

# Measuring the Internet Technical Efficiency: A Ranking for the World Wide Web Pages

K. Nepomuceno, T. Nepomuceno, and D. Sadok

**Abstract**—The Internet is one of the most important innovations of humanity and has brought many of our social and cultural changes. The web pages have great importance because they are the opening door for which users can interact directly with each other and interact with corporations. A bad performance may result in a significant loss from millions of users. This work aims to evaluate the performance of the Internet web pages through Data Envelopment Analysis (DEA), and provide a list with 82 from the 100 most-accessed addresses in the world ranked by their page load time (output) projection on the size and number of objects (inputs). The decomposition of the technical efficiency is made from an inverse frontier perspective; this way, outputs are contracted and inputs expanded. As a result, the pages reaching the efficiency frontier are explored as benchmarks for better practices. The slacks for the page loading time, size and objects are provided for the inefficient units. We discuss the results and the interesting prospects on the Computing and Web-Communication.

**Index Terms**—Data envelopment analysis, Page load time, Web.

## I. INTRODUÇÃO

DURANTE toda história da humanidade, os meios de comunicação tiveram um papel importante na evolução social. Desde a escrita até a Internet, as tecnologias de comunicação foram evoluindo e se aperfeiçoando cada vez mais. Atualmente temos uma variedade de ferramentas e aparatos tecnológicos que facilitam e melhoram o cotidiano das pessoas. A ITU (*International Telecommunication Union*), estimou que cerca de metade da população do mundo estaria *online* ainda do ano de 2015 [1]. A Internet passou a fazer parte da rotina das pessoas e empresas, mudando hábitos sociais e culturais já existentes.

O presente trabalho teve o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Submetido para apreciação da revista científica IEEE Latin America Transactions em 10/03/2019.

K. T. C. Nepomuceno é pesquisadora do Grupo Pesquisa em Redes e Telecomunicações (GPRT) da Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil, [kten@cin.ufpe.br](mailto:kten@cin.ufpe.br).

T. C. C. Nepomuceno é professor adjunto no Departamento de Engenharia de Produção da Universidade Federal de Pernambuco e pesquisador afiliado ao Departamento de Engenharia informática automática e gestionale Antonio Ruberti (DIAG) da Sapienza University of Rome, Roma, Lazio, Italia, [nepomuceno@diag.uniroma1.it](mailto:nepomuceno@diag.uniroma1.it).

D. F. H. Sadok é professor adjunto no Centro de Informática (CIN) da Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil, [jamel@cin.ufpe.br](mailto:jamel@cin.ufpe.br).

*Corresponding author: Késsia Thais C. Nepomuceno.*

A perspectiva é que esse número aumente com o passar do tempo. Mas para manter toda a estrutura da Internet com um número tão expressivo de usuários é necessário que os protocolos de comunicação, e as outras tecnologias fundamentais para o funcionamento da Internet, continuem a dar suporte de forma efetiva. Caso contrário, a Internet poderá ficar sobrecarregada de dados e também número de usuários.

Uma das atividades cruciais na Internet é a navegação em páginas Web. Seus principais protocolos de comunicação são o TCP e o HTTP. O correto funcionamento e performance dessas páginas e também dos protocolos são de suma importância uma vez que o desempenho das páginas impactam diretamente a experiência do usuário. Desde sua criação, as páginas vêm dando suporte a Internet de forma satisfatória. Entretanto, com o grande crescimento de dados e número de usuários, é possível que seus recursos não sejam suficientes para atender a alta demanda atual. Vendo isso, a comunidade científica, empresas e pesquisadores, averigam o desempenho dos recursos existentes nas páginas Web e criam novas tecnologias com o intuito que essas tecnologias deem suporte ao crescente número de usuários.

Ainda no contexto da Internet, o tempo requerido para carregar uma página é um dos principais fatores para a satisfação e experiência final do usuário. O tempo de carregamento de uma página (*page load time* - PLT) influencia no tempo que o usuário permanecerá no site. Usuários que estão frustrados com o baixo tempo de carregamento do *Website* tendem a não acessá-lo novamente [2]. Além disso, o tempo de carregamento da página também influencia na recomendação do usuário sobre o site [3]. Alguns sites de busca, como o Google, utilizam o PLT em seus algoritmos para exibir os resultados de buscas [4], direcionando os usuários para sites com menor tempo de carregamento. Por esses e outros motivos o tempo de carregamento da página tem uma grande importância na Internet, porém esse tempo tende a aumentar cada vez mais devido ao crescimento dos dados. Um PLT alto tem efeito negativo na navegação Web e pode resultar usuários insatisfeitos.

Esse trabalho tem como objetivo avaliar a performance das páginas Web através da Análise Envoltória de Dados (*Data Envelopment Analysis* – DEA). A Análise Envoltória de Dados é uma técnica de programação matemática não-paramétrica (não exige a predefinição de uma forma funcional) desenvolvida por Charnes, Cooper e Rhodes [5] que vem crescendo exponencialmente com milhares de contribuições metodológicas, aplicações empíricas e desenvolvimentos computacionais [6], [7], [8]. Baseado na

determinação de um conjunto de possibilidades de produção em que unidades tomadoras de decisão (*Decision Maker Units* – DMUs) são comparadas entre si, essa técnica constrói uma fronteira de eficiência não paramétrica por partes envolvendo os dados a fim de avaliar a eficiência das DMUs como uma razão entre os recursos usados e os resultados obtidos [9]. Uma unidade tecnicamente ineficiente pode se tornar eficiente ao aumentar seus resultados (output) mantendo o mesmo uso dos insumos, ou reduzindo os insumos da produção mantendo os mesmos resultados, ou ambos (considerando a existência de múltiplos *inputs* e *outputs*).

A avaliação da comunicação em rede tem sido um tema importante e recorrente na literatura [10], [11], [12], [13]. Muitas aplicações e avaliações de eficiência com o suporte DEA têm percorrido os setores de energia [14], [15], [16], serviços [17], [18] e comunicação [19], [20]. A avaliação de sites, páginas e *webservices* tem sido um ponto de interesse na literatura com aplicações que sugerem a construção de sistemas de avaliação com base no *ranking* da Alexa e experiência de usuários [21] redução no tempo de carregamento das páginas [22], [23], [24] e caracterização do impacto das páginas em função dos objetos, imagens e servidores [25]. Entretanto, ao melhor de nosso conhecimento, existe uma carência em avaliações da eficiência de páginas Web apesar de sua relevante importância. Duas instâncias singulares do uso da DEA em páginas web podem ser consultadas na avaliação das *webpages* das universidades na Jordania baseados nas perspectivas de usabilidade, design e desempenho [26] e da eficiência operacional de páginas para o *e-commerce* com foco na qualidade do serviço [27].

O objetivo deste trabalho é utilizar essa técnica para avaliação da eficiência em páginas Web pela razão entre os recursos utilizados e resultados obtidos. Diferentemente as avaliações citadas, esse trabalho mensura a eficiência técnica das páginas Web com foco em seu PLT, fornecendo um ranque de acordo com suas medidas de eficiência técnica. O tamanho das páginas e quantidade de objetos são tomados como *inputs* de produção, e o tempo de carregamento (PLT) como *output*, portanto, quanto maior o tamanho da página ou a quantidade de objetos (*inputs*), maior será o tempo de carregamento da página Web (*output*). Nesse modelo, as páginas consideradas 100% eficientes serão aquelas que possuem o menor PLT proporcionalmente aos recursos utilizados (quantidade de objetos e tamanho), dada sua razão entre *inputs* e *outputs*. As demais páginas se referenciarão a estas para avaliar seu desempenho. Os resultados apontam as páginas com melhor desempenho. Também são apresentadas as páginas que demandam melhorias para alcançar a fronteira de eficiência e onde podemos melhorá-las (folgas).

Este trabalho está organizado da seguinte forma: A seção 2 descreve a metodologia teórica da Análise Envoltória de Dados ajustada ao contexto e a metodologia utilizada na condução dos experimentos, ou seja, é descrito o cenário de avaliação, as ferramentas utilizadas e como os experimentos foram realizados. São apresentados também os fatores e métricas utilizadas; a seção 3 apresenta e discute os resultados obtidos nos experimentos; e por fim, a seção 4 encerra o trabalho apresentando as considerações finais e sugestões de trabalhos futuros.

## II. METODOLOGIA E DADOS

A obtenção e a análise dos dados experimentais foram conduzidas com base em uma metodologia composta por quatro atividades principais: a) planejamento das medições, definindo como e quais métricas serão utilizadas; b) realização das medições em ambiente controlado, avaliando a configuração e funcionamento das páginas Web; c) aplicação da Análise Envoltória de Dados e tratamentos estatísticos a fim de fornecer uma maior confiabilidade nos resultados; e d) realizar a análise e discussão dos resultados dos experimentos.

Para a execução dos experimentos, o ambiente foi configurado de forma que correspondesse o máximo possível à Internet de hoje. Por meio de pesquisas, foram estabelecidas várias configurações que influenciam e representam a Internet do mundo real. Podem-se ver essas configurações na Subseção B. Nos experimentos, foram utilizadas ferramentas confiáveis e consolidadas na comunidade de software. A metodologia DEA adaptada, a configuração dos experimentos bem como as métricas e fatores utilizados são detalhados nas próximas subseções.

### A. Análise Envoltória de Dados

O conceito tradicional de eficiência técnica segundo Shephard [28], Färe e Lovell [29], Charnes e associados está relacionado à máxima contração ou expansão radial que uma unidade decisória (indústrias, empresas, grupos, pessoas, máquinas, tecnologias) pode obter nos insumos de produção ou em seus resultados, respectivamente. Dessa forma, uma unidade ineficiente poderá se tornar eficiente produzindo mais resultados com a mesma quantidade de insumos, ou poderá se tornar eficiente reduzindo os insumos utilizados mantendo os mesmos resultados. No primeiro caso, denominamos modelos orientados ao *output*. No segundo, modelos orientados a *input*. Considerando o segundo caso, a formulação matemática é dada da seguinte forma:

$$\begin{aligned} \min \theta & \tag{1} \\ \text{Sujeito a: } & \sum_{j=1}^n \lambda_j x_{ij} \leq \theta x_{i0} \quad | \quad i = 1, 2, 3, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{rj} \geq y_{r0} \quad | \quad r = 1, 2, 3, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \quad | \quad j = 1, 2, 3, \dots, n \\ & \lambda_j \geq 0 \end{aligned}$$

Onde:  $x_{ij}$ : é o vetor dos insumos (*inputs*)  $i = 1, 2, \dots, m$  usados por cada unidade  $j$ ,  $y_{rj}$ : é o vetor dos resultados (*outputs*)  $r = 1, 2, \dots, s$  para cada unidade  $j$ ;  $\lambda_j$ : é um multiplicador escalar (pesos dual) que traduz a contribuição de cada unidade  $j$  na otimização linear; e  $\theta$ : é o escore de eficiência técnica que multiplicado pelo vetor de insumo da unidade “o” em análise  $x_{i0}$  determina a redução ótima que a DMU deve buscar a fim de atingir a eficiência.

Através de um processo de otimização de pesos no modelo (1) de programação matemática, essa metodologia permite a construção de uma superfície não paramétrica por comparações pareadas envolvendo os dados, denominada

fronteira de eficiência, onde cada insumo é reduzido pelo mesmo fator de proporcionalidade  $\theta$  [30] [31]. Na representação gráfica da Figura 1 abaixo temos cinco unidades eficientes (localizadas ao longo da curva de eficiência) e cinco unidades ineficientes (envelopadas pela combinação linear daquelas).

Em um modelo com retornos variáveis de escala como este, a unidade “a” poderia se tornar eficiente ao reduzir proporcionalmente seus insumos mantendo o mesmo resultado até atingir seu referencial (benchmark) unidade “c” na fronteira de eficiência. Ou ainda aumentando seus resultados até o ponto onde se localiza a unidade “b” na fronteira de eficiência (contrações ou expansões radiais).

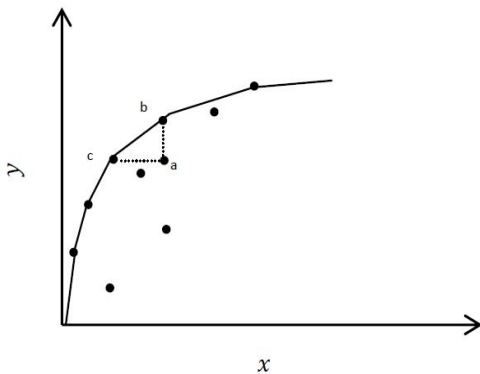


Fig. 1. Fronteira de Eficiência com Retornos Variáveis de Escala.

No contexto da avaliação das páginas Web em que trata esse trabalho, deseja-se otimizar o resultado ao reduzir ao máximo possível o tempo de carregamento das páginas, PLT, e/ou obter o uso máximo possível dos recursos gráficos e digitais (tamanho e número de objetos) dada as restrições de tempo de carregamento. Dessa forma, o modelo de programação linear definido em (1) é ajustado para:

$$\max \phi \tag{2}$$

$$\begin{aligned} \text{Sujeito a:} \quad & \sum_{j=1}^n \lambda_j y_{rj} \leq y_{ro} \quad | r = 1, 2, 3, \dots, s \\ & \sum_{j=1}^n \lambda_j x_{ij} \geq \phi x_{io} \quad | i = 1, 2, 3, \dots, m \\ & \sum_{j=1}^n \lambda_j = 1 \quad | j = 1, 2, 3, \dots, n \\ & \lambda_j \geq 0 \end{aligned}$$

onde:  $x_{ij}$ : é o vetor dos insumos (inputs)  $i = 1, 2, \dots, m$  usados por cada unidade  $j$ ,  $y_{rj}$ : é o vetor dos resultados (outputs)  $r = 1, 2, \dots, s$  para cada unidade  $j$ ;  $\lambda_j$ : é um multiplicador escalar (pesos dual) que traduz a contribuição de cada unidade  $j$  na otimização linear; e  $\phi$ : é o escore de eficiência técnica que multiplicado pelo vetor de insumo da unidade “o” em análise  $x_{io}$  determina a expansão ótima que a DMU deve buscar a fim de atingir a eficiência.

Ou, em função das folgas (slacks) em um modelo aditivo com ambas as orientações [30], [31]:

$$\begin{aligned} \max \quad & \sum_{i=1}^m S_i + \sum_{r=1}^s S_r \tag{3} \\ \text{Sujeito a:} \quad & \sum_{j=1}^n \lambda_j y_{rj} + S_r = y_{ro} \quad | r = 1, 2, 3, \dots, s \\ & \sum_{j=1}^n \lambda_j x_{ij} - S_i = x_{io} \quad | i = 1, 2, 3, \dots, m \\ & \sum_{j=1}^n \lambda_j = 1 \quad | j = 1, 2, 3, \dots, n \\ & \lambda_j \geq 0 \end{aligned}$$

Onde:  $S_i$  e  $S_r$ : são as folgas de produção que devem ser reduzidas ou expandidas com respeito ao insumo  $i$  e resultado  $r$ , respectivamente.

Esse modelo proporciona a construção de uma superfície invertida não paramétrica por partes envolvida pelos dados, tal como representado pela Figura 2. O escore de eficiência para cada DMU que vai de zero a um traduz a necessidade pelos ganhos de eficiência. As folgas nesse modelo são interpretadas no sentido inverso: folgas no tempo de carregamento das páginas PLT representam o que se deseja reduzir em tempo de carregamento (output). Folgas no tamanho ou número de objetos representam o que é esperado expandir comparada a unidades referenciais (páginas benchmark) com tecnologia semelhante.

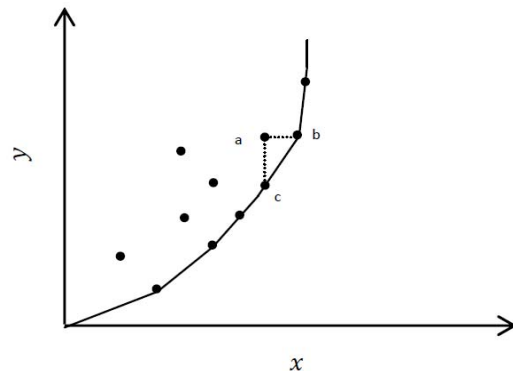


Fig. 2. Fronteira de Eficiência Invertida com Retornos Variáveis de Escala

B. Testbed, Experimentos e Fatores

O cenário aqui descrito foi projetado para refletir as condições reais do tráfego da Internet da maneira mais fiel possível. Portanto, foram usadas as cem páginas mais acessadas do mundo, de acordo com o ranking do site Alexa [32], para executar os experimentos. Os 100 principais sites do Alexa apresentam uma melhor representação da Internet, pois exibe os sites mais acessados do mundo e não estão restritos a um grupo. Entretanto, não foi possível coletar as informações de quantidades de objetos e tamanho das páginas de 18 páginas utilizando a ferramenta Google Big Query por se tratarem, entre outros fatores, de páginas de redirecionamento ou por não estarem mais em funcionamento. Em adição, por esses e outros fatores algumas dessas páginas também não retornaram o tempo de carregamento durante a simulação.

Como o número de 82 unidades é suficiente o bastante para evitar problemas de dimensionalidade e manter a robustez de uma análise com dois insumos e um produto, optamos pela remoção destas páginas e utilizamos apenas as 82 páginas das 100 páginas mais acessadas das quais foram possíveis extrair todas as informações para a análise. A saber, as páginas removidas foram: <http://www.yahoo.com/>, <https://t.co/>, <http://www.paypal.com/>, <http://www.google.com>, <https://www.office.com/>, <http://www.diply.com/>, <https://bongacams.com>, <https://googleusercontent.com>, <http://www.jd.com/>, [chinadaily.com.cn](http://chinadaily.com.cn), [360.com](http://360.com), [coccoc.com](http://coccoc.com), [adnetworkperformance.com](http://adnetworkperformance.com), [haosou.com](http://haosou.com), [adobe.com](http://adobe.com), [china.com](http://china.com), [microsoftonline.com](http://microsoftonline.com), [whatsapp.com](http://whatsapp.com).

Para reproduzir as páginas foi configurado o ambiente com máquinas virtuais usando o VMWare Player, que permitiu paralelizar várias execuções e garantiu exatamente a mesma configuração pois todas as máquinas virtuais eram idênticas. O sistema operacional dessas máquinas foi o Ubuntu 16.04. O navegador utilizado foi o Google Chrome 54.0.28.40.100. As páginas Web são de servidores diferentes. Para entender como o conteúdo é organizado e encontrar o número de objetos nas páginas do Alexa foi usada a ferramenta Google BigQuery. O Google BigQuery contém um grande banco de dados com informações de páginas Web [33]. Assim, foi feito o *upload* das páginas do Alexa no Google BigQuery e, por meio de uma consulta, foi extraído dois dos *inputs* utilizados nos experimentos de análise envoltória dos dados. São eles: o número de objetos dessas páginas; e o tamanho desses objetos medidos em *bytes*.

Para obter o *output* utilizado nos experimentos foi necessário criar um *script* para gravar e reproduzir as páginas Web usando o Mahimahi, uma ferramenta que grava sites e os reproduz sob condições de rede emulada [34]. Com o Mahimahi e através de máquinas virtuais foi possível reproduzir as páginas Web e capturar o *output* utilizado: O *page load time*. O Mahimahi utilizou o mesmo *script* e os mesmos parâmetros para gravação das páginas. Sendo assim, todas as páginas estavam sob a mesma configuração. As páginas foram executadas com RTT de 100 milissegundos e taxa de perda de pacote de 1% com o protocolo de transporte TCP e protocolo de aplicação HTTP 1.1. Esses valores de RTT e perda de pacotes foram escolhidos baseados nos valores médios em [35]. A largura de banda não foi empregada neste trabalho devido à natureza dos experimentos. A janela de controle de congestionamento do protocolo não aumenta para o nível de uso de toda a largura de banda, ou seja, a transmissão termina antes que a conexão a utilize completamente. Diferenças na largura de banda não influenciam os resultados [13]. Foi capturado o arquivo HAR da página e a partir do HAR foi extraído o evento *onLoad* para ser utilizado como *page load time* nos experimentos.

Portanto, para fazer a análise envoltória dos dados, foram avaliados 82 sites com os respectivos valores: Quantidade de objetos e Tamanho da página (*inputs*); Métrica *onLoad* para o *page load time* (*output*). Os *inputs* mencionados foram escolhidos, pois afetam diretamente a performance do site, fazendo com que eles tenham um bom ou mau desempenho comparado aos outros. Através deles é possível visualizar as

páginas com melhores desempenhos e as páginas que precisam de melhoria para alcançar a fronteira de eficiência.

### III. RESULTADOS E DISCUSSÃO

A Tabela 1 apresenta os *inputs*, *output*, escores de eficiência (em termos percentuais) e as respectivas folgas derivadas da análise de eficiência de 82 das 100 páginas Web mais acessadas no mundo. A primeira coluna apresenta as páginas que no contexto DEA simbolizam as unidades decisórias DMUs. Em seguida tem-se a quantidade de objetos e tamanho de cada página (em *bytes*) como *inputs*, ou recursos, utilizados. A quarta coluna apresenta o tempo de carregamento PLT como *output* nesse sistema de produção. Por fim, as últimas colunas são reservadas para a medida de eficiência técnica e as folgas derivadas do modelo. Por exemplo, o site de pesquisa [google.com](http://www.google.com) que é apenas 13,61% eficiente precisa acrescentar 162,4458 objetos (Folga  $x_1$ ) aos objetos já existentes para alcançar a máxima eficiência. A coluna "Folga ( $x_2$ )" apresenta as folgas de tamanho da página em *bytes* que deve aumentar para se tornar eficiente, e por fim "Folga ( $y$ )" apresenta a quantidade em milissegundo que a página deve reduzir para alcançar eficiência máxima.

Essas mudanças nos valores de *inputs* e *output* combinados com outros fatores proporcionam a eficiência ótima em cada página. Uma unidade como o blog [tumblr.com](http://www.tumblr.com), 64,75% eficiente e grande em tamanho, precisa melhorar seu desempenho em 35,25% comparado a páginas que são eficientes com *inputs* e *outputs* semelhantes. Esta página pode obter tal resultado e atingir a fronteira de eficiência ao aumentar 1585 objetos em seu domínio e reduzir em 148 milissegundos o tempo de carregamento. Já uma unidade pequena como a rede social Russa [vk.com](http://www.vk.com), apesar de necessitar de menos objetos, é menos eficiente pois é comparada a páginas com semelhante estrutura produtiva (*inputs* e *outputs*) possui uma distância maior para a fronteira de eficiência.

A partir dos resultados é possível visualizar as páginas que alcançaram a fronteira de eficiência, são elas:

- \_ [amazon.com](http://www.amazon.com);
- \_ [sina.com.cn](http://www.sina.com.cn);
- \_ [bing.com](http://www.bing.com);
- \_ [rakuten.co.jp](http://www.rakuten.co.jp);
- \_ [cntv.cn](http://www.cntv.cn).

Essas páginas possuem eficiência máxima e se encontram sobre a linha da fronteira de eficiência na Figura 3, construída pelo modelo aditivo com ambas as orientações (ver equação 3). A Figura apresenta os recursos utilizados e resultado obtido. O eixo Y representa o output PLT em grandeza de milissegundos. O eixo X simboliza a combinação dos dois *inputs*, quantidade de objetos e tamanho da página. Para que uma das 77 páginas ineficientes atinja a fronteira de eficiência, ela precisa aumentar os recursos (*inputs*) ou então diminuir o resultado (*output*); ou ainda fazer uma combinação dos dois em direção à fronteira de eficiência.

Por exemplo, Pode-se notar 5 pontos sobre a fronteira de eficiência, linha traçada em partes no gráfico da Figura 3, que representam as páginas que atingiram o desempenho ideal ([amazon.com](http://www.amazon.com); [sina.com.cn](http://www.sina.com.cn); [bing.com](http://www.bing.com); [rakuten.co.jp](http://www.rakuten.co.jp); [cntv.cn](http://www.cntv.cn)).

TABELA 1  
TABELA DE EFICIÊNCIA DAS PÁGINAS

Páginas	Objetos (x1)	Tamanho (x2)	PLT (v)	Eficiência	Folga (x1)	Folga (x2)	Folga (v)
amazon.com	199	4639244	8586	100,00%	0	0	0
sina.com.cn	496	19069095	16606	100,00%	0	0	0
bing.com	56	941086	1717	100,00%	0	0	0
rakuten.co.jp	808	10521295	10760	100,00%	0	0	0
cntv.cn	130	1021743	2627	100,00%	0	0	0
msn.com	92	758056	2871	74,19%	5	0	243
pornhub.com	136	2760955	7618	73,34%	38	0	748
xinhuanet.com	128	2577879	7162	68,48%	37	0	291
tumblr.com	92	14845093	25104	64,75%	1585	0	148
ebay.com	174	2624886	9332	62,14%	0	481475	744
baidu.com	18	576698	3240	56,44%	98	0	612
youtube.com	131	1589924	6618	54,75%	2	0	555
amazon.de	302	6225062	12856	54,30%	307	0	378
naver.com	101	1537049	6733	52,93%	51	0	670
flipkart.com	182	2150533	9504	51,41%	0	603182	6
nicovideo.jp	112	1621744	7749	50,00%	0	520856	879
amazon.co.uk	254	5669366	14093	49,11%	420	0	705
sohu.com	385	1573258	12281	47,64%	0	7219496	1521
hao123.com	148	1869267	8890	47,61%	61	0	199
imgur.com	102	1257262	6656	43,29%	6	0	593
microsoft.com	84	2040864	10262	42,64%	157	0	764
github.com	14	1638025	8272	42,60%	265	0	491
amazon.in	307	4408514	11568	41,90%	75	0	808
amazon.co.jp	258	5060789	17499	37,69%	365	0	676
360.cn	165	1504194	9750	37,54%	6	0	149
mail.ru	132	789647	9531	33,84%	0	732181	1648
outbrain.com	91	1507173	9354	32,50%	0	67974	766
xhamster.com	96	869881	7843	30,37%	0	121213	870
pixnet.net	182	3993262	17982	27,95%	380	0	352
alibaba.com	47	1239330	9128	26,33%	101	0	540
google.de	17	678982	5484	24,04%	97	0	331
google.pl	17	672061	6031	23,80%	96	0	878
google.fr	17	670175	5704	23,73%	96	0	551
google.co.uk	17	669256	5625	23,70%	96	0	472
google.it	18	679504	6115	23,39%	165	0	52
google.com.mx	15	647486	5867	22,93%	102	0	714
cnn.com	147	2861726	15879	22,92%	352	0	773
kat.cr	184	387564	10922	22,77%	0	8819384	161
google.co.jp	16	649383	6662	22,36%	170	0	599
google.com.br	15	647513	6593	22,29%	174	0	530
google.ru	19	642695	6513	22,13%	156	0	450
google.co.in	15	641922	6153	22,10%	174	0	90
google.ca	15	641062	6706	22,07%	174	0	643
google.co.id	15	638836	6197	21,99%	173	0	134
imdb.com	206	2144843	15093	21,96%	0	1776714	1705
facebook.com	36	487700	5806	21,42%	0	547324	653
dropbox.com	208	4239393	21627	20,14%	583	0	107
yandex.ru	63	585770	7463	19,93%	0	46424	490
qq.com	220	2314736	18427	19,57%	0	1679670	694
google.com.hk	14	525033	6571	18,08%	164	0	508
google.es	17	679515	6971	18,05%	129	0	101
taobao.com	126	2187887	15157	17,52%	275	0	52
google.com.au	15	645025	6880	17,13%	136	0	10
wikipedia.org	20	53272	3177	15,38%	0	675475	549
soso.com	20	141242	3709	15,38%	0	103670	1081
yahoo.co.jp	57	512902	8779	15,32%	0	578297	89
gmw.cn	141	1669451	14422	15,03%	0	437044	1034
ok.ru	41	578970	9322	14,64%	0	751488	734
reddit.com	53	474662	8809	14,24%	0	594068	118
ask.com	23	276948	5079	14,10%	22	0	734
google.com	16	640444	8693	13,61%	162	0	105
weibo.com	110	940945	12167	13,61%	0	3609626	1407
aliexpress.com	68	1423637	11290	13,53%	305	0	529
twitter.com	35	129214	6075	13,46%	0	1083611	819
pinterest.com	31	1368820	11394	13,00%	569	0	634
go.com	82	1332187	13888	11,54%	227	0	500
vk.com	39	448042	8969	11,41%	30	0	278
stackoverflow.com	26	288626	6116	10,74%	0	217473	53
tianya.cn	40	306612	8625	10,25%	0	75762	742
instagram.com	23	459604	10017	9,60%	114	0	519
apple.com	53	2069649	23528	9,06%	135	0	52
live.com	16	132472	5067	8,60%	0	422842	721
linkedin.com	18	325990	8404	8,47%	85	0	623
craigslist.org	10	162069	4454	8,25%	64	0	109
fc2.com	59	579111	11505	7,30%	0	2590419	745
onlickads.net	12	152909	5255	7,14%	0	682532	102
wordpress.com	20	263552	7994	6,85%	6	0	214
netflix.com	13	702609	15073	5,66%	690	0	878
xvideos.com	54	1279964	26609	5,56%	831	0	743
tmall.com	36	571077	16632	4,27%	134	0	719
blogspot.com	14	78667	8395	3,58%	0	873791	511
logger.com	14	78710	7894	3,58%	0	872593	11

Em termos práticos, para que as outras 77 páginas ineficientes atinjam a fronteira é necessário que as mesmas melhorem seus resultados (ao reduzir o *page load time* no eixo Y) ou seus insumos (ao aumentar o número de objetos e/ou tamanho das páginas no eixo X) de acordo com o percentual de ineficiência obtido entre a distância radial de cada ponto até a sua combinação linear na fronteira de eficiência.

Nesse caso, a página msn.com, por exemplo, que se encontra no canto esquerdo mais inferior do gráfico com escore de 74,19% de eficiência, próximo ao ponto na fronteira que corresponde a cntv.cn, teria que melhorar em 25,81% aumentando em 5 o número de objetos da página e reduzindo em 243 milissegundos o tempo de carregamento (ver Tabela 1). Somente assim, nesse contexto, as páginas conseguem o valor esperado para atingir a eficiência máxima.

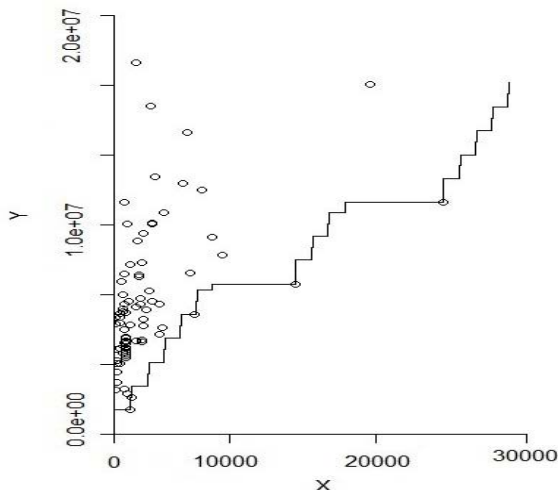


Fig. 3. Fronteira de Eficiência Aditiva Invertida para a Avaliação das Páginas Web (ambas as orientações).

As cinco unidades eficientes (amazona, sina, bing, rakuten e cntv) são consideradas *benchmarks* (referências) para as demais páginas buscarem práticas e tecnologias semelhantes para o alcance da eficiência técnica. Seus valores de folga para melhoria na Tabela 1 estão zerados, pois sua performance se encontra no estado ideal. Para as demais páginas são mostrados quais recursos ou resultado elas precisam modificar para alcançar a fronteira de eficiência. As páginas mais acessadas que apresentam maior ineficiência técnica no carregamento são o blogger, blogspot (ambos de mesmo domínio), tmall, xvideos e netflix. Estas precisam buscar um incremento de eficiência acima de 90%.

#### IV. CONCLUSÃO

Este trabalho realizou uma avaliação da performance da Internet a partir da perspectiva de análise de eficiência das páginas Web. Foram incluídas neste trabalho as páginas da Internet mais acessadas em diferentes condições de rede para avaliar o desempenho dessas páginas. Foi feita a medição de quais páginas apresentam boa performance e quais não apresentam utilizando Análise Envoltória de Dados (DEA). A

partir desta avaliação, foi possível construir um *ranking* contendo as páginas eficientes e ineficientes de acordo com a projeção do tempo de carregamento sobre o número de objetos e tamanho dos endereços acessados como medida de eficiência técnica.

Na prática, DEA é utilizado para otimizar (maximizar ou minimizar) uma função linear de variáveis sujeita a uma série de restrições. Com essa metodologia é possível comparar unidades (páginas) grandes com pequenas, pois o processo de otimização compara cada unidade com uma unidade fictícia que apresente mesmas condições, quantidades de recursos e resultados (*inputs* e *output*). Com os resultados de quais páginas possuem máxima eficiência é possível entender melhor o mecanismo de funcionamento da Internet e estudar fatores que a influenciam e como se pode melhorá-la. No que diz respeito às páginas ineficientes apresentadas neste *ranking*, é possível encontrar pontos de melhoria e procurar estudar a interação dos protocolos de comunicação e tecnologia adotada no desempenho de rede.

Os autores acreditam que este seja um ponto de fundamental importância para entender como páginas hospedadas por grandes corporações com grandes investimentos em rede ainda carecem comparadas a outras menos visadas em condições semelhantes. A influência de diferentes protocolos de comunicação na eficiência de rede no carregamento desses endereços Web deve ser destaque em trabalhos futuros. Para mais estudos de caso, planeja-se definir diferentes níveis dos fatores aqui estudados. Também se planeja adicionar novos recursos de redes que influenciariam a métrica do tempo de carregamento da página. E também realizar uma nova análise do desempenho de eficiência das páginas Web.

#### REFERÊNCIAS

- [1] B. Sanou, "Ict facts and figures," The world in, pp. 1865–2015, 2015.
- [2] Big Commerce, "Big Commerce", <https://www.bigcommerce.com/blog/4-tips-improve-ecommerce-bounce-rate-right-now-sellmore-video/> accessed: Dec-2017.
- [3] PR Newswire, "'customers are won or lost in one second,' finds new aberdeen report," December 2008. [Online]. Available: <http://www.prnewswire.com/news-releases/customers-are-won-or-lost-in-one-second-finds-new-aberdeen-report-65399152.html>.
- [4] Google Inc. "Using site speed in web search ranking", <https://webmasters.googleblog.com/2010/04/using-site-speed-in-web-search-ranking.html/>, accessed: Jan-2017.
- [5] Charnes, A., Cooper, W. W. and Rhodes, E. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429-444. doi:429-444. 10.1016/0377-2217. 1978.
- [6] Daraio, C., Kerstens, K. H., Nepomuceno, T. C. C. and Sickles, R. "Productivity and Efficiency Analysis Software: An Exploratory Bibliographical Survey of the Options". *Journal of Economic Surveys*, 33(1), 85-100. doi: 10.1111/joes.12270. 2019.
- [7] Daraio, C., Kerstens, K., Nepomuceno, T. C. C. and Sickles, R.C. "Empirical surveys of frontier applications: a meta-review". *International. Transactions in Operational Research*, 27: 709-738. doi:10.1111/itor.12649. 2020.
- [8] Emrouznejad, A., G.-I. Yang. A Survey and Analysis of the First 40 Years of Scholarly Literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, 61, 4-8. doi:10.1016/j.seps.2017.01.008. 2018.
- [9] Coelli, T. J., Rao, D. S. P., O'Donnell, C. J. and Battese, G. E. *An introduction to efficiency and productivity analysis*. Springer Science & Business Media. 2005.
- [10] Castellanos, W., Guerri, J. C. and Arce, P. Performance evaluation of scalable video streaming in mobile ad hoc networks. *IEEE Latin*

- America Transactions*, 14(1), 122-129. doi:10.1109/TLA.2016.7430071. 2016.
- [11] Saraiva, N. F., & Soares, A. C. B. A Performance Evaluation Programmable Architecture for Choice of Switching Paradigm in Hybrid Optical Networks. *IEEE Latin America Transactions*, 14(11), 4567-4572. doi: 10.1109/TLA.2016.7795830. 2016.
- [12] Du, Y., Li, C., Hu, Q., Li, X., & Chen, X. Ranking webpages using a path trust knowledge graph. *Neurocomputing*, 269, 58-72. doi: 10.1016/j.neucom.2016.08.142. 2017.
- [13] Nepomuceno, K., de Oliveira, I. N., Aschoff, R. R., Bezerra, D., Ito, M. S., Melo, W. D. F. H. Sadok & Szabó, G. QUIC and TCP: A Performance Evaluation. In *2018 IEEE Symposium on Computers and Communications (ISCC)*. IEEE. pp. 00045-00051. ISSN: 1530-1346 doi:10.1109/ISCC.2018.8538687. 2018.
- [14] Machado, L. G., & Roboredo, M. C. Efficiency evaluation of Brazilian electrical distributors using dea game and cluster analysis. *IEEE Latin America Transactions*, 14(11), 4499-4505. 10.1109/TLA.2016.7795820. 2016.
- [15] Fenerich, F. C., da Costa, S. E. G., & de Lima, E. P. Energy Efficiency in Industrial Environments: Overview and Research Agenda. *IEEE Latin America Transactions*, 15(3), 415-422. doi:10.1109/TLA.2017.7867170. 2017.
- [16] dos Santos Constant, R. Brazilian Electricity Distributors Efficiency Index Based on Non Radial Efficiency. *IEEE Latin America Transactions*, 15(9), 1657-1663. doi:10.1109/TLA.2017.8015049. 2017.
- [17] Nepomuceno, T. C. C., & Costa, A. P. C. "Resource allocation with Time Series DEA applied to Brazilian Federal Saving banks". *Economics Bulletin*, 39(2), 1384-1392. 2019.
- [18] Nepomuceno, T. C. C., Daraio, C., Costa, A. P. C. S. "Combining Multi-Criteria and Directional Distances to Decompose Non-compensatory Measures of Sustainable Banking Efficiency". *Economics Letters*. doi: 10.1080/13504851.2019.1616051. 2019.
- [19] Serrano-Cinca, C., Fuertes-Callén, Y., & Mar-Molinero, C. Measuring DEA efficiency in Internet companies. *Decision Support Systems*, 38(4), 557-573. doi:10.1016/j.dss.2003.08.004. 2005.
- [20] Martínez-Núñez, M., & Pérez-Aguilar, W. S. Efficiency analysis of information technology and online social networks management: An integrated DEA-model assessment. *Information & Management*, 51(6), 712-725. doi:10.1016/j.im.2014.05.009. 2014.
- [21] Li, C., Cheng, B., Chen, J., Gu, P., Deng, N., & Li, D. "A web service performance evaluation approach based on users experience". In *2011 IEEE International Conference on Web Services* (pp. 734-735). IEEE. doi: 10.1109/ICWS.2011.29. 2011.
- [22] Ghosh, P., & Rau-Chaplin, A. "Performance of Dynamic Web Page Generation for Database-driven Web Sites". In *International Conference on Next Generation Web Services Practices* (pp. 56-63). IEEE. doi: 10.1109/NWESP.2006.242006. 2006.
- [23] Mohapatra, P., & Chen, H. "WebGraph: A framework for managing and improving performance of dynamic web content. *IEEE Journal On Selected Areas in Communications*", 20(7), 1414-1425. Doi: 10.1109/JSAC.2002.802072. 2002.
- [24] Sanjaya, R. Web traffic reduction for infrequent update application using Green Ajax. In *2010 2nd IEEE International Conference on Information Management and Engineering* (pp. 170-176). IEEE. Doi: 10.1109/GLOCOM.2016.7841749. 2010.
- [25] Butkiewicz, M., Madhyastha, H. V., & Sekar, V. Characterizing web page complexity and its impact. *IEEE/ACM Transactions on Networking*, 22(3), 943-956. doi: 10.1109/TNET.2013.2269999. 2013.
- [26] Najadat, H., Al-Badarnah, A., Al-Huthaifi, R., Abo-Zaitoon, A., & Al-Omary, Y. Evaluating Jordanian universities' websites based on data envelopment analysis. In *2017 8th International Conference on Information and Communication Systems (ICICS)* (pp. 159-164). IEEE. doi: 10.1109/IACS.2017.7921964. 2017.
- [27] Li, Z., Li, J., & Li, X. A study on website operation efficiency evaluation: Based on DEA model. In *2009 International Conference on Management and Service Science* (pp. 1-6). IEEE. doi: 10.1109/ICMSS.2009.5301119. 2009.
- [28] Shephard, R.W. *Theory of Cost and Production Functions*. Princeton. 1970.
- [29] Färe, R., & Lovell, C. K. Measuring the technical efficiency of production. *Journal of Economic Theory*, 19(1), 150-162. doi:10.1016/0022-0531(78)90060-1. 1978.
- [30] Cook, W. D., & Seiford, L. M. Data envelopment analysis (DEA)—Thirty years on. *European Journal of Operational Research*, 192(1), 1-17. doi:10.1016/j.ejor.2008.01.032. 2009.
- [31] Charnes, A., Cooper, W.W., Golany, B., Seiford, L.M., Stutz, J. Foundations of data envelopment analysis and Pareto-Koopmans empirical production functions. *Journal of Econometrics* 30, 91-107. doi:10.1016/0304-4076(85)90133-2 1985.
- [32] Alexa Site, "The top 500 sites on the web", <http://www.alexa.com/topsites/>, accessed: Jan - 2017.
- [33] Google, "Google Bigquery", <https://cloud.google.com/bigquery/>, accessed: Jan - 2017.
- [34] R. Netravali, A. Sivaraman, S. Das, A. Goyal, K. Winstein, J. Mickens, and H. Balakrishnan, "Mahimahi: Accurate record-and-replay for http." in *USENIX Annual Technical Conference*, 2015, pp. 417-429.
- [35] Xiao Sophia Wang, Aruna Balasubramanian, Arvind Krishnamurthy, and David Wetherall. How speedy is spdy? In *NSDI*, 2014.



**Késsia Thais Cavalcanti Nepomuceno** é mestranda em Ciência da Computação pela Universidade Federal de Pernambuco (UFPE) e possui Bacharelado em Ciência da Computação (2019) pela mesma universidade. Possui experiência em desenvolvimento de software; banco de dados; redes de computadores, atuando principalmente nos temas de protocolos de comunicação e análise de performance. Tem interesse nos temas de IoT e computação em nuvem.



**Thyago Celso Cavalcante Nepomuceno** é Bacharel em Ciências Econômicas (2013), Mestre (2014) e Doutor (2019) em Engenharia de Produção pela Universidade Federal de Pernambuco (UFPE) e PhD em Industrial and Management Engineering pela Sapienza University of Rome (2020). Professor Adjunto no Departamento de Engenharia de Produção da UFPE e membro do Grupo de Pesquisa em Sistemas de Informação e Decisão (GPSID), trabalha nos campos da Análise da Eficiência e Produtividade, Métodos Quantitativos, Sistemas de Informação Geográficos e Mapeamento Bibliométrico.



**Djamel Sadok** recebeu seu PhD de Ciência da Computação na Universidade de Kent em Canterbury, Reino Unido em 1990. Ele é um membro da equipe do Centro de Ciência da Computação da Universidade Federal de Pernambuco desde 1993. Seus interesses de pesquisa incluem sistemas de comunicação, acesso redes, segurança, cloud computing e classificação de tráfego. Ele também está ativamente envolvido no GPRT, um grupo de pesquisa nas áreas de redes e telecomunicações. Atualmente, ele lidera a equipe do GPRT e coordena diversos projetos de pesquisa.