

ANALYSIS OF COAUTHORSHIP NETWORK OF THE IBERIAN LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING

T. M. R. Dias¹, P. M. Dias¹, G. F. Moita²

¹Instituto de Ensino Superior e Pesquisa – INESP, Fundação Educacional de Divinópolis – FUNEDI/UEMG

²Centro federal de Educação Tecnológica de Minas Gerais – CEFET-MG

Abstract. *The social network has a few years, has been studied by different areas of knowledge, with different approaches and interests. In this work the main objective is to conduct a study on collaborative networks, which are a type of social network in which the relationships featuring some kind of professional collaboration among individuals. The goal is to conduct a study on the production network's scientific Iberian-Latin-American Congress on Computational Methods in Engineering with emphasis on the intensity of collaboration between researchers and to identify characteristics of the network as groups or individuals who are influential or have a high degree of importance network.*

Keywords: *Complex Networks, Management of Web data, intensity of collaboration.*

Resumo: *As rede sociais já a alguns anos, tem sido objeto de estudos por diversas áreas do conhecimento, com diferentes enfoques e interesses. Neste trabalho o objetivo principal é realizar um estudo sobre redes de colaboração, que são um tipo de rede social nas quais os relacionamentos caracterizam algum tipo de colaboração profissional entre indivíduos. O objetivo é realizar um estudo em uma rede de produção científica com ênfase na intensidade de colaboração entre pesquisadores e identificar características da rede como grupos ou indivíduos que sejam influentes ou que possuem elevado grau de importância na rede.*

Palavras Chave: *Redes Complexas, Gerência de dados na Web, Intensidade de Colaboração.*

1. INTRODUÇÃO

Grafos, ou redes, são poderosas ferramentas de abstrações que permitem codificar relacionamentos entre pares de objetos, nos quais vértices representam os objetos e arestas os relacionamentos. Em alguns casos os vértices e as arestas correspondem a objetos físicos do mundo real, em outros, os vértices são objetos reais enquanto as arestas correspondem a relacionamentos intangíveis, e ainda existem casos em que vértices e arestas são puras abstrações [1]. Em redes de transporte, por exemplo, o mapa de rotas utilizado por uma transportadora aérea naturalmente forma um grafo, onde os vértices são os aeroportos, e existe uma aresta entre dois vértices se há um voo direto entre dois aeroportos. Já em redes de comunicação, um conjunto de computadores conectados através de uma rede de comunicação pode ser modelado como um grafo, onde cada vértice representa um computador e arestas representam conexões físicas entre eles [1].

Dentre os vários tipos de redes, existem as redes sociais. Uma rede social é um conjunto de pessoas ou grupos que possuem algum tipo de relacionamento entre si [2].

Em [3], é descrito que relacionamentos entre pessoas podem ser de amizade, de parentesco ou de colaboração (por exemplo, coautores em um artigo). Em uma rede social de amizade, o relacionamento entre duas pessoas pode representar uma amizade entre elas. Em uma rede de parentesco, relacionamentos entre pessoas podem indicar que as duas pessoas pertencem à mesma família.

Uma rede de colaboração científica é uma rede onde os vértices são os autores de artigos científicos, e existe uma aresta entre dois autores se eles publicaram juntos, ou seja, colaboraram na produção de um artigo científico [4].

Para [3], redes de colaboração são diferentes de redes de citação, nas quais os nós são documentos e as arestas existem se uma publicação citou a outra. A intensidade dos relacionamentos entre os pesquisadores pode ser medida pelo número total de publicações em conjunto, por exemplo, adicionando peso 1 (um) à aresta para cada publicação feita por um par de autores. Assim, quanto mais publicações estes dois autores tiverem em conjunto, maior será a intensidade do relacionamento, ou peso da aresta.

Ao analisar uma rede de colaboração pode-se descobrir muitas propriedades topológicas da rede, como o número de autores, o número de publicações, o número de colaboradores por artigo, a probabilidade de dois autores terem um colaborador em comum, o menor caminho entre os dois autores mais distantes da rede e o número de componentes conexas. Também é possível identificar outras características importantes que tornam possível o ranqueamento de pesquisadores de acordo com sua importância para um grupo de pesquisa, país ou mundo, ou identificar quais grupos de indivíduos de uma rede são mais importantes [2] [4] [5].

Para este trabalho, a motivação principal é utilizar uma rede de colaboração e baseado na intensidade de relacionamento entre indivíduos desta definir importância de indivíduos, de relações ou conjunto de indivíduos. Desta forma, se torna possível realizar o ranqueamento de indivíduos dentro da rede de colaboração estudada.

2. TRABALHOS CORRELATOS

Redes de colaboração começaram a ser estudadas na Espanha e nos Estados Unidos na tentativa de formar relações de cooperação científica em formato de rede, a partir de indivíduos, grupos e instituições, nacionalmente ou internacionalmente [6].

Uma das redes de colaboração antigas, que ainda hoje é referência na área acadêmica, é a rede de colaboração do grande matemático húngaro Paul Erdős. Através dela, obtém-se o número de Erdős de cada pesquisador [7]. O número de Erdős representa a distância de colaboração entre uma pessoa e Paul Erdős, medida pela autoria de trabalhos acadêmicos.

Para ser atribuído um número de Erdős, um co-autor deve escrever um documento acadêmico com um autor que possua um número finito de Erdős. Paul Erdős é a única pessoa que tem um número de Erdős igual a zero. Para qualquer outro autor, se o menor número de Erdős de todos os seus colaboradores é k , então seu número de Erdős é $k + 1$.

De acordo com [7], Erdős escreveu mais de 1.416 artigos científicos, principalmente em colaboração. Ele tinha 504 colaboradores diretos. Estas são as pessoas com número de Erdős igual 1. Os autores que têm colaborado com eles (mas não com o próprio Erdős) têm um número de Erdős 2 (6.593 pessoas), aqueles que têm colaborado com as pessoas que têm um número de Erdős 2 (mas não com Erdős ou com qualquer um que possua um número de Erdős 1) têm um número de Erdős igual a 3 (33.605 autores), e assim por diante. Uma pessoa que não tem um caminho até Erdős na rede de colaboração tem um número de Erdős infinito.

Outra rede que utiliza a mesma ideia é a The Oracle of Bacon [8], que determina a distância de um ator qualquer de filmes até Kevin Bacon, sendo que o número de Bacon é 1 se o ator contracenou com Bacon em algum filme.

Em [9], Newman, utilizando redes com pesquisadores das áreas de biologia, física e matemática e procurando responder à uma variedade de questões sobre os padrões de colaboração, encontrou vários resultados através dos estudos dessas redes. Dentre eles, constatou que o número de colaboradores na rede de pesquisadores da área de biologia é muito maior que na de matemática devido ao modo de pesquisa (biologia trabalha com experimentos em laboratórios com muitas pessoas, e a matemática é mais teórica, trabalhando poucas pessoas em uma pesquisa). Concluiu, também, que nos últimos anos tem crescido o número de colaborações entre os matemáticos devido às mudanças das organizações sociais na comunidade matemática, ao surgimento de melhores sistemas de comunicação, e às possíveis mudanças nos tipos de problemas estudados e abordagens utilizadas.

Em [10], os autores analisaram a produção científica em três regiões diferentes do mundo, Brasil, América do Norte e Europa, por meio de redes de colaboração obtidas a partir de uma base de dados de publicações em Ciência da Computação, a DBLP. Os resultados obtidos por diferentes métricas indicam que o processo de produção do conhecimento tem mudado diferentemente em cada região. A pesquisa é cada vez mais feita em colaboração nas diferentes sub-áreas da Ciência da Computação. O tamanho da componente conexa gigante indica a existência de grupos de colaboração isolados na rede Europeia, ao contrário do grau de conectividade encontrado no Brasil e na América do Norte. Também foi analisada a evolução temporal das redes sociais que representam as três regiões. O número de autores por artigo aumentou em um período de 12 anos. Observou-se que o número de colaborações entre

os autores cresce mais rápido que o número de autores.

Em [4], Newman abordou métricas para medir a intensidade do relacionamento nas redes de colaboração científica. Esta intensidade é representada através de peso nas arestas da rede de colaboração. Primeiramente mostrou-se uma métrica simples que consiste em adicionar peso 1 a uma aresta para cada artigo que um par de autores possui em conjunto. Ou seja, o peso da aresta corresponde ao número de artigos que dois autores escreveram juntos. Um exemplo desta métrica está ilustrado na Figura 1, onde existem três artigos escritos pelos autores A, B, C, D e E, cada um com diferentes colaborações.

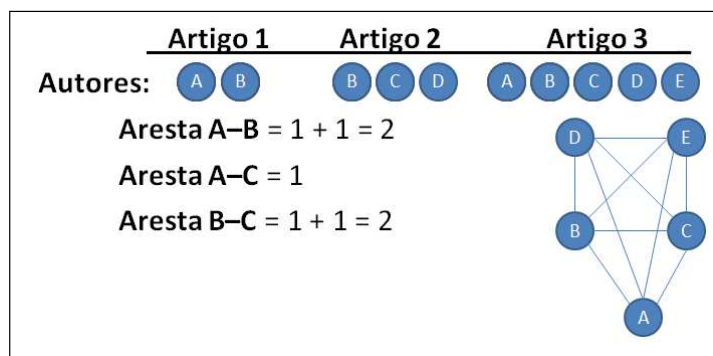


Figura 1: Exemplo do cálculo dos pesos das arestas utilizando uma métrica simples.

De acordo com o exemplo da Figura 1, para calcular o peso da aresta A-B, adiciona-se o peso 1 para cada artigo que eles escreveram juntos. Como eles colaboraram nos artigos 1 e 2, o peso da aresta A-B é 2. Os autores A e C escreveram juntos somente o artigo 2, logo o peso da sua aresta é 1. Já os autores B e C, publicaram juntos os artigos 2 e 3, então o peso da aresta B-C é 2. É importante observar que mesmo o artigo 2 tendo mais autores do que o artigo 3, os pesos das colaborações induzidas por cada artigo tem a mesma importância, gerando os mesmos valores de peso, 1 para cada colaboração.

No mesmo trabalho em questão, Newman introduziu uma nova métrica para medir a intensidade do relacionamento em redes de colaboração científicas, daqui para frente chamada de Métrica de Newman. Ela funciona da seguinte maneira: cada artigo colaborado por um conjunto de autores adiciona $1/n-1$ à intensidade da colaboração, ou seja, ao peso da aresta, onde n é o número de autores do artigo.

Na Figura 2, para calcular o peso da aresta A-B utilizando o método de Newman, adiciona-se $1/n-1$ ao peso por cada artigo que os autores A e B escreveram juntos. Eles escreveram sozinhos o artigo 1, logo o peso correspondente à ele é $1/2-1 = 1$. O artigo 2 foi escrito por 5 autores, então o peso correspondente à ele é $1/5-1 = 0,25$. Logo, o peso desta aresta A-B é 1,25.

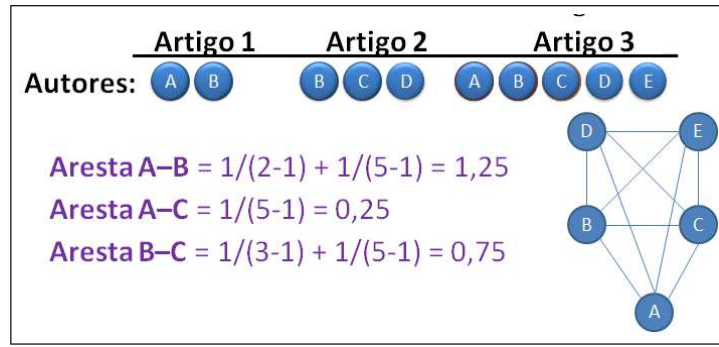


Figura 2: Cálculo dos pesos das arestas utilizando o Método de Newman

É interessante observar os cálculos dos pesos das arestas A-B e B-C, que na métrica simples tiveram o mesmo valor. Já na métrica de Newman, o valor dos pesos foi diferente, pois receberam a influência do número de autores que colaboraram nos artigos. Newman descreve seu método como sendo o quanto os autores se conhecem em cada artigo que eles trabalham juntos. A ideia é que cada artigo tem intensidade constante, que é dividida igualmente entre os co-autores.

Pode-se utilizar a Equação 1 para calcular o peso da aresta w_{ij} entre os autores i e j :

$$w_{ij} = \sum_k \frac{\delta_i^k \delta_j^k}{n_k - 1} \quad (1)$$

onde

$$\delta_i^k = \begin{cases} 1 & \text{se o autor } i \text{ é co-autor do artigo } k; \\ 0 & \text{caso o contrário} \end{cases}$$

e n_k é o número de co-autores do artigo k . Nota-se que artigos com apenas um autor não adiciona peso às arestas.

A Figura 3 mostra a rede de colaboração correspondente aos dados contidos nas figuras 1 e 2. Pode-se observar que somando os pesos das arestas que incidem em cada vértice resulta no número de artigos em que o autor é co-autor, ou seja, no número de artigos que o autor possui com algum outro pesquisador. De agora em diante, essa soma será chamada de peso do vértice.

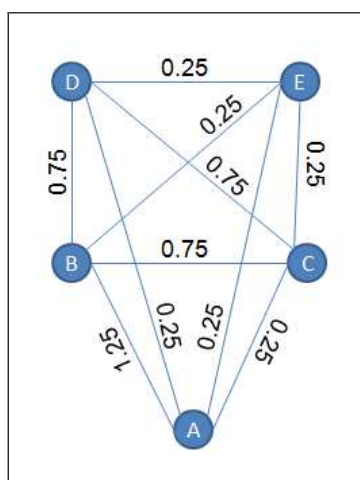


Figura 3: Grafo de colaboração dos autores A, B, C, D e E.

Para o presente trabalho, a métrica utilizada para análise da intensidade de relacionamento da colaboração entre os pesquisadores foi a Métrica de Newman. Com a aplicação da métrica é possível avaliar a intensidade em que um pesquisador está trabalhando em colaboração com outros pesquisadores e qual o impacto das publicações estarem sendo realizadas com poucos ou muitos pesquisadores.

Além disto, a aplicação da métrica, permite que seja comparado a utilização da métrica com a técnica utilizada anteriormente por Newman (método simples), onde cada colaboração recebe um valor constante entre cada par de autor. Com isto, reforça a teoria de avaliar o quanto os autores se conhecem em cada artigo que eles trabalham juntos.

Como o intuito principal é de se avaliar a intensidade de colaboração, artigos que possuem somente um autor em sua publicação não se enquadra a métrica de Newman, já que não se aplica a fórmula para o cálculo da métrica e nem faz parte do escopo do trabalho que deseja avaliar a colaboração.

3. CARACTERIZAÇÃO DA REDE DE COLABORAÇÃO

A rede de colaboração para o presente trabalho, possui dados de publicação de 11 edições (1999 a 2011 com ausência de 2002 e 2009) do CILAMCE - IBERIAN LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING, que está em sua 33a edição, edições estas realizadas no período de 1977 a 2011.

O CILAMCEA tem como objetivo criar um fórum no qual engenheiros, pesquisadores e estudantes podem trocar idéias e informações sobre os métodos computacionais, sistemas disponíveis e melhorias na tecnologia de computação para resolver vários problemas complexos de engenharia práticos e teóricos.

Em todas as suas edições, o CILAMCE tem desempenhado um papel importante na divulgação das mais recentes aplicações e desenvolvimentos computacionais em engenharia entre os profissionais, pesquisadores e estudantes da comunidade ibero latino-americana.

Outras formas possíveis de visualizar a rede é com a aplicação de filtros que possibilitam apresentar a rede com os pesquisadores com o maior número de publicações ou mesmo por grupos de pesquisadores de uma mesma instituição.

Para o presente trabalho, foi aplicado o método de Newman descrito na seção 2 a fim de verificar a eficiência do método para calcular a intensidade de colaborações e verificação do estado de colaboração do CILAMCE nas 11 edições analisadas.

4. APLICAÇÃO DA MÉTRICA DE NEWMAN

É importante ressaltar que várias métricas poderiam ser aplicadas para estudo da rede de colaboração. No entanto, como o enfoque do presente trabalho é verificar a intensidade de colaboração entre pares de autores, optou-se pela métrica de Newman por apresentar bons resultados em trabalhos correlatos.

Outro ponto a se destacar é que como a métrica tem como enfoque mediar a intensidade de colaboração, artigos que foram publicados tendo somente um autor, são irrelevantes para a métrica já que o cálculo é realizado através de pares de pesquisadores e o peso para publicação com apenas um autor é 0 (zero) conforme descrito na seção 2.

Para realizar o cálculo dos pesos em cada par de vértices, foi desenvolvido um algoritmo que tem como entrada a rede disponibilizada e considera o artigo publicado, ano de publicação, quantidade de autores e identificação de cada autor.

Em seu processamento são calculados quantos autores participam de cada artigo utilizando os dados acima descritos e faz o cálculo do peso de cada um dos artigos seguindo a métrica de Newman $1/n-1$, sendo n a quantidade de autores de cada um dos artigos. Como resultado do cálculo é identificado o peso que cada artigo contribui para cada um dos autores. Logo, a aresta que faz a ligação para cada par de autores deste artigo recebe como peso o valor calculado.

Caso seja identificado a colaboração entre 2 autores em mais de um artigo, o peso da arestas destes 2 autores é incrementado com o valor já existente da(s) colaborações anteriores. Tabela 1.

Tabela 1: Exemplo da Matriz de Intensidade de colaboração gerada pelo algoritmo implementado

	2145_	2390_	2432_	282_	2621_	2174_
011_	0,5					
0110_						

0112_						
0117_						
0118_						
012_	0,5					
075_		1	0,33	0,5	0,25	1

A rede de colaboração possui algumas características como a quantidade de 5.000 triângulos e um total de 210 componentes conectados que faz com que em rede de colaboração as ideias de um autor possa chegar aos demais através de um ou múltiplos saltos. No entanto, estes valores após a aplicação do método não são essenciais já que o enfoque está no peso das arestas de pares de autores e não em grupos de autores.

Sendo assim, o peso das arestas e vértices são o principal alvo de estudo. Inicialmente utilizando o peso 1 (um) para cada par de autores (arestas), o peso total de todas as arestas da rede era de 12.500, após a aplicação da métrica passou a ter um peso total de 4.310,00.

Apesar da média apresentada, um detalhe importante que pode ser verificado é que um percentual de 60,2 % dos integrantes da rede possuem somente uma publicação e a soma do peso de suas arestas tem valores bem inferiores a média. No entanto, um percentual bem inferior de pesquisadores possuem um valor alto em seus pesos que eleva a média e mostra o quão importantes estes são na intensidade de colaboração na rede do CILAMCE.

Percebe-se que independente do total de artigos publicados por um determinado pesquisador, a métrica destaca os pesquisadores que possuem uma quantidade elevada de artigos, mas que principalmente considera o número de participantes em cada um dos artigos, não favorecendo as publicações com elevado número de autores.

Na rede de colaboração do CILAMCE, verifica-se que o pesquisador com maior valor no peso de suas arestas, possuem um alto grau de colaboração com outros mas que principalmente a sua participação nos artigos, são geralmente com um número moderado de colaboradores, em média 3.

Artigos com elevado número de autores, contribuem muito pouco para mediar a importância da intensidade de colaboração já que a participação efetiva e uniforme em um grupo com uma quantidade alta de elementos não é trivial e a colaboração uniforme de todos nem sempre é garantida.

O cálculo dos valores com a métrica de Newman se mostra coesa e efetiva quando o intuito principal é apresentar o grau de intensidade de colaboração e análise de como esta colaboração acontece. É possível identificar que a quantidade de artigos publicados por determinado pesquisador não representa efetivamente aquele que tem maior intensidade de colabo-

ração como descrito anteriormente.

O ano com maior número de publicações ocorreu em 2007 contando com aproximadamente 630 publicações mas escrita por muitos autores o que justifica um número baixo de intensidade de colaborações. Este também foi o ano com o maior número de autores. Já no ano de 2000, foram publicados 236 artigos o que elevou a média deste ano se comparado aos anos anteriores.

É possível observar que com algumas exceções como os anos 1999 e 2000, todos os outros anos houve uma média constante de colaboração entre os autores que se aproxima de 3,3.

A partir de 2005 esta média superou 3,4 e está se mantendo nestes valores. Apesar de modesto crescimento, é possível observar que a média de autores por artigo publicados nos últimos anos do CILAMCE tem evoluído. Isto representa uma evolução na colaboração entre pesquisadores da área de métodos computacionais em engenharia.

5. CONCLUSÃO

Dentre os vários tipos de redes, existem as redes sociais que possuem um alto nível de importância por serem capazes de representar fatos e relacionamentos humanos. Vários autores defendem que as características topológicas das redes são capazes de refletir o comportamento de seus integrantes. As redes de colaboração científica como a rede do CILAMCE apresentada neste trabalho, vem sendo objeto de pesquisa de vários pesquisadores com o objetivo de entender não somente as características delas, mas também como criar novos serviços com base no entendimento da evolução e estrutura das mesmas.

Os relacionamentos nas redes de colaboração científica são caracterizados por possuírem diferentes intensidades. Para realizar o cálculo destas intensidades, existem diversas métricas como número de publicações ou quantidade de colaborações, cada qual o um foco em específico. Com a aplicação destas métricas é possível definir o quanto um participante ou grupo de participantes da rede é importante ou influente, realizando para isto uma pontuação seguindo critérios da métrica utilizada.

Neste trabalho foi utilizado a métrica de Newman[4], apresentada na seção 2, cujos resultados foram descritos anteriormente. Nesta métrica o enfoque principal está na intensidade de relacionamento entre os integrantes da rede, que foi aplicado na rede de 11 edições do CILAMCE (1999 a 2011 – exceto 2002 e 2009).

Ao analisar os resultados após a aplicação da rede, é possível identificar que a quantidade de publicações de um determinado autor não significa necessariamente a sua importância ou influencia quando se pretende mediar a intensidade de colaboração entre pares. Um artigo publicado com um número alto de colaboradores, não necessariamente significa que estes possuem uma mesma intensidade que um artigo publicado com um número menor de colaboradores.

É possível ainda identificar que existem pesquisadores que são extremamente influentes em suas publicações e que são autores extremamente importante para o grupo a que per-

tence e para a rede como um todo. Além disto, percebe-se que nos últimos anos o percentual de colaboração entre os autores do CILAMCE tem aumentado.

Novas características podem ser consideradas e agregadas a métrica a fim de se obter resultados mais efetivos como considerar a idade da interação (idade da publicação) quando se for definir a intensidade dos relacionamentos e ainda verificar como a propagação da informação ocorre, já que intuitivamente a informação se propaga mais rapidamente entre os relacionamentos de maior intensidade.

6. REFERÊNCIAS

- [1] KLEINBERG, J., TARDOS, É. Algorithm desing. Pearson Education, 2006. ISBN: 0-321-29535-8.
- [2] NEWMAN, M. E. J. The structure of scientific collaboration networks. In: Proc. Natl. Acad. Sci. USA, v. 98, pp. 404-409, jan. 2001.
- [3] FREIRE, V. P. M. Uma métrica para ranqueamento em redes de colaboração baseada em intensidade de relacionamento. Dissertação de Mestrado. COPPE. UFRJ. 2010.
- [4] NEWMAN, M. E. J. Who Is the Best Connected Scientist ? A Study of Scientific Coauthorship Networks. v. 650/2004, Lecture Notes in Physics, pp. 337-370, Springer Berlin / Heidelberg, ago. 2004. ISBN: 978-3-540-22354-2. doi: 10.1007/b98716.
- [5] WAGNER, C., LEYDESDORFF, L. Network structure, self-organization, and the growth of international collaboration in science, Research Policy, v. 34, n. 10, pp. 1608-1618, December 2005. ISSN: 00487333. Dispo- nível em: <<http://dx.doi.org/10.1016/j.respol.2005.08.002>>.
- [6] OLIVEIRA, E. F. T., DA SILVA SANTAREM, L. G., SEGUNDO, J. E. S. Análise das redes de colaboração científica através do estudo das co-autorias, nos cursos de pós-graduação do brasil no tema tratamento temático da informação, Nuevas perspectivas para la difusión y organización del conocimiento: actas del congreso / coord. por Nuria Lloret Romero, v. 2, pp. 986-1000, set. 2009.
- [7] GROSSMAN, J., ION, P. The Erdös Number Project. Disponível em: <<http://www.oakland.edu/enp/>>.
- [8] REYNOLDS, P. The Oracle of Bacon. Disponível em: <[http:// oracleofbacon.org](http://oracleofbacon.org)>.
- [9] NEWMAN, M. E. J. Coauthorship networks and patterns of scientific collaboration. In: Proc. Natl. Acad. Sci. USA, v. 101, pp. 5200-5205, abr. 2004. doi: 10.1073/pnas.0307545100.
- [10] MENEZES, G. V., ZIVIANI, N., LAENDER, A. H. F., etal. A geographical analysis of knowledge production in computer science. In: Proceedings of the 18th international conference on World wide web, pp. 1041-1050, 2009.