

Descartelizando: Uso de *Machine Learning* e Estatística para Detecção de Indícios de Cartel em Processos Licitatórios

Paulo Jhonny S. da C. Rosa¹, Walmes M. Zeviani²

¹Aluno do programa de Especialização em Data Science & Big Data, pjsdcrosa@mppr.mp.br;

²Professor do Departamento de Estatística - DEST/UFPR, walmes@ufpr.br

Introdução

Um processo licitatório é o meio administrativo pelo qual a Administração Pública adquire bens, produtos e serviços indispensáveis ao cumprimento de suas obrigações, tendo como objetivo escolher, dentre vários concorrentes, a proposta mais vantajosa no que concerne aos aspectos de preço e qualidade. No entanto, em virtude do grande volume de recursos públicos envolvidos, o princípio fundamental da competitividade pode ser infringido mediante a formação de cartel, que é uma espécie de acordo entre empresas no sentido de combinar preços para eliminar a concorrência. Ishii (2009) (1) explica que a atuação de cartéis implica em contratos com valor mais alto e dificilmente inferior a 90% ou 95% de um valor de referência. Ainda, estimativas da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) demonstram que cartéis geram um sobrepreço estimado entre 10% e 20% comparado ao preço de um mercado competitivo.

No Ministério Público do Estado do Paraná (MPPR), as investigações relacionadas a essa prática são iniciadas, na maioria das vezes, a partir de denúncias anônimas realizadas por terceiros, geralmente empresas prejudicadas pela existência de conduta anticompetitiva no certame do qual participaram. Além disso, o crescimento exponencial do volume de procedimentos licitatórios e a inexistência de mecanismos ágeis para identificação de atos colusivos impõem desafios à atuação proativa dos guardiões da lei. Por conseguinte, torna-se imperiosa a implementação de estudos científicos que permitam a produção de conhecimento pautado à formação de convicção fundamentada, visando subsidiar tomadores de decisão no enfrentamento desse problema. Diante do exposto, buscou-se desenvolver neste trabalho uma metodologia para detecção de indícios de cartel em licitações relativas à aquisição de produtos, obras e serviços de engenharia, considerando a gestão municipal paranaense de 2013 a 2016.

Materiais e Métodos

A partir de dados de 25.169 licitações extraídos de fontes públicas e do *software R*, foram realizados os seguintes procedimentos metodológicos:

- **Método 1** - Inicialmente, utilizou-se da técnica de clusterização *k-means* para agrupar empresas com base no número de participações em licitações nos 11 gepatrias do Paraná (unidades do MPPR responsáveis pela defesa do Patrimônio Público). Na sequência, cada gepatria foi associado ao grupo com maior média dos percentuais de participação das empresas, formando-se, então, mercados de licitação;
- **Método 2** - Considerando o Paraná e cada grupo de gepatrias, aplicou-se o algoritmo *Apriori* de descoberta de regras de associação visando detectar potenciais cartéis caracterizados pela atuação frequente e conjunta de empresas. Em seguida, além do uso de alguns critérios de poda, foram propostas duas funções de avaliação de regras, extensões de Silva (2011) (2) e Fraga (2017) (3), quais sejam: a probabilidade de o grupo de empresas firmar um contrato com a Administração Pública (Val_1) e a mediana das razões entre os valores dos contratos celebrados pelo grupo e os valores orçados pela Administração Pública (Val_2). Os resultados foram representados por um indicador de risco de cartel, que serviu para classificar as licitações;
- **Método 3** - Foi realizada uma análise associativa entre o indicador criado e fatores de risco estabelecidos a partir de características de licitações, contratos e empresas. Para tanto, fez-se uso de uma rede neural artificial com aprendizado não supervisionado, denominada *Self-Organizing Maps* (SOM) (KIND, 2013 (4)).

Identificação de Mercados de Licitação

A partir do método 1, foram identificados três grupos distintos de gepatrias, como pode ser visualizado na Figura 1. Nota-se que, mesmo não tendo considerado uma continuidade espacial, os grupos encontrados foram formados por gepatrias vizinhos geograficamente. Nas Figuras 2, 3 e 4, é possível observar os fluxos origem-destino das empresas participantes de licitação nos grupos 1, 2 e 3, respectivamente. Observa-se que empresas mais distantes dos respectivos grupos de atuação participaram de poucas licitações, enquanto que as mais próximas, inclusive localizadas na região, tiveram uma quantidade de participação mais elevada.



Figura 1: Mercados de Licitação

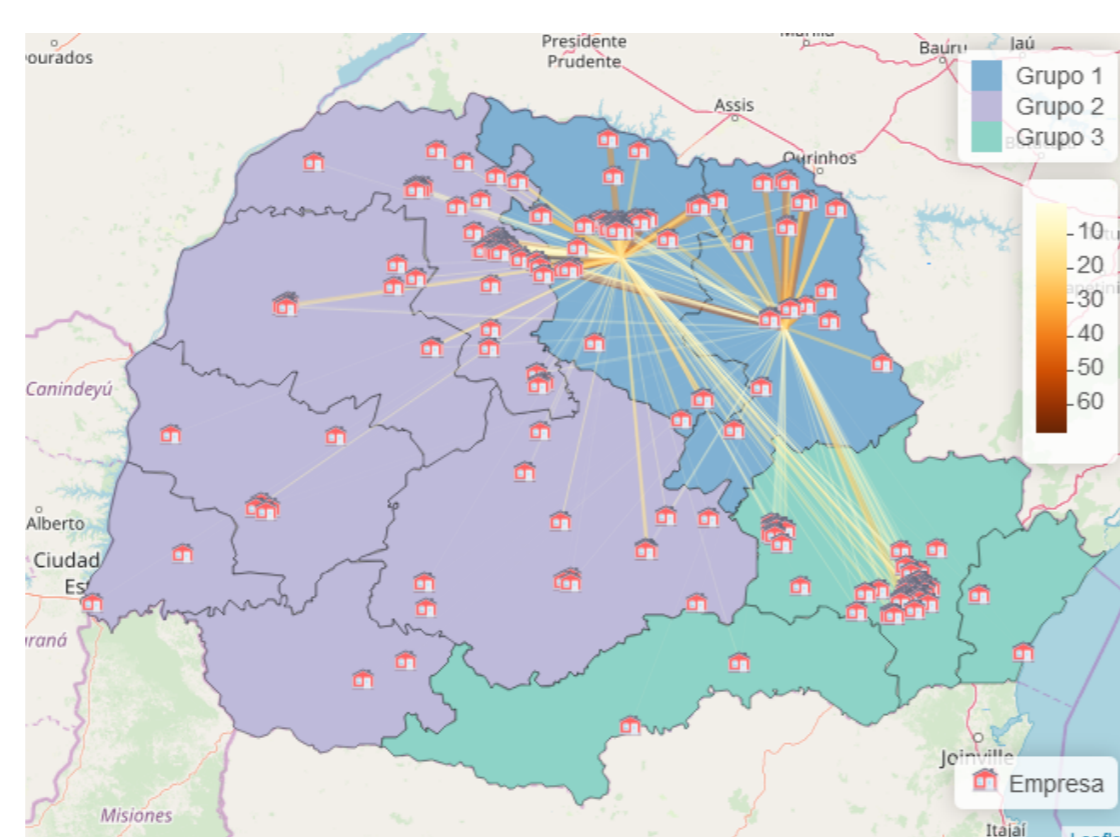


Figura 2: Fluxos origem-destino (G1)

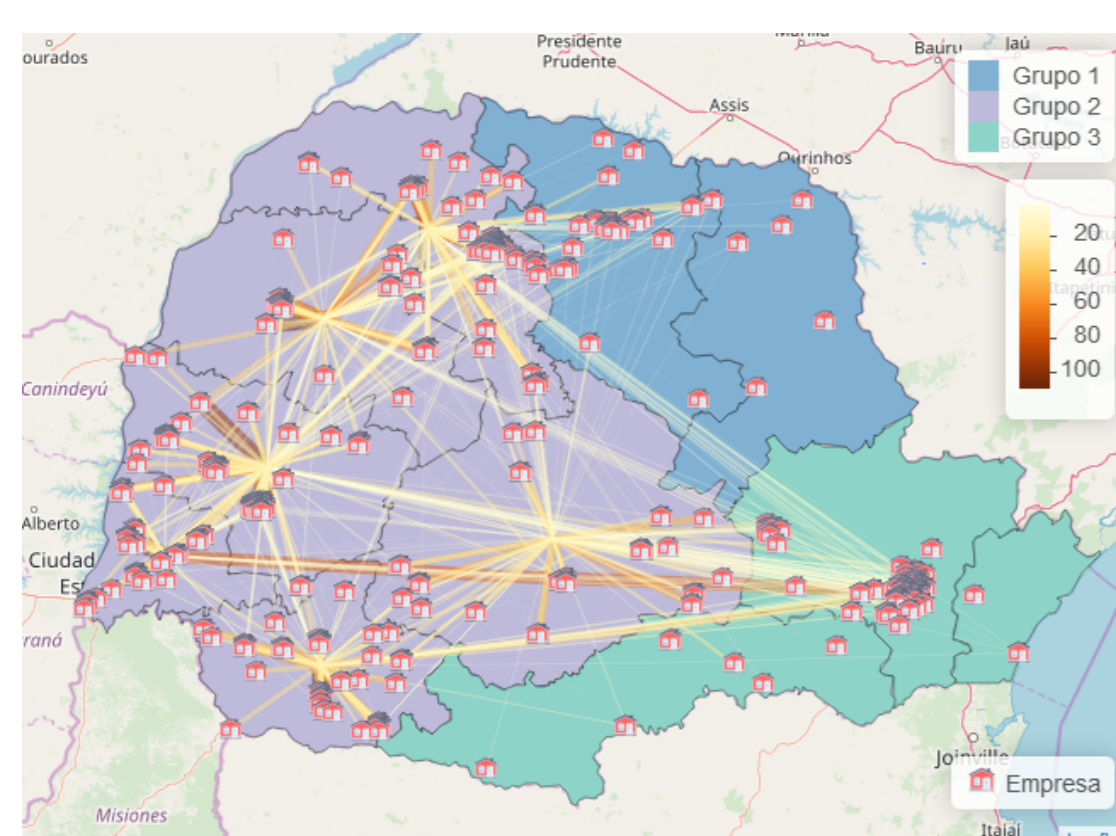


Figura 3: Fluxos origem-destino (G2)

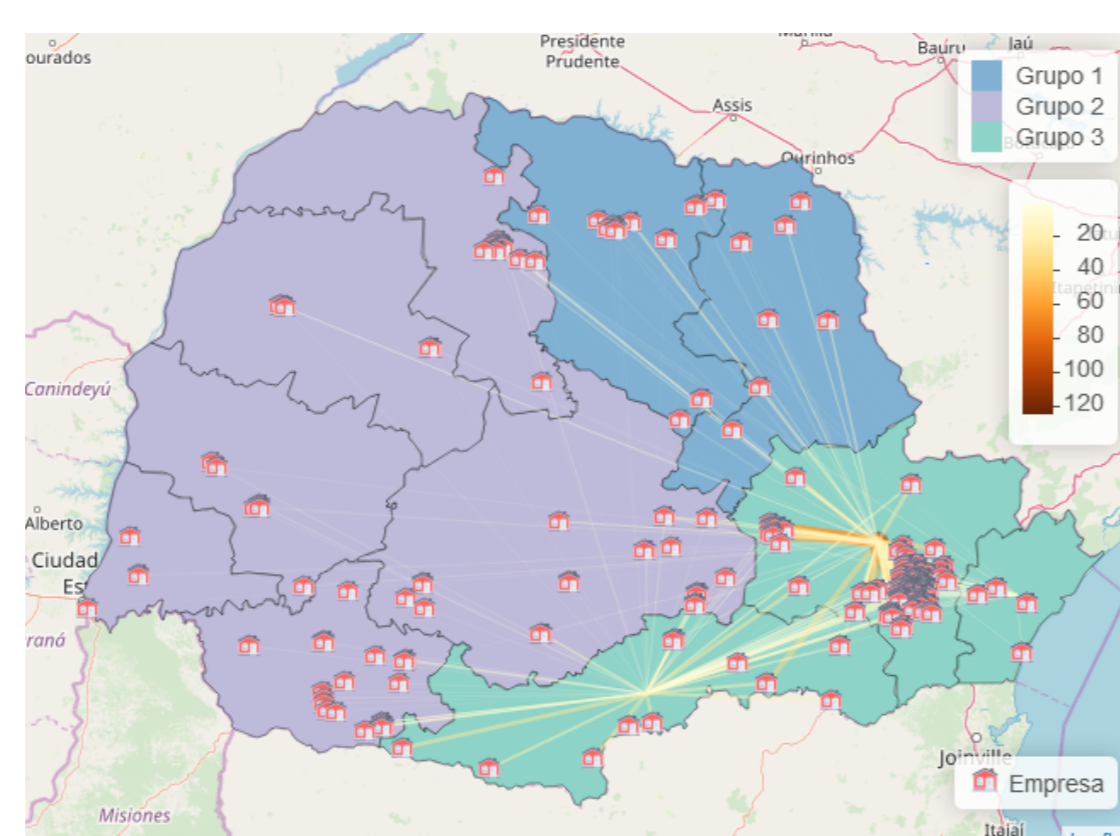


Figura 4: Fluxos origem-destino (G3)

Detecção de Potenciais Cartéis

Na Tabela 1, são apresentados os resultados relativos ao processo de descoberta de indícios de cartel por meio do treinamento de regras de associação, considerando o estado do Paraná e os três mercados de licitação. Foram testados três valores de suporte para o estado como um todo e apenas o suporte com no mínimo cinco participações em licitações para análise dos *clusters*.

Tabela 1: Quadro resumo da aplicação do método 2

	Paraná (suportes)			G1	G2	G3
	5	10	15	5	5	5
	12567	12567	12567	2169	7947	2451
Total de regras	769	64	21	83	310	403
Selecionadas	245	22	7	33	149	76
Val_1 médio	0,72	0,88	0,91	0,70	0,81	0,57
Val_2 médio	0,83	0,87	0,92	0,84	0,85	0,81
Risco médio	0,79	0,86	0,91	0,77	0,83	0,72

No que concerne ao Paraná, das 25.169 licitações, 1.364 (3.631 contratos) tiveram pelo menos uma das 245 regras aplicadas (5,4%). Em relação aos mercados de licitação, foram detectadas 230 licitações (5,3%) com indícios de cartel no G1, 814 (4,9%) no G2 e 279 (6,6%) no G3. Além disso, 191 licitações suspeitas selecionadas por meio da análise macro do estado do Paraná não foram identificadas via análise dos mercados de licitação. Em contrapartida, 150 processos licitatórios com indícios de conluio foram detectados somente a partir dos grupos de gepatrias. As contratações públicas oriundas de licitações suspeitas no Paraná custaram aos cofres públicos o equivalente a R\$ 659.638.750,00 milhões, cerca de 7,5% do total contratado pelos municípios no período. Ainda, considerando apenas as licitações suspeitas identificadas a partir dos mercados de licitação, o custo passou a ser de R\$ 658.521.696,00 milhões.

Como pode ser observado na Figura 5, além de as distribuições de densidade serem parecidas, há uma grande massa de dados concentrada em um patamar acima do limiar de 0,7 de risco, indicando fortes evidências progressivas de conluio na maioria das licitações selecionadas via regras de associação, principalmente aquelas cujo risco se encontra acima de 0,9 (365 licitações no Paraná, 56 no G1, 274 no G2 e 88 no G3). A média de risco de cartel considerando o estado como um todo foi de 0,838. Nos mercados de licitação G1, G2 e G3, esse valor passou a ser de 0,83, 0,839 e 0,842, respectivamente. Isso demonstra que, muito embora 341 licitações supostamente cartelizadas foram descobertas em apenas um dos recortes espaciais, 191 no estado e 150 nos grupos, no geral, os resultados de ambas as configurações foram parecidos.

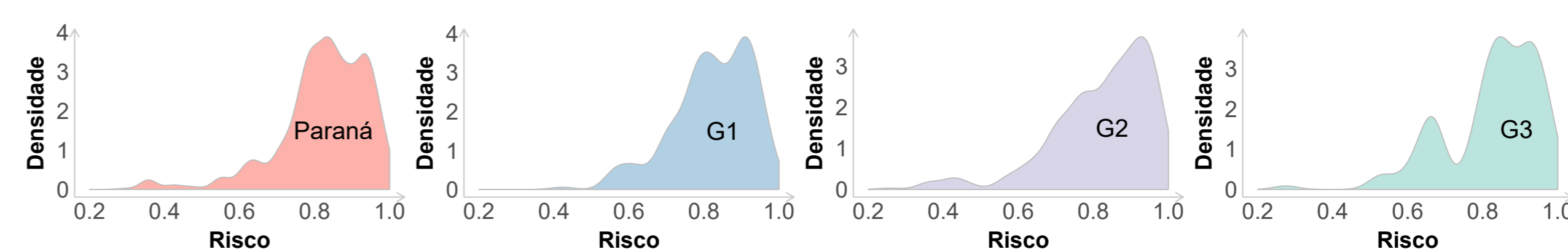


Figura 5: Distribuições de densidade do indicador de risco de cartel no estado e nos três grupos

Relação com Fatores de Risco

Os *heatmaps* da Figura 6 mostram as relações entre os indicadores de indícios de cartel e dez tipologias de risco. A graduação de cores, do amarelo ao azul, representa a variação de cada atributo, e a posição de cada neurônio é a mesma em todos os *heatmaps*. Nos *heatmaps* 1 e 2, é possível perceber uma associação entre os riscos de cartel identificados a partir do Estado e dos três mercados de licitação, ficando evidente a semelhança de resultados de ambos os espaços de solução, corroborando com o observado na Figura 5. Além disso, ocorre uma tendência de associação negativa entre a quantidade de participantes na licitação (*quantPartic*) e os riscos de cartel, o que já era esperado uma vez que ambientes com poucos competidores contribuem para uma atuação colusiva de um cartel.

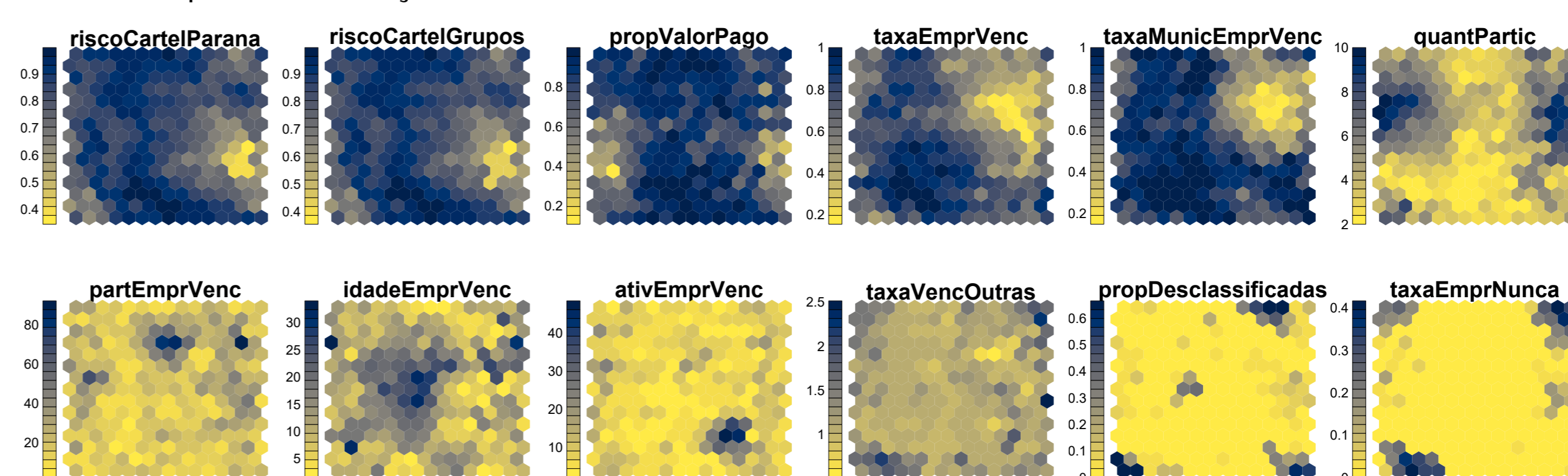


Figura 6: Relação de dez tipologias de risco com os indicadores de cartel

Conclusões

O uso de métodos estatísticos e de *machine learning* se mostrou bastante promissor para a detecção de padrões de possíveis cartéis. A identificação de fornecedores cooperantes entre si, a partir de procedimentos criteriosos e científicos, respalda a possibilidade de uso do conhecimento adquirido no MPPR, principalmente no assessoramento da atividade-fim ministerial, quando da tomada de decisão, propiciando uma atuação proativa e otimizada em relação à atividade colusiva de empresas. Entretanto, é necessário frisar que o aprendizado alcançado fornece tão somente indícios de suspeição da existência de cartéis, podendo ser utilizado apenas para fins de inteligência e não para produção de provas em processos judiciais.

Principais Referências

- [1] R. Ishii, *Favor exchange in collusion: Empirical study of repeated procurement auctions in Japan*, International Journal of Industrial Organization, pag. 137-144 (2009)
- [2] C. V. S. Silva, C. G. Ralha, *Agentes de Mineração e sua Aplicação no Domínio de Auditoria Governamental*, Universidade de Brasília, Instituto de Ciências Exatas, Departamento de Ciência da Computação (2011)
- [3] A. A. Fraga, H. M. B. Ramalho, A. T. C. Almeida, *Detecção de casos suspeitos de fraudes em licitações realizadas nos municípios da Paraíba: uma aplicação de técnicas de mineração de dados*, Universidade Federal da Paraíba, Centro de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Economia do Setor Público (2017)
- [4] M. C. Kind, R. J. Brunner, *SOMz: photometric redshift PDFs with self organizing maps and random atlas*, Department of Astronomy, University of Illinois, Urbana, USA (2013)