

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

CÁTIA SOUZA DO NASCIMENTO

**PANDORA - Uma Ferramenta para
Visualização Incremental e Análise de
Redes Sociais Acadêmicas**

Dissertação apresentada como requisito parcial
para a obtenção do grau de
Mestre em Ciência da Computação

Profa. Dra. Carla Dal Sasso Freitas
Orientadora

Profa. Dra. Renata de Matos Galante
Co-orientadora

Porto Alegre, março de 2013

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Nascimento, Cátia Souza do

PANDORA - Uma Ferramenta para Visualização Incremental e Análise de Redes Sociais Acadêmicas / Cátia Souza do Nascimento. – Porto Alegre: PPGC da UFRGS, 2013.

84 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2013. Orientadora: Carla Dal Sasso Freitas; Coorientadora: Renata de Matos Galante.

1. Redes Sociais. 2. Visualização Interativa. 3. Análises de Métricas de Redes Sociais. I. Freitas, Carla Dal Sasso. II. Galante, Renata de Matos. III. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Pró-Reitor de Coordenação Acadêmica: Prof. Rui Vicente Oppermann

Pró-Reitora de Pós-Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do PPGC: Prof. Luigi Carro

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

“O degrau de uma escada não serve simplesmente para que alguém permaneça em cima dele, destina-se a sustentar o pé de um homem pelo tempo suficiente para que ele coloque o outro um pouco mais alto.”

— SR THOMAS HUXLEY

AGRADECIMENTOS

A caminhada até aqui foi longa, mas, enfim, consegui alcançar meu objetivo. Muito tenho a agradecer a todos que me ajudaram nessa jornada. Em primeiro lugar, a Deus, por ter me dado forças para continuar trabalhando mesmo diante das adversidades. Agradeço à minha família, meus pais e minhas irmãs que, mesmo distantes, me deram apoio total em todos os momentos que passei. Agradeço a Lucas por ter me acompanhado em toda minha trajetória. Aos meus amigos que estão longe, mas que sempre estiveram perto em pensamento. Também agradeço às minhas orientadoras Carla Dal Sasso e Renata Galante por terem me acolhido como orientanda em um momento complicado e por terem me dado a oportunidade de continuar trabalhando aqui na UFRGS. Por fim agradeço a todos que aqui me deram apoio, principalmente aos professores que agregaram conhecimentos à minha vida.

SUMÁRIO

LISTA DE FIGURAS	7
LISTA DE TABELAS	10
RESUMO	11
ABSTRACT	12
1 INTRODUÇÃO	13
1.1 Motivação, Objetivo e Contribuição	14
1.2 Organização do Texto	15
2 FUNDAMENTAÇÃO	16
2.1 Métricas	17
2.1.1 Grau	17
2.1.2 Distância	17
2.1.3 Excentricidade	18
2.1.4 Centralidade de Grau	19
2.1.5 Centralidade de <i>Closeness</i> (Proximidade)	19
2.1.6 Centralidade de <i>Betweenness</i> (Intermediação)	20
2.1.7 Coeficiente de <i>Clustering</i> (Agrupamento)	21
2.1.8 Densidade	22
2.1.9 Coeficiente de Colaboração	22
2.1.10 Diâmetro	22
2.1.11 Assortatividade	23
2.2 Considerações Finais	24
3 TRABALHOS RELACIONADOS	25
3.1 Análises de Redes Sociais	25
3.2 Ferramentas de Visualização	27

3.2.1	COMMETRIX	27
3.2.2	COSBILab Graph	30
3.2.3	SocialAction	32
3.2.4	SocialNetSense	33
3.2.5	ArnetMiner	37
3.2.6	igraph	38
3.2.7	Pajek	40
3.2.8	UCINET	41
3.2.9	NodeXL	44
3.2.10	Estudo Comparativo	45
4	DESCRIÇÃO DA FERRAMENTA	50
4.1	Visão Geral	50
4.1.1	Visualização de Grafos	52
4.1.2	Geração de Gráficos para Análises de Dados	54
4.2	Implementação de PANDORA	55
4.2.1	Funcionalidades de PANDORA	56
4.2.2	Limitações de PANDORA	64
4.3	Considerações Finais	65
5	ESTUDOS DE CASO	69
5.1	Estudo de Caso I - Grupos dentro do PPGC	69
5.2	Estudo de Caso II - Pesquisadores Individuais PPGC	74
5.3	Estudo de Caso III - Grupo Completo PPGC 2004 a 2011	75
6	CONCLUSÃO E TRABALHOS FUTUROS	80
	REFERÊNCIAS	82

LISTA DE FIGURAS

2.1	Grafo correspondente à rede de coautoria do grupo de Computação Gráfica da UFRGS e seus colaboradores no ano de 2008.	16
2.2	Distâncias entre nós e diâmetro do grafo	18
2.3	Um dos possíveis caminhos para representar o diâmetro do grafo.	23
3.1	Interface da ferramenta COMMETRIX	28
3.2	Métricas geradas pela COMMETRIX: (a) Métricas gerais da rede e (b) Métricas de um nó específico	29
3.3	Interface da ferramenta COSBILab Graph	30
3.4	Características básicas das estruturas em COSBILab Graph	31
3.5	SocialAction: Mudança de rótulos dos nós de acordo com as métricas de análise de redes sociais	32
3.6	Grafo mostrando dois grupos de senadores americanos e a frequência de votação em conjunto	33
3.7	Gráficos gerados com a SocialAction que permitem que o usuário compare duas métricas	34
3.8	Esquema proposto para o <i>framework</i> SocialNetSense	34
3.9	Representação visual do SocialNetSense	35
3.10	<i>Interface</i> do RBS	36
3.11	Grafo de coautoria gerado pela ArnetMiner formado por dois pesquisadores escolhidos e com seus colaboradores	37
3.12	Interface da ferramenta online ArnetMiner	38
3.13	Interface criada utilizando a linguagem R que possui suporte para a biblioteca <i>igraph</i>	39
3.14	Interface inicial da ferramenta Pajek	40
3.15	Um grafo gerado usando a Pajek	41
3.16	Interface da ferramenta UCINET	42
3.17	Exemplo de arquivo FreemanDegree, um dos arquivos de saída gerado UCINET	43

3.18	Grafo criado com o NetDraw	43
3.19	Grafos gerados com o nodexl: (a) Grafo Completo e (b) Grafo Filtrado	44
3.20	Criação de uma restrição para gerar grupo de nós de acordo com seus atributos	45
4.1	Visão geral da proposta de PANDORA: o usuário entra com um arquivo XML, contendo a produção bibliográfica, que é usado pelo sistema para gerar uma visualização de grafo, bem como dados analíticos sobre este grafo. A seta em azul indica interação entre o usuário e o sistema.	51
4.2	Esquema do fluxo de execução da aplicação. O fluxo superior representa a abordagem de visualização do grafo completo, em que a aplicação recebe um arquivo XML com todos os pesquisadores, para em seguida gerar um grafo completo juntamente com gráficos e planilhas. O fluxo inferior representa a abordagem incremental em que um pesquisador é escolhido pelo usuário; a aplicação recebe vários arquivos XML e cria um grafo utilizando o XML referente ao pesquisador escolhido pelo usuário, juntamente com planilhas e gráficos.	53
4.3	Esquema da abordagem incremental: ao clicar em um nó qualquer, o grafo é expandido caso o pesquisador tenha colaboradores que não estejam visíveis ainda.	54
4.4	Representação dos módulos de PANDORA incluindo interação com o usuário.	55
4.5	Arquivo XML da pesquisadora Carla Freitas no ano 2008.	57
4.6	Trechos do arquivo de entrada utilizado na aplicação.	58
4.7	Exemplo de grafo completo mostrado ao usuário.	59
4.8	Colorações que representam as diferentes categorias para os pesquisadores. Na ordem das imagens: Discente, Docente e Participante Externo.	59
4.9	Relação Grau x Quantidade de Publicações correspondente ao grafo completo.	61
4.10	Gráfico Assortatividade (Grau x Média dos graus dos colaboradores) para dados de um grafo completo.	61
4.11	Visualização gerada depois do usuário selecionar um pesquisador específico no grafo em exibição no momento.	62
4.12	Grafo de artigos após algumas expansões no grafo de Coautoria.	63
4.13	Coefficiente de Colaboração calculado sobre os dados que se referem aos pesquisadores representados no grafo visualizado.	64
5.1	Grafo de artigos do grupo de Sistemas da Informação no triênio de 2007 a 2009.	71
5.2	Grafo de artigos do grupo de Computação Gráfica no triênio de 2004 a 2006.	71
5.3	Grafo de artigos do grupo de Sistemas Embarcados no triênio de 2007 a 2009.	72
5.4	Evolução dos grafos de colaborações nos anos de 2004 a 2011 do grupo de Sistemas de Informação. Legenda: A - triênio de 2004 a 2006, B - triênio de 2007 a 2009 e C - triênio incompleto contendo 2010 e 2011.	72

5.5	Evolução nos grafos de colaboração entre os anos de 2004 a 2011 para o grupo de Sistemas de Informação. Legenda: A - triênio de 2004 a 2006, B - triênio de 2007 a 2009 e C - triênio incompleto contendo 2010 e 2011.	73
5.6	Gráficos de barras com métricas para cada um dos pesquisadores escolhidos. .	76
5.7	Gráficos das métricas referentes aos anos 2004 a 2010.	77
5.8	Gráfico Grau X Quantidade de Publicações para o PPGC no ano de 2011. . . .	79
5.9	Gráfico de Assortatividade do ano de 2011 para todo o PPGC.	79

LISTA DE TABELAS

3.1	Tabela com características de visualização e interatividade.	48
3.2	Tabela com resumo das métricas calculadas por cada ferramenta.	49
4.1	Tabela com características de visualização e interatividade com acréscimo de PANDORA.	67
4.2	Tabela com resumo das métricas calculadas por cada ferramenta acrescido de PANDORA.	68
5.1	Coefficientes de Colaboração dos grupos em cada triênio.	70
5.2	Métricas dos pesquisadores individuais escolhidos.	74

RESUMO

A análise de redes sociais através de ferramentas visuais permite ao usuário extrair informações importantes sobre os indivíduos e seus relacionamentos. Através dela é possível entender de que forma os grupos se organizam. Para o caso de redes de coautoria, algumas conclusões como qual o pesquisador com maior prestígio na rede ou qual pesquisador tem o perfil mais parecido com o seu podem ser obtidas.

Uma grande quantidade de ferramentas têm sido desenvolvidas para visualização de redes sociais. Algumas delas permitem que análises sejam feitas sobre os dados que compõem as redes, porém a maioria não mostra os resultados dessas análises sobre o próprio grafo, em geral apresentando-as como informações textuais.

Neste trabalho, inicialmente, foi feito um estudo comparativo de diversas ferramentas visuais que além de mostrar redes sociais de forma gráfica, fazem algum tipo de análise sobre a rede. O estudo subsidiou o desenvolvimento de Pandora, uma ferramenta que permite visualização incremental de redes acadêmicas além de interação com o grafo e cálculo de algumas métricas, como centralidades, assortatividade e coeficiente de colaboração.

Palavras-chave: Redes Sociais, Visualização Interativa, Análises de Métricas de Redes Sociais.

ABSTRACT

The analysis of social networks through visual tools allows to extract important information about each individual and their relationships. Through it we can understand how groups are organized. For the case of co-authorship networks, some conclusions about which researcher has greater prestige in the network or where the researcher has profile more like his can be obtained.

A lot of tools have been developed for visualizing social networks. Some of them allow analyzes are made about the data that make up the networks, but most do not show the results of these analyzes on the graph itself, usually presenting them as textual information.

This work was initially done a comparative study of various visual tools that show beyond social networks graphically. This tools generally do some kind of analysis on the network. In this work Pandora was developed, a visualization tool that allows incremental academic networks and interaction with the graph and calculate some metrics, such as centralities, assortativity and collaboration coefficient.

Keywords: Social Networks, Interactive Visualization, Metrics of Analysis.

1 INTRODUÇÃO

A crescente disponibilidade de ferramentas na Web, que possibilitam a troca de um volume também crescente de informações entre pessoas ou organizações, está também permitindo que cientistas analisem em larga escala as maneiras pelas quais as conexões entre as pessoas são estabelecidas e como as chamadas redes sociais evoluem ao longo do tempo. Há muitos anos, pesquisadores de ciências sociais estudam essas redes buscando compreender as estruturas que se formam e os diferentes papéis que surgem (WASSERMAN; FAUST, 1994), contribuindo para um melhor entendimento das relações nessas redes.

De forma simplificada, pode-se dizer que uma rede social é um conjunto de indivíduos com conexões entre eles que dependem dos relacionamentos existentes. Computacionalmente, toda rede pode ser representada como um grafo no qual os nós estão interligados por uma ou mais arestas (relacionamentos). Recentemente, redes sociais se tornaram bastante populares, em especial, com o grande crescimento dos sites de relacionamentos disponíveis na Internet, como: Orkut, Facebook, Twitter e MySpace.

A necessidade de representar visualmente redes sociais existe há muitos anos (FREEMAN, 2009; NORTHWAY, 1952), pois permite, através de uma maneira natural, expressar a conectividade e facilitar a identificação de padrões na rede. Nesse sentido, a visualização e a análise das redes sociais vem atraindo grande interesse tanto na área de análise de dados quanto na de visualização de informações. Esse interesse é motivado, também, pela dimensão que essas redes podem atingir, o que introduz problemas interessantes de pesquisa.

Uma das formas de contornar a dificuldade em lidar com redes de grandes dimensões é através de mecanismos de interação, que surgem como um interessante aliado para auxiliar o usuário a entender e analisar o grafo, bem como formar o mapa mental da rede social de forma mais clara e objetiva. As principais técnicas de visualização de redes incluem diversos recursos como: busca, *zoom*, filtros e manipulação direta do tipo arrastar e soltar, dentre outros (HERMAN; MELANCON; MARSHALL, 2000).

Um aspecto importante na análise de redes sociais apoiada por visualização é a possibilidade de expandir a rede à medida que os relacionamentos de seus indivíduos são explorados. Existem algumas operações de conjunto que podem ser aplicadas sobre grafos, como, por exemplo, união,

composição, soma, produto cartesiano, entre outras (NASCIMENTO et al., 2011). Considerando um grafo que representa uma rede de colaboração científica, a operação de união, por exemplo, equivale a unir uma rede, que esteja sendo visualizada, com outras redes obtidas a partir de nós da rede atual. Esse tipo de visualização permite construir análises sobre um tema, navegando gradualmente pelos diversos relacionamentos de coautoria.

Este trabalho foca em redes sociais de colaboração científica, em especial redes de coautoria, as quais representam as interações no meio acadêmico. Em redes de coautoria, os nós são os pesquisadores enquanto suas publicações são representadas pelas ligações, ou arestas, entre os nós que representam os autores de uma publicação. No desenvolvimento deste trabalho, as redes sociais de colaboração científica são utilizadas como fontes de dados para experimentos e avaliação de resultados. Cabe ressaltar que, embora a estrutura dessas redes possa ser extremamente complexa (HERMAN; MELANCON; MARSHALL, 2000), analisá-las permite identificar os diversos tipos de conexões existentes entre as pessoas dentro e fora de suas instituições, permitindo o planejamento de novas cooperações que podem causar impacto direto na qualidade da produção científica de indivíduos.

1.1 Motivação, Objetivo e Contribuição

Ao estudar algumas ferramentas de visualização e análises de redes sociais, notou-se um deficit nessas ferramentas. Percebeu-se que existem muitos trabalhos analisam o desempenho de grupos de pós-graduação espalhados no Brasil, porém muitos deles fazem apenas análises numéricas e outros utilizam ferramentas que servem para uso geral em rede sociais ou grafos em geral, não sendo específicos para redes acadêmicas. Com isso, algumas métricas de redes acadêmicas deixam de ser calculadas automaticamente. Além disso, as ferramentas utilizadas na maioria das vezes possuem visualização simplificada sem muita interação, até por se tratarem de ferramentas que foram desenvolvidas há muito tempo sem muitos recursos visuais, e que deixaram de ser mantidas e evoluídas com o passar dos anos.

O objetivo do presente trabalho é a proposta de uma ferramenta de visualização e análise de redes de coautoria. Sendo que as principais contribuições desse trabalho dizem respeito a visualização interativa e de forma gradual de grafos e ao cálculo de métricas (centralidades, assortatividade, quantidade de publicações, etc) sobre grafos de forma que um usuário do sistema possa analisar o comportamento de uma rede de coautoria, sendo que essas análises possam ser feitas de forma gradual (em subredes) ou para um grafo completo. Para os estudos de caso, foram utilizados dados provenientes dos relatório CAPES que foram gerados dos anos de 2004 a 2011 para o PPGC da UFRGS.

1.2 Organização do Texto

O texto está organizado como segue. No Capítulo 2 apresentam-se as descrições das métricas estudadas para o desenvolvimento do trabalho com o auxílio de um exemplo de grafo que representa o grupo de computação gráfica da UFRGS no ano de 2008. No Capítulo 3 é feita a revisão de trabalhos correlatos tanto de análises de redes sociais quanto de visualização de redes. No Capítulo 4 é detalhada a aplicação que foi implementada para validar a idéia proposta. No Capítulo 5 são mostrados os estudos de caso que foram utilizados. Por fim, o Capítulo 6 apresenta a conclusão e discute as sugestões de trabalhos futuros além da conclusão do trabalho.

2 FUNDAMENTAÇÃO

Nesse capítulo são apresentadas as definições básicas que fundamentam teoricamente este trabalho. Assim, a Seção 2 conceitua as principais métricas utilizadas nos trabalhos relacionados a redes sociais representadas a partir de grafos. Na Seção 2.2 é feito o fechamento do capítulo relacionando-o às contribuições deste trabalho.

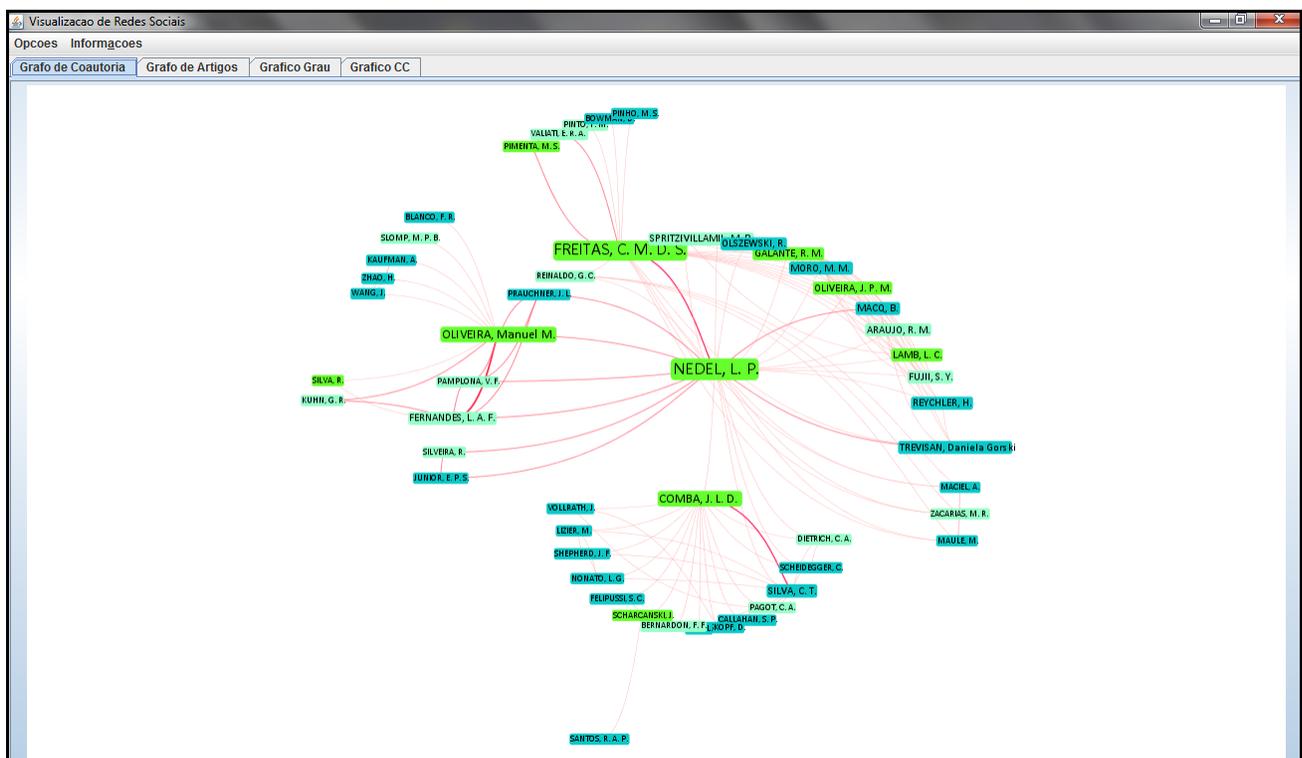


Figura 2.1: Grafo correspondente à rede de coautoria do grupo de Computação Gráfica da UFRGS e seus colaboradores no ano de 2008.

2.1 Métricas

Para esse trabalho as métricas mais empregadas na construção de ferramentas de análise de redes sociais foram selecionadas, estudadas, sendo apresentadas a seguir. Foram tomados como base para esses conceitos os trabalhos de Wasserman e Faust, Freeman e Jamakovic e Uhlig (WASSERMAN; FAUST, 1994; FREEMAN, 1979; JAMAKOVIC; UHLIG, 2008).

Para auxiliar a compreensão das definições, um pequeno grupo de pesquisadores foi selecionado, sendo gerada uma rede social de coautoria. Cada métrica será exemplificada neste capítulo aplicada sobre essa rede social gerada. Os dados são reais e referem-se ao ano de 2008, envolvendo os seguintes pesquisadores: Carla Freitas, Luciana Nedel, João Comba, Manoel Oliveira, Jacob Sharcanski, integrantes do grupo de Computação Gráfica da UFRGS no ano de 2008, e são representados no grafo pelos rótulos: FREITAS, C.M.D.S., COMBA, J.L.D., OLIVEIRA, Manuel M. e SCHARCANSKI, J., respectivamente. É importante ressaltar que os dados incluem os colaboradores desses pesquisadores também. O grafo correspondente a esse grupo pode ser visto na Figura 2.1, tendo sido gerado utilizando a ferramenta implementada nesse trabalho com o objetivo de ilustrar a rede de colaboração.

2.1.1 Grau

O grau do nó representa a quantidade de nós que estão interligados a ele, podendo variar de zero (quando o nó é chamado de isolado) até $n-1$, onde n denota a quantidade total de nós no grafo. O grau do nó mostra a atividade do nó no grafo. Calculando-se essa métrica para os pesquisadores integrantes do grupo de Computação Gráfica apresentados no grafo da Figura 2.1, têm-se os seguintes valores:

$$\begin{aligned} g(\text{Carla}) &= 22 \\ g(\text{Luciana}) &= 27 \\ g(\text{João}) &= 15 \\ g(\text{Manoel}) &= 11 \\ g(\text{Jacob}) &= 3 \end{aligned}$$

Do ponto de vista da semântica da rede, os valores indicam que a pesquisadora Luciana é a que tem maior número de colaboradores dentro do grupo e o pesquisador Jacob tem menor número de colaboradores.

2.1.2 Distância

A distância geodésica, ou simplesmente distância, mede o número de saltos para chegar de um nó a outro no grafo. A distância pode ser definida também como o menor caminho entre dois nós. Se não for possível alcançar um nó a partir de outro, no caso de existirem duas componentes desconexas, então a distância entre eles os nós tende ao infinito.

Dado um grafo G , a distância entre os nós n_1 e n_2 (d_{12}) é igual à distância entre n_2 e n_1 (d_{21}). Um exemplo do cálculo das distâncias entre os nós no grafo pode ser visto na Figura 2.2.

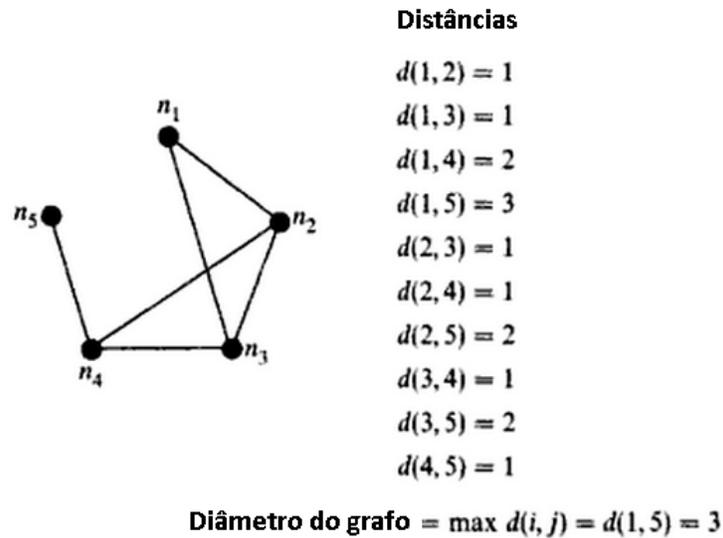


Figura 2.2: Distâncias entre nós e diâmetro do grafo (WASSERMAN; FAUST, 1994).

Aplicando o conceito de distância ao grafo mostrado na Figura 2.1, apenas aos pares que envolvem os professores integrantes do grupo de Computação Gráfica sem envolver seus colaboradores são encontrados os seguintes valores:

$$\begin{aligned}
 d(\text{Carla, Luciana}) &= 1 \\
 d(\text{Carla, João}) &= 2 \\
 d(\text{Carla, Manoel}) &= 2 \\
 d(\text{Carla, Jacob}) &= 3 \\
 d(\text{Luciana, João}) &= 1 \\
 d(\text{Luciana, Manoel}) &= 1 \\
 d(\text{Luciana, Jacob}) &= 2 \\
 d(\text{João, Manoel}) &= 2 \\
 d(\text{João, Jacob}) &= 1 \\
 d(\text{Manoel, Jacob}) &= 3
 \end{aligned}$$

2.1.3 Excentricidade

A excentricidade de um nó representa a maior distância geodésica entre um nó e todos os outros nós do grafo. Tomando-se como exemplo a Figura 2.1, para calcular a excentricidade de cada nó na rede, deve-se calcular a distância deste nó a todos os outros nós do grafo. Desta forma, os valores encontrados para excentricidade dos nós representantes dos integrantes do grupo de Computação Gráfica são:

$$e(\text{Carla}) = 4$$

$$\begin{aligned}
 e(\text{Luciana}) &= 3 \\
 e(\text{João}) &= 3 \\
 e(\text{Manoel}) &= 4 \\
 e(\text{Jacob}) &= 4
 \end{aligned}$$

Assim, é possível perceber que os pesquisadores Luciana e João estão mais próximos de todos os nós do grafo do que os outros 3 pesquisadores. Para Luciana, por exemplo, a maior quantidade de saltos necessários para alcançar um nó qualquer no grafo partindo do nó que a representa é 3. Quanto menor a Excentricidade de um nó, melhor é o seu relacionamento com os outros nós do grafo. Isto significa que quanto mais relacionamentos diretos o nó possui no grafo, menor é a necessidade de nós intermediadores entre ele e os outros nós.

A excentricidade de cada nó pode ser visualizada facilmente usando o leiaute radial, colocando o nó desejado no centro e contando quantos níveis o grafo possui naquela visualização.

2.1.4 Centralidade de Grau

A centralidade de grau é a forma mais simples de medir a importância de um nó no grafo e tem o mesmo valor numérico do grau do nó, ou seja, conta-se a quantidade de nós ligados ao nó no grafo. Assim, é possível medir o nível de atividade de um determinado nó, no sentido de que ele tem mais relacionamentos com outros nós. Para o exemplo da Figura 2.1, calculando as centralidades de grau para cada um dos nós pertencentes ao grupo tomado como exemplo, tem-se:

$$\begin{aligned}
 cg(\text{Carla}) &= 0,43 \\
 cg(\text{Luciana}) &= 0,53 \\
 cg(\text{João}) &= 0,29 \\
 cg(\text{Manoel}) &= 0,22 \\
 cg(\text{Jacob}) &= 0,06
 \end{aligned}$$

Esses valores estão normalizados pois alguns autores normalizam o valor do grau dividindo pela quantidade de nós na rede, ou seja, o máximo possível de ligações para um nó, para poder comparar os nós entre si, assim essa métrica diferencia-se do grau do nó. A normalização é feita dividindo o grau do nó pela quantidade de nós do grafo menos 1, pois esse é o máximo de ligações possíveis para um nó. Para o grafo de exemplo usado neste trabalho, a quantidade de nós é 52.

Assim, o pesquisador com maior centralidade de grau no grupo é Luciana e Jacob o menor. Ou seja, Luciana tem mais possibilidades de publicar com pesquisadores diferentes do que Jacob.

2.1.5 Centralidade de *Closeness* (Proximidade)

Esta métrica revela o quão próximo um nó está de todos os outros. Para o cálculo dessa métrica, são somadas todas as distâncias geodésicas entre o nó e os outros nós do grafo. Em seguida, é calculado o inverso deste valor, já que quanto mais distante o nó estiver de outro nó, menor será a proximidade.

A fórmula utilizada para o cálculo da centralidade de proximidade de um nó n_i é a seguinte:

$$C_p(n_i) = \left(\sum_{j=1}^n d(n_i, n_j) \right)^{-1}, \quad (2.1)$$

onde n representa a quantidade de nós e n_j um nó qualquer no conjunto de nós, exceto n_i . Para normalizar o valor, autores sugerem que o valor de $C_p(n_i)$ seja multiplicado por $(n-1)$ (WASSERMAN; FAUST, 1994).

Calculando a centralidade de proximidade para cada um dos pesquisadores do grupo de Computação Gráfica, os valores aproximados obtidos são:

$$\begin{aligned} \text{cp}(\text{Carla}) &= 0,51 \\ \text{cp}(\text{Luciana}) &= 0,67 \\ \text{cp}(\text{João}) &= 0,52 \\ \text{cp}(\text{Manoel}) &= 0,47 \\ \text{cp}(\text{Jacob}) &= 0,35 \end{aligned}$$

Assim, Luciana mostra-se mais próxima dos outros nós do grafo do que os outros calculados, isso significa que ela através do nó representado por ela é possível alcançar os outros nós do grafo de forma mais rápida. Em teoria, seria mais fácil para ela colaborar com mais pesquisadores que estão em outros níveis no grafo, uma vez que ela teria mais afinidade com os outros nós em relação às colaborações em comum.

2.1.6 Centralidade de *Betweenness* (Intermediação)

A centralidade de um nó mede a importância deste nó para o grafo, a partir da quantidade de caminhos mínimos entre os outros nós que passam por este nó. Assim, calcula-se todos os caminhos geodésicos entre os pares de nós nos quais o nó, ao qual se está calculando a métrica, está incluso. A fórmula utilizada para o cálculo da centralidade de intermediação de um nó n é a seguinte:

$$C_i(n_i) = \left(\sum_{j < k} \frac{d_{jk}(n_i)}{d_{jk}} \right)^{-1}, \quad (2.2)$$

onde j e k representam nós no grafo, $d_{jk}(n_i)$ representa as distâncias mínimas entre os nós j e k que envolvem n_i e d_{jk} representa todos os caminhos mínimos que ligam os nós j e k .

A normalização, segundo os autores, é feita dividindo o $C_i(n_i)$ por $\frac{(n-1)*(n-2)}{2}$ que representa todas as combinações de pares de nós no grafo exceto pelo nó n_i . Assim, é possível obter a proporção de caminhos mínimos em que o nó n_i está envolvido com $C_i(n_i)$ e deste valor obter a proporção em relação a quantidade total de caminhos possíveis entre todos os pares de nós. As centralidades de intermediação para os pesquisadores do grupo de Computação Gráfica são:

$$\begin{aligned}
 ci(\text{Carla}) &= 0,22 \\
 ci(\text{Luciana}) &= 0,71 \\
 ci(\text{João}) &= 0,31 \\
 ci(\text{Manoel}) &= 0,22 \\
 ci(\text{Jacob}) &= 0,04
 \end{aligned}$$

Luciana apresenta maior centralidade de intermediação, 0,71, ou seja, ela participa de 71% dos caminhos mínimos entre todos os nós do grafo. A centralidade de intermediação mostra o quão importante o nó é para o grafo no sentido de que se for retirado do grafo quebrará muitas relações dificultando o fluxo no grafo. Ou seja, o nó intermediador estaria sendo removido o que dificultaria, no caso do grafo de colaboração, o surgimento de novas colaborações entre colaboradores diferentes de dois pesquisadores que são colaboradores entre si.

2.1.7 Coeficiente de *Clustering* (Agrupamento)

Esta métrica quantifica o quão densamente conectados estão os vizinhos de um nó. Este cálculo é feito através da razão entre a quantidade de ligações existentes entre os vizinhos e a quantidade de ligações possíveis que podem ser formadas entre eles.

A fórmula que representa esse cálculo é dada por:

$$C_a(n_i) = \frac{2 * e_{jk}}{k_i * (k_i - 1)}, \quad (2.3)$$

onde e_{jk} representa a quantidade de arestas existentes entre os vizinhos de n_i , j 's e k 's, e $\frac{k_i * (k_i - 1)}{2}$ a quantidade de arestas máximas possíveis entre os nós e k_i representa a quantidade de vizinhos para o nó n_i , para um grafo não direcionado.

$$\begin{aligned}
 ca(\text{Carla}) &= 0,24 \\
 ca(\text{Luciana}) &= 0,19 \\
 ca(\text{João}) &= 0,21 \\
 ca(\text{Manoel}) &= 0,20 \\
 ca(\text{Jacob}) &= 0,33
 \end{aligned}$$

Para a pesquisadora Carla, por exemplo, têm-se 55 arestas envolvidas entre seus vizinhos, sendo que se todos estivessem conectados existiriam 231, o que faz com que $C_c = 55/231 = 0,238$. Isto significa cerca de 23,8%, uma conectividade relativamente baixa. Já Jacob apresentou 33,3% de conectividade entre seus vizinhos, porém deve-se levar em conta que ele possui apenas 3 colaboradores no grafo, enquanto os outros têm no mínimo 10 colaboradores, ou seja, a probabilidade de estarem ligados entre si é menor.

2.1.8 Densidade

A densidade do grafo é medida calculando a razão entre a quantidade de relacionamentos existentes no grafo e a quantidade total de relacionamentos que podem ser formados. Isto mostra a coesão do grafo. Quanto maior a densidade do grafo, mais ligado ele se mostra, e melhor será o fluxo de dados dentro dele. No caso do exemplo da Figura 2.1, que é uma rede de coautoria, a maior densidade representa a maior quantidade de publicações produzidas em colaboração. Então, para esse exemplo, assumindo que RE = quantidade de relacionamentos existentes, RP = quantidade de relacionamentos possíveis e DE = densidade, tem-se que:

$$\begin{aligned} DE &= RE/RP \\ RE &= 140 \\ RP &= 1326 \\ DE &= 0,106 \end{aligned}$$

Sabendo-se que a densidade da rede poderia variar entre 0 e 1, a densidade DE calculada representa apenas cerca de dez por cento das possibilidades.

2.1.9 Coeficiente de Colaboração

O Coeficiente de Colaboração é uma métrica que relaciona a quantidade de publicações (contadas nas arestas do grafo de coautoria) e número de autores para medir o índice de colaboração em um grupo. Para o cálculo é feito um somatório, onde para cada quantidade de autores de artigos, são contadas quantas publicações possuem aquela quantidade de autores. Por exemplo, se temos um conjunto de artigos, separamos em grupos de acordo com a quantidade de autores que cada publicação possui. Assim é feito um somatório dividindo pela quantidade de artigos no total.

$$C_{co} = 1 - \frac{\sum_{j \in 1, n} (1/j) F_j}{N}, \quad (2.4)$$

onde N representa a quantidade total de publicações no grupo, e F_j a quantidade de publicações com j autores.

Essa métrica deve ser calculada para um conjunto de publicações de um conjunto de autores, não somente para um pesquisador. Desta forma, para o exemplo da Figura 2.1 tem-se que o Coeficiente de Colaboração encontrado é de 0,67.

2.1.10 Diâmetro

Para encontrar o diâmetro de um grafo é preciso tomar todas as maiores distâncias geodésicas entre todos os pares de nós no grafo, e a partir daí selecionam a maior dessas distâncias que será o diâmetro do grafo propriamente dito. Em outras palavras, é feita a seleção da maior excentricidade dos nós no grafo. Para o grafo da Figura 2.1, tem-se que o diâmetro D é igual a 5. Um dos caminhos possíveis para representar o diâmetro do grafo é partir de SANTOS R.A.P. e chegar a PINTO, F.M.

passando por SCHARCANSKI, J., COMBA, J.L.D., NEDEL L.P., FREITAS, C.M.D.S e PINTO, F.M., nessa ordem, como pode ser visto na Figura 2.3.

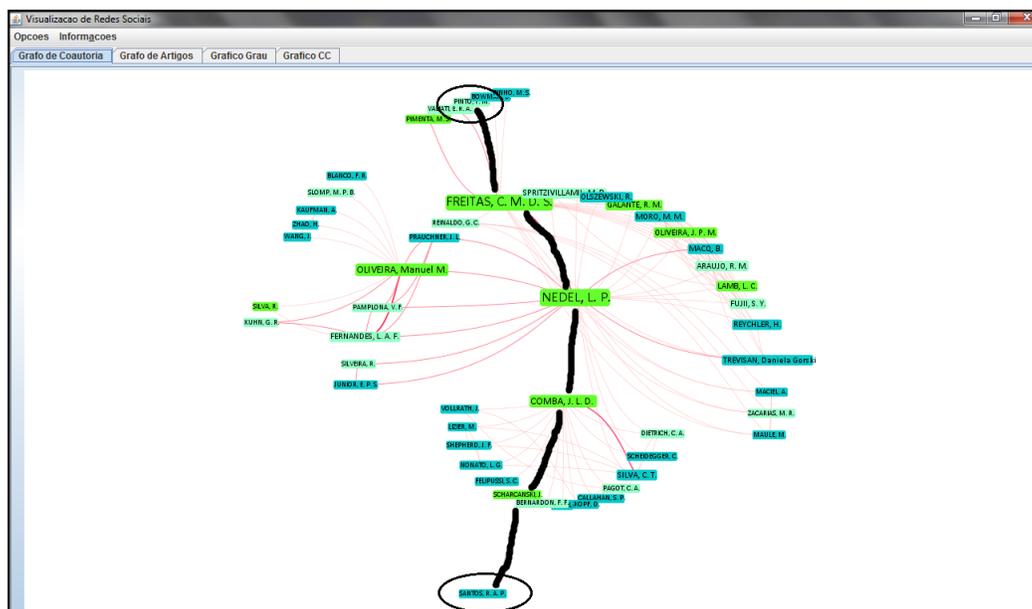


Figura 2.3: Um dos possíveis caminhos para representar o diâmetro do grafo.

2.1.11 Assortatividade

Newman desenvolveu trabalhos relacionados à "*Mixing Pattern*" em redes, que calculam medidas que representam o padrão de mistura que ocorre de acordo com alguma característica específica dos nós (NEWMAN, 2002). Por exemplo, no trabalho de (SOUSA; RUIZ, 2011), essa métrica foi calculada com o intuito de indicar a relação entre nós que representavam doenças mais frequentes e nós que representavam doenças raras. A forma mais empregada dessa métrica é para verificar relacionamentos entre nós que possuem o mesmo grau. A fórmula utilizada para o cálculo da Assortatividade está relacionada à correlação entre os graus de vértices adjacentes e é mostrada a seguir:

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i (1/2)(j_i + k_i)]^2}{M^{-1} \sum_i (1/2) j_i^2 k_i^2 - [M^{-1} \sum_i (1/2)(j_i + k_i)]^2} \quad (2.5)$$

Uma das formas de caracterizar uma rede como assortativa ou disassortativa, segundo o trabalho de (BENEVENUTO; ALMEIDA; SILVA., 2011), é calcular a média dos graus ligados a cada grau k presente no grafo. A partir daí é feita a avaliação das médias, onde valores crescentes, ou seja, nós com grau maior se relacionando com outros nós de grau maior ou igual ao dele indicam uma rede assortativa. Devido à facilidade em entender gráficos visualmente e como já existiam ou-

tros gráficos que eram produzidos nesse trabalho, escolheu-se analisar a assortatividade dos grafos através de um gráfico também.

2.2 Considerações Finais

Este capítulo apresentou as principais métricas referentes às redes sociais, as quais permitem entender aspectos da estrutura das redes. Cada métrica pode ser utilizada de forma individual ou em conjunto, olhando mais de uma métrica ao mesmo tempo, para analisar uma rede. Algumas delas possuem interpretações semelhantes como grau e centralidade de grau, desta forma o usuário deve escolher as métricas que deseja avaliar de acordo com o seu interesse. Por isso, nesse trabalho, foi dada a liberdade ao usuário de escolher quais métricas ele deseja visualizar nos arquivos '.xls' gerados. Entre as métricas descritas, Assortatividade e Coeficiente de Colaboração não são em geral calculadas por ferramentas de análise e visualização de redes sociais acadêmicas, sendo este um diferencial deste trabalho.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta o contexto no qual a dissertação está inserida: análise de redes sociais e visualização de grafos que representam redes sociais. As áreas de pesquisa são descritas, identificando as principais necessidades e desafios visando delimitar o escopo do problema tratado na dissertação. Este capítulo está dividido em 2 seções. Na primeira seção, são apresentados trabalhos que envolvem estudos teóricos sobre análise de redes sociais, sendo que alguns destes utilizam programas de geração de grafos para mostrar saídas gráficas. Na segunda seção, são descritos os trabalhos que propõem ferramentas de visualização de grafos e que calculam métricas para análise dos grafos gerados, e é apresentada uma análise comparativa dos trabalhos descritos. Deve-se ressaltar que os trabalhos descritos nesse capítulo consideram redes sociais de forma abrangente (grafos dentro de vários domínios de problema), enquanto a proposta descrita nesta dissertação tem seu foco em redes sociais acadêmicas de coautoria. O objetivo do estudo em abrangência é, principalmente, aplicar a análise e visualização de grafos na solução dos problemas de redes sociais acadêmicas de coautoria.

3.1 Análises de Redes Sociais

O esforço em decifrar as redes sociais e suas características se aplica a diversas áreas de conhecimento. Desta forma, alguns trabalhos desenvolvidos em outras áreas que utilizavam análise de redes sociais foram pesquisados e são apresentados a seguir.

Em 2006, Oliveira Silva et al. (OLIVEIRA SILVA et al., 2006) realizaram um estudo sobre as redes de colaboração entre professores de programas de Pós-Graduação do Brasil na área de Ciência da Informação. As informações para tal estudo vêm de dados da RedeCI ¹. O objetivo do trabalho foi pesquisar sobre correlações entre trabalhos de pesquisa entre diferentes programas de Pós-Graduação no Brasil. Outra questão foi descobrir quais as características dessa rede, se é densa, se existem colaborações entre professores de diferentes linhas de pesquisa. A partir daí foi feito um estudo bibliométrico das publicações verificando quais tinham colaboração e, assim, foram aplicadas medidas de centralidade. Um grafo foi gerado com tais pesquisadores com o au-

¹www.redeci.netic.com.br

xílio da ferramenta de visualização UCINET². Além disso, uma tabela com centralidade de grau, proximidade e intermediação foi gerada com o objetivo de confirmar o destaque de alguns pesquisadores no grafo com respeito às colaborações, mostrando os pesquisadores que são fundamentais para ligações entre algumas universidades. Como resultado, o estudo mostrou baixa cooperação entre os grupos estudados.

No mesmo campo (Ciência da Informação), foi realizado outro trabalho, em busca de evidências comportamentais dos pesquisadores e tendências evolutivas nas redes de coautoria no ramo da ciência da informação no Brasil (BRANDÃO; PARREIRAS; SILVA, 2007). Para o estudo das tendências foram coletados dados referentes a publicações de diversos eventos, revistas, etc, em determinados períodos de tempo e volumes. Já, para avaliar a evolução, foi feita a divisão da coleção em cinco períodos de tempo. A partir de um gráfico construído com os dados obtidos, os autores concluíram inicialmente que o número de publicações com autoria individuais é superior aos trabalhos feitos em coautoria. Porém perceberam que a diferença entre a quantidade de trabalhos nas duas modalidades vem caindo, o que leva a crer que, no futuro, os trabalhos em colaboração devem superar os individuais. Porém, para trabalhos em nível global (da rede como um todo e não somente local quando envolve colaboração dentro das sub-redes) o nível de colaboração vem diminuindo, devido aos autores "novatos" que entram na rede, incorporando baixo índice de colaboração.

Em 2009, Lima relatou um estudo sobre as redes de colaboração ligadas ao grupo da Pós-Graduação na área de Geociências na UFRGS (LIMA, 2009). A principal questão que motivou o trabalho foi estudar quais elementos influenciam na evolução estrutural das redes de colaboração entre pesquisadores pertencentes ao grupo estudado em um determinado período de tempo. Para isso foram construídas três redes de coautoria usando dados da base CAPES em três intervalos de tempo diferentes. Os resultados do estudo mostraram que o conjunto de atores centrais é formado em sua maioria por docentes, não havendo previsão de mudança desse padrão. Além disso, foi verificada reincidência no padrão de parceiros em colaboração. Outra conclusão foi que tanto grupos, linhas e projetos de pesquisa influenciam na evolução estrutural das redes de colaboração do PPGGeo da UFRGS.

Já em 2008, Fátima e Caregnato aplicaram cálculos bibliométricos às redes de colaboração formadas pelos pesquisadores do grupo da Pós-Graduação em Epidemiologia da Universidade Federal de Pelotas (UFPeL) (FÁTIMA S. MAIA; CAREGNATO, 2008). Para tal estudo foram utilizados trabalhos realizados no período de 1991 a 2002. As autoras constataram que os pesquisadores publicaram mais trabalhos em conjunto do que individualmente. Como métodos de avaliação, as autoras mostraram a distribuição de artigos e o número de autores, além de outro que mostra os coeficientes de colaboração dos professores para cada ano dentro do intervalo de tempo já citado anteriormente. Além disso, alguns grafos foram gerados e analisados. A UCINET foi usada para calcular as centralidades de grau, proximidade e intermediação, para assim comparar os pesquisadores entre si. Como trabalhos futuros, as autoras sugeriram analisar as visibilidades dos artigos.

²<http://www.analytictech.com/ucinet/>

Desta forma as publicações internacionais e nacionais podem ser comparadas de acordo com os seus fatores de impacto.

No trabalho de Hayashi, Hayashi e Lima, em 2008, foi feito um estudo das redes de coautoria formadas por pesquisadores da área de Educação Especial ou entre esses pesquisadores e outros do país e do exterior (HAYASHI; HAYASHI; LIMA, 2008). A partir da análise das colaborações foram identificados possíveis colaboradores futuros e o grau de colaboração dos pesquisadores. Para esse estudo, foram analisadas publicações de pesquisadores da Pós-Graduação de Educação Especial da Universidade Federal de São Carlos (UFSCar) no período de 2001 à 2006, sendo os dados extraídos da base Lattes (arquivos 'xml' individuais de cada pesquisador). Os grafos foram gerados também com o auxílio da UCINET, são eles: grafo de artigos, de livros, capítulos de livros e trabalhos completos. A partir desses grafos, foram feitas análises sobre as componentes dos grafos, comparando seus tamanhos, assim como também foram feitas análises sobre os graus dos nós. O objetivo do trabalho era identificar práticas de coautoria na área de Educação Especial.

Como pode ser visto nesta seção, a maioria dos estudos utilizam grafos para ilustrar as redes estudadas, facilitando a visualização para os leitores. A UCINET mostrou-se bastante utilizada tanto para geração dos grafos das redes quanto para calcular métricas de centralidade.

3.2 Ferramentas de Visualização

Para a realização deste trabalho, algumas ferramentas de visualização de redes sociais foram estudadas com o objetivo de identificar suas características e as métricas utilizadas para análises das redes. Assim, as próximas subseções são dedicadas ao detalhamento das ferramentas analisadas, sendo que cada uma é descrita de acordo com a seguinte ordem: idéias gerais e principais características, métricas calculadas e considerações finais. Por fim, uma análise comparativa é apresentada para o fechamento da seção.

3.2.1 COMMATRIX

COMMATRIX (TRIER; BOBRIK, 2009) é um *framework* para análise e visualização de redes sociais de propósito geral, podendo ser utilizado para representar diversos tipos de redes, como: coautorias, colaboração entre organizações ou rastreamento em redes de roteadores. A interface com o usuário permite que seja feito um exame exploratório para identificação e observação de estruturas relevantes na rede, ou seja, através de filtros é possível visualizar apenas estruturas que satisfaçam uma busca em um determinado momento. Além disso, é permitido configurar as variáveis visuais tais como rótulo, tamanho e cor dos nós. É possível visualizar a evolução da rede em um período de tempo passando como parâmetro de entrada a informação do período de tempo desejado. O usuário pode exportar dados para tabelas em formato .csv e .xcl, para registrar a evolução da rede e que permita posterior análise. O leiaute utilizado para visualizar o grafo é o leiaute baseado em forças, podendo ser ajustado de acordo com as preferências do usuário. O *download* da versão *trial* pode ser feito em <http://www.commetrix.de/>.

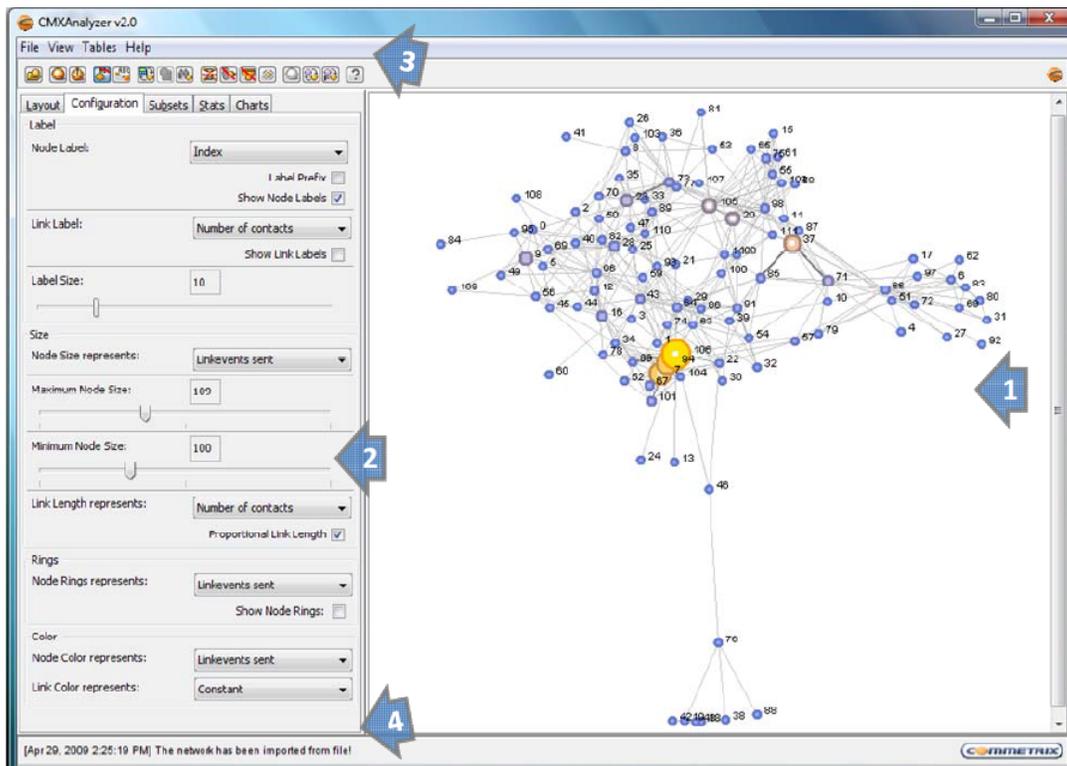


Figura 3.1: Interface da ferramenta COMMETRIX. Fonte: (TRIER; BOBRIK, 2009).

A Figura 3.1 ilustra um exemplo de grafo gerado com a ferramenta COMMETRIX. As setas destacam as principais características da ferramenta. A seta 1 mostra a tela principal da aplicação, na qual o usuário pode marcar, ou desmarcar um nó clicando nele. O nó pode ser reposicionado, selecionando-o e clicando com o botão direito sobre a nova posição desejada. A seta 2 mostra o painel de controle da ferramenta, no qual o usuário pode alterar as características do grafo. Por exemplo, dizer qual característica será usada como rótulo para cada nó, definir o que representa o tamanho dos nós, entre outras. A seta 3 mostra um menu com vários botões que permitem ao usuário carregar dados novos, especificar o tipo de visualização desejada (2D ou 3D), acionar filtros, acessar a ajuda, dentre outras opções. A seta 4 mostra informações sobre o processo atual da operação que está sendo executada, através de uma barra de progresso.

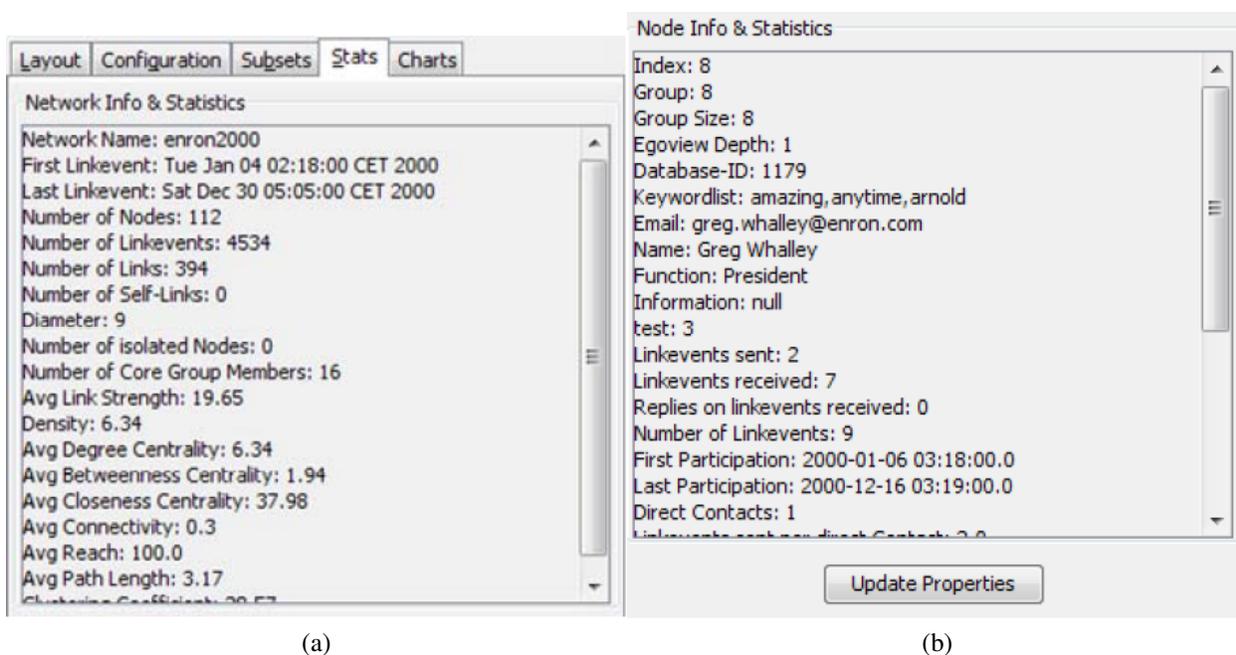


Figura 3.2: Métricas geradas pela COMMETRIX: (a) Métricas gerais da rede e (b) Métricas de um nó específico. Fonte: (TRIER; BOBRIK, 2009).

As métricas calculadas e visualizadas pela COMMETRIX são referentes à rede como um todo ou a um nó específico. As propriedades calculadas para a rede são mostradas na Figura 3.2a, dentre elas: nome da rede, número de nós e relacionamentos, diâmetro do grafo, quantidade de nós isolados (nós de grau zero), quantidade de membros mais ativos, densidade da rede, centralidades, média de alcance, média do tamanho dos caminhos e coeficiente de *clustering*.

Para um determinado nó, as características podem ser vistas na Figura 3.2b. As principais características são: índice, identificador de grupo, tamanho do grupo do qual o nó participa, pro-

fundidade de visão, listas de palavras-chave (três palavras mais usadas para referenciá-lo), atributos do nó (nome, e-mail, etc), *linkevents sent* (quantidade de transações iniciadas/ enviadas por ele), *linkevents received* (quantidade de transações recebidas por ele), alcance do nó, medidas de centralidade, quantidade de contatos diretos, etc.

COMMETRIX mostra-se uma ferramenta eficiente no que tange aos seus objetivos, permitindo grande interação com o usuário. Além disso, o usuário pode ajustar a visualização de diversas formas para deixar o grafo da forma que mais lhe agrada. Pode ser utilizada com diversos tipos de redes, rede de e-mails para verificar a taxa de envio e os nós mais envolvidos dentro de conjunto, ou de uma rede de coautoria. Uma limitação da COMMETRIX é o fato de ser paga, fazendo com que a quantidade de usuários seja restrita. Inclusive, para gerar os arquivos de entrada, é necessária uma outra ferramenta, que exige outra licença para ser utilizada.

3.2.2 COSBILab Graph

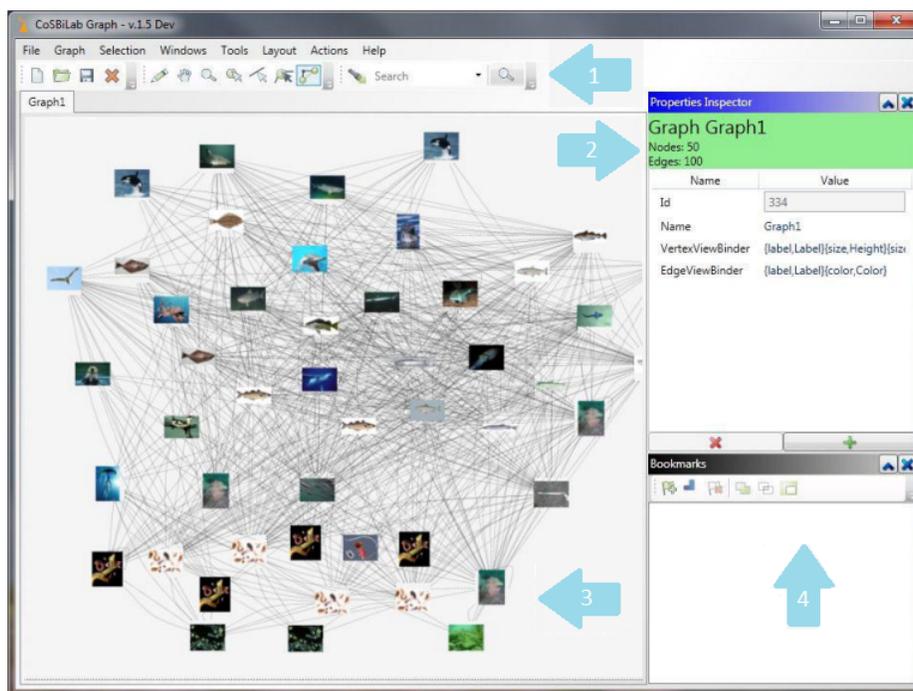


Figura 3.3: Interface da ferramenta COSBILab Graph. Fonte: (VALENTINI; JORDÁN, 2010).

A COSBILab Graph (VALENTINI; JORDÁN, 2010) é uma ferramenta que permite construção, visualização e modificação de grafos. A Figura 3.3 mostra a interface da COSBILab. A seta 1 indica a barra de ferramentas, onde o usuário pode gerar um novo grafo, selecionar os nós, alterar o leiaute (randômico, árvore, espiral, grid, circular, etc), calcular métricas e exportar o grafo. A seta 2 indica uma aba onde é possível visualizar informações de um item selecionado. A seta 3

mostra a tela onde o usuário visualiza e interage com o grafo. O conteúdo da aba indicada pela seta 4 pode variar, mostrando uma tela em que o usuário pode navegar pelo grafo que aparece apontado pela seta 3. Nesta tela aparece um retângulo azul que representa a porção do grafo que está sendo visualizada atualmente pelo usuário; ou ainda um *palette* onde o usuário pode escolher qual o tipo de item que será inserido, por exemplo, se é um nó com ou sem arestas. A COSBILab Graph pode ser obtida gratuitamente para uso não-comercial no site <http://www.cosbi.eu/>.

Por padrão, ao criar os componentes básicos do grafo e o grafo em si, são criadas estruturas (nós, arestas e o grafo) que podem ser alteradas pelo usuário, adicionando ou removendo atributos aos componentes, como pode ser visto na Figura 3.4. Assim, a COSBILab Graph pode ser aplicada em qualquer tipo de rede, sendo adaptada de acordo com o interesse do usuário. Além disso, pode-se alterar atributos gráficos de um nó ou do grafo, por exemplo, rótulo, cor, largura, altura, largura das arestas. A COSBILab Graph também permite que nós possam ser agrupados para melhorar a visualização do grafo.

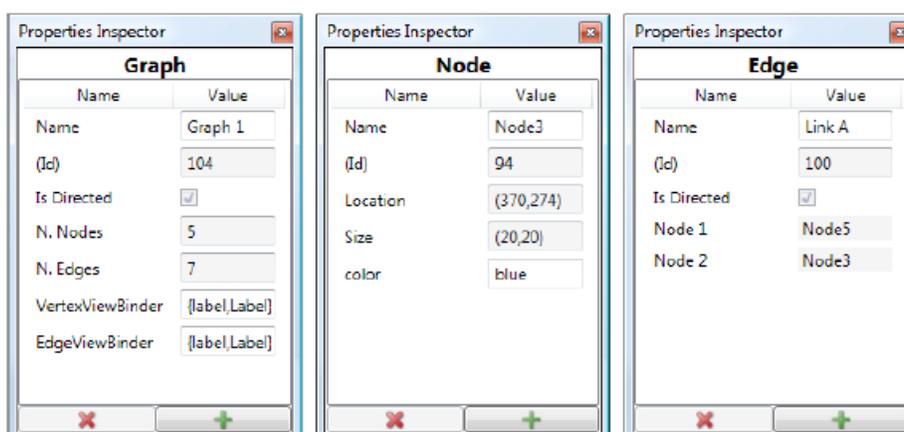


Figura 3.4: Características básicas das estruturas em COSBILab Graph. Fonte: (VALENTINI; JORDÁN, 2010).

As principais métricas calculadas pela COSBILab Graph são as relativas ao grafo: se é ou não direcionado, se possui ciclos, árvores ou florestas, número de arestas e nós, quantidade de subgrafos, coeficiente de *clustering*, medidas de centralidade, quantidade de ciclos, número de componentes do grafo, grau dos nós, etc.

Além de criar um grafo manualmente, o usuário pode escolher importar arquivos em cinco formatos diferentes, incluindo saídas geradas pelo BlenX Environment, DOT GraphViz e UciNet DL. A saída além de ser gerada de forma gráfica para o usuário, também pode ser exportada para '.png'.

Uma das vantagens de utilizar a COSBILab Graph é sua generalidade pois pode ser aplicada a diversos tipos de grafo e a diversas áreas.

3.2.3 SocialAction

SocialAction é uma ferramenta de análise criada para testar o poder de integração entre cálculos de métricas e visualização de redes sociais (PERER; SHNEIDERMAN, 2008). A contribuição sugerida pelos autores da SocialAction é mostrar que essa integração melhora a análise exploratória de dados e que sua metodologia de avaliação para estudos de longo prazo caso captam as estratégias de investigação de analistas de dados. Com as métricas é possível simplificar a compreensão de algumas visualizações complexas, já que permite que o usuário mantenha o foco nos nós que são mais relevantes para ele. A implementação da SocialAction utiliza a biblioteca Prefuse para geração dos grafos.

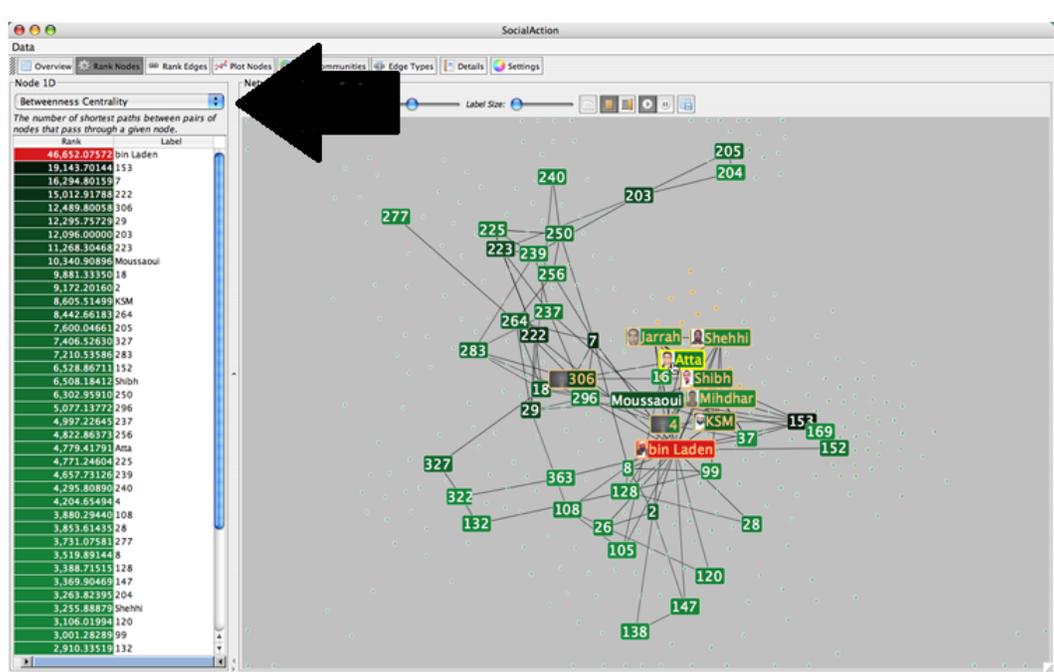


Figura 3.5: SocialAction: Mudança de rótulos dos nós de acordo com as métricas de análise de redes sociais. Fonte: (PERER; SHNEIDERMAN, 2008).

SocialAction permite que o usuário escolha uma dentre as métricas mais comuns de análise de redes sociais (centralidades de grau, proximidade e intermediação, etc) e ordene-a para usar como rótulos dos nós. Além disso, os nós também recebem uma coloração dentro de uma faixa de acordo com seu nível no *ranking* (verde - baixo *ranking*, preto - médio *ranking* ou vermelho - alto *ranking*). Na Figura 3.5, a seta indica o local onde o usuário pode modificar a métrica utilizada para colorir os nós. A SocialAction também permite arrastar e dar *zoom* no grafo. Os rótulos dos nós estão sempre presentes, sendo que o usuário pode alterar o tamanho da fonte e o comprimento dos rótulos. Ao selecionar um nó, o usuário pode visualizar informações sobre

ele. Além disso, ao deixar o *mouse* sobre o nó, os nós vizinhos são destacados alcançando o *NetViz Nirvana* (capacidade de visualizar para cada nó suas arestas e percorrê-las até encontrar cada nó vizinho) para aquele nó. Outras métricas e características visuais estão disponíveis na *SocialAction*, como “ranqueamento” das arestas para encontrar relacionamentos fortes, habilitar algoritmos de agrupamento, analisar diferentes tipos de relacionamento.

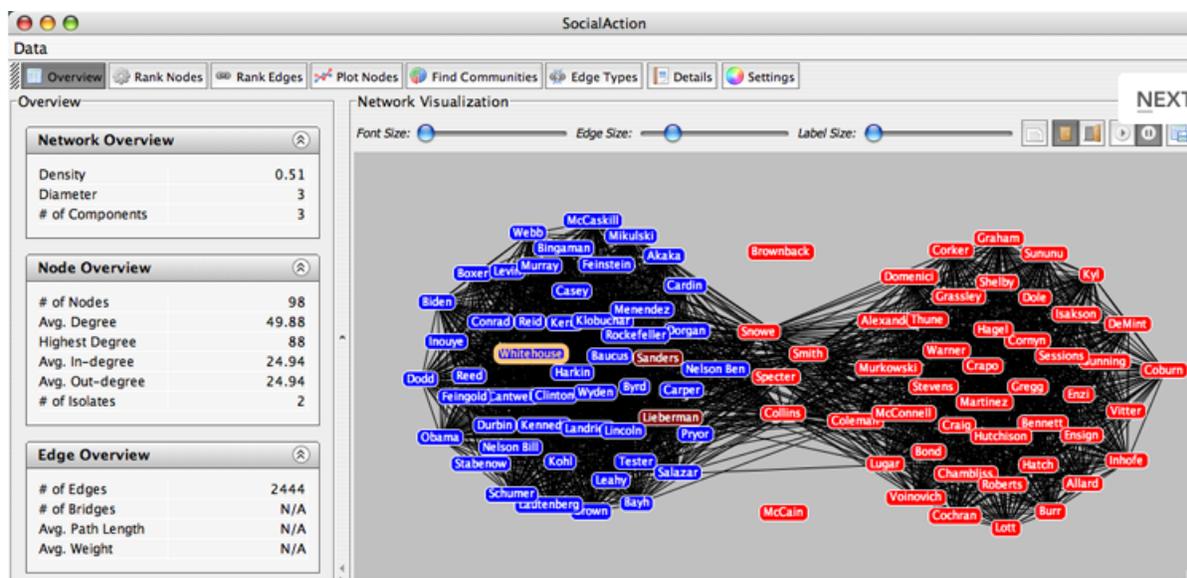


Figura 3.6: Grafo mostrando dois grupos de senadores americanos e a frequência de votação em conjunto. Fonte: (PERER; SHNEIDERMAN, 2008).

Como estudo da efetividade da ferramenta, foi desenvolvida uma metodologia que envolve quatro passos: entrevista, treinamento, utilização da ferramenta, resultado. Os terceiro e quarto passos envolvem a utilização da ferramenta. Para validação do trabalho foram desenvolvidos quatro estudos de caso que foram executados com diversos tipos de usuários. A Figura 3.6 mostra um dos estudos de caso onde as cores azul e vermelha determinam a que grupo os senadores participam, e as arestas representam a frequência em que dois senadores votaram juntos. A *SocialAction* permite que sejam gerados gráficos que comparam duas métricas, por exemplo, grau e centralidade de intermediação.

Como vantagem em utilizar a *SocialAction* pode-se apontar o fato de se aplicar a vários tipos de rede. Já uma das desvantagens é a impossibilidade de exploração de dados durante determinados períodos de tempo, como dias, meses e anos.

3.2.4 SocialNetSense

Gou et al. propôs um *framework* que permite integração entre atributos de redes sociais para entendê-las estrutural e socialmente (GOU et al., 2012). Os autores consideraram que os trabalhos

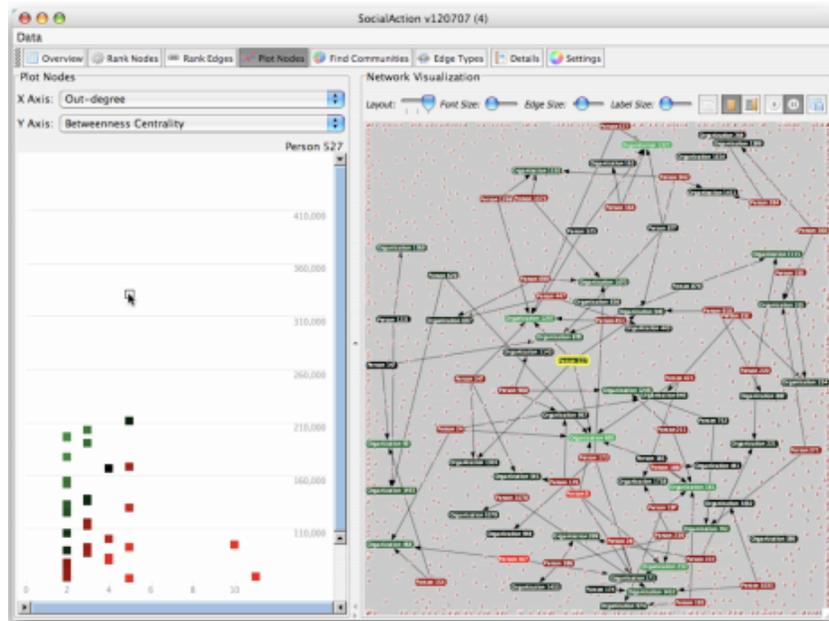


Figura 3.7: Gráficos gerados com a SocialAction que permitem que o usuário compare duas métricas. Fonte: (PERER; SHNEIDERMAN, 2008).

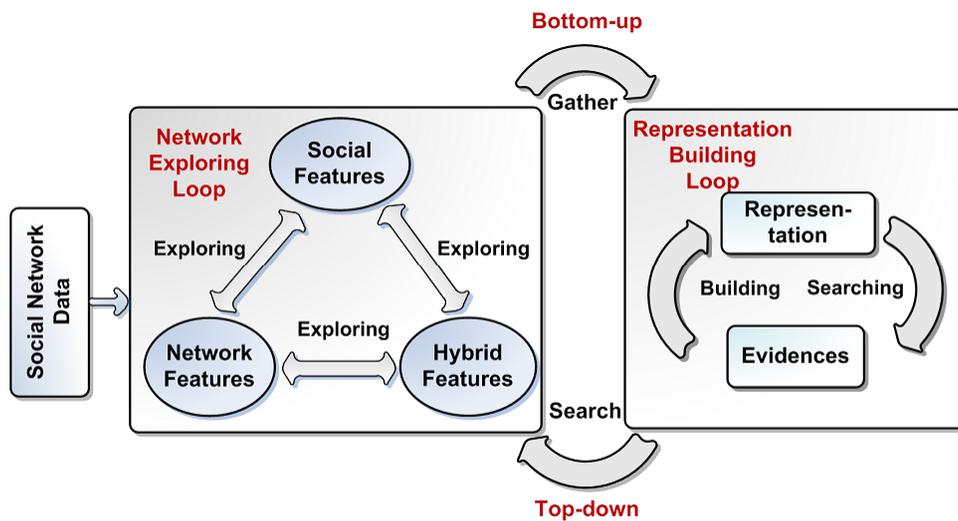


Figura 3.8: Esquema proposto para o *framework* SocialNetSense. Fonte: (GOU et al., 2012).

atualmente desenvolvidos apresentam fraco suporte para redes altamente complexas e normalmente precisam da ajuda do usuário para reunir, sintetizar e organizar informações das características extraídas da rede. Além do *framework*, foi desenvolvido um protótipo, SocialNetSense que se baseia na proposta deste *framework* para permitir visualização e análise de redes sociais juntamente com interação com o usuário.

A Figura 3.8, mostra o esquema do framework proposto por Gou et al.. O esquema é dividido em três partes: dados da rede social, *loop* de exploração de rede e *loop* de construção da representação. Inicialmente, os dados são processados e inseridos em estruturas de dados de acordo com suas características sociais, de rede e híbridas. No *loop* de exploração, os usuários procuram informações desejáveis explorando as características disponíveis nessa fase. As características sociais se referem aos atributos sociais dos atores como classificação na hierarquia da rede. As características de rede estão relacionadas à estrutura da rede. Já as características híbridas mostram o impacto dos atributos sociais sobre a estrutura da rede. Um outro *loop* externo une os apresentados anteriormente, são os processos *top-down* e *bottom-up*. No primeiro, o usuário parte das representações para orientar suas pesquisas no *loop* de exploração. No *bottom-up*, o usuário reúne informações de interesse para dar suporte à sua representação.

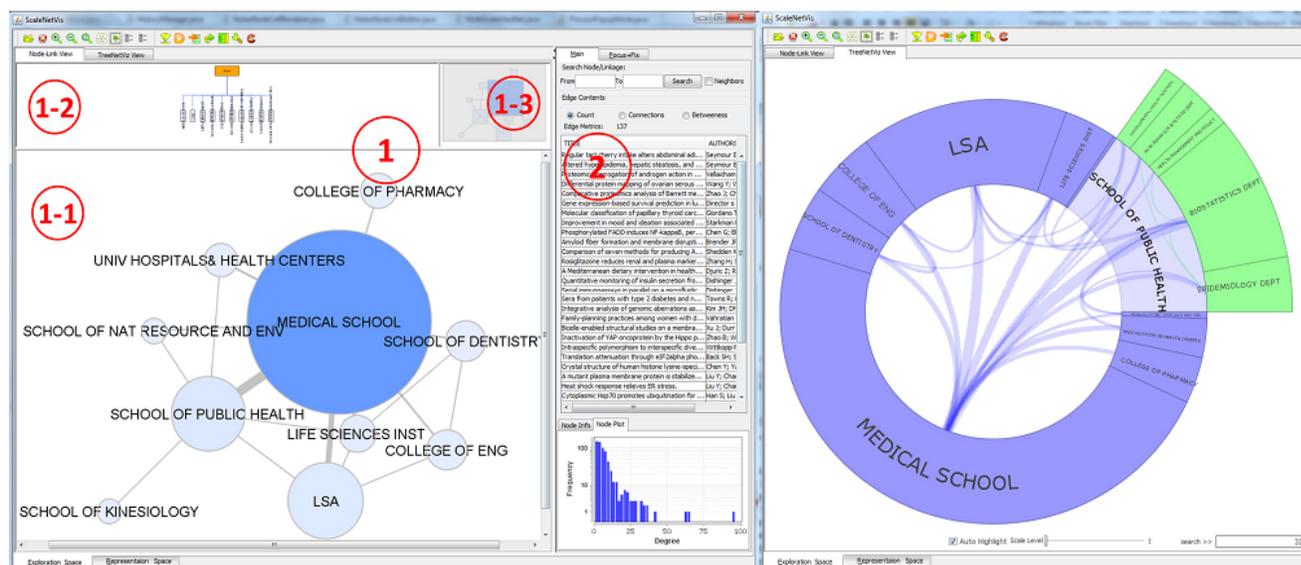


Figura 3.9: Representação visual do SocialNetSense. Fonte: (GOU et al., 2012).

A partir das idéias propostas foi construído um protótipo que permite exploração de características sociais, exploração e análises de características de rede, exploração de características híbridas, construção da representação e gestão de processos que aumentam a conscientização dos usuários em relação ao histórico de ações realizadas por eles. A Figura 3.9, mostra a interface do *Network Exploring Sense* (NES). O Painel 1 mostra a visualização da rede social, o Painel 2

mostra o painel de controle onde estão presentes ferramentas analíticas. O painel 1-1 permite a visualização do grafo da rede; o 1-2 contém uma árvore que representa a hierarquia entre os nós; e o painel 1-3 mostra uma visão geral da rede e permite rápidas navegações na rede. Além do NES, o TreeNetViz (GOU; ZHANG, 2011), mostrado na Figura 3.9, ao lado direito, foi utilizado para mostrar características híbridas de padrões de rede sobre a hierarquia social. Na visualização do TreeNetViz é possível expandir ou agrupar setores ou redimensionar setores.

A partir dos resultados obtidos com o NES, o RBS (*Representation Building Space*) organiza as evidências e estabelece relações entre elas. A Figura 3.10 mostra o painel 1 que contém o *workspace* principal. No painel 2, são mostrados históricos das execuções do usuário guardando o *timestamp* de cada ação. E no painel 3 é mostrada a lista dos elementos presentes no *workspace*. Além do *timestamp*, o usuário pode adicionar notas ou agrupar os elementos selecionados. Através do histórico, o usuário pode escolher voltar para um passo intermediário sem precisar reiniciar todo processo do início.

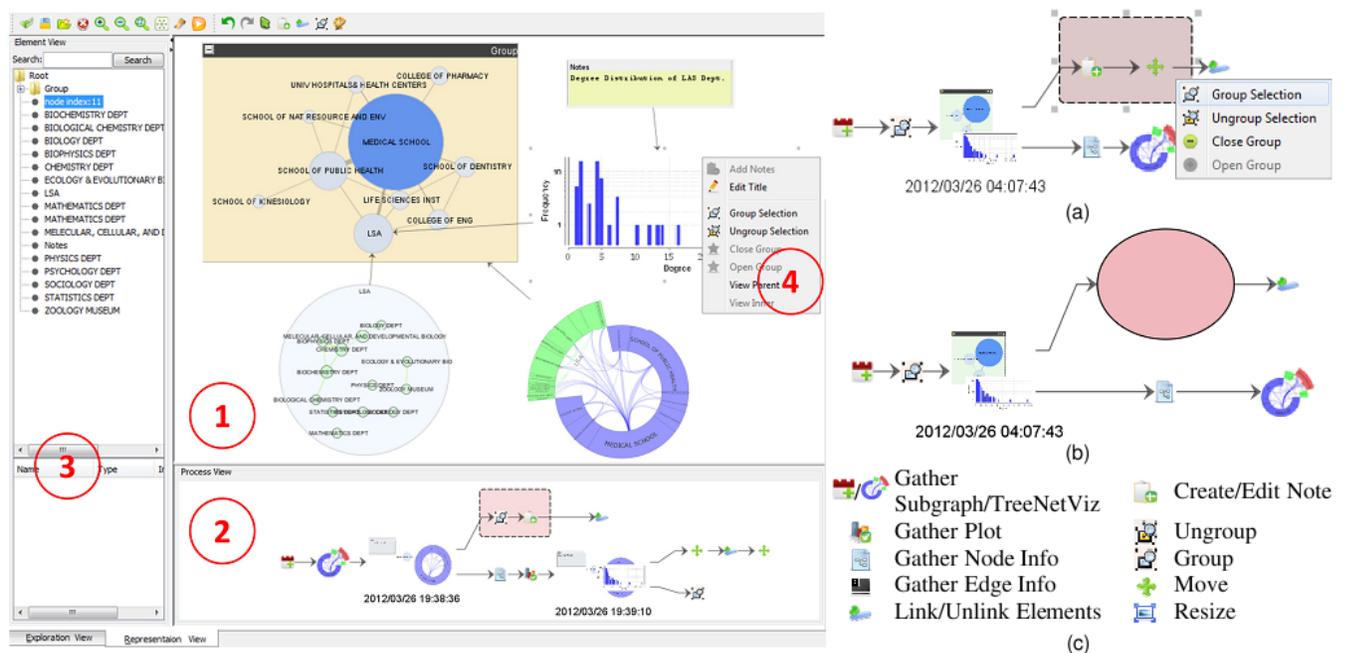


Figura 3.10: *Interface* do RBS. Fonte: (GOU et al., 2012).

Apesar de apresentar grande quantidade de informações para o usuário, isso pode ser um ponto negativo para um usuário leigo ou que queira tirar simples conclusões sobre a rede visualizada. Algumas limitações do SocialNetSense se referem ao fato de funcionar bem com dados incompletos, prejudicando a construção de redes e o suporte de exploração social. Além disso, o armazenamento das ações que formam o gráfico com o histórico possui apenas informações contidas na visualização no momento em que o usuário grava a ação, ou seja, pode perder informações que não estejam

visíveis apenas pelo tamanho da tela, por exemplo.

3.2.5 ArnetMiner

ArnetMiner (TANG et al., 2008) é uma ferramenta que também trabalha com mineração e extração de dados aplicadas às redes sociais acadêmicas. Inicialmente, são extraídos perfis dos pesquisadores automaticamente da *Web*. Em seguida, é feita a integração dos dados referentes às publicações existentes em bibliotecas digitais e das páginas pessoais dos pesquisadores. Após o tratamento da redundância de dados, é feita a modelagem da rede acadêmica envolvendo os pesquisadores extraídos da consulta inicial juntamente com suas informações, ou seja, seus colaboradores e suas publicações. Por fim, são oferecidos serviços de busca dentro da rede (TANG et al., 2008). A Figura 3.11 mostra o exemplo de um grafo de coautoria entre dois pesquisadores (Andrew Chih Yao e John E. Hopcroft) e seus colaboradores, onde podem ser vistos os colaboradores em comum. Para utilizar a ArnetMiner basta acessar o site <http://arnetminer.org/>.

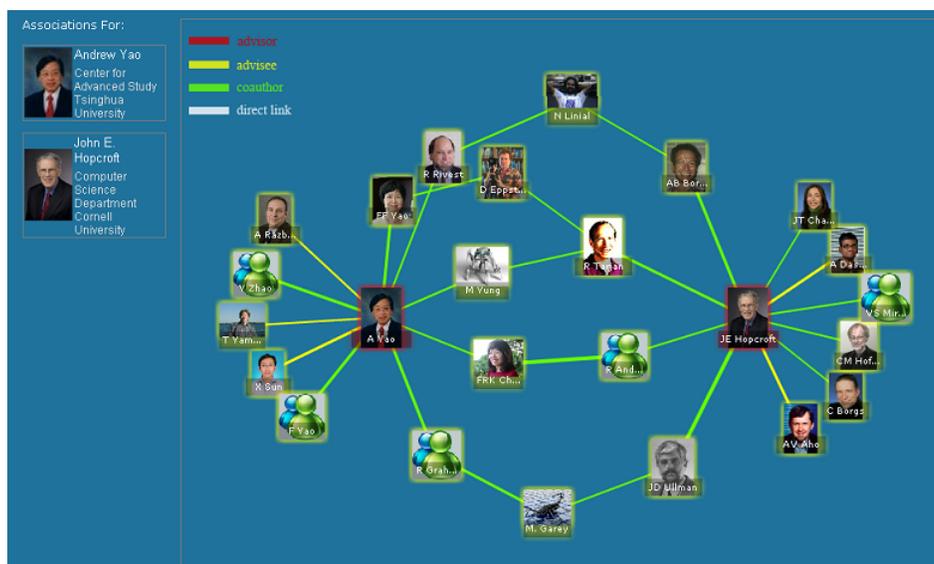


Figura 3.11: Grafo de coautoria gerado pela ArnetMiner formado por dois pesquisadores escolhidos e com seus colaboradores. Fonte: <http://arnetminer.org>.

Além dos grafos de coautoria gerados pela ArnetMiner, também é possível fazer buscas como: busca por um perfil de usuário criado pelo sistema a partir de técnicas de extração de informação; busca de especialista, podendo ser feita a partir de um nome de pesquisador ou a partir do nome de uma área específica ou qualquer outro campo. Então, dependendo da consulta digitada pelo usuário a busca retornará os especialistas naquela área; análise de conferência, onde o usuário consegue buscar pelos pesquisadores que são mais ativos num determinado evento; busca por disciplinas, onde a partir de uma consulta por uma área, por exemplo, mineração de dados, o sistema retorna

quem está ministrando as disciplinas mais relevantes para essa área. O usuário também pode criar um *login* para alterar perfis extraídos, fornecer *feedbacks* para os resultados das pesquisas, acompanhar pesquisadores da ArnetMiner e criar uma página para ele. A Figura 3.12 mostra a interface da ArnetMiner onde o usuário pode interagir com o sistema.

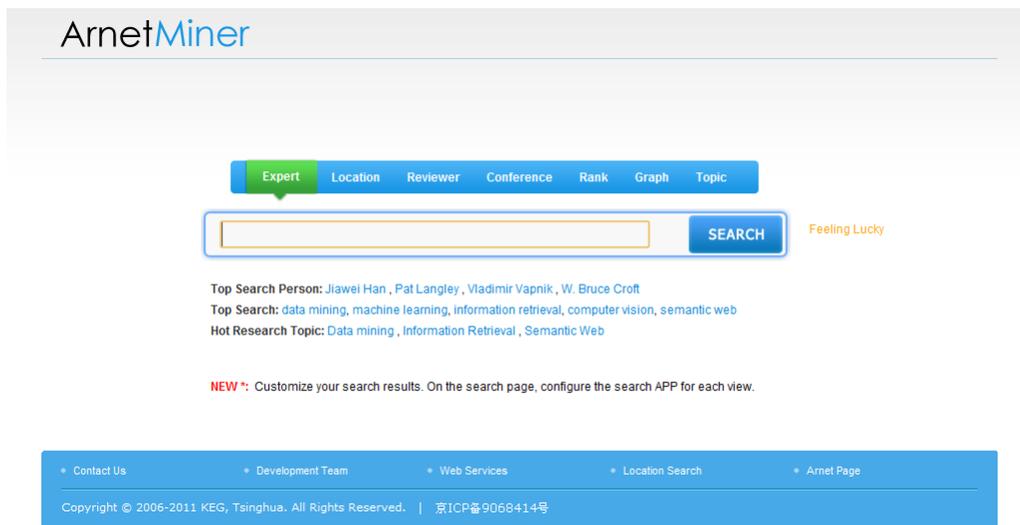


Figura 3.12: Interface da ferramenta online ArnetMiner. Fonte:<http://arnetminer.org/>.

As métricas calculadas pela ArnetMiner são: número total de citações e publicações, H-index, longevidade (tamanho da vida acadêmica de um pesquisador), fator de impacto (reflete a importância do local da publicação), atividade (baseada nos artigos publicados nos últimos anos pelo pesquisador, considerando a importância de cada artigo), diversidade (calculada de acordo com as áreas em que o pesquisador trabalha), sociabilidade (definido pela quantidade de coautores do pesquisador), *uptrend* (métrica utilizada para definir o grau de crescimento de um pesquisador) e *New star* (tem o *score* baseado na atividade do pesquisador).

ArnetMiner é uma ferramenta muito indicada para trabalhar com redes sociais acadêmicas, já que calcula diversas métricas que auxiliam na avaliação da produção de pesquisadores. Além disso, as formas de interação com o grafo são intuitivas e a interface amigável para o usuário. A maior desvantagem em utilizar a ArnetMiner é a necessidade de estar conectado à internet para buscar os dados para formar as redes. Do mesmo modo não é possível passar um arquivo de entrada específico para representar uma rede específica. Os dados, assim, não são completos, restringindo-se aos coletados pela própria ferramenta.

3.2.6 igraph

A igraph (CSARDI; NEPUSZ, 2006) é uma biblioteca para criação e manipulação de grafos não-direcionados e direcionados. Essa biblioteca pode ser incluída em implementações desen-

volvidas em linguagem R ou Python, para auxiliar no cálculo de métricas em análises de grafos. A Figura 3.13 mostra um grafo de uma rede gerada com linguagem R juntamente com a biblioteca igraph. A caixa apontada pela seta mostra alguns leiautes (randômico, circular e leiautes baseados em forças) possíveis para visualização do grafo. Nesta imagem, o tamanho de cada nó está relacionado ao seu grau. O *download* da igraph pode ser feito em: <http://igraph.sourceforge.net/download.html>.

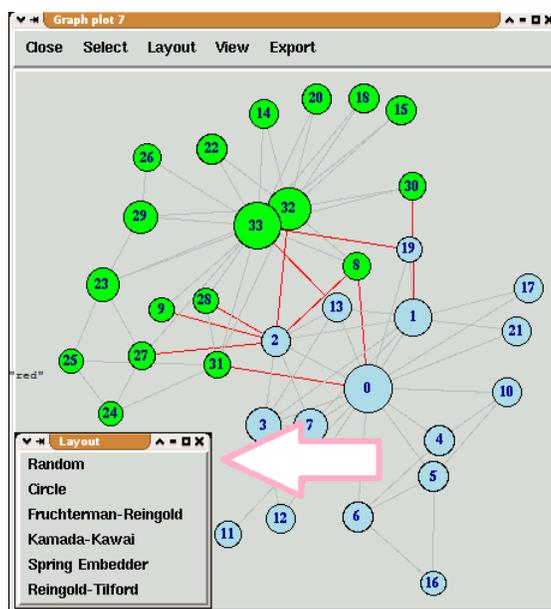


Figura 3.13: Interface criada utilizando a linguagem R que possui suporte para a biblioteca igraph. Fonte: <http://igraph.sourceforge.net/screenshots.html>.

Por se tratar de uma biblioteca, a igraph pode ser aplicada a qualquer campo de estudo que o usuário deseje, sendo necessário apenas adaptar as estruturas de dados que o usuário utilizará para representar a rede social. O arquivo de entrada pode ser qualquer tipo de arquivo que usuário deseje desde que programe corretamente para preencher as estruturas de dados internas implementadas na biblioteca.

A igraph implementa muitos algoritmos que envolvem problemas clássicos de teoria dos grafos, tais como fluxo na rede e árvores geradoras mínimas. As métricas calculadas pela igraph incluem caminhos mínimos (usando algoritmo de Dijkstra), diâmetro, quantidade de vizinhos, subcomponentes do grafo, cálculo de componentes biconectadas, medidas de centralidade, cálculo da densidade, cliques do grafo, entre outras.

Uma das maiores vantagens em utilizar essa biblioteca é a opção de adaptar o grafo de acordo com o problema aumentando as possibilidades de aplicação da biblioteca. Como desvantagem, pode-se dizer que o usuário precisa criar uma interface para que possa haver interação com o grafo, já que a biblioteca gera somente o grafo, permitindo análise sobre este grafo.

3.2.7 Pajek

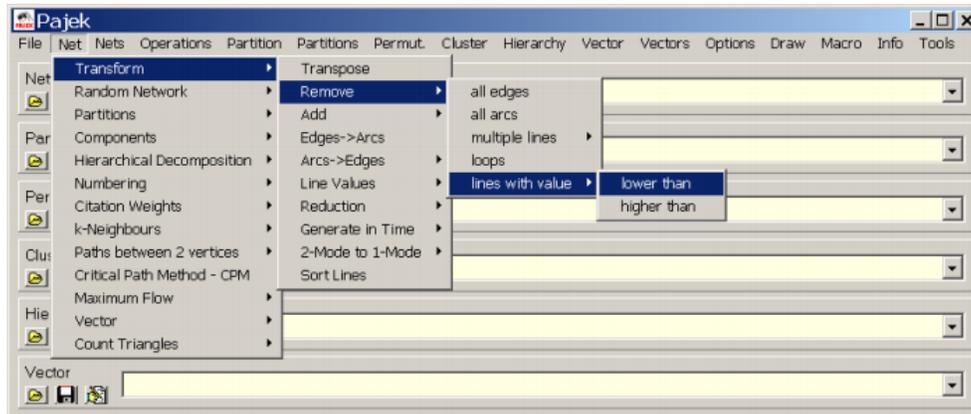


Figura 3.14: Interface inicial da ferramenta Pajek. Fonte: (BATAGELJ; MRVAR, 2011).

Pajek (BATAGELJ; MRVAR, 2011) é uma ferramenta que foi desenvolvida para plataforma Windows que permite visualizar e analisar redes em geral com milhões de nós (redes de colaboração, organização de moléculas, genealogias, redes de *Internet*, etc). Além de ter suporte a grafos direcionados, não-direcionados e mistos, também é possível trabalhar com grafo multi-relacionais, redes de duas modas (grafos bipartites formado entre conjunto de nós disjuntos) e grafos temporais (que variam em determinados períodos de tempo). A Figura 3.14 mostra a janela inicial da Pajek, que está disponível de forma livre para uso não-comercial no site: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.

Pajek usa até seis tipos de objetos de dados:

- Redes (*Networks*) - são os dados principais e contém as arestas e nós. Podem ser representados como aresta, como listas de arestas, formato de matriz ou formato UCINET, GEDCOM. Extensão padrão é .net;
- Partições (*Partitions*) - determina a qual partição cada nó pertence. Extensão padrão é .clu;
- Permutações (*Permutations*) - permite reordenar os nós. Extensão padrão é: .per;
- Agrupamentos (*Clusters*) - subconjunto de nós. Extensão padrão: .cls;
- Hierarquias (*Hierarchies*) - nós ordenados hierarquicamente. Extensão padrão: .hie;
- Vetores (*Vectors*) - para uma determinada propriedade numérica, eles guardam o valor para cada nó do grafo. Extensão padrão: .vec;

Para cada um desses tipos de arquivo, a Pajek consegue lê-los e salvá-los no formato ASCII, sendo que é possível exportar matrizes no formato EPS. Também é possível ler arquivos no formato Pajek .paj, que podem conter os seis tipos de dados. Um exemplo de grafo gerado usando o Pajek pode ser visto na Figura 3.15.

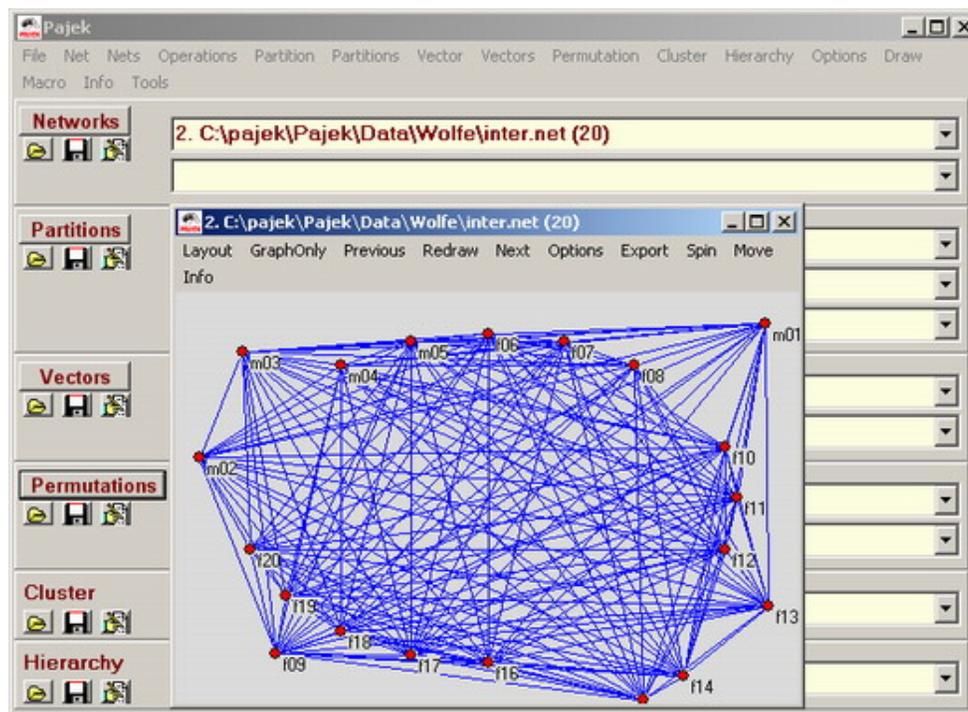


Figura 3.15: Um grafo gerado usando a Pajek. Fonte: <http://imasters.com.br/artigo/4177/sql-server/software-livres-relacionados-a-banco-de-dados>.

A partir das estruturas básicas (nó e aresta) foram implementados algoritmos para calcular algumas métricas, como grau, profundidade, p-cliques, componentes fortes, fracas ou biconectadas. Também é possível calcular caminhos entre os nós do grafo, fluxo máximo entre dois nós, k-vizinhos, extrair subredes, ordenação topológica, remover laços, etc.

Uma das limitações da ferramenta Pajek é ter sido desenvolvida apenas para Windows, não podendo ser utilizada por usuários Linux e Mac OS por exemplo. Por outro lado, é possível aplicá-la a redes com grandes quantidades de nós, sendo essa uma limitação de muitas outras ferramentas.

3.2.8 UCINET

UCINET (BORGATTI; FREEMAN, 2002) é um pacote para análise de redes sociais. A Figura 3.16 mostra a interface do UCINET. Ao clicar no botão indicado pela seta da caixa 1 é possível modificar o diretório padrão, cujo caminho é indicado pela caixa 2. A caixa 3 indica um botão que referencia um editor de texto. A caixa 4 aponta para o editor de tabelas do projeto. A caixa

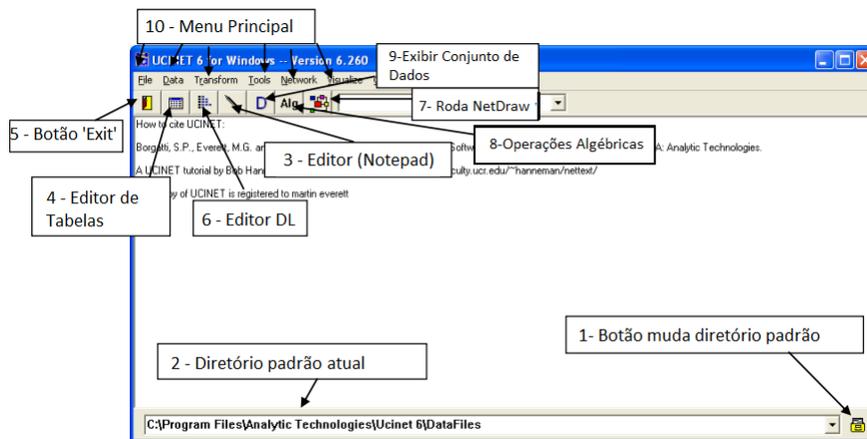


Figura 3.16: Interface da ferramenta UCINET. Fonte: <http://www.analytictech.com/ucinet>.

5 indica o botão de saída do UCINET. A caixa 6 indica o botão para o editor de formato DL, que é um dos formatos mais utilizados pelo pacote UCINET. A caixa 7 indica um botão para rodar o NetDraw³, que associado ao UCINET gera a visualização da rede em forma de grafo. A Figura 3.18 mostra um grafo sendo exibido pelo Netdraw depois de clicar no botão referente a esse aplicativo e escolher o arquivo. Depois disso, é possível interagir com o grafo e alterar características dos nós, arestas e do grafo em geral. A caixa 8 indica um botão que permite que operações algébricas sejam executadas sobre o conjunto de dados de entrada e saída do UCINET. A caixa 9 mostra o botão que permite exibir arquivos gerados pelo UCINET depois de rodar alguma entrada. Por exemplo, um arquivo contendo as medidas de centralidade calculadas para um grafo, como pode ser visto na Figura 3.17. A caixa 10 aponta para o Menu Principal do UCINET, que permite acessar todas as rotinas do UCINET, tais como: *File*, *Data*, *Transform*, *Tools*, *Network*, *Visualize*, *Options* e *Help*. A versão *trial* pode ser obtida em <http://www.analytictech.com/ucinet/download.htm>.

Com a UCINET é possível trabalhar com grafos que possuem grande quantidade de nós, cerca de 10000, ainda de forma eficiente. Além disso, a UCINET permite exportação dos dados para outras ferramentas, por exemplo, Pajek e NodeXL.

As análises permitidas pelo UCINET são: detecção de cliques (subconjunto de nós, no qual a cada dois nós pertencentes a ele, existe uma aresta que conecta os dois nós) no grafo; *clustering* hierárquico, que cria uma matriz da distância geodésica; centralidades de grau, proximidade, intermediação e *eigenvector* dos nós; construção de uma tabela correlacionando as distâncias, para logo depois fazer um ranking dos nós comparando as distâncias; centralidade da rede; densidade de uma rede egocêntrica.

Como vantagem pode-se apontar a possibilidade de criar grafos que representem qualquer tipo de rede social, aumentando a gama de aplicações possíveis. Uma desvantagem é a utilização que

³<http://www.analytictech.com/Netdraw/netdraw.htm>

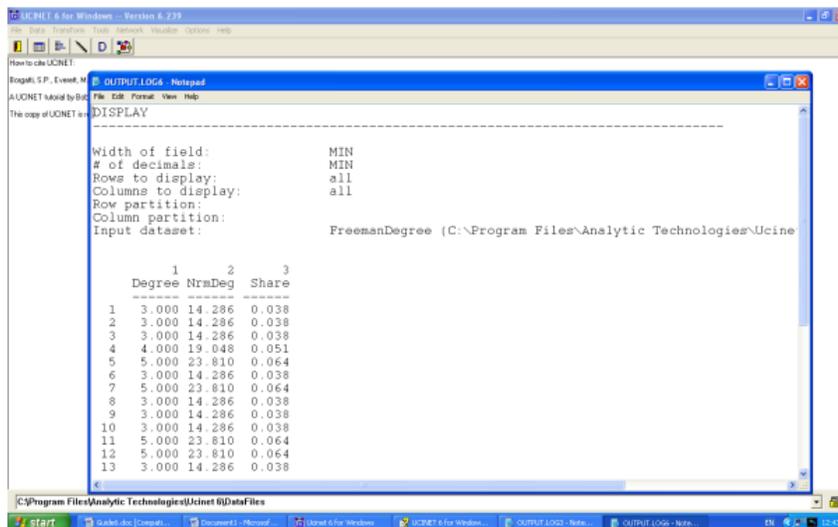


Figura 3.17: Exemplo de arquivo FreemanDegree, um dos arquivos de saída gerado UCINET.
 Fonte: <http://www.analytictech.com/ucinet>.

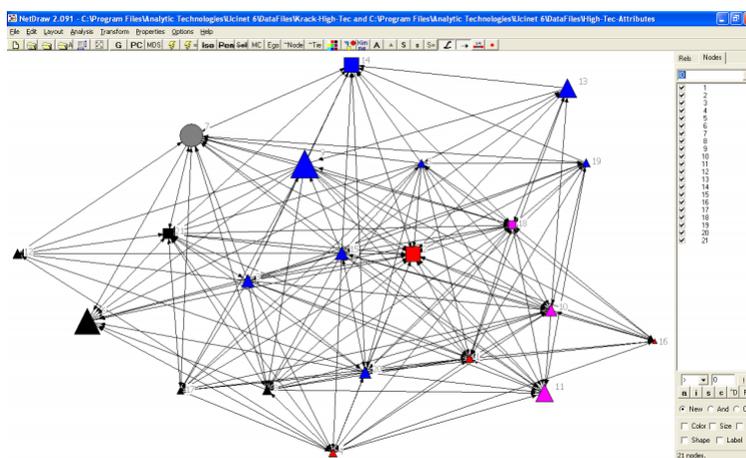


Figura 3.18: Grafo criado com o NetDraw. Fonte: <http://www.analytictech.com/ucinet>.

só pode ser feita na plataforma Windows.

3.2.9 NodeXL

O NodeXL (SMITH et al., 2009) é um *template* desenvolvido para trabalhar com Excel 2007 e 2010, com o intuito de ajudar o usuário a explorar grafos de maneira mais fácil. O principal objetivo do NodeXL é facilitar a usabilidade de forma que não seja necessário programar ou ser profissional em manipulação de grafos e seus dados. Para gerar os grafos, basta passar como entrada uma lista de nós e arestas, ou simplesmente passar uma lista de arestas numa tabela. O NodeXL permite que o usuário possa alterar algumas características dos seus componentes, tais como: largura, cor, nível de transparência das arestas, ou ainda cor, tamanho, posição e até as imagens que representam os nós. O *download* do NodeXL pode ser feito em: <http://nodexl.codeplex.com/releases/view/83654>.

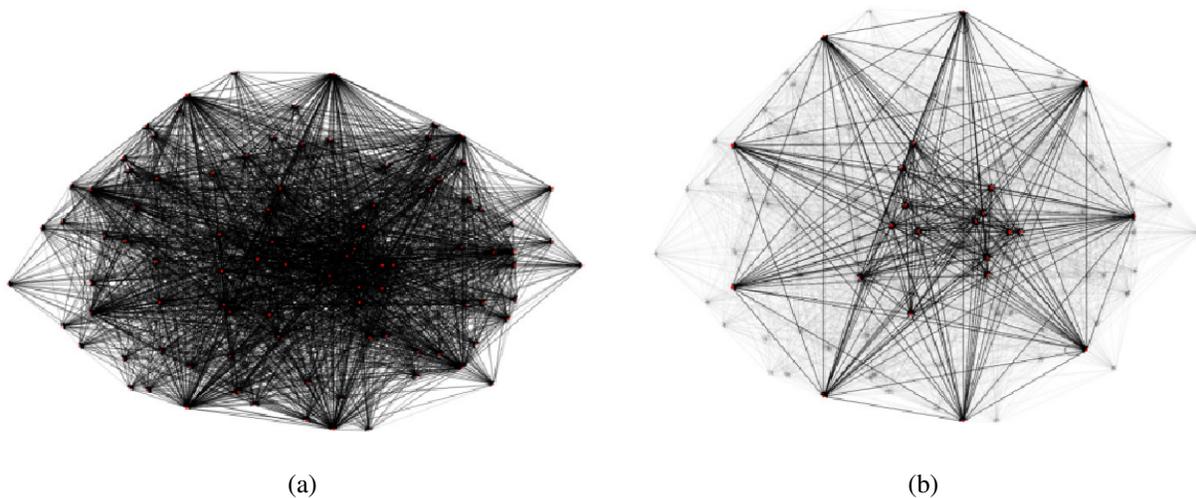


Figura 3.19: Grafos gerados com o nodexl: (a) Grafo Completo e (b) Grafo Filtrado. Fonte: (SMITH et al., 2009).

A Figura 3.19a mostra um exemplo de grafo com dados extraídos de um aplicativo de rede social usado por uma empresa de médio porte. Como pode-se perceber, o grafo parece bastante ilegível, sendo que a rede possui densidade de 26% de todas as arestas possíveis no grafo. Para lidar com esse problema, o NodeXL permite que sejam aplicados filtros, ajustando *sliders* com restrições, por exemplo, mostrar apenas nós com grau maior que dois, como pode ser visto na Figura 3.19b. Pode-se também destacar e diferenciar nós de acordo com suas características, por exemplo, mudar o tamanho do nó de acordo com a quantidade de arestas incidentes. É possível também agrupar nós selecionando um conjunto deles, sendo que o grupo formado tem algumas propriedades como: podem ser expandidos e agrupados, podem ser operados como um conjunto, as

métricas da rede podem ser calculadas para cada grupo. O usuário pode criar os grupos a partir de algum atributo dos nós, como pode ser visto na Figura 3.20, ou seja, escolhendo uma das colunas da tabela para agrupar os nós. Outra forma de agrupar a partir das componentes fortemente conectadas encontradas pelo próprio NodeXL, onde, para cada grupo, todos os nós pertencentes a ele são representados por uma mesma cor. A última forma de agrupamento ocorre através de algoritmos de agrupamento, sendo que o NodeXL dispõe de três algoritmos de agrupamento da biblioteca SNAP⁴: Wakita and Tsurumi, Girvan-Newman ou Clauset-Newman-Moore. Em geral, esses algoritmos tentam posicionar os nós analisando os conjuntos de nós mais fortemente conectados e separando-os em grupos. O NodeXL permite interações como *zoom*, aumentar/diminuir a escala, atribuir cor, opacidade, tamanho, rótulos aos nós e arestas, arrastar nós, além de permitir alterar os leiautes que variam entre algoritmos que implementam o leiaute baseado em forças.

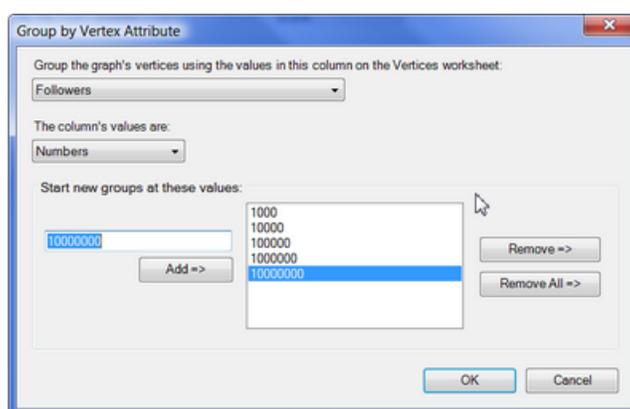


Figura 3.20: Criação de uma restrição para gerar grupo de nós de acordo com seus atributos. Fonte: <http://nodexl.codeplex.com/documentation>.

As principais métricas calculadas pelo NodeXL são: grau de entrada e saída de um nó (quantidade de arestas incidentes e ascendentes do nó), coeficiente de *clustering*, centralidades *closeness*, *betweenness* e *eigenvector*. Outras métricas podem ser adicionadas por usuários avançados, criando uma nova coluna na tabela.

O NodeXL é pago, uma desvantagem. Após o *trial* de noventa dias, é necessário pagar taxas de acordo com o propósito do usuário. Além disso, por ser um template desenvolvido para Excel, só pode ser utilizado no Windows. O NodeXL permite importar e exportar dados nos formatos: GraphML, Pajek, UCINET e matrizes, sendo essa uma vantagem em usar o NodeXL.

3.2.10 Estudo Comparativo

O objetivo desta seção é analisar criticamente as características de cada ferramenta descrita na seção anterior.

⁴<http://snap.stanford.edu>

A Tabela 3.1 apresenta uma síntese de características julgadas importantes para a visualização e interação de redes sociais e que podem, ou não, ser encontradas nas ferramentas analisadas. Foi analisada a disponibilidade dos seguintes recursos:

- **Zoom** - permite afastar ou aproximar um nó sem perder o foco;
- **Distância mínima entre nós** - permite mensurar a distância mínima entre os nós, para mantê-los afastados;
- **Comprimento da aresta** - permite que o usuário ajuste o comprimento das arestas;
- **Alterar os rótulos** - permite que o usuário configure os rótulos, dizendo o que deseja visualizar como rótulo;
- **Alterar o tamanho dos rótulos dos nós** - permitem alterar o tamanho dos rótulos dos nós;
- **Alterar o tamanho dos nós** - permite modificar o tamanho dos nós;
- **Alterar a largura das arestas** - permite que o usuário modifique a largura das arestas, seja configurando a partir do arquivo de entrada ou de forma gráfica;
- **Filtrar** - permite que o usuário crie filtros para limitar a visualização apenas para os nós e arestas que possuam determinadas características;
- **Consultar** - permite que o usuário faça consultas sobre algum nó no grafo, sejam consultas por um nó ou criadas pelo próprio usuário;
- **Expandir** - permite que o usuário expanda o grafo. A partir de cliques nos nós ou através de algum mecanismo que faça com que novos nós e arestas sejam acrescentados ao grafo, de acordo com a vontade do usuário;
- **Agrupar** - permite que sejam definidos grupos com os nós com características semelhantes para diminuir a quantidade de nós, tornando assim a imagem mais "limpa";
- **Representação** - define a estrutura utilizada para a representação gráfica da rede social. Todas as ferramentas estudadas usam a representação em forma de grafo.

Ao analisar a Tabela 3.1, pode-se perceber que nenhuma das ferramentas estudadas permite que seja feito o agrupamento de nós de forma visual. Vale lembrar aqui que o agrupamento aqui não está se referindo ao cálculo do coeficiente de *Clustering*, que é outra definição. Nesse item, o que está sendo avaliado é a possibilidade de formações de grupos visualmente para o usuário. Outro item que chama atenção ao analisar essa tabela é a linha referente ao item de Expansão, sendo que a única ferramenta que permite que o grafo seja expandido de forma incremental é o ArnetMiner.

A Tabela 3.2 contém as métricas mais utilizadas e calculadas pelas ferramentas de visualização de redes sociais analisadas (detalhes sobre as métricas podem ser vistos na Seção 2). Quando um item está marcado com 'Sim', significa que a ferramenta possui uma determinada característica; quando marcado com 'Não', significa que a ferramenta não possui a característica e quando marcado com '?' significa que não é possível afirmar se a ferramenta possui ou não uma característica.

Na Tabela 3.2, o que chama atenção novamente é a coluna referente à ferramenta ArnetMiner, apesar de não calcular as métricas mostradas na Tabela 3.2, calcula outras métricas julgadas pelos autores como mais relevantes para redes de coautoria, são elas: sociabilidade, h-index, longevidade, diversidade, entre outras. Já as outras ferramentas estudadas se aplicam a campos genéricos, de forma que o usuário pode visualizar qualquer tipo de rede, não sendo especializadas em nenhuma área.

	Ferramentas								
Características	COMMETRIX	COSBILab Graph	ArnetMiner	SocialNetSense	Pajek	igraph	UCINET	NodeXL	SocialAction
<i>Zoom</i>	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim
Distância mínima entre nós	Sim	Não	Não	Não	Sim	Sim	Não	Não	Não
Comprimento das arestas	Sim	Não	Não	Não	?	Sim	Não	Não	Não
Alterar rótulos	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim
Alterar tamanho dos rótulos	Sim	?	Não	Sim	Sim	Sim	Sim	Sim	Sim
Alterar tamanho dos nós	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim
Alterar largura das arestas	Sim	Sim	Não	Não	Sim	Sim	Sim	?	Não
Filtros	Sim	Não	Não	Sim	?	Sim	Sim	Sim	Sim
Consulta	Sim	Sim	Não	Não	Não	Não	Não	Não	Não
Expande	Não	Não	Sim	Não	Não	Não	Não	Sim	Não
Agrupa	Não	Não	Não	Não	Não	Não	Não	Sim	Sim
Representação	Grafo	Grafo	Grafo	Grafo	Grafo	Grafo	Grafo	Grafo	Grafo

Tabela 3.1: Tabela com características de visualização e interatividade.

	Ferramentas								
Métricas	COMMETRIX	COSBILab Graph	ArnetMiner	SocialNetSense	Pajek	igraph	UCINET	NodeXL	SocialAction
Distância	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim
Excentricidade	Não	Não	Não	Não	Não	Não	Não	Não	Não
Grau	Sim	Sim	Não	Sim	Sim	Sim	?	Sim	Sim
Diâmetro	Sim	Sim	Não	?	Sim	Sim	?	Sim	Sim
Densidade	Sim	Não	Não	Sim	Não	Sim	Sim	Sim	Sim
Centralidade de Grau	Sim	Não	Não	Sim	Não	Sim	Sim	Não	Sim
Centralidade de <i>Closeness</i>	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim
Centralidade de <i>Betweenness</i>	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim
Coefficiente de <i>Clustering</i>	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim
Assortatividade	Não	Não	Não	Não	Não	Não	Não	Não	Não
Coefficiente de Colaboração	Não	Não	Não	Não	Não	Não	Não	Não	Não

Tabela 3.2: Tabela com resumo das métricas calculadas por cada ferramenta.

4 DESCRIÇÃO DA FERRAMENTA

Este capítulo é dedicado ao detalhamento da ferramenta desenvolvida neste trabalho e está organizado da seguinte forma: a Seção 4.1 mostra uma visão geral das contribuições e propostas de visualização dos grafos das redes de coautoria. Na Seção 4.2, é detalhada a implementação e funcionalidades do protótipo. Finalizando o capítulo, é apresentada uma seção de considerações finais.

4.1 Visão Geral

PANDORA permite visualizar redes de coautoria a partir de arquivos XML que representam produção bibliográfica. A visualização pode ser feita de forma incremental, a partir de expansões no grafo, e permite gerar dados analíticos a respeito da rede que está sendo visualizada.

A Figura 4.1 mostra a visão geral da proposta deste trabalho. A partir de um arquivo fornecido pelo usuário, neste caso um arquivo XML, é gerada uma visualização do grafo da rede de coautoria que permite interação com o usuário de tal forma que ele possa expandir o grafo de acordo com o que deseja visualizar. Ou seja, à medida que o usuário clica em algum nó novos nós e arestas são adicionados ao grafo, caso o nó clicado ainda não tenha sido explorado. Nesta figura, as arestas que foram adicionadas após o clique do usuário estão representadas em vermelho e tracejadas. Cada nó é apresentado sempre que pelo menos uma de suas arestas estiver visível. Por exemplo, na Figura 4.1, os nós que possuem apenas arestas tracejadas tornaram-se visíveis só após o clique do usuário que fez surgir as arestas tracejadas.

Outra funcionalidade de PANDORA é gerar gráficos referentes às métricas calculadas para o grafo. Neste trabalho, os gráficos sugeridos são os de Grau x Quantidade de Publicações, gráfico de Coeficiente de Colaboração e gráfico de Assortatividade. O primeiro gráfico mostra a quantidade de pesquisadores para cada par de coordenadas (grau, quantidade de publicações), assim é possível avaliar a produtividade do grupo visualizado no momento, verificando se os pesquisadores que possuem grau alto também possuem grande quantidade de publicações. O gráfico de Coeficiente de Colaboração indica, para o grafo visualizado no momento, o índice de colaboração entre cada pesquisador e seus colaboradores. À medida que o grafo é expandido, essa métrica é

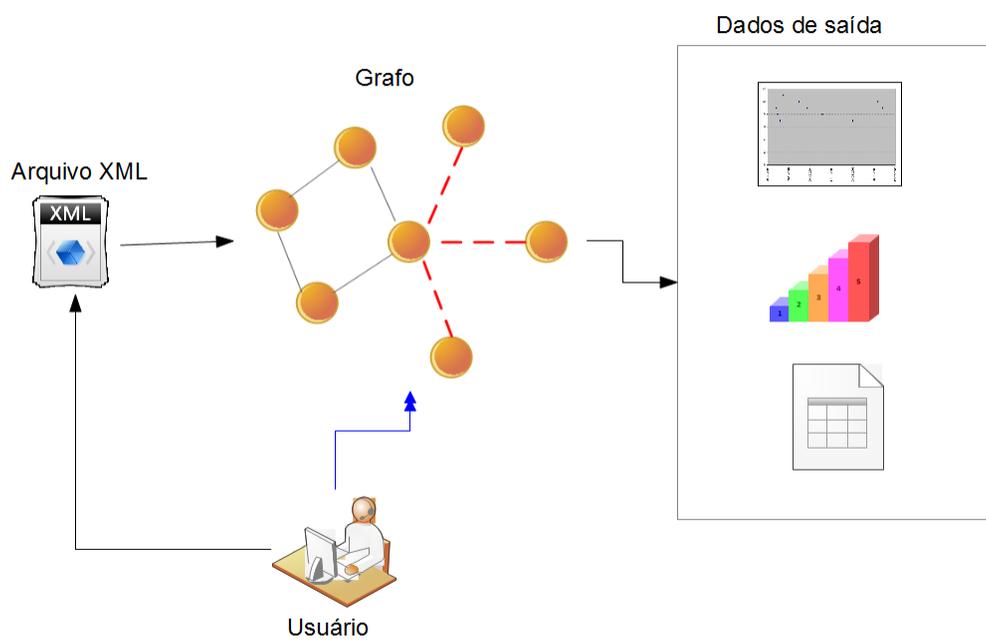


Figura 4.1: Visão geral da proposta de PANDORA: o usuário entra com um arquivo XML, contendo a produção bibliográfica, que é usado pelo sistema para gerar uma visualização de grafo, bem como dados analíticos sobre este grafo. A seta em azul indica interação entre o usuário e o sistema.

recalculada para o novo conjunto de pesquisadores. O gráfico de Assortatividade mostra a relação entre quantidade de colaboradores e média das quantidades de colaboradores dos colaboradores de um pesquisador.

4.1.1 Visualização de Grafos

Este trabalho propõe duas abordagens para visualização de grafos que representam redes sociais acadêmicas de coautoria. A primeira abordagem permite visualizar o grafo completo envolvendo todos os pesquisadores desejados. A segunda abordagem permite incrementar gradualmente o grafo com novos nós que representam pesquisadores colaboradores dos que já estavam sendo visualizados. À medida que o usuário clica em um dos nós no grafo, o grafo é ampliado com todos os coautores do nó clicado.

A Figura 4.2 mostra um esquema no qual podem ser vistas as duas abordagens propostas. O fluxo superior seguindo pela seta 1 representa a abordagem de visualização do grafo completo, já o fluxo a partir da seta 4 representa a abordagem incremental.

Seguindo a partir da seta 1 os seguintes passos são executados:

- Nesse caso, o usuário escolheu visualizar um grafo que contém todos os pesquisadores do arquivo passado como entrada (seta 1), a ferramenta gera então apenas um arquivo XML, a partir do GraphML que representa toda a rede de colaboração entre os pesquisadores (seta 2);
- A partir daí é gerada uma visualização do grafo, com leiaute baseado em forças, e gráficos com métricas extraídas do próprio grafo (seta 2);
- A seta 3 representa a interação que o usuário pode ter com o grafo gerando planilhas, gerando imagens do que está se visualizando dentre outros tipos de interação.

Seguindo a partir da seta 4 os seguintes passos são executados:

- Inicialmente o usuário escolhe um dos pesquisadores disponíveis para visualizar seu grafo de colaboração (seta 4);
- A aplicação gera então diversos arquivos XML um para cada pesquisador no arquivo inicial, sendo que cada arquivo contém informações sobre o pesquisador principal ao qual pertence o arquivo e seus colaboradores (seta 5);
- Depois de separados os arquivos XML, o arquivo referente ao pesquisador que foi escolhido pelo usuário é separado pela ferramenta e usada para gerar a visualização do grafo desse pesquisador, com leiaute radial. Desta forma, a aplicação gera também gráficos referentes ao grafo visualizado (seta 6);

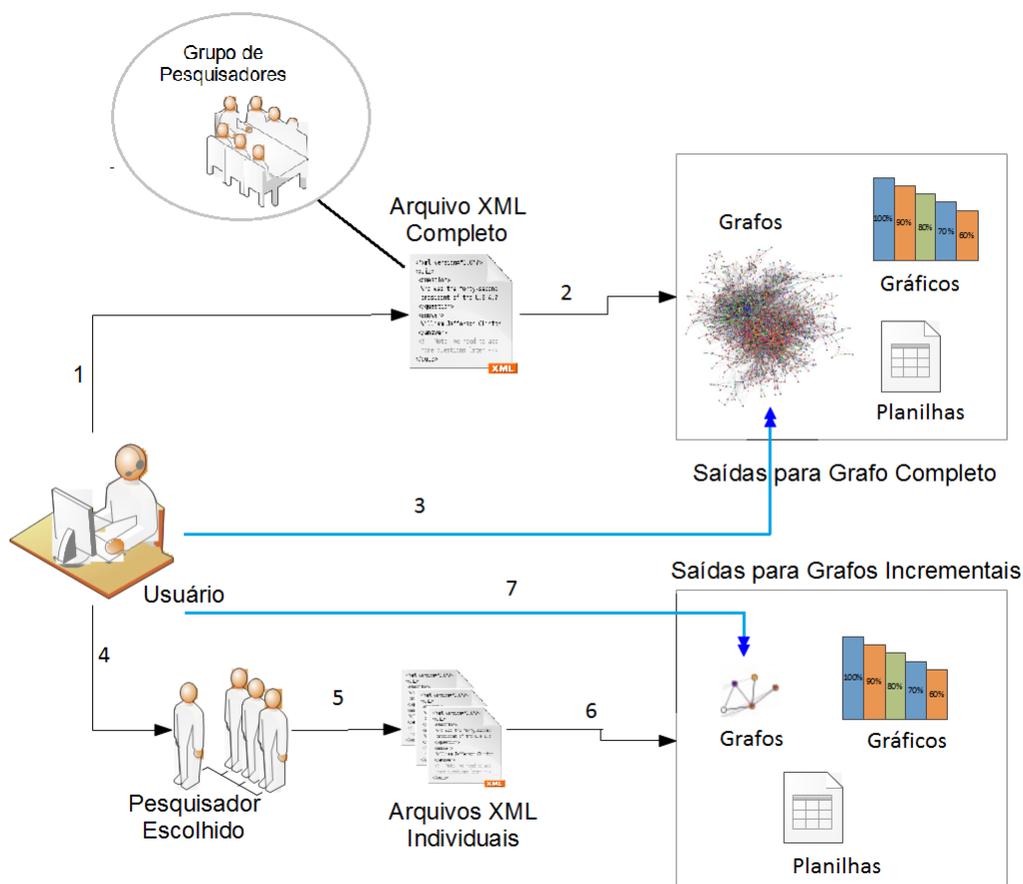


Figura 4.2: Esquema do fluxo de execução da aplicação. O fluxo superior representa a abordagem de visualização do grafo completo, em que a aplicação recebe um arquivo XML com todos os pesquisadores, para em seguida gerar um grafo completo juntamente com gráficos e planilhas. O fluxo inferior representa a abordagem incremental em que um pesquisador é escolhido pelo usuário; a aplicação recebe vários arquivos XML e cria um grafo utilizando o XML referente ao pesquisador escolhido pelo usuário, juntamente com planilhas e gráficos.

- A seta 7 representa a interação com o usuário que pode expandir o grafo, gerar planilhas com as métricas, etc. Ao expandir o grafo, outros arquivos XML são lidos para que novos nós e arestas sejam adicionados.

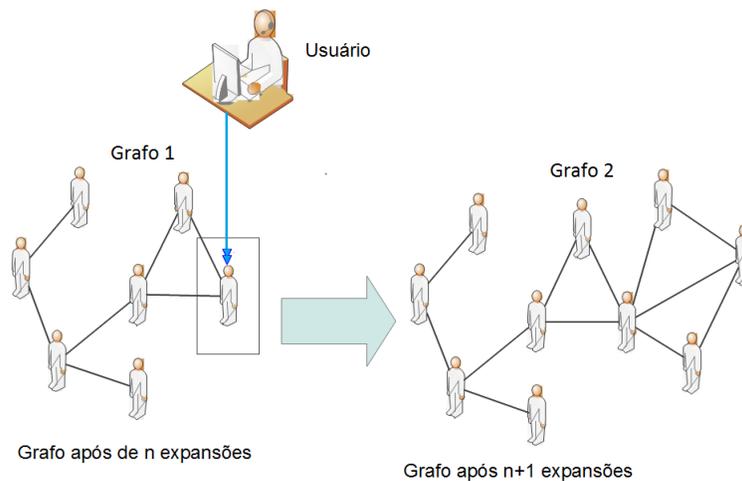


Figura 4.3: Esquema da abordagem incremental: ao clicar em um nó qualquer, o grafo é expandido caso o pesquisador tenha colaboradores que não estejam visíveis ainda.

A abordagem incremental pode ser vista na Figura 4.3, que representa um exemplo, onde após n expansões, o usuário obtém o Grafo 1. Em seguida, ao clicar em um nó que ainda não foi expandido, no caso o nó destacado no Grafo 1, é obtido o Grafo 2. Neste último grafo são adicionados colaboradores que antes não eram visualizados, assim como as arestas que representam colaborações entre eles. A seta em azul representa a interação e intervenção do usuário sobre o grafo.

4.1.2 Geração de Gráficos para Análises de Dados

Para analisar os grafos gerados pela PANDORA, além das métricas calculadas usando JUNG¹, que são mostradas ao deixar o *mouse* sobre os nós no grafo radial, são também criados três gráficos: gráfico Grau x Quantidade de Publicações, gráfico Coeficiente de Colaboração e gráfico de Assortatividade.

No gráfico Grau x Quantidade de Publicações, as métricas utilizadas para sua construção são os graus de cada pesquisador e a quantidade de publicações que ele possui. É importante dizer que nesse gráfico estão sendo levados em conta apenas os pesquisadores que possuem todas suas publicações e colaboradores mostradas no grafo e que tenham sido escolhidos (clicados) pelo usuário. Isso porque para um pesquisador representado por um nó que ainda não havia sido expandido, o cálculo das métricas não estaria condizente com a realidade.

¹<http://jung.sourceforge.net/>

No gráfico de Coeficiente de Colaboração, são calculados os coeficientes de colaboração do grupo que está sendo visualizado no momento. A cada expansão, o pesquisador selecionado é adicionado ao gráfico e a métrica é recalculada para o grupo atual. Desta forma, é possível analisar, para um determinado grupo, de que forma a adição de colaboradores e publicações de um pesquisador, no caso o clicado pelo usuário, influencia, no índice de colaboração do grupo.

O gráfico de Assortatividade é calculado apenas na visualização de grafos completos. Para cada grau encontrado no grafo são calculadas as médias dos graus dos colaboradores do pesquisador. Assim, é possível verificar se pesquisadores com grande quantidade de colaboradores tendem a se relacionar com pesquisadores que também se relacionam com muitos colaboradores.

4.2 Implementação de PANDORA

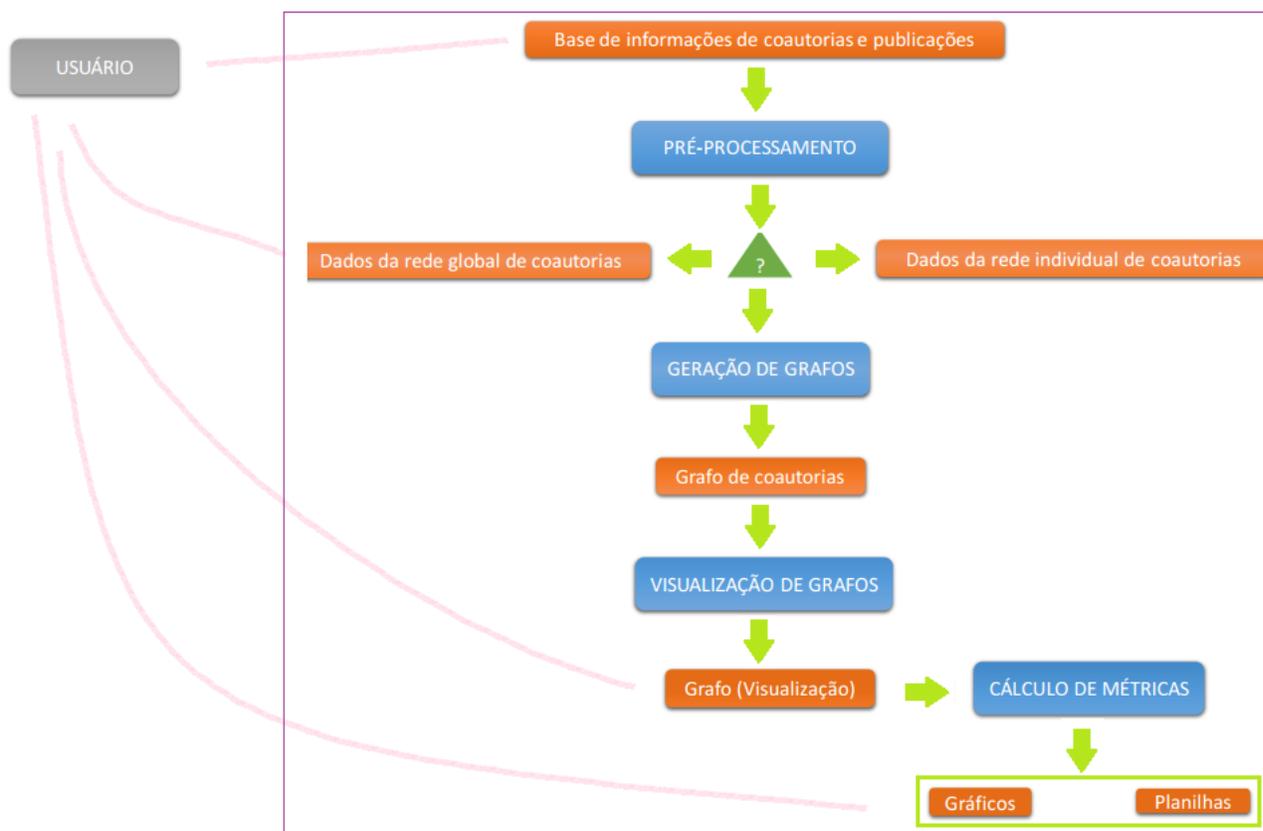


Figura 4.4: Representação dos módulos de PANDORA incluindo interação com o usuário.

O objetivo da implementação de PANDORA é validar a proposta apresentada na Seção 4.1 e mostrar sua viabilidade. A Figura 4.4 mostra o fluxograma que representa os módulos implementados em PANDORA. As caixas azuis representam cada módulo e as de cor laranja representam

dados de entrada e saída de cada módulo. O usuário está representado por uma caixa cinza e para cada dado que ele tem acesso foi adicionada uma linha rosa relacionando-os. A primeira interação com usuário acontece no momento em que ele fornece um arquivo contendo informações de coautorias e publicações. A partir desse arquivo GraphML é feito um pré-processamento, onde são geradas estruturas de dados internas, contendo listas de pesquisadores, publicações e relacionamentos de colaboração. Estas estruturas são usadas para construção dos arquivos XML e, posteriormente, esses arquivos são usados para geração dos grafos. Se o usuário escolhe visualizar o grafo completo, então é gerado um arquivo XML contendo todos os pesquisadores, e são gerados dados da rede global de coautorias. Caso contrário, são criados arquivos XML individualizados, contendo o pesquisador escolhido pelo usuário e seus colaboradores, gerando redes de coautoria.

Para cada pesquisador, é gerado um arquivo XML, como o mostrado na Figura 4.5. Cada arquivo XML contém apenas publicações que envolvem o pesquisador que o usuário selecionou como autor. A caixa em vermelho, na Figura 4.5, mostra as definições do XML, onde cada atributo dos nós e arestas é definido. Dentro da caixa azul, são mostradas informações do pesquisador principal dessa rede, ou seja, esse arquivo se refere a esse pesquisador. O exemplo, refere-se à pesquisadora Carla Maria Dal Sasso Freitas. A caixa em amarelo apresenta as informações sobre os colaboradores de outro pesquisador principal, no exemplo: Ana Lucia Cetertich Bazzan, Denise de Oliveira e Kai Nagel. A caixa rosa apresenta as arestas que mostram quais nós elas estão conectando, a quantidade de publicações em comum e quais são estas publicações.

Depois de gerados os arquivos XML, o módulo de geração de grafos, gera uma estrutura de dados que representa o grafo, a qual é usada para calcular as métricas, assim como para gerar outras visualizações do grafo. Vale ressaltar que as métricas calculadas se referem ao grafo atual visualizado pelo usuário. A partir das métricas são gerados gráficos para análise.

O código da aplicação foi desenvolvido usando linguagem Java, com apoio da biblioteca gráfica Prefuse para facilitar a geração dos grafos. Além disso, a biblioteca JUNG foi usada para auxiliar no cálculo das métricas mais comuns. Já para a geração dos gráficos, foi utilizada a biblioteca JFreeChart.

4.2.1 Funcionalidades de PANDORA

A Figura 4.6 mostra um exemplo de arquivo GraphML usado como entrada para execução da aplicação desenvolvida. O trecho do arquivo circundado em verde contém as definições dos atributos pertencentes aos elementos que definem os nós (pesquisadores), arestas e publicações. No trecho em azul, é mostrada a definição de um nó representante da pesquisadora Carla Maria Dal Sasso Freitas, que contém um *id* que serve como referência para indicar nas arestas quais são os nós que esta estará ligando. No trecho em rosa, é possível ver a definição de duas arestas, onde *target* e *source* recebem o valor dos *ids* referentes aos pesquisadores que estão ligados pela aresta. Estas arestas estão relacionadas à pesquisadora Carla Freitas, já que o *source*, destacado em amarelo, possui o valor do *id* dessa pesquisadora. Caso dois pesquisadores possuam mais de uma publicação em comum, a aresta terá *n data_element*, um para cada publicação. Por fim, o trecho

```

<?xml version="1.0" encoding="UTF-8"?>
<graphml>
  <graph edgedefault="undirected">
    <!--nodes-->
    <key id="id" for="node" attr.type="string" attr.name="id"/>
    <key id="category" for="node" attr.type="string" attr.name="category"/>
    <key id="inbooks" for="node" attr.type="string" attr.name="inbooks"/>
    <key id="name" for="node" attr.type="string" attr.name="name"/>
    <key id="pubcount" for="node" attr.type="string" attr.name="pubcount"/>
    <key id="injournals" for="node" attr.type="string" attr.name="injournals"/>
    <key id="abbreviature" for="node" attr.type="string" attr.name="abbreviature"/>
    <key id="inproceedings" for="node" attr.type="string" attr.name="inproceedings"/>
    <key id="image" for="node" attr.type="string" attr.name="image"/>
    <key id="commonpubs" for="edge" attr.type="int" attr.name="commonpubs"/>
    <key id="publications" for="edge" attr.type="string" attr.name="publications"/>
    <node id="58">
      <data key="id">58</data>
      <data key="category">Docente</data>
      <data key="inbooks">2</data>
      <data key="name">Carla Maria Dal Sasso Freitas</data>
      <data key="pubcount">2</data>
      <data key="injournals">0</data>
      <data key="abbreviature">FREITAS, C. M. D. S.</data>
      <data key="inproceedings">0</data>
      <data key="image">usuario.png</data>
    </node>
    <node id="57">
      <data key="id">57</data>
      <data key="category">Docente</data>
      <data key="inbooks">2</data>
      <data key="name">Taisy Silva Weber</data>
      <data key="pubcount">2</data>
      <data key="injournals">0</data>
      <data key="abbreviature">WEBER, T. S.</data>
      <data key="inproceedings">0</data>
      <data key="image">usuario.png</data>
    </node>
    <node id="56">
      <data key="id">56</data>
      <data key="category">Participante Externo</data>
      <data key="inbooks">2</data>
      <data key="name">Mirella Moura Moro</data>
      <data key="pubcount">2</data>
      <data key="injournals">0</data>
      <data key="abbreviature">MORO, M. M.</data>
      <data key="inproceedings">0</data>
      <data key="image">usuario.png</data>
    </node>
    <!--edges-->
    <edge target="58" source="57">
      <data key="commonpubs">2</data>
      <data key="publications">29 24</data>
    </edge>
    <edge target="58" source="56">
      <data key="commonpubs">2</data>
      <data key="publications">29 24</data>
    </edge>
  </graph>
</graphml>

```

Definições

Pesquisador
PrincipalPesquisadores
Colaboradores

Arestas

Figura 4.5: Arquivo XML da pesquisadora Carla Freitas no ano 2008.

```

<?xml version="1.0" standalone="true"?>
<!DOCTYPE graphml SYSTEM "http://graphml.graphdrawing.org/DTDs/1.0rc/graphml.dtd">
<graphml>
- <graph id="graphml" edgedefault="undirected">
  <key id="event" attr.multiple="false" attr.type="string" attr.name="event" for="composite"/>
  <key id="country" attr.multiple="false" attr.type="string" attr.name="country" for="composite"/>
  <key id="title" attr.multiple="false" attr.type="string" attr.name="title" for="composite"/>
  <key id="publisher_city" attr.multiple="false" attr.type="string" attr.name="publisher_city" for="composite"/>
  <key id="book_title" attr.multiple="false" attr.type="string" attr.name="book_title" for="composite"/>
  <key id="notes" attr.multiple="false" attr.type="string" attr.name="notes" for="composite"/>
  <key id="event_city" attr.multiple="false" attr.type="stream" attr.name="event_city" for="composite"/>
  <key id="issn" attr.multiple="false" attr.type="string" attr.name="issn" for="composite"/>
  <key id="language" attr.multiple="false" attr.type="string" attr.name="language" for="composite"/>
  <key id="edition_number" attr.multiple="false" attr.type="int" attr.name="edition_number" for="composite"/>
  <key id="level" attr.multiple="false" attr.type="string" attr.name="level" for="composite"/>
  <key id="last_page" attr.multiple="false" attr.type="int" attr.name="last_page" for="composite"/>
  <key id="publisher" attr.multiple="false" attr.type="string" attr.name="publisher" for="composite"/>
  <key id="editor_name" attr.multiple="false" attr.type="string" attr.name="editor_name" for="composite"/>
  <key id="journal_name" attr.multiple="false" attr.type="string" attr.name="journal_name" for="composite"/>
  <key id="volume" attr.multiple="false" attr.type="int" attr.name="volume" for="composite"/>
  <key id="proceedings" attr.multiple="false" attr.type="string" attr.name="proceedings" for="composite"/>
  <key id="line_of_research" attr.multiple="false" attr.type="string" attr.name="line_of_research" for="composite"/>
  <key id="page_amount" attr.multiple="false" attr.type="int" attr.name="page_amount" for="composite"/>
  <key id="url" attr.multiple="false" attr.type="url" attr.name="url" for="composite"/>
  <key id="nature" attr.multiple="false" attr.type="string" attr.name="nature" for="composite"/>
  <key id="field" attr.multiple="false" attr.type="string" attr.name="field" for="composite"/>
  <key id="isbn" attr.multiple="false" attr.type="int" attr.name="isbn" for="composite"/>
  <key id="type" attr.multiple="false" attr.type="string" attr.name="type" for="composite"/>
  <key id="first_page" attr.multiple="false" attr.type="int" attr.name="first_page" for="composite"/>
  <key id="year" attr.multiple="false" attr.type="int" attr.name="year" for="composite"/>
  <key id="inproceedings" attr.multiple="false" attr.type="int" attr.name="Number of Publications in Proceedings" for="node"/>
  <key id="abbreviature" attr.multiple="false" attr.type="string" attr.name="Abbreviature" for="node"/>
  <key id="category" attr.multiple="false" attr.type="string" attr.name="Category" for="node"/>
  <key id="pubcount" attr.multiple="false" attr.type="int" attr.name="Number of Publications" for="node"/>
  <key id="name" attr.multiple="false" attr.type="string" attr.name="Name" for="node"/>
  <key id="injournals" attr.multiple="false" attr.type="int" attr.name="Number of Publications in Journals" for="node"/>
  <key id="inbooks" attr.multiple="false" attr.type="int" attr.name="Number of Publications in Books" for="node"/>
  <key id="publications" attr.multiple="true" attr.type="composite" attr.name="Common Publications" for="edge"/>
  <key id="commonpubs" attr.multiple="false" attr.type="int" attr.name="Number of Common Publications" for="edge"/>
- <composite id="0">
  <data key="title">A High Performance Approach for Inner Structures Visualization in Medical Data</data>
  <data key="journal_name">International Journal Of Computer Applications In Technology</data>

```

Atributos

```

- <node id="2">
  <data key="inproceedings">5</data>
  <data key="abbreviature">FREITAS, C. M. D. S.</data>
  <data key="category">Docente</data>
  <data key="pubcount">6</data>
  <data key="name">Carla Maria Dal Sasso Freitas</data>
  <data key="injournals">1</data>
  <data key="inbooks">0</data>
</node>

```

Dados Pesquisador

```

- <edge id="2#3" target="3" source="2">
  <data key="publications">
    <data_element>0</data_element>
  </data>
  <data key="commonpubs">1</data>
</edge>
- <edge id="2#4" target="4" source="2">
  <data key="publications">
    <data_element>0</data_element>
  </data>
  <data key="commonpubs">1</data>
</edge>

```

Arestas

```

- <composite id="0">
  <data key="title">A High Performance Approach for Inner Structures Visualization in Medical Data</data>
  <data key="journal_name">International Journal Of Computer Applications In Technology</data>
  <data key="issn">09528091</data>
  <data key="field">SISTEMAS DE COMPUTACAO</data>
  <data key="nature">Trabalho Completo</data>
  <data key="last_page">33</data>
  <data key="notes">internacional C</data>
  <data key="type">Artigo em Periodico</data>
  <data key="volume">22</data>
  <data key="first_page">23</data>
  <data key="language">Ingles</data>
  <data key="year">2005</data>
</composite>

```

Dados Publicação

Figura 4.6: Trechos do arquivo de entrada utilizado na aplicação.

circundado em amarelo mostra informações de uma publicação, como título, ano de publicação, etc.

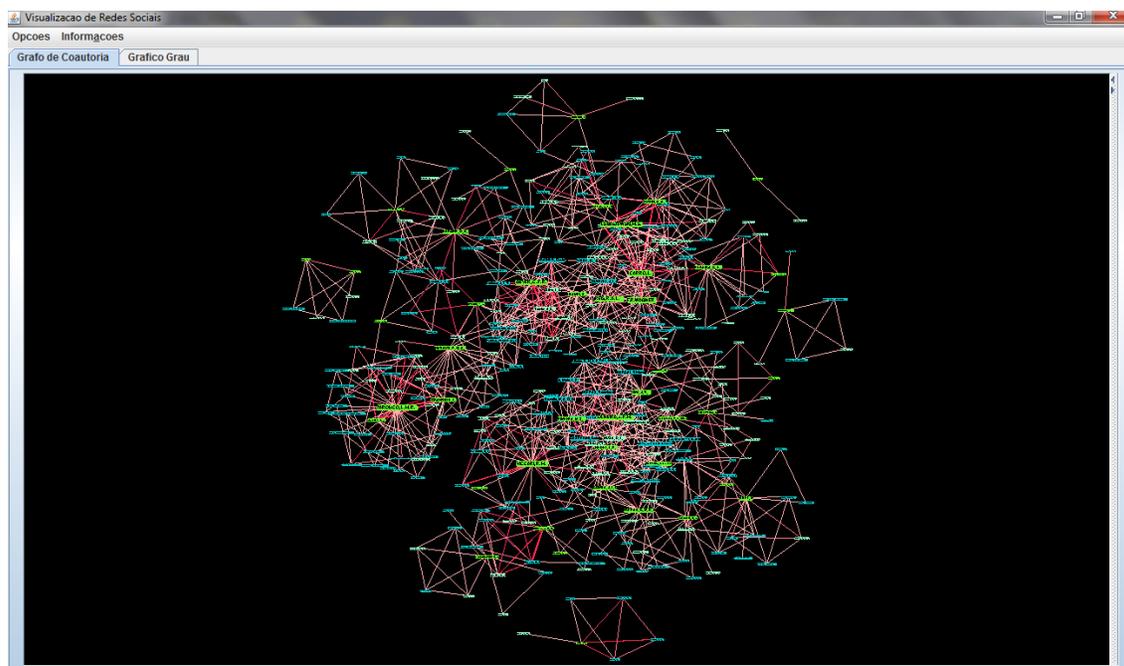


Figura 4.7: Exemplo de grafo completo mostrado ao usuário.

Inicialmente, o usuário escolhe um arquivo GraphML, como o mostrado na Figura 4.6. Em seguida, o usuário deve informar se deseja visualizar apenas o grafo completo ou, partir de um determinado pesquisador, expandir o grafo incrementalmente clicando em cada nó desejado. Para o primeiro caso, é gerado um grafo, como pode ser visto na Figura 4.7, contendo todos os pesquisadores presentes no arquivo XML passado como entrada, além de dois gráficos: um gráfico relacionando grau do nó e quantidade de publicações e outro evidenciando a assortatividade (Grau x Média dos graus dos colaboradores).



Figura 4.8: Colorações que representam as diferentes categorias para os pesquisadores. Na ordem das imagens: Discente, Docente e Participante Externo.

Na primeira aba, é mostrado o grafo completo, representado com leiaute baseado em forças, onde cada nó é influenciado por forças geradas pela interação entre as molas (no caso, as arestas são

tratadas como molas), força de repulsão entre os nós e força de arraste, que é similar à resistência do ar. Neste grafo, as cores dos nós significam as diferentes categorias de nós: Discente (verde claro), Docente (verde grama) ou Participante Externo (verde azulado), como pode ser visto na Figura 4.8. O tamanho dos nós mostra a quantidade de arestas que estão ligadas ao nó, ou seja, seu grau ou, no caso, a quantidade de colaboradores que ele tem. As cores e largura das arestas mostram a quantidade de publicações envolvidas na relação entre dois pesquisadores. Quanto mais escura for a cor da aresta, mais publicações estão sendo representadas; quanto mais grossa, mais publicações estão envolvidas (foram usados dois tipos de representação para chamar mais atenção do usuário). Está disponível uma variação de quatro cores que representam:

- Quantidade de publicações igual a 1 - cor mais clara;
- Quantidade de publicações igual a 2 - segunda cor mais clara;
- Quantidade de publicações igual a 3 - cor escura;
- Quantidade de publicações maior que 3 - cor mais escura.

Vale ressaltar que a largura das arestas segue o mesmo padrão citado para as cores, variando em quatro valores.

Na segunda aba é mostrado um gráfico Grau x Quantidade de Publicações com todos os pesquisadores do conjunto de entrada. Neste gráfico, segundo a legenda, cada ponto possui uma cor e uma forma geométrica que significa a quantidade de pesquisadores que possuem um par de valores (grau, número de publicações). Um exemplo deste gráfico pode ser visto na Figura 4.9.

Na terceira aba é mostrado o gráfico de Assortatividade, onde para cada grau dos nós encontrado no grafo, é calculada a média dos graus dos colaboradores do pesquisador representado pelo nó. Caso apareça um grau para o qual já tenha sido calculada a média, essa é novamente calculada agora incluindo os novos nós. Um exemplo deste tipo de gráfico pode ser visto na Figura 4.10.

Para o caso em que o usuário deseja, inicialmente, visualizar o grafo egocêntrico, a partir de um único pesquisador, será exibida uma lista com todos os pesquisadores do conjunto de entrada para que o usuário possa selecionar um deles. O usuário pode expandir o grafo de forma incremental, clicando nos nós que representam co-autores. A cada clique são acrescentados os nós dos colaboradores daquele que foi selecionado, e que não pertenciam ao grafo, juntamente com as arestas correspondentes às publicações que não estavam sendo visualizadas.

A partir desse momento, estão disponíveis quatro visualizações em abas específicas da interface:

- Grafo de Coautoria - grafo contendo o pesquisador inicialmente escolhido pelo usuário juntamente com seus colaboradores. Neste grafo, estão representadas apenas as relações correspondentes às publicações que envolvam o pesquisador inicial. Ao expandir um nó, são adicionadas as arestas que representam publicações que envolvam o pesquisador expandido.

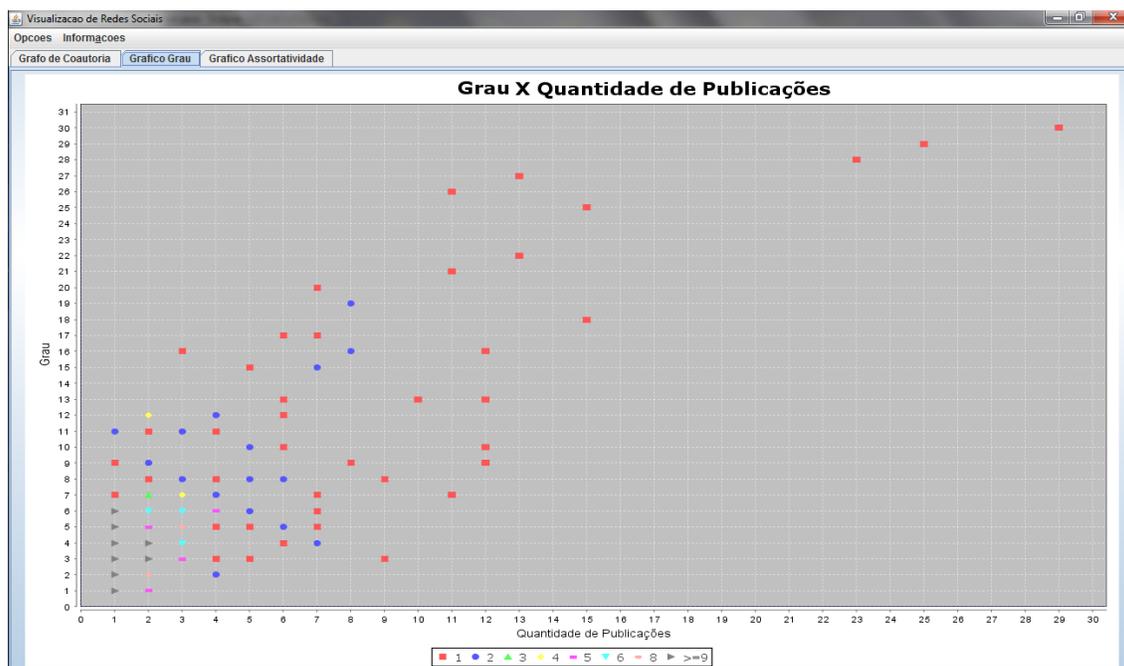


Figura 4.9: Relação Grau x Quantidade de Publicações correspondente ao grafo completo.

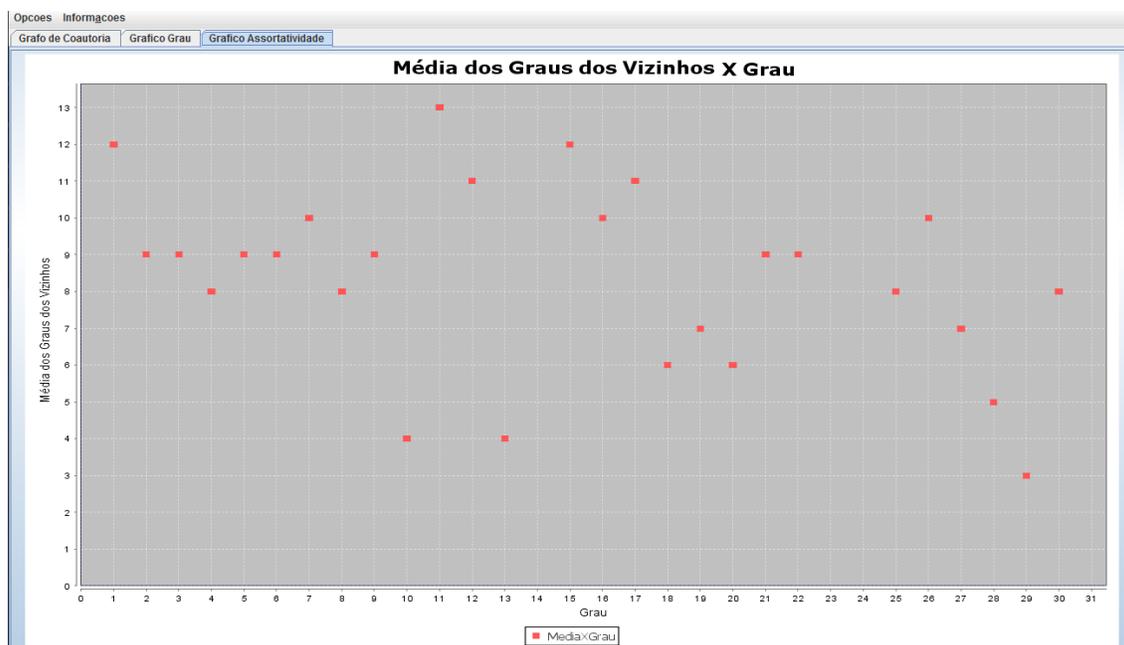


Figura 4.10: Gráfico Assortatividade (Grau x Média dos graus dos colaboradores) para dados de um grafo completo.

Neste grafo, os nós e arestas são representados pelas mesmas cores e larguras citadas anteriormente para o grafo completo. O grafo é mostrado com leiaute radial, em que o nó central mostra o pesquisador inicial ou o último nó clicado para expandir o grafo, já que o Prefuse rearranja o leiaute à medida que o usuário clica em um nó. Ao deixar o *mouse* sobre um nó, são mostradas informações referentes àquele nó como quantidade de publicações, quantidade de publicações ao ser expandido, categoria, métricas calculadas como grau, *betweenness*, etc (Figura 4.11).

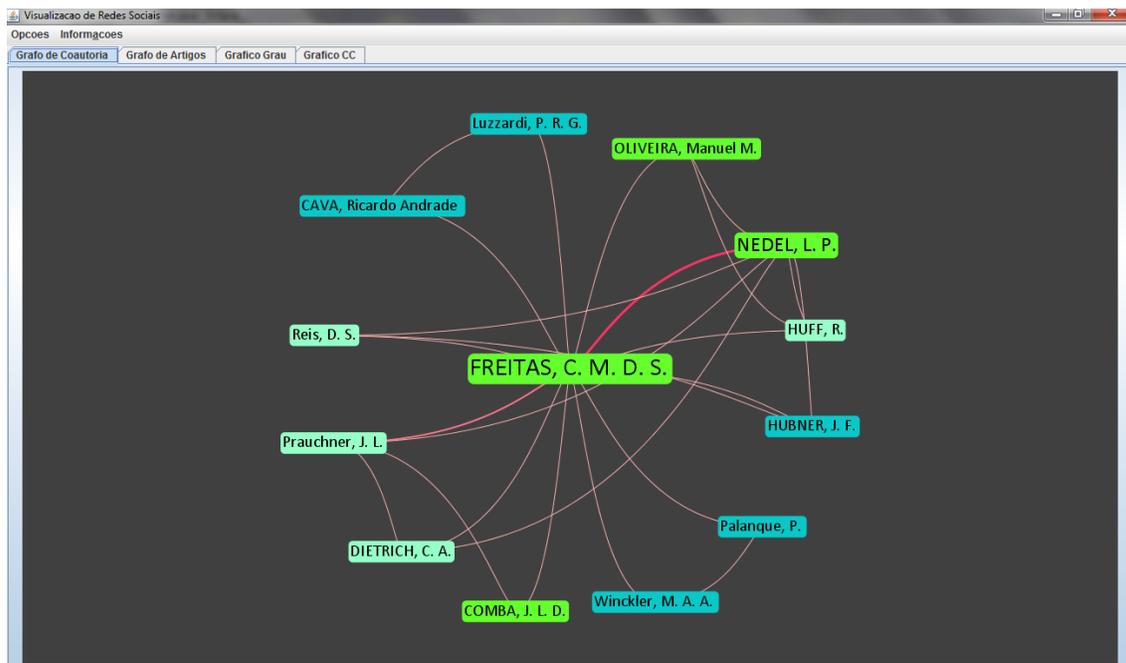


Figura 4.11: Visualização gerada depois do usuário selecionar um pesquisador específico no grafo em exibição no momento.

- Grafo de Publicações - grafo que mostra todas as publicações representadas no grafo de co-autoria, usando leiaute de forças. O grafo subjacente tem as arestas unindo artigos do mesmo autor. Cada publicação é representada por um quadrado em que o tamanho varia de acordo com a quantidade de autores. Para verificar quais artigos tem autores em comum, basta deixar o *mouse* sobre um dos artigos, os que tiverem autores em comum serão destacados em vermelho. Ao deixar o *mouse* sobre um nó nesse grafo, são mostrados todos os autores que participaram da produção daquela publicação, como pode ser visto na Figura 4.12.
- Gráfico Grau x Quantidade de Publicações - gráfico que representa o grau e a quantidade de publicações dos pesquisadores expandidos a partir do pesquisador inicial. Este gráfico utiliza a mesma representação do gráfico para o grafo completo citado anteriormente (Figura

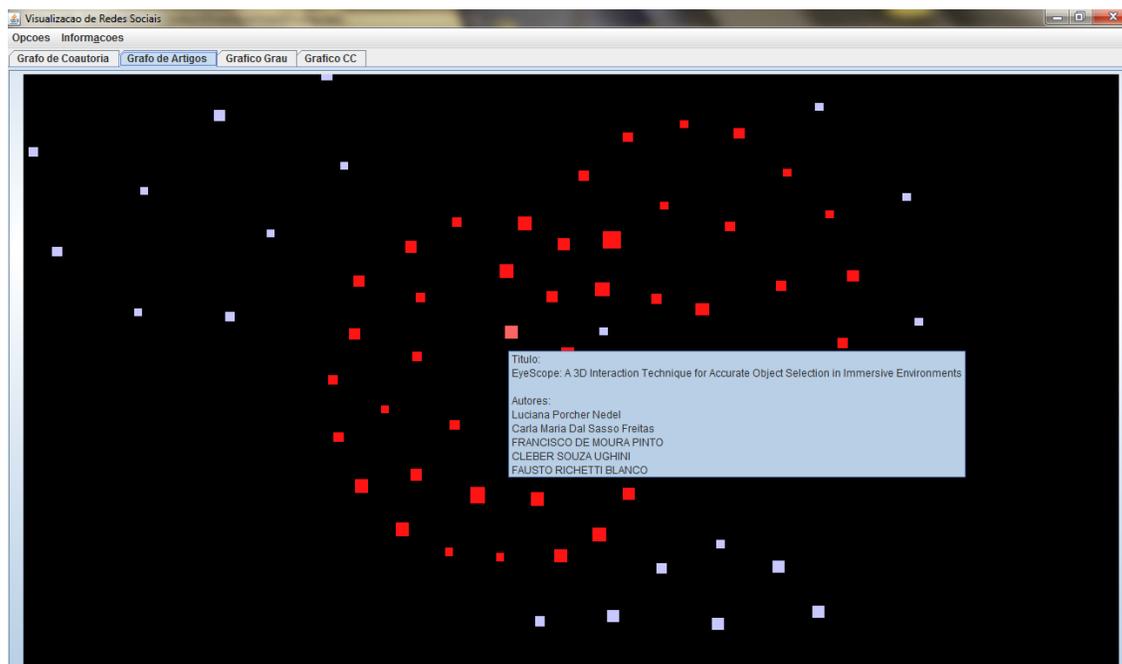


Figura 4.12: Grafo de artigos após algumas expansões no grafo de Coautoria.

4.9). A partir desse gráfico é possível inferir se os autores que possuem maior quantidade de colaboradores (maior grau), em geral, publicam mais que os demais.

- Gráfico de Coeficiente de Colaboração - são calculados os coeficientes de colaboração para o grupo de pesquisadores representado no grafo, naquele momento, considerando apenas as publicações entre eles. Ao expandir o grafo, novas coautorias (e novas publicações) são inseridas e o coeficiente de colaboração é recalculado. O eixo das abscissas representa os nomes dos pesquisadores que foram selecionados até o momento, como pode ser visto na Figura 4.13.

Para ambas as visualizações do grafo completo ou grafo expandido de forma incremental, existem funções na aplicação: "Opções" e "Informações". A partir de "Opções", o usuário pode:

1. reiniciar a visualização do grafo a partir do mesmo arquivo ou de um outro arquivo no mesmo formato;
2. gerar imagens que representam o grafo ou qualquer um dos gráficos gerados;
3. gerar arquivos com os dados dos nós e do grafo completo, que podem ser abertos em planilhas como o Excel, podendo gerar outros gráficos a partir das planilhas, sendo possível também filtrar as características que aparecerão no arquivo.

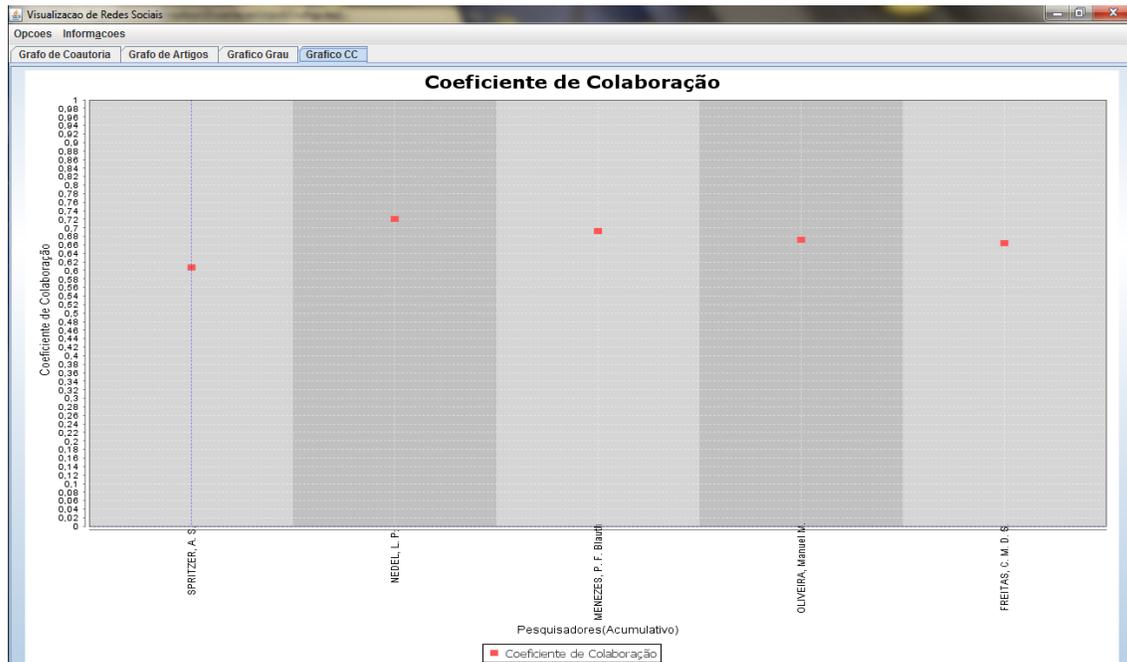


Figura 4.13: Coeficiente de Colaboração calculado sobre os dados que se referem aos pesquisadores representados no grafo visualizado.

Já no botão "Informações", o usuário pode ver os créditos e informações sobre as funcionalidades da ferramenta.

Algumas funcionalidades presentes na ferramenta se referem aos tipos de interação permitidos com a visualização do grafo. É possível aumentar e diminuir o *zoom*. No grafo de visualização completo, pode-se alterar os parâmetros (comprimento das arestas, coeficiente gravitacional e nível de profundidade para visualização) para ajustar o grafo de forma que fique o mais visível possível.

Foi criada nesse trabalho uma variação do coeficiente de colaboração para medir o coeficiente de colaboração apenas entre Docentes. Para tal cálculo foram levados em conta apenas os autores de trabalhos que fossem Docentes, desconsiderando Discentes e Participantes Externos.

4.2.2 Limitações de PANDORA

Atualmente uma das limitações de PANDORA está na geração de alguns gráficos (aqueles que mostram informações em diferentes anos), já que só alguns deles são gerados automaticamente. Isso se deve ao fato de não ser possível ler vários arquivos XML de diversos anos ao executar a aplicação apenas uma vez. Para obter o grafo referente a diversos anos, uma alternativa é passar como entrada um arquivo XML único contendo informações dos anos desejados.

4.3 Considerações Finais

Este capítulo apresentou uma proposta de análise e visualização das redes sociais acadêmicas de forma que seja possível calcular métricas, fazer análises e posteriormente com auxílio da visualização tirar conclusões sobre a rede estudada. A maior contribuição desse trabalho é permitir uma análise incremental da rede. Essa forma de representação permite que o usuário construa um mapa mental não perdendo os passos das alterações do grafo e calcule as métricas sempre que o grafo é alterado.

A maioria das ferramentas estudadas como fundamentação para esse trabalho apresentam métricas e técnicas de interação de visualização de forma individual, ou seja, calculam uma métrica e outra não. Neste trabalho, tentou-se agregar as métricas consideradas como as mais importantes para avaliação das redes sociais aplicadas estritamente ao campo de redes acadêmicas de coautoria. As tabelas 3.1 e 3.2 foram alteradas, gerando agora as tabelas 4.1 e 4.2, para cada uma foi adicionada uma linha referente à PANDORA para comparação com as outras ferramentas estudadas.

A Tabela 4.1 mostra as propriedades interativas de cada ferramenta. Das propriedades apontadas, a PANDORA não possui alteração dos rótulos, tamanhos dos nós, já que se pensou que, por se tratar de uma rede de coautoria, bastaria colocar os sobrenomes dos pesquisadores como rótulos. As consultas e o agrupamento de nós atualmente não são permitidos, mas pensa-se nisso como trabalho futuro. A Tabela 4.2 mostra as métricas calculadas por cada ferramenta. A excentricidade, diâmetro e densidade apesar de não serem calculados na versão atual, podem ser facilmente agregados, uma vez que a biblioteca JUNG permite esses cálculos.

As ferramentas estudadas são dedicadas à visualização de redes de modo geral e, desta forma, acabam não sendo especializadas em redes sociais, em especial as redes de coautoria, exceto pela SocialNetSense e ArnetMiner, esta última se dedica apenas à análise e visualização de redes acadêmicas. A ArnetMiner, porém, busca os dados de diversas fontes da Web, integrando-os, mas não permite que o usuário insira um arquivo de entrada para visualizar uma rede específica. Já a SocialNetSense se mostra um pouco mais ampla que a aplicação desse trabalho no sentido de permitir que os nós sejam não somente indivíduos mas grupo de pesquisa ou instituições, podendo o usuário dar *Zoom* e verificar quais pesquisadores pertencem ao grupo. Desta forma o SocialNetSense é mais aplicável a redes de coautoria em um sentido mais amplo que envolva, por exemplo, diferentes instituições.

Como vantagens em usar PANDORA, em relação às outras ferramentas estudadas, pode-se citar:

- Permite visualizar uma rede de coautoria específica, lendo um arquivo de entrada passado como entrada pelo usuário;
- Permite visualizar de forma geral as publicações de um grupo de pesquisadores mostrando a quantidade de publicações produzidas naquele ano;
- É possível gerar gráficos que possibilitam analisar o comportamento da rede;

- Diversas métricas são calculadas para cada nó, sendo algumas gerais (calculadas pela maioria das ferramentas de análise de redes), e outras que não foram encontradas nas outras ferramentas estudadas como, por exemplo, coeficiente de colaboração e coeficiente de colaboração apenas com Docentes.

Durante o estudo de trabalhos sobre análise de redes acadêmicas, verificou-se que muitos calculam métricas e, para visualização dos dados, utilizam a ferramenta UCINET com NetDraw ou Pajek, não tendo opção para utilizar outras ferramentas. Essas ferramentas foram desenvolvidas para visualizar e calcular métricas de redes em geral, não sendo específicas para redes acadêmicas.

Com PANDORA, um professor ou coordenador de um curso de pós-graduação, graduação ou qualquer grupo que envolva redes acadêmicas, pode analisar seu próprio desempenho ou o desempenho de um grupo em um ano específico. Além disso, exportar os dados em '.csv' permite que o usuário gere gráficos com os dados que desejar, filtrando as métricas e características do nó, ou seja, escolhendo quais métricas devem ser exportadas para o arquivo.

	Ferramentas									
Características	COMMETRIX	COSBILab Graph	ArnetMiner	SocialNetSense	Pajek	igraph	UCINET	NodeXL	SocialAction	PANDORA
<i>Zoom</i>	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim
Distância mínima entre nós	Sim	Não	Não	Não	Sim	Sim	Não	Não	Não	Sim
Comprimento das arestas	Sim	Não	Não	Não	?	Sim	Não	Não	Não	Sim
Alterar rótulos	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Não
Alterar tamanho dos rótulos	Sim	?	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não
Alterar tamanho dos nós	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não
Alterar largura das arestas	Sim	Sim	Não	Não	Sim	Sim	Sim	?	Não	Sim
Filtros	Sim	Não	Não	Sim	?	Sim	Sim	Sim	Sim	Sim
Consulta	Sim	Sim	Não	Não	Não	Não	Não	Não	Não	Não
Expande	Não	Não	Sim	Não	Não	Não	Não	Sim	Não	Sim
Agrupa	Não	Não	Não	Não	Não	Não	Não	Sim	Sim	Não
Representação	Grafo	Grafo	Grafo	Grafo	Grafo	Grafo	Grafo	Grafo	Grafo	Grafo

Tabela 4.1: Tabela com características de visualização e interatividade com acréscimo de PANDORA.

Métricas	Ferramentas									
	COMMETRIX	COSBILab Graph	ArnetMiner	SocialNetSense	Pajek	igraph	UCINET	NodeXL	SocialAction	PANDORA
Distância	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim
Excentricidade	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
Grau	Sim	Sim	Não	Sim	Sim	Sim	?	Sim	Sim	Sim
Diâmetro	Sim	Sim	Não	Sim	Sim	Sim	?	Sim	Sim	Não
Densidade	Sim	Não	Não	Sim	Não	Sim	Sim	Sim	Sim	Não
Centralidade de Grau	Sim	Não	Não	Não	Não	Sim	Sim	Não	Sim	Sim
Centralidade de <i>Closeness</i>	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Centralidade de <i>Betweenness</i>	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Coefficiente de <i>Clustering</i>	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim	Não
Assortatividade	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim (usando gráficos)
Coefficiente de Colaboração	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim

Tabela 4.2: Tabela com resumo das métricas calculadas por cada ferramenta acrescido de PANDORA.

5 ESTUDOS DE CASO

Como estudos de caso, escolheu-se trabalhar com diferentes grupos e mostrar o nicho de aplicações e estudos que podem ser feitos utilizando este trabalho. O primeiro estudo de caso, envolve alguns grupos de pesquisa dentro do PPGC da UFRGS para analisar o comportamento do mesmos, verificando se existe alguma influência devido à área de pesquisa em relação às publicações. No segundo estudo de caso, foram escolhidos alguns pesquisadores dentro do grande grupo do PPGC supondo que estes tivessem diferentes comportamentos em relação à pesquisa, por exemplo, publicar mais com alunos ou participantes externos. O terceiro e último avalia todo o PPGC como um grupo completo.

5.1 Estudo de Caso I - Grupos dentro do PPGC

Para esse estudo de caso, os docentes integrantes do PPGC da UFRGS foram divididos nos grupos de pesquisa do qual participam. Os grupos formados foram: Computação Gráfica, Inteligência Artificial, Processamento Paralelo, Sistemas Embarcados e Sistemas de Informação.

O objetivo inicial desse estudo foi calcular o coeficiente de colaboração de cada grupo, através das expansões disponíveis na PANDORA, e assim verificar de que forma ocorre a colaboração dentro dos grupos. Inicialmente, pensou-se em utilizar todo o intervalo de 2004 a 2011, mas por causa de limitações do tamanho do arquivo buscou-se a alternativa de utilizar os triênios (2004 a 2006; 2007 a 2009 e 2010 a 2011). Desta forma, seria possível avaliar de que forma os grupos se comportaram ao longo dos anos, verificando se ao final de cada triênio houve alguma alteração. Vale lembrar que o último triênio não contém o ano de 2012 pois os dados ainda serão contabilizados.

Para cada grupo, então, iniciou-se com um pesquisador, escolhido aleatoriamente e, a partir deste foram feitas expansões sempre buscando outros Docentes integrantes do grupo ao qual o primeiro pertence. Porém, percebeu-se que alguns deles seriam inalcançáveis a partir do primeiro. Isso porque existem pesquisadores que, em nenhum momento, publicaram com algum docente participante do seu mesmo grupo, ou com seus colaboradores. Ou seja, não existe caminho que ligue os dois pesquisadores no grafo completo, pois eles participam de componentes desconexas. Outro fator notado é que muitos dos pesquisadores pertencentes ao grupos não apareciam ou apareciam

Grupos	Anos		
	2004 a 2006	2007 a 2009	2010 a 2011
Computação Gráfica	0,685 (4 professores)	0,67 (5 professores)	0,68 (8 professores)
Sistemas Embarcados	0,7 (4 professores)	0,73 (4 professores)	0,76 (4 professores)
Sistemas de Informação	0,66 (6 professores)	0,69 (7 professores)	0,73 (9 professores)

Tabela 5.1: Coeficientes de Colaboração dos grupos em cada triênio.

como participante externo, no conjunto de publicações de entrada para PANDORA, mesmo que tenham se tornado Docentes ao longo do tempo.

Para lidar com o problema dos grupos com componentes desconexas, para esse estudo, foram escolhidos apenas os grupos que correspondiam a apenas uma componente conexa representando os docentes. Assim sendo, os grupos escolhidos para essa análise foram: Computação Gráfica, Sistemas Embarcados e Sistemas da Informação. A tabela 5.1 mostra os valores dos coeficientes de colaboração calculados. Para cada valor, está marcada também a quantidade de pesquisadores Docentes e participantes daquele grupo que foram expandidos. Cada coeficiente de colaboração teve uma pequena variação, apesar de terem aumentado a cada triênio nos grupos de Sistemas Embarcados e Sistemas de Informação. No grupo de Computação Gráfica houve uma queda nos anos de 2007 a 2009, mas voltou a aumentar de 2010 a 2011. É importante ressaltar que para o cálculo desses coeficientes de colaboração, as quantidades de autores levam em conta todos os tipos de pesquisadores possíveis (Participantes Externos, Docentes, Discentes). Isso significa que o aumento do coeficiente de colaboração não significa que os Docentes pertencentes ao grupo tenham aumentado a colaboração entre si necessariamente.

O grupo que manteve mais professores publicando proximamente foi o grupo de Sistemas de Informação, que aumentou a cada triênio a quantidade de pesquisadores e também o coeficiente de colaboração. O grupo de Inteligência Artificial foi o que apresentou mais fragmentação sendo que, no último triênio, apresentou 6 grupos separados para os pesquisadores do grupo, sendo que a maioria continha 2 dos pesquisadores.

Apesar do coeficiente de colaboração dos pesquisadores do grupo de Sistemas de Informação ter aumentado, através do grafo de Artigos é possível perceber que os artigos de cada Docente não costumam incluir outros pesquisadores do grupo, como pode ser visto na Figura 5.1. Pode-se ver que existem círculos mais separados, isto significa que os artigos não têm muitos autores em comum, ou seja, o coeficiente de colaboração entre Docentes do grupo é pequeno.

Em comparação ao grupo de Sistemas de Informação, pode-se observar o grafo de Artigos, na Figura 5.2, do grupo de Computação Gráfica no triênio de 2004 a 2006. Neste grafo é possível perceber que fica mais difícil separar os círculos correspondentes a cada pesquisador, ou seja, estão mais integrados, com maiores intersecções, o que significa maior colaboração entre os Docentes é maior. No triênio de 2007 a 2009 o grupo de Computação Gráfica se comportou da mesma forma que o grupo de Sistemas da Informação de 2004 a 2006.

No grupo de Sistemas Embarcados foi encontrado o pesquisador com maior número de pu-

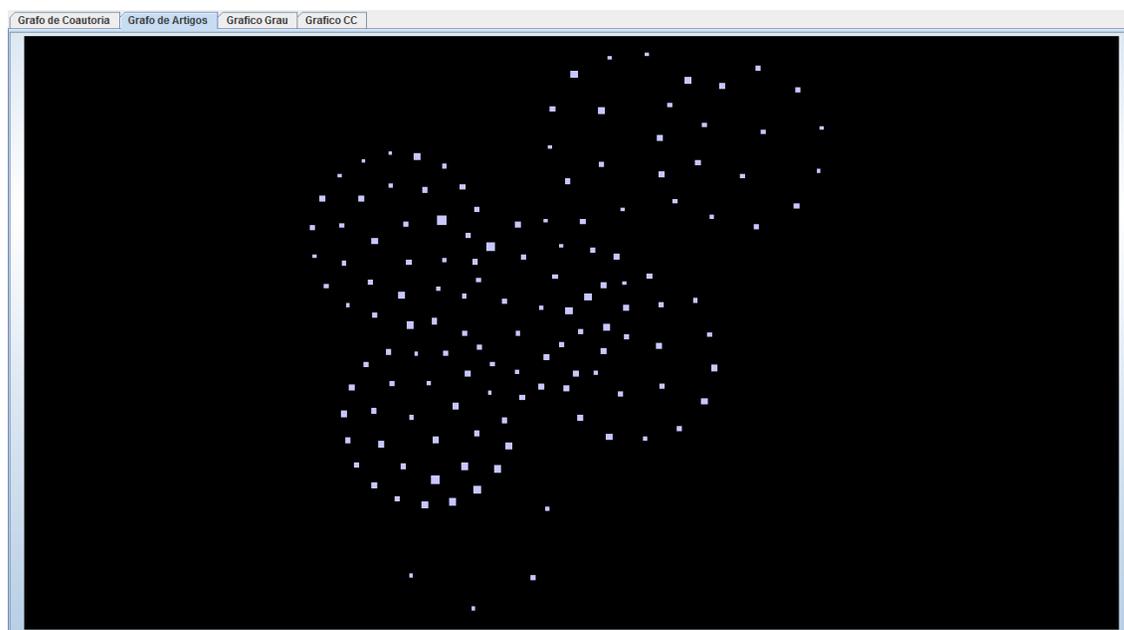


Figura 5.1: Grafo de artigos do grupo de Sistemas da Informação no triênio de 2007 a 2009.

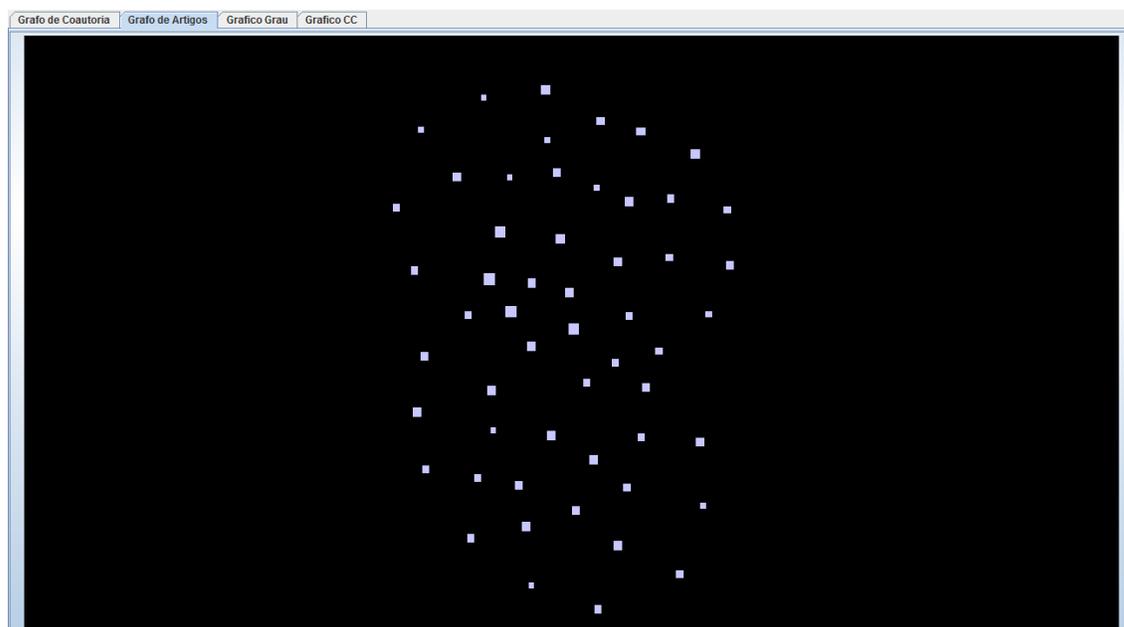


Figura 5.2: Grafo de artigos do grupo de Computação Gráfica no triênio de 2004 a 2006.

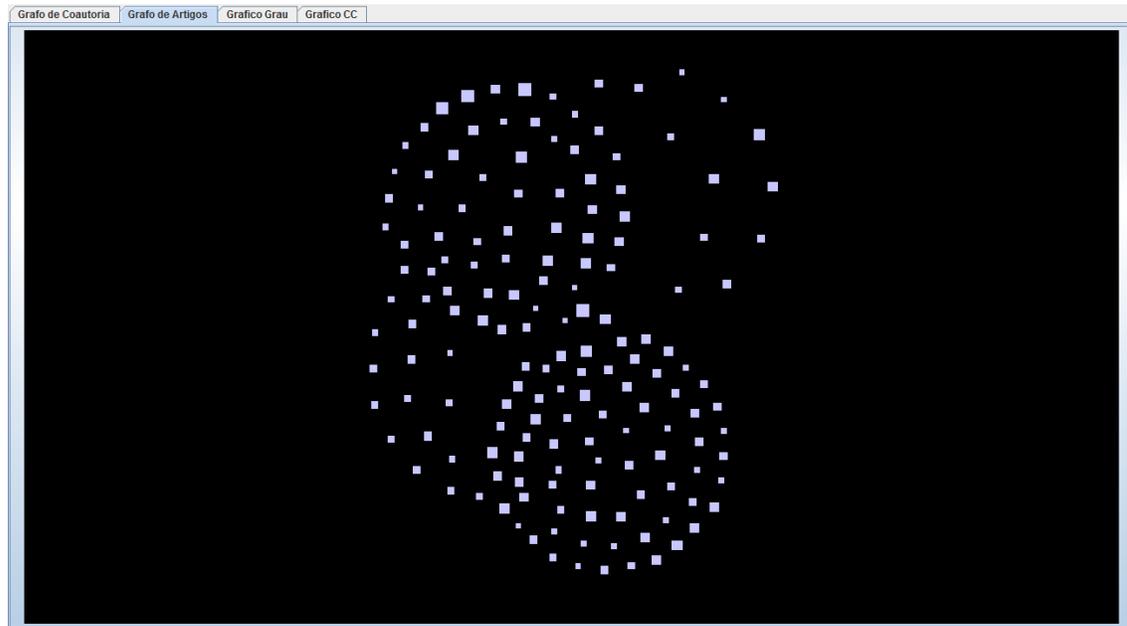


Figura 5.3: Grafo de artigos do grupo de Sistemas Embarcados no triênio de 2007 a 2009.

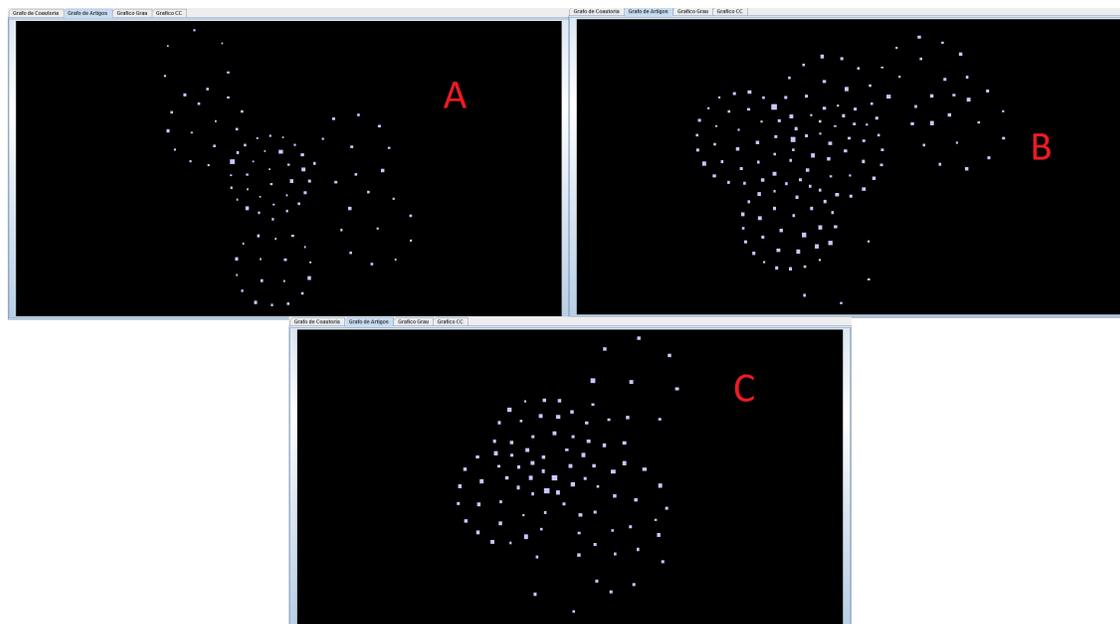


Figura 5.4: Evolução dos grafos de colaborações nos anos de 2004 a 2011 do grupo de Sistemas de Informação. Legenda: A - triênio de 2004 a 2006, B - triênio de 2007 a 2009 e C - triênio incompleto contendo 2010 e 2011.

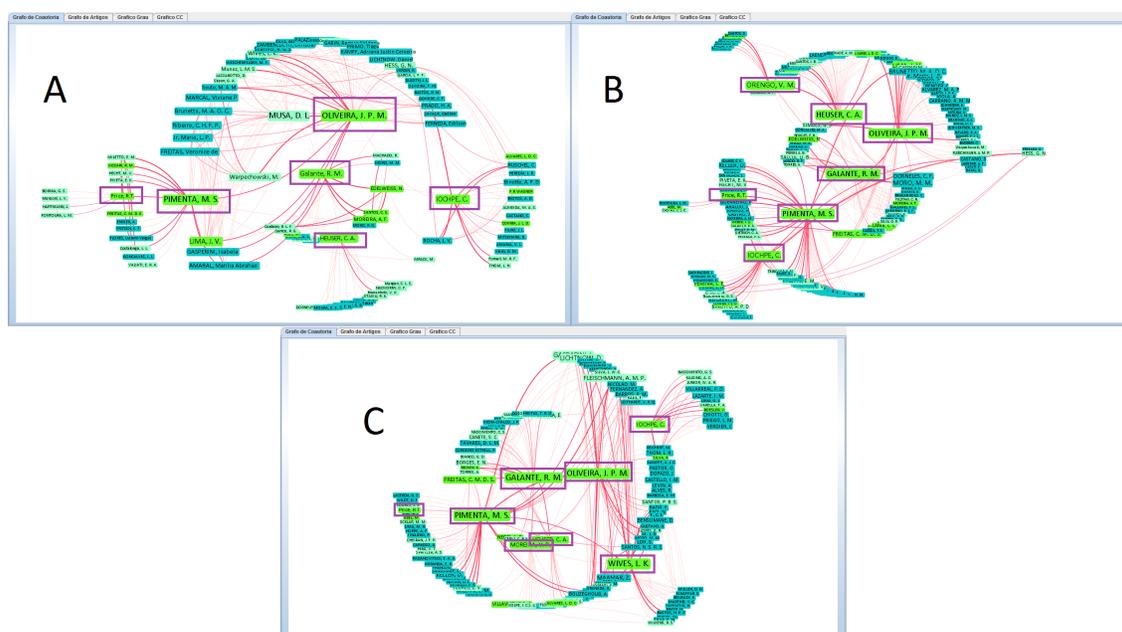


Figura 5.5: Evolução nos grafos de colaboração entre os anos de 2004 a 2011 para o grupo de Sistemas de Informação. Legenda: A - triênio de 2004 a 2006, B - triênio de 2007 a 2009 e C - triênio incompleto contendo 2010 e 2011.

blições em um triênio, Luigi Carro, com 84 publicações, isso pode ser visto na Figura 5.3. Os círculos mostram-se muito mais densos, mostrando que existem muitas publicações com um determinado pesquisador como autor.

Para mostrar as comparações sobre evoluções de um grupo durante os triênios, os grafos de publicações dos triênios são mostrados na Figura 5.4 para o grupo de Sistemas de Informação, já que para esse grupo as mudanças no grafo são visíveis. Além disso, o grupo de Sistemas Embarcados possui uma concentração muito grande de publicações em dois pesquisadores em especial, como foi mostrado na Figura 5.3. Já o grupo de Computação Gráfica apresentou pequena variação no coeficiente de colaboração nos três triênios e como a intenção nesse momento é mostrar a relação entre o grafo de publicações e o coeficiente de colaboração, este não seria um bom exemplo. A Figura 5.4, mostra os grafos de publicações para o grupo de Sistemas de Informação em cada triênio. A cada triênio a colaboração entre os pesquisadores do grupo aumentou. No primeiro triênio, os círculos estão mais dispersos, sendo mais unidos no segundo triênio e mais ainda no último. Isso é comprovado na tabela 5.1, em que o ano em que foi registrado o maior coeficiente de colaboração entre o grupo.

A Figura 5.5, mostra os grafos de colaboração em cada triênio estudado para o grupo de Sistemas de Informação. Os pesquisadores destacados com caixas roxas representam os pesquisadores que formam o grupo de Sistemas de Informação. Com o passar dos anos os docentes do grupo se tornaram mais próximos em relação às colaborações entre si.

5.2 Estudo de Caso II - Pesquisadores Individuais PPGC

Para esse gráfico são levados em conta os anos de 2004 a 2011, calculando as métricas de cada um dos 6 pesquisadores. Por questões de sigilo, eles serão chamados de Professor A, Professor B, Professor C, Professor D, Professor E e Professor F. Para cada um desses pesquisadores foram calculadas métricas utilizando a PANDORA. A tabela 5.2 mostra as médias de quantidades de pesquisadores e da quantidade de colaboradores, assim como a porcentagem da quantidade de colaboradores que são Participantes Externos.

Métricas	Pesquisadores					
	Professor A	Professor B	Professor C	Professora D	Professora E	Professora F
Média da quantidade de publicações	20	8	6	13	8	12
Média da quantidade de colaboradores	35	13	9	16	20	18
Participantes Externos (em porcentagem)	47,7%	48,5%	52%	54,8%	43,5%	56,8%
Docentes (em porcentagem)	22,3%	25,2%	21,3%	12,9%	18,4%	19,9%

Tabela 5.2: Métricas dos pesquisadores individuais escolhidos.

O pesquisador que obteve maior quantidade média de publicações foi o Professor A, assim como também obteve a maior quantidade de publicações. Mas não necessariamente isso ocorre.

Por exemplo, o Professor E apesar de possuir maior quantidade média de colaboradores que o Professor D, possui menor quantidade média de publicações por ano. Para o Professor A, a maior quantidade de publicações registrada aconteceu em 2005, enquanto para os outros pesquisadores isso aconteceu no ano de 2007. A quantidade de Participantes Externos em suas colaborações chamou atenção, esperava-se que os pesquisadores se diferenciasssem mais em relação ao tipo de colaboradores, mas para todos os seis escolhidos, que teoricamente possuem diferentes comportamentos em relação ao tipo de colaboradores, a quantidade de colaboradores como Participantes Externos foi maior. Observa-se, entretanto, que alguns dos Participantes Externos declarados nos cadastros, são na verdade ex-alunos que antes eram orientandos. O ideal seria de alguma forma ter uma quarta categoria ou tratá-los como Discentes até um tempo pré-determinado. Para a média de colaboradores como Docentes, percebe-se que a maior porcentagem foi de 25%. Isso mostra que os pesquisadores Docentes publicam muito pouco em colaboração com outros Docentes. O coeficiente de colaboração entre Docentes se manteve no máximo em 0,5, exceto para o Professor A no ano de 2010. O coeficiente de colaboração geral variou de 0,54 a 0,83, sendo que o maior coeficiente foi registrado em 2005 para o Professor E.

A Figura 5.6 mostra gráficos de barras com as métricas calculadas para cada um dos seis pesquisadores escolhidos. Foi utilizada escala logarítmica para que as métricas pudessem ser colocadas no mesmo gráfico, uma vez que algumas das métricas são calculadas de forma normalizada e outras não tem como serem normalizadas. Vale ressaltar que as métricas calculadas pelo JUNG (centralidades), levam em conta apenas para o cálculo a componente em que o nó está, isso porque para o cálculo dessas métricas são utilizadas as distâncias e, se não é possível alcançar um nó a partir de outro, então a distância entre eles não é relevante, é considerada 0. Então, existem pesquisadores que terão altos valores para centralidades apesar de estarem conectados a poucos nós, mas como suas componentes possuem poucos nós, terão alto valor para centralidade.

5.3 Estudo de Caso III - Grupo Completo PPGC 2004 a 2011

A Figura 5.7 mostra que, em 2007, foi registrada a maior quantidade de Anais e Artigos e que, em 2009, foi registrada a maior quantidade de capítulos de livros publicada. Em geral a quantidade de publicações e de colaboradores varia de maneira semelhante, não sendo esse o único fator determinante para maior quantidade de publicações.

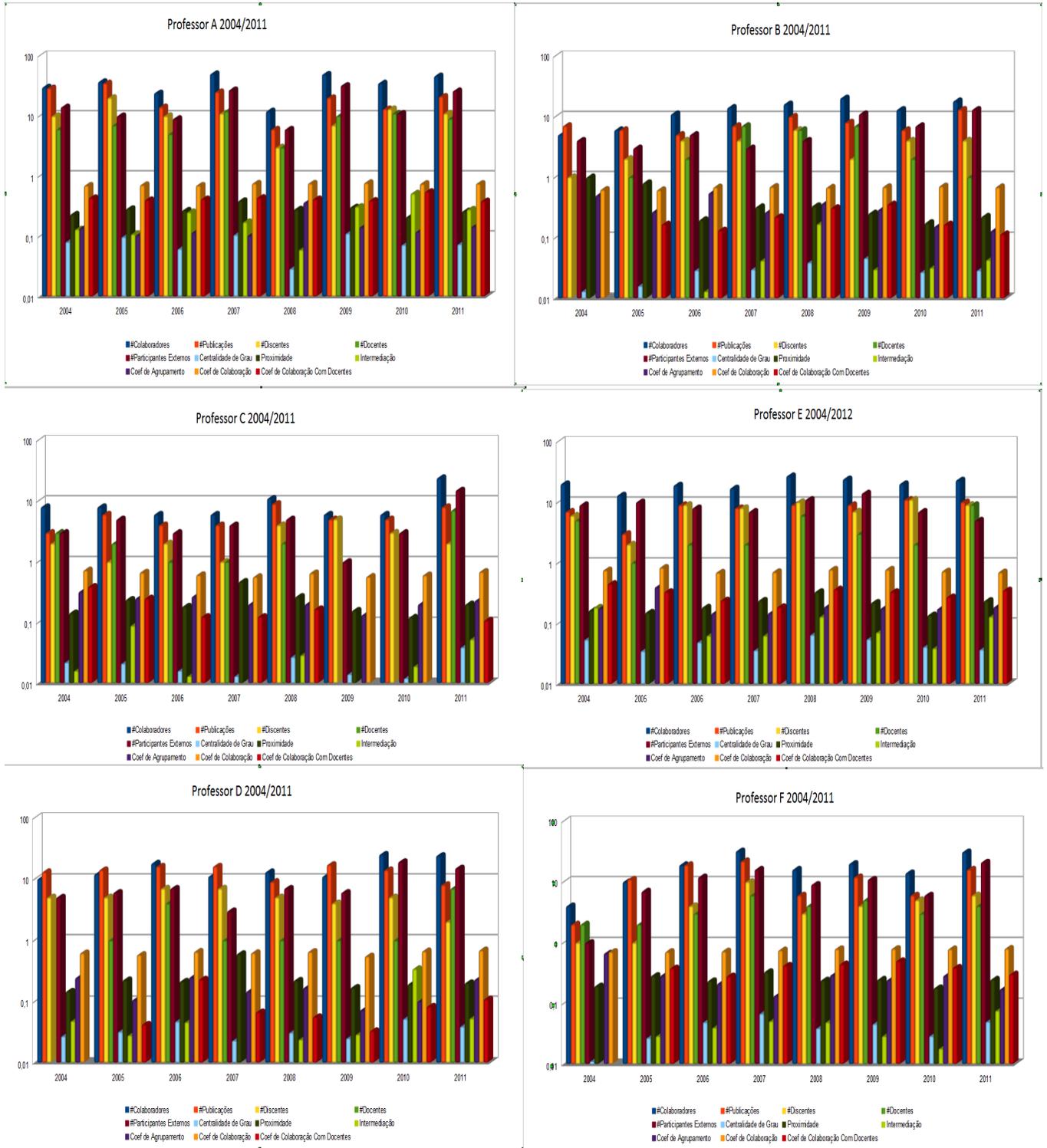


Figura 5.6: Gráficos de barras com métricas para cada um dos pesquisadores escolhidos.

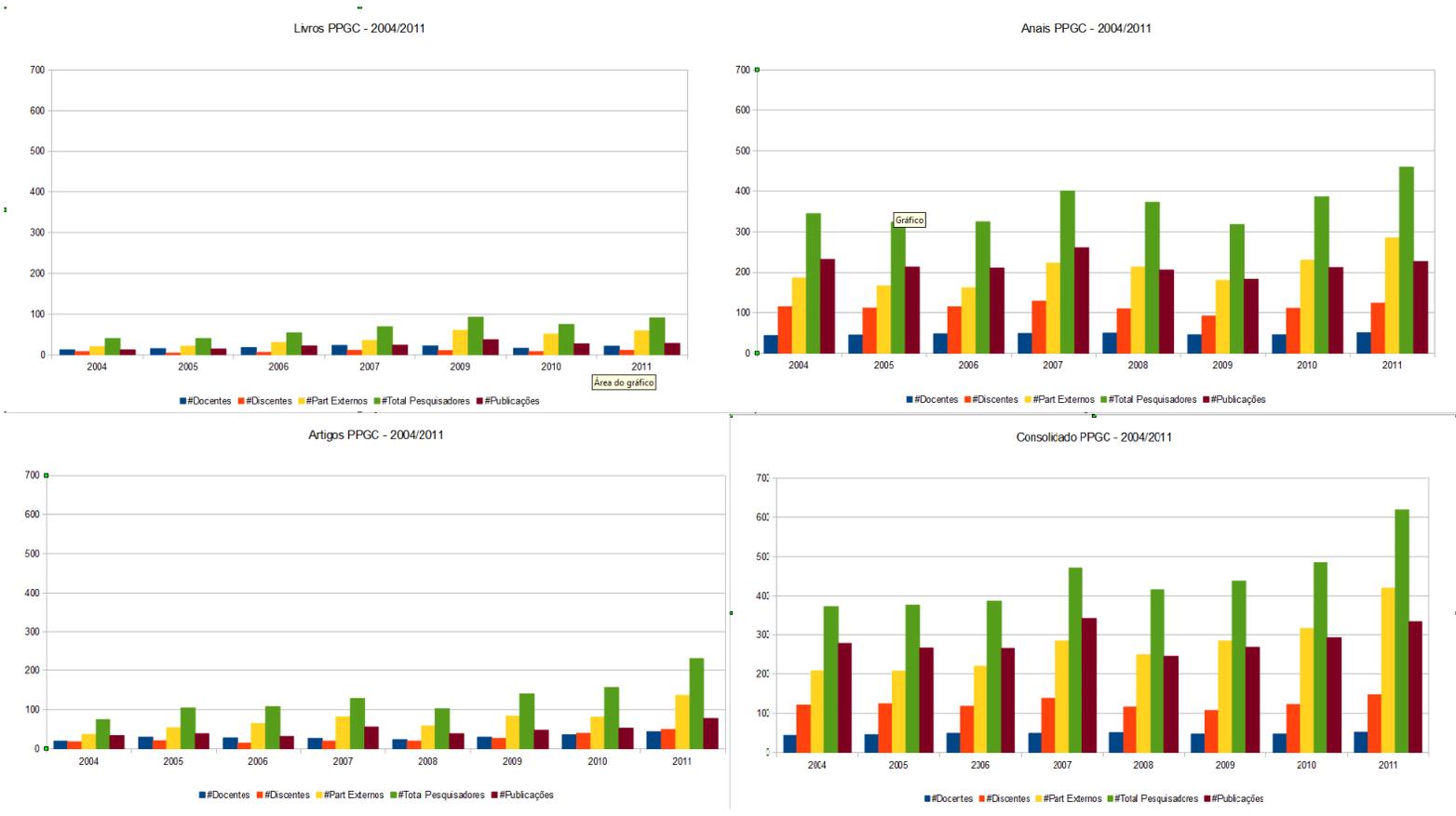


Figura 5.7: Gráficos das métricas referentes aos anos 2004 a 2010.

Na maior parte do tempo, a quantidade de Docentes envolvidos no desenvolvimento de capítulos de livros é de no máximo metade da média dos envolvidos com Anais. Isso mostra a maior importância dada aos Anais e Livros. Em geral, a quantidade de Discentes envolvidos com capítulos de livros é cerca de 10% do valor dos envolvidos com Anais. A quantidade de Participantes Externos é sempre superior às outras categorias de pesquisadores, o que já era esperado.

Os coeficientes de colaboração entre pesquisadores em geral varia entre 0,6 e 0,7, o que é razoável. Porém, a taxa de colaboração entre Docentes varia de apenas 0,1 a 0,2 o que reforça a idéia de que pouco se publica entre Docentes.

O coeficiente de colaboração para os anais se manteve praticamente constante nos 7 anos, já a quantidade de Docentes envolvidos com capítulos de livros se manteve praticamente constante com cerca de 50 pesquisadores em média durante o período. Em 2007, foi registrado o maior número de publicações, sendo registrado aumento da quantidade de pesquisadores, em relação aos anos anteriores. Apesar terem sido registrados nos anos de 2010 e 2011 valores maiores de pesquisadores mas as quantidades de publicações nesses anos ainda foram menores do que em 2007. A maior influência na quantidade de pesquisadores, em geral, é a quantidade de Participantes Externos, sendo que os dois gráficos são muito parecidos em relação ao crescimento. Os dois anos que registraram maior quantidade de publicações foram também os anos em que foram calculados os maiores coeficientes de colaboração entre Docentes, 2007 e 2011.

De 2005 a 2009 houve crescimento considerável na quantidade de pesquisadores em geral envolvidos com capítulos de livros, acompanhado da quantidade de publicações e participantes externos. Entre 2005 e 2007 o coeficiente de colaboração entre docentes envolvidos com capítulos de livros cresceu. Já em 2009 e 2010 houve queda nos gráficos de quantidades de publicações e pesquisadores, mas o coeficiente de colaboração geral apresentou pequeno crescimento.

Em 2006 e 2008, houve queda na quantidade de publicações, participantes externos, docentes e pesquisadores em geral envolvidos com artigos. Em 2006, foi registrada a menor quantidade de publicações realizadas, sendo que em 2007 foi registrada o maior valor para a mesma métrica. Em 2010, o coeficiente de colaboração para artigos foi o menor registrado. No intervalo de 2008 a 2010 a quantidade de artigos cresceu de forma considerável, de 100 para 160. O menor coeficiente de colaboração entre Docentes para artigos registrado aconteceu em 2010.

A Figura 5.8 mostra o gráfico Grau x Quantidade de Publicações e mostra que a maior quantidade de pesquisadores se concentra na faixa de até 2 publicações e até grau 13, que é onde encontra-se a maior parte de retângulos em vermelho escuro. Apenas 10 pesquisadores encontram-se na faixa demarcada de acima de 14 publicações e grau maior que 21. O gráfico de Assortatividade, mostrado na Figura 5.9, mostra que a média dos graus dos vizinhos está acima de 7, o que mostra que a maioria dos pesquisadores tendem a se relacionar com pesquisadores que tenham grande quantidade de colaboradores. Para essas duas análises foram utilizados gráficos referentes ao ano de 2011 que é o mais recente.

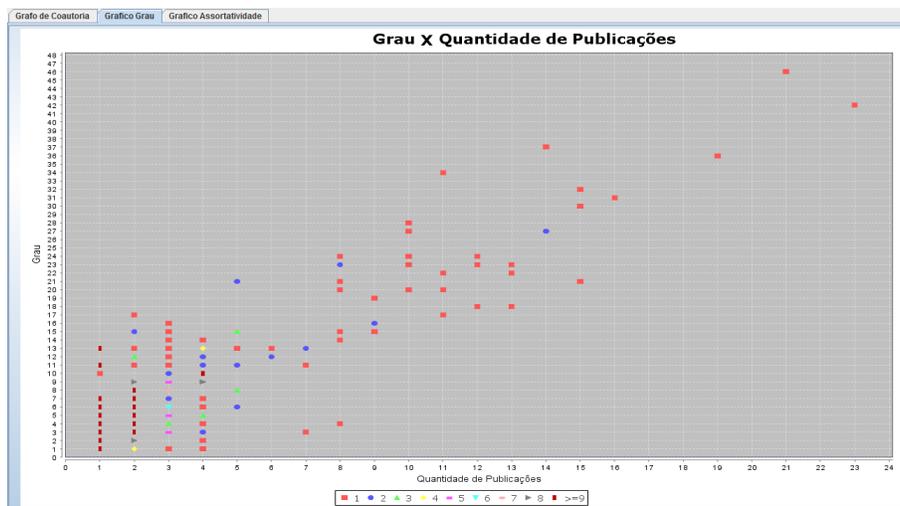


Figura 5.8: Gráfico Grau X Quantidade de Publicações para o PPGC no ano de 2011.

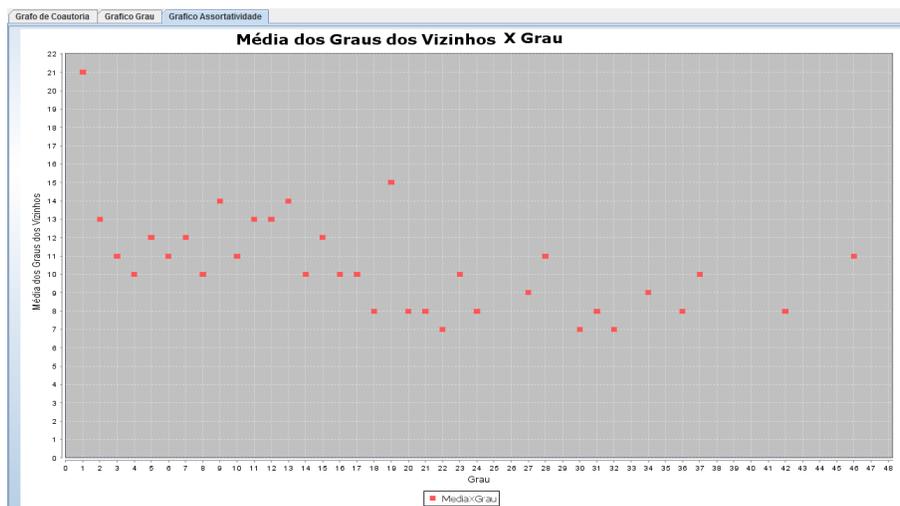


Figura 5.9: Gráfico de Assortatividade do ano de 2011 para todo o PPGC.

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho propõe um método para análise e visualização de grafos que representam redes sociais acadêmicas. Com esse método, o usuário pode visualizar grafos de duas formas: grafo completo com todos os pesquisadores ou, incrementalmente, a partir de um pesquisador com expansões mostrando apenas os colaboradores do pesquisador escolhido pelo próprio usuário.

Dado um arquivo GraphML, arquivo XML específico para representar grafos, a ferramenta gera a visualização do grafo completo ou individual, de acordo com a escolha do usuário. A partir daí são permitidas interações, desde o *zoom* e visualização de informações até alterações nas forças que atuam sobre os nós. Além disso, são gerados dados que podem ser utilizados para interpretação do grafo e para posteriores estudos sobre o comportamento dos indivíduos na rede.

Como estudos de caso, foram utilizados dados referentes ao grupo do PPGC da UFRGS, já que estavam totalmente disponíveis para acesso. Assim, foram desenvolvidos 3 estudos de caso. O primeiro dividiu o grupo de acordo com as sub-áreas de pesquisa, o segundo estudo de caso comparou alguns pesquisadores escolhidos sugerindo que tivessem diferentes comportamentos em relação à escolha dos seus colaboradores (Discentes, Docentes ou Participantes Externos). Já o terceiro estudo de caso analisou o grupo completo do PPGC da UFRGS, no período de 2004 a 2011.

Os resultados apresentados por esse trabalho mostraram algumas conclusões específicas para a rede do PPGC da UFRGS que podem ser usados para melhorar o desempenho de seus integrantes. Além disso, pode ajudar os pesquisadores a perceber características próprias sobre seu comportamento em relação às suas colaborações. Desta forma, estes pesquisadores podem melhorar sua produção analisando os grafos e verificando o que poderia melhorar para chegar nesse objetivo, por exemplo, diversificar sua colaboração, ou buscar colaborações com colaboradores de seus colaboradores, ou buscar produzir mais trabalhos com colegas de sua área de pesquisa.

Diversas idéias podem ser incorporações futuras à PANDORA. Uma delas seria facilitar a entrada de vários arquivos xml simultaneamente para gerar grafos temporais permitindo verificar evoluções em anos. Outra idéia seria automatizar a geração dos gráficos que aqui foram produzidos com o auxílio do Excel. Também poderia ser possível filtrar as publicações para visualizar apenas artigos, periódicos, etc. Sugestões de colaboradores poderiam ser feitas automaticamente

pela aplicação. Uma última sugestão de trabalho futuro seria diferenciar os alunos que saíram há pouco tempo, tratando-os numa categoria especial, de modo que a mudança de papel passa a ser acompanhada.

REFERÊNCIAS

- BATAGELJ, V.; MRVAR, A. **Program for Analysis and Visualization of Large Networks**. [S.l.: s.n.], 2011.
- BENEVENUTO, F.; ALMEIDA, J.; SILVA., A. **Explorando Redes Sociais Online: da coleta e análise de grandes bases de dados às aplicações**. Campo Grande, Brazil., 2011.
- BORGATTI, M. E. S.; FREEMAN, L. **UCINET for Windows**. 2002.
- BRANDÃO, W. C.; PARREIRAS, F. S.; SILVA, A. B. de Oliveira e. **Redes em Ciência da Informação: evidências comportamentais dos pesquisadores e tendências evolutivas das redes de coautoria**. **Informação & Informação**, [S.l.], v.12, 2007.
- CSARDI, G.; NEPUSZ, T. The igraph software package for complex network research. **InterJournal**, [S.l.], v.Complex Systems, p.1695, 2006.
- FREEMAN, L. Centrality in social networks conceptual clarification. **Social networks**, [S.l.], v.1, n.3, p.215–239, 1979.
- FREEMAN, L. C. Social Network Visualization, Methods of. In: **Encyclopedia of Complexity and Systems Science**. [S.l.: s.n.], 2009. p.8345–8363.
- FÁTIMA S. MAIA, M. de; CAREGNATO, S. E. Co-autoria como indicador de redes de colaboração científica. **Perspectivas em Ciência da Informação**, [S.l.], v.13, 2008.
- GOU, L. et al. Semantic Interaction for Sensemaking: inferring analytical reasoning for model steering. **VizWeeK**, [S.l.], 2012.
- GOU, L.; ZHANG, X. L. TreeNetViz: revealing patterns of networks over tree structures. **IEEE Trans. Vis. Comput. Graph.**, [S.l.], v.17, n.12, p.2449–2458, 2011.
- HAYASHI, M. C. P. I.; HAYASHI, C. R. M.; LIMA, M. Y. de. Análise de redes de co-autoria na produção científica em educação especial. **Liinc em revista**, [S.l.], v.4, 2008.

HERMAN, I.; MELANCON, G.; MARSHALL, M. S. Graph visualization and navigation in information visualization: a survey. **IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS**, [S.l.], v.6, n.1, p.24–43, 2000.

JAMAKOVIC, A.; UHLIG, S. On the relationships between topological measures in real-world networks. **NHM**, [S.l.], v.3, n.2, p.345–359, 2008.

LIMA, M. Y. de. **Redes de Co-autoria Científica no Programa de Pós-Graduação em Geociências da UFRGS**. 2009. Dissertação (Mestrado em Ciência da Computação) — UFRGS.

NASCIMENTO, C. S. do et al. Visualização interativa de redes sociais: um estudo de caso em redes de colaboração científica. **Conferencia Latinoamericana de Informática - CLEI**, [S.l.], 2011.

NEWMAN, M. E. J. Assortative Mixing in Networks. **Physical Review Letters**, [S.l.], v.89, n.20, p.208701, Oct. 2002.

NORTHWAY, M. **A primer of sociometry**. [S.l.]: University of Toronto Press, 1952.

OLIVEIRA SILVA, A. B. de et al. Redes de co-autoria dos professores da Ciência da Informação: um retrato da colaboração científica dessa disciplina no brasil. **VII ENANCIB - Encontro Nacional de Pesquisa em Ciência da Informação**, [S.l.], Novembro 2006.

PERER, A.; SHNEIDERMAN, B. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In: **SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS**, New York, NY, USA. **Proceedings...** ACM, 2008. p.265–274. (CHI '08).

SMITH, M. et al. Analyzing (social media) networks with NodeXL. In: **COMMUNITIES AND TECHNOLOGIES**. **Proceedings...** ACM, 2009. p.255–264.

SOUSA, F. S.; RUIZ, E. E. S. Aplicação da teoria de redes complexas no estudo de relacionamento entre doenças em casos de óbito do paciente. **SBI 2011**, [S.l.], 2011.

TANG, J. et al. ArnetMiner: extraction and mining of academic social networks. In: **ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING**, 14., New York, NY, USA. **Proceedings...** ACM, 2008. p.990–998. (KDD '08).

TRIER, M.; BOBRIK, A. Social Search: exploring and searching social architectures in digital networks. **IEEE Internet Computing**, [S.l.], v.13, n.2, p.51–59, 2009.

VALENTINI, R.; JORDÁN, F. Software, Data and Modelling News: cosbilab graph: the network analysis module of cosbilab. **Environ. Model. Softw.**, Amsterdam, The Netherlands, The Netherlands, v.25, n.7, p.886–888, July 2010.

WASSERMAN, S.; FAUST, K. **Social network analysis**: methods and applications. [S.l.]: Cambridge University Press, 1994. (Structural analysis in the social sciences).