

Estudo comparativo de métricas de ranqueamento em redes complexas utilizando coeficientes de correlação

João Gabriel Rocha Silva
Carolina Ribeiro Xavier
Vinícius da Fonseca Vieira
Departamento de Ciência da Computação
Universidade Federal de São João del Rei
São João del Rei, Brasil 36307-352
carolinaxavier@ufsj.edu.br

Iago Augusto de Carvalho
Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil 31270-901
iagoac@ufmg.br

Resumo—A classificação e o ranqueamento de vértices é um tema muito estudado em redes complexas. Existem na literatura diversas métricas utilizadas na classificação de vértices em uma rede. Este trabalho visa comparar as diferentes métricas calculando o coeficiente de correlação entre elas. Resultados demonstram que as métricas Grau e Hub apresentam a maior correlação, ranqueando os vértices de maneira mais similar, enquanto as métricas Hub e PageRank obtiveram o menor coeficiente de correlação.

Keywords—Coeficiente de correlação, Redes Complexas, Ranqueamento, Comparação, Comunidades.

I. INTRODUÇÃO

Sistemas reais, em geral, podem ser representados com uma rede, onde os vértices são os elementos e as arestas são as relações entre eles. Entre as principais redes estudadas estão as redes de colaboração entre os pesquisadores [1] ou a rede de colaboração de atores em filmes [2].

Redes podem ser divididas em comunidades, um subconjunto de vértices com um grande conjunto de arestas conectando seus vértices e um pequeno conjunto de arestas conectando a comunidade a vértices externos a ela [3], [4].

O estudo de redes complexas em sua forma aplicada mostra-se de natureza interdisciplinar, visto a sua gama de aplicação em áreas do conhecimento variadas como biologia, ciências sociais, matemática e engenharia. Sistemas provenientes das mais diversas áreas de conhecimento podem ser facilmente modelados como redes complexas, devido a particularidades das redes complexas que permitem esta fácil utilização, adaptação e aplicação [5], [6], [3].

Estes estudos podem ser realizados através do ranqueamento de vértices. O ranqueamento dos vértices de uma rede a partir de algum critério de classificação pode ser utilizada para obter uma série de dados relevantes sobre o problema em questão [7]. Em redes sociais é possível inferir quais são os usuários mais influentes da rede, suas conexões e o modelo de propagação de informações entre os grupos sociais. Em redes de Internet podemos descobrir quais são as páginas mais relevantes, com o objetivo estudar como estes *websites* relacionam-se entre si e com o restante da rede.

Existem na literatura uma série de métricas que podem ser utilizadas como formas de mensurar as centralidades das com-

plexas estruturas destas redes. Dentre elas, podemos destacar o Grau [8], o PageRank [9], o Hub [10], [11] e o Authority [10], [11]. Onde as quatro métricas citadas geram como saída um *score* para cada vértice. Estes vértices, quando ordenados por seus *scores*, geram um *ranking*.

Este trabalho tem como objetivo analisar o *score* dos vértices de cada rede utilizando quatro diferentes métricas: Grau, PageRank, Hub e Authority. Esta análise é realizada através da comparação dos coeficientes de correlação dos *rankings* gerados pela comparação das métricas em redes do contexto social, de conexões *peer-to-peer* e de colaboração científica. Para cada rede, é calculado o coeficiente de correlação das métricas utilizadas, par a par. Além disto, é realizada uma análise das correlações de diferentes comunidades de uma mesma rede, utilizando as mesmas métricas.

II. REDES COMPLEXAS

O estudo de *redes complexas* tem origem na teoria dos grafos. A teoria dos grafos focou-se inicialmente em *grafos regulares*, sendo que a partir do ano de 1950 grandes redes sem nenhuma configuração evidente foram descritas como grafos aleatórios por [8]. Esta foi a primeira e mais simples descrição de uma rede complexa.

Estruturas modeladas como redes complexas podem descrever uma grande variedade de sistemas. Uma célula é melhor descrita como uma rede complexa de elementos químicos conectados por reações químicas. A Internet pode ser modelada como uma rede complexa de computadores e roteadores interconectados por vários enlaces físicos ou *wireless*. O relacionamento interpessoal pode ser representado como uma rede complexa de indivíduos conectados por suas relações de amizade. A *Web* pode ser descrita como uma grande rede complexa de páginas conectadas por seus *hyperlinks* [12], [8].

A. Elementos básicos

Sendo um campo que emergiu da teoria dos grafos, as redes complexas herdaram muitos de seus elementos básicos [13]. Dado uma rede complexa, define-se como um vértice ou nó da rede sua unidade fundamental, que pode representar objetos de um conjunto, seja ele abstrato ou real. Uma aresta pode ser definida como uma conexão ou relação entre dois vértices

de uma rede. Mais além, uma rede pode ser direcionada ou não direcionada. A primeira é obtida quando arestas não possuem uma direção, e a relação entre dois vértices de uma rede é transitiva. Já uma rede não direcionada possui arestas direcionadas e indicam relações intransitivas entre os vértices de uma rede. Por último, pode-se definir o grau de um vértice como sendo o número de conexões de um vértice com o restante da rede. Esta medida é única para redes não direcionadas ou dois valores (grau de entrada e grau de saída) para redes direcionadas.

Além destes conceitos básicos, pode-se definir a resistência de uma rede como a medida de robustez perante a falha de vértices, i.e. o quão robusta é a rede em relação a retirada de vértices de tal forma que a rede não seja comprometida. Pode-se definir também o coeficiente de agrupamento como um coeficiente que avalia o grau com que os vértices de uma rede tendem a se agrupar, e é baseado no conceito de trios de vértices conectados entre si. Um trio consiste em três vértices que se encontram conectados por duas ou três arestas. Um trio conectado por duas arestas forma um trio aberto, enquanto um trio conectado por três arestas forma um trio fechado. Um triângulo consiste em três trios fechados, um em cada vértice de um trio. O coeficiente de agrupamento mede o número total de trios fechados sobre o número total de trios.

Um último conceito aqui definido é o coeficiente de assortatividade da rede. Ele pode ser delimitado como a tendência de um vértice se ligar a outros vértices com graus similares. O valor de assortatividade pode variar no intervalo $[-1; 1]$. Um coeficiente de assortatividade negativo significa que os vértices tendem a se conectar com vértices de graus diferentes. Já um coeficiente positivo significa que os vértices tendem a se conectar com outros vértices de grau próximos ao seu.

B. Comunidades

A noção de comunidade, e a primeira formalização deste conceito em redes, foi proposta no âmbito das Ciências Sociais [14]. Dado um grafo, define-se uma comunidade como um subgrafo $G'(V', A')$ tal que os vértices em V' estão fortemente conectados entre si pelas arestas A' e os vértices em $V \setminus V'$ estão fracamente conectados aos vértices em V' [12]. Uma rede pode possuir inúmeras diferentes comunidades.

O conceito de modularidade [15] de uma rede representa a medida da qualidade da partição de uma rede em comunidades. Atualmente, a modularidade é o conceito mais utilizado para avaliar esta divisão da rede [7].

A divisão de uma rede complexa em comunidades não é uma tarefa simples. O problema de maximização da modularidade de uma rede foi provado ser fortemente \mathcal{NP} -completo em [16]. Desta forma, uma série de algoritmos heurísticos foram propostos visando a maximização da modularidade. Foram utilizadas abordagens por aglomeração gulosa [17], divisão espectral [18], Evolução Diferencial [19], dentre outros. Além destes, uma abordagem por programação matemática foi proposta em [20].

III. METODOLOGIA

Com o objetivo de comparar diferentes métricas para classificação e ranqueamento de vértices em redes complexas as medidas de Grau, PageRank, Hub e Authority foram

aplicadas em uma série de redes de diferentes tipos. Os tipos de bases de redes complexas utilizadas foram bases de redes sociais, de conexões *peer-to-peer* (P2P) e redes de colaboração de pesquisas científicas. Onde para cada tipo de bases, quatro diferentes redes foram analisadas.

A. Métricas utilizadas

Cada rede foi avaliada utilizando quatro diferentes métricas. Os vértices de cada rede foram ranqueados por seu *score* de forma decrescente, obtendo quatro diferentes listas ordenadas, cada lista relativa a uma métrica.

Estão presentes na literatura uma série de métricas que retornam centralidades de um determinado vértice em uma rede, no entanto as métricas selecionadas apresentam o critério eficiência e custo relativamente mais baixo que outras medidas como por exemplo *Betweenness* (intermediação) e *Closeness* (centralidade de proximidade), este último gerando valores muito próximos para a maioria dos nós da rede, gerando um *ranking* inviável.

Os algoritmos foram implementados em Python versão 2.7.9 com a utilização da biblioteca *iGraph* [21] para construir as listas ordenadas de vértices de cada uma das diferentes métricas.

1) *Grau*: O grau de um vértice pode ser definido como na Seção II. O cálculo para esta métrica consiste apenas na contagem de vértices que são adjacentes ao vértice avaliado.

2) *PageRank*: A métrica *PageRank* foi proposta em [9]. Ela foi utilizada pelo *Google* para avaliar a importância de páginas *Web* e estendida para outros contextos em redes complexas em diversos trabalhos. O *Google* utilizava o *PageRank* para distinguir a qualidade das citações e, conseqüentemente, obter uma medida de importância para uma página levando em consideração as importâncias das páginas que possuem *hyperlinks* para ela e a importância das páginas para a qual ela possui *hyperlinks*.

Esta métrica classifica os vértices de acordo com seu grau e a classificação dos vértices adjacentes a ele. Para obter um *score* alto utilizando esta métrica é necessário se relacionar com vértices de *score* alto e que estes vértices, por sua vez, também se relacionem com vértices de alto *score* alto.

3) *Hub e Authority*: As métricas *Hub* e *Authority* são interdependentes. Para cada vértice $i \in V$, o coeficiente de *Hub* é obtido realizando-se a soma do *score Authority* dos vértices que o vértice i aponta. De forma similar, o coeficiente de *Authority* de um vértice $j \in V$ é obtido realizando-se a soma dos *scores Hub* dos vértices que apontam para o vértice j .

Estes identificadores foram utilizados a priori para aplicações em modelos de páginas *Web*, onde para a extração da autoridade da rede a análise são levados em consideração definir um sub-grafo da *WWW* a ser utilizado pelo algoritmo, no qual o sub-grafo deve unir características como boa parte das grandes autoridades e autoridades relevantes, assim sendo, serão retornadas as páginas autoridades (muito referenciadas) e páginas que referenciam páginas autoridades (*Hubs*), o algoritmo responsável por este cálculo é o *HITS* [10], onde considera-se um bom *Hub* o vértice que aponta para vértices

com alto grau de *Authority*, e um bom *Authority* aquele vértice que é apontado por bons *Hubs*.

B. Bases de dados

Para a realização deste trabalho foram utilizadas três diferentes tipos de bases. Devido a natureza comparativa deste trabalho, fez-se necessária a escolha de redes de diferentes contextos. Para tanto, selecionamos redes sociais, redes de conexões P2P e redes de colaborações em trabalhos científicos. Todas as bases foram coletadas do *website* SNAP [22]. As bases foram escolhidas devido a diversidade de características entre suas topologias. Desta forma é possível avaliar o comportamento de cada contexto perante as correlações de métricas em uma série de ambientes heterogêneos.

As bases de dados de redes sociais utilizadas são redes direcionadas. Uma rede social não possui tendências de ligações entre seus indivíduos, tornando seu valor de assortatividade próximo a 0. Em contrapartida, pode-se afirmar que pessoas influentes em uma rede social tendem a se conectar com outras pessoas influentes usuárias da mesma rede.

Quatro redes sociais foram utilizadas. A rede *S1* foi obtida em Novembro de 2008, extraída de um site de notícias denominado *Slashdot*. Neste site é permitido que se relacionem pelo portal. A rede *S2* foi obtida em Fevereiro de 2008, extraída também das relações entre usuários do site *Slashdot*. A rede *S3* é uma rede social *on-line* do site de avaliação geral dos consumidores *Epinions.com*. A rede *S4* representa a rede de votação para administradores do site *Wikipedia*. Os dados referentes as quatro redes sociais estão dispostos na Tabela I.

Base	Número de nós	Número de Arestas	Assortatividade
<i>S1</i>	77,360	905,468	-0.0488503420029
<i>S2</i>	82,168	948,464	-0.0516365922677
<i>S3</i>	75,888	508,837	-0.0412863005313
<i>S4</i>	8,298	103,689	-0.0832445577169

Tabela I. REDES SOCIAIS

Assim como as redes sociais, as redes de conexão P2P são exemplos de redes direcionadas. Nestas redes, um computador pode estar conectado a outro sem que exista reciprocidade nesta conexão. Redes P2P não possuem tendências de conexão, de forma que sua assortatividade tende a 0. Nestas redes, computadores de grau alto se conectam com computadores de diferentes valores de grau. Desta forma, quando houver alguma falha na rede, esta não sofre fortes danos.

Quatro redes de conexão P2P são utilizadas, provenientes do serviço de compartilhamento de arquivos *Gnutella*. Cada base representa um *snapshot* da rede em um determinado estado de tempo, a partir de Agosto de 2002. Os dados referentes a estas quatro redes P2P são apresentados na Tabela II.

Base	Número de nós	Número de Arestas	Assortatividade
<i>P1</i>	26,518	65,369	-0.0055829202754
<i>P2</i>	22,687	54,705	-0.0062192970313
<i>P3</i>	36,682	88,328	-0.0214005150226
<i>P4</i>	62,586	147,892	-0.0062850609742

Tabela II. REDES P2P

Redes de colaboração de pesquisas científicas diferem fortemente das redes anteriores. São redes não direcionadas, e seu valor de assortatividade é relativamente alto, indicando que vértices de graus similares conectam entre si. Em termos gerais, isto pode indicar que pesquisadores importantes colaboram com pesquisadores importantes. A mesma afirmação é válida para pesquisadores não tão influentes dentro do meio científico. Por se tratar de redes não direcionadas, não é calculado o coeficiente de correlação entre *Hub* e *Authority*, pela própria definição das métricas.

Quatro redes de colaboração científica são utilizadas. A rede *C1* representa a colaboração científica de autores entre os trabalhos submetidos a revista *Relatividade Geral e Cosmologia Quântica*. Já a rede *C2* modela as colaborações científicas entre autores de trabalhos submetidos a revista *Física de Altas Energias* na categoria teórica. A rede *C3* foi extraída das colaborações científicas entre autores de trabalhos submetidos a revista *Astronomia e Física*. A rede *C4* representa a colaboração científica entre autores de trabalhos submetidos a revista *Física de Altas Energias*. Os dados referentes a estas quatro redes de colaboração científica são apresentados na Tabela III.

Base	Número de nós	Número de Arestas	Assortatividade
<i>C1</i>	26,197	28,980	0.659286800266
<i>C2</i>	68,746	51,971	0.267944054894
<i>C3</i>	133,280	396,160	0.205187143464
<i>C4</i>	89,209	237,010	0.632286309198

Tabela III. REDES DE COLABORAÇÃO CIENTÍFICA

C. Cálculo dos coeficientes de correlação

Uma forma de obter uma comparação concreta entre *rankings* gerados por diferentes métricas em uma base é através do cálculo do *coeficiente de correlação*, proposto por Charles Spearman [23]. Ele indica o quão relacionado um elemento encontra-se na classificação por uma métrica em relação a outra, considerando seu posto no *ranking*. Ele leva em consideração a diferença de um posto em dois métodos de classificação distintos bem como o número de pares de valor que compõe a amostra. O cálculo do coeficiente de correlação é dado como na Equação (1).

Onde d_i representa a diferença de classificação de um mesmo elemento em duas métricas, i.e. a diferença de posto, e n representa o número de pares correspondentes nas amostras, i.e. o número de elementos de cada *ranking*.

Além do coeficiente de correlação, pode-se utilizar também o *coeficiente de Kendall-tau* [24]. Ele leva em consideração o número de pares concordantes e discordantes entre os conjuntos de dados, i.e. a quantidade de elementos que estão posicionados no mesmo posto em ambos os *rankings* e a quantidade de elementos que estão posicionados em postos diferentes. O coeficiente de Kendall-tau é descrito na Equação (2).

$$\rho = 1 - \frac{6 \sum d_i^2}{(n^3 - n)} \quad (1)$$

onde n_c representa o número de pares concordantes; n_d

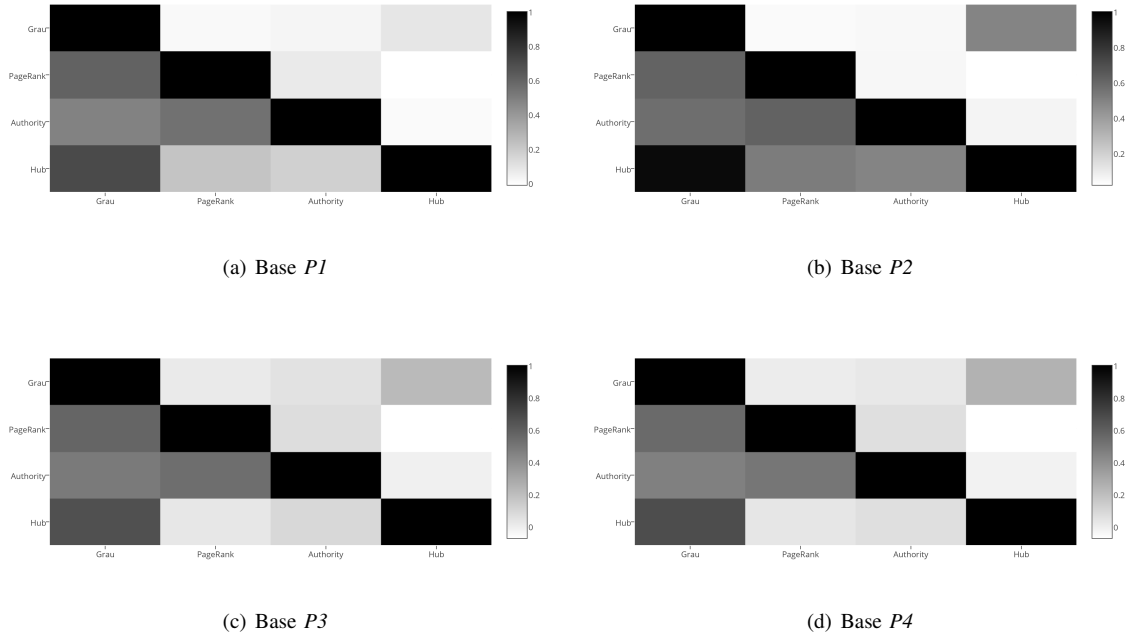


Figura 1. Gráficos de calor para bases de conexões P2P

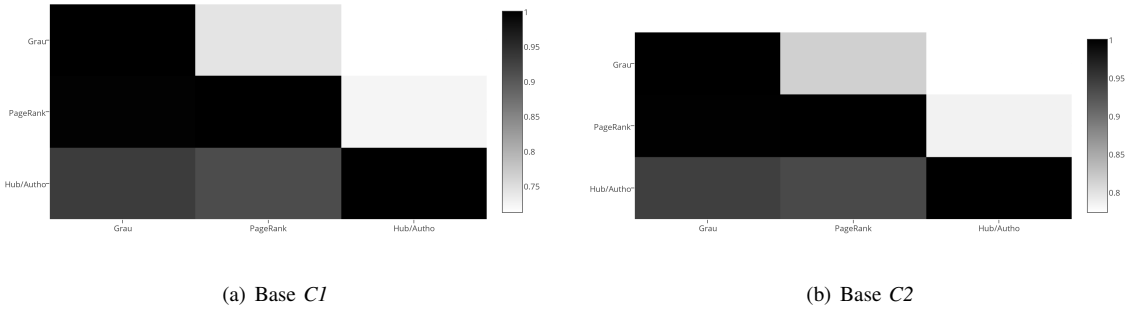


Figura 2. Gráficos de calor para bases de redes de colaboração científica

representa o número de pares discordantes; e n o número de pares da amostra.

Neste trabalho, ambos coeficientes são utilizados por possuírem diferentes estruturas de cálculo. Desta forma, é possível obter uma maior precisão na análise do comportamento das métricas a serem comparadas.

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (2)$$

Pela estrutura das Equações (1) e (2), é simples perceber que o coeficiente de Kendall-tau é mais sensível a variações nas posições dos elementos do que o coeficiente de Spearman. O coeficiente de Spearman leva em conta o tamanho da variação da posição entre um mesmo elemento em diferentes *rankings*, enquanto o coeficiente de Kendall-tau não leva em conta a distância entre um elemento em duas diferentes métricas - ele

simplesmente considera um elemento como concordante ou discordante.

IV. RESULTADOS E DISCUSSÕES

Estão representados nas Figuras 1, 2 e 3 os gráficos de calor dos coeficientes de correlação para cada uma das bases estudadas. Cada gráfico representa o valor da correlação entre duas diferentes métricas em uma mesma base. A área triangular superior de cada diagrama representa os resultados referentes ao coeficiente de Kendall-tau. A área triangular inferior representa os valores obtidos através do coeficiente de Spearman.

Um estudo mais detalhado foi conduzido para duas bases de colaboração científica. Estas bases foram particionadas em comunidades utilizando-se o algoritmo proposto em [25]. Cada comunidade foi avaliada de acordo com as mesmas três métricas. Os resultados para esta avaliação estão representados na Figura 4.

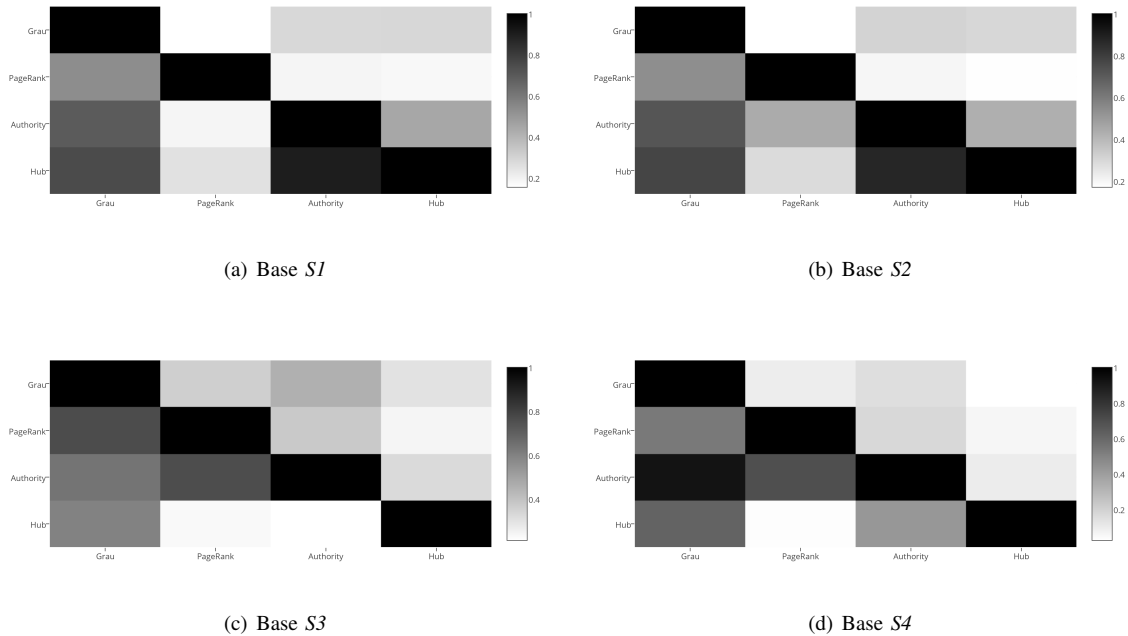


Figura 3. Gráficos de calor para bases de redes sociais

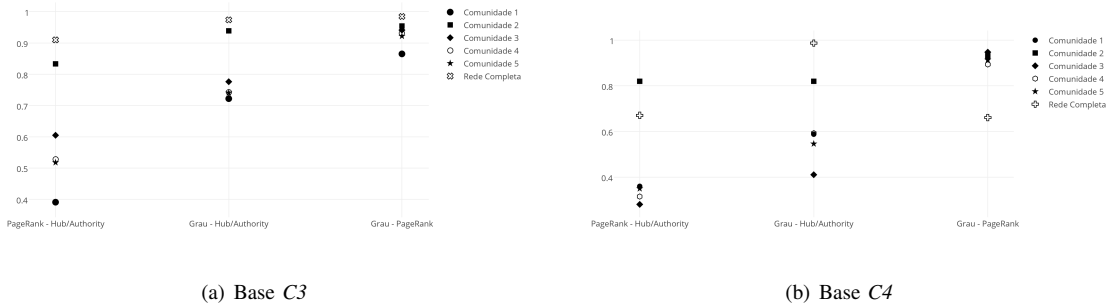


Figura 4. Avaliação de redes de colaboração científica e suas cinco maiores comunidades

As Tabelas IV e V dispõe os dados relativos as análises das cinco maiores comunidades das redes $C3$ e $C4$. As comunidades estão ordenadas em ordem decrescente, da maior para a menor, conforme relatado na Coluna 1.

Base	Número de nós	Número de Arestas	Assortatividade
Comunidade 1	1,492	14,714	0.467634588636
Comunidade 2	1,492	38,360	-0.023267131652
Comunidade 3	1,392	13,038	0.140122813972
Comunidade 4	1,256	13,943	0.221623231319
Comunidade 5	1,224	13,571	0.314617094023

Tabela IV. CARACTERÍSTICAS DAS COMUNIDADES DA REDE DE COLABORAÇÃO $C3$

É notório nas Figuras 1, 2 e 3 que o coeficiente de Kendall-tau realmente obtém valores de correlação menores que a correlação obtida pelo coeficiente de Spearman. A correlação entre a mesma métrica utilizando ambos os coeficientes é máxima, pois elas comparam duas listas ordenadas de vértices iguais. As quatro bases de redes de conexão P2P, como repre-

Base	Número de nós	Número de Arestas	Assortatividade
Comunidade 1	1,248	7,152	0.368494238681
Comunidade 2	1,141	18,092	0.355138874674
Comunidade 3	1,095	5,554	0.23398820513
Comunidade 4	783	5,784	0.50191580333
Comunidade 5	675	4,330	0.610672296924

Tabela V. CARACTERÍSTICAS DAS COMUNIDADES DA REDE DE COLABORAÇÃO $C4$

sentado na Figura 1, apresentam correlações muito próximas entre elas, demonstrando que a estrutura da rede não se modifica com o passar do tempo. As métricas *Grau* e *Hub*, ou *Grau* e *Hub/Authority* demonstraram obter a maior correlação quando analisamos ambas as três bases, obtendo maiores valores para o coeficiente de Spearman, sendo que a correlação entre *Grau* e *PageRank* também obteve altos valores, principalmente para as bases de colaboração científica, obtendo um valor de correlação próximo a 1. A menor correlação foi obtida entre as métricas *Hub* e *PageRank*. A métrica *PageRank* se assemelha

a *Authority*, pois leva em conta a quantidade e qualidade dos vértices que apontam para um determinado nó da rede.

Conhecer quais métricas possuem alta correlação pode levar à inferências sobre a escolha da métrica mais adequada para determinadas aplicações, se para redes de um determinado tipo a correlação de uma medida qualquer possui alta correlação com o grau, usar o grau é mais vantajoso, pois a obtenção da medida é direta. Para um motor de busca na Web, por exemplo, não é vantajoso a combinação de métricas que possuem correlação parecida, visto o fato de que a escolha de outras métricas com informações distintas, como *Hub* e *Pagerank* é mais adequado, por possuírem uma menor correlação para a grande maioria dos testes realizados, trazem informações diferentes sobre os vértices.

Outra conclusão importante da pesquisa é que, visto o fato de Grau e *PageRank* e Grau e *Hub* possuírem correlações altas, a utilização da métrica Grau pode ser mais interessante, visto o fato da sua simplicidade de cálculo, realizada simplesmente pela soma de uma linha na matriz de adjacência.

De acordo com a Figura 4, pode-se notar que existem grandes diferenças entre diferentes comunidades de uma rede. Mais além, comunidades de uma rede não possuem o mesmo coeficiente de correlação da rede. A correlação entre as métricas *PageRank* e *Hub/Authority* apresentou as maiores discrepâncias entre cada comunidade e a totalidade da rede. A correlação que apresentou a menor discrepância entre cada comunidade e entre as comunidades e a rede completa foi a correlação entre Grau e *PageRank*. Isto pode ser explicado devido a grande diferença entre os valores de assortatividade de cada comunidade, como demonstrado nas Tabelas IV e V.

V. CONCLUSÃO

Neste trabalho foram realizadas análises de correlação entre diferentes métricas em três tipos de bases de redes complexas diferentes. Resultados demonstraram que existe uma grande correlação entre as métricas Grau e *Hub*, classificando os vértices em listas ordenadas muito semelhantes entre si. Os vértices destas duas listas apresentam pequenas diferenças de posto, o que colabora para o baixo coeficiente de correlação utilizando o coeficiente de Kendall-tau.

Ao avaliarmos diferentes comunidades de uma mesma rede, a mesma relação é constatada, sendo que a correlação entre as métricas Grau e *Hub/Authority* apresenta menor discrepância em relação as correlações das outras métricas.

AGRADECIMENTOS

Os autores agradecem às agências de fomento: Capes, CNPq e FAPEMIG pelo apoio financeiro.

REFERÊNCIAS

- [1] W. W. Powell, D. R. White, K. W. Koput, and J. Owen-Smith, "Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences," *American journal of sociology*, vol. 110, no. 4, pp. 1132–1205, 2005.
- [2] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [3] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

- [4] A. Lancichinetti, M. Kivela, J. Saramaki, and S. Fortunato, "Characterizing the community structure of complex networks," *PLoS one*, vol. 5, no. 8, 2010.
- [5] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the north american power grid," *Physical review E*, vol. 69, no. 2, p. 025103, 2004.
- [6] A. Arenas, L. Danon, A. Diaz-Guilera, P. M. Gleiser, and R. Guimera, "Community analysis in social networks," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 373–380, 2004.
- [7] V. d. F. Vieira, C. R. Xavier, N. F. F. Ebecken, and A. G. Evsukoff, "Performance evaluation of modularity based community detection algorithms in large scale networks," *Mathematical Problems in Engineering*, vol. 2014, 2014.
- [8] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.
- [9] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web." 1999.
- [10] J. M. Kleinberg, "Hubs, authorities, and communities," *ACM Computing Surveys*, vol. 31, no. 4, p. 5, 1999.
- [11] —, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.
- [12] M. E. Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.
- [13] P. Erdős and A. Rényi, "On the Evolution of Random Graphs," *Publ. Math. Inst. Hungar. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
- [14] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.
- [15] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.
- [16] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hofer, Z. Nikoloski, and D. Wagner, "Maximizing modularity is hard," *arXiv preprint physics/0608255*, 2006.
- [17] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol. 70, no. 6, p. 066111, 2004.
- [18] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.
- [19] T. P. Leal, A. C. Goncalves, V. da F. Vieira, and C. R. Xavier, "Decode-differential evolution algorithm for community detection," in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 4635–4640.
- [20] G. Agarwal and D. Kempe, "Modularity-maximizing graph communities via mathematical programming," *The European Physical Journal B*, vol. 66, no. 3, pp. 409–418, 2008.
- [21] G. Csárdi and T. Nepusz, "The igraph software package for complex network research," *InterJournal Complex Systems*, vol. 1695, p. 1695, 2006. [Online]. Available: <http://igraph.sf.net>
- [22] J. Leskovec. (2015, Abr) Snap: Stanford network analysis platform. [Online]. Available: <http://snap.stanford.edu/snap/>
- [23] C. Spearman, "Correlation calculated from faulty data," *British Journal of Psychology*, vol. 3, no. 3, pp. 271–295, 1910.
- [24] A. Ghent, "Kendall's 'tau' coefficient as an index of similarity in comparisons of plant or animal communities," *The Canadian Entomologist*, vol. 95, no. 06, pp. 568–575, 1963.
- [25] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," no. 10, 2008.