



**E**l 18 de octubre de 2019 (en adelante 18O) pasó a la historia de Chile. Motivos sobran para entender un evento que sigue sucediendo, un octubre que todavía no termina, una historia que se está escribiendo aún.

El 18O tiene una característica que otras revoluciones sociales no han tenido. Atrás quedó ese dicho "la historia la escriben los vencedores." Dejando de lado que aquí no hay vencedores, ya que todos hemos perdido y podemos afirmar desde ya: parte de la historia se está registrando en miles de dispositivos conectados a la Internet, utilizando la Web. Se está difundiendo lo que sucede y lo que sucederá a otros millones de dispositivos conectados, a través de redes sociales y de sistemas de mensajería privada. Cada hora se publican miles de opiniones, fotos y videos testimoniales. Algunos de estos posts se viralizan y marcan la pauta de la discusión.

Desde la investigación, esta fuente de datos presenta potencial porque nos permite entender cómo se siente la gente, y más aún, como expresa ese sentir. Qué es lo que opina, en lo individual, y qué es lo que importa --en términos de generar reacciones-- en lo global. Un factor relevante que podría explicar parte del 18O es que quienes concentran el poder económico y político no han sido capaces de entender el sentir de las personas. Pues bien, las redes sociales son una fuente de dicho sentir.

No es un secreto que las redes sociales tienen sesgos, así que no representan a todos. Son más utilizadas por gente joven, de áreas urbanas, con acceso a tecnología de gama media o alta, y con la educación tecnológica suficiente para expresarse en medios multimedia. Pero esto no implica que no sea una fuente rica de información, siempre y cuando se tengan estos sesgos en cuenta.

Twitter es una de las redes más usadas para entender la percepción de las personas, puesto que cualquiera puede escuchar lo que se dice en público. Se puede construir un robot (un programa) que esté pendiente de lo que se dice que incluya ciertas palabras, y luego de un tiempo, se dispone de un conjunto de tweets publicados que cumplan los criterios de búsqueda.

Parte de mi tesis doctoral utilizó este esquema: tener un robot escuchando lo que se conversaba en Twitter sobre Chile, en períodos intermitentes en los años 2012 y 2013, y continuamente desde el 2015 hasta hoy. El robot registra tweets que, en su texto, incluyen Chile, Santiago, Concepción, Valparaíso, y otros nombres de ciudades; Bachelet,

**O**ctober 18, 2019 (henceforth 18O) went down in Chilean history. Many reasons can help us understand an event that is still happening, an October that is not yet over, a history that is still being written.

*The 18O has a characteristic that other social revolutions have not had, and it has a story that is still being written. Gone is that saying, "History is written by the victors", leaving aside the fact that there are no winners here, because we've all lost. But we can say right now that history is being recorded on thousands of devices connected to the Internet, using the Web. They are spreading what is happening and what will happen to millions of other connected devices, through social networks and private messaging systems. Thousands of opinions, photos, and video testimonials are posted every hour. Some of these posts become viral and set the tone for discussion.*

*From the research, this data source is very rich because it facilitates the understanding of how individuals within society feel about the issue, and most importantly, how they express that emotion. They tell what is their opinion individually, and what is important, in terms of generating reactions – globally. Specifically, a relevant factor that explains part of the 18O is that those who concentrate the economic and political power have not been able to understand people's feelings. Thus, the social networks are a source of such sentiment.*

*It is no secret that social networks have biases; so, they don't represent everyone. They are more used by younger people from urban areas who have more access to mid- or high-end technology, and with sufficient technological education to express themselves in multimedia. Unlike the instruments that typically measure opinion in the country, these are not fully representative of the general population. However, this does not mean that their multimedia use is not a rich source of information, as long as these biases are taken into account.*

*Twitter is one of the most used networks for understanding people's perception, since anyone can listen to what is being said in public. You can build a robot (a program) that is aware of what is being said and that includes certain words; and, after a while, you have a set of published tweets that meet the search criteria.*

*Part of my doctoral thesis used this scheme: to have a robot listen to what was being discussed on Twitter about Chile, in intermittent periods in 2012 and 2013, as well as continuously from 2015 until today. The robot registers tweets that, in their text,*

Piñera, Pinochet, Alessandri, y otros apellidos de políticos/as; lucro, aborto, educación, inmigración, Fonasa, isapres, AFP, pensiones, carabineros, y otras palabras asociadas a temas coyunturales. Con estos datos, en la UDD hemos utilizado Twitter para estudiar la percepción del transporte en Santiago (tesis de Paula Vásquez-Henríquez, Magíster en Ingeniería [[@vasquez2019characterizing](#)]) y las actitudes ante la inmigración (tesis de Yerka Freire, Doctorado en Ciencias de la Complejidad Social [[@freire2019characterization](#)]).

En colaboración internacional, junto a Ricardo Baeza-Yates y Mounia Lalmas (Spotify) hemos publicado estudios sobre como la centralización en Chile afecta la difusión de información [[@graells2014balancing](#)] y qué se puede hacer para mejorar esa situación [[@graells2016encouraging](#)]; sobre la discusión entre personas que opinan distinto, y como podemos acercarlas utilizando algoritmos [[@graells2016data](#)]; y también sobre la representatividad de la discusión de aborto en Chile [[@graells2019representative](#)].

En Chile hay más hombres que mujeres participando en la red social, incluso en un debate asociado a género. Si bien la población Twittera no tiene la misma distribución en el país (por eso tiene sesgos), si se utilizan las técnicas y modelos adecuados puede entregar resultados similares a los que tienen encuestas de opinión pública como la CEP.

Entonces, ¿qué se puede aprender de Chile y el 18O a través de Twitter? En este artículo vamos a explorar dos temas:

- Primero, ¿Quiénes participan de la discusión del 18O en los 90 días posteriores?
- Segundo, ¿Qué caracteriza al debate en términos emocionales?

*include Chile, Santiago, Concepción, Valparaíso, and other city names; Bachelet, Piñera, Pinochet, Alessandri, and other last names of politicians; profit, abortion, education, immigration, Fonasa, isapres, AFP, pensions, police and other words associated with current affairs. With these data, at the UDD we have used the data from Twitter to study the perception of transportation in Santiago (Vásquez-Henríquez, Paula. Thesis, Master in Engineering) and attitudes towards immigration (Freire, Yerka, PhD thesis, from the Doctorate in Sciences of Social Complexity ([@freire2019characterization](#))).*

*In international collaboration with Ricardo Baeza-Yates and Mounia Lalmas (Spotify), we have published studies on how centralization in Chile affects the dissemination of information ([@graells2014balancing](#)). The research shows what can be done to improve this situation; ([@graells2016encouraging](#)); about the discussion between people who have different opinions and how we can bring them together using algorithms ([@graells2016data](#)). We have also shown the representation of the abortion discussion in Chile ([@graells2016data](#)).*

*In Chile there are more men than women participating in the social networks, including in a gender-related debate. Although the Twitter population does not have an equal distribution in the country (that is why it is biased), if the appropriate techniques and models are used, it can deliver results similar to those of public opinion surveys such as the CEP.*

*So, what can be learned from Chile and the 18O through Twitter? In this article, we will explore two issues.*

- *First, who participates in the discussion of the 18O in the 90 days following it?*
- *Second, what characterizes the debate in emotional terms?*

# One might say that when the country is in crisis, people turn to Twitter

## PARTICIPACIÓN EN LA DISCUSIÓN DEL 18O

El conjunto de tweets que analizamos comprende 11,9 millones de tweets publicados por 666 mil personas. Éste es el data set después de ser pre-procesado para eliminar "ruido y suciedad". Por ejemplo, si uno busca "Chile", también encontrará tweets provenientes de México que hablan de ají (en ese país es un ingrediente fundamental); o cuando hay trending topics, mucha gente (y robots) publican tweets con publicidad que incluyen múltiples temas populares, como una manera de inyectar contenido en la discusión. Entonces es necesario contar con mecanismos que identifiquen y descarten este tipo de contenido. No es un proceso automático, sino que requiere mucho trabajo de exploración y validación manual.

Después de la limpieza de la data set, nuestro esquema de análisis predice el sexo de una persona (hombre o mujer, lamentablemente es una clasificación binaria), su ubicación a nivel país (así podemos quedarnos con la conversación de gente en Chile solamente) y un rango de edad. Para ello utiliza el clasificador State-of-the-Art XGBoost [[@chen2016xgboost](#)] (para detalles, ver referencia [[@graells2020representativeness](#)]).

En resumen, se utiliza la información que cada perfil disponibiliza públicamente (nombres, ubicaciones), más los retweets y menciones del perfil (una cuenta tiende a relacionarse con cuentas del mismo lugar). Por ejemplo, si alguien hace Re tweet a un medio de Punta Arenas, es probable que ese perfil sea de la Región de Magallanes. Un aspecto relevante de clasificar es que un programa siempre da una respuesta. Podemos entregarle un perfil de fantasía y aun así trataría de predecir sus atributos. Algunos clasificadores incluyen una medida de confianza en la predicción, y en nuestro caso, usamos esta medida para decidir si quedarnos con la predicción o no.

## PARTICIPATION IN THE 18O DISCUSSION

*The set of tweets analyzed comprises 11.9 million tweets published by 666,000 people, which is the data set after being pre-processed to eliminate "noise and dirt". For example, if you search for "Chile," you will also find tweets from Mexico that speak of chili pepper (in that country it is a basic ingredient); or when there are trending topics, many people (and robots) post tweets with advertising that includes multiple popular topics, as a way of injecting content into the discussion. So, it is necessary to have mechanisms that identify and discard this type of content. It is not an automatic process, but it requires a lot of explorative and manual validation work.*

*After cleaning the data set, our analysis scheme predicts a person's sex (male or female; unfortunately, this is a binary classification), its location at the country level (so we can focus on people in Chile only), and an age range. The analysis uses the State-of-the-Art XGBoost classifier [2] (for details, see reference [6]).*

*In summary, we use the information that each profile makes publicly available (names, locations) plus the retweets and mentions of the profile (an account tends to relate to accounts in the same place). For example, if someone retweets to someone in Punta Arenas, it is likely that this profile is from the Magallanes Region. An important aspect of classifying is that a program always provides an answer. We can give it a fantasy profile and still try to predict its attributes. Some classifiers include a measure of confidence in the prediction, and in our case, we use the measure to decide whether or not to keep the prediction.*

*Once location (mandatory), sex, and age for profiles have been predicted, 660,000 profiles remained. Only*

## TASA DE PARTICIPACIÓN EN TWITTER (ESTALLIDO SOCIAL) POR EDAD Y SEXO

Número de cuentas bajo análisis: 667k, con tweets entre el 18 octubre 2019 y 18 enero 2020

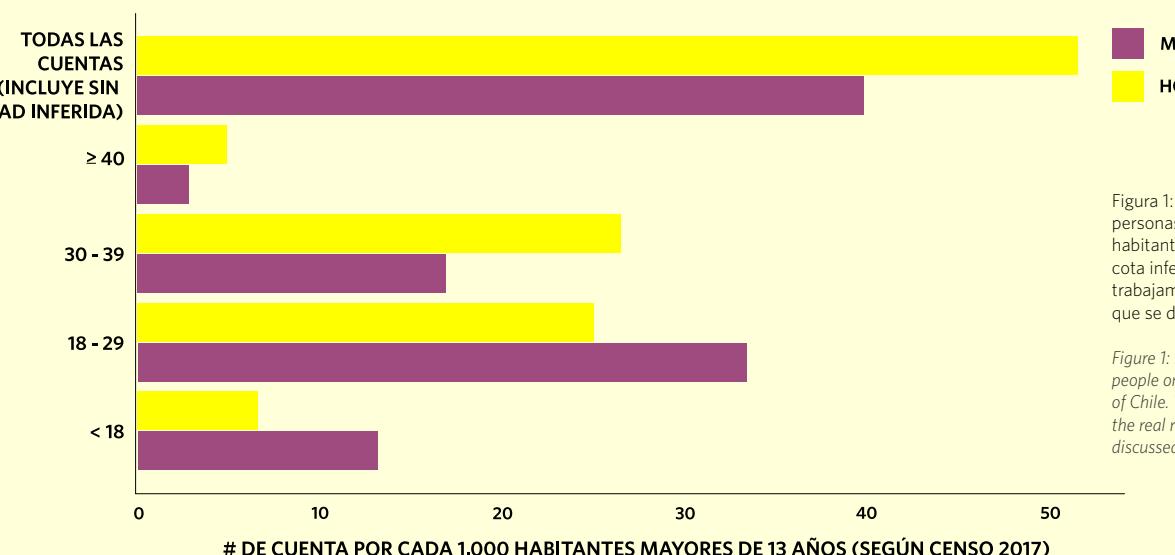


Figura 1: Tasas de representatividad de personas en Twitter por cada 1,000 habitantes de Chile. Estas tasas son una cota inferior de la tasa real, puesto que trabajamos con un subconjunto de lo que se discute en la red social.

Figure 1: Representativeness rates of people on Twitter per 1,000 inhabitants of Chile. These rates are a lower level of the real rate, given it is a subset of what is discussed in the social network.

## PORCENTAJE DE LAS CUENTAS POR FECHA DE CREACIÓN (ANUAL)

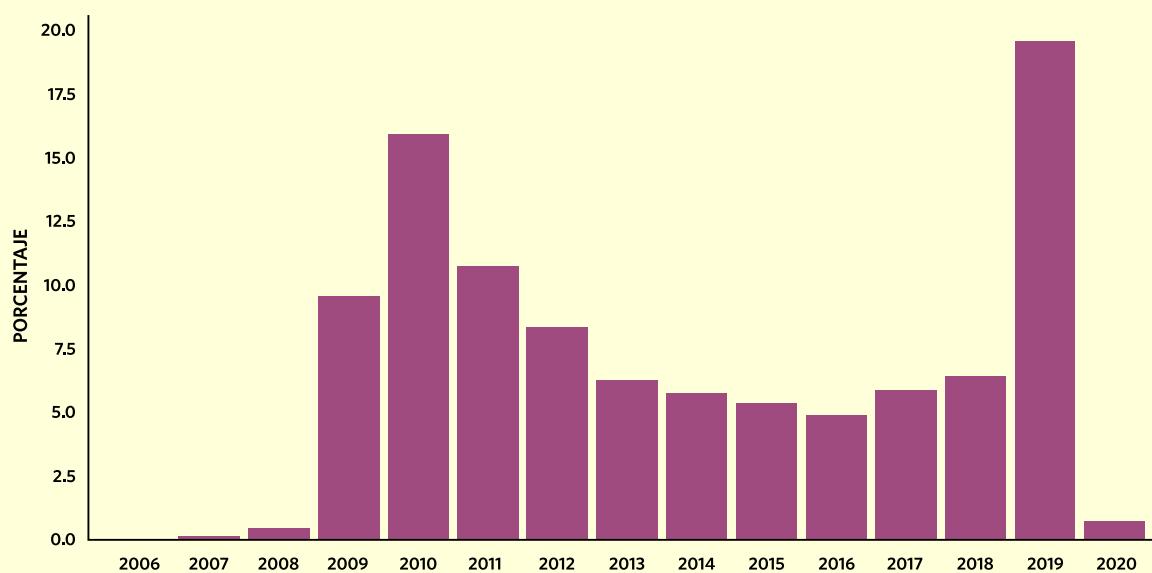


Figura 2: Porcentaje de cuentas que participan en el data set, distribuidas por año de creación.

*Figure 2: Percentage of accounts participating in the data set, distributed by year of creation.*

Una vez que se ha predecido ubicación (obligatoria), sexo y edad para los perfiles, quedan 666 mil perfiles. Se usan solamente los perfiles que se puede determinar con cierta seguridad que pertenecen a gente que reside en Chile, porque la participación en Chile es lo que interesa. Los otros atributos demográficos los consideramos opciones.

En la Figura 1, cada grupo etáreo (por sexo) muestra la cantidad de personas twitteras que hay por cada 1,000 habitantes en el país. Si Twitter fuese representativo de la población general, veríamos números similares en cada categoría. Además, la representatividad de hombres y mujeres no es igual. En general, en Twitter (Chile) hay más hombres que mujeres, y también hay más adultos jóvenes que gente de 40 años o más. Los menores de edad también participan, sin embargo, su presencia no es tan fuerte en comparación a las de otras redes sociales. Ahora bien, en los grupos menores de 30 años sí hay más mujeres que hombres. Estas diferencias generacionales serán interesantes de estudiar en el futuro.

Las tasas de representatividad no cuentan toda la historia. La Figura 2 muestra la proporción de las personas (en el data set) que se registraron cada año. Los dos años que han tenido más registros han sido el 2010 y el 2019. Aunque el gráfico no lo muestra, las dos semanas que han tenido más registros han sido las del 27 de febrero de 2010 (terremoto), y la del 18 de octubre de 2019. Cuando el país está en crisis, se acude a Twitter.

Una posible fuente de confusión es la aparición de robots que intentan influir en la discusión. Estos bots usualmente

profiles that can be determined with certainty as belonging to people residing in Chile are used because participation in Chile is what matters. We consider other demographic attributes optional.

In Figure 1, each age group (by sex) shows the number of tweeters per 1,000 inhabitants in the country. If Twitter were representative of the general population, we would see similar numbers in each category. However, that is not the case. In addition, the representation of men and women is not equal. In general, on Twitter (Chile) there are more men than women, and there are also more young adults than people aged 40 or older. Minors also participate, but their presence on Twitter is not as strong compared to that of other social networks. Note that there are more women in groups under 30 years of age than men. These generational differences will be interesting to study in the future.

Rates of representation do not tell the whole story. Figure 2 shows the proportion of people (in the data set) who were recorded each year. The two years of the most records were 2010 and 2019. Although the graph does not show it, the two weeks with the most records were those of February 27, 2010 (earthquake) and October 18, 2019. Thus, one might say that when the country is in crisis, people turn to Twitter.

A possible source of confusion is the appearance of robots trying to influence the discussion. These bots are usually automatic accounts that have a short duration, since Twitter removes them when they are reported and detected. Although 2019, the year with the largest number of accounts created

Figura 3: Nube de palabras con las hashtags más frecuentes en el data set.  
*Figure 3. Clouds of words in the data set with the more frequent hashtags.*

son cuentas automáticas que tienen poca duración, puesto que Twitter las elimina cuando son reportadas y detectadas. Aunque el año 2019, con la mayor cantidad de cuentas creadas en el data set, no representa ni un quinto del total de las cuentas, por lo que es difícil que exista una influencia de bots en la discusión. Presencia sí, y ciertamente es un tema que debe ser investigado.

Una de las polémicas que se han dado en el 18O ha sido una posible influencia extranjera en la discusión. El esquema utilizado para clasificar la ubicación de las personas (un proceso llamado geolocalización) es similar al de cualquier otra plataforma (técnicas en el estado del arte). Efectivamente, en este experimento se han encontrado cuentas extranjeras que participan, pero su volumen y su influencia no es relevante.

## CARACTERIZACIÓN EMOCIONAL

La Figura 3 muestra los hashtags más frecuentes durante este largo octubre (entre el 18O y el 18 de enero de 2020). Los hashtags de las primeras semanas dominan la nube de palabras, no solamente porque fue el período en el que se registró más gente, sino que también fue el período en que la discusión estaba más centralizada en unos pocos temas.

*our data set does not even represent a fifth of total number of accounts, it is difficult for bots equally to influence the discussion. Presence does, however, and it is certainly an issue that should be investigated.*

One of the controversies that occurred in the discussion. The scheme used to classify the location of people (a process called geolocation), similar to that of any other platform (state-of-the-art technique). Indeed, foreign accounts have been found involved in this experiment, but their volume and influence are not relevant.

## OPTIONAL CHARACTERIZATION

Figure 3 shows the most frequent hashtags during a long October (between 180 and January 18, 2020). Hashtags of the first few weeks dominate word cloud, not only because it was the period when more people posted, but was also because it was the period when the discussion was most used on a few topics.

Later, the diversification of the discussion led specific aspects that were not captured by the robot. For example, the robot does hear "Mixed Convention" or "Constituent Convention", or "Vote to prove", or "Reject to Reform", among others that are more frequent today. So, the robot hears

Más adelante, la diversificación de la discusión derivó en aspectos puntuales que no fueron capturados por el robot. Por ejemplo, el robot no escucha "Convención Mixta" o "Convención Constituyente", ni "Vota Apruebo", ni "Rechazar para Reformar", entre otros que son más frecuentes hoy. Entonces, esos temas son escuchados de manera más superficial. Esto no es una limitante, sino que es una decisión de diseño experimental, puesto que en el futuro se quiere poder comparar distintos períodos de la discusión en Chile, y para ello se necesita que los data sets sean comparables en aspectos como el proceso de captura de datos.

Para analizar los casi 12 millones de tweets del periodo en relación al 18 octubre se utilizó la psicolingüística, una rama de estudio que abarca el uso del lenguaje y su relación con las emociones [@tausczik2010psychological]. En resumen, existen conjuntos de palabras que las personas usan cuando se sienten emociones. La agrupación de los conjuntos se llaman lexicón y se pueden utilizar para analizar texto, incluyendo lo que publican las personas en Twitter.

**¿Cómo funcionan los lexicones?** El mecanismo es directo: se cuenta la cantidad de veces que son utilizadas las palabras de cada categoría en la unidad de análisis. Por ejemplo, para analizar personas en específico, por cada persona se construye una tabla que contiene todas las categorías bajo análisis, y la cantidad de palabras en cada una de esas categorías que utilizó cada persona. Pero ese es solo el primer paso, dado que existen múltiples sesgos en la generación de estas tablas. Por ejemplo, algunas personas publican mucho más que otras [@baeza2018bias], y, por tanto, una persona comunicativa, alegre y prolífica en Twitter puede tener más palabras tristes que una persona introvertida y triste que solamente lee las noticias y responde de vez en cuando a alguien. Si analizamos a nivel de ubicaciones o de días, los sesgos de representatividad hacen que algunos grupos tengan mayor prevalencia que otros, respecto a su representación en el mundo físico. Entonces, el siguiente paso es definir una manera de medir las fluctuaciones o asociaciones en las palabras de cada categoría a través de la unidad de análisis. Por ejemplo, es común que en una comunidad siempre haya mayor frecuencia de palabras de una categoría específica. Lo que importa no es esa magnitud, sino la variación relativa entre una y otra. Una manera de medir esa variabilidad es a través de la estandarización, diciendo que en cada unidad de análisis se utiliza una categoría más que en el promedio de esa misma categoría. Otra manera es en medir la variabilidad y frecuencia de las categorías, y utilizar esas mediciones para determinar si una unidad de análisis específica tiene mayor asociación a las palabras considerando también la variabilidad y frecuencia de las otras categorías. Este enfoque se conoce como Log-Odds Ratio with Uninformative Dirichlet Prior [@monroe2008fightin] as was utilized in this analysis.

Esta sección analiza la emoción día a día usando LIWC con toda la población que participó en la discusión. Para cada día

*those topics in a superficial way. This is not a limitation, but an experimental design decision, since we want to compare different periods of the discussion in Chile in the future. For this, the data sets need to be comparable in aspects such as the process of data capture process.*

*Psycholinguistics, a branch of study that covers the use of language and its relationship with emotions, was used to analyze the nearly 12 million tweets in the period in relation to October 18th [@tausczik2010psychological]. In short, there are sets of words that people use when they feel emotions. These sets are called lexicons and can be used to analyse text, including what people post on Twitter.*

**How do lexicons work?** The mechanism is straightforward: it counts the number of times words in each category in the unit of analysis are used. For example, to analyze specific people, for each person a table is constructed, containing all the categories under analysis and the number of words in each of those categories that each person used. However, that is only the first step, since there are multiple biases in the generation of these tables. For example, some people publish much more than others [@baeza2018bias], and, therefore, a happy, communicative and prolific person on Twitter may have more sad words than an introvert and sad person who only reads the news and responds occasionally. If we analyze the level of locations or days, representative biases make some groups more prevalent than others, with respect to their representation in the physical world. The next step is to define a way to measure fluctuations or associations in the words of each category through the unit of analysis. For example, it is common for a community always to have a greater frequency of words from one category over another. One way to measure this variability is through standardization, saying that each unit of analysis uses one category more than the average for that same category. Another way is by measuring the variability and frequency of the categories and using those measurements to determine whether a specific unit of analysis has a greater association with the words by also considering the variability and frequency of the other categories... This approach is known as Log-Odds Ratio with Uninformative Dirichlet Prior [@monroe2008fightin] as was utilized in this analysis.

*This section analyses the day-to-day emotion using Linguistic Inquiry and Word Count LIWC with the entire population that participated in the discussion. For each day between October 18th and January 18th, we counted the number of words that are available in each semantic category and then converted that count into proportions (e.g., 80% of the categorizable words belong to a category). Then we estimated the association of each day through the Log-Odds Ratio technique. The study focused on three emotions: emotional anxiety (anx in LIWC), anger, and positive feelings (posfeel). By considering each day as a unit of analysis, one can see how much each category is associated with each day.*

### TRES MESES DE EMOCIONES DESDE EL 18/10

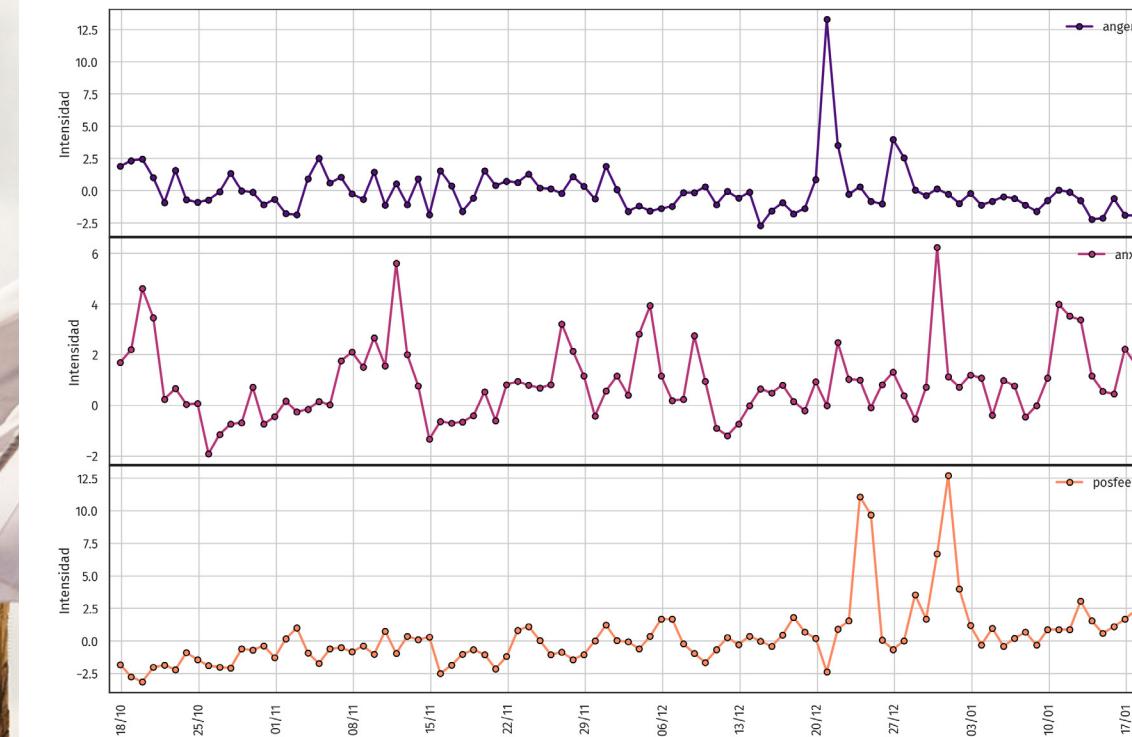


Figura 4: Un "emocograma" de tres emociones entre el 18 de octubre y el 18 de enero.  
Figure 4: An "electrocardiogram" of three emotions between October 18th and January 18th.



# Data science has a challenge to turn this body of information into knowledge

se cuenta la cantidad de palabras disponibles en cada categoría semántica, y después convertimos ese conteo en proporciones (por ejemplo, 80% de las palabras categorizables pertenece a una categoría x). Luego se estima la asociación de cada día con cada categoría a través de la técnica de Log-Odds Ratio.

El estudio se enfocó en tres emociones: ansiedad (anx en LIWC), enojo (anger) y sentimientos positivos (posfeel). Al considerar cada día como unidad de análisis, se puede ver qué tan asociada está cada categoría a cada día. Valores mayores a 0 implican una asociación intensa, una presencia fuerte de esa categoría. Valores cercanos a 0 indican neutralidad. Valores negativos implican ausencia de la categoría. La magnitud está medida en puntos porcentuales.

Con los resultados podemos diseñar un "emociograma" (Figura 4) donde es directo identificar los peaks emocionales y sus respectivas fechas. En el enojo o rabia, hay una fecha que destaca sobre las demás: el 21 de diciembre. Puede haber múltiples explicaciones para ese peak, sin embargo, es factible pensar que se debe al atropello (en la tarde anterior) de un manifestante por un carro lanza gases ("zorillo") de Carabineros.

La ansiedad fue el sentimiento dominante los primeros días de la crisis, y niveles similares de ansiedad aparecen en otras fechas, en un ciclo entre tres y cuatro semanas. Este vaivén de ansiedad muestra una señal preocupante, relacionada con el efecto en la salud mental de no poder salir de una crisis social.

En sentimientos positivos solamente hay dos peaks: navidad y año nuevo. Dos fechas especiales en las que, a pesar de la crisis, se podía compartir con familiares y amigos. Fuera de esas fechas, los sentimientos positivos están principalmente ausentes de la discusión, aunque esta negatividad

*Values greater than 0 imply an intense association, a strong presence of that category. Values close to 0 indicate neutrality. Negative values imply the absence of the category. The magnitude is measured in percentage points.*

*With the results, we can design an "emotional electrocardiogram" (Figure 4) that directly identifies emotional peaks and their respective dates. In anger or rage, there is one date that stands out above the others: December 21. There may be multiple explanations for that peak; however, it is feasible to think that it is due to the outrage (on the previous afternoon) of a protester run over by a police vehicle called a launcher or a "skunk".*

*Anxiety was the dominant feeling the first days of the crisis, and similar levels of anxiety appear at other times, in a cycle of three to four weeks. This anxiety swing shows a worrying sign, related to the effect on mental health of not being able to get out of a social crisis.*

*In positive feelings, there are only two peaks, Christmas and New Year, two special dates in which, despite the crisis, one could share with family and friends. Outside of these dates, positive feelings are mainly absent from the discussion although this negativity tends to diminish over time. We also see that emotions are not exclusive, and that in the vicinity of the new year it is possible to be anxious and have joy. This ambivalence calls for further study of this phenomenon.*

## IT IS POSSIBLE TO MEASURE THE EMOTIONAL FLUCTUATION IN TWITTER

*The 18O has not ended yet, but thanks to the social networks it is possible to study what is being recorded, in quasi-real time. In that context the "possible" was explored with Twitter data. It shows that the number of people discussing the issue is*

tiene una tendencia a disminuir con el tiempo. También se ve que las emociones no son excluyentes, y que en las cercanías de año nuevo es posible estar ansioso y tener alegría. Esta ambivalencia llama a estudiar con mayor profundidad este fenómeno en el futuro.

## ES POSIBLE MEDIR LA FLUCTUACIÓN EMOCIONAL EN TWITTER

El 18O no ha terminado aún, pero gracias a las redes sociales es posible estudiar lo que se va registrando, en un tiempo casi real. En ese contexto, se exploró lo posible con datos de Twitter. Se muestra que la cantidad de gente que discute el tema es considerable, sin embargo, antes de sacar conclusiones sobre lo discutido, hay que considerar las características socio-demográficas de esas personas. La red de Twitter tiene sesgos, algunos explícitos, y otros implícitos. Las personas también los tienen. Y las plataformas también los tienen --- a fin de cuentas, Twitter es un negocio, y lo que busca optimizar son sus ganancias.

El análisis revela que es posible medir la fluctuación emocional en la plataforma. Cuando es necesario entender qué es lo que siente un sector de la sociedad, y ese sector está representado en Twitter, entonces este tipo de "emociograma" permite explorar dicho sentir. Lo presentado aquí es a nivel de todos los participantes en la red, por tanto, tiene un sesgo hacia los hombres y hacia la gente joven con acceso a tecnología. Eso no impide repetir el ejercicio para subconjuntos de la población, por ejemplo, se puede hacer un cálculo para hombres y otro para mujeres, de modo de comparar las diferencias en intensidad y también en ubicación de puntos críticos.

Hay más temas que se pueden explorar con estos datos. Desde las redes de difusión, la fragmentación de las comunidades, hasta los abusos policiales y las violaciones a los derechos humanos. De hecho, la cantidad de tweets con fotos y videos es cercana al 10%. Ese porcentaje incluye memes, noticias, y también registros invaluables de lo que sucede no mostrado en los medios ni en las versiones oficiales.

La historia de un nuevo Chile está escribiéndose. A diferencia de tiempos pasados, se escribe en múltiples plataformas, con distintos medios. Es un desafío para la ciencia de datos y las ciencias sociales poder tomar todo este cuerpo de información y convertirlo en conocimiento que nos ayude a mejorar el entendimiento de nuestro pasado, presente y futuro.



*considerable; however, before drawing conclusions about what was discussed, the socio-demographic characteristics of these people has to be considered. The Twitter platform has biases, some explicit, some implicit. People have them, too. Other social network platforms have them, too--after all, Twitter is a business that seeks to optimize its profits.*

*The analysis reveals that it is possible to measure emotional fluctuation on the platform. When it is necessary to understand what a segment of society feels, and that sector is represented in Twitter, then this type of "emociogram" enables us to explore that feeling. What is presented here is at the level of all the participants in the network; therefore, it has a bias towards men and young people with access to technology. This does not prevent repeating the exercise for subsets of the population; for example, a calculation can be made for men and another for women, in order to compare the differences in intensity, and also in the location of critical points.*

*More topics can be explored with this data from outreach networks to fragmentation of communities, to police abuse and human rights violations. In fact, the number of tweets with photos and videos is close to 10%. That percentage includes memes, news, and also invaluable records of what is happening not shown in the media or official versions.*

*The history of a new Chile is being written. Unlike in the past, it is written on multiple platforms, with different media. It is a challenge for data science and social sciences to be able to take this entire body of information and turn it into knowledge that helps us to better understand our past, present, and future.*