

PROGRAMA DE PESQUISA
ESCOLA DE CONTAS E GESTÃO – ECG-TCE/RJ

RELATÓRIO FINAL

Título do Projeto:

Descoberta de Conhecimento em Bases de Dados, com base em técnicas de Análise Preditiva, visando ao auxílio e ao aprimoramento da atividade de controle externo desempenhada pelo TCE-RJ

Grupo de Pesquisa:
Pesquisador(es)

Pesquisador-coordenador: BRUNO MATTOS SOUZA DE SOUZA MELO
Pesquisador: MARCOS FERREIRA DA SILVA
Pesquisador Avaliador: MARCONI CANUTO BRASIL

Processo administrativo nº:

ESCOLA DE CONTAS E GESTÃO – ECG-TCE/RJ

Diretora-Geral

PAULA ALEXANDRA NAZARETH

Coordenadora-Geral de Estudos e Pesquisas

ROSA MARIA CHAISE

Comissão Permanente de Estudos e Pesquisas – COPEP

ROSA MARIA CHAISE (Presidente)

CLÁUDIO MARTINELLI MURTA

ELAINE FARIA DE MELO

GUILHERME PINTO DE ALBUQUERQUE

JEAN MARCEL DE FARIA NOVO

JORGE HENRIQUE MUNIZ DA CONCEIÇÃO

MARCELLO LEONI ARRUFAT TORRES

MARIA BERNADETTE DAYRELL DE MIRANDA

PAULA ALEXANDRA C. DE PAIVA NAZARETH

PAULO CÉSAR BESSA NEVES

RENATA DE OLIVEIRA RAZUK

SÉRGIO PAULO VIEIRA VILAÇA

SÉRGIO LINO DA SILVA CARVALHO

Pesquisador-avaliador

MARCONI CANUTO BRASIL

Data

26/02/2016

Sumário

RESUMO	4
INTRODUÇÃO	5
PROBLEMA E OBJETIVOS	6
REFERENCIAL TEÓRICO.....	7
METODOLOGIA.....	8
CRONOGRAMA REALIZADO.....	11
RESULTADOS	12
CONCLUSÕES.....	23
REFERÊNCIAS.....	25
APÊNDICES	27
PARECER DO PESQUISADOR-AVALIADOR	31

RESUMO

A pesquisa levada a efeito objetivou avaliar a possibilidade de que técnicas de mineração de dados possam ser utilizadas em auditorias governamentais para a detecção de fraudes. Nesta pesquisa foram utilizados dados colhidos junto à Secretaria de Estado de Planejamento (SEPLAG) relativos a licitações e contratos celebrados no período de agosto de 2011 a agosto de 2015. Além desses dados, foram utilizados também dados de participação societária das empresas que participaram das licitações e contratos objeto desta pesquisa. A detecção de fraudes em licitações constitui um objetivo importante a ser alcançado pela fiscalização realizada pelos Tribunal de Contas, e a possibilidade de se dispor de mecanismos que permitam direcionar os esforços de fiscalização para as licitações e contratações que apresentem um maior risco de fraude. Técnicas de mineração de dados têm sido utilizadas para auxiliar na detecção de fraudes em diversos domínios, tais como: seguros, cartões de crédito, telecomunicações etc. Aplicações na área de auditoria governamental também têm sido objeto de pesquisa. Conquanto a pesquisa tenha evidenciado que é possível introduzir esta metodologia na rotina de uma auditoria governamental, algumas dificuldades foram identificadas, tais como: (a) baixa qualidade dos dados disponíveis para utilização; (b) inadequação do *hardware* disponível para uso; (c) incompatibilidade entre o tempo usualmente destinado à execução de uma auditoria e o que seria necessário para trabalhar os dados colhidos, aplicar as técnicas de mineração de dados adequadas ao trabalho em execução e interpretar os resultados obtidos de forma que estes possam ser utilizados na auditoria. A adoção, pelo TCE-RJ, de uma estratégia de controle que tenha por objetivo disseminar a utilização de métodos quantitativos como ferramenta de auditoria governamental deve considerar investir fortemente em treinamento em métodos quantitativos e uso de *softwares* adequados, visto que a cultura de controle existente é fortemente baseada em análise documental.

Palavras-chaves: Auditoria Governamental, Mineração de Dados, Detecção de Fraudes.

INTRODUÇÃO

Inovações tecnológicas são responsáveis por enormes transformações que têm impactado irreversivelmente a sociedade. Os avanços na área da Tecnologia da Informação, em particular na coleta, transmissão e armazenamento de dados, propiciam a criação e o compartilhamento de grandes bases de dados, complexas e heterogêneas.

A legislação tem refletido estas transformações, exigindo que diversos dados de interesse da sociedade sejam disponibilizados em formato adequado para consumo por máquinas, no que hoje se conhece por dados abertos.

A pesquisa ilustra como as técnicas de mineração de dados podem indicar indícios de irregularidades, utilizando dados oriundos dos sistemas de informação utilizados pela administração pública estadual.

A pesquisa utilizou dados relativos a contratações públicas objetivando identificar risco de ocorrência das seguintes irregularidades:

- Cartel ou Rodízio de empresas em licitações;
- Fracionamento de despesas;
- Licitantes distintos possuindo sócios ou endereços em comum;

Foram utilizadas regras de associação, clusterização, mineração de texto e análise de redes sociais. A utilização destas técnicas representa um avanço em relação às técnicas de análise de dados atualmente empregadas em auditorias de dados realizadas no âmbito do TCE-RJ, que consistem basicamente em técnicas baseadas em verificação de regras de negócio ou cruzamento de dados.

Neste relatório serão apresentados os resultados obtidos para a primeira das três tipologias de irregularidade acima citadas por ter sido a mais complexa e validada por duas abordagens distintas e independentes (Regras de Associação e Análise de Redes Sociais). As demais, embora tenham sido concluídas, não farão parte do corpo deste relatório sendo que os métodos utilizados na sua elaboração e os resultados alcançados serão brevemente apresentados no Apêndice III.

Estas técnicas foram implementadas com a utilização do *software* estatístico *R* (www.r-project.org) em conjunto com o *SQLServer*. No apêndice, apresenta-se um rol de *softwares open source* que podem ser utilizados para implementar diversas técnicas de análise de dados. Objetiva-se, com essa iniciativa, colocar à disposição dos técnicos ferramentas que viabilizem a aplicação de técnicas de análise de dados sem necessidade de vultosos investimentos.

A pesquisa também evidencia que é possível utilizar uma abordagem de auditoria direcionada por dados, pró-ativa, em oposição a uma abordagem reativa, direcionada apenas pela intuição e experiência do auditor.

PROBLEMA E OBJETIVOS

Problema da pesquisa:

É viável a implementação de técnicas de detecção de fraudes com a utilização de mineração de dados nas auditorias realizadas pelo TCE-RJ?

Objetivo Geral:

Demonstrar a viabilidade da implementação de técnicas de mineração de dados para detecção de fraudes em auditorias governamentais.

Objetivos específicos:

Identificar as condições necessárias para a utilização das técnicas de mineração de dados em processos de fiscalização pelos técnicos do TCE-RJ;

Revisão da literatura sobre aplicação de técnicas de mineração de dados em detecção de fraudes no âmbito de auditorias governamentais.

REFERENCIAL TEÓRICO

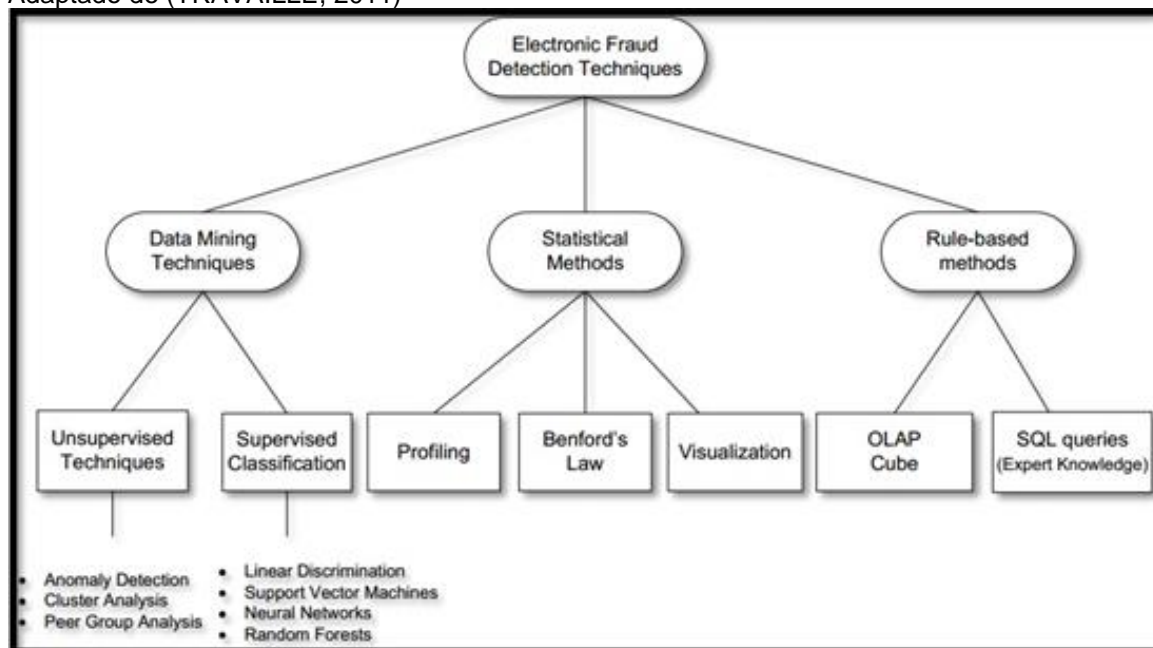
Já há muito que a auditoria financeira tradicional vem sendo objeto de críticas em razão de sua aparente incapacidade de detectar fraudes ocorridas em demonstrações financeiras de diversas corporações (ALBRECHT, 2003). Este fato tem levado ao que se denomina “lacuna de expectativas” que se traduz numa divergência entre o que o usuário das demonstrações financeiras espera dos auditores no que diz respeito à identificação de fraudes e o que os auditores acreditam ser sua responsabilidade neste mister.

Albrecht (2003) conclama os pesquisadores na área de detecção de fraude, ciência da computação e estatística a aplicar suas técnicas e conhecimentos para a pesquisa de formas pelas quais as fraudes possam ser detectadas durante a realização de uma auditoria financeira.

Propõe que a análise seja feita em toda a população de itens a serem examinados e não mais sobre amostras de itens extraídas da população e que se adote uma postura mais pró-ativa em relação à detecção de fraudes (uma abordagem conhecida como *data-driven fraud detection*) o que exige uma análise específica de cada área de negócio e a indicação de potenciais fraudes que possam ocorrer. Então, utilizando *scripts* e *queries* especificamente desenvolvidas para detectar as fraudes previstas, minera-se toda a população de dados em busca dos sintomas de fraude (ALBRECHT, ALBRECHT, DUNN, 2001; ALBRECHT, ALBRECHT, 2002).

A adoção de técnicas de mineração de dados permite ir além da abordagem proposta pelo autor em referência, a qual se baseia em regras. A Figura 1 ilustra as abordagens existentes para detecção de fraudes via análise de dados.

Figura 1 - Taxonomia das Técnicas de Detecção Eletrônica de Fraudes.
Adaptado de (TRAVAILLE, 2011)



No âmbito da auditoria governamental, alguns trabalhos já têm mostrado a necessidade de se integrar o uso de técnicas de mineração de dados ao processo auditorial. Nessa linha cabe citar os trabalhos de BALANIUK (2010), MARTINS JÚNIOR e BRAZ (2010), SILVA e ROCHA (2010), SALES e CARVALHO (2014) e BODINI JUNIOR (2009).

METODOLOGIA

Tomando-se por base a taxonomia apresentada por VERGARA (2014), a qual qualifica a pesquisa científica em relação aos aspectos relativos aos fins e aos meios, o presente trabalho pode ser classificado como uma pesquisa exploratória, metodológica, aplicada, no que se refere aos fins e, quanto aos meios, como uma pesquisa bibliográfica e um estudo de caso.

Quanto aos fins, a pesquisa classifica-se como exploratória uma vez que a aplicação de técnicas de mineração de dados no âmbito das auditorias governamentais é algo novo e pouco explorado. Se considerarmos o universo dos órgãos de controle brasileiros, os que se situam na esfera federal encontram-se num estágio mais avançado, já possuindo algumas iniciativas e trabalhos publicados. No que se refere ao TCE-RJ, tal iniciativa é ainda inédita. Conclui-se, pois que estamos diante de um campo do conhecimento com amplo espaço de desenvolvimento e ainda em busca de consolidação.

Este trabalho é também uma investigação do tipo metodológica, dado que visa à proposição de novos procedimentos e técnicas a serem aplicadas nas diferentes fases da auditoria governamental, sobretudo no seu planejamento e execução. O rótulo de pesquisa aplicada é igualmente adequado visto que as metodologias propostas foram efetivamente empregadas num caso concreto (numa auditoria) a fim de atestar sua viabilidade e aplicabilidade revestindo-se, portanto numa prova de conceito.

No que concerne aos meios, a pesquisa está fundamentada num extenso levantamento bibliográfico e culminou na realização de um estudo de caso circunscrito numa auditoria governamental real objetivando o aprofundamento e o detalhamento da investigação científica.

Quanto a aplicabilidade em auditorias governamentais, as técnicas de mineração de dados enquadram-se no rol de procedimentos substantivos. De fato, conforme KOSKIVAARA (2007), várias técnicas e métodos podem ser utilizados para a realização de procedimentos de revisão analítica, sendo que estas técnicas podem ir de simples comparações a análises complexas.

A referida autora menciona que pesquisadores na área de auditoria tem adotado diferentes classificações para estas técnicas de auditoria, mencionando as classificações de BLOCHER E PATTERSON (1996) e de FRASER et al (1997). Os primeiros autores identificam três tipos de procedimentos de revisão analítica: análise de tendências, análise de quocientes e baseadas em modelos, enquanto os últimos autores as classificam em não quantitativas ou subjetivas, quantitativas simples e quantitativas avançadas. Como exemplos de técnicas quantitativas avançadas menciona a análise de regressão e as redes neurais artificiais, duas técnicas de mineração de dados bastante conhecidas e utilizadas.

Fica claro que, a utilização de mineração de dados em auditoria dá-se no âmbito da utilização dos procedimentos de revisão analítica, conforme a norma de auditoria NBC TA 520 – Procedimentos Analíticos, para a qual “Podem ser usados diversos métodos para executar procedimentos analíticos. Esses métodos variam desde a realização de comparações simples até análises complexas usando técnicas estatísticas avançadas.”

Os dados utilizados nesta pesquisa consistiram em dados colhidos no âmbito de uma fiscalização levada a efeito pela Coordenadoria de Auditorias Temáticas e Operacionais da Secretaria Geral de Controle Externo do Tribunal de Contas do Estado do Rio de Janeiro referentes a licitações e contratos ocorridos no período de janeiro de 2011 a agosto de 2015. A relação dos arquivos de dados utilizados pode ser consultada no Apendice II a este relatório de pesquisa.

Dadas as características dos dados utilizados nesta pesquisa, as técnicas de mineração de dados utilizadas consistiram em técnicas não supervisionadas (ou descritivas) já que não possuem uma variável “target” (alvo) a ser prevista. Especificamente foram utilizadas nesta pesquisa as seguintes técnicas: regras de associação, mineração de texto, clusterização e análise de redes sociais.

Far-se-á, a seguir, uma rápida apresentação de cada uma destas técnicas.

Regras de Associação

Esta técnica consiste na detecção de associações que ocorrem com frequência entre itens (AGRAWAL, IMIELINSKI et al. 1993 *apud* BAESENS et al. 2015). Esta técnica foi desenvolvida inicialmente com o objetivo de detectar quais itens eram adquiridos juntos em compras realizadas em lojas, daí a denominação pela qual a técnica também é conhecida: *market basket analysis*.

As associações identificadas pela técnica consistem em implicações da forma $X \rightarrow Y$ onde X e Y , denominados *itemsets*, são subconjuntos de um conjunto $I = \{i_1, i_2, \dots, i_m\}$ de itens e $X \cap Y = \emptyset$. Itens podem ser produtos comprados em um supermercado, vídeos alugados em uma locadora ou, como no caso desta pesquisa, empresas que participam de licitações.

Exemplificando, uma regra é uma implicação tipo $\{\text{licitante}_A, \text{licitante}_B\} \rightarrow \{\text{licitante}_C\}$ significando que se os licitantes A e B participaram de uma licitação é provável que o licitante C também tenha participado. Regras de associação constituem a base do que se conhece como sistemas de recomendação, que consiste em recomendar ao cliente produtos com base em compras anteriores ou consultas a itens, algo do tipo: “pessoas que viram o filme A, também viram o filme B”.

No contexto desta pesquisa, a técnica foi utilizada objetivando identificar licitantes que frequentemente participam juntos em licitações, o que pode ser um indício do que CARVALHO (2007) define como rodízio de empresas, situação onde as empresas combinam com antecedência quem será o vencedor da licitação, pactuando “por fora” os valores que serão ofertados. O objetivo é gerar regras do tipo: $\{\text{licitante}_A, \text{licitante}_B\} \rightarrow \{\text{licitante}_C\}$ indicando os licitantes que, frequentemente, participam juntos em licitações.

Deve-se notar que nem todas as regras geradas pela técnica serão de interesse. Assim, algumas medidas de qualidade das regras geradas são propostas na literatura, sendo as mais importantes o *support* (suporte), a *confidence* (confiança) e o *lift* (elevação).

O suporte de uma regra $X \rightarrow Y$ é dado pela proporção em que a os itens contidos na regra ocorrem em conjunto em toda a base de dados. Assim, o suporte da regra $X \rightarrow Y$ é dado por: $supp(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$ onde $\sigma(X \cup Y)$ é a quantidade de vezes em que os *itemsets* em X e Y ocorreram juntos no conjunto de todas as transações, dado por N .

A confiança (*confidence*) é a proporção de vezes em que o conseqüente da regra nos itens em que o antecedente ocorre; é uma probabilidade condicionada que informa a probabilidade de obter o conseqüente: dado o antecedente. Matematicamente a confiança da regra $X \rightarrow U$ é dada por: $conf(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}$

O *lift* é uma medida de associação entre os *itemsets* X e Y de uma regra, indicando o afastamento da hipótese de independência entre o antecedente (*itemset* X) e o conseqüente (*itemset* Y) de uma regra. Matematicamente o *lift* é dado por $lift(X, Y) = \frac{supp(X \rightarrow Y)}{supp(X) \times supp(Y)}$. Se o *lift* for igual a 1, então X e Y são independentes, significando que a ocorrência dos itens em X não tem relação com os itens em Y .

SILVA e ROCHA (2010) mencionam um estudo de caso em que aplicam a técnica a uma base de dados relativa a licitações extraída do ComprasNet contendo informações sobre os pregões realizados pelo Governo Federal no período de 2005 a 2008 para a aquisição de um determinado tipo de serviço.

Mineração de Texto

A mineração de texto é uma técnica de mineração que se ocupa da extração de conhecimento do processamento de dados textuais, não estruturados. O instrumento básico da mineração de texto é a matriz de documentos e termos (ou a matriz de termos e documentos) que consiste em uma matriz cujas linhas representam os documentos contendo os textos em análise e as colunas as palavras contidas nos textos. O conteúdo da matriz consiste na frequência de cada palavra em cada um dos documentos. A partir desta matriz, diversas medidas podem ser calculadas, como a *tf-idf* (*term frequency - inverse document*

frequency), que é uma medida do quão importante uma palavra é em um documento integrante de um *corpus* (conjunto de documentos).

Clusterização

Conforme BAESENS *et al.* (2015) o objetivo da clusterização (agrupamento) é dividir um conjunto de observações em segmentos de tal forma que a homogeneidade dentro de cada segmento seja maximizada e a heterogeneidade entre os segmentos também seja maximizada. Sua aplicação em detecção de fraudes consiste em agrupar as anomalias em pequenos e esparsos *clusters*.

Análise de Redes Sociais

As Redes Sociais consistem em estruturas que representam entidades e relacionamentos entre estas. Entidades podem ser pessoas, empresas, ou quaisquer coisas que possam possuir algum tipo de relação entre si.

A Análise de Redes Sociais permite representar as redes sociais através da representação dos nós (entidades) e das ligações entre eles. Os nós da rede social representam os atores dessa rede (indivíduos ou organizações). As ligações representam as relações entre os atores componentes da rede representada.

Este ramo da mineração de dados representa as relações sociais em termos da Teoria de Grafos, permitindo o estudo, por meio da identificação dos atores e suas ligações, das relações entre os mesmos de forma a poder identificar as formas ou padrões de interação entre si. Para BAESENS *et al.* (2015), redes sociais são um elemento importante na análise da fraude, visto que quando as técnicas analíticas tradicionais falham na detecção de fraudes em razão da falta de evidências, a análise de redes sociais pode trazer novos *insights* em razão da investigação de como as pessoas influenciam e são influenciadas umas pelas outras.

Recursos Tecnológicos

Em termos de recursos tecnológicos, foram utilizados os equipamentos e softwares que já integram a estrutura de tecnologia da informação existente no TCE-RJ e *softwares open source*

já disponíveis aos técnicos do TCE-RJ e *softwares open source* ou que já integram a estrutura de Tecnologia da Informação existente no TCE-RJ. Vê-se, portanto, que a utilização de técnicas de mineração de dados não exigirá vultosos investimentos em software. Especificamente, neste projeto, foram utilizados os seguintes *softwares*: R (*open source*) e *SQLServer* (integrante da plataforma de softwares do TCE-RJ).

No R, foram utilizados os seguintes pacotes:

- *arules*, *arulesViz*, (regras de associação);
- *igraph*, *vizNetwork*, (redes sociais);
- *foreach*, *parallel*, (processamento em paralelo);
- *quanteda* (mineração de texto);
- *dplyr*, *openxlsx*, *RODBC*, *sqldf*, *plyr*, *stringr*, *datatable* (importação e pré-processamento de dados).

CRONOGRAMA REALIZADO

Atividades (*)	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan	Fev
Etapa 1 - Pesquisa bibliográfica / revisão da literatura		X	X										
Etapa 2 - Identificação e avaliação das ferramentas de MD disponíveis			X	X	X	X							
Etapa 3 - Identificação de áreas de aplicação no âmbito do TCE-RJ			X	X	X	X							
Etapa 4 - Identificação das bases de dados disponíveis			X	X	X	X							
Etapa 5 - Pesquisa de maturidade analítica no âmbito da CTO			X	X	X	X	X	X	X				
Etapa 6 - Desenvolvimento da prova de conceito				X	X	X	X	X	X	X	X	X	X

RESULTADOS

Nesta parte do relatório de pesquisa serão apresentados os resultados obtidos em nossa prova de conceito relativa a possibilidade de utilização de técnicas de mineração de dados nas auditorias governamentais realizadas por técnicos do Tribunal de Contas do Estado do Rio de Janeiro – TCE-RJ.

Identificação de possíveis cartões por meio de Regras de Associação

A aplicação da técnica de regras de associação na base de dados de licitações utilizada nesta pesquisa gerou, para um suporte de 0.1% e confiança de 70%, 21 regras, das quais apresentamos as quatro primeiras a seguir:

Rules	support	confidence	lift
{10.910.334/0001-56} => {30.806.616/0001-15}	0,013	0,611	28,878
{61.418.042/0001-31} => {01.012.073/0001-66}	0,011	0,725	7,754
{28.242.238/0001-42} => {01.012.073/0001-66}	0,012	0,674	7,208
{04.521.468/0001-82} => {06.788.843/0001-80}	0,011	0,806	28,292

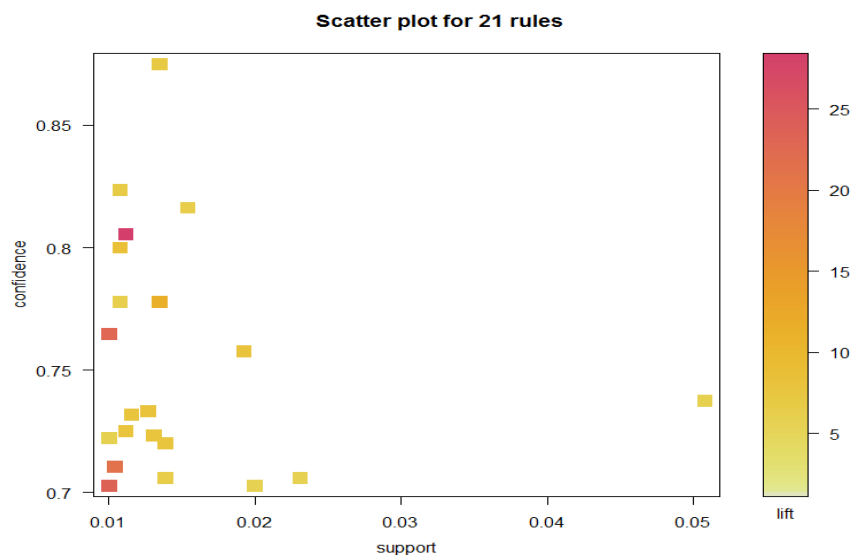
Para a geração das regras utilizou-se apenas os CNPJ das empresas que participaram das licitações, excluindo-se as duplicidades, já que uma mesma empresa poderia aparecer mais de uma vez em uma mesma licitação se estivesse apresentando preços para mais de um item. As regras foram geradas com a função *apriori()* do pacote *arules* do *R*.

Da forma como apresentadas acima, não é possível fazer uma boa avaliação das regras geradas. Para contornar este inconveniente, o pacote *aruleViz* do *R* oferece um conjunto recursos que permitem visualizações gráficas das regras geradas, facilitando a interpretação das regras.

Serão apresentadas a seguir algumas visualizações oferecidas pelo pacote em referência.

No gráfico 1 cada ponto representa uma regra, onde a posição do ponto é dada pelo suporte e confiança da regra e a cor corresponde ao *lift*.

Gráfico 1 – Diagrama de dispersão das regras geradas



Embora este gráfico permita a identificação das regras que possam ser de interesse, já que reúne numa mesma visualização as três principais medidas que auxiliam na seleção de regras que sejam interessantes, não permite identificar os relacionamentos eventualmente existentes entre as empresas no que diz respeito à participação em conjunto em licitações..

O gráfico 2 é especialmente interessante por mostrar como os itens e as regras estão associadas. O gráfico mostra dois agrupamentos de regras e itens (empresas licitantes) que resume-se no quadro a seguir:

Tabela 1 Grupos de empresas identificados por regras de associação

Grupo 1 – Serviços de Saúde

26.921.908/0001-21 – HOSPFAR INDUSTRIA E COMÉRCIO DE PRODUTOS HOSPITALARES LTDA
 07.070.062/0001-19 - MICFARMA COMERCIO DE MEDICAMENTOS E PRODUTOS HOSPITALARES LTDA EPP
 09.660.958/0001-83 - S3MED DISTRIBUIDORA DE MEDICAMENTOS LTDA
 00.085.822/0001-12 - ESPECIFARMA COMERCIO DE MEDICAMENTOS E PRODUTOS HOSPITALARES LTDA
 00.304.559/0001-05 - HUNTER CIENTIFICA COMERCIAL E SERVIÇOS LTDA
 04.344.630/0001-34 - MEGABIO HOSPITALAR LTDA
 48.791.685/0001-68 - CBS MÉDICO CIENTÍFICA COMÉRCIO E REPRESENTAÇÃO LTDA
 01.012.073/0001-66 - HOUSE MED PRODUTOS FARMACEUTICOS E HOSPITALARES LTDA
 61.418.042/0001-31 - CIRURGICA FERNANDES LIMITADA
 32.573.503/0001-42 - HERLAU ATACADISTA DE PRODUTOS HOSPITALARES LTDA
 03.951.140/0001-33 - DE PAULI COMÉRCIO REPRESENTAÇÃO IMPORTAÇÃO E EXPORTAÇÃO LTDA

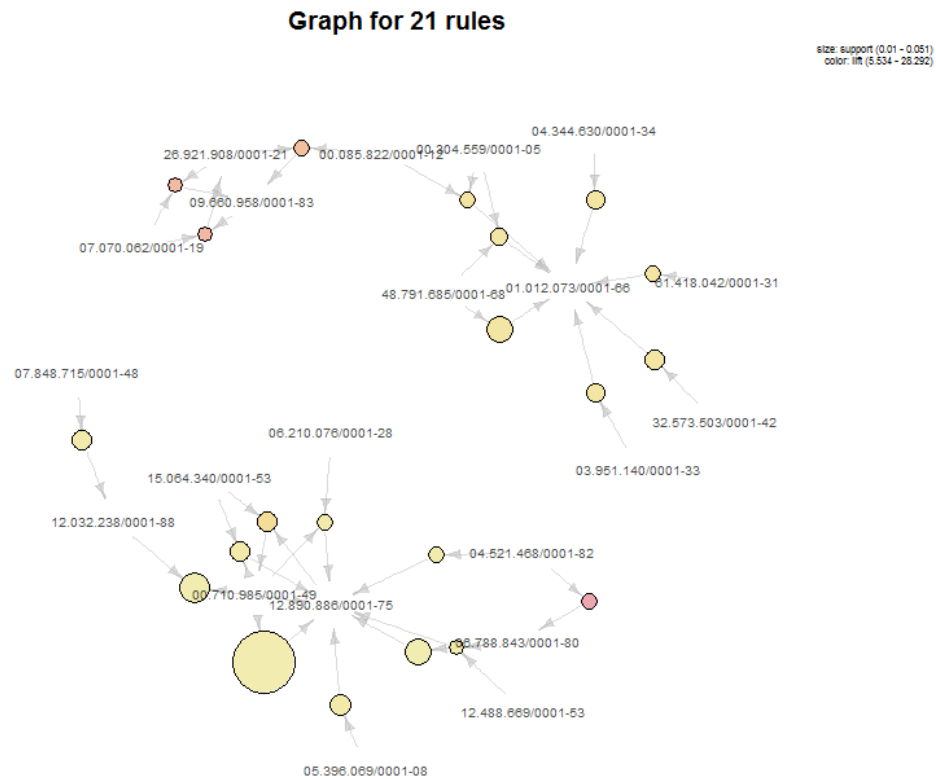
Grupo 2 – Serviços e suprimentos de informática

07.848.715/0001-48 – K S MAX INFORMÁTICA LTDA
 12.032.238/0001-88 - EMAY MATERIAIS E SERVIÇOS LTDA
 06.210.076/0001-28 - QUALYTECK TECNOLOGIA EM INFORMÁTICA EIRELI EPP (ANTIGA QUALYTECK TECNOLOGIA EM INFORMÁTICA LTDA.EPP)
 15.064.340/0001-53 - EFRAIM MATERIAIS E SERVIÇOS LTDA-ME
 00.710.985/0001-49 - VIBHUTI COMERCIO LTDA-MÉ
 12.890.886/0001-75 - VIXNU COMERCIO LTDA-EPP
 04.521.468/0001-82 - LEMAR INK FRANQUIAS LTDA
 12.488.669/0001-53 - QUALYTECK RJ TECNOLOGIA EM INFORMÁTICA EIRELI EPP
 05.396.069/0001-08 - NOVO PONTO DIGITAL COMÉRCIO LTDA
 06.788.843/0001-80 - BENEDITINOS BAZAR E PAPELARIA LTDA.

No gráfico 2 cada círculo representa uma regra, sendo o tamanho do círculo proporcional ao suporte da regra representada; quanto maior o suporte, maior o círculo. A intensidade da cor do círculo é proporcional ao *lift*; quanto mais escuro o círculo, maior o *lift*.

As setas indicam os antecedentes e consequentes da regra. Assim, observando-se um determinado círculo, as setas que “saem” dos círculos, apontam para as empresas que figuram nos consequentes da regra, enquanto as setas que chegam ao círculo referem-se às empresas que figuram nos antecedentes da regra.

Gráfico 2 – Grafo das regras geradas



Com base nas regras geradas, que indicam a frequência de co-ocorrência de empresas em licitações, o auditor poderá focar o trabalho nas licitações em que estas empresas participaram.

Identificação de cartéis e prática de rodízio em licitações por meio de Análise de Redes Sociais

Foram empregadas técnicas de Redes Sociais para identificar situações com elevada probabilidade de serem ocorrências (achados) representativas desta tipologia de irregularidade.

Para isso utilizou-se a base de contratações e licitações estaduais obtida junto à SEPLAG. Esta base contém informações abrangendo o período de janeiro de 2011 a agosto de 2015, com informações de empresas participantes de licitações (inclusive com informações de vencedoras e perdedores), dos itens licitados, dos contratos etc.

A partir destas informações foram construídas redes sociais (representadas por construções matemáticas denominadas grafos¹).

Estruturas que podem ser representadas por grafos estão em toda parte e muitos problemas de interesse prático podem ser formulados como questões sobre grafos. A estrutura de ligação entre páginas da WEB pode ser representada por um grafo direcionado (Gráfico 3): os vértices são as páginas e existe uma aresta da página D para a página A se e somente se D contém um *link* para A.

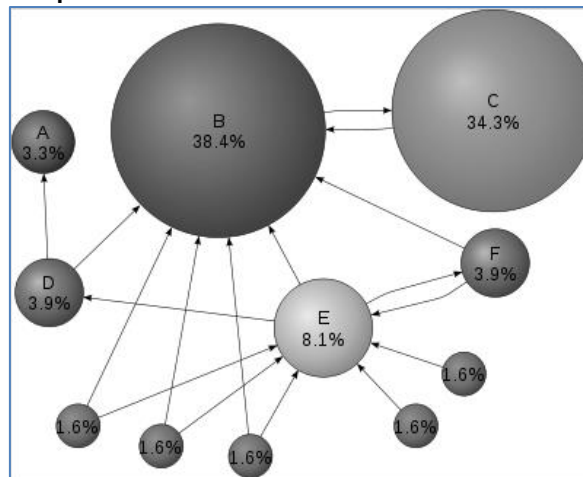
1. A **teoria dos grafos** é um ramo da matemática que estuda as relações entre os objetos de um determinado conjunto.

2. **PageRank** é um algoritmo usado para classificar *sites* em resultados de pesquisas. Foi concebido por Larry Page, um dos fundadores do

Utilizamos grafos direcionados para representar a relação entre as empresas participantes das licitações, da seguinte forma:

1. cada empresa é representada por um nó;
2. as empresas que participaram de um mesmo certame estarão associadas por relações do tipo “perdedor-vencedor”. Tal relação é representada por uma aresta que se inicia no nó representativo da empresa participante perdedora para o nó representativo da licitante vencedora.

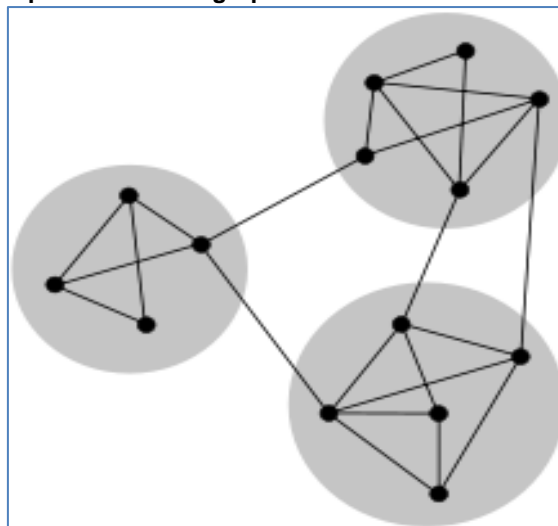
Gráfico 3 - Exemplo de um grafo direcional (contendo arestas direcionais) além de pesos representando a importância relativa (ranqueamento) dos objetos (nós) por ele representado.



Outra ferramenta utilizada foi uma técnica de clusterização específica da teoria de redes sociais denominada detecção de comunidades.

No contexto das redes sociais, a estrutura da comunidade refere-se à ocorrência de grupos de nós que estão mais densamente ligados internamente do que com o resto da rede. Esta heterogeneidade de conexões sugere que a rede tem certas divisões naturais dentro dela (Gráfico 4).

Gráfico 4 - Pequena rede social contendo uma estrutura de comunidade, composta de três grupos de nós com conexões internas densas e conexões esparsas entre os grupos.



A identificação da estrutura de comunidades tem por objetivo “recortar” a rede social original a qual continha, em um único grafo, todas as empresas licitantes e seus respectivos relacionamentos estabelecidos por meio de licitações em que participaram em todo o período (cinco anos) considerado.

A partir desta segmentação é possível delimitar os diferentes “mercados” (medicamentos, alimentos, suprimentos de informática etc.), dado que as empresas que se especializam no fornecimento de uma determinada classe de produtos ou serviços tendem a se relacionar (competirem entre si) com maior frequência (densidade).

Além disso, reduz-se significativamente a demanda por processamento computacional ao se manipular redes sociais menores e menos complexas (sobretudo pela redução da quantidade de conexões).

Por fim, tomou-se emprestado das Redes Sociais a técnica de ranqueamento, a qual foi utilizada para a determinação da importância relativa das empresas participantes das licitações analisadas. Para isso, adotou-se o método *PageRank*².

Este método foi desenvolvido visando avaliar a importância de uma página da WEB, tendo como premissa subjacente a hipótese de que páginas relevantes tendem a ser alcançadas por uma maior quantidade de *links*. De um modo geral, este método atribui maior relevância a uma página que:

1. É alcançada por uma grande quantidade de *links*;
2. É alcançada por páginas relevantes; e
3. Aponta para páginas relevantes.

O peso acumulado pela página é maior no primeiro item e decresce na direção do terceiro item (da lista acima). O grafo representado no Gráfico 3 teve os pesos (relevância) de seus nós determinados pelo algoritmo *PageRank*. Nesta figura é possível ter uma ideia mais clara do quanto cada uma das situações elencadas nos itens acima de fato influencia no *rank* atribuído a um nó. Nela é possível observar que neste método de ranqueamento a soma dos pesos de todos os nós de uma rede social é sempre 1 (ou 100%).

Ao aplicarmos este método de ranqueamento na rede social construída para representar os participantes das licitações contidas na base estudada, atribuiremos maior relevância às empresas que:

1. Sagram-se vencedoras com uma frequência acima da média, configurando uma grande concentração ou monopólio de mercado;
2. Vencem licitações das quais também participaram outras empresas relevantes (casos de rodízio de empresas que detém forte fatia de um determinado mercado e atuam de forma coordenada, alternando-se como vencedoras em licitações, são casos típicos em que esta situação acontece); e
3. Costumam participar de certames em que licitantes relevantes sagram-se vencedoras. Isto ocorre em situações em que o papel preponderante desempenhado por determinada empresa é o de figurante ou de “coelho” (uma empresa atua de forma a elevar os preços do item licitado sem ter, no entanto, a real intenção vencer a licitação).

Portanto o método utilizado para identificar possíveis situações de formação de cartéis em licitações consistiu na:

- Criação de grafos (redes sociais) direcionadas contendo informações de participantes (vencedores e perdedores) de todas as licitações estaduais contidas na base de dados de licitações e contratos da administração pública estadual no período de jan/2011 a ago/2015;
- Segmentação da rede social construída na etapa anterior por meio de técnicas de detecção de comunidades (clusterização);
- Ranqueamento dos nós (empresas) das redes sociais segmentadas com base no algoritmo *PageRank*; e
- Seleção das redes sociais em que no máximo 20% das empresas possuem relevância acumulada

2. **PageRank** é um algoritmo usado para classificar *sites* em resultados de pesquisas. Foi concebido por Larry Page, um dos fundadores do *Google*.

superior a 40%, indicando uma elevada concentração do mercado em uma pequena quantidade de fornecedores.

Esta abordagem permitiu a identificação de dez situações (redes sociais) com alta probabilidade de constituírem cartéis de empresas, conforme mostra a Tabela 1.

A título de ilustração mostraremos três das dez situações acima encontradas as quais têm forte concordância com os resultados obtidos por meio da abordagem de Regras de Associação. As situações apresentadas foram destacadas na Tabela 2.

Tabela 2 - Relação das situações candidatas encontradas.

Id_Situação	Num_Licitacoes	Num_Empresas	Rank_Acumulado
2	41	5	0,42
4	145	11	0,40
7	146	7	0,41
10	529	16	0,40
11	61	6	0,43
12	37	3	0,42
14	13	3	0,44
17	33	8	0,40
51	17	2	0,43
61	9	3	0,44

Possível formação de carteis para o fornecimento de medicamentos

Foram encontradas duas situações de possível formação de cartel no fornecimento de medicamentos para o ERJ. Estes casos correspondem às situações 4 e 7 da Tabela 1.

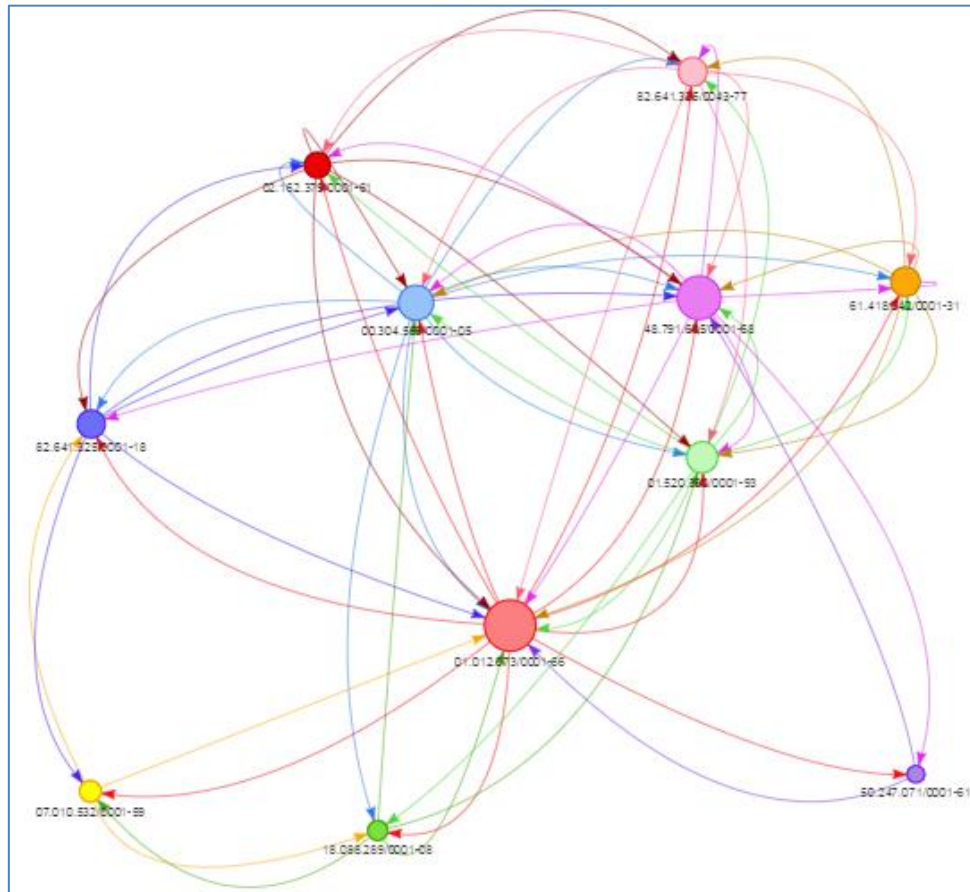
A rede social 4 é composta por 11 empresas. Vale frisar que 4 destas empresas coincidem com as empresas do Grupo 2 identificado pela abordagem de Regras de Associação. A Tabela 3 lista seus CNPJs bem como a relevância (*rank*) calculada para cada uma delas e as coincidências percebidas em relação ao Grupo 1.

Tabela 3 - Relação das empresas que compõem a rede social de medicamentos (Id_Situacao = 4 da Tabela 1).

CNPJ	NOME	RANK	COINCIDENCIA
01.012.073/0001-66	HOUSE MED PRODUTOS FARMACEUTICOS E HOSPITALARES LTDA	19,31%	X
48.791.685/0001-68	CBS MÉDICO CIENTÍFICA COMÉRCIO E REPRESENTAÇÃO LTDA	15,62%	X
00.304.559/0001-05	HUNTER CIENTIFICA COMERCIAL E SERVIÇOS LTDA	11,48%	X
01.520.390/0001-93	MADRIMED PRODUTOS MEDICOS LTDA	9,60%	
61.418.042/0001-31	CIRURGICA FERNANDES LIMITADA	8,68%	X
82.641.325/0043-77	CREMER S/A	8,45%	
82.641.325/0001-18	CREMER S/A	7,97%	
02.162.379/0001-61	TOP LINE COMERCIAL LTDA ME	7,00%	
07.010.532/0001-59	J.PINHEIRO MATERIAIS MEDICOS E ODONTOLOGICOS LTDA.	4,92%	
18.088.289/0001-08	DENT SERV COMÉRCIO E SERVIÇOS CORRELATOS DE SAÚDE LTDA	4,09%	
50.247.071/0001-61	MOGAMI IMPORTACAO E EXPORTACAO LTDA	2,88%	

O Gráfico 5 representa a rede social do possível cartel de fornecedores de medicamentos, correspondente à rede social detalhada na Tabela 2.

Gráfico 5 - Rede social representando possível cartel de empresas fornecedoras de medicamentos. O tamanho dos nós é proporcional ao *rank* (terceira coluna da Tabela 2)



A rede social 7 é composta por 7 empresas. Deste total, 3 empresas coincidem com as empresas do Grupo 2 identificado pela abordagem de Regras de Associação. A Tabela 4 lista seus CNPJs bem como a relevância (*rank*) calculada para cada uma delas e as coincidências percebidas em relação ao Grupo 1.

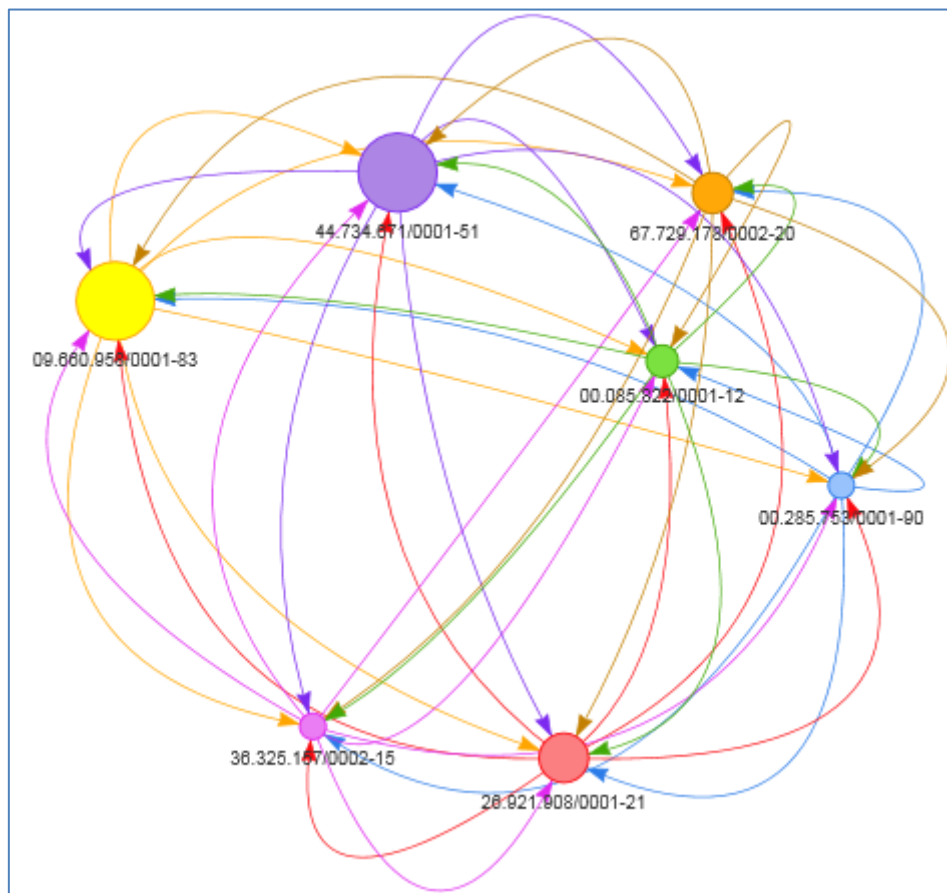
Tabela 4 - Relação das empresas que compõem a rede social de medicamentos (Id_Situacao = 7 da Tabela 1)

CNPJ	NOME	RANK	COINCIDENCIA
44.734.671/0001-51	CRISTÁLIA PRODUTOS QUÍMICOS FARMACÊUTICOS LTDA	21,99%	
09.660.958/0001-83	S3MED DISTRIBUIDORA DE MEDICAMENTOS LTDA	21,82%	X
26.921.908/0001-21	HOSPFAR INDUSTRIA E COMÉRCIO DE PRODUTOS HOSPITALARES LTDA	14,81%	X
67.729.178/0002-20	COMERCIAL CIRÚRGICA RIOCLARENSE LTDA	12,64%	
00.085.822/0001-12	ESPECIFARMA COMERCIO DE MEDICAMENTOS E PRODUTOS HOSPITALARES LTDA	10,54%	X

36.325.157/0002-15	COSTA CAMARGO COMERCIO DE PRODUTOS HOSPITALARES LTDA (FILIAL)	9,12%	
00.285.753/0001-90	VENANCIO PRODUTOS FARMACEUTICOS LTDA	9,08%	

O Gráfico 6 representa a rede social do possível cartel de fornecedores de medicamentos, correspondente à rede social detalhada na Tabela 3.

Gráfico 6 - Rede social representando possível cartel de empresas fornecedoras de medicamentos. O tamanho dos nós é proporcional ao *rank* (terceira coluna da Tabela 3)



Uma constatação interessante a ser mencionada é que numa comparação um pouco mais atenta entre as redes sociais aqui identificadas e o Gráfico 2, que representa todas as 21 regras selecionadas pela técnica de Regras de Associação percebe-se uma clara relação, respectivamente, entre as situações 4 e 7 as regiões esquerda e direita do agrupamento correspondente ao Grupo 1, definido naquela abordagem.

Possível formação de cartel para o fornecimento de suprimentos de informática

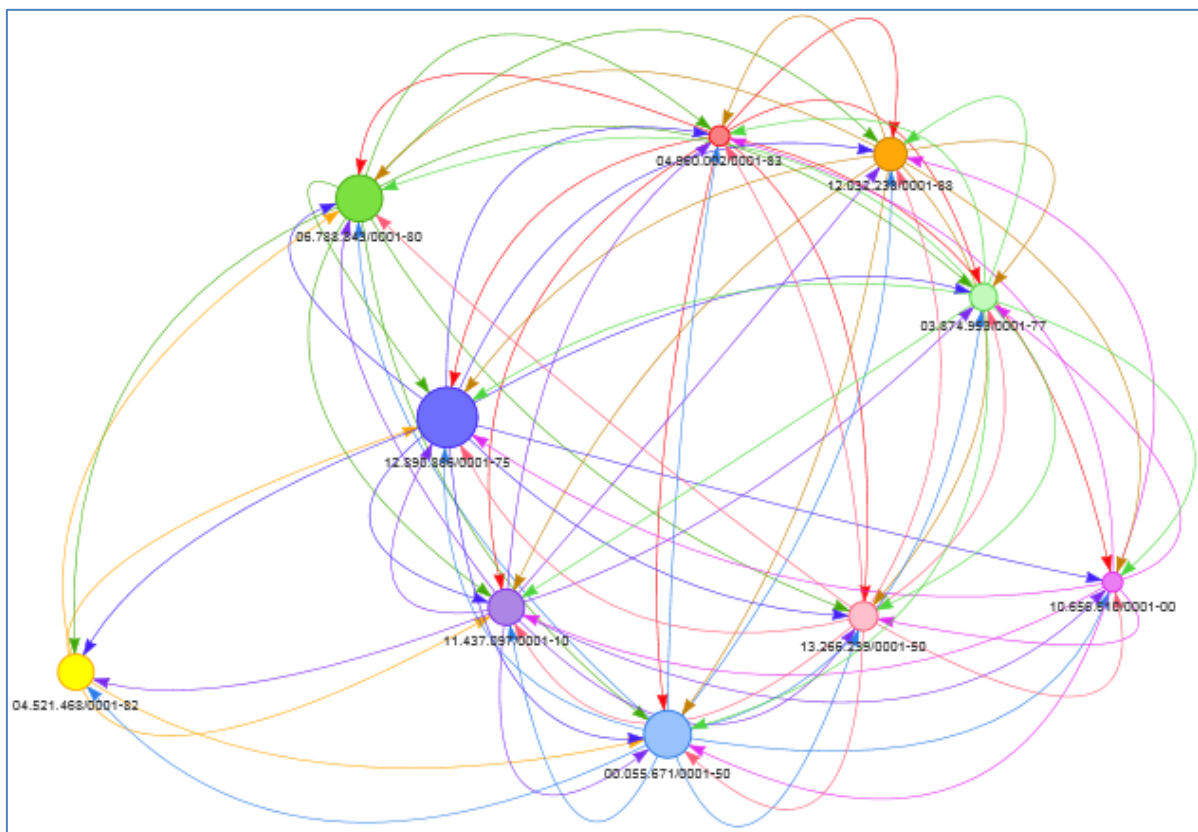
Este caso corresponde à situação 10 da Tabela 2. Esta rede social é composta por 16 empresas que atuam no ramo de suprimentos de informática. Destas empresas, 6 coincidem com as empresas do Grupo 2 identificado pela abordagem de Regras de Associação. A Tabela 5 lista seus CNPJs bem como a relevância (*rank*) calculada para cada uma delas e as coincidências percebidas em relação ao Grupo 2.

Tabela 5 – Relação das empresas que compõem a rede social de suprimentos de informática (id_Situação = 10 da Tabela 1). As empresas sombreadas estão representadas no grafo do Gráfico 3.

CNPJ	NOME	RANK	COINCIDÊNCIA
12.890.886/0001-75	VIXNU COMERCIO LTDA-EPP	14,58%	X
00.055.671/0001-50	VAN MEX COMERCIAL E SERVIÇOS LTDA ME (ANTIGA) PAPELARIA VAN MEX LTDA	10,16%	
06.788.843/0001-80	BENEDITINOS BAZAR E PAPELARIA LTDA.	8,94%	X
11.437.097/0001-10	2W COMERCIO DE INFORMÁTICA LTDA	7,83%	
12.032.238/0001-88	EMAV MATERIAIS E SERVIÇOS LTDA.	7,32%	X
13.266.239/0001-50	MAPPE BRASIL LTDA	6,64%	
04.521.468/0001-82	LEMAR INK FRANQUIAS LTDA	6,27%	X
03.874.953/0001-77	SIERDOVSKI & SIERDOVSKI LTDA-ME	5,73%	
06.011.105/0001-22	AP NET INFORMATICA LTDA	5,18%	
10.656.610/0001-00	ARDO DO BRASIL DISTRIBUIDORA DE PECAS E SERVICOS LTDA	4,84%	
04.960.002/0001-83	DIBOA COMERCIAL LTDA	4,42%	
07.848.715/0001-48	K S MAX INFORMÁTICA LTDA.	4,21%	X
07.324.021/0001-01	ROS RIO MATERIAIS E COMÉRCIO LTDA-EPP (ANTIGA ROS RIO MATERIAIS E REPRESENTAÇÕES LTDA EPP)	4,07%	
00.710.985/0001-49	VIBHUTI COMERCIO LTDA-ME	4,00%	X
15.153.416/0001-17	J N S COMERCIO E SERVIÇOS LTDA	3,60%	
01.521.643/0001-43	DANJAC DISTRIBUIDORA LTDA.	2,22%	

O Gráfico 7 representa a rede social do possível cartel de fornecedores de suprimentos de informática, contemplando as 10 empresas mais relevantes segundo o *PageRank* (Tabela 4).

Gráfico 7– Rede social representando possível cartel de empresas fornecedoras de suprimentos de informática. Esta rede foi ligeiramente simplificada ao se retirar do grafo as 6 empresas de menor relevância, segundo o *PageRank*, a fim de tornar sua leitura mais clara. O tamanho dos nós é proporcional ao *rank* (terceira coluna da Tabela 3).



Diante dos resultados aqui apresentados e de sua comparação com os resultados obtidos por meio da técnica de Regras de Associação, é possível chegar às seguintes conclusões preliminares, a serem confirmadas por pesquisas mais detalhadas:

1. A forte coincidência percebida nos produtos oriundos de duas abordagens distintas (Análise de Redes Sociais e Regras de Associação), completamente independentes, reforça mutuamente a validade dos resultados obtidos;
2. Em relação à abordagem de Regras de Associação, a utilização de ferramentas de análise de Redes Sociais, proposta neste trabalho, aparentemente é mais robusta (identificando um maior número de situações candidatas) e mais seletiva (identificando redes sociais que correspondiam a subgrupos dos grupos identificados pela outra abordagem).

CONCLUSÕES

Análise da Pesquisa e Conclusões

A pesquisa evidenciou ser possível a incorporação de técnicas de mineração de dados nas auditorias realizadas pelo TCE-RJ. Embora factível, diversos obstáculos se opõem a este objetivo, sendo o principal obstáculo a pouca familiaridade dos técnicos com as técnicas de mineração de dados e com os softwares necessários à implementação destas técnicas. Este fato exigirá grande investimento em capacitação, tanto nas técnicas quanto em softwares adequados à implementação das mesmas.

Os resultados alcançados, materializados nas indicações de risco de irregularidades, ainda que necessitando de comprovação, podem ser um importante subsídio ao modelo de controle atualmente em desenvolvimento pela Secretaria Geral de Controle Externo - SGE, denominado Controle Contínuo de Atos – CCA, modelo este que toma por bases dados de licitações e contratações provenientes dos sistemas de informação internos e dos jurisdicionados.

Como dito, os resultados obtidos necessitam serem validados para que se verifique a necessidade de se aperfeiçoar os métodos utilizados. Esta validação poderá ser feita mediante a realização de auditorias nas licitações das quais as empresas identificadas tivessem participado.

Sugestões

O potencial para a utilização de técnicas de mineração de dados em auditoria é promissor. Reconhecendo este fato, o Tribunal de Contas da União explicita em seu plano estratégico de 2015-2021 “a importância de se desenvolver capacidade organizacional ampla para trabalhar com recursos tecnológicos emergentes e analisar grandes bases de dados (*big data*)”.

Objetivando atingir este objetivo “em 2015, o ISC e a Seginf desenvolveram várias ações com o objetivo estratégico de apoiar o desenvolvimento de capacidade organizacional, dentre elas podemos citar: a aquisição de grande quantidade de livros e materiais didáticos digitais sobre esse tema pela Biblioteca, a manutenção e incentivo de uma comunidade de prática, a análise sobre a viabilidade de um curso de pós-graduação em análise de dados para o controle; e o desenvolvimento e a oferta do Programa de Capacitação em Análise de Dados. No referido Programa, dentre outras ações, foram capacitados mais de 100 servidores em análise de dados utilizando a linguagem R”.(TCU, 2016)

Estas são, sem dúvidas, ações que ajudariam o TCE-RJ a incorporar o uso de técnicas de mineração de dados em suas auditorias.

Nota: Instituto Serzedelo Correa (ISC) e Secretaria de Gestão de Sistemas e de Informação para o Controle (Seginf) são unidades do TCU. O ISC tem paralelo com a Escola de Contas e Gestão do TCE-RJ e a Seginf não tem paralelo na estrutura do TCE-RJ.

Pesquisas Futuras

O conjunto de dados utilizado mostrou-se adequado para a realização de testes utilizando outras técnicas de mineração de dados. As técnicas utilizadas nesta pesquisa foram essencialmente técnicas descritivas já que não se dispunha de uma variável que classificasse a licitação em irregular ou não. Não obstante, pesquisa nos bancos de dados do TCE-RJ por auditorias ou análises processuais podem indicar licitações e contratos que tenham sido considerados irregulares, podendo estas informações serem acrescidas às bases de dados.

Também se podem fazer pesquisas na internet acerca de empresas que tenham se envolvido em fraudes, ou empresas que tenham sido declaradas inidôneas e impedidas de participar em licitações públicas.

Outra área onde pode ser possível avançar a partir desta pesquisa é na implementação das auditorias preditivas. Na visão de KUENKAIKAEW e VASARHELYI (2013), a auditoria preditiva é uma abordagem de auditoria contínua que utiliza metodologia preditiva na realização das auditorias. A metodologia preditiva consiste em técnicas de mineração de dados. Em resumo, trata-se de utilizar técnicas de mineração de dados com o objetivo de prever riscos, tendências, nível e fluxos de controle e outros parâmetros do processo de negócio.

Não obstante, nada disso poderá ser implementado de forma satisfatória caso o TCE-RJ não possua um elevado nível de maturidade analítica. A definição de uma escala de maturidade analítica, a identificação do nível de maturidade analítica do TCE-RJ e a proposição de medidas que visem aumentar o nível de maturidade analítica identificado é uma necessidade que precisará ser satisfeita em algum momento, caso se deseje de fato um Tribunal mais eficiente.

REFERÊNCIAS

- ALBRECHT, C. C.; ALBRECHT, W. STEVE (2002): *Strategic fraud detection: a technology-based model* – Rolins Center for eBusiness, Working paper.
- ALBRECHT, C. C.; W. STEVE ALBRECHT; DUNN, J.G. (2001): *Conducting a pro-active fraud audit: a case study*, Journal of Forensic Accounting vol. II p. 2013-218
- ALBRECHT, CONAN. C. (2003): *Audit at a crossroads* – In 35th Symposium on the Interface. Salt Lake City, UT
- BAENSENS, B.; VLASSELAER, V.V.; VERBEKE, W. (2015): *Fraud analytics: using descriptive, predictive, and social network techniques: a guide to data science for fraud detection* – Wiley
- BALANIUK, R. (2010): *A mineração de dados como apoio ao controle externo*, Revista do TCU, vol. 117, p. 79-86
- BODINI JÚNIOR, A. C. (2009): *Utilização de técnicas de data mining na detecção de outliers em auxílio à auditoria operacional com um estudo de caso com dados do sistema de informações hospitalares*; Tese de Doutorado, COPPE UFRJ
- CARVALHO, JOSÉ CARLOS de O. (2007): *Por dentro das fraudes: como são feitas, como denunciá-las, como evitá-las*, Lex Editora, São Paulo.
- CARVALHO, R. N. (2014): *Banco de preços: metodologia par cálculo de preços médios nas compras do governo brasileiro*, XIX Congresso Internacional del CLAD sobre la Reforma del Estado y de la Administracion Pública, Quito, Equador.
- DEBRECENY, R; GRAY, GLEN, L. (2013): *A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits*, University of Waterloo Symposium on Information Integrity and Information Systems Assurance.
- KOSKIVAARA, E. (2007): *Integrating analytical procedures into the continuous audit environment*. *Journal of Information Systems and Technology Management*, Vol 3, No 3, p. 331-346
- MARTINS JUNIOR, A.; BRAZ, MÁRCIO R. (2010): *O controle externo por meio de bases de dados*, Revista do TCU, vol. 117, p. 69-78
- PHUA, C.; LEE, V.; SMITH, K.; GAYLER, R. (2010): *A comprehensive survey of data mining based fraud detection research*.
- SALES, L. J., CARVALHO, R. S. (2014): *Análise multivariada de dados aplicada na previsão de irregularidades em entratos do governo brasileiro*, XIX Congresso Internacional del CLAD sobre la Reforma del Estado y de la Administracion Pública, Quito, Equador.
- SILVA, C. V. S.; ROCHA, H. A. da (2010): *Técnicas de mineração de dados como apoio às auditorias governamentais*, Revista da CGU, Ano V, Nº 08 out/2010 p. 29-40,
- TCU (2016): *Boas razões para participar do Diagnóstico de Competências em Análise de Dados*
url:< <http://portal.tcu.gov.br/innovatcu/noticias/boas-razoes-para-participar-do-diagnostico-de-competencias-em-analise-de-dados.htm>> Acesso em 25/02/2016.
- TÓTH, B.; FAZEKAS, M.; CZIBIK, A.; TÓTH, I. J. (2015): *Toolkit for detecting collusive bidding in public procurement: with examples from Hungary*; Corruption Research Center – Budapest
- TRAVAILLE, P. (2011): *Eletronic fraud detection in the U.S. medicaid healthcare program*, MSc Thesis – University of Twente.










VERGARA, S. C. (2014): *Projetos e relatórios de pesquisa em administração*, São Paulo, Atlas, 15ª Edição.




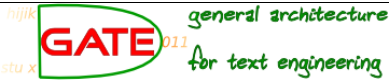


APÊNDICES

APENDICE I – RELATÓRIO DE FERRAMENTAS DE MINERAÇÃO DE DADOS

Este apêndice constitui um breve inventário dos softwares disponíveis que podem ser utilizados no TCE-RJ com vistas a implementar tarefas de análise de dados.

Deu-se preferência por inventariar *softwares open source*, visto que estão prontamente disponíveis para uso sem a necessidade de investimentos de recursos financeiros.

	<p>Nome: R url: https://www.r-project.org/</p>
	<p>Nome: Orange url: http://orange.biolab.si/</p>
	<p>Nome: Knime url: https://www.knime.org/</p>
	<p>Nome: Weka url: http://www.cs.waikato.ac.nz/ml/weka/</p>
	<p>Nome: Rattle url: http://rattle.togaware.com/</p>
	<p>Nome: Scikit-learn url: http://scikit-learn.org/stable/index.html</p>
	<p>Nome: RapidMiner url: https://rapidminer.com/</p>
	<p>Nome: Gephi url: http://gephi.github.io/</p>
	<p>Nome: Pajek url: http://mrvar.fdv.uni-lj.si/pajek/</p>

	<p>Nome: neo4j url: http://neo4j.com/</p>
	<p>Nome: MongoDB url: https://www.mongodb.org/</p>
	<p>Nome: SQLite url: https://www.sqlite.org/</p>
	<p>Nome: NLTK url: http://www.nltk.org/</p>
	<p>Nome: Gate url: https://gate.ac.uk/</p>
	<p>Nome: ProM url: http://www.promtools.org/doku.php</p>
	<p>Nome: Python url: https://www.python.org/</p>

APENDICE II – ARQUIVOS UTILIZADOS NA PESQUISA

Relação dos arquivos de dados utilizados nesta pesquisa.

Arquivos obtidos junto à SEPLAG

Arquivo	Descrição
RELATORIOS_TCE_2015.10.19 - Tab_Contratacao.csv	Arquivo contendo dados relativos às contratações realizadas.
RELATORIOS_TCE_2015.10.19 - Tab_Itens.csv	Arquivo contendo dados relativos aos itens licitados.
RELATORIOS_TCE_2015.10.19 - Tab_Participantes.csv	Arquivo contendo dados relativos aos participantes das licitações.
RELATORIOS_TCE_2015.10.19 - Tab_RDC_BID_BIRD.csv	Arquivo contendo dados relativos às licitações realizadas sob o regime diferenciado de contratações com recursos provenientes do BID e do BIRD.
Tab_Licitacao.csv	Arquivo contendo dados relativos às licitações realizadas pelo executivo estadual entre .jan/2011 e ago/2015

Arquivos obtidos junto à JUCERJA

Arquivo	Descrição
JUCERJA.txt	Arquivo contendo informações sobre as empresas do Estado do Rio de Janeiro, inclusive a composição acionária das mesmas.

APENDICE III –SÍNTESE DOS RESULTADOS OBTIDOS NA ANÁLISE DAS TIPOLOGIAS COMPLEMENTARES

Tipologia	Descrição	Resultado	Métodos Utilizados
Ausência de competição em licitação realizadas	Existência, nas empresas participantes da licitação, de sócios em comum , com indício de simulação licitatória, fraude e violação ao sigilo das propostas	Foram identificadas 99 licitações realizadas por órgãos do Estado em que tal situação ocorreu. Destas, 34 foram vencidas por empresas que continham sócios ou controladores em comum com alguma concorrente no mesmo certame	Análise de Redes Sociais (Grafos Bipartidos)
Ausência de competição em licitação realizadas	Existência, nas empresas participantes da licitação, de empresas com endereço em comum , com indício de simulação licitatória, fraude e violação ao sigilo das propostas	Foram identificadas 46 licitações realizadas entre 2010 e 2014 por órgãos do Estado que foram vencidas por empresas que continham endereços em comum com alguma concorrente no mesmo certame	Análise de Redes Sociais (Grafos Bipartidos) Mineração de Texto (TF-IDF)
Fracionamento de Despesas	Planejamento prévio inadequado dos gastos anuais, permitindo a ocorrência de fracionamento de despesas de mesma natureza pela não observância do valor limite das modalidades licitatórias dado que o mesmo é cumulativo ao longo de um exercício financeiro	Ao longo dos exercícios de 2012 a 2014, foi contabilizado um total de R\$ 3.419.346,17 em despesas fracionadass	Mineração de Texto (TF-IDF)

PARECER DO PESQUISADOR-AVALIADOR

A pesquisa teve o objetivo de demonstrar a aplicabilidade de ferramentas de análise de dados a auditorias a cargo deste Tribunal de Contas. O trabalho contou com forte suporte de *softwares* específicos, com código aberto e/ou gratuito, o que ressalta a proposta de aplicação nos procedimentos desta Casa, já que nenhum custo relevante seria necessário, no que tange à aquisição de aplicativos, para a imediata utilização das ferramentas aqui demonstradas.

Técnicas e métodos de análise de dados são o que de mais avançado se tem hoje na detecção de fraudes e indícios de irregularidades, dentre outros aspectos. A adoção dessas técnicas, pelo TCE-RJ, permite, ainda, prever determinados desvios em potencial.

A equipe de pesquisadores trabalhou com afinco na proposta, com a certeza de que a aplicação das técnicas apresentadas traria resultados efetivos no planejamento de auditorias governamentais. Ela realmente acredita da proposta. Assim, os resultados foram, de fato, auspiciosos.

Cabe destacar que os pesquisadores já aplicam algumas das técnicas nas auditorias deste TCE-RJ, com resultados igualmente excelentes.

Diante de tudo, tenho a certeza de que a equipe de pesquisadores atingiu o objetivo proposto e que ele trará muitos benefícios ao processo de trabalho desta Casa, na realização de auditorias governamentais e nos trabalhos de inteligência de controle externo.

Pesquisador-avaliador

Nome: MARCONI CANUTO BRASIL

Rio de Janeiro, ____ de _____ de ____.

(assinatura do Pesquisador-avaliador)