

Universidade Federal de Pernambuco

Graduação em Ciência da Computação

Centro de Informática

2011.2

Recomendação de usuários baseada na topologia

Trabalho de Graduação

Aluno: Tiago Ferreira Lima (tfl2@cin.ufpe.br)

Orientador: Tsang Ing Ren (tir@cin.ufpe.br)

Recife, Dezembro de 2011

Universidade Federal de Pernambuco

Graduação em Ciência da Computação

Centro de Informática

2011.2

Recomendação de usuários baseada na topologia

Trabalho de Graduação

Monografia apresentada ao Centro de Informática da Universidade Federal de Pernambuco como requisito parcial da obtenção do grau de Bacharel em Ciência da Computação. Orientador Prof. Tsang Ing-Ren (Ph.D.)

Recife, Dezembro de 2011

Universidade Federal de Pernambuco

Graduação em Ciência da Computação

Centro de Informática

2011.2

Recomendação de usuários baseada na topologia

Trabalho de Graduação

Monografia apresentada ao Centro de Informática da Universidade Federal de Pernambuco como requisito parcial da obtenção do grau de Bacharel em Ciência da Computação. Orientador Prof. Tsang Ing-Ren (Ph.D.)

Aprovada em

BANCA EXAMINADORA

Tsang Ing-Ren (Ph.D.)

Universidade Federal de Pernambuco – UFPE

George Darmiton Cavalcanti (Ph.D.)

Universidade Federal de Pernambuco – UFPE

Recife, Dezembro de 2011

Agradecimentos

Meus agradecimentos por este trabalho e por toda a minha graduação vão, primeiramente, aos meus pais por tudo que eles fizeram por mim, não há palavras para descrever tal esforço e dedicação. Aos meus irmãos, a minha vó e os meus familiares, que estiveram sempre ao meu lado dando conselhos e me incentivando de alguma maneira, principalmente me aliviando do estresse do dia-a-dia. A minha namorada Juliana, que está sempre ao meu lado tomando decisões junto comigo e sem ela provavelmente não teria concluído o curso.

Em segundo lugar, ao meu orientador, que o apoio e o entusiasmo me deram motivação em momentos que este trabalho não parecia ter sido uma boa ideia.

Agradeço aos meus amigos que contribuíram de alguma forma com a minha graduação e com este trabalho, em especial para os que leram dezenas de vezes este trabalho me dando sugestões de melhoria.

A todos citados acima agradeço por me aguentarem e por terem de alguma forma me ajudado.

Aos meus pais, Emanuel e Ivanete.

Não tentes ser bem sucedido, tenta antes ser um homem de valor.
(Albert Einstein)

Resumo

Construir relacionamentos é um dos tópicos mais importantes em um Site de Rede Social, e ao considerar um Site de Rede Social Educacional é ainda mais relevante. Isto porque são nos relacionamentos que a colaboração acontece e com isso a construção do conhecimento se concretiza.

O presente trabalho analisa a viabilidade e propõe a construção de um mecanismo de recomendação de usuários para uma Rede Social Educacional, o qual se utiliza da unicidade da topologia da Rede Social Educacional e teorias de redes complexas. Os resultados obtidos demonstram que existe potencial no uso da especificidade de uma rede para melhorar a eficiência de um mecanismo de recomendação.

Palavras-chave: rede social educacional, sistema de recomendação de usuários, redes complexas, topologia, especificidade.

Abstract

Build relationships is one of the most important topics in a Social Network Site, and when considering a Social Network Educational Site it is even more relevant. This is because it is within relationships that collaboration happens and with it the construction of knowledge is concretized.

This paper analyzes feasibility and proposes the construction of a mechanism of recommendation of users to a Social Network Educational Site, which uses the uniqueness of the Social Network Educational Site's topology and theories of complex networks. Results obtained show that there is potential in the use of the specificity of a network to improve the efficiency of recommendation engine.

Keywords: social educational network, recommendation system of users, complex networks, topology, specificity

Lista de Ilustrações

Figura 1 - Perfil de um usuário no Redu.....	6
Figura 2 - Hierarquia detalha do Redu.....	8
Figura 3 - Relacionamento entre os amigos em comum.....	15
Figura 4 - Relacionamentos entre os amigos dos usuários n_i e n_j	16
Figura 5 - Representação gráfica dos valores da função de avaliação com o passar das iterações	17
Figura 6 - Fluxo do mecanismo de recomendação, fase de cálculo dos índices e o começo da segunda fase	20

Lista de Tabelas

Tabela 1 - Resultados do experimento da comparação de Julho e Dezembro de 2011	23
Tabela 2 - Resultados obtidos no segundo experimento.....	24

Sumário

1. Introdução	1
1.1 Motivação	2
1.2 Objetivo.....	2
1.3 Próximas Sessões	3
2. Rede Social	4
2.1 Rede Social Educacional.....	5
2.2 Redu	6
3. Sistema de Recomendação.....	9
3.1 Algoritmos de recomendação de usuários	9
3.1.1 Content Matching.....	9
3.1.2 Content-plus-link	10
3.1.3 Friends-of-Friends (FoF)	11
3.1.4 SONAR	12
3.1.5 Algoritmo de Silva.....	13
3.2 Conclusões	18
4. Sistema de Recomendação para uma Rede Social Educacional.....	19
4.1 Considerações Iniciais	19
4.2 Algoritmo.....	19
4.3 Experimento e Resultado	22

5. Conclusão.....	25
Referências Bibliográficas	26

1. Introdução

Com a quantidade de usuários que utilizam a Internet crescendo em ritmo acelerado a uma taxa de 14% ao ano, chegando a 46,3 milhões de usuários¹ no Brasil. Cresce também a quantidade de usuários que utilizam Sites de Redes Sociais. No Brasil, esse número chega a 42 milhões de usuários, o que corresponde a cerca de 91% de usuários da Internet brasileira.

Segundo Ahn *et al* (2007), um Site de Rede Social (SRS) é um serviço que fornece um espaço online privado para os indivíduos e ferramentas para interação com outras pessoas na Internet. O que torna os SRSs únicos e de grande atrativo não é por apenas permitir que indivíduos conheçam estranhos, mas permitir que os usuários articulem e tornem visíveis suas redes sociais (boy e Ellison 2007). Incentivar a criação de relacionamentos nas redes sociais é de fundamental importância, pois é neste ponto que o crescimento da Rede Social virtual se concretiza e as trocas de experiências crescem.

No contexto educacional, existe a *Computer-supported collaborative learning (CSCL)*, que é um ramo emergente das ciências da aprendizagem que estuda como as pessoas podem aprender em grupo com o auxílio do computador (Stahl, Koschmann e Suthers 2006). Com a CSCL, cresceu-se uma tendência de se utilizar um Site de Rede Social para o ensino, como Cho *et al* (2009) explicita, rede social é um elemento central em ambientes de aprendizagem colaborativa. Da perspectiva de uma rede social, aprendizagem é um resultado social e coletivo alcançado através de conversas sem sentido, práticas compartilhadas e as conexões sociais. Esta tendência vem crescendo cada vez mais, dando a origem a Redes Sociais Educacionais, tais como o Redu².

¹ <http://www.cgi.br/publicacoes/pesquisas/govbr/cgibr-nicbr-censoweb-govbr-2010.pdf>

² <http://ww.redu.com.br>

1.1 Motivação

Redes Sociais Educacionais utilizam o compartilhamento como fundamental premissa para desenvolvimento e criação do conhecimento. De acordo com Cho *et al* (2009), conhecimento não é um objeto estático adquirido por um usuário e sim algo construído através de colaboração de experiências de muitos usuários em um ambiente de aprendizagem. Deste modo, a criação de relacionamentos tem que ser a maior possível, para que o alcance da colaboração aumente.

Para criar relacionamentos, os SRSs utilizam-se de várias abordagens: o usuário pode navegar através do site pela lista de amigos dos seus amigos, através de mecanismos de buscas ou por meio do catálogo de usuários. Todas estas abordagens funcionam em um escopo pequeno de usuários, mas quando o escopo aumenta para milhares ou milhões de usuários, a busca e o catálogo de usuários se tornam ineficientes pela quantidade de casos repetidos, enquanto que navegar na lista de amigos torna-se demorado e tedioso.

Outra abordagem já usada em várias SRSs como Facebook, Orkut, Twitter é um Sistema de Recomendação de Usuários (SRU), que é um mecanismo que busca a recomendação de conexões que seja do interesse do usuário. Tal abordagem pode ser melhorada se aplicada de forma a explorar mais as especificidades de cada tipo de Rede Social, assim como Chen *et al* (2009) fez ao aplicar uma SRU em uma rede social corporativa da IBM obtendo um mecanismo de recomendação mais eficiente.

1.2 Objetivo

O objetivo do presente trabalho é desenvolver um sistema de recomendação de usuários baseando-se na topologia de uma Rede Social Educacional (Redu). Será utilizada a estrutura hierárquica que o Redu oferece, para se construir uma rede diferente das outras com intuito de tornar o mecanismo de recomendação mais eficiente. Espera-se que o mecanismo de recomendação facilite o usuário a encontrar contatos e com isso aumente a quantidade de relacionamentos consideravelmente, pois isso promoverá a construção de conhecimento através da comunicação entre

pessoas conectadas. Esta é uma importante estratégia para ajudar pessoas a se engajar em uma aprendizagem colaborativa (Melo 2010).

1.3 Próximas Sessões

No próximo capítulo será definida e descrita uma Rede Social Educacional, com intuito de se entender mais um tipo de Rede Social. No capítulo 3 será apresentada uma revisão de Sistemas de Recomendação, em que alguns algoritmos e suas respectivas características serão apresentados. Após a revisão, será proposto um novo mecanismo de recomendação para o Redu, neste mesmo capítulo uma avaliação será feita comparando o mecanismo proposto com técnicas já consagradas. Por fim, será apresentada a conclusão do presente trabalho, com intuito de se avaliar as lições aprendidas e mostrar futuras melhorias.

2. Rede Social

De acordo com Boyd e Ellison (2010), em tradução livre: “sites de redes sociais são serviços baseados na web que permitem aos indivíduos (1) construir um perfil público ou semi-público dentro de um sistema limitado, (2) articular uma lista de outros usuários com quem partilham uma conexão, e (3) ver e percorrer sua lista de conexões e aquelas feitas por outros dentro do sistema. A natureza e a nomenclatura destas conexões variam de site para site”. Outros termos utilizados para caracterizar serviços como estes são Tecnologias Sociais Digital, Mídia Participativa e Mídia Social (Greenhow 2009).

Estes sites de redes sociais podem refletir círculos sociais já existentes ou a criação de novas redes sociais. Por exemplo, LinkedIn¹ é uma rede social direcionada para profissionais e a base social por trás do LinkedIn é uma rede social formada por colegas de trabalho ou clientes, ou seja não é uma rede social previamente existente. Enquanto que Redes Sociais como Facebook² e MySpace³, possuem características que permitem ao usuário manter seus laços sociais já existentes (Boyd e Ellison 2010).

Após entrarem em um Site de Rede Social, usuários são solicitados a identificarem outros que estão no sistema e que possuem um relacionamento. A legenda para o relacionamento varia dependendo do site - termos populares incluem “Amigos”, “Contatos”, e “Fãs”. A maioria dos SRSs precisam de confirmação bidirecional para estabelecimento de um relacionamento, mas algumas não. As ligações unidirecionais são algumas vezes chamadas de “Fãs” ou “Seguidores”, mas outras vezes são chamadas simplesmente de "Amizades" (Boyd e Ellison 2010).

¹ <http://www.linkedin.com>

² <http://www.facebook.com>

³ <http://www.myspace.com>

2.1 Rede Social Educacional

Rede Social Educacional é uma rede social que possui características de um Sistema de Gestão de Aprendizagem (SGA). SGAs são softwares desenvolvidos com intuito de prover ensino e aprendizagem virtual ou semipresencial. De forma geral, uma Rede Social Educacional é um *Computer-supported collaborative learning (CSCL)*, que é um ramo da ciência da aprendizagem que estuda como as pessoas podem aprender com a ajuda dos computadores (Cho *et al*, 2009).

A Rede Social foi incluída no processo da aprendizagem, pois segundo Cho *et al* (2009) é um elemento central em um ambiente de aprendizagem colaborativa. Esta aprendizagem é um resultado social e coletivo alcançado através de conversas sem sentido, práticas compartilhadas, e as redes de conexões sociais. Conhecimento gerado pela aprendizagem, neste sentido, não é um objeto estático adquirido por um único indivíduo, mas é ativamente co-construído através de contínuas trocas sociais e colaborações entre alunos inseridos em redes sociais (Cho *et al*, 2009).

Para o usuário, a quantidade de relacionamentos é um indicativo de seu grau de poder, atividade e visibilidade, e ainda é positivamente associado ao desempenho individual. Isto ocorre, pois o usuário que tem alto grau de relacionamentos, ou de centralidade, tem acesso a muitos recursos e isto pode ser importante de diversas maneiras (Cho *et al*, 2009), como por exemplo maior acesso a informação e maior poder de disseminação de conteúdos.

Nota-se que a criação de relacionamentos é um ponto fundamental em qualquer Rede Social, mas no tocante de uma Rede Social Educacional é ainda mais importante. A troca de experiência entre pares é fundamental para construção do conhecimento e a difusão do mesmo. Neste contexto, facilitar a criação de relacionamentos é essencial para o processo de aprendizagem.

2.2 Redu

O Redu¹ é uma rede social educacional criada por alunos do Centro de Informática² (CIn) da Universidade Federal de Pernambuco³ (UFPE). Seu objetivo é promover a aprendizagem através de colaboração e trocas sociais, ou seja, utilizar o ambiente de uma rede social para facilitar trocas e construção do conhecimento.

Tiago Lima

Engenheiro de Software em Redu
Bacharelado em Ciência da Computação pela UFPE
Ensino Médio em CPI - Colégio Preparatório Integrado
22 anos
Fala Português, Inglês
Natural de Recife, Pernambuco
Vive em Recife, Pernambuco

Sobre mim: Sou estudante de Ciência da Computação, trabalho no Redu e gosto de mexer em coisas de tecnologia.

Interesses: Desenvolvimento Web, IA, Linguagens de Programação

Celular: +55 (88) 9928-4017

<http://www.facebook.com/filtiago>
<http://twitter.com/filtiago>
<http://www.linkedin.com/pub/tiago-ferreira/21/800/b31>

Experiência: **Estagiário** em **Ávila Soluções**
July 2010 - December 2010 (5 meses)
Desenvolvia com as linguagens C#, VB. Utilizava SQL Server 2008 para manipulação de dados.

Engenheiro de Software em Redu
July 2010 - Atual (mais de 1 ano)
Construção da plataforma Redu para educação.

Educação: **Bacharelado em Ciência da Computação** pela UFPE
2011

Ensino Médio em CPI - Colégio Preparatório Integrado
2006

Cursos Inscritos: 18

- BB
- Introdução Aluno
- Programação Web
- Padrões Web Aluno
- ER-FCAP Escola do Recife
- Primeiro Ano do Ensino Médio 2011 Aluno
- Redu Dev
- Internationalization Administrador do ambiente
- Python Administrador do ambiente
- Nova Hierarquia Administrador do ambiente

Ver mais
Mostrando os 4 últimos ambientes de Tiago

Figura 1 - Perfil de um usuário no Redu

Por ser um SRS, o Redu apresenta a entidade usuário que possui um perfil associado, relacionamentos entre usuários e lista de amigos. Como mostra a Figura 1, o usuário pode cadas-

¹ <http://www.redu.com.br>

² <http://www.cin.ufpe.br>

³ <http://www.ufpe.br>

trar diversos tipos de informações, como: uma breve descrição sobre sua personalidade, interesses, celular, links para outras redes sociais, experiências profissionais, idiomas, nacionalidade e histórico educacional. Também pode ser observado na Figura 1, a lista dos contatos e a lista dos cursos que usuário participa.

Vendo do ponto de vista educacional, o Redu é dividido em Ambientes Virtuais de Aprendizagem (AVA), sendo uma unidade macro tal qual uma universidade (e.g UFPE). Uma universidade é composta de diversos cursos (e.g Ciência da Computação) os quais irão possuir diferentes disciplinas (e.g Aprendizagem de Máquina). Estas disciplinas contêm módulos de conteúdos; então, para a disciplina de Aprendizagem de Máquina, um módulo existente seria o de Classificadores que possuiria uma aula sobre o algoritmo k -Nearest Neighbor. Da maneira análoga é constituída uma AVA, com um conjunto de cursos que possuem diversas disciplinas. A Figura 02 ilustra toda hierarquia presente no Redu.

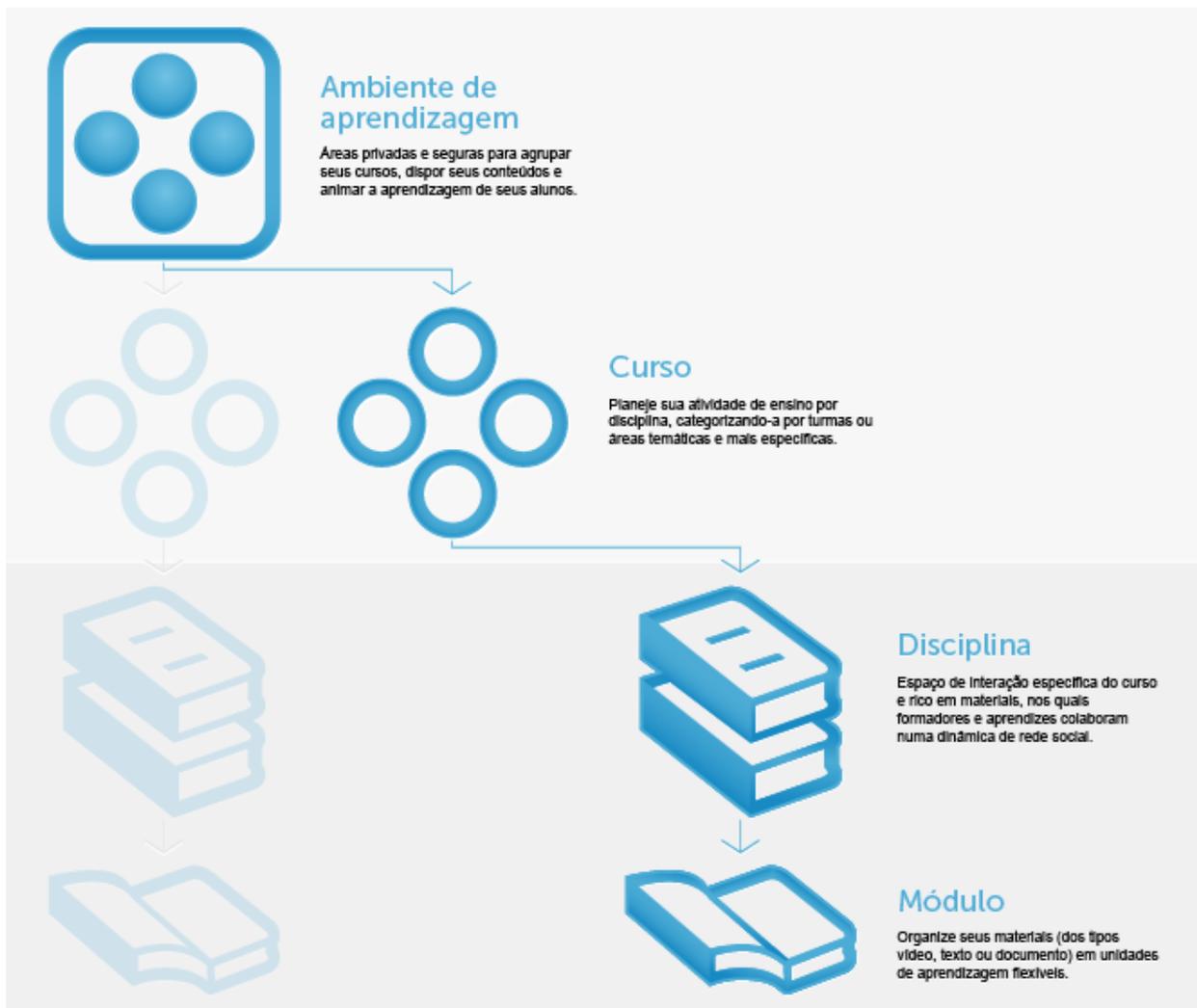


Figura 2 - Hierarquia detalha do Redu

Assim como na Universidade, o usuário se matricula nos Cursos do Redu, e tem acesso a todos os recursos para aquele curso. Isto significa que a moderação de entrada e saída dos usuários é feita através dos cursos.

3. Sistema de Recomendação

Neste capítulo será mostrada uma visão geral das pesquisas relacionadas a esse tema, as quais foram insumos para a construção da solução. Com intuito de explicar a motivação por trás de cada decisão, serão mostrados algoritmos que inspiraram a solução e suas características. Por fim, será relatado a conclusão das pesquisas e o intuito de utilizar ou não utilizar cada algoritmo.

3.1 Algoritmos de recomendação de usuários

Com intuito de dar embasamento à solução deste trabalho, cinco algoritmos que tratam de recomendações serão revistos. Os quatro primeiros são bem conhecidos e foram tema de pesquisa de Chen *et al* (2009), que explica e explicita as características de cada. O último algoritmo é uma contribuição de Silva *et al* (2010), na qual descreve um sistema de recomendação utilizando-se como base a teoria de rede Mundo Pequeno e a técnica de Algoritmo Genético.

A contribuição de Chen *et al* (2009) se baseia em uma Rede Social Corporativa da IBM, Beehive. Nesta rede, foram testados quatro algoritmos que focam na estrutura da rede e são baseados na similaridade de conteúdos. O algoritmo proposto por Silva *et al* (2010), utiliza-se de uma Rede Social Corporativa do Centro de Estudos e Sistemas Avançados do Recife (C.E.S.A.R), chamada Oro-aro.

3.1.1 Content Matching

O algoritmo *Content Matching* é baseado na teoria “se ambos os usuários postam conteúdos em tópicos similares, eles devem ter interesse em conhecer um ao outro”. Ou seja, o algoritmo se esforça em unir usuários que tenham conteúdos em comum.

Utilizando-se de técnicas comuns de extração de informação, criou-se uma nuvem de palavras para cada usuário, que foram obtidas de dentro da Rede Beehive e de um diretório corporativo da IBM. Da rede Beehive, extraiu-se palavras advindas do perfil do usuário, mensagens de status dos usuários, bem como título, descrição, palavras-chave e qualquer outro conteúdo textual associado às fotos e listas compartilhadas. Do diretório corporativo, extraiu-se o cargo do usuário e a cidade onde o usuário trabalha. Todas as palavras extraídas foram originadas pelo algoritmo *Porter Stemming*, que é um algoritmo responsável por remover sufixos e prefixos das palavras em inglês. Após a geração das palavras, um filtro é aplicado usando-se um catálogo de 550 palavras comum em inglês. Todas as palavras que passam pelo filtro são associadas com o usuário u e são usadas para criar um vetor de palavras $V_u = (v_u(w_1), \dots, v_u(w_m))$, para descrever u , onde m é o número total de palavras distintas usadas pelo usuário e cada $v_u(w_m)$ descreve a força do u na palavra w_i .

O valor de w_i é calculado usando-se *term-frequency inverse-user-frequency weighting*, uma direta adaptação de TF-IDF (Salton e Buckley 1988):

$$TF_u(w_i) = (\text{\#usos de } w_i \text{ por } u) \div (\text{\#todas as palavras usadas por } u)$$

$$IDF_u(w_i) = \log[(\text{\#todos os usuários}) \div (\text{\#usuários que usam } w_i \text{ ao menos uma vez})]$$

$$v_u(w_i) = TF_u(w_i) \cdot IDF_u(w_i)$$

a similaridade entre os dois usuários a e b é então mensurada através do cosseno dos dois vetores de palavras V_a e V_b .

Intuitivamente isto significa que a e b serão considerados similares se eles compartilham muitas palavras em comum, e ainda mais se apenas eles compartilham essas palavras. Para critério de criação da lista de recomendação, usuários semelhantes ao usuário u serão recomendados em ordem decrescente de similaridade.

3.1.2 Content-plus-link

O algoritmo *Content-plus-link* expande o algoritmo *Content Matching* com informação de link social derivada da estrutura da rede social. A motivação por trás deste algoritmo é mostrar um

caminho da rede para um laço fraco ou usuário desconhecido, para dar mais insumos ao recipiente da recomendação, fazendo com que o mesmo seja mais favorável à aceitação da recomendação. O algoritmo *Content-plus-link* computa a similaridade da mesma maneira que *Content matching* descrito na seção anterior. Porém, ao invés de recomendar os usuários com as maiores pontuações de similaridade, aumenta-se a similaridade de um candidato c e u em 50% se um link social válido for encontrado. Ou seja, dois usuários que possuem poucas palavras em comum, mas com um link social válido pode ser classificado melhor do que se tivessem mais palavras em comum.

Um link social válido é definido como uma sequência de relacionamento ou interação entre três ou quatro usuários, o primeiro sendo o destinatário da recomendação e o último sendo o usuário recomendado. Cada dois usuários consecutivos a e b na sequência devem satisfazer ao menos uma das seguintes condições:

1. a está conectado com b
2. a comentou algo de b
3. b está conectado de a

Esta definição garante que um link social entre dois usuários existe se e apenas se ao menos um mínimo de interação e colaboração existir entre ambos. Um exemplo desse link entre dois usuários Alice e Charles seria “Alice comentou no mural de Bob, que é considerado amigo de Charles”.

Aumentar a pontuação de similaridade dos candidatos à recomendação com links válidos favorece os candidatos com pequena proximidade da rede social do usuário em relação às pessoas distantes da rede social do mesmo. Ainda mais, pode-se utilizar essa informação de caminho entre usuários com links válidos como argumento de recomendação, ao mostrar o caminho que o algoritmo percorreu para gerar a sugestão.

3.1.3 Friends-of-Friends (FoF)

Todos os dois algoritmos anteriores possuem análise sobre o conteúdo que o usuário produzia, *Friends-of-Friends* se baseia unicamente em informações da rede social. *FoF* parte da intuição

de quê “se muitos dos meus amigos consideram Alice como amiga, talvez Alice possa ser minha amiga também”. Várias análises de redes sociais adotaram ideias similares para encontrar vizinhos e caminhos em uma rede (Geyer et al 2008, Groh e Ehmig 2007 e Granovetter 1973). Esta variante diz que recomendar amigos de um amigo é interessante não por apenas ser bastante intuitiva, mas também por, segundo o blog oficial do Facebook¹, ser uma das métricas fundamentais utilizadas no “*People You May Know*”, que é uma funcionalidade do Facebook utilizada para recomendar usuários. Esta funcionalidade é uma das poucas implantadas em um site de rede social de larga escala e que tem uma efetividade comprovada.

Do lado formal, pode-se definir um predicado $F(a, b)$ que será verdadeiro se e somente se b for amigo de a e ambos a e b forem usuários de Beehive. O algoritmo assim pode ser descrito: para um usuário u que será o recipiente das recomendações, será candidato recomendado se:

$$RC(u, c) = \{ \text{usuário } c \mid \exists \text{ usuário } a \text{ de tal forma que } (u, a) \text{ e } F(a, c) \}.$$

Para cada candidato $c \in RC(u)$, são amigos em comum² se

$$AC(u, c) = \{ \text{usuário } a \mid F(u, a) \text{ e } F(a, c) \},$$

que representa os amigos de u que são conectados com c e servem como ponte entre u e c . A quantidade de amigos em comum é definido pelo tamanho do conjunto $AC(u, c)$, que é utilizado como pontuação para a classificação dos candidatos. Estes são recomendados para o usuário u em ordem decrescente de amigos em comum.

3.1.4 SONAR

Este algoritmo é baseado no sistema SONAR, que agrega todas as informações de relacionamento social de diferentes fontes de dados usados na IBM (Guy et al, 2008). Para a realização da recomendação foram utilizadas sete fontes de dados: (1) organograma, (2) base de dados de pu-

¹ <http://blog.facebook.com/>

² No caso de Beehive, como a existe a ideia de relacionamentos uni-direcionais, Amigos em Comum também são considerados aqueles que só possuem um lado do recipiente.

blicações, (3) base de dados de patentes, (4) sistema de amizades, (5) sistemas de palavras-chave dos usuários, (6) projeto wiki, e (7) sistema de blogs. Um relacionamento é indicado se dois usuários interagiram entre si de alguma maneira, como por exemplo, sendo coautor de uma publicação ou deixando comentários no blog do outro. Nota-se que o sistema SONAR é algo próprio da IBM, o que se instituiu o uso de algo bem específico para construção do sistema de recomendação de usuários.

Para realizar a recomendação com informação de cada uma destas fontes de dados, o algoritmo SONAR computa uma pontuação de relacionamento normalizada no intervalo de $[0,1]$ entre duas pessoas, onde 0 indica nenhum relacionamento e 1 indica o grau máximo de relacionamento. Estas pontuações são então agrupadas em uma única pontuação ponderando-se igualmente todas as fontes de dados (Chen *et al*, 2009). Dado um usuário u , SONAR retorna uma lista de usuários ligados a u e suas pontuações agregadas de cada relacionamento.

Simplificando, SONAR incorpora todas as informações disponíveis dentro da IBM que podem ser utilizadas para adquirir conhecimento entre duas pessoas, e então as classifica baseada na força e frequência das interações entre os usuários.

3.1.5 Algoritmo de Silva

Este algoritmo é baseado nas propriedades da estrutura da rede social. As características topológicas, as informações e as métricas são derivadas da teoria de redes complexas, que dão insumos ao algoritmo. Os sites de redes sociais são tipos de redes que são observados como sendo do tipo Mundo Pequeno (*Small-world*) ou Livre de Escala (*Scale free*) (Ahn *et al*, 2007 e Mislove *et al*, 2007), que são modelos de redes complexas.

Redes livres de escala, de forma bem resumida, são redes que simplesmente possuem uma distribuição de grau de nós seguindo uma lei de potência. Isto significa que muitos usuários possuem poucos amigos, enquanto que poucos usuários possuem muitos amigos. Enquanto que as redes Mundo Pequeno são redes que possuem um caminho médio entre dois usuários com valor baixo, ou seja, para que o usuário se conecte com qualquer outro na rede existem poucos usuários entre eles, além de ter as características das redes Livres de Escala.

Para o processo de recomendação, o relacionamento do usuário central e o usuário a ser recomendado são avaliados para a produção de um único valor que será dito como a força do relacionamento. Este valor é a média ponderada de três índices independentes. Estes foram utilizados devido sua simplicidade e a ideia intuitiva por trás deles (Silva *et al*, 2010). Os índices escolhidos foram: *Friends-of-Friends* (índice já discutido anteriormente) e outros dois índices que se inspiram na ideia de redes mundo pequeno e no conceito de coeficiente de clusterização, que será mostrado mais adiante. Para encontrar os valores dos pesos utilizados na ponderação dos índices, foi utilizada a técnica para problemas de busca e otimização, denominada Algoritmo Genético.

Alguns conceitos e variáveis são definidas para gerar os índices usados no sistema de recomendação. A variável C_i é definida como um conjunto de nós adjacentes ao nó v_i , por exemplo, e D_{C_i} será definido como a densidade adjacente entre o nós contidos no conjunto C_i . A densidade é calculada pela divisão da quantidade de relacionamentos contidos C_i em pela quantidade total de relacionamentos que poderiam ser criados. Pensando de maneira formal, seria a quantidade de arestas existentes em C dividido pela quantidade de arestas pelo clique (grafo completo) de um grafo.

$$D_C = \frac{\sum_{i \in C} (\sum_{j \in C} (M_{ij}))}{(|C| * (|C| - 1)) / 2}$$

Onde M_{ij} é o elemento que pertence a C .

1) Primeiro Índice

É simplesmente a quantidade de amigos em comum, ou seja, é quantidade de nós adjacentes, que são ligados simultaneamente ao nó i e o nó j .

$$I_{1ij} = |C_i \cap C_j|$$

2) Segundo Índice

É a densidade do resultado medido pelo primeiro índice.

$$I_{2ij} = D_{C_i \cap C_j}$$

Este índice mede o grau de relacionamento entre os amigos em comum dos nós de i e j , a Figura 03 ilustra os nós que serão analisados. Um valor baixo desse índice significa que a pequena rede formada pelos amigos em comum dos nós i e j , não é bem relacionada.

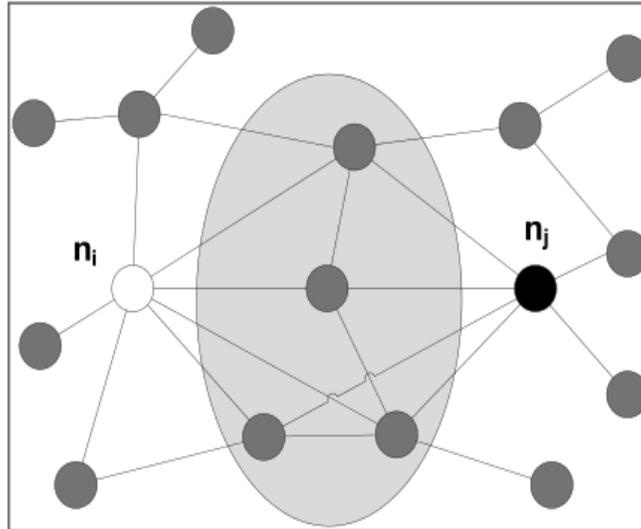


Figura 3 - Relacionamento entre os amigos em comum

3) Terceiro Índice

O terceiro índice é uma variação do segundo índice, ao invés de levar em conta os amigos em comum, ou seja, a interseção entre os nós adjacentes, leva em conta a união dos nós adjacentes, como é mostrada na Figura 04.

$$I_{3ij} = D_{C_i \cup C_j}$$

Este índice mede o grau de coesão do grande grupo formado pelos amigos dos nós i e j .

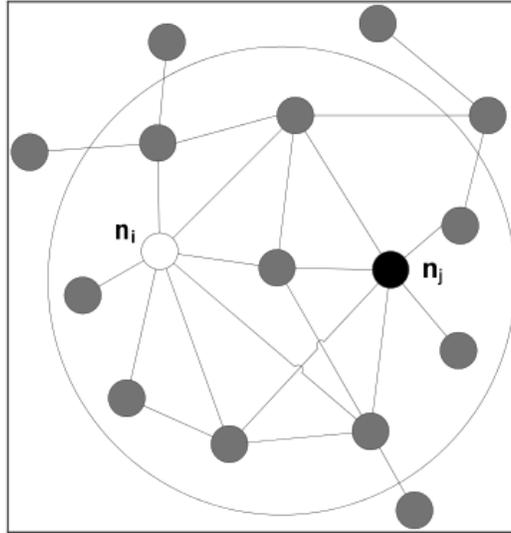


Figura 4 - Relacionamentos entre os amigos dos usuários n_i e n_j

A etapa de calibragem dos pesos é parte fundamental na construção da recomendação, pois ela gerará, segundo a ponderação dos três índices, um único valor que medirá o grau de relacionamento entre dois usuários.

A calibragem dos pesos de cada índice deve ser ajustada para se obter um resultado otimizado. Neste caso, a otimização significa classificar os mais importantes usuários no começo da lista de recomendação. A importância de um usuário na rede social depende do contexto, este contexto pode mudar dependendo do usuário que estiver procurando pela recomendação. Ou seja, a função de avaliação deve considerar a existente estrutura de relacionamentos do usuário (Silva *et al*, 2010).

A função de avaliação, que será utilizada para dizer o quanto os pesos estão se aproximando de uma solução ótima no algoritmo genético, é a média das posições que já estão ligadas ao nó central. Estas posições são calculadas quando os pesos e os índices são gerados e ponderados para formar um único valor. Este valor posiciona as recomendações seguindo a lógica de que o menor valor é a melhor ponderação a ser considerada. Na figura abaixo está representado a função de avaliação com o passar das iterações do algoritmo genético. A linha preta está repre-

sentando o valor da função de avaliação, as linhas cinza são os nós que já possuem relacionamento com o usuário e as brancas aquelas que não possuem.

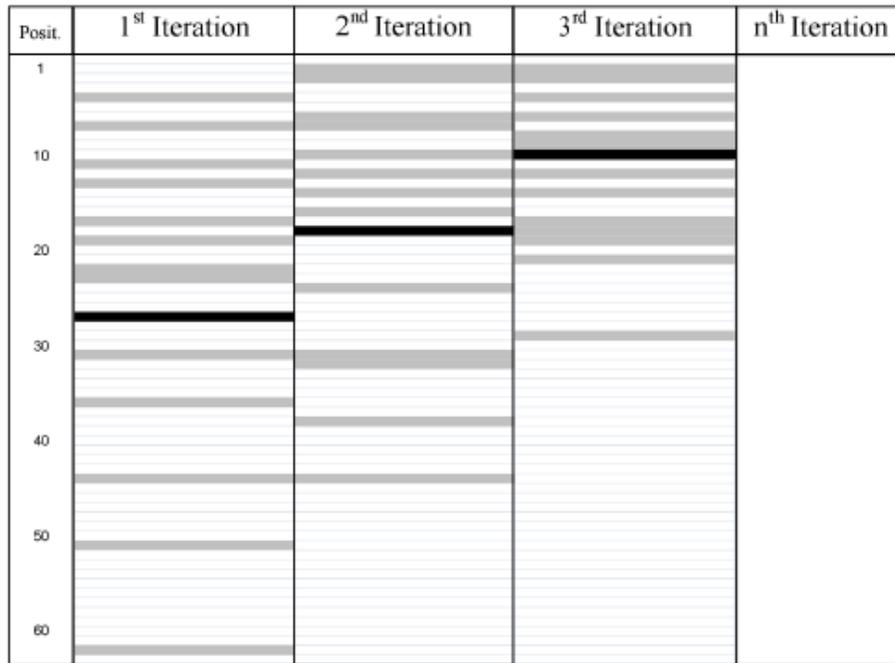


Figura 5 - Representação gráfica dos valores da função de avaliação com o passar das iterações

Para calcular a função de avaliação é preciso que haja a unificação dos pesos e dos índices com intuito de forma um único valor. Para tal, utiliza-se uma função de otimização baseada na soma ponderada dos pesos e índices, representada pela fórmula a seguir:

$$M(n, w) = I_1(n_c, n) \cdot w_1 + I_2(n_c, n) \cdot w_2 + I_3(n_c, n) \cdot w_3$$

para critério de parada do algoritmo genético, estipulou-se que o valor da função de otimização não se modificasse por quatro iterações.

A lista final da recomendação é formada pelos usuários não ligados ao usuário que procura a recomendação, com a ordem crescente da função de otimização.

3.2 Conclusões

Neste capítulo foram examinados cinco algoritmos de complexidade e características diferentes, quatro deles foram extraídos da pesquisa de Chen *et al* (2009) que os utilizou na rede social corporativa Beehive e fez uma análise do impacto e das características de cada algoritmo. O último algoritmo foi extraído da pesquisa de Silva *et al* (2010) que o utilizou na rede social corporativa Oro-Aro e executou experimentos comparando-o com o algoritmo *Friends-of-Friends*.

No artigo Chen *et al* (2009) realizou-se experimentos de recomendação de relacionamentos e constou-se que os algoritmos baseados nas informações que constituem a rede social (nós e arestas) foram capazes de produzir recomendações mais bem aceitas e encontraram mais contactos conhecidos, enquanto que os algoritmos baseado na similaridade entre conteúdo criado pelo usuário foi melhor em recomendar novas amizades. Quanto a eficácia, Chen *et al* (2009) mostra que o algoritmo SONAR é o algoritmo com a menor taxa de rejeição das recomendações entre os quatro examinados no artigo, e foi o único a utilizar as especificidades da Rede Social Beehive em conjunto com fontes de dados da IBM.

A pesquisa de Silva *et al* (2010) mostra a eficácia do uso de teorias de redes complexas, como redes Mundo Pequeno e Livre de Escala e o coeficiente de clusterização, como uma excelente abordagem para construção de um sistema de recomendação. Comparou-se o algoritmo proposto por Silva *et al* (2010) com o algoritmo *FoF* e constatou-se a eficácia da abordagem mostrando uma taxa de aceitação da recomendação melhor para o algoritmo de Silva *et al* (2010).

4. Sistema de Recomendação para uma Rede Social Educacional

Neste capítulo será proposto um novo algoritmo para recomendação de usuários, utilizando-se como base os algoritmos explicados no [Capítulo 3](#) e o Redu (Rede Social Educacional). Na próxima seção serão descritos a motivação e o porquê das escolhas que foram feitas para a elaboração do sistema de recomendação. Na seção seguinte, será descrito o algoritmo do sistema de recomendação. E por último, será feita uma discussão em torno do experimento realizado.

4.1 Considerações Iniciais

O algoritmo proposto neste trabalho foi baseado nas pesquisas realizadas por Chen *et al* (2009) e Silva *et al* (2010). Foi escolhida a abordagem de considerar unicamente a topologia da rede social, pois como Chen *et al* (2010) descreve, é melhor em encontrar os contatos conhecidos e como o primeiro comportamento do usuário é procurar os contatos conhecidos, tal abordagem foi considerada alinhada com os objetivos do Redu neste primeiro momento. O mecanismo de recomendação foi baseado na contribuição de Silva *et al* (2010), com algumas adições para se adequar melhor a especificidade do Redu. Assim como Chen *et al* (2010) construiu o algoritmo SONAR, onde se utiliza de fontes de dados da IBM para construir algo mais específico.

4.2 Algoritmo

O algoritmo proposto por este trabalho possui duas fases: a primeira é o cálculo dos valores dos índices e a segunda é a calibragem dos pesos que serão utilizados para dar mais importância a determinados índices, dependendo da estrutura da rede social do usuário.

Para o cálculo dos valores dos índices, foram utilizados os três índices já citados na sessão 3.1.5, investigados por Silva *et al* (2010), e mais um índice que será descrito mais adiante. De forma bem resumida, os três índices são: (1) a quantidade de amigos em comum dos usuários comparados, (2) o quão relacionado os amigos em comum e (3) quão relacionado são os amigos dos dois usuários (considerando todos os amigos não só os em comum).

O índice adicionado foi criado para utilizar a especificidade da hierarquia do Redu. O índice consiste na quantidade dos cursos em comum de dois usuários, assim como *FoF*, mas ao invés de levar em consideração contatos leva-se em consideração os cursos onde os usuários estão matriculados. Para entendermos de forma mais clara, vamos considerar que C_i é o conjunto de cursos que pertencem ao usuário v_i , com isto, de maneira formal pode-se definir o índice da seguinte maneira:

$$I_{4ij} = | C_i \cap C_j |$$

Onde C_j é o conjunto de cursos pertencentes ao usuário v_j .

A etapa de ponderação segue o mesmo princípio definido por Silva *et al* (2010), mas algumas modificações foram feitas para adequar o Algoritmo Genético a mais um índice. Abaixo será mostrado um fluxograma do funcionamento do algoritmo genético, com a intenção de mostrar todos os passos inclusos no algoritmo.



Figura 6 - Fluxo do mecanismo de recomendação, fase de cálculo dos índices e o começo da segunda fase

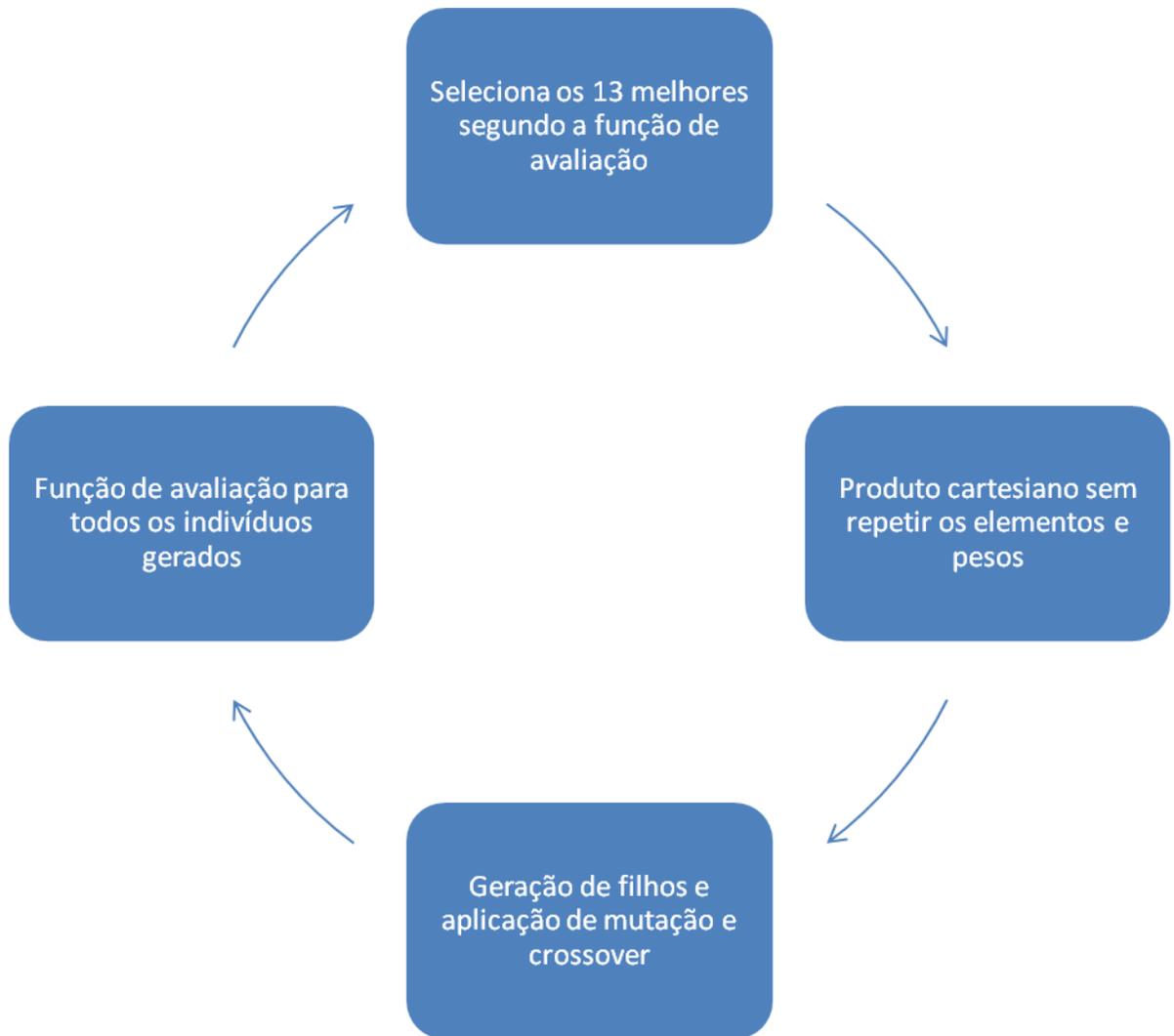
Supondo que o usuário u será o utilizado como o destinatário das recomendações, o primeiro passo é calcular todos os índices para os nós adjacentes a u e os adjacentes dos adjacentes. Com

os índices calculados, passamos para o segundo passo que é o da calibragem dos pesos. Nesta etapa o algoritmo genético gerará 200 indivíduos com pesos produzidos de forma aleatória, e para cada indivíduo será calculada a função de otimização definida na fórmula abaixo,

$$M(n, w) = I_1(n_c, n).w_1 + I_2(n_c, n).w_2 + I_3(n_c, n).w_3 + I_4(n_c, n).w_4$$

Com o uso da função de otimização, haverá o cálculo da função de avaliação da mesma maneira explicada por Silva *et al* (2010). Após o cálculo da função de avaliação, os indivíduos são ordenados de forma ascendente da função de avaliação e os treze primeiros são escolhidos.

Com estes treze usuários, será aplicado o produto cartesiano para criação de pares dos indivíduos de elite. Nestes, será aplicado o *crossover*, onde o ponto de corte é único e aleatório, que produzirá novos indivíduos. A mutação será aplicada aos elementos resultantes do *crossover* e no próximo passo haverá a combinação dos filhos e pais para eliminação dos duplicados. Como critério de parada do Algoritmo genético, foi adicionado um novo critério no qual se o conjunto resultando do plano cartesiano for vazio, o último resultado será retornado. Isto foi adicionado, pois o algoritmo de Silva *et al* (2010) em alguns momentos não convergia e nenhum resultado era retornado. Ao final, a melhor combinação de pesos para o indivíduo será retornada.



4.3 Experimento e Resultado

Para a avaliação do algoritmo proposto foram realizados dois experimentos comparando o Algoritmo Proposto neste trabalho com o algoritmo *FoF* (*Friends-of-Friends*) e o algoritmo proposto por Silva et al (2010). Os dois experimentos foram realizados no Redu, o primeiro foi uma comparação temporal realizada com a base de dados de Julho de 2011 com a de Outubro de 2011. O segundo experimento foi um *Survey* realizado com 47 usuários do Redu, os quais responderam retornando dentre quinze usuários recomendados quais eles conheciam ou se conectariam.

Algoritmo Usado	Usuários recomen- dados	Usuários Aceitos	Taxa de aceitação
Algoritmo de Silva	535	93	17,4%
Algoritmo Proposto	535	93	17,4%
FoF	134	17	12,7%

Tabela 1 - Resultados do experimento da comparação de Julho e Dezembro de 2011

O primeiro experimento avaliou 37 usuários para o algoritmo proposto por este trabalho e para o algoritmo de Silva *et al* (2010), para o algoritmo FoF avaliou-se 9 usuários. Foram geradas recomendações para os usuários, o objetivo era gerar quinze recomendações para cada usuário, mas em alguns casos não se tinha quantidade suficientes de nós para fazê-lo, o que culminou em alguns usuários receberem menos de 15 usuários recomendados. As recomendações foram geradas com usuários da base de dados de Julho de 2011 e comparou-se se alguma recomendação foi concretizada em Outubro de 2011. Não houve uma apresentação da recomendação para o usuário, o que houve foi uma comparação temporal e o quanto cada algoritmo previa de maneira correta as recomendações.

Na Tabela 1 estão os resultados do primeiro experimento realizado com a comparação das bases de Julho e Outubro. Observando-se a taxa de aceitação, nota-se que o Algoritmo Proposto obteve a taxa de aceitação igual ao do proposto por Silva *et al* (2010) e acima da taxa de aceitação de *FoF*.

No segundo experimento foi realizado um *survey* para avaliação dos algoritmos com a base de dados atual do Redu. Para a divisão entre os algoritmos, dezesseis usuários foram para o algoritmo proposto e o mesmo valor para o algoritmo de Silva. Para o *FoF* foram quinze usuários, a divisão foi feita de forma aleatória. Um e-mail foi enviado para os quarenta e sete usuários que possuíam mais de 15 contatos e acessaram o Redu nos últimos cinco dias, isto foi feito para aumentar a quantidade de usuários que iriam responder o e-mail. Os e-mails foram enviados no

dia 27 de novembro de 2011, e até o presente momento (dia 04 de dezembro de 2011) onze pessoas responderam aos e-mails.

Algoritmo Usado	Usuários recomendados	Usuários Aceitos	Taxa de aceitação
Algoritmo de Silva	30	5	16%
Algoritmo Proposto	90	32	35%
FoF	45	31	68%

Tabela 2 - Resultados obtidos no segundo experimento

Na tabela 2 estão os resultados obtidos para o segundo experimento. Nota-se que a taxa de algoritmo Silva *et al* (2010) foi bem abaixo do que o esperado, assim como a taxa do algoritmo proposto. Diante disso, uma entrevista informal foi realizada com dois usuários que receberam a recomendação do algoritmo proposto. De forma resumida, constatou-se que a escolha de usuários para testes não levou em conta o tipo de usuário e o objetivo dele na rede. Os dois usuários eram administradores do Redu e estavam matriculados em vários cursos que não participavam efetivamente, portanto o algoritmo proposto sugeriu muitos usuários que estavam no mesmo curso que os usuários, como era de se esperar, porém estes não tiveram nenhum contato de fato. Para o algoritmo de Silva *et al* (2010) que teve o menor retorno entre os e-mails e apresentou a menor taxa de aceitação (16%), nota-se que ocorreu a mesma coisa que o algoritmo proposto, um dos usuários era administrador do Redu e estava conectado com muitas pessoas para suporte e pesquisa com o usuário.

O Redu por se tratar de uma rede social educacional bem nova, 10 meses de existência, ainda não conseguiu estabelecer um grupo de usuários fixo e que utiliza a ferramenta no dia-a-dia. E para o experimento, a resposta de somente onze usuários faz com que cada usuário tenha um papel muito forte no resultado final.

5. Conclusão

Viu-se que a importância em criar relacionamentos para as redes sociais é muito grande e no contexto educacional é maior ainda, visto que, são nos relacionamentos que a colaboração existirá e assim o conhecimento irá se propagar. Portanto, um sistema de recomendação de usuários é essencial para todo site de rede social, ainda mais se esse site for de uma rede social educacional, como é o caso do Redu.

Baseado nesta ideia foi proposto um sistema de recomendação de usuários para a Rede Social Educacional Redu, com intuito de aumentar os relacionamentos e com isso a colaboração. No primeiro momento viu-se que utilizar a especificidade do Redu poderia ser uma boa abordagem, tendo como base as ideias de Chen *et al* (2009). Nos primeiros resultados obtidos demonstrou-se que isso estava certo, mas foi necessária uma abordagem mais concreta para avaliar de maneira mais incisiva.

Os resultados obtidos no segundo experimento demonstraram que o tamanho da base e os critérios de escolha dos indivíduos para testes não foram bem definidos. Em consequência, a solução proposta e o algoritmo de Silva *et al* (2010) obtiveram resultados bem abaixo do que se esperava, os quais destoaram com os do primeiro experimento. Conclui-se que ainda não é possível definir a qualidade do algoritmo proposto, pois uma avaliação robusta e suficiente não foi feita. Isto ocorreu, pois não foi encontrada uma base suficientemente grande com características semelhantes do Redu, mas há indícios de que tal abordagem é de grande significância para a recomendação.

Para trabalhos futuros, é interessante analisar o algoritmo proposto em uma rede social de maior porte e que tenham características parecidas com o Redu, como por exemplo, o Facebook (ao estabelecer a semelhança entre Cursos do Redu e Grupos do Facebook). Além disso, utilizar uma nova função de avaliação ou até mesmo mudar a técnica de busca da melhor solução para torna a recomendação mais eficiente.

Referências Bibliográficas

AHN, Y.-Y.; HAN, S.; KWAK, H.; MOON, S.; JEONG, H. **Analysis of topological characteristics of huge online social networking services**. Proceedings of the 16th international conference on World Wide Web. *Anais...*, WWW '07. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1242572.1242685>>. Acesso em: 24 nov. 2011. , 2007

BOYD, D. M.; ELLISON, N. B. Social network sites: definition, history, and scholarship. *IEEE Engineering Management Review*, v. 38, n. 3, p. 16-31, Third Quarter 2010.

CHEN, J.; GEYER, W.; DUGAN, C.; MULLER, M.; GUY, I. **Make new friends, but keep the old: recommending people on social networking sites**. Proceedings of the 27th international conference on Human factors in computing systems. *Anais...*, CHI '09. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1518701.1518735>>. Acesso em: 19 nov. 2011. , 2009

CHO, H.; GAY, G.; DAVIDSON, B.; INGRAFFEA, A. Social networks, communication styles, and learning performance in a CSCL community. *Comput. Educ.*, v. 49, n. 2, p. 309–329, set 2007.

GEYER, W.; DUGAN, C.; MILLEN, D. R.; MULLER, M.; FREYNE, J. **Recommending topics for self-descriptions in online user profiles**. Proceedings of the 2008 ACM conference on Recommender systems. *Anais...*, RecSys '08. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1454008.1454019>>. Acesso em: 1 dez. 2011. , 2008

GRANOVETTER, M. **Strength of weak ties**. *Amer. J. Sociology* 78 (1973), 1360-1380, 1973.

GREENHOW, C. **Social networking and education: emerging research within CSCL**. Proceedings of the 9th international conference on Computer supported collaborative learning - Volume 1. *Anais...*, CSCL'09. [S.l.]: International Society of the Learning Sciences. Disponível em: <<http://dl.acm.org/citation.cfm?id=1600053.1600119>>. Acesso em: 19 nov. 2011. , 2009

GROH, G.; EHMIG, C. **Recommendations in taste related domains: collaborative filtering vs. social filtering.** Proceedings of the 2007 international ACM conference on Supporting group work. *Anais...*, GROUP '07. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1316624.1316643>>. Acesso em: 1 dez. 2011. , 2007

GUY, I.; JACOVI, M.; MESHULAM, N.; RONEN, I.; SHAHAR, E. **Public vs. private: comparing public social network information with email.** Proceedings of the 2008 ACM conference on Computer supported cooperative work. *Anais...*, CSCW '08. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1460563.1460627>>. Acesso em: 1 dez. 2011. , 2008

GUY, I.; JACOVI, M.; SHAHAR, E. *et al.* **Harvesting with SONAR: the value of aggregating social network information.** Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. *Anais...*, CHI '08. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1357054.1357212>>. Acesso em: 1 dez. 2011. , 2008

MELO, C. D. A. **Scaffolding of Self-Regulated Learning in Social Networks.** Universidade Federal de Pernambuco. Recife, p. 109. 2010.

MISLOVE, A.; MARCON, M.; GUMMADI, K. P.; DRUSCHEL, P.; BHATTACHARJEE, B. **Measurement and analysis of online social networks.** Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. *Anais...*, IMC '07. New York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1298306.1298311>>. Acesso em: 1 dez. 2011. , 2007

SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. **Information Processing & Management**, v. 24, n. 5, p. 513-523, 1988.

SILVA, N. B.; ING-REN TSANG; CAVALCANTI, G. D. .; ING-JYH TSANG. **A graph-based friend recommendation system using Genetic Algorithm.** 2010 IEEE Congress on Evolutionary Computation (CEC). *Anais...* [S.l.]: IEEE. , 18 jul 2010

STAHL, G.; KOSCHMANN, T.; & SUTHERS, D. (2006). **Computer-supported collaborative learning: An historical perspective.** In R. K. Sawyer (Ed.), Cambridge handbook of the learn-

ing sciences (pp. 409-426). Cambridge, UK: Cambridge University Press. Available at http://GerryStahl.net/cscl/CSCL_English.pdf in English