

Caracterização de Padrões de Acesso de Usuários do Orkut*

Kênia Carolina
Gonçalves
Universidade Federal de
Minas Gerais
Belo Horizonte, Minas Gerais
keniacarolina@dcc.ufmg.br

Humberto Torres
Marques-Neto
Pontifícia Universidade
Católica de Minas Gerais
Belo Horizonte, Minas Gerais
humberto@pucminas.br

Jussara Marques de
Almeida
Universidade Federal de
Minas Gerais
Belo Horizonte, Minas Gerais
jussara@dcc.ufmg.br

ABSTRACT

Recently, there has been a huge growth in popularity of online social networking, which has facilitated the interaction between Internet users. The characterization of the users' access patterns of these applications is an important task to support the optimization project and new systems that best fit the user needs. This paper presents a characterization of users' access patterns of Orkut, the more popular social networking in Brazil. The characterization was done with real data traffic from a residential broadband provider, was presented from two perspectives: the intra-Orkut analysis, which emphasizes the pattern of users' navigation through various features of Orkut, and the inter-social network analysis, which covers the access from Orkut to other social networks such as YouTube, Fotolog, MySpace, Last.fm and Facebook. The results show the presence of three distinct profiles of navigation through the features of Orkut, and a predominance of sessions using the Orkut and Fotolog together with YouTube to share videos and photos.

RESUMO

Recentemente, nota-se um grande crescimento da popularidade das redes sociais online, o que tem facilitado a interação entre os usuários de Internet. A caracterização dos padrões de acesso dos usuários a estas aplicações é tarefa importante para subsidiar o projeto de otimizações e novos sistemas que melhor se adequam às necessidades dos usuários. Este trabalho apresenta uma caracterização dos padrões de acesso dos usuários do Orkut, uma das redes sociais de maior popularidade no Brasil. A caracterização, realizada com dados reais de tráfego de um provedor de banda larga residencial, é apresentada sob duas perspectivas: a análise intra-Orkut, que enfatiza o padrão de navegação dos usuários pelas diver-

*Characterization of User Access Patterns in Orkut.

Esta pesquisa é parcialmente financiada pelo Instituto Nacional de Ciência e Tecnologia para a Web - INCTWeb (MCT/CNPq 573871/2008-6), CNPq, FAPEMIG e CAPES.

sas funcionalidades do Orkut, e a análise inter-rede social, que aborda os acessos a partir do Orkut para outras redes sociais tais como YouTube, Fotolog, MySpace, Last.fm e Facebook. Os resultados mostram a presença de três perfis distintos de navegação pelas funcionalidades do Orkut, bem como uma predominância de sessões que utilizam o Orkut juntamente com o Fotolog e o YouTube para compartilhamento de fotos e vídeos.

Categories and Subject Descriptors

C.4 [Computer Systems Organization]: Performance of Systems - Measurement techniques; H.3.5 [Information Storage and Retrieval]: Online Information Services - Web-based services

General Terms

Human Factors, Measurement

Keywords

Online social networks, characterization of user behavior

1. INTRODUÇÃO

A importância da Web como um ambiente para comunicação e interação entre as pessoas cresce continuamente de forma acelerada. Atualmente, existem vários sistemas da chamada Web 2.0 que promovem a interação entre as pessoas, permitindo que cada um divulgue suas próprias informações e interesses para serem compartilhados. Esses sistemas permitem a criação de redes sociais online, ambiente virtual onde pessoas se conectam para comunicar e interagir com outros participantes da rede [12]. Exemplos de sistemas de redes sociais online com grande popularidade no mundo inteiro são o YouTube, um sistema de compartilhamento de vídeos que também possibilita a criação de várias redes de relacionamento (e.g., rede de amigos, rede de assinaturas, etc), e o Facebook, uma aplicação de rede social de caráter geral. No Brasil, o Orkut é uma das aplicações de redes sociais mais populares, com mais de 24 milhões de participantes, o que representa mais de 70% da população brasileira que acessa a Internet¹. Além disto, o Orkut também figura entre os 3 sites que mais geram tráfego na

¹<http://www.orkut.com.br/html/advertise/BR/overview.html>.

Internet brasileira². Um usuário do Orkut pode, principalmente, compartilhar informações com outros usuários desta rede e participar de comunidades com interesses comuns [4].

Dado o crescimento do uso de aplicações de redes sociais on-line (ou simplesmente redes sociais, no restante do artigo), entender o comportamento dos usuários desses sistemas a partir da caracterização de padrões de acesso é tarefa importante para prover subsídios para o projeto e construção de novas aplicações que melhor se adequam aos padrões típicos e às necessidades dos seus usuários. A caracterização dos acessos a um sistema particular pode ser feita sob duas perspectivas. Uma *visão intra-rede social* contempla os padrões de acesso dos usuários a diferentes funcionalidades do sistema alvo. Já uma *visão inter-rede social* aborda os padrões de navegação de um usuário a partir do sistema alvo para diferentes aplicações de redes sociais.

A caracterização a partir da visão intra-rede social gera conhecimento sobre a navegação e a utilização do sistema pelos seus usuários. Como exemplo, ela pode ajudar a identificar, dentre as funcionalidades atualmente disponíveis, quais são mais populares, quais são raramente utilizadas e quais são frequentemente utilizadas conjuntamente (e.g., uma após a outra). Este conhecimento pode subsidiar a otimização do sistema analisado e/ou o projeto de novos sistemas que melhor atendam às necessidades dos usuários. Já a análise da navegação dos usuários de um sistema específico para outras redes sociais pode fornecer informação sobre como as funcionalidades do sistema em foco estão sendo utilizadas pelos usuários em conjunto com outras redes sociais. Por exemplo, suponha um sistema, como o Orkut, que possui uma funcionalidade de envio de recados, mas, não possui o recurso de disponibilização ou envio de vídeos. Se a análise dos padrões de navegação de seus usuários mostra que constantemente eles saem da página de envio de recados para uma outra rede social de compartilhamento de vídeos (e.g.: YouTube), isto pode ser um indício de que talvez a integração dos dois sistemas, ou então a disponibilização do compartilhamento de vídeos no Orkut, podem aumentar a utilização do sistema pelos seus associados.

Neste contexto, este trabalho caracteriza os padrões de acesso de usuários ao Orkut, considerando ambas visões intra e inter-rede social. Para tanto, foi utilizado um log de tráfego coletado de um provedor regional de acesso de Internet de banda larga residencial de uma companhia de TV a cabo brasileira. Desse log foram extraídas as transações para aplicações de redes sociais, particularmente para o Orkut, e estas foram agrupadas em sessões. A duração das sessões e o padrão de navegação pelas diversas funcionalidades do Orkut, bem como destas para outras aplicações de redes sociais, foram caracterizadas em cada sessão identificada. Em particular, o padrão de navegação dentro de cada sessão foi representado por um *Customer Behavior Model Graph* (CBMG) [7], um modelo de grafo de transições entre estados. Os CBMGs, gerados para cada sessão, foram então agrupados utilizando o algoritmo *k-means* para identificação de grupos/perfis de navegação de usuários dentro do Orkut, bem como deste para os sistemas Facebook, YouTube, MySpace, Last.fm e Fotolog.

Como resultado do agrupamento dos CBMGs, foi possível identificar três perfis de acessos intra-Orkut diferentes. O perfil *Objetivos*, que corresponde a 6,59% das sessões caracte-

terizadas, consiste de sessões em que o usuário navega pouco pelo sistema, concentrando-se na página principal do Orkut. Já o perfil *Moderados*, presente em 19,93% das sessões, corresponde a sessões em que os usuários navegam um pouco mais pelo sistema, fazendo uso de outras funcionalidades como Fotos. Por fim, o perfil *Navegadores* corresponde a sessões em que os usuários exploram amplamente as várias funcionalidades do sistema. Além disto, a grande maioria das sessões, particularmente as com perfis *Moderados* e *Navegadores*, utilizam o Fotolog e o YouTube em conjunto com o Orkut, indicando uma possível necessidade dos usuários de compartilhamento de fotos e vídeos.

O entendimento de como o usuário utiliza o sistema garante que novas ferramentas sejam criadas ou as existentes sejam atualizadas de tal forma que a utilização da rede social se potencialize. Como exemplo e de acordo com os resultados da análise inter-rede social desse trabalho, seria interessante o Orkut agregar ao sistema a possibilidade de que as fotos dos perfis sejam publicadas como álbums para os associados e não associados ao Orkut, assim como pode ser feito no Fotolog. Já na análise das funcionalidades do Orkut, outra possibilidade é que o usuário possa editar sua página principal com aquilo que ele mais utiliza, como se fossem *gadgets*.

O restante deste artigo está organizado da seguinte maneira: a Seção 2 apresenta os principais trabalhos relacionados, enquanto uma visão geral dos dados caracterizados é apresentada na Seção 3. Os padrões de navegação intra-Orkut e deste para outras aplicações são descritos nas Seções 4 e 5, respectivamente. Conclusões e trabalhos futuros são apresentados na Seção 6.

2. TRABALHOS RELACIONADOS

Vários trabalhos de caracterização de carga de trabalho, padrões de acesso e de interação entre usuários estão disponíveis na literatura. Em Tavares et. al. [10] foi feita uma caracterização dos dados trocados entre serviços Web e provedores de serviços. Especificamente sobre redes sociais, em [8] os autores analisaram as características estruturais das redes de relacionamentos do Orkut, Flickr, Live Journal e Youtube, confirmando a presença das propriedades de lei de Potência, livre de escala e mundo-pequeno destas redes. Ahn et. al. [1] estudaram as propriedades de duas redes sociais do Cyworld, sistema bastante popular na Coreia: a rede estabelecida explicitamente pela relação de amizade e a rede criada implicitamente a partir das mensagens trocadas pelos usuários dentro do sistema. Os autores mostraram que as duas redes compartilham semelhanças estruturais tais como as distribuições de grau de entrada e de saída dos usuários. Outros estudos enfatizaram a caracterização de outros sistemas como o Twitter [6] e o Facebook [5].

Em [2] é apresentada uma caracterização da carga de trabalho de serviços de vídeos do maior provedor de conteúdo da América Latina, o UOL. Benevenuto et. al [2] exploraram as requisições realizadas dentro de sessões usuários e apresentaram os CBMGs (*Customer Behavior Model Graph*) de utilização dos serviços. Especificamente focados no Orkut, Chen et. al. [4] caracterizaram as comunidades do Orkut, e exploraram o uso de técnicas de *Association Rule Mining* e *Latent Dirichlet Allocation* para recomendação personalizada de comunidades. Spertus et. al. [9] analisaram o conteúdo das informações trocadas entre os usuários para gerar recomendações e sugestões para os próprios usuários.

²<http://www.alexa.com>.

O Orkut também é foco de trabalhos de otimização de busca em redes sociais explorando os relacionamentos de amizades entre os usuários [11].

O trabalho anterior que mais se aproxima da proposta desse trabalho é aquele apresentado em [3] analisam os padrões de acesso dos usuários do Orkut. Entretanto, Benevenuto et. al. [3] caracterizaram dados coletados a partir de um agregador de redes sociais, enquanto nesse trabalho foram utilizados dados do tráfego total de um provedor de banda larga. Os autores de [3] fizeram uma análise das funcionalidades *Principal e Amigos* como uma só, enquanto nesse trabalho elas foram separadas e outras foram apresentadas separadamente, como: Feeds, Fãs-Favoritos, Configurações e Aplicações. Como resultado, a caracterização desse trabalho provê uma visão muito mais representativa dos padrões de acesso ao Orkut. Além disto, Benevenuto et. al. [3] não trataram de diferentes perfis de padrões de acesso dentro do Orkut e nem da análise inter-rede social, como feito por esse trabalho. Além disso, a base de dados desse trabalho é muito maior: ela contém quase 4 vezes mais sessões do Orkut, geradas por quase 2 vezes mais usuários, além de incluir um número muito mais expressivo de transações para outras redes sociais

3. VISÃO GERAL DA BASE DE DADOS

A caracterização de padrões de acesso dos usuários do Orkut foi feita sobre uma base de dados coletada na infraestrutura de um provedor regional de acesso de Internet (ISP - *Internet Service Provider*) de banda larga residencial de uma companhia de TV a cabo brasileira. Esta base inclui os logs de tráfego do provedor, que registram as transações de serviços Par-a-Par, correio eletrônico, VoIP, HTTP (Hypertext Transfer Protocol), entre outros. Cada linha do log, correspondente a uma transação, contém 46 campos, dos quais se destacam a URL da requisição, o protocolo, duração da transação, quantidade de bytes transferidos, etc.

De acordo com o escopo e objetivos propostos neste trabalho, apenas as transações que utilizam o protocolo HTTP para a comunicação e transferências de dados Web foram processadas. Dentre os 46 campos de cada transação, existe um campo que registra o domínio da transação e outro que registra a URL acessada pelo usuário do provedor. Por exem plo: Domínio: *www.orkut.com.br* e URL acessada: */Communities.aspx*. O domínio da transação foi utilizado para filtrar apenas transações para aplicações de redes sociais, identificadas com base em uma lista de 142 aplicações de redes sociais apresentada no site *Alexa.com*.

Os logs utilizados na caracterização correspondem ao período de 01 a 31 de Junho de 2008, e incluem 8.126.592 transações para aplicações de redes sociais. A Tabela 1 mostra a quantidade de transações para as 15 redes sociais com maior número de transações. É possível observar que quase 8 milhões de transações, o que corresponde a 97% das transações de redes sociais na base de dados utilizada, são para o Orkut, confirmando a grande popularidade desta aplicação no Brasil. Além do Orkut, nota-se também um volume significativo de transações para as aplicações YouTube, MySpace e Fotolog. A quantidade de transações inclui o acesso direto à página Web, imagens e/ou scripts transferidos a partir do domínio da rede social para o usuário.

Após processamento dos logs, foram identificados 55.416 usuários distintos do provedor de acesso a Internet banda larga, sendo que cerca de 90% desses usuários realizaram

Tabela 1: Número de transações por rede social

Rede Social	#Transações
Orkut	7.928.634
Youtube	82.578
MySpace	18.373
Fotolog	13.532
Ning	9.976
Flickr	9.613
Facebook	8.189
Multiply	7.864
DeviantART	7.463
Habbo	7.240
Vox	4.719
Sonico.com	3.565
Netlog	3.477
Last.fm	2.510
hi5	2.503

algum acesso ao Orkut no período analisado. Cada usuário é identificado unicamente pelo *MAC Address* da interface de rede do computador que utiliza para acessar a Internet através da infra-estrutura do ISP. Por questões de privacidade, os dados dos logs não permitem identificar os usuários: eles são tratados na base de dados por identificadores únicos disponibilizados para a pesquisa através do *MAC Address* do usuário-cliente do ISP.

Para caracterização dos padrões de acesso dos usuários do Orkut, utilizou-se o conceito de sessão. Uma sessão começa no momento em que é identificado o acesso ao Orkut por um usuário, independente da URL de início, e termina quando ele realiza logout ou se seu próximo acesso ao Orkut for após 24 horas do último acesso realizado. Dessa foram, foram identificadas 109.141 sessões de Orkut.

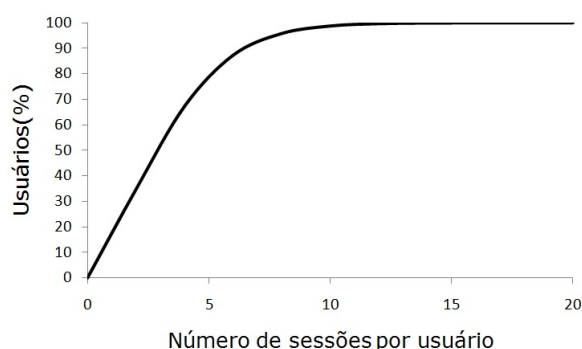


Figura 1: Distribuição Acumulada do Número de Sessões do Orkut por Usuário

A Figura 1 apresenta a distribuição acumulada do número de sessões do Orkut por usuário único, identificados nos logs. A grande maioria dos usuários realizou poucas sessões no período analisado. Em particular, 80% dos usuários realizaram até 6 sessões apenas. A base de dados analisada é resumida na Tabela 2 com os números de usuários, transações e sessões caracterizados.

4. ANÁLISE DE PADRÕES DE ACESSO NO ORKUT

Nesta seção são apresentadas quais as funcionalidades do Orkut que serão analisadas, como os dados foram agrupados

Tabela 2: *Sumário da base de dados*

Período de coleta	01/06/2008 a 31/06/2008
# Usuários totais	55.416
# Usuários que acessaram redes sociais	50.085
# Transações relacionadas a páginas do Orkut	1.111.150
# Sessões do Orkut	190.141

e as características dos grupos de usuários encontrados. Foi feita uma análise de padrões de acesso dos usuários dentro do próprio Orkut, chamada neste trabalho por intra-Orkut, e foram agrupadas as sessões de usuários em grupos/perfis de usuários.

4.1 Funcionalidades do Orkut

A arquitetura do Orkut permite a identificação da funcionalidade ou ferramenta utilizada pelos usuários através da análise da URL. Para a concepção de uma análise intra-rede social é necessário que funcionalidades ou páginas do sistema sejam definidas e em seguida mapeadas em relação as transações realizadas pelos usuários. A definição do mapa de funcionalidades do Orkut foi realizada com base nas URLs presentes nos logs.

Foram encontradas 78 tipos de transações de páginas diferentes do Orkut, as quais foram agrupadas e nomeadas de acordo com suas respectivas finalidades. São elas: (1) Principal - página inicial do Orkut e as páginas que exibem perfil do próprio usuário; (2) Amigos - páginas dos contatos amigos; (3) Comunidades - páginas das comunidades; (4) Depoimentos - criação depoimentos para amigos e/ou acessos aos depoimentos; (5) Fãs-Favoritos - funcionalidades que indicam fãs e/ou favoritos entre os contatos; (6) Vídeos - páginas de vídeos; (7) Fotos - imagens e albúms dos usuários; (8) Configurações - configurações da conta do usuário, segurança e privacidade; (9) Busca - páginas relacionadas as buscas feitas no sistema; (10) Login/Logout - páginas de logon/logout; (11) Recados - páginas de envio de mensagens públicas e rápidas; (12) Mensagens - mensagens enviadas para grupos ou um usuário; (13) Feeds - "News Feed" da página principal; (14) Aplicações - aplicações personalizadas dentro do sistema.

A quantidade de transações registradas para cada funcionalidade do Orkut está representada pela Figura 2. Percebe-se que a grande concentração de transações está na funcionalidade Principal, seguida de Amigos, Comunidades e Fotos. Com as sessões definidas e as funcionalidades do Orkut nomeadas e mapeadas, os conjuntos de transações por sessão de usuário foram agrupados para a identificação de padrões de acesso dos usuários.

4.2 Agrupamento das sessões de usuários

A representação dos padrões de acesso dos usuários às funcionalidades do Orkut foi realizada com um grafo de transição de estados denominado CBMG (*Customer Behavior Model Graph*) [7]. No contexto deste trabalho, os estados de um CBMG representam as funcionalidades do Orkut, as arestas representam as transições entre esses estados, ou seja, a navegação dentro do Orkut, e os pesos dessas arestas são as probabilidades de transição entre os estados. A maioria das arestas dos CBMGs apresentados neste trabalho com peso inferior à precisão de duas casas decimais foi retirada para a simplificação da apresentação visual destes grafos.

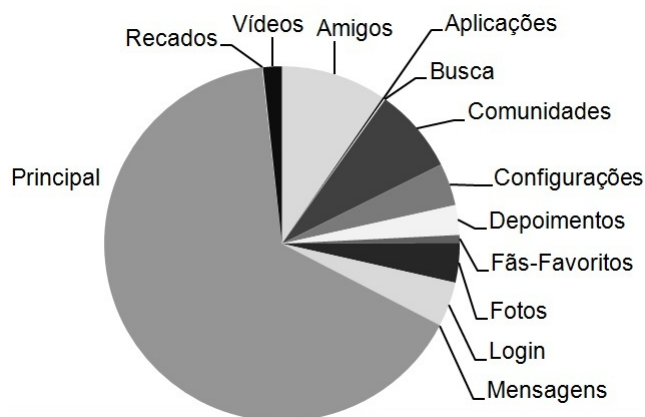


Figura 2: Número de transações por funcionalidades do Orkut

Os estados que possuem baixa conectividade para a precisão adotada são apresentados sem nenhuma conexão nos CBMGs.

As transações das sessões foram contabilizadas em relação a funcionalidade do Orkut (estado) de origem com a funcionalidade do Orkut de destino. Essa quantidade de transações entre os estados foi submetida a um algoritmo de agrupamento que permitiu identificar três grupos distintos de usuários e suas frequências de transições. A caracterização dos dados tornou-se necessária para identificar padrões de acesso e grupos de usuários com perfis semelhantes. Dentro das técnicas de agrupamento existentes, optou-se pelo uso do algoritmo *k-means* que atua nos dados de modo não supervisionado e com critérios de separação por distância dos dados - *Partitional Clustering* [7].

A quantidade de clusters para a execução do *k-means* foi definida examinando as seguintes métricas: distância média entre pontos dos clusters e seus centróides - CV_{Intra} ; e a distância média entre os centróides - CV_{Inter} . Foi selecionada uma relação que permite a menor variação entre os intra-clusters e a maior entre os inter-clusters. A razão - βCV - entre intra-clusters e inter-clusters determina a qualidade do processo de agrupamento [7]. Foi feita uma análise dos coeficientes de variação onde ao passar de 3 clusters, o βCV decresce significativamente indicando que o melhor número de clusters é três.

4.3 Grupos de Usuários

Os três grupos de usuários identificados pelo agrupamento de CBMGs mostram os padrões de acesso dos usuários intra-Orkut. O Cluster 1, denominado neste trabalho por *Objetivos*, contém 9.645 usuários, o Cluster 2, os *Moderados*, e o Cluster 3, os *Navegadores* possuem um número maior de usuários, com 14.020 e 26.420, respectivamente.

Os usuários *Objetivos* são responsáveis por 19,26% das transações dos logs. Já os *Moderados* realizaram 27,99% das transações, onde sua principal característica é ter um comportamento intermediário entre os usuários *Objetivos* e os *Navegadores*. O perfil *Navegadores* caracteriza os usuários que apresentam um grande volume de transações nos logs, com 52,75% do total.

A Figura 3 mostra a distribuição acumulada das durações das sessões, onde foi observado dois padrões de comporta-

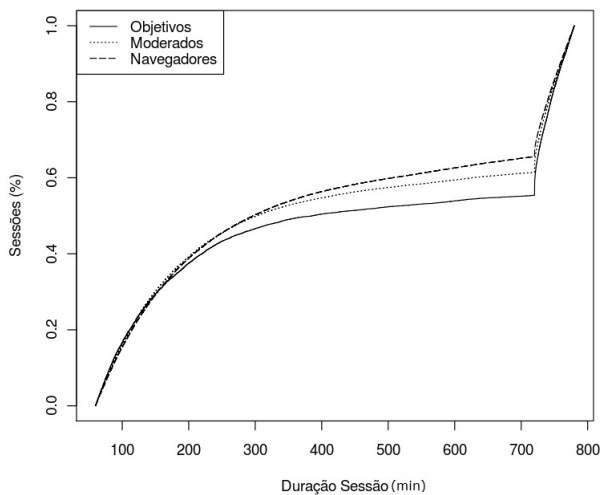


Figura 3: Duração das sessões por grupo de usuário

mento diferentes. Enquanto uma parcela significativa das sessões (36%) dos três grupos de usuários duram até 200 minutos, aproximadamente 40% das sessões duram mais de 12 horas (720 minutos). O mesmo padrão foi observado nos três perfis - *Objetivos*, *Moderados* e *Navegadores*. Entretanto, as sessões do perfil *Objetivos* tendem a ter uma duração ligeiramente maior, porque podem realizar mais acessos dentro do limite máximo adotado de 24 horas.

A Figura 4 é o CBMG do padrão de acesso dos usuários *Objetivos*, os quais possuem a característica de concentrarem o uso do Orkut em determinadas funcionalidades. Eles navegam pouco na rede social e logo realizam *logout*. Esse perfil de usuário pode ser caracterizado como usuários que visualizam a página principal do Orkut para utilizar funcionalidades como um controle de navegação. Pode-se dizer que os *Objetivos* buscam uma visão geral e rápida da rede social, uma vez que na funcionalidade Principal do Orkut é possível visualizar as atualizações dos amigos, novas solicitações dos amigos, visitas recentes ao perfil, próximos aniversários e quantidade de mensagens, recados, fotos e amigos on-line. Pelo CBMG apresentado na Figura 4 é possível observar que em 70% das transições que tem como origem a funcionalidade Principal (a mais utilizada) são para o estado final, ou seja, para o fim da sessão.

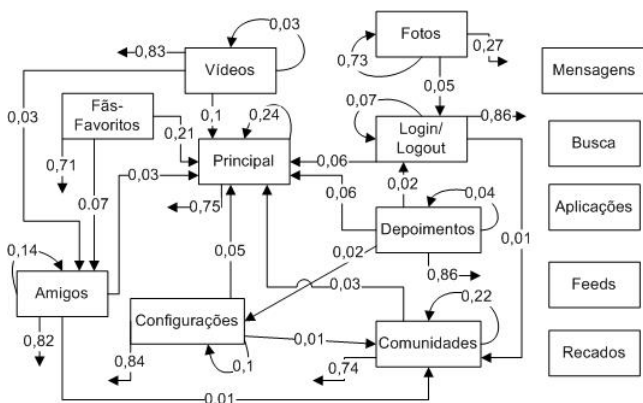


Figura 4: CBMG dos *Objetivos* intra-Orkut

A Figura 5 mostra o grafo do comportamento dos usuários classificados como *Moderados*. Eles navegam por um maior número de páginas da rede social do que os usuários *Objetivos*. Os usuários *Moderados* utilizam seguidamente a funcionalidade Principal do Orkut, pois 40% das transições com origem nesta funcionalidade é para ela mesma. Outra funcionalidade muito utilizada em sequência por estes usuários é a funcionalidade de Fotos, caracterizada pela auto-transição de 0,78 do CBMG.

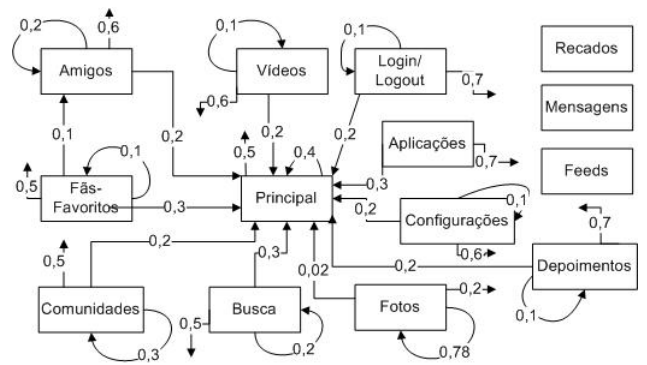


Figura 5: CBMG dos *Moderados* intra-Orkut

O perfil *Navegadores*, Figura 6, apresenta uma dispersão maior no uso das funcionalidades do Orkut. Ou seja, estes usuários exploram e utilizam várias funcionalidades da rede social em suas respectivas sessões. Considerando a quantidade de usuários deste grupo, quase 50% do total, e analisando o CBMG da Figura 6, percebe-se que muitos usuários do Orkut navegam bastante dentro da rede social utilizando com frequência as funcionalidades de Fotos, Comunidades, Amigos e Fãs-Favoritos, além da funcionalidade Principal.

A predominância dos usuários, independente do grupo, está nas funcionalidades Principal, Fotos, Amigos e Comunidades, conforme Figura 2. O mesmo fato pode ser observado no trabalho de Benevenuto et. al. [3], onde as funcionalidades Principal e Amigos são denominadas por "Profile & Friends". Em [3] o comportamento dos usuários apresentado mostra que a probabilidade de transição de "Profiles & Friends" para a própria "Profiles & Friends" é de 69%, porém neste trabalho essa probabilidade é baixa, enquanto a probabilidade de encerramento da sessão é mais alta que a apresentada por Benevenuto et. al. [3]. Observa-se que nos CBMGs encontrados para os três perfis de usuários que, em média, os *Objetivos* finalizam 63% de suas sessões quando estão usando essas funcionalidades. Isto também ocorre para 45% das sessões dos *Moderados* e em 27% das sessões dos *Navegadores*. Na Seção 5 será apresentado os padrões de navegação desses usuários em relação a outras redes sociais a partir das funcionalidades do Orkut.

5. ANÁLISE DOS PADRÕES DE ACESSO DO ORKUT PARA OUTRAS REDES SOCIAIS

A partir da caracterização dos grupos de usuários dentro da rede social Orkut, caracterizou-se também o relacionamento desses usuários com outras redes sociais, bem como de que maneira eles as utilizam a partir de uma sessão do Orkut. Portanto, foi realizado um estudo das transições de uma sessão de usuário do Orkut para outras redes sociais.

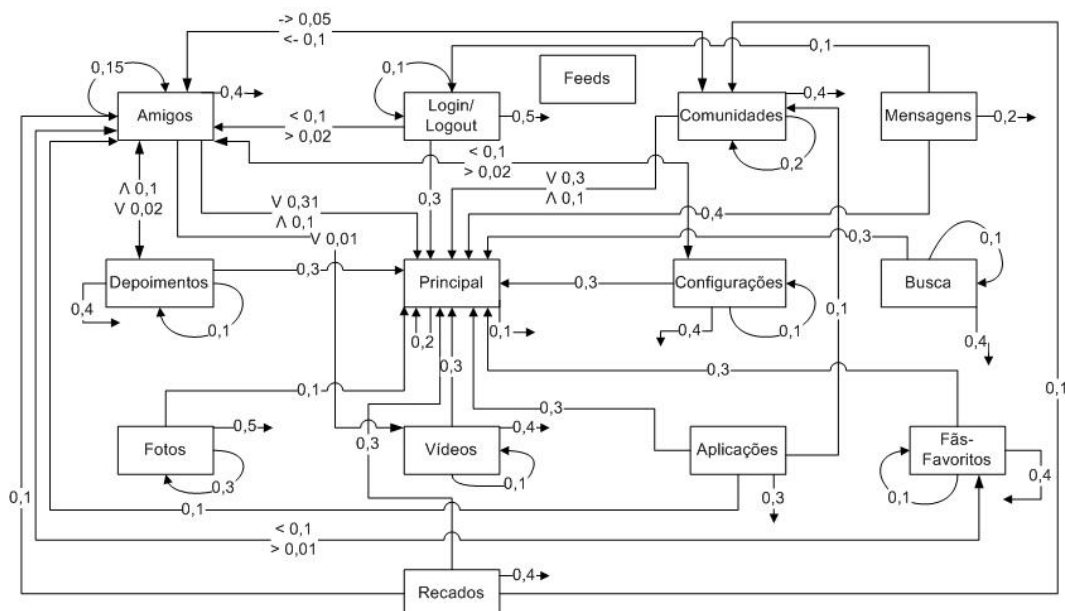


Figura 6: CBMG dos Navegadores intra-Orkut

As próximas sub-seções mostram como essa análise foi realizada e discute os resultados encontrados.

5.1 Definição das redes sociais

Dentre as 15 redes sociais com o maior número de transações na base de dados utilizada nesta caracterização, destacam-se 5 redes com propósitos diferentes. São elas: Facebook - Rede social de caráter geral, semelhante ao Orkut; Youtube - Compartilhamento de vídeos; MySpace - Criações de páginas Web personalizadas e interligadas; Last.fm - Tópicos relacionados a música; Fotolog - Compartilhamento de fotos. A Tabela 3 apresenta a quantidade de usuários, o número de sessões e a quantidade de transações que representa o uso das redes sociais.

Tabela 3: Sumário das redes sociais

Rede Social	# Usuários	#Transações
Youtube	26.351	36.483
MySpace	4.779	11.321
Last.fm	1.193	893
Facebook	2.322	5.075
Fotolog	32.753	73.423

Para caracterizar as transações das redes sociais definidas (Youtube, MySpace, Fotolog, Last.fm e Facebook) é necessário identificar quais as transações que fazem parte das sessões de usuários do Orkut que iniciam o uso dessas outras redes sociais. Por exemplo, só consideramos transações de uso do Facebook por parte de um usuário quando tal transação é realizada enquanto há uma sessão do Orkut aberta desse mesmo usuário.

5.2 Grupos de usuários na navegação em outras redes sociais dentro de sessões do Orkut

Partindo da caracterização dos grupos de usuários do Orkut, foi identificado como e a partir de qual funcionalidade do Orkut os usuários acessam as outras redes sociais. Para

facilitar a análise desse comportamento de acesso dos usuários do Orkut, foram gerados três novos CBMGs para representação do acesso inter-redes sociais a partir do Orkut dos grupos de usuários identificados na análise intra-Orkut. A Figura 7 mostra o comportamento dos usuários *Objetivos* na navegação do Orkut para as outras redes sociais. Observa-se que usuários *Objetivos* do Orkut acessam o Fotolog, o Last.fm e o Youtube a partir da funcionalidade Principal. Nota-se que o uso do Fotolog ocorre a partir das funcionalidades Amigos e Configurações.

O padrão de acesso de outras redes sociais a partir de sessões de usuários *Moderados* do Orkut pode ser observado na Figura 8. Percebe-se a existência do acesso ao Youtube a partir da funcionalidade Vídeos do Orkut. Provavelmente, isso ocorre devido a propriedade do Orkut em disponibilizar vídeos do Youtube em sua página. Percebe-se também que os usuários acessam o Orkut a partir do Youtube e também a partir do Fotolog. Isto pode indicar a necessidade de utilização de imagens e vídeos, seguido da necessidade de uso de uma rede social de objetivo geral como o Orkut. Os usuários *Navegadores*, Figura 9, assim como os *Moderados* navegam em maior parte do tempo do Orkut para o Youtube e para o Fotolog.

6. CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho foi realizada uma caracterização dos padrões de acesso dos usuários do Orkut sob duas perspectivas: a análise intra-Orkut, que enfatiza o padrão de navegação dos usuários pelas diversas funcionalidades do Orkut, e a análise inter-rede social, que aborda os acessos a partir do Orkut para outras redes sociais tais como YouTube, Fotolog, MySpace, Last.fm e Facebook. Para isso foi utilizado uma base de dados real coletada de um provedor de acesso de Internet banda larga residencial. Através dos dados coletados foram gerados os *Customer Behaviour Model Graph* (CBMGs) para representação do padrão de uso dos usuários

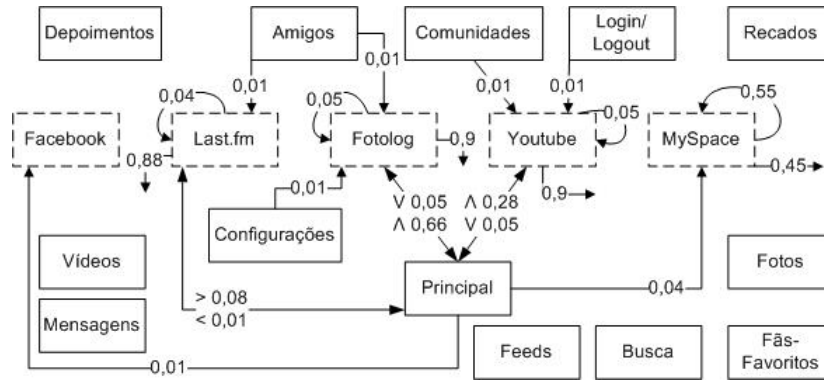


Figura 7: CBMG dos *Objetivos* com as redes sociais

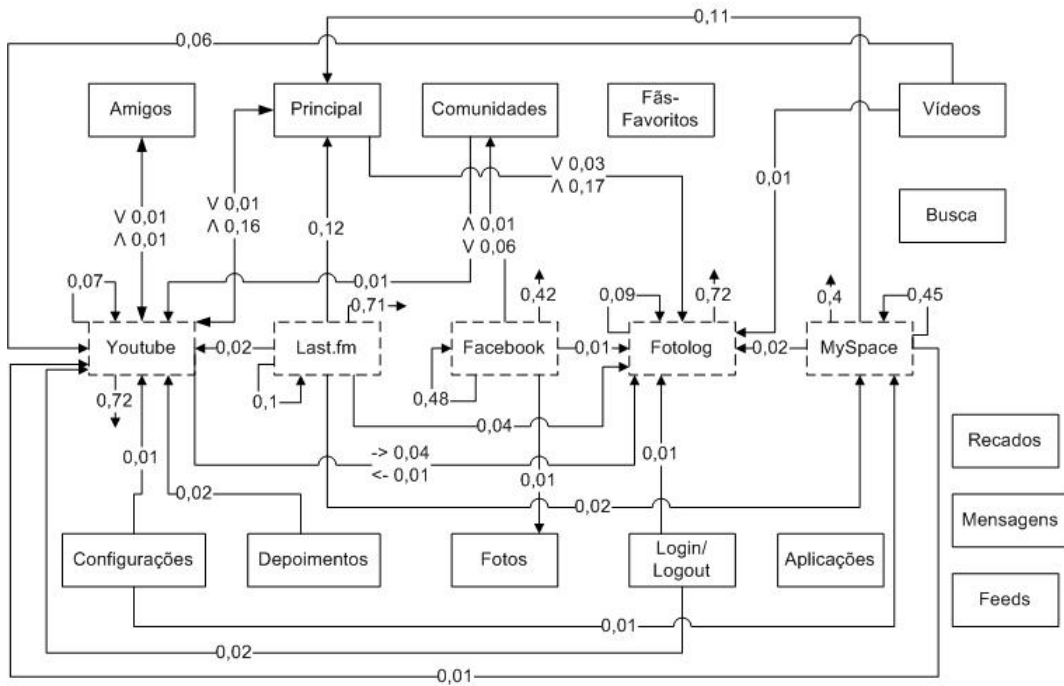


Figura 8: CBMG dos *Moderados* com as redes sociais

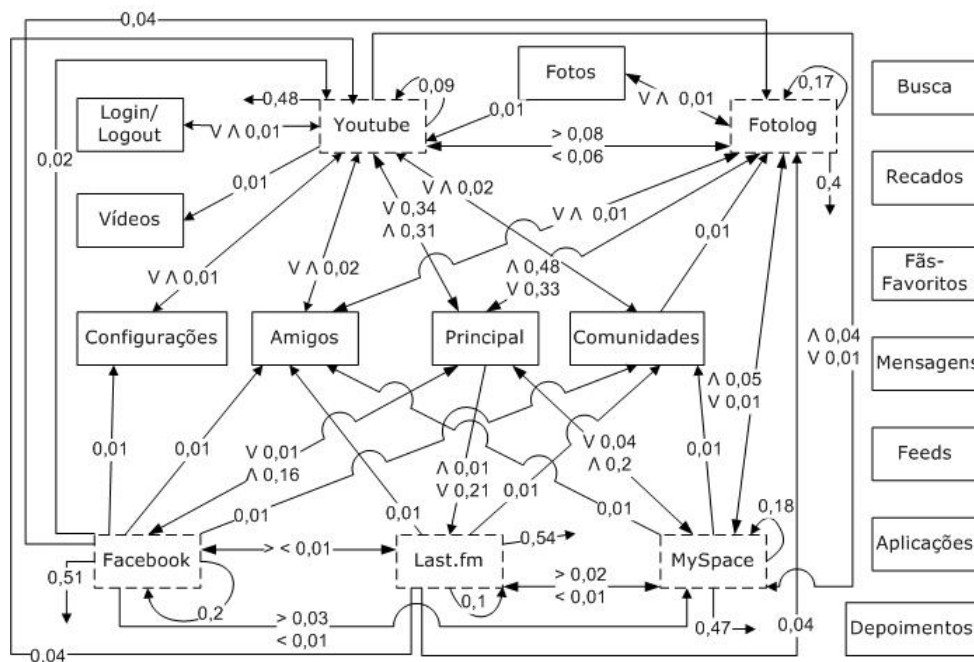


Figura 9: CBMG dos Navegadores com as redes sociais

do Orkut. Os CBMGs foram agrupados utilizando *k-means* e, então classificados em três perfis de usuários: *Objetivos*, *Moderados* e *Navegadores*. Em geral todos os usuários se concentram nas funcionalidades Amigos, Principal, Comunidades e Fotos do Orkut. Os *Objetivos* possuem a probabilidade de encerrarem suas sessões dentro do próprio estado que a iniciaram, enquanto os *Navegadores* já navegam por um número maior de páginas. A grande maioria dos acessos analisados utilizam o Fotolog e Youtube em conjunto com Orkut, indicando uma possível necessidade dos associados do Orkut para o compartilhamento de fotos e vídeos.

Como trabalho futuro, sugerimos a comparação dessa análise com novos dados reais de um provedor de acesso para identificar o crescimento e dinamismo dos usuários de redes sociais. Através de uma comparação entre as navegações dentro de um determinado prazo de tempo, se pode entender melhor a curva de crescimento e ou a aprendizagem dos usuários em relação a rede social utilizada.

7. REFERÊNCIAS

- [1] Yong-Yeol Ahn, Seungyeop Han, Haewoon Kwak, Sue Moon, and Hawoong Jeong. Analysis of topological characteristics of huge online social networking services. In *ACM - WWW'07*, Banff, Canada, 2007.
- [2] Fabrício Benevenuto, Adriano Pereira, Tiago Rodrigues, Virgílio Almeida, Jussara Almeida, and Marcos Gonçalves. Avaliação do perfil de acesso e navegação de usuários em ambientes web de compartilhamento de vídeos. In *ACM - WebMedia'09*, Fortaleza, Brazil, 2009.
- [3] Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. Characterizing user behavior in online social networks. In *ACM - IMC '09*, Chicago, USA, 2009.
- [4] Wen-Yen Chen, Jon-Chyuan Chu, Junyi Luan, Hongjie Bai, Yi Wang, and Edward Chang. Collaborative filtering for orkut communities: discovery of user latent behavior. In *ACM - WWW '09*, Madrid, Spain, 2009.
- [5] Minas Gjoka, Michael Sirivianos, Athina Markopoulou, and Xiaowei Yang. Poking facebook: characterization of osn applications. In *ACM - WOSP'08*, Seattle, USA, 2008.
- [6] Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. A few chirps about twitter. In *ACM - WOSP'08*, Seattle, USA, 2008.
- [7] Daniel Menasce and Virgílio Almeida. *Scaling for E-Business: Technologies, Models, Performance, and Capacity Planning*. Prentice Hall, 2007.
- [8] Alan Mislove, Massimiliano Marcon, Krishna Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *IMC '07: ACM SIGCOMM*, San Diego, USA, 2007.
- [9] Ellen Spertus, Mehran Sahami, and Orkut Buyukkokten. Evaluating similarity measures: a large-scale study in the orkut social network. In *ACM - KDD '05*, Chicago, USA, 2005.
- [10] Thiago Tavares, Regina Santana, Marcos Santana, and Júlio Estrella. Caracterização dos dados trocados entre serviços web e provedores de serviços. In *ACM - WebMedia'08*, Vila Velha, Brazil, 2008.
- [11] Monique Vieira, Bruno Fonseca, Rodrigo Damazio, Paulo Golgher, Davi Reis, and Berthier Ribeiro-Neto. Efficient search ranking in social networks. In *ACM - CIKM'07*, Lisbon, Portugal, 2007.
- [12] Yadong Zhou, Xiaohong Guan, Zhefei Zhang, and Beibei Zhang. Predicting the tendency of topic discussion on the online social networks using a dynamic probability model. In *ACM - WebScience '08*, Pittsburgh, USA, 2008.