

# Cor, Tamanho, Forma e Posição: elementos-Chave da Análise Exploratória de Redes Sociais

Manoel Camilo de Sousa Netto<sup>1</sup>; Adilson Luiz Pinto<sup>2</sup>

## Introdução

A análise exploratória visual é uma técnica apropriada para obter *insights* (intuições) sobre redes, especialmente quando leiautes de grafos são associados a características exteriorizadas nos vértices e nas arestas, tais como cores, tamanhos, formas e posições. Esses atributos, quando transfazem o visual dos grafos, não alteram o âmago da rede: os dados e as estatísticas afetas permanecem incólumes. São revelações de facetas ocultas ao usuário contemplador, olhares diferentes reiteradamente sonogados pelos dados massivos de natureza textual e tabular. Em uma análise exploratória visual, os *insights* são obtidos por aspectos aparentes de assimilação simplificada. A percepção seletiva dessas intuições dá-se pelo uso de habilidades rudimentares e sumárias que, por óbvio, serão alvo de crivo quantitativo e confirmatório, esses mais formais e elaborados. O objetivo do presente trabalho, de natureza quali-quantitativa, é demonstrar como, por meio de

---

1 UFSC, Polícia Federal do Brasil. [manoelcamilo@gmail.com](mailto:manoelcamilo@gmail.com)

2 UFSC. [adilson.pinto@ufsc.br](mailto:adilson.pinto@ufsc.br)

elementos-chave, aplicar a análise exploratória visual para obter *insights* em grafos. Como estudo de caso, o estudo aplica as técnicas para revelar anomalias em uma rede de contratações estatais. Essas anomalias indicam cenários mais vulneráveis a crimes de corrupção.

## **Fundamentação teórica**

A fundamentação teórica apresenta alguns conceitos tais como análise de redes sociais e centralidade de grau, correlação e regressão, análise exploratória de dados e em grafos, elementos-chave na análise exploratória e leiautes geográficos. Também apresenta os conceitos de estatísticos de correlação e regressão, usados como supedâneo nas confirmações quantitativas.

### **Análise exploratória de dados**

Segundo Pinheiro (2009, p. 12), a análise exploratória é um conjunto de técnicas de tratamento de dados que, sem implicar em uma fundamentação matemática mais rigorosa, contribui na sondagem do terreno, ou seja, na obtenção de um primeiro contato com a informação disponível. É livre o uso de recursos visuais, tais como grafos e gráficos. O caso concreto ditará quais ferramentas visuais são mais adequadas.

Sobre a análise exploratória de dados, Bernabeu, Castro e Godino (1991, p. 2) afirmam:

Como indicamos, nos deparamos com uma nova filosofia na aplicação de métodos de análise de dados, embora,

juntamente com ela, algumas técnicas específicas tenham sido desenvolvidas para sua aplicação. Essa filosofia consiste no estudo de dados de todas as perspectivas e com todas as ferramentas possíveis, inclusive as existentes. O objetivo é extrair o máximo de informação possível, gerar novas hipóteses, no sentido construir conjecturas sobre as observações que dispomos.

Bernabeu, Castro e Godino (1991, p. 1) afirmam que antes da análise exploratória a análise dos dados baseava-se principalmente no cálculo estatístico, levando diminuição da importância visual da representação dos dados, e atribuindo-a exclusivamente aos cálculos; em segundo lugar, a análise foi equiparada ao modelo confirmatório. Essa afirmação indica que os estudos recaíam imediatamente sobre os cálculos e as métricas, o que vai de encontro a real intenção preliminar, que não é confirmar, mas sim conjecturar. As percepções visuais iniciais, muito poderosas, eram deixadas ao segundo plano. Enxergar uma rede em determinados leiautes e cujas entidades assumem tamanhos, cores, formas e posições pode ser uma técnica exploratória eficaz que potencialmente guie os cálculos confirmatórios posteriores.

#### **Elementos-chave na análise exploratória de grafos: cor, tamanho, forma e posição**

Quando se projeta uma visualização de rede, o objetivo é responder algumas hipóteses que versam sobre a temática do campo de estudo. O aspecto visual deve destacar propriedades que potencialmente podem levar a conclusões

acerca dos temas de interesse. Essas conclusões devem ser destacadas por elementos de natureza visual. Eles são bem descritos por Ognyanova (2019, p. 3), quando afirma que nas redes, há vários elementos-chave que controlam o resultado. Os principais são cor, tamanho, forma e posição, com uma menção honrosa para as setas que apontam a direção e os *labels* que contêm identificação, conforme indica a na Figura 1 a seguir:

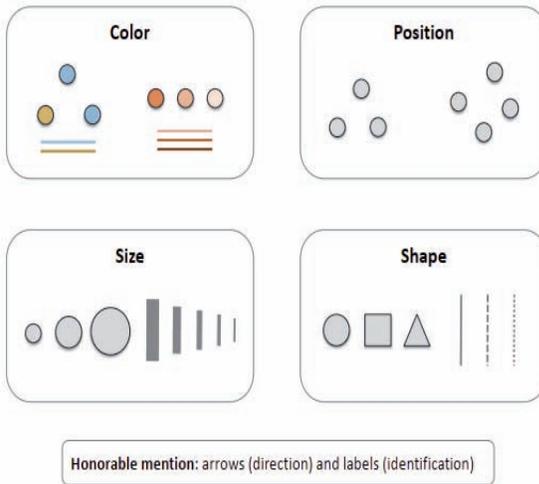


Figura 1. Controles de visualização de rede  
Fonte: extraído de Ognyanova (2019, p. 3)

### Análise exploratória em grafos

Uma das técnicas de análise exploratória é o uso de leiautes de grafos. Eles representam os formatos que essas redes podem assumir em um espaço de plotagem e que têm a aptidão de dotá-lo de algum conteúdo semântico captável pela análise exploratória, a depender da estética. Cherven (2015, p. 28) descreve a importância dos leiautes:

A seleção de um layout apropriado pode fazer a diferença entre criar um grafo impenetrável, que não comunica uma história, e uma visualização facilmente acessível que não apenas comunica, mas também tem apelo estético. Ao ler a literatura de grafos de rede, é provável que se depare com o termo *hairball*, uma descrição para uma rede muito densa com muitas conexões que são praticamente indecifráveis quando se usa os algoritmos comuns de grafos.

Em raciocínio acerca da melhor estética dos grafos, Hu (2006, p. 37) afirma:

A utilidade da representação depende da estética do desenho. Embora não existam critérios estritos para a estética, aceita-se que o mínimo cruzamentos entre arestas, vértices uniformemente distribuídos e a representação da simetria gráfica são desejáveis.

O uso dos leiautes é uma estratégia de análise de redes representadas por grafos na busca por simetria, menos cruzamentos de ligações (arestas) e distribuições uniformes. Essas características podem conferir significado a uma rede que, representada de modo aleatório, pouco informa em seu aspecto visual. A pesquisa conhece da existência das métricas de rede e que elas podem ser obtidas facilmente em formato tabular por qualquer software de análise de redes sociais. Esse formato, entretanto, não expressa semântica obtida de forma rudimentar, concentrada, intuitiva e inaugural. Essa é uma habilidade, entretanto, afeta à análise exploratória.

## Leiautes geográficos

Dentre os vários leiautes de grafos disponíveis, os nominados como geográficos aplicam uma abordagem que permitem o uso de georreferenciamento e redes. São grafos sobrepostos a mapas que podem tirar proveito da capacidade inata dos usuários de visualizar e interpretar dados espaciais. Os leiautes geográficos, entretanto, são apenas uma das opções disponíveis. Eles foram escolhidos tão-somente por seu uso no estudo de caso, então qualquer leiaute que possa evidenciar as características temáticas pode substituí-lo.

Para que os vértices sejam projetados em um leiaute geográfico, os dados de latitude e longitude devem estar presentes entre os seus atributos. Abaixo, a Figura 2 ilustra a malha aérea dos Estados Unidos da América usando um grafo onde os aeroportos são vértices e as arestas são as rotas aéreas que os interligam.

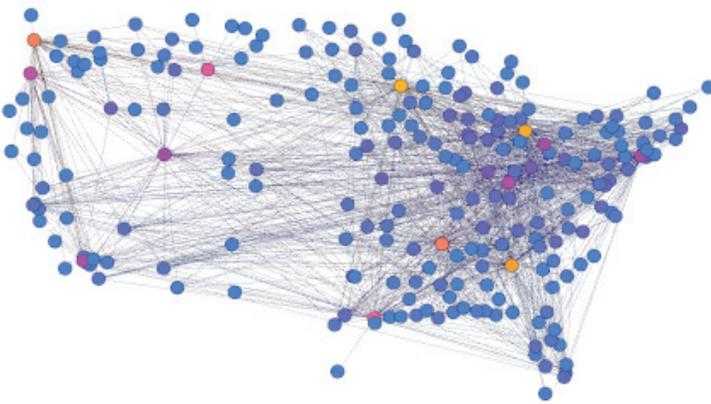


Figura 2. Leiaute geográfico de malha aérea dos Estados Unidos  
Fonte: Extraído de Heymann (2011)

Outra possibilidade de visualização de redes é o uso da composição de dois leiautes em um só grafo, de forma que as vantagens de ambos podem ser concentradas em uma única rede. Por exemplo, um leiaute geográfico pode ser combinado a um leiaute circular.

### **Análise de redes sociais e centralidade de grau**

Segundo Latora, Nicosia e Russo (2017, p. 64), a análise de redes sociais (ARS) é a disciplina que fornece métodos e ferramentas quantitativos para entender os sistemas sociais, representando-os e estudando-os como redes. Robins (2015, p. 13) também afirma que para as redes sociais, as entidades estudadas são mais frequentemente pessoas, embora possam ser outros atores sociais, como organizações. Uma rede social pode ser representada por um grafo, entretanto essa é uma das formas disponíveis para ilustrar uma rede. Portanto, o grafo não é a rede social em essência, mas sim uma forma de representá-la.

Wasserman e Faust (1994, p. 3) afirmam que, do ponto de vista da análise de redes sociais, o ambiente social pode ser expresso como regularidades nas relações entre unidades de interação. Essas unidades podem possuir graus de importância variáveis dentro da rede. A centralidade é um conjunto de métricas que representam algumas formas de obter esses graus de importância, pois as medidas de centralidade indicam os atores mais importantes na rede.

Segundo Latora, Nicosia e Russo (2017, p. 31) o conceito de centralidade e as primeiras medidas relacionadas ao tema foram introduzidas no contexto da ARS. Sobre centralidade de atores, Wasserman e Faust (1994, p. 3)

afirmam que atores proeminentes são aqueles que estão amplamente envolvidos no relacionamento com outros atores. Esse envolvimento os torna mais visíveis.

Erciyés (2015, p. 71) afirma que a centralidade é uma medida da importância de um vértice ou uma aresta em uma rede complexa. Assim, atribui-se um valor de importância para cada vértice ou aresta com base em sua posição topológica na rede. Segundo Gündüz-Ögüdücü & Etaner-Uyar (2014, p. 4), a centralidade mede a importância relativa de um vértice e fornece uma indicação sobre a sua influência na rede, sendo as seguintes suas variações: centralidade de grau, centralidade de intermediação, centralidade de proximidade e centralidade de autovetor. A pesquisa ora conduzida utiliza apenas a centralidade de grau como fundamento. É necessário, portanto, o entendimento acerca do conceito da métrica nominada grau e das suas variações.

Muitas métricas da ARS estão relacionadas ao grau de um vértice, cujo conceito Caldarelli (2007, p.14) expõe como sendo o número de arestas de um vértice, ao tempo em que esclarece que a soma de todos os graus em um grafo é o dobro do número de suas arestas. Isso ocorre porque cada aresta contribui duas vezes na contagem do grau: uma unidade para cada um dos vértices aos quais se conecta. Dada uma matriz de adjacência  $A(n,n)$ , Caldarelli (2007, p. 14) relata que o grau pode ser expresso por:

$$k_i = \sum_{j=1,n} a_{i,j}$$

Latora, Nicosia e Russo (2017, p. 31) afirmam que, se o grafo for direcionado, o grau do vértice possui dois

componentes: o número de links de saída  $k_i^{out}$  (grau de saída ou *Outdegree*), e o número de links de entrada  $k_i^{in}$  (grau de entrada ou *Indegree*) do nó  $i$ . O grau total do nó é então definido como:

$$k_i = k_i^{out} + k_i^{in}$$

Nos grafos ponderados –cuas arestas possuem pesos– também podem ser computados o *Weight Indegree* e o *Weight Outdegree*. Caldarelli (2007, p. 14) afirma que essas duas métricas são uma extensão do conceito de grau e contabilizam os pesos das arestas ao invés da simples contagem delas, sendo calculadas por:

$$k_i^w = \sum_{j=1,n} a_{ij}^w$$

A Figura 3 ilustra os conceitos dos diversos graus de centralidade (que acima foram apresentados de forma matemática) de forma concisa:

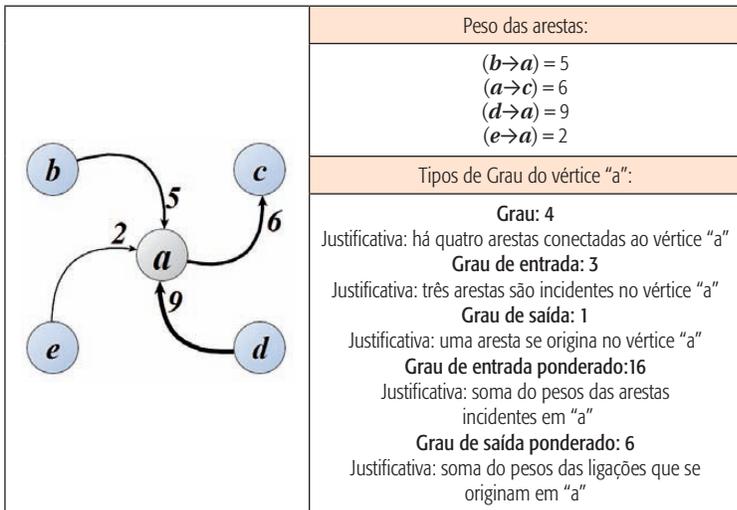


Figura 3. Diversos tipos de grau de um vértice. Fonte: Elaborada pelo autor

## Correlação e regressão

Os conceitos de correlação e regressão ora apresentados se relacionam com o estudo de caso, porque os métodos quantitativos disponíveis usados para confirmação da análise exploratória são diversificados e podem variar caso-a-caso.

Segundo Larson & Farber (2016, p. 446), uma correlação  $r$  é uma relação entre duas variáveis representadas por pares ordenados  $(x, y)$ , sendo  $x$  a variável independente (ou explanatória) e  $y$  a variável dependente (ou resposta). Afirma aquele autor que a variação do coeficiente de correlação é de  $-1$  a  $1$ , sendo que, quando  $x$  e  $y$  têm uma correlação linear positiva forte,  $r$  está próximo de  $1$ . Esse coeficiente pode ser calculado por:

$$r = \frac{n \sum xy - (\sum x) (\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$

Abaixo, a Figura 4 exemplifica alguns exemplos de correlações positivas:

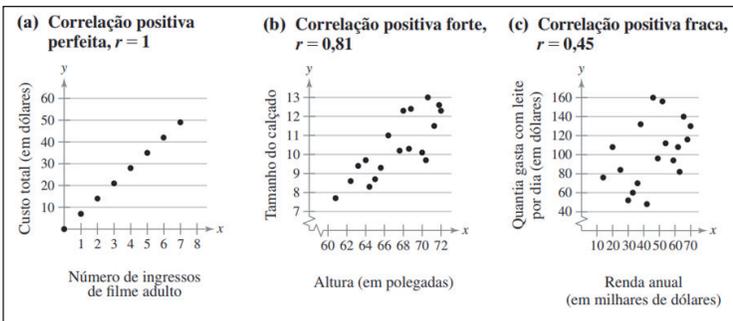


Figura 4. Tipos de Correlação entre duas variáveis  
Fonte: extraído e adaptado de Larson e Farber (2016, p. 442)

É possível traçar um gráfico da linha de regressão para duas variáveis correlacionadas. Segundo Agresti, Franklin e Klingenberg, B. (2018, p. 139) a linha de regressão prediz

o valor para a variável de resposta y e como uma função linear do valor x. Supondo  $y_n$  denotar o valor previsto de y, a equação para a linha de regressão, onde a indica a interceptação em y e b determina a inclinação da reta, conforme a seguir:

$$y_n = a + bx$$

Nesta fórmula, a denota a interceptação em y e b indica a inclinação. A Figura 5, abaixo, exemplifica uma reta de regressão entre duas variáveis com correlação negativa:

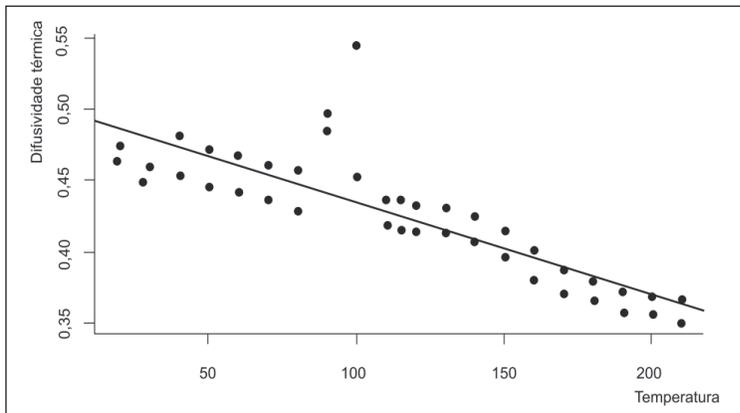


Figura 5.

Reta de regressão entre variáveis correlacionadas temperatura e difusividade térmica

Fonte: Extraído de Pinheiro, Carvajal, da Cunha, & Gomes (2012, p. 260)

## Metodologia

A metodologia aplicou as técnicas propostas a uma rede social de contratações públicas. Originalmente, essa rede é formada por milhares de vértices e arestas o que, em tese, dificulta a análise visual. Assim, antes de executar a análise exploratória, a rede foi filtrada segundo três cri-

térios que a tornam mais vulnerável. Para cada um desses critérios há uma justificativa baseada no empirismo policial. Após, na rede já filtrada, foram aplicadas métricas e atributos dos vértices aos elementos-chave de aspecto visual. A rede foi então alvo da análise exploratória, os insights foram obtidos e, por fim, submetidos à análise quantitativa confirmatória.

### **Escolha justificada para os três critérios de vulnerabilidade na contratação pública de empresas privadas**

Antes de citar e justificar os critérios, é relevante conhecer a rede social de contratações públicas e seus atores sociais componentes.

Uma **prefeitura** é o órgão governamental local responsável pela administração das cidades. Como representante do poder executivo local, cada uma delas acumula atribuições relacionadas à educação, saúde, preservação do patrimônio público, segurança, dentre outras responsabilidades legais. Como o rol de atividades públicas atribuída a esse ator social é diversificada, a ampla gama de bens e serviços exigíveis à sua atuação não pode ser suprida unicamente pela sua capacidade limitada. Assim, as prefeituras necessariamente interagem com outros atores sociais: as empresas privadas fornecedoras de bens e serviços, cujas áreas de atuação são tão diversificadas quanto a própria capacidade humana de empreender.

As prefeituras, portanto, contratam e realizam pagamentos a empresas privadas. Assim, os órgãos podem cumprir suas obrigações administrativas relacionadas à prestação de bens e serviços à população das cidades que

administram. **Prefeituras, pagamentos e empresas**, portanto, compõem os principais entes da rede social do estudo de caso. Essa rede pode apresentar vulnerabilidades que a tornem mais suscetível ao cometimento de crimes de corrupção relacionados às contratações das empresas. Em particular, as prefeituras da rede são todas de um único estado da federação brasileiro: o Piauí.

Crítérios de Vulnerabilidade (CVs) são pré-requisitos que, quando atendidos, tendem a deixar determinada rede social mais suscetível a uma determinada situação desfavorável. Para a rede de contratações estatais ora analisada, três Critérios de Vulnerabilidade (CV) foram escolhidos e justificados com base no empirismo policial. Cada um deles é explanado a seguir.

O primeiro CV é existência, no quadro societário de uma empresa contratada por uma prefeitura, de ao menos um sócio que possua acusação criminal prévia. Os dados coletados, antes mesmo de serem analisados como uma rede, revelaram que há um grande percentual de acusações criminais contra pessoas que outrora já foram acusadas (reincidentes de acusação). Para chegar a tal conclusão, foram analisados os dados das acusações criminais do estado do Piauí. Nesse repositório, das 1.327 (mil trezentas e vinte e sete) pessoas acusadas em pelo menos um procedimento policial<sup>3</sup>, 307 delas (trezentas e sete), ou seja, 23,14% do total, já possuíam outra acusação prévia, o que significa que são reincidentes nessas acusações. E, embora

---

<sup>3</sup> Inquérito policial é a nomenclatura usada para o procedimento policial que apura autoria e materialidade de crimes.

representem uma minoria do total de pessoas acusadas, os reincidentes respondem a um total de 1.056 (mil e cinquenta e seis) acusações, ou seja, 50,89% de um total de 2.075 (dois mil e setenta e cinco). Assim, na população considerada, os reincidentes, os quais representam 23,14% das pessoas respondem por mais de 50% de todas as acusações. A reincidência, portanto, se apresenta como um fator que aumenta a probabilidade criminal, considerando que as acusações recaem, em sua maioria, face a pessoas que foram anteriormente acusadas.

O segundo CV é o fato de que altos ganhos financeiros são um atrativo ao crime. Obter grandes quantias monetárias é um objetivo de grande parte dos criminosos como forma de compensação aos perigos penais e sociais decorrentes do delito (reprovação social, prisão e condenação). O criminoso que atua nas contratações públicas direciona seu comportamento a obter os maiores contratos, onde o retorno financeiro é compensatório aos temerários riscos. No Brasil é comum o uso de empresas privadas como beneficiárias de pagamentos integrais por serviços parcialmente prestados. Isso gera obras e serviços de má qualidade ou inacabados. A economia decorrente desse ardil é repartida entre empresários e gestores públicos em benefício próprio.

O terceiro CV é a hipercontratação como indicativo de vulnerabilidade. A motivação da escolha desse critério se relaciona com a alta seletividade das contratações públicas no Brasil. Contemporaneamente, ser contratado por inúmeros órgãos públicos é uma tarefa hercúlea, pois as seleções dos fornecedores, em regra, são realizadas por certames públicos, em geral de alta concorrência em que

os participantes disputam em igualdade de condições: as licitações.

Segundo Justen Filho (2014, p. 495):

A licitação é um procedimento administrativo disciplinado por lei e por um ato administrativo prévio, que determina critérios objetivos visando a seleção da proposta de contratação mais vantajosa e a promoção do desenvolvimento nacional sustentável, com observância do princípio da isonomia, conduzido por um órgão dotado de competência específica.

Assim, as empresas que colecionam muitos órgãos públicos como clientes são excepcionalíssimas. Entretanto, conluíus criminosos entre agentes públicos e privados podem impulsionar uma empresa a colecionar múltiplos contratantes estatais.

O 2º e o 3º CVs, ambos de natureza quantitativa utilizaram o conceito de *Outlier* Estatístico para selecionar as empresas que lhes atendem. Para entender o que é um *Outlier*, apresentam-se alguns conceitos a seguir. Devore (2006, p. 28) relata que os quartis dividem o conjunto de dados estatísticos em quatro partes iguais em que aproximadamente  $1/4$  dos dados recai sobre ou abaixo do primeiro quartil  $q1$ , metade dos dados sobre ou abaixo do segundo quartil  $q2$  (a mediana) e aproximadamente  $3/4$  dos dados sobre ou abaixo do terceiro quartil  $q3$ . As faixas delimitadas pelos três quartis, portanto, contém 25%, 50% e 75% do quantitativo geral de observações estatísticas. Larson e Farber (2016, p. 81) afirmam que a distância (ou

amplitude) interquartil ( $dq$ ) de um conjunto de dados é uma medida de variação que fornece a amplitude da porção central (aproximadamente metade) dos dados. A  $dq$ , portanto, é a diferença entre o terceiro e o primeiro quartis, conforme abaixo:

$$d_q = q_3 - q_1$$

Segundo Bussab e Morettin (2010, p. 48), *Outliers* são valores maiores do que o limite superior ( $Ls$ ) e menores que o limite inferior ( $Li$ ) os quais, por sua vez, são determinados respectivamente por:

$$\begin{cases} Ls = q_3 + 1,5d_q \\ Li = q_1 - 1,5d_q \end{cases}$$

Em relação 2º e 3o CVs, as empresas que os atendem são aquelas que são *Outliers* estatísticos. Os valores que receberam e a quantidade de prefeituras contratantes estão acima dos respectivos limites superiores ( $Ls$ ). As empresas que atendem os três CVs serão nominadas, doravante, como empresas atípicas. Após a aplicação dos critérios de vulnerabilidade, a rede pode ser representada por um grafo.

### **Criação de uma rede de contratações públicas expressa por um grafo**

Os dados da Tabela 1 expressam apenas quantidade de vértices e arestas que permaneceram na rede após a aplicação dos três critérios de vulnerabilidade.

Tabela 1: Dados utilizados na pesquisa

UNIDADE DE INFORMAÇÃO	Quantidade	Representação no Grafo
Empresas: Empresas privadas contratadas para fornecerem bens e/ou serviços às prefeituras.	78	Vértice
Prefeituras: Órgãos públicos estatais que contratam algumas empresas para vender bens ou prestar serviços	224	Vértice
Pagamento: Soma de todos os pagamentos de uma prefeitura para uma empresa	1874	Aresta

Nota: Fonte: Elaborado pelo(a) autor(a).

### Aplicação de leiautes circular e geográfico

Antes da aplicação do leiaute, é necessário definir quais métricas de rede serão utilizadas. Essas métricas devem ser aplicadas nos tamanhos, cores e posição dos vértices, espessura e cores das arestas, dentre outras propriedades que permitam distinguir características relacionadas aos critérios de vulnerabilidade ao crime.

Tabela 2: Mapeamento de características e elementos-chave selecionados para evidenciá-las

Característica que se quer evidenciar	Métrica ou atributo	Elemento-chave utilizado para evidenciar característica			
		Posição	Cor	Tamanho	Forma
Quantidade de empresas atípicas contratadas por uma prefeitura	Outdegree			X	
População das prefeituras	Atributo do vértice		X		
Coordenada geográfica da cidade administrada pela prefeitura	Atributo do vértice	X			
Montante de recursos públicos recebidos (\$) por uma empresa	Weight InDegree	X		X	
Tipo de crime cometido pelos sócios da empresa	Atributo do vértice		X		X

Fonte: Elaborado pelo autor

Segundo Lusher, Koskinen, & Robins (2013, p. 120), uma rede bipartida consiste em dois conjuntos de vértices, com vínculos definidos apenas entre, mas não dentro dos dois conjuntos. Como na rede de contratações ora construída as prefeituras se conectam apenas às empresas e vice-versa, a rede é considerada bipartida.

O grafo gerado pela aplicação de leiautes está exposto pela Figura 6. Nele há dois leiautes independentes. O primeiro deles, representado pelo círculo externo, contém apenas as empresas atípicas<sup>4</sup> contratadas pelas prefeituras. Doravante serão realizados comentários acerca de relevantes insights visuais obtidos a partir dessa imagem.

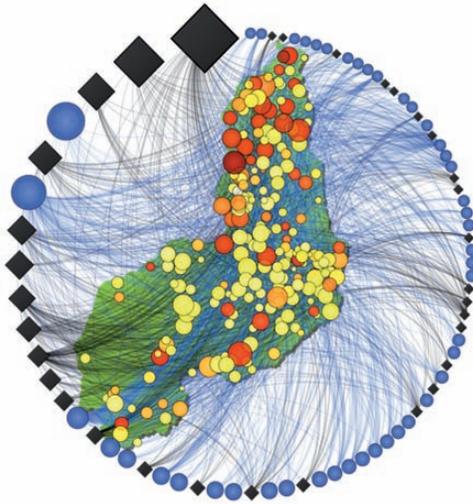


Figura 6. Composição de leiautes circular e geográfico – Contratações públicas das prefeituras do Piauí – Brasil  
Fonte: Elaborado pelo autor.

---

<sup>4</sup> Aquelas que atenderam aos três critérios de vulnerabilidade.

Nesse leiaute, o tamanho dos vértices é proporcional à métrica de centralidade *weight in-degree* que, empiricamente, representa o valor monetário total recebido pelas empresas de cada prefeitura a qual se conecta. Cada vértice de empresa está disposto na ordem ascendente dessa métrica, no sentido horário no círculo. Por sua vez, os elementos-chave cor e forma dos vértices no leiaute circular representam o tipo de crime que o sócio da empresa atípica cometeu: são pretos e em forma de losango se ao menos um dos seus sócios seja acusado de crimes em contratações públicas. São azuis e circulares caso os crimes sejam de outra categoria.

O segundo leiaute, o geográfico, contém as prefeituras. Elas estão representadas por vértices georreferenciados dentro do mapa, de acordo com as coordenadas geográficas das cidades que administram. O tamanho desses vértices são proporcionais à centralidade *Outdegree*. Empiricamente, essa métrica representa a quantidade de empresas atípicas que a prefeitura contratou. Ainda no leiaute geográfico, a cor dos vértices de prefeituras diz respeito à população da cidade e obedece a uma escala de calor que varia do amarelo, passando por laranja, vermelho claro e chegando ao vermelho escuro nas cidades mais populosas. Assim, prefeituras na cor amarela atendem a uma pequena população. No outro extremo, as prefeituras em vermelho escuro atendem às maiores populações do estado. Opcionalmente, o nome das cidades como *labels* dos vértices poderia ser inserido. Entretanto, como a pesquisa não visa analisar casos pontuais e concretos, a informação foi omitida.

**Análise qualitativa e exploratória: a contratação anômala de empresas atípicas**

Quanto maior a população de uma cidade, maior a quantidade de serviços públicos devidos aos cidadãos. Isso implica, presume-se, no incremento da quantidade de contratações de empresas pelas prefeituras que administram as cidades mais populosas. De forma oposta, há minoração presumida dessas quantidades nas cidades menos populosas.

As mesmas afirmações poderiam servir em relação a contratação de empresas atípicas: elas seriam, em tese e em regra, menos contratadas por cidades com menores populações e mais contratadas pelas mais populosas. Isso ocorre porque se a quantidade de empresas contratadas aumenta, amplificada será a probabilidade de elas integrarem o rol daquelas classificadas como atípicas. Uma análise exploratória visual e preliminar do grafo, entretanto, não parece confirmar essa hipótese. Há uma quantidade considerável de prefeituras de cidades pouco populosas – representadas, no mapa, por vértices em tons de amarelo – que contratam muitas empresas atípicas. Esse fato é evidenciado pela dimensão considerável dos seus vértices que, quando comparado aos das prefeituras das cidades mais populosas (em tons de vermelho), possuem tamanho similar.

Para exemplificar tal fato, tome-se como referência a prefeitura da cidade<sup>5</sup>  $\alpha$ , a mais populosa do estado (861.442 pessoas). Naturalmente, como  $\alpha$  é a maior contratante de empresas em geral, também estará entre os maiores con-

---

<sup>5</sup> Os nomes das cidades foram proposadamente omitidos, visto que a pesquisa não objetiva a indicação de casos concretos pontuais.

tratantes daquelas classificadas como atípicas: foram, ao todo, 23 (vinte e três) delas contratadas por  $\alpha$ . Esse fato é esperado, considerando a suposta correlação entre quantidade de contratações à população da cidade.

A anormalidade surge ao analisar a prefeitura da cidade  $\beta$ , também contratante de empresas atípicas. Apesar de administrar uma circunscrição com uma população equivalente a 0,62% de  $\alpha$ ,  $\beta$  contratou o equivalente a 86,95% da quantidade de empresas atípicas que  $\alpha$  contratou. E, ao menos visualmente, esse caso não é excepcional, visto que há várias outras prefeituras que a análise exploratória permitiu identificar com o mesmo comportamento anômalo de  $\beta$ . Essa hipótese será confirmada pela análise quantitativa *a posteriori*.

#### **Análise exploratória: a prevalência dos acusados por crimes de licitações dentre os mais bem pagos pelo poder público**

Na Figura 6, os vértices das empresas atípicas estão dispostos<sup>6</sup> numa ordem crescente da métrica Weight In-degree. Ela contém os montantes em recursos públicos recebidos por essas empresas. Considerando que os elementos-chave forma e cor aplicados a esses vértices estão de acordo com o tipo de crime de que seus sócios foram acusados, surge um padrão observável: há concentração das empresas dos acusados crimes de licitação nos 3º e 4º quadrantes do círculo, o que aponta que eles são ampla maioria em posições do grafo onde os ganhos financeiros são mais significativos.

---

<sup>6</sup> Uso do elemento-chave POSIÇÃO.

## Resultados

Os resultados da pesquisa apresentam as confirmações quantitativas dos insights obtidos na análise exploratória do estudo de caso, bem como uma sugestão de um fluxo de tarefas necessárias à análise exploratória em redes sociais.

### **Análise quantitativa: correlação população versus quantidade de empresas contratadas**

A análise exploratória presumiu que cidades com maior população tendem a contratar uma maior quantidade de empresas. A hipótese pôde ser confirmada pela pesquisa através do cálculo do coeficiente de correlação  $r$  entre essas duas variáveis: a população das cidades (variável independente  $x$ , atributo dos vértices) e a quantidade de empresas contratadas (variável dependente  $y = f(x)$ , obtida pelo `outDegree` dos vértices). O valor obtido do coeficiente de correlação foi  $r = 0,860$ . Por ser próximo de 1 (um), afirma-se que há correlação linear  $r$  positiva e forte entre ambas as variáveis. Há, portanto, uma tendência de que cidades com maiores populações contratem mais empresas, conforme demonstra o valor de  $r$ . No gráfico da Figura 7 a correlação positiva forte permite mostrar a tendência por meio de uma reta de regressão linear. A hipótese aventada na análise qualitativa, portanto, é válida.

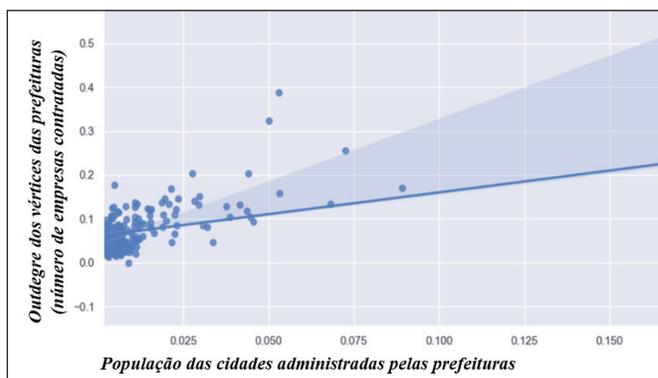


Figura 7. Correlação forte e positiva entre população das cidades e contratações de empresas pelas prefeituras. Fonte: Elaborado pelo autor

### Análise quantitativa: confirmação de anomalia na contratação de empresas atípicas

De modo diverso às variáveis correlacionadas anteriormente, estão as variáveis população versus contratação de empresas atípicas. O coeficiente de correlação entre ambas não indica tendências, pois cai significativamente para  $r = 0,620$ . Em consequência, o gráfico de dispersão não apresenta uma tendência linear aparente, conforme a Figura 8 apresenta:

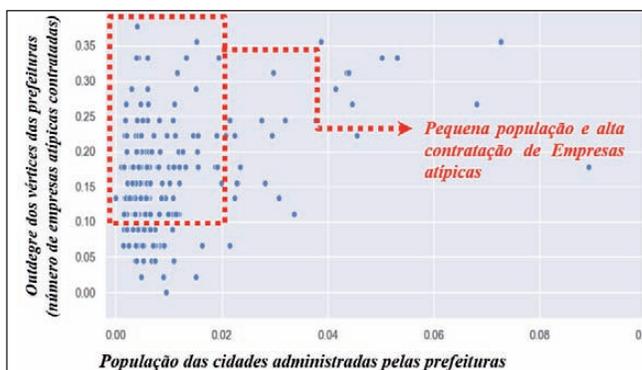


Figura 8. Ausência de correlação entre população das cidades e contratações de empresas atípicas pelas prefeituras 7 Fonte: Elaborado pelo autor

Apesar de ausência de correlação linear forte, o gráfico da Figura 8 apresenta 146 das 224 prefeituras dentro da região indicada pela linha tracejada em vermelho. Esse grupo de prefeituras contrata muitas empresas atípicas e administra cidades com pequenas populações. Assim, 65,17% das prefeituras apresentam esse comportamento anômalo.

Confirma-se, dessa forma, o insight da análise exploratória do grafo: ainda que contratem poucas empresas porque as demandas por bens e serviços são menores, grande parte das prefeituras que administram pequenas populações contratam muitas empresas atípicas, o que representa o aumento da vulnerabilidade da rede social de contratações estatais.

#### Análise quantitativa: preponderância de empresas atípicas nos quadrantes onde os ganhos são mais significativos

O gráfico da Figura 9, a seguir, ilustra a presença percentual de empresas atípicas de acordo com o tipo de crime que cometeram, em cada um dos quadrantes do leiaute circular do grafo:

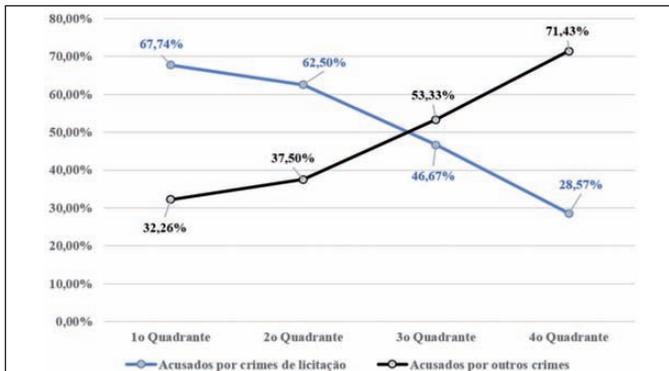


Figura 9. Presença de atípicos de acordo com o tipo de crime por quadrante do leiaute circular. Fonte: Elaborado pelo autor.

A presença das empresas atípicas cujos sócios são acusados de crimes de licitação é maior nos quadrantes superiores do leiaute circular, onde os ganhos também são maiores. Essa presença chega a 71,43% no último quadrante. Assim, a percepção obtida pela análise exploratória do grafo foi confirmada. Nesse cenário, os vértices pretos concentrados no 3º e 4º quadrantes compõem um rol perigoso: seus sócios são acusados por crimes de licitações e, concomitantemente, são os mais bem pagos num grafo de contratações estatais, justamente o tipo de rede que é alvo desse tipo de crime.

### Fluxo de tarefas sugeridas para uma análise exploratória de redes sociais

Com supedâneo nas etapas que foram seguidas, os passos sugeridos para execução de uma análise exploratória de redes sociais seriam: 1º) Preparação, 2º) Análise exploratória e 3º) Confirmação. Essas etapas estão iconograficamente descritas na Figura 10:

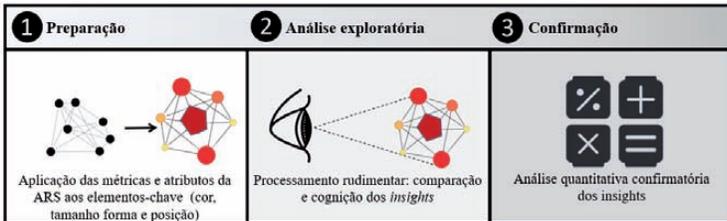


Figura 10. Etapas da análise exploratória em Redes Sociais baseada nos elementos-chave cor, tamanho, forma e posição. Fonte: Elaborado pelos autores

## Conclusão

As métricas e atributos dos vértices e arestas que compõem uma rede social podem ser aplicados aos elementos-chave cor, tamanho, forma e posição. Em seguida, estará viabilizada a análise exploratória aplicada a Redes Sociais, cujos insights visuais não são fruto de olhares aleatórios sobre uma representação pictográfica dos grafos que os representam. Eles são intuições dirigidas e propositadas a captar as –outrora mascaradas– características manifestamente visíveis e evidenciadas pelos elementos-chave. O processo de análise torna-se facilitado porque exige do observador tão-somente as comparações rudimentares (igual ou diferente, maior ou menor, esquerda ou direita, acima ou abaixo). Por sua simplicidade, o reconhecimento desses elementos-chave são habilidades que o homem adquire ainda em tenra idade e, regra geral, aplica na fase adulta com destreza. A depender da etapa confirmatória e quantitativa, as intuições obtidas podem ser relevantes. Assim, é possível afirmar que a migração da análise baseada em percepções mais sofisticadas para uma análise exploratória baseada nas comparações rudimentares dos elementos-chave desonera o observador. Isso porque esse processo o desobriga dos processamentos cognitivos mais apurados (leituras e cálculos) nas etapas iniciais da análise de redes sociais.

## Referências

- Agresti, A., Franklin, C., & Klingenberg, B. (2018). *Statistics: the art and science of learning from data*. Harlow: Pearson Education Limited.
- Bernabeu, M. del C. B., Castro, A. E., & Godino, J. D. (1991). "Análisis exploratorio de datos: sus posibilidades en la enseñanza secundaria". *Revista Suma*. Retrieved from <http://revistasuma.es/revistas/9-otono-1991/analisis-exploratorio-de-datos-sus.html>
- Caldarelli, G. (2013). *Scale-free networks: complex webs in nature and technology*. Oxford: Oxford University Press.
- Cherven, K. (2015). *Mastering Gephi network visualization: produce advanced network graphs in Gephi and gain valuable insights into your network datasets*. Birmingham: Packt publishing.
- Devore, J. L. (2006). *Probabilidade e estatística para engenharia e ciências*. São Paulo: Cengage Learning.
- Erciyes, K. (2014). *Complex Networks: An Algorithmic Perspective*. CRC Press.
- Heymann, S. (2015, February 25). *New Tutorial: Layouts in Gephi*. Retrieved March 10, 2020, from <https://gephi.wordpress.com/2011/06/13/new-tutorial-layouts-in-gephi/>
- Hu, Yifan (2005). Efficient, high-quality force-directed graph drawing. *The Mathematica Journal*, 37-71.
- Filho, M. J. (2014). *Curso de Direito Administrativo (10th ed.)*. São Paulo: Revista dos Tribunais.
- Gündüz-Ögüdücü S üle S ima, & Etaner-Uyar A. S ima. (2014). *Social networks: analysis and case studies*. New York: Springer
- Larson, R., Farber, B. (2016). *Estatística aplicada (6th ed.)*. São Paulo: Prentice Hall.
- Latora, V., Nicosia, V., & Russo, G. (2017). *Complex networks: principles, methods and applications*. Cambridge, United Kingdom: Cambridge

- University Press.
- Lusher, D., Koskinen, J., & Robins, G. (2013). Exponential Random graph models for social networks: theory, methods, and applications. New York: Cambridge University Press.
- Morettin, P. A., & Bussab, W. de O. (2010). Estatística básica. São Paulo: Saraiva.
- Ognyanova, K. (2019). Academia.edu. In Academia.edu. Retrieved from [https://www.academia.edu/30316883/Network\\_visualization\\_with\\_R?email\\_work\\_card=view-paper](https://www.academia.edu/30316883/Network_visualization_with_R?email_work_card=view-paper)
- Pinheiro João Ismael D. (2009). Estatística básica: a arte de trabalhar com dados. Rio de Janeiro: Campus.
- Pinheiro, J. I., Carvajal, S. R., da Cunha, S. B., & Gomes, G. C. (2012). Probabilidade e Estatística. São Paulo: Elsevier Editora Ltda.
- Robins, G. (2015). Doing social network research: network-based research design for social scientists. Los Angeles, CA: Sage.
- Wasserman, S., & Faust, K. L. M. (1994). Social network analysis: methods and applications. Cambridge: Cambridge University Press.