

# Análise de relacionamentos de um banco de dados de investigados da Polícia Federal

Dênis Muller Borges  
Sebastián Gonçalves (Orientador)

## Introdução

Analizamos um banco de dados de relacionamentos entre indivíduos investigados pela Polícia Federal. Os dados coletados entre 2012 e 2014 foram encriptados e repassados para o nosso grupo a fim de desenvolver uma cooperação, da qual o presente trabalho faz parte. Utilizando as relações do banco de dados como conexões entre pares de indivíduos construímos uma rede complexa cuja maior componente conexa é mostrada na Figura 1 e seus parâmetros na Tabela 1.

O trabalho foi dividido em três etapas: filtragem e padronização dos termos que qualificam a função de cada indivíduo nas relações; classificação dessas funções de acordo com suas naturezas (amoroso, criminal, familiar, amizade, casual, duvidoso); descrição topológica e representação gráfica da rede de relacionamentos.

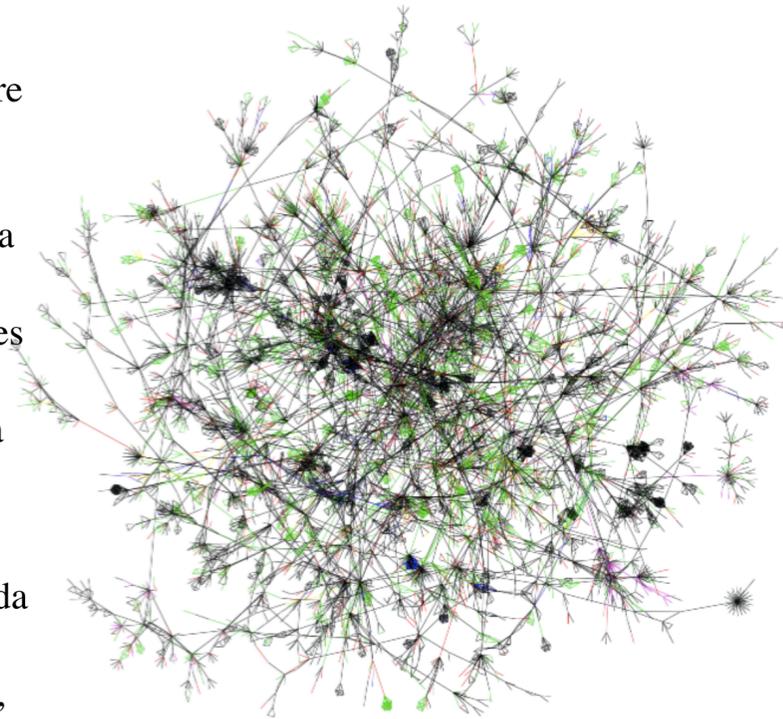


Figura 1. Rede de relações do banco de dados. Grupos de classificação identificados por cores (como na Figura 4).

Tabela 1. Parâmetros topológicos da rede.

Parâmetros	
número de nós	8699
número de conexões	17329
diâmetro	46
coeficiente de agrupamento	0.435
densidade	4.5E-04
characteristic path length	15.196
Modularidade	0.96

Tabela 2. Frequência de tipos de relações dos nodos estratégicos

Bridges	Frequência(%)
criminal	80,30
familiar	4,80
casual	1,29
amoroso	2,09
amizade	0,06
duvidoso	11,04

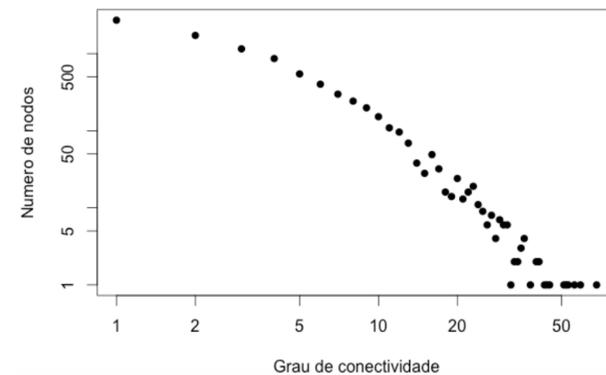


Figura 2. Distribuição do grau de conectividade da rede.

## Processo de filtragem e classificação

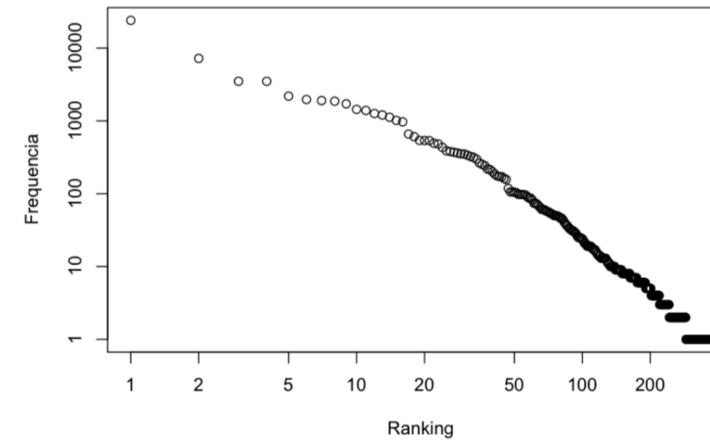


Figura 3. Distribuição de frequências das palavras do banco de dados, após a filtragem.

Tabela 3. Palavras mais frequentes no banco de dados

Ranking	Palavra	Frequência(%)
1	comparsa	33,87
2	irmao	10,17
3	contato	4,93
4	associado	4,92
5	socio	3,08
6	fornecedor	2,77
7	filho	2,67
Total		70.982

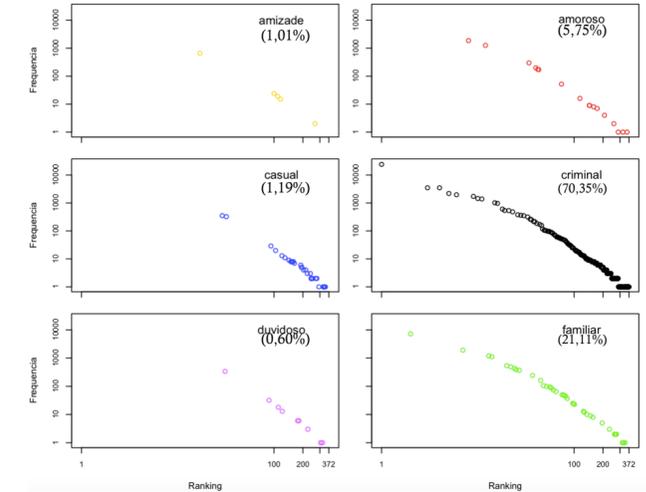


Figura 4. Distribuição de frequências das palavras do banco de dados por grupo de classificação.

Tabela 4. Frequência de tipos de relações no banco de dados

Relação	Frequência(%)
criminal	69,40
familiar	21,82
amoroso	5,91
casual	1,24
amizade	1,03
duvidoso	0,59
Número de Relações	33.922

## Conclusões

- A maior componente conexa da rede é aleatória, relativamente heterogênea e com alto grau de modularidade.
- A distribuição de frequências das palavras que caracterizam os indivíduos segue aproximadamente uma lei de potências (lei de Zipf).
- A maioria das relações são do tipo criminal (~70%).
- Entre as ligações que conectam módulos a presença de relacionamentos criminais é maior que entre as outras ligações.