

UMA ESTRATÉGIA PARA VISUALIZAÇÃO DA INFORMAÇÃO GERADA A PARTIR DE DADOS ALTAMENTE CONECTADOS

Deusdete Vieira Inácio, Alexandre Cardoso e Edgard Lamounier Jr.

Universidade Federal de Uberlândia, Faculdade de Engenharia Elétrica

Programa de Pós-graduação em Engenharia Elétrica

Uberlândia - MG, Brasil.

deusdete@iftm.edu.br, {alexandre, lamounier}@ufu.br

Resumo – Para que se possa explorar os dados em toda sua granularidade a fim de localizar padrões, tendências e gerar conhecimento, deve-se considerar o aumento do uso dos recursos computacionais, uma vez que os mesmos são usados para armazenar desde transações complexas a simples logs de usuários. Isso faz o número de conexões entre os dados aumentar consideravelmente, neste sentido, é necessário buscar novas formas e métodos de visualização da informação para que se possa apresentá-los ao usuário sem a perda de conectividade, ou até mesmo que informações importantes sejam ocultadas. Este trabalho apresenta um sistema que utiliza uma estratégia de visualização, gerada a partir de dados altamente conectados, que acessa um SGBD de uma instituição de ensino, explora, cria a representação visual e apresenta a informação do ponto de interesse ao usuário.

Palavras-Chaves: altamente conectados, rede de dados, visualização da informação, instituição de ensino.

A STRATEGY FOR INFORMATION VISUALIZATION GENERATED FROM HIGHLY CONNECTED DATA

Abstract - In order to use the data in all its granularity to locate patterns, trends and generating knowledge, one must consider the increasing use of computing resources, since they are used to store transactions from the single complex logs users. This makes the number of connections between the data increase considerably in this regard, it is necessary to seek new forms and methods for visualizing information so you can present them to the user without loss of connectivity, or even that important information is hidden. This paper presents a system that uses a visualization strategy, generated from highly connected data, which accesses a DBMS for an educational institution, explores, creates a visual representation and presents information from the user point of interest.

Keywords – data highly connected, network data, visualization of information, institution education

I. INTRODUÇÃO

O crescente uso dos recursos de informática no dia a dia da sociedade, coletando e armazenando dados, está fazendo com o que o volume de dados armazenado e suas conexões cresçam exponencialmente, seja de forma centralizada ou distribuída. Isso exige cada vez mais das técnicas de visualização da informação, pois estes dados podem se apresentar de forma não linear (conexões um para um ou um para n) e sim da forma n para n, com características especiais de ligação, invólucro ou contenção. Estas características e a heterogeneidade dos dados a serem recuperados podem provocar uma sobrecarga de informações que é um dos principais fatores que dificultam a recuperação de resultados relevantes à demanda do usuário,[13]. Os Sistemas de Recuperação da Informação (SRI) são os responsáveis por recuperar e disseminar informações relevantes provenientes de dados armazenados previamente, mas o crescente uso destes recursos e o aumento dos dados a serem interpretados, de fato, faz com que seja exigido cada vez mais destes sistemas. Nesse sentido, pesquisadores apontam como um grande desafio, a integração de diferentes áreas da computação para desenvolver soluções que trate, recupere e dissemine informação relevante, a partir de volumes de dados [1]. Este desafio é consoante com o objetivo da área de Visualização da Informação (V.I), uma área emergente da ciência que estuda formas de apresentar dados visualmente, de tal modo que relações entre os mesmos sejam melhor compreendidas ou novas informações possam ser descobertas, [6].

Estratégias têm sido propostas no sentido de resolver o problema de manipulação (entrada/saída/verificação) de dados com aplicação de técnicas adequadas de Visualização da Informação. Entretanto, os sistemas muitas das vezes apresentam as informações em formato estático que não oferecem interação com o usuário, essa forma de apresentação não permite explorar os dados contidos nas apresentações gráficas. Um exemplo clássico desse contexto pode ser observado na Figura 1.1. Nesta figura, apresenta-se o número de alunos de um determinado curso e os campi que os oferece. Neste caso, seria de muita valia para o gestor da instituição pública se ele pudesse a partir desde mesmo gráfico, ser provido de ferramentas que permitissem navegar por informações conexas tais como: quem são os alunos

matriculados neste curso, quais as matérias que eles cursam e até mesmo traçar perfis socioeconômicos dos cursos. Todas as interconexões, se exploradas de forma correta, fornece subsídios de suma importância para a tomada de decisão. Entretanto, sistemas atuais voltados para esta natureza de gestão estão desprovidos destas funcionalidades [3].

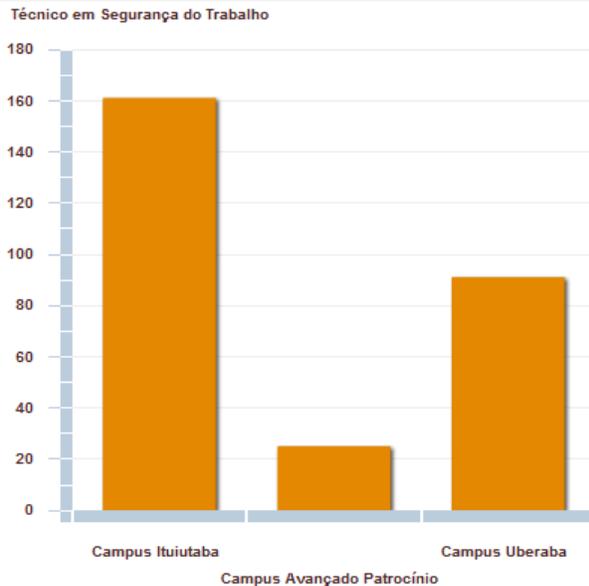


Figura 1.1 Gráfico hipotético que mostra a relação do nº de alunos de um determinado curso do Instituto Federal do Triângulo Mineiro

A ampla disponibilidade de computadores combinado com os avanços da computação gráfica, torna possível o acesso aos sistemas, interagindo, explorando e manipulando representações visuais, que podem se apresentar de variadas formas, em tempo real.

Neste sentido, esta pesquisa propõe investigar estratégias computacionais que possibilitem explorar dados brutos de um determinado domínio de aplicação e todas as suas conexões e ainda, integrar técnicas de Visualização da Informação com sistemas de gerenciamento de banco de dados (SGDB), a fim de suportar, de maneira mais eficiente, Sistemas de Recuperação de Informação.

II. FUNDAMENTOS

A. Visualização da Informação

A visualização é uma atividade que os seres humanos constroem de forma interna na mente, [8], é algo que não pode ser impresso em papel ou exibido em uma tela de computador, sendo um processo cognitivo. Os computadores podem facilitar o processo com algumas ferramentas e técnicas de representação, mas o processo de visualização continua a ser uma atividade que ocorre na mente humana.

Dos computadores, podemos extrair dados estatísticos, financeiros, contábeis, mas isolados. A Informação é um produto da mente humana, que com a ajuda de algum dado, adiciona a este elemento o seu conhecimento pessoal e outras informações para gerar o produto de que o sujeito precisa,

[7]. A informação é criada a partir dos dados representados de forma visual, estes dados podem ser em grande quantidade e de numerosas fontes. Além disso, a informação pode ser transformada em conhecimento e, finalmente, em sabedoria, [8], veja figura 2.1, logo a seguir serão exemplificadas as etapas descritas.

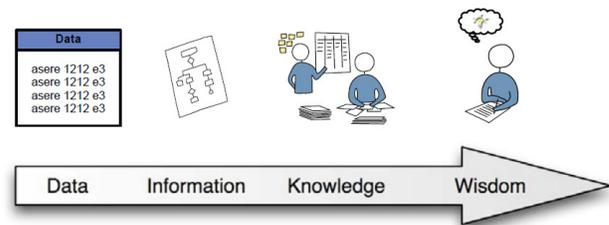


Figura 2.1 – Do dado a sabedoria, [8].

▪ Dados (*Data*)

Os dados sozinhos não são suficientes para estabelecer um processo comunicativo, eles devem primeiramente ser processados, organizados e apresentados em um formato adequado. Esta transformação, apresentação e a comunicação ao seu redor produzem a **informação (Information)**.

▪ Conhecimento (*knowledge*)

Quando a informação é integrada a experiência, se cria conhecimento, pois só adquirimos o conhecimento com que somos capazes de entender, este é objetivo principal de qualquer processo de comunicação.

▪ Sabedoria (*wisdom*)

Sabedoria é o mais alto nível de compreensão, pode ser definida como a fase em que uma pessoa adquiriu um nível avançado de conhecimento dos processos e relacionamentos, [8], a sabedoria é auto induzida através da contemplação, o estudo e interpretação do conhecimento, mas, ao contrário do conhecimento, não pode ser diretamente transmitido ou ensinado.

Seguindo esta ótica, o homem pode buscar a informação, acrescentar a ela seus próprios conhecimentos e extrair o produto que é a tomada de decisão. Para que isso ocorra são necessários formas e métodos mais claros e objetivos de apresentação dos dados.

B. Rede de Dados

Os dados organizados em uma estrutura de rede de dados podem ser representados naturalmente por grafos, onde as representações visuais em que os pontos, chamados de nós ou vértices, representam instancias dos dados que são correlacionados por ligações que indicam as relações entre as instancias. As arestas dos grafos podem também ter direção e valores, que são tratados como pesos no caso de valores numéricos ou etiquetas caso textuais (Figura 2.2).

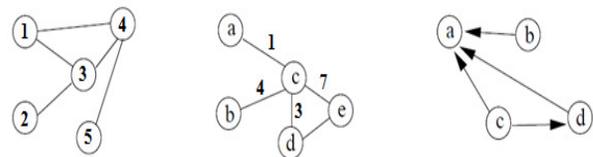


Figura 2.2 - Representações simples de estruturas de redes de dados.

Este tipo de representação pode se tornar crítica se o dimensionamento não for satisfatório e o número de nós e arestas forem muito elevados, pois bordas e nós sobrepostos podem tornar a representação visual ilegível. Quando isso ocorre e o número de nós é muito grande, estas técnicas podem mostrar seus limites, tanto para espaço reduzido de apresentação ou para complexidade dos algoritmos de disposição, impactando também no tempo de processamento computacional. Assim, a única opção que se tem é criar uma versão reduzida do gráfico o que pode sacrificar um pouco a informação, mas ao mesmo tempo preserva a estrutura geral dos dados, permitindo a identificação dos padrões principais. Como exemplo de técnicas de aperfeiçoamento do gráfico a ser mostrado pode-se citar uma técnica simples chamada redução de link [5], que consiste em visualizar apenas as arestas com pesos acima de um determinado valor, ou que atendem a critérios específicos, desta forma, apenas as arestas que poderia ser de interesse para o usuário são representados.

III. TRABALHOS CORRELATOS

Esta seção apresenta alguns trabalhos relacionados a esta pesquisa dos quais foram analisados as características de facilidade na navegação da informação, interatividade, utilização de outras técnicas de visualização, bem como a parte intuitiva.

A. VIZSTER (visualizing online social networks)

O Vizster, Figura 3.1, é uma ferramenta de visualização para redes sociais online, permitindo a exploração da estrutura da comunidade de serviços de redes sociais como friendster.com [10], tribe.net [11], e orkut [12]. Tais serviços fornecem os meios pelos quais os usuários podem articular publicamente a sua "amizade" comum sob a forma de ligações de amizade, formando um grafo não direcionado no qual os usuários são os nós e as ligações de amizade são as bordas.

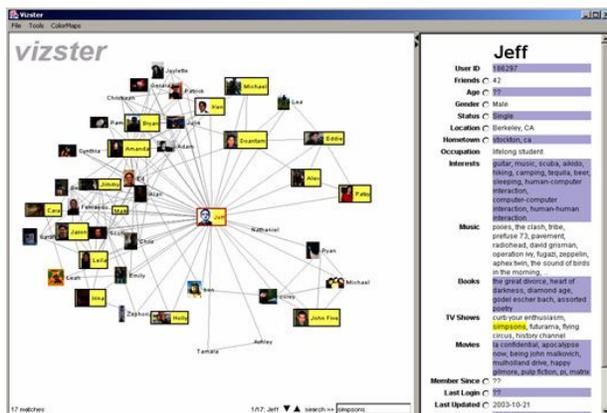


Figura 3.1 – Sociograma Vizster, [4].

Esses serviços também permitem que os usuários descrevam a si mesmos em um perfil, incluindo atributos como idade, estado civil, orientação sexual e interesses diversos.

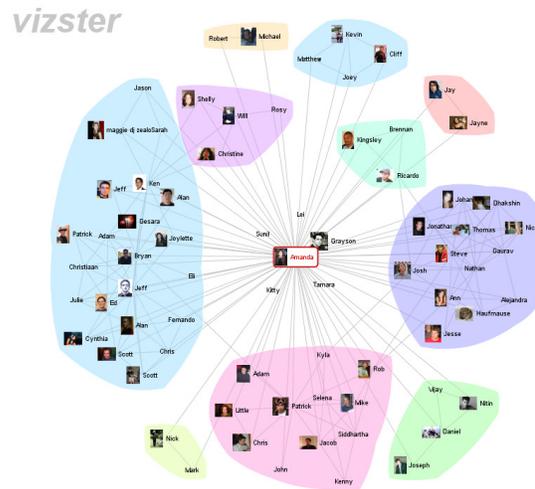


Figura 3.2 – Visualização dos grupos e comunidades de um indivíduo no Vizster [4].

O Vizster se baseia em layouts nó-link de rede e contribui com técnicas personalizadas para explorar a conectividade em estruturas de gráficos grandes, dando apoio à procura e análise visual, e automaticamente identificar e visualizar as estruturas comunitárias.

B. Net Mining and Social Network Analysis

Esta técnica foi usada com o intuito de fornecer informações em análise de redes sociais, conhecida como sociograma, onde a representação visual mostra as conexões entre os atores, conforme Figura 3.4, [3].

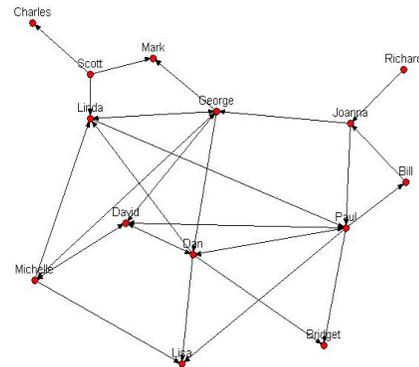


Figura 3.4 – Estrutura de um Sociograma, [3].

Na construção desta técnica, é ressaltado que os atores são tipicamente indivíduos, mas podem também ser empresas ou organizações, dependendo do nível em que a análise está sendo conduzida [3]. Como exemplo, foi aplicado esta estratégia para mapear as edições de páginas em wiki pages (Figura 3.5), com intuito de mapear a uma relação entre editores, indicando através de edição de mesma pagina uma relação de afinidade.

Com isso poderia sugerir aos editores participar de um determinado grupo por afinidade, direcionar avisos ou

propagandas da área de interesse entre outros, tudo isso com base nas relações dos dados do ponto de interesse.

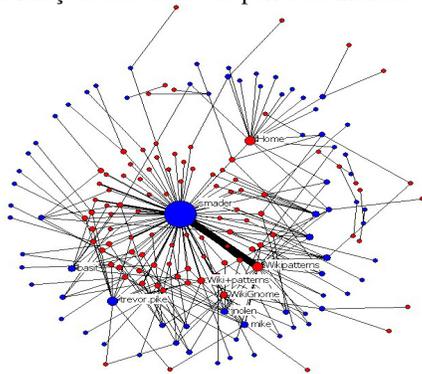


Figura 3.5 - Associação de páginas – Wik Editor, [3].

No exemplo da Figura 3.2 os nós em azul identificam os editores de páginas e/ou criadores, os vermelhos são páginas. Neste contexto, pode-se observar que à medida que a busca das informações cresce, a estratégia de visualização aqui utilizada tende a ser não intuitiva do ponto de vista do usuário, em função do acúmulo de imagens apresentadas, ocorrendo sobreposição de informação.

C. Intermap

O sistema InterMap, faz uso da estratégia de Redes de Dados (Seção 2.2), a fim de representar graficamente as interações ocorridas entre participantes do ambiente de aprendizado TelEduc [8].

No exemplo da Figura 3.6, os vértices de um grafo representam participantes de um curso oferecido via TelEduc. As arestas representam o envio de mensagens entre esses participantes na ferramenta de correio em um determinado período. Isto gera uma figura que traz à tona a informação de trocas de mensagens entre participantes do curso.

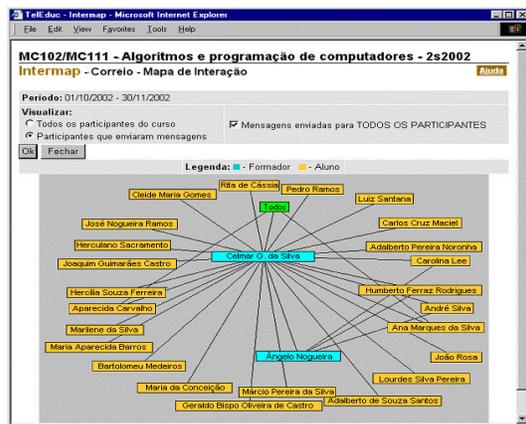


Figura 3.6 - Interação entre participantes de um curso, [9].

O InterMap descreve as informações de forma clara e objetiva quanto ao ponto de interesse. Entretanto, sua representação é estática não permitindo interação com o gráfico criado. Além disso, pode ocorrer, em caso de uma consulta, o retorno de muitos nós, gerando sobreposição de informação com dificuldade de navegação. Finalmente, o

sistema permite gerar imagens de tamanho não suportado na área de visualização do usuário, o que prejudicaria o entendimento.

IV. ARQUITETURA PROPOSTA

A criação da estratégia proposta foi criada a partir de estudos de técnicas de visualização, levando-se em conta as características dos dados armazenados e o ponto de interesse do usuário. Foram avaliadas estratégias aplicadas a estruturas de dados com características semelhantes, no intuito de verificar suas contribuições no contexto proposto, bem como a tecnologia utilizada e também o nível de interatividade.

Com base nas informações identificadas e com o auxílio de *stakeholders*, iniciou-se a elaboração de um modelo estrutural para a criação do sistema, responsável pela extração e apresentação da informação do ponto de interesse. A Figura 4.1 apresenta de forma geral a arquitetura do sistema proposto.

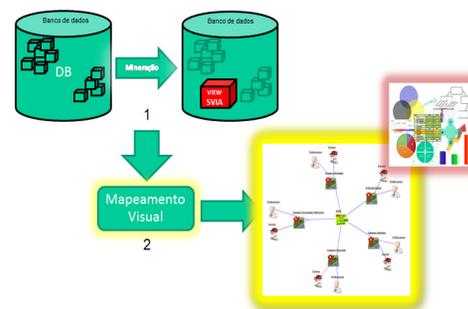


Figura 4.1 - Arquitetura do sistema.

Como pode ser observado, o modelo proposto é baseado em três principais componentes. São eles:

1 - Mineração dos dados: extração, tratamento e conversão dos dados brutos em um formato estruturado para serem utilizados pelo software de visualização, representado em uma visão de mesmo nome do sistema nome SVIA (Sistema de Visualização de Indicadores Acadêmicos).

Esta etapa do processo é a responsável por filtrar os dados, fornecendo ao software SVIA dados já refinados, evitando que conexões não tratadas no contexto da aplicação sejam geradas, o que dificultaria a representação, manipulação e a compreensão dos dados apresentados.

2 - Mapeamento Visual das estruturas: define-se o dimensionamento e as conexões, a fim de alimentar os componentes que gerarão a apresentação gráfica.

3 - Representação Visual da Informação: resultado do mapeamento das estruturas do ponto de interesse.

A. A tecnologia Utilizada

Para criação dos componentes visuais da aplicação utilizou-se o Adobe Flash Builder, pela facilidade de se criar representações visuais dinâmicas, interativas e o funcionamento das aplicações geradas, por este, ser baseado em navegadores de internet que são independentes da plataforma (sistemas operacionais) utilizada. Integrou-se a aplicação scripts na linguagem PHP, responsáveis pela

conexão e extração dos dados da view mineradora SVIA, criada previamente no SGBD da instituição IFTM.

B. Apresentação da Ferramenta

As estruturas administrativas de ensino necessitam cada vez mais de ferramentas que lhe forneçam suporte a tomada de decisão, e os dados contidos em seus SGBDs possuem grande interconectividade, o que oferece um cenário ideal para um estudo de caso. Neste contexto, foi escolhido o Instituto Federal do Triângulo Mineiro e, com base nas pesquisas levantadas e análise das estratégias correlatas, criou-se uma ferramenta que utiliza a técnica de redes de dados para navegação e, mesclando técnicas usuais de Visualização da Informação, como gráfico de pizza, barras, área e linha do tempo, facilitam a apresentação e a manipulação dos dados do ponto de interesse do usuário.

A Figura 4.2 abaixo mostra a interface do sistema, onde a interatividade foi fator primordial que norteou o seu desenvolvimento, assim como a integração com outras técnicas de visualização da informação que dão suporte à tomada de decisão.

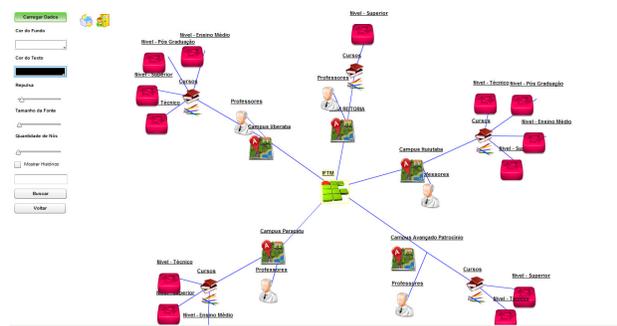


Figura 4.2 – Tela principal da aplicação SVIA.

A estrutura de navegação conta com um nó central representando a instituição IFTM, a esta, são conectados seus Campi e os demais pontos de interesse. Com base nas necessidades da administração, que foram definidas pelo *stakeholder*.

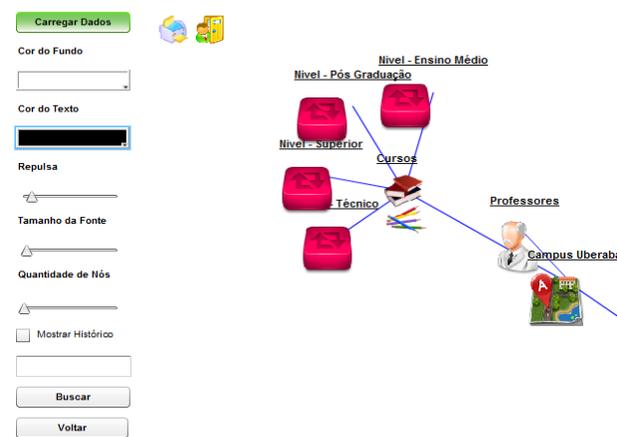


Figura 4.3 – Painel de controle da aplicação.

Em cada nó representante de um determinado ponto de interesse é possível explorar a granularidade da informação, seja por meio de gráficos, tabelas ou novos nós gerados na representação (veja figura 4.4).

Na tela principal da aplicação é possível também, alterar algumas propriedades visuais da apresentação, Figura 4.3, como tamanho da fonte, repulsão (distância) entre os nós apresentados, bem como a quantidade de nós a ser mostrado ao mesmo tempo na área de visualização. As alterações aplicadas a essas características são visíveis em tempo de execução da aplicação, não sendo necessário redesenhar os objetos do cenário e nem recarregar os dados da aplicação.

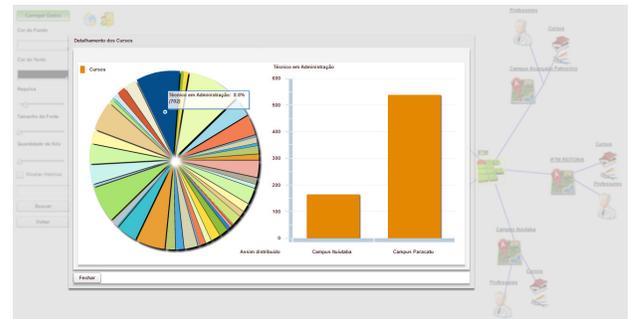


Figura 4.4 – Integração da Navegação com outras técnicas de VI.

A estratégia de Redes de Dados aqui aplicada, fornece navegação pelos pontos de interesse da administração, possibilitando a interatividade com o usuário e a visualização das informações de forma gráfica com apenas cliques de mouse ou telas tangíveis. Isto providencia uma maior flexibilização de sua utilização, uma vez que a tecnologia empregada na aplicação é multiplataforma e poucas ou, em alguns casos, nenhuma informação de entrada é exigida, sendo necessários apenas poucos cliques do mouse. Abaixo na Figura 4.5 pode-se observar um exemplo da funcionalidade de aumento de número de nós, onde a visão da rede de dados torna-se evidente, mas a informação é quase que nula, pois o algoritmo de disposição é forçado a apresentar um número de informação muito elevado.

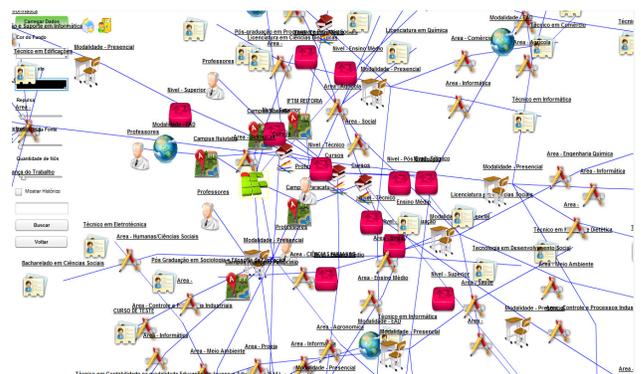


Figura 4.5 – Expansão do número de nós.

O sistema filtra, já na mineração de dados, os pontos de interesse do usuário evitando a representação de ligações de dados existentes, mas desnecessárias ao ponto de interesse,

permitindo também caso necessário explorar a níveis mais detalhados as conexões omitidas.

Toda a área visualizada pode ser arrastada de forma a possibilitar interação com todos os nós da rede de dados.

De fato, a ferramenta foi testada com Coordenadores de curso nos Campi do IFTM e os mesmo confirmaram que a estratégia proposta neste artigo, apresenta de forma mais rápida e natural os resultados (ou relatórios) por eles requeridos. Ainda, os mesmos puderam confirmar que com uma quantidade significativamente menor de interações com o sistema, a informação desejada foi alcançada. Principalmente, quando comparado com métodos tradicionais.

V. CONCLUSÕES E TRABALHOS FUTUROS

Após analisar o uso da ferramenta por parte de potenciais usuários, pôde se observar que a mesma oferece mecanismos de busca mais interativos e mais rápidos quando comparado com os sistemas relacionados.

Tal funcionalidade é de grande importância para sistemas de gestão de informação, principalmente, aqueles que trabalham com o bem público. Isto porque é importante que um governo seja capaz de apresentar de forma eficiente e transparente a forma que sua receita é gerenciada. Na verdade, esta é uma premissa que toda sociedade espera de seus governantes.

As avaliações dos usuários foram positivas e o sistema lhes permitiu extrair informações importantes para suporte à tomada de decisão. Como trabalho futuro, os autores estão, atualmente, investigando as seguintes possibilidades:

- Expansão dos pontos de interesse.
- Criação de novos gráficos, tabelas e outras representações visuais, de acordo com os moldes da instituição foco.
- Inserção em tempo real de novos parâmetros durante o processo de busca.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Carvalho, A. C. P. de L. F. et al. (2006). *Grandes Desafios da Pesquisa em Computação no Brasil – 2006 – 2016*.
- [2] Freitas, C. M. D. S.; Chubachi, O. M.; Luzzardi, P. R. G.; Cavar. A.. *Introdução à Visualização de Informações*. RITA – Rev. de Informática Teórica e Aplicada, Instituto de Informática UFRGS, Porto Alegre, RS, v. VIII, n. 2, p. 143-158, outubro, 2001. <http://www.inf.ufrgs.br/~revista/docs/rita08/> (12/Janeiro/2012)
- [3] Lee, Lauren Lock.; Matheson, James; Mader, Stewart. (2006). *WikiMining, WikiNetworks and Wikinomics*. <http://www.optimice.com.au/publications.php> (11/Fevereiro/2012).
- [4] Heer, Jeffrey; Vizster: Visualizing Online Social Networks. <http://vis.berkeley.edu/papers/vizster/2005-Vizster-InfoVis.pdf> (12/Janeiro/2012).
- [5] Mazza, Riccardo. (2008) *Introduction to Information Visualization*. University of Lugano, Switzerland, Springer-Verlag London, 2008. ISBN: 978-1-84800-218-0 e-ISBN: 978-1-84800-219-7
- [6] Nascimento, H. A. do; Ferreira, C. B. R. *Visualização de Informações - Uma Abordagem Prática*. In: UNISINOS, 2005. Rio Grande do Sul. XXV Congresso da Sociedade Brasileira de Computação. Rio Grande do Sul: São Leopoldo, 2005, p. 1262-1312.
- [7] Ragioto, Maria Regina Theobaldo. (2007) *Tecnologia e a ecologia da informação*. http://200.159.236.207/users/59/10/1059/informacao_tecnologia.pdf (15/Março/2012)
- [8] Robert Jacobson, editor. *Information Design*. MIT Press, Cambridge, MA, 1999.
- [9] Silva, C. G. (2006). *Exploração de bases de dados de ambientes de Educação a Distância por meio de ferramentas de consulta apoiadas por Visualização de Informação*. Tese de Doutorado. Instituto de Computação, Universidade Estadual de Campinas. <http://www.ic.unicamp.br/~celmar/tese> (03/Dezembro/2011).
- [10] boyd, d. *Friendster and Publicly Articulated Social Networks*. CHI 2004, Vienna, Austria, 1279-1282.
- [11] JUNG: Java Universal Network/Graph Framework. <http://jung.sf.net/> (15/Março/2012)
- [12] Harrison, S., P. Dourish. *Re-place-ing Space: The Roles of Place and Space in Collaborative Systems*. CSCW 1996, Boston, MA, 67-76.
- [13] Souza, R. R. (2006). *Sistemas de Recuperação de Informação e Mecanismos de Busca na web: panorama atual e tendências*. *Perspectivas em ciência da informação*. Belo Horizonte, V.11, n. 2, p. 161-173.