

# josedezas

software engineering and data science



## contact

Porto, Portugal  
+351 939 828 043

jose Luis devezas@gmail.com  
(preferred method of contact)

<http://www.josedezas.com>  
[linkedin://jldevezas](https://www.linkedin.com/in/jldevezas)

## languages

portuguese (native)  
english (fluent)

## programming

♥ Scala Python R SQL  
JavaScript D3.js  
Sigma.js HTML5 CSS3  
Java C# Gremlin

## education

2003–2010 **Masters** in Informatics and Computing Engineering FEUP, Universidade do Porto  
Classification: 14 out of 20 (fourteen out of twenty)

### *Link Ecosystem of the Portuguese Blogosphere*

This thesis explored the blogosphere as an ecosystem of blogs comprised not only of a set of posts and their individual content, but also of the interactions established through the hyperlinks connecting the posts. Based on this study, we discovered a correlation between the blog's centrality and the length and number of posts.

Thesis Classification: 18 out of 20 (eighteen out of twenty)

2000–2003 **Technological Specialization** in Informatics Colégio de Gaia  
Classification: 18 out of 20 (eighteen out of twenty)

## experience

2013–2015 **INTERRELATE** Porto, Portugal  
*Software Engineer / Data Scientist*

### *Data Science:*

- Collection and analysis of several types of data, mainly originating from the web, particularly from social media.
- Highly proficient in R, Python and SQL (from data collection to reporting), focusing on tasks such as topic tracking, outlier and anomaly detection, statistical smoothing, regression, classification, clustering and network analysis.
- Skilled in dynamic reporting using R Markdown and brew.
- Developed several data visualization widgets using technologies like d3.js and sigma.js.
- Analyzed social networks and other graphs using Gephi and R packages such as igraph.

### *Software Engineering:*

- Software design of web applications and analytics pipelines.
- Technology evaluation and selection.
- Supporting the team with new technologies or languages.
- INTERRELATE Backoffice:
  - Devising strategies for automatic and periodical data collection and information extraction from unstructured user-generated content;
  - Systems integration, including team coordination (4 person project).

### *Programming:*

- Worked in nearly every company project, using various different languages, including Scala, Java, R, Python, Perl, PHP, JavaScript, SQL (PostgreSQL & PL/pgSQL, H2, SQLite, etc.).
- Developed several web applications and web services using frameworks such as Play Framework or Flask.
- INTERRELATE Backoffice:
  - Web crawler development;
  - Development of data visualization widgets for internal statistics;
  - Front-end design and development.

### *Sysadmin:*

- Manage all of the company servers, including security, backups and product deploys.
- Maintain firewall rules and SMTP server.
- DNS configurations (including SPF, for SPAM control).
- Nginx configurations.
- Company services maintenance (internal and external).
- Log monitoring and bug reporting.
- Real-time e-mail log error monitoring for critical systems and web service mobile monitoring for downtime or connection issues.
- Support Ticket System configuration.
- PayPal payment services.

2012–2013	<b>LABORATÓRIO SAPO/U.PORTO</b> <i>Researcher / Developer</i> Research and development of music discovery and recommendation techniques supported on a graph-based neighborhood approach, as well as on latent factor models based on matrix factorization algorithms. Development of methodologies to combine knowledge from different dimensions, including not only user profile, content and context features, but also social features, such as community structure. Related tasks required work on data mining, multimodal network analysis and community detection, and information retrieval.	Porto, Portugal
2011–2012	<b>CRACS/INESC TEC</b> <i>Researcher / Developer</i> Enhancement, adaptation and implementation of state of the art algorithms for network analysis, data mining and data visualization, in the context of intelligent information systems for cyber journalism, as part of the Breadcrumbs Project (UTA-Est/MAI/0007/2009).	Porto, Portugal
2010–2011	<b>LABORATÓRIO SAPO/U.PORTO</b> <i>Researcher / Developer</i> Work in the areas of data mining, network analysis, community detection, information retrieval, real-time data analytics and data visualization.	Porto, Portugal
2002–2003	<b>NÚCLEO PORTUGUÊS DE FREEBSD</b> <i>Documentation / Programming</i> Translation of the official FreeBSD documentation to portuguese, creation of how-to guides in portuguese, porting open source applications to FreeBSD and programming and content management of the hint subsystem.	Coimbra, Portugal

## awards

2012	<b>Certificate of Merit</b> The 2012 IAENG International Conference on Data Mining and Applications Awarded to the best papers published in the conference.
------	--

## communication skills

2013	<b>Oral Presentation</b>	DEI-FEUP
	Did a short presentation on the Juggle project for the department, in the context of DEI Talks.	
2013	<b>Oral Presentation</b>	SAPO
	Presented the Juggle project, including the developed graph-based recommendation algorithms, for the SAPO technical team.	
2012	<b>Oral Presentation</b>	International Conference on Knowledge Discovery and Information Retrieval
	Presented an interactive news clips visualization tool for applications in journalistic research and knowledge discovery.	
2012	<b>Oral Presentation</b>	IAENG International Conference on Data Mining and Applications
	Presented a methodology, based on community detection in multidimensional networks, for the creation of news context from a folksonomy of web clipping.	
2011	<b>Poster</b>	International AAAI Conference on Weblogs and Social Media
	Presented research work on using the h-index to estimate blog authority.	
2010	<b>Oral Presentation</b>	Laboratório SAPO/U.Porto
	Presented Ciclope project for the Portugal Telecom's CEO and the media, as part of the inaugural presentation of Laboratório SAPO/U.Porto.	
2010	<b>Poster</b>	KDD Workshop on Social Media Analytics
	Presented the blog popularity study conducted during my Masters degree.	

## interests

**professional:** data analysis, social network analysis, community detection, web development, data visualization **personal:** music, television, guitar, cooking

**additional information:** Please check out my personal website, where I've shared some of my research output, mostly on the topic of community-driven music discovery:  
<http://josedevezas.com/#/individual-listening-behavior>.

## publications

### articles in peer-reviewed journals

The Community Structure of a Multidimensional Network of News Clips

José Luís Devezas, Álvaro Reis Figueira

International Journal of Web Based Communities 9.3 (2013) pp. 411–429. 2013

Finding Language-Independent Contextual Supernodes on Coreference Networks

José Devezas, Álvaro Figueira

IAENG International Journal of Computer Science 39.2 (2012) pp. 200–207. 2012

### book chapters

Álvaro Figueira, José Devezas, Nuno Cravino, Luis-Francisco Revilla. "Creating and Analysing a Social Network Built From Clips of Online News". In: *Information Systems and Technology for Organizations in a Networked Society*. IGI Global, 2012.

### international peer-reviewed conferences/proceedings

Large-scale Crossmedia Retrieval for Playlist Generation and Song Discovery

Filipe Coelho, José Devezas, Cristina Ribeiro

Proceedings of the 10th International Conference in the RIAO series (OAIR 2013), 2013

### Using the Overlapping Community Structure of a Network of Tags to Improve Text Clustering

Nuno Cravino, José Devezas, Álvaro Figueira

*Proceedings of the 23rd ACM Conference on Hypertext and Social Media (HT 2012)*, 2012, Milwaukee, WI, USA

### Creating News Context From a Folksonomy of Web Clipping

José Devezas, Henrique Alves, Álvaro Figueira

*Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2012 (IMECS 2012)*, 2012, Hong Kong

### Studying a Personality Coreference Network in a News Stories Photo Collection

José Devezas, Filipe Coelho, Sérgio Nunes, Cristina Ribeiro

*Lecture Notes in Computer Science: Proceedings of the 34th European Conference on Information Retrieval (ECIR 2012)*, 2012, Barcelona, Spain

### Interactive Visualization of a News Clips Network: A Journalistic Research and Knowledge Discovery Tool

José Devezas, Álvaro Figueira

*Proceedings of the 4th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2012)*, 2011, Barcelona, Spain

### Using the H-index to Estimate Blog Authority

José Devezas, Sérgio Nunes, Cristina Ribeiro

*Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*, 2011, Barcelona, Spain

### Studying Blog Features over Link Popularity

José Devezas, Cristina Ribeiro, Sérgio Nunes

*Proceedings of the First Workshop on Social Media Analytics (SOMA 2010), held in conjunction with the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010)*, 2010, Washington, DC, USA

### FEUP at TREC 2010 Blog Track: Using h-index for blog ranking

José Devezas, Sérgio Nunes, Cristina Ribeiro

*The Nineteenth Text REtrieval Conference Proceedings (TREC 2010)*, 2010, Gaithersburg, MD, USA

## thesis reports

José Devezas. “Link Ecosystem of the Portuguese Blogosphere”. Master’s Thesis. Faculdade de Engenharia da Universidade do Porto, 2010, pp. 1–50.

## research reports

### An Overview of the Graph Database Paradigm - Breadcrumbs Project

José Devezas

*Tech. rep.*, 2011

### Overlapping Community Detection - Labs SAPO/UP Report

José Devezas, Sérgio Nunes, Cristina Ribeiro

*Tech. rep.*, 2011