

# Ranking Universities Based on Career Outcomes of Graduates

Navneet Kapur<sup>a</sup> Nikita Lytkin<sup>b</sup> Bee-Chung Chen<sup>b</sup>  
Deepak Agarwal<sup>b</sup> Igor Perisic<sup>b</sup>

<sup>a</sup>GoFundMe

<sup>b</sup>LinkedIn Corporation

Apresentado por Eduardo Elias Ribeiro Junior  
[edujrrib@gmail.com](mailto:edujrrib@gmail.com)

16 de novembro de 2016

# Sumário

1. Publicação
2. Artigo
3. Considerações

1

# Publicação

- ▶ **Artigo: Ranking Universities Based on Career Outcomes of Graduates**  
Publicado em agosto/2016  
(Elaborado em 2014)

- ▶ **Periódico: In Proceedings of the 22nd ACM SIGKDD (KDD '16)**  
Apresentado dia 15 agosto (Paper Applied Data Science Tracks)

- ▶ **Autores: Kapur N., Lytkin N., Chen B., Agarwal D., Perisic I.**

- ▶ **Divulgação: Resumo e vídeo promocional no 22nd ACM SIGKDD (KDD '16)**  
<http://www.kdd.org/kdd2016>

## Ranking Universities Based on Career Outcomes of Graduates

Naveet Kapur\*  
GoFundMe  
Redwood City, CA, USA  
nkapur@gofundme.com

Nikita Lytkin  
LinkedIn Corporation  
Mountain View, CA, USA  
nlytkin@linkedin.com

Bee-Chung Chen  
LinkedIn Corporation  
Mountain View, CA, USA  
bchen@linkedin.com

Deepak Agarwal  
LinkedIn Corporation  
Mountain View, CA, USA  
dagarwal@linkedin.com

Igor Perisic  
LinkedIn Corporation  
Mountain View, CA, USA  
iperisic@linkedin.com

### ABSTRACT

Every year, millions of new students enter higher educational programs. Publicly available rankings of academic programs play a key role in prospective students' decisions regarding which universities to apply to and attend in. While surveys indicate that majority of freshmen enter college to get good jobs after graduation, established methodologies for ranking universities rely on indirect indicators of career outcomes such as reputational assessments of the universities among academic peers, acceptance and graduation rates, learning environment, and availability of research funding. In addition, many of these methodologies rely on arbitrary choices of weighting factors for the different ranking indicators, and suffer from lack of analyses of statistical stability. In this paper, we address these challenges indirectly by developing a novel methodology for ranking and recommending universities for different professions on the basis of career outcomes of professionals who graduated from those schools. Our methodology incorporates a number of techniques for achieving statistical stability, and represents a step towards personalized educational recommendations based on interests and ambitions of individuals. We have applied this methodology on LinkedIn's Education Graph data of over 400 million professionals from around the world. The resulting university rankings have been made available to the public and demonstrate there are new valuable insights to be gleaned from professional career data on LinkedIn.

### Keywords

Educational Recommendations; University Rankings; Company Rankings; Statistics

\*Work done while at LinkedIn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the name and full address on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. Any other forms of copying, reprinting or redistribution requires explicit permission from the publisher. For more information, contact the publisher.

© 2016 ACM. August 13 - 17, 2016, San Francisco, CA, USA.  
© 2016 Copyright held by the owner(s). Publication rights licensed to ACM.  
978-1-4503-2202-2/16/08...\$15.00  
DOI: <http://dx.doi.org/10.1145/2908072.2908701>

### 1. INTRODUCTION

Millions of high school students (2 million in the US alone in 2015) apply for higher education every year. The search starting college students, the application process starts with selecting schools to apply to based on the student's career interests and academic performance. A recent survey [14] conducted by Higher Education Research Institute on hundreds of thousands of entering freshmen found that 80% of freshmen attend college to get a good job while 91% state the desire to be very well off financially as one of their personal goals. Thus, the ability to recommend schools on the basis of careers and eventually in a personalized manner has potential to provide tremendous value.

On LinkedIn.com, millions of professionals across the world enter rich information about their careers. We propose to leverage this valuable data and convert it into actionable information for LinkedIn's youngest users and drive change through actionable insights at higher education institutions. In this paper, we for the first time present in full detail our novel approach to ranking and recommending universities given a choice of a profession, on the basis of career outcomes of professionals who graduated from those schools.

The notion of ranking universities in itself is not a new concept. Ranking agencies such as US News & World Report, Times Higher Education and QS produce university lists each year – overall and by major. These rankings assess schools on the basis of indicators such as percentage of accepted students who go on to enroll, graduation rates, average SAT scores in addition to research inclusion indicators like reputational assessments by peers at other universities. However, we believe that a more objective way to evaluate a degree program with respect to career outcomes is to measure performance of its graduates in industry. We achieve this by first developing an approach for identifying most desirable companies for different professions. We then present a methodology for ranking universities based on the rates at which their graduates are able to obtain jobs at these desirable companies in a given profession. Such data-driven rankings are a complete data product which require careful consideration for a number of statistical aspects including representation bias and statistical robustness of ranking. In following sections, we present our methodology and the practices used for correcting potential representation biases

2

# Artigo

1. Introdução
2. Metodologia
3. Aplicação
4. Conclusões

2.1

Artigo

**Introdução**

# Ranking de Universidades

- ▶ Primeiro passo de um estudante do ensino médio que deseja se aplicar ao ensino superior é escolher as possíveis universidades;
- ▶ A classificação das universidades tem papel fundamental na escolha dos aspirantes ao ensino médio;
- ▶ Dentre milhares de estudantes do primeiro ano do ensino superior 88% desejam obter um bom emprego enquanto 81% tem como objetivo pessoal a estabilidade financeira.



## Rankings atuais

São geralmente baseados em

- ▶ Número de matriculados;
- ▶ Valor de financiamento de pesquisas;
- ▶ Produção acadêmica;
- ▶ Reputação por pares acadêmicos.

Alguns desses itens são complicados pois

- ▶ São subjetivos;
- ▶ São auto-influenciados

Além disso os rankings não incorporam o desempenho profissional dos graduados.

# Proposta de Ranking

O [LinkedIn.com](https://www.linkedin.com) conta com milhões de perfis com informações sobre suas carreiras profissionais e acadêmicas.

Incorporar os dados do LinkedIn dos graduados com perfil no LinkedIn para ranquear as universidades

# Objetivos

- ▶ Ranquear as universidades com base no desempenho de seus estudantes egressos na indústria (mercado de trabalho);
  - ▶ Identificar as empresas mais desejáveis para diferentes profissões;
  - ▶ Ranquear as universidades com base na proporção de egressos que estão empregados em empresas desejáveis;
- ▶ Incorporar aspectos estatísticos para representação de viés e robustez dos resultados;

2.2

Artigo

**Metodologia**

## Visão Geral

Passos genéricos da metodologia:

1. Classificar as empresas mais desejáveis;
2. Escolher as K empresas mais desejáveis;
3. Verificar a proporção de graduados de uma universidade que são empregados nas K empresas mais desejáveis;
4. Ranquear as universidades com base nessa proporção.

Algoritmos propostos:

**CompanyRanker:** Ranquear as empresas para mensurar o desempenho profissional dos graduados de uma universidade;

**SchoolRanker:** Ranquear as universidades com base no desempenho de seus graduados.

# Esquematização

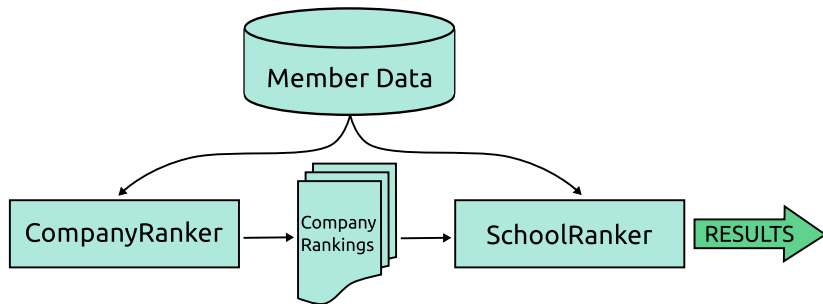


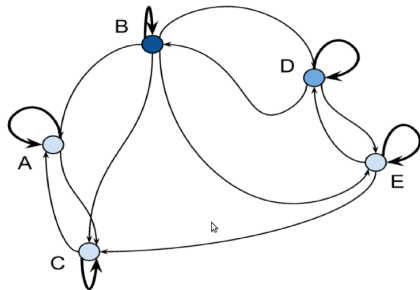
Figura : Visão geral da arquitetura do algoritmo

## CompanyRanker - Talent Flow Graph

- ▶ Nodos representam as empresas;
- ▶ Arestas representam as transições entre companhias;
- ▶ Auto-loops são incluídos para todas as empresas, com pesos da forma

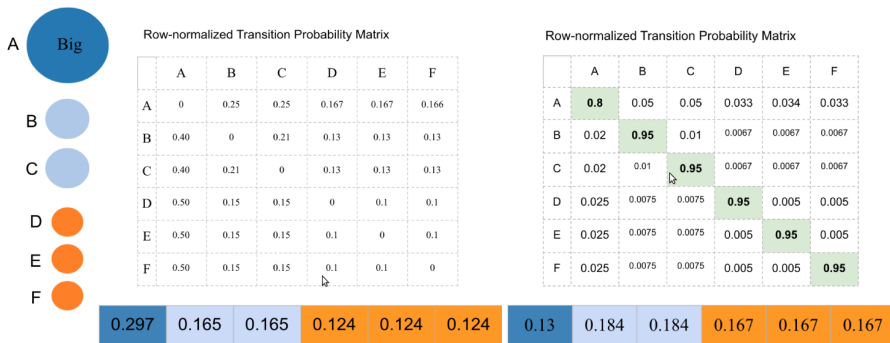
$$\sum_{x \in R_P(A)} t_P(x, A)$$

$R_P(A)$  é o conjunto de funcionários da empresa A com atuação maior que a mediana de atuação da profissão P  
 $t_P(x, A)$  é a atuação do funcionário A dividido pela mediana de atuação da profissão.



**Figura :** Ilustração de um Talent Flow Graph para uma profissão.

# CompanyRanker - Probabilidades de transição



**Figura :** Matrizes de probabilidades de transição representando TFG's desconsiderando a retenção (esquerda) e considerando a retenção (direita). Os resultados abaixo correspondem ao escore do PageRank.



## SchoolRanker - Escore de sucesso da universidade

$$r = \sum_{x \in X} m(x) \frac{q(x)}{p(x)} \quad s = \sum_{x \in X} n(x) \frac{q(x)}{p(x)} \quad \theta = \frac{s}{r}$$

Em que:

- $X$  o conjunto de combinações dos atributos gênero, ano de graduação, grau de escolaridade e universidade;
- $p(x)$  a proporção de graduados do LinkedIn com atributos  $x$ ,  $x \in X$ ; e
- $q(x)$  a proporção de graduados de uma base externa com atributos  $x$ ,  $x \in X$ .
- $m(x)$  número de graduados com atributos  $x$ ,  $x \in X$ , em empresas relevantes para dada profissão.
- $n(x)$  número de graduados com atributos  $x$ ,  $x \in X$ , empregados em empresas top para dada profissão.

## SchoolRanker - Escore de sucesso da universidade

Alguns problemas com o método:

- ▶ Informações incorretas nos perfis;
- ▶ Viés de representação também influencia na classificação das empresas;
- ▶ Não há uma base de dados sólida dos empregados de todas as empresas.

Propostas para minimizar ou considerar esses problemas:

- ▶ Rigoroso teste de spam, usados somente perfis aprovados no teste;
- ▶ Técnica de reamostragem Monte Carlo

# SchoolRanker - Reamostragem Monte Carlo

Ideia:

- ▶ Usar um grande número de *conjuntos perturbados* para que o viés de representação seja diluído pela aleatoriedade.

Reamostragem:

- ▶ Dado o ranqueamento das empresas e K número de empresas top:
  - ▶ Substitua um subconjunto das empresas top (5% ou 10%) por empresas não-top;
  - ▶ A seleção das empresas não-top substitutas pe realizada de forma proporcional a sua medida de desejabilidade (escore PageRank).

## SchoolRanker - Classificação

- ▶ Para cada conjunto perturbado ranquea-se as universidades com base em  $\theta$  e armazena-se sua posição.
- ▶ O ranqueamento é então realizado com base no percentil de 95% das posições ocupadas em cada conjunto perturbado.
- ▶ Caso ocorra empate olha-se o percentil de 75%, caso persista o empate as universidades são classificadas na mesma posição.

## Escolha do K

- ▶ Minimizar as mudanças nos ranks devido a únicas empresas

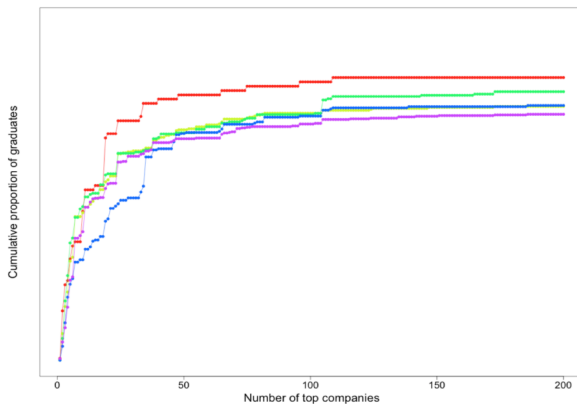


Figura : Distribuição acumulada empírica dos graduados de cinco universidades empregados nas top empresas (eixo x).

## Escolha do K

- ▶ Defini-se um grid de K
- ▶ Para cada K reamostramos os conjuntos perturbados
- ▶ Calcule a concordância média entre pares

$$\sum_N \frac{\#(\cap \mathcal{U}_N)}{N}, \text{ para } N = \{3, 5, 10, 25\}$$

Exemplo para  $N = 3$ .

Posição	Classificação com base nas reamostras							
1ª	Ua	Uc	Ue	Ua	Ug		Uf	
2ª	Uc	Ua	Uf	Ub	Ua	...	Uc	→ $\frac{1}{3}$
3ª	Uf	Ub	Ua	Ug	Uh		Ua	

## Escolha do K

- ▶ Deve-se escolher o K que maximiza a concordância.

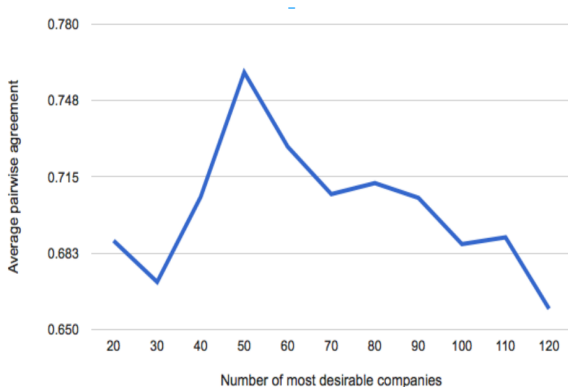


Figura : Exemplo da concordância média entre os conjuntos reamostrados para diferentes escolhas de K número de empresas top.

2.3

Artigo  
**Aplicação**



- ▶ Aplicação a dados do LinkedIn (LinkedIn's Economic Graph data);
- ▶ Base com mais de 400 milhões de profissionais;
- ▶ Discutidos apenas duas profissões Investment Bankers e Software Developers at Startups considerando os EUA;
- ▶ Uma interface para alunos estava disponível em <https://www.linkedin.com/edu/> (recurso foi descontinuado).

# Ranqueamento das empresas

**Table 1: Top 20 companies for Investment Bankers (in alphabetical order).**

Company
American Express
Bank of America
Citi
Credit Suisse AG
Deutsche Bank
Goldman Sachs
Jefferies
JPMorgan Chase & Co
KPMG
Morgan Stanley
New York Life Insurance Company
Nomura
Piper Jaffray
Raymond James Financial, Inc.
RBC Capital Markets
TIAA-CREF
UBS AG
USAA
Wells Fargo
William Blair

**Table 2: Top 20 startups for Software Developers (in alphabetical order)**

Company	Type
Airbnb	Hospitality
Braintree	Payments
Clever	Education
CoreOS	Operating Systems
Couchbase	Database Systems
Counsyl	Healthcare
Dropbox	Cloud Storage
Fitbit	Fitness
Flatiron Health	Healthcare
GitHub	Software Development Tools
Lyft	Transportation and Delivery
Oscar Health	Healthcare
PlanGrid	Industrial Design Software
PlanSource	HR Management Systems
Riot Games	Entertainment
Sovrn Holdings	Online Advertising
Square	Payments
Tapad	Online Advertising
Uber	Transportation and Delivery
Zen Payroll	Payroll Management

↳

# Ranqueamento das universidades

**Table 3: Top ranking undergraduate programs for Investment Bankers**

School	75th Percentile	95th Percentile	Final Rank
Georgetown University	1	2	1
University of Pennsylvania	3	4	2
Yale University	4	7	3
Columbia University in the City of New York	7	9	4
Duke University	7	9	4
Princeton University	9	11	6
New York University	11	11	7
Wellesley College	4	12	8
Cornell University	10	12	9
Dartmouth College	10	13	10

# Ranqueamento das universidades

**Table 4: Top ranking undergraduate programs for Software Developers at Startups**

School	75th Percentile	95th Percentile	Final Rank
Stanford University	1	1	1
Massachusetts Institute of Technology	2	2	2
University of California, Berkeley	3	4	3
Carnegie Mellon University	4	5	4
Brown University	6	7	5
Cornell University	6	10	6
The University of Texas at Austin	13	15	7
Rochester Institute of Technology	14	18	8
University of Illinois at Urbana-Champaign	15	18	9
University of Maryland College Park	7	19	10

2.4

Artigo

**Conclusões**

- ▶ Nova metodologia de ranqueamento baseada em desempenho de carreira (mais fidedigno com o interessante dos estudantes);
- ▶ Em contraste com outros ranqueamentos não há a utilização de avaliações feitas por profissionais;
- ▶ O ranqueamento conta com procedimento para garantir robustez dos resultados.

### Desenvolvimentos futuros

- ▶ Usar dados mais granulares sobre as transições de emprego;
- ▶ Incorporar dados de remuneração salarial;
- ▶ Incorporar dados adicionais de especialistas da indústria.



3

# Considerações

- ▶ Grande contribuição do artigo é com relação a motivação;
- ▶ Ótima ideia para considerar as empresas no ranqueamento (PageRank);
- ▶ Algumas escolhas não foram bem esclarecidas (limite inferior de  $\theta$ , percentis 95% e 75%);
- ▶ Poderia se aproveitar melhor os resultados do PageRank (a metodologia considera somente a ordenação dos escores e não os escores);
- ▶ Pouco reportado os resultados da aplicação.



## Referências

-  N. KAPUR, N. LYTKIN, B. CHEN, D. AGARWAL, AND I. PERISIC. (2016). Ranking Universities Based on Career Outcomes of Graduates. **In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)**. ACM, New York, NY, USA, 137-144. DOI: <http://dx.doi.org/10.1145/2939672.2939701>
-  L. PAGE, S. BRIN, R. MOTWANI, AND T. WINOGRAD. (1999). The pagerank citation ranking: Bringing order to the web. **Technical Report 1999-66**. Stanford InfoLab

Obrigado!