

CredibilityRank: um Arcabouço para Projeto e Avaliação de Modelos de Credibilidade de Serviços da Web*

Sara Guimarães, Arlei Silva, Wagner Meira Jr.
Universidade Federal de Minas Gerais (UFMG)
Depart. de Ciência da Computação
Av. Antônio Carlos 6627
31270-010 - BH, MG, Brasil
{sara,arlei,meira}@dcc.ufmg.br

Adriano Pereira
Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)
Departamento de Computação
Av. Amazonas, 7675
30510-000 - BH, MG, Brasil
adriano@decom.cefetmg.br

RESUMO

A popularização das aplicações Web tem feito surgir novos serviços a cada dia, bem como tem demandado mecanismos que assegurem a credibilidade desses serviços. Até o presente momento, muito pouco foi feito no sentido de medir e entender a credibilidade dos serviços neste complexo ambiente da Web, o que por si só representa um grande desafio de pesquisa. Com base nas dificuldades relacionadas a essa tarefa de atribuir um valor de credibilidade a um serviço online da Web 2.0, propomos um arcabouço para o projeto, a implementação e a avaliação de modelos de credibilidade. Denominamos modelo de credibilidade uma função capaz de atribuir um valor de credibilidade a um serviço da Web, considerando diferentes critérios associados a esse serviço e a seu fornecedor. Para validar o arcabouço, realizamos experimentos usando dados reais de um serviço de comércio eletrônico, avaliando modelos de credibilidade baseados em diferentes critérios. Os resultados mostram que o modelo é aplicável, sendo capaz de dar suporte à tomada de decisão de usuários de serviços na Web.

ABSTRACT

The popularization of Web applications has given rise to new services every day, demanding mechanisms to ensure the credibility of these services. Since now, little has been done to measure and understand the credibility of services on this complex Web environment, which itself is a major research challenge. From the challenges related to the task of assigning a credibility value to an online service in Web 2.0 applications, we propose a framework for the design, implementation and evaluation of credibility models. We call a credibility model a function capable of assigning a credibility value to any transaction of a service on the Web, considering different criteria of this service and its supplier. To validate the framework, we perform experiments using an actual dataset, from

*CredibilityRank: a Framework for Design and Evaluation of Credibility Models in Web Applications

which we evaluated different credibility models using distinct types of information sources, and it allows to compare and evaluate these credibility models. The results show that the credibility framework has applicability and is capable to support decision of users of Web services.

Categories and Subject Descriptors

K.4.4 [Computers and Society]: Electronic Commerce; H.3.5 [Online Information Services]: Web-based services; H.4.m [Information Systems]: Miscellaneous

General Terms

Experimentação, Confiabilidade

Keywords

Credibilidade, Modelos de Credibilidade, Serviços da Web, Mercados Eletrônicos, Web 2.0

1. INTRODUÇÃO

A popularização das aplicações da Web 2.0, onde usuários podem interagir criando e compartilhando conteúdo, negociando produtos e formando comunidades, representou uma grande revolução da forma como usuários e corporações utilizam a Web. Essa revolução trouxe desafios relacionados à credibilidade dessas aplicações. Como consequência, mecanismos que permitam aos usuários avaliar de forma mais fidedigna a credibilidade dos serviços da Web são essenciais nos dias de hoje.

Bibliotecas digitais, mercados eletrônicos [5], sistemas de compartilhamento e geração de conteúdo [3, 4] são exemplos de aplicações que demandam mecanismos de avaliação da credibilidade. Diversas dessas aplicações já provêem sistemas para lidar com isso, tais como sistemas de reputação. Em sistemas de compartilhamento de conteúdo, por exemplo, é comum atribuir uma reputação aos usuários com base em um histórico de *feedbacks* recebidos. Outras informações disponibilizadas por essas aplicações podem ser vistas como critérios de credibilidade, tais como o tempo de cadastro de um vendedor em um mercado eletrônico [7], ou o número de citações de um autor em uma biblioteca digital. Entretanto, agregar diferentes critérios de credibilidade, provendo suporte à tomada de decisão do usuário de forma efetiva, é um desafio.

A tarefa de avaliar e quantificar a credibilidade de um serviço da Web representa um grande desafio de pesquisa. Dentre as principais dificuldades associadas à essa tarefa, podemos destacar: o

grande número de variáveis envolvidas e a baixa confiabilidade das informações disponíveis. Com base nessas dificuldades, propomos um arcabouço, denominado *CredibilityRank*, que permite projetar, executar e avaliar modelos de credibilidade capazes de gerar uma escala de credibilidade (*ranking*) dos serviços com base em diferentes critérios.

Modelos de credibilidade diferem dos modelos de reputação, já vastamente estudados na literatura [16, 26] porque não consideram apenas o *feedback* dos usuários como forma de avaliação de um serviço, mas um conjunto de atributos de credibilidade, relacionados ao serviço provido e também ao usuário fornecedor do serviço, como forma de enriquecer a avaliação desse serviço da *Web*.

A partir da formulação do problema da credibilidade apresentada neste artigo, diferentes modelos podem ser propostos. Consequentemente, a avaliação de um modelo de credibilidade se torna essencial para a comparação de diferentes modelos para uma dada aplicação. Portanto, nosso arcabouço pode ser utilizado para projetar, implementar e avaliar modelos de credibilidade. Essa avaliação é baseada em uma amostra significativa de serviços com *feedbacks* dos usuários e numa escala de credibilidade gerada pelo modelo. Quanto maior a capacidade do modelo de posicionar serviços satisfatórios sob o ponto de vista do usuário em posições superiores nessa escala, maior é a sua qualidade.

É importante explicar que, apesar dos problemas existentes nos sistemas de reputação [23], é necessário usar informações de *feedback* para que a opinião do usuário possa ser mensurada em relação a um serviço da *Web*, que em um modelo de credibilidade está descrito por diferentes características chamadas atributos de credibilidade. Além disso, existem trabalhos específicos que lidam com o problema de melhorar a qualidade dos sistemas de reputação, tais como os que buscam identificar fraudadores desses sistemas [19], que inclusive foi utilizado na mesma aplicação real adotada neste trabalho.

Neste trabalho, realizamos experimentos usando dados reais de um serviço da *Web* - mercado eletrônico, com o qual avaliamos diferentes modelos de credibilidade usando diferentes tipos de fonte de informação, tais como atributos relacionados às ofertas, expertise do vendedor e sua qualificação. Os resultados mostram a aplicabilidade do arcabouço proposto através da análise de diferentes modelos de credibilidade. Além disso, a análise dos resultados permitiu obter observações que favorecem o entendimento do estudo de caso e direções futuras para a pesquisa.

O restante deste artigo está organizado da seguinte forma. A seção 2 descreve alguns trabalhos relacionados. Na seção 3 é apresentada uma formulação para o problema da credibilidade e o arcabouço para a implementação e avaliação de modelos de credibilidade. A seção 4 descreve o estudo de caso, usando uma amostra significativa de dados de um grande provedor latino-americano de serviços da *Web*. Finalmente, a seção 5 apresenta as conclusões e trabalhos futuros.

2. TRABALHOS RELACIONADOS

O que faz com que um usuário A ou B confie mais ou menos em um serviço da *Web* fornecido pelo usuário C ou D? Perguntas como essa motivaram o desenvolvimento de pesquisas na área de credibilidade. O conceito de credibilidade foi introduzido no final da década de 90 para avaliar a capacidade de um usuário acreditar em um sistema computacional, e definido como a percepção do usuário em relação ao objeto de análise. A maioria dos pesquisadores concorda que essa percepção é resultado da análise de várias dimensões do objeto de interesse, sendo as duas principais a confiança e a expertise (experiência, competência) [27].

Nos últimos anos, o conceito de credibilidade passou a ser es-

tudado na *Web*, visando medir o quanto um usuário confia em um serviço ou informação disponível. É consenso na literatura que a credibilidade pode ser subjetiva ao usuário, mas também depende de medidas objetivas. A credibilidade na *Web* se tornou um assunto multidisciplinar, com pesquisadores das áreas de comunicação focando em uma avaliação mais qualitativa (e subjetiva) da credibilidade [8], enquanto pesquisadores da área de ciência da computação tem focado em métricas mais objetivas. Os métodos propostos na área de ciência da computação são fortemente baseados em confiança (*trust*) e reputação e redes de citação [12], além de *rankings* de credibilidade que levam em conta a fonte da informação [1] e conteúdo [17].

Mecanismos de reputação são baseados em opiniões virtuais, fornecidas por pessoas que geralmente não se conhecem pessoalmente. Portanto a confiança neste meio eletrônico é mais difícil de ser estabelecida, se comparado ao mundo real. Em geral, a reputação de um comprador representa a probabilidade de pagamento do bem adquirido e a reputação de um vendedor representa a probabilidade de entrega bem sucedida do item anunciado (produto que foi comprado) após o pagamento [14]. Essas probabilidades estão relacionadas com confiança [20].

Mercados eletrônicos estão se tornando cada dia mais populares. Muitos trabalhos investigam sistemas de reputação e como eles induzem ao comportamento cooperativo. Dellarocas [6] realizou uma profunda revisão sobre isso. Ao mesmo tempo que fornece incentivos para um bom comportamento dos usuários, sistemas de reputação também podem fomentar comportamentos enganosos.

Klos et al. [18] analisaram o efeito da confiança e da reputação sobre os lucros obtidos através de intermediários em relações comerciais eletrônicas. Diferentes esquemas de propagação de confiança e desconfiança em comércio eletrônico foram estudados e avaliados em Guha et al. [13]. Resnick et al. [24] mostraram que os vendedores com reputação mais elevada conseguem vender seus produtos mais facilmente, entretanto os ganhos nos preços finais são reduzidos.

Resnick e outros autores [23] explicam que esses sistemas de reputação apresentam três problemas principais: (i) os compradores têm pouca motivação para fornecer *feedback* aos vendedores; (ii) é difícil de obter *feedback* negativo, pois é comum que as pessoas procurem negociar e resolver os problemas antes de preencher a avaliação no sistema; (iii) é difícil assegurar avaliações honestas. Uma vez que é muito simples se cadastrar em tais sistemas, é muito fácil criar uma identidade falsa que pode ser usada para comercializar com outros usuários e distorcer o sistema de reputação.

Como o sistema de *feedback* é a base da reputação nos mercados eletrônicos e fornece a informação que é usada antes que a transação aconteça, é fácil para fraudadores realizarem transações artificiais para que eles possam alcançar uma boa pontuação no sistema de reputação. Basicamente, essa pontuação artificial pode ser usada para enganar compradores que pagam e não recebem o produto certo ou pode ser usada como meio para vender mais, já que o vendedor poderá ter uma reputação favorável em relação a outros no mercado [23]. Considerando esta situação, os mercados devem dispor de ferramentas para identificar fraudadores a fim de proteger os usuários honestos. Os usuários que interagem com fraudadores podem também ter sua reputação afetada [20]. Gavish e Tucci [9] mostram que os compradores que são vítimas de fraudes reduzem o seu volume de transações, o que traz prejuízos para os mercados.

Os trabalhos aqui apresentados sugerem a crescente necessidade de prover novos modelos de credibilidade que forneçam subsídios a usuários de serviços online para que possam atuar com mais segurança e confiança na *Web*.

3. DEFINIÇÃO DO ARCABOUÇO

Nesta seção nós descrevemos o *Credibility Rank*, um novo arcabouço para o projeto e análise de modelos de credibilidade. Esse arcabouço, baseado em uma nova modelagem do problema da credibilidade, é a principal contribuição deste trabalho, sendo que neste trabalho estudamos o problema da credibilidade no contexto de uma aplicação *Web*.

O arcabouço concebido é um conjunto de módulos de software elaborados como classes (seguindo paradigma de orientação a objetos) capaz de executar algumas tarefas descritas a seguir: descrever e instanciar um modelo de credibilidade, ler um conjunto de atributos de credibilidade, processar esses atributos e aplicar um modelo de credibilidade a estes dados ou transações, produzindo um *ranking* de credibilidade. Além disso, existe um módulo responsável por realizar a avaliação da qualidade de um determinado modelo, utilizando para isso alguma variável resposta, como o *feedback* das transações no contexto do estudo de caso que realizamos. A proposta desse arcabouço é evoluir gradativamente, tornando-se um ferramental mais completo para projetar, executar e avaliar modelos de credibilidade de serviços baseados em transações.

A credibilidade de um serviço é associada a alguma informação que caracteriza ou descreve esse serviço. Essa informação pode estar relacionada com a configuração do serviço e do usuário que a oferece. Quando alguém usa um serviço, uma transação, que contém informações sobre o serviço e o seu fornecedor, é registrada. A partir de cada transação de um serviço é possível extrair um conjunto de informações que pode ser usado para modelar a credibilidade. Além disso, a partir da qualificação que um usuário provê a uma transação, é possível medir o seu grau de satisfação com ela. A partir dessa explicação geral, descrevemos algumas entidades que serão usadas para definir conceitos importantes sobre credibilidade e que serão adotados neste trabalho:

- **Aplicação:** é o site intermediário onde serviços são providos por fornecedores, como uma biblioteca digital ou um mercado eletrônico.
- **Serviço:** é o serviço online sendo oferecido, que pode ser caracterizado por um conjunto de atributos. Um serviço é descrito em termos de suas características (como preço e duração em um comércio eletrônico, ou número de visualizações e *tags* (rótulos) em um sistema de compartilhamento de vídeos online) e informações do fornecedor (como reputação, tempo de cadastro).
- **Fornecedor:** usuário que fornece um serviço. Exemplos podem ser o dono de uma oferta de um produto em um comércio eletrônico, o autor de um conteúdo publicado em um *site* de *microblogging* ou um fórum.
- **Usuário:** quem usa um serviço, o que gera uma transação, como comprar uma oferta de um produto em um mercado eletrônico ou assistir um vídeo online.
- **Transação:** é o registro do uso de um serviço, que gera um evento que está relacionado com um conjunto de informações associadas ao serviço e a seu fornecedor.
- **Feedback:** é a opinião do usuário ou o grau de satisfação com o serviço usado. Cada transação tem um *feedback*.

Por exemplo, um mercado eletrônico como o eBay é uma aplicação em que vendedores são fornecedores, ofertas são serviços, negociações são transações e compradores são usuários. Em um site de compartilhamento de vídeos, como o Youtube, um vídeo é

um serviço provido por um fornecedor e uma visualização é uma transação entre um usuário e o fornecedor. Muitos *websites* rentáveis e populares, tais como Youtube, Wikipedia, Flickr, Myspace e eBay suportam essas interações entre usuários e fornecedores como intermediários. Consequentemente, mecanismos eficientes que permitam que o usuário avalie a credibilidade de serviços são de fundamental importância para o sucesso dessas aplicações.

A fim de estimar a qualidade dos mecanismos de avaliação da credibilidade, nós propomos um arcabouço para aplicação e análise de modelos de credibilidade. Um modelo de credibilidade é uma função que permite a comparação entre serviços em termos de credibilidade. Um modelo de credibilidade efetivo é capaz de indentificar quais são os serviços mais confiáveis e denunciar quais os de menor confiabilidade para um dado usuário. Entretanto, o projeto de modelos de alta qualidade apresenta vários desafios, tais como:

1. **Informação limitada:** A comunicação através da *Web* é bastante limitada quando comparada ao mundo real. A maioria das interações na *Web* são puramente virtuais, baseadas em descrições textuais, fotos, identidades online e outras informações providas pela aplicação. A falta de informação considerando tanto serviços quanto fornecedores é uma das principais dificuldades relacionadas à credibilidade em serviços online [23, 24, 25].
2. **Inexistência de classes bem definidas:** A confiabilidade não é uma variável binária sob a perspectiva do usuário. A decisão do usuário é normalmente baseada em um compromisso entre fatores. Em um comércio eletrônico, por exemplo, compradores podem estar dispostos a assumir riscos na credibilidade em troca de menores preços ou escolher dar prêmios de preço para um vendedor com alta reputação [25]. Dessa forma, uma escala que permite a avaliação da credibilidade pode assistir as decisões de usuários de serviços *Web*.
3. **Necessidade de agregar informação de credibilidade:** Aplicações *Web* que suportam interações entre usuários geralmente possuem informações relacionadas à credibilidade do serviço. Por exemplo, muitas bibliotecas digitais mostram o número de visualizações e citações de artigos publicados e também informações sobre autores, como afiliação, número de publicações e citações. Embora essas informações possam ser usadas a fim de avaliar a credibilidade do serviço na *Web*, a tarefa de agregar tais informações é normalmente deixada para o usuário.
4. **Informações pouco confiáveis:** A maior parte da informação sobre serviços online providos na *Web* está sujeita à inconsistências e fraudes. Reputações artificialmente infladas e troca de pseudônimos, por exemplo, têm sido usados para promover usuários e apagar históricos de *feedbacks* negativos. Além disso, descrições enganosas de serviços são causas comuns de transações mal sucedidas na *Web* [9, 10, 11, 24].

A maioria das aplicações *Web* existentes que suportam interações entre usuário provêem mecanismos para medir a credibilidade de usuários e serviços (ex. sistemas de reputação) [23]. Entretanto, tais mecanismos têm sua eficácia afetada pelos desafios descritos acima. Considerando o ponto de vista do usuário, selecionar um serviço confiável na *Web* tem se tornado uma tarefa difícil, reduzindo a satisfação dos usuários e limitando a popularidade dessas aplicações [10, 11]. Além disso, quando serviços ruins podem resultar em perdas financeiras aos usuários, como em um mercado eletrônico, assegurar a credibilidade de serviços providos se torna ainda mais crítico.

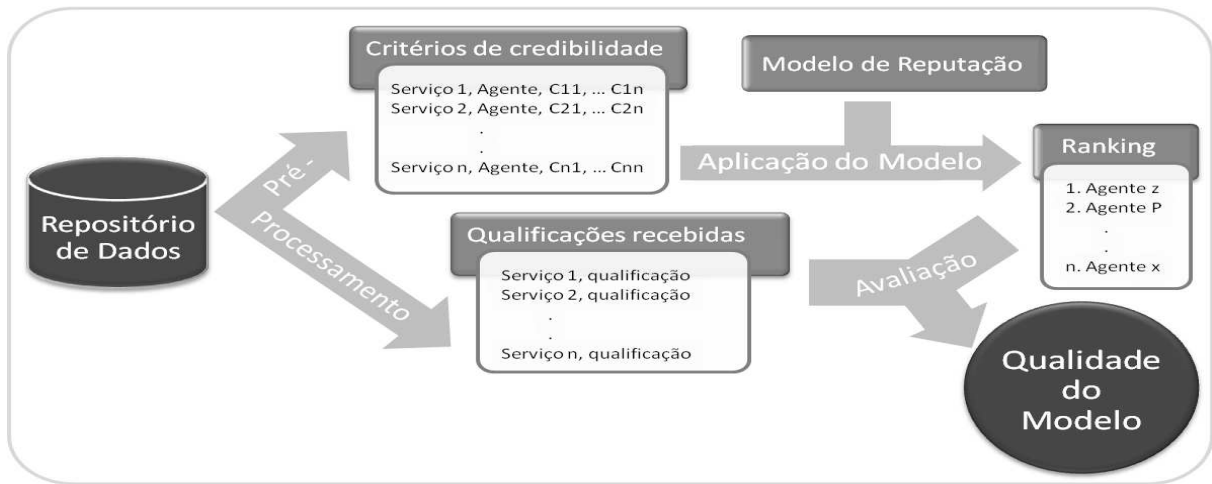


Figura 1: Esquema do Arcabouço *CredibilityRank*

O primeiro desafio mencionado está mais relacionado ao projeto de novas interfaces para reduzir a barreira na comunicação entre usuários que interagem na *Web*, o que é um problema importante na área de Interações Humano-Computador (IHC) e Trabalho Colaborativo Apoiado por Computador (CSCW). Embora a falta de informação tenha um papel importante na medição da credibilidade de um serviço, projetar novas interfaces *Web* não é o foco deste trabalho. Os desafios 2, 3 e 4 tratam do uso efetivo da informação disponível acerca de serviços e fornecedores com o propósito de prover mecanismos para avaliação da credibilidade na *Web*. Para avaliar esses problemas, nós propomos uma definição formal de um modelo de credibilidade como uma função capaz de ordenar serviços em termos de credibilidade. Com base nessa formulação, nós projetamos um arcabouço para a aplicar e avaliar modelos de credibilidade.

Definição Um modelo de credibilidade \mathcal{M} é uma função que recebe um conjunto de serviços $S = s_1, s_2, \dots, s_n$, sendo n o número de serviços e s_i a tupla de atributos do serviço i , e retorna uma ordenação (*ranking*) R , onde serviços são posicionados em termos de credibilidade baseado nos seus atributos. Um *ranking* pode ser descrito por uma função $R : i \rightarrow \mathbb{Z}$, em que $R(i)$ é a posição do serviço i e $0 \leq R(s_j) < n, \forall s_j \in S$. Quanto maior a credibilidade de um dado serviço i de acordo com \mathcal{M} , menor o valor de $R(i)$.

Uma tupla s_i pode ser composta de diferentes atributos do serviço relacionados ao seu fornecedor e também ao próprio serviço avaliado. Por exemplo, em um site de compartilhamento de vídeo, um serviço s_i pode conter informação sobre há quanto tempo o fornecedor que disponibilizou o vídeo está registrado no site, quantos vídeos ele disponibilizou e também o número de visualizações e avaliações positivas recebidas pelo vídeo. Baseado nessas informações, o modelo de credibilidade pode permitir ao usuário distinguir quais vídeos são *spams* ou discordam da descrição mostrada.

O principal objetivo de um modelo de credibilidade é agregar atributos do serviço a fim de prover uma avaliação efetiva da credibilidade. Além disso um modelo de alta qualidade deve basear-se nos atributos mais confiáveis a fim de atender com maior exatidão as expectativas do usuário. O *ranking* de serviços em termos de credibilidade pode ser visto como uma escala de credibilidade, o que não se restringe a uma classificação binária (confiável, não confiável), mas provê uma comparação entre serviços concor-

rentes. Além disso, a identificação de *rankings* é um problema amplamente estudado em Recuperação de Informações e Aprendizado de Máquinas.

Dado um modelo de credibilidade \mathcal{M} , é importante avaliar sua efetividade. Com base na avaliação de modelos de credibilidade, é possível comparar modelos existentes e selecionar o mais eficaz para uma aplicação específica. Para quantificar a qualidade de um modelo de credibilidade, nós propomos um novo arcabouço experimental baseado na ordenção de serviços em termos de credibilidade. O arcabouço proposto utiliza um *log* de transações de serviços em aplicações *Web* para avaliar os modelos de credibilidade.

O arcabouço recebe um *log* de transações \mathcal{D} (repositório de dados) e um modelo de credibilidade \mathcal{M} . O *log* \mathcal{D} contém transações de serviços e avaliações de credibilidade dadas por usuários. Os serviços são avaliados através de *feedbacks* atribuídos por usuários para a respectiva transação. Sistemas de *feedback* são frequentemente empregados para gravar opiniões de usuários a respeito de serviços providos através de aplicações *Web*. O arcabouço proposto faz uso desses *feedbacks* para avaliar a efetividade dos modelos de credibilidade. Dado um *ranking* de serviços baseado em um modelo de credibilidade \mathcal{M} , o arcabouço avalia \mathcal{M} em termos de sua capacidade de posicionar serviços que recebem altas qualificações no topo das escalas (*rankings*).

A Figura 1 ilustra os passos seguidos por nosso arcabouço. Ele simula a interação do usuário com o modelo de credibilidade a fim de avaliar a qualidade dos *rankings* gerados. Primeiramente, um histórico de transações é dividido em tuplas de serviço (ex. os atributos do serviço que originaram cada transação) e qualificações (*feedbacks*) recebidos. A seguir as tuplas de serviço são ordenadas pelo modelo de credibilidade. A qualidade do modelo é avaliada usando a qualificação dos usuários em um passo final.

Uma importante característica dos atributos do serviço é que a maioria deles são acumulativos. O número de visualizações em um site de compartilhamento de vídeo, a reputação de um vendedor em um mercado eletrônico, o número de citações de um autor em uma biblioteca digital, entre outros, são variáveis cujo valor aumenta ao longo do tempo, o que torna irrealística a aplicação de modelos de credibilidade em serviços providos em diferentes épocas. Comparar a reputação de dois vendedores que negociaram em diferentes anos, por exemplo, não é justo, já que uma reputação alta no primeiro ano do mercado pode ser considerada muito baixa alguns anos depois. A fim de evitar esse efeito temporal, nós fix-

amos um intervalo de ativação para cada serviço e ordenamos apenas serviços que estavam ativos durante o mesmo intervalo. Para cada tupla de serviços s_i , nós ordenamos o conjunto de serviços ativos enquanto s_i está ativo e gravamos ambas as posições de s_i no *ranking* e o *feedback* recebido por cada transação de s_i . A avaliação global de um modelo de credibilidade é baseada nos *feedbacks* recebidos pelas transações dos serviços que estão no topo em todos os *rankings* gerados.

Algoritmo 1: Arcabouço CredibilityRank

Entrada: \mathcal{M}, \mathcal{D}
Saída: Avaliação de \mathcal{M}
 $\mathcal{F} \leftarrow \text{extrai_feedbacks}(\mathcal{D});$
 $\mathcal{T} \leftarrow \text{extrai_transacoes}(\mathcal{D});$
 $R \leftarrow \emptyset;$
for $t \in \mathcal{T}$ **do**
 $A \leftarrow \text{services_ativos}(t, \mathcal{T});$
 $s \leftarrow \text{obtem_servico}(t);$
 $R \leftarrow \mathcal{M}(s, A);$
 $S \leftarrow S \cup (\text{posicoes}(s, R), F_t);$
Avalia(S);

O Algoritmo 1 é uma descrição em alto-nível de *CredibilityRank*, o arcabouço proposto neste trabalho. Além da extração da tupla de serviços e da avaliação (*feedback*) do log de transações \mathcal{D} , o pseudo-código descreve como o arcabouço itera sobre o conjunto de transações em \mathcal{D} . Para cada transação t , o algoritmo identifica o conjunto de serviços A que estavam ativos quando t foi efetuado. Também é necessário indentificar o serviço específico s associado a t . Através da aplicação do modelo de credibilidade \mathcal{M} recebendo A e o serviço s , nós obtemos a ordenação (*ranking*) R . As posições dos serviços no *ranking* ($\text{posicoes}(s, R)$) e *feedbacks* F_t recebidos por cada transação são incluídos em um conjunto S , que será usado na avaliação de \mathcal{M} .

O arcabouço não define um procedimento específico de avaliação, já que isso depende da aplicação *Web* analisada. Entretanto, várias métricas de avaliação baseadas em *rank* podem ser usadas, dependendo do cenário de aplicação. Exemplos de métricas incluem:

- Número de *feedback* negativo/positivo nas top-K posições do *rank*¹ [2, 21].
- *Feedback* médio nas top-K posições do *rank*.
- Métricas baseadas em ganho cumulativo descontado, tais como NDCG (*Normalized Discounted Cumulative Gain*), que dá maiores pontuações a resultados corretos que apareceram anteriormente no *ranking* [15].

A próxima seção apresenta um estudo de caso onde aplicamos e avaliamos alguns modelos de credibilidade baseados em *rank* utilizando o arcabouço *CredibilityRank* e uma base de dados atual e representativa.

4. ESTUDO DE CASO

Esta seção apresenta o nosso estudo de caso onde aplicamos e avaliamos alguns modelos de credibilidade baseados em *ranking* utilizando o arcabouço *CredibilityRank* e dados reais de um mercado eletrônico brasileiro.

4.1 Descrição da Base de Dados

Esta seção descreve o *TodaOferta*¹, um mercado eletrônico de-

¹Similar à precisão em N (P@N).

¹<http://www.todaoferta.com.br>

envolvido pelo maior provedor de serviços da Internet na América Latina, chamado Universo Online Inc. (UOL)². Ela também define alguns conceitos básicos relacionados ao mercado.

O *TodaOferta* [22] é um *website* para compra e venda de produtos através da *Web*. A Tabela 1 mostra um resumo da base de dados do *TodaOferta*, que diz respeito a um número significativo de usuários, anúncios e negociações. Porém, devido à confidencialidade dos dados, é importante ressaltar que informações quantitativas não serão apresentadas neste artigo.

Cobertura (tempo)	Jun/2007 a Jul/2009
#categorias	32
#sub-categorias	2.189
Média de anúncios por vendedor	42,48

Tabela 1: Resumo da amostra de dados

No *TodaOferta*, compradores são usuários, ofertas são serviços e vendedores são fornecedores. O sistema de reputação do *TodaOferta* é relativamente simples. Após a negociação, compradores e vendedores são convidados a qualificar a outra parte de forma positiva (1), negativa (-1) ou neutra (0). A reputação dos usuários é definida como a soma de todas as qualificações recebidas por ele. Os *feedbacks* de um mesmo usuário só são consideradas uma única vez para computar uma reputação. Sistemas de reputação são úteis para transmitir confiança em aplicações de comércio eletrônico. Entretanto, o *TodaOferta* provê outras informações sobre os vendedores e compradores que podem ser utilizadas para identificar serviços confiáveis ou não (ex. tempo de cadastro do vendedor, comentários deixados por usuários que negociariam com o vendedor).

4.2 Experimentos

Nesta seção nós aplicamos o arcabouço *CredibilityRank* para avaliar alguns modelos de credibilidade usando a base de dados do *TodaOferta*. O objetivo principal dessa avaliação experimental usando uma base de dados real é promover um estudo da efetividade de diferentes informações existentes sobre vendedores e ofertas como fontes de credibilidade.

Dado um atributo a_i , definimos um modelo de credibilidade \mathcal{M}_i que ordena os serviços de acordo com a_i . A menos que seja dito o contrário, quanto maior o valor de a_i para um dado serviço s , mais alta a posição de s no *ranking*. Os resultados obtidos por esses modelos que consideram apenas um atributo são importantes para motivar o projeto de novos modelos de credibilidade que combinem atributos do serviço a fim de quantificar com exatidão a credibilidade dos serviços. É importante enfatizar que os atributos para modelar a credibilidade podem ser um conjunto de atributos combinados, não apenas modelos únicos. Nós selecionamos os 12 atributos a seguir, divididos em 5 categorias, para serem usados no *ranking* de serviços:

- Características da Oferta
 - **Preço:** preço do produto/serviço sendo oferecido.
 - **Duração:** duração da oferta (anúncio) estabelecido pelo vendedor (em dias).
 - **Destaque:** indica se a oferta foi anunciada com destaque (pacotes de anúncio especiais).
 - **Visualizações:** o número de visualizações da oferta.
- Transação Segura

²<http://www.uol.com.br>

- **Oferta com Pagamento Seguro:** indica se uma oferta tem a opção de utilizar o mecanismo de pagamento seguro provido pelo mercado eletrônico.
- **Transação Segura:** identifica a transação que foi efetuada adotando o mecanismo de pagamento seguro.
- Expertise do Vendedor
 - **Itens Vendidos:** o montante total de itens que o vendedor já vendeu no mercado.
 - **Tempo de cadastro:** há quanto tempo o vendedor está registrado no mercado eletrônico.
- Qualificação do Vendedor
 - **Qualificação Positiva:** o montante de qualificações positivas que o usuário (vendedor) recebeu.
 - **Reputação:** a soma das qualificações do vendedor, considerando diferentes tipos de qualificações recebidas.
- Tipo de Vendedor
 - **Megavendedor:** indica se o usuário é considerado um vendedor de destaque pelo *TodaOferta*.
 - **Selo Qualidade:** denota o vendedor que possui um certificado de qualidade, provido por uma empresa à parte.

Cada um dos atributos descritos é usado por um diferente modelo de credibilidade. Um modelo ordena os serviços ativos, considerando as transações (ex. uma venda feita no mercado eletrônico), de acordo com o atributo dado. Para atributos binários, como *Megavendedor* e *Transação Segura*, o modelo considera um vendedor que apresenta o critério mais confiável que um que não o apresenta.

A métrica aplicada na avaliação dos modelos de credibilidade é a probabilidade de receber *feedback* negativo nas $X\%$ primeiras posições do *ranking*. Essa é uma métrica de avaliação intuitiva que apresenta como importante característica o fato de ser mais confiável que a de *feedback* positivo na maioria dos mercados eletrônicos, já que ataques típicos tentam inflacionar a reputação de vendedores através de *feedbacks* positivos. Como a métrica selecionada considera apenas informação sobre *feedbacks* negativos, o que pode ser assumido como confiável, ela pode prover uma análise com maior acurácia da qualidade dos modelos de credibilidade.

A Figura 2 mostra os resultados obtidos pelos modelos de credibilidade de atributos não binários. Os gráficos estão agrupados de acordo com os critérios. Os resultados para atributos binários (ex. *Destaque*, *Selo de Qualidade*) serão apresentados separadamente. É importante notar que os resultados são acumulativos de acordo com as posições do *ranking* (i.e. de todos os *feedbacks* negativos recebidos, quantos aparecem nas primeiras $X\%$ posições). É útil comparar os resultados obtidos por cada critério com a função $y = x$, que provê os resultados de um modelo de credibilidade hipotético e randômico. Um modelo que gera uma curva similar à do modelo randômico pode ser considerado inefetivo.

A Figura 2(a) apresenta a probabilidade de receber *feedback* negativo nas primeiras $X\%$ posições do *ranking*, considerando os atributos relacionados à característica da oferta (exceto o atributo binário *Destaque*). O valor de *Preço* é normalizado para diferentes produtos baseado nas suas categorias. Ofertas de produtos com menores preços estão localizadas nas posições do topo do *ranking*. O resultado mostra que ofertas com preços altos estão normalmente associadas com transações que recebem *feedback* negativo. Levantamos três hipóteses para esse resultado: (1) Dado um produto P , os anúncios que o ofertam por menores preços são mais confiáveis, o que é

contra-intuitivo, (2) usuários têm maior tendência de dar *feedback* negativo quando compram produtos mais caros e (3) itens baratos, que representam grandes volumes de venda, atraem vendedores profissionais, que são mais capazes de prover serviços satisfatórios. Considerando a *Duração* da oferta, podemos ver que as primeiras posições do *ranking* indicam uma pequena porção de qualificações negativas. Por outro lado, as últimas posições mostram uma taxa crescente da probabilidade de *feedback* negativo. Esse resultado sugere que usuários confiam mais em anúncios de produtos que dêem mais tempo para que o comprador decida comprar ou não o produto, e não agir compulsivamente devido ao pequeno período de duração. Analizando o atributo *Visualizações*, concluímos que ele não é um bom critério para qualificar a credibilidade das transações nesse mercado, já que nas posições do topo do *ranking* há um número significativo de *feedback* negativo, que continua crescendo constantemente e depois apresenta um pequeno decréscimo no final.

Considerando a expertise do vendedor, a Figura 2(b) mostra a probabilidade de receber qualificação negativa para dois outros atributos: *Itens Vendidos*, e *Tempo de Cadastro*. Ambos mostram resultados similares, uma pequena probabilidade de *feedback* negativo nas posições do topo do *ranking* e também um pequeno crescimento dessa probabilidade nas próximas posições dele. Nas posições finais do *ranking* (após 70% dele - eixo X) podemos observar um alto crescimento dessa probabilidade. Comparando-os, podemos ver que *Itens Vendidos* é melhor que *Tempo de Registro* até o meio do *ranking*, exceto nas primeiras posições. Também podemos ver que ambos estão abaixo da curva linear representada por $y=x$.

A Figura 2(c) apresenta a probabilidade de *feedback* negativo para atributos agrupados de acordo com a *Qualificação do Vendedor*. Ambos são bons indicadores de credibilidade na primeiras posições do *ranking*, como pode ser visto no gráfico. Comparando-os, podemos ver que *Qualificação Positiva* é melhor, já que concentra um maior número de *feedbacks* negativos nas últimas posições do *ranking*. Além disso, os dois estão abaixo da curva linear $y=x$.

A Tabela 2 apresenta uma análise comparativa entre os 7 atributos não-binários discutidos. Para cada comparação proposta, a tabela mostra a porcentagem de *feedback* negativo nas 20% primeiras e últimas posições do *ranking*, divididas entre intervalos iguais de 10% cada. Um bom atributo de credibilidade tem baixa probabilidade de receber *feedback* negativo nas primeiras posições e uma alta porcentagem nas últimas posições do *ranking*, já que isso indica que as qualificações negativas estão mais concentradas nas posições finais, e há poucas no topo do *ranking*.

Atributos	Probabilidade de <i>Feedback</i> Negativo (%)			
	Primeiras Posições (%)		Últimas Posições (%)	
	0-10	10-20	90-100	80-90
Preço	5.48	11.25	19.23	44.42
Duração	4.97	7.91	22.81	10.09
Visualizações	18.08	9.86	7.28	8.63
Itens Vendidos	9.04	4.76	18.61	14.68
Tempo de Cadastro	7.81	7.93	20.08	15.18
Feedback Positivo	8.21	3.69	35.50	14.42
Reputação	9.05	4.52	14.21	4.11

Tabela 2: Porcentagem de *Feedback* Negativo nas Primeiras e Últimas posições do *Ranking* de Credibilidade

Os resultados são consistentes com os obtidos na Figura 2. *Preço* é um bom indicador, já que nas 10% últimas posições do *ranking* há 44.42% de todos os *feedbacks* negativos, e nas suas primeiras posições, só 5.48%. O atributo *Duração* também é um bom indicador de credibilidade, com 4.97% de *feedbacks* negativos nas

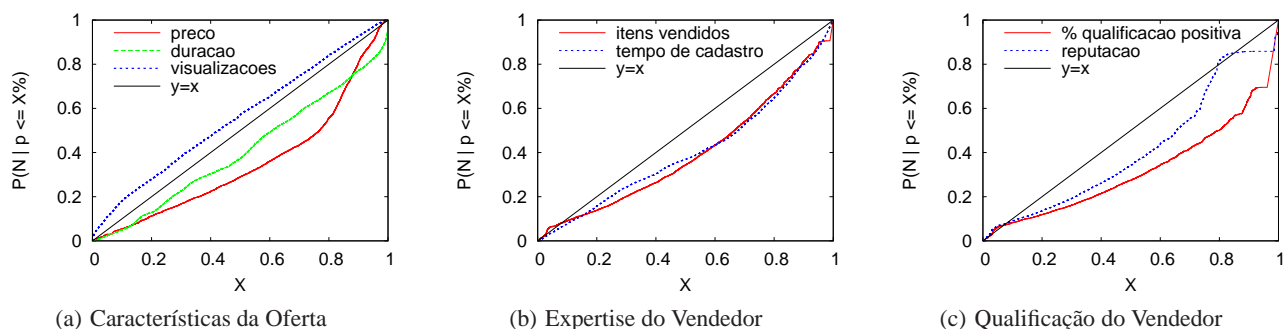


Figura 2: Modelos de Credibilidade - Probabilidade de feedback negativo nas primeiras X% posições do ranking

primeiras 10% posições do ranking e quase 23% nas últimas. A *Qualificação Positiva* é um indicador muito bom também, apresentando baixa porcentagem de *feedback* negativo nas primeiras 10% posições (8.21%) e uma porcentagem muito alta nas últimas posições do ranking (35.50%). Semelhantemente, podemos concluir que *Itens Vendidos* e *Tempo de Cadastro* são bons indicadores, mas não tanto quanto os descritos anteriormente. A *Reputação* não pode ser considerada um atributo bom ou ruim para modelar a credibilidade quando comparados os valores do topo e da parte inferior do ranking. Já o atributo *Visualizações* tem um resultado diferente. Ele não pode ser visto como bom indicador de credibilidade, pois tem altas porcentagens de *feedbacks* negativos nas primeiras posições e baixas nas últimas.

A Tabela 3 mostra a análise de atributos binários de credibilidade. Para cada um deles, calculamos a probabilidade de receber *feedback* negativo considerando a situação em que esses atributos estão ativos (ou verdadeiros). Observamos que os atributos do grupo *Tipo de Vendedor* são os melhores. Vendedores que possuem o *Selo de Qualidade* têm apenas 2.99% de probabilidade de receber *feedback* negativo. *Megavendedores* têm 17.78% dessa probabilidade, o que também é um bom indicador. Considerando os mecanismos de pagamento seguro, pode-se observar que prover essa opção não é garantia de credibilidade, entretanto transações que realmente adotaram o mecanismo apresentam uma porcentagem relativamente pequena de receber *feedback* negativo. O resultado do atributo *Destaque* não é tão conclusivo como os demais.

Atributos	Probabilidade de <i>Feedback</i> Negativo (%)
Selo de Qualidade	2.99
Megavendedor	17.78
Transação Segura	20.06
Destaque	30.32
Oferta com Pagamento Seguro	48.90

Tabela 3: Porcentagem de *Feedback* Negativo para Atributos Binários

Nesta seção nós apresentamos nosso estudo de caso e a análise dos resultados obtidos pela aplicação do *CredibilityRank* usando modelos diferentes, cada um deles usando um atributo específico para gerar um ranking de credibilidade das transações de uma aplicação *Web*, que é um mercado eletrônico. A partir dos rankings gerados, nós propomos um método de avaliação que consiste na quantificação do total de *feedbacks* recebidos nas primeiras X% posições para cada modelo de credibilidade proposto. Os resultados mostram que quanto menor a porcentagem nas posições do topo e maior o montante de *feedback* negativo nas posições finais, melhor o modelo é. A partir do conjunto de modelos de credibilidade que nós avaliamos, aqueles que geraram os melhores resul-

tados são os relacionados à *Expertise do Vendedor*, *Qualificação do Vendedor* e *Tipo do Vendedor*, juntamente com alguns atributos individuais, como *Preço*, *Duração* e *Transação Segura*. Os atributos *Visualizações* e *Oferta com Pagamento Seguro* não foram tão conclusivos, ou até mesmo mostraram-se ruins para indicar a credibilidade do serviço. Esses resultados motivam trabalhos futuros, mostrando que há muito mais a analisar e concluir sobre esses modelos de credibilidade e como podemos combiná-los para gerar outros modelos que podem ser mais confiáveis e ajudar usuários a executar transações seguras na *Web*.

5. CONCLUSÃO

Neste trabalho, estudamos o problema da credibilidade em serviços da *Web*. Nesse contexto, **provedores** oferecem **serviços** a **usuários** em qualquer **aplicação Web 2.0**. A tarefa de avaliar a credibilidade de um serviço na *Web* representa um problema de pesquisa com vários desafios, como: informação limitada, inexistência de classes bem definidas, necessidade de agregar informação de credibilidade e baixa confiabilidade das informações disponíveis.

O principal objetivo da nossa pesquisa é propor o arcabouço (nomeado *CredibilityRank*) para projetar, implementar e avaliar de modelos de credibilidade. Denominamos um modelo de credibilidade uma função capaz de atribuir um valor de credibilidade a uma transação de um serviço da *Web*, considerando diferentes critérios associados a esse serviço e seu fornecedor. Essa modelagem permite a definição ou projeto de sistemas (ou modelos) de credibilidade que podem gerar escalas (*rankings*) de credibilidade baseados em vários critérios diferentes. Os valores de credibilidade dos serviços, representados por transações, criam uma escala que nos permite avaliar esses modelos baseado em métricas de avaliação de *rankings*, amplamente utilizadas em Recuperação de Informações.

Nós executamos experimentos para avaliar diversos modelos de credibilidade usando tipos diferente de fontes de informação e uma base de dados atual de um mercado eletrônico - o *TodaOferta*. Os resultados mostraram a aplicabilidade do arcabouço proposto. Além disso, a análise dos resultados proveu informações que contribuem para entender o cenário do estudo de caso e para identificar direções para trabalhos futuros.

Essa nova modelagem do problema da credibilidade em serviços da *Web* visa promover mecanismos de credibilidade mais eficientes, uma vez que permite a combinação de vários atributos considerados relevantes no cálculo da confiabilidade do serviço. Essa formulação considerando um ranking é mais apropriada pois não restringe o problema a simplesmente classificar o serviço como confiável ou não. Além disso, o arcabouço permite a comparação e avaliação de diversos modelos de credibilidade. Isso é importante já que difer-

entes modelos podem ser propostos para uma dada aplicação.

Como trabalhos futuros, pretendemos gerar e avaliar novos modelos de credibilidade que combinem alguns dos modelos que apresentamos e analisamos neste trabalho. A partir dos resultados é possível inferir quais combinações têm maior chance de serem bem sucedidas. Além disso, implementaremos novos modelos de credibilidade baseados em algumas técnicas já propostas na literatura, como SVM-Rank [28]. Por fim, queremos também aplicar o arcabouço em outros cenários da *Web*, como bibliotecas digitais, redes sociais e aplicações que provêem conteúdo para usuários.

6. AGRADECIMENTOS

Esta pesquisa foi parcialmente patrocinada por Universo OnLine S. A. - UOL (www.uol.com.br) e parcialmente apoiada pelo Instituto Nacional de Ciência e Tecnologia para a Web (CNPq no. 573871/2008-6), CAPES, CNPq, Finep e Fapemig.

7. REFERÊNCIAS

- [1] A. Amin, J. Zhang, H. Cramer, L. Hardman, and V. Evers. The effects of source credibility ratings in a cultural heritage information aggregator. In *WICOW '09: Proceedings of the 3rd workshop on Information credibility on the web*, pages 35–42, New York, NY, USA, 2009. ACM.
- [2] R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.
- [3] F. Benevenuto, A. Pereira, T. Rodrigues, V. Almeida, J. Almeida, and M. Gonçalves. Avaliação do perfil de acesso e navegação de usuários em ambientes web de compartilhamento de vídeos. In *WebMedia '09: Proceedings of the 15th Brazilian Symposium on Multimedia and the Web*, New York, NY, USA, 2009. ACM.
- [4] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and M. Gonçalves. Detectando usuários maliciosos em interações via vídeos no youtube. In *WebMedia '08: Proceedings of the 14th Brazilian Symposium on Multimedia and the Web*, pages 138–145, New York, NY, USA, 2008. ACM.
- [5] S. C. Cazella, E. Reategui, and L. O. C. Alvares. E-commerce recommenders' authority: applying the user's opinion relevance in recommender systems. In *WebMedia '06: Proceedings of the 12th Brazilian Symposium on Multimedia and the web*, pages 71–78, New York, NY, USA, 2006. ACM.
- [6] C. Dellarocas. Reputation mechanisms. In *Handbook on Economics and Information Systems*, pages 629–660. Elsevier Publishing, 2006.
- [7] D. Duarte, A. Pereira, and W. Meira Jr. Práticas de comercialização em mercados eletrônicos. In *Workshop de Trabalhos de Iniciação Científica - WTIC/WebMedia '09: Proceedings of the 15th Brazilian Symposium on Multimedia and the Web*, New York, NY, USA, 2009. ACM.
- [8] A. J. Flanagan and M. J. Metzger. The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media Society*, 9(2):319–342, April 2007.
- [9] B. Gavish and C. L. Tucci. Reducing internet auction fraud. *Commun. ACM*, 51(5):89–97, 2008.
- [10] D. G. Gregg and J. E. Scott. The role of reputation systems in reducing on-line auction fraud. *Int. J. Electron. Commerce*, 10(3):95–120, 2006.
- [11] D. G. Gregg and J. E. Scott. A typology of complaints about ebay sellers. *Commun. ACM*, 51(4):69–74, 2008.
- [12] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 403–412, New York, NY, USA, 2004. ACM.
- [13] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 403–412, New York, NY, USA, 2004. ACM.
- [14] D. Houser and J. Wooders. Reputation in auctions: Theory, and evidence from ebay. *Journal of Economics & Management Strategy*, 15(2):353–369, 06 2006.
- [15] K. Järvelin and J. Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 41–48, New York, NY, USA, 2000. ACM.
- [16] A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decis. Support Syst.*, 43(2):618–644, 2007.
- [17] A. Juffinger, M. Granitzer, and E. Lex. Blog credibility ranking by exploiting verified content. In *WICOW '09: Proceedings of the 3rd workshop on Information credibility on the web*, pages 51–58, New York, NY, USA, 2009. ACM.
- [18] T. Klos and F. Alkemade. Trusted intermediating agents in electronic trade networks. In *AAMAS '05: Proc. of the 4th intl joint conference on Autonomous agents and multiagent systems*, pages 1249–1250, New York, USA, 2005. ACM.
- [19] R. Maranzato, A. Pereira, A. P. do Lago, and M. Neubert. Fraud detection in reputation systems in e-markets using logistic regression. In *SAC '10: Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 1454–1459, New York, NY, USA, 2010. ACM.
- [20] M. I. Melnik and J. Alm. Does a seller's ecommerce reputation matter? evidence from ebay auctions. *Journal of Industrial Economics*, 50(3):337–49, September 2002.
- [21] A. Moffat and J. Zobel. Rank-biased precision for measurement of retrieval effectiveness. *ACM Trans. Inf. Syst.*, 27(1):1–27, 2008.
- [22] A. M. Pereira, D. Duarte, W. M. Jr., V. Almeida, and P. Góes. Analyzing seller practices in a brazilian marketplace. In *18th International World Wide Web Conference*, pages 1031–1041, April 2009.
- [23] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman. Reputation systems. *Commun. ACM*, 43(12):45–48, 2000.
- [24] P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *The Economics of the Internet and E-Commerce*, Elsevier Science B.V.:127–157, 2002.
- [25] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on ebay: A controlled experiment. *School of Information, University of Michigan*, Ann Arbor, Michigan, USA:34, 2003.
- [26] J. Sabater and C. Sierra. Review on computational trust and reputation models. *Artif. Intell. Rev.*, 24(1):33–60, 2005.
- [27] S. Tseng and B. J. Fogg. Credibility and computing technology. *Commun. ACM*, 42(5):39–44, 1999.
- [28] H. Yu. Svm selective sampling for ranking with application to data retrieval. In *KDD '05: Proc. of the 11th ACM SIGKDD intl conference on Knowledge discovery in data mining*, pages 354–363, New York, USA, 2005. ACM.