

# Red Flagging et détection automatique de la corruption dans les marchés publics : que détecte-t-on ?

## Regards croisés sur la corruption et le favoritisme

Pierre-Henri Morand

Avignon Université

19 janvier 2024



## Un consortium de recherche

- .JPEG - A.U. (+ CRED, P.II, CRESE, U.F.C.)
- CRA - A.M.U.
- LIA - A.U. ( + IPHC, Saint-Etienne)
- Dataactivist

## Une approche méthodologique

- utiliser les données ouvertes
  - Open data de la commande publique
  - Open data des décisions de justice
  - RNE, RBE
- exploiter l'aspect relationnel des données
  - approches classiques (économétrie)
  - détection d'anomalie dans des graphes

# Ce que l'on pourrait faire

si tout allait bien...

- 1 identifier tous les cas de fraude jugés (vérité terrain)
- 2 les "tagger" dans le graphe des marchés publics
- 3 identifier leurs caractéristiques spécifiques (leur "motif" spécifique)
- 4 détecter des cas similaires dans l'ensemble des données passées et actuelles

# Les cas suspicieux laissent-ils des traces ?

nb : souvent, pb = pas de marchés !



## Guillaume Peltier : dons et services rendus, plongée dans ses micropartis

Le député et numéro deux du parti Les Républicains a créé plusieurs mouvements, dont les plus gros donateurs sont des entrepreneurs implantés dans sa circonscription, et auxquels il a parfois rendu des services. Un cas qui illustre l'ambiguïté de ces structures prisées par nombre d'élus et permettant de recevoir des dons anonymes et détaxés.

L'élus et l'homme d'affaires se rencontrent rapidement après l'élection de M. Peltier à Neung-sur-Beuvron, en 2014. Ils sympathisent. Selon nos informations, M. Roussel devient rapidement un donateur de l'élus. En 2017, Neung-sur-Beuvron paye 182 000 euros à Girard-Sudron pour réaliser une étude « technico-économique » sur le remplacement de l'éclairage public du village par des luminaires utilisant la technologie LED. Dès le début de l'année 2017, M. Peltier a



## Garde à vue d'Éric Piolle et droit des marchés publics, que comprendre ?

**OPINION.** Selon la Chambre régionale des comptes (CRC) Auvergne-Rhône-Alpes, les éléments pour dégraver aux obligations de mise en concurrence n'ont pas été suffisamment prouvés par l'exécutif grenoblois. Par Jean-Baptiste Vila, Université de Bordeaux (\*)



# Ces traces apparaissent-elles dans les données ?

Avis No 13-32047    publié le 22/02/2013    **Date limite de réponse : 27/02/2013**

Dépt. : 94    Institut national audiovisuel - ina.

## Prestation de conseil pour le compte de l'INA à Bry-Sur-Marne

Avis de marché    Procédure NC

*Attribution du marché ou du lot :*  
- lot 1: Lot unique.

*Nom du titulaire/organisme :* Roland Berger Strategy Consultants, 11 rue de Prony 75017 Paris.

*Date d'attribution du marché :* 26 avril 2013.

*Nombres d'offres reçues :* 5.

*Instance chargée des procédures de recours :* Tribunal de grande instance de paris 4 boulevard du palais 7501 Paris tél. : 01-44-32-51-51.

Avis No 10-259855    publié le 04/12/2010

Dépt. : 94    Institut national audiovisuel - ina

## Prestation de conseil en développement stratégique pour l'Institut National Audiovisuel de Bry-sur-Marne

Résultat de marché    Procédure NC

*V.3) Nom et adresse de l'opérateur économique en faveur duquel une décision d'attribution du marché a été prise :*

Roland Berger Strategy Consultants, 11 rue du Pony, 75017 Paris.

*V.4) Informations sur le montant du marché :*

*Estimation initiale du montant du marché :*

Valeur : 240 000 EUR.

## Marseille : cinq ans de prison pour le haut fonctionnaire corrompu

Renaud Chervet touchait des pots-de-vin contre l'attribution de marchés publics du Département

Par Laurent D'ANCONA



# Y'a-t-il des signaux d'alerte ?

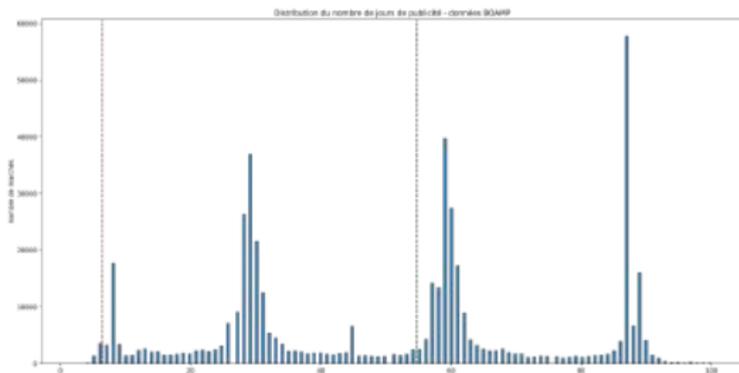


Figure: Publicity FR

publicityDuration	
count	2179.0
mean	46.0
std	11.0
min	15.0
10%	32.0
20%	36.0
30%	42.0
40%	43.0
50%	45.0
60%	49.0
70%	51.0
80%	54.0
90%	59.0
max	101.0

Figure: publicity CD13

# Y'a-t-ils des signaux d'alerte ?

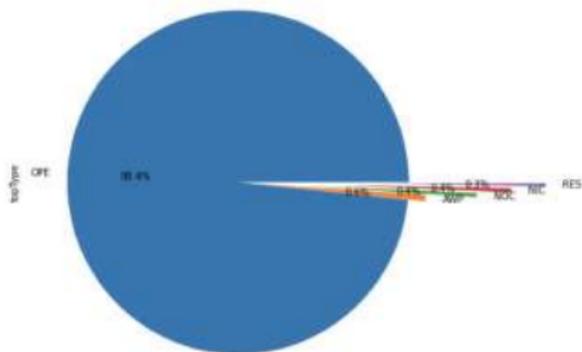


Figure: Type of procedure / CD13

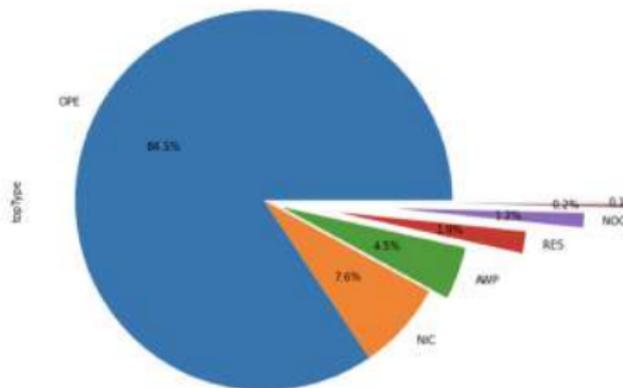


Figure: Type of procedure / FR

# Y'a-t-ils des signaux d'alerte ?

Price_weight	
<b>count</b>	2288.0
<b>mean</b>	52.0
<b>std</b>	19.0
<b>min</b>	10.0
<b>25%</b>	40.0
<b>50%</b>	50.0
<b>75%</b>	60.0
<b>max</b>	100.0

Figure: Obj. criteria / CD13

Price_weight	
<b>count</b>	706632.0
<b>mean</b>	47.0
<b>std</b>	17.0
<b>min</b>	0.0
<b>25%</b>	40.0
<b>50%</b>	40.0
<b>75%</b>	60.0
<b>max</b>	100.0

Figure: Obj. criteria / FR

# Y'a-t-ils des signaux d'alerte ?

WIN_share	
count	2322.00
mean	26.93
std	29.63
min	0.59
10%	1.69
20%	3.85
30%	6.67
40%	10.00
50%	14.29
60%	23.08
70%	33.33
80%	50.00
90%	83.33
max	100.00

Figure: Concentration / CD13

WIN_share	
count	1013410.00
mean	29.05
std	31.78
min	0.03
10%	1.59
20%	3.75
30%	6.58
40%	10.32
50%	16.56
60%	25.00
70%	33.33
80%	50.00
90%	100.00
max	100.00

Figure: Concentration / FR

# Sur quoi peuvent porter les signaux d'alerte ?

compte tenu des données accessibles

- des éléments de procédure
  - durée / mode de publicité
  - critères / poids des critères
  - choix de la procédure
  - ...
- le niveau de concurrence obtenu (nb d'offres)
- des prix anormaux
- la faible diversité des entreprises sélectionnées
- la présence de liens (familiaux / politiques etc...)

# A la recherche de la vérité terrain

Quelle variable expliquée ?

- cas pénaux
- cas administratifs (TA, TAE, CE, CRC etc...)
- proxy (local ? taux départemental d'atteinte à la probité) ?

## BDD décisions de justice administrative

- 1 2008-2021
- 2 T.A., C.A.A., C.E.
- 3 435 entries

Manquement à l'obl d'écarter une candidature irrecevable, une offre irrégulière, inacceptable ou inappropriée	81
Manquement à l'obl d'information des candidats	71
Imprécision, erreur ou irrégularité des documents de la consultation ou du marché	38
Avantage d'un candidat ou manquement au principe d'égalité	31
Marché/ DSP passé sans publicité ni mise en concurrence	31
Erreur d'appréciation dans l'évaluation d'une candidature ou d'une offre	29
Méthode de notation irrégulière	29
Méconnaissance des spécifications légales applicables au marché	26
Manquement à l'obl de contrôle des capacités professionnelles, techniques et financières	24
Manquement aux règles de publicité	23
Interdiction d'écarter une candidature ou une offre valide	23
Modification des caractéristiques du marché, modification de l'offre ou du contrat	23
Critère ou sous-critère irrégulier	20

# Vérité terrain dans DeCoMaP

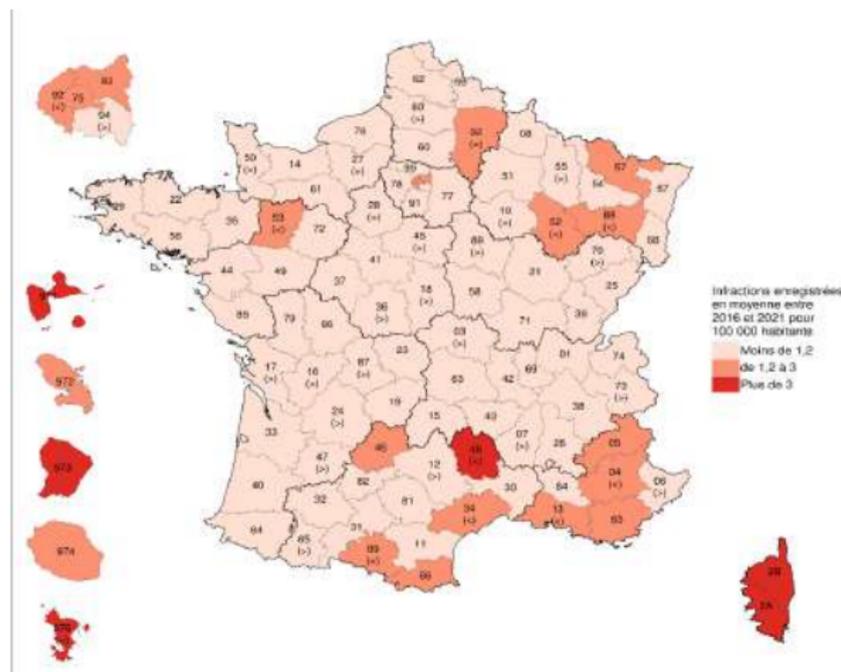


Figure: Atteintes à la probité / 100 000h hab

# Données : que cherchons-nous à / pouvons nous regarder ?

## Variables explicatives

- données des marchés
  - procédure, publicité, critères, niveau de la concurrence...
- données agrégées à l'échelle des acteurs
  - concentration, diversité des fournisseurs, parts de marché
- graphe des acteurs
  - liens contractuels
  - liens de contrôle des structures publics, RNE
  - liens de propriété
  - Décision CJUE sur la Base du Registre des Bénéficiaires effectifs

# Comprendre les liens entre variable explicative et expliquée définir les redflags

Evolution du taux d'atteinte à la probité (échelle départementale) / aux Red Flags

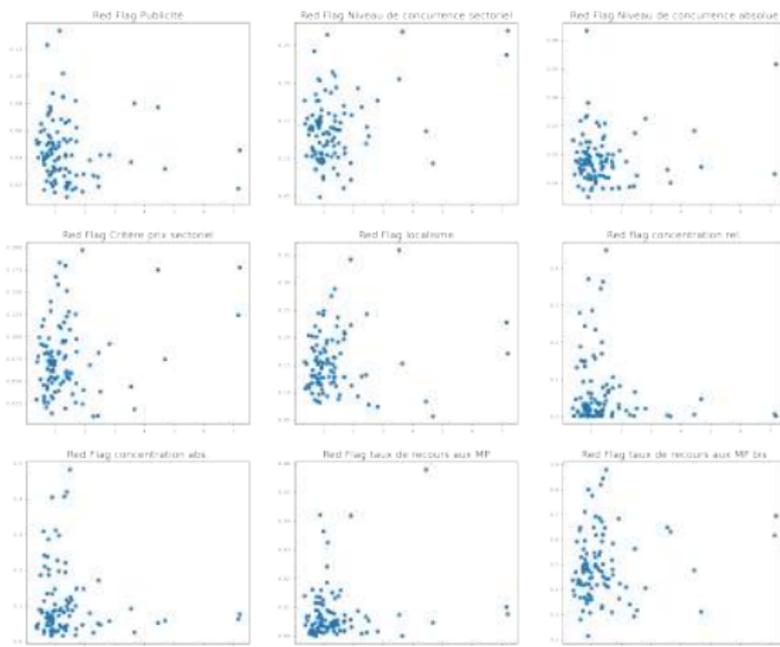


Figure: relations entre atteinte à la probité - échelle dep et tx de redflag

# Sinon quel proxy ?

Maréchal, F., & Morand, P. H. (2022). Are social and environmental clauses a tool for favoritism? Analysis of French public procurement contracts. *European Journal of Political Economy*

	<i>Dependent variable:</i>
	<b>local</b>
<b>Allotment</b>	<b>0.618*** (0.523, 0.713)</b>
<b>Publicity</b>	<b>-0.512*** (-0.627, -0.396)</b>
Accelerated Procedure	-0.249 (-0.669, 0.170)
Negotiated Procedure	-0.263 (-0.601, 0.075)
<b>Competitive Procedure with Negotiation</b>	<b>0.555*** (0.453, 0.658)</b>
Restricted Procedure	-0.395 (-1.281, 0.491)
<b>Social clause</b>	<b>-0.269** (-0.475, -0.064)</b>
<b>Environmental clause</b>	<b>-0.219** (-0.376, -0.061)</b>
Constant	0.107 (-0.047, 0.261)
Observations	7,048
Log Likelihood	-4,747.380
Akaike Inf. Crit.	9,512.760

Note: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01

# Lier proxy de corruption et red flags

proxy : taux départemental d'atteintes à la probité

OLS Regression Results

---

Dep. Variable:	taux	R-squared:	0.525
Model:	OLS	Adj. R-squared:	0.498
Method:	Least Squares	F-statistic:	19.68
Date:	Thu, 20 Jul 2023	Prob (F-statistic):	3.73e-13
Time:	17:56:55	Log-Likelihood:	-100.30
No. Observations:	95	AIC:	212.6
Df Residuals:	89	BIC:	227.9
Df Model:	5		
Covariance Type:	nonrobust		

---

	coef	std err	t	P> t	[0.025	0.975]
const	0.7242	0.640	1.132	0.261	-0.547	1.996
RedFlag_PublicityDuration	19.4324	11.622	1.672	0.098	-3.660	42.524
RedFlag_tenderNumber	3.5785	0.861	4.156	0.000	1.868	5.289
RedFlag_Price_weight	9.5387	1.519	6.278	0.000	6.520	12.558
RedFlag_taux_JOUE	-0.1784	0.576	-0.310	0.757	-1.322	0.966
RedFlag_taux_WIN_share	-1.3672	0.496	-2.759	0.007	-2.352	-0.383

---

Omnibus:	11.317	Durbin-Watson:	1.792
Prob(Omnibus):	0.003	Jarque-Bera (JB):	13.340
Skew:	0.638	Prob(JB):	0.00127
Kurtosis:	4.319	Cond. No.	227.

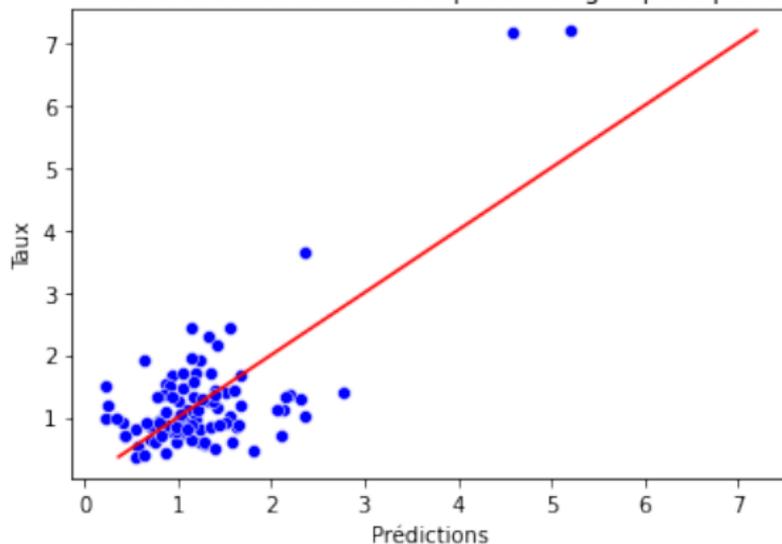
---

Figure: par département, les redflags expliquent-ils le taux d'atteinte à la probité ?

# Lier proxy de corruption et red flags

proxy : taux départemental d'atteintes à la probité

Régression linéaire avec les variables explicatives groupées par département



# Même logique, méthodes nouvelles : Graph-based anomaly detection

## Definition 1 (Hawkins' Definition of Outlier, 1980)

“An outlier is an observation that differs so much from other observations as to arouse suspicion that it was generated by a different mechanism.”

## Definition 2 (General Graph Anomaly Detection Problem)

Given a (plain/attributed, static/dynamic) graph database, Find the graph objects (nodes/edges/substructures) that are rare and that differ significantly from the majority of the reference objects in the graph.

## Différence juridique, continuum économique

- je favorise une entreprise que je sais performante
- je favorise une entreprise qu'on m'a dit performante
- je favorise une entreprise en phase avec mes objectifs politiques
- je favorise une entreprise d'un ami
- je favorise une entreprise parce que c'est celle d'un ami
- je favorise une entreprise parce qu'elle me rémunère pour cela
- je favorise une entreprise parce qu'elle offre un emploi (fictif)
- ...

- en vous remerciant de votre attention...
- et dans l'attente de vos questions !

# Graph: Consulting markets graphs in Potin et al. (2022)

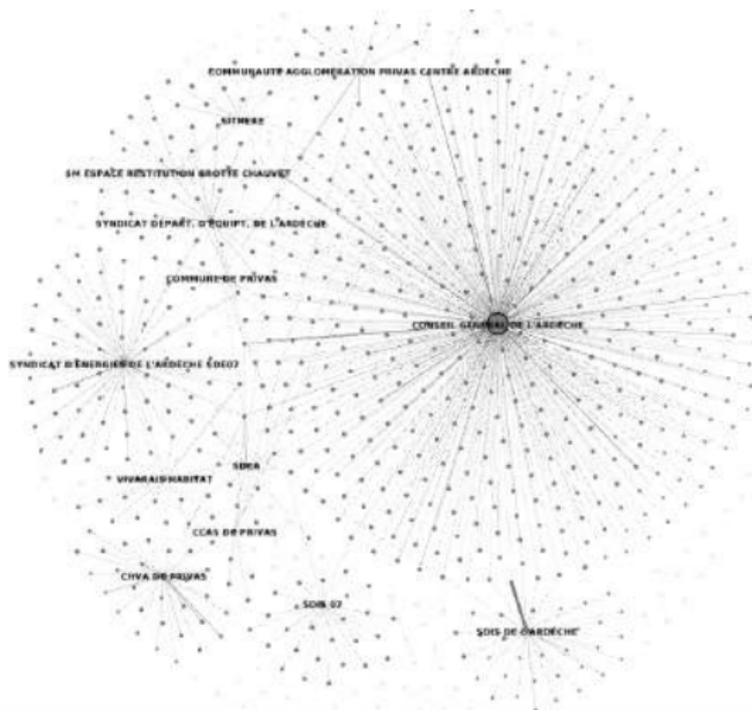


Figure: the network of consulting markets

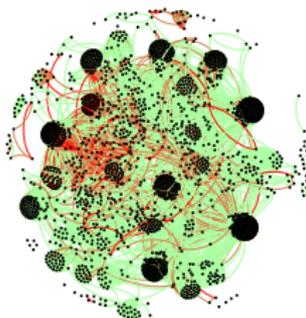


Figure: Red flags in the projected graph

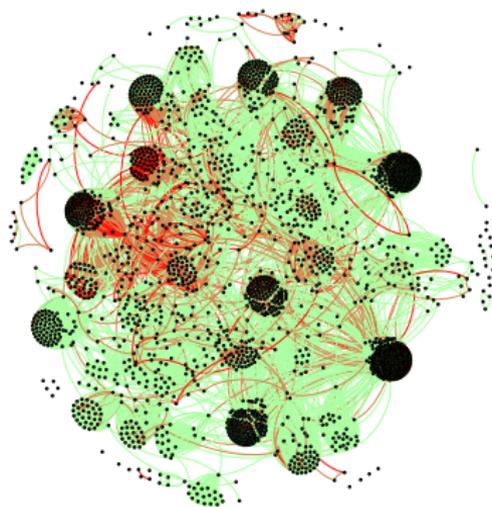


Figure: Global anomalies detected

**Nouvelle approche** : trouver des structures dans les graphes liées à des comportements redflags et non redflags

- Construction d'une collection de graphes redflags ( $\mathcal{G}_r$ ) et non redflags ( $\mathcal{G}_{nr}$ ).
- Identification des sous graphes discriminants dans chaque collection ( $\mathcal{P}_r$  et  $\mathcal{P}_{nr}$ )
- Représentation des graphes via ces sous graphes
- Classification des graphes en redflags ou non redflags, sans connaissance de l'existence des redflags.

### Identification de motifs discriminants:

Entrée: Ensemble de graphe  $\mathcal{G} = \mathcal{G}_r \cup \mathcal{G}_{nr}$

Sortie: Ensemble de motifs  $\mathcal{P} = \mathcal{P}_r \cup \mathcal{P}_{nr}$

### Recherche de motifs:

Entrée: Ensemble de graphe  $\mathcal{G}$

: Ensemble de motifs  $\mathcal{P}$

Sortie: Ensemble de vecteurs booléens  $\mathcal{X}$ , tel que  $X_{ij} = 1$

: si le motif  $P_j$  est présent dans le graphe  $G_i$

### Classification:

Entrée: Ensemble de vecteurs booléens  $\mathcal{X}$

Output: Associer à chaque vecteur un label redflag ou non redflag

# Work in progress

## Predictive model

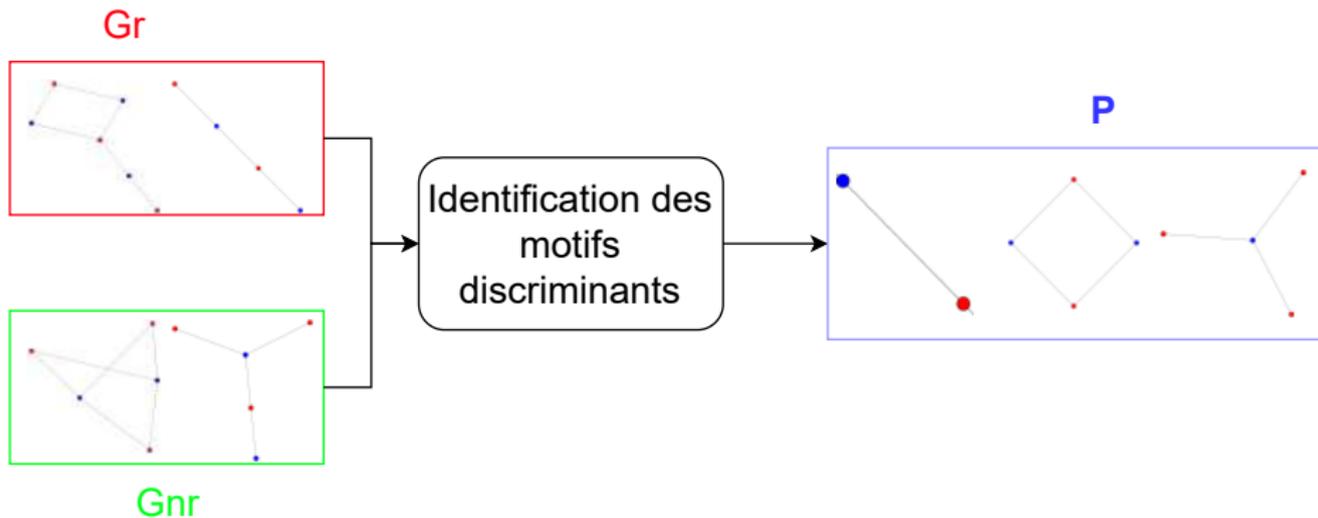
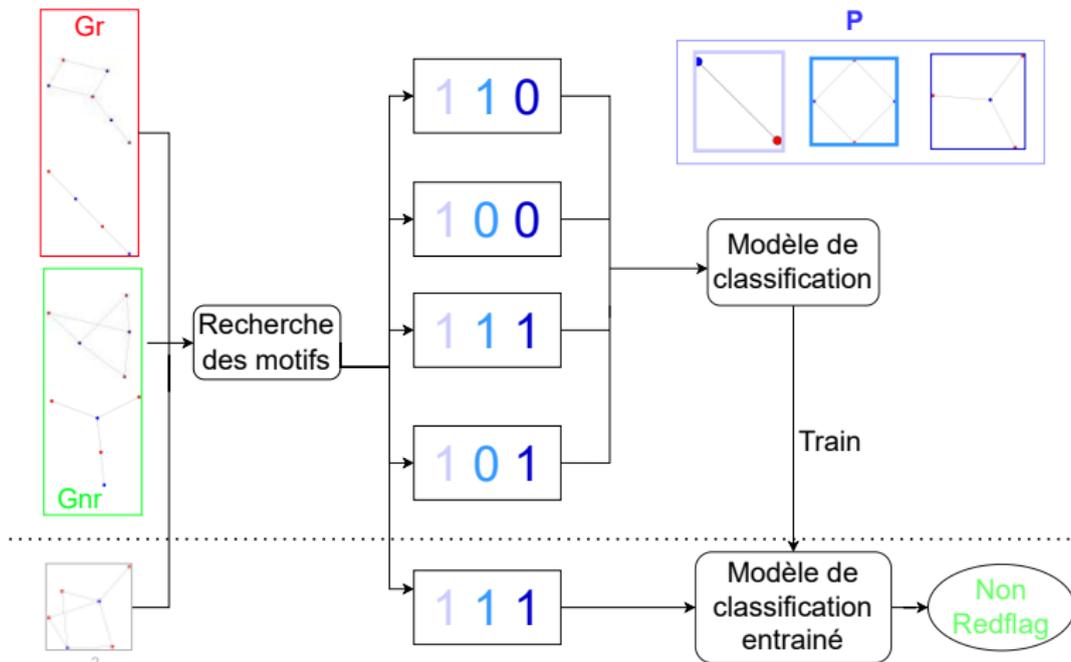


Figure: Partie identification des sous graphes

# Work in progress

## Predictive model



- Collection de 660 graphes (330 redflags et 330 non redflags)
- Filtrage sectoriel (marché de travaux), temporel (périodes d'un an), et géographique (départemental).
- Méthode de recherche de motifs : TKG.
- Méthode de classification : Random Forest

# Work in progress

## Predictive model

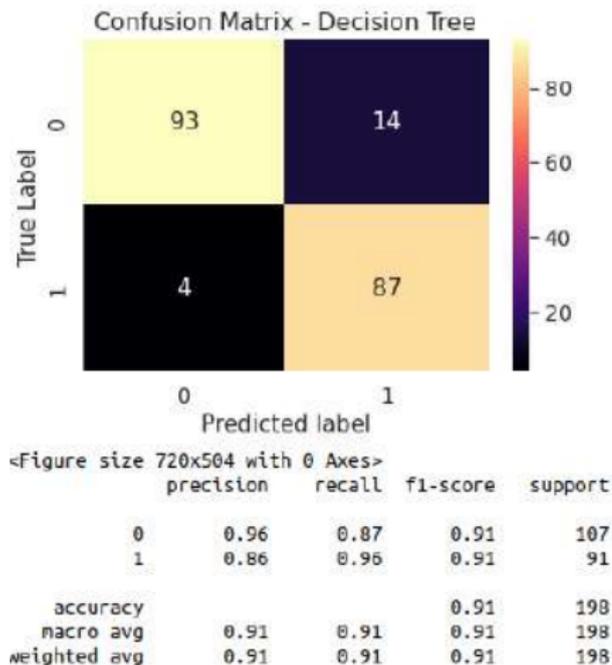


Figure: Résultats

# What are we looking for?

a mechanism design approach

- The degree  $\alpha_i \in [0, 1]$  to which the public buyer takes firm  $i$ 's profit,  $\Pi_i$ , into account.
- The public buyer assessment,  $S_i$ , of the fact that firm  $i$  is carrying out the contract.
- The surplus  $V(q_i)$ , generated by the fact that the contract is carried out with a quality level  $q_i$ .
- When choosing a level of quality  $q_i$ , firm  $i$  incurs a cost  $c(q_i, \theta_i)$ , where  $\theta_i$  represents its private efficiency parameter.  $\theta_i \sim F(\cdot)$
- The potential bribe, equal to a fraction,  $\gamma_i \geq 0$ , of the payment of the contract, given by firm  $i$  to the public buyer.

$$[x_i(\Theta), t_i(\Theta)]_{i=1, \dots, n}$$

# What are we looking for?

a mechanism design approach

The generic objective function:

$$E_{\Theta} W = \int_{\Theta} \sum_i \left[ \left( S_i + V(q_i) \right) x_i - (1 + \lambda) t_i + \alpha_i \left( (1 - \gamma_i) t_i - c(q_i, \theta_i) x_i \right) + \gamma_i t_i \right] dF(\Theta)$$

The generic modified objective function (i.e incentives into account)

$$E_{\Theta} W = \int_{\Theta} \sum_i x_i \left[ \left( S_i + \hat{V}(q_i) \right) - \frac{1 + \lambda - \gamma_i}{1 - \gamma_i} c(q_i, \theta_i) - \left( \frac{1 + \lambda - \gamma_i}{1 - \gamma_i} - \alpha_i \right) \frac{F_i(\theta_i)}{f_i(\theta_i)} c_{\theta}(q_i, \theta_i) \right] dF(\Theta)$$

Given its parameter  $\theta_i$ , firm  $i$  is induced to choose an optimal quality

$$q^*(\theta_i) = \arg \max_{q_i} V(q_i) - J_i(q_i, \theta_i)$$

with  $J_i(q_i, \theta_i) = \frac{1 + \lambda - \gamma_i}{1 - \gamma_i} c_i(q_i, \theta_i) + \left( \frac{1 + \lambda - \gamma_i}{1 - \gamma_i} - \alpha_i \right) \frac{F_i(\theta_i)}{f_i(\theta_i)} c_{\theta}(q_i, \theta_i)$ .

# What are we looking for?

a mechanism design approach

Firm  $t$  e.g. is awarded the contract if

$$S_t + V(q^*(\theta_t)) - J_t(q^*(\theta_t), \theta_t) > S_s + V(q^*(\theta_s)) - J_s(q^*(\theta_s), \theta_s) \quad \forall t, s.$$

from Marechal & Morand (2022)

The mere observation of procedural choices does not theoretically reveal the preferences of the public purchaser.

- $\gamma_i$  VS  $\alpha_i$  ?
- lawful or unlawful preferences?
- lawful or unlawful selection rule?