

**Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**

---

## Desenvolvimento de um algoritmo de programação para análise das palavras de Normas Técnicas de Geotecnia

Iury Costa Barros <sup>1</sup>

Instituto Federal de Goiás, Aparecida de Goiânia, GO

Sávio Aparecido dos Santos Pereira <sup>2</sup>

Instituto Federal de Goiás, Aparecida de Goiânia, GO

Arlam Carneiro Silva Junior <sup>3</sup>

Instituto Federal de Goiás Aparecida de Goiânia, GO

Diogo Gonçalves Dias <sup>4</sup>

Instituto Federal de Goiás, Aparecida de Goiânia, GO

### 1 Introdução

Ciência de Dados é a área responsável pela coleta, análise e compreensão de grandes dados, atendendo as necessidades da sociedade atual. A Ciência de Dados é uma área que incorpora elementos de várias outras áreas e utiliza teorias de áreas básicas já consolidadas, como as Engenharias [4]. A produção de relatórios da análise de dados aperfeiçoa a tomada de decisões nas diversas áreas, minimizando o erro e racionalizando a resposta. A necessidade de normatização de técnicas e metodologias de Engenharia Civil contribuiu para que uma grande quantidade de dados fosse acumulada colaborando para o desenvolvimento da Engenharia em termos, principalmente, de segurança e qualidade. [3] explicam que, no princípio, à ausência dessas normas restringia o uso de novos processos e quando utilizados, resultavam em experiências desastrosas. Porém, esta grande quantidade de informações acumuladas favorece uma maior complexidade da relação entre as técnicas envolvidas tornando-se a prática mais desafiante. Diante deste cenário, a Ciência de Dados pode contribuir com metodologias que ajudem a entender a relação entre as normas, e a sua aplicação em áreas específicas da Engenharia Civil, além de indicar formas para facilitar as relações entre as normas e a elaboração de novos conteúdos. Para a criação do grafo que indica a força entre as normas técnicas é necessário antes entender termos como redes complexas e medidas de centralidade. As redes complexas são redes com estrutura irregular, complexa e dinamicamente evoluindo no tempo [2]. Ou seja, são redes que não seguem um padrão regular. No âmbito da teoria dos grafos e da análise de redes, surgem

---

<sup>1</sup>iucosbarros@gmail.com

<sup>2</sup>savioaparecido1@gmail.com

<sup>3</sup>arlamjr@gmail.com

<sup>4</sup>diogodias.gd@gmail.com

medidas de centralidade para quantificar a importância de nós em redes sob diferentes aspectos [1]. Nesta teoria, os grafos são representados por  $G(V, E)$ , onde  $V$  é um conjunto não vazio de objetos denominados vértices, ou nós, e  $E$  é um conjunto de pares não ordenados de  $V$ , denominados arestas.

## 2 Objetivos e Resultados

O presente trabalho consiste na avaliação das relações entre palavras obtidas de uma rede de normas técnicas ligadas a Engenharia Civil, com foco na área da Geotecnia. Foram selecionadas normas técnicas da área de Geotecnia para serem avaliadas, e desenvolvido um algoritmo para buscar e analisar dados específicos. As normas foram caracterizadas conforme o tema, tamanho e aplicação prática. Em seguida, utilizando esses dados, é possível a geração de um grafo (são formados por dois conjuntos, os vértices e os arcos, onde cada arco está associado a dois vértices) mostrando as forças entre as palavras obtidas das normas técnicas. Neste processo de análise de rede, sua estrutura é então analisada e, em seguida, contextualizada e interpretada. O algoritmo desenvolvido realiza a leitura dos dados textuais e os importa para serem analisados. Inicialmente foi realizada uma limpeza de dados, retirando palavras que são irrelevantes para a análise, as chamadas stop words (como por exemplo, "de", "para", "uma", entre outras), com o auxílio da biblioteca NLTK do Python. Após isso, foi realizada uma busca por uma palavra específica e a sua posição no texto armazenado em uma lista, cada vez que aparece. Com isso foi possível fazer a busca por palavras vizinhas e também armazená-las em uma lista. Obtidos esses dados, foi realizada então uma contagem da quantidade de vezes que os vizinhos aparecem no texto, obtendo assim as palavras de maior relevância para o texto. Com a criação do grafo a partir da relação entre essas palavras, é possível observar de melhor maneira o comportamento delas no texto, sendo possível assim fazer uma avaliação, de forma estatística, das relações entre os conceitos e as palavras. Sendo assim, é possível a sugestão de novas revisões, junções de conteúdo ou até mesmo sugerir focos específicos para estudo, de forma a melhorar e simplificar o que for possível.

## Referências

- [1] T. G. Almeida, F. G. Nakamura e E. F. Nakamura. Uma Abordagem Baseada em Redes Complexas para Análise de Depoimentos Legais, *36º Concurso de Trabalhos de Iniciação Científica (CTIC 2017)*, Vol. 36, 2017.
- [2] S. Boccaletti, V. Latora, Y. Moreno and M. Chavez. Complex networks: Structure and dynamics, *Physics reports*, 134:175-308, 2006.
- [3] O. M. Gonçalves, V. M. Jhon, F. A. Picchi e N. M. Sato. Normas técnicas para avaliação de sistemas construtivos inovadores para habitações, *Coletânea habitare*, 12:42-53, 2003.
- [4] F. Porto e A. Ziviani. Ciência de dados, *III Seminário de Grandes Desafios da Computação no Brasil*, 2014.