

## Initialization

### Set paths

Determinamos os caminhos, para salvar e processar documentos, para utilizar pacotes e dicionários:

```
In[*]:= Clear[path, path2, path3, path4, path5]
path = "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\datasets\\";
path2 = "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\processing\\";
path3 = "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\packages\\";
path4 =
  "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\dictionaries\\";
path5 = "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\pictures\\";
```

### Importando os datasets sobre Marielle Franco em 2020 em espanhol, inglês e português.

```
Clear[authors, fullauthors]
authors = {"MFES_2020", "MFEN_2020", "MF_2020"};
```

Importamos o csv e o chamamos data[i], em que i são as autoras

```
In[*]:= Clear[data]
Do[data[i] = Import[path <> i <> ".csv"], {i, authors}];
```

Eliminamos todas as linhas sem informação

```
In[*]:= Clear[data1]
Do[data1[i] = Select[data[i], Not[# == {}] &];
  Print[i <> " → " <> ToString[Length[data[i]]]], {i, authors}];

MFES_2020 → 18776
MFEN_2020 → 13531
MF_2020 → 351757
```

```
In[*]:= Clear[data2]
Do[data2[i] = Drop[data1[i], 1], {i, authors}];
```

Do dataset inicial, extraímos o texto, as datas e as linguagens.

```
In[*]:= Clear[data3]
Do[data3[i] = data2[i] [[All, {2, 4, 8}]], {i, authors}];
```

```

Clear[languages, tweets]
languages = {"es", "en", "pt"};
tweets["es"] = Select[data3["MFES_2020"], #[[3]] == "es" &];
tweets["en"] = Select[data3["MFEN_2020"], #[[3]] == "en" &];
tweets["pt"] = Select[data3["MF_2020"], #[[3]] == "pt" &];
Table[Length[tweets[j]], {j, languages}]

```

```

Out[ ] =
{18775, 13530}

```

Criamos uma série de regras para construir termos significativos e as aplicamos aos tweets com e sem retweets:

```

In[ ] := Clear[EncodeRules, DecodeRules]
EncodeRules = {"Brazil" → "Brasil", ("Buenos Aires" | "Baires" | "BsAs") → "BuenosAires",
  ("Sao Paulo" | "São Paulo" | "San Pablo") → "SaoPaulo",
  ("Silvina Ocampo" | "Silvina") → "SilvinaOcampo",
  ("Rio de Janeiro" | "RioJaneiro" | "RiodeJaneiro") → "RioDeJaneiro",
  ("NYC" | "New York City" | "NewYorkCity" | "NewYork" |
    "Nova Yorque" | "New York" | "Nueva York" | "Nova York") → "NewYork",
  ("Estados Unidos de América" | "Estados Unidos") → "EstadosUnidos",
  ("United States of America" | "United States") → "UnitedStates",
  ("Latin America" | "Latin American") → "LatinAmerica",
  "Clarice Lispector" → "ClariceLispector",
  ("Simone de Beauvoir" | "Simone Beauvoir") → "SimonedeBeauvoir",
  "Marielle Franco" → "MarielleFranco",
  ("Eva Perón" | "Eva Peron" | "Evita") → "EvaPeron", "Sueli Carneiro" → "SueliCarneiro",
  "Conceição Evaristo" → "ConceiçãoEvaristo", "Victoria Ocampo" → "VictoriaOcampo",
  ("Lélia González" | "Lelia Gonzalez" | "Lélia Gonzalez" | "Lelia González" |
    "Lélia de Almeida González" | "Lélia Almeida González") →
    "LéliaGonzalez", "Alfonsina Storni" → "AlfonsinaStorni",
  ("Carolina Maria de Jesus" | "Carolina Maia de Jesus" | "Carolina María de Jesus" |
    "Carolina de Jesus" | "Carolina María de Jesús") → "CarolinaMariadeJesus",
  "Alejandra Pizarnik" → "AlejandraPizarnik", "Elza Soares" → "ElzaSoares",
  "Rita Segato" → "RitaSegato", "Santiago Maldonado" → "SantiagoMaldonado",
  "Simón Bolívar" → "SimónBolívar", "Anderson Gomes" → "AndersonGomes",
  ("Angela Davis" | "Ângela Davis") → "AngelaDavis", "Nina Simone" → "NinaSimone",
  ("Beatriz Nascimento" | "Maria Beatriz Nascimento") → "BeatrizNascimento",
  ("bell hooks" | "Bell Hooks" | "Bell hooks" | "bell Hooks") → "bellhooks",
  ("Djamila Ribeiro" | "djamila Ribeiro" | "Djamila ribeiro" | "Djamilla Ribeiro") →
    "DjamilaRibeiro", ("Dona Ivone Lara" | "Ivone Lara") → "IvoneLara",
  ("Teresa de Benguela" | "Tereza de Benguela") → "TerezadeBenguela",
  "Maya Angelou" → "MayaAngelou", "Jurema Werneck" → "JuremaWerneck",
  ("Chimamanda Ngozi Adichie" | "Chimamanda Adichie" | "Chimamanda Ngozi") →
    "ChimamandaNgoziAdichie", "James Baldwin" → "JamesBaldwin",
  "Octavia Butler" → "OctaviaButler", "Toni Morrison" → "ToniMorrison",
  ("Lula da Silva" | "Inácio Lula da Silva" | "Luiz Inácio Lula da Silva") → "LuladaSilva",

```

"Audre Lorde" → "AudreLorde", "Companhia das Letras" → "CompanhiadasLetras",  
 "Machado de Assis" → "MachadodeAssis", "Lima Barreto" → "LimaBarreto",  
 ("Abdias do Nascimento" | "Abdias Nascimento") → "AbdiasdoNascimento",  
 ("Maria Firmina dos Reis" | "Maria Firmina") → "MariaFirminadosReis",  
 ("Quarto de Despejo" | "quarto de despejo" | "Quarto De Despejo" | "Quarto de despejo") →  
 "Quartodedespejo", "Casa de Alvenaria" → "CasadeAlvenaria",  
 "Cuarto de desechos" → "Cuartodedesechos", "Milton Santos" → "MiltonSantos",  
 "Child of the Dark" → "ChildofTheDark", "Luiza Bairros" → "LuizaBairros",  
 "Paulo Freire" → "PauloFreire", "Darcy Ribeiro" → "DarcyRibeiro",  
 "Benedita da Silva" → "BeneditadaSilva",  
 "Viola Davis" → "ViolaDavis", ("Malcom X" | "Malcolm X") → "MalcomX",  
 ("Luther King" | "Martin Luther King") → "MartinLutherKing",  
 "Frantz Fanon" → "FrantzFanon", "Morgan Freeman" → "MorganFreeman",  
 "Desmond Tutu" → "DesmondTutu", "Clóvis Moura" → "ClóvisMoura",  
 "Édison Carneiro" → "EdisonCarneiro", "Nei Lopes" → "NeiLopes",  
 "Muniz Sodré" → "MunizSodré", "Guerreiro Ramos" → "GuerreiroRamos",  
 "Grada Kilomba" → "GradaKilomba", "Silvio Almeida" → "SilvioAlmeida",  
 ("Joice Berth" | "Joyce Berth") → "JoiceBerth",  
 "Kabengele Munanga" → "KabengeleMunanga", "Dandara dos Palmares" → "DandaradosPalmares",  
 "Antonieta de Barros" → "AntonietadeBarros", "André Rebouças" → "AndréRebouças",  
 "Milton Nascimento" → "MiltonNascimento", "José do Patrocínio" → "JosédoPatrocínio",  
 "Cidinha da Silva" → "CidinhadaSilva", "Ailton Krenak" → "AiltonKrenak",  
 "Patricia Hill Collins" → "PatriciaHillCollins",  
 "Florestan Fernandes" → "FlorestanFernandes",  
 ("Gloria Anzaldua" | "Gloria Anzaldúa" | "Glória Anzaldúa" | "Glória Anzaldua") →  
 "GloriaAnzaldua", "Nelson Mandela" → "NelsonMandela",  
 "Aimé Césaire" → "AiméCésaire", "Assad Haider" → "AssadHaider",  
 "Achille Mbembe" → "AchilleMbembe", "Ana Maria Gonçalves" → "AnaMariaGonçalves",  
 "Eliana Alves Cruz" → "ElianaAlvesCruz", "Carla Akotirene" → "CarlaAkotirene",  
 "Lázaro Ramos" → "LázaroRamos", "Gilberto Freyre" → "GilbertoFreyre",  
 ("Virginia Bicudo" | "Virginia Leone Bicudo") → "VirginiaBicudo",  
 "Aimé Césaire" → "AiméCésaire", "Zumbi dos Palmares" → "ZumbidosPalmares",  
 "Fundação Palmares" → "FundaçãoPalmares", "Flávia Rios" → "FláviaRios",  
 "Movimento Negro Unificado" → "MovimentoNegroUnificado",  
 "Ruth de Souza" → "RuthdeSouza", "Diário de Bitita" → "DiáriodeBitita",  
 ("Audálio Dantas" | "Audalio Dantas") → "AudalioDantas", "Jorge Amado" → "JorgeAmado",  
 "Caetano Veloso" → "CaetanoVeloso", "Solano Trindade" → "SolanoTrindade",  
 ("Vera Eunice de Jesus" | "Vera Eunice") → "VeraEunicedeJesus",  
 "Frida Kahlo" → "FridaKahlo"};

DecodeRules = { "BuenosAires" → "Buenos Aires", "SaoPaulo" → "São Paulo",  
 "RioDeJaneiro" → "Rio de Janeiro", "NewYork" → "New York",  
 "SilvinaOcampo" → "Silvina Ocampo", "EstadosUnidos" → "Estados Unidos",  
 "UnitedStates" → "United States", "LatinAmerica" → "Latin America",  
 "ClariceLispector" → "Clarice Lispector", "MarielleFranco" → "Marielle Franco",  
 "EvaPeron" → "Eva Peron", "SueliCarneiro" → "Sueli Carneiro",  
 "ConceiçãoEvaristo" → "Conceição Evaristo", "VictoriaOcampo" → "Victoria Ocampo",  
 "LéliaGonzalez" → "Lélia Gonzalez", "AlfonsinaStorni" → "Alfonsina Storni",

"CarolinaMariadeJesus" → "Carolina Maria de Jesus",  
 "AlejandraPizarnik" → "Alejandra Pizarnik", "ElzaSoares" → "Elza Soares",  
 "RitaSegato" → "Rita Segato", "SantiagoMaldonado" → "Santiago Maldonado",  
 "SimónBolívar" → "Simón Bolívar", "AndersonGomes" → "Anderson Gomes",  
 "AngelaDavis" → "Angela Davis", "NinaSimone" → "Nina Simone",  
 "BeatrizNascimento" → "Beatriz Nascimento",  
 "bellhooks" → "bell hooks", "DjamilaRibeiro" → "Djamila Ribeiro",  
 "IvoneLara" → "Ivone Lara", "TeresadeBenguela" → "Teresa de Benguela",  
 "MayaAngelou" → "Maya Angelou", "JuremaWerneck" → "Jurema Werneck",  
 "ChimamandaNgoziAdichie" → "Chimamanda Ngozi Adichie", "JamesBaldwin" → "James Baldwin",  
 "OctaviaButler" → "Octavia Butler", "ToniMorrison" → "Toni Morrison",  
 "LuladaSilva" → "Lula da Silva", "AudreLorde" → "Audre Lorde",  
 "CompanhiadasLetras" → "Companhia das Letras", "MachadodeAssis" → "Machado de Assis",  
 "LimaBarreto" → "Lima Barreto", "AbdiasdoNascimento" → "Abdias do Nascimento",  
 "MariaFirminadosReis" → "Maria Firmina dos Reis",  
 "Quarto de Despejo" → "Quarto de despejo", "CasadeAlvenaria" → "Casa de Alvenaria",  
 "Cuartodedesechos" → "Cuarto de desechos", "MiltonSantos" → "Milton Santos",  
 "ChildoftheDark" → "Child of The Dark", "LuizaBairros" → "Luiza Bairros",  
 "PauloFreire" → "Paulo Freire", "DarcyRibeiro" → "Darcy Ribeiro",  
 "BeneditadaSilva" → "Benedita da Silva", "ViolaDavis" → "Viola Davis",  
 "MalcomX" → "Malcom X", "MartinLutherKing" → "Martin Luther King",  
 "FrantzFanon" → "Frantz Fanon", "MorganFreeman" → "Morgan Freeman",  
 "DesmondTutu" → "Desmond Tutu", "ClóvisMoura" → "Clóvis Moura",  
 "ÉdisonCarneiro" → "Edison Carneiro", "NeiLopes" → "Nei Lopes",  
 "MunizSodré" → "Muniz Sodré", "GuerreiroRamos" → "Guerreiro Ramos",  
 "GradaKilomba" → "Grada Kilomba", "SilvioAlmeida" → "Silvio Almeida",  
 "JoiceBerth" → "Joice Berth", "KabengeleMunanga" → "Kabengele Munanga",  
 "DandaradosPalmares" → "Dandara dos Palmares",  
 "AntonietadeBarros" → "Antonietta de Barros", "AndréRebouças" → "André Rebouças",  
 "MiltonNascimento" → "Milton Nascimento", "JosédoPatrocínio" → "José do Patrocínio",  
 "CidinhadaSilva" → "Cidinha da Silva", "AiltonKrenak" → "Ailton Krenak",  
 "PatriciaHillCollins" → "Patricia Hill Collins",  
 "FlorestanFernandes" → "Florestan Fernandes", "GloriaAnzaldua" → "Gloria Anzaldua",  
 "NelsonMandela" → "Nelson Mandela", "AiméCésaire" → "Aimé Césaire",  
 "AssadHaider" → "Assad Haider", "AchilleMbembe" → "Achille Mbembe",  
 "AnaMariaGonçalves" → "Ana Maria Gonçalves", "ElianaAlvesCruz" → "Eliana Alves Cruz",  
 "CarlaAkotirene" → "Carla Akotirene", "LázaroRamos" → "Lázaro Ramos",  
 "GilbertoFreyre" → "Gilberto Freyre", "VirginiaBicudo" → "Virginia Bicudo",  
 "AiméCésaire" → "Aimé Césaire", "ZumbidosPalmares" → "Zumbi dos Palmares",  
 "FundaçãoPalmares" → "Fundação Palmares", "FláviaRios" → "Flávia Rios",  
 "MovimentoNegroUnificado" → "Movimento Negro Unificado",  
 "RuthdeSouza" → "Ruth de Souza", "DiáriodeBitita" → "Diário de Bitita",  
 "AudalioDantas" → "Audalio Dantas", "JorgeAmado" → "Jorge Amado",  
 "CaetanoVeloso" → "Caetano Veloso", "SolanoTrindade" → "Solano Trindade",  
 "VeraEuniceJesus" → "Vera Eunice de Jesus",  
 "SimonedeBeauvoir" → "Simone de Beauvoir"};

Eliminamos os retweets

```
In[*]:= Clear[NoRT]
Do[NoRT[j] = Select[tweets[j],
  Not[StringContainsQ[#[[2]], StartOfString ~~ "RT", IgnoreCase -> True]] &], {j, languages}]

In[*]:= Table[i -> Dimensions[NoRT[j]], {j, languages}]
Out[*]:=
{i -> {18774, 3}, i -> {13524, 3}}

In[*]:= Clear[Enc]
Do[Enc[j] = Map[{StringReplace[#[[1]], EncodeRules, IgnoreCase -> True], #[[2]]} &,
  Tally[NoRT[j][[All, 2]]], {j, languages}];
```

## Operações adicionais de limpeza dos tweets

Limpamos o corpus de links, nomes de usuários e outros caracteres não alfanuméricos.

```
In[*]:= Clear[CleanText, ProperWords]
CleanText[text_] := StringReplace[text,
  {RegularExpression["[http]*[s]*[:] [\\\/\\\/] [a-z0-9\\.\\\/]+[ ]*" ] -> " ",
    RegularExpression["[rt: ]*[rt ]* [\\@] [a-z0-9]*[:]*" ] -> " ",
    RegularExpression["@[a-z0-9]*[:]*" ] -> " "}, IgnoreCase -> True];
ProperWords[lw_] :=
  Select[lw, And[StringFreeQ[#, RegularExpression["\\W"]], StringLength[#] > 4] &];

In[*]:= Clear[ttweets]
Do[ttweets[j] = SortBy[Map[{ProperWords[TextWords[CleanText[#[[1]]]], #[[2]]} &, Enc[j]],
  -#[[2]] &], {j, languages}];
```

Partindo de uma lista de stopwords, ou palavras sem valor léxico, (artigos, pronomes, preposições) em português, inglês e espanhol, eliminamos todas essas palavras de nosso corpus.

```
In[*]:= Clear[swes, swpt, swen, stopwords, RemoveStopWords, cleantweets0]
swes = Get[path4 <> "sw-es.txt"];
swpt = Get[path4 <> "sw-pt.txt"];
swen = Get[path4 <> "sw-en.txt"];
stopwords = Join[swes, swpt, swen];
RemoveStopWords[s_] :=
  Select[s, Not[StringMatchQ[#, Alternatives @@ stopwords, IgnoreCase -> True]] &];

In[*]:= Do[cleantweets0[j] = Map[{RemoveStopWords[#[[1]], #[[2]]} &, ttweets[j]], {j, languages}];
```

Depois consolidamos nosso corpus agrupando aqueles tweets que tenham ficado iguais após a eliminação de stopwords.

```
In[*]:= Clear[ConsolidateTally]
ConsolidateTally[tally_] :=
  Map[{#[[1, 1]], Total[#[[All, 2]]]} &, Gather[tally, SameQ[#1[[1]], #2[[1]]] &]];

In[*]:= Do[cleantweets[j] = ConsolidateTally[cleantweets0[j]], {j, languages}];
```

A última operação de limpeza tem a ver com obter as raízes das palavras nas três línguas e agrupar as palavras de acordo com a palavra mais frequente da mesma raiz

Primeiro computamos a quantidade de repetições das palavras com uma função que considera a multiplicidade de tweets.

```
In[*]:= Clear[DistributelList, ConsolidateTally, wordtally]
DistributelList[{l_List, x_}] := Map[{#, x} &, l];
ConsolidateTally[tally_] :=
  Map[{#[[1, 1]], Total[#[[All, 2]]]} &, Gather[tally, SameQ[#1[[1]], #2[[1]]] &]];

In[*]:= Do[wordtally[j] = SortBy[ConsolidateTally[
  Join@@Map[DistributelList, cleantweets[j]]], -#[[2]] &], {j, languages}];
```

Definimos um comando que obtém o termo mais comum de uma classe de termos para cada língua. Fazemos isso através das funções WordStem, WordStemES e WordStemPT. A primeira é uma função do Wolfram que obtém a raiz das palavras. As duas últimas foram implementadas por nós no Wolfram Mathematica a partir do algoritmo para obter raízes dessas línguas presente aqui: <http://snowball.tartarus.org/algorithms/portuguese/stemmer.html> e <http://snowball.tartarus.org/algorithms/spanish/stemmer.html>.

```
In[*]:= Clear[MeaningfulTermsRules]
MeaningfulTermsRules[ts_, "en"] := Module[{cat, f},
  cat =
    Gather[ts, SameQ[WordStem[ToLowerCase[#1[[1]]]], WordStem[ToLowerCase[#2[[1]]]]] &];
  f[c_] := Map[#[[1]] → Last[SortBy[c, Last]][[1]] &, c];
  Select[Union[Flatten[Map[f, cat]]], Not[SameQ[#[[1]], #[[2]]]] &]];
MeaningfulTermsRules[ts_, "es"] := Module[{cat, f},
  cat = Gather[ts,
    SameQ[WordStemES[ToLowerCase[#1[[1]]]], WordStemES[ToLowerCase[#2[[1]]]]] &];
  f[c_] := Map[#[[1]] → Last[SortBy[c, Last]][[1]] &, c];
  Select[Union[Flatten[Map[f, cat]]], Not[SameQ[#[[1]], #[[2]]]] &]];
MeaningfulTermsRules[ts_, "pt"] := Module[{cat, f},
  cat = Gather[ts,
    SameQ[WordStemPT[ToLowerCase[#1[[1]]]], WordStemPT[ToLowerCase[#2[[1]]]]] &];
  f[c_] := Map[#[[1]] → Last[SortBy[c, Last]][[1]] &, c];
  Select[Union[Flatten[Map[f, cat]]], Not[SameQ[#[[1]], #[[2]]]] &]];
```

```
In[*]:= Clear[mrules]
Do[mrules[j] = MeaningfulTermsRules[wordtally[j], j], {j, languages}];
```

Vamos substituir as palavras pela palavra mais frequente do grupo dado pela mesma raiz, de acordo com as regras já computadas.

```
In[*]:= Clear[mtweets]
Do[mtweets[j] = Map[{Sort[#[[1]]], #[[2]]} &, Replace[cleantweets[j], mrules[j], {3}]],
  {j, languages}];
```

Vamos computar as palavras que só aparecem uma vez em cada conjunto de tweets (por intelectual e língua), que consideraremos triviais para cómputo dos temas principais.

```
In[*]:= Clear[trivialwords]
Do[trivialwords[j] = Select[Tally[Flatten[mtweets[j] [[All, 1]]], #[[2]] < 2 &] [[All, 1]],
  {j, languages}];
```

Computamos os tweets significativos eliminando as palavras que aparecem só uma vez, as quais consideraremos triviais para extrair os temas principais.

```
In[*]:= Clear[Mtweets]
Do[Mtweets[j] = ConsolidateTally[
  Select[Map[{Sort[Complement[#[[1]], trivialwords[j]]], #[[2]]} &, mtweets[j]],
    Length[#] ≥ 1 &]], {i, authors}, {j, languages}];
```

Tabelas para o número de tweets significativos únicos, com ou sem retweets:

```
In[*]:= TableForm[Table[Length[Mtweets[j]], {j, languages}], TableHeadings → {languages}]
Out[*]//TableForm=
es|1854
en|1653
```

Recuperamos esse corpus textual significativo:

```
Do[Put[{Enc[j], ttweets[j], cleantweets0[j], cleantweets[j],
  wordtally[j], mrules[j], mtweets[j], trivialwords[j], Mtweets[j]}],
  path2 <> "TM-processed-tweets-LAWomen-2023" <> ToString[j] <> ".dat"], {j, languages}];
```

```
Do[{Enc[j], ttweets[j], cleantweets0[j], cleantweets[j],
  wordtally[j], mrules[j], mtweets[j], trivialwords[j], Mtweets[j]} = Get[
  path2 <> "TM-processed-tweets-LAWomen-2023" <> ToString[j] <> ".dat"], {j, languages}];
```

Com os processamentos resultantes, elaboramos uma modelagem de tópicos para cada linguagem.

Computamos a DTM com retweets.

```

In[*]:= Clear[docs, terms, dtm]
Do[docs[j] = Replace[Mtweets[j], DecodeRules, {3}], {j, languages}];
Do[terms[j] = Union[Flatten[docs[j][[All, 1]]], {j, languages}];
Do[dtm[j] = Table[d[[2]]  $\times$  Count[d[[1]], t], {d, docs[j]}, {t, terms[j]}], {j, languages}];

Do[Put[{docs[j], terms[j], dtm[j]},
  path2 <> "TM-NNMF-dtm-Twitter-2023" <> ToString[j] <> ".dat"], {j, languages}];

```

Fazemos uma tabela com os números de documentos únicos e termos únicos.

```

In[*]:= TableForm[Table[Length[docs[j]], {j, languages}], TableHeadings -> {languages}]
TableForm[Table[Length[terms[j]], {j, languages}], TableHeadings -> {languages}]

```

```

Out[*]//TableForm=
  pt | 29 593
  en | 1653
  es | 1854

```

```

Out[*]//TableForm=
  pt | 9199
  en | 1426
  es | 1696

```

## Utilizamos a técnica de Modelagem de Tópicos - NonNegativeMatrixFactorization

```

In[*]:= Dimensions[dtm["en"]]

```

```

Out[*]=
{1653, 1426}

```

```

In[*]:= Needs["NonNegativeMatrixFactorization", path3 <> "NonNegativeMatrixFactorization.m"]

```

```

In[*]:= Clear[T, W, H]
T[j_] = 10;
Do[{W[j], H[j]} = NonNegativeMatrixFactorization[dtm[j], T[j]], {j, languages}];
Do[{W[j], H[j]} = RightNormalizeMatrixProduct[W[j], H[j]], {j, languages}];

```

```

In[*]:= Do[Put[{T[j], W[j], H[j]}, path2 <>
  "TM-NNMF-RT-Matrices-Twitter-2023-10-ESEN" <> ToString[j] <> ".dat"], {j, languages}];

```

```

In[*]:= {T["pt"], W["pt"], H["pt"]} = Get[path2 <> "TM-NNMF-RT-Matrices-Twitter-2023-10-PTpt.dat"];

```

```

In[*]:= Manipulate[
  If[Min[Dimensions[H[j]]]  $\geq$  T[j], TabView[Table["Topic " <> ToString[n] -> TableForm[
    Map[Reverse, BasisVectorInterpretation[H[j][[n]], 15, terms[j]]], {n, 1, 10}]],
    "Insufficient data to perform analysis."], {{j, languages[[1]], "Language"}, languages},
  ControlPlacement -> Top, LabelStyle -> Directive[FontSize -> 20]]

```

## Grafos para os tópicos extraídos a partir de uma seleção (manual) dos mais relevantes.

Colocamos os tópicos para cada grafo :



```

Clear[topics]
topics["pt"] = {2, 5, 7, 9, 10};
topics["es"] = {2, 3, 4, 8, 10};
topics["en"] = {2, 4, 6, 7, 9};

```

Colocamos os nomes dos tópicos no grafo :

```

Clear[topiclabels]
topiclabels["pt"] =
  {"1", "Saber", "3", "4", "Respostas", "6", "Assassinato", "8", "Bolsonaro", "Valores"};
topiclabels["en"] =
  {"1", "Accountable", "3", "Black women", "5", "Activist", "Killed", "8", "Police", "10"};
topiclabels["es"] =
  {"1", "Cumplen", "Negra", "Concejala", "5", "6", "7", "Mujeres", "9", "Asesinato"};

```

Construção dos gráficos extraídos dos tweets. ( $i = \text{"EP", "RS"}$   $j = \text{en, es, pt}$ , e  $k$  é o número de termos a considerar para cada tópico.

```

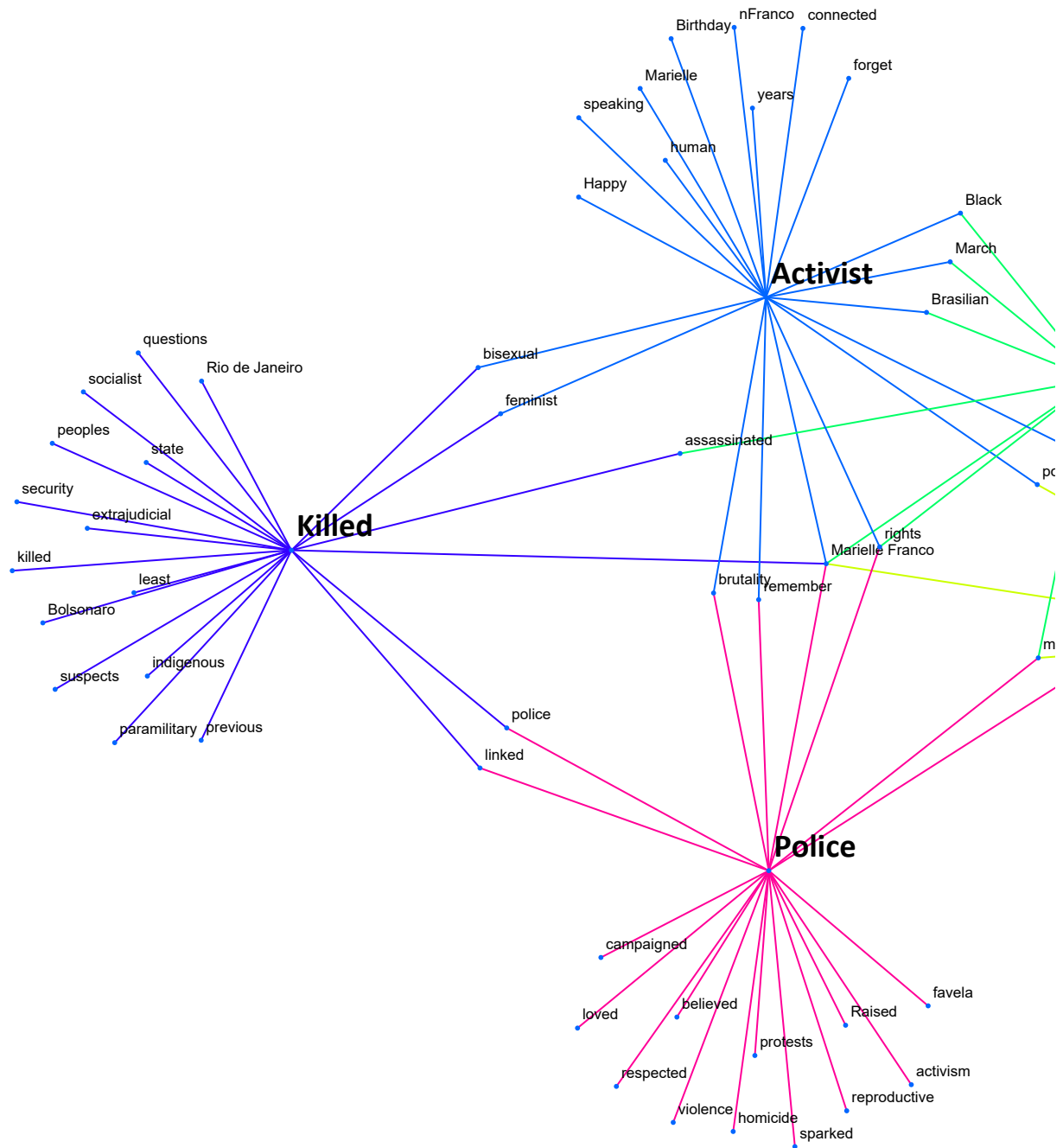
In[ ]:= Clear[TweetsTopicGraph]
TweetsTopicGraph[j_, k_ : 10] := Module[{t, mvert, edges, alledges, g},
  Do[t[n] = BasisVectorInterpretation[H[j][n], k, terms[j]]][All, 2];
  mvert[n] = Style[Capitalize[topiclabels[j][n]],
    FontFamily -> "Calibri", FontSize -> 20, Bold], {n, topics[j]};
  alledges = Flatten[
    Table[edges[n] = Map[UndirectedEdge[#, mvert[n]] &, t[n][1 ;;]], {n, topics[j]};
  g = Graph[alledges, VertexLabels -> "Name", ImageSize -> 1000, PlotTheme -> "LargeGraph"];
  HighlightGraph[g, Table[Map[Style[#, Hue[n / 10]] &, edges[n]], {n, topics[j]}]]
];

```

Marielle Franco em inglês com 20 palavras por tópico:

```
In[ ]:= MFES = TweetsTopicGraph["en", 20]
```

```
Out[ ]:=
```



```
In[ ]:= Export["MF2020EN.png", MFES]
```

```
Out[ ]:=
```

MF2020EN.png