Títulos

Comprovante Graduação 1/3

República Federativa do Brasil
Ministério da Educação
Universidade Federal de Campina Grande

Certificado

O Reitor da Universidade Federal de Campina Grande, no uso de suas atribuições e tendo em vista a conclusão do Curso de ***Graduação em Ciência da Computação***, em 27 de outubro de 2007, confere o título de ***Bacharel em Ciência da Computação*** a ***FLAVIO VINICIUS DINIZ DE FIGUEIREDO,*** cédula de identidade nº 2782251-SSP-PB, e lhe outorga o presente Certificado a fim de que possa gozar de todos os direitos e prerrogativas legais, até que seja substituído pelo competente Diploma, devidamente registrado.

Campina Grande, 27 de outubro de 2007.

Pró-Reitor de Ensino

Coordenador do Curso

Vicemário Simões
Pró-Reitor de Graduação
MAT. SIAPE 330202

REPÚBLICA FEDERATIVA DO BRASIL
ORDEM E PROGRESSO

Curso reconhecido pela Portaria do MEC nº 60, publicada no D.O.U de 15 de janeiro de 1980.

Comprovante Graduação 2/3

República Federativa do Brasil
Ministério da Educação
Universidade Federal de Campina Grande

Diploma

O Reitor da Universidade Federal de Campina Grande, no uso de suas atribuições e tendo em vista a conclusão do Curso de Ciência da Computação, em 27 de outubro de 2007, confere o título de **Bacharel em Ciência da Computação** a **Flavio Vinicius Diniz de Figueiredo,** brasileiro, nascido em 30 de setembro de 1985, em Campina Grande-PB, cédula de identidade nº 2782251 SSP/PB, e lhe outorga o presente Diploma a fim de que possa gozar de todos os direitos e prerrogativas legais.

Campina Grande, 31 de outubro de 2007.

Flavio Vinicius Diniz de Figueiredo

Diplomado

Coordenador de Controle Acadêmico
Clebert José Alves

Reitor
Thompson Fernandes Mariz

Comprovante Graduação 3/3

MINISTERIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
PRÓ-REITORIA DE ENSINO
COORDENAÇÃO DE CONTROLE ACADÊMICO

Diploma registrado sob o n.º 1262, do livro A-06, fls. 1262, por delegação de competencia nos termos do art. 48 da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as Diretrizes e Bases da Educacão Nacional.

Processo n.º 23096.023926/07-10 PRG

Campina Grande, 31 de outubro de 2007

Ezimar Patricio
Portaria R/GR/ nº 002/2002

Vicemário Simões
PRÓ-REITOR

Curso Reconhecido pela PORTARIA Nº 60 de 15/01/1980, publicado no D.O.U. de 16/01/1980

Nº 07886

República Federativa do Brasil

UNIVERSIDADE FEDERAL DE MINAS GERAIS

O Reitor da Universidade Federal de Minas Gerais, Professor Jaime Arturo Ramírez, no uso de suas atribuições, confere a

Flavio Vinicius Diniz de Figueiredo

o grau de Mestre em Ciência da Computação

e outorga-lhe o presente diploma, nos termos da legislação vigente.

Belo Horizonte, 8 de agosto de 2014.

Reitor

Diretor da Unidade

Pró-Reitor de Pós-Graduação

Diplomado

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INCIPIT VITA NOVA
7 DE SETEMBRO DE 1927

DIPLOMADO: **FLAVIO VINICIUS DINIZ DE FIGUEIREDO**

Filiacao: EDUARDO ENEAS DE FIGUEIREDO
EVELINE DINIZ DE FIGUEIREDO

Data de Nascimento: 30 de setembro de 1985

Naturalidade: Paraíba

Nacionalidade: Brasileira

Documento de Identidade: 2782251/PB

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE MINAS GERAIS

Registro efetuado nos termos da
Lei Nº 9.394/96 de 20/12/1996 -- Artigo 48 -- Parágrafo 1º

Número 1238 Livro RD.2015/2
Proc. nº 2015/09.00435
Belo Horizonte, 16 setembro 2015

Luciana Moreira de Avelar
Diretora da Divisão de Registro de Diplomas

Daniele C. M. Fagundes Zárate
Diretora do Departamento de Registro e Controle Acadêmico

Data da defesa: 18 de maio de 2010.

Curso de Pós-Graduação em Ciência da Computação, nível Mestrado

Avaliação da Capes referente ao Triênio 2007-2009
Reconhecimento homologado pelo CNE (Portaria MEC 1077, de 31/08/2012, DOU 13/09/2012, seç. 1, p. 25.)

Setor de Expedição de Diplomas/PRPG

Comprovante Mestrado 3/4

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PPGCC
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DCC-UFMG

DECLARAÇÃO

Declaramos, para os devidos fins, que **Flavio Vinicius Diniz de Figueiredo**, matrícula 2008659458, concluiu o mestrado no Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Minas Gerais, Brasil, tendo defendido sua dissertação intitulada *Evidências de qualidade de atributos textuais na Web 2.0* no dia 18/05/2010. O processo de expedição do diploma encontra-se em andamento.

Belo Horizonte, 07 de outubro de 2015.

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO – UFMG

Prof. Luiz Chaimowicz
Coordenador do Programa de Pós-Graduação
em Ciência da Computação - ICEx - UFMG

Prof. Luiz Chaimowicz
Coordenador do Colegiado do Programa de
Pós-Graduação em Ciência da Computação
ICEx - UFMG / Portaria 8571

Comprovante Mestrado 4/4

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Evidências de qualidade de atributos textuais na Web 2.0

FLAVIO VINICIUS DINIZ DE FIGUEIREDO

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

Jussara Marques de Almeida

PROFA. JUSSARA MARQUES DE ALMEIDA - Orientadora
Departamento de Ciência da Computação - UFMG

Marcos André Gonçalves

PROF. MARCOS ANDRÉ GONÇALVES - Co-orientador
Departamento de Ciência da Computação - UFMG

Cristina Duarte Murta

PROFA. CRISTINA DUARTE MURTA
CEFET - MG

PROF. EDLENO SILVA DE MOURA
Departamento de Ciência da Computação - UFAM

Pável Pereira Calado

PROF. PÁVEL PEREIRA CALADO
Instituto de Engenharia de Sistemas e Computadores - INESC

Belo Horizonte, 18 de maio de 2010.

Comprovante Doutorado 1/2

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Understanding, modeling and predicting the popularity of online content on social media applications

FLAVIO VINICIUS DINIZ DE FIGUEIREDO

Tese defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROFA. JUSSARA MARQUES DE ALMEIDA GONÇALVES - Orientadora
Departamento de Ciência da Computação - UFMG

PROFA. ANA PAULA COUTO DA SILVA
Departamento de Ciência da Computação - UFMG

PROF. ANIRBAN MAHANTI
NICTA - Austrália

PROF. CAETANO TRAINA JÚNIOR
Instituto de Ciências Matemáticas e de Computação - USP

PROF. FABRÍCIO BENEVENUTO DE SOUZA
Departamento de Ciência da Computação - UFMG

PROF. MARCOS ANDRÉ GONÇALVES
Departamento de Ciência da Computação - UFMG

PROF. RENATO MARTINS ASSUNÇÃO
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 01 de junho de 2015.

Comprovante Doutorado 2/2

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PPGCC
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DCC-UFMG

DECLARAÇÃO

Declaramos, para os devidos fins, que **Flavio Vinicius Diniz de Figueiredo**, matrícula 2010718237, concluiu o doutorado no Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Minas Gerais, Brasil, tendo defendido sua tese intitulada *Understanding, modeling and predicting the popularity of online content on social media applications* no dia 01/06/2015.

Informamos ainda, que, o processo de expedição do diploma encontra-se em andamento.

Belo Horizonte, 29 de outubro de 2015.

Prof. Luiz Chaimowicz
Coordenador do Programa de Pós-Graduação
em Ciência da Computação - ICEx - UFMG

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO UFMG

Comprovante Sanduíche CMU

Carnegie Mellon University

Christos Faloutsos
Professor
Computer Science Department
412-268-1457
christos@cs.cmu.edu

September 10, 2013

Dear Sir/Madam:

I am writing to inform you that I have accepted to advise the student, Flavio Figueiredo, for one year starting in October, 2013.

Flavio will perform activities related to his Ph.D. as part of the Scientists without Borders Ph.D. Sandwich Program.

Also, base on our previous contacts I acknowledge that Flavio is sufficient enough in English to perform his activities at Carnegie Mellon University.

Should you need additional information feel free to contact me at 412-268-1457.

Sincerely yours,

Christos Faloutsos
Professor
School of Computer Science
Carnegie Mellon University

Experiência Profissional

Comprovante Bolsa Pos-Doc UFCG (SET-C)

Ministério da Ciência e Tecnologia **CNPq**	**COMPROVANTE DE RENDIMENTO PAGOS E DE RETENÇÃO DE IMPOSTO DE RENDA NA FONTE ANO BASE 2015**

1.Fonte Pagadora Pessoa Jurídica ou Pessoa Física

Nome	CNPJ/CGC	
Conselho Nacional de Desenvolvimento Científico e Tecnológico	*33.654.831/0001-36*	
Endereço *SHIS Quadra 01 Conjunto B - Edifício Santos Dumont - Lago Sul*		
Cidade *Brasília*	UF *DF*	Telefone *(61) 3211-9000*

2.Pessoa Física Beneficiária dos Rendimentos

CPF	Nome Completo
058.784.734-45	*Flavio Vinicius Diniz de Figueiredo*
Natureza do Rendimento *Bolsa de Estudo*	

3.Rendimentos Tributáveis, Deduções e Imposto Retido na Fonte (em R$)

01.Total dos Rendimentos (inclusive férias)	0,00
02.Contribuição Previdenciária Oficial	0,00
03.Contribuição à Previdencia Privada e ao FAPI	0,00
04.Pensão Alimentícia (informar o beneficiário no campo 06)	*0,00*
05.Dependentes	0,00
06.Imposto Retido na Fonte	0,00

4.Rendimentos Isentos e não Tributáveis (em R$)

01.Salário Família	0,00
02.Parcela isenta dos proventos de Aposentadoria, Reforma e Pensão (65 anos)	0,00
03.Diárias e Ajudas de Custo	0,00
04.Pensão ,Proventos de Aposentadoria ou Reforma por Molestia Grave e Aposentadoria ou Reforma por Acidente em Serviço	0,00
05.Lucro e Dividendo Apurado a partir de 1996 pago por PJ (Lucro Real , Presumido ou Arbitrado).	0,00
06.Valores Pagos ao Titular ou Sócio de Microempresa ou Empresa de Pequeno Porte exceto Pro Labore, Aluguéis ou Serviços Prestados	0,00
07.Outros	
Bolsa de Estudo	*22.500,00*

5.Rendimentos Sujeitos à Tributação Exclusiva (Rendimento Líquido, em R$)

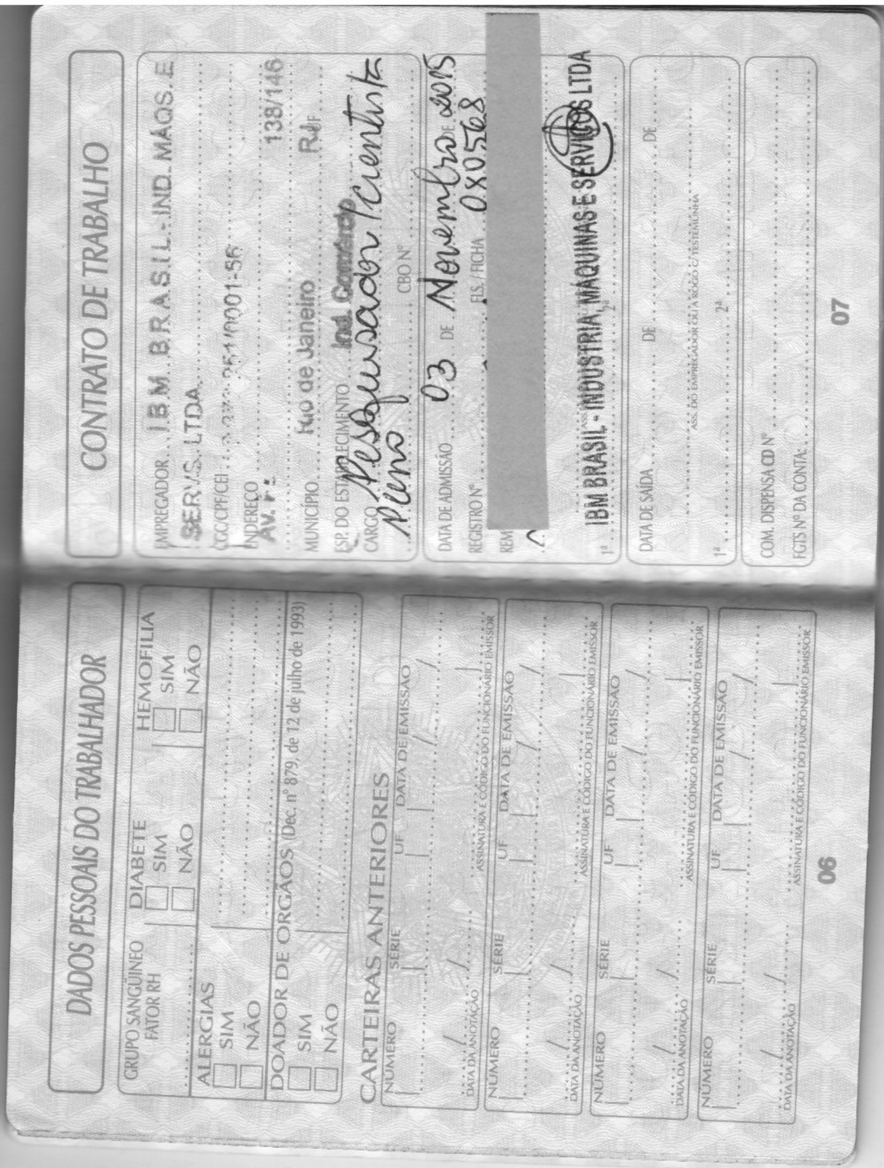
01.Décimo terceiro salário	0,00
02.Outros	0,00

6.Informações Complementares

Processos do Beneficiário : 350950/2015-6.

7.Responsável pelas Informações

Nome *Coordenação Geral de Execução do Fomento*



DADOS PESSOAIS DO TRABALHADOR

GRUPO SANGÜÍNEO FATOR RH ....

DIABETE ☐ SIM ☐ NÃO

HEMOFILIA ☐ SIM ☐ NÃO

ALERGIAS ☐ SIM ☐ NÃO

DOADOR DE ORGÃOS (Dec. nº 879, de 12 de julho de 1993) ☐ SIM ☐ NÃO

CARTEIRAS ANTERIORES

NÚMERO	SÉRIE	UF	DATA DE EMISSÃO
			/ /
DATA DA ANOTAÇÃO / /			ASSINATURA E CÓDIGO DO FUNCIONÁRIO EMISSOR
			/ /
DATA DA ANOTAÇÃO / /			ASSINATURA E CÓDIGO DO FUNCIONÁRIO EMISSOR
			/ /
DATA DA ANOTAÇÃO / /			ASSINATURA E CÓDIGO DO FUNCIONÁRIO EMISSOR
			/ /
DATA DA ANOTAÇÃO / /			ASSINATURA E CÓDIGO DO FUNCIONÁRIO EMISSOR

06

CONTRATO DE TRABALHO

EMPREGADOR I.B.M. B.R.A.S.I.L.-IND. MÁQS. E SERVS. LTDA.

CGC/CPF/CEI [illegible]251/0001-56

ENDEREÇO AV. [illegible] 138/146

MUNICÍPIO Rio de Janeiro RJ UF

ESP. DO ESTABELECIMENTO Ind. Comércio

CARGO Pesquisador Cientista Pleno

CBO Nº

DATA DE ADMISSÃO 03 DE Novembro DE 2015

REGISTRO Nº FLS./FICHA 080568

REM [redacted]

IBM BRASIL - INDUSTRIA, MÁQUINAS E SERVIÇOS LTDA

1ª 2ª

DATA DE SAÍDA .... DE .... DE ....

ASS. DO EMPREGADOR OU A ROGO C/ TESTEMUNHA

1ª .... 2ª ....

COM. DISPENSA CD Nº ....

FGTS Nº DA CONTA: ....

07

Comprovante Pesquisador IBM

Experiência Ensino

Comprovante Monitoria UFCG

República Federativa do Brasil
Ministério da Educação
Universidade Federal de Campina Grande

CERTIFICADO

Certificamos para os devidos fins que o aluno FLÁVIO VINICIUS D. DE FIGUEIREDO, exerceu a função de monitor bolsista da disciplina PROGRAMAÇÃO I integrante do Projeto de Iniciação á Docência intitulado "MELHORIA DO ENSINO DE GRADUAÇÃO NO CCT/UFCG" desenvolvido no Centro de Ciências e Tecnologia da UFCG, correspondendo ao ano letivo de 2004, com carga horária semanal de 12 horas.

Campina Grande-PB, 13 de outubro de 2005

PROFA. BETÂNIA MARIA DE OLIVEIRA
Coordenadora de Programas e Estágios
PRE/UFCG

PROF. ANTONIO PEDRO FERREIRA SOUSA
Assessor de Monitoria do CCT/UFCG

Comprovante Curso Extensão UFMG

Fundação de Desenvolvimento da Pesquisa

Plano de Atividades

Modalidade do estágio: ☐ OBRIGATÓRIO ☒ NÃO-OBRIGATÓRIO

Dados do Estagiário (a)

Nome: FLÁVIO VINÍCIUS DINIZ DE FIGUEIREDO

CPF: 058.784.734-45

Curso: Ciência da Computação Período: Mestrado

Projeto: 3642-13 ICEx/DCC/CURSOS/Java

Dados da Instituição de Ensino

Nome da Instituição de Ensino: Universidade Federal de Minas Gerais

Orientador do estágio (a ser preenchido pela Instituição de Ensino: Profa. Jussara Marques de Almeida

Formação do Orientador do Estágio : Doutora

Dados da Concedente/Supervisor

Fundação de Desenvolvimento da Pesquisa - FUNDEP

Supervisor do estágio: Prof. Antônio Otávio Fernandes

Formação : Doutor

Objetivos do estágio

- Lecionar a Linguagem de Programação JAVA através da plataforma J2SE. A disciplina ministrada foi Introdução a J2SE.
- Aulas Expositivas: O estagiário foi responsável por ministrar aulas presenciais e expositivas por um mês.
- Aulas de exercícios: Acompanhamento da turma em aulas práticas.
- Acompanhamento do monitor: O estagiário é responsável por acompanhar as atividades do monitor da disciplina.

Descrição das atividades (Descrição das atividades que deverão ser desenvolvidas pelo aluno no campo de estágio, com detalhamento dos procedimentos em conformidade com a área de formação do estagiário)

Belo Horizonte, 10 / 11 / 2008 .

FUNDEP
SUPERVISOR DO ESTÁGIO

ESTAGIÁRIO (A)

ORIENTADOR DO ESTÁGIO
INSTITUIÇÃO DE ENSINO

Av. Antônio Carlos, 6627 - Campus UFMG - Un. Adm. II
Campus UFMG Pampulha - Caixa Postal 856 - CEP 30123-970
Belo Horizonte - MG - Brasil

Prêmios

8th International ACM Web Science Conference 2016

Leibniz Universität Hannover

WebSci'16
Hannover
May 22nd - 25th, 2016

Association for Computing Machinery

BEST PAPER AWARD

kindly sponsored by facebook and presented to

Mariana Arantes, Flavio Figueiredo and Jussara M. Almeida

for the paper

Understanding Video-Ad Consumption on YouTube: A Measurement Study on User Behavior, Popularity, and Content Properties

Wendy Hall — Wolfgang Nejdl

General Chairs, Web Science 2016

WebSci - Best Paper

PPGCC
PROGRAMA DE PÓS-GRADUAÇÃO
EM CIÊNCIA DA COMPUTAÇÃO
DCC-UFMG

Semana de Seminários 2015

CERTIFICADO

Certificamos que o trabalho ***Understanding, Modeling and Predicting the Popularity of Online Content on Social Media Applications,*** de autoria do aluno ***Flávio Vinícius Diniz de Figueiredo***, orientado pela professora ***Jussara Marques de Almeida Gonçalves,*** foi premiado como *Menção Honrosa,* categoria doutorado, na Semana de Seminários do Programa de Pós-Graduação em Ciência da Computação, realizada no período de 10 a 12 de junho de 2015.

Belo Horizonte, 08 de outubro de 2015

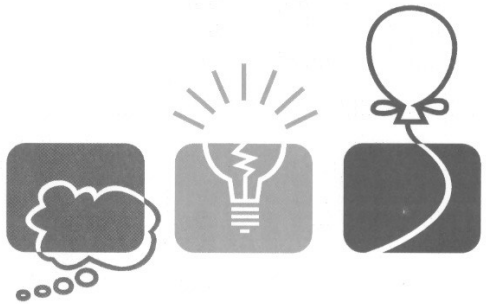
UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

UFMG
UNIVERSIDADE FEDERAL DE MINAS GERAIS

Luiz Chaimowicz
Coordenador do PPGCC

Ana Paula Couto da Silva
Coordenadora do Evento

Certificate of Achievement



acm International Collegiate Programming Contest

IBM event sponsor

awarded to Coach

Flavio Figueiredo

Universidade Federal de Minas Gerais

Honorable Mention

South America Brazil

Programming Contest

Universo September 20, 2008

William B. Poucher

William B. Poucher, Ph. D.
ICPC Executive Director

Menção Honrosa - ICPC

IBM

IBM Ph.D. FELLOWSHIP AWARD 2012

Recipient

Flavio Figueiredo

Universidade Federal de Minas Gerais

Departamento de Ciência da Computação

On behalf of Bernard Meyerson,
IBM Fellow, VP Innovation

Daniel M. Dias,
Distinguished Engineer, Director IBM Research - Brazil

Daniel M. Dias

IBM PhD Fellow

Evaluation

ECML/PKDD - Melhor Posição 2/3 tarefas

The Predictive Challenge 2014 has finished. Here are the results!

	Visits		Twitter		Facebook	
1st	**Flavio Figueiredo**	0.989812941222	**Gitte Vanwinckelen**	0.651184742657	**Flavio Figueiredo**	1.37830671254
2nd	Gitte Vanwinckelen	0.990598535376	Flavio Figueiredo	0.666506696220	Gitte Vanwinckelen	1.38076993166
3rd	Marc Boullé	1.053206073250	Marc Boullé	0.675431833637	Marc Boullé	1.42624047963
4th	Joao Palotti	1.067825597200	Shestakoff	0.815422997464	Joao Palotti	1.58924526319
5th	Shestakoff	1.105847877280	Joao Palotti	0.815867344020	Shestakoff(_ts)	1.74397954800
6th	Sergey Kovalchuk	1.416778188380	Sandeep	0.886336371563	Sandeep	2.23005056168
7th	Sandeep	1.734745713910	Sergey Kovalchuk	1.001072229550	Sergey Kovalchuk	2.54174057720

Flavio Figueiredo is with Universidade Federal de Minas Gerais, Brazil. Gitte Vanwinckelen is with the Department of Computer Science, KU Leuven, Belgium.

Data availability

The deadline for the competition have passed, but you still can get access to the dataset. You need to print, sign, and send to ecmlpkdd@chartbeat.com a scanned copy of this agreement in order to be able to get the training and testing data.

Evaluation

Evaluation of task 1 was made in terms of RMSE (Root Mean Squared Error) in the prediction of `log(1+SUM_VISITS_48H)`; participants will be ranked by decreasing RMSE. Evaluation of tasks 2 and 3 wasmade also in terms of RMSE, and participants ranked respectively by decreasing RMSE on `log(1+SUM_TWITTER_48H)` and `log(1+SUM_FACEBOOK_48H)`.

The code that computes the logged RMSE is available: rmse_code.txt

After the competition ends, participants will be able to see the distribution of errors per website in order to gain further insights on their systems.

Paper

Participants will be asked to submit a description of their systems on a paper, for details see workshop »»

~~Submitting your prediction (deadline passed)~~

~~You need to upload to http://hacks.chartbeat.com/ecml/ three files in comma-separated format, containing records of the form <PAGE_ID, SUM_VISITS_48H>, <PAGE_ID, SUM_TWITTER_48H> and <PAGE_ID, SUM_FACEBOOK_48H> for each of the 30,000 URLs in the secret evaluation data, example:~~

```
H000P000,101
H000P001,50
...
```

~~SUM_VISITS_48H is the prediction of the total visits received in the first 48 hours by the webpage (in this case, page H000P000 is predicted by you to acquire 101 visits in its firs 48 hours), SUM_TWITTER_48H is the same for Twitter, and SUM_FACEBOOK_48H the same for Facebook.~~

~~You can submit intermediary predictions. Intermediary predictions will be used to create a ranking that will allow you to see how their system compares with the one of the competitors. The last prediction submitted by each person/team will be the one taken into account for deciding on the winner.~~

License_agreement_ChartbeatEcmlPkdd_dataset.pdf (82k)	Carlos Castillo, Mar 20, 2014, 4:22 AM	v.1	
rmse_code.txt (1k)	Carlos Castillo, Jul 13, 2014, 4:58 AM	v.1	

Comments

Google PhD Awards (2013) - 1/3

Gmail **Flavio Figueiredo <flaviovdf@gmail.com>**

Re: Google Brazil Focused Research Grants Program

1 message

Jussara Almeida <jussara@dcc.ufmg.br> Fri, Jan 4, 2013 at 8:45 PM
To: Brian Zaki <brianz@google.com>
Cc: Berthier Ribeiro-Neto <berthier@google.com>, Flavio Vinicius <flaviovdf@gmail.com>

Excellent! These are great news!

Jussara

Enviado via iPhone

Em 04/01/2013, às 21:08, Brian Zaki <brianz@google.com> escreveu:

> Dear Profa. Jussara,
>
> It is our pleasure to inform you that your research proposal entitled "Popularity Evolution of User Generated Videos: Characterization, Prediction and Applications" has been approved for financial support from Google. We will contact you in the upcoming weeks with instructions on how to proceed for the implementation of this research grant.
>
> Congratulations and Happy New Year!
>
> Sincerely,
> Brian Zaki & Berthier Ribeiro-Neto
>
> --
>
> Brian C. Zaki | Principal Engineering Recruiter, Latin America | brianz@google.com | +55 (31) 2128-6789
>
> ** Learn more about Google's Belo Horizonte Office - English I Portuguese
> ** G+ Hangout on Air: Google Belo Horizonte

Google PhD Awards (2013) - 2/3

REQUERIMENTO DE BOLSA DE PESQUISA

Data de emissão do Documento: 15/02/2013

Dados do Projeto

Pedido: 652065
Projeto: 19571 - GOOGLE/FUNDEP/EPEX/PROGRAMA DE BOLSAS **Subprojeto:** 1 - GOOGLE/FUNDEP/EPEX/PROGRAMA DE BOLSAS **Rubrica / Item Apoiado:** 190: Bolsa De Pesquisa / 1 : Bolsa De Pesquisa - Doutorando.
Analista: VICTOR MORANDI CHAMONE **Suporte:** Ana Rita De Melo **Equipe:** GAP - EPEX

Dados do Bolsista

Tipo Bolsa: BOLSA DE PESQUISA
CPF: 05878473445
Nome: FLAVIO VINICIUS DINIZ DE FIGUEIREDO
Identidade: 2782251
Orgão Expedidor: SSPPB
Data Expedição: 20/06/2000
Profissao: ESTUDANTE
País Nacionalidade: BRASIL

Atividade

Data de início: 25/03/2013
Data de término: 28/02/2014
Carga Horaria Mensal: 80

Local da atividade

Escola/Faculdade: UFMG
Unidade: DCC-DEPARTAMENTO DE CIENCIA DA COMPUTAÇÃO
Departamento: POS-GRADUAÇÃO
Cidade: Belo Horizonte - MG

Descrição da atividade da bolsa:

1) coleta de dados; 2) avaliar se os mercados publicitários podem se beneficiar dos modelos de previsão de popularidade; 3) caracterizar a importância do conteúdo e características sociais, para a evolução da popularidade de vídeos; 4) estudar a previsão de Medidas de popularidade (isto é, acessos) e as tendências.

Endereço

Endereço: RUA Professor Domício Murta
Número: 420 **Comp.:** APTO. 102, BL 2
Bairro: Ouro Preto
Cidade: Belo Horizonte - **UF:** MG
Cep: 31330670
Telefone: 3193904118 **Tel. Alternativo:**
E-Mail: flaviovdf@gmail.com

Vínculo Institucional

Instituição: UFMG
Categoria: Discente
Cargo:
Matricula SIAPE:2008659458
Endereço: AVENIDA Avenida Presidente Antônio Carlos
Número: 6627 **Comp.:**
Bairro: Pampulha
Cidade: Belo Horizonte - MG
Cep: 31270901

Pagamentos (bruto)

Modalidade: Mensal Automático
Número de parcelas: 12
Valor Total: R$36.000,00 (trinta e seis mil reais)
Data de inicio: 25/03/2013
Data de término: 25/02/2014
Parcela: 1 / 12 - **Valor:** 3000 - **Data:** 25/03/2013 **Parcela:** 2 / 12 - **Valor:** 3000 - **Data:** 25/04/2013 **Parcela:** 3 / 12 - **Valor:** 3000 - **Data:** 25/05/2013 **Parcela:** 4 / 12 - **Valor:** 3000 - **Data:** 25/06/2013 **Parcela:** 5 / 12 - **Valor:** 3000 - **Data:** 25/07/2013 **Parcela:** 6 / 12 - **Valor:** 3000 - **Data:** 25/08/2013 **Parcela:** 7 / 12 - **Valor:** 3000 - **Data:** 25/09/2013 **Parcela:** 8 / 12 - **Valor:** 3000 - **Data:** 25/10/2013 **Parcela:** 9 / 12 - **Valor:** 3000 - **Data:** 25/11/2013 **Parcela:** 10 / 12 - **Valor:** 3000 - **Data:** 25/12/2013 **Parcela:** 11 / 12 - **Valor:** 3000 - **Data:**

Google PhD Awards 2013 - 3/3

25/01/2014 **Parcela:** 12 / 12 - **Valor:** 3000 - **Data:** 25/02/2014

Bolsa de Pesquisa

1. Sem Retenção de Imposto de Renda:
Bolsa concedida a servidores da UFMG ou de outras instituições federais de ensino superior ou de pesquisa científica e tecnológica nos termos da Lei nº 8958/94, com vista a auxiliar o beneficiário na realização de estudos ou pesquisa, obedecidas as seguintes condições: (1) que os resultados dessas atividades não revertam em vantagens econômicas para a FUNDEP ou para a instituição financiadora da bolsa, e; (2) que não importem em contraprestação de serviços.

Art. 39, inciso VII do Decreto número 3000/99 - Regulamentação do IR

2. Com Retenção de Imposto de Renda:
Bolsa concedida a servidores da UFMG ou de outras instituições federais de ensino superior ou de pesquisa científica e tecnológica, com vista a auxiliar o beneficiário na realização de estudos ou pesquisa, quando a pesquisa representa contraprestação de serviços ou trouxer vantagens para o financiador (Exemplo: Pesquisa encomentada).

Art. 43, inciso I do Decreto número 3000/99 - Regulamentação do IR

Disposições Gerais

1. O valor supra definido poderá ser alterado, conforme disponibilidade financeira do projeto acadêmico a que a bolsa se vincula.
2. Em atendimento ao decreto 7.423/10 art. 7º § 4º: O limite máximo da remuneração, retribuições e bolsas percebidas pelo docente, em qualquer hipótese não poderá exceder o maior valor recebido pelo funcionalismo público federal, nos termos do art. 37, XI, da constituição. (O valor máximo atual é de R$ 26.723,13).
3. Fica acertado que a presente bolsa não gerará qualquer vínculo empregatício entre a FUNDEP e o BOLSISTA, que desempenhará suas atribuições vinculadas ás atividades de extensão da UFMG, sem subordinação á FUNDEP, consoante estabelece o art. 4º, §, da lei nº 8.958/94.
4. O presente instrumento poderá ser rescindido a qualquer tempo, por iniciativa unilateral de qualquer das partes, mediante comunicação prévia de 10 (dez) dias. Nessa hipótese, caso a Bolsa ainda não tenha sido paga integralmente, a FUNDEP não terá obrigatoriedade de completar o valor restante.
5. A bolsa requerida pressupõe que as atividades do bolsista serão realizadas sem prejuízo das atividades funcionais a que o servidor está sujeito.
6. Fica eleito o foro da comarca de Belo Horizonte para dirimir qualquer litígio decorrente do presente instrumento.

Caracterização do Projeto

DE ACORDO

_____
Chefe do Depto.

_____
Coordenador: VICTOR MORANDI CHAMONE

_____
Bolsista: FLAVIO VINICIUS DINIZ DE FIGUEIREDO

_____
FUNDEP: Ana Rita De Melo

Para uso da Gerência de Pessoal

Recebido em _____/_____/_____ Processado em _____/_____/_____
Por: _____ Por: _____

Av. Antonio Carlos, 6627 - Unidade Administrativa II, 4º andar - Campus Pampulha -31270.901 - Belo Horizonte - MG - Brasil
Telefone: +55 31 3409-4200 - fundep@fundep.ufmg.br - www.fundep.ufmg.br

FUNDEP Espaço do Coordenador

REQUERIMENTO DE BOLSA DE PESQUISA

Google PhD Award 2014 - 1/2

Data de emissão do Documento: 02/04/2014

Dados do Projeto

Pedido: 778699
Projeto: 19571 - GOOGLE/FUNDEP/EPEX/PROGRAMA DE BOLSAS **Subprojeto:** 1 - GOOGLE/FUNDEP/EPEX/PROGRAMA DE BOLSAS **Rubrica / Item Apoiado:** 190: Bolsa De Pesquisa / 1 : Bolsa De Pesquisa - Doutorando.
Analista: VICTOR MORANDI CHAMONE **Suporte:** Ana Rita De Melo **Equipe:** GAP - EPEX

Local da atividade

Escola/Faculdade: UFMG
Unidade: DCC-DEPARTAMENTO DE CIENCIA DA COMPUTAÇÃO
Departamento: POS-GRADUAÇÃO
Cidade: Belo Horizonte - MG

Forma de pagamento

CC - Depósito em Conta Corrente
Banco:001 - BANCO DO BRASIL S/A
Agência: 1591-1
Conta: 8907-9

Descrição da atividade da bolsa:

1) coleta de dados; 2) avaliar se os mercados publicitários podem se beneficiar dos modelos de previsão de popularidade; 3) caracterizar a importância do conteúdo e características sociais, para a evolução da popularidade de vídeos; 4) estudar a previsão de Medidas de popularidade (isto é, acessos) e as tendências.

Dados do Bolsista

Tipo Bolsa: BOLSA DE PESQUISA
CPF: 05878473445
Nome: FLAVIO VINICIUS DINIZ DE FIGUEIREDO
Identidade: 2782251
Orgão Expedidor: SSPPB
Data Expedição: 20/06/2000
Profissao: ESTUDANTE
País Nacionalidade: BRASIL

Endereço

Endereço: RUA Professor Domício Murta
Número: 420 **Comp.:** APTO. 102, BL 2
Bairro: Ouro Preto
Cidade: Belo Horizonte - **UF:** MG
Cep: 31330670
Telefone: 3193904118 **Tel. Alternativo:**
E-Mail: flaviovdf@gmail.com

Vínculo Institucional

Instituição: UFMG
Categoria: Discente
Cargo: ESTUDANTE
Matricula SIAPE: 2008659458
Endereço: AVENIDA Avenida Presidente Antônio Carlos
Número: 6627 **Comp.:**
Bairro: Pampulha
Cidade: Belo Horizonte - MG
Cep: 31270901

Atividade

Data de início: 03/03/2014
Data de término: 28/02/2015
Carga Horaria Mensal: 80

Pagamentos (pagamento imediato)

Modalidade: Mensal Automático
Número de parcelas: 12
Valor Total: R$36.000,00 (trinta e seis mil reais)
Data de inicio: 02/04/2014
Data de término: 02/03/2015
Pagamento imediato - Valor: 3000 **Parcela:** 2 / 12 - **Valor:** 3000 - **Data:** 25/04/2014 **Parcela:** 3 / 12 - **Valor:** 3000 - **Data:** 23/05/2014 **Parcela:** 4 / 12 - **Valor:** 3000 - **Data:** 25/06/2014 **Parcela:** 5 / 12 - **Valor:** 3000 - **Data:** 25/07/2014 **Parcela:** 6 / 12 - **Valor:** 3000 - **Data:** 25/08/2014 **Parcela:** 7 / 12 - **Valor:** 3000 - **Data:** 25/09/2014 **Parcela:** 8 / 12 - **Valor:** 3000 - **Data:** 24/10/2014 **Parcela:** 9 / 12 - **Valor:** 3000 - **Data:** 25/11/2014 **Parcela:** 10 / 12 - **Valor:** 3000 - **Data:** 23/12/2014 **Parcela:** 11 / 12 - **Valor:** 3000 - **Data:** 23/01/2015 **Parcela:** 12 / 12 - **Valor:** 3000 - **Data:** 25/02/2015

Google PhD Award 2014 - 2/2

Bolsa de Pesquisa

1. Sem Retenção de Imposto de Renda:
Bolsa concedida a servidores da UFMG ou de outras instituições federais de ensino superior ou de pesquisa científica e tecnológica nos termos da Lei nº 8958/94, com vista a auxiliar o beneficiário na realização de estudos ou pesquisa, obedecidas as seguintes condições: (1) que os resultados dessas atividades não revertam em vantagens econômicas para a FUNDEP ou para a instituição financiadora da bolsa, e; (2) que não importem em contraprestação de serviços.

Art. 39, inciso VII do Decreto número 3000/99 - Regulamentação do IR

2. Com Retenção de Imposto de Renda:
Bolsa concedida a servidores da UFMG ou de outras instituições federais de ensino superior ou de pesquisa científica e tecnológica, com vista a auxiliar o beneficiário na realização de estudos ou pesquisa, quando a pesquisa representa contraprestação de serviços ou trouxer vantagens para o financiador (Exemplo: Pesquisa encomentada).

Art. 43, inciso I do Decreto número 3000/99 - Regulamentação do IR

Disposições Gerais

1. O valor supra definido poderá ser alterado, conforme disponibilidade financeira do projeto acadêmico a que a bolsa se vincula.
2. Em atendimento ao decreto 7.423/10 art. 7º § 4º: O limite máximo da remuneração, retribuições e bolsas percebidas pelo docente, em qualquer hipótese não poderá exceder o maior valor recebido pelo funcionalismo público federal, nos termos do art. 37, XI, da constituição. (O valor máximo atual é de R$ 28.059,29).
3. Fica acertado que a presente bolsa não gerará qualquer vínculo empregatício entre a FUNDEP e o BOLSISTA, que desempenhará suas atribuições vinculadas ás atividades de extensão da UFMG, sem subordinação á FUNDEP, consoante estabelece o art. 4º, §, da lei nº 8.958/94.
4. O presente instrumento poderá ser rescindido a qualquer tempo, por iniciativa unilateral de qualquer das partes, mediante comunicação prévia de 10 (dez) dias. Nessa hipótese, caso a Bolsa ainda não tenha sido paga integralmente, a FUNDEP não terá obrigatoriedade de completar o valor restante.
5. A bolsa requerida pressupõe que as atividades do bolsista serão realizadas sem prejuízo das atividades funcionais a que o servidor está sujeito.
6. Fica eleito o foro da comarca de Belo Horizonte para dirimir qualquer litígio decorrente do presente instrumento.

Caracterização do Projeto

DE ACORDO

_____
Chefe do Depto.

_____
Coordenador: VICTOR MORANDI CHAMONE

_____
Bolsista: FLAVIO VINICIUS DINIZ DE FIGUEIREDO

_____
FUNDEP: Ana Rita De Melo

Para uso da Gerência de Pessoal

Recebido em _____/_____/_____

Processado em _____/_____/_____

Por: _____

Por: _____

Av. Antonio Carlos, 6627 - Unidade Administrativa II, 4º andar - Campus Pampulha -31270.901 - Belo Horizonte - MG - Brasil
Telefone: +55 31 3409-4200 - fundep@fundep.ufmg.br - www.fundep.ufmg.br

Elap Fellowship - Estudos no Canadá - 1/2

Gmail

Flavio Figueiredo <flaviovdf@gmail.com>

IMPORTANT INFORMATION - Emerging Leaders in the Americas Program (ELAP)

Go Global <go.global@ubc.ca> Thu, Jul 29, 2010 at 2:38 PM
To: "flaviov@dcc.ufmg.br" <flaviov@dcc.ufmg.br>

Dear Mr. Flavio Figueiredo,

Congratulations! The University of British Columbia, on behalf of The Canadian Bureau for International Education (CBIE) and Foreign Affairs and International Trade Canada (DFAIT), is pleased to inform you that you have been awarded a scholarship under the **Emerging Leaders in the Americas Program (ELAP).**

Please follow these steps to ensure an efficient processing of the scholarship:

1. Read and sign the attached Privacy Notice Statement. Scan and email it to <go.global@ubc.ca> **by noon (PST) Friday August 6, 2010.**

2. Keep the copy of the agreement attached - this should be included as proof of funding when applying for the work, visit or study permit.

To Receive Your Scholarship

1. Please email a scanned copy of your visa or letter issued by the Canadian Embassy or High Commission in your country **and** proof of medical coverage in Canada to <go.global@ubc.ca> **as soon as you receive these documents**.

2. Regarding medical insurance, CBIE accepts both international and national proof of medical coverage.

3. CBIE will only issue a cheque to UBC for the amount of the scholarship when the above mentioned documents are received.

4. You will access your scholarship through reimbursement of receipted expenses – airfare, visa costs, medical insurance, airport transportation and other related expenses – with the balance paid out to you via a cheque. Please make sure to **keep all your receipts** with you and to bring them to your supervisor as soon as possible.

5. Please note that the reimbursement process can take up to three weeks. Therefore, we strongly recommend that you are aware of everyday expenses and come prepared to finance your first weeks with personal funds and that you budget accordingly. The following website will provide you with helpful information housing in Vancouver http://www.housing.ubc.ca/other-housing/helpful-links.

6. Please note that CBIE does not require financial reporting.

Elap Fellowship - Estudos Canadá - 2/2

Visa information

For visa information please visit the **Citizenship and Immigration Canada website**: http://www.cic.gc.ca and contact your closest **Canadian Representative Abroad** for more information on the visa process: http://www.cic.gc.ca/english/information/offices/index.asp#international

Congratulations again! If you have any questions about this scholarship, feel free to contact me at the email below. We look forward to welcoming you and wish you luck in your learning at UBC.

Kind regards,

Iris Amuto

go.global@ubc.ca

Privacy Notice Statement - English.pdf
68K

Best Paper Nominee (Candidato) - ACM DL 2014 - 1/2

Programme

Registrations for the Digital Libraries Conference have now closed. Any queries can be directed to the events team (events@city.ac.uk (mailto:events@city.ac.uk)). Please be advised it is unlikely we will be able to process late registrations.

Details of our accepted papers are available from the accepted papers page (http://www.city.ac.uk/digital-libraries-2014/accepted-papers). (http://www.city.ac.uk/__data/assets/excel_doc/0008/228374/Schedule-of-DL-Conference_updated-v2-website-28-aug-14.xlsx)
Download a draft conference timetable. (http://www.city.ac.uk/__data/assets/pdf_file/0003/230727/Final-DL-Schedule-for-website.pdf)

Monday, 8th September: Tutorials and Doctoral Consortium, City University London

Tutorial 3: **From Preserving Data to Preserving Research; Curation of Process and Context** Rudolf Mayer, Stefan Proell, Angela Dappert, Raul Palma, Kevin Page, Daniel Garijo and Paul Gooding Morning Session - Half Day Room: D221 Tutorial 5: **Tools and Techniques for Revisiting Online Scholarly Content** Martin Klein, Herbert Van De Sompel, Michael Nelson and Richard Wincewicz Afternoon Session - Half Day Room: D222 Tutorial 6: **Introduction to Digital Libraries** Edward A. Fox Full Day Session Room: D104

Doctoral Consortium Andy Macfarlane Full Day Session Room: A214 Workshop 5: **Digital Preservation Sustainability on the EU Policy Level** Full Day Session Rooms: Great Hall

Tuesday 9th September: Milton Court, Barbican Centre

8.30am-9am: Welcome Coffee and Registration

9am-9.30am: Opening and Welcome

9.30am-10.45am: Keynote Speaker: **Prof. Dieter Fellner**, Frauenhofer Institute

10.45am-11.15am: Coffee Break

11.15am-12.30pm: Parallel Sessions (2 tracks)

Options 1Track 1: Preservation Strategies (Session 1) Session chair: Herbert Van de Sompel

233 - Chuck Cartledge and Michael Nelson (Best paper nominees). When Should I Make Preservation Copies of Myself?

184 - Catherine Marshall and Frank Shipman. An Argument for Archiving Facebook as a Heterogeneous Personal Store

37 - Michael Day, Ann MacDonald, Akiko Kimura and Maureen Pennock. Implementing Digital Preservation Strategy: Developing Content Collection Profiles at the British Library

Best Paper Nominee (Candidato) - ACM DL 2014 - 2/2

33 - Mat Kelly, Michael Nelson and Michele C. Weigle. The Archival Acid Test: Evaluating Archive Performance on Advanced HTML and JavaScript

Track 2: Recommendation and Indexing (Session 13)Session chair: Ingo Frommholz

88 - Monika Akbar, Clifford Shaffer, Weiguo Fan and Edward Fox. Recommendation Based on Deduced Social Networks in an Educational Digital Library

131 - Tim Gollub, Matthias Hagen, Michael Völske and Benno Stein. Dynamic Taxonomy Composition via Keyqueries

53 - Paul Clough, Arantxa Otegi and Eneko Agirre. Personalized Page Rank for Making Recommendations in Digital Cultural Heritage Collections

216 - Tanja Friedrich and Andreas Oskar Kempf. Making Research Data Findable in Digital Libraries: A Layered Model for User-Oriented Indexing of Survey Data

12.30pm-1.45pm: Lunch

1.45pm-3pm: Parallel Session and Panel discussion (2 tracks)

Options 2*Track 1: Publications impacts (Session 4)*Session chair: Jose Borbinha

108 - Glauber Dias Gonçalves, Flavio Vinicius Diniz de Figueiredo, Marcos Andre Goncalves and Jussara Marques de Almeida (Best student paper nominees). Characterizing Scholar Popularity: A Case Study in the Computer Science Research Community

124 - Thiago H. P. Silva, Mirella M. Moro, Ana Paula C. Silva, Wagner Meira Jr. and Alberto Laender. Community-based Endogamy as an Influence Indicator

113 - Denilson Pereira, Eduardo Silva and Ahmed Esmin. Disambiguating Publication Venue Titles using Association Rules

Track 2: Digital Libraries: Evolving from Collections to Communities? (Panel 1)

Moderators: Deanna Marcum, Karen Calhoun

Panelists: Brian Beaton, Jill Cousins, Herbert Van de Sompel

3pm-3.30pm: Coffee Break

3.30pm-4.45pm: Parallel Sessions (2 tracks)

Options 3*Track 1: Personal DL Design (Session 5)*Session chair: Cathy Marshall

235 - Sandra Trullemans and Beat Signer. From User Needs to Opportunities in Personal Information Management: A Case Study on Organisational Strategies in Cross-Media Information Spaces

90 - Su Inn Park and Frank Shipman. PerCon: A Personal Digital Library for Heterogeneous Data

153 - Annika Hinze, Claire Timpany, Nicholas Vanderschantz and Hayat Alqurashi. Social Information Behaviour in Physical Libraries: Implications for the design of digital libraries

Publicações

EXPLORING THE LATENT STRUCTURE OF COLLABORATIONS IN MUSIC RECORDINGS: A CASE STUDY IN JAZZ

Nazareno Andrade
Universidade Federal de Campina Grande

Flavio Figueiredo
IBM Research - Brazil

ABSTRACT

Music records are largely a byproduct of collaborative efforts. Understanding how musicians collaborate to create records provides a step to understand the social production of music. This work leverages recent methods from trajectory mining to investigate how musicians have collaborated over time to record albums. Our case study analyzes data from the Discogs.com database from the Jazz domain. Our analysis examines how to explore the latent structure of collaboration between leading artists or bands and instrumentists over time. Moreover, we leverage the latent structure of our dataset to perform large-scale quantitative analyses of typical collaboration dynamics in different artist communities.

1. INTRODUCTION

Collaboration is a major component of musical creation. Examining who has collaborated in a record is a common method to understand their style, content, and process of creation. Collaborators leave a mark in the music, and may affect the style of the leading artists themselves. For example, the fact that Miles Davis collaborated with Charlie Parker in the beginning of his career can be seen as an important influence in the development of his style.

Looking at a larger picture, understanding the string of collaborators of a musician over his or her career is also prolific source of information to understand the career itself. Reusing the same example, it is possible to partly describe changes in Miles Davis' style in the 70s by describing how he changed the musicians recording with him. At the same time, similarities in the sequence of collaborators for two artists may denote similarities in the artists themselves. Complementarily, identifying common sequences of leading artists with which different instrumentists have performed also helps understanding how styles and communities of musical creation evolve.

From a quantative standpoint, collaboration patterns have often been studied through the use of methods from graph analysis to large-scale collaboration networks (e.g. [1, 9, 16, 18, 21]). However, although these methods provide valuable insights, they fail to focus on longitudinal views of collaboration. This way, they do not allow for examining patterns in collaboration trajectories.

This work leverages recent methods proposed for mining trajectories in object consumption to study collaborations trajectories among musicians. We use TribeFlow [7], a method recently shown to accurately and expressively discover latent spaces of consumption sequences in the Web domain [7]. Our work explores how this model can also be used to discover latent structures in the trajectories of musicians as they collaborate with the leading artists or bands in records. This exploration is done through a case study with Jazz records. Collaborators and musicians as extracted from the Discogs.com collaborative database of discographic information.

The rest of this paper is organized as follows. In the next section we discuss related work. An overview of the TribeFlow model is described in Section 3. This is followed by a description of our datasets in Section 4. Our main results are discussed in Sections 5 to 7. Section 5 discusses the latent trajectories (collaboration spaces) extracted with TribeFlow. Here, we discuss how the method extracts a semantically meaningful latent representation of our datasets. In Section 6 we compare the collaboration spaces of different artists. Section 7 discusses how artists move between collaboration spaces over time. Finally, in Section 8 we conclude the paper.

2. RELATED WORK

Our cultural products, music being no exception, are strongly tied of our social interactions, and in particular to the dynamics of such interactions. Realizing the importance of understanding networks of interacting collaborators, various research efforts have looked into large-scale creation, dissemination and curation of information by groups of individuals [2,3,5,10–12,15,17,20,21]. Some efforts have also specifically focused on understanding musical recordings as a collaborative effort [1, 8, 9, 16, 18, 19]. Nevertheless, much less attention has been given to the dynamics of collaborations trajectories as we do.

With regards to musical production, very recently Bae et. al. [1] looked into the network properties and community structure of the ArkivMusic [1] database. This database, contains meta-data on classical music records. The authors looked into complex network properties such as power-law distributions and the small world effect [5] that exist

[1] http://www.arkivmusic.com

Attribution: Nazareno Andrade, Flavio Figueiredo. "Exploring The Latent Structure of Collaborations in Music Recordings: A Case Study in Jazz", 17th International Society for Music Information Retrieval Conference, 2016.

MINING ONLINE MUSIC LISTENING TRAJECTORIES

Flavio Figueiredo[1] **Bruno Ribeiro**[2] **Christos Faloutsos**[3]
Nazareno Andrade[4] **Jussara M. Almeida**[5]
[1] IBM Research - Brazil [2] Purdue University [3] Carnegie Mellon University
[4] Universidade Federal de Campina Grande [5] Universidade Federal de Minas Gerais

ABSTRACT

Understanding the listening habits of users is a valuable undertaking for musicology researchers, artists, consumers and online businesses alike. With the rise of Online Music Streaming Services (OMSSs), large amounts of user behavioral data can be exploited for this task. In this paper, we present SWIFT-FLOWS, an approach that models user listening habits in regards to how user attention transitions between artists. SWIFT-FLOWS combines recent advances in trajectory mining, coupled with modulated Markov models as a means to capture both how users switch attention from one artist to another, as well as how users fixate their attention in a single artist over short or large periods of time. We employ SWIFT-FLOWS on OMSSs datasets showing that it provides: (1) semantically meaningful representation of habits; (2) accurately models the attention span of users.

1. INTRODUCTION

Is it possible to create expressive yet succinct representations of individuals' music listening habits? Are there common patterns on how music is listened to across different genres and different artists that have highly different popularity? For a long time such questions have attracted the attention of researchers from different fields. In the fields of psychology and musicology [10, 20, 21], researchers exploit musical preferences to study social and individual identity [20], mood regulation [23], as well as the underlying factors of preferences [21]. Computer scientists are also tackling such questions as they become central to develop music recommender systems [3, 4, 7].

With the rise of Online Music Streaming Services (OMSSs) over the last decade, large datasets of user [1] behavior can be used to shed light on questions like the ones above. More specifically, digital traces of the listening habits of individuals are readily available to researchers.

In this paper, we focus on the online listening habits of users as trajectories [7] (or trails [24]). Given that a user, u, listens to music by switching attention between different artists, a trajectory captures the sequence of artists or songs visited by a user when listening to music. The main contribution of this paper is to present the SWIFT-FLOWS [2] model, a general technique designed to study user trajectories in OMSSs. We tackle several challenges that stem from the complexity of user behavior, such as:

(a) *Asynchronous users with mixed but similar behavior*: Users that consume music from a set of artists will not start their playlists at the same time or listen to songs in the same order.
(b) *Repeated consumption*: Users tend to listen to artists in bursts, more than what one would expect at random in a shuffled playlist.
(c) *Biased Observations & Small Subpopulations*: User behavior datasets are naturally sparse and biased towards more popular artists. Nevertheless, we still want to be able to analyze underrepresented subpopulations of users and artists.

SWIFT-FLOWS effectiveness is evaluated in large datasets, with results showing that SWIFT-FLOWS: (1) captures semantically meaningful representation of artist transitions; (2) accurately models the attention span of users.

2. RELATED WORK

Understanding the listening habits of individuals has attracted interest from different research fields. Among other problems, musicologists and social psychologists have looked into the latent factors that explain musical preferences [20, 21], factors that affect listener experience (e.g., Music itself, Situational Factors and the Listener him/herself) [10], as well as the relationships between musical imagination and human creativity [10].

Regarding the material methods listeners exploit to listen to music, Nowak [16] discussed the social-material relations of music consumption. The authors conclude that even the same user still relies on multiple forms of listening to music (e.g., legal and illegal downloading, streaming services, CDs, etc). These various forms of consumption were also discussed by Bellogin *et al.* [1]. Here, the au-

[1] Since our case study is on Online Music Streaming Services (OMSSs), we use the terms users and listeners interchangeably.

[2] Switch and Fixation Trajectory Flows

Understanding Video-Ad Consumption on YouTube

A Measurement Study on User Behavior, Popularity, and Content Properties

Mariana Arantes[1] Flavio Figueiredo[2] Jussara M. Almeida[1]
[1]Universidade Federal de Minas Gerais - Brazil
[2]IBM Research - Brazil
[1]{mariana.arantes, jussara}@dcc.ufmg.br [2]flaviovdf@br.ibm.com

ABSTRACT

Faced with the challenge of attracting user attention and revenue, social media websites have turned to video advertisements (video-ads). While in traditional media the video-ad market is mostly based on an interaction between content providers and marketers, the use of video-ads in social media has enabled a more complex interaction, that also includes content creator and viewer preferences. To better understand this novel setting, we present the first data-driven analysis of video-ad exhibitions on YouTube.

CCS Concepts

•**Information systems** → **Social advertising;** *Traffic analysis;*

Keywords

YouTube; Video Ads; Popularity; User-Behavior

1. INTRODUCTION

From search engines to social media applications, advertising has become an ubiquitous commerce on the most popular Internet websites. In these websites, users are given access to a wide range of content (e.g., YouTube videos), whereas the content providers that maintain the application exploit user behavior and content data to create online ad-auctions to earn profits [1–5, 8, 11, 12, 14–17].

Social media applications, in particular, allowed for a novel advertisement market. In these applications, any user can take the role of content creator, viewer or marketer. In contrast, in print or television advertising, the creation and selection of advertisement placements is done by select individuals. Taking YouTube as an example, any user can create videos and market them as advertisements to be streamed to other users. This market allows any user to profit from ads, be it as a content producer that receives monetary shares when ads are streamed before their videos, or even as viewers that can gain from well placed ads that meet their preferences.

We here present a measurement study of how video-ads present on the currently most popular video streaming application, YouTube, are consumed by users. Video-ads, i.e., advertisements presented to the user in the form of videos, is rising as one of the most important (in terms of investments and profits) means of online publicity [23–25]. Thus, to better understand the video-ads market on YouTube, in this paper we study three research questions (RQs):

RQ1: How do users consume video-ads? YouTube allows users to skip watching the ad entirely, usually after an initial exhibition period (e.g., 5 seconds), jumping directly to the requested video. In RQ1, we analyze how users consume video-ads, focusing on their "skipping" behavior. Specifically, we characterize user behavior as to whether they tend to consume video-ads in full or skip them, and the fraction of the ad exhibited to the user prior to the skipping. By tackling RQ1, we aim at drawing insights into how users often respond to video-ad exhibitions, and how effective these ads are in terms of drawing and *keeping* user attention, particularly if compared to other forms of online advertising [15, 22].

RQ2: How does video-ad popularity evolve over time? We aim at understanding the properties of video-ad popularity by analyzing the distributions of the *number of views* and *exposure time*. Whereas the former captures the amount of accesses to each video-ad, the latter captures the amount of time that users were exposed to its content. While measuring the effectiveness of ads is a controversial research issue, both number of views and exposure time have been used as proxies of success of ad campaigns [7, 8, 11, 18]. We look into how bursty the popularity evolution of video-ads is, the time it takes for a video-ad to peak in popularity, and the different profiles of video-ad popularity evolution. Thus, in RQ2, we deepen our analyses of the effectiveness of video-ads on YouTube, correlating the profile of popularity evolution followed by a video-ad with its ultimate success (in terms of popularity).

RQ3: What are the relationships (if any) between a video-ad and the video-contents with which it is associated? In ad-auctions, a video-ad is paired with a piece of content (a YouTube video in our case, or simply a video-content) to be displayed to the user. Our aims in RQ3 are twofold. First, we analyze whether more popular video-ads tend to be paired with videos that are also very popular. Secondly, we assess the extent to which video-ads that are paired with more similar content have a tendency to be more effective (popular), thus uncovering evidence of whether contextual advertising [16] increases the effectiveness of video-ads on YouTube.

The main contribution of this study is to provide an in-depth view of different properties of video-ads on YouTube. Our findings offer a *novel*, *broad* and *timely* look into the ecosystem of video advertisements, drawing valuable insights that motivate the design of more cost-effective strategies to make online video-ads potentially more profitable. Such insights should be of interest to content producers, content providers and marketers, who financially benefit from the success of ad campaigns. Our findings are also of interest to YouTube users in general since they are subject to video-ads.

WebSci '16, May 22-25, 2016, Hannover, Germany

DOI: http://dx.doi.org/10.1145/2908131.2908159

TribeFlow: Mining & Predicting User Trajectories

Flavio Figueiredo[1,2], Bruno Ribeiro[4,5], Jussara Almeida[3], Christos Faloutsos[5]
[1]UFCG - Brazil, [2]IBM Research - Brazil, [3]UFMG - Brazil, [4]Purdue University, [5]Carnegie Mellon University
{flaviov,jussara}@dcc.ufmg.br, ribeiro@cs.purdue.edu, christos@cs.cmu.edu

ABSTRACT

Which song will Smith listen to next? Which restaurant will Alice go to tomorrow? Which product will John click next? These applications have in common the prediction of user trajectories that are in a constant state of flux over a hidden network (e.g. website links, geographic location). Moreover, what users are doing now may be unrelated to what they will be doing in an hour from now. Mindful of these challenges we propose TribeFlow, a method designed to cope with the complex challenges of learning personalized predictive models of non-stationary, transient, and time-heterogeneous user trajectories. TribeFlow is a general method that can perform next product recommendation, next song recommendation, next location prediction, and general arbitrary-length user trajectory prediction without domain-specific knowledge. TribeFlow is more accurate and up to $413\times$ faster than top competitors.

Keywords

User Trajectory Recommendation; Latent Environments;

1. INTRODUCTION

Web users are in a constant state of flux in their interactions with products, places, and services. User preferences and the environment that they navigate determine the sequence of items that users visit (links they click, songs they listen, businesses they visit). In this work we refer to the sequence of items visited by a user as the user's trajectory. Both the environment and user preferences affect such trajectories. The underlying navigation environment may change or vary over time: a website updates its design, a suburban user spends a weekend in the city. Similarly, user preferences may also vary or change over time: a user has different music preferences at work and at home, a user prefers ethnic food on weekdays but will hit all pizza places while in Chicago for the weekend.

The above facts result in user trajectories that over multiple time scales can be non-stationary (depend on wall clock times), transient (some visits are never repeated), and time-heterogeneous (user behavior changes over time); please refer to Section 5 for examples. Unfortunately, mining non-stationary, transient, and time-heterogeneous stochastic processes is a challenging task. It would be easier if trajectories were stationary (behavior is independent of wall clock times), ergodic (visits are infinitely repeated), and time-homogeneous (behavior does not change over time).

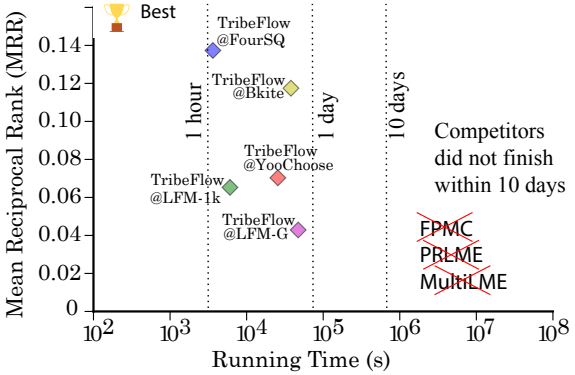


Figure 1: TribeFlow is at least an order of magnitude faster than state-of-the-art methods for next-item predictions.

In this work we propose TribeFlow to tackle the problem of mining and predicting user trajectories. TribeFlow takes as input a set of users and a sequence items they visit (user trajectories), including the timestamps of these visits if available, and outputs a model for personalized next-item prediction (or next $n > 1$ items). TribeFlow can be readily applied to personalized trajectories from next check-in recommendations, to next song recommendations, to product recommendations. TribeFlow is highly parallel and nearly two orders of magnitude faster than the top state-of-the-art competitors. In order to be application-agnostic we ignore application-specific user and item features, including time-of-day effects, but these can be trivially incorporated into TribeFlow.

To illustrate the performance of TribeFlow consider Figure 1, where we seek to compare the Mean Reciprocal Rank (MRR) of TribeFlow over datasets with up to 1.6 million items and 86 million item visits (further details about this dataset is given in Section 4) against that of state-of-the-art methods such as Multi-core Latent Markov Embedding (MultiLME) [40], personalized ranking LME (PRLME) [13], and Context-aware Ranking with Factorizing Personalized Markov Chains [45] (FPMC). Unfortunately, MultiLME, PRLME, and FPMC cannot finish any of these tasks in less than 10 days while for TribeFlow it takes between one and thirteen hours. In significantly sub-sampled versions of the same datasets we find that TribeFlow is at least 23% more accurate than its competitors.

TribeFlow works by decomposing potentially non-stationary, transient, time-heterogeneous user trajectories into *very* short sequences of random walks on latent environments that are stationary, ergodic, and time-homogeneous. An intuitive way to understand TribeFlow

WWW 2016, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4143-1/16/04.
http://dx.doi.org/10.1145/2872427.2883059.

TrendLearner: Early Prediction of Popularity Trends of User Generated Content

Flavio Figueiredo[1,*], Jussara M. Almeida[a], Marcos A. Gonçalves[a], Fabricio Benevenuto[a]

[a]*Department of Computer Science, Universidade Federal de Minas Gerais*
Av. Antônio Carlos 6627, CEP 31270-010, Belo Horizonte - MG, Brazil. Phone: +55 (31) 3409-7541, Fax: +55 (31) 3409-5858

Abstract

Predicting the popularity of user generated content (UGC) is a valuable task to content providers, advertisers, as well as social media researchers. However, it is also a challenging task due to the plethora of factors that affect content popularity in social systems. Here, we focus on the problem of predicting the popularity *trend* of a piece of UGC (object) *as early as possible*. Unlike previous work, we explicitly address the inherent tradeoff between prediction accuracy and remaining interest in the object after prediction, since, to be useful, accurate predictions should be made *before* interest has exhausted. Given the heterogeneity in popularity dynamics across objects, this tradeoff has to be solved on a per-object basis, making the prediction task harder. We tackle this problem with a novel two-step learning approach in which we: (1) extract popularity trends from previously uploaded objects, and then (2) predict trends for newly uploaded content. Our results for YouTube datasets show that our classification effectiveness, captured by F1 scores, is 38% better than the baseline approaches. Moreover, we achieve these results with up to 68% of the views still remaining for 50% or 21% of the videos, depending on the dataset.

Keywords: popularity, trends, classification, social media, ugc, prediction

1. Introduction

The success of Internet applications based on user generated content (UGC)[1] has motivated questions such as: How does content popularity evolve over time? What is the potential popularity a piece of content will achieve after a given time period? How can we predict popularity evolution of a particular piece of UGC? For example, from a system perspective, accurate popularity predictions can be exploited to build more cost-effective content organization and delivery platforms (e.g., caching systems, CDNs). They can also drive the design of better analytic tools, a major segment nowadays [20, 34], while online advertisers may benefit from them to more effectively place contextual advertisements. From a social perspective, understanding issues related to popularity prediction can be used to better understand the human dynamics of consumption. Moreover, being able to predict popularity on an automated way is crucial for marketing campaigns (e.g. created by activists or politicians), which increasingly often use the Web to influence public opinion.

Challenges: However, predicting the popularity of a piece of content, here referred to as an *object*, in a social system is a very challenging task. This is mostly due to the various phenomena affecting the popularity prediction of social media – which were observed on the datasets we use (as well as others) [11, 22, 33] – as well as the diminishing interesting in objects over time, which implies that popularity predictions must be timely to capture user interest and be useful in real work settings. Both challenges can be summarized as follows:

1. Due to the easiness with which UGC can be created, many factors can affect an object's popularity. Such factors include, for instance, the object's content, the social context in which it is inserted (e.g., social neighborhood or influence zone of the object's creator), the mechanisms used to access the content (e.g., searching, recommendation, toplists), or even an external factor, such as a hyperlink to the content in a popular blog or website. These factors can cause spikes in the surge of interest in objects, as well as information propagation cascades which affect the popularity trends of objects.
2. To be useful in a real scenario, a popularity prediction approach must identify popularity trends *before the user interest in the object has severely diminished.* To illustrate this point, Figure 1 shows the popularity evolution of two YouTube videos: the video on the left receives more than 80% (shaded region) of all views received during its lifespan in the first 300 days since upload, whereas the other video receives only about half of its total views in the same time frame. If we were to monitor each video for 300 days, most potential views of the first video would be lost. In other words, not all objects require the same monitoring period, as assumed by previous work, to produce accurate predictions: for some objects, the prediction can be made earlier. Thus, the tradeoff should be solved on a *per-object* basis, which implies that determining the duration of the monitoring period that leads to a good solution of the tradeoff for *each object* is part of the problem.

*Corresponding author
Email addresses: flaviov@dcc.ufmg.br (Flavio Figueiredo), jussara@dcc.ufmg.br (Jussara M. Almeida), mgoncalv@dcc.ufmg.br (Marcos A. Gonçalves), fabricio@dcc.ufmg.br (Fabricio Benevenuto)

[1] YouTube, Flickr, Twitter, and so forth

Building a Question-Answering Corpus using Social Media and News Articles

Paulo Cavalin, Flavio Figueiredo, Maíra de Bayser,
Luis Moyano, Heloisa Candello, Ana Appel, and Renan Souza

IBM Research - Brazil

Abstract. Is it possible to develop a reliable QA-Corpus using social media data? What are the challenges faced when attempting such a task? In this paper, we discuss these questions and present our findings when developing a QA-Corpus on the topic of Brazilian finance. In order to populate our corpus, we relied on opinions from experts on Brazilian finance that are active on the Twitter application. From these experts, we extracted information from news websites that are used as answers in the corpus. Moreover, to effectively provide rankings of answers to questions, we employ novel word vector based similarity measures between short sentences (that accounts for both questions and Tweets). We validated our methods on a recently released dataset of similarity between short Portuguese sentences. Finally, we also discuss the effectiveness of our approach when used to rank answers to questions from real users.

Keywords: Question and Answer, Social Media, Finance

1 Introduction

The avalability of corpora to drive and sustain Question-Answering (QA) systems [8] is of fundamental importance. Such corpora are generally obtained from various sources, normally large collections of text, such as online news [1], or Wikipedia [6]. Some authors have put forward the advantages of using social media in the construction of certain types of corpora, e.g., in [3], the authors propose building comparable corpora from social networks, in particular, Twitter. In the same line, the authors of [5] propose using Twitter as an alternative in short sentences settings. Finally, authors in [7] address QA in social media mainly as a characterization effort on the type and frequency of questions and answers that may be found in Twitter, even though they don't address the construction of a corpus nor the targeting of any specific domain.

Motivated by the above setting, we study the potential of using social media data to create a QA corpus for specific fields. We refer to the corpus simply as QA-Corpus. In details, we study the viability of how can social media information be explored to create a QA-Corpus on the topic of Brazilian finance. To our knowledge, this is the first work to study the possibility of building a domain-specific corpus from social media. We believe that social media data can provide a proxy to field experts who can provide timely, possibly spam free

Uma Caracterização dos Padrões de Navegação de Usuários em uma Aplicação Social de Streaming de Vídeo

Mariana Arantes, Flavio Figueiredo, Jussara M. Almeida

[1]Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

{mariana.arantes, flaviov, jussara}@dcc.ufmg.br

Abstract. *In this paper, we present a characterization of user behavior in video streaming applications. Our work is focused on providing insights to the following questions: (1) What are the most common websites that lead users to video content in social media video applications? (2) What is the browsing behavior of users within the application until they leave? (3) How exposed are users to ads in these applications? Using user browsing data from a large university campus in Brazil, we focus our case-study application on YouTube, the largest video application online. Different from previous studies, we characterize the browsing behavior of individual users. Moreover, while tackling the third question, we also characterize user exposure to video ads, a new kind of advertisement that was mostly unaccounted for by previous work. Our main results show that: (1) users usually reach content using different links depending on the kind of content (e.g., comedy videos vs news videos); (2) users tend to use search engines and related video lists to follow up on new content after viewing a video; and (3) video ads account for a much higher exposure to ad content than link ads.*

Resumo. *Neste trabalho, é apresentada uma caracterização do comportamento de usuários em aplicações de mídia social de streaming de vídeo online. A caracterização é feita com o objetivo de responder a três perguntas motivadoras: (1) Quais fontes externas (websites) mais frequentemente levam usuários para vídeos? (2) Como é o comportamento de navegação de usuários dentro da aplicação de streaming de vídeo? (3) Quão expostos os usuários estão a diferentes tipos de propaganda em tais aplicações? Usando uma base de dados de navegação de usuários de um grande campus universitário brasileiro, estudou-se o comportamento de usuários no YouTube, a maior aplicação de streaming de vídeo atualmente. Diferente de estudos passados, neste trabalho é caracterizado o comportamento individual de usuários na aplicação. Além disso, o acesso aos dados possibilitou a análise do comportamento de usuários quando expostos a um novo tipo de propaganda online, as propagandas em formato de vídeo. Os principais resultados mostram que: (1) os links que mais frequentemente levam usuários para vídeos do YouTube variam dependendo da categoria do vídeo, (2) após visualizarem um vídeo, usuários tendem a usar máquinas de busca e listas de vídeos relacionados para continuarem navegando na aplicação, e (3) propagandas no formato de vídeo tendem a atrair maior atenção dos usuários do que propagandas tradicionais em links.*

Esta pesquisa é financiada pelo INCTWeb (MCT/CNPq 573871/2008-6), CNPq, CAPES e FAPEMIG.

NetSciX 2015 - Sem Qualis

Modeling and Mining Information Popularity Online

Flavio Figueiredo Jussara M. Almeida
Computer Science Department - Universidade Federal de Minas Gerais, Brazil
{flaviov, jussara}@dcc.ufmg.br

1. INTRODUCTION

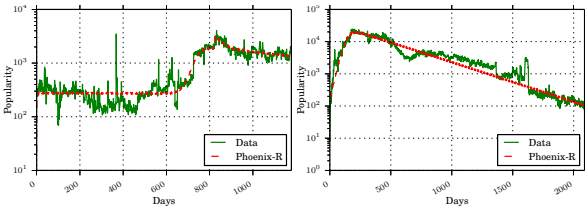
Nowadays, there is an unprecedented amount of user generated content being produced online. This fact is one of the driving forces of what is know as the *social media* phenomenon. Social media shifted how information is produced and propagated. While in traditional media select individuals are responsible for the production, curation and propagation of information, in the social media setting anyone can produce and share information online. One major question in this setting is: *What drives the popularity of information in social media?* This is an interesting question since even with the overload of information that accompanies this mass production of content, some pieces of information manage to attract the attention of millions of users, while the majority remain obscure.

One example of the complexity behind social media popularity is the YouTube channel of *Henri, le Chat-Noir*[1]. The first video of Henri was uploaded in 2007 and remained in obscurity for years. However, in 2012 a user of the Tumblr social network found the video and posted online[2]. Currently, the video and channels has millions of visits from a wide range of different sources (e.g., OSNs, search engines, word-of-mouth and so forth). We can use this single example to motivate our research. Important questions that we raised and approached were as follows:

What is the impact of incoming links on the popularity of online information? There are multiple forms through which users can reach content and, thus, there are multiple driving forces that may impact the popularity of information. Identifying such forces is crucial for designing more cost-effective content dissemination strategies. For instance, should a content creator invest time on perfecting the keywords describing content (for better search rankings) or focus on campaigning videos in OSNs? Our current results show that search engines and social propagation inside a service (say YouTube) are major factors in driving popularity [4].

How does information popularity evolve over time? Here, we aim at answering if there are different patterns which capture the major trends in which information popularity evolves over time. In a birds-eye-view, we found that there exists a combination of two trends governing the popularity of information. One trend consists of contents that tend to remain attractive over time with an always increasing or steady-state popularity [4]. The other, accounts for content that tend to peak in popularity for a short while, with three different popularity decay characteristics after the peak. Examples of both trends are shown in Figure 1.

How do users perceive the quality of popular and unpopular information? Most research in online popularity neglect the users perception of the information being disseminated. We studied if users tend to like very popular content or dislike very unpopular content. Based on a user based study on Mechanical Turk and selected videos from YouTube, we showed that, while users perceptions of content are highly subjective, when the majority of users like a certain YouTube video that video tends to be popular. This was interesting since it shows that there is more than social propagation to popularity. Based on our results [2] we hypothesize that videos like the Henri example would likely be less popular (regardless of OSN propagation) if their content did not appeal to users.



(a) Rock Song (b) User Dancing Video

Figure 1: Different YouTube videos as captured by our model.

Can we model and predict the future popularity of information? Two of our most recent results showed that we can model [5] the popularity of information over time and predict the future popularity [1]. Based on our previous findings that we discussed, we developed the Phoenix-R model which can capture the long term popularity evolution as showed in Figure 1. Also, we combined social network propagation and early view patterns of news media to develop prediction models with the user of machine learning tools [1]. These two results show the applicability of our results to mining tasks such as popularity prediction.

With these questions we summarize some of the work on information popularity online that we are pursuing. Our results show that this a promising and new area of research. Currently, we are working on optimizing how early can we predict popularity [3] and practical applications for models and predictions (e.g., search engine rankings or advertising).

[1] `http://www.youtube.com/user/HenriLeChatNoir`
[2] `http://knowyourmeme.com/memes/henri-le-chat-noir`

2. REFERENCES

[1] F. Figueiredo, J. Almeida, Y. Matsubara, B. Ribeiro, and C. Faloutsos. Revisit Behavior in Social Media: The Phoenix-R Model and Discoveries. In *Proc. ECML/PKDD*, 2014.

[2] F. Figueiredo, F. Benevenuto, J. Almeida, and K. Gummadi. Does Content Determine Information Popularity in Social Media? A Case Study of YouTube Videos' and their Popularity. In *Proc. CHI.*, 2014.

[3] F. Figueiredo, F. Benevenuto, M. Gonçalves, and J. Almeida. TrendLearner: Early Prediction of Popularity Trends of User Generated Content. *Online at: http://arxiv.org/abs/1402.2351*, 2014.

[4] F. Figueiredo, F. Benevenuto, M. Gonçalves, and J. Almeida. On the Dynamics of Social Media Popularity: A YouTube Case Study. *ACM Transactions on Internet Technology*, To appear.

[5] F. Figueiredo, M. Gonçalves, and J. Almeida. Improving the Effectiveness of Content Popularity Prediction Methods using Time Series Trends. In *ECML/PKDD Discovery Challenge*, 2014.

Improving the Effectiveness of Content Popularity Prediction Methods using Time Series Trends

Flavio Figueiredo Marcos Gonçalves Jussara M. Almeida
Computer Science Department - Universidade Federal de Minas Gerais, Brazil
{flaviov, mgoncalv, jussara}@dcc.ufmg.br

ABSTRACT

We here present a simple and effective model to predict the popularity of web content. Our solution, which is the winner of two of the three tasks of the ECML/PKDD 2014 Predictive Analytics Challenge, aims at predicting user engagement metrics, such as number of visits and social network engagement, that a web page will achieve 48 hours after its upload, using only information available in the first hour after upload. Our model is based on two steps. We first use time series clustering techniques to extract common temporal trends of content popularity. Next, we use linear regression models, exploiting as predictors both content features (e.g., numbers of visits and mentions on online social networks) and metrics that capture the distance between the popularity time series to the trends extracted in the first step. We discuss why this model is effective and show its gains over state of the art alternatives.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

General Terms

Algorithms; Measurement

Keywords

analytics; predictive; challenge; web page; popularity; host

1. INTRODUCTION

With the ever-growing production of online content, characterizing and predicting user engagement (e.g., number of visits or social engagement such as Facebook likes) on web content may have multiple beneficial values such as: (1) understanding the human dynamics of information consumption; (2) supporting the decisions of content producers and providers on different tasks (e.g., marketing and content filtering); and, (3) understanding the physical processes that govern the growth of viewership on the Web. Several previous studies [2–4, 11] have characterized some of the factors that cause the popularity growth of different kinds of web content. Complementarily, various others [1, 6, 7, 11] have focused on the the task of popularity prediction. We focus here on the latter task, aiming at predicting the popularity of a piece of content.

Popularity prediction is a difficult and important task since it mostly translates into income and profits for content providers, creators and consumers alike. For example, more visitors to a web page may lead to more ad-clicks and sales. Moreover, content provisioning to a large amount of users may require decisions such as geographical sharding of content to servers (due to the increased traffic). Thus, if planning is not performed correctly, longer latencies and loading times, and thus, fewer users may be expected. Finally, accurate and early predictions can lead to better services to the end consumer, such as search engine rankings [8].

We here present a simple, and yet effective, model for predicting the popularity of online content. More specifically, we present the winning model of two of the three tasks of the ECML/PKDD 2014 Predictive Analytics Challenge. In the challenge, different features related to the popularity of 30,000 web pages from 100 different hosts were provided. The goal of the challenge is to predict the popularity of 30,000 other pages from the same 100 hosts 48 hours after their upload. The features provided for the task were measured in the first hour after upload for each page.

Our model exploits the temporal features related to web pages (e.g., past visits and social engagement), as well as typical popularity (i.e., number of visits) time series trends which exist in the dataset. Such trends are extracted via unsupervised learning methods. Specifically, it combines the temporal features with features that capture the distances between the popularity time series for each web page and the extracted trends. We present a data characterization that motivates the design of our solution, and show the gains in prediction accuracy (ranging from 15% to 27%) when it is compared to state of the art alternatives.

The rest of this paper is organized as follows. We formally describe the prediction problem and present the state of the art baseline methods in Section 2. In Section 3 we introduce our proposed solution, whereas our experiments and results are presented in Section 4. Finally, Section 5 concludes the paper.

2. BACKGROUND

We start this section by defining the content popularity prediction problem (Section 2.1). In this definition, as throughout the rest of the paper, we refer to a particular piece of content as a web page[1]. Next, we discuss existing state of the art solutions used as baselines in our experimental study (Section 2.2).

[1]We here focus on web page popularity prediction, given the goal of the ECML/PKDD Challenge. However, our models are general and can be applied to other types of online content.

ECML/PKDD Discovery Challenge on Predictive Analytics
Nancy, France, September 2014.

Does Content Determine Information Popularity in Social Media?

A Case Study of YouTube Videos' Content and their Popularity

Flavio Figueiredo[1], **Jussara M. Almeida**[1], **Fabrício Benevenuto**[1]
[1]Department of Computer Science, UFMG, Brazil
{flaviov,jussara,fabricio}@dcc.ufmg.br

Krishna P. Gummadi[2]
[2]MPI-SWS, Germany
gummadi@mpi-sws.org

ABSTRACT

We here investigate *what drives the popularity of information on social media platforms*. Focusing on YouTube, we seek to understand the extent to which content by itself determines a video's popularity. Using mechanical turk as experimental platform, we asked users to evaluate pairs of videos, and compared users' relative perception of the videos' content against their relative popularity reported by YouTube. We found that in most evaluations users could not reach consensus on which video had better content as their perceptions tend to be very subjective. Nevertheless, when consensus was reached, the video with preferred content almost always achieved greater popularity on YouTube, highlighting the importance of content in driving information popularity on social media.

Author Keywords

Content popularity; social media; user study

ACM Classification Keywords

H.5.4 Hypertext/Hypermedia: User issues.

INTRODUCTION

What drives the popularity of information in social media? Recently, this question has attracted a lot of research attention as social media sites become increasingly popular. An unresolved part of this question is about the relative roles of two primary forces that drive the popularity of a piece of information: (i) its content, i.e., the interestingness, topicality, or quality of the information *as perceived by users*, and (ii) its dissemination mechanisms, such as propagation by word-of-mouth, blogs or mass media channels. It stands to reason that both factors matter, but the extent to which they impact the popularity of a piece of information remains an open issue.

Many previous studies on how information becomes popular in social media sites focused on dissemination related factors (e.g, social influence, mechanisms that expose content to users, time of upload) [2, 4, 7, 9, 10], ignoring the role of the content itself. Other efforts, instead, analyzed social media content focusing on data mining tasks such as popularity prediction [11] and video classification [5], analyzing popularity differences in content duplicates [2], and exploring content importance as parameter of popularity evolution models [8]. In this paper we take a different and complementary approach, focusing on understanding the extent to which content matters for popularity of videos on YouTube.

Our methodology attempts to assess *users' relative perceptions* of the contents of pairs of videos through user surveys conducted over Amazon mechanical turk. Users in our experiments are exposed only to the video content, and are not subjected to other factors (inherent to the YouTube site) that may impact their perceptions of content (e.g., user comments, social links, appearance of content in external sites). Specifically, we present to users pairs of videos from the same major topic and uploaded around the same date, and ask them to choose which one: (1) *they enjoyed more*, (2) *they would be more willing to share with friends*, and (3) *they predicted would become more popular on YouTube*. These questions target the user's individual perception of content interestingness and of the interests of her social circle (and thus the chance of the content spreading through it), as well as the user's expectations on a global scale. Our goals are to assess, for each of these questions, whether users reach consensus, and, when there is consensus, whether user perceptions match the relative popularity achieved by the videos on YouTube.

We find that users could not reach consensus in many evaluations, even when the popularity (on YouTube) of the evaluated videos differs by orders of magnitude. The lack of consensus is more striking for sharing and liking choices. It also depends on the video topic. This suggests that users' perceptions about content are quite subjective and that content may not be the most important factor that drives popularity in many cases. However, whenever participants reached consensus, their choices mostly match the video with largest popularity on YouTube, suggesting that, in these cases, content has a significant impact and predictive power on video popularity.

The goals of our study complement previous work. In particular, Salganik *et al.* [9] also relied on a user study to understand popularity dynamics. However, they focused on the impact of social influence on popularity, whereas we focus on the role of content and rely on users to evaluate the content in a setup that is isolated (to the extent possible) from dissemination mechanisms that might influence popularity. To our knowledge, the human perceptions of content and how they correlate to popularity in a social media site have not been

Acknowledgments: Research supported by InWeb - Institute of Science and Technology for Web Research and by individual grants from CNPq, Capes and Fapemig.

CHI'14, April 26–May 1, 2014, Toronto, Canada.

http://dx.doi.org/10.1145/2556288.2557285

Revisit Behavior in Social Media: The Phoenix-R Model and Discoveries

Flavio Figueiredo[1], Jussara M. Almeida[1],
Yasuko Matsubara[2,3], Bruno Ribeiro[3], and Christos Faloutsos[3]

[1] Department of Computer Science, Universidade Federal de Minas Gerais
[2] Department of Computer Science, Kumamoto University
[3] Department of Computer Science, Carnegie Mellon University

Abstract. How many listens will an artist receive on a online radio? How about plays on a YouTube video? How many of these visits are new or returning users? Modeling and mining popularity dynamics of social activity has important implications for researchers, content creators and providers. We here investigate the effect of revisits (successive visits from a single user) on content popularity. Using four datasets of social activity, with up to tens of millions media objects (e.g., YouTube videos, Twitter hashtags or LastFM artists), we show the effect of revisits in the popularity evolution of such objects. Secondly, we propose the PHOENIX-R model which captures the popularity dynamics of individual objects. PHOENIX-R has the desired properties of being: (1) parsimonious, being based on the minimum description length principle, and achieving lower root mean squared error than state-of-the-art baselines; (2) applicable, the model is effective for predicting future popularity values of objects.

1 Introduction

How do we quantify the popularity of a piece of content in social media applications? Should we consider only the audience (unique visitors) or include revisits as well? Can the revisit activity be explored to create more realistic popularity evolution models? These are important questions in the study of social media popularity. In this paper, we take the first step towards answering them based on four large traces of user activity collected from different social media applications: Twitter, LastFM, and YouTube[4].

Understanding the popularity dynamics of online content is both a challenging task, due to the vast amount and variability of content available, as it can also provide invaluable insights into the behaviors of human consumption [6] and into more effective engineering strategies for online services. A large body of previous work has investigated the popularity dynamics of social media content, focusing mostly on modeling and predicting the *total number of accesses* a piece of content receives [5, 6, 9, 17, 21].

However, a key aspect that has not been explored by most previous work is the effect of revisits on content. The distinction between audience (unique users), revisits (returning users), and popularity (the sum of the previous two) can have large implications for different stakeholders of these applications - from content providers to content

[4] `http://twitter.com` `http://lastfm.com` `http://youtube.com`

On the Dynamics of Social Media Popularity: A YouTube Case Study

FLAVIO FIGUEIREDO, JUSSARA M. ALMEIDA, MARCOS ANDRÉ GONÇALVES and FABRÍCIO BENEVENUTO, Universidade Federal de Minas Gerais, Brazil

Understanding the factors that impact the popularity dynamics of social media can drive the design of effective information services, besides providing valuable insights to content generators and online advertisers. Taking YouTube as case study, we analyze how video popularity evolves since upload, extracting popularity trends that characterize groups of videos. We also analyze the referrers that lead users to videos, correlating them, features of the video and early popularity measures with the popularity trend and total observed popularity the video will experience. Our findings provide fundamental knowledge about popularity dynamics and its implications for services such as advertising and search.

Categories and Subject Descriptors: H.1.2 [**User/Machine Systems**]: Human Factors

General Terms: Measurement, Human Factors

Additional Key Words and Phrases: youtube, social media, characterization, referrers, popularity growth

ACM Reference Format:
Flavio Figueiredo, Jussara M. Almeida, Marcos André Gonçalves, Fabricio Benevenuto, 2014. On the Dynamics of Social Media Popularity: A YouTube Case Study *ACM Trans. Internet Technol.* 1, 1, Article 01 (October 2014), 22 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

User generated content (UGC) has emerged as the predominant form of online information sharing nowadays. The unprecedented amount of information being produced is one of the driving forces behind the success of the social media phenomenon [Kaplan and Haenlein 2010; Cormode and Krishnamurthy 2008]. This phenomenon is a shift from the traditional media where, instead of content being produced mostly by a few selected individuals, anyone, in theory, can produce and share content online. However, the "information overload" that accompanies the huge amount of social media being produced has its drawbacks. For example, it is ever-so-difficult to find and filter relevant content to oneself. Nevertheless, some pieces of content (or *objects*) succeed in attracting the attention of millions of users, while most remain obscure. This leads to the heavy tailed characteristic of content popularity [Sinha and Pan 2007; Clauset et al. 2009], where a few objects become very popular while most of them attract only a handful of views. *What makes one particular object become hugely popular while the majority receive very little attention? Which factors affect how the popularity of an object will evolve over time?* These are major questions in the social media context that drive our present work.

This research is partially funded by the Brazilian National Institute of Science and Technology for Web Research (MCT/CNPq/INCT Web Grant Number 573871/2008-6), and by the authors' individual grants from Google, CNPq, CAPES and Fapemig. We also thank Caetano Traina, Renato Assunção, Virgilio Almeida, Elizeu Santos-Neto, Matei Ripeanu, and the anonymous reviewers for discussions on drafts of this work.

DOI: http://dx.doi.org/10.1145/0000000.0000000

Characterizing Scholar Popularity: A Case Study in the Computer Science Research Community

Glauber D. Gonçalves, Flavio Figueiredo, Jussara M. Almeida, Marcos A. Gonçalves
Department of Computer Science, Universidade Federal de Minas Gerais, Brazil
{ggoncalves,flaviov,jussara,mgoncalv}@dcc.ufmg.br

ABSTRACT

A common live debate among scholars regards the popularity, productivity and impact of research. This paper aims to contribute to such discussion by quantifying the impact of various academic features on a scholar popularity throughout her career. Using a list of over 2 million publications in the Computer Science research area obtained from two large digital libraries, we analyze how features that capture the number and rate of publications, number and quality of publication venues, and the importance of the scholar in the co-authorship network relate to the scholar popularity. We also investigate the temporal dynamics of scholar popularity, identifying a few common profiles, and characterizing scholars in each profile according to their academic features.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

Citation analysis, scholar popularity dynamics, academic scholar features

1. INTRODUCTION

What factors contribute to a successful career in research? This is a fundamental and broad question that draws the attention of all scholars. Success in research can be assessed in terms of various measures. The acknowledgement by peers of the value of a researcher's publications (e.g., by citing them) is one of the most sought-after measures of scholarly success, as it can be seen as an estimate of her influence and visibility in the community, and ultimately of her scholarly popularity [10].

Investigating the factors that impact the popularity of a scholar can shed light into proactive actions that might guide decisions to shape a career in research. Moreover, from a system standpoint, it can draw useful insights into the design of cost-effective popularity prediction methods, which, in turn, can be exploited for improving various services (e.g. expert or academic collaboration recommendation services [4]). Some questions of interest are: *to which extent do the quantity and the quality of the publication venues impact a scholar popularity? To which extent is the number of publications related to popularity? What is the role that the co-authorship network plays in the scholar popularity?*

Influence, productivity and popularity in research has already been tackled in various prior studies since the early 20^{th} century [6, 25]. For example, many authors have characterized influential publications [22, 32], some with the goal of designing models to predict the number of times a particular piece of work will be cited [5, 7, 21, 34]. Others have analyzed various impact factors to evaluate the influence of publication venues (e.g., journals) [3, 12] and researchers [15, 11, 29] from citations.

In this paper, our goal is to investigate and *quantify* the factors that impact the popularity of a scholar during her career. Complementing prior analyses [31, 10, 34], we intend to assess the impact of various academic characteristics, or *features*, on the popularity a scholar achieves during the career. We also want to study the temporal popularity dynamics of various scholars, identifying a few common profiles, and characterizing scholars in each profile according to their popularity and academic features.

As in [27, 13, 31], we use the *total number of citations* to estimate the popularity of a scholar. Although other indices, such as weighted metrics [32, 10] and PageRank [3, 33], could be adopted, our choice is based on two factors. Firstly, some authors [27] have argued that citation counts are better indicators of the scientific contribution of researchers, disciplines or nations than impact factors (such as the h-index [15]). Secondly, considering qualitative aspects of the research is very subjective and would require the use of metrics with debatable biases and criticisms [14, 12, 18]. Thus, we assess scholar popularity by the total number of citations, leaving the analysis of other indices to future work.

Specifically, we focus on scholars of a specific research field - Computer Science, and crawl statistics about their publication records from two large digital libraries, namely ArnetMiner[1] and Microsoft Academic[2]. We start by studying how various academic features[3] are correlated with the popularity of a scholar. The following features are analyzed: total number of publications, yearly rate

[1]http://arnetminer.org
[2]http://academic.research.microsoft.com
[3]Features that reflect the academic activity of a scholar.

JCDL'14, September 8–12, 2014, London, England, UK.
978-1-4799-5569-5/14/$31.00 ©2014 IEEE.

On the Prediction of Popularity of Trends and Hits for User Generated Videos

Flavio Figueiredo
flaviov@dcc.ufmg.br
Computer Science Department
Universidade Federal de Minas Gerais

ABSTRACT

User generated content (UGC) has emerged as the predominant form of media publishing on the Web 2.0. Motivated by the large adoption of video sharing on the Web 2.0, the objective of our work is to understand and predict popularity trends (e.g, will a video be viral?) and hits (e.g, how may views will a video receive?) of user generated videos. Such knowledge is paramount to the effective design of various services including content distribution and advertising. Thus, in this paper we formalize the problem of predicting trends and hits in user generated videos. Also, we describe our research methodology on approaching this problem. To the best of knowledge, our work is novel in focusing on the problem of predicting popularity trends complementary to hits. Moreover, we intend on evaluating efficacy of our results not only based on common statistical error metrics, but also on the possible online advertising revenues our predictions can generate. After describing our proposal, we here summarize our latest findings regarding (1) uncovering common popularity trends; (2) measuring associations between UGC features and popularity trends; and (3) assessing the effectiveness of models for predicting popularity trends.

Categories and Subject Descriptors

C.4 [**Computer Systems Organization**]: Performance of Systems—*Measurement techniques*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

Keywords

UGC; video; popularity; trends

General Terms

Human Factors; Design; Measurement

1. INTRODUCTION

On the Web 2.0, user generated content (UGC) has become the de-facto form of media publishing on some of the most popular Internet applications nowadays [6]. Focusing on video content, websites such as YouTube [1] receive over 800 million unique users monthly, attracting over 1 million different advertisers [23]. Even niche applications, such as Vimeo [2], which focuses on independent filmmakers, manage to attract over 70 million unique users monthly [19].

Given the success of such applications and the current large volume of videos consumed daily, understanding how users find such content and how content popularity evolves provide valuable insights for content generators, online advertisers and Internet service providers (ISPs), amongst others. For instance, from a systems perspective, understanding these properties may drive the design of better analytic tools, a major market segment nowadays. Online advertisers may also benefit from this information to better place contextual advertisements, while ISPs could exploit it to develop more cost-effective content delivery platforms and caching systems. From a social perspective, understanding the properties of video popularity trends could be used to better comprehend the human dynamics of consumption processes [7]. Also, content producers could use insights on how user collaboration and collaborative social activities on Web 2.0 applications may impact content popularity, providing information on aspects related to their own fame on video sharing applications.

Most previous efforts, which are focused on predicting the popularity of a piece of content measured at a specific future date [16–18, 22], are still preliminary, as they provide limited knowledge on which features and system mechanisms (e.g., search, related videos, etc) contribute the most to popularity growth. Analyzing the importance of such features to popularity growth is key to provide scalable alternatives to service design, as solutions based on content analysis are less scalable in (user generated) videos. Moreover, there is little effort towards predicting popularity evolution (or trends), which may also provide valuable knowledge. For instance, online advertisers and content delivery systems could benefit more from predicting not only a final popularity measure for UGC, but also whether its popularity trend is increasing and how stable it is likely to be over time.

In sum, our proposed research aims at understanding the importance and utility of various features, particularly referrers (i.e, incoming links to videos), on the popularity evolution of individual user generated videos and exploiting them to develop methods to predict future popularity measures and trends of those videos.

The rest of this paper is organized as follows. Section 2 describes our problem statement and research goals. We describe our current methodology on addressing our goals on Section 3. The current state of our research is described in Section 4 while our related work is addressed in Section 5. Section 6 concludes this paper.

[1] `http://youtube.com`
[2] `http://vimeo.com`

WSDM'13, February 4–8, 2013, Rome, Italy.

ARTICLE IN PRESS

Information Processing and Management xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Assessing the quality of textual features in social media

Flavio Figueiredo [a], Henrique Pinto [a], Fabiano Belém [a], Jussara Almeida [a], Marcos Gonçalves [a,*], David Fernandes [b], Edleno Moura [b]

[a] Universidade Federal de Minas Gerais, Department of Computer Science, Belo Horizonte, MG, Brazil
[b] Universidade Federal do Amazonas, Department of Computer Science, Manaus, AM, Brazil

ARTICLE INFO

Article history:
Received 22 March 2011
Received in revised form 12 March 2012
Accepted 14 March 2012
Available online xxxx

Keywords:
Web 2.0
Textual features
Information quality
Social media

ABSTRACT

Social media is increasingly becoming a significant fraction of the content retrieved daily by Web users. However, the potential lack of quality of user generated content poses a challenge to information retrieval services, which rely mostly on textual features generated by users (particularly tags) commonly associated with the multimedia objects. This paper presents what, to the best of our knowledge, is currently the most comprehensive study of the relative quality of textual features in social media. We analyze four different features, namely, TITLE, TAGS, DESCRIPTION and COMMENTS posted by users, in four popular applications, namely, YouTube, Yahoo! Video, LastFM and CiteULike. Our study is based on an extensive characterization of data crawled from the four applications with respect to usage, amount and semantics of content, descriptive and discriminative power as well as content and information diversity across features. It also includes a series of object classification and tag recommendation experiments as case studies of two important information retrieval tasks, aiming at analyzing how these tasks are affected by the quality of the textual features. Classification and recommendation effectiveness is analyzed in light of our characterization results. Our findings provide valuable insights for future research and design of Web 2.0 applications and services.

1. Introduction

The advent and rapid growth of a variety of Web 2.0 applications, enabling and fostering the establishment of online communities and social networks, have contributed to the creation and dissemination of a massive amount of social media content. Social media refers to content created and disseminated via social interactions, and is thus typically associated with *user generated content*. In general, social media is increasingly becoming a significant fraction of the content searched for and retrieved daily by Web users. Take YouTube,[1] the currently most popular social video sharing application, as an example. With reportedly 24 h of videos uploaded per minute and 2 billion video views a day,[2] YouTube often figures among the top four applications in volume of traffic over the Internet.[3]

Social media typically includes a main object, stored in one of various media types (e.g., text, image, audio or video), as well as a variety of other sources of information related to the object, which we refer to as the object's associated *features*. *Content features* are sources of information which can be extracted from the object itself, such as the color histogram of an

* Corresponding author.
E-mail addresses: flaviov@dcc.ufmg.br (F. Figueiredo), hpinto@dcc.ufmg.br (H. Pinto), fmuniz@dcc.ufmg.br (F. Belém), jussara@dcc.ufmg.br (J. Almeida), mgoncalv@dcc.ufmg.br (M. Gonçalves), david@dcc.ufam.edu.br (D. Fernandes), edleno@dcc.ufam.edu.br (E. Moura).

[1] http://www.youtube.com.
[2] http://www.youtube.com/t/fact_sheet.
[3] http://www.alexa.com.

0306-4573/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ipm.2012.03.003

WSDM 2011 - Qualis B1

The Tube over Time: Characterizing Popularity Growth of YouTube Videos

Flavio Figueiredo Fabrício Benevenuto Jussara M. Almeida
{flaviov, fabricio, jussara}@dcc.ufmg.br
Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte - Brazil

ABSTRACT

Understanding content popularity growth is of great importance to Internet service providers, content creators and online marketers. In this work, we characterize the growth patterns of video popularity on the currently most popular video sharing application, namely YouTube. Using newly provided data by the application, we analyze how the popularity of individual videos evolves since the video's upload time. Moreover, addressing a key aspect that has been mostly overlooked by previous work, we characterize the types of the referrers that most often attracted users to each video, aiming at shedding some light into the mechanisms (e.g., searching or external linking) that often drive users towards a video, and thus contribute to popularity growth. Our analyses are performed separately for three video datasets, namely, videos that appear in the YouTube top lists, videos removed from the system due to copyright violation, and videos selected according to random queries submitted to YouTube's search engine. Our results show that popularity growth patterns depend on the video dataset. In particular, copyright protected videos tend to get most of their views much earlier in their lifetimes, often exhibiting a popularity growth characterized by a viral epidemic-like propagation process. In contrast, videos in the top lists tend to experience sudden significant bursts of popularity. We also show that not only search but also other YouTube internal mechanisms play important roles to attract users to videos in all three datasets.

Categories and Subject Descriptors

C.4 [**Computer Systems Organization**]: Performance of Systems—*Measurement techniques*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

General Terms

Human Factors, Measurement

WSDM'11, February 9–12, 2011, Hong Kong, China.

Keywords

YouTube, video popularity, popularity growth, referrer

1. INTRODUCTION

Understanding content popularity growth on the Internet is of great relevance to a broad range of services, from technological, economical and social perspectives. Such understanding can drive the design of cost-effective caching and content distribution mechanisms as well as uncover potential bottlenecks in system components such as search engines [6]. Moreover, predicting popularity is also important not only for supporting online and viral marketing strategies as well as effective information services (e.g., content recommendation and searching services) [12] but also because it may uncover new (online and offline) business opportunities. From a sociological point of view, a deep study of popularity evolution may also reveal properties and rules governing collective user behavior [10].

Online Social Networks (OSNs) are currently a major segment of the Internet. Considering video sharing OSNs, YouTube[1] is the one with the largest number of registered users [1], who upload and share their videos at a staggering rate. Indeed, it has been reported that the amount of content uploaded to YouTube in 60 days is equivalent to the content that would have been broadcasted for 60 years, without interruption, by NBC, CBS and ABC altogether [2]. Moreover, YouTube has reportedly served over 100 million users only on January 2009 [1], with a video upload rate equivalent to 10 hours per minute [2]. At such unprecedented user and content growth rates, understanding video popularity on YouTube becomes a challenge of utmost importance, as the myriad of different contents make user behavior and attention span highly variable and unpredictable [6].

As argued by Willinger *et al.* [20], most previous analyses of OSNs have treated such systems as static. Most of them focus on analyzing structural properties of single snapshots of relationship networks (e.g., friendship network) that emerge in such systems [3, 5, 15]. However, since OSNs are inherently dynamic, these studies fail to address key properties of the underlying system dynamics. Regarding one such property, namely popularity, a few studies have analyzed YouTube with respect to video popularity characteristics [6, 9, 10] and prediction [14, 19]. However, most of them, despite covering a rich set of popularity properties and their implications for system design, focused on only a single or

[1] `http://www.youtube.com`
[2] `http://www.youtube.com/t/fact_sheet`

Content Popularity Evolution in Online Social Networks

Flavio Figueiredo Fabricio Benevenuto Jussara M. Almeida
{flaviov, fabricio, jussara}@dcc.ufmg.br
Universidade Federal de Minas Gerais (UFMG)

ABSTRACT

Understanding content popularity growth on Online Social Networks (OSNs) is of great importance to Internet service providers, content creators and online marketers. However, most previous studies of OSNs are based on static views of the system, thus neglecting the temporal evolution of the network, and a possible correlation with content popularity growth. Moreover, previous analyses also greatly neglect the impact of the referrers (i.e., incoming links from external sites) on content popularity. We here provide some initial results on the analysis of content popularity growth in YouTube videos. Our study is based on three video datasets, namely popular videos, randomly collected videos, and copyright protected videos, with distinct characteristics in terms of temporal popularity evolution. We also characterize the different referrers that most often lead users to YouTube videos. Our results shed some light into aspects that impact content popularity growth.

Categories and Subject Descriptors

C.4 [**Computer Systems Organization**]: Performance of Systems—*Measurement techniques*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

General Terms

Human Factors, Measurement

Keywords

OSNs, YouTube, video popularity, popularity growth

1. THEORETICAL BACKGROUND

This paper describes a PhD work, being developed at the Universidade Federal de Minas Gerais. The work started in July 2010 and is expected to finish by July 2015.

Given that Online Social Networks (OSNs) are currently a major segment of the Internet, understanding content popularity growth on these networks is of great relevance to a broad range of services, from technological, economical and social perspectives. Such understanding can drive the design of cost-effective caching and content distribution mechanisms as well as uncover potential bottlenecks in system components such as search engines [5]. Moreover, predicting popularity is also important not only for supporting online and viral marketing strategies as well as effective information services (e.g., content recommendation and searching services) but also because it may uncover new (online and offline) business opportunities. From a sociological point of view, a deep study of popularity evolution may also reveal properties and rules governing collective user behavior [6].

Webmedia-WTD'11, October 3–6, 2011, Florianopolis, Brasil.

2. OBJECTIVES

The main objective of our work is to understand the diffusion and evolution of content popularity in large scale OSNs. In particular, we are interested on dealing with OSNs which focus on user created content (UGC) [1], due to the volume [5, 6] and more complicated nature of such media [4]. One representative example of such OSNs is YouTube[2], being the largest video sharing network nowadays.

In broader terms, we aim at understanding the evolution of content popularity with respect to three main research challenges (RC): (1) popularity growth patterns, which are related to the different patterns of popularity evolution across UGC content; (2) the referrers (i.e. incoming links) of UGC, which deals with how users find content on OSN and how this impacts popularity evolution; and finally, (3) how changes in the structure of the OSN affect popularity.

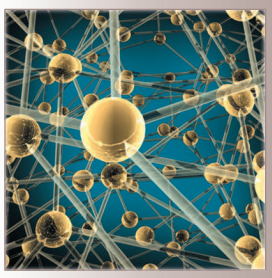
We begin our study with a review of the related literature in Section 3. A description of each challenge is presented in Section 4, while our research methodologies are presented in Section 5. In order to provide initial insights on *RC 1-2*, we characterized the growth patterns of video popularity on YouTube [8]. Using newly provided data by the application, we analyzed how the popularity of individual videos evolves since the video's upload time. We also characterize the different referrers for each video. Our results reveal differences in popularity evolution patterns depending on different video samples (top, random and copyrighted). These are presented in Section 6. Section 7 concludes the work.

3. RELATED WORK

Static views of popularity: There have been a few studies that address content popularity on OSNs, and, particularly, on video sharing systems. Cha *et al.* [5] presented

[1]Online radios, such as LastFM (`http://www.last.fm`), are examples OSNs which does not deal with UGC.
[2]`http://www.youtube.com`

IEEE Internet Computing 2010 - Qualis A1



On the Quality of Information for Web 2.0 Services

Most Web 2.0 applications let users associate textual information with multimedia content. Despite each application's lack of editorial control, these textual features are still the primary source of information for many relevant services such as search. Previous efforts in assessing the quality of these features primarily target single applications and mainly focus on tags, thus neglecting the potential of other features. The current study assesses and compares the quality of four textual features (title, tags, description, and comments) for supporting information services using data from YouTube, YahooVideo, and LastFM.

Jussara M. Almeida, Marcos André Gonçalves, Flavio Figueiredo, Henrique Pinto, and Fabiano Belém
Universidade Federal de Minas Gerais, Brazil

One key characteristic of Web 2.0 applications, such as YouTube, is the primary role users play in creating and sharing content. Although a significant amount of the content in these applications is multimedia, users are commonly encouraged to associate pieces of textual information – *textual features* – with the multimedia objects. Common examples are title, tags, description, and user comments.

Because these textual features are user generated, however, the respective applications have no editorial control and thus can't guarantee quality, neither in terms of syntactic correctness nor of the text's semantic relationship with the object. This poses a challenge to services such as search and advertising that primarily rely on textual features as sources of information about the objects' contents. This happens because the use of multimedia information retrieval mechanisms in Web 2.0 is still limited, possibly because state-of-the-art techniques are often ineffective under the low quality of most content and don't scale well to the size of several applications.[1]

Previous efforts toward assessing the quality of textual features primarily focused on tags, investigating how to use them to support search, recommendations, and object classification.[2–4] However, researchers haven't reached a consensus regarding their quality.[5–7] Moreover, they

1089-7801/10/$26.00 © 2010 IEEE

Assessing the Value of Contributions in Tagging Systems

Elizeu Santos-Neto[1], Flavio Figueiredo[2], Jussara Almeida[2], Miranda Mowbray[3], Marcos Gonçalves[2], Matei Ripeanu[1]

Elec. & Computer Engineering Department[1]
University of British Columbia
Vancouver, BC, Canada
{elizeus,matei}@ece.ubc.ca

Dep. de Ciência da Computação[2]
Univ. Federal de Minas Gerais
Belo Horizonte, MG, Brazil
{flaviovdf,jussara,mgoncalv}@dcc.ufmg.br

Hewlett-Packard Labs[3]
Bristol, United Kingdom
miranda.mowbray@hp.com

Abstract — **Assessing the value of individual users' contributions in peer-production systems is paramount to the design of mechanisms that support collaboration and improve users' experience. For instance, to incentivize contributions, file-sharing systems based on the BitTorrent protocol equate value with volume of contributed content and use a prioritization mechanism to reward users who contribute more. This approach and similar techniques used in resource-sharing systems rely on the fact that the physical resources shared among users are easily quantifiable.**
In contrast, information-sharing systems, like social tagging systems, lack the notion of a physical resource unit (e.g., content size, bandwidth) that facilitates the task of evaluating user contributions. For this reason, the issue of estimating the value of user contributions in information sharing systems remains largely unexplored. This paper introduces this problem and takes the first steps towards a solution. More precisely, it presents a framework to design algorithms that estimate the value of user contributions in tagging systems, proposes three complementary success criteria for potential solutions, and outlines the methodological evaluation challenges.

Keywords - tagging systems; information value; entropy; web;

I. INTRODUCTION

The wide adoption of blogs, wikis, and tagging systems has transformed online information production from a centralized, proprietary, and hierarchical editorial model to a decentralized, non-proprietary, and collaborative model. Benkler [1] defines systems with the above characteristics as *commons-based peer-production systems* (or, simply, peer production systems).

Quantifying the value of individual user contributions in peer-production systems is instrumental for a number of mechanisms that enable their efficient functioning and even their long-term survival. One direct application of methods to quantify the value of individual user contributions is to support incentive mechanisms to boost participation and collaboration. For instance, in *offline* peer-production systems like car pooling, drivers have an incentive to share their cars (i.e., give rides to other people), as cars with a higher occupancy are allowed to use a dedicated faster lane [1]. Similarly, in *online* peer-production systems, like BitTorrent [6], for instance, users who contribute more (i.e., they upload more) have higher download priority. Similarly, the value of contributions can also be used to deter malicious and opportunistic users by marginalizing those users who do not contribute at all.

In peer-production systems designed for the shared use of a single type of resource, which embeds an easily measurable physical quantity, quantifying users' contributions is generally straightforward. In *Folding@Home*[1], for instance, users donate CPU cycles, thus user contributions can simply be calculated by estimating the number of (normalized) CPU/hours donated by a user. Analogously, in BitTorrent-based systems contributions can be evaluated by estimating the volume of uploaded traffic. Thus, to a great extent, quantifying contributions in these systems reduces to counting the units of donated resources. Even in peer-production systems that deal with multiple types of resources, accounting for the amounts of physical resources produced by one user and consumed by others is at the core of techniques that quantify the value of contributions [2].

However, some systems lack a clear mapping between the amount of physical resources donated and the value of a contribution. This is the case of peer-production systems designed to support the production of information goods, such as social tagging systems (e.g., *del.icio.us* or *CiteULike.org*) and wikis (e.g., *Wikipedia.org*). In particular, the lack of a quantifiable resource unit creates a new challenge: the value of each user's contributions cannot anymore be directly linked to some amount of resource used to produce them. For example, whereas a user may produce a large number of tags, only a few of them may indeed help their peers (i.e., other users of the system) in particular tasks such as navigating the list of items or organizing their item collections. Moreover, the value of information is contextual to the user. Some users may find the tag 'agneta' valuable when searching for information about a specific item in the system (e.g., Pedro Juan Gutierrez's novel Tropical Animal where Agneta is one of the characters), while others may

This work was supported in part by IDRC/AUCC (LACREG 2009/10), INCT-Web (MCT/CNPq grant 57.3871/2008–6), the Institute for Computing, Information and Cognitive Systems (ICICS) at UBC, and the authors' individual grants from BC Innovation Council (BCIC), UBC, CNPq, and FAPEMIG.

[1] http://folding.stanford.edu

Evidence of Quality of Textual Features on the Web 2.0

Flavio Figueiredo
Dep. of Computer Science
Fed. Univ. of Minas Gerais
flaviov@dcc.ufmg.br

Fabiano Belém
Dep. of Computer Science
Fed. Univ. of Minas Gerais
fmuniz@dcc.ufmg.br

Henrique Pinto
Dep. of Computer Science
Fed. Univ. of Minas Gerais
hpinto@dcc.ufmg.br

Jussara Almeida
Dep. of Computer Science
Fed. Univ. of Minas Gerais
jussara@dcc.ufmg.br

Marcos Gonçalves
Dep. of Computer Science
Fed. Univ. of Minas Gerais
mgoncalv@dcc.ufmg.br

David Fernandes
Dep. of Computer Science
Fed. Univ. of Minas Gerais
davidf@dcc.ufmg.br

Edleno Moura
Dep. of Computer Science
Fed. Univ. of Amazonas
edleno@dcc.ufam.edu.br

Marco Cristo
FUCAPI - Analysis, Research
and Tech. Innovation Center
marco.cristo@gmail.com

ABSTRACT

The growth of popularity of Web 2.0 applications greatly increased the amount of social media content available on the Internet. However, the unsupervised, user-oriented nature of this source of information, and thus, its potential lack of quality, have posed a challenge to information retrieval (IR) services. Previous work focuses mostly only on tags, although a consensus about its effectiveness as supporting information for IR services has not yet been reached. Moreover, other textual features of the Web 2.0 are generally overseen by previous research.

In this context, this work aims at assessing the relative quality of *distinct* textual features available on the Web 2.0. Towards this goal, we analyzed four features (TITLE, TAGS, DESCRIPTION and COMMENTS) in four popular applications (CiteULike, Last.FM, Yahoo! Video, and Youtube). Firstly, we characterized data from these applications in order to extract evidence of quality of each feature with respect to usage, amount of content, descriptive and discriminative power as well as of content diversity across features. Afterwards, a series of classification experiments were conducted as a case study for quality evaluation. Characterization and classification results indicate that: 1) when considered separately, TAGS is the most promising feature, achieving the best classification results, although its absence in a non-negligible fraction of objects may affect its potential use; and 2) each feature may bring different pieces of information, and combining their contents can improve classification.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services

General Terms

Experimentation, Measurement

Keywords

Web 2.0, Textual Features, Social Media

CIKM'09, November 2–6, 2009, Hong Kong, China.

1. INTRODUCTION

Web 2.0 applications have grown significantly in diversity and popularity in recent years. Popular examples include Youtube and Yahoo! Video[1] (or simply YahooVideo), two social video sharing applications, Last.FM[2] (or simply LastFM), an online radio and music community website, and CiteULike[3], a scholarly reference management and discovery service. By distributing mostly *user generated content* and enabling the establishment of online communities and social networks, these applications make use of collaborative knowledge to increase the amount and diversity of content offered. Youtube, for example, is currently the largest video database in the world[4], and the second most searched Website[5]. Although the musical content available in LastFM is not *generated* by its users, it is currently one of the most popular Internet radio stations, due to its community based organization, which gives users the ability to describe musical content and interact over social networks.

Social media is here used to refer to the content, most commonly generated by users, available in Web 2.0 applications. This typically comprises a main object, stored in one of various media types (text, image, audio or video), as well as several other sources of information, commonly in textual form (and thus referred to as *textual features*), associated with it (e.g., tags). Being often unsupervised sources of data, social media offers no guarantee of quality, thus posing a challenge to information retrieval (IR) services such as

[1] http://youtube.com and http://video.yahoo.com
[2] http://last.fm
[3] http://www.citeulike.org
[4] http://www.comscore.com/press/release.asp?press=1929
[5] http://bits.blogs.nytimes.com/2008/10/13/search-ads-come-to-youtube
URLs were last accessed on June 2009

Webmedia 2009 - Qualis B3

Caracterizando o Uso e a Qualidade de Atributos Textuais em Aplicações da Web 2.0

Flavio Figueiredo
Dep. of Computer Science
Fed. Univ. of Minas Gerais
flaviov@dcc.ufmg.br

Fabiano Belém
Dep. of Computer Science
Fed. Univ. of Minas Gerais
fmuniz@dcc.ufmg.br

Henrique Pinto
Dep. of Computer Science
Fed. Univ. of Minas Gerais
hpinto@dcc.ufmg.br

Jussara Almeida
Dep. of Computer Science
Fed. Univ. of Minas Gerais
jussara@dcc.ufmg.br

Marcos Gonçalves
Dep. of Computer Science
Fed. Univ. of Minas Gerais
mgoncalv@dcc.ufmg.br

David Fernandes
Dep. of Computer Science
Fed. Univ. of Minas Gerais
davidf@dcc.ufmg.br

Edleno Moura
Dep. of Computer Science
Fed. Univ. of Amazonas
edleno@dcc.ufam.edu.br

Virgilio Almeida
Dep. of Computer Science
Fed. Univ. of Minas Gerais
virgilio@dcc.ufmg.br

Marco Cristo
FUCAPI - Analysis, Research and Tech. Innovation Center
marco.cristo@gmail.com

ABSTRACT

Apesar da grande quantidade de conteúdo multimídia em aplicações Web 2.0, a maioria de seus serviços de Recuperação de Informação (RI) fazem uso apenas dos atributos textuais associados ao conteúdo (por exemplo, etiquetas ou *tags*). Entretanto, por serem tipicamente gerados pelos usuários, tais atributos não oferecem garantias de qualidade para serviços de RI. Neste trabalho, são investigadas evidências de qualidade de atributos textuais em aplicações populares da Web 2.0 em relação a três aspectos: utilização; poder descritivo e discriminativo. Para tanto, foi feita uma caracterização do uso de quatro atributos textuais (título, descrição, *tags* e comentários) nos sistemas Youtube, YahooVideo, LastFM e CiteULike. Alguns de nossos resultados, que podem ser considerados no projeto de serviços de RI na Web 2.0 são: (1) atributos textuais colaborativos, embora não sejam explorados significativamente em algumas aplicações, contêm a maior quantidade de informação quando presentes; (2) há uma diversidade significativa de informação entre os atributos textuais; e (3) o título e as *tags* dos objetos parecem ser os atributos mais promissores para serviços de RI, visto que o primeiro é quase sempre presente e tem alto poder descritivo, e o segundo, quando utilizado, apresenta alto poder descritivo e discriminativo.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services

General Terms

Experimentação, Caracterização

WebMedia'09, October 5–7, 2009, Fortaleza, CE, Brazil.

Keywords

Web 2.0, Atributos Textuais

1. INTRODUÇÃO

Aplicações Web 2.0 têm crescido significativamente em diversidade e popularidade nos últimos anos. Exemplos incluem o Youtube e o Yahoo! Video[1] (ou simplesmente YahooVideo), dois sistemas sociais de compartilhamento de vídeo, o Last.FM[2] (ou simplesmente LastFM), uma estação de rádio *online* que possibilita a interação social em torno de tópicos relacionados a música, e o CiteULike[3], um serviço de gerência e descoberta de referências bibliográficas voltado para a comunidade acadêmica. Fazendo uso da distribuição de conteúdo gerado por usuários, do estabelecimento de redes sociais, e de colaboração, estas aplicações oferecem uma maior quantidade e diversidade de conteúdo em relação às aplicações da Web tradicional. O Youtube, por exemplo, é atualmente uma das maiores bases de dados[4], e o segundo sistema de busca mais utilizado[5] do mundo. Embora o conteúdo musical disponível no LastFM não seja gerado por usuários, esta aplicação é uma das estações de rádio mais populares na Internet, devido a sua organização baseada em comunidades, que permite ao usuário descrever conteúdo musical e interagir em redes sociais.

O termo "mídia social" refere-se ao conteúdo produzido por usuários e disponibilizado na Web 2.0. Este compreende tanto o conteúdo principal (denominado *objeto*), que pode ser armazenado em forma de texto, imagem, áudio ou vídeo, quanto todas as informações textuais associadas ao conteúdo (denominadas *atributos textuais*). Por ser gerada pelos usuários, não há garantias de que a mídia social ofereça a qualidade demandada por outros usuários e, em particular, por serviços de Recuperação de Informação (RI), tais como

[1] http://youtube.com e http://video.yahoo.com
[2] http://last.fm
[3] http://www.citeulike.org
[4] http://www.comscore.com/press/release.asp?press=1929
[5] http://bits.blogs.nytimes.com/2008/10/13/search-ads-come-to-youtube

On the Impact of Choice in Multi-Service P2P Grids

Álvaro Coêlho, Paulo Ditarso, Flavio Figueiredo, David Maia, Francisco Brasileiro

Universidade Federal de Campina Grande
Departamento de Sistemas e Computação
Laboratório de Sistemas Distribuídos
Av. Aprígio Veloso, s/n, Bloco CO
58.109-970, Campina Grande - PB, Brazil
Emails: {degas, pmaciel, flaviov, davidcmm, fubica}@lsd.ufcg.edu.br

***Abstract*—In this paper we consider a peer-to-peer grid system which provides multiple services to its users. In this system, an incentive mechanism promotes collaboration among peers. It has been shown that the use of a reciprocation mechanism in such a system is able to prevent free riding and, at the same time, promotes the clustering of peers that have mutually profitable interactions. However, when peers are subject to a budget limitation, each peer must select a subset of all services that can possibly be offered. In this work we show that the received utility is strongly dependent of the offered services. The main contributions of this work are a methodology to evaluate the impact of service changes in the obtained utility and how much different sets of offered services impact in the peer's utility. These results indicate that further research is needed, particularly for the development of heuristics to choose the best services to offer.**

I. Introduction

With the popularity of the Internet, peer-to-peer (P2P) systems are becoming an interesting way to obtain files, storage, processing and many other computational resources with high levels of availability.

P2P systems can be defined as networks of computers in which participants play both the role of a client and that of a server [1]. This is to say that wherever a new client is added, there is a new server that is also added to the system. In these systems, peers who are acting as servers must schedule the use of their resources in order to provide services to the other peers that are acting as clients at that time.

This scheduling process has been often done using two approaches: market-based and sharing-based [2]. Market-based approaches have been used in many systems [3], [4], [5]. Although market-based solutions give more strict guarantees on the quality of service provided, the nature of this approach implies in higher transaction costs for participating in such economy. This occurs because this kind of solution relies on the existence of a currency distribution system, banking services, auditing, and accurate pricing [6]. On the other hand, no currency schemes or any trustworthy central institution are used in sharing systems, and peers exchange resources based on a reciprocation scheme. In this environment, peers can make use of social mechanisms for monitoring and enforcement, because the information is loosely structured (and therefore, easier to obtain). So, this kind of solution decreases the transaction costs [7]. Reciprocation between peers is achieved through the use of incentive mechanisms.

Reciprocation-based mechanisms have been explored in previous works [6], [8], [9]. These mechanisms presume a single good to be exchanged, making the reciprocation schema simpler. However, they are not appropriate for the case in which peers need to exchange multiple goods (or services). In this case some additional considerations must be made, because peers may exchange a kind of service by another and peers may value services differently [10]. Consequently, the partnership between peers is based not only in the peer's behavior but also in the profitability of their interactions.

The Extended Network of Favors (ExtNoF) mechanism was proposed in [10] as a variation of the original Network of Favors. ExtNoF allows peers to exchange multiple services using the past interactions to find the more profitable partnerships, generating clusters of peers having compatible affinities that lead to mutually profitable interactions.

Services can be thought at the resource level (e.g. CPU, disk, bandwidth, etc), but also at a higher level, such as the execution of a particular software. Given the multitude of services that can be possibly offered, and the fact that peers have a limited budget, not all services can be offered by all peers. This imposes the need for each peer to choose a set of services to offer among all possible services. Each possible set has a specific cost and returns different profits.

In this paper we make an analysis of a P2P system in which multiple services are exchanged. We show that, for each peer, there are different set of services to be offered and each set gives different results. Moreover, we show also the impact of different choices of services in the peers' final utilities.

The rest of the paper is organized as follows. Related work to ours is discussed in Section II. Section III gives a formal description of the problem, and describes a utility function for the multi-service P2P grid system. A numerical evaluation of the system is presented in Section IV, in which different scenarios were considered. Finally, Section V closes the paper with our concluding remarks and perspectives of future work.

978-1-4244-2191-6/08/$25.00 ©2008 IEEE

NOMS 2008 - Qualis A2

On the Planning of a Hybrid IT Infrastructure

Paulo Ditarso, Flavio Figueiredo, David Maia, Francisco Brasileiro, Álvaro Coêlho
Universidade Federal de Campina Grande
Departamento de Sistemas e Computação
Laboratório de Sistemas Distribuídos
Av. Aprígio Veloso, s/n, Bloco CO
58.109-970, Campina Grande - PB, Brasil
Emails: {pmaciel, flaviov, davidcmm, fubica, degas}@lsd.ufcg.edu.br

***Abstract*—With the emergence of utility computing and the continuous search for reducing the cost of running Information Technology (IT) infrastructures, we will soon experience an important change on the way these infrastructures are assembled, configured and managed. In this paper we consider the problem of managing a hybrid high-performance computing infrastructure whose processing elements comprise in-house dedicated machines, a utility computing service provider, and idle machines from a best-effort peer-to-peer grid. This infrastructure supports the execution of both best-effort and real-time applications. Real-time applications use primarily computing power from the in-house machines and any processing power that can be attained from the best-effort grid. Extra capacity required to meet deadlines is purchased from the utility computing service provider. This extra capacity is reserved for future use through short term contracts which are negotiated with no human intervention. We take a business-driven approach for the management of this hybrid infrastructure and propose heuristics that can be used by a contract planner agent to reduce the cost of running the applications at the same time that guarantees that deadlines are met. In particular, we show that constructing an estimation for the behavior of the grid is essential for making contracts that lead to high efficiency in the use of the hybrid infrastructure.**

I. Introduction

Utility computing is becoming reality with several companies offering solutions for its implementation and others already providing computation on demand based on these solutions. Those promoting this kind of solution place as one of their main selling arguments the possibility of providing their customers with substantial reductions on the total cost of ownership (TCO) of their Information Technology (IT) infrastructures.

Although economical advantages certainly play an important role on the adoption of a technology, other factors also have their saying. For instance, it is likely that the migration of services that are supported by in-house dedicated IT infrastructure to one offered by an external utility computing service provider will face strong resistance from the internal IT management staff. Also, it may well be the case that running certain types of services on a utility computing infrastructure will not yield the cost reduction sought. More importantly, for some services it may not be desirable to have them executing on a third-party infrastructure (e.g. strategic, sensitive and critical services). Finally, having some in-house capacity may minimize possible undesirable effects of price fluctuations due to transient instabilities in the utility computing market.

On the other hand, the market-based utility computing model is not the only alternative to reduce TCO. Among other solutions proposed, peer-to-peer grid computing has been suggested as a way to enable a simpler economy for the trading of idle cycles [1]. Markets rely on the existence and efficiency of contract negotiation, norm enforcement, banking and accounting mechanisms. For several scenarios in distributed computing (and also outside computing) implementing such mechanisms is complex, costly or inefficient [2]. On the other hand, in these situations sharing systems may be efficient, as they can use information which is loosely structured and therefore easier to obtain, they can make use of social mechanisms for monitoring and enforcement, and they have lower marginal transaction costs [2]. However, unlike the solutions based on markets, these solutions generally give no guarantees on the quality of the service provided. Nevertheless, they have been successfully used to increase the cost effectiveness of IT infrastructures in a number of settings [3].

Given all that, in the near future we envision that many IT infrastructures will comprise a mix of computing power provided by in-house dedicated infrastructure and computing power provided from external utility computing service providers. Moreover, both kinds of approaches will provide different guarantees, varying from those with well defined quality of service to those provided in a best-effort basis.

This conjecture is also supported by our own experience. In our effort to disseminate the use of the OurGrid middleware we have fostered the creation of the OurGrid Community (http://www.ourgrid.org/), an open peer-to-peer grid that is in production since December, 2004. (See http://status.ourgrid.org/ for an up to date snapshot of the running system.) The OurGrid Community has been used in a variety of areas, from engineering [4] to bioinformatics [5], from computer science [6] to financial applications [7]. In particular, OurGrid supports the cooperative work of a community of meteorologists and hydrologists, both in the academia and in the government [4], [8]. Some members of this community provide daily weather forecasts as a public service (see, for instance, http://www.cptec.inpe.br/). Moreover, the capacity required at critical times (when real-time applications are run) is normally much larger than that required at other times. So, if for one side over provisioning the IT infrastructure of these public agencies to cope with the high demand at critical times

978-1-4244-2066-7/08/$25.00 ©2008 IEEE

Um portifolio de segurança para um sistema de grade entre pares de livre entrada

Flavio Figueiredo, Matheus Gaudêncio, Thiago Emmanuel, Rodrigo Miranda, Francisco Brasileiro

[1]Universidade Federal de Campina Grande
Departamento de Sistemas e Computação
Laboratório de Sistemas Distribuídos
Av. Aprígio Veloso, s/n, Bloco CO
58.109-970, Campina Grande - PB, Brazil

{flaviov,matheusgr,thiagoepdc,vilar,fubica}@lsd.ufcg.edu.br

Abstract. *Security is an important aspect in grid computing, due to the necessity to protect the grid resources, users and middleware. In free-to-join peer-to-peer grids, the security issues are more complex, because there are not strong identities in the system. In fact, there is not a "one size fits all" solution for the users of this kind of grid. This paper adapts the security technologies available in the other kinds of grids and distributed systems, applying then to the free-to-join peer-to-peer grids, in order to build a security portfolio for the distinct user types in this kind of system. The portfolio was developed and validated on OurGrid.*

Resumo. *Segurança é um aspecto importante na computação em grade, pois é preciso proteger os recursos, os usuários e o próprio middleware da grade. Nas grades entre pares e de livre entrada, as questões de segurança se tornam mais complexas, devido à ausência de identidades fortes no sistema. Na realidade, não existe uma solução única que satisfaça os requisitos de todos os usuários deste tipo de grade. Neste artigo, as tecnologias de segurança disponíveis para os outros tipos de grades e sistemas distribuídos são adaptadas e aplicadas às grades entre pares e de livre entrada, a fim de formar um portifolio de segurança disponível para os diferentes perfis de usuários deste tipo de sistema. O portifolio foi implementado e validado no OurGrid.*

1. Introdução

É cada vez mais comum o uso de grades computacionais para o auxílio de pesquisas científicas que exijam um grande poder computacional. Há na literatura propostas de diversos modelos de grades, cada um apresentando requisitos distintos de segurança para atender às necessidades dos seus respectivos usuários. Em geral, existem dois problemas de segurança em grades: (i) a segurança dos usuários da grade contra recursos maliciosos; e (ii) a segurança dos recursos da grade contra usuários maliciosos. Uma forma comum de garantir a proteção de ambas as partes é fazer uso de um esquema de segurança baseado na confiança mútua entre os participantes [Foster et al. 1998].

O OurGrid é um *middleware* de grades computacionais que permite que qualquer usuário possa participar do sistema. Em `http://status.ourgrid.org` pode-se

Bridging the High Performance Computing Gap: the OurGrid Experience

Francisco Brasileiro, Eliane Araújo, William Voorsluys, Milena Oliveira, Flávio Figueiredo
{fubica, eliane, william, milena, flaviov}@lsd.ufcg.edu.br

Universidade Federal de Campina Grande
Departamento de Sistemas e Computação
Laboratório de Sistemas Distribuídos
Av. Aprígio Veloso, s/n, Bloco CO
58.109-970, Campina Grande - PB, Brazil

Abstract

High performance computing is currently not affordable for those users that cannot rely on having a highly qualified computing support team. To cater for these users' needs we have proposed, implemented and deployed OurGrid. OurGrid is a peer-to-peer grid middleware that supports the automatic creation of large computational grids for the execution of embarrassingly parallel applications. It has been used to support the OurGrid Community - a public free-to-join grid that is in production since December 2004. In this paper we show how the OurGrid Community has been used to support the execution of a number of applications. Further we discuss the main benefits brought up by the system and the difficulties that have been faced by the system developers and the users and managers of the OurGrid Community.

Keywords: *grid computing; peer-to-peer; free-to-join; public grid; bag-of-tasks applications.*

1 Introduction

New developments in communication and computing technologies have substantially impacted the way scientific research is conducted. Not only these technologies allow for an unprecedent level of interaction among researchers, but also the use of computers for data analysis, simulations, visualization of results etc play nowadays a fundamental role in the working methodology adopted by many research groups. As a consequence, having access to high performance computing facilities has become a must for many research areas.

Due to this demand, research in computing science has, for some time now, sought ways to expand the reach of high performance computing. One of the first initiatives in this sense provided a way to aggregate unused computing power in a local area network [17]. The next step was to increase the scale on the number of resources aggregated by harvesting the idle computing power in the Internet [4, 2], what has been dubbed *public resource computing* or *voluntary computing*. More recently, *grid computing* has been proposed as a way to build virtual organizations aggregating computing resources that are under different administrative domains [14].

Despite being successful in providing non-trivial amounts of computing power, these mainstream technologies are not affordable to most users with demands for high performance computing. In the case of voluntary computing, it is necessary to set up a large control center that will be responsible for managing the millions of public resources that are contributed to the system. Moreover, a lot of effort needs to be placed in convincing resource owners to install the software that will allow them to contribute their resources to the system. On the other hand, most grids in production (eg. EGEE - www.eu-egee.org, TeraGrid - www.teragrid.org, NAREGI - www.naregi.org, NGS - www.grid-support.ac.uk, APAC - www.apac.edu.au, PRAGMA - www.pragma-grid.net) use some flavor of the Globus toolkit [1]. Installing, configuring, and customizing the Globus middleware is not a trivial task, and today requires a highly skilled support team. Moreover, joining such grids involves a negotiation process that consumes time and may place players with smaller amounts of resources in a disadvantageous position in relation to larger players.

Therefore, if for one side the massive use of computers by researchers fosters, at an ever-faster pace, amazing developments for the society at large, for the other side it con-

Seventh IEEE International Symposium on Cluster Computing and the Grid(CCGrid'07)
0-7695-2833-3/07 $20.00

Implementing a Distributed Execution Service for a Grid Broker

Flavio V. D. de Figueiredo, Francisco V. Brasileiro, Andrey E. M. Brito

Universidade Federal de Campina Grande
Departamento de Sistemas e Computação
Laboratório de Sistemas Distribuídos
Av. Aprígio Veloso, 882 - Bloco CO, Bodocongó
CEP 58109-970, Campina Grande - PB, Brasil

flaviov@lsd.ufcg.edu.br, {fubica, andrey}@dsc.ufcg.edu.br

Abstract. *Grid middleware such as OurGrid offer solutions for executing parallel tasks on a grid system. In such systems, users submit their applications for executions through a client broker. MyGrid is the client broker used for the OurGrid system; it is in charge of managing task executions that a user has submitted. Although the broker is able to detect task failures and reschedule them, MyGrid itself constitutes a single point of failure from the user perspective. If it fails, all knowledge of task executions is lost. Moreover, MyGrid is also a bottleneck, since hundreds, or even thousands, of executions could potentially be spawned by an application and need to be managed at the same time by a single broker. In this paper we present the design and implementation of a fault-tolerant distributed execution service that allows for load balancing and improves MyGrid performance. A checkpointing mechanism is used to ease the implementation of the service and to further increase system reliability.*

1. Introduction

A grid system comprises a set of distributed, heterogeneous resources, such as: personal computers, storage space, clusters, servers, etc., connected through a network. Grid computing has been proposed to cater for the high computational demand in different areas of research, such as physics, biology, astronomy and computer science itself. A grid user gains access to a grid system by using a grid middleware; with it the user can access a variety of services, such as: resource management, security services, monitoring services and execution services [Foster et. al., 2001][Foster and Iamnitchi, 2003].

The OurGrid system [Cirne et. al., 2006] is a peer-to-peer grid middleware in which sites lend their idle resources to gain access to resources from other sites when they need. The goal behind the use of the system is the execution of Bag of Tasks (BoT) applications, those parallel applications in which tasks do not need to communicate among each other. The OurGrid middleware has been used to support OurGrid's free-to-join community, which is in production since December 2004 (see http://status.ourgrid.org/ for a fresh snapshot of the running system).

Three main services compose the OurGrid system; these are the brokers, resources and peers. Peers are in charge of delivering resources, known as Grid Machines or simply GuMs, to the broker, known as MyGrid [Cirne et. al., 2003]. MyGrid runs in the user machine, named the *home machine*. GuMs offer an environment for the execution of tasks, computation requested by the broker is done on these . A machine is available as a GuM if the User Agent service is running at it, this service is available for Linux, Windows or using OurGrid´s sandboxing solution called SWAN[Cirne et. al., 2006]. Peers also exchange resources amongst themselves, using the Network of Favours incentive mechanism [Cirne et. al., 2006]. An OurGrid site is a Local Area Network with Peers, Brokers and GuMs running on the machines that compose it.

Current fault-tolerance in OurGrid is only available for the GuMs, such failures are treated with checkpointing mechanisms [OurGrid Team, 2006] or with task replication [Cirne et.

Girafas: Uma Ferramenta Computacional para Apoio no Ensino de Algoritmos Genéticos

Flávio Roberto Santos[1], **Flavio V D de Figueiredo**[2], **Vinicius F C Florentino**[3], **Joseana Macêdo Fechine**[4]

[1,2,3,4]Departamento de Sistemas e Computação,

Centro de Engenharia Elétrica e Informática

Universidade Federal de Campina Grande

flaviors@dsc.ufcg.edu.br, flaviov@lsd.ufcg.edu.br,
vinicius@lsd.ufcg.edu.br, joseana@dsc.ufcg.edu.br

Abstract. Genetic Algorithms make use of evolutionary concepts observed in nature to solve computational problems in less time than it would be necessary by the use of traditional algorithms. These types of algorithms are generally used to solve scheduling and timetable problems. Nowadays, these algorithms are becoming more popular in various areas, many outside computer science. This work describes the construction of a tool, Girafas, with studies purpose which intends to apply and validate the use of genetic algorithms in a ludic environment. The tool uses a virtual world composed by giraffes and trees and the evolution is observed through time, with the possibility of manipulating variables that model the environment.

Resumo. Algoritmos Genéticos utilizam o conceito de evolução observado na natureza para solucionar problemas computacionais de forma mais rápida do que com o uso de algoritmos tradicionais. Os problemas comuns que utilizam algoritmos genéticos são os de escalonamento e alocação. Atualmente, esta técnica é utilizada em diversas áreas, inclusive fora do âmbito da computação. Este trabalho descreve a construção de uma ferramenta, Girafas, com fins didáticos que visa aplicar e validar o uso de algoritmos genéticos em um ambiente lúdico. A ferramenta utiliza um cenário virtual composto por girafas e árvores, no qual é observada a evolução das girafas no tempo, a partir da manipulação das variáveis que modelam o ambiente.

1 Introdução

Os Algoritmos Genéticos (AG) são uma família de modelos computacionais inspirados na evolução que utilizam esse conceito para solucionar, por exemplo, problemas computacionais que demandariam muito tempo com o uso de algoritmos tradicionais. Os problemas mais comuns que utilizam algoritmos genéticos são os de escalonamento e alocação. Atualmente, esta técnica também está sendo usada em diversas áreas, inclusive fora do âmbito da computação (GOUNARES e SIKCHI, 2002; RUSSEL and NORVIG, 2003).

Dentro desse contexto, o objetivo do trabalho é o desenvolvimento de uma

Bancas e Teses

Defesa Anna 1/2

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PPGCC
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DCC-UFMG

DCC.149/2015 - Pós-Graduação
Belo Horizonte, 24 de agosto de 2015.

Ilmo. Sr.
Dr. Flavio Vinicius Diniz de Figueiredo
Pós-Doc/ DSC - UFCG

Prezado Senhor,

O Colegiado do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Minas Gerais tem a satisfação de convidar V. Sa. para participar da Banca Examinadora da 979ª defesa de Dissertação de Mestrado do Programa.

Título: CHARACTERIZING AND MODELING THE DYNAMICS OF ONLINE KNOWLEDGE-SHARING NETWORKS

Aluna: **Anna Christina de Carvalho Guimarães**

Data da Defesa: 27/08/2015 às 10:00 horas

Local: Sala 2077 do ICEx - UFMG
Av. Antônio Carlos, 6627 - Campus - Pampulha

Banca Examinadora:

PROFA. JUSSARA MARQUES DE ALMEIDA GONÇALVES - Orientadora
Departamento de Ciência da Computação - UFMG

PROF. ANA PAULA COUTO DA SILVA - Coorientadora
Departamento de Ciência da Computação - UFMG

PROF. ARTUR ZIVIANI
Laboratório Nacional de Computação Científica - CNPq

DR. FLAVIO VINICIUS DINIZ DE FIGUEIREDO
Pós-Doc/ DSC - UFCG

PROF. PEDRO OLMO STANCIOLI VAZ DE MELO
Departamento de Ciência da Computação - UFMG

Atenciosamente,

Prof. Luiz Chaimowicz
Coordenador do Programa de Pós-Graduação
em Ciência da Computação - ICEx - UFMG

Defesa Anna 2/2

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PPGCC
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DCC-UFMG

DECLARAÇÃO

Declaramos, para os devidos fins, que o dr. **Flavio Vinicius Diniz de Figueiredo** participou da banca de defesa de dissertação de mestrado de **Anna Christina de Carvalho Guimarães**, aluna do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Minas Gerais, intitulada *Characterizing And Modeling The Dynamics Of Online Knowledge-sharing Networks.*

Belo Horizonte, 27 de agosto de 2015.

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO UFMG

Prof. Luiz Chaimowicz
Coordenador do Programa de Pós-Graduação
em Ciência da Computação - ICEx - UFMG

Eventos

CERTIFICADO

Certificamos que **Flávio Figueiredo** apresentou o artigo **Content Popularity Evolution in Online Social Networks** no Workshop de Teses e Dissertações em Banco de Dados do **XXVI Simpósio Brasileiro de Banco de Dados**, realizado em 06 de outubro de 2011 em Florianópolis – SC.

Florianópolis, 06 de outubro de 2011.

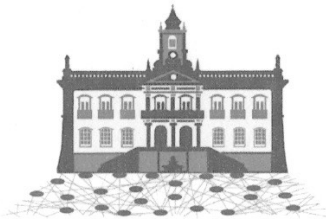
Carina F. Dorneles

Roberto Willrich

Ronaldo S. Mello

Coordenadores Gerais

SBC



XXX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos

CERTIFICADO

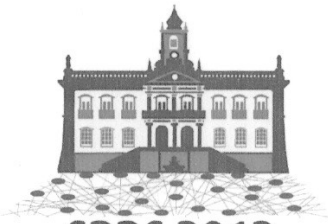
Conferimos o presente certificado a **Flávio Vinícius Diniz de Figueiredo** por sua participação como voluntário na organização do **XXX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, realizado de 30 de abril a 4 de maio de 2012 em Ouro Preto - MG.

Ouro Preto, 4 de maio de 2012

Jussara Marques de Almeida — Dorgival O. Guedes Neto

Coordenadores Gerais do SBRC 2012

SBC



XXX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos

CERTIFICADO

Conferimos o presente certificado a **Flávio Vinícius Diniz de Figueiredo** por sua participação no **XXX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, realizado de 30 de abril a 4 de maio de 2012 em Ouro Preto - MG, totalizando uma carga horária de 40 horas.

Ouro Preto, 4 de maio de 2012

Jussara M. Almeida — Dorgival O. Guedes Neto

Coordenadores Gerais do SBRC 2012

SBC

Certificado

Certificamos que **FLAVIO VINICIUS DINIZ DE FIGUEIREDO** participou do *XXIV Simpósio Brasileiro de Redes de Computadores*, promovido pela Sociedade Brasileira de Computação e realizado de 29 de maio a 2 de junho de 2006 em Curitiba PR, tendo também participado dos seguintes eventos, realizados em conjunto com o simpósio:

Mini-curso 5: Segurança em redes P2P: princípios, tecnologias e desafios (4h)
VII Workshop da Rede Nacional de Pesquisa (16h)
VII Workshop de Testes e Tolerância a Falhas (8h)

Curitiba PR, 29 de maio de 2006

Carlos Alberto Maziero
Elias Procópio Duarte Jr
Keiko Ono Fonseca
Coordenação Geral do SBRC 2006

SBC

Certificamos que FLAVIO VINICIUS DINIZ DE FIGUEIREDO participou do 23º Simpósio Brasileiro de Redes de Computadores (SBRC2005), realizado em Fortaleza-CE, de 09 a 13 de Maio de 2005.

Atividades:

Sessões Técnicas
Salão de Ferramentas
Palestras
Painéis
WTF: Workshop de Tolerância a Falhas

Rossana Maria de Castro Andrade

Joaquim Celestino Júnior

SBC

www.sbc.org.br

NetSci X Rio de Janeiro

International School and Conference on Network Science | Brazil, January 14–16 2015

CERTIFICATE

This is to certify that

Flavio Figueiredo and Jussara Almeida

had presented a paper entitled "Modeling and Mining Information Popularity Online " at the **NetSci-X 2015** from 14th to 16th of January 2015 in Rio de Janeiro, Brazil.

Alexandre G. Evsukoff
General Chair

SEBRAE

Certificado

Conferido a *Flávio V. D. de Figueiredo*

pela participação Na Uffizi PB 2006 – Oficinas para Negócios Inovadores de TIC da Paraíba

realizado no período de 30 de novembro a 01 de dezembro de 2006

com carga horária 15 horas.

Ministrado por

João Pessoa 01 de dezembro de 2006

Farol Digital
Apoiando Empreendimentos de Tecnologia da Informação e Comunicação na Paraíba

Uffizi PB 2006
Oficinas para Negócios Inovadores de Empresas de Tecnologias da Informação e Comunicação da Paraíba
30 de Novembro e 01 de Dezembro

SEBRAE

Ivani Costa

UNICENP

CENTRO UNIVERSITÁRIO POSITIVO

CERTIFICADO

Confere-se este certificado a **Flavio V. D. de Figueiredo**, pela apresentação do artigo **Girafas: Uma Ferramenta Computacional para Apoio no Ensino de Algoritmos Genéticos (#24787)** no **III Simpósio Brasileiro de Sistemas de Informação**, realizado no câmpus do Centro Universitário Positivo, em Curitiba, de 8 a 10 de novembro de 2006.

Curitiba, 10 de novembro de 2006.

Francisco Javier Kantek Garcia Navarro
Coordenador Geral do SBSI 2006

Alexandre Reis Graeml
Coordenador Científico do SBSI 2006

香港 中国旅行社 有限公司
CTS CHINA TRAVEL SERVICE (HONG KONG) LIMITED
商務會獎中心 牌照號碼：350273
Business Travel & MICE Centre
香港德輔道中 71 號永安集團大廈 601- 2 室
Suite 601- 2, Wing On House, 71 Des Voeux Road Central, HONG KONG
電話 Tel: (852) 2160 5055 傳真 Fax: (852) 3547 6072 / (852) 2851 1426

發 票
OFFICIAL RECEIPT

參考編號 **Reference Number:**	**CIKM 400**	收款日期 **Payment Date:**	9/3/2009 (MM/DD/YY)
茲收到 **This is an acknowledgement of receipt from:**	**Mr. Figueiredo Flavio**		
機構 **Organization:**	Federal University of Minas Gerais		
會議名稱 **Conference Name:**	**The 18th ACM Conference on Information and Knowledge Management**		
會議日期 **Date of Conference:**	**2-6 November 2009**		
會議登記費明細 **Payment detail:**	**Main Conference Registration**	**USD**	**640 .00**
	Workshop with Conference Registration	**USD**	**0 .00**
	Workshop Registration	**USD**	**0 .00**
	Half-day Tutorial Registration	**USD**	**0 .00**
	Two half-day Tutorial Registration	**USD**	**0 .00**
	總計 **Payment Total:**	**USD**	**640 .00**

備注
Remark:

商務會獎中心
Business Travel & MICE Centre

Gmail **Flavio Figueiredo <flaviovdf@gmail.com>**

Registration Confirmation

6 messages

events@regonline.com <events@regonline.com> Thu, Nov 18, 2010 at 4:18 PM
Reply-To: sofmac@fetch.com
To: Flavio Figueiredo <flaviov@dcc.ufmg.br>

ACM International Conference on Web Search and Data Mining (WSDM 2011)

Thank you for registering.

Flavio,

You are confirmed for:
ACM International Conference on Web Search and Data Mining (WSDM 2011)

Review your registration record or make any changes.
Registration Number: 28789566
Name: Flavio Figueiredo

ACM International Conference on Web Search and Data Mining (WSDM 2011)

For a map and directions to the event, click here.

Contact Information:
Email: wsdm2011@gmail.com

Thanks.

We look forward to seeing you at the event.

When

Add to My Calendar

Where
Sheraton Hotel and Towers, Kowloon

Hong Kong,

Privacy Policy

ACM International Conference on Web Search and Data Mining WSDM 2013

Monday, February 04, 2013 - Friday, February 08, 2013

Rome
Viale Manzoni, 1
Rome, 00185
Italy
Event Details
Email: registration@wsdm2013.org
Email Us

Personal Info | WSDM 2013 registration | Merchandise | Checkout | **Confirmation**

Personal Info

Registration ID: 49954282
Registrant: Flavio Figueiredo
Universidade Federal de Minas Gerais
Av. Antônio Carlos 6627 - ICEx - Sala 4010
Pampulha
Belo Horizonte - MG, 31270-010
Brazil
Registration Date: 11/26/2012 5:53 PM
Registrant Type: Student
Status: Confirmed
Work Phone: +55 31 3409 5850
Email: flaviov@dcc.ufmg.br

WSDM 2013 registration

Conference only - ACM/SIG non members

Monday, February 04, 2013 - Friday, February 08, 2013
€275.00

Fees

Fee	Quantity	Unit Price	Amount
Fee			
Conference only - ACM/SIG non members	1	€275.00	€275.00
Subtotal:			€275.00
Total:			€275.00

Transactions

Transaction Type	Date	Amount	Balance
Transaction Amount	11/26/2012	€275.00	€275.00
Online Credit Card Payment (**********5028) Details	11/26/2012	-€275.00	€0.00
Current Balance:			**€0.00**

Payment Method

Payment Method: Credit Card (Visa)

The online credit card payment for this event will be listed on your credit card statement with the name **WSDM 2013 Organizing Comi**.

Gmail **Flavio Figueiredo <flaviovdf@gmail.com>**

CHI 2014 - Registration Confirmation

1 message

CHI 2014 Registration <events@regonline.com> Tue, Feb 4, 2014 at 3:43 PM
Reply-To: CHI 2014 Registration <ylopez@executivevents.com>
To: Flavio Figueiredo <flaviov@dcc.ufmg.br>

Thank you for your CHI 2014 Registration, we look forward to seeing you in Toronto! We are reviewing your record and will be in touch if we see missing information.

Click here to reserve a room at the conference hotel(s). To receive the conference rate you must book by 17 March, 2014.

For those of you paying by check or wire transfer, please be reminded that payment must be made by the early-bird deadline (14 March) or if you have registered at the regular rate, prior to the event.

Please make your check payable to ACM/CHI 2014 and send your check to:

CHI2014
c/o Executive Events
6325 Gunpark Dr., Suite C
Boulder, CO 80301.

If you have elected to pay by wire transfer, please follow these instructions:

Note on your wire documents: CHI 2014, Your Name, Your Registration ID Number

Wire the funds to:
JP Morgan Chase Bank NA
Bank Address: 4 Metrotech Center, 8th Floor, Brooklyn, NY 11245
Account Name: Association For Computing Machinery, Inc
Routing Number (ABA) is 021000021
Swift Code: CHASUS33
Account Number is 301170895565

As a reminder, your registration fee will include the following

Full Conference Registration: If you registered for the Full Conference, you will have access to all conference activities and events Monday, 28 April through Thursday 1 May, 2014 including: paper and poster presentations, job fair, exhibits, and coffee breaks. Also included are the following items: 1 conference reception ticket, conference bag, final program, conference proceedings on USB.

One Day Registration: A one day registration includes: access to all conference activities and events (paper and poster presentations, courses, job fair, exhibits, coffee breaks) for the day purchased. It does NOT include technical materials, proceedings, conference bag, or reception ticket, although these items may be purchased separately.

Course Selection: If you registered for any courses, the $25 charge reflects a material fee per course. Course material will be distributed when you pick up your badge at the conference and will serve as your "ticket" for the session.

Workshop Registration: If you are an invited workshop attendee, workshop fees include coffee breaks and any material provided by the instructor.

Exhibit Booth Staff Registration: This registration pass allows you access to work in your company's exhibit booth only (no courses or other conference activities are included)

Please note the following Terms & Conditions you have agreed to:

Cancellation/Refund Policy:

Cancellations will be accepted until 14 March, 2014. Refunds will be made in full less a $100 cancellation fee. There will be no refunds issued after 14 March, 2014.

Registration Changes & Transfers:

Transfer requests must be made in writing and received by 18 April in order to be processed. Please email chiregistration@executivevents.com outlining the transfer information.

Student Non-Member Registrants - Send proof of student status to **CHI Student Proof.** (ACM or SIGCHI student members do not need to send a copy of their student ID)

International attendees who require a visa please visit for more information.

Warm regards,
CHI 2014 Registration Team

Reference Number: 64875230
Name: Flavio Figueiredo

Attendee Record and Invoice: View details

CHI 2014 Attendee Registration

Date: 4/26/2014 - 5/1/2014 - Add to Calendar
Location:
Metro Toronto Convention Centre
416-585-8000
255 Front St W
Toronto, Ontario M5V 2W6
Canada

Contact Information:

Registration Contact:
Executive Events
1-888-526-1242 Toll Free
1-303-530-4683 International Direct Dial
1-303-530-2691 Fax

Siret : 180 089 047 00047 - APE : 7219Z - TVA N° FR 45 180 089 047

DATE Tuesday, September 23th, 2014

FACTURE / INVOICE

N° 000407/ECML-ILP/14/NAN/COL
Référence à rappeler lors du règlement
Reference to be mentioned with payment

Réf. Commande : EcmlPkdd2014_BT PNC Bank

ADRESSE DE FACTURATION / *INVOICE ADDRESS*

Mr Flavio FIGUEIREDO
Universidade Federal de Minas Gerais
Av. Antônio Carlos 6627
31270-010 Belo Horizonte
BRAZIL

ECML PKDD 2014 & ILP 2014

The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases
International Conference on Inductive Logic Programming

Désignation	Qté	Prix U.	TVA	Montant HT
ECML PKDD student registration	1	308,33	20,00 %	308,33

T.V.A. : L'INRIA est assujetti à la T.V.A.
Délai de règlement : à réception de la facture.

Payment deadline: upon receipt of invoice.

Total HT :	308,33€
dont TVA à 20,00 %	61,67€
Total TTC :	370,00€

TOUT PAIEMENT DOIT ETRE EFFECTUE EN EUROS
ALL PAYMENTS SHOULD BE MADE IN EUROS

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE
Inria Nancy Grand Est
615 rue du jardin botanique
CS 20101
54603 Villers-lès-Nancy cedex
France

Gmail **Flavio Figueiredo <flaviovdf@gmail.com>**

ICML 2016 Registration Confirmation

1 message

Jody Anagnos <events@regonline.com> Tue, May 17, 2016 at 4:16 PM
Reply-To: Jody Anagnos <jody@executivevents.com>
To: Flavio Figueiredo <flaviovdf@gmail.com>

You are receiving this email because you registered for ICML 2016.
If you are having trouble viewing this email, try viewing it in a browser.

ICML @ NYC

International Conference on Machine Learning

ICML 2016

Thank you for registering for the ICML to be held June 19 – 24, 2016 in New York, NY. This email confirms that we have received your registration for the event. You may click on the link below to view your detailed registration record.

View details

Venue and Lodging:
Conference registration does not include your hotel room.

A group rate has been organized with the Marriott Marquis in New York. Rates start at $332/night. You can call the hotel for the group rate at 877-303-0104 or locally at 212-398-1900. You do not need an access code to receive this rate. The web link is

http://icml.cc/2016/?page_id=27.

Questions: If you have any questions regarding your registration, please contact us at 855-290-2441.

See you in New York!

When
6/19/2016 - 6/24/2016

Contact

Email Contact:
jody@executivevents.com
Registration Questions:
855-290-2441

Name: Flavio Figueiredo
Confirmation Number: 92815764

REGISTRATION SUMMARY

Thank you for your registration to WWW2016.

We will proceed with the verification of your payment information.
We reserve the right to cancel the registration without notice in the case of inconsistencies or anomalies.
Please keep in mind that payment receipts will only be provided at the conference in April.

Event name : WWW2016
Date : April 11th - 15th, 2016
Location : Montréal - CANADA

First name : Flavio
Last name : Figueiredo
Title : Researcher
Affiliation : IBM Research Brazil
Accepted paper number : 745

Registration : Regular/Full Pass
Payment method : AMEX
Bill number : 448
Date & time : 2016-02-03 11:22:23
Authorisation number : 120197

Total before taxes : 995$ CAD
Discount : %
Gala Dinner : 0$ CAD
Total after discount : 995$ CAD
GST(5%)(108161159RT0001) : 49.75$ CAD
QST(9.975%)(1006120021TQ0012) : 0$ CAD
Total paid : 1044.75$ CAD

Email : flaviovdf@gmail.com
Phone : +5521969883152
Address : Arnaldo Quintela, 10, apt 702
Country : Brazil
City : Rio de Janeiro
Province : RJ
Postal code : 22280070

Do you have food allergies or intolerances? : No

I agree to receive news releases and updates from UQAM and the WWW2016 team : Yes

Please print this page for your documents. A confirmation email will be sent to you.

If you experience technical problems while completing the registration form contact us by email at registration@www2016.ca or by phone +1 514 987 6516.

Thank you and see you at the conference!