Explorando Redes Sociais Online: Da Coleta e Análise de Grandes Bases de Dados às Aplicações

Fabrício Benevenuto
UFOP

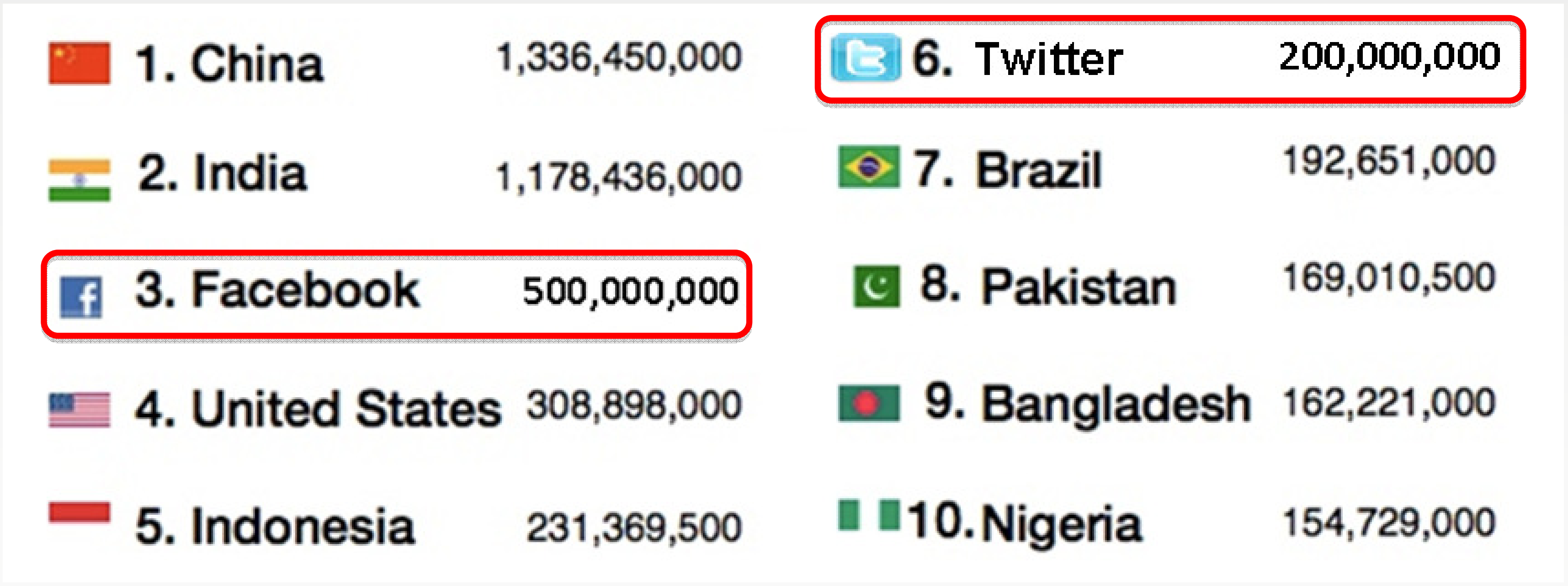
Jussara Almeida
UFMG

Altigran S. Silva
UFAM/UFMG

Organização

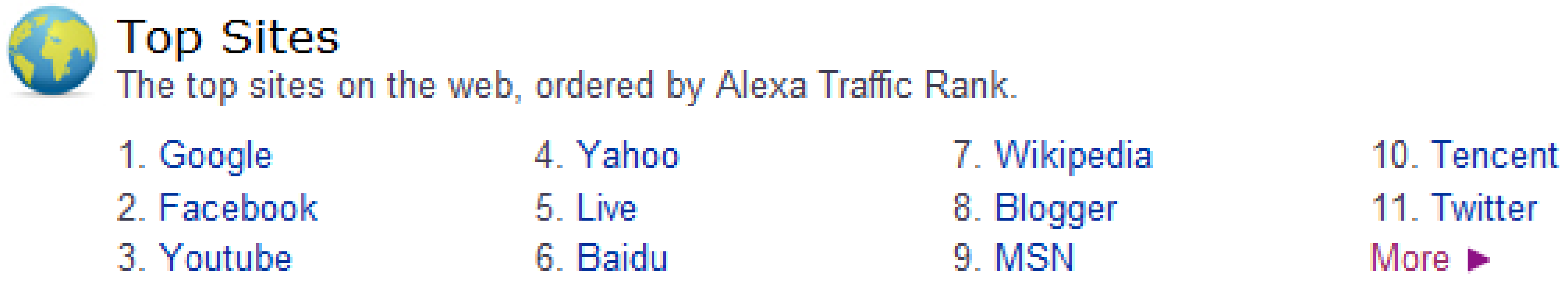
- Introdução
  - redes complexas e redes sociais online
- Coleta e obtenção de dados de redes sociais
  - Diferentes formas de obtenção de dados
  - Principais trabalhos que realizaram diferentes tipos de coletas
  - Técnicas, APIs, códigos

Se o Facebook e o Twitter fossem países



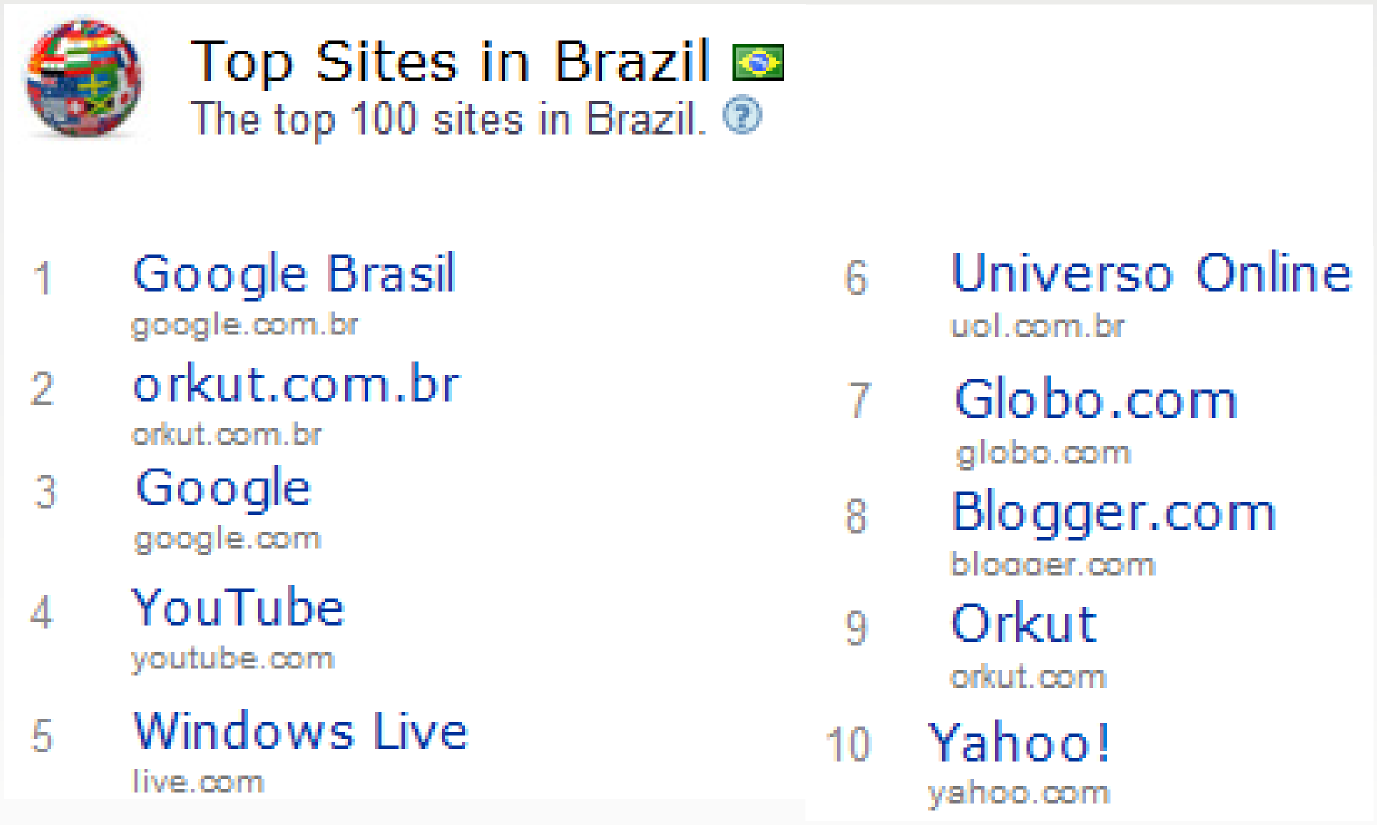
1. China	1,336,450,000	6. Twitter	200,000,000
2. India	1,178,436,000	7. Brazil	192,651,000
3. Facebook	500,000,000	8. Pakistan	169,010,500
4. United States	308,898,000	9. Bangladesh	162,221,000
5. Indonesia	231,369,500	10. Nigeria	154,729,000

Quais os sites mais populares da Web?


Top Sites
The top sites on the web, ordered by Alexa Traffic Rank.

1. Google	4. Yahoo	7. Wikipedia	10. Tencent
2. Facebook	5. Live	8. Blogger	11. Twitter
3. Youtube	6. Baidu	9. MSN	More ▶

Quais os sites mais populares do Brasil?

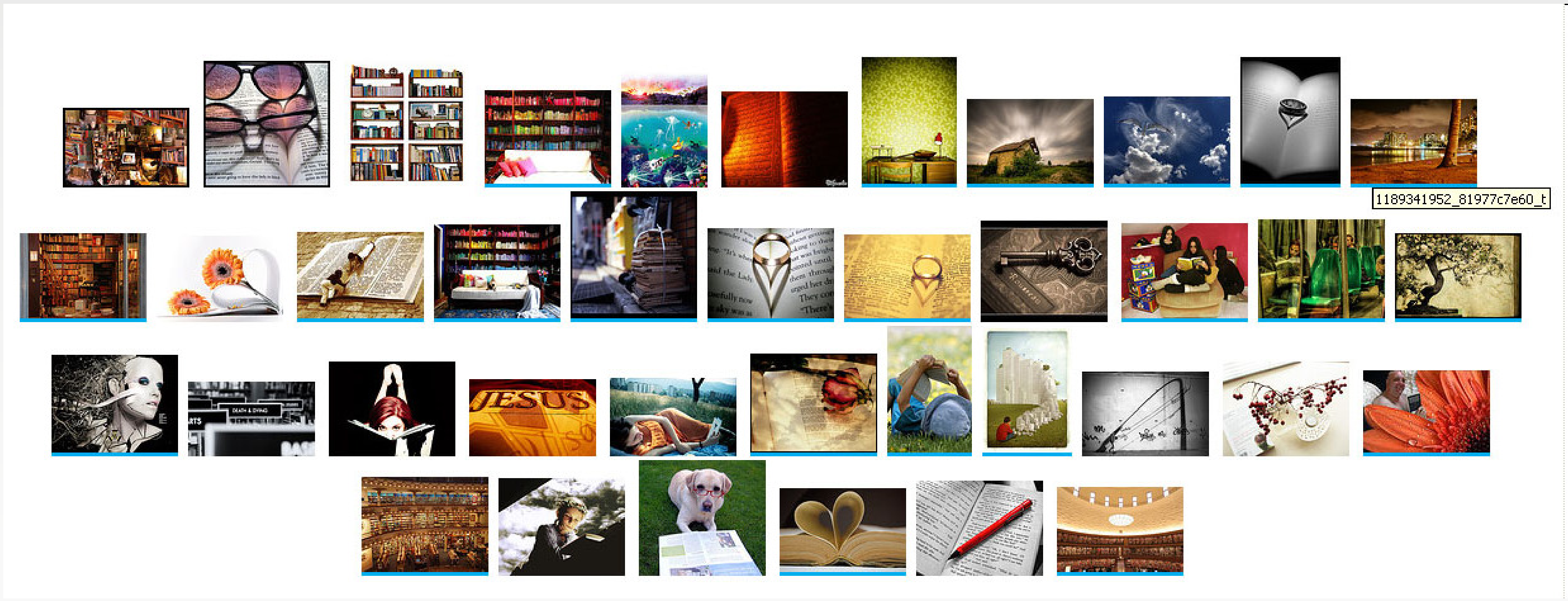


Top Sites in Brazil
The top 100 sites in Brazil.

1 Google Brasil
google.com.br
2 orkut.com.br
orkut.com.br
3 Google
google.com
4 YouTube
youtube.com
5 Windows Live
live.com
6 Universo Online
uol.com.br
7 Globo.com
globo.com
8 Blogger.com
blogger.com
9 Orkut
orkut.com
10 Yahoo!
yahoo.com

YouTube

- 2 bilhões de vídeos vistos por dia
- 24 horas de vídeos recebidos por minuto



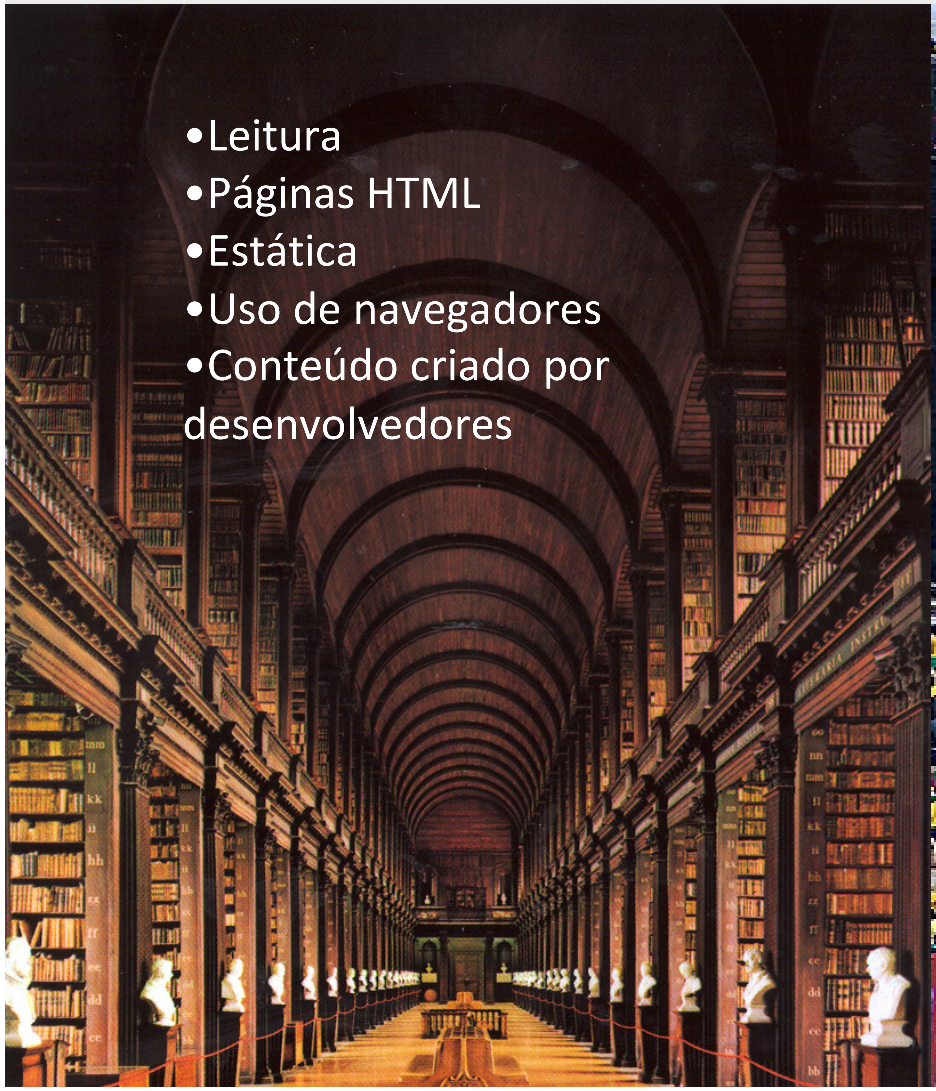
4 Bilhões de imagens no Flickr

Mais estatísticas

- Redes sociais são mais populares do que email
- Usuários do Facebook navegam 700 bilhões de minutos por mês
- Orkut possui 100 milhões de usuários, a maioria brasileiros
- Twitter recebe 65 milhões de tweets por dia

Mudança de perspectiva da Web

Web 1.0



- Leitura
- Páginas HTML
- Estática
- Uso de navegadores
- Conteúdo criado por desenvolvedores

Web 2.0

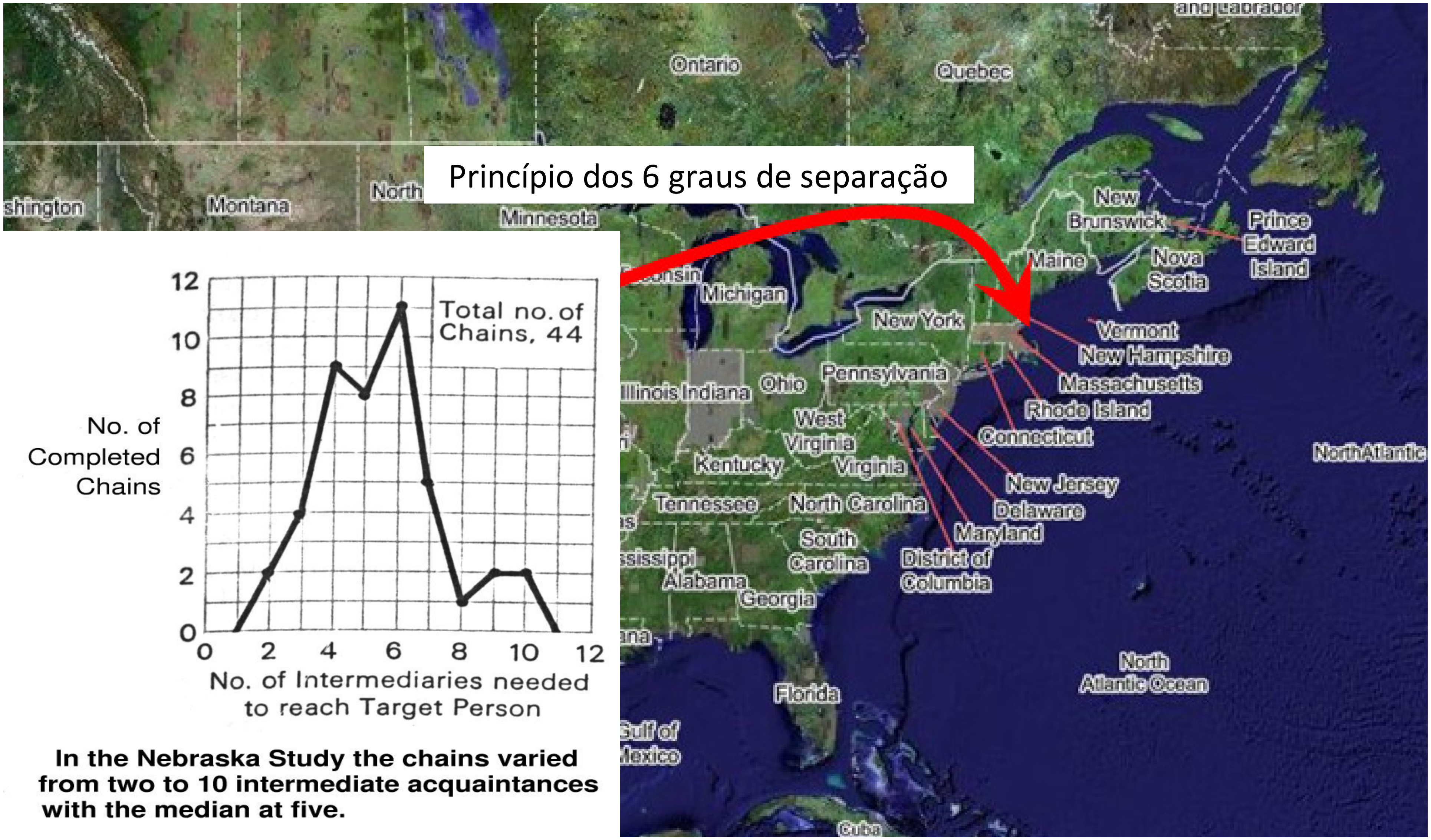
- Leitura/escrita
- Páginas/postagens/mídia
- Dinâmica
- Navegadores, RSS, etc.
- Conteúdo criado por usuários

Por que redes sociais online?

- Propósito comercial
  - 1.2 bilhões de dólares investidos em propaganda em 2007
  - Campanhas políticas
- Grande volume de dados
  - Recuperação e distribuição de conteúdo
- Aspectos sociológicos
  - Reprodução do comportamento humano
  - Registro de interações sociais

Experimento de Milgram (1967):

Dado um indivíduo em Boston, passe a mensagem para uma pessoa que você conhece que é a mais próxima do alvo até que a mensagem atinja o alvo.

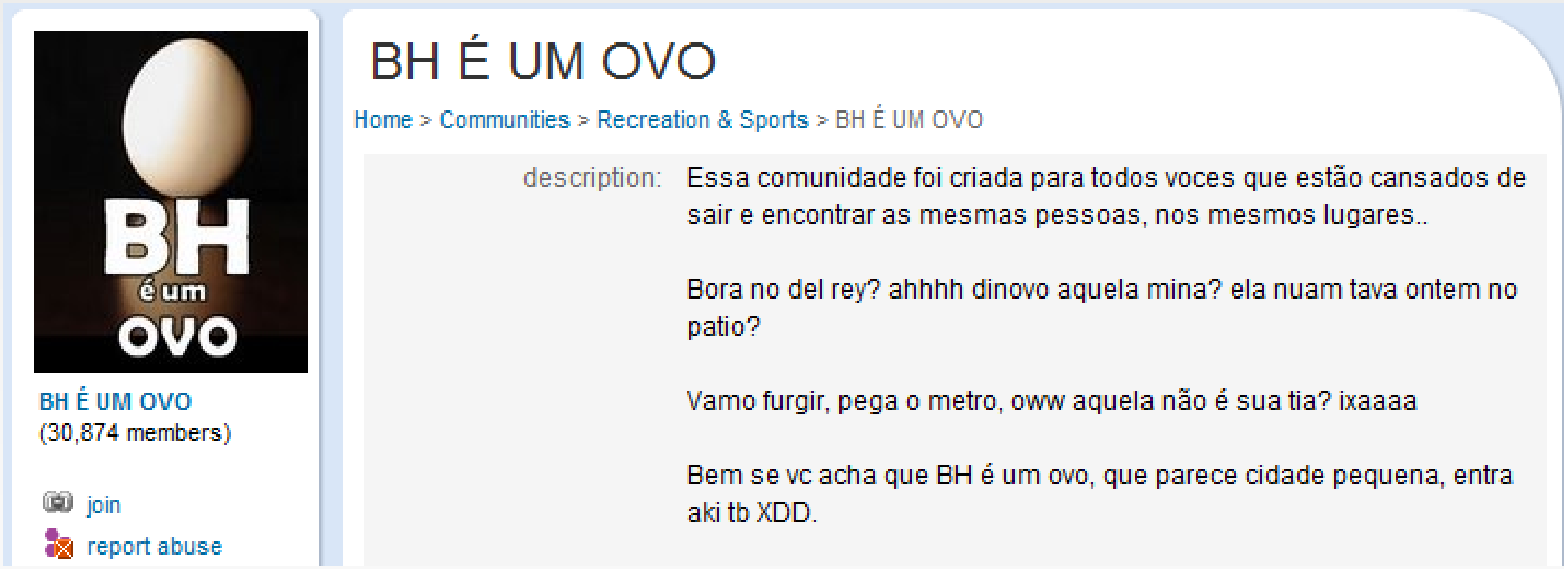


In the Nebraska Study the chains varied from two to 10 intermediate acquaintances with the median at five.

Redes small world

- Pessoa alvo trabalhava em Boston como corretor
- 296 enviaram cartas
- 20% alcançaram o alvo
- comprimento médio da seqüência = 6.5
- **Os seis graus de separação**

Redes small world



BH É UM OVO

Home > Communities > Recreation & Sports > BH É UM OVO

description: Essa comunidade foi criada para todos voces que estão cansados de sair e encontrar as mesmas pessoas, nos mesmos lugares..

Bora no del rey? ahhhh dinovo aquela mina? ela nuam tava ontem no patio?

Vamo furgir, pega o metro, oww aquela não é sua tia? ixaaaa

Bem se vc acha que BH é um ovo, que parece cidade pequena, entra aki tb XDD.

BH É UM OVO
(30,874 members)

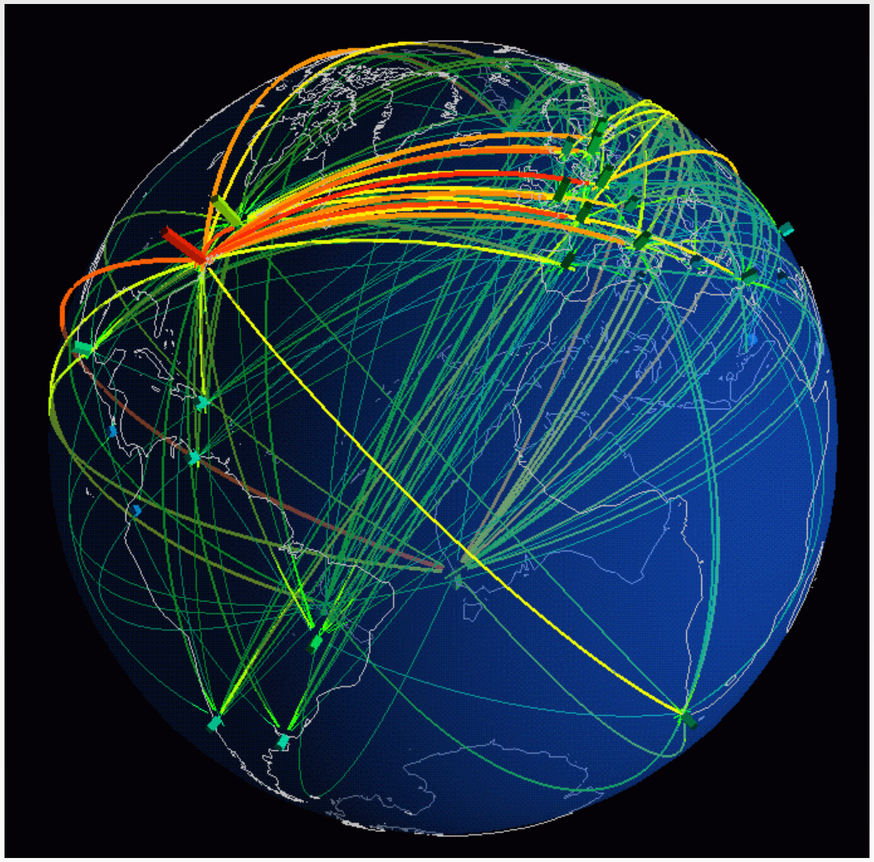
join
report abuse

Redes small world

• Experimento de email Dodds, Muhamad, Watts, Science 301, (2003)

• 18 alvos, 13 países diferentes

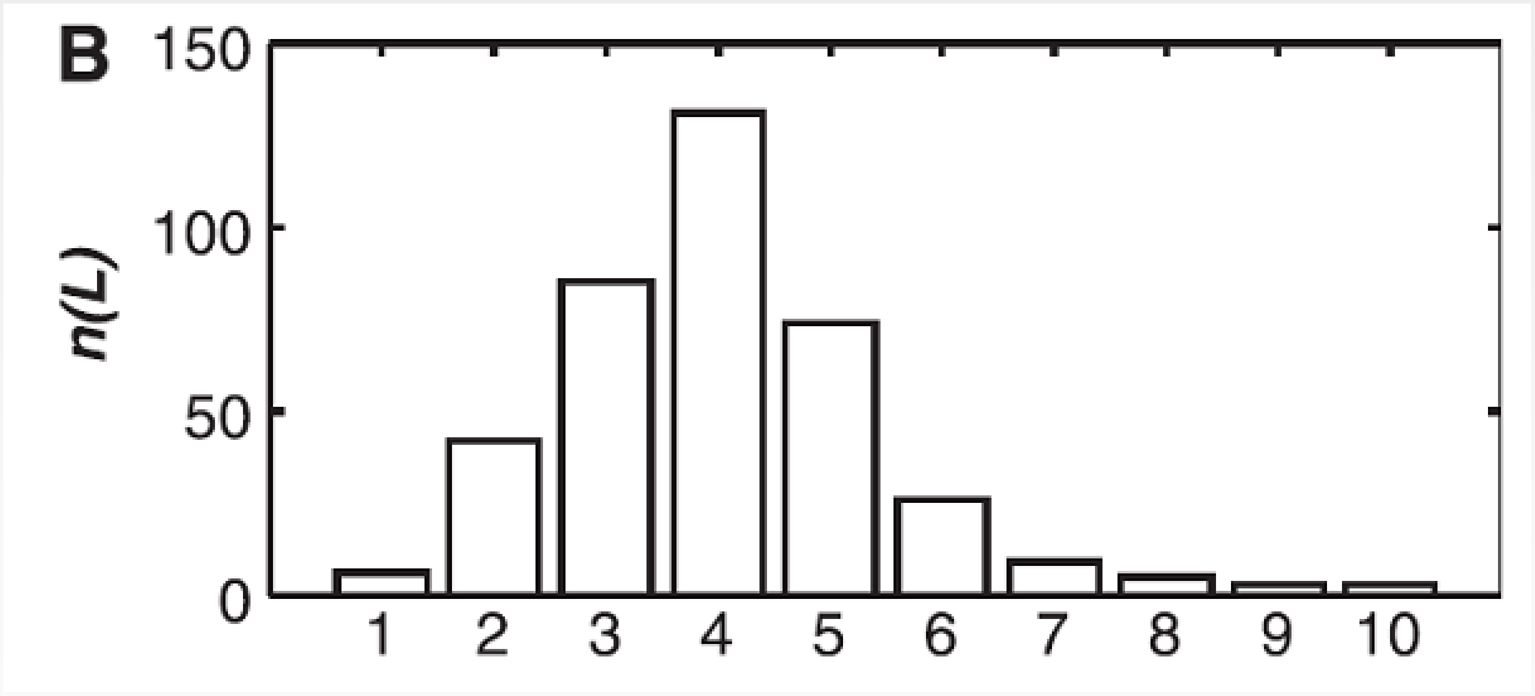
• 60.000+ participantes
• 24.163 seqüência de mensagens
  - 384 alcançaram alvos
  - Maior parte dos caminhos médios entre 2 e 7

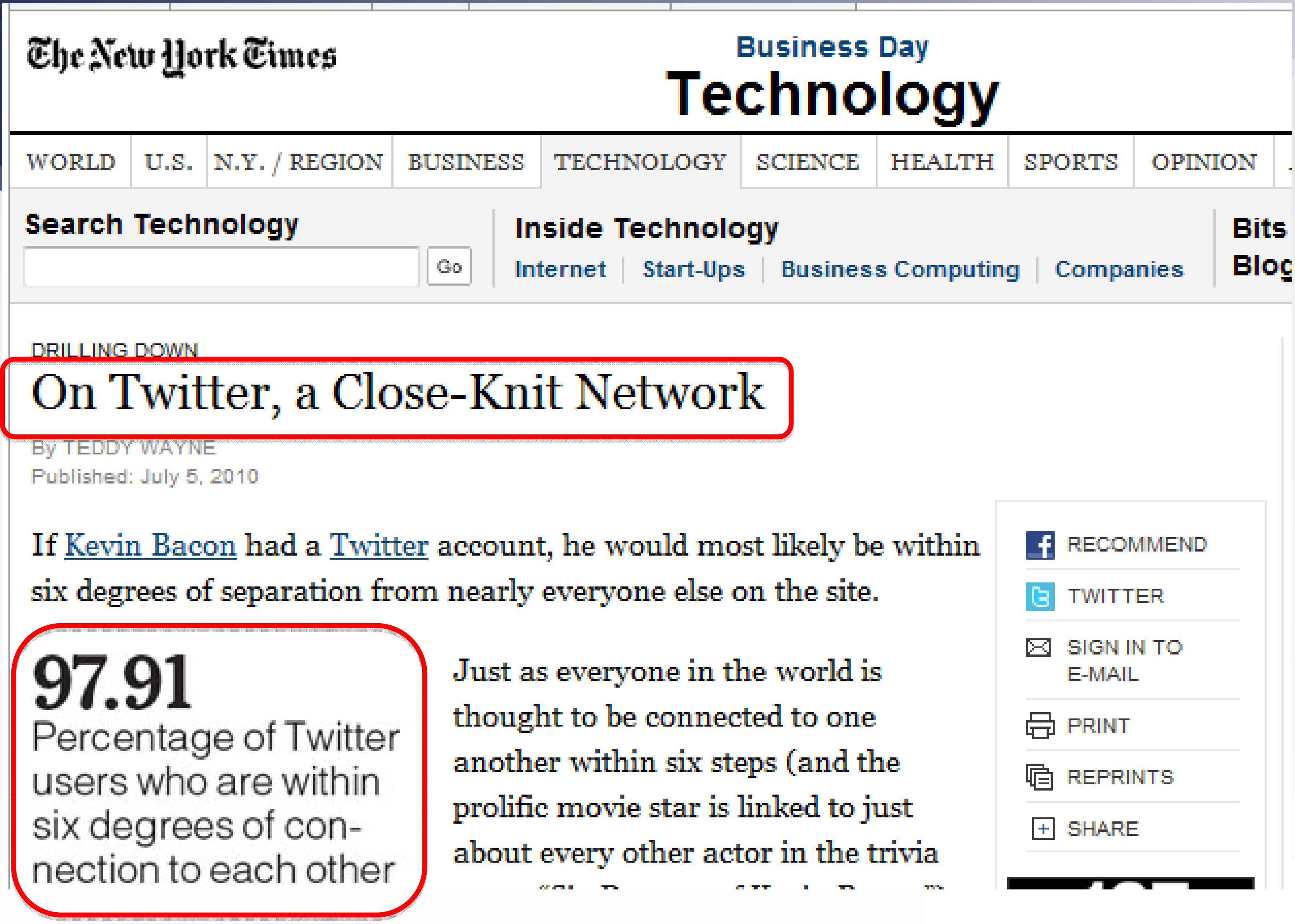


Redes small world

- Alvos do experimento
  - Um professor na Ivy League university
  - Um inspetor na Estônia,
  - Um consultor tecnológico na Índia
  - Um policial na Austrália
  - Um veterinário no exército Norueguês

Redes small world




The New York Times

Business Day
Technology

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

Search Technology [Go]

Inside Technology
Internet | Start-Ups | Business Computing | Companies

Bits
Blog

DRILLING DOWN

On Twitter, a Close-Knit Network

By TEDDY WAYNE
Published: July 5, 2010

If Kevin Bacon had a Twitter account, he would most likely be within six degrees of separation from nearly everyone else on the site.

97.91
Percentage of Twitter users who are within six degrees of con-nection to each other

Just as everyone in the world is thought to be connected to one another within six steps (and the prolific movie star is linked to just about every other actor in the trivia

RECOMMEND
TWITTER
SIGN IN TO E-MAIL
PRINT
REPRINTS
SHARE

Estudo da Web em larga escala

Getting information off the Internet is like taking a drink from a fire hydrant.
Mitchell Kapor
Adapted from http://www.flickr.com/photos/josephrobertson/127758523

Desafios tecnológicos

Coleta de dados → **Medição e Análise** → **Construção de aplicações**

Coleta de dados
- Elos sociais, usuários, páginas, postagens, blogs, vídeos, fotos, metadados, etc.
- Coleta precisa ser rápida e completa
- Armazenamento

Medição e Análise
- Processamento distribuído e paralelo
- Memória compartilhada
- Análise, simulação, visualização

Construção de aplicações
- Algoritmos
- Modelos
- Sistemas

Temas em redes sociais online

Análise e modelagem de comportamento social

Detecção de comportamento oportunista

Predição de popularidade, evolução temporal de redes sociais

Propagação de informação, influência social, comunidades

Teorias e modelos sobre comportamento coletivo

Sistemas de recomendação, ranking e recuperação de conteúdo (tempo real)

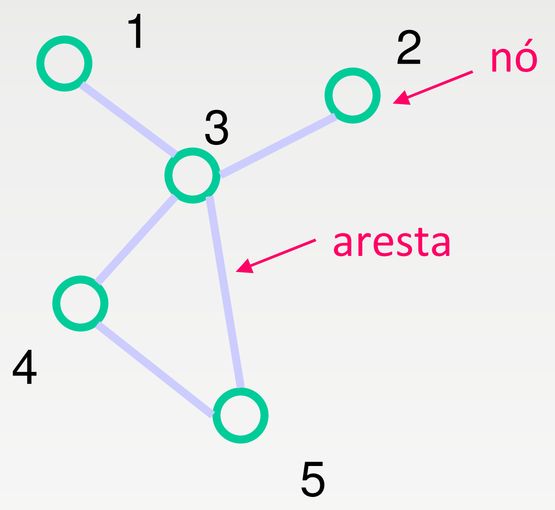
Análise de sentimentos e mineração de opiniões

Segurança, privacidade, riscos e confiança

Mashups e agregação de conteúdo

Parelelismo, algoritmos para grandes grafos

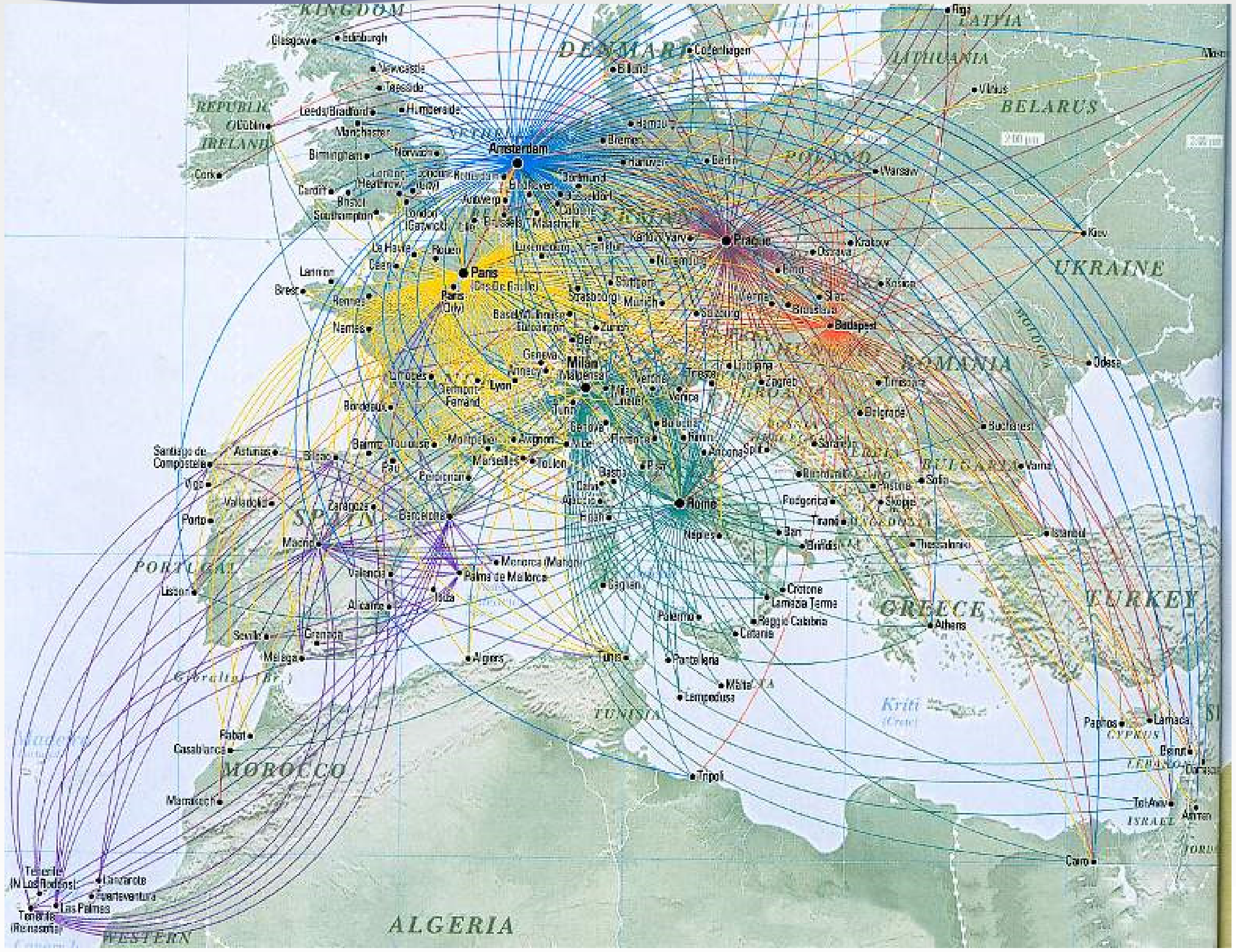
Teoria de Grafos e Redes complexas



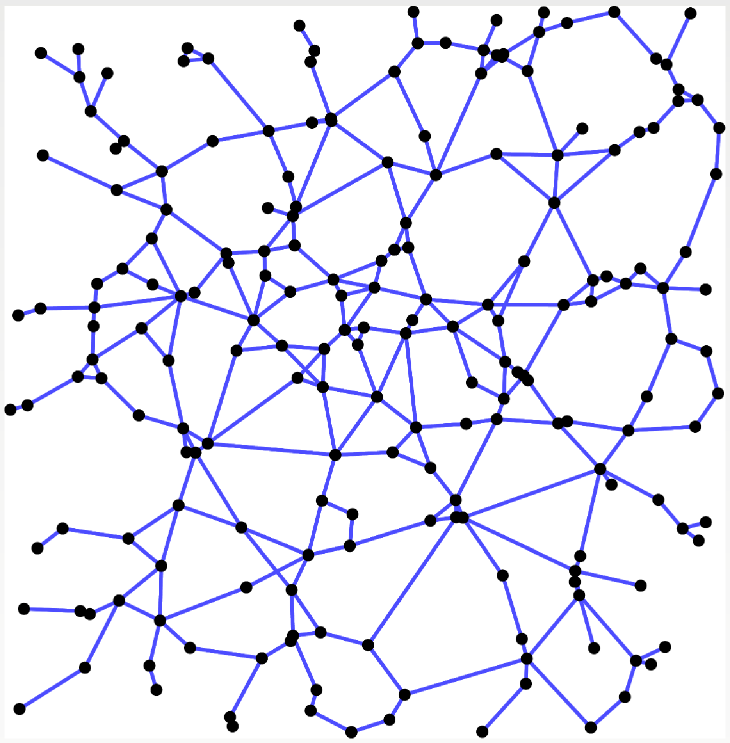
“Rede” ≡ “Grafo”

Pontos	**Linhas**	
vértices	Arcos, arestas	matemática
nós	Links, arestas	Ciência Comp.
atores	ligações, relações	sociologia

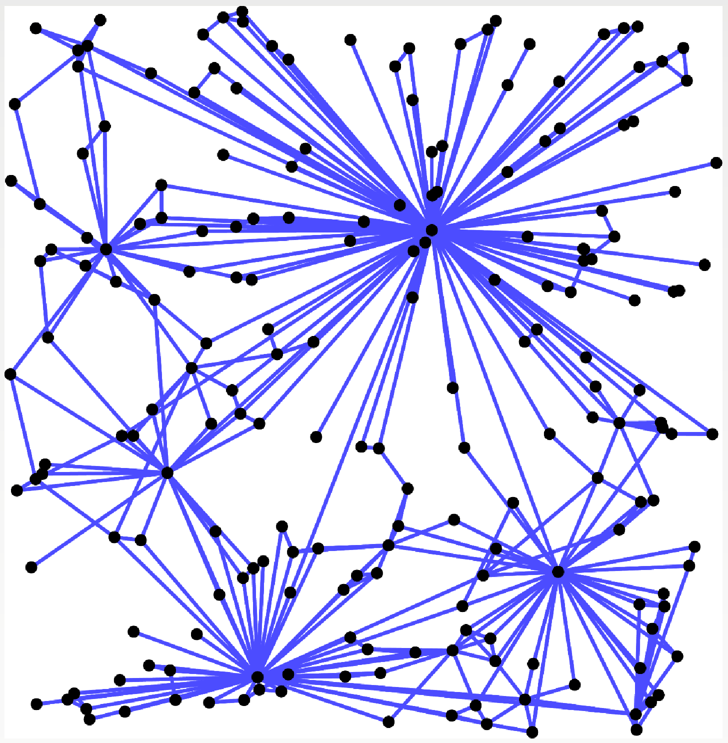
Redes de transporte: linhas aéreas



Redes de transporte



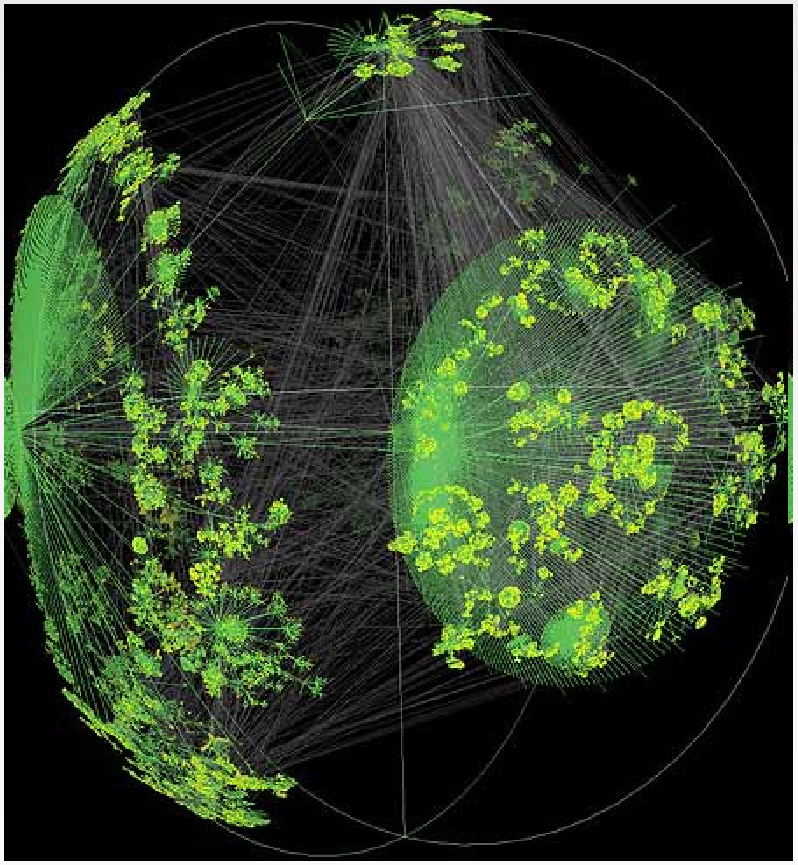
Estradas



Rotas Aéreas

Internet – mapa de IPs

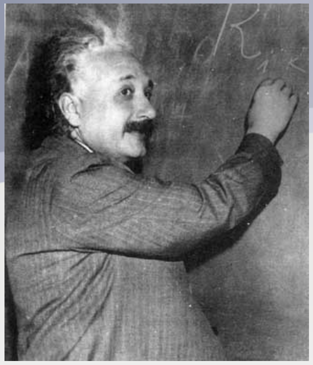
- Uma rede de computadores e roteadores
- Nós são máquinas físicas
- Arestas conexões entre máquinas



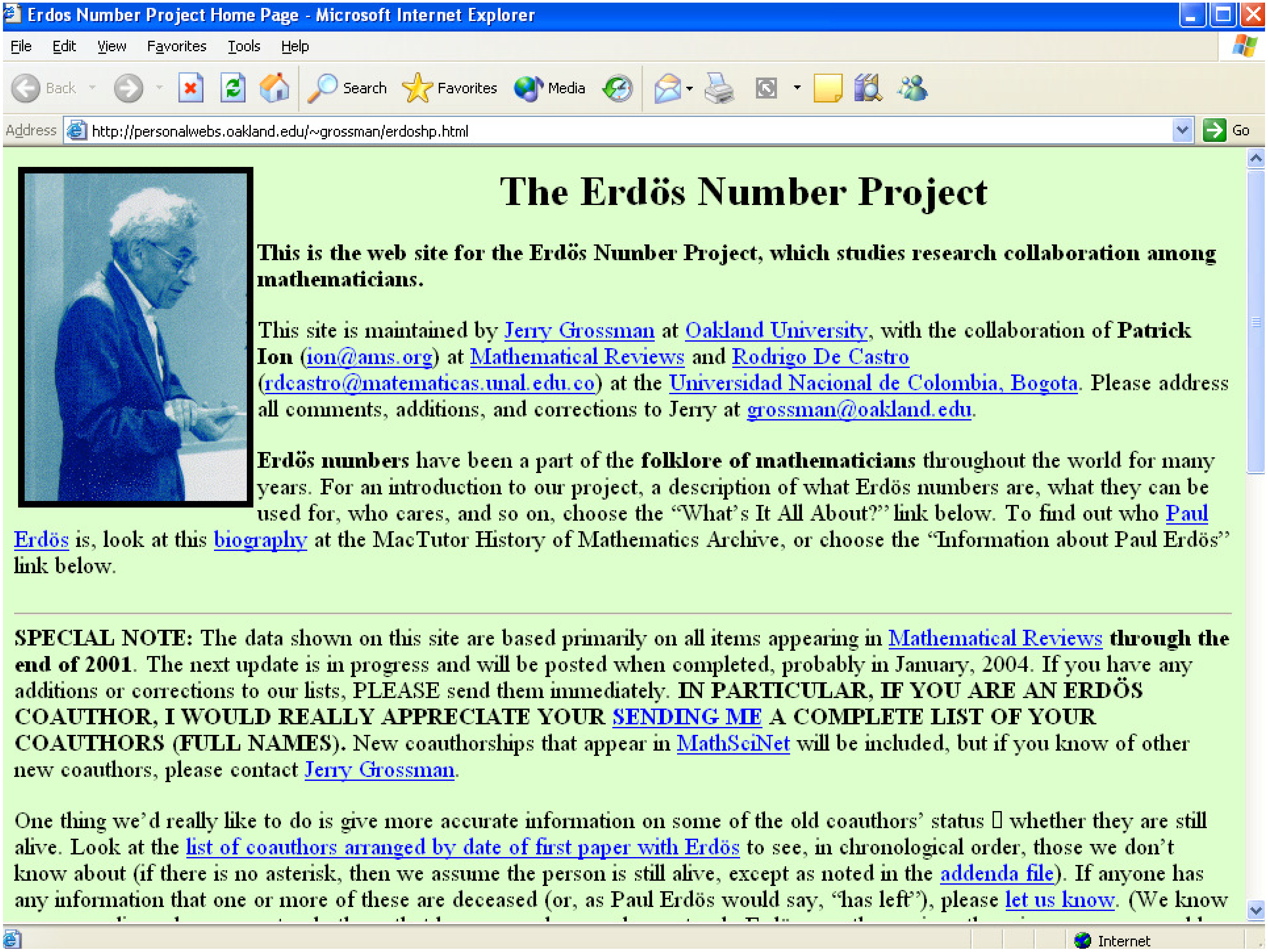
Rede de colaboração científica

- Paul Erdös (1913-1996)
  - Oliver Sacks: "A mathematical genius of the first order, Paul Erdös was totally obsessed with his subject - he thought and wrote mathematics for nineteen hours a day until the day he died. He traveled constantly, living out of a plastic bag, and had no interest in food, sex, companionship, art - all that is usually indispensable to a human life."
  - The Man Who Loved Only Numbers (Paul Hoffman, 1998)
  - Erdös publicou > 1,400 papers com > 500 co-autores durante sua vida

Números de Erdös de premios Nobel de física



Max von Laue	1914	4
Albert Einstein	**1921**	**2**
Niels Bohr	1922	5
Louis de Broglie	1929	5
Werner Heisenberg	1932	4
Paul A. Dirac	1933	4
Erwin Schrödinger	1933	8
Enrico Fermi	1938	3
Ernest O. Lawrence	1939	6
Otto Stern	1943	3
Isidor I. Rabi	1944	4
Wolfgang Pauli	1945	3
Frits Zernike	1953	6
Max Born	1954	3
Willis E. Lamb	1955	3
John Bardeen	1956	5
Walter H. Brattain	1956	6
William B. Shockley	1956	6
Chen Ning Yang	1957	4
Tsung-dao Lee	1957	5
Emilio Segrè	1959	4
Owen Chamberlain	1959	5
Robert Hofstadter	1961	5
Eugene Wigner	1963	4
Richard P. Feynman	1965	4
Julian S. Schwinger	1965	4
Hans A. Bethe	1967	4
Luis W. Alvarez	1968	6
Murray Gell-Mann	1969	3
John Bardeen	1972	5
Leon N. Cooper	1972	6
John R. Schrieffer	1972	5
Aage Bohr	1975	5
Ben Mottelson	1975	5
Leo J. Rainwater	1975	7
Steven Weinberg	1979	4
Sheldon Lee Glashow	1979	2
Abdus Salam	1979	3
S. Chandrasekhar	1983	4
Norman F. Ramsey	1989	3


Erdos Number Project Home Page - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://personalwebs.oakland.edu/~grossman/erdoshp.html
Go
Internet

The Erdös Number Project

This is the web site for the Erdös Number Project, which studies research collaboration among mathematicians.

This site is maintained by Jerry Grossman at Oakland University, with the collaboration of **Patrick Ion** (ion@ams.org) at Mathematical Reviews and Rodrigo De Castro (rdcastro@matematicas.unal.edu.co) at the Universidad Nacional de Colombia, Bogota. Please address all comments, additions, and corrections to Jerry at grossman@oakland.edu.

Erdös numbers have been a part of the **folklore of mathematicians** throughout the world for many years. For an introduction to our project, a description of what Erdös numbers are, what they can be used for, who cares, and so on, choose the "What's It All About?" link below. To find out who Paul Erdös is, look at this biography at the MacTutor History of Mathematics Archive, or choose the "Information about Paul Erdös" link below.

SPECIAL NOTE: The data shown on this site are based primarily on all items appearing in Mathematical Reviews **through the end of 2001**. The next update is in progress and will be posted when completed, probably in January, 2004. If you have any additions or corrections to our lists, PLEASE send them immediately. **IN PARTICULAR, IF YOU ARE AN ERDÖS COAUTHOR, I WOULD REALLY APPRECIATE YOUR SENDING ME A COMPLETE LIST OF YOUR COAUTHORS (FULL NAMES).** New coauthorships that appear in MathSciNet will be included, but if you know of other new coauthors, please contact Jerry Grossman.

One thing we'd really like to do is give more accurate information on some of the old coauthors' status whether they are still alive. Look at the list of coauthors arranged by date of first paper with Erdös to see, in chronological order, those we don't know about (if there is no asterisk, then we assume the person is still alive, except as noted in the addenda file). If anyone has any information that one or more of these are deceased (or, as Paul Erdös would say, "has left"), please let us know. (We know

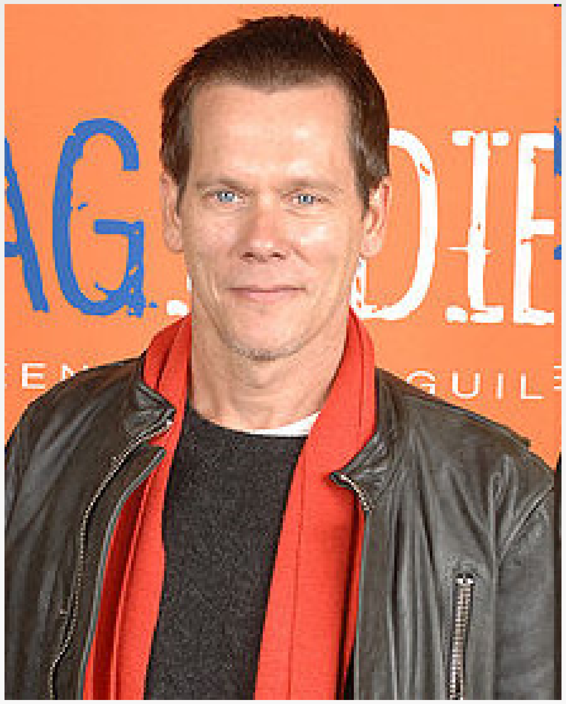
Rede de colaboração científica

- Meu número de Erdos = 4

Fabrício Benevenuto -> Miranda Mowbray -> Jonathan Jedwab -> Joe Gillis -> Paul Erdös

Fabrício Benevenuto -> Hamed Haddadi -> Andrew Thomason -> Fan Chung -> Paul Erdös

Rede de atores



Bacon number

Elvis Presley -> Edward Asner -> Kevin Bacon

Austin Powers: The spy who shagged me

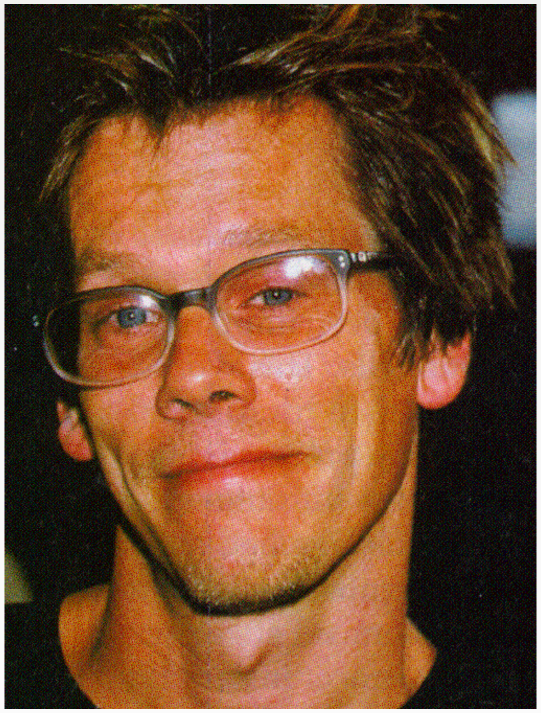
Robert Wagner

Let's make it legal

Wild Things

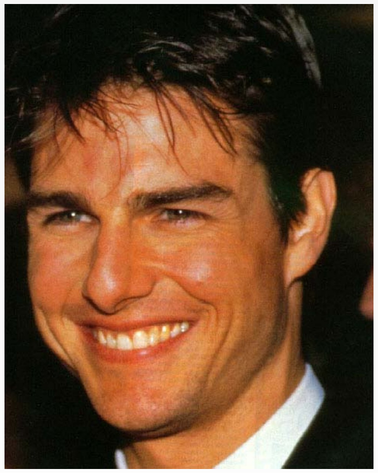
What Price Glory

Internet Movie Database



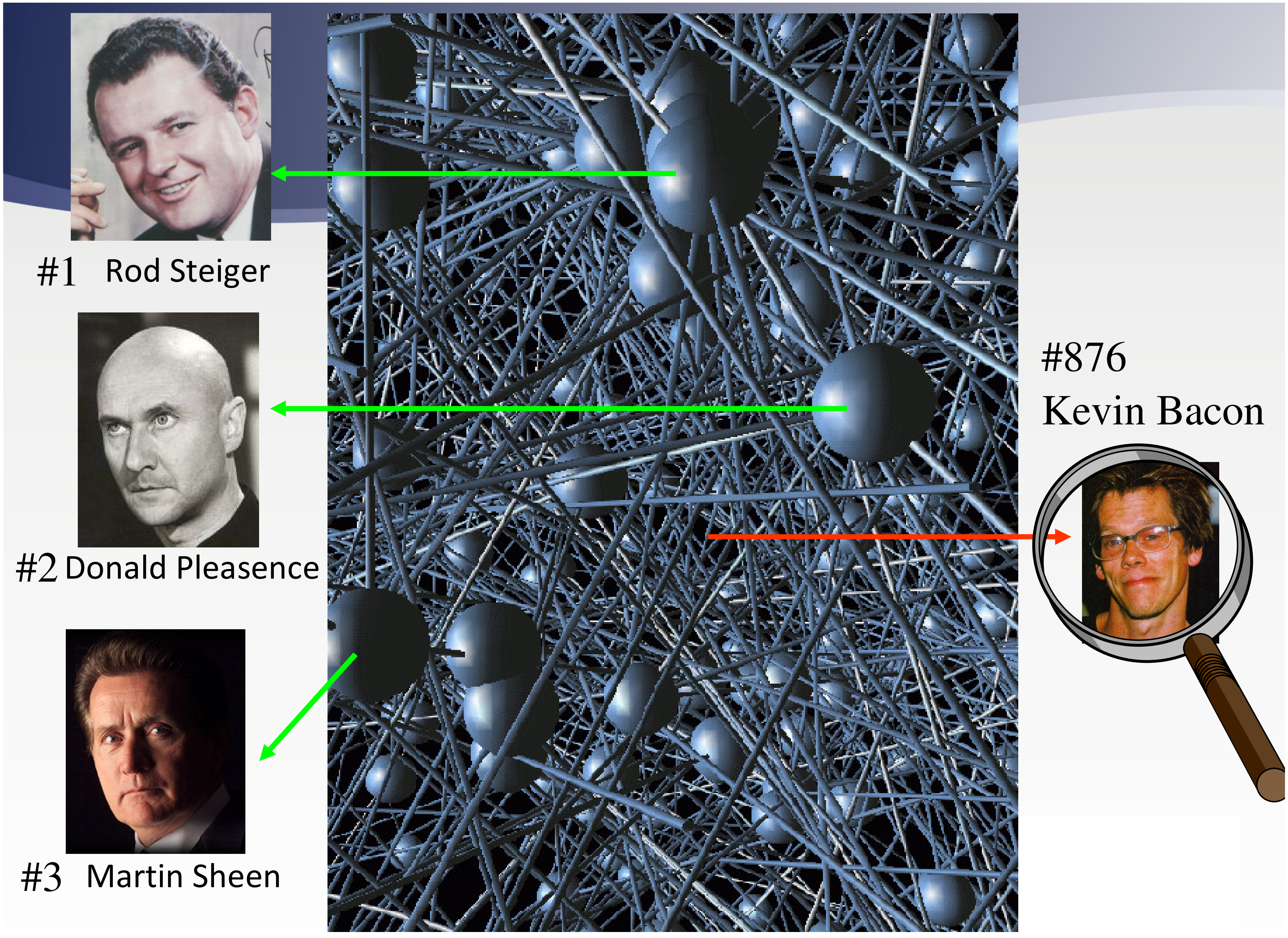
Barry Norton

A Few Good Man



Monsieur Verdoux

http://www.cs.virginia.edu/oracle/

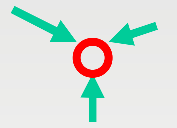

#1 Rod Steiger
#2 Donald Pleasence
#3 Martin Sheen
#876
Kevin Bacon

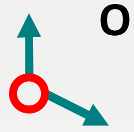
Métricas de redes

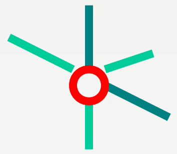
- Grau
- Coeficiente de Clusterização
- Componentes
- Distância Média
- PageRank

Grau dos nós

- In-degree = grau de entrada
- Out-degree = grau de saída
- Degree = grau



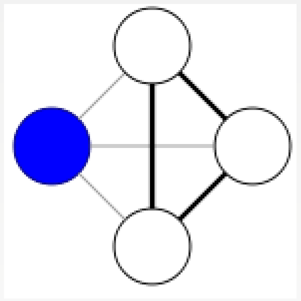




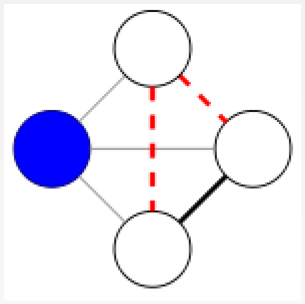
O que significa o grau no Orkut e no Twitter?

Coeficiente de clusterização

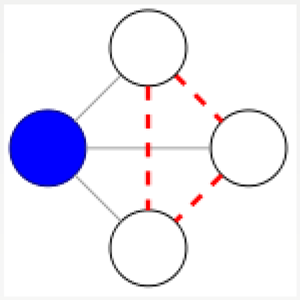
$$CC_i = \frac{\text{número de conexões entre os vizinhos de i}}{\text{número máximo de conexões possíveis entre os vizinhos de i}}$$



cc = 1



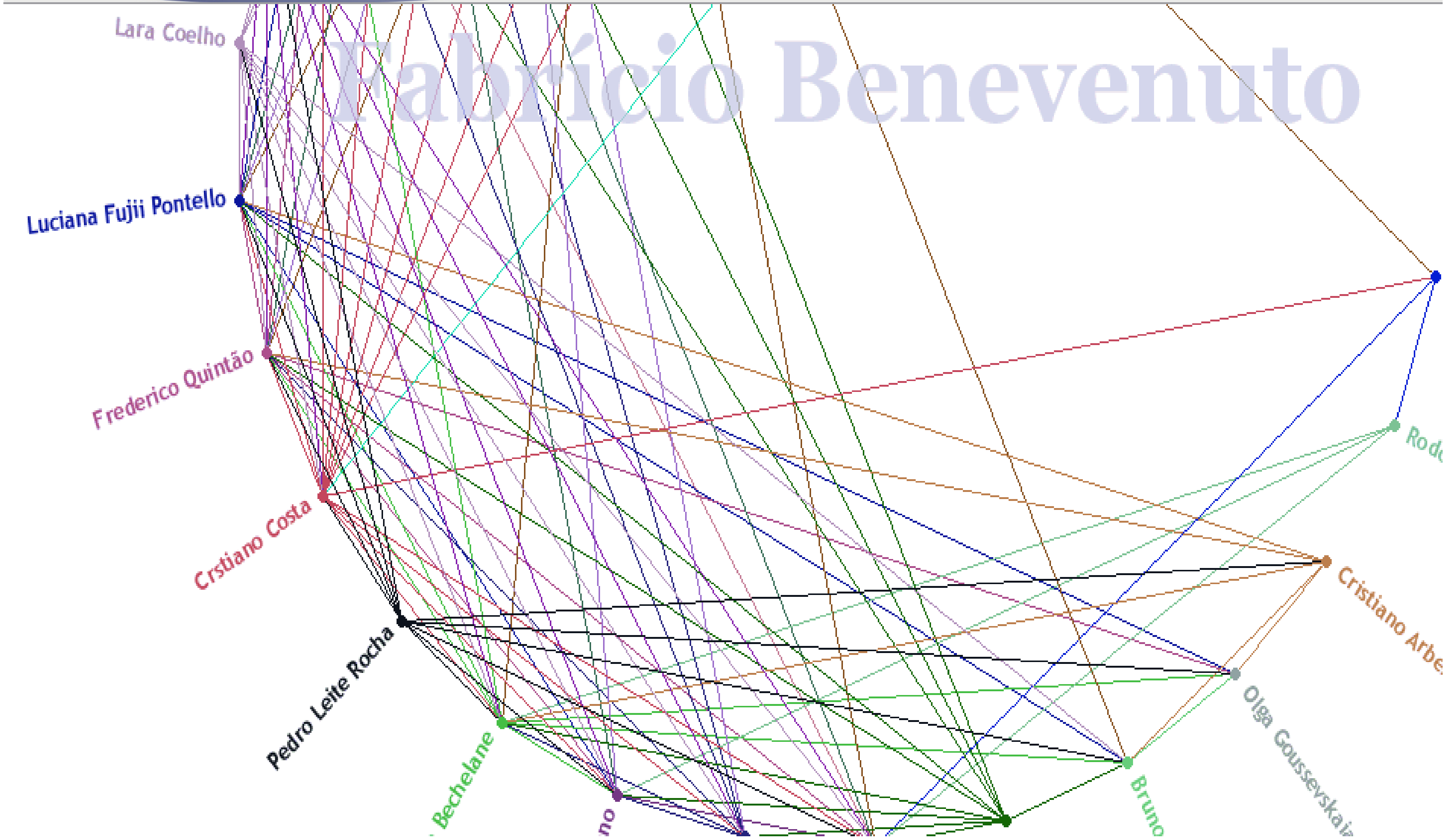
cc = 1/3



cc = 0

Os amigos dos spammers estão conectados entre si?

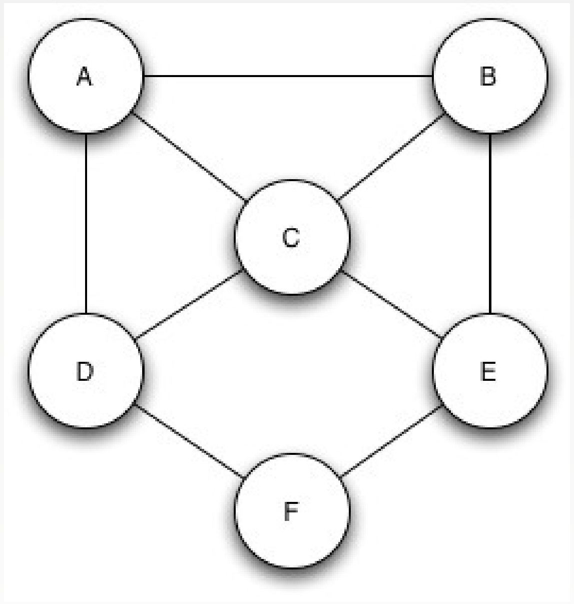
Facebook – Friend Weel



Coeficiente de Clusterização Global

Média sobre todos nós *n*

$$C = \frac{1}{n}\sum_i C_i$$



$CC_A = 2/3$

$CC_B = 2/3$

$CC_C = 1/2$

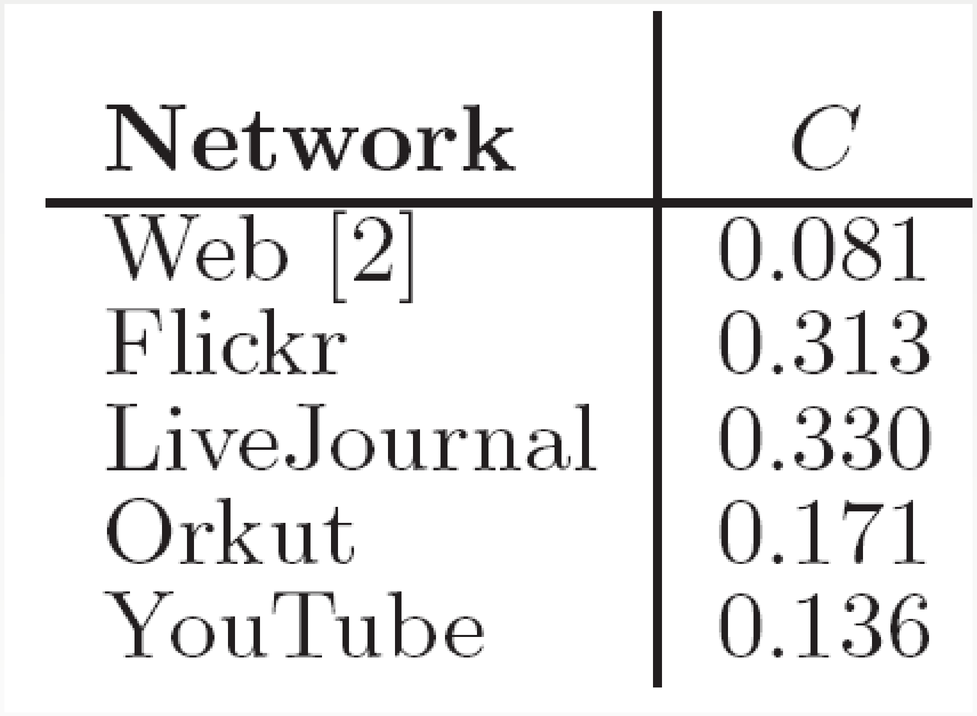
$CC_D = 1/3$

$CC_E = 1/3$

$CC_F = 0$

$CC = 5/12$

O que significa o coeficiente de clusterização?



Network	C
Web [2]	0.081
Flickr	0.313
LiveJournal	0.330
Orkut	0.171
YouTube	0.136

Componentes conectados

Componentes <u>fortemente</u> conectados: Strongly connected components (SCC): cada nó dentro do componente pode ser alcançado de outro nó do componente seguindo arestas orientadas.

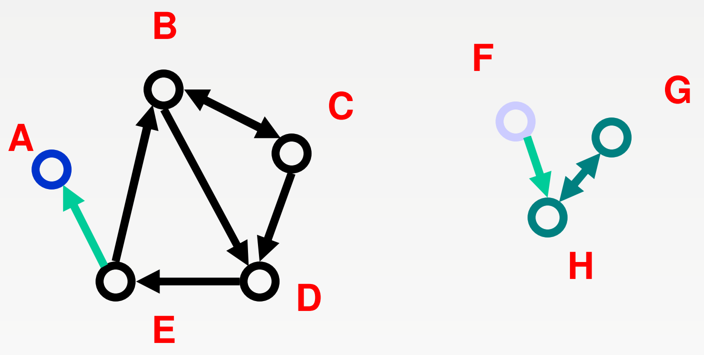
Componentes: SCC

B C D E

A

G H

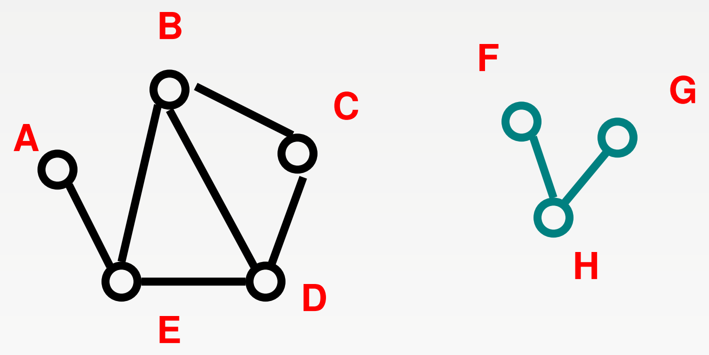
F



Métricas das Redes: componentes conectados

Componentes fracamente conectatos (Weakly connected components WCC): cada nó pode ser alcançado a partir de qualquer outro nó seguindo arestas em qualquer direção.

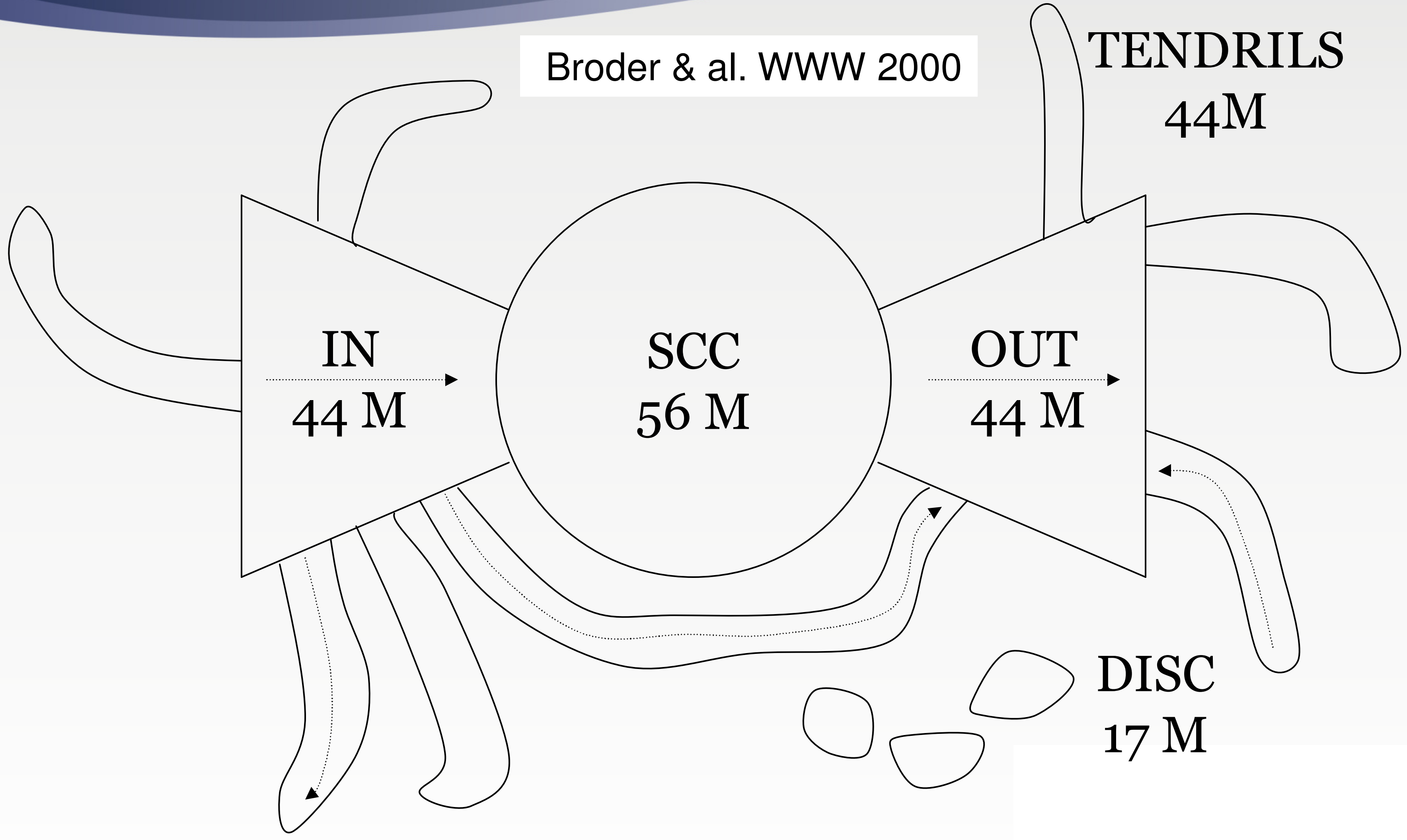
- WCC
    - **A B C D E**
    - **G H F**



Em redes não orientadas, simplesmente refere-se a componentes conectados

Estrutura Bow-tie da web

Broder & al. WWW 2000



Caminho mínimo - shortest paths

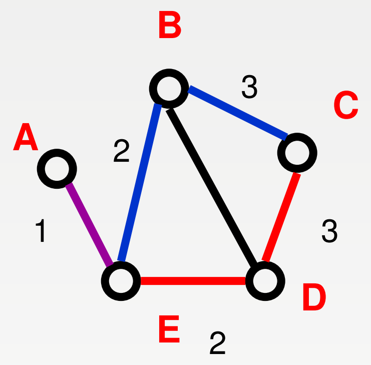
Caminho mínimo: a menor seqüência de arestas conectando dois nós.

Nem sempre única

A e C são conectados por 2 shortest paths

A – E – B - C

A – E – D - C



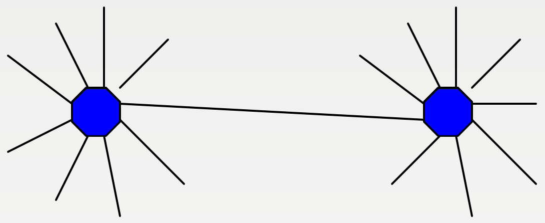
Diâmetro: a maior distância geodésica no grafo

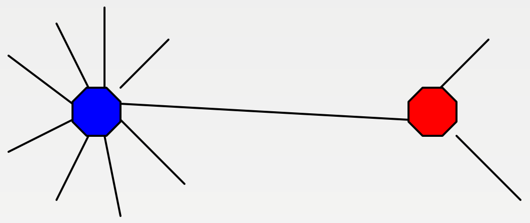
Diâmetro neste grafo = A-C = 3

Distância média: caminho mínimo médio entre todos os nós da rede

Assortatividade

Assortative networks **Disassortative networks**





•Redes reais sempre exibem uma das duas tendências,

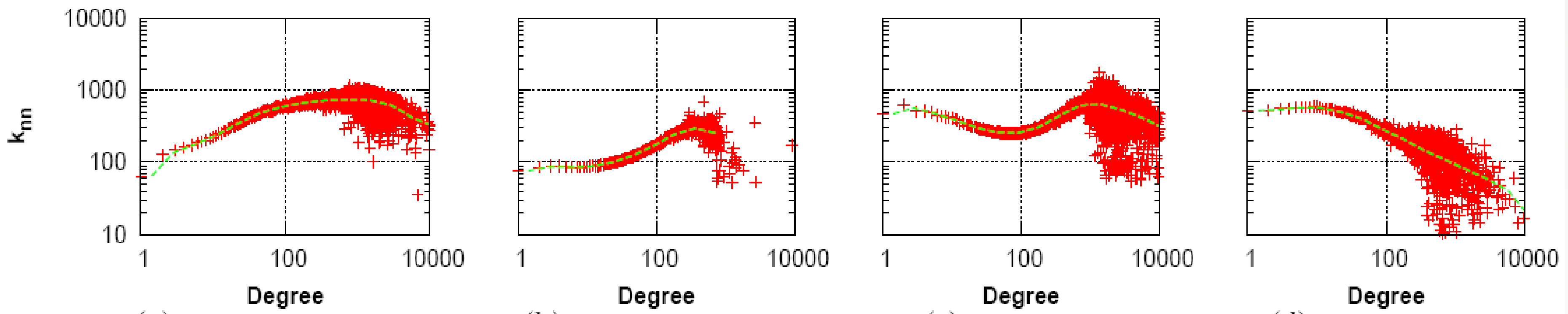
• redes “similares” exibem comportamentos “similares” .

⇧ ⇧

Social networks Techological networks

Assortatividade

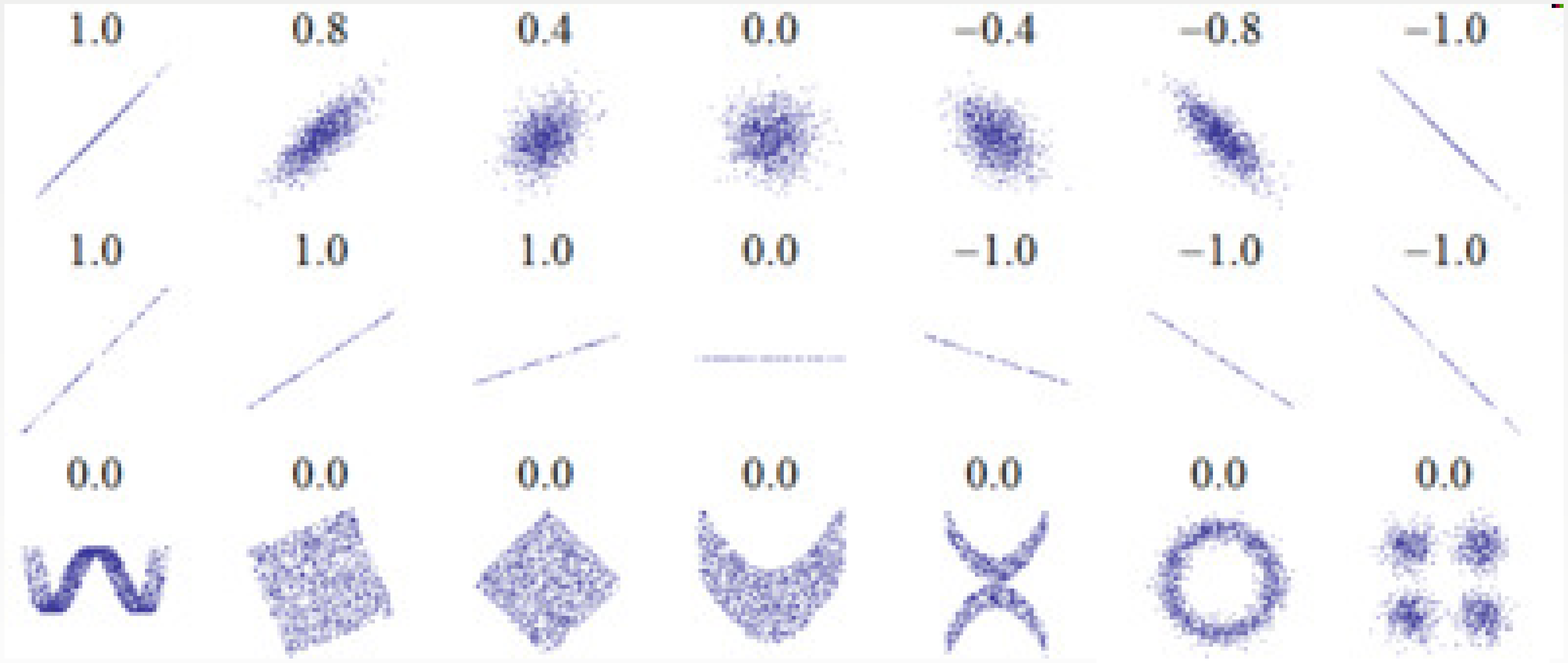
- Knn (k): média do grau dos vizinhos dos nós de grau k
- Se knn é uma função crescente de k: → Assortative
- Se knn é uma função decrescente de k: → Disassortative



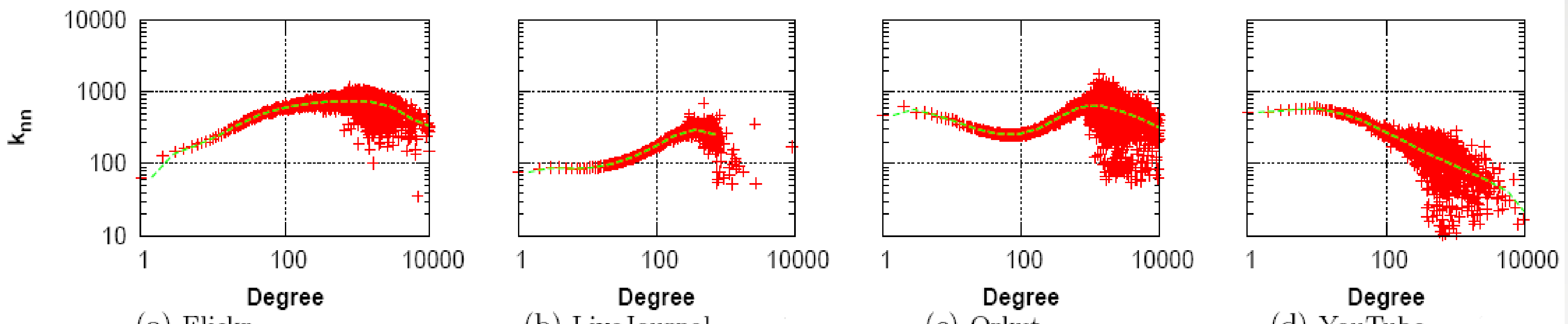
(a) Flickr (b) LiveJournal (c) Orkut (d) YouTube

Assortatividade

Coeficiente de Pearson (r): número que representa a função Knn (k).

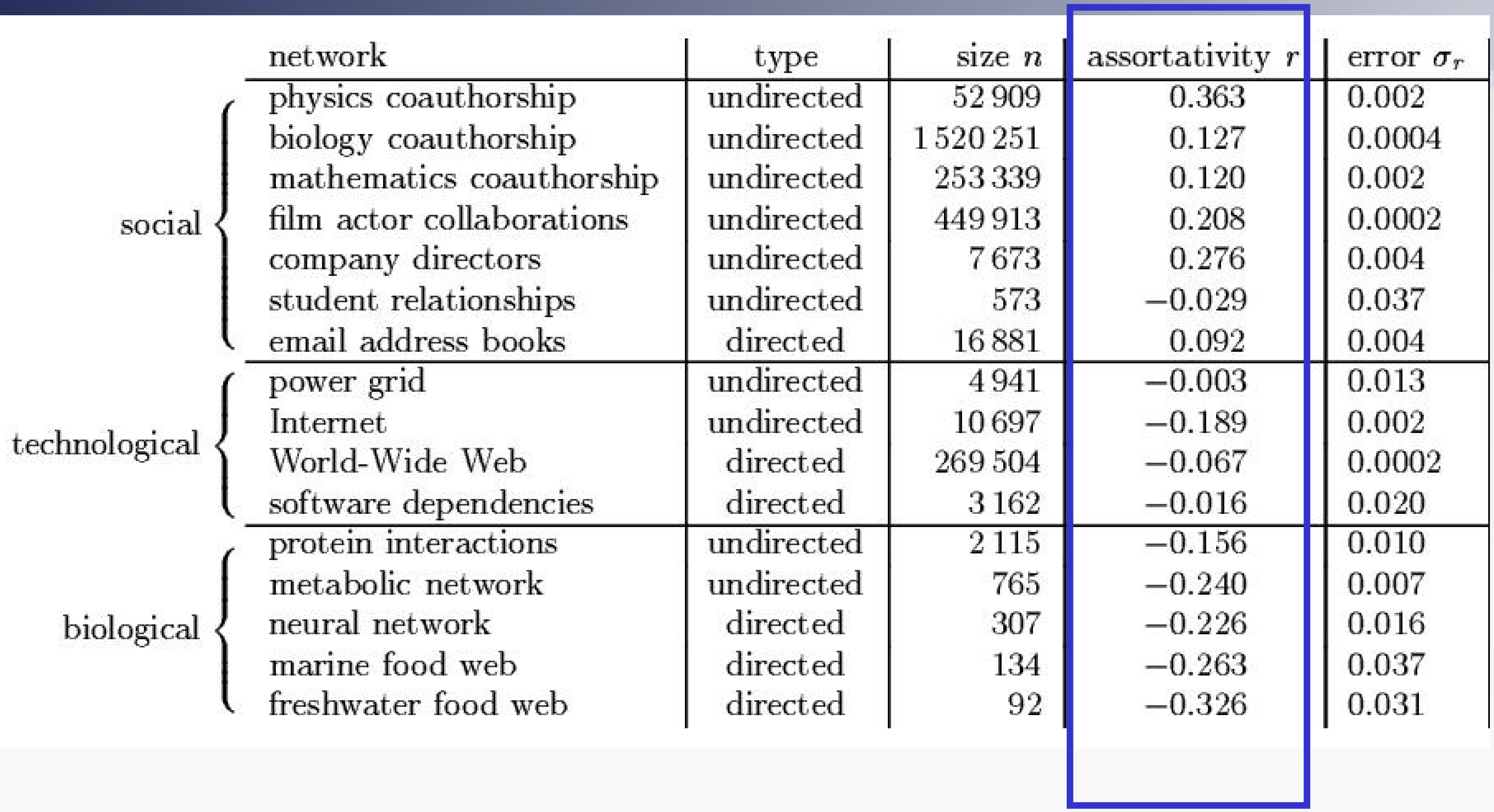


Assortatividade



(a) Flickr (b) LiveJournal (c) Orkut (d) YouTube

- Flckr r = 0.202
- LiveJournal r = 0.179
- Orkut r = 0.072
- YouTube r = -0.033
- Web r = -0.067
- Internet r = -0.189

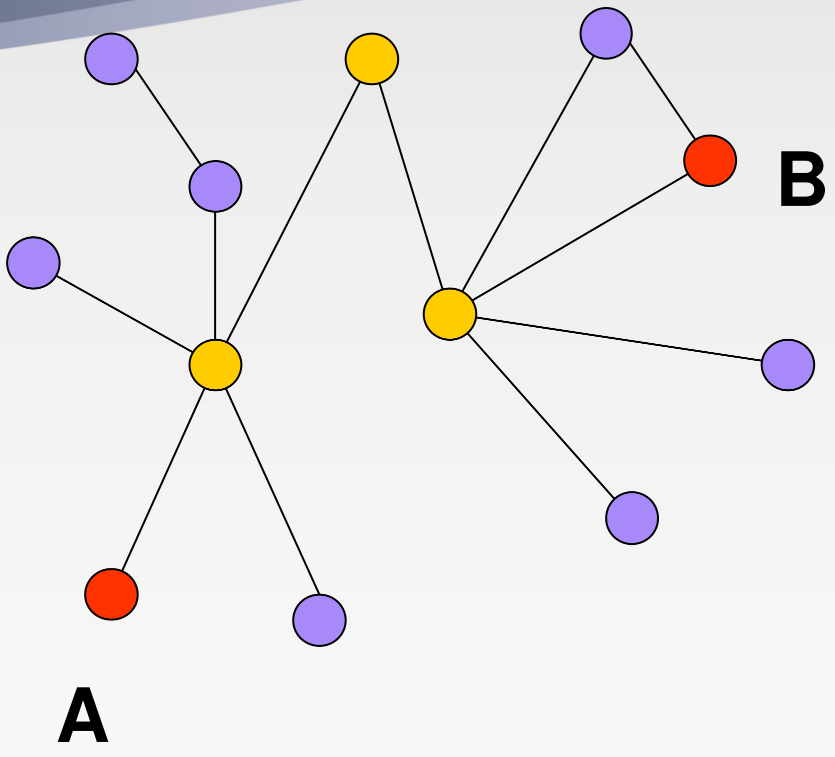


	network	type	size n	assortativity r	error σ_r
social	physics coauthorship	undirected	52 909	0.363	0.002
	biology coauthorship	undirected	1 520 251	0.127	0.0004
	mathematics coauthorship	undirected	253 339	0.120	0.002
	film actor collaborations	undirected	449 913	0.208	0.0002
	company directors	undirected	7 673	0.276	0.004
	student relationships	undirected	573	−0.029	0.037
	email address books	directed	16 881	0.092	0.004
technological	power grid	undirected	4 941	−0.003	0.013
	Internet	undirected	10 697	−0.189	0.002
	World-Wide Web	directed	269 504	−0.067	0.0002
	software dependencies	directed	3 162	−0.016	0.020
biological	protein interactions	undirected	2 115	−0.156	0.010
	metabolic network	undirected	765	−0.240	0.007
	neural network	directed	307	−0.226	0.016
	marine food web	directed	134	−0.263	0.037
	freshwater food web	directed	92	−0.326	0.031

Conseqüências da assortatividade: - Disseminação de Epidemias, Comunidades Isoladas....

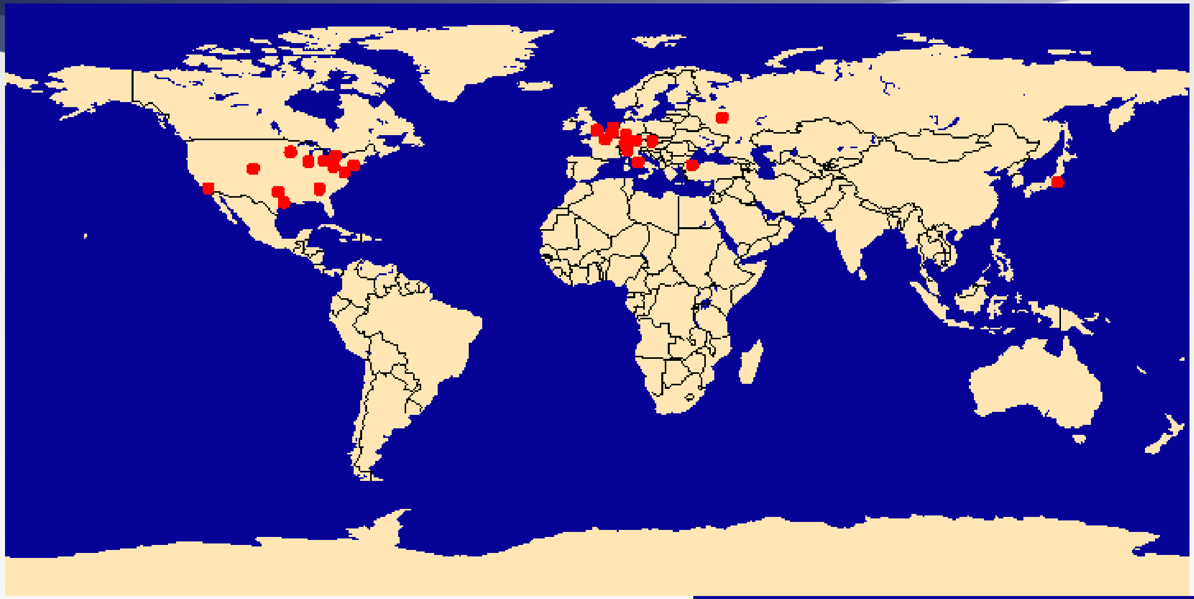
➢ Newman, *PRE,* bf 67 : 026126 , (2003).

Betweenness mede a centralidade de nós



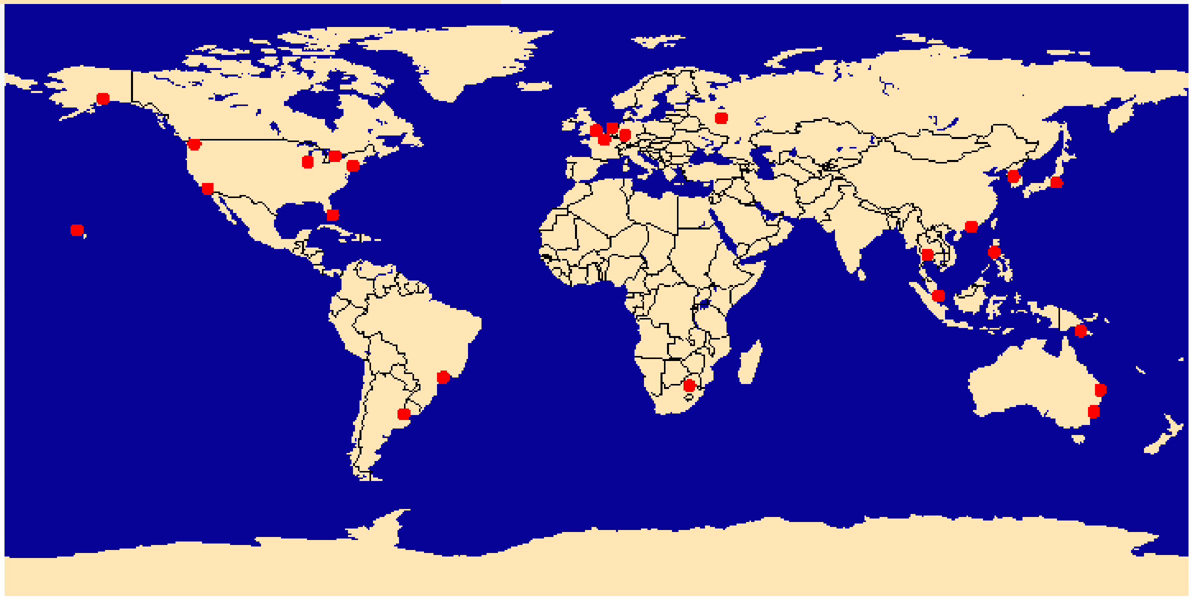
O betweenness b_i do nó i é o número de caminhos mínimos entre pares de nós que passa pelo nó i.

Aeroportos: cidades mais conectadas não são as mais centrais



Cidades Mais **conectadas**

Mais **centrais**



PageRank

PageRank e o Google

- Google foi fundada em 1998 por Larry Page e Sergey Brin

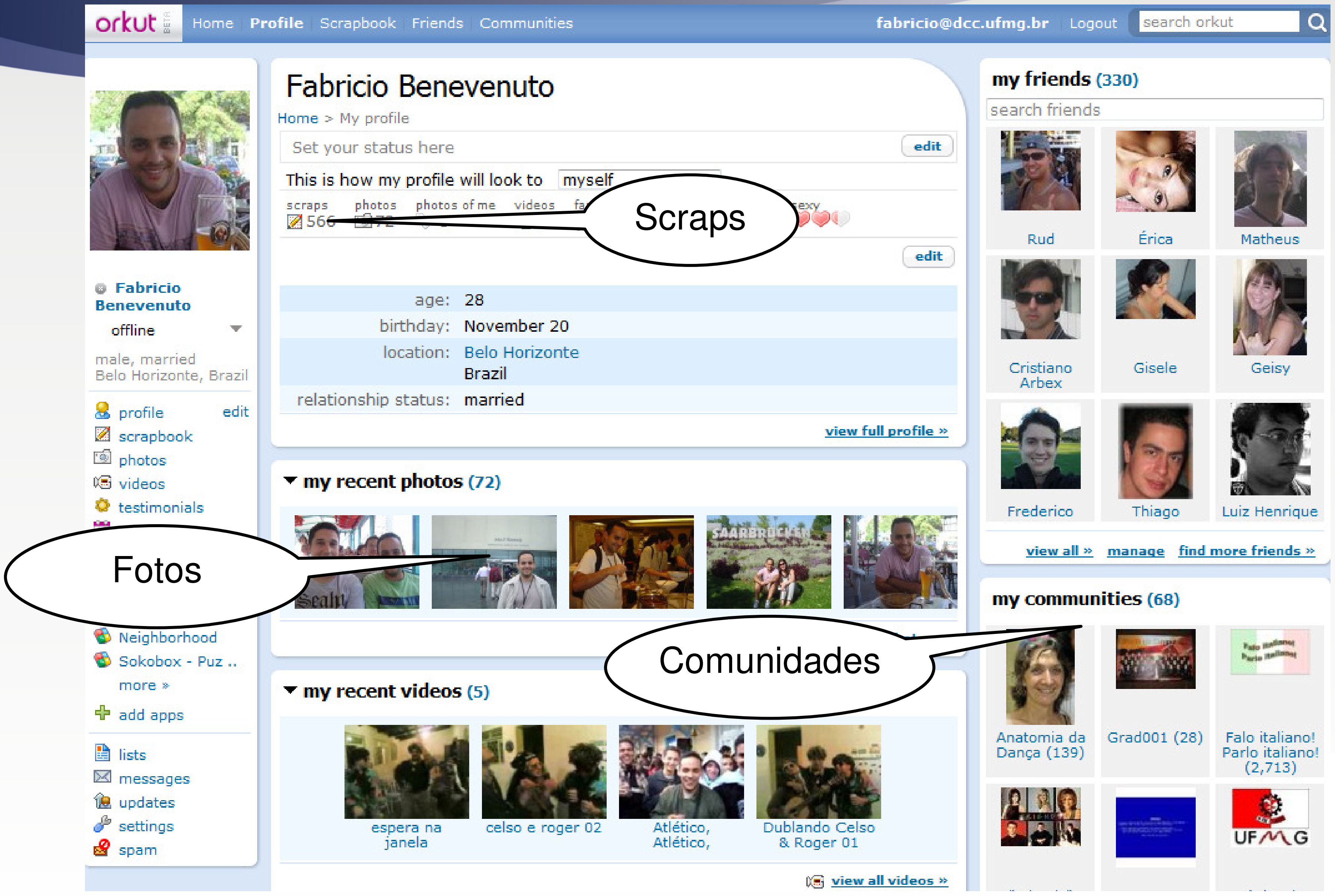
- Utiliza o pagerank para ordernar páginas de busca

- PageRank foi lançado em um artigo científico, parte de uma tese de doutorado em Stanford

Sistemas sociais populares

- Orkut
- Facebook
- YouTube
- Flickr
- Last FM
- Twitter
- Wikipedia

Orkut



Fabricio Benevenuto
offline
male, married
Belo Horizonte, Brazil

Fabricio Benevenuto

Home > My profile

Set your status here edit

This is how my profile will look to myself

scraps 566 photos 72 photos of me videos fa... sexy

edit

age: 28
birthday: November 20
location: Belo Horizonte
Brazil
relationship status: married

my recent photos (72)

my recent videos (5)

espera na janela
celso e roger 02
Atlético, Atlético,
Dublando Celso & Roger 01

my friends (330)

search friends

Rud
Érica
Matheus
Cristiano Arbex
Gisele
Geisy
Frederico
Thiago
Luiz Henrique

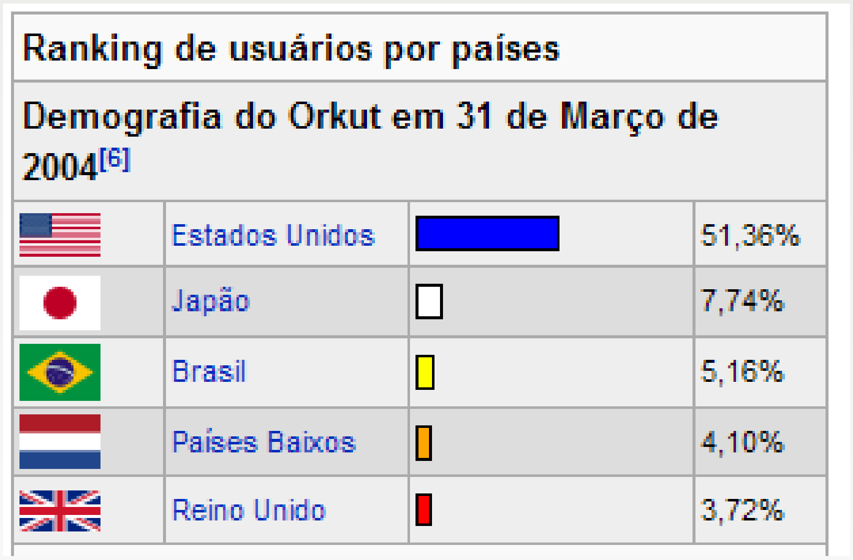
my communities (68)

Anatomia da Dança (139)
Grad001 (28)
Falo italiano! Parlo italiano! (2,713)

Como surgiu o Orkut?

- Rede Social do Google
- Criada por Orkut Buyukkokten

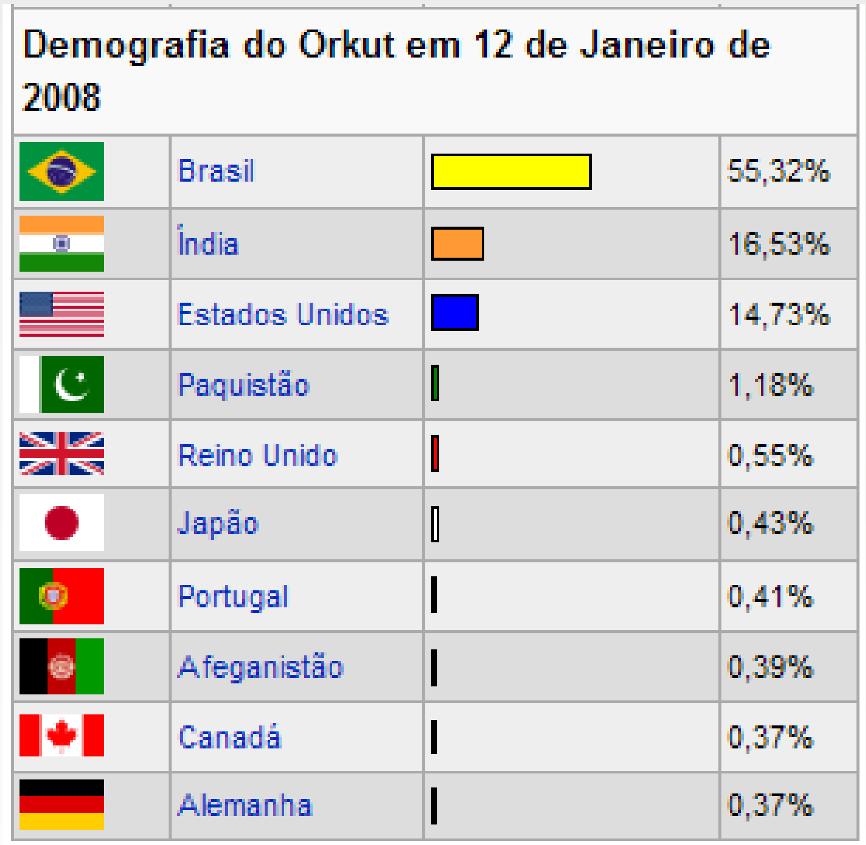
Popularidade do Orkut



Ranking de usuários por países

Demografia do Orkut em 31 de Março de 2004[6]

Estados Unidos	🟦	51,36%
Japão	⬜	7,74%
Brasil	🟨	5,16%
Países Baixos	🟧	4,10%
Reino Unido	🟥	3,72%



Demografia do Orkut em 12 de Janeiro de 2008

Brasil	🟨	55,32%
Índia	🟧	16,53%
Estados Unidos	🟦	14,73%
Paquistão	🟩	1,18%
Reino Unido	🟥	0,55%
Japão	⬜	0,43%
Portugal	▮	0,41%
Afeganistão	▮	0,39%
Canadá	▮	0,37%
Alemanha	▮	0,37%

Orkut no Brasil

- Por que deu tão certo no Brasil?
  - Fenômeno chamado no exterior de “*Brazilian Takeover*”
  - Efeito cascata? Se todo mundo tem orkut, quero ver o que é isso.
  - Característica da cultura brasileira?
  - Invasão da língua portuguesa
    - **Grande número de comunidades**
    - **Postagens em comunidades existentes**
- Primeira comunidade vendida
  - “Eu amo Floripa” – R$ 2.000,00

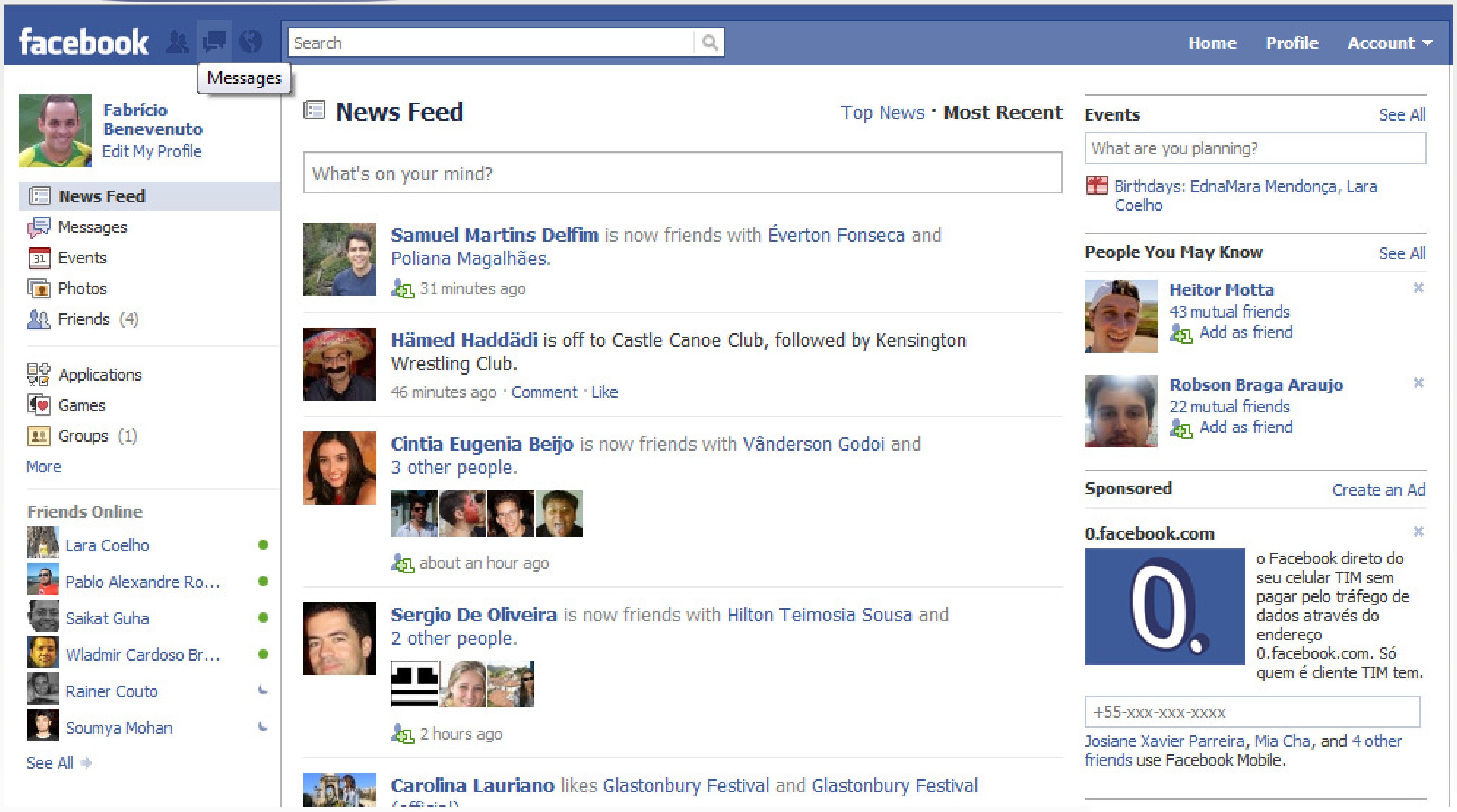
Orkut: Termo de adesão

- O que eles podem fazer com seus dados ?
  - O site passa a ser dono de absolutamente tudo o que você escreve e publica por lá (está no termo de adesão)
  - O termo de adesão diz que informações pessoais não serão vendidas, emprestadas ou alugadas
- O mesmo vale para vários outros sistemas
  - Humoristas não querem postar piadas no Twitter

Facebook

- Começou com acesso restrito
  - Primeiro Harvard, depois Stanford, Columbia e Yale
- Investimentos de capital de risco
  - Primeiro 500 mil, depois 12,7 milhões e depois 27,5 milhões
  - Microsoft comprou 1.6% do Facebook por 246 milhões (em 2007)
  - Foco da empresa não é vender
- Estatísticas de acesso
  - 500 milhões de usuários registrados

Facebook



Fabrício Benevenuto
Edit My Profile

News Feed
Messages
Events
Photos
Friends (4)

Applications
Games
Groups (1)
More

Friends Online
Lara Coelho
Pablo Alexandre Ro...
Saikat Guha
Wladmir Cardoso Br...
Rainer Couto
Soumya Mohan
See All

News Feed — Top News · Most Recent

What's on your mind?

Samuel Martins Delfim is now friends with Éverton Fonseca and Poliana Magalhães.
31 minutes ago

Hämed Haddädi is off to Castle Canoe Club, followed by Kensington Wrestling Club.
46 minutes ago · Comment · Like

Cintia Eugenia Beijo is now friends with Vânderson Godoi and 3 other people.
about an hour ago

Sergio De Oliveira is now friends with Hilton Teimosia Sousa and 2 other people.
2 hours ago

Carolina Lauriano likes Glastonbury Festival and Glastonbury Festival

Events — See All
What are you planning?
Birthdays: EdnaMara Mendonça, Lara Coelho

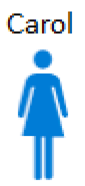
People You May Know — See All
Heitor Motta
43 mutual friends
Add as friend
Robson Braga Araujo
22 mutual friends
Add as friend

Sponsored — Create an Ad
0.facebook.com
o Facebook direto do seu celular TIM sem pagar pelo tráfego de dados através do endereço 0.facebook.com. Só quem é cliente TIM tem.
+55-xxx-xxx-xxxx
Josiane Xavier Parreira, Mia Cha, and 4 other friends use Facebook Mobile.

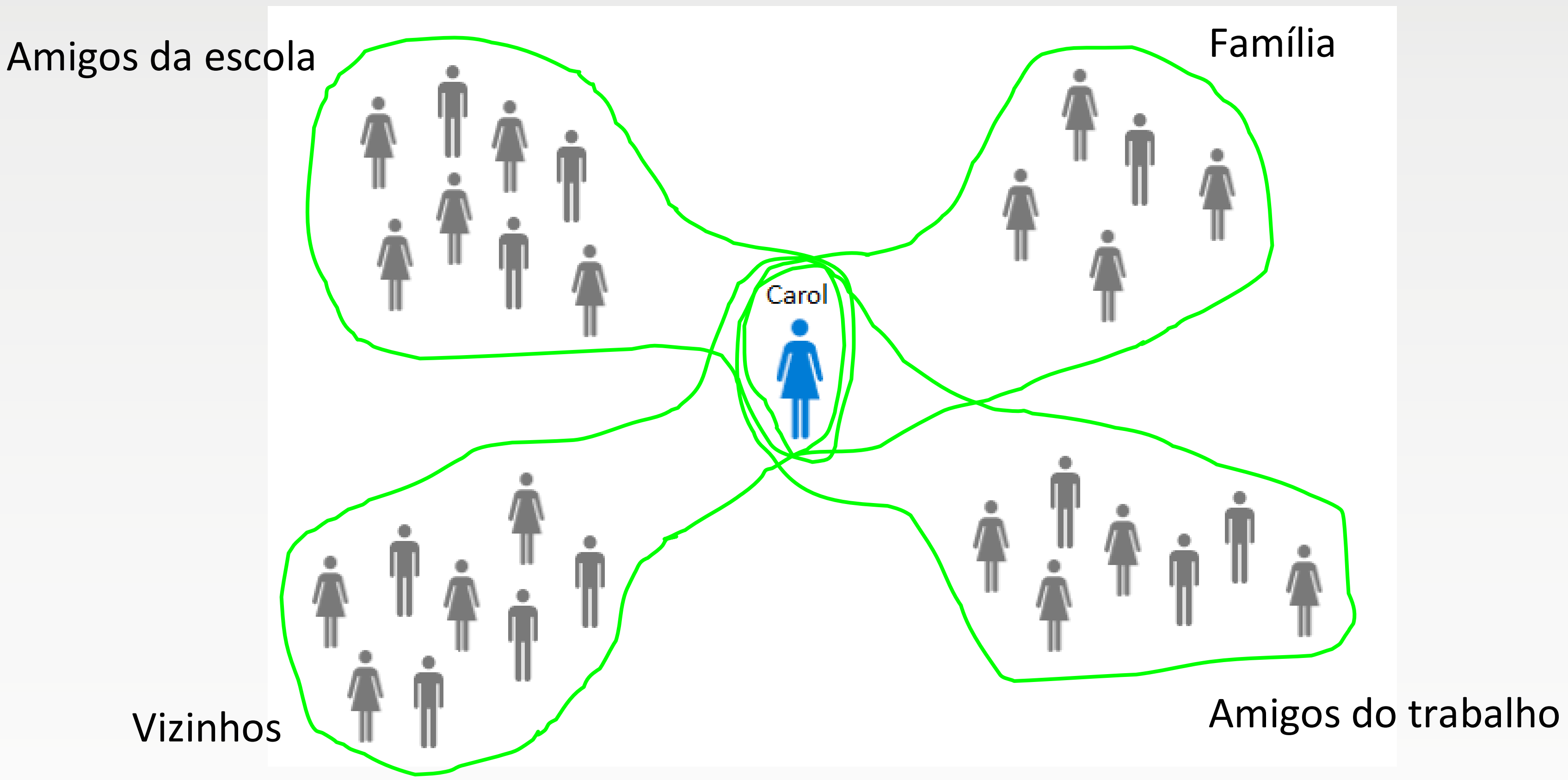
Problemas no Orkut e Facebook

- Spam, propagandas, phishing
- Usuários falsos
  - Celebridades ou não
- Comunidades ofensivas
  - Apologia às drogas, racismo, terrorismo, etc.
- Privacidade
  - Fotos postadas por amigos com tags
  - Mesmo que somente amigos possam acessar conteúdo

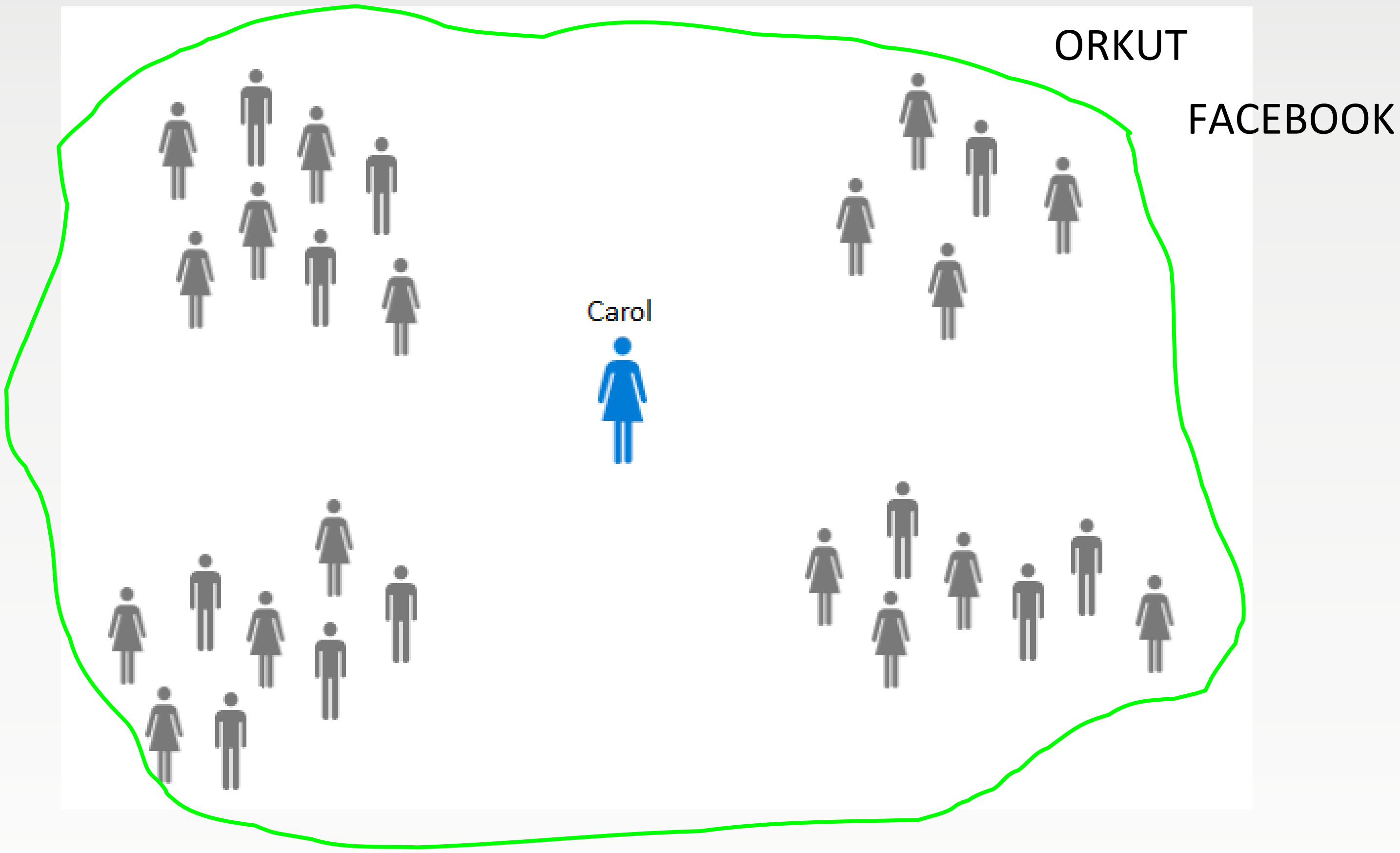
Exposição em redes sociais



Exposição em redes sociais



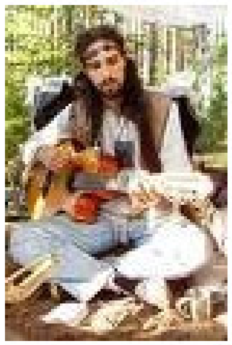
Exposição em redes sociais



Exposição em redes sociais

- Fotos bem identificadas (com tags)
  - Eu, João, José, etc.
- Comunidades indicando lugares onde estudou, gostos particulares, etc

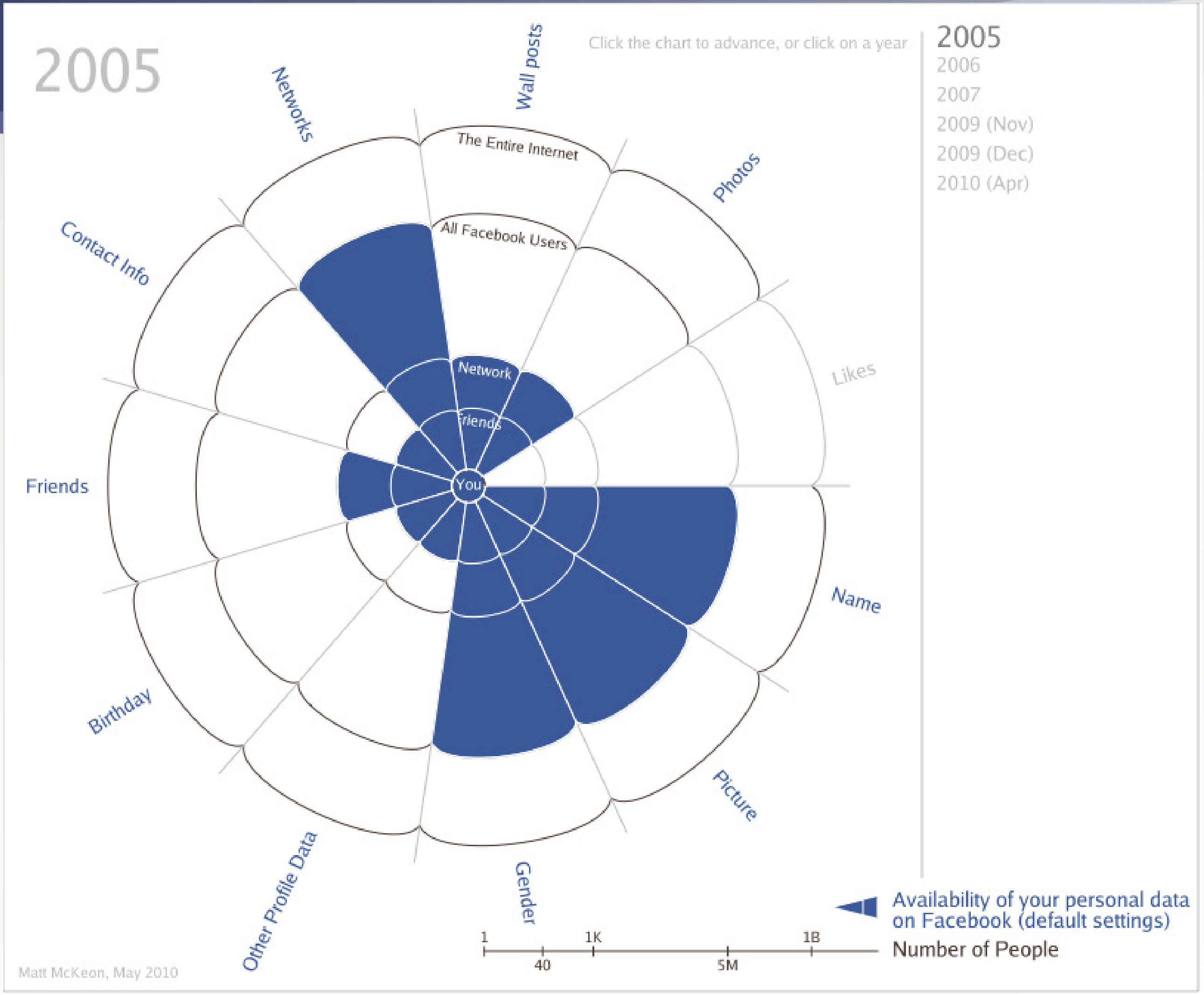
EU ODEIO MEU CHEFE


Se não der certo! viro hippie

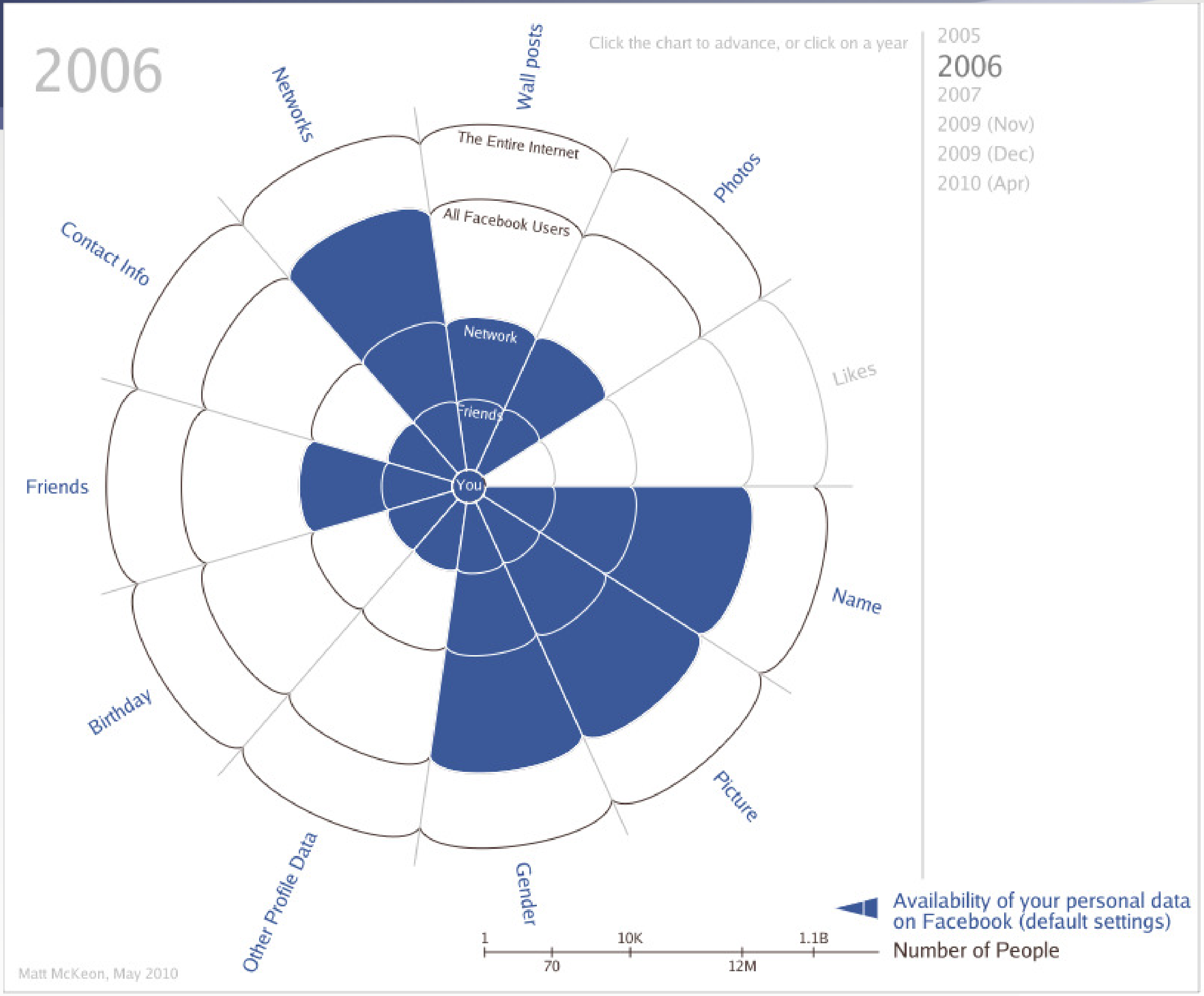
Eu Odeio Acordar Cedo
(5,967,619 members)

Tenho medo da Gina dos palitos

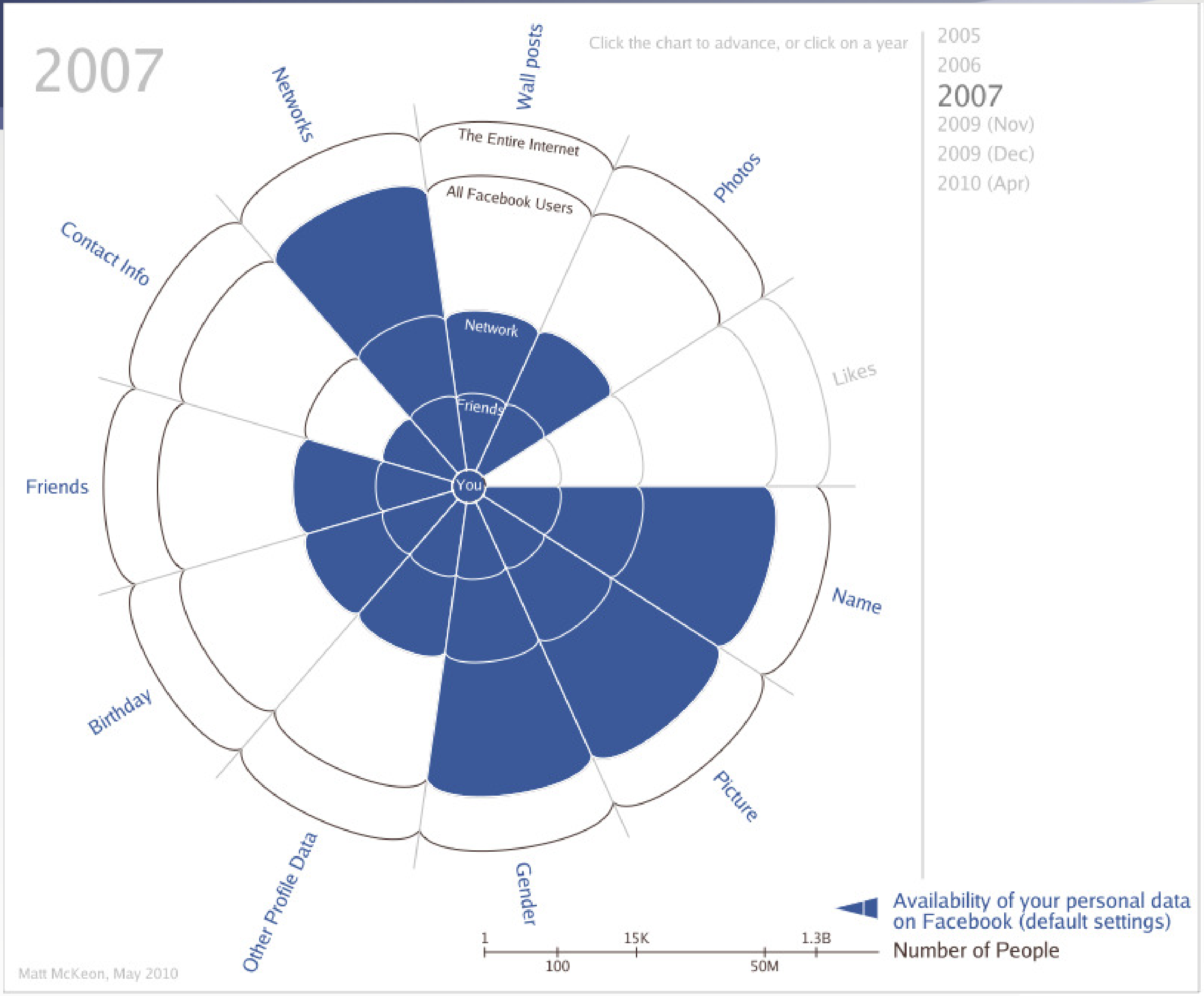
Privacidade no Facebook



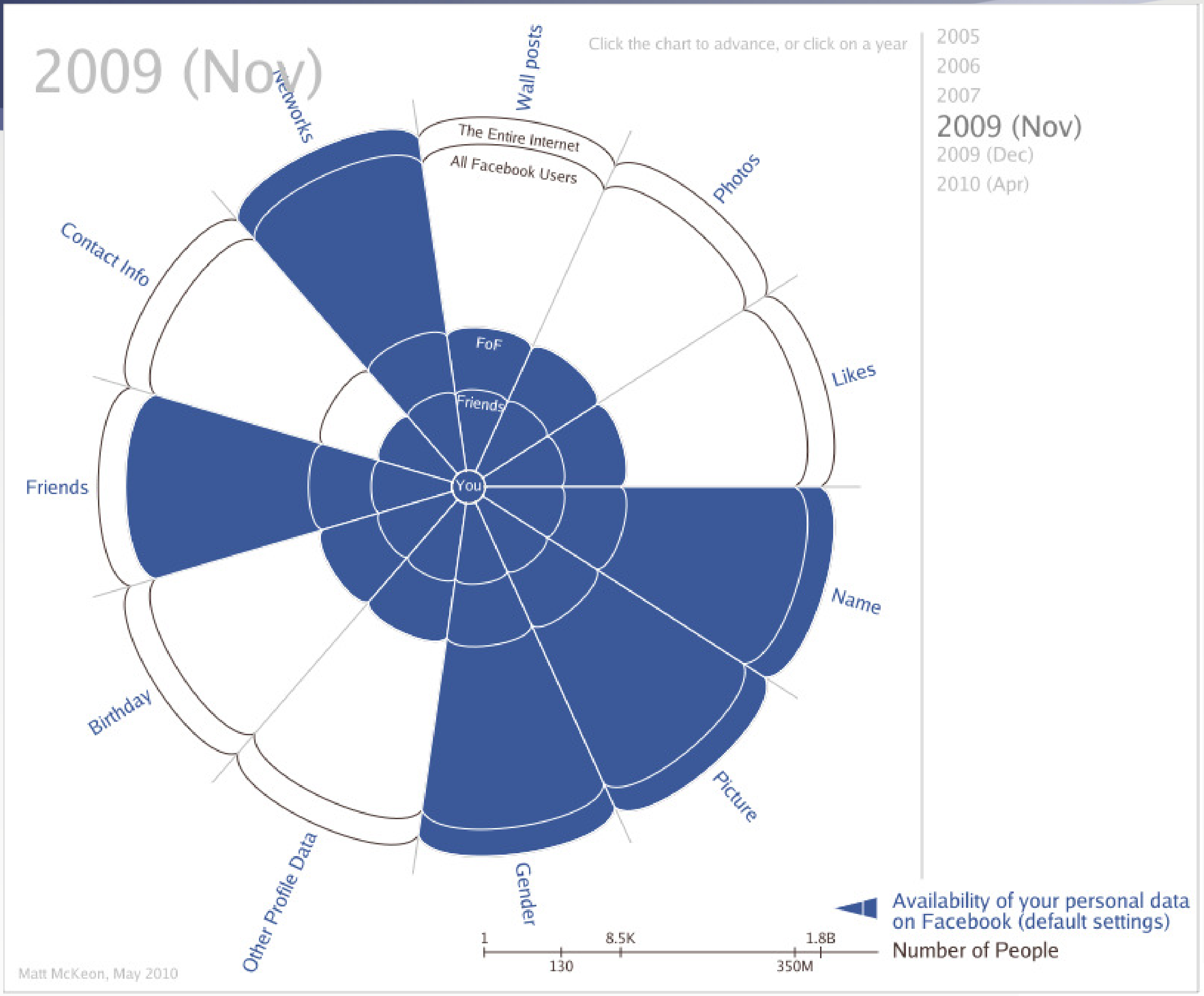
Privacidade no Facebook



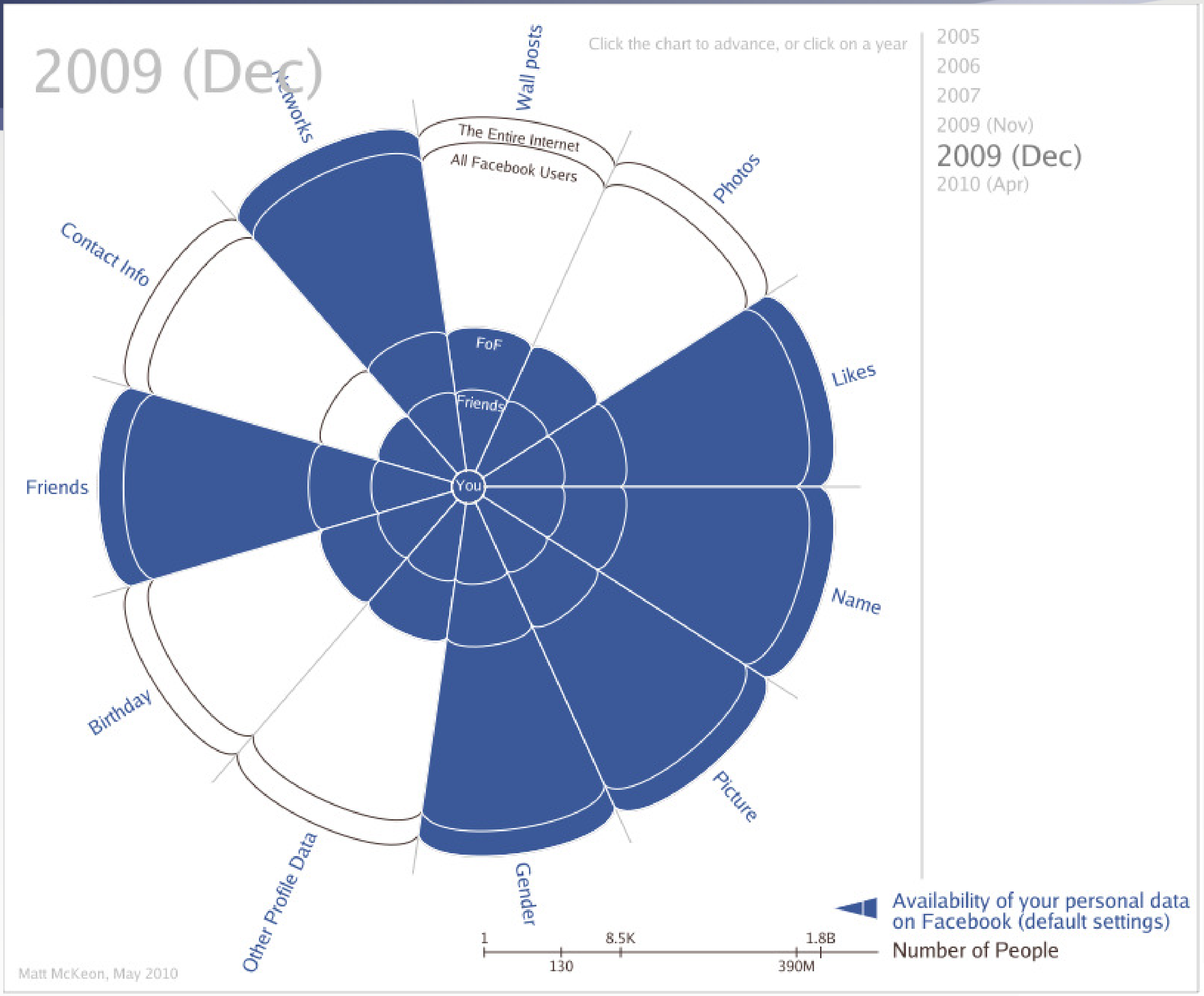
Privacidade no Facebook



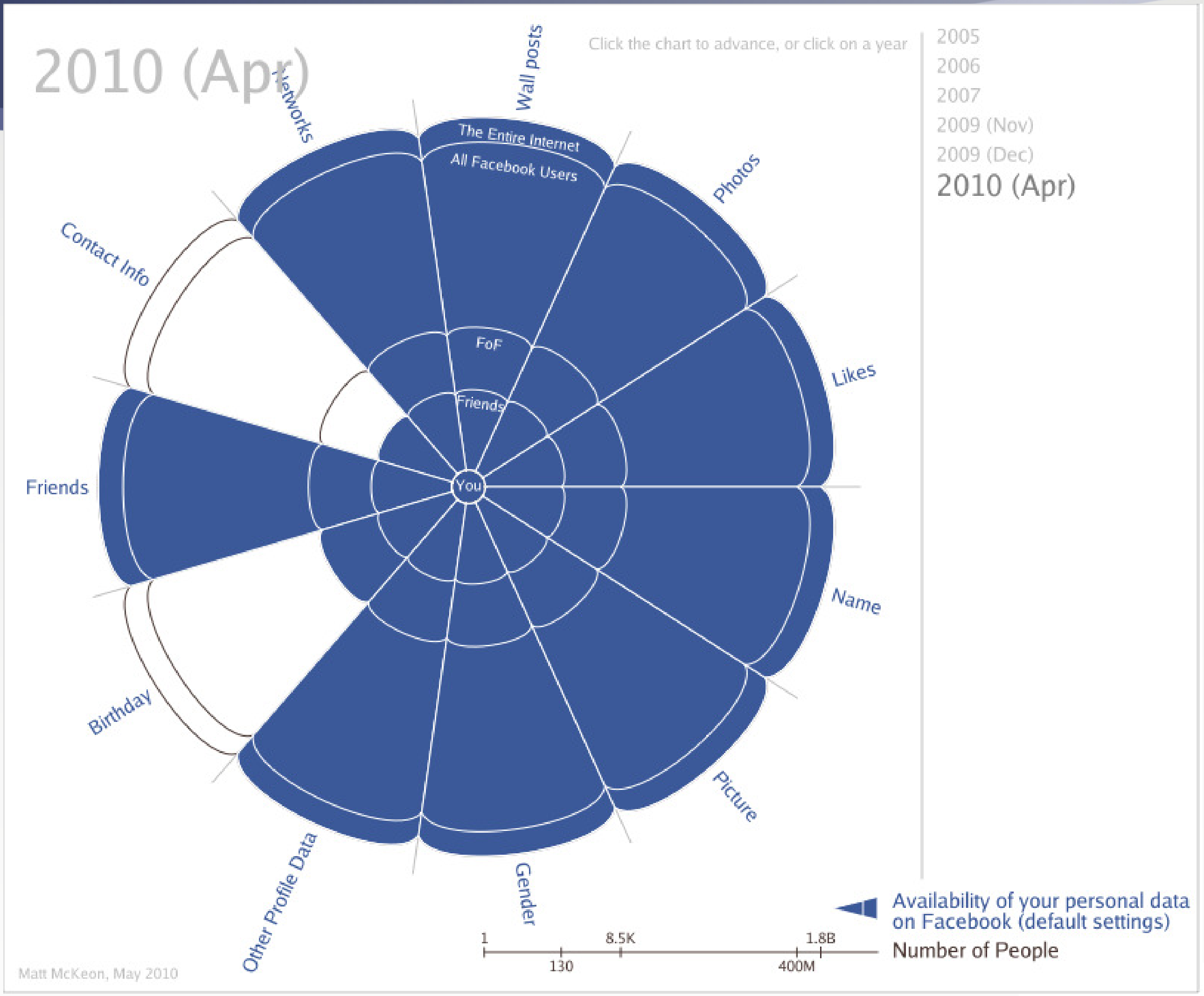
Privacidade no Facebook



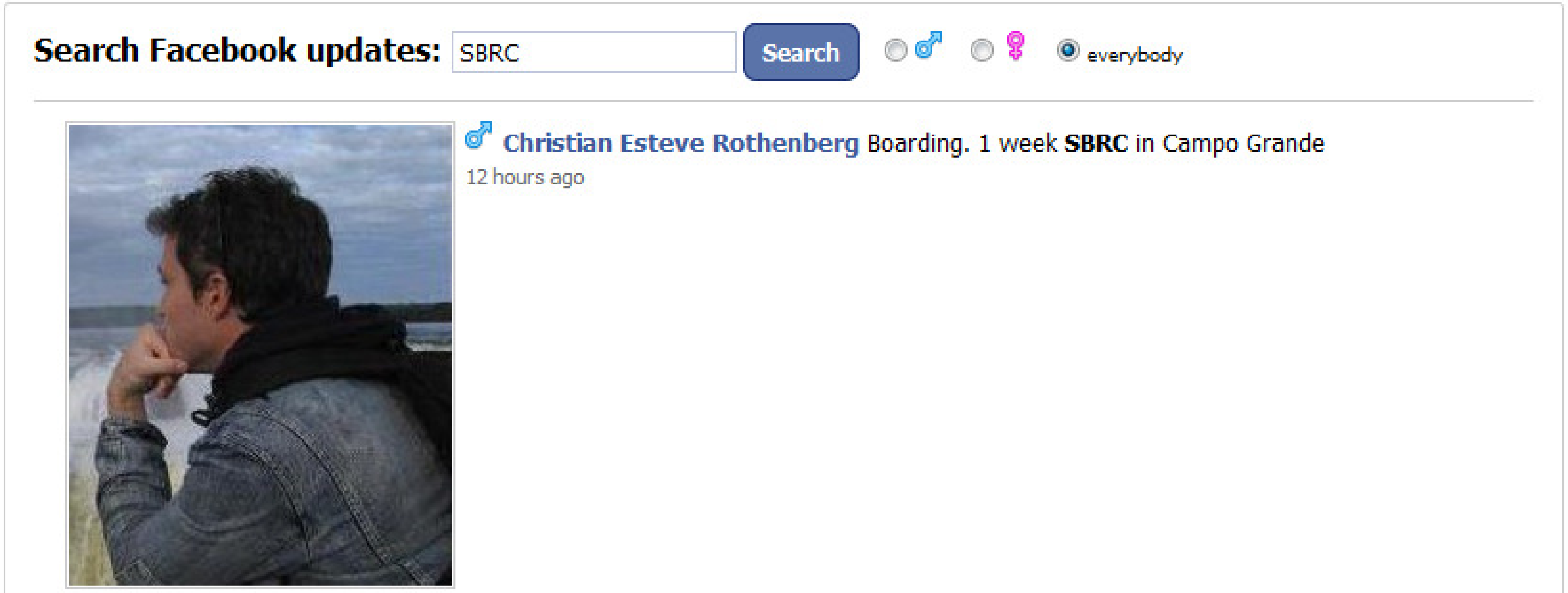
Privacidade no Facebook



Privacidade no Facebook



Openbook http://youropenbook.org/

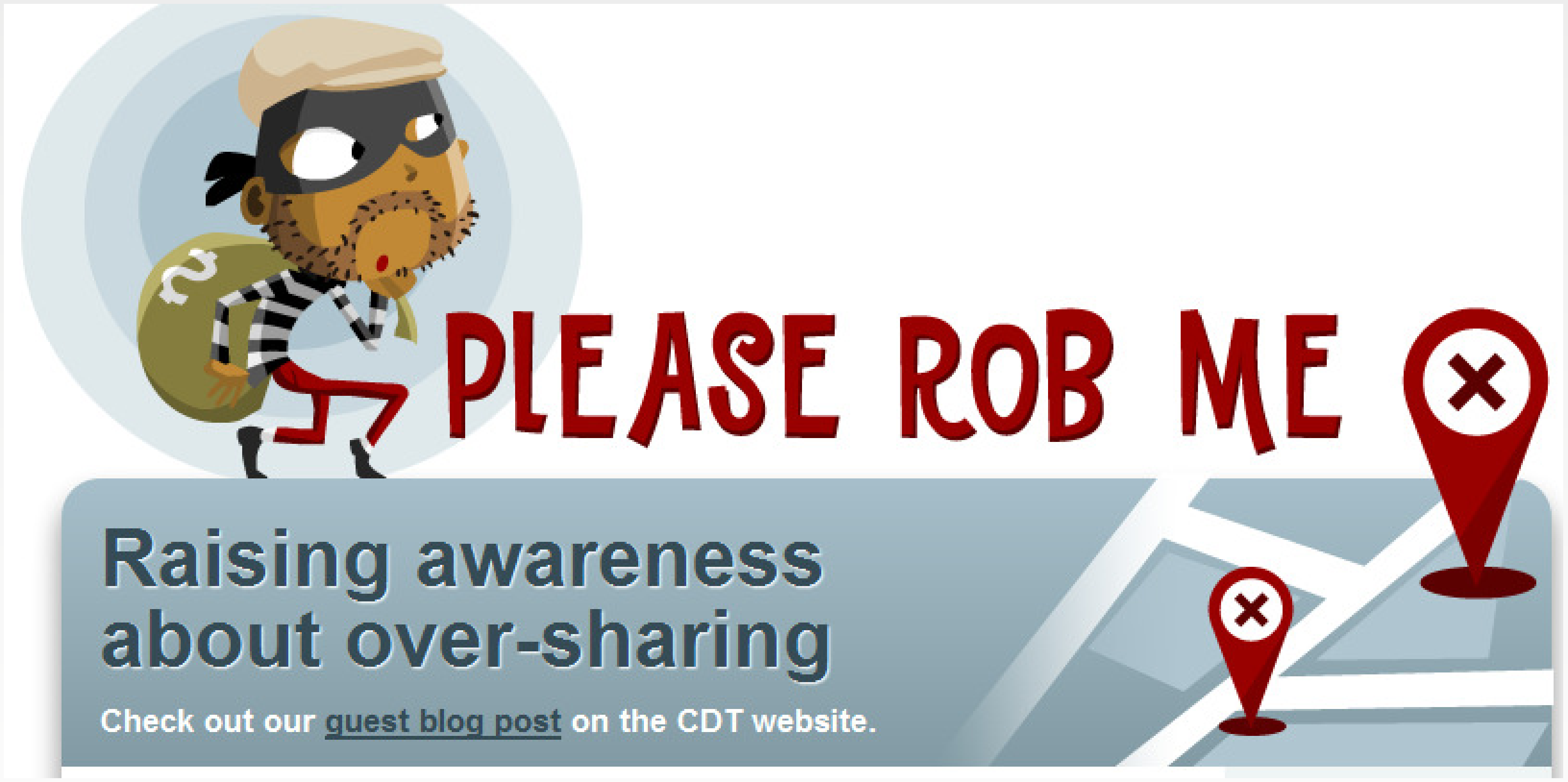


Search Facebook updates: SBRC

Christian Esteve Rothenberg Boarding. 1 week **SBRC** in Campo Grande
12 hours ago

Please rob me

http://pleaserobme.com/



Raising awareness about over-sharing

Check out our guest blog post on the CDT website.

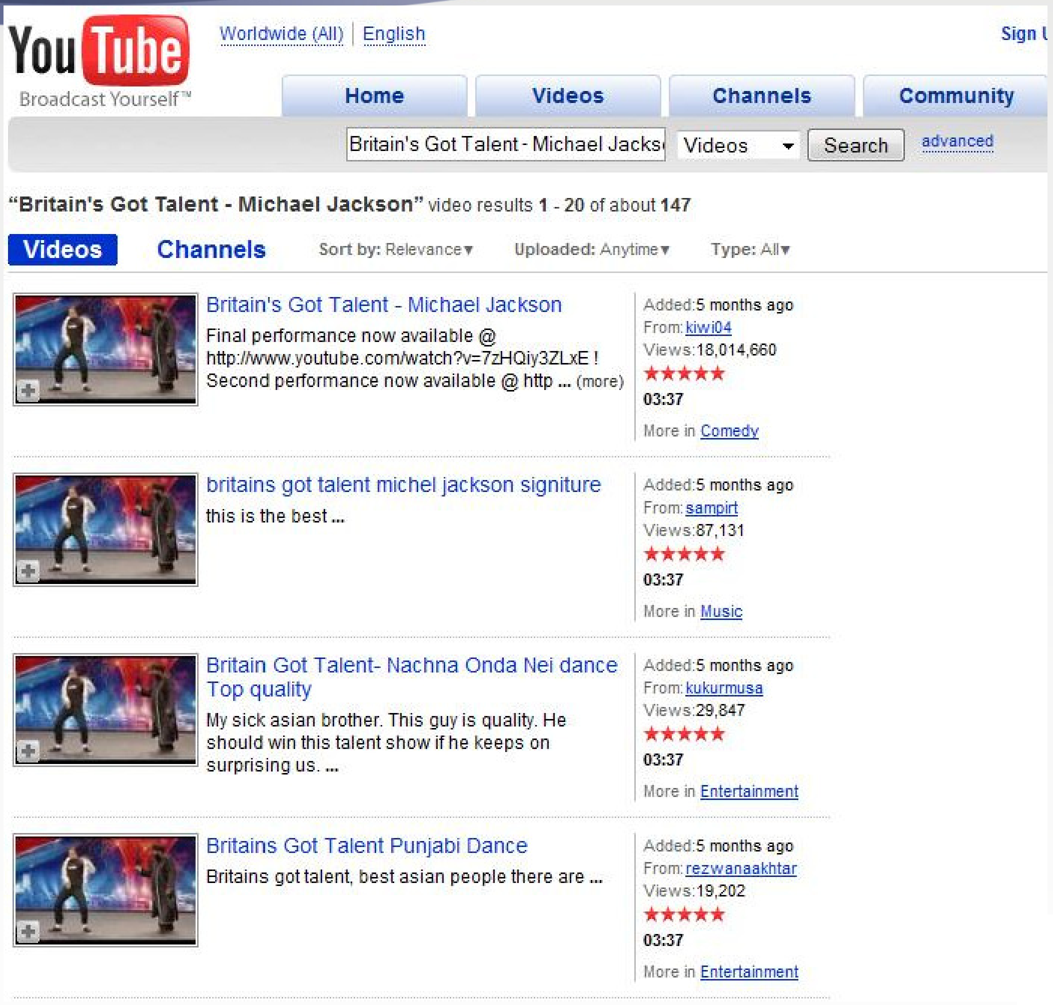
YouTube

- Pioneiro em compartilhamento de vídeos
- Formato: Flash Vídeo (Macromedia flash vídeo)
- Comprado pelo Google por 1.6 bilhões
- Recebe 10 horas de vídeos a cada minuto
- Vídeo sobre a infraestrutura do YouTube
  - http://video.google.com/videoplay?docid=-6304964351441328559
  - Thumbnails consomem muitos recursos
- Listas mais discutidos, respondidos, vistos, etc.
  - Música "My hot hot Sexy" chegou entre o mais vistos
  - Site dos fãs da Avril Lavigne tentando tornar um vídeo dela o mais popular do YouTube

Problemas no YouTube

- Desempenho
- Problemas com copyright
  - Vídeo da Cicarelli. Bloqueio do YouTube no Brasil
  - Propagandas com lucros para os donos do vídeo
  - Parceria com globo, BBC e outras grandes
- Video Spam, promoção do conteúdo, contas falsas, scripts automáticos
- Metadados que não descrevem bem o conteúdo
- Pornografia
- Vídeo duplicado
- Associação de propagandas

Duplicatas no YouTube



Worldwide (All) | English

Home | Videos | Channels | Community

Britain's Got Talent - Michael Jacks | Videos | Search | advanced

"Britain's Got Talent - Michael Jackson" video results 1 - 20 of about 147

Videos | Channels | Sort by: Relevance | Uploaded: Anytime | Type: All

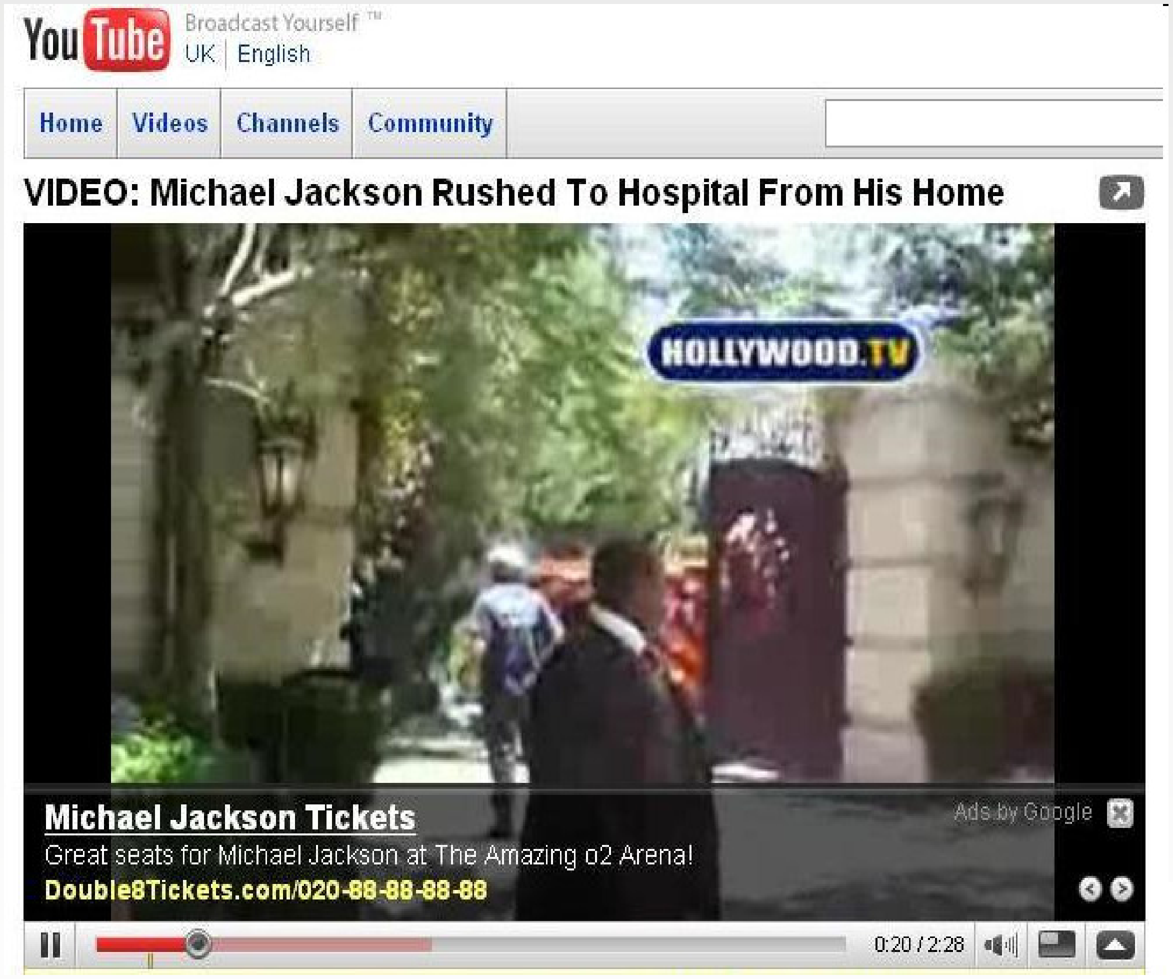
Britain's Got Talent - Michael Jackson
Final performance now available @ http://www.youtube.com/watch?v=7zHQiy3ZLxE ! Second performance now available @ http ... (more)
Added: 5 months ago
From: kiwi04
Views: 18,014,660
★★★★★
03:37
More in Comedy

britains got talent michel jackson signiture
this is the best ...
Added: 5 months ago
From: sampirt
Views: 87,131
★★★★★
03:37
More in Music

Britain Got Talent- Nachna Onda Nei dance Top quality
My sick asian brother. This guy is quality. He should win this talent show if he keeps on surprising us. ...
Added: 5 months ago
From: kukurmusa
Views: 29,847
★★★★★
03:37
More in Entertainment

Britains Got Talent Punjabi Dance
Britains got talent, best asian people there are ...
Added: 5 months ago
From: rezwanaakhtar
Views: 19,202
★★★★★
03:37
More in Entertainment

Propagandas no YouTube



VIDEO: Michael Jackson Rushed To Hospital From His Home

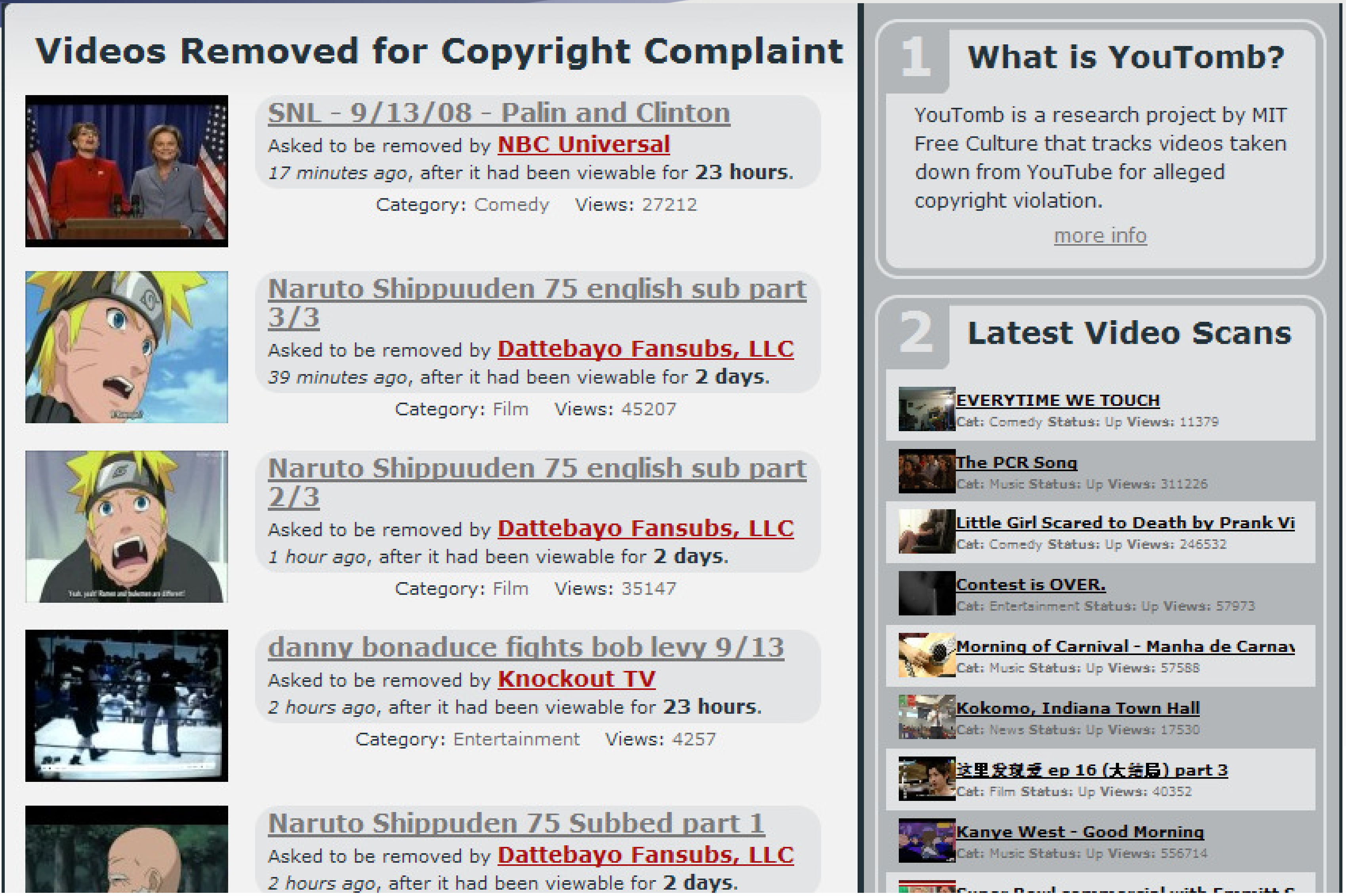
Michael Jackson Tickets

Great seats for Michael Jackson at The Amazing o2 Arena!

Double8Tickets.com/020-88-88-88-88

YouTomb

http://youtomb.mit.edu/



Videos Removed for Copyright Complaint

SNL - 9/13/08 - Palin and Clinton
Asked to be removed by **NBC Universal**
17 minutes ago, after it had been viewable for **23 hours**.
Category: Comedy Views: 27212

Naruto Shippuuden 75 english sub part 3/3
Asked to be removed by **Dattebayo Fansubs, LLC**
39 minutes ago, after it had been viewable for **2 days**.
Category: Film Views: 45207

Naruto Shippuuden 75 english sub part 2/3
Asked to be removed by **Dattebayo Fansubs, LLC**
1 hour ago, after it had been viewable for **2 days**.
Category: Film Views: 35147

danny bonaduce fights bob levy 9/13
Asked to be removed by **Knockout TV**
2 hours ago, after it had been viewable for **23 hours**.
Category: Entertainment Views: 4257

Naruto Shippuden 75 Subbed part 1
Asked to be removed by **Dattebayo Fansubs, LLC**
2 hours ago, after it had been viewable for **2 days**.

1 What is YouTomb?

YouTomb is a research project by MIT Free Culture that tracks videos taken down from YouTube for alleged copyright violation.

more info

2 Latest Video Scans

EVERYTIME WE TOUCH
Cat: Comedy **Status:** Up **Views:** 11379

The PCR Song
Cat: Music **Status:** Up **Views:** 311226

Little Girl Scared to Death by Prank Vi
Cat: Comedy **Status:** Up **Views:** 246532

Contest is OVER.
Cat: Entertainment **Status:** Up **Views:** 57973

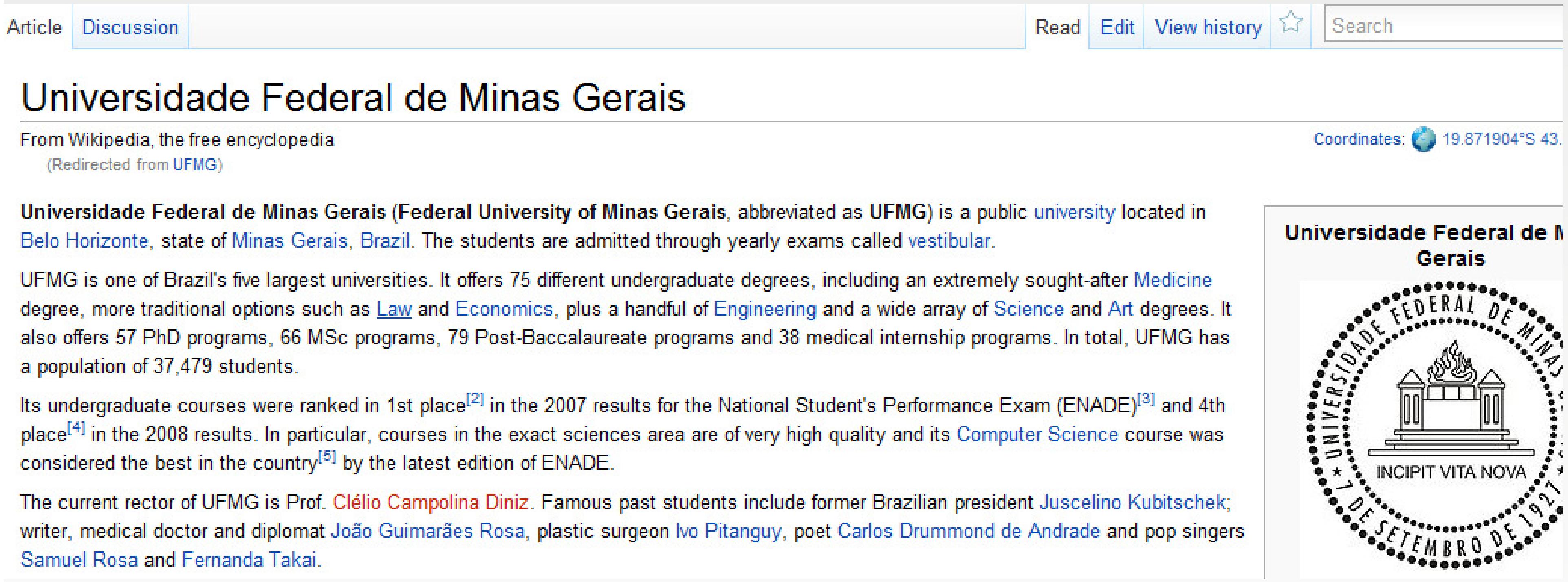
Morning of Carnival - Manha de Carnav
Cat: Music **Status:** Up **Views:** 57588

Kokomo, Indiana Town Hall
Cat: News **Status:** Up **Views:** 17530

这里发现爱 ep 16 (大结局) part 3
Cat: Film **Status:** Up **Views:** 40352

Kanye West - Good Morning
Cat: Music **Status:** Up **Views:** 556714

Wikipedia



Article Discussion Read Edit View history Search

Universidade Federal de Minas Gerais

From Wikipedia, the free encyclopedia

(Redirected from UFMG)

Coordinates: 19.871904°S 43.

Universidade Federal de Minas Gerais (**Federal University of Minas Gerais**, abbreviated as **UFMG**) is a public university located in Belo Horizonte, state of Minas Gerais, Brazil. The students are admitted through yearly exams called vestibular.

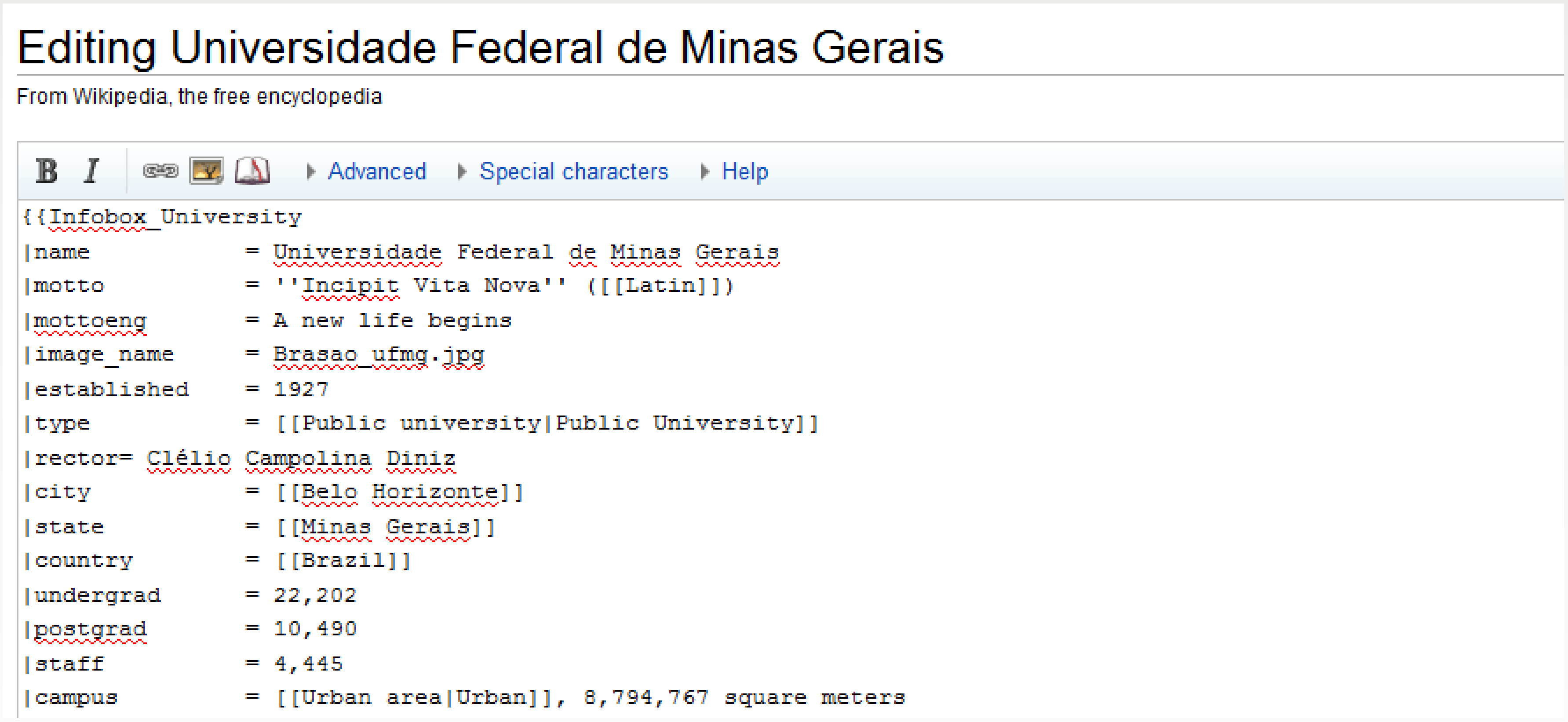
UFMG is one of Brazil's five largest universities. It offers 75 different undergraduate degrees, including an extremely sought-after Medicine degree, more traditional options such as Law and Economics, plus a handful of Engineering and a wide array of Science and Art degrees. It also offers 57 PhD programs, 66 MSc programs, 79 Post-Baccalaureate programs and 38 medical internship programs. In total, UFMG has a population of 37,479 students.

Its undergraduate courses were ranked in 1st place[2] in the 2007 results for the National Student's Performance Exam (ENADE)[3] and 4th place[4] in the 2008 results. In particular, courses in the exact sciences area are of very high quality and its Computer Science course was considered the best in the country[5] by the latest edition of ENADE.

The current rector of UFMG is Prof. Clélio Campolina Diniz. Famous past students include former Brazilian president Juscelino Kubitschek; writer, medical doctor and diplomat João Guimarães Rosa, plastic surgeon Ivo Pitanguy, poet Carlos Drummond de Andrade and pop singers Samuel Rosa and Fernanda Takai.

Universidade Federal de M Gerais

Wikipedia



Editing Universidade Federal de Minas Gerais

From Wikipedia, the free encyclopedia

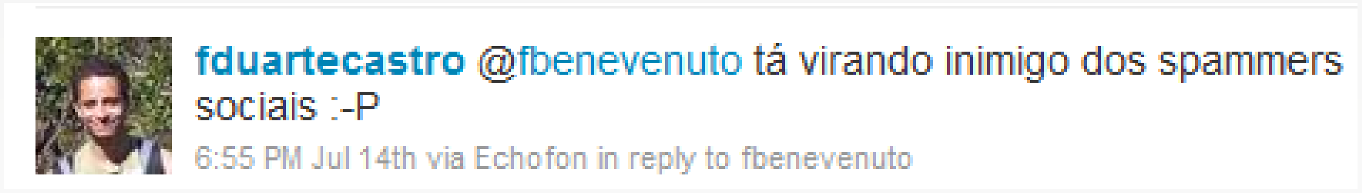
```
{{Infobox_University
|name           = Universidade Federal de Minas Gerais
|motto          = ''Incipit Vita Nova'' ([[Latin]])
|mottoeng       = A new life begins
|image_name     = Brasao_ufmg.jpg
|established    = 1927
|type           = [[Public university|Public University]]
|rector= Clélio Campolina Diniz
|city           = [[Belo Horizonte]]
|state          = [[Minas Gerais]]
|country        = [[Brazil]]
|undergrad      = 22,202
|postgrad       = 10,490
|staff          = 4,445
|campus         = [[Urban area|Urban]], 8,794,767 square meters
```

Problemas no Wikipedia

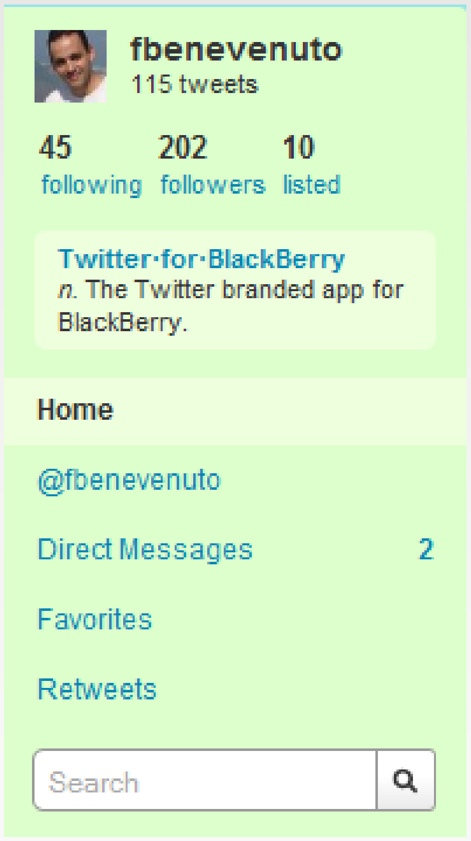
- Vandalismo
  - Apagar uma página existente
  - Editar uma página e colocar um conteúdo não correspondente ao assunto
    - **Spam, Links externos, links internos**
  - Uso de contas falsas
- Atitudes contra vandalismo
  - Patrulhamento de mudanças recentes
  - Bloqueio de IP em caso de detecção

Twitter

- Micro-blog: mensagens de no máximo 140 caracteres
- Muitas celebridades utilizam
- Busca em tempo real


fduartecastro @fbenevenuto tá virando inimigo dos spammers sociais :-P

6:55 PM Jul 14th via Echofon in reply to fbenevenuto


fbenevenuto
115 tweets

45 following · 202 followers · 10 listed

Twitter·for·BlackBerry
n. The Twitter branded app for BlackBerry.

Home

@fbenevenuto

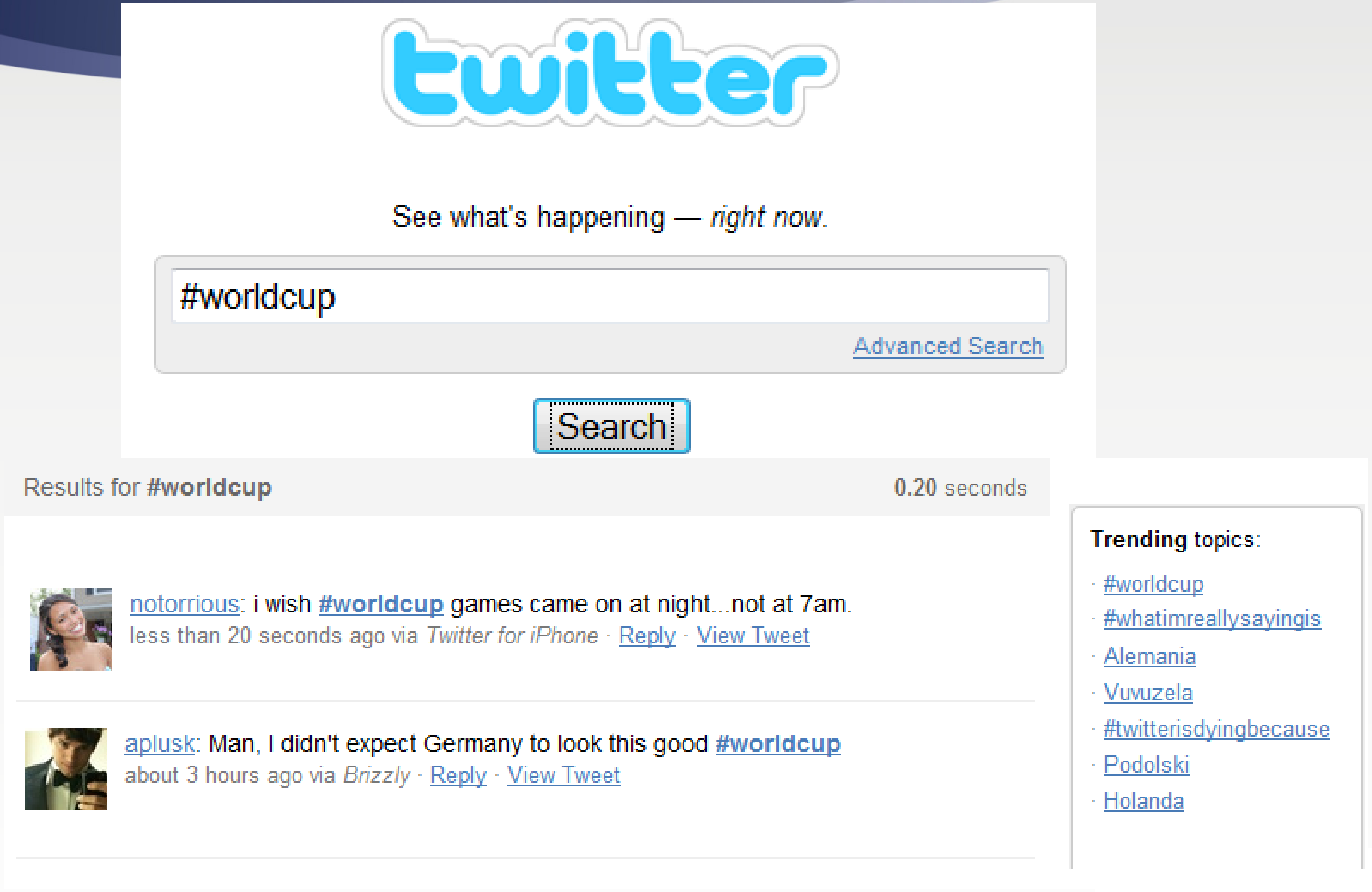
Direct Messages 2

Favorites

Retweets

Search

Busca em tempo real no Twitter



See what's happening — *right now.*

#worldcup

Results for **#worldcup** **0.20** seconds

notorrious: i wish **#worldcup** games came on at night...not at 7am.
less than 20 seconds ago via *Twitter for iPhone* · Reply · View Tweet

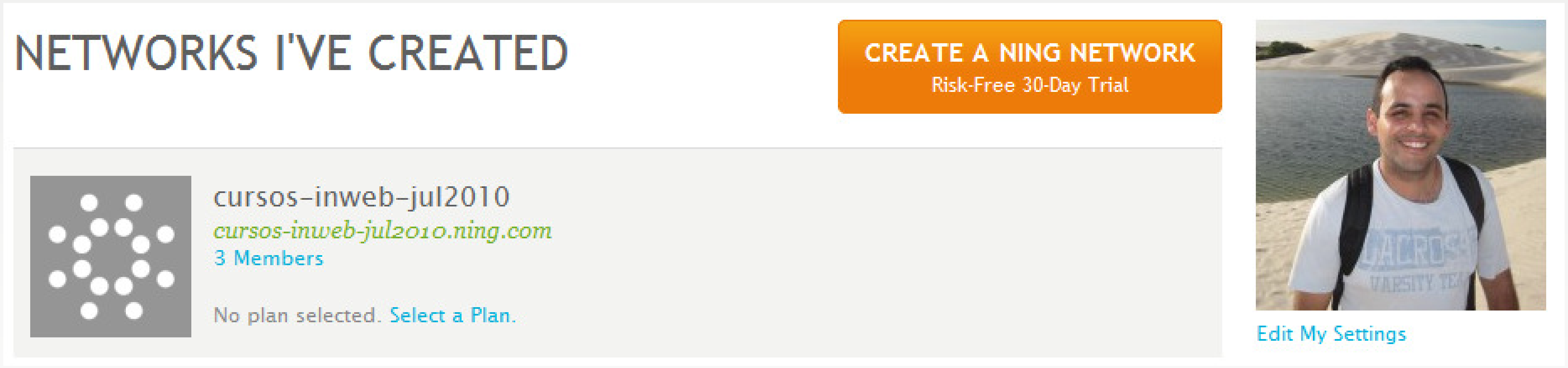
aplusk: Man, I didn't expect Germany to look this good **#worldcup**
about 3 hours ago via *Brizzly* · Reply · View Tweet

Trending topics:

- #worldcup
- #whatimreallysayingis
- Alemania
- Vuvuzela
- #twitterisdyingbecause
- Podolski
- Holanda

Ning

- Plataforma que permite a criação de redes sociais individualizadas

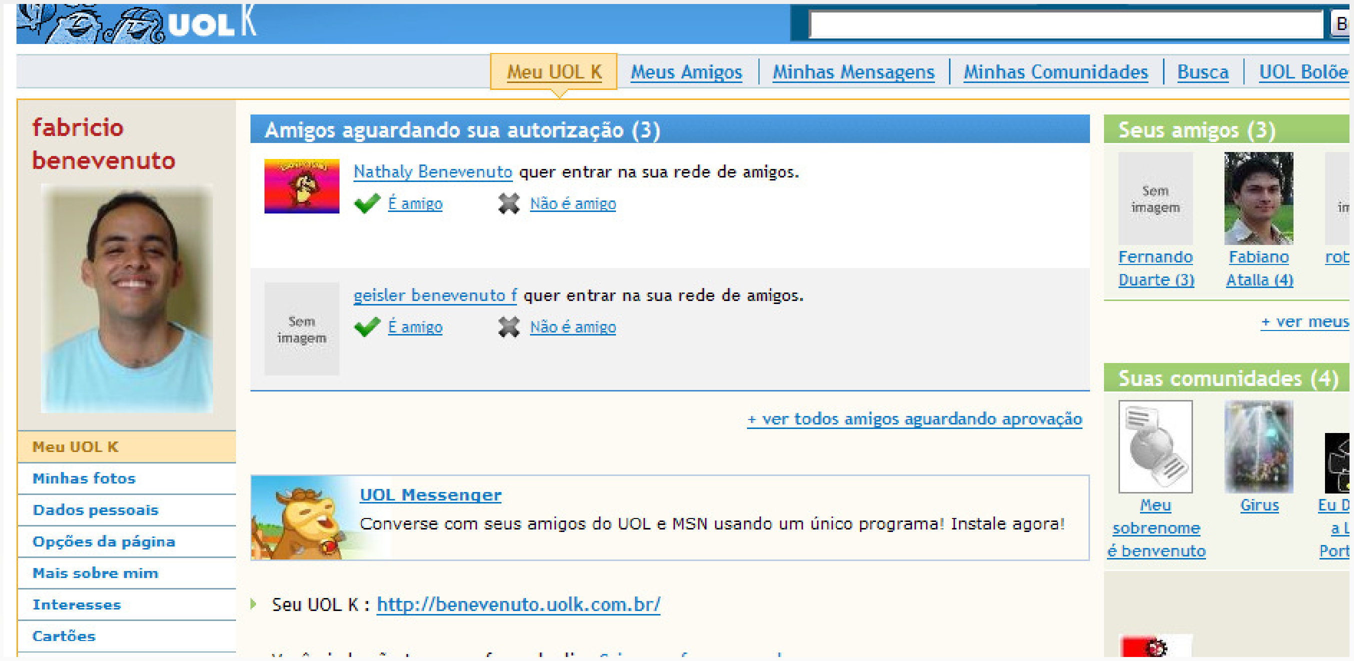

NETWORKS I'VE CREATED

cursos-inweb-jul2010
cursos-inweb-jul2010.ning.com
3 Members

No plan selected. Select a Plan.

E no Brasil?

- UOL: blogs, videolog, UOLk



Meu UOL K | Meus Amigos | Minhas Mensagens | Minhas Comunidades | Busca | UOL Bolõe

fabricio benevenuto

Meu UOL K
Minhas fotos
Dados pessoais
Opções da página
Mais sobre mim
Interesses
Cartões

Amigos aguardando sua autorização (3)

Nathaly Benevenuto quer entrar na sua rede de amigos.
É amigo Não é amigo

geisler benevenuto f quer entrar na sua rede de amigos.
É amigo Não é amigo

+ ver todos amigos aguardando aprovação

UOL Messenger
Converse com seus amigos do UOL e MSN usando um único programa! Instale agora!

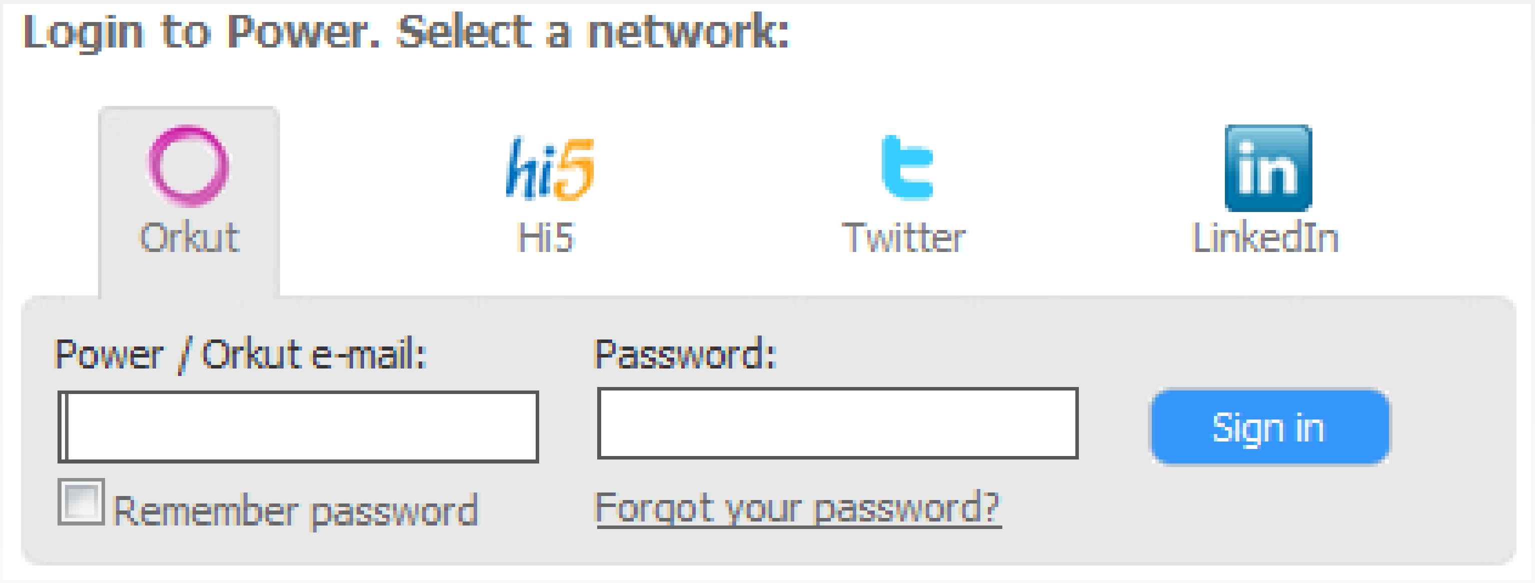
Seu UOL K : http://benevenuto.uolk.com.br/

Seus amigos (3)
Fernando Duarte (3) Fabiano Atalla (4) rob

+ ver meus

Suas comunidades (4)
Meu sobrenome é benvenuto Girus Eu D a L Port

Power.com


Login to Power. Select a network:

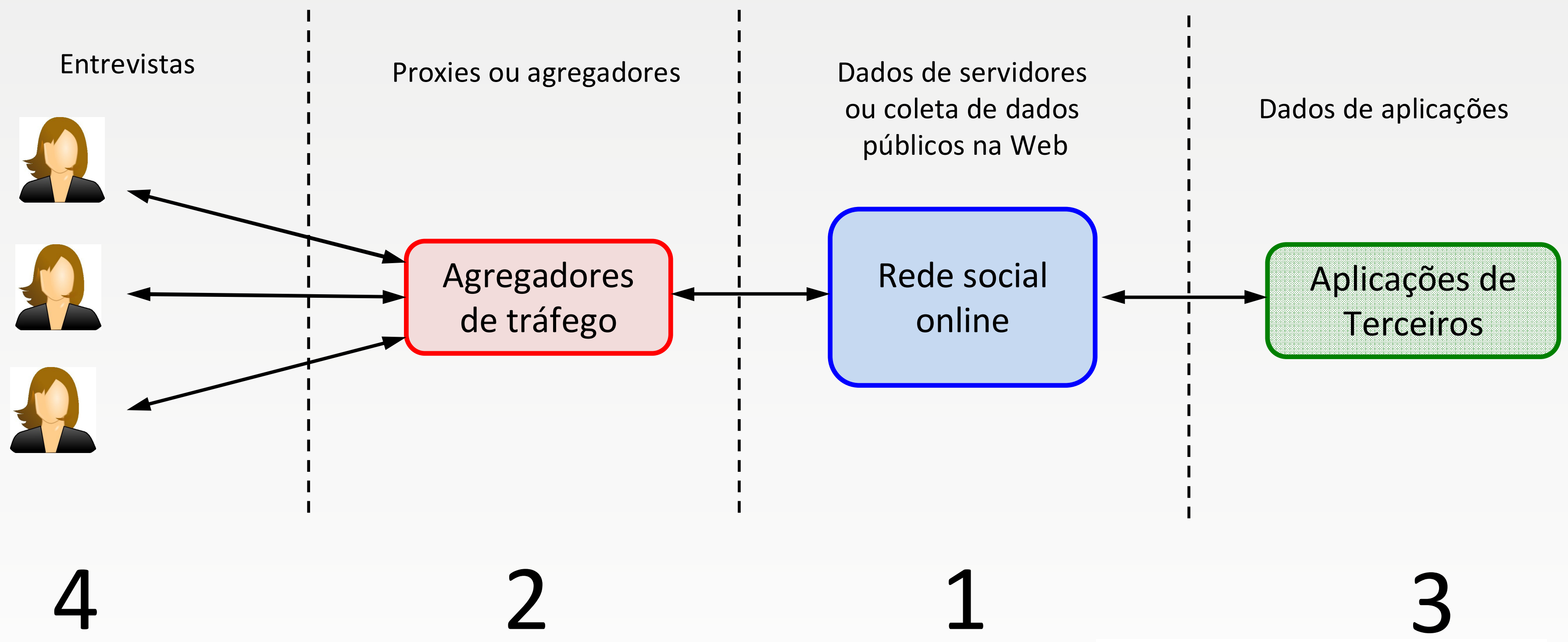
Power / Orkut e-mail:

Password:

Remember password

Forgot your password?

Formas de coleta de dados

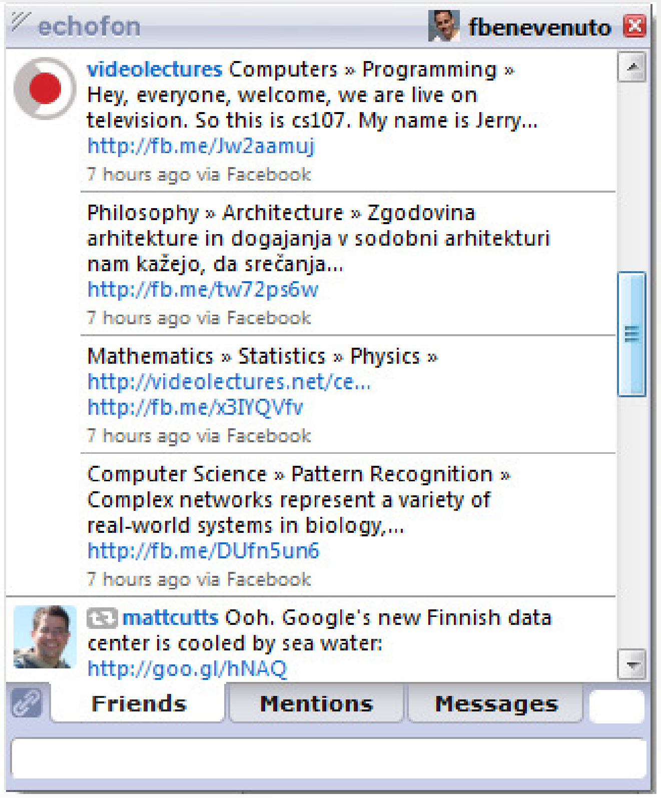


Coletores

- Coleta de IDs sequencias
  - APIs, scripts em perl e python
  - Measuring User Influence in Twitter: The million Follower Fallacy. **ICWSM’10**
- Coleta em tempo real
  - APIs
  - Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. **WWW’10**
- Coleta de chamadas ocultas com Firebug
  - Firebug e coleta de chamadas ocultas
  - The Tube over Time: Characterizing Popularity Growth of YouTube Videos. **WSDM’11**
- Coleta do WCC, distribuída e por snowball
  - Measurement and Analysis of Online Social Networks. **IMC’07**

API do Twitter

- Permitem a construção de aplicações, mas podem ser utilizadas por crawlers
  - statuses/filter
  - statuses/sample
  - trends
  - trends/daily
  - trends/weekly
  - statuses/retweets_of_me
  - statuses/mentions
  - account/rate_limit_status



videolectures Computers » Programming » Hey, everyone, welcome, we are live on television. So this is cs107. My name is Jerry... http://fb.me/Jw2aamuj
7 hours ago via Facebook

Philosophy » Architecture » Zgodovina arhitekture in dogajanja v sodobni arhitekturi nam kažejo, da srečanja... http://fb.me/tw72ps6w
7 hours ago via Facebook

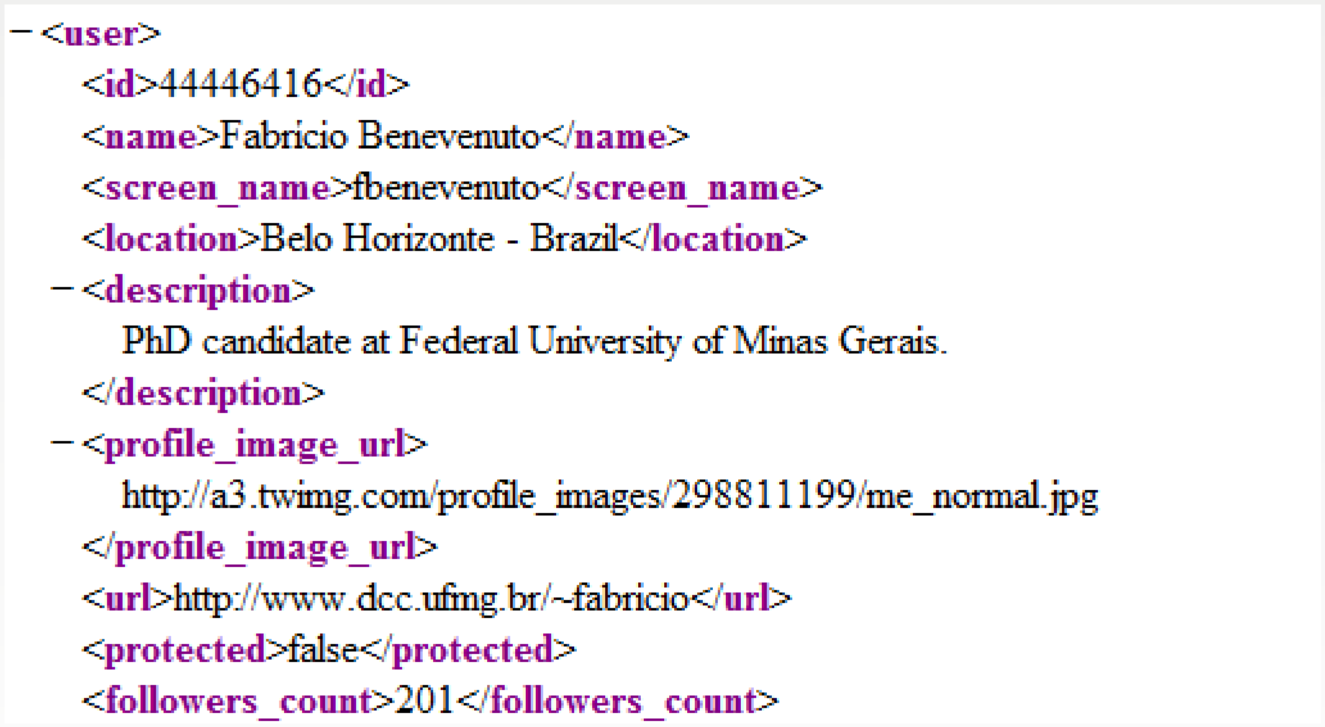
Mathematics » Statistics » Physics » http://videolectures.net/ce... http://fb.me/x3IYQVfv
7 hours ago via Facebook

Computer Science » Pattern Recognition » Complex networks represent a variety of real-world systems in biology,... http://fb.me/DUfn5un6
7 hours ago via Facebook

mattcutts Ooh. Google's new Finnish data center is cooled by sea water: http://goo.gl/hNAQ

API do Twitter

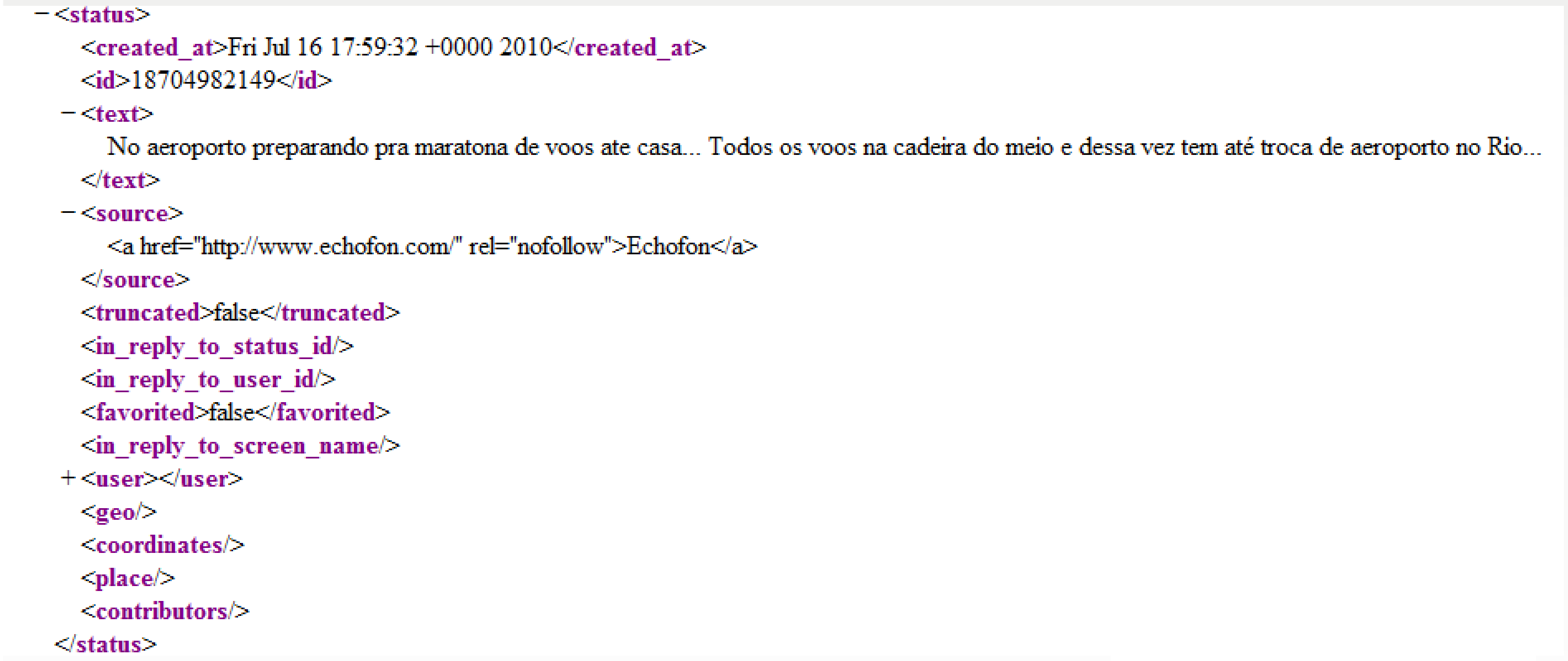
- Profile do usuário: http://twitter.com/users/show/44446416.xml



```
– <user>
    <id>44446416</id>
    <name>Fabricio Benevenuto</name>
    <screen_name>fbenevenuto</screen_name>
    <location>Belo Horizonte - Brazil</location>
  – <description>
      PhD candidate at Federal University of Minas Gerais.
    </description>
  – <profile_image_url>
      http://a3.twimg.com/profile_images/298811199/me_normal.jpg
    </profile_image_url>
    <url>http://www.dcc.ufmg.br/~fabricio</url>
    <protected>false</protected>
    <followers_count>201</followers_count>
```

API do Twitter

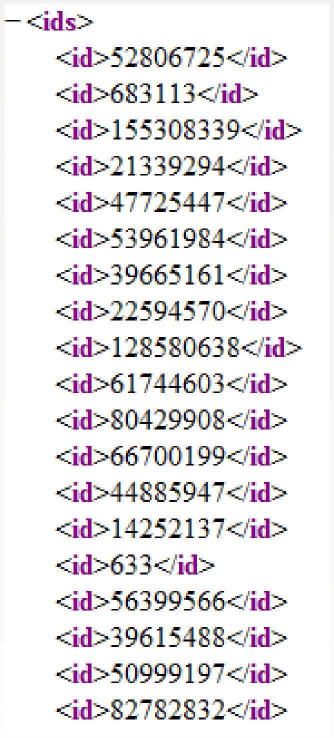
- Tweets: http://twitter.com/statuses/user_timeline.xml?user_id=44446416&count=200&page=1



```
– <status>
    <created_at>Fri Jul 16 17:59:32 +0000 2010</created_at>
    <id>18704982149</id>
  – <text>
      No aeroporto preparando pra maratona de voos ate casa... Todos os voos na cadeira do meio e dessa vez tem até troca de aeroporto no Rio...
    </text>
  – <source>
      <a href="http://www.echofon.com/" rel="nofollow">Echofon</a>
    </source>
    <truncated>false</truncated>
    <in_reply_to_status_id/>
    <in_reply_to_user_id/>
    <favorited>false</favorited>
    <in_reply_to_screen_name/>
  + <user></user>
    <geo/>
    <coordinates/>
    <place/>
    <contributors/>
  </status>
```

API do Twitter

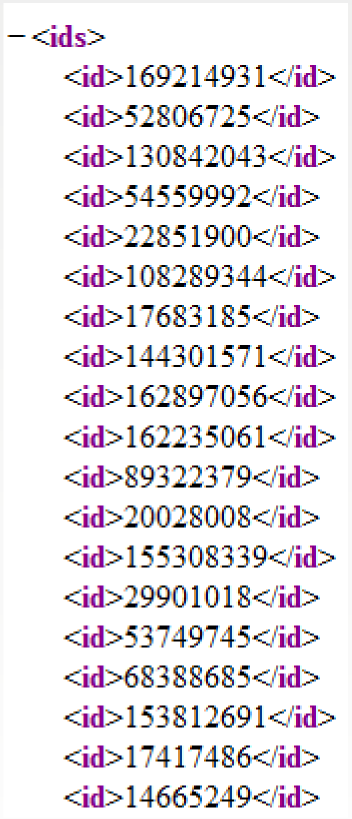
- Followees: Provê 5000 IDs por requisição
- http://twitter.com/friends/ids/44446416.xml?page=1



```
- <ids>
    <id>52806725</id>
    <id>683113</id>
    <id>155308339</id>
    <id>21339294</id>
    <id>47725447</id>
    <id>53961984</id>
    <id>39665161</id>
    <id>22594570</id>
    <id>128580638</id>
    <id>61744603</id>
    <id>80429908</id>
    <id>66700199</id>
    <id>44885947</id>
    <id>14252137</id>
    <id>633</id>
    <id>56399566</id>
    <id>39615488</id>
    <id>50999197</id>
    <id>82782832</id>
```

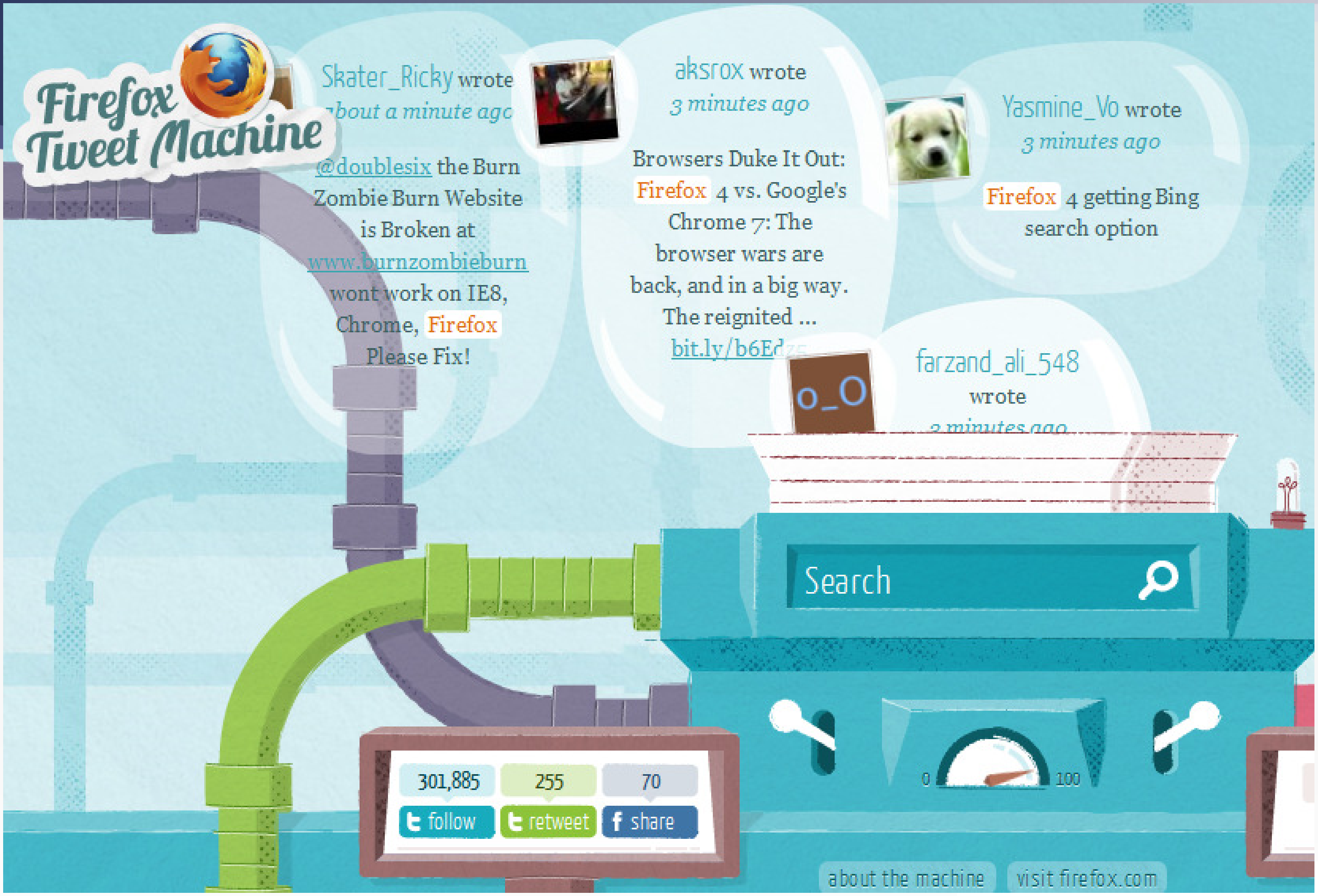
API do Twitter

- Followers: http://twitter.com/followers/ids/44446416.xml?page=1



```
- <ids>
    <id>169214931</id>
    <id>52806725</id>
    <id>130842043</id>
    <id>54559992</id>
    <id>22851900</id>
    <id>108289344</id>
    <id>17683185</id>
    <id>144301571</id>
    <id>162897056</id>
    <id>162235061</id>
    <id>89322379</id>
    <id>20028008</id>
    <id>155308339</id>
    <id>29901018</id>
    <id>53749745</id>
    <id>68388685</id>
    <id>153812691</id>
    <id>17417486</id>
    <id>14665249</id>
```

http://firefoxtweetmachine.com/



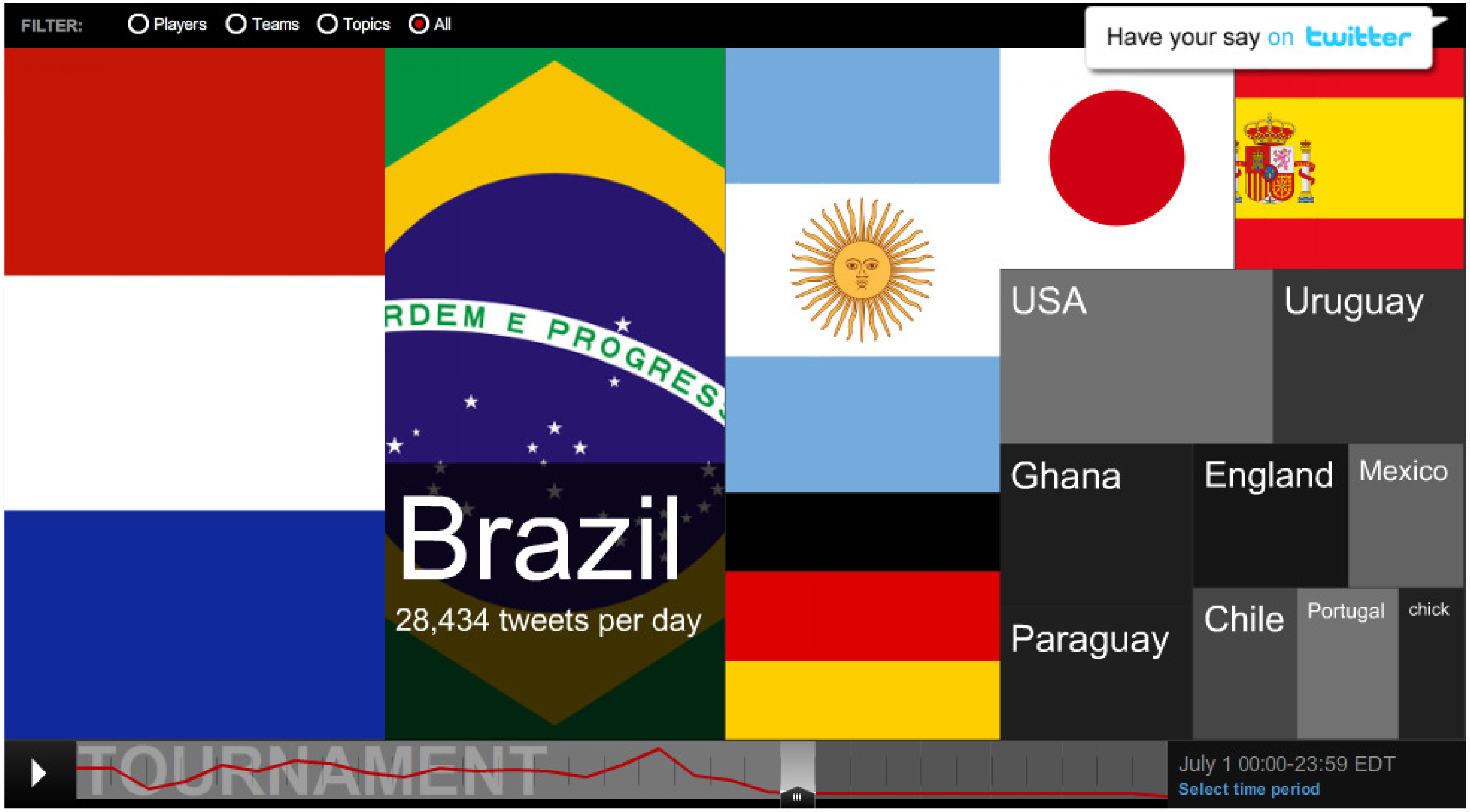
Skater_Ricky wrote
about a minute ago
@doublesix the Burn Zombie Burn Website is Broken at www.burnzombieburn wont work on IE8, Chrome, Firefox Please Fix!

aksrox wrote
3 minutes ago
Browsers Duke It Out: Firefox 4 vs. Google's Chrome 7: The browser wars are back, and in a big way. The reignited ... bit.ly/b6Ed

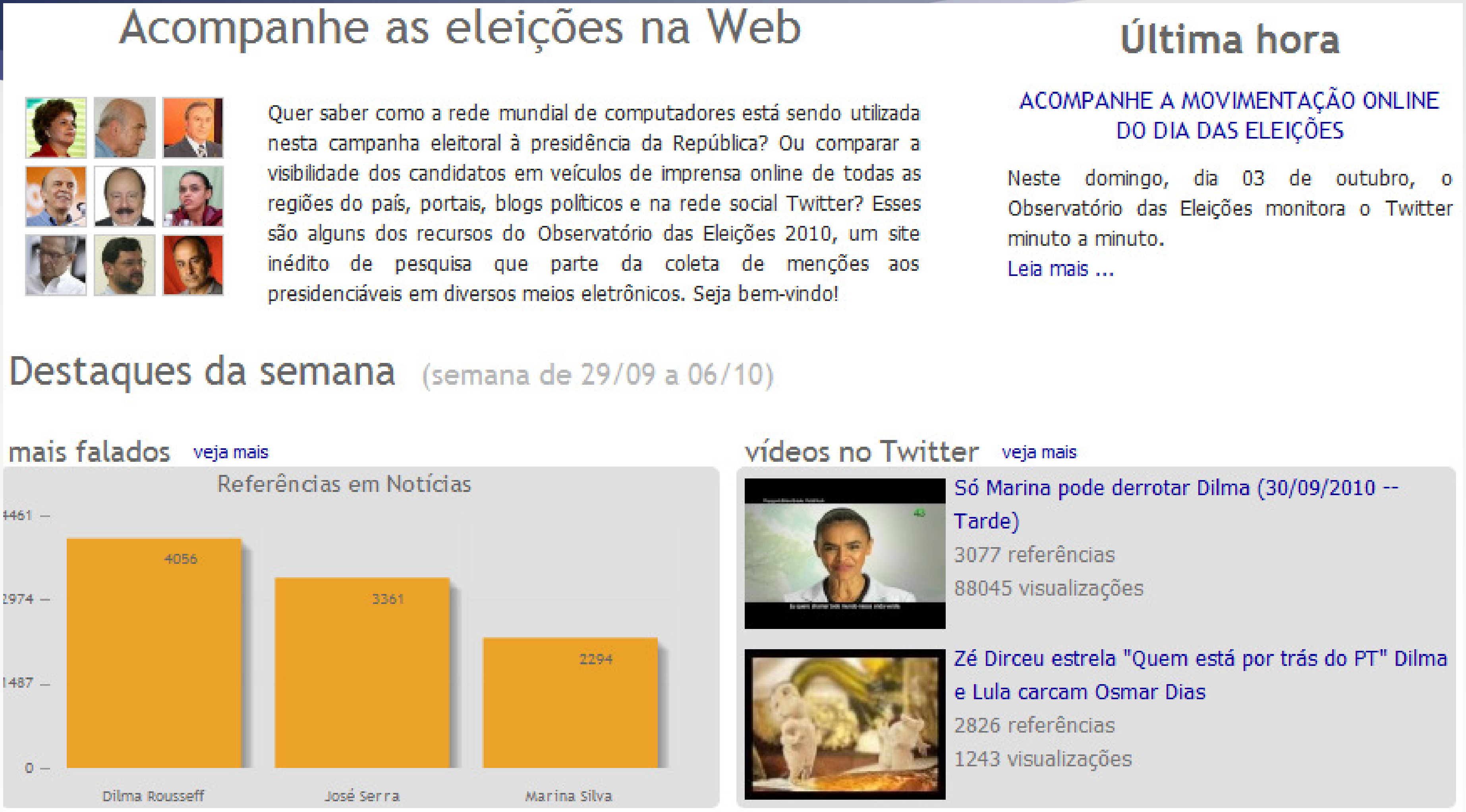
Yasmine_Vo wrote
3 minutes ago
Firefox 4 getting Bing search option

farzand_ali_548 wrote
3 minutes ago

301,885	255	70
follow	retweet	share


FILTER:
Players
Teams
Topics
All
Have your say on twitter
Brazil
28,434 tweets per day
USA
Uruguay
Ghana
England
Mexico
Paraguay
Chile
Portugal
chick
TOURNAMENT
July 1 00:00-23:59 EDT
Select time period

http://observatorio.inweb.org.br/eleicoes2010/destaques/



Acompanhe as eleições na Web

Quer saber como a rede mundial de computadores está sendo utilizada nesta campanha eleitoral à presidência da República? Ou comparar a visibilidade dos candidatos em veículos de imprensa online de todas as regiões do país, portais, blogs políticos e na rede social Twitter? Esses são alguns dos recursos do Observatório das Eleições 2010, um site inédito de pesquisa que parte da coleta de menções aos presidenciáveis em diversos meios eletrônicos. Seja bem-vindo!

Última hora

ACOMPANHE A MOVIMENTAÇÃO ONLINE DO DIA DAS ELEIÇÕES

Neste domingo, dia 03 de outubro, o Observatório das Eleições monitora o Twitter minuto a minuto.
Leia mais ...

Destaques da semana (semana de 29/09 a 06/10)

mais falados veja mais

Referências em Notícias

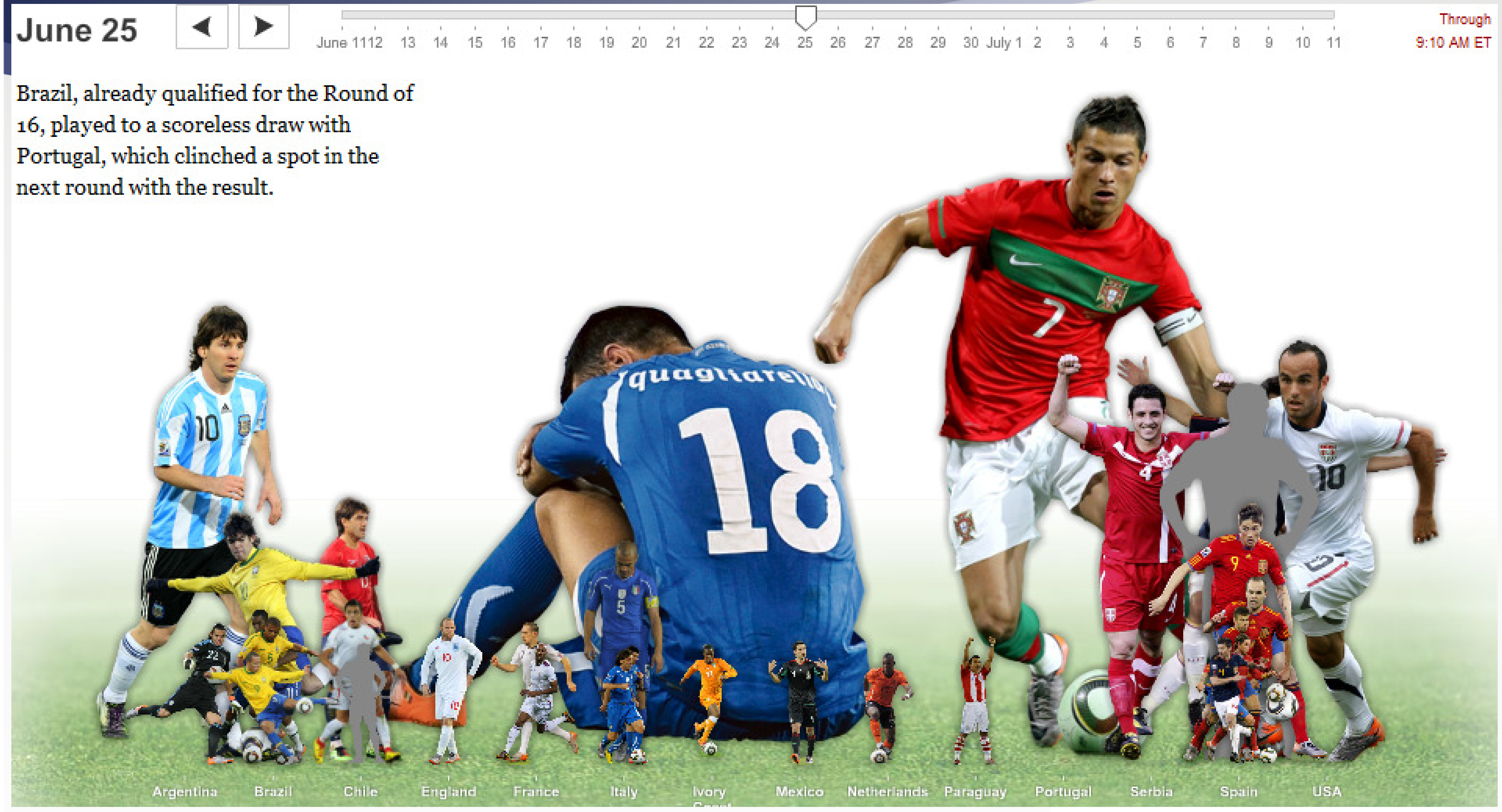
Dilma Rousseff	4056
José Serra	3361
Marina Silva	2294

vídeos no Twitter veja mais

Só Marina pode derrotar Dilma (30/09/2010 -- Tarde)
3077 referências
88045 visualizações

Zé Dirceu estrela "Quem está por trás do PT" Dilma e Lula carcam Osmar Dias
2826 referências
1243 visualizações

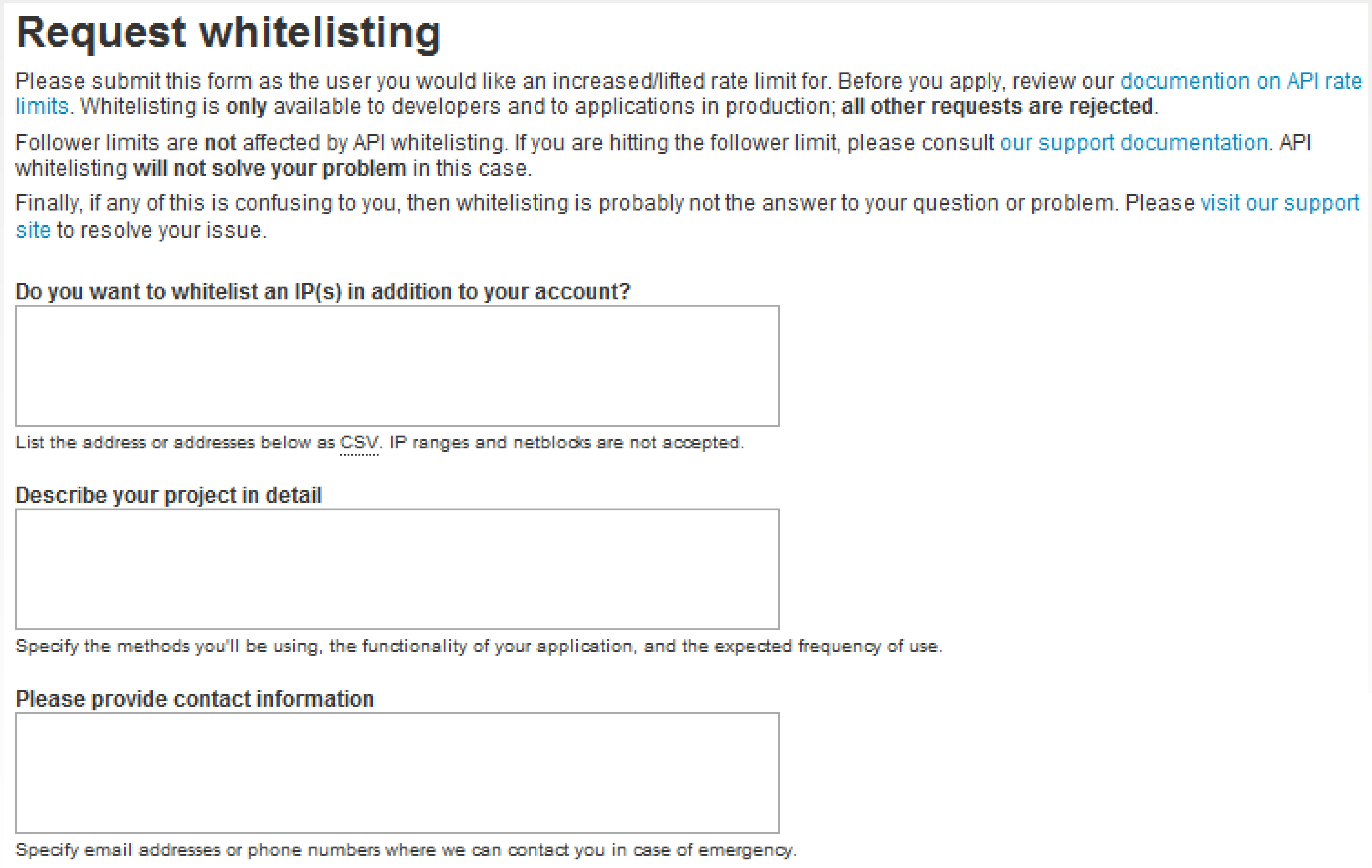
http://www.nytimes.com/interactive/2010/07/02/sports/soccer/facebook-worldcup.html/



Brazil, already qualified for the Round of 16, played to a scoreless draw with Portugal, which clinched a spot in the next round with the result.

API do Twitter

- http://twitter.com/help/request_whitelisting



Request whitelisting

Please submit this form as the user you would like an increased/lifted rate limit for. Before you apply, review our documention on API rate limits. Whitelisting is **only** available to developers and to applications in production; **all other requests are rejected**.

Follower limits are **not** affected by API whitelisting. If you are hitting the follower limit, please consult our support documentation. API whitelisting **will not solve your problem** in this case.

Finally, if any of this is confusing to you, then whitelisting is probably not the answer to your question or problem. Please visit our support site to resolve your issue.

Do you want to whitelist an IP(s) in addition to your account?

List the address or addresses below as CSV. IP ranges and netblocks are not accepted.

Describe your project in detail

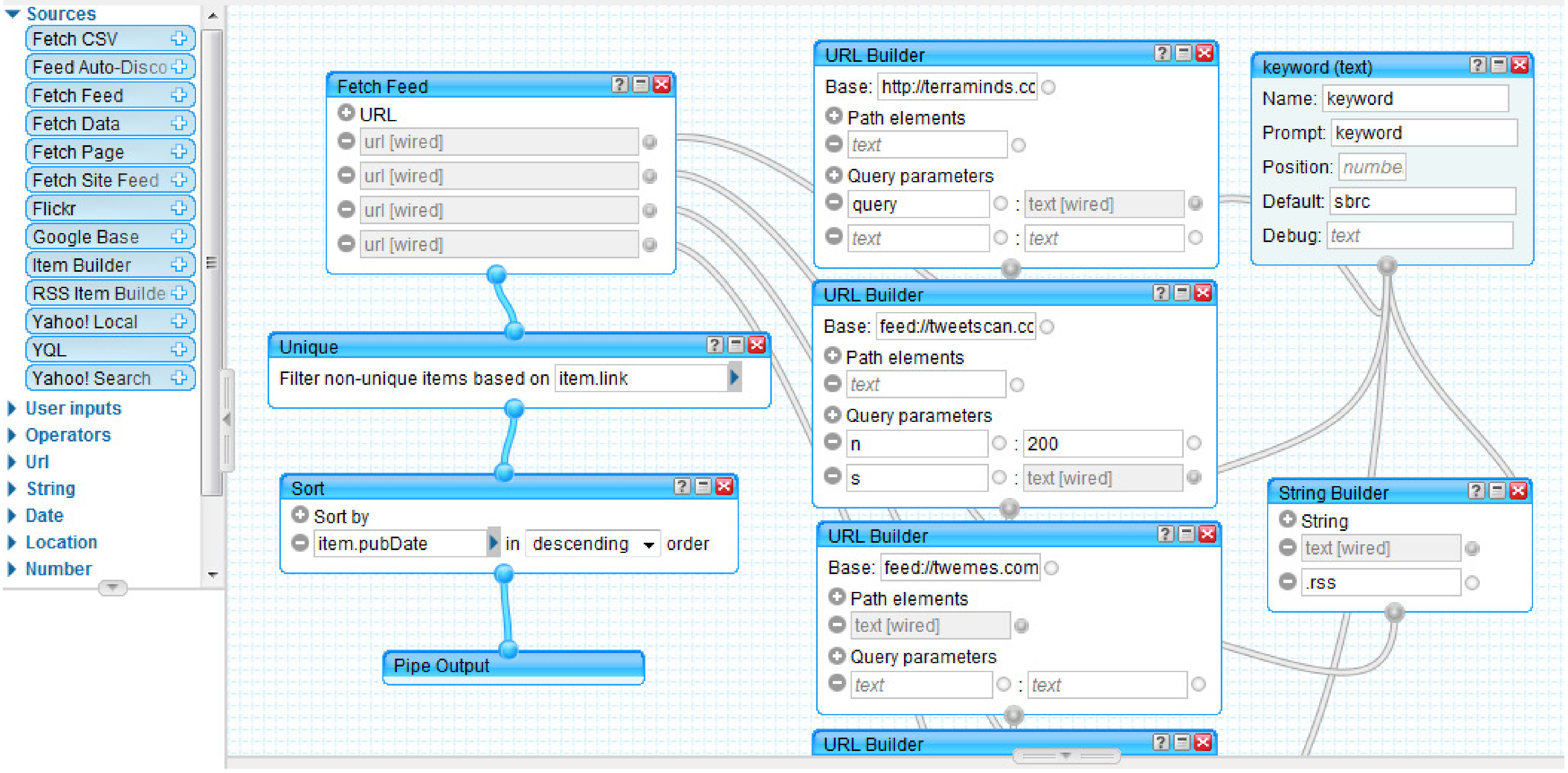
Specify the methods you'll be using, the functionality of your application, and the expected frequency of use.

Please provide contact information

Specify email addresses or phone numbers where we can contact you in case of emergency.

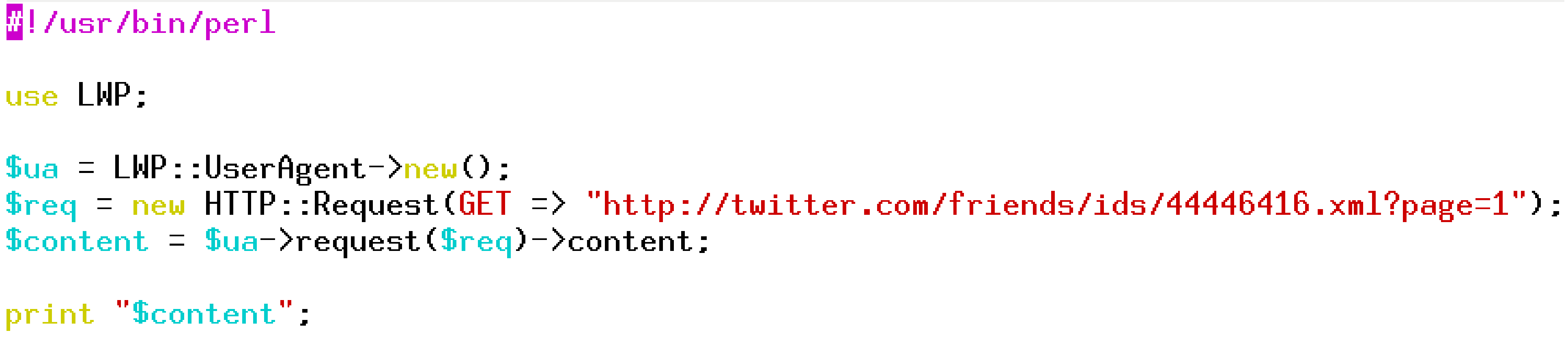
Mashups

- Aplicação que mistura várias APIs
- Yahoo Pipes: Pode ser útil também para coletar dados



Crawler – código em perl

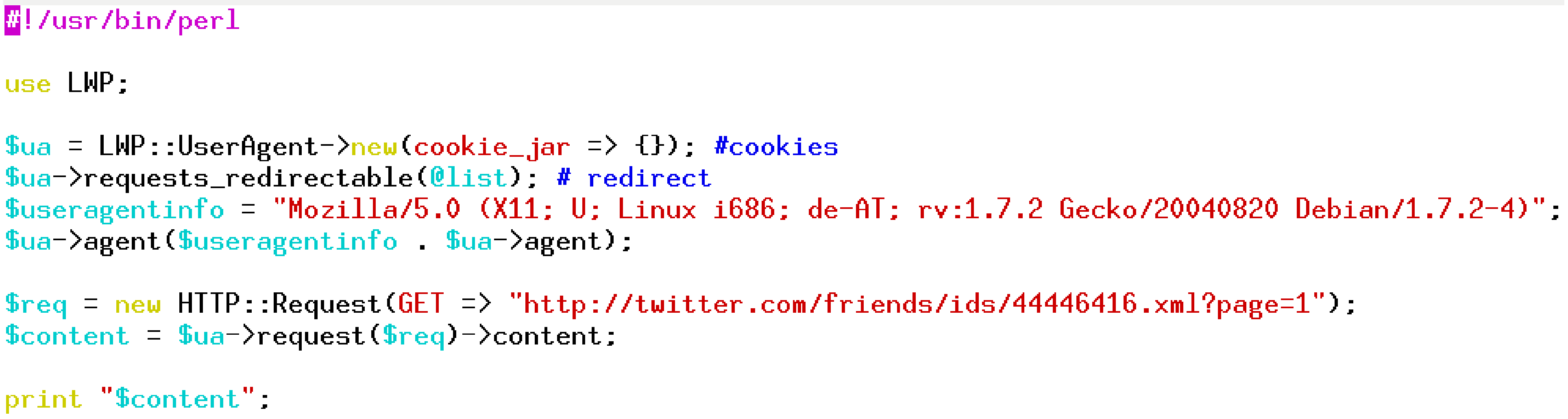
- Biblioteca LWP da linguagem PERL



```perl
#!/usr/bin/perl

use LWP;

$ua = LWP::UserAgent->new();
$req = new HTTP::Request(GET => "http://twitter.com/friends/ids/44446416.xml?page=1");
$content = $ua->request($req)->content;

print "$content";
```

Crawler – código em perl

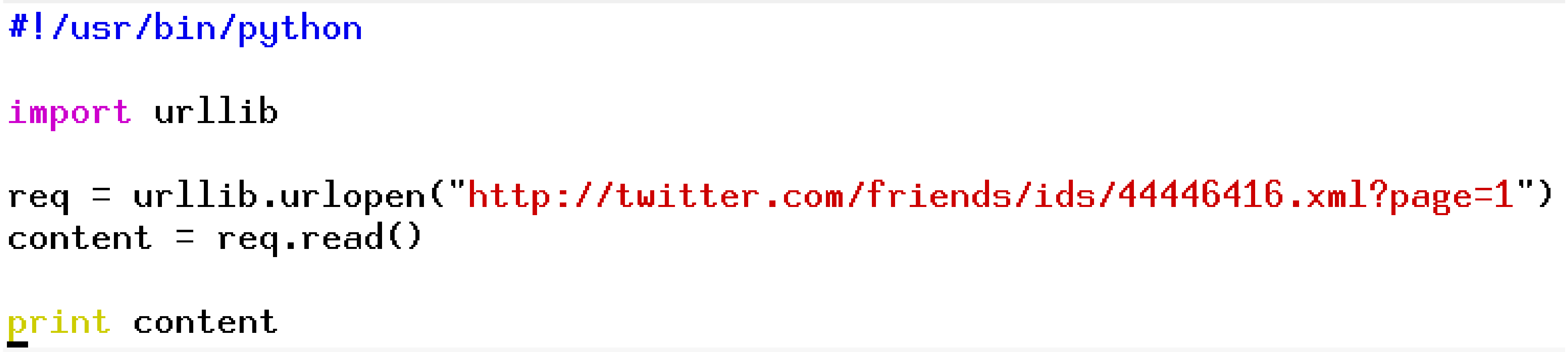
- Com mais detalhes no cabeçalho



```
#!/usr/bin/perl

use LWP;

$ua = LWP::UserAgent->new(cookie_jar => {}); #cookies
$ua->requests_redirectable(@list); # redirect
$useragentinfo = "Mozilla/5.0 (X11; U; Linux i686; de-AT; rv:1.7.2 Gecko/20040820 Debian/1.7.2-4)";
$ua->agent($useragentinfo . $ua->agent);

$req = new HTTP::Request(GET => "http://twitter.com/friends/ids/44446416.xml?page=1");
$content = $ua->request($req)->content;

print "$content";
```

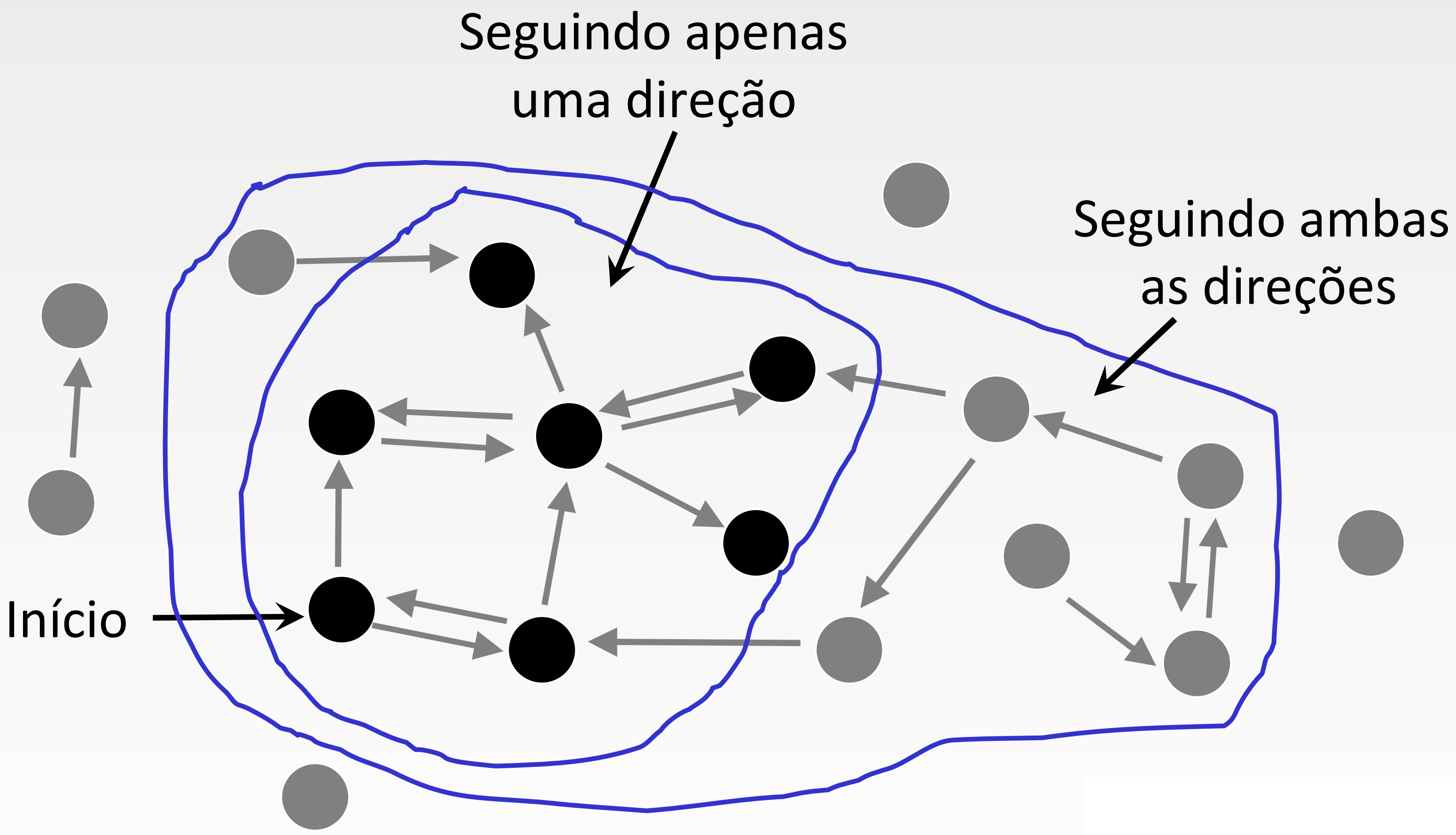
Crawler – código em python

- Biblioteca urllib da linguagem PYTHON

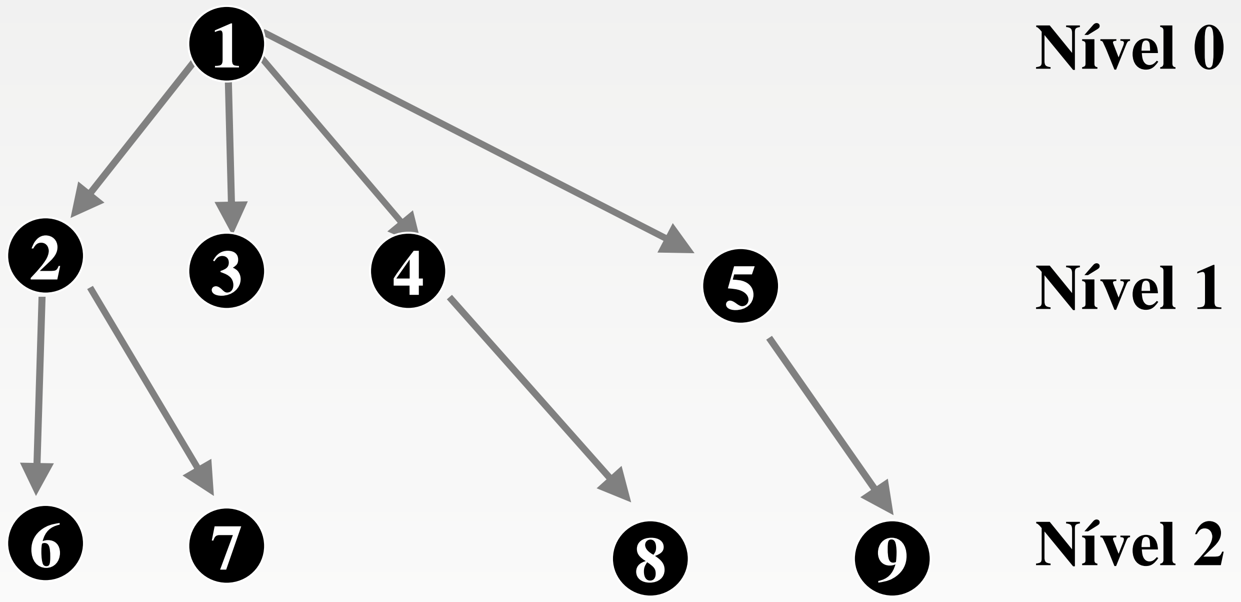


```python
#!/usr/bin/python

import urllib

req = urllib.urlopen("http://twitter.com/friends/ids/44446416.xml?page=1")
content = req.read()

print content
```

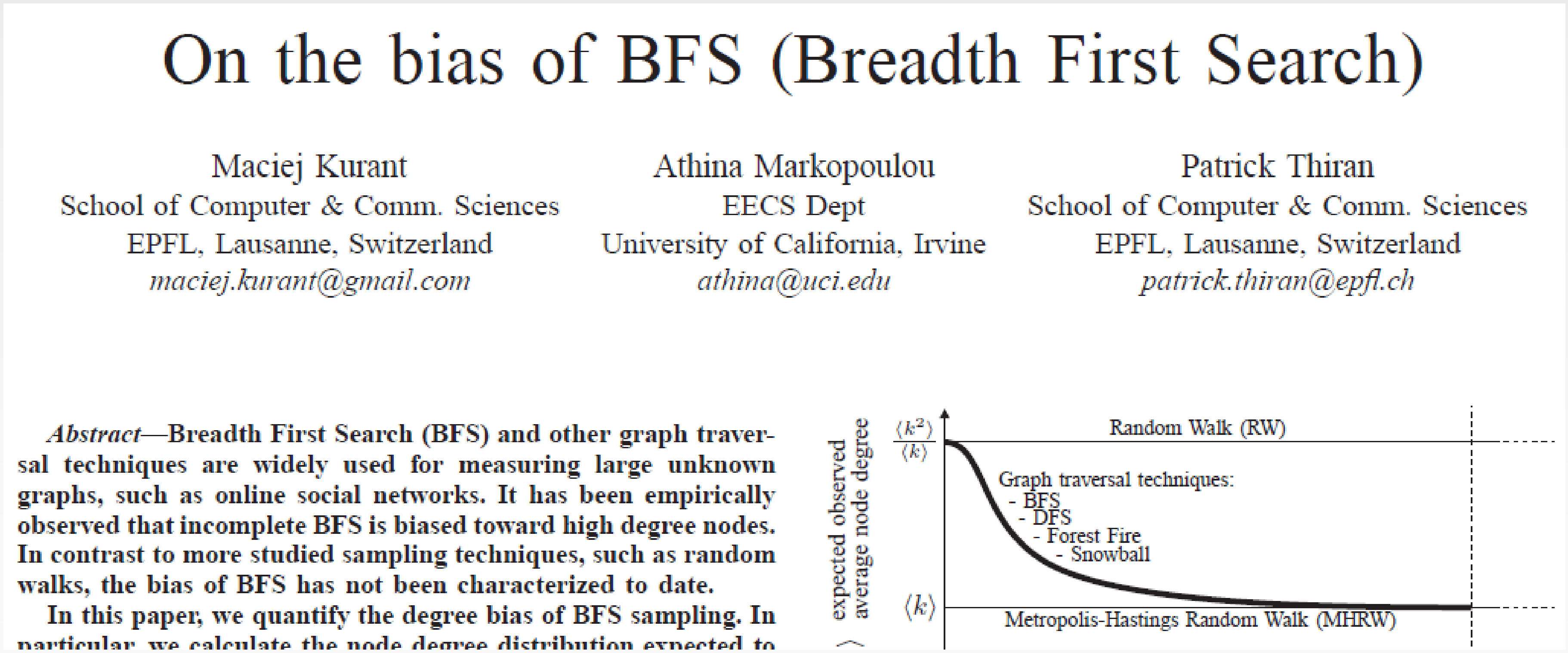
Coleta do WCC



Amostragem com Snowball



Problemas com Snowball



On the bias of BFS (Breadth First Search)

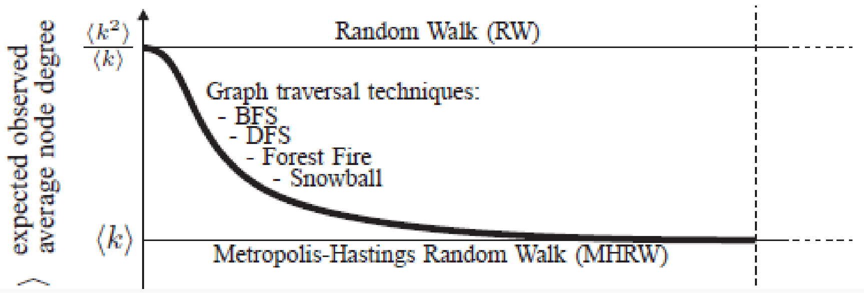
Maciej Kurant
School of Computer & Comm. Sciences
EPFL, Lausanne, Switzerland
maciej.kurant@gmail.com

Athina Markopoulou
EECS Dept
University of California, Irvine
athina@uci.edu

Patrick Thiran
School of Computer & Comm. Sciences
EPFL, Lausanne, Switzerland
patrick.thiran@epfl.ch

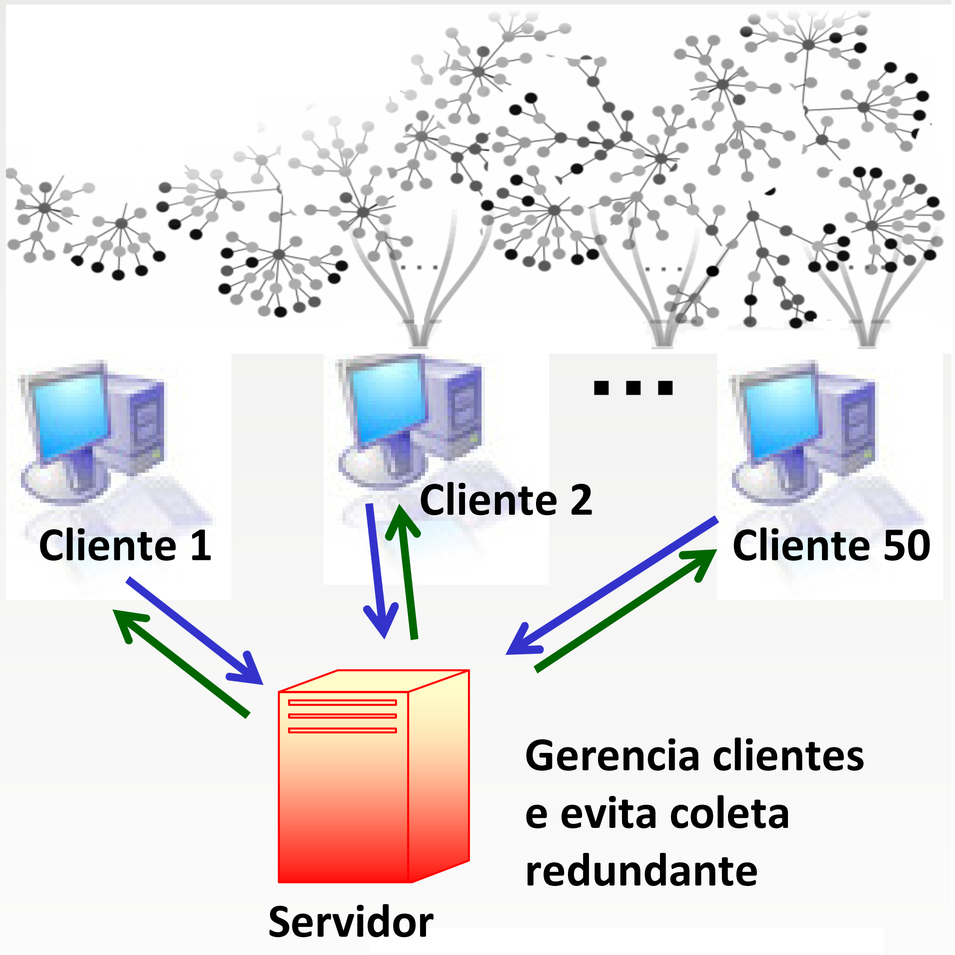
***Abstract*—Breadth First Search (BFS) and other graph traversal techniques are widely used for measuring large unknown graphs, such as online social networks. It has been empirically observed that incomplete BFS is biased toward high degree nodes. In contrast to more studied sampling techniques, such as random walks, the bias of BFS has not been characterized to date.**

In this paper, we quantify the degree bias of BFS sampling. In particular, we calculate the node degree distribution expected to



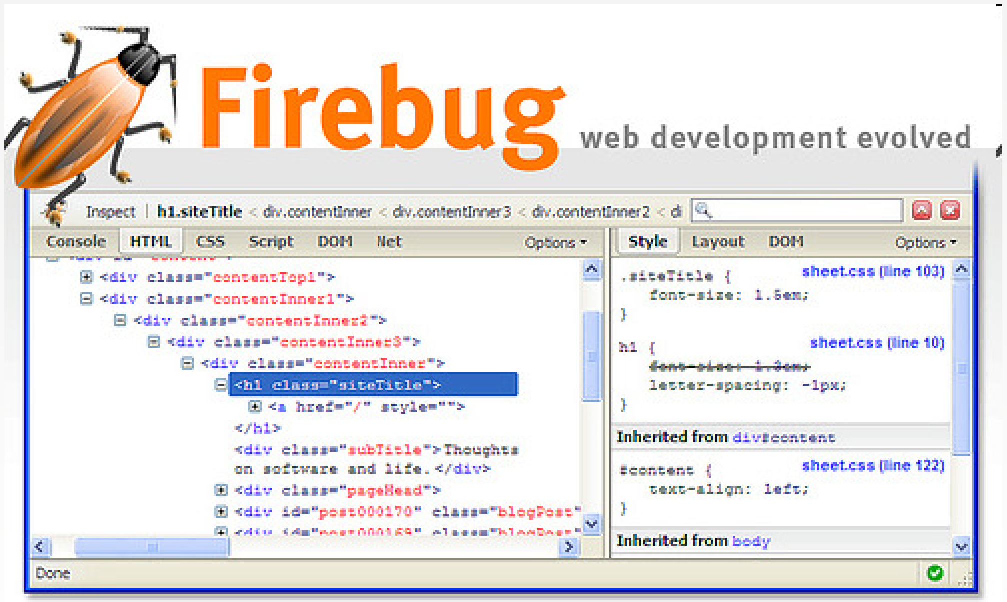
Crawlers distribuídos

- Clientes
  - Recebem páginas do servidor para coletar
  - Coletam páginas
  - Encontram novas páginas a serem coletadas e devolvem ao servidor
- Servidor
  - coordena clientes
  - evita redundância
  - O servidor pode ser um simples banco de dados



Firebug/tcpdump

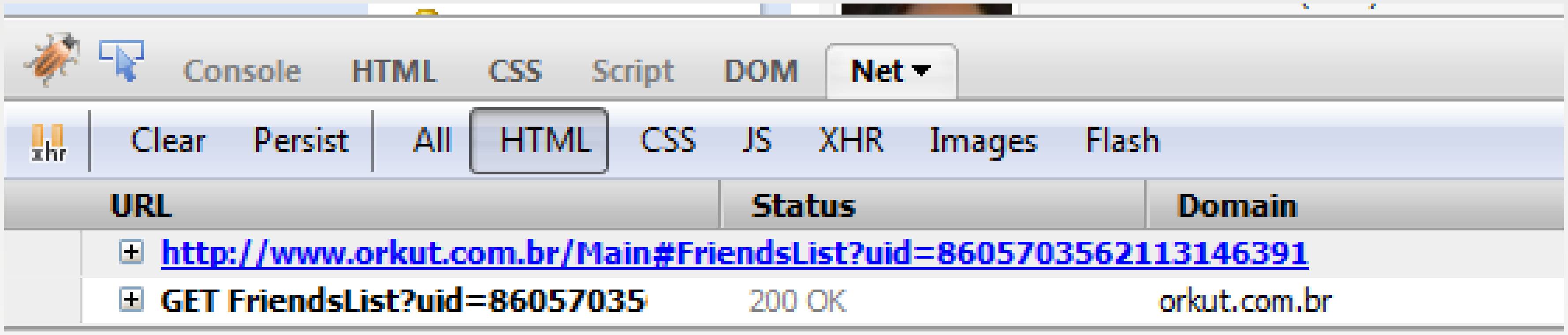
- JavaScript e Ajax muitas vezes escondem o HTML que procuramos com os crawlers.
- O Firebug é um add on do firefox que pode ajudar
- Qualquer ferramenta tcpdump like também pode resolver


Firebug
web development evolved
Inspect | h1.siteTitle < div.contentInner < div.contentInner3 < div.contentInner2 < di
Console HTML CSS Script DOM Net Options
Style Layout DOM Options
Done

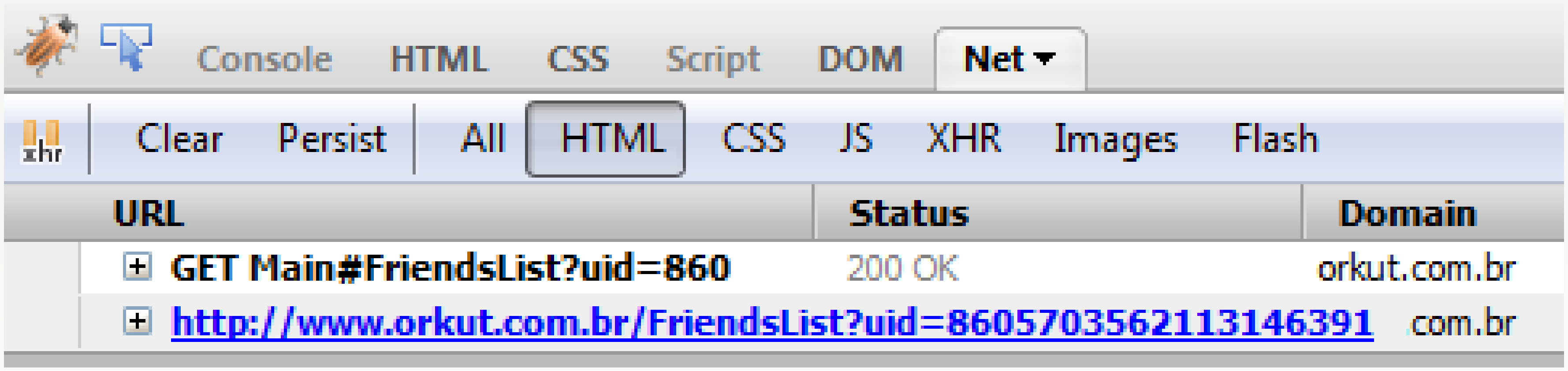
```
<div class="contentTop1">
<div class="contentInner1">
  <div class="contentInner2">
    <div class="contentInner3">
      <div class="contentInner">
        <h1 class="siteTitle">
          <a href="/" style="">
        </h1>
        <div class="subTitle">Thoughts
        on software and life.</div>
        <div class="pageHead">
        <div id="post000170" class="blogPost"
```

```
.siteTitle {                sheet.css (line 103)
    font-size: 1.5em;
}

h1 {                        sheet.css (line 10)
    ~~font-size: 1.3em;~~
    letter-spacing: -1px;
}
Inherited from div#content
#content {                  sheet.css (line 122)
    text-align: left;
}
Inherited from body
```

Coletando o Orkut



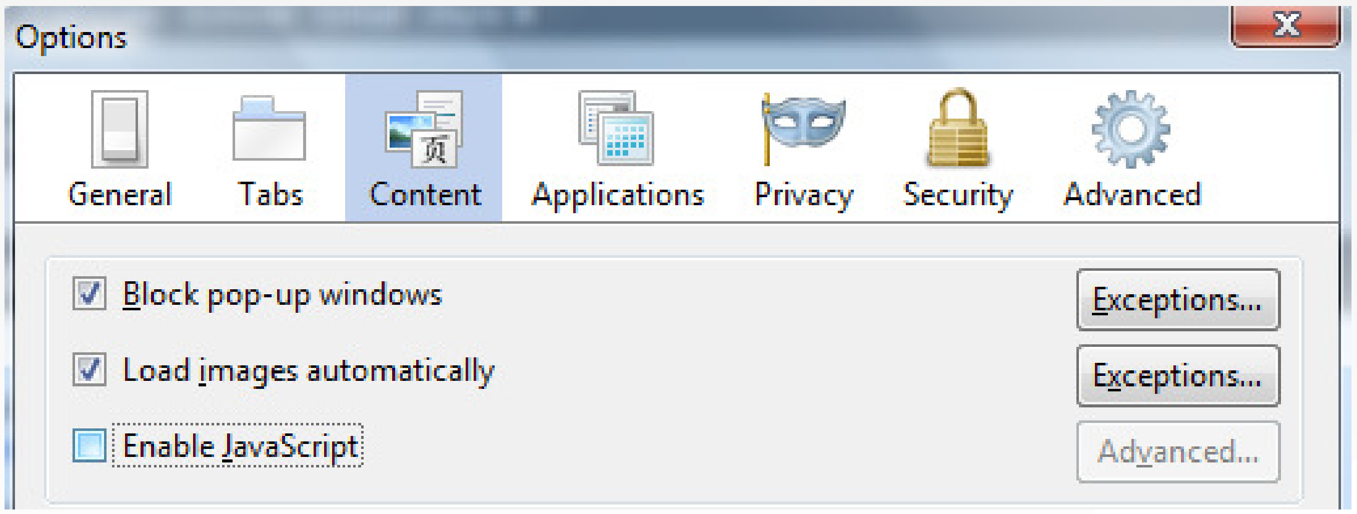
URL	Status	Domain
http://www.orkut.com.br/Main#FriendsList?uid=860570356211314639 1		
GET FriendsList?uid=86057035	200 OK	orkut.com.br



URL	Status	Domain
GET Main#FriendsList?uid=860	200 OK	orkut.com.br
http://www.orkut.com.br/FriendsList?uid=8605703562113146391		com.br

Coletando o Orkut

- Main# não permite que o fonte das páginas sejam visualizadas
  - http://www.orkut.com.br/**Main#**FriendsList?uid=8605703562113146391
- Solução: desabilitar Javascript e não utilizar o Main#
  - http://www.orkut.com.br/FriendsList?uid=8605703562113146391



[✓] Block pop-up windows
[✓] Load images automatically
[ ] Enable JavaScript

Coleta de IDs sequenciais

- IDs dos usuários são sequenciais no Twitter
  - Inspecionamos 80M de usuários, coletando perfil, todos os elos e tweets
  - Nenhum ID nas listas de seguidores/seguidos era superior a 80M
- Total de **55M de usuários, 2B de elos e 1.8B de tweets**
  - Cerca de 2 TB coletados
  - Lista branca para 58 máquinas no MPI-SWS
  - 20.000 requisições/hora em cada máquina
- Grafo de 55 milhões de nodos e 2 bilhões de arestas
  - Como armazenar um grafo desses?

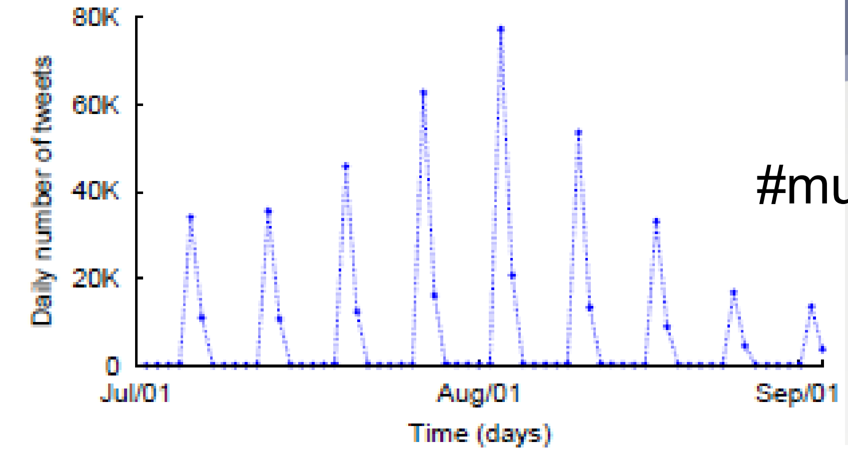
Informações coletadas

- **Informação do usuário:**
  userid, screen_name, nfollowers, nfollowees, ntweets, utc_offset, verified
  name, date, timezone, location

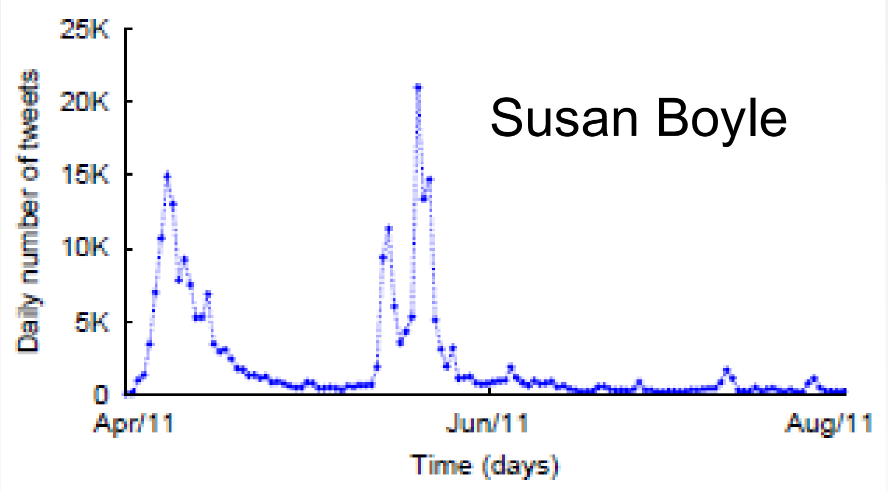
- **Informação dos links (seguidor/seguido):**
  userid_from userid_to

- **Informação dos tweets:**
  userid, tweetid, tweetid_replied, user_replied, date, source, text,
  screen_name, nfollowers, nfollowees, ntweets, utc_offset, protected
  verified, name, date, timezone, location

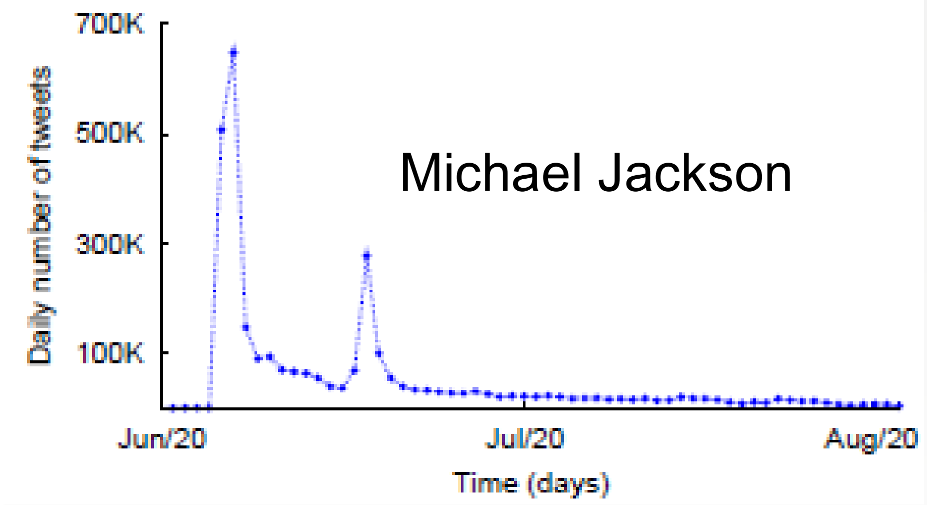
Permite reprodução de eventos



#musicmonday



Susan Boyle



Michael Jackson

ICWSM 2010

Measuring User Influence in Twitter: The Million Follower Fallacy

M. Cha[1], H. Haddadi[2], F. Benevenuto[3], K. Gummadi[4]

[1]Korea Advanced Institute of Science and Technology (KAIST)

[2]Unviersity of London

[3]Universidade Federal de Ouro Preto

[4]Max Planck Institute for Software Systems (MPI-SWS)

Our goal

Characterize influence in social media and study its dynamics (Influence: potential to cause others to engage in a certain act)

1. How can we **measure influence** of a single user?

2. Does influence of a user hold **across topics?**

3. What behaviors make ordinary users influential?

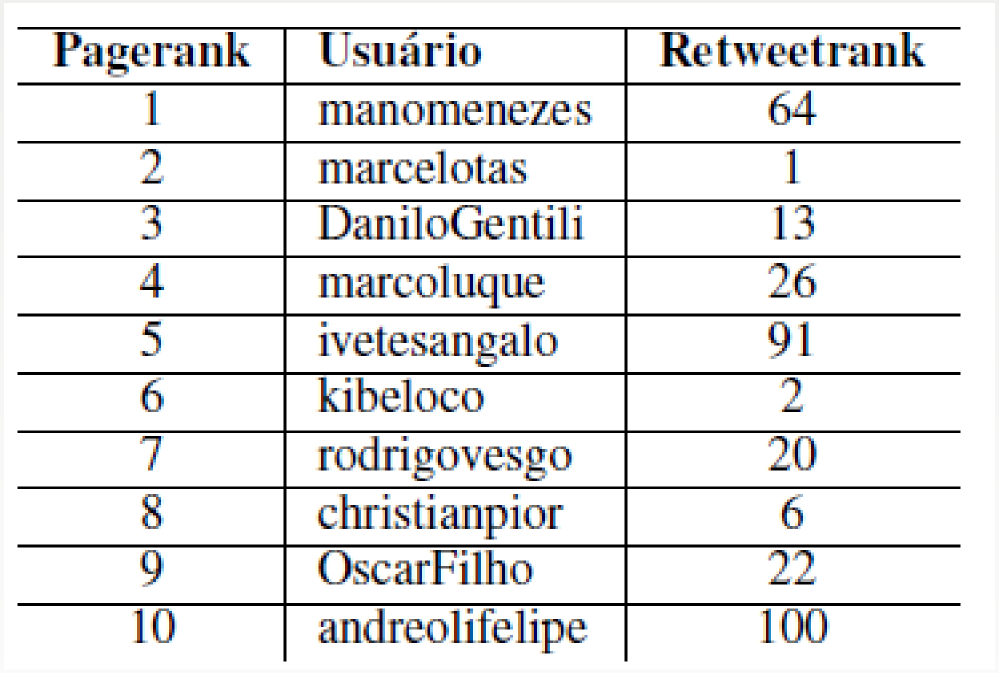
Considered Twitter as a medium of influence for our study

Example from the top 100 users

Indegree	rank 1 3.3M	rank 4 2.6M	rank 2 3.1M
Mentions	rank 6	-	rank 71
Retweets	rank 7	rank 24	-

The million follower fallacy!

E os usuários mais influentes brasileiros? (artigo submetido ao Webmedia 2011)



Pagerank	Usuário	Retweetrank
1	manomenezes	64
2	marcelotas	1
3	DaniloGentili	13
4	marcoluque	26
5	ivetesangalo	91
6	kibeloco	2
7	rodrigovesgo	20
8	christianpior	6
9	OscarFilho	22
10	andreolifelipe	100

ACM SIGCOMM WOSN 2009

Hot Today, Gone Tomorrow: On the Migration of MySpace Users

Mojtaba Torkjazi[1], Reza Rejaie[1], Walter Willinger[2]

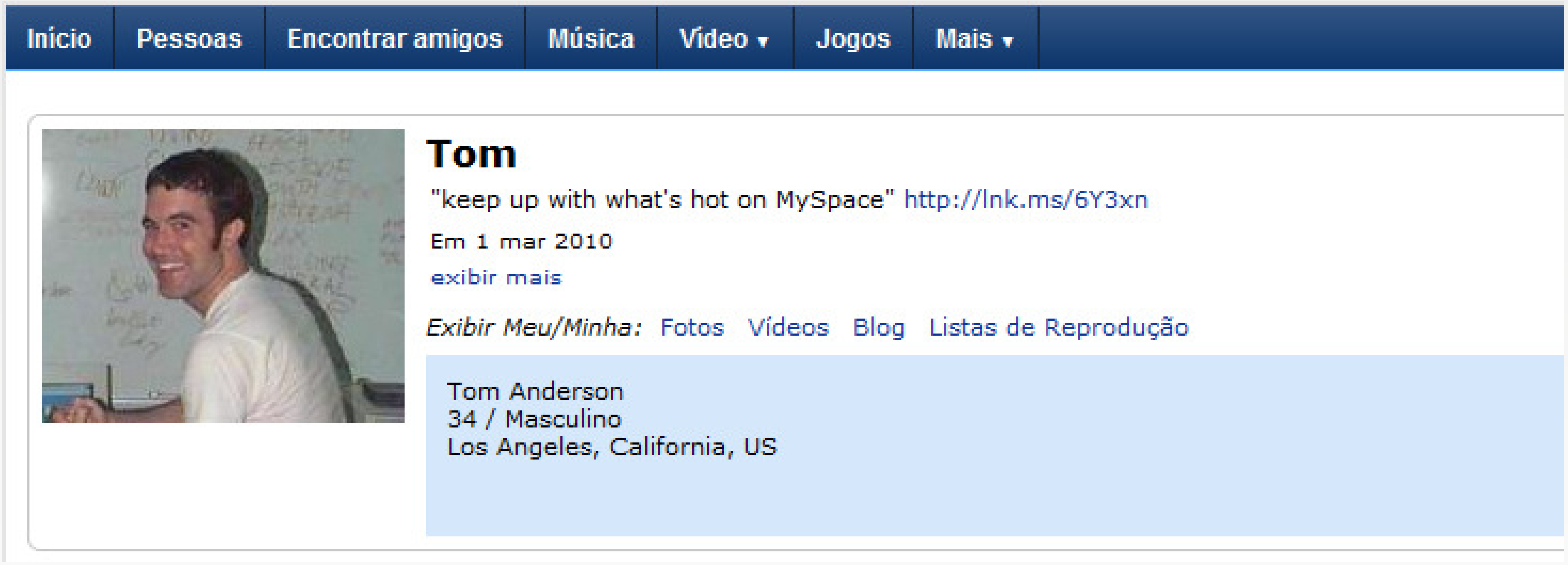
[1] University of Oregon

[2] AT&T Labs-Research

MySpace Features

- Provides explicit profile status
  - Public
  - Private
  - Invalid
- Availability of users' last login
  - Enables assessment of the level of activity among users
  - Importantly, allows inference of population growth of MySpace (see later for details)
- Global visibility
  - http://www.myspace.com/user_id
- Monotonic assignment of numeric ID

Todos tem pelo menos 1 amigo no MySpace



Tom

"keep up with what's hot on MySpace" http://lnk.ms/6Y3xn

Em 1 mar 2010

exibir mais

Exibir Meu/Minha: Fotos Vídeos Blog Listas de Reprodução

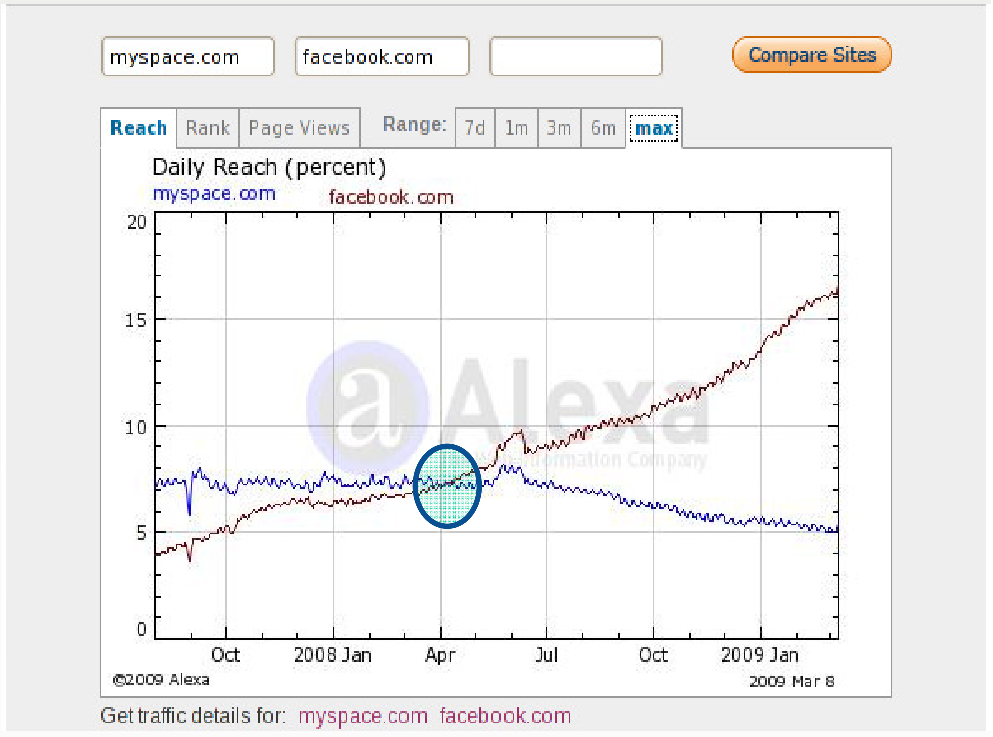
Tom Anderson
34 / Masculino
Los Angeles, California, US

Measurement

- Feb. 26th 2009: MySpace ID space [1 … 455,881,700]
- 50 parallel samplers to collect 360K users in less than 12 hours (0.1% of MySpace population)
- Using HTML parser to post-process the downloaded profiles and extract
  - User s' profile status (invalid, public, private)
  - Users' last login date
  - Users' friend list (only for public profiles)
- Unable to parse last login info for 0.96% of public and 0.08% of private profiles
  - Last login info is not provided or is provided with obvious errors (e.g. 1/1/0001)

MySpace Life Cycle

- *Possible reasons behind MySpace's decline?*
  - Slow-down in the growth rate of MySpace is related to emergence of Facebook
  - Informal evidence (Alexa.com): Daily accesses to Facebook surpassed that of MySpace, at around April 2008



Get traffic details for: myspace.com facebook.com

ACM SIGIRG/SIGKDD WSDM 2011

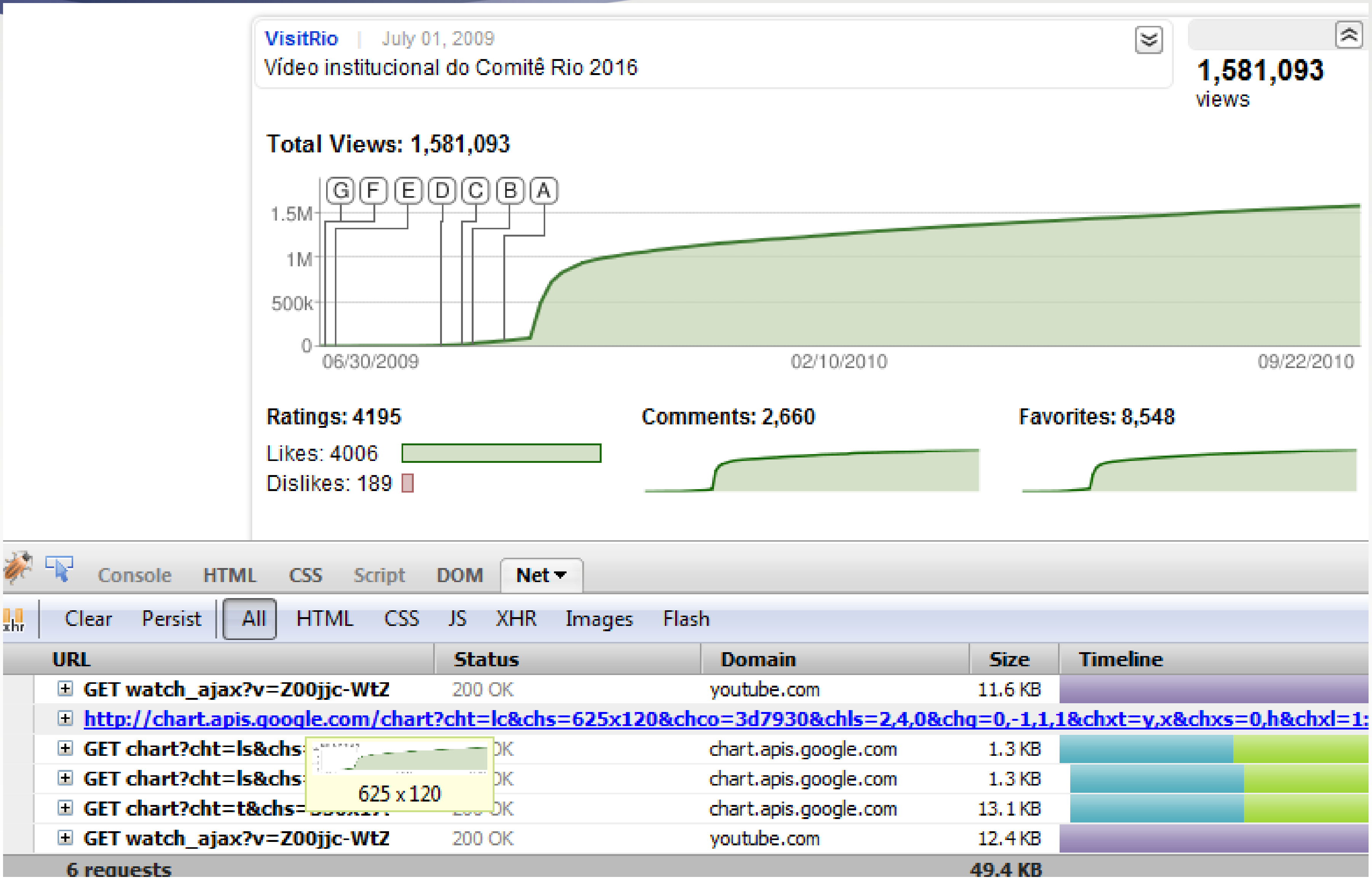
The Tube over Time: Characterizing Popularity Growth of YouTube Videos

F. Figueiredo[1], F. Benevenuto[2], J. Almeida[1]

[1]Universidade Federal de Minas Gerais (UFMG)

[2]Universidade Federal de Ouro Preto (UFOP)

Ajax no YouTube



VisitRio | July 01, 2009

Vídeo institucional do Comitê Rio 2016

1,581,093 views

Total Views: 1,581,093

Ratings: 4195

Likes: 4006

Dislikes: 189

Comments: 2,660

Favorites: 8,548

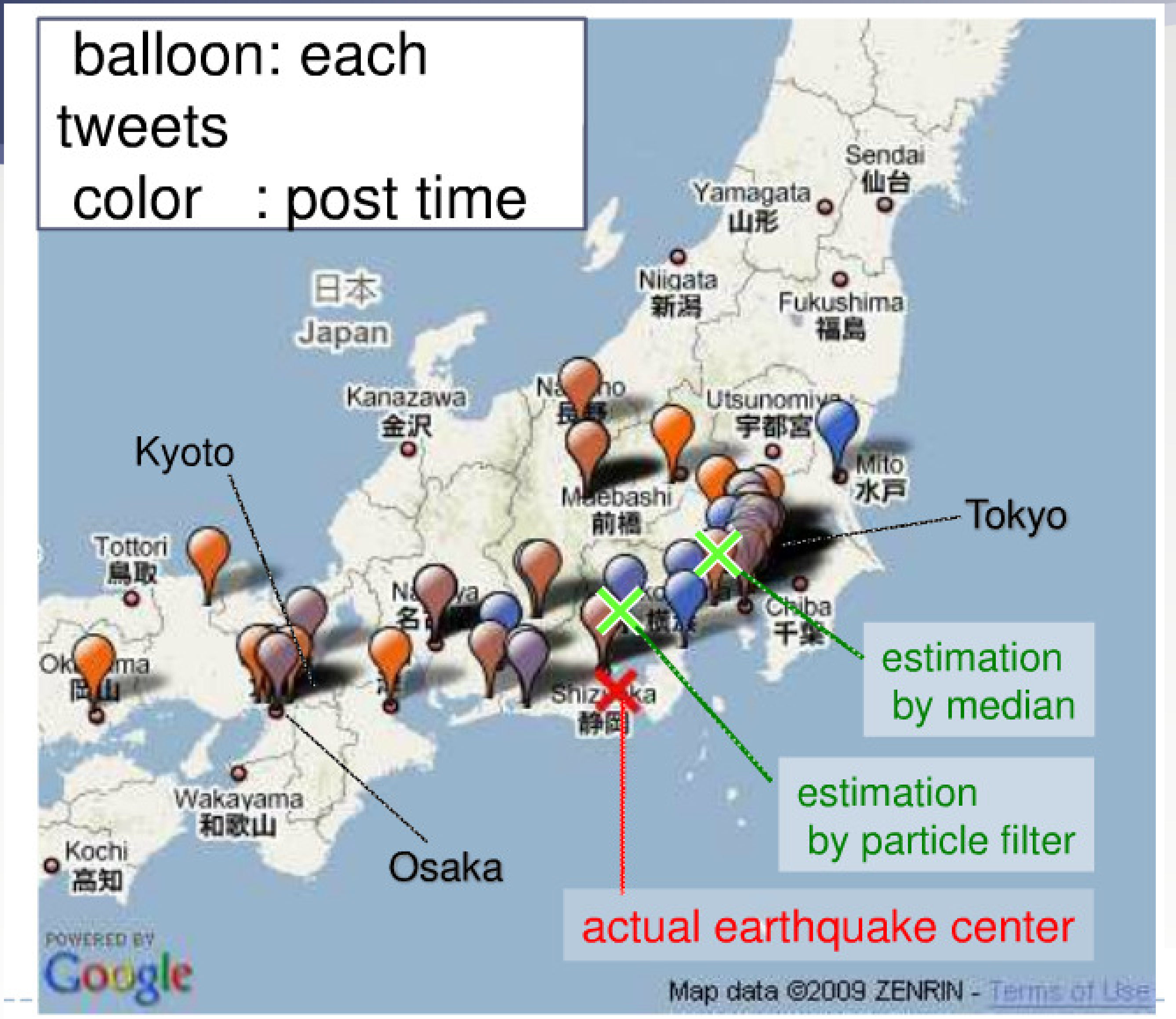
URL	Status	Domain	Size	Timeline
GET watch_ajax?v=Z00jjc-WtZ	200 OK	youtube.com	11.6 KB	
http://chart.apis.google.com/chart?cht=lc&chs=625x120&chco=3d7930&chls=2,4,0&chg=0,-1,1,1&chxt=y,x&chxs=0,h&chxl=1:				
GET chart?cht=ls&chs	OK	chart.apis.google.com	1.3 KB	
GET chart?cht=ls&chs	OK	chart.apis.google.com	1.3 KB	
GET chart?cht=t&chs=	OK	chart.apis.google.com	13.1 KB	
GET watch_ajax?v=Z00jjc-WtZ	200 OK	youtube.com	12.4 KB	
6 requests			49.4 KB	

625 x 120

WWW 2010

Earthquake Shakes Twitter User:
Analyzing Tweets for Real-Time Event Detection

Takehi Sakaki @tksakaki
Makoto Okazaki @okazaki117
Yutaka Matsuo @ymatsuo
the University of Tokyo

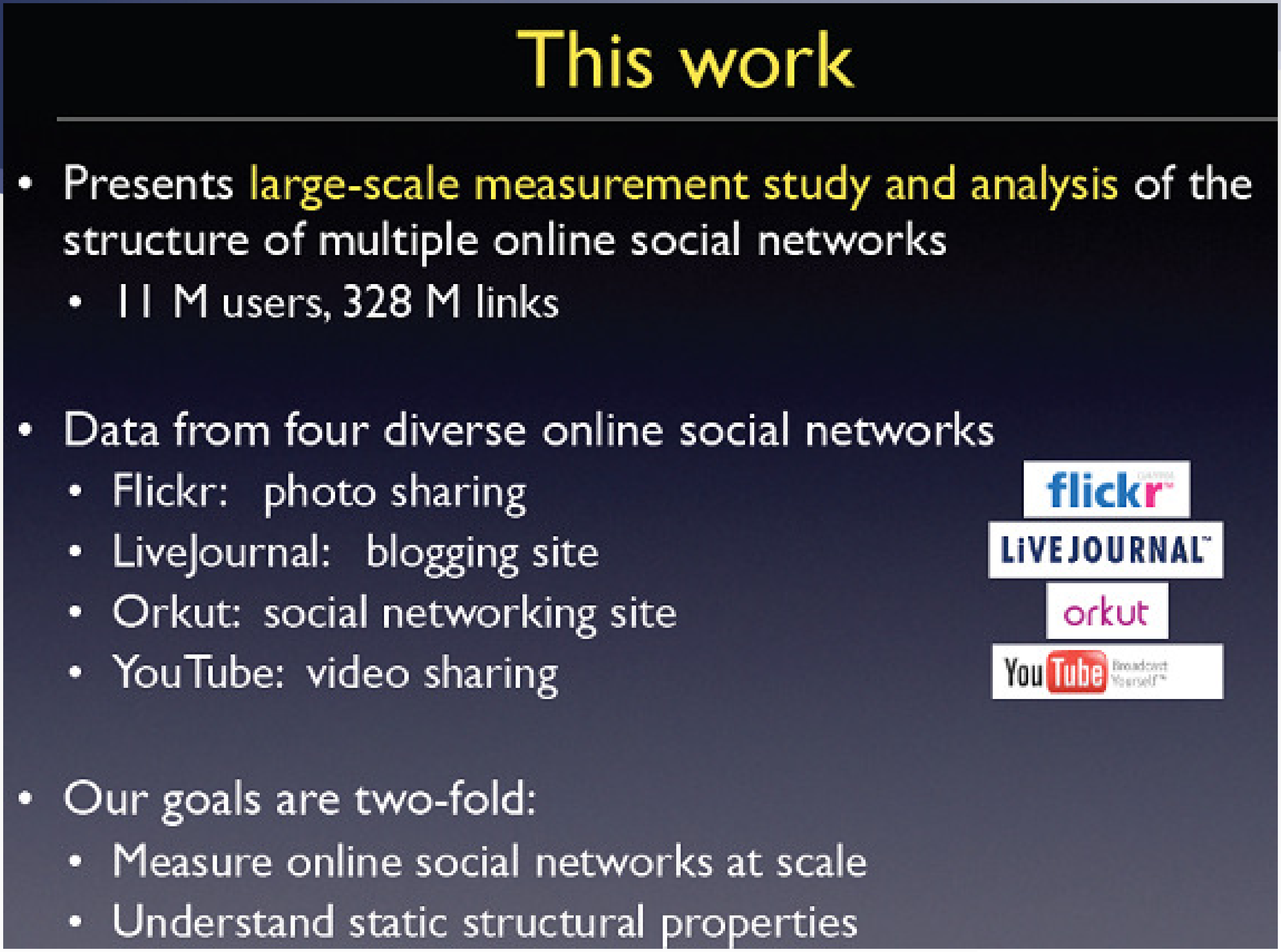

balloon: each tweets
color : post time
日本
Japan
Kyoto
Osaka
Tokyo
Sendai 仙台
Yamagata 山形
Niigata 新潟
Fukushima 福島
Kanazawa 金沢
Utsunomiya 宇都宮
Mito 水戸
Maebashi 前橋
Tottori 鳥取
Okayama 岡山
Chiba 千葉
Shizuoka 静岡
Wakayama 和歌山
Kochi 高知
estimation by median
estimation by particle filter
actual earthquake center
POWERED BY Google
Map data ©2009 ZENRIN - Terms of Use

ACM IMC 2007

Measurement and Analysis of Online Social Networks

Alan Mislove, Massimiliano Marcon, Krishna Gummadi,
Peter Druschel, Bobby Bhattacharjee

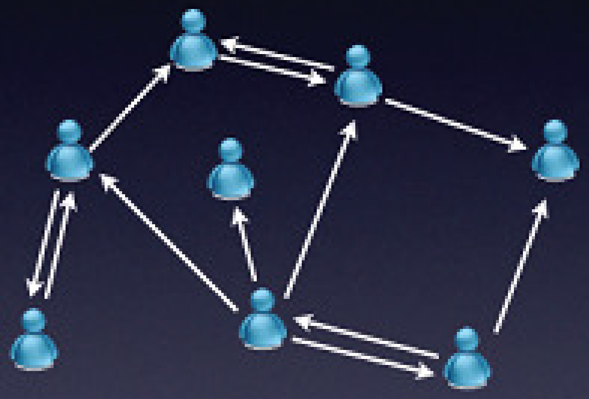
Max Planck Institute for Software Systems (MPI-SWS)



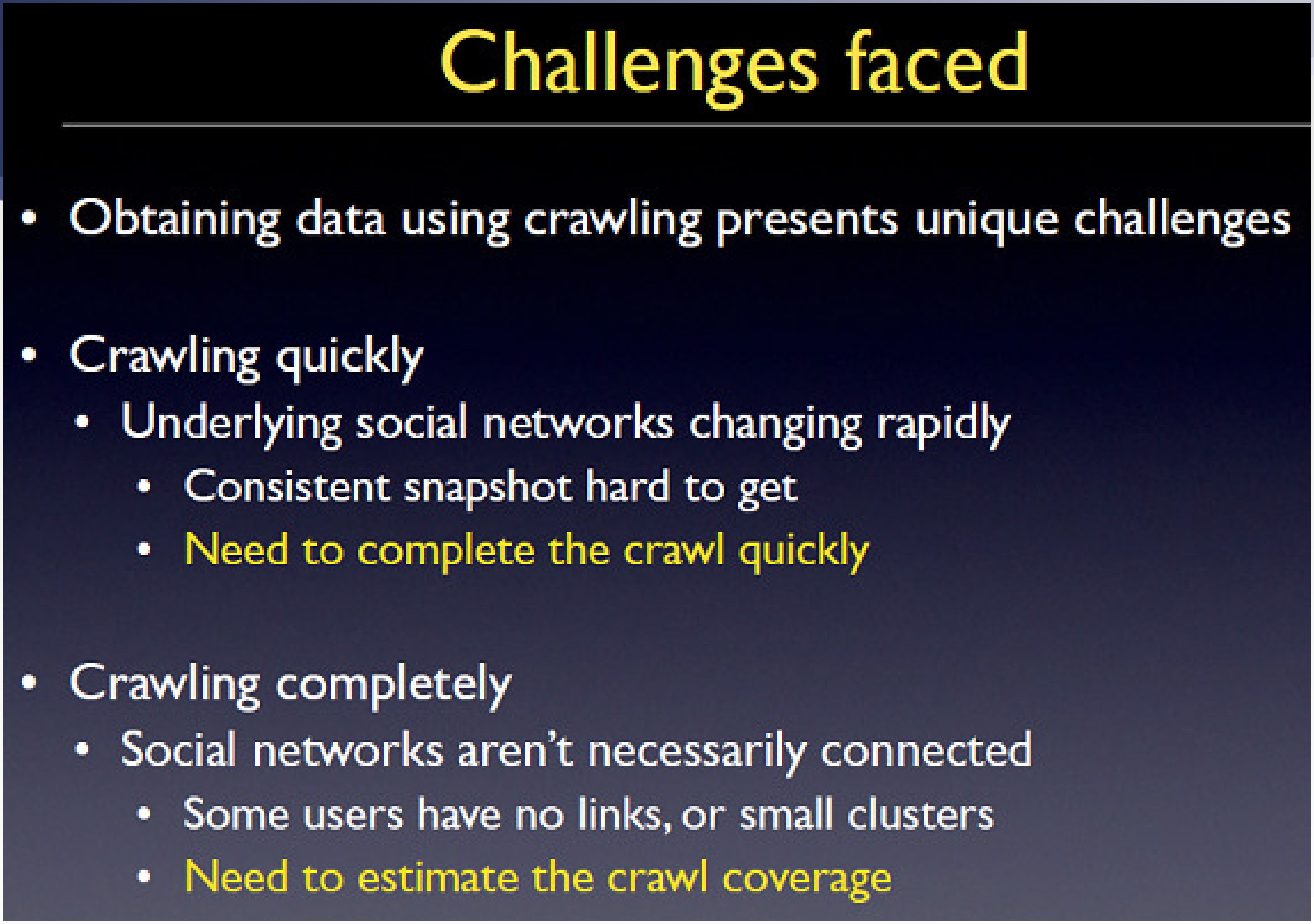
This work

- Presents large-scale measurement study and analysis of the structure of multiple online social networks
  - 11 M users, 328 M links
- Data from four diverse online social networks
  - Flickr: photo sharing
  - LiveJournal: blogging site
  - Orkut: social networking site
  - YouTube: video sharing
- Our goals are two-fold:
  - Measure online social networks at scale
  - Understand static structural properties

Medição de OSNs



- Sites reluctant to give out data
  - Cannot enumerate user list
  - Instead, performed crawls of user graph
- Picked known seed user
  - Crawled all of his friends
  - Added new users to list
- Continued until all known users crawled
- Effectively performed a BFS of graph

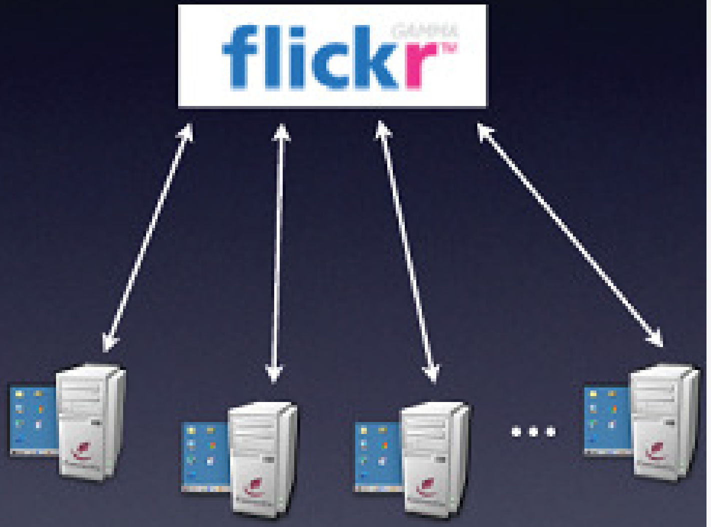


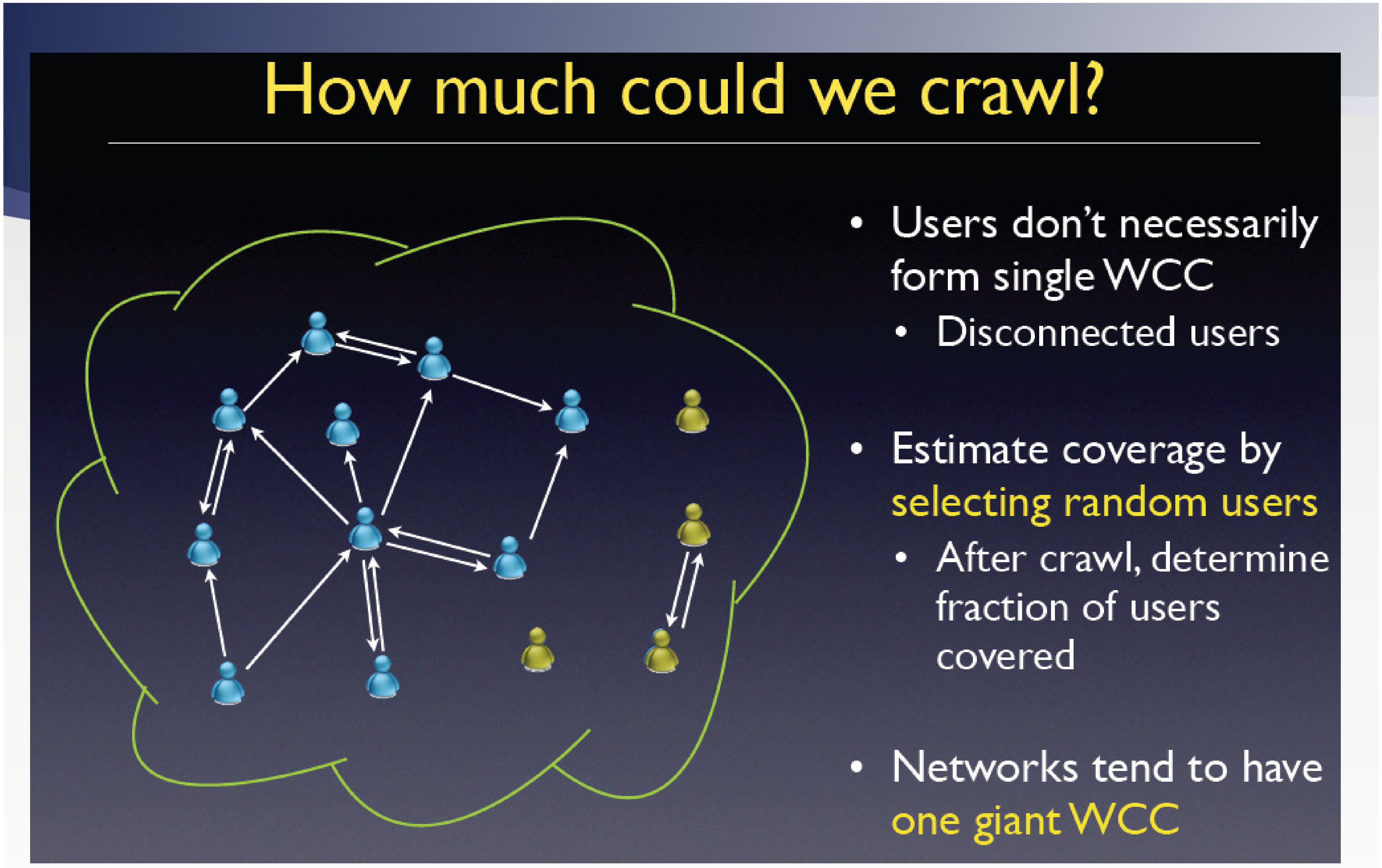
Challenges faced

- Obtaining data using crawling presents unique challenges
- Crawling quickly
  - Underlying social networks changing rapidly
    - Consistent snapshot hard to get
    - Need to complete the crawl quickly
- Crawling completely
  - Social networks aren't necessarily connected
    - Some users have no links, or small clusters
    - Need to estimate the crawl coverage

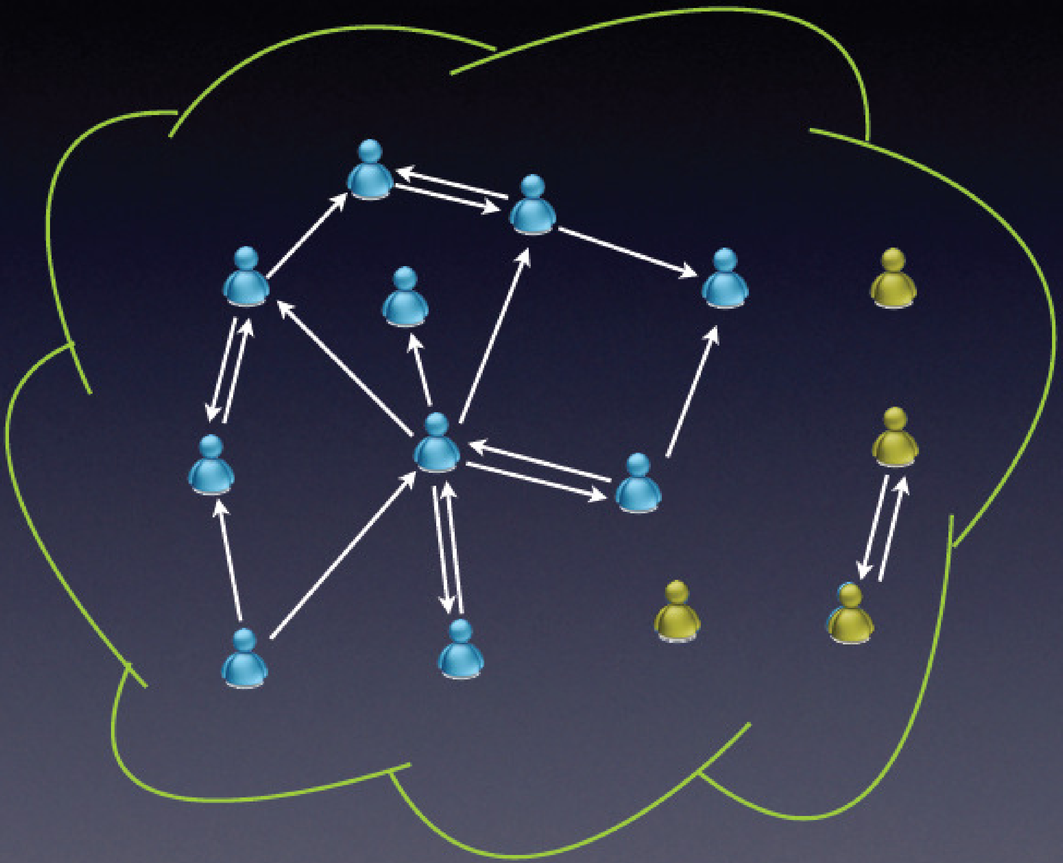
How fast could we crawl?

- Crawled using cluster of 58 machines
  - Used APIs where available
  - Otherwise, used screen scraping
- Crawls took varying times
  - Flickr, YouTube: 1 day
  - LiveJournal: 3 days
  - Orkut (partial): 39 days
- Crawls subject to rate-limiting
  - Discovered appropriate rates





How much could we crawl?



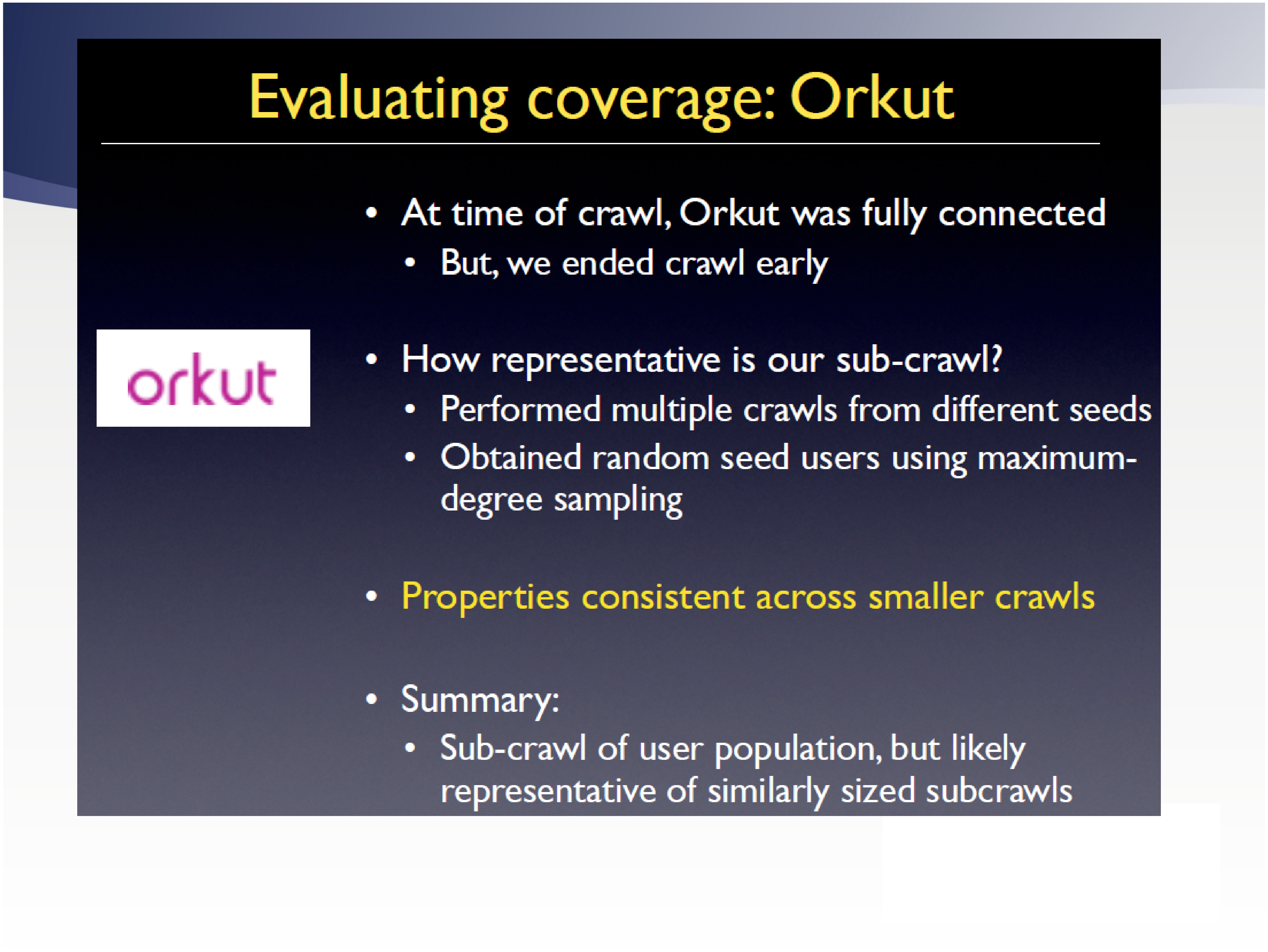
- Users don't necessarily form single WCC
  - Disconnected users
- Estimate coverage by selecting random users
  - After crawl, determine fraction of users covered
- Networks tend to have one giant WCC

Evaluating coverage: Flickr

- Obtained random users by guessing usernames (########@N00)
- Fraction of disconnected users is 73%
- But, disconnected users have very low degree
  - 90% have no outgoing links, remaining 10% have few links
- Summary:
  - Covered 27% of user population, but remaining users have very few links

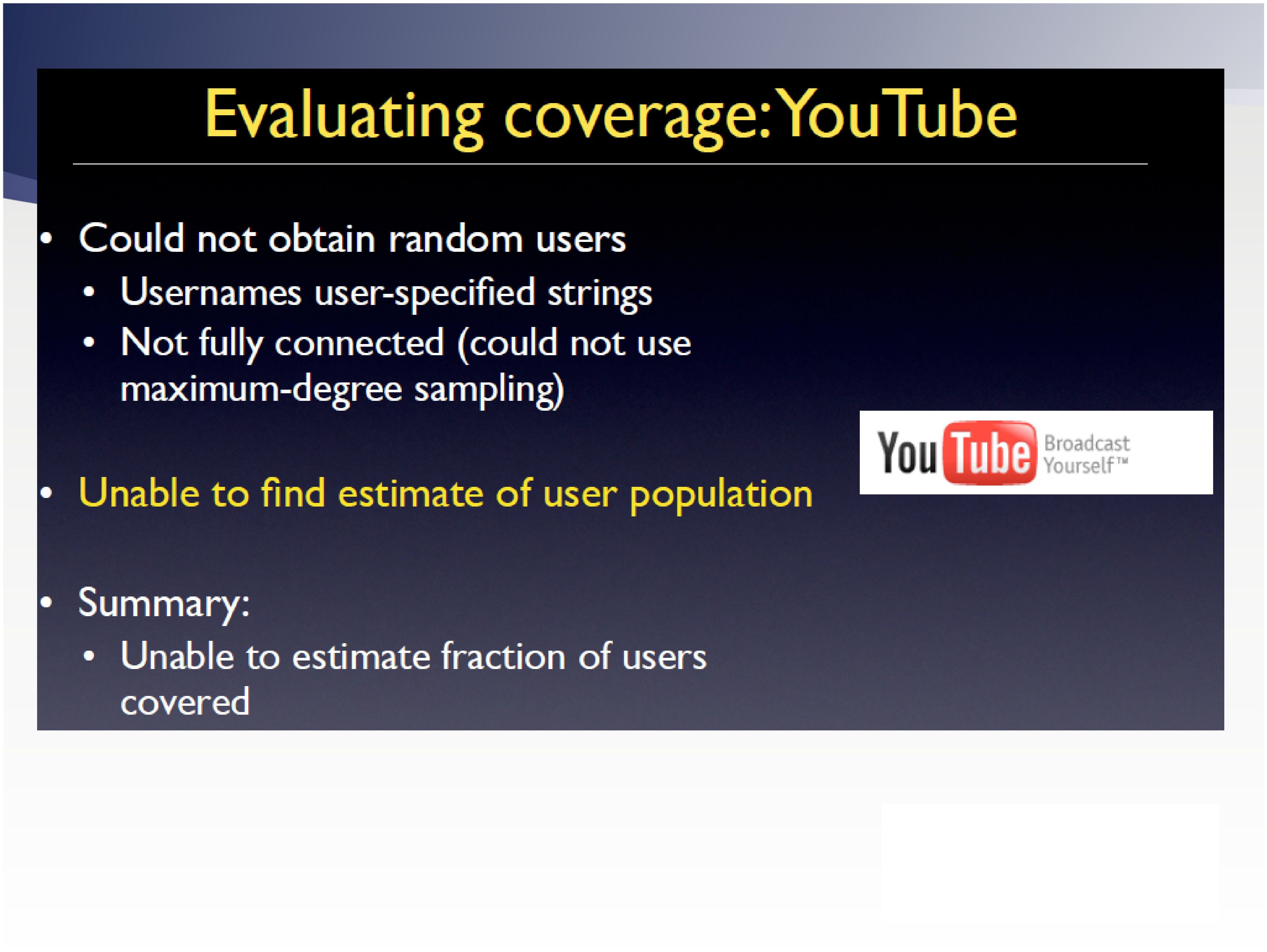
Evaluating coverage: LiveJournal

- Obtained random users using special URL
  - `http://www.livejournal.com/random.bml`
- Fraction of disconnected users is only 5%
- Summary:
  - Crawl covered 95% of user population



Evaluating coverage: Orkut

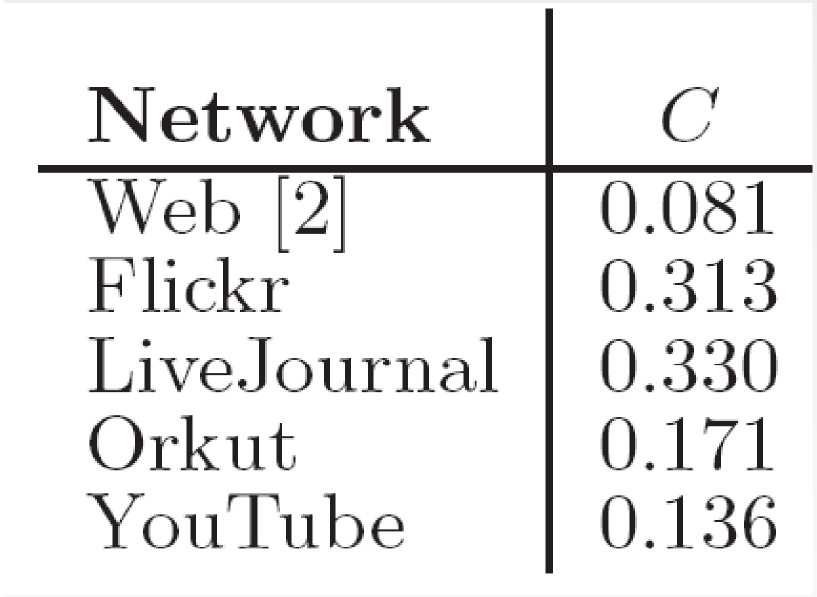
- At time of crawl, Orkut was fully connected
  - But, we ended crawl early
- How representative is our sub-crawl?
  - Performed multiple crawls from different seeds
  - Obtained random seed users using maximum-degree sampling
- Properties consistent across smaller crawls
- Summary:
  - Sub-crawl of user population, but likely representative of similarly sized subcrawls



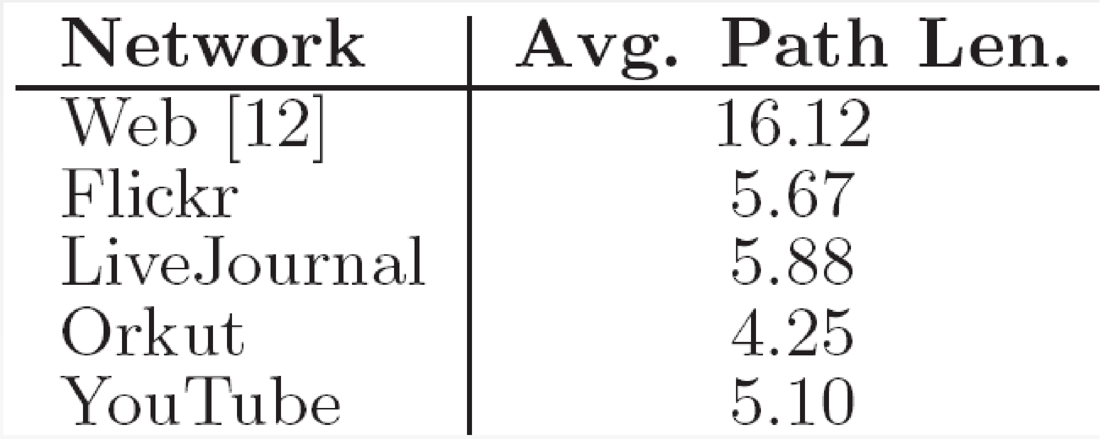
Evaluating coverage: YouTube

- Could not obtain random users
  - Usernames user-specified strings
  - Not fully connected (could not use maximum-degree sampling)
- Unable to find estimate of user population
- Summary:
  - Unable to estimate fraction of users covered

Confirmou propriedades small-world



Network	C
Web [2]	0.081
Flickr	0.313
LiveJournal	0.330
Orkut	0.171
YouTube	0.136



Network	Avg. Path Len.
Web [12]	16.12
Flickr	5.67
LiveJournal	5.88
Orkut	4.25
YouTube	5.10

Redes sociais online possuem características Small World

ACM TOMCCAP 2009

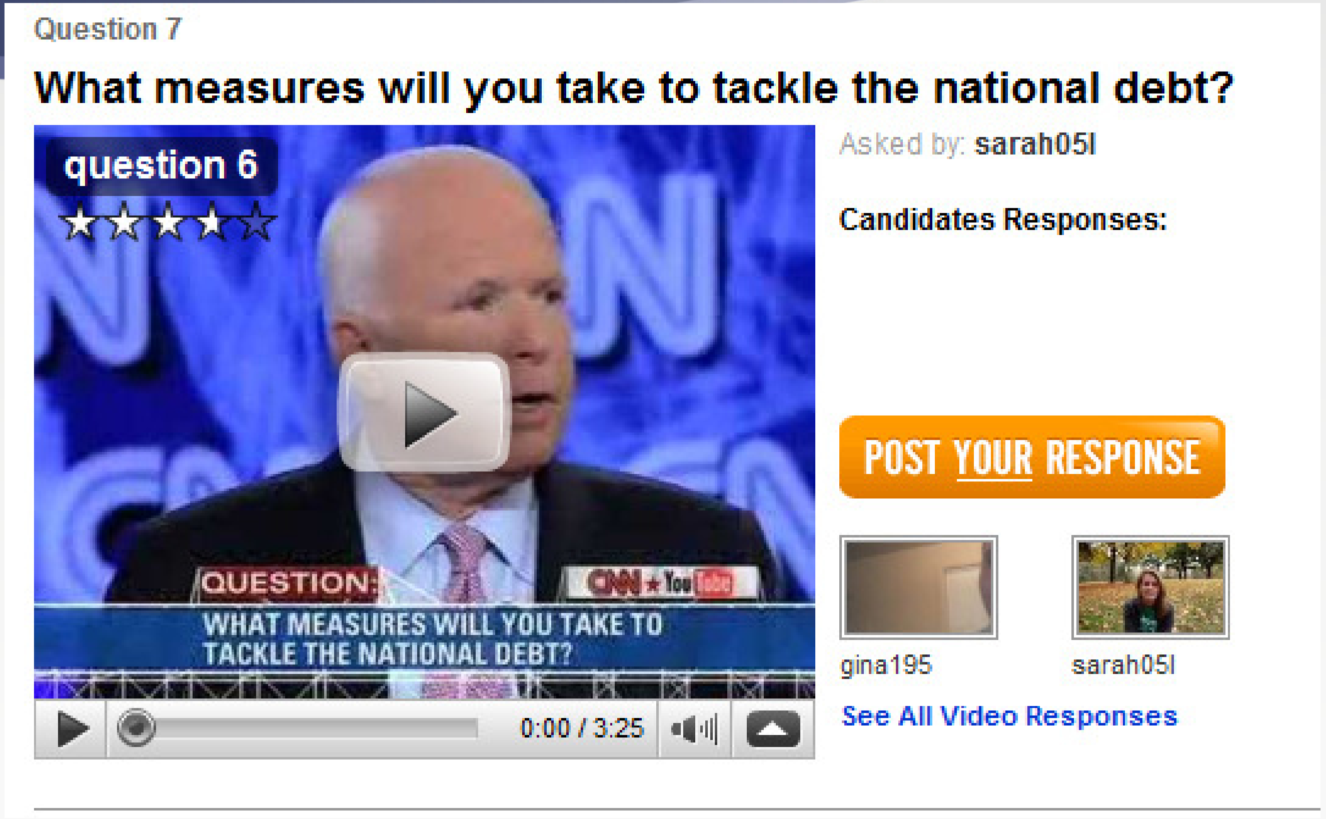
Video interactions in Online Video Social Networks

F. Benevenuto[1], T. Rodrigues[1], V. Almeida[1], J. Almeida, K. Ross[2]

[1]Universidade Federal de Minas Gerais

[2]Polytechnic Institute of NYU

Detecção de usuários oportunistas



Question 7

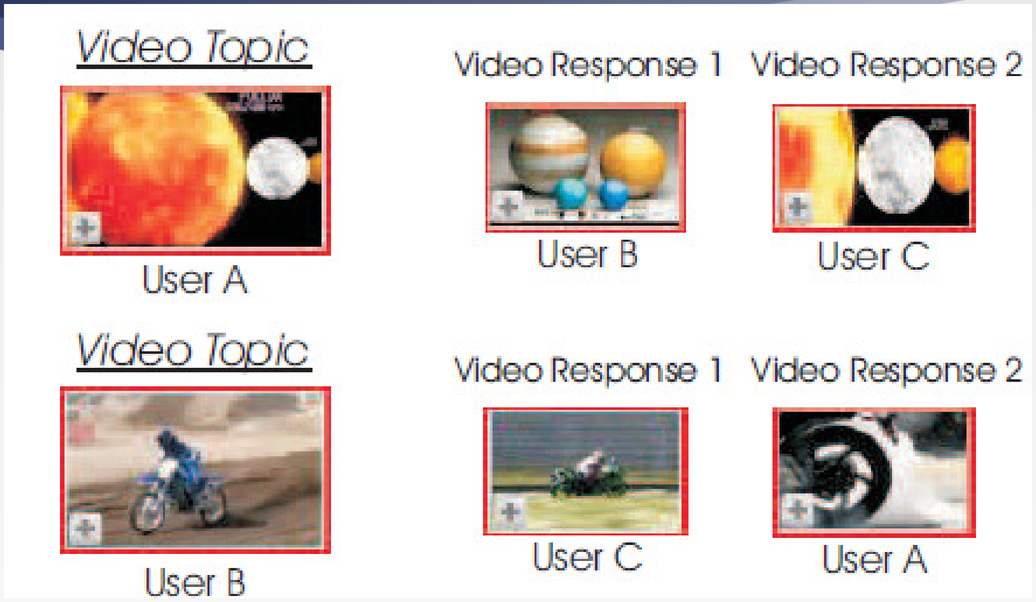
What measures will you take to tackle the national debt?

Asked by: sarah05l

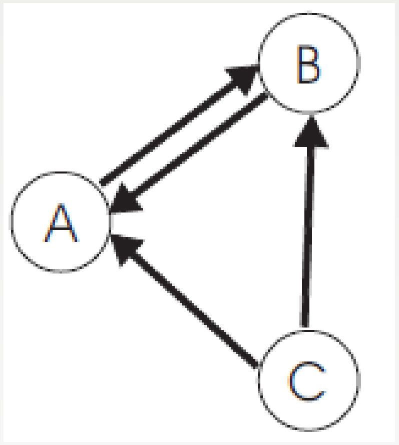
Candidates Responses:

Longas discussões em alguns tópicos

Coleta de vídeo respostas



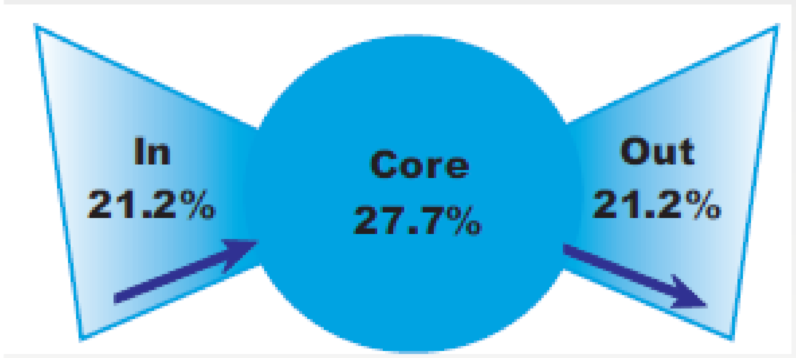
Video response user graph



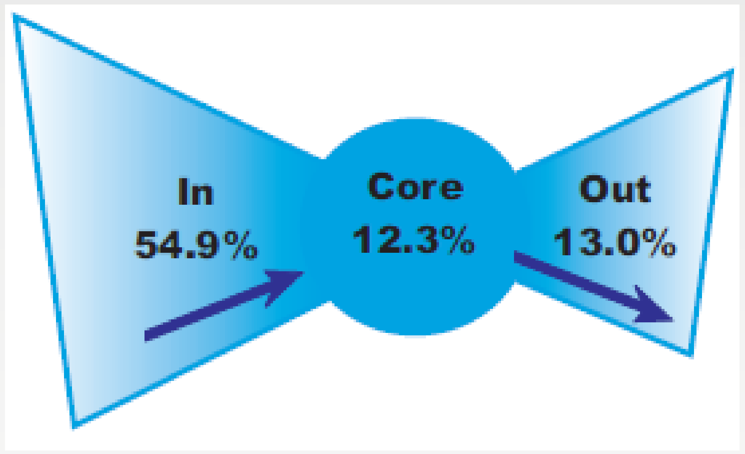
- Effective performed a BFS of our graph
- Collect entire weakly connected components (WCCs)
- **417,759** video responses, **223,851** video topics, and**160,765** users
- Validation with random searches

Bow-tie structure

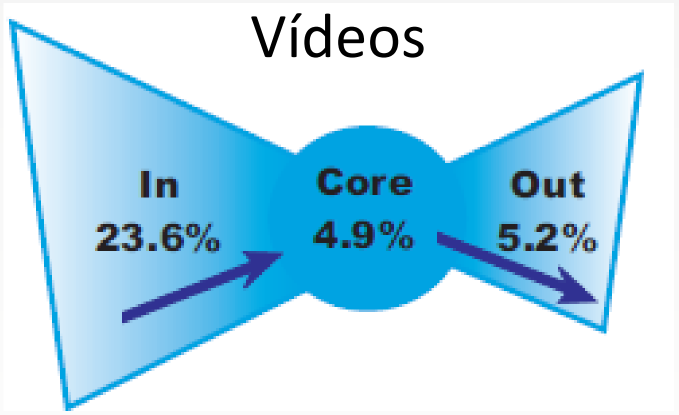
Web



Java Fórum



Vídeos



ACM SIGCOMM WOSN 2009

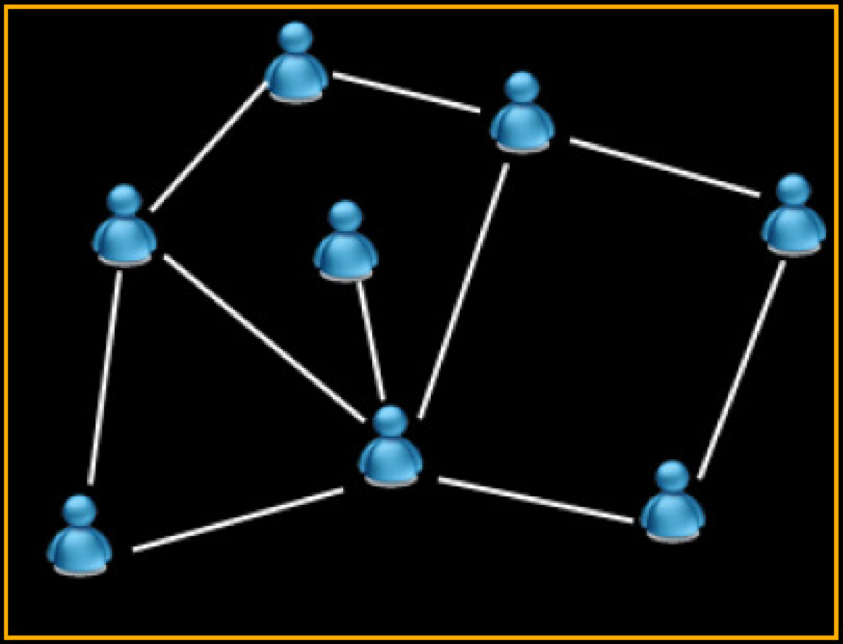
On the Evolution of User Interactions in Facebook

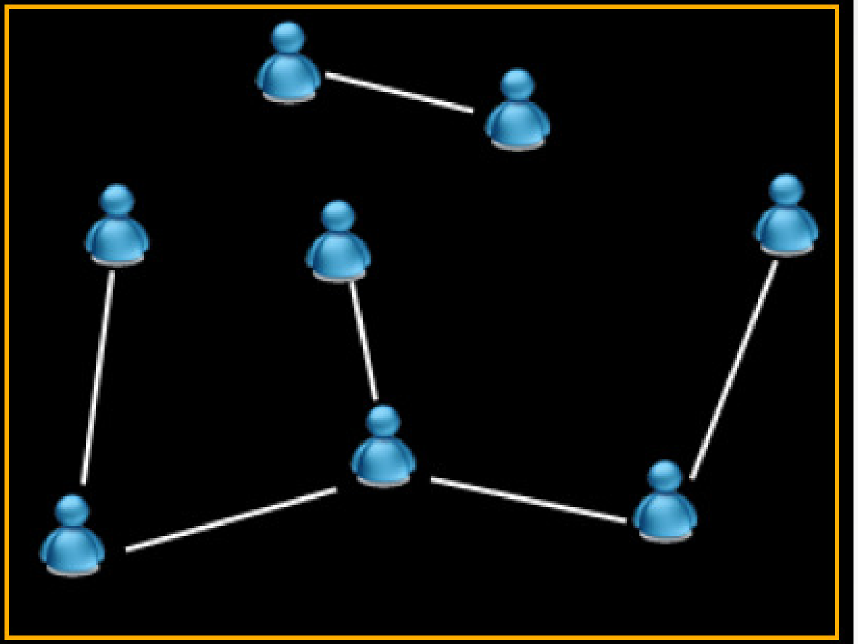
B. Viswanath, A. Mislove, M. Cha, K. Gummadi

Max Planck Institute for Software Systems (MPI-SWS)

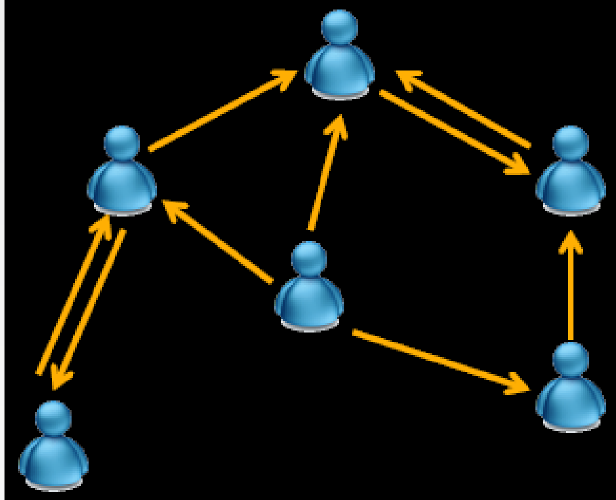
Gauging the strength of social links?

* Idea: Use interaction to differentiate strong and weak links


Social network

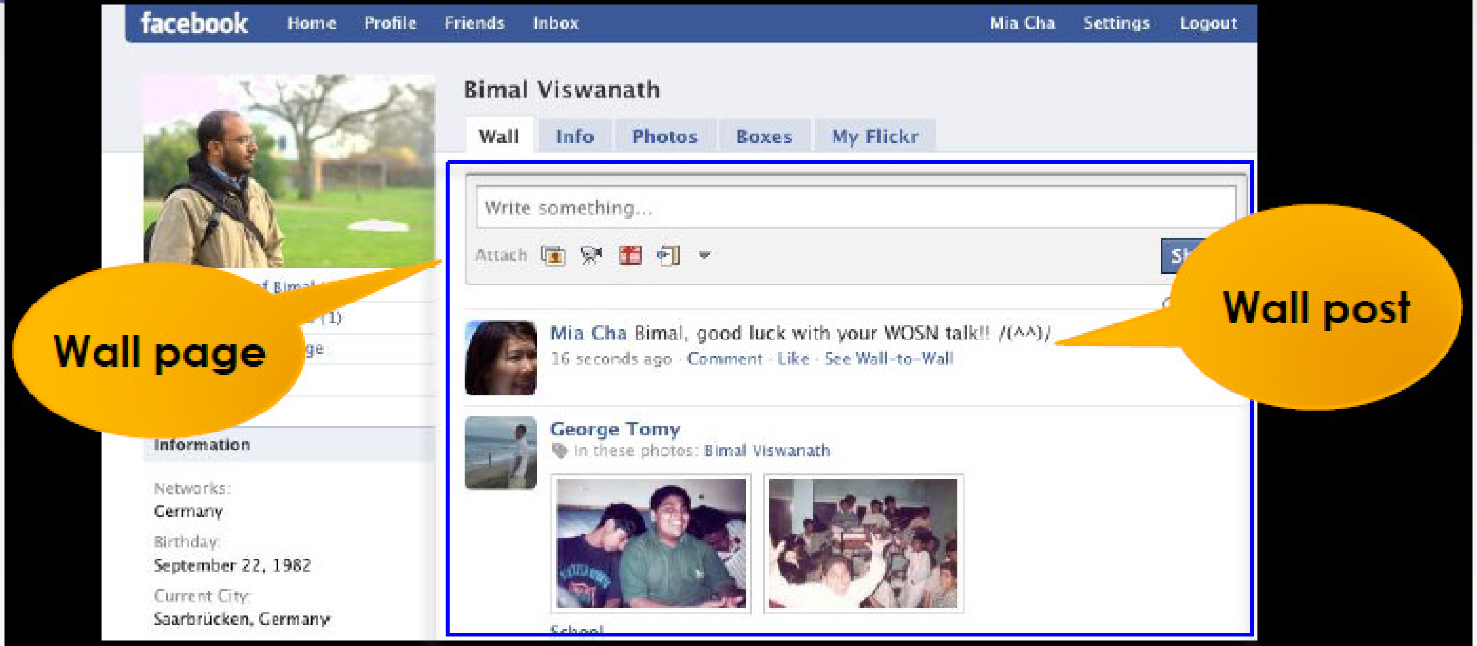

Interaction network

Crawling Facebook



- Facebook reluctant to give out data
  - Performed crawl of user graph
- Picked known seed user
  - Crawled all of his friends
  - Add new users to list
- Continued until all reachable users crawled
- Crawled Facebook New Orleans regional network
  - Over 90,000 users, 3M social links
- We could create many crawling accounts

Collected interaction data



Bimal Viswanath

Mia Cha Bimal, good luck with your WOSN talk!! /(^^)/
16 seconds ago · Comment · Like · See Wall-to-Wall

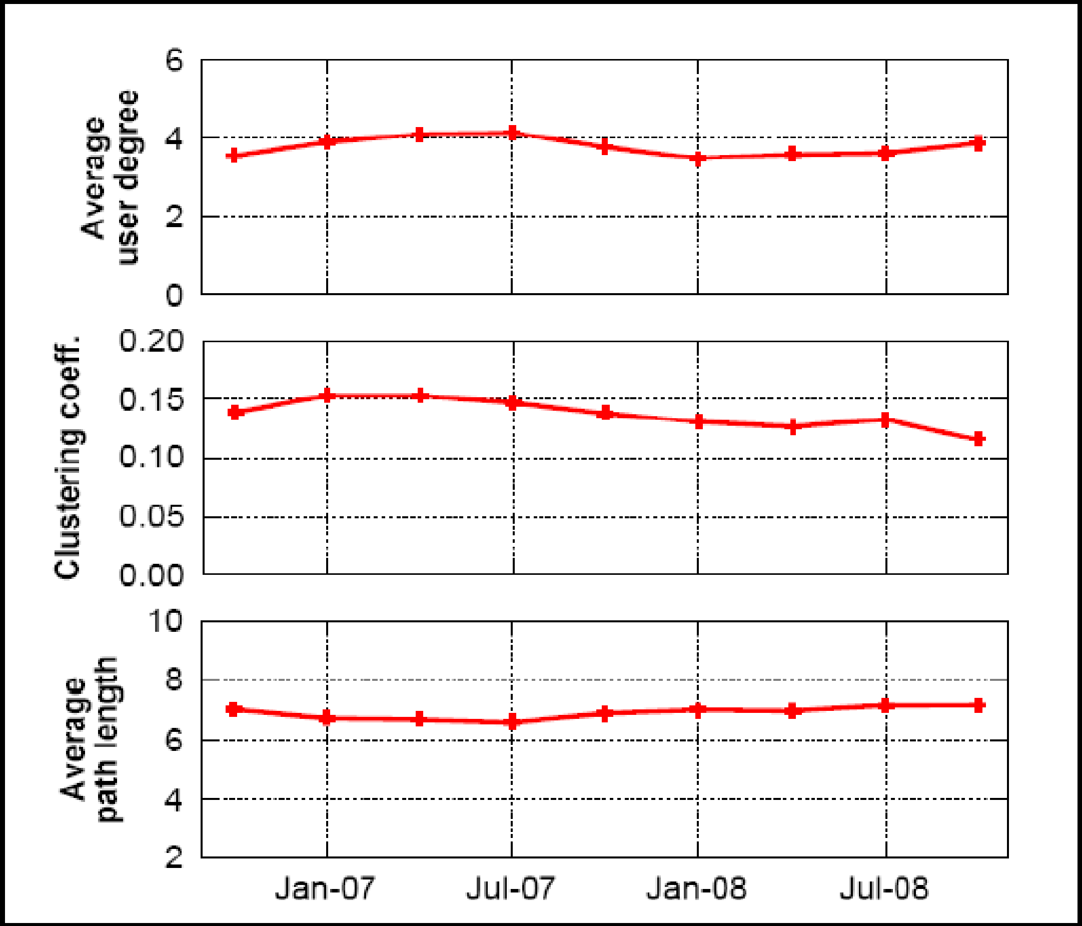
George Tomy
In these photos: Bimal Viswanath

Information

Networks:
Germany
Birthday:
September 22, 1982
Current City:
Saarbrücken, Germany

- Able to download entire wall history
- 800,000 wall posts
- Link creation time known from wall page

Evolution of structural properties



Graph properties remarkably stable

Ética dos crawlers

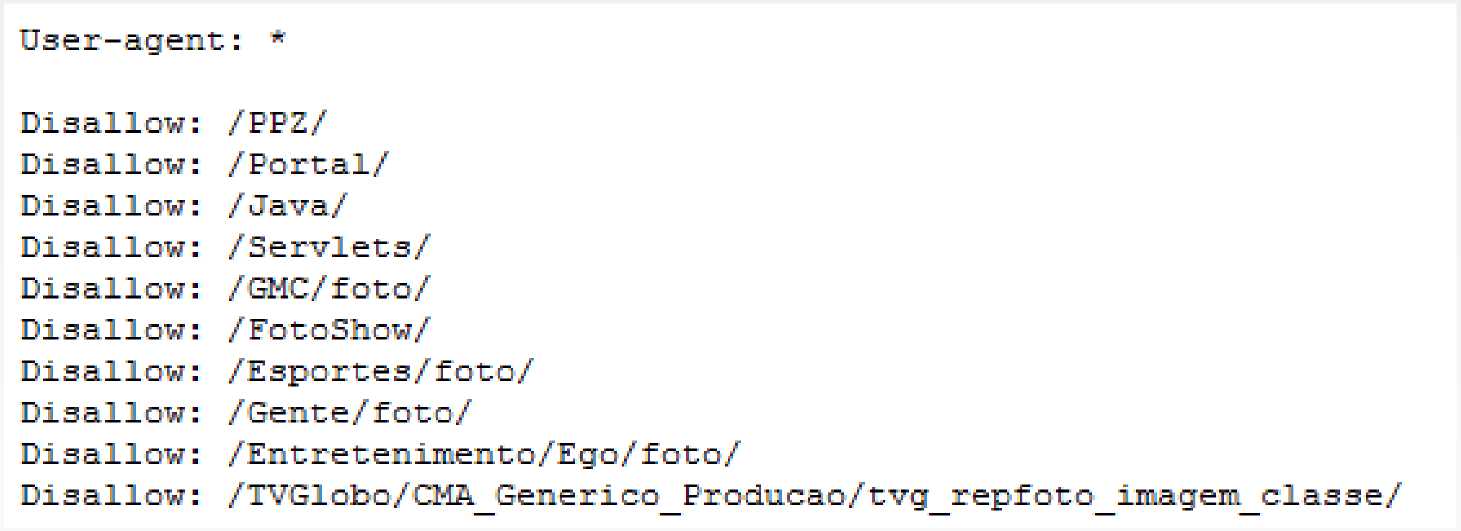
- Possibilidade de bloquear crawlers: **robots.txt**
  - Especifica diretórios e páginas que podem ou não podem ser coletadas com o uso de crawler

```
User-agent: Googlebot
Disallow: /confidencial
Disallow: /protegido
```

```
User-agent: *
Disallow: /temp
```

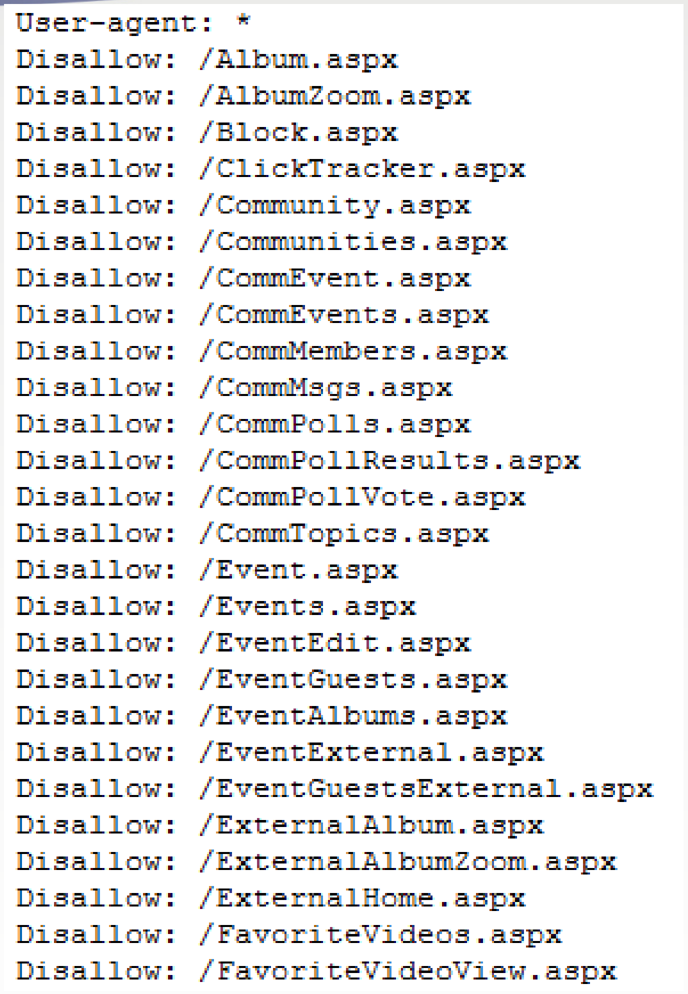
- Mais detalhes
  - http://www.robotstxt.org/wc/robots.html
  - http://pt.wikipedia.org/wiki/Robots.txt

Robots.txt – globo.com



```
User-agent: *

Disallow: /PPZ/
Disallow: /Portal/
Disallow: /Java/
Disallow: /Servlets/
Disallow: /GMC/foto/
Disallow: /FotoShow/
Disallow: /Esportes/foto/
Disallow: /Gente/foto/
Disallow: /Entretenimento/Ego/foto/
Disallow: /TVGlobo/CMA_Generico_Producao/tvg_repfoto_imagem_classe/
```

Robots.txt – orkut



```
User-agent: *
Disallow: /Album.aspx
Disallow: /AlbumZoom.aspx
Disallow: /Block.aspx
Disallow: /ClickTracker.aspx
Disallow: /Community.aspx
Disallow: /Communities.aspx
Disallow: /CommEvent.aspx
Disallow: /CommEvents.aspx
Disallow: /CommMembers.aspx
Disallow: /CommMsgs.aspx
Disallow: /CommPolls.aspx
Disallow: /CommPollResults.aspx
Disallow: /CommPollVote.aspx
Disallow: /CommTopics.aspx
Disallow: /Event.aspx
Disallow: /Events.aspx
Disallow: /EventEdit.aspx
Disallow: /EventGuests.aspx
Disallow: /EventAlbums.aspx
Disallow: /EventExternal.aspx
Disallow: /EventGuestsExternal.aspx
Disallow: /ExternalAlbum.aspx
Disallow: /ExternalAlbumZoom.aspx
Disallow: /ExternalHome.aspx
Disallow: /FavoriteVideos.aspx
Disallow: /FavoriteVideoView.aspx
```

Agregadores de tráfego

- Proxies: reconstrução de transações e sessões
  - YouTube Traffic Characterization: A view from the Edge. **IMC'07**
  - Understanding Online Social Networks Usage from a Network Perspective. **IMC'09**

- Agregadores de redes sociais
  - Characterizing User Behavior in Online Social Networks. **IMC'09**

ACM IMC 2007

YouTube Traffic Characterization: A View From the Edge

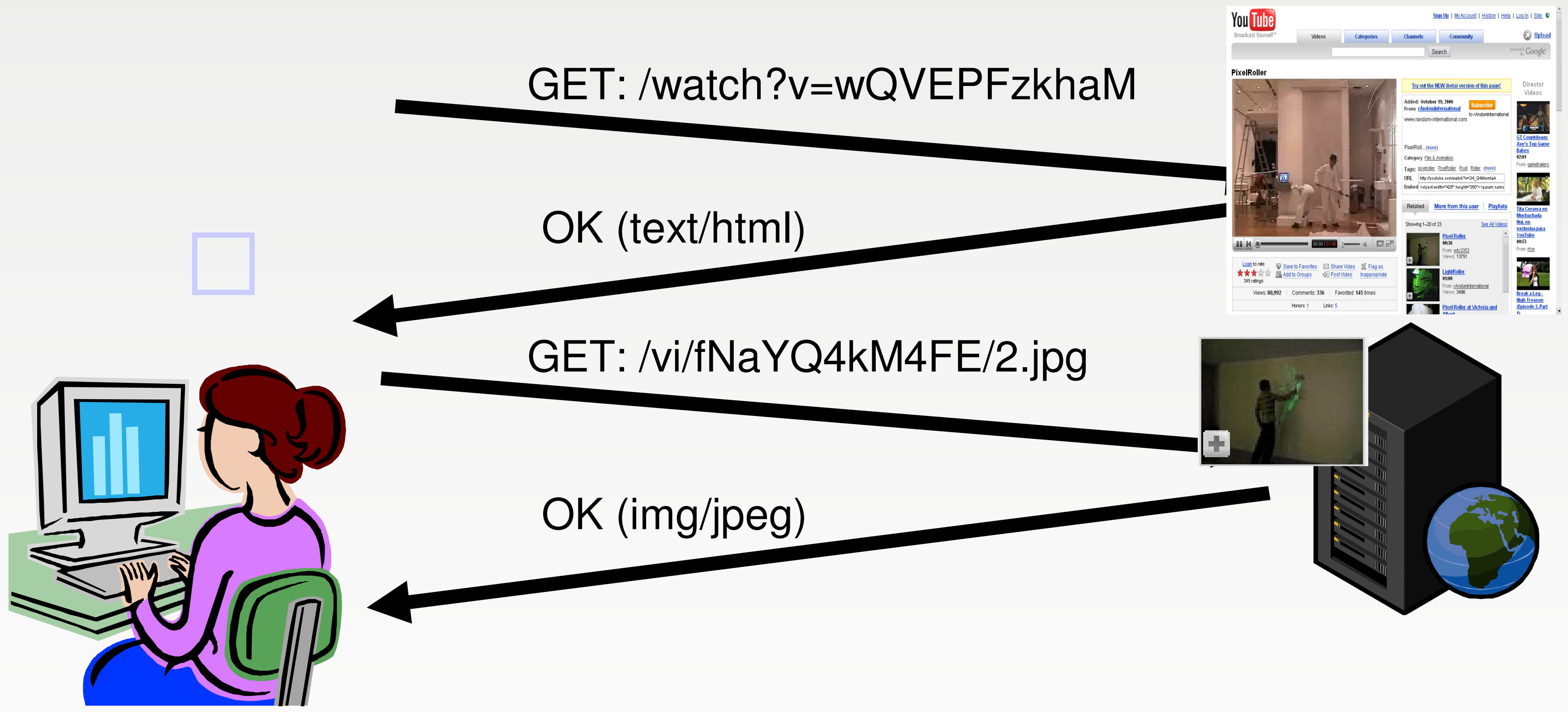
Phillipa Gill[1], Martin Arlitt[2][1],

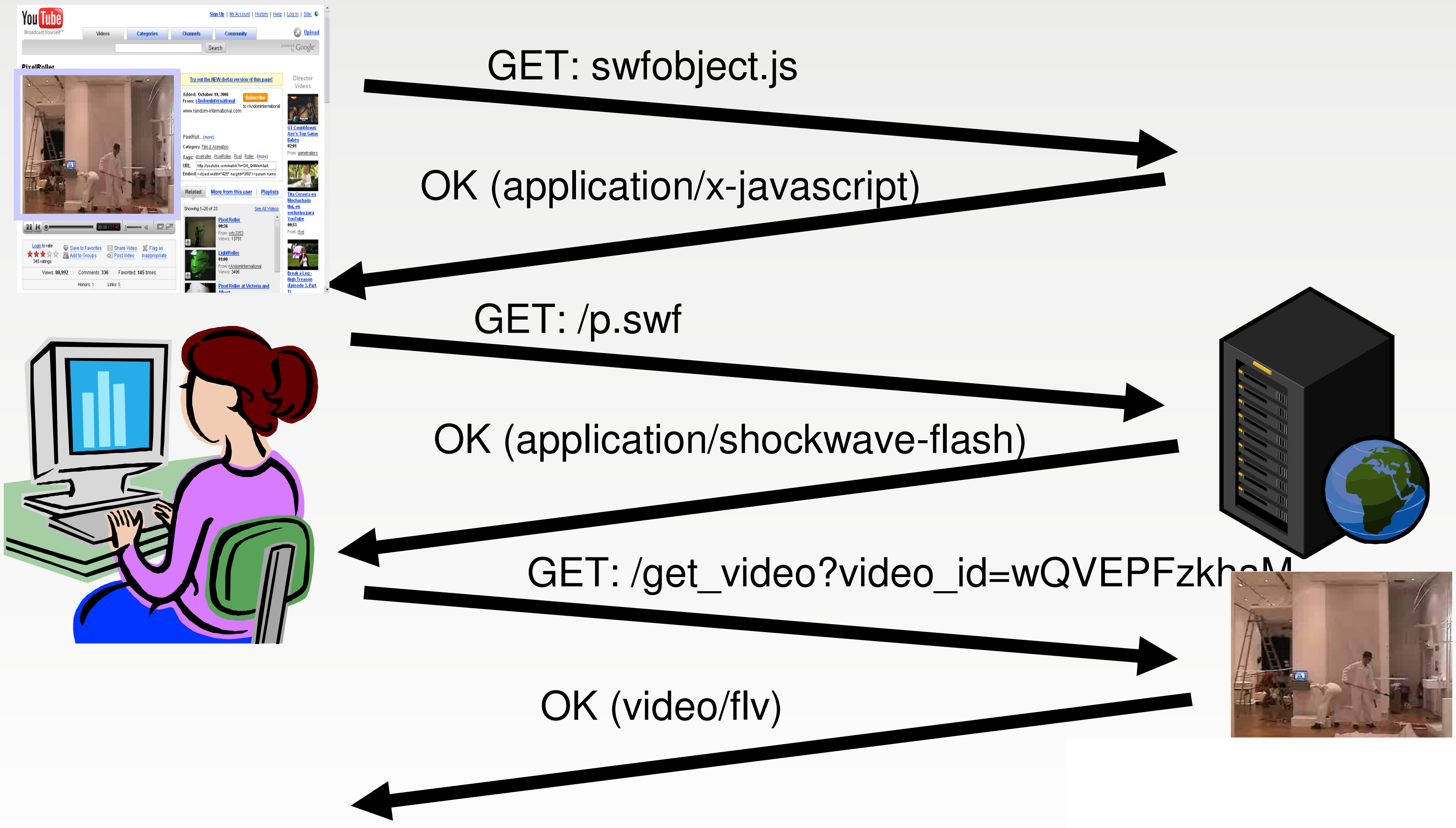
Zongpeng Li[1], Anirban Mahanti[3]

[1]Dept. of Computer Science, University of Calgary, Canada

[2]Enterprise Systems & Software Lab, HP Labs, USA

[3]Dept. of Computer Science and Engineering, IIT Delhi, India


GET: /watch?v=wQVEPFzkhaM
OK (text/html)
GET: /vi/fNaYQ4kM4FE/2.jpg
OK (img/jpeg)


GET: swfobject.js
OK (application/x-javascript)
GET: /p.swf
OK (application/shockwave-flash)
GET: /get_video?video_id=wQVEPFzkhaM
OK (video/flv)

Edge = Campus de uma universidade

Campus

- 28.000 estudantes e 5.300 professores e funcionários
- Link de 300Mb/s full-duplex

Objetivo:

- Coletar o uso do YouTube em todo o campus
- Obter dados de um período extenso
- Proteger a privacidade dos usuários

Desafios:

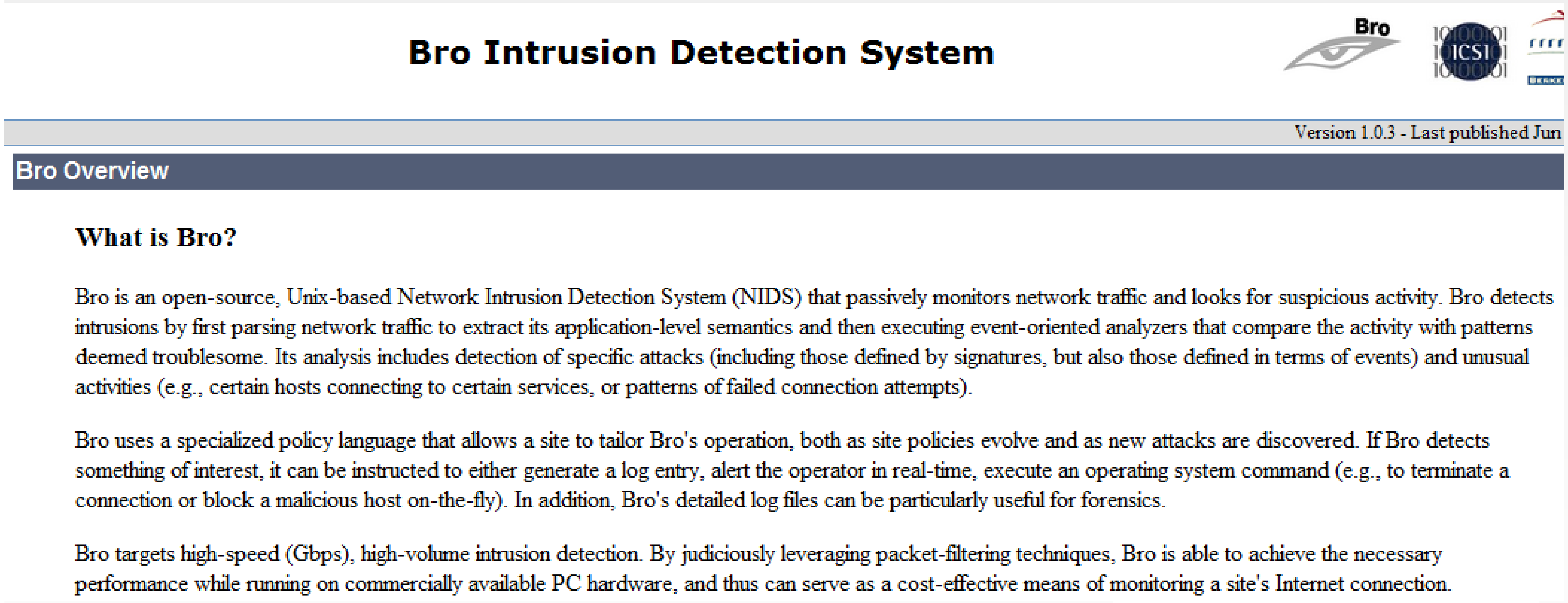
- Popularidade do YouTube
- Limitação dos monitores de tráfego
- Volume do uso da Internet do campus

Metodologia

- Identificar servidores provendo conteúdo do YouTube
- Utilizar **bro** para sumarizar cada transação HTTP em tempo real
- Reiniciar **bro** diariamente e comprimir o log diariamente
- Mapear cada visitante a um ID único

Bro

http://www.bro-ids.org/


Bro Intrusion Detection System

Version 1.0.3 - Last published Jun

Bro Overview

What is Bro?

Bro is an open-source, Unix-based Network Intrusion Detection System (NIDS) that passively monitors network traffic and looks for suspicious activity. Bro detects intrusions by first parsing network traffic to extract its application-level semantics and then executing event-oriented analyzers that compare the activity with patterns deemed troublesome. Its analysis includes detection of specific attacks (including those defined by signatures, but also those defined in terms of events) and unusual activities (e.g., certain hosts connecting to certain services, or patterns of failed connection attempts).

Bro uses a specialized policy language that allows a site to tailor Bro's operation, both as site policies evolve and as new attacks are discovered. If Bro detects something of interest, it can be instructed to either generate a log entry, alert the operator in real-time, execute an operating system command (e.g., to terminate a connection or block a malicious host on-the-fly). In addition, Bro's detailed log files can be particularly useful for forensics.

Bro targets high-speed (Gbps), high-volume intrusion detection. By judiciously leveraging packet-filtering techniques, Bro is able to achieve the necessary performance while running on commercially available PC hardware, and thus can serve as a cost-effective means of monitoring a site's Internet connection.

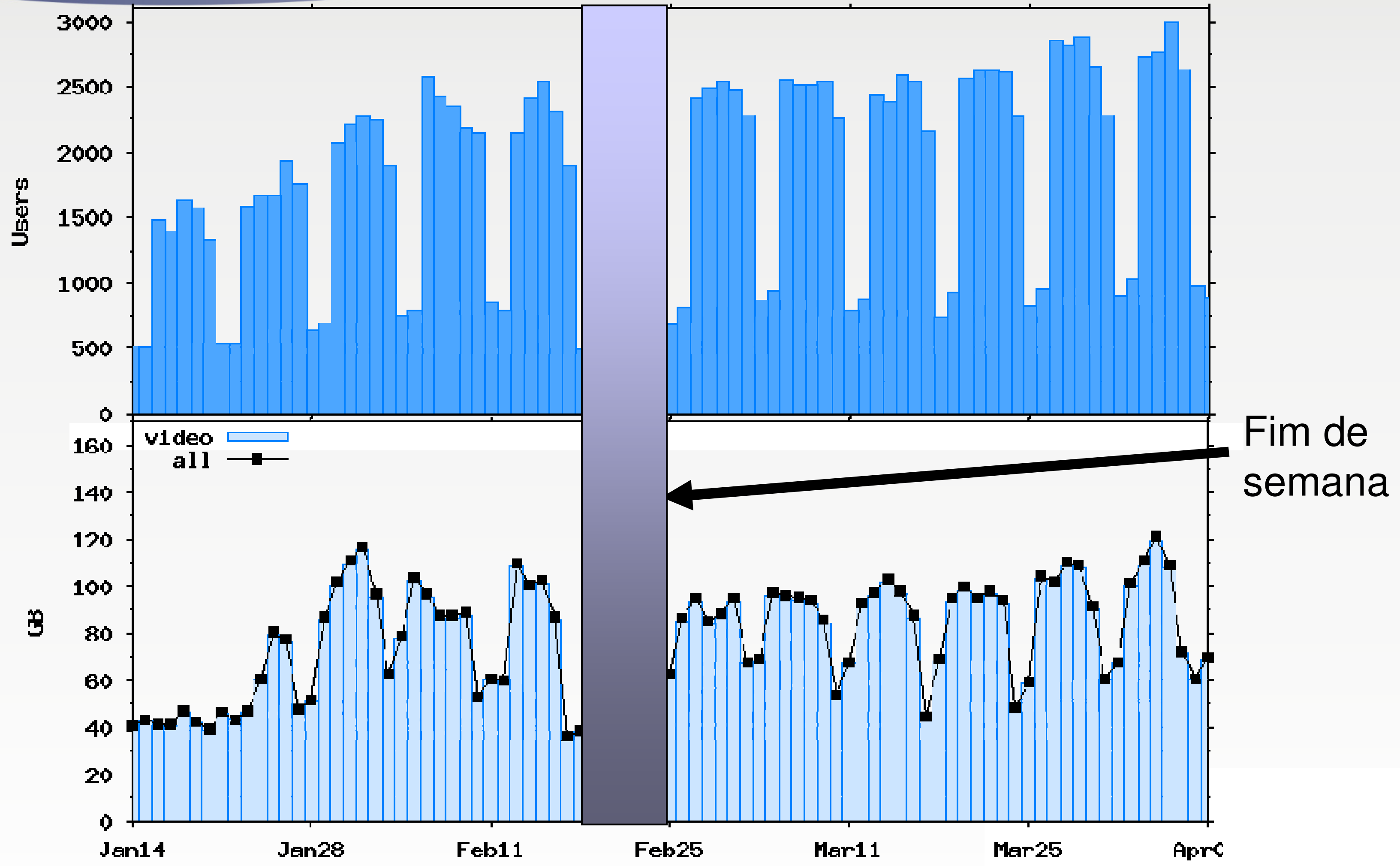
Sumário dos dados

Start Date:	Jan. 14, 2007
End Date:	Apr. 8, 2007
Total Valid Transactions:	23,250,438
Total Bytes:	6.54 TB
Total Video Requests:	625,593
Total Video Bytes:	6.45 TB
Unique Video Requests:	323,677
Unique Video Bytes:	3.26 TB

HTTP Response Codes

Code	% of Responses	% of Bytes
200 (OK)	75.80	89.78
206 (Partial Content)	1.29	10.22
302 (Found)	0.05	0.00
303 (See Other)	5.33	0.00
304 (Not Modified)	17.34	0.00
4xx (Client Error)	0.19	0.00
5xx (Server Error)	0.01	0.00

Campus Usage Patterns



ACM IMC 2009

Characterizing User Behavior in Online Social Networks

Fabrício Benevenuto[1], Tiago Rodrigues[1],
Meeyoung Cha[2], Virgílio Almeida[1]

[1]Universidade Federal de Minas Gerais
[2]Max Planck Institute for Software Systems (MPI-SWS)

O que os usuários fazem nas redes sociais

Post status

Watch videos

Search

Send messages

Upload videos and pictures

Browse list of friends

Join communities

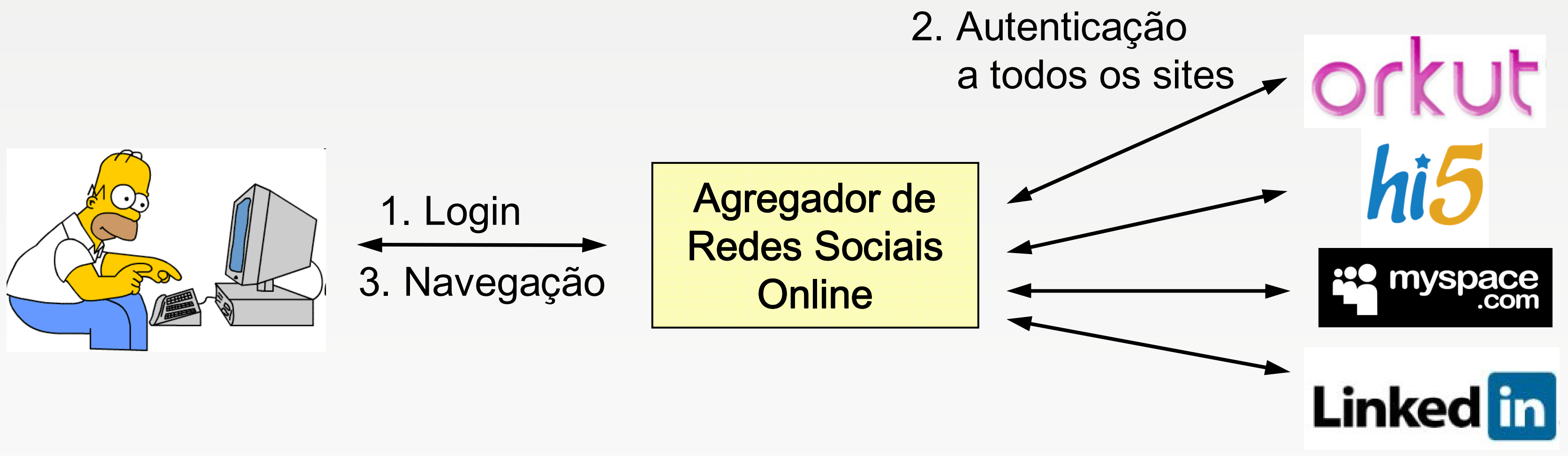
Use applications

Browse profiles and pictures

Entender navegação e interação dos usuários através de todas as atividades

Agregador de tráfego

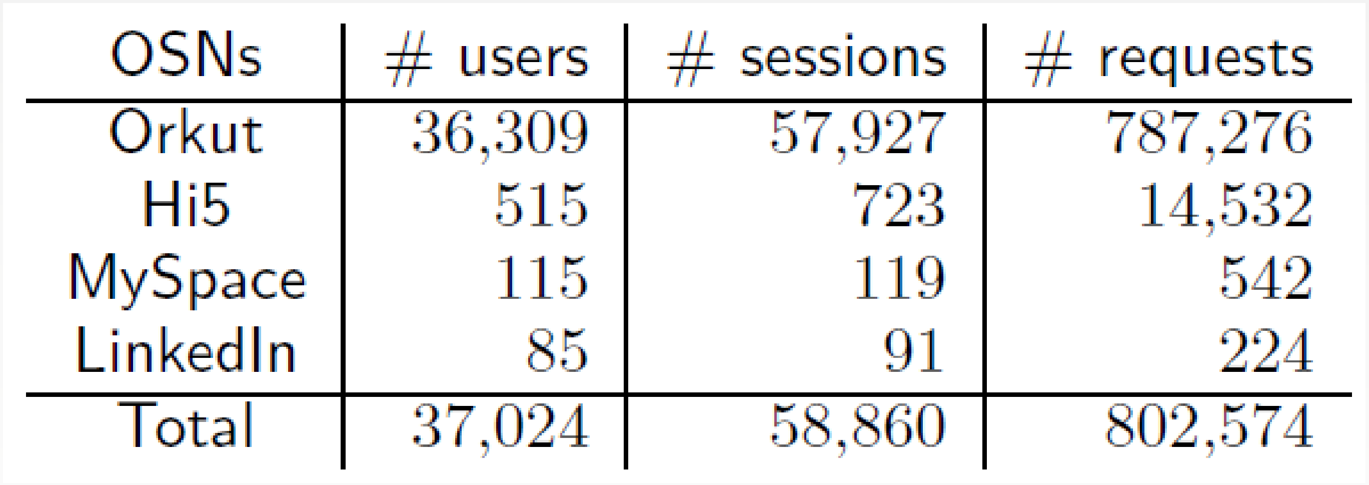
Dados podem ser coletados de um agregador de redes sociais



1. Login
3. Navegação
Agregador de Redes Sociais Online
2. Autenticação a todos os sites
orkut
hi5
myspace.com
LinkedIn

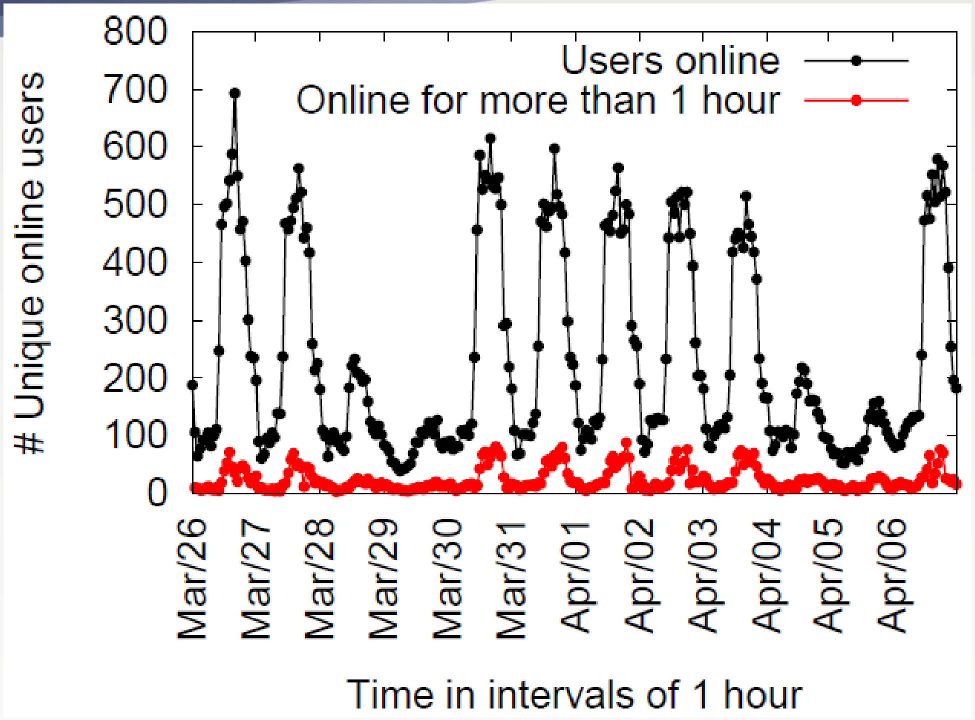
Dados obtidos

- 12 dias (26 de março a 6 de abril de 2009)
- Sumários de sessões HTTP
  - User ID, session ID, URL, timestamp, IP address, traffic bytes



OSNs	# users	# sessions	# requests
Orkut	36,309	57,927	787,276
Hi5	515	723	14,532
MySpace	115	119	542
LinkedIn	85	91	224
Total	37,024	58,860	802,574

Padrões de acesso



- Best fittings para várias medidas
  - inter-session time, inter-request time, session duration

Atividades no Orkut

Profile & Friends
Browse profile, homepage, list of friends, friend updates, members of communities, fans, etc.

Scrapbook
write
browse

Photos
Edit/Organize photos
browse photos, albums, photos, list of albums, comments in photos, photos tagged

Messages
write
browse

Communities
join/leave
post in topics
browse communities, topics, list of communities, etc.

Search

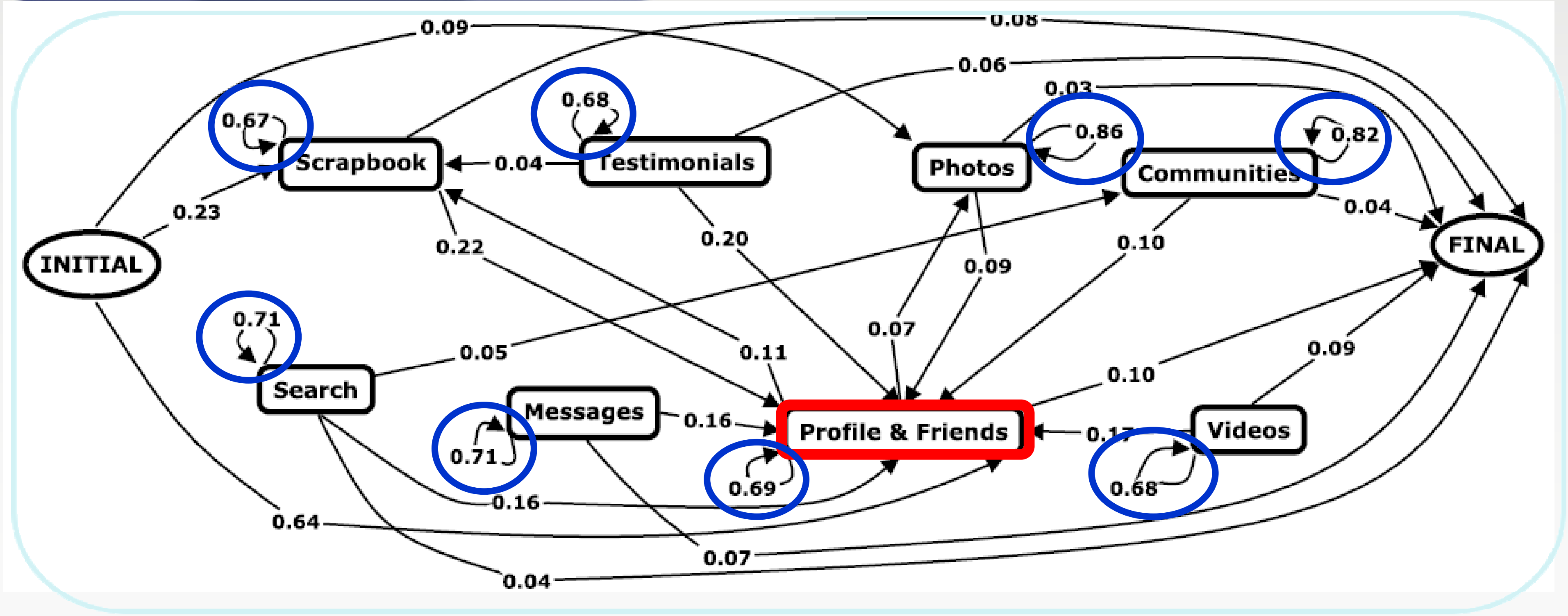
Testimonials
write
browse written and received

Videos
browse list of favorites
watch a video

Others
Applications
user settings

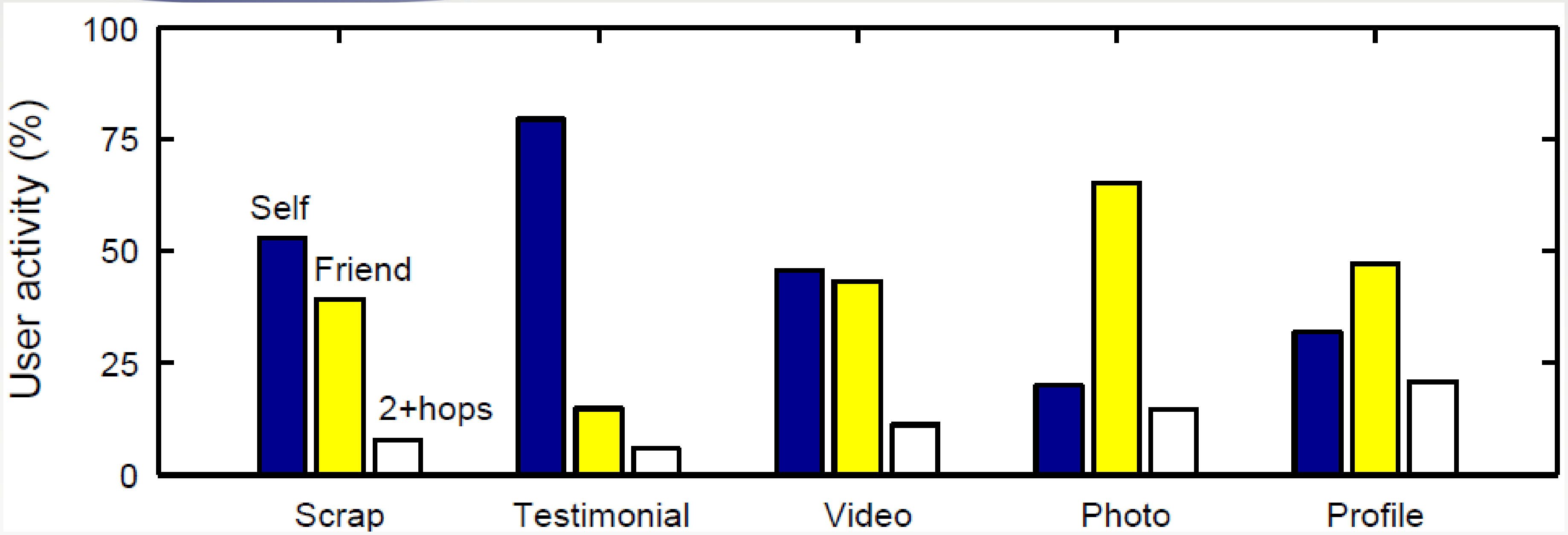
Browsing corresponde a 92% das atividades!

Seqüência das atividades



- Profile & Friends são centrais
- Self-loops são dominantes em todas as categorias

Interações no Orkut



- Usuários acessam mais as páginas de seus amigos
- Interação com desconhecidos é alta

Interações no Orkut

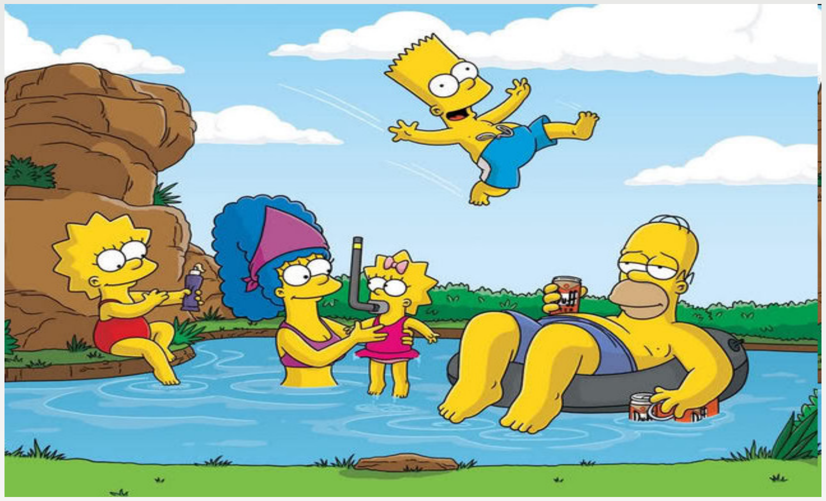
1) Marge faz upload de uma foto

2) Homer recebe a atualização

3) Um amigo de Marge comenta a foto

4) O comentário também aparece para Homer

5) E Homer fica curioso para saber quem é esse cara que comentou na foto de sua esposa!

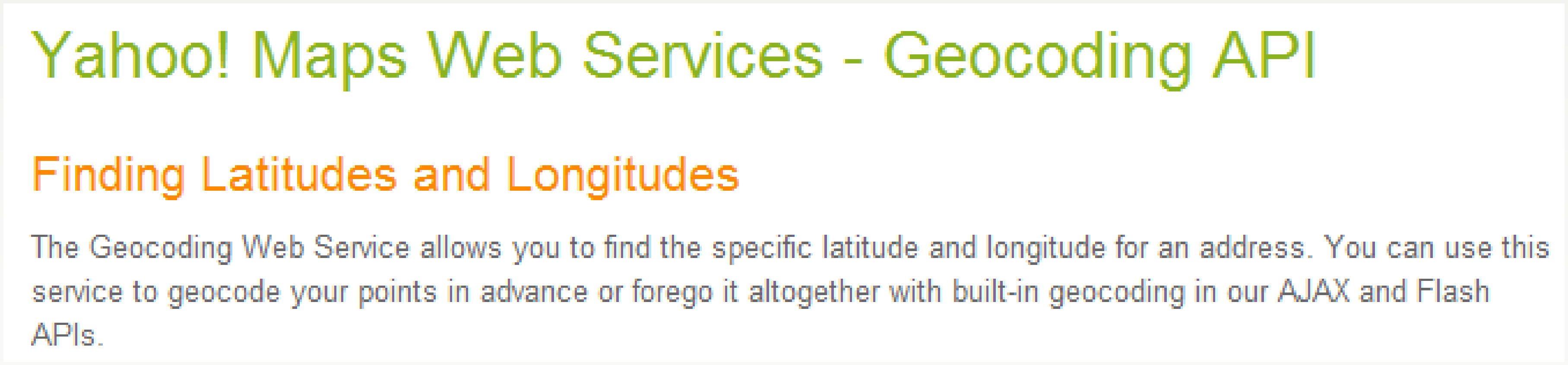


Nice picture, Marge.

- Descoberta de conteúdo através de elos sociais
  - Acessos vêm da homepage e do scrapbook

Obtendo informações geográficas

- Informações geográficas são muitas vezes texto livre
  - Usuários podem preencher qualquer coisa. Ex. Sampa, BH, Marte
  - http://developer.yahoo.com/maps/rest/V1/geocode.html

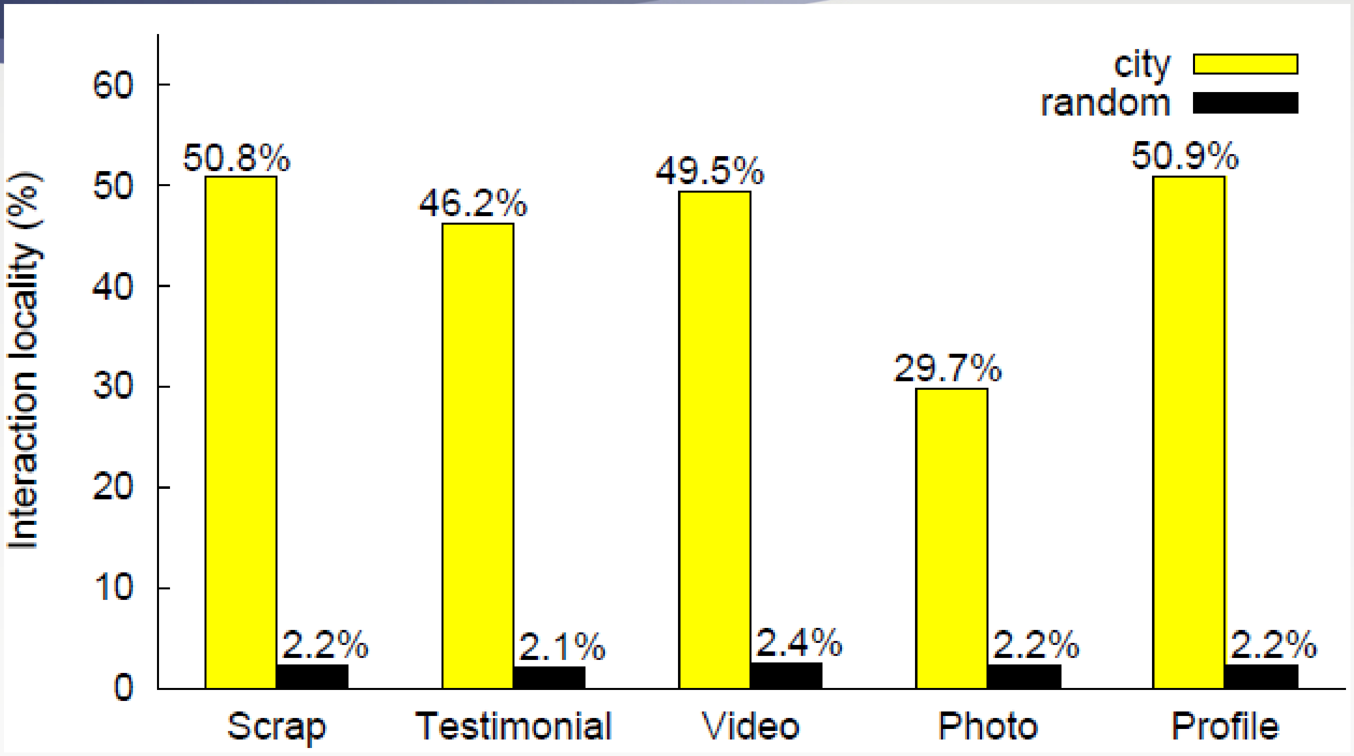


Yahoo! Maps Web Services - Geocoding API

Finding Latitudes and Longitudes

The Geocoding Web Service allows you to find the specific latitude and longitude for an address. You can use this service to geocode your points in advance or forego it altogether with built-in geocoding in our AJAX and Flash APIs.

Interações através da distância física



Conteúdo produzido e consumido localmente

ACM IMC 2009

Understanding Online Social Network Usage from a Network Perspective

Fabian Schneider[1], Anja Feldmann[1],

Balachander Krishnamurthy[1], Walter Willinger[2]

[1]Technische Universtit¨at Berlin / Deutsche Telekom Laboratories

[2]AT&T Labs–Research

General Approach

1. Reconstruct OSN clickstreams from anonymized packet-level traces
    - Anonymized HTTP header traces from two large ISPs
    - Used Bro[1] to extract HTTP request-response pairs (rr-pairs)

2. Map rr-pairs into sessions
    - Sessions identified via SessionIDs (from HTTP `Cookie` header)
    - Track logins and logouts $\Rightarrow$ Authenticated or offline state
    - Cookies help if login or logout not observed

3. Classify rr-pairs
    - Active (rr-pair resulting from user action) or
      Indirect (e.g. followup/embedded via HTTP `Referer` chain)
    - Determine user actions, group into 13 categories

OSN Selection criteria:

- OSNs focussing on profiles (e. g., no YouTube, . . .)
- 2 globally popular
- 2 locally popular (well represented at one ISP)

[✓ facebook]

studiVZ
DAS STUDIVERZEICHNIS

LinkedIn

hi5

Category Examples

Home
All actions on the homepage once authenticated

Photos
Uploading, tagging, and managing photos

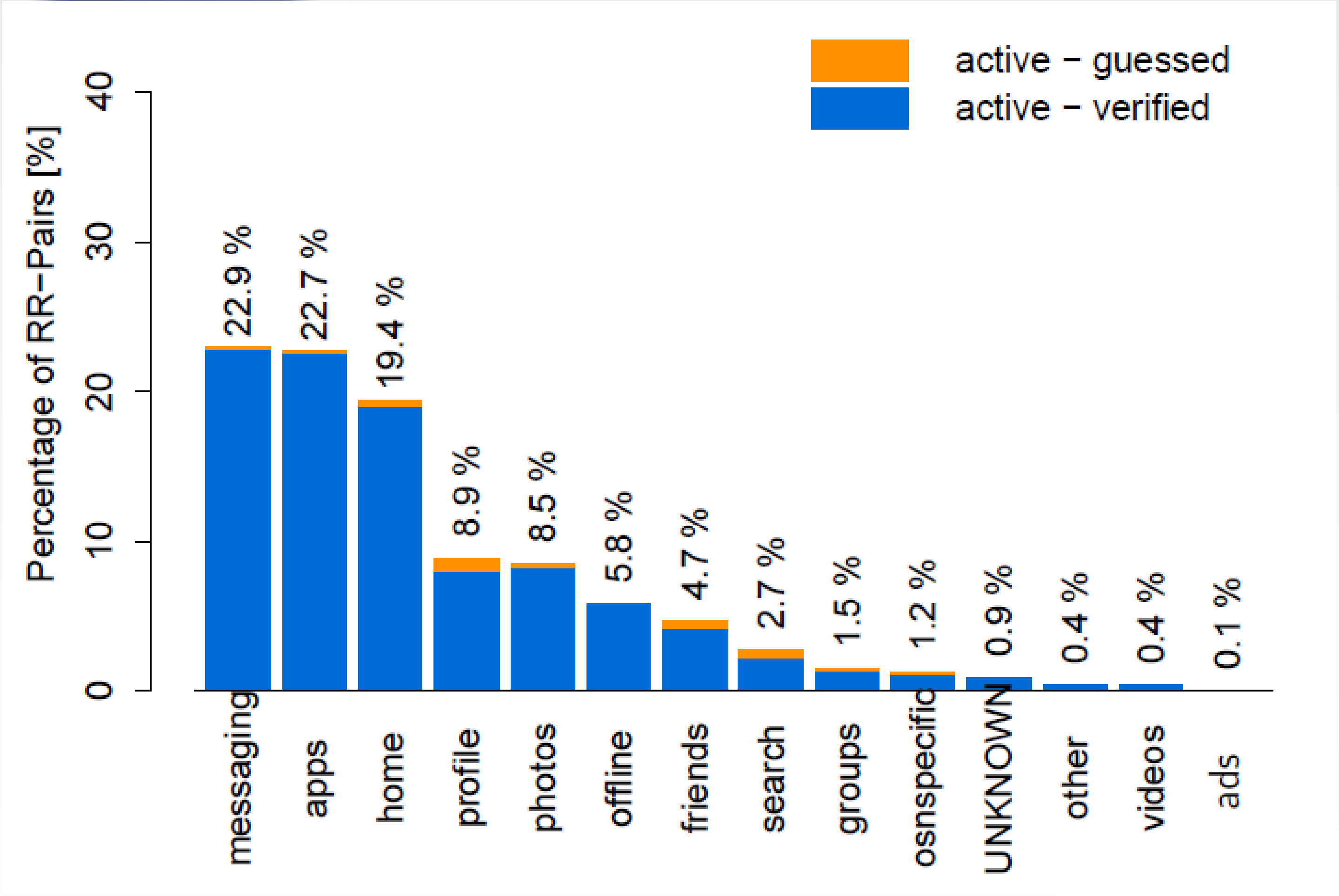
Profile
Accessing and changing profiles, posting on walls, privacy settings

Friends
Browsing, inviting, and accepting friends

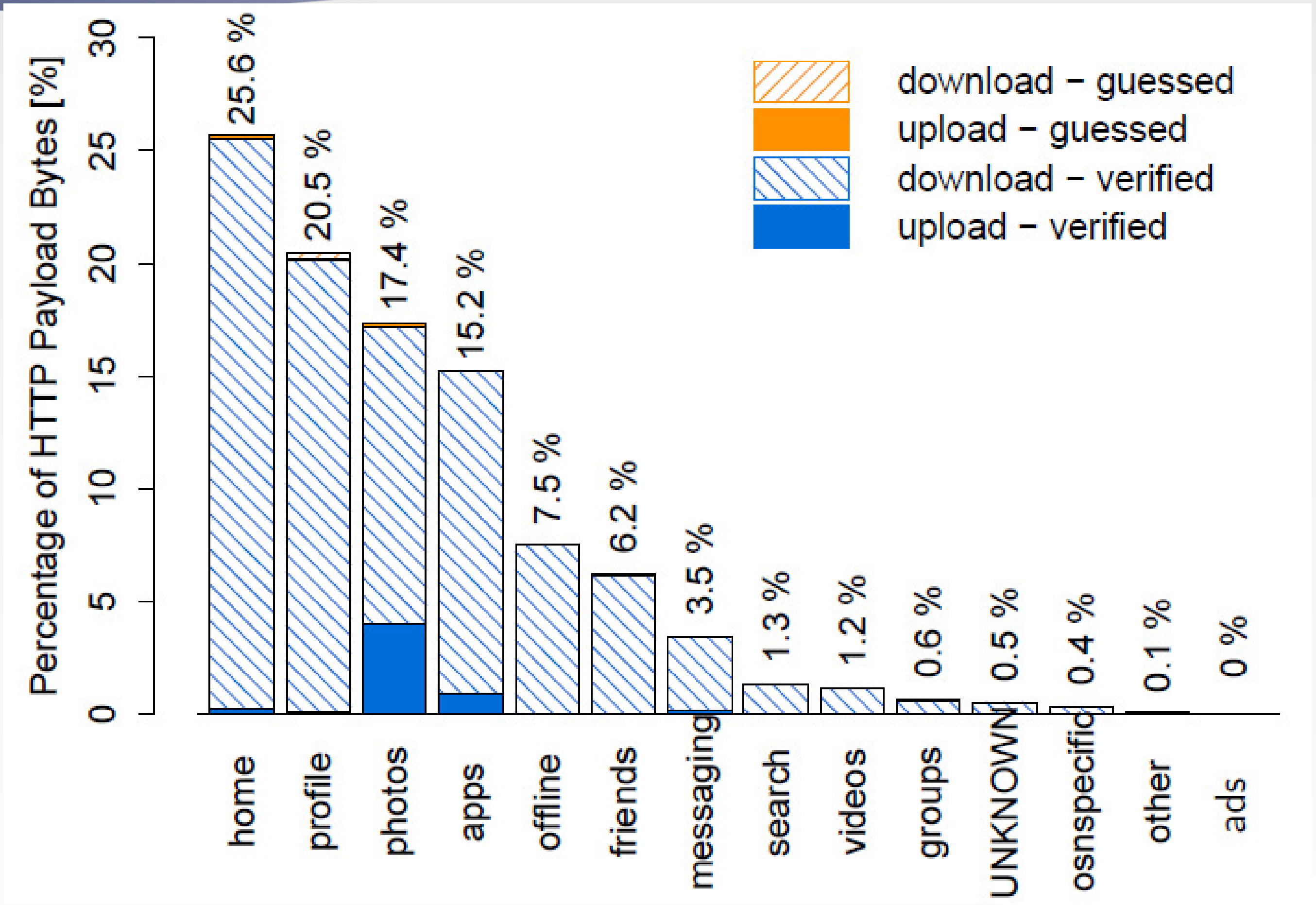
Apps
Applications (external and internal), **only** rr-pairs directed towards OSN servers

Offline
All actions while unauthenticated, e. g., public profile browsing, registering

Popularidade das atividades

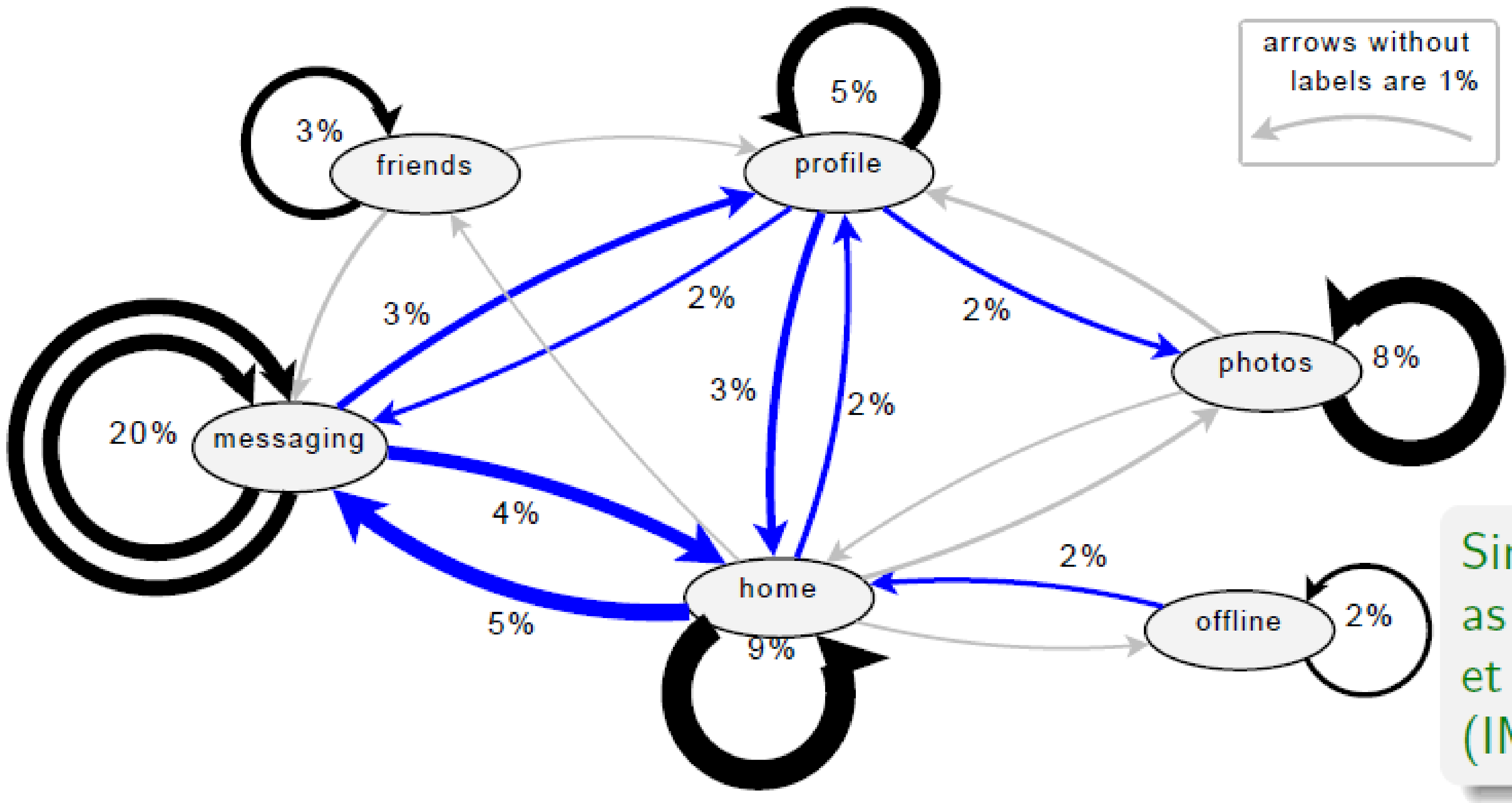


Volume por categoria



Seqüência de atividades

Click sequences of Facebook for ISP-A2: Global transition probabilities



Similar findings as Benevenuto et al for Orkut (IMC'09)

Findings

⇒ Messaging traps users; Home, Photos and Profile attract users to stay

Aplicações e jogos online

- Funcionamento e construção de aplicações em redes sociais
  - Unveiling Facebook: A measurement study of social network based applications. IMC’08
- Jogos Online
  - Social influence and the diffusion of user-created content. EC’09.

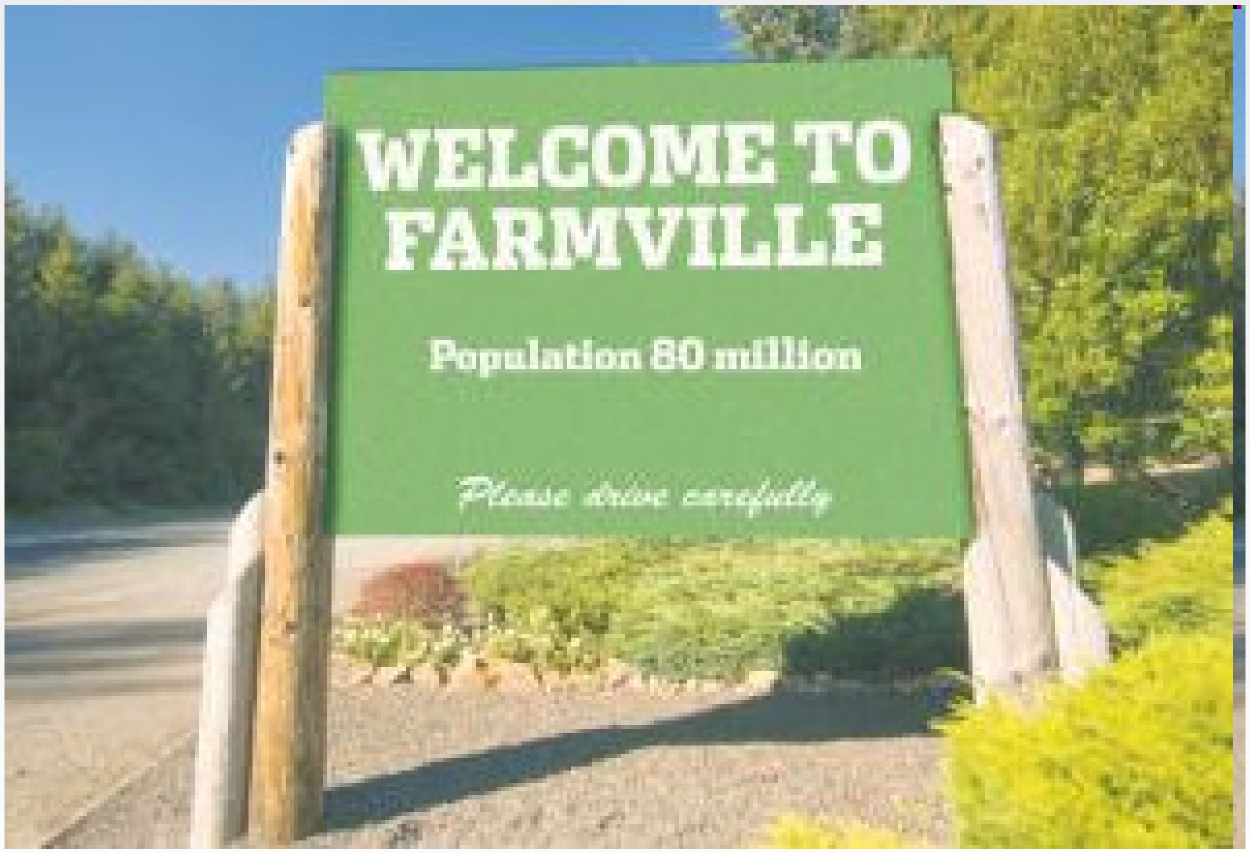
Aplicações

- Dominante em vários sistemas
  - Facebook, Orkut, Hi5, MySpace
- Duas plataformas maiores
  - Facebook Developer Platform (FDP)
  - OpenSocial

Facebook - aplicações

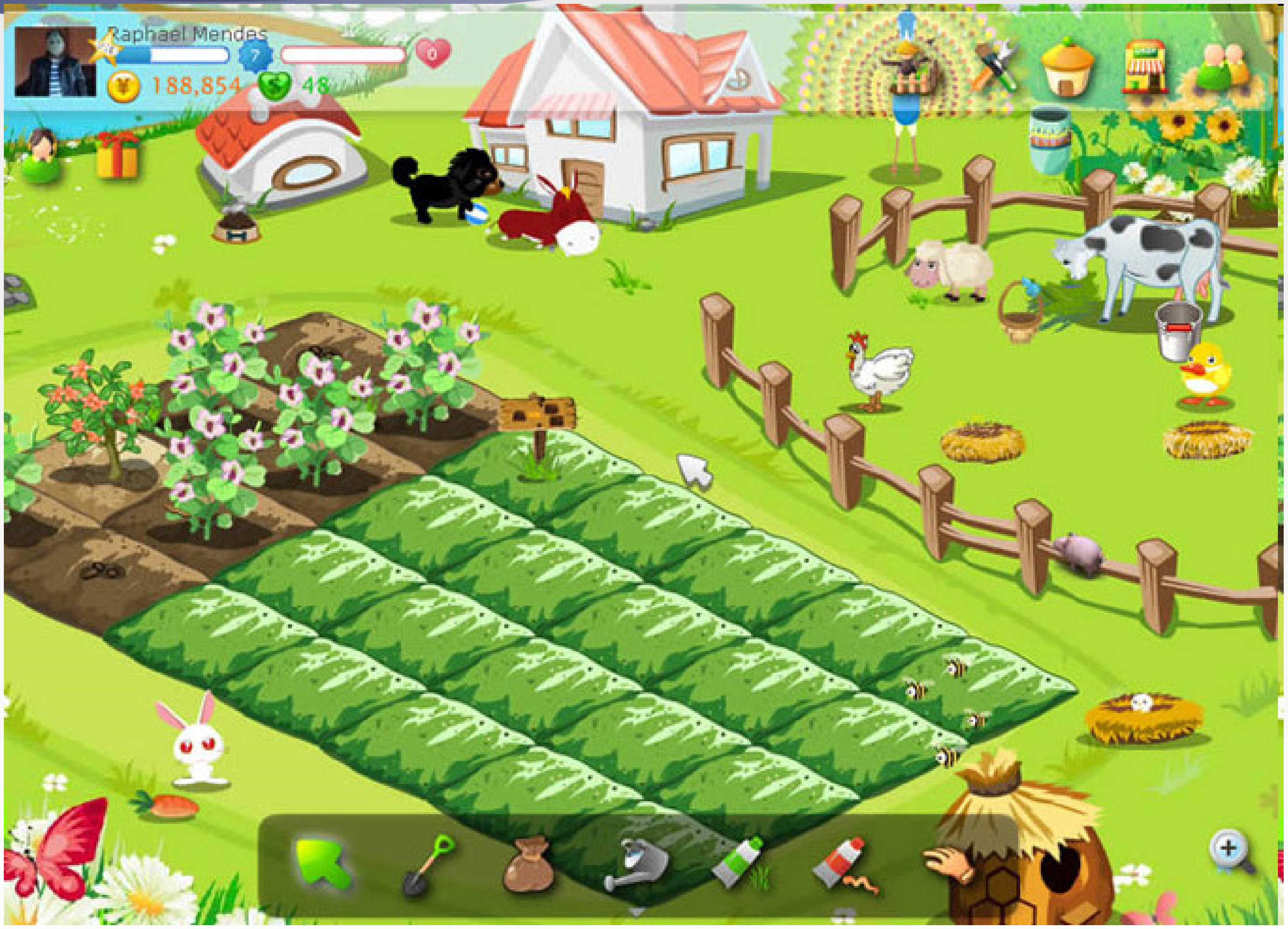
- Mais de 1 milhão de desenvolvedores em 180 países
- Mais de 550 mil aplicações ativas
- Mais de 100 milhões de usuários utilizando aplicações

Facebook - aplicações

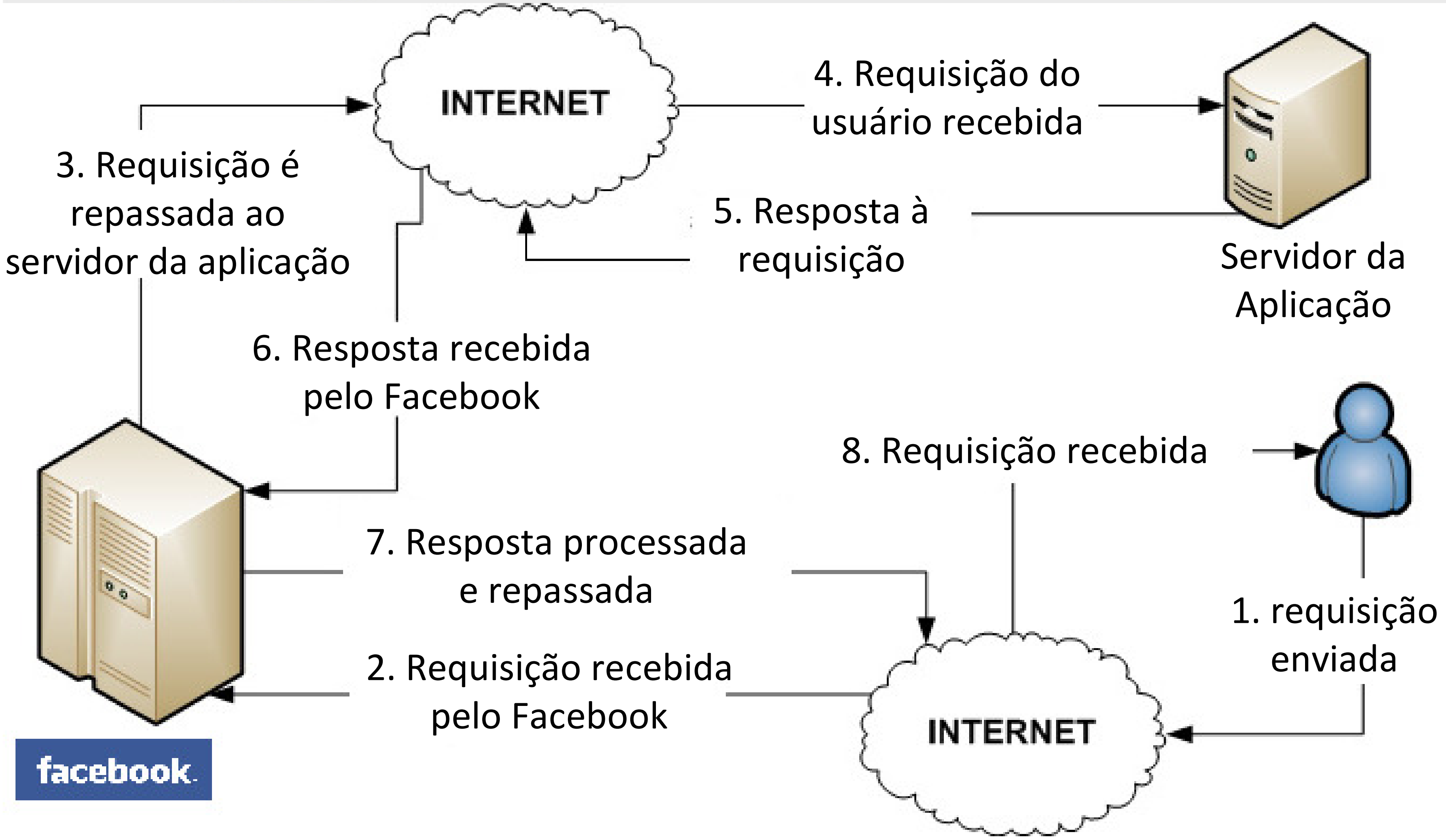


Facebook - Aplicações

Orkut - Aplicações



Aplicações

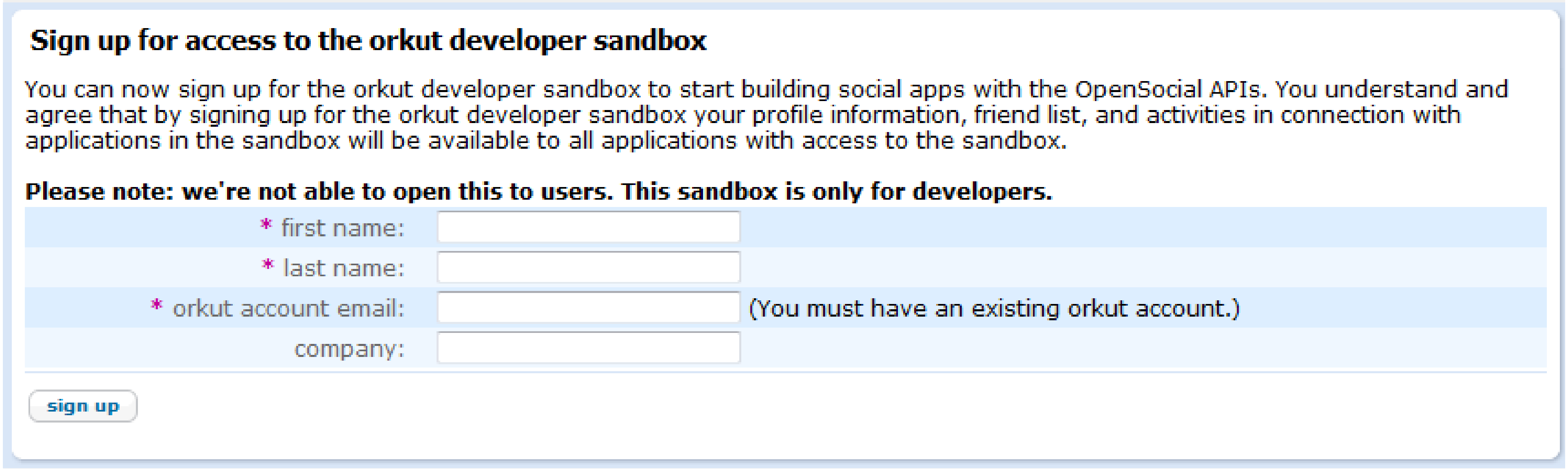

INTERNET
4. Requisição do usuário recebida
3. Requisição é repassada ao servidor da aplicação
5. Resposta à requisição
Servidor da Aplicação
6. Resposta recebida pelo Facebook
8. Requisição recebida
7. Resposta processada e repassada
1. requisição enviada
2. Requisição recebida pelo Facebook
INTERNET
facebook

Como criar uma aplicação no Orkut?

- Crie uma conta no sandbox.orkut
- Determine um local onde sua aplicação vai ficar
  - Página pessoal, repositórios, etc.
- Entendimento da API do Orkut
- Crie uma aplicação que seja legal
- Se o Orkut aprovar, a aplicação se torna pública.

Como criar uma aplicação no Orkut?

- http://sandbox.orkut.com/SandboxSignup.aspx



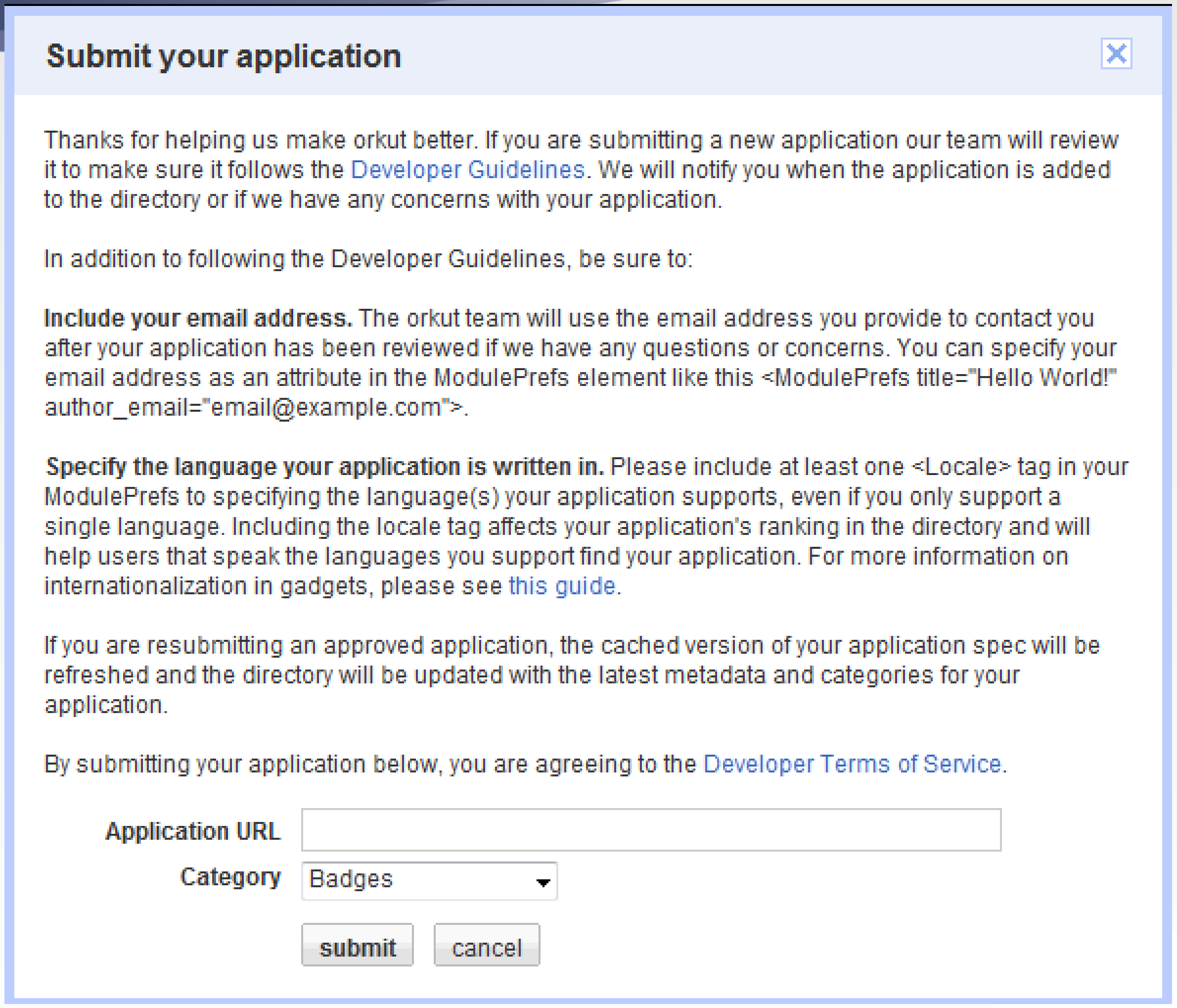
Sign up for access to the orkut developer sandbox

You can now sign up for the orkut developer sandbox to start building social apps with the OpenSocial APIs. You understand and agree that by signing up for the orkut developer sandbox your profile information, friend list, and activities in connection with applications in the sandbox will be available to all applications with access to the sandbox.

Please note: we're not able to open this to users. This sandbox is only for developers.

* first name:		
* last name:		
* orkut account email:		(You must have an existing orkut account.)
company:		

Como criar uma aplicação no Orkut?



Submit your application

Thanks for helping us make orkut better. If you are submitting a new application our team will review it to make sure it follows the Developer Guidelines. We will notify you when the application is added to the directory or if we have any concerns with your application.

In addition to following the Developer Guidelines, be sure to:

Include your email address. The orkut team will use the email address you provide to contact you after your application has been reviewed if we have any questions or concerns. You can specify your email address as an attribute in the ModulePrefs element like this <ModulePrefs title="Hello World!" author_email="email@example.com">.

Specify the language your application is written in. Please include at least one <Locale> tag in your ModulePrefs to specifying the language(s) your application supports, even if you only support a single language. Including the locale tag affects your application's ranking in the directory and will help users that speak the languages you support find your application. For more information on internationalization in gadgets, please see this guide.

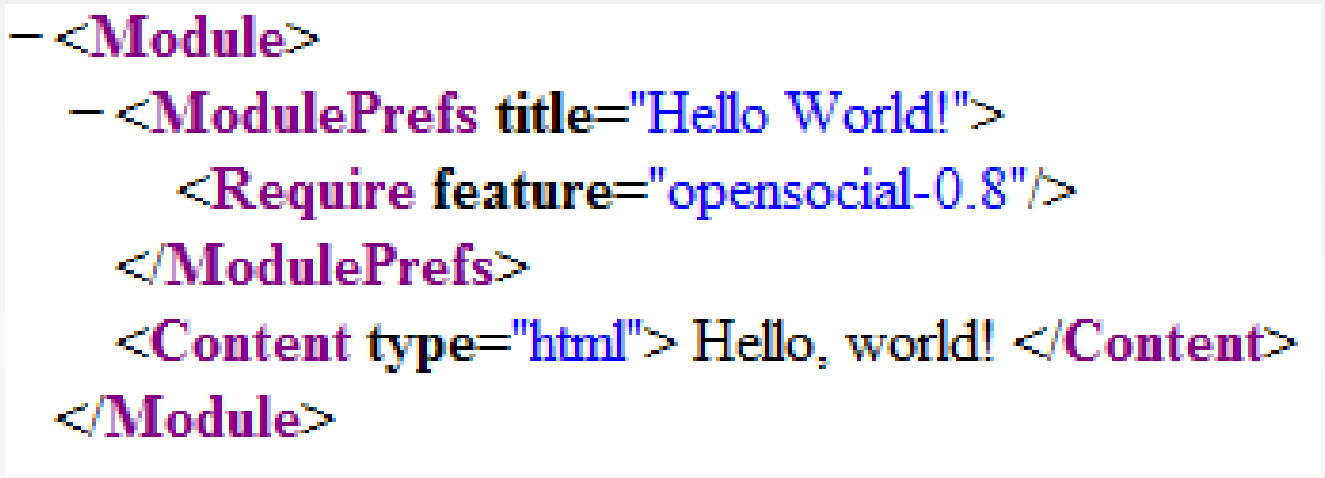
If you are resubmitting an approved application, the cached version of your application spec will be refreshed and the directory will be updated with the latest metadata and categories for your application.

By submitting your application below, you are agreeing to the Developer Terms of Service.

Application URL	
Category	Badges

Como criar uma aplicação no Orkut?

http://homepages.dcc.ufmg.br/~fabricio/hello.xml



```
- <Module>
  - <ModulePrefs title="Hello World!">
      <Require feature="opensocial-0.8"/>
    </ModulePrefs>
    <Content type="html"> Hello, world! </Content>
  </Module>
```

- Mais informações:
  - http://code.google.com/apis/orkut/articles/tutorial/tutorial.html#gadget-basics

ACM SIGCOMM IMC 2008

Unveiling Facebook: A measurement study of social network based Applications

A. Nazir, S. Raza, C. Chuah

University of California, Davis

Our Applications

- We deployed three applications on Facebook:

Fighters' Club (FC, 3.4M+, Jun 2007)

Social Gaming

Got Love? (GL, 4M+, Nov 2007)

Hugged (0.7M+, Feb 2008)

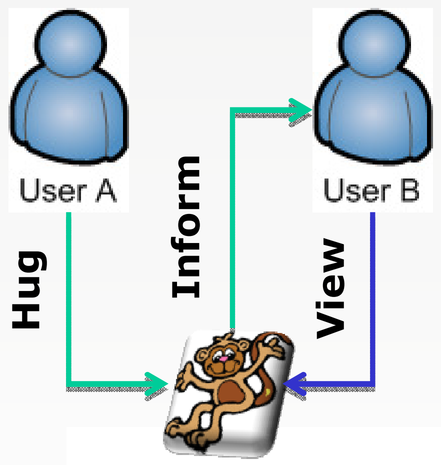
Social Utility

GL, HUGGED: SOCIAL UTILITY APPLICATIONS

- *GL*: friend-friend, one request per target friend

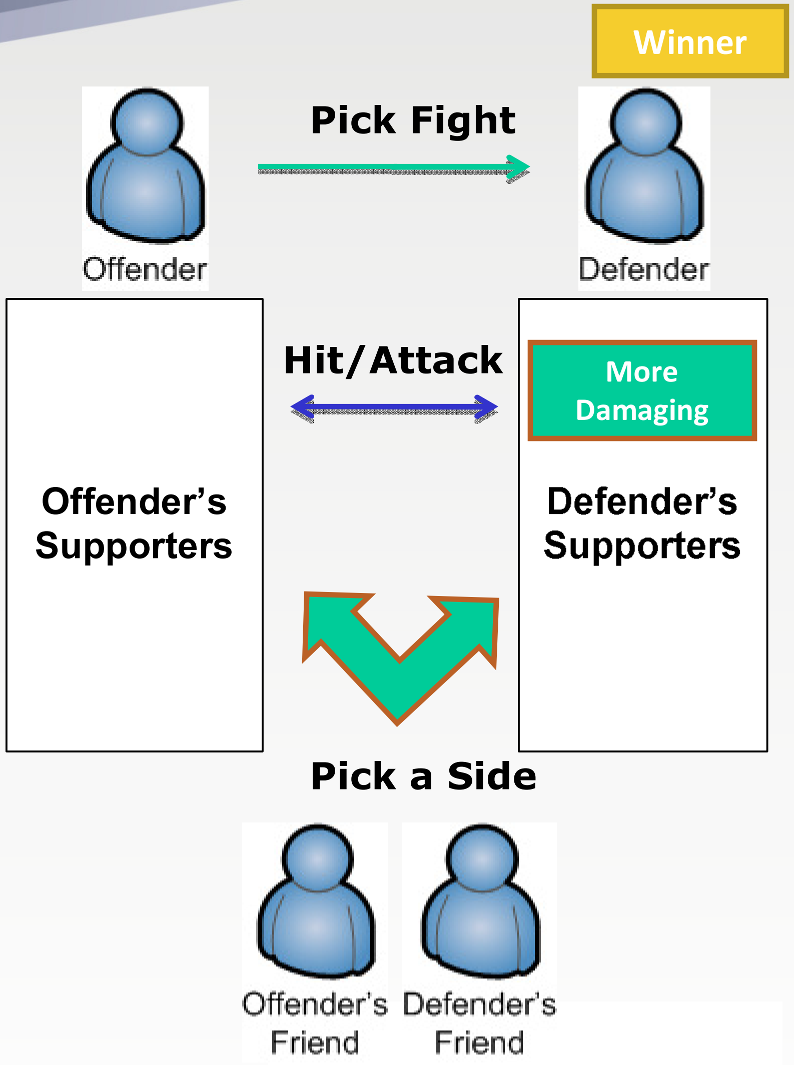
- *Hugged*: friend-friend, multiple requests per target friend

- Similar functionality:
  - User A hugs/loves (friend) User B
  - User B accepts/ignores hug/love



FIGHTERS' CLUB: A GAMING APPLICATION

- Friend-friend, non-friend to non-friend interaction
- Number of blows limited through points system



DATA SET SUMMARY

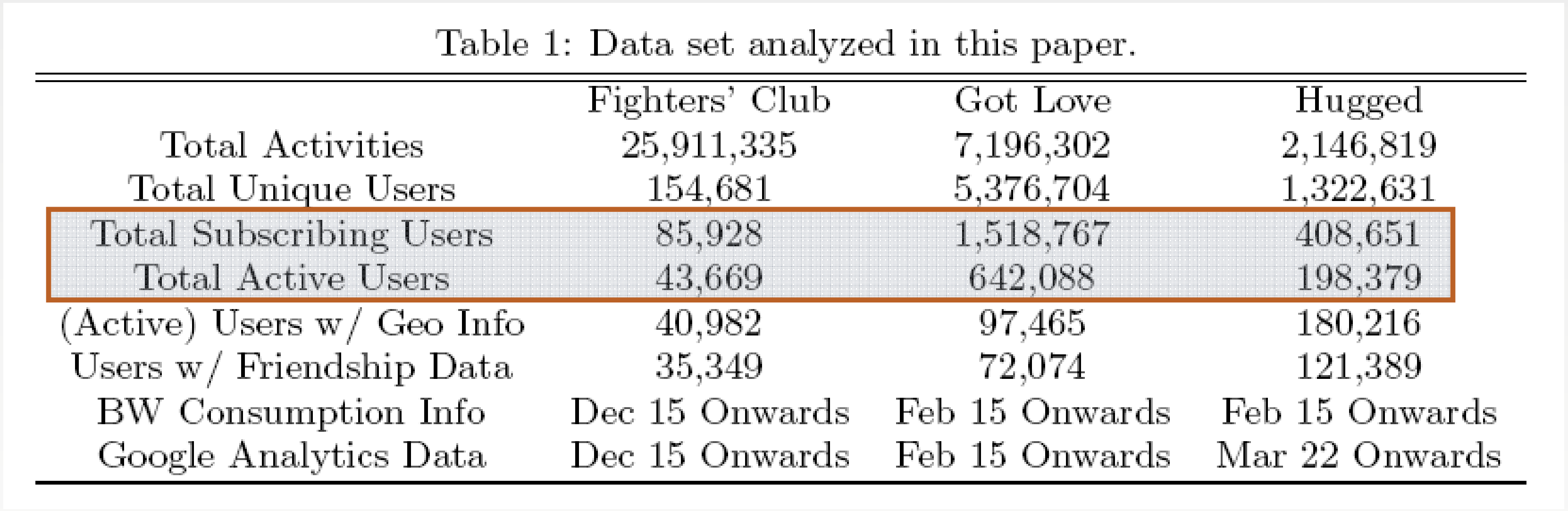
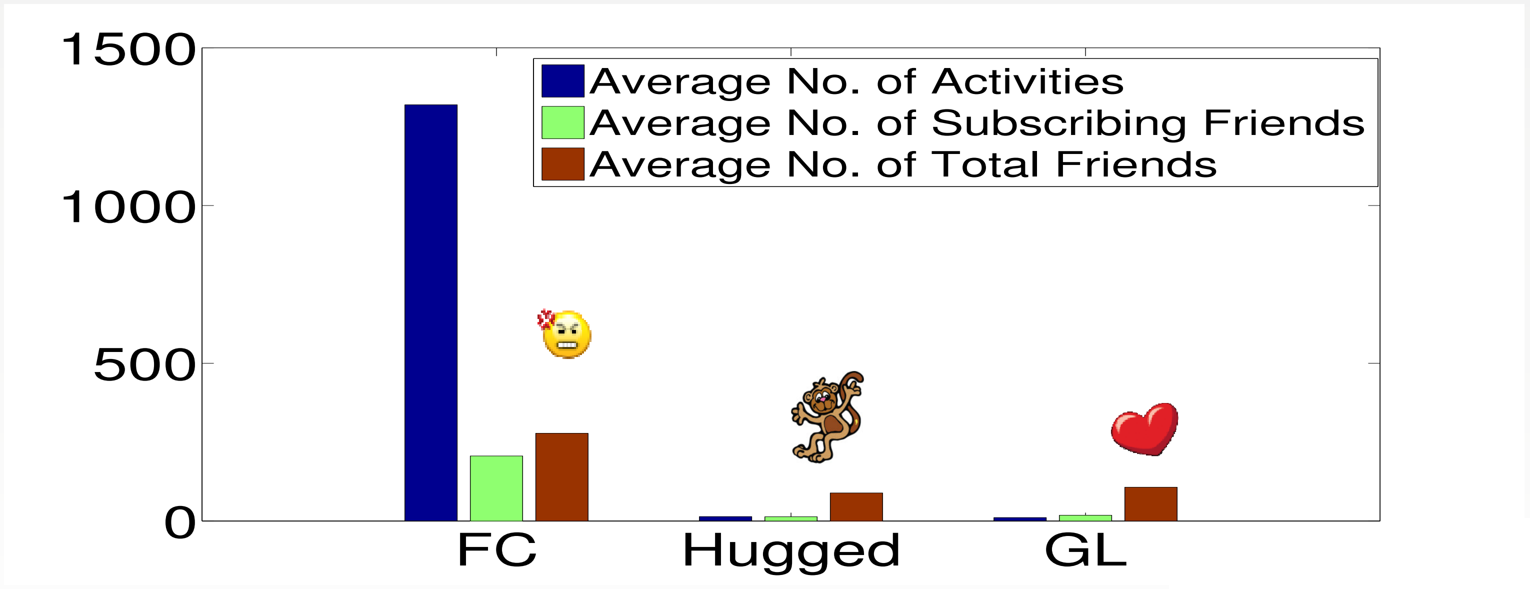


Table 1: Data set analyzed in this paper.

	Fighters' Club	Got Love	Hugged
Total Activities	25,911,335	7,196,302	2,146,819
Total Unique Users	154,681	5,376,704	1,322,631
Total Subscribing Users	85,928	1,518,767	408,651
Total Active Users	43,669	642,088	198,379
(Active) Users w/ Geo Info	40,982	97,465	180,216
Users w/ Friendship Data	35,349	72,074	121,389
BW Consumption Info	Dec 15 Onwards	Feb 15 Onwards	Feb 15 Onwards
Google Analytics Data	Dec 15 Onwards	Feb 15 Onwards	Mar 22 Onwards

SOCIAL GAMING VS. SOCIAL UTILITY APPLICATIONS: RESULTS

- Other differences:
  - Average number of activities higher on *FC* than on *GL*, *Hugged*
  - Average number of friends on application, total number of friends on Facebook, significantly higher for *FC* than GL, *Hugged*



INTERACTION GRAPHS: DATA AND RESULTS SUMMARY

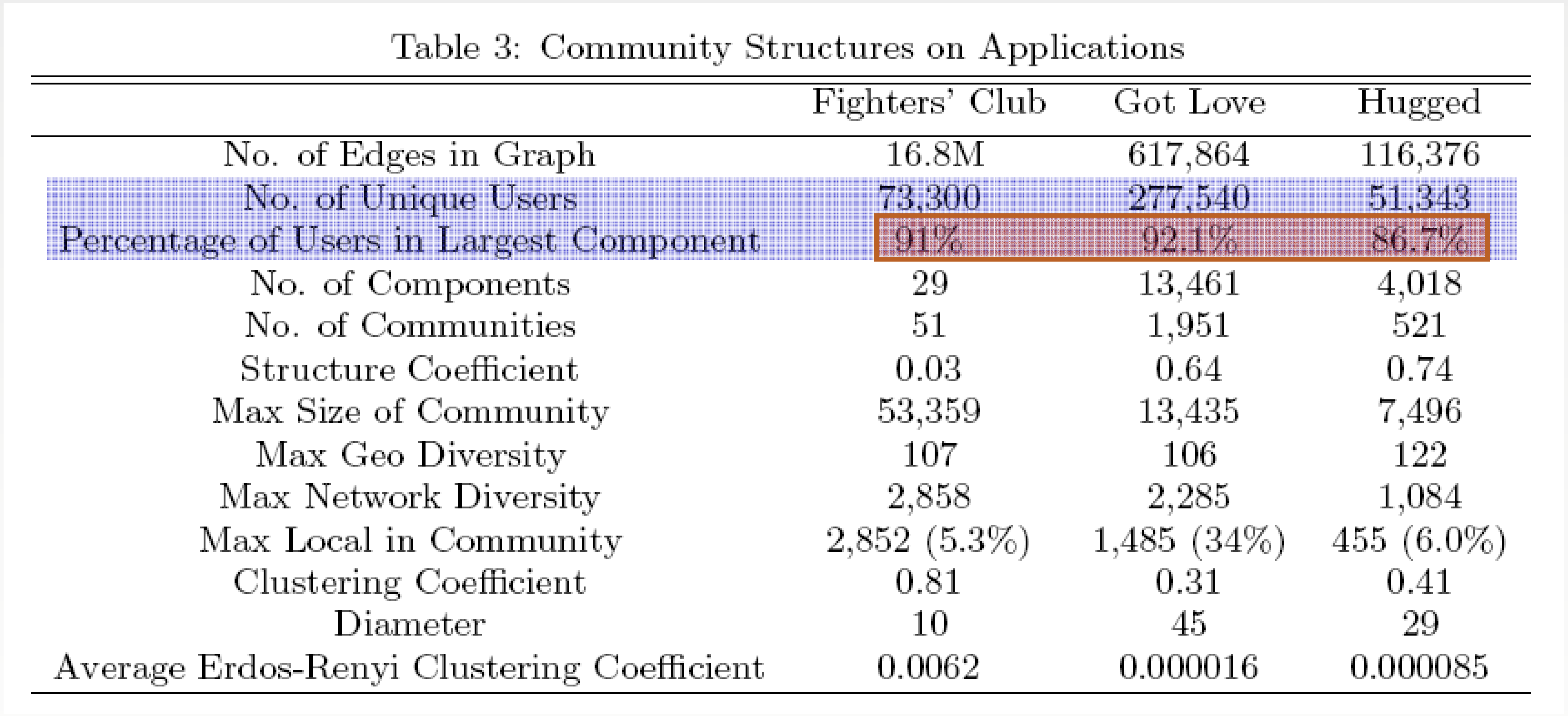


Table 3: Community Structures on Applications

	Fighters' Club	Got Love	Hugged
No. of Edges in Graph	16.8M	617,864	116,376
No. of Unique Users	73,300	277,540	51,343
Percentage of Users in Largest Component	91%	92.1%	86.7%
No. of Components	29	13,461	4,018
No. of Communities	51	1,951	521
Structure Coefficient	0.03	0.64	0.74
Max Size of Community	53,359	13,435	7,496
Max Geo Diversity	107	106	122
Max Network Diversity	2,858	2,285	1,084
Max Local in Community	2,852 (5.3%)	1,485 (34%)	455 (6.0%)
Clustering Coefficient	0.81	0.31	0.41
Diameter	10	45	29
Average Erdos-Renyi Clustering Coefficient	0.0062	0.000016	0.000085

INTERACTION GRAPHS: DATA AND RESULTS SUMMARY

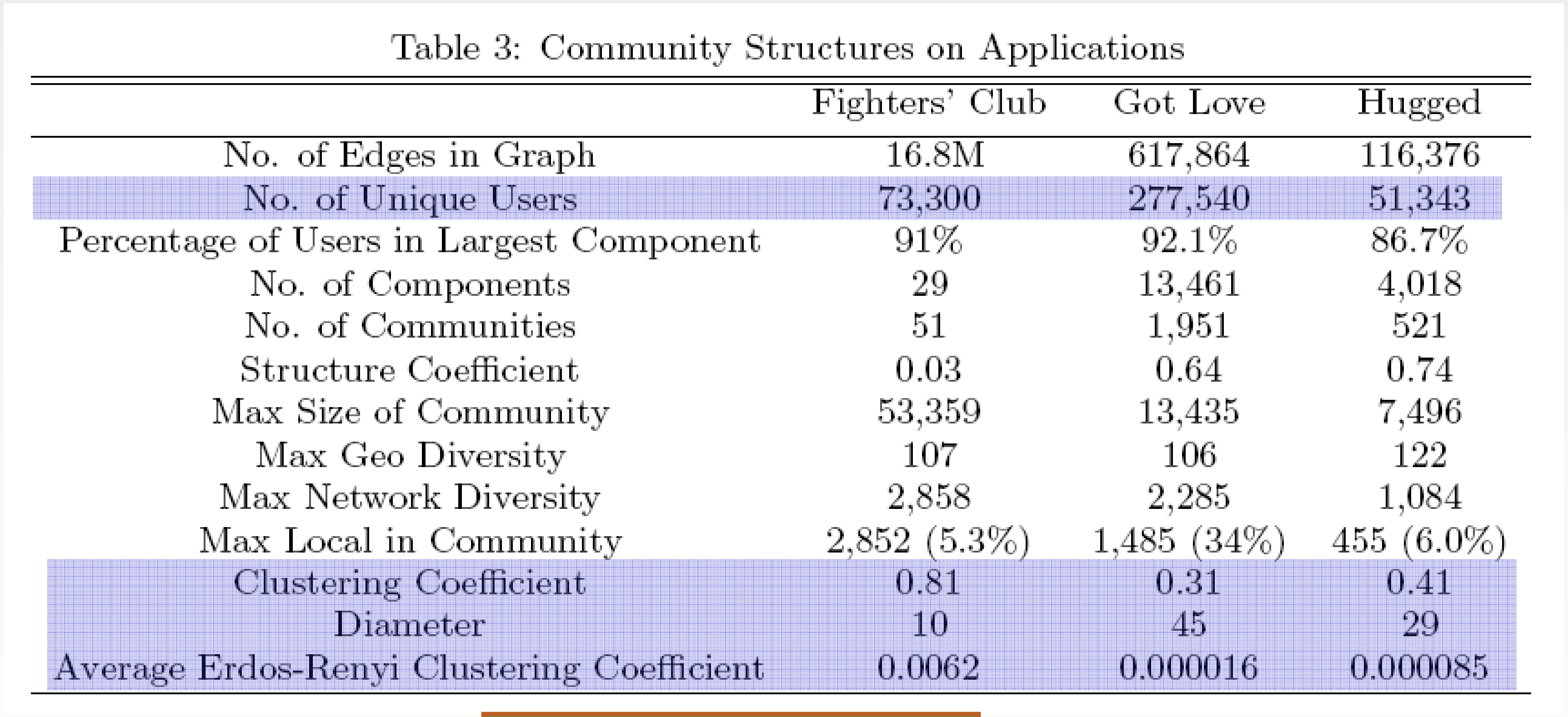


Table 3: Community Structures on Applications

	Fighters' Club	Got Love	Hugged
No. of Edges in Graph	16.8M	617,864	116,376
No. of Unique Users	73,300	277,540	51,343
Percentage of Users in Largest Component	91%	92.1%	86.7%
No. of Components	29	13,461	4,018
No. of Communities	51	1,951	521
Structure Coefficient	0.03	0.64	0.74
Max Size of Community	53,359	13,435	7,496
Max Geo Diversity	107	106	122
Max Network Diversity	2,858	2,285	1,084
Max Local in Community	2,852 (5.3%)	1,485 (34%)	455 (6.0%)
Clustering Coefficient	0.81	0.31	0.41
Diameter	10	45	29
Average Erdos-Renyi Clustering Coefficient	0.0062	0.000016	0.000085

Actually Small World Networks!

ACM EC 2009

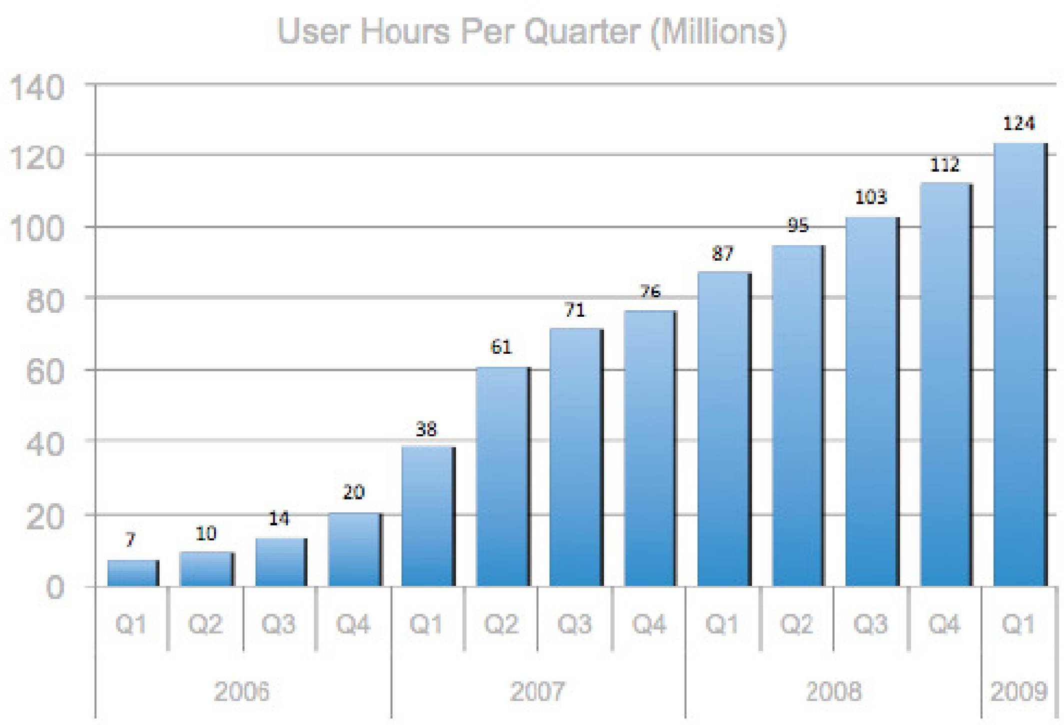
Social Influence and the Diffusion of User-created Content

E. Bakshy, B. Karrer, L. Adamic

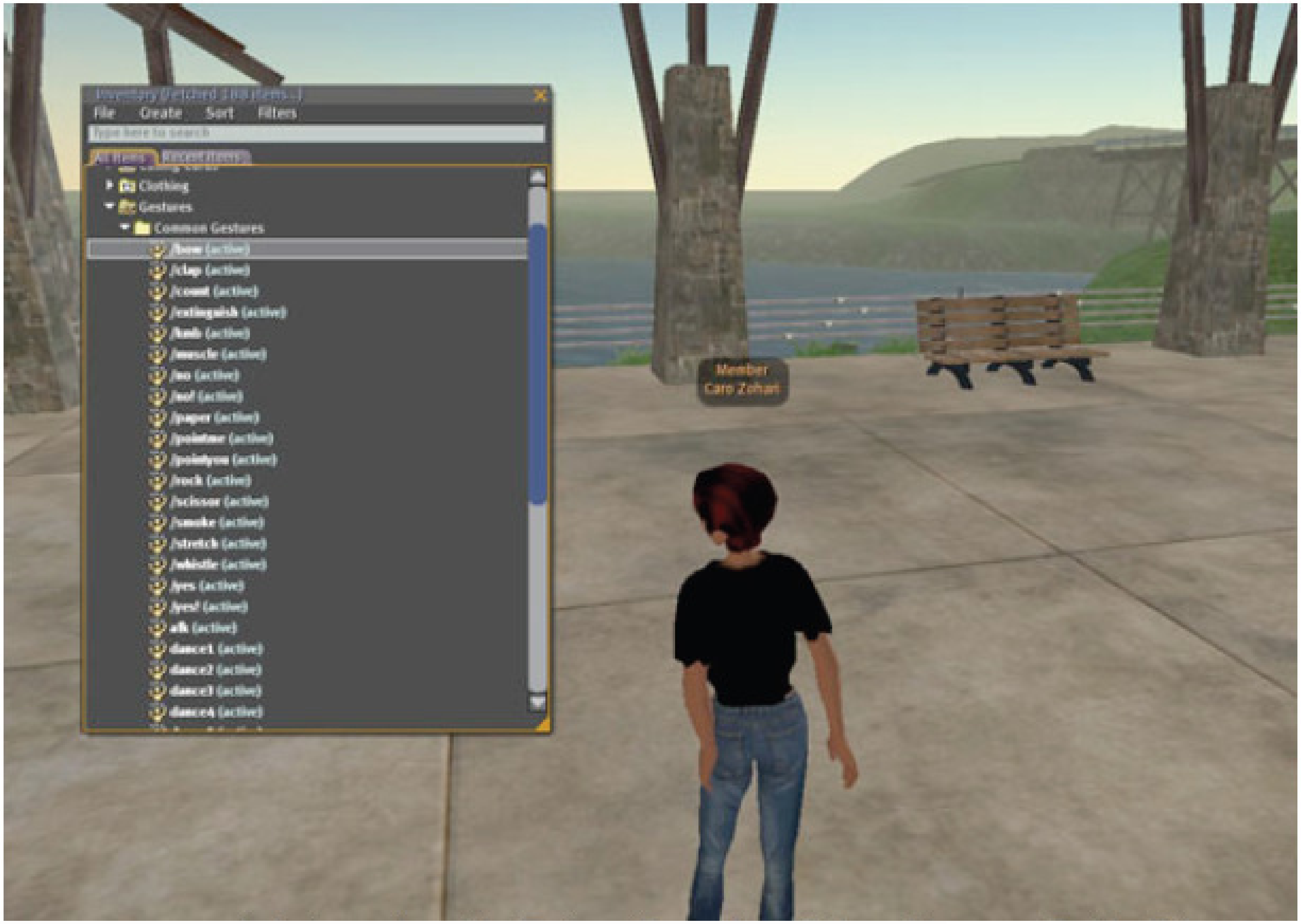
University of Michegan

Why study second life?

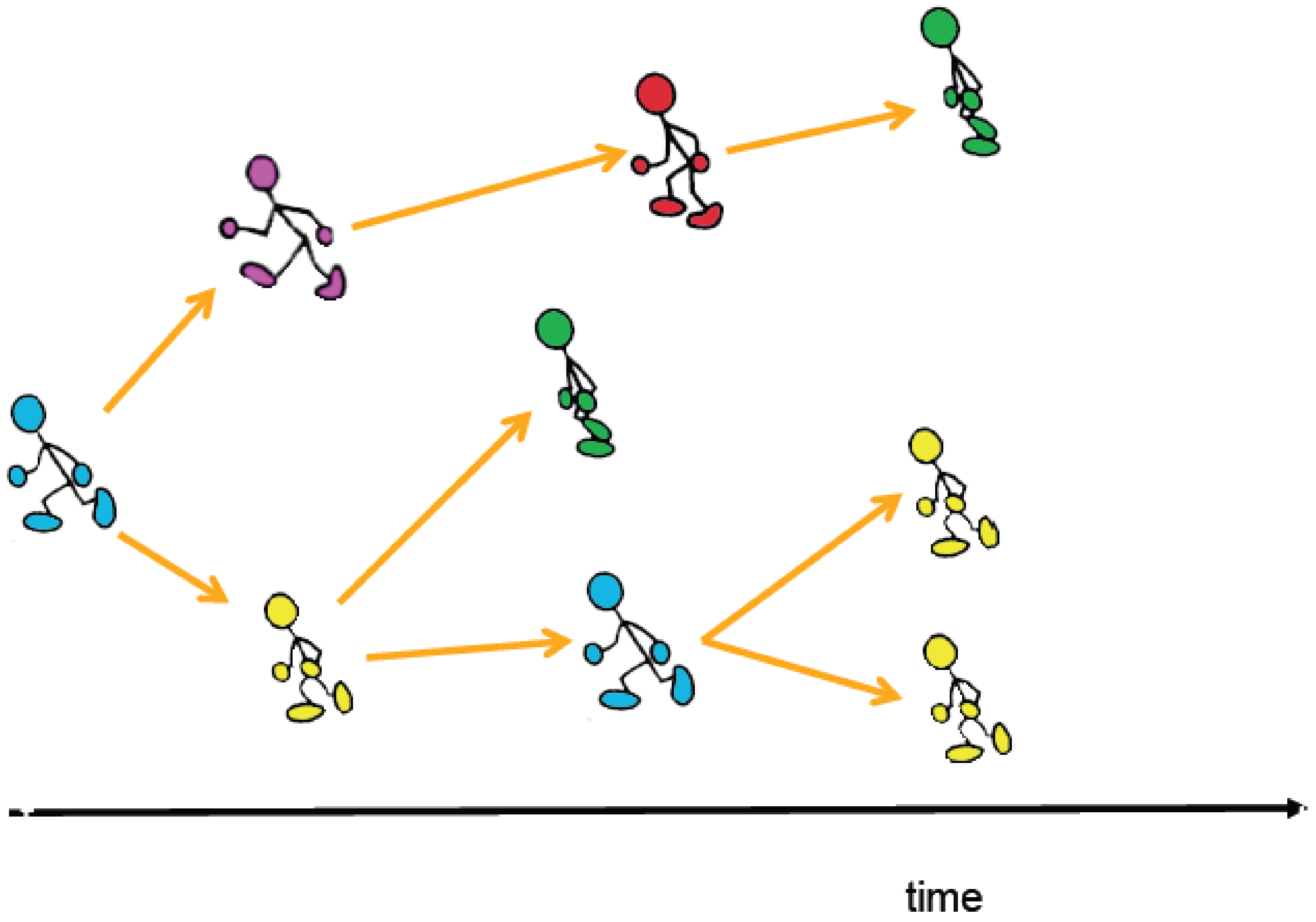
- digital traces!
- content is user-created
- content is shared and traded



gestures in second life



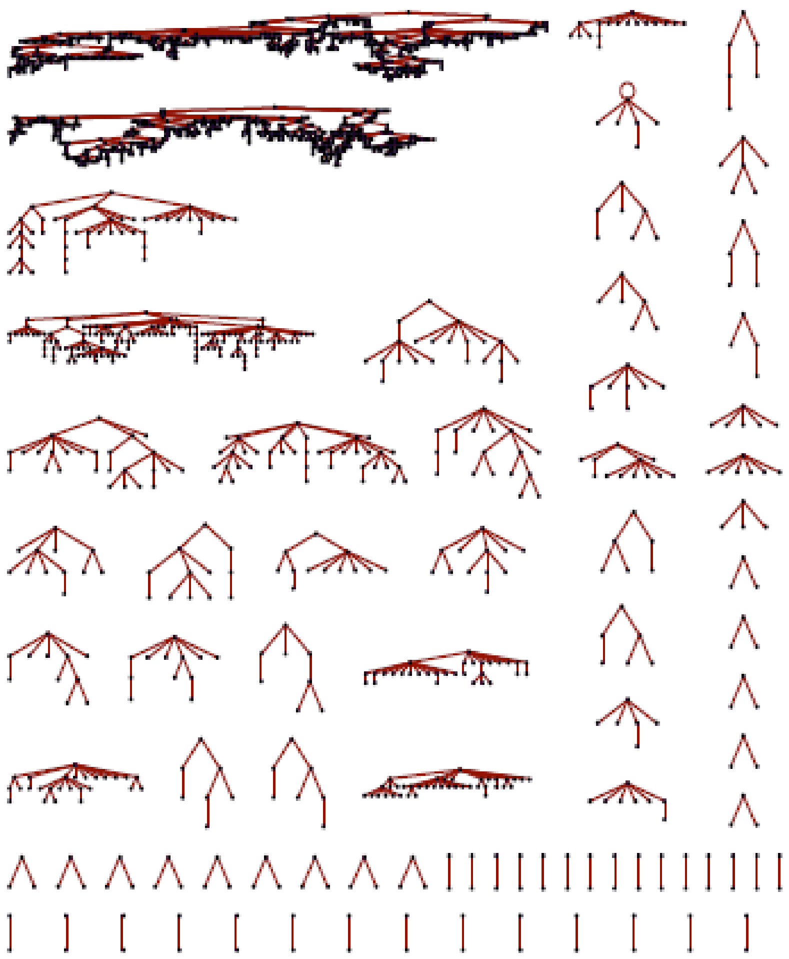
How do gestures spread?



Dataset

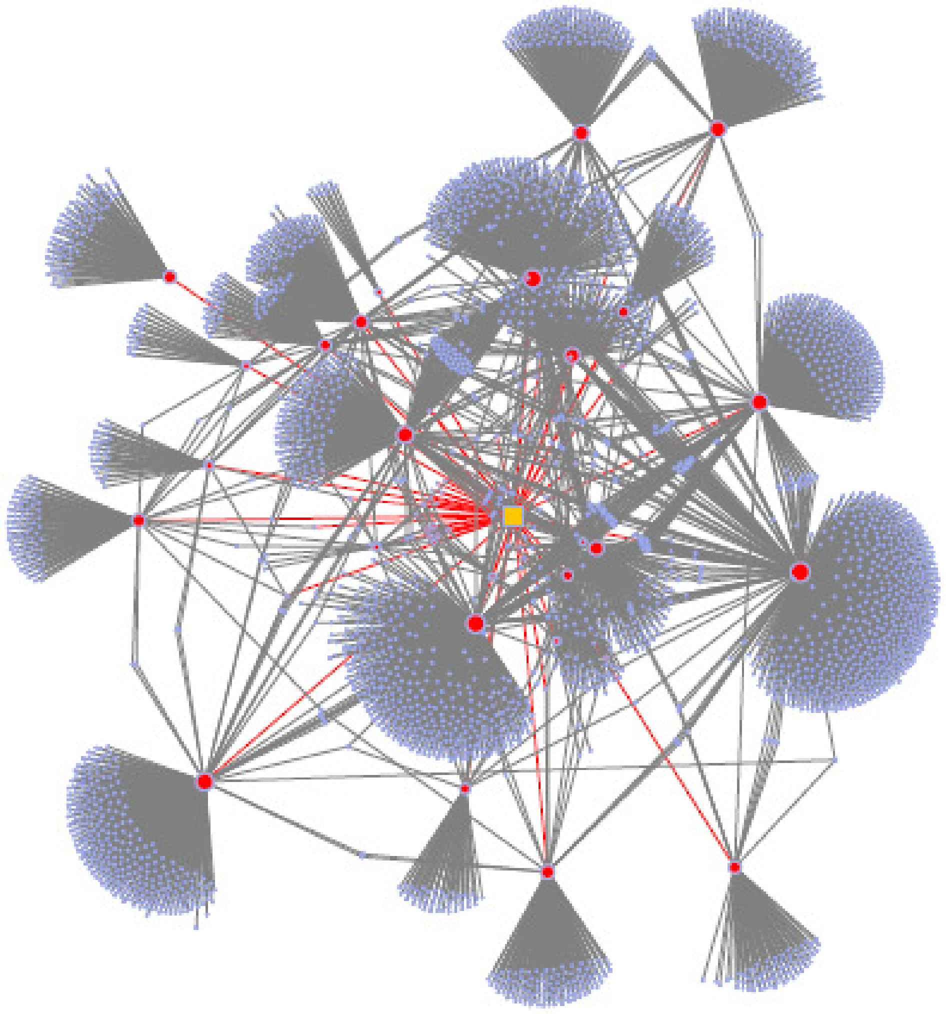
- gesture transfers 9/2008-1/2009
- 100,229 users who exchanged at least 1 object
- 106,499 assets with at least 16 unique owners & not distributed by Linden Lab

spread of an aerosmith gesture



the role of the social network

- weekly snapshots
- direct influence:
  - 48% of transfers occur between friends
- indirect influence:
  - of the remainder 38% occur after at least one friend has adopted



Dados dos servidores

- Dados do MSN
  - Planetary-Scale Views on a Large Instant-Messaging Network. **WWW'08**.
- Dados do CyWorld
  - Comparison of Online Social Relations in Terms of Volume vs. Interaction: A Case Study of Cyworld. **IMC'08**.
- Dados do YouTube
  - Video Suggestion and Discovery for YouTube: Taking Random Walks Through The View Graph**. WWW'08**.
- Dados do UOL
  - Characterization and Analysis of User Profiles in Online Video Sharing Systems. **JIDM'10**.

WWW 2008

Planetary-Scale Views on a Large Instant-Messaging Network

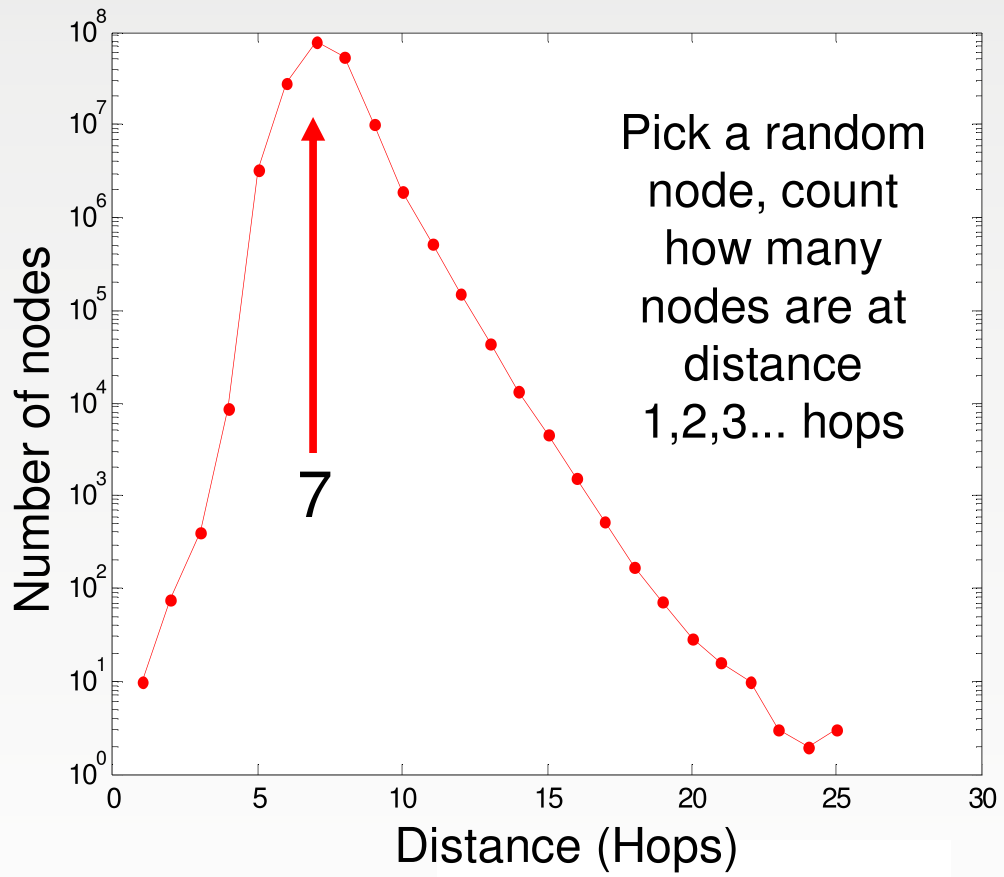
Jure Leskovec and Eric Horvitz

Carnegie Mellon University

Microsoft Research

Small-world effect

- Microsoft Messenger network
  - 180 million people
  - 1.3 billion edges
  - Edge if two people exchanged at least one message in one month period



WWW 2008

Comparison of Online Social Relations in Terms of Volume vs. Interaction: A Case Study of Cyworld

Hyunwoo Chun, Haewoon Kwak, Young-Ho Eom, Yong-Yeol Ahn, Sue Moon, Hawoong Jeong

KAIST

Cyworld

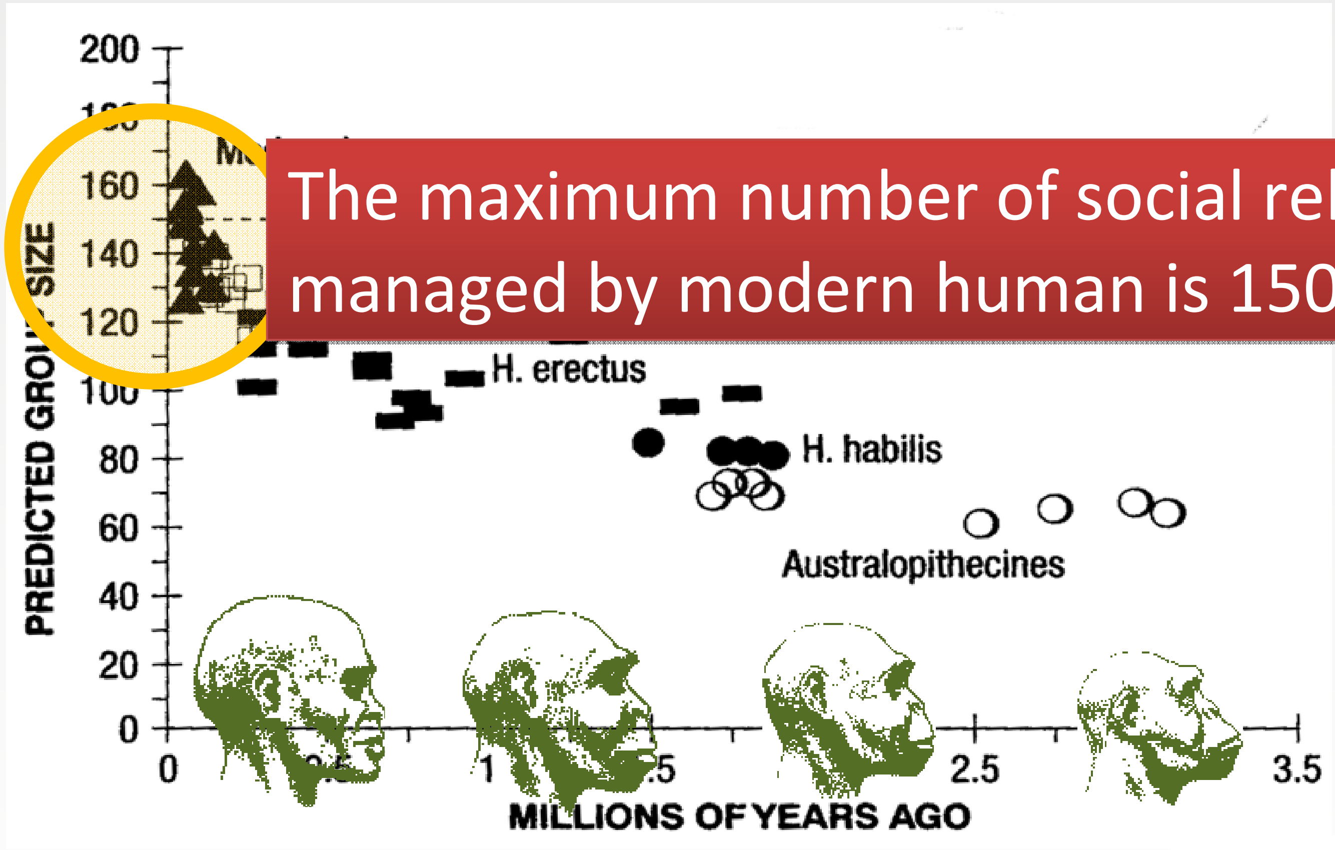
- Most popular OSN in Korea (22M users)

- Guestbook is the most popular feature
- Each guestbook message has 3 attributes
  - < From, To, When >

- We analyze 8 billion guestbook msgs of 2.5yrs

Dunbar's number

Behavioral and brain scineces, 16(4):681–735, 1993



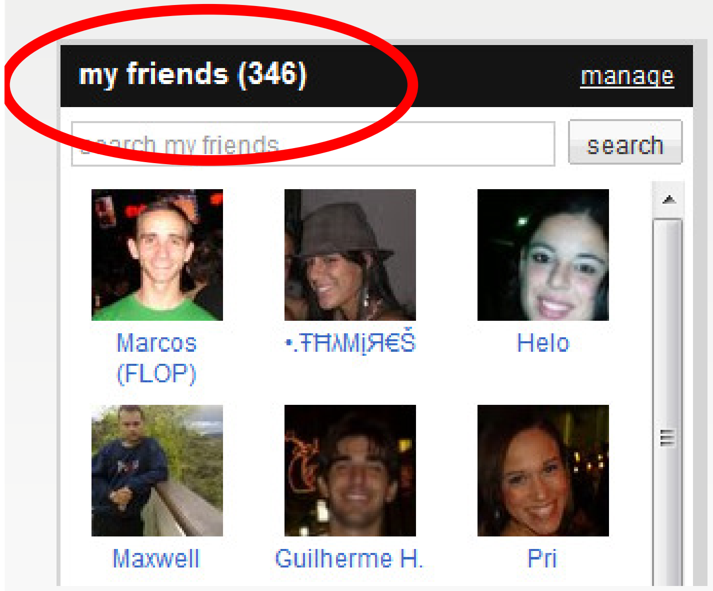
The maximum number of social relations managed by modern human is 150.

Cyworld 200 vs. Dunbar’s 150

- Has human networking capacity really grown?
  - Yes, technology helps users to manage relations
  - No, it is only an inflated number

Dunbar's number

Orkut



my friends (346) manage

search my friends search

Marcos (FLOP)

•.ŦĦΛMịЯ€Š

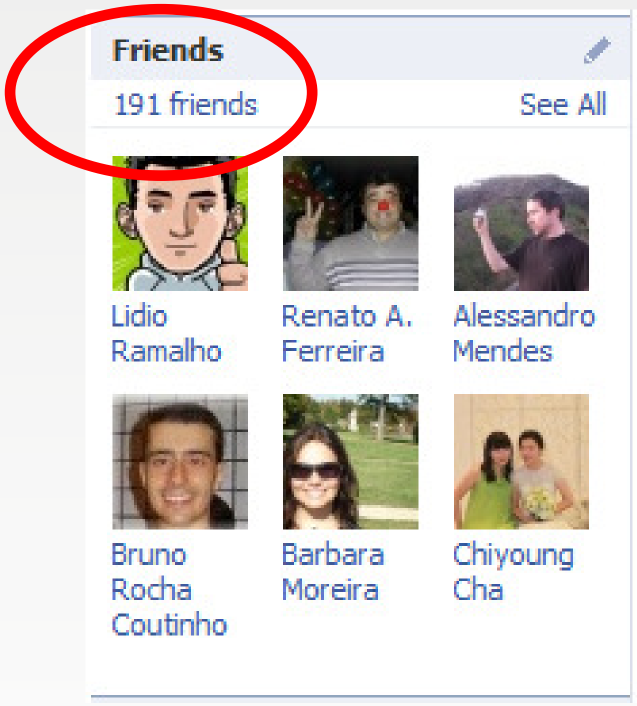
Helo

Maxwell

Guilherme H.

Pri

Facebook



Friends

191 friends See All

Lidio Ramalho

Renato A. Ferreira

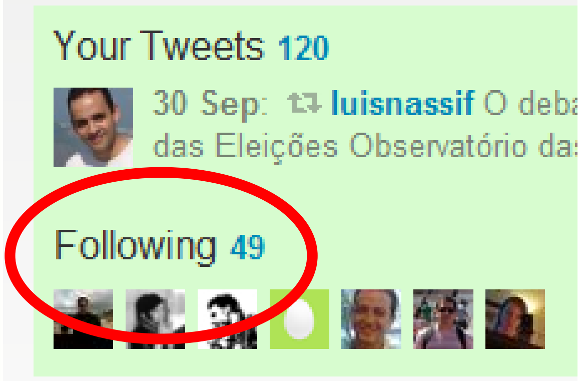
Alessandro Mendes

Bruno Rocha Coutinho

Barbara Moreira

Chiyoung Cha

Twitter



Your Tweets 120

30 Sep: luisnassif O deba das Eleições Observatório da

Following 49

WWW 2008

Video Suggestion and Discovery for YouTube: Taking Random Walks Through The View Graph

S. Baluja and R. Seth and D. Sivakumar and Y. Jing and J. Yagnik and S. Kumar and D. Ravichandran and M. Aly

Google

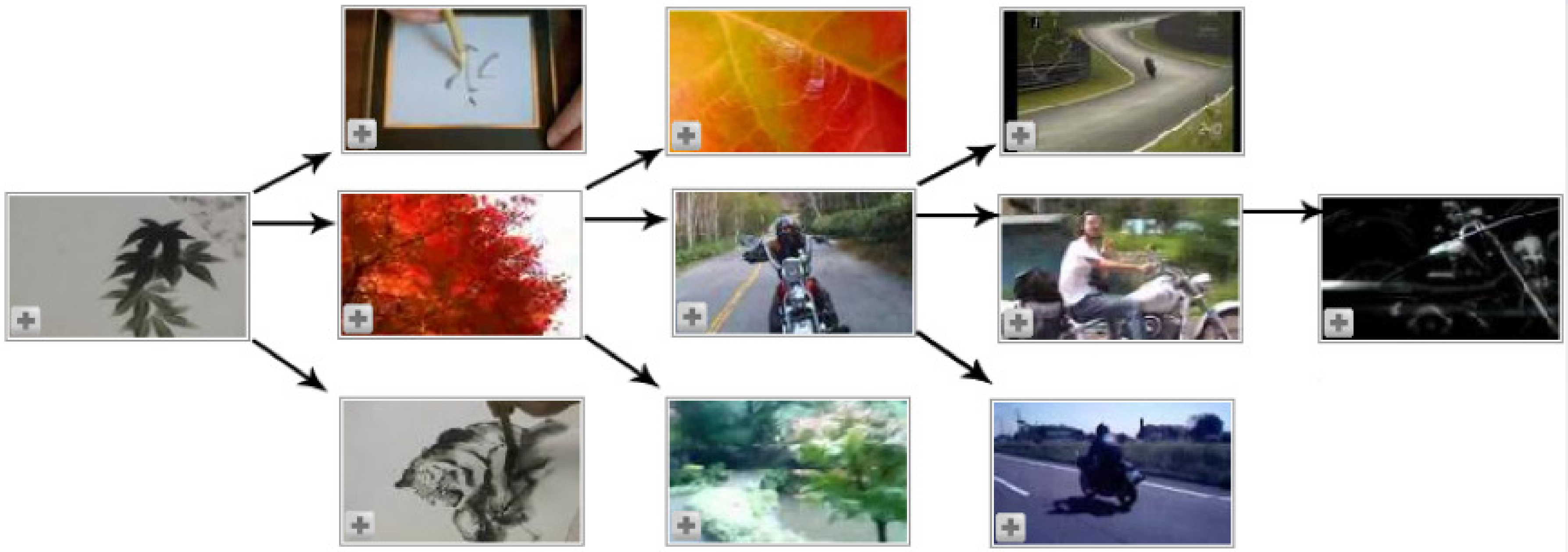
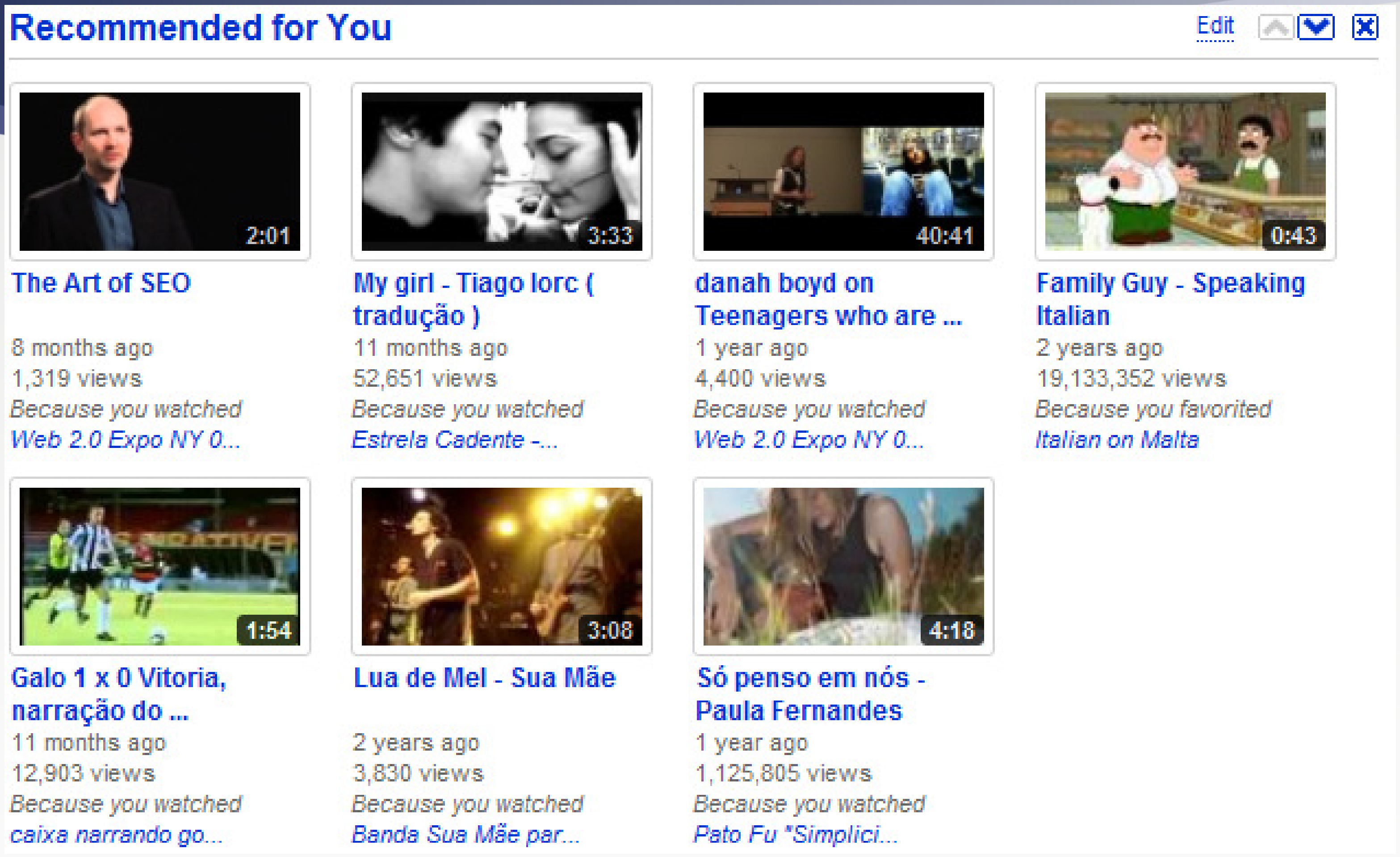


Figure 1: Video-Video Co-View Graph. Each video is a vertex in the graph that is linked to other videos often co-viewed. Often, only links with some minimum number of views are instantiated.



Recommended for You

The Art of SEO
2:01
8 months ago
1,319 views
Because you watched
Web 2.0 Expo NY 0...

My girl - Tiago Iorc (tradução)
3:33
11 months ago
52,651 views
Because you watched
Estrela Cadente -...

danah boyd on Teenagers who are ...
40:41
1 year ago
4,400 views
Because you watched
Web 2.0 Expo NY 0...

Family Guy - Speaking Italian
0:43
2 years ago
19,133,352 views
Because you favorited
Italian on Malta

Galo 1 x 0 Vitoria, narração do ...
1:54
11 months ago
12,903 views
Because you watched
caixa narrando go...

Lua de Mel - Sua Mãe
3:08
2 years ago
3,830 views
Because you watched
Banda Sua Mãe par...

Só penso em nós - Paula Fernandes
4:18
1 year ago
1,125,805 views
Because you watched
Pato Fu "Simplici...

JIDM 2010

Characterization and Analysis of User Profiles in Online Video Sharing Systems

Fabrício Benevenuto[1], Adriano Pereira[2], Tiago Rodrigues[1],

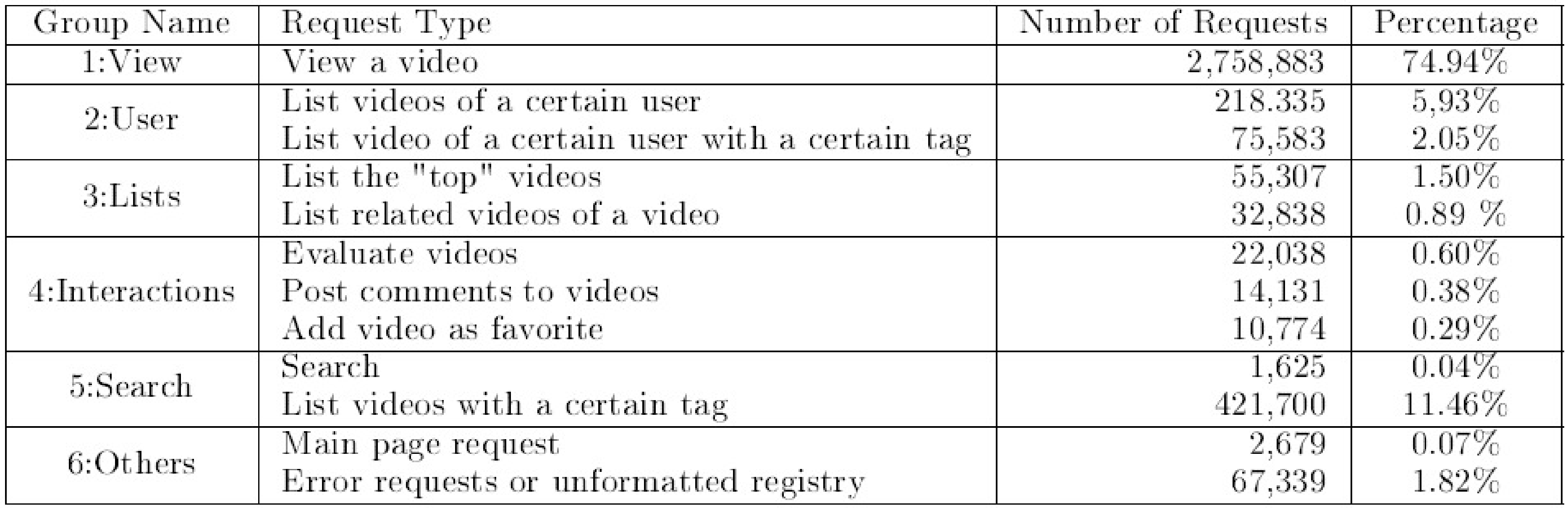
Virgílio Almeida[1], Jussara Almeida[1], Marcos Gonçalves[1]

[1]UFMG

[2]CEFET-MG

UOL Video Service Dataset

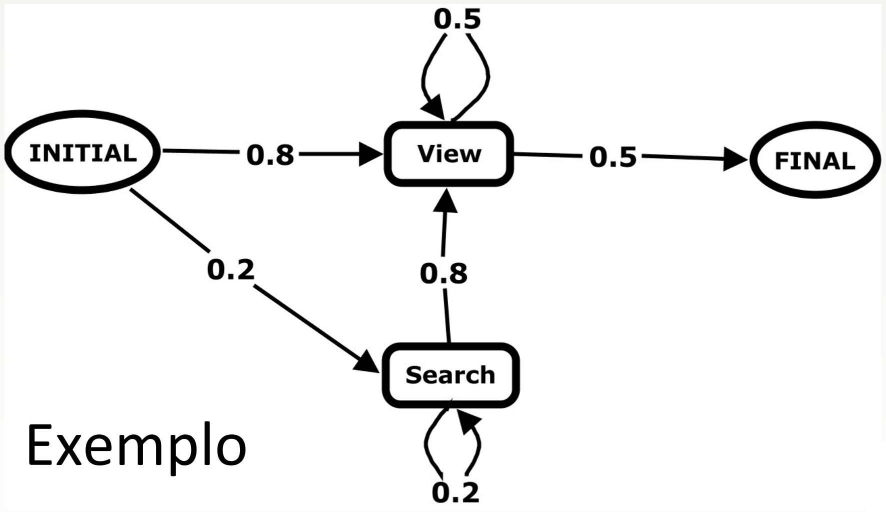
- Logs from the OVSN service from UOL
- Period: 12/12/2007 a 01/07/2008
- 3,681,232 requests from 1,127,537 different IPs
- Each line contains IP, time, request type, status, size, referee, and user-agent (anonymized)



Group Name	Request Type	Number of Requests	Percentage
1:View	View a video	2,758,883	74.94%
2:User	List videos of a certain user	218.335	5,93%
	List video of a certain user with a certain tag	75,583	2.05%
3:Lists	List the "top" videos	55,307	1.50%
	List related videos of a video	32,838	0.89 %
4:Interactions	Evaluate videos	22,038	0.60%
	Post comments to videos	14,131	0.38%
	Add video as favorite	10,774	0.29%
5:Search	Search	1,625	0.04%
	List videos with a certain tag	421,700	11.46%
6:Others	Main page request	2,679	0.07%
	Error requests or unformatted registry	67,339	1.82%

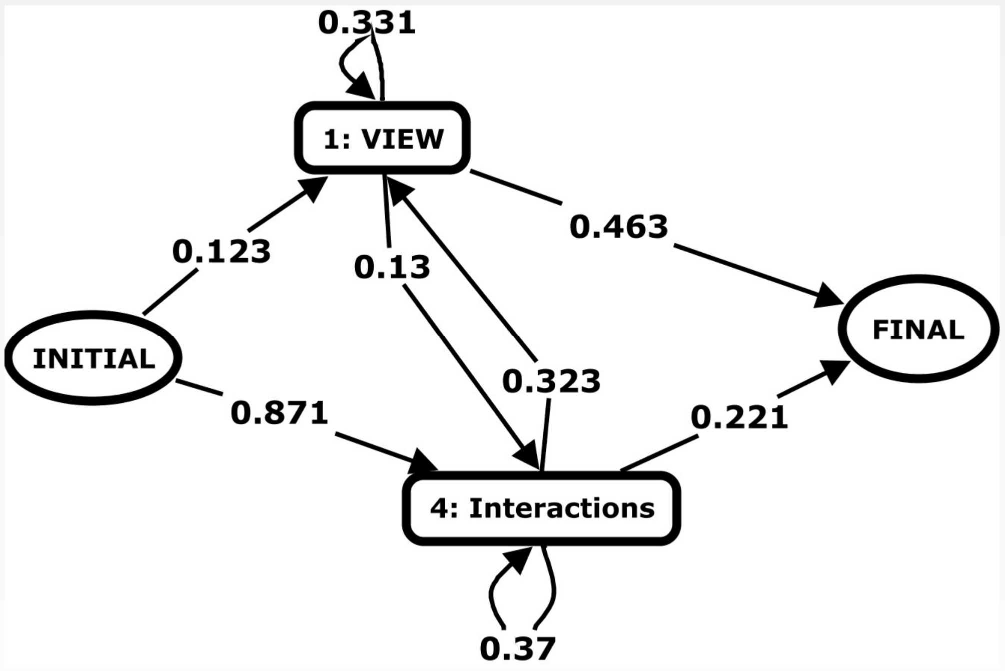
Navegação de tipos de usuários

- Probabilistic direct graph
  - **Nodes** = types of user requests. **Direct edges** = probability of navigation
  - Compute individual graphs based on all sessions of the user. Apply a clustering technique to identify different groups of users
  - Use X-means to define suitable number of groups



User Navigation Model Graphs

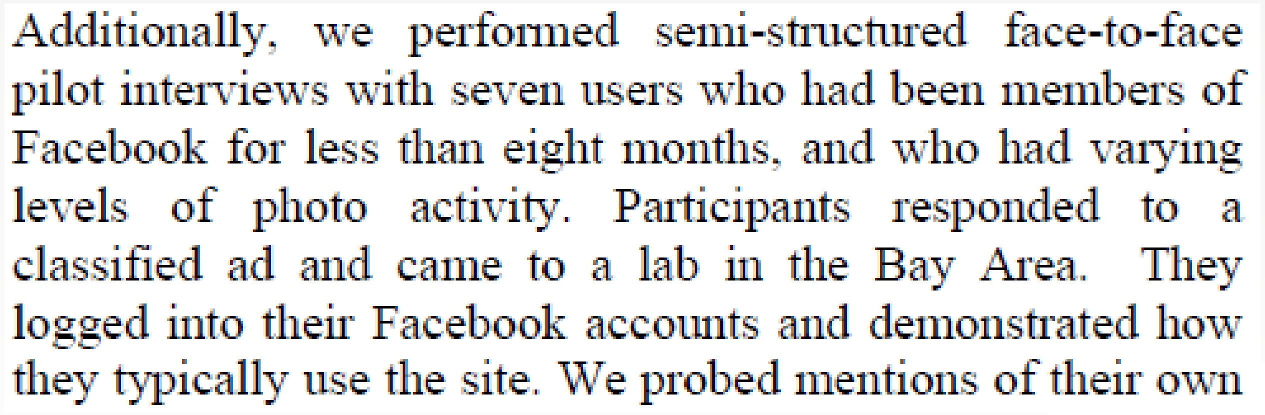
- Found 15 groups of users (also useful for service differentiation)
- Found a group of suspect users



Entrevistas formatadas

- Usuários respondem questionários formatados ou entrevistas, visando validar/refutar hipóteses
- Vários artigos do CHI. http://www.chi2010.org/

Feed Me: Motivating Newcomer Contribution in Social Network Sites.
M. Burke, C. Marlow, and T. Lento. CHI'2009.


Additionally, we performed semi-structured face-to-face pilot interviews with seven users who had been members of Facebook for less than eight months, and who had varying levels of photo activity. Participants responded to a classified ad and came to a lab in the Bay Area. They logged into their Facebook accounts and demonstrated how they typically use the site. We probed mentions of their own

Honeypots e coleções rotuladas

- Honeypots
  - Uncovering Social Spammers: Social Honeypots + Machine Learning. **SIGIR'10**
- Coleções rotuladas
  - Detecting Spammers on Twitter. **CEAS'10**

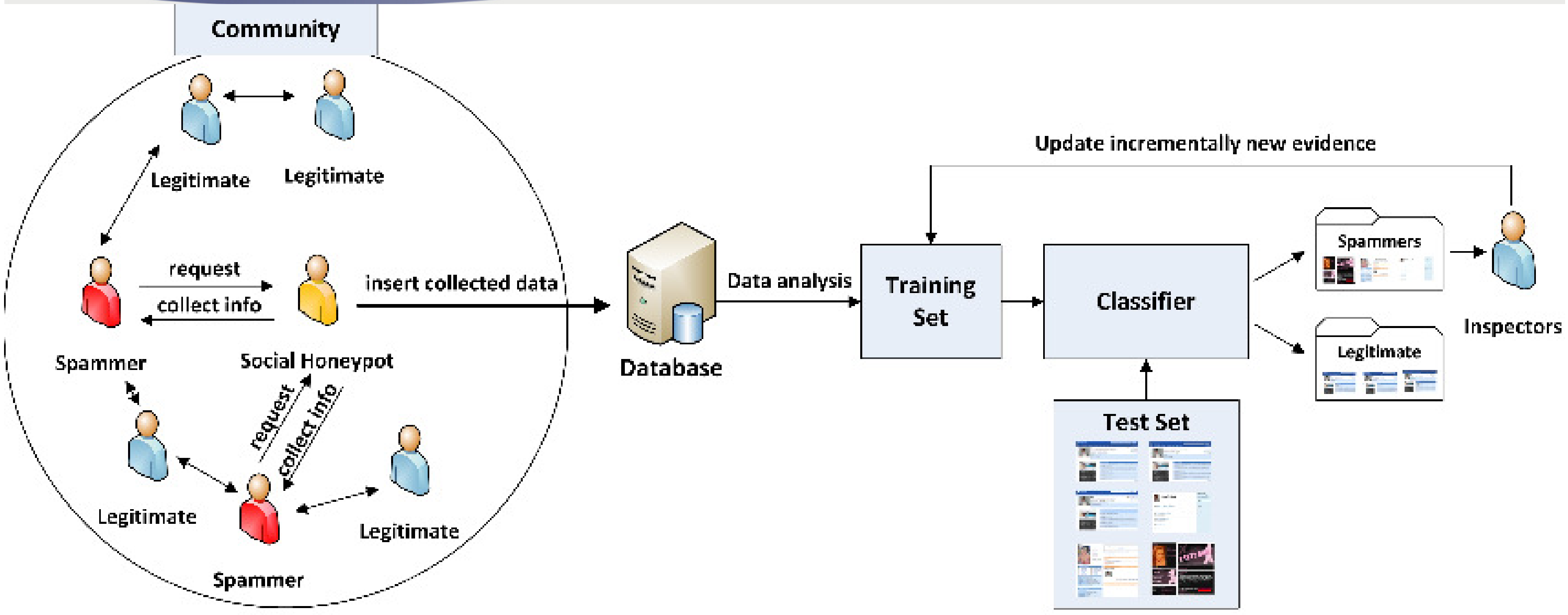
ACM SIGIR 2010

Uncovering Social Spammers: Social Honeypots + Machine Learning

K. Lee, J. Caverlee, and S. Webb

Texas A&M University

Abordagem



- Honeypots em dois sistemas: Myspace e Twitter

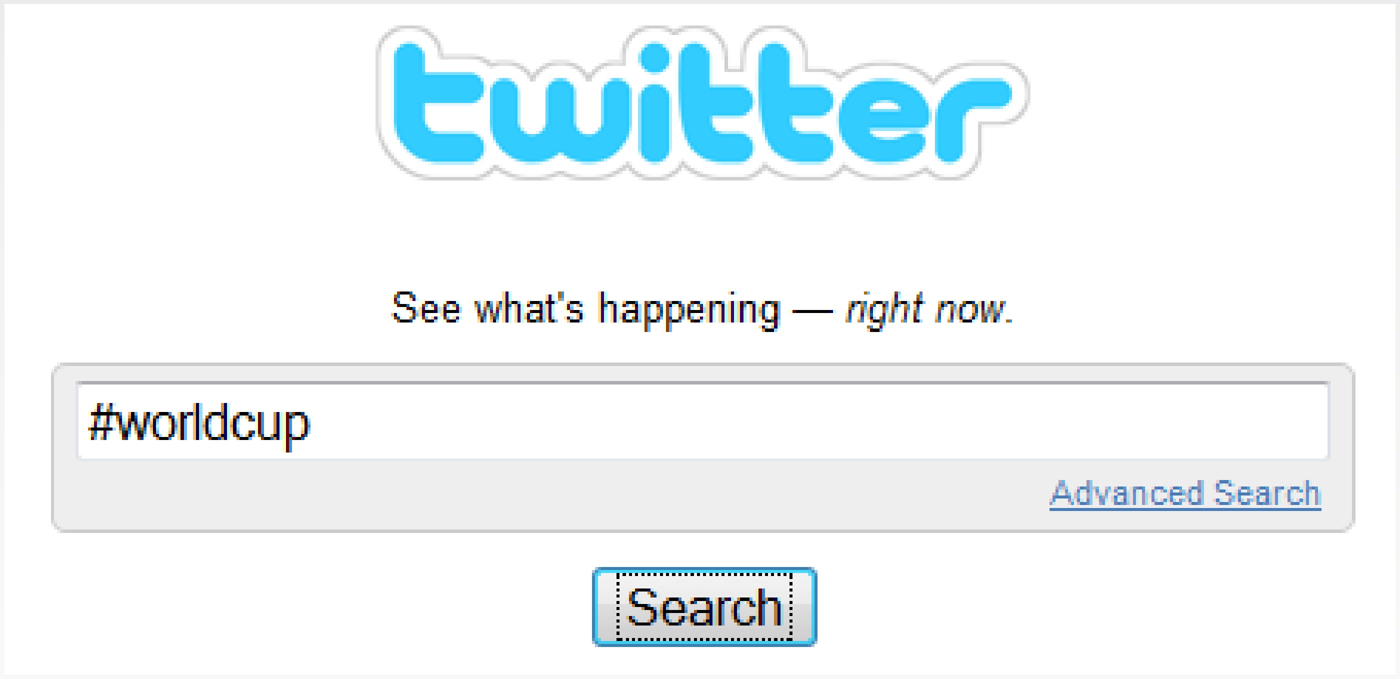
CEAS 2010

Detecting Spammers on Twitter

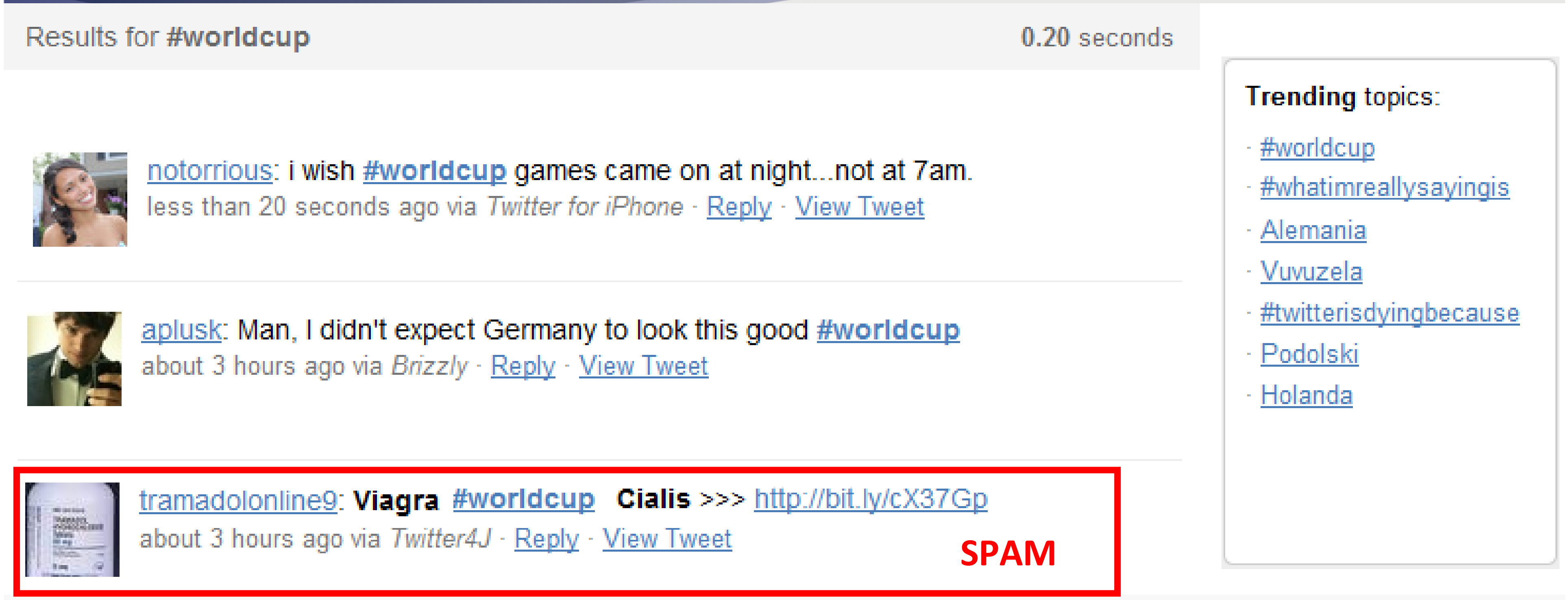
F. Benevenuto, A. Veloso, G. Magno, T. Rodrigues, V. Almeida

Universidade Federal de Minas Gerais

Spam no Twitter



Spam no Twitter



Results for **#worldcup** 0.20 seconds

notorrious: i wish #worldcup games came on at night...not at 7am.
less than 20 seconds ago via *Twitter for iPhone* · Reply · View Tweet

aplusk: Man, I didn't expect Germany to look this good #worldcup
about 3 hours ago via *Brizzly* · Reply · View Tweet

tramadolonline9: **Viagra** #worldcup **Cialis** >>> http://bit.ly/cX37Gp
about 3 hours ago via *Twitter4J* · Reply · View Tweet

SPAM

Trending topics:

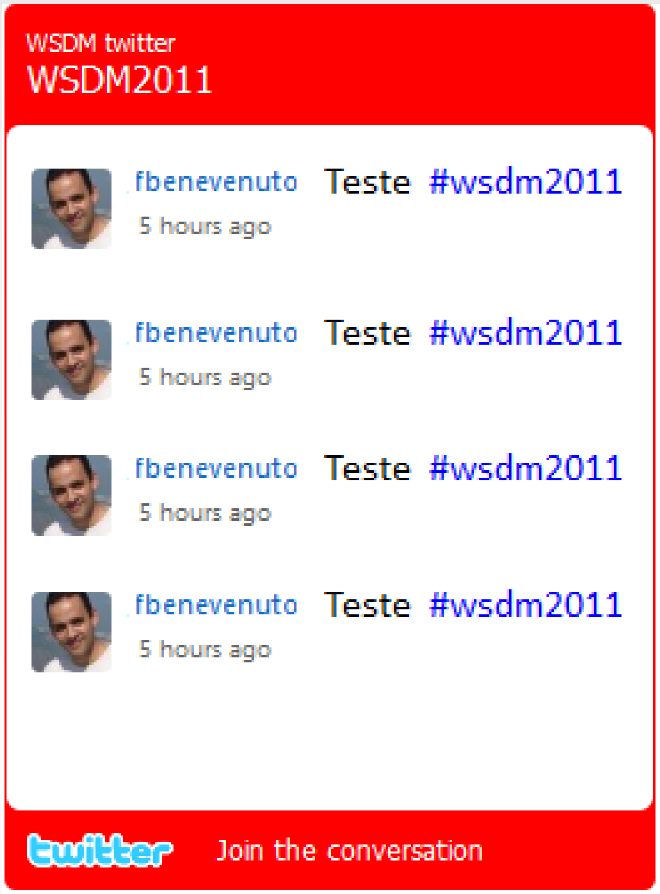
- #worldcup
- #whatimreallysayingis
- Alemania
- Vuvuzela
- #twitterisdyingbecause
- Podolski
- Holanda

Usuários postam URLs não relacionadas ao conteúdo

Spam on Twitter

Afeta mashups e ferramentas meme-tracking

- E.g. Conferences:
  - http://www.wsdm2011.org/
- E.g. Observatório da Web:
  - http://observatorio.inweb.org.br/



WSDM twitter
WSDM2011

fbenevenuto Teste #wsdm2011
5 hours ago

fbenevenuto Teste #wsdm2011
5 hours ago

fbenevenuto Teste #wsdm2011
5 hours ago

fbenevenuto Teste #wsdm2011
5 hours ago

twitter Join the conversation

Objetivos e Metodologia

1. Coleta do Twitter e criação de uma coleção de usuários manualmente rotulados como spammers ou não spammers

2. Caracterização do comportamento dos usuários

- Identificação de características capazes de distinguir spammers de não spammers

3. Criação de um método de detecção de spammers que utiliza as características do comportamento dos usuários

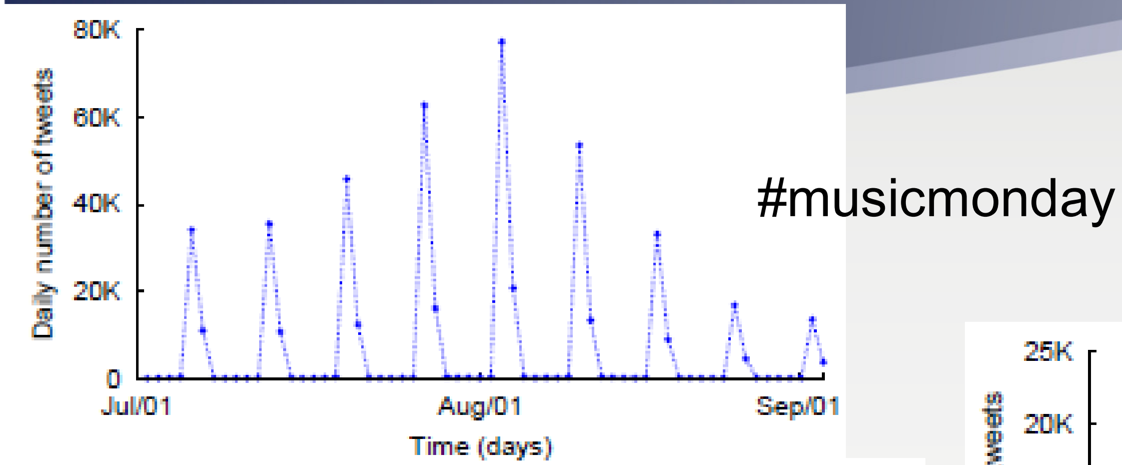
Propriedades desejáveis da coleção rotulada

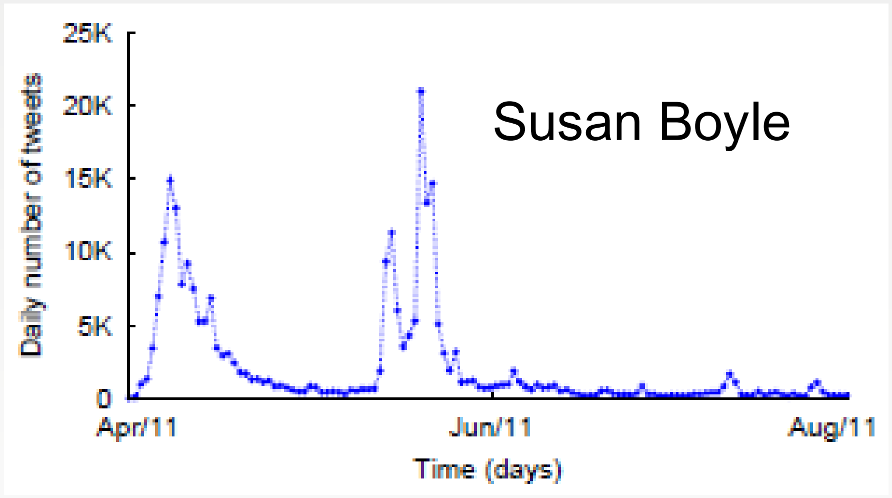
1) Ter um número significativo de spammers e usuários legítimos

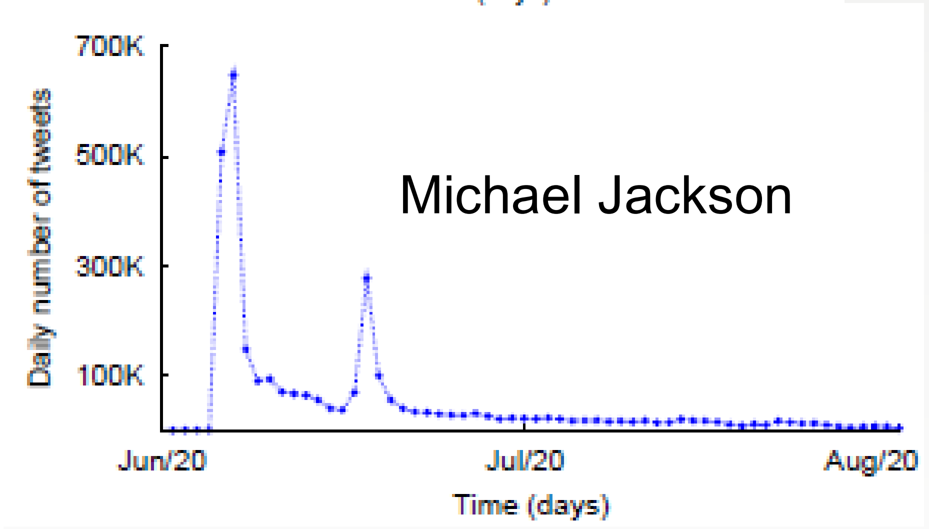
2) Incluir spammers que são agressivos em suas estratégias

3) Escolher usuários aleatoriamente e não baseados em suas características

Coleção rotulada

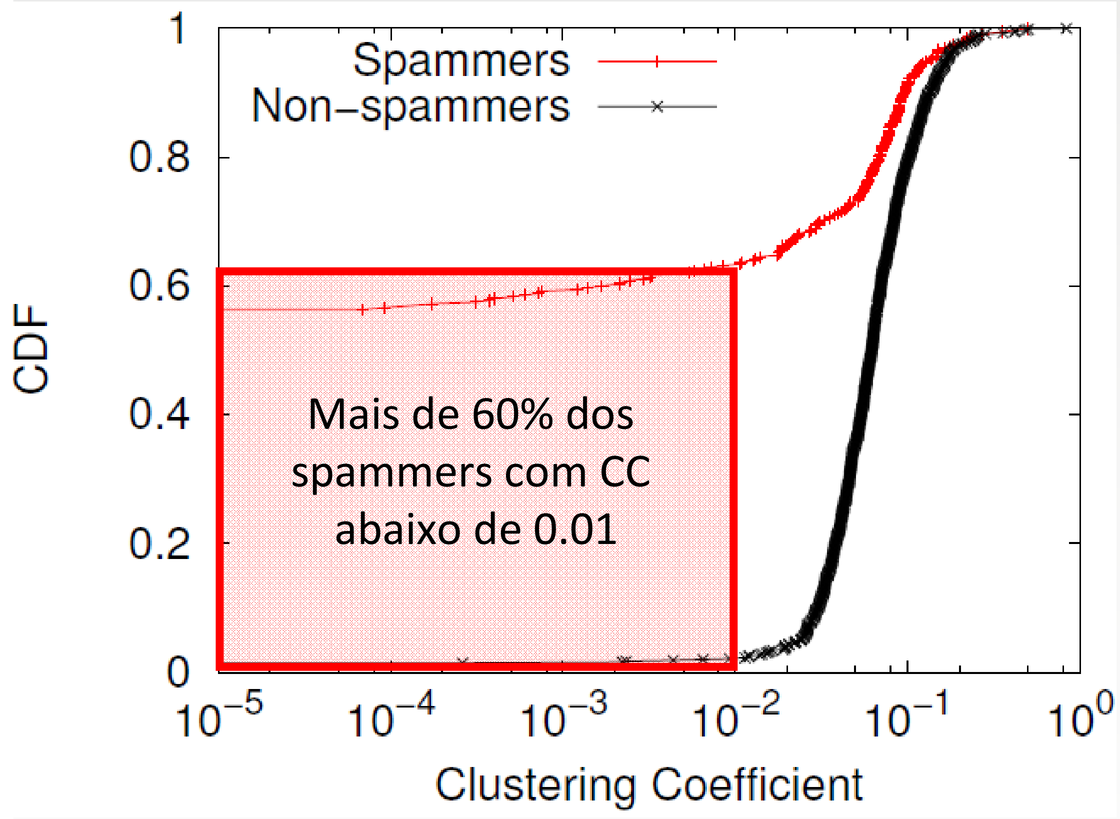






8207 usuários analisados dos quais 355 são spammers

Coeficiente de Clusterização



Coeficiente de clusterização probabilidade dos vizinhos de um nodo estarem conectados

Amigos dos spammers não estão conectados entre si

Obrigado!

- Slides e texto do curso na minha página
- Colaborações, datasets, mestrado na UFOP....
- SBRC 2012 será em Ouro Preto e tem social networks no CFP

Fabrício Benevenuto

e-mail: benevenuto@gmail.com

www.dcc.ufmg.br/~fabricio