

MEDIDAS DE CENTRALIDADE E DETECÇÃO DE COMUNIDADES EM REDE DE CO-AUTORIA

Diego Andrés de Barros Lima Barbosa
Leonardo Borges Avelino
Rarylson Freitas de Souza
Camila Cristina Gomes Ferreira de Oliveira
Claudia Justel

Instituto Militar de Engenharia-IME
Praça General Tibúrcio, 80 – Praia Vermelha - RJ
diego.andres.barbosa@gmail.com , lborgav@gmail.com, rarylson@gmail.com,
cjustel@ime.eb.br , milinhacristina@hotmail.com .

RESUMO

Neste trabalho é apresentada uma análise de uma rede de co-autoria construída a partir de dados reais da *Plataforma Lattes* do CNPq. A análise consiste em determinação dos vértices centrais, considerando as medidas de centralidade mais usadas: grau, proximidade, intermediação, autovetor e *Pagerank*. Um estudo sobre comunidades nesta rede também é apresentado. Os resultados obtidos com os algoritmos clássicos de Girvan e Newman, e Clauset, Newman e Moore para detecção de comunidades na rede proposta são apresentados.

PALAVRAS CHAVE. Teoria espectral de grafos, algoritmos, particionamento de grafos, redes sociais.

Área de classificação principal (Teoria e Algoritmos em Grafos).

ABSTRACT

This work presents an analysis of a co-authorship network constructed from real information of the Lattes Platform hosted in CNPq homepage. It consists in identifying the central vertices of the network by using the usual centrality measures: degree, closeness, betweenness, eigenvector and pagerank. Moreover, a study about communities is introduced. The results obtained with the classical algorithms of Girvan and Newman and Clauset, Newman and Moore are presented.

KEYWORDS. Spectral graph theory, algorithms, graph partitioning, social networks.

Main area (Theory and Algorithms in Graphs).

1. Introdução

As redes complexas, e em particular redes de co-autoria, tem sido exaustivamente estudadas nos últimos anos ([11], [15]). Diversos trabalhos foram publicados sobre redes de co-autoria obtidas a partir de dados reais. Uma análise do grafo de colaboração de Erdős é apresentada em [3]. O trabalho de Newman de 2001 ([13]) estuda grafos de colaboração obtidos de áreas de conhecimento diferentes. Em [14], o autor propõe uma rede construída a partir da bibliografia de

dois *surveys* sobre redes complexas. *Microsoft Academic Research* ([12]) oferece um serviço de busca na área acadêmica que permite obter os co-autores de um pesquisador determinado, a partir de uma base de dados de periódicos e anais de congressos, mantidos desde 2009, mostrando uma visualização dos resultados. Já a *Plataforma Lattes* do CNPq ([7]) disponibiliza, para cada pesquisador cadastrado, uma rede de colaboração construída a partir de algumas informações declaradas no currículo do pesquisador. Neste artigo propomos construir uma rede de co-autoria contendo dados reais a partir das informações públicas sobre artigos publicados em periódicos de um subconjunto de pesquisadores da *Plataforma Lattes*. Faremos uma análise dessa rede utilizando medidas de centralidade e finalmente faremos um estudo sobre comunidades na mesma rede.

Este trabalho está organizado da seguinte forma. A Seção 2 contém as definições e conceitos básicos sobre grafos e redes necessários para a compreensão do texto. A Seção 3 apresenta as medidas de centralidade definidas para redes. Na Seção 4 é introduzida a definição de comunidade e os algoritmos para detecção de comunidades abordados neste trabalho são descritos brevemente. A construção do grafo de co-autoria proposto e os resultados obtidos para as medidas de centralidade e detecção de comunidades do grafo proposto são tratados na Seção 5. Finalmente, a Seção 6 traz as conclusões, e a Seção 7 a bibliografia utilizada.

2. Conceitos básicos e definições

2.1 Conceitos básicos em grafos

Um **grafo não orientado** $G=(V,E)$ é formado por dois conjuntos V e E , onde V é um conjunto finito e não vazio, $|V|=n$, cujos elementos são denominados **vértices** e E , $|E|=m$, é um conjunto de subconjuntos de dois elementos de V cujos elementos são denominados **arestas**. Cada aresta é denotada por $e=(v_i,v_j)$, onde $v_i, v_j \in V$. O **grau** de um vértice v_i , denotado $grau(v_i)$, é o número de arestas incidentes a ele e dois vértices são denominados **adjacentes** se existe uma aresta entre eles.

Um **caminho** é uma sequência de vértices distintos, v_1, \dots, v_k , tal que $(v_i, v_{i+1}) \in E$ para $1 \leq i \leq$

$k-1$, $k > 1$. O **comprimento do caminho** v_1, \dots, v_k é igual a $k-1$. Um grafo é **conexo** quando existe um caminho entre qualquer par de seus vértices. Caso contrário, ele é denominado **desconexo**.

A **distância entre dois vértices** v e w do grafo G , denotada $d_G(v,w)$, é o comprimento do menor caminho entre v e w . A **excentricidade de um vértice** v , é definida por $e(v) = \max_{\{w \in V\}} d_G(v,w)$. O valor máximo das distâncias entre todos os pares de vértices do grafo é o **diâmetro do grafo**, $diam(G) = \max_{\{v,w \in V\}} d_G(v,w)$. O **centro do grafo** G , é o subconjunto de V formado pelos vértices de excentricidade mínima.

Um grafo não direcionado G com n vértices pode ser representado através de uma matriz $A(G) = (a_{i,j})$ de dimensão $n \times n$, denominada **matriz de adjacência** de G , onde a entrada $a_{i,j}$ da matriz é igual a 1 se v_i e v_j são adjacentes ou igual a 0 caso contrário. Um número complexo λ é **autovalor** da matriz M de dimensão $n \times n$ quando existe um vetor $x \in C^n$ não nulo tal que: $Mx = \lambda x$. Quando a matriz M é simétrica, todos os autovalores são reais. O autovalor com maior valor absoluto de uma matriz simétrica é denominado **raio espectral** (no caso da matriz de adjacência de um grafo, o raio espectral coincide com o maior autovalor da matriz, que é um número positivo). Uma matriz com entradas positivas e não redutível admite um autovetor associado ao raio espectral com todos os componentes positivos, denominado **vetor de Perron**. A matriz de adjacência de um grafo não direcionado é não redutível e portanto o vetor de Perron, associado ao maior autovalor dessa matriz, possui todos os elementos maiores que zero.

Outros conceitos sobre grafos, algoritmos e complexidade podem ser obtidos em [20]. Definições e conceitos básicos em teoria espectral de grafos são apresentados em [1].

2.1 Conceitos básicos em redes

Segundo Kleinberg e Easley ([11]), podemos considerar uma **rede** como uma coleção de objetos na qual, alguns pares deles estão relacionados por conexões. Essa definição é muito flexível e depende da forma na qual as relações são utilizadas para definir as ligações. Devido a essa

flexibilidade, existem exemplos de redes em diversos domínios, podemos citar dentre eles as redes de comunicação, redes sociais, redes biológicas e redes de informação.

Nos últimos anos foram desenvolvidas várias linhas de pesquisa na área de redes. Por exemplo:

- 1) análise da estrutura da mesma por meio de teoria de grafos.
- 2) estudo do comportamento interdependente de grupos de indivíduos da rede, por meio da teoria de jogos.
- 3) relacionamento entre a dinâmica da rede e suas consequências na estrutura da mesma.

A teoria de grafos, como modelo matemático da estrutura da rede, permite analisar aspectos topológicos da mesma. Neste trabalho abordamos dois problemas importantes na análise da estrutura de uma rede: a determinação dos elementos mais importantes da mesma utilizando medidas de centralidade e a determinação de grupos de elementos fortemente relacionados entre si dentro da rede por meio da detecção de comunidades. Essa análise será aplicada numa rede de co-autoria, que é um tipo particular de rede social.

Outros conceitos sobre as redes abordadas neste trabalho podem ser consultados nos livros de Kleinberg e Easley ([11]) e Newman ([15]).

3. Medidas de centralidade

Nesta seção são apresentadas as definições das medidas de centralidade de vértice mais usadas na análise de redes. Existem algumas variantes nas definições adotadas para essas medidas. Por exemplo em alguns casos, a medida pode ser ou não normalizada (caso da centralidade de grau, autovetor e proximidade). Neste trabalho foram escolhidas as definições achadas na literatura que melhor atendem às idéias originais, sempre respeitando o princípio que medidas com valor maior determinam vértices mais centrais. Seja $G=(V,E)$ um grafo onde $|V|=n$ e $|E|=m$. Seja $v \in V$ um vértice do grafo.

A **centralidade de grau** é a medida mais simples. Esta medida está baseada na seguinte idéia: vértices que possuem grau maior podem estar em uma posição mais privilegiada da rede. A centralidade de grau de um vértice v é definida como $C_D(v) = grau(v)/(n-1)$.

A **centralidade de proximidade** (*closeness*) é definida a partir da média das distâncias geodésica do vértice v e todos os outros vértices alcançáveis por ele, $C_C(v) = (n-1) / \sum_{t \neq v \in V} d_G(v,t)$. Vértices com maior centralidade de proximidade assim definida ocupam uma posição próxima do centro do grafo.

A **centralidade de intermediação** (*betweenness*) considera todos os caminhos mínimos entre pares de vértices de uma rede, e os vértices que pertencem a um número maior de caminhos mínimos são os que possuem maior intermediação. A centralidade de intermediação de um vértice v é definida como $C_B(v) = \sum_{\{s \neq v \neq t \in V\}} \sigma_{st}(v) / \sigma_{st}$, onde $\sigma_{st}(v)$ denota o número de caminhos mínimos entre s e t , que passam por v e σ_{st} denota o número de caminhos mínimos entre s e t .

Dado um grafo G onde os vértices estão numerados de 1 até $n = |V|$ e seja $A = (a_{ij})$ a matriz de adjacência do grafo G . Seja x_i , $1 \leq i \leq n$, o *score* do i -ésimo vértice do grafo definido da maneira seguinte: $x_i = (\sum_{j=1, \dots, n} a_{ij} x_j) / \lambda$, onde λ é o maior autovalor da matriz A . A centralidade de autovetor determina um *score* para cada vértice que é proporcional aos *scores* dos seus vértices vizinhos. Portanto, a **centralidade de autovetor** é definida como $C_A(v) = (\sum_{j=1, \dots, n} a_{ij} x_j) / \lambda$, ou o componente do vetor de Perron da matriz A (autovetor associado ao maior autovalor, λ , da matriz de adjacência do grafo).

O valor de **Pagerank** de um vértice do grafo numerado i , $1 \leq i \leq n$ é definido pelo componente x_i , $1 \leq i \leq n$, do vetor que é solução da equação

$$x = D(D - \alpha A)^{-1} \mathbf{1}, \quad (1)$$

onde $A = (a_{ij})$ é a matriz de adjacência do grafo, $D = (d_{ii})$ a matriz diagonal dos graus dos vértices do grafo, $\mathbf{1}$ é o vetor com todos os componentes iguais a um e α é um parâmetro a ser ajustado pelo usuário ($\alpha=0,85$ é um valor frequentemente usado). O **Pagerank** é considerado uma

generalização do valor de centralidade de autovetor, pois da equação $x_i = (\sum_{j=1, \dots, n} a_{ij} x_j) / \lambda$ ou na forma vetorial $(1/\lambda)Ax = x$, é generalizada pela adição de constantes α e β e a matriz D^{-1} na forma $x = \alpha(AD^{-1})x + \beta I$, para o valor de $\beta=I$ (equivalente com a equação dada em (1)).

4 Detecção de comunidades

Segundo Fortunato ([8]), uma **comunidade**, também denominada *cluster* ou módulo, é formada por um grupo de vértices que possuem propriedades comuns, e/ou desempenham uma função similar dentro de uma rede dada. As comunidades são utilizadas em diversas aplicações no contexto de redes reais e permitem fazer uma análise estrutural da mesma. No caso de redes de grande porte, determinar comunidades permite o armazenamento eficiente da rede num computador. Identificar módulos e seus limites permite classificar os vértices que o compõem, desde o ponto de vista da posição estrutural dentro do módulo. Outro aspecto importante relacionado com a estrutura de comunidade de uma rede é a hierarquia determinada entre os integrantes da rede. Geralmente, as redes reais são formadas por comunidades que contém comunidades menores, que por sua vez contém outras comunidades. Determinar comunidades em grafos é um assunto bastante explorado na área de computação. A formulação matemática deste problema é denominada particionamento de grafos (*graph clustering*).

O problema de determinar um k -particionamento mínimo de um grafo é um problema de otimização combinatória que consiste em, dado um conjunto de dados D e uma função de distância $d : D \times D \rightarrow N$ tal que d satisfaz a desigualdade triangular, determinar uma partição do conjunto D em k *clusters* C_1, \dots, C_k de maneira tal que C_i, C_j são disjuntos; para i distinto de j e tal que a maior distância entre pares de *clusters* seja minimizada (Shaeffer, [18]).

4.1 Algoritmos de Girvan e Newman e de Clauset, Newman e Moore

Os dois algoritmos descritos a seguir são muito utilizados na análise de redes sociais para detecção de comunidades.

Girvan e Newman ([9]) propõem um algoritmo heurístico baseado no conceito de intermediação, parecido com o definido na Secção 2.2, para determinar *clusters* de um grafo não direcionado $G = (V, E)$. Os autores assumem que arestas com maior valor de intermediação (generalização da medida de centralidade definida para vértices) são ligações entre *clusters*. Dessa forma, a rede é dividida eliminando, uma por vez, as arestas do grafo com maior valor de intermediação, até atingir o critério de parada. Após a remoção de uma aresta, os valores de intermediação e os caminhos mínimos são recalculados, caso seja necessário. O algoritmo proposto por Girvan e Newman pode ser implementado com complexidade de tempo $O(n m^2)$ para um grafo com n vértices e m arestas.

Outro algoritmo foi proposto por Clauset, Newman e Moore ([6]) que apresenta complexidade de tempo $O(n \log^2(n))$ para grafos não direcionados e esparsos com n vértices. Um índice para medir a qualidade de um particionamento de um grafo que representa uma rede é denominado modularidade e foi introduzido por Newman e Girvan em 2004 ([16]). Dada uma partição do grafo em *clusters* C_1, \dots, C_k , a **modularidade** da partição, notada por $M(C_1, \dots, C_k)$, é definida como: $M(C_1, \dots, C_k) = \sum_{i=1, \dots, k} \varphi_{i,i} - \sum_{i \neq j, i,j \in \{1, \dots, k\}} \varphi_{i,j}$ onde $\varphi_{i,j} = \sum_{(u,v) \in E, u \in C_i, v \in C_j} w(e)$, onde $w(e)=1$ se o grafo não tiver peso associado às arestas. Dessa forma o grau interno e o grau externo de cada *cluster* pode ser definido, em termos desta medida de modularidade, como $\text{grau}_{\text{int}}(C_i) = \varphi_{i,i}$ e $\text{grau}_{\text{ext}}(C_i) = \sum_{i \neq j, i,j \in \{1, \dots, k\}} \varphi_{i,j} - \varphi_{i,i}$ para $1 \leq i \leq k$. O problema de determinar um particionamento do grafo com valor máximo de modularidade é NP-Completo [5]. O método proposto por Clauset et al. faz uma otimização gulosa da modularidade, utilizando estruturas de dados mais sofisticadas que o algoritmo anterior.

Tanto o algoritmo de Girvan e Newman quanto o algoritmo de Clauset et al. são métodos heurísticos para particionar um grafo G . Os mesmos determinam um conjunto de *clusters* C_1, \dots, C_k que formam uma partição do conjunto de vértices de G de maneira tal que $C_i \cap C_j$ é vazio; quando $i \neq j$.

5 Resultados

5.1 Construção do grafo de co-autoria

Para a construção do rede de co-autoria deste trabalho, foram retirados dados da *Plataforma Lattes* do site do *Conselho Nacional de Pesquisa* (CNPq). Nessa plataforma é possível obter dados dos pesquisadores como: área de pesquisa, formação acadêmica, publicações em periódicos, publicações em congressos, etc. Para determinar a escolha dos pesquisadores a ser incluídos no grafo, foi realizada uma busca avançada no site [7]. Buscou-se a palavra chave "grafos" dentre o conjunto de pesquisadores com bolsa de produtividade em pesquisa. Foram selecionados os pesquisadores com resultados de grau de aderência maior ou igual que 90% em 11/10/2010. Dessa forma, foram escolhidos 13 pesquisadores. Para cada pesquisador, extraiu-se da seção "Artigos completos publicados em periódicos" do respectivo currículo *Lattes* a informação correspondente aos artigos publicados em periódicos, incluindo os co-autores dos mesmos. A partir dessa informação, foi criado um o grafo de co-autoria, contendo 207 vértices e 520 arestas, cujos vértices são os 13 pesquisadores e seus co-autores e existe uma aresta entre dois pesquisadores, se eles tiveram pelo menos uma publicação conjunta.

O um desenho do grafo obtido é apresentado na Figura 1. As informações completas sobre a rotulação do grafo estão disponíveis em [2].

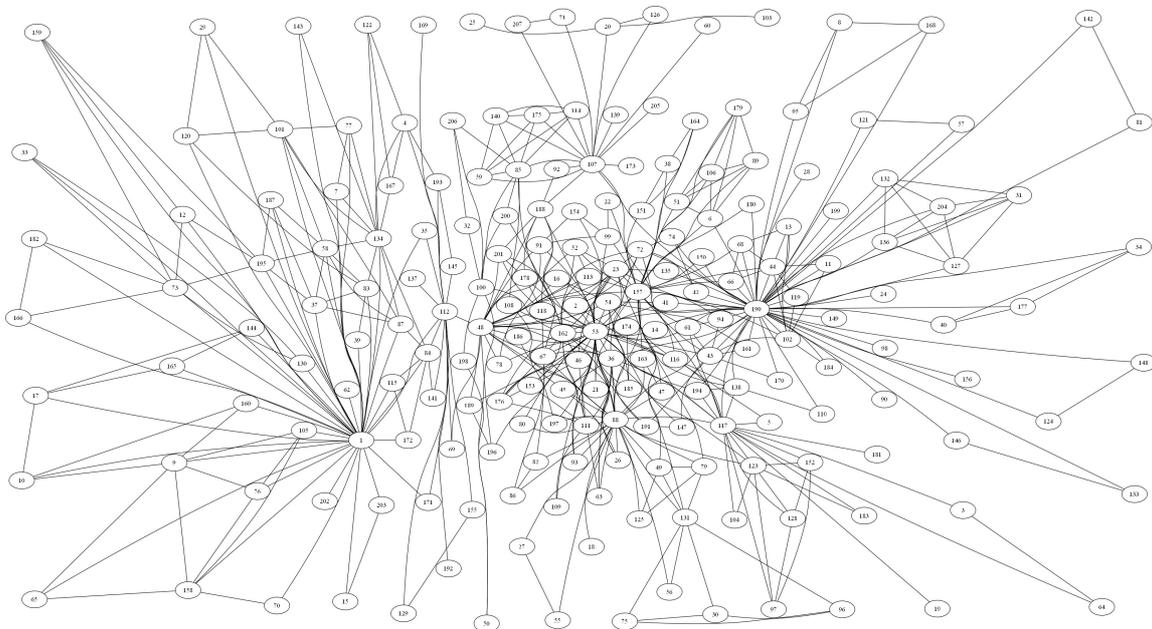


Figura 1: grafo de co-autoria

5.2 Medidas de centralidade no grafo de co-autoria

Para determinar os valores das medidas de centralidade apresentadas na Seção 2.2, foi utilizada a biblioteca *SNAP* (*Stanford Network Analysis Platform*, [19]). A biblioteca *centr.h* da plataforma *SNAP* determina, para o grafo de entrada em formato de lista de arestas, o valor correspondente a cada vértice para cada uma das cinco medidas de centralidade. O cálculo das medidas de centralidade utilizados nos algoritmos implementados nesta biblioteca tem algumas particularidades. A medida de proximidade utiliza a fórmula definida na Seção 3. No cálculo do *Pagerank*, o valor *default* para o parâmetro α é 0,85 e pode ser alterado pelo usuário. A medida de proximidade é calculada usando o algoritmo de Busca em Largura (*Breadth First Search*, *BFS*). Como a *BFS* determina caminhos com menor número de arestas desde um vértice dado, a medida de proximidade implementada na biblioteca unicamente pode ser usada no caso de grafos sem pesos nas arestas. As outras quatro medidas podem também ser usadas no caso de grafos com pesos nas arestas.

Usando a biblioteca mencionada, foram calculados os valores das medidas para cada vértice do grafo de co-autoria descrito na seção anterior. A Tabela 1 apresenta os vértices do grafo de co-autoria com maior valor de centralidade para cada uma das 5 medidas (ordenados pelos valores das respectivas medidas, em ordem decrescente). As informações completas sobre as medidas de centralidade para todos os vértices do grafo da Figura 1 estão disponíveis em [2].

grau	proximidade	intermediação	autovetor	PageRank
190	53	48	53	190
53	48	112	190	1
1	88	190	88	53
157	157	1	157	157
88	190	53	48	88

Tabela 1: Os 5 vértices com maior valor para cada medida de centralidade (ordenados)

5.3 Obtenção de comunidades no grafo de co-autoria

Utilizando a biblioteca *cmt.h* da plataforma *SNAP* foram determinadas as comunidades para o grafo de co-autoria anterior, segundo o algoritmo de Girvan e Newman (GN) e segundo o algoritmo de Clauset, Newman e Moore (GNM). Os arquivos de saída dos dois algoritmos para o grafo da Figura 1 estão disponíveis em [2].

O software para detectar comunidades usando o algoritmo GN permite calcular o valor exato da intermediação de cada aresta do grafo (cuja definição é semelhante à definição de intermediação de um vértice), ou escolher um cálculo aproximado da mesma para o caso de grafos de tamanho maior.

O algoritmo GN produziu 8 comunidades para distribuir os vértices do grafo de co-autoria. Uma figura do grafo mostrando as 8 comunidades foi elaborada a partir da saída do algoritmo GN, usando o *Graphviz* ([10]), um *software* de visualização gráfica, *Open Source*, desenvolvido pela *AT&T Labs Research*. O desenho da saída produzida pelo algoritmo GN é mostrado na Figura 2.

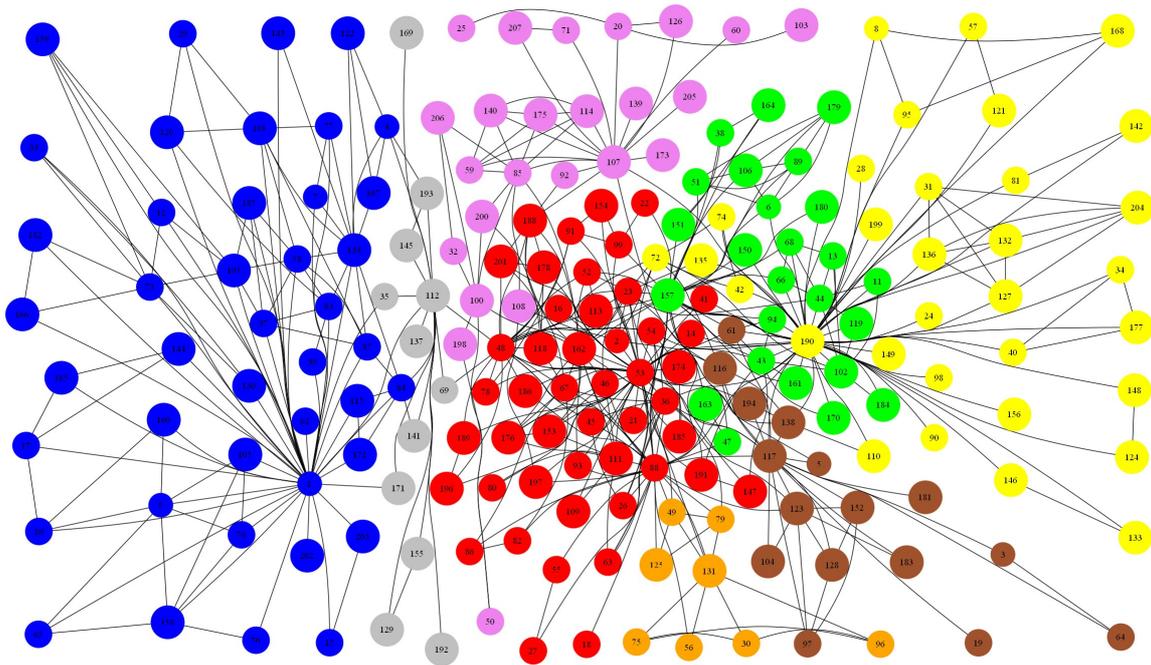


Figura 2: visualização das 8 comunidades obtidas pelo algoritmo GN

Já o algoritmo CNM obteve 7 comunidades para distribuir os vértices do mesmo grafo. A Figura 3 mostra um desenho do grafo obtido a partir da saída do algoritmo CNM.

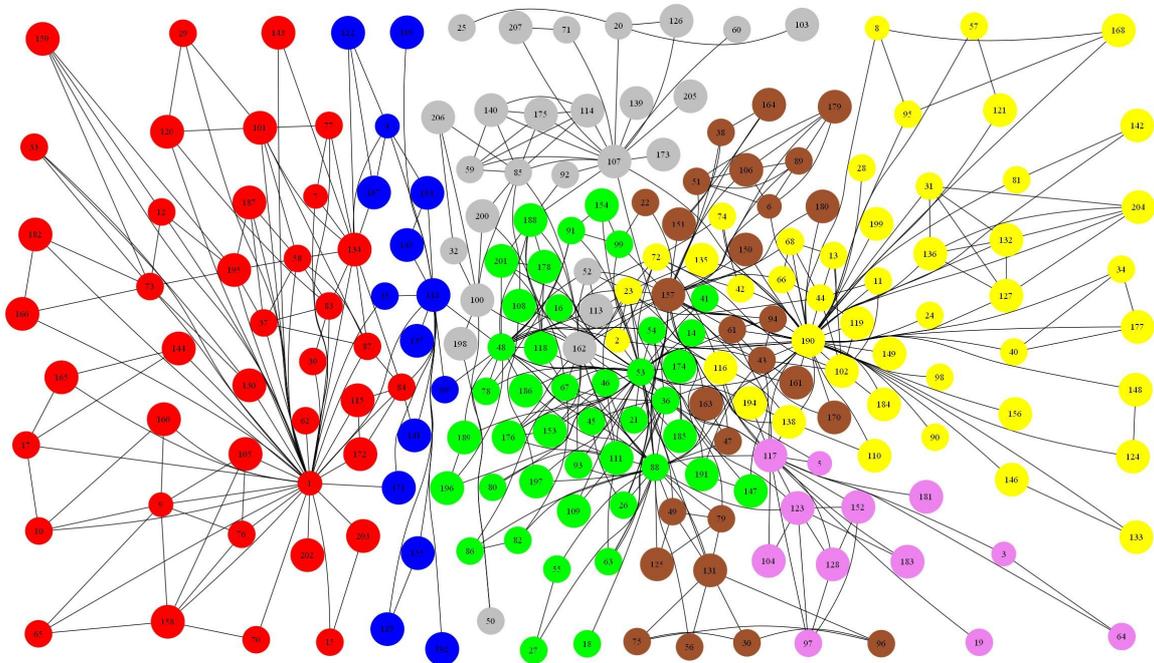


Figura 3: visualização das 7 comunidades obtidas pelo algoritmo CNM

Os dois algoritmos utilizados para detectar comunidades produziram número de comunidades diferentes. O algoritmo CNM determinou uma comunidade a menos que o algoritmo GN. Os vértices pertencentes a cada comunidade diferem para cada algoritmo. Podemos observar que os vértices 190, 53 são os dois vértices mais centrais para todas as medidas, o vértice 157 dentre os 5 vértices mais centrais para 4 medidas (todas menos intermediação) e o vértice 1 dentre os 5 vértices mais centrais para 3 medidas (grau, intermediação e *pagerank*). Esses quatro vértices pertencem a comunidades diferentes pelo resultado dado nos dois algoritmos. Já o vértice 88, está entre os 5 vértices mais centrais para 4 medidas (grau, proximidade, autovetor e *pagerank*) e pertence a mesma comunidade que o vértice 53, no resultado obtido pelos dois algoritmos.

Nas duas tabelas a seguir, é apresentado um resumo das informações obtidas sobre os vértices mais centrais com respeito às comunidades obtidas para cada algoritmo as quais eles pertencem. Na Tabela 2, cores iguais determinam vértices na mesma comunidade obtida pelo algoritmo de Girvan e Newman:

grau	proximidade	intermediação	autovetor	PageRank
190	53	48	53	190
53	48	112	190	1
1	88	190	88	53
157	157	1	157	157
88	190	53	48	88

Tabela 2: Os 5 vértices mais centrais (ordenados) vs. comunidades (GN)

Na Tabela 3 cores iguais determinam vértices na mesma comunidade obtida pelo algoritmo de Clauset, Newman e Moore:

grau	proximidade	intermediação	autovetor	PageRank
190	53	48	53	190
53	48	112	190	1
1	88	190	88	53
157	157	1	157	157
88	190	53	48	88

Tabela 3: Os 5 vértices mais centrais (ordenados) vs. comunidades (CNM)

A maior parte das comunidades obtidas pelos dois algoritmos, diferem em alguns vértices. Observar também que os vértices numerados 4, 108, 52, 113, 162, 22 foram alocados em comunidades diferentes pelos dois algoritmos. Podemos considerar que esses vértices pertencem à fronteira das comunidades obtidas pelos dois algoritmos.

Os vértices 96, 75, 131, 30, 49, 56, 79, 125 formam uma comunidade diferente na saída do algoritmo GN, fato que não acontece com a saída do CNM (na saída deste último, esses 8 vértices pertencem à mesma comunidade que o vértice 157). Neste caso podemos interpretar que devido a diferencia de critérios para determinar comunidades utilizados pelos algoritmos de GN e CNM (eliminação de arestas com maior valor de intermediação e maximização da modularidade de maneira aproximada) os resultados sejam diferentes. Um estudo dos vértices

que compõem as comunidades em ambos os casos pode determinar qual dos resultados é mais significativo.

6 Conclusões

A análise da estrutura da rede de co-autoria, construída neste trabalho, determinou os vértices de maior valor de centralidade, para as 5 medidas definidas na Seção 2 e detectou comunidades na mesma, utilizando dois algoritmos clássicos para tal fim descritos na Seção 4.

A construção do grafo foi realizada extraindo dados de um arquivo de texto criado a partir das informações declaradas pelos 13 pesquisadores do CNPq, e seus co-autores. Foi necessário realizar um trabalho de correção do grafo obtido, devido a que as informações declaradas no currículo dos pesquisadores não são padronizadas. Foram detectados vários erros de digitação nos nomes dos co-autores.

Para visualizar os resultados obtidos pelos algoritmos de detecção de comunidades no grafo de co-autoria foi necessário criar uma interface que permitiu adaptar os resultados obtidos com o pacote da biblioteca *SNAP* para obter uma entrada do software de visualização de grafos, *Graphviz*. A implementação desta interface foi realizada na linguagem de programação *C++*.

Acreditamos que, do ponto de vista prático, os resultados obtidos com o estudo das comunidades e as medidas de centralidade podem ser utilizados para determinar os principais autores, grupos e, como consequência, os principais assuntos pesquisados no Brasil na área de grafos. O estudo da composição dos grupos poderia mostrar a participação de pesquisadores das diferentes regiões e instituições do país, assim como também a participação de pesquisadores do exterior. Os vértices mais centrais dentro de uma mesma comunidade podem indicar características dos trabalhos dos pesquisadores. Assim como os vértices na fronteira das comunidades podem destacar interlocutores entre os grupos de pesquisa.

Como trabalhos futuros podemos mencionar a continuação do estudo de comunidades no mesmo grafo de co-autoria, utilizando outros algoritmos. Os algoritmos de Newman ([14]), Bondel et. al ([4]) e Rosval e Bergitom ([17]), utilizam abordagens diferentes das apresentadas neste trabalho. Propomos implementar os algoritmos referenciados anteriormente e fazer uma análise comparativa com os resultados obtidos neste artigo. Outro aspecto do trabalho que não foi abordado neste artigo, consiste na utilização de métricas para avaliar comunidades ([5], [15]). Propomos também como trabalho futuro fazer uma análise das comunidades obtidas por diferentes algoritmos, utilizando as métricas mais adequadas existentes na literatura.

Agradecimentos: Os autores agradecem ao CNPq (Projeto N° 474754/2010-3) e CAPES-DS pelo financiamento parcial dessa pesquisa.

Este trabalho faz parte do Projeto de Final de Curso de graduação dos três primeiros autores e da pesquisa da dissertação de mestrado da quarta autora.

7. Referências bibliográficas

[1] Abreu, N.M.M., Del-Vecchio, R.R., Vinagre, C.T.M., Stevanovic, D. (2007) *Introdução à Teoria Espectral de Grafos com Aplicações*. SBMAC.

[2] Barbosa, D.A.B.L., Avelino, L.B., Sousa, R.F., Justel, C.M. (2011) Resultados experimentais em grafo de co-autoria: medidas de centralidade e detecção de comunidades. *Monografias em Sistemas e Computação 02/2011*, Seção de Engenharia de Computação, SE8, Instituto Militar de Engenharia.

http://www.comp.ime.eb.br/techreports/repositorio/2011_02.pdf

- [3] **Batagelj, V. Mrvar, A.** (2000) Some analyses of Erdős collaboration graph. *Social Networks* 22, pp. 173-186.
- [4] **Blondel, V. D., Guillaume, J-L., Lambiotte, R., Lefebvre, E.** (2008) Fast unfolding of communities in large networks, *J. Stat. Mech.* P10008.
- [5] **Brandes, U., Delling, D., Gaertler, M., Görke, R., Hofer, M., Nikoloski, Z., Wagner, D.** (2008) On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering* 20(2), pp.172-188.
- [6] **Clauset, A., Newman, M.E.J., Moore, C.** (2004) Finding community structure in very large networks. *Physical Review E*, 70:066111.
- [7] **CNPq, Plataforma LATTES** disponível em: <http://buscatextual.cnpq.br/buscatextual/busca.do?metodo=apresentar>. Acesso em out. 2010.
- [8] **Fortunato, S.** (2010) Community detection in graphs, *Physics Reports* 486, pp. 75-174, arXiv:0906.0612v2.
- [9] **Girvan, M., Newman, M.E.J.** (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), pp.7821–7826.
- [10] **Graphviz** disponível em: www.graphviz.org Acesso em mar. 2011.
- [11] **Kleinberg, J. Easley, D.** (2010) *Networks, crowds and markets*, Cambridge University Press.
- [12] **Microsoft Academic Search** disponível em: <http://academic.research.microsoft.com/>. Acesso em abr. 2011.
- [13] **Newman, M.E.J.** (2001) Scientific collaboration networks. I. Network construction and fundamental results, *Physical Review E* 64, 016131.
- [14] **Newman, M.E.J.** (2006) Finding community structure in networks using the eigenvectors of matrices, *Physical Review E* 74(3), 036104.
- [15] **Newman, M.E.J.** (2010) *Networks: an Introduction*, Oxford University Press.
- [16] **Newman, M.E.J., Girvan, M.** (2004) Finding and evaluating community structure in networks, *Physical Review E* 69, 026113
- [17] **Rosvall, M., Bergstrom, C.T.** (2008) Maps of random walks on complex networks reveal community structure, *PNAS*, Vol. 105 No. 4 pp. 1118-1123.
- [18] **Shaeffer, S.E.** (2007) Graph clustering, *Computer Science Review I*, pp. 27-64.
- [19] **SNAP**, disponível em: www.snap.stanford.edu. Acesso em mar. 2011.
- [20] **Szwarcfiter, J.L.** (1988) *Grafos e algoritmos computacionais*, Ed Campus.