

# Aplicação de modelos para redes sociais em bases textuais

Bolsista: Pedro Henrique Sodré Puntel

Orientador: Gustavo da Silva Ferreira

maio/2019

## 1 Objetivo do Projeto

Este projeto de iniciação científica analisa dados de artistas a fim de permitir uma visualização da sua estrutura de popularidade com base em um ranqueamento semanal produzido pelo site *www.billboard.com*. Utilizando modelos de espaços latentes para redes sociais, investigou-se como se davam as associações entre artistas que já ocuparam e feriram ranqueamento considerado. Para tal, três modelos foram considerados, sendo o primeiro deles responsável por modelar as citações entre artistas, identificando ainda claros agrupamentos dos mesmos a partir dos padrões latentes

## 2 Relevância e Contextualização do Projeto

A popularização das redes sociais traz consigo a necessidade de avanços na modelagem das relações entre indivíduos que compõem uma rede. Não obstante, os objetivos das análises de redes (duos) compõem papel central, bem como a sua identificação e de grupos subjacentes na rede.

Sob este prisma, o avanço dos modelos estatísticos tem permitido análises a cerca de diversos aspectos de uma única rede, como por exemplo: opiniões, hábitos, desejos e graus de relacionamento. A aplicação destes modelos.

Não obstante, em paralelo ao avanço computacional, grandes bases de dados textuais encontram-se agora disponíveis na Internet, sendo a maioria destas encontradas de forma não-estruturada. De acordo com *Gantz et al. (2008)*, em 2007, cerca de 90% de toda a informação contida no universo digital era composta por dados não estruturados como textos, imagens e vídeos. Neste sentido, a habilidade de coletar, estruturar e analisar parte dessa grande quantidade de dados é essencial, sendo então o principal objeto de estudo durante o início deste projeto.

Para lidar com a não-estruturação das bases de dados textuais, diversas ferramentas de *Web scraping* foram desenvolvidas nos últimos anos, permitindo novos níveis de integração dos modelos matemáticos/estatísticos para com as informações latentes disponíveis na Internet.