



ANÁLISE DE REDE SOCIAL: ESTUDO COMPARATIVO ENTRE FERRAMENTAS E PROPOSTA DE APLICAÇÃO NA ÁREA DE EDUCAÇÃO BÁSICA

Trabalho de Conclusão de Curso

Engenharia da Computação

José Maurílio da Silva Júnior

Orientador: Prof. Dr. Genésio Gomes Cruz Neto

Co-orientador: Prof. Dr. Ricardo Bastos Cavalcante Prudêncio



**Universidade de Pernambuco
Escola Politécnica de Pernambuco
Graduação em Engenharia de Computação**

JOSÉ MAURÍLIO DA SILVA JÚNIOR

**ANÁLISE DE REDE SOCIAL:
ESTUDO COMPARATIVO ENTRE
FERRAMENTAS E PROPOSTA DE
APLICAÇÃO NA ÁREA DE
EDUCAÇÃO BÁSICA**

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Recife, novembro de 2011.

MONOGRAFIA DE FINAL DE CURSO

Avaliação Final (para o presidente da banca)*

No dia 20 de Dezembro de 2011, às 9:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente JOSÉ MAURILIO DA SILVA JUNIOR, orientado pelo professor Genésio Gomes da Cruz Neto, sob título Análise de rede social: estudo comparativo entre ferramentas e propostas de aplicação na educação básica, a banca composta pelos professores:

Edison de Queiroz Albuquerque

Genésio Gomes da Cruz Neto

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

Aprovada

Aprovada com Restrições*

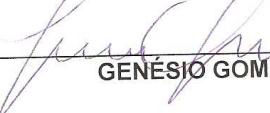
Reprovada

e foi-lhe atribuída nota: 9,0 (MVL)

*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 7 dias para entrega da versão final da monografia a contar da data deste documento.


EDISON DE QUEIROZ ALBUQUERQUE


GENÉSIO GOMES DA CRUZ NETO

* Este documento deverá ser encadernado juntamente com a monografia em versão final.

Dedico este trabalho a meus pais, que nunca mediram esforços para me garantir uma educação de qualidade.

Agradecimentos

Agradeço à minha família e à minha namorada pelo apoio irrestrito em todos os aspectos de minha vida. Desejo também registrar meu agradecimento a:

- Meus colegas de trabalho, que me deram significativa ajuda na concepção desta monografia.
- Meu orientador e a meu co-orientador pela dedicação a fim de que meu trabalho se realizasse da melhor maneira possível.
- Joy Street por ter permitido que abordasse algumas ideias concebidas dentro da empresa.
- Meus colegas e professores da universidade que de alguma forma tenham contribuído para minha formação.
- Meu amigo e companheiro de curso Rodrigo Mendes, que teve grande importância para a minha graduação.

Resumo

O conceito de rede social está em grande evidência atualmente e isso tem despertado o interesse de muitos estudos nessa área. Esse campo de estudo é chamado de Análise de Rede Social (ARS) e ele tem fundamental importância no entendimento dos fenômenos que envolvem essas agregações sociais. Nesse cenário, são necessárias ferramentas que dêem suporte a essas análises, são as chamadas ferramentas de Análise de Rede Social. Diante da escassez de material didático (principalmente escrito em português) para profissionais e demais interessados em utilizar ferramentas de análise de rede social, este trabalho pretende servir como um guia inicial e um incentivo ao uso desses softwares. Para isso, será realizado um estudo comparativo entre algumas das diversas ferramentas existentes e as informações relevantes serão condensadas em forma de quadro comparativo. Além disso, este trabalho apresentará a proposta de implantação de uma ferramenta de ARS em uma rede social com fins educacionais, o que servirá para melhor compreender aplicações de ferramentas de análise de rede social para fins práticos.

Abstract

The concept of social network is much in evidence nowadays, and this has aroused the interest of many studies in this area. This field of study is called Social Network Analysis (SNA) and it has fundamental importance in the understanding of phenomena involved in these social aggregations. In this scenario, some tools are necessary for supporting these analysis, they are called Social Network Analysis tools. Given the scarcity of didactic material (mainly written in Portuguese) for professionals and others interested in using social network analysis tools, this paper aims to serve as an initial guide and an incentive to use of these kind of software. To achieve this, here will be performed a comparative study of some of the many existing tools and relevant information will be condensed in the form of comparative table. In addition, this work shows a proposal of deployment of SNA tool in a educational purposes social network, what will give the reader better comprehension about the use of such tools for practical purposes.

Sumário

Capítulo 1 Introdução	1
1.1 Objetivos da Pesquisa.....	2
1.1.1 Objetivo Geral:	2
1.1.2 Objetivos Específicos:	2
1.2 Metodologia e Estratégia de Ação.....	3
1.3 Resultados e Impactos Esperados	3
1.4 Organização do texto.....	4
Capítulo 2 Análise de Rede Social	5
2.1 Evolução das Redes Sociais	5
2.2 Conceitos de Teoria dos Grafos.....	6
2.2.1 Grafo Não Direcional	7
2.2.2 Grafo Direcional (Digrafo)	8
2.3 Análise de Rede Social.....	9
2.3.1 Grau (<i>degree</i>).....	11
2.3.2 Intermediação (<i>betweenness</i>):.....	12
2.3.3 Proximidade (<i>closeness</i>).....	14
2.3.4 Outras medidas de centralidade.....	15
2.4 Ferramentas e Aplicações de ARS	16
2.4.1 Exemplos de Análise de Rede Social em Aplicativos Sociais	16
2.5 Considerações finais	17
Capítulo 3 Comparativo de Ferramentas	18
3.1 Ferramentas de Análise de Rede Social	18
3.2 Gephi.....	20
3.3 Cytoscape	24
3.4 JUNG	26
3.5 GUESS.....	28
3.6 Comparativo das ferramentas.....	30
3.6.1 Critérios de comparação.....	31
3.6.2 Quadro Comparativo das Ferramentas	32
3.6.3 Análise dos Resultados.....	33
3.7 Considerações Finais	34
Capítulo 4 Proposta de Ferramenta para a Rede Social da OjE.....	36

4.1	O que é a OjE.....	36
4.2	Proposta de Ferramenta de ARS da OjE.....	38
4.2.1	Criando e exibindo a visualização do grafo da rede social	38
(a)	Extração dos dados	39
(b)	Criação do arquivo do grafo	40
(c)	Integração do GEXF Explorer com código HTML e exibição no navegador.....	41
4.2.2	Funcionalidades da ferramenta de ARS da OjE	43
4.3	Considerações Finais	48
Capítulo 5 Conclusão e Trabalhos Futuros.....		49
5.1	Conclusão.....	49
5.2	Trabalhos Futuros	49
Bibliografia		51
Apêndice A		54

Índice de Figuras

Figura 1 - Grafo de relação “mora perto de” entre seis pessoas	8
Figura 2 - Exemplo de grafo direcionado (digráfo)	9
Figura 3 – Exemplos de diferentes topologias de rede.	11
Figura 4 - Exemplo de análise de rede com o software Gephi.....	20
Figura 5 - Interface do Gephi e dados analíticos da rede.....	21
Figura 6 - Realce dos adjacentes de um nó.....	22
Figura 7 - Descoberta de caminho mais curto pelo Gephi	22
Figura 8 - Rede modelada com o algoritmo force-based.	23
Figura 9 - Agrupamento de nós pelo Gephi	23
Figura 10 - Filtros de nós e arestas.....	24
Figura 11 - Exemplos de customização de nós no Cytoscape.....	26
Figura 12 - Exploração de grafo misto e algoritmos de layout com o JUNG.	27
Figura 13 - Aplicativo que utiliza o JUNG e mostra o menor caminho entre nós.	28
Figura 14 - Tela de exportação de arquivos do GUESS	29
Figura 15 - Tela principal de uma aplicação baseada no GUESS.....	30
Figura 16 – Tela de um dos jogos da plataforma.	37
Figura 17 – Tela de um dos enigmas da plataforma.	37
Figura 18 – Passos para geração e exibição do grafo da rede.	39
Figura 19 – Diagrama Entidade Relacionamento do banco criado (a), dados na tabela usuário (b) e dados da tabela relacionamento (c)	40
Figura 20 – Código HTML para exibir rede no navegador.	42
Figura 21 - Página HTML que exibe o grafo da rede.	43
Figura 22 – Tela principal da ferramenta de monitoramento de OjE	44
Figura 23 – Representação gráfica da rede da cidade de Recife.....	45

Figura 24 – Modo tela cheia da ferramenta exibindo grafo da rede de Recife 47

Índice de Tabelas

Tabela 1. Tabela comparativa entre ferramentas.....	32
--	----

Tabela de Símbolos e Siglas

API - Application Programming Interface

AR - Análise de Redes

ARS - Análise de Redes Sociais

HITS - Hypertext Induced Topic Search

URI - Uniform Resource Locator

XML - Extensible Markup Language

Capítulo 1

Introdução

Sob uma ótica ampla, uma rede social pode ser descrita como um conjunto de atores e os relacionamentos estabelecidos entre eles. De acordo com Wasserman et al. [1], esses atores podem ser uma pessoa, um conjunto discreto de pessoas agregadas em uma unidade social, uma empresa, agências de serviço público de uma cidade, estados-nações do mundo dentre outros. As redes sociais não são dependentes do ciberespaço, portanto os atuais e tão populares aplicativos sociais - ou redes sociais *on line*, ou *sites* de rede social - são apenas um novo local em que se pode desenvolver o comportamento social [2].

Atualmente, esse conceito de rede social está em bastante evidência, notadamente em função da popularidade desses aplicativos sociais. *Sites* como Orkut [3] e Facebook [4] são só alguns dos responsáveis por essa popularização. É importante ressaltar que o uso dessas aplicações não tem se restringido apenas a fins meramente de entretenimento. Pequenas empresas, grandes corporações, governos e diversas outras organizações têm despertado para o fato de que *sites* de rede social são um forte aliado na interação com determinado público de interesse.

A grande quantidade de acessos a esses *sites* produz bases de dados com informações preciosas, que muitas vezes podem ser utilizadas em estudos que envolvam os usuários dessas redes e as interações entre eles. Com isso, tomadores de decisão podem realizar ações mais acertadas baseando-se em dados extraídos dessas análises. São de grande importância tais estudos, pois, segundo Kumar et al. [5], o poder de interação entre as pessoas pode ser a razão da falha ou do sucesso de uma organização.

Devido a essa grande quantidade de participantes de uma rede social *on line* e dos dados referentes a ela, é necessário usarmos alguma ferramenta computacional que nos dê suporte quando quisermos fazer algum estudo sobre esse tema. Quando esse estudo envolve uma análise quantitativa e/ou qualitativa dos dados ou necessita da análise visual dos indivíduos da rede e das ligações estabelecidas entre eles, são usadas as chamadas ferramentas de Análise de Rede

Social (ARS)¹ [6]. Diante da também elevada quantidade de ferramentas de ARS disponíveis, pode ser uma tarefa não muito trivial escolher uma adequada às necessidades do usuário, o que pode dificultar, ou até mesmo inviabilizar, o uso de tais softwares [7].

Embora os softwares de ARS já venham sendo utilizados com certa frequência em estudos há um tempo considerável, poucos são os materiais (sobretudo em português) que abordam o tema de uma forma que possa ser acessível a usuários menos experientes. Técnicas relevantes para fins práticos como, por exemplo, a integração dessas ferramentas a web sites de redes sociais são pouco difundidas. Diante disso, este trabalho pretende servir como um guia inicial para aqueles interessados em utilizar ferramentas de Análise de Rede Social para alguma aplicação prática. Para isso, será feito um estudo comparativo entre algumas ferramentas open source de ARS disponíveis e será apresentada uma proposta de implantação de um software desse tipo na área de educação básica (ensino fundamental e médio).

1.1 Objetivos da Pesquisa

Este trabalho tem os seguintes objetivos:

1.1.1 Objetivo Geral:

Apresentar como as ferramentas de análise de rede social podem ser usadas na prática para monitorar aplicativos sociais e apresentar seus benefícios.

1.1.2 Objetivos Específicos:

1. Fazer um estudo bibliográfico sobre o tema de Análise de Redes Sociais e suas aplicações.
2. Realizar um estudo comparativo entre diversas ferramentas open source de Análise de Rede Social existentes.

¹ Neste trabalho, “rede social” e “Análise de Rede Social” fazem menção às redes sociais como um conceito amplo, e não somente a sites de redes sociais.

3. Apresentar uma proposta de implantação de ferramenta de ARS em um aplicativo de rede social com fins educacionais. Proposta esta que possui um protótipo funcional e que contou com a participação do autor deste trabalho na equipe de concepção e desenvolvimento.

1.2 Metodologia e Estratégia de Ação

Para alcançar o primeiro objetivo do trabalho, a metodologia aplicada será o estudo da literatura e de artigos científicos relacionados ao tema, bem como *sites*, *blogs* e tutoriais oriundos de fontes relevantes na área.

Para o estudo comparativo será necessário realizar uma coleta de informações sobre as diversas ferramentas em suas respectivas páginas web e em tutoriais de terceiros. Os critérios usados na comparação serão obtidos de métricas presentes em literaturas reconhecidas ou de entrevistas com profissionais experientes no assunto. A comparação será apresentada em forma de tabela e um texto explicativo desta relacionando os dados com referenciais teóricos.

O terceiro e último objetivo a ser conseguido é mostrar um protótipo de ferramenta de ARS que teve o autor deste trabalho como integrante da equipe de concepção e desenvolvimento. O ambiente em que protótipo foi implantando é uma rede social *on line* que incorpora uma plataforma de jogos, ambos destinados a alunos do ensino fundamental e médio da rede estadual de ensino. Para isso, será necessário que haja uma interação entre a empresa na qual a ideia foi concebida e o autor do trabalho, a fim de que sejam mais bem compreendidos o contexto em que o software de ARS será aplicado e a utilidade prática da ferramenta.

1.3 Resultados e Impactos Esperados

É notável o grande crescimento do uso de aplicativos sociais e o interesse de organizações em utilizar esse meio para obter alguma vantagem. As organizações podem utilizar redes sociais para, por exemplo, realizar recrutamento de pessoas (ex: *LinkedIn* [8]), fazer divulgação de produtos (ex: Anúncios no Orkut [3]), ou simplesmente interagir com seu público alvo (ex: Twitter [9]). Dada a grande quantidade de dados decorrentes do uso dessas redes, faz-se necessário o uso de

ferramentas que permitam tratá-los a fim de serem retiradas informações importantes.

Diante desse contexto, os resultados deste trabalho servirão de material didático com uma linguagem mais acessível para pessoas menos experientes no campo de estudo de ARS. Além disso, após a leitura desta monografia, o leitor poderá ter melhor discernimento entre as potencialidades de quatro ferramentas *open source* de ARS. Pretende-se também que este trabalho sirva ao propósito de ser um incentivador e um guia inicial de como aplicar conceitos de Análise de Rede Social para fins práticos, como, por exemplo, o monitoramento de sites de rede social.

1.4 Organização do texto

Este trabalho possui seu texto estruturado em 5 capítulos. No capítulo 1, é feita uma introdução sobre o problema e é descrita a metodologia de desenvolvimento. O capítulo 2 aborda alguns conceitos básicos de Teoria dos Grafos e de Análise de Rede Social que serão importantes para o entendimento do restante do trabalho. No capítulo 3, é realizado um estudo comparativo entre quatro ferramentas *open source* de ARS e algumas das informações são condensadas em forma de tabela comparativa. O capítulo 4 aborda uma proposta de ferramenta de ARS aplicada à área de ensino escolar básico. O último capítulo (capítulo 5) descreve as conclusões deste trabalho.

Capítulo 2

Análise de Rede Social

Este capítulo aborda, na seção 2.1, alguns conceitos básicos de Redes Sociais e sua evolução. Aqui também são expostos, na seção 2.2, alguns conceitos básicos de Teoria dos Grafos, já que esse campo de estudo é de suma importância para a área de Análise de Redes Sociais, que é o tema da seção 2.3. A seguir, começamos falando dos fundamentos de Redes Sociais.

2.1 Evolução das Redes Sociais

Uma rede social pode ser entendida como um conjunto de atores e os relacionamentos estabelecidos entre eles. Esses atores podem ser uma pessoa, um conjunto discreto de pessoas agregadas em uma unidade social, uma empresa, agências de serviço público de uma cidade, estados-nações do mundo dentre outros [1]. As redes sociais não são dependentes da internet, portanto os atuais e tão populares aplicativos sociais - ou redes sociais *on line*, ou *sites* de rede social - são apenas um novo local em que se pode desenvolver o comportamento social [2].

Na concepção de Boyd et al. [10], *sites* de redes sociais são serviços baseados na *web* que permitem aos indivíduos (1) construir um perfil público ou semi-público dentro de um sistema limitado, (2) articular uma lista de outros usuários com quem eles compartilham uma conexão, e (3) ver e pesquisar sua lista de conexões e aquelas feitas por outros dentro do sistema. A natureza e a nomenclatura dessas conexões podem variar de *site* para *site*.

De acordo com Boyd et al. [10], o primeiro *site* que se encaixa no conceito de rede social *on line* foi o SixDegrees.com, lançado em 1997. O SixDegrees.com autodenominava-se uma ferramenta que possibilitava as pessoas se conectarem entre si e trocarem mensagens. Muitos desses aplicativos sociais tentam focar um público alvo específico, baseado na sua localização geográfica ou na língua falada. No entanto, nem sempre isso é conseguido. O Orkut [3], por exemplo, foi lançado

nos EUA e só possuía interface gráfica em língua inglesa, mas isso não impediu que os brasileiros formassem a maior fatia dos seus usuários [10].

Atualmente, duas das mais utilizadas redes sociais são o Twitter [9] e o Facebook [4]. Segundo os dados de uma pesquisa realizada em 2010 pela Hitwise (empresa que monitora atividades na web) e divulgada em seu *blog*, o Facebook superou o *site* do Google em quantidade de acessos nos Estados Unidos em março desse mesmo ano [11]. Para se ter uma ideia da atual importância dos aplicativos sociais, somente com jogos que funcionam integrados a essas plataformas, a previsão é que as empresas faturem em torno de um bilhão de dólares em 2011, como mostra o *site* E-Commerce News citando uma pesquisa realizada pela empresa eMarketer [12].

Além do caráter econômico citado anteriormente, *sites* de rede social estão desempenhando papéis de cunho social bastante importantes na atualidade. Um exemplo recente é o crucial papel que a utilização do Twitter teve nas revoltas do mundo árabe, como mostra uma matéria no *site* da revista Veja [13]. A seguir, são explicados alguns conceitos de grafos, que são de grande importância para o estudo das redes sociais.

2.2 Conceitos de Teoria dos Grafos

Em Análise de Rede Social, as redes são normalmente modeladas por meio de grafos, o que evidencia a grande importância da Teoria dos Grafos nesse campo de estudo [14]. Em virtude dessa importância, as duas subseções seguintes têm por objetivo esclarecer alguns conceitos-chaves essenciais ao entendimento de ARS. Vale destacar que a área de Análise de Redes Sociais não necessita e nem se utiliza da teoria dos grafos por completo, já que esse é um campo muito amplo e possui uma subárea da matemática dedicada somente a ele. O que os cientistas sociais utilizam se restringe àquilo que eles julgam como importante para descrever e analisar padrões de relações sociais [15]. A seguir são explicados alguns conceitos de grafos.

2.2.1 Grafo Não Direcional

De forma simples, um grafo não direcional – ou somente grafo - consiste em um conjunto de linhas (também chamadas de arestas ou arcos) conectando pares de pontos (também chamados de nós ou vértices). No que se refere a aplicativos sociais *on line*, essa estrutura pode ser utilizada para descrever, por exemplo, redes como Orkut, Facebook e outros. Segundo uma descrição mais formal, essa estrutura é formada por dois conjuntos de informações: um conjunto de nós, $\mathcal{N} = \{n_1, n_2, \dots, n_g\}$, e um conjunto de linhas, $\mathcal{L} = \{l_1, l_2, \dots, l_L\}$, entre pares de nós. Neste caso, há g pontos e L arestas. Em um grafo, cada linha é um par não ordenado de nós distintos, $l_k = (n_i, n_j)$. Uma vez que as arestas são pares não ordenados, a linha entre os nós n_i e n_j é idêntica à que liga n_j e n_i ($l_k = (n_i, n_j) = (n_j, n_i)$). Quando uma aresta liga um nó a ele próprio, (n_i, n_i) , ela é chamada de laço (*loop*) ou ligação reflexiva (*reflexive tie*). Um grafo que não possui laços e nem arestas redundantes é chamado de *grafo simples* [1].

Dado um grafo qualquer, pode-se denotá-lo em termos de seu conjunto \mathcal{N} de nós e do conjunto \mathcal{L} de arestas da seguinte maneira: $\mathcal{G} = (\mathcal{N}, \mathcal{L})$. Dois nós, n_i e n_j , são ditos adjacentes se existe uma linha $l_k = (n_i, n_j)$ no conjunto de linhas \mathcal{L} . Um ponto e uma aresta são incidentes entre si caso o nó faça parte do par não ordenado que define a aresta. Por exemplo, os nós n_1 e n_2 são incidentes com a linha $l_1 = (n_1, n_2)$. Cada linha é incidente com os dois nós no par não ordenado que a define [1].

Um grafo $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ pode, também, ser representado em forma de um diagrama em que pontos denotam nós e uma linha conecta dois pontos se existe no conjunto de arestas \mathcal{L} um par não ordenado formado por esses pontos. Utilizando grafo não direcional, apenas relações simétricas podem ser representadas [14]. A figura 1 mostra o exemplo de uma dessas estruturas. Na imagem, é possível ver que existem seis nós (n_1, n_2, \dots, n_6), que representam pessoas, e seis arestas (l_1, l_2, \dots, l_6), que denotam uma relação de “mora perto de”.

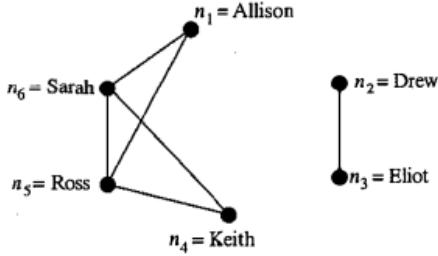


Figura 1 - Grafo de relação “mora perto de” entre seis pessoas

Dado um par ordenado de nós, (n_i, n_j) , diz-se que eles são alcançáveis (*reachable*) entre si se, e somente se, existe um caminho - sequência de uma ou mais arestas, $(n_i, n_a), (n_a, n_b), (n_b, b_c) \dots (n_z, n_j)$, que começa em n_i e termina em n_j , talvez passando por nós intermediários, n_a, n_b, n_z . Quando um caminho começa e termina no mesmo ponto ele é chamado de ciclo (*cycle*) [16]. Na figura 1, o caminho (n_1, n_5) , (n_5, n_6) , (n_6, n_1) é um ciclo. Outro conceito importante é a ideia de conectividade. Quando todo ponto for alcançável por todos os outros, o grafo é dito conectado, o que não é o caso da figura 1.

Associado a cada caminho há uma distância, que corresponde ao número de arestas nesse caminho. Os menores caminhos ligando um dado par de nós são chamados de geodésicas. Nós que estão presentes na única ou em todas as geodésicas ligando determinado par de pontos são ditos estar *entre* (*between*) os extremos (*end points*) [16].

2.2.2 Grafo Direcional (Digrafo)

Um grafo direcionado, ou digrafo, $\mathcal{G}_d = (\mathcal{N}, \mathcal{L})$ é formado por dois conjuntos: um conjunto de nós $\mathcal{N} = \{n_1, n_2, \dots, n_3\}$, e um conjunto de arcos, $\mathcal{L} = \{l_1, l_2, \dots, l_L\}$. Cada arco é um par ordenado de nós distintos, $l_k = \langle n_i, n_j \rangle$. O arco $\langle n_i, n_j \rangle$ é direcionado de n_i (origem ou remetente) para n_j (término, receptor) [1]. Esse tipo estrutura pode ser usado para descrever, por exemplo, redes como a do Tweeter ou qualquer tipo de rede que possua a noção de direção de relacionamento.

Em um digrafo, o número de arcos possíveis é dado por $g(g-1)$, onde g é o número de nós. Assim como em um grafo, um nó é dito incidente a um arco quando ele está presente no par ordenado que define esse arco. Por exemplo, no arco $l_k = \langle n_i, n_j \rangle$, ambos os nós são incidentes a ele. No que se refere a adjacência entre nós, os digrafos têm um conceito mais complicado quando comparados aos grafos

comuns. Aqui, um nó n_i é adjacente a um nó n_j se $\langle n_i, n_j \rangle$ pertence a \mathcal{L} . Do mesmo modo, um nó n_j é adjacente a um nó n_i se $\langle n_j, n_i \rangle$ é elemento de \mathcal{L} [1].

Quando digrafos são representados como diagrama, seus nós são expressos como pontos e seus arcos, como setas. O arco $\langle n_i, n_j \rangle$ é representado por uma seta que vai do ponto n_i ao ponto n_j [1]. A figura 2 exemplifica um grafo dessa natureza. Na imagem, existem seis pessoas (representadas por seis nós) e as relações de amizades entre elas (representadas por arcos). Existirá um arco ligando dois nós sempre que o primeiro escolher o segundo como amigo. Por exemplo, Ross (n_5) escolhe Sarah (n_6) como um amigo, então o arco $\langle n_5, n_6 \rangle$ é incluído no grafo [1].

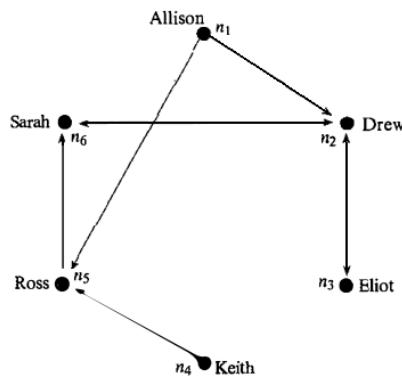


Figura 2 - Exemplo de grafo direcionado (digrafo)

2.3 Análise de Rede Social

O estudo analítico das redes sociais é chamado de Análise de Redes Sociais (ARS) ou, em inglês, *Social Network Analysis* (SNA). Essa é uma abordagem que possui suas raízes na Sociologia, na Psicologia Social e na Antropologia [17]. Esse estudo e suas técnicas também podem ser aplicados para entender os indivíduos e seus relacionamentos estabelecidos em aplicativos sociais, já que estes são apenas uma variação do espaço onde pode se desenvolver o comportamento social [2].

De acordo com a definição dada por Ortiz-Arroyo [14]

A Análise de Rede Social compreende o estudo das relações, laços, padrões de comunicação, desempenho e comportamento dentro dos grupos sociais. Em ARS, uma rede social é comumente modelada por um grafo composto por nós e arestas. Os nós no grafo

representam atores sociais e os links, a relação ou laços entre eles.
(original em Inglês)

Sob uma visão mais ampla, a análise de rede social na verdade é uma subárea de um campo de estudo maior chamado de Análise de Redes (AR), que, por sua vez, também engloba as ditas redes complexas. Na AR, são utilizados métodos de análise numéricas e estatísticas adequados para o estudo de relações entre entidades. Assim, os métodos e conceitos de análise de rede proveem uma fundamentação matemática para a modelagem do relacionamento entre entidades [6].

Além das áreas do conhecimento citadas acima, as técnicas de análise matemática utilizadas pela análise de redes também podem ser empregadas em diversas outras, como na Economia, na Física, na Computação e na Ciência da Informação [6]. As entidades a serem estudadas podem ser dos mais diversos tipos e variam de acordo com o campo de estudo em que se esteja trabalhando. Páginas na web, documentos, atores sociais (pessoas, organizações, países etc.) são só alguns dos exemplos. No caso destes últimos, em que lidamos com atores sociais, os métodos de análise são chamados Análise de Redes Sociais. O que permite esse compartilhamento de técnicas é o fato de todas essas áreas utilizarem uma modelagem e fundamentação matemática comum, que é, na maioria das vezes, através de grafos e matrizes [6].

Há um consenso entre os sociólogos de que poder é uma propriedade fundamental das estruturas sociais. No entanto, quando se discute o seu conceito e as formas de analisar suas causas e consequências nem sempre as opiniões convergem. Na tentativa de contornar esse problema, a Análise de Rede tem sido usada para melhor embasar algumas definições e para propiciar formas de medir essa propriedade mais precisamente [15].

Analistas de redes preferem dizer que trabalham a ideia de centralidade em vez de dizer que estão medindo o poder dos nós na rede. As medidas de centralidade baseadas em grau, proximidade e intermediação servem para descrever quão próximo os indivíduos estão do "centro" da ação na rede, embora as concepções do que significa estar no "centro" também sejam um pouco divergentes. No entanto, de forma geral, há muitas razões para associar posições centrais a

posições de poder [15]. As quatro subseções a seguir abordam algumas das principais medidas de centralidade utilizadas em análise de redes sociais.

2.3.1 Grau (*degree*)

Quando olhamos a topologia em estrela da figura 3, é intuitivo notar que A está em vantagem em relação aos demais nós. Se o ator D decide não fornecer um recurso a A, este possui diversos outros locais de onde obtê-lo; no entanto, se D decide não receber de A, então não poderá receber de nenhum outro. Neste caso, a medida de poder que está sendo utilizada é o grau, que mede a quantidade de arestas incidentes a um nó. Agora, tomando como exemplo a topologia em forma de círculo, percebemos que todos os atores possuem grau dois, o que nos leva a concluir que todos estão na mesma situação de vantagens e desvantagens (considerando-se especificamente o grau do nó como medida de poder). Já quando observamos a formatação em forma de linha, vemos que A e G estão em desvantagem e os demais têm aparentemente igual situação. Na maioria das vezes, os nós que estão mais centralizados na estrutura, no sentido de terem maior grau, tendem a ter posições favorecidas e, consequentemente, maior poder [15].

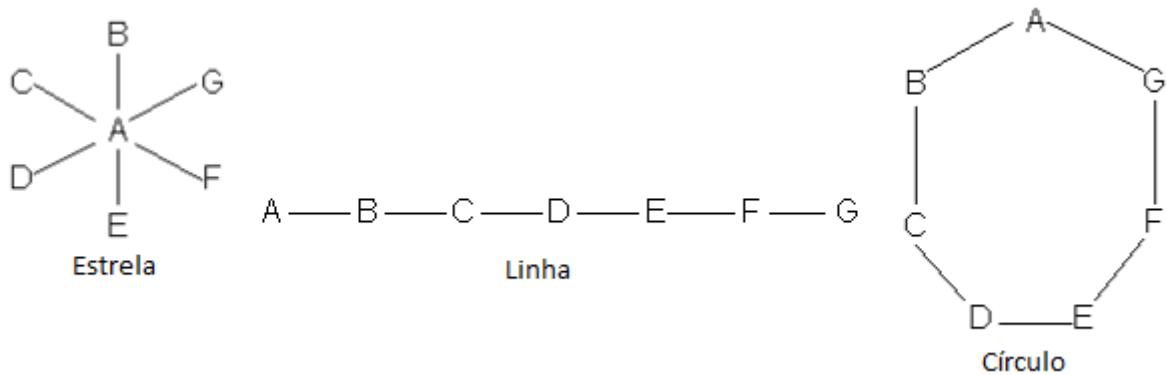


Figura 3 – Exemplos de diferentes topologias de rede.

Matematicamente, o grau de um nó p_k é definido de acordo com a seguinte equação:

$$C_d(p_k) = \sum_{i=1}^n a(p_i, p_k),$$

em que $a(p_i, p_k) = 1$ se, e somente se, p_i e p_k são conectados por uma linha e igual a 0 se não forem. Caso p_k seja diretamente ligado a muitos nós, $C_d(p_k)$ assumirá valores maiores; do contrário, assumirá valores menores [16]. Além disso, em situações em que p_k esteja totalmente isolado dos demais nós, $C_d(p_k)$ será igual a 0.

A equação anterior é, de certo modo, sensível ao tamanho da rede à qual está sendo aplicada. Em determinadas aplicações, no entanto, é mais adequado que a magnitude dos valores assumidos pela equação seja independente do tamanho da rede. Isso acontece, por exemplo, nos casos em que queremos comparar a centralidade de nós de redes distintas [16]. Para que se chegue a essa independência, modifica-se a equação anterior para

$$C'_d(p_k) = \frac{\sum_{i=1}^n a(p_i, p_k)}{n - 1},$$

que mede a proporção de outros pontos ligados diretamente a p_k e onde $n - 1$ denota a máxima quantidade desses pontos. $C_d(p_k)$ e $C'_d(p_k)$ são, portanto, medidas estruturais de centralidade baseadas no grau do ponto p_k e servem como indicador de potencial de atividade de comunicação [16].

2.3.2 Intermediação (*betweenness*):

Considerando o grafo em estrela da figura 3, A possui capacidade de, por exemplo, intermediar contato entre os outros nós ou de isolá-los. Isso se deve ao fato de que o referido nó está entre todos os pares formados pelos demais pontos e de A estar diretamente ligado a todos os nós. Se, por exemplo, A quiser contactar F, ele simplesmente o faz; no entanto, se F quiser contactar B, ele deve fazê-lo através de A. Tratando-se da rede circular, vê-se que cada ator está em exatamente um dos dois caminhos que unem quaisquer outros dois atores. Novamente, neste caso, há uma equidade em relação às vantagens ou às desvantagens do posicionamento dos pontos na rede. Já no caso da rede linear, os pontos nos extremos (A e G) estão em desvantagem em relação aos demais, que possuem poder de intermediação em decorrência de suas localizações entre pares de nós. Em resumo, quando um nó está entre os demais, ele se encontra em posição de vantagem estrutural [15].

O potencial de um ponto p_k controlar a comunicação entre p_i e p_j pode ser definido em termos da probabilidade de p_k estar em uma geodésica aleatoriamente

escolhida que liga p_i e p_j . O valor dessa probabilidade é calculado com a seguinte equação:

$$b_{ij}(p_k) = \frac{g_{ij}(p_k)}{g_{ij}},$$

onde $g_{ij}(p_k)$ é a quantidade de geodésicas que ligam p_i e p_j passando por p_k e g_{ij} é o total de geodésicas ligando p_i e p_j .

Estendendo a equação anterior, é possível calcular o valor global de intermediação de p_k . Para isso, somam-se todos os valores parciais de intermediação de p_k em relação a todos os pares não ordenados de pontos. Essa extensão é denotada por

$$C_B(p_k) = \sum_{i < j}^n \sum_{j < i}^n \frac{g_{ij}(p_k)}{g_{ij}},$$

onde $k \neq i \neq j$ e n é o total de nós da rede [16]. Na equação, o valor de $C_B(p_k)$ é acrescido de 1 unidade sempre que p_k estiver na única geodésica ligando um par de pontos. Caso existam geodésicas alternativas, esse valor cresce proporcionalmente em relação à frequencia de ocorrência de p_k .

De forma similar à seção anterior, para a medida de intermediação também se faz necessário definir uma equação que seja independente do tamanho do grafo. Aqui, essa normalização será realizada dividindo a equação anterior pelo maior valor de intermediação possível para o nó de uma rede. Segundo Freeman [16], esse valor máximo é atingido para um ponto central em uma topologia em estrela e é definido pela seguinte equação:

$$\frac{n^2 - 3n + 2}{2}.$$

Desse modo, a equação final que mede relativamente o poder de intermediação de um nó qualquer p_k tem a seguinte formatação:

$$C'_B(p_k) = \frac{2C_B(p_k)}{n^2 - 3n + 2}.$$

Em resumo, $C_B(p_k)$ e $C'_B(p_k)$ são medidas de centralidade cujo objetivo é medir o potencial que determinado ponto tem de controlar a comunicação de outros.

2.3.3 Proximidade (*closeness*)

Outro aspecto pelo qual A pode ser considerado mais importante que os demais na figura em forma de estrela é o da medida de proximidade. Em outras palavras, esse ator é o que está mais próximo de todos os outros. Neste caso, o poder é exercido em forma de negociação direta e intercâmbio de informação. Pontos que conseguem atingir outros através de caminhos de comprimento mais curto ou que são mais alcançáveis pelos demais também por caminho de comprimento mais curto têm posição privilegiada. Ainda utilizando o exemplo do grafo em forma de estrela, percebe-se que A está a uma distância geodésica de um em relação aos demais nós; ao passo que estes estão localizados a uma geodésica de dois em comparação aos outros (exceto de A). Agora, considerando a rede circular em termo de proximidade dos seus atores, podemos dizer que todos os nós estão a distâncias distintas de cada outro nó na rede. Todavia, todos eles possuem distribuições de proximidade idênticas, o que nos faz concluir que são idênticos em termos de posição estrutural. Na figura do grafo linear, o ator do meio (D) é o que está mais perto de todos os outros, enquanto que os nós nos extremos estão em certa desvantagem [15].

A medida de proximidade pode ser entendida como o poder que um nó tem de não precisar de outros como intermediadores. Ou seja, quanto mais centralizado o nó estiver no grafo, mais ele consegue contatar os demais de modo menos custoso e em menor tempo. O valor dessa métrica relativo a um ponto p_k pode ser calculado somando-se as distâncias geodésicas dele a todos os outros pontos da rede. Como esse valor cresce quando os nós se encontram mais distantes, na verdade, essa é uma medida de descentralidade ou de centralidade inversa [16].

Seja $d(p_i, p_k)$ o número de arestas da geodésica que liga p_i e p_k , então a medida de descentralidade de p_k será dada por:

$$C_C(p_k)^{-1} = \sum_{i=1}^n d(p_i, p_k).$$

Vale salientar que essa é uma métrica válida somente para grafos conectados, uma vez que se ele fosse não conectado, todos os nós estariam a uma distância infinita de pelo menos um ponto, o que tornaria $\sum_{i=1}^n d(p_i, p_k) = \infty$ para todo p_k .

Assim como nas duas seções anteriores, para o cálculo da proximidade também é importante uma forma que seja independente do tamanho da rede. Para tal, a centralidade relativa de p_k é definida como:

$$C'_C(p_k) = \left[\frac{\sum_{i=1}^n d(p_i, p_k)}{n-1} \right]^{-1} = \frac{n-1}{\sum_{i=1}^n d(p_i, p_k)},$$

em que $n-1$ denota a mínima soma de distâncias para um ponto que é adjacente a todos os outros [16]. $C'_C(p_k)$ assume valor 1 caso p_k seja adjacente a todos os outros nós e tende a diminuir nos casos em que p_k está distante dos demais.

$C_C(p_k)^{-1}$ e $C'_C(p_k)$ são, portanto, indicadores de centralidade baseados na proximidade do nó. Ambos podem ser utilizados nos casos em que se queira determinar o grau de independência ou de eficiência de comunicação do nó[16].

2.3.4 Outras medidas de centralidade

Além das medidas principais de centralidade citadas anteriormente, outras técnicas também são utilizadas em Análise de Redes Sociais. A seguir, são dadas informações breves sobre algumas delas.

- HITS (*Hypertext Induced Topic Search*): o funcionamento básico desse algoritmo consiste em analisar a estrutura de *links* direcionais de uma rede para atribuir às páginas *web* pesos que serão utilizados para classificar a importância delas. O algoritmo baseia-se na premissa de que toda página serve para os propósitos de fornecer informações sobre um tópico e de referenciar outras páginas. Isso leva a dois conceitos de categorização: autoridade e *hub*. Para ser considerada uma autoridade, uma página deve fornecer boas informações sobre determinado tópico; para ser considerada um *hub*, deve referenciar boas autoridades em certo assunto. Desse modo, HITS pode ser definido como um algoritmo que serve para quantificar o valor de cada página como uma autoridade e como um *hub* [18].
- PageRank: essa técnica foi apresentada em 1998 pelos fundadores do Google e possui similaridades com o HITS (anterior ao PageRank). Esse algoritmo é atualmente utilizado na indexação de resultados de buscas realizadas com o site Google.com. Neste caso, cada página

web é representada por um nó e cada *hiperlink*, por uma aresta. Neste tipo de análise (que só é aplicável a grafos direcionais), trabalha-se a ideia de prestígio através do *in-degree* (quantidade de arestas que incidem em um nó) e *outdegree* (quantidade de arestas que saem de um nó) dos pontos do grafo. Nós que são referenciados por outros nós importantes são chamados de autoridade (*authority*) e nós que referenciam outros relevantes são chamados de *hubs*. No que se refere à Análise de Rede Social, pode-se citar, por exemplo, o trabalho de Noordhuis et al. [19], que realizaram um estudo aplicando o PageRank para analisar características da rede social Tweeter.

2.4 Ferramentas e Aplicações de ARS

Existem diversas ferramentas que foram desenvolvidas para dar suporte ao campo de estudo de Análise de Rede Social. Elas permitem analisar teoricamente a rede como também representá-la graficamente. Isso se dá através do cálculo de diversos indicadores da estrutura da rede, dos relacionamentos entre os atores e da posição desses atores.

Além disso, alguns desses *softwares* também permitem fazer um comparativo entre diversas redes. Dentre essas ferramentas, por serem bastante presentes na literatura e em artigos, podemos destacar: UCINet [20] e Pajek [21] (ambas não *open source*). Uma lista bastante completa desse tipo de *software* pode ser encontrada no site da INSNA [22] (*International Network for Social Network Analysis*), uma associação de pesquisadores de ARS. O capítulo 3 deste trabalho apresenta quatro ferramentas *open source* de ARS e realiza um comparativo entre elas. Lá é possível ter uma visão mais aprofundada das potencialidades desse tipo de *software*.

2.4.1 Exemplos de Análise de Rede Social em Aplicativos Sociais

Com o grande crescimento do uso de *sites* de rede social, muitos estudos estão usando os conceitos de ARS para entender as propriedades desses aplicativos. Mais especificamente, essas pesquisas tentam compreender as práticas,

implicações, cultura de uso e o engajamento dos usuários com esses softwares [23]. Alguns desses estudos são citados a seguir.

- Adamic et al. [24] estudaram uma rede social *on line* na Universidade de Stanford. Eles descobriram que a rede exibia um padrão de mundo pequeno² e de agrupamento local.
- Kumar et al. [5] estudaram dois aplicativos (Flickr e Yahoo) e descobriram que ambos possuem grandes grupos altamente conectados.
- Girvan and Newman [25] concluíram que usuários desses aplicativos tendem a formar grupos muito unidos.
- Backstrom et al. [26] estudou a participação dos usuários em grupos do LiveJournal e apresentou um modelo de crescimento desses grupos.

Todos esses pesquisadores pontuam que esse tipo de estudo de aplicativos sociais é importante para entender aspectos culturais, sociais e econômicos. O entendimento desses três aspectos é crucial para atividades de planejamento e de tomada de decisão [23].

2.5 Considerações finais

Este capítulo apresentou uma evolução das redes sociais, bem como alguns conceitos básicos de grafos e do campo de estudo de Análise de Rede Social. Além disso, foram mostrados alguns estudos que aplicam conceitos de ARS para entender o comportamento das redes formadas em aplicativos sociais.

² A teoria de mundo pequeno diz que as pessoas estão no máximo a seis graus de separação. Ou seja, pode-se alcançar qualquer indivíduo no mundo através de uma rede de contatos de no máximo tamanho seis.

Capítulo 3

Comparativo de Ferramentas

O objetivo deste capítulo é mostrar as principais características e funcionalidades de quatro ferramentas *open source* de Análise de Rede Social, bem como fazer um comparativo entre elas. As ferramentas aqui abordadas serão o Gephi [27], Cytoscape [28], JUNG [29] e GUESS [30]. Os seguintes critérios foram utilizados para escolha dessas ferramentas:

- Existência de artigos científicos que as mencionem ou que as utilizaram no trabalho
- Apresentarem métricas básicas de Análise de Rede Social
- Existir documentação básica em *sites*, tutoriais ou artigos
- Possuírem funcionalidade de visualização de grafo de rede
- Serem *Open Source*.

Este trabalho foca especificamente em ferramentas *open source* pelo fato de elas serem grátis, o que não demandaria gastos financeiros para aqueles interessados em iniciar suas atividades nesse campo de estudo. Outro fato importante é que, com esse tipo de filosofia *open source*, pessoas interessadas podem modificar essas ferramentas e incorporá-las em algum aplicativo de seu interesse.

Este capítulo tem a seguinte estruturação: a seção 3.1 explica o que são ferramentas de ARS; as seções 3.2, 3.3, 3.4 e 3.5 falam, respectivamente, sobre as ferramentas Gephi, Cytoscape, JUNG e GUESS; a última seção faz um comparativo entre essas ferramentas estudadas.

3.1 Ferramentas de Análise de Rede Social

As redes sociais podem chegar a possuir milhões de integrantes (indivíduos, organizações, instituições etc.), portanto muitas vezes é necessário usarmos algum recurso computacional que nos dê suporte quando quisermos fazer algum estudo sobre esse tema. Quando esse estudo envolve uma análise da estrutura da rede e

dos relacionamentos entre atores, são usadas as chamadas Ferramentas de Análise de Rede Social.

Dentre as funcionalidades oferecidas por grande parte dessas ferramentas, podem ser destacadas:

- Visualização do grafo da rede (integrantes representados por nós e os relacionamentos entre eles, por arestas)
- Cálculo de diversos indicadores relacionados aos nós em específico ou ao grafo como um todo
- Detecção de comunidades (*clustering*)

Nesses programas de computador, os dados são normalmente guardados, descritos e manipulados por meio de matrizes. Segundo Barbosa *et al.* [7], o uso de matrizes para descrever relacionamentos em grupos sociais teve um grande crescimento a partir do ano de 1940. As operações matriciais passaram a ser amplamente utilizadas para definições e cálculos em análise de redes sociais, sendo a representação primária para a maioria dos atuais softwares utilizados nessas atividades.

No que se refere à parte de visualização, muitas ferramentas implementam diversos algoritmos de layout de grafos, que podem ser aplicados de forma simultânea em uma mesma rede. Os algoritmos do tipo *force-based* têm inspiração em fenômenos físicos e seu funcionamento consiste em repelir nós isolados para regiões vazias e agrupar nós mais bem conectados. Dentre algumas dessas técnicas *force-based*, podemos citar a de Fruchterman Reingold [31] e a de Kamada-Kawai [32], ambas bastante utilizadas em estudos de redes. Outra funcionalidade oferecida por grande parte dessas ferramentas é a mudança de cor e/ou de forma dos nós de acordo com as características específicas de cada grupo, o que gera um cruzamento de vários dados delineantes da rede. A Figura 4 foi obtida a partir do software Gephi (abordado na próxima seção) e mostra um grafo que representa terremotos acontecidos em algumas localidades e as relações entre eles.

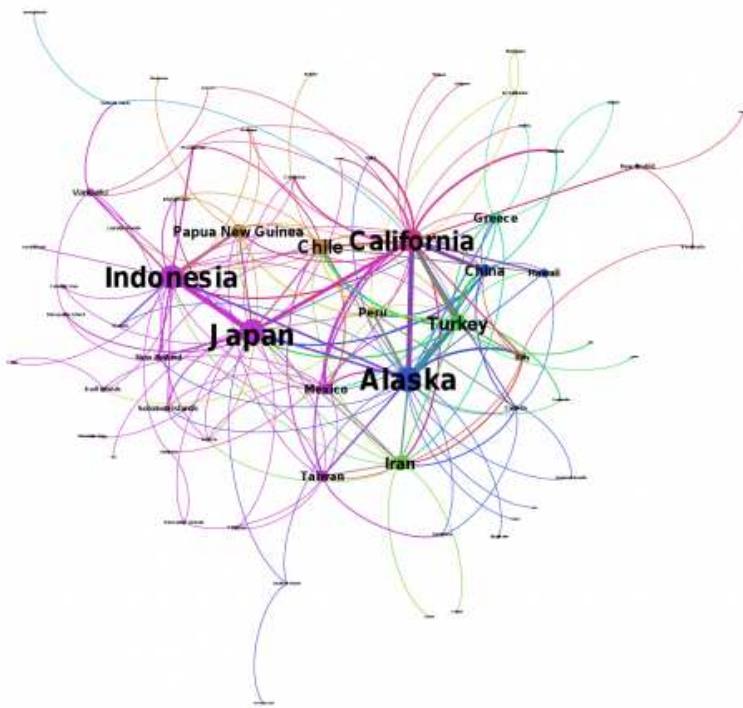


Figura 4 - Exemplo de análise de rede com o software Gephi.

As quatro seções a seguir abordam detalhadamente cada uma das quatro ferramentas que são o foco deste estudo comparativo.

3.2 Gephi

O Gephi é um *software* gratuito (*free*) e de código aberto (*open source*) para exploração e manipulação de redes. Os grafos que simulam essas estruturas podem possuir milhares de nós e arestas e podem ser direcionais, não-direcionais ou mistos. É desenvolvido em linguagem Java e funciona em ambiente Windows, Linux e Mac OS X. Essa ferramenta tem sua interface gráfica baseada em NetBeans e possui uma arquitetura multitarefa que permite trabalhar com diversas redes ao mesmo tempo. Um detalhe muito importante é que essa ferramenta possui uma comunidade de usuários bastante ativa que troca experiências via fórum e disponibiliza diversos tutoriais em *blog* e *Wiki*. A arquitetura modular do Gephi permite que desenvolvedores implementem novas funcionalidades como métricas, algoritmos de leiaute de grafo ou fontes de dados e as disponibilizem na forma de *plug-ins* para outros usuários [33].

No que se refere a métricas de análise de redes, o Gephi possui as tradicionais medidas de centralidades: grau (*degree*), intermediação (*betweenness*) e proximidade (*closeness*). Além dessas de centralidade, também estão disponíveis diversas outras como: densidade (*density*), comprimento de caminho (*path length*), diâmetro (*diameter*), HITS, modularidade (*modularity*) e coeficiente de agrupamento (*clustering coefficient*). A Figura 5 mostra a interface do programa e uma das formas como essas métricas podem ser visualizadas, que é a de relatório HTML [33].

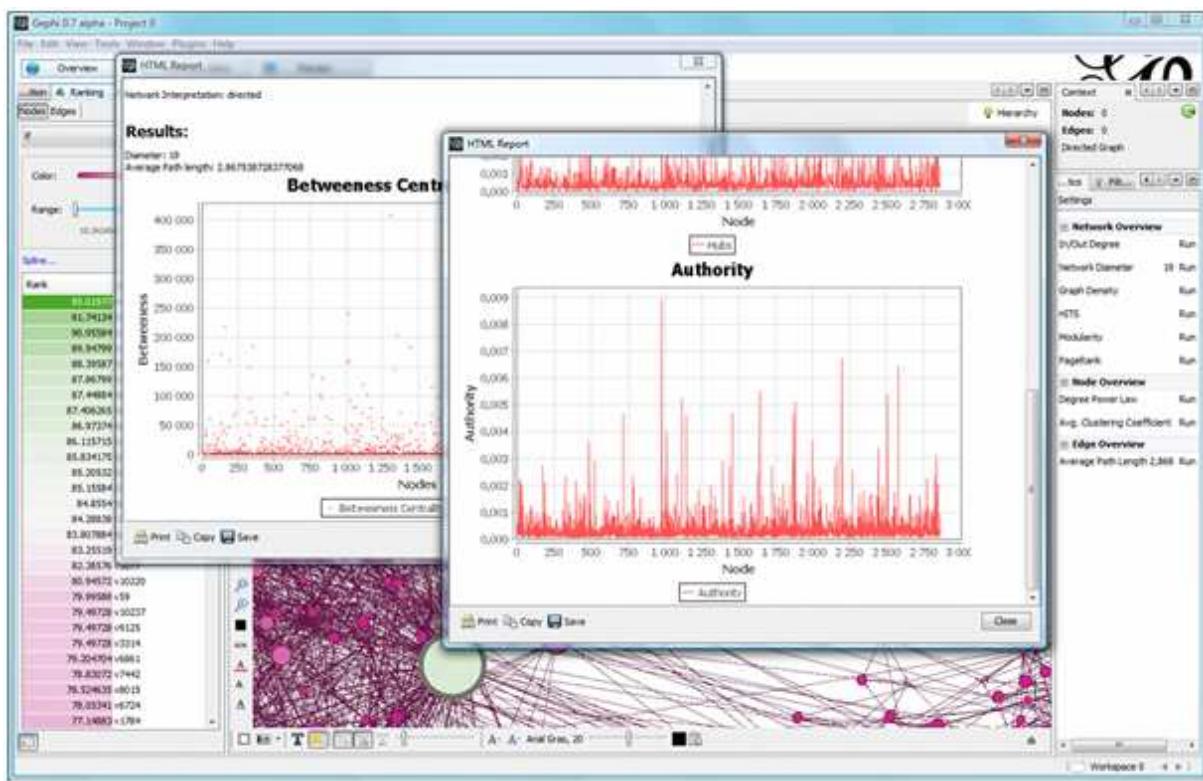


Figura 5 - Interface do Gephi e dados analíticos da rede

Essa ferramenta é bem versátil e pode trabalhar com diversos tipos de arquivos de entrada e de saída. Os arquivos de entrada podem ser dos tipos: GEXF, GDF, GML, GraphML, Pajek NET, GraphViz DOT, CSV, UCINET DL, Tulip TPL, Netdraw VNA e Planilha Excel. Uma funcionalidade muito interessante é que possível buscar os dados de entrada diretamente dos bancos de dados MySQL e PostgreSQL. Em relação aos arquivos de saída, esse software pode exportar a rede para três tipos de arquivos que são bastante difundidos: PDF, JPG e SVG [33].

A parte de visualização e manipulação da rede possui diversas funcionalidades. Dentre elas, vale destacar:

- Possibilidade de selecionar um ou vários nós para realçar os que são adjacentes a ele(s) e realçar as arestas que os conectam. A Figura 6 é um exemplo que mostra os vizinhos de um nó selecionado.

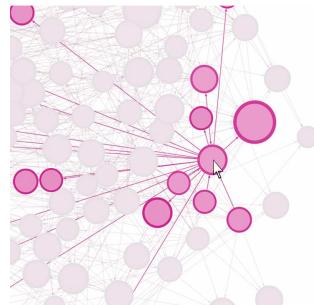


Figura 6 - Realce dos adjacentes de um nó.

- Pode-se realçar o caminho mais curto que liga dois nós. A Figura 7 mostra o exemplo de realce do caminho mais curto ligando dois aeroportos nos EUA através de linhas aéreas.

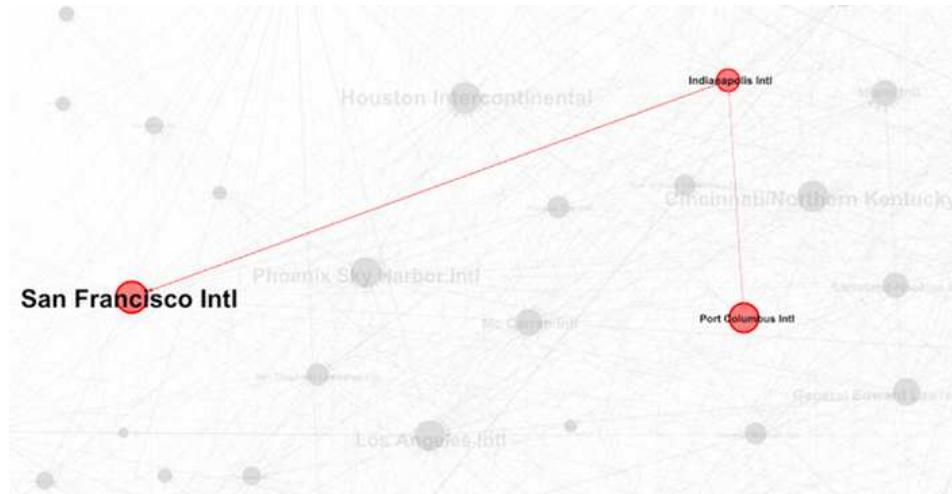


Figura 7 - Descoberta de caminho mais curto pelo Gephi

- Diversos algoritmos de *layout* estão disponíveis e outros podem ser facilmente instalados como *plug-in* através da própria ferramenta. A Figura 8 mostra uma rede modelada com o algoritmo de layout de grafo *force-based*.

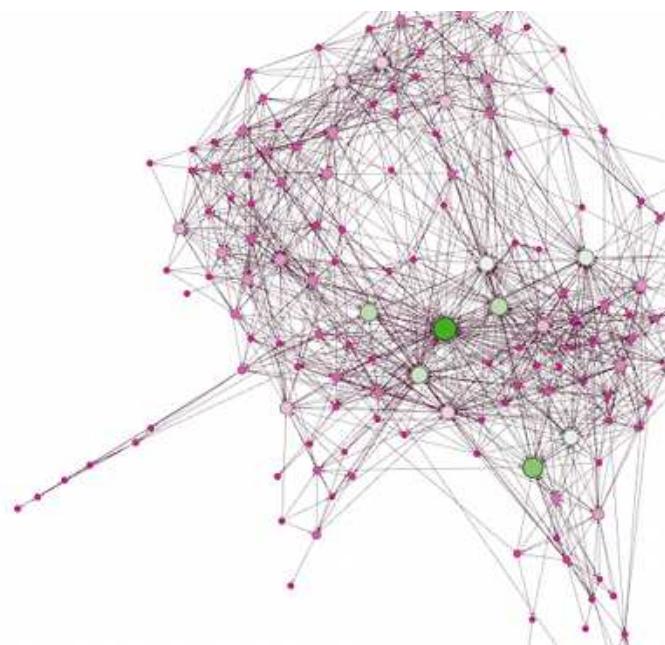


Figura 8 - Rede modelada com o algoritmo *force-based*.

- É possível visualizar a rede com os nós variando de cor e/ou tamanho de acordo com determinada métrica escolhida ou com atributos dos nós. Ou seja, podemos, por exemplo, fazer com que os nós mais conectados (maior grau) sejam maiores que os demais, a exemplo dos nós em verde da Figura 8.
- É possível agrupar os nós de acordo com seus atributos ou utilizando alguns dos algoritmos de clusterização disponíveis. Além disso, podemos agrupar subgrafos em meta-nós. A Figura 9 mostra o agrupamento de nós e o gráfico em pizza descrevendo os grupos, ambos elaborados com a ferramenta.

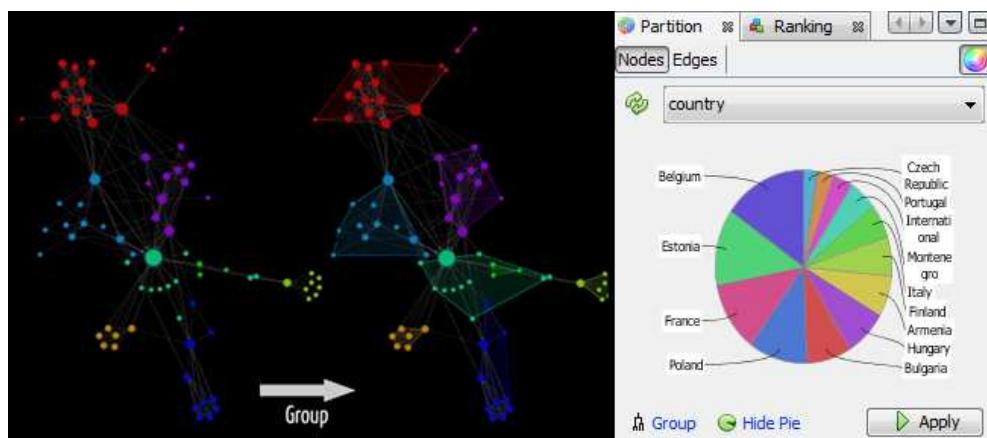


Figura 9 - Agrupamento de nós pelo Gephi

- Possui filtros que possibilitam selecionar nós e/ou arestas de acordo com a estrutura da rede ou dos dados. Para realizar essa filtragem não é necessário escrever nenhum *script*, basta apenas ir selecionando os filtros desejados. Pode-se filtrar, por exemplo, pelo grau do nó, peso da aresta e atributos dos elementos da rede. A Figura 10 mostra alguns dos elementos visuais de filtragem disponíveis.

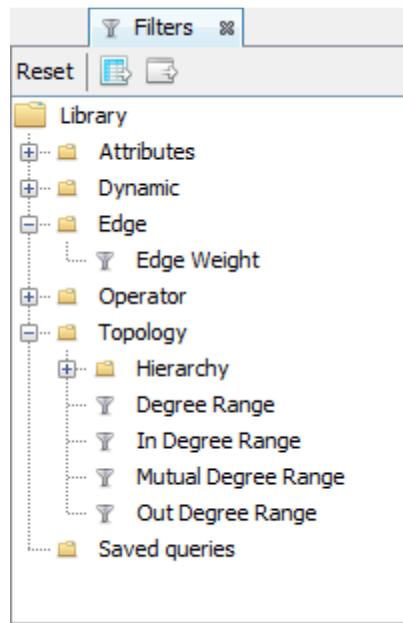


Figura 10 - Filtros de nós e arestas

3.3 Cytoscape

O Cytoscape é plataforma de software *open source* que foi concebida originalmente para permitir a visualização de interações moleculares no campo de estudo da bioinformática. No entanto, ela agora se tornou uma ferramenta geral para análise e visualização de redes complexas. Essa plataforma pode ser empregada em diversos domínios de problemas, incluindo, por exemplo, o campo original da bioinformática, análise de rede social e web semântica. É possível baixar *plug-ins* no site da ferramenta ou até mesmo desenvolvê-los utilizando uma API - Interface de Programação de Aplicativos, do Inglês, *Application Programming Interface* - Java. Utilizando esses *plug-ins* pode-se estender as funcionalidades do programa com novos algoritmos de leiaute, suporte a novos formatos de arquivo, linguagens de *script* e conexão com banco de dados [28].

Quando se trabalha na área de análise de redes, com o Cytoscape é possível:

- Visualizar e analisar grandes redes de relacionamentos interpessoais.
- Montar redes sociais a partir de tabelas e formulários.
- Coletar interações sociais da *web* através de *web crawlers* [34] dentro da própria ferramenta ou de API's como a *Social Graph API* [35].
- Calcular estatísticas da rede utilizando *plug-ins* como NetworkAnalyzer ou CentiScaPe.
- Trabalhar em conjunto com o *software* estatístico R [36] para análises mais avançadas.
- Fazer descobertas de caminho mais curto entre dois nós da rede.
- Descoberta de grupos através dos algoritmos hierárquico, k-Medóides, AutoSOME, k-means, MCODE e outros. [37]
- Importar arquivos de outros programas populares de Análise de Rede, incluindo o igraph, o Pajek e o Graphviz
- Valores de atributos podem ser usados para controlar aspectos visuais (forma, cor e tamanho) dos nós e arestas como também para realizar buscas dentro da rede, operações de filtragem e outras análises.
- Os nós podem ser customizados com imagens carregadas a partir de um *browser* padrão de arquivo ou através de uma URL (*Uniform Resource Locator*). A Figura 11 mostra exemplos de customizações de nós realizadas com essa ferramenta. O item (A) exibe nós em forma de gráfico circular (pizza), o item (B), nós com gráficos gerados pela API do Google Chart e os itens (C) e (D) exibem um rede de interação proteína-proteína em que cada nó contém uma imagem 3D da estrutura da proteína representada pelo nó.

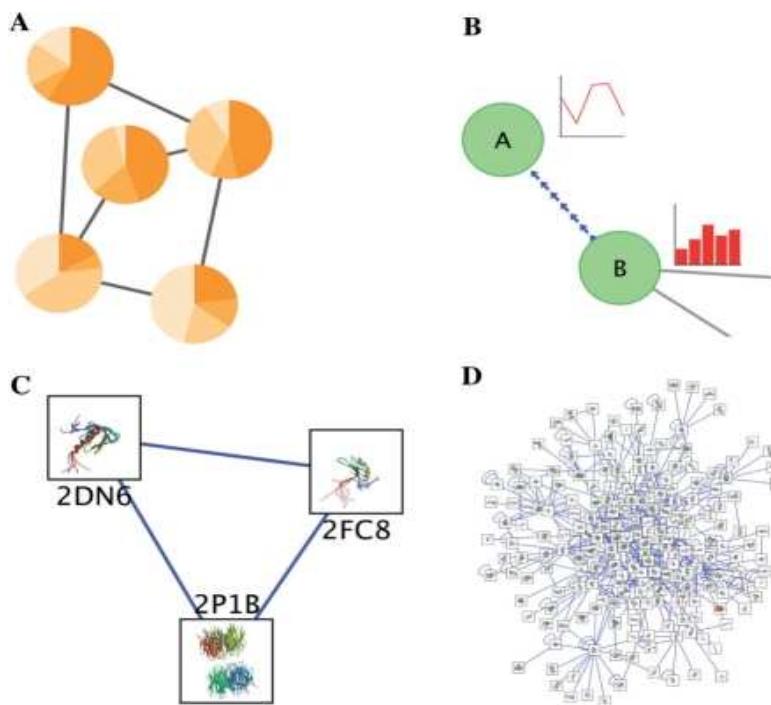


Figura 11 - Exemplos de customização de nós no Cytoscape

3.4 JUNG

JUNG significa Java Universal Network/Graph Framework e, como o nome sugere, é uma biblioteca de *software* em Java que permite modelar, analisar e visualizar dados que podem ser representados como um grafo ou rede. Por ser escrito em Java, as aplicações baseadas no *framework* do JUNG podem fazer uso das API's dessa linguagem como também bibliotecas de terceiros. Os tipos de grafo com que se pode trabalhar são: grafos direcionais e não direcionais, grafos multimodais, grafos com arestas paralelas e hipergrafos. A Figura 12 mostra o exemplo de um grafo misto (arestas direcionais e não direcionais) e algumas opções de algoritmos de leiaute de grafo em um aplicativo feito com o JUNG [29].

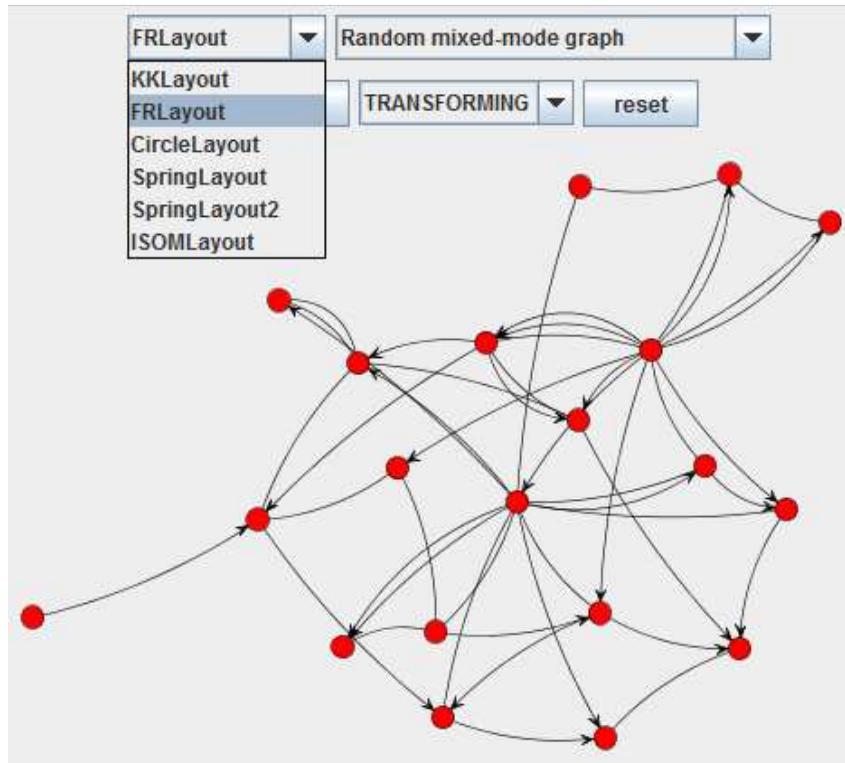


Figura 12 - Exploração de grafo misto e algoritmos de leiaute com o JUNG.

A versão atual do JUNG possui implementações de algoritmos para os campos de teoria dos grafos, mineração de dados e análise de redes sociais. Para este último, existem rotinas para agrupamentos, decomposição, geração de grafos aleatórios, análise estatística, cálculo de distâncias da rede e medidas de importância (centralidade, PageRank e HITS). A Figura 13 mostra o exemplo de um applet Java que usa este *framework* e cuja função é calcular o caminho mais curto entre dois nós.

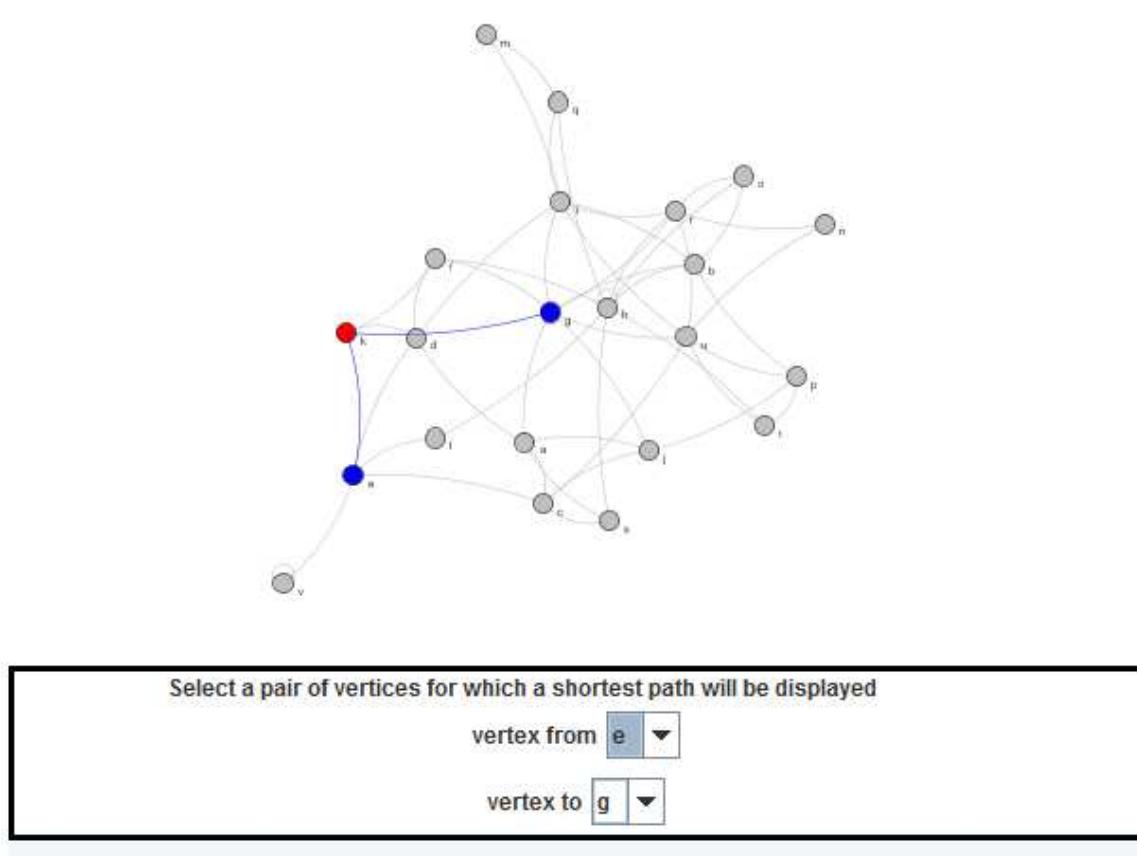


Figura 13 - Aplicativo que utiliza o JUNG e mostra o menor caminho entre nós.

No que se refere à parte de visualização, o JUNG oferece um *framework* cuja proposta é facilitar a construção de ferramentas para a exploração interativa de dados da rede. Para se configurar a visualização do grafo é possível utilizar um dos algoritmos de layout fornecidos ou, se preferido, implementar algum de seu interesse.

3.5 GUESS

A proposta do GUESS é ser uma ferramenta de análise e visualização de grafos e redes. Esse sistema é desenvolvido em Java e possui incorporada uma linguagem de domínio específico chamada de Gython (uma extensão de Python) que possibilita a manipulação da estrutura dos grafos estudados. Na interface gráfica da ferramenta, existe um interpretador onde é possível digitar os comandos escritos em Gython e ver as alterações realizadas no grafo em tempo real. No que se refere a exportação de arquivos, com o GUESS é possível gerar diversos formatos, como

mostra a Figura 14. Dentre os formatos de exportação, merecem destaque o JPEG, JPG, GIF, SFW (flash) e PDF. [30]

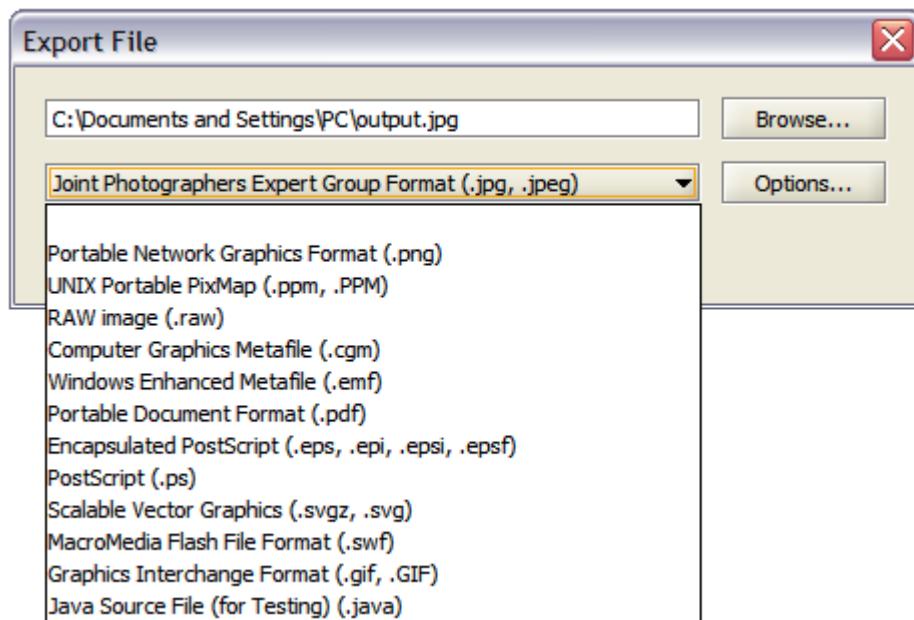


Figura 14 - Tela de exportação de arquivos do GUESS

Pelo fato de ser *open source*, é possível também construir novas aplicações e applets modificando ou reutilizando o código original, o que dá mais flexibilidade para adequar o aplicativo às necessidades de quem o utilizará. Quanto aos formatos de arquivos que podem ser lidos pela ferramenta, os atualmente suportados são do tipo graphml e gml, ambos formatos de arquivo para descrição de grafos.

A interface gráfica do programa é baseada no Piccolo (API gráfica de alto nível para Java), mas também se pode modificar o código para trabalhar com interfaces de outros sistemas, como a do Prefuse e a do TouchGraph. O GUESS também pode ser integrado para trabalhar com o JUNG e com o R, o que fornece mais opções para novos algoritmos de leiaute, comandos de análise de grafo e suporte a análises estatísticas mais elaboradas. Além das funcionalidades citadas anteriormente, essa ferramenta também fornece suporte para se trabalhar com dados temporais que permitem mostrar os estados do grafo em forma de animação. A Figura 15 mostra a tela principal deste *software* em que é exibida uma rede exemplo qualquer. Na imagem, a seta vermelha indica um painel onde são mostradas informações de um nó selecionado com o clique do mouse; a seta verde mostra a área em que se pode digitar comandos para manipular a visualização do

grafo, que, nesse exemplo, teve o tamanho dos nós e o algoritmo de leiaute modificados através de alguns comandos mostrados na figura. Apesar de o site da ferramenta ter sido atualizado pela última vez no ano de 2007 (segundo consta na página inicial), existe um grupo de discussão da ferramenta que ainda está ativo e onde é possível tirar dúvidas e dividir experiências com outros usuários.

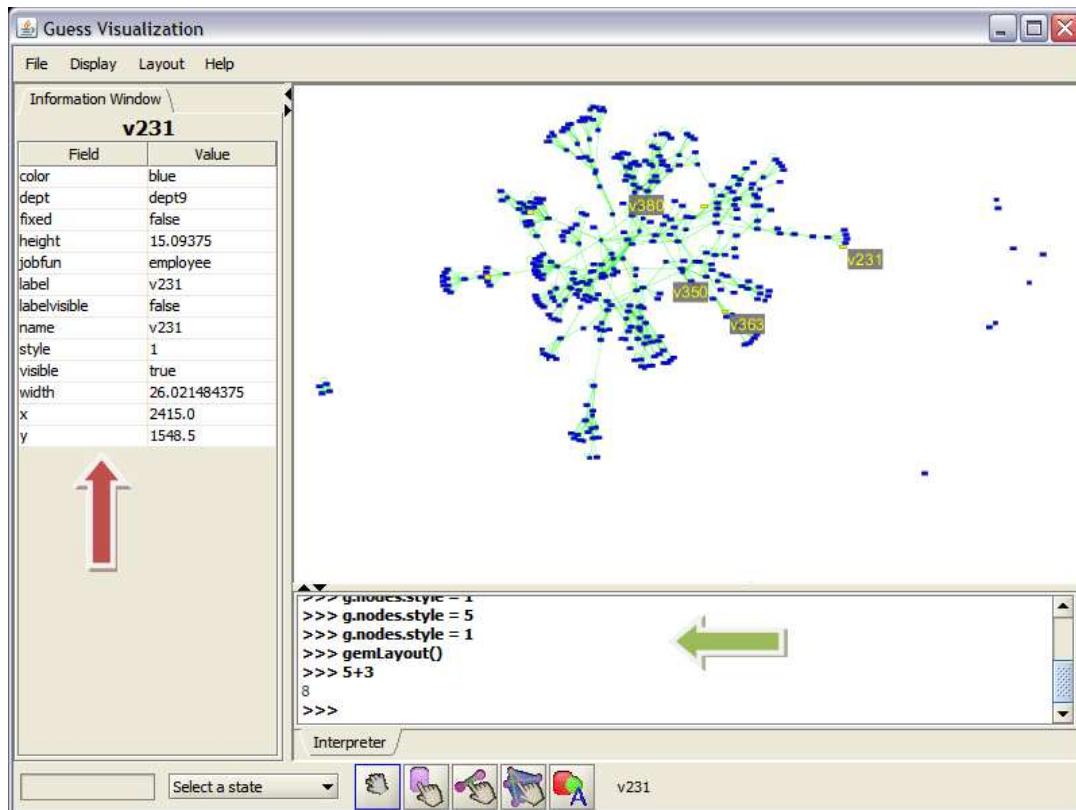


Figura 15 - Tela principal de uma aplicação baseada no GUESS.

3.6 Comparativo das ferramentas

Esse estudo comparativo das ferramentas abordadas nas seções anteriores tem por objetivo condensar algumas informações relevantes a respeito delas e tornar mais fácil a escolha de uma em detrimento de outra. Para isso, informações importantes foram retidas diretamente das fontes oficiais de cada uma. Os critérios de comparação foram baseados na literatura de ARS e em aspectos práticos relacionados ao uso dessas ferramentas. As seguir são detalhados os critérios e resultados da comparação.

3.6.1 Critérios de comparação

Nesse estudo comparativo, os critérios que serão levados em consideração serão:

- Tipo da ferramenta: classifica se a ferramenta é uma aplicação de *software* completa ou se é um *framework* a partir do qual se deve implementar o *software*.
- Plataforma: especifica em quais sistemas operacionais a ferramenta pode ser utilizada.
- Quantidade suportada de nós: especifica quantos nós, aproximadamente, a ferramenta consegue tratar.
- Site com tutoriais e fórum: verifica se a ferramenta possui um *site* oficial que conta com tutoriais de uso e fórum de discussões.
- Arquivo de entrada: especifica quais os formatos de arquivos de entrada que a ferramenta suporta.
- Leiaute de Grafo: especifica se a ferramenta suporta algum algoritmo de leiaute do tipo *force-based* e se é possível instalar *plugins* de novos algoritmos.
- Métricas: especifica quais métricas comumente utilizadas em ARS são disponíveis na ferramenta.
- Conexão com BD's: especifica se a ferramenta possui capacidade de se conectar aos bancos de dados MySQL e PostgreSQL.

Existem três formas básicas de representar a estrutura de uma rede. São elas: matriz de adjacências; lista de adjacências (para grafos direcionais), que mostra o nó seguido por todos os demais a que ele direciona um arco; pares de vértices. Os formatos comparados serão:

- a) GEXF: tipo de arquivo proprietário do *software* Gephi [33]. É baseado em xml e possibilita armazenar informações de posicionamento, de forma e de cor dos elementos do grafo.
- b) NET: tipo de arquivo proprietário do programa Pajek [21]. Apesar de não muito bem documentado é muito utilizado por softwares de ARS.
- c) DL: tipo de arquivo padrão do *software* UCINet [20]. A representação das arestas nesse tipo de arquivo pode ser por matriz, por lista de arcos ou por lista de arestas.

d) GraphML: tipo de arquivo de descrição de grafos também baseado em xml. Foi desenvolvido por uma comunidade dedicada à representação visual de grafos.

3.6.2 Quadro Comparativo das Ferramentas

A tabela a seguir condensa o resultado desse comparativo. Os critérios citados na subseção anterior estão posicionados na coluna mais à esquerda, enquanto que as ferramentas comparadas estão na primeira linha da tabela.

Tabela 1. Tabela comparativa entre ferramentas

Ferramenta	Gephi	Cytoscape	JUNG	GUESS
Tipo	Software	Software	Framework	Software
Plataforma	Windows, Mac OS e Linux	Windows, Mac OS e Linux	Java (multiplataforma)	Windows, Mac OS e Linux
Quantidade Suportada de nós	< 1.000.000	150.000	> 150.000	700
Site com tutoriais e fórum	Sim	Sim	Sim	Sim
Arquivo de entrada				
GEXF (Gephi)	Sim	Não	Não	Não
NET (Pajek)	Sim	Sim	Sim	Sim
DL (UCINet)	Sim	Não	Não	Não
GraphML	Sim	Com plugin	Sim	Sim
Leiaute de Grafo				
Force-Based	Sim	Sim	Sim	Sim
Aceita Plugins	Sim	Sim	Sim	Não
Métricas				
Grau	Sim	Sim	Sim	Sim
Intermediação	Sim	Não	Sim	Sim
Proximidade	Sim	Não	Sim	Não
HITS	Sim	Não	Sim	Sim
Page Rank	Sim	Não	Sim	Sim
Conexão com BD's				
MySQL	Sim	Não	Não	Não
PostgreSQL	Sim	Não	Não	Não

3.6.3 Análise dos Resultados

Com relação às conclusões que podemos tirar desse comparativo, abaixo são abordados cada um dos critérios pesquisados.

Com exceção do JUNG, que é um *framework*, todas as ferramentas já são prontas para uso e contam com uma interface gráfica. O Gephi possui também a biblioteca Gephi Toolkit, que empacota módulos essenciais dessa ferramenta e permite que programas Java utilizem-se das suas funcionalidades. O GUESS não conta com um instalador executável, por isso necessita de um pouco de conhecimento em linhas de comando. Para utilizar o JUNG é necessário algum conhecimento de Java, já que para manipular grafos mesmo de forma simples é preciso escrever código nessa linguagem.

Todas essas ferramentas são executáveis em três das principais plataformas de sistemas operacionais atuais (Windows, Mac OS e Linux). Há relatos de dificuldade de instalação do GUESS no MAC OS, por esse motivo foi disponibilizado no *site* da ferramenta um tutorial para ajudar nessa tarefa.

As informações sobre a quantidade de nós que pode ser tratada pelas ferramentas são um pouco divergentes. O *site* do GUESS não mencionava esse dado, então foi tomado como valor o tamanho da base de dados que o *site* disponibiliza para fins de exemplo. As demais ferramentas sugerem valores bastante razoáveis e informam também que essa quantidade varia em função da configuração do *hardware*.

Todos os *sites* contam com tutoriais, exemplos e fóruns de discussão. Alguns mais bem elaborados que outros. Merece destaque, por exemplo, o *site* do Gephi, que é bastante atualizado e conta com fóruns de discussão muito utilizados. Por outro lado, o *site* do GUESS, pelo que consta na própria página, não é atualizado desde o ano 2007.

Dentre os formatos de arquivos pesquisados, os únicos aceitos por todas as ferramentas são o Net (Pajek) e o GraphML, embora este último necessite de *plugin* para ser utilizado no Cytoscape. O Gephi se destaca por trabalhar com todos os formatos comparados, já o Cytoscape é o menos flexível a esse respeito. Vale salientar que Cytoscape pode tratar diversos arquivos de entrada, no entanto eles

têm um domínio muito específico, que é o da bioinformática, e não são interessantes para esse estudo comparativo. Embora o GUESS se proponha a suportar os formatos NET e GraphML, o site informa que eles não são completamente compatíveis e indica o uso do formato padrão do programa (o GDF).

Todas as ferramentas comparadas possuem algoritmos de leiaute de grafo. Além disso, três delas permitem incorporar novos através de *plugins*. O JUNG possui dentre os algoritmos implementados o de Fruchterman Reingold [31] e o de Kamada-Kawai [32]. Os web sites dessas três ferramentas incentivam os usuários/desenvolvedores a implementarem novos leiautes disponibilizá-los como novos *plugins*.

No que se refere a conexão com banco de dados, o Gephi é a única ferramenta que já é pronta para se conectar ao MySQL e ao PostgreSQL. No entanto, para o JUNG, possivelmente, não seria uma tarefa muito complicada implementar essa funcionalidade, uma vez que ele é baseado em Java e poderia usar as potencialidades dessa linguagem de programação.

Com relação às métricas de centralidade, o Cytoscape se destaca negativamente, uma vez que só implementa a medida de grau. As demais ferramentas, em relação a esse aspecto, ficam em uma certa situação de igualdade.

3.7 Considerações Finais

Este capítulo abordou de forma abrangente as principais características de ferramentas de ARS. Também tratou de forma mais específica quatro ferramentas *open source*, que tiveram alguns critérios relevantes comparados e condensados na forma de tabela.

A partir desse estudo comparativo, é possível dizer que o Gephi é uma boa opção de ferramenta *open source*. Ela se mostrou bastante completa em relação aos critérios abordados, notadamente no quesito de suporte ao aprendizado do seu uso. O Cytoscape se mostrou uma ferramenta bastante robusta para o campo original da bioinformática, mas esse viés penaliza um pouco a parte voltada para ARS e pode dificultar o uso do software por pessoas menos familiarizadas com Análise de Redes Sociais. O JUNG atende uma quantidade razoável dos critérios, no entanto exige

que o usuário possua alguma familiaridade com a linguagem de programação Java. O GUESS, como o próprio *site* atesta, ainda carece de muitas implementações de funcionalidades, o que só permitiria operações muito elementares de ARS. A seguir, o capítulo 4 propõe uma ferramenta de ARS voltada para o monitoramento de um aplicativo de rede social com fins educacionais.

Capítulo 4

Proposta de Ferramenta para a Rede Social da OjE

Este capítulo mostra uma proposta de ferramenta de Análise de Rede Social voltada para um aplicativo social da área de ensino básico. Pretende-se aqui que, ao final do capítulo, o leitor obtenha uma visão básica de como implantar uma ferramenta dessa natureza em redes sociais *on line* e de como a Análise de Rede Social pode ser utilizada para extrair informações relevantes. A seção 4.1 explica brevemente o que é a OjE. A seção 4.2 aborda essa ferramenta de ARS detalhando seu processo de concepção e implementação, bem como exibe algumas de suas funcionalidades.

4.1 O que é a OjE

A OjE (Olimpíada de Jogos Digitais e Educação) é uma plataforma *web* de rede social que possui diversos desafios apresentados aos jogadores na forma de jogos casuais e enigmas – questões construídas no formato das do ENEM (Exame Nacional do Ensino Médio). Essa ferramenta educacional tem como principal objetivo estimular o processo de aprendizagem e o diálogo entre alunos e professores na escola de Ensino Básico.

Atualmente, a plataforma tem sido implementada na rede pública de ensino de dois estados: Pernambuco (desde 2008) e Acre (desde 2011). Com base nesse sistema *web*, são realizadas competições anuais envolvendo toda a rede estadual. Os alunos formam equipes entre si e com seus professores para responder aos desafios e pontuar em diferentes disputas, a fim de ganharem prêmios virtuais e reais em eventos *online* e ao vivo durante o ano [38]. As figuras 16 e 17 mostram, respectivamente, a tela de um dos jogos e a de um dos enigmas.

Capítulo 4 -
Proposta de Ferramenta para a Rede Social da Oje

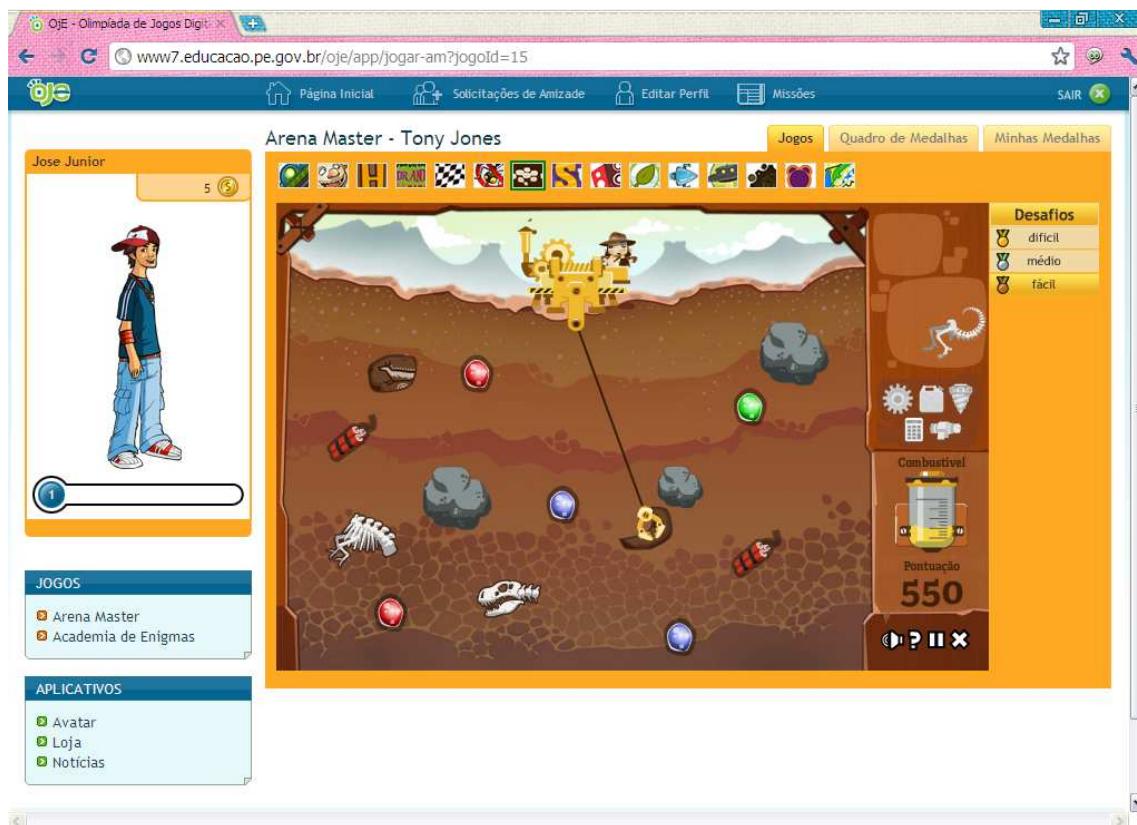


Figura 16 – Tela de um dos jogos da plataforma.

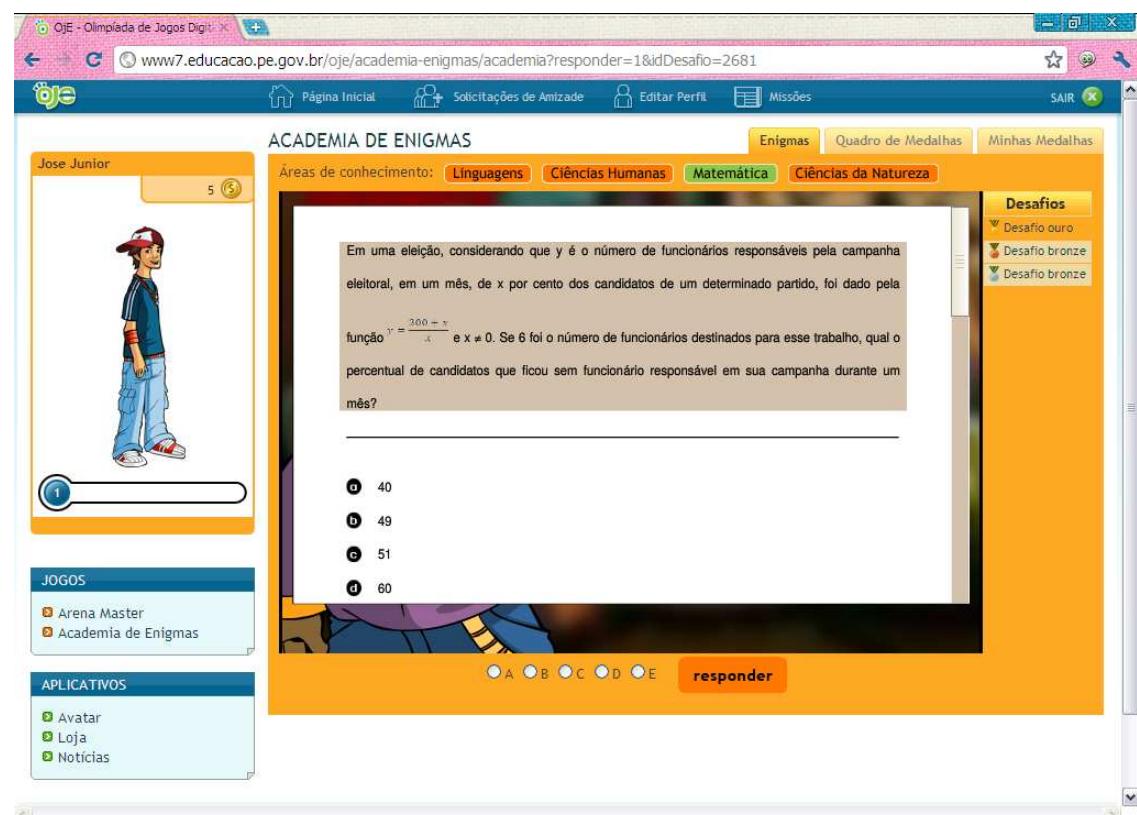


Figura 17 – Tela de um dos enigmas da plataforma.

4.2 Proposta de Ferramenta de ARS da OjE

A fim de monitorar as características da rede social da OjE, a equipe de pesquisa da Joy Street, empresa responsável pela implementação dessa olimpíada, está desenvolvendo uma ferramenta de Análise de Rede Social, que será integrada ao módulo de gerenciamento e acompanhamento da OjE, chamado internamente de Painel de Controle. Essa ferramenta, que possui o autor deste trabalho como um dos integrantes da equipe de concepção, pesquisa e desenvolvimento, já possui um protótipo funcional e possivelmente terá uma versão final em ambiente de produção no segundo semestre de 2011.

A intenção é fazer com que a ferramenta de monitoramento da rede seja acessível às Secretarias de Educação, aos professores das escolas e àqueles colaboradores da Joy Street responsáveis por decisões estratégicas referentes à Olimpíada. A considerável quantidade de futuros usuários, a não necessidade de instalação no computador cliente e a relativa facilidade de atividades manutenção levaram à escolha da web como plataforma de suporte para a ferramenta de análise.

A subseção 4.2.1 mostra uma abordagem simplificada da realizada na implementação de parte da ferramenta de monitoramento. A seção 4.2.2 mostra parte da versão final do protótipo e algumas de suas funcionalidades, bem como algumas informações que podem ser extraídas com essa ferramenta.

4.2.1 Criando e exibindo a visualização do grafo da rede social

Visando melhor demonstrar como pode ser implantada uma ferramenta de ARS em um ambiente web, esta subseção elabora um modelo simplificado seguindo a mesma abordagem utilizada no caso da ferramenta da OjE. A intenção é de que se obtenha uma visão básica porém abrangente dos passos necessários para esse tipo de projeto.

Neste modelo, para que consigamos criar e exibir a visualização da rede, serão necessários três passos principais: em primeiro lugar, devemos extrair informações sobre os usuários da rede e os relacionamentos estabelecidos entre eles; em segundo, devemos criar um arquivo com extensão gexf (arquivo proprietário do programa Gephi) que contém informações sobre a estrutura do grafo e as métricas calculadas; em terceiro lugar, devemos mostrar no navegador esse

arquivo criado. A figura 18 dá uma visão geral desses passos, representados, respectivamente, pelas setas nomeadas por “a”, “b” e “c”.

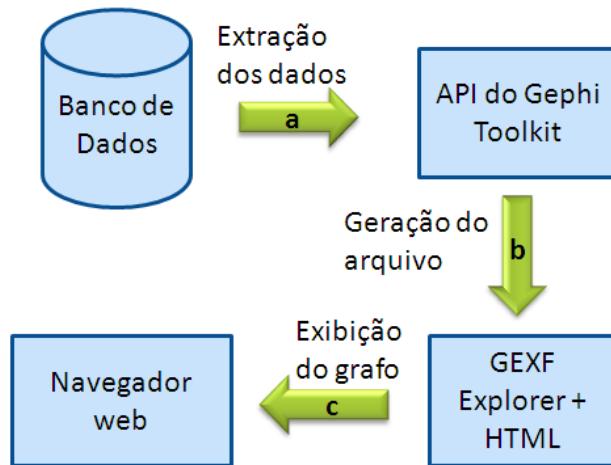


Figura 18 – Passos para geração e exibição do grafo da rede.

Os três itens a seguir abordam, detalhadamente, cada um desses passos.

(a) Extração dos dados

Um aplicativo real de rede social certamente deverá possuir as informações sobre seus usuários e os relacionamentos estabelecidos entre eles armazenados em algum local. Para fins de exemplificação, aqui serão criados dados fictícios que serão armazenados no gerenciador de banco de dados relacional MySQL [39]. Serão necessárias duas tabelas: uma que contém dados relacionados ao usuário (com os campos id, nome e cidade), chamada de ‘usuario’, e outra que guardará informações de relacionamentos (com os campos usuario_um e usuario_dois), chamada de ‘relacionamento’. Algumas características desse banco criado são exibidas na Figura 19, em que o item (a) mostra o Diagrama Entidade Relacionamento e os itens (b) e (c), os dados inseridos na tabela ‘usuario’ e ‘relacionamento’, respectivamente.

De posse dessas informações previamente armazenadas no banco de dados, podemos, então, extraí-las para que componham as propriedades do grafo. Os dados da tabela ‘usuario’ serão usados como atributos dos nós; os da tabela ‘relacionamento’, como arestas do grafo.

Essa fase de extração de dados foi separada das demais apenas por questões didáticas. Na verdade, a consulta ao banco de dados acontece na classe

Java que é implementada no próximo passo. Os códigos SQL de consulta ao banco já utilizados são: "select u.id as id, u.nome as label, u.cidade from usuario u" para extração das informações dos nós (usuários) e "SELECT r.usuario_um AS source, r.usuario_dois AS target FROM relacionamento r" para extração das arestas (relacionamentos).

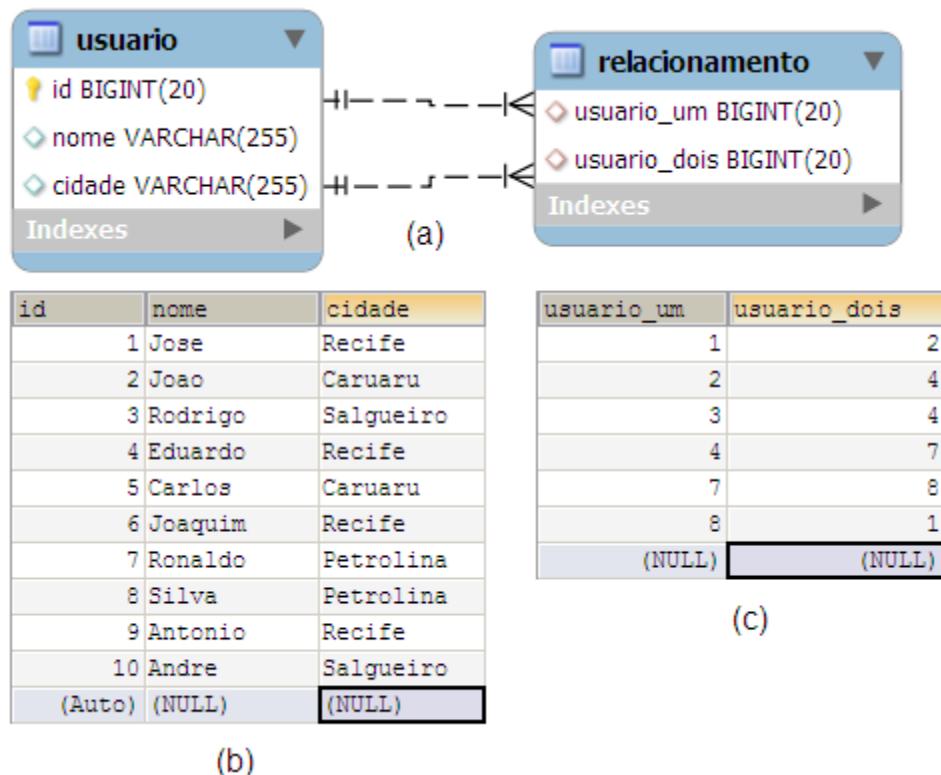


Figura 19 – Diagrama Entidade Relacionamento do banco criado (a), dados na tabela usuário (b) e dados da tabela relacionamento (c)

A seguir, o item (b) aborda a fase de criação do arquivo que descreve o grafo.

(b) Criação do arquivo do grafo

Esta fase é muito importante, pois é nela onde são aplicados os conceitos existentes no campo de estudo de ARS. Aqui é onde podem ser aplicadas, por exemplo, as medidas de centralidade estudadas anteriormente. Além disso, também estamos interessados em exibir em forma de grafo as características de nossa rede. Desse modo, é necessário usarmos algoritmos de leiaute de grafo (um grafo pode receber diversos algoritmos ao mesmo tempo).

De posse dos dados do item anterior, parte-se, então, para a geração do arquivo que organiza a rede em forma de grafo e que contém informações sobre os nós, as arestas e valores calculados para alguma(s) métrica(s) de interesse. A extensão desse arquivo é gexf, padrão do software Gephi. O gexf é baseado na linguagem de marcação XML (*Extensible Markup Language*) e ele pode ser criado tanto através da interface gráfica do próprio Gephi ou utilizando a sua biblioteca Java, chamada de Gephi Toolkit [40]. No entanto, se for desejável automatizar essa tarefa de criação, então é preciso trabalhar com a segunda opção. A seguir são mostrados os passos para a sua geração.

1. Fazer o download do Gephi Toolkit e criar um projeto Java que o referencia.
2. Implementar a classe que se conectará ao banco de dados e criará o arquivo gexf. Esta segunda etapa foi realizada utilizando o ambiente de desenvolvimento Eclipse Helios [41] e o código criado especificamente para este exemplo (anexo A). O código do exemplo possui comentários acerca do seu funcionamento para facilitar seu entendimento. Neste exemplo criado, os nós têm tamanho proporcional ao seu grau e os mais conectados têm cor de tonalidade vermelha. Para a criação do código fonte é bastante indicado se basear em alguns dos exemplos que acompanham o Gephi Toolkit, o que foi o caso neste trabalho.

(c) Integração do GEXF Explorer com código HTML e exibição no navegador

Depois de criado o arquivo da rede, para que seja possível exibi-lo no navegador será utilizado um aplicativo Flash *open source* e *free* chamado de GEXF Explorer [42]. Esse aplicativo, que por ser *open source* pode ser customizado (o que foi o caso no projeto da OjE), deve, então, ser incorporado ao código HTML da página que exibirá todo o conteúdo de interesse. O código HTML da Figura 20 é um exemplo básico criado para exibir o grafo da seção anterior.

```
1 <html>
2 <head>
3 <meta http-equiv="Content-Type" content="text/html; charset=utf-8">
4 <title>Exemplo de Ferramenta</title>
5 </head>
6 <body bgcolor="#E6E8FA">
7 <div id="conteudo">
8 <object width="550" height="400">
9   <param name="movie"
10    value="GexfExplorer1.0.swf?path=../test.gexf&fps=
11    true&labelsColor=0xFFEEDD&font=Arial&clickableNodes=true"/>
12   <param name="allowFullScreen" value="true" />
13   <param name="bgcolor" value="#FFFFFF" />
14   <embed src="GexfExplorer1.0.swf?path=../teste.gexf&fps=
15    true&labelsColor=0xFFEEDD&font=Arial&clickableNodes=true"
16    allowFullScreen="true" bgcolor="#111111" width="550" height="400">
17 </embed>
18 </object>
19 </div>
20 </body>
21 </html>
```

Figura 20 – Código HTML para exibir rede no navegador.

A seguir, são explicadas as principais características do código acima.

- Linha 4: é definido o título da página.
- Linha 6: define-se a cor do fundo da página, que, neste caso, será cinza.
- Linha 8: é a partir daqui até a linha 18 onde se configura o aplicativo Flash que exibirá a rede.
- Linha 10: nessa linha, são configurados os caminhos do aplicativo Flash [...] value="GexfExplorer1.0.swf? [...]) e do arquivo da rede [...] path=../test.gexf& [...]), que estão, especificamente neste exemplo, localizados na mesma pasta do arquivo HTML. Além disso, essa linha também sinaliza que será mostrado um indicador de quadros por segundo (fps) ([...] fps=true [...]).
- Linha 11: aqui se configura a cor e o tipo de fonte ([...] &labelsColor=0xFFEEDD&font=Arial& [...]) e sinalizamos se o nó responderá a eventos de clique ([...] &clickableNodes=true" [...]).
- Linha 12: esta parte do código diz se o aplicativo Flash irá possuir o modo tela cheia.

- Linha 13: nessa linha configuramos a cor do fundo do aplicativo Flash.
- Linhas 14 a 16: essas linhas têm a mesma função do código presente entre as linhas 8 e 13. O intuito dessa repetição é driblar a incompatibilidade dos navegadores servindo como um código alternativo caso não seja possível executar o primeiro (linhas 8 a 13).

A figura 21 mostra o resultado final exibido no navegador.

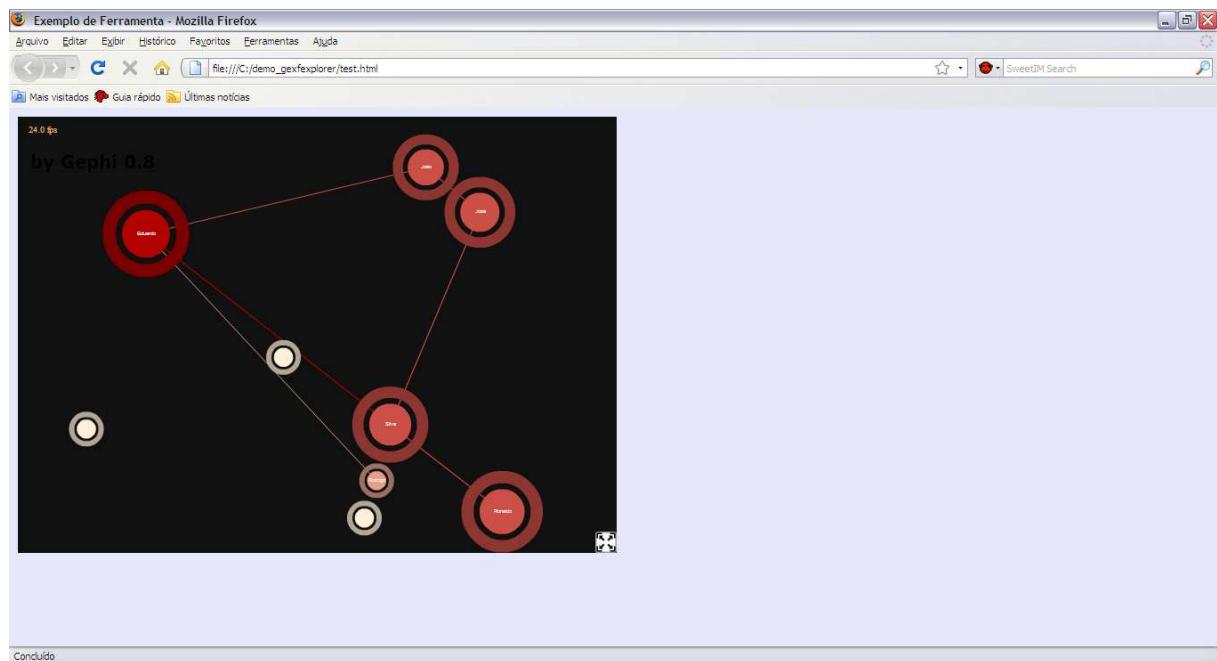


Figura 21 - Página HTML que exibe o grafo da rede.

4.2.2 Funcionalidades da ferramenta de ARS da OjE

A figura 22 exibe a tela principal da ferramenta. A região indicada pela seta vermelha é o mapa do estado de Pernambuco, em que cada cidade possui um marcador colorido de acordo com o valor de uma métrica que foi selecionada. As possíveis métricas são: número de inscritos na rede social, percentual de escolas que possuem participantes da OjE (esta é a métrica exibida na figura) e quantidade de acessos ao site da rede social na última semana, no último mês ou sem restrição de tempo.

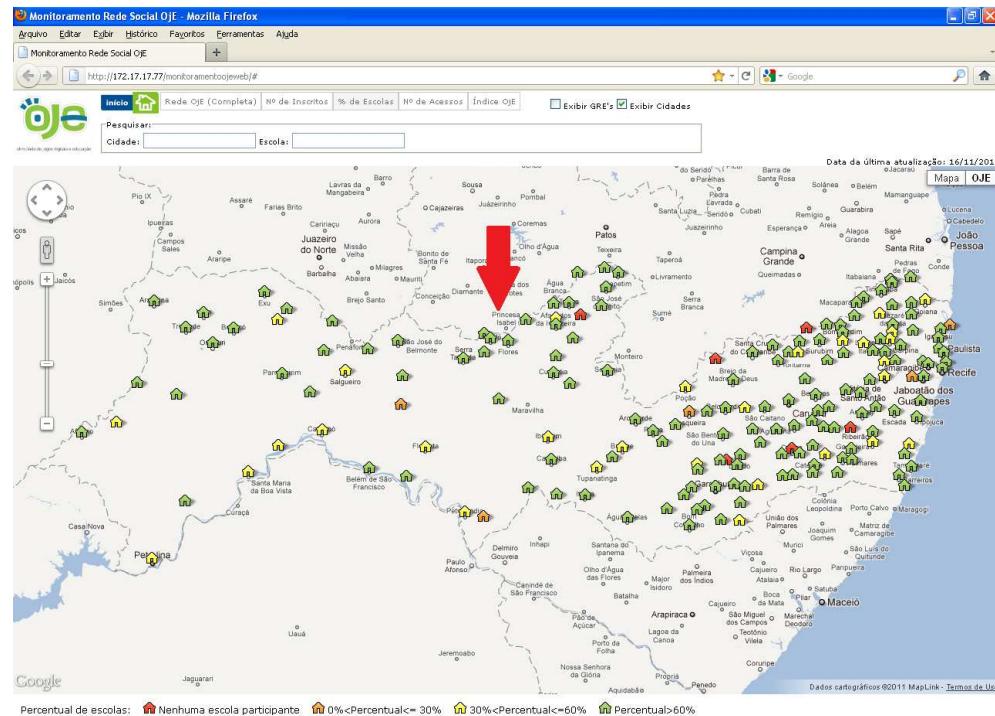


Figura 22 – Tela principal da ferramenta de monitoramento de OjE

A partir da interface gráfica exibida na figura 22, já seria possível retirar algumas informações relevantes sobre o uso do aplicativo de rede social. Poderíamos, por exemplo, verificar que:

- A maior concentração de cidades que participam da OjE se dá na região da Zona da Mata e na região Agreste do Estado.
- Caso houvesse algum padrão de cor em alguma sub-região, ele seria facilmente perceptível e indicaria características similares de comportamento naquela área.
- As cidades cujos marcadores estão na cor vermelha (no total de 6) mostram que não há nenhuma escola participante, isso demandaria uma ação de marketing nas escolas a fim de divulgar o produto da OjE.

Ao clicar em algum dos marcadores de cidades, abre-se uma tela que mostra o grafo da rede daquela localidade. A figura 23 exibe a tela que é aberta em resposta a esse evento de clique no marcador da cidade de Recife.

O layout do grafo em destaque nessa figura possui as seguintes propriedades:

- Cada usuário (aluno ou professor) da cidade é representado por um nó, cujo tamanho varia de acordo com a quantidade de amigos, ou, mais precisamente, com o grau do nó.
- Cada eixo do grafo representa os usuários pertencentes a uma mesma escola.
- Os nós (usuários) de cada eixo (escola) são ordenados de acordo com a sua quantidade de acessos ao site, de forma que os que possuem maior valor fiquem mais próximos do centro do grafo.
- A cor de cada nó varia de acordo com o número de acessos à rede social. As tonalidades presentes são: vermelho (menos acessos), amarelo (quantidade mediana de acessos) e verde (alta quantidade de acessos).

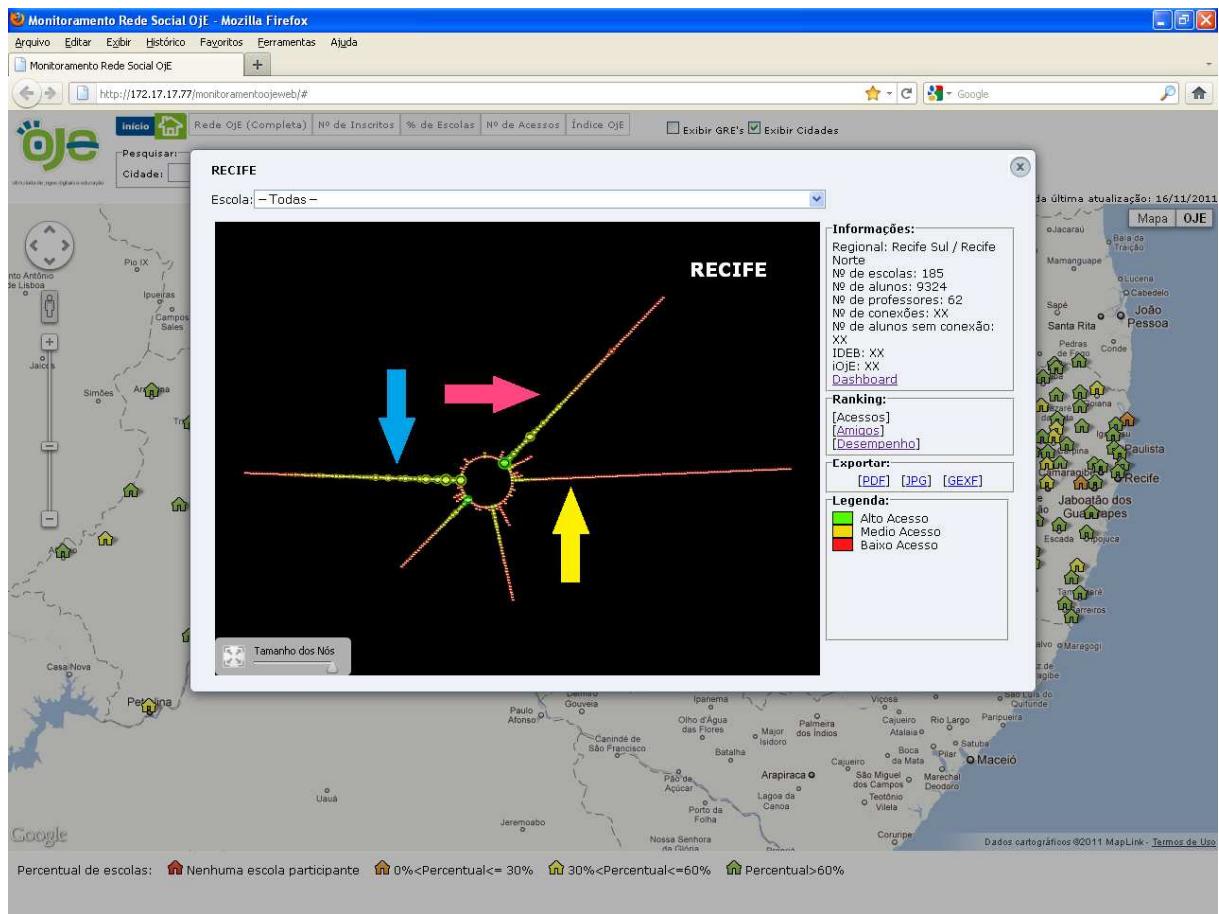


Figura 23 – Representação gráfica da rede da cidade de Recife

A partir desse cruzamento de informações existente no grafo é possível ter um panorama bastante útil do comportamento da rede da OjE. Algumas informações relevantes que podemos obter da análise desse grafo podem ser:

- Conseguimos distinguir facilmente três escolas da cidade de Recife que possuem mais usuários cadastrados que as demais. Na figura, elas são indicadas pelas três setas coloridas.
- A escola indicada pela seta amarela possui relativamente uma grande quantidade de usuários. No entanto, a média de acessos dos seus usuários é baixa, isso é evidenciado pelas tonalidades de cor dos nós, que são, na maioria, vermelhos ou amarelos. Além disso, podemos concluir que os professores e alunos dessa instituição de ensino ainda não possuem muitos amigos no aplicativo social, isso é perceptível pelo pequeno tamanho dos nós.
- Na cidade de Recife, os usuários que mais possuem amigos pertencem às escolas que mais possuem usuários cadastrados. Isso é perceptível pelo fato de que os nós de maior tamanho (maior grau) encontram-se nos maiores eixos (escolas) do grafo.
- A escola indicada pela seta azul possui o maior número de usuários mais bem engajados socialmente (o eixo do grafo possui maior quantidade de nós com tamanhos maiores).

Quando o modo de visualização da rede é expandido para tela cheia (figura 24), ao se clicar em um nó, abre-se um painel que contém informações daquele usuário, bem como as conexões com seus amigos são realçadas no grafo.

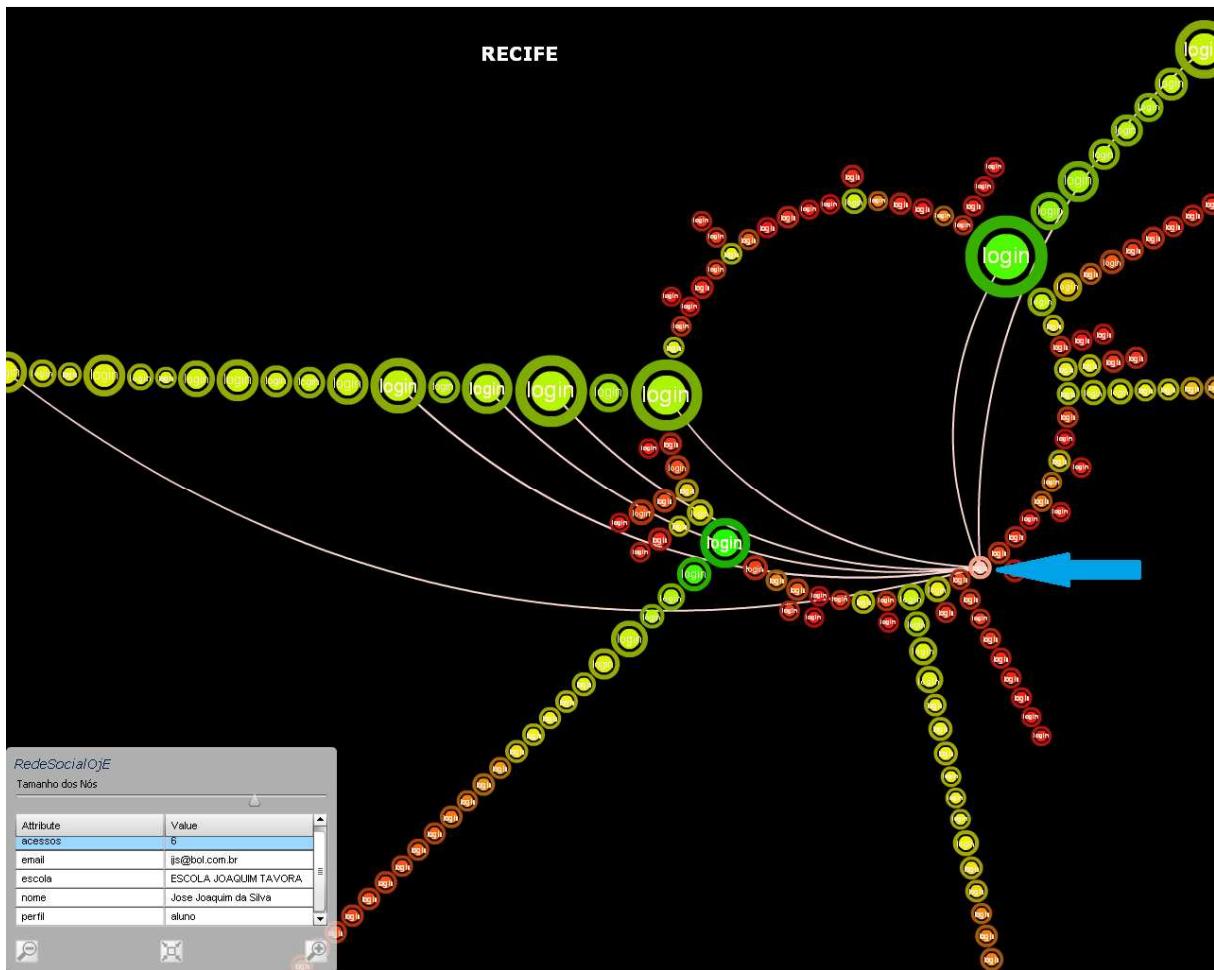


Figura 24 – Modo tela cheia da ferramenta exibindo grafo da rede de Recife

A partir dessa interface gráfica exibida nessa figura, podemos, por exemplo:

- Saber algumas informações cadastrais (nome, perfil, escola, email etc) e quantitativas dos usuários (por exemplo, o número de acessos). A seta em azul indica um aluno cujas informações estão aparecendo no painel à esquerda.
- Ver quem são os amigos de cada usuário da rede. Na imagem, as arestas exibidas mostram os relacionamentos do usuário indicado pela seta azul. Percebe-se que ele se relaciona com nós de duas das escolas que mais possuem usuários em Recife.
- Podemos inferir que em Recife os usuários mais engajados socialmente também estão entre os que mais acessam o *web site*. Isso pode ser percebido pelo fato de que nós de maior tamanho possuem tonalidades de cor mais próximas do verde e estão mais próximos do centro do grafo.

Vale salientar que outras métricas também foram incorporadas na ferramenta. Além das medidas de número de amigos e quantidade de acessos, também foram implementadas visualizações baseadas no desempenho individual dos usuários em enigmas e jogos. Com isso, é possível exibir um cruzamento de informações semelhante ao do exemplo anterior e monitorar a performance dos participantes por cidade e por escola.

4.3 Considerações Finais

Neste capítulo, foi apresentado como uma ferramenta de Análise de Rede Social pode ser incorporada a um aplicativo social. Além disso, mostrou-se também como informações relevantes podem ser retiradas utilizando essa ferramenta. Tudo isso serviu para dar um panorama básico de como a Análise de Rede Social pode ser utilizada para fins práticos e de como é possível fazer a implantação customizada desse tipo de ferramenta em uma plataforma *web*.

Segundo a visão do autor deste trabalho, a maior dificuldade encontrada na concepção e na utilização dessa ferramenta está na transferência de conceitos teóricos de Análise de Rede Social para a aplicação prática. No entanto, embora essa ferramenta incorpore apenas uma pequena parcela da imensidão de conceitos desse campo de estudo, ela mostra que é possível, sim, utilizar ARS para mapear o comportamento da rede em aplicativos sociais. Na sequência serão apresentados algumas conclusões e trabalhos futuros desta monografia.

Capítulo 5

Conclusão e Trabalhos Futuros

A seção 5.1 deste capítulo apresenta a conclusão do trabalho e a seção 5.2, sugestões de trabalhos futuros.

5.1 Conclusão

Este trabalho teve como contribuições permitir o entendimento de conceitos elementares de ARS, auxiliar na escolha de ferramentas *open source* de ARS e mostrar como esse campo de estudo pode ser aplicado para monitorar *sites* de rede social. O capítulo 2 serve como uma referência para aqueles interessados em conceitos teóricos de Análise de Rede Social, como medidas de centralidade, por exemplo. O capítulo 3 elicitou diversos critérios importantes de existirem em uma ferramenta de ARS e os comparou para quatro ferramentas, das quais o Gephi se mostrou uma das mais completas. O capítulo quatro expôs uma proposta de ferramenta de ARS voltada para a OjE, uma plataforma de rede social com fins educacionais, o que permitirá que interessados em aplicações prática possam ter uma ideia geral de como isso é possível.

5.2 Trabalhos Futuros

Durante o desenvolvimento desta monografia surgiram algumas ideias de trabalhos futuros. São elas:

- Incluir novas ferramentas no estudo comparativo, abordando, inclusive, ferramentas pagas, o que daria um leque muito maior de possibilidades de escolha de ferramentas de ARS.
- Incluir algoritmos de detecção de comunidades na ferramenta proposta.
- Realizar um estudo qualitativo com usuários de aplicativos de rede social e profissionais da área de Análise de Rede Social. Estudo esse visando eliciar melhores requisitos para a ferramenta proposta.

- Incluir novos leiautes de grafo na ferramenta proposta.

Bibliografia

- [1] WASSERMAN, S. e FAUST, K. **Social Network Analysis: METHODS AND APPLICATIONS.** CAMBRIDGE UNIVERSITY PRESS, 1994.
- [2] KERBAUY, M. T. M. e SANTOS, V. M. Dos. Educação e tecnologias: reflexão, inovação e práticas. In: BARROS, D. M. (Ed.). Lisboa: BARROS, D.M.V. et al., 2011. p. 978-989.
- [3] **Orkut.** Disponível em: <www.orkut.com>.
- [4] **Facebook.** Disponível em: <www.facebook.com>.
- [5] KUMAR, R.; NOVAK, J. e TOMKINS, A. Structure and evolution of online social networks. In: PROCEEDINGS OF THE 12TH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING - KDD '06. Anais... New York, New York, USA: ACM Press. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1150402.1150476>>, 2006.
- [6] MATHEUS, R. F. e SILVA, A. B. de O. E. **Fundamentação Teórica Para a Análise de Redes com Ênfase na Análise de Redes Sociais. Ciência da Informação.** 2005.
- [7] BARBOSA, T. de A. e RODRIGUEZ, M. V. R. SOFTWARES PARA ANÁLISE DE REDES SOCIAIS - ARS. In: VI CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO: ENERGIA, INOVAÇÃO, TECNOLOGIA E COMPLEXIDADE PARA A GESTÃO SUSTENTÁVEL. Anais... Niterói: SBC, 2010.
- [8] **LinkedIn.** Disponível em: <www.linkedin.com>.
- [9] **Twitter.** Disponível em: <www.twitter.com>.
- [10] BOYD, danah m. e ELLISON, N. B. Social Network Sites: Definition, History, and Scholarship. **Journal of Computer-Mediated Communication.**
- [11] **Hitwise.** Disponível em: <http://weblogs.hitwise.com/heather-dougherty/2010/03/facebook_reaches_top_ranking_i.html>. Acesso em: 8 out. 2011.
- [12] **E-Commerce News.** Disponível em: <<http://ecommerce-news.com.br/noticias/pesquisas-noticias/games-em-redes-sociais-podem-faturar-us-1-bilhao-em-2011>>. Acesso em: 9 out. 2011.
- [13] **Veja.** Disponível em: <<http://veja.abril.com.br/noticia/vida-digital/o-twitter-so-nao-faz-revolucao-mas-ajuda>>. Acesso em: 6 dez. 2011.
- [14] ORTIZ-ARROYO, D. Discovering Sets of Key Players in Social Networks. In: ABRAHAM, A.; HASSANIEN, A.-E.; SNÄSEL, V. (Eds.). **Computational Social Network Analysis.** [S.I.]: Springer, 2010. p. 485.
- [15] HANNEMAN, R. A. Introduction to Social Network Methods. **Network**, 1998. .
- [16] FREEMAN, L. C. Centrality in social networks conceptual clarification. **Social Networks**, 1978. v. 1p. 215-239.
- [17] MATHEUS, R. F. e SILVA, A. B. de O. E. **Análise de redes sociais como método para a Ciência da Informação. ARTIGO 03,** 2006.

- [18] FARAHAT, A. et al. Authority rankings from hits, pagerank, and salsa: existence, uniqueness, and effect of initialization □†. **Creativity**, p. 1-20, 2001.
- [19] NOORDHUIS, P.; HEJKOOP, M. e LAZOVIK, A. Mining Twitter in the Cloud: A Case Study. **2010 IEEE 3rd International Conference on Cloud Computing**, p. 107-114, doi:10.1109/CLOUD.2010.59, 2010.
- [20] BORGATTI, S. P.; EVERETT, M. G. e FREEMAN, L. C. Ucinet for Windows: Software for Social Network Analysis. **Harvard Analytic Technologies**, v. 2006, p. SNA Analysis software, 2002.
- [21] BATAGELJ, V. e MRVAR, A. Pajek-program for large network analysis. **Connections**, v. 21, n. 2, p. 47-57, 1998.
- [22] **INSNA**. Disponível em: <<http://www.insna.org/software/index.html>>. Acesso em: 4 nov. 2011.
- [23] ANDREA, A. D.; FERRI, F. e GRIFONI, P. An Overview of Methods for Virtual Social Networks Analysis. **Social Networks**, p. 3-25, doi:10.1007/978-1-84882-229-0, 2010.
- [24] ADAMIC, L. A. e BUYUKKOKTEN, O. **A social network caught in the Web**, 2003.
- [25] GIRVAN, M. e NEWMAN, M. E. J. Community structure in social and biological networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 12, p. 7821-6, doi:10.1073/pnas.122653799, 2002.
- [26] BACKSTROM, L.; HUTTENLOCHER, D. e KLEINBERG, J. **Group Formation in Large Social Networks□: Membership , Growth , and Evolution. Science**. [S.l: s.n.], 2006.
- [27] BASTIAN, M.; HEYMANN, S. e JACOMY, M. Gephi□: An Open Source Software for Exploring and Manipulating Networks. In: AMERICAN JOURNAL OF SOCIOLOGY. **Anais...** 2009.
- [28] **Cytoscape**. Disponível em: <<http://www.cytoscape.org/>>.
- [29] **JUNG**. . [S.l: s.n.]. Disponível em: <<http://jung.sourceforge.net/>>.
- [30] **GUESS**. Disponível em: <<http://www.graphexploration.cond.org/>>.
- [31] FRUCHTERMAN, T. M. J. e REINGOLD, E. M. Graph drawing by force-directed placement. **Software: Practice and Experience**, v. 21, n. 11, p. 1129-1164, doi:10.1002/spe.4380211102, 1991.
- [32] KAMADA, T. e KAWAI, S. AN ALGORITHM FOR DRAWING GENERAL UNDIRECTED GRAPHS. **Information Processing Letters**, v. 31, n. April, p. 7-15, 1989.
- [33] **Gephi**. Disponível em: <<http://gephi.org/>>.
- [34] **Cytoscape Wiki**. Disponível em: <<http://wiki.cytoscape.org/ScriptingPlugins>>, [S.d.].
- [35] **Social Graph API**. Disponível em: <<http://code.google.com/intl/en/apis/socialgraph/>>.
- [36] **R Project**. Disponível em: <<http://www.r-project.org/>>.
- [37] **RBVI Cytoscape Plugins**. Disponível em: <<http://www.cgl.ucsf.edu/cytoscape/cluster/clusterMaker.html>>.
- [38] MEIRA, L. et al. **Lições de uma olimpíada de jogos digitais para o monitoramento de redes de ensino**. 2011.

- [39] **MySQL Server**. Disponível em: <<http://www.mysql.com>>, [S.d.].
- [40] **Gephi Toolkit**. Disponível em: <<http://gephi.org/toolkit>>. Acesso em: 5 nov. 2011.
- [41] **Eclipse Helios**. Disponível em: <<http://www.eclipse.org/helios>>, 2010.
- [42] **GEXF Explorer**. Disponível em: <<http://gexf.net/explorer>>. Acesso em: 5 nov. 2011.

Apêndice A

Código usado para gerar o arquivo do Gephi

```

import java.awt.Color;
import java.io.File;
import java.io.IOException;
import org.gephi.data.attributes.api.AttributeColumn;
import org.gephi.data.attributes.api.AttributeController;
import org.gephi.data.attributes.api.AttributeModel;
import org.gephi.graph.api.GraphController;
import org.gephi.graph.api.GraphModel;
import org.gephi.io.database.drivers.MySQLDriver;
import org.gephi.io.exporter.api.ExportController;
import org.gephi.io.importer.api.Container;
import org.gephi.io.importer.api.EdgeDefault;
import org.gephi.io.importer.api.ImportController;
import org.gephi.io.importer.plugin.database.EdgeListDatabaseImpl;
import org.gephi.io.importer.plugin.database.ImporterEdgeList;
import org.gephi.io.processor.plugin.DefaultProcessor;
import org.gephi.layout.plugin.forceAtlas.ForceAtlasLayout;
import org.gephi.project.api.ProjectController;
import org.gephi.project.api.Workspace;
import org.gephi.ranking.api.Ranking;
import org.gephi.ranking.api.RankingController;
import org.gephi.ranking.plugin.transformer.AbstractColorTransformer;
import org.gephi.ranking.plugin.transformer.AbstractSizeTransformer;
import org.gephi.statistics.plugin.GraphDistance;
import org.openide.util.Lookup;
import org.gephi.ranking.api.Transformer;

public class MYSQLImportExport {

    public void script() {
        //Inicia um projeto e uma workspace
        ProjectController pc =
        Lookup.getDefault().lookup(ProjectController.class);
        pc.newProject();
        Workspace workspace = pc.getCurrentWorkspace();

        //Inicializa os controladores e o modelo
        ImportController importController =
        Lookup.getDefault().lookup(ImportController.class);
        GraphModel graphModel =
        Lookup.getDefault().lookup(GraphController.class).getModel();
        AttributeModel attributeModel =
        Lookup.getDefault().lookup(AttributeController.class).getModel();

        //Importa os dados do banco de dados
        EdgeListDatabaseImpl db = new EdgeListDatabaseImpl();
        db.setDBName("baserede");
    }
}

```

```

db.setHost("localhost");
db.setUsername("root");
db.setPasswd("mysql");
db.setSQLDriver(new MySQLDriver());
db.setPort(3306);
db.setNodeQuery("select u.id as id, u.nome as label, u.cidade from
usuario u");
db.setEdgeQuery("SELECT r.usuario_um AS source, r.usuario_dois AS
target FROM relacionamento r");
ImporterEdgeList edgeListImporter = new ImporterEdgeList();
Container container = importController.importDatabase(db,
edgeListImporter);
container.setAllowAutoNode(false); //Não cria nós para substituir
os ausentes
container.getLoader().setEdgeDefault(EdgeDefault.UNDIRECTED); //O
grafo será não direcional

//Adiciona os dados importados à GraphAPI
importController.process(container, new DefaultProcessor(),
workspace);

//Cria uma escala de cor para colorir os nós de acordo com seu grau
RankingController rankingController =
Lookup.getDefault().lookup(RankingController.class);
Ranking degreeRanking =
rankingController.getModel().getRanking(Ranking.NODE_ELEMENT,
Ranking.DEGREE_RANKING);
AbstractColorTransformer colorTransformer =
(AbstractColorTransformer) colorTransformer

rankingController.getModel().getTransformer(Ranking.NODE_ELEMENT,
Transformer.RENDERABLE_COLOR);

colorTransformer.setColors(new Color[]{new Color(0xFEF0D9), new
Color(0xB30000)});
rankingController.transform(degreeRanking,colorTransformer);

//Calcula medida de centralidade
GraphDistance distance = new GraphDistance();
distance.setDirected(false);
distance.execute(graphModel, attributeModel);

//Modifica o tamanho de cada nó de acordo com a medida de
centralidade calculada anteriormente
AttributeColumn centralityColumn =
attributeModel.getNodeTable().getColumn(GraphDistance.BETWEENNESS);
Ranking centralityRanking =
rankingController.getModel().getRanking(Ranking.NODE_ELEMENT,
centralityColumn.getId());
AbstractSizeTransformer sizeTransformer = (AbstractSizeTransformer)

rankingController.getModel().getTransformer(Ranking.NODE_ELEMENT,
Transformer.RENDERABLE_SIZE);
sizeTransformer.setMinSize(20);
sizeTransformer.setMaxSize(50);
rankingController.transform(centralityRanking,sizeTransformer);

//Executa o algoritmo de layout
ForceAtlasLayout secondLayout = new ForceAtlasLayout(null);
secondLayout.setGraphModel(graphModel);

```

```
secondLayout.resetPropertiesValues();
for (int i = 0; i < 10 && secondLayout.canAlgo(); i++)
    secondLayout.goAlgo();

//Exporta o arquivo criado para a pasta raiz do projeto
ExportController ec =
Lookup.getDefault().lookup(ExportController.class);
try {
    ec.exportFile(new File("teste.gexf"));
} catch (IOException ex) {
    ex.printStackTrace();
    return;
}
// Método principal que cria o objeto que gerará o grafo
public static void main(String args[]){
    new MySQLImportExport().script();
}
```