

Detecção de Casos Suspeitos de Conluio em Licitações Públicas: uma Aplicação do Algoritmo a Priori de Aprendizado de Máquina para o Estado da Paraíba

Detection of Suspected Collusion Cases in Public Biddings: an Application of the a Priori Machine Learning Algorithm for the State of Paraíba

Hilton Martins Brito Ramalho

Universidade Federal da Paraíba – UFPB – Brasil

hilton@lema.ufpb.br

ORCID: 0000-0003-1501-8652

Aléssio Tony Cavalcanti de Almeida

Universidade Federal da Paraíba – UFPB – Brasil

alessio@lema.ufpb.br

ORCID: 0000-0003-0436-359X

Alcimar Alves Fraga

Tribunal de Contas do Estado da Paraíba – TCE-PB – Brasil

alcimarfraga70@gmail.com

ORCID: 0000-0003-1676-5190

Submetido em 30/03/2020; **Aprovado** em 09/07/2020.

Resumo

Objetivo: Este artigo objetiva identificar casos potencialmente suspeitos de conluio em licitações de gestões municipais da Paraíba de 2005 a 2016. **Metodologia:** Para tanto, foram utilizados dados do Tribunal de Contas do Estado da Paraíba e o algoritmo de aprendizado de máquina Apriori para a construção de regras de associação em uma amostra de 104 mil licitações, com 60 mil proponentes. As regras de associação foram avaliadas considerando suas consistências com padrões de estratégias cooperativas em jogos repetidos, especialmente pela mensuração da probabilidade de vitória de grupos de empresas com atuações cíclicas e da média de concorrentes. As empresas mapeadas nas regras de associação foram ranqueadas pelo indicador de suspeição da empresa (ISE), diretamente relacionado à probabilidade de vitória, à associação com falsos concorrentes e à atuação espacialmente concentrada. O ISE é uma contribuição proposta por esta pesquisa. **Resultados:** Os resultados revelaram indicações de suspeição de conluio para várias empresas dos ramos de alimentação, serviços de limpeza, locação de veículos, copiadoras, palcos, banheiros químicos, fornecimento de combustíveis, peças automotivas, material esportivo, medicamentos e material hospitalar, consultorias nas áreas de contabilidade, engenharia civil e publicidade, considerando cada um dos três períodos de gestão municipal avaliados (2005–2008, 2009–2012 e 2013–2016). Ademais, foram encontrados vários padrões de associação entre licitantes com evidências de simulação de concorrência e de atuação concentrada em poucos municípios. **Contribuição:** Os resultados gerados, particularmente pelo índice de suspeição de empresa, fornecem uma triagem inicial para otimização de processos de fiscalização e de auditorias em órgãos de controle, cuja identificação incontestada de potencial fraude nas licitações requer, necessariamente, avaliações processuais complementares e específicas para cada caso.

Palavras-chave: Licitações. Conluio. Aprendizado de máquina. Regras de associação. Paraíba.

Abstract

Purpose: This paper aims to identify potentially collusive bidding schemes in municipal government biddings of Paraíba state during 2005 to 2016. **Methodology:** For this purpose, data from the Court of Auditors of the State of Paraíba and the Apriori machine learning algorithm were used to build association rules in a sample of 104 thousand biddings, with 60 thousand bidders. The association rules were evaluated considering patterns of cooperative strategies in repeated games, taking account the bid winner likelihood and the bidders competition on average. How companies mapped in the association rules

were ranked by the firm's suspicion indicator (ISE), directly related to the probability of victory, association with false tests and spatially concentrated performance. ISE is a contribution proposed by this research. **Findings:** The results revealed indications of suspected collusion for various companies such as meals suppliers, cleaning services, rental of vehicles, copiers, stages, chemical toilets, fuel suppliers, automotive parts, sports equipments, medicines and hospital material, consulting in accounting, civil engineering and advertising, considering each of the three municipal management periods (2005–2008, 2009–2012 and 2013–2016). We also found several patterns of association between bidders with evidence of competition simulation and concentrated activity in a few municipalities. **Contribution:** the results, particularly by the firm's suspicion indicator, provide important initial guidelines to optimize inspection and auditing processes in justice control bodies, whose undisputed identification of a potential bidding fraud, necessarily, requires complementary and specific procedural evaluations for each case.

Keywords: Procurement Auctions. Collusion. Machine learning. Rules association. Paraíba.

1 Introdução

Segundo relatório da Organização para Cooperação e Desenvolvimento Econômico (OCDE), em 2008, os processos de licitações públicas executados nos países-membros responderam por cerca de 15% do Produto Interno Bruto (PIB) e por 30% dos gastos governamentais em média (OCDE, 2012). Apesar de sua importância na economia do setor público, os mercados de licitações são vulneráveis a ações fraudulentas e à corrupção, sobretudo em países em desenvolvimento. Em 2015, o relatório sobre transparência internacional constatou que 67% dos 168 países avaliados apresentaram sérios problemas de corrupção, inclusive todos os países-membros do BRICS (Brasil, Rússia, Índia, China e África do Sul), com o Brasil ocupando a 76ª posição. De acordo com esse relatório, países pobres chegam a perder anualmente em torno de US\$ 1 trilhão em decorrência da corrupção (Transparency International, 2016).

No Brasil, a administração pública tem se preocupado cada vez mais com a qualidade na oferta de serviços públicos, a disponibilidade de ferramentas para controle social e a prestação de contas com transparência. No entanto, apesar de vários esforços de fiscalização, tem se observado que os mercados de licitações estão sujeitos a práticas anticompetitivas por parte de grupos de empresas, muitas vezes com a participação de agentes públicos (Campos, 2008). Segundo dados da Secretaria de Defesa Econômica do Ministério da Justiça, considerando-se o valor médio anual de R\$ 300 bilhões para compras e aquisições diversas em processos licitatórios, estima-se que práticas fraudulentas no Brasil gerem um prejuízo em torno de R\$ 25 a R\$ 40 bilhões, o que representa uma média de 10% de desvio de recursos públicos (Campos, 2008).

Ao se considerar o quadro de persistência de desigualdades regionais e sociais no Brasil, os estados mais pobres da federação podem ser mais prejudicados com esquemas de fraudes em processos licitatórios. O estado da Paraíba, por exemplo, se enquadra nessa situação. Conforme o Instituto Brasileiro de Geografia e Estatística (IBGE, 2018), a Paraíba registrava Produto Interno Bruto (PIB) de R\$ 59 bilhões a preços correntes no ano de 2016, valor equivalente a 3% do PIB de São Paulo (aproximadamente R\$ 2 trilhões), estado mais rico do país. Por outro lado, segundo dados do Tribunal de Contas do Estado da Paraíba (TCE-PB) para o mesmo período, os valores licitados pelos municípios paraibanos totalizaram cerca de 23% do PIB estadual.

Diante do exposto, urge a necessidade de pesquisas científicas que forneçam bases teóricas e metodológicas para identificação de comportamentos de empresas e prestadores de serviços condizentes com potenciais ações fraudulentas nos mercados de licitações. Vale observar que o crescimento do volume de processos licitatórios específicos em intervalos de tempo mais curtos impõe vários desafios no âmbito dos órgãos de controle. O uso de tarefas de aprendizagem de máquina tem sido cada vez mais recorrente na análise de grandes volumes de dados (*big data*). Na literatura internacional, por exemplo, há vários estudos que procuraram atuar nesse campo de investigação, sobretudo as abordagens teóricas de Porter e Zona (1993), Bajari e Ye (2003) e análises empíricas de Hüscherlath e Veith (2008), Blankenburg, Geist e Kholodilin (2012) e Ormosi (2014). No Brasil, destacam-se os trabalhos de Da Silva et al. (2005), Balaniuk et al. (2012) e Ralha e Silva (2012), que aplicaram técnicas de aprendizado de máquina para descoberta de conhecimento a partir de cruzamento de informações, detecção de padrões específicos, identificação de observações aberrantes e associações entre variáveis.

Dessa forma, o objetivo principal deste estudo é identificar padrões de associação entre empresas

licitantes consistentes com comportamentos teoricamente esperados em ambientes anticompetitivos. Para tanto, foram realizados experimentos de aprendizado de máquina baseados no algoritmo de análise combinatória *Apriori*, desenvolvido por Agrawal e Srikant (1994), para licitações municipais ocorridas no estado da Paraíba durante o período de 2005 a 2016, que cobre três ciclos completos de gestão (2005-2008, 2009-2012 e 2013-2016). A presente pesquisa pretende contribuir para o campo de investigação ao estender a função de avaliação proposta por Ralha e Silva (2012), de modo que a seleção de padrões comportamentais descobertos por tarefas de aprendizado de máquina considere não apenas a tendência de vitória de empresas licitantes, mas o ambiente de competição (quantidade de concorrentes). Aspectos como o grau de associação entre empresas, a recorrência em diversos grupos (rodízios), a tendenciosidade de vitória, a simulação de competição e a concentração regional são considerados na construção de um indicador de suspeição por empresa, sendo este último uma contribuição inédita da pesquisa. Em linhas gerais, procura-se avançar no instrumental metodológico, concebendo resultados mais robustos para a literatura especializada e uso futuro por especialistas de órgãos de controle, como análise exploratória inicial de potenciais riscos em processos licitatórios.

2 Revisão da literatura

No intuito de reduzir custos de aquisição de bens públicos e de minimizar desperdícios de recursos, a administração pública, inclusive no Brasil, tem feito uso do instrumento de licitação. Segundo o Tribunal de Contas da União (TCU), o processo licitatório é um procedimento administrativo formal em que a administração pública convoca, por meio de condições estabelecidas em ato próprio (edital ou convite), empresas interessadas na apresentação de propostas para o fornecimento de bens e serviços, sendo selecionada a proposta mais vantajosa para celebração de contrato (Brasil, 2010).

Conforme ressaltado por Heimler (2012), os mercados de licitações públicas são bastante vulneráveis a ações coordenadas por parte de empresas. Ao contrário de mercados tradicionais, o mecanismo de licitações não permite ajustamentos de oferta mediante variações de preços (oferta fixa), possibilitando ganhos atrativos via conluios. As empresas que praticam comportamentos colusivos (anticompetitivos) nesses mercados costumam articular estratégias de preços, simular concorrência e interagir com agentes públicos no intuito de obter vantagens econômicas. Quando relacionadas a práticas de corrupção, as ações colusivas agravam os prejuízos na alocação de recursos públicos.

As ações anticompetitivas acima mencionadas podem se manifestar de diversas formas. Segundo relatório da OCDE (2009), destacam-se os seguintes casos: i) apresentação de propostas fictícias ou de cobertura; ii) direcionamento de licitação; iii) divisão de mercado; iv) fixação de preços; v) rodízio de vencedores ou ofertantes; vi) subcontratação de empresas do grupo ou entes públicos; vii) supressão de propostas; viii) consórcios entre concorrentes.

Conforme Gibbons (1992), na prática, uma licitação pública pode ser entendida como um jogo que pode se repetir de forma indefinida e com possibilidade de comunicação entre os jogadores (proponentes), haja vista a grande frequência de ofertas nos mercados de licitações e a presença de informação assimétrica, o que dificulta o controle de órgãos fiscalizadores sobre o comportamento das empresas fornecedoras. As empresas, nesse mercado de licitações públicas, têm como escopo central a maximização do retorno econômico, isto é, fechar contrato com a administração pública em condições vantajosas. Nessa abordagem, elas podem decidir por estratégias cooperativas ou não cooperativas.

A literatura documenta vários estudos que usam modelos baseados em teoria dos jogos para entender o resultado de ações estratégicas de empresas participantes de processos de licitações/leilões sob diversas condições de esquemas de rodízio e de comunicação. Destacam-se os estudos de McAfee e McMillan (1992), Athey e Bagwell (2001), Aoyagi (2003), Rachmilevitch (2014) e Wang (2016). McAfee e McMillan (1992), de forma pioneira, desenvolveram um modelo que considera a comunicação entre participantes de leilões. Eles demonstraram que, quando se permitem transferências monetárias entre ofertantes, esquemas de rodízio podem representar um equilíbrio de Nash eficiente para um amplo leque de valores de taxa de desconto intertemporal. Contudo, no caso em que não há transferências, o comportamento colusivo pode não resultar em equilíbrio em jogo estático.

Em linhas gerais, os modelos teóricos documentados na literatura especializada sugerem que disputas em mercados de licitações públicas podem ser tratadas como jogos que se repetem indefinidamente em um ambiente de monitoramento imperfeito. Nesse sistema, os participantes (jogadores) po-

dem ter incentivos para adotar comportamentos colusivos baseados na crença de que ações atuais podem afetar ações futuras, isto é, de que acordos implícitos podem ser sustentados mediante estratégias punitivas, especialmente quando há alta probabilidade de novas interações.

Na literatura empírica há vários estudos que aplicam diferentes metodologias e variáveis na tentativa de identificar fraudes em licitações governamentais. As seguintes informações são comumente usadas: (a) preço ofertado pelas empresas participantes (Abrantes-Metz et al., 2006; Bolotova, Connor & Miller, 2008), (b) preço da oferta vencedora (Lundberg, 2005), (c) razão preço de oferta/preço da oferta vencedora (Porter & Zona, 1993; Pesendorfer, 2000; Bajari & Ye, 2003), e (d) padrão de rodízios de ofertantes e vencedores (Ralha & Silva, 2012). A maioria desses trabalhos se limitaram à análise de séries de tempo para preços ofertados e aplicação de testes para distribuição probabilística (Padhi & Mohapatra, 2011).

No sentido de investigar a possível presença de manipulação de preços em licitações para construção de rodovias na cidade de Long Island nos Estados Unidos durante a década de 1980, Porter e Zona (1993) aplicaram um modelo de regressão linear para analisar a relação entre preços de oferta dos participantes dos certames e custos de execução dos projetos. Seus achados indicaram a presença de cartel, uma vez que o ranqueamento dos lances não coincidiu com o ranqueamento dos custos das empresas. Já Bajari e Ye (2003) desenvolveram um modelo teórico para identificar e testar a manipulação de lances em processos de licitação. Seus resultados mostraram as condições para a distribuição de lances em condições competitivas, comparando com equilíbrio em caso de conluio. Sob condições competitivas, os lances devem ser independentes após controlar todo conjunto de informações sobre custos observados pelas empresas. Já no caso de conluio, espera-se uma correlação entre os lances quando membros do cartel manipulam preços na tentativa de simular competitividade.

A partir de uma base de dados contendo mais de 340.000 transações no mercado de cimento da Alemanha, Hüschelrath e Veith (2008) investigaram a presença de cartéis. Estes autores empregaram uma análise de quebra estrutural na série de índice de preço juntamente com um modelo de regressão para controlar efeitos de custo e demanda sobre diferenciação de preços. Seus achados indicaram que consumidores que adquiriram cimento de grandes fornecedores pagaram, em média, um preço maior que aqueles que compraram de pequenas empresas. Em outro estudo para a Alemanha, Blanckenburg, Geist e Kholodilin (2012) analisaram a distribuição de variações de preços em série de tempo entre 1976 e 2009 para os 11 maiores cartéis industriais. A partir de testes estatísticos para diferença de distribuições de preços, suas evidências indicaram que 9 dos 11 casos estudados tiveram forte influência no controle de preços durante anos.

No Brasil, Balaniuk et al. (2012) usaram dados do TCU sobre contratos públicos e privados ao longo do período de 1997 e 2011. Eles aplicaram experimentos de aprendizado de máquina com o algoritmo Naïve Bayes de classificação, identificando 2.560 contratos com elevada probabilidade de fraude dentro de um conjunto de 795.954. Já Ralha e Silva (2012) empregaram um modelo multiagente com experimentos de aprendizado de máquina para investigar a presença de práticas de conluio a partir de dados de pregões eletrônicos do Governo Federal no período de 2005 e 2008 em todos os estados da federação, extraídos dos sistemas ComprasNet. Os achados indicaram *clusters* com fortes indicações de cartelização com casos específicos reforçados por regras de associação. Assim, os autores em destaque reforçam que, de acordo com a expertise dos especialistas na área, os rodízios em licitações públicas de âmbito nacional são, em geral, tipicamente regionais. Ou seja, seria pouco provável este tipo de prática, por parte de tais empresas, numa atuação recorrente e concomitante em todos os estados brasileiros.

3 Procedimentos metodológicos

Dado o objetivo geral deste estudo, o uso de ferramentas de aprendizado de máquina para a descoberta de Regras de Associação (RA) mostra-se pertinente, uma vez que possibilita identificação de casos suspeitos de fraude por meio de padrões atípicos entre empresas participantes de licitações. O algoritmo Apriori, desenvolvido por Agrawal e Srikant (1994), é um dos mais usados quando se trata de tarefas de aprendizado de máquina para descoberta de RA. Ele parte do seguinte princípio para reduzir o espaço de busca por RA: se um conjunto de itens não é frequente, todos os seus subconjuntos não podem ser frequentes (Lantz, 2013). Destaca-se que toda a parte de leitura, tratamento e processamento dos dados desta pesquisa foi realizada por meio do *software* R (R Core Team, 2020), em sua versão 4.0.0, usando em especial a biblioteca *arules* (Hahsler et al., 2011), versão 1.6-6, como pacote principal para identificação das regras de associação.

Se, por um lado, bases de dados que envolvem transações de itens costumam ser imensas, por outro, a exploração de associações relevantes entre tais itens envolve a identificação e a contabilização de inúmeros conjuntos ou subconjuntos que aparecem com alguma regularidade e que registram algum padrão de correlação. Nesse contexto, o número de potenciais conjuntos de itens que podem formar padrões distintos cresce exponencialmente com o total de características a serem analisadas. Tal natureza combinatória inviabiliza qualquer descoberta de conhecimento implícito aos dados por parte do ser humano. Apenas o uso de algoritmos com regras heurísticas para aprendizado de máquina é capaz de fornecer os resultados almejados.

Uma regra de associação (RA) formaliza padrões de relacionamento entre dois conjuntos de itens na forma de premissas e de resultado. Para fins de ilustração, considere-se a seguinte RA: $\{X\} \Rightarrow \{Y\}$, onde $\{X\}$ é o conjunto de premissas (antecedentes), e $\{Y\}$ o resultado condicionado a estas premissas (consequente). Para avaliar a qualidade de uma RA, o Quadro 1 apresenta algumas estatísticas básicas usadas nessa modelagem.

Quadro 1 - Estatísticas comumente usadas para mensurar a qualidade das regras de associação

Nome	Estatística	Descrição
Suporte	$supp(X) \equiv p(X) = \frac{n_x}{N}$	Frequência relativa de uma empresa X no total de certames (probabilidade incondicional).
Confiança	$conf(X \Rightarrow Y) \equiv p(Y X) = \frac{p(X,Y)}{p(X)}$	Probabilidade de a empresa Y participar do certame dado que a empresa X participa (probabilidade condicional).
Lift	$lift(X \Rightarrow Y) = \frac{supp(X,Y)}{supp(X) \cdot supp(Y)}$	Índice de desvio do suporte da RA em relação ao suporte esperado em caso de independência entre os conjuntos X e Y, onde $lift > 1$ indica elevado grau de associação.
Convicção	$conv(X \Rightarrow Y) = \frac{1 - supp(X,Y)}{1 - conf(X \Rightarrow Y)}$	Probabilidade de o conjunto X ser observado sem o conjunto Y, onde $conv = 1$ implica independência entre X e Y, ao passo que $conv > 1$ sugere forte dependência.

Fonte: McNicholas e Zhao (2009).

Legenda: n_x = total de vezes que o conjunto X é observado; N = total de transações na base de dados.

O algoritmo Apriori seleciona RA com base em parâmetros mínimos de suporte e de confiança para definição do espaço de busca. Por exemplo, em uma amostra de 5.000 licitações, a escolha de um suporte mínimo de 1% implica que uma empresa precisa aparecer ao menos em 50 certames para que seja identificada uma RA. Há, contudo, um dilema na escolha dos parâmetros mínimos para a aplicação do algoritmo Apriori. Valores de suporte muito altos podem não produzir RA (aumentando exponencialmente o número de subconjuntos de empresas) ou RA muito genéricas (grandes empresas fornecedoras que participam com frequência de certames), enquanto valores muito baixos podem gerar poucas RA (ignorando outros subconjuntos relevantes) ou RA não realistas (baixo indicador de confiança).

Conforme destaca Lantz (2013), um conjunto de dados pode ser reestruturado de forma transacional (matriz esparsa), em que cada linha representa um processo licitatório e cada coluna representa uma empresa. Assim, a matriz de transações é formada por m licitações e n empresas, de modo que o elemento/atributo a_{ij} recebe os seguintes valores binários:

$$a_{ij} = \begin{cases} 1, & \text{se a empresa } j \text{ participou da licitação } i \\ 0, & \text{se a empresa } j \text{ não participou da licitação } i. \end{cases} \quad (1)$$

Na primeira fase de iterações, o Apriori gera, para cada iteração i , um conjunto de todos os i subconjuntos que atendem o suporte mínimo da regra. Em seguida, todos os conjuntos de itens da iteração i são combinados para a próxima iteração $i + 1$, sendo eliminados conjuntos não frequentes. O procedimento é encerrado quando não é possível produzir novos conjuntos de dados (Lantz, 2013). Portanto, regras de associação com baixo suporte (caso atípico dos dados) e alta confiança (participação muito provável de uma empresa Z na presença de outras empresas X e Y) podem identificar padrões suspeitos de conluio entre empresas concorrentes.

Embora as estatísticas de suporte e de confiança sejam úteis para a seleção de RA de interesse, na prática, tais informações não são suficientes para assegurar que padrões cíclicos de empresas signifiquem comportamento colusivo. Nesse sentido, uma forma de selecionar RA mais fortes é o uso de funções de validação que consideram outras informações discriminantes de comportamentos inidôneos à luz da Teoria dos Jogos. Neste trabalho, faz-se uma extensão da função de avaliação proposta por Ralha e Silva

(2012), a fim de aumentar a robustez dos padrões de risco identificados.

Uma vez identificadas RA que atendam parâmetros mínimos de suporte e de confiança, faz-se uma avaliação delas a partir da seguinte função:

$$F(RA) = I(RA) \times M(RA), \quad (2)$$

em que

$$I(RA) = \begin{cases} 1, \text{ se } \frac{n(j)}{N} \times T < q(0,95) \forall j \in r \\ 0, \text{ caso contrário} \end{cases}$$

$$M(RA) = \begin{cases} 4 \text{ (risco alto), se } p_r > \bar{p} \text{ e } c_r > \bar{c} \\ 3 \text{ (risco médio), se } p_r > \bar{p} \text{ e } c_r \leq \bar{c} \\ 2 \text{ (risco baixo), se } p_r \leq \bar{p} \text{ e } c_r > \bar{c} \\ 1 \text{ (risco muito baixo), se } p_r \leq \bar{p} \text{ e } c_r \leq \bar{c} \end{cases}$$

O termo $I(RA)$ é uma função indicadora que filtra RA cujos membros não são proponentes frequentes, isto é, assume o valor 1 se, e somente se, o total de licitações que a empresa j da RA participou $-\frac{n_j}{N} \times T$ – for inferior ao quantil $q_{(0,95)}$ da distribuição de participação anual de fornecedores e 0 caso contrário, onde $\frac{n_j}{N}$ é o suporte da empresa j – frequência relativa – e T o total de licitações na matriz de transações. Esse termo procura, portanto, eliminar padrões de associação casuais que podem surgir em razão do envolvimento de grandes fornecedores. Já o termo $M(RA)$ mapeia zonas de risco de fraude para RA filtradas, ponderando as RA pela chance de vitória de ao menos uma empresa quando da atuação conjunta do grupo suspeito (jogo cooperativo) e pela presença de outros concorrentes das empresas identificadas na regra (restrição de oferta ou simulação de concorrência). Note-se que os valores 1, 2, 3 e 4 indicam, respectivamente, os seguintes níveis de risco de conluio: muito baixo, baixo, médio e alto; p_r é a probabilidade de ao menos uma empresa da RA vencer uma licitação quando todo grupo envolvido concorre; c_r é o número médio de concorrentes dos membros da RA em licitações; \bar{p} e \bar{c} são parâmetros limiares (críticos) de probabilidade do rodízio induzir vitória e de média de concorrentes¹. Conforme Ralha e Silva (2012), a probabilidade $p_r = v_r/n_r$ é dada pela razão entre o número de vezes que algum fornecedor do grupo r venceu licitações em que todo o grupo participou (v_r) e o total de licitações envolvendo participação conjunta do grupo (n_r).

A zona de risco alto de conluio é caracterizada por RA cujo grupo de proponentes registra elevada probabilidade de formalizar contrato com unidades gestoras municipais e tem concorrentes acima da média geral. Ou seja, padrões cíclicos de participação conjunta e com forte propensão de vitórias entre empresas são consistentes com resultados esperados em jogos repetidos, nos quais os jogadores (empresas) tem incentivos para adotar ações cooperativas. Na zona de risco médio de conluio encontram-se RA em que as empresas têm alta probabilidade de vitória e possuem concorrentes aquém da média geral. As regras mapeadas na chamada zona de risco baixo de conluio costumam registrar alta média de concorrentes e baixa probabilidade de vitória de membro da RA (baixa probabilidade de ações cooperativas). Finalmente, as RA classificadas na zona de risco muito baixo de conluio incluem regras (grupos) que combinam baixo número de concorrentes e baixa probabilidade de vitória de algum membro, sendo pouco provável a sustentação da hipótese de conluio.

No intuito de ranquear empresas envolvidas em RA selecionadas e classificadas conforme a equação (2), é calculado o Indicador de Suspeição da Empresa² (ISE) a partir de três dimensões: a) pontuação pela recorrência da empresa em grupos com alta associação de membros e tendenciosidade de vitória; b) pontuação pela recorrência da empresa em grupos com indicações de simulação de concorrência; e c) pontuação pela recorrência da empresa em grupos com atuação concentrada em determinados municípios. Formalmente:

¹ Embora sejam parâmetros arbitrários de delineamento de espaços de solução, neste estudo emprega-se o limiar de risco de jogo cooperativo $\bar{p} = 0,50$ e o limiar de concorrência \bar{c} igual a média histórica de proponentes por licitação.

² Contribuição inédita desta pesquisa.

$$ISE_j = \overline{conf}(r_j^i) \times \left\{ \sum_{i=1}^I M(r_j^i) p_r^i + \sum_{k=1}^K M(r_j^k) p_r^k + \sum_{l=1}^L M(r_j^l) p_r^l \right\}, \quad (3)$$

Onde $j = 1, \dots, J$ indexa cada empresa na matriz de transação; $i = 1, \dots, I$ indexa cada RA em que a empresa j se faz presente; $\overline{conf}(r_j^i) = \sum_{i=1}^I conf(r_j^i) I^{-1}$ é a média da estatística de confiança da regra de associação r em que a empresa j se faz presente, considerando um total de I grupos distintos, em que $conf(\cdot)$ é uma probabilidade condicional que mensura a probabilidade de uma empresa participar do certame dada a participação de outra; $k = 1, \dots, K$ indexa cada RA em que a empresa j aparece junto com ao menos uma empresa h tipicamente perdedora, inclusive para $h = j$ (RA com simulação de concorrência); $l = 1, \dots, L$ indexa cada RA em que a empresa j participou e cujas disputas foram todas concentradas em um único município³ (RA com concentração regional); $M(\cdot)$ é uma função que atribui pesos crescentes para cada regra de associação segundo sua posição nos quadrantes de risco de conluio e p_r é a probabilidade de uma empresa envolvida na regra r vencer um certame.

Na equação (3), o termo $\sum_{i=1}^I M(r_j^i) p_r^i$ acumula pontuação para a empresa j conforme seu envolvimento em diferentes RA; $\sum_{k=1}^K M(r_j^k) p_r^k$ soma pontos para a empresa j segundo sua recorrência em RA (grupos) que incluem fornecedores quase sempre perdedores (no máximo uma vitória) nos certames em que o grupo r participou e $\sum_{l=1}^L M(r_j^l) p_r^l$ agrega pontos conforme o total de casos em que a empresa j esteve envolvida em grupos cujas atuações sempre ocorreram em um mesmo município. Note-se que, em cada uma das somatórias de recorrências, os pontos são crescentes segundo a posição da RA nos quadrantes de risco de conluio e grau de tendenciosidade – probabilidade de vitória de uma empresa do grupo. Ao final, a somatória total de pontos é ponderada pela estatística de confiança (média) da(s) RA em que o fornecedor j participou⁴.

3.1 Dados

Os dados empregados na análise empírica desta pesquisa foram coletados junto ao TCE-PB, incluindo informações de licitações municipais (data, modalidade, descrição de objeto não estruturada) e empresas participantes (inclusive número de Cadastro de Pessoa Física – CPF e Cadastro de Pessoa Jurídica – CNPJ), compreendendo o período de janeiro de 2005 até setembro de 2016.

Cada licitação foi unicamente identificada a partir da concatenação dos seguintes campos presentes na base de dados: número da licitação (por exemplo, 122010 – licitação 12 do ano de 2010); código da modalidade de licitação (número inteiro entre 0 e 13, correspondente a cada modalidade presente na base de dados: Adesão a Registro de Preço, Chamada Pública, Concorrência, Convite, Dispensa por outros motivos, Dispensa por Valor Inexigível, Leilão, Pregão Eletrônico, Pregão Eletrônico e Presencial, Pregão Presencial, Sem Licitação e Tomada de Preços); código da unidade gestora (por exemplo, 101095 é código da Prefeitura Municipal de João Pessoa).

No intuito de não se explicitar nos resultados do artigo os dados de empresas, foi gerado um novo código sequencial (único) e correspondente a cada CPF/CNPJ. A amostra selecionada consiste em certames pertencentes às modalidades Pregão Presencial, Pregão Eletrônico, Carta Convite, Tomada de Preços e Concorrência de todas as unidades gestoras municipais. As licitações dessas modalidades correspondem a 73,5% de todos os registros históricos da base de dados. Portanto, a amostra final é composta por 103.086 licitações e de um total de 280.274 ofertas.

Os períodos selecionados para formação das matrizes de transações correspondem a ciclos de gestão municipal: (a) período 2005–2008; (b) período 2009–2012; e (c) período 2013–2016. Destarte, foram criados quatro conjuntos de dados na forma transacional (matriz esparsa), cujas linhas incluem os códigos de licitações, as colunas códigos de empresas, e as entradas valores binários (1 caso o proponente participou do certame, 0 caso contrário).

No tocante aos recortes temporais acima mencionados, vale ressaltar que tal estratégia procura

³ Todas as medidas estatísticas de concentração foram calculadas por meio do coeficiente de Gini (Hahsler, 2011), que estabelece limiares entre 0 e 1 para avaliação da distribuição dos dados, onde valores iguais a 0 indicam distribuição equitativa dos dados e valores iguais a 1 indicam perfeita concentração.

⁴ O indicador ISE_j pode ser normalizado pela sua importância na soma de todos os valores para o conjunto de empresas, assegurando sua escala de 0 a 100%.

melhor definir espaços de busca para a construção de RA, uma vez que diferenças de perfis de gestão podem afetar o volume ofertado de licitações, os tipos de bens e serviços licitados e os grupos de empresas concorrentes (demandas seletivas), aumentando a probabilidade de se encontrar RA mais significativas. Não obstante, RA mapeadas nesses recortes temporais podem ser melhor relacionadas a padrões espaciais de concentração, levantando indicações de possíveis relações de práticas de conluís com os ciclos políticos municipais.

Vale ressaltar que, no intuito de assegurar a comparação de resultados em todos os experimentos, o algoritmo Apriori considerou os respectivos parâmetros mínimos de suporte e de confiança⁵, 0,02% e 80%, assim como licitações com, no mínimo, 2 e, no máximo, 10 proponentes para a formação de RA. Tais parâmetros permitem a identificação de um considerável número de RA, possibilitando uma aplicação mais ampla da função de avaliação – equação (2) – na filtragem de resultados de interesse e, assim, capturando com maior chance indicadores de irregularidades. A Tabela 1 resume os conjuntos de dados transacionais selecionados para experimentos de aprendizado de máquina segundo períodos de gestão municipal.

Tabela 1 – Dados selecionados para experimentos de mineração por período de gestão municipal

Período	Ofertas	Proponentes	Proponentes (vitória em ao menos 2 licitações)	Proponentes (vitória em ao menos 5 licitações)	Licitações
2005-2008	96.666	19.667	7.678	1.827	30.381
2009-2012	103.364	22.162	9.472	2.356	35.747
2013-2016	80.244	17.801	8.557	2.301	36.958
Total	280.274	59.630	25.707	6.484	103.086

Fonte: Elaboração própria a partir de dados do TCE-PB.

Nota: Ofertas são referentes ao total de registros de lances por licitação na base de dados.

Assim, o mesmo proponente pode ofertar em diferentes licitações.

Conforme pode ser observado, 43% das empresas participantes dos processos licitatórios em todo o período considerado venceram ao menos 2 certames. Quando se consideram apenas empresas vencedoras de ao menos 5 licitações, tal percentual diminui para 11%.

4 Resultados

Esta seção apresenta os resultados gerados a partir de experimentos de aprendizado de máquina no intuito de se detectar grupos de empresas suspeitas de fraudes em licitações municipais no estado da Paraíba durante o período de 2005 a 2016. A análise é direcionada para a identificação de fornecedores cujos comportamentos são consistentes com ações anticompetitivas (colusivas), em particular: grupos com padrões cíclicos de participação/vitória; empresas presentes em diferentes grupos com padrões de associação (rodízios); padrões de associação que caracterizam simulação de concorrência e/ou concentração geográfica. A maior parte dos resultados está organizada por três períodos de gestão municipal: 2005–2008, 2009–2012 e 2013–2016.

4.1 Experimentos segundo ciclos políticos

A Tabela 2 informa o total de RA encontradas e selecionadas segundo períodos de análise e os valores médios de estatísticas de suporte, confiança, *lift* e convicção para as regras selecionadas. Estas estatísticas medem, neste estudo, a frequência relativa, o grau de associação entre empresas e a independência entre antecedentes e consequentes. Considerando todo o período, de 2005 a 2016, o algoritmo Apriori mapeou um total de 1.439 RA, sendo o intervalo de 2009-2012 aquele com maior registro de regras (653 casos ou 45,4% do total). Os critérios iniciais usados para formação desses padrões basearam-se, particularmente, nas estatísticas de suporte (ao menos 0,02%) e de confiança (ao menos 80%).

Tabela 2 – Estatísticas descritivas das regras de associação descobertas por períodos de gestão municipal

⁵ A escolha dos valores de suporte e confiança neste estudo foi feita com base na experimentação do modelo para diferentes limiares dessas estatísticas, cuja seleção se deu por meio da observação empírica dos diferentes conjuntos de regras identificadas e do julgamento dos resultados através da experiência de um corpo técnico de auditores de contas públicas. Nesse cenário, foi decidido, após análises exploratórias e observação dos resultados, para o caso das licitações municipais da Paraíba, que um maior nível de confiança (pelo menos 80%) e um nível de suporte de, pelo menos, 0,02% produziram uma amostra satisfatória de conjuntos de regras para serem avaliados, posteriormente, pela função de avaliação expressa na equação (2).

Conjunto de dados	Regras descobertas	Regras selecionadas	Média - regras selecionadas			
			Suporte %	Confiança %	Lift	Convicção
2005-2008	526	155	0,0323	92,0710	1.657,7358	8,0719
2009-2012	653	81	0,0300	92,8185	1.735,3286	8,9599
2013-2016	260	34	0,0292	92,5911	1.894,6878	8,1846

Fonte: Elaboração própria a partir de dados do TCE-PB.

Nota: Suporte mínimo: 0,02%. Confiança mínima: 80%.

Considerando o total de 1.439 RA descobertas a partir de algoritmo Apriori segundo os períodos representativos de ciclos políticos municipais, o recorte seguinte foi baseado na função de avaliação de risco de conluio da equação (2). Com base nessa função de avaliação, assegurou-se a sobrevivência de RA cujos membros não são proponentes frequentes, a fim de eliminar riscos de padrões de associação casuísticas em decorrência de grandes fornecedores, resultando na seguinte distribuição: 155 RA para o período 2005-2008, 81 RA para o período 2009-2012; e 34 RA para o período 2013-2016, ou seja, respectivamente 29%, 12% e 13%.

Os dados indicam que as regras selecionadas nos experimentos apresentaram, em média, baixos valores de suporte e altos valores para as estatísticas de confiança, *lift* e convicção. Os valores altos para a estatística de confiança informam que as regras descobertas possuem elevada probabilidade de ocorrência do conseqüente/resultado, dado o conjunto de antecedentes/premissa. Igualmente, no que diz respeito aos valores médios da estatística *lift*, foram constatados valores muito superiores à unidade, indicando elevado grau de dependência. Ressaltem-se, ainda, os valores também elevados para a estatística de convicção (superiores a 1), os quais sugerem uma forte relação de dependência entre empresas envolvidas nas regras de associação.

A Tabela 3 apresenta a distribuição das regras filtradas por zonas de risco de conluio conforme a interação entre indicadores de probabilidade de vitória de, ao menos, um membro do grupo de empresas e de média de concorrentes das firmas envolvidas nos grupos.

Tabela 3 – Distribuição de regras de associação selecionadas segundo períodos de gestão municipal e por faixas de risco de conluio

Período	Total de regras por risco de fraude				Regras selecionadas
	Alto	Médio	Baixo	Muito Baixo	
2005-2008	9	111	5	30	155
	5,8%	71,6%	3,2%	19,4%	100%
2009-2012	5	64	0	12	81
	6,2%	79,0%	0,0%	14,8%	100%
2013-2016	20	8	4	2	34
	58,8%	23,5%	11,8%	5,9%	100%

Fonte: Elaboração própria a partir de dados do TCE-PB.

Os achados dispostos na Tabela 3 são sugestivos no tocante às gradações de risco de conluio. É possível se depreender que a redução da quantidade de regras de associação encontradas no período de gestão de 2013-2016, em relação aos demais, não significou uma redução proporcional de regras fortes com alto risco de conluio. Nesse aspecto, percebe-se uma maior incidência relativa de RA com alto risco para o período de gestão municipal mais recente (2013-2016), ou seja, 20 regras num total de 34, o que representa 58,8% das regras, em comparação com 5,8% no período de 2005-2008 e 6,2% referente ao período 2009-2012.

Em linhas gerais, também vale observar que a sólida maioria das regras selecionadas a partir da função de avaliação se concentraram entre os níveis de risco alto e médio de comportamento colusivo, cujo montante percentual ultrapassou a faixa de 78% em todos os experimentos realizados. Com exceção do período 2013-2016, nos demais percebe-se que cerca de 71% a 79% das RA foram classificadas com risco médio de conluio. Doravante são analisadas apenas as regras de associação classificadas com alto e médio risco de conluio, cujas elevadas probabilidades de vitória são consistentes com a hipótese de jogo cooperativo.

4.2 Padrões suspeitos em ambientes de maior concorrência

A Tabela 4 apresenta os grupos de empresas⁶ identificados com alto risco de conluio nos períodos 2005–2008, 2009–2012 e 2013–2016. São destacadas as estatísticas de suporte, confiança, probabilidade de a associação induzir vitória de ao menos uma firma, média de concorrentes das empresas suspeitas⁷ em todo o conjunto de licitações (independentemente de atuarem em grupo ou não) e média de concorrentes nas licitações quando o grupo suspeito atua em conjunto.

Tabela 4 – Regras de associação com alto risco de conluio segundo períodos de análise

RA	Suporte %	Confiança %	Probabilidade de vitória % (forne- cedor)	Média de concorrentes (forne- dores)	Média de concorrentes (grupo)
Período 2005–2008					
{57617} => {43040}	0,0296	100,00	100,00	2,8	1,0
{57617} => {40290}	0,0263	88,89	100,00	3,2	1,0
{36846} => {954}	0,0230	87,50	100,00	3,5	3,4
{36846} => {42216}	0,0230	87,50	100,00	3,2	3,4
{36846} => {40400}	0,0263	100,00	100,00	3,2	3,2
{954} => {40400}	0,0296	100,00	100,00	3,1	3,3
{42216} => {40400}	0,0296	81,82	100,00	2,9	3,3
{41903} => {40274}	0,0362	91,67	100,00	2,8	3,4
{23238} => {4250}	0,0230	87,50	100,00	6,3	6,3
Período 2009–2012					
{57474} => {47578}	0,0280	83,33	90,00	2,7	3,4
{991} => {27028}	0,0252	90,00	77,78	3,9	4,6
{16291} => {39616}	0,0224	100,00	100,00	4,9	7,1
{41448} => {39616}	0,0224	88,89	100,00	4,5	7,1
{40495} => {46014}	0,0392	100,00	100,00	4,5	5,7
Período 2013–2016					
{24966} => {41448}	0,0216	100,00	100,00	4,9	3,1
{49798} => {41448}	0,0216	100,00	100,00	4,9	3,1
{40761} => {41448}	0,0216	100,00	100,00	4,9	7,0
{16291} => {41448}	0,0216	100,00	100,00	4,9	7,0
{46847} => {43678}	0,0352	86,67	92,31	3,1	3,5
{46847} => {40471}	0,0325	80,00	100,00	3,0	3,8
{2254} => {41135}	0,0216	100,00	100,00	6,2	8,4
{16243} => {35753}	0,0244	90,00	100,00	9,8	9,9
{23719} => {18254}	0,0216	80,00	100,00	9,8	10,9
{23719} => {35753}	0,0216	80,00	100,00	10,9	11,0
{38258} => {40143}	0,0271	83,33	100,00	8,3	9,9
{38258} => {37953}	0,0271	83,33	100,00	9,8	10,5
{38258} => {49250}	0,0325	100,00	100,00	9,0	9,6
{38258} => {40423}	0,0325	100,00	100,00	8,0	9,6
{41000} => {40071}	0,0460	89,47	100,00	5,8	5,6
{40143} => {49250}	0,0379	82,35	100,00	8,3	9,1
{40143} => {40423}	0,0406	88,24	100,00	7,6	8,6
{37953} => {49250}	0,0433	94,12	100,00	8,7	10,3
{37953} => {40423}	0,0433	94,12	100,00	7,9	10,0
{49250} => {40423}	0,0622	85,19	91,30	7,4	9,2

Fonte: Elaboração própria a partir de dados do TCE-PB.

Nota: Suporte mínimo: 0,02%. Confiança mínima: 80%.

Dentre as RA selecionadas, verificam-se baixos valores da estatística de suporte. Por exemplo, nos experimentos realizados com os conjuntos de dados para os períodos de 2005–2008, 2009–2012 e 2013–2016, os valores de suporte situam-se no intervalo de 0,02% a 0,06%. Tais valores são muito próximos ao limiar mínimo de suporte usado para definir a construção de RA pelo algoritmo Apriori

⁶ Para garantir o sigilo da informação, empresas receberam um código numérico gerado por algoritmo, apenas a título de diferenciá-las entre si. Tais códigos correspondem às sequências numéricas que estão entre chaves e se encontram dispostos nas tabelas a seguir, no campo (coluna) denominado de “RA” ou “Grupo suspeito”.

⁷ O termo ‘suspeito’ é usado apenas para indicar que a empresa forma uma RA com risco de conluio, cuja identificação efetiva da suspeição de fraude requer análises complementares e qualitativas do processo licitatório.

(0,02%), sugerindo que as RA mapeadas não são suficientemente frequentes nos conjuntos de dados para caracterizar padrões genéricos.

No tocante aos valores da estatística de confiança (probabilidade condicionada), cabe notar que as RA identificadas registram, em geral, patamares superiores ao parâmetro mínimo de 80% considerado pelo algoritmo de aprendizado de máquina. Note-se que os resultados encontrados para o período 2005–2008 indicam que 8 RA, dentre as 9 selecionadas, registram estatística de confiança superior a 87%, isto é, uma elevada chance (probabilidade) de a empresa consequente (lado direito) participar do certame quando as antecedentes (lado esquerdo) estiverem presentes. Já quando se observam as RA filtradas para o período 2009–2012, percebe-se que 4 das 5 RA mapeadas têm confiança igual ou superior a 88%, enquanto no período mais recente (2013–2016), de 20 RA descobertas, 10 apresentaram confiança igual ou acima de 90%. Destarte, em linhas gerais, as RA selecionadas por períodos de gestão municipal registram alto padrão de associação entre empresas.

Outro aspecto relevante nos dados registrados na Tabela 4 em destaque são os elevados valores de probabilidade de vitória de membros de grupos suspeitos em licitações, indicador sugestivo de possíveis rodízios na repartição de valores contratuais. Note-se que, independentemente do período analisado, a grande maioria das RA mapeadas indicam grupos cujos membros têm probabilidade superior a 90% de fechar contrato com órgãos municipais. Note-se que, do total de RA registradas, apenas 4 não apresentaram probabilidade de vitória de 100%.

No experimento de aprendizado de máquina realizado para o período 2005–2008, os fornecedores mais recorrentes nas regras de associação de alto risco de conluio foram as empresas: a) 40400 e 36846, envolvidas em três rodízios, sendo um deles com participação das duas e outros associados com os fornecedores 954 e 42216; e b) a empresa 57617 – também presente em dois outros rodízios – cuja participação condicionou a presença das empresas 40290 e 43040. Em geral, essas empresas participaram de licitações cujo objeto foi a aquisição de gêneros alimentícios direcionados a escolas, e os contratos firmados por elas envolveram cerca de R\$ 5,1 milhões⁸.

Os resultados obtidos para o conjunto de dados referentes ao período 2009–2012 também apontaram suspeitas de práticas anticompetitivas. Cabe destacar a empresa 39616 – mapeada em duas RA – cuja participação está fortemente relacionada à presença das firmas 16291 e 41448. Nos certames vencidos pelas empresas aqui identificadas, o valor total ofertado chegou a R\$ 1,7 milhões para o fornecimento de gêneros alimentícios e material de limpeza, merenda, gás de cozinha e carne e hortifrutigranjeiros para hospitais, creches e escolas. As regras de associação mapeadas para o período 2013–2016 permitem identificar alguns fornecedores que participaram de diferentes relações: as empresas 38258, 40423, 41448 e 49250 – presentes em quatro RA; 37953 e 40143 em três RA; 23719 e 46847 em duas RA. Vale notar que, nesses casos, além da participação de uma empresa estar relacionada à presença de outras, quando elas atuam em conjunto a probabilidade de vitória de ao menos uma é próxima a 100%.

4.3 Evidências de simulação de concorrência

Nesta subseção, faz-se uma exploração de casos suspeitos de simulação de concorrência a partir da filtragem de padrões de associação envolvendo empresas que recorrentemente não vencem licitações, considerando todos os experimentos realizados e diferentes classificações de risco de conluio. A Tabela 5 registra RA cuja probabilidade de vitória de ao menos um participante é de 80% ou mais (associação tendenciosa) – RA de médio ou alto risco de conluio – e que há ao menos um fornecedor que nunca venceu ou que venceu apenas um certame quando das atuações conjuntas do grupo suspeito. Em geral, apenas são reportados os casos mais recorrentes, isto é, aqueles em que os grupos participaram conjuntamente de 10 ou mais certames.

Conforme pode ser observado, nas RA elencadas há sempre uma empresa que tem sua participação muito associada à outra, porém nunca venceu ou apenas venceu um certame quando o grupo atuou em conjunto. A maioria dos casos suspeitos foi detectada nas tarefas de aprendizado de máquina para os conjuntos de dados referentes aos ciclos políticos de 2005–2008 e 2009–2012. No experimento realizado para o período de 2013–2016, apenas um grupo foi detectado.

Para o período de gestão municipal de 2005–2008, tem-se um total de 11 regras identificadas. A

⁸ Doravante cabe ressaltar que todos os valores monetários citados estão a preços constantes de 2016, corrigidos pelo Índice Geral de Preços-Disponibilidade Interna (IGP-DI).

fim de obter uma melhor compreensão dos resultados da Tabela 5, exemplifica-se o caso da empresa 39961, que teve sua participação em certames associada às empresas 41502 e 44294. Isto pode ser observado em duas RA distintas, a 6ª e a 10ª regra do mencionado período. Essas empresas participaram de licitações para aquisição de gêneros alimentícios, verduras, carnes e ovos, destinados à merenda escolar da rede municipal, creches e hospitais. Cabe destacar que, em 11 certames nos quais 39961 e 41502 atuaram juntas, a primeira venceu em 10 ocasiões, e a segunda perdeu todas. Já em 10 licitações que 39961 e 44294 participaram conjuntamente, 39961 venceu todas. Ademais, quando se comparam as participações totais e vitórias (histórico de cada fornecedor), percebe-se que 41502 participou de outras 4, perdendo todas; 44294 atuou em mais 10 licitações, vencendo apenas em duas oportunidades; e 39961, em duas outras atuações, não venceu nenhuma delas.

Tabela 5 – Regras de associação com médio e alto risco de conluio e indicadores de simulação de concorrência

Grupo	Atuação conjunta do grupo		Histórico individual	
	Licitações (Total)	Vitórias de cada fornecedor	Participações de cada fornecedor	Vitórias de cada fornecedor
2005-2008				
{57206,58039}	18	57206 (0), 58039 (15)	57206 (21), 58039 (23)	57206 (2), 58039 (16)
{33978,32547}	12	33978 (0), 32547 (12)	33978 (12), 32547 (26)	33978 (0), 32547 (22)
{44662,45977}	11	44662 (0), 45977 (11)	44662 (11), 45977 (15)	44662 (0), 45977 (12)
{45861,46442}	11	45861 (0), 46442 (9)	45861 (11), 46442 (17)	45861 (0), 46442 (13)
{42580,58536}	11	42580 (9), 58536 (1)	42580 (12), 58536 (33)	42580 (10), 58536 (11)
{39961,41502}	11	39961 (10), 41502 (0)	39961 (12), 41502 (15)	39961 (10), 41502 (0)
{21248,15770}	11	21248 (0), 15770 (10)	21248 (12), 15770 (25)	21248 (0), 15770 (16)
{42510,48724}	10	42510 (1), 48724 (7)	42510 (10), 48724 (25)	42510 (1), 48724 (14)
{43160,40491}	10	43160 (0), 40491 (8)	43160 (10), 40491 (31)	43160 (0), 40491 (16)
{39961,44294}	10	39961 (10), 44294 (0)	39961 (12), 44294 (20)	39961 (10), 44294 (2)
{38103,38751}	10	38103 (0), 38751 (10)	38103 (12), 38751 (32)	38103 (2), 38751 (23)
2009-2012				
{48959,58082}	24	48959 (1), 58082 (23)	48959 (25), 58082 (32)	48959 (2), 58082 (29)
{48382,58082}	21	48382 (1), 58082 (20)	48382 (22), 58082 (32)	48382 (1), 58082 (29)
{47521,50488}	12	47521 (11), 50488 (0)	47521 (12), 50488 (27)	47521 (11), 50488 (0)
{57072,58548}	12	57072 (0), 58548 (11)	57072 (15), 58548 (25)	57072 (0), 58548 (16)
{45659,50988}	11	45659 (1), 50988 (11)	45659 (11), 50988 (20)	45659 (1), 50988 (16)
{939,49737}	11	939 (1), 49737 (10)	939 (12), 49737 (22)	939 (1), 49737 (19)
{11073,20246}	10	11073 (0), 20246 (10)	11073 (10), 20246 (19)	11073 (0), 20246 (19)
{41124,44919}	10	41124 (0), 44919 (10)	41124 (11), 44919 (20)	41124 (0), 44919 (19)
{51165,53523}	10	51165 (0), 53523 (10)	51165 (11), 53523 (25)	51165 (0), 53523 (15)
2013-2016				
{50716,48447}	8	50716 (0), 48447 (8)	50716 (8), 48447 (13)	50716 (0), 48447 (13)

Fonte: Elaboração própria a partir de dados do TCE-PB.

Notas: Apenas RA com probabilidade de vitória de igual ou superior a 80% e com maior frequência de atuação conjunta do grupo suspeito. Entre parênteses: Total – Número de licitações em que o grupo suspeito atuou em conjunto. Vitórias (atuação conjunta do grupo) – Número de vitórias de cada fornecedor condicionado à participação conjunta do grupo suspeito. Participações (histórico individual) – Número de licitações de que o fornecedor participou (inclusive em atuação fora do grupo). Vitórias (histórico individual) – Número de vitórias de cada fornecedor (inclusive em atuação fora do grupo). Em certames loteados, a soma do total de vitórias de cada empresa pode ultrapassar o total de certames.

Já com relação ao período de 2009–2012, uma das regras que despertaram mais atenção diz respeito ao caso das firmas 48959 e 58082, que disputaram em conjunto 24 licitações para fornecimento de material esportivo com valores contratados de R\$ 2,3 milhões. A firma 48959 venceu apenas uma delas, enquanto 58082 saiu vitoriosa em 23 ocasiões. Todavia, quando ambas concorreram abertamente com outras empresas, 48959 venceu uma nova licitação, e 58082 conquistou 6 das outras 8 licitações que disputou. No período de 2013 a 2016, a única RA encontrada envolve as empresas 50716 e 48447. Ao participarem conjuntamente de 8 certames não loteados sobre a aquisição de móveis de escritório, 48447 venceu todos eles. Salienta-se que estes certames foram as únicas licitações que a firma 50716 disputou, tendo perdido todas. Todavia, quando 48447 disputou mais 5 licitações com outras empresas conseguiu vencer todas elas novamente. Logo, os resultados sugerem que a firma 50716 está recorrentemente simulando concorrência e influenciando intensamente a participação de 48447, que ganhou todas as licitações de que participou.

4.4 Evidências de concentração regional

A Tabela 6, a seguir, apresenta resultados de RA selecionados segundo o grau de concentração de atuações em determinados municípios paraibanos. São registrados apenas os casos em que a associação tem probabilidade de vitória igual ou acima de 80% e há concentração de 100% de atuações em determinados municípios, isto é, casos em que todas as participações conjuntas do grupo suspeito ocorreram nas cidades adiante relacionadas.

Tabela 6 – Regras de associação com indicações de potencial concentração regional – casos selecionados

Grupo	Licitações (Total)	Município com 100% de concentração	Vitórias no município	Distribuição de vitórias	Risco de conluio
2005-2008					
{41448,58011}	8	Brejo do Cruz	87,50	41448 (7), 58011 (4)	Médio
{42061,50427}	7	Campina Grande	100,00	42061 (1), 50427 (7)	Médio
{23238,4250}	7	Catolé do Rocha	100,00	23238 (7), 4250 (4)	Alto
{15909,4163}	7	Gado Bravo	100,00	15909 (4), 4163 (3)	Médio
{32254,32643}	7	Guarabira	100,00	32254 (2), 32643 (7)	Médio
{52350,48775}	8	Guarabira	100,00	52350 (8), 48775 (7)	Médio
{36846,954}	7	Marizópolis	100,00	36846 (7), 954 (7)	Alto
{36846,42216}	7	Marizópolis	100,00	36846 (7), 42216 (7)	Alto
{36846,40400}	8	Marizópolis	100,00	36846 (8), 40400 (7)	Alto
{954,40400}	9	Marizópolis	100,00	954 (9), 40400 (9)	Alto
{42216,40400}	9	Marizópolis	100,00	42216 (9), 40400 (9)	Alto
{57617,40290}	8	Paulista	100,00	57617 (8), 40290 (0)	Alto
{34104,29305}	7	Pirpirituba	85,71	34104 (3), 29305 (3)	Médio
{52354,48244}	7	Santa Rita	100,00	52354 (1), 48244 (7)	Médio
{37909,40562}	7	São J. do Rio do Peixe	85,71	37909 (5), 40562 (1)	Médio
{41903,40274}	11	Sousa	100,00	41903 (11), 40274 (11)	Alto
{41903,58664}	10	Sousa	100,00	41903 (9), 58664 (10)	Médio
2009-2012					
{16291,39616}	8	Brejo do Cruz	100,00	16291 (8), 39616 (8)	Alto
{41448,39616}	8	Brejo do Cruz	100,00	41448 (8), 39616 (8)	Alto
{57505,48751}	10	Caiçara	80,00	57505 (4), 48751 (4)	Médio
{56960,58597}	9	Campina Grande	100,00	56960 (9), 58597 (0)	Médio
{57474,47578}	10	Catolé do Rocha	90,00	57474 (7), 47578 (8)	Alto
{52350,41515}	16	Guarabira	100,00	52350 (16), 41515 (16)	Médio
{52350,48775}	18	Guarabira	100,00	52350 (18), 48775 (18)	Médio
{58651,44735}	8	Manáira	100,00	58651 (0), 44735 (8)	Médio
{50667,49618}	10	Nova Floresta	100,00	50667 (9), 49618 (10)	Médio
{42478,42174}	9	Paulista	100,00	42478 (9), 42174 (8)	Médio
{40495,46014}	14	Sousa	100,00	40495 (14), 46014 (13)	Alto
2013-2016					
{24966,41448}	8	Belém do Brejo do Cruz	100,00	24966 (8), 41448 (8)	Alto
{49798,41448}	8	Belém do Brejo do Cruz	100,00	49798 (8), 41448 (8)	Alto
{40761,41448}	8	Brejo do Cruz	100,00	40761 (8), 41448 (8)	Alto
{16291,41448}	8	Brejo do Cruz	100,00	16291 (8), 41448 (8)	Alto
{46847,43678}	13	Catolé do Rocha	92,31	46847 (12), 43678 (11)	Alto
{46847,40471}	12	Catolé do Rocha	100,00	46847 (11), 40471 (11)	Alto
{16243,35753}	9	Monteiro	100,00	16243 (9), 35753 (9)	Alto
{23719,18254}	8	Monteiro	100,00	23719 (8), 18254 (8)	Alto
{23719,35753}	8	Monteiro	100,00	23719 (8), 35753 (8)	Alto
{51030,49618}	8	Nova Floresta	100,00	51030 (2), 49618 (8)	Médio
{2254,41135}	8	Rio Tinto	100,00	2254 (8), 41135 (8)	Alto
{57956,38245}	9	São Domingos do Cariri	100,00	57956 (5), 38245 (4)	Médio

Fonte: Elaboração própria a partir de dados do TCE-PB.

Notas: Apenas RA com probabilidade de vitória igual ou superior a 80% e taxa de concentração de 100%. Entre parênteses: Vitórias (histórico individual) – Número de vitórias de cada fornecedor (inclusive em atuação fora do grupo). Em certames loteados, a soma do total de vitórias de cada membro pode ultrapassar o total de certames.

Considerando o período de gestão política entre 2005 e 2008, constata-se uma concentração regional predominante no sertão da Paraíba, com algumas poucas exceções, vejamos: Santa Rita (região da Mata Paraibana); Campina Grande, Gado Bravo, Guarabira e Pirpirituba (região do Agreste); e Brejo do Cruz, Catolé do Rocha, Marizópolis, Paulista, São João do Rio do Peixe e Sousa (região do Sertão).

Os dados mostram que os grupos suspeitos formados pelas empresas 40400, 36846, 954 e 42216, além da sua classificação de alto risco de conluio, apresentam grande concentração regional na região do sertão paraibano, em especial no município de Marizópolis, para o período de gestão de 2005–2008. São empresas que se revezam muito entre si quando da atuação em disputas licitatórias nesta edilidade. Vale ressaltar que, a despeito de a distribuição de vitórias nas licitações ser bem equilibrada entre essas empresas, mesmo que atuando alternadamente em grupos diferentes, não afasta a possibilidade de acertos escusos e alternância de vitórias, como uma espécie de loteamento de contratos com a administração pública.

Diferentemente do que aconteceu com alguns outros grupos deste mesmo período de gestão (2005–2008), a regra formada pelas empresas 42061 e 50427, cuja concentração regional se deu em Campina Grande, apresenta certa atipicidade de padrão, pelo fato de que 42061 venceu apenas 1 das 7 licitações que disputou em conjunto com 50427, enquanto esta saiu vitoriosa em todas. Igualmente com relação aos grupos formados pelas empresas {32254,32643}, {57617,40290} e {52354,48244}, com concentração também de 100%, respectivamente, nas cidades de Guarabira, Paulista e Santa Rita, uma das firmas venceu a maioria ou todos os certames de que participou em conjunto, enquanto suas correspondentes parceiras apresentaram tímida atuação, perdendo todas ou a maioria das disputas. Depreende-se, dessa forma, que, além da concentração espacial em determinadas regiões do estado, tais regras também carregam consigo indicações potenciais de simulação de concorrência e rodízio em licitações, ainda que com médio risco de conluio.

Já com relação ao período de 2009–2012, tem-se a seguinte distribuição espacial com 100% de concentração nos seguintes municípios, apenas do Agreste e do Sertão do estado: Caiçara, Campina Grande, Guarabira e Nova Floresta (Agreste) e Brejo do Cruz, Catolé do Rocha, Manaíra, Paulista e Sousa (Sertão). Para o período de gestão 2013–2016, foram constatados grupos de empresas com atuação mais presente nos seguintes municípios, de acordo suas regiões: Rio Tinto (região da Mata); Nova Floresta (Agreste); Monteiro e São Domingos do Cariri (Borborema); e Belém do Brejo do Cruz, Brejo do Cruz e Catolé do Rocha (Sertão).

4.5 Ranqueamento de empresas suspeitas de conluio

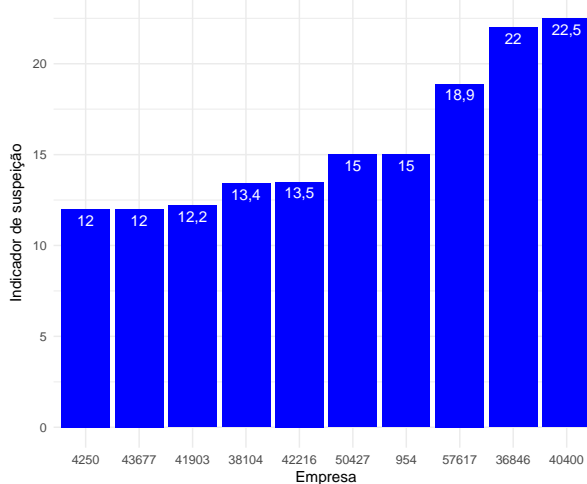
A Figura 1 apresenta uma síntese dos resultados analisados até o momento, ao registrar as 10 empresas com maiores indicadores de suspeição de conluio de acordo com os distintos experimentos de aprendizado de máquina por períodos de gestão municipal. Os indicadores foram calculados a partir da equação (3), considerando a recorrência de cada empresa em grupos com grande tendência de vitória, grupos com presença de empresas perdedoras e grupos com atuação concentrada em determinado município.

No tocante ao experimento realizado para o período 2005–2008, os resultados apontam elevado grau de suspeição para as seguintes empresas: a) 40400, 36846, 57617, 954, 42216, 41903 e 4250 (fornecedores de gêneros alimentícios); b) 50427 (fornecedor de órteses e próteses); c) 38104 (fornecedor de *softwares* e de serviços de manutenção de sistemas); e d) 43677 (fornecedor de material gráfico e de serviços de copiadora).

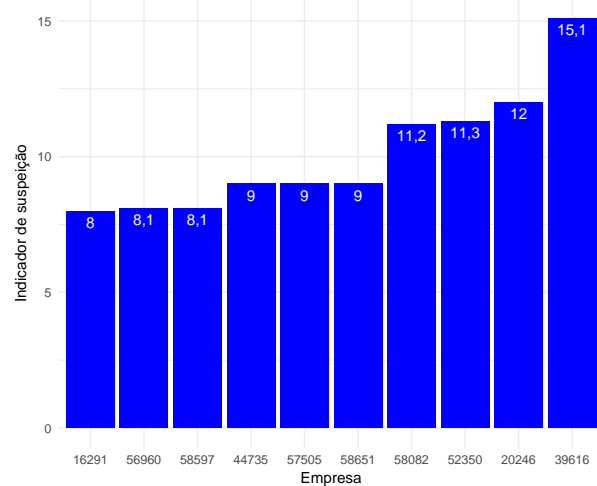
Os achados para o conjunto de dados de 2009–2012 apontam medidas estatísticas iniciais de suspeitas para as empresas: a) 39616, 57505 e 16291 (fornecedores de gêneros alimentícios, produtos de limpeza, carne e hortifrutigranjeiros); b) 20246 (prestador de serviços técnicos especializados na área de engenharia civil); c) 52350 (fornecedor de materiais de construção, cal, tintas e solventes); d) 58082 (fornecedor de material esportivo); e) 58651 e 44735 (fornecedores de combustível, lubrificantes e óleos); f) 56960 e 58597 (prestadores de serviços de publicidade: propaganda, comunicação e marketing).

No rol de empresas com alto grau de suspeição de conluio detectadas no experimento para o período 2013–2016, destacam-se: a) 41448, 38258, 40423, 49250, 46847, 37953, 40143 e 24966 (fornecedores de gêneros alimentícios, hortifrutigranjeiros e material de limpeza); e b) 35753 e 23719 (prestadores de serviços de transporte autônomo). Nesse contexto, chama atenção a forte discrepância de pontuação da firma 41448 em relação às demais. Tal fato pode ser justificado pelo seu envolvimento em 4 regras de associação com alto risco de conluio, especialmente concentrado nos municípios de Belém do Brejo do Cruz e Brejo do Cruz.

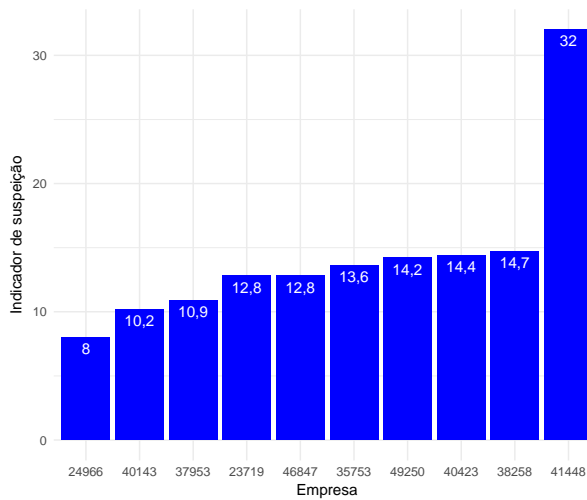
Figura 1 – Empresas com maiores índices de suspeição de conluio



(a) 2005–2008



(b) 2009–2012



(c) 2013–2016

Fonte: Elaboração própria a partir dos dados do TCE-PB.

5 Implicações dos resultados para administração pública

Neste estudo, a utilização de ferramentas de aprendizado de máquina se mostrou importante para o mapeamento de padrões de associação entre empresas licitantes mediante grandes bases de dados transacionais (*big data*), em que problemas de crescimento exponencial de combinações de elementos na formação de conjuntos precisam ser superados. A avaliação criteriosa de conjuntos de empresas com participação associada à luz da Teoria dos Jogos possibilitou a geração de conhecimento com efetiva possibilidade de aplicação em ambientes institucionais, em especial para atividades ligadas à auditoria e à fiscalização de licitações do setor público e dos contratos delas decorrentes.

É cabível ressaltar que o índice de suspeição proposto nesta pesquisa apresenta um mapeamento inicial de riscos de potenciais irregularidades nos processos de licitações específicos de uma dada empresa, indicando trilhas de auditoria para os órgãos de controle e alertando, inclusive, o setor de compras das gestões públicas sobre a necessidade de uma atenção especial ao certame, dado o histórico de participantes com recorrência de comportamentos potencialmente tendenciosos.

A despeito de tais achados não consistirem em lastro probatório suficiente para atestar a existência de fraudes, eles fornecem indicações iniciais de suspeição, podendo ser usados como meio de se otimizar o trabalho exploratório de triagem e localização de risco por parte da auditoria e controle do gasto público e, conseqüentemente, melhor carrear o dispendioso esforço técnico necessário à fiscalização do erário.

6 Conclusão

Este trabalho teve por objetivo geral identificar grupos de empresas suspeitas de práticas potencialmente anticompetitivas em licitações ocorridas nos municípios paraibanos durante o período de 2005 a 2016. Para tanto, foram realizadas tarefas de aprendizado de máquina por ciclos de gestão municipal – busca de regras de associação – a partir de massa de informações de processos licitatórios no estado da Paraíba. O uso do algoritmo Apriori possibilitou, a partir de regras heurísticas e de análise combinatória, a identificação preliminar de vários conjuntos de empresas (regras de associação) cujos membros têm alta associação estatística em termos de atuações conjuntas em licitações.

O processo de descoberta de conhecimento foi aprofundado a partir da aplicação de uma função de avaliação das regras de associação, com o escopo de minimizar o problema das associações espúrias. Nesta fase foi possível eliminar regras de associação em que ao menos um membro era fornecedor frequente (regras casuísticas) e classificar grupos de empresas conforme risco de práticas potencialmente anticompetitivas, considerando a tendenciosidade de vitória dos membros de cada grupo e o total de concorrentes das empresas envolvidas. A partir desta filtragem e classificação foram investigadas outras características dos grupos suspeitos: distribuição de participações de cada empresa, distribuição de vitórias e distribuição geográfica de atuações. Tais informações foram usadas no cômputo de um indicador de suspeição para cada empresa, considerando o acúmulo de pontos por recorrência em diferentes regras de associação (rodízios), o envolvimento em grupos com presença de falsos concorrentes e a participação em grupos com atuações restritas a um único município.

Os resultados permitiram a identificação de vários padrões de associação entre empresas licitantes. No âmbito dos grupos de empresas classificadas com alto risco de conluio foram observadas associações entre fornecedores de gêneros alimentícios e produtos/serviços de limpeza, transportadores autônomos e fornecedores de materiais gráficos. Tais grupos não apenas registraram forte correlação na participação conjunta de empresas, mas também elevada propensão de vitória de determinados participantes, ainda que na presença de elevado número de concorrentes. Já os achados referentes aos grupos classificados com médio risco de conluio, isto é, regras de associação cujos membros têm alta propensão de vitória e baixo número de concorrentes, revelaram o envolvimento de empresas de diversos ramos de atividade econômica, a saber: fornecedores de reagentes para exames bioquímicos; fornecedores de medicamentos, materiais médicos hospitalares e odontológicos; fornecedores de cadeiras de rodas, prótese e aparelhos ortopédicos; fornecedores de serviços gráficos, insumos de informática e locação de copiadoras; fornecedores de materiais de construção; prestadores de consultorias nas áreas de contabilidade e de engenharia civil; fornecedores de pneus, peças e serviços mecânicos; fornecedores de combustíveis e lubrificantes; prestadores de serviços de locação de veículos; fornecedores de material esportivo e fardamento; prestadores de serviço de locação de palcos, banheiros químicos, sonorização e iluminação; e prestadores de serviços na área laboratorial.

As tarefas de aprendizado de máquina também produziram evidências iniciais sobre simulação de concorrência em vários grupos de empresas com alta associação na participação de certames. Alguns comportamentos chamaram atenção, a saber: i) firmas que só atuaram em licitações quando da participação da parceira na regra de associação (empresa que vincula a participação da outra no certame); ii) empresas que normalmente perderam todas (ou quase todas) as licitações quando da participação conjunta com uma firma parceira e que continuaram com esse comportamento ao longo das demais disputas com rivais fora da regra de associação, gerando indicadores potenciais de simulação de concorrência; iii) licitantes que apresentaram acentuada inversão de comportamento, ou seja, perdiam sempre nas disputas com empresas parceiras em atuação conjunta (regra de associação) e passaram a ganhar um expressivo número de certames em licitações que competiram com outras firmas; iv) empresas que venciam sempre quando disputavam em conjunto com a parceira da regra de associação, contudo, passaram a perder em licitações com outras firmas.

No tocante às evidências sobre concentração regional, os resultados mostraram que é no município de Guarabira onde há uma maior persistência em termos de atuação de grupos com maior índice de suspeição, uma vez que a empresa 48775, por exemplo, fechou contratos com esta municipalidade em todos os três períodos de gestão política considerados, em particular na atuação conjunta com a firma 5230, formando grupo nos ciclos políticos subsequentes de 2005–2008 e 2009–2012. Ademais, também se destacaram os municípios de Marizópolis, Brejo do Cruz e Belém do Brejo do Cruz. Todas estas localidades concentraram vários grupos de empresas com altos padrões de associações. Nesse rol, foram

achadas medidas estatísticas de suspeição envolvendo as empresas 40400, 36846, 39616 e 41448, haja vista o envolvimento destas em distintas regras de associação.

O ranqueamento de empresas conforme grau de suspeição de risco produziu resultados consistentes com aqueles identificados nas regras de associação avaliadas. Em geral, as firmas com maiores valores no indicador de suspeição atuam nos ramos de gêneros alimentícios, fornecimento de órteses e próteses, fornecimento de *softwares* e de serviços de manutenção de sistemas, fornecimento de material gráfico e de serviços de copiadora, prestação de serviços técnicos especializados na área de engenharia civil, prestação de serviços técnico-contábeis, fornecimento de materiais de construção, cal, tintas e solventes, fornecimento de material esportivo, fornecimento de combustível, lubrificantes e óleos, prestação de serviços de publicidade: propaganda, comunicação e marketing, e prestação de serviços de transporte autônomo.

Por fim, destacamos que os resultados e os instrumentais apresentados nesta pesquisa precisam ser vistos com cautela, tendo em vista que as regras de associação e o ranqueamento de empresas pelo índice de suspeição precisam de análises qualitativas complementares dos órgãos competentes (como Tribunal de Contas, Ministério Público e Polícia Federal) para se identificar a efetiva prática de fraude. Não obstante, as evidências trazidas podem ser úteis como uma triagem inicial e exploratória para análise de casos com maiores riscos potenciais de práticas anticompetitivas, sendo relevantes para otimizar os recursos e subsidiar a decisão de quais processos podem ser priorizados para uma avaliação específica. Ademais, em investigações futuras, pretendemos avançar com pesquisas de avaliação de impacto e mensuração da qualidade preditiva do modelo de suspeição de empresa, bem como integrar as informações dos padrões de associação com bases de inidoneidade de empresas, interseção de quadro societário (1º e 2º graus) e outras fontes informacionais, com o escopo de calibrar e aumentar a precisão da detecção de práticas anticompetitivas em licitações públicas.

Referências

- Abrantes-Metz, R. M., Froeb, L., Geweke, J., Taylor, C. (2006). A variance screen for collusion. *International Journal of Industrial Organization*, 24(3), 467–486.
- Agrawal, R., Srikant, R. (1994). Fast Algorithms for Mining Association Rules. In: Proceedings of the 20th VLDB Conference, Santiago, 487–499.
- Aoyagi, M. (2003). Bid rotation and collusion in repeated auctions. *Journal of Economic Theory*, 112(1), 79–105.
- Athey, S., Bagwell, K. (2001). Optimal collusion with private information. *RAND Journal of Economics*, 32(3), 423–428.
- Bajari, P., Ye, L. (2003). Deciding Between Competition and Collusion. *Review of Economics and Statistics*, 85(4), 971–989.
- Balaniuk, R. et al. (2012). Risk based government audit planning using naïve bayes classifiers. *Frontiers in Artificial Intelligence and Applications*, 243(1), 1313–1323.
- Blanckenburg, K., Geist, A., Kholodilin, K. A. (2012). The Influence of Collusion on Price Changes: New Evidence from Major Cartel Cases. *German Economic Review*, 13(3), 245–256.
- Bolotova, Y., Connor, J. M., Miller, D. J. (2008). The impact of collusion on price behavior: Empirical results from two recent cases. *International Journal of Industrial Organization*, 26(6), 1290–1307.
- Brasil. Tribunal de Contas da União. (2010). *Licitações e contratos: orientações e jurisprudência do TCU*. Brasília: TCU.
- Campos, F. (2008). As práticas de conluio nas licitações públicas à luz da teoria dos jogos. *Revista Análise Econômica*, 50(26), 185–206, 2008.
- Da Silva, J. C. et al. (2005). Distributed data mining and agents. *Engineering Applications of Artificial Intelligence*, 18(7), 791–807.
- Gibbons, R. (1992). *Game Theory for Applied Economists*. New Jersey: Princeton University Press.
- Hahsler, M., Chelluboina, S., Hornik, K., Buchta C. (2011). The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Datasets. *Journal of Machine Learning Research*, 12, 1977–1981.
- Heimler, A. (2012) Cartels in public procurement. *Journal of Competition Law & Economics*, 8(4), 849–862.

- Hüschelrath, K., Veith, T. (2008). Cartel Detection in Procurement Markets. *Managerial and Decision Economics*, 35(1), 404–422.
- IBGE – Instituto Brasileiro de Geografia e Estatística. (2018). Produto Interno Bruto dos Municípios. Rio de Janeiro.
- Lantz, B. (2013). *Machine Learning with R*. UK: Packt Publishing.
- Lundberg, S. (2005). Restrictions on competition in municipal competitive procurement in Sweden. *International Advances in Economic Research*, 11(3), 329–342, 2005.
- Mcafee, P., Mcmillan, J. (1992). Bidding rings. *The American Economic Review*, 82(3), 579–599.
- McNicholas, D., Zhao, Y. (2009). *Association rules: An overview*. London: IGI Global.
- OCDE (2009). Diretrizes para combater o conluio entre concorrentes em contratações públicas. OCDE, 19, 2009. Disponível em: <<http://www.oecd.org/daf/competition/cartels/44162082.pdf>>.
- OCDE (2012). Government at a Glance. OCDE, 2011. 18. 254 p. Disponível em: <http://www.oecd-ilibrary.org/governance/government-at-a-glance-2011_gov_glance-2>.
- Ormosi, L. (2014). A tip of the iceberg? The probability of catching cartels. *Journal of Applied Econometrics*, 29(1), 549–566.
- Padhi, S. S., Mohapatra, K. J. (2011). Detection of collusion in government procurement auctions. *Journal of Purchasing and Supply Management*, 17(4), 207–221.
- Pesendorfer, M. (2000). A Study of Collusion in First-Price Auctions. *The Review of Economic Studies*, 67(3), 381–411.
- Porter, R. H., Zona, J. D. (1993). Detection of Bid Rigging in Procurement Auctions. *Journal of Political Economy*, 101(3), 518–538.
- Rachmilevitch, S. (2014). First-best collusion without communication. *Games and Economic Behavior*, 83, 224–230.
- Ralha, C. G., Silva, C. S. (2012). A multi-agent data mining system for cartel detection in Brazilian government procurement. *Expert Systems with Applications*, 39(14), 11642–11656.
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em < <https://www.r-project.org/>>.
- Transparency International. (2016). *Corruption Perceptions Index 2015*. Berlin: Transparency International Press.
- Wang, H. (2016). Optimal implicit collusion in repeated procurement auctions. *Journal of Economics*, 117(3), 259–284.