LÍRICA MUSICAL CHILENA, SU ESTRUCTURA TEMÁTICA Y EVOLUCIÓN EN LOS ÚLTIMOS 25 AÑOS: EVIDENCIA DESDE EL MODELAMIENTO ESTRUCTURAL DE TÓPICOS

Jornadas Nacionales de Estadística 2026

Tomás Barrientos (tbarrientos@uc.cl)

Gonzalo Bustamante Rojas (gbustamantr@uc.cl)

Motivación

Intuición

La música, como un artefacto estético, contiene mensajes, tonalidades y estéticas propias del momento y contexto en el que se producen.

La experiencia estética de la música es mutable y representa la experiencia de un momento en el tiempo, dejando ver sus ideales sobre el presente y el futuro (Fisher, 2014)



Violeta Parra
est. 1938



Los Jaivas
est. 1963



Los Bunkers
est. 1999



Mon Laferte
est. 2003



Álex Anwandter
est. 2005



Pailita
est. 2019

Lo que dice la literatura

1 Charts anglosajones y alemanes, medidos con diccionarios

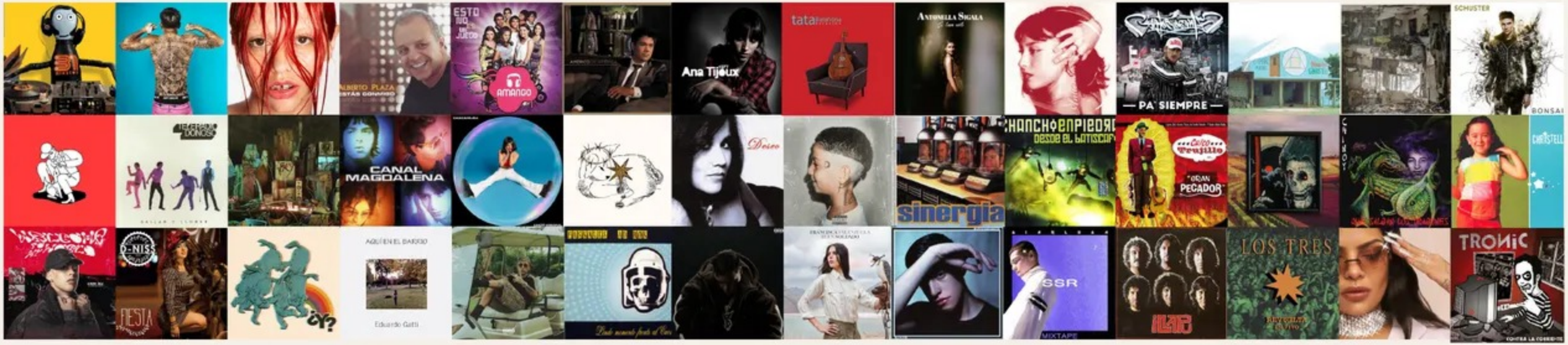
- LIWC, DICTION, codificación manual y métricas de complejidad (Pettijohn & Sacco, 2009; Christenson et al., 2019; Varnum et al., 2021; Anglada-Tort et al., 2021)
- Unidad de análisis: el chart, no la escena

2 Incipientes modelos de tópicos y ninguno en español

- LDA en listas alemanas (Hunke et al., 2025); BERTopic sobre Genius en inglés (Chen et al., 2025)
- Ninguno toma el campo documentado de un país

3 Literatura chilena: casos y escenas

- Hauntología (Allimant, 2024); indie 2005–2018 (Figueroa-Bustos, 2021); música urbana (Vásquez-Mejías et al., 2025)
- A nuestro saber, sin análisis corpus-level entre géneros que describa cambio por períodos



Pregunta de investigación y objetivos

Pregunta de investigación

¿Cómo cambió la composición temática de las letras del campo musical chileno documentado por musicapopular.cl entre 2000 y 2025?

Objetivo general

Estimar el cambio temático en las letras de la música popular chilena de 2000 a 2025

Objetivo específico

Identificar los temas de las letras y describir su evolución en el período.

Estimar el cambio temático con distintas unidades de análisis

Separar cuánto del cambio se explica por el recambio de cohortes de artistas y cuánto ocurre dentro de cada cohorte.

Corpus y modelado

Marco

MusicaPopular.cl
la enciclopedia de la música chilena

Definición de artistas activos entre el 2000 y 2025

Artistas = 2.177

Recolección

GENIUS

Recolección con API

Artistas = 458
Documentos =11.558

- **Unidad de Análisis**: La memoria activa de la lírica musical chilena
- **Unidad de Observación**: Letra de canción descompuesta en tokens

Preprocesamiento

- Lematización
- Bigrams
- Poda frecuencia <5
- Correcciones lematización
- Voseo
- "Chilenismos"

Corpus

Artistas= 363
Documentos= 8.572

Metadatos

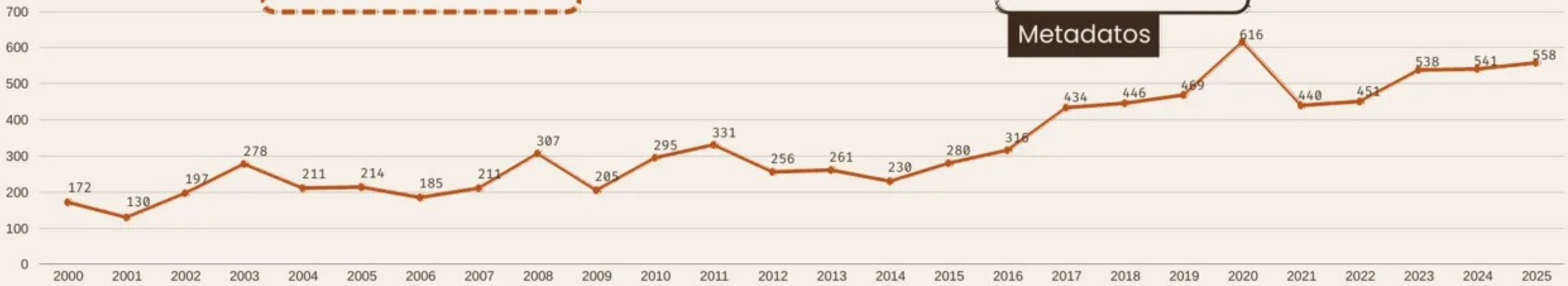
MusicBrainz
WIKIDATA

- Años de actividad
- Localidad de origen
- Géneros musicales
- Tipo artista

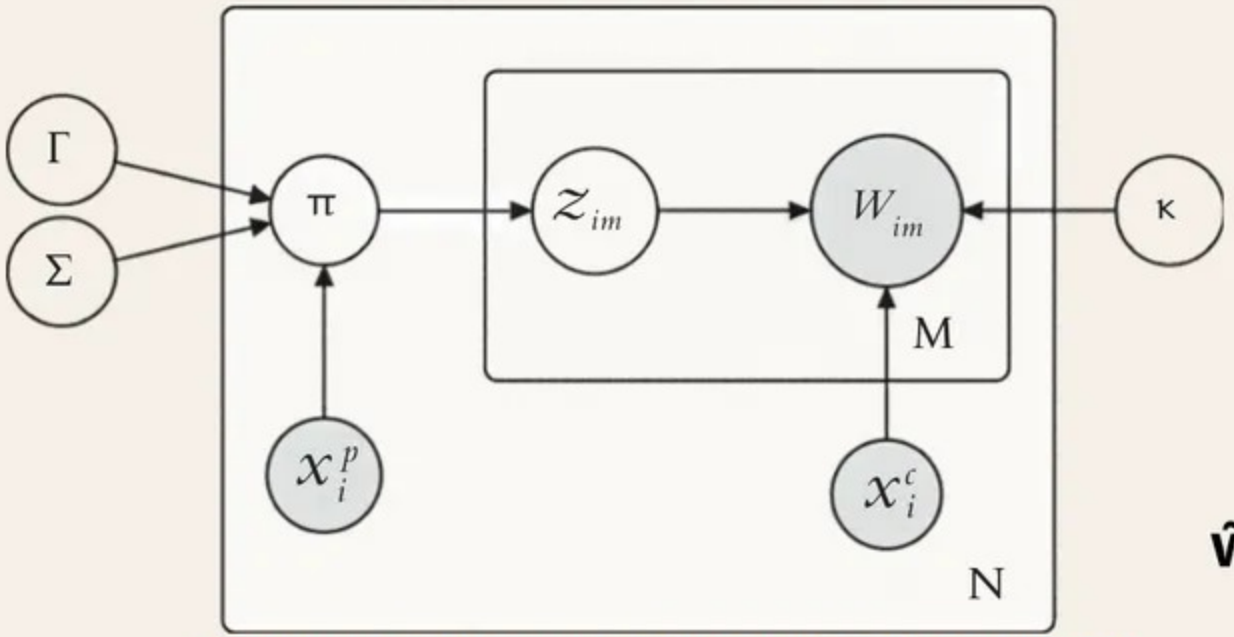
Modelado

STM

- K = 8
- prevalencia ~ s(año, df=10)
- Test de Wald sobre el efecto del año
- *Entropía de mezcla temática*



Structural Topic Modeling



En nuestro caso, al especificar la prevalencia por año estamos definiendo:

X^c_i Año de publicación

Donde

Γ, Σ: Parámetros globales: Γ conecta metadatos con prevalencia; Σ modela la covariación entre tópicos.

π_i: Proporciones de tópicos del documento i; resumen su mezcla temática.

X^P_i: Covariables de prevalencia: explican cuánto aparece cada tópico en el documento i.

z_{im}: Tópico latente asignado al token m del documento i.

X^c_i: Covariables de contenido: explican cómo se expresa un tópico en distintos grupos.

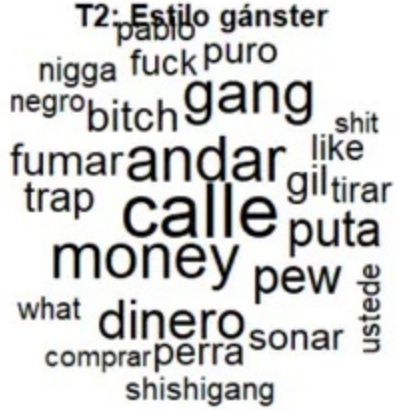
$\tilde{W}_{im}$ + K: Palabra observada + tasas globales que alimentan la distribución tópico–palabra.

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A Model of Text for Experimentation in the Social Sciences. Journal of the American Statistical Association, 111(515), 988-1003. https://doi.org/10.1080/01621459.2016.1141684

T1: Naturaleza y ensueño



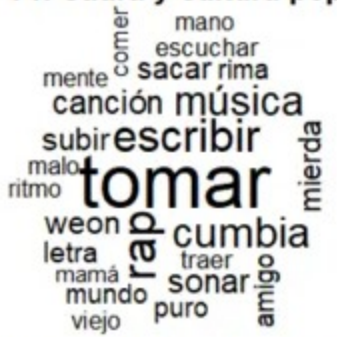
T2: Estilo gánster



T3: Fiesta y baile



T4: Sátira y cultura pop



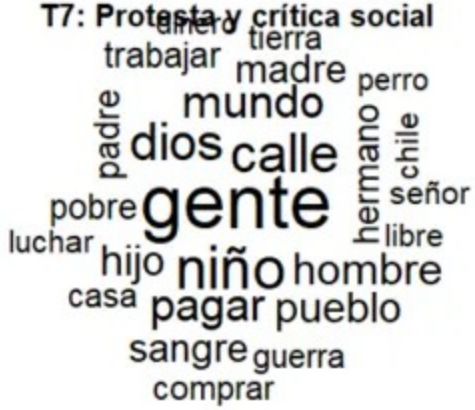
T5: Perreo y deseo



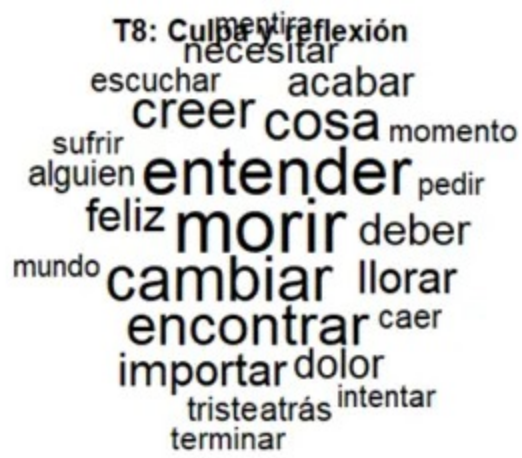
T6: Desamor y nostalgia

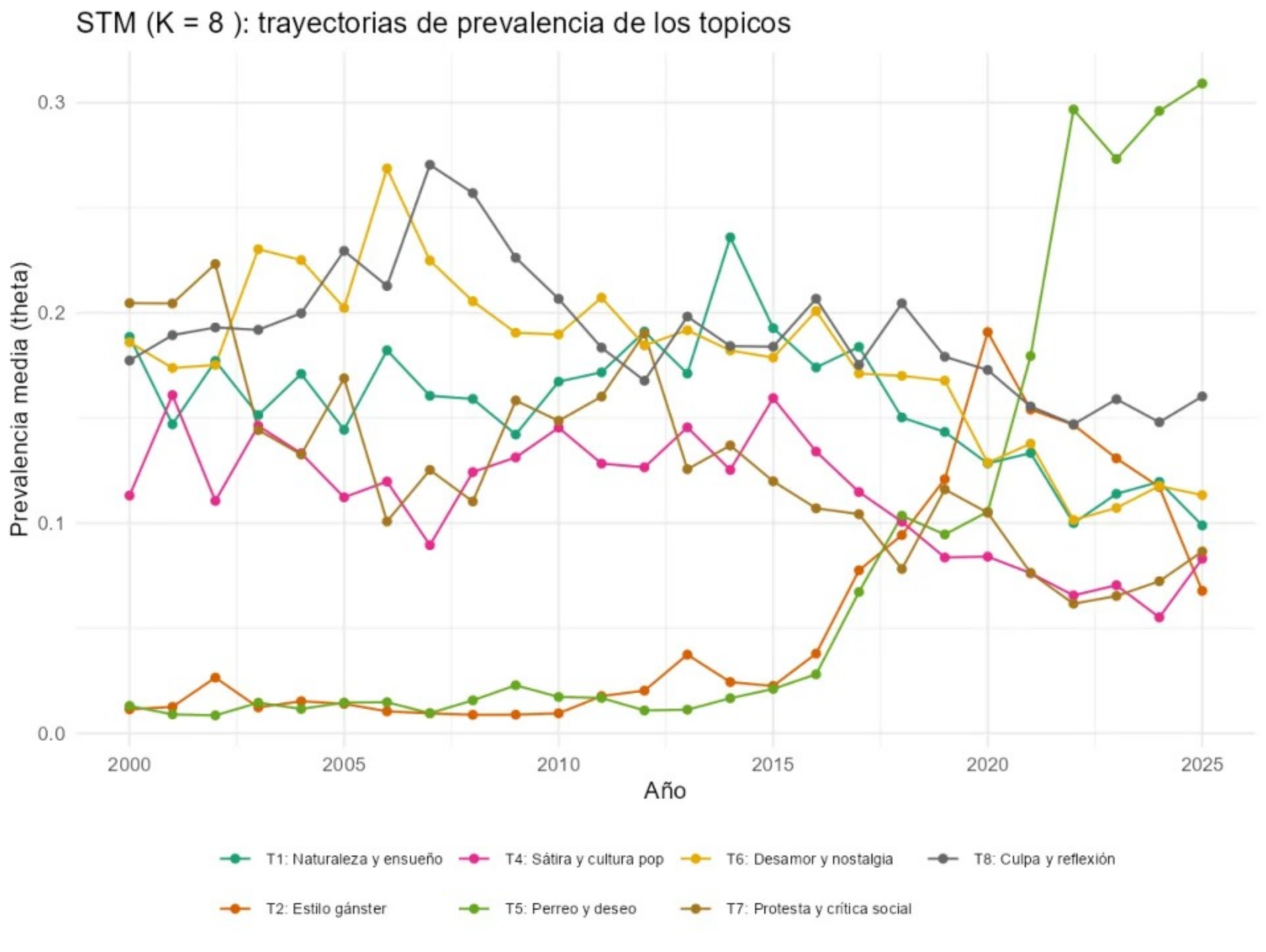


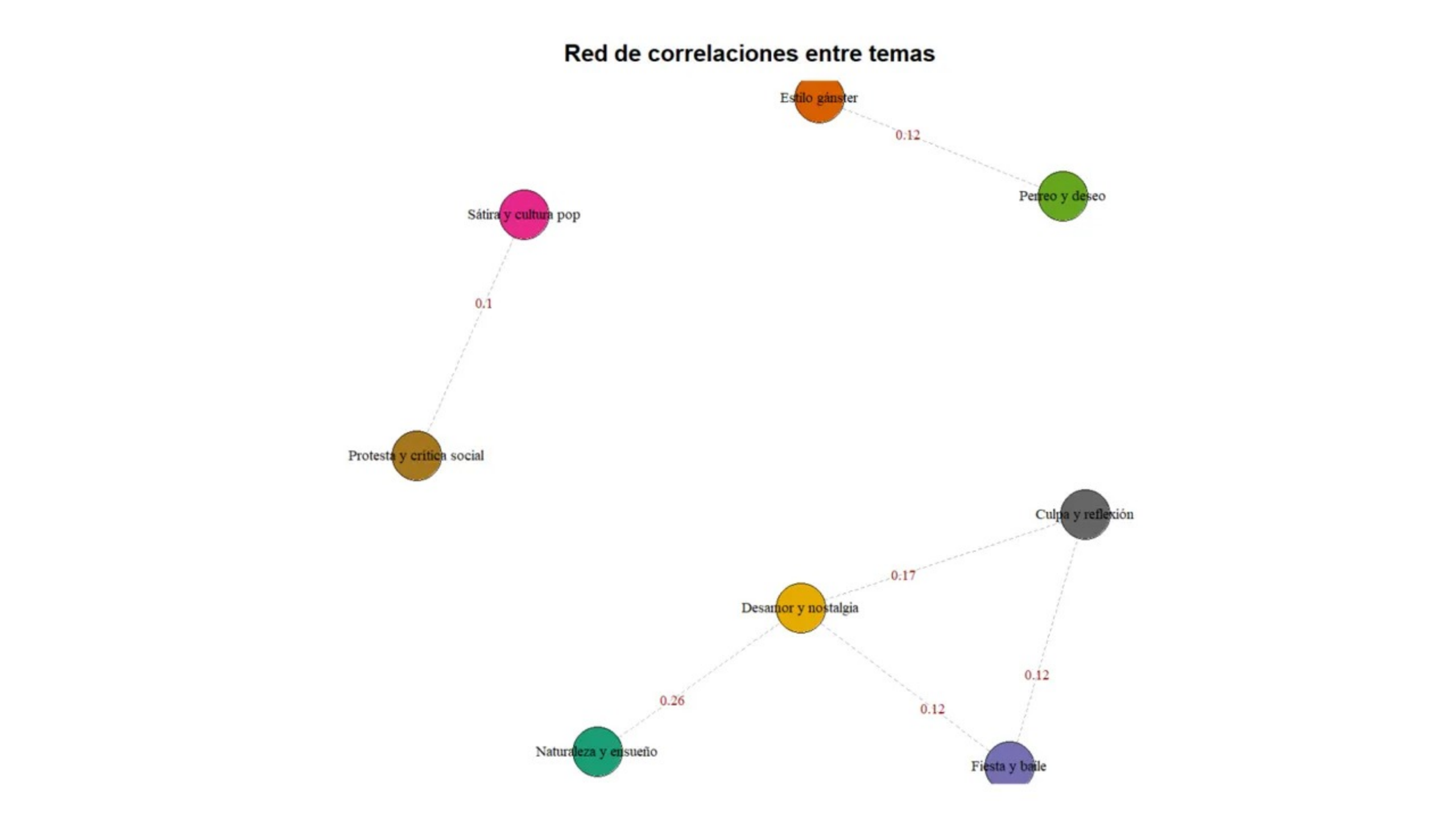
T7: Protesta y crítica social

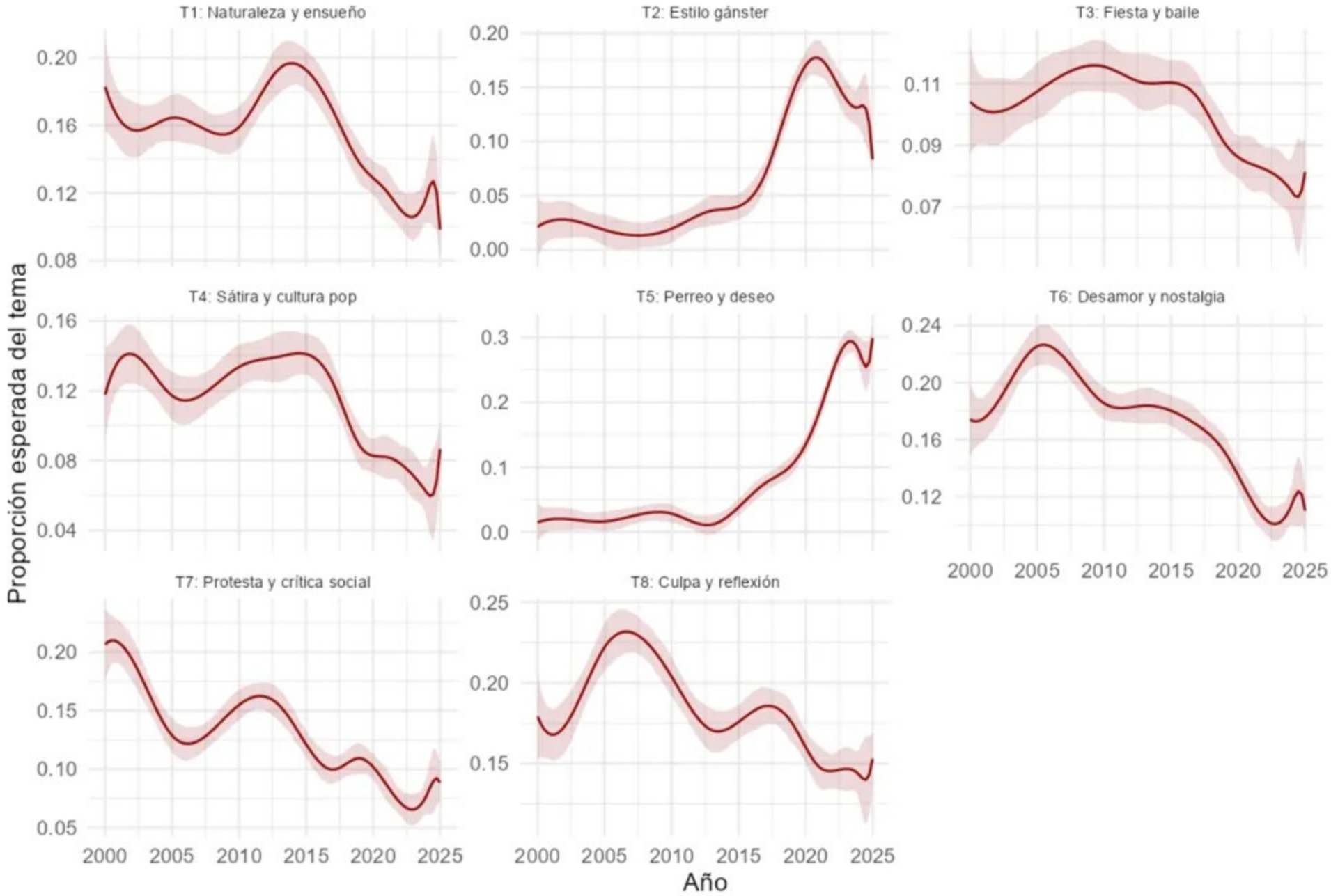


T8: Culpa y reflexión




STM (K = 8): trayectorias de prevalencia de los topicos
Prevalencia media (theta)
0.0
0.1
0.2
0.3
2000
2005
2010
2015
2020
2025
Año
T1: Naturaleza y ensueño
T4: Sátira y cultura pop
T6: Desamor y nostalgia
T8: Culpa y reflexión
T2: Estilo gánster
T5: Perreo y deseo
T7: Protesta y crítica social


Red de correlaciones entre temas
Estilo gánster
0.12
Perreo y deseo
Sátira y cultura pop
0.1
Protesta y crítica social
Culpa y reflexión
0.17
Desamor y nostalgia
0.12
0.26
0.12
Naturaleza y ensueño
Fiesta y baile


STM (K = 8): prevalencia de temas según el modelo a lo largo del tiempo
prevalencia ~ s(año, df = 10), con banda de confianza del 95%
T1: Naturaleza y ensueño
T2: Estilo gánster
T3: Fiesta y baile
T4: Sátira y cultura pop
T5: Perreo y deseo
T6: Desamor y nostalgia
T7: Protesta y crítica social
T8: Culpa y reflexión
Proporción esperada del tema
Año

Implicancias y limiaciones

Limitaciones

- Objeto: el campo indexado (musicapopular.cl a Genius), no toda la producción chilena. 363 de 2.177 artistas llegan al corpus final.
- Instrumento: STM registra léxico; pierde negación, ironía, voz narrativa y novedad formal. Un tópico es un síntoma, no un indicador.
- Decisiones analíticas: preprocesamiento, K = 8 y nombres de tópicos se fijaron con los modelos candidatos.
- Inestabilidad de las estimaciones por la cantidad de tokens en el corpus
- Potenciales problemas de replicabilidad

Implicancias

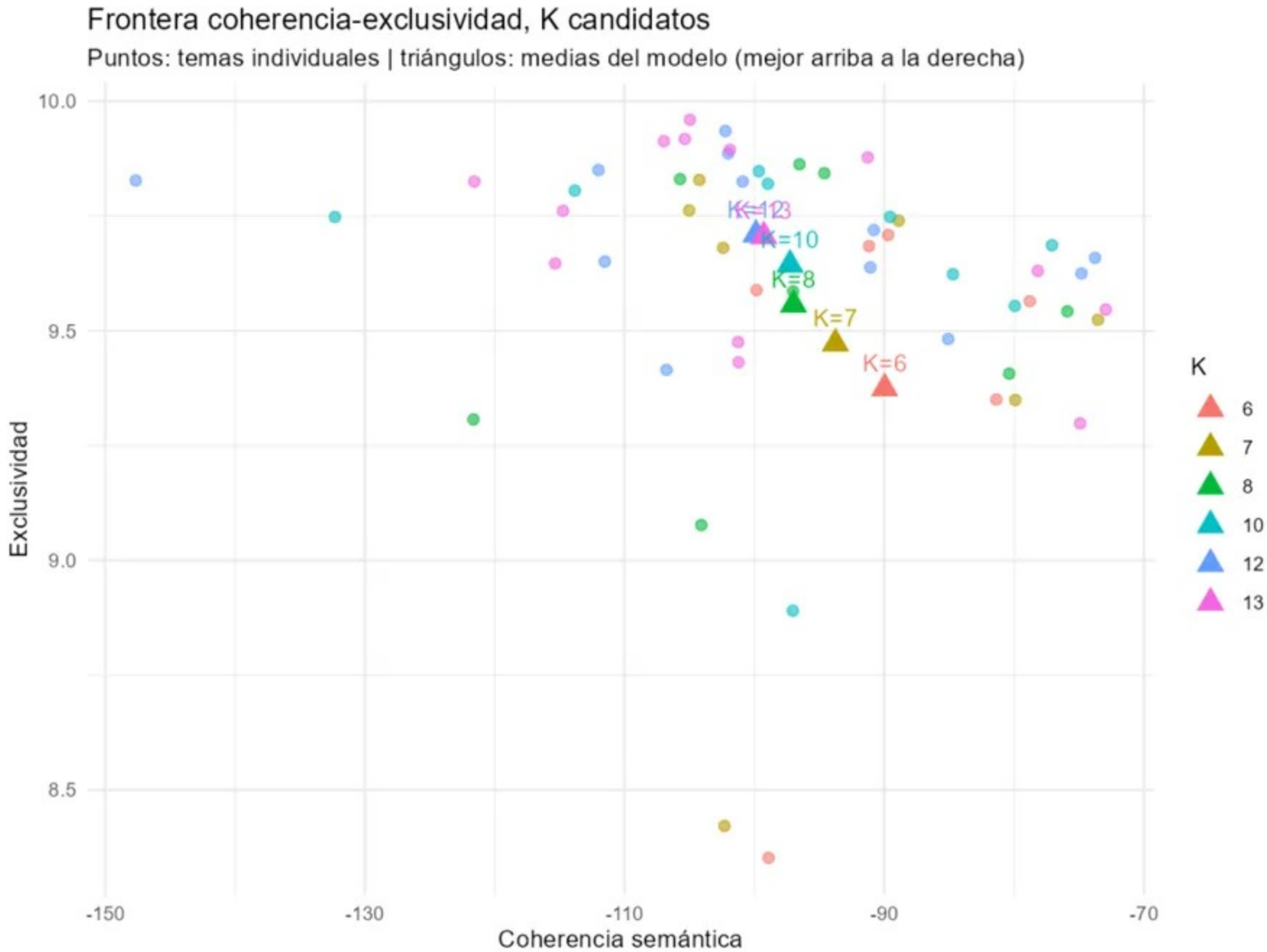
- Establecimiento de una descripción base de la estructura temática y su variación por período.
- Punto de partida para una etapa de medición e inferencia, con medidas y rivales definidos antes de mirar los datos y evidencia de más de un registro.
- A nuestro saber, primer análisis corpus-level con STM de letras chilenas entre géneros.

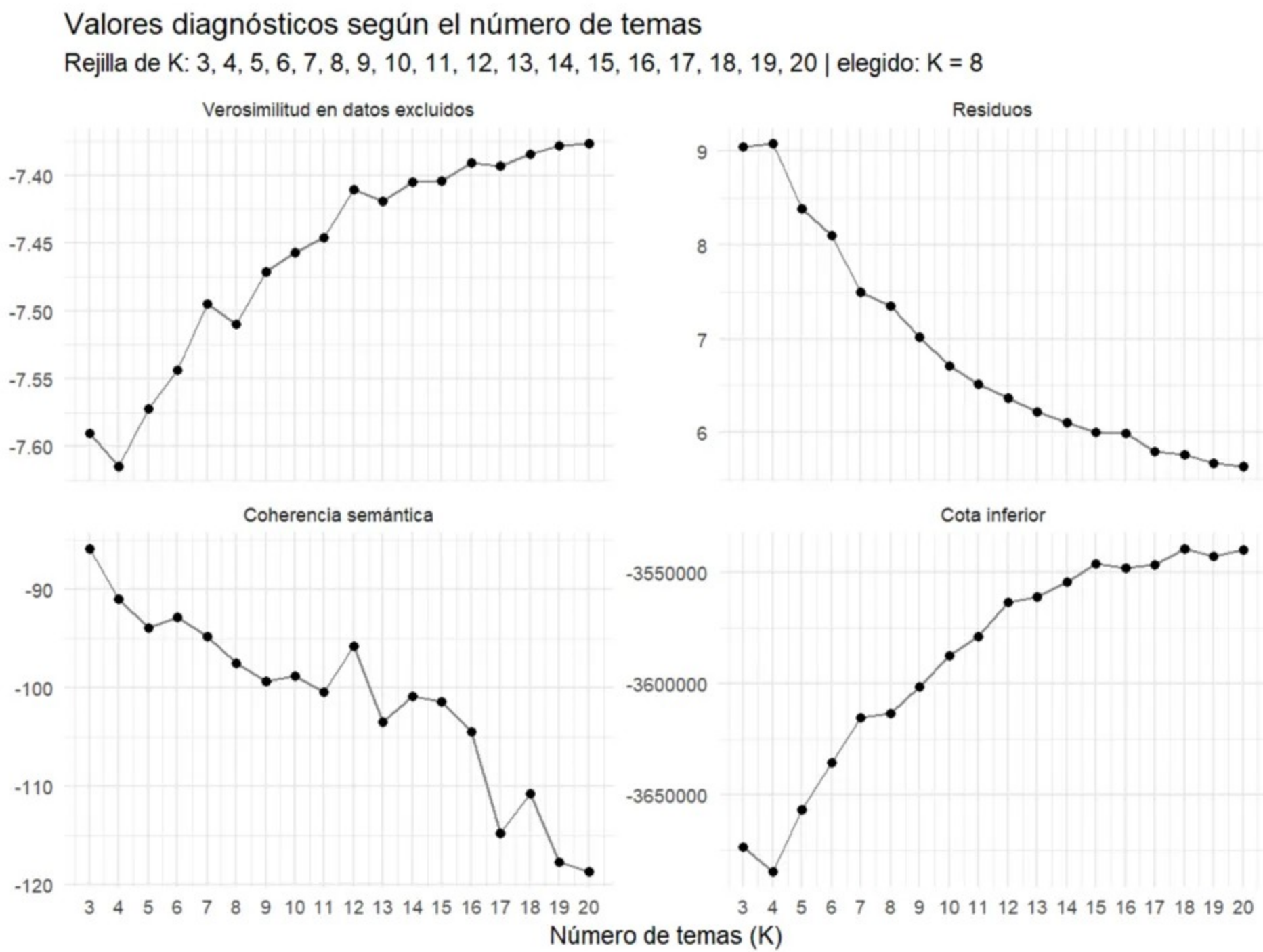
Referencias

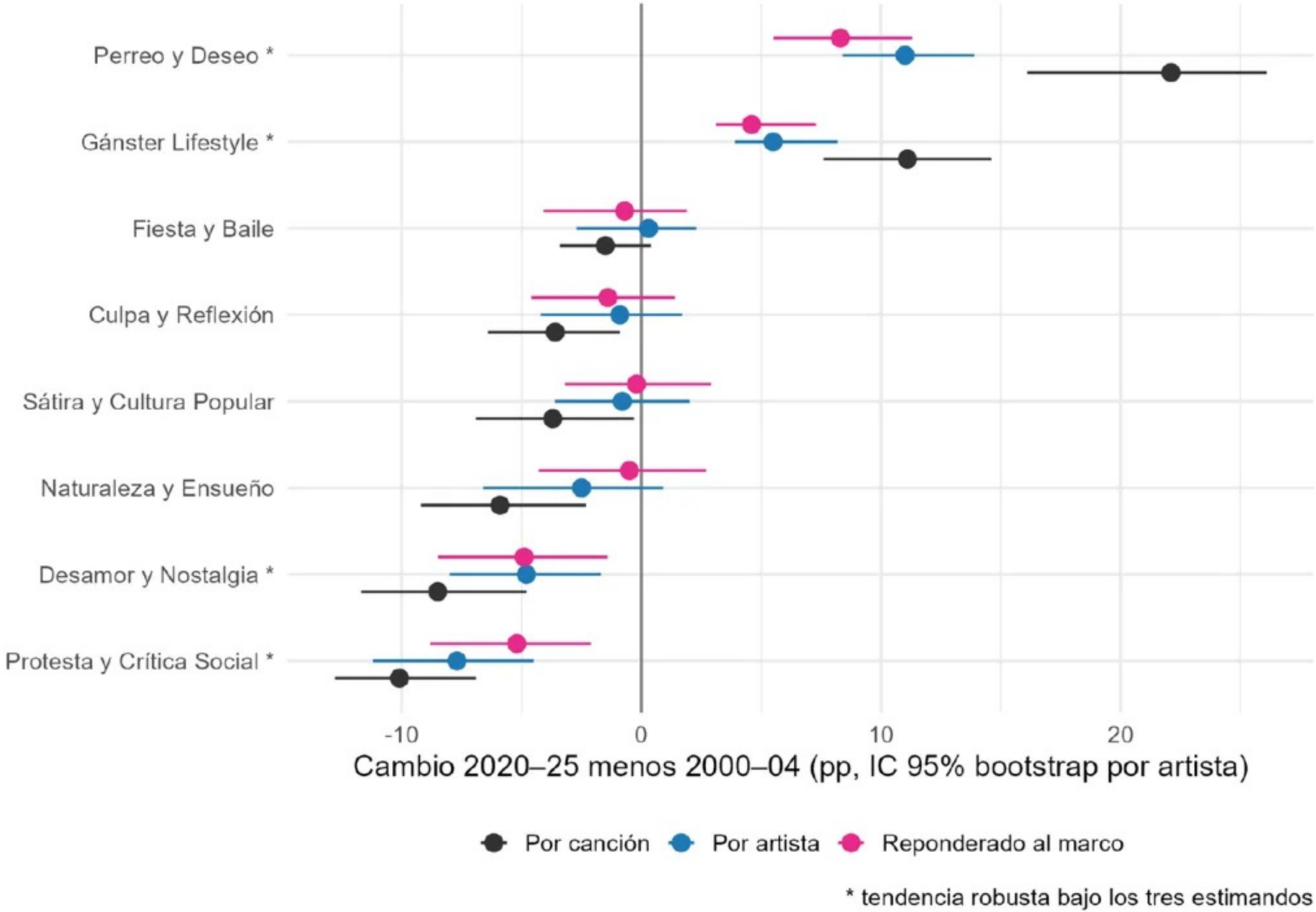
- Allimant, I. H. (2024). El trabajo de la memoria en la música popular chilena: espectros, horizontes perdidos, estéticas y sonoridades retro. Journal of Iberian and Latin American Research, 30(1), 4–16. https://doi.org/10.1080/13260219.2024.2393168
- Anglada-Tort, M., Krause, A. E., & North, A. C. (2021). Popular music lyrics and musicians' gender over time: A computational approach. Psychology of Music, 49(3), 426–444. https://doi.org/10.1177/0305735619871602
- Chen, D., Satish, A., Khanbayov, R., Schuster, C., & Groh, G. (2025). Tuning into bias: A computational study of gender bias in song lyrics. In Proceedings of the 9th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2025) (pp. 117–129). https://doi.org/10.18653/v1/2025.latechclfl-1.12
- Christenson, P. G., de Haan-Rietdijk, S., Roberts, D. F., & ter Bogt, T. F. M. (2019). What has America been singing about? Trends in themes in the U.S. top-40 songs: 1960–2010. Psychology of Music, 47(2), 194–212. https://doi.org/10.1177/0305735617748205
- Figueroa-Bustos, A. A. (2021). Antes del estallido: comunicación de lo político en letras de canciones chilenas indie (2005–2018). Comunicación y Medios, 30(44). https://doi.org/10.5354/0719-1529.2021.60801
- Fisher, M. (2014). Ghosts of my life: Writings on depression, hauntology and lost futures. Zero Books.
- Hunke, T., Huber, F., & Steffens, J. (2025). The evolution of song lyrics: An NLP-based analysis of popular music in Germany from 1954 to 2022. Music & Science, 8, 20592043251331155. https://doi.org/10.1177/20592043251331155
- Pettijohn, T. F., & Sacco, D. F. (2009). The language of lyrics: An analysis of popular Billboard songs across conditions of social and economic threat. Journal of Language and Social Psychology, 28(3), 297–311. https://doi.org/10.1177/0261927x09335259
- Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. Journal of the American Statistical Association, 111(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684
- Varnum, M. E. W., Krems, J. A., Morris, C., Wormley, A., & Grossmann, I. (2021). Why are song lyrics becoming simpler? A time series analysis of lyrical complexity in six decades of American popular music. PLOS ONE, 16(1), e0244576. https://doi.org/10.1371/journal.pone.0244576
- Vásquez-Mejías, A., Riffo-Pavón, I., & Zarzuri, R. (2025). No todo lo que reluce es narco: análisis temático de la música urbana en Chile. Resonancias Revista de investigación musical. https://doi.org/10.7764/res.2025.57.8

Muchas gracias!

Si te interesan estas temáticas o quieres
colaborar no dudes en contactarnos


Frontera coherencia-exclusividad, K candidatos
Puntos: temas individuales | triángulos: medias del modelo (mejor arriba a la derecha)
10.0
9.5
9.0
8.5
Exclusividad
-150
-130
-110
-90
-70
Coherencia semántica
K=12
K=13
K=10
K=8
K=7
K=6
K
6
7
8
10
12
13


Valores diagnósticos según el número de temas
Rejilla de K: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 | elegido: K = 8
Verosimilitud en datos excluidos
-7.40
-7.45
-7.50
-7.55
-7.60
Residuos
9
8
7
6
Coherencia semántica
-90
-100
-110
-120
Cota inferior
-3550000
-3600000
-3650000
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Número de temas (K)


Perreo y Deseo *
Gánster Lifestyle *
Fiesta y Baile
Culpa y Reflexión
Sátira y Cultura Popular
Naturaleza y Ensueño
Desamor y Nostalgia *
Protesta y Crítica Social *
-10
0
10
20
Cambio 2020–25 menos 2000–04 (pp, IC 95% bootstrap por artista)
Por canción
Por artista
Reponderado al marco
* tendencia robusta bajo los tres estimandos