

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**Sistema de Recomendação de Recursos
Educativos: uma abordagem baseada em
difusão de informação em redes sociais**

Felipe Neves Braz

JUIZ DE FORA
JULHO, 2017

Sistema de Recomendação de Recursos Educativos: uma abordagem baseada em difusão de informação em redes sociais

FELIPE NEVES BRAZ

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Sistemas de Informação

Orientador: Victor Ströele de Andrade Menezes

JUIZ DE FORA

JULHO, 2017

SISTEMA DE RECOMENDAÇÃO DE RECURSOS
EDUCACIONAIS: UMA ABORDAGEM BASEADA EM DIFUSÃO
DE INFORMAÇÃO EM REDES SOCIAIS

Felipe Neves Braz

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
BACHAREL EM SISTEMAS DE INFORMAÇÃO.

Aprovada por:

Victor Ströele de Andrade Menezes
Doutor em Engenharia de Sistemas e Computação (UFRJ)

Fernanda Cláudia Alves Campos
Doutora em Engenharia de Sistemas e Computação (UFRJ)

Jairo Francisco de Souza
Doutor em Informática (PUC-RIO)

JUIZ DE FORA
05 DE JULHO, 2017

Resumo

Com a crescente utilização de redes sociais e disponibilização de recursos na Web, muitos trabalhos focam em diferentes métricas para previsão de difusão de informação em redes sociais. Essas redes auxiliam na caracterização de perfis de usuários, permitindo uma absorção mais detalhada de suas preferências, auxiliando Sistemas de Recomendação no processo de busca de conteúdo relevante, tendo em vista esses interesses. Há uma grande dificuldade em recomendar conteúdos, de forma que o usuário tenha interesse em repassá-los à outros. O presente trabalho busca estudar e identificar as métricas de difusão de informação em redes sociais que mais se adequam em um contexto de sistemas de recomendação de recursos educacionais, visando o aumento da difusão desses recursos, quando recomendados.

Palavras-chave: Sistemas de Recomendação, Redes Sociais, Difusão de Informação.

Abstract

With the increasing use of social networks and availability of resources on the Web, many papers focus on different metrics for predicting the diffusion of information on social networks. These networks assist in the characterization of user profiles, allowing a more detailed absorption of their preferences, assisting Recommender Systems in the process of searching for relevant content in view of these interests. There is a great difficulty in recommending content, so that the user has an interest in passing it to others. The present work seeks to study and identify the metrics of information dissemination in social networks that are more appropriate in a context of recommendation systems of educational resources, aiming to increase the diffusion of these resources, when recommended.

Keywords: Recommender Systems, Social Networks, Information Diffusion.

Agradecimentos

Agradeço primeiramente a Deus, pelo dom da vida.

A minha família pelo apoio e sustento, e de modo especial a minha mãe, pelos conselhos ao longo dessa caminhada e o ombro amigo que sempre se mostrou presente.

Ao professor Victor pela orientação, amizade, confiança e principalmente, pela paciência, sem a qual este trabalho não se realizaria.

Aos meus amigos, pelos momentos divertidos e pelos conselhos.

Aos professores do Departamento de Ciência da Computação pelos seus ensinamentos.

Por fim, agradeço as oportunidades que tive durante essa caminhada, e pelos diálogos que, de algum modo, contribuíram para meu enriquecimento pessoal e profissional, pois como dito por Khalil Gibran “as flores desabrocham para continuar a viver, pois reter é perecer”.

“Uma mente necessita de livros da mesma forma que uma espada necessita de uma pedra de amolar, se quisermos que se mantenha afiada.”

George R. R. Martin

Conteúdo

Lista de Figuras	6
Lista de Tabelas	7
1 Introdução	8
1.1 Justificativa	9
1.2 Objetivos	10
1.3 Metodologia	10
2 Revisão Bibliográfica	12
2.1 Redes Sociais	12
2.2 Educação Informal	13
2.3 Sistemas de Recomendação	14
2.4 Trabalhos Relacionados	15
3 Solução Proposta	18
3.1 Arquitetura	18
3.2 Ferramentas Utilizadas	20
3.2.1 Linguagens	21
3.2.2 Banco de Dados	22
3.2.3 Twitter API	23
3.2.4 YouTube Data API	24
3.2.5 DBpedia	24
3.2.6 Google Scholar	25
3.3 Fluxo de Trabalho	25
3.3.1 Construção das Relações	27
3.3.2 Métricas	28
3.3.3 Escolha dos Usuários	30
3.3.4 Extração de Preferências de Perfil	31
3.3.5 Busca de Recursos Educacionais	33
4 Avaliação	36
4.1 Análise das Recomendações	39
4.2 Considerações Finais do Capítulo	43
5 Considerações Finais	45
Bibliografia	47

Lista de Figuras

2.1	Exemplo da estrutura de um grafo e rede social.	13
3.1	Arquitetura conceitual da pesquisa.	19
3.2	Estrutura do Twitter.	21
3.3	Exemplo de uma consulta Cypher.	23
3.4	Fluxo de trabalho	25
3.5	Workflow de desenvolvimento.	26
3.6	Modelagem do banco de dados em grafo.	27
3.7	Comportamento de difusão no Twitter.	28
3.8	Cypher query para recuperação dos ids de usuários que mais publicaram.	29
3.9	Cypher query para recuperação dos ids de usuários que mais tiveram retweets.	29
3.10	Query de busca SPARQL na base da DBpedia para extração de termos. (PEREIRA, 2015)	33
3.11	Query de busca SPARQL na base da DBpedia para recuperação dos recursos. (PEREIRA, 2015)	33
4.1	Primeira forma de pedido de colaboração estabelecido.	36
4.2	Tweet feito por um potencial influenciador identificado.	37
4.3	Resposta a uma requisição realizada.	37
4.4	Resposta com restrição para seguir o perfil.	37
4.5	Nova forma de pedido para colaboração.	38
4.6	Home page do perfil recomendador após novos pedidos de colaboração.	38
4.7	<i>Feedback</i> à recomendação.	41
4.8	Comportamento automático de retweet.	41
4.9	Post feito pelo usuário com o recurso recomendado.	42
4.10	Resultado do processo de difusão.	42

Lista de Tabelas

3.1	Informações extraídas do Twitter	28
4.1	Análise quantitativa por métrica	39
4.2	Relacionamento de usuários e métricas	39
4.3	Resultados gerais dos tweets por usuário.	40

1 Introdução

A facilidade na disponibilização de recursos na Web (World Wide Web), proporcionou um volume muito expressivo de informações apresentadas aos usuários. Perante esse fato, os usuários demonstram dificuldade em encontrar recursos que são de seu interesse e que sejam aderentes às suas necessidades.

Segundo Burke (2002), um Sistema de Recomendação (SR) pode ser definido como “qualquer sistema que produza recomendações individualizadas ou que tenha o efeito de guiar o usuário de forma personalizada para objetos relevantes ou que lhes sejam úteis dentre as diversas opções possíveis”. Dessa forma, os SR realizam a filtragem de informações, analisando o perfil e os interesses do usuário, para posteriormente recomendar conteúdo ao mesmo ou ao grupo ao qual pertence. A grande dificuldade está em como recomendar conteúdos de forma que tenha a ação esperada, que consiste no usuário consumir o recurso recomendado, que o recurso seja realmente relevante e que o usuário tenha interesse em repassar o recurso a outros.

Nos últimos anos, nota-se o aumento na procura por recursos educacionais através da Web, seja para cursos formais, que se desenvolvem à distância ou presencialmente, ou até mesmo pela procura informal por conteúdos de interesse. Essas buscas muitas vezes são feitas em diferentes repositórios de dados, que possuem uma vasta quantidade de recursos educacionais com finalidades distintas e em diferentes formatos, tornando a escolha desses recursos uma tarefa difícil para os usuários.

Recursos educacionais se caracterizam por qualquer tipo de objeto cujo principal objetivo, seja de cunho educacional, como por exemplo, vídeo-aulas, livros, artigos e slides. Esses recursos podem ser encontrados em Repositórios de Recursos Educacionais, em bases de Vídeos, como o YouTube¹, e em base de dados ligados, como a DBpedia².

Redes Sociais, de acordo com Marteleto (2001) representam “[...] um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses com-

¹<https://www.youtube.com>

²<http://www.dbpedia.pt>

partilhados”. Partindo desse pensamento, a utilização de redes sociais auxilia no processo de caracterização do perfil de interesse do indivíduo, possibilitando uma recomendação mais rica em conteúdo relevante.

Interno às redes sociais, ocorre o fenômeno de difusão de informação, que, conforme descrito por Taxidou e Fischer (2014), é formado quando um usuário, presente na rede social, passa uma mesma informação, sem modificação, para frente, ou seja, compartilha em seu perfil a informação postada pelo usuário “raiz”, que consiste naquele usuário que originou a informação.

Pesquisas apontam o uso de diferentes métricas e abordagens para previsão desse fenômeno, como Taxidou e Fischer (2014), Kempe, Kleinberg e Tardos (2003) e Guille, Hacid e Favre (2013). A maioria dos trabalhos são teóricos, visando expor novas formas de calcular e prever a difusão de informação, baseados nas conexões entre os indivíduos presentes nessas redes. Isso motiva a análise das métricas para difusão de recursos educacionais, possibilitando a identificação das métricas que são mais adequadas para a maximização de difusão de informação, fazendo com que mais pessoas acessem os conteúdos educacionais recomendados.

1.1 Justificativa

Redes sociais oferecem um amplo conjunto de informações suscetíveis à análise, possibilitando a descoberta de informações referentes ao relacionamento entre indivíduos, interesses em comum, dentre inúmeras outras análises que podem ser feitas através de dados extraídos dessas redes (TAXIDOU; FISCHER, 2014).

A recomendação de algo relevante é de suma importância para que conteúdos de qualidade sejam compartilhados, aumentando assim a disseminação de informação, seja ela referente a qualquer tipo de nicho. Logo, recomendando um recurso de forma acertada, ou seja, a recomendação sendo útil ao usuário faz com que o conteúdo tenha uma maior possibilidade de sofrer difusão através de seu compartilhamento ou indicação por aquele que recebeu a recomendação.

Na educação informal, os conhecimentos são partilhados entre os indivíduos, onde os ensinamentos se desenvolvem de forma espontânea, facilitando o aprendizado e o tor-

nando prazeroso (CASCAIS; FACHÍN-TERÁN, 2013). Com o advento da Internet, isso se mostra ainda mais possível, possibilitando interações instantâneas entre usuários de modo a propagar mais rapidamente as informações.

Portanto, a proposta se motiva pela dinamicidade no consumo e compartilhamento de informações por usuários de redes sociais, e pela disseminação de informação através dessas redes. Além da possibilidade de extrair automaticamente características de perfil e de contexto educacional do usuário, visando a recomendação individualizada e personalizada.

1.2 Objetivos

O trabalho apresenta um modelo de recomendação de recursos educacionais baseado na aprendizagem informal e métricas de difusão de informação. A aprendizagem informal caracteriza-se pela própria busca do aprendiz por objetos de interesse, tornando a maneira de adquirir conhecimento mais agradável (GOHN, 2006). A pesquisa propõe uma forma de otimizar o alcance de pessoas à recursos educacionais de interesse, visando a obtenção de conhecimento, com isso, utilizando sistemas de recomendação.

Assim, são recomendados recursos educacionais objetivando a disseminação dos mesmos, tendo como base as preferências de mídia e de conteúdo dos usuários, na rede social Twitter. Para tal, é feita a extração das publicações e relacionamentos dos perfis dos usuários para identificar e recomendar para aqueles mais propensos à gerar difusão dos recursos recomendados. Por fim, é feita uma avaliação das recomendações, visando analisar qual métrica de escolha de difusores mostra maior alcance do recurso na rede.

1.3 Metodologia

O presente trabalho possui natureza aplicada e abordagem quantitativa, pois busca desenvolver uma melhor forma de recomendar conteúdo a pessoas, utilizando redes sociais como base de estudo e proporcionando o melhoramento da escolha de pessoas mais propensas a difundir informação dentro da rede.

O tipo de pesquisa utilizada no trabalho é a pesquisa bibliográfica, onde é con-

cebida a partir de materiais já publicados, mesclada com a pesquisa experimental. No que diz respeito a esse trabalho, refere-se a quando e porque uma recomendação é mais propensa a ocorrer do que em outras situações.

A utilização dessas duas metodologias de pesquisa fornece tanto uma base teórica consistente, quanto possibilita a obtenção de informação nova através dos experimentos nos dados coletados, além de observar os efeitos produzidos no estudo.

Como o universo de redes sociais é muito extenso em sua totalidade, a pesquisa utiliza da técnica de amostragem, sendo extraídas parte das informações e relacionamentos pertinentes aos perfis dos usuários. Como dito anteriormente, o ambiente que se pretende realizar a extração dos dados e a recomendação dos recursos educacionais é a rede social Twitter, onde será possível realizar a criação de uma base de dados e aplicar as métricas e técnicas de difusão de informação e centralidade estudadas, buscando recomendar e analisar a propagação de recursos educacionais nesta rede.

2 Revisão Bibliográfica

Visando os conceitos abrangentes ao tema do trabalho, serão abordadas na revisão bibliográfica questões como a análise de redes sociais, o desenvolvimento e evolução da educação informal a partir do uso da Internet, bem como os conceitos ligados à Sistemas de Recomendação.

2.1 Redes Sociais

Em análise de redes sociais, são combinadas estatística e matemática discreta com sistemas complexos para explorar os processos e fenômenos, bem como a difusão de informação que ocorrem dentro de uma organização (MCMAHON; MILLER; DRAKE, 2001). Uma definição simples de rede social é dada por Emirbayer e Goodwin (1994), sendo “um conjunto de relações sociais ou laços sociais entre um conjunto de atores (e os atores ligados entre si)”. Esses atores podem ser compreendidos como pessoas, organizações ou grupos de pessoas. A partir dessa definição, podemos perceber que redes sociais abrangem um número vasto de possibilidades de nichos de representações, reforçando o que McMahon, Miller e Drake (2001) dizem, que “A análise de redes sociais é estudada cientificamente em diversas disciplinas, como sociologia, antropologia, psicologia, geografia, matemática, estatística e ciência da computação”.

Redes sociais ou sociograma normalmente são representados por grafos, onde um grafo, é um par $G=(N,E)$, tendo N como um conjunto de vértices ou nós, e E o conjunto de ligações ou arestas presentes em G . Essa notação se aplica em redes direcionadas ou não-direcionadas (HELLMANN; STAUDIGL, 2014). A utilização de grafos para representação de redes como um todo é melhor aplicada, pois permite a compreensão e visualização mais facilitada de toda a estrutura de ligações, conforme ilustrado na Figura 2.1

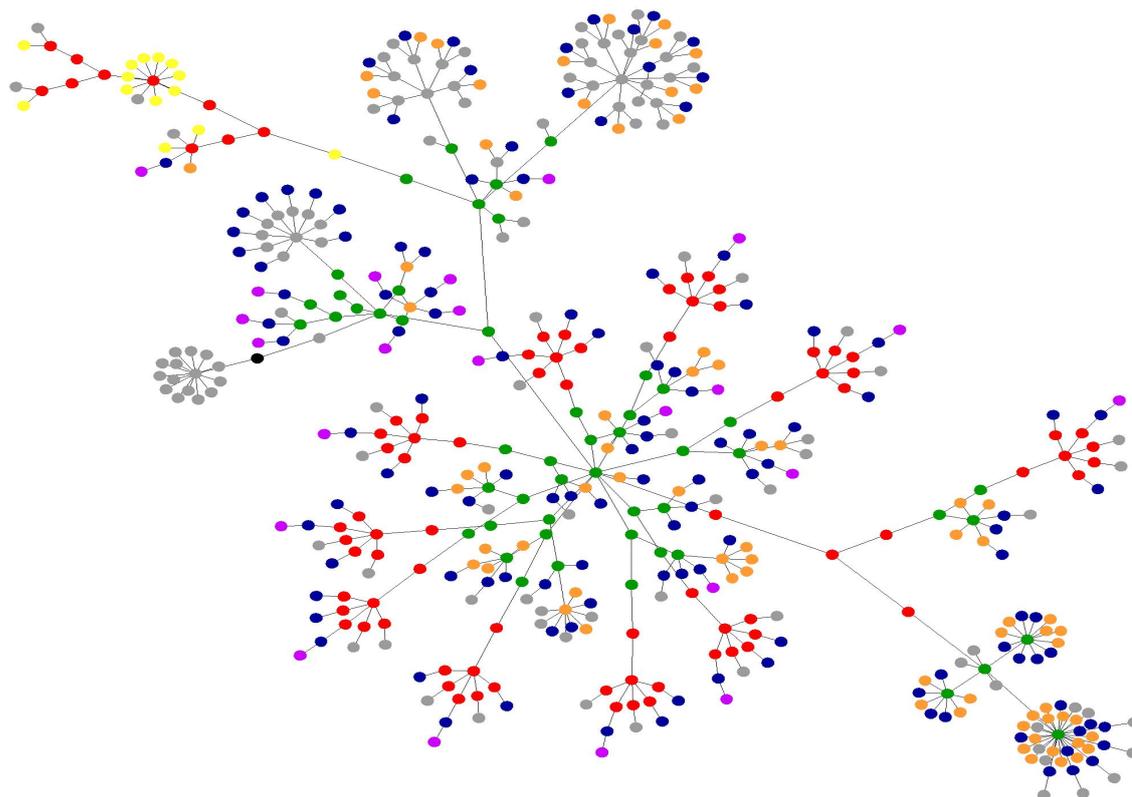


Figura 2.1: Exemplo da estrutura de um grafo e rede social.

As redes sociais são as responsáveis pela interação entre pessoas e o compartilhamento de ideias. As em formato digital favorecem esse comportamento em uma escala global, dando a possibilidade de interação com maior número de pessoas e a difusão de informações pelas mesmas.

2.2 Educação Informal

A educação, desde o início das civilizações, está presente na base da sociedade, sendo entendida como um processo de desenvolvimento da capacidade intelectual do ser humano (GASPAR, 2002) onde é responsável pela criação e transmissão de cultura que molda as características de um grupo, além de gerar a possibilidade de evolução das civilizações intelectual e estruturalmente.

Ao contrário da educação formal, que segundo Gohn (2006) é relativa “ao ensino e aprendizagem de conteúdos historicamente sistematizados, normatizados por leis”,

dependendo assim “de uma diretriz educacional centralizada como o currículo, com estruturas hierárquicas e burocráticas” (GADOTTI, 2005); a educação informal é aquela onde as pessoas adquirem conhecimento através da vivência do dia-a-dia, como, por exemplo, em conversas entre amigos, jornais, programas de TV, sites e em vídeos online. Ela não possui vínculo à local de ensino ou nível de conhecimento; está diretamente ligada à valores e a cultura de cada localidade; possuindo uma única especificidade, ter quem queira ensinar e quem queira aprender (GASPAR, 2002).

Hargis (2001) mostra que programas de computador juntamente com a Internet proveem acesso a uma gama extensa de informações, conteúdos e mídias em diferentes formatos. A utilização dessas tecnologias online, nas quais possuem ferramentas de busca, auxiliam no encontro de materiais necessários para obtenção de conhecimento, isto, através da aprendizagem informal. Porém, diante de tantas possibilidades, as pessoas não encontram o que desejam e necessitam de forma facilitada.

Esses recursos podem ser encontrados em Repositórios de Recursos Educacionais, que têm como propósito agrupar objetos educacionais que se caracterizam por qualquer tipo de mídia, como por exemplo textos, imagens, vídeos, dentre outras, de forma a ser disponibilizado em um determinado meio (MARCHIORI, 2012).

2.3 Sistemas de Recomendação

Os sistemas de recomendação são largamente utilizados, principalmente, em sites de comércio eletrônico e, com o desenvolvimento do serviço de *streaming* da Netflix³, vem sendo utilizado também nesta plataforma. Esses sistemas possuem o intuito de recomendar conteúdos aos usuários a partir da análise de seu perfil e contexto, tendo em vista suas preferências e seus interesses. Para essa recomendação são utilizadas técnicas que auxiliam na caracterização do perfil e contexto dos usuários, na alocação dos mesmos em grupos com necessidades semelhantes, na localização de recursos que atendem às necessidades dos usuários, dentre outras.

Sistemas de Recomendação, quando expostos em contextos educacionais apresentam algumas limitações (COSTA; AGUIAR; MAGALHÃES, 2013), fazendo-se necessário

³<https://www.netflix.com>

analisar, além das preferências do usuário, o quanto o recurso irá ajudá-lo a alcançar seus objetivos de aprendizagem (TANG; MCCALLA, 2004). Para isso, existem, dentre outras, três técnicas básicas de filtragem de informação na literatura, que auxiliam na recomendação desses recursos, são elas: a filtragem baseada em conteúdo, que se caracteriza por utilizar informações de preferência anteriores do usuário como base para a recomendação; a filtragem colaborativa, que busca analisar perfis parecidos ao do usuário que se quer recomendar, recomendando objetos que esses perfis similares preferiram no passado, além da abordagem híbrida, que busca combinar as duas técnicas anteriores, tentando minimizar seus pontos fracos (BURKE, 2002; REATEGUI; CAZELLA, 2005).

2.4 Trabalhos Relacionados

O estudo de técnicas para recomendação e difusão em redes sociais, embora seja recente, já conta com vários trabalhos com abordagens distintas para resolver o problema. O trabalho de Guille, Hacid e Favre (2013), tenta prever a difusão de informação na rede social Twitter utilizando um modelo baseado em difusão ao longo do tempo. Nele, é levado em consideração a quantidade de postagens, bem como a frequência das mesmas, para calcular qual indivíduo é mais propenso a difundir informações durante o tempo. Esse estudo faz perceber que existe uma influência de um indivíduo para outro, sendo expressa pela reputação do mesmo na rede ou até mesmo pela quantidade de seguidores que possui.

Algumas abordagens realizam um estudo de como a informação atinge uma distância maior no fluxo de difusão, sem levar em conta o porquê da difusão acontecer. Como em Kempe, Kleinberg e Tardos (2003), que tenta atingir nós mais rapidamente, interagindo com os mesmos, fazendo-os visualizar a informação através de diferentes tipos de abordagens, como escolha randômica dos nós ou através da escolha baseada em um peso de influência entre os vizinhos. Esse tipo de abordagem é rebatida no trabalho de Wang, Chen e Wang (2012), onde é feita uma análise comparativa de diferentes modelos para predição de *cascade*, que consiste no fluxo de difusão. É mostrado que a probabilidade de um nó atingir outro está diretamente ligada à quantidade de caminhos entre dois nós, sendo necessário analisar a influência de um perante outro com base em suas

ligações. A influência que um nó faz sobre outro nó diminui conforme existem outros nós no caminho entre ambos.

Outro trabalho interessante, que estuda o comportamento da rede social para submeter recomendações, é o realizado por Yu (2012), que compara algoritmos de recomendação por meio de análise de similaridade entre perfis, bem como os algoritmos que utilizam ranqueamento de influência entre usuários. Seu propósito é mostrar uma abordagem de recomendação para usuários no Twitter, baseando-se na natureza de mudança de ligações da rede social, buscando selecionar melhores resultados através de um processo dinâmico de competição entre os algoritmos. Este trabalho conclui que redes sociais não são mais utilizadas apenas para fazer amizades e, sim, estão evoluindo para um novo tipo de redes de informação, mostrando a importância de se encontrar e sugerir informação útil para usuários nas redes sociais.

Pesquisas em sistemas de recomendação já estão inseridas no ambiente educacional, como é mostrado por Costa, Aguiar e Magalhães (2013). Nesse contexto, são apresentadas questões nas quais os sistemas de recomendação devem levar em consideração, sendo necessário validar o recurso a ser recomendado para o usuário. Propostas como Casagrande, Kozima e Willrich (2013) e Ribeiro, Fonseca e Freitas (2013), mostram como esses sistemas podem ser utilizados para educação, podendo as vezes ser de forma simplificada e escalável, possibilitando atender uma grande quantidade de alunos e professores, quando inseridos dentro de instituições, ou com a realização de mapeamentos de recursos por meio de *hashtags* para que o usuário interaja com aqueles recursos que são de seu interesse.

Outros trabalhos abordam a inserção de sistemas de recomendação de recursos educacionais em redes sociais, como mostrado em Pereira (2015), Almeida et al. (2015) e Almeida et al. (2016), onde são levantadas diferentes abordagens no contexto de sistemas de recomendação, tendo em vista grupos de usuários em redes sociais; contribuindo de forma à explorar informações geradas por esses grupos, auxiliando no processo de busca e seleção de recursos educacionais de interesse comum para os participantes.

Este trabalho busca estudar e identificar as métricas de difusão de informação em redes sociais que mais se adequam em um contexto de sistemas de recomendação de

recursos educacionais, visando o aumento da difusão desses recursos quando recomendados e avançando as pesquisas em relação aos trabalhos desenvolvidos na área de Sistemas de Recomendação na Educação.

3 Solução Proposta

Neste capítulo serão tratados os aspectos do desenvolvimento, como a estruturação, modelagem e as ferramentas utilizadas no presente trabalho, bem como o processo de recomendação dos recursos educacionais.

3.1 Arquitetura

Visando solucionar os problemas identificados neste estudo, foi elaborada uma arquitetura baseada em serviços, cuja representação pode ser vista na Figura 3.1 e a descrição de suas camadas será feita ao longo deste capítulo.

A proposta foi pensada para trabalhar com arquitetura baseada em serviços web; possibilitando a integração de partes do sistema com outros os quais precisem de suas funcionalidades.

Serviços web, por padrão, são sistemas que possuem uma única funcionalidade geral: prover alguma informação para as requisições que recebe. Essas requisições seguem o padrão HTTP, nas quais são processadas em um servidor que armazena o serviço e são encaminhadas novamente ao cliente com um formato padrão de estruturação de dados chamado JSON (*JavaScript Object Notation*). Este formato para transmissão de dados na web é largamente utilizado em comunicações cliente e servidor. Dessa forma, a padronização das comunicações entre partes do sistema facilita sua integração com outros que, não necessariamente, precisam estar escritos na mesma linguagem.

A arquitetura proposta, mostrada na Figura 3.1, é modelada em camadas, onde cada camada constitui um subsistema. Cada subsistema fornece um conjunto de serviços que serão detalhados no texto subsequente.

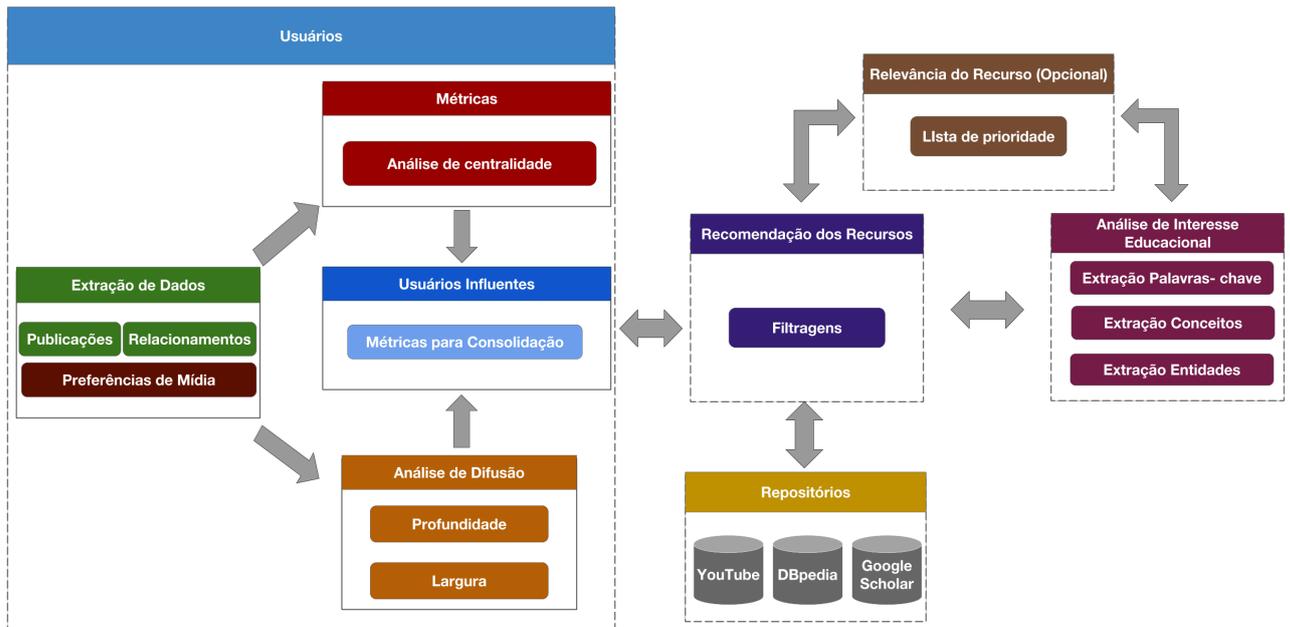


Figura 3.1: Arquitetura conceitual da pesquisa.

A camada de Extração de Dados é responsável por obter os dados necessários para definição de perfil dos usuários. Assim, nesta camada, são extraídas as publicações postadas pelos usuários, seus relacionamentos com outros usuários na rede e suas preferências de formato de mídia. A escolha da preferência de mídia serve como base para definição do tipo de recurso a ser recomendado.

Ligadas à camada de extração de dados estão as camadas de cálculo de métricas, responsável pelos cálculos de centralidade dos usuários; e a camada de análise de difusão, que analisa a importância de cada usuário na rede, em termos de difusão de informação, tanto em profundidade, quanto em largura. Essas duas camadas são responsáveis por realizar os cálculos que precedem à escolha dos usuários que irão receber as recomendações, gerando, cada uma, uma lista de usuários, os quais as métricas apontaram como sendo os mais difusores. A escolha, ocorre em uma quarta camada, denominada Usuários Influentes. Esta, por sua vez, faz a consolidação das listas obtidas de ambas as camadas anteriores, ordenadas de forma decrescente, segundo as métricas calculadas.

Para melhor descrever o conceito da arquitetura, as camadas que estão diretamente relacionadas entre si e com a rede social foram representadas em conjunto, de forma a compor um serviço, ou seja, as camadas de extração de dados, cálculo de métricas, análise de difusão e usuários influentes não funcionam separadamente. As demais cama-

das são pensadas para trabalhar de forma independente, também caracterizando-se como serviços.

A camada de recomendação de recursos é a que está mais diretamente relacionada com as demais, sendo a responsável por realizar a filtragem dos conteúdos baseada em modelos de recomendação e realizar a escolha dos conteúdos que melhor aderem às preferências educacionais dos usuários.

Para a escolha desses conteúdos é necessário fazer um mapeamento das preferências dos usuários. Esse mapeamento é realizado pela camada de extração de conceitos, que extrai conceitos dos perfis dos usuários, possibilitando a definição das preferências de conteúdo dos mesmos.

Os recursos alvo podem ser encontrados em repositórios educacionais online. Os utilizados neste trabalho foram: YouTube, Google Scholar e DBpedia. Em síntese, a camada de recomendação de recursos realiza a busca e a entrega de recursos educacionais aos usuários, ordenando esses recursos por relevância.

Além dessas, pensou-se em uma camada de relevância de recursos, que é opcional, pois alguns repositórios já devolvem os recursos por ordem de relevância. É destinada para aqueles que não possuem essa categorização, como a DBpedia, por exemplo, sendo necessário incorporá-la para realizar uma validação dos recursos e categorizá-los por nível de relevância, antes de encaminhá-los para os usuários.

3.2 Ferramentas Utilizadas

Com o intuito de obter os dados necessários para as análises e recomendações planejadas, o Twitter foi a rede social que mais se mostrou capaz de oferecer os tipos de dados e relacionamentos que esta pesquisa procura. Dessa forma, essa rede social foi escolhida devido sua estrutura e proposta, que é a propagação de informação.

Segundo Kwak et al. (2010), a estrutura do Twitter, representada na Figura 3.2, é diferente de outras redes sociais, como Facebook ou MySpace, onde os usuários não precisam de reciprocidade na relação seguir e ser seguido. A relação de seguir no Twitter significa que o usuário seguidor receberá todas as mensagens, que são denominadas tweets, dos usuários nos quais segue. O Twitter por motivos de escalabilidade, limita a quantidade

de caracteres presentes em descrições de perfil, nome de usuário, e, principalmente, em tweets. Nestes são permitidos no máximo 140 caracteres. Nesse contexto, para informar aos usuários sobre o objetivo da pesquisa e a metodologia utilizada, optou-se por criar um site⁴ e, para realizar as recomendações, foi criado um perfil recomendador no Twitter.

Interno aos tweets, existe uma estrutura que permite a definição das relações entre *posts*, como é o caso dos termos “RT” que demonstra que o *post* é um retweet de uma mensagem publicada por outra pessoa; “@” seguido do nome de usuário, demonstra que o *post* faz menção a este usuário e por fim “#”, responsável por fazer uma marcação de algum termo ou assunto, de modo a facilitar sua identificação quando submetido a uma pesquisa na rede social (KWAK et al., 2010).

O Twitter também suporta comentários e favoritos nos tweets, sendo o primeiro, de certa forma, um tweet que possui um comportamento de resposta ou reação a um outro e, o segundo é uma demonstração quantitativa de quantas pessoas gostaram da postagem, semelhante ao “like” em outras redes sociais. Esses dados se mostram importantes, pois oferecem um grande número de possibilidades de análise para escolha dos usuários-chave para recomendação posterior.

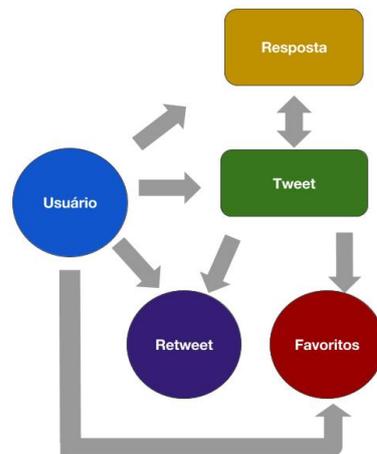


Figura 3.2: Estrutura do Twitter.

3.2.1 Linguagens

Para análise de uma quantidade expressiva de dados, se faz necessário o uso de linguagens que ofereçam processamento mais rápido e objetivo. Tendo isso em mente,

⁴<https://sites.google.com/view/e-recoSYS/home>

foram escolhidas as linguagens Python e R, que oferecem tanto a rapidez no processamento dos dados, quanto na parte de cálculos estatísticos necessários na análise do grafo, que representa a rede, através de suas bibliotecas.

Python e R são largamente utilizadas na área de processamento de dados e *business intelligence*, possuindo muitas funcionalidades referentes a cálculos e amostragens, além de oferecerem muitas bibliotecas gráficas com as quais se faz possível uma melhor visualização dos dados a serem analisados; isso sempre de forma otimizada e consequentemente mais rápida; quando comparada a outras linguagens. Além desses fatores, a escolha dessas linguagens foi devido à fácil programação que oferecem, possibilitando um trabalho mais objetivo e com maior foco nas análises.

As bibliotecas utilizadas nos scripts do trabalho possuem uma ampla base para as computações necessárias, compreendendo as conexões com as bases de dados do Twitter de forma facilitada, através de suas interfaces. São 3 bibliotecas principais de extração de dados utilizadas: a TwitterAPI e tweepy são duas bibliotecas para conexão com a API de busca e de streaming do Twitter, ambas são para a linguagem Python. Outra biblioteca utilizada é a py2neo em sua versão 2.0, que provê conexão com o banco de dados Neo4j, permitindo a manipulação dos dados.

Além dessas, são utilizadas bibliotecas para cálculo de métricas. Ao realizar testes, percebeu-se que o cálculo é feito de maneira mais rápida quando processado utilizando a linguagem R. Dessa forma, com o auxílio de uma biblioteca em Python chamada *rpy2*, responsável por gerar uma interface para o R, podemos calcular as métricas de maneira mais otimizada. E para o cálculo são utilizadas duas bibliotecas em R, sendo uma a *igraph*, responsável por estruturar o grafo em memória e prover as funções para cálculo das métricas de centralidade; e por fim, a biblioteca RNeo4j, responsável por fazer a ligação com o banco e o R, permitindo a recuperação dos nós e relacionamentos armazenados.

As versões das linguagens Python e R utilizadas são 2.7.12 e 3.3.3, respectivamente.

3.2.2 Banco de Dados

Foi escolhido um banco de dados NOSQL em forma de grafos, pois sua estrutura

facilita a representação de redes sociais, as quais podem ser representadas por grafos. O padrão NOSQL traz consigo a característica de escalabilidade mais facilitada, além de fornecer uma robustez maior, sendo importantes fatores quando se trabalha com análise de redes sociais e grande volume de dados.

Esse tipo de banco de dados oferece uma otimização de análise que os bancos relacionais não conseguem exprimir de maneira natural, sendo necessário muita duplicação de informação nas tabelas, que, para um volume de dados muito grande, se torna inviável. Assim, o banco escolhido foi o Neo4j, um banco *Open Source*, escalável, possuindo suporte para importação de dados facilitada, além de permitir a construção de uma modelagem flexível. A versão do Neo4j que foi utilizada é a 3.1.3.

Cypher

O Neo4j, por ser um banco com modelagem baseada em grafos, possui uma linguagem própria de consulta e manipulação de dados, a Cypher. Ao contrário da linguagem SQL, comum aos bancos relacionais; a Cypher é uma linguagem baseada em padrões de ligação, o que torna mais fácil a obtenção das relações e propriedades das mesmas. Um exemplo de uma consulta Cypher pode ser conferido na Figura 3.3.

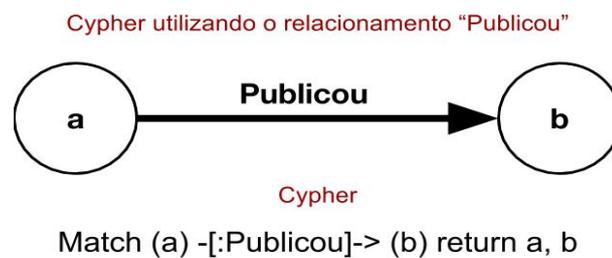


Figura 3.3: Exemplo de uma consulta Cypher.

A consulta na Figura 3.3 irá retornar os nós *a* e *b*, os quais possuem ligação conforme padrão explícito na consulta, onde *a* publicou *b*, juntamente com suas propriedades.

3.2.3 Twitter API

O Twitter possui APIs que possibilitam a extração de informações presentes em sua base de dados. A Search API é a principal API do Twitter, pois é a que oferece maiores

métodos de obtenção de informação. Streaming API e Ads API, são outras duas APIs que o Twitter contém, sendo que a primeira oferece fluxo contínuo de informação, dado uma lista de identificadores de usuários ou de palavras-chave; onde são recuperados os dados enquanto houver informação a ser consumida. Caso não tenha informação disponível para ambas as listas, o fluxo de informação é cortado e a conexão fechada. A segunda, por sua vez, é de uso predominante de empresas de publicidade e propaganda, sendo possível gerenciar campanhas de forma programada. As utilizadas neste trabalho são a Streaming API e Search API, que consistem em APIs de busca. Todas as informações recuperadas pelas APIs são de teor público, preservando assim a privacidade dos usuários.

3.2.4 YouTube Data API

A YouTube Data API⁵ fornece acesso a busca de vídeos, bem como estatísticas sobre os mesmos, como compartilhamentos, marcações de favoritos ou comentários realizados. Além dessas funcionalidades, a API do YouTube permite inserir, atualizar e excluir vídeos publicados na conta autenticada. Para o presente trabalho foi interessante a utilização de uma abordagem de recomendação audiovisual, pois muitos usuários podem preferir esses tipos de mídia para adquirir conhecimento. Dessa forma, a possibilidade de se realizar uma busca personalizada aos conteúdos de interesse através da filtragem de vídeos, cujo o conteúdo se enquadra na categoria de interesse educacional do usuário, e se mostra uma boa estratégia para recomendação de conteúdos audiovisuais.

3.2.5 DBpedia

A DBpedia é um projeto cujo foco é na conversão de conteúdo da Wikipédia em uma base de conhecimento estruturada, sendo possível a extração de conteúdos correlacionados.

Segundo Auer et al. (2007), são providos três mecanismos de acesso às bases de dados da DBpedia: Dados Ligados, protocolo SPARQL e por download da base em RDF. A forma utilizada neste trabalho é através do protocolo SPARQL.

Através de aplicações clientes a busca com o protocolo SPARQL é feita através

⁵<https://developers.google.com/youtube/v3/>

do endereço <http://dbpedia.org/sparql>, sendo uma interface mais apropriada quando a aplicação cliente sabe exatamente qual informação espera receber.

Na seção 3.3.5, serão mostradas as queries utilizadas para busca de recursos nessas bases.

3.2.6 Google Scholar

O Google Scholar ou Google Acadêmico⁶ é uma ferramenta de pesquisa da Google que auxilia na busca por recursos educacionais, como livros, artigos, jornais universitários, dentre outros tipos de obras.

Foi utilizado um módulo que provê classes e parser para a busca de recursos no Google Scholar. Esse módulo foi construído por Christian Kreibich que pode ser encontrado para download em sua página do GitHub⁷.

Esse módulo funciona com linha de comando Linux, sendo necessário realizar a tradução dos resultados para o padrão desejado. Essa tarefa foi implementada em uma classe na camada de Recomendação de Recursos, onde realiza a extração de URLs com índice de relevância maior.

3.3 Fluxo de Trabalho

Este trabalho segue a estrutura da arquitetura criada, obtendo os dados da rede social Twitter, através de suas APIs; identificando os principais candidatos a serem recomendados os recursos educacionais e analisando o por quê desses usuários serem difusores, tendo como base métricas de difusão. A Figura 3.4 apresenta o fluxo de trabalho realizado nesta pesquisa.



Figura 3.4: Fluxo de trabalho

⁶<https://scholar.google.com.br/>

⁷<https://github.com/ckreibich/scholar.py>

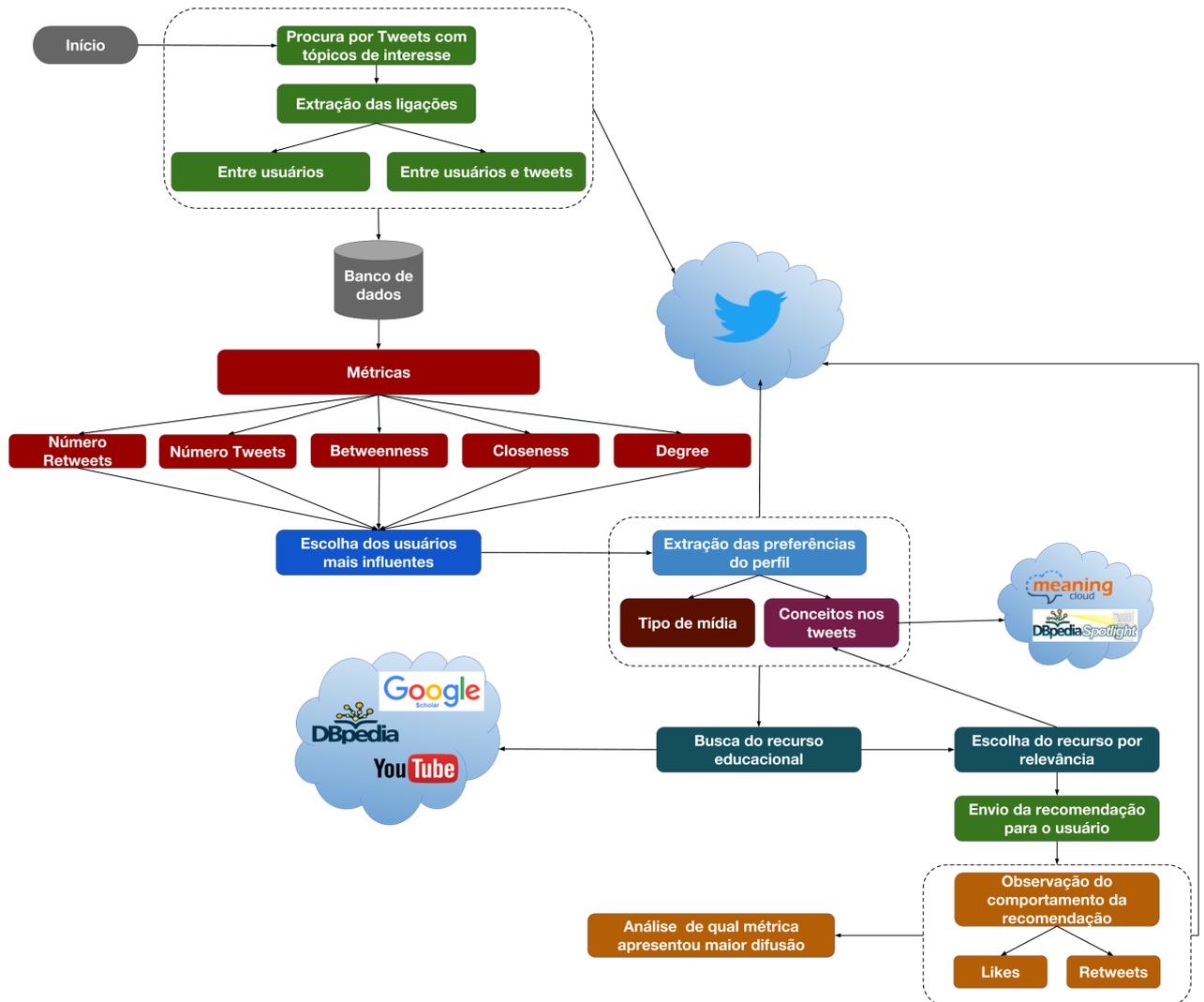


Figura 3.5: Workflow de desenvolvimento.

Primeiramente, é feita a coleta de parte das relações de perfis públicos do Twitter, onde serão escolhidos os usuários mais influentes a partir de métricas de difusão de informação. Para esses usuários é feita a análise de suas preferências educacionais, baseada nas publicações em seus perfis.

A partir da identificação das preferências educacionais dos usuários influentes, são procurados recursos aderentes à essas preferências em repositórios de recursos educacionais e, posteriormente, é feito o envio dos mesmos aos usuários e é analisado o comportamento de difusão que esses recursos sofrem no Twitter.

A Figura 3.5 apresenta a arquitetura de desenvolvimento utilizada neste trabalho, que será detalhada nas subseções seguintes.

3.3.1 Construção das Relações

A construção das relações ocorre ao mesmo tempo em que é realizada a busca dos dados (tweets) na rede social. Ao receber as respostas da API do Twitter, é feita a inserção das informações no banco de dados seguindo o modelo apresentado na Figura 3.6.

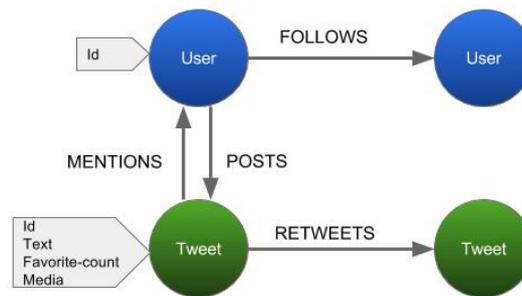


Figura 3.6: Modelagem do banco de dados em grafo.

Através dos parâmetros que podem ser utilizados nas chamadas da API, podemos selecionar o tipo de dado que queremos buscar, como, por exemplo, um tweet específico de algum usuário, tweets sobre algum assunto específico, podendo limitar tempo de postagem, região, mínimo de retweets que obteve, dentre outros parâmetros que auxiliam na busca por informação. As mensagens recebidas da API são em formato JSON, um formato bastante usual na manipulação de informação pela web.

Ao receber um tweet com o comportamento de retweet, uma propriedade está inclusa no objeto JSON obtido. Dessa forma, é possível verificar se é um tweet normal ou um retweet. Caso seja verificado que se trata de um retweet, é possível, dentro da estrutura do objeto, ver qual tweet foi retweetado. Assim, caso o mesmo já conste no banco, apenas é feita a ligação entre eles seguindo o modelo da Figura 3.7. Caso contrário, o tweet é adicionado e a ligação feita em seguida. Neste trabalho, foram escolhidos os temas “*data mining*” e “*data science*” para a busca dos recursos educacionais, os quais são bastante utilizados no meio em que esta pesquisa está inserida, motivando a utilização dos mesmos como base de estudo.

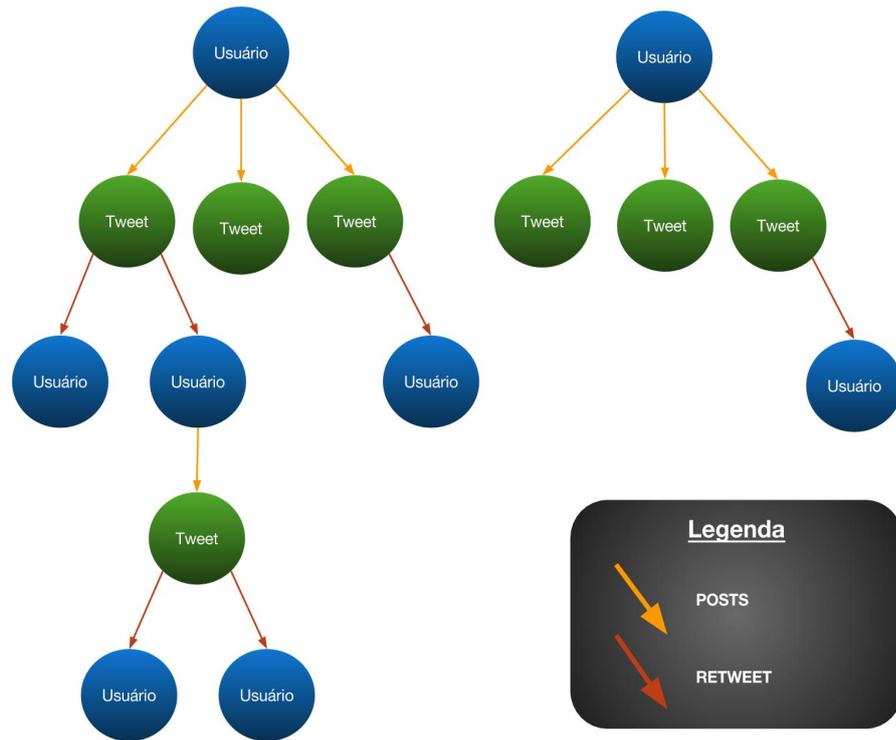


Figura 3.7: Comportamento de difusão no Twitter.

Os dados recuperados são tanto dos usuários que publicaram as mensagens, quanto as mensagens propriamente ditas. Assim, são selecionadas apenas as informações identificadoras de cada um e as relações entre ambos. As informações extraídas podem ser conferidas na Tabela 3.1.

Tabela 3.1: Informações extraídas do Twitter

Entidade	Informação
Usuário	id
Tweet	id
	text
	favorite_count
	media

3.3.2 Métricas

Para que o processo de análise de perfil e escolha dos usuários se realize, é necessário fazer uma avaliação quantitativa da importância do mesmo na rede social. Essa

avaliação é feita através de métricas de centralidade, presentes na teoria dos grafos, onde é possível ver a relação do nó com os demais da rede. Em conjunto a essas métricas, também é avaliada a quantidade de tweets realizados pelos usuários sobre os tópicos escolhidos para análise, que, como dito anteriormente, são “*data mining*” e “*data science*”. Por fim, também são avaliados o número de retweets que cada usuário sofreu nessas publicações, o que nos mostra quão influente é o usuário ou quão relevantes são as publicações do mesmo para os demais usuários da rede.

Os cálculos das métricas são feitos, em sua maioria, através das biblioteca descritas na seção 3.2.1.

Número de Tweets

Nesta métrica é realizada uma consulta simples ao banco de dados, onde são retornados ordenadamente os nós com maior número de publicações. A consulta através da linguagem Cypher pode ser conferida na Figura 3.8.

```
1 MATCH (k:User)-[r:POSTS]->(t) where toInt(k.id) is not null
2 return k.id as id, count(r) order by count(r) desc limit 6
```

Figura 3.8: Cypher query para recuperação dos ids de usuários que mais publicaram.

Número de Retweets(RT)

Esta métrica é semelhante à localizada acima, porém, nela são buscados os usuários que mais difundiram informação, ou seja, aqueles cuja quantidade de retweets é a mais expressiva na base. A consulta pode ser conferida na Figura 3.9.

```
1 MATCH p=(k:User)-[r:POSTS]->(t)<-[rt:RETWEETS]-() where toInt(k.id) is not null
2 with count(p) as ret, k match (k:User)-[r:POSTS]->(t)
3 return toString(k.id) as id, count(r), sum(ret) order by sum(ret) desc limit 6
```

Figura 3.9: Cypher query para recuperação dos ids de usuários que mais tiveram retweets.

Centralidade Betweenness

Essa centralidade é a responsável por dizer quão central é o nó em relação a todo

o grafo, sendo sua fórmula definida como a que segue:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Onde, σ_{st} é o total de caminhos mais curtos entre o nó s e o nó t . E $\sigma_{st}(v)$ é o número desses caminhos que passam pelo nó v (WASSERMAN; FAUST, 1994).

Centralidade Closeness

Esta, por sua vez, calcula quão próximo um indivíduo está para com os demais no grafo, ou seja, quão próximo um nó está dos demais.

$$C_C(x) = \frac{N}{\sum_y d(y,x)}$$

Onde, N é o número de vértices presentes no grafo e $d(y,x)$ é a distância entre os vértices y e x (WASSERMAN; FAUST, 1994).

Degree(Grau)

E, por fim, esta é a medida na qual calcula a quantidade de ligações que cada nó possui, sendo o conjunto dessas ligações chamado de grau.

$$C_D(v) = deg(v)$$

Onde, $deg(v)$ é o número de arestas presentes no vértice v (WASSERMAN; FAUST, 1994).

As análises de qual métrica se observa um melhor comportamento de difusão, estão na seção 4.1.

3.3.3 Escolha dos Usuários

Visando não recomendar para um número muito expressivo de pessoas de uma única vez, o que torna o processo de recomendação mais custoso, foi feita a seleção de 6 usuários para cada métrica descrita anteriormente. Os resultados das métricas são ordenados de forma decrescente, auxiliando a etapa de consolidação com os referentes às outras métricas.

A consolidação ocorre pela junção das listas geradas por cada métrica em uma única lista. Como cada métrica é calculada separadamente, pode coincidir de um usuário obter maior peso em diferentes métricas, sendo assim, para retirar qualquer possibilidade

de duplicação na junção das listas, é feito um mapeamento nas mesmas, retirando as duplicatas, porém preservando qual métrica a originou.

3.3.4 Extração de Preferências de Perfil

O trabalho visa a recomendação com base no modelo de filtragem baseada em conteúdo. Dessa forma, é necessário fazer um mapeamento das preferências do usuário a partir de seu perfil. Através da API do Twitter é possível extrair a *timeline* do usuário com seus 200 tweets mais recentes, quantidade máxima permitida pela API.

Como dito anteriormente, as respostas em formato JSON recebidas da API do Twitter possuem propriedades que permitem a verificação de imagens nas publicações. De acordo com essa informação, é possível quantificar a preferência de mídia do usuário. Foi considerado que, se o usuário possui em suas 200 últimas publicações maior número de publicações com presença de imagens ou vídeos, a recomendação que procederá será a do tipo vídeo. Caso contrário, é escolhida a opção de envio de recursos textuais.

Logo, são consolidados os tweets presentes na *timeline* do usuário, em um único texto, no qual será passado para a camada de recomendação de recursos, juntamente com a preferência de mídia do usuário, para a busca dos recursos mais relevantes.

Ao ser submetido o texto com os tweets juntamente com a preferência de mídia para a camada de recomendação de recursos, é necessário fazer a extração de conceitos e palavras-chave no texto recebido a fim de criar um perfil de interesse. Na arquitetura criada, existem duas possibilidades de extração de tópicos do texto. Uma é através da API MeaningCloud para extração de conceitos; e outro, é através da DBpedia Spotlight, uma ferramenta da DBpedia que realiza anotações de nomes de entidades presentes no texto, nas quais fazem referência a recursos encontrados na base da DBpedia.

MeaningCloud

MeaningCloud⁸ é uma empresa nova iorquina especializada em softwares de análise semântica. Possui diversas APIs para extração de informação de textos. Dentre elas, há extração de sentimentos, conceitos, entidades e classificação de texto. A utilizada por essa

⁸<https://www.meaningcloud.com/>

pesquisa é a API *Topic Extraction*⁹. Nela é possível fazer a extração de conceitos e entidades de textos, onde ordena por relevância os tópicos encontrados. Essa característica é importante, pois auxilia na busca de recursos educacionais, fornecendo tópicos como metadados nessas buscas.

No desenvolvimento foram apenas consideradas as entidades extraídas nas quais apresentam valores de relevância maiores ou iguais a 70%. Dessa forma, aumenta-se a probabilidade de relevância dos conteúdos, cujos metadados sejam equivalentes a essas entidades.

As APIs da MeaningCloud oferecem suporte a 6 diferentes idiomas como o português, inglês, francês, espanhol, italiano e catalão. Nesta pesquisa, foram submetidos textos apenas em inglês, pois o ambiente de pesquisa possui maiores contribuições nesta linguagem.

Essa API é utilizada na busca no YouTube e no Google Scholar, pois é a que oferece maior confiança na extração de conceitos. Como o YouTube possui uma base muito extensa de vídeos, que podem ser classificados como a pessoa desejar, torna difícil a busca correta pelo recurso, assim, utilizando um método mais confiável de palavras-chave na busca aumenta a probabilidade de acerto na recomendação.

Dessa forma ela caracteriza uma maneira mais completa de análise; porém, existe um limite de requisições mensais no plano gratuito, sendo somente de 40000 requisições. Mas, para finalidades desta pesquisa, a quantidade atendeu de forma adequada.

DBpedia Spotlight

A DBpedia Spotlight¹⁰ extrai as entidades presentes no texto baseada em um valor de confiança, que no caso da pesquisa optou-se pelo valor 0.7, buscando aumentar a relevância dos recursos encontrados. Ela faz anotações no texto submetido, relacionando os termos com os existentes na base da DBpedia. O valor de confiança representa o nível de especificidade que se quer realizar a extração, sendo que, quanto menor for a porcentagem submetida, termos mais genéricos serão recuperados, caso contrário, termos mais específicos perante o texto serão retornados. Essa API é utilizada na busca de

⁹<https://www.meaningcloud.com/developer/topics-extraction>

¹⁰<http://wiki.dbpedia.org/projects/dbpedia-spotlight>

conteúdos na DBpedia, pois já fornece links para os mesmos a partir do conceito extraído.

3.3.5 Busca de Recursos Educacionais

A busca de recursos educacionais é feita através do serviço que a camada Repositórios oferece, onde são buscados os recursos nos quais possuem metadados semelhantes aos conceitos extraídos previamente.

Para os recursos do tipo audiovisual, a busca é feita utilizando a YouTube Data API, onde são ranqueados os recursos por relevância automaticamente. Já para os recursos do tipo texto, é realizada a busca na base da DBpedia pelo protocolo SPARQL com o auxílio de duas queries SPARQL, que podem ser conferidas nas figuras 3.10 e 3.11; uma responsável por recuperar os termos que a DBpedia possui referentes aos conceitos extraídos e outra para a busca dos recursos utilizando esses termos.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
select ?b
where {
  {?subject skos:prefLabel ?object FILTER regex(?object, '%s', 'i') .
   ?subject skos:narrower ?a .
   ?a skos:prefLabel ?b }
  union
  {?subject skos:prefLabel ?b FILTER regex(?b, '%s' , 'i')}
}
```

Figura 3.10: Query de busca SPARQL na base da DBpedia para extração de termos. (PEREIRA, 2015)

```
PREFIX dbpedia-owl:<http://dbpedia.org/ontology/>
PREFIX dbpprop:<http://dbpedia.org/property/>
PREFIX dbpprop-rsc:<http://dbpedia.org/resource/>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX dcterms:<http://purl.org/dc/terms/>
PREFIX dc:<http://purl.org/dc/elements/1.1/>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
select ?t where{ <http://dbpedia.org/resource/%s> dbpedia-owl:wikiPageExternalLink ?t}
```

Figura 3.11: Query de busca SPARQL na base da DBpedia para recuperação dos recursos. (PEREIRA, 2015)

A variável %s presente em ambas as queries, é substituída pelo valor do conceito,

na primeira query, e para o termo recuperado da DBpedia, na segunda.

Relevância do recurso educacional

Segundo Pereira et al. (2015), existe uma grande dificuldade em priorizar os recursos recuperados da DBpedia, pois os mesmos não possuem uma catalogação, que auxiliem na determinação da relevância do recurso para o usuário. Para que ocorra um envio automatizado das recomendações, é necessário que se determine a relevância do recurso. Essa determinação acontece na camada Relevância do Recurso. Nela são buscados os recursos recuperados na camada de recomendação, através de suas URLs; eliminando os que não são mais alcançáveis, ou seja, aqueles que não estão mais disponíveis no endereço recebido. É feita a extração dos textos presentes no recurso online e, posteriormente, esse texto é submetido para a camada de extração de conceitos, visando observar a compatibilidade dos conceitos encontrados no texto, com os encontrados no perfil do usuário. A equação utilizada para determinar a relevância é a que segue:

$$Relevância(x, y) = \frac{|Conceitos(x) \cap Conceitos(y)| * 100}{|Conceitos(y)|}$$

Onde, $Conceitos(x)$ representa o conjunto de conceitos extraídos do texto do recurso, enquanto $Conceitos(y)$, representa o conjunto de conceitos extraídos do perfil do usuário.

O resultado dessa equação evidencia a relevância em porcentagem do recurso x para o usuário y ; sendo mais fácil e confiável, a recomendação automatizada com os recursos da DBpedia. Após calculadas as relevâncias para os recursos, é feito um ranqueamento dos mesmos por ordem de relevância decrescente e, são enviados para a camada de recomendação, somente os recursos mais relevantes. Essa camada é utilizada apenas para os recursos da DBpedia, pois os demais repositórios já fazem ranqueamento automático por relevância, dessa forma ele se torna opcional na arquitetura.

Caso a camada de relevância não encontre uma compatibilidade entre os conceitos do texto e do perfil do usuário, é feita a busca de outros recursos textuais, com os mesmos conceitos, agora, no repositório do Google Scholar, sendo que o mesmo já ordena por relevância os resultados, assim são submetidos estes para recomendação.

Há duas possibilidades de ocorrer a não recomendação de recursos textuais. Uma

na DBpedia, quando os recursos recuperados através dos conceitos, não obtiveram uma relevância maior que 0%, ou seja, os recursos não são relevantes para o usuário de forma alguma. E, a outra possibilidade é através da combinação dos conceitos no Google Scholar onde, uma combinação muito extensa de conceitos não retorna resultados. Dessa forma, para aumentar a probabilidade de encontrar recursos neste repositório, foi fixado a quantidade dos 8 primeiros conceitos extraídos do perfil do usuário, como meio de busca.

4 Avaliação

Para recomendar recursos educacionais, foram enviadas previamente às recomendações mensagens aos usuários considerados influentes. Caso eles seguissem o perfil recomendador¹¹ do projeto no Twitter, seria considerado que os mesmos aceitaram participar da pesquisa. Assim, as recomendações foram submetidas a apenas aos usuários que seguem o perfil recomendador, utilizando a abordagem descrita anteriormente.

Visando validar a arquitetura criada, foi pedido, como mostra a Figura 4.1, em uma primeira tentativa, aos usuários presentes na base, considerados difusores pelas métricas calculadas, para que fizessem parte do experimento seguindo o perfil recomendador.

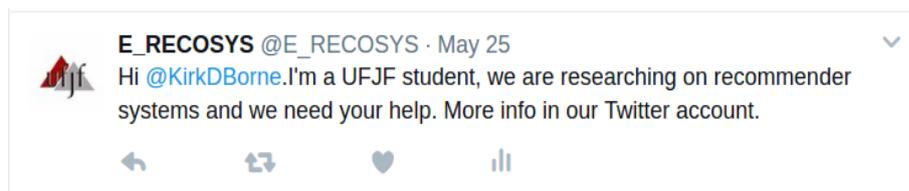


Figura 4.1: Primeira forma de pedido de colaboração estabelecido.

A quantidade de pedidos enviados foram 36. Como resultado, apenas 2 dessas 36 pessoas seguiram o perfil, mostrando interesse em receber conteúdos educacionais baseados em seus tweets. Um terceiro usuário, considerado um dos mais influentes na base, não seguiu o perfil E_RECOSYS, porém tweetou sobre a proposta que a pesquisa possui, recebendo 24 retweets e 15 curtidas (Figura 4.2), o que mostra que a proposta tem uma aceitação boa no meio de estudo escolhido. Porém, um usuário disse que verificou o site criado e que não via sentido na proposta, não seguindo o perfil recomendador. A mensagem pode ser conferida na Figura 4.3.

¹¹https://twitter.com/E_RECOSYS

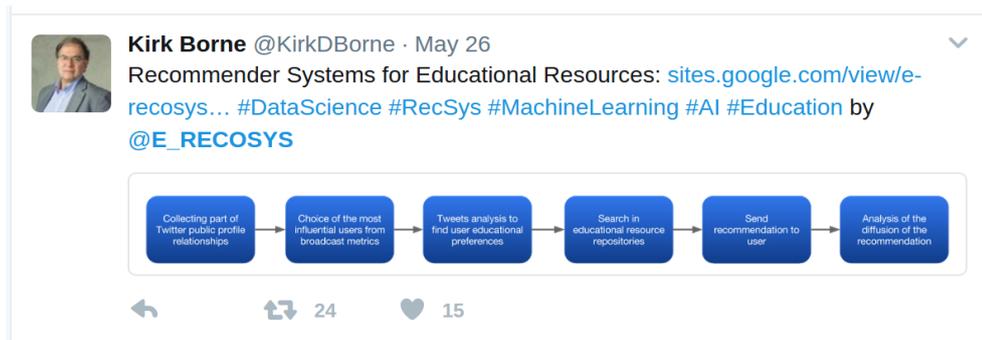


Figura 4.2: Tweet feito por um potencial influenciador identificado.

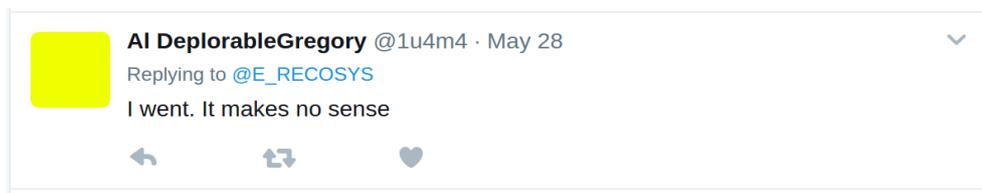


Figura 4.3: Resposta a uma requisição realizada.

Já outro usuário, não mostrou aversão à proposta, porém colocou empecilho, dizendo que somente seguiria o perfil, se o mesmo o seguisse primeiro, como pode ser visto na Figura 4.4. Dessa forma, optamos por adotar outra estratégia, em busca de conseguir novos seguidores para proceder com as recomendações. Optamos por seguir todo perfil que a arquitetura encontrasse como os mais influentes. Assim, em uma segunda tentativa, pedimos novamente aos usuários que nos seguissem, agora seguindo-os de volta. A forma de pedido para colaboração pode ser observado na Figura 4.5.

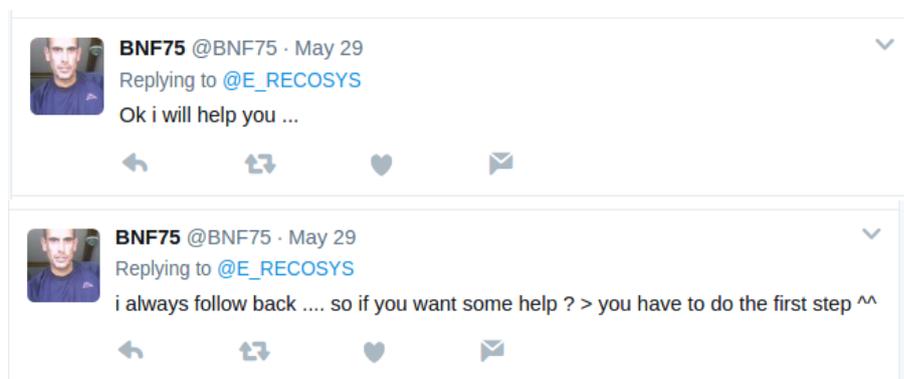


Figura 4.4: Resposta com restrição para seguir o perfil.



Figura 4.5: Nova forma de pedido para colaboração.

Nessa segunda tentativa, obtivemos um total de 12 perfis que nos seguiram de volta, dos 36 pedidos. A *home page* do perfil recomendador após os novos pedidos de colaboração pode ser vista na Figura 4.6. Dentre esses seguidores, se encontram o KirkD-Borne, que tweetou sobre o perfil recomendador e sua proposta, bem como BNF75, que pediu para segui-lo primeiro. Outros 10 usuários, por meio da influência dos que nos seguiram e por acesso ao perfil recomendador no Twitter, mostraram interesse em receber recomendações de recursos educacionais, seguindo o perfil recomendador. Mantendo a proposta inicial, a recomendação se procedeu para aqueles usuários considerados, pelas métricas, os mais difusores e que estavam seguindo o perfil.



Figura 4.6: Home page do perfil recomendador após novos pedidos de colaboração.

A Tabela 4.1 apresenta a quantidade de requisições feitas por métrica, englobando

as duas estratégias de pedido de colaboração, bem como o número de recomendações realizadas para cada uma delas. Por fim, apresenta a quantidade de retweets observados em recomendações feitas, caracterizando um processo de difusão de informação.

Tabela 4.1: Análise quantitativa por métrica

Métricas	Requisições	Recomendação	Difusão
Número de Retweets	12	8	43
Número de Tweets	12	0	0
Betweenness	12	1	0
Closeness	12	4	0
Degree	12	0	0

4.1 Análise das Recomendações

Como descrito em Kwak et al. (2010), retweets tendem a ocorrer dentro de 1 dia, sendo que conforme o tempo passa, as chances de um retweet acontecer diminui, devido a baixa visualização do mesmo na timeline do usuário, sobreposto por outros tweets posteriores. Portanto, para analisar a difusão dos recursos recomendados, esperou-se cerca de 2 dias para finalizar a análise após a recomendação. As recomendações ocorreram para os usuários da Tabela 4.2 seguindo as métricas utilizadas e atribuídas aos mesmos.

Tabela 4.2: Relacionamento de usuários e métricas

Usuário	Métrica
KirkDBorne	Número de Retweet
DBaker007	Número de Retweet
ipfconline1	Número de Retweet
DeepLearn007	Número de Retweet
Afolabibayor	Betweenness
AdjoaBawuah	Closeness
4politics2	Closeness

Não foram todos os usuários que seguiram o perfil recomendador, dessa forma não

foi possível avaliar as outras métricas no processo de difusão. E, além disso, para os outros 5 usuários, os quais não constam na tabela, mas que seguem o perfil recomendador, não foi possível recomendar, pois as informações extraídas dos perfis não foram significantes o suficiente para gerar conceitos e, para aqueles que obtiveram conceito, não foi encontrado conteúdos na DBpedia ou no Google Scholar relevantes o suficiente para o usuário, assim não se pôde proceder com as recomendações para esses usuários. A Tabela 4.3 apresenta os resultados gerais das recomendações por usuário.

Tabela 4.3: Resultados gerais dos tweets por usuário.

Usuário	Interações	Detalhes	Click Link	Ver Perfil	RT	Curtidas
KirkDBorne	20	2	7	9	0	1
DBaker007	7	2	2	1	0	1
ipfconline1	12	2	3	2	5	0
DeepLearn007	8	1	5	1	0	0
Afolabibayor	4	2	1	1	0	0
AdjoaBawuah	8	3	3	2	0	0
4politics2	15	4	6	4	0	0

A coluna Interações, mostra a quantidade de interações que os tweets com recomendações obtiveram. Essa quantidade está diretamente ligada a popularidade e nível de pesquisa, que o usuário recomendado possui. A coluna Detalhes, apresenta a quantidade de vezes que os tweets foram visualizados de maneira destacada, ou seja, obtiveram interesse, por parte dos usuários, à visualizar suas informações gerais. Como os tweets continham links para os recursos recomendados, a quantidade de vezes que esses links foram clicados é apresentada na coluna Click Link.

A coluna Ver Perfil, por sua vez, apresenta a quantidade de vezes que o perfil recomendador é acessado através da recomendação. As colunas RT e Curtidas, apresentam a quantidade de retweets que as recomendações, feitas aos usuários, obtiveram e, seu número de curtidas, respectivamente.

Das recomendações feitas, apenas o recurso recomendado ao usuário KirkDBorne gerou um *cascade*, não automático, da recomendação. Para outros usuários, como DBa-

ker007, obtivemos um *feedback* através de curtidas ou respostas à recomendação, como pode ser observado na Figura 4.7.



Figura 4.7: *Feedback* à recomendação.

Observamos que sempre que o usuário ipfconline1 era mencionado, 2 perfis retweetavam a menção, como pode ser observado na Figura 4.8, dessa forma apresenta um comportamento automático desses perfis, onde os mesmos possuem o papel de difusores de informação, difundindo informação a partir de um determinado tema ou de determinado usuário, como neste caso. Porém, o usuário recomendado não demonstrou *feedback* para a recomendação.



Figura 4.8: Comportamento automático de retweet.

A Figura 4.10, apresenta o resultado do processo de difusão, tanto para o *post* apresentado na Figura 4.2, como para o apresentado na Figura 4.9.

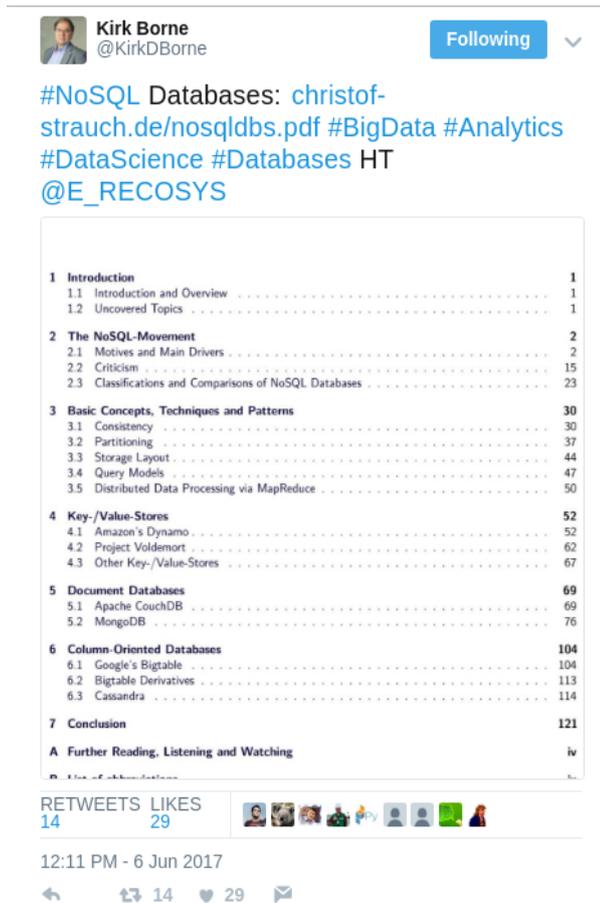


Figura 4.9: Post feito pelo usuário com o recurso recomendado.

Na Figura 4.10a, não consta todos os retweets que a publicação recebeu, pois alguns dos retweets, tiveram a identificação de seu publicante bloqueada pelo mesmo. Mas a quantidade total de retweets pode ser conferida na Figura 4.2.

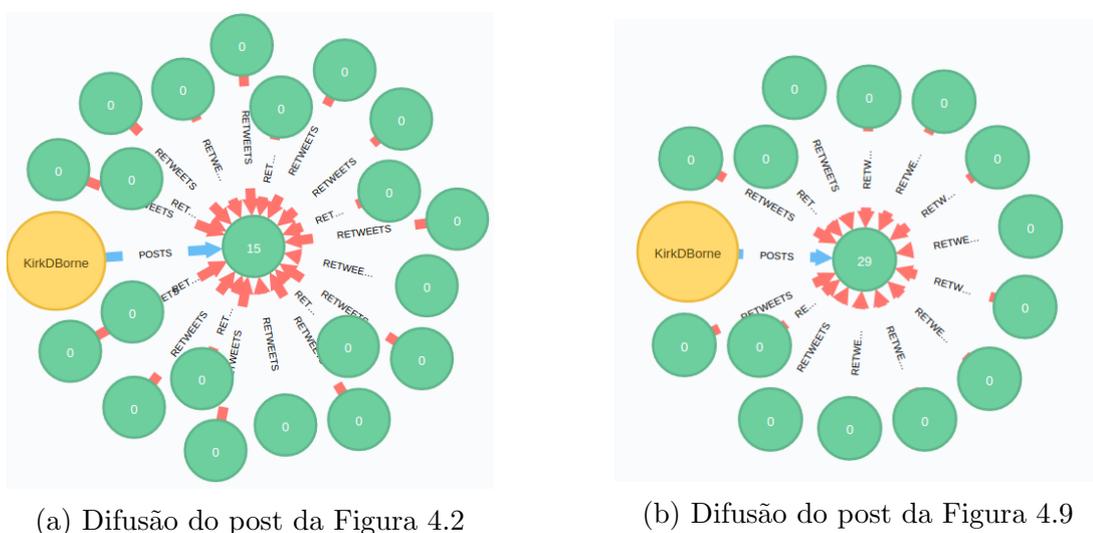


Figura 4.10: Resultado do processo de difusão.

Como citado em Kwak et al. (2010), o processo de *cascade* não ocorre de maneira muito profunda no Twitter, ou seja, não gera árvores de difusão com alturas muito extensas. Das alturas analisadas nos perfis considerados difusores, o valor máximo encontrado foi 2 níveis de *cascade*. Porém, como é mostrado em Tong et al. (2016), o *cascade* pode ocorrer de várias formas, bem como a influência de um nó para com outro. Na rede social trabalhada, percebeu-se que a difusão, na maioria das vezes, ocorre horizontalmente, em apenas um nível de *cascade*, porém a quantidade de pessoas que são atingidas é muito grande, pois todos os seguidores do usuário que postou o conteúdo, têm acesso àquela informação.

Assim, dentre as métricas que foram possíveis de serem analisadas, as que apresentaram melhores resultados para o encontro de potenciais difusores, foram as quantitativas, ou seja, pelo número de postagens e pelo número de retweets que o perfil possui, onde a forma de avaliação se deu pela dimensão de *cascade* obtido através da recomendação, sendo os usuários que foram os propulsores da difusão, aqueles que estavam presentes em ambas as métricas.

Percebeu-se que a influência e o peso do usuário na rede, quando visualizado o perfil do mesmo, normalmente as variáveis que são levadas em consideração pelos outros usuários do Twitter são o número de seguidores e a quantidade de postagens que o perfil publica sobre uma determinada área ou assunto. Conforme as recomendações foram ocorrendo e as pessoas vinham a ter acesso às mesmas, outros usuários começaram a seguir o perfil recomendador, decorrente do peso e poder de influência do perfil recomendado. Alguns desses usuários deixaram de seguir o perfil posteriormente, quando viram que não foram seguidos de volta, devido a política que criamos de seguir somente os perfis considerados difusores de informação.

4.2 Considerações Finais do Capítulo

Pôde-se perceber que os usuários no Twitter estão agrupados por assuntos de interesse, como o caso desta pesquisa, que se concentrou em buscar usuários relacionados aos termos “*data mining*” e “*data science*”. Esse comportamento de grupo faz com que a difusão ocorra mais horizontalmente, atingindo mais diretamente quem segue o perfil que

esteja relacionado ao assunto de interesse, sendo mais difícil observar difusões verticais de informação. Isso é devido ao fato de que muitos dos seguidores dos perfis, presentes neste grupo, não possuem interesse naquele determinado assunto, logo, não compartilham as informações que estes grupos de usuários publicam.

A validação das métricas, por sua vez, obteve um bom resultado dentro das possibilidades de análise, em vista que não foram todos os usuários requisitados que seguiram o perfil recomendador e, em alguns perfis, não foi possível extrair preferências educacionais. A busca por recursos, demonstrou um bom funcionamento, porém, demonstra pontos que devem ser melhorados em prol de uma recomendação mais rica em relevância para os usuários, como uma melhor combinação de conceitos para pesquisa no repositório do Google Scholar e, uma forma mais refinada de calcular a relevância dos recursos, para aqueles extraídos da DBpedia, independentemente de formato.

5 Considerações Finais

O presente trabalho, buscou apresentar uma nova proposta de inserção de sistemas de recomendação de recursos educacionais em redes sociais, fora de ambientes formais de Educação, concentrando as recomendações a perfis considerados influentes e difusores de informação, através de métricas de difusão.

O processo de recomendação de recursos educacionais foi realizado conforme arquitetura criada, apresentada na seção 3.1, onde é feita a análise do perfil do usuário, buscando extrair as preferências educacionais do mesmo, de modo a realizar a recomendação de recursos que mais se enquadrem aos seus interesses. Com o experimento, a arquitetura foi validada com sucesso, onde se obteve resultados favoráveis a proposta. A proposta obteve uma boa aceitação no meio de pesquisa escolhido, onde 22 pessoas demonstraram interesse em receber conteúdos educacionais baseados nos seus interesses.

Dentre as métricas avaliadas, as que apresentaram melhor eficácia foram as quantitativas, sendo expressa pelo número de retweets e número de publicações feitas. Esse fato é devido ao peso do perfil, sendo visualizado de forma influenciadora por outros usuários que visitem seu perfil. Quando esse usuário influente possui uma frequência de postagens de conteúdos de interesse, há uma maior chance de que as informações postadas sofram difusão pelos seguidores do usuário.

A partir das análises e recomendações feitas, foi observado o potencial da pesquisa em atingir outras massas de usuários, tendo em vista como ponto de partida a recomendação para grupo de usuários com interesses educacionais em comum, onde, com a difusão de informação e pela influência dos difusores presentes neste grupo, outros perfis demonstraram interesse em receber conteúdo educacional. Dessa forma, a rede social do sistema desenvolvido, cresceria de forma mais heterogênea, possibilitando abordar outros interesses educacionais, distintos do grupo inicial de usuários. Das recomendações realizadas apenas três sofreram difusão, sendo duas de forma automática. Dessa forma, há a necessidade de se aprimorar a busca de conteúdo relevante nos repositórios utilizados, fazendo uma melhor combinação dos conceitos extraídos dos perfis dos usuários.

Outras possibilidades foram identificadas ao longo do desenvolvimento da pesquisa, como a recomendação baseada em predição de difusão temporal, onde encontraria os usuários mais propensos a difundir informação, em uma escala de tempo maior; e a realização de cálculos para predição de quais nós são mais propensos a serem atingidos por uma recomendação, feita a outro usuário. Além da possibilidade de utilização de uma abordagem de recomendação colaborativa, analisando os recursos recomendados para usuários similares aos que se quer recomendar conteúdo.

Bibliografia

- ALMEIDA, R. et al. Sistemas de recomendação de recursos educacionais para grupos de redes sociais: um mapeamento sistemático. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2015. v. 26, n. 1, p. 1022.
- ALMEIDA, R. et al. Recomendação de recursos educacionais para grupos: buscando soluções em redes sociais. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2016. v. 27, n. 1, p. 996.
- AUER, S. et al. Dbpedia: A nucleus for a web of open data. *The semantic web*, Springer, p. 722–735, 2007.
- BURKE, R. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, Springer, v. 12, n. 4, p. 331–370, 2002.
- CASAGRANDE, M. F. R.; KOZIMA, G.; WILLRICH, R. Técnica de recomendação baseada em metadados para repositórios digitais voltados ao ensino. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2013. v. 24, n. 1, p. 677.
- CASCAIS, M. d. G. A.; FACHÍN-TERÁN, A. Educação formal, informal e não formal em ciências: Contribuições dos diversos espaços educativos. *Novas perspectivas de ensino de Ciências em espaços não formais amazônicos*. Manaus, AM: UEA Edições, 2013.
- COSTA, E.; AGUIAR, J.; MAGALHÃES, J. Sistemas de recomendação de recursos educacionais: conceitos, técnicas e aplicações. *Jornada de Atualização em Informática na Educação*, v. 1, n. 1, 2013.
- EMIRBAYER, M.; GOODWIN, J. Network analysis, culture, and the problem of agency. *American journal of sociology*, JSTOR, p. 1411–1454, 1994.
- GADOTTI, M. A questão da educação formal/não-formal. *Sion: Institut Internacional des Droits de 1^o Enfant*, p. 1–11, 2005.
- GASPAR, A. A educação formal e a educação informal em ciências. *Massarani L., Moreira IC, Brito F. orgs*, p. 171–183, 2002.
- GOHN, M. d. G. Educação não-formal, participação da sociedade civil e estruturas colegiadas nas escolas. *Rio de Janeiro: Revista Ensaio-Avaliação e Políticas Públicas em Educação*, SciELO Brasil, v. 14, n. 50, p. 11–25, 2006.
- GUILLE, A.; HACID, H.; FAVRE, C. Predicting the temporal dynamics of information diffusion in social networks. *arXiv preprint arXiv:1302.5235*, 2013.
- HARGIS, J. Can students learn science using the internet? *Journal of Research on Computing in Education*, Taylor & Francis, v. 33, n. 4, p. 475–487, 2001.
- HELLMANN, T.; STAUDIGL, M. Evolution of social networks. *European Journal of Operational Research*, Elsevier, v. 234, n. 3, p. 583–596, 2014.

- KEMPE, D.; KLEINBERG, J.; TARDOS, É. Maximizing the spread of influence through a social network. In: ACM. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.], 2003. p. 137–146.
- KWAK, H. et al. What is twitter, a social network or a news media? In: ACM. *Proceedings of the 19th international conference on World wide web*. [S.l.], 2010. p. 591–600.
- MARCHIORI, P. Z. Bibliotecas digitais e repositórios de objetos de aprendizagem. *Informação & Sociedade*, Universidade Federal da Paraíba-Programa de Pós-Graduação em Ciência da Informação, v. 22, n. 2, 2012.
- MARTELETO, R. M. Análise de redes sociais: aplicação nos estudos de transferência da informação. *Ciência da informação*, SciELO Brasil, v. 30, n. 1, p. 71–81, 2001.
- MCMAHON, S. M.; MILLER, K. H.; DRAKE, J. Networking tips for social scientists and ecologists. *Science*, American Association for the Advancement of Science, v. 293, n. 5535, p. 1604–1605, 2001.
- PEREIRA, C. K. *Extração de Características de Perfil e de Contexto em Redes Sociais para Recomendação de Recursos Educacionais*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, 2015.
- PEREIRA, C. K. et al. Extração de características de perfil e de contexto em redes sociais para recomendação de recursos educacionais. *Revista Brasileira de Informática na Educação*, v. 23, n. 3, 2015.
- REATEGUI, E. B.; CAZELLA, S. C. Sistemas de recomendação. In: CITESEER. *XXV Congresso da Sociedade Brasileira de Computação*. [S.l.], 2005. p. 306–348.
- RIBEIRO, F. A. A.; FONSECA, L. C. C.; FREITAS, M. de S. Recomendando objetos de aprendizagem a partir das hashtags postadas no moodle. In: *Anais do Simpósio Brasileiro de Informática na Educação*. [S.l.: s.n.], 2013. v. 24, n. 1, p. 82.
- TANG, T.; MCCALLA, G. Beyond learners’ interest: personalized paper recommendation based on their pedagogical features for an e-learning system. In: SPRINGER. *Pacific Rim International Conference on Artificial Intelligence*. [S.l.], 2004. p. 301–310.
- TAXIDOU, I.; FISCHER, P. M. Online analysis of information diffusion in twitter. In: ACM. *Proceedings of the 23rd International Conference on World Wide Web*. [S.l.], 2014. p. 1313–1318.
- TONG, C. et al. A novel information cascade model in online social networks. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 444, p. 297–310, 2016.
- WANG, C.; CHEN, W.; WANG, Y. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, Springer Science & Business Media, v. 25, n. 3, p. 545, 2012.
- WASSERMAN, S.; FAUST, K. *Social network analysis: Methods and applications*. [S.l.]: Cambridge university press, 1994. v. 8.
- YU, S. J. The dynamic competitive recommendation algorithm in social network services. *Information Sciences*, Elsevier, v. 187, p. 1–14, 2012.