

**Uma Abordagem Baseada em Ontologias para a Predição de  
Ligações entre Pesquisadores em Redes de Colaboração Científica**

Thiago Henrique Dias Araujo

TEXTO APRESENTADO  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
O EXAME DE QUALIFICAÇÃO  
DE  
MESTRE EM CIÊNCIAS

Programa: Mestrado em Ciência da Computação

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Renata Wassermann

São Paulo, março de 2016

# **Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica**

Esta é a versão original do texto elaborado pelo candidato Thiago Henrique Dias Araujo para o exame de qualificação apresentado ao Instituto de Matemática e Estatística da Universidade de São Paulo como requisito para obtenção de título de Mestre em Ciências.

# Resumo

ARAÚJO, T. H. D. **Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica**. 2016. 20 f. Exame de qualificação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016.

A comunidade científica pode ser vista como uma rede onde cada pesquisador se relaciona com outros pesquisadores através de colaborações, como na co-autoria de um artigo científico. Alguns trabalhos utilizam ontologias para modelar o domínio das redes de colaboração, e outros aplicam técnicas de aprendizado de máquina para prever ligações entre pessoas dentro dessas redes. Entretanto, algumas limitações existem nessas metodologias por utilizarem formas pouco expressivas de representação dessas relações, ou por não aproveitarem características específicas das entidades na análise desse domínio. A proposta do presente trabalho é criar uma Ontologia capaz de indicar características próprias dessas pessoas, e descrever relações entre elas, aplicando esse conhecimento em um modelo de aprendizado capaz de descobrir e prever novas relações entre esses pesquisadores.

**Palavras-chave:** redes de colaboração científica, ontologia, aprendizado de máquina, predição de ligações.



# Abstract

ARAUJO, T. H. D. **Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica**. 2016. 20 f. Exame de qualificação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016.

[illegible]

**Keywords:** scientific collaboration networks, ontology, machine learning, link prediction.



# Sumário

<b>Lista de Abreviaturas</b>	<b>ix</b>
<b>Lista de Figuras</b>	<b>xi</b>
<b>Lista de Tabelas</b>	<b>xiii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	2
1.2 Objetivos . . . . .	2
1.3 Metodologia . . . . .	3
1.4 Contribuições . . . . .	3
1.5 Organização do Trabalho . . . . .	3
<b>2 Conceitos</b>	<b>5</b>
2.1 Aprendizado Supervisionado . . . . .	5
<b>3 Trabalhos Correlatos</b>	<b>7</b>
3.1 scriptLattes . . . . .	7
3.2 Predição de Ligações . . . . .	7
3.2.1 Importância de um pesquisador, financiamento, e quantidade de colaboradores em um projeto . . . . .	7
3.2.2 Predição utilizando Aprendizado Supervisionado . . . . .	8
3.2.3 Grafos Relacionais com Atributos . . . . .	8
3.2.4 Predição de Citações . . . . .	8
3.3 Ontologias . . . . .	9
3.3.1 Plataforma Lattes . . . . .	9
3.4 Justificativa . . . . .	9
<b>4 Conclusões</b>	<b>11</b>
<b>5 Cronograma</b>	<b>13</b>
<b>Referências Bibliográficas</b>	<b>15</b>





# Lista de Abreviaturas

SPARQL	( <i>SPARQL Protocol and RDF Query Language</i> )
SQL	( <i>Structured Query Language</i> )
SRL	( <i>Statistical Relational Learning</i> )
SVM	( <i>Support Vector Machine</i> )
SWRL	( <i>Semantic Web Rule Language</i> )



# Lista de Figuras



# Lista de Tabelas

5.1 Cronograma de atividades . . . . . 13



# Capítulo 1

## Introdução

A internet não é apenas um repositório quase infinito de informações aleatórias em páginas da web. Em uma análise mais apurada, ela passa a ilustrar imensas cadeias de significados. Com o surgimento da área de Mineração de Dados, esse imenso conjunto de dados começou a ser analisado e utilizado como base para a construção de novos conhecimentos.

Por essa razão, quando falamos da quantidade maciça de informação disponível, conhecida como *Big Data*, devemos pensar também que é possível extrair muito **significado** desses dados. Para [Kay \(2014\)](#), o que chamamos de *Big Data* não é algo relevante pela sua quantidade, mas sim pelo que ele chama de *Big Meaning*, isto é, toda a riqueza de significados que pode ser extraída desses dados.

Nesse panorama, surgem novas possibilidades de extração de informação e de construção do conhecimento de forma automatizada. Através de modelos de classificação e de análise, tornou-se possível, segundo [Halevy et al. \(2009\)](#), construir sistemas que entendem uma frase em determinada língua e a traduzem para outra, ou que interpretam e classificam textos, tudo isso usando como exemplo apenas os inúmeros textos contidos na web.

Ainda como desdobramento dessas novas técnicas, tornou-se possível analisar as interações entre seres humanos em redes sociais, como nas redes de colaboração científica, que contêm informações de autores, publicações e seus temas, em diversas áreas do saber. É possível identificar comunidades interessadas em temas similares, grupos de pesquisadores, e mudanças nos interesses dessas comunidades. Através disto, é possível saber mais a respeito do dia-a-dia da produção científica, e do desenvolvimento da Ciência em geral ao longo do tempo.

Com especial interesse, podemos analisar essas redes de colaboração, que formam um subconjunto da comunidade científica, e avaliar o desenvolvimento da Ciência em todo o mundo, descobrindo quais são as principais contribuições e tendências, como se dá o relacionamento entre pesquisadores, e qual a importância e influência de grupos de pesquisa em suas comunidades. Esse tipo de análise é interessante também em levantamentos sobre produtividade, impacto de uma pesquisa, relevância e popularidade de um projeto ou uma área, como formas de se mensurar e quantificar o progresso científico.

Várias técnicas de análise podem ser empregadas nesse problema, desde as mais simples, como relatórios de produtividade, até as mais complexas, como os agrupamentos de pesquisadores em categorias específicas ou a classificação de trabalhos de forma automática. Na literatura, encontramos algumas ferramentas muito úteis para essa análise. Uma delas trata do problema da predição de ligações entre pessoas, que permite prever, com boa segurança, novas colaborações entre pesquisadores, através da verificação de suas características e dos seus trabalhos publicados.

Como é usual no método científico, encontramos algumas limitações nessas ferramentas, geralmente derivadas de métodos da área de Aprendizado de Máquina. Algumas também não levam em consideração características dos pesquisadores, mas apenas da estrutura da rede. Existem representações mais expressivas que poderiam ser aplicadas ao domínio com facilidade, e várias técnicas relacionadas à área de Sistemas Baseados em Conhecimento que simplificariam essa tarefa, e aumentariam sua eficácia.

Com o presente trabalho, esperamos contribuir para esse tipo de análise, aplicando técnicas

de Sistemas Baseados em Conhecimento, e comparando o resultado com outros trabalhos, para demonstrar a eficácia das técnicas propostas.

## 1.1 Motivação

Uma rede social é uma estrutura relacional que podemos facilmente representar por um grafo com atributos (Cervantes (2014)), ou por uma ontologia (Anauê e Yamate (2009)), ou até mesmo como um banco de dados. Uma rede de colaboração científica possui uma natureza multi-relacional, pois uma publicação pode ter vários tipos de relacionamento diferentes, como citações, autoria, veículo de publicação, dentre outros. E um autor pode se relacionar com outro via coautoria de um artigo, participação em banca ou conferência, fazendo parte do mesmo grupo de pesquisa, ou tendo interesses similares, dentre inúmeras outras relações possíveis. Todos esses atributos podem contribuir na análise da estrutura dessa rede.

Entretanto, se desejamos aplicar um modelo de aprendizado para analisar alguma característica da rede, geralmente utilizamos modelos que recebem como entrada tabelas de atributos e valores. Por causa disso, é preciso transformar uma representação relacional mais rica (como um grafo) em uma representação mais simples (como uma tabela). Portanto, podemos dizer que uma ferramenta de aprendizado supervisionado desse tipo utiliza representações pouco expressivas. Segundo Raedt (2008), existem várias representações que podem ser utilizadas para modelar um problema de aprendizado, algumas mais expressivas, outras menos. Ele mostra em seu trabalho que existe uma hierarquia dessas representações em uma ordem crescente de expressividade.

Além da análise da estrutura da rede, podemos investigar outras características, como o perfil de um pesquisador, seu currículo, sua experiência, e quem são seus pares. Ou a área de pesquisa onde estão inseridas as suas publicações mais importantes. E não é nada fácil representar esse tipo de informação em tabelas de atributos e valores.

A principal motivação do trabalho é a possibilidade de se utilizar conhecimento prévio do domínio, também chamado de *background-knowledge*. A proposta é construir uma ontologia que representa a rede de colaboração, capaz de inferir novas relações, regras, e derivar novas características das entidades pertencentes ao domínio. Assim, é possível extrair esse conhecimento (*background-knowledge*) e transformá-lo em novos atributos da rede de colaboração, enriquecendo os dados a serem explorados pelo modelo de predição. Esperamos, com isso, aumentar a eficiência do modelo.

A principal vantagem do uso de uma ontologia é que o conhecimento extraído é declarativo, compacto e de alto-nível, o que simplifica a sua análise e validação. Sem contar que esse conhecimento a respeito de uma entidade do domínio pode se espalhar para outras, e gerar novos atributos.

## 1.2 Objetivos

O objetivo do presente trabalho é demonstrar que um modelo de predição de ligações em uma rede de colaboração científica, quando enriquecido com conhecimento prévio (*background knowledge*) extraído de uma ontologia, se torna mais eficaz em sua tarefa de predição. Para isso, espera-se cumprir os seguintes passos:

- Modelar uma Ontologia para representar o conhecimento a respeito de uma rede de colaboração científica, contendo informações sobre pesquisadores, publicações e colaborações, e instituições.
- Construir consultas em SPARQL, que utilizam regras de inferência na Ontologia, que serão utilizadas como conhecimento prévio. Algumas dessas consultas são:
  - Qual a área de pesquisa de um pesquisador?
  - Quem são os pesquisadores que ele orientou?
  - Há quantos anos um dado pesquisador trabalha em uma instituição?



- Construir um modelo de predição de ligações baseado no trabalho de [Cervantes \(2014\)](#).
- Enriquecer o modelo de predição com o conhecimento prévio extraído da ontologia, e comparar a eficácia dos dois modelos.
- Propor novas consultas à ontologia que possam ser utilizadas como conhecimento prévio aplicado ao modelo, e fazer novos testes.

### 1.3 Metodologia

Os objetivos traçados serão alcançados através da execução dos seguintes itens:

1. **Estudo dos resultados de [Cervantes \(2014\)](#):**

O objetivo desse item é entender o modelo de representação das redes de colaboração como grafos com atributos e o modelo de predição utilizando aprendizado supervisionado para que possa ser implementado e enriquecido.

2. **Estudo sobre modelagem de Ontologias e Consultas via SPARQL:**

O objetivo desse item é ...

3. **Propor uma forma de extração de conhecimento prévio da ontologia:**

O objetivo desse item é propor um método de extração das informações obtidas via consulta da ontologia, construção de uma base de conhecimento prévio, e utilização desse conhecimento no enriquecimento do modelo, via adição de novos atributos aos vértices da rede de colaboração científica .

4. **Executar a ideia proposta:**

Desenvolver um algoritmo que recebe como entrada um conjunto de dados sobre pesquisadores, suas publicações e seus coautores, que seja capaz de popular a ontologia e extrair dela o conhecimento prévio, enriquecer o modelo de predição, e executá-lo.

5. **Realizar teste da ideia com informações da Plataforma Lattes:**

Preparar um ambiente de testes, que contenha um conjunto de dados extraídos da Plataforma Lattes através do *scriptLattes* e a ontologia. Popular a ontologia, e extrair o conhecimento prévio. Rodar o modelo de predição com e sem o enriquecimento com conhecimento, e comparar os resultados.

### 1.4 Contribuições

A principal contribuição esperada deste trabalho é a construção de um modelo que utilize aprendizado supervisionado e seja capaz de receber informação (*background-knowledge*) extraída de uma ontologia e de tratar o problema de predição de ligações em uma rede de colaboração científica de forma eficiente. O interesse de tal contribuição é a proposta de uma forma de representação mais compacta e expressiva do conhecimento prévio, algo que pode permitir uma fácil validação, expansão, generalização e reutilização desse conhecimento, e uma consequente melhoria da capacidade de predição desse modelo.

### 1.5 Organização do Trabalho

No Capítulo 2, apresentamos alguns conceitos teóricos que servem para dar um melhor entendimento do trabalho. Já no Capítulo 3, exploramos alguns trabalhos relacionados, algumas questões que ainda não foram exploradas em outros trabalhos, e justificamos a importância deste projeto de pesquisa. Finalmente, no Capítulo 4, delimitamos o escopo do projeto e discutimos algumas conclusões. O Capítulo 5 trata do cronograma de atividades necessárias para a execução do projeto.



# Capítulo 2

## Conceitos

Alguns conceitos teóricos são importantes para o entendimento do presente projeto. Aqui, discutimos os fundamentos mais importantes.

### 2.1 Aprendizado Supervisionado

O aprendizado supervisionado é um tipo de aprendizado de máquina em que o agente, durante a fase de treinamento, observa alguns exemplos de entradas (*inputs*) e de saídas (*outputs*) esperadas e aprende a classificá-las, através de uma função de transformação de entradas em saídas. Depois disso, esse agente observa uma base de testes com inúmeros exemplos e classifica esses dados de acordo com o seu modelo interno.

Segundo [Russell e Norvig \(2009\)](#)...



## Capítulo 3

# Trabalhos Correlatos

Neste capítulo, apresentamos algumas contribuições importantes que tratam do problema de análise de redes de colaboração científica e predição de ligações, e outros trabalhos relacionados que serviram de inspiração para este projeto de pesquisa.

### 3.1 scriptLattes

A Plataforma Lattes é uma plataforma vinculada ao CNPq, e a mais importante base integrada de currículos, grupos de pesquisa e instituições de ensino do Brasil, registrando informações valiosas sobre as atividades de pesquisa e o perfil de pesquisadores de diversas áreas da Ciência e Tecnologia em todo o país.

O projeto *scriptLattes*, proposto por [Mena-Chalco e Junior \(2009\)](#), é um sistema capaz de fazer mineração dos dados de currículos presentes na Plataforma Lattes e gerar vários relatórios acadêmicos, além de disponibilizar informações sobre as publicações dos pesquisadores brasileiros, fazendo desambiguação dos autores e artigos, e exportando dados sobre coautoria e outros tipos de colaboração.

O presente trabalho vai utilizar uma base de mais de 4 milhões de currículos extraídos da Plataforma Lattes, gentilmente cedida pelo professor Jesús P. Mena-Chalco, como base de testes para o modelo aqui proposto.

### 3.2 Predição de Ligações

Vários trabalhos presentes na literatura exploram o problema da predição de ligações (*link prediction*) em redes sociais. Esse problema possui diversas aplicações, como na análise e reconstrução de redes e em sistemas que utilizam informações pessoais para sugerir novos contatos ou novos amigos. Outros pretendiam detectar ligações entre pessoas em redes terroristas, com o intuito de prevenir ataques. Uma aplicação interessante tinha por intuito encontrar artigos ou documentos relacionados e sugerir citações.

Discutimos a seguir alguns trabalhos relevantes.

#### 3.2.1 Importância de um pesquisador, financiamento, e quantidade de colaboradores em um projeto

Em [Ebadi e Schiffauerova \(2015\)](#), encontramos uma análise muito apurada sobre uma rede de colaboração científica, que serviu para determinar os fatores que determinam a importância de um pesquisador nessa rede, sua centralidade e sua produtividade. Foram feitas várias constatações interessantes. Descobriram, por exemplo, que grupos ligados a organizações que produziam muitos artigos importantes acabavam tendo uma melhor performance do que outros grupos.

Outro fato interessante é que a experiência e os anos de trabalho de um indivíduo faz com que ele seja mais conhecido em sua comunidade. Tendo acesso a dados sobre financiamento e fomento

à pesquisa, constataram que indivíduos que participam de pesquisas fomentadas por indústrias e empresas acabam atraindo mais colaboradores.

Essa análise também foi baseada em medidas feitas na estrutura da rede, como a centralidade dos vértices, *eigenvectors* e coeficientes de agrupamento, número médio de coautores (ligações) de um dado pesquisador, dentre outros indicadores. Com essas medidas, foi possível encontrar os líderes de diversas comunidades, que são pessoas com grande influência local.

Descobriram, ainda, que os pesquisadores mais produtivos e com o trabalho de melhor qualidade também eram os mais colaborativos. Também mostraram a influência que o financiamento causa nas redes locais de colaboração ao longo do tempo, fazendo com que os cientistas que recebem financiamento e ocupam posições de liderança busquem coautores em outras comunidades mais distantes em vez de aumentar suas conexões na comunidade local.

### 3.2.2 Predição utilizando Aprendizado Supervisionado

Hasan *et al.* (2006) aplicou e comparou a eficácia de diversas técnicas de aprendizado supervisionado, tais como *SVMs*, *Árvores de Decisão*, *Multilayer Perceptrons*, modelos de classificação utilizando kNN (*k-nearest neighbors*), *Naïve Bayes*, e *RBF Networks*.

O uso de *SVMs* mostrou-se melhor do que as outras técnicas, por obter uma taxa de acerto um pouco maior, e também por possibilitar a escolha das características importantes a serem utilizadas pelo modelo, através de um método automatizado de escolha de características.

O aspecto mais importante desse projeto é que a escolha do algoritmo não parece fazer tanta diferença assim na eficácia do modelo. Podemos, com isso, concluir que uma boa escolha das características são um fator mais importante na eficácia do modelo.

### 3.2.3 Grafos Relacionais com Atributos

Um experimento muito interessante e inspirador foi explorado por Cervantes (2014), que modelou a rede de colaboração científica através de grafos relacionais com atributos para mostrar as relações de coautoria entre pesquisadores. Foi também criado um modelo baseado em *SVMs* capaz de prever novas colaborações (ligações) entre pesquisadores a partir de dados de treinamento extraídas da Plataforma Lattes. Essa base foi extraída através do *scriptLattes*.

O modelo proposto permitiu uma série de outras análises relevantes da estrutura dessa rede, como a identificação de pesquisadores mais importantes, ou mais colaborativos, que correspondem aos vértices com mais conexões. Também foi possível identificar comunidades dentro de diferentes áreas, formadas por componentes conexos desse grafo.

Por sua flexibilidade e alto nível de expressividade, esse modelo em grafo com atributos pode ser facilmente expandido. Entretanto, o modelo de aprendizado utilizado recebe como entrada tabelas de atributos e valores. Por causa disso, é preciso transformar uma representação relacional mais expressiva (um grafo) em uma representação mais simples (uma tabela) para que o algoritmo funcione. Podemos dizer que essa ferramenta de aprendizado supervisionado utiliza representações menos expressivas.

### 3.2.4 Predição de Citações

Uma aplicação do Aprendizado Estatístico Relacional foi feita por Alexandrin Popescul e Ungar (2003) em um sistema de predição de citações, que nada mais é do que um modelo de predição de ligações entre documentos. Analisando informações a respeito de duas publicações, o sistema calculava a probabilidade de uma publicação citar a outra. Com isso, o sistema poderia sugerir trabalhos correlatos que pudessem ser citados por um autor durante a escrita de sua publicação.

O grande problema desse modelo de aprendizado é a seleção de atributos (*features*) que auxiliem na tarefa de classificação. Como o problema está modelado de forma relacional, a quantidade de atributos pode crescer indefinidamente. É preciso, portanto, encontrar uma forma inteligente de adicionar um atributo e avaliar sua relevância, através de algum tipo de verificação. No trabalho de

Popescul e Ungar (2007) é sugerido um modelo interessante de seleção de atributos: a geração de novos atributos se dá via busca no conjunto de atributos possíveis, e sua seleção é feita através de um teste que avalia a significância estatística da inclusão desse atributo no modelo de predição. No caso desse trabalho, que foi aplicado ao mesmo problema de predição de citações, o conjunto de atributos possíveis surgiam de consultas SQL relacionadas às entidades pertencentes a um banco de dados relacional, e o algoritmo explorava esse conjunto e selecionava os atributos mais relevantes segundo sua significância estatística. O interessante desse trabalho é que o algoritmo, ao gerar os novos atributos, adicionava essas informações ao banco de dados como novas relações entre entidades, enriquecendo o modelo.

Nossa proposta é parecida, pois também desejamos gerar novos atributos para essas entidades, explorando algumas consultas possíveis, extraíndo conhecimento e formalizando esse resultado como *background knowledge*, que será usado para enriquecer o modelo de aprendizado. A principal diferença está no uso de uma ontologia em vez de um banco de dados relacional.

### 3.3 Ontologias

Uma rede social é uma estrutura relacional que podemos facilmente representar por um grafo, por uma ontologia, ou até como um banco de dados relacional. E uma rede de colaboração científica possui uma natureza multi-relacional, pois uma publicação pode ter vários tipos de relacionamento diferentes, como citações, autoria, veículo de publicação, dentre outros. E um autor pode se relacionar com outro via coautoria de um artigo, participação em banca ou conferência, fazer parte do mesmo grupo de pesquisa, ou ter interesses similares, dentre inúmeras outras relações possíveis.

Alguns trabalhos modelaram ontologias para tratar de problemas relacionados ao uso da Plataforma Lattes e outras redes de colaboração.

#### 3.3.1 Plataforma Lattes

No caso da Plataforma Lattes, Anauê e Yamate (2009) construíram uma ontologia capaz de responder a diversas questões de competência. O usuário consulta a base usando linguagem natural, e a ontologia consegue encontrar a informação desejada, através de inferência. A vantagem dessa metodologia é a facilidade na modelagem, nas ferramentas de inferência, e na expressividade da representação. Galego (2013) também desenvolveu uma ontologia para esse domínio, mas seu interesse era trabalhar com Web Semântica, gerar relatórios e detectar inconsistências.

### 3.4 Justificativa

TODO TODO TODO





## Capítulo 4

## Conclusões

[illegible]



## Capítulo 5

# Cronograma

O cronograma de atividades está descrito na tabela a seguir.

	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov
1	X								
2	X	X							
3		X	X						
4			X	X					
5					X	X	X		
6					X	X	X	X	
7									X

**Tabela 5.1:** *Cronograma de atividades*

*Item 1.* Estudar o domínio e modelar uma Ontologia.

*Item 2.* Modelar as consultas à ontologia que servirão para enriquecer o modelo de predição.

*Item 3.* Construir um modelo de predição de ligações baseado nos trabalhos da literatura, fazendo experimentos com diferentes técnicas de aprendizado supervisionado.

*Item 4.* Propor e implementar um algoritmo de extração de conhecimento da ontologia e enriquecimento do modelo de predição.

*Item 5.* Montar ambiente de testes e elaborar alguns testes de comparação.

*Item 6.* Analisar os resultados e escrever a dissertação.

*Item 7.* Apresentar a dissertação.



# Referências Bibliográficas

- Alexandrin Popescul e Ungar(2003)** Rin Popescul Alexandrin Popescul e Lyle H. Ungar. Statistical relational learning for link prediction. Em *In Proceedings of the Workshop on Learning Statistical Models from Relational Data at IJCAI-2003*. Citado na pág. 8
- Anauê e Yamate(2009)** Costa Anauê e Fabio Yamate. Semantic Lattes: uma ferramenta de consulta de informações acadêmicas da base Lattes baseada em ontologias, 2009. URL <http://www.pcs.usp.br/{~}pcspf/2009/Trabalhos/Cooperativo/G4/monografia.pdf>. Citado na pág. 2, 9
- Cervantes(2014)** Everlyn Perez Cervantes. Análise de Redes de Colaboração Científica: Uma Abordagem Baseada em Grafos Relacionais com Atributos, 2014. Citado na pág. 2, 3, 8
- Ebadi e Schiffauerova(2015)** Ashkan Ebadi e Andrea Schiffauerova. How to become an important player in scientific collaboration networks? *Journal of Informetrics*, 9(4):809–825. ISSN 17511577. doi: 10.1016/j.joi.2015.08.002. URL <http://dx.doi.org/10.1016/j.joi.2015.08.002><http://linkinghub.elsevier.com/retrieve/pii/S1751157715000565>. Citado na pág. 7
- Galego(2013)** Eduardo Ferreira Galego. Extração e Consulta de Informações do Currículo Lattes baseada em Ontologias Eduardo Ferreira Galego Dissertação apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo para obtenção do título de Mestre em Ciências da Computação. página 70. Citado na pág. 9
- Halevy et al.(2009)** Alon Halevy, Peter Norvig e Fernando Pereira. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2):8–12. ISSN 1541-1672. doi: 10.1109/MIS.2009.36. URL <http://dx.doi.org/10.1109/MIS.2009.36>. Citado na pág. 1
- Hasan et al.(2006)** Mohammad Al Hasan, Vineet Chaoji, Saeed Salem e Mohammed Zaki. Link prediction using supervised learning. Em *In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security*. Citado na pág. 8
- Kay(2014)** Alan Kay. The future doesn't have to be incremental, 2014. URL <https://www.youtube.com/watch?v=gTAghAJcO1o>. DEMO Enterprise 2014. Citado na pág. 1
- Mena-Chalco e Junior(2009)** Jesús Pascual Mena-Chalco e Roberto Marcondes Cesar Junior. scriptlattes: an open-source knowledge extraction system from the lattex platform. *Journal of the Brazilian Computer Society*, 15(4):31–39. ISSN 0104-6500. doi: 10.1007/BF03194511. URL <http://dx.doi.org/10.1007/BF03194511>. Citado na pág. 7
- Popescul e Ungar(2007)** Alexandrin Popescul e Lyle H. Ungar. Feature generation and selection in multi-relational statistical learning. Em Lise Getoor e Ben Taskar, editors, *Introduction to Statistical Relational Learning*, chapter 16, páginas 453–475. The MIT Press. Citado na pág. 9
- Raedt(2008)** Luc De Raedt. *Logical and Relational Learning*. Cognitive Technologies. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-540-20040-6. doi: 10.1007/978-3-540-68856-3. URL <http://www.springer.com/us/book/9783540200406><http://link.springer.com/10.1007/978-3-540-68856-3>. Citado na pág. 2

**Russell e Norvig(2009)** Stuart Russell e Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edição. ISBN 0136042597, 9780136042594. Citado na pág. [5](#)