Temas de pesquisa na pós-graduação em filosofia e sua relação com gênero:

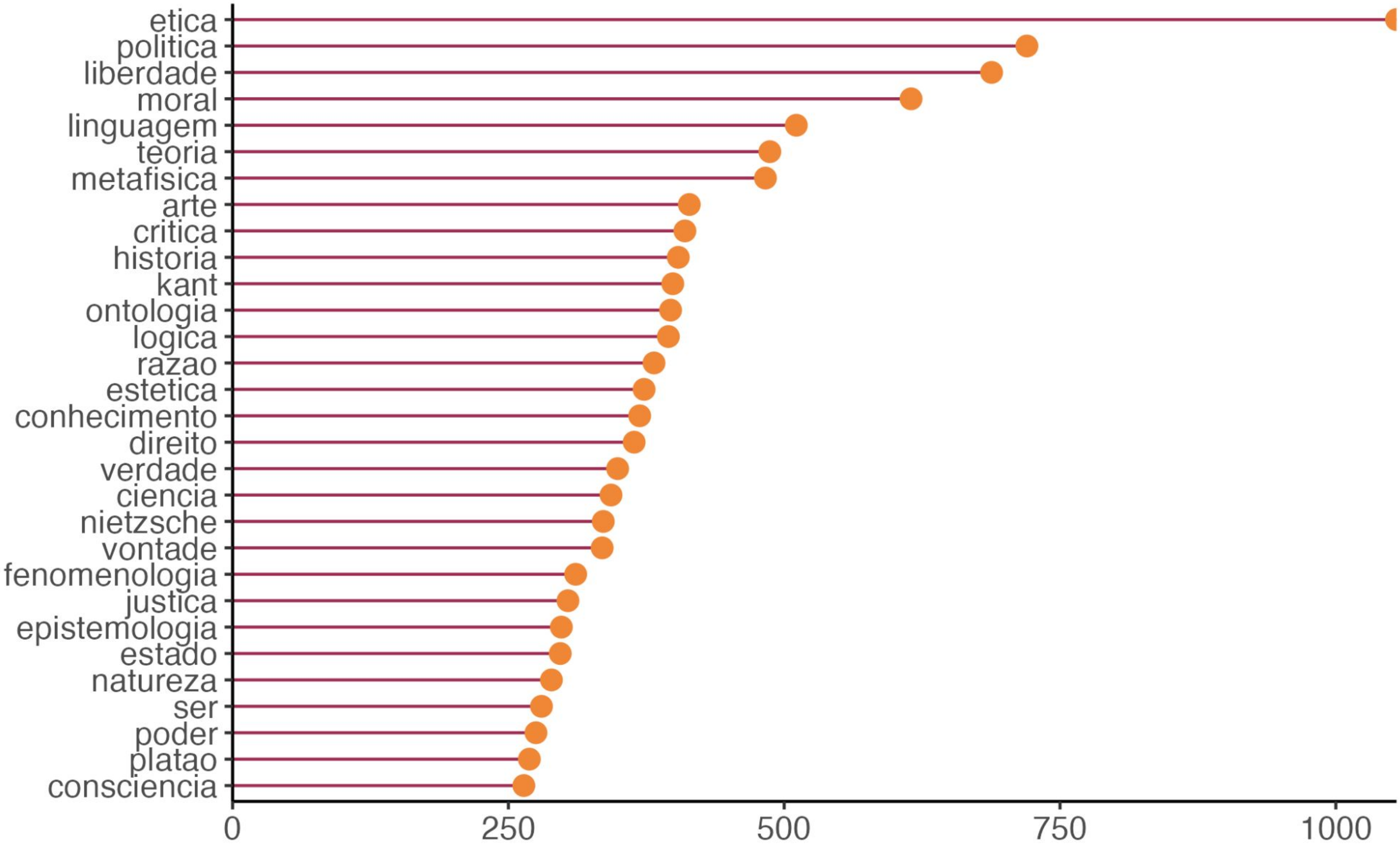
uma análise de teses e dissertações a partir da modelagem estruturada de tópicos

Marcos Fanton (UFSM) | marcos.fanton@ufsm.br

Modelagem Estruturada de Tópicos (STM)

30 palavras mais frequentes em palavras-chaves

Teses e Dissertações de Filosofia (1987-2020)



Dados: CAPES | Elaboração: Dataphilo

STM :: Conceito

Modelagem de tópicos é uma classe de métodos de clusterização (não-supervisionada) que permite inferir tópicos específicos a partir de um *corpus* de documentos e das palavras que cada documento possui.

A análise estruturada permite adicionar covariáveis (metadados de cada documento, como ano de publicação, gênero, raça, ideologia do autor) ao modelo para aprimorar a qualidade da inferência e da interpretação dos tópicos.

BLEI (2012) Probabilistic topic models.

ROBERTS *et al.* (2014) Structural topic models for open-ended survey responses.

ROBERTS *et al.* (2016) A model of text for experimentation in the social sciences.

STM :: Usos recentes na filosofia

MALATERRE et al. (2019). *What is this thing called philosophy of science? A computational topic-modelling perspective, 1934-2015*. HOPOS.

MALATERRE et al. (2020) *The recipes of philosophy of science: characterizing the semantic structure of corpora by means of topic associative rules*. PLOS ONE.

NOICHL (2021) *Modeling the structure of recent philosophy*. Synthese. 198.

WEATHERSON (2022) *A History of Philosophy Journals. Volume 1: Evidence from Topic Modeling, 1876-2013*. Bookdown.

Modelo *bag of words* (*bow*)

Representação simplificada de cada documento através da frequência de palavras contidas nele. Tudo o que importa é o **número de palavras** - ordem, contexto e outras nuances gramaticais e sintáticas não são captadas.

Modelo *bag of words* (*bow*)

Etapas:

1. Escolha da unidade de análise :: resumo de trabalhos
2. *Tokenização* :: separação do resumo em palavras
3. Redução da complexidade:
3.1. Transformação em caixa baixa
3.2. Remoção da pontuação
3.3. Formação de *bi-* e *trigrams* ('razao', 'pura' = 'razaopura')
3.4. Remoção de *stopwords* ('de', 'a', 'o', 'para', etc.)
3.5. Criação de classes equivalentes (Schofield, Mimmo, 2016):
3.5.1. Transformação em *stems*
3.5.2. Transformação em *lemmas*
3.6. Filtragem de *tokens* por frequência ('filosofia', 'ideia')
3.6.1. Filtragem por falta de relevância semântica ('capitulo', 'tese')
3.6.2. Filtragem por classe gramatical da palavra (*POS - parts of speech tagging*)
3.7. Criação da matriz de termos e documento

Adaptado de GRIMMER; ROBERTS; STEWART. (2022) *Text as Data*. Princeton.

Resumo: “Como entender a afirmação de Kant de que os conceitos puros do entendimento são derivados do entendimento puro? Este problema se impõe na medida em que as categorias são conceitos e os conceitos são cognições cuja forma é a mesma para todos e cuja matéria é sempre oriunda da sensibilidade. Por sua vez, tal dificuldade nos deixa outro problema: compreender como podemos distinguir duas categorias puras, sendo que elas não possuem matéria e têm a mesma forma comum. [...]”

	doc_id	word	ano	gorientador	n
1	4223	afirmacao	2008	Female	1
2	4223	atentar	2008	Female	1
3	4223	categoria	2008	Female	3
4	4223	categorias	2008	Female	5
5	4223	causa	2008	Female	1
6	4223	cognicoes	2008	Female	1
7	4223	comum	2008	Female	1
8	4223	deixa	2008	Female	1
9	4223	derivacao	2008	Female	1
10	4223	derivada	2008	Female	1
11	4223	derivados	2008	Female	1
12	4223	distinguir	2008	Female	3
13	4223	entendimento	2008	Female	3
14	4223	exemplo	2008	Female	1
15	4223	exposto	2008	Female	1

	doc_id	diferenca	experiencia	fenomenal	platao	parmenides	categoria
1	4223	.	.	.	.	.	3
2	4224						

Corpus de teses e dissertações

Total de resumos - Código de Área ‘70100004’

n: 11766

↓

Resumos insuficientes (< 15 palavras)

n: 9050

↓

Resumos em outros idiomas

n: 9039

↓

Trabalhos duplicados e sem título

n: 9028

↓

Trabalhos com gênero de orientador não identificado

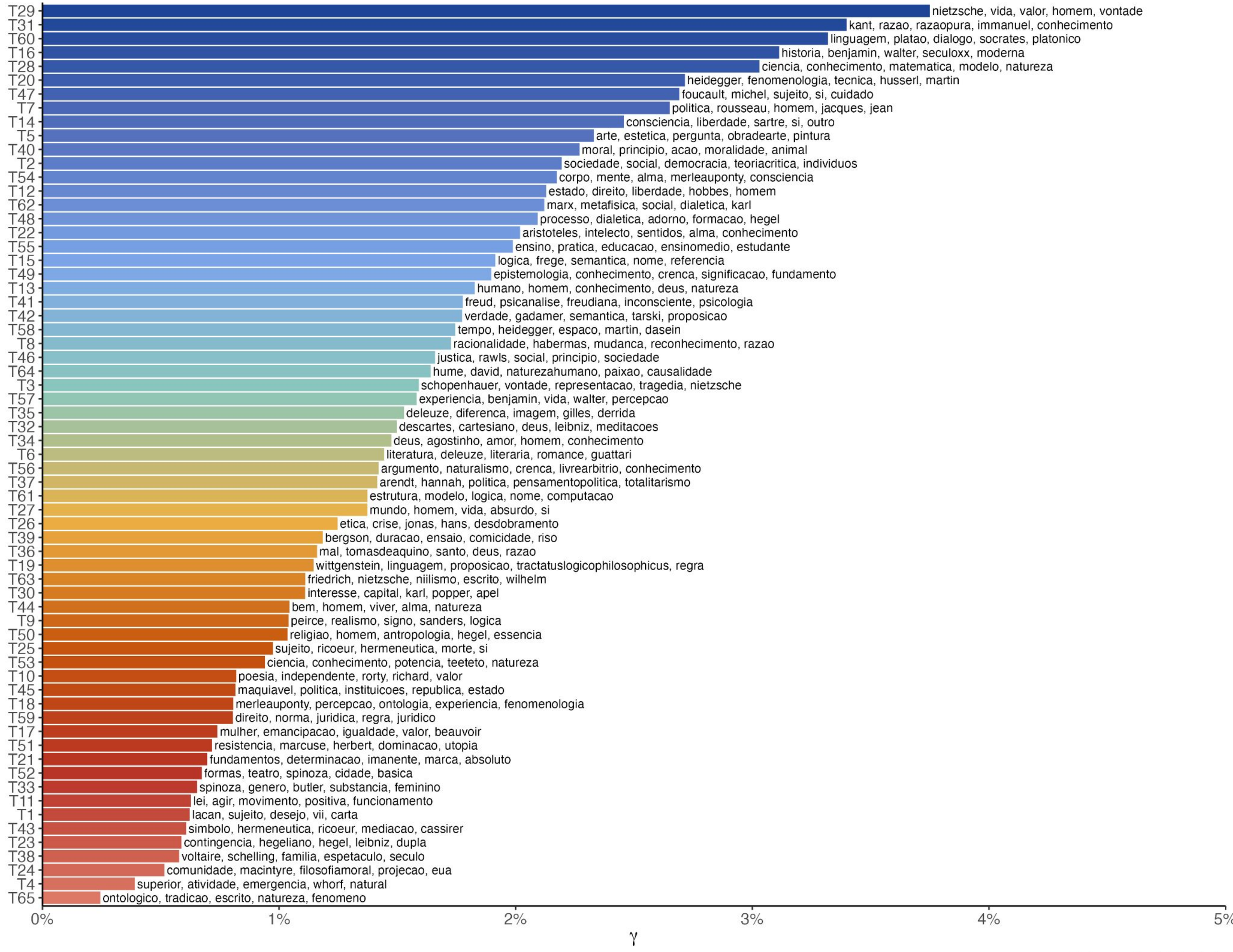
n: 8643

Resultados

Teses e Dissertações no modelo final: 8642
Número de tópicos: 65 tópicos
Dimensão: 30276 *tokens*
Trabalhos por gênero do orientador:
Homens (6900, 80%), Mulheres (1743, 20%)

65 tópicos do *corpus* de teses e dissertações de Filosofia (n: 8642)

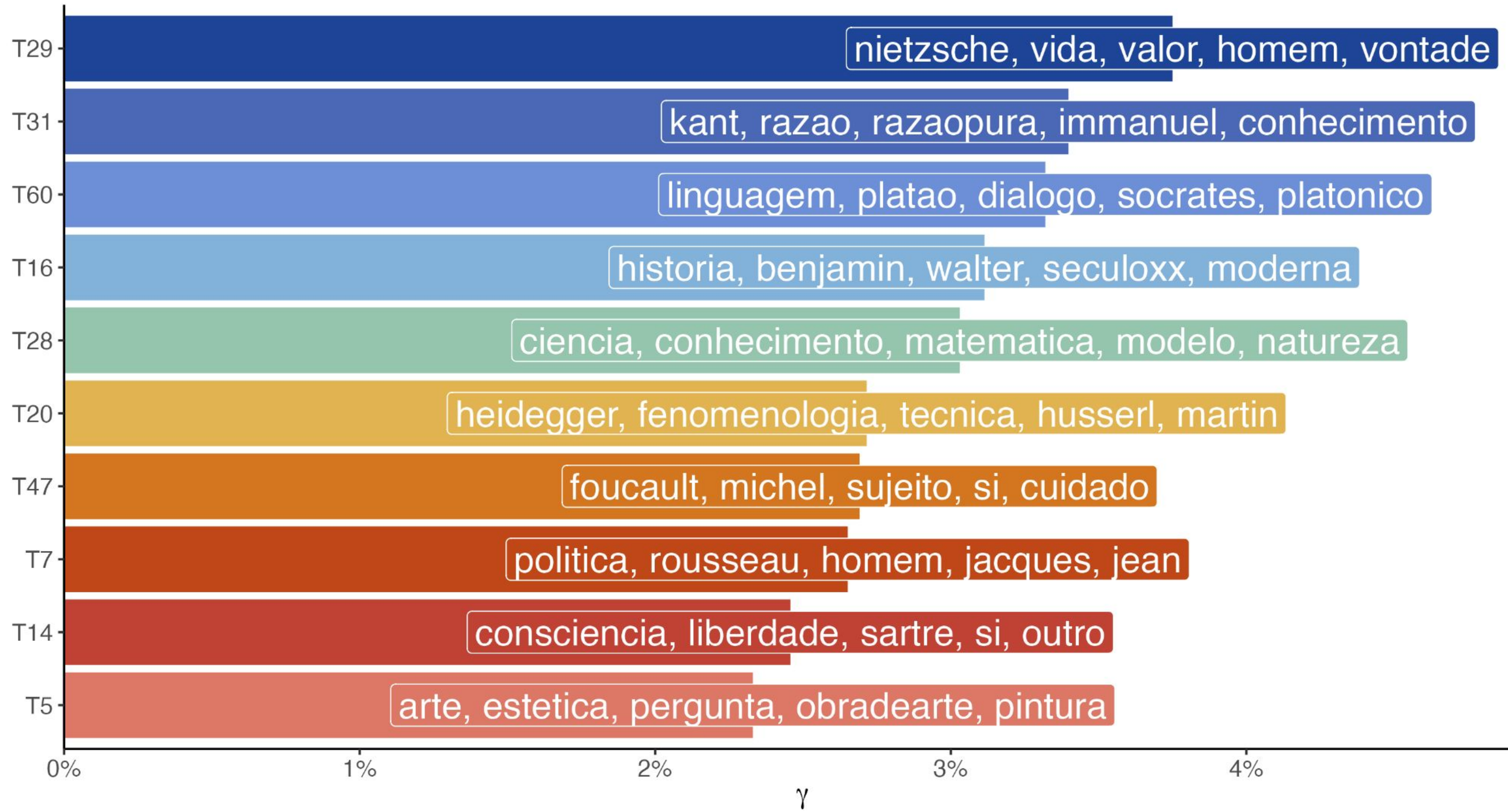
Com a probabilidade média esperada para cada tópico (γ)



Elaboração: Dataphilo

Os 10 tópicos com maior probabilidade média esperada (γ)

Corpus de teses e dissertações de Filosofia (n: 8642)

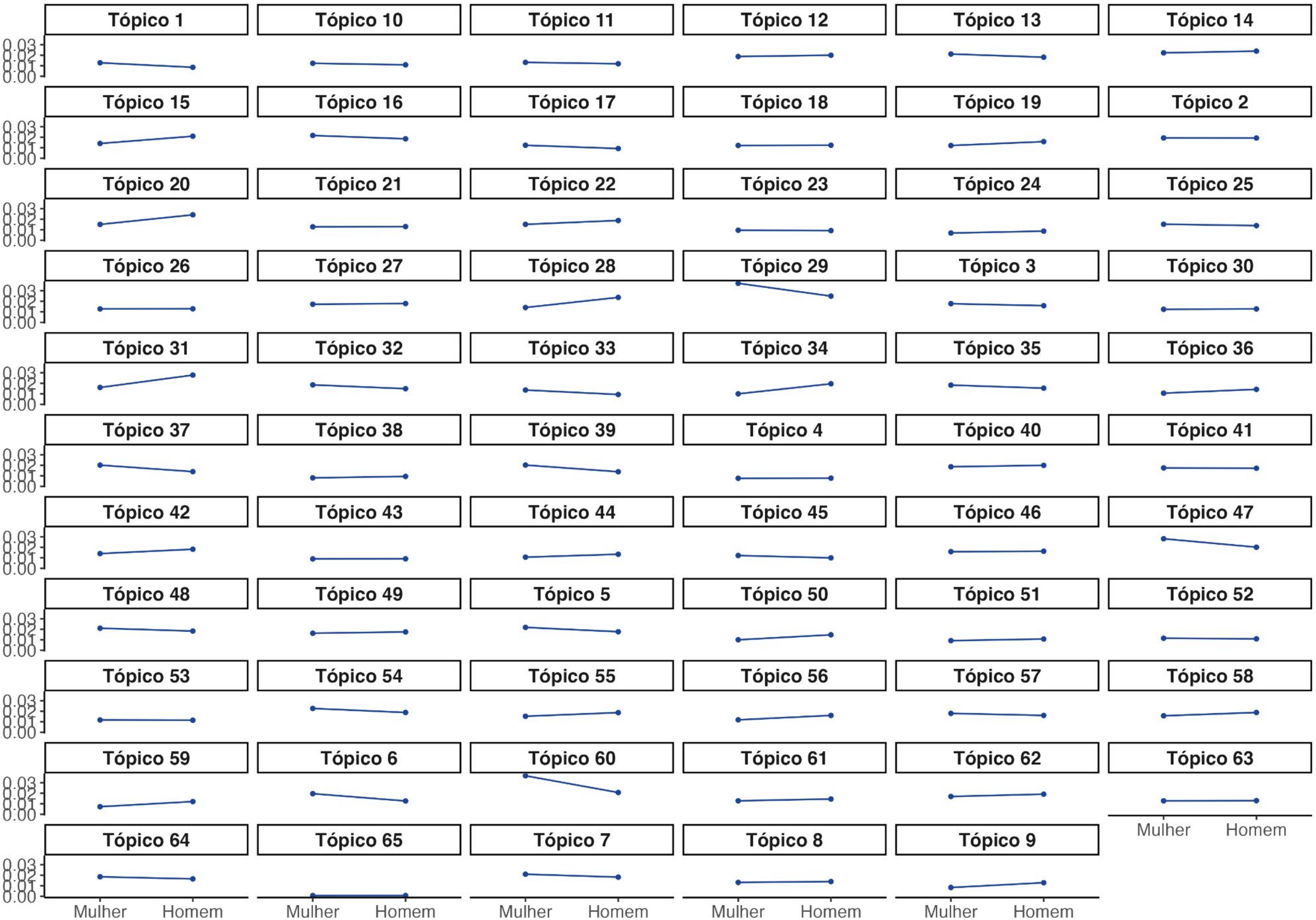


Elaboração: Dataphilo

Efeito de estimação pontual e intervalos de confiança dos tópicos ao longo do tempo

Elaboração: Dataphilo

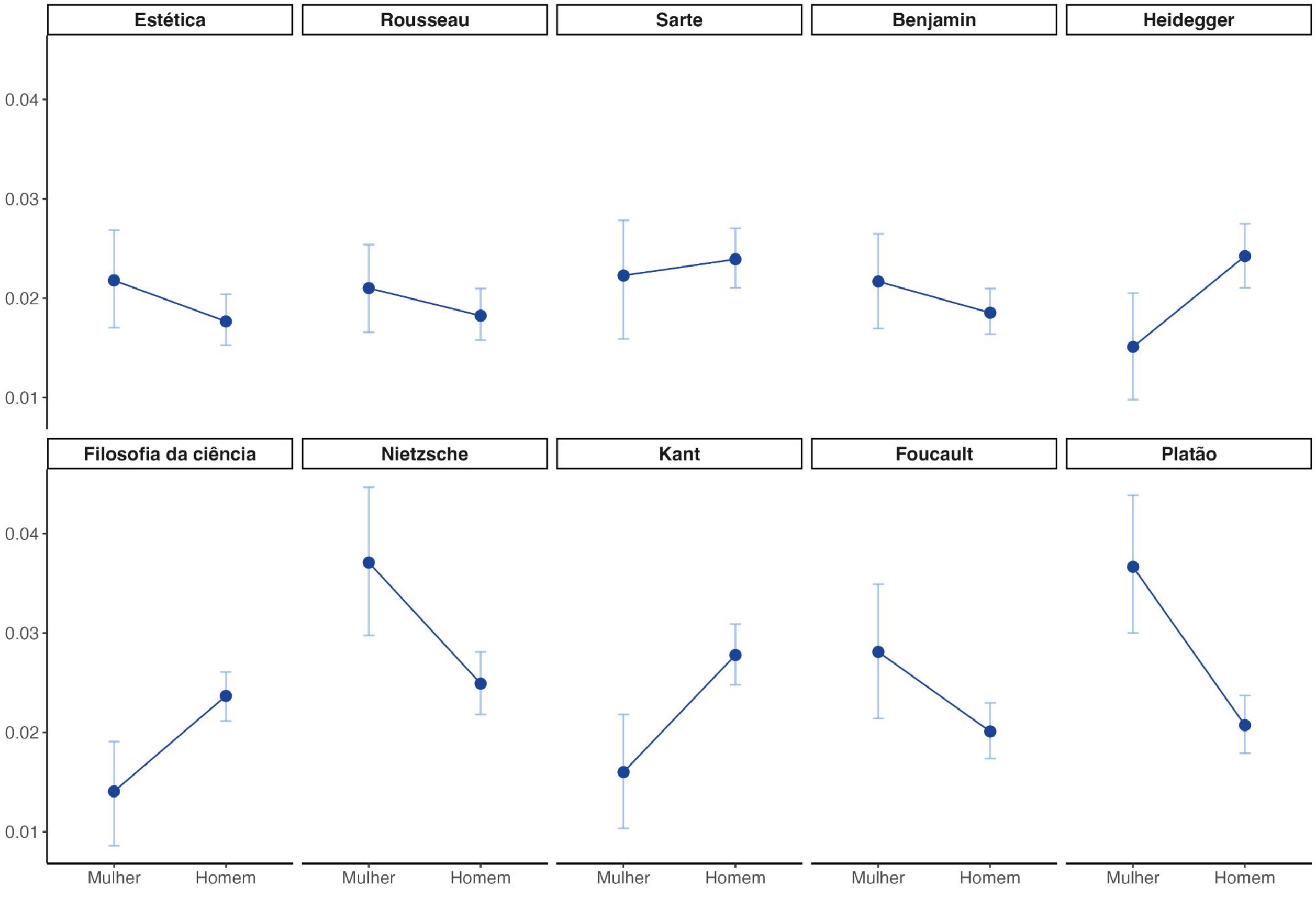
Efeito de estimação pontual dos tópicos por gênero do orientador



Elaboração: Dataphilo

Efeito de estimação pontual e intervalos de confiança dos tópicos por gênero do orientador

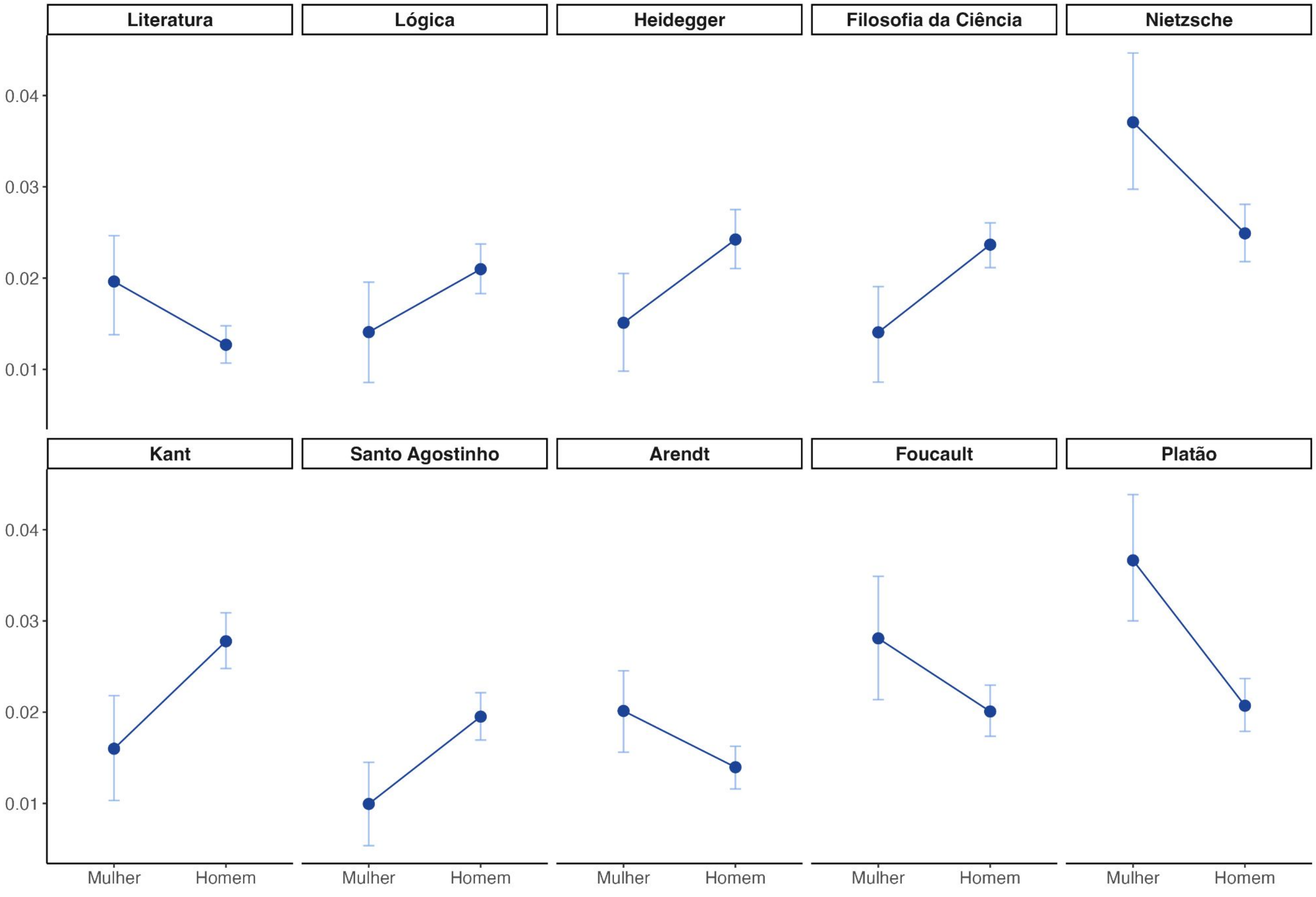
Os 10 tópicos com maior prevalência



Elaboração: Dataphilo

Efeito de estimação pontual e intervalos de confiança dos tópicos por gênero do orientador

Os 10 tópicos com maior diferença de gênero



Elaboração: Dataphilo

Para resumir:

1. A modelagem estruturada de tópicos encontrou 65 tópicos a partir do resumo de teses e dissertações da filosofia (de 1987-2020).

2. A STM permitiu analisar a tendência dos tópicos ao longo dos anos.

3. A STM permitiu explorar desigualdades de gênero 3.1. na produção de trabalhos finais; e 3.2. em tópicos específicos da pós-graduação em filosofia.

Futuros passos:

1. Redução da dimensão de tópicos através da criação de categorias (*clusters* de tópicos)
2. Análise da relação entre tópicos para validação do modelo
3. Análise dos efeitos das covariáveis (ano e gênero do orientador) nas categorias