

# Topic Change Analysis

*Joaquim Neto*

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
setwd("~/PyCharm-Workspace/tribunaldb/")
```

```
load('ts_data.RData')
```

## Objetivos:

- 1) Descobrir possíveis pontos de virada, aonde o *ofensor* passa de um comportamento normal para um comportamento tóxico
- 2) Descobrir os principais tópicos para cada *timeslice*
- 3) Descobrir quais as principais tendências de troca de tópico

(ie. topico a tende a mudar para topico b)

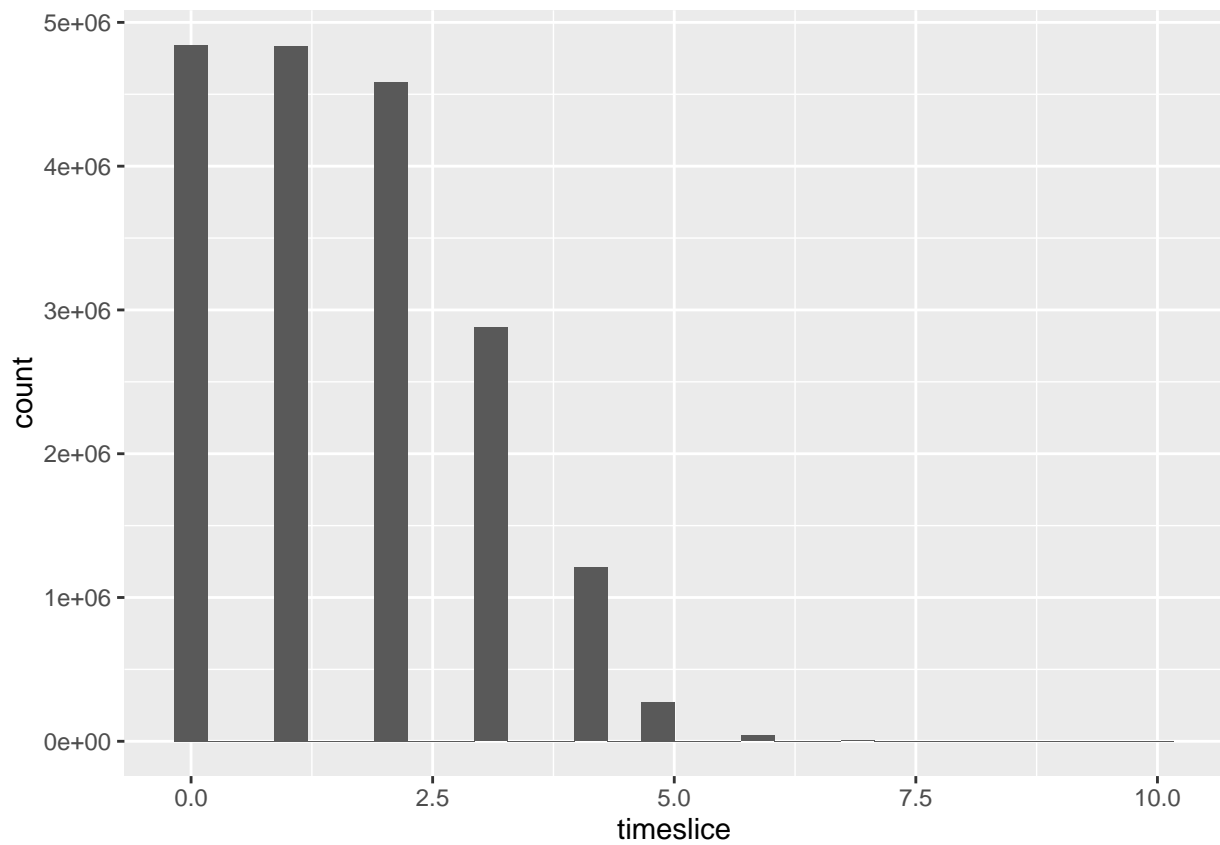
## Parâmetros

**Tamanho de cada timeslice:**

600s ##### Clusterizador e dados usados: LDA, sem divisão de grupos por timeslice, dataset NA completo.  
##### Peso usado na interpretação tf(cluster)-idf(corpus) aplicado nas top-100 palavras de cada cluster.

## Preprocessamento

- 1) Removido timeslices acima de 60min(engloba ~95% das tslices)

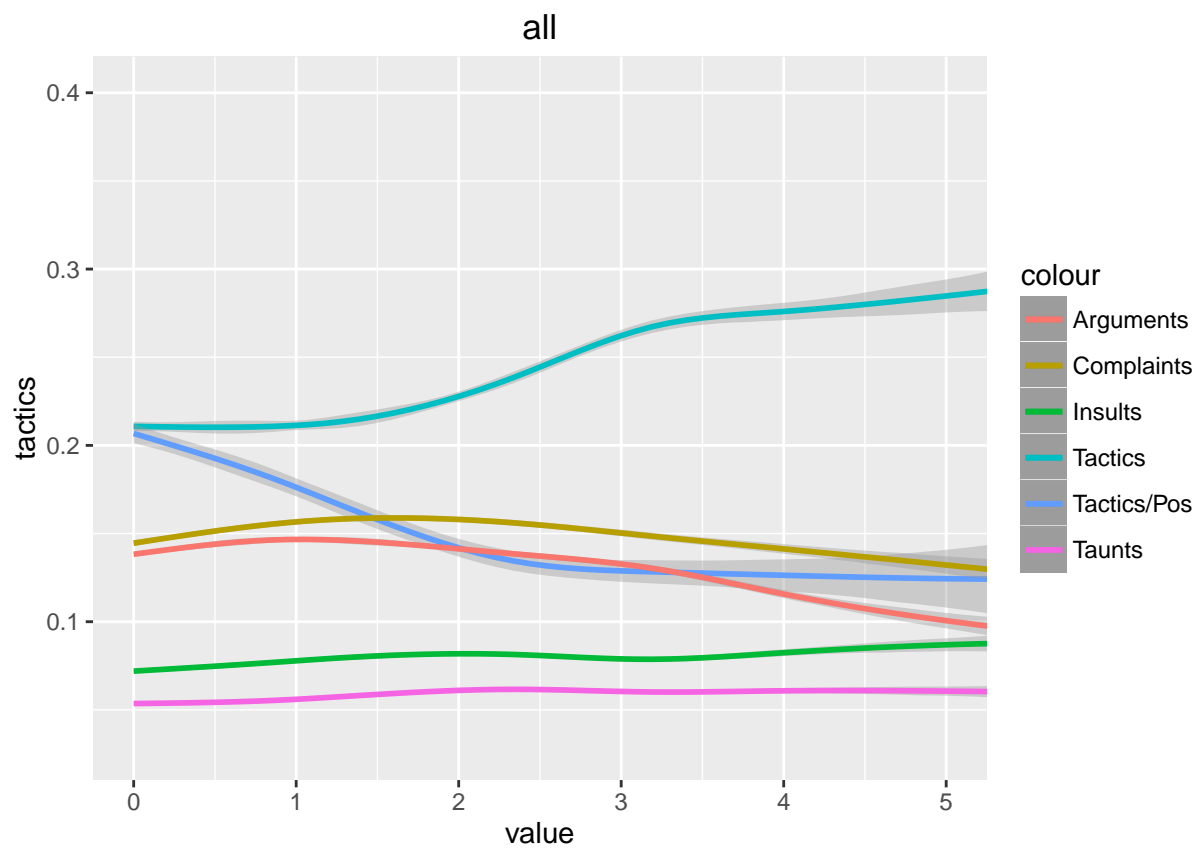


## Análise

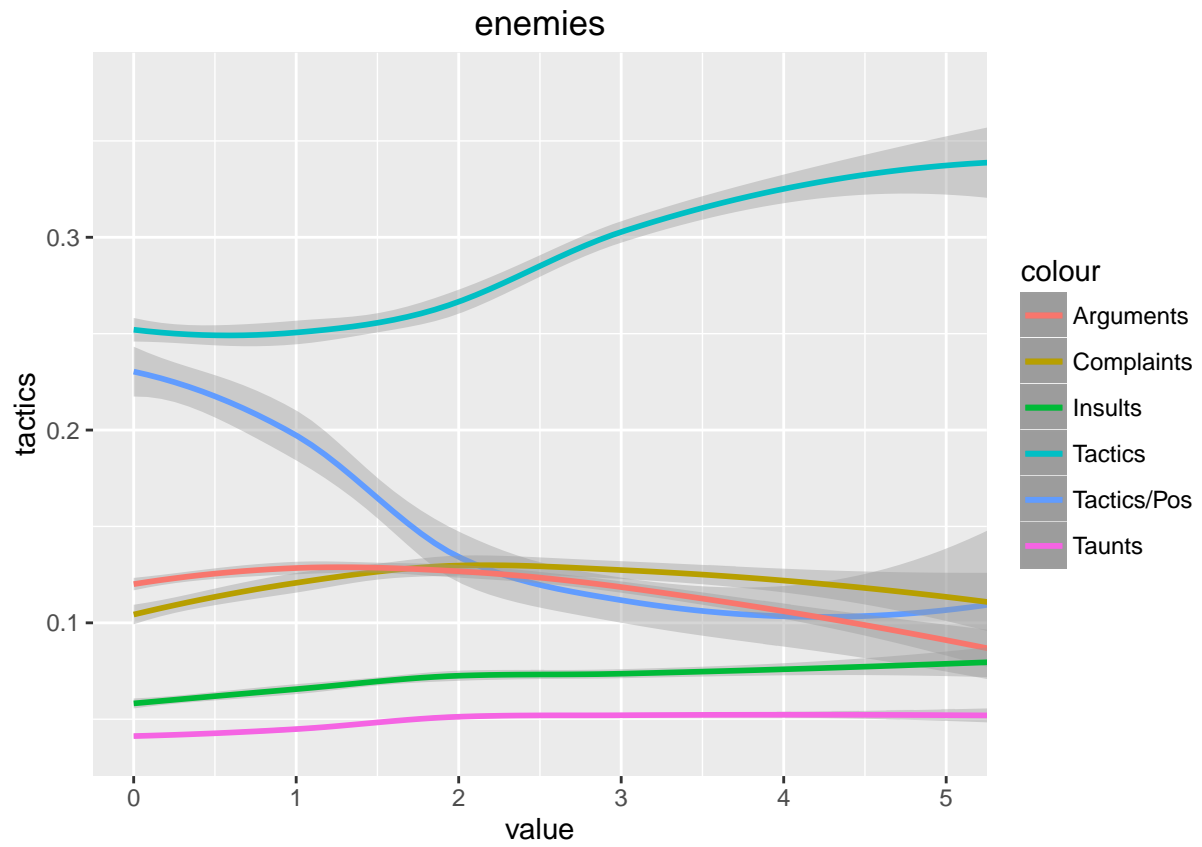
Uso dos tópicos através dos timeslices

Método:

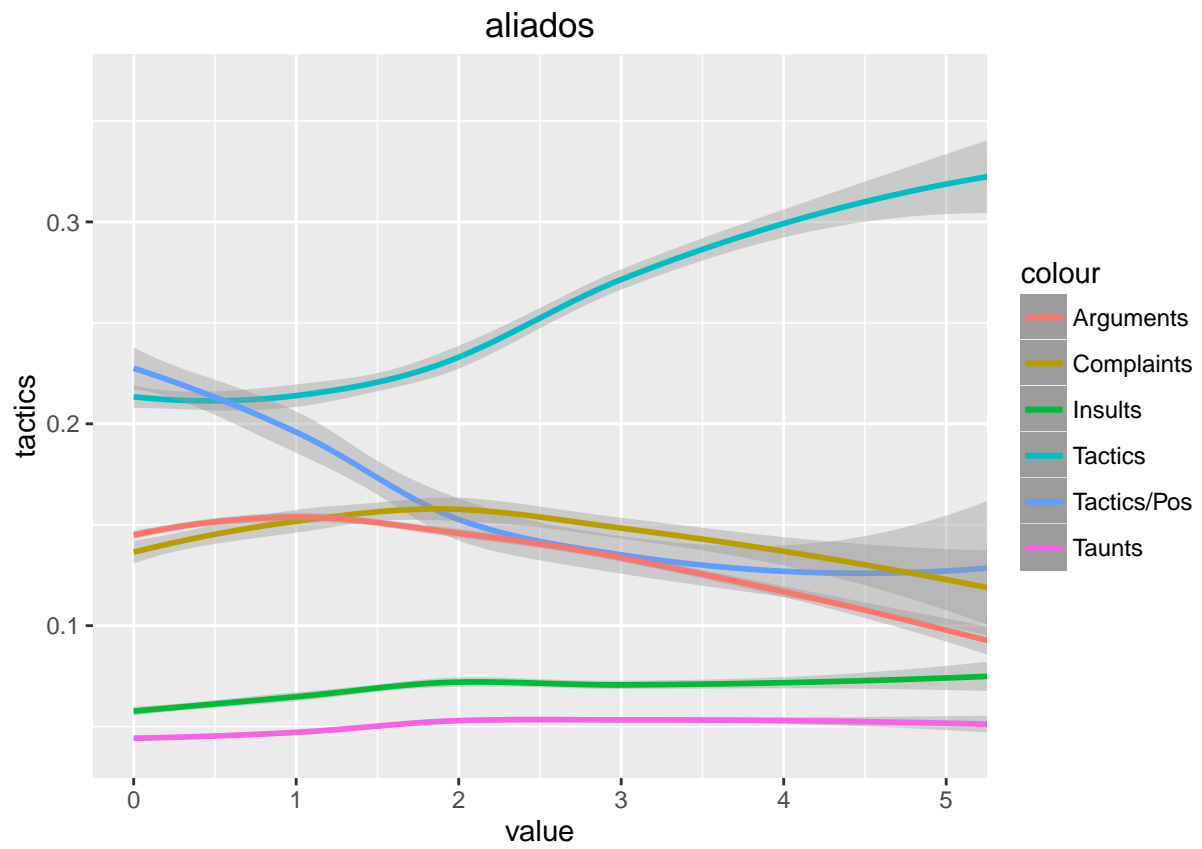
Todos os grupos



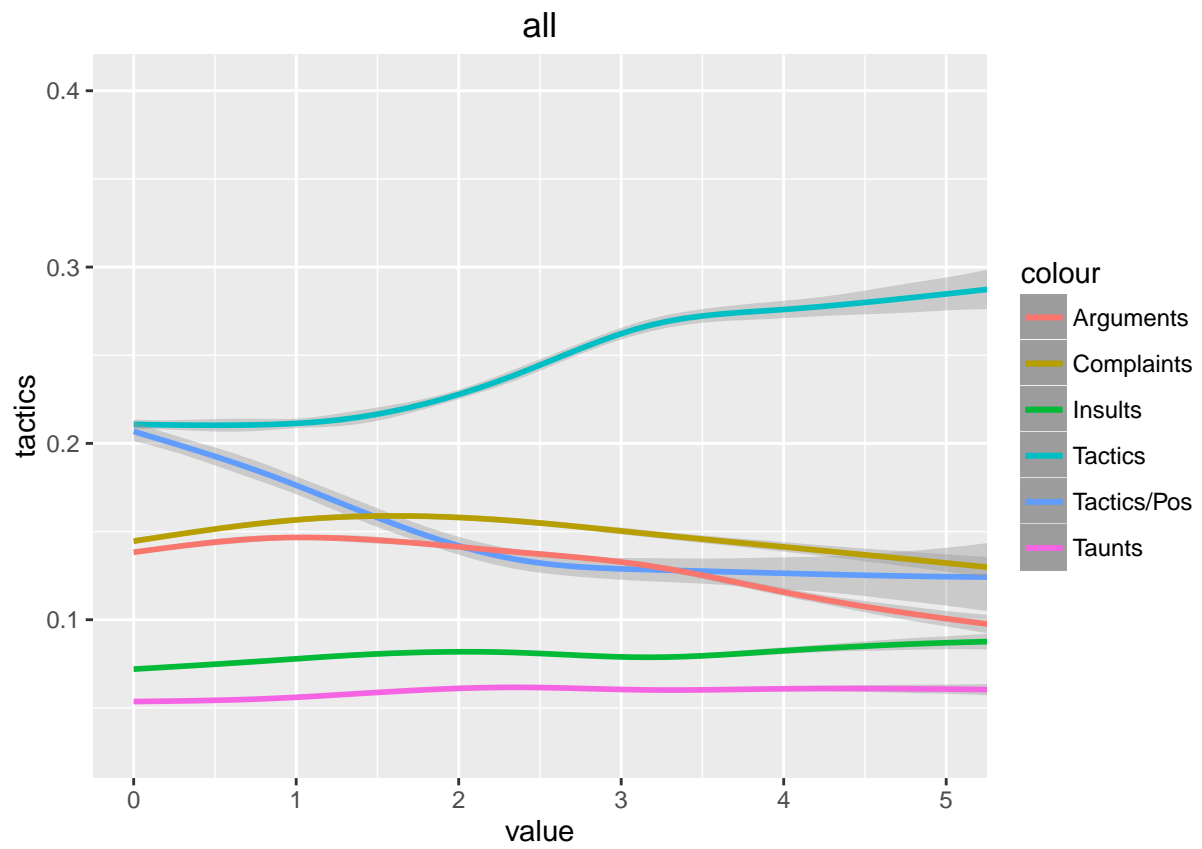
Grupos inimigos



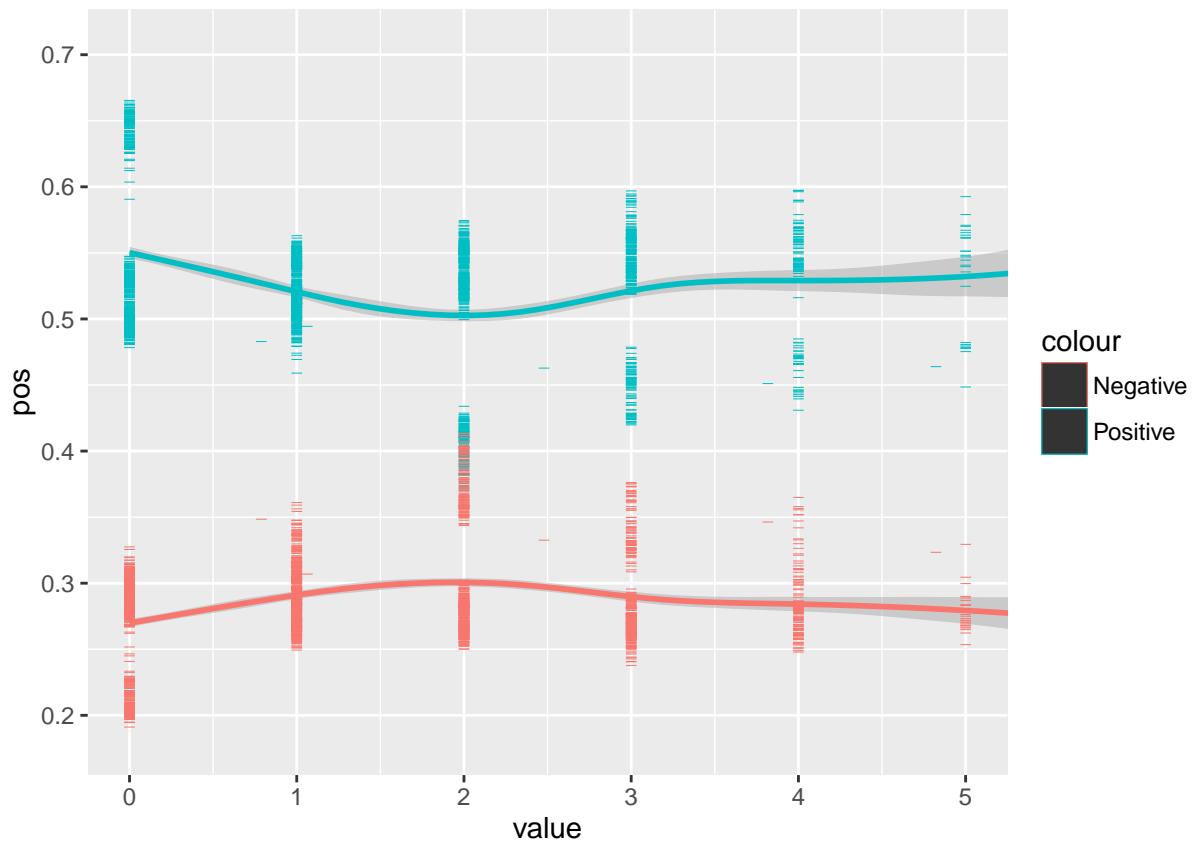
Grupos aliados



Grupos ofensores



Todos os grupos, divisão em pos/neg



## Relação entre tópicos em diferentes timeslices

### Experimento 1:

Apriori com 50% de confiança e 33% de suporte, limitado as timeslices 1 e 2, intervalo no qual os gráficos apresentam alguma variação

**ts.1=Tactics**

ts2.tactics -> ts.1.tactics com 38% de suporte e 100% de confiança.

**ts.1=Tactics.Pos,**

Nenhum resultado relevante.

**ts.1=small.talk**

Nenhum resultado relevante.

**ts.1=complaints**

Nenhum resultado relevante.

**ts.1=arguments**

Nenhum resultado relevante.

**ts.1=insults**

Nenhum resultado relevante.

**ts.1=taunts**

Nenhum resultado relevante.

## **Conclusões**

Não parecem existir relações significantes

## **Experimento 2:**

Apriori com 50% de confiança e 33% de suporte, limitado as timeslices 2 e 3, intervalo no qual os gráficos apresentam alguma variação

**ts.2=Tactics**

Nenhum resultado relevante.

**ts.2=Tactics.Pos,**

$\{ts.3=empty\} \Rightarrow \{ts.2=tactics.pos\}$ , com 44% de suporte

**ts.2=small.talk**

$\{ts.3=empty\} \Rightarrow \{ts.2=small.talk\}$  com 40% de suporte

**ts.2=complaints**

$\{ts.3=empty\} \Rightarrow \{ts.2=complaints\}$  com 42% de suporte

**ts.2=arguments**

$\{ts.3=empty\} \Rightarrow \{ts.2=arguments\}$  com 38% de suporte

**ts.2=insults**

$\{ts.3=empty\} \Rightarrow \{ts.2=insults\}$  com 43% de suporte

**ts.2=taunts**

$\{ts.3=empty\} \Rightarrow \{ts.2=taunts\}$  com 43% de suporte



### **Conclusões**

Esses resultados não parecem significar muita coisa além de uma quantidade maior de ‘empty’(sem conversas) em ts.3 do que em ts.2, o que de fato acontece(43% vs 14%).

### **Experimento 3:**

Apriori com 33% de confiança e 5% de suporte, usando todas as timeslices, e todos os tópicos. Suporte é baixo devido aos desbalanceamentos entre os grupos.

### **Conclusões**

Sem resultados conclusivos, de 300+ regras, a maioria esmagadora dizia de algum modo que empty -> empty e topic -> topic