

# Avaliação da qualidade de textos por meio de Redes Complexas

Ronan de Souza Castro, Alexander Cristian Menten Roncalli

Departamento de Ciência da Computação – Universidade Federal de São João Del Rei

**Abstract.** *This article discusses how complex networks can be used in texts, so that it is possible to analyze their textual quality, analyzing their agglomeration coefficient and their dissertation level from the point of view of networks.*

**Resumo.** *Este artigo descreve como pode ser feito o uso de redes complexas em textos, para que seja possível analisar a qualidade textual destes, analisando seu coeficiente de aglomeração e seu nível dissertativo no ponto de vista de redes.*

## 1. Introdução

O nosso trabalho consiste em analisar a qualidade do texto baseado no coeficiente de aglomeração do texto, este tipo de análise pode ser utilizado para classificar a legibilidade de um texto, a variedade de vocabulário do texto e poderia ser usado para classificar a qualidade do texto.

## 2. Objetivo

Nosso objetivo é conseguir analisar o texto a partir de um coeficiente de aglomeração de e em seguida tentar encontrar uma correlação de notas de redação serem inversamente proporcionais ao coeficiente de aglomeração. Para isso utilizamos uma abordagem de redes complexas para analisar o texto.

## 3. Metodologia

### 3.1- Origem dos Dados

Como o trabalho foca em analisar a qualidade do texto, podemos escolher qualquer texto, porém para obter uma base inicial utilizamos o poema “no meio do caminho” de Drummond de Andrade. Os outros textos nós pegamos de um banco de redações da uol, que tem notas dedicadas a cada texto.

### 3.2- Processamento dos Dados

Os dados utilizados no programa são as palavras do texto, só que para obtermos uma análise mais específica utilizamos uma biblioteca de Python (nltk) para remover palavras não significativas, como conjunções, pontuação e advérbios, para que possamos utilizar as palavras que tem um valor significativo na análise.

### 3.3- Modelagem da Rede

A rede foi modelada a partir das palavras restantes do texto após o processamento dos dados, das quais cada palavra foi definida como um vértice da rede, e as arestas entre os vértices representam uma relação de adjacência entre duas palavras no texto, não importando a ordem delas, caso durante a montagem da rede as mesmas duas palavras coincidam na adjacência novamente, é adicionado mais um ao peso da aresta entre os vértices dessas palavras.

### 3.4- Análise Descritiva Básica

Com a rede já formada o programa busca encontrar cliques, para que em seguida possamos calcular o coeficiente de aglomeração do texto.

### 3.5- Análise Coeficiente de Aglomeração

O cálculo do coeficiente de aglomeração é feito a partir da média de peso das arestas dos cliques e do número total de vértices.

$$\left( \frac{(\sum \text{peso das arestas dos cliques})}{\text{número de cliques}} \right) \\ \text{número de vértices}$$

Figura 1: equação do coeficiente de aglomeração

Desta forma o coeficiente indicará que o texto é menos coeso se houver uma grande repetição das mesmas palavras, ou seja, se o somatório dos peso das arestas for muito maior que o número de cliques. e o coeficiente aumenta ainda mais se estas variáveis acontecerem ao mesmo tempo que não existe um grande vocabulário no texto, resultando realmente em uma aglomeração, pois há um conglomerado enorme de palavras inter relacionando em uma rede pequena de palavras.

## 4. Discussão

Nesta parte gostaríamos de discutir sobre os resultados que encontramos em nossos testes.

#### 4.1- No meio do Caminho

No meio do caminho tinha uma pedra  
tinha uma pedra no meio do caminho  
tinha uma pedra  
no meio do caminho tinha uma pedra.

Nunca me esquecerei desse acontecimento  
na vida de minhas retinas tão fatigadas.  
Nunca me esquecerei que no meio do caminho  
tinha uma pedra  
tinha uma pedra no meio do caminho  
no meio do caminho tinha uma pedra

Figura 2: No meio do caminho

No teste feito neste texto podemos analisar que ele é pouco dissertativo, ou seja, ele tem um coeficiente de aglomeração muito alto, podemos observar na imagem abaixo:

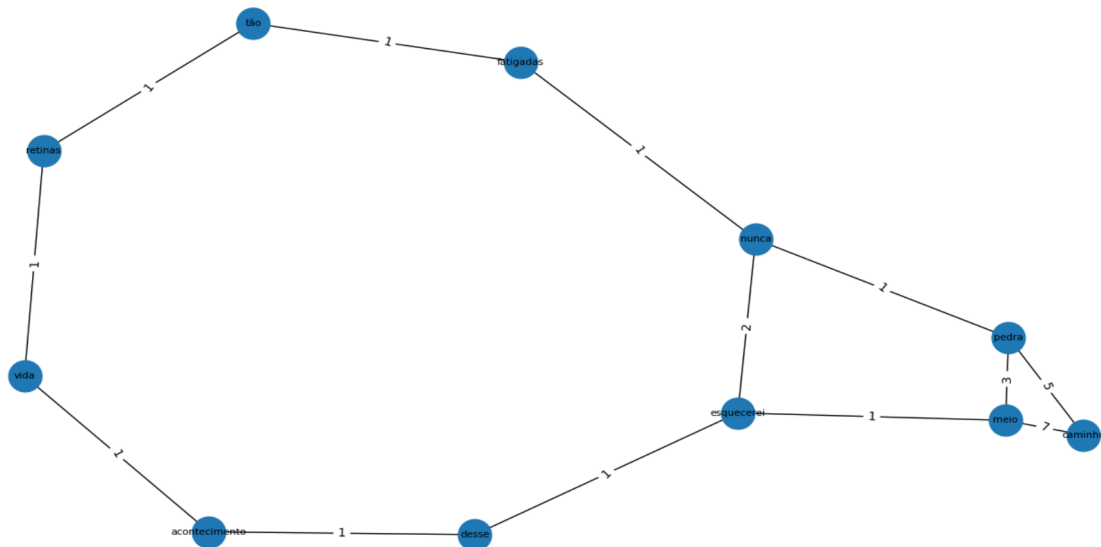


Figura 3: Rede do texto no meio do caminho

Ao analisar esta imagem vemos que o clique composto por pedra, caminho e meio tem peso 15 e que compõem o único clique da rede. Então aplicando o cálculo do coeficiente de aglomeração chegaremos em:  $(15/1)/11$ . Que resultaria num coeficiente de 1.363636.

## **4.2- Redações**

Após termos uma base do que seria um texto totalmente pouco dissertativo e pouco coeso, fizemos testes em redações que obtivemos do site do uol, das quais essas já haviam sido avaliadas com uma nota por um profissional. Para validar a nossa ideia de utilizar a rede para avaliar textos, utilizamos o programa para comparar a nota em relação ao coeficiente de aglomeração.

### ***4.2.1- Redação nota 960***

O resultado do coeficiente foi nulo, pois não havia cliques na rede.

### ***4.2.2- Redação nota 600***

O resultado do coeficiente foi nulo, pois não havia cliques na rede.

### ***4.2.3- Redação nota 400***

O resultado do coeficiente deste texto foi de 0,02190

### ***4.2.4- Redação nota 0***

O resultado do coeficiente deste texto foi de 0,05952

Analisando estes resultados podemos perceber que por meio deste programa podemos fazer uma análise superficial, porém verídica, uma vez que o coeficiente de aglomeração teve um resultado inversamente proporcional a nota. Mas ao mesmo tempo vemos que só o coeficiente de aglomeração não é suficiente para uma correção, uma vez que as redações 4.2.1 e 4.2.2 tiveram notas diferentes e o mesmo coeficiente. Isso ocorre pois o coeficiente está relacionado à repetição de palavras entre si, e não ao significado delas e ao sentido delas no texto.

## **6. Conclusão**

Neste trabalho realizamos a implementação de um simples programa para analisar a qualidade de um texto, e percebemos que isso realmente é possível, ainda que através de um programa mais básico, consequentemente a análise de qualidade não é muito avançada. Com a sequência deste trabalho poderíamos fazer um programa que analisasse um texto mais profundamente de tal modo que poderia ser utilizado como base para correções de grandes quantidades de textos, seja de forma parcial ou completa, seja no meio empresarial, no ramo de vestibulares ou auxiliando alunos na educação.

## 7. Referências

- Antiqueira, L. Nunes, M. Oliveira, O. Costa, L (2005) “Modelando Textos como Redes Complexas”, <http://www.nilc.icmc.usp.br/til/til2005/arq0054.pdf>, Abril.
- Metz, J. Calvo, R. Seno, E. Romero, R. Liang, Z. (2007) “Redes Complexas: conceitos e aplicações”, [encurtador.com.br/zAS45](http://encurtador.com.br/zAS45) , Abril.
- BANCO DE REDAÇÕES. Uol. Disponível em: <https://educacao.uol.com.br/bancoderedacoes/> . Abril.