UNIVERSIDADE SALVADOR – UNIFACS
DEPARTAMENTO DE ENGENHARIA E ARQUITETURA
DEPARTAMENTO DE CIÊNCIAS EXATAS E DE COMUNICAÇÃO
MESTRADO EM REDES DE COMPUTADORES

IZABEL CRISTINA ANDION CASTRO

UMA FERRAMENTA PARA ANÁLISE DA COMUNICAÇÃO ORGANIZACIONAL ATRAVÉS DE REDES SOCIAIS

Salvador
2005

IZABEL CRISTINA ANDION CASTRO

UMA FERRAMENTA PARA ANÁLISE DA COMUNICAÇÃO ORGANIZACIONAL ATRAVÉS DE REDES SOCIAIS

Dissertação apresentada ao Curso de Mestrado Profissional em Redes de Computadores, Universidade Salvador – UNIFACS, como requisito parcial para obtenção do grau de Mestre.

Orientador: Prof. Dr. Manoel Gomes de Mendonça Neto

Salvador
2005

TERMO DE APROVAÇÂO

IZABEL CRISTINA ANDION CASTRO

UMA FERRAMENTA PARA ANÁLISE DA COMUNICAÇÃO ORGANIZACIONAL ATRAVÉS DE REDES SOCIAIS

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Redes de Computadores, Universidade Salvador – UNIFACS, do Curso de Mestrado Profissional em Redes de Computadores, pela seguinte banca examinadora:

Prof. Manoel Gomes de Mendonça Neto (Orientador)
Doutor em Ciência da Computação. University of Maryland at College Park, UMCP, Estados Unidos.
Universidade Salvador - UNIFACS

Prof. Joberto Sergio Barbosa Martins
Doutor em Informática, Universite de Paris VI (Pierre et Marie Curie). U.P. VI, França.
Universidade Salvador – UNIFACS

Prof. Cláudio Guimarães Cardoso
Doutor em Comunicação e Cultura Conteporânea, Universidade Federal da Bahia, Brasil.
Universidade Federal da Bahia - UFBA

Salvador, 30 de setembro de 2005

AGRADECIMENTOS

Agradeço inicialmente a meus queridos pais, Carolina e Hipólito, que me deram oportunidade de evolução pessoal e profissional através de muito carinho, educação e de princípios éticos que sempre conduziram minha vida.

Aos meus filhos Ricardo e Filipe e a meu companheiro de todos os momentos Carlos por terem apoiado minha decisão de mudança profissional na busca de novos desafios.

Ao meu irmão Eugênio que com sua competência e observações contribuiu para o êxito do estudo de caso.

Aos meus colegas do NUPERC (Núcleo de Pesquisas em Redes e Computação da Universidade Salvador) que me apoiaram em vários momentos com conselhos e orientações técnicas, principalmente Manuel Marques Neto e Marcio Almeida que permitiram o reuso de parte do seu código e compartilharam experiências de sucesso.

Agradeço especialmente ao meu orientador Manoel Mendonça e ao professor Celso Seibel que através de idéias, trocas de conhecimentos e discussões contribuíram decisivamente para o enriquecimento do trabalho.

Finalmente, agradeço aos diretores da ADN por terem confiado em mim e no meu trabalho, disponibilizando horas de seus funcionários e principalmente seus dados, para que eu pudesse comprovar a aplicabilidade da ferramenta proposta.

RESUMO

Esta dissertação descreve uma abordagem para a análise de redes de comunicação em organizações através de uma ferramenta computacional. A ferramenta modela a comunicação da organização por meio de grafos, onde os colaboradores são representados por nós e as trocas de fluxo de informações e conhecimento são representadas por arestas. Isto cria um mapa matematicamente analisável da rede de colaboração de uma organização. A ferramenta permite a execução de algoritmos de Análise de Redes Sociais (*SNA*) e a exploração visual do grafo modelado. Se comparada a outros produtos de *SNA*, a ferramenta desenvolvida incorpora maior semântica às análises. Isto é alcançado graças à definição dinâmica dos atributos que são associados aos objetos estudados (colaboradores e fluxos de informações). Simplifica também a exploração das redes modeladas e das métricas obtidas através do uso de técnicas de mineração visual de dados. Os grafos construídos podem ser explorados interativamente, e o mapeamento entre seus atributos visuais e os atributos dos objetos modelados pode ser dinamicamente modificado. Atributos visuais, em particular cores, também são usados para apresentar os valores das métricas computadas pelos algoritmos de análise de redes sociais. Isto permite ao usuário a interpretação visual de relacionamentos e papéis na rede. O ferramental desenvolvido foi utilizado em um estudo de caso, onde foi analisada a rede de comunicação organizacional associada à operação do *Help Desk* de uma empresa de médio porte. O estudo indicou que a ferramenta facilita a identificação de comunidades de prática, de padrões de comunicação e de indivíduos responsáveis pela dinamização da rede modelada.

Palavras-chave: Tecnologia da Informação: Gestão do Conhecimento, Análise de Redes Sociais e Mineração Visual de Dados.

.

ABSTRACT

This work describes an approach to analyze organizational communication networks through the use of a computing tool. This tool models the communication as a graph, where people are represented as nodes and information flows are represented as edges. This creates a map of the organization collaboration network, which can be mathematically analyzed. The tool implements Social Network Analysis (*SNA*) algorithms and allows the visual exploration of the modeled graph. Compared to other *SNA* products, the developed tool incorporates greater semantic capability through the dynamic definition of the attributes associated with the objects under study, i.e. people and communication flows. The tool also simplifies the exploration of the modeled networks and the results obtained by the *SNA* algorithms through the use of visual data mining techniques. The built graphs can be interactively explored and the mapping between their visual attributes and the attributes of the modeled objects can be dynamically changed. Visual attributes, colors in particular, can also be used to represent the values of the metrics computed by the *SNA* algorithms. This enables users to visually interpret relationships and roles over the modeled network. The developed tool was applied in a case study aimed at analyzing the communication network associated with a *Help Desk* in a medium sized organization. The study indicated that the tool facilitates identification of communities of practice, communication patterns, and key individuals over the modeled network.

Keywords: Information Technology: Knowledge Management, Social Network Analysis and Visual Data Mining.

LISTA DE FIGURAS

Figura 1. Processos de conversão do conhecimento adaptado de Nonaka e Takeuchi (1997). 21
Figura 2. Modelo para Gerenciamento do Conhecimento traduzido de Earl (1997). 24
Figura 3. Categorização de ferramentas para Gestão do Conhecimento. 25
Figura 4. Representação de uma rede social com grafos gerada pela ferramenta *UCINET 6* (BORGATTI;EVERRET;FREEMAN, 2002). 32
Figura 5. Rede egocêntrica de Marcos utilizando ferramenta *UCINET 6.* 34
Figura 6. Representação de grafo não direcionado. 35
Figura 7. Representação de grafo direcionado. 35
Figura 8. Grafo e um subgrafo. 36
Figura 9. Grafo direcional conexo. 37
Figura 10. Grafo não direcional desconexo. 37
Figura 11. Grafo formado de dois componentes conexos S1 e S2. 37
Figura 12. Grafo fortemente conexo. 38
Figura 13. Componentes fortemente conexos. 38
Figura 14. Ponto de articulação e ponte. 39
Figura 15. Grafo (a) representado por lista de adjacência (b) e matriz de adjacência (c). 40
Figura 16. Grafo e sua matriz de distâncias. 43
Figura 17. Grafo e matriz de distâncias adaptado de Chachra (1997). 43
Figura 18. Um grafo e seus *cliques* adaptado Waserman e Faust (1999). 44
Figura 19. Grafo ilustrando *n-clique* e *n-clã* adaptado Waserman e Faust (1999). 45
Figura 20. Um *3-clique* e um *3-plex* extraídos de Scoot (2000). 46
Figura 21. Exemplo de Cross e Prusak (2002) utilizando ferramenta *UCINET 6.* 53
Figura 22. Exemplo adaptado de Cross e Prusak (2002) utilizando ferramenta *UCINET 6.* 54
Figura 23. Modelo de processos *SNA* 56
Figura 24. Garfo apresentado pela ferramenta *UCINET 6* e alguns recursos de interação. 60
Figura 25. Processo de *KDD* e seus passos (MENDONÇA, 2001). 64
Figura 26. Modelo de referência de visualização (CARD; MACKINLAY; SHNEIDERMAN, 1999) 67
Figura 27. Técnica orientada a pixel – padrões recursivos (KEIM, 2002) 68
Figura 28. Coordenadas paralelas (KEIM, 2002). 69

Figura 29. Imagem da Região dos Grandes Lagos (ANKERST, 2000 apud MARQUES NETO, 2004) 70
Figura 30. TreeMap ferramenta para visualização hierárquica. 71
Figura 31. Cálculo do melhor tamanho das arestas em força-dirigida (NIGGEMANN, 2001) 72
Figura 32. O três passos do desenho orientado a nível (NIGGEMANN, 2001) 73
Figura 33. Modelo de dados gerado pelo produto *MS ACCESS* 81
Figura 34. Visualização através de grafos usando força-dirigida (TOUCHGRAPH, 2004). 87
Figura 35. Arquitetura da ferramenta *FlowMiner* 88
Figura 36. Diagrama de caso do uso da ferramenta *FlowMiner* 90
Figura 37. Diagrama de classes que implementam o módulo de *Entrada de Dados* 91
Figura 38. Tela inicial do *FlowMiner* mostrando subgrupos encontrados na base de dados. 92
Figura 39. Tela principal do *FlowMiner* mostrando subgrupo 1 escolhido. 93
Figura 40. Diagrama das classes do *TouchGraph* estendidas para *FlowMiner* 94
Figura 41. Diagrama das classes que implementam detalhes dos objetos da cena visual. 95
Figura 42. Foco no painel de *Cenários* preenchido com os termos escolhidos 96
Figura 43. Resultado na cena visual apresentando apenas os elementos selecionados 96
Figura 44. Foco no painel de *Controles* e *Filtros* 97
Figura 45. Diagrama de caso do uso de *Controles de Interação* 97
Figura 46. Foco nos grafos antes e depois da utilização do controle de aproximação 98
Figura 47. Diagrama das classes que implementam controle de aproximação 99
Figura 48. Foco nos grafos antes e depois da utilização do controle de rotação 99
Figura 49. Diagrama das classes que implementam controle de rotação 100
Figura 50. Foco no grafo com raio de exibição quatro e com raio de exibição um. 100
Figura 51. Diagrama das classes do controle de raio de exibição 101
Figura 52. Foco no painel de *Filtros* com opções disponíveis. 101
Figura 53. Foco no grafo e no respectivo controle de *Filtros*. 102
Figura 54. Diagrama das classes de *Filtros* 103
Figura 55. Foco na aba *Consultas Visuais* 103
Figura 56. Tela modificada após a seleção do atributo *Departamento* de atores 104
Figura 57. Diagrama da classe *Histograma* 105
Figura 58. Diagrama de caso de uso do módulo *Métricas*. 106
Figura 59. Foco na quarta área, com as métricas da rede total. 107

Figura 60. Diagrama das classes utilizadas nas métricas estruturais 108
Figura 61. Diagrama das classes envolvidas nas métricas estruturais e por ator 109
Figura 62. Transformação do papel de cada ator a depender da métrica aplicada. 110
Figura 63. Tela principal após acionada a opção métrica de *Centralidade de Proximidade*. 110
Figura 64. Foco na tabela preenchida com *Centralidade de Proximidade* 111
Figura 65. Tela principal mostrando resultado da métrica *Corretagem* pelo ator Blanco. 111
Figura 66. Diagrama das classes envolvidas na *Rede Egocêntrica* 113
Figura 67. Tela inicial com a composição de subgrupos da comunidade. 122
Figura 68. Tela principal mostrando a rede total 123
Figura 69. Aproximando e solicitando detalhe da ligação de maior intensidade 125
Figura 70. Centralidade dos atores por grau de conexões (ligações) 126
Figura 71. Centralidade dos atores por grau de fluxos (mensagens trafegando) 127
Figura 72. Centralidade dos atores por *Centralidade de Intermediação* 129
Figura 73. Rede de Esclarecimento de dúvidas 132
Figura 74. Centralidade por fluxo de saída na comunidade de Esclarecimento de dúvida 133
Figura 75. Foco na estatística da rede ego de Cristiano 135
Figura 76. Foco no mapeamento da rede ego de Cristiano 135
Figura 77. Foco na estatística da rede ego de Mauro 136
Figura 78. Foco no mapeamento da rede ego de Mauro 136

LISTA DE QUADROS

Quadro 1. Fórmula da densidade de uma rede direcional 42
Quadro 2. Fórmula da distância Euclidiana das ligações que saem de um ator 46
Quadro 3. Fórmula de Centralidade de grau. 48
Quadro 4. Fórmula de Centralidade de proximidade. 49
Quadro 5. Fórmula de Centralidade de intermediação. 49

LISTA DE TABELAS

Tabela 1. Cinco maiores pontuações de centralidade de conexões e fluxos 128
Tabela 2. Quatro maiores pontuações em proximidade de envio. 128
Tabela 3. Maiores pontuações antes e após eliminação da não reciprocidade. 130
Tabela 4. Atores críticos da rede 130
Tabela 5. Medidas estruturais de subgrupos ou comunidades por assunto. 133
Tabela 6. Conectores centrais por fluxo de saída para cada subgrupo. 134

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO 15

1.1 OBJETIVO DA DISSERTAÇÃO 16
1.2 ABORDAGEM ADOTADA 17
1.3 TRABALHO DESENVOLVIDO 17
1.4 ORGANIZAÇÃO DA DISSERTAÇÃO 18

CAPÍTULO 2 - GESTÃO DO CONHECIMENTO 20

2.1 ESTRATÉGIAS PARA GESTÃO DO CONHECIMENTO 22
2.2 ANÁLISE DE REDES DE COMUNICAÇÃO E O CAPITAL SOCIAL 26
2.3 ANÁLISE DE REDES SOCIAIS 27
2.4 CONSIDERAÇÕES FINAIS 28

CAPÍTULO 3 - ANÁLISE DE REDES SOCIAIS 29

3.1 CONCEITOS DE REDES SOCIAIS 30
3.1.1 Ator, ligação, relação e conteúdo transacional 30
3.1.2 Modalidades e níveis de análise 32
3.1.3 Múltiplas relações 32
3.1.4 Direção e peso das relações 33
3.1.5 Redes egocêntricas 33
3.2 TEORIA DOS GRAFOS 34
3.2.1 Grafos 34
3.2.2 Adjacência ou vizinhança 35
3.2.3 Grau 36
3.2.4 Subgrafo 36
3.2.5 Cadeia 36
3.2.6 Caminho 36
3.2.7 Grafo conexo 37
3.2.8 Grafo desconexo 37
3.2.9 Componente conexo 37
3.2.10 Grafo fortemente conexo 38
3.2.11 Componentes fortemente conexos 38
3.2.12 Pontos de articulação e ponte 38
3.2.13 Representações computacionais 39
3.2.14 Operações sobre grafos 40
3.3 MÉTRICAS EM REDES SOCIAIS 40
3.3.1 Redes como um todo 41
3.3.2 O papel de cada ator na rede 47
3.3.3 Rede Egocêntrica 50
3.4 REDES SOCIAIS ORGANIZACIONAIS 50
3.5 METODOLOGIA PARA ANÁLISE DE REDES SOCIAIS 56
3.5.1 Coleta de Dados 56
3.5.2 Transformação 58
3.6 FERRAMENTAS DE ANÁLISE DE REDES SOCIAIS 58

3.6.1 UCINET 59
3.6.2 VISONE 60
3.6.3 NetVis 61
3.6.4 Conversation MAP 61
3.7 CONSIDERAÇÕES FINAIS 61
CAPÍTULO 4 - MINERAÇÃO VISUAL DE DADOS 63
4.1 VISUALIZAÇÃO DE INFORMAÇÕES 66
4.1.1 Técnicas orientadas a Pixel 68
4.1.2 Projeção Geométrica 69
4.1.3 Técnicas Iconográficas 69
4.1.4 Técnicas Hierárquicas 70
4.1.5 Técnicas Baseadas em Grafos 71
4.2 MECANISMOS DE INTERAÇÃO 73
4.2.1 Uso de atributos visuais 74
4.2.2 Transformações visuais após interações 74
4.2.3 Consultas dinâmicas 75
4.3 VISUALIZAÇÃO EM REDES SOCIAIS 76
4.4 CONSIDERAÇÕES FINAIS 77
CAPÍTULO 5 - A FERRAMENTA *FLOWMINER* 78
5.1 INOVAÇÕES DA *FLOWMINER* 79
5.2 USUÁRIOS DA FERRAMENTA 80
5.3 MODELO DE DADOS 81
5.4 REQUISITOS PARA UTILIZAÇÃO DA FERRAMENTA 84
5.5 REQUISITOS FUNCIONAIS DA FERRAMENTA 85
5.6 DESENVOLVIMENTO DA FERRAMENTA 86
5.7 ARQUITETURA DA FERRAMENTA 87
5.7.1 Módulo de Entrada de Dados 90
5.7.2 Módulo de Apresentação Visual 93
5.7.3 Módulo de Controle de Consulta 95
5.7.4 Módulo de Controle de Interações 97
5.7.5 Módulo de Métricas de rede social 105
5.7.6 Sobre implementação e reuso 113
5.8 CONSIDERAÇÕES FINAIS 114
CAPÍTULO 6 - ESTUDO DE CASO 116
6.1 CARACTERIZAÇÃO DO ESTUDO DE CASO 116
6.1.1 Caracterização da Organização 117
6.1.2 Metodologia adotada 118
6.1.3 Geração da base de dados 119
6.1.4 Exemplos de Uso da Ferramenta na ADN 121
6.2 DISCUSSÃO DOS RESULTADOS 136
6.3 CONSIDERAÇÕES FINAIS 138
CAPÍTULO 7 - CONCLUSÃO 139
7.1 CONTRIBUIÇÕES 140
7.2 TRABALHOS FUTUROS 141

REFERÊNCIAS 143

ANEXO A 149

ANEXO B 152

CAPÍTULO 1 - INTRODUÇÃO

Em meio à velocidade, variedade e complexidade crescente de informações e conhecimentos que se criam e se transformam na sociedade globalizada, as organizações e os indivíduos passam por uma intensa mudança no seu comportamento e forma do aprendizado.

Mudanças rápidas também significam conhecimentos rapidamente obsoletos e uma necessidade de realizar novas curvas de aprendizado em curtos períodos (PRUSAK, 1997). Há algum tempo atrás o mercado valorizava indivíduos com conhecimento específico no domínio de um assunto, logo depois surgiu a necessidade dos generalistas, aqueles indivíduos que possuem uma visão macro sobre muitos assuntos, apresentando um posicionamento crítico mais abrangente. Atualmente as exigências aumentaram, busca-se o generalista-especialista, aquele indivíduo preparado e qualificado dentro de um perfil multidisciplinar, um profissional com competência técnica e visão global (CASTRO, 2002).

Neste ambiente movido pela agilidade em adquirir novos conhecimentos, a rede de contatos que um indivíduo possui o auxilia na obtenção de alternativas ou soluções para novos problemas, na troca de experiências e na aceleração do aprendizado. Uma teia de contatos bem estabelecida e bem dimensionada permite ao indivíduo uma maior mobilidade na solução de problemas, na busca de conselhos e novas idéias.

Para planejar e gerenciar de forma pró-ativa, obtendo uma maior sintonia e participação da cadeia de colaboração, a gerência também necessita de conhecimento de como a organização funciona e se comunica. A compreensão do papel e da importância que cada colaborador desempenha na rede corporativa tanto formal quanto informal permite uma visão dinâmica e eficiente do capital humano e social disponível.

"A efetividade de um time está baseada no conhecimento e na experiência pessoal de seus membros, acrescido pela comunicação, troca de informações através de suas conexões diretas e indiretas" (KREBS, 1999, p. 8).

Um ambiente propício a atitudes de socialização do conhecimento, é um instrumento poderoso para se alcançar a organização que aprende. Cada indivíduo é um agente do saber, influenciando e sendo influenciado pelas práticas de sua comunidade.

Análise de Redes Organizacionais se propõe a aplicar a ciência de Análise de Redes Sociais (*Social Network Analysis - SNA*) na organização. Esta ciência olha para o complexo sistema humano como um sistema interconectado de nós e arestas, onde os nós são as pessoas ou grupos e as arestas seus relacionamentos que acontecem através das trocas constantes de fluxos de informações e conhecimento (KREBS, 1998). *SNA* está focada em descobrir os padrões de interação entre as pessoas, visualizando seus contatos e suas trocas de informações. O uso de imagens visuais onde a organização ou parte dela é apresentada como um grafo que se modifica em diversos cenários dependendo das dimensões e critérios analisados, possibilita uma visão crítica e auxilia na investigação e entendimento de comportamentos adotados pela comunidade.

Estudos mostraram que gerentes possuem diferentes visões das redes que os rodeiam. Em virtude de sua posição hierárquica ficam fora das interações do dia-a-dia que geram a estrutura informal da organização. Esta percepção é agravada quando estas interações acontecem em ambiente virtual ou de telecomunicações. Análise de Redes Sociais voltada para organizações, foca a organização informal mostrando como a comunicação e troca de informações efetivamente acontecem, em que circunstâncias o conhecimento é compartilhado e os caminhos que são percorridos (CROSS e outros, 2001).

"A organização formal é o esqueleto da empresa enquanto que a informal é o sistema nervoso central dirigindo o pensamento, ações e reações coletivas às necessidades do negócio". (KRACKHARDT; HANSON, 1993, p. 37).

1.1 Objetivo da Dissertação

O objetivo desta dissertação é a implementação de uma ferramenta que apóie a análise de redes sociais em organizações através da construção e da representação gráfica de mapas de comunicação a partir de fontes de informação já existentes. A principal contribuição deste trabalho será oferecer um mecanismo para apoiar o estudo do processo de socialização do

conhecimento, auxiliando na identificação de indivíduos que exercem papéis de líderes, formadores de opiniões e especialistas emergentes, facilitando as trocas de experiências e aprendizado e apoiando o entendimento dos padrões de colaboração organizacional.

Neste contexto, o trabalho pretende ainda explorar a aplicabilidade da ferramenta no auxílio ao entendimento e análise do comportamento de comunicação de uma comunidade organizacional.

1.2 Abordagem Adotada

Como decorrência dessa proposta, as características de uma ferramenta para mapeamento da comunicação, denominada *FlowMiner*, são apresentadas. Esta ferramenta é genérica o suficiente para que seja independente de uma organização ou domínios específicos. Um modelo de dados relacional expansível foi adotado como entrada para aumentar a semântica e flexibilidade na qualificação dos objetos estudados, ficando a critério de cada organização a quantidade e a descrição das categorias necessárias. Ao fim do trabalho um estudo de caso foi executado, com o intuito de subsidiar a avaliação da aplicabilidade da ferramenta em organizações.

1.3 Trabalho Desenvolvido

A ferramenta *FlowMiner* foi desenvolvida na plataforma *Java*™ (SUN, 2004), utilizando um arcabouço livre denominado *TouchGraph* (TOUCHGRAPH, 2004) também em *Java*, com a finalidade de oferecer uma representação em grafo para exploração visual de redes de comunicação organizacional.

A ferramenta *FlowMiner* disponibiliza diversos controles visuais que permitem interação com o grafo através de seleções de objetos por valores de atributos. Após estas interações, transformações são executadas estabelecendo novos cenários para análise. Nestes cenários surgem comunidades de práticas, onde métricas de análise de redes sociais podem ser aplicadas para avaliar a estrutura geral, entender o papel de cada indivíduo na manutenção da comunidade e como cada um é influenciado por ela.

Um estudo de caso aplicando a ferramenta foi desenvolvido, utilizando como dados os registros da comunicação entre colaboradores de uma empresa de consultoria de informática para resolver problemas e esclarecer dúvidas sobre manutenção de sistemas. Os dados coletados corresponderam a quatro meses de comunicação entre uma população de 67

indivíduos envolvidos em processo de manutenção de sistemas. Durante este período trafegaram na rede 1 734 mensagens de assuntos variados.

Após a coleta, os dados passaram por processos de transformações até se tornarem aptos à inclusão na base de dados da ferramenta. Uma vez inserido os dados, foi executada a análise da rede social apresentada utilizando a ferramenta *FlowMiner*. A análise constou de quatro passos: análise da rede como um todo, análise por ator, análise dos cenários de comunidades e análise das redes egocêntricas.

Ao final da análise e entendimento dos resultados apresentados, a ferramenta possibilitou as seguintes identificações e constatações:

- Hierarquização da rede;
- Necessidade de treinamento;
- Padrões de comportamentos;
- Identificação de comunidades de práticas;
- Identificação de indivíduos mais atuantes;
- Maiores colaboradores por assunto;

1.4 Organização da Dissertação

Esta dissertação está organizada em sete capítulos. Após este capítulo de introdução são apresentados mais três contendo a fundamentação teórica e revisão da literatura relacionada ao tema do trabalho. Nos três seguintes estão a descrição da ferramenta desenvolvida, o estudo de caso e a conclusão. Por fim aparecem as referências bibliográficas e eletrônicas e dois anexos contendo pseudo-códigos utilizados para implementação de algoritmos e os dados utilizados para comprovar a correção das métricas obtidas no estudo de caso. Sendo assim, esta dissertação está dividida nos seguintes capítulos:

- CAPÍTULO 1 – descreve a motivação, objetivos e escopo da pesquisa, como também a organização do documento da dissertação;
- CAPÍTULO 2 – apresenta a fundamentação teórica e a revisão de literatura sobre o tema Gestão de Conhecimento;
- CAPÍTULO 3 – apresenta as linhas teóricas e algumas ferramentas existentes de Análise de Redes Sociais, no intuito de contextualizar a ferramenta desenvolvida;

- CAPÍTULO 4 – apresenta as principais abordagens técnicas e mecanismos utilizados em Mineração Visual de Dados e que contribuições esta tecnologia pode agregar ao ferramental de Análise de Redes Sociais;
- CAPÍTULO 5 – descreve detalhadamente a ferramenta *FlowMiner* e sua implementação;
- CAPÍTULO 6 – faz uma avaliação da aplicabilidade da ferramenta *FlowMiner* através de um estudo de caso;
- CAPÍTULO 7 – apresenta as conclusões deste trabalho e comenta a evolução da ferramenta, relacionando os trabalhos futuros que podem ser realizados;
- REFERÊNCIAS – contêm uma lista composta dos documentos padronizados e utilizados na elaboração deste trabalho;
- ANEXO A – mostra os pseudo-códigos utilizados para desenvolver os principais algoritmos na ferramenta *FlowMiner*.
- ANEXO B – mostra as estatísticas geradas por uma ferramenta comercial utilizando a mesma rede do estudo de caso, com a finalidade de validar os resultados calculados pela ferramenta *FlowMiner* para as principais métricas estudadas.

CAPÍTULO 2 - GESTÃO DO CONHECIMENTO

Este capítulo desenvolve conceitos e teorias encontrados na literatura sobre o tema Gestão do Conhecimento e como estas idéias influenciaram e balizaram o desenvolvimento deste trabalho.

Nonaka e Takeuchi (1997) e Sveiby (1998) definem o conhecimento humano como um processo social entre indivíduos que é criado e expandido tanto em qualidade quanto em quantidade através de interações sociais. Categoriza-se o conhecimento tácito como aquele que é adquirido pelos indivíduos através da observação e experiência e ampliado com seus valores e habilidades. É um ativo do ser humano e não das organizações, sendo desta forma de difícil documentação e transmissão. Já o conhecimento conhecido como explícito é aquele adquirido pela educação formal, registrado e sistematizado.

Davenport e Prusak (1998) argumentam que a construção do conhecimento é um processo longo e confuso, e as maneiras de utilizá-los são múltiplas e imprevisíveis. Um ponto fundamental é que os seres humanos e suas relações são essenciais para conceber, interpretar e interiorizar esse complexo tipo de informação.

De acordo com Nonaka e Takeuchi (1997), a criação do conhecimento organizacional é um processo interminável que provém de constantes interações entre os conhecimentos tácitos e explícitos, exigindo uma atualização e inovação contínua das intenções e valores organizacionais. Uma empresa criadora de conhecimento opera num sistema aberto onde existem constantes intercâmbios de conhecimento no ambiente interno e externo. Os indivíduos não recebem passivamente o novo conhecimento, eles os interpretam ativamente, adequando-os à sua própria situação e perspectivas. A esse processo os autores denominam de

conversão do conhecimento que se processa em quatro etapas: socialização, exteriorização, combinação e interiorização como mostrado na Figura 1.

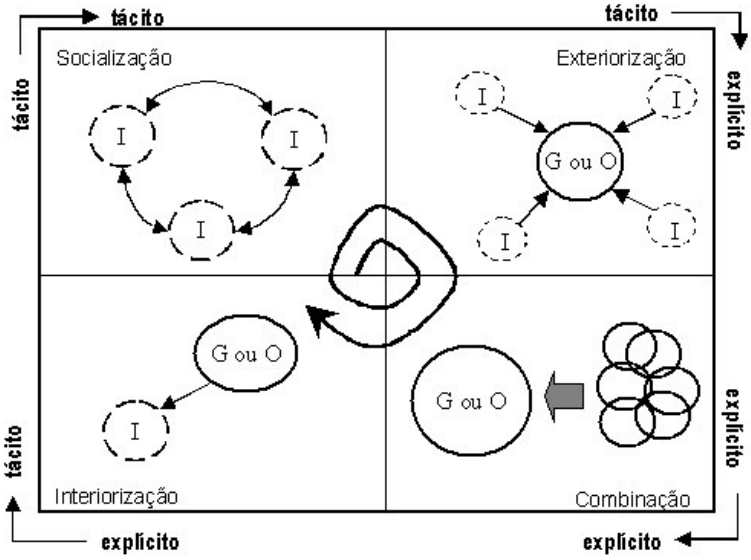

Figura 1. Processos de conversão do conhecimento adaptado de Nonaka e Takeuchi (1997).

A *socialização* desenvolve um ambiente propício à interação, facilitando o compartilhamento de experiências e práticas entre seus membros. Neste processo o conhecimento tácito é adquirido através da prática, observação e imitação. A comunicação e o diálogo são componentes importantes para reorientar os modelos mentais na mesma direção, em busca de sintonia.

A *exteriorização* é a transformação do conhecimento tácito em conceitos explícitos, considerado o processo chave para a criação do conhecimento. Através dele o conhecimento passa a ser representado por modelos, hipóteses e conceitos que devem ser expressos em uma linguagem sistêmica e lógica coerente, permitindo facilmente sua transmissão e compartilhamento.

A *combinação* é a sistematização do conhecimento associando diferentes formas do conhecimento explícito. As informações existentes são classificadas, acrescentadas, categorizadas levando à geração de novos conhecimentos. Este processo é facilitado com o uso de redes de comunicação e banco de dados.

Por fim a *interiorização* é a forma de incorporar o conhecimento explícito em tácito, está fortemente ligado ao aprendizado através da experiência, do experimento, podendo ser também adquirido através de relatos históricos de experiências bem sucedidas.

Nonaka e Takeuchi (1997) esclarecem que a criação do conhecimento organizacional é um processo em espiral que circula continuamente entre as quatro etapas citadas, e para ser alcançado deve começar no nível individual e ir se ampliando através da interação entre os grupos e comunidades atravessando toda a empresa. Um ambiente propício ao compartilhamento deve existir permitindo que cada indivíduo possa ser influenciado e ao mesmo tempo influenciar a espiral do conhecimento, desenvolvendo desta forma a cultura organizacional. A "organização que aprende" se alimenta com suas próprias experiências e influenciada pelo ambiente social gera e acumula novos conhecimentos. O termo Gestão do Conhecimento é utilizado por alguns autores para descrever uma proposta de instrumentação para a aprendizagem organizacional e está relacionado com as novas tecnologias da informação e comunicação (MESQUITA, 2003). Um grande desafio desta abordagem é a disponibilização de mecanismos eficientes e eficazes para capturar, armazenar e manter o conhecimento, bem como, uma infra-estrutura técnica propícia à criação, divulgação e disseminação do conhecimento.

Lévy (1999) observa que um ponto positivo da Gestão do Conhecimento é o enriquecimento humano, aumentando a competência dos indivíduos e grupos, promovendo a sociabilidade e o reconhecimento recíproco, proporcionando ferramentas de autonomia e criando a diversidade. O regime de produção e distribuição do saber depende não apenas das especialidades do sistema cognitivo humano, mas também dos modos de organização coletiva e dos instrumentos de comunicação e tratamento das informações.

As novas tecnologias de informação e comunicação têm sinalizado com ferramentas e artefatos que auxiliam e apóiam no processo de aprendizado organizacional. A busca pela gestão do conhecimento mostra que organizações estão valorizando os indivíduos, seus recursos humanos como fonte de conhecimento e que pretendem encontrar mecanismos para direcionar este conhecimento para o aumento da capacidade de inovação, de produtividade, além de tentar reter o conhecimento como corporativo (CIANCONI, 2001).

Um dos principais objetivos da Gestão de Conhecimento no meio empresarial é tornar claro o valor do capital humano da organização (PRUSAK, 2001).

2.1 Estratégias para Gestão do Conhecimento

Para Hansen, Nohia e Tierney (1999) existem duas estratégias de implementação de gestão do conhecimento nas organizações que são codificação e personalização. A codificação é uma estratégia centrada na TI que possibilita o registro do conhecimento em

grandes bases de dados corporativas para que os membros da organização possam buscá-lo e acessá-lo sem a necessidade do especialista que gerou seu conteúdo. Na década de 90 muitos sistemas voltados a esta estratégia foram desenvolvidos e implantados em grandes corporações trazendo muita controvérsia sobre o assunto, pois os resultados obtidos não condisseram com a expectativa gerada. Num trabalho de mestrado acadêmico da UFBA, Mesquita (2003) abordou este tema e concluiu que:

> [..]a ênfase na codificação de conhecimento em repositórios informatizados pode condicionar percepções de perpetuação dos tradicionais processos organizacionais. Além disso, argumentou-se que tal ênfase poderia restringir a capacidade inovadora dos trabalhadores envolvidos.[...] Para catalisar inovações, sejam elas micro ou macro-inovações, seriam necessárias outras ferramentas além da codificação de conhecimento em repositórios informatizados. Essas inovações estariam ligadas à emancipação dos sujeitos, e obtida através de técnicas sinérgicas de comunicação[...] (MESQUITA , 2003, p. 109-112)

Prusak (1997) também esclarece que a vantagem competitiva de organizações depende não só do que ela conhece, mas também de como usa o conhecimento e quão rápido pode conhecer algo novo. A produção de inovação é resultado do conhecimento do grupo em mercados não explorados e em novas possibilidades técnicas, operações efetivas vêm do conhecimento compartilhado de como as coisas funcionam e como poderiam funcionar.

A segunda estratégia de Gestão do Conhecimento apontada por Hansen, Nohia e Tierney (1999) é a personalização. A utilização do termo personalização é uma tradução direta do texto, observa-se que socialização seria mais bem empregado nesta situação, pois a explicação dos autores é que nessa estratégia a TI é utilizada como um elemento auxiliar, viabilizando ferramentas de comunicação que permitam o contato entre as pessoas para compartilhamento do conhecimento. O propósito desta estratégia é propiciar e otimizar o processo de socialização e colaboração entre seus membros.

"Nessa abordagem as redes passam a ter um papel fundamental no lugar dos bancos de dados, pois sua intenção é a circulação e disseminação das informações e não seu armazenamento" (MESQUITA, 2003, p. 76).

Earl (1997) fornece outra visão estratégica interessante sobre o mesmo assunto. Neste artigo ele descreve o tema Conhecimento como Estratégia tomando como base dois estudos de caso, o da Skandia International e o da Shorko Films.

Para o autor a construção do conhecimento requer a combinação de tecnologia e ação social. Uma estratégia adequada requer a construção de um modelo com quatro componentes

básicos: Sistema do Conhecimento, Redes, Trabalhadores do Conhecimento e Aprendizado Organizacional. Ver Figura 2.

Figura 2. Modelo para Gerenciamento do Conhecimento traduzido de Earl (1997).

O componente *Sistema do Conhecimento* é responsável pela captura das experiências, registro e disponibilização em bases de dados que serão analisadas e exploradas através de ferramentas de apoio a decisão. As *Redes* são significativas na captura, construção e disseminação do conhecimento. A construção do conhecimento é facilitada pela troca de documentação, hipóteses, dados e mensagens através da rede. Redes facilitam a eficiência da administração interna e contribuem para uma maior criatividade através do compartilhamento de problemas e soluções, desenvolvimento de produtos e construção de alianças.

O componente *Trabalhadores do Conhecimento*, segundo o autor, é o desafio pessoal de cada colaborador que necessita continuamente adquirir conhecimento para o crescimento de seu perfil e de sua experiência, tornando-o um bem mais valioso para a organização. Esta tendência sugere um movimento para o mérito nas organizações, não só para quem adquire novos conhecimentos, mas para quem promove o seu compartilhamento. Esta tendência a nível individual tem implicações a nível organizacional, o conhecimento só é maximizado se a organização aprende como um todo.

O último componente se refere à *Organização que Aprende*. Para que a organização aprenda, ele cita três indicadores necessários. O primeiro se refere à necessidade de colaboração no seu funcionamento, a existência de sistematização em rede promove a

elevação do nível de colaboração. Segundo o autor, a organização em rede, atravessa fronteiras internas e externas para trocar informações, quebrando posições estabelecidas, permitindo a localização de perfis relevantes ou experientes, criando sinergia de talentos compartilhados. O segundo indicador apontado é a criação de demandas de treinamento e desenvolvimento pessoal. A necessidade não é só de treinamento para utilizar a tecnologia, mas também no processo de aumentar as habilidades das pessoas em análise, raciocínio e dedução. A educação em aprendizagem organizacional é contínua. O último indicador diz respeito ao incentivo e controle à ética nas atividades de colaboração. Um problema que afeta a credibilidade do processo se refere ao oportunismo de alguns indivíduos que podem reter, distorcer ou usar em proveito próprio o conhecimento adquirido.

Combinando as idéias destes autores e acrescentando novas visões propomos uma nova categorização das ferramentas de TI para auxílio à gestão do conhecimento.

Figura 3. Categorização de ferramentas para Gestão do Conhecimento.

Na Figura 3 podemos observar dois níveis de ferramentas: as que possibilitam apoio à gestão do conhecimento e as que analisam o uso da gestão do conhecimento.

O apoio ao conhecimento explícito se dá através das ferramentas de codificação que disponibilizam informações e conhecimento em bases de dados para acesso coletivo. Já o conhecimento tácito é trocado através de ferramentas de socialização que propiciam a comunicação e o compartilhamento de experiências como: correio eletrônico, *chats*, *work group*, *help desk,* entre outros.

O segundo nível diz respeito a ferramentas que analisam e entendem a utilização do primeiro nível. Na análise das ferramentas de codificação pode-se avaliar a freqüência de uso, o perfil de utilização dos usuários, necessidade de atualização, feedback de usuários, enfim, avaliar a utilização efetiva do conhecimento armazenado. Na análise das ferramentas de redes

de comunicação pode-se avaliar o comportamento das comunicações, quem se comunica com quem, com que intensidade e freqüência e quais os assuntos de interesse.

2.2 Análise de Redes de Comunicação e o Capital Social

A proposta deste trabalho se enquadra na análise de redes de comunicação, especificamente na análise de redes de comunicação em organizações e está fortemente ancorada na teoria de Análise de Redes Sociais que será descrita no próximo capítulo. Estas ferramentas não enfatizam a codificação, classificação e armazenamento de conhecimento, mas sim o mapeamento e análise das interações sociais entre os indivíduos onde existem trocas de informações, conhecimentos, idéias e conselhos.

Borgatti e Foster (2003) esclarecem que existe um forte deslocamento em pesquisas organizacionais a partir da segunda metade do século 20, onde explicações individualistas, essenciais e atômicas estão sendo substituídas por um entendimento relacional, contextual e sistêmico. O crescimento apontado por eles nas pesquisas de redes organizacionais está na busca do capital social, um conceito que tem forte sintonia e aumenta o interesse na teoria da Análise de Redes Sociais. Existem duas conceituações clássicas para capital social

> A primeira está associada aos sociólogos Ronald Burt, Nan Lin e Alejandro Portes, que se refere aos recursos como informações, idéias, apoios que os indivíduos são capazes de procurar em virtude de suas ligações com outras pessoas. Estes recursos são sociais na medida em que são acessíveis dentro e por meio das relações, diferente do capital físico (ferramentas e tecnologias) e humano (educação e habilidades) que são propriedades do indivíduo. A estrutura da rede de um indivíduo, ou seja, com quem ele se relaciona, têm um papel fundamental no fluxo de recursos. Aqueles que ocupam posições estratégicas e têm ligações que alcançam vários grupos, potencialmente têm mais mobilidade e possuem mais recursos que os outros (GROOTAERT e outros, 2003, p. 5).

A segunda abordagem, associada a Robert Putman (2000 apud GROOTAERT e outros, 2003), refere-se à natureza e a extensão do envolvimento do indivíduo com várias redes informais e formais. Capital social nesta abordagem caracteriza-se nas variadas maneiras pelas quais os membros de uma comunidade interagem, é uma qualidade do grupo, são os cruzamentos de interconexões entre todos os membros do grupo. Resumindo, a primeira abordagem foca o indivíduo e a segunda foca o grupo (BORGATTI; JONES; EVERETT, 1998; GROOTAERT e outros, 2003).

Generalizando, o conceito de capital social se baseia no valor das conexões para fins produtivos, é a riqueza que pode derivar de contatos, ligações e da capacidade de trabalhar bem com outras pessoas. Um valioso aspecto do conhecimento humano está inserido nas

relações sociais. Várias teorias e trabalhos sobre gerenciamento organizacional tentam explicar através do capital social como ligações pessoais ou posição na rede interferem no significado de poder, liderança, mobilidade, empregabilidade, desempenho individual, criatividade e desempenho em grupo (BORGATTI; FOSTER, 2003).

"Pesquisas mostraram que quem você conhece, tem um impacto significante em quem você virá a conhecer. Seus relacionamentos são críticos para obter informações, resolver problemas e aprender como se trabalha" (CROSS e outros, 2001, p. 100).

2.3 Análise de Redes Sociais

Conhecimento existe e cresce principalmente nas complexas estruturas de redes internas e suas comunidades, tornando estas estruturas uma das unidades de análise mais produtiva em trabalhos do conhecimento (PRUSAK, 2001).

Numa organização, as ferramentas de Análise de Redes Sociais não possibilitam apenas a apresentação das redes hierárquicas formalmente conhecidas, na qual observa-se quem trabalha onde e quem se reporta a quem, mas também mostra o mapeamento do contato informal entre as pessoas e as informações compartilhadas entre elas, respondendo a questões como quem conhece quem (capital social) e quem conhece o que (capital humano).

Como observou Prusak (2001), Gestão do Conhecimento usa a idéia do capital humano como uma vantagem financeira obtida com o investimento em educação e treinamento dos funcionários, no desenvolvimento de suas competência e habilidades, trazendo um retorno em produtividade, desenvolvimento de perfis, capacidade de inovação e mobilidade no trabalho. Contudo o maior objetivo é no conhecimento do grupo e nos processos do capital social.

Análise de Redes Sociais ilumina a organização informal, mostrando como a comunicação e a troca de informações efetivamente acontecem dentro da organização (CROSS e outros, 2001). Olha o que as pessoas fazem, as circunstâncias em que elas compartilham o conhecimento, os caminhos que usam, mudam ou ignoram (PRUSAK, 2001).

O mapeamento da rede de comunicação consiste em mostrar como os fluxos de informações fluem na organização através de seus colaboradores, apontando elementos que influem neste processo de forma positiva ou negativa, criando assim oportunidades para corrigir e melhorar este ambiente a fim de propiciar uma maior troca e disseminação de conhecimentos entre os indivíduos, facilitando a socialização, integrando pessoas, processos e projetos.

A socialização é um processo de compartilhamento de experiências, de criação de habilidades e modelos mentais comunitários (NONAKA; TAKEUCHI, 1997). O mapeamento da rede de comunicação permite a identificação de especialistas em assuntos de interesse que contribuem com freqüência na distribuição do saber, que fazem a diferença na corrente de colaboração, criando desta forma oportunidades para maior utilização destes indivíduos.

Muitos executivos e pesquisadores voltaram seus focos para esse tipo de instrumento como um elemento facilitador da socialização, de práticas comunitárias, de troca de conhecimento e informações, permitindo que pessoas intra e inter-organizações possam compartilhar interesses comuns.

Os mapas como instrumentos conceituais e técnicos permitirão redefinir os grupos humanos como comunidades de conhecimento e aprendizagem mútua, ou sujeitos coletivos do saber (LÉVY; AUTHIER, 1995).

2.4 Considerações Finais

Este capítulo discutiu a Gestão de Conhecimento, o papel da socialização na construção do capital social da organização e a importância do mapeamento de suas redes sociais para o entendimento de seus processos de socialização. No próximo capítulo serão apresentados: a conceituação teórica sobre Análise de Redes Sociais, suas principais abordagens, como estas técnicas podem ser aplicadas em organizações e algumas ferramentas existentes.

CAPÍTULO 3 - ANÁLISE DE REDES SOCIAIS

O assunto Análise de Redes Sociais é interdisciplinar, ele engloba diversos campos do conhecimento acadêmico como: Sociologia, Psicologia Social, Antropologia, Estudos de Organização, Estatística, Matemática e Ciência da Computação (BORGATTI, 2002). Ao longo dos anos, diversas pesquisas em diferenciados campos acadêmicos fortaleceram as teorias associadas às redes sociais.

De acordo com Scott (2000), três linhas de pesquisas precederam a atual teoria de redes sociais.

1 - os analistas sociométricos que na década de 1930 trabalharam em pequenos grupos e produziram muitos avanços técnicos utilizando métodos da teoria dos grafos;

2 - os pesquisadores de Harvard que nos anos 1930 exploraram padrões de relações inter-pessoais informais e a formação de subgrupos;

3 - os antropólogos de Manchester que baseados nas duas linhas anteriores investigaram a estrutura das relações comunitárias em sociedades tribais e vilas.

Estas três vertentes foram reunidas novamente em Harvard nas décadas de 1960 e 1970 onde a moderna teoria de análise de redes sociais foi desenvolvida.

Foi a partir dos trabalhos de um grupo de pesquisadores da Universidade de Manchester que a teoria de análise de rede social tomou impulso. Deste grupo participava Clyde Mitchell que desenvolvendo idéias já lançadas evoluiu a teoria de análise de rede social em 1969, permitindo sistematizações posteriores (SILVA, 2003).

A concepção de Mitchell em 1969 definiu o conceito de esfera da análise social. A esfera da análise social para ele se define como os padrões de interação que os indivíduos têm com um conjunto de pessoas e as ligações que estas pessoas têm entre elas. Estas redes inter-pessoais são construídas a partir de dois tipos de ações: a ação de comunicação e a ação transacional. A comunicação envolve a transferência de informações entre indivíduos, o estabelecimento de normas sociais e a criação de certo grau de consenso. Já a ação transacional envolve a transferência de bens e serviços entre pessoas. Mitchell voltou-se para a Matemática, mais precisamente para a teoria dos grafos e concentrou-se nas características da organização informal e inter-pessoal (SCOTT, 2000).

Além da teoria dos grafos, a análise de redes sociais foi fortemente influenciada por mais uma teoria matemática: a álgebra matricial. O uso de matrizes para denotar relações em grupos sociais cresceu a partir do ano 1940. As operações matriciais são amplamente usadas para definições e cálculos em análise de redes sociais, sendo a representação primária para a maioria dos softwares de análise de redes sociais (SILVA, 2003).

3.1 Conceitos de Redes Sociais

As definições conceituais contidas nesta seção foram extraídas de Wasserman e Faust (1999) e Hanneman (2001) e serão utilizadas ao longo deste trabalho.

3.1.1 Ator, ligação, relação e conteúdo transacional

Os conceitos fundamentais em uma rede social são: atores e ligações. Um *ator* pode ser uma pessoa, um grupo, uma empresa, ou seja, qualquer unidade social, enquanto que uma *ligação* é a conexão entre dois atores. Ao conjunto de todos os atores entre os quais serão medidas as ligações chama-se o *grupo*, e quando atores e ligações são parte ou subconjunto do grupo são denotados de *subgrupo*. A delimitação do grupo, ou seja, a escolha da amostra da população para o estudo é um processo básico e necessário para a viabilização da análise.

Ao conjunto de ligações de um tipo específico entre os membros de um grupo chama-se *relação*. Por exemplo, numa rede social organizacional podem existir entre os atores ligações de amizade e ligações profissionais, caracterizando-se assim duas relações distintas.

Em redes sociais, uma importante característica é que existem diversas razões que contribuem para a composição das ligações. As ligações se originam das várias atividades que cada indivíduo participa, ou seja, os diversos papéis que o indivíduo assume como participante de um grupo. O conceito sociológico de papel corresponde às normas e

expectativas que se aplicam ao ocupante de uma determinada posição (BOISSEVAN, 1974 apud SILVA, 2003). As pessoas assumem papéis específicos: funcionário, gerente, especialista e amigo. Através de cada papel o indivíduo entra em contato com outros para compartilhar assuntos e atividades de interesse. A estratificação dos papéis desempenhados pelos indivíduos ajuda a categorizar o tipo de relação. No caso do exemplo citado acima, os papéis de funcionário, gerente e especialista refere-se às relações profissionais e o papel de amigo à relação de amizade.

Através de uma relação específica existem trocas de *conteúdo transacional*. Define-se conteúdo transacional como os elementos materiais e não-materiais trocados entre dois atores em uma situação ou relação particular (BOISSEVAN, 1974 apud SILVA, 2003). Os conteúdos transacionais trocados depende não só do papel, mas também da forma com que cada ator específico desempenha este papel. Assim sendo nem todos os atores trocam conteúdos transacionais na mesma quantidade e intensidade, vai depender do desempenho de cada um. O conteúdo transacional serve para mediar a qualidade e intensidade da relação. Ele demonstra qual o investimento ou a importância que os atores estão atribuindo a uma relação (SILVA, 2003).

Resumindo, uma rede social é um grupo de atores (também conhecido na teoria dos grafos como nós) conectados por ligações (arestas), ver Figura 4. Os atores podem possuir vários atributos chamados de variáveis de composição que servirão para enriquecer semanticamente o ator, definir suas principais características. Já as ligações possuem pelo menos uma variável, conhecida como variável estrutural que caracteriza o tipo de relação estudada. Os conceitos e as motivações teóricas do estudo de uma rede específica, usualmente determinam que variáveis medir e que técnicas são apropriadas para suas medidas.

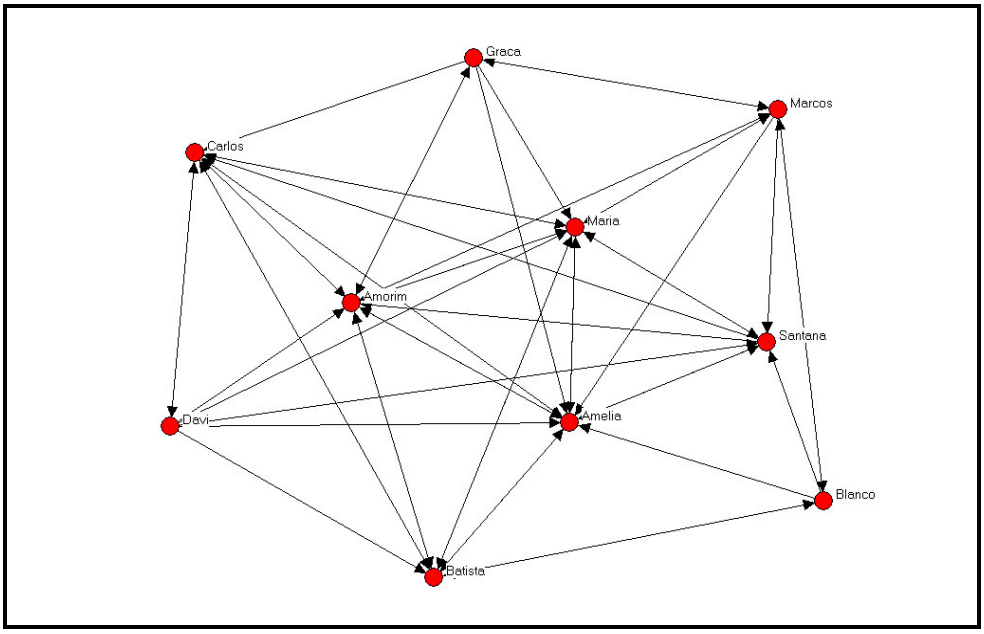

Figura 4. Representação de uma rede social com grafos gerada pela ferramenta *UCINET 6* (BORGATTI;EVERRET;FREEMAN, 2002).

3.1.2 Modalidades e níveis de análise

Em algumas redes é possível se estabelecer vários níveis de análise nos atores, ou seja, estabelecer duas ou mais entidades que os atores pertencem e podem ser analisadas. Normalmente estas entidades são agrupamentos dos atores em estruturas hierárquicas ou aninhadas, como funcionários são agrupados em setores, departamentos, diretorias e unidades. Após o agrupamento dos atores, as ligações também são agrupadas para mostrar como acontecem as interações entre os grupos ou entidades. Uma rede social pode possuir vários níveis de análise a depender da população que está sendo estudada, pois os atores podem estar envolvidos com várias entidades ao mesmo tempo. Na literatura quando este evento acontece diz-se que a rede tem múltiplos níveis de análise e os dados têm múltiplas modalidades. A proposta de ferramenta apresentada neste trabalho se foca apenas um nível de análise onde os atores são os colaboradores de uma organização, não foi implementado o agrupamento de atores.

3.1.3 Múltiplas relações

Atores em redes sociais podem ser caracterizados por vários atributos que são conhecidos na teoria como variáveis de composição. Já as ligações são medidas por variáveis estruturais que são relações de um tipo específico entre um par de atores. Para ligações é comum descrever apenas um tipo de relação. Esta abordagem se contrapõe à situação real em

que um ator normalmente apresenta múltiplos relacionamentos e sua posição em um relacionamento pode reforçar ou contradizer sua posição em outro. Embora seja muito interessante a utilização de vários atributos nas ligações, esta área é relativamente inexplorada em redes sociais e a maioria das ferramentas utilizam apenas um tipo de relação para análise. A proposta de ferramenta deste trabalho apresenta uma abordagem que permite a análise de múltiplas relações através da classificação ou categorização de seus conteúdos transacionais. Esta á uma importante contribuição do nosso trabalho.

3.1.4 Direção e peso das relações

Cada ligação entre os atores pode ser direcionada, isto é, indicar a origem e o destino entre um par de atores. Quando a ligação apresenta as duas direções então acontece a reciprocidade na ligação. Em redes onde a ligação se dá através de troca de fluxo de informações, a dupla direção caracteriza a conversação. A importância de uma ligação poderá ser avaliada pelo seu peso ou intensidade, onde medidas como quantidade e a freqüência de fluxo de informação são relevantes para a análise.

Neste trabalho, como são tratadas as trocas de fluxos de informações na comunicação entre os atores, as métricas adotadas referem-se a redes direcionais.

3.1.5 Redes egocêntricas

Redes egocêntricas ou locais são redes que focam no indivíduo (ego), fazem parte desta rede o ator em análise, os atores que se relacionam diretamente com ele e medidas de ligação entre o ego, seus vizinhos e entre eles. Esta modalidade de análise busca capturar a influência do grupo sobre o indivíduo, que posição cada ator ocupa na estrutura social local e também entender as oportunidades e restrições que cada indivíduo possui como resultado da forma em que ele está inserido na rede. A abordagem centrada no ego não traz informações sobre toda a rede, são dados de micro redes, amostras de áreas locais que fazem parte da rede como um todo. Ver Figura 5.

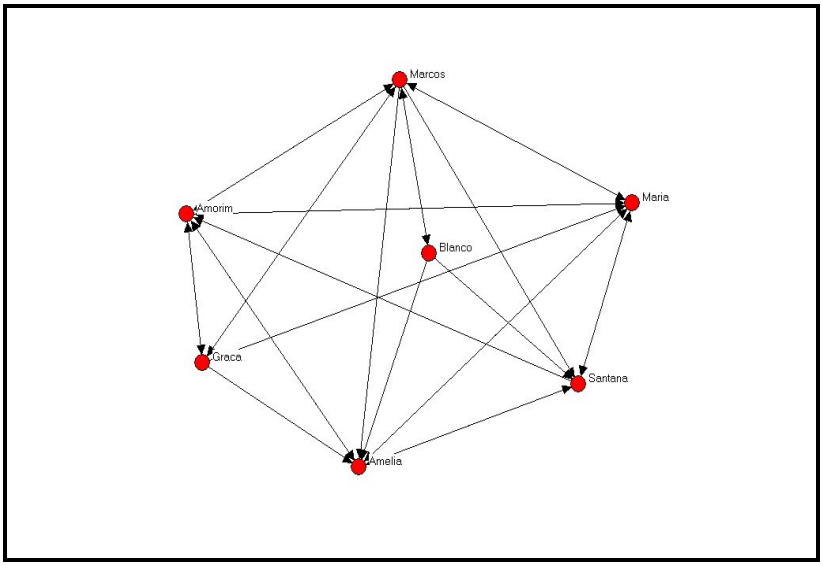

Figura 5. Rede egocêntrica de Marcos utilizando ferramenta *UCINET 6.*

Como dito anteriormente, os métodos matemáticos formais utilizados para representar as redes sociais, são os grafos (teoria dos grafos) e as matrizes (álgebra matricial). Na área da sociologia os grafos são também conhecidos como *sociogramas.* Estas duas técnicas, grafos e matrizes, foram bem aceitas na área, pois apresentam e sumarizam informações de forma fácil e rápida, forçam uma abordagem sistêmica e completa, permitem o uso da computação, têm regras e convenções que promovem uma clara comunicação dos resultados. Na próxima seção apresentamos conceitos e algoritmos básicos utilizados na teoria dos grafos.

3.2 Teoria dos grafos

As definições que constam desta seção ajudarão no entendimento da seção seguinte onde serão abordadas as métricas para análise de redes sociais. Os conceitos utilizados foram extraídos de Ziviane (2004), Chachra, Ghare e Moore (1979), Weisstein (1999) e Mariani (2005). A principal finalidade desta seção é introduzir as teorias, conceitos e formulações matemáticas de grafos que são utilizados nas métricas e medidas para análise de redes sociais.

3.2.1 Grafos

Um grafo é constituído de um conjunto de vértices e um conjunto de arestas conectando pares de vértices. Um grafo é representado matematicamente por $G=(V,A)$, onde V é o conjunto não vazio de vértices, nós ou nodos do grafo e A é o conjuntos de arestas, arcos ou ligações do grafo que representam pares $A=(v,w)$, onde v e $w \in V$.

Um grafo ***G(V,A)*** dado por: ***V = { p | p é uma pessoa }*** e ***A = { (v,w) | v é amigo de w }*** esta representado na Figura 6 onde V={Sara, José, João, Ana) e A = { (Sara, José) , (Sara, João) , (José, Ana) , (João, José) } (MARIANI, 2005).

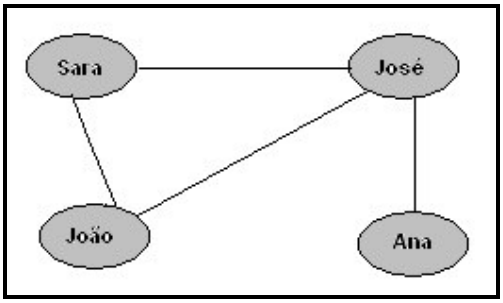

Figura 6. Representação de grafo não direcionado.

Neste exemplo considera-se que a relação: ***v é amigo de w*** é uma relação simétrica, ou seja, se ***v é amigo de w*** então ***w é amigo de v.*** Como conseqüência, as arestas que ligam os vértices não possuem qualquer orientação.

Um grafo onde as relações podem ser assimétricas é dito grafo orientado, direcional ou dígrafo. Num grafo definido por ***V = { p | p é uma pessoa da família Santos }*** e ***A = { (v,w) | v é pai/mãe de w }***, cujo exemplo é V = { Sara, Mario, Maria, Pedro, João, José} e A = {(Mario, Sara), (Maria, Sara), (Maria, João), (José, João), (João, Pedro)}, ver Figura 7. Neste exemplo existe claramente uma orientação na relação, pois se ***v é pai/mãe de w***, o mesmo não pode-se afirmar do contrário, logo, não existe correspondência na relação (MARIANI, 2005).

Figura 7. Representação de grafo direcionado.

3.2.2 Adjacência ou vizinhança

Um conceito muito utilizado em grafos é adjacência ou vizinhança. Dois vértices ***v*** e ***w*** são ditos adjacentes ou vizinhos, se existir uma aresta ***a=(v,w)*** em ***G***. Num grafo simples como na Figura 6 esta aresta é dita ser incidente a ambos, ***v*** e ***w***. Neste exemplo Sara incide em João e João em Sara. No caso do grafo ser dirigido, a adjacência ou vizinhança é especializada em: sucessor e antecessor. O vértice ***w*** é sucessor de ***v*** se existir uma aresta que parte de ***v*** e chega em ***w***. No exemplo da Figura 7, Sara é sucessor de Mario e Maria. Já o vértice ***v*** é antecessor de ***w*** se existir uma aresta que parte de ***v*** e chega em ***w***, então João é antecessor de Pedro.

3.2.3 Grau

O grau de um vértice é dado pelo número de arestas incidentes sobre este vértice. Na Figura 6 o grau de José é três e o de Sara é dois. Em grafos direcionais, o grau se especializa em grau de saída ou de emissão, que corresponde ao número de arestas partindo do vértice, e grau de entrada ou recepção que corresponde ao número de arestas chegando ao vértice. Na Figura 7 o grau de entrada de João é dois e de saída é um. Em grafos direcionais o grau corresponde à soma do grau de entrada com o grau de saída. Se G é o grau, Ge é o grau de entrada e Gs o grau de saída, então: $G = Ge + Gs$.

3.2.4 Subgrafo

Um grafo $Gx(Vx,Ax)$ é dito ser subgrafo de um grafo $G(V,A)$ quando $Vx \subseteq V$ e $Ax \subseteq A$. Na Figura 8 observa-se um grafo na esquerda e um subgrafo na direita.

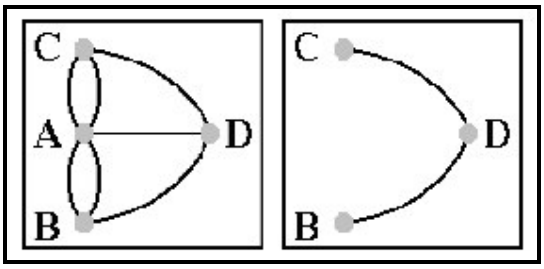

Figura 8. Grafo e um subgrafo.

3.2.5 Cadeia

Uma cadeia é uma seqüência qualquer de arestas adjacentes que ligam dois vértices. O conceito de cadeia vale também para grafos orientados, bastando que se ignore o sentido da orientação dos arcos. O tamanho ou comprimento de uma cadeia é o número de arestas (arcos) que a compõe.

3.2.6 Caminho

Um caminho é uma cadeia na qual todas as arestas possuem a mesma orientação. Aplica-se, portanto, somente a grafos orientados. Na Figura 7 seqüência de vértices (Maria, João, Pedro) é um exemplo de caminho. O tamanho de um caminho é igual ao número de arestas percorridas durante este caminho, o que também é igual ao número de vértices percorridos menos um. O tamanho do caminho Maria, João, Pedro é dois.

3.2.7 Grafo conexo

Um grafo é dito conexo se existir pelo menos uma cadeia ligando cada par de vértices. O grafo da Figura 9 é um exemplo de um grafo conexo direcional, neste caso se desconsidera a orientação.

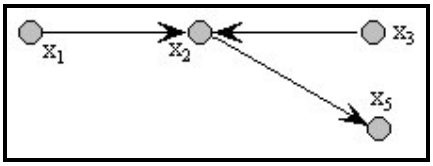

Figura 9. Grafo direcional conexo.

3.2.8 Grafo desconexo

Um grafo é considerado desconexo se existir pelo menos um par de vértices que não está ligado por nenhuma cadeia. Na Figura 10 um exemplo de um grafo não direcional desconexo.

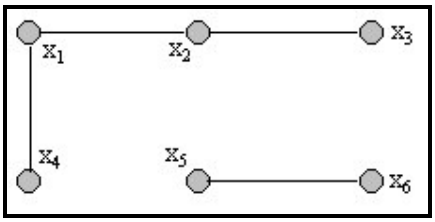

Figura 10. Grafo não direcional desconexo.

3.2.9 Componente conexo

Um grafo desconexo é formado por pelo menos dois subgrafos conexos, disjuntos em relação aos vértices e máximos em relação à inclusão. Ver Figura 11. Cada um destes subgrafos conexos é chamado de um componente conexo do grafo. Um subgrafo conexo é o mesmo que um componente conexo (MARIANI, 2005).

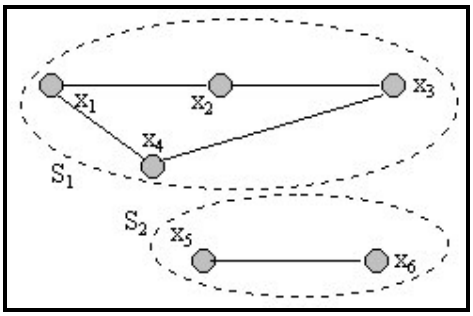

Figura 11. Grafo formado de dois componentes conexos S1 e S2.

3.2.10 Grafo fortemente conexo

No caso de grafos orientados, um grafo é chamado de fortemente conexo se todo par de vértices está ligado por pelo menos um caminho em cada sentido. Isto significa que cada vértice pode ser alcançável partindo-se de qualquer outro vértice do grafo. Na Figura 12 vemos o exemplo de um grafo fortemente conexo, todos os vértices são alcançáveis de qualquer outro (MARIANI, 2005).

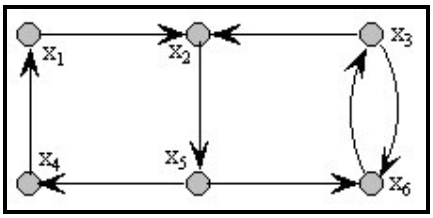

Figura 12. Grafo fortemente conexo.

3.2.11 Componentes fortemente conexos

Um grafo que não é fortemente conexo é formado por pelo menos dois subgrafos fortemente conexos, disjuntos em relação aos vértices e máximos em relação à inclusão. Cada um destes subgrafos é dito ser um componente fortemente conexo do grafo (MARIANI, 2005).

Os componentes fortemente conectados ou conexos de um grafo direcional são compostos de vértices mutuamente alcançáveis (ZIVIANI, 2004). Na Figura 13 (a) podemos observar três componentes fortemente conectados: {Sara, José, João, Ana}, {Rosa} e {Pedro}. Todos os pares em {Sara, José, João, Ana} são mutuamente alcançáveis, já os vértices {4,5} não são fortemente conectados porque Pedro não alcança Rosa. No grafo da Figura 13 (b) existem dois componentes fortemente conectados.

Figura 13. Componentes fortemente conexos.

3.2.12 Pontos de articulação e ponte

Um ponto de articulação ou pontos de corte é um vértice cuja remoção (juntamente com as arestas a ele conectadas) provoca uma redução na conectividade do grafo. Em geral um ponto de articulação quando removido aumenta o número de componentes. Uma ponte é uma aresta cuja remoção provoca uma redução na conectividade do grafo (MARIANI, 2005; WEISSTEIN, 1999).

Na Figura 14 o vértice D é um ponto de articulação, se removido com suas arestas provoca uma desconexão no grafo. A aresta que liga o vértice D e E é uma ponte, se removida também provoca uma desconexão no grafo.

Figura 14. Ponto de articulação e ponte.

3.2.13 Representações computacionais

Segundo Ziviane (2004) algoritmos de manipulação de grafos são de fundamental importância na ciência da computação, várias aplicações necessitam considerar conjunto de conexões entre pares de objetos e os relacionamentos derivados dessas conexões para responder diversas questões. Dentre as aplicações citadas estão: modelagem da *Web* como um grafo, onde os objetos são os documentos e seus *links* as arestas para permitir busca na *Web;* mapas de pessoas e instituições, onde suas aplicações são as conexões com objetivo de descobrir pessoas interessadas e posições disponíveis; caminhos mais curtos em visitas de turistas a cidades de uma região.

A viabilização de algoritmos de manipulação de grafos em programas de computadores inclui a criação de uma estrutura de dados definida pelos tipos de dados que representarão os grafos e os algoritmos que executarão as operações definidas sobre esta estrutura. As representações internas de estruturas de dados mais usadas em grafos são: *matrizes de adjacências* e *listas de adjacências*.

A representação de um grafo $G=(V,A)$ com N vértices como uma matriz de adjacências corresponde a um arranjo bidimensional de $(N \times N)$ com as seguintes propriedades: $A(i,j) = 1$ se existir uma aresta (Vi,vj) e $A(i, j) = 0$ caso contrário.

A maior desvantagem da utilização de matrizes para representar grafos é que a matriz necessita de N^2 de espaço. Por este motivo recomenda-se sua utilização em grafos densos, grafos que possuem muita conectividade.

Uma solução para o problema de espaço é a adoção de listas de adjacências. Neste tipo de estrutura as linhas da matriz são representadas por listas ligadas e um vetor Vi é utilizado como cabeçalho para estas listas. A representação de um grafo $G(V,A)$ por listas de adjacências consiste de um arranjo de V listas, uma para cada vértice em V. Cada lista fornece os vértices para os quais Vi é adjacente ou vizinho (ZIVIANI, 2004). Uma lista de adjacência representando um grafo direcional de N vértices necessita de um espaço de $(N + E)$.

Na Figura 15 o grafo apresentado na letra (a) está representado como lista de adjacência na letra (b) e como matriz de adjacência na letra (c).

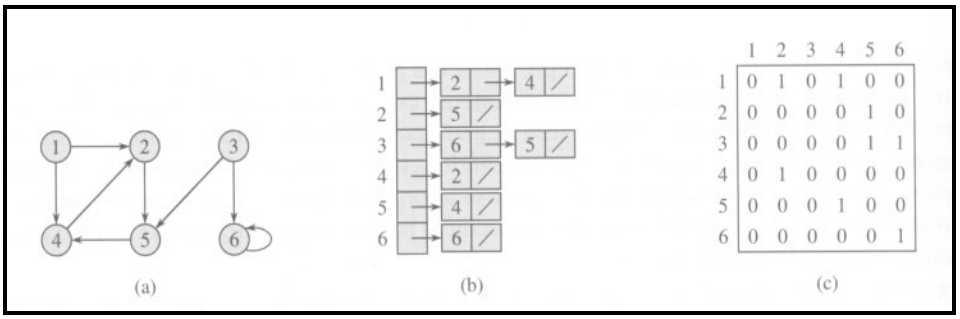

Figura 15. Grafo (a) representado por lista de adjacência (b) e matriz de adjacência (c).

3.2.14 Operações sobre grafos

Uma das operações mais comuns em grafos é caminhar sobre o grafo. O percurso inicia a partir de um vértice ***Vi*** que pertence ao grafo e deve visitar todos os vértices alcançáveis a partir de ***Vi***. Para realizar esta operação existem dois algoritmos: *busca em profundidade* e *busca em largura* (ZIVIANI, 2004).

Outra operação importante é a descoberta do caminho mais curto em grafos direcionais. Esta operação determina para cada vértice ***Vi*** de um grafo direcional ***G(V,A)*** qual o caminho mais curto de ***Vi*** para todos os vértices de ***G***. A busca do caminho mais curto em grafos é aplicada em diversas situações, por este motivo foram propostos um número diverso de algoritmos dentre os quais os mais famosos são os algoritmos de Dijkstra e o de Floyd (ZIVIANI, 2004). Neste trabalho optamos pela utilização do algoritmo de Dijkstra (DIJKSTRA, 1959), amplamente referenciado na literatura sobre grafos e conhecido com um algoritmo de código simples com desempenho $O(N^2)$ onde ***N*** é o número de vértices do grafo. O pseudo-código deste algoritmo implementado na ferramenta encontra-se ANEXO A.

3.3 Métricas em Redes Sociais

Como dito anteriormente, os métodos matemáticos formais utilizados para representar as redes sociais, são os grafos (teoria dos grafos) e as matrizes (álgebra matricial). Na área da sociologia os grafos são também conhecidos como *sociogramas.*

Estas duas técnicas, grafos e matrizes, foram bem aceitas na área, pois apresentam e sumarizam informações de forma fácil e rápida, forçam uma abordagem sistêmica e completa, permitem o uso da computação, têm regras e convenções que promovem uma clara

comunicação dos resultados (SILVA, 2003). Nesta seção serão explorados tópicos importantes das métricas de redes sociais.

A Análise de Redes Sociais se interessa nas relações entre os atores, a utilização e interpretação de suas métricas vão depender da comunidade que está sendo estudada. As informações constantes nesta seção foram baseadas nas seguintes referências: Wasserman e Faust (1999), Hanneman (2001), Borgatti (2002) e Silva (2003).

Existem três visões possíveis para a análise das características estruturais de uma rede social. A primeira examina a rede como um todo, métricas avaliam como a rede é composta e qual o grau de interação entre seus componentes. A segunda visão foca cada ator e seu papel na estrutura da rede, utilizando as medidas de centralidade. A terceira que é uma variante da primeira, analisa a rede de um ator específico, rede egocêntrica. Nela analisa-se o ator em foco, as ligações com seus vizinhos e entre eles, buscando entender a influência do grupo sobre o indivíduo.

3.3.1 Redes como um todo

Para a primeira abordagem estrutural, análise da rede como um todo, são utilizados critérios vindos da teoria dos grafos como tamanho, densidade, distância média entre os atores, diâmetro, alcançabilidade e reciprocidade da rede social estudada.

3.3.1.1 Critérios fundamentais

O tamanho da rede é um fator importante para as possibilidades de ligações, quanto maior a rede mais difícil fica conhecer os outros e trocar informações, isto limita a capacidade que cada participante tem de construir e manter os relacionamentos. Quando uma rede se torna muito grande aumenta também a possibilidade de segmentação ou particionamento em subgrupos. O ***tamanho*** de uma rede é a soma de todos os seus atores de uma rede direcional. Se N é o tamanho da rede, $N * (N-1)$ é o número de possíveis ligações entre os atores. O número de ligações possíveis cresce na ordem N^2 enquanto o número de atores cresce linearmente.

A ***densidade*** de uma rede direcionada é a relação entre o número de ligações existentes E e o número de ligações possíveis $N*(N-1)$. Logo, a densidade de uma rede R direcionada pode ser representada por:

Quadro 1. Fórmula da densidade de uma rede direcional

A densidade é um índice do potencial de comunicação, uma medida que demonstra o quanto ainda a rede poderá crescer em conexões.

Outra propriedade importante das redes é ***distância*** entre os atores. Distância entre dois atores é definida como a menor soma dos pesos de uma seqüência de ligações na mesma direção entre um ator de origem e um de destino. Nele se aplica o cálculo do menor caminho da teoria dos grafos, para isso os algoritmos calculam a matriz de distância. Nesta matriz existem os menores caminhos de cada vértice para todos os outros. Ver Figura 16.

Um ***caminho*** é a seqüência de atores ligados numa mesma direção que começam e terminam em dois atores desejados. O ***menor caminho*** entre dois atores, também conhecido como ***caminho mais curto***, é a menor soma dos pesos das ligações entre atores consecutivos encontrados entre dois atores. Ele identifica as ligações de menor custo entre dois atores. Nas redes de comunicação abordadas neste trabalho não se aplicam pesos nas ligações para o cálculo da distância, logo cada ligação terá peso um na sua contabilização. Resumindo o termo distância se refere ao menor caminho entre dois atores. Sendo ***v*** e ***t*** dois atores de uma rede, ***d(v,t)*** é a distância ou menor caminho entre ***v*** e ***t***.

Caminhos mais curtos são conexões mais eficientes entre atores. As medidas de caminhos fornecem informações importantes para analisar as distâncias entre os atores. Distâncias podem dar idéia do poder das ligações. Atores conectados com pequenas distâncias ou através de vários caminhos são fortemente conectados, suas conexões estão menos sujeitas a descontinuidade sendo mais estáveis e confiáveis. Redes com pequena distância média são consideradas densas, isto sugere que a informação pode viajar mais rapidamente através da rede. Redes com distâncias médias grandes entre seus atores levam mais tempo para difundir a informação entre a população. A distância também pode ser um fator de estratificação e diferenciação. Atores mais próximos estão habilitados a exercer mais poder do que os que estão mais distantes.

Na Figura 16 podemos observar que o ator A possui dois caminhos para chegar a D. Um direto e outro através de B. O menor caminho é o direto, pois apenas uma ligação é contabilizada como peso. Este caminho é o escolhido pelo algoritmo para preencher a

respectiva célula da matriz de distâncias, já o ator C só será alcançado por A através de B contabilizando o valor dois para no seu menor caminho.

	A	**B**	**C**	**D**
A	0	1	2	1
B	2	0	1	1
C	∞	∞	0	∞
D	1	2	3	0

Figura 16. Grafo e sua matriz de distâncias.

O ***diâmetro*** é a maior distância entre dois atores quaisquer da rede e diz quão grande é uma rede. Para determinar o diâmetro, seleciona-se o maior valor de cada linha da matriz de distância. O maior de todos estes valores é o diâmetro. Então, o diâmetro é o maior número de atores que deve ser transpassado para percorrer de um ator a qualquer outro na rede, na mesma direção. Sendo ***v*** e ***t*** dois atores quaisquer, $\boldsymbol{max_{v,t}\, d(v,t)}$ é a máxima distância entre ***v*** e ***t*** e o diâmetro será a maior distância entre todos os atores (WEISSTEIN, 1999). Na Figura 17 o diâmetro do grafo é quatro, pois esta é a máxima distância entre todos os vértices, no caso a distância de *A* para *F* e de *F* para *A*. Hanneman (2001) caracteriza uma rede social como compacta, se esta tem um diâmetro menor ou igual a três.

	A	**B**	**C**	**D**	**E**	**F**	$\boldsymbol{max_{v,t}}$
A	0	1	2	3	3	4	4
B	1	0	1	2	2	3	3
C	2	1	0	1	1	2	2
D	3	2	1	0	2	3	3
E	3	2	1	2	0	1	3
F	4	3	2	3	1	0	4

Figura 17. Grafo e matriz de distâncias adaptado de Chachra (1997).

A ***alcançabilidade*** mede uma seqüência de ligações que permite traçar um caminho da de um ator fonte a outro ator destino. Se algum ator na rede não consegue alcançar outros, isto indica uma divisão potencial da rede ou indica que a população estudada é composta de mais de um subgrupo. Na Figura 16 o ator *C* não consegue alcançar nenhum dos outros.

Alguns atores são predominantemente fontes de informações, outros são receptores e outros podem ser ambos, transmissores e receptores. Atores recíprocos são aqueles que se comunicam nas duas direções, caracterizando uma conversação. A ***reciprocidade*** evidencia o grau de conversação estabelecido numa rede de comunicação. Ela pode ser medida por ator ou para toda a rede. Por ator será verificada a simetria, o balanceamento dos relacionamentos dos atores em relação a sua vizinhança, evidenciando que percentual da comunicação entre os pares está ocorrendo de forma recíproca. Um percentual alto de atores simétricos em uma rede sugere uma estrutura de comunicação não hierárquica.

Além dos critérios descritos acima, outra proposta de análise da abordagem estrutural centra-se na coesão da rede social que busca identificar os subgrupos coesos existentes na rede.

3.3.1.2 Critérios de coesão

Existem três formas de analisar a coesão de um grupo ou subgrupo. A primeira está baseada na reciprocidade que pode ser medida através de *cliques*. Um ***clique*** é um subgrafo com três ou mais atores, onde todos os seus atores têm ligações nas duas direções e não existe nenhum outro ator que seja vizinho a todos os membros do *clique*. Na Figura 18 mostra um grafo e a lista de *cliques* encontrados (WASERMAN; FAUST, 1999).

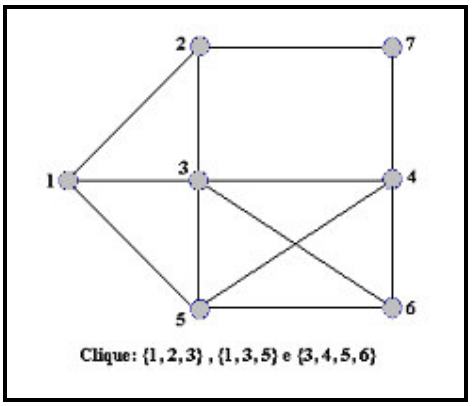

Figura 18. Um grafo e seus *cliques* adaptado Waserman e Faust (1999).

Grupos isolados e fortemente conectados formam um *clique*. Dentro de grupos altamente coesos, os indivíduos tendem a ter crenças homogêneas. (COLLINS, 1988 apud WASSERMAN; FAUST, 1999)

A segunda forma leva em consideração a alcançabilidade que é medida através das menores distâncias e diâmetro da rede, e pode ser medida pelas métricas *n-clique* e *n-clã*.

O ***n-clique*** estabelece um valor ***n*** como a máxima menor distância permitida entre dois atores dentro de um grupo, para que o grupo seja considerado coeso. Um *n-clique* permite ***n-1*** intermediários entre dois atores, o que amplia o conceito de *clique* que só permite atores vizinhos e recíprocos.

O ***n-clã*** é um subgrupo cujo diâmetro não pode ser superior a ***n***, com esta definição procurou-se resolver o problema de um *n-clique* onde intermediários podem não pertencer ao *clique*. Logo, o *n-clã* permite a intermediação, desde que os atores intermediários estejam dentro do grupo.

Na Figura 19 podemos observar que o 2-*clique {1, 2, 3, 4, 5}* tem um diâmetro de três, pois como o nó *6* não pertence ao *2-clique*, dentro dele o menor caminho entre o nó *4* e *5* é o caminho *4-2-3-5* que tem o tamanho três. Entretanto, para calcular o *n-clique* considera-se os caminhos que envolvem os nós não pertencentes ao *n-clique.* O caminho *4-5-6* foi considerado, mesmo o nó *6* não pertencendo ao *n-clique*.

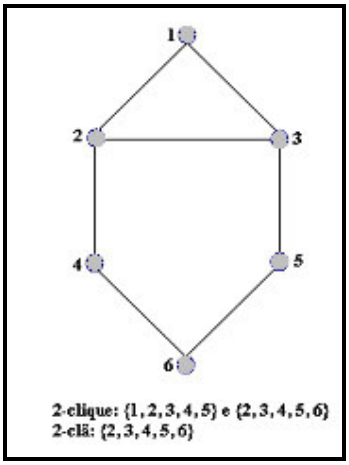

Figura 19. Grafo ilustrando *n-clique* e *n-clã* adaptado Waserman e Faust (1999).

A terceira forma, denominada ***k-plex***, baseia-se em propriedades relacionadas à adjacência entre os atores de um grupo, medida em termos do grau de conectividade de cada ator. Um *k-plex* é um conjunto de atores de tal forma que cada ator é vizinho de quase todos os outros, onde ***k*** é o número de exceções. O conceito de *k-plex* elimina a necessidade da existência de ligações entre todos os atores de um grupo e considera apenas as ligações diretas como indicativas de coesão. Se ***k=1***, *1-plex* equivale a *1-clique,* sendo o subgrafo mais coeso, em que cada membro está conectado a ***n-1*** outros membros. Quando ***k=2***, todos os membros de um *2-plex* estão conectados com pelo menos ***n-2*** dos outros membros, embora *2-plex* pode não ser um *2-clique* (SCOTT, 2000).

Na Figura 20 o grafo (a) é um *3-clique*, todos os pares se conectam a uma distância de três ou menos. Contudo não é um *3-plex*, pois A, C, E e F estão conectados com menos que três outros membros. O grafo (b) é um *3-clique* e um *3-plex*.

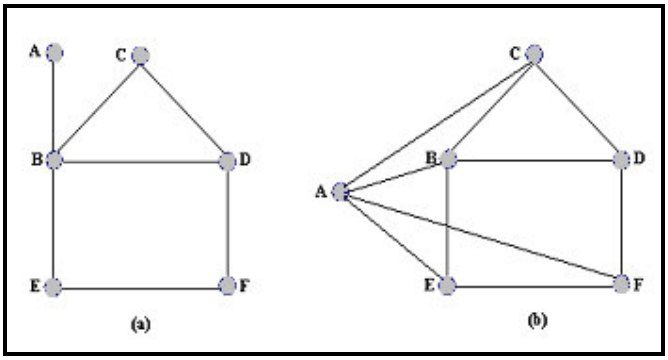

Figura 20. Um *3-clique* e um *3-plex* extraídos de Scoot (2000).

Uma última e mais nova linha de pesquisa na abordagem estrutural de uma rede estuda as similaridades entre as posições que os atores ocupam, medida conhecida como equivalência estrutural.

3.3.1.3 Equivalência Estrutural

Este critério caracteriza os subgrupos através das posições similares que seus atores ocupam na rede. Wasserman e Faust (1999) definem que a noção de posição se refere a um conjunto de atores que são similares em atividades sociais, ligações ou interações. Diferentes autores desenvolveram diferentes medidas para estabelecer os níveis de similaridade estrutural de um conjunto de atores, entre as quais a mais difundida é similaridade baseada na distância euclidiana desenvolvida por Ronald Burt (1987 apud WASSERMAN; FAUST, 1999).

Sendo $\boldsymbol{x_{ik}}$ o valor da ligação entre o ator $\boldsymbol{i}$ e o ator $\boldsymbol{k}$, uma medida para o nível de similaridade entre os atores $\boldsymbol{i}$ e $\boldsymbol{j}$ é definida como a distância euclidiana entre as ligações que saem e partem destes atores. Para os atores $\boldsymbol{i}$ e $\boldsymbol{j}$, esta é a distância entre as linhas $\boldsymbol{i}$ e $\boldsymbol{j}$ e as colunas $\boldsymbol{i}$ e $\boldsymbol{j}$ da matriz de adjacência, sendo a fórmula para tal cálculo dada pela seguinte equação, para $\boldsymbol{i \neq k}$ e $\boldsymbol{j \neq k}$:

$$d_{ij} = \sqrt{\sum_{k=1}^{n} [(x_{ik} - x_{jk})^2 + (x_{ki} - x_{kj})^2]}$$

Quadro 2. Fórmula da distância Euclidiana das ligações que saem de um ator

Se os atores ***i*** e ***j*** forem estruturalmente equivalentes, as entradas em suas respectivas linhas e colunas da matriz de adjacência serão idênticas e a distância euclidiana entre eles será igual a zero. Na medida em que estes dois atores não forem estruturalmente equivalentes, a distância euclidiana entre eles será maior. As distâncias euclidianas são calculadas entre todos os pares de atores numa rede.

3.3.2 O papel de cada ator na rede

A segunda visão da rede olha para cada ator da rede. O objetivo principal é encontrar o papel que cada um desempenha na manutenção e expansão da estrutura da rede. A centralidade é a principal medida para avaliar um ator, nela se mede quão acessível e qual a importância de um ator para os demais atores de uma rede. A centralidade é uma função, não apenas do tamanho da rede, mas também da maneira como os diversos atores estão interligados.

Pesquisas e metodologias foram desenvolvidas para permitir o entendimento da participação dos atores na rede e seu grau de influência e importância. O poder de um ator depende do grau com que ele monopoliza o fluxo de informação, favores e serviços para e entre os membros de uma rede. O número e tamanho das ligações são importantes para entender regras individuais, oportunidades, comportamentos e potencial da rede. As conexões que um indivíduo possui na rede lhe impõem limites e/ou oportunidades. A diferença de conectividade, possibilitando maiores oportunidades, demonstra estratificação entre os grupos. Estas diferenças podem ajudar a entender difusão, homogeneidade, solidariedade e outras propriedades.

O poder é uma propriedade fundamental das redes sociais, ele é uma conseqüência dos padrões de relacionamentos e de posicionamento na rede. Atores com mais importância ou mais proeminentes usualmente estão em posições estratégicas da rede. Atores inseridos em redes sociais de alta densidade e alto acoplamento têm maior potencialidade para exercer poder.

As três medidas de centralidade: grau, proximidade, intermediação que serão apresentadas a seguir são citadas na literatura como os principais indicadores de poder na análise de redes sociais.

A ***centralidade de grau*** mede a quantidade de ligações de um ator. Quanto mais ligações um ator possui maior suas oportunidades de escolha, maior sua autonomia em

relação aos outros e conseqüentemente maior o seu poder. Para redes que levam em consideração a direção das ligações, grau de entrada elevado representa prestígio e popularidade embora possam sofrer sobrecarga, interferências ou ruídos ao receber mensagens de várias fontes, enquanto que alto grau de saída representa colaboração e influência.

A forma como os indivíduos estão conectados afetam seu comportamento. Indivíduos mais conectados influenciam mais e são mais influenciados. Populações mais conectadas têm maior mobilidade de recursos e maiores perspectivas de resolução de seus problemas. A conectividade entre os indivíduos são indicadores de maior possibilidade de colaboração e ao mesmo tempo complexidade social. Atores com poucas ou sem ligações têm menos possibilidade de aprender e obter apoio. O número e tipos de ligações que um ator possui na rede são fatores chave para analisar seu comportamento, oportunidades, influência e poder. Sendo a um ator e c uma conexão ou ligação então, centralidade de grau será (GAERTLER; WAGNER, 2004):

Quadro 3. Fórmula de Centralidade de grau.

Uma variação da centralidade de grau já exposta foi proposta por Phillip Bonacichi. Nesta nova medida de centralidade, ele argumenta que um ator que possui ligações com atores de prestígio elevado não equivale a um ator que tem ligações com atores periféricos, ou seja, os prestígios individuais das ligações indiretas de um ator podem aumentar ou diminuir a medida de seu próprio prestígio. Bonacich criou uma fórmula em que fatores de atenuação e amplificação alteram a centralidade de grau de acordo com os prestígios individuais das ligações indiretas (LAZEGA, 1998 apud SILVA, 2003). A centralidade e o poder de um ator nesta medida são funções das conexões dos seus vizinhos. Quanto mais conexões os vizinhos possuem mais central o ator será e quanto menos conexões os vizinhos possuírem maior será o poder do ator sobre os vizinhos (HANNEMAN, 2001).

A ***centralidade de proximidade*** de um ator se mede pelo número mínimo de passos que ele deve executar para entrar em contato com os outros atores da rede. Quanto mais central for um ator, mais ele está próximo dos outros, e mais rapidamente ele entra em contato ou interage com os outros. É uma medida da autonomia, da independência em relação ao

controle exercido pelos outros. Esta medida leva em consideração o somatório dos menores caminhos de um ator para com todos os outros. Sendo ***A*** o conjunto de todos os atores, e ***d(v,t)*** a distância entre dois atores quaisquer representados por ***v*** e ***t,*** *a* melhor situação de proximidade acontecerá quando o somatório das distâncias de um ator para com todos os outros é igual ao número total de atores menos um, $\sum d(v,t) = n - 1$, isto significa que o ator em foco é adjacente ou vizinho de todos os outros**.** O valor relativo da proximidade é representado por (GAERTLER; WAGNER, 2004):

Quadro 4. Fórmula de Centralidade de proximidade.

A ***centralidade de intermediação*** está baseada no controle exercido por um ator sobre as interações entre dois outros atores. Desde que dois atores não sejam adjacentes ou vizinhos, eles dependem de outros atores do grupo para realizar suas trocas. Uma característica negativa é que caso ocorra fala na intermediação, o fluxo entre os atores não adjacentes será interrompido. A posição de intermediação determina influência através da possibilidade de controle do fluxo e indica uma situação preferencial de um ator, pois a intermediação permite a combinação de diferentes idéias que podem influir para a criação de soluções inovadoras. Atores com alta intermediação possuem uma boa visibilidade do que está acontecendo na rede. O maior valor de intermediação que poderá ser encontrado quando o número de atores de uma rede for ***N*** será $Ci_{máx} = (N-1)*(N-2)$, neste caso o ator exerce intermediação em todos os caminhos da rede. Seja ***A*** um conjunto de atores, ***s***, ***t*** e ***v*** atores distintos pertencentes a ***A***, $\sigma G(s,t|v)$ o número de menores caminhos entre ***s*** e ***t*** que passam pelo ator ***v*** e $\sigma G(s,t)$ o número de todos os menores caminhos entre ***s*** e ***t***, então centralidade de intermediação será (GAERTLER; WAGNER, 2004):

Quadro 5. Fórmula de Centralidade de intermediação.

3.3.3 Rede Egocêntrica

Numa rede egocêntrica serão analisados apenas o ator em foco e sua vizinhança. A vizinhança de um ator é o grupo de atores com que ele interage diretamente. Para analisar uma rede egocêntrica serão utilizadas algumas das medidas estruturais usadas para medir toda a rede como tamanho e densidade. O tamanho da rede ego significa com quantas pessoas o indivíduo se relaciona diretamente, um valor alto nesta medida é positivo, pois quanto mais contatos o indivíduo possui maior será sua chance de obter o recurso de que necessita. Alta densidade em rede ego é um fator negativo, pois se todos os vizinhos são ligados entre si, isto representa uma redundância de relacionamentos.

Neste tipo de rede a heterogeneidade e composição são mais dois fatores importantes de serem observados. A heterogeneidade representa a variedade de vizinhos relacionados a atributos de categorização, quanto mais variado em termos de valores de atributos são os vizinhos, maior será a abrangência de informações na rede. Outro fator importante é a composição da rede ego que determina a qualidade dos vizinhos, ou seja, quanto mais bem posicionados forem os vizinhos em termo de centralidade, melhor será a rede do indivíduo (BORGATTI; JONES; EVERETT, 1998).

3.4 Redes Sociais Organizacionais

A utilização de análise de redes sociais em organizações tem sido objeto de vários trabalhos de técnicos e acadêmicos. Rede social organizacional é uma comunidade profissional onde as pessoas se comunicam através de diversos meios. Os meios de comunicação utilizados por organizações modernas são cada vez mais baseados em tecnologia de informação ou comunicação. Estas redes podem ser analisadas através de métodos que acompanham as comunicações. A análise das redes sociais organizacionais permite visualizar a organização em vários níveis ou camadas, como um todo, em grupos ou um indivíduo. Comunidades de práticas podem ser reveladas através de seus problemas, interesses e clientes comuns. Estas comunidades guardam, compartilham e ampliam competências e os indivíduos aprendem e enriquecem seu conhecimento participando de uma ou mais destas comunidades (KREBS, 1998).

Lesser e Prusak (2000) definem as comunidades de prática como coleções de indivíduos ligados por relacionamentos informais que compartilham regras de trabalhos

similares num contexto comum. Eles salientam que a palavra comunidade não denota posição geográfica, unidade de negócio ou fronteiras funcionais, mas sim tarefas, contextos e interesses comuns de trabalho. A palavra prática implica em conhecimento em ação, como o indivíduo executa suas tarefas básicas do dia-a-dia, são processos dinâmicos aprendidos pelo indivíduo de como fazer seu trabalho. As comunidades de prática diferem das unidades formais das organizações como equipes e grupos de trabalho. Estes são orientados a tarefas e atendem a requisitos formais para serem membros, já as comunidades possuem membros informais, são de natureza auto-organizacional e de alta fluidez.

A análise de redes organizacionais permite identificar as comunidades de prática que influenciam positivamente nas organizações. Em ambientes geograficamente dispersos que utilizam basicamente a comunicação virtual, a identificação de comunidades focando em seus principais elementos facilita o contato direto, aponta os especialistas permitindo a expansão de sua atuação. Muitas organizações estão interessadas em identificar estes indivíduos e dar a eles mais tempo e recursos para criar e compartilhar eficientemente conhecimento com outros na organização.

As redes sociais organizacionais são redes nas quais as ligações se dão através de várias relações. Cada colaborador participa da rede assumindo vários papéis, hora como amigo, hora como especialista técnico, hora como conselheiro político. Krackhardt e Hanson (1993) sugeriram que as análises das redes informais organizacionais devem partir de três tipos principais de relação:

- Relação de aconselhamento – mostra os colaboradores influentes, de quem outros dependem para resolver problemas e receber informações técnicas.

- Relação de confiança – mostra como os membros da comunidade compartilham as delicadas informações políticas.

- Relação de comunicação – revela as pessoas que trocam informações relacionadas ao trabalho.

Eles demonstram em estudos de caso, como mapas de cada um dos tipos de relações podem melhorar o entendimento dos gerentes e proporcionar alternativas para solução de problemas organizacionais através da reestruturação da organização formal com

características positivas encontradas na organização informal, mantendo ambas em sintonia com os objetivos da empresa.

Na categorização das relações sugerida pelos pesquisadores cada colaborador pode assumir três papéis distintos. Através destes papéis serão trocados diversos elementos materiais e não materiais denominados conteúdo transacional. Nas relações categorizadas pelos autores, existem trocas de informações técnica, políticas e de trabalho respectivamente.

Krackhardt e Hanson (1993) sinalizam que muitos executivos falham em processos de reestruturação organizacional porque olham apenas para a organização formal e não dão atenção às redes informais da organização. As redes informais de relacionamentos não são vistas com bons olhos por tais executivos, para eles estas redes possuem um estigma de serem desreguladoras da organização formal, de possibilitarem desvirtuação do caminho traçado, introduzindo pontos negativos nos relacionamentos, além de acreditarem que a análise do comportamento utilizando tal ferramenta seja pouco científica. No artigo os pesquisadores tentam provar que estas crenças são mitos e afirmam que rede informal de colaboração é o centro nervoso da organização que dirige a coletividade através de processos, ações e reações às necessidades do negócio. Os gerentes utilizando simulações através do mapeamento da rede podem antecipar como decisões estratégicas afetarão a organização. Os autores citam como exemplo, a necessidade de se formar um time que precisa remover um funcionário chave de suas atividades do dia a dia. O mapa ajudará a identificar se a remoção do funcionário implicará em isolamento de grupos, o que poderá levar o gerente a reconsiderar sua estratégia.

Este mesmo tema foi abordado novamente por Cross e Prusak (2002), neste artigo eles relatam que depois de analisar mais de 50 redes de grandes organizações, chegaram à conclusão de que o mundo real em muitas empresas é feito de informalidades, através de contatos pessoais e sugerem que nestas redes existem quatro elementos comuns cujos bons desempenhos são críticos para a produtividade de qualquer organização.

O primeiro é o ***conector central***, aquele indivíduo que liga muitas pessoas com outras na rede. Não são líderes formais, mas fornecem informações críticas ou expertise para fazer o trabalho acontecer. Nas experiências relatadas pelos autores os conectores centrais servem as organizações de uma forma positiva, ligando colegas e aumentando a produtividade, em alguns casos eles podem criar gargalos que podem engessar ou sobrecarregar a rede. Um

gerenciamento saudável deve acompanhar o nível de carga destes conectores centrais e tentar minimizá-lo, além de estimular e incentivar a criação de novos conectores centrais através de motivações e recompensas.

O segundo elemento é o ***especialista periférico***, aquele indivíduo que está intencionalmente na borda da rede, e que possui um tipo informação ou conhecimento técnico específico que será passado para qualquer membro do grupo apenas quando necessário. Estes indivíduos podem ser pesquisadores e cientistas que desejam participar de forma periférica da rede.

Na Figura 21 podemos observar a presença destes dois papéis. Alan representa o conector central, pois é a pessoa mais considerada e procurada dentro da rede, apesar de Lisa ser a gerente do departamento. Para Paul pode ser atribuído o papel de especialista periférico, pois está na borda da rede e apesar de não estar conectado com muitos do grupo oferece expertise a alguns membros da rede (CROSS; PRUSAK, 2002).

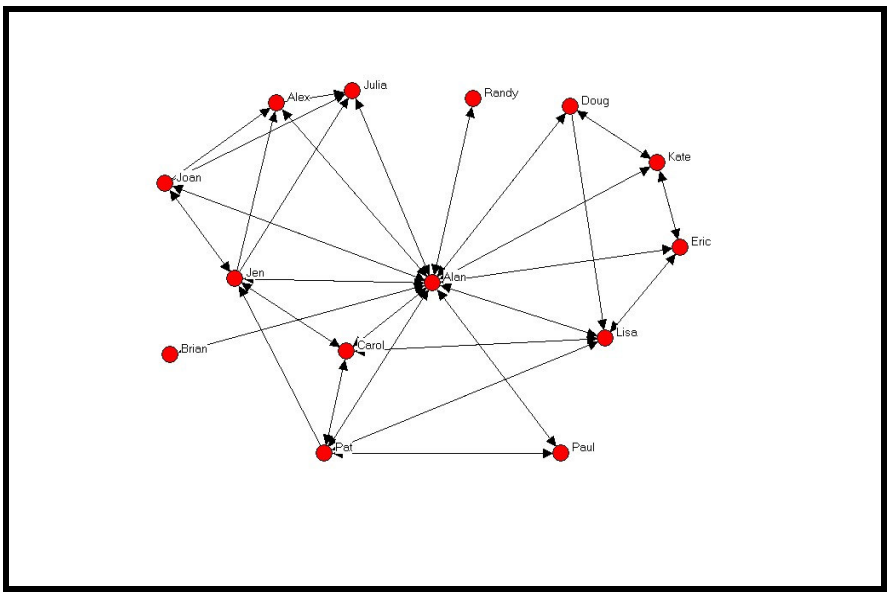

Figura 21. Exemplo de Cross e Prusak (2002) utilizando ferramenta *UCINET 6.*

O terceiro elemento é o ***expansor de fronteiras***, que conecta a rede com outras partes da organização ou com redes de outras organizações. O expansor de fronteiras serve como um condutor de informações, sua posição permite receber e combinar informações e idéias de mais de um subgrupo ou grupo, dando-lhe uma visão mais completa e um controle maior sobre as informações que trafegam.

O quarto e último elemento é o ***corretor de informações.*** Esta figura permite que diferentes subgrupos fiquem interligados, impedindo a segmentação e fragmentação da rede. Papel semelhante ao expansor de fronteira exceto por estar sempre dentro da própria rede.

Na Figura 22 podemos identificar no ator da posição central o papel de corretor de informações, pois ele faz com que dois subgrupos se comuniquem, caso seja removido da rede ela se dividirá em dois subgrupos disjuntos que não terão nenhum canal de comunicação.

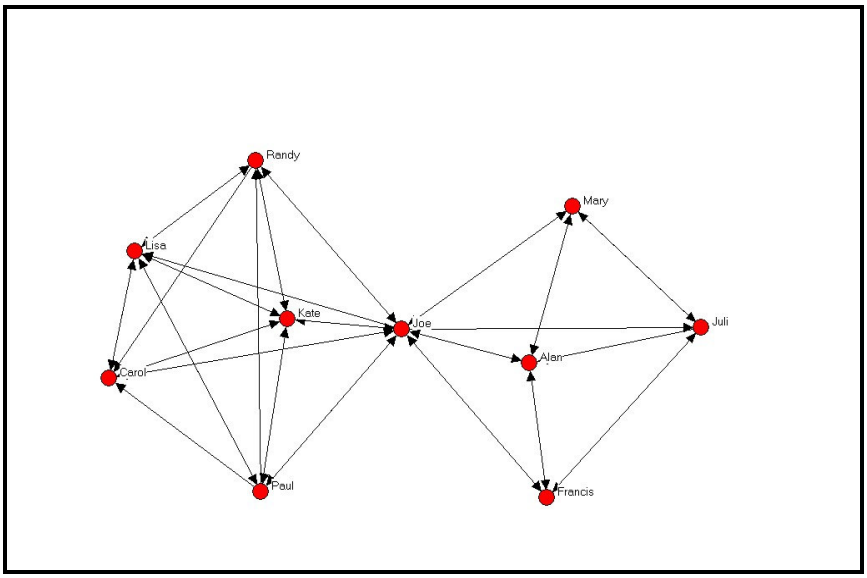

Figura 22. Exemplo adaptado de Cross e Prusak (2002) utilizando ferramenta *UCINET 6.*

A identificação destes indivíduos segundo os pesquisadores nutre o capital social. A abertura e a sistematização das redes informais tornam o trabalho em grupo mais efetivo.

Indivíduos com fortes redes de contatos ficam mais satisfeitos com seu trabalho e permanecem mais nas organizações.

Borgatti, Jones e Everret (1998) divulgaram um trabalho onde identificam que medidas de análise de redes sociais podem ser usadas para formalizar a noção de capital social. Eles consideraram as abordagens de Putman e Burt, citadas na seção 2.2, sobre capital social para separar as unidades de análise. A análise de Putman idealiza o capital social com qualidade do grupo, já Burt concebe o capital social como o valor de relacionamentos sociais individuais.

Quando aplicando as medidas de análise de redes sociais ao indivíduo, a abordagem de Burt pode ser adotada. Para medir os relacionamentos do indivíduo com seus vizinhos podemos utilizar as medidas da rede egocêntrica como: tamanho, densidade, heterogeneidade e composição.

Quando aplicando as medidas de análise de redes sociais a um subgrupo de atores dentro de uma rede, podemos analisar as duas dimensões: relacionamentos dentro do subgrupo ou do subgrupo para fora. A dimensão relacionamentos dentro do subgrupo corresponde a abordagens de Putman e para medi-la utiliza-se as medidas de estrutura da rede: tamanho, densidade, distância media e as medidas de centralidade. Já a dimensão relacionamento do subgrupo para fora, podem ser vistos através da abordagem de Burt se olharmos agora para cada subgrupo como uma entidade única que passa a se relacionar com entidades do mesmo tipo (subgrupo), desta forma as medidas utilizadas poderão ser medidas da rede egocêntrica, onde cada subgrupo será considerado um ator.

Como vimos nos estudos expostos muitos pesquisadores mostraram e comprovaram a eficiência da análise social aplicada a organizações. Resumindo as aplicações desta tecnologia podemos citar:

- Gestão do Conhecimento
- Construção de times
- Gerenciamento de projetos
- Localização de expertise
- Planejamento organizacional

- Difusão de inovações
- Desenvolvimento de líderes
- Descoberta de formadores de opinião
- Descoberta de práticas comunitárias
- Mapeamento e medidas do fluxo de informações

3.5 Metodologia para Análise de Redes Sociais

Para utilizar uma ferramenta de Análise de Redes Sociais será necessário executar alguns processos anteriores que vão garantir a coleta, transformação e disponibilização dos dados em arquivos ou banco de dados. O escopo da proposta deste trabalho não trata os processos iniciais do *SNA*, mas sim os processos após o armazenamento destes dados já transformados em ambiente de banco de dados, ver Figura 23. As técnicas e todos os passos utilizados para obtenção dos dados no estudo de caso serão apresentados no capítulo 6.

Figura 23. Modelo de processos *SNA*

3.5.1 Coleta de Dados

A coleta de dados é um processo básico e importante na análise de redes sociais, embora não seja apresentada uma solução nem uma abordagem obrigatória, as alternativas para este processo serão introduzidas e relatadas, indicando seus pontos positivos e negativos.

A coleta dos dados tem um papel fundamental e crítico na análise de redes sociais, pois neste processo será escolhida a amostra que representará a população ou o público que será estudado. A escolha correta será decisiva para a produção de bons resultados, pois os dados coletados representarão as verdades ou fatos nos quais o produto se baseará para estabelecer a análise através de métricas e apresentar os resultados aos peritos no assunto. O papel dos peritos na verificação e validação dos resultados obtidos pelas ferramentas é a certificação da consistência e sincronismo com a realidade dos fatos.

O processo de coleta dos dados poderá ser feito de duas maneiras, através de procedimentos convencionais ou procedimentos automatizados. A coleta dos dados por procedimentos como: questionários, entrevistas e inspeções são os mais utilizados, estes

instrumentos buscam coletar as informações nas fontes, ou seja, nos indivíduos ou atores envolvidos na rede social estudada. O maior problema relatado com estes procedimentos é seu trabalho intenso que consome um grande tempo, se tornando proibitivo em grandes organizações (TYLER;WILKINSON;HUBERMAM, 2003).

A coleta de dados automatizada é abordada em estudos de caso e em ferramentas de análise de redes sociais mais recentes (TYLER; WILKINSON; HUBERMAM, 2003), (SCHWARTZ; WOOD, 1993), (CAMPBELL e outros, 2003), (GLOOR e outros, 2003), (OHSAWA e outros, 2002), (GUIMERA e outros, 2002) e (CHANG; CHEN; CHUANG, 2002). Nesta abordagem se utilizam bases de dados já existentes nas organizações, para coletar as informações sobre a rede social. Bases de dados de ferramentas que registram a interação de comunicação entre indivíduos servem para obtenção das ligações existentes na rede social. Ferramentas com correio eletrônico, *chats, newsgroups, groupware, workgroup, workflow, HelpDesk,* entre outras, possuem estas características. Estas ferramentas possuem extensas bases de dados, onde são registradas as comunicações entre os indivíduos ou atores de uma rede social. Especificamente ferramentas de correio eletrônico são as principais representantes deste tipo de registro.

Correio eletrônico vem se tornando um meio de comunicação cada vez mais importante na sociedade da informação. Ele possibilita comunicações de negócios, sociais e técnicas e passou a ser tema relevante de pesquisa em comunidades e redes sociais (TYLER; WILKINSON; HUBERMAN, 2003). Considerada a aplicação primária de produtividade, no mundo dos negócios e o mais usado aplicativo de comunicação nos EUA e no Canadá, uma pesquisa revela que 97% dos trabalhadores trocam mensagens eletrônicas diariamente (ROHALL, 2001). A coleta de informações sobre as mensagens que trafegam no ambiente de correio eletrônico fornece uma amostra da comunicação entre os colaboradores muito próxima da realidade. A estes dados poderão ser acrescentadas de forma tradicional as informações sobre o uso dos outros meio de comunicações como: comunicação face-a-face e telefônica.

As informações resultantes da coleta automática trazem um grande benefício que é a confiabilidade e a veracidade das informações, uma vez que é um registro da realidade. Apesar de trazer este grande benefício, introduz um componente preocupante que diz respeito à ética da privacidade.

Analisando a questão ética, uma abordagem proposta para a implantação de uma ferramenta de análise de redes sociais que utilize processos automatizados é a transparência e a autorização. Para implementar e fazer uso adequado deste tipo de abordagem, buscando socialização e colaboração dos indivíduos, a coleta automática deve ser obtida através da conscientização dos participantes da intenção de utilização das informações e garantias do uso destas informações apenas se houver adesão explicita do indivíduo ao processo. Cada participante terá plena consciência e autorizará o uso das informações, não de uma forma imposta mais através de adesão por entendimento dos benefícios.

Após o processo da coleta automática, os dados precisam ser transformados em informações que poderão ser tratadas pelas ferramentas de *SNA*.

3.5.2 Transformação

Os dados extraídos de bases de dados já existentes, normalmente não estão preparados para serem usadas diretamente nas ferramentas de *SNA*. Para isso será necessário um processo de transformação, onde os dados serão filtrados, reduzidos, consolidados e representados nos formatos e fontes de dados aceitas como entrada pela ferramenta utilizada. A maioria destas ferramentas aceita arquivos convencionais de dados em formato *ASCII*.

3.6 Ferramentas de Análise de Redes Sociais

A busca por mecanismos e ferramentas de visualização que apóiem os processos de análise em redes sociais vem de muito tempo, desde os anos de 1930. Na seção 4.3 deste trabalho existe um breve relato do histórico do surgimento destas ferramentas e como a Tecnologia de Informação tem assumido o papel de apoio a estudos nesta área. Os produtos desenvolvidos para este fim utilizam linguagens de desenvolvimento como *Java*, *C++, Pascal* e *Delphi*. A grande maioria aceita arquivos convencionais no formato *ASCII* com entrada de dados.

As ferramentas encontradas e pesquisadas trazem pouca capacidade de associação semântica aos objetos estudados: atores e ligações. Algumas permitem associação fixa de poucos atributos nos atores. Às ligações, associa-se somente o peso ou intensidade. Outro fato observado é a pouca capacidade de exploração visual dos dados nas ferramentas. Em algumas ferramentas os resultados das métricas utilizadas não aparecem na cena visual (grafo), sendo apresentados em tabelas estatísticas em outra janela. A busca de subgrupos ou comunidades

se dá através de métricas de agrupamentos ou *clique,* não atribuindo nenhum significado ou semântica sobre o interesse da comunidade.

Apesar das redes sociais apresentarem relações diversas entre seus atores, a maioria das ferramentas de análise disponível trata apenas uma relação por vez e não menciona o conteúdo transacional trocado.

Nas seções seguintes serão apresentadas algumas destas ferramentas e suas principais características.

3.6.1 UCINET

De todas as ferramentas pesquisadas é a mais conhecida, mais citada academicamente, e a que apresenta uma solução de métricas mais completa, contendo um variado conjunto de medidas. Esta ferramenta é utilizada como referência em trabalhos dos principais pesquisadores e acadêmicos do assunto. É uma ferramenta no ambiente Windows distribuída pela Analytic Technologies, Inc (BORGATTI; EVERETT; FREEMAN, 2002). O programa pode ser carregado livremente por 30 dias e para maior duração possui preços diferenciados para fins educacionais, governamentais e corporativos.

A ferramenta foi desenvolvida com a colaboração de vários especialistas conhecidos na literatura sobre redes sociais: Steven Borgatti, David Krackhardt, Stan Wasserman, Martin Everett , Linton Freeman entre outros. Ela faz uso de rotinas desenvolvidas em Pascal. Os dados são armazenados e tratados através de matrizes. Como entrada de dados aceita vários tipos de arquivos no formato *ASCII*.

O programa utiliza algoritmos rápidos que possibilitam manusear grandes volumes de dados. As métricas disponibilizadas calculam propriedades como: centralidade, identificação de subgrupos, coesão, análise de papéis através da centralidade utilizando, funções da teoria dos grafos e da álgebra matricial.

Para apresentação de mapas em grafos utiliza o *NETDRAW* um programa de livre utilização para visualizar redes sociais, ver Figura 24. O *NETDRAW* manipula múltiplas relações e permite a utilização de atributos apenas nos nós para atribuir cores, formatos e tamanho. Os recursos de interações e visualizações são pouco desenvolvidos, não introduzem

maiores significados aos objetos apresentados. As métricas não são mostradas visualmente nos grafos e são apresentadas aos usuários apenas no formato de tabelas.

Figura 24. Garfo apresentado pela ferramenta *UCINET 6* e alguns recursos de interação.

3.6.2 VISONE

Visual Social Network é uma ferramenta desenvolvida na Universidade de Konstanz e Universisdade de Passau, ambas na Alemanha, tendo como pesquisadores Dorothea Wagner e Ulrink Brands respectivamente. Implementada em *C++* e utilizando a biblioteca *LEDA* (*The Library of Efficient Data Types and Algoritms*) é um software livre para propósitos acadêmicos, roda em ambiente *Linux, Solaris* e *Windows* (BRANDES; WAGNER, 2003).

Utiliza uma classe *GraphWin* para editoração de grafos. A entrada de dados é feita através de uma sublinguagem *XML* denominada *GraphML (Graph Markup Language)*, mas aceita também outros formatos através de importação. Esta linguagem permite a definição de atributos de estilo, largura, cor e intensidade para as arestas e cor, largura, altura, nome e valor para os nós. Permite diversas formas de a apresentação do grafo e possibilita a visualização da importância de cada ator através de regras como aumento de tamanho, mudança de cor ou posição. Possui mecanismo para seleção de atributos especificados nos nós e arestas, que facilitam a exploração permitindo incluir, alterar e esconder os objetos. Não identifica o

conteúdo transacional das ligações. Na atual versão a ferramenta apresenta analise apenas para o nível do indivíduo ou do ator.

3.6.3 NetVis

NetVis é uma ferramenta livre de código aberto baseada na *Web,* em ambiente *Apache/mySQL/php* que também pode executar no *Windows* ou *MAC*, para analisar e visualizar redes sociais (CUMMINGS, 2004). Usa como entrada arquivos *CSV* e calcula métricas de centralidade, medidas de coesão e analisa redes egocêntricas. Possui uma *Applet Java* para visualização da rede através de grafos em 2D e 3D. Desenvolvida por Jonathan Cummings, professor da MIT (*Massachusetts Institute of Technology*), permite interações para incluir ou apagar atores e ligações. Interage com os atributos dos atores utilizando cores. Quando solicitada alguma métrica, representa a diferença de valores entre os atores por mudança de cores ou tamanho. Não permite atributos nem interação com as ligações.

3.6.4 Conversation MAP

Ferramenta desenvolvida por Waren Sack da MIT Media Laboratory que se propõe a analisar textos de mensagens arquivadas por ferramentas *Usenet newsgroups* e gerar saídas gráficas que podem ser utilizadas para pesquisar e ler este arquivamento. O sistema incorpora uma série de procedimentos de análise que detalha quem está falando com quem no grupo, o que estão falando, ou seja, qual o tema das discussões e que temas centrais em discussão estão associados com outros temas. Os procedimentos são escritos linguagem *PERL*, os resultados são armazenados como *HTML* e mostrados em uma *Applet Java*. (SACK, 2000).

Entre as ferramentas observadas, *UCINET* é a mais consistente e a que apresenta métricas e algoritmos mais desenvolvidos. Técnicas e abordagem de diversos autores foram introduzidas ao produto permitindo a escolha da técnica mais adequada ao problema estudado.

3.7 Considerações Finais

Neste capítulo foram apresentados os conceitos e as métricas utilizadas na Análise de Redes Sociais. Discorreu-se sobre como esta teoria está sendo vista e utilizada por pesquisadores no auxílio à gestão empresarial. Foram apresentadas algumas ferramentas *SNA* existentes e suas principais características. No próximo capítulo serão descritas questões relacionadas às técnicas de mineração visual de dados, suas principais abordagens,

mecanismos utilizados e como estas técnicas poderão auxiliar no entendimento e na assimilação dos resultados obtidos em ferramentas *SNA*.

CAPÍTULO 4 - MINERAÇÃO VISUAL DE DADOS

Nas últimas três décadas a tecnologia de banco de dados contribuiu fortemente para o aumento do volume de informações digitalizadas nas organizações. A variedade e a quantidade de estruturas armazenadas atendiam principalmente a sistemas transacionais para administrar e controlar os processos do dia-a-dia. No decorrer deste período novas necessidades de informações para apoio a tomada de decisão ampliaram os objetivos de manipulação destes dados. Bases de dados já existentes nas organizações passaram a ser vistas como repositórios de informações históricas que poderiam executar processamentos analíticos para encontrar associações, categorização e padrões escondidos com a finalidade de predizer ou descobrir tendências de comportamento. Desde então novos mecanismos e técnicas vêm surgindo com a finalidade de adequar e melhorar a capacidade de manipulação destas informações para obtenção dos resultados pretendidos.

Descoberta de conhecimento em bancos de dados, também conhecida como KDD (*Knowledge Discovery in Database*) é uma área de pesquisa voltada para este fim. Considerado como um processo global de identificação de padrões, modelos ou estruturas válidas nos dados que sejam potencialmente úteis e interpretáveis, reuniu vários campos de pesquisas como: banco de dados, aprendizado de máquina, reconhecimento de padrões, estatística, inteligência artificial, sistemas especialistas, visualização de dados e computação de alto desempenho para a obtenção do atual estágio de sua tecnologia (FAYYAD, 1998; FAYYAD; PAITETSKY-SHAPIRO; SMYTH, 1996).

Em muitos textos sobre o assunto confunde-se a definição de KDD com a de Mineração de Dados. Neste trabalho utiliza-se a mesma abordagem referenciada por Fayyad

(1998), onde KDD é apontado como o processo geral de descoberta do conhecimento útil dos dados e Mineração de Dados se refere a um passo particular do processo KDD.

Segundo Fayyad (1998), para realização de um processo de KDD vários passos deverão ser executados: seleção, pré-processamento, mineração de dados e assimilação dos resultados (Figura 25).

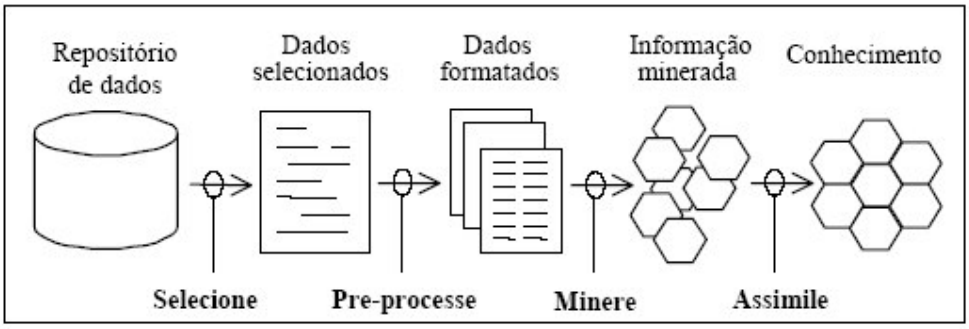

Figura 25. Processo de *KDD* e seus passos (MENDONÇA, 2001).

O primeiro passo conhecido por seleção consta da captura dos dados ou da amostra de dados que será utilizada para a descoberta de conhecimento. Estes dados que normalmente já estão armazenados nas organizações em diversas fontes como: arquivos ou bases de dados deverão ser acessados e extraídos de seu local original para um repositório específico do processo em questão. Um esforço razoável é despendido na execução desta tarefa a fim de conseguir autorização para os acessos, de transportar estes dados dos locais originais e organizá-los no local de destino, possibilitando assim a execução dos passos seguintes do *KDD*.

> Dependendo da complexidade, da semântica, das fontes e do repositório destino dos dados, este procedimento pode variar de uma simples consulta a um banco de dados, a um procedimento complexo de migração, incluindo intrincadas conversões de tipos e formatos de dados (MENDONÇA, 2001, p. 8).

O próximo passo é o *pré-processamento*, ele consiste da transformação dos dados originais, que podem estar em diversos formatos, para o formato desejado pelo passo conhecido como *mineração*. Para pré-processar os dados várias técnicas poderão ser utilizadas: padronização de caracteres, concatenação de vários atributos numa única cadeia caracteres, formatação de representação para converter os dados em formatos aceitos pelo algoritmo de mineração, limpeza de dados a fim de eliminar dados inválidos ou não

interessantes, categorização dos dados para transformar valores numéricos em categóricos representando faixas de valores, conversão de unidades, entre outras.

O passo denominado mineração de dados é encarado como o núcleo do processo *KDD*, sendo responsável pelo mapeamento dos dados para algum tipo de informação ou modelo, representando 15 a 20% do processo de *KDD* (ANDRIENKO, 1999 apud BRANCO, 2003). Este passo acontece após o pré-processamento, quando os dados já estão prontos para serem submetidos aos algoritmos. As técnicas e os algoritmos utilizados são variados, e dependem da aplicação e da necessidade de cada usuário.

As grandes áreas da mineração de dados são: construção de modelos, extração automática de padrões e exploração visual de dados (MENDONÇA, 2001).

A área de construção de modelos envolve a criação de modelos de classificação, estimação e predição. Nestes casos o conhecimento está codificado dentro do próprio modelo.

A área de extração automática de padrões envolve a aplicação de ferramentas que determinam padrões de aglomeração ou associação entre variáveis de interesse. A obtenção de conhecimento se dá quando um perito consegue interpretar e utilizar os padrões detectados.

A área de exploração visual de dados se concentra em disponibilizar aos peritos mecanismos de visualização e exploração interativa de dados. Os peritos utilizarão estes mecanismos sobre os dados para detectar padrões e extrair conhecimentos por eles próprios.

Apesar da importância de todos os outros passos, a mineração de dados é o processo chave para a descoberta do conhecimento. O foco deste trabalho está na mineração de dados, especificamente na área de exploração visual de dados.

Na concepção de ferramentas efetivas de exploração visual de dados ou mineração visual de dados (MVD), a presença de dois mecanismos é de fundamental importância: a visualização e a interatividade A visualização diz respeito à representação dos dados e informações em formatos gráficos objetivando a ampliação da cognição. A interatividade se define como a capacidade do usuário interagir com as cenas visuais e com os dados que estão sendo explorados, eliminando informações desnecessárias, selecionando as de interesse e buscando informações mais detalhadas.

Estes dois mecanismos reúnem métodos que permitem interagir com representações visuais, exibidas como imagens para ganhar entendimento e introspecção sobre os dados e os fenômenos que eles representam (SCHROEDER,1998 apub BRANCO, 2003)

A visualização dos dados permite que informações complexas sejam apresentadas através de telas visuais melhorando a capacidade de entendimento e interpretação. Quando aliada à exploração interativa permite também a apresentação rápida de novos cenários, fornecendo ao perito maior flexibilidade e facilidade para modificar as cenas visuais e examinar conjuntos de dados específicos. Muitas ferramentas modernas de visualização combinam a capacidade de construir complexas cenas visuais com controles de seleção iterativa dos dados mostrados. Estas funcionalidades combinadas permitem que um perito possa, ele próprio, iterativamente explorar os dados (MENDONÇA; ALMEIDA, 2001).

Para Keim e Kriegel (1996) a presença do ser humano neste processo é de vital importância, pois só assim é possível combinar a flexibilidade, criatividade e conhecimento geral do ser humano com a capacidade de armazenamento e o poder de processamento dos computadores. Eles argumentam que o sistema perceptivo humano processa os dados de forma flexível, analisa eventos complexos com rapidez, reconhece propriedades não usuais e desconsidera propriedades sem interesse, além de utilizar o conhecimento geral para produzir conclusões.

4.1 Visualização de informações

Visualização é o processo de mapear os dados e informações no formato gráfico, com a finalidade de comunicar uma percepção do elemento observado. Historicamente representações gráficas têm sido utilizadas como instrumento de comunicação em diversas áreas do conhecimento humano. Seu objetivo original era comunicar idéias pré-concebidas, mas atualmente novas abordagens se propõem a utilizar mapas para criar ou descobrir novos conhecimentos. Com a evolução da computação, a geração de imagens gráficas sofreu inovações significativas e impactantes, se tornando mais eficientes a um custo cada vez mais reduzido (BRANCO, 2003).

O modelo de referência apresentado pela Figura 26 representa as etapas essenciais a serem consideradas no uso da visualização. Nele pode-se observar que os dados passam por várias etapas de transformação através de interações com o usuário até que se alcance o objetivo desejado, ou seja, a geração da imagem visual que representará o significado a ser comunicado (CARD; MACKINLAY; SHNEIDERMAN, 1999).

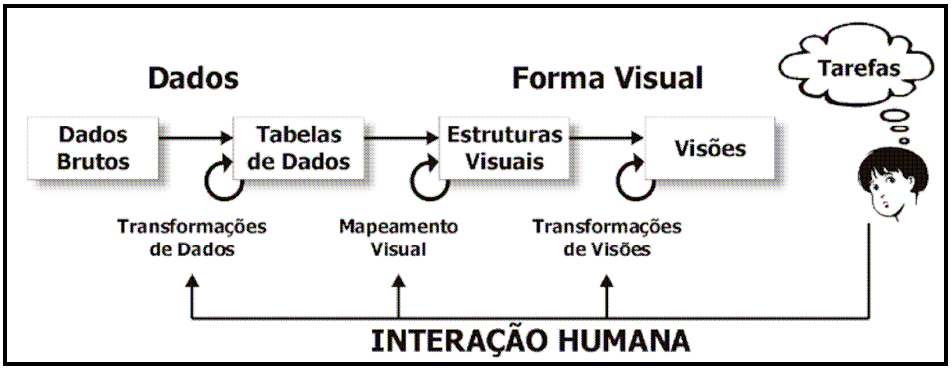

Figura 26. Modelo de referência de visualização (CARD; MACKINLAY; SHNEIDERMAN, 1999)

Neste modelo o objetivo é mapear tabelas de dados em uma metáfora visual. Na primeira etapa os dados originais são selecionados dinamicamente e passam por um processo de transformação e filtragem. O principal objetivo desta interação é selecionar os dados que farão parte da estrutura visual a ser analisada pelo usuário. No passo seguinte, o mapeamento visual, ocorre a escolha de como os dados serão associados aos atributos visuais: cor, forma, legenda, tamanho, entre outros. Por fim, as transformações visuais alteram a organização dos dados na cena visual com a utilização de recursos como: aproximação, afastamento, rotação e translação. Cada um destes passos pode acontecer várias vezes para facilitar a interpretação e o entendimento do usuário. Estes três tipos de interação ocorrem concorrentemente durante o processo de exploração visual dos dados.

Nas ferramentas modernas de exploração visual, os controles gráficos de interação são usados para filtragem, mapeamento, transformação visual com poucos cliques do mouse, permitindo os usuários explorarem e interpretarem os dados sob análise.

O conjunto de valores assumidos pelos dados pode ser classificado em dois formatos básicos: categóricos ou nominal e quantitativo. Os valores categóricos ou nominais não têm relação de ordem (ex: Inglês, Português, Alemão), já os ordinais apresentam relação de ordem (ex: pequeno, médio e grande). Os valores quantitativos representam valores numéricos da escala racional ou absoluta. Ao contrário das escalas categóricas estas escalas implicam na existência de uma unidade de medição associada aos valores das variáveis.

Sejam numéricos ou categóricos os dados a serem analisados são geralmente colocados em tabelas, onde cada linha representa um registro de dados e cada coluna um campo ou variável deste registro. Dados que envolvem múltiplos tipos de registros são representados em múltiplas tabelas, onde cada tabela representa uma entidade. Normalmente estes dados são

representados em um modelo relacional e armazenados em Sistema Gerenciadores de Banco de Dados.

Para mapear estes dados em representações visuais, as bases de dados relacionais são consideradas genericamente como um conjunto de dados multidimensionais, onde os atributos das relações correspondem às dimensões (KEIM; KRIEGEL, 1996).

Keim e Kriegel (1996) definiram visualização de informações como um conjunto de técnicas usadas para mapear graficamente informações abstratas multidimensionais armazenadas em grandes bases de dados e enumeraram estas técnicas de representação visual separando-as em algumas categorias que serão apresentadas a seguir.

4.1.1 Técnicas orientadas a Pixel

Consiste em utilizar um *pixel* para representar cada valor de um atributo, colorindo-o de uma cor pré-definida de acordo com o domínio do atributo. Para conjunto de dados com ***m*** atributos ou dimensões, a tela é dividida em ***m*** janelas, uma janela para representar cada um dos atributos. Correlações e dependências funcionais são detectadas pela análise de regiões correspondentes nas múltiplas janelas. A grande vantagem desta representação é a grande quantidade de informações exibidas simultaneamente. (KEIM; KRIEGEL, 1996; KEIM, 2000; KEIM, 2002).

Na Figura 27, a técnica de padrão recursivo apresentada é baseada em um arranjo recursivo de *pixel.*

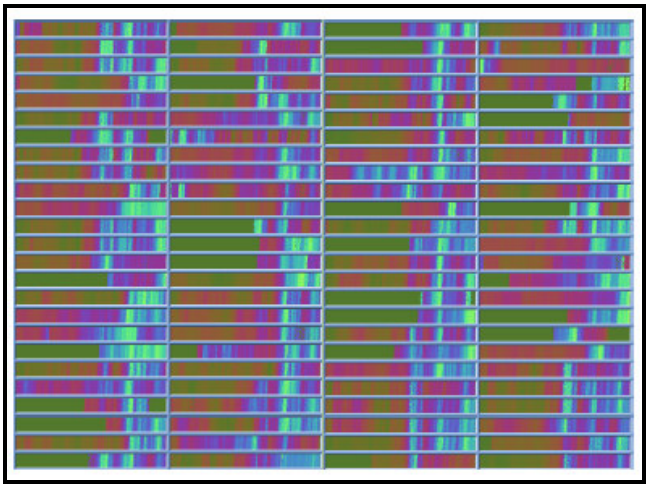

Figura 27. Técnica orientada a pixel – padrões recursivos (KEIM, 2002)

A figura mostra um conjunto de dados agrupados por um atributo, no caso uma série temporal. Neste exemplo estão representados 20 anos de preços diários de 100 itens de estoque de uma empresa.

4.1.2 Projeção Geométrica

Tem como objetivo identificar projeções de interesse em conjunto de dados multidimensionais. Uma técnica muito utilizada e chamada: *Coordenadas Paralelas* (INSELBERG; DIMSDALE, 1990). Esta técnica mapeia um espaço dimensional *K*, em um espaço visual bidimensional, usando ***K*** eixos eqüidistantes e paralelos a um eixo principal (x ou y). Cada eixo representa uma dimensão dos dados, sobre o qual será mapeado linearmente, do menor ao maior, o intervalo de valores de dados correspondentes. Cada item de dado é exibido como uma linha poligonal, que intercepta cada eixo no ponto correspondente ao valor do atributo associado ao eixo. Ver Figura 28.

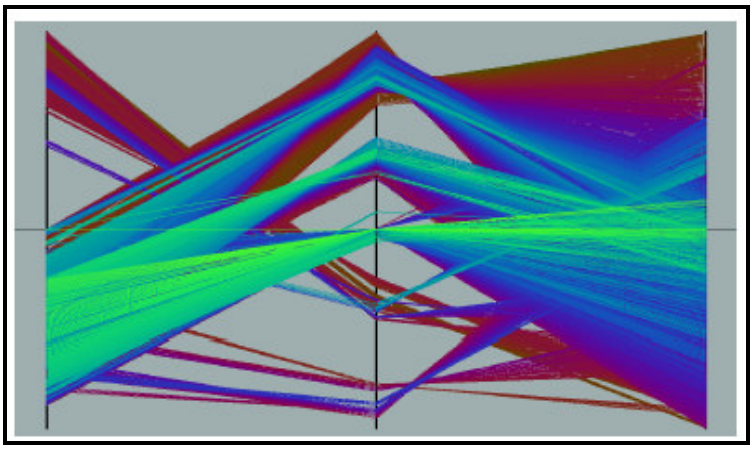

Figura 28. Coordenadas paralelas (KEIM, 2002).

Esta técnica permite a identificação de diferenças na distribuição dos dados e correlação entre os atributos. O maior problema desta técnica é a possibilidade de sobreposições de linhas em grandes volumes de dados, limitando assim sua utilização para representação de no máximo 1 000 itens de dados (KEIM; KRIEGEL, 1996).

4.1.3 Técnicas Iconográficas

Utiliza ícones para mapear os valores dos atributos de um item de dado multidimensional. Vários ícones podem ser utilizados como: pequenas faces, ícones tipo estrela, ícones de varetas e ícones de cores (KEIM, 2002). Um dos ícones mais adequados

para mineração de dados e que permite a representação de um maior volume de dados é o ícone de varetas (KEIM; KRIEGEL, 1996; PICKETT; GRINSTEIN, 1988). Para representar as dimensões usa-se variação de cores, espessura e posição das varetas.

Quando mapeados na tela os ícones formam texturas que variam de acordo com as características dos dados, explorando a capacidade humana de destingir texturas em imagens complexas.

Na Figura 29 os ícones representam cinco dimensões (variáveis), através das variações das texturas.

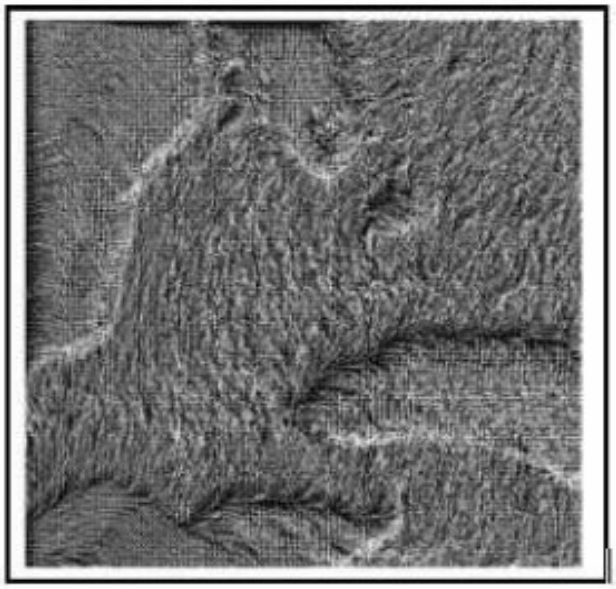

Figura 29. Imagem da Região dos Grandes Lagos (ANKERST, 2000 apud MARQUES NETO, 2004)

Esta técnica é limitada quanto à quantidade de dimensões representadas. A utilização de muitas dimensões pode provocar diminuição ou perda na capacidade de perceber os dados (WONG; BERGERON, 1997).

4.1.4 Técnicas Hierárquicas

O espaço n-dimensional é dividido em subespaços que serão organizados e exibidos de forma hierárquica, projetando ou embutindo estes espaços uns dentro do outros. O uso desta técnica é interessante na detecção de agrupamentos, de pontos com comportamentos discrepantes e de padrões (HOFFMAN; GRINSTEIN, 2002). Possui limitações quanto ao número de atributos e de valores distintos a serem visualizados. A eficácia da análise depende

do arranjo hierárquico dos atributos e dos critérios de categorização dos dados (KEIM; KRIEGEL, 1996; WONG; BERGERON, 1997).

A Figura 30 mostra um exemplo de uma técnica hierárquica chamada de Mapa em Árvore (JOHNSON; SHNEIDERMAN, 1991). Essa técnica usa retângulos recursivamente aninhados para representar hierarquias. A abordagem utiliza 100% do espaço disponível para transmitir informações, permitindo a representação de dezenas de milhares de itens em uma única cena visual.

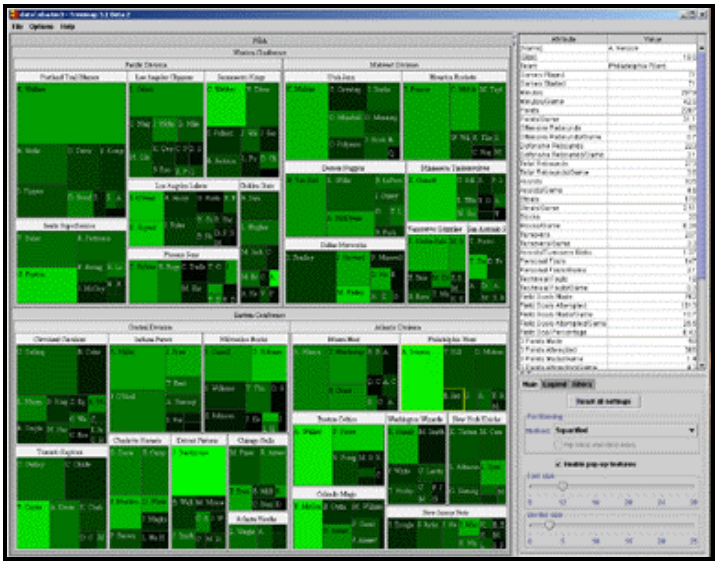

Figura 30. TreeMap ferramenta para visualização hierárquica.

4.1.5 Técnicas Baseadas em Grafos

Um grafo consiste de grupos de objetos, chamados nós, conectados entre eles através de outros objetos chamados arestas. No grafo os nós são os objetos estudados e as arestas a representação de seus relacionamentos. Situações de alto inter-relacionamento como troca de mensagens entre pessoas em ambiente de comunicação, comportamento de compras, uso de *hyperlinks* em ambiente *Web* e outros, são bem representadas através de grafos (KEIM, 2002).

As principais características visuais a serem consideradas neste tipo de representação são: número máximo de cruzamentos de arestas, exibição ótima de simetria entre vértices e arestas, distribuição uniforme dos vértices e comprimento uniforme das arestas (MARQUES NETO, 2004).

De acordo com Niggemann (2001) muitas pesquisas foram desenvolvidas recentemente na área de desenho de grafos e as mais populares famílias de algoritmos são a *força-dirigida* e o *desenho orientado a nível*.

Niggeman (2001) esclarece que o método da força-dirigida se baseia em encontrar o tamanho ótimo das arestas que ligam os nós para desenhar um grafo de forma que a figura resultante seja visualizada facilmente pelo ser humano. Isto pode ser ilustrado fazendo analogia com o campo da física, onde os nós correspondem a pequenos corpos que se repelem enquanto que as arestas correspondem a molas que os conectam. Inicialmente os nós são dispostos de forma que as molas estejam da forma mais relaxada possível, então todas as forças que afetam os nós são calculadas e interativamente aplicadas aos corpos.

A Figura 31 ilustra um exemplo de força-dirigida, no lado direito o grafo aparece com os tamanhos da aresta de forma não otimizada. As setas demonstram as forças aplicadas em cada nó. Estas forças serão utilizadas para calcular o tamanho ótimo das arestas que por fim será aplicado para ligar os nós (lado direito da figura).

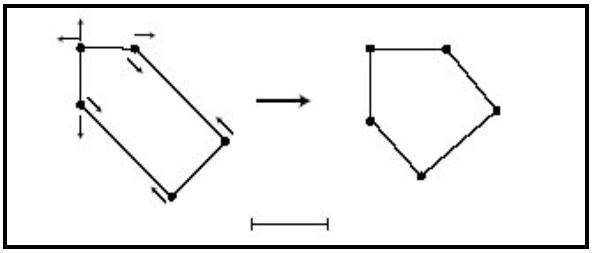

Figura 31. Cálculo do melhor tamanho das arestas em força-dirigida (NIGGEMANN, 2001)

A maior vantagem do método da força-dirigida é boa representação do desenho resultante. Como desvantagens estão: a modificação das posições dos elementos a cada interação e a complexidade interna dos algoritmos que são da ordem de $O(N^3)$ passos, devido às várias interações, enquanto que nos outros algoritmos são da ordem de $O(N)$ (MARQUES NETO, 2004).

O outro método, o desenho orientado a nível á utilizado normalmente para grafos dirigidos. O algoritmo executa três passos para a geração do grafo. No primeiro passo os nós são distribuídos em camadas horizontais, depois são organizados para minimizar os cruzamentos e só então a coordenada horizontal é ajustada para cada um dos nós (NIGGEMANN, 2001). Ver Figura 32.

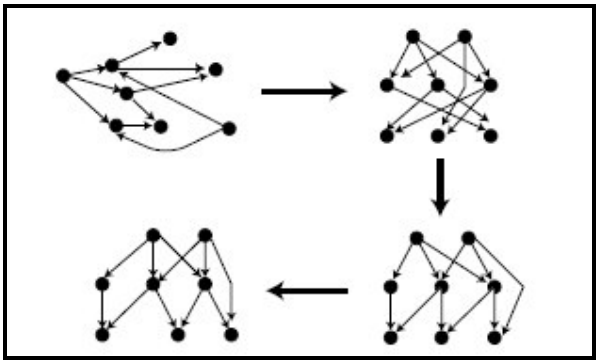

Figura 32. O três passos do desenho orientado a nível (NIGGEMANN, 2001)

As vantagens são a visualização intuitiva para grafos com muitas arestas e a boa representação hierárquica. A maior desvantagem é uma representação pobre em grafos não dirigidos, pois neste caso não existem hierarquias.

4.2 Mecanismos de interação

Associados aos mecanismos de visualização apresentados, uma área de grande importância para a exploração e descoberta de novas informações é a interatividade. Mecanismos de interação visual criam interfaces com imagens gráficas que permitem que o usuário resolva consultas de forma dinâmica utilizando apenas o mouse.

Eick (2000) argumenta que estruturas de dados, metáforas visuais e operações interativas utilizadas na visualização da informação ajudam o usuário a enxergar tendências e anomalias no modelo de negócio. Quando os dados são observados por várias perspectivas, é possível explorar os pontos fortes e minimizar os pontos fracos das diferentes técnicas de visualização. Os problemas de escalabilidade visual provocado pela limitação física dos dispositivos para representação de grandes volumes de dados, podem ser amenizados, permitindo que apenas partes dos dados, no lugar do todo, sejam analisadas.

Na exploração de informações, a combinação de múltiplas visualizações do mesmo conjunto de dados de forma coordenada, permite que usuários rapidamente explorem informações complexas (NORTH; SHNEIDERMAN, 1999).

Mendonça e Almeida (2001) assinalam que implementações eficientes de interação fornecem aos usuários uma forma simples e fácil de detectar novos padrões sem precisar de algoritmos específicos de detecção de padrões. Para isso as ferramentas devem apresentar as seguintes funcionalidades:

- Permitir o uso de diversos atributos visuais para produzir gráficos multidimensionais facilmente interpretáveis;
- Permitir a navegação interativa sobre a representação visual e o controle interativo de suas características;
- Permitir o controle interativo e dinâmico dos dados analisados.

4.2.1 Uso de atributos visuais

A fim de melhorar a identificação de aspectos relevantes nas imagens, deve-se ter preocupação com a forma de exibição das informações. Algumas formas de representações como: orientação, forma, concavidade, tamanho, curvatura e delimitações, permitem destaque de informações através da diferenciação, estes atributos visuais promovem uma fácil distinção dos registros de dados sendo atualizados (MENDONÇA, 2001).

Outro recurso muito utilizado é a cor. A abordagem mais comum é aplicar matrizes de cores diferentes para promover distinção entre valores nominais, enquanto que valores quantitativos devem receber tons e luminosidades diferentes (BRANCO, 2003).

Levkowitz e Herman (1992 apud BRANCO, 2003) acrescentam que a escolha da escala de cores para valores numéricos deve ter três propriedades desejáveis: ordem, conformidade, distância representativa e limites. A ordem das cores deve manter a ordem dos valores dos dados, cores ligadas a valores inferiores devem ser percebidas anteriores a valores superiores. A uniformidade e distância das cores devem preservar a distância entre os valores que representam. A escala de cores não deve criar percepção de limites que não existam nos dados numéricos que representam, devendo manter uma relação contínua.

4.2.2 Transformações visuais após interações

O sistema de visualização deve oferecer meios que permitam limitar a quantidade de informações que o usuário recebe, mas ao mesmo tempo mantê-lo informado sobre o espaço total de informações (FREITAS e outros, 2001). Desta forma, é necessário possibilitar formas de manipular o conjunto de dados para representá-lo como um todo ou apenas parte dele.

Shneiderman (1996) identificou sete etapas que aplicações de visualização devem promover através de interações:

Visão geral – Representação de toda a coleção de dados, está sujeita a parâmetros limitantes de equipamentos gráficos e do sistema perceptivo humano.

Zooming – Visão em detalhe de um item ou conjunto de itens de interesse. O zoom pode ser geométrico, quando for necessário visualizar maiores detalhes da imagem ou semântico quando se deseja visualizar uma porção da coleção de dados que interessa (HERMAN; MELANÇON; MARSHALL, 2000 apud BRANCO, 2003).

Filtragem – Elimina itens sem interesse, para reduzir o tamanho do conjunto de dados observado, focando apenas os itens de interesse.

Detalhes sob demanda – Apresentação de detalhes de um item ou grupo de itens após seleção. Usualmente utiliza-se o clique no objeto para apresentar os valores de seus atributos.

Relações – Observação de relacionamentos entre itens. Na descoberta de um item o usuário pode pedir outros com atributos similares

Histórico – Manutenção do histórico de ações a fim de dar suporte a ações de desfazer, repetir ou refinamentos progressivos.

Extração – Extração de resultados de operações de visualização é a capacidade de salvar em arquivos para apoiar usos posteriores.

4.2.3 Consultas dinâmicas

Nesta técnica controles visuais são associados a parâmetros para a execução de consultas num conjunto de dados. O resultado reflete imediatamente no mapeamento utilizado para visualização, mostrando apenas os itens de dados que satisfazem às seleções especificadas.

Esta técnica inicialmente apresenta todo o domínio na visualização e em seguida permite restringir ou destacar os elementos que mapeiam a informações correspondentes às consultas. Os três tipos de controles mais utilizados são: botões, controles de faixas contínuas (*range slider)* e controles de dados nominais (*check box).* Os botões são utilizados para representar valores de dados em dois estados: verdadeiro e falso, os controles de faixas continuas para representar dados numéricos, quantitativos e os controles nominais para dados discretos (BRANCO, 2003).

4.3 Visualização em Redes Sociais

Segundo Freeman (2000) as imagens computacionais e as métricas baseadas em conceitos matemáticos são os dois fatores responsáveis pelo crescimento em pesquisas de redes sociais. Os debates e pesquisas utilizando imagens para representar redes sociais originam-se nos anos 1930, quando o psiquiatra J. L. Moreno publicou a abordagem chamada *sociometria* (*Who Shall Survive?*), onde as imagens gráficas eram produzidas manualmente apresentando pontos para representar as pessoas e linhas unindo estes pontos representavam as relações. As imagens produzidas não possuíam ainda uma estrutura adequada e seu sucesso variava com a capacidade de entendimento e perfil de seu criador. A evolução da visualização das redes sociais a partir deste ponto passou por mais quatro fases distintas.

A segunda fase aconteceu nos anos 1950, quando pesquisadores iniciaram a utilização de procedimentos de computação para produção de imagens. A teoria dos grafos trouxe o mecanismo para a construção de modelos estruturais mais genéricos, uma grande colaboração na área que juntamente com o aumento da disponibilidade dos computadores permitiu produzir imagens automaticamente na década de 1970.

Finalmente a partir de 1990 com a popularização de aplicações distribuídas, redes podiam ser analisadas remotamente. Neste caso a *WWW* (*World Wide Web*) passou a ser a escolha natural como plataforma de análise.

Atualmente o uso de cores, animação e maior flexibilidade permitem o usuário interagir com as imagens que recebe. Através da evolução destas ferramentas as imagens se tornaram um instrumento crítico e fundamental no auxílio das investigações, entendimento e comunicação dos pesquisadores em comportamentos das redes sociais.

Freeman (2000) enfatiza que apesar dos avanços dos programas de visualização, a real maturidade deste ferramental ocorrerá quando associarmos uma camada de acesso à banco de dados modernos, facilitando desta forma o registro e recuperação dos dados de redes sociais.

Apesar de existirem várias ferramentas para Análise de Redes Sociais, a grande maioria ainda não disponibiliza uma adequada visualização interativa, normalmente as métricas são apresentadas em tabelas sem refletir seus resultados nas cenas visuais. Na maioria das ferramentas existem poucos atributos nos atores, apenas um tipo de relação é analisado e não existe registro do conteúdo transacional. Outra observação nas pesquisas das

ferramentas existentes é a não utilização de gerenciadores bases de dados como infra-estrutura de armazenamento e consulta dos dados. A grande maioria das ferramentas analisadas utiliza arquivos convencionais para leitura e gravação dos dados.

4.4 Considerações Finais

Neste capítulo foram apresentadas a fundamentação teórica e a revisão de literatura sobre o tema Mineração Visual de Dados. Termos e conceitos, importantes para contextualizar o trabalho como *KDD* e Mineração de Dados foram apresentados, definidos e inter-relacionados. Por fim foi relatado um breve histórico de como a computação tem sido utilizada ao longo dos anos por ferramentas de Análise de Redes Sociais. No próximo capítulo será apresentada a proposta deste trabalho, e que contribuições ela introduz as ferramentas de *SNA*.

CAPÍTULO 5 - A FERRAMENTA *FLOWMINER*

Instrumentos para apoio à gestão de conhecimento e auxilio à aprendizagem organizacional têm despertado o interesse e atraído a atenção de pesquisadores e executivos. O crescimento de investimentos voltados para valorização do capital intelectual e social mostra a preocupação destes executivos nos recursos humanos como fonte de conhecimento e a crença de que práticas de cooperação e solidariedade aumentam a produtividade e criam um ambiente propício à criatividade.

A ferramenta *FlowMiner* foi desenvolvida com o objetivo de oferecer uma forma de representação visual para exploração de redes sociais intra-organizacionais de comunicação. A partir da determinação dos colaboradores que compõem esta rede e de seus relacionamentos através da troca de informações e conhecimento é traçado um mapa em forma de grafos, onde os nós são os colaboradores e as arestas suas ligações de comunicação. A ferramenta utiliza controles que permitem interagir com a cena visual fazendo seleção por valores de atributos ou filtros, transformando-a em cenários escolhidos pelos usuários. Nestes cenários podem ser aplicadas métricas de redes sociais para fazer a análise estrutural da rede e o entendimento do papel de cada colaborador na comunidade em estudo. A apresentação dos resultados estatísticos obtidos com as métricas incorpora técnicas de visualização. Estas técnicas utilizam padrões de cores que são refletidos na cena visual para representar a intensidade dos valores obtidos por cada colaborador ou ligação.

5.1 Inovações da *FlowMiner*

A ferramenta proposta traz três contribuições em relação às ferramentas de Análise de Redes Sociais existentes.

1. Ela aumenta a capacidade semântica dos objetos estudados.

Foi concebido um modelo de dados que além de possuir atores e ligações, introduziu um novo objeto, o conteúdo transacional. Como definidos anteriormente conteúdos transacionais são os recursos materiais e não materiais trocados pelos atores através das ligações. Na abordagem de redes sociais organizacionais estes recursos são representados pelos fluxos de informações trocados entre os atores. Neste modelo de dados os objetos atores e conteúdo transacionais (fluxos) passam a ser qualificados e categorizados por um número diverso e dinâmico de atributos. As ligações são tratadas como tubos ou canais de conexões direcionados através dos quais trafegam os fluxos. Para implementar esta abordagem os dados coletados são representados em um modelo de dados expansível e armazenados em um banco de dados. A variedade de atributos e qualificadores dos objetos dependerá do tipo e necessidade da organização estudada.

2. Ela utiliza técnicas de exploração visual para representar os resultados encontrados.

O uso de metáforas visuais torna os resultados apresentados mais simples e fáceis de compreender. Com poucos cliques do mouse, detalhes dos objetos visualizados poderão ser apresentados, objetos poderão ser retirados da cena visual ou centralizados. Os cenários apresentados em grafos poderão sofrer aproximações ou afastamentos, rotação e diminuição ou aumento do raio de exibição para melhorar a visualização. A apresentação das ligações em diversos tons de azul mostra a qualidade das ligações entre os atores, enfatizando ligações com maior intensidade de fluxos. Cores também são utilizadas para apresentar os resultados das métricas que atuam sobre a cena visual apresentada, realçando com tons diferenciados de vermelho os valores obtidos e evidenciando os atores mais bem pontuados.

3. Ela incorpora interatividade baseada nos vários atributos que qualificam os objetos.

Atores e conteúdos transacionais (fluxos) poderão ser selecionados na cena visual através da combinação de valores dos seus atributos. Este processo permite a identificação de comunidades de prática que se reúnem ou trocam informações sobre os principais assuntos de interesse das mensagens. Sobre estas comunidades que formam a cena visual após a seleção, poderão ser aplicadas as métricas que identificam atores que se destacam do grupo. Isto permite a localização para um dado cenário dos conectores centrais, corretores de informações ou expansores de fronteiras, conceitos já descritos na seção 3.4.

5.2 Usuários da ferramenta

O público alvo da ferramenta *FlowMiner* são gerentes e funcionários, uma vez que um dos seus objetivos é facilitar o processo de socialização do aprendizado dentro da organização. As informações obtidas com o uso da ferramenta permitirão aos gerentes planejarem a execução de ações corretivas e melhorias no ambiente de trabalho visando facilitar e motivar a colaboração e a disseminação de conhecimentos e práticas de trabalho, mantendo desta forma um ambiente produtivo e saudável. Os funcionários utilizarão a ferramenta para conhecer e entender seu papel nas comunidades das quais participam e enxergar que novas ligações poderiam implementar em seus contatos para melhorar seu envolvimento e aumentar sua capacidade de recorrência. Para isso, a representação visual da rede de comunicação é disponibilizada, bem como a navegação através dos seus elementos, definidos a partir de diversas formas de pesquisa sobre os valores de atributos contidos na base de dados. Estas pesquisas são implementadas através de mecanismos que permitem a identificação de comunidades que trocam informações sobre um determinado cliente, projeto ou assunto de interesse, e/ou sobre indivíduos que se comunicam e possuem valores de atributos comuns como departamento, função, competências e habilidades.

A ferramenta tem como foco principal o conteúdo transacional, ou seja, os fluxos de informações trocados através das ligações de comunicação que deverão estar devidamente categorizados e sumarizados por assunto de interesse. A partir desta categorização a ferramenta interage com os atributos e identifica seu domínio de valores possibilitando iluminar as informações que fluem entre os colaboradores para execução de suas tarefas do dia-a-dia. A ferramenta permite a identificação e visualização de comunidades de prática através de interações sobre o conteúdo transacional trocado e a aplicação de métricas para estabelecer o papel que cada colaborador desempenha na comunidade estudada. Os papéis que

são apontados estão descritos originalmente no artigo de Cross e Prusak (2002) e foram discutidos na seção 3.4.

5.3 Modelo de Dados

A ferramenta usa como entrada um modelo relacional de dados, composto de cinco entidades básicas e seus relacionamentos, representado na Figura 33.

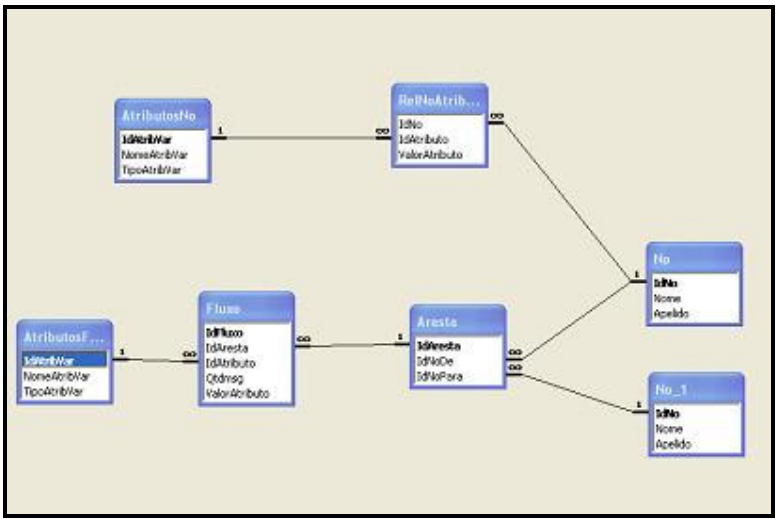

Figura 33. Modelo de dados gerado pelo produto *MS ACCESS*

As principais entidades do modelo são:

- Nó – Contém informações sobre um colaborador da organização estudada, pode ser um funcionário, terceiro, fornecedor ou cliente. Como atributos fixos possui id, nome, apelido.
- Aresta – Contém informações sobre uma ligação entre dois colaboradores. Como atributos fixos possui id, node (identificador do colaborador origem da ligação) e nopara (identificação do colaborador destino da ligação).
- Fluxo – conteúdo transacional das ligações, ou seja, mensagens consolidadas por categoria que passam através de uma ligação. Como atributos um fluxo possui id, idaresta, idatributo, qtdmsg e valoratributo.

- AtributosNo – Possíveis atributos de um colaborador. Como atributos possui id, nome e tipo. São as dimensões do ator que podem ser analisados. Esta entidade permite a criação de vários atributos, quantos forem necessários, para qualificar um colaborador.
- AtributosFluxos – Possíveis atributos que categorizam os fluxos. Como atributos possui id, nome, tipo. São as dimensões ou categorias que os fluxos podem ser analisados. Esta entidade permite criar quantos atributos forem necessários para qualificar os fluxos.

Existem relacionamentos entre a entidade *Nó* e *AtributosNo, Fluxo* e *AtributosFluxo*, exatamente para permitir dimensões e atributos variados nos colaboradores e nos fluxos.

Todas as entidades e um relacionamento são representados por tabelas no modelo do banco de dados relacional. Ao todo existem seis tabelas: *Nó, Aresta, AtributosNo, Fluxo, AtributosFluxos e RelNosAtributos.*

Na tabela de *Nó* estão registrados os atores da população estudada. Os atores poderão ser pessoas, grupos, equipes, empresas ou qualquer entidade que se deseje analisar. A versão atual da ferramenta pode representar apenas um nível de atores por vez. Para representar um ator o registro é formado de três atributos fixos que são: número de identificação do ator; nome completo do ator e apelido ou nome mais conhecido do ator. Exemplo de uma tupla: 1, Maria de Fátima Torres, Fátima.

A tabela de *Aresta* representa a existência de ligação ou canal de comunicação entre dois atores quaisquer. Basta haver um conteúdo transacional (no caso de comunicação: uma mensagem) trocado entre dois atores e uma ligação deverá ser estabelecida. Para registrar uma ligação ou aresta existem três atributos fixos: número de identificação da aresta, número de identificação do ator origem e número de identificação do ator destino. Exemplo de uma tupla: *1, 1, 2*. Significando que a aresta de número *1* vai do ator *1* para o ator *2*.

A tabela *AtributosNo* representa os atributos variáveis e dinâmicos que cada ator poderá possuir. A definição destes atributos dependerá de cada organização, estes atributos representam as características ou qualidades dos atores que se deseja analisar. A variedade destes atributos significa maior semântica e flexibilidade no entendimento da composição dos recursos humanos. Para registrar uma entrada nesta tabela existem três atributos fixos:

identificação do atributo, nome do atributo e tipo do atributo. Exemplo de uma tupla: *1, Departamento, String*. Significando que o atributo 1 de *Nó* é *Departamento* e este atributo receberá valores do tipo *String*.

A tabela *RelNoAtributo* representa o relacionamento entre o *Nó* e seus atributos variáveis. Cada Nó poderá estar associado a nenhum ou a todos os atributos da tabela anterior. Não existe obrigatoriedade nos atributos variáveis. Cada entrada nesta tabela representa a característica ou qualificação do ator em foco. Exemplo de uma tupla: *1, 1, Produção*. Significando que o ator de identificação *1*, que é *Fátima*, tem para o atributo *1* que é *Departamento*, o valor *Produção*, ou seja, ela está alocada no departamento de Produção. Desta forma vários atributos poderão caracterizar e qualificar os atores. Permitindo inclusive mais de um valor por atributo. No exemplo do atributo *Habilidades*, um ator poderá estar associado a vários valores como: *Negociador*, *Comunicador*, entre outros.

Na tabela *AtributosFluxos* é semelhante a *AtributosNo*, ela representa os atributos variáveis e dinâmicos que os fluxos possuirão. O fluxo representa uma consolidação de mensagens por assunto, são mensagens de um mesmo assunto que passam por uma aresta específica. Os assuntos ou categorias que serão registrados nesta tabela dependem dos interesses de análise da organização. Para registrar uma entrada nesta tabela existem três atributos fixos: identificação do atributo, nome do atributo e tipo do atributo. Exemplo de uma tupla: *1, Tecnologia, String*. Significa que o fluxo será categorizado pelo assunto de *Tecnologia* cujo valor é do tipo *String*.

Finalmente a tabela de *Fluxos* representa o relacionamento entre *Aresta* e *AtributosFluxos*. Uma aresta poderá ter vários fluxos passando por ela, cada fluxo representa um grupo de mensagens consolidado por um assunto. A versão atual da ferramenta esta limitada a um assunto por fluxo, mas já existe um planejamento para na próxima versão permitir um número ilimitado de atributos por fluxo, esta restrição não está no modelo de dados e sim nas estruturas internas utilizadas no programa. Para registrar uma entrada na tabela de Fluxos existem cinco atributos fixos: identificação do fluxo, identificação da aresta onde o fluxo passa, identificação do atributo que qualifica o fluxo, quantidade de mensagens e valor do atributo. Exemplo de uma tupla: *1, 1, 1, 10, Banco de Dados*. Significando que o fluxo *1* da aresta *1* (do ator *1* para o ator *2*) diz respeito à categoria *Tecnologia* e que possui *10* mensagens sobre o assunto *Banco de Dados*. Desta forma vários fluxos com intensidade

(quantidade de mensagens) de vários assuntos podem ser registrados numa aresta ou na comunicação entre dois atores.

5.4 Requisitos para utilização da ferramenta

Vários processos devem ser executados para viabilizar a análise de redes sociais. Este trabalho é uma proposta de ferramenta que engloba apenas o processo de visualização interativa e análise de redes intra-organizacionais, os processos anteriores: de seleção da amostra, captura, transformação e classificação dos dados no modelo de dados apresentado não fazem parte deste trabalho. No capítulo 6 onde está descrito o estudo de caso da aplicação da ferramenta, será abordado como estes processos foram executados para obtenção dos dados utilizados.

A utilização da ferramenta numa organização, pré-supõe que alguns requisitos deverão ser atendidos. A seguir serão apresentados os principais requisitos:

1. Identificar a amostra da população que será estudada;
2. Coletar, classificar e organizar os dados de interesse da organização para categorização dos colaboradores escolhidos como amostra. Estes dados poderão ser coletados de bases de dados ou arquivos corporativas ou através de formulários ou questionários, onde serão levantados todos os atributos necessários para qualificar cada colaborador. A quantidade destes atributos pode variar de acordo com a necessidade da organização;
3. Identificar e registrar uma taxonomia plana de categorias que interessam para qualificar os fluxos de informações. A taxonomia retratará os assuntos de interesse para descoberta de comunidades de prática e dizem respeito aos domínios de conhecimento da organização com assuntos técnicos, assuntos administrativos, projetos em desenvolvimento, clientes, entre outros;
4. Coletar, classificar e consolidar os fluxos de informações entre os colaboradores. Estes fluxos podem ser coletados através de questionários ou entrevistas com os colaboradores ou poderão ser extraídos utilizando o texto das mensagens de ferramentas de comunicação como correio eletrônico, *chats*, *newsgroups, workflow e HelpDesk*. Para categorizar ou classificar os textos, técnicas de

mineração de dados, especificamente classificação, poderão ser adotadas, usando como atributos os termos da taxonomia do item anterior;

5. Disponibilizar o modelo de banco de dados apresentado preenchido com as informações coletadas no item 2 e 4;
6. Conscientizar os colaboradores envolvidos, estimulando sua participação através da divulgação dos benefícios a serem obtidos. Por questões éticas já apresentadas, a participação dos colaboradores só poderá ser obtida por adesão.

Os requisitos que vão do item 1 ao 5 dizem respeito a construção do modelo de dados utilizado pela ferramenta. O requisito 6 se refere ao processo de apoio à implantação da ferramenta.

5.5 Requisitos funcionais da ferramenta

Para que os objetivos definidos para a ferramenta *FlowMiner* sejam alcançados se faz necessária a definição de alguns requisitos funcionais relacionados a sistemas de análise de redes sociais combinados a técnicas de exploração visual de dados, conforme descrito a seguir.

1. Visualizar em estrutura de grafos os relacionamentos de comunicação entre os atores da comunidade;
2. Permitir um entendimento dos elementos que compõem o grafo através da exploração de detalhes de seus atributos;
3. Reduzir ou ampliar a comunidade estudada através de cenários de grafos variados dependendo dos valores dos atributos dos atores e dos fluxos de informações que trafegam entre eles;
4. Utilizar métricas de *SNA* para entender e analisar a estrutura da rede como um todo, de uma comunidade de prática ou da rede individual de cada ator;
5. Ajudar na localização de atores críticos na rede, ou seja, conectores centrais, corretores de informações e expansores de fronteiras;
6. Facilitar a socialização, ajudando coordenadores e colaboradores a entenderem sua participação na rede, sua rede de contatos e que ligações seriam interessantes serem ampliadas;

7. Apoiar o processo de avaliação do capital social da organização buscando motivar e estimular a troca de conhecimento e sua disseminação;
8. Facilitar a socialização, integrando pessoas, e criando oportunidades para aumento do aprendizado organizacional.

Os três primeiros requisitos se referem ao modelo de representação da comunicação em organizações que deve ser estabelecido pela ferramenta. Os requisitos 4 e 5 estão apoiados nas métricas e medidas presentes em ferramentas de análise de redes sociais. Os três últimos requisitos visam melhoria do processo de comunicação através da interpretação dos resultados fornecidos pelo mapeamento.

5.6 Desenvolvimento da ferramenta

A ferramenta foi desenvolvida na plataforma *Java*™, pois este ambiente utiliza o paradigma de orientação a objeto, além de promover a portabilidade de código e interação com diversas plataformas, inclusive com plataformas *WWW*. O paradigma de orientação a objetos direciona o desenvolvimento para a reutilização de código aumentando a produtividade na construção de sistemas e programas.

A ferramenta integra três abordagens importantes: uso de banco de dados, exploração visual de dados e algoritmos para métricas de redes sociais. Na questão banco de dados poderá ser utilizado qualquer produto gerenciador de banco de dados relacional, compatível com a *API* (*Application Program Interface*) *JDBC* (*Java DataBase Connectivity)* (SUN, 2003), isto trás flexibilidade e robustez para a gerência de dados da aplicação.

Para exploração visual de dados foram construídos diversos controles que utilizam como arcabouço visual a ferramenta *TouchGraph*. O *TouchGraph* é um arcabouço (*framework*) livre escrito em *Java* para construção de visualizações interativas que utilizam grafos como paradigma visual. As classes do *TouchGraph* permitem aos programadores facilmente incorporarem às suas aplicações funcionalidades de desenho, navegação, aproximação, rotação e modificação do raio de exibição dos grafos apresentados (TOUCHGRAPH, 2004).

Este arcabouço foi escolhido por suas características técnicas, por ser escrito em *Java*, ser usado por uma série de ferramentas como sua base de construção e por já ter sido testado com sucesso no NUPERC. O *TouchGraph* utiliza algoritmo de força-dirigida para desenhar os

grafos, seu processo de desenho e apresentação cria redes visualmente atrativas e fáceis de interpretar pelos usuários. A Figura 34 mostra um exemplo de um grafo desenhado pela ferramenta.

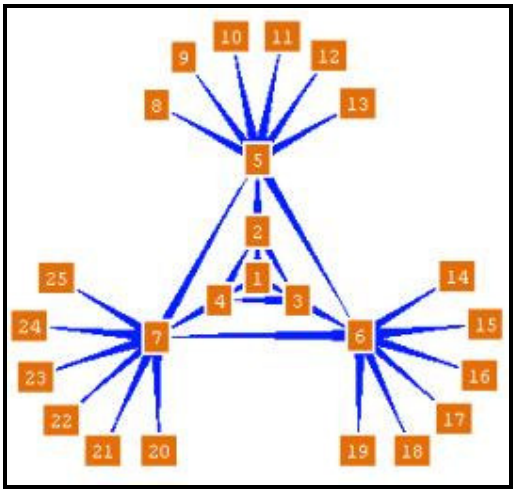

Figura 34. Visualização através de grafos usando força-dirigida (TOUCHGRAPH, 2004).

Os algoritmos para métricas de Redes Sociais que foram incorporados à ferramenta *FlowMiner* são um subconjunto daqueles encontrados na revisão de literatura *SNA* e foram escolhidos baseados nas suas características de menor custo computacional, desempenho e popularidade. Estes algoritmos estão identificados e referenciados na descrição dos módulos que compõem a arquitetura da ferramenta. Todos os algoritmos incorporados na ferramenta *FlowMiner* foram implementados pela autora e os resultados comparados contra aqueles produzidos pela ferramenta *UCINET 6* (BORGATTI; EVERETT; FREEMAN, 2002) para várias amostras de dados.

5.7 Arquitetura da ferramenta

A ferramenta *FlowMiner* foi construída e desenvolvida utilizando cinco módulos lógicos. Estes módulos se relacionam formando o conjunto de métodos funcionais que atendem a todo processamento do produto. Ver Figura 35.

1. Modulo de entrada de dados
2. Modulo de apresentação visual
3. Modulo de controle de consultas
4. Modulo de controle de interações

5. Modulo de métricas da rede social

Figura 35. Arquitetura da ferramenta *FlowMiner*

Os módulos da ferramenta foram implementados utilizando-se um grupo de classes cuja união é representada por componentes de software. Estes componentes interagem uns com outros através de interfaces públicas e possuem propriedades importantes, tais como o fato de serem independentes, autocontidas e reutilizáveis.

O módulo de entrada de dados é responsável pela recuperação das informações no banco de dados e armazenamento destas informações em estruturas na memória que serão passadas para os outros módulos. Após a leitura dos dados, o módulo de apresentação visual mapeia as informações lidas em forma de grafo.

O módulo de controle de interações provê a funcionalidade de transformação da cena visual apresentada utilizando aproximação, afastamento, rotação, mudança do raio de exibição para que o usuário controle o posicionamento do grafo na tela. Ele também possibilita a aplicação de filtros que removem atores ou ligações não desejados na cena. Os filtros são aplicados aos atributos quantitativos dos atores e fluxos. Consultas visuais mostram também a distribuição dos valores dos atributos dos atores e ligações. Além disso, o módulo de interação permite o uso do mouse sobre qualquer elemento do grafo mostrado para obter uma descrição detalhada do mesmo.

O módulo de controle de consultas age sobre a apresentação visual permitindo que escolhas do usuário por valores de atributos variáveis de atores e de fluxos modifiquem a cena visual e mostrem cenários diversos destes elementos.

O módulo de métricas de rede social provê a aplicação de algoritmos de Análise de Redes Sociais sobre a representação visual para obter métricas no cenário montado. Estas métricas atuam sobre a cena visual pintando seus elementos em tons variados de uma mesma cor para enfatizar a diferença dos valores calculados entre eles e apresenta as estatísticas encontradas mostrando o valor calculado para cada ator em uma tabela.

Estes módulos estão associados às seguintes funções básicas da ferramenta:

- Obtenção das informações relativas aos elementos atores, ligações e fluxos;
- Representação e visualização dos elementos através de mapeamento por grafo;
- Seleção de cenários por valores de atributos de atores e fluxos;
- Interação com os elementos do cenário do grafo;
- Aplicação de métricas de redes sociais no cenário do grafo;
- Obtenção da distribuição dos atributos dos atores e ligações;

O usuário poderá interagir com os seguintes casos de uso, ver Figura 36.

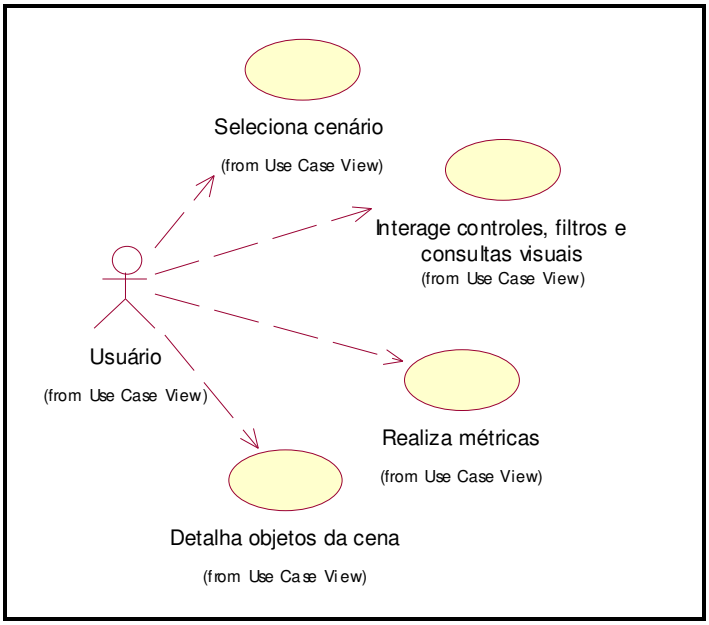

Figura 36. Diagrama de caso do uso da ferramenta *FlowMiner*

5.7.1 Módulo de Entrada de Dados

O módulo de entrada de dados foi desenvolvido com o objetivo de permitir a integração com qualquer software banco de dados relacional. Este módulo é responsável por todo o acesso às bases de dados e pela carga destas informações em estruturas de dados específicas. Para acesso aos dados existe a classe concreta *AcessaBD* que provê a conexão com o banco de dados e posterior recuperação dos dados a partir das consultas pré-estabelecidas. As classes *EstruturasAresta*, *EstruturasNo*, *ListaAdj, Matriz* e *MatrizND* utilizam a classe *AcessaBD* para solicitar consultas ao banco de dados e carregar os dados nas respectivas estruturas de dados que serão manipuladas pelas próximos módulos da ferramenta. Ver Figura 37.

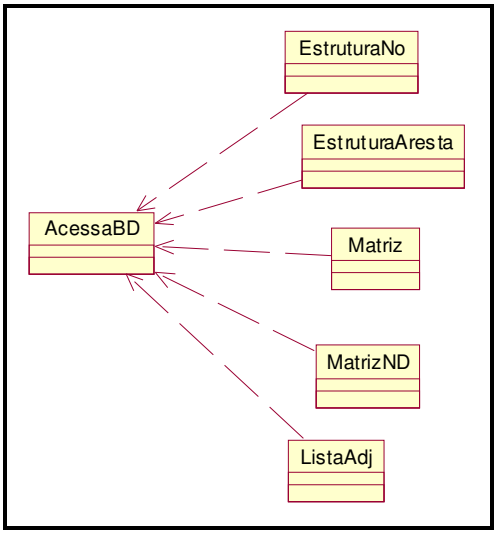

Figura 37. Diagrama de classes que implementam o módulo de *Entrada de Dados*

A classe *Aplicação (main class)* que é acionada no início do programa é responsável pela leitura de todo o banco de dados através do módulo de entrada e carregá-lo em uma matriz de adjacência onde todas as ligações são duplicadas para se tornarem recíprocas, neste método a matriz é transformada em um a matriz não direcionada. Esta matriz será lida para descobrir os componentes fortemente conectados, ou seja, se o grafo está representado por um único grupo ou está subdividido em subgrupos disjuntos. Na teoria de grafos os componentes fortemente conectados são conjuntos máximos de vértices, tal que, para cada par de vértices eles sejam mutuamente alcançáveis (ZIVIANI, 2004). A matriz de adjacência onde todas as ligações são recíprocas é obtida através da execução da técnica conhecida com *simetrização pelos máximos* (BORGATT; EVERETT; FREEMAN, 2002). Esta rotina encontrada no produto *UCINET 6* foi reproduzida na ferramenta *FlowMiner* através da classe MatrizND.

Através da estrutura de subgrupos ou componentes encontrados e exibidos na tela em forma de grafo, o usuário deverá escolher qual o subgrupo que ele deseja explorar, vide Figura 38. Caso o subgrupo possua apenas um ator o usuário receberá uma mensagem solicitando a escolha de outro subgrupo. A tela mostrada na Figura 38 apresenta do lado direito uma tabela onde para cada subgrupo encontrado estão expostas as quantidades de atores e ligações existentes.

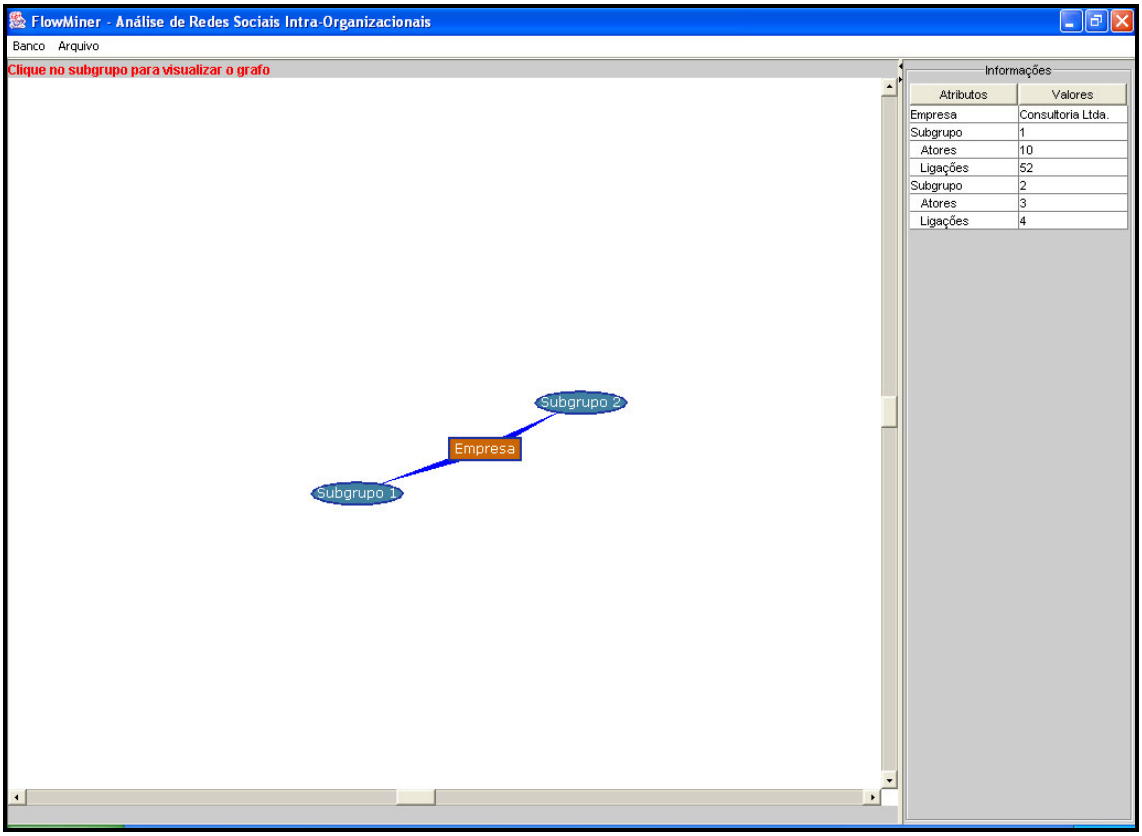

Figura 38. Tela inicial do *FlowMiner* mostrando subgrupos encontrados na base de dados.

Após, escolhido o subgrupo com o clique do mouse, o controle será passado para a classe *MostraGrafo* que através do módulo de entrada de dados seleciona no banco de dados todos os registros das tabelas relacionados com os elementos do subgrupo escolhido, instancia as respectivas estruturas de dados e apresenta visualmente o subgrupo construindo um grafo com todos os seus elementos: atores, ligações e fluxos de informações, usando para isso o módulo de apresentação. Ver Figura 39.

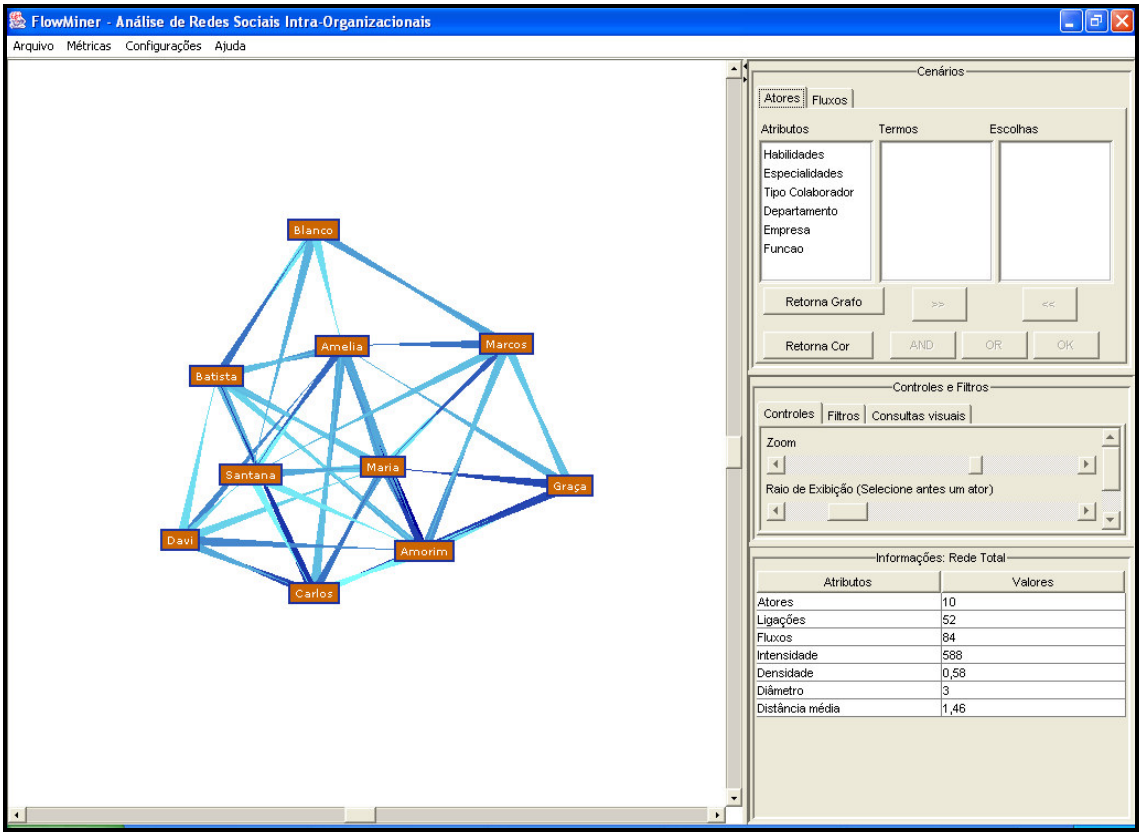

Figura 39. Tela principal do *FlowMiner* mostrando subgrupo 1 escolhido.

5.7.2 Módulo de Apresentação Visual

Na Figura 39 observa-se a apresentação das estruturas lidas após a escolha do subgrupo 1. Nesta tela, que a partir de agora será referenciada como tela principal, existem quatro áreas distintas.

A primeira área, a maior que se encontra à esquerda da tela, apresenta o grafo representando o mapa de comunicação do subgrupo escolhido. O grafo é construído usando o arcabouço *TouchGraph.* Este arcabouço é composto por um conjunto de classes e interfaces que fazem a representação de uma coleção de objetos na estrutura visual de um grafo. A classe, que compõe o painel desta área *FWPainel,* herda da classe *GLPanel* do *TouchGraph* todas as suas características. No desenvolvimento deste trabalho adaptações foram introduzidas no *TouchGraph* para que o grafo pudesse exibir informações específicas dos dados explorados. A Figura 40 mostra um diagrama das modificações feitas no arcabouço.

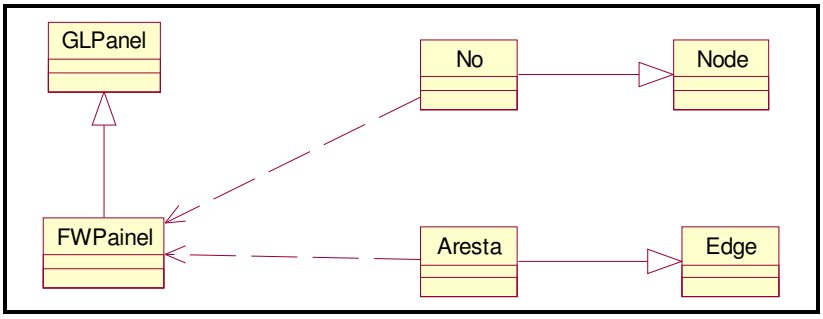

Figura 40. Diagrama das classes do *TouchGraph* estendidas para *FlowMiner*

Após a seleção do subgrupo, cada elemento que o compõe será lido das tabelas do banco e carregado como objetos de sua respectiva classe: *No, Aresta* e *Fluxo*. Desta forma todos os elementos selecionados e seus respectivos atributos e relacionamentos são representados como nós e arestas na estrutura visual. A classe *FWPainel*, herdeira da *GLPanel*, permite o desenho do grafo que será feito com a inserção dos objetos do tipo *Node* e *Edge* contidos no *TouchGraph*. O objeto *Fluxo* não possui uma representação visual no grafo, para apresentá-lo foi utilizado o recurso visual de tonalidade do objeto *Aresta*. A tonalidade é obtida através do valor do atributo intensidade que pertence ao objeto Aresta. Este atributo representa quantas mensagens no total trafegam por uma aresta específica, quanto maior for esta quantidade mais forte será a tonalidade, mostrando a intensidade ou importância da ligação entre os atores.

A segunda área da tela principal, ver Figura 39, representa as opções de consulta que o usuário tem para modificar o cenário do grafo. O módulo de controle de consulta discutido adiante na seção 5.7.3 é responsável por estas funcionalidades.

A terceira área mostra os controles de interações que são compostos de controles, filtros e consultas visuais. A responsabilidade da implementação desta funcionalidade está no módulo de controle de interações que será discutido na seção 5.7.4.

A quarta e última área, diz respeito a tabelas variadas que são preenchidas a depender o procedimento adotado. Sempre que um grafo é apresentado na primeira área, a quarta área mostrará as métricas de rede social relativas a este grafo. A seção 5.7.5, sobre o módulo métricas da rede, explicará este procedimento.

Todos os objetos contidos na primeira área do grafo são sensíveis ao *mouse*, se o usuário pressionar o *mouse* sobre os objetos nós ou arestas, a tabela da quarta área será

preenchida com detalhes do respectivo objeto. Para viabilizar este procedimento métodos pertencentes ao *TouchGraph* foram adaptados para mostrar através de um objeto *ModeloTabela* os detalhes do nó ou aresta selecionada com o *mouse*. Para implementar esta tabela que a partir de agora será utilizada por todos os módulos para detalhar os resultados, foram criadas duas classes *ModeloTabela* e *EsquemaTabela*. A classe *ModeloTabela* é apenas um repositório dos dados passados como parâmetro à classe *EsquemaTabela* que é responsável pela exibição. Ver Figura 41.

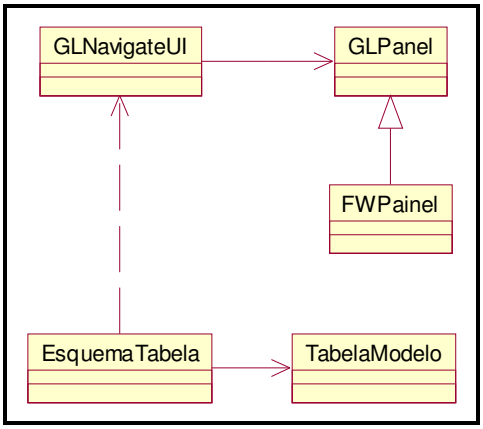

Figura 41. Diagrama das classes que implementam detalhes dos objetos da cena visual.

5.7.3 Módulo de Controle de Consulta

A classe principal deste módulo chama-se *Cenário*, que é uma extensão da classe *JPanel* do pacote *Java Swing* (SUM, 2004). Neste painel são apresentadas duas listas, preenchidas com os atributos encontrados para os atores e fluxos lidos das respectivas tabelas do banco de dados. O usuário poderá escolher termos do domínio de cada atributo e utilizar operadores lógicos (AND, OR) para executar as consultas. Para voltar à cena inicial do grafo o usuário pode utilizar o botão *Retorna Grafo*.

Na Figura 42 o cenário escolhido foi por fluxos sobre o assunto *Tecnologia* que apresentam os termos (categorias) *Banco de Dados* ou *Linguagens,* o operador lógico escolhido foi OR. Após a execução das seleções, modificações são introduzidas no grafo para deixar na cena visual apenas os elementos que satisfazem a consulta. Consultas podem se repetir utilizando o botão *AND*. A visão de cenários por variação nos atributos e termos permite iluminar comunidades por assunto de interesse, projetos, clientes, ou qualquer outro assunto capturado na taxonomia, fornecendo uma visão clara de como os indivíduos interagem para executar suas tarefas.

Figura 42. Foco no painel de *Cenários* preenchido com os termos escolhidos

Na Figura 43 pode-se observar a modificação da cena visual levando em consideração a seleção executada na Figura 42 e a conseqüente atualização da quarta área com as novas métricas sobre a estrutura geral do grafo.

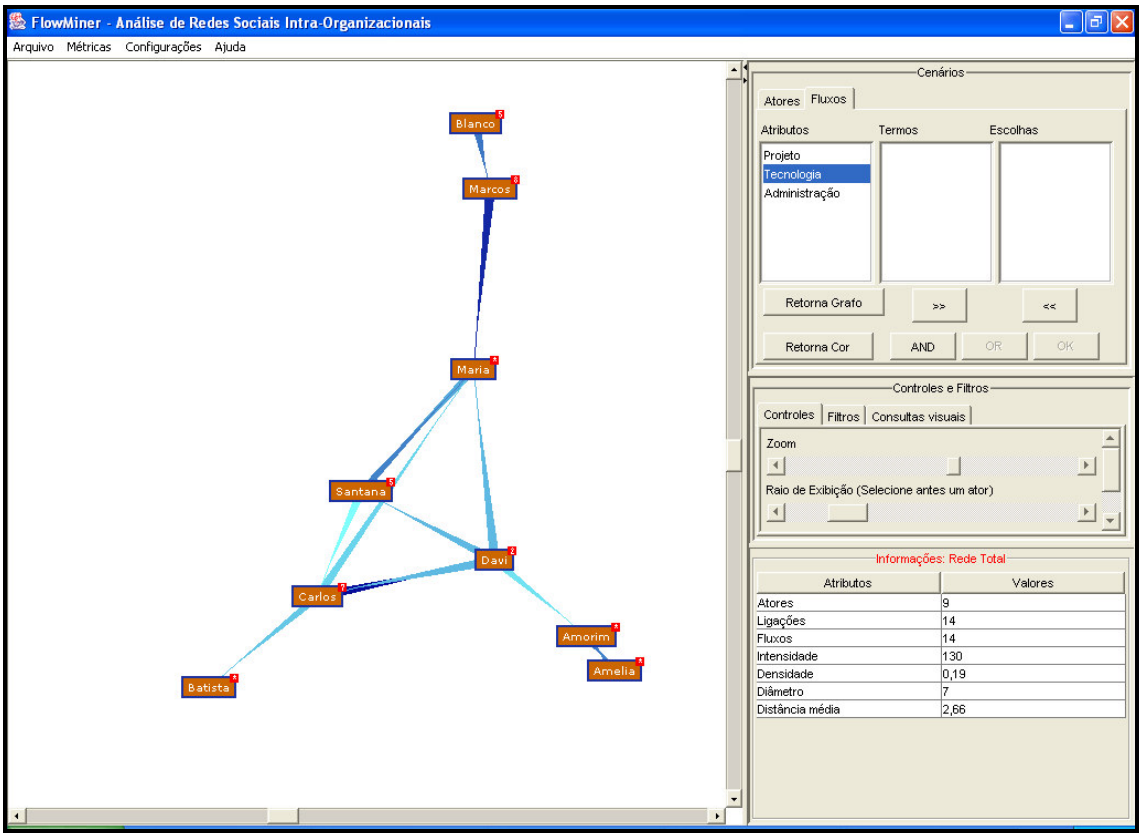

Figura 43. Resultado na cena visual apresentando apenas os elementos selecionados

5.7.4 Módulo de Controle de Interações

Interações com a estrutura visual facilitam a exploração do modelo. Existem basicamente três formas de interações na terceira área da tela principal que estão representadas pelas abas: *Controles, Filtros e Consultas Visuais*. Ver Figuras 44 e 45.

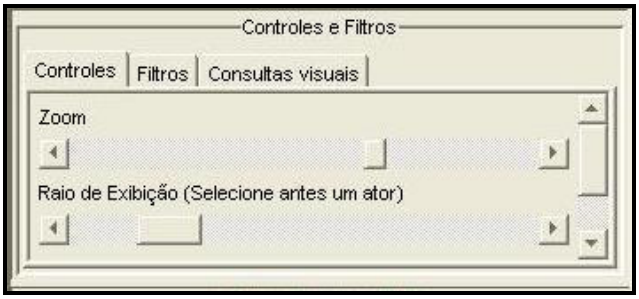

Figura 44. Foco no painel de *Controles* e *Filtros*

Usando estas opções o usuário poderá aplicar controles visuais para melhorar a visualização do grafo, filtros que identificam atores por quantidade de conexões e arestas por quantidade de fluxos trafegando e consultas visuais para visualmente observar a distribuição de valores dos atributos dos atores e das ligações. A Figura 45 apresenta os três casos de uso dos controles de interação.

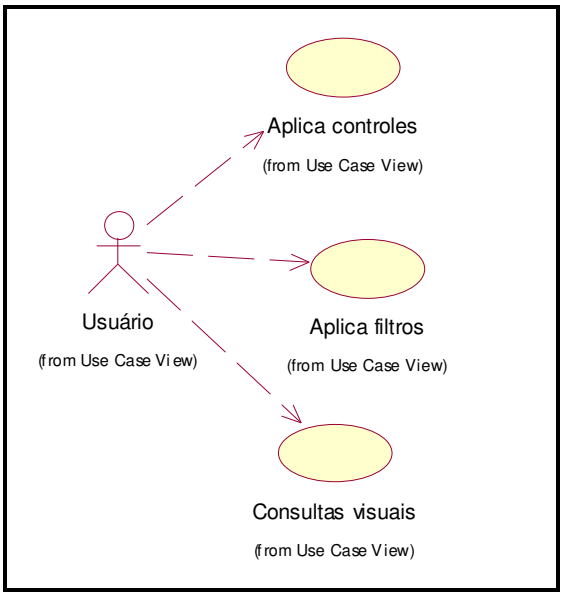

Figura 45. Diagrama de caso do uso de *Controles de Interação*

5.7.4.1 Aba de Controles

Na opção *Controles* existem funcionalidades que permitem modificações visuais, mas não estruturais, no grafo apresentado. A seguir serão descritas as funcionalidades implementadas nesta opção:

1. **Controle de aproximação (Zoom)**

Permite aumentar ou diminuir o tamanho das arestas distanciando ou aproximando os nós apresentados. Esta função facilita a visualização e entendimento do grafo diminuindo o cruzamento de arestas ou a sobreposição de nós conforme exibido na Figura 46.

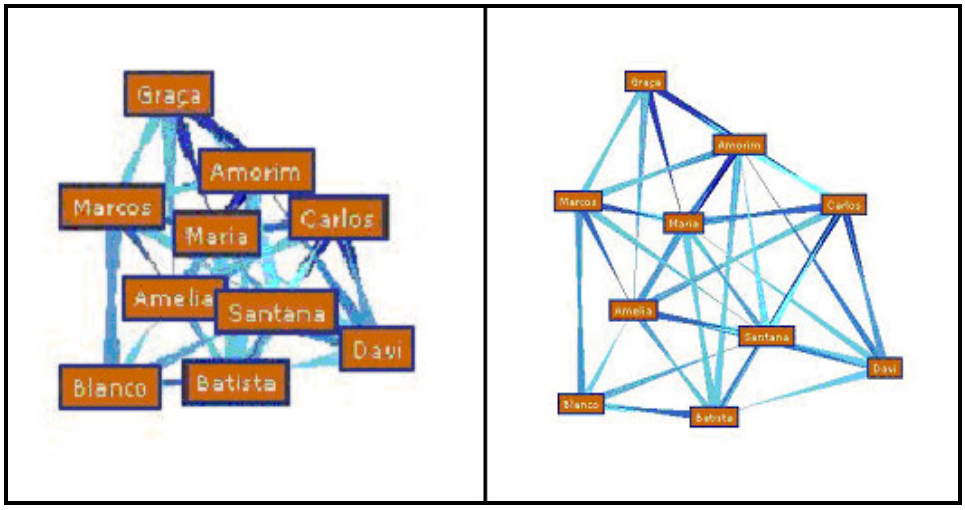

Figura 46. Foco nos grafos antes e depois da utilização do controle de aproximação

O controle de aproximação foi implementado na ferramenta *FlowMiner* através da classe *Controle* usando um controle visual conhecido como barra de rolagem, classe *JScrollBar* do *Java*. Esta barra pode variar de –100 a +100 unidades, ver a barra na Figura 44. A variação da unidade é passada para o método *setZoomValue* da classe *GLPanel* do *TouchGraph.* O relacionamento entre estas classes pode ser visto na Figura 47.

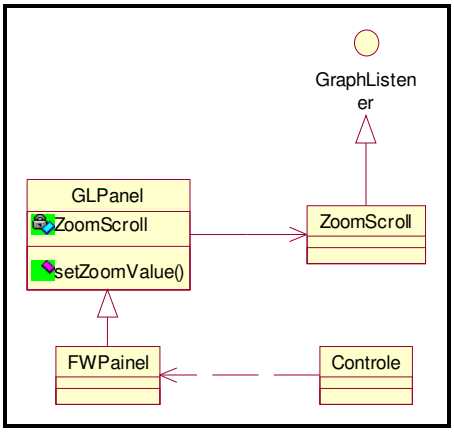

Figura 47. Diagrama das classes que implementam controle de aproximação

2. Controle de Rotação

Este controle permite o usuário girar o grafo em torno de um nó central. O usuário deve centralizar o nó desejado clicando com o *mouse* sobre ele. Para acionar o controle de rotação uma barra de rolagem semelhante à do controle de aproximação faz o valor do ângulo variar de 0 a 359 graus. A Figura 48 apresenta um grafo centralizado no nó Blanco (amarelo) antes e depois da aplicação da rotação de 180 graus.

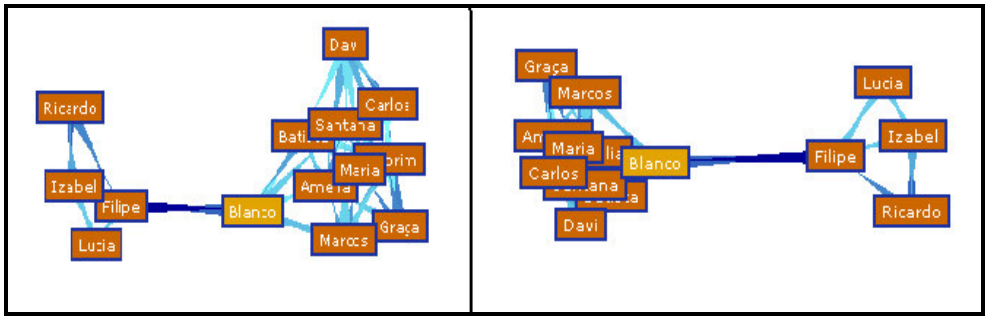

Figura 48. Foco nos grafos antes e depois da utilização do controle de rotação

O diagrama de classes apresentado na Figura 49 mostra que o método *setRotateScroll* também é um método herdado do *TouchGraph* através da classe *RotateScroll* que a classe *Controle* utiliza. As classes *RotateScroll* e *ZoomScroll* implementam a interface *GraphListener*. A chamada do método *setRotationAngle* é feita a partir do evento gerado pela barra de rolagem da mesma forma como acontece no controle de aproximação.

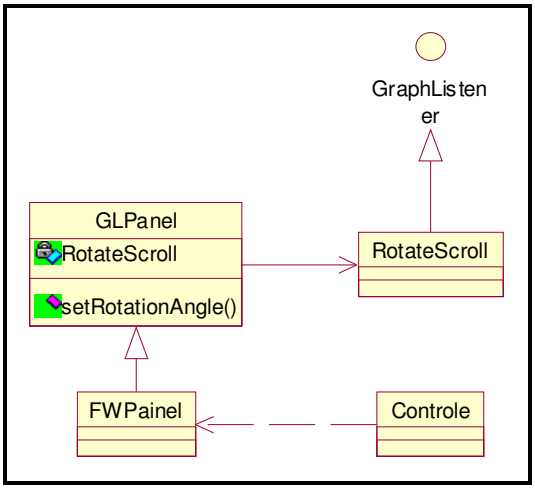

Figura 49. Diagrama das classes que implementam controle de rotação

3. Controle de Raio de Exibição

O controle do raio de exibição tem o mesmo comportamento do de aproximação e rotação. Neste controle também é necessário marcar um nó como o centro do grafo. Após acionar este controle através da barra de rolagem, que fornecerá valores entre um e quatro, serão exibidos apenas os nós cuja distância para o nó central for menor ou igual ao valor informado. Se o valor for zero apenas o nó central aparecerá, se o valor for um, serão mostrados apenas os nós vizinhos ao central, ou seja, nós adjacentes. Se for quatro, maior valor do raio, todos os nós aparecerão na cena visual. Ver Figura 50.

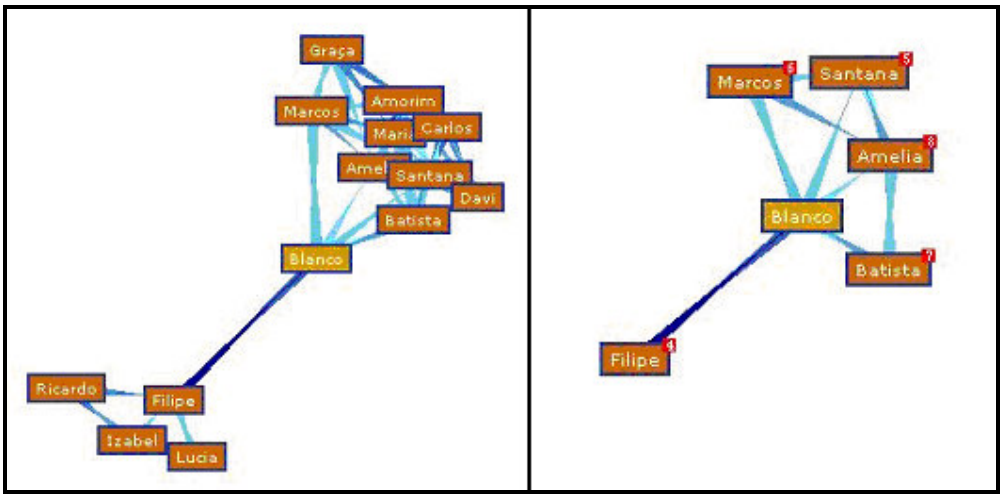

Figura 50. Foco no grafo com raio de exibição quatro e com raio de exibição um.

O diagrama de classes da Figura 51 mostra o uso do método *setLocalityRadius* herdado da classe *GLPanel* que usa o atributo *localityScroll* do tipo *LocalityScroll*

para definir o nível de visualização do grafo. A classe *LocalityScroll* contém métodos que permitem exibir ou ocultar nós a partir da distância entre os nós.

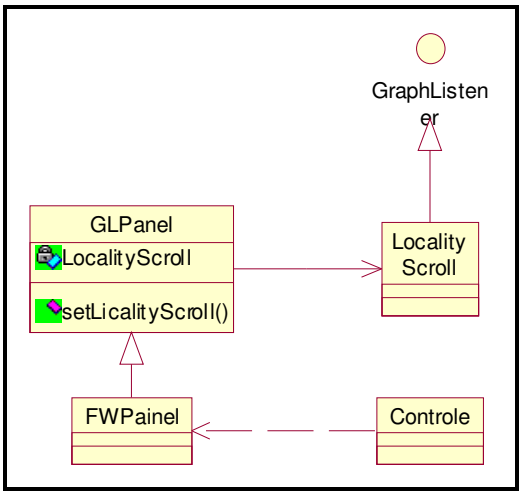

Figura 51. Diagrama das classes do controle de raio de exibição

5.7.4.2 Aba de Filtros

Na aba *Filtros* duas formas de modificações visuais da cena foram introduzidas na ferramenta.Ver na Figura 52 as possibilidades.

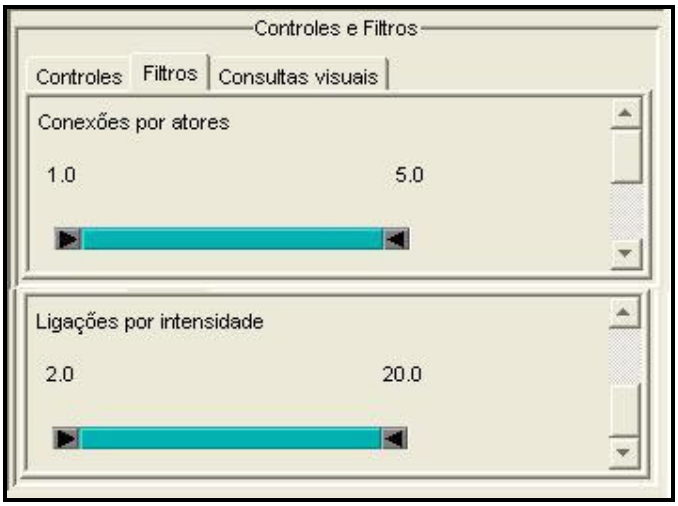

Figura 52. Foco no painel de *Filtros* com opções disponíveis.

Na opção *Conexão por atores* uma barra com dois extremos é disponibilizada informado a quantidade mínima e máxima deste atributo. Ao movimentar a seta da esquerda para a direita o usuário poderá observar a mudança de cor no grafo, para uma cor mais clara, dos atores que possuem quantidade conexões fora do intervalo apresentado. No grafo permanecerão com a cor normal apenas os atores que possuem quantidade de conexões no

intervalo mostrado na barra, inclusive os extremos. Quando movimentando a seta da direita para esquerda observa-se o retorno à cor original dos atores que tiverem quantidade de conexões no intervalo mostrado.

Desta forma o usuário pede variar o controle para fazer clarear e escurecer atores por variação da quantidade de conexões. Na opção *Ligações por intensidade* o mesmo procedimento é executado agora levando em consideração a mínima e a máxima quantidade de mensagens trafegando pelas ligações. Quando movendo o controle da esquerda para a direita o efeito será clarear as ligações com intensidade (quantidade de mensagens) inferior ao intervalo apresentado e movendo da direita para a esquerda fará as ligações voltarem à cor original de acordo com o intervalo. Ver Figura 53.

Através destes filtros o usuário poderá observar que atores são possuidores de muitas conexões, muitos contatos, os que realmente usam o instrumento de comunicação para resolver seus problemas ou ajudar a comunidade, por outro lado poderá perceber pontos na comunidade que precisam ser trabalhados para ampliar a inclusão e participação.

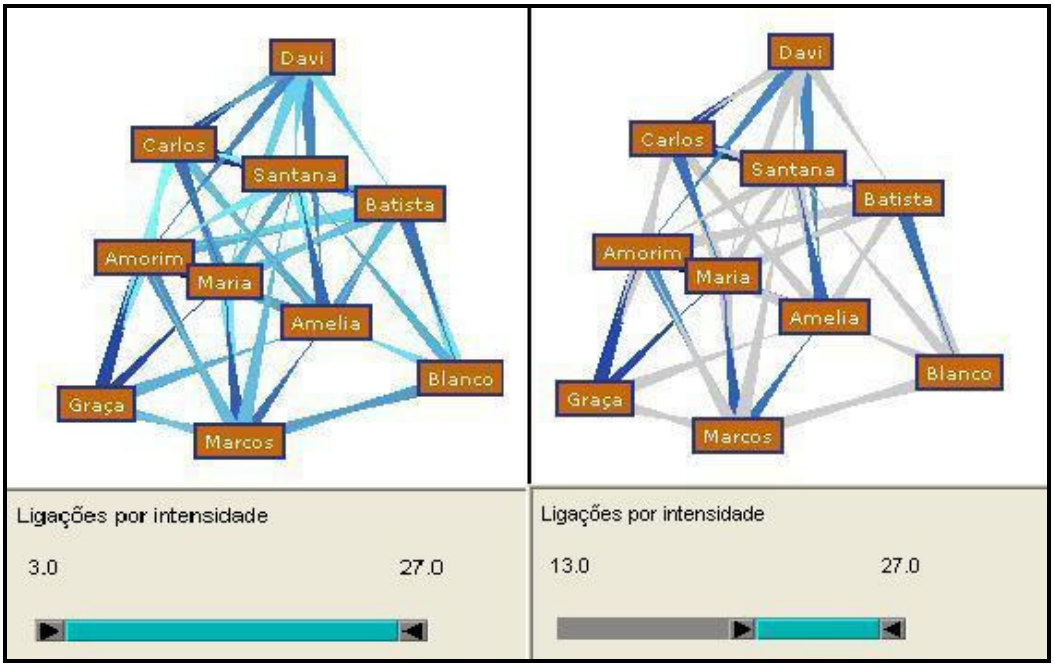

Figura 53. Foco no grafo e no respectivo controle de *Filtros*.

A Figura 53 mostra o efeito no grafo da utilização da opção *Ligações por intensidade.* Nele observa-se a qualidade das conexões através da intensidade das ligações, ou seja, qual a

importância das ligações para os atores, ligações com alta intensidade identificam conexões confiáveis onde ao atores gastam energia por acreditarem no benefício obtido.

As classe *Filtro* implementa este procedimento instanciando a classe *doubleSlideBar* que é responsável pela realização das barras com dois extremos e seus eventos. Eventos acionados nas barras executam métodos de *Filtros* que identificam os objetos e trocam suas cores na cena visual. Ver Figura 54.

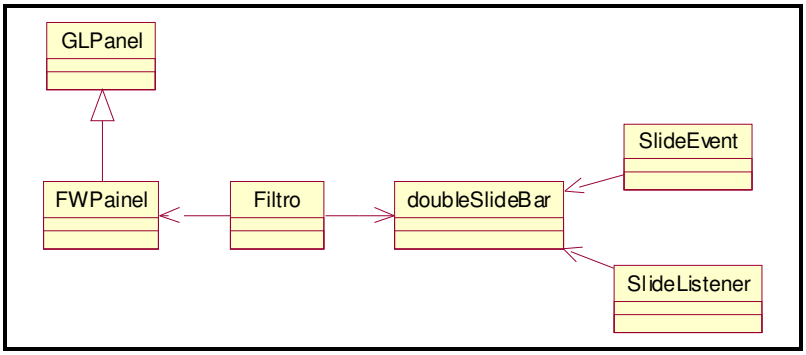

Figura 54. Diagrama das classes de *Filtros*

5.7.4.3 Aba de Consultas Visuais

A aba *Consultas Visuais* mostra a distribuição dos valores dos atributos dos atores e das ligações como na Figura 55.

Figura 55. Foco na aba *Consultas Visuais*

O clique do *mouse* no botão *Atributos Atores* ou no botão *Atributo Ligações* faz surgir um *menu* onde o usuário poderá escolher qual atributo deseja conhecer a distribuição.

A escolha feita pelo usuário fará colorir os elementos do grafo que possuem o mesmo valor de domínio do atributo escolhido com uma cor específica e mostrará na quarta área do grafo uma tabela com as quantidades encontradas para cada valor do atributo selecionado

sincronizando o valor do atributo com a cor dos atores. Na Figura 56 podemos observar o resultado para a escolha atributo *Departamento* para atores.

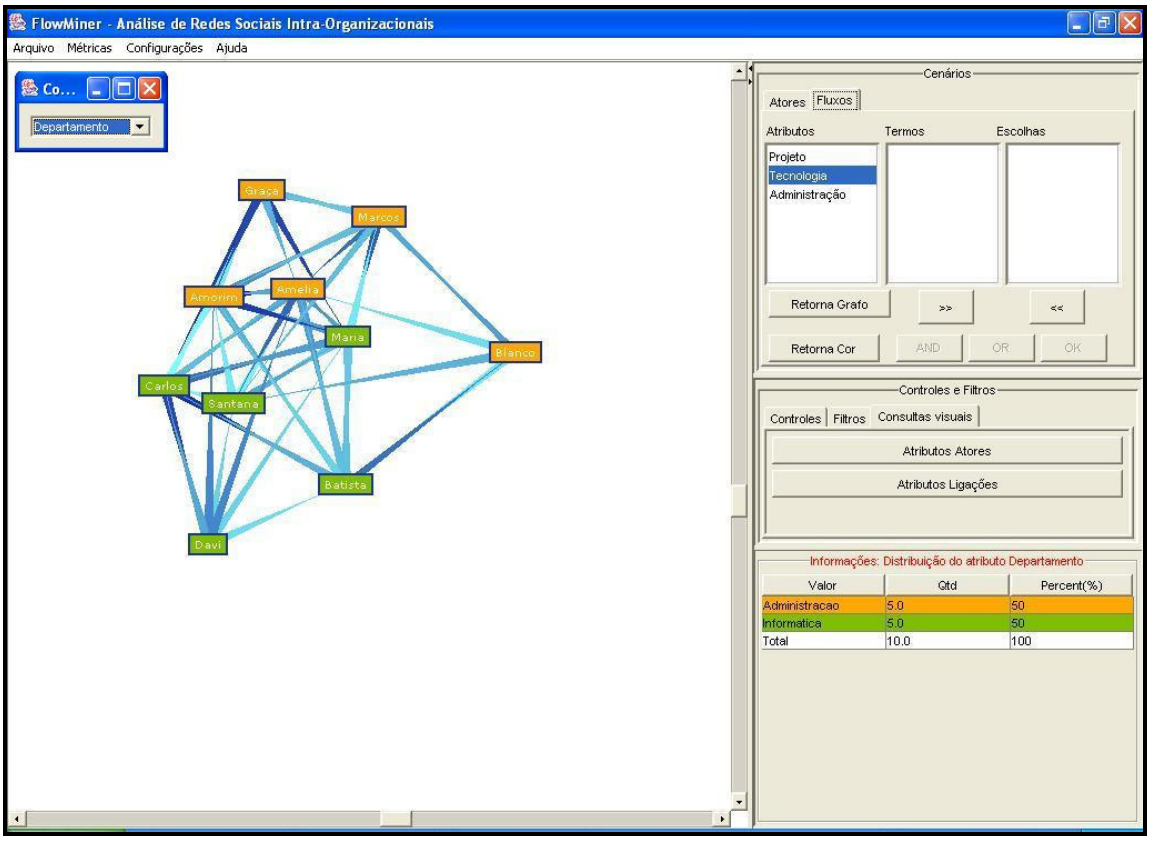

Figura 56. Tela modificada após a seleção do atributo *Departamento* de atores

A classe *Histograma* implementa este procedimento instanciando a classe *FrameAtributo* que é responsável pela realização do *frame* onde será selecionado o atributo. O evento acionado na escolha do atributo executa métodos de *FrameAtributo* que identifica os objetos e troca suas cores na cena visual, além de acionar métodos em *EsquemaTabela* para preenchimento da respectiva tabela na quarta área da tela principal. Diagrama de classes na Figura 57.

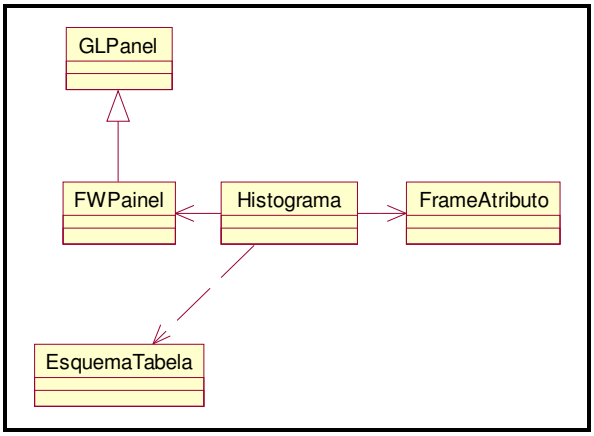

Figura 57. Diagrama da classe *Histograma*

5.7.5 Módulo de Métricas de rede social

Ao longo de anos de estudo e pesquisas diversas métricas foram desenvolvidas para analisar e avaliar redes sociais. As principais métricas de *SNA* foram discutidas na contextualização teórica deste trabalho e apresentadas no capítulo 3. Para implementação e utilização neste trabalho foram escolhidas as métricas mais básicas e que são freqüentemente citadas na literatura. Em especial as métricas que trariam significado se aplicadas a redes intra-organizacionais. As funções implementadas na ferramenta permitem:

- Analisar a rede como um todo, levando em consideração métricas da estrutura global da rede;
- Analisar o papel de cada ator na rede, levando em consideração medidas de centralidade;
- Analisar uma rede egocêntrica. Esta rede apresenta a estrutura da rede individual de um ator específico.

Na Figura 58 pode-se observar o diagrama de casos para utilização de métricas na ferramenta *FlowMiner*.

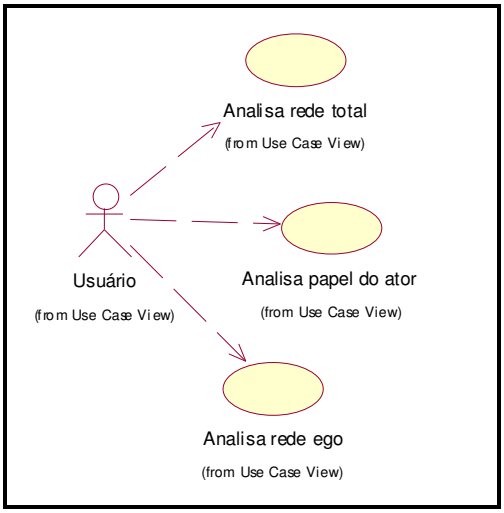

Figura 58. Diagrama de caso de uso do módulo *Métricas*.

Como o desenvolvimento teórico sobre Análise de Redes Sociais já foi explorado no capítulo 3 deste trabalho, aqui serão acrescentados as formas de implementação e os algoritmos de *SNA* utilizados no desenvolvimento da ferramenta.

Funcionalmente a execução das métricas são obtidas pelo usuário utilizando a opção *Métricas* do *MenuBar*. A opção *Métricas* mostra dois itens: *Estruturais* e *Por Ator*.

No item *Estruturais* existe dois subitens *Rede Total,* que mede a rede como um todo e *Rede Egocêntrica*, que a partir da apresentação de um *Menu* permite ao usuário escolher o ator desejado para realizar a visualização de sua rede egocêntrica.

No item *Por Ator* existem mais três subitens *Centralidade*, *Corretor e Elimina não reciprocidade*. Em *Centralidade* é permitida a escolha da métrica de centralidade que será calculada para todos os atores. As métricas de centralidade implementadas são: grau de conexão, grau de fluxo, proximidade e intermediação. Em *Corretor* será calculado para cada ator se sua retirada do grupo fará com que o grupo se divida em subgrupos. A opção *Elimina não reciprocidade* permite a eliminação das ligações que tem apenas uma direção, pois os resultados de algumas métricas podem ser distorcidos por estas ocorrências. A eliminação das ligações não simétricas, deixando apenas as ligações recíprocas corrige estas distorções.

A aplicação de cada uma destas métricas provocará mudanças no grafo, mostrando através da variação de tonalidade da cor a graduação dos valores calculados para cada ator. A tabela da quarta área será preenchida com os valores estatísticos calculados.

A seguir será descrito como foi implementada cada uma destas opções.

5.7.5.1 Análise da rede total

Na ferramenta *FlowMiner* a rede total ou rede como um todo é considerada a partir do subgrupo escolhido pelo usuário para ser analisado. Após o clique do *mouse* no subgrupo escolhido, a tela principal será apresentada com um grafo na primeira área e as medidas da rede total na quarta área. A Figura 59 mostra em detalhe as métricas apresentadas na quarta área do painel principal.

Informações: Rede Total

Atributos	Valores
Atores	10
Ligações	52
Densidade	0,58
Diâmetro	3
Distância média	1,46

Figura 59. Foco na quarta área, com as métricas da rede total.

Nesta tabela são apresentadas as seguintes informações:

- Tamanho da rede – total de atores e ligações. Redes muito grandes se tornam mais complexas para análise;
- Densidade da rede – relação entre o total de ligações existentes, e o total de ligações possíveis. Dependendo da rede, um valor baixo nesta métrica identifica oportunidades de promoção de maior cooperação e solidariedade;
- Diâmetro da rede – maior distância entre dois nós da rede. Este valor significa quantos atores no máximo deverão ser atravessados para se alcançar o ator mais extremo da rede. Valores baixos neste parâmetro identificam redes compactas;
- Distância média – valor médio referente a quantos atores no mínimo um ator tem que atravessar para alcançar um ator específico, um valor alto nesta medida representa uma rede muito hierarquizada com alta intermediação no fluxo de informações. Valores pequenos representam redes com alta coesão entre seus membros.

O usuário tem mais duas formas para obter o preenchimento desta tabela, uma é clicando na área que compõe a fundo do grafo e a outra é utilizando a opção, *Métricas*, *Estruturais*, *Rede Total* da barra de *Menu*.

Para obter e calcular as informações apresentadas na tabela, estruturas como matrizes e lista de adjacências são utilizadas. Três classes são responsáveis pela implementação destas medidas: *ListaAdj, MatrizND e Metricas*, nestas classes existem métodos que permitem a realização de todas estas métricas. No cálculo das distâncias, ou seja, do menor caminho entre dois nós, a ferramenta implementou o algoritmo de Dijkstra (DIJKSTRA,1959), a classe responsável por esta implementação é a *Dijkstra.* Ver diagrama na Figura 60.

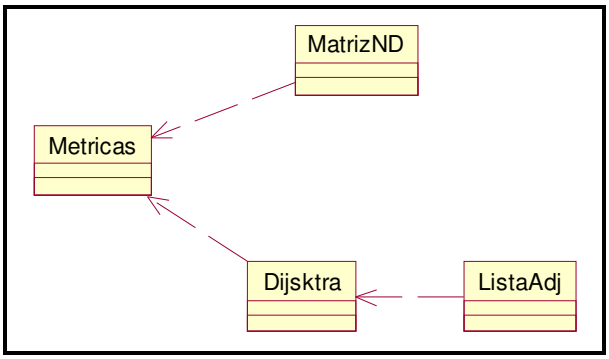

Figura 60. Diagrama das classes utilizadas nas métricas estruturais

5.7.5.2 Análise do papel de cada ator

Neste procedimento procura-se o papel que cada ator desempenha na manutenção da estrutura da rede. A principal medida é a centralidade. Esta medida permite avaliar a importância de um ator na comunidade e de que forma os diversos atores estão interligados. A três medidas principais são: centralidade de grau, centralidade de proximidade e centralidade de intermediação.

A *centralidade de grau* foi implementada usando métodos clássicos da teoria dos grafos, onde através de uma estrutura de matriz ou lista de adjacências são calculados os somatórios das ligações que cada ator tem com os outros (ZIVIANI, 2004).

A *centralidade de proximidade* foi implementada através do somatório das menores distâncias entre um ator e todos os outros. No cálculo das menores distâncias utilizou-se o algoritmo de Dijsktra, já referenciado.

A *centralidade de intermediação* foi implementada através do algoritmo de Ulrik Brandes (2001), devido a sua característica de menor consumo de tempo e espaço. Os algoritmos anteriores gastavam $O(N^3)$ de tempo e $O(N^2)$ de espaço para calcular as centralidades em uma rede de N atores e M ligações, enquanto que o algoritmo de adotado gasta $O(N+M)$ de espaço e executa em $O(N*M)$ de tempo em grafos sem considerar pesos (BRANDES, 2001). A implementação deste algoritmo está representado na ferramenta *FlowMiner* pela classe *Brandes.* No ANEXO A encontra-se o pseudo-código deste algoritmo.

Outra medida que pode ser calculada é a *corretagem.* Esta medida avalia se a retirada de cada ator provoca subdivisão na rede.

As classes envolvidas na implementação destas métricas são: *ListAdj, Métricas, Dijsktra, e Brandes.* A Figura 61 mostra como estas classes se relacionam.

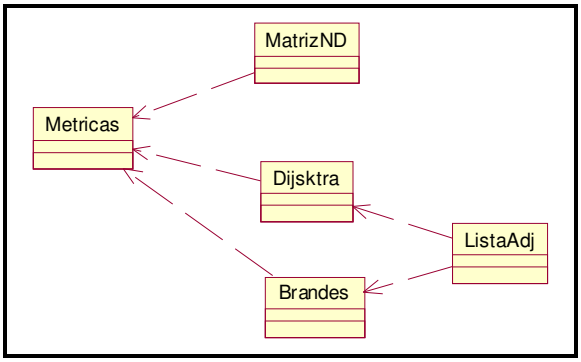

Figura 61. Diagrama das classes envolvidas nas métricas estruturais e por ator

Na atual versão da ferramenta estão implementadas as seguintes métricas de centralidade: *Grau de conexão, Grau de conexão de entrada, Grau de conexão de saída, Grau de fluxo, Grau de fluxo de entrada, Grau de fluxo de saída, Proximidade de envio, Intermediação* e *Corretagem*. Estas métricas podem ser aplicadas no grafo total ou em cenas visuais geradas após consultas por atributos dos atores e das ligações.

A Figura 62 mostra as transformações no grafo através da mudança de tons de vermelho para as métricas: *Grau de conexão, Grau de fluxos, Proximidade de envio* e *Intermediação* respectivamente. As graduações dos tons de vermelho demonstram as diferença de valores obtidos na métrica aplicada a cada ator.

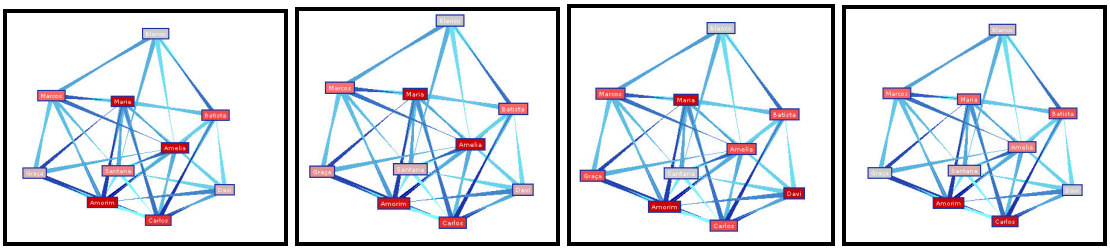

Figura 62. Transformação do papel de cada ator a depender da métrica aplicada.

O efeito sobre a tela principal da aplicação quando executada a métrica *Proximidade de envio* pode ser visto na Figura 63.

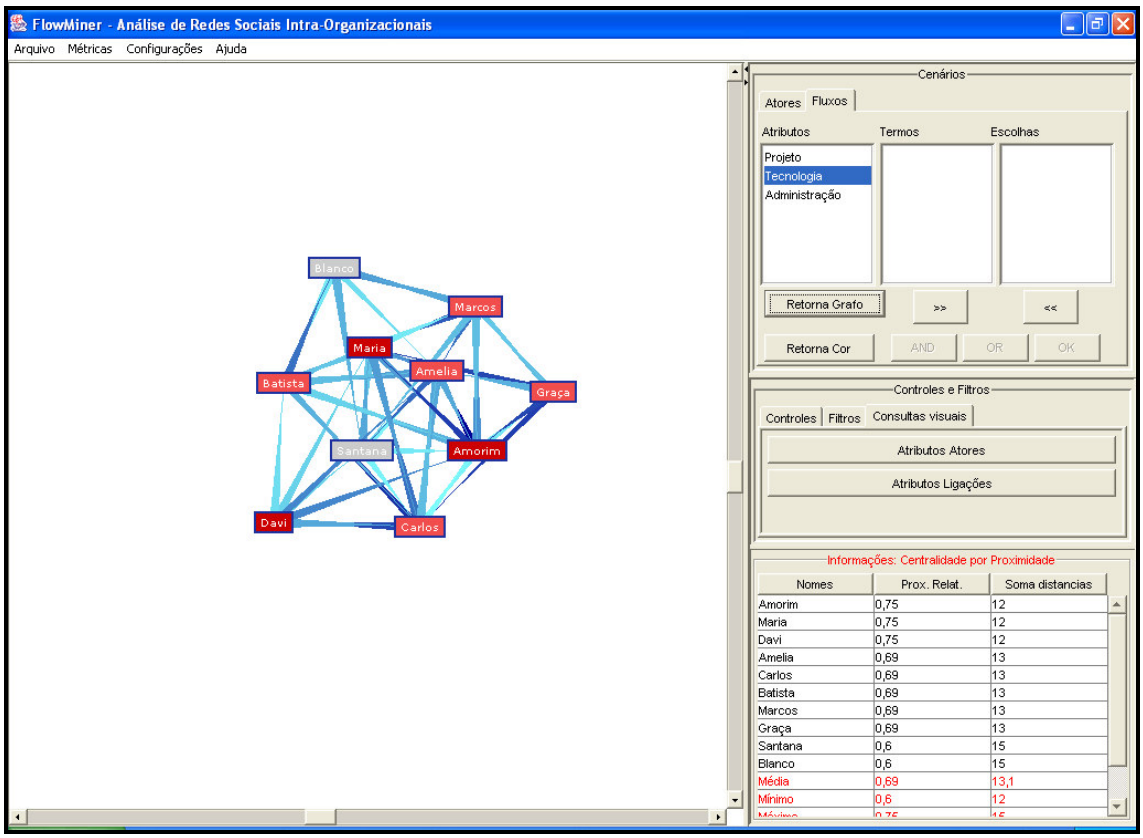

Figura 63. Tela principal após acionada a opção métrica de *Centralidade de Proximidade*.

Nesta tela observa-se que a tonalidade de cada nó varia com o valor da medida calculada. Cores mais fortes representam maiores valores e vice-versa. Na quarta área da tela uma tabela está preenchida com os valores encontrados para a medida de cada ator, trazendo também medidas de valor mínimo, máximo, média, variância e desvio padrão. Na Figura 64 foco na tabela com os valores.

Informações: Centralidade por Proximidade

Nomes	Proximidade	Soma das distancias
Davi	0,75	12
Amorim	0,75	12
Maria	0,75	12
Graça	0,69	13
Batista	0,69	13
Amelia	0,69	13
Marcos	0,69	13
Carlos	0,69	13
Blanco	0,6	15
Santana	0,6	15
Média	0,69	13,1
Mínimo	0,6	12
Máximo	0,75	15
Variancia	0	1,21
Desvio Padrão	0,05	1,1

Figura 64. Foco na tabela preenchida com *Centralidade de Proximidade*

Já a Figura 65 mostra a métrica *Corretagem* e sua atuação em uma cena visual já selecionada. A tabela da quarta área informará por ator em quantos subgrupos sua retirada dividirá a rede. Um clique do *mouse* na linha que representa as informações de um ator, no caso Blanco, promoverá a pintura dos subgrupos que o nó escolhido divide a rede, em cores diferentes.

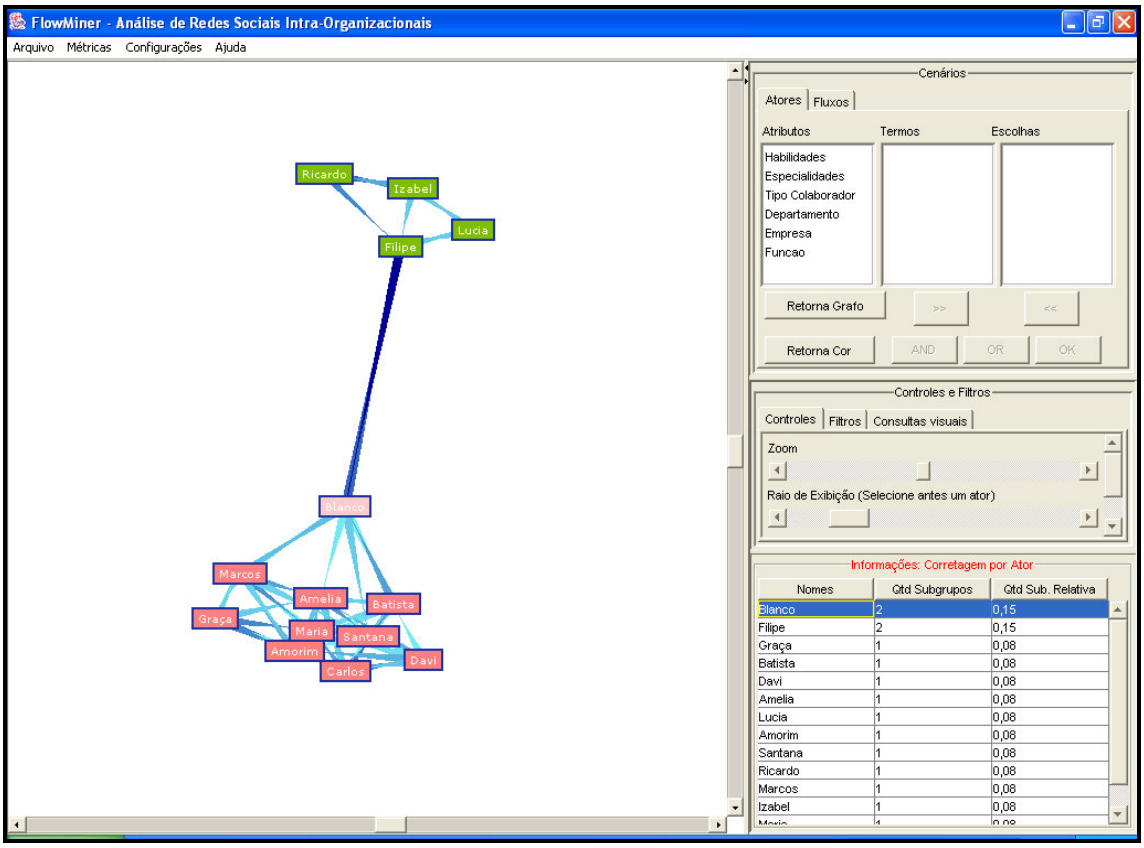

Figura 65. Tela principal mostrando resultado da métrica *Corretagem* pelo ator Blanco.

5.7.5.3 Rede egocêntrica

O papel da rede egocêntrica é mostrar a rede individual de um ator. Nesta rede observam-se os vizinhos de um ator, que ligações seus vizinhos possuem com outros atores e entre eles. A rede individual de cada ator é importante para conhecer sua rede de contatos, entender como a comunidade influi no indivíduo e qual sua chance de obter recursos. A heterogeneidade é uma medida importante, pois revela a variedade da vizinhança, quanto maior variedade maior abrangência nas informações circulando. A composição da vizinhança é outra medida importante, diz respeito à importância de cada vizinho com relação às medidas de centralidade, quanto mais bem posicionados forem seus vizinhos melhor será a rede do indivíduo.

Na ferramenta *FlowMiner* a rede egocêntrica poderá ser obtida de duas maneiras, através do subitem *Rede Egocêntrica*, do item *Estruturais*, da opção *Métricas* da barra de *Menu*, ou do duplo clique em um ator na cena visual. Após um dos dois procedimentos, a cena visual se transformará deixando visível apenas o nó escolhido, seus vizinhos e as ligações entre eles, e uma tabela na quarta área do painel principal mostrará as medidas estruturais desta rede. As métricas executadas são as mesmas da rede total, apenas levando em consideração a rede de um indivíduo.

Para avaliar a heterogeneidade da rede ego, basta verificar os detalhes dos atributos de cada vizinho clicando em cada um deles, ou executar a distribuição de valores do atributo desejado em *Consultas Visuais*. Para avaliar a composição da rede, basta executar as métricas de centralidade que atuarão na rede como um todo revelando a participação de cada vizinho na rede total.

A rede ego foi implementada na classe *FrameEgo que* é uma extensão da classe *JFrame* do pacote *Swing Java.* Esta classe foi criada para permitir ao usuário a escolha de um nó individual. Para visualizar a rede ego após a escolha foi utilizado o método *slowScrollToCenter* da classe *HVScroll* que centraliza o nó selecionado, e o método *setLocaly* da classe *LocalityScroll* que mostra apenas os nós adjacentes do nó central e suas ligações, ambas as classes são do *TouchGraph*. Além disso, métodos específicos foram criados na classe *MostraGrafo* para preparar as estruturas de forma que métricas da rede como um todo atuem apenas sobre os atores da rede ego e as métricas de centralidade atuem sobre todos os

atores da rede. As tabelas são apresentadas com os resultados das métricas usando a classes *ModeloTabela* e *EsquemaTabela.* Ver Figura 66.

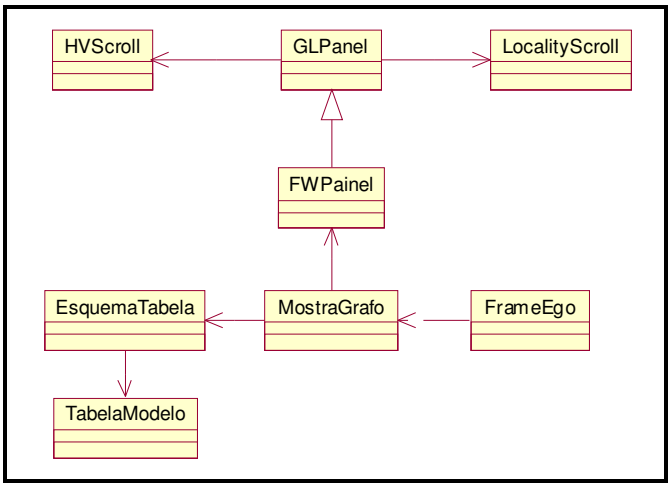

Figura 66. Diagrama das classes envolvidas na *Rede Egocêntrica*

5.7.6 Sobre implementação e reuso

O processo de implementação dos módulos apresentados na Figura 35 e descritos ao longo desta seção foi baseado no conceito de software reusável, que previlegia a reutilização de componentes e funções de produtos de software já existentes (SOMMERVILLE, 2003). Para tal, foram observadas e combinadas experiências anteriores bem sucedidas no desenvolvimento de ferramentas de mineração visual de dados efetuadas no NUPERC, no qual esta dissertação de mestrado foi desenvolvida.

Como fonte principal de reuso de código foram utilizados os trabalhos de Marques Neto (2004) e Almeida (2003), ambos desenvolveram ferramentas de mineração visual de dados utilizando conceitos de consultas interativas, transformação de dados em paradigmas visuais e obtenção de detalhes sob demanda, como discutido na seção 4.2. Apesar do reuso de código dos trabalhos citados, as idéias sobre exploração visual de dados empregados nesta dissertação têm sua origem em trabalhos prévios desenvolvidos por pesquisadores do NUPERC em conjunto com a Universidade de Maryland (SEAMAN e outros, 1999, 2003).

No desenvolvimento da ferramenta *FlowMiner* foram implementadas 43 classes, dentre as quais aproximadamente 80% foram desenvolvidas e 20% foram reusadas das fontes citadas no parágrafo anterior de forma integral ou com algumas modificações para adaptá-las aos objetivos desejados. Além disso, na composição da ferramenta foi incorporada a

biblioteca *TouchGraph* versão 1.22 (TOUHGRAPH, 2004), que é composta de 42 classes, dentre as quais 10% sofreram pequenas alterações para adequá-las às necessidades da aplicação.

É importante realçar que apesar do reuso de mecanismos de visualização, controle de interações e mapeamento de cores, o conceito global implementado na ferramenta *FlowMiner*, isto é, análise visual de comunicação organizacional é original. Os trabalhos mencionados anteriormente trataram de análise visual de repositório de documentos usando diagramas de dispersão (SEAMAN e outros, 1999, 2003), análise visual de dados hierárquicos (ALMEIDA, 2003) e análise visual de dados armazenados em bases relacionais (MARQUES NETO, 2004).

5.8 Considerações Finais

FlowMiner é uma ferramenta voltada a visualização de redes sociais desenvolvida como parte desta dissertação. Utiliza como metáfora visual um grafo que tem como objetivo montar uma representação da rede de comunicação entre colaboradores de uma organização. Esse grafo é construído de forma gráfica e navegacional a partir de mecanismos de acesso a uma base de dados específica.

Seguindo a tendência da maioria das ferramentas de visualização baseada em grafos, *FlowMiner* foi construída usando algoritmos de *força-dirigida*. Como base para a construção desse tipo de algoritmo foi utilizado o arcabouço *TouchGraph.*.

As principais contribuições da *FlowMiner* em relação às outras ferramentas de *SNA* estão na utilização de banco de dados como entrada das informações, a possibilidade de atribuir semântica aos objetos representados e a utilização de vários controles de visualização que facilitam o entendimento e a capacidade de inspeção dos dados. Através desses controles visuais é possível:

- Ocultar ou exibir objetos, permitindo construção de cenas variadas usando consultas por atributos;

- Aplicar métricas de análise de redes sociais sobre as cenas geradas. A partir delas refletir mudanças de cores nos objetos e mostrar tabelas estatísticas para apresentar os resultados obtidos;

- Filtrar através de mudança de cores os atributos quantitativos (numéricos) dos objetos;
- Exibir sob demanda os valores dos atributos dos objetos da cena visual;
- Manipular o grafo, mudando o posicionamento de seus objetos para melhorar a visualização.

Neste capítulo discutiu-se a implementação da ferramenta, suas funcionalidade, os principais artefatos produzidos e suas contribuições. A fundamentação teórica para a sua construção foi apresentada nos capítulos 2, 3 e 4. No próximo capítulo será introduzido o estudo de caso desenvolvido com o objetivo de avaliar a aplicabilidade da ferramenta.

CAPÍTULO 6 - ESTUDO DE CASO

Este capítulo aborda as características operacionais da ferramenta, assim como sua avaliação através da realização de um estudo de caso. O estudo foi selecionado no contexto empresarial e envolveu uma empresa de software e consultoria de TI, a ADN Tecnologia de Sistemas Ltda, que se caracteriza pelo desenvolvimento de projetos tecnológicos.

O principal objetivo dos testes realizados durante o desenvolvimento do estudo de caso, foi a avaliação da ferramenta como instrumento para o entendimento e análise do comportamento e de padrões de comunicação da comunidade estudada como apoio ao processo de gestão do conhecimento.

Neste capítulo serão caracterizados o estudo de caso e a organização envolvida, para isso são descritos: o domínio da aplicação, os dados envolvidos, o processo de desenvolvimento, resultados obtidos e como foram comparados com os da ferramenta *UCINET*. Ao final deste capítulo é feita uma avaliação geral da ferramenta *FlowMiner* apontando seus pontos fortes, fracos e melhorias identificadas ao longo do processo.

6.1 Caracterização do Estudo de Caso

Nesta seção serão descritos a metodologia utilizada e o desenvolvimento do estudo de caso que envolveu levantamento, coleta das informações, análise e comparação dos resultados.

O levantamento e escolha da comunidade a ser estudada surgiram do meu conhecimento da organização, onde trabalhei durante cinco anos e meio a partir de meados de 1996. A necessidade era identificar uma comunidade que executasse a comunicação através

de uma ferramenta automatizada para poder analisá-la e conhecer seus padrões de comportamentos. A escolha recaiu sobre a ferramenta *HelpDesk.* O *HelpDesk* é um sistema criado pela ADN que executa em ambiente *Intranet* e *Extranet.* Desenvolvido em *ASP* (*Active Server Page*) da Microsoft com uma base de dados armazenada em um servidor que utiliza o sistema gerenciador de banco de dados *MS SQL Server.* O *HelpDesk* é utilizado por funcionários da área de TI dos clientes e por desenvolvedores da ADN para comunicar e resolver dúvida ou problema envolvendo a manutenção de sistemas criados ou integrados pela ADN. A escolha da ferramenta *HelpDesk* se baseou em algumas características importantes para a viabilização da utilização da ferramenta *FlowMiner:*

- A comunidade escolhida executa sua comunicação através de uma ferramenta automatizada;
- Toda conversação sobre o assunto está registrada e armazenada em bases de dados;
- Os registros dos históricos dos chamado de esclarecimento de dúvidas ou resolução de problemas já estão categorizados através de uma taxonomia criada pela organização;
- Todos os registros de comunicação se referem a assuntos técnicos de acesso coletivo, não existindo nenhum problema de privacidade ou violação de informações sigilosas.

O sistema *HelpDesk* não possui informações sobre o perfil dos atores envolvidos, foi necessário então definir os atributos necessários ao *FlowMiner* com a ajuda dos coordenadores e analistas, que acompanharam o estudo de caso, e extraí-los de outras bases de dados disponíveis na organização.

6.1.1 Caracterização da Organização

A ADN é uma organização brasileira de médio porte com 150 profissionais, entre funcionários e consultores. Sediada em Salvador, com uma filial em São Paulo, atua na área de Tecnologia da Informação desde 1990, prestando consultoria, desenvolvendo projetos tecnológicos e fornecendo produtos e serviços em informática. Dentre estes serviços estão o suporte a ambientes computacionais e a integração e comercialização de aplicativos.

A ADN tem por objetivo prestar, de forma integrada, todos os serviços requeridos à gestão e otimização da TI nas empresas. Propõe-se a prover soluções completas que atendam toda a cadeia de agregação de valor ao negócio, desde os processos de retaguarda até os processos de relacionamento com os clientes. Ela está estruturada para atuar em diversas linhas de negócio relacionadas ao desenvolvimento de projetos de tecnologia de informação, desenvolvimento de aplicações para o mercado e soluções de informática baseadas em integração com softwares de mercado.

Nos projetos de manutenção de sistemas a ADN atua com uma equipe de apoio local por cliente para o suporte de primeiro nível, caso este serviço seja contratado ou o primeiro nível fica a cargo da própria equipe de TI do cliente. Uma equipe centralizada na Matriz é responsável pelo apoio de segundo nível às equipes locais da ADN no cliente ou à equipe TI do cliente. O sistema de *HelpDesk* funciona como instrumento de comunicação e registro de todas as ocorrências de esclarecimento de dúvidas e solução de problemas de manutenção dos sistemas entre equipe de TI do cliente ou equipe de apoio local e equipe da matriz. Uma ocorrência é denominada *chamado*, ou seja, um chamado é um registro de um esclarecimento de dúvida ou a descrição de um problema de sistema para solução. A equipe da Matriz conhecida de Retaguarda é dividida em grupos por sistema ou cliente dependendo do tamanho e complexidade e para cada grupo existe um coordenador. Através do *HelpDesk* os coordenadores dos grupos da equipe da Matriz acompanham e participam do processo aprovando e direcionando chamados para os analistas responsáveis pela resolução.

6.1.2 Metodologia adotada

Uma vez estabelecida a fonte de informações, este trabalho foi conduzido em três etapas. Na primeira etapa foi desenvolver a geração da base de dados a ser utilizada pela ferramenta *FlowMiner. N*esta base foram registrados os atores e seus atributos, bem como as comunicações realizadas durante um período escolhido, de forma categorizada e consolidada. Na segunda etapa, foi realizada uma análise dos dados utilizando os recursos disponíveis na ferramenta *FlowMiner* já discutidos no capítulo 5. Na terceira e última etapa, os resultados foram apresentados a dois coordenadores da ADN para interpretação e avaliação, nesta etapa os dados foram também lançados em forma de planilha no software de análise de redes sociais *UCINET 6.0* para checar as métricas obtidas. Os resultados das métricas do software *UCINET* constam do ANEXO B. Nas próximas seções serão descritas em detalhes as etapas da metodologia.

6.1.3 Geração da base de dados

As informações envolvidas neste estudo são provenientes das bases de dados dos sistemas *HelpDesk*, *Recursos Humanos*, *Folha de Pagamento* e *Clientes*. Os dados coletados foram extraídos destas diversas fontes e armazenados em uma área de transição para possibilitar a execução das etapas de seleção e pré-processamento. Os nomes dos colaboradores e dos clientes foram trocados ou generalizados por questões de preservação de privacidade.

Inicialmente foi feita uma análise das informações existentes na organização que poderiam compor os atributos dos atores. Os atributos definidos para compor os atores foram: nome, apelido, empresa, departamento, função, atuação e qualificação.

O atributo atuação representa o papel que o ator desempenho no processo de manutenção de sistemas. No sistema de *HelpDesk*, os papéis são: usuário, apoio, coordenador ou desenvolvedor. Um usuário apenas registra solicitações ou chamados, o apoio esclarece dúvidas e registra chamado, o coordenador avalia a necessidade do chamado vindo do usuário ou apoio e o repassa para o desenvolvedor responsável por resolvê-lo, o desenvolvedor pode repassá-lo para outro ou resolver o chamado. A resolução do chamado pode implicar em alterações no código do sistema em questão.

O atributo qualificação assinala a escala de qualificação por tempo e nível da função do ator, podendo ser sênior, pleno, júnior e *trainee*. Os outros atributos são auto-explicativos.

A amostra dos atores foi selecionada da utilização do *HelpDesk* durante aproximadamente quatro meses, no período de 15 de março a 12 de julho de 2005. Esta amostra representa os principais técnicos envolvidos na manutenção de sistemas: usuários, apoios, desenvolvedores e coordenadores da Matriz, envolvendo desta forma todas as classes de atores que participam da manutenção de sistemas.

Na seleção foram extraídos 67 atores. A etapa de pré-processamento fez a transformação dos dados para adequá-los à estrutura da base local da ferramenta *FlowMiner*. Esta etapa englobou a complementação dos atributos de cada ator através da combinação das bases de dados utilizadas no sistema de *Recursos Humanos* e *Folha de Pagamento*.

Após a inserção dos atores e seus atributos foram carregadas as informações sobre as comunicações mantidas entre eles no período estudado.

No sistema *HelpDesk* existem duas tabelas que registram esta comunicação: *Chamado* e *HistoricoChamado*. Os chamados estão classificados por tipo e causa. Esta categorização foi também adotada pelo sistema *FlowMiner* para o conteúdo transacional, ou seja, as mensagens trocados entre os atores da comunidade. Foram inseridos então, como atributos dos fluxos, as seguintes categorias: Suporte técnico, Esclarecimento de dúvidas, Manutenção corretiva fora garantia, Manutenção corretiva em garantia, Manutenção evolutiva, Evolução do produto, Manutenção legal, Orçamentação, Chamado indevido e Retorno de erro. Para cada um deste tipo de causa existem subclassificações de descrição da causa que foram transformadas em valores ou domínios destes atributos.

Um chamado do sistema *HelpDesk* é auditado numa tabela conhecida como *HistoricoChamado*. Nesta tabela está armazenado o trâmite de todos os chamados, ou seja, os atores por onde o chamado passou até ser resolvido. O próximo passo foi transformar o chamado e seu histórico em várias mensagens. Cada ator por onde o chamado transitou se torna uma mensagem com ator origem, ator destino, tipo de causa e descrição da causa. O objetivo é identificar o caminho que o chamado trafegou. Observar que neste momento o caminho se transformará em diversas mensagens todas com a mesma identificação de causa e descrição da causa, mudando apenas os atores por onde trafegou. A partir da criação das mensagens é possível identificar as ligações (arestas) e os fluxos que são os objetos a serem carregados nas bases do *FlowMiner*.

As ligações ou arestas são obtidas através das mensagens, se existir pelo menos uma mensagem entre dois atores quaisquer, a combinação ator origem e ator destino se transformará em uma ligação, levando em consideração a direção, ou seja, uma dupla de atores poderá no máximo ter duas arestas, uma em cada direção. Resumindo, uma ligação é estabelecida entre um ator origem e destino se pelo menos uma mensagem for encontrada entre eles.

Um fluxo é criado para cada grupo de mensagens de mesma ligação (que corresponde a ator origem e ator destino), atributo (que corresponde a tipo da causa) e valor (que corresponde a causa). A quantidade de mensagens encontrada por ligação, atributo e valor se transforma na intensidade do fluxo.

O menor grau de detalhe do *FlowMiner* é o fluxo, que tem como informações a ligação por onde passa, o atributo e valor que o categoriza e a intensidade que corresponde ao somatório das mensagens encontradas. Exemplo: se entre o ator 1 e 2 foram emitidas 10 mensagens de *Esclarecimento de Dúvidas* (atributo) sobre *Banco de Dados* (valor) e 5 mensagens de *Suporte Técnico* (atributo) sobre *Performance* (valor), a transformação para o *FlowMiner* será a inclusão de um registro de ligação entre ator 1 e 2 e a inclusão de dois fluxos para esta ligação, um fluxo de *Esclarecimento de Dúvida* sobre *Banco de Dados* com intensidade 10 e outro de *Suporte Técnico* sobre *Performance* com intensidade 5.

A geração das bases de dados do *FlowMiner* englobou as operações de: exclusão de atributos ou registros não relevantes, redução de escala para a transformação de campos numéricos ou alfanuméricos em campos categóricos e classificação por categorias para inclusão de consolidações. Após a etapa de pré-processamento, as informações puderam ser inseridas na base de dados local da ferramenta *FlowMiner* através de funcionalidades de importação automática disponíveis nas ferramentas de banco de dados.

6.1.4 Exemplos de Uso da Ferramenta na ADN

Após a criação da base de dados como descrito na seção anterior foi iniciado o uso da ferramenta *FlowMiner* com a finalidade de analisar o comportamento da comunidade estudada.

Como primeira tela a ferramenta apresenta um grafo mostrando se a comunidade está ou não dividida em subgrupos disjuntos que são subgrupos que não se comunicam. Neste caso a comunidade possui apenas um subgrupo composto de 67 atores e 323 ligações. Ver Figura 67.

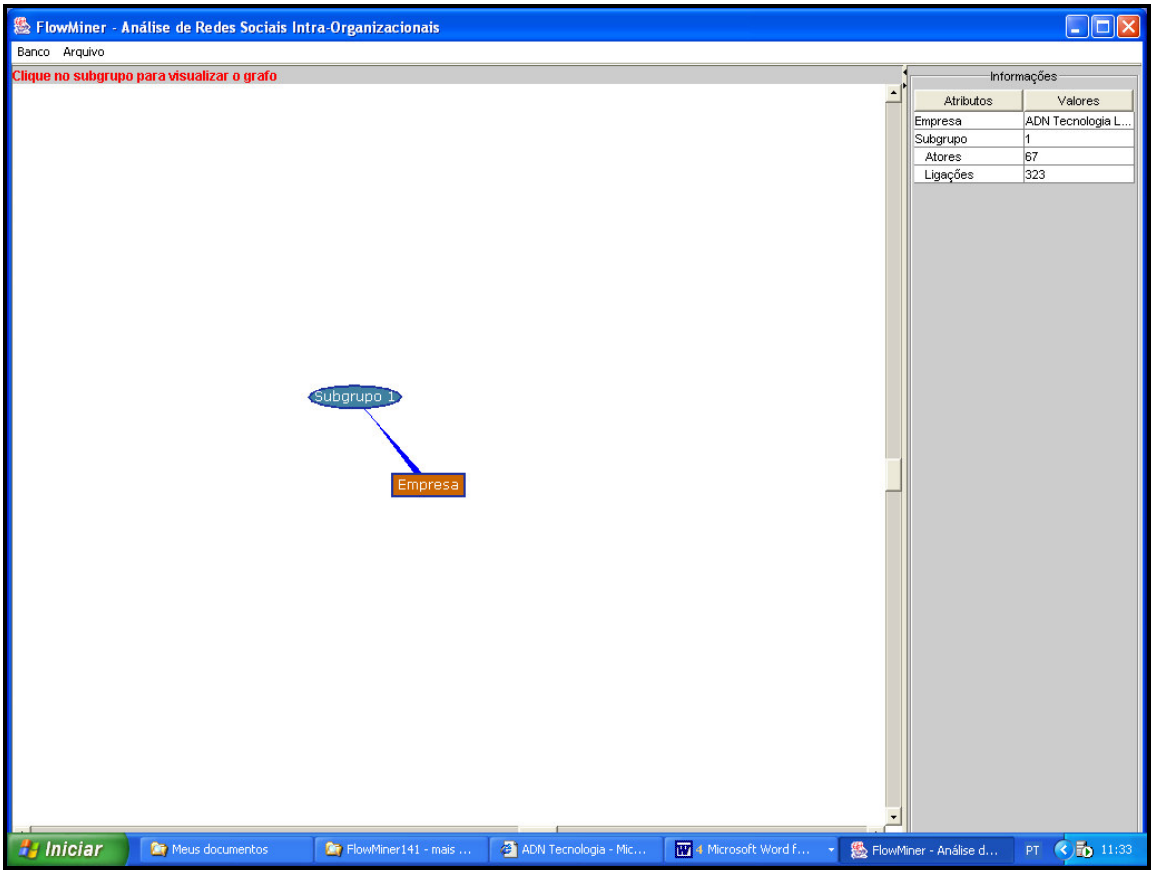

Figura 67. Tela inicial com a composição de subgrupos da comunidade.

Após a escolha do subgrupo, a tela principal é apresentada com o mapa da rede como um todo. Observar que os nomes dos atores que aparecerão nas telas a seguir não correspondem à realidade, eles foram modificados para proteger a privacidade. Ver Figura 68.

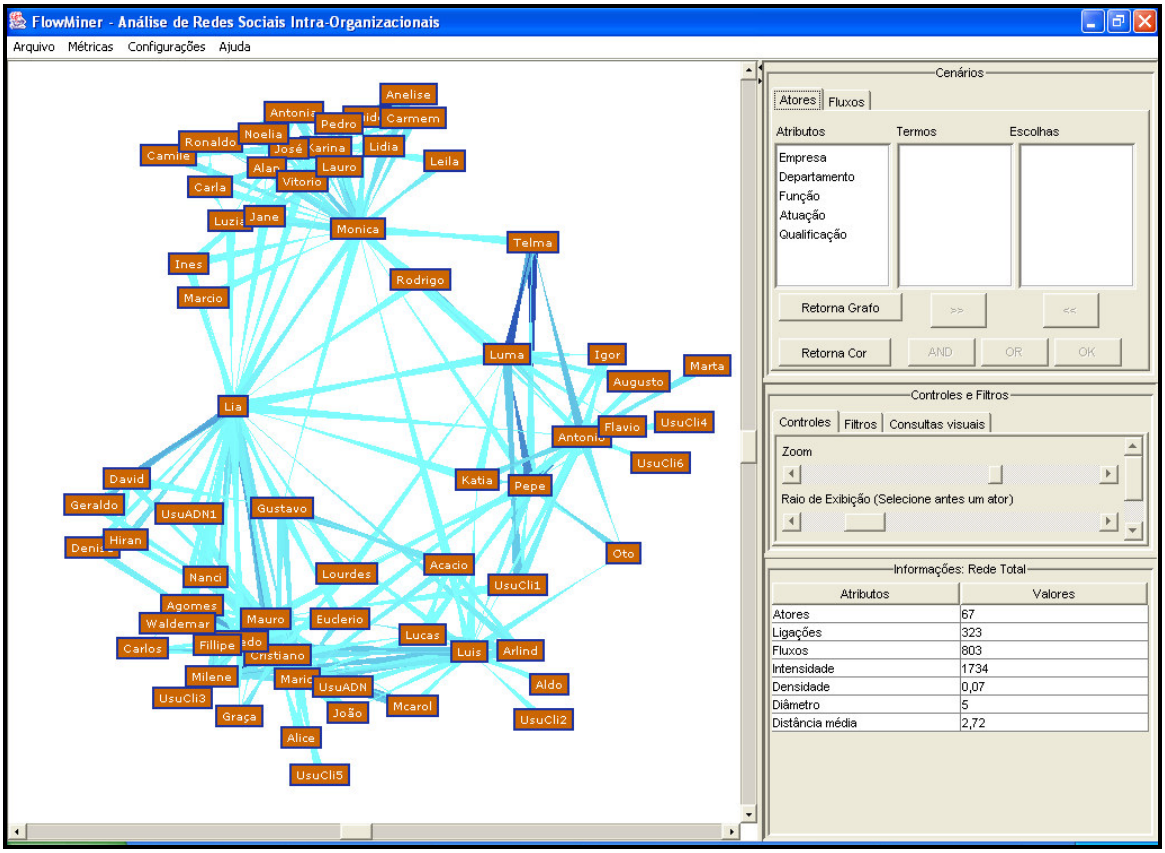

Figura 68. Tela principal mostrando a rede total

6.1.4.1 Análise da rede como um todo

Esta análise fornece uma visão global da rede estudada, mostrando a estrutura total de comunicação entre seus atores. As métricas informam números gerais sobre os elementos analisados e as principais características estruturais da rede.

No mapa apresentado podemos observar que a rede possui 67 atores, interligados por 323 ligações por onde trafegam 803 fluxos de intensidade total 1 734. A soma das intensidades neste caso demonstra o total de mensagens encontradas, 1 734 mensagens foram consolidadas por ligação, atributo e valor em 803 fluxos.

Apesar do mapa dar a sensação de que existe um grande número de ligações a densidade calculada demonstra que apenas 7% do potencial da rede está sendo utilizado. Isto se justifica, pois um chamado na maioria das vezes é submetido pelo usuário para o coordenador do grupo ao qual está inserido, por sua vez o coordenador o encaminha

normalmente para os desenvolvedores do seu grupo. A divisão destes grupos se dá normalmente por cliente ou por sistema atendido.

A distância média apresentada 2,7 mostra uma intermediação média de aproximadamente dois contatos a mais para alcançar qualquer outro na rede. Justificado pela característica de que a maioria dos chamados passa pelo coordenador, para depois ser encaminhado ao desenvolvedor responsável pela solução, que pode ainda em alguns casos repassá-lo para outro desenvolvedor. Isto demonstra uma rede hierárquica, onde os coordenadores concentram os chamados para distribuição.

O diâmetro da rede para os atores alcançáveis foi calculado em cinco ligações, ou seja, a distância máxima a separar quaisquer dois atores nesta rede não é superior a cinco contatos. Este diâmetro demonstra uma rede pouco conectada.

O mapa identifica também as ligações que possuem maior intensidade de fluxo através da coloração mais forte de suas ligações. Fazendo uma aproximação através dos controles de zoom e clicando na ligação podemos observar detalhes sobre este elemento.

Na Figura 69 a ligação na direção de Milene para Cristiano apresenta a maior intensidade do grafo, significando que 71 mensagens trafegaram de Milene para Cristiano. No detalhe da ligação, observamos os 13 fluxos categorizados por atributos com suas respectivas intensidades, estes fluxos compõem as 71 mensagens. Mais de 49% das mensagens dizem respeito a Esclarecimento de Dúvidas, demonstrando que Michele cuja atuação é de apoio está repassando mensagens que deveria estar resolvendo, para um coordenador da equipe Matriz.

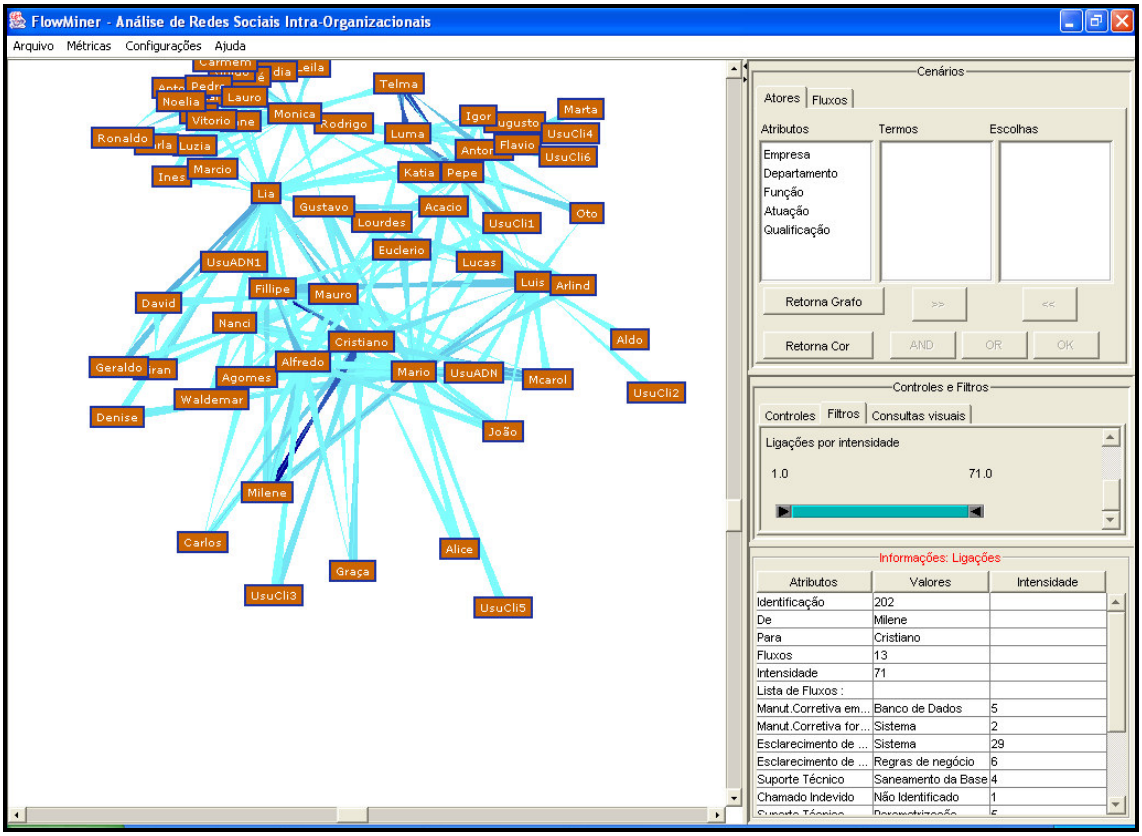

Figura 69. Aproximando e solicitando detalhe da ligação de maior intensidade

6.1.4.2 Análise por ator

Nesta análise serão identificados os personagens responsáveis pela dinamização da rede, aqueles indivíduos que ocupam posições estratégicas. Na seção 3.4 deste trabalho foram definidos os papéis de conectores centrais, corretores de informações e expansores de fronteiras, que segundo Cross e Prusak (2002) são os atores críticos para o funcionamento da rede. A análise da rede através das métricas de centralidade dos atores permite atribuir aos atores seus respectivos papéis. Calcular a centralidade de um ator significa identificar a posição em que ele se encontra em relação às trocas de informações, o que traz consigo a idéia de poder (MARTELETO, 2001).

Na Figura 70 foi executada a opção *Métrica – Por Ator – Centralidade – Grau conexão* e como resultado os atores Lia, Mônica, Cristiano, Mauro e Mario são pontuados como os atores que possuem maior número de conexões ou ligações. Estes mesmos atores também aparecem se revezando nas maiores pontuações das opções *Grau de conexões de entrada* e *Grau de conexão de saída*, mostrando que muitos atores enviam e recebem

informações deles. Lia, Mônica, Mario e Cristiano atuam como coordenadores e Mauro é um desenvolvedor da equipe Matriz.

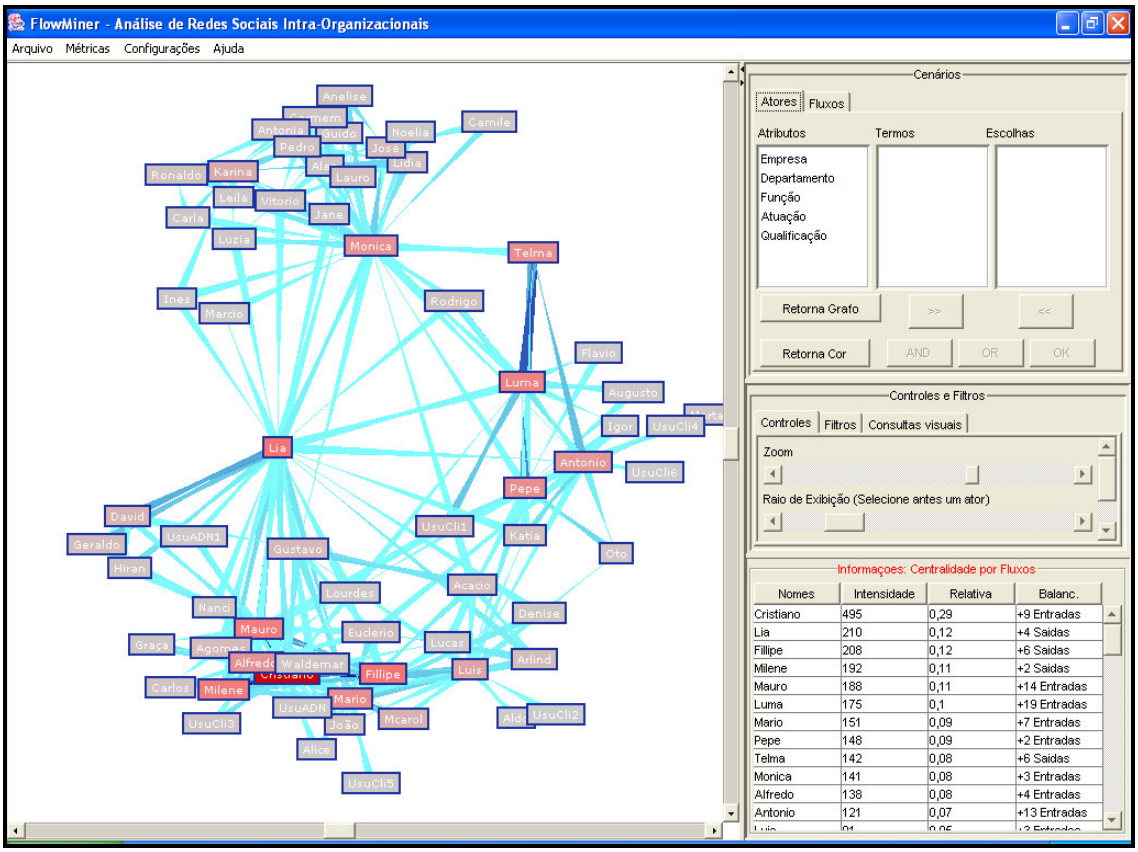

Figura 70. Centralidade dos atores por grau de conexões (ligações)

A Figura 71 mostra a opção *Métrica – Por Ator – Centralidade – Grau de fluxos* que avalia a quantidade de fluxos trafegando pelas ligações. Nesta medida Cristiano apresenta o maior índice, 29% das mensagens da rede. Por este motivo ele foi considerado o *conector central* desta comunidade, pois é o terceiro ator mais procurado em termos de ligações, difere de três ligações do primeiro e uma do segundo, enquanto que em volume de mensagens sua quantidade representa 2,3 vezes a quantidade do segundo ator. Seu balanceamento de conexões demonstra que existe conversação entre ele e seus vizinhos e o balanceamento de fluxos mostra que nove mensagens estão ainda sem repasse ou solução.

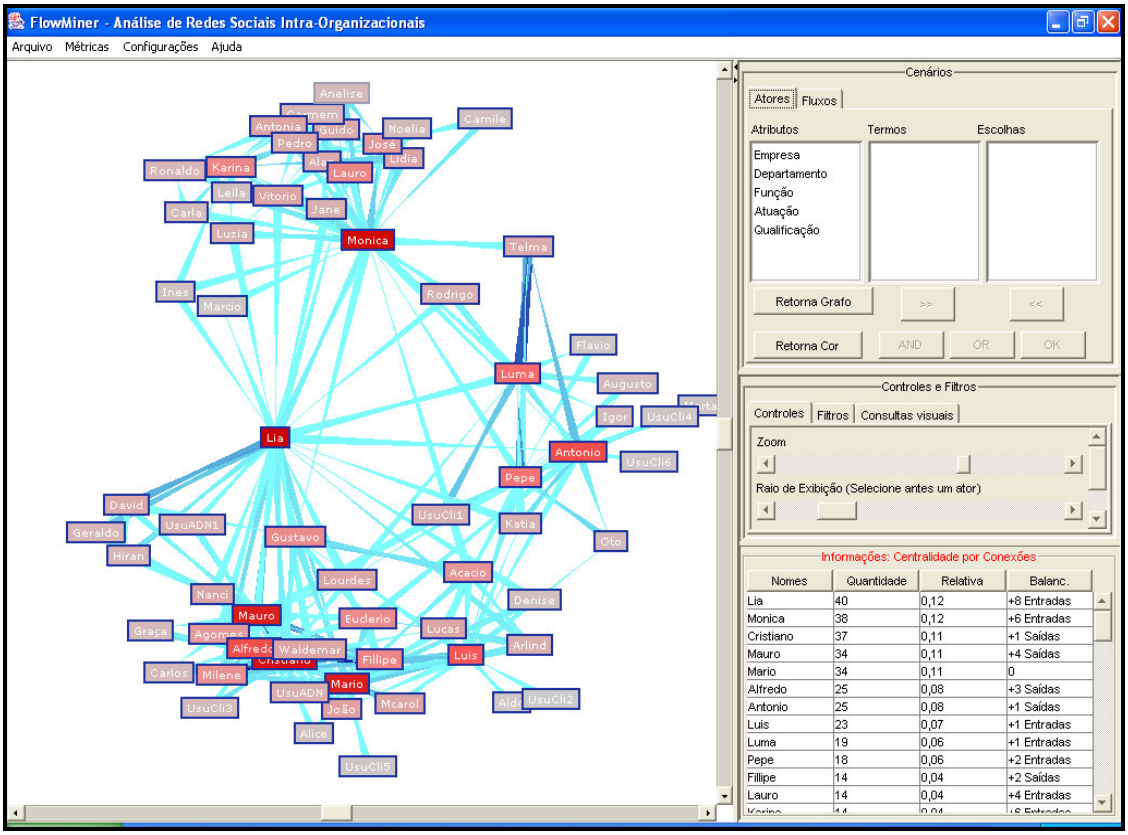

Figura 71. Centralidade dos atores por grau de fluxos (mensagens trafegando)

Analisando os desenvolvedores da equipe Matriz, o *conector central* desta função é Mauro, que se destaca em número de conexões e em quantidade de fluxos de informações. Por Mauro trafegam 11% das mensagens, ele é o desenvolvedor que concentra maior recepção e transmissão de chamados. Apesar de resolver o maior número de chamados, os chamados foram classificados pelos coordenadores como de baixa complexidade.

Entre os apoios, Filipe e Lia que são de clientes distintos, Cliente8 e Cliente7 respectivamente, aparecem com altos volumes de fluxos de entrada e saída. Analisando esta situação verificou-se que estes atores funcionam como cadastradores de chamados de vários usuários do cliente. Este papel em redes sociais é reconhecido como o *expansor de fronteira.* Ele é aquele ator que funciona como receptor e disseminador das informações para outra comunidade, neste caso os usuários do cliente. Esta indicação de padrão de comportamento dos dois apoios coincide com o papel que cada um desempenha nas equipes locais dos respectivos clientes, centralizando todos os chamados dos usuários. O resumo das pontuações obtidas pela ferramenta está apresentado na tabela da Tabela 1.

Tabela 1. Cinco maiores pontuações de centralidade de conexões e fluxos

Grau de conexões		Grau de conexões de entrada		Grau de conexões de saída	
Ator	**Pontuação**	**Ator**	**Pontuação**	**Ator**	**Pontuação**
Lia	40	Lia	24	Cristiano	19
Mônica	38	Mônica	22	Mauro	19
Cristiano	37	Cristiano	18	Mario	17
Mauro	34	Mario	17	Lia	16
Mario	34	Mauro	15	Mônica	16
Grau de fluxo		**Grau de fluxo de entrada**		**Grau de fluxo de saída**	
Ator	**Pontuação**	**Ator**	**Pontuação**	**Ator**	**Pontuação**
Cristiano	495	Cristiano	252	Cleber	243
Lia	210	Lia	103	Lia	107
Filipe	208	Filipe	101	Filipe	107
Milene	192	Mauro	101	Milene	97
Mauro	188	Luma	97	Mauro	87

A próxima métrica de centralidade analisada foi *Proximidade de envio*. Nesta medida Jane e Inês aparecem como os atores com a menor distância para todos os outros. Estes atores possuem apenas ligação de entrada para a rede, nenhum outro ator os alcança. A presença deste tipo de ligação interfere e distorce o resultado desta medida. Para corrigir esta distorção um processo conhecido de *simetrização pelos mínimos* da matriz de adjacência**,** elimina todas as ligações assimétricas ou não recíprocas**.** A execução do processo *Elimina não reciprocidade* e a re-execução da métrica de proximidade de envio, corrige estas distorções e aponta Lia, Cristiano, Mauro e Mario como os quatro atores mais bem posicionados na rede para o envio de informações. Eles estão mais próximos de todos os outros atores. Isto mede a independência destes atores em relação ao controle dos outros. Todos são coordenadores exceto Mauro que é desenvolvedor.Ver Tabela 2.

Tabela 2. Quatro maiores pontuações em proximidade de envio.

Proximidade de envio (total)		Proximidade de envio (ligações recíprocas)	
Ator	**Pontuação**	**Ator**	**Pontuação**
Jane	483	Lia	218
Inês	488	Cristiano	226
Lia	510	Mauro	229
Anelise	511	Mario	233

Para analisar os atores críticos no papel de *corretor de informações*, são usados os resultados dos cálculos de centralidade de intermediação da Tabela 3. Estes atores atuam como *pontos de articulação* facilitando o fluxo de informações. Um indivíduo pode não ter muitos contatos, ligações com baixa intensidade, mas ter importância na mediação de trocas.

“O papel de mediador traz em si o poder de controlar as informações que circulam na rede e o trajeto que elas podem percorrer” (MARTELETO, 2001, p. 79).

Na Figura 72 onde foi aplicada a opção *Métricas – Por Ator – Centralidade – Intermediação,* encontramos Mônica, Lia, Antonio, Mario e Cristiano como os cinco atores com maior índice de intermediação. Eliminando as ligações não recíprocas e re-executando a métrica obtemos uma nova lista, nesta nova lista os mesmos atores se revezam. As duas listas apontam para a importância destes atores no papel de *corretores de informações*. A retirada de um destes atores deixa outros atores sem contato com parte da rede ou sem ligações com a rede. Quanto maior for este índice, mais forte é considerado o papel de *corretor de informações* desempenhado pelo respectivo ator.

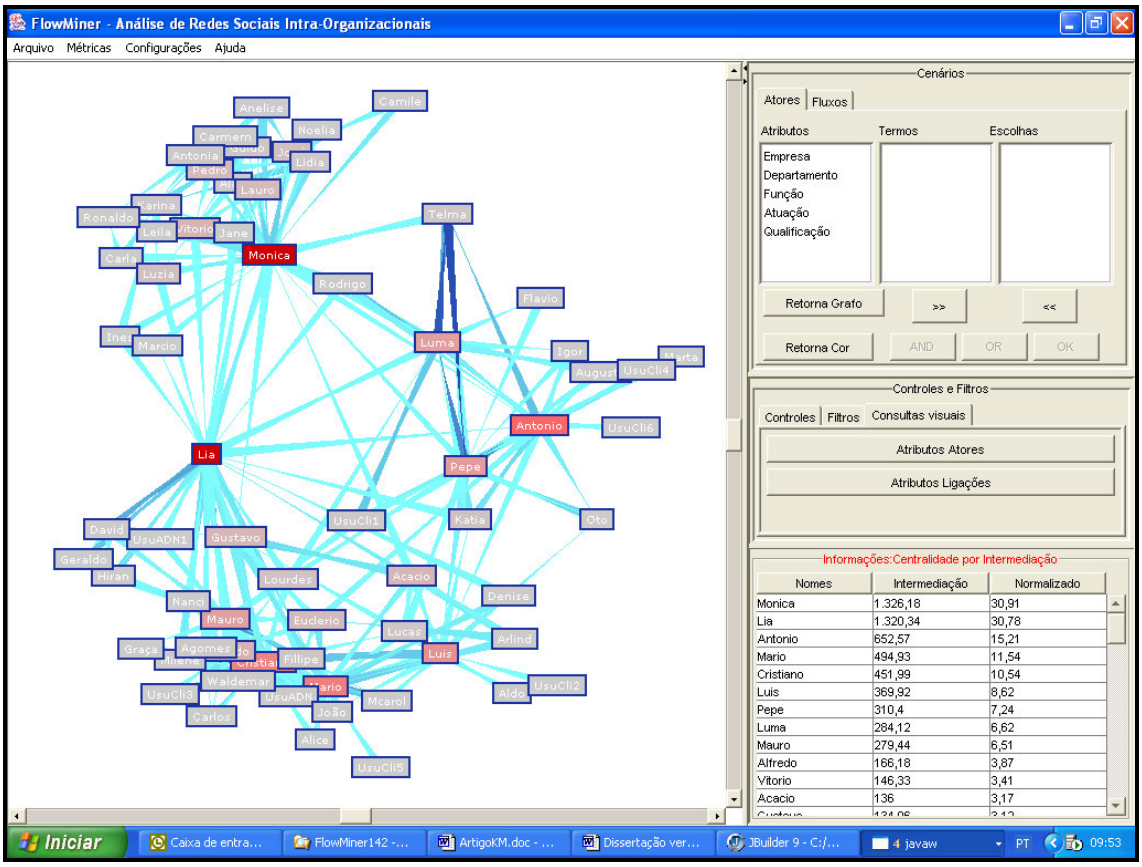

Figura 72. Centralidade dos atores por *Centralidade de Intermediação*

A Tabela 3 resume as cinco primeiras posições em *Centralidade de intermediação* com e sem as ligações não recíprocas.

Tabela 3. Maiores pontuações antes e após eliminação da não reciprocidade.

Intermediação (total)		Intermediação (ligações recíprocas)	
Ator	**Pontuação**	**Ator**	**Pontuação**
Mônica	1.326,18	Lia	873,01
Lia	1.320,34	Mônica	809,45
Antonio	652,57	Mario	589,83
Mario	494,93	Antonio	582,60
Cristiano	451,99	Cristiano	513,12

A presença de Antonio nesta lista onde só existem coordenadores foi uma informação inesperada. Analisando a situação verificou-se que o comportamento de Antonio na rede difere de outros desenvolvedores, pois ele recebe chamados de vários coordenadores bem como diretamente de usuários, participando desta forma de chamados originados por vários atores.

O resumo dos resultados obtidos na análise onde são apresentados os atores críticos nos três papéis relevantes: *conector central, corretor de informações* e *expansor de fronteira* com a descrição de sua atuação estão registrados na Tabela 4.

Tabela 4. Atores críticos da rede

Papel	Atores críticos	Atuação
Conector central	Cristiano	Coordenador
	Mauro	Desenvolvedor
Corretor de informações	Lia	Coordenador
	Mônica	Coordenador
Expansor de Fronteira	Filipe	Apoio
	Milene	Apoio

6.1.4.3 Análise de cenários da comunidade

"Nos espaços informais as redes são iniciadas a partir da tomada de consciência de uma comunidade de interesses ou valores comuns entre os participantes" (MARTELETO, 2001, p. 73). Na análise dos subgrupos, ou comunidades de prática, a ferramenta *FlowMiner* se difere da maioria dos produtos de *SNA*, nos outros produtos esta análise é feita através do método de busca de *cliques,* que foi descrito na seção 3.3.1.2 este trabalho.

Para descobrir subgrupos por interesses a ferramenta *FlowMiner* faz uso da execução de pesquisas dinâmicas através dos atributos dos fluxos, buscando identificar subgrupos que

se comunicam e se movimentam em torno de um determinado assunto. A seleção por categoria de fluxos forma vários cenários, mapeando subgrupos e estabelecendo valores quantitativos sobre as conexões estabelecidas e os fluxos que trafegam. No estudo em questão foram identificados cinco grandes subgrupos na comunidade em ordem de quantidade de mensagens trafegando. Estes subgrupos ou comunidades de prática englobam 94% das mensagens.

1. Subgrupo Manutenção em Garantia
2. Subgrupo Suporte Técnico
3. Subgrupo Manutenção Evolutiva
4. Subgrupo Esclarecimento de Dúvida
5. Subgrupo Orçamentação

A Figura 73 mostra a comunidade que troca mensagens sobre *Esclarecimento de Dúvidas*, nela observa-se 3 subgrupos disjuntos que no total possuem 38 atores, 88 ligações, por onde trafegam 114 fluxos referentes a 273 mensagens.

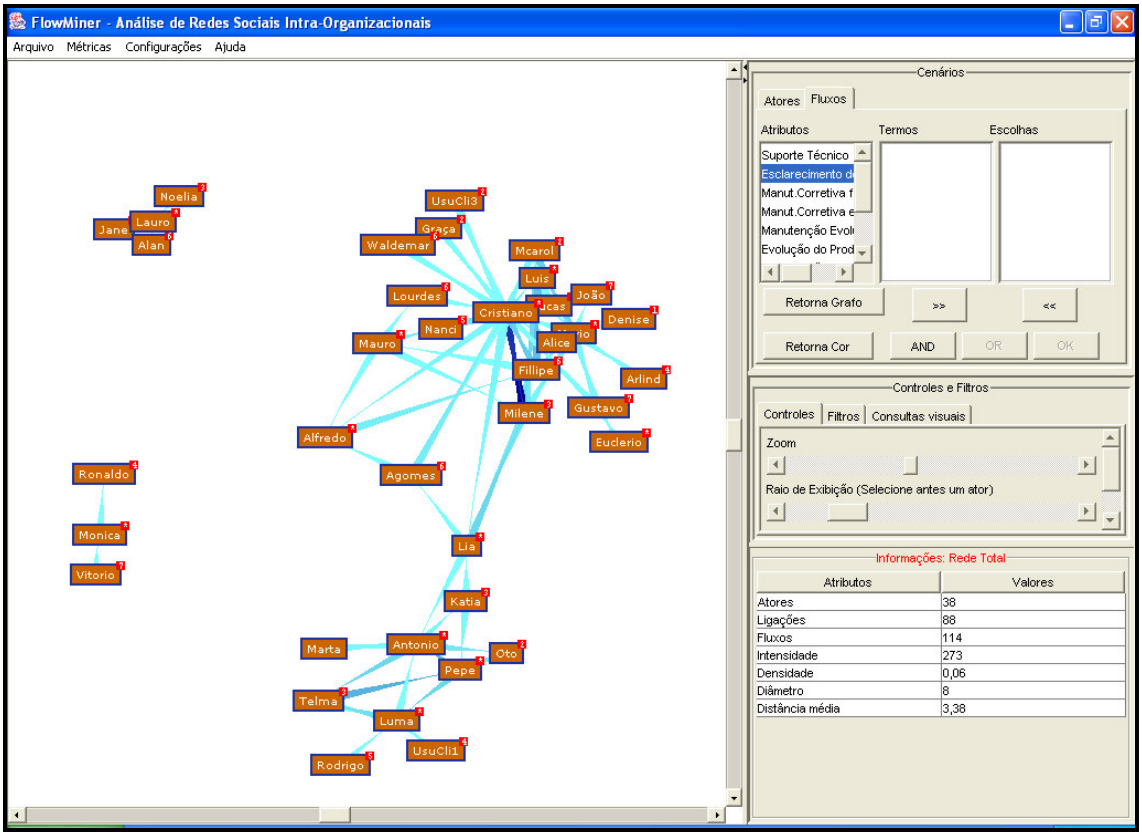

Figura 73. Rede de Esclarecimento de dúvidas

Já a Figura 74 mostra a aplicação da métrica *Centralidade por fluxo de saída* no grafo da Figura 73. Esta medida aponta a colaboração de cada ator. Um resumo dos quatro atores mais bem pontuados está na Tabela 6 e são: Cristiano, Milene, Filipe e Mario. Como Milene e Filipe são apoios da equipe local de clientes, esta informação demonstra que eles necessitam de treinamento, pois deveriam estar resolvendo diretamente as dúvidas e não repassando para a equipe da Matriz. Isto foi comprovado, pois neste período estes dois apoios estavam iniciando o suporte aos respectivos clientes e por este motivo sobrecarregando o coordenador Cristiano de mensagens sobre este tema.

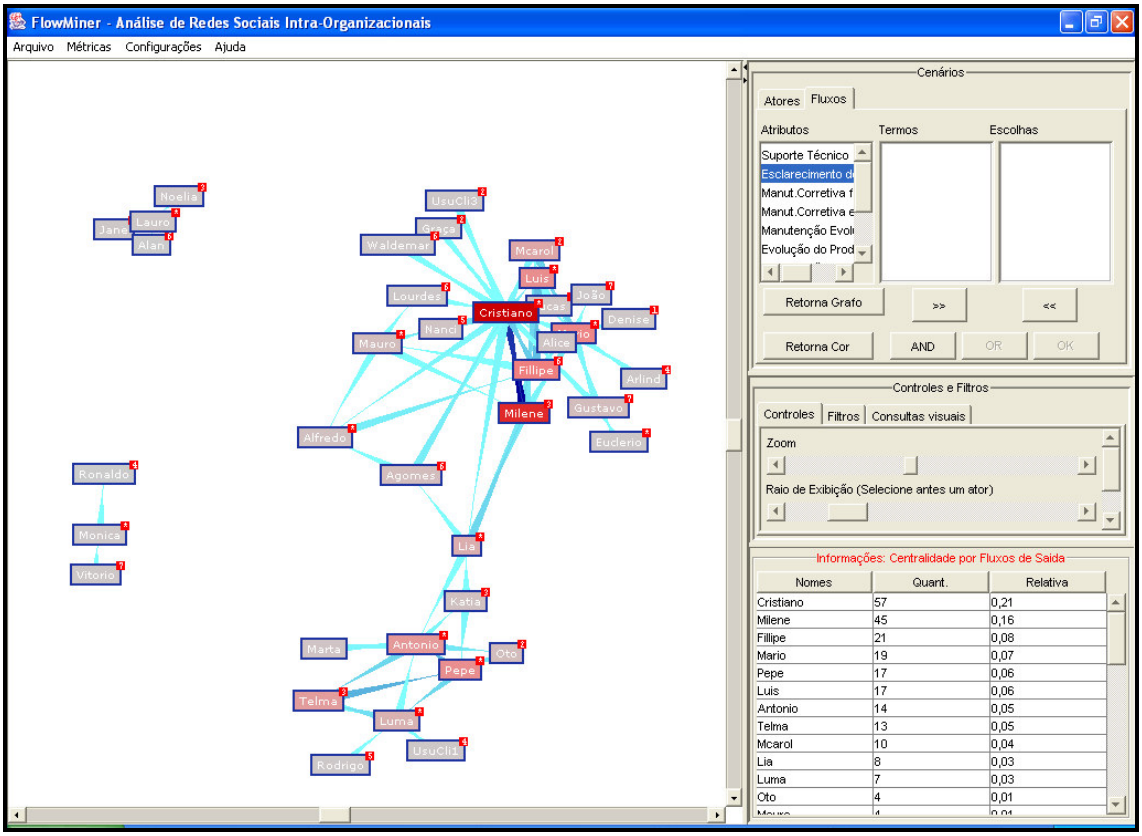

Figura 74. Centralidade por fluxo de saída na comunidade de Esclarecimento de dúvida

Após analisar detalhadamente cada subgrupo foi elaborada a Tabela 5 com o resumo dos quantitativos referentes à estrutura da rede encontrada para cada subgrupo.

Tabela 5. Medidas estruturais de subgrupos ou comunidades por assunto.

Rede por Assunto	Atores	Ligações	Intensidade
Manutenção em Garantia	56	188	645
Suporte Técnico	44	114	321
Manutenção Evolutiva	45	108	291
Esclarecimento de Dúvidas	38	88	273
Orçamentação	25	53	113

Na Tabela 6 encontram-se os conectores centrais encontrados para cada subgrupo, levando em consideração a *Centralidade por Fluxo de Saída.* Esta métrica avaliou o grau de expansividade e de colaboração de cada participante do grupo.

Tabela 6. Conectores centrais por fluxo de saída para cada subgrupo.

Conectores centrais por fluxos de saída				
Manut. em garantia	**Suporte Técnico**	**Manut. Evolutiva**	**Escl. de Dúvida**	**Orçamentação**
Cristiano	Cristiano	Lia	Cristiano	Cristiano
Lia	Luma	Cristiano	Milene	Filipe
Mauro	Taiana	Mauro	Filipe	Álvaro
Filipe	Milene	Monica	Mario	Mario

Dependendo da atuação de cada ator esta medida pode representar padrões de comportamentos diferenciados: se coordenador significa repasse ou resolução de chamados, se desenvolvedor representa resolução de chamados e se apoio ou usuário representa envio de chamado, embora a medida em todos represente o grau da atuação de cada ator no papel que desempenha na rede. A análise dos resultados se torna relativa aos atributos dos atores, comprovando que quanto mais atributos forem introduzidos nos objetos analisados maior será o entendimento do usuário na análise dos dados.

Através das medidas e da observação das redes traçadas, foi possível verificar os comportamentos que movimentam os subgrupos. Foram identificados indivíduos que são responsáveis por estabelecerem relações e facilitarem as trocas de informações entre e os demais. A presença dos atores Filipe e Milene que são apoio, com volume de mensagens de saída sobre *Esclarecimento de Dúvidas,* demonstra necessidade de treinamento destes elementos que atuando como apoio de clientes distintos, pois eles deveriam executar esta tarefa e não repassá-la para a equipe da Matriz.

6.1.4.4 Análise das redes egocêntricas

Para observar o comportamento e padrões adotados por cada indivíduo podemos analisar sua rede egocêntrica, nestas redes a unidade de análise não será os atributos individuais de cada um, mas o conjunto de relações que o indivíduo estabelece através de suas interações com os outros.

A estrutura da rede egocêntrica é vista como uma rede de relações e de limitações que pesa sobre as escolhas, as orientações, os comportamentos e as opiniões do indivíduo analisado (MARTELETO, 2001).

A análise da rede egocêntrica pode ser realizada para cada ator da rede. A seguir análise resumida dos dois conectores centrais da rede: Cristiano e Mauro.

Nas figuras 75 e 76 podemos observar a rede egocêntrica de Cristiano. Na sua rede trafegam 49,5% das mensagens de toda rede. Uma característica importante neste tipo de rede é a composição. Ela determina a qualidade dos vizinhos. Vários atores de relevância, Lia, Mauro, Filipe e Milene, foram encontrados na composição da rede ego de Cristiano. Cristiano é o coordenador mais articulado da rede, com uma forte teia de contatos, por onde trafegam aproximadamente metade do fluxo de informações da rede, demonstrando que este coordenador possui um bom conhecimento sobre os problemas e soluções encontradas e que exerce uma forte influência e controle sobre a comunidade.

Informações: Rede Egocêntrica

Atributos	Valores
Atores	21
Ligações	120
Fluxos	355
Intensidade	859
Densidade	0,29
Diâmetro	3
Distância média	1,76

Figura 75. Foco na estatística da rede ego de Cristiano

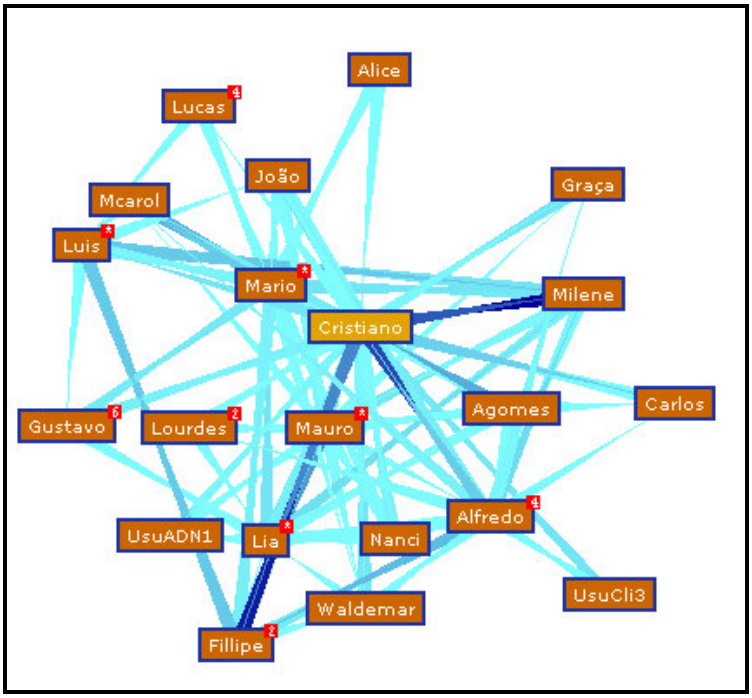

Figura 76. Foco no mapeamento da rede ego de Cristiano

Nas Figuras 77 e 78 podemos observar a rede ego de Mauro, apontado com o desenvolvedor que recebe e manda mais mensagens. Na sua rede ego trafegam 44,4% das mensagens da rede. A composição apresenta também a presença de vizinhos relevantes em colaboração como: Cristiano, Lia, Milene e Filipe.

Informações: Rede Egocêntrica

Atributos	Valores
Atores	21
Ligações	100
Fluxos	282
Intensidade	770
Densidade	0,24
Diâmetro	3
Distância média	1,82

Figura 77. Foco na estatística da rede ego de Mauro

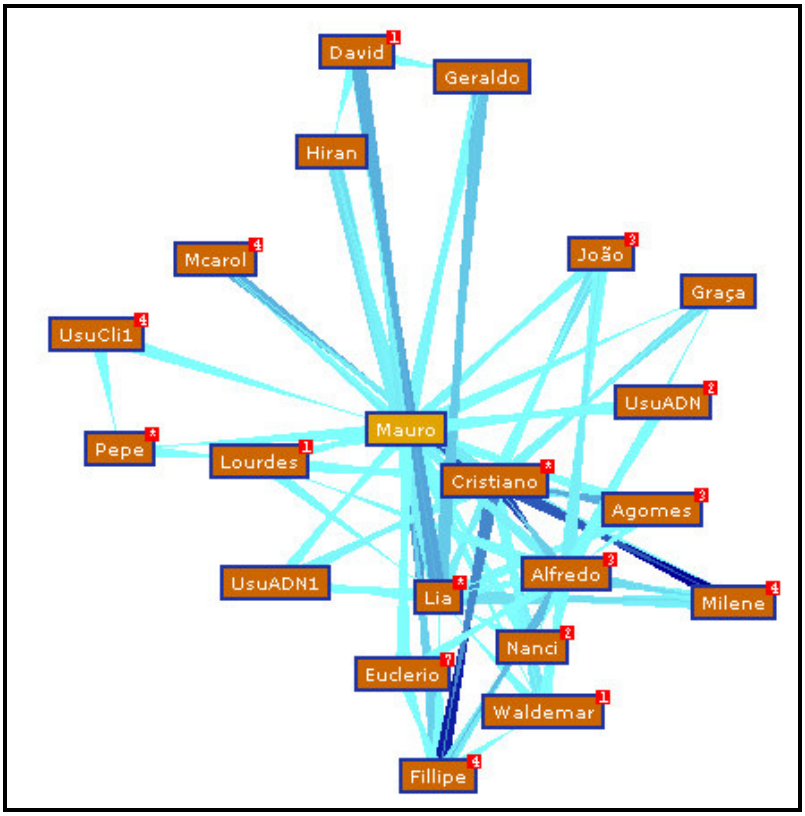

Figura 78. Foco no mapeamento da rede ego de Mauro

6.2 Discussão dos resultados

O uso da metodologia de análise de redes sociais juntamente com as características de exploração visual disponibilizadas na ferramenta *FlowMiner* permitiram e facilitaram a

combinação de elementos para entendimento e avaliação do modo de comunicação, troca de informações e conhecimentos na comunidade estudada.

Os padrões de comunicação dos atores que participam da rede foram avaliados e identificados os atores que são responsáveis pela dinamização da rede. Estes indivíduos atuam como mediadores ou facilitadores do fluxo de informações, colaborando com a disseminação de informações e fortalecendo as ligações entre a rede como um todo.

A visão fornecida pela ferramenta complementou as informações geradas pelo sistema *HelpDesk*, pois introduziu uma visão do grupo e de seu comportamento de comunicação. As análises subsidiadas pela ferramenta permitiram as seguintes constatações e identificação:

- A rede estudada tem uma estrutura hierarquizada, em que cada ator primeiro acessa o coordenador de sua equipe para depois alcançar qualquer outro elemento na rede;
- Dois apoios locais funcionam como expansores de fronteira, entre a comunidade da rede e a comunidade do cliente;
- Estes expansores de fronteira necessitam de treinamento;
- Existe um desenvolvedor com comportamento de intermediador;
- Identificou-se o coordenador e desenvolvedor mais dinâmicos da rede;
- Identificou-se as comunidades de prática por assunto e seus maiores colaboradores.

Para execução dos levantamentos de dados do *HelpDesk*, de dados dos atores, apresentação e interpretação dos resultados da ferramenta *FlowMiner* foram feitas 5 visitas à ADN e gastas 20 horas. Nas apresentações da ferramenta e suas funcionalidades foram envolvidos três coordenadores e na coleta dos dados foi envolvido o Administrador de Banco de Dados (*DBA*). Para a interpretação e análise dos resultados obtidos foi envolvida a coordenadora responsável pelo controle e gerenciamento da ferramenta *HelpDesk*. Três bases de dados de períodos crescentes foram carregadas até a versão final dos dados, onde estão registrados quatro meses de comunicação do grupo.

A ferramenta foi considerada ao final dos testes como de fácil utilização e sua forma de apresentação dos resultados adequada para o que se propõe.

Os resultados encontrados com a ferramenta *FlowMiner* foram analisados e interpretados por alguns coordenadores que conhece e utilizam o sistema *HelpDesk,* eles acompanharam o desenvolvimento do estudo de caso. As métricas foram confrontadas com as métricas da ferramenta *UCINET* utilizando o mesmo conjunto de dados, para garantir a correção dos valores obtidos. Os resultados extraídos das métricas da ferramenta *UCINET* se encontram no ANEXO B.

6.3 Considerações Finais

Neste capítulo foi apresentado um estudos de caso com a ferramenta *FlowMiner*, como uma forma de demonstrar e avaliar sua aplicabilidade em organizações. No próximo capítulo será apresentada a conclusão desta dissertação, incluindo considerações sobre os trabalhos futuros e as evoluções previstas para a ferramenta.

CAPÍTULO 7 - CONCLUSÃO

A principal motivação para a adoção de uma solução de análise de redes sociais que foque em interesses compartilhados é a busca por uma alta conectividade de seus participantes com a finalidade de promover um aumento na oportunidade de colaboração e compartilhamento de recurso, informações e conhecimentos.

O aumento do volume em pesquisas de gerenciamento através de redes sociais demonstra um direcionamento para o entendimento relacional, contextual e sistêmico em que procura-se analisar e entender como as pessoas interagem e encontram informações para executarem tarefas do dia-a-dia.

Este trabalho apresentou uma abordagem sobre mapeamento de redes de comunicação em organizações, inserindo a ferramenta como um produto de auxílio e apoio à Gestão do Conhecimento, permitindo que gerentes tenham uma visão mais real de como se processa a interação e troca de informações entre seus colaboradores e a partir deste entendimento possam corrigir distorções, introduzir mudanças, ampliar recursos, ou seja, agirem como facilitadores de uma maior cooperação e solidariedade entre seus membros. Já os colaboradores podem identificar e avaliar seu posicionamento nas comunidades e a partir desta análise promover ampliações nos contatos ou fortalecimento nas ligações.

Um ambiente aberto à colaboração, adaptável, com ênfase na potencialização da comunicação e na eficiência do fluxo de informações, torna-se um ambiente confiável e propício ao aprendizado organizacional.

A ferramenta *FlowMiner,* proposta neste trabalho, realiza uma estratégia caracterizada pela incorporação de maior semântica aos dados através de qualificadores nos objetos

observados e na exploração de variados mecanismos de visualização e interação com os mesmos. Esta estratégia traz como maior beneficio melhorar o entendimento e simplificar os resultados apresentados, facilitando o julgamento dos critérios adotados.

Pode-se concluir que as linhas teóricas desenvolvidas ao longo desta dissertação foram comprovadas no estudo de caso realizado, que identificou elementos críticos e padrões de comportamento adotados pela comunidade seguindo metodologia de análise de redes organizacionais sugerida neste trabalho.

7.1 Contribuições

As organizações eficientes e competitivas são aquelas que possuem redes de relacionamentos entre seus colaboradores. A proposta desta dissertação é a disponibilização de uma ferramenta que apóie a gestão do conhecimento organizacional através da representação visual dos relacionamentos entre os indivíduos, construídos a partir da rede de comunicação.

Dentre as principais contribuições deste trabalho podem ser destacadas.

- A identificação de colaboradores chaves que precisam ser mantidos e estimulados a disseminar informações e conhecimento;
- A identificação de comunidades de práticas e o acompanhamento da dinâmica de seus relacionamentos;
- Tornar o fluxo de comunicação transparente, gerando um ambiente de trabalho mais aberto e confiável, fortalecendo atitudes positivas de compartilhamento e colaboração ampliando e compreendendo a formação do capital social;
- Estimular coordenadores, gerentes e funcionários a construírem uma boa rede pessoal para aumentar produtividade e eficiência;
- Descobrir oportunidade para inserir mudanças, ampliar a conectividade e estimular membros periféricos da rede a se incluírem em comunidades de interesse.

7.2 Trabalhos Futuros

O estudo de caso apresentado no decorrer deste trabalho e sugestões vindas dos usuários que participaram de sua elaboração, apontaram novas funcionalidades que podem vir a serem implementadas à ferramenta em trabalhos futuros. As principais extensões sugeridas foram:

- A inclusão de um processo para criar múltiplos níveis de análise nos atores, permitindo o agrupamento destes atores em entidades organizacionais que por sua vez podem ser analisadas;
- Permitir a qualificação de um fluxo em mais de uma categoria;
- A auditoria do processo de seleção das comunidades com a finalidade de possibilitar retorno a cenários intermediários;
- O registro de observações encontradas pelo analisador ao longo do processo.

Além disso, podem ser listados alguns trabalhos futuros relacionados à evolução da ferramenta:

- Integrar a ferramenta com outra ferramentas capazes de automatizar a captura de mensagens trocadas em produtos de correio eletrônico, que por sua vez se integrará com uma ferramenta de classificação textual para categorização do conteúdo das mensagens, gerando de forma automática tabelas para serem incorporadas à base de dados de entrada da ferramenta *FlowMiner*;
- Melhorar o desempenho da ferramenta para permitir trabalhar com maiores volumes de dados;
- Inserir a dimensão tempo na análise dos dados.

Este trabalho apresentou uma abordagem que envolve a construção de uma ferramenta para mapeamento de redes de comunicação em organizações, cujo objetivo é se tornar um instrumento útil para as organizações que precisam gerenciar o seu capital social. Durante este trabalho foram apresentados a fundamentação teórica e revisão da literatura relacionadas ao tema, a descrição detalhada do trabalho desenvolvido, um estudo de caso com a ferramenta

FlowMiner, as conclusões e contribuições obtidas no decorrer dessa dissertação, como também as perspectivas futuras para a evolução da ferramenta construída.

REFERÊNCIAS

ALMEIDA, Marcio. **Uma ferramenta para mineração visual de dados usando mapas em árvores e suas aplicações.** 2003. 57 f. Dissertação (Mestrado) - Redes de Computadores, Universidade Salvador, Salvador, 2003.

BORGATTI, Stephen P.; JONES, Candace; EVERETT, Martin G. **Network Measures of Social Capital.** 1998**.** Disponível em: <http://www.analytictech.com/borgatti/publications.htm>. Acesso em: mar. 2004.

BORGATTI, Stephen P.; EVERETT, Martin G.; FREEMAN, Linton C. **UCINET 6.0 Versão 1**. 2002. Disponível em: <http://www.analytictech.com/ucinet.htm>. Acesso em: mar. 2005.

BORGATTI, Stephen P. **Basic Social Network Concepts**, 2002**.** Disponível em: <http://www.analytictech.com/borgatti/.com/ucinet.htm>. Acesso em: fev. 2004.

BORGATTI, Stephen P.; FOSTER, Pacey C. The Network Paradigm in Organizational Research: A Review and Typology. **Journal of Management**, 2003. 29(6), p. 991-1013. Disponível em: <www.analytictech.com/borgatti/papers/borgattifoster.pdf>. Acesso em: mar. 2004.

BRANCO, V. **Visualização como Suporte à Exploração de uma Base de Dados Pluviométricas.** 2003. 87 f. Dissertação (Mestrado), Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo, 2003.

BRANDES, Ultrik. A Fast Algorithm for Betweenness Centrality. **Journal of Mathematical Sociology**, 2001, 25(2), p. 163-177. Disponível em: <www.inf.uni-konstanz.de/algo/publications/b-fabc-01.pdf>. Acesso em: mar. 2004.

BRANDES, Ulrik; WAGNER, Dorothea. **VISONE - Analysis and Visualization of Social Network**. 2003. Disponível em: <http://www.visone.de/>. Acesso em: mar. 2004.

CAMPBELL, Christopher S. et al. Expertise identification using email communications. In: CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 12., 2003, New Orleans. **Proceedings of the twelfth international conference on Information and knowledge management.** New York: ACM Press, 2003, p. 528 - 531

CARD, Stuart; MACKINLAY, Jock; SHNEIDERMAN, Ben. **Readings in Information Visualization:** using visualization to think. San Francisco: Morgan Kaufmann Publ, 1999. 671 p.

CASTRO, Cláudio de Moura. **Examinando as novas tendências e transformações no mundo do trabalho.** In**:** Seminário Internacional Mercado de Trabalho e Dinâmica Ocupacional, 2002, Belo Horizonte. Informativo do SENAI/DN - Sistema CNI. Disponível em: <http://www.dn.senai.br/sb/sb59/novas_tendencias59.htm>. Acesso em: jul. 2005.

CHACHRA, Vinod; GHARE, Prabhakar; MOORE, James. **Applications of Graph Theory Alghorithms.** New York: North Holland, 1979. 421 p.

CHANG, Chin-Lung; CHEN, Ding-Yi; CHUANG, Tyng-Ruey. Browsing Newsgroups with A Social Network Analyzer. In: INTERNATIONAL CONFERENCE ON INFORMATION VISUALIZATION, 6., 2002, **Sixth Information Visualisation (IV'02).** IEEE Publish, 2002. p. 750 - 755.

CIANCONI, Regina. **Gestão da Informação na Sociedade do Conhecimento.** 2. ed. Brasília. CNI/SENAI, 2001. 119 p.

CROSS, Rob et al. Knowing What We Know: Supporting Knowledge Creation and Sharing in Social Networks. **Organizational Dynamics,** v. 30, n. 2, p.100-120, 2001. Disponível em: <www.organizational-dynamics.com>. Acesso em: abr. 2004.

CROSS, Rob; PRUSAK, Laurence. The people who make organization go - or stop. **Havard Business Review**, Boston, v. 80, n. 6, p.105-112, 2002.

CUMMINGS, Jonathan. **NetVis – Dynamic Visualization of Social Network.** Disponível em: <em http://www.netvis.org/index.php>. Acesso em: maio 2004.

DAVENPORT, Thomas; PRUSAK, Laurence. **Ecologia da Informação:** por que só a tecnologia não basta para o sucesso na era da informação. 6. ed. São Paulo: Futura, 1998. 316 p.

DIJKSTRA, Edsger**.** A note on two problems in connetion with graphs. In: **Numerische Mathematik**, Mathematisch Centrum, Amsterdam, 1, S. 269–271, 1959. Disponível em: <http://www.garfield.library.upenn.edu/classics1983/A1983QA19900001.pdf>. Aceso em: 23 abr. 2004.

EARL, Michael. **Knowledge as Strategy: Reflections on Skandia International and Shorko Films**, Knowledge in Organization, Laurence Prusak (ed.), Butterworth-Heinemann, p 1-13, 1997.

EICK, Stephen G. Visual Discovery and Analysis. **IEEE Transaction on Visualization and Computer Graphics**, v.6, n.1, p. 44-58, 2000.

FAYYAD, Usama; PAITETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. The KDD process for Extrating Useful Knowledge for Volume of Data. In: **Communication of the ACM**, v. 39, n. 11, 1996.

FAYYAD, Usama. Mining Database: Towards Algorithms for Knowledge Discovery. In: **Bulletin of the IEEE Computer Society Technical Committee on Data Engineering,** v. 21, no. 1, March 1998.

FREEMAN, Linton C. Visualizing Social Networks. **Jornal of Social Structure,** v. 1, 2000**.** Disponível em: <http://www.cmu.edu/joss/content/articles/volume1/Freeman.html>. Acesso em fevereiro 2004.

FREITAS, Carla M. et al. **Introdução à visualização de informações.** 2001**.** Disponível em: <http://www.inf.ufrgs.br/~revista/docs/rita08/rita_v8_n2_p143a158.pdf>. Acesso em: 20 dez. 2004.

GAERTLER, Marco; WAGNER, Dorothea. **Algorithms for representing centrality, groups and density and clusteres graph representation.** Disponível em: <www.cosin.org/extra/deliverables/D6.pdf>. Acesso em: 15 ago. 2004.

GLOOR, Peter A. et al. Visualization of Communication Patterns in Collaborative Innovation Networks - Analysis of Some W3C Working Groups. In: CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 12., 2003, New Orleans. **Proceedings of the twelfth international conference on Information and knowledge management.** New York: ACM Press, 2003. p. 750 - 755.

GROOTAERT, Christian at al. **Questionário Integrado para Medir Capital Social.** Banco Mundial, 2003. 73 p. Disponível em: <http://poverty2.forumone.com/library/view/11998>. Acesso em: set. 2004.

GUIMERÀ, R. et al. **Self-Similar community structure in organization.** 2002. Disponível em: <arxiv.org/abs/cond-mat/0211498>. Acesso em: dez. 2004.

HANNEMAN, Robert A. **Introdution to Social Methods.** 2001**.** Department of Sociology, University of California, Riverside. Disponível em: <http://www.analytictech.com/ucinet.htm> Acesso em: mar. 2005.

HANSEN, Morten T.; NOHRIA, Nitin; TIERNEY, Thomas. What´s Your Strategy for Managing Knowledge? **Harvard business review on organizational learning**, Boston, Harvard Business School Press, v. 1, p.61-86, 1999.

HOFFMAN, Patrick; GRINSTEIN, George. A survey of visualization for high-dimensional data mining, 2002. In: **Information Visualization in Data Mining and Knowledge Discovery**, U. Fayyad, G. G. Grinstein, and A. Wierse, Eds. Morgan Kaufmann Publishers, San Francisco, CA, 47-82.

INSELBERG, Alfred; DIMSDALE Bernad. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In: IEEE VISUALIZATION, 1., 1990, San Francisco, California. **Proceedings of the 1st conference on Visualization '90.** Los Alamitos, CA: IEEE Computer Society Press, 1991. v. 1, p. 361 - 378.

JOHNSON, Brian; SHNEIDERMAN Ben. Tree-Maps: a space-filling approach to the visualization of hierarchical information structures. In: IEEE VISUALIZATION, 2., 1991, San Diego. **Proceedings of the 2nd conference on Visualization '91.** Los Alamitos, CA, USA: IEEE Computer Society Press, 1991. v. 1, p. 284 – 291, , ISBN:0-8186-2245-8 (PAPER).

KEIM, Daniel A.;KRIEGEL, Hans-Peter. Visualization Techniques for Mining Large Database: A Comparison. **IEEE Transaction on Knowledge and Data Engeneering**, v. 8, n.6, p. 923-938, 1996, ISSN:1041-4347**.**

KEIM, Daniel A. Designing pixel-oriented visualization techniques: Theory and Application. **IEEE Transaction on Visualization and Computer Graphics**, v.6, n.1., p.59-78, 2000**,** ISSN:1077-2626.

KEIM, Daniel A. Information Visualization and Visual Data Mining. **IEEE Transactions on Visualization and Computer Graphics**, New Jersey, v. 8, n. 1, p.1-8, jan. 2002. ISSN:1077-2626.

KRACKHARDT, David; HANSON, Jeffrey. Informal Networks: The Company behind the chart, **Harvard Business Review**, Boston, EUA, v. 71, n. 4, p. 37-49, 1993.

KREBS, Valdis. **Mapping and Measuring Knowledge Creation**, **Re-use and Flow,** 1998. Disponível em <http://www.orgnet.com>. Acesso : em fevereiro 2004

_____. **Managing the Connected Organization,** 1999**.** Disponível em <http://www.orgnet.com.>. Acesso em: fev 2004

LESSER, Eric; PRUSAK, Laurence. Communities of Practice, Social Capital and Organization Knowledge. In: Lesser, E.; Fontaine A.; Slusher J., **Knowledge and communities,** p.123-131, Boston, Butterworth Heinemann, 2000.

LÉVY, Pierre. **Inteligência Coletiva**: por uma antropologia do ciberespaço. 2. ed. SP: Loyola, 1999, 212 p.

LÉVY, Pierre; AUTHIER, Michel. **As árvores de conhecimentos**. 2 ed., SP: Escuta, 1995, 192 p.

MARIANI, Antonio C. **Teoria dos Grafos**. Disponível em: <http://www.inf.ufsc.br/grafos/livro.html>. Acesso em: ago 2004.

MARTELETO, Regina Maria. Análise de redes sociais - aplicação nos estudos de transferência da informação. **Revista Ciência da Informação**, v. 30, n. 1, p. 71-81, 2001, IBICT.

MENDONÇA NETO, Manoel Gomes de. Mineração de Dados. In: ESCOLA REGIONAL DE INFORMÁTICA DA SBC REGIONAL DE SÃO PAULO, 6., 2001, São Carlos. **Anais: Minicursos - Coletâneas e textos.** São Carlos: USP-ICMC, 2001. p. 189 - 218.

MENDONÇA NETO, Manoel Gomes de; ALMEIDA, Marcio. Uso de Interfaces Abundantes em Informação para Exploração Visual de Dados. In: IV WORKSHOP SOBRE FATORES HUMANOS EM SISTEMAS COMPUTACIONAIS, 4., 2001, Florianópolis. **Anais IHC 2001.** p. 256 - 268.

MESQUITA, Maria. **Tecnologias de Gestão do Conhecimento e sua relação com a inovação nas organizações**: o caso de uma multinacional de consultoria. 2003. 123 f. Dissertação (Mestrado), Administração de Empresas, UFBA, Salvador, 2003.

MARQUES NETO, Manoel. **GraphMiner:** Uma ferramenta para visualização de dados em banco de dado**s.** 2004. 103 f. Dissertação (Mestrado), Redes de Computadores, Universidade Salvador, Salvador, 2004.

NIGGEMANN, Oliver. **Visual Data Mining of Graph-Based Data**. 2001. 150 f. Tese (Doutorado). Departament of Mathematics and Computet Science, University of Paderborn, Germany, 2001.

NONAKA, Ikujiro; TAKEUCHI, Hirotaka. **Criação de Conhecimento na Empresa:** como as empresas japonesas geram a dinâmica da inovação. 10. ed. RJ: Campus, 1997. 358 p.

NORTH, Chris; SHNEIDERMAN, Ben. **Snap-Together Visualization:** A User Interface for Coordinating Visualizations via Relational Schemata**.** 1999, University of Maryland. Disponível em: <http://www.cs.umd.edu/hcil/snap/snap-avi00.pdf>. Acesso em: maio 2005.

OHSAWA, Yukio et al. Featuring Web Communities based on Word Co-occurrence Structure of Communications. In: WWW2002, 11, 2002, Honolulu, Hawaii. **The eleventh International World Wide Web Conference,** v. 1, p. 1 - 15. Disponível em: <http://www2002.org/CDROM/refereed/105/>. Acesso em: fev. 2004.

PICKETT, Ronald; GRINSTEIN, George. Iconographic displays for visualizing multidimensional data. In: **IEEE Conference on Systems, Man and Cybernetics**, 1988, IEEE Press, Psicataway, NJ, pages 514-519.

PRUSAK, Laurence. **Knowledge in Organizations:** Resources for the Knowledge-Based Economy, 1 ed., Butterworth, Heinemam, 1997, 253 p.

PRUSAK, Laurence. Where did Knowledge Management come from? **IBM System Journal**, v. 40, n. 4, p.1002-1007, maio 2001.Knowledge Management

ROHALL, Steven et al. Email Visualization to Aid Communication. In: INFOVIS 2001, San Diego. **IEEE Symposium on Information Visualization**: IEEE Publications, 2001. v.1, p. 1-4. Disponível em: <http://domino.research.ibm.com/cambridge/research.nsf/0/62fd8b3a73c2cfff85256ad4004e07ac/$FILE/remail-steve%20rohall.pdf>. Acesso em: 15 fev. 2004.

SACKS, Warren. Email Visualization to Aid Communication. In: **International Conference on Intelligent User Interfaces**, 5., 2000, New Orleans. New York: ACM Press, 2000. v. 1, p. 233-240

SCOTT, John. **Social Network Analysis.** 2. ed. London: Sage, 2000. 208 p.

SCHWARTZ, Michael F.; WOOD, David C. M. Discovering Shared Interest Among People Using Graph Analysis of Global Eletronic Mail Traffic, In: **Communications of the ACM**, volume 36, issue 8, p. 78-89, 1993

SEAMAN, Carolyn et al. An Experience Management System for a Software Consulting Organization. In: NASA SEL, 1999. **Proceedings of the 24th NASA SEL Software Engineering Workshop** (SEW´24), Greenbelt, MD: NASA Goddard Space Flight Center..

SEAMAN, Carolyn et al. User Interface Evaluation and Empirically-Based Evolution of a Prototype Experience Management Tool. **IEEE Transaction on Software Engineering**, v.29, n.9, p. 838-850, 2003.

SHNEIDERMAN, Ben. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: VISUAL LANGUAGE, 1., 1996, Washington. **Proceedings of the 1996 IEEE Symposium on Visual Languages.** Washington: IEEE Computer Society, 1996. v. 1, p. 336 - 336.

SILVA, Marcus César Marinho da. **Redes Sociais Intra-organizacionais Informais e Gestão:** um estudo nas áreas de manutenção e operação da Planta HYCO-8, Camaçari, BA. 2003. 213 f. Dissertação (Mestrado), Administração de Empresas, UFBA, Salvador, 2003.

SOMMERVILLE, Ian. **Engenharia de Software.** 6. ed. São Paulo: Addison Wesley, 2003. 592 p.

SVEIBY, Karl Erik. **A nova riqueza das Organizações**, 6 ed, Rio De Janeiro: Campus, 1998. 280 p.

SUN Microsystem, Inc. **The Source for Java Technology.** Disponível em: <http://java.sun.com>. Acesso em: mar. 2004.

TYLER, Joshua; DENNIS, Wilkinson; HUBERMAN, Bernardo. **Email as Spectroscopy: Automated Discovery of Community Structure within Organizations.** 2003. HP Labs, Palo Alto, CA. Disponível em: <http://www.hpl.hp.com/research/idl/papers/email/email.pdf>. Acesso em: fev. 2004.

TOUCHGRAPH. **TouchGraph llc**. Disponível em: <http://www.touchgraph.com> .Acesso em: fev. 2004.

WALDURA, Renaud. **Dijkstra's Shortest Path Algorithm in Java.** Disponível em: <http://renaud.waldura.com/doc/java/dijkstra/>. Acesso em: ago. 2001.

WASSERMAN, Stanley; FAUST, Katherine. **Social Network Analysis:** methods and applications. Cambridge: Cambridge University Press, 1999. 825 p.

WEISSTEIN, Eric. **Graph Diameter**. 1999. In: MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/GraphDiameter.html Acesso em: agosto/2004

WONG, Pak Chung; BERGERON, Daniel. **30 Years of Multidimensional Multivariate Visualization.** Washington: IEEE Computer Society, 1997. 30 p. ISBN:0-8186-7777-5.

ZIVIANI, Nivio. **Projetos de algoritmos:** com implementação em Pascal e C. 2. ed. São Paulo: Thompson, 2004. 552 p.

ANEXO A

A seguir serão apresentados pseudo-códigos dos dois principais algoritmos utilizados pela ferramenta *FlowMiner*:

1 - Algoritmo de Dijkstra implementado por Renaud Waldura (WALDURA, 2003)

O algoritmo de Dijkstra particiona os vértices em dois grupos distintos, o grupo de vértices não marcados e o grupo de vértices marcados. Inicialmente todos os vértices são não marcados, e o algoritmo termina quando todos os vértices estão marcados. Um vértice passa do estado não marcado para um estado marcado, quando o algoritmo encontra o menor caminho deste vértice para a fonte.

As seguintes estruturas de dados são utilizadas pelo algoritmo:

d - guarda o melhor cálculo da menor distancia da fonte para cada vértice

π – guarda o predecessor de cada vértice no menor caminho da fonte

S – grupo de vértice marcados, os vértices cujo menor caminho da fonte tenham sido encontrados

Q – o grupo dos vértices não marcados

A partir destas definições o algoritmo será descrito em alto nível. Sendo **s** o vértice fonte:

```
// initialize d to infinity, π and Q to empty
  d = (∞)
  π = ()
  S =Q = ()
  Add s to Q
  d(s) = 0
  while Q is not empty
        { u = extract-minimum(Q)
          add u to S
          relax-neighbors(u) }
```

relax-neighbors(u)

```
 { for each vertex v adjacent to u, v not in S
        { if d(v) > d(u) + [u, v]   // a short distance exists
             { d(v) = d(u) + [u, v]
               π(v) = u
               add  v to Q
              }
        }
  }
```

extract-minimum(Q)

```
 { find the smallest (as defined by d) vertex in Q
   remove it from Q
   return it
  }
```

2 - Algoritmo de Brandes para calcular intermediação em grafos sem peso (BRANDES, 2001)

Algorithm 1: Betweenness centrality in unweighted graphs

```
C_B[v] ← 0, v ∈ V;
for s ∈ V do
    S ← empty stack;
    P[w] ← empty list, w ∈ V;
    σ[t] ← 0, t ∈ V;   σ[s] ← 1;
    d[t] ← −1, t ∈ V;   d[s] ← 0;
    Q ← empty queue;
    enqueue s → Q;
    while Q not empty do
        dequeue v ← Q;
        push v → S;
        foreach neighbor w of v do
            // w found for the first time?
            if d[w] < 0 then
                enqueue w → Q;
                d[w] ← d[v] + 1;
            end
            // shortest path to w via v?
            if d[w] = d[v] + 1 then
                σ[w] ← σ[w] + σ[v];
                append v → P[w];
            end
        end
    end
    δ[v] ← 0, v ∈ V;
    // S returns vertices in order of non-increasing distance from s
    while S not empty do
        pop w ← S;
        for v ∈ P[w] do δ[v] ← δ[v] + (σ[v]/σ[w]) · (1 + δ[w]);
        if w ≠ s then C_B[w] ← C_B[w] + δ[w];
    end
end
```

ANEXO B

A seguir relatórios estatísticos gerados pela ferramenta *UCINET 6* (BORGATTI; EVERETT; FREEMAN, 2002) para as principais métricas utilizadas no estudo de caso. Durante a fase de testes da ferramenta *FlowMiner* os resultados obtidos foram confrontados diversas vezes com os resultados do produto UCINET, utilizando em ambas as ferramentas o mesmo conjunto de dados com a finalidade de avaliar a acurácia das informações calculadas. A seguir resultados obtidos com o último conjunto de dados utilizado no estudo de caso. Todos resultados das métricas apresentadas nas figuras, no decorrer deste trabalho, utilizaram este mesmo conjunto de dados.

1 - Cálculo da densidade da rede total

BLOCK DENSITIES OR AVERAGES

Input dataset: C:\Arquivos de programas\Ucinet 6\DataFiles\HelpDeskFinal

Relation: 1

Density (matrix average) = 0.0730
Standard deviation = 0.2602

Use MATRIX>TRANSFORM>DICHOTOMIZE procedure to get binary image matrix.
Density table(s) saved as dataset Density
Standard deviations saved as dataset DensitySD
Actor-by-actor pre-image matrix saved as dataset DensityModel

---
Running time: 00:00:01

Output generated: 02 set 05 14:27:43
Copyright (c) 1999-2005 Analytic Technologies

2 - Cálculo do grau de conexão de cada ator

FREEMAN'S DEGREE CENTRALITY MEASURES

Diagonal valid? NO
Model: ASYMMETRIC
Input dataset: C:\Arquivos de programas\Ucinet 6\DataFiles\HelpdeskFinal

	1 OutDegree	2 InDegree	3 NrmOutDeg	4 NrmInDeg
25 Mauro	19.000	15.000	28.788	22.727
7 Cristiano	19.000	18.000	28.788	27.273
21 Mario	17.000	17.000	25.758	25.758
24 Monica	16.000	22.000	24.242	33.333
23 Lia	16.000	24.000	24.242	36.364
1 Alfredo	14.000	11.000	21.212	16.667
3 Antonio	13.000	12.000	19.697	18.182
20 Luis	11.000	12.000	16.667	18.182
17 Luma	9.000	10.000	13.636	15.152
11 Filipe	8.000	6.000	12.121	9.091
29 Pepe	8.000	10.000	12.121	15.152
2 Acacio	6.000	5.000	9.091	7.576
6 AGomes	6.000	4.000	9.091	6.061
10 Euclerio	6.000	7.000	9.091	10.606
55 Jane	6.000	0.000	9.091	0.000
26 Milene	6.000	6.000	9.091	9.091
48 Antonia	6.000	2.000	9.091	3.030
61 Lauro	5.000	9.000	7.576	13.636
8 David	5.000	2.000	7.576	3.030
13 Gustavo	5.000	8.000	7.576	12.121
4 Arlindo	4.000	2.000	6.061	3.030
47 Anelise	4.000	0.000	6.061	0.000
19 Lucas	4.000	4.000	6.061	6.061
16 Lourdes	4.000	4.000	6.061	6.061
57 Licia	4.000	10.000	6.061	15.152
18 Katia	4.000	2.000	6.061	3.030
27 Nanci	4.000	4.000	6.061	6.061
32 UsuCli1	4.000	2.000	6.061	3.030
49 Carmen	4.000	1.000	6.061	1.515
30 Rodrigo	4.000	2.000	6.061	3.030
64 Pedro	4.000	3.000	6.061	4.545
15 João	4.000	5.000	6.061	7.576
63 Noelia	4.000	0.000	6.061	0.000

67 Vitorio	4.000	4.000	6.061	6.061
58 Lidia	4.000	4.000	6.061	6.061
50 Carla	4.000	0.000	6.061	0.000
53 Guido	3.000	4.000	4.545	6.061
33 UsuADN	3.000	3.000	4.545	4.545
39 Waldemar	3.000	5.000	4.545	7.576
56 Jose	3.000	7.000	4.545	10.606
22 MCarol	3.000	5.000	4.545	7.576
31 Telma	3.000	4.000	4.545	6.061
54 Ines	3.000	0.000	4.545	0.000
66 UsuADN1	3.000	3.000	4.545	4.545
5 Augusto	2.000	2.000	3.030	3.030
9 Denise	2.000	1.000	3.030	1.515
35 UsuCli3	2.000	2.000	3.030	3.030
45 Alan	2.000	5.000	3.030	7.576
41 Geraldo	2.000	3.000	3.030	4.545
60 Luzia	2.000	4.000	3.030	6.061
59 Leila	2.000	1.000	3.030	1.515
46 Alice	2.000	1.000	3.030	1.515
51 Camile	2.000	1.000	3.030	1.515
42 Hiran	2.000	3.000	3.030	4.545
14 Igor	2.000	2.000	3.030	3.030
36 UsuCli4	1.000	1.000	1.515	1.515
12 Graça	1.000	3.000	1.515	4.545
52 Carlos	1.000	3.000	1.515	4.545
37 UsuCli5	1.000	1.000	1.515	1.515
44 Aldo	1.000	1.000	1.515	1.515
28 Oto	1.000	3.000	1.515	4.545
38 UsuCli6	1.000	1.000	1.515	1.515
43 Marta	1.000	1.000	1.515	1.515
40 Flavio	1.000	1.000	1.515	1.515
65 Rinaldo	1.000	4.000	1.515	6.061
62 Marcio	1.000	1.000	1.515	1.515
34 UsuCli2	1.000	0.000	1.515	0.000

DESCRIPTIVE STATISTICS

	1 OutDegree	2 InDegree	3 NrmOutDeg	4 NrmInDeg
1 Mean	4.821	4.821	7.304	7.304
2 Std Dev	4.465	5.140	6.765	7.787
3 Sum	323.000	323.000	489.394	489.394
4Variance	19.938	26.416	45.772	60.642
5 SSQ	2893.000	3327.000	6641.414	7637.741
6 MCSSQ	1335.851	1769.851	3066.691	4063.018
7Euc Norm	53.787	57.680	81.495	87.394
8 Minimum	1.000	0.000	1.515	0.000
9 Maximum	19.000	24.000	28.788	36.364

Network Centralization (Outdegree) = 21.809%
Network Centralization (Indegree) = 29.500%

Actor-by-centrality matrix saved as dataset FreemanDegree

Running time: 00:00:01
Output generated: 02 set 05 14:24:11
Copyright (c) 1999-2005 Analytic Technologies

3 - Cálculo do grau de intermediação de cada ator

FREEMAN BETWEENNESS CENTRALITY

Input dataset: C:\Arquivos de programas\Ucinet 6\DataFiles\HelpDeskFinal
Important note: this routine binarizes but does NOT symmetrize.
Un-normalized centralization: 81922.260

		1	2
		Betweenness	nBetweenness
24	Monica	1326.183	30.913
23	Lia	1320.343	30.777
3	Antonio	652.568	15.211
21	Mario	494.933	11.537
7	Cristiano	451.988	10.536
20	Luis	369.918	8.623
29	Pepe	310.399	7.235
17	Luma	284.123	6.623
25	Mauro	279.445	6.514
1	Alfredo	166.182	3.874
67	Vitorio	146.334	3.411
2	Acacio	135.999	3.170
13	Gustavo	134.056	3.125
64	Pedro	109.427	2.551
61	Lauro	107.594	2.508
10	Euclerio	101.883	2.375
60	Luzia	70.912	1.653
56	Jose	66.200	1.543
16	Lourdes	57.728	1.346
18	Katia	52.453	1.223
53	Guido	37.252	0.868
31	Telma	33.374	0.778
19	Lucas	28.900	0.674
11	Filipe	26.624	0.621
57	Licia	25.260	0.589
26	Milene	21.478	0.501

48	Antonia	19.233	0.448
58	Lidia	16.479	0.384
6	AGomes	16.259	0.379
32	UsuCli1	14.487	0.338
15	João	10.021	0.234
65	Rinaldo	7.602	0.177
33	UsuADN	6.889	0.161
49	Carmen	5.601	0.131
22	MCarol	5.377	0.125
27	Nanci	4.387	0.102
4	Arlindo	4.298	0.100
28	Oto	2.559	0.060
39	Waldemar	2.096	0.049
45	Alan	1.921	0.045
9	Denise	1.661	0.039
8	David	0.398	0.009
14	Igor	0.392	0.009
40	Flavio	0.392	0.009
5	Augusto	0.392	0.009
30	Rodrigo	0.000	0.000
35	UsuCli3	0.000	0.000
38	UsuCli6	0.000	0.000
43	Marta	0.000	0.000
46	Alice	0.000	0.000
47	Anelise	0.000	0.000
50	Carla	0.000	0.000
51	Camile	0.000	0.000
54	Ines	0.000	0.000
55	Jane	0.000	0.000
36	UsuCli4	0.000	0.000
37	UsuCli5	0.000	0.000
42	Hiran	0.000	0.000
59	Leila	0.000	0.000
52	Carlos	0.000	0.000
12	Graça	0.000	0.000
62	Marcio	0.000	0.000
63	Noelia	0.000	0.000
44	Aldo	0.000	0.000
41	Geraldo	0.000	0.000
66	UsuADN1	0.000	0.000
34	UsuCli2	0.000	0.000

DESCRIPTIVE STATISTICS FOR EACH MEASURE

		1	2
		Betweenness	nBetweenness
1	Mean	103.463	2.412
2	Std Dev	250.178	5.832

3	Sum	6932.000	161.585
4	Variance	62589.066	34.008
5	SSQ	4910671.000	2668.248
6	MCSSQ	4193467.500	2278.551
7	Euc Norm	2216.003	51.655
8	Minimum	0.000	0.000
9	Maximum	1326.183	30.913

Network Centralization Index = 28.93%

Output actor-by-centrality measure matrix saved as dataset FreemanBetweenness

---
Running time: 00:00:01
Output generated: 02 set 05 14:29:14
Copyright (c) 1999-2005 Analytic Technologies

4 - Cálculo do grau de proximidade de cada ator

CLOSENESS CENTRALITY

Input dataset: C:\Arquivos de programas\Ucinet 6\DataFiles\HelpDeskFinal
Method: Geodesic paths only (Freeman Closeness)
Output dataset: C:\Redes\Mestrado\Redes Sociais\Produtos apresentação\Anexos\Closeness

Note: Data not symmetric, therefore separate in-closeness & out-closeness computed.

The network is not connected. Technically, closeness centrality cannot be computed, as there are infinite distances.

Closeness Centrality Measures

	1 inFarness	2 outFarness	3 inCloseness	4 outCloseness
23 Lia	118.000	510.000	55.932	12.941
24 Monica	124.000	522.000	53.226	12.644
3 Antonio	134.000	543.000	49.254	12.155
13 Gustavo	136.000	547.000	48.529	12.066
17 Luma	138.000	534.000	47.826	12.360
21 Mario	140.000	538.000	47.143	12.268
7 Cristiano	144.000	528.000	45.833	12.500
25 Mauro	146.000	534.000	45.205	12.360
20 Luis	147.000	547.000	44.898	12.066
26 Milene	150.000	541.000	44.000	12.200
61 Lauro	153.000	574.000	43.137	11.498

1	Alfredo	153.000	542.000	43.137	12.177
10	Euclerio	157.000	551.000	42.038	11.978
29	Pepe	159.000	534.000	41.509	12.360
6	AGomes	162.000	544.000	40.741	12.132
31	Telma	166.000	563.000	39.759	11.723
66	UsuADN1	169.000	552.000	39.053	11.957
28	Oto	171.000	593.000	38.596	11.130
2	Acacio	173.000	538.000	38.150	12.268
11	Filipe	174.000	539.000	37.931	12.245
22	MCarol	175.000	575.000	37.714	11.478
15	João	175.000	572.000	37.714	11.538
19	Lucas	176.000	559.000	37.500	11.807
57	Licia	176.000	599.000	37.500	11.018
41	Geraldo	177.000	558.000	37.288	11.828
18	Katia	177.000	549.000	37.288	12.022
42	Hiran	177.000	558.000	37.288	11.828
8	David	178.000	554.000	37.079	11.913
16	Lourdes	179.000	545.000	36.872	12.110
30	Rodrigo	179.000	539.000	36.872	12.245
56	Jose	181.000	577.000	36.464	11.438
65	Rinaldo	182.000	609.000	36.264	10.837
62	Marcio	182.000	580.000	36.264	11.379
45	Alan	183.000	580.000	36.066	11.379
58	Lidia	183.000	576.000	36.066	11.458
53	Guido	183.000	600.000	36.066	11.000
33	UsuADN	184.000	575.000	35.870	11.478
64	Pedro	185.000	596.000	35.676	11.074
60	Luzia	185.000	554.000	35.676	11.913
39	Waldemar	187.000	553.000	35.294	11.935
27	Nanci	188.000	546.000	35.106	12.088
59	Leila	189.000	580.000	34.921	11.379
14	Igor	195.000	583.000	33.846	11.321
5	Augusto	195.000	579.000	33.846	11.399
32	UsuCli1	195.000	551.000	33.846	11.978
12	Graça	197.000	587.000	33.503	11.244
36	UsuCli4	199.000	602.000	33.166	10.963
38	UsuCli6	199.000	602.000	33.166	10.963
43	Marta	199.000	602.000	33.166	10.963
52	Carlos	201.000	587.000	32.836	11.244
40	Flavio	202.000	592.000	32.673	11.149
35	UsuCli3	202.000	584.000	32.673	11.301
4	Arlindo	204.000	552.000	32.353	11.957
9	Denise	205.000	554.000	32.195	11.913
37	UsuCli5	205.000	597.000	32.195	11.055
46	Alice	205.000	580.000	32.195	11.379
67	Vitorio	208.000	551.000	31.731	11.978
44	Aldo	212.000	606.000	31.132	10.891
48	Antonia	245.000	573.000	26.939	11.518
51	Camile	246.000	579.000	26.829	11.399
49	Carmen	249.000	575.000	26.506	11.478

47 Anelise	4422.000	511.000	1.493	12.916
55 Jane	4422.000	483.000	1.493	13.665
54 Ines	4422.000	488.000	1.493	13.525
63 Noelia	4422.000	511.000	1.493	12.916
50 Carla	4422.000	512.000	1.493	12.891
34 UsuCli2	4422.000	541.000	1.493	12.200

Statistics

	1	2	3	4
	inFarness	outFarness	inCloseness	outCloseness
1 Mean	559.552	559.552	34.366	11.827
2 Std Dev	1211.628	28.492	11.758	0.618
3 Sum	37490.000	37490.000	2302.496	792.379
4 Variance	1468043.500	811.799	138.256	0.382
5 SSQ	119336528.000	21032004.000	88389.820	9396.738
6 MCSSQ	98358912.000	54390.566	9263.124	25.618
7 Euc Norm	10924.126	4586.066	297.304	96.937
8 Minimum	118.000	483.000	1.493	10.837
9 Maximum	4422.000	609.000	55.932	13.665

Network centralization not computed for unconnected graphs
Output actor-by-centrality measure matrix saved as dataset C:\Redes\Mestrado\Redes Sociais\Produtos apresentação\Anexos\Closeness

---
Running time: 00:00:01
Output generated: 02 set 05 14:31:52
Copyright (c) 1999-2005 Analytic Technologies