

Centralidade em Redes P2P de Transmissão ao Vivo

João F. A. e Oliveira, Pedro de C. Gomes, Alex B. Vieira, Sérgio V. A. Campos

¹Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte, Minas Gerais, Brasil 30332-0250

{holiver,pcgomes,borges,scampos}@dcc.ufmg.br

Abstract. *In peer-to-peer live streaming systems a few nodes have a special role on content transmission as they have much higher upload rate. Understanding how this pattern emerge is important in several ways, such as to defend these networks against attacks like data pollution or denial of service. This work addresses the problem of identifying these nodes through the use of centrality-like metrics on a real, SopCast-based, p2p live streaming environment. Results show that degree and closeness centrality gives the best correlation with upload rate from all the metrics tested.*

Resumo. *Em redes par-a-par de transmissão ao vivo existem alguns poucos pares que desempenham um papel especial, eles apresentam uma taxa de transmissão significativamente maior que a dos demais. Entender como pares especiais se formam é importante de várias maneiras, como na defesa da rede contra ataques de poluição de conteúdo ou negação de serviço. Este trabalho trata o problema de identificação dos nós especiais através do uso de métricas de centralidade. Os experimentos foram conduzidos em canais criados na rede SopCast, uma das aplicações p2p mais populares para transmissão ao vivo. Os resultados indicam que a centralidade de grau e closeness, entre todas as métricas utilizadas, tem as melhores correlações com a taxa de transmissão de um participante.*

1. Introdução

Ao longo dos últimos anos, redes par-a-par, ou p2p, tem sido usadas para a distribuição de diferentes tipos de mídias através de compartilhamento de arquivo ou na forma de conteúdo sob demanda. Muitos desses sistemas, como o Gnutella [Ripeanu et al. 2002], Kazaa [Leibowitz et al. 2003], eDonkey [Tutschku 2004] e BitTorrent [Cohen 2003], são hoje famosos, consolidados. Seus respectivos sucessos, enquanto sistemas distribuídos, são oriundos de importantes características atribuídas a seus projetos em p2p como tolerância a falha, escalabilidade e otimização do uso de recursos. Todavia, a principal característica de redes p2p é o fato do modelo de transmissão ser cooperativo, sendo que durante a entrega do conteúdo os pares assumam comportamento tanto de clientes quanto de servidores [Schollmeier 2001].

Esse trabalho estuda um tipo de rede p2p que é utilizada para a distribuição ao vivo de conteúdo. Esta tem todas as vantagens e desvantagens de um projeto p2p, com algumas restrições adicionais tais como as necessidades de baixa latência de exibição, características de tempo real, baixa resiliência a poluição e um ponto central de falha associado à fonte da mídia. As redes p2p de distribuição de mídia estudadas nesse artigo são as do tipo *mesh-pull*, cuja descrição será dada nos parágrafos a seguir.

A maioria dos sistemas par-a-par de transmissão ao vivo são baseadas no sistema do Coolstreaming, conhecido também como rede sobreposta orientada a dados (*data-driven overlay*), onde um nó sempre encaminha dados para outros que os esperam, sem papéis prescritos como pai e filho. Em outras palavras, em vez de determinar uma estrutura específica que restrinja a transmissão, a disponibilidade do dado é o que guia a direção do fluxo [Zhang et al. 2005].

O gerenciamento de parcerias, que dá o tipo de organização da rede, indica que os pares devem ao entrar na rede requisitar parceiros de um *bootstrap* que responde com um subconjunto aleatório de pares já presentes na rede. Apesar desse método de inicialização da rede parecer gerar grafos aleatórios foi observado em [Wu et al. 2008] que de fato as redes formadas são do tipo "mundo pequeno"[Watts 2004]. Este nome ilustra características desse tipo de rede, que apresenta uma distância pequena entre dois vértices quaisquer e participantes que se agrupam em pequenas comunidades.

Sobre o fluxo contínuo de mídia, segundo o protocolo, deve ser dividido em pequenos pedaços, chamados *chunks*, que são as unidades trocadas pelos pares. A fonte da mídia gera continuamente esses *chunks*, cada par da rede (incluindo a fonte) anuncia para seus vizinhos quais *chunks* eles tem e, finalmente, todos escalonam quais pedaços irão pegar de quais parceiros embasados nas suas necessidades individuais e no mapa de *chunks* dos parceiros.

É importante observar que esse esquema de transmissão implica uma característica: para cada *chunk* disseminado é formado um grafo de distribuição *tree-like* com raiz no servidor de mídia. Assim, os dados são injetados na rede através da fonte de conteúdo, os vizinhos da fonte podem requisitar tais dados assim que souberem que eles estão disponíveis, depois os vizinhos dos vizinhos e assim por diante.

Enfim, é possível observar um grafo de propagação de dados em forma de árvore, uma *overlay* com modelo mundo pequeno e, se a rede estiver sobre a Internet, provavelmente uma *underlay* com modelo livre de escala [Faloutsos et al. 1999]. O interessante é que nenhum modelo isoladamente consegue explicar uma característica observada em experimentos reais: mais da metade de todos os dados é transmitida por praticamente cinco por cento dos nós da rede, suas taxas de *upload* são comumente maiores que a da fonte e tais pares serão chamados de super nós de agora em diante. Esse comportamento foi observado em diversas medições usando SopCast¹ sobre o PlanetLab² sem nenhuma tendência para que um par específico pertencesse ao conjunto dos super, o que traz a questão: Por que e como surgem os super nós?

Na verdade, essa questão apareceu primeiramente enquanto estudando a propagação de poluição e o papel essencial dos nós com alta taxa de *upload* no ataque. Um super poluidor pode deteriorar mais de 30% de todo *download* da rede e alcançar mais de 50% dos nós [Oliveira et al. 2009]. Responder a essa questão é importante porque se for conhecido como os super nós se formam será possível identificá-los e, então, construir mecanismos de defesa, de incentivo e de prioridade na entrega de dados mais eficientes e conscientes dessa classificação.

Finalmente, o objetivo e contribuição desse artigo é o levantamento de quais

¹<http://www.sopcast.com/>

²<http://www.planet-lab.org/>

métricas topológicas de centralidade podem ser melhores na determinação dos super nós, sendo este resultado obtido através da correlação de *ranks* de taxa de *upload* e de tais métricas calculados sobre dados reais e simulados. Na próxima seção serão apresentados os trabalhos relacionados. Na seção 3, é exposto o ambiente de estudo, um pouco sobre o sistema real estudado, o PlanetLab e sobre a simulação. A seção 4 trata diversas definições sobre redes complexas e procedimentos adotados nesta pesquisa. Os resultados sobre correlação são abordados na seção 5. E, finalmente, a seção 6 exhibe as conclusões do trabalho.

2. Trabalhos Relacionados

Sobre a aplicação de métricas de centralidade em sistemas de distribuição ao vivo P2P, destaca-se [Wu et al. 2008], que durante uma caracterização do sistema UUsee³, apresenta o uma correlação de taxa de *upload* e distribuição de graus entre todos os nós de um sistema. Este tipo de mensuração é amplamente aplicada em redes sociais tanto na Internet (Flicker⁴, Last.FM⁵, Facebook⁶ e outras como YouTube⁷ [Santos et al. 2009, Kumar et al. 2006]), quanto fora dela em cenários de resolução de problemas em grupo, política, desenvolvimento urbano e projetos de organizações como descrito no trabalho de [Freeman 1979].

Relativo ao tipo de rede estudada, alguns trabalhos, como [Ali et al. 2006, Hei et al. 2007, Silverston and Fourmaux 2007, Tran et al. 2004], realizam medições de parâmetros do comportamento dos pares. Entretanto, nesses trabalhos os dados capturados ficaram basicamente restritos à experiência dos coletores, ou participantes, e pouco abordam sobre visões mais abstratas da rede como um todo.

A proposta deste artigo foi motivada por pesquisas anteriores encontradas em [Oliveira et al. 2009], que aborda um ataque de poluição à rede na perspectiva de um nó que tem alta taxa de *upload* média. A noção de super nó implica um forte ponto de ataque para esse tipo de sistema.

3. Ambiente

As próximas subseções explicarão um pouco sobre as ferramentas usadas nesse trabalho, o SopCast, o PlanetLab e o simulador criado sobre a plataforma Oversim(OMNet++).

3.1. SopCast

Parte do conjunto de dados reais usado nesse experimento é o mesmo de [Oliveira et al. 2009]. Todos os dados foram gerados a partir da aplicação SopCast, um dos mais populares sistemas de distribuição de vídeo ao vivo. Um canal privado foi criado a fim de transmitir um fluxo de 120 *kpbs*. Somente pares que sabiam o *id* do canal poderiam ingressar na *overlay* específica e assistir seu conteúdo, ou seja, supõe-se um ambiente isolado de agentes externos (não haviam usuários reais, somente *crawlers* ou *bots*). Além disso, o *churn* da rede é garantido na forma de parcerias promíscuas, ao invés dos nós entrando e saindo, dado que trocas de parceiros são muito frequentes.

³<http://www.uusee.com/>

⁴<http://www.flickr.com>

⁵<http://last.fm>

⁶<http://www.facebook.com>

⁷<http://www.youtube.com>

Para cada rodada do experimento, esta metodologia de captura compreendeu, inicializar uma versão Linux do cliente SopCast, em um conjunto de 400 nós do PlanetLab. Ao final de cinco minutos todos os participantes deixavam a rede ao mesmo tempo. Apesar da aplicação não ter código-fonte aberto, a coleta é representada pelos *logs* do tráfego na camada de rede obtidas através da execução simultânea do Wireshark⁸.

3.2. PlanetLab

PlanetLab é um consórcio mundial de instituições de pesquisa que mantém um ambiente global para o desenvolvimento e testes de aplicações distribuídas. Cada instituição mantém um ou mais nós na Internet que são servidores de máquinas virtuais. Uma conta de acesso ao PlanetLab é chamada de *slice*. Se uma instituição mantiver pelo menos um nó em operação lhe é concedido o direito de criar *slices* e cada *slice* tem o poder de controlar um conjunto de máquinas virtuais em outros nós da rede.

O uso do PlanetLab trouxe muitas vantagens no que se refere a aquisição de dados. A mais óbvia foi a disponibilidade de centenas de nós, o que ajuda a recuperar dados mais representativos e confiáveis. Este experimento, por exemplo, se baseia em dados de cerca de 400 nós. Outra vantagem é que os nós do PlanetLab são dispersos tanto em redes distintas quanto geograficamente, isso evita que aspectos de localização mascarem o comportamento da rede. Por fim, o fato dos nós do PlanetLab possuírem IPs reais e não sofrerem filtragem de pacote evita a necessidade de tratar o problema do NAT (*Network Address Translation*) [Bellovin 2002], onde participantes da transmissão que estão atrás do mesmo *firewall* são identificados como um único participante.

3.3. OverSim

A simulação foi criada com base no Coolstreaming [Zhang et al. 2005] que descreve diversos algoritmos para o projeto de um sistema de distribuição de conteúdo p2p como gerenciamento de participantes, representação e troca do mapa de *bits* do *buffer* e escalonamento de *chunks*. Uma rodada da simulação usa 400 nós, simulando cinco minutos de interações da transmissão de um fluxo de 120kbps. O simulador foi desenvolvido com o suporte dos *frameworks* OMNet++⁹ e OverSim¹⁰ [Baumgart et al. 2007], eles oferecem suporte para desenvolvimento de simuladores de redes sobrepostas com abstrações de *churn*, canais de transmissão, largura de banda, *bootstrapping* e, inclusive, roteadores e *backbones* para um *underlay*, se necessário.

4. Características de Redes Complexas

Este trabalho fundamenta-se em várias métricas de teoria dos grafos e no estudo de propriedades topológicas das redes complexas. Nesta seção são apresentados os conceitos destas áreas que são determinantes para compreensão do texto.

A rede *overlay* que dissemina um canal do SopCast pode ser modelada por um grafo, onde os vértices representam os participantes da transmissão, e as arestas as parcerias entre eles. Ao interpretar que uma parceria abre uma comunicação bilateral considera-se o grafo não direcionado. Dois vértices são chamados vizinhos se há uma aresta que os

⁸<http://www.wireshark.org>

⁹<http://www.omnetpp.org/>

¹⁰<http://www.oversim.org/>

conecte, ou no caso da rede p2p, parceiros. O grau de um vértice é o número de arestas que incidem sobre ele, ou neste caso, o seu número de parceiros. O grau médio é a média aritmética do grau de todos os participantes na rede. Note que este valor é igual a duas vezes o número de arestas, dividido pelo número de vértices, uma vez que cada aresta contribui para o grau de dois vértices. O valor do grau médio é um dos indicativos do nível de conectividade do grafo.

4.1. Métricas de Centralidade

A teoria dos grafos e análise de rede definem várias medidas de centralidade de um vértice em um grafo que permitem determinar sua importância relativa. Três dessas métricas, que são usadas nesse trabalho, são a centralidade de grau, o *closeness* e o *betweenness*. Elas foram criadas para analisar qualquer tipo de grafo, contudo, para se adaptar ao cenário distinto de distribuição foram criadas algumas variações das mesmas. Por exemplo, todos os dados se originam de um único nó, a fonte ou servidor de mídia, e, por isso, foram calculadas duas variações das métricas de *betweenness* e *closeness*, onde somente consideraram-se os caminhos mínimos de todos os vértices para o nó fonte.

A primeira métrica, centralidade de grau, é uma medida relativa ao número de arestas que cada vértice tem. Uma alta centralidade de grau está normalmente associado a uma maior chance de qualquer dado que trafega na rede passar pelo dado vértice. Além disso, a centralidade de grau pode ser calculada para grafos não direcionados e direcionados, onde nesse último caso são considerados o grau de entrada (arestas que chegam no vértice) e grau de saída (arestas que saem do vértice). Já para os experimentos desta pesquisa, onde a rede forma um grafo não direcionado, temos, segundo [Newman 2003], que a métrica pode ser calculada como o grau de cada vértice sobre a quantidade de vértices do grafo menos um:

$$C_{Grau}(v) = \frac{grau(v)}{n - 1}$$

A métrica de *closeness* [Sabidussi 1966], é uma medida topológica de proximidade espacial. Trazendo essa medida para a teoria de grafos ela define quão perto um vértice está de todos os outros através das conexões estabelecidos no grafo e passa a ser descrita como a distância geodésica (caminho mínimo) média entre o vértice v e todos os outros vértices alcançáveis a partir dele ($t \in V \setminus v$).

$$C_{Closeness}(v) = \frac{1}{\sum_{t \in V \setminus v} d_G(v, t)}$$

Outra métrica, baseada no *closeness*, levou em consideração que o único vértice alcançável que importa é a fonte, logo, desconsiderando a variação de t . Enfim, essa medida se tornou somente o inverso do caminho mínimo até o servidor de mídia, portanto, foi considerado somente como a distância.

$$C_{Distancia}(v, fonte) = \frac{1}{d_G(v, fonte)}$$

O *betweenness* [Brandes 2001] é uma medida descritiva de quão interno um vértice é no grafo. O valor dele para um vértice v é a quantidade de caminhos mínimos entre os nós s e t que passam por v . Na fórmula a seguir σ_{st} é o número de caminhos mínimos de s a t , e $\sigma_{st}(v)$ é o número de caminhos mínimos de s a t que passam pelo vértice v :

$$C_{Betweenness}(v) = \sum_{\substack{s \neq v \neq t \in V \\ s \neq t}} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Por fim, essa fórmula sofreu uma alteração para formar uma nova métrica e t passou a ser percebido como um único vértice específico, a fonte de mídia. Assim, foi calculado um *betweenness* especial, partindo de todos os nós somente até o servidor que passem por v , como:

$$C_{B.Fonte}(v, fonte) = \sum_{s \neq v \neq fonte \in V} \frac{\sigma_{s fonte}(v)}{\sigma_{s fonte}}$$

5. Correlações de Upload e Centralidade

O objetivo principal deste trabalho é a identificação de super nós na *overlay* de um sistema p2p de transmissão de mídia ao vivo. Através de métricas de centralidade [Koschützki et al. 2005], como visto na seção 4 e com base na forma com a qual o conteúdo é disseminado através da rede, supõe-se que será possível revelar os super nós dado que a função de tais métricas é determinar a importância relativa de um vértice num grafo (no caso, um nó numa rede).

De cada experimento foram calculados ou observados, para cada nó, as seguintes métricas: taxa de *upload*, grau, *closeness*, distância ao servidor de conteúdo, *betweenness* normal e dos nós à fonte. Essas medidas foram ordenadas para formar *ranks*, ou seja, é associado o número um ao nó mais importante para uma dada métrica, ao segundo mais importante, dois, e assim por diante. Para a taxa de *upload*, grau, *closeness* e ambos os *betweenness*, os nós mais próximos da primeira posição são aqueles com maiores valores para a medida, enquanto que para distância, são os que tem os menores valores. Por exemplo, para distância, o nó mais perto da fonte é possivelmente o mais valioso por que receberá a mídia primeiro, portanto, sua posição no *rank* terá um valor mais próximo de um.

A criação dos *ranks* foi feita de forma a não ter um critério de desempate, se mais de um par tiver o mesmo valor para uma dada métrica a posição deles no *rank* será igual e a posição do próximo nó com valor de métrica diferente será incrementado do número de nós presentes na posição anterior do *rank*. Isso funciona para quase todas as medidas, contudo, distância dos participantes à fonte é uma medida que variou muito pouco, o que implica em muitos nós empatados numa mesma posição do *rank*. Por isso, a fim de melhorar o resultado da medida, foi criada uma métrica baseada na ordenação do par de métricas (distância, grau) tornando o grau o critério de desempate.

Os experimentos se encaixam em três cenários, com cinco repetições de cada: o real, sobre a rede SopCast, e dois simulados, no Oversim, representando redes onde os

	Grau	Close.	Distância	(Dist.,Grau)	Betw.	B.Fonte
Real SopCast	0.8893	0.8869	0.7994	0.8345	0.8319	0.7562
Sim. Heterogêneo	0.6864	0.6925	0.2960	0.5725	0.7204	0.6991
Sim. Homogêneo	0.1683	0.2316	0.7378	0.7213	0.1236	0.9446

Tabela 1. Coeficientes de Correlação de Pearson.

pares tem uma distribuição de largura de banda heterogênea e homogênea. A opção por dois cenários de simulação foi embasada em dois fatos: no desconhecimento da largura de banda dos nós da rede real e nas diferenças comportamentais do protocolo no que diz respeito ao escalonador de *chunks*. Dado que o algoritmo de escalonamento de *chunks* atua de forma a solicitá-los do parceiro mais rápido, se todo nó tem a mesma velocidade, o escalonamento realiza uma seleção aleatória entre as possíveis origens para cada *chunk*, caso contrário é muito provável que o algoritmo selecione sempre os mesmos vizinhos durante as requisições. O objetivo de trazer cenários simulados foi tentar comparar e validar o que foi observado no cenário real.

Foram remontados os grafos da rede a partir dos *logs* de troca de fluxo de mídia, ignorando a parte inicial da coleta a fim de eliminar instabilidades típicas da inicialização da rede. Aqui, as métricas descritas anteriormente foram calculadas a cada 5 segundos sobre um grafo temporalmente dinâmico com base no intervalo entre 100 e 300 segundos dos *logs*. Depois, os *ranks* são obtidos sobre o valor médio de cada métrica em cada nó. Finalmente, para apresentação dos resultados foram calculadas as médias entre os *ranks* das métricas de cada experimento.

A tabela de coeficientes de Pearson (tabela 1) sumariza os resultados e mostra que grau e *closeness* dão *ranks* sistematicamente mais correlatos que as demais métricas. Essa resposta é encontrada também nos experimentos individuais onde não existe uma dominância clara sobre qual das duas é a melhor. Os resultados da simulação heterogênea validam esse resultado, indicando ainda que o *betweenness* é também uma boa métrica, apesar de não ter sido tão interessante para esse conjunto de dados reais. É importante observar que o cenário real e simulado heterogêneo são os que se equivalem e que geram conclusões mais correlatas entre si. Finalmente, o *betweenness* dos nós somente para a fonte emergiu como uma excelente métrica num cenário levemente diferente onde a rede tem uma característica homogênea de pares.

A primeira medida analisada foi centralidade de grau, representada pela figura 1. A imagem 1(a) é uma das que apresentam melhores correlações visuais entre as métricas para os dados reais. A tabela 1 também confirma esse dado numericamente através do coeficiente de correlação de Pearson entre taxa de *upload* e grau. Além disso, os experimentos reais indicam individualmente correlações moderadas no aspecto analisado enquanto a média destaca uma correlação bem forte. É possível observar que a dispersão dos pontos dos experimentos reais no início do *rank* (até 100) tem menos precisão do que o resto dos pontos e baixa exatidão comparado com o resultado esperado, o que implica uma deficiência da métrica em identificar super nós. Apesar disso, em 80% dos experimentos o nó de maior taxa de *upload* teve também o maior grau. Os dados simulados da rede heterogênea aparentam ter dois crescimentos diferentes por que alguns experimentos tiveram pouca variação da taxa de *upload* e muitos nós (+ de 50%) não fizeram *upload*

portanto saturando o *rank*. No entanto, a correlação no início da curva (até 50) é visualmente linear. Este comportamento é melhor explicado pelo modelo de disseminação em árvore, um nó que possui mais parceiros tem grandes chances de emergir como um super nó, além disso, pode ser inferido que a forma de disseminação supera os modelos de organização da *overlay* e da *underlay*. Finalmente, na distribuição de *ranks* de graus da simulação homogênea da figura 1(c) observamos uma forma de sino indicando uma correlação quase inexistente entre as métricas para esse caso.

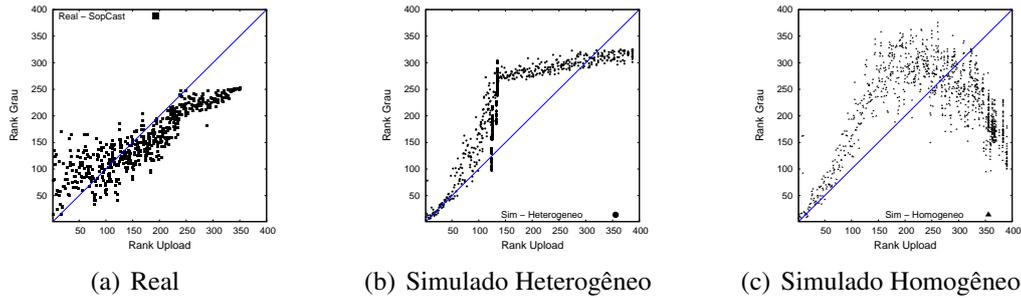


Figura 1. Correlação entre ranks de Grau e Upload.

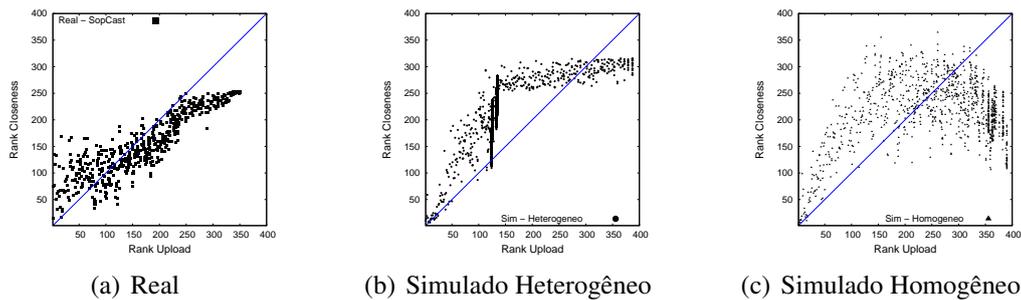


Figura 2. Correlação entre ranks de Closeness e Upload.

O estudo de *closeness* (figura 2) obteve resultados similares ao de grau inclusive no que diz respeito a análises mais finas como: o par de maior *closeness* ser o de maior taxa de *upload*. Numericamente é possível observar na tabela de coeficientes que essa métrica é tão boa quanto a da centralidade de grau.

A terceira métrica, distância, presente na figura 3, apresentou a pior representação visual apesar de posar numericamente como uma correlação pouco mais fraca que as últimas para o caso real. Individualmente nos experimentos reais essa métrica é sistematicamente ruim comparada com as melhores métricas, sendo isso validado através da simulação heterogênea. No gráfico 3(a) a distribuição real teve poucos patamares por conta da média dos experimentos, contudo, geralmente esta métrica teve pouca variação, assim, aglomerando a análise em níveis. De toda forma, isso permitiu deduzir que a árvore de disseminação tenta se manter curta mas com carga baixa na raiz, uma vez que, para distância do nó-fonte > 1 , *ranks* baixos (menores distâncias) concentram mais pontos. Ademais, distância não aparenta ser uma boa métrica neste cenário onde a rede não permanece estática com o tempo e, apesar de numericamente a simulação homogênea encontrar um bom resultado, é possível observar graficamente que é impraticável identificar super nós através dessa métrica.

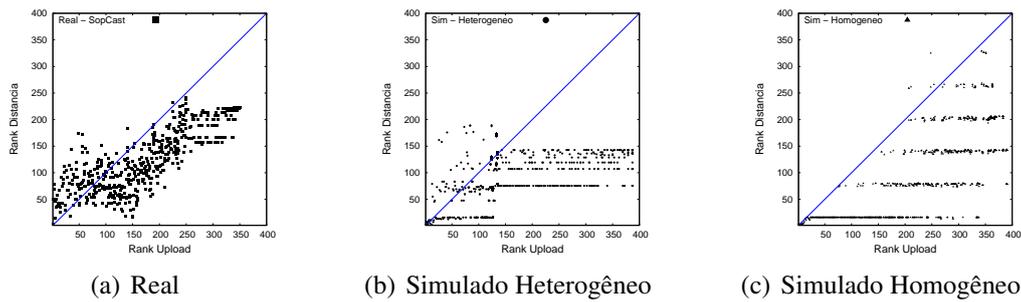


Figura 3. Correlação entre ranks de Distância e Upload.

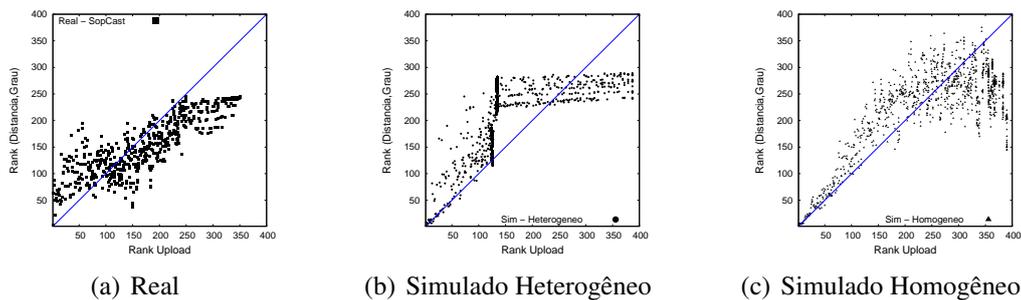


Figura 4. Correlação entre ranks de (Distância, Grau) e Upload.

Na figura 4, é mostrado a métrica que usa o par (distância, grau) ordenado para formar o *rank*. Essa medida foi capaz de melhorar a variação de valores do eixo Y, especialmente sobre a média dos experimentos, mas o resultado geral não melhorou muito. Visualmente alguns dos níveis observados na figura 3 são ainda um pouco presentes na figura 4 e muito presente nos experimentos individuais. Numericamente a métrica composta alcançou um resultado intermediário comparado com seus componentes.

O gráfico do *rank* de *betweenness* pode ser visto na figura 5. A análise é novamente similar à das métricas de *closeness* e grau, trazendo poucas diferenças práticas. Todavia, de acordo com a tabela 1 essa métrica é menos efetiva que as citadas para o caso real apesar de ser até melhor que as demais de acordo com a simulação heterogênea.

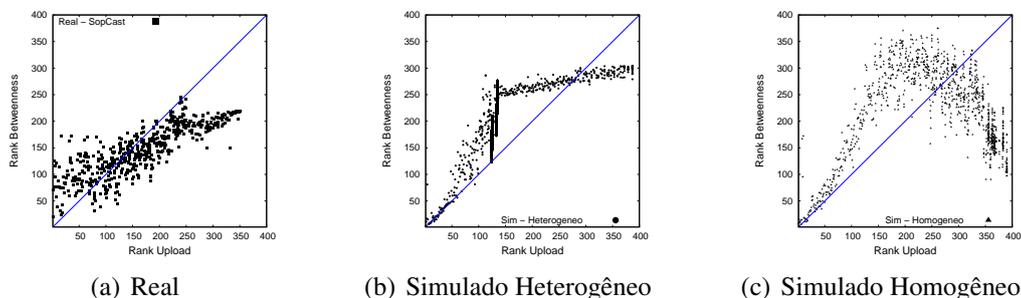


Figura 5. Correlação entre ranks de Betweenness e Upload.

Finalmente, a última métrica analisada foi o *betweenness* de todos os nós somente até a fonte que deu resultados bem diferentes para os três cenários, como mostra a figura 6 e a tabela 1. Essa exceção era esperada justamente por conta da diferença comportamental

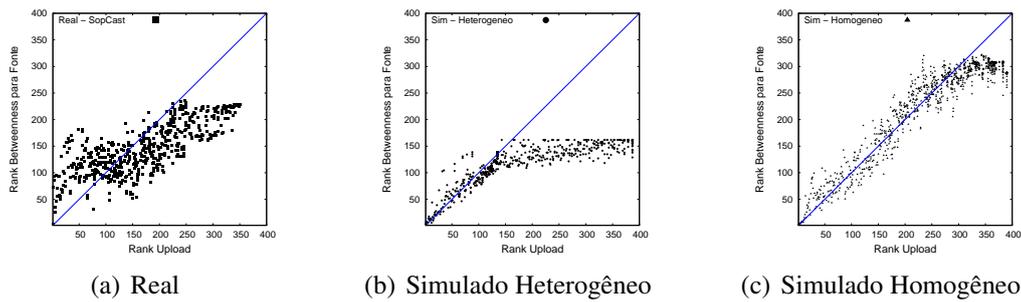


Figura 6. Correlação entre ranks de Betweenness para Fonte e Upload.

do escalonador. Nos cenários heterogêneos é mais provável que os nós mais distantes do servidor façam nenhum ou pouco envio e que sejam aqueles que tem os menores valores para essa métrica. Além disso, os caminhos de distribuição de dados são mais fixos que no homogêneo. A rede analisada é remontada a partir dos *logs* de distribuição, e nos cenários heterogêneos a métrica tem menor variabilidade já que ao longo do tempo o grafo sobre o qual se faz as análises muda pouco. Ou seja, como podemos verificar através da figura 6(c), essa métrica criada é muito boa para identificar o *rank* de *upload* dos nós num cenário de seleção de *chunks* de pares aleatório, mas não para o cenário real que é heterogêneo e o escalonamento leva banda em consideração.

6. Conclusões e Trabalhos Futuros

O trabalho conclui que tanto a centralidade de grau quanto o *closeness* oferecem as melhores correlações com taxa de *upload* entre todas as métricas de centralidade testadas (grau, *closeness*, distância, o par (distância, grau), *betweenness* e *betweenness* dos nós para a fonte). No cenário real, apesar do coeficiente de correlação dessas métricas serem altos para a média dos experimentos, individualmente eles são moderados, próximos de 0.50. Através do conjunto de dados da simulação com pares heterogêneos foi possível validar os resultados obtidos e induzir ao entendimento de que o modelo de disseminação supera tanto o modelo de *overlay* quanto o de *underlay*, ou seja, o algoritmo de escalonamento de *chunks* influencia mais na correlação do que a organização das redes sobreposta e física.

Através do *betweenness* dos nós para a fonte, uma modificação da métrica de *betweenness* original, foi possível encontrar a melhor correlação num ambiente de pares homogêneos, ou no qual o algoritmo de escalonamento escolha parceiros aleatórios que, minimamente, tenham o *chunk* desejado. Esse resultado, aliado às outras diferenças entre os coeficientes de correlação de Pearson para os cenários simulados, indica que a largura de banda dos pares influencia significativamente os resultados. Isso motiva novos experimentos e testes, incluindo o *rank* de larguras de banda nas medidas e outras variações como o par (largura de banda, grau) ou (largura de banda, *closeness*).

Para os trabalhos futuros planeja-se uma identificação mais sistemática de super nós apoiada por uma combinação destes resultados. Além disso, espera-se adquirir informação sobre a distribuição da largura de banda entre os nós do Planetlab. Com isso será possível comparar cenários reais e simulados através da nova métrica proposta anteriormente de forma que leve a uma conclusão definitiva sobre o comportamento emergente de super nós.

Referências

- Ali, S., Mathur, A., and Zhang, H. (2006). Measurement of commercial peer-to-peer live video streaming. In *Proc. of Workshop in Recent Advances in Peer-to-Peer Streaming*.
- Baumgart, I., Heep, B., and Krause, S. (2007). OverSim: A Flexible Overlay Network Simulation Framework. In *Proceedings of 10th IEEE Global Internet Symposium (GI '07) in conjunction with IEEE INFOCOM 2007, Anchorage, AK, USA*, pages 79–84.
- Bellovin, S. M. (2002). A technique for counting natted hosts. In *IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement*, pages 267–272, New York, NY, USA. ACM.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177.
- Cohen, B. (2003). Incentives build robustness in bittorrent.
- Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the internet topology. In *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 251–262, New York, NY, USA. ACM.
- Freeman, L. (1979). Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239.
- Hei, X., Liang, C., Liang, J., Liu, Y., and Ross, K. (2007). A Measurement Study of a Large-Scale P2P IPTV System. *Multimedia, IEEE Transactions on*, 9(8):1672–1687.
- Koschützki, D., Lehmann, K. A., Peeters, L., Richter, S., PODEHL, D. T., and Zlotowski, O. (2005). *Centrality Indices*, volume 3418/2005 of *Lecture Notes in Computer Science: Network Analysis*, chapter Part I: Elements, pages 16–61. Springer Berlin / Heidelberg.
- Kumar, R., Novak, J., and Tomkins, A. (2006). Structure and evolution of online social networks. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 611–617, New York, NY, USA. ACM.
- Leibowitz, N., Ripeanu, M., and Wierzbicki, A. (2003). Deconstructing the kaza network. In *WIAPP '03: Proceedings of the The Third IEEE Workshop on Internet Applications*, page 112, Washington, DC, USA. IEEE Computer Society.
- Newman, M. (2003). The structure and function of complex networks. *Arxiv preprint cond-mat/0303516*.
- Oliveira, J., Vieira, A., and Campos, S. (2009). Poluição de conteúdo em sistemas p2p live streaming. In *Simpósio Brasileiro de Sistemas Multimídia e Web - Webmedia*.
- Ripeanu, M., Iamnitchi, A., and Foster, I. (2002). Mapping the gnutella network. *IEEE Internet Computing*, 6(1):50–57.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603.
- Santos, R., Rocha, B., Rezende, R., and Loureiro, A. (2009). Characterizing the YouTube video-sharing community. <http://security1.win.tue.nl/bpontes/pdf/yt.pdf>, 12.

- Schollmeier, R. (2001). A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. *Peer-to-Peer Computing, IEEE International Conference on*, 0:0101.
- Silverston, T. and Fourmaux, O. (2007). Measuring P2P IPTV Systems. In *Proc. of ACM NOSSDAV*.
- Tran, D., Hua, K., and Do, T. (2004). A peer-to-peer architecture for media streaming. *Selected Areas in Communications, IEEE Journal on*, 22(1):121–133.
- Tutschku, K. (2004). A measurement-based traffic profile of the eDonkey filesharing service. *Lecture notes in computer science*, pages 12–21.
- Watts, D. J. (2004). *Six Degrees: The Science of a Connected Age (Open Market Edition)*. W.W. Norton & Co.
- Wu, C., Li, B., and Zhao, S. (2008). Exploring large-scale peer-to-peer live streaming topologies. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 4(3):1–23.
- Zhang, X., Liu, J., Li, B., and Yum, T. (2005). CoolStreaming/DONet: A data-driven overlay network for efficient live media streaming. In *proceedings of IEEE Infocom*, volume 3, pages 13–17. Citeseer.