Linguagens de consulta para bases de dados em grafos: um mapeamento sistemático

Query languages for graph databases: a systematic mapping

Simone de Oliveira Santos [1]
Martin A. Musicante [2]
Mirian Halfeld Ferrari Alves [3]

Data de submissão: 09/03/2015, Data de aceite: 04/05/2016

Resumo: A popularização das redes sociais, associada à necessidade de analisar e sumarizar grandes volumes de dados oriundos das mesmas tem favorecido o uso de bases de dados em grafos. As linguagens de consulta a este tipo de base de dados devem, ao mesmo tempo, ter expressividade suficiente para a realização de consultas complexas e possibilitar o processamento eficiente de grandes volumes de dados. Este artigo apresenta um mapeamento sistemático sobre as linguagens de consulta para bases de dados em grafos, com foco nas suas características principais como paradigma ou capacidade de agregação de dados. O foco deste mapeamento é investigar e quantificar as publicações referentes às linguagens de consulta, caracterizando-as, identificando possíveis áreas de pesquisa, tendências e desafios.

***P*alavras-chave:** bases de dados em grafos, linguagem de consulta, mapeamento sistemático

Abstract: The widespread use of digital social networks, together with the need for the analysis and summarisation of their contents is at the base of the widespread use of graph databases. Query languages for this kind of databases need to be expressive enough to express complex queries to be efficiently performed over enormous amounts of data. In this paper, we perform a systematic mapping about query languages for graph databases, covering aspects such as their paradigm or aggregation capabilities. We focus on the quantitative analysis of publications on the subject, to better identify trends and future challenges.

***K*eywords:** graph database, query language, systematic mapping

[1]PPgSC Programa de Pós-graduação em Sistemas e Computação, UFRN Universidade Federal do Rio Grande do Norte - Natal, Rio Grande do Norte, Brasil. `{simone82@ppgsc.ufrn.br,mam@dimap.ufrn.br}`
[2]PPgSC - Programa de Pós-graduação em Sistemas e Computação, DIMAp Departamento de Informática e Matemática aplicada, UFRN Universidade Federal do Rio Grande do Norte - Natal, Rio Grande do Norte, Brasil
[3]Université d'Orléans, INSA CVL, LIFO EA - Orléans, France. `{mirian@univ-orleans.fr}`

1 Introdução

Grafos são populares para representar dados. Eles podem ser usados para modelar relacionamentos entre objetos e recursos. Grafos podem ser definidos como uma coleção de vértices e arestas, onde os vértices representam entidades, e as arestas representam a forma como as entidades se relacionam [78]. As bases de dados em grafos fazem parte de um conjunto de tecnologias criadas ou redescobertas para suprir algumas limitações do modelo de dados relacional [1].

O modelo de dados em grafo pode ser definido como aquele em que tanto o esquema quanto as instâncias são modelados como grafos [3]. Nos últimos anos, este modelo de dados tem recebido muita atenção da indústria e da academia. Com o avanço da Internet, o compartilhamento de dados se tornou um tópico importante e uma grande quantidade de grafos está disponível em bases de dados para uso experimental [12]. Os grandes grafos ganharam destaque e cada vez mais aplicações são feitas para lidar com esse tipo de volume.

É possível notar o crescente interesse pelas bases de dados em grafos pela literatura, inclusive conferências específicas dedicadas ao assunto[4]. Estudos relacionados às bases de dados em grafos podem ser direcionados para a forma de armazenamento dos dados (estrutura dos dados), para a forma de manipulação desses dados através de uma linguagem de consulta e também nas restrições de integridade apropriadas para o contexto.

Este artigo apresenta um mapeamento sistemático [74] sobre as linguagens de consulta para bases de dados em grafos. Para atingir esse objetivo, também será investigado o retorno ao uso do modelo de dados em grafos, apresentando os números referentes a este tópico e as razões para este comportamento. Esta investigação envolve identificar o tipo de contribuição que os artigos se propõem, como também que tipo de problema eles tentam resolver, e que técnicas eles usam para atingir os seus objetivos.

Um mapeamento sistemático é um método para analisar um tópico de interesse seguindo passos bem definidos. Ao final da análise, deve-se responder um conjunto de questões de pesquisa que são lançadas no início do processo. Toda a investigação dos números do mapeamento devem servir de base para responder as questões lançadas. O objetivo deste trabalho é investigar e quantificar as publicações referentes às linguagens de consulta, caracterizando-as, identificando possíveis áreas de pesquisa, tendências e desafios.

Este artigo é organizado como segue. A Seção 2 apresenta conceitos referentes ao método usado para a confecção deste mapeamento sistemático, e considerações referentes ao modelo de dados em grafos e trabalhos relacionados. A Seção 3 apresenta todo o processo de mapeamento, etapa a etapa. A Seção 4 apresenta as análises quantitativas extraídas do

[4]Por exemplo, as conferências GraphConnect (http://graphconnect.com/) e o International Workshop on Querying Graph Structured Data (http://www.isgroup.unimo.it/graphq2016/).

mapeamento e a Seção 5 contém as conclusões finais.

2 Revisão Teórica

Esta seção é dedicada a apresentar o processo de mapeamento sistemático que será aplicado neste trabalho e apresentar conceitos referentes ao modelo de dados em grafos e linguagens de consulta.

2.1 Mapeamento sistemático

Um mapeamento sistemático é um tipo de estudo que fornece uma visão geral sobre uma área específica de pesquisa [74]. De forma geral, o processo é feito coletando artigos referentes a um assunto, publicados em uma época determinada. A partir dos dados coletados é realizado um mapeamento entre os diferentes aspectos dos artigos coletados, de forma a poder fazer uma análise quantitativa do conjunto, verificando relações e tendências. O mapeamento é feito geralmente a partir da leitura dos títulos e resumos dos artigos. A análise quantitativa é ilustrada com gráficos e tabelas. O método utilizado neste trabalho foi feito de acordo com [74], que contém cinco grandes etapas, como mostrado na Figura 1, e cada etapa será descrita a seguir.

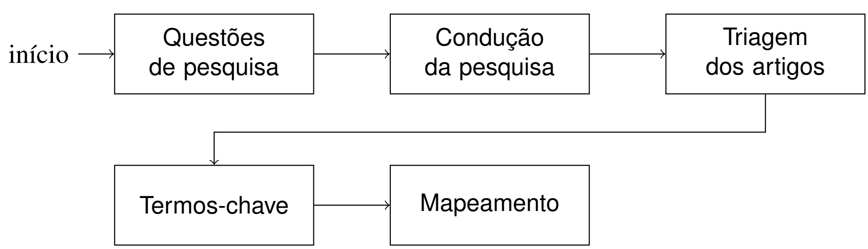

Figura 1. Etapas do mapeamento sistemático aplicadas neste trabalho.

2.1.1 Questões de pesquisa A primeira parte do mapeamento consiste em definir o objetivo principal do mesmo. As questões de pesquisa definem o assunto e o objetivo do mapeamento. As questões indagam sobre o assunto escolhido e indicam quais são os interesses específicos. As questões de pesquisa são importantes também para definir quais artigos devem fazer parte do mapeamento.

2.1.2 Condução da pesquisa Este passo refere-se ao processo de coletar artigos nas bases de dados que reúnem as publicações de conferências e periódicos científicos. A condução da pesquisa consiste em três etapas: definição dos termos de consulta; escolha das bases de dados; e a execução da consulta nas bases de dados.

A definição dos termos de consulta refere-se à criação de palavras ou termos (*string* de busca) usados para consultar os artigos nas bases de dados. A escolha das bases de dados deve levar em conta as bibliotecas mais populares e abrangentes. Atualmente as consultas geralmente são feitas em bibliotecas digitais *on-line* que reúnem publicações das principais conferências e periódicos científicos. A execução da consulta nas bases de dados consiste em utilizar a *string* de busca nas bibliotecas digitais e recuperar todos os artigos que casaram com a consulta feita.

2.1.3 Triagem dos artigos A fase de triagem é feita após a recuperação dos artigos das bases de dados de bibliotecas digitais previamente selecionadas. Os artigos recuperados são organizados em listas e a triagem serve para escolher quais artigos devem ser selecionados para fazer parte do mapeamento ou não. A triagem é feita com a leitura dos títulos e resumos dos artigos somente, com base em critérios de inclusão e exclusão de artigos.

Os critérios de inclusão e exclusão são definidos baseando-se nos objetivos do mapeamento, principalmente usando as questões de pesquisa. Os critérios definem quais artigos são relevantes ou não para o mapeamento. É importante incluir nos critérios de exclusão a verificação de artigos repetidos, haja vista que um mesmo artigo pode estar incluído em mais de uma base de dados.

2.1.4 Termos-chave A busca por termos-chave é usada para criar uma classificação dos artigos. Os termos-chave são identificados durante a leitura dos títulos e resumos dos artigos. Estes termos são aqueles que identificam a contribuição do trabalho, pontos importantes como o objetivo, os métodos e tecnologias usados. A classificação neste contexto, é o ato de combinar os termos identificados e definir um conjunto de categorias. A divisão em categorias é importante e ajuda no momento de fazer uma análise e tirar conclusões. Os artigos selecionados para o mapeamento devem ser distribuídos dentro das categorias definidas.

2.1.5 Mapeamento Nesta etapa os artigos devem estar distribuídos entre as categorias definidas. A classificação não precisa ser estática, ela pode evoluir por exemplo com a criação ou mesclagem de categorias. Esta é a fase onde o mapeamento toma forma e as análises são feitas. A análise dos resultados deve ter como foco o número de publicações em cada categoria. Isso permite visualizar a progressão das publicações, como elas se distribuem aos longo dos anos, ou se algum assunto está em ênfase ou caindo em desuso. Neste ponto, também é oportuno identificar se há uma relação entre as categorias, e ilustrá-las se for o caso. Ao final da análise é importante que as questões de pesquisa sejam respondidas com base nos resultados obtidos com o mapeamento.

2.2 Modelo de dados em grafos e linguagens de consulta

Modelos de dados são usados para representar a forma ou esquema dos dados de uma organização. Um modelo de dados consiste em (i) um conjunto de tipos estruturados de dados, (ii) um conjunto de operadores ou regras de inferências, e (iii) um conjunto de regras de integridade [1]. Este trabalho tem maior foco no conjunto de operadores, ou seja, nas linguagens de consulta que são usadas para manipular os dados. O uso das bases de dados em grafos tem captado o interesse da indústria e academia na última década [101]. É possível observar um aumento no número de artigos publicados neste contexto a partir da década de 2000. Este comportamento pode ser observado na Seção 4.

O surgimento de novas aplicações relacionadas com a Internet, como as redes sociais digitais e a web semântica, demandaram novas tecnologias. Dentre estas novas tecnologias, o *Resource Description Framework* (RDF) [22] surgiu como um formato para representar metadados na *Web*, mas atualmente ele é bastante usado como uma extensão do modelo de dados em grafos [2]. Um documento do tipo RDF é uma coleção de triplas no padrão *sujeito-predicado-objeto* representando, dessa forma, os arcos de um grafo (o predicado), ligando um par de nodos (sujeito e objeto). Este tipo de grafo, que é considerado por alguns como um modelo de dados, é uma das tecnologias usadas na web semântica que provê um ambiente onde aplicações podem manipular (consultar/alterar) dados e fazer inferências [86]. A linguagem de consulta SPARQL é a linguagem recomendada pela W3C[8] para grafos RDF, mas de forma geral, não há uma linguagem de consulta padrão para bases de dados em grafos [12].

Outro tipo de tecnologia voltada para o modelo de dados em grafos são os sistemas nativos. Os sistemas nativos consideram a estrutura de grafos tanto no armazenamento físico dos dados quanto no processamento de consultas [72]. Estes sistemas são feitos especificamente para gerenciar grafos e usam listas de adjacência para processá-los. Um exemplo para esse tipo de sistema é o Neo4j [5], que utiliza a linguagem de consulta Cypher, que também dá suporte ao SPARQL.

A agregação é uma função bastante usada no modelo de dados relacional. Uma função de agregação usa o resultado de uma consulta como valor intermediário para fazer uma outra operação. As funções de agregação mais comuns são soma, média, maior valor e menor valor. O uso desse tipo de função nas linguagens de consulta é observado neste mapeamento.

2.3 Trabalhos relacionados

Existem vários trabalhos que fazem comparativos, ilustram o estado-da-arte, ou fazem revisões da literatura dedicados às bases de dados em grafos. Esses trabalhos normal-

[5]http://neo4j.com/

mente exploram os diversos modelos de dados e diferentes tipos de aplicações. Alguns deles dedicam-se exclusivamente às linguagens de consulta para grafos. Peter Wood mostra em [101] uma visão geral das linguagens de consulta para grafos desenvolvidos nos últimos 25 anos. O autor explora diferentes sintaxes e aspectos das funcionalidades.

O *survey* [5] tem como objetivo apresentar linguagens de consulta propostos para os principais formalismos da web semântica, tais como RDF e outros. O artigo discorre sobre várias linguagens e tenta destacar importantes aspectos sobre elas. Os autores Angles e Gutierrez escreveram o *survey* [3] que é uma referência para o modelo de dados em grafos. Eles dedicam uma seção para descrever as linguagens de consulta para grafos que servem como referência na área.

Este presente mapeamento pode ser visto como um complemento aos *surveys* já existentes e com o diferencial de ser apresentado em um formato (mapeamento sistemático) que não foi encontrado abordando o assunto em questão. Um mapeamento tem também como objetivo quantificar as publicações encontradas em um determinado período, dando a oportunidade de fazer conjecturas baseando-se nestes números. A maior diferença entre os *surveys* e o mapeamento sistemático está no fato de que o mapeamento dá uma ênfase maior nos números de publicações, divisão em grupos, análise quantitativa.

3 Processo de mapeamento sistemático

A execução deste mapeamento foi feita seguindo o método proposto em [74], o qual foi apresentado na Seção 2.1. Nesta seção, será descrito em detalhes como foi desempenhada cada etapa do processo de mapeamento. A última etapa, referente à extração de dados e mapeamento será apresentada na Seção 4.

3.1 Questões de pesquisa

O objetivo deste mapeamento sistemático é investigar quais linguagens de consulta para bases de dados em grafos são mais usadas e quais suas principais particularidades. Características das linguagens como a agregação estão entre as contribuições buscadas e ainda, quais paradigmas são mais usados para a construção dessas linguagens.

Baseado neste objetivo foram definidas três questões de pesquisa para a condução deste mapeamento:

- *QP1: Quais são as linguagens de consultas mais usadas no modelo de dados em grafos ao longo dos anos?* Esta pergunta ajudará a identificar as linguagens mais usadas dentro do modelo de dados em grafos, se há novas linguagens sendo propostas e também será possível analisar, de acordo com o número de publicações, se certas linguagem estão definindo tendências.

- *QP2: Quantas dessas linguagens possuem suporte a agregação?* As funções de agregação são muito úteis quando é necessário fazer algum tipo de cálculo ou medida que envolve o agrupamento do resultado de uma consulta. O modelo de dados relacional tem suporte para a agregação, então é importante verificar em números o que está sendo feito para o modelo de dados em grafos para suprir essa necessidade.

- *QP3: Quais paradigmas são mais usados na definição das linguagens de consulta para o modelo de dados em grafos?* Esta questão tem o objetivo de identificar as tendências na definição das linguagens, e também se isso vem mudando ao longo dos anos.

3.2 Condução da pesquisa

A primeira parte consistiu em definir a *string* de busca a partir das questões de pesquisa. As principais palavras-chave definidas foram: "grandes volumes de dados", "linguagem de consulta", "modelo de dados em grafo" e "funções de agregação". A busca deve recuperar artigos publicados nas principais bases de dados. Dessa forma, os termos devem ser buscados em inglês, haja vista que as principais publicações são feitas nesta língua. Assim sendo, as palavras-chave da busca usadas foram: *big data*, *query language*, *graph database* e *aggregation*. A partir destas palavras foram criadas variações destas para que a abrangência da consulta fosse maior.

As variações devem ser criadas porque para chegar à *string* de consulta final antes são feitos alguns testes. Normalmente as primeiras consultas não trazem resultados satisfatórios, ou a consulta é abrangente demais e retorna um número muito alto de resultados, ou a consulta é restrita demais e retorna um número muito baixo de resultados. As variações incluem adicionar o plural de alguns termos, e/ou adicionar sinônimos.

Baseado nas palavras-chave foi criada a seguinte consulta: *((“big data” OR “large data” OR “large dataset” OR “social network” OR “social networks”) AND (“query language”) AND (“graph database” OR “graph databases” OR “graph data” OR “large graph”) AND (“aggregation” OR “summarisation” OR “aggregate”)).*

A segunda parte consistiu na escolha das bases de dados a serem consultadas. Foram escolhidas três bases de dados: *Science Direct*, *IEEE Xplore*, e *ACM digital library*. Estas bases foram escolhidas porque (i) elas são disponibilizadas na Internet, (ii) são populares para a área da Ciência da Computação, e (iii) agregam um grande número de conferências e periódicos científicos. Estas bases são disponibilizadas pela plataforma Portal de Periódicos

Tabela 1. Resultado da execução da consulta nas bibliotecas digitais.

Base de dados	Endereço eletrônico	Artigos recuperados
Science Direct	*www.sciencedirect.com*	67
IEEE Xplore	*ieeexplore.ieee.org*	89
ACM digital library	*dl.acm.org*	105
Total		**261**

CAPES[6], facilitando assim o acesso aos artigos. A plataforma Google Acadêmico[7] também foi considerada para ser incluída nas bases, mas ela foi descartada pelo fato de que grande parte dos resultados retornados já constarem nas outras bibliotecas digitais selecionadas, e ela também retorna muitos artigos de baixa representatividade.

A terceira parte desta etapa do processo consistiu na execução da *string* de busca em cada uma das bases de dados escolhidas. A execução da consulta resultou em um total de 261 artigos. A quantidade de artigos recuperados por base de dados pode ser visualizado na Tabela 1. Na tabela, a primeira coluna indica o nome da base de dados, seguido do endereço eletrônico desta base na Internet, e a quantidade de artigos recuperados.

3.3 Triagem dos artigos

Esta fase do mapeamento consiste em identificar quais artigos farão parte do mapeamento sistemático ou não, levando em conta os critérios estabelecidos. Os critérios de inclusão e exclusão de artigos foram definidos baseados no objetivo do mapeamento, levando em conta as questões de pesquisa. Os critérios definidos para este mapeamento podem ser lidos na Tabela 2.

Dentre os critérios de inclusão, os trabalhos devem mencionar alguma relação com linguagem de consulta e ter sido publicado em algum periódico, conferência ou *workshop*. Já os critérios de exclusão, cada um deles é um item eliminatório, ou seja, se um artigo se enquadra em algum dos critérios ele é eliminado.

Os trabalhos que não mencionam o uso de linguagens de consulta foram excluídos porque fugiam do escopo do mapeamento. Os capítulos de livros não foram considerados porque apresentam conteúdo que não é interessante para os números do mapeamento e sim para o uso no referencial teórico. A opção pela língua inglesa ocorreu para limitar os artigos que fossem possíveis de ser lidos pela autora, como também para eliminar possíveis artigos de mais baixa expressividade, escritos em língua nativa. Os anais completos abrangem uma

[6]http://www.periodicos.capes.gov.br/
[7]https://scholar.google.com.br/

Tabela 2. Critérios de inclusão e exclusão de artigos.

Inclusão
Trabalhos que mencionam linguagens de consulta em grafos nos resumos.
Artigos publicados em periódicos científicos, conferências ou *workshops*.
Exclusão
Trabalhos que não mencionam o uso de linguagens de consulta em grafos nos resumos.
Capítulos de livros.
Artigos escritos em língua diferente da língua inglesa.
Anais de congressos completos e índices.
Artigos repetidos.
Artigos sem resumo disponível.

grande quantidade de artigos com temas diversos que fogem ao escopo do mapeamento e os índices somente trazem os títulos, sem resumos. Os artigos sem resumo não entram no mapeamento, porque este mapeamento é feito baseado nos títulos e resumos. Um dos critérios de exclusão é a verificação de artigos repetidos, e esta verificação foi a primeira etapa da triagem.

Com os artigos organizados em listas, uma lista por base de dados, foi feita uma triagem em busca de trabalhos repetidos. Esta triagem foi feita com a leitura dos títulos dos artigos e a comparação entre as listas. Artigos com títulos iguais eram separados para uma verificação mais precisa do conteúdo. Estes eram verificados se apresentavam o(s) mesmo(s) autor(es) e o mesmo resumo. Ao constatar que o artigo era o mesmo, este era marcado como excluído, mantendo apenas um dos exemplares. Alguns artigos foram identificados como repetidos apenas após a leitura do resumo. Isso acontece porque há artigos em que os títulos são diferentes mas trata do mesmo trabalho. Muitas vezes o segundo artigo é uma versão estendida para um periódico ou uma continuação de um trabalho inicialmente já publicado. Em casos onde a diferença não é substancial o artigo é marcado como repetido. Ao todo foram identificados e excluídos 11 artigos repetidos.

Em seguida, a triagem continuou com base na leitura do título e dos resumos dos artigos e levando em conta os demais critérios de inclusão e exclusão. Após a finalização da triagem foram excluídos um total de 149 artigos. A totalização dos artigos pode ser vista na Tabela 3. A lista final contém um total de 101 artigos.

Tabela 3. Totalização após inclusão e exclusão de artigos.

	Science Direct	**IEEE**	**ACM**	**Total**
Artigos recuperados	67	89	105	261
Artigos repetidos	2	1	8	11
Artigos excluídos	44	53	52	149
Total	21	35	45	**101**

3.4 Termos-chave

A classificação dos artigos foi feita baseando-se em termos-chave que foram definidos, levando-se em conta as questões de pesquisa. A classificação levou à criação de três categorias. Cada categoria foi dividida em subcategorias para facilitar o processo de análise. As categorias e suas respectivas subcategorias são descritas a seguir.

Tipo de contribuição Esta categoria refere-se à principal contribuição do trabalho que está sendo analisado, o que está sendo proposto como objetivo principal. As divisões propostas para esta categoria foram (a) *linguagem* que indica se o trabalho apresenta uma nova linguagem ou a extensão de uma já existente; (b) *método* que indica se o trabalho apresenta um método para a resolução de um problema; (c) *survey/estado-da-arte/comparativo*; e ainda duas subcategorias complementares que indicam se o trabalho realizou (d) *experimentos* e disponibilizou (e) *ferramenta*.

Tipo de problema Refere-se ao tipo de problema que os autores pretendem abordar. Esta categoria apresenta as seguintes divisões: (a) *processamento de consultas* que indica se o foco da contribuição está em melhorias para o processamento de consultas; (b) *manipulação de grandes grafos* que indica se o trabalho está apto a lidar com grafos muito grandes (*big data*). (c) *funções de agregação* que indica se a linguagem implementa funções de agregação; (d) *consultas por palavras-chave* que indica o uso de palavras-chave para fazer consultas; (e) *top-k* que são aquelas linguagens que fazem consultas com um número limitado (preestabelecido) de resultados; (f) *integração de dados* quando o objetivo é unir dados de fontes diferentes; (g) *otimização de desempenho* que indica a implementação de alguma técnica de melhora no desempenho do algoritmo.

Paradigma tecnológico Refere-se ao modelo de programação, paradigma de programação ou ao tipo de tecnologia empregada a qual o problema envolve. Esta categoria apresenta as seguintes divisões: (a) *RDF* que é um modelo de dados muito usado para disponibilizar informações na *Web*; (b) *paralelismo* que indica o modelo paralelo de programação; (c) *distribuição de dados* que indica manipulação de dados distribuídos; (d) *linguagem declarativa* que indica o uso do paradigma declarativo.

Os artigos incluídos no mapeamento foram classificados de acordo com as três categorias acima listadas. O resultado da classificação foi sumarizado nas Tabelas 4 a 6 que estão em um Anexo no final do artigo. As tabelas contém as referências dos artigos separados por categoria e subcategorias. Os dados foram analisados quantitativamente levando em conta a classificação proposta e enfatizando o número de publicações em cada categoria. A próxima seção mostrará a última etapa do processo de mapeamento sistemático, onde os resultados das análises feitas serão mostrados e as questões de pesquisa respondidas.

4 Análise e Discussões

Nesta seção serão apresentadas análises quantitativas e qualitativas do mapeamento sistemático relacionado às linguagens de consulta em grafos. A análise quantitativa apresentará os números de acordo com a classificação que foi indicada na seção anterior. A análise qualitativa apresentará relações entre os resultados quando for o caso, e discussões acerca dos números.

Recentemente, numerosos projetos para processamento, consulta e análise de bases de dados em grafos tem surgido [78]. Este retorno de interesse por esse modelo pode ser observado no gráfico da Figura 2. Este gráfico, lista a quantidade de artigos incluídos no mapeamento publicados por ano.

No gráfico é possível observar um crescimento do número de publicações a partir de 2008. Este mapeamento sistemático recuperou artigos publicados entre 1989 e 2015. Do total de artigos utilizados neste mapeamento 85% deles foram publicados entre 2009 e 2014. A condução da pesquisa de artigos nas bibliotecas foi feita até o dia 06/11/2014, por essa razão a quantidade de artigos publicados em 2015 está baixa. Antes de 2009 as publicações referentes às linguagens de consulta eram mais numerosas na área de XML[10]. Pode-se observar no gráfico que o número de artigos publicados nesta época na área de linguagens de consulta em grafos não foi expressivo. Antes da década de 2000 o número de publicações foi pequeno. Os anos de 1989, 1990, 1992, 1993, 1996 e 1999 tiveram um artigo em cada ano, e dois artigos no ano 2000.

O gráfico da Figura 3 apresenta o número de publicações por ano e por base de dados. O gráfico indica as três bases de dados usadas nesse mapeamento: SD (*Science Direct*), IEEE (*IEEE explore*) e ACM (*ACM digital library*). É possível observar que a base ACM se mantém com o maior número de publicações ao longo dos anos. Os artigos repetidos e que foram removidos não tem influência nesse resultado já que foi desta base a maior quantidade de artigos repetidos que foram excluídos. Isso pode ser constatado na Tabela 3.

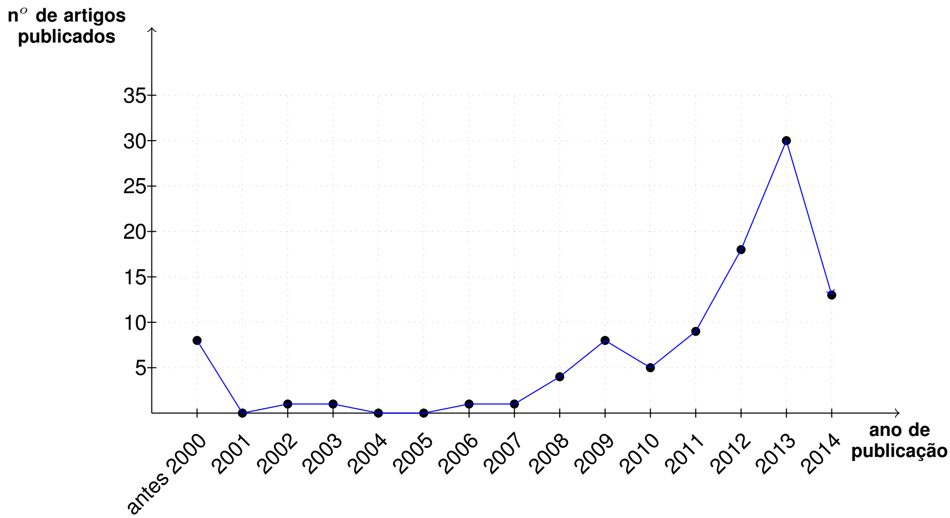

Figura 2. Publicações por ano.

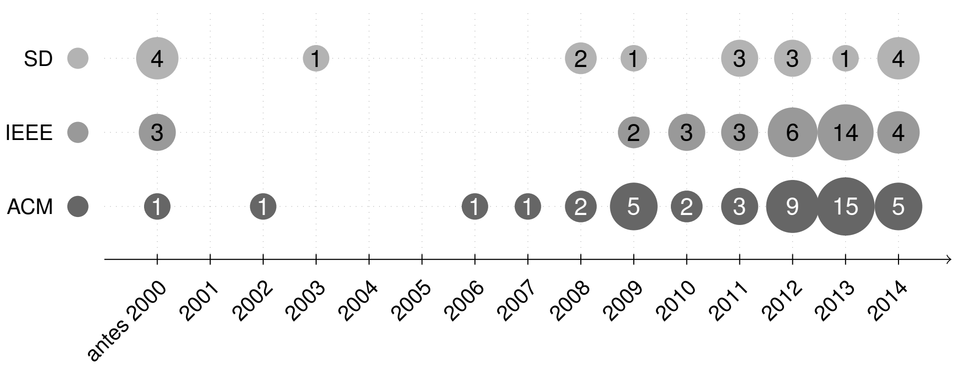

Figura 3. Publicações por ano e editora.

Há vários fatores que podem explicar esse retorno do modelo de dados em grafos. As redes sociais digitais, que ganharam popularidade nos últimos anos, são comumente representadas pelo modelo de dados em grafos. Os grafos são uma boa opção para as redes sociais por causa da sua característica de interconectividade, onde os vértices são entidades e as arestas são os relacionamentos entre as entidades. Nesses contextos a topologia, ou seja a estrutura do grafo, é importante [3]. Neste mapeamento os tópicos relacionados às redes sociais são abordados nos resumos de 25% dos artigos selecionados.

Os grandes volumes de dados ou grandes grafos são gerados pelas redes sociais digitais e também estão presentes em diversos outros segmentos como análise social e *business intelligence* [40] [69] [34] [9] [90], farmacologia [37], partição e gerenciamento de dados [98] [87] [19] [81] [105], catálogos de referências bibliográficas [80], entre outros. Os tópicos relacionados a *big data* são abordados nos resumos de 48% dos artigos listados.

4.1 Análise das categorias

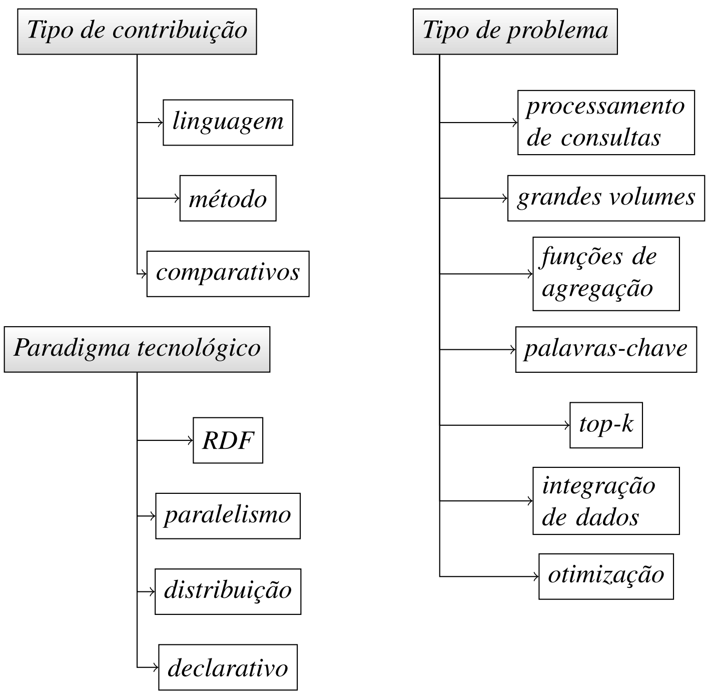

Figura 4. Divisão das categorias

Nesta seção serão analisadas as categorias de forma quantitativa, e também serão destacadas percepções a partir dos números apresentados. A Figura 4 apresenta um organograma de como as categorias foram divididas.

4.1.1 Tipo de contribuição Esta categoria teve como propósito indicar a principal contribuição do trabalho, o seu objetivo. A classificação do tipo de contribuição atribuía a um artigo se ele descrevia uma nova linguagem, um método ou conduzia um estudo comparativo. Como complemento a classificação também indicava se o artigo apresentava experimentos, e disponibilizava uma ferramenta.

Os artigos que apresentam uma nova linguagem de consulta ou uma extensão de uma linguagem já existente totalizaram 23%. Os artigos que apresentavam um método totalizaram 62%, e os estudos comparativos, que englobam *surveys* e estado-da-arte (visão atual sobre uma determinada área de estudo) totalizaram 17%. De todos esses artigos, 36% conduziram e apresentaram experimentos, e apenas 7% disponibilizavam ferramentas.

O número de artigos que conduziram e apresentaram experimentos foi considerada pequena pelos autores, sendo que menos da metade os apresentaram (36%). O experimentos são importantes na validação de um método ou ferramenta. Eles também são importantes para dar a possibilidade da audiência poder replicar o experimento e fazer comparativos.

4.1.2 Tipo de problema Esta categoria teve como propósito indicar o tipo de problema que o artigo pretende atacar. A classificação do tipo de problema se deu da seguinte forma: técnicas de processamento de consultas, gerenciamento de grandes volumes de dados, implementação de funções de agregação, técnicas de consultas por palavras-chave, técnicas de consultas *top-k*, integração de dados, e métodos de otimização de desempenho.

Os artigos que tem como foco técnicas de aperfeiçoamento do processamento de consultas são a maioria, representam 90%. Uma grande parte dos trabalhos está interessado em resolver problemas com o gerenciamento de grandes volumes de dados, estes representam 48%. A implementação de funções de agregação foi citada em 11% dos artigos. Consultas por palavras-chave e integração de dados representam 11% cada. Técnicas de consultas *top-k*, onde o número de resultados de uma consulta é limitado a um valor *k* representam 10%. E por fim, os artigos que implementam métodos de otimização de desempenho somaram 12%.

Como o foco principal deste mapeamento está nas linguagens de consulta para grafos era esperado que a maioria dos artigos tivessem como foco algum método ou algoritmo para melhoramento do processamento de consultas de uma forma geral. Dentro desta categoria, também foi possível observar que dentre todos os artigos que lidam com grandes volumes de dados (48% do total), 66% destes também implementam métodos para otimizar o seu desempenho. Lidar com um grande volume de dados frequentemente vem associado à busca de soluções para a melhoria do desempenho.

O aparecimento de diversos artigos que focam em consultas *top-k* mostra que nem sempre resultados completos são necessários. Dependendo do contexto, limitar estes resultados pode ser uma boa solução. Alguns trabalhos que lidam com *big data* como [71] e

[51] implementam consultas *top-k* que mostram apenas os resultados mais relevantes, seja porque os resultados completos não são necessários ou por uma questão de desempenho do algoritmo.

4.1.3 Paradigma tecnológico Esta categoria teve como propósito identificar o modelo de programação, paradigma de programação ou o tipo de tecnologia empregada. O paradigma de interesse para este mapeamento foi o declarativo, também entraram na classificação os grafos do tipo RDF, processamento paralelo e a distribuição de dados.

O paradigma declarativo foi detectado em 37% do total de artigos. Dentre os artigos que apresentam novas linguagens, 70% destas novas linguagens usam o paradigma declarativo. Os artigos que usam o modelo de dados RDF totalizaram 23%. Artigos que utilizam processamento paralelo somaram 12% e a distribuição de dados foi identificada em 8%. O uso do processamento paralelo, embora bastante popular, apresentou baixa percentagem neste mapeamento. Isso é explicado pelo fato de que o contexto deste trabalho está mais voltado para bancos de dados. O uso de termos como "linguagem de consulta" (*query language*) limita o aparecimento de trabalhos mais voltados para o processamento paralelo.

4.2 Relação entre as categorias

As Figuras 5 e 6 apresentam dois gráficos que fazem uma relação entre as categorias. O gráfico da Figura 5 apresenta uma relação entre a categoria "Tipo de contribuição" com as categorias "Tipo de problema" e "Paradigma tecnológico". O gráfico da Figura 6 apresenta uma relação entre as categorias "Tipo de problema" e "Paradigma tecnológico". Nestes gráficos é possível observar o número de publicações relacionando as subcategorias entre si.

Através do gráfico mostrado na Figura 5 é possível visualizar um maior volume de publicações relacionadas à melhoria no processamento de consultas e situações que envolvem grandes volumes de dados (lado esquerdo do gráfico). No lado direito do gráfico é possível visualizar que os grafos RDF também tem presença expressiva, especialmente com relação subcategoria "Método". Também é expressiva a relação entre as contribuições "Linguagem" e "Método" e o paradigma "Declarativo", onde 70% (16/23) das linguagem e 30% (18/62) dos métodos estão relacionados ao paradigma declarativo.

No gráfico da Figura 6 é possível fazer a seguinte observação: 61% (14/23) dos trabalhos que usam grafos RDF como base de dados estão associados com os grandes grafos. E ainda, (6/11) 55% implementam funções de agregação.

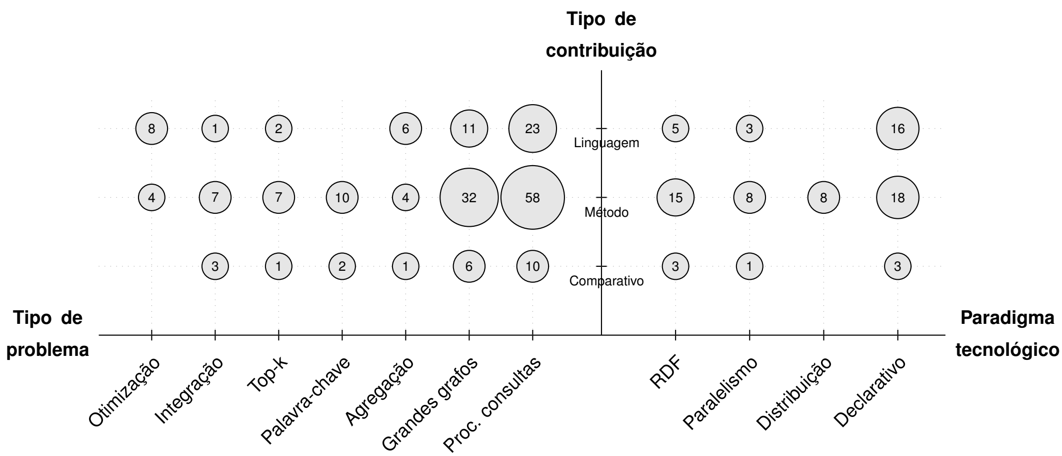

Figura 5. Relação entre a categoria "Tipo de contribuição" e as categorias "Tipo de problema" e "Paradigma tecnológico".

4.3 Questões de pesquisa

Nesta seção será apresentada uma análise baseada nas questões de pesquisa que foram lançadas no início deste mapeamento.

- *QP1: Quais são as linguagens de consulta mais usadas no modelo de dados em grafo ao longo dos anos?*

  Como não há uma linguagem padrão para uma base de dados em grafo, cada linguagem é criada ou estendida para atender ao seu problema específico [12]. A linguagem mais usada como base para a criação de novas linguagens é SQL (20% do total), a linguagem padrão para o modelo relacional, ou pelo menos estas novas linguagens usam uma sintaxe muito similar. Dentre elas podemos citar SQL/TC [23], BiQL [30], SoQL [79], SCOPE [108], e o trabalho apresentado em [97]. Dentre as baseadas em Datalog, uma linguagem de consulta para base de dados dedutivos, podemos citar SociaLite [84], DatalogFS [65], Glog [34]. Dentre as baseadas em SPARQL, a linguagem de consulta recomendada para RDF, tem-se G-SPARQL [80], MashQL [46], SPARankQL [15]. A linguagem Cypher, que é a linguagem de consulta para sistemas Neo4j, foi usada como referência em [95][60][14]. Cypher também tem sido usada em comparativos, como em [42][4][43]. Outras linguagens que serviram como referência para novas linguagens são OQL e XPath.

- *QP2: Quantas dessas linguagens possuem suporte à agregação?*

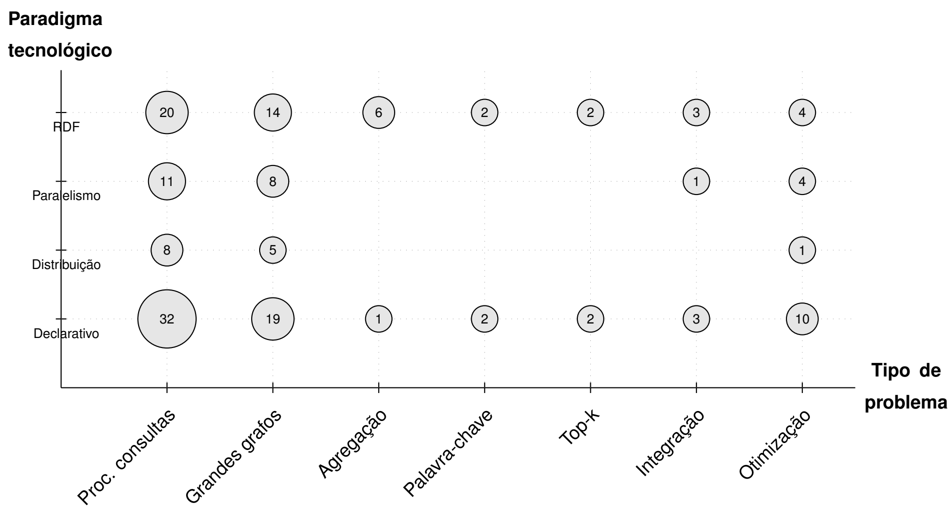

Figura 6. Relação entre "Tipo de problema" e "Paradigma tecnológico".

A agregação é uma função bastante usada no modelo relacional, mas que não tem a mesma popularidade no modelo de dados para grafo. Uma função de agregação usa o resultado de uma consulta como valor intermediário para fazer uma outra operação. As funções de agregação mais comuns são soma, média, maior valor e menor valor. Os trabalhos que apontaram para o uso de funções de agregação somam apenas 11% do total. Aqui pode-se sinalizar uma área com potencial a ser explorado.

- *QP3: Quais paradigmas são mais usados na definição das linguagens de consulta para o modelo de dados em grafos?*

  Das linguagens que foram identificadas neste mapeamento, a maioria delas, 70% são linguagens declarativas. Dentre estas, a maior parte usam o estilo de consultas "`SELECT -WHERE -FROM`" que é conhecido como o estilo SQL. O uso desse estilo é justificado muitas vezes pelo fato de ser muito conhecido, o que tornaria a compreensão da nova linguagem mais rápida.

5 Conclusões

Este mapeamento sistemático apresentou uma investigação acerca de linguagens de consulta para base de dados em grafos. Foram usados um total de 101 artigos que foram distribuídos em categorias que facilitaram a extração e análise de informações sobre o tema.

A partir desta análise foi concluído que o uso da base de dados em grafos teve um aumento no número de publicações a partir da década de 2000. Com o avanço da internet e das redes sociais, a busca por diferentes tecnologias além do modelo relacional ficou evidenciada.

O modelo de dados em grafos não possui uma linguagem de consulta padrão, a maioria das contribuições são específicas para o domínio em que é aplicado. No entanto, os grafos RDF que possuem uma recomendação (*W3C recomendation*) para linguagem de consulta, usam o SPARQL como referência para novas linguagens e extensões.

Os números deste mapeamento destacaram o uso do paradigma declarativo para a construção de novas linguagens, sem indicativo claro de uma mudança nesse sentido para os próximos anos. Também foi identificado que as funções de agregação não são muito exploradas nas linguagens de consulta, indicando que este tópico tem potencial para trabalhos futuros.

Além das conclusões específicas deste estudo, o mesmo serviu para repertoriar as principais contribuições da área de linguagens de consulta para bases de dados em grafos, servindo, assim, como fonte de referências bibliográficas da área, em complemento a outras fontes como [101][3][5].

Contribuição dos autores:

- Simone de Oliveira SANTOS: Conduziu a pesquisa e escreveu o artigo.

- Mirian Halfeld Ferrari Alves e Martin A. Musicante: Supervisão da condução da pesquisa. Os co-autores indicaram bases de dados, trabalharam nas formulação das strings de busca e revisaram a escrita do artigo. Também foram os responsáveis pelo direcionamento referente aos tipos de gráficos apresentados no trabalho e associações a serem feitas.

Referências

[1] R. Angles. A comparison of current graph database models. In *IEEE 28th International Conference on Data Engineering Workshops (ICDEW)*, pages 171–177, April 2012.

[2] Renzo Angles and Claudio Gutierrez. Querying RDF data from a graph database perspective. In Asunción Gómez-Pérez and Jérôme Euzenat, editors, *The Semantic Web: Research and Applications*, volume 3532 of *Lecture Notes in Computer Science*, pages 346–360. Springer Berlin Heidelberg, 2005.

[3] Renzo Angles and Claudio Gutierrez. Survey of graph database models. *ACM Comput. Surv.*, 40(1):1:1–1:39, February 2008.

[4] Renzo Angles, Arnau Prat-Pérez, David Dominguez-Sal, and Josep-Lluis Larriba-Pey. Benchmarking database systems for social network applications. In *First International Workshop on Graph Data Management Experiences and Systems*, GRADES '13, pages 15:1–15:7, New York, NY, USA, 2013. ACM.

[5] James Bailey, François Bry, Tim Furche, and Sebastian Schaffert. Web and semantic web languages: A survey. In Norbert Eisinger and Jan Maluszynski, editors, *Reasoning Web*, volume 3564 of *Lecture Notes in Computer Science*, pages 35–133. Springer Berlin Heidelberg, 2005.

[6] Pablo Barceló Baeza. Querying graph databases. In *Proceedings of the 32nd Symposium on Principles of Database Systems*, PODS '13, pages 175–188, New York, NY, USA, 2013. ACM.

[7] Yijun Bei, Zhen Lin, Chen Zhao, and Xiaojun Zhu. HBase system-based distributed framework for searching large graph databases. In *14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 151–156, July 2013.

[8] Tim Berners-Lee and Jeffrey Jaffe. World Wide Web Consortium (W3C).

[9] Dritan Bleco and Yannis Kotidis. Business intelligence on complex graph data. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, EDBT-ICDT '12, pages 13–20, New York, NY, USA, 2012. ACM.

[10] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau, and John Cowan. Extensible Markup Language (XML) 1.1. W3C Recomendation. August 2006.

[11] M. Brocheler, A. Pugliese, and V.S. Subrahmanian. COSI: Cloud oriented subgraph identification in massive social networks. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 248–255, Aug 2010.

[12] Mike Buerli and Cal Poly San Luis Obispo. The current state of graph databases. *Department of Computer Science, Cal Poly San Luis Obispo, mbuerli@calpoly.edu*, 2012.

[13] V.G. Castellana, A. Tumeo, O. Villa, D. Haglin, and J. Feo. Composing data parallel code for a SPARQL graph engine. In *International Conference on Social Computing (SocialCom)*, pages 691–699, Sept 2013.

[14] Ciro Cattuto, Marco Quaggiotto, André Panisson, and Alex Averbuch. Time-varying social networks in a graph database: A Neo4J use case. In *First International Workshop on Graph Data Management Experiences and Systems*, GRADES '13, pages 11:1–11:6, New York, NY, USA, 2013. ACM.

[15] Juan P. Cedeño and K. Selçuk Candan. R2DF framework for ranked path queries over weighted RDF graphs. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 40:1–40:12, New York, NY, USA, 2011. ACM.

[16] Alan Chappell, Sutanay Choudhury, John Feo, David Haglin, Alessandro Morari, Sumit Purohit, Karen Schuchardt, Antonino Tumeo, Jesse Weaver, and Oreste Villa. Toward a data scalable solution for facilitating discovery of scientific data resources. In *Proceedings of the 2013 International Workshop on Data-Intensive Scalable Computing Systems*, DISCS-2013, pages 55–60, New York, NY, USA, 2013. ACM.

[17] Yi Chen, Wei Wang, and Ziyang Liu. Keyword-based search and exploration on databases. In *IEEE 27th International Conference on Data Engineering (ICDE)*, pages 1380–1383, April 2011.

[18] M. P. Consens and A. O. Mendelzon. Expressing structural hypertext queries in graphlog. In *Proceedings of the Second Annual ACM Conference on Hypertext*, HYPERTEXT '89, pages 269–292, New York, NY, USA, 1989. ACM.

[19] O. Cure, F. Kerdjoudj, Chan Le Duc, M. Lamolle, and D. Faye. On the potential integration of an ontology-based data access approach in NoSQL stores. In *Third International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)*, pages 166–173, Sept 2012.

[20] Michael Curtiss, Iain Becker, Tudor Bosman, Sergey Doroshenko, Lucian Grijincu, Tom Jackson, Sandhya Kunnatur, Soren Lassen, Philip Pronin, Sriram Sankar, Guanghao Shen, Gintaras Woss, Chao Yang, and Ning Zhang. Unicorn: A system for searching the social graph. *Proc. VLDB Endow.*, 6(11):1150–1161, August 2013.

[21] A. Cuzzocrea and P. Serafino. A reachability-based theoretical framework for modeling and querying complex probabilistic graph data. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1177–1184, Oct 2012.

[22] Richard Cyganiak, David Wood, Markus Lanthaler, Graham Klyne, Jeremy J. Carroll, and Brian McBride. Resource Description Framework (RDF): Concepts and abstract syntax. W3C Recomendation. February 2014.

[23] S. Dar and R. Agrawal. Extending SQL with generalized transitive closure. *IEEE Transactions on Knowledge and Data Engineering*, 5(5):799–812, Oct 1993.

[24] Roberto De Virgilio, Antonio Maccioni, and Riccardo Torlone. Converting relational to graph databases. In *First International Workshop on Graph Data Management Experiences and Systems*, GRADES '13, pages 1:1–1:6, New York, NY, USA, 2013. ACM.

[25] Renaud Delbru, Stephane Campinas, and Giovanni Tummarello. Searching web data: An entity retrieval and high-performance indexing model. *Web Semantics: Science, Services and Agents on the World Wide Web*, 10(0):33 – 58, 2012. Web-Scale Semantic Information Processing.

[26] Saumen Dey, Víctor Cuevas-Vicenttín, Sven Köhler, Eric Gribkoff, Michael Wang, and Bertram Ludäscher. On implementing provenance-aware regular path queries with relational query engines. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 214–223, New York, NY, USA, 2013. ACM.

[27] Ying Ding. Community detection: Topological vs. topical. *Journal of Informetrics*, 5(4):498 – 514, 2011.

[28] Yerach Doytsher, Ben Galon, and Yaron Kanza. Querying geo-social data by bridging spatial networks and social networks. In *Proceedings of the 2Nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, LBSN '10, pages 39–46, New York, NY, USA, 2010. ACM.

[29] Anton Dries and Siegfried Nijssen. Analyzing graph databases by aggregate queries. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, MLG '10, pages 37–45, New York, NY, USA, 2010. ACM.

[30] Anton Dries, Siegfried Nijssen, and Luc De Raedt. A query language for analyzing networks. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 485–494, New York, NY, USA, 2009. ACM.

[31] V. Dritsou, P. Constantopoulos, A. Deligiannakis, and Y. Kotidis. Shortcut selection in RDF databases. In *IEEE 27th International Conference on Data Engineering Workshops (ICDEW)*, pages 194–199, April 2011.

[32] Wenfei Fan, Xin Wang, and Yinghui Wu. Incremental graph pattern matching. *ACM Trans. Database Syst.*, 38(3):18:1–18:47, September 2013.

[33] L. Fegaras. Supporting Bulk Synchronous Parallelism in Map-Reduce queries. In *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:*, pages 1068–1077, Nov 2012.

[34] Jun Gao, Jiashuai Zhou, Chang Zhou, and J.X. Yu. GLog: A high level graph analysis system using mapreduce. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 544–555, March 2014.

[35] A. González-Beltrán, P. Milligan, and P. Sage. Range queries over skip tree graphs. *Computer Communications*, 31(2):358 – 374, 2008. Special Issue: Foundation of Peer-to-Peer Computing.

[36] Stewart Greenhill and Svetha Venkatesh. Semantic data modelling and visualisation using noetica. *Data & Knowledge Engineering*, 33(3):241 – 276, 2000.

[37] Paul Groth, Antonis Loizou, Alasdair J.G. Gray, Carole Goble, Lee Harland, and Steve Pettifer. Api-centric linked data integration: The open PHACTS discovery platform case study. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2014.

[38] Amarnath Gupta and Simone Santini. On querying OBO ontologies using a DAG pattern query language. In *Proceedings of the Third International Conference on Data Integration in the Life Sciences*, DILS'06, pages 152–167, Berlin, Heidelberg, 2006. Springer-Verlag.

[39] Hao He, Haixun Wang, Jun Yang, and Philip S. Yu. BLINKS: Ranked keyword searches on graphs. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, SIGMOD '07, pages 305–316, New York, NY, USA, 2007. ACM.

[40] J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 51–60, Oct 2011.

[41] Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres, and Stefan Decker. Searching and browsing linked data with SWSE: The semantic web search engine. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4):365 – 401, 2011. {JWS} special issue on Semantic Search.

[42] Aidan Hogan, Jürgen Umbrich, Andreas Harth, Richard Cyganiak, Axel Polleres, and Stefan Decker. An empirical survey of linked data conformance. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14(0):14 – 44, 2012. Special Issue on Dealing with the Messiness of the Web of Data.

[43] Florian Holzschuher and René Peinl. Performance of graph query languages: Comparison of cypher, gremlin and native access in Neo4J. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 195–204, New York, NY, USA, 2013. ACM.

[44] T. Härder, B. Mitschang, and H. Schöning. Query processing for complex objects. *Data & Knowledge Engineering*, 7(3):181 – 200, 1992.

[45] H.M. Jamil. Design of declarative graph query languages: On the choice between value, pattern and object based representations for graphs. In *IEEE 28th International Conference on Data Engineering Workshops (ICDEW)*, pages 178–185, April 2012.

[46] M. Jarrar and M.D. Dikaiakos. Querying the data web: The MashQL approach. *Internet Computing, IEEE*, 14(3):58–67, May 2010.

[47] Ruoming Jin, Yang Xiang, Ning Ruan, and David Fuhry. 3-HOP: A high-compression indexing scheme for reachability query. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, SIGMOD '09, pages 813–826, New York, NY, USA, 2009. ACM.

[48] S. Jouili and V. Vansteenberghe. An empirical comparison of graph databases. In *International Conference on Social Computing (SocialCom)*, pages 708–715, Sept 2013.

[49] Andrey Kan, Jeffrey Chan, James Bailey, and Christopher Leckie. A query based approach for mining evolving graphs. In *Proceedings of the Eighth Australasian Data Mining Conference - Volume 101*, AusDM '09, pages 139–150, Darlinghurst, Australia, Australia, 2009. Australian Computer Society, Inc.

[50] G. Karvounarakis, A. Magganaraki, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, and K. Tolle. Querying the semantic web with RQL. *Computer Networks*, 42(5):617 – 640, 2003. The Semantic Web: an evolution for a revolution.

[51] G. Kasneci, M. Ramanath, M. Sozio, F.M. Suchanek, and G. Weikum. STAR: Steiner-tree approximation in relationship graphs. In *IEEE 25th International Conference on Data Engineering - ICDE'09*, pages 868–879, March 2009.

[52] K. Kaur and R. Rani. Modeling and querying data in NoSQL databases. In *IEEE International Conference on Big Data*, pages 1–7, Oct 2013.

[53] A. Khan, Yinghui Wu, and X. Yan. Emerging graph queries in linked data. In *IEEE 28th International Conference on Data Engineering (ICDE)*, pages 1218–1221, April 2012.

[54] Arijit Khan, Nan Li, Xifeng Yan, Ziyu Guan, Supriyo Chakraborty, and Shu Tao. Neighborhood based fast graph search in large networks. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, pages 901–912, New York, NY, USA, 2011. ACM.

[55] Benny Kimelfeld and Yehoshua Sagiv. Efficiently enumerating results of keyword search over data graphs. *Information Systems*, 33(45):335 – 359, 2008. Selected Papers from the Tenth International Symposium on Database Programming Languages (DBPL 2005.

[56] Taekyong Lee, Lei Sheng, T. Bozkaya, N.H. Balkir, Z. Meral Ozsoyoglu, and Z.M. Ozsoyoglu. Querying multimedia presentations based on content. *IEEE Transactions on Knowledge and Data Engineering*, 11(3):361–385, May 1999.

[57] M. Leida and A. Chu. Distributed SPARQL query answering over RDF data streams. In *IEEE International Congress on Big Data (BigData Congress)*, pages 369–378, June 2013.

[58] Xiang Lian, Eugenio De Hoyos, Artem Chebotko, Bin Fu, and Christine Reilly. k-nearest keyword search in RDF graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, pages 40 – 56, 2013.

[59] Vitaliy Liptchinsky, Benjamin Satzger, Rostyslav Zabolotnyi, and Schahram Dustdar. Expressive languages for selecting groups from graph-structured data. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 761–770, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[60] Yunkai Liu and T.M. Vitolo. Graph data warehouse: Steps to integrating graph databases into the traditional conceptual structure of a data warehouse. In *IEEE International Congress on Big Data (BigData Congress)*, pages 433–434, June 2013.

[61] Michel Mainguenaud. Constraint-based queries in a geographical database for network facilities. *Computers, Environment and Urban Systems*, 20(2):139 – 151, 1996.

[62] Federica Mandreoli, Riccardo Martoglia, Giorgio Villani, and Wilma Penzo. Flexible query answering on graph-modeled data. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 216–227, New York, NY, USA, 2009. ACM.

[63] Norbert Martinez-Bazan and David Dominguez-Sal. Using semijoin programs to solve traversal queries in graph databases. In *Proceedings of Workshop on GRAph Data Management Experiences and Systems*, GRADES'14, pages 6:1–6:6, New York, NY, USA, 2014. ACM.

[64] Yosi Mass and Yehoshua Sagiv. Language models for keyword search over data graphs. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 363–372, New York, NY, USA, 2012. ACM.

[65] Mirjana Mazuran, Edoardo Serra, and Carlo Zaniolo. Extending the power of datalog recursion. *The VLDB Journal*, 22(4):471–493, August 2013.

[66] Jayanta Mondal and Amol Deshpande. EAGr: Supporting continuous ego-centric aggregate queries over large dynamic graphs. In *Proceedings of the 2014 ACM SIGMOD*

International Conference on Management of Data, SIGMOD '14, pages 1335–1346, New York, NY, USA, 2014. ACM.

[67] A. Morari, V.G. Castellana, D. Haglin, J. Feo, J. Weaver, A. Tumeo, and O. Villa. Accelerating semantic graph databases on commodity clusters. In *IEEE International Conference on Big Data*, pages 768–772, Oct 2013.

[68] W.E. Moustafa, A. Kimmig, A. Deshpande, and L. Getoor. Subgraph pattern matching over uncertain graphs with identity linkage uncertainty. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 904–915, March 2014.

[69] A. Naseer, L. Laera, and T. Matsutsuka. Enterprise biggraph. In *46th Hawaii International Conference on System Sciences (HICSS)*, pages 1005–1014, Jan 2013.

[70] María Constanza Pabón, Claudia Roncancio, and Martha Millán. Graph data transformations and querying. In *Proceedings of the 2014 International C* Conference on Computer Science & Software Engineering*, C3S2E '14, pages 20:1–20:6, New York, NY, USA, 2014. ACM.

[71] Chang-Sup Park and Sungchae Lim. Efficient processing of keyword queries over graph databases for finding effective answers. *Information Processing & Management*, 51(1):42 – 57, 2015.

[72] Raqueline RM Penteado, Rebeca Schroeder, Diego Hoss, Jaqueline Nande, Ricardo M Maeda, Walmir O Couto, and Carmem S Hara. Um estudo sobre bancos de dados em grafos nativos. X ERBD - Escola Regional de Banco de Dados.

[73] A. Petermann, M. Junghanns, R. Muller, and E. Rahm. BIIIG: Enabling business intelligence with integrated instance graphs. In *IEEE 30th International Conference on Data Engineering Workshops (ICDEW)*, pages 4–11, March 2014.

[74] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. Systematic mapping studies in software engineering. In *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*, EASE'08, pages 68–77, Swinton, UK, 2008. British Computer Society.

[75] Letao Qi, H.T. Lin, and V. Honavar. Clustering remote RDF data using SPARQL update queries. In *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 236–242, April 2013.

[76] C. Raguenaud, J. Kennedy, and P.J. Barclay. The Prometheus taxonomic database. In *Proceedings IEEE International Symposium on Bio-Informatics and Biomedical Engineering*, pages 63–70, 2000.

[77] Padmashree Ravindra, HyeongSik Kim, and Kemafor Anyanwu. To nest or not to nest, when and how much: Representing intermediate results of graph pattern queries in mapreduce based processing. In *Proceedings of the 4th International Workshop on Semantic Web Information Management*, SWIM '12, pages 5:1–5:8, New York, NY, USA, 2012. ACM.

[78] Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases*. O'Reilly Media, Inc., 2013.

[79] R. Ronen and O. Shmueli. SoQL: A language for querying and creating data in social networks. In *ICDE'09. IEEE 25th International Conference on Data Engineering*, pages 1595–1602, March 2009.

[80] Sherif Sakr, Sameh Elnikety, and Yuxiong He. Hybrid query execution engine for large attributed graphs. *Information Systems*, 41(0):45 – 73, 2014.

[81] Semih Salihoglu and Jennifer Widom. GPS: A graph processing system. In *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*, SSDBM, pages 22:1–22:12, New York, NY, USA, 2013. ACM.

[82] Mohamed Sarwat, Sameh Elnikety, Yuxiong He, and Mohamed F. Mokbel. Horton+: A distributed system for processing declarative reachability queries over partitioned graphs. *Proc. VLDB Endow.*, 6(14):1918–1929, September 2013.

[83] Timos K. Sellis, Nick Roussopoulos, and Raymond T. Ng. Efficient compilation of large rule bases using logical access paths. *Information Systems*, 15(1):73 – 84, 1990. Knowledge Engineering.

[84] Jiwon Seo, S. Guo, and M.S. Lam. SociaLite: Datalog extensions for efficient social network analysis. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 278–289, April 2013.

[85] Hong-Han Shuai, De-Nian Yang, P.S. Yu, Chih-Ya Shen, and Ming-Syan Chen. On pattern preserving graph generation. In *IEEE 13th International Conference on Data Mining (ICDM)*, pages 677–686, Dec 2013.

[86] Steve Speicher, John Arwe, and Ashok Malhotra. Linked data platform 1.0. W3C Recomendation. February 2015.

[87] S.R. Spillane, J. Birnbaum, D. Bokser, D. Kemp, A. Labouseur, P.W. Olsen, J. Vijayan, Jeong-Hyon Hwang, and Jun-Weon Yoon. A demonstration of the G* graph database system. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 1356–1359, April 2013.

[88] Dan Suciu. Distributed query evaluation on semistructured data. *ACM Trans. Database Syst.*, 27(1):1–62, March 2002.

[89] Zhao Sun, Hongzhi Wang, Haixun Wang, Bin Shao, and Jianzhong Li. Efficient subgraph matching on billion node graphs. *Proc. VLDB Endow.*, 5(9):788–799, May 2012.

[90] Yuanyuan Tian, Jignesh M. Patel, Viji Nair, Sebastian Martini, and Matthias Kretzler. Periscope/GQ: A graph querying toolkit. *Proc. VLDB Endow.*, 1(2):1404–1407, August 2008.

[91] Alberto Tonon, Gianluca Demartini, and Philippe Cudré-Mauroux. Combining inverted indices and structured search for ad-hoc object retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 125–134, New York, NY, USA, 2012. ACM.

[92] Thanh Tran, Haofen Wang, and Peter Haase. Hermes: Data web search on a pay-as-you-go integration infrastructure. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):189 – 203, 2009. The Web of Data.

[93] Ying Tu and Han-Wei Shen. GraphCharter: Combining browsing with query to explore large semantic graphs. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pages 49–56, Feb 2013.

[94] M. Uddin, R. Stadler, M. Miyazawa, and M. Hayashi. Graph search for cloud network management. In *Network Operations and Management Symposium (NOMS), 2014 IEEE*, pages 1–5, May 2014.

[95] Raoul-Gabriel Urma and Alan Mycroft. Source-code queries with graph databases with application to programming language usage and evolution. *Science of Computer Programming*, 97, Part 1(0):127 – 134, 2015. Special Issue on New Ideas and Emerging Results in Understanding Software.

[96] Ramakrishna Varadarajan, Vagelis Hristidis, Louiqa Raschid, Maria-Esther Vidal, Luis Ibáñez, and Héctor Rodríguez-Drumond. Flexible and efficient querying and ranking on hyperlinked data sources. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 553–564, New York, NY, USA, 2009. ACM.

[97] Fan Wang and Gagan Agrawal. Answering complex structured queries over the deep web. In *Proceedings of the 15th Symposium on International Database Engineering & Applications*, IDEAS '11, pages 115–123, New York, NY, USA, 2011. ACM.

[98] Rui Wang and K. Chiu. A stream partitioning approach to processing large scale distributed graph datasets. In *IEEE International Conference on Big Data*, pages 537–542, Oct 2013.

[99] Jesse Weaver, Vito Giovanni Castellana, Alessandro Morari, Antonino Tumeo, Sumit Purohit, Alan Chappell, David Haglin, Oreste Villa, Sutanay Choudhury, Karen Schuchardt, and John Feo. Toward a data scalable solution for facilitating discovery of science resources. *Parallel Computing*, 40(10):682 – 696, 2014.

[100] Klara Weiand, Fabian KneiSSl, Wojciech Lobacz, Tim Furche, and François Bry. pest: Fast approximate keyword search in semantic data using eigenvector-based term propagation. *Information Systems*, 37(4):372 – 390, 2012. Semantic Web Data Management.

[101] Peter T. Wood. Query languages for graph databases. *SIGMOD Rec.*, 41(1):50–60, April 2012.

[102] Wenlei Xie, Guozhang Wang, David Bindel, Alan Demers, and Johannes Gehrke. Fast iterative graph computation with block updates. *Proc. VLDB Endow.*, 6(14):2014–2025, September 2013.

[103] P. Yadav and V. Samala. Benchmarking over a semantic repository. In *Second International Conference on Advanced Computing (ICoAC)*, pages 51–59, Dec 2010.

[104] Shengqi Yang, Yanan Xie, Yinghui Wu, Tianyu Wu, Huan Sun, Jian Wu, and Xifeng Yan. SLQ: A user-friendly graph querying system. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD ’14, pages 893–896, New York, NY, USA, 2014. ACM.

[105] Shengqi Yang, Xifeng Yan, Bo Zong, and Arijit Khan. Towards effective partition management for large graphs. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD ’12, pages 517–528, New York, NY, USA, 2012. ACM.

[106] Hilmi Yildirim, Vineet Chaoji, and Mohammed J. Zaki. GRAIL: A scalable index for reachability queries in very large graphs. *The VLDB Journal*, 21(4):509–534, August 2012.

[107] Kai Zeng, Jiacheng Yang, Haixun Wang, Bin Shao, and Zhongyuan Wang. A distributed graph engine for web scale RDF data. In *Proceedings of the 39th international conference on Very Large Data Bases*, PVLDB’13, pages 265–276. VLDB Endowment, 2013.

[108] Jingren Zhou, Nicolas Bruno, Ming-Chuan Wu, Per-Ake Larson, Ronnie Chaiken, and Darren Shakib. SCOPE: Parallel databases meet MapReduce. *The VLDB Journal*, 21(5):611–636, October 2012.

[109] Linhong Zhu, Wee Keong Ng, and James Cheng. Structure and attribute index for approximate graph matching in large graphs. *Information Systems*, 36(6):958 – 972, 2011.

[110] Lei Zou, M. Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gStore: A graph-based SPARQL query engine. *The VLDB Journal*, 23(4):565–590, August 2014.

ANEXO

Tabela 4: Categoria: tipo de contribuição.

Referências dos artigos da subcategoria Linguagem
- Sherif Sakr, Sameh Elnikety, and Yuxiong He. Hybrid query execution engine for large attributed graphs. *Information Systems*, 41(0):45 – 73, 2014
- G. Karvounarakis, A. Magganaraki, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, and K. Tolle. Querying the semantic web with RQL. *Computer Networks*, 42(5):617 – 640, 2003. The Semantic Web: an evolution for a revolution
- R. Ronen and O. Shmueli. SoQL: A language for querying and creating data in social networks. In *ICDE'09. IEEE 25th International Conference on Data Engineering*, pages 1595–1602, March 2009
- J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 51–60, Oct 2011
- Jiwon Seo, S. Guo, and M.S. Lam. SociaLite: Datalog extensions for efficient social network analysis. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 278–289, April 2013
- C. Raguenaud, J. Kennedy, and P.J. Barclay. The Prometheus taxonomic database. In *Proceedings IEEE International Symposium on Bio-Informatics and Biomedical Engineering*, pages 63–70, 2000
- Jun Gao, Jiashuai Zhou, Chang Zhou, and J.X. Yu. GLog: A high level graph analysis system using mapreduce. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 544–555, March 2014
- P. Yadav and V. Samala. Benchmarking over a semantic repository. In *Second International Conference on Advanced Computing (ICoAC)*, pages 51–59, Dec 2010
- Taekyong Lee, Lei Sheng, T. Bozkaya, N.H. Balkir, Z. Meral Ozsoyoglu, and Z.M. Ozsoyoglu. Querying multimedia presentations based on content. *IEEE Transactions on Knowledge and Data Engineering*, 11(3):361–385, May 1999
- S. Dar and R. Agrawal. Extending SQL with generalized transitive closure. *IEEE Transactions on Knowledge and Data Engineering*, 5(5):799–812, Oct 1993
- M. Jarrar and M.D. Dikaiakos. Querying the data web: The MashQL approach. *Internet Computing, IEEE*, 14(3):58–67, May 2010
- Anton Dries, Siegfried Nijssen, and Luc De Raedt. A query language for analyzing networks. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 485–494, New York, NY, USA, 2009. ACM
- Anton Dries and Siegfried Nijssen. Analyzing graph databases by aggregate queries. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, MLG '10, pages 37–45, New York, NY, USA, 2010. ACM
- Fan Wang and Gagan Agrawal. Answering complex structured queries over the deep web. In *Proceedings of the 15th Symposium on International Database Engineering & Applications*, IDEAS '11, pages 115–123, New York, NY, USA, 2011. ACM
- M. P. Consens and A. O. Mendelzon. Expressing structural hypertext queries in graphlog. In *Proceedings of the Second Annual ACM Conference on Hypertext*, HYPERTEXT '89, pages 269–292, New York, NY, USA, 1989. ACM

- Vitaliy Liptchinsky, Benjamin Satzger, Rostyslav Zabolotnyi, and Schahram Dustdar. Expressive languages for selecting groups from graph-structured data. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 761–770, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee
- Mirjana Mazuran, Edoardo Serra, and Carlo Zaniolo. Extending the power of datalog recursion. *The VLDB Journal*, 22(4):471–493, August 2013
- Mohamed Sarwat, Sameh Elnikety, Yuxiong He, and Mohamed F. Mokbel. Horton+: A distributed system for processing declarative reachability queries over partitioned graphs. *Proc. VLDB Endow.*, 6(14):1918–1929, September 2013
- Amarnath Gupta and Simone Santini. On querying OBO ontologies using a DAG pattern query language. In *Proceedings of the Third International Conference on Data Integration in the Life Sciences*, DILS'06, pages 152–167, Berlin, Heidelberg, 2006. Springer-Verlag
- Yerach Doytsher, Ben Galon, and Yaron Kanza. Querying geo-social data by bridging spatial networks and social networks. In *Proceedings of the 2Nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, LBSN '10, pages 39–46, New York, NY, USA, 2010. ACM
- Juan P. Cedeño and K. Selçuk Candan. R2DF framework for ranked path queries over weighted RDF graphs. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 40:1–40:12, New York, NY, USA, 2011. ACM
- Jingren Zhou, Nicolas Bruno, Ming-Chuan Wu, Per-Ake Larson, Ronnie Chaiken, and Darren Shakib. SCOPE: Parallel databases meet MapReduce. *The VLDB Journal*, 21(5):611–636, October 2012
- Michael Curtiss, Iain Becker, Tudor Bosman, Sergey Doroshenko, Lucian Grijincu, Tom Jackson, Sandhya Kunnatur, Soren Lassen, Philip Pronin, Sriram Sankar, Guanghao Shen, Gintaras Woss, Chao Yang, and Ning Zhang. Unicorn: A system for searching the social graph. *Proc. VLDB Endow.*, 6(11):1150–1161, August 2013
.

Referências dos artigos subcategoria Método

- Linhong Zhu, Wee Keong Ng, and James Cheng. Structure and attribute index for approximate graph matching in large graphs. *Information Systems*, 36(6):958 – 972, 2011
- Chang-Sup Park and Sungchae Lim. Efficient processing of keyword queries over graph databases for finding effective answers. *Information Processing & Management*, 51(1):42 – 57, 2015
- Renaud Delbru, Stephane Campinas, and Giovanni Tummarello. Searching web data: An entity retrieval and high-performance indexing model. *Web Semantics: Science, Services and Agents on the World Wide Web*, 10(0):33 – 58, 2012. Web-Scale Semantic Information Processing
- Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres, and Stefan Decker. Searching and browsing linked data with SWSE: The semantic web search engine. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4):365 – 401, 2011. {JWS} special issue on Semantic Search
- Thanh Tran, Haofen Wang, and Peter Haase. Hermes: Data web search on a pay-as-you-go integration infrastructure. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):189 – 203, 2009. The Web of Data

- T. Härder, B. Mitschang, and H. Schöning. Query processing for complex objects. *Data & Knowledge Engineering*, 7(3):181 – 200, 1992
- Timos K. Sellis, Nick Roussopoulos, and Raymond T. Ng. Efficient compilation of large rule bases using logical access paths. *Information Systems*, 15(1):73 – 84, 1990. Knowledge Engineering
- Benny Kimelfeld and Yehoshua Sagiv. Efficiently enumerating results of keyword search over data graphs. *Information Systems*, 33(45):335 – 359, 2008. Selected Papers from the Tenth International Symposium on Database Programming Languages (DBPL 2005
- Raoul-Gabriel Urma and Alan Mycroft. Source-code queries with graph databases with application to programming language usage and evolution. *Science of Computer Programming*, 97, Part 1(0):127 – 134, 2015. Special Issue on New Ideas and Emerging Results in Understanding Software
- Klara Weiand, Fabian KneiSSl, Wojciech Lobacz, Tim Furche, and François Bry. pest: Fast approximate keyword search in semantic data using eigenvector-based term propagation. *Information Systems*, 37(4):372 – 390, 2012. Semantic Web Data Management
- Xiang Lian, Eugenio De Hoyos, Artem Chebotko, Bin Fu, and Christine Reilly. k-nearest keyword search in RDF graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, pages 40 – 56, 2013
- Michel Mainguenaud. Constraint-based queries in a geographical database for network facilities. *Computers, Environment and Urban Systems*, 20(2):139 – 151, 1996
- Stewart Greenhill and Svetha Venkatesh. Semantic data modelling and visualisation using noetica. *Data & Knowledge Engineering*, 33(3):241 – 276, 2000
- H.M. Jamil. Design of declarative graph query languages: On the choice between value, pattern and object based representations for graphs. In *IEEE 28th International Conference on Data Engineering Workshops (ICDEW)*, pages 178–185, April 2012
- Hong-Han Shuai, De-Nian Yang, P.S. Yu, Chih-Ya Shen, and Ming-Syan Chen. On pattern preserving graph generation. In *IEEE 13th International Conference on Data Mining (ICDM)*, pages 677–686, Dec 2013
- A. Petermann, M. Junghanns, R. Muller, and E. Rahm. BIIIG: Enabling business intelligence with integrated instance graphs. In *IEEE 30th International Conference on Data Engineering Workshops (ICDEW)*, pages 4–11, March 2014
- A. Morari, V.G. Castellana, D. Haglin, J. Feo, J. Weaver, A. Tumeo, and O. Villa. Accelerating semantic graph databases on commodity clusters. In *IEEE International Conference on Big Data*, pages 768–772, Oct 2013
- Yijun Bei, Zhen Lin, Chen Zhao, and Xiaojun Zhu. HBase system-based distributed framework for searching large graph databases. In *14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 151–156, July 2013
- A. Naseer, L. Laera, and T. Matsutsuka. Enterprise biggraph. In *46th Hawaii International Conference on System Sciences (HICSS)*, pages 1005–1014, Jan 2013
- M. Uddin, R. Stadler, M. Miyazawa, and M. Hayashi. Graph search for cloud network management. In *Network Operations and Management Symposium (NOMS), 2014 IEEE*, pages 1–5, May 2014

- V.G. Castellana, A. Tumeo, O. Villa, D. Haglin, and J. Feo. Composing data parallel code for a SPARQL graph engine. In *International Conference on Social Computing (SocialCom)*, pages 691–699, Sept 2013
- W.E. Moustafa, A. Kimmig, A. Deshpande, and L. Getoor. Subgraph pattern matching over uncertain graphs with identity linkage uncertainty. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 904–915, March 2014
- M. Brocheler, A. Pugliese, and V.S. Subrahmanian. COSI: Cloud oriented subgraph identification in massive social networks. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 248–255, Aug 2010
- A. Cuzzocrea and P. Serafino. A reachability-based theoretical framework for modeling and querying complex probabilistic graph data. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1177–1184, Oct 2012
- Rui Wang and K. Chiu. A stream partitioning approach to processing large scale distributed graph datasets. In *IEEE International Conference on Big Data*, pages 537–542, Oct 2013
- S.R. Spillane, J. Birnbaum, D. Bokser, D. Kemp, A. Labouseur, P.W. Olsen, J. Vijayan, Jeong-Hyon Hwang, and Jun-Weon Yoon. A demonstration of the G* graph database system. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 1356–1359, April 2013
- Yunkai Liu and T.M. Vitolo. Graph data warehouse: Steps to integrating graph databases into the traditional conceptual structure of a data warehouse. In *IEEE International Congress on Big Data (BigData Congress)*, pages 433–434, June 2013
- L. Fegaras. Supporting Bulk Synchronous Parallelism in Map-Reduce queries. In *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:*, pages 1068–1077, Nov 2012
- O. Cure, F. Kerdjoudj, Chan Le Duc, M. Lamolle, and D. Faye. On the potential integration of an ontology-based data access approach in NoSQL stores. In *Third International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)*, pages 166–173, Sept 2012
- M. Leida and A. Chu. Distributed SPARQL query answering over RDF data streams. In *IEEE International Congress on Big Data (BigData Congress)*, pages 369–378, June 2013
- Ying Tu and Han-Wei Shen. GraphCharter: Combining browsing with query to explore large semantic graphs. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pages 49–56, Feb 2013
- V. Dritsou, P. Constantopoulos, A. Deligiannakis, and Y. Kotidis. Shortcut selection in RDF databases. In *IEEE 27th International Conference on Data Engineering Workshops (ICDEW)*, pages 194–199, April 2011
- Letao Qi, H.T. Lin, and V. Honavar. Clustering remote RDF data using SPARQL update queries. In *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 236–242, April 2013
- G. Kasneci, M. Ramanath, M. Sozio, F.M. Suchanek, and G. Weikum. STAR: Steiner-tree approximation in relationship graphs. In *IEEE 25th International Conference on Data Engineering - ICDE'09*, pages 868–879, March 2009
- Ruoming Jin, Yang Xiang, Ning Ruan, and David Fuhry. 3-HOP: A high-compression indexing scheme for reachability query. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, SIGMOD '09, pages 813–826, New York, NY, USA, 2009. ACM

- Kai Zeng, Jiacheng Yang, Haixun Wang, Bin Shao, and Zhongyuan Wang. A distributed graph engine for web scale RDF data. In *Proceedings of the 39th international conference on Very Large Data Bases*, PVLDB'13, pages 265–276. VLDB Endowment, 2013
- Andrey Kan, Jeffrey Chan, James Bailey, and Christopher Leckie. A query based approach for mining evolving graphs. In *Proceedings of the Eighth Australasian Data Mining Conference - Volume 101*, AusDM '09, pages 139–150, Darlinghurst, Australia, Australia, 2009. Australian Computer Society, Inc
- Hao He, Haixun Wang, Jun Yang, and Philip S. Yu. BLINKS: Ranked keyword searches on graphs. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, SIGMOD '07, pages 305–316, New York, NY, USA, 2007. ACM
- Dritan Bleco and Yannis Kotidis. Business intelligence on complex graph data. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, EDBT-ICDT '12, pages 13–20, New York, NY, USA, 2012. ACM
- Alberto Tonon, Gianluca Demartini, and Philippe Cudré-Mauroux. Combining inverted indices and structured search for ad-hoc object retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 125–134, New York, NY, USA, 2012. ACM
- Roberto De Virgilio, Antonio Maccioni, and Riccardo Torlone. Converting relational to graph databases. In *First International Workshop on Graph Data Management Experiences and Systems*, GRADES '13, pages 1:1–1:6, New York, NY, USA, 2013. ACM
- Dan Suciu. Distributed query evaluation on semistructured data. *ACM Trans. Database Syst.*, 27(1):1–62, March 2002
- Jayanta Mondal and Amol Deshpande. EAGr: Supporting continuous ego-centric aggregate queries over large dynamic graphs. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 1335–1346, New York, NY, USA, 2014. ACM
- Zhao Sun, Hongzhi Wang, Haixun Wang, Bin Shao, and Jianzhong Li. Efficient subgraph matching on billion node graphs. *Proc. VLDB Endow.*, 5(9):788–799, May 2012
- Wenlei Xie, Guozhang Wang, David Bindel, Alan Demers, and Johannes Gehrke. Fast iterative graph computation with block updates. *Proc. VLDB Endow.*, 6(14):2014–2025, September 2013
- Ramakrishna Varadarajan, Vagelis Hristidis, Louiqa Raschid, Maria-Esther Vidal, Luis Ibáñez, and Héctor Rodríguez-Drumond. Flexible and efficient querying and ranking on hyperlinked data sources. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 553–564, New York, NY, USA, 2009. ACM
- Federica Mandreoli, Riccardo Martoglia, Giorgio Villani, and Wilma Penzo. Flexible query answering on graph-modeled data. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 216–227, New York, NY, USA, 2009. ACM
- Semih Salihoglu and Jennifer Widom. GPS: A graph processing system. In *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*, SSDBM, pages 22:1–22:12, New York, NY, USA, 2013. ACM
- Hilmi Yildirim, Vineet Chaoji, and Mohammed J. Zaki. GRAIL: A scalable index for reachability queries in very large graphs. *The VLDB Journal*, 21(4):509–534, August 2012

- María Constanza Pabón, Claudia Roncancio, and Martha Millán. Graph data transformations and querying. In *Proceedings of the 2014 International C* Conference on Computer Science & Software Engineering*, C3S2E '14, pages 20:1–20:6, New York, NY, USA, 2014. ACM
- Lei Zou, M. Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gStore: A graph-based SPARQL query engine. *The VLDB Journal*, 23(4):565–590, August 2014
- Wenfei Fan, Xin Wang, and Yinghui Wu. Incremental graph pattern matching. *ACM Trans. Database Syst.*, 38(3):18:1–18:47, September 2013
- Yosi Mass and Yehoshua Sagiv. Language models for keyword search over data graphs. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 363–372, New York, NY, USA, 2012. ACM
- Arijit Khan, Nan Li, Xifeng Yan, Ziyu Guan, Supriyo Chakraborty, and Shu Tao. Neighborhood based fast graph search in large networks. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, pages 901–912, New York, NY, USA, 2011. ACM
- Saumen Dey, Víctor Cuevas-Vicenttín, Sven Köhler, Eric Gribkoff, Michael Wang, and Bertram Ludäscher. On implementing provenance-aware regular path queries with relational query engines. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 214–223, New York, NY, USA, 2013. ACM
- Yuanyuan Tian, Jignesh M. Patel, Viji Nair, Sebastian Martini, and Matthias Kretzler. Periscope/GQ: A graph querying toolkit. *Proc. VLDB Endow.*, 1(2):1404–1407, August 2008
- Shengqi Yang, Yanan Xie, Yinghui Wu, Tianyu Wu, Huan Sun, Jian Wu, and Xifeng Yan. SLQ: A user-friendly graph querying system. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 893–896, New York, NY, USA, 2014. ACM
- Ciro Cattuto, Marco Quaggiotto, André Panisson, and Alex Averbuch. Time-varying social networks in a graph database: A Neo4J use case. In *First International Workshop on Graph Data Management Experiences and Systems*, GRADES '13, pages 11:1–11:6, New York, NY, USA, 2013. ACM
- Padmashree Ravindra, HyeongSik Kim, and Kemafor Anyanwu. To nest or not to nest, when and how much: Representing intermediate results of graph pattern queries in mapreduce based processing. In *Proceedings of the 4th International Workshop on Semantic Web Information Management*, SWIM '12, pages 5:1–5:8, New York, NY, USA, 2012. ACM
- Alan Chappell, Sutanay Choudhury, John Feo, David Haglin, Alessandro Morari, Sumit Purohit, Karen Schuchardt, Antonino Tumeo, Jesse Weaver, and Oreste Villa. Toward a data scalable solution for facilitating discovery of scientific data resources. In *Proceedings of the 2013 International Workshop on Data-Intensive Scalable Computing Systems*, DISCS-2013, pages 55–60, New York, NY, USA, 2013. ACM
- Shengqi Yang, Xifeng Yan, Bo Zong, and Arijit Khan. Towards effective partition management for large graphs. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, pages 517–528, New York, NY, USA, 2012. ACM
- Norbert Martinez-Bazan and David Dominguez-Sal. Using semijoin programs to solve traversal queries in graph databases. In *Proceedings of Workshop on GRAph Data Management Experiences and Systems*, GRADES'14, pages 6:1–6:6, New York, NY, USA, 2014. ACM

.

Referências dos artigos subcategoria Comparação

- Jesse Weaver, Vito Giovanni Castellana, Alessandro Morari, Antonino Tumeo, Sumit Purohit, Alan Chappell, David Haglin, Oreste Villa, Sutanay Choudhury, Karen Schuchardt, and John Feo. Toward a data scalable solution for facilitating discovery of science resources. *Parallel Computing*, 40(10):682 – 696, 2014
- Renaud Delbru, Stephane Campinas, and Giovanni Tummarello. Searching web data: An entity retrieval and high-performance indexing model. *Web Semantics: Science, Services and Agents on the World Wide Web*, 10(0):33 – 58, 2012. Web-Scale Semantic Information Processing
- Aidan Hogan, Jürgen Umbrich, Andreas Harth, Richard Cyganiak, Axel Polleres, and Stefan Decker. An empirical survey of linked data conformance. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14(0):14 – 44, 2012. Special Issue on Dealing with the Messiness of the Web of Data
- A. González-Beltrán, P. Milligan, and P. Sage. Range queries over skip tree graphs. *Computer Communications*, 31(2):358 – 374, 2008. Special Issue: Foundation of Peer-to-Peer Computing
-
- Paul Groth, Antonis Loizou, Alasdair J.G. Gray, Carole Goble, Lee Harland, and Steve Pettifer. Api-centric linked data integration: The open PHACTS discovery platform case study. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2014
- Ying Ding. Community detection: Topological vs. topical. *Journal of Informetrics*, 5(4):498 – 514, 2011
- A. Khan, Yinghui Wu, and X. Yan. Emerging graph queries in linked data. In *IEEE 28th International Conference on Data Engineering (ICDE)*, pages 1218–1221, April 2012
- S. Jouili and V. Vansteenberghe. An empirical comparison of graph databases. In *International Conference on Social Computing (SocialCom)*, pages 708–715, Sept 2013
- K. Kaur and R. Rani. Modeling and querying data in NoSQL databases. In *IEEE International Conference on Big Data*, pages 1–7, Oct 2013
- R. Angles. A comparison of current graph database models. In *IEEE 28th International Conference on Data Engineering Workshops (ICDEW)*, pages 171–177, April 2012
- Yi Chen, Wei Wang, and Ziyang Liu. Keyword-based search and exploration on databases. In *IEEE 27th International Conference on Data Engineering (ICDE)*, pages 1380–1383, April 2011
- Renzo Angles, Arnau Prat-Pérez, David Dominguez-Sal, and Josep-Lluis Larriba-Pey. Benchmarking database systems for social network applications. In *First International Workshop on Graph Data Management Experiences and Systems*, GRADES '13, pages 15:1–15:7, New York, NY, USA, 2013. ACM
- Florian Holzschuher and René Peinl. Performance of graph query languages: Comparison of cypher, gremlin and native access in Neo4J. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 195–204, New York, NY, USA, 2013. ACM
- Peter T. Wood. Query languages for graph databases. *SIGMOD Rec.*, 41(1):50–60, April 2012
- Pablo Barceló Baeza. Querying graph databases. In *Proceedings of the 32nd Symposium on Principles of Database Systems*, PODS '13, pages 175–188, New York, NY, USA, 2013. ACM

- Renzo Angles and Claudio Gutierrez. Survey of graph database models. *ACM Comput. Surv.*, 40(1):1:1–1:39, February 2008

.

Referências dos artigos subcategoria Experimentos

- Linhong Zhu, Wee Keong Ng, and James Cheng. Structure and attribute index for approximate graph matching in large graphs. *Information Systems*, 36(6):958 – 972, 2011
- Chang-Sup Park and Sungchae Lim. Efficient processing of keyword queries over graph databases for finding effective answers. *Information Processing & Management*, 51(1):42 – 57, 2015
- Thanh Tran, Haofen Wang, and Peter Haase. Hermes: Data web search on a pay-as-you-go integration infrastructure. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):189 – 203, 2009. The Web of Data
- A. González-Beltrán, P. Milligan, and P. Sage. Range queries over skip tree graphs. *Computer Communications*, 31(2):358 – 374, 2008. Special Issue: Foundation of Peer-to-Peer Computing
- Raoul-Gabriel Urma and Alan Mycroft. Source-code queries with graph databases with application to programming language usage and evolution. *Science of Computer Programming*, 97, Part 1(0):127 – 134, 2015. Special Issue on New Ideas and Emerging Results in Understanding Software
- Klara Weiand, Fabian KneiSSl, Wojciech Lobacz, Tim Furche, and François Bry. pest: Fast approximate keyword search in semantic data using eigenvector-based term propagation. *Information Systems*, 37(4):372 – 390, 2012. Semantic Web Data Management
- Xiang Lian, Eugenio De Hoyos, Artem Chebotko, Bin Fu, and Christine Reilly. k-nearest keyword search in RDF graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, pages 40 – 56, 2013
- Hong-Han Shuai, De-Nian Yang, P.S. Yu, Chih-Ya Shen, and Ming-Syan Chen. On pattern preserving graph generation. In *IEEE 13th International Conference on Data Mining (ICDM)*, pages 677–686, Dec 2013
- A. Petermann, M. Junghanns, R. Muller, and E. Rahm. BIIIG: Enabling business intelligence with integrated instance graphs. In *IEEE 30th International Conference on Data Engineering Workshops (ICDEW)*, pages 4–11, March 2014
- A. Morari, V.G. Castellana, D. Haglin, J. Feo, J. Weaver, A. Tumeo, and O. Villa. Accelerating semantic graph databases on commodity clusters. In *IEEE International Conference on Big Data*, pages 768–772, Oct 2013
- S. Jouili and V. Vansteenberghe. An empirical comparison of graph databases. In *International Conference on Social Computing (SocialCom)*, pages 708–715, Sept 2013
- Yijun Bei, Zhen Lin, Chen Zhao, and Xiaojun Zhu. HBase system-based distributed framework for searching large graph databases. In *14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 151–156, July 2013
- J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 51–60, Oct 2011

- W.E. Moustafa, A. Kimmig, A. Deshpande, and L. Getoor. Subgraph pattern matching over uncertain graphs with identity linkage uncertainty. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 904–915, March 2014
- Jiwon Seo, S. Guo, and M.S. Lam. SociaLite: Datalog extensions for efficient social network analysis. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 278–289, April 2013
- Rui Wang and K. Chiu. A stream partitioning approach to processing large scale distributed graph datasets. In *IEEE International Conference on Big Data*, pages 537–542, Oct 2013
- Jun Gao, Jiashuai Zhou, Chang Zhou, and J.X. Yu. GLog: A high level graph analysis system using mapreduce. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 544–555, March 2014
- Ying Tu and Han-Wei Shen. GraphCharter: Combining browsing with query to explore large semantic graphs. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pages 49–56, Feb 2013
- P. Yadav and V. Samala. Benchmarking over a semantic repository. In *Second International Conference on Advanced Computing (ICoAC)*, pages 51–59, Dec 2010
- Kai Zeng, Jiacheng Yang, Haixun Wang, Bin Shao, and Zhongyuan Wang. A distributed graph engine for web scale RDF data. In *Proceedings of the 39th international conference on Very Large Data Bases*, PVLDB'13, pages 265–276. VLDB Endowment, 2013
- Andrey Kan, Jeffrey Chan, James Bailey, and Christopher Leckie. A query based approach for mining evolving graphs. In *Proceedings of the Eighth Australasian Data Mining Conference - Volume 101*, AusDM '09, pages 139–150, Darlinghurst, Australia, Australia, 2009. Australian Computer Society, Inc
- Roberto De Virgilio, Antonio Maccioni, and Riccardo Torlone. Converting relational to graph databases. In *First International Workshop on Graph Data Management Experiences and Systems*, GRADES '13, pages 1:1–1:6, New York, NY, USA, 2013. ACM
- Jayanta Mondal and Amol Deshpande. EAGr: Supporting continuous ego-centric aggregate queries over large dynamic graphs. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 1335–1346, New York, NY, USA, 2014. ACM
- Zhao Sun, Hongzhi Wang, Haixun Wang, Bin Shao, and Jianzhong Li. Efficient subgraph matching on billion node graphs. *Proc. VLDB Endow.*, 5(9):788–799, May 2012
- Vitaliy Liptchinsky, Benjamin Satzger, Rostyslav Zabolotnyi, and Schahram Dustdar. Expressive languages for selecting groups from graph-structured data. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 761–770, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee
- Wenlei Xie, Guozhang Wang, David Bindel, Alan Demers, and Johannes Gehrke. Fast iterative graph computation with block updates. *Proc. VLDB Endow.*, 6(14):2014–2025, September 2013
- Federica Mandreoli, Riccardo Martoglia, Giorgio Villani, and Wilma Penzo. Flexible query answering on graph-modeled data. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 216–227, New York, NY, USA, 2009. ACM

- Semih Salihoglu and Jennifer Widom. GPS: A graph processing system. In *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*, SSDBM, pages 22:1–22:12, New York, NY, USA, 2013. ACM
- Lei Zou, M. Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gStore: A graph-based SPARQL query engine. *The VLDB Journal*, 23(4):565–590, August 2014
- Mohamed Sarwat, Sameh Elnikety, Yuxiong He, and Mohamed F. Mokbel. Horton+: A distributed system for processing declarative reachability queries over partitioned graphs. *Proc. VLDB Endow.*, 6(14):1918–1929, September 2013
- Wenfei Fan, Xin Wang, and Yinghui Wu. Incremental graph pattern matching. *ACM Trans. Database Syst.*, 38(3):18:1–18:47, September 2013
- Arijit Khan, Nan Li, Xifeng Yan, Ziyu Guan, Supriyo Chakraborty, and Shu Tao. Neighborhood based fast graph search in large networks. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, pages 901–912, New York, NY, USA, 2011. ACM
- Saumen Dey, Víctor Cuevas-Vicenttín, Sven Köhler, Eric Gribkoff, Michael Wang, and Bertram Ludäscher. On implementing provenance-aware regular path queries with relational query engines. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 214–223, New York, NY, USA, 2013. ACM
- Juan P. Cedeño and K. Selçuk Candan. R2DF framework for ranked path queries over weighted RDF graphs. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 40:1–40:12, New York, NY, USA, 2011. ACM
- Shengqi Yang, Xifeng Yan, Bo Zong, and Arijit Khan. Towards effective partition management for large graphs. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, pages 517–528, New York, NY, USA, 2012. ACM
.

Referências dos artigos subcategoria Ferramenta

- Stewart Greenhill and Svetha Venkatesh. Semantic data modelling and visualisation using noetica. *Data & Knowledge Engineering*, 33(3):241 – 276, 2000
- R. Ronen and O. Shmueli. SoQL: A language for querying and creating data in social networks. In *ICDE'09. IEEE 25th International Conference on Data Engineering*, pages 1595–1602, March 2009
- A. Naseer, L. Laera, and T. Matsutsuka. Enterprise biggraph. In *46th Hawaii International Conference on System Sciences (HICSS)*, pages 1005–1014, Jan 2013
- Ying Tu and Han-Wei Shen. GraphCharter: Combining browsing with query to explore large semantic graphs. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pages 49–56, Feb 2013
- Hilmi Yildirim, Vineet Chaoji, and Mohammed J. Zaki. GRAIL: A scalable index for reachability queries in very large graphs. *The VLDB Journal*, 21(4):509–534, August 2012
- Yuanyuan Tian, Jignesh M. Patel, Viji Nair, Sebastian Martini, and Matthias Kretzler. Periscope/GQ: A graph querying toolkit. *Proc. VLDB Endow.*, 1(2):1404–1407, August 2008
- Jingren Zhou, Nicolas Bruno, Ming-Chuan Wu, Per-Ake Larson, Ronnie Chaiken, and Darren Shakib. SCOPE: Parallel databases meet MapReduce. *The VLDB Journal*, 21(5):611–636, October 2012

Tabela 5: Categoria: tipo de problema.

Referências dos artigos subcategoria Processamento de consultas
- Sherif Sakr, Sameh Elnikety, and Yuxiong He. Hybrid query execution engine for large attributed graphs. *Information Systems*, 41(0):45 – 73, 2014
- Jesse Weaver, Vito Giovanni Castellana, Alessandro Morari, Antonino Tumeo, Sumit Purohit, Alan Chappell, David Haglin, Oreste Villa, Sutanay Choudhury, Karen Schuchardt, and John Feo. Toward a data scalable solution for facilitating discovery of science resources. *Parallel Computing*, 40(10):682 – 696, 2014
- Linhong Zhu, Wee Keong Ng, and James Cheng. Structure and attribute index for approximate graph matching in large graphs. *Information Systems*, 36(6):958 – 972, 2011
- Chang-Sup Park and Sungchae Lim. Efficient processing of keyword queries over graph databases for finding effective answers. *Information Processing & Management*, 51(1):42 – 57, 2015
- Renaud Delbru, Stephane Campinas, and Giovanni Tummarello. Searching web data: An entity retrieval and high-performance indexing model. *Web Semantics: Science, Services and Agents on the World Wide Web*, 10(0):33 – 58, 2012. Web-Scale Semantic Information Processing
- Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres, and Stefan Decker. Searching and browsing linked data with SWSE: The semantic web search engine. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4):365 – 401, 2011. {JWS} special issue on Semantic Search
- Thanh Tran, Haofen Wang, and Peter Haase. Hermes: Data web search on a pay-as-you-go integration infrastructure. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):189 – 203, 2009. The Web of Data
- T. Härder, B. Mitschang, and H. Schöning. Query processing for complex objects. *Data & Knowledge Engineering*, 7(3):181 – 200, 1992
- Timos K. Sellis, Nick Roussopoulos, and Raymond T. Ng. Efficient compilation of large rule bases using logical access paths. *Information Systems*, 15(1):73 – 84, 1990. Knowledge Engineering
- Raoul-Gabriel Urma and Alan Mycroft. Source-code queries with graph databases with application to programming language usage and evolution. *Science of Computer Programming*, 97, Part 1(0):127 – 134, 2015. Special Issue on New Ideas and Emerging Results in Understanding Software
- Klara Weiand, Fabian KneiSSl, Wojciech Lobacz, Tim Furche, and François Bry. pest: Fast approximate keyword search in semantic data using eigenvector-based term propagation. *Information Systems*, 37(4):372 – 390, 2012. Semantic Web Data Management
- Xiang Lian, Eugenio De Hoyos, Artem Chebotko, Bin Fu, and Christine Reilly. k-nearest keyword search in RDF graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, pages 40 – 56, 2013
- Michel Mainguenaud. Constraint-based queries in a geographical database for network facilities. *Computers, Environment and Urban Systems*, 20(2):139 – 151, 1996
- Stewart Greenhill and Svetha Venkatesh. Semantic data modelling and visualisation using noetica. *Data & Knowledge Engineering*, 33(3):241 – 276, 2000

- G. Karvounarakis, A. Magganaraki, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, and K. Tolle. Querying the semantic web with RQL. *Computer Networks*, 42(5):617 – 640, 2003. The Semantic Web: an evolution for a revolution
- H.M. Jamil. Design of declarative graph query languages: On the choice between value, pattern and object based representations for graphs. In *IEEE 28th International Conference on Data Engineering Workshops (ICDEW)*, pages 178–185, April 2012
- A. Khan, Yinghui Wu, and X. Yan. Emerging graph queries in linked data. In *IEEE 28th International Conference on Data Engineering (ICDE)*, pages 1218–1221, April 2012
- A. Petermann, M. Junghanns, R. Muller, and E. Rahm. BIIIG: Enabling business intelligence with integrated instance graphs. In *IEEE 30th International Conference on Data Engineering Workshops (ICDEW)*, pages 4–11, March 2014
- R. Ronen and O. Shmueli. SoQL: A language for querying and creating data in social networks. In *ICDE'09. IEEE 25th International Conference on Data Engineering*, pages 1595–1602, March 2009
- Yijun Bei, Zhen Lin, Chen Zhao, and Xiaojun Zhu. HBase system-based distributed framework for searching large graph databases. In *14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 151–156, July 2013
- J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 51–60, Oct 2011
- A. Naseer, L. Laera, and T. Matsutsuka. Enterprise biggraph. In *46th Hawaii International Conference on System Sciences (HICSS)*, pages 1005–1014, Jan 2013
- M. Uddin, R. Stadler, M. Miyazawa, and M. Hayashi. Graph search for cloud network management. In *Network Operations and Management Symposium (NOMS), 2014 IEEE*, pages 1–5, May 2014
- V.G. Castellana, A. Tumeo, O. Villa, D. Haglin, and J. Feo. Composing data parallel code for a SPARQL graph engine. In *International Conference on Social Computing (SocialCom)*, pages 691–699, Sept 2013
- W.E. Moustafa, A. Kimmig, A. Deshpande, and L. Getoor. Subgraph pattern matching over uncertain graphs with identity linkage uncertainty. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 904–915, March 2014
- Jiwon Seo, S. Guo, and M.S. Lam. SociaLite: Datalog extensions for efficient social network analysis. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 278–289, April 2013
- M. Brocheler, A. Pugliese, and V.S. Subrahmanian. COSI: Cloud oriented subgraph identification in massive social networks. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 248–255, Aug 2010
- A. Cuzzocrea and P. Serafino. A reachability-based theoretical framework for modeling and querying complex probabilistic graph data. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1177–1184, Oct 2012
- Rui Wang and K. Chiu. A stream partitioning approach to processing large scale distributed graph datasets. In *IEEE International Conference on Big Data*, pages 537–542, Oct 2013

- K. Kaur and R. Rani. Modeling and querying data in NoSQL databases. In *IEEE International Conference on Big Data*, pages 1–7, Oct 2013
- S.R. Spillane, J. Birnbaum, D. Bokser, D. Kemp, A. Labouseur, P.W. Olsen, J. Vijayan, Jeong-Hyon Hwang, and Jun-Weon Yoon. A demonstration of the G* graph database system. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 1356–1359, April 2013
- R. Angles. A comparison of current graph database models. In *IEEE 28th International Conference on Data Engineering Workshops (ICDEW)*, pages 171–177, April 2012
- Yunkai Liu and T.M. Vitolo. Graph data warehouse: Steps to integrating graph databases into the traditional conceptual structure of a data warehouse. In *IEEE International Congress on Big Data (BigData Congress)*, pages 433–434, June 2013
- C. Raguenaud, J. Kennedy, and P.J. Barclay. The Prometheus taxonomic database. In *Proceedings IEEE International Symposium on Bio-Informatics and Biomedical Engineering*, pages 63–70, 2000
- Jun Gao, Jiashuai Zhou, Chang Zhou, and J.X. Yu. GLog: A high level graph analysis system using mapreduce. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 544–555, March 2014
- L. Fegaras. Supporting Bulk Synchronous Parallelism in Map-Reduce queries. In *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:*, pages 1068–1077, Nov 2012
- O. Cure, F. Kerdjoudj, Chan Le Duc, M. Lamolle, and D. Faye. On the potential integration of an ontology-based data access approach in NoSQL stores. In *Third International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)*, pages 166–173, Sept 2012
- M. Leida and A. Chu. Distributed SPARQL query answering over RDF data streams. In *IEEE International Congress on Big Data (BigData Congress)*, pages 369–378, June 2013
- Ying Tu and Han-Wei Shen. GraphCharter: Combining browsing with query to explore large semantic graphs. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pages 49–56, Feb 2013
- P. Yadav and V. Samala. Benchmarking over a semantic repository. In *Second International Conference on Advanced Computing (ICoAC)*, pages 51–59, Dec 2010
- Taekyong Lee, Lei Sheng, T. Bozkaya, N.H. Balkir, Z. Meral Ozsoyoglu, and Z.M. Ozsoyoglu. Querying multimedia presentations based on content. *IEEE Transactions on Knowledge and Data Engineering*, 11(3):361–385, May 1999
- S. Dar and R. Agrawal. Extending SQL with generalized transitive closure. *IEEE Transactions on Knowledge and Data Engineering*, 5(5):799–812, Oct 1993
- V. Dritsou, P. Constantopoulos, A. Deligiannakis, and Y. Kotidis. Shortcut selection in RDF databases. In *IEEE 27th International Conference on Data Engineering Workshops (ICDEW)*, pages 194–199, April 2011
- Letao Qi, H.T. Lin, and V. Honavar. Clustering remote RDF data using SPARQL update queries. In *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 236–242, April 2013
- G. Kasneci, M. Ramanath, M. Sozio, F.M. Suchanek, and G. Weikum. STAR: Steiner-tree approximation in relationship graphs. In *IEEE 25th International Conference on Data Engineering - ICDE'09*, pages 868–879, March 2009

- M. Jarrar and M.D. Dikaiakos. Querying the data web: The MashQL approach. *Internet Computing, IEEE*, 14(3):58–67, May 2010
- Ruoming Jin, Yang Xiang, Ning Ruan, and David Fuhry. 3-HOP: A high-compression indexing scheme for reachability query. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, SIGMOD '09, pages 813–826, New York, NY, USA, 2009. ACM
- Kai Zeng, Jiacheng Yang, Haixun Wang, Bin Shao, and Zhongyuan Wang. A distributed graph engine for web scale RDF data. In *Proceedings of the 39th international conference on Very Large Data Bases*, PVLDB'13, pages 265–276. VLDB Endowment, 2013
- Andrey Kan, Jeffrey Chan, James Bailey, and Christopher Leckie. A query based approach for mining evolving graphs. In *Proceedings of the Eighth Australasian Data Mining Conference - Volume 101*, AusDM '09, pages 139–150, Darlinghurst, Australia, Australia, 2009. Australian Computer Society, Inc
- Anton Dries, Siegfried Nijssen, and Luc De Raedt. A query language for analyzing networks. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 485–494, New York, NY, USA, 2009. ACM
- Anton Dries and Siegfried Nijssen. Analyzing graph databases by aggregate queries. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, MLG '10, pages 37–45, New York, NY, USA, 2010. ACM
- Fan Wang and Gagan Agrawal. Answering complex structured queries over the deep web. In *Proceedings of the 15th Symposium on International Database Engineering & Applications*, IDEAS '11, pages 115–123, New York, NY, USA, 2011. ACM
- Renzo Angles, Arnau Prat-Pérez, David Dominguez-Sal, and Josep-Lluis Larriba-Pey. Benchmarking database systems for social network applications. In *First International Workshop on Graph Data Management Experiences and Systems*, GRADES '13, pages 15:1–15:7, New York, NY, USA, 2013. ACM
- Hao He, Haixun Wang, Jun Yang, and Philip S. Yu. BLINKS: Ranked keyword searches on graphs. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, SIGMOD '07, pages 305–316, New York, NY, USA, 2007. ACM
- Dritan Bleco and Yannis Kotidis. Business intelligence on complex graph data. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, EDBT-ICDT '12, pages 13–20, New York, NY, USA, 2012. ACM
- Alberto Tonon, Gianluca Demartini, and Philippe Cudré-Mauroux. Combining inverted indices and structured search for ad-hoc object retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 125–134, New York, NY, USA, 2012. ACM
- Dan Suciu. Distributed query evaluation on semistructured data. *ACM Trans. Database Syst.*, 27(1):1–62, March 2002
- Jayanta Mondal and Amol Deshpande. EAGr: Supporting continuous ego-centric aggregate queries over large dynamic graphs. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 1335–1346, New York, NY, USA, 2014. ACM
- Zhao Sun, Hongzhi Wang, Haixun Wang, Bin Shao, and Jianzhong Li. Efficient subgraph matching on billion node graphs. *Proc. VLDB Endow.*, 5(9):788–799, May 2012

- M. P. Consens and A. O. Mendelzon. Expressing structural hypertext queries in graphlog. In *Proceedings of the Second Annual ACM Conference on Hypertext*, HYPERTEXT '89, pages 269–292, New York, NY, USA, 1989. ACM
- Vitaliy Liptchinsky, Benjamin Satzger, Rostyslav Zabolotnyi, and Schahram Dustdar. Expressive languages for selecting groups from graph-structured data. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 761–770, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee
- Mirjana Mazuran, Edoardo Serra, and Carlo Zaniolo. Extending the power of datalog recursion. *The VLDB Journal*, 22(4):471–493, August 2013
- Wenlei Xie, Guozhang Wang, David Bindel, Alan Demers, and Johannes Gehrke. Fast iterative graph computation with block updates. *Proc. VLDB Endow.*, 6(14):2014–2025, September 2013
- Ramakrishna Varadarajan, Vagelis Hristidis, Louiqa Raschid, Maria-Esther Vidal, Luis Ibáñez, and Héctor Rodríguez-Drumond. Flexible and efficient querying and ranking on hyperlinked data sources. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 553–564, New York, NY, USA, 2009. ACM
- Federica Mandreoli, Riccardo Martoglia, Giorgio Villani, and Wilma Penzo. Flexible query answering on graph-modeled data. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 216–227, New York, NY, USA, 2009. ACM
- Semih Salihoglu and Jennifer Widom. GPS: A graph processing system. In *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*, SSDBM, pages 22:1–22:12, New York, NY, USA, 2013. ACM
- Hilmi Yildirim, Vineet Chaoji, and Mohammed J. Zaki. GRAIL: A scalable index for reachability queries in very large graphs. *The VLDB Journal*, 21(4):509–534, August 2012
- María Constanza Pabón, Claudia Roncancio, and Martha Millán. Graph data transformations and querying. In *Proceedings of the 2014 International C* Conference on Computer Science & Software Engineering*, C3S2E '14, pages 20:1–20:6, New York, NY, USA, 2014. ACM
- Lei Zou, M. Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gStore: A graph-based SPARQL query engine. *The VLDB Journal*, 23(4):565–590, August 2014
- Mohamed Sarwat, Sameh Elnikety, Yuxiong He, and Mohamed F. Mokbel. Horton+: A distributed system for processing declarative reachability queries over partitioned graphs. *Proc. VLDB Endow.*, 6(14):1918–1929, September 2013
- Wenfei Fan, Xin Wang, and Yinghui Wu. Incremental graph pattern matching. *ACM Trans. Database Syst.*, 38(3):18:1–18:47, September 2013
- Yosi Mass and Yehoshua Sagiv. Language models for keyword search over data graphs. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 363–372, New York, NY, USA, 2012. ACM
- Arijit Khan, Nan Li, Xifeng Yan, Ziyu Guan, Supriyo Chakraborty, and Shu Tao. Neighborhood based fast graph search in large networks. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, pages 901–912, New York, NY, USA, 2011. ACM

- Saumen Dey, Víctor Cuevas-Vicenttín, Sven Köhler, Eric Gribkoff, Michael Wang, and Bertram Ludäscher. On implementing provenance-aware regular path queries with relational query engines. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 214–223, New York, NY, USA, 2013. ACM
- Amarnath Gupta and Simone Santini. On querying OBO ontologies using a DAG pattern query language. In *Proceedings of the Third International Conference on Data Integration in the Life Sciences*, DILS'06, pages 152–167, Berlin, Heidelberg, 2006. Springer-Verlag
- Florian Holzschuher and René Peinl. Performance of graph query languages: Comparison of cypher, gremlin and native access in Neo4J. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 195–204, New York, NY, USA, 2013. ACM
- Yuanyuan Tian, Jignesh M. Patel, Viji Nair, Sebastian Martini, and Matthias Kretzler. Periscope/GQ: A graph querying toolkit. *Proc. VLDB Endow.*, 1(2):1404–1407, August 2008
- Peter T. Wood. Query languages for graph databases. *SIGMOD Rec.*, 41(1):50–60, April 2012
- Yerach Doytsher, Ben Galon, and Yaron Kanza. Querying geo-social data by bridging spatial networks and social networks. In *Proceedings of the 2Nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, LBSN '10, pages 39–46, New York, NY, USA, 2010. ACM
- Pablo Barceló Baeza. Querying graph databases. In *Proceedings of the 32nd Symposium on Principles of Database Systems*, PODS '13, pages 175–188, New York, NY, USA, 2013. ACM
- Juan P. Cedeño and K. Selçuk Candan. R2DF framework for ranked path queries over weighted RDF graphs. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 40:1–40:12, New York, NY, USA, 2011. ACM
- Jingren Zhou, Nicolas Bruno, Ming-Chuan Wu, Per-Ake Larson, Ronnie Chaiken, and Darren Shakib. SCOPE: Parallel databases meet MapReduce. *The VLDB Journal*, 21(5):611–636, October 2012
- Shengqi Yang, Yanan Xie, Yinghui Wu, Tianyu Wu, Huan Sun, Jian Wu, and Xifeng Yan. SLQ: A user-friendly graph querying system. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 893–896, New York, NY, USA, 2014. ACM
- Renzo Angles and Claudio Gutierrez. Survey of graph database models. *ACM Comput. Surv.*, 40(1):1:1–1:39, February 2008
- Ciro Cattuto, Marco Quaggiotto, André Panisson, and Alex Averbuch. Time-varying social networks in a graph database: A Neo4J use case. In *First International Workshop on Graph Data Management Experiences and Systems*, GRADES '13, pages 11:1–11:6, New York, NY, USA, 2013. ACM
- Padmashree Ravindra, HyeongSik Kim, and Kemafor Anyanwu. To nest or not to nest, when and how much: Representing intermediate results of graph pattern queries in mapreduce based processing. In *Proceedings of the 4th International Workshop on Semantic Web Information Management*, SWIM '12, pages 5:1–5:8, New York, NY, USA, 2012. ACM
- Alan Chappell, Sutanay Choudhury, John Feo, David Haglin, Alessandro Morari, Sumit Purohit, Karen Schuchardt, Antonino Tumeo, Jesse Weaver, and Oreste Villa. Toward a data scalable solution for facilitating discovery of scientific data resources. In *Proceedings of the 2013 International Workshop on Data-Intensive Scalable Computing Systems*, DISCS-2013, pages 55–60, New York, NY, USA, 2013. ACM

- Shengqi Yang, Xifeng Yan, Bo Zong, and Arijit Khan. Towards effective partition management for large graphs. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, pages 517–528, New York, NY, USA, 2012. ACM
- Michael Curtiss, Iain Becker, Tudor Bosman, Sergey Doroshenko, Lucian Grijincu, Tom Jackson, Sandhya Kunnatur, Soren Lassen, Philip Pronin, Sriram Sankar, Guanghao Shen, Gintaras Woss, Chao Yang, and Ning Zhang. Unicorn: A system for searching the social graph. *Proc. VLDB Endow.*, 6(11):1150–1161, August 2013
- Norbert Martinez-Bazan and David Dominguez-Sal. Using semijoin programs to solve traversal queries in graph databases. In *Proceedings of Workshop on GRAph Data Management Experiences and Systems*, GRADES'14, pages 6:1–6:6, New York, NY, USA, 2014. ACM
.

Referências dos artigos subcategoria Grandes grafos

- Sherif Sakr, Sameh Elnikety, and Yuxiong He. Hybrid query execution engine for large attributed graphs. *Information Systems*, 41(0):45 – 73, 2014
- Jesse Weaver, Vito Giovanni Castellana, Alessandro Morari, Antonino Tumeo, Sumit Purohit, Alan Chappell, David Haglin, Oreste Villa, Sutanay Choudhury, Karen Schuchardt, and John Feo. Toward a data scalable solution for facilitating discovery of science resources. *Parallel Computing*, 40(10):682 – 696, 2014
- Linhong Zhu, Wee Keong Ng, and James Cheng. Structure and attribute index for approximate graph matching in large graphs. *Information Systems*, 36(6):958 – 972, 2011
- Chang-Sup Park and Sungchae Lim. Efficient processing of keyword queries over graph databases for finding effective answers. *Information Processing & Management*, 51(1):42 – 57, 2015
- Renaud Delbru, Stephane Campinas, and Giovanni Tummarello. Searching web data: An entity retrieval and high-performance indexing model. *Web Semantics: Science, Services and Agents on the World Wide Web*, 10(0):33 – 58, 2012. Web-Scale Semantic Information Processing
- Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres, and Stefan Decker. Searching and browsing linked data with SWSE: The semantic web search engine. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4):365 – 401, 2011. {JWS} special issue on Semantic Search
- Aidan Hogan, Jürgen Umbrich, Andreas Harth, Richard Cyganiak, Axel Polleres, and Stefan Decker. An empirical survey of linked data conformance. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14(0):14 – 44, 2012. Special Issue on Dealing with the Messiness of the Web of Data
- Timos K. Sellis, Nick Roussopoulos, and Raymond T. Ng. Efficient compilation of large rule bases using logical access paths. *Information Systems*, 15(1):73 – 84, 1990. Knowledge Engineering
-
- Klara Weiand, Fabian KneiSSl, Wojciech Lobacz, Tim Furche, and François Bry. pest: Fast approximate keyword search in semantic data using eigenvector-based term propagation. *Information Systems*, 37(4):372 – 390, 2012. Semantic Web Data Management

- Paul Groth, Antonis Loizou, Alasdair J.G. Gray, Carole Goble, Lee Harland, and Steve Pettifer. Api-centric linked data integration: The open PHACTS discovery platform case study. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2014
- G. Karvounarakis, A. Magganaraki, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, and K. Tolle. Querying the semantic web with RQL. *Computer Networks*, 42(5):617 – 640, 2003. The Semantic Web: an evolution for a revolution
- Ying Ding. Community detection: Topological vs. topical. *Journal of Informetrics*, 5(4):498 – 514, 2011
- Hong-Han Shuai, De-Nian Yang, P.S. Yu, Chih-Ya Shen, and Ming-Syan Chen. On pattern preserving graph generation. In *IEEE 13th International Conference on Data Mining (ICDM)*, pages 677–686, Dec 2013
- R. Ronen and O. Shmueli. SoQL: A language for querying and creating data in social networks. In *ICDE'09. IEEE 25th International Conference on Data Engineering*, pages 1595–1602, March 2009
- Yijun Bei, Zhen Lin, Chen Zhao, and Xiaojun Zhu. HBase system-based distributed framework for searching large graph databases. In *14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 151–156, July 2013
- J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 51–60, Oct 2011
- A. Naseer, L. Laera, and T. Matsutsuka. Enterprise biggraph. In *46th Hawaii International Conference on System Sciences (HICSS)*, pages 1005–1014, Jan 2013
- V.G. Castellana, A. Tumeo, O. Villa, D. Haglin, and J. Feo. Composing data parallel code for a SPARQL graph engine. In *International Conference on Social Computing (SocialCom)*, pages 691–699, Sept 2013
- M. Brocheler, A. Pugliese, and V.S. Subrahmanian. COSI: Cloud oriented subgraph identification in massive social networks. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 248–255, Aug 2010
- Rui Wang and K. Chiu. A stream partitioning approach to processing large scale distributed graph datasets. In *IEEE International Conference on Big Data*, pages 537–542, Oct 2013
- S.R. Spillane, J. Birnbaum, D. Bokser, D. Kemp, A. Labouseur, P.W. Olsen, J. Vijayan, Jeong-Hyon Hwang, and Jun-Weon Yoon. A demonstration of the G* graph database system. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 1356–1359, April 2013
- Jun Gao, Jiashuai Zhou, Chang Zhou, and J.X. Yu. GLog: A high level graph analysis system using mapreduce. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 544–555, March 2014
- Ying Tu and Han-Wei Shen. GraphCharter: Combining browsing with query to explore large semantic graphs. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pages 49–56, Feb 2013
- P. Yadav and V. Samala. Benchmarking over a semantic repository. In *Second International Conference on Advanced Computing (ICoAC)*, pages 51–59, Dec 2010

- V. Dritsou, P. Constantopoulos, A. Deligiannakis, and Y. Kotidis. Shortcut selection in RDF databases. In *IEEE 27th International Conference on Data Engineering Workshops (ICDEW)*, pages 194–199, April 2011
- Letao Qi, H.T. Lin, and V. Honavar. Clustering remote RDF data using SPARQL update queries. In *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 236–242, April 2013
- G. Kasneci, M. Ramanath, M. Sozio, F.M. Suchanek, and G. Weikum. STAR: Steiner-tree approximation in relationship graphs. In *IEEE 25th International Conference on Data Engineering - ICDE'09*, pages 868–879, March 2009
- M. Jarrar and M.D. Dikaiakos. Querying the data web: The MashQL approach. *Internet Computing, IEEE*, 14(3):58–67, May 2010
- Ruoming Jin, Yang Xiang, Ning Ruan, and David Fuhry. 3-HOP: A high-compression indexing scheme for reachability query. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, SIGMOD '09, pages 813–826, New York, NY, USA, 2009. ACM
- Andrey Kan, Jeffrey Chan, James Bailey, and Christopher Leckie. A query based approach for mining evolving graphs. In *Proceedings of the Eighth Australasian Data Mining Conference - Volume 101*, AusDM '09, pages 139–150, Darlinghurst, Australia, Australia, 2009. Australian Computer Society, Inc
- Anton Dries, Siegfried Nijssen, and Luc De Raedt. A query language for analyzing networks. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 485–494, New York, NY, USA, 2009. ACM
- Dritan Bleco and Yannis Kotidis. Business intelligence on complex graph data. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, EDBT-ICDT '12, pages 13–20, New York, NY, USA, 2012. ACM
- Alberto Tonon, Gianluca Demartini, and Philippe Cudré-Mauroux. Combining inverted indices and structured search for ad-hoc object retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 125–134, New York, NY, USA, 2012. ACM
- Jayanta Mondal and Amol Deshpande. EAGr: Supporting continuous ego-centric aggregate queries over large dynamic graphs. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 1335–1346, New York, NY, USA, 2014. ACM
- Zhao Sun, Hongzhi Wang, Haixun Wang, Bin Shao, and Jianzhong Li. Efficient subgraph matching on billion node graphs. *Proc. VLDB Endow.*, 5(9):788–799, May 2012
- Wenlei Xie, Guozhang Wang, David Bindel, Alan Demers, and Johannes Gehrke. Fast iterative graph computation with block updates. *Proc. VLDB Endow.*, 6(14):2014–2025, September 2013
- Semih Salihoglu and Jennifer Widom. GPS: A graph processing system. In *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*, SSDBM, pages 22:1–22:12, New York, NY, USA, 2013. ACM
- Hilmi Yildirim, Vineet Chaoji, and Mohammed J. Zaki. GRAIL: A scalable index for reachability queries in very large graphs. *The VLDB Journal*, 21(4):509–534, August 2012
- Lei Zou, M. Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gStore: A graph-based SPARQL query engine. *The VLDB Journal*, 23(4):565–590, August 2014

- Mohamed Sarwat, Sameh Elnikety, Yuxiong He, and Mohamed F. Mokbel. Horton+: A distributed system for processing declarative reachability queries over partitioned graphs. *Proc. VLDB Endow.*, 6(14):1918–1929, September 2013
- Wenfei Fan, Xin Wang, and Yinghui Wu. Incremental graph pattern matching. *ACM Trans. Database Syst.*, 38(3):18:1–18:47, September 2013
- Yuanyuan Tian, Jignesh M. Patel, Viji Nair, Sebastian Martini, and Matthias Kretzler. Periscope/GQ: A graph querying toolkit. *Proc. VLDB Endow.*, 1(2):1404–1407, August 2008
- Shengqi Yang, Yanan Xie, Yinghui Wu, Tianyu Wu, Huan Sun, Jian Wu, and Xifeng Yan. SLQ: A user-friendly graph querying system. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 893–896, New York, NY, USA, 2014. ACM
- Alan Chappell, Sutanay Choudhury, John Feo, David Haglin, Alessandro Morari, Sumit Purohit, Karen Schuchardt, Antonino Tumeo, Jesse Weaver, and Oreste Villa. Toward a data scalable solution for facilitating discovery of scientific data resources. In *Proceedings of the 2013 International Workshop on Data-Intensive Scalable Computing Systems*, DISCS-2013, pages 55–60, New York, NY, USA, 2013. ACM
- Shengqi Yang, Xifeng Yan, Bo Zong, and Arijit Khan. Towards effective partition management for large graphs. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, pages 517–528, New York, NY, USA, 2012. ACM
- Michael Curtiss, Iain Becker, Tudor Bosman, Sergey Doroshenko, Lucian Grijincu, Tom Jackson, Sandhya Kunnatur, Soren Lassen, Philip Pronin, Sriram Sankar, Guanghao Shen, Gintaras Woss, Chao Yang, and Ning Zhang. Unicorn: A system for searching the social graph. *Proc. VLDB Endow.*, 6(11):1150–1161, August 2013
.

Referências dos artigos subcategoria Funções de agregação

- A. González-Beltrán, P. Milligan, and P. Sage. Range queries over skip tree graphs. *Computer Communications*, 31(2):358 – 374, 2008. Special Issue: Foundation of Peer-to-Peer Computing
- Michel Mainguenaud. Constraint-based queries in a geographical database for network facilities. *Computers, Environment and Urban Systems*, 20(2):139 – 151, 1996
- J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 51–60, Oct 2011
- Jiwon Seo, S. Guo, and M.S. Lam. SociaLite: Datalog extensions for efficient social network analysis. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 278–289, April 2013
- S. Dar and R. Agrawal. Extending SQL with generalized transitive closure. *IEEE Transactions on Knowledge and Data Engineering*, 5(5):799–812, Oct 1993
- Anton Dries and Siegfried Nijssen. Analyzing graph databases by aggregate queries. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, MLG '10, pages 37–45, New York, NY, USA, 2010. ACM
- Fan Wang and Gagan Agrawal. Answering complex structured queries over the deep web. In *Proceedings of the 15th Symposium on International Database Engineering & Applications*, IDEAS '11, pages 115–123, New York, NY, USA, 2011. ACM

- Dritan Bleco and Yannis Kotidis. Business intelligence on complex graph data. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, EDBT-ICDT '12, pages 13–20, New York, NY, USA, 2012. ACM
- Jayanta Mondal and Amol Deshpande. EAGr: Supporting continuous ego-centric aggregate queries over large dynamic graphs. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 1335–1346, New York, NY, USA, 2014. ACM
- M. P. Consens and A. O. Mendelzon. Expressing structural hypertext queries in graphlog. In *Proceedings of the Second Annual ACM Conference on Hypertext*, HYPERTEXT '89, pages 269–292, New York, NY, USA, 1989. ACM
- Lei Zou, M. Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gStore: A graph-based SPARQL query engine. *The VLDB Journal*, 23(4):565–590, August 2014
.

Referências dos artigos subcategoria Palavras-chave

- Chang-Sup Park and Sungchae Lim. Efficient processing of keyword queries over graph databases for finding effective answers. *Information Processing & Management*, 51(1):42 – 57, 2015
- Renaud Delbru, Stephane Campinas, and Giovanni Tummarello. Searching web data: An entity retrieval and high-performance indexing model. *Web Semantics: Science, Services and Agents on the World Wide Web*, 10(0):33 – 58, 2012. Web-Scale Semantic Information Processing
- Thanh Tran, Haofen Wang, and Peter Haase. Hermes: Data web search on a pay-as-you-go integration infrastructure. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):189 – 203, 2009. The Web of Data
- Benny Kimelfeld and Yehoshua Sagiv. Efficiently enumerating results of keyword search over data graphs. *Information Systems*, 33(45):335 – 359, 2008. Selected Papers from the Tenth International Symposium on Database Programming Languages (DBPL 2005
- Xiang Lian, Eugenio De Hoyos, Artem Chebotko, Bin Fu, and Christine Reilly. k-nearest keyword search in RDF graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, pages 40 – 56, 2013
- M. Uddin, R. Stadler, M. Miyazawa, and M. Hayashi. Graph search for cloud network management. In *Network Operations and Management Symposium (NOMS), 2014 IEEE*, pages 1–5, May 2014
- Yi Chen, Wei Wang, and Ziyang Liu. Keyword-based search and exploration on databases. In *IEEE 27th International Conference on Data Engineering (ICDE)*, pages 1380–1383, April 2011
- Hao He, Haixun Wang, Jun Yang, and Philip S. Yu. BLINKS: Ranked keyword searches on graphs. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, SIGMOD '07, pages 305–316, New York, NY, USA, 2007. ACM
- Alberto Tonon, Gianluca Demartini, and Philippe Cudré-Mauroux. Combining inverted indices and structured search for ad-hoc object retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 125–134, New York, NY, USA, 2012. ACM

- Yosi Mass and Yehoshua Sagiv. Language models for keyword search over data graphs. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 363–372, New York, NY, USA, 2012. ACM
- Shengqi Yang, Yanan Xie, Yinghui Wu, Tianyu Wu, Huan Sun, Jian Wu, and Xifeng Yan. SLQ: A user-friendly graph querying system. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 893–896, New York, NY, USA, 2014. ACM
.

Referências dos artigos subcategoria Top-k

- Chang-Sup Park and Sungchae Lim. Efficient processing of keyword queries over graph databases for finding effective answers. *Information Processing & Management*, 51(1):42 – 57, 2015
- A. Cuzzocrea and P. Serafino. A reachability-based theoretical framework for modeling and querying complex probabilistic graph data. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1177–1184, Oct 2012
- V. Dritsou, P. Constantopoulos, A. Deligiannakis, and Y. Kotidis. Shortcut selection in RDF databases. In *IEEE 27th International Conference on Data Engineering Workshops (ICDEW)*, pages 194–199, April 2011
- G. Kasneci, M. Ramanath, M. Sozio, F.M. Suchanek, and G. Weikum. STAR: Steiner-tree approximation in relationship graphs. In *IEEE 25th International Conference on Data Engineering - ICDE'09*, pages 868–879, March 2009
- Anton Dries and Siegfried Nijssen. Analyzing graph databases by aggregate queries. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, MLG '10, pages 37–45, New York, NY, USA, 2010. ACM
- Hao He, Haixun Wang, Jun Yang, and Philip S. Yu. BLINKS: Ranked keyword searches on graphs. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, SIGMOD '07, pages 305–316, New York, NY, USA, 2007. ACM
- Ramakrishna Varadarajan, Vagelis Hristidis, Louiqa Raschid, Maria-Esther Vidal, Luis Ibáñez, and Héctor Rodríguez-Drumond. Flexible and efficient querying and ranking on hyperlinked data sources. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 553–564, New York, NY, USA, 2009. ACM
- Federica Mandreoli, Riccardo Martoglia, Giorgio Villani, and Wilma Penzo. Flexible query answering on graph-modeled data. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 216–227, New York, NY, USA, 2009. ACM
- Arijit Khan, Nan Li, Xifeng Yan, Ziyu Guan, Supriyo Chakraborty, and Shu Tao. Neighborhood based fast graph search in large networks. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, pages 901–912, New York, NY, USA, 2011. ACM
- Juan P. Cedeño and K. Selçuk Candan. R2DF framework for ranked path queries over weighted RDF graphs. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 40:1–40:12, New York, NY, USA, 2011. ACM

.

Referências dos artigos subcategoria Integração de dados

- Jesse Weaver, Vito Giovanni Castellana, Alessandro Morari, Antonino Tumeo, Sumit Purohit, Alan Chappell, David Haglin, Oreste Villa, Sutanay Choudhury, Karen Schuchardt, and John Feo. Toward a data scalable solution for facilitating discovery of science resources. *Parallel Computing*, 40(10):682 – 696, 2014
- Renaud Delbru, Stephane Campinas, and Giovanni Tummarello. Searching web data: An entity retrieval and high-performance indexing model. *Web Semantics: Science, Services and Agents on the World Wide Web*, 10(0):33 – 58, 2012. Web-Scale Semantic Information Processing
- Thanh Tran, Haofen Wang, and Peter Haase. Hermes: Data web search on a pay-as-you-go integration infrastructure. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):189 – 203, 2009. The Web of Data
- Paul Groth, Antonis Loizou, Alasdair J.G. Gray, Carole Goble, Lee Harland, and Steve Pettifer. Api-centric linked data integration: The open PHACTS discovery platform case study. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2014
- G. Karvounarakis, A. Magganaraki, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, and K. Tolle. Querying the semantic web with RQL. *Computer Networks*, 42(5):617 – 640, 2003. The Semantic Web: an evolution for a revolution
- A. Petermann, M. Junghanns, R. Muller, and E. Rahm. BIIIG: Enabling business intelligence with integrated instance graphs. In *IEEE 30th International Conference on Data Engineering Workshops (ICDEW)*, pages 4–11, March 2014
- J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 51–60, Oct 2011
- W.E. Moustafa, A. Kimmig, A. Deshpande, and L. Getoor. Subgraph pattern matching over uncertain graphs with identity linkage uncertainty. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 904–915, March 2014
- Yunkai Liu and T.M. Vitolo. Graph data warehouse: Steps to integrating graph databases into the traditional conceptual structure of a data warehouse. In *IEEE International Congress on Big Data (BigData Congress)*, pages 433–434, June 2013
- Federica Mandreoli, Riccardo Martoglia, Giorgio Villani, and Wilma Penzo. Flexible query answering on graph-modeled data. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 216–227, New York, NY, USA, 2009. ACM
- María Constanza Pabón, Claudia Roncancio, and Martha Millán. Graph data transformations and querying. In *Proceedings of the 2014 International C* Conference on Computer Science & Software Engineering*, C3S2E '14, pages 20:1–20:6, New York, NY, USA, 2014. ACM

.

Referências dos artigos subcategoria Otimização de desempenho

- Jiwon Seo, S. Guo, and M.S. Lam. SociaLite: Datalog extensions for efficient social network analysis. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 278–289, April 2013

- Jun Gao, Jiashuai Zhou, Chang Zhou, and J.X. Yu. GLog: A high level graph analysis system using mapreduce. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 544–555, March 2014 - V. Dritsou, P. Constantopoulos, A. Deligiannakis, and Y. Kotidis. Shortcut selection in RDF databases. In *IEEE 27th International Conference on Data Engineering Workshops (ICDEW)*, pages 194–199, April 2011 - Letao Qi, H.T. Lin, and V. Honavar. Clustering remote RDF data using SPARQL update queries. In *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 236–242, April 2013 - G. Kasneci, M. Ramanath, M. Sozio, F.M. Suchanek, and G. Weikum. STAR: Steiner-tree approximation in relationship graphs. In *IEEE 25th International Conference on Data Engineering - ICDE'09*, pages 868–879, March 2009 - M. Jarrar and M.D. Dikaiakos. Querying the data web: The MashQL approach. *Internet Computing, IEEE*, 14(3):58–67, May 2010 - Fan Wang and Gagan Agrawal. Answering complex structured queries over the deep web. In *Proceedings of the 15th Symposium on International Database Engineering & Applications*, IDEAS '11, pages 115–123, New York, NY, USA, 2011. ACM - Mirjana Mazuran, Edoardo Serra, and Carlo Zaniolo. Extending the power of datalog recursion. *The VLDB Journal*, 22(4):471–493, August 2013 - Semih Salihoglu and Jennifer Widom. GPS: A graph processing system. In *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*, SSDBM, pages 22:1–22:12, New York, NY, USA, 2013. ACM - Mohamed Sarwat, Sameh Elnikety, Yuxiong He, and Mohamed F. Mokbel. Horton+: A distributed system for processing declarative reachability queries over partitioned graphs. *Proc. VLDB Endow.*, 6(14):1918–1929, September 2013 - Juan P. Cedeño and K. Selçuk Candan. R2DF framework for ranked path queries over weighted RDF graphs. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 40:1–40:12, New York, NY, USA, 2011. ACM - Jingren Zhou, Nicolas Bruno, Ming-Chuan Wu, Per-Ake Larson, Ronnie Chaiken, and Darren Shakib. SCOPE: Parallel databases meet MapReduce. *The VLDB Journal*, 21(5):611–636, October 2012 .

Tabela 6: Categoria: paradigma tecnológico.

Referências dos artigos subcategoria RDF
- Jesse Weaver, Vito Giovanni Castellana, Alessandro Morari, Antonino Tumeo, Sumit Purohit, Alan Chappell, David Haglin, Oreste Villa, Sutanay Choudhury, Karen Schuchardt, and John Feo. Toward a data scalable solution for facilitating discovery of science resources. *Parallel Computing*, 40(10):682 – 696, 2014 - Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres, and Stefan Decker. Searching and browsing linked data with SWSE: The semantic web search engine. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4):365 – 401, 2011. {JWS} special issue on Semantic Search

- Aidan Hogan, Jürgen Umbrich, Andreas Harth, Richard Cyganiak, Axel Polleres, and Stefan Decker. An empirical survey of linked data conformance. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14(0):14 – 44, 2012. Special Issue on Dealing with the Messiness of the Web of Data
- Klara Weiand, Fabian KneiSSl, Wojciech Lobacz, Tim Furche, and François Bry. pest: Fast approximate keyword search in semantic data using eigenvector-based term propagation. *Information Systems*, 37(4):372 – 390, 2012. Semantic Web Data Management
- Xiang Lian, Eugenio De Hoyos, Artem Chebotko, Bin Fu, and Christine Reilly. k-nearest keyword search in RDF graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, pages 40 – 56, 2013
- Paul Groth, Antonis Loizou, Alasdair J.G. Gray, Carole Goble, Lee Harland, and Steve Pettifer. Api-centric linked data integration: The open PHACTS discovery platform case study. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2014
- G. Karvounarakis, A. Magganaraki, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, and K. Tolle. Querying the semantic web with RQL. *Computer Networks*, 42(5):617 – 640, 2003. The Semantic Web: an evolution for a revolution
- A. Morari, V.G. Castellana, D. Haglin, J. Feo, J. Weaver, A. Tumeo, and O. Villa. Accelerating semantic graph databases on commodity clusters. In *IEEE International Conference on Big Data*, pages 768–772, Oct 2013
- V.G. Castellana, A. Tumeo, O. Villa, D. Haglin, and J. Feo. Composing data parallel code for a SPARQL graph engine. In *International Conference on Social Computing (SocialCom)*, pages 691–699, Sept 2013
- A. Cuzzocrea and P. Serafino. A reachability-based theoretical framework for modeling and querying complex probabilistic graph data. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1177–1184, Oct 2012
- Rui Wang and K. Chiu. A stream partitioning approach to processing large scale distributed graph datasets. In *IEEE International Conference on Big Data*, pages 537–542, Oct 2013
- O. Cure, F. Kerdjoudj, Chan Le Duc, M. Lamolle, and D. Faye. On the potential integration of an ontology-based data access approach in NoSQL stores. In *Third International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)*, pages 166–173, Sept 2012
- M. Leida and A. Chu. Distributed SPARQL query answering over RDF data streams. In *IEEE International Congress on Big Data (BigData Congress)*, pages 369–378, June 2013
- P. Yadav and V. Samala. Benchmarking over a semantic repository. In *Second International Conference on Advanced Computing (ICoAC)*, pages 51–59, Dec 2010
- V. Dritsou, P. Constantopoulos, A. Deligiannakis, and Y. Kotidis. Shortcut selection in RDF databases. In *IEEE 27th International Conference on Data Engineering Workshops (ICDEW)*, pages 194–199, April 2011
- Letao Qi, H.T. Lin, and V. Honavar. Clustering remote RDF data using SPARQL update queries. In *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 236–242, April 2013
- M. Jarrar and M.D. Dikaiakos. Querying the data web: The MashQL approach. *Internet Computing, IEEE*, 14(3):58–67, May 2010

- Kai Zeng, Jiacheng Yang, Haixun Wang, Bin Shao, and Zhongyuan Wang. A distributed graph engine for web scale RDF data. In *Proceedings of the 39th international conference on Very Large Data Bases*, PVLDB'13, pages 265–276. VLDB Endowment, 2013
- Alberto Tonon, Gianluca Demartini, and Philippe Cudré-Mauroux. Combining inverted indices and structured search for ad-hoc object retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 125–134, New York, NY, USA, 2012. ACM
- Lei Zou, M. Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gStore: A graph-based SPARQL query engine. *The VLDB Journal*, 23(4):565–590, August 2014
- Amarnath Gupta and Simone Santini. On querying OBO ontologies using a DAG pattern query language. In *Proceedings of the Third International Conference on Data Integration in the Life Sciences*, DILS'06, pages 152–167, Berlin, Heidelberg, 2006. Springer-Verlag
- Juan P. Cedeño and K. Selçuk Candan. R2DF framework for ranked path queries over weighted RDF graphs. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 40:1–40:12, New York, NY, USA, 2011. ACM
- Padmashree Ravindra, HyeongSik Kim, and Kemafor Anyanwu. To nest or not to nest, when and how much: Representing intermediate results of graph pattern queries in mapreduce based processing. In *Proceedings of the 4th International Workshop on Semantic Web Information Management*, SWIM '12, pages 5:1–5:8, New York, NY, USA, 2012. ACM
.

Referências dos artigos subcategoria Paralelismo

- Jesse Weaver, Vito Giovanni Castellana, Alessandro Morari, Antonino Tumeo, Sumit Purohit, Alan Chappell, David Haglin, Oreste Villa, Sutanay Choudhury, Karen Schuchardt, and John Feo. Toward a data scalable solution for facilitating discovery of science resources. *Parallel Computing*, 40(10):682 – 696, 2014
- A. Morari, V.G. Castellana, D. Haglin, J. Feo, J. Weaver, A. Tumeo, and O. Villa. Accelerating semantic graph databases on commodity clusters. In *IEEE International Conference on Big Data*, pages 768–772, Oct 2013
- V.G. Castellana, A. Tumeo, O. Villa, D. Haglin, and J. Feo. Composing data parallel code for a SPARQL graph engine. In *International Conference on Social Computing (SocialCom)*, pages 691–699, Sept 2013
- Jun Gao, Jiashuai Zhou, Chang Zhou, and J.X. Yu. GLog: A high level graph analysis system using mapreduce. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 544–555, March 2014
- L. Fegaras. Supporting Bulk Synchronous Parallelism in Map-Reduce queries. In *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:*, pages 1068–1077, Nov 2012
- Letao Qi, H.T. Lin, and V. Honavar. Clustering remote RDF data using SPARQL update queries. In *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 236–242, April 2013
- Zhao Sun, Hongzhi Wang, Haixun Wang, Bin Shao, and Jianzhong Li. Efficient subgraph matching on billion node graphs. *Proc. VLDB Endow.*, 5(9):788–799, May 2012

- Mohamed Sarwat, Sameh Elnikety, Yuxiong He, and Mohamed F. Mokbel. Horton+: A distributed system for processing declarative reachability queries over partitioned graphs. *Proc. VLDB Endow.*, 6(14):1918–1929, September 2013
- Jingren Zhou, Nicolas Bruno, Ming-Chuan Wu, Per-Ake Larson, Ronnie Chaiken, and Darren Shakib. SCOPE: Parallel databases meet MapReduce. *The VLDB Journal*, 21(5):611–636, October 2012
- Padmashree Ravindra, HyeongSik Kim, and Kemafor Anyanwu. To nest or not to nest, when and how much: Representing intermediate results of graph pattern queries in mapreduce based processing. In *Proceedings of the 4th International Workshop on Semantic Web Information Management*, SWIM '12, pages 5:1–5:8, New York, NY, USA, 2012. ACM
- Alan Chappell, Sutanay Choudhury, John Feo, David Haglin, Alessandro Morari, Sumit Purohit, Karen Schuchardt, Antonino Tumeo, Jesse Weaver, and Oreste Villa. Toward a data scalable solution for facilitating discovery of scientific data resources. In *Proceedings of the 2013 International Workshop on Data-Intensive Scalable Computing Systems*, DISCS-2013, pages 55–60, New York, NY, USA, 2013. ACM
- Shengqi Yang, Xifeng Yan, Bo Zong, and Arijit Khan. Towards effective partition management for large graphs. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, pages 517–528, New York, NY, USA, 2012. ACM
.

Referências dos artigos subcategoria Distribuição de dados

- Yijun Bei, Zhen Lin, Chen Zhao, and Xiaojun Zhu. HBase system-based distributed framework for searching large graph databases. In *14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 151–156, July 2013
- M. Brocheler, A. Pugliese, and V.S. Subrahmanian. COSI: Cloud oriented subgraph identification in massive social networks. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 248–255, Aug 2010
- Rui Wang and K. Chiu. A stream partitioning approach to processing large scale distributed graph datasets. In *IEEE International Conference on Big Data*, pages 537–542, Oct 2013
- S.R. Spillane, J. Birnbaum, D. Bokser, D. Kemp, A. Labouseur, P.W. Olsen, J. Vijayan, Jeong-Hyon Hwang, and Jun-Weon Yoon. A demonstration of the G* graph database system. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 1356–1359, April 2013
- M. Leida and A. Chu. Distributed SPARQL query answering over RDF data streams. In *IEEE International Congress on Big Data (BigData Congress)*, pages 369–378, June 2013
- Kai Zeng, Jiacheng Yang, Haixun Wang, Bin Shao, and Zhongyuan Wang. A distributed graph engine for web scale RDF data. In *Proceedings of the 39th international conference on Very Large Data Bases*, PVLDB'13, pages 265–276. VLDB Endowment, 2013
- Dan Suciu. Distributed query evaluation on semistructured data. *ACM Trans. Database Syst.*, 27(1):1–62, March 2002
- Semih Salihoglu and Jennifer Widom. GPS: A graph processing system. In *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*, SSDBM, pages 22:1–22:12, New York, NY, USA, 2013. ACM
.

Referências dos artigos subcategoria Linguagem declarativa
- Sherif Sakr, Sameh Elnikety, and Yuxiong He. Hybrid query execution engine for large attributed graphs. *Information Systems*, 41(0):45 – 73, 2014
- Jesse Weaver, Vito Giovanni Castellana, Alessandro Morari, Antonino Tumeo, Sumit Purohit, Alan Chappell, David Haglin, Oreste Villa, Sutanay Choudhury, Karen Schuchardt, and John Feo. Toward a data scalable solution for facilitating discovery of science resources. *Parallel Computing*, 40(10):682 – 696, 2014
- Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres, and Stefan Decker. Searching and browsing linked data with SWSE: The semantic web search engine. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4):365 – 401, 2011. {JWS} special issue on Semantic Search
- Aidan Hogan, Jürgen Umbrich, Andreas Harth, Richard Cyganiak, Axel Polleres, and Stefan Decker. An empirical survey of linked data conformance. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14(0):14 – 44, 2012. Special Issue on Dealing with the Messiness of the Web of Data
- Timos K. Sellis, Nick Roussopoulos, and Raymond T. Ng. Efficient compilation of large rule bases using logical access paths. *Information Systems*, 15(1):73 – 84, 1990. Knowledge Engineering
- Klara Weiand, Fabian KneiSSl, Wojciech Lobacz, Tim Furche, and François Bry. pest: Fast approximate keyword search in semantic data using eigenvector-based term propagation. *Information Systems*, 37(4):372 – 390, 2012. Semantic Web Data Management
- Xiang Lian, Eugenio De Hoyos, Artem Chebotko, Bin Fu, and Christine Reilly. k-nearest keyword search in RDF graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, pages 40 – 56, 2013
- Paul Groth, Antonis Loizou, Alasdair J.G. Gray, Carole Goble, Lee Harland, and Steve Pettifer. Api-centric linked data integration: The open PHACTS discovery platform case study. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2014
- G. Karvounarakis, A. Magganaraki, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, and K. Tolle. Querying the semantic web with RQL. *Computer Networks*, 42(5):617 – 640, 2003. The Semantic Web: an evolution for a revolution
- H.M. Jamil. Design of declarative graph query languages: On the choice between value, pattern and object based representations for graphs. In *IEEE 28th International Conference on Data Engineering Workshops (ICDEW)*, pages 178–185, April 2012
- R. Ronen and O. Shmueli. SoQL: A language for querying and creating data in social networks. In *ICDE'09. IEEE 25th International Conference on Data Engineering*, pages 1595–1602, March 2009
- A. Morari, V.G. Castellana, D. Haglin, J. Feo, J. Weaver, A. Tumeo, and O. Villa. Accelerating semantic graph databases on commodity clusters. In *IEEE International Conference on Big Data*, pages 768–772, Oct 2013
- J. Heer and A. Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 51–60, Oct 2011

- V.G. Castellana, A. Tumeo, O. Villa, D. Haglin, and J. Feo. Composing data parallel code for a SPARQL graph engine. In *International Conference on Social Computing (SocialCom)*, pages 691–699, Sept 2013
- Jiwon Seo, S. Guo, and M.S. Lam. SociaLite: Datalog extensions for efficient social network analysis. In *IEEE 29th International Conference on Data Engineering (ICDE)*, pages 278–289, April 2013
- A. Cuzzocrea and P. Serafino. A reachability-based theoretical framework for modeling and querying complex probabilistic graph data. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1177–1184, Oct 2012
- Rui Wang and K. Chiu. A stream partitioning approach to processing large scale distributed graph datasets. In *IEEE International Conference on Big Data*, pages 537–542, Oct 2013
- C. Raguenaud, J. Kennedy, and P.J. Barclay. The Prometheus taxonomic database. In *Proceedings IEEE International Symposium on Bio-Informatics and Biomedical Engineering*, pages 63–70, 2000
- Jun Gao, Jiashuai Zhou, Chang Zhou, and J.X. Yu. GLog: A high level graph analysis system using mapreduce. In *IEEE 30th International Conference on Data Engineering (ICDE)*, pages 544–555, March 2014
- O. Cure, F. Kerdjoudj, Chan Le Duc, M. Lamolle, and D. Faye. On the potential integration of an ontology-based data access approach in NoSQL stores. In *Third International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)*, pages 166–173, Sept 2012
- M. Leida and A. Chu. Distributed SPARQL query answering over RDF data streams. In *IEEE International Congress on Big Data (BigData Congress)*, pages 369–378, June 2013
- P. Yadav and V. Samala. Benchmarking over a semantic repository. In *Second International Conference on Advanced Computing (ICoAC)*, pages 51–59, Dec 2010
- Taekyong Lee, Lei Sheng, T. Bozkaya, N.H. Balkir, Z. Meral Ozsoyoglu, and Z.M. Ozsoyoglu. Querying multimedia presentations based on content. *IEEE Transactions on Knowledge and Data Engineering*, 11(3):361–385, May 1999
- S. Dar and R. Agrawal. Extending SQL with generalized transitive closure. *IEEE Transactions on Knowledge and Data Engineering*, 5(5):799–812, Oct 1993
- V. Dritsou, P. Constantopoulos, A. Deligiannakis, and Y. Kotidis. Shortcut selection in RDF databases. In *IEEE 27th International Conference on Data Engineering Workshops (ICDEW)*, pages 194–199, April 2011
- Letao Qi, H.T. Lin, and V. Honavar. Clustering remote RDF data using SPARQL update queries. In *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 236–242, April 2013
- M. Jarrar and M.D. Dikaiakos. Querying the data web: The MashQL approach. *Internet Computing, IEEE*, 14(3):58–67, May 2010
- Kai Zeng, Jiacheng Yang, Haixun Wang, Bin Shao, and Zhongyuan Wang. A distributed graph engine for web scale RDF data. In *Proceedings of the 39th international conference on Very Large Data Bases*, PVLDB'13, pages 265–276. VLDB Endowment, 2013
- Fan Wang and Gagan Agrawal. Answering complex structured queries over the deep web. In *Proceedings of the 15th Symposium on International Database Engineering & Applications*, IDEAS '11, pages 115–123, New York, NY, USA, 2011. ACM

- Alberto Tonon, Gianluca Demartini, and Philippe Cudré-Mauroux. Combining inverted indices and structured search for ad-hoc object retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 125–134, New York, NY, USA, 2012. ACM
- M. P. Consens and A. O. Mendelzon. Expressing structural hypertext queries in graphlog. In *Proceedings of the Second Annual ACM Conference on Hypertext*, HYPERTEXT '89, pages 269–292, New York, NY, USA, 1989. ACM
- Mirjana Mazuran, Edoardo Serra, and Carlo Zaniolo. Extending the power of datalog recursion. *The VLDB Journal*, 22(4):471–493, August 2013
- Lei Zou, M. Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gStore: A graph-based SPARQL query engine. *The VLDB Journal*, 23(4):565–590, August 2014
- Mohamed Sarwat, Sameh Elnikety, Yuxiong He, and Mohamed F. Mokbel. Horton+: A distributed system for processing declarative reachability queries over partitioned graphs. *Proc. VLDB Endow.*, 6(14):1918–1929, September 2013
- Saumen Dey, Víctor Cuevas-Vicenttín, Sven Köhler, Eric Gribkoff, Michael Wang, and Bertram Ludäscher. On implementing provenance-aware regular path queries with relational query engines. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 214–223, New York, NY, USA, 2013. ACM
- Amarnath Gupta and Simone Santini. On querying OBO ontologies using a DAG pattern query language. In *Proceedings of the Third International Conference on Data Integration in the Life Sciences*, DILS'06, pages 152–167, Berlin, Heidelberg, 2006. Springer-Verlag
- Juan P. Cedeño and K. Selçuk Candan. R2DF framework for ranked path queries over weighted RDF graphs. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 40:1–40:12, New York, NY, USA, 2011. ACM
- Jingren Zhou, Nicolas Bruno, Ming-Chuan Wu, Per-Ake Larson, Ronnie Chaiken, and Darren Shakib. SCOPE: Parallel databases meet MapReduce. *The VLDB Journal*, 21(5):611–636, October 2012
- Padmashree Ravindra, HyeongSik Kim, and Kemafor Anyanwu. To nest or not to nest, when and how much: Representing intermediate results of graph pattern queries in mapreduce based processing. In *Proceedings of the 4th International Workshop on Semantic Web Information Management*, SWIM '12, pages 5:1–5:8, New York, NY, USA, 2012. ACM