

Universidade Federal do ABC
Programa de Iniciação Científica - UFABC

Heurísticas computacionais para identificação e redução de potenciais conflitos de interesse em bancas de concursos

Relatório Final - Iniciação Científica - Modalidade voluntário

Edital N° 01/2017

Aluno

Brian Alves Andreossi

RA 11060215

brian.alves@aluno.ufabc.edu.br

Orientador

Jesús P. Mena-Chalco

jesus.mena@ufabc.edu.br

Santo André, 06 de Setembro de 2018

CARTA DE AVALIAÇÃO

À Ilustríssima Pró-Reitora de Pesquisa
Profa. Dra. Sônia Maria Malmonge

Encaminho o relatório final do aluno Brian Alves Andreossi referente ao projeto de pesquisa junto ao programa de Iniciação Científica, edital N° 01/2017.

É importante destacar que, nesse período de pesquisa, o aluno: (1) apresentou ótima autonomia acadêmica, e (2) teve um ótimo desempenho no desenvolvimento das atividades de estudo/aprendizado dos algoritmos considerados no Projeto de Iniciação Científica.

Atenciosamente,

Jesús P. Mena-Chalco (orientador)
Professor Adjunto - CMCC
SIAPE 1934625
jesus.mena@ufabc.edu.br

Resumo

Um problema recorrente na definição de bancas é a possibilidade de um avaliador conhecer um candidato e comprometer o julgamento de um processo seletivo. Por vezes, algumas instituições que realizam avaliações acabam sendo prejudicadas; muitas vezes sem mesmo saber que houve a chance de uma avaliação unilateral. Outras demoram meses, se usado de trabalho manual, conectando pessoas e definindo uma banca genuinamente justa. Este projeto de pesquisa aborda este problema considerando uma abordagem computacional. Almeja-se reduzir potenciais conflitos de interesse em processos que necessitam de seleção prévia de avaliadores. Dado um grafo relacionando objetos, o algoritmo seleciona avaliadores que não conheçam diretamente algum membro do grupo de candidatos, e então, combina-os e classifica-os de acordo com parâmetros interessantes para o usuário e, por fim, apresenta todas as bancas possíveis e idôneas, de forma a respeitar a imparcialidade do processo e a independência no que diz respeito as relações entre Avaliador e Candidato. Em uma análise inicial, avançou-se em estudos de como é definido um conflito de interesse, sua viabilidade, como identificá-lo e seguiu-se ao desenvolvimento da heurística. Percebeu-se as características do conflito de interesse assim que se tem um avaliador selecionado para banca e um candidato qualquer que se conheçam, e minuciou-se isto para a heurística. Com o uso de grafos, obtém-se as relações entre os elementos de um grupo feito da união dos subgrupos. Esta conexão é analisada, respeitando a neutralidade das relações, em detrimento do procedimento e gerando uma lista com a melhor seleção capaz de realizar a ação de forma imparcial.

Palavras-chave: conflito de interesse, imparcialidade, heurísticas computacionais.

Sumário

1	Introdução	5
1.1	Problematização	6
2	Objetivos e Metas	6
2.1	Objetivo Geral	6
2.2	Objetivos Específicos	7
3	Conceitos Básicos	7
3.1	Conflito de Interesse	7
3.1.1	Sobre o Termo ‘Conflito’	7
3.1.2	Definição de Conflito de Interesse	7
3.1.3	Exemplos de Conflito de Interesse	8
3.2	Análise Combinatória	9
3.2.1	Combinação	9
3.2.2	Tratamento Computacional	9
3.3	Grafos	10
3.3.1	Definições	10
3.3.2	Algoritmo de Floyd-Warshall	11
4	Método	12
4.1	Dados de Entrada	13
4.2	Cálculo da Matriz Distância	14
4.3	Obtenção dos Avaliadores Isolados	14
4.4	Obtenção dos Avaliadores Elegíveis	14
4.5	Cálculo do Número de Grupos	15
4.6	Geração dos Grupos	15
4.7	Classificação dos Grupos conforme os Parâmetros	16
5	Resultados e Discussões	17
5.1	Resultados Iniciais	17
5.1.1	Simulações em outros Casos	19
5.2	Resultados Finais	20
6	Cronograma	26
7	Trabalhos Correlatos	27
8	Considerações Finais	27

1 Introdução

Posterior à evolução da internet e dos meios de comunicação, percebe-se como as conexões sociais se tornaram maiores, e a troca de informações mais rápidas. De acordo com Werthein (2000), como a informação é parte integrante de toda atividade humana, essas atividades tendem a ser afetadas diretamente pela evolução de tecnologias informacionais, como é o caso da internet. Desta forma, o efeito é o crescimento de núcleos sociais, os quais interagem entre si e entre outros núcleos, além de evoluírem individualmente (Recuero, 2004).

Não se deve permitir que essas tais conexões interfiram nos processos que levam a escolha dos objetos capazes de fazer a diferença no local onde estão inseridos. Em outras palavras, essas conexões tem de ser deixadas a parte em prol do critério justo no momento do julgamento de certos processos. Ser candidato a uma profissão e conhecer um avaliador pode modificar o julgamento citado anteriormente.

Partindo disso, torna-se comum propostas como a de Adan Dunn, pesquisador do Instituto Australiano de Inovação em Saúde, que sugeriu a criação de um registro de nível mundial e público onde seria armazenado ligações de autores de pesquisas com partes potencialmente interessadas. A pesquisa citada por Dunn estima que 43% a 69% de artigos e relatórios sobre teste clínico contém falhas no que foi transmitido por causa de conflito de interesses (FAPESP, 2016).

No mês de Julho de 2016 foram abertos 130 concursos públicos no Brasil. Foram mais de 223 mil vagas¹ em todo o país. Para cada vaga aberta, existe um perfil de profissional ideal. Portanto, é necessário a avaliação de cada candidato para conferir se este profissional é o almejado pela instituição. Algumas vezes, essas avaliações são feitas por bancas, principalmente no âmbito de instituições de ensino superior, onde essas são constituídas por uma junta do corpo docente, escolhida previamente.

É preciso responder as perguntas que seguem. Como definir uma “boa banca” de avaliação de candidatos? Baseadas em informações de colaboração acadêmica (e.g., coautoria de publicações), pode rapidamente ser identificado um grupo de pessoas/instituições avaliadoras que seja totalmente independente ou isônomas das pessoas/instituições avaliadas?

Visando problemas dessa natureza -que é ainda mais pertinente à medida que se cresce a rede de influência por meio da evolução tecnológica- é que esta iniciação científica está inserida. A proposta é receber essa rede de influência e decidir conjuntos (baseados em critérios como distância, senioridade, grupo social) de quantos avaliadores sejam necessários e que estejam menos instigados a tornar real o que era somente um possível conflito de interesse. Fornecidos um grafo que conecte os personagens de

¹O número de vagas abertas por meio de concursos públicos no mês de Julho de 2016 está disponível em <http://g1.globo.com/economia/concursos-e-emprego/noticia/2016/07/mais-de-130-concursos-publicos-abertos-tem-223-mil-vagas-no-brasil.html>, último acesso em 27 de fevereiro de 2017.

um processo seletivo de qualquer natureza, e informações que ficam a critério do usuário, será possível identificar grupos que serão capazes de avaliar de forma mais imparcial os candidatos de um processo seletivo.

1.1 Problematização

Ao se realizar um concurso, abre-se inscrição para candidatos poderem participar e concorrer às vagas que são destinadas a estes participantes. A intenção é dentre diversas possibilidades, escolher o perfil mais apropriado para o preenchimento do loco, seja este acadêmico ou do mercado de trabalho. O segundo passo para o concurso é, então, a avaliação, baseado em critérios decididos pela instituição, dos candidatos e a classificação dos mesmos para que se possa escolher qual está mais apto à vaga ofertada. Para isso, é necessário que se encontre avaliadores que possuam a capacidade de caracterizar e observar, em cada um dos candidatos, separadamente, e principalmente de forma igualitária, para que seja preservado o interesse primário, que é a contratação do candidato melhor encaixado nos critérios de avaliação. Para que isto ocorra, é necessário que o grupo de candidatos e avaliadores não possuam uma relação direta, seja ela acadêmico-profissional ou social. Até então, as relações pesquisadas e o processo de combinação de avaliadores elegíveis é feito manualmente, que além de ser caro do ponto de vista econômico, é passível de erros e demorado.

Neste contexto, é que esta Iniciação Científica se viabiliza, dado que se mensurado as relações e processado pela heurística, têm-se de maneira rápida e barata os avaliadores elegíveis, e os mais diversos grupos que se pode formar com este conjunto, analisando também critérios interessantes e desejáveis nestes avaliadores, do ponto de vista da instituição que realiza o concurso.

2 Objetivos e Metas

2.1 Objetivo Geral

Demonstrar que a utilização de informações de relação socioprofissional, representadas em grafos, conjuntamente com heurísticas computacionais de seleção de grupos de avaliadores permite identificar e reduzir potenciais conflitos de interesse.

A classificação dos grupos de avaliadores, após sua identificação, será realizada considerando informações sobre diversidade do grupo, por exemplo, senioridade, área de atuação, localização geográfica, e informações topológicas.

2.2 Objetivos Específicos

- Desenvolver um método inclinado a precisar e qualificar elementos imparciais a partir da análise de grafos.
- Identificar todos os grupos de avaliadores com um número de elementos definido previamente.
- Classificar todos os grupos de avaliadores conforme informações topológicas e/ou socioculturais.

3 Conceitos Básicos

3.1 Conflito de Interesse

É necessário clarificar o que este projeto chama de Conflito de Interesse, uma vez que a heurística envolvida tenta evitá-los. E para isso, torna-se obrigatória a definição de conflito.

3.1.1 Sobre o Termo ‘Conflito’

Não há um consenso unânime na literatura sobre a concepção de conflito. A descrição defendida por [Thomas \(1992\)](#) entende que um conflito é um processo que se inicia assim que se percebe que alguma parte se frustrou, ou está para se frustrar, no que diz respeito a algum interesse. Se definido desta forma, é possível generalizar o conceito, o que é interessante pois dá uma base satisfatória para a conceituação de conflito de interesse que é defendida por [Thompson \(1993\)](#), como será mostrado na Seção [3.1.2](#).

3.1.2 Definição de Conflito de Interesse

Conflito de interesse é um cenário de condições no qual um julgamento profissional a cerca de um interesse primário tende a ser influenciado indevidamente por um interesse secundário ([Thompson, 1993](#)), mesmo que não haja conhecimento da influência por parte do indivíduo que a realiza ([Rothman, 1993](#)). Se um julgamento é dito influenciado de forma indevida, é certo que alguma parte se frustrou sobre a decisão tomada a favor daquele interesse primário, exatamente como caracterizado por [Thomas \(1992\)](#).

[Thompson \(1993\)](#) ainda evidencia aspectos sobre os interesses (primário e secundário). O primário é determinado pelo problema a ser julgado. Por mais que haja controvérsias sobre os interesses primários, geralmente há consenso sobre quais são e que devem ser uma consideração primária de qualquer decisão profissional. No caso estudado, seria o primário a definição de um perfil ideal de

candidato e a semelhança do perfil do candidato com o ideal, que, se julgado de maneira justa, leva o candidato à vaga. Quanto ao interesse secundário, este nem sempre é ilegítimo. Muitas vezes é parte necessária ou desejável da prática profissional, o que também é defendido por Rothman (1993); não é necessário que haja má fé onde há conflito de interesse, mas onde há possibilidade de tal influência, há razões para se preocupar que talvez esse potencial implique em um aumento da probabilidade de ato errôneo. O fato é que seu peso relativo na decisão profissional é problemático. Esse segundo interesse pode até ser justo, mas não deve dominar (ou mesmo parecer dominar) perante o interesse primário na construção da decisão ideal.

Alguns estudiosos procuram fornecer razões para que sejam evitados os conflitos de interesse, o que viabiliza esta iniciação científica. Segundo Thompson (1993), a regulação por meio de regras é a solução. A justificativa seria a conservação da integridade do julgamento profissional: “É seguro além de eticamente mais responsável decidir enquanto exclui os possíveis fatores que tendem a nos distrair dos objetivos médicos e acadêmicos”.

3.1.3 Exemplos de Conflito de Interesse

Existem muitos casos na área médica, como cita Thompson (1993). Cientistas priorizam trabalhos para indústrias de cosméticos e medicamentos que pagam para ter suas patentes estudadas.

Ainda na área médica, Goldim (2002) aponta para alguns ocorridos dentro de hospitais universitários: o interesse primário dos profissionais é atender devidamente os pacientes que são internados naquele hospital. Porém, os profissionais têm também o dever de ensinar os aspirantes, o que caracteriza o interesse secundário. Quando esse segundo interesse se torna prioritário, ou mesmo interfere na decisão final, o conflito de interesse se estabelece. De uma forma prática: um médico mantém um paciente internado por mais tempo para que um número maior de alunos possa estudar o caso. O trabalho de Goldim (2002) ainda alerta que há conflitos de interesse nas áreas de ensino, assistência e pesquisa.

Na área de pesquisa, Rego & Palácios (2008) mostram dados de cientistas estadunidenses que afirmam ter se comportado de forma inapropriada e que esse fenômeno pode ser descrito genericamente como uma manifestação dos conflitos de interesse no campo. São os comportamentos: falsificação de resultados, uso de ideias de outros sem o devido crédito, uso de informação confidencial sem autorização, falha em apresentar resultados contraditórios com pesquisas anteriores, mudança de desenho, método ou resultados de um estudo em resposta a pressões do patrocinador, uso de desenhos de pesquisa inadequados ou inapropriados.

3.2 Análise Combinatória

Existe um número de avaliadores elegíveis em uma rede de influência, e existe um número de vagas disponíveis na banca. Com a ferramenta de cálculo de conjuntos chamada Combinação e com o tratamento computacional sugerido por Knuth (2011), descritos nas seções seguintes, é possível contabilizar e montar cada um dos grupos de avaliadores possíveis em uma lista.

3.2.1 Combinação

Frequentemente é necessário saber quantos grupos diferentes de k objetos podem ser formados de um total de n objetos, tendo em vista que a posição de cada objeto no grupo é irrelevante (não contabiliza novo grupo).

De forma geral, $n(n-1)(n-2)\dots(n-k+1)$ representa o número total de grupos quando se tem n objetos escolhidos de k em k , quando importa a ordem em que os objetos estão no grupo. Isto significa que o número de grupos será contado $k!$ vezes a mais do que o grupo que interessa (que são os grupos diferentes onde não se repetem elementos), uma vez que tem fatorial de k repetições de um grupo com os mesmos elementos. Então o número total de grupos é:

$$\frac{n(n-1)(n-2)\dots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

Quando é necessário saber o número de grupos quando se tem n objetos, cada grupo com k , pode-se denotá-lo da seguinte forma:

$$\binom{n}{k}$$

Também podemos dizer que $\binom{n}{k}$ representa o número de combinações possíveis de n objetos em grupos de k elementos de cada vez (Ross, 2009).

3.2.2 Tratamento Computacional

Nesta seção será demonstrado a ideia inicial de como computar conjuntos de elementos de tamanho fixo a partir de um conjunto de cardinalidade maior. Knuth (2011) propõe listar os objetos de duas maneiras principais. A primeira é expondo um a um os elementos que estão no conjunto; e a segunda é mostrar cadeias com k binários enfileirados, representando cada binário, um objeto. Se o objeto estiver presente naquele grupo, o seu binário é 1; caso contrário é 0.

A forma escolhida para o problema desta Iniciação é a primeira, demonstrando os elementos um a um; e não as cadeias de binários. Knuth mostra somente o algoritmo para obtenção das cadeias de binários, uma vez que a posição deles são como índices, o que ajuda na obtenção da identificação do objeto tratado.

O Algoritmo usado é o L, que gera todas as combinações, se $n \geq k \geq 0$, como segue:

- (L1) [Inicializar.] Defina $c_j \leftarrow j - 1$ para $1 \leq j \leq k$; também defina $c_{k+1} \leftarrow n$ e $c_{k+2} \leftarrow 0$.
- (L2) [Visitar.] Visite a combinação $c_k \dots c_2 c_1$.
- (L3) [Achar j.] Defina $j \leftarrow 1$. Então, enquanto $c_j + 1 = c_{j+1}$, defina $c_j \leftarrow j - 1$ e $j \leftarrow j + 1$; eventualmente acontecerá da condição $c_j + 1 \neq c_{j+1}$ ser satisfeita.
- (L4) [Feito?] Conclua o algoritmo se $j > k$.
- (L5) [Incremente c_j .] Defina $c_j \leftarrow c_j + 1$ e retorne para (L2).

c_{j+1} e c_{j+2} são usados como sinalizadores.

3.3 Grafos

3.3.1 Definições

- (a) Grafo é definido por um conjunto (V, A) , onde V denota o conjunto de vértices e A denota o conjunto de arestas. Para esta iniciação científica, V é restringido a um conjunto finito de vértices. Eles representam redes de interações: quando dois elementos de um grupo (que são representados por vértices) tem alguma interação direta, então existe uma aresta os ligando (Feofiloff *et al.*, 2011). Um grafo é dito ponderado quando suas arestas tem pesos, e outra definição possível é: se dois vértices estão ligados por uma aresta em um grafo ponderado, não necessariamente a distância entre eles é igual a 1.
- (b) Vértices são elementos que podem ou não estar ligados por arestas. Eles representam os elementos do problema real. Um vértice é dito isolado quando nenhuma aresta incide nele (Feofiloff *et al.*, 2011). Porém, para o problema estudado, o termo isolado é mais comumente usado quando não existe um caminho entre um avaliador e um candidato. Assim, um avaliador é isolado se e somente se em sua componente conexa não há candidatos, somente avaliadores.
- (c) Arestas são ligações entre dois vértices. Uma aresta como (v, w) , $v, w \in V$, pode ser denotado por vw . Uma aresta vw incide em v e em w , de tal forma que v e w são as pontas da aresta. Quando uma aresta incide em dois vértices v e w , dizemos que eles são vizinhos ou adjacentes. E como restrição para esta pesquisa, excluiu-se grafos com arestas ditas paralelas (quando existem

duas arestas incidindo nos mesmos dois vértices) e quando as pontas da aresta coincidem ao mesmo vértice, a qual damos o nome de laço (Feofiloff *et al.* , 2011).

- (d) Componentes conexas é um subgrafo, de tal forma que existe um caminho mínimo entre cada um dos vértices deste subgrafo, 2 a 2 (Feofiloff *et al.* , 2011).
- (e) Subgrafo de um grafo é todo grafo tal que tanto as arestas quanto os vértices estejam contidos no grafo de origem (Feofiloff *et al.* , 2011).
- (f) Matriz-Adjacência (denotada M_{adj}) é uma das formas como um grafo é representado computacionalmente. A matriz têm, em suas linhas e colunas, os vértices. Quando existe uma aresta entre dois vértices v e w , então a M_{adj} na linha v , coluna w e na linha w , coluna v recebem um valor de controle que representa a presença de aresta. Para grafos ponderados, este valor é o peso da aresta, e se o grafo for não-ponderado, geralmente a presença é caracterizada por 1 e a ausência por 0. É uma matriz quadrada, de ordem igual ao número de vértices. Os elementos da diagonal principal recebem o valor de controle que representa a ausência de aresta (já que para essa Iniciação Científica, os laços foram desconsiderados), assim como todas as outras posições que não tem arestas os ligando.
- (g) Caminho Mínimo é o menor caminho entre dois vértices. Para grafos não ponderados, como os utilizados neste projeto de pesquisa, dois vértices que se ligam diretamente tem caminho mínimo igual a 1.

3.3.2 Algoritmo de Floyd-Warshall

Utiliza-se o algoritmo de Floyd-Warshall (Atallah, 1999) quando é necessário encontrar em um grafo a menor distância possível entre 2 vértices. Se o gráfico é ponderado, um vértice que se liga ao outro com peso p não necessariamente tem menor distância de valor p . A saída desse algoritmo é a Matriz Distância. O pseudocódigo do algoritmo está descrito em Algoritmo 1.

Algoritmo 1: ALGORITMO DE FLOYD-WARSHALL

Entrada: a matriz adjacência do grafo

Saída: a matriz com as distâncias mínimas

```
1 início
2   Para cada linha e cada coluna de  $M_{adj}[][]$  faça
3     if linha  $\neq$  coluna e  $M_{adj}[linha][coluna] == 0$  then
4        $M_d[linha][coluna] = \text{número de vértices} /* A maior distância é igual ao número de$ 
5         vértices*/
6     end
7     if  $M_{adj}[linha][coluna] \neq 0$  then
8        $M_d[linha][coluna] = 1 /* Se para 2 vértices, existe uma aresta, então, a menor$ 
9         distância entre eles é 1.*/
10    end
11  fim-Para
12  Para cada  $k^a$  entre 0 e número de vértices faça
13    Para cada linha da  $M_{adj}$  faça
14      Para cada coluna da  $M_{adj}$  faça
15        if  $M_d[linha][coluna] > M_d[linha][k] + M_d[k][coluna]$  then
16           $M_d[linha][coluna] = M_d[linha][k] + M_d[k][coluna]$ 
17           $M_d[coluna][linha] = M_d[linha][k] + M_d[k][coluna]$ 
18        end
19      fim-Para
20    fim-Para
21  fim-Para
22  retorna  $M_d[][]$ 
23 fim
```

^ak é um caminho que passa por outro vértice

4 Método

Inicialmente se tem um grafo e uma lista com os rótulos, classificando os vértices em Avaliadores e Candidatos. É necessário descobrir a distância entre cada par de vértices do grafo, para se ter precisamente o quão próximo eles são. Tendo as distâncias de cada par, deve-se analisar os rótulos e excluir do processo os avaliadores que tem ligação direta com os candidatos. A premissa é que se um avaliador é diretamente ligado a um candidato, então o sistema deve retirá-lo do processo de seleção, uma vez que seu julgamento tem maior chance de ser imparcial. A partir desse momento, usar o algoritmo de Knuth, descrito na Seção 3.2.2, para tratamento computacional da combinação e enumeração dos grupos. Posteriormente, classificar os grupos, criando um ranking com as opções. Para isto, o método deste projeto de pesquisa é executado em seis etapas, a saber:

1. Identificação do caminho mínimo entre vértices;
2. Obtenção de avaliadores isolados;
3. Obtenção de avaliadores elegíveis;
4. Cálculo do número de grupos;

5. Geração dos grupos de avaliadores;
6. Classificação dos grupos de avaliadores conforme os parâmetros de *ranking*.

A Figura 1 clarifica as seis etapas do processo descrito acima.

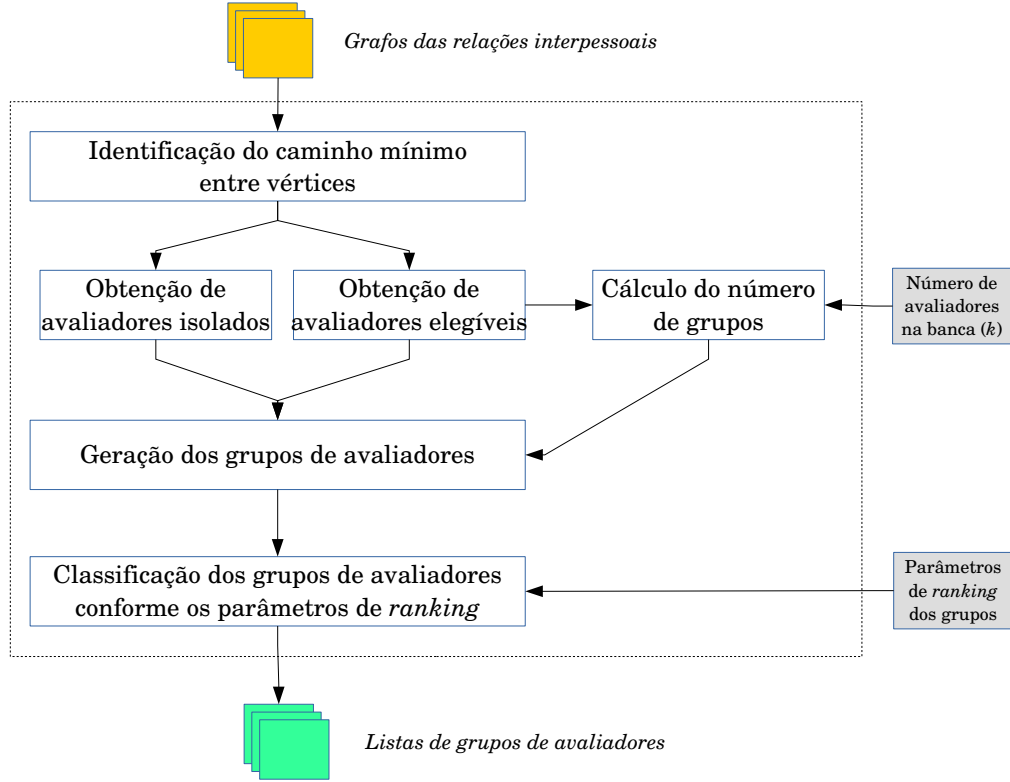


Figura 1: Fluxograma explicativo das ações realizadas nessa Iniciação Científica.

4.1 Dados de Entrada

Como parâmetros de entrada do método, tem-se dois conjuntos de dados: uma matriz-adjacência e rótulos classificatórios para os elementos (avaliadores e candidatos):

- A **Matriz-Adjacência** é o representante computacional do grafo que reúne todos os elementos do processo (que para esta Iniciação Científica são os avaliadores e candidatos). Se existe x trabalhos realizados em conjunto entre M e N , então tanto na linha M , coluna N , quanto na linha N , coluna M , deve estar na matriz o valor x . Este valor x pode relacionar o quão próximo os elementos M e N são, o que pode ser utilizado como parâmetro para criar um ranking de quais grupos tem avaliadores mais próximos ou menos próximos, dependendo do que é mais interessante como resultado final. A partir desta matriz é possível calcular a Matriz-Distância, usando o algoritmo de Floyd-Warshall. Esta matriz torna possível saber a proximidade entre os vértices, o que é interessante para o processo de obtenção dos Avaliadores Elegíveis e dos Avaliadores Isolados.

-
- Os **Rótulos**, que associam informação sobre o tipo de atuação no processo seletivo. Se o elemento M é um Avaliador, então na linha M estará escrito *Avaliador*. Senão, estará *Candidato*. Além disto, estarão outras informações, como Senioridade, Região ao qual pertence e etc.

Uma forma de utilizar o protótipo era por meio da disponibilização dos Conjuntos de Dados de Entrada pelo usuário, junto com os parâmetros de saída. Com o desenvolvimento e evolução da pesquisa, uniformizou-se a entrada com a saída do *ScriptLattes* (Mena-Chalco & Junior, 2009). É contribuição desta Iniciação um artefato que realiza mineração e coleta automática de dados, por meio da saída do *ScriptLattes*.

4.2 Cálculo da Matriz Distância

Para realizar este cálculo, foi utilizado o Algoritmo de Floyd-Warshall. Destaca-se como importância da Matriz Distância a facilidade de:

- (a) Determinar as Componentes Conexas do grafo, que por sua vez são importantes para encontrar os avaliadores elegíveis, e os inelegíveis; e entre os elegíveis, quem são os isolados;
- (b) Determinar parâmetros como soma das distâncias, média das distâncias, variância, desvio-padrão, etc.

4.3 Obtenção dos Avaliadores Isolados

A obtenção dos avaliadores isolados é importante para a classificação dos grupos, uma vez que pode ser utilizado como parâmetro. Em geral, um grupo que possua mais avaliadores isolados é mais imparcial. Ele é obtido analisando as Componentes Conexas: um avaliador é isolado se em sua componente conexa só houver avaliadores. Se existe ao menos um avaliador isolado na Componente Conexas, então todos os elementos dessa Componente também são avaliadores isolados.

4.4 Obtenção dos Avaliadores Elegíveis

Para a obtenção dos Avaliadores Elegíveis, representados por n , é analisado cada aresta de cada vértice: todos os avaliadores que tem arestas ligadas diretamente a candidatos devem ser excluídos do processo, se tornando inelegíveis; os que sobram são elegíveis.

4.5 Cálculo do Número de Grupos

Tendo obtido o número n de avaliadores elegíveis e recebido o valor k , o número de grupos possíveis é igual a $\binom{n}{k}$. O tempo de execução depende de quão grande $\binom{n}{k}$ é, pois quanto mais grupos, mais métricas são utilizadas para caracterizá-los e mais processamento é realizado. No escopo deste projeto, $\binom{n}{k}$ é tratado em tempo aceitável, como será descrito na seção 5.

4.6 Geração dos Grupos

Tendo calculado quantos grupos serão, se enfileirava uma carreira de bits de tamanho igual ao número de professores elegíveis e permutava k bits 1. Cada combinação representava um grupo de avaliadores possíveis.

Em uma primeira análise, foi proposto o Algoritmo 2 para gerar cada uma das $\binom{n}{k}$ combinações. A ideia propunha calcular o maior número que se podia escrever com n bits, e para cada valor em um intervalo de 1 a este número convertê-lo para binário, contar o número de dígitos ‘1’ e se fosse igual a k , então a posição de cada bit representaria o avaliador que estaria no grupo. Este algoritmo é lento, em especial entre a linha 10 e 17, uma vez que para um certo número n , sua saída demora algo da grandeza de $O(2^n)$, ou seja, exponencial em relação ao número de avaliadores elegíveis.

A observação e estudo do problema levou a uma solução mais eficiente, que não analisa ou transforma números, mas literalmente combina cada um dos elementos, como apresentado no algoritmo 3. Foi posicionado justamente entre a linha 10 e 17 do algoritmo, que é o laço de grandeza exponencial, na função Enumeração de Conjuntos. Este algoritmo é classificado como método, já que não tem saída explícita, porém seu objetivo é atualizar a variável S , que é a própria lista de grupos.

Algoritmo 2: ENUMERAÇÃO DE CONJUNTOS

Entrada: a matriz adjacência do grafo e a lista com os rótulos de cada índice, e o número de membros da banca

Saída: uma lista com todos os conjuntos possíveis do grupo elegível

```
1 início
2   numL = contarLinhas(arquivo matriz adj) // conta quantas linhas (elementos) o arquivo
   tem
3   M[ ][ ] = construirMatriz(arquivo matriz adj) // coleta os valores do arquivo e guarda em
   uma matriz M
4   D[ ][ ] = FloydWarshall(M) // cria uma matriz D com as distâncias, obtida com o
   algoritmo de Floyd-Warshall
5   A[ ] = Avaliadores(numL, arquivo com rótulos) // cria um vetor que informa quais índices
   são avaliadores
6   E[ ] = Elegíveis(D, P, numL) // cria um vetor que informa quais avaliadores são elegíveis
7   n = NumConj(E, número de membros da banca) // calcula o número de grupos possíveis
8   total = Total(número de elegíveis) // calcula qual o maior número decimal que se escreve
   com essa quantidade de binários

9   // trecho substituído pelo Método Combinações
10  Para cada número menor que total faça
11    aux = TransfBin(total) // transforma total da base decimal para a binária
12    i = Conta1(aux) // conta a quantidade de 1 presentes nos caracteres
13    if i == k then
14      aux = CompBin(aux) // Adiciona uma quantidade de 0 para que o //binário possua o
        mesmo número de avaliadores elegíveis
15      verifica a posição(ões) do(s) avaliador(es) e adiciona o grupo à Lista
16    end
17 fim-Para
18 fim
```

Algoritmo 3: COMBINAÇÕES

Entrada: Elegíveis[], k, n, v[], S[][]

```
1 ctrl = 0; //variável global que controla quantos grupos já foram escritos.
2 início
3   if k == 0 then
4     Para cada i de 0 a |v|-1 faça
5       S[ctrl][i] = v[i];
6     fim-Para
7     ctrl++;
8     retorne;
9   end
10  Para cada j de n até |Elegiveis|-k faça
11    v[|v|-k] = Elegiveis[j];
12    Combinações(Elegiveis, k-1, j+1, v, S);
13  fim-Para
14 fim
```

4.7 Classificação dos Grupos conforme os Parâmetros

A classificação dos grupos se dá a partir de parâmetros que devem ser informados ou extraídos das informações passadas, e acontece no momento em que se tem a lista contendo todos os grupos de avaliadores possíveis. A proposta inicial era de definir alguns que fossem relevantes para a diferenciação

de grupos para que se fizesse possível o ranqueamento.

A primeira análise deste problema levou a parâmetros gerados principalmente a partir de medidas topológicas do grafo, sempre desenhadas levando em conta cada elemento da banca em potencial. São eles: Distância Mínima e Máxima de um Avaliador para um Candidato; Soma, Média, Variância e Desvio Padrão de todas as Distância de um Avaliador a um Candidato e número de Avaliadores Isolados em relação a Candidatos. Tais parâmetros conseguiam distinguir grupos gerados a partir de amostras pequenas, mas se mostravam insuficientes para grupos maiores, com grafos pouco conexos.

Insistiu-se em diferenciar os elementos baseados em informações topológicas, com parâmetros como Média dos Graus dos Vértices, número de Avaliadores Isolados em relação aos outros Avaliadores do grupo, coeficiente de clusterização dos Avaliadores, mas o problema da baixa diferenciação dos grupos por conta do número de conexões não foi resolvido.

Então, considerou-se parâmetros baseados nas informações inerentes a cada candidato da banca, como Mínimo e Máximo das Senioridades (a qual é estimada pelo ano de conclusão do primeiro doutorado de um avaliador-pesquisador) e número de regiões diferentes por grupo que é definido como territórios que se distinguem por alguma apelidação ([Região, 2008](#)). Vale dizer que o tipo do território deve ser equivalente para todos os elementos, não só do grupo, mas do caso (se for ambientes de trabalho, por exemplo, todos as regiões serão parametrizadas desta maneira). Ao analisar informações do âmbito sociocultural, a pluralidade dos grupos se sobressaiu.

Apesar das informações socioculturais tenham sido mais determinantes, as medidas topológicas também têm sua importância, porque o isolamento entre Avaliadores e entre Avaliadores e Candidatos tem correlação direta com a presença ou não de arestas em um grafo. Portanto, a união dos melhores indicativos de cada uma das fases será capaz de classificar e, portanto, ranquear os grupos de pesquisadores que, no futuro, podem ser utilizados para a composição de uma banca de avaliação.

5 Resultados e Discussões

5.1 Resultados Iniciais

Foi possível, além de obter todos os conjuntos e fornecer algumas de suas características, criar certas ferramentas de análise de grafos (contabilizar número de subgrafos e enumerá-los, contabilizar e enumerar vértices isolados), que tiveram contribuição considerável na obtenção do resultado final, que são grupos de avaliadores mais independentes em relação aos candidatos em uma banca. Veja na [Figura 2](#) que as arestas ao centro (as que incidem em candidatos, representados por amarelo, e avaliadores, representados por azul) retiram esses vértices do grupo avaliadores elegíveis.

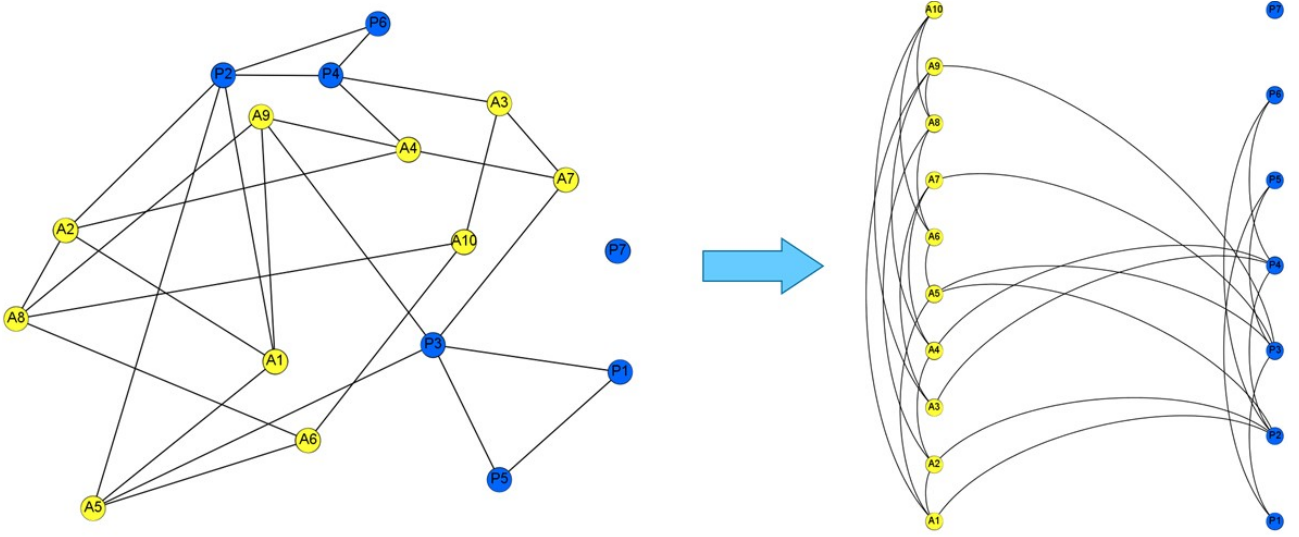


Figura 2: *Análise visual da percepção dos grupos pelo algoritmo no grafo sintético.*

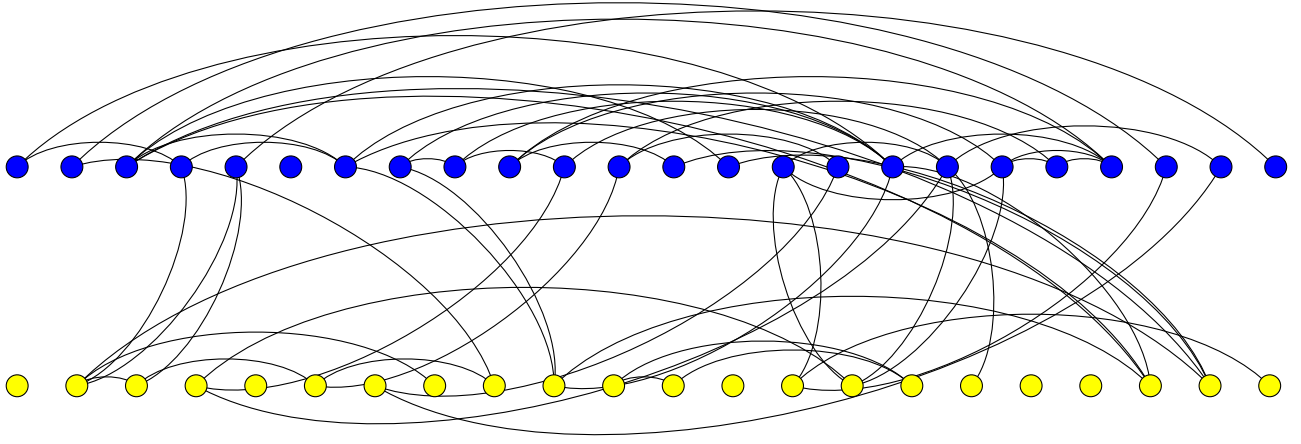
Como casos-teste, foram utilizados 3 grafos, como se pode ver na Tabela 1.

- (a) G1 corresponde a um grafo de coautoria entre 578 doutores associados aos Programas de Pós-graduação Economia do Brasil (professores permanentes). Este grafo foi considerado como estudo de caso inicial na versão preliminar da implementação do protótipo. Para efeito de estudo, definiram-se 100 avaliadores e 478 candidatos, de forma arbitrária (Haddad *et al.*, 2017).
- (b) G2 corresponde a um grafo de coautoria entre bolsistas de Produtividade em Pesquisa do CNPq (com vigência em 2016). Este grafo contém 24 doutores classificados como 1A e 22 classificados como 1B. No sistema de avaliação hierárquica, doutores 1A têm preferência aos doutores 1B e então a avaliação para bolsas deve ser feita por doutores superiores no sistema. Este exemplo retrata um caso real.
- (c) G3 corresponde a um grafo artificial com 7 avaliadores e 10 candidatos, criado para fins de teste.

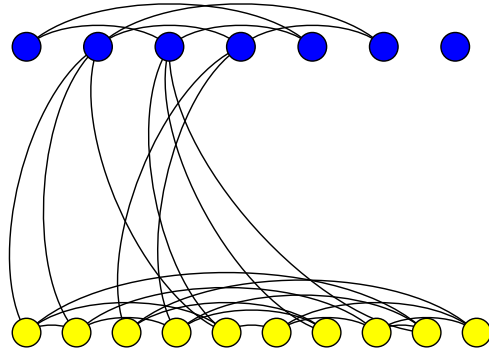
Tabela 1: *Principais informações dos casos de teste já realizados, com o número de avaliadores/banca fixo. (k).*

Grafo	Vértices	Avaliadores	Candidatos	Av. Elegíveis	Av. Isolados	k	CC	Grupos	Tempo
G1	578	100	478	27	22	4	125	17550	2,7 min
G2	46	24	22	7	1	4	8	35	~ 0 min
G3	17	7	10	4	1	4	2	1	~ 0 min

Para efeito de comparação da sua eficiência de tempo, foi coletado também dados variando k , somente do G1, para que suas características sejam evidenciadas. Lembra-se que, variando k , não se alteram: número de vértices, avaliadores, candidatos, avaliadores elegíveis, avaliadores isolados e Componentes Conexas. Estes dados estão presentes na Tabela 2.



(a)



(b)

Figura 3: Grafos de exemplo: (a) bolsistas por produtividade, (b) caso sintético. Amarelo representa os candidatos; Azul, os avaliadores.

Tabela 2: Informações das variações de grupos e tempo no G1 conforme variação do número de avaliadores/banca.

k	Grupos	Tempo
1	27	1 min e 17 seg
2	351	1 min e 17 seg
3	2925	1 min e 25 seg
4	17550	2 min e 43 seg
5	80730	29 min e 25 seg

5.1.1 Simulações em outros Casos

Com a atualização dos algoritmos para gerar conjuntos e a criação de parâmetros de classificação, o desenvolvimento da integração com o *ScriptLattes* (Mena-Chalco & Junior, 2009) e a disponibilização de quatro novos grafos de casos reais para testes, se coletou mais dados para validação funcional do protótipo. Para efeito de comparação, foi gerado um grafo artificial com base no grafo G1, descrito na Seção 5.1, simulando senioridade e região. Este grafo será nomeado como G1*.

As informações de cada um dos testes são as que seguem na Tabela 3. A coluna Melhores Grupos é definida pelo número de grupos se filtrados maximizando o número de avaliadores isolados em relação a candidatos no grupo e número de regiões diferentes; um parâmetro topológico e outro sociocultural.

Quanto menor for o conjunto dos melhores grupos em relação aos outros, mais sobressaltado fica a diferença entre eles.

Tabela 3: *Informações dos atributos dos grafos simulados no protótipo e classificação por filtro.*

Grafos	Candidatos	Avaliadores	Elegíveis	Isolados	k	Grupos	Melhores	%Melhores
C1	20	13	12	4	5	792	5	0,63%
C2	10	12	12	12	5	792	8	1,01%
C3	32	6	5	5	3	10	3	30,00%
C4	12	8	8	8	3	56	6	10,71%
G1*	478	100	27	22	5	80730	9460	11,72%

5.2 Resultados Finais

Durante o último semestre de trabalho, foram implementados aprimoramentos no protótipo responsável pelo cálculo dos grupos. Apesar do ganho ter sido muito considerável, não houveram quaisquer mudanças nos algoritmos, mas sim em como era realizado as operações de entrada e saída deste protótipo. O maior tempo gasto era em virtude de movimentar conteúdos de variáveis que se tornavam grandes durante o processamento e demandava maior tempo para serem atualizadas a cada vez que um novo grupo tinha suas métricas computadas. A solução se baseou em desprender a necessidade de uma variável de tamanho indefinido, fazendo mais operações de saída. Uma outra forma possível de solução era usar um objeto próprio para ser incrementado, porém, teria difícil tradução para linguagens que não fossem orientadas ou que não tivessem classes específicas para o incremento de variáveis do tipo cadeia de caracteres. Na Tabela 5 é possível ver o tempo que se demorou para executar todo o código para cada um dos k e algumas áreas.

Além disso, a implementação de novos critérios (como comentado anteriormente, a exemplo de senioridade e região proveniente do avaliador) e atribuição de métricas baseadas nestes proporcionou formas diversas de caracterizar e diferenciar os grupos. Nos dados reais, provenientes do concurso vigente de 2018, supondo bancas de 5 avaliadores, foi possível separar valores entre 0,03% e 38,71% de grupos com todos os avaliadores elegíveis e isolados para cada área², como se pode ver na Tabela 5. A escolha de k menor ou igual a 5 se justifica para este concurso, onde metade das áreas tiveram sucesso em formar bancas para todos os valores de k menores ou iguais a 5 e a outra metade não.

A Tabela esta disposta com 17 áreas (as que foram capazes de gerar pelo menos um grupo com 5 avaliadores); as relações R_k entre Número de grupos com k avaliadores isolados dividido pelo total de grupos com k avaliadores, em porcentagem, e os tempos t_k para gerar cada um dos $\binom{n}{k}$ grupos, posteriormente normalizados. Isto indica uma diminuição da amostragem de grupos para decisão

²Algumas áreas não conseguiram gerar bancas com 5 candidatos

do possível melhor. O melhor caso de Rk é o menor possível, mas diferente de zero, pois caso seja zero, não haveria um grupo totalmente isolado e o parâmetro para escolha devia ser outro, como por exemplo senioridade e região, por exemplo.

Algumas informações dessas áreas estão na Tabela 4. É possível ver um crescimento do número de grupos conforme cresce k ; a função $\binom{n}{k}$ cresce para k pequeno em relação a n , o que faz sentido para o problema, pois é esperado que haja um número bem maior de elegíveis do que de locus nas bancas (ainda que para certas áreas isto não seja possível).

Tabela 4: Principais informações das Áreas capazes de gerar grupos com pelo menos 5 avaliadores isolados.

Área	Rótulo	Avaliadores			Grupos com k avaliadores			
		Total	Elegíveis	Isolados	k = 2	k = 3	k = 4	k = 5
Antropologia	A1	68	29	8	406	3654	23751	118755
Arquitetura e Urb.	A2	37	17	14	136	680	2380	6188
Artes	A3	49	31	26	465	4495	31465	169911
Ciência Política	A4	53	28	14	378	3276	20475	98280
Comunicação	A5	51	14	6	91	364	1001	2002
Economia	A6	78	28	10	378	3276	20475	98280
Educação	A7	182	51	13	1275	20825	249900	2349060
Eng. Biomédica	A8	40	18	7	153	816	3060	8568
Eng. Elétrica	A9	147	25	5	300	2300	12650	53130
Filosofia	A10	77	49	35	1176	18424	211876	1906884
Geografia	A11	44	30	14	435	4060	27405	142506
História	A12	102	44	22	946	13244	135751	1086008
Letras	A13	105	69	35	2346	52394	864501	11238513
Linguística	A14	96	23	7	253	1771	8855	33649
Matemática	A15	141	40	10	780	9880	91390	658008
P. Urbano e Regional	A16	37	24	14	276	2024	10626	42504
Sociologia	A17	106	55	23	1485	26235	341055	3478761

Tabela 5: Relação entre Números de Grupos inteiramente isolados em porcentagem e tempo para executar todo o protótipo para aquele caso.

Área	R2	t2	R3	t3	R4	t4	R5	t5
A1	6,89	0,0004	1,53	0,0009	0,29	0,0023	0,04	0,0103
A2	66,91	0,0002	53,52	0,0002	42,05	0,0004	32,35	0,0008
A3	69,89	0,0002	57,84	0,0006	47,51	0,0025	38,71	0,0107
A4	24,07	0,0003	11,11	0,0007	4,88	0,0018	2,03	0,0074
A5	16,48	0,0001	5,49	0,0003	1,49	0,0003	0,29	0,0007
A6	11,9	0,0008	3,66	0,0010	1,02	0,0025	0,25	0,0098
A7	6,11	0,0015	1,37	0,0033	0,28	0,0255	0,05	0,2843
A8	13,72	0,0000	4,28	0,0002	1,14	0,0004	0,24	0,0008
A9	3,33	0,0007	0,43	0,0011	0,03	0,0023	0	0,0069
A10	50,59	0,0007	35,52	0,0024	24,71	0,0143	17,02	0,1370
A11	20,91	0,0003	8,96	0,0005	3,65	0,0023	1,4	0,0102
A12	24,41	0,0008	11,62	0,0021	5,38	0,0115	2,42	0,0999
A13	25,36	0,0008	12,49	0,0047	6,05	0,0658	2,88	1,0000
A14	8,3	0,0004	1,97	0,0006	0,39	0,0016	0,06	0,0048
A15	5,76	0,0012	1,21	0,0024	0,22	0,0101	0,03	0,0679
A16	32,97	0,0001	17,98	0,0003	9,42	0,0011	4,71	0,0035
A17	17,03	0,0008	6,75	0,0029	2,59	0,0254	0,96	0,2843

Tais informações também são úteis para evidenciar a dificuldade de gerar bancas sem relação com cada um dos avaliadores em algumas áreas, e a facilidade em outras. Uma hipótese é que seja mais fácil conforme a ciência se enquadre em *Soft* e mais difícil conforme se enquadre em *Hard Science*. Isto porque, em tese, ciências com maior rigor metodológico (*Hard Sciences*) possivelmente tenham menos avaliadores e portanto menos grupos possíveis. Existe ainda uma relação de difícil previsão com o grau de atividade que possivelmente exista entre os membros dos avaliadores. O que se prevê é que a medida que diminui a relação entre avaliadores, maior será os avaliadores elegíveis e portanto $\binom{n}{k}$ será maior.

Na Tabela 6, foram escolhidas as 10 áreas com maior número de avaliadores, e algumas informações de suas características de bancas. Para este concurso (relacionado com os bolsistas de produtividade vigente em 2018), é importante ressaltar que:

- Os Doutores de Educação e Matemática tem pouco contato com seus candidatos;
- É garantido que não há grupos de qualquer tamanho totalmente isolados em Química e Medicina Veterinária; e
- Existe muita contribuição científica em cooperação na área de Agronomia, a ponto de obter menos de 1% de avaliadores elegíveis para o concurso.

Tabela 6: Características importantes das 10 áreas com maior número de avaliadores em 2018.

Área	Avaliadores	Elegíveis	El./Tot. %	Isolados	Is./Tot. %	Reg. \neq
Física	449	26	5,79	4	0,89	23
Agronomia	354	3	0,85	1	0,28	28
Química	287	5	1,74	0	0,00	24
Medicina	249	5	2,01	1	0,40	18
Geociências	213	12	5,63	1	0,47	22
Educação	182	51	28,02	13	7,14	22
Eng. Elétrica	147	25	17,01	5	3,40	17
Matemática	141	40	28,37	10	7,09	18
Med. Veterinária	141	3	2,13	0	0,00	19
C. da Computação	139	7	5,04	3	2,16	21

A relação entre Elegíveis e Avaliadores do curso de Agronomia pode levantar hipóteses sobre a disseminação do curso no país. Todas as Unidades Federativas brasileiras tem representantes no conjunto de avaliadores. Ao mesmo tempo, dos 354 bolsistas, somente 3 estão distantes suficiente para avaliar. Isto pode representar uma migração dos avaliadores para o desenvolvimento de publicações em cooperação.

Outro dado importante é a relação de avaliadores elegíveis em relação ao número total de avaliadores presentes no concurso para cada área. Foram destacados na Tabela 7 as 10 áreas com a maior quantidade de avaliadores elegíveis no concurso de 2018 de bolsistas. É essencial perceber que das 10 áreas, 9 pertencem a *Soft Sciences*. Além disso, Matemática e Educação aparecem como áreas

onde há maior conectividade entre os membros do processo, ao mesmo tempo que Artes não há tanta comunicação entre diferentes níveis da Academia.

Tabela 7: *Características importantes das 10 áreas com maior número de avaliadores elegíveis em 2018.*

Área	Avaliadores	Elegíveis	El./Tot. %	Isolados	Is./Tot. %	Regiões \neq
Letras	105	69	65,71	35	33,33	21
Sociologia	106	55	51,89	23	21,70	19
Educação	182	51	28,02	13	7,14	22
Filosofia	77	49	63,64	35	45,45	17
História	102	44	43,14	22	21,57	21
Matemática	141	40	28,37	10	7,09	18
Artes	49	31	63,27	26	53,06	14
Geografia	44	30	68,18	14	31,82	15
Antropologia	68	29	42,65	8	11,76	19
Ciência Política	53	28	52,83	14	26,42	11

Uma outra medida importante retirada desta pesquisa é a dificuldade que certas áreas tem de gerarem bancas. De 85 áreas estudadas, 14 não tiveram ao menos 2 avaliadores elegíveis. Destas, 11 se encontram no grupo das *Hard Sciences*. Em complemento a estes dados, 52% não obtiveram resultados positivos nos testes, já que não montaram bancas de 5 avaliadores. Isto é evidência da dificuldade que se tem de formar bancas, principalmente na área das Exatas, e que este tema deve ser tratado com mais cuidado, já que a avaliação desses candidatos, quando positiva, representa em linhas gerais a renovação do conjunto de avaliadores em bancas futuras.

Para demonstrar a classificação (ou ranqueamento), foi utilizado PCA (Principal Components Analysis) no conjunto de avaliadores de Ciências de Computação com k valorado em 5, com intuito de mostrar graficamente as diferenciações possíveis de cada grupo, como mostra a Figura 4. Foram gerados 21 grupos. O melhor grupo depende dos parâmetros de entrada do usuário, como maximização/minimização da idade média, maximização/minimização de regiões diferentes ou maximização/minimização de avaliadores isolados no grupo.

Na Figura 5, é demonstrado alguns dos vetores que representam as variáveis usadas para classificar os grupos. A somatória vetorial das métricas de cada grupo posicionam o grupo no ponto em cada ponto que está representado na Figura 4.

Para o fim de entendimento geral, a Tabela 8 mostra algumas características das bancas $B3$, $B6$, $B15$, $B20$ e $B21$ todas estas descritas na Figura 4. Estas bancas foram geradas a partir de 7 avaliadores elegíveis, de um total de 139 avaliadores e 268 candidatos em 21 regiões diferentes. Haviam somente 3 isolados.

Crerios Topológicos são variáveis que são mais lógicas a sua maximização ou minimização. Há uma certa concordância que, ao aumentar o números de isolados em um banca, o seu juízo é mais imparcial. O mesmo ocorre com a média das Distâncias, que é a soma da distância de todo Avaliador

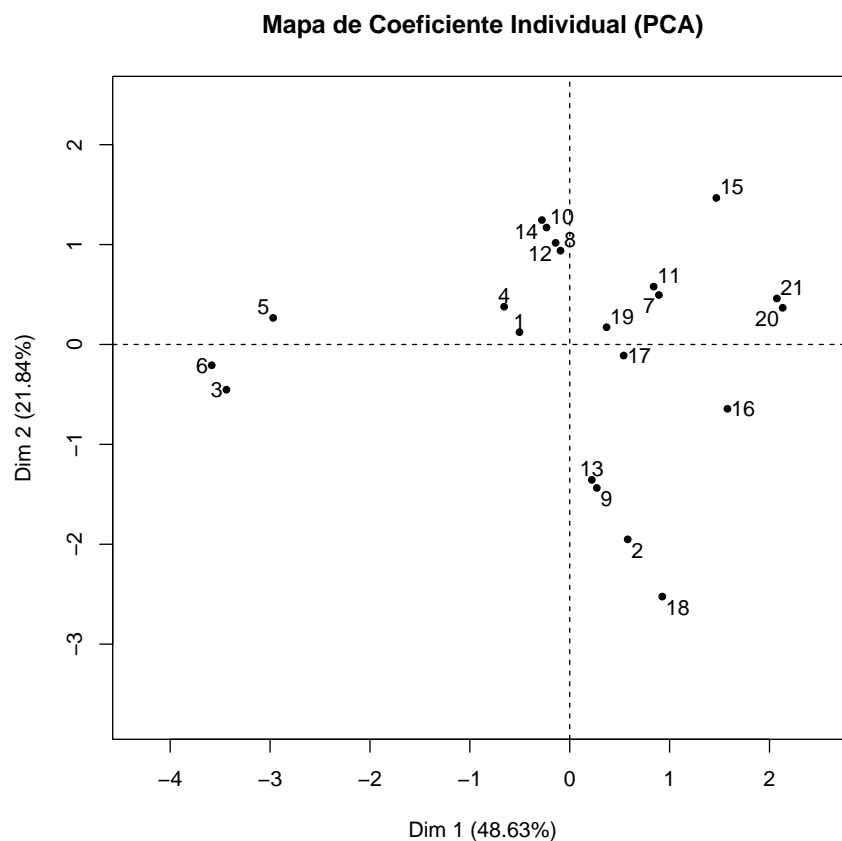


Figura 4: *Coeficiente Individual das bancas de 5 avaliadores possíveis para a área de Ciência de Computação.*

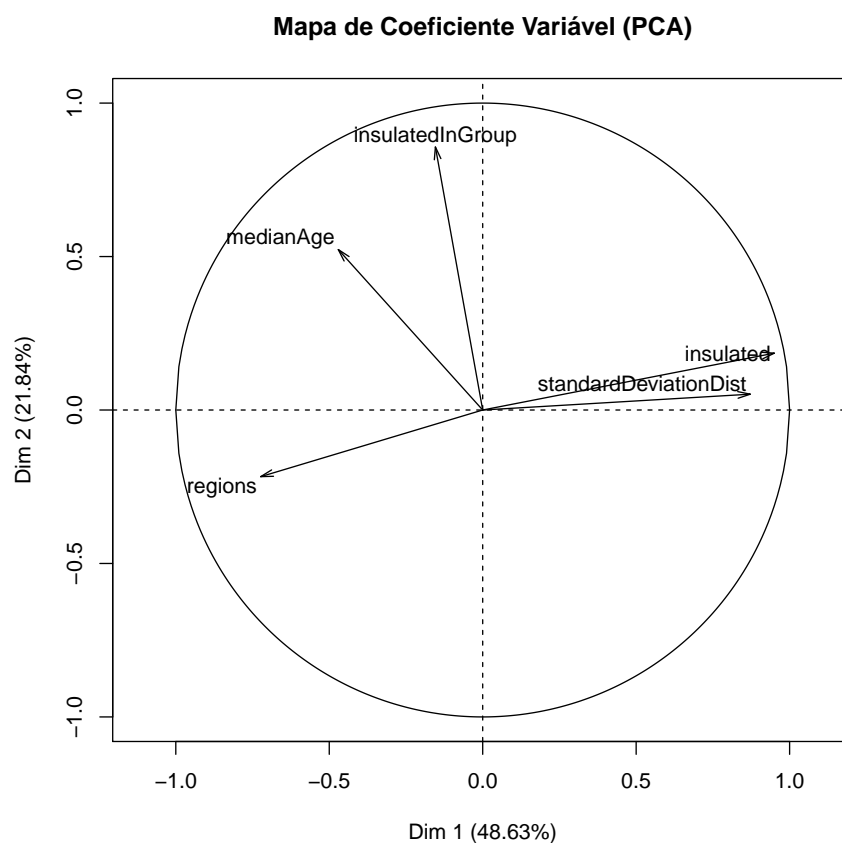


Figura 5: *Vetores de algumas Variáveis classificadoras dos grupos para a área de Ciência de Computação.*

Tabela 8: Características principais de bancas destoantes da Área de Ciências da Computação no concurso de 2018.

Rótulo	Medidas Topológicas			Medidas Sociológicas				
	<i>Distâncias</i>	Nº Isolados	<i>Grau</i>	< Idade	> Idade	<i>Idade</i>	Var. das Idades	Regiões
B3	27203.4	1	0.0	15	42	24.2	108.16	5
B6	27203.4	1	0.0	16	42	25.4	91.84	5
B15	68136.8	3	0.4	15	42	26.6	83.44	3
B20	68132.8	3	0.4	15	30	21.4	31.44	3
B21	68142.6	3	0.4	15	30	21.8	27.76	3

da banca para todo Candidato dividido por k , que é melhor conforme aumenta, pois quanto mais distante está os avaliadores dos candidatos, melhor. Porém, esta regra não é óbvia para medidas sociológicas. Alguns critérios pode-se querê-los maximizados, outros minimizados.

Portanto, o julgamento de que a banca $B3$ e $B6$ são menos aptas a avaliar é óbvio ao avaliar o número de isolados, visto que há grupos com esta característica mais ressaltada. O mesmo critério não é tão simples de ser visualizado nas bancas $B15$, $B20$ e $B21$, que possuem dados muito próximos. E é neste momento que se vê a importância das medidas sociológicas, que são capazes de diferir grupos que em tese são muito parecidos.

Para o caso das 3 bancas, há de se escolher qual critério deve-se maximizar ou minimizar. Se a média da idade deve ser mais alta, a $B15$ se torna a melhor escolha. Se é necessário avaliadores mais próximos no quesito idade, a $B20$ e $B21$ se apresentam como opções, se diferindo. Outros critérios deste perfil devem ser coletados para uma melhor diferenciação (e.g., gênero, área de estudo, número de publicações).

6 Cronograma

Este projeto de Iniciação Científica foi particionado em nove atividades que estão vigentes entre o período de Agosto de 2017 e Agosto de 2018, período o qual foi dividido em 3 quadrimestres (Tabela 9).

Por consequência direta do estudo de disciplinas relacionadas com o aprimoramento de algoritmos, o curso natural da pesquisa foi em direção ao aperfeiçoamento do algoritmo de Enumeração de Conjuntos, que muito se relaciona com as atividades de Geração de Conjuntos e depende da manutenção e testes do protótipo. Portanto, a prioridade esteve no estudo de classificação dos avaliadores e no aprimoramento do cálculo do número de grupos.

Dando continuidade ao projeto, considerou-se uma última fase de Aprimoramento da etapa de geração dos grupos de avaliadores, no sentido de classificar e ranquear cada banca gerada pela heurística seguindo uma resultante dos parâmetros. Esta etapa foi resolvida a partir de maximização de parâmetros. Tendo realizada a etapa, se prendeu a atenção a aprimoramentos de análise assintótica dos algoritmos. Posteriormente, fez-se manutenção do protótipo baseada no aprimoramento supracitado e testes do mesmo.

Tabela 9: Cronograma de atividades de trabalho. Os meses estão representados com números de 2 algarismos. A linha marcada com *P* representa o que foi projetado e a linha marcada com *R* o tempo que demorou para ser realizado.

Atividade		Quadrimestre 1				Quadrimestre 2				Quadrimestre 3			
		01	02	03	04	05	06	07	08	09	10	11	12
Aprofundamento no aprendizado de técnicas de análise de grafos	P												
	R												
Aprimoramento da etapa de identificação dos caminhos mínimos	P												
	R												
Estudo e aprimoramento da etapa de obtenção dos av. isolados e elegíveis	P												
	R												
Estudo e aprimoramento da etapa de cálculo do número de grupos	P												
	R												
Aprofundamento no aprendizado de algoritmos de enumeração	P												
	R												
Aprimoramento da etapa de geração dos grupos de avaliadores	P												
	R												
Análise, manutenção e teste do protótipo já desenvolvido	P												
	R												
Escrita de artigo científico para evento nacional	P												
	R												
Escrita de relatório	P												
	R												

7 Trabalhos Correlatos

Existem muitos trabalhos descritos na literatura que essencialmente discutem e apresentam definições ou estudos de casos sobre conflito de interesse, mas raramente descrevem uma solução para o problema, e quando descrevem, abordam soluções de formas manuais (Viswanathan *et al.*, 2013), usando questionários. Além disso, os estudos sobre conflitos de interesses estão focados principalmente na área médica, como apresentado por Friedberg *et al.* (1999), Strom (2004) e outros; ou na área financeira, visto por Puri (1996). Conflitos de interesses na área acadêmica estão pouco explorados.

Não foram identificadas soluções computacionais similares à proposta nesta iniciação o que, segundo o nosso conhecimento, evidencia a grande importância por métodos que resolvam o problema para a área acadêmica.

8 Considerações Finais

O propósito desta Iniciação Científica se construiu em cima de um método capaz de diminuir potenciais conflitos de interesses em bancas de concursos, entendendo primeiramente do que se trata um conflito de interesse, como ele se caracteriza, e de que forma é possível evitá-lo usando heurísticas computacionais baseadas em grafos.

O estudo dos conceitos levou ao entendimento do porquê conflitos de interesse precisam ser evitados, e para isto, se apresentou um conjunto de algoritmos base que permite nomear cada um dos grupos de avaliadores possíveis, dado um grafo apresentando as relações interpessoais entre os candidatos e avaliadores. Os próximos passos consistiram em melhorias nos métodos e no ranqueamento a partir dos parâmetros já recebidos e coletados.

Houve também consenso e desenvolvimento no sentido de integralizar a saída do *ScriptLattes* (Mena-Chalco & Junior, 2009) com a entrada desta heurística, em uma tentativa sucedida de transformar este projeto em um módulo auxiliar de automatização de geração de bancas, com o objetivo de instaurar um sistema *online* de consulta de vida acadêmica de pesquisadores e análise de conflito de interesses em casos reais. Esta integração se dá por meio do reconhecimento de padrões das informações na saída do *ScriptLattes*, usando Expressões Regulares para identificar e coletar os dados para construção dos arquivos de entrada deste projeto. Tal automatização é artefato funcional deste projeto.

Este projeto pode também ser aplicado em licitações. Proprietários de empresas que fossem ligados entre si poderiam combinar propostas, o que causa transtorno (é uma prática ilegal). Com o uso deste método, tal empresa seria, se não excluída do processo, ao mínimo monitorada para evitar tal

situação. Também seria possível adaptar para o uso policial. Uma análise de possíveis nomes em uma investigação, onde a lista gerada seria uma espécie de rede de interações, o que talvez possa identificar autores e coautores de crimes.

Interessa à comunidade científica responder se é possível encontrar grafos mais completos, com mais interligações (olhando, se possível, relacionamentos a parte do mundo acadêmico que, vez ou outra, cruza as fronteiras e interfere na complementação da comunidade científica) e avaliar a influência de terceiros no processo de avaliação, para se entender se realmente existem tantos avaliadores elegíveis a avaliação de candidatos quanto esta pesquisa mostrou até aqui.

A contribuição desta Iniciação Científica pode, além de automatizar um processo de escolha de banca, fornecer medidas quantitativas e qualitativas do relacionamento acadêmico entre membros que supostamente estão em níveis diferentes do conhecimento em certa área. Essas medidas podem corroborar teorias que ajudem a trilhar o caminho que a comunidade científica deve tomar para que ela se torne mais independente de falhas individuais e que cresça a parte de conflitos que podem manchar sua maior virtude, a razão.

Algumas das questões aqui tratadas são: quão difícil é encontrar bancas independentes? Quão próximo estão os avaliadores dos candidatos? Quão próximos estão entre si? Uma área se conecta com qual frequência? Há diferenças significativas nestas respostas em ciências classificadas como *soft* e *hard*? Estas respostas dependem de mais conjuntos de dados e uso da heurística e novas fases de estudos em cima do tema.

Referências

- Atallah, M. J. 1999. Basic Graph Algorithms. *Chap. 6, pages 136–158 of: Algorithms and theory of computation handbook*, 1 edn. Chapman & Hall/CRC Applied Algorithms and Data Structures series. CRC Press. 11
- FAPESP. 2016. Conflitos de interesse mais transparentes. *Revista Pesquisa FAPESP*, **244**, 9. <http://revistapesquisa.fapesp.br/2016/06/15/conflitos-de-interesse-mais-transparentes>. 5
- Feofiloff, P., Kohayakawa, Y., & Wakabayashi, Y. 2011. *Uma introdução sucinta à teoria dos grafos*. Sociedade Brasileira de Matemática. 10, 11
- Friedberg, M., Saffran, B., Stinson, T. J., Nelson, W., & L.Bennett, C. 1999. Evaluation of conflict of interest in economic analyses of new drugs used in oncology. *The Journal of the American Medical Association*, **282**(15), 1453–1457. 27
- Goldim, J. R. 2002 (Janeiro). *Conflito de interesses na área da saúde*. <https://www.ufrgs.br/bioetica/confliit.htm>. Último acesso em 05/09/2016. 8
- Haddad, E., Mena-Chalco, J.P., & Sidone, O. 2017. Produção Científica e Redes de Colaboração dos Docentes Vinculados aos Programas de Pós-graduação em Economia no Brasil. *Estudos Econômicos*, 1–61. (aceito para publicação). 18
- Knuth, D. E. 2011. *The Art of Computer Programming, Volume 4A: Combinatorial Algorithms, Part 1*. Addison-Wesley Professional. 9
- Mena-Chalco, J.P., & Junior, R. 2009. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, **15**(4), 31–39. 14, 19, 27
- Puri, M. 1996. Commercial banks in investment banking conflict of interest or certification role? *Journal of Financial Economics*, **40**(3), 373–401. 27
- Recuero, R. C. 2004. Teoria das redes e redes sociais na internet: considerações sobre o Orkut, os weblogs e os fotologs. In: *XXVII Congresso Brasileiro de Ciências da Comunicaç ao. XXVII INTERCOM*. 5
- Região. 2008. *Dicionário Aurélio*. Editora Positivo. 17
- Rego, S., & Palácios, M. 2008. Conflitos de interesses e a produção científica. *Revista Brasileira de Educação médica*, **32**(3), 281–282. 8
- Ross, S. 2009. *Probabilidade: um curso moderno com aplicações*. Bookman Editora. 9
- Rothman, K. J. 1993. Conflict of interest: the new McCarthyism in science. *The Journal of the American Medical Association*, **269**(21), 2782–2784. 7, 8
- Strom, B. L. 2004. Potential for conflict of interest in the evaluation of suspected adverse drug reactions: a counterpoint. *The Journal of the American Medical Association*, **292**(21), 2643–2646. 27
- Thomas, K. W. 1992. Conflict and conflict management: Reflections and update. *Journal of organizational behavior*, **13**(3), 265–274. 7
- Thompson, D. F. 1993. Understanding financial conflicts of interest. *New England Journal of Medicine*, **329**, 573–576. 7, 8
- Viswanathan, M., Carey, T. S., Belinson, S. E., Berliner, E., Chang, S., Graham, E., Guise, J., Ip, S. S., Maglione, M. A., McCrory, D., & et al. 2013. Identifying and managing nonfinancial conflicts of interest for systematic reviews. *Agency for Healthcare Research and Quality (US)*, 3–6. 27
- Werthein, J. 2000. A sociedade da informação e seus desafios. *Ciência da informação, Brasília*, **29**, 71–77. 5