



**APLICACIONES DE CIENCIA DE LOS DATOS Y
REDES SOCIALES A FENÓMENOS SOCIOPOLÍTICOS**

FACULTAD DE ESTUDIOS ESTADÍSTICOS

**Polarización en la movilización universitaria: un
estudio de ciencia de datos sobre el debate de Vito
Quiles en X**

Docente : Arturo Rodríguez Sáez

Estudiante: Pablo Galarón Mateo

Fecha: Noviembre de 2025

ÍNDICE

Resumen.....	2
1. Introducción.....	2
2. Objetivo y preguntas de investigación.....	3
3. Marco teórico, antecedentes y fundamento metodológico.....	3
3.1 Marco teórico	3
▪ Polarización.....	4
▪ Populismo.....	4
▪ Derecha Radical y Extrema Derecha.....	5
▪ Movimientos sociales y movilización estudiantil...5	
▪ Análisis de Redes Sociales (SNA).....	6
3.2 Antecedentes y estudios previos.....	6
3.3 Fundamentación teórica de las decisiones metodológicas.....	7
4. Hipótesis y variables.....	7
4.1 Hipótesis.....	7
4.2 Variables.....	9
5. Diseño metodológico.....	9
5.1 Fuentes de datos.....	10
▪ API X.....	10
▪ Web Scraping.....	11
▪ Consideraciones éticas.....	11
5.2 Herramientas y técnicas de análisis.....	11
6. Ventajas, Limitaciones y consideraciones éticas.....	14
7. Análisis con ejemplos de tuits reales.....	15
8. Tratamiento y exploración con datos reales.....	17
8.1 Carga de datos.....	18
8.2 Aplicación del modelo.....	19
8.3 Nube de palabras.....	20
8.4 Gráfico SNA.....	21
9. Bibliografía.....	23

Resumen.

Este trabajo analiza el debate que se generó en X a raíz de las visitas de Vito Quiles a distintas universidades españolas y de las movilizaciones estudiantiles que surgieron como respuesta. El objetivo principal es comprobar si la conversación en la plataforma estuvo polarizada y, además, identificar qué otros elementos como marcos populistas, discursos de derecha radical/extrema derecha y la propia movilización estudiantil influyeron en esa división.

Para ello, se recopilarán tuits publicados en torno a las fechas de los eventos, seleccionando palabras clave, menciones y hashtags relevantes. A partir de estos datos se aplicarán diferentes herramientas de ciencia de datos, como modelos capaces de identificar ciertos marcos discursivos y métodos de análisis de redes sociales. Con estas técnicas será posible observar qué grupos de usuarios participaron, qué cuentas tuvieron mayor influencia en la difusión del debate y cómo se estructuraron las interacciones.

El trabajo también tiene en cuenta cuestiones éticas, como evitar información que pueda identificar a los usuarios y reconocer que los datos de redes sociales siempre tienen límites. En resumen, el objetivo es entender cómo un conflicto universitario puede pasar a las redes y hacerse más visible a través de la manera en que se comparte y se comenta en X.

1 Introducción.

En los últimos años, X se ha convertido en uno de los lugares donde más se discuten temas políticos y sociales. Lo que antes quedaba dentro de una universidad o de un grupo reducido ahora puede llegar a miles de personas en cuestión de minutos. Esto ocurre porque las redes funcionan con las conocidas “tres V” del Big Data: volumen (se generan muchos mensajes), velocidad (se publican de forma constante y muy rápida) y variedad (hay opiniones, vídeos, memes, comentarios políticos, etc.). Por eso, cualquier conflicto local puede amplificarse y transformarse en un debate mucho más grande.

Las visitas de Vito Quiles a varias universidades españolas son un ejemplo claro de este proceso. Sus intervenciones generaron protestas de estudiantes, debates sobre libertad de expresión y también muchas reacciones en X. Vito Quiles es una figura relacionada a menudo con discursos de derecha radical/extrema derecha y con un estilo comunicativo que incluye elementos del populismo, como la división entre “la gente” y “las élites(esteste gobierno actual)”, o la idea de que ciertos grupos controlan las instituciones. Esta forma de comunicar, sumada a la tensión que hubo en los campus, hizo que la discusión en X se polarizara rápidamente entre quienes apoyaban su presencia y quienes la rechazaban.

Este trabajo analiza cómo se desarrolló ese debate en X. El interés está en ver si realmente hubo polarización entre los usuarios. Analizando a su vez si aparecieron mensajes con marcos propios del populismo o de la derecha radical/extrema derecha y qué grupos se formaron dentro de la conversación. También resulta importante observar cómo se difundió el debate: qué cuentas tuvieron más influencia, cómo se conectaron los distintos usuarios y qué tipo de argumentos dominaban.

El objetivo final es entender cómo un conflicto universitario puede empezar de forma local, trasladarse a una red social como X y amplificarse gracias a las dinámicas propias de esta plataforma. Este tipo de casos muestra cómo las redes no solo reproducen debates políticos, sino que también ayudan a moldearlos y a darles una dimensión mucho mayor que la que tienen fuera de internet.

2 Objetivo y preguntas de investigación.

El objetivo de este trabajo es analizar cómo se desarrolló el debate sobre la visita de Vito Quiles a varias universidades españolas dentro de la plataforma X y qué dinámicas políticas aparecieron durante esos días. En concreto, se quiere observar si la conversación mostró señales de polarización. Para ello se analizará si aparecieron discursos relacionados con el populismo o la derecha radical/extrema derecha, y cómo se organizaron los usuarios dentro de la red.

Preguntas de investigación principales.

- ¿La conversación en X sobre las visitas de Vito Quiles mostró niveles claros de polarización entre usuarios con posturas opuestas?
- ¿Se identifican mensajes o marcos discursivos que puedan relacionarse con el populismo o con ideas propias de la derecha radical/extrema derecha? Si es así, ¿es derecha radical o extrema derecha?
- ¿Cómo se estructuró la interacción entre usuarios durante el debate (grupos, nodos influyentes, conexiones, etc.)?

3 Marco teórico, antecedentes y fundamento metodológico.

3.1 Marco teórico.

El caso de las visitas de Vito Quiles a universidades españolas conecta directamente con varios conceptos trabajados en clase, en este trabajo me ha gustado combinar varios de estos conceptos a modo de trabajo final. Para entender cómo se desarrolla un conflicto político dentro de una red social como X, es necesario apoyarse en nociones como polarización, populismo, derecha radical/extrema derecha, movimientos sociales, Big Data y análisis de redes sociales (SNA). A continuación se recogen los principales marcos teóricos que permiten interpretar este fenómeno.

Polarización.

La polarización se produce cuando las posiciones políticas se alejan cada vez más hasta formar dos bandos enfrentados y con poco espacio intermedio. No solo aparece en el plano ideológico, sino también en el emocional por ejemplo. Por eso se suelen distinguir tres tipos principales:

- Polarización ideológica, centrada en las diferencias en temas políticos y sociales.
- Polarización afectiva, basada en los sentimientos positivos hacia el propio grupo y el rechazo hacia el contrario.
- Polarización de identidad, que mezcla política con valores culturales, estilos de vida o identidades más profundas.

Este marco encaja muy bien con el caso que propongo de Vito Quiles, porque el debate en X no fue solo sobre ideas, sino también sobre emociones y pertenencias. La dinámica de “nosotros contra ellos” apareció de forma clara entre quienes defendían su presencia y quienes la criticaban, reforzando tanto la distancia ideológica como la afectiva. En este debate también aparecen expresiones que se han popularizado en los últimos años a causa de la creciente polarización política en España, como la idea de “las dos Españas”. Este marco encaja muy bien con lo que ocurrió en el caso de Vito Quiles, donde se vio una división clara entre dos posiciones enfrentadas que apenas compartían espacios comunes.

Populismo.

El populismo puede entenderse como una forma de comunicación política que simplifica los conflictos y los plantea como enfrentamientos entre dos grupos muy claros. En este tipo de discurso suelen aparecer tres elementos básicos:

- Retórica anti-elitista, que señala a ciertos grupos (políticos, instituciones, universidades o medios) como responsables de los problemas.
- Apelación a un “pueblo” idealizado, presentado como la parte moralmente legítima de la sociedad.
- Invocación de la voluntad popular, defendiendo que “la mayoría” debe imponerse por encima de cualquier mediación.

A esto se añaden rasgos como el uso de emociones fuertes, la simplificación de conflictos complejos y la tendencia a dividir a la sociedad entre “los que dicen la verdad” y “los manipuladores”. Estos marcos aparecieron con facilidad en la conversación sobre Vito Quiles, donde algunos mensajes hablaban de universidades “controladas por el gobierno Sanchista” o se presentaban como la “voz del pueblo frente a la dictadura Sanchista”.

Derecha Radical y Extrema Derecha.

La derecha radical combina nativismo, autoritarismo y conservadurismo social, aunque sin romper del todo con la democracia. La extrema derecha lleva estas ideas más lejos, cuestionando principios básicos de la democracia liberal y promoviendo posturas más excluyentes o jerárquicas. En redes sociales, estos discursos suelen aparecer a través de mensajes sobre amenazas culturales, críticas a instituciones que consideran progresistas, defensa de valores tradicionales o discursos antiinmigración.

Dado que Vito Quiles es una figura asociada con estos marcos, no es raro que en X surgieran mensajes que encajan con la derecha radical e incluso algunos más cercanos a la extrema derecha. Analizar estos patrones permite entender mejor la orientación ideológica del debate.

Movimientos sociales y movilización estudiantil.

Los movimientos sociales son acciones colectivas que buscan un cambio social o político y que se basan en un sentimiento compartido de agravio y en la creación de una identidad común. Sus mensajes suelen cumplir funciones distintas:

- informativa (convocatorias, horarios, explicaciones)
- movilizadora
- emocional (indignación, apoyo, preocupación)
- culpabilizadora (señalar a los responsables)

En las protestas universitarias vinculadas a Quiles aparecieron exactamente estas funciones: había avisos de concentración, vídeos, mensajes de apoyo entre estudiantes y críticas dirigidas a distintos actores. En X estas dinámicas se amplificaron rápidamente, permitiendo que la movilización local se transformara en un debate mucho más amplio.

Análisis de Redes Sociales (SNA).

El análisis de redes sociales estudia cómo se conectan los usuarios dentro de una plataforma. Utiliza conceptos como:

- nodos (usuarios),
- enlaces (retuits, menciones),
- centralidad (quién influye más),
- intermediación (quién conecta grupos),
- modularidad y comunidades (cómo se agrupan los usuarios).

Estas herramientas permiten ver si el debate se fragmentó en grupos, qué perfiles actuaron como líderes, qué mensajes circularon más y si los dos bandos interactuaron o prácticamente estaban aislados. Esto ayuda a complementar el análisis del contenido con una visión más estructural del conflicto.

3.2 Antecedentes y estudios previos.

En los últimos años se han publicado varios estudios que ayudan a entender cómo funcionan debates políticos tensos en redes como X y por qué tienden a generar posiciones muy enfrentadas.

- “Twitter and Affective Polarisation: Following Political Leaders in Spain”

Este estudio analiza cómo X contribuye a la polarización afectiva en España, especialmente cuando los usuarios siguen a líderes políticos. Muestra que la exposición constante al discurso político digital aumenta la hostilidad entre los dos bandos.

Lorenzo-Dus, N., & Bou-Franch, P. (2022). *Twitter and affective polarisation: Following political leaders in Spain* (Tesis de máster). Universitat Pompeu Fabra.

https://repositori.upf.edu/bitstream/handle/10230/60443/Lorenzo_sesp_twit.pdf

- “The Use of Social Media in the Populism of Chega and VOX”

Compara cómo los partidos de derecha radical Chega (Portugal) y Vox (España) usan redes sociales para difundir discursos populistas y nativistas, mostrando que estas estrategias aumentan polarización y movilización.

Lisi, M., & Borg, S. (2023). *The use of social media in the populism of Chega and Vox*. Social Sciences, 12(5), 255.

<https://doi.org/10.3390/socsci12050255>

- “Populismo en redes sociales. Análisis del discurso del movimiento 15M, Podemos y VOX en Twitter”

Analiza cómo 15M, Podemos y Vox utilizan X para difundir discursos populistas basados en “pueblo vs élites”, mostrando que los mensajes emocionales y confrontativos generan mayor interacción.

Pérez, L. (2021). *Populismo en redes sociales: Análisis del discurso del movimiento 15M, Podemos y VOX en Twitter* (Tesis doctoral). Universidad Pontificia Comillas.

<https://repositorio.comillas.edu/xmlui/handle/11531/83501>

3.3 Fundamentación teórica de las decisiones metodológicas.

La metodología del trabajo se basa en las teorías explicadas en el marco teórico. Por un lado, analizar el contenido del debate en X tiene sentido porque la polarización, el populismo y los discursos de derecha radical se identifican sobre todo a través del lenguaje y de los marcos que usan los usuarios cuando comentan un tema político. La teoría muestra que estos discursos se reconocen en ideas como “pueblo vs élites”, el uso de identidad política o la confrontación entre dos bandos, por lo que revisar los mensajes es necesario para ver si aparecen estas dinámicas.

Por otro lado, el uso del análisis de redes sociales (SNA) está justificado porque la polarización no solo se expresa en lo que se dice, sino también en cómo se organiza la conversación. Según la teoría, cuando hay un debate muy tenso, las interacciones tienden a separarse en dos comunidades con poca conexión entre ellas. SNA permite comprobar si esto ocurre realmente en el caso de Vito Quiles, ver qué grupos se formaron y qué usuarios tuvieron más influencia.

Además, tiene sentido usar datos de X porque sus características (volumen, velocidad y variedad) hacen que sea un buen lugar para estudiar cómo un conflicto local se amplifica rápidamente y se convierte en un debate mucho más grande.

4 Hipótesis y variables.

En pocas horas la conversación se llenó de mensajes emocionales, acusaciones cruzadas de ambos bandos y lecturas políticas del conflicto universitario. Este comportamiento encaja con muchos de los fenómenos explicados en la teoría, así que planteo las siguientes hipótesis y variables para comprobar si realmente aparecieron estas dinámicas.

4.1 Hipótesis.

Primera hipótesis(PRINCIPAL): Polarización.

El debate sobre Vito Quiles en X mostrará polarización, tanto en el contenido como en la estructura de la red. Se espera encontrar dos comunidades muy separadas:

- una que lo defiende (“los estudiantes están adoctrinados”, “Vito dice la verdad que nadie quiere oír”),
- y otra que lo critica (“este tipo no pinta nada en la universidad”, “las protestas son legítimas”).

Mini ejemplo:

- Después de que Quiles subiera vídeos desde el campus, muchos usuarios respondieron con mensajes como “normal que lo echen, viene solo a provocar”, mientras que otros replicaban “la universidad pública está tomada por la izquierda, bien por ir allí a decir lo que otros callan”. Dos posturas sin punto medio ->> polarización clara.

Segunda Hipótesis: Populismo.

En el debate aparecerán marcos populistas, pero no en el sentido clásico de “pueblo vs élites económicas”, sino marcos populistas adaptados al contexto, donde la élite señalada es la universidad, los estudiantes organizados o incluso el gobierno. Es decir, planteo que Vito Quiles y sus seguidores se presentan como quienes “dicen la verdad” frente a unas élites progresistas que intentarían censurarlo.

Mini ejemplo:

- “En la universidad solo hablan los de siempre. Cuando va alguien que piensa distinto, lo intentan callar.”

Tercera Hipótesis: ¿Derecha Radical o Extrema Derecha?

Una parte importante de los mensajes a favor de Vito Quiles mostrará rasgos de derecha radical, ya que este discurso encaja con el perfil ideológico de muchas cuentas que lo apoyaron. Lo más común serán críticas a la universidad por “adoctrinamiento”, acusaciones de sesgo ideológico y referencias a valores tradicionales. En algunos casos aparecerán señales de extrema derecha, como símbolos identitarios duros o banderas preconstitucionales, pero serán minoritarios. La tendencia dominante en el debate será la de derecha radical.

Mini ejemplo:

- “Las universidades están totalmente inclinadas hacia la izquierda; así no hay pluralidad.”(derecha radical, más predominante)
- Usuarios que añaden banderas preconstitucionales o símbolos ultras en su perfil.(extrema derecha, casos puntuales)

Cuarta Hipótesis(SNA): ¿Cómo será la organización de la red?

Se espera que el debate sobre Vito Quiles en X se organice en dos comunidades casi aisladas, una que lo apoya y otra que lo critica. En cada bloque aparecerán algunos perfiles influyentes como comentaristas políticos, medios afines o asociaciones estudiantiles que actuarán como nodos centrales.

4.2 Variables.

Antes de definir las variables, conviene aclarar que este trabajo no se centra solo en medir la polarización. En realidad, lo que busca es entender cómo se mezclan varios fenómenos dentro del debate sobre Vito Quiles en X: los marcos populistas, los discursos de derecha radical, la movilización estudiantil y la forma en que los usuarios se agrupan en la red.

Aun así, para organizar el análisis y para utilizar adecuadamente los futuros modelos de Ciencia de Datos, la polarización se toma como variable dependiente, porque es el resultado final de todo lo anterior: si el debate acaba dividido en dos bloques, si apenas hay interacción entre ellos y si las posiciones se vuelven más extremas. El resto de los elementos (populismo, derecha radical/extrema derecha, estructura de la red y las protestas en los campus) se usan como variables independientes, ya que ayudan a explicar por qué se produce esa división y qué factores la refuerzan. Se ilustra todo esto en la siguiente tabla:

Tipo de variable	Nombre	Qué mide	Cómo se reconoce en el debate en X
Variable dependiente (VD)	Polarización del debate	Nivel de división entre dos grupos enfrentados en X	Cuando el debate se organiza en dos bloques que apenas interactúan y defienden posiciones opuestas sin puntos intermedios
Variable independiente (VI1)	Marcos populistas	Discursos que simplifican el conflicto como “ellos contra nosotros”	Mensajes que hablan de “los de siempre”, “no dejan hablar a quien piensa distinto” o “la universidad está controlada”
Variable independiente (VI2)	Discursos de derecha radical	Críticas a la universidad por sesgo ideológico o defensa de valores tradicionales	Tuits sobre “adoctrinamiento”, “campus dominados por la izquierda”, símbolos conservadores, etc.
Variable independiente (VI3)	Organización de la red (SNA)	Cómo se conectan los usuarios en X y qué grupos se forman	Retuits dentro del mismo grupo, aparición de nodos influyentes, dos comunidades separadas.
Variable independiente (VI4)	Movilización estudiantil	Peso de los mensajes relacionados con protestas y acciones en los campus	Videos de manifestaciones, pancartas, convocatorias, grabaciones desde la universidad

5 Diseño metodológico.

El diseño metodológico de este trabajo es teórico, pero se plantea como si fuera un estudio real sobre el debate generado en X por las visitas de Vito Quiles a distintas universidades. El objetivo es mostrar qué pasos se seguirían para analizar polarización, marcos populistas, discursos de derecha radical y la organización de la red, utilizando técnicas de ciencia de datos y análisis de redes sociales (SNA).

5.1 Fuentes de datos.

API de X.

La API de X permite descargar tuits de manera controlada usando palabras clave, hashtags o menciones. Este sería el método principal porque es la forma más directa y ordenada de obtener datos.

¿Qué elementos recopilaría y por qué?

Tipo de elemento	Ejemplos aplicados al caso	Para qué sirve en el análisis
Hashtags	#VitoQuiles, #Universidad, #LibreExpresión, #Adoctrinamiento	Detectar eventos, debates activos y picos de conversación
Palabras clave	“estudiantes”, “protesta universitaria”, “le echaron de la facultad”, “manifestación en campus”, “vito”, “Vito”, “Quiles”.....	Recoger mensajes relacionados aunque no usen hashtags
Menciones	@VitoQuiles, cuentas de asociaciones estudiantiles, medios que comentaron el conflicto	Identificar actores relevantes y conexiones entre usuarios
Metadatos recopilados	Fecha del tuit, número de retuits, respuestas, menciones entre usuarios	Son esenciales para análisis de redes sociales (SNA) porque permiten ver quién interactúa con quién y qué mensajes se difunden más

Web Scraping.

El **web scraping** consiste en extraer información directamente de una página web mediante un programa que la lee de forma automática. No sería el método principal, pero puede servir en casos concretos, por ejemplo:

- cuando la API no devuelve tuits muy virales
- para capturar capturas de hashtags que no se pueden descargar
- o para recopilar conversaciones públicas que aparecen en la web de X sin necesidad de iniciar sesión.

Consideraciones éticas.

Aunque este estudio es teórico(aunque al final haré un trabajo con unos pocos datos), trabajar con datos de X siempre implica ciertas precauciones. La idea no es señalar a usuarios anónimos, sino entender cómo funciona la conversación en general. Aun así, hay que tener en cuenta lo siguiente:

- Anonimizar a cualquier usuario normal, evitando nombres, fotos o tuits literales.
- Sí se pueden mencionar cuentas públicas (periodistas, medios, asociaciones, políticos) para el SNA, porque forman parte del debate público.
- Ser consciente de que X tiene mucho sesgo, así que lo que se ve allí no representa a toda la sociedad.
- Interpretar los datos con cuidado, sin sacar conclusiones exageradas ni estigmatizar a ningún colectivo.

5.2 Herramientas y técnicas de análisis.

Después de recoger y limpiar los tuits mediante las fuentes ya explicadas (API de X y, si hiciera falta, scraping), el siguiente paso sería analizar si el debate está polarizado. Para ello se combinarían dos métodos: mirar el contenido de los mensajes y estudiar cómo se organiza la red de usuarios.

En la parte del contenido, la idea sería entrenar un modelo de clasificación que distinga entre tuits “a favor”, “en contra” o “neutros”. Para esto se podría usar un modelo sencillo como Naive Bayes, que funciona bien con texto, o un SVM, que separa muy bien categorías cuando las posiciones están muy marcadas. Primero habría que etiquetar manualmente un pequeño conjunto de tuits para que el modelo aprenda qué tipo de mensajes van en cada clase. Si luego, al aplicarlo al resto de los datos, casi todos los tuits acaban en los dos extremos y apenas aparecen mensajes moderados, eso sería una señal clara de polarización discursiva.

En paralelo se analizaría la estructura del debate construyendo una red a partir de retuits y menciones. Aquí cada usuario sería un nodo y cada interacción un enlace. Con esta red se podrían usar técnicas típicas de análisis de redes, como la detección de comunidades y el modularidad, para ver si los usuarios se agrupan en dos bloques que apenas interactúan entre sí. Si aparecen dos comunidades grandes y muy separadas, eso indicaría polarización estructural.

Juntando ambos análisis, el del contenido y el de la red, se puede decidir si la variable dependiente, la polarización, está realmente presente en el debate. Si el modelo muestra dos posiciones muy marcadas y la red aparece claramente partida en dos, se podría afirmar que la conversación sobre Vito Quiles en X está polarizada.

Después de comprobar si el debate está polarizado, las demás variables independientes se analizarían con modelos parecidos, pero adaptados a cada aspecto del contenido. Para los marcos populistas, lo más útil sería entrenar un modelo sencillo como Naive Bayes o SVM, capaz de reconocer frases del estilo “no dejan hablar al que piensa distinto”, “la universidad está controlada por los mismos” o cualquier mensaje que plantee el conflicto como un choque entre “los del gobierno Sanchista” y “los seguidores de Vito”. En el caso de los discursos de derecha radical, la idea sería usar otro modelo parecido pero centrado en expresiones muy típicas de ese espacio ideológico: “adoctrinamiento ideológico”, “universidad tomada”, “recuperar valores”, etc. Y si quisiera distinguir cuándo eso roza la extrema derecha, por ejemplo, cuando aparecen símbolos, banderas preconstitucionales o mensajes mucho más agresivos, ahí sí tendría sentido tirar de un modelo zero-shot, que ayuda a detectar categorías más específicas sin tener que entrenar desde cero.

La movilización estudiantil sería más fácil de identificar: bastaría con clasificar tuits que hablan de protestas, convocatorias, vídeos del campus, pancartas, o mensajes de apoyo entre estudiantes. Un clasificador simple podría separar bien estos mensajes del resto del debate. Y, por último, el tema de los usuarios influyentes va por la vía del análisis de redes: una vez construida la red de retuits y menciones, solo hay que ver qué cuentas aparecen recibiendo más interacciones, quiénes conectan a más gente o quiénes actúan como punto de referencia dentro de cada grupo. Sin complicarse, básicamente ver quién mueve más conversación y a quién le hacen caso dentro de cada bloque.

Al final, todas estas variables independientes sirven para explicar por qué el debate acaba tan dividido. Si cada bloque tiene sus propios discursos (populistas, identitarios, estudiantiles), sus propios referentes y casi no se cruzan, es mucho más fácil entender cómo la conversación en torno a Vito Quiles termina tan marcada por la polarización.

Además de estudiar estas variables por separado, sería muy interesante cruzar los datos de contenido con la estructura de la red. El objetivo sería comprobar si los usuarios que ocupan posiciones centrales, es decir, los más influyentes, son precisamente aquellos que utilizan los discursos más radicales. Esto permitiría ver si en este caso la polarización no es solo una cuestión de opiniones, sino una estrategia eficaz para ganar relevancia y liderazgo dentro de cada grupo.

Asimismo, conviene analizar la evolución temporal del debate. Si se superponen los datos de publicación con los horarios de las visitas a los campus, es muy probable que aparezcan picos de actividad simultáneos. Este análisis confirmaría que la dinámica en X es puramente reactiva, donde lo que ocurre en el mundo físico se traslada y amplifica en la red de manera casi instantánea.

Para poder hacer todo el análisis anterior (clasificar tuits, detectar discursos, construir la red, ver si hay polarización, etc.), es necesario apoyarse en varias herramientas de ciencia de datos. Cada una cumple su función y permite resolver una parte distinta del trabajo. No hace falta usarlas todas a la vez en un estudio real, pero sí explicarlas para que quede claro cómo podría llevarse a cabo el análisis del debate. A continuación muestro la tabla indicando todas las herramientas que alguna ya he mencionado y otras que no he mencionado todavía pero sí utilizaría para este análisis sobre Vito Quiles.

Herramienta	Para qué se usaría	Cómo se aplicaría al caso de Vito Quiles
Python (pandas, scikit-learn, nltk, networkx)	Para limpiar datos, procesar texto, entrenar modelos y analizar redes.	pandas para ordenar los tuits, nltk para limpiar texto, scikit-learn para clasificar mensajes (populismo, derecha radical...); networkx para construir la red de retuits/menciones.
R (visualización)	Para hacer gráficos más claros y detallados.	Crear gráficos de actividad por días, comparar uso de palabras clave o visualizar diferencias entre grupos del debate.
Modelos de ML (Naive Bayes, SVM, redes neuronales)	Para clasificar texto según categorías (a favor/en contra, populismo, derecha radical).	Detectar qué discursos aparecen en el debate y cómo se distribuyen entre los dos bloques.
Zero-shot classification	Para clasificar tuits sin necesidad de entrenar un modelo desde cero.	Separar mensajes populistas, de derecha radical o de protesta estudiantil sin tener un dataset enorme.
SNA (detección de comunidades, centralidad, modularidad)	Para analizar cómo se organiza la red y ver si hay polarización estructural.	Ver si el debate se divide en dos grupos aislados, identificar qué cuentas son influyentes dentro de cada bloque.
Visualización (grafos, mapas de calor, etc.)	Para mostrar patrones de forma clara y visual.	Grafo del debate con las dos comunidades, mapa de calor de palabras típicas en cada bloque.

6 Ventajas, Limitaciones y consideraciones éticas.

A la hora de analizar un debate como el de Vito Quiles en X, usar ciencia de datos tiene cosas muy buenas, pero también algunos límites importantes. Además, al trabajar con mensajes de personas reales, siempre hay que tener en cuenta ciertas normas éticas básicas.

Ventajas

- Permite analizar miles de tuits en muy poco tiempo, algo imposible de hacer a mano.
- Los modelos ayudan a detectar patrones que no siempre se ven a simple vista (polarización, marcos populistas, discursos ideológicos...).
- El análisis de redes sociales muestra cómo se organizan los usuarios y quién influye más, dando una visión más completa que solo mirar el contenido.
- Se pueden combinar herramientas diferentes (ML, SNA, visualización), lo que da una imagen mucho más rica del debate.
- Es especialmente útil cuando un tema se mueve muy rápido, como ocurre en protestas o debates políticos en X.

Limitaciones

- X no representa a toda la sociedad; suele hablar gente más politizada o muy activa en debates públicos.
- La API tiene límites y puede no mostrar todos los tuits, lo que deja lagunas en el análisis.
- Los modelos pueden equivocarse y clasificar mal algunos mensajes, especialmente cuando el tono es irónico o ambiguo.
- Los datos pueden estar sesgados (por bots, cuentas muy activas, cámaras de eco...).
- La estructura de la red cambia muy rápido y un análisis puede quedarse “viejo” en pocas horas.

Consideraciones éticas(aunque ya las comenté antes en fuentes de datos pero lo pongo otra vez)

- Anonimizar siempre a los usuarios que no sean figuras públicas: nada de nombres, fotos o tuits literales de gente normal.
- No perseguir ni señalar perfiles individuales, solo analizar patrones generales.
- Respetar los términos de uso de la plataforma y no acceder a contenido privado.
- Dejar claro que los modelos no son perfectos y pueden cometer errores, así que las conclusiones deben tomarse con prudencia.

7 Análisis con ejemplos de tuits reales.

Para ilustrar cómo funcionaría el modelo de clasificación del trabajo, voy a analizar uno a uno varios tuits que he encontrado en X. La idea es simular lo que haría el modelo, ver si cada mensaje es **a favor**, **en contra** o **neutral**, y además detectar si aparece **populismo**, **derecha radical** o cualquier otro rasgo relevante para el análisis. Esto sirve como ejemplo práctico de cómo se aplicarían las variables del estudio (polarización, populismo, ideología, etc.) en un dataset que fuese real.

Filon de Alejandria 🇪🇸 · 26 oct
En respuesta a @jpurias



17 Me gusta 216 Retweets

Clasificación del modelo: A favor de Vito Quiles

Rasgos detectados: Derecha radical / lenguaje despectivo

Por qué:

Usa un insulto ideológico (“zurdo”), típico de la derecha radical. Polarización clara.



Ricardo
@Ricardo31917849

Seguir

Respondiendo a @joanmarcfons y a @jpurias

Cómo si no supiéramos todos, que la universidad ha sido tomada por el comunismo, desde hace décadas, y esa basura humana es la que determina quien puede o no ha hablar.

12:58 · 26 oct 25 · 184 Visualizaciones

8 Me gusta

17 Me gusta 216 Retweets

Clasificación del modelo: A favor de Vito Quiles

Rasgos detectados: Populismo + derecha radical

Por qué: Habla de “universidad tomada por el comunismo” típico marco populista de “instituciones secuestradas por una élite progresista”. Además, usa lenguaje agresivo (“basura humana”).



Javi94 🇪🇸 · 26 oct
@JaviDD1994

Seguir

En respuesta a @jpurias

En que momento ha incitado a la violencia??? No como estás verdad 🙄



13:41 · 26 oct 25 · 543 Visualizaciones

71 Me gusta

17 Me gusta 216 Retweets

Clasificación del modelo: A favor de Vito Quiles

Rasgos detectados: Derecha radical y antiizquierda.

Por qué:

Niega que haya violencia y saca una foto fuera de contexto sin hacer referencia al momento haciendo alusiones a Irene Montero, denostándola de violenta y diciendo que los violentos son “ellos”. Muy claro para la polarización.



Clasificación del modelo: A favor de Vito Quiles

Rasgos detectados: Populismo suave + derecha radical light

Por qué:

Divide entre “todos” (el pueblo) vs “comunistas” (la élite o grupo que supuestamente domina la uni). Un clásico.



Clasificación del modelo: A favor de Vito Quiles

Rasgos detectados: Derecha radical muy clara.

Por qué:

“El enemigo” es la universidad, se habla de “pocilga”, insulto directo.

El modelo lo marcaría como discurso ideológico duro.



Clasificación del modelo: En contra de Vito Quiles

Rasgos detectados: Crítica a la ultraderecha + señalamiento de símbolos franquistas.

Por qué:

Asocia a Vito Quiles con grupos falangistas, se considera totalmente del grupo contrario.

Nota: Ya empezamos a ver muestras de la extrema derecha que aunque hayamos dicho que eran minoritarias también pueden aparecer. En este caso es en contra pero mucha gente subió también vídeos con estas banderas estando a favor de ello.



Vito Quiles no ha llegado a pisar Somosaguas. Lo diremos las veces que haga falta: ni el machismo, ni el racismo ni los discursos reaccionarios no tienen cabida en nuestras aulas.



Clasificación del modelo: En contra de Vito Quiles

Rasgos detectados: Activismo estudiantil + marco antifascista.

Por qué:

Lenguaje de condena directa yo creo que muy claro para el modelo.



Clasificación del modelo: En contra de Vito Quiles

Rasgos detectados: Ataque personal + denuncia política.

Por qué:

Insultos, descalificación y crítica a su capacidad de convocatoria, claro bloque de anti Vito Quiles.

Después de ver todos los tuits, la polarización es clarísima. No aparecen casi mensajes intermedios: o defienden a Quiles con ideas típicas de la derecha radical (“adoctrinamiento”, “universidad tomada por la izquierda”, burlas tipo “lágrimas de zurdo”) o lo critican con etiquetas como “ultraderechista”, “fascista” o denunciando banderas franquistas. Son dos bloques totalmente enfrentados, cada uno con su propio lenguaje y sus propias narrativas. Aunque esto no sea un modelo ejecutado de verdad, ya deja bastante claro que, con datos así, un clasificador separaría muy rápido los dos polos y la red seguramente saldría partida en dos comunidades casi sin contacto entre ellas.

8 Tratamiento y exploración con datos reales.

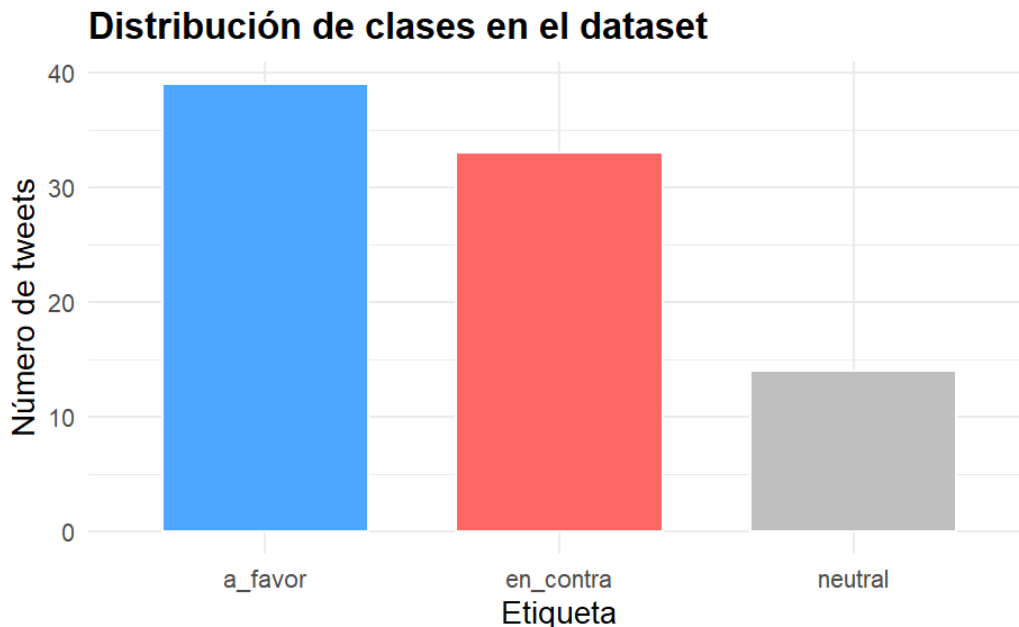
A partir de este punto dejo atrás la teoría y empiezo a trabajar con datos reales. En el planteamiento inicial definí un marco bastante completo: polarización como variable dependiente y, como independientes, populismo, derecha radical, movilización estudiantil y la estructura de la red. Sobre el papel queda muy bien, pero llevar todo eso a la práctica es bastante costoso. No solo porque no tenga acceso a la API de X (que también limita mucho), sino porque cada una de esas variables requeriría modelos específicos, anotación manual y un volumen de datos mucho mayor del que puedo manejar en este trabajo.

Por eso, para esta parte práctica he decidido centrarme en lo que realmente puedo abordar, la polarización. El objetivo es comprobar si, en un conjunto real de mensajes, aparecen dos bloques claramente diferenciados como los que describo en la teoría. Para hacerlo, he recopilado manualmente un dataset pequeño de 86 tweets relacionados con el caso de Vito Quiles. No es un conjunto enorme ni pretende ser representativo de toda la conversación, pero sí es suficiente para una primera aproximación.

La idea aquí es aplicar un análisis exploratorio sencillo y utilizar un modelo similar al que describí en la parte teórica, por ejemplo un clasificador SVM básico, para ver si los mensajes tienden a agruparse en dos posturas enfrentadas o si hay más matices de los que parece. A partir de esta sección, por tanto, paso a mostrar y comentar los resultados obtenidos con estos datos reales, centrándome exclusivamente en si existe o no polarización en el debate.

8.1 Carga de datos.

Antes de empezar a trabajar en R, primero recopilé manualmente los tweets y los fui clasificando en tres categorías: a favor, en contra y neutral. Cuando terminé esa parte, guardé todo en un archivo CSV para poder procesarlo mejor. A partir de ahí ya pasé a R, donde cargo el dataset y hago una visualización inicial para ver cuántos tweets hay de cada tipo y cómo queda distribuido el conjunto.

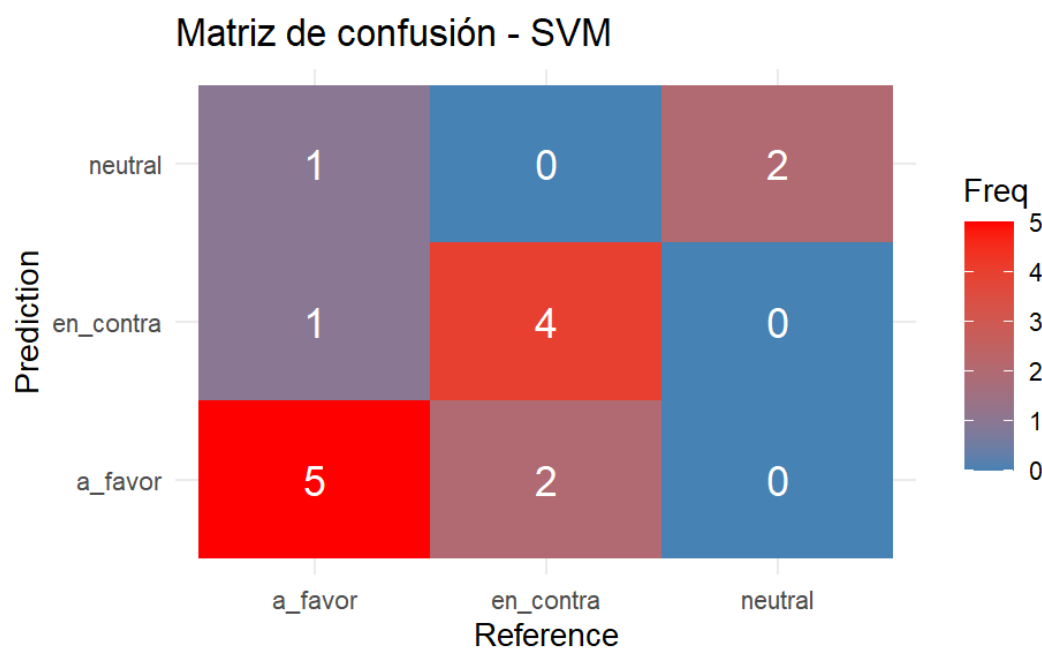


Este gráfico muestra claramente cómo se distribuyen las tres categorías del dataset. La clase más numerosa es la de tweets a favor de Vito Quiles, seguida bastante de cerca por los en contra. La categoría neutral es la menos frecuente, algo que encaja con lo esperado en un debate tan polarizado, la mayoría de los mensajes toman partido de una forma u otra. Aunque el conjunto es pequeño, esta primera visualización ya deja ver que las posiciones extremas dominan la conversación y que los mensajes neutros son minoría.

Antes de aplicar el modelo, es necesario preparar los datos. Primero divido el dataset en dos partes, una para entrenar el clasificador y otra para probarlo (train y test). Además, proceso el texto de los tweets para dejarlo en un formato adecuado: paso todo a minúsculas, elimino puntuación, números, espacios repetidos, stopwords en español e incluso restos de emojis o símbolos que no aportan contenido. Todo esto es importante porque el modelo solo puede aprender bien si el texto está limpio y homogenizado. Una vez hecho este preprocesamiento, ya puedo construir la matriz de términos y avanzar hacia el entrenamiento del clasificador SVM. Consultar todos estos pasos intermedios en el script.

8.2 Aplicación del modelo.

Después de preparar los datos, ya puedo ejecutar el modelo SVM para ver si es capaz de distinguir entre las tres categorías que marqué a mano. El objetivo aquí no es obtener una precisión espectacular, sino comprobar si, incluso con un dataset pequeño, el modelo es capaz de captar la polarización del debate. Una vez entrenado, obtengo la matriz de confusión que se muestra a continuación, donde se puede ver cuántos aciertos y fallos ha tenido en cada clase en test.



A partir de esta matriz, se observa que el modelo funciona razonablemente bien. Acerta la mayoría de los tweets a favor y en contra, que son justo las dos posiciones polarizadas del debate. También identifica correctamente los pocos tweets neutros, aunque esa categoría tiene menos datos. Lo importante es que el modelo no confunde masivamente las clases polarizadas entre sí: distingue bastante bien cuándo un mensaje es favorable a Vito Quiles y cuándo es crítico con él.

En conjunto, la precisión ronda el 73%, que para un dataset tan reducido es más que suficiente para responder a la pregunta principal: sí, el modelo detecta polarización. Es decir, la separación entre mensajes a favor y en contra es lo bastante clara como para que un algoritmo sencillo, como un SVM lineal, pueda aprenderla sin problemas. Esto confirma que incluso con pocos datos reales, la conversación sobre Vito Quiles se organiza en dos bloques bien diferenciados.

8.3 Nube de palabras.

Además he generado esta nube de palabras para ver de un vistazo qué términos dominan en el debate y qué tipo de lenguaje aparece en los tweets reales.



Lo que se observa es que las palabras más visibles están muy relacionadas con el contexto universitario y con la figura de Vito Quiles. También aparecen términos con una carga ideológica evidente, como “comunistas”, “censura”, “fascista” o “zurdo”, que muestran el tono confrontativo de la conversación. Junto a esto, se mencionan varias ciudades vinculadas a los eventos, lo que indica que el debate se extiende por distintos campus. En conjunto, el vocabulario refleja una discusión marcada por posiciones muy enfrentadas, algo que encaja con la idea de polarización que ya he observado y confirmado en este trabajo.

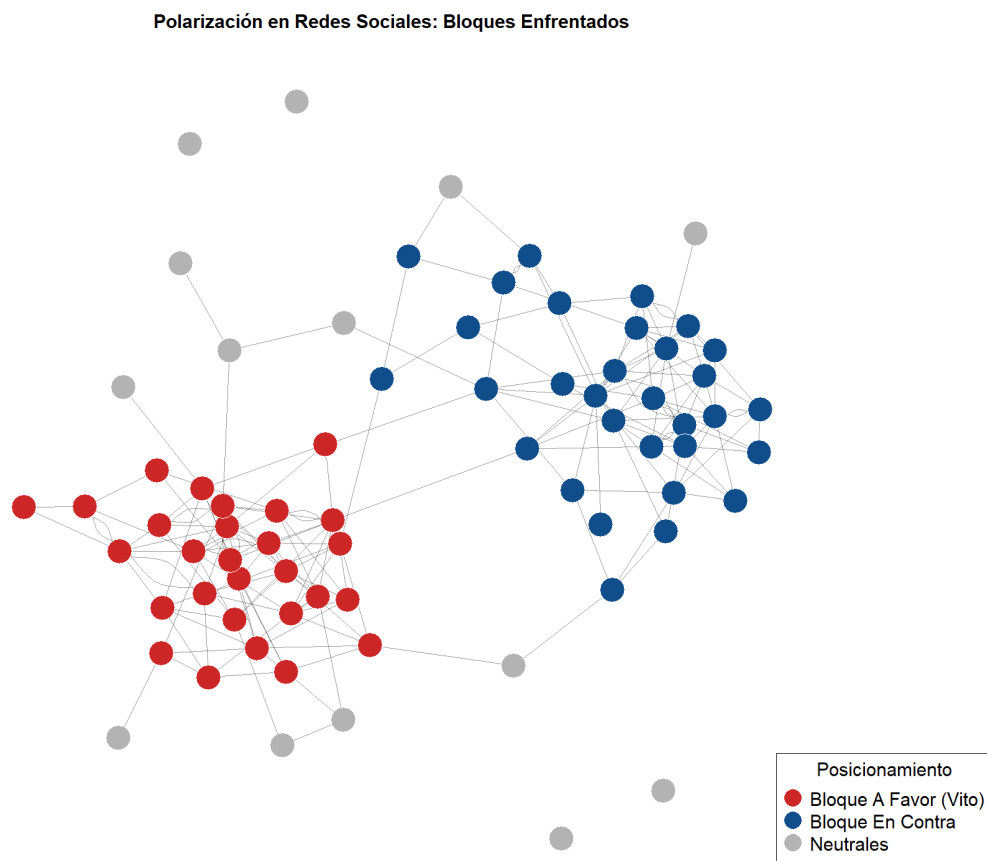
A favor

En contra



Y si observamos las nubes diferenciando los dos bandos, vemos que ambos utilizan un lenguaje muy marcado y casi no comparten vocabulario. En los tweets a favor aparecen términos elogiosos y ligados a marcos identitarios (España, libertad, verdad), mientras que en los tweets en contra predominan palabras críticas y despectivas (propaganda, ruido, show, provocar). En resumen, cada lado habla desde su propio marco y refuerza su posición, lo que vuelve a mostrar que el debate está claramente polarizado.

8.4 Gráfico SNA.



En este gráfico se aprecia que los usuarios del debate se organizan en dos bloques que quedan claramente diferenciados: por un lado, los tweets a favor de Vito Quiles (en rojo) y, por otro, los que van en contra (en azul). Dentro de cada bloque hay bastantes conexiones internas, lo que indica que los mensajes de un mismo grupo tienden a relacionarse entre sí. En cambio, entre los dos lados apenas aparecen enlaces, lo que refleja que las interacciones cruzadas son muy escasas. Los nodos grises, que representan posiciones más neutras, están más dispersos y no se integran del todo en ninguno de los dos bloques.

De todas formas, este resultado hay que tomarlo con cierta cautela, porque el dataset no viene de la API, sino que lo recopilé manualmente. Eso hace que falten menciones y retuits reales, lo cual puede influir en cómo queda dibujada la red y hacer que parezca más separada de lo que estaría con datos completos. Aun así, el gráfico sirve para hacerse una idea general de la división del debate y de cómo se diferencian las dos posiciones principales, mostrando la clara polarización que vengo observando a lo largo de todo el trabajo.

Por último, si tuviera que definir ‘polarización’ con una sola imagen, elegiría esta, que además pertenece a este mismo caso y cuadra exactamente con lo que he observado del gráfico con las dos comunidades:



“Dos bandos enfrentados, sin término medio.”

9 **Bibliografía.**

- Rodríguez Sáez, A. (2025). *Polarización* [Material docente]. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid.
- Rodríguez Sáez, A. (2025). *Populismo y Ciencia de Datos* [Material docente]. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid.
- Rodríguez Sáez, A. (2025). *Derecha radical* [Material docente]. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid.
- Rodríguez Sáez, A. (2025). *SNA I* [Material docente]. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid.
- Rodríguez Sáez, A. (2025). *Big Data y Ciencias Sociales* [Material docente]. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid.
- Rodríguez Sáez, A. (2025). *Movimientos sociales* [Material docente]. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid.
- Rodríguez Sáez, A. (2025). *Presentación: Tres P's, manada y rich-clubs* [Material docente]. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid.
- Rodríguez Sáez, A. (2025). *Centralidad* [Material docente]. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid.
- Lorenzo-Dus, N., & Bou-Franch, P. (2022). *Twitter and affective polarisation: Following political leaders in Spain* (Tesis de máster). Universitat Pompeu Fabra.
- Lisi, M., & Borg, S. (2023). *The use of social media in the populism of Chega and Vox*. Social Sciences, 12(5), 255.
- Pérez, L. (2021). *Populismo en redes sociales: Análisis del discurso del movimiento 15M, Podemos y VOX en Twitter (Tesis doctoral)*. Universidad Pontificia Comillas.
- Montero, J. R., & Torcal, M. (2018). *La polarización política en España: debates y perspectivas*. Centro de Investigaciones Sociológicas.