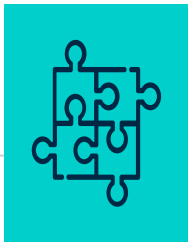


A close-up photograph of a computer keyboard. A magnifying glass with a black frame is positioned over a white key that has the words "FAKE NEWS" printed in bold, black, sans-serif capital letters. Two thin, parallel pink lines are drawn across the key, one above and one below the text. The background shows other keys on the keyboard, including ones with accented characters like "é", "è", "à", and symbols like "ç" and "@".

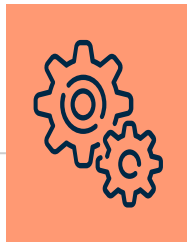
**FAKE  
NEWS**

**Análise de Tópicos**



1

CONHECIMENTO E  
PRÉ-PROCESSAMENTO



2

VISUALIZAÇÕES  
GRÁFICAS  
EXPLORATÓRIAS



3

ANÁLISE DE TÓPICOS -  
LDA

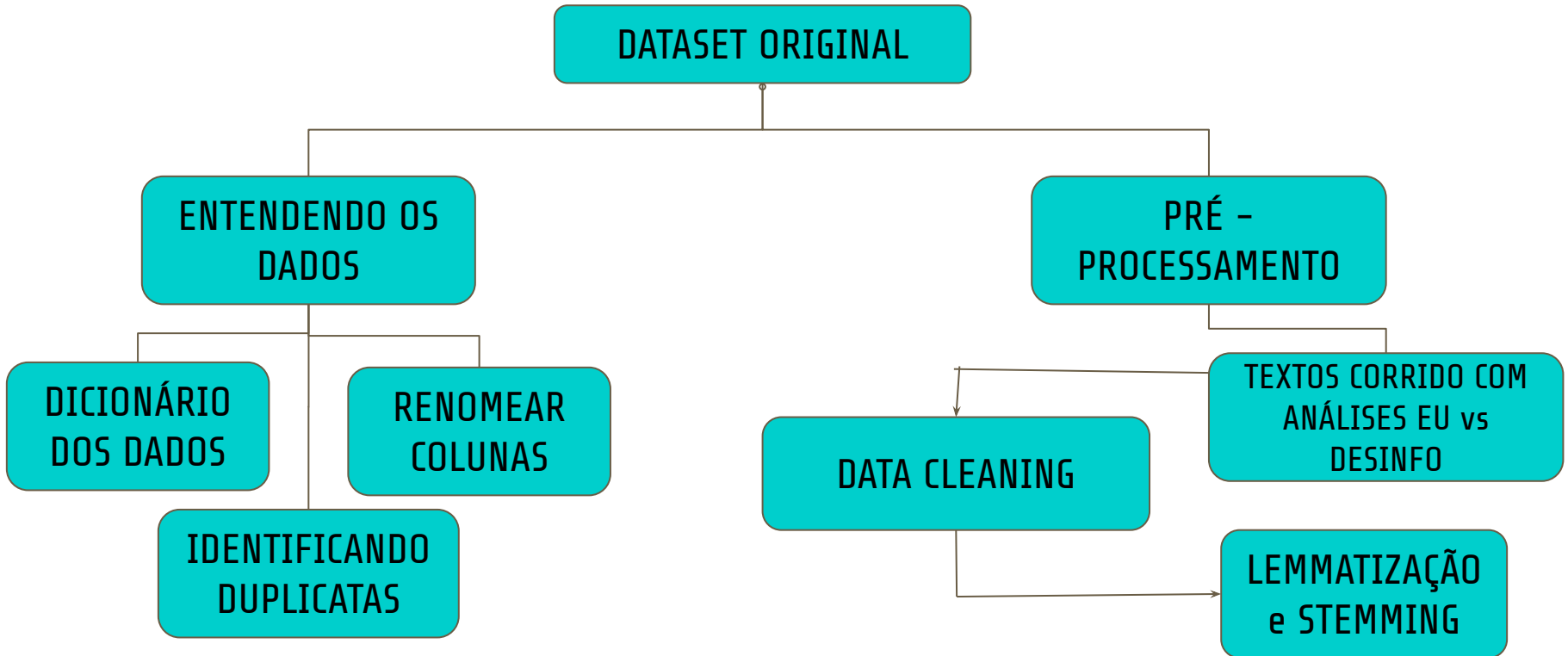
# CONHECENDO A FONTE



Dados coletados através de web scrap do banco de dados do site EU vs Desinfo que, por sua vez é baseado em análises do East Stratcom Task Force.

A East Stratcom Task Force é uma organização para explicar melhor os valores, interesses e políticas da União Europeia nos países parceiros da europa oriental.

# CONHECENDO A BASE E PRÉ-PROCESSANDO



# VISUALIZAÇÃO EXPLORATÓRIA

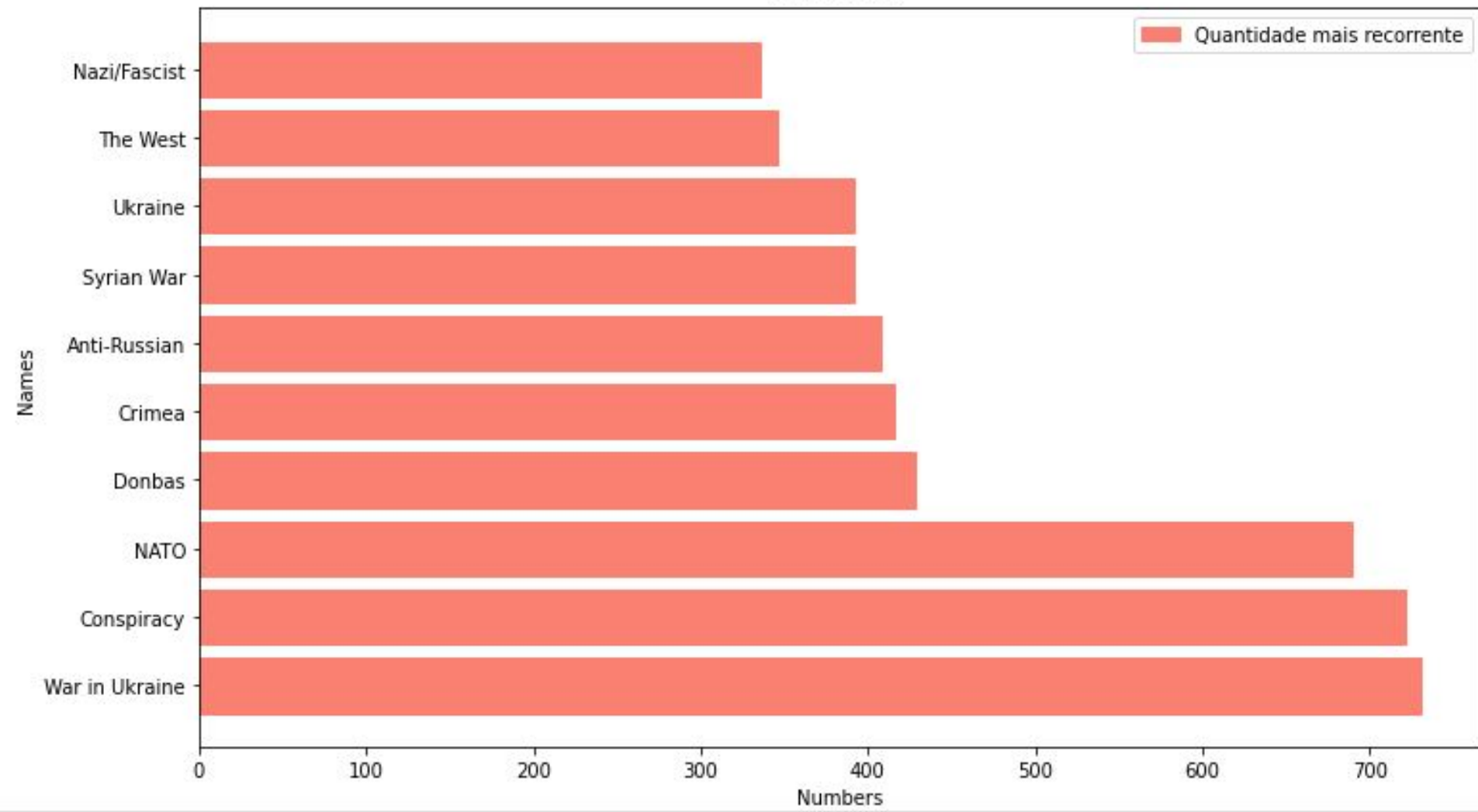
WORD CLOUD E BARCHART – KEYWORDS



[illegible]



## FAKE NEWS



# VISUALIZAÇÃO DOS DADOS

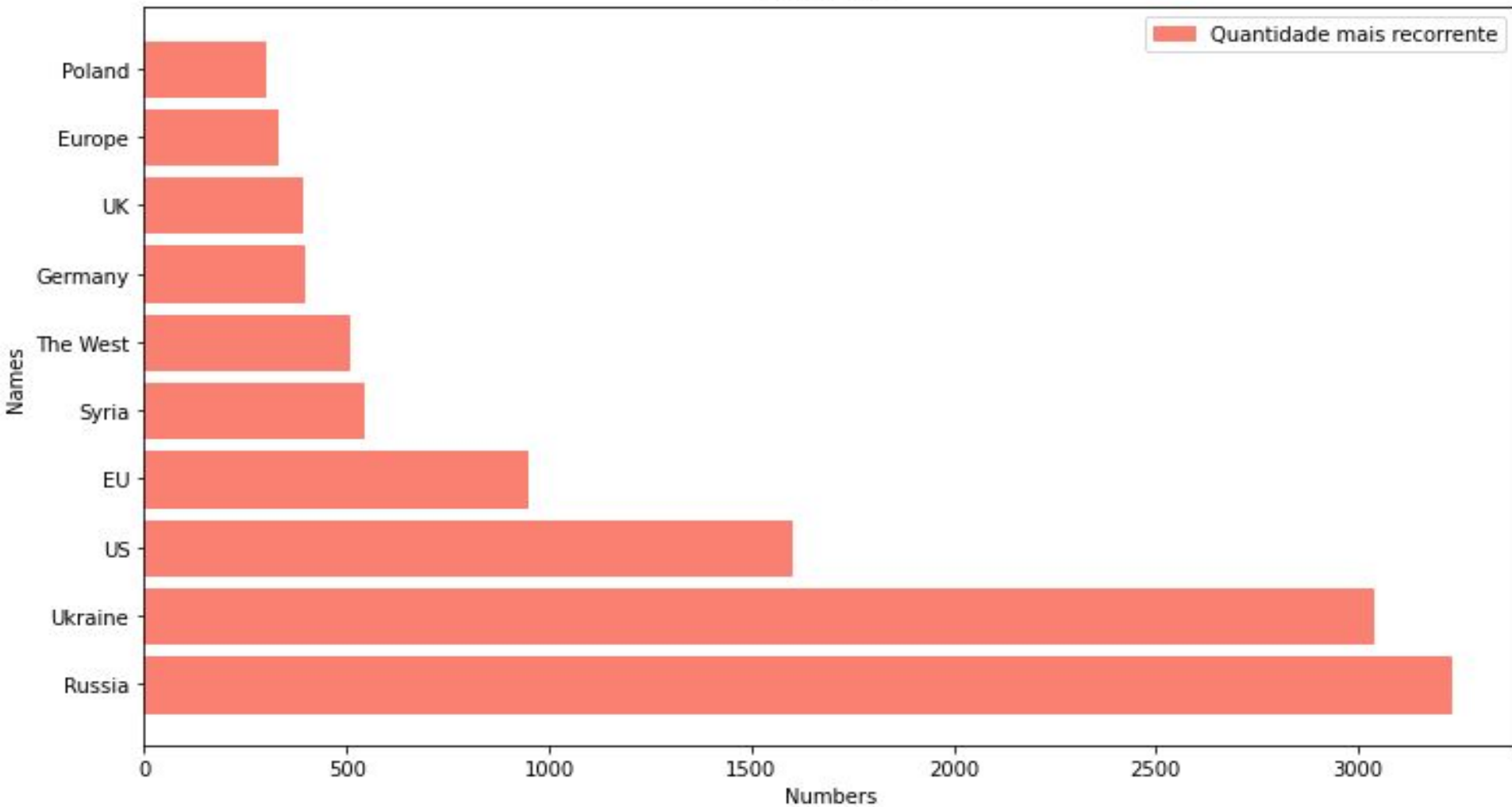
WORD CLOUD E BARCHART – COUNTRIES





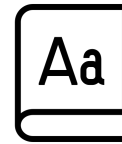


## FAKE NEWS



# VISUALIZAÇÃO DOS DADOS

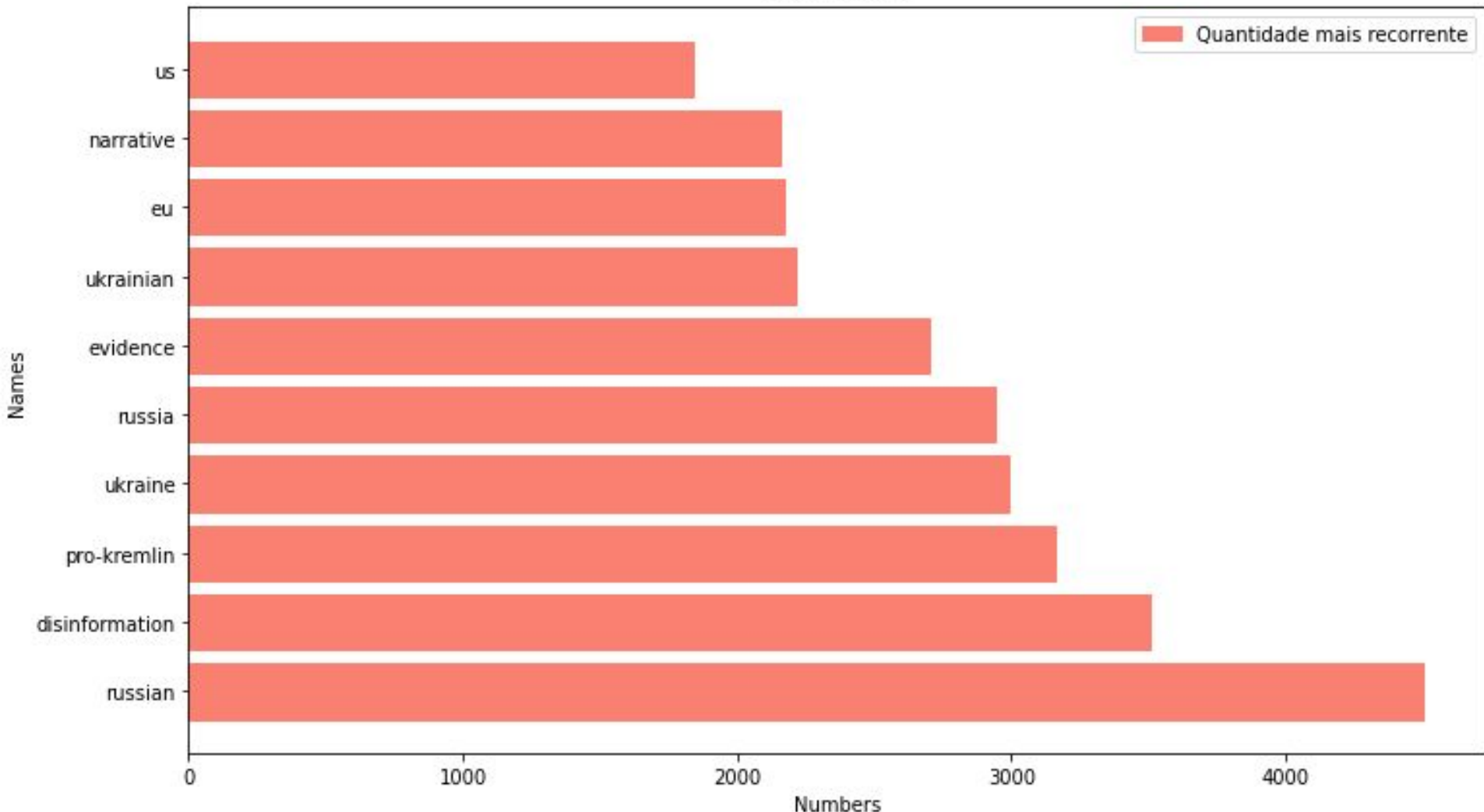
WORD CLOUD E BARCHART – CONCLUSION



[illegible]



## FAKE NEWS



# ANÁLISE DE TÓPICOS

## LDA

Cada documento é descrito pela distribuição de um conjunto de tópicos e cada tópico pode ser descrito por uma distribuição de palavras.

Verifica a co-ocorrência das palavras para separá-las por tópicos.

Contabiliza a ocorrência de palavras por documento e seus respectivos tópicos.

Determina a probabilidade de um texto pertencer a cada tópico.

# ANÁLISE DE TÓPICOS

## COERÊNCIA UMass

Logaritimo que verifica a qualidade dos tópicos

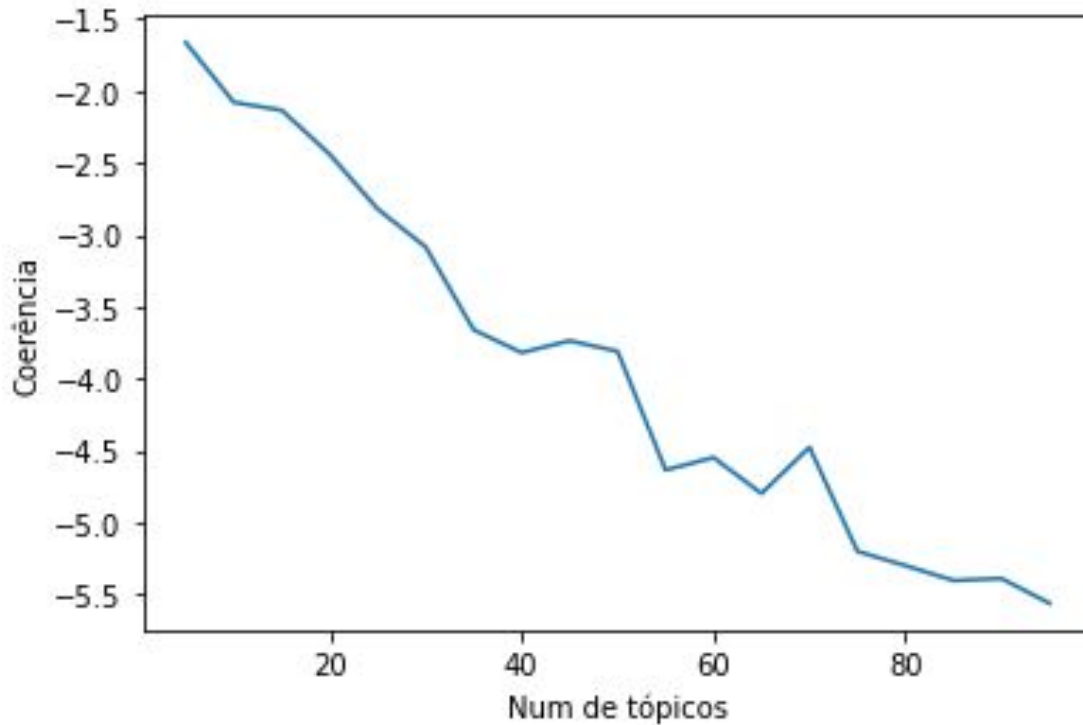
Tópicos perfeitos:  
 $UMass = 0$

Verifica a co-ocorrência das palavras nos documentos, utiliza o cálculo de quantas vezes elas aparecem juntas.

As palavras ganham um score maior se aparecerem com mais frequência juntas.



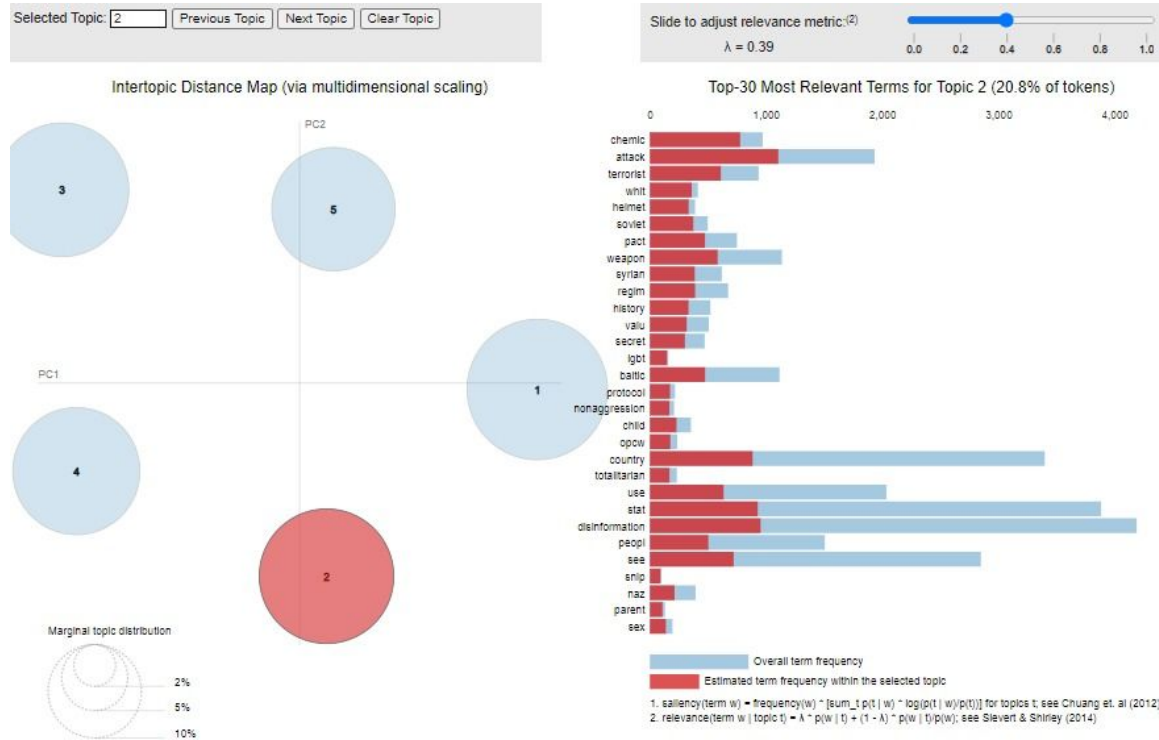
# ANÁLISE DE TÓPICOS



Número de tópicos  
usada: 5  
Valor de Coerência:  
-1.6633

# ANÁLISE DE TÓPICOS

## Guerra Civil Síria



## Tópico 2 1358 notícias

'0.011 ----- "attack"  
0.009 ----- "disinformation"  
0.009 ----- "stat"  
0.009 ----- "country"  
0.008 ----- "chemic"  
0.008 ----- "prokremlin"  
0.007 ----- "see"  
0.007 ----- "evidenc"  
0.006 ----- "use"  
0.006 ----- "terrorist"

# ANÁLISE DE TÓPICOS

## Tensões na Crimeia

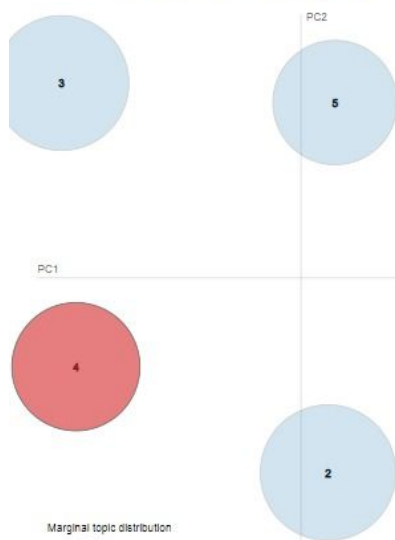
Selected Topic:  Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:<sup>(2)</sup>

$\lambda = 0.39$



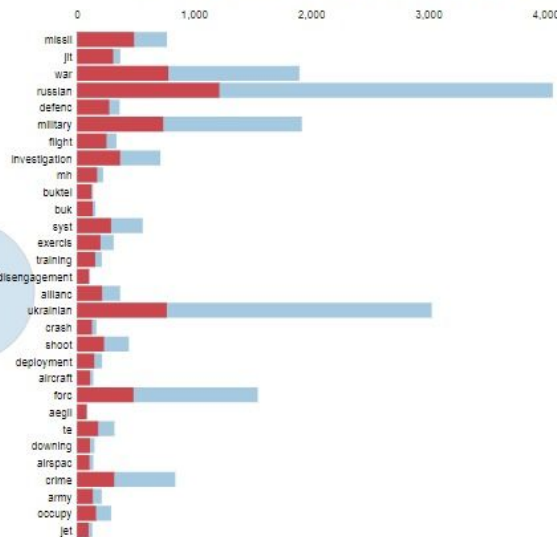
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 4 (18.6% of tokens)



Overall term frequency  
Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) \* [sum\_t p(t | w) \* log(p(t | w)/p(t))]  
2. relevance(term w | topic t) =  $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$ ; see Slevert & Shrivay (2014)

Tópico 4:  
1358 notícias

0.017 ----- "russian"  
0.011 ----- "war"  
0.011 ----- "ukrainian"  
0.011 ----- "military"  
0.008 ----- "narrativ"  
0.007 ----- "missil"  
0.007 ----- "forc"  
0.007 ----- "stat"  
0.007 ----- "country"  
0.006 ----- "disinformation"

# TRABALHOS FUTUROS

Utilizar a análise de similaridade semântica entre os textos, clusterizar e comparar com os resultados obtidos anteriormente com as métricas do LDA.

Testar os resultados das avaliações do site Eu vs desinfo construindo um classificador a partir das notícias originais. Esse classificador pode conter análise da estrutura do texto e do espalhamento.

Verificação das palavras chaves nas notícias originais, se elas são as mesmas utilizadas nas avaliações (principalmente resumo das notícias) do banco de dados EU vs Desinfo. Verificando a possibilidade de viés humano de cunho político na produção das avaliações.