

Base de Dados Orientada a Grafos: Um Experimento aplicado na Análise Social

Graphs-Oriented Databases: A Experiment applied on Social Analysis

Guilherme Martins Alvarez¹
Flávio Ceci²

Artigo recebido em para publicação em jun./2014 e aceito para publicação em out./2014

RESUMO

O grande desafio para os engenheiros de software é produzir tecnologias que consigam armazenar e analisar de forma eficiente a massa de dados gerada atualmente. Verificou-se, então, que o modelo relacional tem se mostrado ineficiente, pois o volume de dados e seus relacionamentos afetam diretamente o seu desempenho. Portanto, utilizar um modelo flexível, como o orientado a grafos, pode ser uma boa alternativa para a criação de modelos complexos de dados. Este trabalho tem por objetivo modelar e implementar uma base de dados orientada a grafos, a fim de identificar suas diferenças em relação à abordagem relacional. Para isso, foi elaborado um experimento baseado em análise social, no qual bases de dados foram carregadas com dados obtidos através do Facebook, para verificar o desempenho dos modelos em consultas recursivas. Os resultados apresentados pelo experimento foram satisfatórios, pois forneceram informações sobre as principais diferenças na utilização desses modelos.

Palavras-chave: Modelagem Orientada a Grafos; Modelagem Relacional; Banco de Dados; Teoria de Grafos e Big Data.

ABSTRACT

The great challenge for software engineers is to develop new technologies and data models that are able to store and analyze efficiently the large amount of data created nowadays. It was found that the relational model has proved inefficient because the volume of data and their relationships directly affect the performance. Therefore, use a flexible model, as oriented graphs, can be a good alternative for creating complex data models. This paper aims at modeling and implementing a graph database, in order to identify their differences to the relational approach. For this, an experiment was developed based on social analysis which databases were loaded with data obtained from Facebook, in order to verify the performance of the two data models in recursive queries. The results showed by performance tests were satisfied, giving the information about the main differences in the use of relational databases and graph databases.

Keywords: *Graphs-Oriented Modeling; Relational Modeling; Database; Graph Theory and Big Data.*

¹ Possui graduação em Ciência da Computação pela Universidade do Sul de Santa Catarina (2013). Tem experiência na área de Ciência da Computação, atuando principalmente nos seguintes temas: modelagem de dados, teoria de grafos, bancos de dados, engenharia do conhecimento e representação de conhecimento. E-mail: guilherme.alvarez@gmail.com.

² Possui graduação em Ciência da Computação pela Universidade do Sul de Santa Catarina (2007) e mestrado em Engenharia e Gestão do Conhecimento pela Universidade Federal de Santa Catarina (2010). É doutorando em Engenharia do Conhecimento pela UFSC. Atualmente é professor titular da Universidade do Sul de Santa Catarina e pesquisador do Instituto Stela. Tem experiência na área de Ciência da Computação, com ênfase em Sistemas de Informação, atuando principalmente nos seguintes temas: recuperação da informação, reconhecimento de entidades, engenharia do conhecimento, extração de conhecimento e análise de sentimento. E-mail: flavio.ceci@unisol.br.

INTRODUÇÃO

Com a popularização dos sistemas de informação, a expansão de suas bases de dados e a necessidade de se obter o melhor desempenho ao menor custo, aumentou o nível de exigência e expectativa dos consumidores de tecnologia.

A quantidade de dados armazenados em sistemas de informação tem crescido diariamente e analisar esse grande volume de dados tornou-se um requisito importante de negócio. Segundo Corrigan et al (2013), muitos analistas do campo da gestão de informação têm trabalhado para traduzir dados estruturados e não estruturados em conhecimentos úteis para as organizações, porém essas informações demandam um esquema diferente do modelo relacional. Nesse sentido, para Vicknair et al. (2010) utilizar o paradigma orientado a grafos pode ser uma solução, pois o volume de dados e seus relacionamentos não são um impeditivo para um ótimo desempenho.

Verificou-se que o modelo de dados está evoluindo, mas o esquema acaba não acompanhando essa evolução. Desse modo, o modelo relacional acaba se tornando uma barreira, pois impõe uma resistência a mudanças rápidas que os sistemas atuais necessitam.

Outro ponto a ser avaliado é o desempenho das consultas realizadas nos bancos de dados. Algumas, envolvendo os objetos armazenados, estão apresentando deficiência no tempo de resposta, pois, no modelo relacional, o desempenho das consultas é diretamente proporcional ao número de relacionamentos envolvidos e aos tamanhos da base de dados.

Sobre isso, Robinson, Webber e Eifrem (2013) afirmam que a análise de grandes volumes de dados, *Big Data*, é complexa e demorada, utilizando o modelo relacional devido ao alto nível de processamento necessário, mas se torna eficiente em um modelo de dados que utiliza grafos.

Os grafos são estruturas que podem ser utilizadas em diversas áreas de interesse e são úteis para modelagem de objetos e suas interações. Utilizar um banco de dados orientado a grafos traz algumas vantagens, pois, além de simplificar a identificação e modelagem de dados, a estrutura livre de índice facilita as consultas de alto desempenho.

Este trabalho tem como objetivo modelar e implementar uma base de dados orientada a grafos, a fim de identificar suas diferenças em relação à abordagem

relacional. Por meio de uma pesquisa aplicada, envolvendo levantamento bibliográfico e um experimento, serão elencadas e discutidas algumas questões envolvendo o desempenho e a modelagem de dados, tanto a orientada a grafos quanto a relacional.

O trabalho está organizado em cinco seções principais. Após a introdução, a segunda apresenta o referencial teórico relacionado a banco de dados orientado a grafos e análise social. A terceira seção descreve a solução proposta. Na quarta, é apresentada a avaliação de resultados da solução proposta através de um experimento. Por fim, na última seção, são apresentadas as considerações finais.

REFERENCIAL TEÓRICO

A presente seção descreve alguns elementos necessários para o entendimento deste artigo. São abordados conceitos básicos relacionados a banco de dados orientado a grafos e análise social.

BANCO ORIENTADO A GRAFOS

Para Foggia, Sansome e Vento (2001), os grafos são estruturas de dados dotados de um poder tão expressivo que tornam a sua utilização rentável nas mais díspares áreas. Ao atribuir um significado adequado para os vértices e arestas de um grafo, é possível alcançar representações completas, de interesse em muitos domínios de aplicação, que vão desde a área científica à área humanística.

Tais modelos de dados podem ser caracterizados como aqueles nos quais as estruturas de dados são modeladas na forma de grafos, e a manipulação de dados é expressa por operações assim orientadas. Esses modelos tiveram início no final dos anos oitenta e início dos noventa, em paralelo aos modelos orientados objeto. (ANGLES; GUTIERREZ, 2008)

De acordo com Batra e Tyagi (2012), em uma base de dados orientada a grafos, os dados são representados por nós, arestas e propriedades. Os nós representam as entidades ou objetos; as arestas demonstram a relação entre os nós; e as propriedades retratam as características específicas dos nós e arestas.

Bases de dados orientadas a grafo são eficientes quando se lida com áreas onde a interconectividade dos dados é importante. Com o modelo de dados

orientado a grafos, pode-se manter a estrutura original de alta fidelidade do grafo, enquanto adicionam-se elementos que atendam às novas necessidades. (ROBINSON; WEBBER; EIFREM, 2013)

Recuperar a informação a partir de um grafo requer algo que é conhecido como travessia, que envolve o conceito de caminhar ao longo dos vértices do grafo. Uma diferença importante entre uma travessia e uma consulta SQL é que travessia é uma operação localizada, na qual não há índice de adjacência global. Cada vértice e aresta no grafo armazenam um índice dos objetos conectados a ele, e, dessa maneira, o tamanho do grafo não ocasiona a perda de desempenho em uma travessia. (MILLER, 2013)

A seção seguinte apresenta conceitos relacionados à análise social, necessários para o entendimento deste trabalho.

ANÁLISE SOCIAL

Para Cook e Holder (2007), a análise social é o estudo das interações sociais humanas. Então, a análise das redes sociais pode ser utilizada para investigar padrões, estruturas de uma comunidade, organização de outras redes sociais. etc. Essas redes podem ser associadas a diversas áreas, como empresas, grupos familiares, redes acadêmicas e outras. Na sua forma mais simples, um grafo de rede social contém nós, que representam atores, e arestas, que representam as relações ou interações entre os atores.

As redes sociais *online* expandiram os horizontes da análise social na última década, pois redes sociais como *Twitter*, *LinkedIn*, e *Facebook*, tem se tornado cada vez mais populares. (AGGARWAL, 2011). Segundo Abraham, Hassanien e Snásel (2009), as redes sociais virtuais permitem que as pessoas troquem informações por meio de uma interação fácil, universal, barata e confiável. Essas interações estão ligadas à possibilidade de acessar informações específicas sobre as diferentes áreas de interesse. Portanto, essas fontes de informação são socialmente úteis, pois permitem que as pessoas estabeleçam com facilidade um contato entre si.

A representação através de grafos é natural para os dados extraídos de *sites* de redes sociais, nos quais as pessoas criam uma rede de amigos, colegas ou parceiros de negócios. Esse tipo de modelagem permite a aplicação da teoria de grafos, métodos tradicionais de análise de redes sociais e de mineração de dados

em grafos. No entanto, o enorme tamanho de um grafo utilizado para representar as redes sociais pode apresentar desafios para o processamento automatizado de dados devido à necessidade de uma poderosa estrutura de *hardware*. Outro desafio na aplicação de processos automatizados para a mineração de dados em redes sociais é a constante mudança de conteúdo e estrutura. (AGGARWAL, 2011).

MATERIAIS E MÉTODO

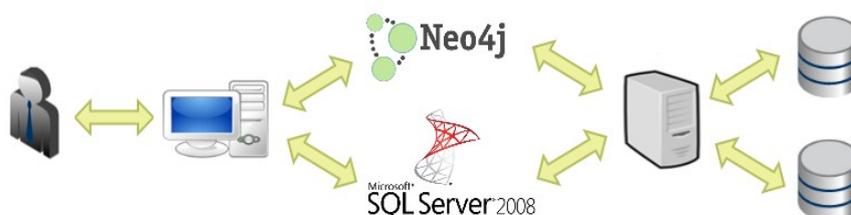
Nesta seção, objetiva-se apresentar o método adotado nesta pesquisa, bem como os elementos que fazem parte do experimento desenvolvido.

Para realizar o experimento proposto, são implementadas duas bases de dados com as mesmas informações armazenadas, uma utilizando o modelo relacional, e a outra, utilizando o modelo orientado a grafos.

A base de dados relacional utiliza o SQL Server 2008 Enterprise como Sistema Gerenciador de Banco de Dados, e o banco de dados orientado a grafo utiliza o Neo4J. As bases de dados ficarão armazenadas em um mesmo servidor para que haja as mesmas condições nas consultas e testes necessários.

A Figura 1 ilustra o modelo desenvolvido como proposta de solução para o problema.

Figura 1 – Solução Proposta



Fonte: Autoria própria

A solução proposta utiliza informações retiradas de uma rede social como fonte de informações para o desenvolvimento dos bancos de dados. Os dados selecionados para montar o grafo social foram o nome, sobrenome, sexo e código dos usuários no *Facebook*, além das relações de amizade formadas na rede social, estabelecendo-se como as arestas do grafo social.

Essas informações vinculadas a pessoas e suas relações de amizade são modeladas conforme os paradigmas de bancos de dados relacional e orientado a grafos. Além disso, estabeleceu-se um dos autores deste trabalho como nó de

referência para as buscas. Dessa forma, todas as consultas realizadas teriam esse autor como ponto inicial.

Após a conclusão do desenvolvimento das bases de dados, são aplicadas consultas recursivas entre os dados armazenados, para que os resultados obtidos sejam analisados e comparados.

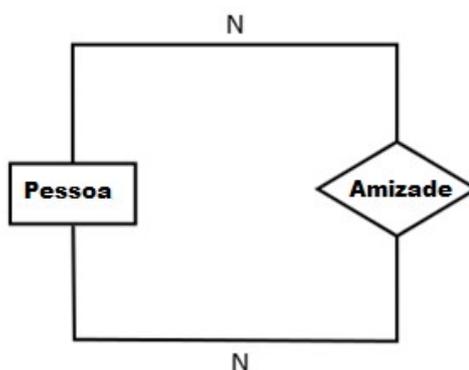
PROJETO DE EXPERIMENTO

Nesta seção, são apresentados os modelos de dados utilizados neste trabalho, o histórico do desenvolvimento do experimento e os resultados dos testes comparativos entre a abordagem relacional e a orientada a grafos.

MODELO ENTIDADE-RELACIONAMENTO

A Figura 2 ilustra a modelagem E.R, entidade-relacionamento, do banco de dados relacional utilizado neste trabalho.

Figura 2 – Modelo Entidade-Relacionamento



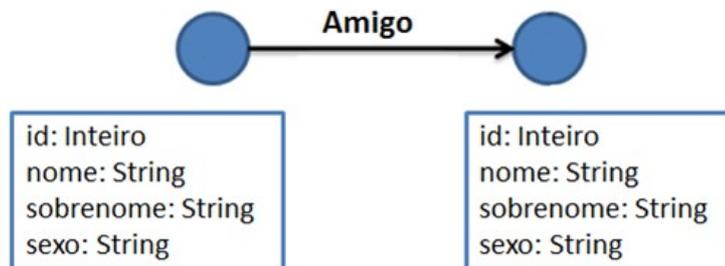
Fonte: Autoria própria

Pode-se verificar que se trata de um auto-relacionamento, no qual a entidade Pessoa se relaciona com ela mesma através do relacionamento Amizade. Dessa forma, o relacionamento Amizade irá estabelecer uma conexão entre os indivíduos da rede social.

MODELO ORIENTADO A GRAFOS

A Figura 3 ilustra a modelagem da base de dados orientada a grafos utilizada neste trabalho. Nela, pode-se identificar as entidades Pessoa, representadas pelos vértices do grafo, ligadas através do relacionamento Amigo. Os atributos das entidades Pessoa são relacionados como suas propriedades.

Figura 3 – Modelo Orientado a Grafos



Fonte: Autoria própria

O modelo de dados orientado a grafos é uma ferramenta eficaz para a modelagem de dados quando o foco no relacionamento entre as entidades é o ponto principal da concepção do modelo de dados.

HISTÓRICO DE DESENVOLVIMENTO

Escolheu-se o *Facebook* como a rede social desse estudo, pelo fato da sua API ser de fácil acesso e por estar extremamente difundido no Brasil e no mundo. As relações de amizade, parentesco, trabalho, estudo, interesses e outras, poderiam ser analisadas enriquecendo a rede construída nesse projeto e dando abertura a trabalhos futuros.

Utilizando a Graph API, biblioteca fornecida pelo *Facebook*, os dados das consultas realizadas na base de dados da rede social são retornados em forma de arquivo JSON, dessa forma podem ser explorados e integrados a diversos sistemas.

Durante os estudos da Graph API, verificou-se que a rede social impõe algumas regras e necessidades de algumas autorizações como medida de segurança e de garantia da privacidade de seus usuários. Portanto, para que os dados da lista de amigos fossem disponibilizados pela base de dados da rede social, seria necessária a autorização do indivíduo que quisesse colaborar com a pesquisa.

Realizar a configuração de segurança para liberação de dados de forma manual na ferramenta Graph API Explorer, fornecida pelo *Facebook*, seria um impeditivo para os usuários participarem desta pesquisa fornecendo seus dados, pois muitos não possuem acesso nem conhecimento às ferramentas de desenvolvimento disponibilizadas pela rede social.

Assim, com intuito de facilitar o processo de autorização de acesso aos dados dos usuários do *Facebook*, foi desenvolvido um web site que conecta a base de dados da rede social e envia uma solicitação de autorização ao usuário automaticamente. Após obter a permissão, o mesmo envia uma solicitação com a consulta de dados ao *Facebook*, que retorna o resultado em forma de dados JSON. Os arquivos gerados com as informações dos usuários são armazenados no servidor de hospedagem do *web site*.

Como os bancos de dados utilizados neste trabalho não possibilitam leitura, identificação nem persistência direta dos dados através de um arquivo JSON, houve a necessidade de implementar uma aplicação para realizar essa tarefa. Foi desenvolvido um sistema em JAVA que faz a leitura dos arquivos JSON disponibilizados pelo Facebook e adiciona os dados dos usuários e seus relacionamentos nas bases de dados Neo4J e SQL Server 2008. O sistema teve por objetivo automatizar a leitura e inserir dados nas bases a serem utilizadas nos testes.

Durante o desenvolvimento do sistema, verificou-se que seria mais eficiente utilizar uma biblioteca para converter arquivos JSON em objetos JAVA. Após levantamento e análise das bibliotecas disponíveis na web, a escolhida foi a GSON, desenvolvida pela Google. A utilização dessa biblioteca simplificou e facilitou o desenvolvimento dessa aplicação, além de trazer segurança a persistência de dados.

AVALIAÇÃO DOS RESULTADOS

Dentro do contexto de análise das redes e das relações sociais, existem dois tipos primários de dados que podem ser analisados:

- Análise de conteúdo da rede, baseada nas informações contidas na rede de dados.
- Análise estrutural da rede, baseada na estrutura de ligação da rede de dados.

No experimento apresentado neste trabalho, somente a primeira forma de análise será abordada, demonstrando as vantagens da utilização de bases de dados orientadas a grafo.

Na base relacional utilizada neste experimento, a tabela Pessoa contém os dados dos usuários do *Facebook*, enquanto a tabela Amizade armazena a relação

de amizade entre os indivíduos, referenciando a tabela Pessoa através de chave estrangeira.

O *Facebook* apresenta sugestões aos usuários de possíveis amigos até um determinado grau de profundidade, baseando-se nas relações entre os indivíduos da rede. Porém, a busca desse tipo de dados em bases de dados relacionais não se trata de algo simples, pois, para encontrar os amigos de amigos de um usuário, é necessário o uso das operações de junção de dados.

Nos casos em que se deseja localizar amigos em comum ou amigos de amigos, o objetivo da pesquisa é consultar apenas os amigos de amigos de um determinado usuário. Mas, ao utilizar as junções, a base de dados avalia todos os dados contidos na tabela Amizade, descartando posteriormente os dados desnecessários, o que gera uma perda de desempenho.

No Quadro 1, pode-se visualizar os resultados obtidos nos testes de desempenho realizados na base de dados relacional, nos quais foram utilizados diferentes profundidades e quantidades de dados.

Profundidade	10095 Usuários	20349 Usuários	30522 Usuários	40053 Usuários	57673 Usuários
2	10ms	11ms	13ms	15ms	24ms
3	24ms	25ms	41ms	59ms	48ms
4	41ms	71ms	82ms	119ms	193ms

Quadro 1 – Testes Base de Dados Relacional
Fonte: Autoria própria

Pode-se verificar que a base de dados relacional trata as consultas com profundidades 2 e 3 de maneira aceitável, mas, ao utilizar profundidades acima de 4, há uma queda significativa no desempenho. Dessa forma, é possível identificar que, em consultas que envolvem várias junções recursivas, as bases de dados relacionais normalmente apresentam essa queda.

Já ao realizar uma pesquisa para encontrar todos os amigos de amigos de um indivíduo em profundidade 4, o banco de dados relacional gera o produto cartesiano da tabela Amizade quatro vezes. Aplicando esse contexto à tabela Pessoa utilizada no experimento, com 57.673 registros, o conjunto resultante terá 57.673^4 , ou seja, 11.063.439.658.308.155.000 linhas. Para obter esse tipo de resultado, o banco de dados necessita de muito tempo de processamento, que acaba sendo desperdiçado, pois a base de dados descarta mais de 95% dos registros para retornar apenas os dados relacionados aos amigos de amigos de um indivíduo até a profundidade 4.

No modelo de dados orientado a grafos, os usuários do *Facebook* são os nós do grafo e as suas informações são as propriedades desse nó, as relações de amizade estabelecidas na rede social são os relacionamentos entre os nós, e os índices são as chaves únicas utilizadas para identificar esses dados.

Ao realizar a travessia, a base de dados acessa um conjunto destes movendo-se entre os nós através das arestas formadas pelos seus relacionamentos. Dessa forma, a travessia só considera os dados necessários para a consulta, sem a necessidade de envolver o conjunto inteiro de dados.

No Quadro 2, pode-se visualizar os resultados obtidos nos testes de desempenho realizados na base de dados orientada a grafos, nos quais foram utilizados diferentes profundidades e quantidades de dados.

Profundidade	10095 Usuários	20349 Usuários	30522 Usuários	40053 Usuários	57673 Usuários
2	7ms	8ms	8ms	8ms	11ms
3	8ms	9ms	16ms	25ms	38ms
4	9ms	15ms	26ms	39ms	62ms

Quadro 2 – Testes Base de Dados Orientada a Grafos
Fonte: Autoria própria

Observando o quadro apresentado, verifica-se que o desempenho do banco de dados orientado a grafos é semelhante ao relacional na consulta de profundidade 2. No entanto, quando a profundidade e a quantidade de dados envolvidos na consulta aumentam, a base de dados orientada a grafos apresenta um desempenho significativamente melhor que o modelo relacional.

A fim de simular um ambiente com um volume maior de dados, foram criadas duas bases com registros fictícios aplicando a mesma estrutura de rede social já demonstrada neste trabalho. As estruturas geradas possuem 1 milhão de pessoas e cada pessoa possui em torno de 100 relacionamentos de amizade.

No Quadro 3, pode-se visualizar os resultados obtidos nos testes de desempenho realizados em ambos os modelos de bases de dados estudados.

Base de dados relacional		Base de dados orientada a grafos	
Profundidade	Tempo de Execução	Profundidade	Tempo de Execução
2	5586ms	2	16ms
3	8550ms	3	328ms
4	46818ms	4	2497ms
5	638635ms	5	11218ms

Quadro 3 – Teste nas Bases de Dados com um milhão de dados
Fonte: Autoria própria

Aplicar os testes de desempenho utilizando uma maior massa de dados na base de dados relacional se mostrou uma tarefa demorada, devido à estrutura do modelo relacional. Com 1 milhão de indivíduos cadastrados, o produto cartesiano gerado na consulta de profundidade 5 possui 10^{30} registros. Portanto filtrar todos os registros para verificar quais se encaixam nos requisitos da pesquisa é um desperdício de tempo e de desempenho.

Pelos resultados apresentados no Quadro 3, é possível identificar que o desempenho na consulta com 1 milhão de indivíduos na profundidade 2, para a base de dados orientada a grafos, é semelhante ao resultado obtido nos testes iniciais com 57673 indivíduos na mesma profundidade.

A análise de grandes volumes de dados é complexa utilizando o modelo relacional devido ao alto nível de processamento necessário, mas torna-se mais simples em um esquema de dados que utiliza grafos, pois as travessias utilizadas nas consultas tornam a pesquisa de dados mais eficiente quando se trabalha com redes complexas destes.

Independente do número de nós e relacionamentos existentes no grafo, a travessia só irá acessar os nós que são relevantes para a consulta. Quanto maior a profundidade estabelecida nesta, maior será a quantidade de nós que a travessia precisa visitar, tornando a consulta mais lenta. No entanto, este aumento de tempo é linear e independente do tamanho total do grafo.

As consultas foram realizadas com cache quente, ou seja, cada consulta foi repetida 10 vezes e o menor resultado obtido foi apresentado.

CONSIDERAÇÕES FINAIS

Este trabalho apresentou conceitos sobre bancos de dados orientados a grafo e análise social, tendo como objetivo principal a modelagem e a implementação de uma base de dados orientada a grafos, a fim de identificar suas diferenças em relação à abordagem relacional. Outro objetivo deste trabalho foi realizar um experimento sobre análise social a fim de testar o desempenho dos modelos relacional e orientado a grafos ao realizarem consultas que envolvem cálculos relacionais e relacionamentos recursivos.

Para isso, a rede de dados foi construída utilizando de informações fornecidas por usuários da rede social *Facebook* através de uma aplicação elaborada para facilitar a conexão com essa rede e capturar os dados dos usuários dispostos a colaborar com este projeto.

Os testes iniciais de desempenho foram realizados baseando-se nas relações entre os indivíduos da rede, consultando os dados de amigos de amigos até a profundidade 4, tendo um dos autores deste trabalho como vértice inicial da pesquisa.

Ao analisar os resultados obtidos nos testes, é possível identificar a diferença de desempenho entre o modelo relacional e o orientado a grafos em consultas recursivas, visto que as consultas realizadas na base de dados orientada a grafos apresentaram um desempenho superior ao apresentado pelo modelo relacional.

Ao simular o contexto de *Big Data* para verificar a performance das bases de dados frente a um volume maior destes, a diferença de desempenho se mostrou ainda mais significativa. Portanto, pode-se concluir que utilizar bases de dados orientadas a grafo para realizar pesquisas em redes complexas de dados é mais vantajoso quando se buscam soluções de alto desempenho e escalabilidade, independente do cenário utilizado na pesquisa.

A estrutura dos grafos facilita a modelagem dos dados, pois não há necessidade de reestruturar o esquema de dados cada vez que um novo tipo de entidade ou relacionamento é adicionado.

O modelo de dados orientado a grafos se mostrou muito útil para modelagem, armazenamento e análise de redes complexas de dados como as utilizadas na análise social, pois a forma como os dados são modelados nas bases orientadas a grafos é muito próxima à maneira que os mesmos estão estruturados e conectados na vida real, tornando a modelagem em grafos versátil, simples e eficiente.

Após analisar a problemática apresentada neste trabalho, a proposta de solução e os resultados obtidos através dos testes de desempenho, conclui-se que a proposta de utilização de uma base orientada a grafos para análise social é válida, podendo também ser utilizada em outros cenários e não apenas no apresentado no experimento.

REFERÊNCIAS

ABRAHAM, Ajith; HASSANIEN, Aboul-Ella; SNÁSEL, Vaclav. **Computational Social Network Analysis: Trends, Tools and Research Advances**. Londres: Springer, 2009.

AGGARWAL, Charu C. **Social Network Data Analytics**. Nova York: Springer, 2011.

ANGLES, Renzo; GUTIERREZ, Claudio. **Survey of Graph Database Models**. ACM Computing Surveys, 2008.

BATRA, Shalini; TYAGI, Charu. **Comparative analysis of relational and graph databases**. International Journal of Soft Computing, vol 2, 2012.

COOK, Diane. J.; HOLDER, Lawrence. B. **Mining Graph Data**. Nova Jersey: John Wiley & Sons, Inc., 2007

CORRIGAN, David; DEROOS, Dirk; DEUTSCH, Tom; ZIKOPOULOS, Paul; PARASURAMAN, Krishnan; GILES, James. **Harness the Power of Big Data**. McGraw-Hill Companies. 2013.

FOGGIA, P; SANSONE, C; VENTO, M. **A Database of Graphs for Isomorphism and Sub-Graph Isomorphism Benchmarking**, In: 3RD IAPR TC-15 INTERNATIONAL WORKSHOP ON GRAPH-BASED REPRESENTATIONS, 2001, Ischia. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.6803&rep=rep1&type=pdf>>. Acesso em: 01 maio 2013

MILLER, Justin J; **Graph Database Applications and Concepts with Neo4j**. In: Southern Association for Information Systems Conference. 2013, Atlanta. p.141 – 147. Disponível em: < <http://sais.aisnet.org/2013/MillerJ.pdf>>. Acesso em: 05 maio 2013

ROBINSON, Ian; WEBBER, Jim; EIFREM, Emil. **Graph Databases**. O'Reilly Media, Inc. 2013.

VICKNAIR, chad; MACIAS, Michael; ZHAO, Zhendong; NAN, Xiaofei; CHEN, Yixin; WILKINS, Dawn. **A comparison of a graph database and a relational database: A data provenance perspective**. In: 48th Annual Southeast Regional Conference, 2010, Nova York. p. 68–80. Disponível em: <http://www.cs.olemiss.edu/~ychen/publications/conference/vicknair_acmse10.pdf>. Acesso em: 25 de maio 2013.