



XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)
ISSN 2177-3688

GT 4 – Gestão da Informação e do Conhecimento

Comunicação Oral

ANÁLISE DE REDES SOCIAIS NA COLABORAÇÃO CIENTÍFICA¹

SOCIAL NETWORKING ANALISING IN SCIENTIFIC COLLABORATION

Marcos L. Mucheroni, USP
mucheroni.marcosl@gmail.com

Decio Funaro, USP
deciofunaro@hotmail.com

Resumo: Analisa a colaboração científica e, se de fato, formamos redes segundo os procedimentos metodológicos, apresentando de modo sintético em gráficos e tabelas as diversas métricas de Análise de Redes Sociais (ARS). O objetivo é o de estabelecer um conjunto de métricas para esta análise pelo cálculo de indicadores como as centralidades, a hipótese do Mundo Pequeno de Travers e Milgram ou a Lei das Potências de Barabási. Para isto realizou-se alguns desenvolvimentos específicos, incluindo programas de computador, como fonte alimentadora da ferramenta Pajek associados a dados de bases indexadas na construção de métricas da provenientes da ARS. Definiu-se três áreas de estudos: educação, Sociologia e Ciência da Informação, as duas primeiras por se tratarem de importantes para o conjunto da sociedade e CI por se tratar da área em que o estudo foi feito. Estes estudos podem ser expandidos para as demais áreas da comunidade científica e dar continuidade a este trabalho.

Palavras-chave: Análise de Redes Sociais. Colaboração científica. Redes de cocitação.

Abstract: This paper analyzes the scientific collaboration and actually form networks according to the methodological procedures, presenting in summary form in graphs and tables the various metrics of Social Network Analysis (SNA) establishing a set of metrics for this analysis. For that were made some specific developments using Pajek tool, using indexed databases and building metrics ARS. Three areas of study were chosen: education, sociology and Information Science, the first two since they are important to society as a whole and CI because it is the area in which the study was done. These studies can be expanded to other areas of the scientific community and to continue this work.

Keywords: Social Networking Analysis. Scientific Collaboration. Co-citation networking.

¹ O conteúdo textual deste artigo, os nomes e e-mails foram extraídos dos metadados informados e são de total responsabilidade dos autores do trabalho.

1 INTRODUÇÃO

As comunidades científicas, desde os chamados "colégios invisíveis", teceram redes de influência e de relacionamentos paralelamente à produção científica. Embora haja referências a trabalhos de alguns autores sobre a análise teórica e empírica de resultados e sobre as coautorias, o tema ainda carece de estudos aprofundados sobre os tipos de redes que estas comunidades determinam, por quais caminhos foram desenvolvidas e quais informações sobre a comunidade pode-se obter pelo estudo desses relacionamentos.

Métricas bem estabelecidas sobre a "produtividade" de autores foram estabelecidas e seus resultados são importantes para analisar tanto a produtividade científica como tendências de pesquisa em determinadas áreas, o crescimento e a chamada explosão informacional (MEADOWS, 1972) tornou-se fundamental como auxílio a análise do fluxo informacional dentro de áreas diversas áreas, como a da Ciência da Informação.

Diversos estudos estabeleceram também as redes de cocitação, redes colaborativas como são chamados os relacionamentos em redes nas comunidades científicas. Destaca-se sobre o tema o trabalho publicado na revista *Informação & Informação*, de 2007, dedicado ao tema com um levantamento de 12 títulos de teses e dissertações na área feita por Tomaél (2007).

Porém o ambiente científico se faz uso de um modelo genérico de métricas de publicação sem verificar de fato como os diversos autores e trabalhos se relacionam ainda que sejam crescentes os estudos em redes colaborativas em odontologia (FUNARO, 2010) em ambientes digitais e redes colaborativas (POBLACIÓN; MUGNAINI; RAMOS, 2009) feito em colaboração com diferentes pesquisadores em CI e em conjunto com o Centro Latino-Americano e Caribe de Informação em Ciências da Saúde (BIREME).

Entre os trabalhos de cunho interdisciplinar destaca-se o trabalho de Marteleto e Stotz (2009) *Informação, saúde e redes sociais: diálogos de conhecimentos nas comunidades da Maré*, fruto de pesquisas compartilhadas entre pesquisadores de Ciências da Informação e da área da Saúde, estudando a construção compartilhada de saberes sobre as condições de vida.

O presente texto tem por objetivo verificar as efetivas colaborações acadêmicas nas áreas da Educação, da Sociologia e da Ciência da Informação através da metodologia da Análise de Redes Sociais com base nas coautorias. Conquanto apenas as três áreas citadas tenham sido abordadas, este trabalho mantém a perspectiva de futuras aplicações para os muitos segmentos do conhecimento, bem como a constatação da formação de redes colaborativas entre autores de áreas distintas, evidenciando indícios de transdisciplinaridade.

Muitos autores demonstram a eficácia do uso de ferramentas e métricas nas ciências

sociais, em estudos inter, multi ou trans disciplinar (WASSERMAN; FAUST, 1994, p. 10).

Para tratar os dados obtidos das bases indexadas específicas das três áreas e obter os resultados a serem analisados usou-se a metodologia da Análise de Redes Sociais.

2 REDES COM LEI DE POTÊNCIA E SEM-ESCALA COMO MÉTRICAS DE COAUTORIA

As redes sociais são estruturas sociais formadas por um grupo de indivíduos – chamados "nós" ou "vértices" – e suas relações de interdependência, tais como gostos, clube de coração, amizades, inimizades, relações financeiras, hierarquia ou credos. Caracterizam-se por valorizarem os elos informais e as relações em detrimento das estruturas hierárquicas (MARTELETO, 2001). Sob o ponto de vista de uma metodologia interdisciplinar, encontramos o trabalho de Marteleto e Tomaél (2005), enquanto estudos quantitativos e métricas podem ser encontrados em Wasserman e Faust (1994).

Os princípios discutidos a seguir fundamentam-se nas pesquisas relacionadas à estrutura da rede social, especialmente no que trata de valorar determinadas relações ou nós, inferindo ou, até mesmo, explicando o posicionamento de determinado nó como peça que dá à rede social a característica de um grafo conexo, ou seja, onde sempre existe um caminho que conecta quaisquer dois atores na rede.

Os estudos de Barabási (1999) fundamentam-se na estatística, embasando cientificamente suas conclusões e reforçando o seu distanciamento do empirismo especulativo. Em uma análise mais recente, Barabási (2003) constrói seu raciocínio em modelos estatísticos aplicados aos nós e suas ligações através de distribuições estatísticas, especialmente a exponencial e a Poisson.

Pensando nos aspectos humanos, as relações entre os indivíduos, por aspectos diversos, como a amizade, o parentesco, a profissão ou o clube de coração, formam redes sociais que, sob o enfoque da modelagem, não difere das relações entre os sítios da internet, das linhas de transmissão de energia, das linhas de transmissão de voz e dados ou de uma malha rodoviária.

As redes sem escala (*Networks without Scale*) (BARABÁSI; BONABEAU, 2003), por seu conceito, se assemelham às redes de alta escalabilidade (FU et al., 2014), ou seja, aquelas que podem crescer com facilidade, durante mais de 40 anos foram tratadas como sistemas complexos aleatórios.

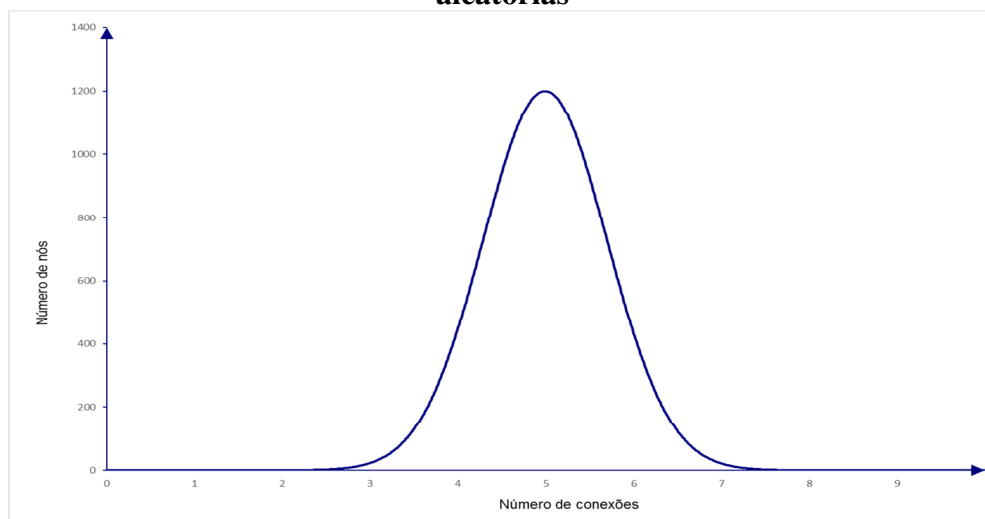
Em 1959, os matemáticos húngaros Paul Erdős e Alfréd Rényi empenhados em modelar as redes de comunicações e as ciências da vida, sugeriram que tais sistemas seguiam um modelo em que ligações entre os nós aconteciam de forma aleatória, fazendo uso e trazendo estudos

sobre a teoria dos grafos a novo destaque e com enfoque nas redes aleatórias (BARABÁSI; BONABEAU, 2003). Uma inferência inicial sobre as redes aleatórias é a de que são democráticas, independentemente da distribuição das ligações, nas quais a maioria dos nós, têm aproximadamente o mesmo número de conexões.

As ligações em uma rede aleatória seguem uma distribuição de Poisson (cujas forma lembra um sino), como na Figura 1 a seguir, evidenciando que são raros os nós contendo mais ou menos ligações do que a média.

As redes aleatórias são também chamadas exponenciais, já que a probabilidade que um nó esteja conectado a k outros nós é inversa e exponencialmente proporcional a k .

Figura 1 - Gráfico típico de uma distribuição de Poisson, característico das redes aleatórias



Fonte: Adaptação de Barabási e Bonabeau (2003).

Um importante estudo aparece na edição de maio de 2003 da *Scientific American*:

Os cientistas descobriram recentemente que vários sistemas complexos têm uma arquitetura subjacente regida por princípios de organização compartilhados. Esta visão tem implicações importantes para uma série de aplicações, desde o desenvolvimento de medicamentos até a segurança na internet (BARABÁSI; BONABEAU, 2003, p.50).

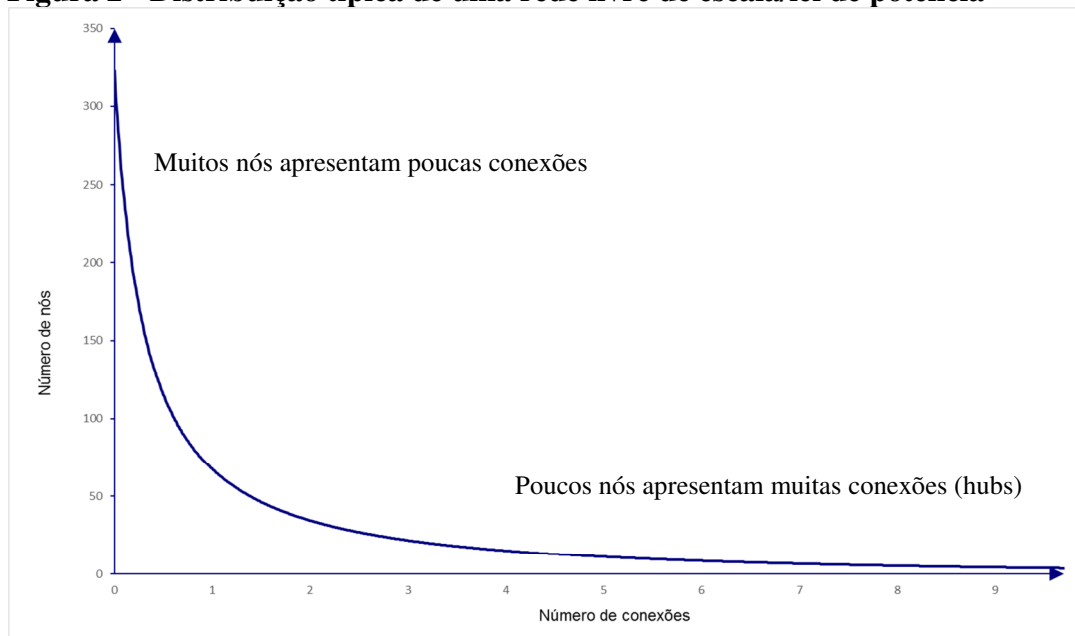
Dando continuidade à esta pesquisa Barabási e Bonabeau (2003), em parceria com Hawoong Jeong e Réka Albert da Universidade de Notre Dame, iniciaram em 1998 um estudo com o objetivo de mapear a Web, com a expectativa de se tratar de uma rede aleatória, já que os interesses pessoais em terem seus sítios referenciados por outros, à grande diversidade de interesses de todos e o número elevado de páginas por eles escolhidas induzia a acreditar as conexões seguiam um padrão aparentemente aleatório. Todavia, os resultados não confirmaram

as expectativas. Por via de um robô (programa), os sítios foram sendo visitados e coletadas suas conexões a outros, observando-se que páginas com grande número de conexões a elas essencialmente davam sustentação à Web como um todo. Das páginas visitadas, mais de 80% tinham menos do que quatro conexões, enquanto uma pequena parte (0,01%) dos nós apresentava mais que 1000 conexões e, em novo experimento, obteve-se um nó com mais de dois milhões de referências.

Calculando-se o número de páginas com exatamente k conexões, obteve-se a chamada "*power law*" (lei de potências), ou seja, a probabilidade de que qualquer nó estivesse conectado a outros k nós era proporcional a $1/k^n$. Como as conexões entrantes (n) eram aproximadamente iguais a 2, cada nó teria aproximadamente quatro vezes mais chances (22) de ter apenas a metade do número de ligações entrantes que outro nó. Isso mostra que a lei de potências difere em sua distribuição das redes aleatórias com suas distribuições em forma de sino (Poisson).

A Figura 2 mostra as redes aleatórias com suas distribuições em forma de sino (Poisson), com um gráfico típico de uma distribuição das leis de potências.

Figura 2 - Distribuição típica de uma rede livre de escala/lei de potência



Fonte: Adaptação de Barabási e Bonabeau (2003).

Barabási em parceria com Hawoong Jeong e Réka Albert (BARABÁSI, BONABEU, 2003), da Universidade de Notre Dame, iniciaram em 1998 um estudo com o objetivo de mapear a Web, com a expectativa de se tratar de uma rede aleatória, já que os interesses pessoais em terem seus sítios referenciados por outros, à grande diversidade de interesses de todos e o número elevado de páginas por eles escolhidas induzia a acreditar as conexões seguiam um padrão aparentemente aleatório. Todavia, os resultados não confirmaram as

expectativas. Por via de um robô (programa), os sítios foram sendo visitados e coletadas suas conexões a outros, observando-se que páginas com grande número de conexões em toda Web.

Das páginas visitadas, mais de 80% tinham menos do que quatro conexões, enquanto uma pequena parte (0,01%) dos nós apresentava mais que 1000 conexões e, em novo experimento, obteve-se um nó com mais de dois milhões de referências.

Nas redes aleatórias os *hubs* são inexistentes, e é a sua presença a característica notável das redes com lei de potência, aquelas em que alguns nós concentram grande número de ligações, fenômeno este batizado de *scale-free* (livre-de-escala) por admitir a existência e a possibilidade de crescimento indefinidamente de alguns poucos nós.

As redes livres de escala existem em abundância. A *World Wide Web* exemplifica as características das redes com lei de potência através dos *hiperlinks* (endereços eletrônicos) referenciando e sendo referenciados pelas páginas. Pesquisas feitas pelos irmãos Michalis Faloutsos da Universidade da Califórnia em Riverside, Petros Faloutsos da Universidade de Toronto e Christos Faloutsos da Universidade Carnegie Mellon (BARABÁSI; BONABEAU, 2003) estudaram a Web sob aspectos físicos, analisando a topologia às quais os roteadores se conectavam através das linhas de dados, e tal estrutura também seguia o modelo livre de escala.

Algumas pesquisas apontaram na direção de que algumas redes sociais também seguem o modelo livre-de-escala (BARABÁSI; BONABEAU, 2003). Como exemplo, pesquisadores da Universidade de Boston e da Universidade de Estocolmo mostraram que a rede de relacionamentos sexuais entre suecos segue a lei de potência: enquanto a maioria dos indivíduos têm experiências sexuais com poucos parceiros durante sua vida, alguns poucos têm relações com um grande número de indivíduos. Outro exemplo foi exposto na pesquisa feita por Stefan Bornholdt da Universidade de Kiel na Alemanha, mostrando que a rede social formada pelos endereços de e-mail também segue o padrão livre-de-escala.

Assim também segue a lei de potência, a rede de citações em publicações científicas, como mostrado por Sidney Redner da Universidade de Boston (REDNER, 2005).

Estudando as redes colaborativas em diversas disciplinas, incluindo médicos e cientistas da computação, o pesquisador Mark Newman da Universidade do Michigan em Ann Arbor concluiu que essas também seguem o padrão *livre-de-escala*, confirmando pesquisas semelhantes feitas por Barabási envolvendo matemáticos e neurologistas, as quais apontaram para Erdős como coautor com mais de 1400 trabalhos envolvendo aproximadamente 500 outros pesquisadores (NEWMAN, 2004). As redes livres-de-escala também ocorrem no ambiente de negócios (NEWMAN, 2004). Uma pesquisa realizada por Walter W. Powell da Universidade de Stanford, Douglas R. White da Universidade da Califórnia em Irvine, Kenneth W. Koput da

Universidade do Arizona e Jason-Owen Smith da Universidade do Michigan envolvendo empresas de biotecnologia americanas mostrou a existência de *hubs*.

3 MÉTRICAS DE REDES SOCIAIS EM COMUNIDADES CIENTÍFICAS

Pesquisadores italianos realizaram uma análise profunda naquele segmento, observando que, através de dados obtidos da base de dados farmacológicos da Universidade de Siena, composto por não menos que 20.000 itens contratuais entre 7200 firmas, os estudos citados de Powell, White, Koput e Smith indicavam que a existência daqueles *hubs* levava à conclusão de que se tratava de uma rede livre-de-escala.

Como definida nos parágrafos anteriores, uma rede aleatória tem como principal característica a inexistência de *hubs* (NEWMAN, 2004). Ao examinar mais profundamente as pesquisas de Erdős e Rényi, observam-se duas razões para que nas redes democráticas não ocorram *hubs*.

A primeira está no fato de que nas pesquisas de Erdős e Rényi os nós estavam previamente definidos antes das ligações entre eles, contrariando a dinâmica de páginas da internet ou outras redes, com inclusões (exclusões em menor número) de páginas com grande frequência.

Diversas métricas são utilizadas para análise em Redes Sociais, entre estas são utilizadas neste trabalho temos para uma rede de n nós, a rede completa $A(n)$ é expressa como tendo o seguinte número de arestas:

$$A(n) = C(n, 2) = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}$$

O grau de densidade $d(V,A)$ para a rede pode ser definido pela densidade de uma rede com n vértices e a arestas:

$$d(V, A) = \frac{a}{A(n)} = \frac{2a}{n(n-1)}$$

A centralidade de grau pode ser classificada entre grau entrante (*indegree*), relacionada àquelas ligações que têm suas direções voltadas para o nó e, em contrapartida, o grau saínte (*outdegree*), temos a centralidade de grau $C_D(v)$ para um nó v como:

$$C_D(v) = grau(v) \text{ ou normalizado } C_D(v) = \frac{grau(v)}{n}$$

A centralidade de intermediação (*betweenness*) razão entre o número de geodésicas que contêm x_i e o número total de geodésicas entre x_j e x_k , chamada *betweenness* ou centralidade do nó x_i com relação aos vértices x_j e x_k é dada por:

$$b_{jk}(x_i) = \frac{g_{jk}(x_i)}{g_{jk}}$$

A centralidade de proximidade de um nó v como a média entre as distâncias geodésicas entre v e todos os outros N nós:

$$C_c(v) = \frac{1}{\sum_{j=1}^N d(v,j)}$$

Foram desenvolvidas ainda métricas a partir dos estudos dos trabalhos de Barabási e Bonabeau (2003) para redes livre-de-escala (*scale-free*).

A coleta de dados das bases para as áreas de Sociologia, Educação e Ciência da Informação propiciaram a construção das três redes colaborativas e a consequente criação dos arquivos de dados de entrada do software PAJEK (arquivos com a extensão ".net").

O PAJEK (NOOY; MRVAR; BATAGELJ, 2005), aplicação desenvolvida para a análise de redes sociais, permite que sejam calculados grande diversidade de índices para as redes, além de possibilitar a visualização de clusters e gráficos.

Os resultados consideraram os principais aspectos de uma rede colaborativa, como centralidade e proximidade, fornecidas pela aplicação. A geração dos arquivos de dados do PAJEK foi feita por aplicação desenvolvida pelo autor em VBScript, focada exclusivamente na criação dos arquivos ".net" e na criação de arquivos ".csv" (Comma-Separated Values), acessíveis pelo Microsoft Excel, com a finalidade de obter informações como a centralidade de grau por vértice (autor) e a determinação daqueles com maior número de ligações.

A Tabela 1, obtida através do uso do *Excel*, mostra as intersecções entre as três áreas estudadas, evidenciando a possibilidade de que um autor pertença a uma determinada área e publique com outro autor de outra, tornando-se assim um vértice que conecta as duas redes.

Tabela 1 - Número de autores em comum para as áreas (intersecções)

	Número de autores em comum
CI x Educação	108
CI x Sociologia	67
Educação x Sociologia	156
CI x Educação x Sociologia	4

Fonte: Os autores (2015)

Note-se que nada se pode afirmar sobre as redes pela simples análise da tabela, considerando não haver relação entre as intersecções e a dinâmica das redes, já que todas as três podem ter como característica a representação por um grafo desconexo, o que ficará claro mais à frente.

3.1 ANÁLISE NAS ÁREAS PELA CONEXÃO ENTRE AUTORES

Uma observação importante está no fato de que os nomes dos autores apresentam variações internas às bases. Não há padronização no cadastramento, ocorrendo para um mesmo autor o cadastramento de iniciais simples ou seguidas por um ponto ou, em casos extremos, adicionado uma alcunha entre nome e sobrenome. Este fato nos leva a acreditar que há variação nas quantidades de autores, um mesmo sendo considerado diversas vezes e, sendo assim, podendo alterar os dados e, conseqüentemente, as relações e as intersecções.

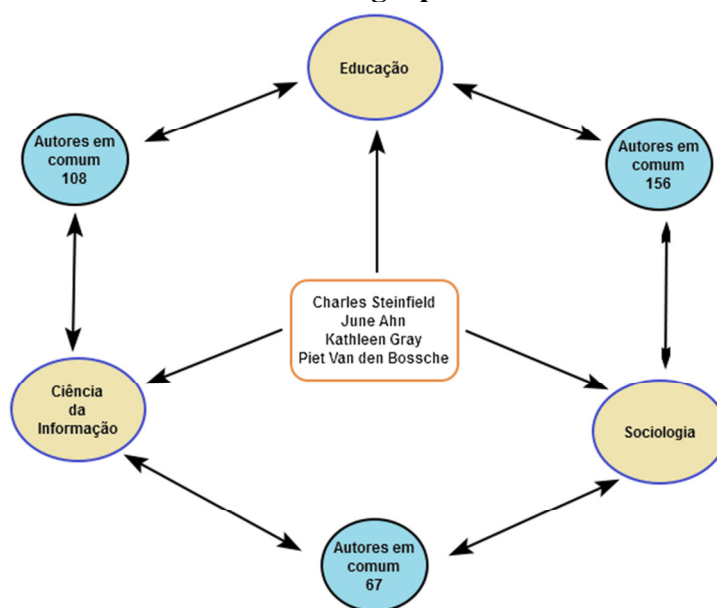
Após análise visual dos dados e a aplicação de algoritmo programado para essa finalidade, verificou-se que a quantidade de autores com problemas no nome não invalidaria os resultados dentro de suas áreas, por representarem baixo percentual. No entanto, deve-se observar que coautores de trabalhos que envolvem o nome de um autor nas condições acima descritas podem ser inseridos ou excluídos de seus verdadeiros clusters, o que implicaria na possibilidade de desconexão entre partes realmente conectadas para cada área e para a totalidade.

Para que seja possível certificar-se de que há autores em comum entre as três áreas estudadas e representadas por trabalhos diferentes (ou seja, o autor faz parte de duas ou três áreas, de fato, apresentando diferentes trabalhos), foram analisadas cada nome nas três bases à procura de seus artigos e em quais seriam diferentes, criando efetivamente uma ligação entre as áreas através do autor. Na Figura 3 estão relacionados os autores e as áreas das quais fazem parte de uma rede colaborativa através da conexão por aqueles pesquisadores.

Ainda que existam autores que conectem as três áreas, não se pode concluir que o conjunto forma uma rede conexa ou um cluster, sendo necessário o exame minucioso das publicações e coautores para que seja possível inferir algo sobre as relações globais entre as áreas, ou mais especificamente, entre autores de redes sociais com trabalhos publicados e inseridos nas bases pesquisadas. Um exemplo é desenvolvido a seguir, para ilustrar a conexão entre autores nas três áreas investigadas, enquanto Charles Steinfield e June Ahn aparecem na conexão de autores entre Ciência da Informação e Educação, e Ciência da Informação e Sociologia, a autora Kathleen Gray aparece apenas na Conexão de Educação e Sociologia.

Os resultados obtidos com o uso do Pajek foram sintetizados na Tabela 1 para as três áreas estudadas (Sociologia, Educação e Ciência da Informação), mostrando simultaneamente índices como o número de autores da base, o número de conexões, a quantidade de artigos, a média de autores por artigo e a média de coautorias para cada cientista. Esses índices têm a finalidade de fornecer argumentos para determinar o grau de relacionamento que os autores têm entre si e características da rede gerada para cada uma das áreas.

Figura 3 - Visualização gráfica das áreas de CI, Educação e Sociologia exemplificando a conexão de um grupo de autores



Fonte: Os autores (2015)

A visualização dos grafos gerados pelo software Pajek contribui para a compreensão de alguns conceitos, como a posição do principal vértice (o autor com o maior grau, ou número de parcerias), a distribuição dos componentes da rede (aqueles grupos em que não há arestas entre eles), os componentes fracos e os fortes ou os vértices adjacentes ao nó de maior grau.

3.2 MÉTRICAS NAS ÁREAS DE CONEXÃO ENTRE AUTORES

As tabelas visam comparar os índices relativos às redes colaborativas para as áreas e a identificação comparativa dos índices de maior componente em cada área de Sociologia, Educação e Ciência da Informação.

Da Tabela 2, nota-se sua densidade 22 vezes menor que a da CI e 26 vezes menor que a da Sociologia, indicando haver uma quantidade muito menor de conexões entre os autores se comparada com todos os artigos recuperados da área.

Das Tabelas 2 e 3, conclui-se ser a área da Educação, embora originalmente um dos

principais propulsores da disciplina em sua gênese, aquela que apresenta os menores índices de colaboração e integração entre os autores que utilizam a ARS em seus trabalhos.

Tabela 2 - Índices relativos às redes colaborativas para as áreas de Sociologia, Educação e CI

	CI		Educação		Sociologia	
Número de autores	3285		3194		4097	
Número de ligações	3233		138		5975	
Número de artigos	2244		3520		1763	
Média de autores por artigo	1,46		0,91		2,32	
Densidade	0,000599373		0,000027063		0,000712102	
<i>Closeness</i> (centralidade de proximidade)	Não computável		Não computável		Não computável	
Grau médio	1,96834		0,086412		2,916768	
Grau máximo	25		6		49	
Número de vértices grau 0 (isolados)	915	27,85	3084	96,56%	442	13,84%
Número de vértices grau 1	735	22,37	28	0,88%	843	26,39%
<i>Network All Degree Centralization</i> (Centralidade de Grau da Rede)	0,00701757		0,00185321		0,01125628	
<i>All degree centrality Maior</i> (Maior centralidade de grau)	25		6		49	
<i>All degree centrality Menor</i> (Menor centralidade de grau)	0		0		0	
<i>Network Betweenness Centralization</i> (Centralidade de Intermediação)	0,00004121		0,00000157		0,00019591	
<i>Average distance among reachable pairs</i> (Distância média entre os vértices)	1,44603		1,05479		1,62963	
Maior distância entre vértices	6		2		6	
Maior Prestígio	0,0081		0,0019		0,0122	

Fonte: Os autores (2015).

A Tabela 3 contém as principais características e índices obtidos dos componentes com o maior número de vértices das áreas de Sociologia, Educação e Ciência da Informação. Sua construção foi fundamentada em software desenvolvido para essa finalidade e o uso do PAJEK.

Tabela 3 - Comparativo entre os índices do maior componente de cada área

Maior Componente	CI	Educação	Sociologia
Vértices	34	7	79
Vértices/Total (%)	1,04	0,22	1,93
Arestas	143	13	297
Maior distância	4	2	5
Distância média	1,99109	1,38095	2,43460
Maior grau	25	6	49
Menor grau	3	2	1
Grau médio	8,41176471	3,71428571	7,51898734
Centralidade de grau (Degree)	0,53409091	0,53333333	0,54562105
Centralidade de proximidade (Closeness)	0,60800074	0,66000000	0,44608387
Centralidade de intermediação (Betweenness)	0,40275373	0,53333333	0,53567148

Fonte: Os autores (2015)

A quantidade de autores que publicam isoladamente representa 96,56% da totalidade, muito maior que da CI (27,85%) e da Sociologia (10,79%). O grau médio de cada autor, ou seja, a quantidade de média de coautoria, de 0,086 para a Educação, segundo mais este critério, coloca a área abaixo dos níveis de colaboração apresentados pela CI (1,97) e Sociologia (2,92).

O maior componente, ou seja, o conjunto que apresenta o maior número de autores em que há um caminho entre dois quaisquer, apresentado na Tabela 3, situa a Educação como aquela com o menor componente, isto representando apenas 0,22% do total de autores, contra 1,93% para o maior componente da Sociologia, ou seja, quase nove vezes, em termos percentuais, que a Educação.

Ainda da Tabela 3, observa-se que a CI apresenta o maior grau médio de 8,41, realçando a diferença entre seu maior componente e o total da base, com grau médio de 1,97, ou seja, os autores presentes neste componente possuem mais laços internos do que a base como um todo, similar à Sociologia, com grau médio do maior componente aproximadamente 7,51 contra 2,92 da base em sua totalidade.

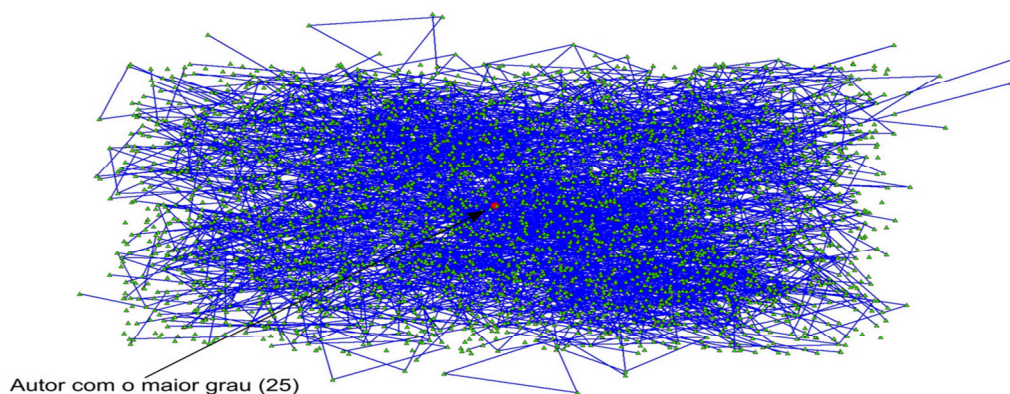
Observa-se também a diferença da colaboração entre as áreas, mas para uma análise mais aprofundada, utilizou-se de gráficos de análise criados com o PAJEK, alimentado por dados obtidos de programa escrito exclusivamente com a finalidade de converter listas de coautores para um formato de rede.

Os Gráficos 1 e 2 (CI) representam a rede formada por todos os autores e suas relações de coautoria na área de CI, exibindo os vértices por suas concentrações e marcando em vermelho o vértice de maior grau (escolha do autor feita pelo aplicativo criado em *VBScript* e inserido no arquivo de dados do PAJEK).

Observa-se dos gráficos, em especial o de componentes (Gráfico 2), que a rede não é constituída por um grafo conexo, o que permite confirmar as conclusões feitas com base na Tabela 2 que aponta 27,85% de autores isolados na área da CI (indicados na parte inferior do gráfico). Ainda com referência ao Gráfico 2, fica visualmente evidenciado que 22,37% (Tabela 2 – Número de vértices de grau 1) publicaram apenas com um coautor, ou seja, 49,77% publicaram com mais que um autor. O Gráfico 3, obtido da Tabela 3 representa o maior componente da rede para a Ciência da Informação, ou seja, o maior conjunto de autores do qual sempre há um caminho entre qualquer par. O número de autores da CI neste componente, 34 autores, representam 1,04% do total de autores da base (Tabela 3).

O Gráfico 1 destaca o autor com o maior grau (25) na base adotada para a Ciência da Informação, representado como um círculo vermelho

Gráfico 1 - Representação gráfica dos vértices (autores) e suas relações (coautorias) para a área de CI, destacando o autor com o maior número de coautores (25)

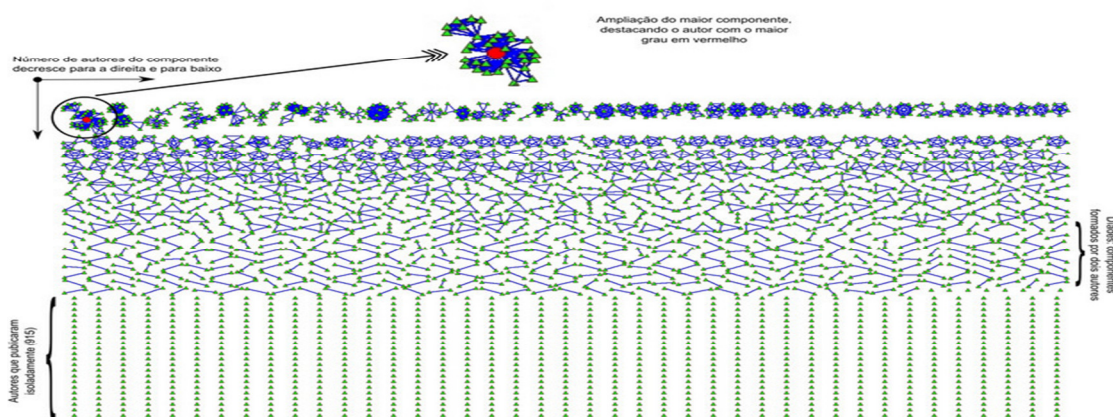


Fonte: Os autores (2015)

Pode-se perceber claramente a dificuldade de separar os componentes visualmente em razão do número de autores (3285). Os outros autores foram representados por triângulos verdes e suas ligações por linhas azuis. O gráfico de componentes permite que se distingam visualmente conjuntos relevantes com relação à quantidade de autores em cada componente.

Os componentes da rede composta pelos autores, nos critérios definidos pela pesquisa, são mostrados no Gráfico 2. O maior componente está destacado na parte superior do gráfico, contendo 34 vértices, incluindo o de maior grau da base, com 25 coautorias. Como o número de autores dos componentes diminui nos sentidos para a direita e para baixo, na parte inferior estão representados os autores individuais (27,85%) e, acima desses, as díades, como indicado.

Gráfico 2 - Representação dos componentes da área de CI, com o maior componente localizado na parte superior à esquerda, com o autor de maior grau indicado em vermelho e autores individuais abaixo.

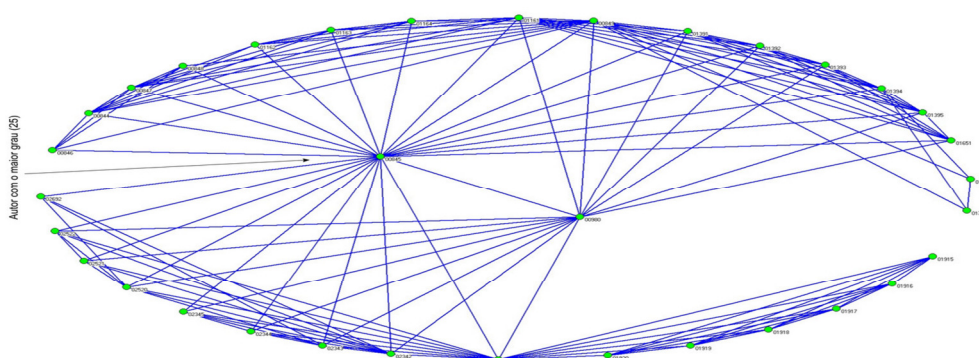


Fonte: Os autores (2015)

Uma alternativa de visualização do maior componente é o Gráfico 3, realçando o autor de maior grau, e dispondo os outros autores de forma elíptica, representados por círculos verdes.

O Gráfico 2 (componentes para a CI), em sua parte superior, apresenta os componentes com o maior número de autores, decrescendo verticalmente para baixo, passando pelas díades (autores que publicaram unicamente com outro) e chegando aos autores individuais, representados por triângulos verdes na parte inferior.

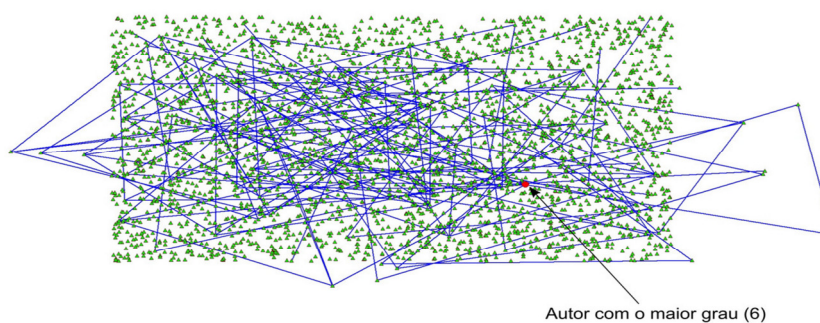
Gráfico 3 - Representação gráfica do maior componente da área de CI composto por 34 autores, ou 1,04% do total de autores da base pesquisada e indicação ao autor de maior grau (25) do componente.



Fonte: Os autores (2015)

O maior componente da área da Ciência da Informação (Gráfico 3) é composto por 34 autores e o autor com a maior centralidade de grau (25) é mostrado ao centro ao lado de outro autor com 16 ligações. Das Tabelas 2 e 3, verifica-se que os autores que concentram o maior número de parceiros em coautoria na base como um todo e no maior componente têm o mesmo grau (25).

Gráfico 4 - Representação gráfica do grafo da Educação, onde se pode observar a existência de muitos autores individuais (não conectados a outros) esparsos pela figura, e destacando o autor de maior grau, ou seja, o de maior número de coautores (6)



Fonte: Os autores (2015)

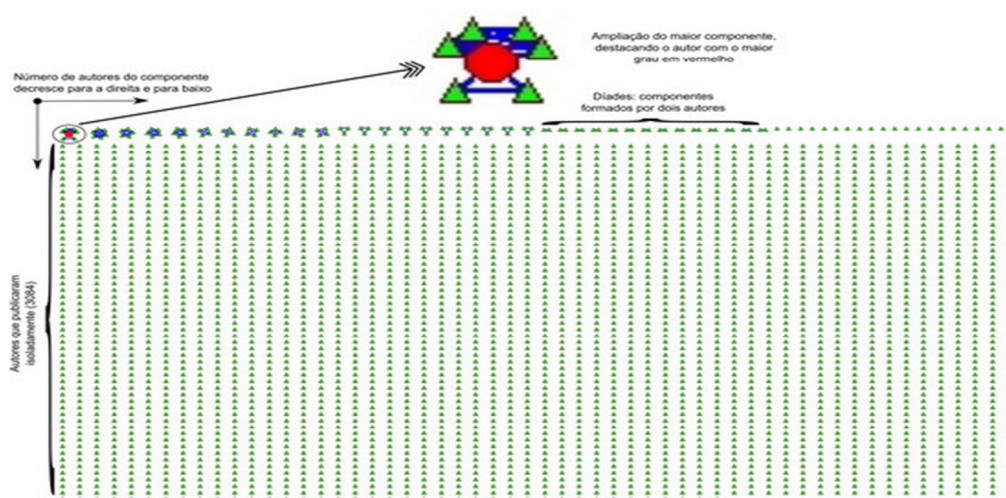
O Gráfico 4 mostra a distribuição de autores e ligações na Educação enquanto o Gráfico 5 de componentes, confirma visualmente a existência de um grande número de autores isolados.

Através do Gráfico 4 e da Tabela 3 pode-se visualizar a distribuição e características do maior componente existente na base da Educação (artigos que utilizam ARS), contendo apenas 7 vértices (0,22%) da totalidade, com seu vértice de maior grau sendo 6, igual ao da rede da base completa.

No Gráfico 4 observa-se a grande quantidade de autores individuais (3084 triângulos verdes sem conexões) e o autor com o maior grau (centralidade de grau igual a 6) em um círculo vermelho.

Pode-se construir a representação gráfica dos componentes de Educação, que tem um número extraordinário de autores individuais (3084), obtido a partir da Tabela 2, mas que é sintetizado em um único gráfico onde aparece alguns autores com 7 e 6 coautores.

Gráfico 5 - Representação gráfica dos componentes da Educação, evidenciando a grande quantidade de autores individuais (3084) e exibe o maior componente na parte superior à esquerda da figura, destacando com um ponto vermelho o autor com o maior número de coautores (6)



Fonte: Os autores (2015)

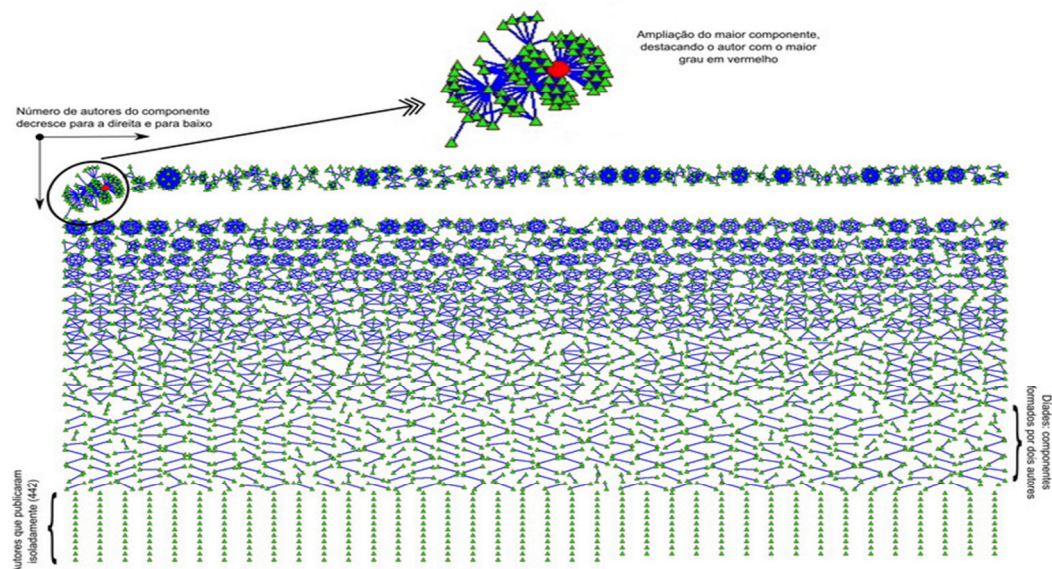
Assim como para a Ciência da Informação, o maior componente da Educação inclui um autor com grau 6, o mesmo que para a base completa, destacado no Gráfico 5. A Sociologia foi a área com os maiores índices entre as três estudadas (Tabela 2), onde se observa que o número de autores (4097) é o maior das três áreas, enquanto o número de artigos o menor (1763), o que nos fornece um índice de 2,32 autores por artigo, a maior média. O grau médio de cada autor é de 2,92 contra 1,97 da CI, o segundo maior índice. O índice de autores que publicaram isoladamente representa 10,79%, que se comparado à CI (27,85%), é mais que o dobro. Já

68,64% dos autores publicaram em parceria, chegando um de seus representantes a 49, praticamente o dobro da CI, com 25. Seu componente com o maior número de vértices (79 autores) também segue a tendência e mais que dobra o índice da CI (34).

O Gráfico 6 exibe os componentes para a Sociologia, permitindo observar visualmente a grande quantidade de componentes compostos por vários autores (linhas de cima do gráfico) e a pouca quantidade de isolados (linhas de baixo do gráfico).

O maior componente da Sociologia tem como características, ainda, o diâmetro (2,43) maior que o da rede de Sociologia. Pode-se construir o gráfico que obtêm o componente com maior grau (49).

Gráfico 6 - Os componentes da Sociologia, com o maior deles contendo 79 autores na parte superior à esquerda, indicando seu autor de maior grau (49) em vermelho e mostrando os 442 autores que publicaram isoladamente na parte inferior.



Fonte: Os autores (2015).

Uma das características marcantes das redes colaborativas está no fato de que estas seguem o modelo de livre-de-escala (*scale-free*), com a presença de *hubs* e apresentando a distribuição de atores por conexões uma variante da distribuição exponencial (*power law*).

Para as três redes de coautoria adotadas para a realização do estudo (Sociologia, Educação e Ciência da Informação), há o particular interesse em delimitar o diâmetro de cada uma delas, ou seja, o comprimento do maior caminho entre dois autores constantes da base de dados, o que fornecerá uma ideia do comportamento dos pesquisadores com relação aos seus pares.

4 AS REDES DE COAUTORIA NA PERSPECTIVA LIVRE-DE-ESCALA

Para as redes de citação em publicações científicas, Sidney Redner da Universidade de Boston (BARABÁSI; BONABEAU, 2003) mostrou que estas seguem a power law, ou seja, nelas a existência de hubs pode ser verificada e a distribuição de ligações segue a regra de que alguns poucos autores (nós) apresentam uma grande quantidade de citações, enquanto a grande maioria recebe poucas citações.

Seguindo os mesmos princípios adotados por Barabási e Bonabeau (2003), poderemos inferir que as redes de coautoria também seguem aos princípios de livre de escala.

Para isso, lançaremos mão das bases analisadas na Sociologia, na Educação e na Ciência da Informação com o objetivo de modelar as redes de coautoria ao livre-de-escala.

É possível verificar, mais adiante, a constatação da ocorrência de *hubs*, autores com grande número de coautores em seus trabalhos, não necessariamente simultâneos, e que, na medida em que a rede cresce, tornam-se nós concentradores de ligações.

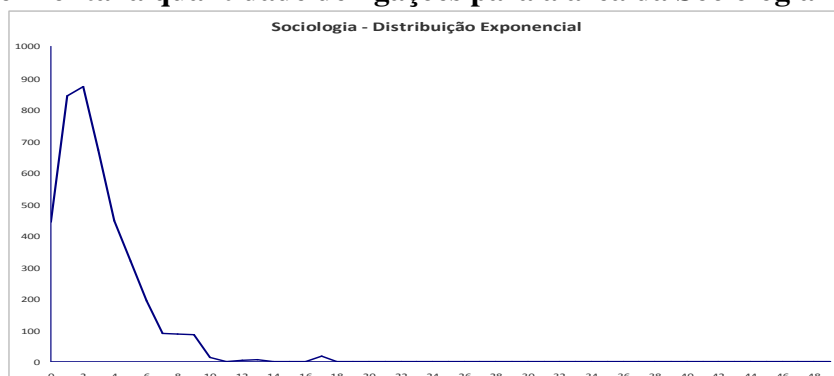
Após a tabulação dos dados possibilitaram a criação dos gráficos 7, 8 e 9, onde se pode observar que: a) para a Sociologia, há uma divergência nos primeiros dados observados (tabelados), onde o número de autores que possuem 0 (zero) ligações (coautorias) é de 442, seguido por 843 que possuem 1 ligação e o ponto máximo vem a seguir, o qual os autores com 2 ligações totalizam 871, diminuindo para 671 autores com 3 ligações e seguindo o modelo exponencial.

Tabela 4 – Distribuição da quantidade de autores por ligação para as áreas da Sociologia, Educação e CI

Sociologia		Educação		Ciência da Informação	
Quantidade de Ligações (Coautoria)	Quantidade de Autores (Nós)	Quantidade de Ligações (Coautoria)	Quantidade de Autores (Nós)	Quantidade de Ligações (Coautoria)	Quantidade de Autores (Nós)
0	442	0	3084	0	915
1	843	1	28	1	735
2	871	2	32	2	712
3	671	3	24	3	392
4	448	4	19	4	222
48	0	5	6	24	0
49	1	6	1	25	1

Fonte: Os autores (2015).

Gráfico 7 - Gráfico relativo à Tabela 4, onde o eixo vertical representa o número de autores e o horizontal a quantidade de ligações para a área da Sociologia

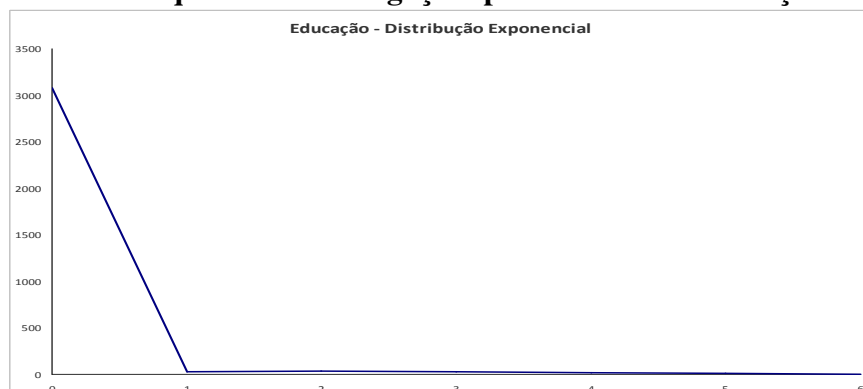


Fonte: Os autores (2015)

A partir daí, com outras variações, como o número de autores com 17 ligações totalizando 18, entre zero autores com 16 ou 18 coautorias, o que mostra a tendência da curva se aproximar de zero à medida que cresce o número de coautorias (máximo 49). A soma dos autores para as 11 primeiras linhas do tabelamento foi responsável por 4064 autores de um total de 4097, ou seja, 99,2% dos autores da área de Sociologia nas bases da pesquisa têm menos do que 10 ligações, enquanto 33 autores possuem entre 11 e 49 (máximo) coautorias, ou seja, 0,8%, caracterizando uma distribuição exponencial, onde poucos autores possuem muitas ligações e muitos possuem poucas ligações.

Semelhante o raciocínio anterior tabelando os dados para o Gráfico 8, tem-se o número de autores que escrevem isoladamente, sem coautoria, é de 3084 de um total de 3194 autores, ou 96,6%, enquanto apenas 110 possuem coautorias entre 1 e 6 (3,4%), e, como mostrado na Figura 2, o número decresce com o aumento do número de coautorias, chegando ao máximo de 6 coautorias para um autor. Aqui, observa-se novamente a tendência exponencial na distribuição dos dados, o que nos permite concluir a modelagem da coautoria na Educação como livres-de-escala.

Gráfico 8 - Gráfico relativo à Tabela 4, onde o eixo vertical representa o número de autores e o horizontal a quantidade de ligações para a área da Educação

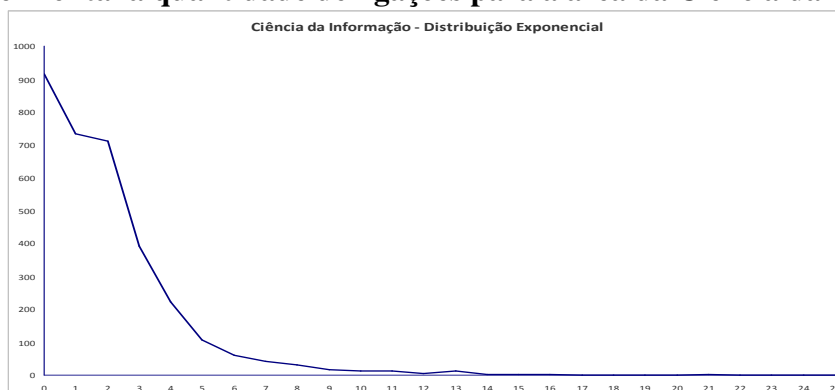


Fonte: Os autores (2015).

O mesmo pode ser feito para os dados compilados para a área de CI a partir da tabela 6 mostra-se que 3083 autores de um total de 3285 (93,9%) possuem 5 ou menos coautorias e que 202 autores (6,1%) possuem entre 6 e 25 coautorias, com apenas um autor com 25 (máximo).

O Gráfico 9 mostra a distribuição da quantidade de autores (eixo vertical) em função da quantidade de coautorias (eixo horizontal), evidenciando a aparência exponencial da curva.

Gráfico 9 - Gráfico relativo à Tabela 4, onde o eixo vertical representa o número de autores e o horizontal a quantidade de ligações para a área da Ciência da Informação



Fonte: Os autores (2015)

5 EXISTEM AUTORES DE DESTAQUE NAS ÁREAS DA EDUCAÇÃO, DA SOCIOLOGIA E DA CIÊNCIA DA INFORMAÇÃO NOS CRITÉRIOS ADOTADOS POR ESTE TEXTO?

As pesquisas de Newman (2004) envolveram a análise de bases de dados bibliográficos, tendo a estatística como ferramenta fundamental associada à Análise de Redes Sociais (ARS), que possibilitou obter informações como o número de autores por artigo, o número de artigos publicados pelos autores, o número de colaboradores para cada cientista, as distâncias típicas entre dois autores e outras medidas inerentes às redes como centralidade de grau, de proximidade e *betweenness*.

A pesquisa de Newman (2004) não levou a conclusões sobre o potencial de ligações nas redes colaborativas e propõe medidas para a força das redes colaborativas ao analisar pares de autores e o número de cientistas os quais o par tenha trabalhado. Através dos resultados da pesquisa, Newman (2004) procurou responder à seguinte questão: "Quem é o cientista melhor relacionado?", o presente trabalho pretende ser uma contribuição a esta resposta.

Nos últimos anos, as redes sociais tornaram-se alvo de estudos em diversas áreas, incluindo negócios, educação, comunidades étnicas ou religiosas (NEWMAN, 2004),

tomando como parte da metodologia o desenvolvimento da base de dados pelo preenchimento de questionários ou de entrevistas, tarefa que apresenta dificuldades na tabulação e análise dos dados obtidos e torna precárias as medidas estatísticas em função do tamanho das amostras.

REFERÊNCIAS

BARABÁSI, A. L.; BONABEAU, E. Scale-free networks. **Scientific American**, New York, v. 288, n. 5, p. 50-59, May 2003. DOI:10.1038/scientificamerican0503-60.

FU, Z. et al. **Parallel breadth first search on GPU clusters**. Utah: University of Utah, 2014. Technical Report. Disponível em: <<http://mapgraph.io/papers/UUSCI-2014-002.pdf>> Acesso em: 19 mar. 2015.

FUNARO, V. M. B. O. **Rede colaborativa entre autores em odontologia**: docentes dos programas de pós-graduação credenciados em universidades participantes do Sistema de Informação Especializado na área de Odontologia (SIEO), 2010. 184f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2010.

MARTELETO, R. M. Análise de redes sociais – aplicação nos estudos de transferência da informação. **Ciência da Informação**, Brasília, DF, v. 30, n. 1, p. 71-81, jan./abr. 2001.

_____.; STOTZ, E. N. **Informação, saúde e redes sociais**: diálogos de conhecimentos nas comunidades da Maré. Rio de Janeiro: Fiocruz, Belo Horizonte: UFMG, 2009. p. 27-42.

_____.; TOMAÉL, M. I. A metodologia de análise das redes sociais. In: VALENTIM, M. L. P. (Org.). **Métodos qualitativos de pesquisa em ciência da informação**. São Paulo: Polis, 2005. cap. 4, p. 81-100.

MEADOWS, D. H. **The limits to growth**. New York: The New American Library, 1972.

NEWMAN, M. E. J. Who is the best connected scientist? A study of scientific coauthorship networks. **Lecture Notes in Physics**, Berlin, v. 650, p. 337-370, 2004. DOI: 10.1007/978-3-540-44485-5_16.

NOOY, W.; MRVAR, A.; BATAGELJ, V. **Exploratory social networks analysis with Pajek**. Cambridge: Cambridge University Press, 2005.

POBLACIÓN, D. A.; MUGNAINI, R.; RAMOS, L. M. S. V. C. **Redes sociais e colaborativas em informação científica**. São Paulo: Angellara, 2009.

REDNER, S. Citation statistics from 110 years of physical review. **Physics Today**, New York, June 2005. Disponível em: <<http://physics.bu.edu/~redner/pubs/pdf/PT.pdf>> Acesso em 14 mar. 2015.

TOMAÉL, M. I. Redes sociais, conhecimento e inovação localizada. **Informação & Informação**, Londrina, v. 12, n. 0, 2007.

WASSERMAN, S.; FAUST, K. **Social network analysis**. Cambridge: Cambridge University Press, 1994.