

Dime qué **Emoji** usas y te diré qué **sientes**

Eduardo Graells-Garrido

Yerka Freire

Paula Vásquez-Henríquez

Data Science Institute UDD

Center for Social Complexity UDD

StarsConf 2018. Santiago, Chile.

rawpixel@unsplash





> Propuesta: Entender el comportamiento y la opinión de las personas en torno a temas relevantes.

> Caso 1: Transporte
> Caso 2: Ley de Aborto



Métodos Tradicionales

vs

Registro Digital

Métodos Tradicionales



Muestra limitada



Tiempo de análisis:
meses, años



Granularidad temporal
gruesa



Alto costo

Registros Digitales



Cantidad enorme de
personas



Tiempo de análisis:
días



Día a día



Bajo costo

¿Nuevo Paradigma?

Enfoque tradicional:

- > Una vez que se tiene una pregunta, se **recolectan datos** y se obtiene una respuesta

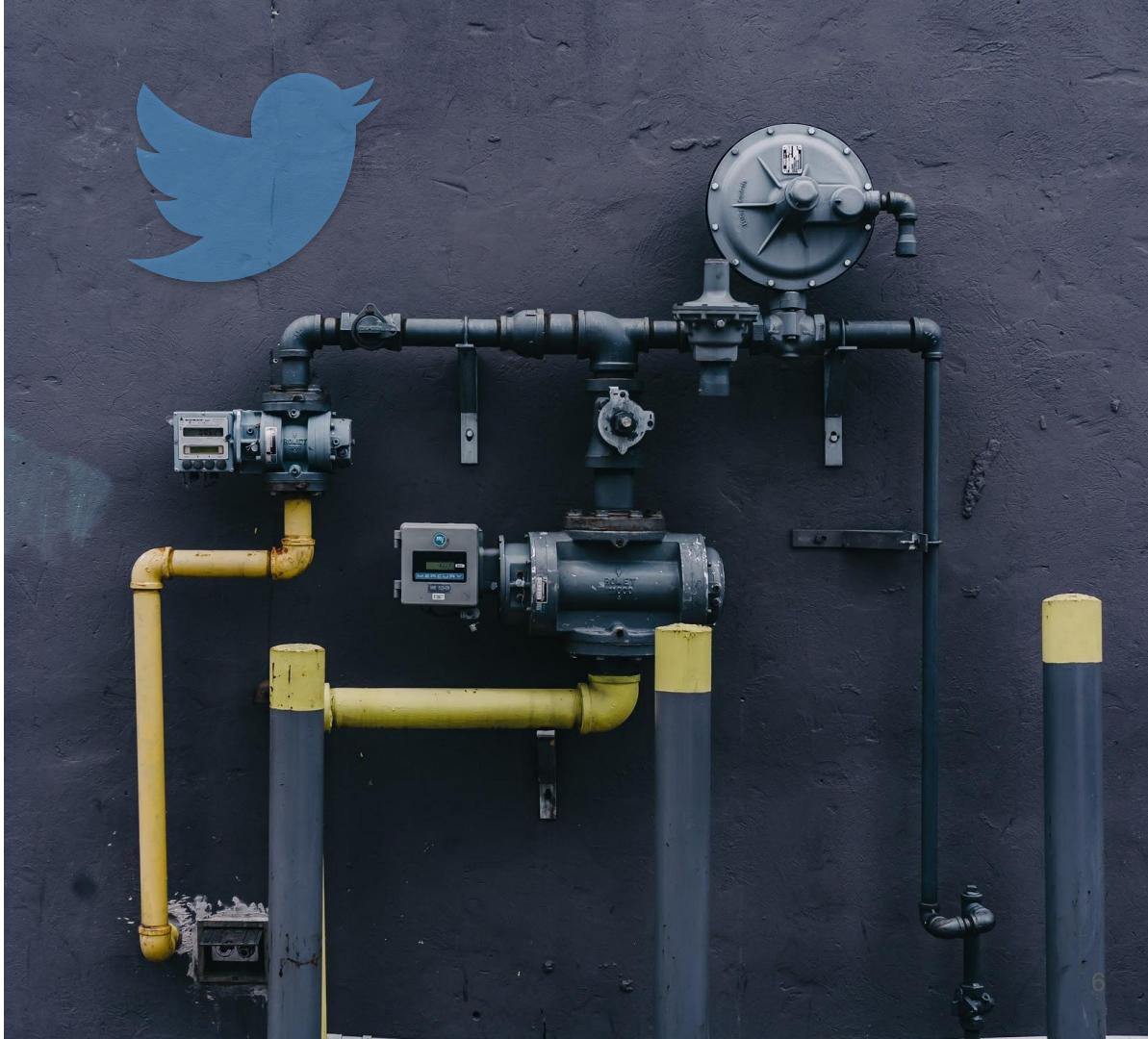
Personas **generan registros** día a día:

- > Opiniones y atributos demográficos son **auto-reportados** de manera pública.
- > ¡Los datos ya están disponibles, les podemos hacer preguntas!

The screenshot shows Javiera Mena's Twitter profile page. At the top, there is a large photo of her performing on stage with red lights, a smaller circular profile picture, and some basic stats: Tweets 13,3 mil, Siguiendo 804, Seguidores 170 mil, Me gusta 5.676, and Listas 2. Below this, the profile section shows her name, handle (@javieramena), a bio mentioning her new album '#ESPEJO', her location (Chile), and her birth date (03 de junio). A button to 'Twittear a Javiera Mena' is present. The timeline section starts with a tweet from 'JavieraMenaFans' retweeting another user about a free mall event. This tweet is highlighted with an orange box. Below it is a GIF of Javiera Mena performing. Further down the timeline, another tweet from 'JavieraMenaFans' is shown, advertising an event at Estación Mapocho. The number '5' is visible in the bottom right corner of the timeline area.

¿Cómo medir los registros digitales?

scottwebb@unsplash

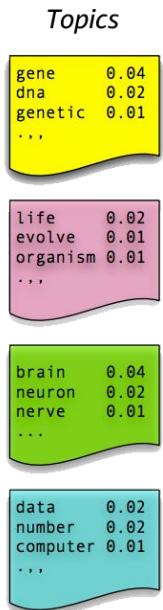


Aprendizaje Automático

La traducción de Machine Learning en castellano es **Aprendizaje Automático**.

> AA es una manera de **construir modelos desde los datos de manera automática**, con un foco en los patrones matemáticos.

> Usaremos un modelo **generativo** llamado **Latent Dirichlet Allocation (LDA)**, con enfoque en una versión **Semi-Supervisada** (guiada)



Documents

Topic proportions & assignments

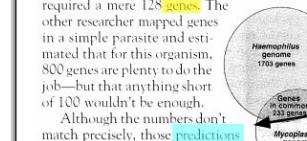
Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,¹⁰ two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions

"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden. He arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains

Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an



Stripping down, Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

MOCHI (WIP)

Caracterización de la Percepción del Transporte

Trabajo junto a
Diego Caro



 **rafa...** @rafael_olea · 5 sept.
Dos micro del **#transantiago** 211 pasaron por fuera del paradero n°2 en mall plaza sur y no entraron. Más encima está la gente del Transantiago con tótem. Cuál es la idea? No saqué foto, la denuncia anterior no me respondieron nada. 
@Transantiago sucedió recién a las 20:12pm

1 1 1 1

 **Sofi** 😊 @_foreverenana_ · 1 h
No les pasa que van **caminando** por la calle con audífonos y de pronto empieza a sonar una canción buenísima y ustedes se sienten como **caminando** en una pasarela. 😂😂

 **Graciela Bernal C.** @gracielabernal6 · 22h
Replying to @rhurtubia
Yo cumplí los 60 y voy a todos lados en **bici**, de hecho voy todos los años a Lo Vásquez y vuelvo sola y acompañada, e ido para el 18 de Sept a Algarrobo sola. Es una maravilla!
Entre más ciclistas en las vías más seguras son.

 **Francisco Fernandez** @tisco_fernandez · 5 sept.
Voy en la micro y el micrero va escuchando Janis Joplin. Si evaluar **#Transantiago** le pongo un 7

 **Nilse Silva** @Nilsesilvaperez · 14h
Nunca podré entender cómo un taxista es capaz de persona de su auto por el simple hecho de pagar cc segunda vez que me pasa. Con estas actitudes lo ú más a **@Uber_Chile** **@Chile_Taxi**

 **Felipe Drogueut** @FelipeDrogueut · Sep 26
1 hora en **micro** y aún no salgo de Quilicura 😕

Translate Tweet

 **Consuelo** @MConsuelofg · Sep 28
La refregá en el hombro en la **micro** 😔
Me pasó cuando era escolar.
#PP

 **Paula Vásquez-Henríquez** @pauvasquezh · 4 h
El chófer de la C02 de la que me acabo de bajar era lo más amable de la vida ❤️
CC: @Transantiago

2 2 2

 **Ariel López** @ariellopez · Sep 26
Cuando tienes que ir de A-B distantes en 1 km y sólo tienes 5 min
-caminar: 15 min ✗
-correr: 5 min + transpiración ✗
-mi bicicleta: no la traje ✗
-micro: no hay servicios ✗
-taxi: congestión + riesgo = 15 ✗
-uber: congestión + espera = 18 ✗
-bicicleta pública: 5 min ✓

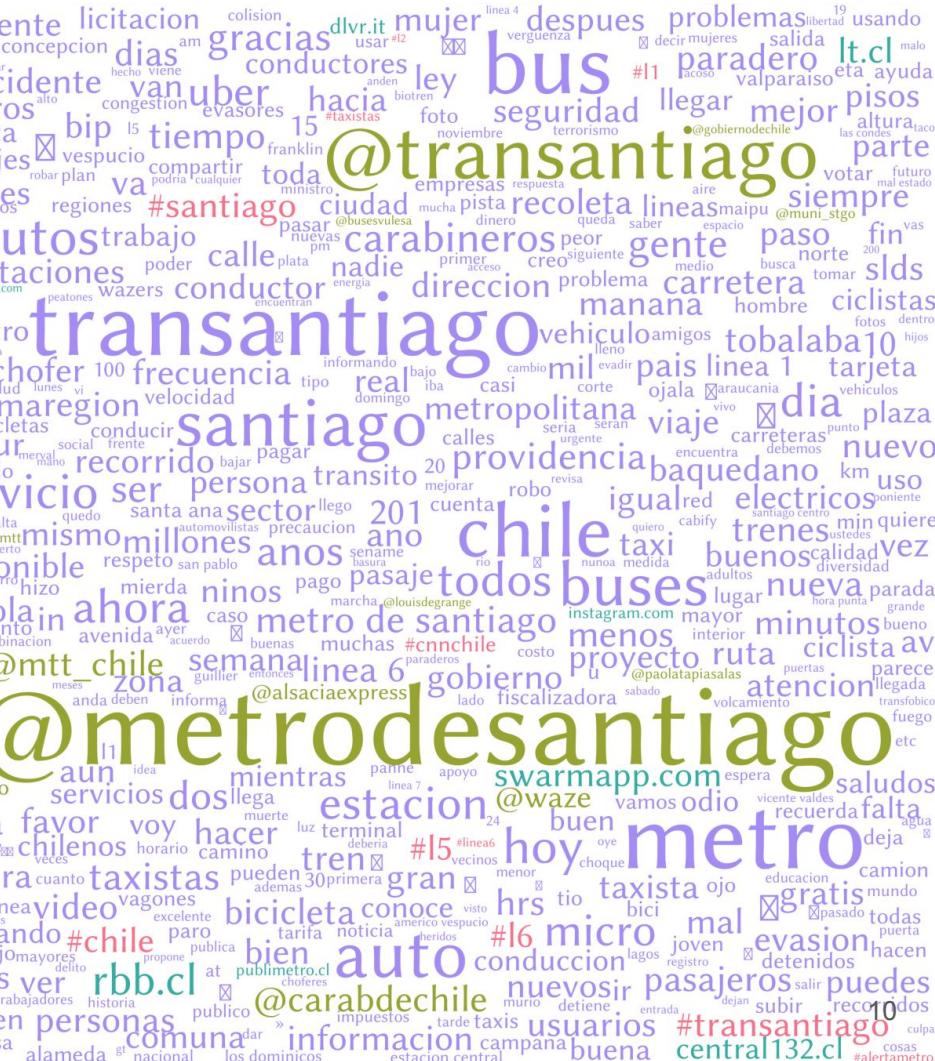
> ¿Qué caracteriza la percepción de cada modo de transporte?

Dataset:

Año 2017

525K Tweets (9.9K semanales)

80K Usuaries



Percepción: Emociones y Discurso

Emociones (EmoLex)

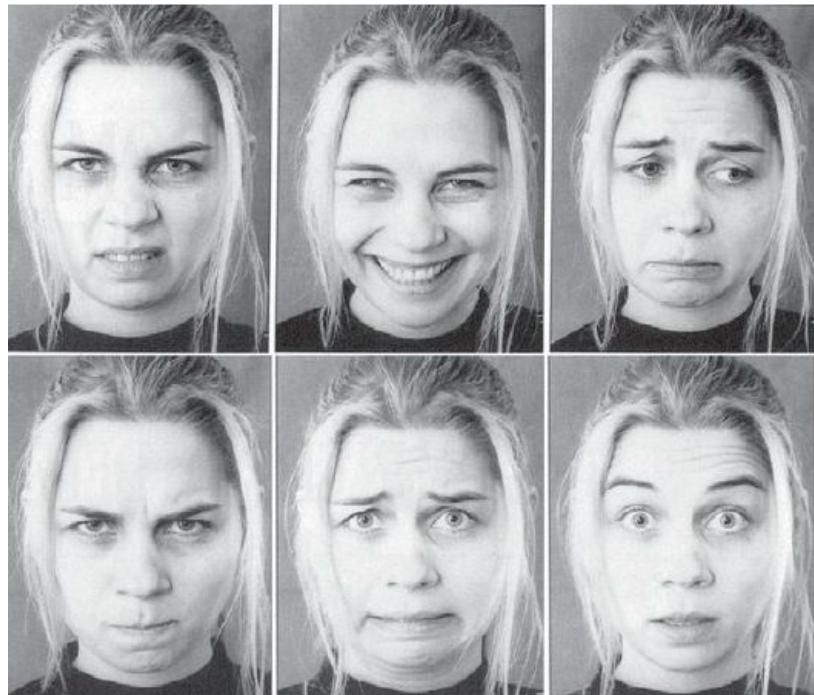
- > Basado en la teoría de **Paul Ekman**, psicólogo:
rabia, disgusto, tristeza, sorpresa, miedo,
alegría

Temáticas (LIWC)

- > Hecho por psico-lingüistas (**Pennebaker**, 1996)
- > sexualidad, cuerpo
- > futuro, pasado
- > familia, sociabilidad
- > trabajo, recreación

Emojis (Emoji Sentiment Ranking)

- > Creado por **Kralj Novak** et al. organiza los emoji por su **polaridad**



Casey C. Bennett and Selma Šabanović

Detección de Tópicos

Cada **modo de transporte** es un **tópico de LDA**:

> Guiado por palabras asociadas a cada uno, especificadas manualmente

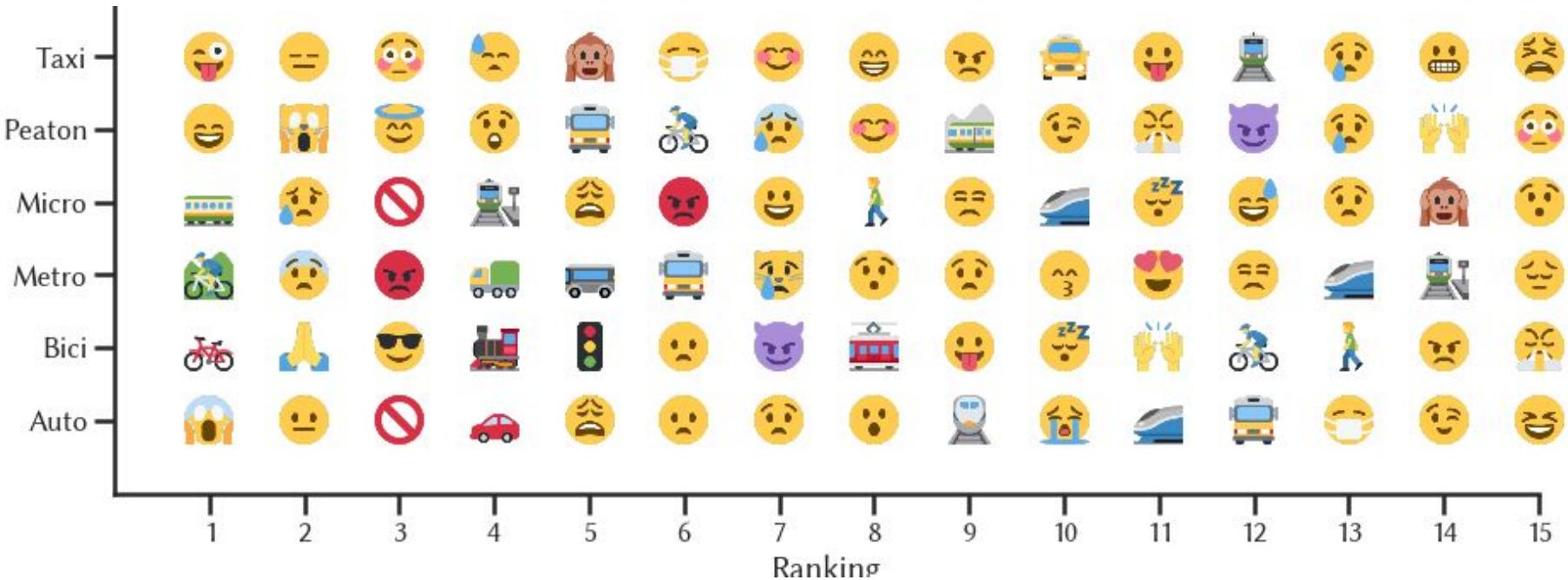
Cada **emoción** es un tópico:

> Guiado por **EmoLex**

Tópico	Palabras Guía	Palabras Relevantes
Metro	#l1, anden, baquedano, metro, andén, vagón, 	conociste, inaugure
Micro	recorrido, torniquete, @metbus, alsacia, @transantiago, 	por favor, exigencias, 
Taxi	taxis, uber, cabify, @uber_chile, 	amarillo
...		
Emoción: Joy	celebración, entretenido, logro, felicidad, tranquilo, 	 , inmediato, amplia
Emoción: Anger	abandono, caótico, ilegal, obligar, decepcionado,  , 	por favor, dentro
Emoción: Disgust	abuso, deshonesto, intolerancia, borracho, expulsar, 	dentro, amarillo
...		

¿Twitter es sólo para quejarse?

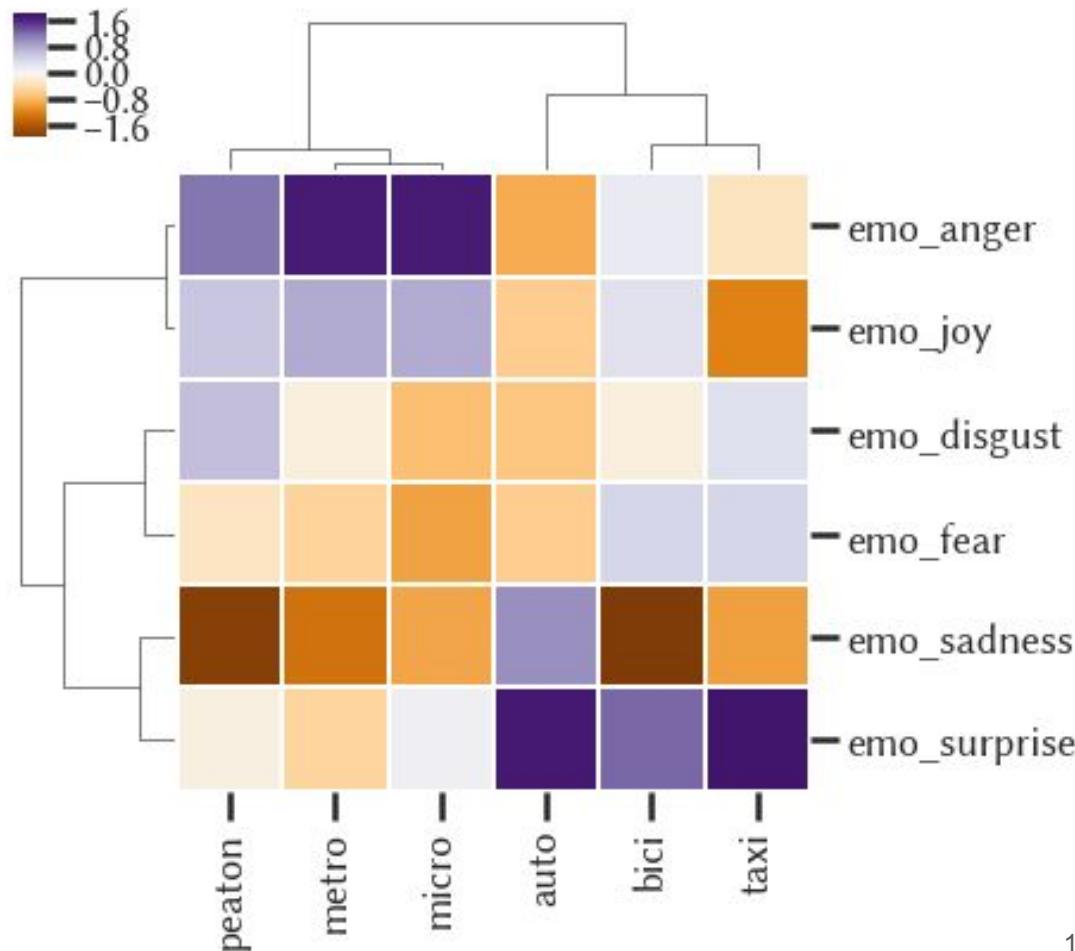
15 emojis más relacionados a cada medio de transporte



Emociones y medios de transporte

Relación entre los tópicos asociados a EmoLex y Modos de Transporte.

> Cada celda de la matriz es la importancia relativa de cada emoción dentro del modo de transporte, en función del lenguaje utilizado por sus usuarios.



> El transporte
no es neutro al
género

- Banco Mundial





FORMAS DE ACOSO CALLEJERO #12 PUNTEOS (PRESIÓN O ROCE DE GENITALES SOBRE EL CUERPO)

“MUCHAS VECES LOS HOMBRES SE PARAN AL LADO
MIENTRAS VOY SENTADA EN LA MICRO Y ROZAN SUS
GENITALES EN MI HOMBRO. AUNQUE ME CORRA,
ELLOS LO SIGUEN INTENTANDO. UN ASCO TOTAL”.

El 33,3% de las chilenas ha sufrido “punteos”
o roces genitales en el espacio público.

*1era Encuesta de Acoso Callejero en Chile

#NoMásAcoso

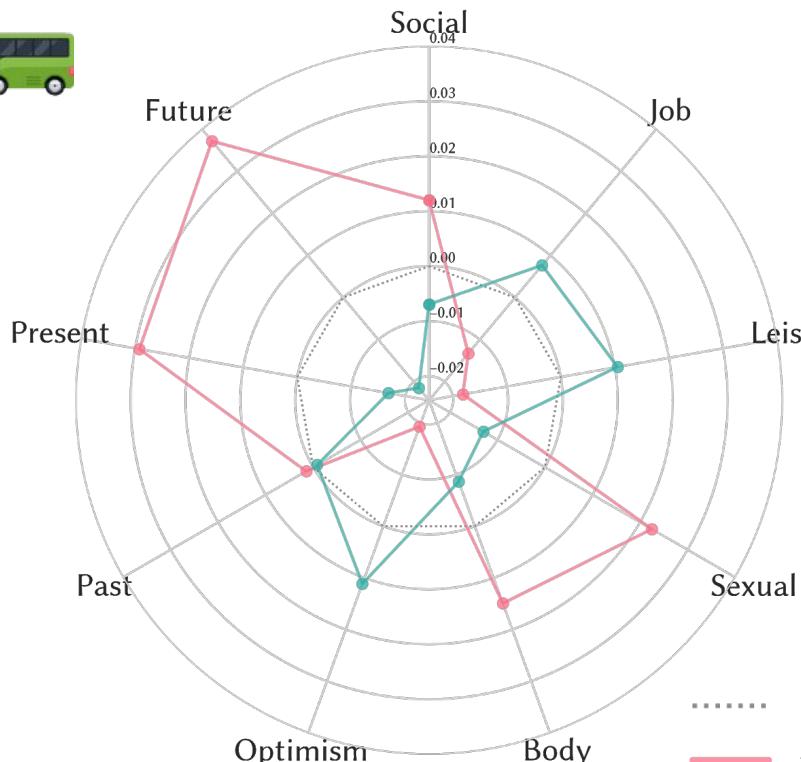


@ocacchile

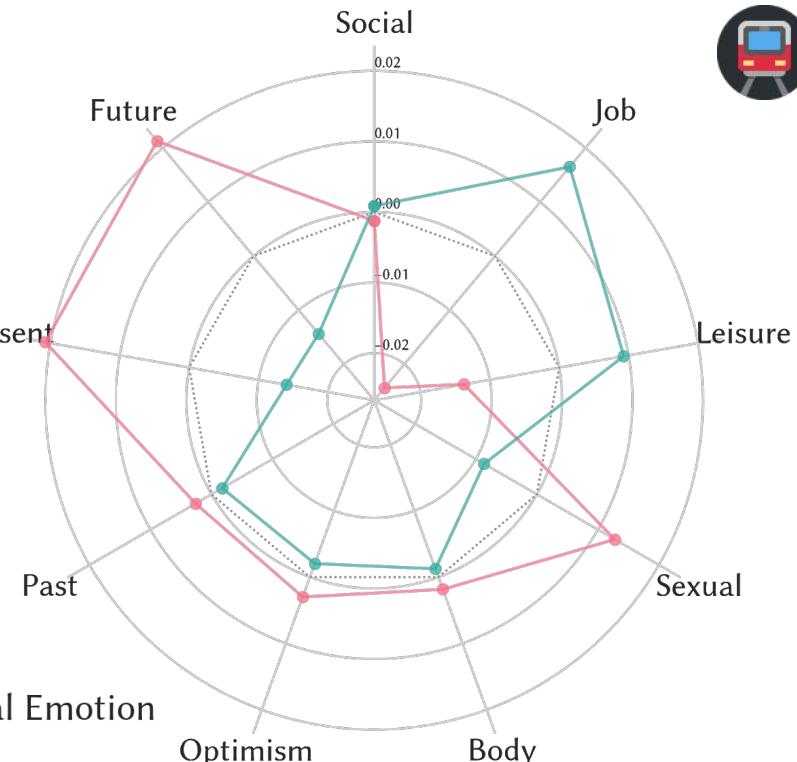


/ocacchile

TEMÁTICAS DESAGREGADAS POR GÉNERO



..... Neutral Emotion
— female
— male



> Esto no es solo investigación:
¡obtuvimos mención honrosa en una hackathon de transporte! :)

Junto a: Ignacio Pérez-Messina y Valentina Prado, Team MOCHI en #HackAmericas.



Dinámicas de la Opinión Pública: **Ley de Aborto** en Chile

Trabajo junto a **Ricardo Baeza-Yates** y
Mounia Lalmas

rakelilla79@twitter



> ¿Cómo reaccionaron las personas en Chile, en las plataformas virtuales, a la legislación de la **Ley de Aborto** entre 2015 y 2017?

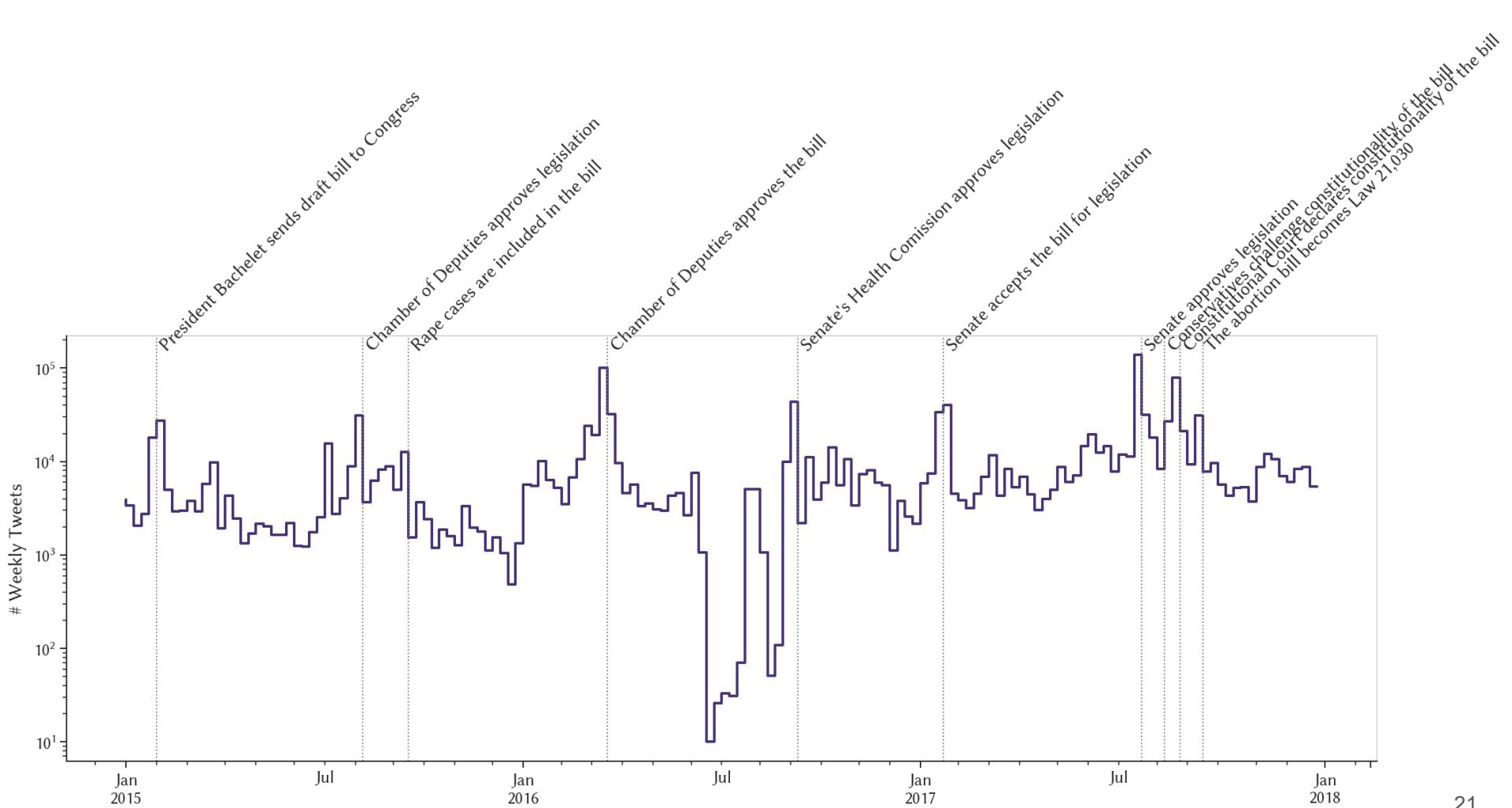
Dataset:
100K usuarie
1M tweets (**9.2K** semanales)



Marcha en defensa de los derechos a abortar en marzo de 2015.
Source: MovilH (CC licensed).



Manifestación pública en oposición a los derechos a abortar en marzo de 2016.
Source: Radio BíoBío.



Posiciones

Tópicos en LDA:

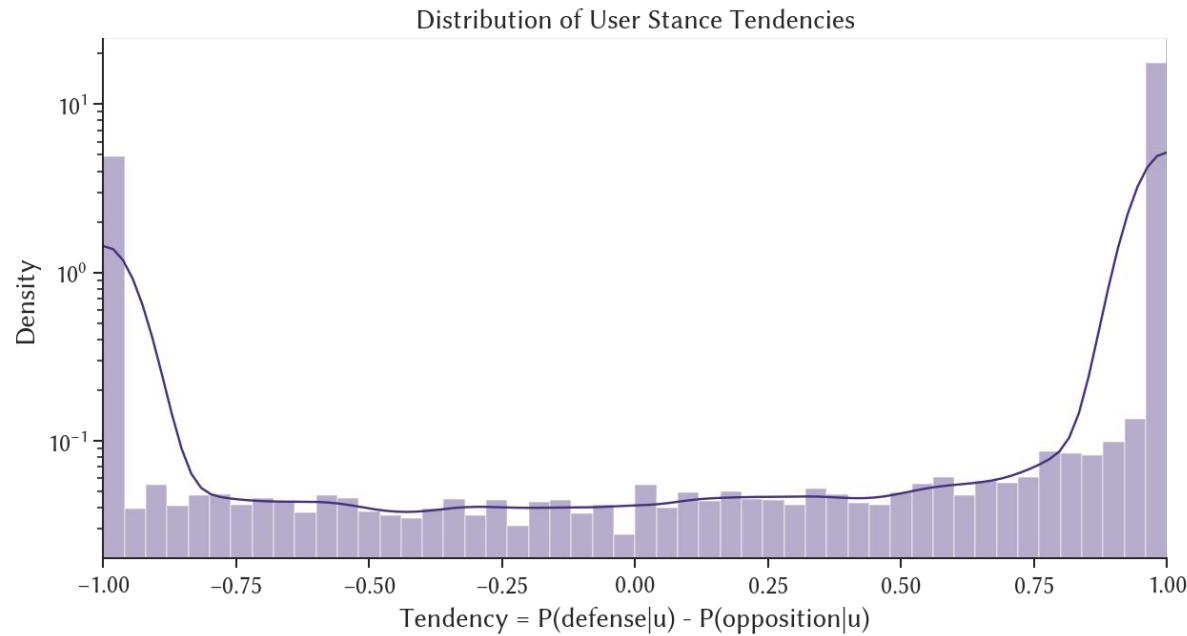
- > En defensa del derecho a abortar
- > En oposición al derecho a abortar

Resultado (coherente con CEP):

- > 73% de perfiles en **defensa**
- > 27% de perfiles en **oposición**

Resultado desagregado:

- > Probabilidades de estar en **oposición a los derechos de abortar es 0.06 mayor en hombres.**



oposición:

#noalabortion
#siempreporlavida
#funeraldechile
#objectionconcienciaesderecho
@quieronacer
#salvemoslas2vidas

...

defensa:

#abortolegal
#abortolibre
@miles_chile
#dejaladecidir
#apoyolas3causales
#obligadasaparir

p(URL | In Defense group)



p(URL | In Opposition group)



Diferencias entre posturas en función de los resultados del modelo **LDA Semi-Supervisado**.

Defensa
Oposición



Ranking de Asociación

Dinámicas del Discurso

Asociadas al grupo en oposición:

- > siempre: causalidad, muerte, trabajo
- > parciales: inhibición, negación, cuerpo, familia, religión

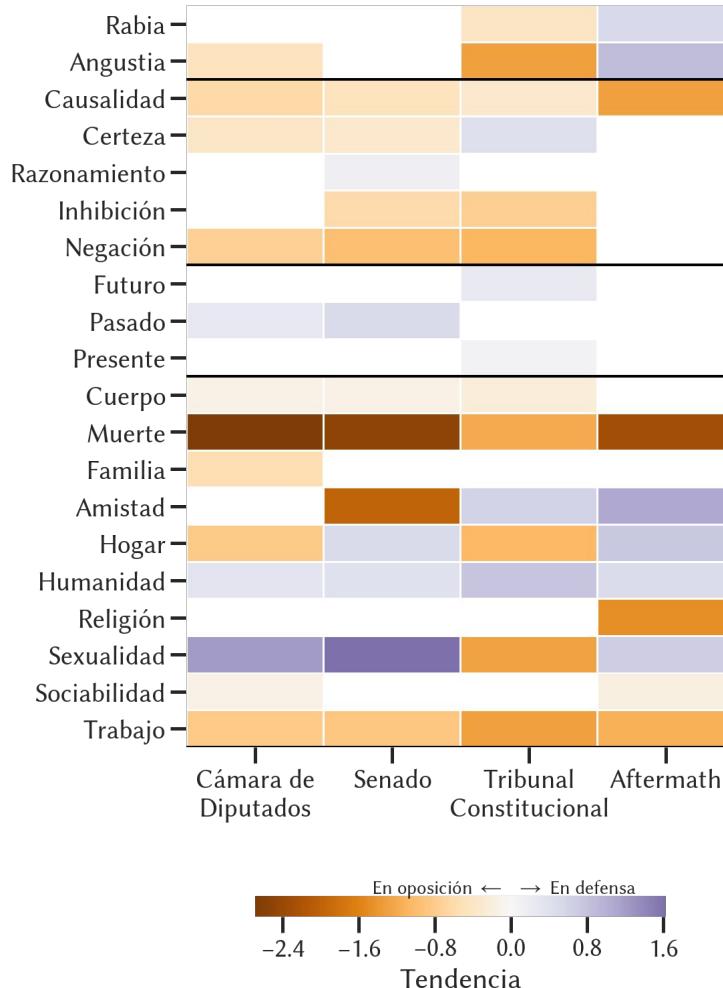
Asociadas al grupo en defensa:

- > siempre: humanos
- > parciales: pasado, futuro y presente

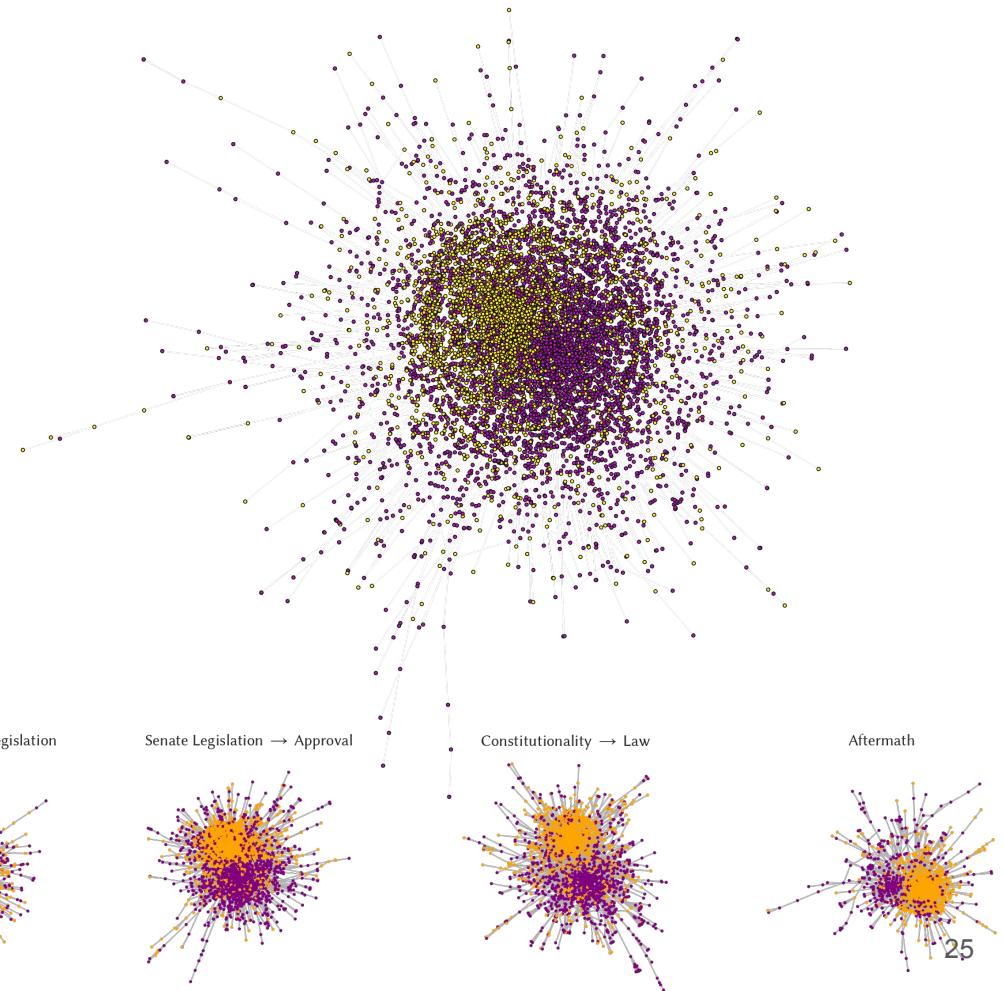
Vaivenes:

- > sexualidad: primariamente defensa
- > cambia en recurso constitucional.

- > rabia y angustia: primariamente oposición
- > cambia con la aprobación de la ley.



> Entender **cómo se sienten las personas**, qué estructura su discurso, y cómo se relacionan, permite comunicarse mejor con ellas.



¿Qué llevarse a casa?

“Take Aways”



Aunque datos no sean representativos del total de la población **sí podemos aprender de comportamiento humano :)**

> **Debemos ser cuidadosas/os con el análisis e incluir los sesgos en él.**

La dificultad no está en la implementación sino en **hacer las preguntas adecuadas.**

> ¡Esas preguntas vienen de **otras disciplinas!**

> **GuidedLDA** implementa LDA Semi-Supervisado.

¿Qué pasa con la **ironía y el sarcasmo?**

> Es un área de investigación activa, sin embargo, analizar **Perfiles en vez de Tweets individuales** permite disminuir su influencia.

¡Gracias! ¿Preguntas?

¿Quieres **colaborar** con nosotros?

¡Escríbenos! :)

egraells@udd.cl

Esta presentación está disponible en:

<https://bit.ly/DimeQueEmojiUsas>



En Twitter:
@pauvasquezh / @carnby / @yerka_freire
(foto por @diegocaro)