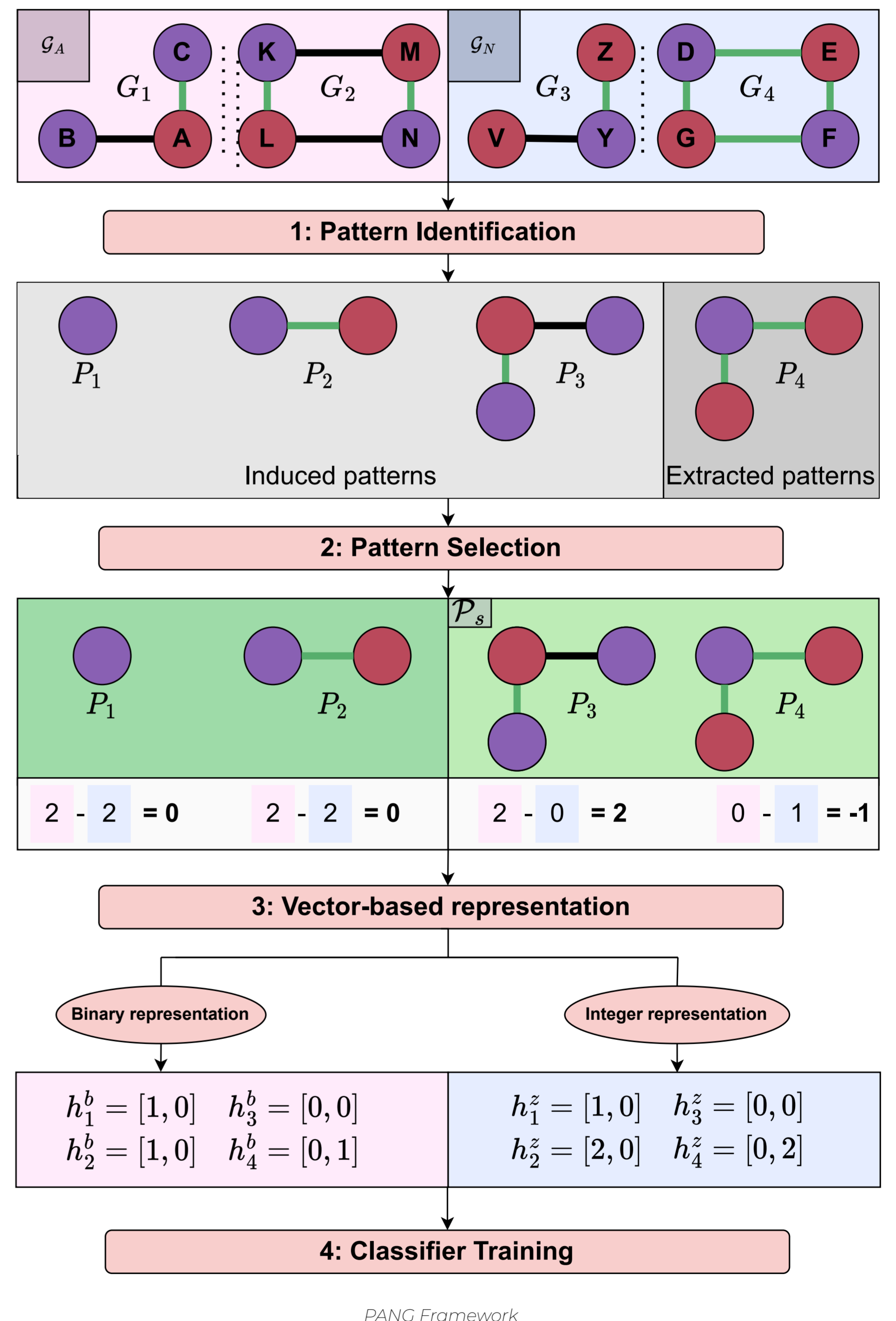
## PANG Framework

Description of graphs in terms of bags-of-subgraphs, called **patterns**.

Computation of a **discriminant** score for patterns indicating presence in one class compared with another

Representation of each graph as a **boolean** or **integer** vector using the most discriminant patterns.

Vector classification using a standard **classifier**



## Perspectives

- Improvement of the **discriminant** score
- Construction of a protocol to identify the **optimal** parameters of the PANG framework according to the **dataset**



ANR-19-CE38-0004



CV HAL



LinkedIn



Lucas Potin

3rd Year PhD Student at Avignon University  
lucas.potin@univ-avignon.fr