

Impacto das Relações Sociais em Sistemas de Compartilhamento de Arquivos*

Thiago Amaral Guarnieri¹, Alex Borges Vieira¹, Ana Paula Couto da Silva²

¹DCC - Universidade Federal de Juiz de Fora – ²DCC - Universidade Federal de Minas Gerais

thiago.guarnieri@ice.ufjf.br, alex.borges@ufjf.edu.br

ana.coutosilva@dcc.ufmg.br

Resumo. Neste artigo são apresentados os resultados preliminares que visam discutir o impacto das redes sociais, particularmente o Facebook, na melhoria da disseminação de dados em redes de compartilhamento de arquivos. São comparados swarms que tiveram seus anúncios pela rede social e outros via site de indexação e uma discussão inicial do impacto no tráfego Inter-AS é realizada. Resultados completos da caracterização preliminar apresentada poderão ser utilizados na definição de novos mecanismos de incentivo que considerem relações sociais e localidade geográfica.

1. Introdução

A adoção do paradigma de comunicação *peer-to-peer* (P2P) tem crescido massivamente desde a década passada. O tráfego gerado por aplicações que usam esse paradigma é dominante na maior parte das redes de computadores. Estima-se que no ano de 2012, mais de 36% de todo volume de dados de *upload* na América do Norte foi gerado pelo Bittorrent [Cohen 2003], uma das principais aplicações. O volume total gerado pelo Bittorrent supera 10% de todo o tráfego de pico na América do Norte¹.

A larga adoção do paradigma P2P está intimamente relacionada a duas características principais: resiliência e escalabilidade. A maior resiliência se deve ao fato da arquitetura não ter somente um ponto de falha central. A maior escalabilidade é alcançada dado que todas as entidades que participam do sistema P2P podem contribuir com uma parcela de seus recursos para atingir um fim comum. Por exemplo, os *peers* de uma aplicação de compartilhamento de arquivos pode contribuir enviando trechos do arquivo compartilhado a seus parceiros. Assim, a disseminação dos arquivos compartilhados pode ser mais rápida e menos custosa do que se dependesse de uma entidade centralizadora.

Desta forma, o sucesso para uma disseminação de conteúdo em sistemas P2P está na colaboração voluntária dos diversos *peers* que compõem a rede. No entanto, a falta de cooperação em tais aplicações impõe um sério desafio para a qualidade do serviço ofertado. Por exemplo, em aplicações de compartilhamento de arquivos, os participantes podem ter o tempo de descarga do arquivo penalizado caso exista poucos *peers* cooperando. Em um pior caso, o arquivo compartilhado pode não estar completo na rede P2P, o que impossibilitará que algum *peer* o tenha por completo.

Há vários mecanismos que tentam incentivar a cooperação dos *peers* de aplicações P2P. Alguns são baseados em reputação dos usuários. Usuários que obtêm boa reputação

*Este trabalho foi financiado pelo CNPq, CAPES, FAPEMIG.

¹Global Internet Phenomena Report: 2H 2012, www.sandvine.com

no sistema são priorizados na entrega de recursos. No entanto, tais mecanismos muitas vezes precisam de um coordenador central ou são facilmente burlados. No outro extremo, o protocolo *Bittorrent* apresenta um mecanismo descentralizado, conhecido como *tit-for-tat*. Os *peers* do sistema se monitoram mutuamente e a quantidade de recursos (i.e. pedaços de arquivo) recebida por um *peer* é proporcional a sua contribuição.

O bom funcionamento do mecanismo de incentivo *tit-for-tat* está intimamente ligado a contribuição de curto prazo entre *peers* na rede P2P. Tal mecanismo não considera incentivo, favorecendo a formação de parcerias que considerem a localização geográfica ou até mesmo a relação de cooperação de longa duração. Considerar esses dois fatores, poderia resultar em uma melhor eficiência de compartilhamento na rede P2P [Andrade et al. 2012].

Por exemplo, no que tange a localização geográfica, a escolha aleatória de parceiros pode gerar um grande volume de tráfego entre diferentes Sistemas Autônomos (AS). Mesmo que existam vários *peers* de um mesmo AS na rede P2P, nada garante que haverá preferência pela interação entre eles (*peers* do mesmo AS). De certa forma, o custo de tráfego inter-AS onera os provedores de serviço de Internet e geram uma discriminação do tráfego P2P. Mais ainda, a utilização de mecanismos de incentivos de longa duração entre *peers* torna-se importante e necessária. Uma vez que uma boa parceria se estabeleça, ela deve ser mantida por um tempo suficiente para favorecer a rápida disseminação de conteúdo (no caso de compartilhamento de arquivos).

Nessa linha, alguns trabalhos mostram a viabilidade do uso de relações sociais dos participantes de uma aplicação P2P para melhorar tanto a localização geográfica do tráfego P2P, quanto fomentar parcerias de longa duração. As relações de similaridade sociais, de redes como *Facebook*² e o *Twitter*³, podem ser utilizadas para tornar mais rápida e segura a disseminação de conteúdo na rede P2P. Há uma confiança e agrupamento inerente aos nós de uma rede social que pode ser explorada, por exemplo, para evitar a incidência de *peers* pouco cooperativos.

De fato, a disseminação de arquivos pelo protocolo *Bittorrent* é melhorada quando se utiliza uma rede social para anunciar o *torrent* associado ao arquivo (*torrent* é o ponto de entrada na rede P2P para obter o arquivo desejado). Por exemplo, em [Wang et al. 2011, Wang et al. 2012], os autores mostram que em *torrents* anunciados por redes sociais, particularmente o *twitter*, há um número considerável de *peers* com comportamento estável. Estes *peers* ingressam na rede P2P (ou *swarm* de disseminação do arquivo) seguindo padrões temporais previsíveis e disseminam conteúdo por um tempo mais longo que o padrão comum. Por participarem por mais tempo da rede P2P, a capacidade de disseminação do conteúdo é melhorada.

Mais ainda, localização de *peers*, mesmo que em um *swarm* P2P em particular, pode ser importante para a melhoria global do sistema. Por exemplo, em [Wang and Liu 2012], os autores mostram que 85% dos *peers* participam de múltiplos *swarms* e que essa característica pode ser utilizada para aumentar a localidade de toda a rede P2P e reduzir tráfego inter-AS. Para tal, é preciso realizar o rastreamento da migração de um *peer* entre os diferentes *trackers* (centralizadores da rede Bittorrent). Como alter-

²www.facebook.com

³www.twitter.com

nativa, pode-se realizar um rastreamento dos peers nas respectivas redes sociais a que eles pertencem. Assim, há uma possibilidade de criar grupos de *peers* que mantenham o tráfego P2P restrito a seus ASs.

A noção de tempo de permanência de longa duração pode ser estendida a grupos de *peers* [Pietiläinen and Diot 2012]. Neste caso, os grupos são definidos como comunidades temporais. O eventual contato oportunista dos *peers* pode ser explorado para melhorar a disseminação de informação na rede P2P. No caso do trabalho citado, foram exploradas as interações sociais entre pessoas em um grande evento. Porém, outros cenários podem ser explorados, como a interação de usuários de uma grande rede social.

Neste artigo são apresentados os resultados preliminares que visam discutir o impacto das redes sociais, particularmente o *Facebook*, na melhoria da disseminação de dados em redes de compartilhamento de arquivos. São comparados *swarms* que tiveram seus anúncios pela rede social e outros que não o tiveram. Nesse sentido é feita uma discussão inicial do impacto no tráfego Inter-AS. Mais ainda, são apresentadas caracterizações preliminares quanto a localização dos *peers* destas redes P2P. Resultados completos da caracterização preliminar apresentada neste artigo poderão ser utilizados na definição de novos mecanismos de incentivo que considerem relações sociais e localidade geográfica.

2. Disseminação de *Torrents* em Redes Sociais

2.1. Metodologia de Coleta de Dados

Para analisar o papel das redes sociais na disseminação de arquivos *torrent* e a correlação com a localização geográfica dos *peers*, foram definidos e coletados dois tipos de *torrents*: (i) *torrents* de interesse regional; (ii) *torrents* de interesse global. Os *torrents* de interesse local são aqueles cujo conteúdo tem apelo à uma parcela específica de pessoas. Como exemplo, podem ser considerados filmes produzidos e realizados por *Bollywood*, onde a maior parte de público tende a se concentrar na Ásia, em regiões próximas à Índia. *Torrents* de interesse global são aqueles cujo conteúdo tendem a não ter um público regionalmente específico, sendo estes consumidos mundialmente. Nesta categoria estão, por exemplo, os filmes produzidos por *Hollywood*.

Para os resultados preliminares apresentados, foram considerados dois conteúdos distintos: um *torrent* com conteúdo regional indiano; e um *torrent* de conteúdo global. Para cada conteúdo foram coletados dados dos respectivos *swarms* anunciados via uma comunidade indiana no Facebook (*swarm social*) e via site de indexação (*swarm não-social*). Essa metodologia tem como objetivo analisar a correlação entre relações sociais, localização geográfica e tráfego inter-ASs. Para mapear o endereço IP dos *peers* em suas respectivas localizações geográficas, foi utilizada a base de dados *GeoLite*⁴.

2.2. Resultados Preliminares e Análise

A análise preliminar dos *swarms* descritos na Seção 2.1 mostra indícios da correlação entre relações sociais e localidade geográfica. Mais precisamente, *torrents* disseminados via redes sociais tendem a agrupar *peers* que estão mais próximos geograficamente (devido a própria natureza das relações interpessoais), além de clusterizar *peers* que tendem a ter um maior nível de confiança mútua.

⁴Maxmind GeoLite, www.dev.maxmind.com/geoip/geolite/

A Figura 1 mostra a porcentagem dos *peers* localizados em diferentes países, para o *torrent* com conteúdo de interesse regional. Considerando o caso da disseminação via rede social, o total de *peers* localizados no entorno da Índia⁵ é um pouco maior (59% contra 56%) quando comparado a disseminação via site de indexação. Como é analisado um conteúdo de apelo regional, é esperado que a localização geográfica dos *peers* não sofra mudanças significativas entre os dois paradigmas de distribuição do conteúdo.

A Figura 2, por sua vez, mostra a porcentagem dos *peers* localizados em diferentes países, considerando o *torrent* que dissemina conteúdo de interesse global. A disseminação via rede social, neste caso, está fortemente correlacionada com a localização geográfica dos *peers*: no *torrent* social, a parcela de *peers* localizados em regiões próximas à Índia é 3 vezes maior (24% contra 8%) do que na disseminação via site.

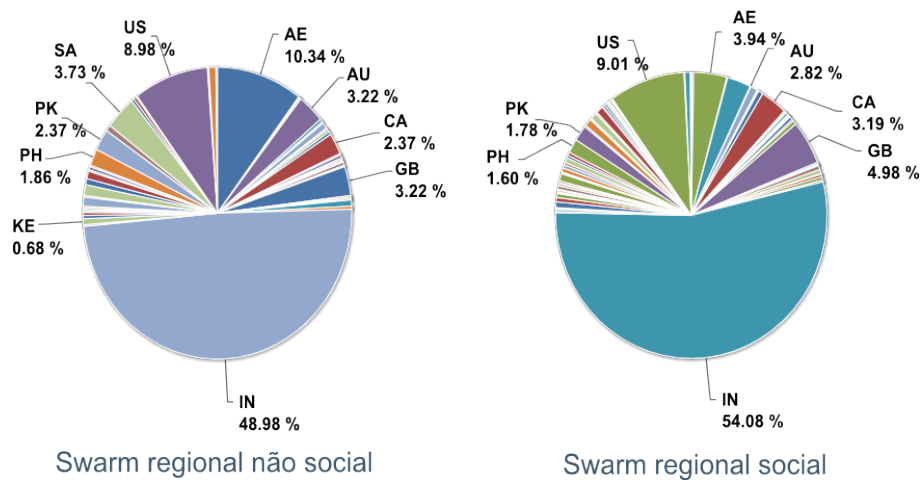


Figura 1. Distribuição de peers em torrent de interesse regional

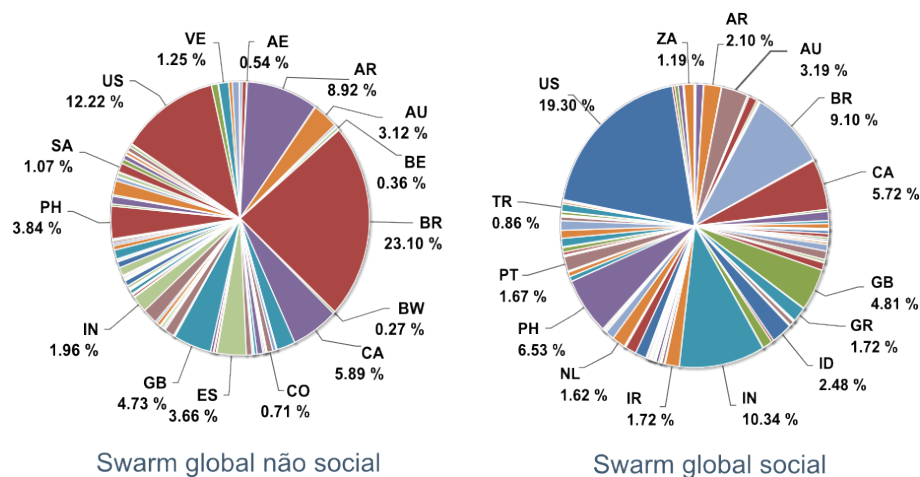


Figura 2. Distribuição de peers em torrent de interesse global

Também foi analisada a distribuição dos *peers* pelos Sistemas Autônomos (Tabelas 1 e 2). No caso do conteúdo regional, a disseminação social obteve 25,6% do número

⁵Paquistão, Índia, China, Irã, Tailândia, Indonésia, Filipinas, Vietnã e Malásia.

de *peers* nos 10 maiores AS contra 29,8% na disseminação não-social. No entanto, como é considerado uma comunidade indiana na rede social, temos efetivamente 24.4% de *peers* em ASs da região no *swarm social*, contra 20.7% do *swarm não-social*. Já no caso do conteúdo global a proporção fica em 7,89% para a disseminação social contra 7,37% para a disseminação não-social. Novamente, como é considerada uma comunidade indiana no Facebook, 5.26% dos *peers* estão localizados próximos (5 ASs) no *swarm social* contra 1.6% no *swarm não-social*.

Swarm social				Swarm não-social			
total	%	Id AS	País	total	%	Id AS	País
147	13,8%	75741	India	74	12,5%	75741	India
28	2,6%	76780	India	18	3%	63986	United Arab Emirates
21	1,9%	17921	India	17	2,8%	498	United Arab Emirates
15	1,4%	77677	India	16	2,7%	76780	India
13	1,2%	63986	United Arab Emirates	15	2,5%	17921	India
11	1,0%	77690	India	9	1,5%	101639	India
11	1,0%	101639	India	9	1,5%	101554	Pakistan
10	0,9%	74893	India	9	1,5%	53316	United Arab Emirates
10	0,9%	17728	India	6	1%	49906	United Kingdom
10	0,9%	101554	Pakistan	5	0,8%	188581	United Arab Emirates
25.6% do total de <i>peers</i> (24.4% no entorno)				29.8% do total de <i>peers</i> (20.7% no entorno)			

Tabela 1. Distribuição de *peers* por AS em conteúdo regional

Swarm social				Swarm não-social			
total	%	Id AS	País	total	%	Id AS	País
65	3,0%	75741	India	11	0,98%	108246	Argentina
20	0,9%	69144	United States	10	0,89%	75741	India
14	0,66%	77074	Philippines	9	0,80%	98420	Brazil
13	0,61%	73824	Philippines	9	0,80%	107349	Argentina
11	0,52%	73822	Philippines	8	0,71%	77074	Philippines
11	0,52%	53669	United Kingdom	8	0,71%	104825	Brazil
10	0,47%	97400	Malaysia	7	0,62%	99150	Brazil
9	0,42%	36309	United States	7	0,62%	103832	Uruguay
9	0,42%	68353	Canada	7	0,62%	108195	Argentina
8	0,38%	53112	Portugal	7	0,62%	50612	Spain
7,89% do total de <i>peers</i> (5.26% no entorno)				7,37% do total de <i>peers</i> (1,6% no entorno)			

Tabela 2. Distribuição de *peers* por AS em conteúdo global

Estes resultados fornecem indícios de que explorar relações sociais podem trazer melhoria no desempenho de sistemas de disseminação de arquivos, principalmente no que tange a questão da localização geográfica. Para a continuidade do trabalho, já foram coletados ≈ 30.700 arquivos *torrents*, entre janeiro e março de 2013. Informações como metadados também foram recuperadas. Estes dados serão a base para o estudo do impacto das relações sociais em sistemas de compartilhamento de arquivos.

3. Discussões

Considerando a metodologia de coleta de dados, os próximos passos são:

1. Navegação pelo grafo social do *facebook*, buscando grupos de divulgação de *torrents* e obtendo diferentes arquivos de metadados. Estes arquivos serão comparados aos que foram coletados via o site de indexação *www.btmon.com*. O objetivo é quantificar o nível de inserção de arquivos *torrents* no *Facebook*.
2. Para definir o nível de clusterização dos *peers*, baseada em localização geográfica, o grafo de conectividade do *swarm*, clientes *BitTorrent* serão instalados no *PlanetLab*⁶. Esses clientes interagirão com *swarms* reais coletando as informações como endereços IP e mensagens com a lista de vizinhos de um *peer*.
3. Para cada *torrent* serão feitas coletas sucessivas, em intervalos pré-determinados de tempo. Assim, poderão ser definidos nós ou grupos de nós que permaneçam no *swarm* um longo período de tempo, contribuindo com um grande número de nós. O objetivo é verificar a existência e caracterizar as comunidades temporais. Estas relações sociais poderão ser exploradas para propostas de mecanismos de incentivo mais eficientes.

Ao final da caracterização, os seguintes resultados são esperados:

1. **Cooperação entre *peers*:** Relações sociais entre os *peers* podem auxiliar na definição de novos mecanismos de incentivo, que desencorajem *peers* não-cooperativos. A caracterização a ser realizada irá verificar se a incidência de *peers* não-cooperativos é a mesma quando *swarms sociais* são considerados.
2. **Localização Geográfica e Contexto Social:** Analisar o impacto, em diferentes métricas de desempenho, da utilização de um mecanismo *tit-for-tat* modificado, considerando *peers* pertencentes ao mesmo AS, com características sociais similares ou relações de longa duração.

Este artigo apresenta resultados iniciais do impacto da disseminação de *torrents* em redes sociais, tendo como particular enfoque o Facebook. Estudos preliminares mostram a existência do agrupamento de nós, baseados em relação social ou localização geográfica. Estas características podem ser utilizadas para melhorar o compartilhamento de conteúdo.

Referências

- Andrade, N., Mowbray, M., Lima, A., Wagner, G., and Ripeanu, M. (2012). Influences on cooperation in bittorrent communities. *ACM SIGCOMM Workshop on Economics of peer-to-peer systems*, pages 111–115.
- Cohen, B. (2003). Incentives build robustness in bittorrent. *Proceedings of 1st Workshop on Economics of Peer-to-Peer Systems (P2PECON)*.
- Pietiläinen, A. and Diot, C. (2012). Dissemination in opportunistic social networks: the role of temporal communities. *Proceedings of the thirteenth ACM international symposium on Mobile Ad Hoc Networking and Computing*, pages 165–174.
- Wang, F., Liu, J., Xu, K. ., and Wu, D. (2012). Torrents on twitter: Explore long-term social relationships in peer-to-peer systems. *IEEE Transactions on Network and Service Management*, pages 1–10.
- Wang, H. and Liu, J. (2012). Exploring peer-to-peer locality in multiple torrent environment. *IEEE Transactions on Parallel and Distributed Systems*, pages 1216–1226.
- Wang, H., Wang, F., and Liu, J. (2011). On long-term social relationships in peer-to-peer systems. *IEEE 19th International Workshop on Quality of Service*, pages 1–8.

⁶www.planet-lab.org