



## Introdução a Data Science (INF 1335)

### Exercício 5 • Rede de Citações

O objetivo deste trabalho é analisar as citações a artigos publicados na série de conferências IHC, com base nos seguintes arquivos:

- `artigo.csv`: os artigos publicados nas diversas edições do evento
- `keyword.csv`: as palavras-chave de cada artigo
- `autor.csv`: os autores de cada artigo
- `pessoa.csv`: autores de artigos ou de referências (ainda não foi processada/gerada uma tabela com os autores das referências)
- `instituicao.csv`: as instituições dos autores de cada artigo
- `referencia.csv`: as referências citadas por cada artigo

Observações:

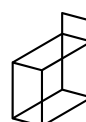
- 1 Quase todas as perguntas devem envolver análise numérica, apresentação visual dos dados e texto descrevendo/explicando/discutindo os dados.
- 2 As perguntas e gráficos a seguir são apenas uma amostra. Vocês devem acrescentar as perguntas e gráficos que julgarem úteis para a análise.

#### Algumas perguntas interessantes para a análise

- 1 Quantos artigos foram publicados em cada edição (ano) do evento?
- 2 Quantos autores publicaram em cada edição do evento?
- 3 Como o número de autores por artigo estão distribuídos a cada edição?
- 4 Como a proporção entre referências em português e em inglês tem variado ao longo dos anos?
- 5 Quais autores mais colaboram (publicam artigos em coautoria)?
- 6 Quais autores mais citam outros autores?
- 7 Quais autores têm maior proporção autocitação<sup>1</sup>/citação?
- 8 Como são distribuídos os números de referências citados por cada artigo? Variam muito de edição para edição?
- 9 Quais conferências são mais citadas pelos artigos publicados no IHC (tipo de referência = C)? E quais periódicos (tipo de referência = J)?

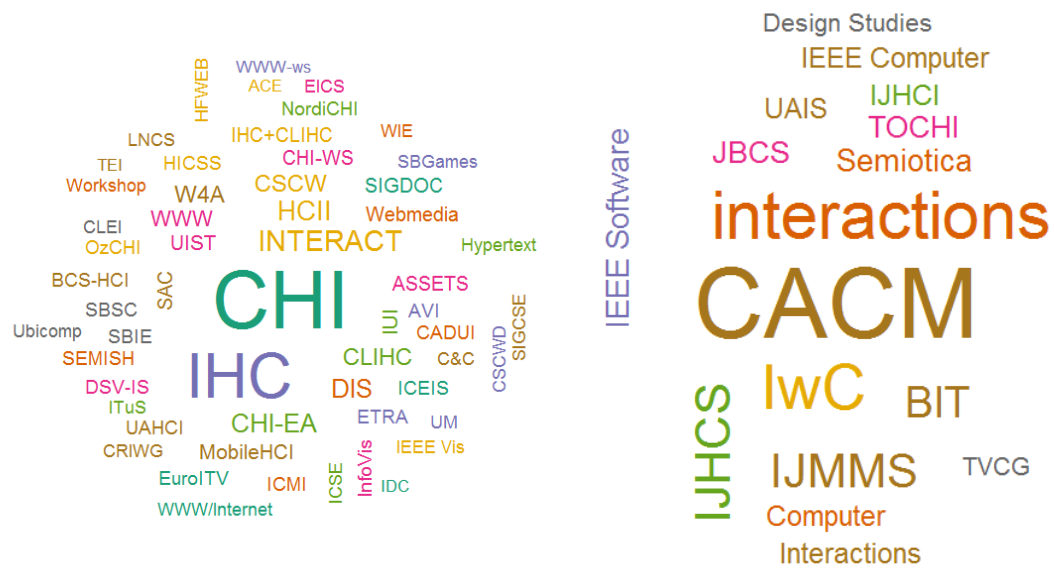
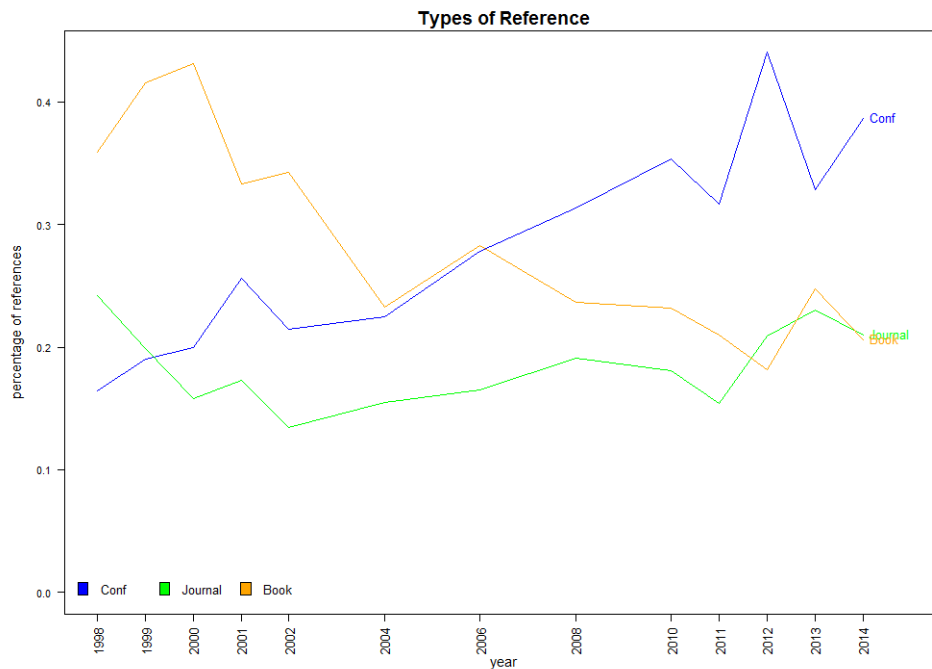
---

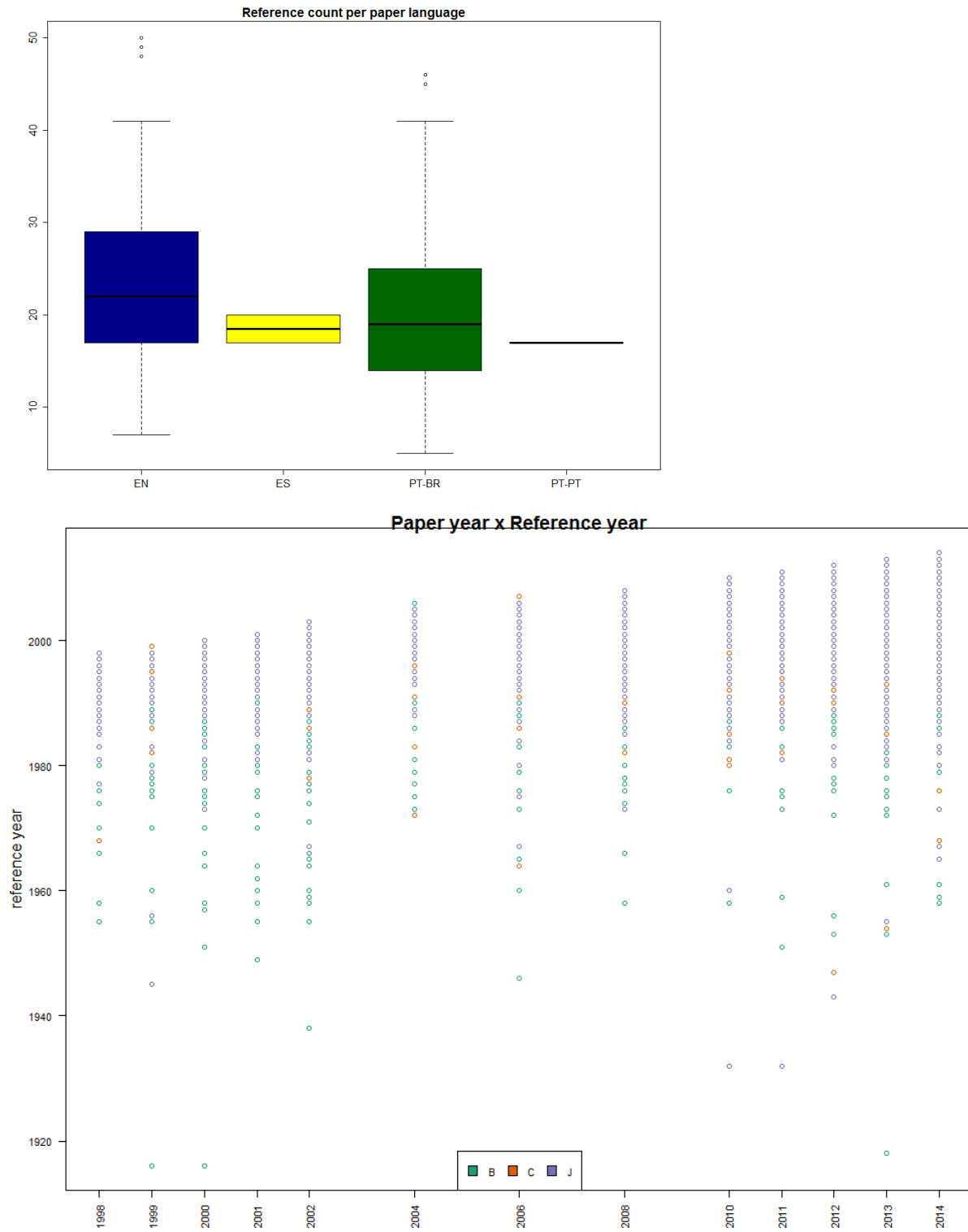
<sup>1</sup> A autocitação ocorre quando o artigo de um autor cita outro artigo redigido por ele próprio (possivelmente em coautoria).

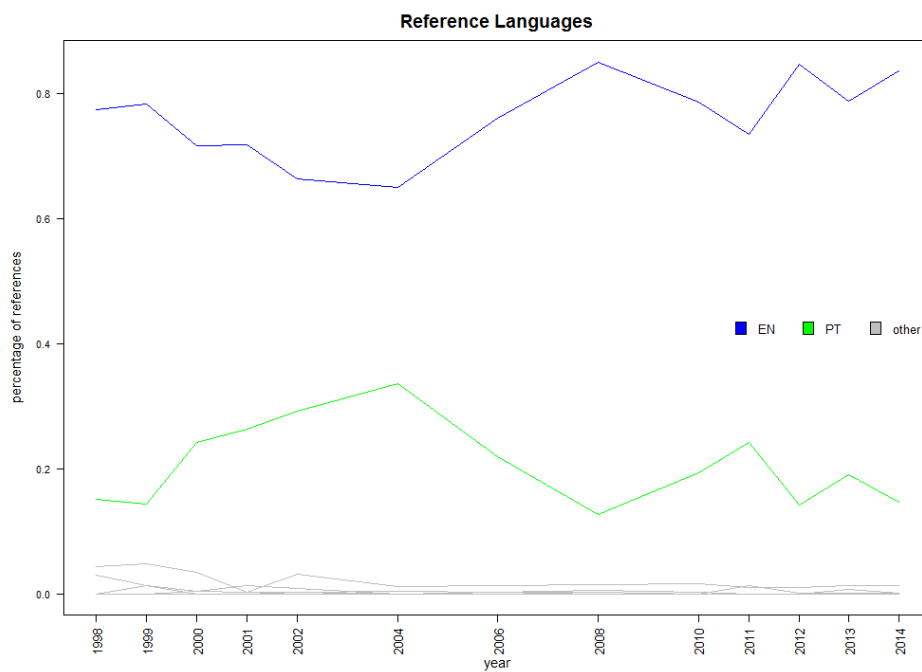
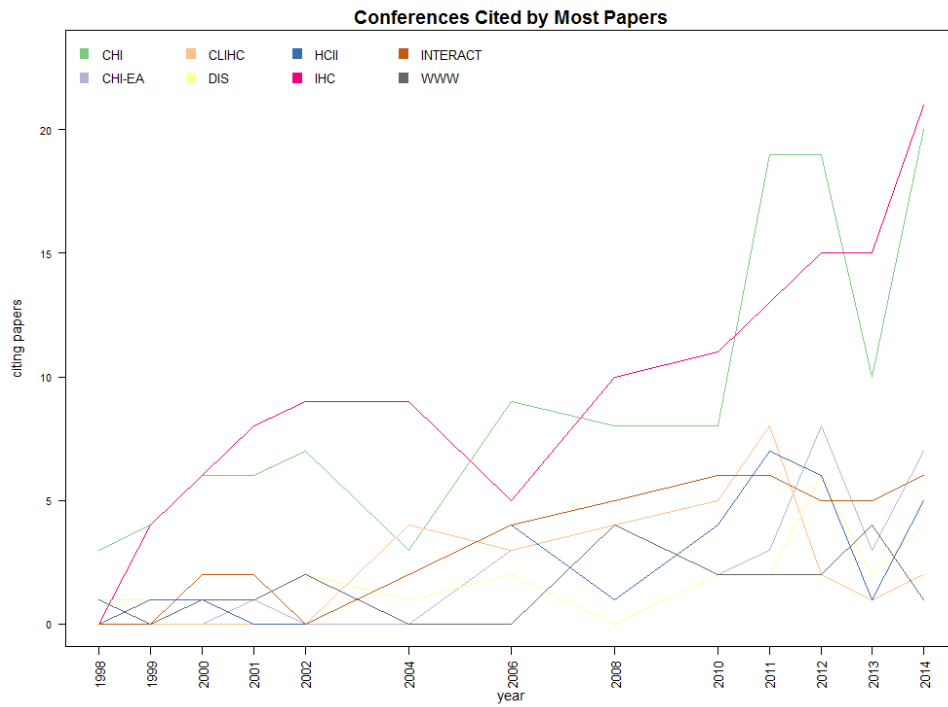


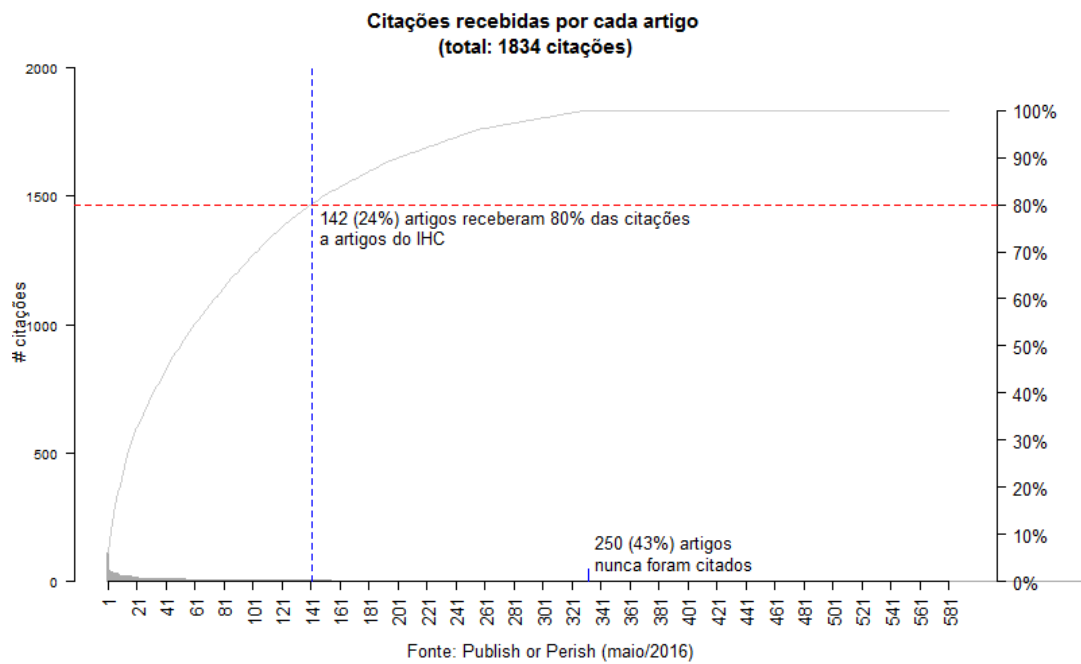
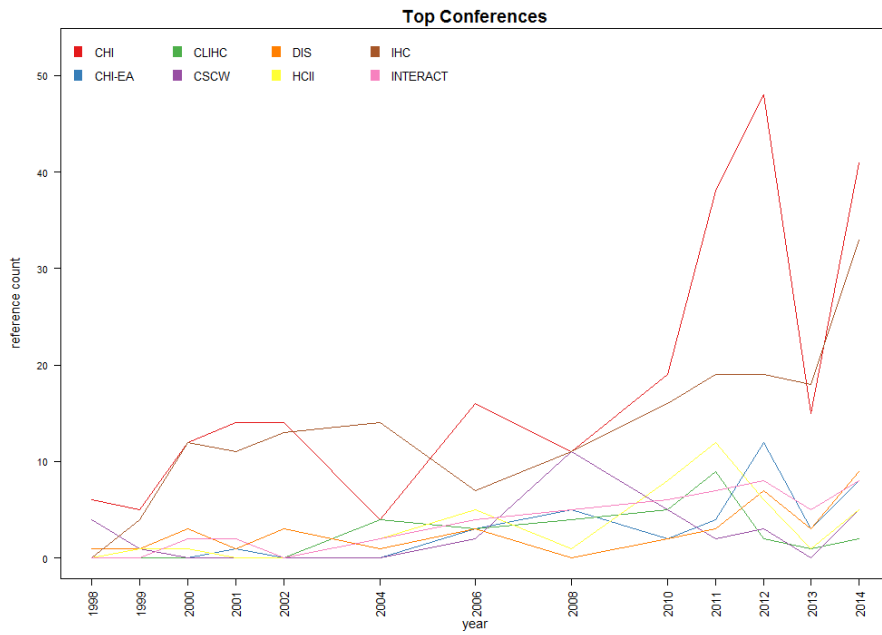
## Alguns gráficos interessantes

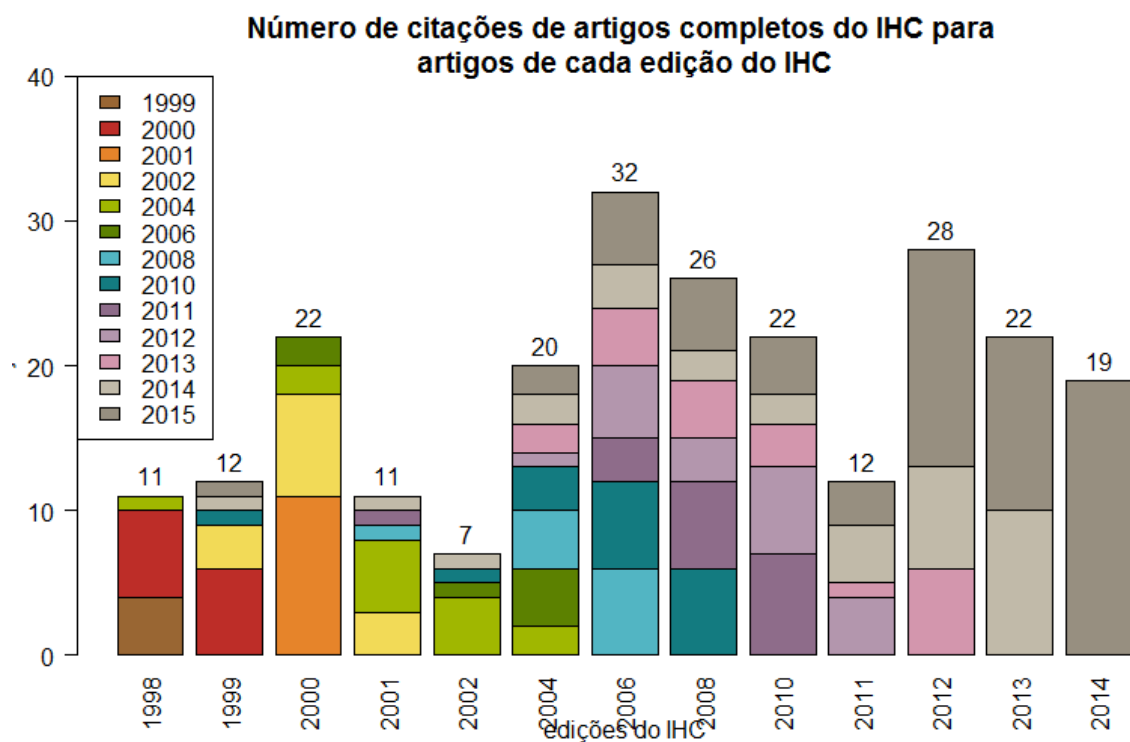
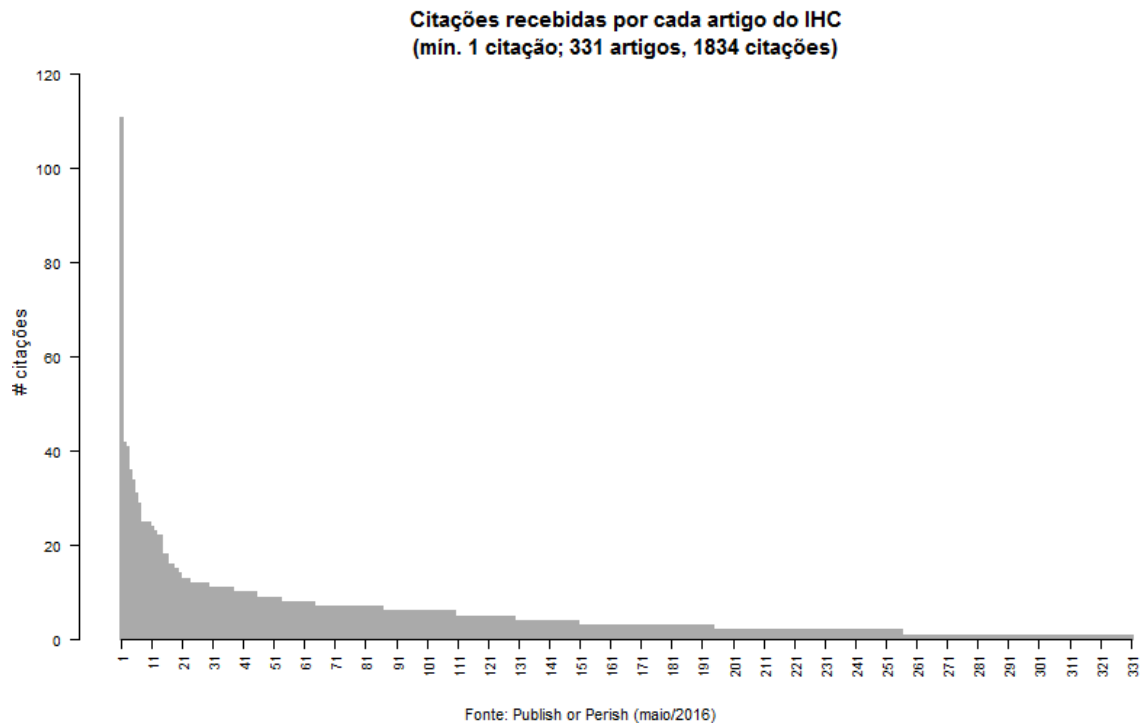
Observação: Esses gráficos foram feitos sobre uma versão antiga da base, e portanto vocês podem obter resultados ligeiramente diferentes.

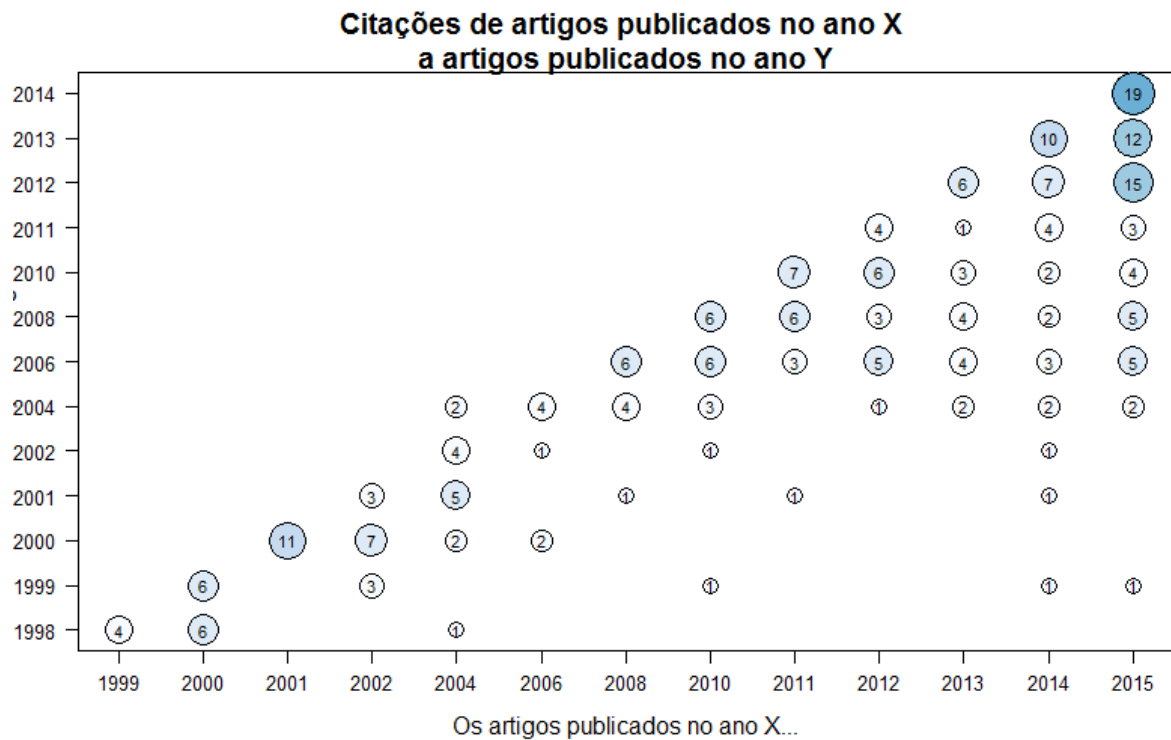




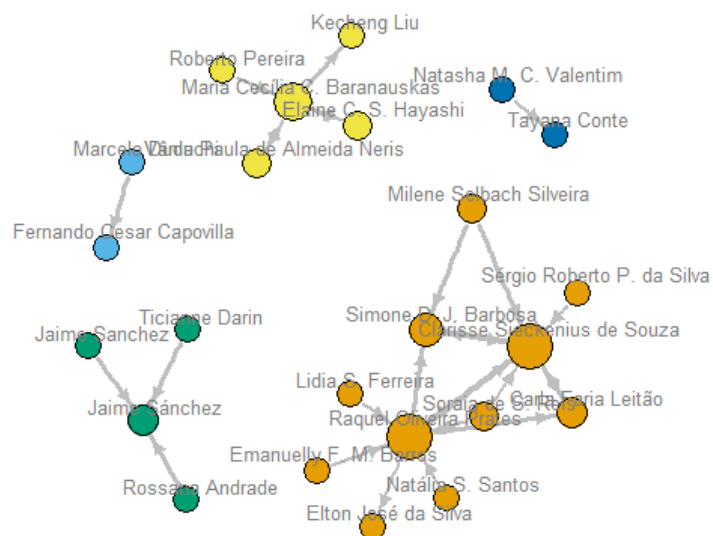








### Citações mais frequentes



\* Arestas com peso  $\geq 15$