

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Coordenação de Pós-Graduação em Engenharia Elétrica

Distribuição de Conteúdo Entre-pares em Perspectiva: Um  
Estudo sobre Carga, Compartilhamento e Qualidade de Serviço  
em Comunidades BitTorrent

Nazareno Andrade

Tese de Doutorado submetida à Coordenação do Curso de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências no domínio da Engenharia Elétrica.

Área de Concentração: Processamento da Informação

Francisco Brasileiro

(Orientador)

Walfredo Cirne

(Orientador)

Campina Grande, Paraíba, Brasil

© Nazareno Andrade, 01/10/2008

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

A553d

2008 Andrade, Nazareno.

Distribuição de conteúdo entre-pares em perspectiva: um estudo sobre carga, compartimento e qualidade de serviço em comunidades BitTorrent / Nazareno Andrade. — Campina Grande, 2008.

91 f. : il.

Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Referências.

Orientadores: Prof. Dr. Francisco Brasileiro, Prof. Dr. Walfredo Cirne

1. Sistemas entre-pares. 2. Distribuição de Conteúdo. 3. Compartilhamento. 4. BitTorrent. I. Título.

CDU-621.3:004.72(043



**DISTRIBUIÇÃO DE CONTEÚDO ENTRE-PARES EM PERSPECTIVA: UM ESTUDO  
DA CARGA, COMPARTILHAMENTO E QUALIDADE DE SERVIÇO EM  
COMUNIDADES BITTORRENT**

**NAZARENO FERREIRA DE ANDRADE**

Tese Aprovada em 27.10.2008



**FRANCISCO VILAR BRASILEIRO, Ph.D., UFCG**  
Orientador

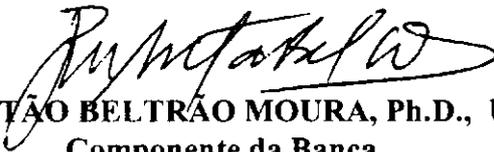
**WALFREDO DA COSTA CIRNE FILHO, Ph.D., UFCG**  
Orientador (Ausência Justificada)



**DORGIVAL OLAVO GUEDES NETO, Dr., UFMG**  
Componente da Banca



**ANTONIO MARINHO BARCELLOS, Dr., UFRGS**  
Componente da Banca



**JOSÉ ANTÃO BELTRÃO MOURA, Ph.D., UFCG**  
Componente da Banca



**DALTON DARIO SEREY GUERRERO, D.Sc., UFCG**  
Componente da Banca

**CAMPINA GRANDE – PB**  
**OUTUBRO - 2008**

## Resumo

A maior enciclopédia do mundo, a Wikipédia, uma infra-estrutura global de telefonia, o Skype, e uma das maiores plataformas de computação paralela do mundo, o SETI@home, são hoje baseados no compartilhamento de recursos possuídos por aqueles que, tradicionalmente, eram apenas clientes desses tipos de serviços. Sistemas baseados em compartilhamento foram projetados para a posição de agentes importantes de nossa sociedade pela prevalência de computadores pessoais e da Internet. Entender esses sistemas e, a partir deste entendimento, melhor projetá-los, é, portanto, fundamental para que realizemos o potencial desta sociedade.

Entender o funcionamento de um sistema de compartilhamento, contudo, é uma tarefa complexa. Além de conhecer seu projeto e implementação, é preciso analisar também a plataforma em que esse sistema executa e a carga a que ele é submetido. Com frequência, essa plataforma e carga dependem da ação coletiva de um grande número de usuários, tornando seu estudo não-trivial.

Neste trabalho, examinamos esses dois aspectos em uma instância amplamente popular de sistemas baseados no compartilhamento: a distribuição de conteúdo entre-pares através do BitTorrent. Nossa abordagem consiste da coleta e análise de dados sobre o funcionamento de diversas comunidades que utilizam o BitTorrent para distribuir conteúdo, a fim de entender o impacto do comportamento de seus usuários no desempenho do sistema.

Nossos principais resultados são (i) a constatação de que a distribuição de popularidade de conteúdo em comunidades BitTorrent não tem cauda longa, o que altera a lógica do dimensionamento de *caches* para o BitTorrent em comparação com a Web e outros sistemas entre-pares, (ii) um modelo mais preciso que o estado-da-arte para o decaimento do interesse da comunidade por um conteúdo com o tempo, o que contribui para avaliações de desempenho mais precisas do BitTorrent e de mecanismos semelhantes, (iii) a análise dos níveis de contribuição dos usuários em diversas comunidades que usam o BitTorrent, que mostra que usuários contribuem mais nessas comunidades que em outros sistemas entre-pares e que os usuários de uma comunidade contribuem mais na presença de mecanismos de incentivo extras ou quando a comunidade compartilha conteúdo estritamente legal e é socialmente coesas, (iv) a observação de que comunidades BitTorrent tipicamente provêm corretamente

virtualmente todas as requisições que recebem, e (v) a análise da eficácia de mecanismos de sanções para o incentivo à contribuição em algumas comunidades que usam o BitTorrent, que mostra que embora esses mecanismos resultem em comunidades equânimes e com mais contribuição, eles não implicam em uma maior proporção de requisições servidas ou em uma mudança notável na contenção por recursos em relação às comunidades coesas que compartilham apenas conteúdo legal.

## Abstract

The world's largest encyclopedia, Wikipedia, a global telephony infrastructure, Skype, and one of the largest parallel computing platforms in the world, SETI@home, currently operate based on resources shared by those which, traditionally, were only clients of similar services. Sharing-based systems were propelled by the prevalence of personal computers and Internet into major agents in our society. Understanding these systems and, from this understanding, better design them, is therefore paramount for attaining the potential of this society.

Understanding the behavior of a sharing-based system, however, is a complex enterprise. Besides understanding its design and implementation, it is also necessary to analyze the platform on which it runs and the workload it serves. Often, both this platform and workload result from the collective action of a large number of users, making their study non-trivial.

In this work we examine these two aspects in a largely popular instance of sharing-based systems: peer-to-peer content distribution through BitTorrent. Our approach consists of collecting and analyzing data about user behavior and system performance in multiple communities that rely on BitTorrent for content distribution, aiming to understand the impact of user behavior on the system's performance.

Our main results are (i) the understanding that the distribution of content popularity in BitTorrent communities is not long-tailed, what changes the trade-off in cache provisioning for BitTorrent when compared to cache provisioning for the Web and for other peer-to-peer systems, (ii) a model for the decaying of the community's interest for content items over time that is more precise than the state-of-the-art and contributes to more precise performance evaluations of BitTorrent and similar systems, (iii) the analysis of user contribution levels in several communities that use BitTorrent, which shows that users contribute more in these communities than in other peer-to-peer systems and that users in a community contribute more in the presence of extra incentive mechanisms or when the community shares only strictly legal content and is socially cohesive, (iv) the conclusion that BitTorrent communities typically serve correctly virtually all requests received, and (v) the analysis of the efficacy of sharing enforcement mechanisms in promoting contribution in some BitTorrent communities, which reveals that although these mechanisms result in equitable communities with higher levels of contribution, they do not lead to a larger proportion of requests served or

to a noticeable change in the level of resource contention of the community when compared to socially cohesive communities that share legal content.

## Agradecimentos

Uma das melhores sensações ao fim desses quatro anos é olhar para trás e perceber a quantidade de pessoas a quem devo agradecimentos. Ao longo desse tempo, tive a sorte de aprender sobre computação, pesquisa, ciência, boêmia e a vida em geral com muitas pessoas. Quaisquer méritos dessa tese são, ao menos em parte, delas.

Em primeiro lugar, tanto agradecimento quanto é possível a Fubica e Walfredo, orientadores, exemplos, colegas e amigos ao longo dos últimos seis anos. As incontáveis vezes que Fubica me pôs os pés no chão, me mostrou que é preciso também lidar com o dróps de anis, foi camarada ou me motivou a levar projetos até o fim me ensinaram sobre o que é ser um pesquisador. Walfredo me contagiou com seu poço borbulhante de energia (para parafrasear o próprio) desde que começamos a trabalhar juntos e manteve sempre claro que pesquisa deve ser prazer e que podemos fazer dela o bicho cúbico.

Além de Walfredo e Fubica, minha formação deve enormemente ao Laboratório de Sistemas Distribuídos da UFCG. Tive o privilégio de aprender muito ao longo dos últimos seis anos trabalhando e me divertindo com pessoas competentíssimas. Devo vários dos resultados dessa tese a Jaíndson, Flávio Barata, David e Elizeu, inúmeras discussões sobre o que significam meus resultados, o que é ciência e outras dúvidas triviais a Lauro, Raquel, Ayla, Matheus, Lívia e Milena, e discussões sobre o por quê de tudo isso e como melhor jogar sinuca a Flávio Peruca, Manel, João Arthur e Marcus. Além disso, minha pesquisa não teria sido feita sem o trabalho de Zane, Randolf, Moisés, Carlinha e Tomás, o time de suporte do LSD, sem Marcelo, Elayne e Kêka mantendo as engrenagens do LSD rodando, e sem sugestões, ensinamentos e cafezinhos dos demais colegas que, espero, entenderão que eu não liste mais dezenas de pessoas nessa página.

Em duas oportunidades tive também o prazer de ser bem recebido e de ter aprendido em outros laboratórios.

No HP Labs Bristol, obrigado por me ensinarem sobre trabalho em equipe, Peter Toft, Richarde Hawkes, David Trastour, Guillaume Belrose e Andrew Byde. Além disso, obrigado a Miranda pela matemática, música e ética que tenho aprendido ao longo desses anos de colaboração e amizade.

O segundo laboratório a quem devo agradecimentos pela hospitalidade e oportunidade de trabalhar entre pessoas brilhantes é o NetSysLab, da UBC. Obrigado a Samer, Abdullah e Elizeu pelas discussões, críticas e diversão. Obrigado também a Matei, por enfrentar a burocracia, contribuir muito para este trabalho e discutir pesquisa velejando.

De volta a Campina Grande, obrigado pelas inúmeras discussões sobre o a vida, o universo e tudo mais a Elizeu, Lauro e Cássio, companheiros na aventura da ciência em Campina Grande pelos últimos seis anos (ou quase).

Devo também agradecimentos à banca avaliadora do trabalho: professores Marinho, Dorigival, Dalton e Antão, pelas sugestões para a melhoria do trabalho, e a Ângela e Pedro, da Copele, por toda a ajuda administrativa ao longo desse trabalho.

Fora do âmbito acadêmico, duas pessoas foram muito importantes para essa tese. Obrigado, Mãe, por toda a força durante esse trabalho e por ter dito que entendeu tudo na minha defesa. Obrigado, Aida, pelo caminho ladeira abaixo, por, mesmo sem entender de computação, discutir o tema de meu doutorado quando eu não sabia o que fazer, ler e corrigir minha tese e por me desaperrear sempre.

Por fim, obrigado a todos cujos impostos financiaram esse trabalho.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Justificativa . . . . .	1
1.2	Objetivos . . . . .	4
1.3	Estrutura deste documento . . . . .	5
<b>2</b>	<b>Distribuição de Conteúdo Entre-pares</b>	<b>7</b>
2.1	Compartilhamento como modalidade de produção econômica . . . . .	7
2.2	Comunidades de distribuição de conteúdo entre-pares . . . . .	9
2.3	O BitTorrent . . . . .	11
<b>3</b>	<b>Revisão Bibliográfica</b>	<b>16</b>
3.1	Caracterização de sistemas entre-pares . . . . .	16
3.1.1	Padrões de demanda em sistemas entre-pares . . . . .	16
3.1.2	Padrões de compartilhamento em sistemas entre-pares . . . . .	19
3.2	Modelagem, análise e medição do BitTorrent . . . . .	23
3.2.1	Análise do protocolo . . . . .	23
3.2.2	Medições de sistemas baseados no BitTorrent . . . . .	25
3.3	Considerações gerais . . . . .	28
<b>4</b>	<b>Metodologia</b>	<b>31</b>
4.1	Terminologia . . . . .	31
4.2	Comunidades estudadas . . . . .	32
4.2.1	Tipos de comunidades . . . . .	32
4.2.2	Sanções por nível de compartilhamento . . . . .	33
4.3	Coleta de dados . . . . .	34

4.4	Reconstruindo a dinâmica das torrentes . . . . .	37
4.4.1	Identificando nós . . . . .	37
4.4.2	Estimando a perda de informação . . . . .	39
4.4.3	Amostrando torrentes completas . . . . .	42
<b>5</b>	<b>Demanda por Recursos em Comunidades BitTorrent</b>	<b>44</b>
5.1	Qual é a distribuição de popularidade das torrentes? . . . . .	44
5.2	Como a popularidade de um arquivo se comporta no tempo? . . . . .	48
<b>6</b>	<b>Compartilhamento em Comunidades BitTorrent</b>	<b>51</b>
6.1	Como o projeto do BitTorrent influencia a contribuição dos usuários? . . . . .	52
6.1.1	Contribuição e recompensa no BitTorrent . . . . .	52
6.1.2	Free riding no BitTorrent . . . . .	54
6.1.3	Nível de Semeadura . . . . .	57
6.1.4	Tempo de semeadura e vazão para envio de dados . . . . .	60
6.1.5	Discussão . . . . .	60
6.2	Qual é a origem dos recursos providos no sistema? . . . . .	63
6.3	Como a população de grandes contribuidores se comporta no tempo? . . . . .	66
<b>7</b>	<b>A Relação entre Demanda e Oferta de Recursos em Comunidades BitTorrent</b>	<b>68</b>
7.1	Existe uma relação entre demanda e a contribuição de um usuário? . . . . .	68
7.2	As contribuições são suficientes para atender à demanda dos usuários? . . . . .	70
7.3	Há contenção por recursos nas torrentes? . . . . .	73
<b>8</b>	<b>Conclusão</b>	<b>77</b>
8.1	Principais resultados . . . . .	77
8.2	Implicações para o projeto de sistemas . . . . .	79
8.3	Trabalhos Futuros . . . . .	82

# Lista de Figuras

5.1	Popularidade de todas as torrentes nos registros (esquerda) e daquelas na amostra $\tau_{30}$ (direita). Ambos os eixos estão em escala logarítmica . . . . .	45
5.2	Adequação da distribuição cumulativa empírica da popularidade das torrentes durante todo nosso registro às distribuições log-normal e Weibull. . . . .	46
5.3	Adequação de $\lambda(t)$ e $\gamma(t)$ a três exemplos de torrentes. As torrentes à esquerda e ao centro são de <i>bitsoup</i> , enquanto a mais à direita é de <i>alluvion</i> . . . . .	49
6.1	Experimento medindo tempos relativos de download. . . . .	54
6.2	CDFs dos tempos de semeadura e vazão para envio de dados providos pelos nós. Amostras omitidas se comportam de maneira semelhante. . . . .	61
6.3	Concentração das contribuições no nível da comunidade para <i>bitsoup</i> ( $\tau_{30}$ ). . . . .	64
6.4	Concentração das contribuições no nível das torrentes nos registros de período. Os 20% principais contribuidores na torrente forneceram $x$ ou menos dos recursos em $y$ das torrentes. . . . .	64
7.1	Volume enviado e recebido por usuários em <i>bitsoup</i> . . . . .	69
7.2	Proporção de requisições servidas em cada torrente (esquerda) e no nível das comunidades (direita). Setas indicam os intervalos de 95% de confiança. . . . .	71
7.3	Número de nós e torrentes ativas em cada observação do registro de <i>alluvion</i> . . . . .	72
7.4	CDF do coeficiente de correlação de Kendall entre a velocidade de envio e recebimento de dados dos nós em cada torrente de $\tau_{30}$ . . . . .	75

# Lista de Tabelas

4.1	Características das comunidades estudadas . . . . .	33
4.2	Dados disponíveis nos relatórios das diferentes comunidades . . . . .	35
4.3	Características do registro comparativo . . . . .	36
4.4	Características dos registros de período. . . . .	37
4.5	Características das amostras $\tau_8$ e $\tau_{30}$ . . . . .	43
5.1	Comparação da percentagem das torrentes em que cada modelo foi equivalente ou superior ao modelo oponente. . . . .	49
6.1	Características das populações gerais e de free riders de <i>etree</i> , <i>easytree</i> , <i>al-luvion</i> e <i>bitsoup</i> com intervalos de confiança de 95%. Recorde que, para <i>easytree</i> , dispomos apenas das idades dos nós em dias. . . . .	56
6.2	Níveis de sementeira nas seis comunidades presentes em nosso registro comparativo. . . . .	58

# Capítulo 1

## Introdução

O objetivo deste trabalho é contribuir para o projeto de sistemas baseados em compartilhamento, permitindo o desenvolvimento de sistemas deste tipo que sejam mais eficientes que os atuais. Com este fim, caracterizamos em detalhe um sistema de distribuição de conteúdo baseado no compartilhamento e usado rotineiramente por milhões de pessoas. Essa caracterização nos permite avançar em diversas dimensões o conhecimento atual sobre os recursos de que esses sistemas dependem, a carga que eles servem e seu atual desempenho. Os resultados obtidos têm impacto no projeto de sistemas de distribuição de conteúdo, no projeto de sistemas baseados em compartilhamento e na operação da infra-estrutura que os serve.

O restante deste capítulo contextualiza este trabalho e detalha a sua relevância (Seção 1.1), seus objetivos (Seção 1.2) e a organização do restante deste documento (Seção 1.3).

### 1.1 Justificativa

Sistemas baseados em compartilhamento são hoje um componente expressivo da economia da informação. A maior enciclopédia do mundo, a Wikipédia [10], é mantida com o tempo e conhecimento de voluntários; uma das principais plataformas de computação paralela do mundo, o SETI@home [8], é formada por recursos compartilhados por seus donos; milhões de pessoas usam diariamente o BitTorrent [32] para distribuir conteúdo para grandes audiências a partir de conexões domiciliares; e, desde 1996, o servidor Web mais popular

da Internet é o Apache [11], um Software Livre, predominantemente desenvolvido por um voluntariado. Esses exemplos ilustram a prevalência dos sistemas baseados em compartilhamento como importantes atores e, por vezes, protagonistas na oferta de serviços centrais de nossa sociedade.

Todavia, embora o potencial desses sistemas seja claro, nosso entendimento sobre eles é limitado. Sistemas baseados em compartilhamento são o resultado, além de um projeto e de uma implementação, da interação entre o conjunto de mecanismos projetados e implementados e a ação coletiva de um grande número de agentes. Do ponto de vista de um sistema computacional, isso implica que o entendimento da carga a que um sistema desse tipo é submetido e da plataforma em que ele executa não é trivial.

Por outro lado, embora não seja trivial, o entendimento desses aspectos é fundamental. O funcionamento de um sistema computacional é entendido a partir de quatro componentes: seu projeto, sua implementação, a carga a que ele é submetido e a plataforma em que ele executa. O projeto e a implementação do sistema definem a eficiência possível, enquanto a carga e plataforma de execução determinam a eficiência atingida pelo sistema. Sem um conhecimento desses quatro aspectos não é possível compreender a qualidade do serviço prestado por um sistema ou otimizar seu projeto para as condições mais comuns em que ele opera, limitando nossa capacidade de projetar sistemas eficientes.

Neste trabalho, contribuímos para o conhecimento sobre a carga e a plataforma de execução de uma instância amplamente utilizada de sistemas baseados em compartilhamento: sistemas de distribuição de conteúdo entre-pares<sup>1</sup>.

Sistemas de distribuição de conteúdo entre-pares são uma alternativa baseada em compartilhamento ao modelo de distribuição cliente-servidor. No modelo cliente-servidor, o servidor é responsável por distribuir o conteúdo para todos os clientes, incorrendo em todo o custo de envio de dados. Como esse custo cresce linearmente com a quantidade de clientes, essa abordagem é inviável para provedores com recursos limitados e grande audiência. Em um sistema de distribuição de conteúdo entre-pares, o servidor divide o custo do envio de

---

<sup>1</sup>Optamos pelo termo *entre-pares* em lugar de *par-a-par* para a tradução de *peer-to-peer*. O termo *par-a-par* pode ser encontrado na literatura em Português, mas julgamos que o uso de *par*, no singular, como sinônimo de *igual* é menos natural que *pares* para designar um conjunto de iguais. Pela mesma razão, utilizamos *nó* em lugar de *par* para designar um participante do sistema.

dados com os consumidores: à medida que estes obtêm partes do arquivo, eles as enviam para outros consumidores, compartilhando sua largura de banda.

Possivelmente, o sistema entre-pares mais utilizado para a distribuição de conteúdo hoje é o BitTorrent [32]. Embora outros sistemas semelhantes tenham sido propostos [38; 1; 54] e alguns estejam disponíveis para o uso [1], nenhum deles tem aceitação tão ampla quanto o BitTorrent. Embora não haja um consenso quanto à sua popularidade, diversas medições colocam o BitTorrent como um dos agentes principais na geração de tráfego na Internet atual. Segundo um estudo da empresa alemã Ipoque focado em seu país, Austrália, Europa Oriental e Sul da Europa, durante 2007, entre 49 e 83% do tráfego na Internet naquelas regiões era gerado por sistemas entre-pares, dentre os quais o BitTorrent era o mais popular [46]. A Ipoque reporta que, por exemplo, na Alemanha, o BitTorrent era responsável por aproximadamente metade de todo o tráfego na Internet. A provedora de serviços de rede Ellacoya, a partir de uma medição em conexões de banda larga estadunidenses em 2007, atribuiu 37% do tráfego medido a sistemas entre-pares, dentre os quais o BitTorrent é o mais popular<sup>2</sup> [35].

A eficiência e a escalabilidade do projeto e implementação do BitTorrent já foram amplamente estudadas através de modelagens analíticas [63; 81; 64], simulações [25; 75] e experimentos [52; 48; 56]. Entretanto, os estudos de sua carga e da plataforma em que ele executa são atualmente limitados em amplitude e profundidade. Embora partes desses dois aspectos tenham sido analisadas a partir de dados sobre o uso do sistema, as análises que compõem o estado-da-arte não investigam a diversidade nos grupos de usuários que compartilham recursos entre si e oferecem um conhecimento limitado da relação entre a oferta de recursos por parte dos usuários e sua demanda por serviço.

Não sabemos hoje, por exemplo, se, como resultado do projeto do BitTorrent e dos hábitos de seus usuários, conteúdo disponibilizado em comunidades baseadas no compartilhamento através do BitTorrent é geralmente distribuído com sucesso; não sabemos a qualidade dessa distribuição nem os fatores que a influenciam; não sabemos como o comportamento dos usuários molda o funcionamento desses sistemas. Sem esse conhecimento, por

---

<sup>2</sup>Cabe ressaltar que ambos os estudos não detalham o método utilizado e que os dados utilizados não são públicos, o que diminui a transparência e verificabilidade dos resultados. Até onde pudemos determinar, contudo, não há estudo científico da popularidade do BitTorrent.

sua vez, não validamos a eficiência do projeto do BitTorrent, não podemos analisar se a abordagem baseada no compartilhamento é viável em comparação a suas alternativas, e perdemos oportunidades valiosas de otimizar esses sistemas.

Os resultados deste trabalho têm relevância, portanto, em três frentes. Primeiro, eles contribuem para o entendimento do funcionamento de um sistema de distribuição de conteúdo baseada no compartilhamento, a partir do qual é possível melhor informar projetistas de sistemas semelhantes. Em segundo lugar, a compreensão do serviço oferecido através dessa abordagem põe em perspectiva a eficiência de sistemas baseados no compartilhamento para a distribuição de conteúdo, uma aplicação fundamental da Internet. Finalmente, a elaboração de meios mais eficientes, escaláveis e descentralizados de distribuição de conteúdo contribui para a democratização do diálogo global permitido pela Internet.

## 1.2 Objetivos

O objetivo geral deste trabalho é investigar a tese de que, através de uma caracterização mais ampla e profunda do que as existentes na literatura sobre a carga imposta e os recursos compartilhados pelos usuários em um sistema de distribuição de conteúdo baseado no compartilhamento, é possível descobrir novos padrões relevantes para projetar sistemas deste tipo que sejam mais eficientes que os atuais e para melhor adequar a estes a infra-estrutura de rede que os serve. Sistemas mais eficientes seriam aqueles capazes de prover uma melhor qualidade de serviço que os atuais a partir dos recursos de que dispõem ou capazes de consumir menos recursos para prover a mesma qualidade de serviço.

Tendo em vista nosso objetivo geral, definimos os seguintes objetivos específicos:

1. ampliar o entendimento atual de como os padrões de consumo de conteúdo em comunidades que utilizam o BitTorrent podem ser usados para torná-lo mais eficiente, em particular com relação ao uso de *caches* para diminuir seu impacto sobre a infra-estrutura de rede;
2. avaliar como o projeto do BitTorrent influencia o comportamento colaborativo dos usuários no compartilhamento de recursos, em particular quando comparado ao comportamento dos usuários em outros sistemas baseados em compartilhamento;

3. examinar que características resultantes dos padrões de comportamento dos usuários ao compartilhar recursos no BitTorrent são relevantes para seu projeto e para o projeto de sistemas semelhantes;
4. criar conhecimento sobre como a oferta e a demanda por recursos dos usuários se relacionam em comunidades que usam o BitTorrent;
5. avaliar quantitativamente a qualidade de serviço provida em comunidades BitTorrent;
6. identificar se características sociológicas em diferentes comunidades de usuários do BitTorrent influenciam o compartilhamento de recursos por parte dos usuários e a qualidade do serviço provido nas comunidades; e
7. investigar se as comunidades que usam o BitTorrent operam tipicamente super ou sub-providas de recursos em relação à sua demanda.

Finalmente, nosso critério de sucesso é a identificação de informação valiosa para o projeto de mecanismos de distribuição colaborativa de conteúdo que seja previamente desconhecida e esteja codificada em padrões de consumo e compartilhamento de recursos no BitTorrent.

### **1.3 Estrutura deste documento**

Este documento está estruturado em mais sete capítulos além desta introdução. Esses capítulos estão organizados da seguinte forma:

*Capítulo 2: Distribuição de conteúdo entre-pares.* Neste capítulo, definimos o objeto de nosso estudo. Para tanto, discutimos a classe geral de sistemas em que ele se insere e detalhamos o funcionamento do BitTorrent.

*Capítulo 3: Revisão da bibliografia relacionada.* No terceiro capítulo, contextualizamos nosso trabalho junto à bibliografia da área. Discutimos primeiro análises e medições de outros sistemas entre-pares e, em seguida, caracterizações do BitTorrent.

*Capítulo 4: Metodologia.* Para que possamos realizar a caracterização dos comportamentos de consumo e colaboração dos usuários do BitTorrent, precisamos de registros do seu uso em produção. Nesse capítulo, apresentamos nossa metodologia de coleta de dados,

os registros que obtivemos e o método empregado na interpretação das medições obtidas que será utilizada nos capítulos subseqüentes.

*Capítulo 5: Demanda por recursos em comunidades BitTorrent.* Nesse capítulo, investigamos os registros que obtivemos para caracterizar a chegada de requisições por um arquivo durante o tempo e a popularidade de arquivos no sistema.

*Capítulo 6: Compartilhamento em comunidades BitTorrent.* Nesse capítulo, identificamos e analisamos padrões de comportamento colaborativo dos usuários do BitTorrent, comparando a proporção de usuários que não contribui nas diferentes comunidades que usam o BitTorrent com outros sistemas entre-pares e entre si, comparando os níveis de contribuição nas diferentes comunidades, examinando que parcela dos usuários é responsável pela maior parte dos recursos de que o sistema depende, e examinando como o conjunto de principais contribuidores muda com o tempo.

*Capítulo 7: A relação entre demanda e oferta de recursos em comunidades BitTorrent.* No último capítulo de nossa caracterização, investigamos se o comportamento do usuário com relação à sua contribuição é relacionado à demanda que ele impõe no sistema, quantificamos a proporção de requisições que falham nas comunidades que estudamos, e examinamos se a quantidade de recursos que os usuários contribuem para a distribuição de um arquivo é suficiente para atender a demanda da comunidade.

*Capítulo 8: Conclusão.* Nesse capítulo, encerramos o documento resumizando os resultados obtidos, colocando as contribuições em perspectiva quanto a seu impacto e discutindo os trabalhos futuros que vislumbramos a partir do apresentado neste documento.

## Capítulo 2

# Distribuição de Conteúdo Entre-pares

Neste capítulo, definimos nosso objeto de estudo e discutimos as características deste objeto que motivam nosso trabalho. Para tanto, primeiro contextualizamos sistemas de compartilhamento. Em seguida, definimos sistemas de distribuição de conteúdo entre-pares e apresentamos o funcionamento da instância desse tipo de sistema que estudaremos, o BitTorrent.

### 2.1 Compartilhamento como modalidade de produção econômica

Chamamos de sistema de compartilhamento, ou sistema baseado no compartilhamento, aquele que funciona através da produção social de um serviço ou bem com base em um *rossio*. A produção social, como definida por Benkler [24], é aquela onde os recursos necessários para a produção de um serviço ou bem são alocados com base em relações sociais, sejam essas diretas ou indiretas. Essa forma de produção se contrapõe à produção baseada em mercados e à produção organizada pelo Estado, onde os recursos são alocados através de transações monetárias e decretos governamentais, respectivamente.

Um *rossio*<sup>1</sup>, segundo o dicionário Houaiss, é um “terreno roçado e usufruído em comum”. O conceito remonta a propriedades compartilhadas na Inglaterra medieval sobre as quais um grupo de agricultores tinha direitos coletivos. Mais recentemente, após a privatização da

---

<sup>1</sup>Usamos *rossio* como tradução do termo *commons*, em inglês. Uma outra tradução comum para esse termo é bem de uso comum.

maior parte das terras inglesas, esse termo é usado para designar recursos compartilhados pela população, como ruas, praças, rios e florestas [71].

Dois exemplos de sistemas baseados em compartilhamento que são centrais em alguns nichos de nossa economia e antecedem a prevalência de dispositivos digitais são bancos de sangue e deslocamentos através de caronas. Bancos de sangue são rossios abastecidos por voluntários motivados por convicções pessoais ou atendendo a pedidos de amigos de pacientes que usaram os recursos do banco. Quanto aos deslocamentos através de caronas, mais pessoas recorrem a caronas do que às suas alternativas comerciais (como o ônibus, por exemplo) para se deslocar para o trabalho todos os dias em várias cidades dos Estados Unidos [23].

Benkler argumenta que a sociedade da informação, através da popularização do computador e da Internet, potencializou o compartilhamento como uma forma de produção de bens e serviços economicamente valiosos, tornando esse modelo um agente fundamental em nossa economia [24; 23; 22]. A disseminação do computador de uso pessoal criou um grande excedente de recursos nas mãos de um amplo contingente da população. Como é impossível adquirir um computador que atenda apenas à demanda exata de seu usuário, cada computador implica um excedente de recursos de computação, armazenamento e comunicação para seu proprietário. A Internet, por sua vez, teve dois efeitos. Primeiro, conectou o excedente proveniente dos computadores pessoais. Segundo, reduziu sensivelmente o custo de comunicação entre as pessoas, facilitando o compartilhamento de um outro excedente: o conhecimento das pessoas interconectadas. Esse cenário viabilizou empreitadas coletivas como a Wikipedia, hoje a maior enciclopédia do mundo, e a criação de sistemas computacionais baseados no compartilhamento para, por exemplo, o compartilhamento de arquivos [3; 5; 2], computação paralela [6; 8], telefonia [9] e televisão [4].

A análise de Benkler é fundamental para pôr em perspectiva sistemas baseados no compartilhamento e é complementar à nossa pesquisa. Neste trabalho, examinamos em profundidade o projeto de um sistema de compartilhamento a fim de gerar conhecimento que nos ajude a projetar sistemas de compartilhamento mais eficientes no futuro. Essa análise visa fornecer subsídios para que projetistas de sistemas futuros estejam melhor informados; para que operadores de sistemas atuais melhor entendam o impacto de sistemas de compartilhamento na infra-estrutura que os serve; e para que projetistas, operadores e usuários de

serviços semelhantes compreendam o desempenho da instância de sistemas de compartilhamento que analisamos quando comparada às suas alternativas baseadas em mercados.

## 2.2 Comunidades de distribuição de conteúdo entre-pares

Uma aplicação popular baseada no compartilhamento na Internet é a distribuição de conteúdo entre-pares. Um sistema de distribuição de conteúdo entre-pares é aquele onde os participantes do sistema que obtêm o conteúdo provêm recursos utilizados na distribuição. No caso da distribuição de conteúdo, o recurso necessário para o serviço é principalmente largura de banda. A utilização da largura de banda dos demais nós do sistema, além da do produtor de conteúdo, alivia a carga que seria imposta a este último em um sistema cliente-servidor. Assim, a distribuição torna-se mais escalável, uma vez que novos recursos são adicionados sempre que um novo consumidor se junta ao sistema.

Neste trabalho, focamos em comunidades de compartilhamento que utilizam o BitTorrent para distribuir conteúdo, que chamamos de *comunidades BitTorrent*. Note-se que analisamos essas comunidades sob a óptica de sistemas de distribuição de conteúdo entre-pares. Além disso, as distinguimos do BitTorrent em si por este ser apenas um componente de um sistema de distribuição de conteúdo. O BitTorrent é somente o protocolo de distribuição de dados e não lida, por exemplo, com a descoberta de dados ou com a autenticação e autorização de usuários.

A necessidade freqüente dessas funcionalidades popularizou o uso do BitTorrent em conjunto com sítios Web que contêm listagens do conteúdo disponível, mecanismos de busca e, ocasionalmente, mecanismos de autenticação e autorização. A utilização de sítios Web, por sua vez, leva à segmentação da população de usuários do BitTorrent. Diferentes sítios se especializam na distribuição de diferentes tipos de conteúdo, como, por exemplo, componentes do Linux<sup>2</sup>, música<sup>3</sup> e filmes de animação japoneses<sup>4</sup>. A audiência de cada um desses sítios interage através do sistema de distribuição entre-pares e comumente através de outros mecanismos como fóruns e listas de discussão. Essas audiências são entidades sociais mais coesas que a sociedade de usuários do BitTorrent e podem ser caracterizadas como comuni-

---

<sup>2</sup><http://linuxtracker.org/>

<sup>3</sup><http://bt.etree.org>

<sup>4</sup><http://www.animesuki.com/>

dades online [77].

A relevância do estudo de comunidades BitTorrent tem três frentes: a sua popularidade; a sua diversidade; e a sua preocupação com o incentivo à colaboração dos usuários. A popularidade dessas comunidades pode ser ilustrada pelos números de uma das comunidades mais acessadas do mundo, a PirateBay ([www.piratebay.org](http://www.piratebay.org)). Em setembro de 2008, a página inicial do sítio dessa comunidade relatava cerca de dez milhões de usuários online, participando da distribuição de conteúdo simultaneamente. Um fenômeno dessa magnitude tem influência importante na consideração de como podemos obter ou distribuir conteúdo e em como devemos projetar nossa infra-estrutura de rede. Assim, torna-se relevante pôr em perspectiva o desempenho e a qualidade de serviço dessas comunidades, bem como estudar maneiras de garantir que a infra-estrutura de rede utilizada que as serve o faça eficientemente.

A segunda frente da relevância é relativa à diversidade encontrada nas comunidades BitTorrent. Na maioria dos sistemas de compartilhamento de arquivos entre-pares, como o Kazaa e o Gnutella, todos os usuários formam uma única comunidade, tipicamente pouco coesa. No BitTorrent, a diversidade das comunidades de usuários e sua maior coesão dão margem ao estudo da relação entre características das comunidade — tais como propósito, normas ou identidade — e o uso e funcionamento do sistema de compartilhamento.

Finalmente, a terceira frente diz respeito à existência de mecanismos de incentivo à colaboração implementados nas comunidades BitTorrent. Em qualquer sistema de compartilhamento onde usuários que provêm ou não recursos obtêm o mesmo serviço do sistema e há um custo para a provisão, há um incentivo econômico para que usuários não provejam recurso algum. De fato, duas medições do sistema de compartilhamento de arquivos Gnutella encontraram grandes proporções de nós que apenas consomem recursos do sistema, sem contribuir nada em troca [12; 43]. Nesse sistema, o custo percebido pelos usuários provavelmente é o de que, quando se usa o protocolo de transporte TCP, fluxos de dados enviados tornam mais lentos os de recebimento [36].

Quando a capacidade do sistema é proporcional à quantidade de recursos disponível, baixos níveis de contribuição têm um impacto negativo no sistema. Por esta razão, diversos mecanismos de incentivo à contribuição foram propostos para sistemas entre-pares (eg. [15; 33; 34; 37; 49; 76; 67]). Todavia, embora diversos mecanismos tenham sido projetados e

estudados na literatura, poucos mecanismos foram implantados em sistemas em produção<sup>5</sup>. Dentre os implantados, um número ainda menor foi avaliado no seu uso em produção: até o início deste trabalho, não tínhamos conhecimento de nenhuma análise de um mecanismo em produção. A falta de uma avaliação desse tipo deixa em aberto a questão de o quanto e como um mecanismo de incentivo afeta o comportamento de usuários na prática.

O BitTorrent possui um mecanismo comprovadamente robusto de incentivo à colaboração especificado em seu projeto. Além disso, diversas comunidades BitTorrent operam um mecanismo adicional que incentiva o bom comportamento de longo prazo nas comunidades. O estudo do comportamento dos usuários e do funcionamento do sistema na presença desses dois mecanismos dá à investigação das comunidades BitTorrent a possibilidade de contribuir significativamente para preencher a lacuna do entendimento empírico de mecanismos de incentivo ao compartilhamento.

## 2.3 O BitTorrent

Nesta seção, descrevemos o funcionamento do BitTorrent, com atenção particular ao mecanismo de incentivo especificado no protocolo e ao mecanismo de incentivo extra desenvolvido por operadores de comunidades BitTorrent. Para uma descrição mais detalhada do protocolo e de sua implementação de referência, referimos o leitor para o trabalho de Konrath et al. [50].

Para obter um arquivo usando o protocolo BitTorrent, um usuário precisa se juntar à rede formada por todos os usuários participando da distribuição do arquivo<sup>6</sup>. No vocabulário BitTorrent, essa rede é chamada de *torrente*. Uma torrente é formada por nós que têm uma

---

<sup>5</sup>No início deste trabalho, além do BitTorrent, o OurGrid [31; 16], o eDonkey [2], o Kazaa [5] e o Maze [80] eram quatro outros sistemas entre-pares em produção que dispunham de mecanismos de incentivo. Porém (i) o Kazaa possuía um *hack* conhecido que permitia a usuários burlar o mecanismo [58]; (ii) não havia, em nosso conhecimento, nenhum estudo que comprovasse a eficácia do mecanismo do eDonkey; (iii) o nível de contenção por recursos do OurGrid não era alto o suficiente para uma avaliação prática da eficácia de seu mecanismo; e (iv) Yang et al. [80] reportavam que os usuários do Maze contornavam o mecanismo de incentivo do sistema obtendo novas identidades constantemente.

<sup>6</sup>Essa rede pode também distribuir múltiplos arquivos, porém esses são agrupados e vistos como uma única cadeia de bytes na torrente, tornando sua distribuição idêntica à de um único arquivo. Assim, sem perda de generalidade, nos referimos doravante ao conteúdo distribuído como um arquivo.

cópia incompleta do arquivo — chamados de *sugadores* (*leechers*, em inglês)— e nós que acabaram de baixar o arquivo e ainda estão na torrente, chamados de *semeadores* (*seeders*, em inglês).

Para distribuir um arquivo usando o BitTorrent, o distribuidor do conteúdo cria um arquivo de metadados, normalmente identificado por uma extensão ‘.torrent’. Os metadados descrevem a divisão do arquivo original em pedaços e especificam um ou mais *trackers*, que são o componente do sistema responsável por manter uma lista dos nós participando da torrente em um dado instante. O arquivo de metadados então é distribuído, normalmente através de servidores Web.

De posse do arquivo de metadados que descreve a torrente, um nó  $P$  se comunica com o tracker para se anunciar e receber uma lista de outros nós aos quais deve se conectar. Essa lista é gerada aleatoriamente pelo tracker, a fim de aproximar a topologia da torrente de um grafo aleatório. Os nós aos quais  $P$  está conectado são seus *vizinhos*; sempre que o número de vizinhos fica abaixo de um limite,  $P$  requisita uma nova lista de nós ao tracker.

Após se conectar a seus vizinhos,  $P$  troca com cada vizinho uma lista dos pedaços dos arquivos que cada um possui. Durante a participação de  $P$  na torrente, a lista dos pedaços dos arquivos que cada vizinho de  $P$  possui serve para determinar que vizinhos têm pedaços que interessam a  $P$  e para que  $P$  ordene a lista de pedaços que irá obter. Para maximizar a quantidade de pedaços que  $P$  possui que interessa a seus vizinhos, ele procura obter primeiro os pedaços mais raros em sua vizinhança. Note-se que, como efeito colateral, pedaços menos disponíveis são mais replicados entre os nós.

De posse de uma lista das peças que pode obter e de suas prioridades,  $P$  requisita pedaços dos arquivos a seus vizinhos e responde suas requisições até conseguir todo o arquivo. Quando acaba de baixar o arquivo,  $P$  se torna um semeador e passa a apenas enviar pedaços do arquivo para seus vizinhos durante o restante de sua participação na torrente. Tanto como sugador quanto como semeador,  $P$  informa periodicamente ao tracker as quantidades de dados que enviou e recebeu na torrente, sua velocidade média de envio e recebimento de dados e o tempo durante o qual está ativo na torrente.

### Alocação de recursos e incentivos

Sugadores e semeadores alocam sua largura de banda entre as conexões com seus vizinhos de forma ligeiramente diferente.

A alocação dos sugadores implementa um mecanismo de incentivo à provisão de recursos pelos próprios sugadores com uma política baseada no olho-por-olho (no inglês, *tit-for-tat*). Cada sugador possui um número limitado de conexões  $n$  (tipicamente,  $n = 5$ ) através das quais envia pedaços do arquivo para seus vizinhos. Dentre essas conexões,  $n - 1$  são alocadas para os vizinhos que forneceram as melhores taxas de envio de pedaços do arquivo em um passado recente (da ordem de dezenas de segundos), recompensando-os. A conexão restante é utilizada para a descoberta de novos nós que têm taxas de envio melhores que algum dos  $n - 1$  atualmente escolhidos. Para isso, essa conexão é alocada de forma aleatória e otimista para os nós que não estão entre os  $n - 1$  já selecionados para receber pedaços do arquivo. Caso algum dos nós selecionados aleatoriamente retribua com um serviço melhor que o de algum dos  $n - 1$  atualmente escolhidos, a lista de nós para os quais as conexões de envio são alocadas é atualizada.

As conexões de um semeador são alocadas para os nós conhecidos que tiveram as melhores taxas de recebimento de pedaços em um passado recente. De forma semelhante à alocação do sugador, uma das conexões de envio de dados é utilizada para sondar novos nós e melhorar a alocação de recursos atual. Note-se que o mecanismo que incentiva a provisão de recursos por parte dos sugadores não incentiva a provisão dos semeadores. Como estes não estão mais interessados em obter pedaços do arquivo, não são contemplados pela priorização daqueles.

### Torrentes e comunidades

O BitTorrent lida com o problema da distribuição de um arquivo em uma torrente. Uma comunidade BitTorrent, por sua vez, geralmente distribui centenas ou milhares de arquivos. Assim, uma comunidade é tipicamente constituída de múltiplas torrentes nas quais seus usuários podem estar participando simultaneamente.

Esse cenário cria um problema de alocação de recursos: um usuário que obteve os arquivos distribuídos em diversas torrentes é capaz de semear em todas essas torrentes, porém

tem uma quantidade limitada de largura de banda para dividir entre elas. Além disso, ele pode precisar dividir sua largura de banda entre essas torrentes e outras em que ele esteja atuando como sugador. A decisão de como alocar a largura de banda é feita pelo usuário do BitTorrent e, até onde pudemos determinar, não há um consenso sobre a melhor estratégia para essa alocação.

Se um usuário não é sugador de nenhuma torrente e deseja que sua contribuição seja o mais útil possível para a comunidade, não é claro que estratégia ele deve seguir. Tipicamente, a única informação a que o usuário tem acesso sobre o estado das torrentes é o número de semeadores e sugadores em cada torrente. Não é claro, a partir dessa informação, se e como os usuários podem alocar descentralizadamente seus recursos de forma eficiente: pode ser necessário levar em conta a capacidade contribuída por cada usuário, o tempo durante o qual ele contribuirá ou outras informações indisponíveis para produzir uma alocação eficiente.

Se o usuário é sugador de alguma torrente, por outro lado, o BitTorrent lhe incentiva a alocar a largura de banda entre as torrentes em que o usuário está atuando como sugador. Como não há incentivos para a semeadura e sugadores são recompensados por sua velocidade de envio de dados, essa estratégia provavelmente lhe garantirá um melhor retorno em termos de velocidade de download.

### **Sanções por nível de compartilhamento**

A ausência de incentivos para a semeadura motivou a criação de um mecanismo de incentivo adicional para comunidades BitTorrent. Algumas comunidades consideram que o mecanismo de olho-por-olho é insuficiente e, como solução, usam um mecanismo de *Sanções por Nível de Compartilhamento* (SNC) para aumentar os níveis de cooperação de seus usuários.

O *nível de compartilhamento* de um usuário é a razão entre as quantidades de dados enviada e a recebida por esse usuário em todas as torrentes de que ele participou. O mecanismo de sanções utiliza a informação sobre o progresso do download enviada pelos nós periodicamente durante sua participação nas torrentes para construir um histórico da participação dos usuários na comunidade. A partir desse histórico, o mecanismo verifica se os usuários têm um nível de compartilhamento acumulado mínimo e pune aqueles que não têm. A punição em geral acontece através da proibição do acesso a novos conteúdos e, em algumas situações, leva em consideração uma janela de tempo.

---

Note-se que esse mecanismo requer que os usuários primeiro se registrem em um sítio Web e se autentiquem para acessar uma torrente. Os clientes se identificam com o registro de seus usuários a cada requisição ao tracker, permitindo que o mecanismo mantenha o registro da atividade dos usuários. Para evitar que usuários obtenham novas identidades sempre que são marginalizados no sistema, algumas comunidades têm um limite de registros ativos consideravelmente menor que a demanda por registros, o que faz com que um usuário que tente obter uma nova identidade tenha que esperar em uma fila. Outra alternativa utilizada é impor uma penalidade aos novos usuários até que estes contribuam um mínimo para o sistema.

# Capítulo 3

## Revisão Bibliográfica

Neste capítulo, discutimos a literatura relacionada com o nosso trabalho. Dividimos essa literatura em estudos relacionados com a caracterização de sistemas de compartilhamento entre-pares que não o BitTorrent (Seção 3.1) e estudos que caracterizam o protocolo e as comunidades BitTorrent (Seção 3.2).

### 3.1 Caracterização de sistemas entre-pares

Dentro do universo de sistemas de compartilhamento, delimitamos como literatura relacionada a nosso trabalho, além de trabalhos centrados na análise do BitTorrent, caracterizações da carga e da plataforma de execução de sistemas de compartilhamento e distribuição de arquivos entre-pares. Esses sistemas são particularmente próximos ao que estudamos, pois (i) de maneira semelhante ao BitTorrent, nesses sistemas há compartilhamento, além de conteúdo, de largura de banda; e (ii) esses sistemas permitem, em seu estudo, uma abordagem semelhante à que usamos no contexto de comunidades BitTorrent.

#### 3.1.1 Padrões de demanda em sistemas entre-pares

A carga de um sistema entre-pares é definida pelos padrões na demanda por serviço de seus usuários. No contexto de sistemas de compartilhamento de arquivos, a demanda por serviço dos usuários é representada por suas requisições por arquivos. Em particular, dois aspectos comumente analisados são a distribuição dessas requisições no tempo e nos arquivos

disponíveis no sistema.

Entender se as requisições por um arquivo estão concentradas em um período curto ou estão igualmente divididas ao longo de um período extenso elucidam os requisitos de disponibilidade do sistema: o arquivo deve estar disponível enquanto há demanda por ele. A distribuição das requisições dos usuários entre os arquivos disponíveis tem relação com a utilização da infra-estrutura de rede do sistema. Se essa distribuição mostra muita duplicação nas requisições por um arquivo, a rede com frequência transmite dados duplicados, o que é uma ineficiência do ponto de vista da operação da infra-estrutura. Para um provedor de rede, importar o conteúdo apenas uma vez da Internet e transmiti-lo a partir de um *cache* para clientes em sua rede local é uma forma de aumentar a eficiência de rede.

O uso de caches é comum na distribuição de conteúdo Web e, dado o volume do tráfego de sistemas entre-pares, atrativo para a redução de custos na distribuição entre-pares. No contexto de páginas Web, essa prática é favorecida pela distribuição das requisições dos usuários entre as páginas. A popularidade de páginas Web é bem modelada por uma distribuição Zipf, o que implica que um pequeno número de páginas é muito popular enquanto maioria delas recebe poucas requisições em um período [26]. Essa observação faz com que caches relativamente pequenos tenham alta eficiência, mas a eficiência de um cache não cresça linearmente com seu tamanho.

Leibowitz et al. [53] e Gummadi et al. [40] estudam a demanda por arquivos dos usuários do Kazaa. Leibowitz et al. utilizam um registro<sup>1</sup> de aproximadamente um mês em Fevereiro de 2003 obtido a partir do tráfego de download dos usuários de um provedor de Internet israelense, enquanto Gummadi et al. obtiveram um registro de 200 dias entre maio e dezembro de 2002 instrumentando o proxy de conexão com a Internet da University of Washington.

Ambos os estudos relatam uma grande concentração do tráfego gerado no sistema em requisições por um pequeno número de arquivos. Gummadi et al. reportam que o 1% dos arquivos mais populares são responsáveis por 50% do tráfego no sistema, enquanto Leibowitz et al. reportam que em seu registro os mesmo 1% são responsáveis por 80% desse tráfego.

Quanto à popularidade dos arquivos, Leibowitz et al. reportam apenas que há também uma concentração considerável na distribuição da quantidade de requisições por arquivo: 30% dos downloads são dos 1% de arquivos mais populares. Gummadi et al. observam

<sup>1</sup>Neste documento, usamos *registro* como tradução do termo em Inglês *trace*.

que a distribuição da quantidade de requisições por arquivo é concentrada, porém não é bem representada por uma distribuição Zipf. Na distribuição Zipf, a popularidade do  $i$ -ésimo objeto mais popular é proporcional a  $i^{-\alpha}$ , com  $\alpha$  sendo o coeficiente Zipf da distribuição. Esta distribuição reflete a popularidade de documentos disponíveis na Web e, como mostrado por Gummadi et al., é comumente mencionada como um fenômeno natural na popularidade de conteúdo.

Gummadi et al. identificam que os objetos mais populares no Kazaa são menos populares do que prevê a distribuição Zipf. Os autores comparam essa descoberta com diversos outros estudos sobre a carga de sistemas de distribuição de conteúdo multimídia e sugerem que esse fenômeno é comum. Além disso, eles propõem um modelo para explicá-lo. Segundo esse modelo, embora a preferência dos usuários pelos objetos de conteúdo seja explicada por uma distribuição Zipf, a distribuição das requisições não o é por dois motivos: (i) os objetos são imutáveis e (ii) os usuários acessam tipicamente uma única vez cada objeto.

Saleh e Hefeeda [69] estudaram a popularidade dos arquivos no Gnutella a partir de um registro das buscas por arquivos e das respostas dos potenciais provedores no sistema. Os autores reportam que a popularidade de arquivos no Gnutella é semelhante à descrita por Gummadi et al. para o Kazaa, e sugerem que uma distribuição Mandelbrot-Zipf modela bem essa popularidade. Além disso, Saleh e Hefeeda mostram através de análise e simulação que à medida que a distribuição de popularidade observada se distancia da Zipf — tendo menor concentração de requisições nos itens mais populares — a eficiência do uso de caches no tráfego do Gnutella diminui.

Os estudos de Leibowitz et al. e Gummadi et al. também avaliam a distribuição da popularidade dos arquivos ao longo do tempo. Leibowitz et al. identificam que em seu experimento de aproximadamente um mês há dois tipos de arquivos muito populares: os de popularidade persistente, que correspondem a aproximadamente 15% dos arquivos, e os transientes, cujo tempo de popularidade é mais curto. Gummadi et al. observam um período consideravelmente maior que Leibowitz et al. e identificam que os arquivos mais populares no sistema são substituídos tipicamente em algumas semanas. Comparando a idade dos arquivos com sua popularidade, eles constatam que novos arquivos tendem a ser mais populares que antigos. Contudo, a maioria das requisições ao longo dos 200 dias de registro são para arquivos com mais de um mês de idade.

Finalmente, além da distribuição de requisições no tempo e entre os arquivos, outra perspectiva passível de análise na carga de um sistema de compartilhamento é a semelhança entre os interesses dos consumidores do sistema. Para entender essa perspectiva, Iamnitchi et al. [45] propuseram o estudo do grafo de compartilhamento de acesso a dados dos usuários de um sistema. Iamnitchi e Foster [44] mostram que explorando os padrões nesse grafo, é possível melhorar o desempenho de mecanismos de descoberta de recursos em um grid. De maneira similar, Sripanidkulchai et al. [73] utilizam registros de uso do Gnutella e mostram que é possível reduzir o tráfego e melhorar o desempenho da busca de conteúdo nesse sistema alterando sua topologia para que um nó requisiite primeiro conteúdo a nós que lhe proveram serviço no passado.

### 3.1.2 Padrões de compartilhamento em sistemas entre-pares

Os recursos utilizados em um sistema entre-pares são providos através de compartilhamento. Assim, para entender a plataforma em que o sistema executa, torna-se fundamental entender a disposição dos usuários de compartilhar recursos. Note-se que essa disposição pode ser resultado de motivações intrínsecas ou extrínsecas. Uma motivação intrínseca é aquela que não é causada por um incentivo óbvio externo ao agente. Por exemplo, o desejo de contribuir com caridade normalmente é associado a motivações intrínsecas. Motivações extrínsecas, em contrapartida, são aquelas relacionadas a recompensas ou punições externas ao agente, como o lucro. Em um sistema de compartilhamento sem mecanismos de incentivo, atribui-se a contribuição a motivações intrínsecas; na presença de incentivos, as motivações podem ser intrínsecas ou extrínsecas.

O comportamento dos usuários pode ser entendido num primeiro momento pela distribuição da contribuição entre os usuários. No contexto de sistemas entre-pares, esse problema tem recebido considerável atenção em investigações da quantidade de *free riding* nesses sistemas. *Free riding* é um termo da Economia que define o comportamento de atores que consomem mais que a parte que lhes é devida ou arcam com menos que o custo que lhes cabe em um sistema<sup>2</sup>.

---

<sup>2</sup>É comum na literatura relacionada a sistemas entre-pares a consideração de que um free rider é um nó que não contribui com recursos para o sistema (eg. [12; 43; 37; 42; 56]). Note-se, porém, que um nó que não contribui mas também não consome recursos do sistema não afeta o benefício dos demais nós, não sendo,

Em um artigo amplamente citado na literatura sobre sistemas entre-pares, Adar e Huberman analisam o comportamento de usuários do Gnutella em 2000 e diagnosticam a existência de uma grande proporção dos usuários que não compartilha arquivos com o sistema [12]. Adar e Huberman mediram passivamente o Gnutella e caracterizaram o comportamento colaborativo dos usuários a partir de (i) o número de arquivos disponíveis para download em cada nó observado e (ii) a quantidade de respostas positivas a buscas por arquivos enviada por cada um nó. Uma limitação mencionada pelos autores, contudo, é a impossibilidade de relacionar a colaboração e o consumo de recursos dos nós em seu experimento. Dessa forma, não é possível investigar se os nós que não contribuem de fato consomem recursos como aqueles que contribuem. Adar e Huberman optaram por considerar que todos os nós requisitam arquivos do sistema, notando que a quantidade de free riding real no Gnutella é algo entre o caso considerado e o caso onde os nós que não contribuem também não consomem recursos, uma situação onde haveria menos free riding.

Os principais resultados de Adar e Huberman que são relevantes para o nosso trabalho são que: (i) aproximadamente 66% dos nós do Gnutella não compartilham arquivo algum e (ii) uma pequena parcela dos nós é responsável por prover a maior parte do conteúdo interessante (o 1% de nós que responde mais buscas é responsável por 47% das respostas e os 25% que mais respondem, por 98% das respostas). Além disso, os autores especulam que a ausência de contribuições não é correlacionada com a capacidade da conexão dos usuários.

Em uma medição de 8 dias do Gnutella em 2002, Saroiu et al. [70] reportam que aproximadamente 25% dos clientes contactados não compartilhavam arquivo algum e que os 7% de clientes que mais compartilhavam eram responsáveis pela maioria do conteúdo disponível no sistema.

Em 2005, Hughes et al. [43] repetiram o experimento de Adar e Huberman e reportam que o nível de colaboração no sistema havia caído: 85% dos nós não compartilhavam nenhum arquivo e os 1% que respondiam mais buscas por conteúdo respondiam 50% das requisições, enquanto os 25% que mais respondiam ainda eram responsáveis por 98% das respostas. Além disso, Hughes et al. mostram que há uma correlação entre a capacidade da conexão portanto, um fardo. No restante deste documento, utilizamos a definição que considera free rider apenas os nós que consomem mais que a parte que lhes cabe, especificando as definições consideradas por outros autores quando necessário.

reportada pelo nó e a quantidade de respostas a buscas por conteúdo que esse nó responde.

Em conjunto, os estudos de Adar e Huberman, Saroiu et al. e Hughes et al. indicam que, no Gnutella, as motivações intrínsecas dos usuários não são suficientes para promover amplos níveis de contribuição. Esses estudos não deixam claro se os usuários que mais consomem do sistema são aqueles que mais provêm recursos ou se os recursos providos pela minoria que contribui são suficientes para prover uma qualidade de serviço satisfatória a todos os usuários. Contudo, sob a consideração dos autores de que a maioria dos nós no sistema requisita arquivos, a baixa proporção de contribuidores é um indício de que esse sistema tem um desempenho aquém daquele que poderia ter com mais contribuição.

Indícios semelhantes foram encontrados em outro sistema de compartilhamento de arquivos entre-pares: a rede eDonkey. Handurukande et al. reportam que aproximadamente 80% dos nós identificados em um registro de 50 dias não compartilhavam nenhum arquivo com o sistema [42]. Em outro estudo, Anagnostakis et al. reportam que 75% dos nós analisados no registro de 10 dias realizado pelos autores removeram todos os arquivos que baixaram da pasta de arquivos compartilhados menos de uma hora após obtê-los [14].

Os resultados das medições no eDonkey fornecem uma nova perspectiva sobre o comportamento colaborativo de usuários em sistema entre-pares. O eDonkey utiliza um mecanismo de incentivo que faz com que cada nó priorize as requisições dos nós que lhe serviram mais conteúdo no passado. A pouca contribuição de muitos usuários sugere que esse mecanismo não é eficaz em promover a colaboração.

Acreditamos que o nível de contribuição no eDonkey pode não ser significativamente maior que aquele observado no Gnutella por dois motivos. Primeiro, pode ser que o mecanismo de incentivo dos clientes não proveja benefícios suficientes para os nós que contribuem em comparação com os que não contribuem. A eficácia de um mecanismo baseado na reciprocidade é diretamente ligada à frequência de interação dos participantes, e não há evidência de que a interação entre dois nós no eDonkey seja frequente o suficiente para que o mecanismo implementado seja eficaz<sup>3</sup>. Uma segunda possibilidade é que os fatores que limitam a contribuição dos nós estejam além da decisão racional dos usuários do sistema.

---

<sup>3</sup>Embora a necessidade de interações frequentes para a eficácia da reciprocidade seja intuitiva, referimos o leitor interessado em mais detalhes para a análise detalhada provida por Lai et al. [51] e para a discussão mais ampla feita por Axelrod [20]

Mesmo que o mecanismo de incentivo seja eficaz, a quantidade de serviço que um nó pode contribuir para o sistema está limitada por outros fatores, como a quantidade e a capacidade dos recursos que cada nó dispõe.

Embora os trabalhos de Adar et al., Saroiu et al., Hughes et al., Handurukande et al. e Anagnostakis et al. apontem para altos níveis de não-colaboração em sistemas de compartilhamento de arquivos, essa observação não é um consenso na literatura. Zhao et al. reportam uma medição do Gnutella onde apenas 13% dos nós contactados não compartilhavam arquivo algum [82]. Zhao et al. não discutem a relação entre sua medição e o trabalho de Hughes et al, realizada em um período próximo. Eles atribuem o número menor de usuários que não compartilham arquivos em relação a medições de 2000 e 2002 à propaganda pró-compartilhamento dos distribuidores de software do Gnutella e à maior difusão do acesso a banda larga.

Independente do nível de contribuições, uma possível explicação para o comportamento não-colaborativo de parte dos usuários de um sistema de compartilhamento é que estes percebam um custo na contribuição de recursos para o sistema. Se o benefício de contribuir resultante de suas motivações intrínsecas e extrínsecas não é maior que esse custo, o usuário agindo racionalmente optará por não contribuir. Feldman et al. [36] demonstram através de simulações de fluxos TCP que embora os canais de rede sejam *full-duplex*, fazer upload impacta de forma significativa na qualidade de downloads que acontecem simultaneamente. Isso acontece porque os fluxos de envio de dados competem com as confirmações de pacotes recebidos. Essa competição ocasiona em atrasos das confirmações e potencialmente na retransmissão dos dados. Os autores sugerem que esse impacto é percebido pelos usuários como uma parte significativa do custo de contribuir para o sistema e os motiva ao free riding.

Finalmente, além da proporção de usuários que provê recursos e de quanto recurso cada usuário provê, outra perspectiva do compartilhamento de recursos é o quão estável é o comportamento dos contribuidores. Em um sistema entre-pares de *multicast* no nível da aplicação, Ripeanu et al. relatam que os nós mais propensos a fornecer um bom serviço em um instante são geralmente aqueles que proveram melhor serviço em um passado recente [65]. Embora esse resultado seja razoavelmente intuitivo, Ripeanu et al. mostram que fazendo a topologia do sistema refletir esse padrão, é possível prover, com um sistema menos complexo, um serviço de multicast tão eficiente quanto as alternativas estruturadas em torno

de protocolos mais complexos de manutenção da topologia.

## 3.2 Modelagem, análise e medição do BitTorrent

No contexto do BitTorrent, dividimos a bibliografia relacionada em análises do projeto e implementação do BitTorrent (Subseção 3.2.1) e estudos fundados em medições do comportamento dos usuários (Subseção 3.2.2).

### 3.2.1 Análise do protocolo

Diversos estudos atestam a escalabilidade e eficiência do BitTorrent através de modelagem, simulação e experimentos.

Qiu e Srikant [63] estudam o BitTorrent segundo um modelo de fluidos que mostra a escalabilidade e a alta utilização da capacidade de upload das torrentes. Yang e Veciana [81] usam outro modelo de fluidos e analisam a eficiência no uso da capacidade de upload de um sistema de distribuição colaborativa similar ao BitTorrent. Yang e Veciana reportam que a vazão do sistema tem um crescimento logarítmico em relação ao número de usuários em um surto de popularidade (*flash-crowd*). Caso as requisições cheguem segundo um processo de Poisson, o modelo dos autores aponta para um crescimento linear da capacidade com o número de usuários. Rai et al. [64] utilizam cadeias de Markov para modelar a obtenção de um arquivo por um nó e argumentam que esse processo pode ser modelado com precisão se são consideradas três fases na participação do nó na torrente.

Usando simulações, Bharambe et al. [25] analisam o BitTorrent e verificam que este é escalável, robusto a surtos de popularidade e permite uma alta taxa de utilização na largura de banda disponível. Stutzbach et al. [75] encontram resultados similares analisando um modelo de download colaborativo baseado no BitTorrent.

Experimentalmente, Legout et al. [52] usam torrentes com clientes instrumentados na plataforma de experimentos distribuída PlanetLab [7] para demonstrar algumas propriedades do BitTorrent. Especificamente, eles demonstram (i) a eficácia do mecanismo de incentivo, (ii) o agrupamento de nós com largura de banda similar e (iii) a alta utilização média da capacidade de upload do sistema durante a distribuição do arquivo.

Em contraste com os resultados de Legout et al., alguns estudos experimentais reportam

a possibilidade de free riding no BitTorrent. Um mecanismo de incentivo à colaboração é eficaz contra o free riding quando não permite que um free rider obtenha, durante um determinado período, um benefício maior que aquele de um nó que colabora com o sistema. Considerando que há um custo para a doação de recursos, se colaboradores e free riders obtivessem o mesmo benefício do sistema, a estratégia dominante seria não colaborar.

Jun e Ahamad [48] executaram experimentos no PlanetLab e encontraram situações onde não há correlação entre a quantidade de dados fornecida para o sistema por um nó e a sua velocidade de obtenção do arquivo. Locher et al. [56] alteraram um cliente BitTorrent para não enviar pedaços do arquivo para outros nós e mostraram que, em algumas torrentes reais e em ambiente controlado, um nó rodando o cliente alterado obtém o arquivo mais rápido que um nó que contribui para o sistema. Em outras torrentes, contudo, acontece o contrário.

Atribuimos o comportamento observado nesses dois trabalhos à mesma razão: o mecanismo de incentivo do BitTorrent não é projetado para incentivar a colaboração em qualquer situação. Este mecanismo é de priorização, e mecanismos desse tipo só oferecem incentivos à contribuição quando há um nível mínimo de contenção pelos recursos do sistema. Quando há recursos o suficiente para atender toda a demanda do sistema, nós que contribuem e aqueles que não contribuem obterão o mesmo benefício, uma vez que não há como priorizar nó algum. Caso os contribuidores incorram em algum custo para contribuir, há um incentivo para o free riding<sup>4</sup>.

Um mecanismo de priorização parte, portanto, do pressuposto de que o incentivo é necessário apenas quando há contenção por recursos. Se há recursos para atender a todos os usuários, a colaboração não precisa ser incentivada. Uma possível explicação para a divergência entre resultados de Jun e Ahamad e Locher et al. e os de Legout et al. é que apenas nos experimentos deste último houve contenção por recursos no sistema. Infelizmente, Jun e Ahamad e Locher et al. não fornecem detalhes suficientes sobre seu experimento para comprovar essa hipótese.

---

<sup>4</sup>Para uma análise formal das condições de eficácia de um mecanismo de priorização, referimos o leitor para dois trabalhos em que analisamos um mecanismo análogo ao do BitTorrent no contexto de grids entre-pares [17; 15]

### 3.2.2 Medições de sistemas baseados no BitTorrent

Estudos baseados em simulação, análise e experimentos em ambiente controlado do sistema assumem uma certa carga e uma certa provisão no sistema. Trabalhos de medição do uso em produção do BitTorrent são necessários para descobrir comportamentos de carga e provisão que reflitam aqueles de casos de uso reais.

#### Padrões de consumo de recursos

Assim como para a análise de padrões de consumo em sistemas entre-pares em geral (Seção 3.1.1), nesta seção focamos na distribuição das requisições dos usuários por arquivos no tempo e nos arquivos.

Bellissimo et al. [21] notam no estudo de duas comunidades BitTorrent que a quantidade de requisições que o sistema recebe pelos arquivos não é descrita por uma distribuição Zipf como acontece com o conteúdo disponível na Web. O principal tipo de conteúdo distribuído nessas comunidades é multimídia, o que aproxima a observação de Bellissimo et al. e as análises do Kazaa e Gnutella comentadas na Seção 3.1.1. Entretanto, Bellissimo et al. se limitam à observação da divergência com relação à distribuição Zipf e não comparam seus dados com os estudos mencionados ou propõem um modelo para a popularidade das torrentes.

A distribuição das requisições dos usuários por um arquivo no tempo é o processo de chegada de requisições na torrente que o distribui. Uma descrição desse processo explica a evolução da demanda por um arquivo no tempo e é necessária para a simulação e análise de torrentes. Na ausência de modelos realistas para esse processo, diversos estudos assumem que ele é bem modelado por um processo de Poisson (e.g. [55; 63; 61; 30]). Entretanto, em medições de comunidades BitTorrent diferentes, Pouwelse et al. [62], Bellissimo et al. [21] e Guo et al. [41] identificam que essa consideração não é realista. Guo et al. mostram que a taxa de chegada das requisições por uma torrente decresce exponencialmente com o tempo, sugerindo que o estudo apenas do seu estado estável negligencia a maior parte do tempo de vida da torrente.

### **Padrões de compartilhamento**

No BitTorrent, os usuários compartilham recursos sujeitos a um mecanismo de incentivo, enquanto sugadores, e baseados em motivações intrínsecas, enquanto semeadores.

Com relação ao mecanismo de incentivo do sistema, Izal et al. [47] identificam evidência de sua eficácia analisando uma torrente de cinco meses de duração distribuindo a imagem ISO do Linux RedHat 9 para mais de 180.000 consumidores. Eles observam uma correlação entre as velocidades de obtenção e envio de dados dos nós, corroborando a hipótese de que o mecanismo de incentivo do BitTorrent é eficaz.

Bellissimo et al. [21] identificam em duas comunidades BitTorrent que o tempo que os usuários passam enviando um arquivo é, em média, o dobro do tempo que passam baixando-o. Já na torrente distribuindo a imagem Linux, Izal et al. verificam que semeadores contribuem por seis horas e meia em média antes de abandonar o sistema, enquanto o tempo médio para obter o arquivo é de cerca de treze horas. Em ambos os casos, contudo, a distribuição de tempo de download entre os consumidores varia bastante e obter informações apenas a partir da média é pouco informativo.

Izal et al. documentam também a importância dos semeadores na distribuição do conteúdo: esses nós contribuem mais que o dobro da quantidade de dados que os sugadores proveram na torrente que eles estudaram.

Pouwelse et al. [62] analisam o comportamento de semeadura de 50.000 nós em uma torrente distribuindo uma cópia pirata de um jogo. Os autores constataram que a maioria dos nós desconecta poucas horas após acabar o download: apenas 17% fica mais de uma hora após acabar o download, 3,2% fica mais de 10 horas e 0,34% fica mais de 100 horas. Os autores também notam que o número de semeadores em uma torrente dez dias após seu início não fornece uma boa previsão do tempo durante o qual haverá semeadores nessa torrente. Pouwelse et al. utilizam a observação da distribuição desigual dos tempos de semeadura para explicar essa imprevisibilidade: um seeador que ficará muito tempo em uma torrente pode ser mais importante para a disponibilidade desta do que vários que ficarão pouco.

Piatek et al. [60] examinam a distribuição das contribuições de largura de banda entre os usuários de um grande número de torrentes. Os autores coletaram um registro de cerca de um mês contendo aproximadamente 60.000 torrentes provenientes de diversas comunidades.

A coleta de dados foi feita usando nós instrumentados que acompanhavam o progresso de seus vizinhos, e os autores observam que 80% da largura de banda agregada das torrentes observadas vêm de 10% dos nós.

Além de analisar como as contribuições estão distribuídas entre os usuários, os dados coletados em estudos de torrentes permitem a análise da frequência com que requisições não são corretamente atendidas. Nesse sentido, Guo et al. diagnosticaram que, em média, 10% das requisições em uma torrente falham em uma comunidade ao longo de quatro meses [41]. Piatek et al. reportam que seus clientes instrumentados não conseguiram obter dados em 25% das torrentes analisadas. A causa precisa dessa incapacidade não é discutida por Piatek et al. Por fim, os autores reportam que 11% das torrentes examinadas não tinham semeadores durante a medição.

Além do serviço provido nas torrentes, é possível analisar a disponibilidade de uma comunidade BitTorrent. Pouwelse et al. sugerem que esta é comprometida pela dependência de um ponto central de falhas, o tracker. Essa disponibilidade pode ser aumentada de duas formas: implementando-o de uma forma distribuída e tolerante a falhas ou aumentando a robustez do servidor (ou servidores) que roda(m) o software. Entretanto, consideramos o estudo da disponibilidade do tracker fora do escopo de nosso trabalho, uma vez que não depende do comportamento dos usuários do sistema.

Em 2004, no início deste trabalho, não conhecíamos nenhum estudo de medição que focasse na comparação do comportamento colaborativo de diferentes comunidades de usuários do BitTorrent. Conduzimos em 2004 um trabalho que analisa os níveis de contribuição em diferentes comunidades BitTorrent e a relação desses níveis com diferentes características de torrentes [18]. Em 2006, estendemos esse trabalho considerando o BitTorrent como uma tecnologia para a dádiva [66], que é uma forma de compartilhamento (*gifting* na literatura em inglês). Em 2008, aprofundamos nossa análise dos padrões de compartilhamento e analisamos a relação entre a demanda e a provisão de recursos em comunidades BitTorrent [19]. Esses três trabalhos são a base da caracterização que apresentamos nos Capítulos 5, 6 e 7.

Anagnostakis et al. [14] publicaram uma análise do efeito prático do mecanismo de incentivo do BitTorrent em seu uso. Em particular, os autores analisam a ocorrência de free riding em cerca de 700 torrentes. Anagnostakis et al. definem free riding como o consumo de recursos do sistema sem nenhuma contribuição. Nesses termos, os autores identificam que

em média 10% dos nós em uma torrente são free riders. Além disso, 10% dos nós deixam a torrente logo após acabar de baixar o arquivo, enquanto 50% dos nós enviam pelo menos a mesma quantidade de dados que baixam e 17% dos nós enviam pelo menos duas vezes mais do que recebem do sistema.

O trabalho de Anagnostakis et al. fornece dados importantes sobre o uso do BitTorrent. Entretanto, duas ressalvas devem ser feitas. Em primeiro lugar, a definição de free riding dos autores não considera o quanto é possível contribuir. Nós que se juntam a uma torrente onde há apenas semeadores não têm a possibilidade de enviar pedaços do arquivo para ninguém e, portanto, podem não ser considerados como free riders. Considerá-los como tal implica uma medição pessimista do comportamento colaborativo no sistema. Em segundo lugar, Anagnostakis et al. não diferenciam o comportamento dos usuários das torrentes medidas enquanto sugadores e semeadores. Assim, é difícil argumentar, a partir de seus resultados, qual a influência do mecanismo de incentivo do BitTorrent e de outras motivações de seus usuários na quantidade de colaboração observada no sistema.

### 3.3 Considerações gerais

Embora contribuam significativamente para o entendimento de comunidades BitTorrent, os estudos da demanda por recursos e das contribuições dos usuários no BitTorrent que discutimos neste capítulo incorrem em quatro limitações que motivam nosso trabalho.

Primeiramente, esses trabalhos utilizam identificadores imprecisos para observar os usuários a partir dos dados coletados. Em todos os trabalhos mencionados, o identificador de um usuário em uma torrente é seu endereço IP, algumas vezes somado a um número aleatório gerado pelo cliente em cada sessão. Esse identificador é impreciso a partir do momento que usuários utilizam endereços IP dinâmicos<sup>5</sup> ou NAT<sup>6</sup>. É possível reduzir essa imprecisão com

---

<sup>5</sup>DHCP (*Dynamic Host Configuration Protocol*) é um protocolo que permite a configuração automática de dispositivos em uma rede. Com frequência, esse protocolo é utilizado para atribuir endereços IP dinamicamente em uma rede local.

<sup>6</sup>NAT (*Network address translating*) é um processo pelo qual os endereços IP das requisições feitas por um grupo de computadores são reescritos ao passar por um roteador a fim de que este seja o único dispositivo da rede com um IP público. O termo NAT também é comumente utilizado para descrever o roteador responsável por esse processo.

alguma confiança em uma torrente, a partir de outras informações do progresso do usuário, como o total de dados baixado e enviado que ele reporta periodicamente. Contudo, não há informação semelhante que torne confiável identificar o comportamento de um usuário em diversas torrentes.

Em segundo lugar, os trabalhos discutidos são restritos em escopo, uma vez que analisam poucas torrentes [47; 74], apenas uma comunidade [21; 41; 14] ou não distinguem entre as comunidades que observam [60]. Não há, portanto, um entendimento das regularidades e peculiaridades no comportamento de usuários em comunidades BitTorrent com diferentes características na literatura revisada.

A terceira limitação dos estudos analisados é uma investigação limitada da relação entre a demanda e a oferta de recursos em torrentes [47; 74; 21; 14]. Guo et al. e Piatek et al. avançaram nessa direção, ao examinar falhas na provisão de serviço em torrentes. Argumentamos, porém, que esses resultados ainda não nos permitem entender (i) a qualidade de serviço típica de uma comunidade, uma vez que Guo et al. estudaram apenas uma comunidade e Piatek et al. não consideram esse conceito; (ii) se os usuários que mais provêm recursos nas torrentes e comunidades são também aqueles que mais consomem, ou se esses sistemas são compostos de nós predominantemente provedores e nós predominantemente consumidores; e (iii) se torrentes e comunidades operam tipicamente sub- ou super-providas de recursos.

Finalmente, os estudos mencionados que focam em comunidades BitTorrent têm limitações metodológicas na estimativa de perda de informação que resulta da amostragem realizada [21; 41]. A metodologia de Guo et al. e Bellissimo et al. se baseia em amostrar periodicamente as torrentes ativas em uma comunidade durante um intervalo de tempo. Esse método implica duas potenciais fontes de erro na medição: (i) perda de informação pela periodicidade da amostragem; e (ii) perda de informação pelo desconhecimento da duração de eventos que acabam após o fim do registro. Argumentamos que o erro resultante pode não ser negligenciável e não foi discutido na literatura.

Nosso trabalho expande o estado da arte com relação a essas quatro limitações (a) obtendo e analisando um registro de uma comunidade que usa identificadores únicos para seus usuários; (b) ampliando o escopo da caracterização realizada para diversas comunidades, uma das quais tem mais de 10,000 torrentes e um milhão de downloads durante nosso regis-

tro; (c) aprofundando a análise sobre a relação entre a provisão e demanda de recursos dos usuários, entendendo se os usuários que mais provêm são os que mais consomem recursos do sistema, quantificando aspectos da qualidade de serviço na distribuição de conteúdo e examinando se torrentes tipicamente operam em contenção de recursos ou super-providas; e (d) discutindo em profundidade os erros implicados no método usado.

Os dados que coletamos, a metodologia usada e a discussão dos erros em que nosso método incorre são detalhados a seguir, no Capítulo 4. A caracterização subsequente, que apresenta os avanços no entendimento sobre demanda e compartilhamento de recursos e da relação entre esses fatores em Comunidades BitTorrent é apresentada nos Capítulos 5, 6 e 7. Durante o restante deste documento, contrastamos nossos resultados com os trabalhos discutidos neste capítulo sempre que a comparação é relevante.

# Capítulo 4

## Metodologia

Este capítulo descreve a terminologia usada neste trabalho, as comunidades de distribuição de dados estudadas, o método para a coleta de dados e a metodologia para reconstruir o comportamento dos usuários a partir dos dados. Em nosso conhecimento, nossa metodologia é inovadora no estudo do BitTorrent ao considerar (i) diversas comunidades, (ii) incluir comunidades onde é possível identificar o comportamento de um usuário em diversas torrentes e (iii) considerar o erro inerente ao método de amostragem utilizado.

A Seção 4.1 apresenta o vocabulário que utilizamos deste ponto em diante. Na Seção 4.2, apresentamos as comunidades que usamos em nossas análises e discutimos suas características. A seguir, na Seção 4.3 detalhamos a metodologia utilizada para a coleta de dados dessas comunidades e os registros que construímos nessa coleta. Por fim, na Seção 4.4, consideramos o problema da reconstrução da dinâmica de torrentes e usuários a partir dos dados coletados, bem como o processo de amostrar apropriadamente as torrentes reconstruídas.

### 4.1 Terminologia

No restante deste documento, diferenciamos entre *usuários* e *nós*. Um usuário é um participante em uma comunidade BitTorrent. Cada vez que um usuário participa em uma torrente, ele aparece no sistema como um *nó*. Essa distinção é relevante porque, em algumas das comunidades estudadas, é possível observar com precisão apenas o comportamento dos nós e um usuário pode ser responsável por diversos nós em um instante.

Um nó *chega* em uma torrente na primeira vez que participa dela. Cada nó pode se

utilizar de diversas sessões não concomitantes em uma mesma torrente e *abandona* a torrente quando sai dela e não retorna. O tempo que o nó permanece na torrente após acabar de obter o arquivo e antes de a abandonar é o *tempo de semeadura* do nó. O *início de uma torrente* é o momento da chegada do primeiro nó na torrente. O *fim da torrente* é o momento em que o último nó abandona a torrente. A *duração de uma torrente* é o tempo entre seu início e fim, e uma torrente é *completa* em nossas medições se seu início e fim acontecem no período de nossas medições.

Consideramos dois níveis na análise de uma comunidade BitTorrent: *o nível da comunidade* caracteriza o comportamento dos *usuários* nas diversas torrentes em que eles participam. *O nível das torrentes* é usado para analisar o comportamento dos *nós* em cada torrente, sem agregar esse comportamento para observar usuários. Os níveis da comunidade e das torrentes oferecem visões complementares das comunidades e, novamente, a distinção é relevante por não podermos observar todas as métricas no nível da comunidade para algumas de nossas amostras.

## 4.2 Comunidades estudadas

Coletamos dados de cinco comunidades BitTorrent diferentes: *bt.etree.org*, *piratebay.org*, *torrentportal.com*, *easytree.org*, e *bitsoup.org*. A essas, somamos os dados disponíveis no Umass Trace Repository<sup>1</sup> sobre uma outra comunidade chamada *alluvion.org*. Doravante, nos referimos às seis comunidades como *etree*, *piratebay*, *torrentportal*, *easytree*, *bitsoup* e *alluvion*, respectivamente. A Tabela 4.1 resume as principais características dessas comunidades.

Com essa amostra de comunidades, temos diversidade nas dimensões (i) dos tipos de comunidade, no que diz respeito a seus objetivos e práticas, e (ii) da combinação de mecanismos utilizados nas comunidades.

### 4.2.1 Tipos de comunidades

As comunidades *piratebay*, *torrentportal* e *bitsoup* representam a forma prevalente de comunidade existente, onde não há moderação do tipo de conteúdo compartilhado. Como

---

<sup>1</sup><http://traces.cs.umass.edu/>

Tabela 4.1: Características das comunidades estudadas

Comunidade	Conteúdo	Mecanismos específicos
<i>etree</i>	Música (autorizada)	-
<i>alluvion</i>	Conteúdo gerado por usuários	-
<i>piratebay</i>	Filmes, música, software, etc.	-
<i>torrentportal</i>	Filmes, música, software, etc.	-
<i>easytree</i>	Música (bootleg)	Sanções por Nível de Comp.
<i>bitsoup</i>	Filmes, música, software, etc.	Sanções por Nível de Comp.

conseqüência, nessas comunidades há um grande número de torrents compartilhando material cuja distribuição é ilegal em muitos países, como software pirateado e cópias de filmes protegidos por direitos autorais.

As comunidades *alluvion* e *etree* não permitem a distribuição de conteúdo protegido por direitos autorais e *easytree* é focada em conteúdo cuja distribuição normalmente não é vista como ilegal ou imoral.

*Alluvion* é orientada à distribuição de conteúdo ligado à atividade de um fórum online. Esse conteúdo consiste principalmente de mídia produzida pelos participantes do fórum e conteúdo cuja distribuição é livre, como trailers e demonstrações de jogos. *Etree* é uma comunidade criada em torno de gravações de shows de artistas que autorizam a distribuição não-remunerada destas gravações. *Easytree* é uma comunidade fundada por colecionadores de *bootlegs* e destinada à distribuição não-comercial desse tipo de gravação. *Bootlegs* são gravações não oficiais de shows de artistas que não necessariamente autorizaram os registros. Embora os direitos autorais da gravação ainda pertençam ao artista, historicamente a prática da distribuição não-comercial de *bootlegs* existe mesmo antes da mídia digital [57] e como uma atividade menos ameaçadora para a indústria que a pirataria.

#### 4.2.2 Sanções por nível de compartilhamento

Duas das comunidades estudadas empregam o mecanismo de sanções para incentivar ainda mais a contribuição de seus usuários: *easytree* e *bitsoup*. Recorde que o nível de compartilhamento é a razão entre o contribuído e o consumido por um usuário na comunidade. A comunidade *easytree* exige que usuários tenham um nível de compartilhamento mínimo de

0,25, sob a pena de eles não poderem se juntar a novas torrentes.

*Bitsoup* tem exigências maiores que *easytree*, requerendo dos usuários um nível de compartilhamento mínimo de 0,7 enquanto o total baixado pelo usuário for menor que 9 GB e um nível de compartilhamento de 0,9 desse ponto em diante. Para diminuir o benefício que um usuário que acabou de entrar na comunidade pode ter em não colaborar, usuários que fizeram menos que 5 GB de upload são impedidos de entrar em torrentes recém-publicadas durante algumas horas se seu nível de compartilhamento não está acima de uma outra exigência. Por exemplo, se um nó que baixou menos de 5 GB tem nível de compartilhamento inferior a 0,8 ele só poderá entrar uma torrente 8 horas depois do início da torrente.

### 4.3 Coleta de dados

Os dados sobre a comunidade *easytree* foram fornecidos pelos administradores do sistema a partir do banco de dados do tracker. Para as demais comunidades, a informação dos registros que coletamos e as do registro de *alluvion* são obtidos a partir das páginas HTML de relatórios das comunidades. Essas páginas são geradas pelo tracker operado por cada comunidade a partir de informações que os nós enviam periodicamente sobre seu estado corrente. Nossos dados foram obtidos através de um coletor (*crawler*) que periodicamente baixa todas as páginas com relatórios de um determinado site. O conjunto de relatórios de uma comunidade obtidos em uma execução do coletor representa o estado de uma comunidade em um instante. Chamamos esse conjunto de uma *observação* da comunidade.

Para algumas das comunidades, obtivemos observações periódicas no intuito de estudar a dinâmica de usuários e torrentes. No entanto, como cada observação resulta em milhares de requisições ao servidor Web de uma comunidade, a frequência de observação deve ser moderada, a fim de manter razoável a carga gerada como efeito colateral. Nosso coletor executou para todas as comunidades uma vez por hora. Experimentamos com frequências mais altas (uma observação a cada 15 e a cada 30 minutos), mas a carga resultante foi vista como muito alta pelos administradores das comunidades <sup>2</sup>.

Os relatórios publicados nas seis comunidades possuem diferentes informações; por consequência, nossas observações também. A Tabela 4.2 detalha os dados disponíveis nas

---

<sup>2</sup>Recebemos um pedido formal de *etree* para reduzir a carga que impúnhamos nos servidores da comunidade e fomos bloqueados por *bitsoup*.

observações das diferentes comunidades.

Os dados em comum nas observações de todas as comunidades são as informações relatadas por torrente. Estas são: (i) o tamanho do arquivo sendo distribuído, (ii) a idade da torrente em dias, (iii) o número de nós participando e (iv) o número de semeadores na torrente.

Para *easytree*, *etree*, *alluvion* e *bitsoup*, dispomos também de informações sobre o estado reportado por cada nó da comunidade no momento da observação. Esse estado consiste, para as quatro comunidades, de: (i) o total de dados enviado e recebido, (ii) a idade do nó, representando há quanto tempo o nó se juntou à torrente e (iii) se o nó é um seador. Além disso, para *etree* e *easytree*, sabemos se ele é *conectável*, o que é verdadeiro se o tracker consegue abrir uma conexão para o nó. Se um nó não é conectável, ele está utilizando um firewall ou NAT e outros nós não podem abrir conexão para ele. Ele ainda pode enviar e receber dados por conexões que ele iniciar com nós conectáveis, mas note que dois nós que não são conectáveis não têm como trocar dados entre si.

Tabela 4.2: Dados disponíveis nos relatórios das diferentes comunidades

Comunidade	Por torrente	Por nó				Identificador
		Enviado e recebido	Idade	Seador?	Conectável?	
<i>etree</i>	•	•	•	•	•	Parte IP + porta
<i>easytree</i>	•	•	•	•	•	Ident. de sessão
<i>piratebay</i>	•					
<i>torrentportal</i>	•					
<i>alluvion</i>	•	•	•	•		Hash do IP
<i>bitsoup</i>	•	•	•	•		Ident. único

As quatro comunidades que possuem dados detalhados por nó utilizam diferentes maneiras de identificá-los. Esse fato tem implicações na identificação da atividade de um usuário através de várias observações e discutimos em detalhe como lidamos com essas implicações adiante, na Seção 4.4.

A partir das observações que obtivemos das comunidades, construímos os seguintes registros:

1. **Comparativo:** consiste de uma observação das seis comunidades estudadas. As

observações das quatro primeiras comunidades foram feitas em março de 2005, a observação de *alluvion* é de outubro de 2003 e a de *bitsoup* é de fevereiro de 2007.

2. **Etree:** consiste de observações horárias da comunidade *etree* num período de aproximadamente 10 dias em março de 2005.
3. **Alluvion:** É composto de observações feitas de hora em hora da comunidade *alluvion* ao longo de aproximadamente 50 dias entre outubro e dezembro de 2003.
4. **Bitsoup:** contém observações feitas na comunidade *bitsoup* de hora em hora durante aproximadamente 65 dias entre abril e julho de 2007.

O registro comparativo nos permite confrontar uma figura geral do estado de todas as comunidades. Como essa amostra contém um grande número de torrentes com idades diferentes, esperamos ter uma amostra representativa de todos os diferentes estágios da vida de uma torrente. As principais características do registro comparativo estão resumidas na Tabela 4.3. Note-se que embora esse registro seja composto de uma observação de cada comunidade, é possível, através dessas observações, estimar a variância nas métricas que calculamos e levar esse erro em consideração em eventuais comparações.

Tabela 4.3: Características do registro comparativo

Comunidade	Torrentes	Nós
<i>etree</i>	567	4.492
<i>easytree</i>	2.586	25.687
<i>piratebay</i>	13.054	320.900
<i>torrentportal</i>	10.115	357.428
<i>alluvion</i>	204	2.807
<i>bitsoup</i>	3.376	50.295

Os demais registros nos dão a possibilidade de analisar o comportamento dos nós em detalhe ao longo do tempo. As características desses registros estão resumidas na Tabela 4.4.

Tabela 4.4: Características dos registros de período.

Registro	Duração	Torrentes		Nós	
		total	média	total	média
<i>etree</i>	10 dias, mar. 2005	1,589	835	66,588	4,905
<i>alluvion</i>	50 dias, out.-dez. 2003	1,528	278	227,096	7,312
<i>bitsoup</i>	68 dias, abr.-jul. 2007	13,741	6,633	1,694,243	145,462

## 4.4 Reconstruindo a dinâmica das torrentes

Os registros de períodos são constituídos de observações horárias do estado dos nós e das torrentes em cada comunidade. Antes de analisá-los, é necessário reconstruir o comportamento dos nós e das torrentes ao longo do tempo a partir destas observações. Esse processo implica em três desafios, que discutimos a seguir.

### 4.4.1 Identificando nós

O primeiro é relacionado à identificação imprecisa utilizada por alguns trackers, uma questão já abordada por trabalhos anteriores [21; 41; 74; 47]. Para uma das comunidades que estudamos, *bitsoup*, não há esse problema. Nela, os usuários são obrigados a se registrar e a se autenticar para participar de uma torrente. Isso permite que as páginas de relatório dessa comunidade possuam informação precisa sobre a participação de usuários ao longo do tempo e em diferentes torrentes<sup>3</sup>.

A identificação dos usuários é imprecisa em *alluvion* e *etree*, o que nos permite apenas identificar, através de heurísticas, os nós em cada torrente. Para essas duas comunidades, utilizamos heurísticas semelhantes às descritas em trabalhos anteriores [21; 41; 47] para inferir o comportamento dos nós a partir dos dados nos registros. Note que nessas duas comunidades optamos, portanto, por não identificar um usuário ao longo de diversas torrentes.

<sup>3</sup>Ainda é possível que um usuário forneça seu registro e senha para outrem e portanto que alguns identificadores correspondam ao comportamento agregado de diversos usuários. Consideramos, contudo, que esse comportamento acontece pouco o suficiente para que o negligenciemos.

## Etree

Na comunidade *etree* os nós em uma torrente são identificados de acordo com uma versão semi-anônima da combinação entre seu endereço IP e a porta TCP utilizada pelo cliente BitTorrent. Apenas os três primeiros octetos do endereço IP e a porta do cliente são mostrados nos relatórios, produzindo identificadores semelhantes a “200.168.76.xxx:8001”. Assim, nessa comunidade (i) vários usuários com um mesmo identificador podem estar conectados simultaneamente a uma torrente e (ii) vários usuários com o mesmo identificador podem se conectar a uma torrente ao longo do tempo.

Para identificar os nós através de múltiplas sessões em uma torrente nesse registro, usamos a seguinte heurística: para cada nó ativo na observação  $o_i$ , mantemos (i) uma estimativa de sua velocidade máxima de upload e download, (ii) uma estimativa de sua velocidade média de upload e download e (iii) informação sobre o estado desse nó quanto ao progresso no download do arquivo. Dois nós ativos com o mesmo identificador em  $o_i$  são registrados como respondendo por dois usuários diferentes. Um nó  $p'$  em  $o_{i+1}$  é o mesmo nó observado  $p$  observado em  $o_i$  se (i)  $p$  e  $p'$  têm o mesmo identificador, (ii)  $p'$  não é sugador em  $o_{i+1}$  se  $p$  é semeador em  $o_i$  e (iii)  $p'$  não reporta um total de dados baixado ou enviado menor que  $p$ . Caso exista mais de um  $p'$  que satisfaça as condições para ser considerada como continuação de  $p$ , selecionamos aquele cujo total de download e upload reportados seja mais semelhante à projeção do total de download de  $p$  considerando nossa estimativa de quanto ele teria feito de upload e download até  $o_{i+1}$  com base em suas taxas de download e upload médias até  $o_i$ .

Se um mesmo identificador aparece em  $o_i$ , não aparece em  $o_{i+1}$  e aparece novamente em  $o_{i+k}$ , com  $k > 1$ , a heurística é levemente diferente. Se um nó  $p$  aparece em  $o_i$  e não em  $o_{i+1}$ , ele esteve ativo durante uma quantidade de tempo desconhecida, entre 0 e  $o_{i+1} - o_i$ . Nós consideramos que uma sessão  $j$  em  $o_{i+k}$  é a continuação de um nó  $p$  em  $o_i$  que não apareceu em nenhuma observação entre  $o_i$  e  $o_{i+k}$  se (i)  $j$  e  $p$  têm o mesmo identificador, (ii),  $j$  não anuncia o nó como sugador se  $p$  é semeador, e (iii) os totais de download e upload reportados por  $j$  são no mínimo 0,9 daqueles reportados por  $p$  e no máximo 1,1 do máximo que  $p$  pode ter chegado a baixar ou enviar, estimado segundo nosso registro de suas taxas máximas de upload e download. Os valores de 0,9 e 1,1 flexibilizam o casamento entre as quantidades porque (i) observamos empiricamente que alguns clientes reportam quantidades ligeiramente inferiores de download e upload entre sessões e (ii) nosso registro das taxas

máximas de upload e download dos nós são estimativas.

### Alluvion

No registro de *alluvion*, os nós são identificados através do hash de seu endereço IP e por um identificador aleatório gerado pelo cliente para cada sessão de participação na torrente.

O identificador de sessão nos permite diferenciar múltiplos nós conectados a uma torrente com o mesmo endereço IP simultaneamente, o que pode acontecer caso eles compartilhem um NAT.

Para identificar um mesmo nó através de diversas sessões, seguimos uma modificação da abordagem utilizada por Guo et al. [41] e Izal et al. [47] na análise de registros semelhantes. Primeiro, assim como em *etree*, consideramos em *alluvion* que um nó não muda de endereço IP entre sessões. Após o fim de uma sessão do usuário  $p$  que aparece pela última vez em  $o_i$ , consideramos que o usuário  $p$  pode voltar à torrente como  $p'$  em uma observação  $o_j$ , desde que (i)  $p'$  tenha o mesmo endereço IP que  $p$ , (ii)  $o_j$  seja posterior a  $o_i$ , (iii)  $p'$  não seja sugador em  $o_j$  se ele era semeador em  $o_i$  e (iv) as diferenças entre o total de upload e download de  $p'$  em  $o_j$  e  $p$  em  $o_i$  não seja maior que 100MB. Para *alluvion*, utilizamos um valor absoluto em lugar de uma proporção, pois verificamos que a informação sobre total de upload e download nesse registro varia de forma menos coerente que em *etree*. Experimentamos com outros limites maiores e menores e não observamos grandes variações nas características dos nós obtidos.

#### 4.4.2 Estimando a perda de informação

O segundo desafio na interpretação de nossos registros resulta da frequência de amostragem em nossa coleta de dados. Como as observações são periódicas, informação é perdida à medida que alguns eventos podem ser observados apenas com a granularidade da frequência de nossas observações ou não podem ser observados a partir de observações com a frequência que utilizamos. Nesta seção, discutimos essa perda de informação e suas implicações.

Durante o restante deste trabalho, estamos interessados na medição de quantos nós chegam a cada torrente e de cinco grandezas para cada um desses nós: instante de chegada, velocidade de upload, velocidade de download, total de dados enviado e tempo de semeadura. Nosso objetivo a partir desta medição é comparar (i) as características dos nós entre si e (ii)

as características dos nós em diferentes comunidades. Como nossa meta é a comparação, a propriedade que desejamos no tocante aos erros é que estes sejam independentes do nó ou comunidade medido: se todos os nós estão sujeitos a erros não-tendenciosos, a sua comparação é válida.

Recorde que o ciclo de vida completo de um nó consiste em chegar a uma torrente, obter todo o arquivo, semear por algum tempo e abandonar a torrente. Esse processo pode ser feito em uma ou mais sessões, com períodos offline entre elas. As duas fontes de erros em nossos dados advém: (i) de só pudermos notar eventos que acontecem entre as observações  $o_1$  e  $o_2$  no momento da observação  $o_2$ , e não no instante em que eles ocorrem; e (ii) de não conseguirmos observar algumas sessões dos nós. A contagem da quantidade de nós que chegaram a uma torrente é afetada apenas pelo segundo tipo de erro, que discutimos a seguir. A diferença entre a ocorrência de um evento e sua observação, por sua vez, é relevante para todas as outras medidas.

Podemos contornar a imprecisão gerada pelo primeiro tipo de erro no cálculo do instante de chegada dos nós, de sua velocidade de download/upload e do total de dados enviados. Nesses casos, a imprecisão potencial vem de não observarmos o instante preciso de início ou fim das sessões do nó. A partir da observação de que ele não está online na observação  $o_i$ , sabemos apenas que ele saiu da torrente entre  $o_{i-1}$  e  $o_i$ . Para contornar esse erro ao medir o instante de chegada, utilizamos a informação presente nos registros de a quanto tempo o nó está online no momento da observação. Para as velocidades de download e upload, consideramos apenas os dados enviados e recebidos entre duas medições. Dessa forma, evitamos o problema de termos imprecisão no tempo que os nós passaram online e precisão em qual volume de dados eles obtiveram e enviaram na torrente. Para o total de dados enviados, consideramos o relatado pelo nó em sua última sessão, de forma que a única quantidade de dados não contabilizada devido a esse tipo de erro é aquela enviada entre a última observação do nó e o final de sua última sessão.

O tempo de semeadura merece uma consideração mais detalhada. Primeiro, devido à diferença entre o tempo da observação e a ocorrência dos eventos, podemos não ver um nó como sugador ou não vê-lo como semeador, o que implica que não observamos com precisão tempos de semeadura muito curtos e não podemos determinar que fração dos nós acaba de baixar o arquivo. Além disso, para cada sessão semeando de um nó, há um erro na estimativa

do tempo de sementeira. Ao observarmos o nó online em  $o_{i-1}$  e offline em  $o_i$ , consideramos que ele saiu da torrente no instante de  $o_i$ . Assim, o tempo de sementeira tem um erro de até uma hora para cada sessão do nó, implicando que nós que usam muitas sessões têm um erro maior que nós que usam poucas. Embora seja possível determinar quantas sessões observamos para cada nó, não é possível inferir a quantidade de erro para cada nó sem conhecer um modelo dos tamanhos de sessão dos nós. Em contrapartida, é impossível obter esse modelo a partir de nossos dados.

É possível, entretanto, analisar quão frequentemente acontecem grandes erros e avaliar como eles afetam as comparações que desejamos fazer entre usuários. Analisando a quantidade de sessões no registro de bitsoup, observamos que menos de 3% dos nós têm mais que 20 sessões como semeadores em uma torrente. Assim, em comparações entre nós, precisamos considerar que 3% do nós podem ter seus tempos de sementeira superestimados em 20 horas ou mais. Para os demais nós, o erro pode ser de até 20 horas, e portanto faz sentido considerar apenas diferenças maiores que essa margem de erro nos tempos de sementeira.

A segunda fonte de erro em nossas medições são sessões que não observamos, que ocorrem quando um nó entra e sai da torrente entre  $o_{i-1}$  e  $o_i$ . Assumimos que a chance de não observarmos uma sessão de um nó é a mesma para todos os nós. Essa consideração é verdadeira se o fator predominante na decisão dos hábitos do nó em uma torrente são externos à torrente, tais como a hora do dia e preferências pessoais do usuário. Dado o tamanho de nossa amostra, argumentamos que esses fatores não teriam tendências entre os usuários.

Baseados na consideração de que o erro é imparcial entre usuários, desejamos limitar a quantidade de usuários que não observamos em uma torrente. Considerando que um nó tem largura de banda  $b_i$ , o arquivo tem tamanho  $f$  e  $s_i$  é o tempo de sementeira deste nó, cada nó passará  $f/b_i + s_i$  unidades de tempo na torrente. Como quanto maior o tempo que o nó passa na torrente, maior é a chance de que o observemos, a chance de observar um nó em uma torrente é diretamente proporcional ao tamanho do conteúdo distribuído na torrente e ao tempo de sementeira do nó.

Novamente, sem um modelo preciso da distribuição de tempos de sessão dos nós, não é possível precisar a quantidade de nós que não observamos. Limitamos a quantidade de erro que admitimos, portanto, apenas considerando somente torrentes distribuindo arquivos maiores que 100MB. Na comunidade *bitsoup*, aquela em que as estimativas de largura de

banda de download dos nós é a maior, 50% dos nós demoraria ao menos meia hora baixando um arquivo desse tamanho. Além disso, nas torrentes examinadas no trabalho de Stutzbach e Rejaie [74], a mediana do tempo de sementeira dos nós é uma hora. Combinadas, essas estimativas resultam em um quarto de nós que passa mais de meia hora na torrente, um quarto que passa mais de uma hora, um quarto que passa mais de uma hora e meia e um quarto que passa menos de meia hora.

Embora essa estimativa não resulte em um limite de erro preciso, verificamos empiricamente que os resultados observados nos trechos de registros para os quais temos observações a cada meia hora não são qualitativamente diferentes daqueles para os quais temos observações horárias.

### 4.4.3 Amostrando torrentes completas

A terceira complicação na análise de nossos dados resulta da duração limitada de nossos registros. Para algumas análises, como a caracterização do processo de chegada de nós em uma torrente, é necessário examinar uma amostra de torrentes completas. Além disso, é desejável que essa amostra reflita a população de torrentes da comunidade como um todo. Contudo, uma vez que os dados de cada comunidade foram coletados por períodos limitados (até 68 dias em *bitsoup*), é necessário cuidado ao amostrar as torrentes completas para não produzir uma amostra tendenciosa.

Incluindo todas as torrentes que estão completas em um registro de duração  $\alpha$ , obteríamos uma amostra tendenciosa, pois só incluiríamos uma torrente de duração  $\alpha$  se ele se iniciasse no primeiro instante de nosso registro. Torrentes de duração  $\alpha/2$ , por outro lado, teriam mais chance de ser adicionados, pois seriam incluídos se iniciassem na primeira metade do nosso registro.

Para contornar essa possível tendência com relação às durações das torrentes incluídas em nossa amostra, recorreremos ao método baseado em criação proposto por Roselli et al. [68]. Aplicando esse método, definimos, para um registro entre os tempos  $t_i$  e  $t_f$ , uma janela de tempo de duração  $\tau$ . Incluímos em nossa amostra as torrentes que (i) têm duração máxima de  $\tau$ , (ii) iniciam entre  $t_i$  e  $t_f - \tau$  e (iii) terminam entre  $t_f - \tau$  e  $t_f$ . Esse método nos permite fazer uma amostra com torrentes que refletem a distribuição típica das durações de torrentes que não duram mais que  $\tau$ .

A fim de estudar amostras que podemos comparar, amostramos as comunidades *alluvion*, *bitsoup* e *etree* com  $\tau = 8$  dias. Chamamos essas amostras de amostras  $\tau_8$ . De forma similar, amostramos as duas comunidades para as quais temos registros mais longos com  $\tau = 30$  dias e chamamos as amostras obtidas de amostras  $\tau_{30}$ . A Tabela 4.5 resume as características dessas amostras.

Tabela 4.5: Características das amostras  $\tau_8$  e  $\tau_{30}$ .

Amostra	Torrentes			Nós		
	todas	$\tau_8$	$\tau_{30}$	todos	$\tau_8$	$\tau_{30}$
<i>alluvion</i>	1,247	271	355	187,916	12,291	43,930
<i>bitsoup</i>	10,463	416	1,123	1,351,806	8,400	54,889
<i>etree</i>	284	124	-	11,788	1,764	-

## Capítulo 5

# Demanda por Recursos em Comunidades BitTorrent

A primeira parte de nossa caracterização foca na demanda gerada pelos membros de uma comunidade BitTorrent. Os padrões nessa demanda podem ser explorados por projetistas de mecanismos de distribuição de conteúdo para otimizá-los e pelos operadores da infraestrutura que serve esses mecanismos para baratear sua operação. Para informar essas duas perspectivas, consideramos as chegadas em torrentes (i.e., requisições dos usuários por arquivos) e investigamos as seguintes questões: (i) qual é a distribuição de popularidade das torrentes, medida pela quantidade de chegadas que cada torrente recebe; e (ii) como é a evolução da taxa de chegadas em uma torrente no tempo. A distribuição da popularidade do conteúdo influencia na eficiência de mecanismos de caching que visam diminuir o impacto de mecanismos de distribuição de conteúdo na infra-estrutura da rede. A composição dessa distribuição com a evolução da popularidade do conteúdo no tempo fornece um modelo de carga que pode ser utilizado por projetistas de mecanismos de distribuição para avaliar seus projetos. Para a análise de ambas, utilizamos os registros de período que temos de *alluvion*, *etree* e *bitsoup*.

### 5.1 Qual é a distribuição de popularidade das torrentes?

A popularidade de uma torrente e, por conseguinte, do conteúdo que ela distribui, pode ser medida pelo número de chegadas na torrente durante um período. Em nossos dados, podemos observar a popularidade de todas as torrentes durante os registros e a popularidade

das amostras de torrentes completas  $\tau_{30}$  e  $\tau_8$ . A primeira perspectiva mostra como o interesse dos usuários se distribui no conteúdo disponível em um período, enquanto a segunda revela o número total de usuários que se junta a uma torrente entre seu início e fim. Em ambas as perspectivas, nossa principal observação é a mesma: *a distribuição da popularidade de conteúdo em comunidades BitTorrent não tem cauda longa*.

A Figura 5.1 mostra a popularidade de todo o conteúdo em nossos registros (esquerda) e a popularidade das torrentes em  $\tau_{30}$  (direita). A distribuição da popularidade das torrentes em  $\tau_8$  tem características similares. Em todas as amostras, as curvas têm formato semelhante e claramente diferente de uma distribuição Zipf, comumente referida em modelos de popularidade de conteúdo.

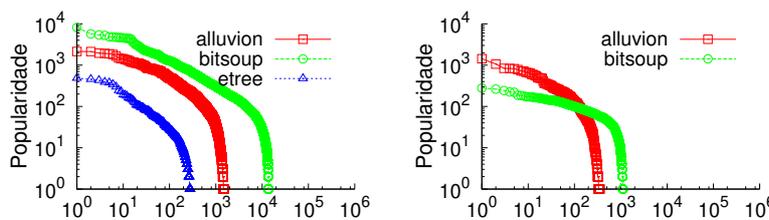


Figura 5.1: Popularidade de todas as torrentes nos registros (esquerda) e daquelas na amostra  $\tau_{30}$  (direita). Ambos os eixos estão em escala logarítmica

Em todas as amostras, uma distribuição Log-normal ou Weibull modela bem os dados empíricos. A Figura 5.2 mostra as funções de distribuição cumulativa empírica e teórica e os gráficos de quantil-quantil (qq-plots) comparando essas duas distribuições para algumas de nossas amostras<sup>1</sup>. As demais amostras se comportam de maneira similar.

As distribuições Weibull e Log-normal são diferentes daquelas observadas em sistemas de compartilhamento de arquivos entre-pares [40; 69] e *streaming* de vídeo [29], mas semelhantes à distribuição da atividade dos usuários nos tópicos em quatro sistemas baseados em produção social analisados por Wilkinson [79] e à distribuição da popularidade de filmes medida por sua renda de bilheteria [72].

<sup>1</sup>Os gráficos de quantil-quantil ou qq-plots são um teste visual do quão semelhante são as distribuições que originaram duas amostras. Se as amostras têm seus quantis idênticos, o qq-plot que compara ambas será uma linha de 45 graus partindo do ponto (0, 0). Para facilitar as comparações, desenhamos essa reta em todos os qq-plots.

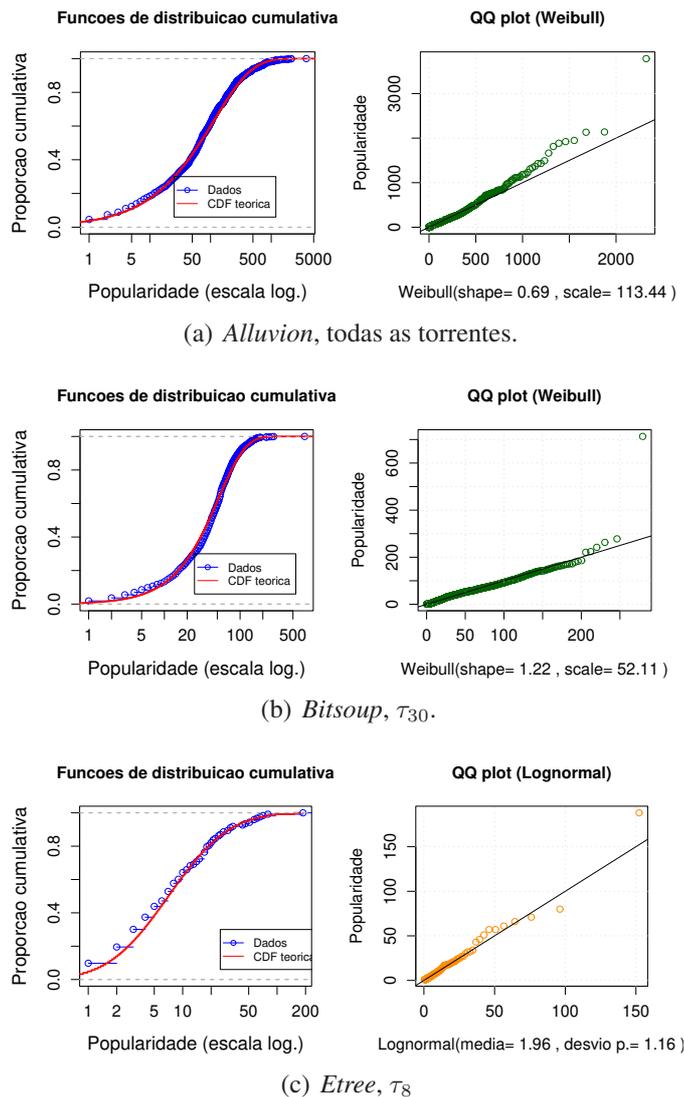


Figura 5.2: Adequação da distribuição cumulativa empírica da popularidade das torrentes durante todo nosso registro às distribuições log-normal e Weibull.

Uma característica particular das distribuições que observamos é que elas não têm cauda longa<sup>2</sup>. Devido a essa característica, as comunidades que observamos têm uma menor proporção de arquivos pouco populares que outros sistemas de compartilhamento de arquivos [40; 69]. Um fator que pode explicar essa diferença é o custo da publicação de conteúdo. É mais custoso para um usuário publicar e manter disponível um arquivo através do BitTorrent do que é fazê-lo usando outros sistemas como, por exemplo, a Web e o Kazaa. Nesses dois sistemas, uma vez que o produtor de conteúdo tenha os recursos necessários

<sup>2</sup>Matematicamente, uma distribuição tem cauda longa (*heavy tail*) se  $P[X > x] \propto x^{-\alpha}$  quando  $x \rightarrow \infty$ , com  $0 < \alpha < 2$ . Nem a distribuição Log-normal nem a Weibull obedecem a essa premissa.

para publicar o arquivo, o custo adicional para a publicação de cada novo arquivo é pequeno: basta colocar o arquivo em uma pasta específica.

No BitTorrent, por outro lado, é necessário que, para cada arquivo, o usuário utilize um aplicativo para separar o conteúdo em pedaços e gerar metadados. Em seguida, o usuário precisa publicar esses metadados e dedicar recursos para semear o arquivo por um período. Ademais, os recursos alocados para semear em uma torrente estão potencialmente indisponíveis para upload em outras torrentes nas quais o usuário está obtendo arquivos e se beneficiaria contribuindo.

### **Discussão**

A ausência de uma cauda longa na popularidade das torrentes tem implicações no projeto de caches para a infra-estrutura de rede que serve o BitTorrent. Por um lado, para caches pequenos, essas distribuições de popularidade levam a caches que são menos eficientes do que seriam caso as distribuições tivessem caudas longas. Por outro lado, grandes caches podem ter uma eficiência consideravelmente maior que no caso de distribuições de cauda longa, uma vez que a porcentagem de itens impopulares é mais limitada nos dados que observamos.

A relação que observamos entre a carga do sistema e o projeto de caches para essa carga contrasta com a observada no caching de páginas Web, cuja popularidade tem cauda longa [39; 27]. Além disso, nossa análise sugere que mecanismos de cache projetados para distribuições de cauda longa documentadas em sistemas de compartilhamento de arquivos entre-pares (e.g. [78; 69]) devem ser reconsiderados antes de ser aplicados a tráfego BitTorrent. Por fim, esses resultados complementam os de Bellissimo et al. Esses autores documentam que a popularidade dos arquivos no registro de *alluvion* que estudamos é significativamente diferente de uma distribuição Zipf [21]; nossos resultados usam uma amostra mais ampla que inclui três comunidades de compartilhamento de conteúdo e sugerem distribuições que modelam os dados analisados.

## 5.2 Como a popularidade de um arquivo se comporta no tempo?

Uma segunda dimensão da demanda dos usuários é revelada pela distribuição das chegadas ao longo da duração das torrentes. Nossa caracterização (i) *reproduz resultados anteriores que mostram que a taxa de chegada de usuários em uma torrente tipicamente decresce rapidamente a partir de seu início*, e (ii) *propõe um modelo para a evolução da taxa de chegada de usuários no tempo que é mais abrangente que o estado-da-arte*.

Estudos anteriores relatam que a taxa de chegada de usuários em uma torrente decresce rapidamente após seu início [21; 62; 41]. Guo et al. [41], baseados no registro de *alluvion* que usamos, argumentam que essa taxa decresce exponencialmente com o tempo, sendo modelada por uma função da forma  $\lambda(t) = \alpha_0 e^{-t/\beta}$ . Nós revisitamos esse modelo usando os dados de três comunidades e explorando a precisão na identificação dos usuários provida pelo registro de *bitsoup*. Note-se que para as comunidades *alluvion* e *etree* usamos heurísticas para identificar os usuários em uma torrente semelhantes às de trabalhos anteriores [21; 41].

Para analisar a evolução da taxa de chegada de usuários às torrentes, examinamos a quantidade de chegadas de usuários por dia após o início de cada torrente. Essa análise mostra que embora uma função exponencial seja capaz de modelar de forma precisa uma proporção significativa das torrentes nas três comunidades, ela falha em representar a chegada de um pequeno número de usuários por dia por vários dias próximo ao término da torrente. Esse fenômeno acontece em uma ampla proporção das torrentes e é particularmente perceptível nas torrentes de *bitsoup*.

A análise da adequação da função  $\lambda(t)$  nos motivou a propor então um novo modelo, baseado em uma função na forma  $\gamma(t) = \alpha_0 / (1 + \beta t)$ , com  $t \in \mathbb{N}$ . Essa função modela uma proporção maior das torrentes de *bitsoup* ao mesmo tempo que modela as torrentes em *alluvion* e *etree* de maneira similar à função exponencial. Assim como no modelo exponencial,  $\alpha_0$  representa a taxa de chegada de usuários inicial e  $\beta$  é um fator que influencia quão rápido essa taxa decresce com o tempo. Diferente de  $\lambda(t)$ , contudo, de acordo com  $\gamma(t)$ , a taxa de chegada de usuários decresce mais lentamente de forma geral, e a uma velocidade que varia com o tempo, sendo o decréscimo menor com a evolução da torrente. As diferenças entre os dois modelos são exemplificadas na Figura 5.3.

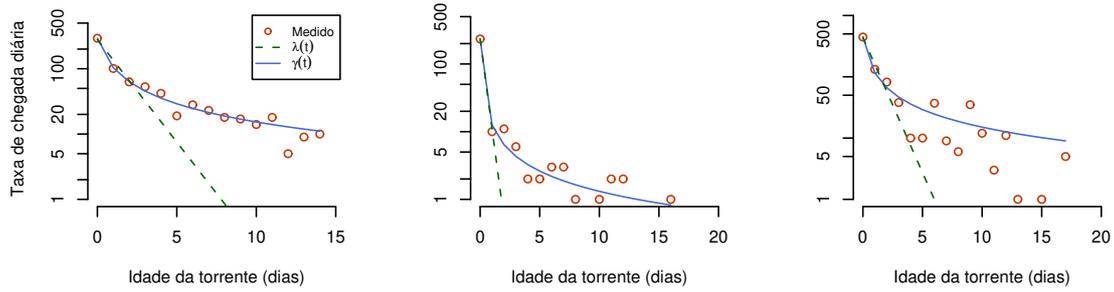


Figura 5.3: Adequação de  $\lambda(t)$  e  $\gamma(t)$  a três exemplos de torrents. As torrents à esquerda e ao centro são de *bitsoup*, enquanto a mais à direita é de *alluvion*.

Uma forma de comparar quantitativamente os modelos  $\lambda(t)$  e  $\gamma(t)$  é utilizando o critério de informação de Akaike (AIC) [28]. Esse critério quantifica o quão bem um modelo estatístico se adequa a um conjunto de dados quando comparado a outro modelo. Comparando os valores do AIC para os dois modelos, é possível diagnosticar ambos como adequados ou concluir que há mais evidência para o uso de um deles.

Utilizamos uma abordagem conservadora e consideramos que um modelo pode ser usado a menos que haja essencialmente nenhum suporte para ele em comparação com o modelo oponente. A seguir, medimos o quão frequentemente cada modelo é adequado para modelar as torrents em nossos registros.

Tabela 5.1: Comparação da porcentagem das torrents em que cada modelo foi equivalente ou superior ao modelo oponente.

Amostra	Nº de torrents	$\lambda(t)$	$\gamma(t)$
<i>etree</i> $\tau_8$	27	100%	100%
<i>alluvion</i> $\tau_{30}$	194	67%	65%
<i>bitsoup</i> $\tau_{30}$	858	40%	79%
<i>chegadas</i> < 50	406	45%	82%
50 ≤ <i>chegadas</i> ≤ 150	430	37%	75%
<i>chegadas</i> ≥ 150	22	9%	91%

A Tabela 5.1 resume a comparação para as torrents em nossos registros que duraram ao menos cinco dias e tiveram um mínimo de dez usuários. Uma pequena fração (5-10%) das torrents em cada registro não se adequou a nenhum dos modelos e não foi incluída na

tabela. Para o restante dos dados, nenhum modelo é o mais adequado para todas as torrentes e os modelos têm cobertura similar para as torrentes em *etree* e *alluvion*. Contudo, o modelo  $\gamma(t)$  se adequa consideravelmente melhor às torrentes no registro de *bitsoup*, especialmente às torrentes mais populares.

Uma possível explicação para a diferença na adequação dos modelos é a escala da comunidade *bitsoup*. *Bitsoup* é significativamente maior que *etree* e *alluvion*, o que pode resultar em usuários chegando a torrentes por períodos mais longos. É também possível que a heurística de identificação de usuários aplicada em *etree* e *alluvion* influencie as taxas de chegadas de usuários observadas nessas comunidades. Entretanto, nossos dados não nos permitem avaliar essa potencial influência.

## Discussão

Nossos resultados demonstram que o modelo  $\gamma(t)$  pode ser uma ferramenta valiosa tanto para informar o projeto quanto para sintetizar cargas úteis à análise de desempenho de sistemas de distribuição de conteúdo.

Para o projeto, nosso modelo complementa o modelo exponencial, representando um fenômeno que acontece em uma parcela significativa das torrentes que observamos e que não é bem representado pelo modelo exponencial. Esse fenômeno precisa ser levado em consideração pelos projetistas de sistemas de distribuição de conteúdo, para que estes sejam capazes tanto de lidar com uma grande demanda inicial pelo arquivo sendo distribuído quanto com um longo período de baixa demanda próximo ao término do interesse da população de usuários pelo arquivo.

Para a síntese de cargas, a representação mais realista de como a popularidade de um grande número de torrentes decresce no tempo leva a análises mais próximas da realidade. Além disso,  $\gamma(t)$  é particularmente adequada para modelar torrentes muito populares, um tipo de torrente frequentemente de interesse na análise de desempenho de mecanismos.

Finalmente, ressaltamos que o decréscimo agudo na taxa de chegada de usuários com o tempo implica que um processo de Poisson não modela precisamente a chegada de usuários no sistema. Diversos estudos (e.g. [55; 63; 61; 30]) têm se baseado nesse modelo e nossos resultados ressaltam a necessidade de reconsiderá-los com modelos mais precisos para a carga imposta no sistema simulado.

## Capítulo 6

# Compartilhamento em Comunidades BitTorrent

Os recursos disponíveis em uma comunidade BitTorrent são fornecidos por seus usuários. As ações dos usuários determinam a oferta de recursos no sistema à medida que eles (i) configuram a largura de banda máxima que seu cliente pode usar para enviar dados; (ii) determinam quanto tempo seus clientes semeiam os arquivos que já baixaram; e (iii) decidem abandonar as torrents em que estão semeando, parando de contribuir nelas.

O comportamento dos usuários influencia portanto três aspectos do sistema: (a) vazão agregada, (b) durabilidade do conteúdo e (c) volume de dados enviado. A vazão agregada do sistema é soma da taxa com que todos os nós transmitem dados em um determinado instante, e os usuários contribuem para essa vazão agregada provendo *vazão* para enviar dados. A durabilidade de um item é o tempo pelo qual ele está disponível no sistema e é influenciada pelo *tempo de semeadura* dos usuários. Finalmente, o *volume enviado* de dados é o total transferido entre todos os usuários, e é resultado de provisão de vazão ou desta em conjunto com tempo de semeadura *quando há demanda por serviço* na comunidade.

Neste capítulo, analisamos a contribuição dos usuários para esses três aspectos do sistema. Num primeiro momento, analisamos como o conteúdo compartilhado e os mecanismos de incentivo utilizados nas comunidades BitTorrent afetam o comportamento colaborativo dos usuários (Seção 6.1). Em seguida, examinamos que fração dos usuários é responsável por que parte dos recursos contribuídos nas comunidades e nas torrents (Seção 6.2). A investigação do comportamento dos usuários revela que uma minoria dos usuários são responsáveis pela maior parte das contribuições, o que nos leva à análise da

dinâmica dessa minoria (Seção 6.3).

## 6.1 Como o projeto do BitTorrent influencia a contribuição dos usuários?

O BitTorrent permite a análise da relação entre algumas opções de projeto de sistemas de compartilhamento e o comportamento colaborativo de seus usuários. Os dados que coletamos permitem uma investigação dessa relação em duas perspectivas: (i) comparando os níveis de contribuição observados no BitTorrent com aqueles observados em outros sistemas de compartilhamento; e (ii) comparando os níveis de compartilhamento de comunidades com diferentes características, seja na dimensão do conteúdo compartilhado, seja no tocante aos mecanismos de incentivo utilizados.

Nossa análise mostra que (i) *há menos free riding no BitTorrent que em outros sistemas entre-pares*, porém (ii) *o free riding não é fortemente influenciado pela presença de mecanismos de incentivo ou pelo conteúdo compartilhado*; por outro lado, (iii) *usuários semeiam mais em comunidades mais coesas e onde a distribuição dos dados não é ilegal* e (iv) *o mecanismo de sanções por nível de compartilhamento promove mais semeadura nas comunidades que o utilizam, embora não promova contribuições substancialmente maiores de vazão*.

### 6.1.1 Contribuição e recompensa no BitTorrent

Antes de medirmos o comportamento colaborativo no BitTorrent, contudo, investigamos os incentivos para a ação racional dos usuários. Um usuário agindo racionalmente procura maximizar o benefício que obtém do sistema, dadas as limitações impostas por este.

Em comunidades que não utilizam sanções por nível de compartilhamento (SNC), não há recompensas instrumentais para que um usuário semeie. Nessas comunidades, esse tipo de contribuição acontece principalmente por motivações intrínsecas do usuário, como o desejo de melhorar o serviço provido pela comunidade ou de obter reconhecimento por sua ação. Na presença de sanções, espera-se que o usuário contribua o necessário para manter seu nível de compartilhamento acima do exigido pela comunidade.

A forma como um usuário racional deve agir enquanto sugador, no entanto, é menos

previsível. Caso o mecanismo de incentivo ao envio de dados do BitTorrent fosse sempre eficaz e não houvesse custo para contribuir, um usuário racional deveria prover tanta banda quanto fosse possível para o sistema. Entretanto, como discutido no Capítulo 3, o mecanismo de incentivo do BitTorrent não é eficaz em toda situação. Quando há recursos suficientes no sistema para atender a toda a demanda dos nós, os nós que colaboram e os que não colaboram para o sistema são atendidos sem distinção. Nesse caso, se há um custo para contribuir, os nós que não contribuem obterão uma maior utilidade ao usar o sistema.

Para entender melhor a escolha necessária aos sugadores, conduzimos o seguinte experimento: selecionamos aleatoriamente 21 torrentes de *etree*, *piratebay*, *btefnet.org* e *suprnova.org*<sup>1</sup> em outubro de 2004 e março de 2005 e inserimos um nó colaborando e um nó alterado para não colaborar (ou seja, um free rider) em cada torrente, medindo o tempo de download de ambos. Em seguida, relacionamos (i) a razão entre o tempo de download do colaborador e o do nó que age como free rider com (ii) o nível de sementeira da torrente no início do experimento. O nível de sementeira é a proporção dos nós em uma torrente que é sementeira em um determinado instante. Embora esse nível não seja o único fator que indica a abundância de banda para upload no sistema, ele certamente é um dos que a influencia, uma vez que sementeiras apenas provêm banda na torrente.

A Figura 6.1 contém os resultados dos experimentos e mostra que embora o BitTorrent consiga penalizar free riders nas torrentes com baixo número de sementeiras, em torrentes com altos níveis de sementeira, free riders tiveram menor tempo de download do que nós que colaboraram na maior parte das torrentes.

Esse resultado demonstra que há um custo para a contribuição no BitTorrent. Esse custo pode ser uma manifestação direta do custo de cooperação no compartilhamento de banda identificado por Feldman et al. [36]: quando um nó faz upload, os fluxos de envio de dados competem com os pacotes de confirmação de recebimento de dados que o nó deve enviar. Essa competição atrasa os pacotes, também atrasando o recebimento de mais dados e diminuindo o desempenho dos colaboradores na obtenção de arquivos. Nesse caso, a existência do custo implica em um incentivo para que sugadores agindo racionalmente sejam free riders em torrentes onde há abundância de recursos.

---

<sup>1</sup>Suprnova.org era uma das comunidades BitTorrent mais populares da Internet no período dos experimentos, porém acabou pouco após esse período. Btefnet.org era uma comunidade centrada na distribuição de episódios de séries de televisão.

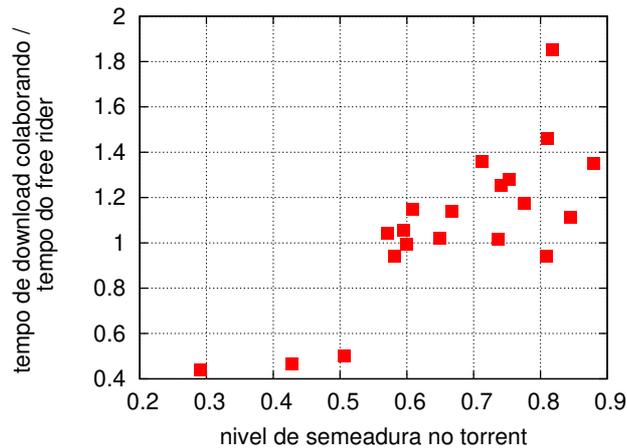


Figura 6.1: Experimento medindo tempos relativos de download.

### 6.1.2 Free riding no BitTorrent

Após a análise da motivação para o free riding, nos concentramos agora em investigar os os níveis de free riding no BitTorrent. Para definir um free rider no BitTorrent, recorreremos à definição de nível de compartilhamento utilizada no mecanismo de sanções por nível de compartilhamento. O nível de compartilhamento de um nó é a quantidade de dados enviada por esse nó sobre a quantidade de dados recebida por ele. A partir dessa definição, conceituamos dois tipos de free riders:

1. *caloteiros*, que são os nós que consomem recursos do sistema mas não contribuem recurso algum em troca, tendo o nível de compartilhamento zero, e
2. *caronas*, que são os nós que contribuem menos que o socialmente aceitável dado seu consumo de recursos do sistema. Os caronas têm um nível de compartilhamento abaixo de um limite que consideramos como 0,25, seguindo a definição de nó pouco colaborativo utilizada na comunidade *easytree*<sup>2</sup>.

<sup>2</sup>O administrador da comunidade *easytree* nos informou o seguinte procedimento de decisão para esse valor (chamado por ele de SRE) em comunicação privada via email:

”First the required minimum share ratio was .5, which lead to a lot of discussion on our Yahoo Mailing List. Pro’s and con’s were discussed and we also set up a poll. The conclusion was to lower SRE to .25 to give everybody a fair chance.”

. Essa e outras citações à comunicação privada com o administrador de *easytree* são feitas com permissão do interlocutor.

Essa distinção é relevante porque identificamos que é mais custoso para um usuário não contribuir recurso algum no BitTorrent do que em outros sistemas entre-pares. Para decidir entre atuar como um nó que colabora, um caloteiro ou um carona, um usuário agindo racionalmente levará em consideração, além do retorno que obterá em cada um dos casos, a dificuldade de agir segundo cada um desses papéis. Para diversos sistemas de compartilhamento de arquivos entre-pares, o custo para agir como um caloteiro é pequeno: basta configurar o software para não compartilhar arquivo algum. No BitTorrent, contudo, esse custo é mais alto. Avaliamos os 10 clientes mais populares nas observações de *etree* e constatamos que em nenhum deles é dada ao usuário a possibilidade de configurar o software para não fazer upload. Caso um usuário deseje esse comportamento, é necessário modificar o código do cliente. Esse esforço é considerável para uma grande parcela dos usuários, o que implica em uma maior dificuldade de atuar como caloteiro<sup>3</sup>.

Para analisar o free riding, utilizamos as quatro comunidades do nosso registro comparativo para as quais há informações individuais sobre os nós: *etree*, *easytree*, *alluvion* e *bitsoup*. A fim de minimizar a influência de torrents onde não é possível para um nó contribuir, consideramos apenas aquelas onde há pelo menos três nós e pelo menos um semeador e apenas os nós que receberam ou enviaram algum dado.

A Tabela 6.1 mostra o resultado de nossa medição da proporção de free riders nas quatro comunidades e compara as características dos caloteiros e caronas com as da população geral, formada por todos os nós.

Em todas as comunidades, há uma proporção pequena de caloteiros: no máximo cerca de 5% dos nós ativos no momento da medição. A proporção de caronas é maior: aproximadamente entre 18-28%. Essas proporções contrastam com o comportamento de usuários reportado em estudos da Gnutella e eDonkey, onde altos níveis de free riding foram documentados [12; 43; 42].

Contudo, embora as proporções de caloteiros em *easytree* e *bitsoup* – as duas comunidades que usam o mecanismo de sanções para incentivar contribuições – sejam menores que a de *etree*, a comunidade com menor proporção de caloteiros é *alluvion*. Essa observação sugere uma independência entre a quantidade de incentivos em uma comunidade e a quanti-

---

<sup>3</sup>Até onde pudemos determinar, essa dificuldade imposta pelas configurações permitidas nos clientes BitTorrent mais populares não havia sido documentada até este trabalho.

Tabela 6.1: Características das populações gerais e de free riders de *etree*, *easytree*, *alluvion* e *bitsoup* com intervalos de confiança de 95%. Recorde que, para *easytree*, dispomos apenas das idades dos nós em dias.

População	Característica	<i>etree</i>	<i>easytree</i>	<i>alluvion</i>	<i>bitsoup</i>
Geral	Nós	4.492	25.687	2.807	50.295
	Conectáveis (%)	49 ± 1	59,6 ± 0,4	-	-
	Idade	Mediana: 9,9h	67% nós ≤ 24h	Mediana: 6,4h	Mediana: 6,6h
Caloteiros	Proporção (%)	4,8 ± 0,4	3,5 ± 0,1	1,8 ± 0,3	3,6 ± 0,1
	Conectáveis (%)	15,4 ± 3,4	35,7 ± 2,2	-	-
	Idade	Mediana: 2,8h	81% nós ≤ 24h	Mediana: 1,0h	Mediana: 2,7h
Caronas	Proporção (%)	27,6 ± 0,9	22,9 ± 0,4	23,5 ± 1,1	17,9 ± 0,2
	Conectáveis (%)	40,5 ± 2	55,5 ± 0,9	-	-
	Idade	Mediana: 4,4h	74% nós ≤ 24h	Mediana: 3,3h	Mediana: 4,5h

dade de caloteiros observada nela. Uma explicação provável para essa independência é que a dificuldade imposta pelos clientes BitTorrent mais populares iniba esse comportamento em todas as comunidades.

As proporções de caronas têm uma variação maior entre as amostras e a comunidade com menor proporção de usuários agindo assim é *bitsoup*. Essa comunidade tem inclusive uma proporção de caronas menor que *easytree*, o que pode ser efeito da configuração mais agressiva de seu mecanismo de sanções por nível de compartilhamento.

Notamos, porém, que a diferença entre as quantidades de caronas de *bitsoup* ou *easytree* e *alluvion* são pequenas, sugerindo que a proporção de caronas também não é fortemente afetada pela presença do mecanismo de sanções por nível de compartilhamento. Uma possibilidade é que a combinação entre a impossibilidade de configurar o cliente para não contribuir e o mecanismo de olho-por-olho entre sugadores sejam os maiores responsáveis pela pequena proporção de caronas entre os usuários no BitTorrent.

Além disso, observamos duas características em comum na população de caloteiros e caronas quando comparadas à população geral. Primeiro, nas comunidades onde há informação sobre a conectividade dos nós, notamos que a proporção de caloteiros e caronas que é conectável é menor que a de nós conectáveis na população geral. Na população geral de *etree* e *easytree*, cerca de 50 e 60% dos nós são conectáveis, respectivamente. Os caronas dessas comunidades têm menos conectividade, por volta de 40 e 55%, respectivamente, e entre os

caloteiros as proporções são ainda menores: apenas 15 e 36% deles são conectáveis.

A relação entre a falta de colaboração e de conectividade dos nós sugere que esta impacta na capacidade dos nós de contribuir para o sistema. Esse impacto pode ser explicado pelo fato de que se dois nós não são conectáveis, eles não podem trocar dados entre si. De fato, posterior a nosso trabalho, Mol et al. [59] criaram um modelo que relaciona a condição dos nós quanto à sua conectividade e sua capacidade de contribuir. Esses autores mostram que, em certas condições, a maior parte da contribuição em uma torrente tem que vir dos nós conectáveis.

A segunda característica comum observada na população de caloteiros e caronas, e que pudemos verificar nas quatro comunidades, é uma idade relativamente baixa em comparação com os nós da população geral. A idade de um nó é o tempo entre sua última entrada na torrente e nossa medição.

Um fator que pode explicar a idade menor dos free riders é que um nó que se juntou a uma torrente mas ainda não conseguiu pedaços do arquivo que interessem a seus vizinhos não pode contribuir para estes.

### 6.1.3 Nível de Semeadura

Após a análise do free riding, voltamos nossa atenção para a semeadura nas diferentes comunidades. Primeiro focamos em torrentes e comunidades, em lugar de nós, para comparar o *nível de semeadura* em diferentes comunidades. O nível de semeadura de uma comunidade ou torrente é o total de semeadores sobre o total de nós ativos na comunidade ou torrente em um instante e é um indício do quanto os nós contribuem como tal. Assim como na seção anterior, utilizamos o registro comparativo das comunidades para essa análise.

Antes de analisar o comportamento do nível de semeadura nas comunidades, verificamos se há correlação entre o nível de semeadura e as demais características das torrentes que medimos. A única correlação significativa nas seis comunidades é uma correlação negativa entre o nível de semeadura e o tamanho do arquivo distribuído na torrente. Para algumas das comunidades, há também uma correlação positiva entre o número de semeadores e a idade da torrente.

Atribuímos a correlação negativa entre o nível de semeadura e o tamanho do arquivo distribuído a (i) o BitTorrent não ser um sistema interativo e (ii) o comportamento padrão

dos clientes BitTorrent ser atuar como semeador após baixar um arquivo. Os arquivos distribuídos usando o BitTorrent são tipicamente grandes, portanto um usuário normalmente não espera que ele seja baixado observando o cliente, como acontece ao navegar na Web. Assim, supondo que um usuário interaja de tempos em tempos com seu cliente e abandone a torrente assim que notar que o arquivo completo foi baixado, quanto menor for o arquivo, maior será o tempo que o cliente atuará como semeador na torrente.

Voltando ao nível de sementeira das comunidades, a Tabela 6.2 compara a proporção de semeadores das seis comunidades do registro comparativo. As diferenças entre o nível de sementeira das diferentes comunidades são significativas com um nível de significância de 0,01. Verificamos através de regressões parciais que as diferenças nos níveis apresentados na Tabela 6.2 não são explicadas pelas diferenças em idade e tamanho de arquivo das torrentes nas diferentes comunidades.

Tabela 6.2: Níveis de sementeira nas seis comunidades presentes em nosso registro comparativo.

Comunidade	Proporção de semeadores	Outras características	
		Estritamente legal	Mecanismos
piratebay	28%		
torrentportal	31%		
alluvion	42%	sim	
etree	52%	sim	
easytree	59%		Sanções por N. C.
bitsoup	76%		Sanções por N. C.

Atribuímos as diferenças na proporção de semeadores observada nas comunidades que investigamos a uma combinação de fatores técnicos e sociais:

1. *O tipo de conteúdo compartilhado tem influência no nível de sementeira da comunidade.* Entre as comunidades que não utilizam sanções por nível de compartilhamento, aquelas que distribuem um tipo de conteúdo estritamente legal (*etree* e *alluvion*) têm mais sementeira.

Outro indício da relação entre o nível de sementeira e o conteúdo compartilhado na comunidade foi relatada pelo moderador de *easytree* em uma entrevista que realizamos

com ele em 2004. Recorde que *easytree* permite apenas a distribuição de gravações bootleg e possui mecanismos em uso para evitar a distribuição de música comercializada. Na entrevista, o moderador atribuiu parte do elevado nível de cooperação dos usuários de *easytree* à cultura de compartilhamento herdada de colecionadores de bootleg offline:

“EZT [*easytree*] is in operation since January 11, 2004. It started off with a small, but dedicated user base which continuously kept growing. When STG (Sharing the Groove [*uma outra comunidade BitTorrent com público semelhante*]) started having problems, more and more users came over to EZT, especially after the article about STG in the New York Times, where EZT was also mentioned. The number of what we call “suck-and-duck-leechers” also increased, unfortunately. Many of the good sharers were rather annoyed about this behaviour.

(...)

We see more and more people starting to use BitTorrent who don't originally come from the “trading circles”, so they don't know anything about trading etiquette. There were (and there still are) lots of bootleg trading circles/ mailing lists, where you could (and can) get bootlegs, but you have to give something in return – either a copy of a bootleg of your own collection or blanks & postage. Or you're given a freebie, but then you're asked to re-offer freebies back to the group. So those coming from the trading circles know that all is about sharing and spreading the music.”

De fato, o tipo de conteúdo compartilhado tem potencialmente relação com outros aspectos sociais da comunidade, como seu tamanho, coesão e identidade. Infelizmente, nossa amostra não nos permite investigar mais a fundo o quanto cada um desses fatores influencia na quantidade de cooperação observada nas comunidades. Assim, apenas ressaltamos a observação de que há características sociológicas que têm relação com as quantidades de contribuição observadas nas comunidades que analisamos.

2. *Sanções por nível de compartilhamento aumentam o nível de sementeira de uma comunidade.* *Easytree* e *bitsoup* têm significativamente mais sementeira que as demais comunidades que estudamos. Atribuímos esse fato ao mecanismo de sanções por nível de compartilhamento. Além disso, *bitsoup* tem consideravelmente mais sementeira que *easytree*, o que atribuímos às maiores exigências de contribuição impostas pela configuração que essa comunidade usa nas sanções.

### 6.1.4 Tempo de sementeira e vazão para envio de dados

Após investigar os níveis de sementeira no registro comparativo, focamos agora na análise dos registros de período, para entender como os nós contribuem *tempo de sementeira* e *vazão* para o envio de dados.

A comparação entre as comunidades só pode ser feita no nível das torrentes, uma vez que é impossível analisar a contribuição em nível de comunidade dos usuários de *etree* e *alluvion*. Além disso, a comparação direta da vazão e tempo de sementeira no nível das torrentes só faz sentido se essas métricas não estiverem correlacionadas com as características das torrentes. Caso contrário, eventuais diferenças observadas nas contribuições podem ser resultado de diferenças nas características das torrentes, e não nos hábitos dos usuários.

Uma análise das correlações entre as contribuições dos nós e as características das torrentes em que eles participam mostra que a vazão e o tempo de sementeira não têm correlação com as características do arquivo sendo distribuído ou com o tempo que um nó leva para obtê-lo. Dessa forma, é possível analisar as contribuições dos nós independentemente da torrente que eles participam e compará-las para as três comunidades.

A Figura 6.2 mostra que as distribuições dos dois tipos de contribuição são bastante concentradas. Além disso, a distribuição da quantidade de vazão contribuída é semelhante em todas as comunidades, enquanto há uma diferença notável quanto ao tempo de sementeira em *bitsoup*, onde os usuários tipicamente semeiam por mais tempo. Em  $\tau_{30}$ , apenas cerca de 5% dos nós contribui grandes quantidades de vazão ( $\geq 100$  KB/s), 20-30% dos nós não contribui banda alguma, e 40% contribui entre 5 e 50 KB/s;  $\tau_8$  tem resultados qualitativamente semelhantes. Com relação ao tempo de sementeira, contudo, observamos que os resultados das três comunidades são menos semelhantes: usuários tipicamente semeiam por mais tempo em *bitsoup* que em *etree* and *alluvion*. Novamente, atribuímos essa diferença ao mecanismo de SNC, corroborando os resultados da Seção 6.1.3.

### 6.1.5 Discussão

Nossas observações sugerem que o projeto do BitTorrent tem sucesso na redução do free riding. Os níveis de free riding observados nas comunidades que medimos foram consideravelmente menores que aqueles reportados em outros sistemas entre-pares. Essa constatação

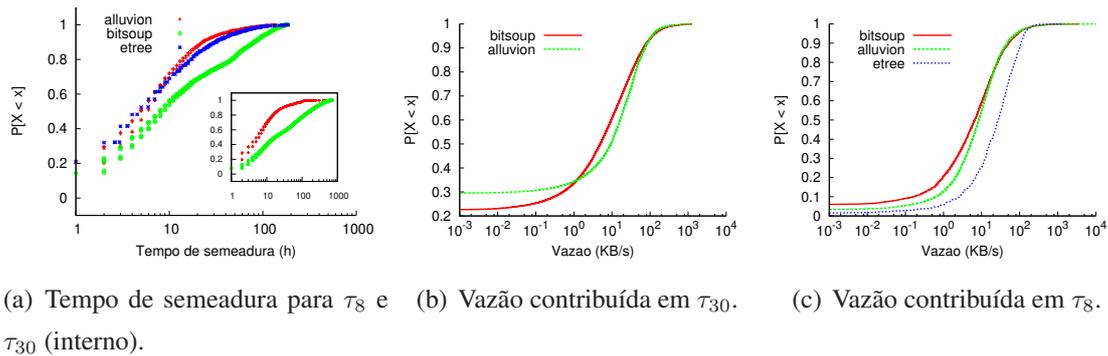


Figura 6.2: CDFs dos tempos de sementeira e vazão para envio de dados providos pelos nós. Amostras omitidas se comportam de maneira semelhante.

é verdadeira tanto considerando free riders como nós que não contribuem recurso algum quanto os considerando nós que contribuem menos que o socialmente aceitável.

Nossa análise sugere, contudo, que os mecanismos de incentivo ao upload não são o único fator que contribui para o comportamento mais colaborativo dos usuários no BitTorrent. Caso esse fator dominasse as motivações para a contribuição, *bitsoup* e *easytree* — as comunidades que empregam mecanismos de incentivos extras — deveriam ter proporções de free riders claramente menores que as demais comunidades. Não observamos essa diferença em nossas amostras. Além disso, entre os free riders, observamos que as proporções de caloteiros são semelhantes entre as comunidades.

Identificamos um outro fator que julgamos também ser importante no nível de contribuição observado no BitTorrent: os clientes mais utilizados em nossas amostras não permitem que o usuário configure o software para não contribuir. Essa prática afeta o comportamento de usuários que não têm conhecimento ou não estão dispostos a investir o esforço necessário para modificar o software. Até onde pudemos determinar, essa característica dos clientes ainda não havia sido documentada e argumentamos que ela é relevante no estudo e projeto do BitTorrent e de sistemas semelhantes.

Uma recomendação natural que surge a partir dessa observação é que a prática de criar um nível mínimo de contribuição em lugar de permitir que usuários facilmente passem a não contribuir pode ser explorada em outros sistemas de compartilhamento entre-pares. Por exemplo, sistemas de compartilhamento de arquivos poderiam criar um espaço mínimo reservado para arquivos que são necessariamente providos para a comunidade, em lugar de dar ao usuário a opção de deixar em um diretório os arquivos que quer compartilhar. Naturalmente,

o custo da contribuição mínima para o sistema precisa ser mantido baixo.

Nossa análise da sementeira mostra que esta é afetada pela presença do mecanismo de sanções por nível de compartilhamento. Porém, embora a presença de incentivos aumente a quantidade de sementeira numa comunidade, outros dois fatores têm um papel relevante na definição da quantidade de colaboração desse tipo que encontramos: as características sociais da comunidade e um comportamento padrão no software de compartilhamento. Argumentamos que esses são dois fatores cuja importância pode ser revista e explorada no projeto de comunidades BitTorrent e de outros sistemas semelhantes.

As características sociais da comunidade de usuários podem ser exploradas para motivar ainda mais o compartilhamento. Nossa análise sugere que, na ausência de sanções por nível de compartilhamento, usuários fazem mais sementeira em comunidades onde é mais seguro contribuir e cuja identidade é clara. Evidenciar essas características em outros sistemas que as possuam pode resultar em mais contribuições.

Juntas, as observações da semelhança nos níveis de vazão contribuídos e da diferença nos nível e no tempo de sementeira nas comunidades que operam o mecanismo de SNC trazem uma nova perspectiva sobre do efeito desse mecanismo. Nossa análise revela que usuários tipicamente tentam aumentar o volume de dados enviado no sistema semeando por mais tempo, e não provendo mais vazão para o sistema. Esse entendimento é importante para os projetistas e operadores de mecanismos de incentivo para o BitTorrent ou sistemas semelhantes, uma vez que provê evidência de qual recursos os usuários estão dispostos a investir como resposta a incentivos para aumentar seu volume de dados enviado.

Finalmente, apesar da diferença no tempo de sementeira nas comunidades, a distribuição do tempo que os nós passam semeando em todas as comunidades é consideravelmente concentrada. Essa regularidade é uma evidência que corrobora com a conjectura de Pouwelse et al. [62] de que o comportamento de usuários individualmente é mais relevante para determinar a longevidade de uma torrente que o número de semeadores online em um dado instante. A partir de um dado momento, é mais provável que um arquivo se mantenha disponível por mais tempo se há um semeador que permanecerá na torrente por um longo período do que se há diversos que permanecerão períodos curtos.

## 6.2 Qual é a origem dos recursos providos no sistema?

Nossa análise das contribuições dos usuários mostra que *no nível da comunidade, as contribuições de todos os tipos estão concentradas em uma pequena proporção dos contribuidores, enquanto no nível das torrentes, as contribuições são tipicamente menos concentradas.*

Para ambas as análises, utilizamos os registros de períodos. Além disso, a análise da contribuição dos usuários no nível da comunidade só é possível no registro de *bitsoup*, uma vez que apenas ele contém identificação precisa dos usuários.

A Figura 6.3 mostra como contribuições de vazão, tempo de semeadura e volume enviado estão concentradas na população de usuários de *bitsoup*. Todos os tipos de contribuição são consideravelmente concentrados: os 20% de usuários que mais contribuem são responsáveis por cerca de 80% dos três tipos de contribuição.

Contudo, embora uma minoria dos usuários forneça a maior parte dos recursos do sistema, a concentração que observamos no BitTorrent é menos pronunciada que aquela encontrada em estudos das redes de compartilhamento de arquivos Gnutella [12; 43] e eDonkey [42]. Hughes et al. [43] reporta que 1% dos usuários no Gnutella eram responsáveis por 50% das respostas a buscas por arquivos e 25% dos usuários respondiam 98% das respostas em 2005. Nos dados que observamos, 1% dos usuários são responsáveis por 36, 44 e 68% das contribuições de volume enviado, banda e tempo de semeadura respectivamente; porém os 25% dos usuários que mais contribuem são responsáveis por 49, 54 e 83% dessas contribuições. Embora a concentração no 1% de usuários mais ativos seja comparável no Gnutella e nos dados que analisamos, a concentração nos 25% principais contribuidores é consideravelmente menor no BitTorrent. Handurukande et al. [42] relatam que 80% dos nós em seus registros não compartilharam arquivos com o sistema. Essa situação implica que 20% dos nós eram responsáveis por 100% das contribuições nesse período, uma concentração também maior que a que observamos em *bitsoup*.

A perspectiva do comportamento agregado na comunidade, contudo, é incompleta. Uma alta concentração no nível da comunidade não implica que contribuições também são concentradas nas torrentes. A investigação no nível das torrentes usa as amostras de torrentes completas das três comunidades. Nós focamos nos 20% principais contribuidores em cada

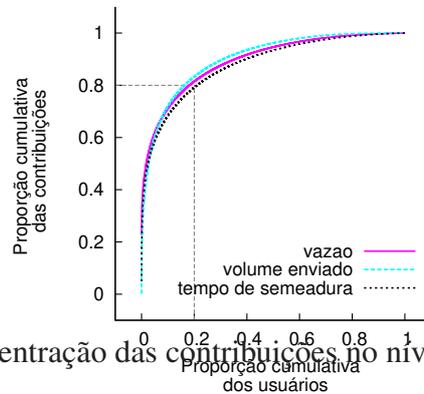


Figura 6.3: Concentração das contribuições no nível da comunidade para *bitsoup* ( $\tau_{30}$ ).

torrente e medimos a proporção da contribuição total feita por eles. A Figura 6.4 mostra a CDF da proporção da contribuição total feita por esses contribuidores em cada torrente: a proporção  $y$  de todas as torrentes teve  $x$  ou menos dos seus recursos fornecidos pelos 20% principais contribuidores.

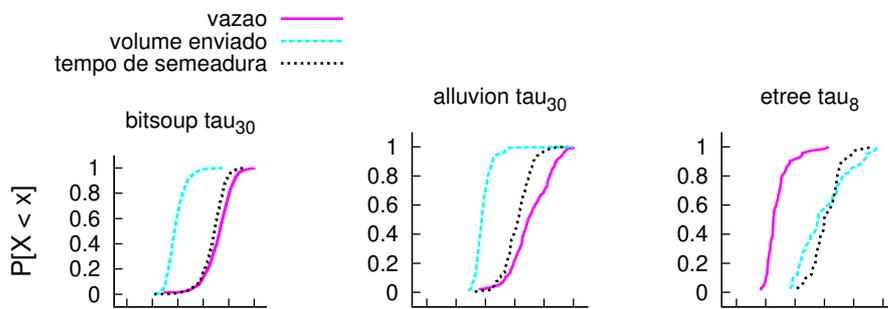


Figura 6.4: Concentração das contribuições no nível das torrentes. Os eixos Y são  $P[X < x]$  e os eixos X são 'Contribuído pelos 20% principais contribuidores'. Os 20% principais contribuidores forneceram  $x$  dos recursos em  $y$  das torrentes.

A concentração das contribuições é menos pronunciada no nível das torrentes que no nível da comunidade em todas as amostras. No nível da comunidade, os 20% maiores contribuidores são responsáveis por 80% ou mais das contribuições; no nível das torrentes os 20% maiores contribuidores são responsáveis por uma proporção similar das contribuições apenas em uma pequena parcela das torrentes. Em particular, considerando o volume enviado, os principais 20% entre os contribuidores raramente são responsáveis pela maior parte do contribuído. A vazão e o tempo de sementeira são tipicamente mais concentrados, embora

ainda menos que no nível da comunidade.

Concentrando a análise na métrica de volume enviado, é possível ampliar o entendimento sobre como os recursos são providos nas comunidades. A concentração considerável de volume enviado no nível da comunidade em contraste com a concentração mais amena no nível das torrentes sugere que os principais contribuidores não atingem esse status como resultado de grandes contribuições em um pequeno número de torrentes. Em lugar disto, esses usuários contribuem em um grande número de torrentes ao longo do tempo.

### **Dicussão**

Por um lado, entender que uma pequena proporção dos usuários são responsáveis pela maior parte dos dados transferidos motiva as comunidades a valorizar seus principais contribuidores, recompensando-os e prezando sua participação. Em particular, uma proporção de 1% dos usuários contribui particularmente muito em comparação com os demais tanto no BitTorrent quanto no Gnutella, o que sugere uma regularidade na distribuição das contribuições através desses sistemas e reforça a possibilidade de benefícios a partir de incentivos direcionados a essa parcela dos usuários em outros sistemas.

Por outro lado, a relativa equidade de contribuições no nível das torrentes pode informar a alocação de recursos na comunidade. Essa informação é útil quando decidindo, seja de maneira centralizada ou por ação coletiva, em quais torrentes cada usuário deve semear. A ausência de grande concentração nas contribuições no nível das torrentes implica que a quantidade de nós contribuindo em uma torrente é um indicador plausível da qualidade do serviço oferecido naquela torrente. Nossos resultados sugerem que a alocação de recursos no nível da comunidade pode usar essa informação simples e de fácil obtenção para decidir para onde direcionar os recursos dos usuários.

Além disso, a concentração das contribuições dos usuários nos diz algo sobre a robustez das comunidades estudadas. No nível da comunidade, nossos dados mostram que a falha de uma proporção de usuários-chave teria impacto significativo sobre as comunidades, enquanto a ausência da maior parte dos usuários pouco afetaria os demais. Contudo, embora os usuários-chave sejam uma minoria do sistema, a concentração do serviço nesses usuários é significativamente menor que em outros sistemas entre-pares, sugerindo que comunidades BitTorrent são mais robustas que esses sistemas à saída de seus usuários.

Finalmente, a falta de concentração acentuada nas contribuições no nível das torrentes sugere que a falha de usuários tem uma pequena chance de ter um impacto significativo na qualidade de serviço da distribuição de um arquivo.

### **6.3 Como a população de grandes contribuidores se comporta no tempo?**

Os resultados na Seção 6.2 mostram que, no nível da comunidade, uma pequena parcela dos usuários provê a maior parte dos recursos usados no sistema. Nesta seção, nos referimos a esses usuários como os *grandes contribuidores* e analisamos se o conjunto de grandes contribuidores na comunidade é estável ou muda com o tempo. Se os usuários se revezam nesse papel, a comunidade é relativamente mais descentralizada e mais robusta à falha desse tipo de usuário.

Para essa investigação, assim como na Seção 6.2, usamos o registro de período de *bitsoup* e conduzimos a análise no nível da comunidade. Para tanto, dividimos o registro em janelas de tempo consecutivas de mesma duração. Em cada janela, ordenamos os usuários de acordo com o volume de dados que eles enviaram durante a janela, e um usuário é um grande contribuidor nesta janela se ele está entre os 20% de usuários que mais enviaram dados. Em seguida, examinamos como o conjunto de grandes contribuidores muda através das janelas. Para evitar o efeito de possíveis sazonalidades diárias ou semanais no comportamento dos usuários, utilizamos janelas de uma semana de duração.

Os resultados dessa análise mostram que *30% da população de usuários participa do conjunto de grandes contribuidores ao menos uma vez durante a duração do registro, enquanto apenas 1,8% dos usuários se mantém no conjunto durante todo o registro*. Essa observação mostra que uma parcela maior que 20% dos usuários se reveza no papel de grande contribuidor e que apenas uma pequena fração destes se mantém como tal por longos períodos. Note-se, contudo, que a duração de nossos registros não permite que analisemos se usuários deixam de ser grandes contribuidores definitivamente.

**Discussão**

A observação de que o conjunto de grandes contribuidores não é estático ao longo do tempo mostra que *bitsoup* é mais robusta que um sistema que depende de um conjunto estático de usuários para prover a maior parte de seus recursos. Embora o desempenho do sistema dependa em grande parte de uma fração relativamente pequena dos usuários durante um período, existe uma redundância nos usuários que participam dessa fração. Essa redundância implica que a falha ou saída de um grande contribuidor tem menos impacto no desempenho do sistema do que teria se o conjunto de grandes contribuidores fosse estático.

## Capítulo 7

# A Relação entre Demanda e Oferta de Recursos em Comunidades BitTorrent

Além de estudar a demanda e a oferta de recursos separadamente, entender a relação entre esses dois aspectos revela informação importante sobre o funcionamento de um sistema computacional. Este capítulo investiga essa relação em comunidades BitTorrent, se concentrando nas seguintes questões: (i) a demanda de um usuário está relacionada com sua contribuição? (ii) a oferta de recursos no sistema é suficiente para atender à demanda agregada de seus usuários? e (iii) tipicamente, há contenção por recursos em uma torrente? Durante todo este capítulo, usamos para tanto os registros de períodos de *alluvion*, *etree* e *bitsoup*.

### 7.1 Existe uma relação entre demanda e a contribuição de um usuário?

Nossa análise das contribuições de recursos no nível da comunidade mostra que uma minoria dos usuários é responsável pela maior parte dos recursos contribuídos (Seção 6.2). Essa observação levanta outra questão: esses agem como servidores para uma maioria que se comporta como clientes ou eles servem primeiramente uma demanda criada por eles próprios? A análise nessa seção mostra em *bitsoup*, a comunidade onde podemos analisar esse aspecto do comportamento dos usuários, os principais contribuidores não agem como servidores apenas: *os principais contribuidores são também os principais consumidores na comunidade*.

A Figura 7.1 mostra um gráfico de dispersão do volume de dados enviado e recebido

pelos usuários de *bitsoup*. A cor de cada ponto representa a quantidade de torrents em que o usuário participou: mais torrents resultam em pontos mais claros, em escala logarítmica no gradiente de preto a vermelho. A figura mostra que os logaritmos do volume enviado e recebido pelos usuários têm uma forte correlação linear (coeficiente de correlação de Pearson de 0,77), implicando que os usuários que são os principais contribuidores da comunidade são também seus principais consumidores. Além disso, notamos no gráfico que os usuários que são os principais consumidores e provedores são também aqueles que participam em mais torrents.

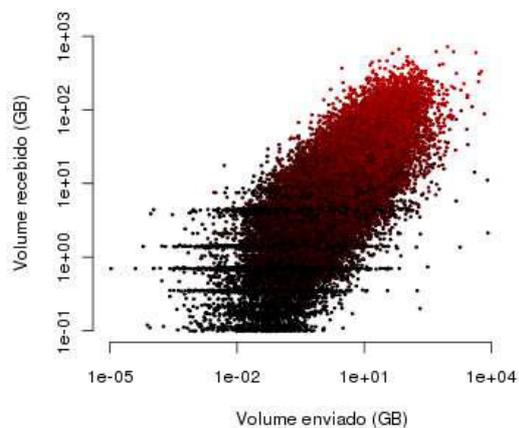


Figura 7.1: Volume enviado e recebido por usuários em *bitsoup*.

## Discussão

Essas observações retratam *bitsoup* como um sistema de compartilhamento equânime. Nossos registros não permitem determinar se essa característica é um resultado do mecanismo de incentivo entre sugadores implementado no BitTorrent ou do mecanismo de sanções aplicado nessa comunidade, mas nossos resultados demonstram a escalabilidade de comunidades adotando ambos os mecanismos. Se as contribuições são proporcionais ao consumo no sistema, então os níveis de contenção por recursos e, por conseguinte, de provisão de recursos, não são afetados pela escala da comunidade. Na ausência de uma correlação como a que observamos em *bitsoup*, o crescimento da população de usuários pode levar a um crescimento equivalente nos níveis de contenção pelos recursos disponíveis, piorando a qualidade do serviço obtido pelos usuários.

Esses resultados também fornecem mais informação sobre o free riding em *bitsoup*, confirmando de forma mais geral que esse comportamento não é a norma: usuários tipicamente consomem uma fração dos recursos do sistema relacionada com sua contribuição.

O entendimento da relação entre o consumo e a contribuição dos usuários permite uma observação mais realista do free riding. As análises de níveis de free riding em Gnutella e eDonkey na literatura (eg. [12; 43; 42]) são normalmente baseadas na hipótese de que todos os usuários consomem do sistema, enquanto apenas uma fração destes contribui durante as medições. No entanto, usuários podem ter diferentes níveis de atividade, incluindo o nível onde eles não consomem recursos do sistema, o que torna os níveis de free riding reportados nesses estudos uma extrapolação pessimista. Nossa análise mostra que no BitTorrent não faz sentido fazer uma consideração semelhante e sugere que trabalhos futuros devem utilizar uma metodologia mais próxima à nossa na análise de free riding em outros sistemas entre-pares.

## 7.2 As contribuições são suficientes para atender à demanda dos usuários?

Embora nossos resultados até aqui mostrem que os usuários em geral contribuem de forma equitativa, não é claro que as contribuições são suficientes para manter um bom nível de qualidade no serviço provido pela comunidade. Esta seção então investiga o quão frequentemente os usuários provêm recursos suficientes para garantir o bom funcionamento<sup>1</sup> das comunidades. Para isso, medimos a proporção das requisições por conteúdo que são servidas em cada torrente e nas comunidades como um todo. Nossa análise mostra que *nas três comunidades, a grande maioria das requisições é corretamente servida*.

Em nossos registros, requisições falham claramente devido à ausência de recursos para servi-la quando uma torrente acaba com um ou mais sugadores e nenhum semeador lhes fornecendo dados. Chamamos de  $s_t$  e  $s_c$  as frações de requisições que não falham dessa forma em uma torrente e em uma comunidade, respectivamente.

A Figura 7.2 (esquerda) mostra a CDF de  $s_t$  para todas as torrentes observadas em nossos

---

<sup>1</sup>Chamamos de *bom funcionamento* a propriedade de *liveness* das comunidades, que estipula que “something good eventually happens during execution” [13].

registros. Considerando apenas torrents completas, os resultados são qualitativamente semelhantes para as três comunidades: a maior parte das torrents serve virtualmente todas as requisições que recebe ( $s_t > 0,99$  para 97% das torrents em *bitsoup* e *etree*, e para 60% das torrents em *alluvion*), enquanto uma pequena fração das torrents falha em servir a maior parte de suas requisições.

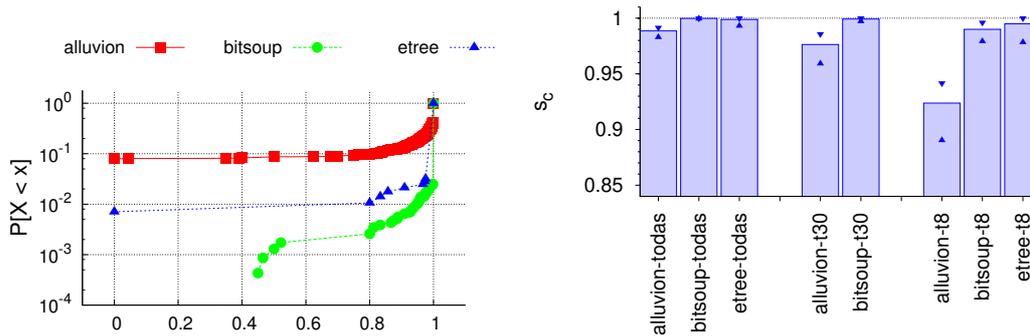


Figura 7.2: Proporção de requisições servidas em cada torrente (esquerda) e no nível das comunidades (direita). Setas indicam os intervalos de 95% de confiança.

Contudo, medir a proporção de requisições servida no nível das torrents não nos dá uma visão precisa da proporção de requisições que são servidas nas comunidades durante um período. De fato, uma análise mais detalhada de  $s_t$  revela que a maioria das torrents que têm baixos valores de  $s_t$  recebe menos de 20 requisições pelo conteúdo que distribui. Essas torrents são uma minoria e servem uma pequena parcela de todos os usuários na comunidade.

Uma perspectiva complementar da análise de  $s_t$  é, portanto, analisar  $s_c$ . A Figura 7.2 (direita) mostra os intervalos de confiança de 95% de  $s_c$  para as três comunidades. Observamos que a proporção de requisições servidas nas três comunidades estudadas é alta: considerando todas as requisições vistas,  $s_c$  não é estatisticamente diferente de 1 para *bitsoup* e *etree*, e é maior que 0,98 para *alluvion*. Notamos, entretanto, que nas amostras de torrents completas,  $s_c$  é significativamente menor em *alluvion* que em *bitsoup* e *etree*.

Além disso, nossos resultados não distinguem estatisticamente  $s_c$  em *bitsoup* e *etree*. Essa constatação indica que o mecanismo de sanções por nível de compartilhamento (SNC) não é necessário para que uma comunidade atinja o alto nível de serviço provido em *bitsoup*.

Note, contudo, que a partir destes resultados, não é possível inferir que o mecanismo de SNC é ineficiente em aumentar o valor de  $s_c$  em uma comunidade, pois *bitsoup* serve significativamente mais requisições que *alluvion*. Entretanto, os níveis de contribuição semelhantes em *etree* e *alluvion* observados no Capítulo 6 e o alto valor de  $s_c$  medido em *etree* sugerem que alguma peculiaridade em *alluvion* é a causa de sua menor proporção de requisições servidas.

Essa conjectura é fortalecida por uma investigação mais aprofundada do registro de *alluvion*. A Figura 7.3 mostra o número de nós em cada observação do registro de *alluvion* que utilizamos, onde é possível notar um influxo extraordinário de nós por volta do dia 30 de outubro. Esse influxo parece estar relacionado à alta proporção de falhas em *alluvion*: removendo o primeiro mês desse registro, a proporção de requisições servidas nas torrentes completas no restante do registro não é significativamente diferente das de *bitsoup* e *etree*. Não é claro como o grande número de nós afeta as falhas e não é possível investigar uma possível causalidade a partir do registro de *alluvion* que obtivemos. Os resultados que temos, contudo, sugerem que o comportamento desse período é atípico na comunidade.

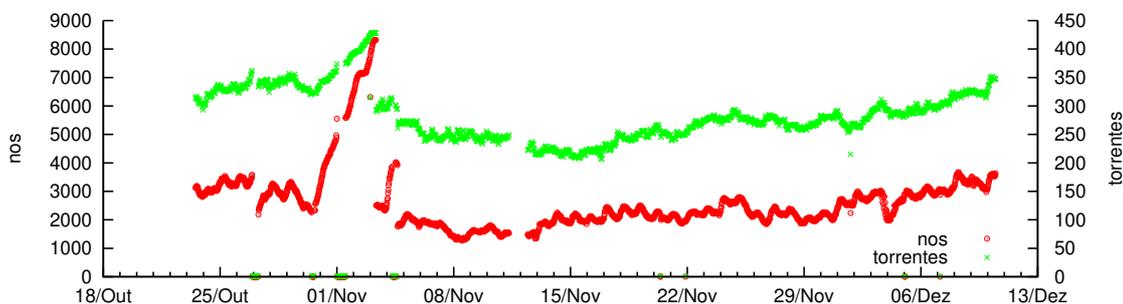


Figura 7.3: Número de nós e torrentes ativas em cada observação do registro de *alluvion*.

Considerando essa observação junto aos resultados anteriores, nossa análise indica que embora o mecanismo de SNC leve a mais sementeira em uma comunidade, ele não é necessário para que comunidades como *etree* e *alluvion* atinjam uma alta qualidade de serviço segundo a métrica que consideramos.

## Discussão

Além de aumentar o entendimento sobre o efeito do SNC nas comunidades, os resultados desta seção também quantificam um aspecto fundamental da qualidade de serviço provida em

comunidades de distribuição de conteúdo entre-pares. Essas comunidades são organizadas de maneira descentralizada e informal como sistemas de produção baseadas em compartilhamento, coordenadas através de relações sociais e necessitam da colaboração de milhares de pessoas distribuídas através do globo, mas ainda assim servem todas ou praticamente todas as requisições que recebem. Nossos resultados provêm evidência quantitativa da eficácia da abordagem baseada no compartilhamento como uma alternativa viável a seus equivalentes no mercado para a distribuição de conteúdo.

Nossa interpretação do nível de serviço provido por comunidades BitTorrent contrasta com as de Guo et al. [41] e Piatek et al. [60]. Guo et al. calcularam a média de  $s_t$  no registro de *alluvion* e interpretaram esse valor (0,9) como um sinal de uma qualidade de serviço tipicamente ruim em torrentes. Observamos que a distribuição de  $s_t$  é bastante concentrada, o que torna sua média uma descrição limitada do comportamento típico das torrentes: 60% das torrentes em *alluvion* servem mais que 99% de suas requisições. Além disso, a qualidade de serviço em *alluvion* difere sensivelmente de outras comunidades e até da mesma comunidade em períodos diferentes. Piatek et al. reportaram que o BitTorrent provê uma qualidade de serviço ruim a seus usuários porque 25% das 50.000 torrentes observadas durante suas medições estavam indisponíveis. A divergência entre a conclusão de Piatek et al. e a nossa é devido a diferentes considerações sobre quando uma torrente deve ser capaz de distribuir dados. Consideramos que o serviço em uma torrente não foi corretamente provido apenas quando há demanda para ele; Piatek et al. não relacionam disponibilidade e demanda.

### 7.3 Há contenção por recursos nas torrentes?

Embora o bom funcionamento do sistema signifique que em algum momento ele provê o serviço esperado, nossa análise até agora não mostra o quão bem a oferta de recursos se adequa à demanda dos usuários. Examinamos agora a relação entre demanda e oferta de recursos no nível das torrentes. Para tanto, investigamos, a partir da contenção por recursos nas torrentes, se estas tipicamente operam sub- ou super-providas de recursos. Em suma, nossa análise mostra que *a contenção de recursos é semelhante nas diferentes comunidades: a maior parte das torrentes opera sob contenção de recursos, enquanto em um quarto das torrentes, não há contenção.*

A contenção e a ausência de contenção em uma torrente expõem desencontros entre a oferta e demanda por recursos. Quando a demanda por recursos é muito maior que a oferta, a velocidade com que os dados são distribuídos fica aquém daquela que a vazão de download coletiva permite. Se a oferta é muito maior que a demanda, os recursos dos provedores são sub-utilizados. Além disso, a oferta e demanda por recursos têm um papel nos incentivos do sistema: priorização, o princípio do mecanismo de incentivo entre sugadores no BitTorrent, só é relevante quando os recursos não são abundantes o suficiente para prover toda a demanda de todos os consumidores no sistema.

Para investigar a contenção nas torrentes, assumimos que o mecanismo de tit-for-tat do BitTorrent é eficaz e que sugadores que provêm mais vazão em uma torrente são priorizados pelos seus pares (Legout et al. [52] apresentam evidência experimental dessa eficácia). Isso implica que quando há contenção por recursos em uma torrente, os usuários que mais contribuem vazão de upload são aqueles que obtêm o arquivo mais rápido. Uma forma de mensurar quão forte é essa relação em uma torrente é medindo o coeficiente de correlação de Kendall entre as velocidades de envio e recebimento de dados dos nós na torrente: quanto maior a correlação, maior a contenção por recursos na torrente. O coeficiente de Kendall é particularmente adequado por ser não-paramétrico. Esse coeficiente mede o grau de concordância entre os postos de listas ordenadas das observações das duas variáveis consideradas, capturando inclusive relações não-lineares entre essas variáveis.

A Figura 7.4 apresenta a CDF do coeficiente de correlação de Kendall entre a velocidade de envio e recebimento de dados dos sugadores em todas as torrentes de *bitsoup* e *alluvion* que têm ao menos cinco nós em nossos registros. A distribuição do grau de correlação por torrente é bastante similar para as duas comunidades. Para a maior parte das torrentes, há ao menos uma correlação pequena ( $\geq 0,3$ ) entre as velocidades de envio e recebimento de dados, indicando contenção. Porém, a correlação é forte ( $\geq 0,6$ ) apenas para uma pequena proporção das torrentes. Além disso, em um quarto das torrentes, os nós provêm vazão suficiente para que todos os consumidores recebam o serviço que demandam, independente de suas contribuições e não há correlação ( $< 0,3$ ).

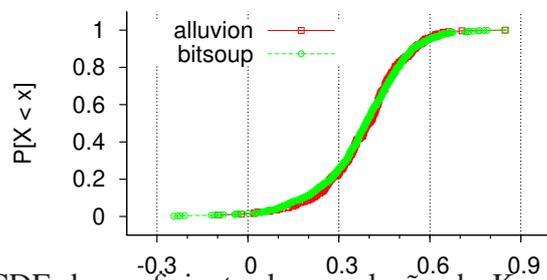


Figura 7.4: CDF do coeficiente de correlação de Kendall entre a velocidade de envio e recebimento de dados dos nós em cada torrente de 730.

### Discussão

A ausência de contenção pelos recursos em um quarto das torrents é particularmente relevante para o projeto de mecanismos de incentivo, pois mecanismos baseados em priorização são irrelevantes nestas torrents. Por outro lado, a existência de contenção na maioria das torrents sugere que o mecanismo de incentivo aos sugadores do BitTorrent é normalmente perceptível para os usuários. Além disso, a correlação e a contenção que ela implica indicam que as contribuições atuais são frequentemente insuficientes para atender a demanda dos nós na distribuição de conteúdo nas comunidades.

É possível entender ainda outra perspectiva da contenção por recursos fazendo uma consideração sobre a conexão de rede dos usuários das comunidades que estudamos. Conexões domésticas são frequentemente assimétricas com relação à largura de banda que fornecem ao usuário para enviar e receber dados. Normalmente, um usuário possui uma largura de banda menor para enviar dados que para recebê-los.

Assumindo que maior parte dos usuários do BitTorrent têm conexões de Internet assimétricas, a demanda dos sugadores só pode ser atendida se há semeadores em uma torrente. Assim, nossos resultados sugerem que (i) os níveis de semeadura que observamos não são suficientes para compensar a assimetria na conexão de Internet dos sugadores; e (ii) as comunidades que observamos poderiam se beneficiar de níveis mais altos de contribuição de vazão ou semeadura na maior parte de suas torrents.

Outro ponto importante em nossos resultados é que os níveis de contenção de recursos não mudam significativamente entre as comunidades. Isso é evidência de que os níveis de

semeadura mais altos observados em *bitsoup* não mudam drasticamente a relação entre oferta e demanda nessa comunidade.

Por fim, o amplo espectro de situações com relação à contenção por recursos nas torrentes que observamos concordam com o experimento de Locher et al. [56]. Os autores examinaram que em algumas torrentes, a velocidade de recebimento de dados de um usuário está relacionada com a velocidade com que ele envia dados, enquanto em outras torrentes, essa relação não existe. Nossos dados quantificam esse fenômeno em uma amostra ampla de torrentes em múltiplas comunidades. Izal et al. [47] observam que há uma correlação positiva entre a velocidade de envio e recebimento de dados em uma torrente de grande escala. Nossas observações indicam como essa correlação varia em comunidades constituídas de diversas torrentes de tamanhos e durações diferentes.

# Capítulo 8

## Conclusão

Neste capítulo, encerramos o documento com uma recapitulação dos principais resultados obtidos (Seção 8.1), uma discussão das implicações de nossos resultados quando considerados em conjunto (Seção 8.2) e sugestões de direções para trabalhos futuros (Seção 8.3).

### 8.1 Principais resultados

Este trabalho contribui para o entendimento da demanda por recursos, do compartilhamento e da relação entre demanda e compartilhamento na distribuição de conteúdo entre-pares. Tal contribuição é realizada através de uma análise mais ampla e profunda que a existente na literatura sobre comunidades BitTorrent. A análise apresentada é mais ampla por considerar diversas comunidades BitTorrent e mais profunda por investigar de forma precisa o comportamento de usuários ao participar de diversas torrentes em uma comunidade e por considerar aspectos da relação entre demanda e oferta de recursos em diversas comunidades.

A seguir, recapitulamos os principais resultados discutidos ao longo deste documento.

#### **Análise da demanda por recursos**

Do ponto de vista de demanda por recursos, os resultados mostram que a distribuição de popularidade do conteúdo não tem cauda longa. Considerando a distribuição das requisições no tempo, nossa caracterização das chegadas dos usuários às torrentes confirma — com uma amostra mais ampla — a constatação de trabalhos anteriores de que a taxa de chegada de usuários em uma torrente decresce rapidamente com o tempo. Além disso, propomos um

modelo para explicar a taxa de chegada de usuários ao longo do tempo que é mais preciso que o estado-da-arte.

### **Análise do compartilhamento de recursos**

Na perspectiva do compartilhamento de recursos, nossa análise resulta no entendimento de diversas características da plataforma de execução de comunidades BitTorrent e da relação entre seu projeto e o comportamento dos usuários. A caracterização mostra que os mecanismos de incentivo utilizados nas comunidades estudadas têm influência na contribuição dos usuários, mas não são os únicos fatores que a determinam. Os níveis de contribuição observados variam de acordo com o tipo do conteúdo compartilhado, evidenciando a influência de fatores sociológicos na contribuição dos usuários. Além disso, embora o mecanismo de incentivo de longo prazo por sanções por nível de compartilhamento resulte em usuários se mantendo por mais tempo como contribuidores, ele não resulta em contribuições significativamente maiores de largura de banda para envio de dados nas comunidades.

Ainda nessa perspectiva, nossos resultados mostram que a contribuição de recursos é concentrada em uma minoria dos usuários para todas as comunidades estudadas, embora a concentração seja moderada em cada torrente. Esse contraste sugere que embora as contribuições agregadas dos usuários sejam heterogêneas, elas são relativamente homogêneas em cada torrente. Finalmente, o conjunto de usuários que são grandes contribuidores não é estático na comunidade em que podemos conduzir essa análise, embora não apresente grande dinâmica ao longo de nossos registros. Essas características apontam para uma robustez moderada das comunidades à falha de seus grandes contribuidores.

### **Análise da relação entre demanda e oferta de recursos**

Estudando a relação entre demanda e oferta de recursos, aprofundamos o conhecimento disponível sobre o comportamento dos usuários, a qualidade de serviço provida por comunidades BitTorrent e o regime em que essas comunidades operam com respeito à provisão de recursos.

Primeiro, nossos resultados mostram que, na comunidade onde podemos analisar o comportamento dos usuários, os grandes contribuidores não são usuários altruístas que servem a demanda dos demais; os usuários que mais contribuem são também aqueles que geram mais

demanda no sistema. Essa observação retrata a comunidade em que estudamos o comportamento dos usuários como um sistema relativamente equânime, contrastando-a com a visão predominante sobre outros sistemas entre-pares, aos quais são atribuídos altos níveis de free riding.

Nossos resultados mostram também que o efeito dos níveis de contribuição observados é que as comunidades servem corretamente a grande maioria das requisições que recebem. Além disso, as comunidades que não usam o mecanismo de sanções por nível de compartilhamento para incentivar a contribuição dos usuários após o fim da obtenção do arquivo atingem níveis equivalentes ao daquela que usa. A quantificação da proporção de requisições corretamente servida contribui para retratar comunidades BitTorrent como serviços confiáveis. Com efeito, essa quantificação é evidência da viabilidade da produção baseada em compartilhamento para a distribuição de conteúdo na Internet.

Examinando as torrentes individualmente, observamos que contenção por recursos é a norma, acontecendo em três quartos das torrentes. Em um quarto das torrentes, por outro lado, há abundância de recursos, o que faz com que os usuários obtenham um serviço que independe da quantidade de recursos contribuída. Esse resultado quantifica a proporção de torrentes em que os incentivos projetados no BitTorrent podem ser eficientes, mostrando que, embora haja espaço para o free riding, esse espaço acontece em uma minoria das torrentes. Por outro lado, nossos resultados indicam que a oferta de recursos nas torrentes geralmente não é suficiente para atender a toda a demanda dos usuários.

## **8.2 Implicações para o projeto de sistemas**

Os resultados de nossa análise corroboram a hipótese inicial deste trabalho, de que seria possível identificar padrões úteis ao projeto de sistemas de distribuição de conteúdo entre-pares a partir de uma caracterização tal qual a que realizamos.

### **Comunidades BitTorrent em perspectiva**

Em um primeiro momento, nossos resultados contribuem para pôr comunidades BitTorrent em perspectiva com relação ao espaço de alternativas para a distribuição de conteúdo na Internet. Nossa caracterização traz evidências de que, embora operem tipicamente com

contenção pelos recursos disponíveis, essas comunidades são alternativas confiáveis para a distribuição de conteúdo. Além disso, na comunidade que podemos analisar em mais detalhe e que usa o mecanismo de sanções por nível de compartilhamento, há equidade na participação dos usuários: aqueles que contribuem mais recursos são aqueles que mais recebem do sistema. Em conjunto, essas observações são subsídios para as decisões de projetistas, operadores e usuários de mecanismos de distribuição de conteúdo, informando sobre a eficácia e eficiência de um dos mecanismos atualmente mais utilizados para a distribuição de conteúdo.

### **O projeto do BitTorrent em contexto**

Num segundo momento, este trabalho tem implicações ao contrastar a relação entre carga, compartilhamento e o projeto e implementação das comunidades BitTorrent com os mesmos aspectos de outros sistemas de compartilhamento. A partir da observação de que as distribuições de popularidade de conteúdo nas comunidades estudadas não têm cauda longa, constatamos que o *trade-off* no dimensionamento de caches é qualitativamente diferente no BitTorrent quando comparado à Web, ao Kazaa e ao Gnutella, três referenciais anteriores. Nossos resultados sugerem que a abordagem utilizada em outros sistemas de buscar um tamanho de cache que, mantendo cópias de um pequeno número de itens muito populares, abarcará uma proporção significativa do tráfego no sistema, terá eficácia limitada no BitTorrent. Por outro lado, quando viável, um cache que busque manter cópias de um grande número de itens tem o potencial de abarcar proporções maiores do tráfego do sistema que na Web, no Kazaa ou no Gnutella.

Na mesma direção, nosso estudo permite o entendimento da eficácia de várias das decisões de projeto do BitTorrent, em particular com relação a outros sistemas entre-pares. O projeto do BitTorrent objetiva promover o compartilhamento entre seus usuários. Nossos resultados mostram que esse projeto, em conjunto com aspectos da implementação do sistema, tem sucesso em reduzir a quantidade de usuários que não contribuem e que contribuem pouco para o sistema. Contudo, nossa análise relativiza o papel dos mecanismos de incentivo usados em comunidades BitTorrent. Primeiro, as características sociais de algumas das comunidades que estudamos têm efeito semelhante ao do mecanismos de sanções por nível de compartilhamento para promover a semadura na comunidade. Segundo, comunidades

que compartilham conteúdo legal e são relativamente coesas atingem virtualmente a mesma proporção de requisições servidas que uma comunidade que utiliza o mecanismo de sanções. Juntas, essas duas observações motivam a consideração de aspectos sociais do sistema de compartilhamento como componente central na promoção de contribuições. Embora não seja possível reduzir essa recomendação a uma receita, dada a variabilidade entre múltiplos fatores nas comunidades e as simplificações de nossa análise de seus aspectos sociológicos, nossos resultados sugerem que a promoção de uma identidade da comunidade e da noção de que é seguro contribuir podem beneficiar o compartilhamento. Note-se que, embora fosse esperado no início deste trabalho que houvesse alguma relação entre as características das comunidades e o comportamento de seus usuários, não era possível prever ou quantificar essa relação ou sua influência quando comparada a mecanismos de incentivo. Além disso, até onde pudemos determinar, essa relação não havia sido documentada em nenhum outro sistema de compartilhamento de recursos entre-pares.

### **Projeto de melhorias para comunidades de distribuição de conteúdo entre-pares**

A terceira perspectiva em que nossa análise tem implicações é no desenvolvimento futuro do BitTorrent e de sistemas semelhantes a ele. Primeiro, o modelo proposto para o decaimento da taxa de chegadas de usuários às torrentes implica em mudanças na avaliação de desempenho do BitTorrent e de suas extensões. Nossos resultados sugerem que trabalhos futuros devem considerar nosso modelo em complemento ou substituição ao previamente proposto.

Além deste ponto, a nossa análise tem duas outras implicações a partir da observação que três quartos das torrentes operam com contenção por recursos enquanto um quarto opera super-provida. A primeira é a quantificação da quantidade de conteúdo em cuja distribuição o mecanismo de incentivo implementado no BitTorrent é eficaz. Nossos resultados dão subsídios para que projetistas e operadores decidam a necessidade de modificações nos mecanismos de incentivo do BitTorrent à luz de medições. A segunda implicação a partir dessa figura é a de que há um potencial para a realocação de recursos em comunidades BitTorrent. Embora não seja possível quantificar o benefício potencial a partir de nossos resultados, estes sugerem que recursos poderiam ser movidos de torrentes super-providas para aquelas sub-providas.

Finalmente, nossos resultados sobre a homogeneidade de contribuições nas torrentes

sugere que mecanismos de alocação de recursos em comunidades BitTorrent podem ser desenvolvidos a partir de informação relativamente simples e de fácil obtenção. Como a contribuição dos nós tipicamente independe de seu comportamento passado ou de outras características dos nós, é possível criar mecanismos que fazem uma alocação precisa usando apenas informações sobre a quantidade de provedores e consumidores de serviço em uma torrente.

### **8.3 Trabalhos Futuros**

Nossos resultados motivam diversas direções para trabalhos futuros:

1. *Avaliação quantitativa de estratégias de caching para a distribuição de conteúdo entre-pares.* Nossos resultados motivam uma análise quantitativa da eficiência de caching para a carga que observamos, com ênfase na avaliação da relação entre o tamanho do cache e a eficiência atingida, a fim de contrastá-la com a relação entre esses fatores observada na Web.
2. *Estudo quantitativo da eficiência de uma alocação de semeadores com base no nível de semeadura das torrentes.* Nossa análise motiva uma investigação quantitativa de métodos de alocação de semeadores nas comunidades que considerem apenas a quantidade de semeadores presentes nas torrentes. Usando informação possível de ser obtida de forma descentralizada e de simples obtenção, esses mecanismos podem ser mais eficientes e escaláveis do que mecanismos que necessitem de informação cuja obtenção é mais complexa.
3. *Investigação de quão geral é o fenômeno do rápido decaimento do interesse dos usuários por um arquivo na distribuição de conteúdo.* A dinâmica da demanda pelo conteúdo de um sistema define os requisitos de disponibilidade do conteúdo. Se outros sistemas apresentam uma dinâmica semelhante à que observamos, nosso modelo pode ser útil para generalizar os fatores que influenciam essa dinâmica.
4. *Avaliação da contribuição dos usuários em relação à sua demanda em outros sistemas entre-pares.* Para uma comparação mais adequada dos níveis de contribuição observados nas comunidades BitTorrent com os de outros sistemas entre-pares, é necessário

que estes sejam analisados com uma abordagem semelhante à nossa. As atuais análises de contribuições em outros sistemas entre-pares se baseiam na consideração de que todos os usuários consomem recursos do sistema. Nossos resultados, por outro lado, indicam que os níveis de demanda dos usuários variam drasticamente e devem ser considerados junto às contribuições dos usuários.

5. *Projeto de aspectos sociológicos de comunidades de compartilhamento de recursos.* Nossa análise traz evidência quantitativa da influência das características de diferentes comunidades construídas em torno da publicação e distribuição de conteúdo no comportamento de seus usuários. O projeto de comunidades futuras pode se beneficiar de estudos que analisem em mais detalhe as práticas que têm sucesso na promoção de contribuição nas comunidades e avaliem sua reproducibilidade.
6. *Análise comparativa da qualidade de serviço das comunidades BitTorrent com suas alternativas baseadas em mercado.* A avaliação da proporção de requisições corretamente servidas nas comunidades que examinamos oferecem um referencial para a comparação do serviço provido através do compartilhamento com suas alternativas comerciais. Essa comparação, tanto pela métrica que consideramos quanto por outras métricas pode contribuir para colocar em perspectiva a qualidade do serviço possível com os sistemas de compartilhamento de que dispomos hoje.
7. *Caracterização de comunidades BitTorrent a partir de registros mais amplos.* Embora nossa caracterização tenha usado registros de até 68 dias e diversas comunidades, ela provê uma visão limitada do comportamento dos usuários no BitTorrent. Em particular, não pudemos analisar torrentes mais longas que 30 dias e investigamos apenas uma comunidade com identificação de usuários precisa. Caracterizar mais comunidades de diferentes características, tanto nas dimensões que consideramos quanto em outras, ainda é necessário para que compreendamos mais completamente como o fator humano molda a distribuição de conteúdo na Internet e como devemos projetar mecanismos que melhor nos sirvam.

# Bibliografia

- [1] Dijjer, <http://dijjer.org/>. Acesso em julho de 2007.
- [2] eMule official website, <http://www.emule-project.net/>. Acesso em outubro de 2008.
- [3] Gnutella, <http://www.gnutella.com>. Acesso em novembro de 2001.
- [4] Joost: <http://www.joost.com/>. Acesso em outubro de 2008.
- [5] KaZaA, <http://www.kazaa.com/>. Acesso em Novembro de 2003.
- [6] OurGrid, open-source software for peer-to-peer grid computing, <http://www.ourgrid.org/>. Acesso em outubro de 2008.
- [7] PlanetLab, an open platform for developing, deploying and accessing planetary-scale services, <http://www.planet-lab.org/>. Acesso em outubro de 2008.
- [8] SETI@home, <http://setiathome.ssl.berkeley.edu/>. Acesso em outubro de 2008.
- [9] Skype: <http://www.skype.com/>. Acesso em outubro de 2008.
- [10] Wikipedia, the free encyclopedia: <http://www.wikipedia.org>. Acesso em outubro de 2008.
- [11] The apache http server project, <http://httpd.apache.org/>, outubro 2008. Acesso em outubro de 2008.
- [12] E. Adar and B. A. Huberman. Free Riding on Gnutella. *First Monday*, 5(10), outubro 2000.
- [13] Bowen Alpern and Fred B. Schneider. Defining liveness. Relatório Técnico TR85-650, Cornell University, Ithaca, EUA, 1984.

- [14] Kostas Anagnostakis, Fotios Harmantzis, Sotiris Ioannidis, and Manaf Zghaibeh. On the Impact of Practical P2P Incentive Mechanisms on User Behavior. Working Paper 06-14, NET Institute, setembro 2006.
- [15] Nazareno Andrade, Francisco Brasileiro, Walfredo Cirne, and Miranda Mowbray. Automatic grid assembly by promoting collaboration in peer-to-peer grids. *Journal of Parallel and Distributed Computing*, 67(8):957–966, 2007.
- [16] Nazareno Andrade, Francisco Brasileiro, Miranda Mowbray, and Walfredo Cirne. *Market Oriented Grid and Utility Computing*, capítulo A Reciprocation-based Economy for Multiple Services in a Computational Grid. Wiley Press, 2009 (aceito).
- [17] Nazareno Andrade, Miranda Mowbray, Walfredo Cirne, and F. Brasileiro. When can an autonomous reputation scheme discourage free-riding in a peer-to-peer system? In *Proceedings of the 2004 IEEE International Symposium on Cluster Computing and the Grid (CCGRID '04)*, p. 440–448, Washington, DC, EUA, abril 2004.
- [18] Nazareno Andrade, Miranda Mowbray, Aliandro Lima, Gustavo Wagner, and Matei Ripeanu. Influences on cooperation in bittorrent communities. In *Proceeding of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems (P2PECON '05)*, p. 111–115, New York, NY, EUA, 2005.
- [19] Nazareno Andrade, Elizeu Santos-Neto, Francisco Brasileiro, and Matei Ripeanu. Resource demand and supply in bittorrent communities. *Computer Networks Journal*, dezembro 2008. (aceito).
- [20] Robert Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.
- [21] Anthony Bellissimo, Brian N. Levine, and Prashant Shenoy. Exploring the use of BitTorrent as the basis for a large trace repository. Relatório Técnico 04-41, University of Massachusetts, junho 2004.
- [22] Y. Benkler. Coase's penguin, or, linux and the nature of the firm. *Yale Law Journal*, (112):369–446, 2002.
- [23] Yochai Benkler. Sharing nicely: On shareable goods and the emergence of sharing as a modality of economic production. *The Yale Law Journal*, 114:273–358, 2004.

- [24] Yochai Benkler. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, New Haven, CT, EUA, 2006.
- [25] Ashwin R. Bharambe, Cormac Herley, and Venkata N. Padmanabhan. Analyzing and improving a bittorrent networks performance mechanisms. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, Barcelona,, abril 2006.
- [26] Lee Breslau, Pei Cao, Li Fan, Graham Phillips, and Scott Shenker. Web Caching and Zipf-like Distributions: Evidence and Implications. In *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, 1999.
- [27] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(1):309–320, junho 2000.
- [28] Kenneth Burnham and David Anderson. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2):261–304, 2004.
- [29] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world’s largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (IMC '07)*, p. 1–14, San Diego, California, EUA, 2007. ACM.
- [30] Alix L.H. Chow, Leana Golubchik, and Vishal Misra. Improving BitTorrent: A simple approach. In *Proceedings of IPTPS*, 2008.
- [31] W. Cirne, F. Brasileiro, N. Andrade, L. Costa, A. Andrade, R. Novaes, and M. Mowbray. Labs of the world, unite!!! *Journal of Grid Computing*, 4(3):225–246, 2006.
- [32] B. Cohen. Incentives Build Robustness in BitTorrent. In *Proc. Workshop on Economics of Peer-to-Peer Systems*, Berkeley CA, EUA, junho 2003.
- [33] Landon P. Cox and Brian D. Noble. Samsara: Honor among Thieves in Peer-to-Peer Storage. In *Proc. 19th ACM Symp. Operating Systems Principles*, Bolton Landing NY, EUA, outubro 2003.

- [34] E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, and P. Samaranti. Managing and Sharing Servents' Reputations in P2P Systems. *IEEE Trans. Data and Knowledge Engineering*, 15(4):840–854, Julho/agosto 2003.
- [35] Ellacoya. Ellacoya data shows web traffic overtakes peer-to-peer (p2p) as largest percentage of bandwidth on the network. Technical report, [www.ellacoya.com/news/pdf/2007/NXTcommEllacoyaMediaAlert.pdf](http://www.ellacoya.com/news/pdf/2007/NXTcommEllacoyaMediaAlert.pdf), 2007.
- [36] Michal Feldman, Kevin Lai, John Chuang, and Ion Stoica. Quantifying Disincentives in Peer-to-Peer Networks. In *Proc. First Workshop on Economics of Peer-to-Peer Systems*, junho 2003.
- [37] Michal Feldman, Kevin Lai, Ion Stoica, and John Chuang. Robust incentive techniques for peer-to-peer networks. In *Proceedings of the 5th ACM conference on Electronic commerce*, p. 102–111, New York, NY, EUA, maio 2004. ACM Press.
- [38] Christos Gkantsidis and Pablo Rodriguez. Network coding for large scale content distribution. In *INFOCOM*, p. 2235–2245. IEEE, 2005.
- [39] Steven Glassman. A caching relay for the world wide web. *Comput. Netw. ISDN Syst.*, 27(2):165–173, 1994.
- [40] Krishna P. Gummadi, Richard J. Dunn, Stefan Saroiu, Steven D. Gribble, Henry M. Levy, and John Zahorjan. Measurement, modeling, and analysis of a peer-to-peer file-sharing workload. In *Proc. Nineteenth ACM symposium on Operating systems principles (SOSP)*, p. 314–329, 2003.
- [41] Lei Guo, Songqing Chen, Zhen Xiao, Enhua Tan, Xiaoning Ding, and Xiaodong Zhang. Measurements, analysis, and modeling of bittorrent-like systems. In *Proc. ACM SIGCOMM/USENIX Internet Measurement Conference*, p. 19–21, outubro 2005.
- [42] S. B. Handurukande, A.-M. Kermarrec, F. Le Fessant, L. Massoulié, and S. Patarin. Peer sharing behaviour in the edonkey network, and implications for the design of server-less file sharing systems. In *Proc. EuroSys*, p. 359–371, New York, NY, 2006.
- [43] D. Hughes, G. Coulson, and J Walkerdine. Freeriding on Gnutella Revisited: the Bell Tolls. *IEEE Distributed Systems Online*, 6(6), fevereiro 2005.

- [44] Adriana Iamnitchi and Ian Foster. Interest-aware information dissemination in small-world communities. In *Proceedings of HPDC*, 2005.
- [45] Adriana Iamnitchi, Matei Ripeanu, and Ian T. Foster. Small-world file-sharing communities. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, 2004.
- [46] Ipoque. Internet study 2007. Technical report, [www.ipoque.de](http://www.ipoque.de), 2007. <http://www.ipoque.com/resources/internet-studies/internet-study-2007>.
- [47] M. Izal, G. Urvoy-Keller, E. W. Biersack, P. Felber, A. Al Hamra, and L. Garcés-Erice. Dissecting BitTorrent: Five Months in a Torrent's Lifetime. In *Proc. Passive and Active Measurements*, abril 2004.
- [48] Seung Jun and Mustaque Ahamad. Incentives in BitTorrent induce free riding. In *Proceedings of the ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems (P2PECON)*, p. 116–121, New York, NY, EUA, agosto 2005.
- [49] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The EigenTrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th international conference on World Wide Web (WWW)*, p. 640–651, New York, NY, EUA, maio 2003. ACM Press.
- [50] Marlom A. Konrath, Marinho P. Barcellos, Juliano F. Silva, Luciano P. Gasparly, and Rafael Dreher. Atacando um exame com um bando de mentirosos: vulnerabilidades em bittorrent. In *XXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2007)*, volume 2, p. 883–896, maio 2007.
- [51] K. Lai, M. Feldman, I. Stoica, and J. Chuang. Incentives for Cooperation in Peer-to-Peer Networks. In *Proc. Workshop on Economics of Peer-to-Peer Systems*, Berkeley, California, EUA, junho 2003.
- [52] Arnaud Legout, Nikitas Liogkas, Eddie Kohler, and Lixia Zhang. Clustering and sharing incentives in BitTorrent systems. *SIGMETRICS Perform. Eval. Rev.*, 35(1):301–312, 2007.

- [53] Nathaniel Leibowitz, Matei Ripeanu, and Adam Wierzbicki. Deconstructing the Kazaa network. In *Proceedings of the 3rd IEEE Workshop on Internet Applications (WIAPP '03)*, page 112, San Jose, CA, EUA, junho 2003.
- [54] Dave Levin, Rob Sherwood, and Bobby Bhattacharje. Fair file swarming with FOX. In *Proc. International Workshop on Peer-to-Peer Systems (IPTPS)*, 2006.
- [55] N. Liogkas, R. Nelson, E. Kohler, , and L. Zhang. Exploiting Bittorrent for fun (but not profit). In *Proc. IPTPS*, fevereiro 2006.
- [56] Thomas Locher, Patrick Moor, Stefan Schmid, and Roger Wattenhofer. Free Riding in BitTorrent is Cheap. In *Proceedings of HotNets*, novembro 2006.
- [57] Gary Warren Melton. An examination of the bootleg record industry and its impact upon popular music consumption. *Tracking*, 4(1), Winter 1991.
- [58] T. Mennecke. The latest Kazaa Hack – version 2.5. Slyck News, from <http://www.slyck.com/news.php?story=217>, agosto 2003.
- [59] J.J.D. Mol, J.A. Pouwelse, D.H.J. Epema, and H.J. Sips. Free-riding, fairness, and firewalls in p2p file-sharing. In K. Wehrle, W. Kellerer, S.K. Singhal, and R. Steinmetz, editors, *8-th IEEE International Conference on Peer-to-Peer Computing*, p. 301–310. IEEE Computer Society, setembro 2008.
- [60] Michael Piatek, Tomas Isdal, Arvind Krishnamurthy, and Thomas Anderson. One hop reputations for peer to peer file sharing workloads. In *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation*, 2008.
- [61] Francesca Lo Piccolo, Giovanni Neglia, and Giuseppe Bianchi. The effect of heterogeneous link capacities in bittorrent-like file sharing systems. In *Proceedings of the 2004 International Workshop on Hot Topics in Peer-to-Peer Systems (HOT-P2P)*, p. 40–47, Washington, DC, EUA, 2004. IEEE Computer Society.
- [62] J A Powelse, P Garbacki, D H J Epema, and H J Sips. Measurement study of the bittorrent peer-to-peer file-sharing system. Relatorio Tecnico PDS-2004-003, Delft U. Technology, abril 2004.

- [63] Dongyu Qiu and Rayadurgam Srikant. Modeling and performance analysis of bittorrent-like peer-to-peer networks. In *Proceedings of ACM SIGCOMM*, p. 367–378, agosto 2004.
- [64] Vivek Rai, Swaminathan Sivasubramanian, Sandjai Bhulai, Pawel Garbacki, and Marten van Steen. A multiphased approach for modeling and analysis of the bittorrent protocol. In *Proceedings of ICDCS*, p. 25–27, 2007.
- [65] Matei Ripeanu, Adriana Iamnitchi, Ian Foster, and Anne Rogers. In search of simplicity: A self-organizing group communication overlay. In *Proceedings of SASO*, p. 371–374, 2007.
- [66] Matei Ripeanu, Miranda Mowbray, Nazareno Andrade, and Aliandro Lima. Gifting technologies: A BitTorrent case study. *First Monday*, 11(11), novembro 2006.
- [67] Bruno Gusmão Rocha, Virgilio Almeida, and Dorgival Olavo Guedes Neto. Increasing qos in selfish overlay networks. *IEEE Internet Computing*, 10(3):24–31, 2006.
- [68] Drew Roselli, Jacob R. Lorch, and Thomas E. Anderson. A comparison of file system workloads. In *USENIX Annual Technical Conference*, Berkeley, CA, EUA, 2000.
- [69] Osama Saleh and Mohamed Hefeeda. Modeling and caching of peer-to-peer traffic. In *ICNP '06: Proceedings of the Proceedings of the 2006 IEEE International Conference on Network Protocols*, p. 249–258, Washington, DC, EUA, 2006. IEEE Computer Society.
- [70] S. Saroiu, P. K. Gummadi, and S. D. Gribble. A Measurement Study of Peer-to-Peer File Sharing Systems. In *Proc. Multimedia Computing and Networking 2002*, San Jose CA, EUA, Jan. 2002.
- [71] Imre Simon and Miguel Said Vieira. O rossio não-rival. Disponível em [http://www.ime.usp.br/is/papir/RNR\\_v9.pdf](http://www.ime.usp.br/is/papir/RNR_v9.pdf), 2008. Acesso em outubro de 2008.
- [72] Sitabhra Sinha and Raj Kumar Pan. *Econophysics and Sociophysics: Trends and Perspectives*, capítulo How a "Hit" is Born: The Emergence of Popularity from the Dynamics of Collective Choice, p. 417–447. Wiley-VCH, 2006.

- [73] Kunwadee Sripanidkulchai, Bruce M. Maggs, and Hui Zhang. Efficient content location using interest-based locality in peer-to-peer systems. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, 2003.
- [74] Daniel Stutzbach and Reza Rejaie. Understanding churn in peer-to-peer networks. In *Proceedings of the 6th ACM SIGCOMM Internet Measurement Conference*, p. 189–202, New York, NY, EUA, 2006. ACM Press.
- [75] Daniel Stutzbach, Daniel Zappala, and Reza Rejaie. The scalability of swarming peer-to-peer content delivery. In *Proceedings of NETWORKING*, p. 15–26, 2005.
- [76] Vivek Vishnumurthy, Sangeeth Chandrakumar, and Emin Gün Sirer. KARMA: A secure economic framework for peer-to-peer resource sharing. In *Proc. Workshop on Economics of Peer-to-Peer Systems*, Berkeley, California, EUA, junho 2003.
- [77] Chris Werry and Miranda Mowbray. *Online Communities: Commerce, Community Action, and the Virtual University*. Prentice Hall PTR, dezembro 2000.
- [78] Adam Wierzbicki, Nathaniel Leibowitz, Matei Ripeanu, and R. Wozniak. Cache replacement policies for peer-to-peer file-sharing protocols. *European Transactions on Telecommunications*, 15(6):559 – 569, 2004.
- [79] Dennis M. Wilkinson. Strong regularities in online peer production. In *EC '08: Proceedings of the 9th ACM conference on Electronic commerce*, p. 302–309, New York, NY, EUA, 2008. ACM.
- [80] M Yang, Z Zhang, Xiaoming Li, and Yafei Dai. An empirical study of free-riding behavior in the maze p2p file-sharing system. In *Proceedings of IPTPS '05*, fevereiro 2005.
- [81] Xiangying Yang and Gustavo de Veciana. Service capacity of peer to peer networks. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, 2004.
- [82] Shanyu Zhao, Daniel Stutzbach, and Reza Rejaie. Characterizing files in the modern gnutella network: a measurement study. In *Proc. of SPIE/ACM Multimedia Computing and Networking*, San Jose, CA, EUA, janeiro 2006.