

Exploração visual multidimensional de redes sociais

G. F. Andery, A. A. Lopes, and R. Minghim

Instituto de Ciências Matemáticas e de Computação - ICMC-USP
{gfandery,alneu,rminghim}@icmc.usp.br

Resumo Redes sociais são normalmente estudadas utilizando grafos como meio de representação visual, os quais destacam relacionamentos entre indivíduos e grupos de indivíduos. Este artigo apresenta três novas abordagens para apoio à exploração visual de redes sociais, propostas e implementadas para aumentar o poder de análise de abordagens baseadas em grafos para este fim. Resultados com dados reais demonstram as possibilidades de identificação de padrões relevantes, que seriam difíceis de identificar por outros métodos de análise exploratória.

Key words: rede social, grafo, projeção multidimensional, mineração visual

1 Introdução

Desde a década de trinta, pesquisadores em ciências sociais buscam compreender as relações humanas na sociedade com o auxílio de representações visuais. O advento do computador e, mais recentemente, da internet, possibilitou o surgimento de um campo que tem despertado a atenção de estudiosos das áreas de visualização de informação e de ciências sociais, o da visualização de redes sociais. A motivação para o estudo de redes sociais é variada, mas o princípio é comum. Existe uma grande quantidade de informações proveniente de relacionamentos humanos em *sites* de relacionamento que pode ser empregada para os mais variados fins, como busca de padrões de preferências para apoio aos setores de comércio e produção, detecção e investigação de atividades ilícitas, dentre outros.

A forma mais comum de representar redes sociais é por meio de grafos, ou seja, diagramas contendo vértices e arestas, os quais representam os indivíduos da rede e as relações entre eles, respectivamente. Grafos são interessantes por constituírem uma forma de representação muito mais expressiva que as tabelas atributo-valor e potencialmente mais eficiente que dados multi-relacionais [1]. Diversas ferramentas para análise visual de redes sociais utilizam essa representação, como as descritas em [2], [3] e [4], que destacam relacionamentos entre indivíduos e grupos de indivíduos. Entretanto, os dados provenientes de redes sociais também possuem atributos nos vértices e arestas, que podem variar desde valores numéricos até textos e imagens, além de relacionamentos múltiplos entre

os vértices. Essas características recebem pouca atenção por grande parte das ferramentas de análise de redes sociais existentes.

Neste artigo apresentamos três abordagens de exploração visual de redes sociais. A primeira abordagem utiliza um grafo híbrido, no qual os vértices representam mais de um tipo de objeto. A segunda abordagem projeta os indivíduos no plano a partir de uma matriz de similaridade entre os mesmos. A terceira abordagem cria duas visualizações simultâneas, utilizando as abordagens anteriores, e as coordena de modo que os indivíduos em uma sejam facilmente localizados na outra e associações entre redes possam ser estabelecidas, identificando informações contidas nos dados não visíveis por qualquer uma das abordagens individualmente.

A principal contribuição deste trabalho é adaptar três abordagens de mapeamento visual à análise de redes sociais em grafos, a saber: grafos híbridos, projeções multidimensionais e coordenação entre essas visualizações.

Os primeiros resultados obtidos são promissores, permitindo identificar padrões nos relacionamentos entre os objetos das redes não percebidos em outras abordagens.

O restante deste artigo está dividido como segue. A próxima Seção descreve os trabalhos relacionados à análise visual de redes sociais. Alguns conceitos básicos e as abordagens propostas são detalhados na Seção 3. A Seção 4 apresenta estudos de casos usando as abordagens propostas. Por fim, a Seção 5 descreve as conclusões e propõe trabalhos futuros.

2 Trabalhos relacionados

Diversos trabalhos têm sido publicados descrevendo métodos para realizar exploração visual de redes sociais. A abordagem mais intuitiva e, conseqüentemente, a mais comum, utiliza grafos como forma de representação desse tipo de dado, pois redes sociais possuem a característica intrínseca de serem formadas por nós e arestas, sendo que os nós representam os indivíduos da rede e as arestas o relacionamento entre eles.

Algumas abordagens utilizando tais elementos têm sido desenvolvidas para analisar domínios específicos, como redes de co-autoria em publicações científicas ([5], [6]), comunicações por e-mail ([7], [8]), e comunidades virtuais em *sites* de relacionamento ([3], [4]).

Ferramentas menos específicas, para análise de domínios variados, são descritas em [2]. Dentre as ferramentas descritas estão o NetDraw, capaz de visualizar grandes redes sociais; o StoCNET, o MultiNet, o UCINet e o Agna, que realizam análises estatísticas sobre as redes sociais; o Blanche e o Condor, capazes de simular a evolução ocorrida nas redes; e uma biblioteca em R¹ para análises estatísticas em redes sociais e representações visuais das mesmas. Entretanto,

¹ R é uma ferramenta e uma linguagem de programação para computação estatística, e está disponível gratuitamente no endereço <http://www.r-project.org/>.

apesar do grande número de ferramentas existentes, poucas possuem funcionalidades para representar os atributos dos vértices e, as que possuem, se limitam a modificar suas propriedades visuais, como forma, cor e tamanho.

Um trabalho semelhante a uma das abordagens propostas neste artigo é descrito em [9]. Nele é permitida a construção de redes com vértices de tipos diferentes, definidos pelo usuário. Apesar de gerar resultados interessantes, essa ferramenta também não é capaz de destacar atributos dos vértices.

3 Exploração visual multidimensional de redes sociais

As abordagens propostas neste artigo visam possibilitar (i) a visualização de redes com diferentes tipos de vértices (ver Seção 3.2); (ii) a visualização de redes baseadas em projeção multidimensional dos vértices, quando os mesmos são descritos por um vetor de atributos – nesse caso, a proximidade entre os vértices é um indicativo da similaridade entre eles (ver Seção 3.3); e (iii) a análise simultânea das redes anteriores via coordenação (ver Seção 3.4). Para destacar a relevância dessas abordagens, a seguir são apresentados conceitos relativos à representação de conhecimento usando grafos.

3.1 Representação de conhecimento em grafos

A representação dos dados é fundamental na descoberta de padrões. Via de regra, os objetos tratados pelas técnicas de descoberta de padrões são representados em tabelas atributo-valor, existindo, no entanto, domínios e aplicações nos quais existem informações não apenas dos objetos, mas também das relações entre eles. Grafos permitem representar tanto os objetos quanto essas relações, situando-se, em termos de expressividade, entre as representações atributo-valor e relacional [1].

O exemplo a seguir, devido a Motta [10], ilustra a expressividade de uma representação baseada em grafos. Na Figura 1 é apresentado um diagrama Entidade-Relacionamento (DER) de um banco de dados de publicações científicas, destacando que um mesmo artigo pode citar e pode ser citado por muitos artigos, e um autor pode participar de muitos artigos assim como um artigo ter muitos autores. Na Tabela 1 são representadas as tabelas com informações dos artigos e dos autores, e uma tabela de relacionamento, representando co-autoria.

O conjunto de tabelas apresentado em Tabela 1 pode ser representado por grafos que modelam as relações de co-autoria. A Figura 2(a) apresenta um grafo cujos vértices representam os autores e as arestas co-autoria. A Figura 2(b) apresenta um grafo híbrido no qual os vértices em forma de círculo representam os autores, os vértices em forma de quadrado representam os artigos, e as arestas representam o relacionamento entre autores e artigos. O grafo híbrido é uma das propostas deste trabalho, apresentado em maiores detalhes na seção a seguir.

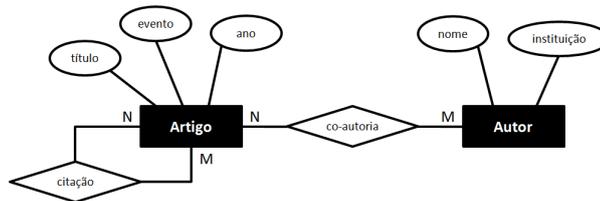


Figura 1. Diagrama Entidade-Relacionamento de um banco de dados de publicações científicas.

(a)

id	título	evento	ano
1	<i>Finding community structure in very large networks</i>	<i>Physical Review E</i>	2004
2	<i>Community structure in social and biological networks</i>	<i>PNAS</i>	2002
3	<i>The Power of Choice in Network Growth</i>	<i>European Journal of Physics B</i>	2007
4	<i>Explosive Percolation in Random Networks</i>	<i>Science</i>	2009

(b)		(c)	
id	nome	instituição	artigo
a	Aaron Clauset	<i>University of New Mexico</i>	1
b	Mark Newman	<i>University of Michigan</i>	1
c	Cristopher Moore	<i>University of New Mexico</i>	1
d	Michelle Girvan	<i>University of Michigan</i>	2
e	Raissa D'Souza	<i>University of California</i>	2
f	Paul Krapivsky	<i>Boston University</i>	3
g	Dimitris Achlioptas	<i>University of California</i>	3
h	Joel Spencer	<i>New York University</i>	3
			4
			4
			4
			4
			4
			4
			4

Tabela 1. Conjunto de tabelas do banco de dados de publicações científicas: (a) artigo, (b) autor e (c) co-autoria.

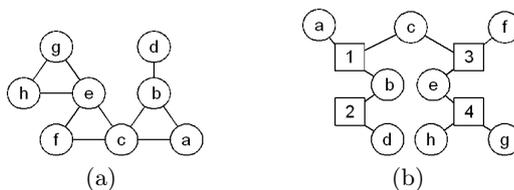


Figura 2. Rede de co-autoria: (a) apenas autores, (b) autores e artigos.

3.2 Grafos híbridos

Denominamos grafo híbrido um grafo no qual os vértices representam mais de um tipo de objeto. Com esse grafo, é possível representar dados relacionais, como tabelas n-m de bancos de dados, como ilustrado na Figura 2(b). Nesse caso, é possível visualizar tanto as arestas que conectam vértices do mesmo tipo quanto as que conectam vértices de tipos diferentes.

Ao utilizar esse grafo em conjunto com funcionalidades de interação típicas em análise de redes sociais, é possível realizar uma exploração muito mais ampla e eficiente, voltada aos interesses do usuário. Uma das funcionalidades mais interessantes nesse contexto é a possibilidade de criar a chamada *rede egocêntrica* de um vértice, que filtra a visualização deixando visível somente o vértice de interesse e aqueles que estão conectados a ele. Além disso, também é importante utilizar atributos visuais para os vértices, como forma, cor e tamanho, para representar os atributos dos dados.

3.3 Projeção multidimensional

Cada indivíduo de uma rede social, tipicamente representado por um vértice, pode ser descrito por um vetor de atributos, ou como um ponto em um espaço de alta dimensionalidade. O processo de projeção multidimensional reduz o número de dimensões do conjunto de dados, mapeando os mesmos em um espaço de baixa dimensionalidade, em geral, de uma a três dimensões, preservando ao máximo as relações de distância entre os indivíduos [11]. A segunda abordagem proposta utiliza esse processo para posicionar os indivíduos no plano de acordo com os valores de seus atributos.

Essa abordagem permite investigar grupos de indivíduos que possuem valores de atributos similares por sua proximidade no plano. Em uma situação na qual as conexões também estão relacionadas com a similaridade entre os vértices, esse posicionamento produzirá menor cruzamento entre as arestas, caso contrário, poderá haver muitos cruzamentos.

3.4 Coordenação

Múltiplas Visões Coordenadas (CMV - *Coordinated and Multiple Views*) refere-se à possibilidade de visualizar e interagir com várias visões simultaneamente, de modo que ações realizadas em uma delas sejam ecoadas para as demais, facilitando a exploração concomitante de padrões. Cada visão pode aplicar uma técnica diferente, tanto sobre o mesmo conjunto de dados quanto sobre conjuntos diferentes. Dessa forma, ao utilizar várias técnicas, as desvantagens de uma podem ser compensadas pelos pontos fortes das outras [12].

A terceira abordagem proposta aqui é a utilização de coordenação para visualizar simultaneamente uma rede social apresentada pelas duas visões geradas com as abordagens anteriores, adicionadas ou não a uma visão convencional baseada em grafos simples de redes. Dessa forma, a exploração permite identificar agrupamentos tanto por semelhança de objetos (considerando suas tabelas atributo-valor) quanto pelas conexões da rede (considerando as tabelas de relacionamentos entre objetos).

4 Estudo de caso

As abordagens apresentadas acima foram implementadas em uma ferramenta de código aberto denominada Projection Explorer (PEX). A ferramenta foi origi-

nalmente construída para realizar a projeção multidimensional de coleções de documentos de texto, tendo sido adaptada para visualizar coleções de imagens, séries temporais, dentre outros, possuindo diversas técnicas de projeção multidimensional e várias funcionalidades já implementadas, como coordenação e busca textual. Este trabalho estende a ferramenta para a visualização de redes sociais, aproveitando suas estruturas de dados e algumas características de exploração. A ela foram adicionadas também funcionalidades de grafos, como a possibilidade de alteração de atributos visuais dos vértices e geração de redes egocêntricas. Essas redes podem ser geradas interativamente, adicionando-se outras redes egocêntricas às já existentes.

A partir de um conjunto de dados sobre redes sociais do Orkut, diversas visualizações foram geradas por meio das abordagens propostas. O conjunto de dados contém uma tabela de indivíduos, constituída por um identificador e um atributo “sexo”, uma tabela de comunidades, formada por um identificador e um atributo “nome”, e uma tabela que relaciona os indivíduos às comunidades a que pertencem.

A Figura 3 exibe um grafo híbrido gerado com o conjunto de dados do Orkut, em que os vértices azuis representam comunidades, os vértices vermelhos indivíduos do sexo masculino, o vértice verde um indivíduo do sexo feminino, e as arestas conectam os indivíduos às suas comunidades. O posicionamento dos vértices foi gerado aleatoriamente. Para esse exemplo, inicialmente foi gerada a rede egocêntrica da comunidade “Palmeiras desde Criancinha”. Esse procedimento permitiu identificar que essa comunidade contém apenas um indivíduo do sexo feminino. Então, foi adicionada a rede egocêntrica desse indivíduo à rede anterior. Tal procedimento permite identificar suas comunidades, bem como as intersecções com a rede anterior.

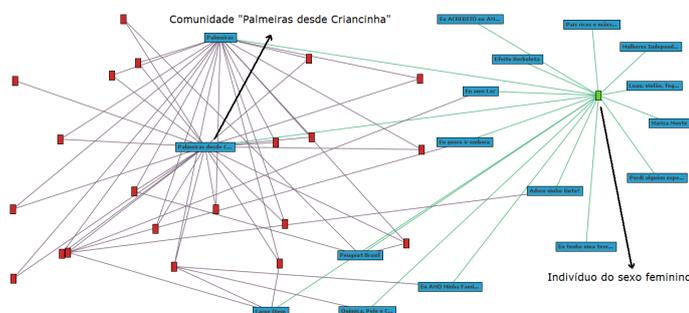


Figura 3. Grafo híbrido de uma rede social do Orkut, em que vértices representam indivíduos ou comunidades e as arestas conectam os indivíduos às comunidades.

Para realizar a projeção multidimensional, foi gerada uma matriz de adjacência de vértices de modo que o valor entre um par de indivíduos é “1” se ambos

estão em uma mesma comunidade, ou “0” caso contrário. Para o atributo “sexo”, foram adicionadas duas dimensões à matriz, uma para o valor “masculino” e outra para “feminino”, que assumem “0” ou “1” de acordo com o sexo do indivíduo. A matriz foi utilizada como tabela de dados estruturada para criar a projeção multidimensional apresentada na Figura 4, que aplica uma técnica denominada IDMAP. Nesse caso, os vértices representam apenas os indivíduos.

No conjunto de dados existem indivíduos em três diferentes comunidades relacionadas à marca Peugeot. Assim, os vértices foram coloridos de acordo com o número de comunidades contendo o termo “Peugeot” das quais os indivíduos fazem parte. Nesse caso, os indivíduos que não fazem parte de nenhuma dessas comunidades aparecem em branco; os indivíduos que pertencem a uma delas em bege; a duas, em verde escuro; e apenas um indivíduo que pertence a todas as três comunidades relacionadas à Peugeot, em preto. Como é possível observar, os vértices ficam separados entre os que estão em comunidades relacionadas à Peugeot e os que não estão. O resultado permite explorar e identificar de quais outras comunidades esses grupos de indivíduos fazem parte.

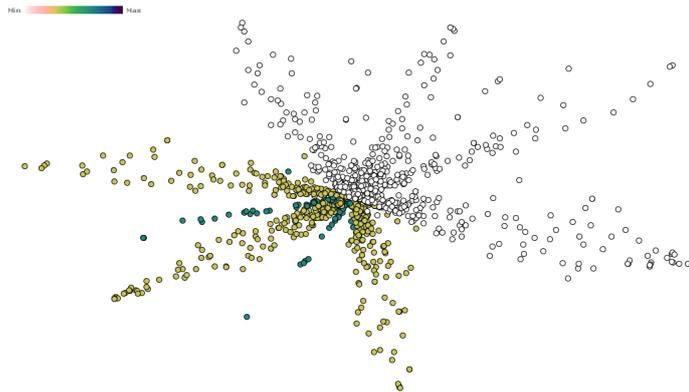


Figura 4. Projeção multidimensional a partir de uma rede social do Orkut, em que a cor indica a quantas comunidades relacionadas à Peugeot os indivíduos pertencem.

A Figura 5 mostra uma coordenação entre as duas abordagens anteriores. Na visão de grafo híbrido, à esquerda, foi criada a rede egocêntrica da comunidade “Eu odeio Fiat !!”, em que a cor indica o tipo do objeto, bege para comunidade e marrom para indivíduo. Na visão de projeção multidimensional, à direita, a cor é a mesma que na Figura 4, ou seja, indica a quantas comunidades relacionadas à Peugeot os indivíduos pertencem. Nesse caso, ao selecionar os membros da comunidade “Eu odeio Fiat!!” na visão de grafo híbrido, os mesmos foram destacados na visão de projeção multidimensional, permitindo verificar que nenhum deles pertence a qualquer comunidade relacionada à Peugeot.

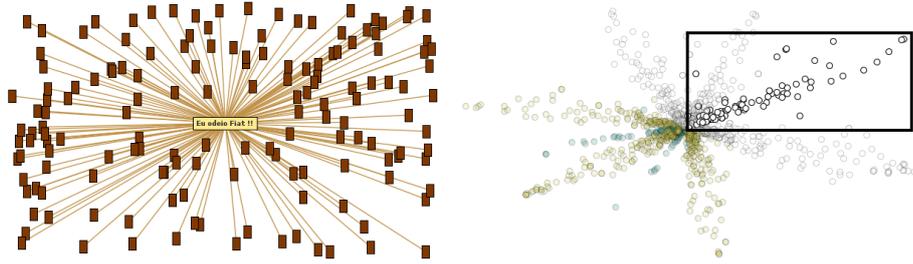


Figura 5. Coordenação entre as duas abordagens anteriores, em que os membros da comunidade “Eu odeio Fiat !!” foram destacados.

Segundo [13], ao analisar uma rede social, é importante identificar comunidades, atores centrais e analisar o papel e as posições dos atores e das conexões. Os dados analisados já possuem informações sobre as comunidades, seus membros e as relações entre eles, sendo que a abordagem de grafo híbrido permite destacar imediatamente todos os membros de qualquer comunidade. Com relação aos papéis e posições dos atores, a abordagem de projeção multidimensional busca justamente agrupar os indivíduos de acordo com a semelhança entre seus atributos. Além disso, ao explorar a projeção multidimensional, é possível perceber que os indivíduos ao centro estão em mais comunidades, sendo parecidos a um grande número de outros indivíduos, que também estão em algumas de suas comunidades. Portanto, o posicionamento na projeção multidimensional reflete, até certo ponto, o quão central é um indivíduo.

5 Conclusão e trabalhos futuros

As abordagens propostas possibilitam novas alternativas para exploração visual de redes sociais. Esses resultados iniciais sugerem que grafos híbridos auxiliam no processo de investigação dirigido, ou seja, quando o usuário tem idéia sobre o que procurar, enquanto que projeções multidimensionais podem fornecer uma visão geral sobre a rede, os indivíduos e seus atributos. Dessa forma, projeções multidimensionais podem ser utilizadas como ponto de partida em uma investigação. A coordenação auxilia o usuário a descobrir padrões de relacionamento não expressos diretamente.

Os próximos passos do trabalho envolvem ampliar o conjunto de métricas de similaridade para diferentes tipos de atributos, implementar técnicas de posicionamento para grafos híbridos e usar as características da rede para alimentar um processo de projeção multidimensional específico para este caso.

Referências

- [1] Washio, T., De Raedt, L., Kok, J.N.: Advances in mining graphs, trees and sequences. *Fundamenta Informaticae* **66**(1-2) (2005) 5–8

- [2] Huisman, M., Duijn, M.A.J.: Software for Social Network Analysis. In: *Models and Methods in Social Network Analysis*. Cambridge University Press (2005) 270–316
- [3] Heer, J., Boyd, D.: Vizster: Visualizing online social networks. In: *Proceedings of the IEEE Symposium on Information Visualization, 2005*, Washington, DC, USA, IEEE Computer Society (2005) 5
- [4] Mutton, P.: Inferring and visualizing social networks on internet relay chat. In: *Proceedings of the Eighth International Conference on Information Visualisation*. (2004) 35–43
- [5] Henry, N., Fekete, J.D.: Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics* **12**(5) (2006) 677–684
- [6] Henry, N., Fekete, J.D., McGuffin, M.: Nodetrix: a hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics* **13**(6) (2007) 1302–1309
- [7] Viegas, F., Boyd, D., Nguyen, D., Potter, J., Donath, J.: Digital artifacts for remembering and storytelling: posthistory and social network fragments. In: *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*. (2004) 10
- [8] Fisher, D., Dourish, P.: Social and temporal structures in everyday collaboration. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, ACM (2004) 551–558
- [9] Shen, Z., Ma, K.L., Eliassi-Rad, T.: Visual analysis of large heterogeneous social networks by semantic and structural abstraction. *IEEE Transactions on Visualization and Computer Graphics* **12**(6) (2006) 1427–1439
- [10] da Motta, R.C.: *Uso de redes complexas na classificação relacional*. Master’s thesis, ICMC - Universidade de São Paulo, São Carlos (2009)
- [11] Tejada, E., Minghim, R., Nonato, L.G.: On improved projection techniques to support visual exploration of multidimensional data sets. *Information Visualization* **2**(4) (2003) 218–231
- [12] Keim, D., Kriegel, H.: Visualization techniques for mining large databases: a comparison. *IEEE Transactions on Knowledge and Data Engineering* **8**(6) (1996) 923–938
- [13] Freitas, C.M.D.S., Nedel, L.P., Galante, R., Lamb, L.C., Spritzer, A.S., Fujii, S., de Oliveira, J.P.M., Araújo, R.M., Moro, M.M.: *Extração de conhecimento e análise visual de redes sociais*. In: *SEMISH - Seminário Integrado de Software e Hardware*, Belém do Pará, Brasil, SBC (2008) 106–120