

# Uma Caracterização dos Padrões de Navegação de Usuários em uma Aplicação Social de Streaming de Vídeo

Mariana Arantes, Flavio Figueiredo, Jussara M. Almeida

<sup>1</sup>Departamento de Ciência da Computação  
Universidade Federal de Minas Gerais

{mariana.arantes, flaviov, jussara}@dcc.ufmg.br

**Abstract.** *In this paper, we present a characterization of user behavior in video streaming applications. Our work is focused on providing insights to the following questions: (1) What are the most common websites that lead users to video content in social media video applications? (2) What is the browsing behavior of users within the application until they leave? (3) How exposed are users to ads in these applications? Using user browsing data from a large university campus in Brazil, we focus our case-study application on YouTube, the largest video application online. Different from previous studies, we characterize the browsing behavior of individual users. Moreover, while tackling the third question, we also characterize user exposure to video ads, a new kind of advertisement that was mostly unaccounted for by previous work. Our main results show that: (1) users usually reach content using different links depending on the kind of content (e.g., comedy videos vs news videos); (2) users tend to use search engines and related video lists to follow up on new content after viewing a video; and (3) video ads account for a much higher exposure to ad content than link ads.*

**Resumo.** *Neste trabalho, é apresentada uma caracterização do comportamento de usuários em aplicações de mídia social de streaming de vídeo online. A caracterização é feita com o objetivo de responder a três perguntas motivadoras: (1) Quais fontes externas (websites) mais frequentemente levam usuários para vídeos? (2) Como é o comportamento de navegação de usuários dentro da aplicação de streaming de vídeo? (3) Quão expostos os usuários estão a diferentes tipos de propaganda em tais aplicações? Usando uma base de dados de navegação de usuários de um grande campus universitário brasileiro, estudou-se o comportamento de usuários no YouTube, a maior aplicação de streaming de vídeo atualmente. Diferente de estudos passados, neste trabalho é caracterizado o comportamento individual de usuários na aplicação. Além disso, o acesso aos dados possibilitou a análise do comportamento de usuários quando expostos a um novo tipo de propaganda online, as propagandas em formato de vídeo. Os principais resultados mostram que: (1) os links que mais frequentemente levam usuários para vídeos do YouTube variam dependendo da categoria do vídeo, (2) após visualizarem um vídeo, usuários tendem a usar máquinas de busca e listas de vídeos relacionados para continuarem navegando na aplicação, e (3) propagandas no formato de vídeo tendem a atrair maior atenção dos usuários do que propagandas tradicionais em links.*

## 1. Introdução

Uma das maiores fontes de renda na Internet hoje em dia são as aplicações de streaming de vídeo online. Em alguns casos mais populares de tais aplicações, como as de vídeos de mídia social, usuários têm acesso gratuito a uma imensa quantidade de conteúdo. Ao mesmo tempo, os provedores de serviço fazem uso de propagandas, que são expostas aos usuários, para sustentarem os seus modelos de negócio. Neste contexto, a manutenção da atenção dos usuários na aplicação, e ao mesmo tempo a maior exposição às propagandas, são etapas cruciais para a manutenção de um sistema economicamente sustentável para provedores, produtores e consumidores de conteúdo.

O foco desse trabalho se encontra nos padrões de comportamento de navegação de usuários em sistemas de streaming de vídeo. Em especial, é apresentada uma caracterização de todo o processo de navegação dos usuários desde o momento em que eles chegam à aplicação até o momento em que eles saem dela. Com esta caracterização, é possível entender: características dos links de entrada para aplicações de streaming de vídeo; o fluxo de atenção dos usuários dentro da aplicação; e características dos links de saída; Além disso, é possível caracterizar a exposição dos usuários às propagandas, a maior fonte de renda em aplicações de streaming [Gill et al. 2013, Krishnan and Sitaraman 2013, Farahat and Bailey 2012].

O estudo é focado nas seguintes perguntas:

- Q1: Quais links da Internet mais frequentemente levam usuários para vídeos? Existe alguma relação entre os links de entrada<sup>1</sup> e o tipo do conteúdo<sup>2</sup>?
- Q2: Como é o comportamento do usuário dentro da aplicação de streaming de vídeo? Os usuários fazem uso dos links de saída expostos na aplicação, isto é, eles clicam em URLs que são apresentadas nas páginas dos vídeos visualizados?
- Q3: Quais as características do tempo de exposição dos usuários às propagandas em vídeo? Existe diferença entre esse novo tipo de publicidade e o mais tradicional (propaganda em links)?

Usando uma base de dados de comportamento de navegação dos usuários de um grande campus universitário do Brasil, estudou-se a aplicação de streaming de vídeo YouTube. Esta base foi coletada a partir do ponto de acesso da universidade, assim representando o comportamento de acessos ao YouTube de toda a comunidade universitária. Inicialmente a base de dados era composta de 9,854,903 requisições relacionadas ao YouTube. Esta base foi posteriormente filtrada, para remover requisições automaticamente geradas pelo YouTube. Após a filtragem, restaram um total de 690,713 requisições para 123,694 vídeos distintos, feitas por 24,509 usuários únicos (identificados pelo endereço IP anonimizado).

Como mencionado, diferente de trabalhos passados (ver Seção 2), neste artigo foi feita uma análise do comportamento individual dos usuários desde o acesso à aplicação até a saída dela. Além disso, com o acesso a base de dados foi possível fazer uma caracterização da exposição dos usuários a vídeo-propagandas, uma nova forma de publicidade online.

---

<sup>1</sup>As URLs dos *websites* que levam usuários para o YouTube.

<sup>2</sup>O tipo de conteúdo é capturado pela categoria do vídeo (e.g., carros, animais, blogs etc) no YouTube.

É importante ressaltar que as respostas das três perguntas acima se complementam, dando assim uma visão geral do comportamento de usuários em aplicações de streaming de vídeo. Q1 foca nas características de como usuários iniciam o acesso ao YouTube. Q2 visa explicar o comportamento dos usuários dentro da aplicação até o momento em que eles saem dela. Enquanto isso, Q3 visa entender a exposição dos usuários às propagandas de vídeo quando dentro do YouTube.

Os principais resultados da caracterização mostram que: (1) os links que mais frequentemente levam usuários para vídeos do YouTube variam dependendo da categoria do vídeo, (2) após visualizarem um vídeo, usuários tendem a usar máquinas de busca e listas de vídeos relacionados para continuarem navegando na aplicação, e (3) propagandas no formato de vídeo tendem a atrair maior atenção dos usuários do que propagandas tradicionais (com URLs).

O restante do trabalho está organizado da seguinte forma: a próxima seção apresenta os trabalhos relacionados e, em seguida, a Seção 3 descreve a coleta de dados e metodologia. A Seção 4 contém os resultados em relação a Q1, Q2, e Q3. Por fim, a Seção 5 apresenta as conclusões e discussões sobre trabalhos futuros.

## 2. Trabalhos Relacionados

Gill *et al.* [Gill et al. 2007] caracterizou os padrões de tráfego para o YouTube a partir de dados de uma rede local. Os autores estavam motivados pela mudança, ocorrida na época, na forma como conteúdo online era gerado, deixando de ser criado por editores e administradores e passando a ser criado majoritariamente pelos próprios usuários. O trabalho focou nas características dos vídeos assistidos no campus de uma universidade Estadunidense e os autores caracterizaram a popularidade, bitrate e os padrões diários/semanais de acesso aos vídeos. De forma complementar, o trabalho de Zink *et al.* [Zink et al. 2009] também focou em uma caracterização de dados de um campus universitário. No entanto, os autores tiveram como maior objetivo ressaltar os ganhos, em redução de tráfego de rede, que sistemas de proxy e redes de distribuição P2P podem obter no streaming de vídeos. Diferente dos trabalhos mencionados, o presente trabalho foca em entender o comportamento de navegação de usuários, como tal comportamento é relacionado ao tipo de conteúdo (capturado pela categoria dos vídeos) e como é a exposição dos usuários a diferentes tipos de propaganda (links e propagandas em forma de vídeo). Assim, é apresentada uma visão mais detalhada de como usuários navegam em sistemas de streaming de vídeo, complementando os trabalhos citados.

Ja os trabalhos de Benevenuto *et al.* [Benevenuto et al. 2009] e Schneider *et al.* [Schneider et al. 2009] caracterizaram os padrões de acessos as redes Orkut e Facebook, respectivamente. Devido ao objetivo desses dois trabalhos ser a caracterização de usuários em redes sociais online, os autores focaram no comportamento dos usuários dentro das aplicações e nas possíveis ações realizadas por eles (curtir, compartilhar, visualizar perfil, etc.) e não aprofundaram no entendimento das relações entre os acessos a conteúdo e o tipo deles. De qualquer forma, é importante mencionar que assim como no trabalho aqui apresentado, o trabalho de Schneider *et al.* [Schneider et al. 2009] contabilizou uma baixa taxa de cliques em propagandas na forma de links. A baixa taxa de cliques também é comum em máquinas de busca online [Farahat and Bailey 2012, Gill et al. 2013].

Além de trabalhos focados em tráfego de redes de campus universitários, vários

outros artigos caracterizaram diferentes propriedades de aplicações de streaming de vídeo utilizando dados públicos da aplicação [Cha et al. 2009, Figueiredo et al. 2011, Broxton et al. 2011, Wattenhofer et al. 2012]. Em especial, o trabalho de Cha *et al.* [Cha et al. 2009] foi um dos primeiros a caracterizar propriedades dos vídeos compartilhados no YouTube. Além desse, o trabalho de Figueiredo *et al.* [Figueiredo et al. 2011] focou no entendimento da evolução de popularidade de vídeos online. É ressaltado também o trabalho de Wattenhofer *et al.* [Wattenhofer et al. 2012], que correlacionou métricas de ações de usuários (número de comentários) com a popularidade dos vídeos no YouTube. Esse último foi feito usando uma base de dados privada da Google. É importante mencionar que nenhum desses trabalhos focou no comportamento de navegação dos usuários, possivelmente pelo difícil acesso a dados públicos que permitem tais estudos (em contrapartida aos dados de rede local de um campus universitário). Assim, nesse artigo é fornecido um novo e importante conhecimento, complementar aos trabalhos que estudam conteúdo de vídeo e acessos de usuários usando bases de dados públicas.

### 3. Metodologia

O objetivo dessa seção é apresentar a metodologia utilizada para a coleta e limpeza de dados. Para isso, a coleta de dados é apresentada na Seção 3.1 e o algoritmo para limpeza dos dados na Seção 3.2.

#### 3.1. Coleta de dados

A análise realizada neste trabalho é feita a partir de dados coletados em uma rede de dados de uma universidade pública brasileira, com aproximadamente 50 mil alunos. Através do uso de um roteador do ponto de acesso da universidade, a ferramenta Tstat [Finamore et al. 2011], de monitoramento de tráfego, foi executada para extrair informações de todos os pacotes de tráfego HTTP originados ou destinados a algum terminal da rede. A coleta capturou dados durante o período de Março a Novembro de 2014.

Os pacotes relacionados ao YouTube foram filtrados através dos campos *host* e *referrer*<sup>3</sup> do cabeçalho HTTP, que informam para qual servidor a requisição foi feita e de que endereço da Internet (uma URI ou link) a requisição foi originada<sup>4</sup>, respectivamente. Dessa forma é possível capturar o comportamento dos usuários desde o momento em que eles chegam à aplicação até o momento em que eles saem dela. Considerando somente as requisições relacionadas ao YouTube, os dados são compostos de 9,854,903 requisições.

Apesar de todas as requisições filtradas serem relacionadas ao YouTube, nem todas representam ações realizadas por usuários. Sempre que uma página é carregada, dezenas de requisições são geradas automaticamente pelo sistema para obter todo o conteúdo necessário, por exemplo imagens, propagandas ou partes de vídeos. Uma vez que o objetivo é caracterizar o comportamento dos usuários, o desafio encontra-se em identificar as requisições que representam os cliques dos usuários, conforme está descrito a seguir.

Para resolver o problema, foram construídas árvores de navegação para cada usuário, através do uso do algoritmo *Resurf* [Xie et al. 2013], que reconstrói atividades de navegação na Web a partir das informações fornecidas nos cabeçalhos HTTP dos pacotes. A próxima seção explica esse algoritmo e sua aplicação ao problema aqui apresentado.

---

<sup>3</sup>Geralmente chamado de *referer* devido a um erro de ortografia no RFC1945.

<sup>4</sup>Isto indica que o usuário clicou em um link para o YouTube a partir da página indicada pelo *referrer*.

### 3.2. Limpeza de Dados e o Algoritmo Resurf

O objetivo do *Resurf* é reconstruir o comportamento dos usuários ao navegarem na Web através dos pacotes de tráfego HTTP coletados, de forma a identificar quais *websites* são visitados por ações (cliques) de usuários e quanto tráfego é gerado.

Para cada usuário do sistema, identificado por um endereço IP anonimizado, o algoritmo cria uma árvore, onde os vértices representam as requisições e as arestas representam as relações entre elas. A árvore é construída com uso dos campos URI e *referrer* do cabeçalho HTTP. Existe uma aresta direcionada entre dois vértices somente se a requisição do vértice final foi originada pela requisição do vértice inicial, ou seja, o campo *referrer* do vértice final for igual ao campo URI do vértice inicial. Em suma, a árvore retrata a sequência de requisições feitas pelo usuário.

Devido ao fato de muitas requisições na Web serem geradas automaticamente, o método também propõe algumas heurísticas para identificar as requisições que interessam, ou seja, aquelas que representam cliques de usuários. Uma delas analisa as informações sobre o tipo e tamanho do conteúdo requisitado, de forma que somente requisições com tamanho acima de um valor determinado e cujo conteúdo seja HTML ou XML sejam aceitas como geradas por usuários. Outra estimativa se baseia na intuição de que páginas realmente acessadas por usuários são complexas e provavelmente geram muitas requisições. Desta forma, se após uma página ter sido requisitada, ela aparecer no campo *referrer* de alguma outra requisição, tal página provavelmente é uma requisição feita por um usuário. Com isso, uma árvore de requisições de usuários pode ser construída, onde cada vértice representa uma URI e um filho de um vértice representa uma URI que foi acessada após o pai (que neste momento é um *referrer*). Por questões de espaço, é sugerido ao leitor a referência original do *Resurf* para mais detalhes sobre o algoritmo [Xie et al. 2013]. De qualquer forma, é importante ressaltar que na implementação do *Resurf* realizada neste trabalho, os nós não foram filtrados de acordo com o tamanho da página. Esta escolha foi devida ao fato da ampla propagação de serviços de redirecionamento, como o bit.ly, que geram requisições de 0 bytes.

Além do uso do *Resurf*, foram usadas também ferramentas de desenvolvedores de navegadores web para identificar os principais tipos de acesso aos vídeos do YouTube. Esta foi uma etapa manual do trabalho onde foram analisados os padrões de URIs que levam ao conteúdo de vídeo. Existem duas formas majoritárias de acesso a vídeos no YouTube: (1) a página do vídeo no YouTube; e, (2) o uso de vídeos embutidos em outras páginas. O primeiro caso acontece quando se assiste a um vídeo na aplicação YouTube. O segundo acontece quando os vídeos são compartilhados em redes sociais, como o Facebook, ou anexados a outras páginas da Internet.

Com o objetivo de focar apenas nas requisições diretamente relacionadas a vídeos do YouTube, foram filtrados da saída do *Resurf* (a árvore de requisições), os vértices que não são ligados (ou como pai ou filhos) a um vértice de acesso a vídeo (casos 1 e 2 acima). A Figura 1 mostra um exemplo de saída do *Resurf* para um usuário. O nó raiz representa o usuário e os nós rotulados como *naosalvo*, *altamenteacido* e *jacarebanguela* representam as aplicações externas que levaram o usuário para o YouTube. Os nós inferiores com rótulo *subscribe\_widget* indicam que o usuário se inscreveu em um canal do YouTube, os nós com rótulo *watch?v* representam o acesso a um vídeo, o nó com rótulo *results?search* indica que o usuário realizou uma busca no YouTube e o nó *doubleclick* é um link de saída

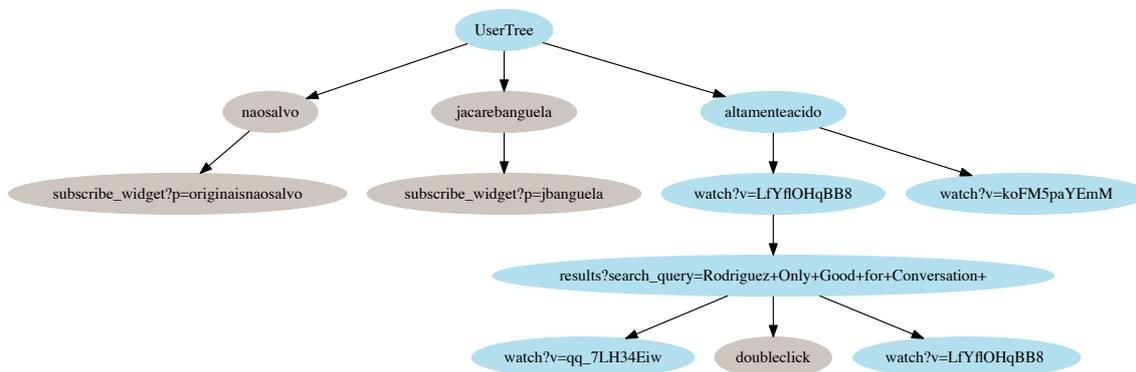


Figura 1. Árvore de um usuário.

Tabela 1. Volume de dados após limpeza no período de Março a Novembro 2014.

	Valor
# de vídeos	123,694
# de usuários (ips anonimizados)	24,509
# de requisições total (relacionadas a conteúdo)	690,713
# de requisições a vídeos via link externo ao YouTube (e.g., facebook.com)	43,967
# de requisições a vídeos via link interno ao YouTube (e.g., busca)	66,759
# de requisições diretas a vídeos (sem link de referer)	115,345
# de requisições com cliques de saída	10,695

que o usuário clicou. Após a aplicação do filtro, os nós em cinza são descartados porque eles não são diretamente relacionados a conteúdo.

Após a filtragem, a base de dados consiste de 690,713 requisições que podem ser consideradas ações de navegação relacionadas a conteúdo. Estas requisições foram feitas para 123,694 vídeos distintos por 24,509 usuários. Dentre essas requisições, 218,576 foram para visualizações de vídeos, sendo que 10,695 dessas visualizações foram sucedidas por cliques em endereços de páginas externas ao YouTube. A Tabela 1 apresenta algumas estatísticas da base de dados filtrada.

Todo vídeo do YouTube é associado a uma categoria que reflete o assunto ou tópico do seu conteúdo. A categoria de um vídeo é selecionada entre uma lista de 18 tópicos pre-definidos. Alguns exemplos são: Música e Esportes. Visando compreender quais tipos de conteúdo são acessados pelos usuários no campus, uma coleta da categoria dos vídeos identificados foi realizada através da API fornecida pelo próprio YouTube.

#### 4. Análise de Resultados

Esta seção apresenta os principais resultados da caracterização, provendo respostas para as três perguntas indicadas na Seção 1. A Seção 4.1 foca na primeira questão e aborda as principais fontes externas (i.e., links) que atraem usuários para o YouTube. A Seção 4.2 mostra os resultados quanto aos padrões de navegação do usuário dentro do YouTube. Por fim a Seção 4.3 aborda a última questão, mostrando resultados sobre a exposição do usuário a propagandas no YouTube, particularmente propaganda em forma de vídeos.

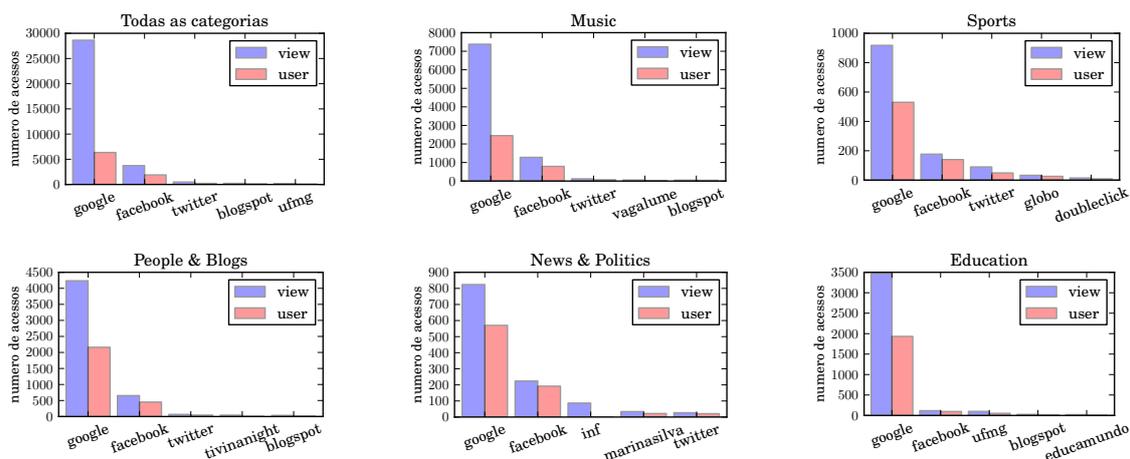


Figura 2. Links que geram acessos para conteúdo.

#### 4.1. Quais fontes externas mais levam usuários para vídeos do YouTube?

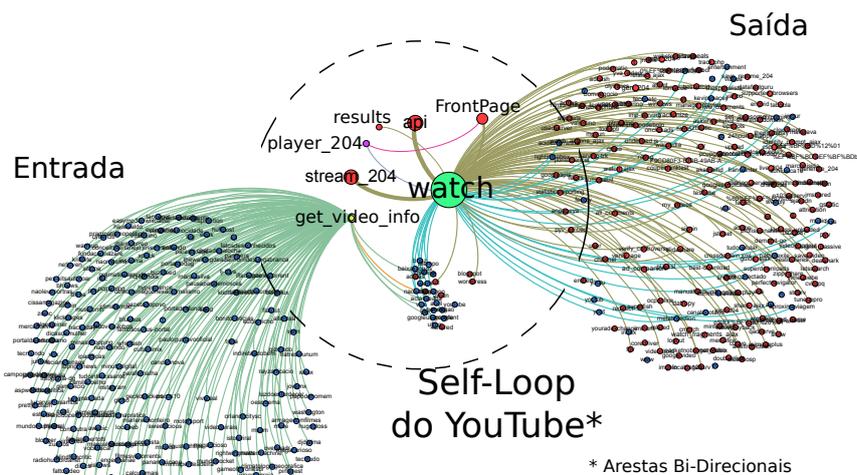
Para responder à primeira questão motivadora deste estudo, o conjunto das árvores de requisições dos usuários foi analisado e todos os vértices que representam acesso a vídeo através de links externos foram considerados. Ao todo foram recebidas 43,967 requisições para conteúdo vindas de 1,124 links diferentes. Essas requisições representam 19.45% do total de requisições para vídeos feitas no campus, mostrando que o volume de requisições geradas por links externos é significativo.

Como citado na Seção 3, as categorias dos vídeos assistidos no campus foram coletadas, possibilitando a análise da relação entre os links externos que geram acessos a conteúdo e os tipos desses conteúdos. A Figura 2 mostra os links que mais contribuíram para o acesso a conteúdo de forma geral e para 5 das 18 categorias existentes, a saber: *News & Politics*, *People & Blogs*, *Music*, *Sports* e *Education*. Essas categorias estão entre as mais populares na base de dados coletada.

O gráfico para todas as categorias mostra que os *websites* Google e Facebook são os que geram a maior quantidade de tráfego, representando 65% e 8.5% do total de requisições, respectivamente. Do mesmo modo é possível observar que ambos os *websites* também são os mais populares para todas as categorias mostradas, com destaque para a categoria de Educação (*Education*), onde o Facebook possui menor importância, representando apenas 2.8% das 4,152 requisições geradas para a categoria e o Google se destaca, representando 83.6% das requisições, o que mostra que o link externo Google tende a levar muitos usuários para vídeos relacionados a Educação.

O gráfico da categoria Música (*Music*) mostra que apesar dos *websites* Google, Facebook e Twitter serem os três mais populares, assim como acontece ao considerar todas as categorias juntas, o *website* brasileiro sobre músicas, conhecido como Vagalume, aparece como o quarto mais popular, indicando que links externos tendem a gerar mais acessos para categorias específicas. Esse indício também pode ser encontrado ao observar o gráfico da categoria de Pessoas e Blogs (*People & Blogs*), em que, apesar de representar apenas 0.7% das 5,545 requisições, o *website* Tivinanight, que realiza a cobertura de shows e eventos, portanto diretamente relacionado ao tema, é o quarto mais popular.

Ao analisar o gráfico de Esportes (*Sports*), é possível notar um maior impacto



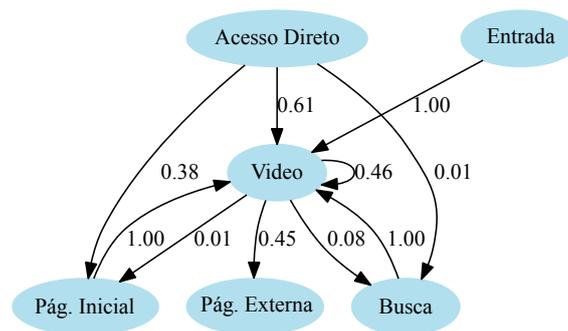
**Figura 3. Estrutura bow-tie do YouTube**

do *website* Twitter, que representa 6.5% dos acessos à categoria enquanto apenas 1.1% dos acessos ao considerar todas as categorias juntas. Por fim, a categoria de Notícias e Política (*News & Politics*) também possui indícios da relação entre links externos e tipos de conteúdo. Nessa categoria, é importante ressaltar que o *website* Twitter apareceu somente como o quinto mais popular e o *website* Facebook apresentou maior destaque do que nas outras categorias, representando 15.8% dos acessos. Outros dois *websites*, Inf e Marinasilva, ambos relacionados a política, foram também importantes para a geração de tráfego para a categoria, representando 6.1% e 2.4% dos 1,419 acessos. Nota-se que, apesar da grande popularidade de algumas aplicações, notavelmente Facebook e Google, como geradoras de acessos para o YouTube, outras fontes externas também são relevantes. Entretanto, as mesmas variam dependendo do tipo de conteúdo (categoria do vídeo).

#### 4.2. Como é o comportamento de navegação de usuários no YouTube?

Após a chegada dos usuários ao YouTube, é importante caracterizar o comportamento deles até o momento em que eles deixam a aplicação, de forma a obter informações sobre a exposição deles a vídeos e propagandas. Todas as ações do usuário durante esse período foram definidas como uma *sessão*. Para a obtenção desses dados, o primeiro passo consiste em utilizar os grupos de requisições de entrada, saída e requisições internas à aplicação, como definido na Seção 3, para a construção de uma estrutura *bow-tie* [Broder et al. 2000] do YouTube. Essa estrutura ajuda no entendimento visual do problema e é formada por três partes: no lado esquerdo estão os *websites* de entrada, no centro encontram-se as requisições internas e no lado direito encontram-se predominantemente as requisições de saída (isso acontece devido ao fato de algumas requisições do YouTube aparecerem como entrada e saída). Os nós da figura representam as requisições e as arestas indicam a sequência em que elas ocorrem. A Figura 3 mostra a estrutura.

Os nós rotulados como *get\_video\_info* e *watch* representam as exibições de vídeos, a primeira de forma embutida em outros *websites* e a segunda através do YouTube. As ações dos usuários de buscar por vídeos e acessar a página inicial da aplicação são representadas respectivamente pelos nós *results* e *FrontPage*. Essa estrutura mostra a existência de um ciclo de ações realizadas pelos usuários (destacado na figura como o *self-loop*),



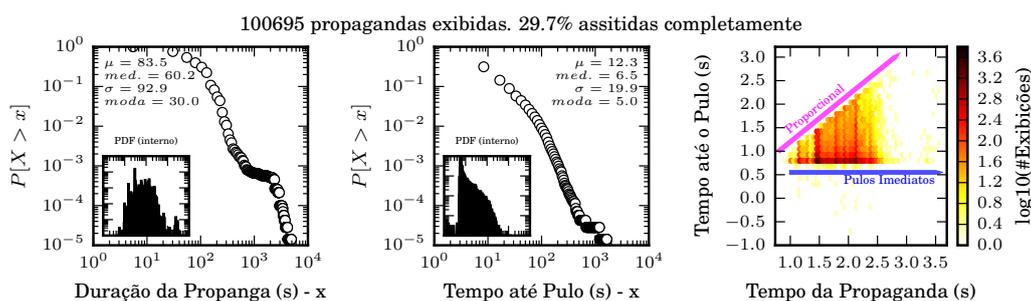
**Figura 4. CBMG das principais ações.**

onde eles basicamente navegam entre vídeos utilizando a máquina de busca do YouTube e acessando os vídeos expostos na página inicial. A estrutura também mostra alguns exemplos interessantes de outras ações que podem ser realizadas pelos usuários, como as representadas pelas requisições *stream\_204* e *player\_204* que são geradas a partir de uma requisição de acesso a vídeo. Através de uma verificação manual, acredita-se que essas requisições representam as interações do usuário com os vídeos, pulando (*skip*) partes ou acessando as propagandas escritas que aparecem dentro deles enquanto são executados.

O *bow-tie* mostra apenas uma visão geral da navegação dos usuários. Para entender o comportamento deles no YouTube de maneira mais detalhada, as transições entre as principais funcionalidades da aplicação foram representadas por um CBMG (*Customer Behavior Model Graph*) [Menasce and Almeida 2000], um modelo de grafo que representa transições entre estados de um sistema. Devido ao fato das ações que podem ser realizadas por usuários no YouTube não serem exclusivas, ou seja, o usuário pode executar várias ações a partir de uma mesma página de vídeo, como publicar um comentário e marcar que gostou do vídeo, foi decidido usar apenas as ações que representam a navegação do usuário entre páginas para criar o CBMG. Essas ações podem ser de cinco tipos: Página de Entrada, Página de Saída (Externa), Página Inicial do YouTube, Página de Busca e Página de acesso a vídeo. A Figura 4 mostra o CBMG gerado a partir dos dados coletados. O rótulo de cada transição representa a probabilidade da dada transição ocorrer (i.e., o usuário executar uma ação após a outra).

Os usuários chegam a vídeos do YouTube através de cinco formas (representadas pelos nós de mesmo nome na figura): acesso direto, página inicial do YouTube, página de busca, página de entrada ou através de outro vídeo. A partir de um vídeo, o usuário pode fazer uma busca por outros vídeos, ir para a página inicial, usar a lista de vídeos relacionados para acessar diretamente outro vídeo ou pode sair da aplicação. É importante ressaltar que existem três ciclos a partir de um acesso a vídeo que mantêm o usuário na aplicação, o primeiro é entre vídeos e é muito popular, pois em 46% das vezes que usuários deixam uma página de um vídeo, eles vão para outro vídeo. O segundo acontece em 8% das vezes e mantém os usuários navegando entre vídeos e páginas de busca. O terceiro, menos frequente, mostra que usuários acessam vídeos e voltam para a página inicial do YouTube e a partir de lá acessam outros vídeos.

Após analisar o comportamento dos usuários dentro da aplicação, o próximo passo consiste em descobrir se esses usuários clicam nos links (URLs) expostos nas páginas do sistema, que na maioria dos casos, são propagandas. Como mencionado da Seção 3, foram feitas 218,576 requisições para visualizações de vídeos no campus e 10,695 dessas



**Figura 5. Exposição dos usuários a propagandas.**

visualizações foram sucedidas por cliques em endereços de saída do YouTube, levando a uma taxa de clique de saída de 0.047. Embora seja um valor pequeno, ele é maior do que o esperado ao comparar com o trabalho de Schneider *et al.* [Schneider et al. 2009], em que a taxa de saída foi menor que 0.1%. Isso pode ser explicado devido a dois motivos: (1) diferenças no comportamento dos usuários na aplicação YouTube em relação ao Facebook (caso de estudo dos autores); (2) ao fato do referido trabalho considerar somente as propagandas e aqui, todos os links de saída do YouTube foram considerados.

### 4.3. Propagandas em forma de vídeos

Visando entender o comportamento dos usuários quando expostos a propagandas em forma de vídeo, os padrões de endereços que representam requisições para tais propagandas foram identificados manualmente. Ao todo, foram encontradas 100,695 requisições, o que implica que 44.54% dos acessos para vídeo exibiram um vídeo de propaganda.

Na Figura 5 são apresentadas informações sobre as requisições de propagandas em formato de vídeo. No lado esquerdo da figura, é apresentada a distribuição complementar de probabilidade cumulativa (CCDF) das durações das propagandas. O eixo-y representa a fração de requisições que obtiveram um valor superior à duração apresentada no eixo-x. Para fins de melhor entendimento, também é apresentado no gráfico uma visualização do formato do histograma dos dados. A mediana nesse gráfico representa o ponto em que  $P[X > x]$  (eixo-y) tem valor igual a 0.5, ou seja, 50% das requisições de propaganda tem tempo maior que 60 segundos. Já a média é igual a 83 segundos e o desvio padrão é 92.9 segundos. Esta média é alta pois existem raros casos (probabilidade menor que 0.01%), em que as propagandas têm duração maior que 1000 segundos (16 minutos aproximadamente). A propaganda mais longa tem duração maior que 10,000 segundos (2.8 horas). Embora esses casos sejam raros, buscando na Internet em *websites* de discussões e no próprio YouTube, foi averiguado que tais propagandas realmente existem<sup>5</sup>. De qualquer forma, a grande maioria das propagandas tem duração igual ou superior a 30 segundos.

Até o momento a discussão focou apenas nas características do tempo das propagandas. Porém, como o YouTube permite o “pulo” de um propaganda, ou seja, ele permite que o usuário pare a exibição da mesma e vá direto para o vídeo, com base no tempo do pulo é possível entender melhor o tempo que usuários possivelmente dedicam assistindo vídeo-propagandas<sup>6</sup>. Inicialmente, foi identificado que 29.7% das pro-

<sup>5</sup><http://www.youtube.com/watch?v=QUs82pvfKV8>

<sup>6</sup>Não é possível saber se o usuário estava visualizando a propaganda durante sua exibição. De qualquer forma, o tempo do pulo pode ser usado como evidência de que o usuário assistiu parte da propaganda, uma vez que ele estava com o navegador aberto na propaganda quando usou a função de pulo.

pagandas foram assistidas completamente. Ou seja, o usuário não utilizou o recurso de pulo em nenhum momento. Esse resultado foi de certa forma surpreendente, considerando a baixa taxa de cliques em links de saída observada nos dados (4.7%), além da taxa ainda mais baixa de cliques em propagandas em outros aplicativos (menos de 0.1%) [Farahat and Bailey 2012, Schneider et al. 2009]. Considerando estes resultados, era esperado que os usuários tivessem bem menos interesse nas vídeo-propagandas. Embora são dois meios distintos de exposição e publicidade (vídeos vs links), a alta taxa de cliques em vídeo-propagandas (29.7%), mais que duas ordens de grandeza maior do que as taxas de cliques em propagandas tradicionais, serve como evidência da maior eficácia desse método como meio de publicidade online.

Com base na distribuição do tempo até o pulo (CCDF na Figura 5 centro), inicialmente foi observado que 35% das propagandas tiveram um pulo em menos de 6 segundos. Além disto, existe um pico na distribuição em torno dos 5 segundos (veja gráfico interno), o tempo mínimo permitido pelo YouTube. Isso indica que quando os usuários usam o pulo, ou seja, perdem interesse na propaganda, isto ocorre com mais frequência assim que permitido pela aplicação. De qualquer forma, o tempo médio até um pulo foi de 12.3 segundos e o desvio padrão foi de 19.9 segundos. Embora exista uma concentração em torno do tempo mínimo, em 19.8% dos casos os usuários assistiram mais do que a média (12.3 segundos). Para entender melhor esse comportamento, o tempo da propaganda e o tempo até o pulo foram correlacionados (Figura 5 na direita). Os dois eixos da figura estão em log e com base nela é possível ver dois comportamentos distintos. Como esperado, uma grande quantidade de pontos está agrupada ao redor dos 5 segundos ( $\log_{10}(5) = 0.68$  no eixo-y), independente do tempo da propaganda. Mais interessante é ver que existe um grupo de pontos que tem uma correlação entre o tempo de pulo e o tempo da propaganda. Esses casos possivelmente são usuários que assistem a propaganda em parte (proporcional ao tempo da mesma), antes de perderem interesse. Tal caso é interessante pois também mostra que propagandas em forma de vídeo aumentam o tempo de exposição mesmo quando os usuários decidem fazer uso da função pulo.

Por fim, é ressaltado uma diferença entre os resultados aqui apresentados e uma recente caracterização de propagandas em *websites* de streaming de vídeo de conteúdo profissional [Krishnan and Sitaraman 2013]. Usando dados proprietários de sites de notícias e esportes, o estudo mostra que usuários tendem a assistir totalmente 73% das propagandas exibidas no início de vídeos e 96% daquelas exibidas no meio dos vídeos. Aqui, os resultados mostram taxas menores, 29.7%. Existe uma diferença entre o YouTube e as aplicações exploradas pelos autores. Nessas aplicações, não se pode pular a propaganda, ou seja, quando um usuário não assiste uma propaganda na totalidade, tal usuário na verdade parou de assistir o vídeo do conteúdo como um todo. Em contrapartida, o YouTube permite o pulo (skip) das propagandas, mostrando um comportamento mais abrangente dos usuários que tem essa capacidade.

## 5. Conclusões

Neste trabalho foi realizada uma caracterização do comportamento dos usuários no YouTube utilizando dados de tráfego coletados de uma grande universidade pública brasileira. A caracterização foi realizada visando entender (1) quais websites mais frequentemente atraem usuários para o YouTube, (2) como e o padrão de navegação dos usuários durante uma sessão do YouTube e (3) quão expostos eles estão a vídeo propagandas. Os resultados

mostram que a quantidade de usuários atraídos para o YouTube através de outros *websites* é significativa, além de existir uma relação entre as categorias dos vídeos e os *websites* que mais frequentemente levam acessos a elas. A partir do momento que os usuários acessam um vídeo do YouTube, foi possível concluir que eles continuam a navegação entre vídeos através da lista de vídeos relacionados e da máquina de busca do YouTube. Os resultados também sugerem que vídeo-propagandas tendem a ser mais efetivas do que as propagandas exibidas nas páginas de vídeos em forma de links. Esses resultados são importantes pois podem ajudar anunciantes a entender o comportamento dos usuários e assim atingir melhores resultados com os anúncios. Como trabalho futuro, pretende-se estender a caracterização feita para outras aplicações (e.g., Facebook) bem como para analisar mais detalhadamente a exposição dos usuários a diferentes tipos de propaganda.

## Referências

- [Benevenuto et al. 2009] Benevenuto, F., Rodrigues, T., Cha, M., and Almeida, V. (2009). Characterizing User Behavior in Online Social Networks. In *Proc. IMC*.
- [Broder et al. 2000] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. (2000). Graph structure in the web. *Computer networks*, 33(1):309–320.
- [Broxton et al. 2011] Broxton, T., Interian, Y., Vaver, J., and Wattenhofer, M. (2011). Catching a viral video. *Journal of Intelligent Information Systems*, pages 1–19.
- [Cha et al. 2009] Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y., and Moon, S. (2009). Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Transactions on Networking*, 17(5):1357–1370.
- [Farahat and Bailey 2012] Farahat, A. and Bailey, M. C. (2012). How effective is targeted advertising? In *Proc. WWW*, page 111.
- [Figueiredo et al. 2011] Figueiredo, F., Benevenuto, F., and Almeida, J. (2011). The tube over time: characterizing popularity growth of youtube videos. In *Proc. WSDM*.
- [Finamore et al. 2011] Finamore, A., Mellia, M., Meo, M., Munafo, M. M., and Rossi, D. (2011). Experiences of internet traffic monitoring with tstat. *Network, IEEE*, 25(3):8–14.
- [Gill et al. 2007] Gill, P., Arlitt, M., Li, Z., and Mahanti, A. (2007). Youtube Traffic Characterization: A View From the Edge. In *Proc. IMC*.
- [Gill et al. 2013] Gill, P., Erramilli, V., Chaintreau, A., Krishnamurthy, B., Papagiannaki, D., and Rodriguez, P. (2013). Follow the Money: Understanding Economics of Online Aggregation and Advertising. In *Proc. IMC*.
- [Krishnan and Sitaraman 2013] Krishnan, S. S. and Sitaraman, R. K. (2013). Understanding the Effectiveness of Video Ads: A Measurement Study. In *Proc. IMC*.
- [Menasce and Almeida 2000] Menasce, D. and Almeida, V. (2000). Scaling for e-business: Technologies, models, performance, and capacity planning.
- [Schneider et al. 2009] Schneider, F., Feldmann, A., Krishnamurthy, B., and Willinger, W. (2009). Understanding online social network usage from a network perspective. In *Proc. IMC*.
- [Wattenhofer et al. 2012] Wattenhofer, M., Wattenhofer, R., and Zhu, Z. (2012). The YouTube social network. In *Proc. ICWSM*.
- [Xie et al. 2013] Xie, G., Iliofotou, M., Karagiannis, T., Faloutsos, M., and Jin, Y. (2013). Resurf: Reconstructing web-surfing activity from network traffic. In *Proc. IFIP*, pages 1–9. IEEE.
- [Zink et al. 2009] Zink, M., Suh, K., Gu, Y., and Kurose, J. (2009). Characteristics of YouTube network traffic at a campus network - Measurements, models, and implications. *Computer Networks*, 53(4):501–514.