

27/12/2016

## Jugada preparada

TXT [PABLO A. GONZÁLEZ](#)

¿Qué diferencia hubo entre las poblaciones que se expresaron en Twitter a favor y en contra del recorte en CONICET?

Siempre decimos que la ciencia es una linterna que ilumina la realidad. Eso implica pensar que apuntarle a un proceso y tratar de desentramarlo nos permite intentar entender lo subyacente, y esta es una de esas veces donde recoger, ordenar y presentar información puede contarnos una historia.

**Esta es la historia de cómo Twitter se expresó respecto del recorte en CONICET, con lo bueno, lo malo y lo feo.**

La posta de hacer visible este evento la tomaron Analia Celeste Luis ([@Analiaci](#)) y Yamila Abbas ([@abbasyami](#)), recogiendo los tweets que contenían referencias al CONICET, entre el 10 y el 21 de diciembre de 2016.

**NOTA agregada 27/12 (13:17):** Es importante remarcar que todo el análisis es válido para el evento particular, y que cualquier otro evento va a requerir un análisis particular (y tal vez hasta cruzarlo con otros eventos para encontrar patrones más grandes).

Para esto, usaron un script de Python sobre la api de Twitter que ellas mismas desarrollaron, bajaron un total de **156.000 interacciones de 29.000 usuarios**, y después se coparon en compartir no sólo su análisis sino también los datos crudos (de la parte de redes, la parte de lenguaje natural nos escapa técnicamente), para que nosotros pudiéramos aportar en esa instancia.

El '52% de los usuarios se expresaron a favor' viene de analizar lenguaje natural, con las limitaciones que eso tiene; pero dado que las limitaciones a priori son para ambos lados y que no tenemos forma automática de analizar esa cantidad de datos sin asumir un margen de error, lo vamos a tomar como 'el mejor dato que tenemos al momento, que está sujeto a mejora, pero que está bueno'.



Vamos de lo general a lo particular

**Puchito metodológico:** Para determinar la postura de los usuarios en torno a los reclamos de la comunidad científica de CONICET, se utilizaron dos métodos principales: como primer paso o criterio tomamos una muestra del 50% de la base total de tweets y analizamos el texto con el programa SPSS Text, esto nos permite clasificar los términos más significativos de acuerdo a su tendencia: favorable o desfavorable en relación al reclamo. Cada contenido, tweet o retweet puede estar clasificado en más de una postura, definiéndose la predominante a partir de la suma de concurrencias. Como segundo criterio o método de clasificación, analizamos los

*usuarios más retweeteados y centrales asignándoles una postura. Esta categorización es manual, ya que no lleva mucho esfuerzo y permite tener un control más acabado de los contenidos cargados de ironía. Por ejemplo, podríamos presumir que un tweet que contiene el término “ñoqui” o “militantes” se encuentra más cerca de estar a favor del recorte en materia de ciencia y tecnología. Sin embargo, ese criterio no es suficiente, ya que en gran parte de la conversación política predomina el sarcasmo.*

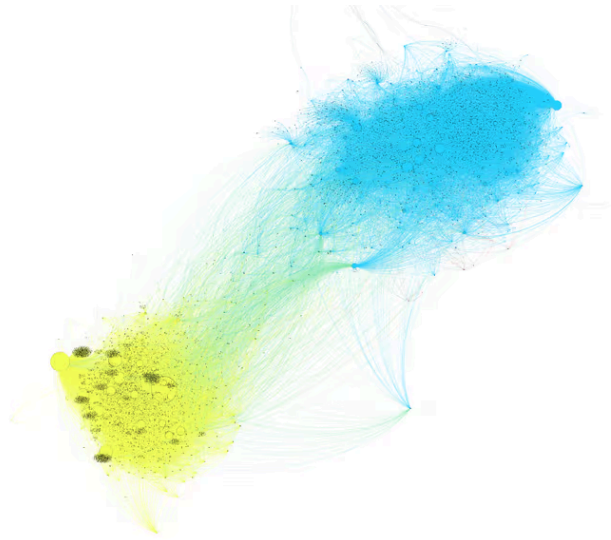
Que una conversación de esa magnitud exista no es poco, porque por lo menos nos dice que **decenas de miles de usuarios querían expresarse respecto del evento**. No es menor ver que el 84% del peso de contenido vino de retweets y no de tweets escritos originalmente. Este dato por ahora parece irrelevante, pero tengámoslo en mente porque nos va a contar un poco de historia sobre la estructura y comportamiento de la red.

Pero recién estamos entrando en calor, porque hay data para hacer dulce de data.

Una de las cosas que más nos interesaba era entender cómo se organizaban las cuentas que empujaban esta conversación, así que lo primero fue descomponer la red de interacción en clusters (grupitos que dependen de la cantidad de menciones y retweets que relacionan a los usuarios entre ellos, y que nos terminan definiendo una especie de agrupación por afinidad). Cada nodo, entonces, es un usuario, y cada arco es un vínculo entre dos usuarios.

***Puchito metodológico:*** *Para analizar la existencia de clusters corrimos un análisis de modularidad en Gephi con distinta restrictividad, o sea que analizamos condiciones distintas donde los grupos se construyen de formas más o menos estrictas, lo que genera grupos más grandes o los subdivide en sus grupos internos.*

Gracias a esa descomposición, pueden verse dos clusters bien definidos: **por un lado, un cluster que en su mayoría se refiere negativamente a CONICET; por el otro, un grupo que en su mayoría se expresa en su defensa, y que critica el recorte en ciencia y técnica.**



*Los unos y los otros*

Pero estas poblaciones no son idénticas en tamaño ni en estructura: una tiene usuarios clave bien definidos y de alto peso (lo que quiere decir que un puñado chico de cuentas se constituía como nodos centrales, o sea que mostraba gran potencia de amplificación), y la otra es una red más descentralizada, donde la distribución del peso en la conversación es más uniforme entre los usuarios.

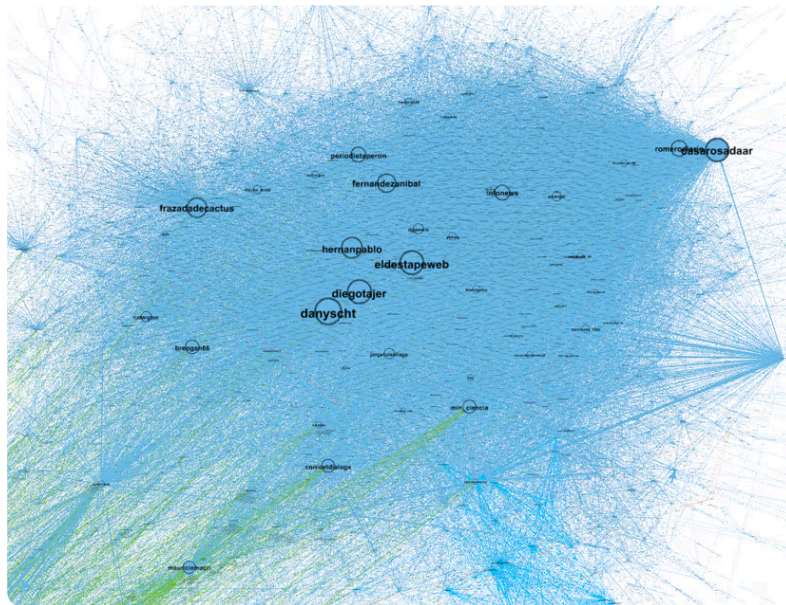


Imagen detallada del cluster que se expresaba en promedio positivamente respecto de CONICET y sus usuarios clave, donde el tamaño del @ es proporcional al peso dentro de la red (la escala es válida para ambos gráficos).

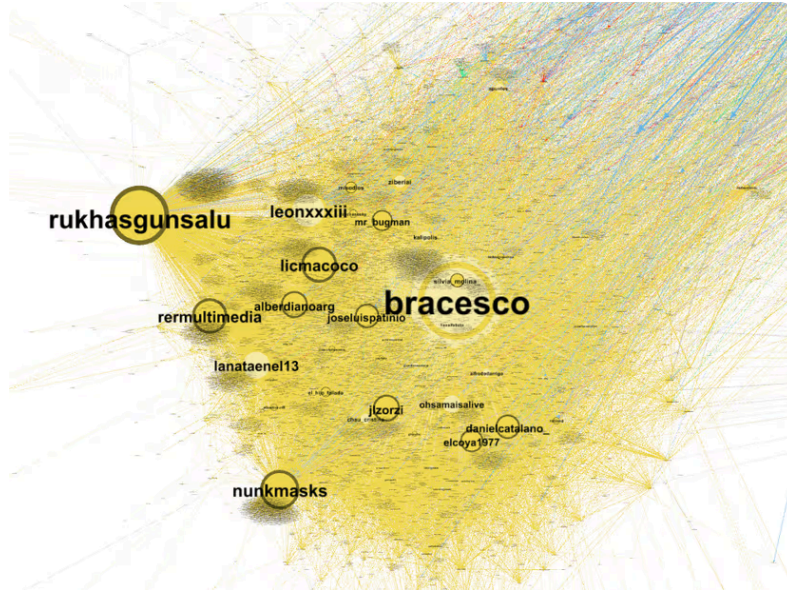
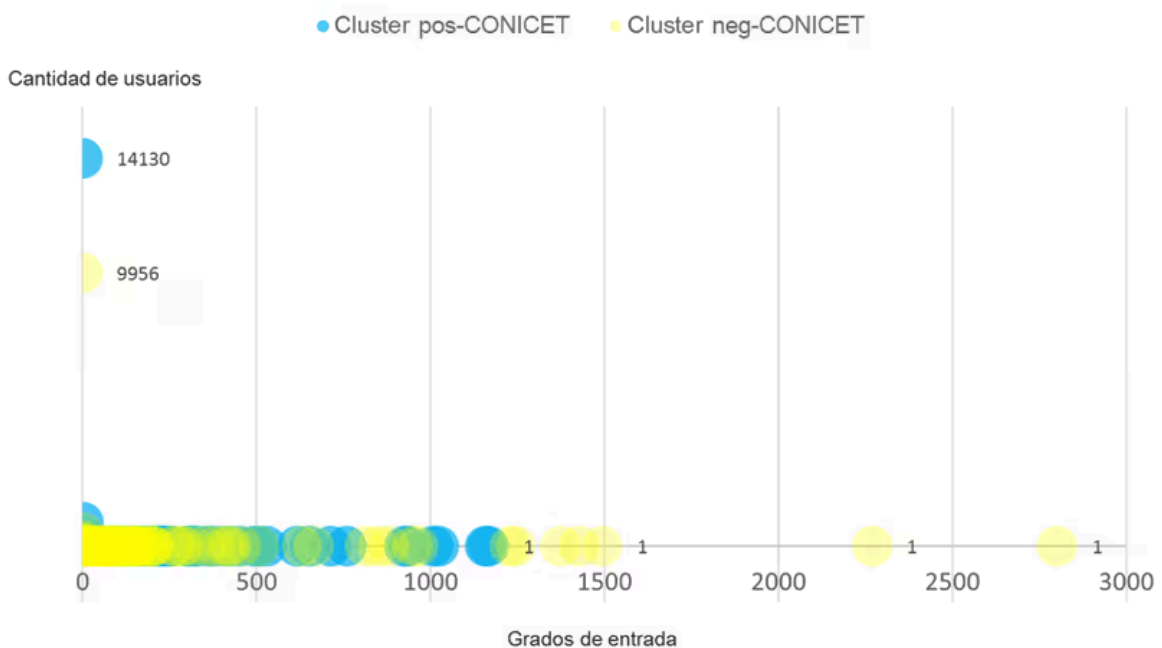


Imagen más detallada del cluster que se expresaba en promedio negativamente respecto de CONICET y sus usuarios clave, donde el tamaño del @ es proporcional al peso dentro de la red (la escala es válida para ambos gráficos).

Esta diferencia de naturaleza en los dos grupos se ve mucho más clara cuando la ordenás en un gráfico con la distribución de los grados de entrada en las distintas comunidades.

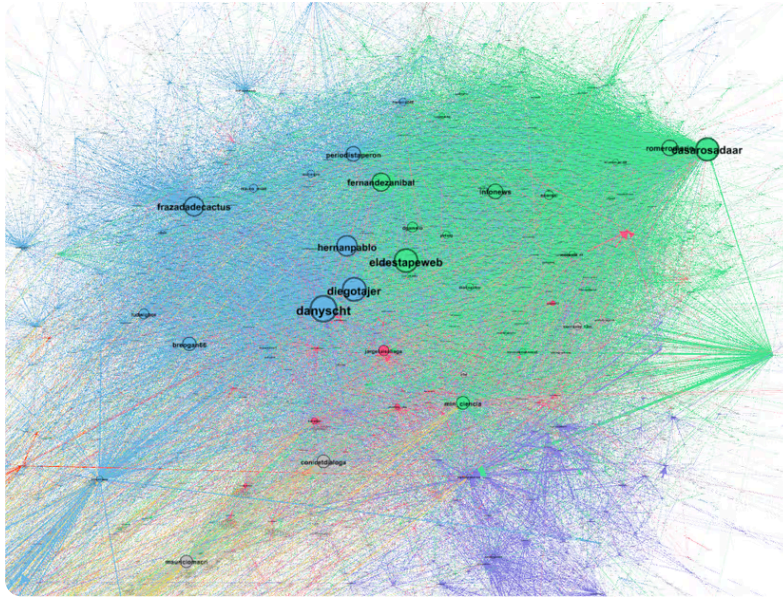
**Metodología:** El grado de entrada sirve para entender cómo es la dinámica de interacción de los usuarios de cada comunidad; básicamente son las menciones y retuits entrantes que tienen los nodos de cada cluster.



En el eje de coordenadas podemos ver cómo los usuarios en favor de la comunidad científica se encuentran más cercanos al eje Y. Esto significa que **hubo muchos más usuarios que no fueron retweeteados o no tuvieron menciones entrantes** (recordemos el punto anterior sobre cómo la conversación fue dominada por los retweets y no por los tweets novedosos). Para ejemplificar un poco más en base al gráfico, hubo **al menos 14 mil usuarios que tuvieron grado de entrada 0. Esto significa que expresaron su opinión activamente, sin aglutinarse en torno a un usuario central.**

Por otro lado, en el cluster que se manifestó en favor del recorte se ven 7 burbujas que se alejan del eje Y superando el grado de entrada 1000. Esto significa que podemos ver que al menos 7 usuarios aglutinaron más de 1000 menciones o retuits, dando lugar a pensar que se trata de una comunidad jerarquizada, organizada a partir de líderes de opinión en su red. Qué cosa hermosa cuando los datos cuentan una historia.

No es menor notar también que los principales agentes del cluster neg-CONICET se muestran a sí mismos como oficialistas y anti-kirchneristas. Dentro del cluster contrario al recorte hay algunas cuentas fácilmente identificables como kirchneristas, pero estas cuentas no dominan el cluster entero, y esto no es menor. Si nos ponemos más finos con el agrupamiento (o sea, más restrictivos, y en vez de armar dos grupos grandes, armamos más), adentro de ese mismo grupo de expresión positiva respecto del CONICET (pero que, recordemos, se agrupó por su nivel de interacción interno, no por su forma de expresarse respecto del conflicto) hay varias poblaciones separadas: la recién mencionada, y una dominada por usuarios como [@frazadadecactus](#), [@breogan66](#), [@danyscht](#) y [@diegotajer](#), a la que no se la puede caracterizar políticamente de forma tan sencilla o unívoca (en el eje kirchnerista/antikirchnerista) [*paréntesis agregado 28/12*].



La misma, misma imagen detallada del cluster que se expresaba positivamente respecto del CONICET y sus usuarios clave, donde el @ es proporcional al peso dentro de la red, pero esta vez con un criterio de agrupamiento más restringido para poder ver comunidades dentro de la comunidad (la escala es válida para con los gráficos anteriores). Dentro del cluster de apoyo, identificamos dos poblaciones, cada una con sus referentes (verde y celeste).

Entonces, tenemos dos (o tres) poblaciones claras. Nuestro interés llegado ese punto fue evaluar el funcionamiento interno de cada una de las dos grandes poblaciones originales (neg/pos CONICET).

Para eso, empezamos por tratar de entender de qué habla cada grupo. Cuando el análisis se da sobre el cluster que se expresa negativamente respecto de CONICET, aparece esto:



Montonero Mufasa, RENUNCIE

Así se hacen claros los memes dominantes del cluster negativo, con mucha presencia de **Mufasa, la NASA y Perón**; con las palabras ‘militantes’ y ‘ñoquis’ como calificativos preferidos y con un principal interés en atacar Sociales,



Dado que la conversación muestran este comportamiento tan diferente durante la semana, podemos empezar a pensar que ese patrón esconde algún tipo de característica distintiva entre las poblaciones (aunque bien podría ser simplemente simplemente que es fin de semana). Tercera idea a sostener, junto con el porcentaje de retweets y la estructura distintiva de ambas redes.

Algo que nos llamó mucho la atención es que muchos de los usuarios inmediatamente adyacentes a los centrales del cluster negativo tienden a no tener ningún tipo de identificación, estética o estilo personal discursivo que nos haga pensar que son personas expresando su opinión en redes. Exceptuando el caso del conocido ex periodista de Clarín @Bracesco, en la red de usuarios neg-CONICET abundan las cuentas con estética despersonalizada: caricaturas, fotos de bancos de datos y portadas en baja resolución(\*). Podrán pensar “OK, yo tampoco tengo una estética que sea fácilmente identificable”, pero esto contrastó fuertemente con los usuarios de mayor centralidad en la red (**NOTA sobre un comentario:** *este patrón también apareció en los adyacentes, no solamente en los centrales, pero está pendiente un análisis mucho mejor sobre esta comparación*) que se manifestó en apoyo a la comunidad científica de CONICET. Los usuarios con mayor volumen de menciones y retweets en esta red fueron referentes académicos, investigadores y algunos portales digitales. El punto es que la opinión del cluster que apoya al CONICET suele estar más referenciada: sabemos quién nos habla y desde dónde.

(\*)



[Christmas Costumes & Santa Claus Suits - HalloweenCostumes.com](http://www.halloweencostumes.com)



[www.halloweencostumes.com](http://www.halloweencostumes.com) › [Holiday Costumes](#) › [Traducir esta página](#)  
250 x 357 - Our Christmas costumes are perfect for spreading the holiday spirit. Get one of these **adult Santa Claus costumes** for Christmas parties or any other event.

Quiero ser un niño de verdad! NOTA 27/12, 13:25 @pablоторres es un mal ejemplo acá porque sí es un usuario real y bancamos que diga LO QUE QUIERA. Esto no quiere decir que no se encuentre (quiera o no) en una estructura que contiene cuentas que podrían no ser de personas reales).

De nuevo, esto no es análisis, es dato. Y es correlación, no causalidad. Porque correlación no implica causalidad. Correlación sí implica correlación. Muy tautológico todo.

Y esto ya sería un montón, pero hay otra posibilidad de amasar los datos y tratar de atacar una historia más. **Muchos acusan la existencia de una estructura organizada de operadores, y otros consideran que eso es un mito**, así que veamos si hay razones o no para pensar que esa idea de manipular la opinión pública usando cuentas que no responden a personas reales se sostiene o se debilita con datos. A lo que ya sabemos sobre la distribución jerárquica y centralizada de usuarios del cluster neg-CONICET y sobre su alta tasa de retweets sobre tweets originales, podemos sumarle el dato de que **el clúster negativo tenía alto peso relativo de lunes a viernes y muy bajo los fines de semana**, y podemos agregar un dato más: la fecha en la que esas cuentas fueron creadas.



*Me viá abrir una cuenta en el Tuitter*

Bueno.

Si bien esto no es una confirmación de que exista una actividad organizada para generar la amplificación de determinada información (en este caso en particular, opinión en contra de CONICET), sí es interesante y sí está bueno que tengamos en mente y a la vista este comportamiento y nos preguntemos qué modelos explican mejor esta acumulación de evidencia (y hasta qué medir para descartar

hipótesis alternativas). Un gran paso siguiente podría ser ahondar minuciosamente en el comportamiento de cada una de estas cuentas (patrones como la tasa de tweets originales contra los retweets, un ejercicio divertido para cuentas como [@gabrielluna007](#), [@azulinamorena](#) o [@esthermdq](#), protagonistas del cluster neg-CONICET, que son básicamente cuentas de retweets sin generación de contenido propio pero al mismo tiempo con decenas de RTs diarios, lo que nos hace pensar en un sistema organizado de amplificación de contenido, o en usuarios reales muy entusiasmados con retweetear, pero nunca expresarse personalmente).

***Nota 27/11:** Hay un par de comentarios excelentes sobre explicaciones alternativas a estas conductas. Desde la idea de que es más fácil oponerse a las cosas en general desde lo anónimo y a favor desde lo personal a la 'hipótesis oficinista' de que los retuiteadores lo hacen durante su horario de trabajo normal y no lo hacen durante los fines de semana. Otra idea interesante es suprimir cualquier tipo de intuición de intención organizada y suponer que los comportamientos que asumimos automatizados no son parte de una estructura (formal o semi formal), sino de una población de usuarios reales que se expresan en tasas de decenas o centenas de RTs por día. Tal vez esta sea la hipótesis más desesperanzadora, pero no por eso menos válida o menos atendible de confirmar. Tendremos que buscar los métodos para abordarla.*

Esto no significa que tengamos certezas. Desde ya que hay explicaciones alternativas al escenario organizado para la generación de opinión concertada. Hay escenarios de entusiasmados retweeteadores a los que no les copa poner su foto, que encima no opinan los fines de semana y que justo se dio que el día que abrieron la cuenta había habido un evento político relevante. Pero, al que le quepa el Bayes, que se lo ponga.

Toda esta información (y su análisis) debe ser vista con la cautela de quien presenta un preliminar y asumiendo buena voluntad y capacidad de refinar cada una de las estimaciones, pero al mismo tiempo como una señal extremadamente clara de cómo **Twitter como soporte y medio no es ajeno a la posibilidad de ser analizado, entendido, expuesto y tal vez hasta manipulado** para tomar ventaja de algunos de los sesgos más íntimos del humano: nuestra capacidad de acordar y dar por válida información que se solapa con lo que creemos (aun sin chequearla,

ni a ella ni a sus emisores), nuestra tendencia a expedirnos con seguridad sobre áreas de las que no tenemos idea (sobreestimando nuestra capacidad de análisis y el ámbito en el que somos idóneos), y nuestra reticencia a admitir como válida evidencia que pueda ir en contra de las posturas que ya decidimos tomar.

Entonces, hablemos abiertamente de todos los temas que entendamos importantes. Hablemos del recorte. Hablemos del papel de la ciencia en la industria, la salud y la educación, de lo imprescindible de la ciencia básica, de lo igual de necesaria que es la aplicada, de cómo hacer ciencia para la sociedad plena, de cómo evaluar a nuestros científicos profesionales. Nada es perfecto ni sagrado, CONICET tampoco. Discutamos qué hace y cómo lo hace. Discutamos si necesita modificaciones y cuáles. Sobre todo y más allá de esa institución, hablemos de si existe una dirección pensada, estudiada y expresada abiertamente que muestre una visión sostenida y a largo plazo para el desarrollo de la ciencia, la tecnología y la educación. Pero no por discutir todo esto dejemos de hablar de cómo priorizamos la ciencia en el gran arco argumental de un plan país. Porque **la charla de fondo a tener es esa: ciencia sí o no, y cuánto.**

Precisamente por eso es que todas esas conversaciones tenemos que poder tenerlas hablando claro y hablando cada uno. Sabiendo que los que opinan son otros como nosotros, ciudadanos comunes, que fuimos habitando un medio que supuso en algún momento la posibilidad de democratizar el acceso a decir lo que pensamos; hasta que alguien se dió cuenta de que Twitter es un medio como cualquier otro, y que hecho el medio, hecha la rosca.

Aprendimos a leer los diarios como de quien vienen, a tratar de entender las intenciones por detrás. Quizás les toca ahora a las redes sociales ser observadas de manera de descubrir si hay piolines que hacen que las marionetas bailen, griten y retweeteen. Puede ser la única forma que tengamos de reclamar la posibilidad de discutir temas importantes sin que nadie sarpe la conversación gritando o usando humo y espejos para hacer que lo que piensa suene como rugido de masa, cuando en realidad es solamente eco.

**NOTA 28/12:** *Cambiamos las nomenclaturas de los clústers a neg-CONICET y pos-CONICET para reflejar lo más unívocamente posible que la categorización se dió por ‘este grupo se expresó en promedio negativamente respecto de CONICET’ y ‘este grupo se expresó en promedio positivamente respecto de CONICET’.*

[elgatoylacaja.com/jugada-preparada](http://elgatoylacaja.com/jugada-preparada)

---

Sumate en   
[eglc.ar/bancar](http://eglc.ar/bancar)