



Análise do Impacto da Rede Social na Marcação de Fotos como Favoritas no Flickr¹

Luiz Alves, Matheus Araújo, Lesandro Ponciano, Andrey Brito
{luiz.filho,matheus.macie1}@ccc.ufcg.edu.br {lesandrop, andrey}@lsd.ufcg.edu.br

Universidade Federal de Campina Grande
Laboratório de Sistemas Distribuídos

Technical Report No. UFCG-LSD-2012-01

2 de setembro de 2012

Abstract. *Understanding user behavior in user generated-content systems (UGCs) is a fundamental building block for maximizing usefulness of these systems. For example, the importance of users' connections and the degree of iteration is needed in order to build meaningful recommendation systems. In this paper, we evaluate the impact of users' network on marking photos as favorites in Flickr considering various levels of distance between users in the network topology. We have observed that users' first-degree contacts are clearly the biggest contributors to the evaluation of hers/his photos (i.e., marking them as favorites). This result indicates that, in contrast to other UGCs such as Twitter and Digg, contact relations are a good indication of content interests in Flickr.*

¹Este relatório apresenta detalhes do artigo "Assessing the impact of the social network on marking photos as favorites in Flickr" [Alves et al. 2012] publicado no 18th Brazilian Symposium on Multimedia and the Web (WebMedia'2012).

1 Introdução

Nos últimos anos tem-se observado grande expansão dos sistemas de conteúdo gerado pelos usuários (*user-generated content systems*, UGCs) como Twitter, Youtube, Flickr e Digg. Esses sistemas são utilizados diariamente por milhões de usuários com o objetivo de compartilhar e/ou consumir conteúdos dos mais diversos temas, como arte, política e economia [Chiarandini 2012, Chiluka et al. 2011, David and Jon 2010, Fortunato 2010].

Diversos estudos em UGCs têm investigado como um conteúdo compartilhado por um usuário é valorado ou disseminado pelos seus contatos na rede social [Bigonha et al. 2010, Doerr et al. 2012, Burke et al. 2011, van Zwol 2007]. Esses estudos têm focado, por exemplo, em caracterizar a forma como um conteúdo é propagado em uma rede de contatos e/ou identificar usuários que produzem conteúdos mais populares e usuários mais influentes [Bigonha et al. 2010]. Alguns estudos têm mostrado que alguns usuários constituem grandes redes de contato, mas essa rede tem pouco efeito sobre os conteúdos que eles consomem [Doerr et al. 2012]. Essa constatação coloca em perspectiva que a criação de um contato entre dois usuários pode não está essencialmente relacionada ao interesse no conteúdo compartilhado [Burke et al. 2011]. Em particular, a partir dos diversos níveis de distância entre os usuários na rede social, procuramos identificar o impacto da rede de contatos na marcação de conteúdos como favoritos.

Neste trabalho analisamos o impacto da rede de contatos na avaliação de fotos utilizando dados da rede social Flickr¹, um dos mais importantes sites de compartilhamento de fotos, que possui milhões de usuários, mais de 6 bilhões² de fotos postadas e milhares de novas fotos sendo postadas diariamente [Chiluka et al. 2011]. No Flickr usuários podem realizar diversas atividades, como criar contatos com outros usuários, postar fotos e marcar fotos como favoritas. Nossa análise considera a rede social formada pelas ligações de contato no Flickr. Consideramos diversos níveis de distância entre os usuários na rede social com o propósito de identificar relações entre a preferência por conteúdos postados e a proximidade entre os usuários na topologia da rede social.

Os resultados obtidos mostram que 70% das fotos marcadas como favoritas pelos usuários são postadas pelos seus contatos. Isso indica que o usuário consome conteúdo dos usuários com quais estabelece contato. Além disso, diferentemente do observado nas redes sociais Digg.com [Doerr et al. 2012] e Twitter [Bigonha et al. 2010], no Flickr a rede de contatos apresenta potencial como forma de disseminação de conteúdo. Nossos resultados mostram que aproximadamente 10% das fotos marcadas como favoritas pelos usuários não foram postadas pelos seus contatos, mas sim disseminadas por eles por meio de favoritos. De uma forma geral, os comportamentos que caracterizamos neste trabalho abrem diversas perspectivas no uso da rede social, como recomendação de conteúdo [Krohn-Grimberghe et al. 2012], descoberta de usuários formadores de opinião e formação de times de especialistas [Anagnostopoulos et al. 2012].

As demais seções deste artigo estão organizadas da seguinte forma. A Seção 2 apresenta os trabalhos relacionados, a Seção 3 apresenta a metodologia de coleta dos dados e de processamento do grafo de contatos. A Seção 5, por sua vez, apresenta os resultados. Finalmente, a Seção 6 apresenta as conclusões e trabalhos futuros.

2 Trabalhos relacionados

Diversos estudos têm se dedicado à análise das atividades de usuários em UGCs [Doerr et al. 2012, Cha et al. 2008, Burke et al. 2011, Bigonha et al. 2010]. De uma forma geral, esses estudos se concentram (i) na análise de diferentes perfis de usuários em termos de produção e consumo de conteúdo [Burke et al. 2011], (ii) na modelagem da disseminação do conteúdo ao longo de uma rede de contatos [Cha et al. 2008] e (iii) na análise da importância da rede de contatos na disseminação de conteúdo [Doerr et al. 2012].

Burke et al. [Burke et al. 2011] investigam a atividade de usuários no Facebook delineando diferentes perfis de usuários e como eles fazem uso da rede social. O estudo mostra que usuários atuam no sistema de diferentes formas em termos de atividade social e uso das mídias disponíveis. Alguns usuários apresentam um comportamento mais ativo, i.e., são mais propensos a produzir do que consumir conteúdos, enquanto outros apresentam um comportamento mais passivo. Neste trabalho analisamos a rede de contatos na perspectiva do usuário e focamos em analisar se os usuários demonstram alguma preferência por conteúdos postados por outros usuários que estão próximos ou distantes na rede de contatos.

Cha et al. [Cha et al. 2008] utilizam uma base de dados do Flickr para analisar como um conteúdo é disseminado ao longo da rede social. Esse estudo se concentra nos conteúdos postados pelos usuários e no conjunto de características que permite classificá-los. Neste trabalho não tratamos das características dos conteúdos postados

¹www.flickr.com

²<http://blog.flickr.net/en/2011/08/04/6000000000/>

pelos usuários, consideramos apenas a quantidade de vezes que o conteúdo postado foi marcado como favorito. Focamos na análise da preferência que os usuários manifestam pelos conteúdos postados em sua rede de contatos.

Doerr et al. [Doerr et al. 2012] analisam a importância de usuários no agregador de notícias Digg.com e mostram que usuários criam redes de contato que exercem pouca importância na disseminação de conteúdo. Comportamento semelhante tem sido observado na rede social Twitter [Bigonha et al. 2010], onde as interações de um usuário com outros é mais representativo de sua influência do que os contatos que ele possui na rede social [Bigonha et al. 2010].

Neste trabalho focamos em analisar o impacto da rede de contatos na avaliação de fotos na rede social Flickr. Essa rede social apresenta diversas diferenças em relação ao Digg e ao Twitter. No Digg os usuários apenas agregam notícias anteriormente produzidas e veiculadas por outros meios, por exemplo, sites de notícias. No Twitter os usuários geralmente postam diversas pequenas mensagens. De outro modo, no Flickr os usuários produzem as fotos postadas e, geralmente, há uma preocupação dos usuários com a qualidade do conteúdo postado.

3 Materiais e métodos

Esta seção descreve os materiais e métodos utilizados para medir o impacto da rede social na marcação de fotos como favoritas no Flickr. É importante notar que a interface do Flickr apresenta aos usuários informações sobre fotos marcadas como favoritas e novas postagens realizadas pelos seus contatos. Entretanto, do mesmo modo, o Flickr também provê aos seus usuários informações sobre fotos que foram postadas e/ou muito favoritas por outros usuários do sistema, que não são seus contatos diretos. Assim, os usuários têm acesso às fotos postadas por seus contatos e àquelas postadas por outros usuários que não estão em sua rede de contatos.

Coleta dos dados. Os dados utilizados neste trabalho foram coletados por meio da API pública do Flickr³. Como método de coleta utilizamos o *Snowball Sampling* [Goodman 1961, Chiluka et al. 2011]. Esse método conduz a uma coleta quase completa de dados para uma vizinhança particular do grafo, que é de interesse para nossa análise. Os dados coletados contêm links usuário-usuário e usuário-item. Os links usuário-usuário indicam a criação de contato entre um usuário e outro. Os links usuário-item indicam que um usuário postou ou marcou uma foto como favorita. A janela de tempo considerada para a coleta de dados é de quatro meses. Um resumo dos dados coletados é apresentado na Tabela 1.

Tabela 1: Visão geral dos dados coletados na rede social Flickr

Período de coleta	de 01/11/2011 a 29/02/2012
Total de usuários	2.241.714
Total de fotos	10.304.604
Marcas de favorito	29.249.205
Links de contato	20.684.373

Questões investigadas. Nossa análise visa delinear (i) como os usuários são afetados por sua rede de contatos e (ii) como eles afetam a rede em que participam. Quando investigamos como os usuários são afetados por sua rede de contatos, medimos em que proporção eles marcam como favoritas as fotos postadas por outros usuários na rede social. Quando investigamos como os usuários afetam sua rede de contatos, consideramos em que proporção os usuários dessa rede marcam como favoritas as fotos postadas pelo usuário. Visamos identificar como os favoritos feitos e os recebidos pelos usuários variam ao longo da rede social, considerando usuários em diferentes distâncias na topologia da rede social. Na Figura 1 mostramos a relação entre favoritos feitos e favoritos recebidos para cada usuário da base de dados utilizada.

Distância entre usuários. Um fator fundamental neste trabalho é medir a distância entre dois usuários na rede social. A rede social pode ser modelada como um grafo direcionado $G = (U, L)$, onde U é o conjunto de usuários e L o conjunto de arestas direcionadas que representam ligações de contato entre usuários. Um exemplo desse tipo de grafo é apresentado na Figura 2. Nesse grafo, uma aresta $(u_1, u_2) \in L$ indica que u_1 possui u_2 como

³<http://www.flickr.com/services/api/>

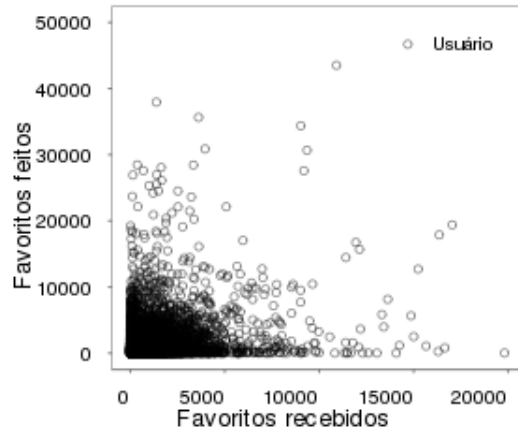


Figura 1: Gráfico de dispersão da quantidade de vezes que cada usuário marcou como favoritas fotos postadas por outros usuários (favoritos feitos) pela quantidade de vezes que esse usuário teve suas fotos marcadas como favoritas (favoritos recebidos).

um contato e que existe uma distância 0 de u_1 a u_2 . Assim, a distância entre dois usuários é o menor número de usuários que estão entre os dois usuários na rede de contatos.

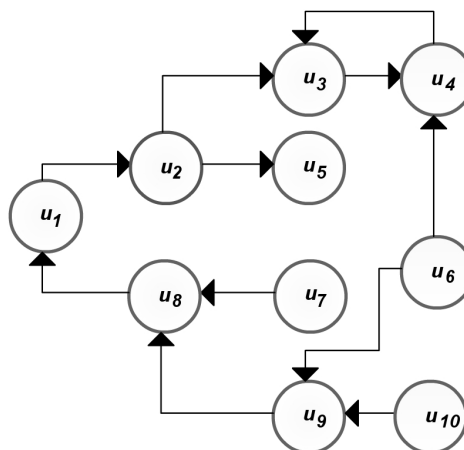


Figura 2: Exemplo de um subgrafo da rede social.

Considere o grafo apresentado na Figura 2. Nessa figura, o usuário u_2 é um contato do usuário u_1 . O usuário u_3 , por sua vez, é um contato do usuário u_2 . Pode-se dizer que u_3 está a uma distância 1 do usuário u_1 . De outro modo, o usuário u_7 não está na rede de contatos do usuário u_1 e não pode ser alcançado por ele por meio do grafo da rede social, mas u_1 está na rede de contatos do usuário u_7 . O cálculo da distância pode ser implementado em uma busca em largura partindo de um usuário de referência até um nível n . Quando uma foto postada pelo usuário u_1 é marcada como favorita pelo usuário u_8 , dizemos que u_1 influenciou um contato que se encontra a uma distância 0. De outro modo, se o usuário u_7 marcar como favorita uma foto postada pelo usuário u_1 , consideramos que o usuário u_1 influenciou um usuário que se encontra distante dele em 1 usuário.

Metodologia. Nosso método visa identificar do total de favoritos recebidos pelas fotos postadas por um usuário de referência, o percentual desses favoritos que foram feitos por usuários a uma distância 0 até uma distância n do usuário de referência. Do mesmo modo, do total de favoritos feitos por um usuário de referência, o percentual desses favoritos que foram feitos em fotos postadas por usuários a uma distância 0 até uma distância n do usuário de referência na topologia da rede social. Para definir os usuários de referência, selecionamos aleatoriamente

um conjunto de 50 usuários. Esse número de usuários mostrou-se suficiente para obter resultados com um erro estatístico satisfatório para um nível de confiança de 95%.

4 Processamento dos dados

É requerido alto poder computacional na construção e processamento do grafo da rede social, verificando as fotos postadas e marcadas como favoritas por cada usuário em diversas distâncias partindo de um usuário de referência. Isso envolve percorrer milhões de usuários e fotos (Tabela 1). Para agilizar o processamento, utilizou-se o modelo de programação paralela MapReduce para execução na plataforma Hadoop.

Os dados utilizados neste trabalho foram coletados por meio da API pública do Flickr⁴. Após a coleta, os dados foram estruturados em uma base de dados que contém o registro de um evento em cada linha. Os eventos podem ser do tipo *user-user*, que indicam a criação de contato entre usuários, ou do tipo *user-item*, que indicam que um usuário postou ou marcou uma foto como favorita. Todos os eventos têm 4 campos separados por “,”. O primeiro e segundo campo variam com o tipo do evento da seguinte forma:

- Evento 0: usuário (*userId*) cria contato com outro usuário (*otherUserId*), estrutura: *userId,otherUserId,0,0*. Todos os eventos deste tipo têm timestamp 0;
- Evento 1: usuário (*userId*) faz upload de uma foto (*photoId*). Esse tipo de evento tem a seguinte estrutura: *userId,photoId,1,photoDateOfUpload*;
- Evento 2: usuário (*userId*) favorita uma foto (*photoId*). Esse tipo de evento tem a seguinte estrutura: *userId,photoId,2,photoDateOfFaved*;

O terceiro campo de cada evento é o tipo do evento. O quarto campo de cada evento é o timestamp em que o evento ocorreu, podendo ser, conforme o evento, o instante em que a foto foi postada no sistema (*photoDateOfUpload*) ou o instante em que a foto foi marcada como favorita (*photoDateOfFaved*). Não é possível obter pela API do Flickr a informação sobre quando um evento *user-user* ocorreu. Em razão disso, em nossos dados todos os eventos deste tipo têm timestamp definidos como 0.

O processamento da base de dados consiste em um workflow de jobs MapReduce. Esse workflow é formado pelas seguintes etapas: (i) *Verificação de links user-user*, nesta etapa são identificadas as relações de contato entre os usuários; (ii) *Amostragem*, nesta etapa são selecionados aleatoriamente um conjunto de usuários para serem utilizados como usuário de referência, i.e., aqueles que serão objeto da análise proposta; (iii) *Transformação*, cria-se uma representação do grafo (todos os usuários de referência e os usuários que estão em suas rede de contatos) e verifica-se as fotos marcadas como favoritas e postadas por cada usuário que compõem esse grafo; (iv) *Verificação de eventos user-item*, nesta etapa, para cada um dos usuários, verifica-se as fotos que ele postou e as fotos que ele marcou como favoritas; (v) *Cálculo das distâncias e verificação da influência*, esta etapa consiste no cálculo distâncias entre os usuários de referência e demais usuários no grafo da rede social. Nesta etapa, calcula-se também a influência do usuário sobre sua rede e, de modo inverso, a influência da rede sobre ele.

No restante desta seção detalhamos o algoritmo utilizado para calcular a influência dos usuários sobre suas redes. Dado um usuário de referência, esse algoritmo consiste em verificar a sua distância para os demais usuários da rede e calcular, das fotos que ele postou, quantas foram marcadas como favoritas por cada um dos demais usuários em diversas distâncias na topologia da rede social – discutiremos no final desta seção que esse algoritmo se mantém para o cálculo da influência da rede sobre o usuário.

Esse algoritmo pode ser implementado como um programa MapReduce que realiza uma busca em largura no grafo da rede social calculando as distâncias de um usuário de referência para todos os outros usuários. Essa implementação consiste em um job iterativo que é executado para cada usuário de referência, que foram previamente selecionados. O usuário de referência é definido pela constante *user* e as fotos postadas por ele são definidas pela constante *fotosPostadas*. Essas constantes são definidas no código *main* do job iterativo. A função *main* consiste em um loop que faz a expansão no grafo até o nível *n*, ou distância máxima entre usuários que se deseja analisar. Neste trabalho a distância máxima (*n*) é 5. A cada execução do loop é realizada uma chamada MapReduce. A saída gerada pela função *reduce* em uma iteração é indicada como entrada para a função *Map* na iteração seguinte.

⁴<http://www.flickr.com/services/api/>

```

public void map(Text key, Text value, OutputCollector<Text, Text> output,
    Reporter reporter) throws IOException {
    String valor = value.toString();
    String[] array = valor.split(";");
    if (array.length == 3) {
        if (!array[0].startsWith("_") && !array[0].contains("s")) {
            int nivelAtual = Integer.valueOf(array[0]);
            String[] contatos = array[2].split("_");
            for (String contato : contatos) {
                output.collect(new Text(contato),
                    new Text(Integer.toString(nivelAtual + 1)));
            }
        }
        output.collect(key, value);
    }
}

```

Listing 1: Função Map

A função *Map* (Código 1)) recebe como chave o identificador de um usuário e, como valor, uma lista que é composta dos seguintes campos separados por “;”: o primeiro campo representa a distância do usuário recebido como chave e o usuário de referência; o segundo campo contém os identificadores das fotos marcadas como favoritas pelo usuário recebido como chave; o terceiro campo, por sua vez, é uma lista de usuários que adicionaram como um contato o usuário que a função recebe como chave. Usuários que não foram adicionados como contato por nenhum dos outros usuários, não têm 3 itens na lista anterior, nesse caso a função Map apenas emite os parâmetros recebidos. A existência de um “s” no primeiro campo indica que foi atingida a distância mínima, este é um artifício para evitar reprocessar buscas para esse usuário. O caractere “_” no campo de distância indica que o usuário em questão ainda não foi visitado, i.e., não se sabe a que distância ele está do usuário de referência. Para usuários não marcados com “s” e que já possuem uma distância calculada, a saída da função é uma atualização da distância entre o usuário de referência e o usuário recebido como chave.

```

public void reduce(Text key, Iterator<Text> values,
    OutputCollector<Text, Text> output, Reporter reporter)
    throws IOException {

    String padrao = "";
    String nivel = "";

    while (values.hasNext()) {
        //identifica lista de itens e novas distâncias
        String valor = values.next().toString();
        if (valor.contains(";")) {
            padrao = valor;
        } else {
            nivel = valor;
        }
    }

    String[] array = padrao.split(";");

    if (!nivel.equals("")) {
        //se alguma nova distância foi encontrada
        if (!padrao.equals("")) {
            array[0] = array[0].replaceAll("_", nivel);
            output.collect(key, new Text(arrayToString(array)));
        }
    } else {

        if (array[0].equals("_")) {
            // primeira chamada do workflow iterativo em que nenhum usuário foi visitado
            if (key.toString().equals(user)) {
                //se o usuário recebido é o usuário de referência
                array[0] = array[0].replaceAll("_", "0");
            }

            if (!array[1].contains("d")) {

```

```

//Verifica a interseção entre as fotos postadas pelo
//usuário de referência e favoritadas pelos demais usuários
String[] fotos = array[1].trim().split(" ");
String[] fotosPostadas = posts.trim().split(" ");
int count = 0;
if (!fotos[0].trim().isEmpty()) {
    for (String foto : fotos) {
        if (!foto.trim().isEmpty()) {
            for (String i : fotosPostadas) {
                if (i.equals(foto)) {
                    count++;
                }
            }
        }
    }
}
array[1] = count + "d";
}

output.collect(key, new Text(arrayToString(array)));
} else {
//quando a distância é encontrada
array[0] = array[0].replaceAll("s", "") + "s";
output.collect(key, new Text(arrayToString(array)));
}
}
}
}

```

Listing 2: Função Reduce

A função *Reduce* (Código 2) atualiza as distâncias entre os contatos. A entrada dela é a saída da função *Map* indicada no Código 1. A entrada consiste em uma chave com o identificador de um usuário corrente e um conjunto de valores que contêm uma lista com formato igual à recebida como um valor pela função *Map* e, em alguns casos, as distâncias encontradas para o usuário. Note que esse *Reduce* é chamado em um workflow iterativo, na primeira chamada ele identifica o usuário de referência e computa a quantidade de fotos que ele postou e que foram marcadas como favoritas pelos demais usuários, inserindo após esse cálculo a letra “d” no segundo campo da lista. A letra “d”, semelhante ao “s”, é usada como artifício para que esse cálculo não se repita. Referente ao “s”, teremos as distâncias entre os contatos, representado no primeiro campo da lista. Quando a primeira distância é encontrada, tem-se o caminho mais curto entre os contatos e, portanto, não haverá necessidade de processar contatos em que as distâncias foram definidas, diminuindo assim o processamento. As demais chamadas no processo iterativo realizam a expansão em largura no grafo da rede social, partindo do usuário de referência.

Esse algoritmo de cálculo da influência do usuário de referência sobre sua rede se mantém para o cálculo da influência da rede sobre os usuários de referência, mudam-se apenas os seguintes parâmetros: (i) a lista de fotos postadas (*fotosPostadas*) pelo usuário de referência passa a ser uma lista de fotos marcadas como favoritas; (ii) no valor recebido pela função *Map*, o segundo campo passa a ser os identificadores das fotos postadas pelo usuário indicado na chave recebida como parâmetro; o terceiro campo, por sua vez, passa a ser a lista de identificadores de usuário que foram adicionados como um contato pelo usuário indicado na chave recebida como parâmetro; (iii) a primeira chamada do workflow iterativo computa a quantidade de fotos marcadas como favoritas pelo usuário de referência e que foram postadas pelos demais usuários.

Encontram-se disponíveis para download⁵ todos os dados utilizados neste trabalho, os *scripts* implementados para coletar os dados por meio da API do Flickr e as implementações MapReduce utilizadas para identificar os favoritos realizados e recebidos nas mais diversas distâncias entre usuários na rede social.

5 Resultados

Os resultados obtidos são apresentados nas Figuras 3 e 4. A Figura 3 tem no eixo *y* o percentual de favoritos feitos pelo usuário em fotos postadas por outros usuários que se encontram até a distância indicada no eixo *x*. Os resultados apresentados nessa figura mostram que 70% das fotos marcadas como favoritas pelos usuários foram

⁵URL para download dos dados <http://redmine.lsd.ufcg.edu.br/projects/itemsrecommendation/wiki>

postadas por usuários que se encontram a uma distância 0 (Fig 3), i.e., aqueles que são seus contatos. Esse resultado é um indicativo de que no Flickr os usuários consomem conteúdos compartilhados pelos usuários com os quais estabelecem contato.

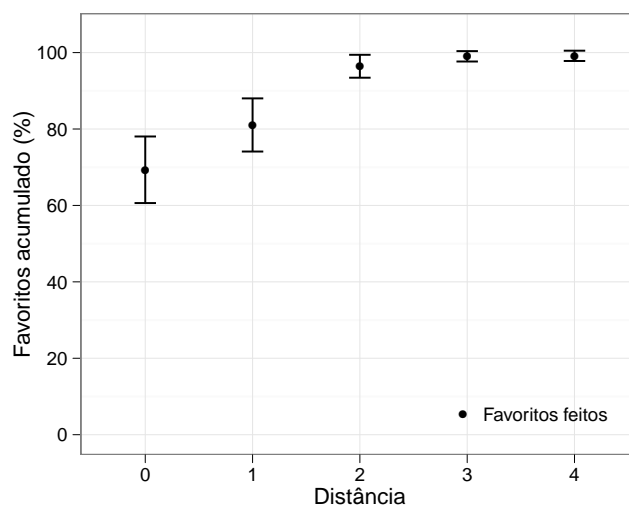


Figura 3: Distância na rede de contatos em que foram postadas as fotos marcadas como favoritas pelos usuários.

Os resultados apresentados na Figura 3 mostram também que aproximadamente 20% das fotos marcadas como favoritas pelos usuários foram postadas por usuários a uma distância 1 e 2. De um modo geral, esses resultados mostram que no Flickr os usuários têm importância na produção de conteúdo relevante para seus contatos e também na disseminação desse tipo de conteúdo.

A Figura 4 tem no eixo *y* o percentual de favoritos recebidos por fotos postadas pelos usuários. Calcula-se de todos os favoritos recebidos que percentual desses favoritos foram feitos por outros usuários que se encontram até o nível de distância indicado no eixo *x*. Os resultados apresentados nessa figura mostram que grande parte (aproximadamente 75% em média) dos favoritos recebidos pelos conteúdos postados pelos usuários advém de seus contatos (distância 0).

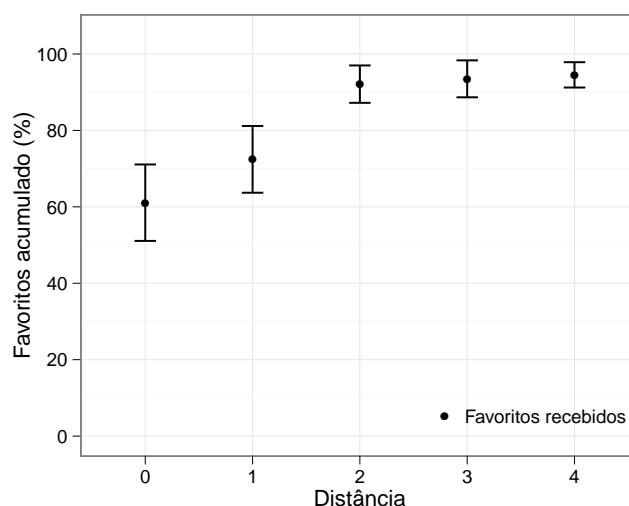


Figura 4: Distância na rede de contatos em que as fotos postadas pelos usuários foram marcadas como favoritas.

Implicações e limitações. Os resultados obtidos neste trabalho indicam que em UGCs, como o Flickr, a rede de contatos apresenta considerável importância para os usuários em termos de conteúdos relevantes, uma vez

que maior parte dos conteúdos que eles marcam como favoritos são postados por usuários com os quais mantêm contato. Esse resultado possui diversas implicações em, por exemplo: (i) sistemas de recomendação de conteúdo considerando fotos marcadas como favoritas mais próximas na rede de contatos; (ii) descoberta de usuários influentes em sua rede de contatos; e (iii) formação de times de usuários especialistas levando em consideração aqueles que mais influenciam a sua rede de contatos.

Uma limitação deste trabalho é não delimitar diferentes perfis de usuários considerando, por exemplo, o número de contatos e a relação entre a quantidade de fotos postadas e a quantidade de favoritos recebidos. Um estudo mais aprofundado pode revelar se os resultados observados podem ter maior ou menor intensidade dependendo dos perfis dos usuários analisados.

6 Conclusão

Neste trabalho analisamos o impacto da rede de contatos na avaliação de fotos na rede social Flickr. Avaliamos as preferências dos usuários por meio de fotos postadas em diferentes distâncias na topologia da rede social. Os resultados obtidos indicam que os usuários demonstram preferência pelos conteúdos postados pelos seus contatos, ou outros usuários que tenham alguma ligação com seus contatos. Juntos, esses representam em média 80% dos favoritos feitos pelos usuários. Além disso, os contatos dos usuários são responsáveis por mais de 60% dos favoritos recebidos nos conteúdos postados.

Como trabalho futuro pretendemos agrupar os usuários da rede social em diferentes perfis, considerando, por exemplo, o número de contatos que eles possuem e a razão entre a quantidade de fotos postadas e a quantidade de favoritos recebidos. Com essa análise, pretendemos identificar se os resultados observados possuem maior ou menor significância dependendo do perfil do usuário. Além disso, dado que nosso resultado mostra que os usuários demonstram preferência pelos conteúdos postados em sua rede de contatos, pretendemos usar essas conexões em um algoritmo de recomendação de fotos.

Agradecimentos Os autores agradecem os valiosos comentários de Nazareno Andrade, Leandro Marinho e dos avaliadores anônimos. Este artigo foi desenvolvido em cooperação com a Hewlett-Packard Brasil Ltda., no contexto do programa de Formação de Recursos Humanos em Analytics HP/UFCG 2012, e utilizou incentivos da Lei de Informática brasileira (lei n. 8.248 de 1991).

Referências

- [Alves et al. 2012] Alves, L., Araújo, M., Ponciano, L., and Brito, A. (2012). Assessing the impact of the social network on marking photos as favorites in flickr. In *18th Brazilian Symposium on Multimedia and the Web*, São Paulo, Brazil. ACM.
- [Anagnostopoulos et al. 2012] Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., and Leonardi, S. (2012). Online team formation in social networks. In *Proceedings of the 21st international conference on World Wide Web*, pages 839–848. ACM.
- [Bigonha et al. 2010] Bigonha, C., Cardoso, T. N. C., Moro, M. M., Almeida, V. A. F., and Gonçalves, M. A. (2010). Detecting Evangelists and Detractors on Twitter. In *18th Brazilian Symposium on Multimedia and the Web*, pages 107–114.
- [Burke et al. 2011] Burke, M., Kraut, R., and Marlow, C. (2011). Social capital on facebook: differentiating uses and users. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 571–580. ACM.
- [Cha et al. 2008] Cha, M., Mislove, A., Adams, B., and Gummadi, K. P. (2008). Characterizing social cascades in flickr. In *Proceedings of the first workshop on Online social networks*, pages 13–18. ACM.
- [Chiarandini 2012] Chiarandini, L. (2012). Exploration and discovery of user-generated content in large information spaces. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 755–756. ACM.

- [Chiluka et al. 2011] Chiluka, N., Andrade, N., and Pouwelse, J. (2011). A link prediction approach to recommendations in large-scale user-generated content systems. In *Proceedings of the 33rd European conference on Advances in information retrieval*, pages 189–200. Springer-Verlag.
- [David and Jon 2010] David, E. and Jon, K. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA.
- [Doerr et al. 2012] Doerr, C., Blenn, N., Tang, S., and Van Mieghem, P. (2012). Are friends overrated? a study for the social news aggregator digg.com. *Comput. Commun.*, 35(7):796–809.
- [Fortunato 2010] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(35):75 – 174.
- [Goodman 1961] Goodman, L. A. (1961). Snowball sampling. *The Annals of Mathematical Statistics*, 32(1):148–170.
- [Krohn-Grimberghe et al. 2012] Krohn-Grimberghe, A., Drumond, L., Freudenthaler, C., and Schmidt-Thieme, L. (2012). Multi-relational matrix factorization using bayesian personalized ranking for social network data. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 173–182. ACM.
- [van Zwol 2007] van Zwol, R. (2007). Flickr: Who is looking? In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 184–190. IEEE Computer Society.