

Práctica 4: Caso Práctico de Análisis y Evaluación de Redes

Davide Jarik De Rosa - Erasmus student

DNI: CA91061DU

Email: davidejarik@correo.ugr.es

December 15, 2024

Contents

1 Pregunta de investigación planteada y conjunto de datos	3
1.1 Los ultimos 10 días de Elon Musk sobre X antes de las elecciones americanas	3
1.2 El conjunto de datos estructural	3
2 Estructura de la red	4
3 Las medidas principales	4
3.1 Tabla y resultados	4
3.2 Diferencias entre una diferente relación cercanía - enlaces	5
3.3 Los enlaces más fuertes en la red	6
4 Las propiedades de la red	6
4.1 Distribuciones de grados y Coeficientes de clustering	6
4.2 ¿Es una red libre de escala?	8
4.3 ¿Es una red de mundo pequeño?	9
5 Las medidas de Centralidad	9
5.1 Centralidad de Grado	9
5.2 Intermediación	9
5.3 Cercanía	10
5.4 Centralidad de Vector Propio	10
6 Descubrimiento de comunidades	11
6.1 El método de Lovaina	11
6.1.1 Comunidad 4 - El soporte a MSG	12
6.1.2 Comunidad 5 - El woke mind virus	12
6.1.3 Comunidad 6 - El soporte a X	13
6.1.4 Comunidad 7 - Kamala y Biden	13
6.1.5 Comunidad 8 - Vote vote vote	14
6.1.6 Comunidad 9 - La ardilla asesinada	14
6.1.7 Comunidad 10 - El soporte a Donald Trump	14
6.1.8 Comunidad 12 - Las empresas de Musk	15
6.2 El método de Girvan-Newman	16
6.3 El método de Leiden	16
6.3.1 La nueva comunidad de la ardilla	17
6.3.2 La nueva comunidad de Starlink	17
6.4 El algoritmo de Ravasz	18
7 Visualización de la red	19
8 Análisis de resultados final	22
8.1 Discusión de los resultados obtenidos	22
8.2 Comparación con los resultados existentes	22
A Las comunidades del algoritmo de Ravasz	25

1 Pregunta de investigación planteada y conjunto de datos

1.1 Los ultimos 10 días de Elon Musk sobre X antes de las elecciones americanas

"The world's richest man was central to 2024's presidential contest. Elon Musk has never been far from political commentary thanks to his business interests and public profile. But in 2024 he placed his considerable cultural and financial weight behind Donald Trump's successful campaign to become US President for a second time [2]."

Esto es lo que escribió Matthew Ward Agius el 8 noviembre 2024, para el periódico DW. El millonario puso casi 120 millones de euros en la campaña electoral de Donald Trump. ¿El dinero fue la única manera de influir los votantes? Elon Musk tiene un account X, su plataforma que compró en abril 2022, con más de 200 millones de followers [1] y más de 60K posts. Es claro que escribió mucho sobre Donald Trump y las elecciones en X, lo que podemos preguntar entonces es: "¿hasta qué nivel Elon Musk usó X para influir los votantes de los Estados Unidos antes de las elecciones americanas?"

El dominio concreto que escogí en esta práctica son los más de 200 posts que Elon Musk escribió con su account X 10 días antes de las elecciones americanas. La pregunta de investigación será:

Donald Trump y las elecciones estadounidenses protagonizan las actividades de Elon Musk sobre X antes del 5 de noviembre 2024 (the election day). También se pueden identificar clústeres semánticos centrados en el apoyo al ex-presidente y en el ataque a sus oponentes.

Es claro que para responder a esta pregunta necesitamos de una red semántica.

1.2 El conjunto de datos estructural

Obtener un conjunto de datos estructural que puede ayudar a responder a la pregunta planteada no es difícil, será importante pero el pre-processing que faremos con los datos antes de construir la red. El modelo que vamos a usar para construir una red semántica es lo mismo de [12], aquí listamos el proceso paso a paso:

- Con la búsqueda avanzada en X, miramos y copiamos en un fichero .txt los posts de Elon Musk desde el 26 de Octubre hasta el 4 de Noviembre (el día antes del election day).
- Con la librería re [14] de python, tomamos cada post y ponemos las palabras en una lista, sin espacios o puntuación.
- Con la librería nltk [18], lematizamos las palabras, poniendo los verbos en la forma infinito y los nombres al masculino singular.
- Sustituimos y fusionamos palabras. Esto es uno step un poquito raro, que en este caso tenemos que hacer: tendremos como siempre en redes semanticas muchas palabras que son una expresión junta, también muchas palabras diferentes que se refieren a lo mismo concepto. La lista en el detalle se puede buscar en el fichero JUPITER "support_code_BT" de la documentación. Aquí podemos poner ejemplos de sustituciones: "harris" : "kamala" (la persona es la misma, usamos el apellido no el nombre), "trump" : "donaldtrump" (una vez se refiere al account, una vez al apellido), 'dem' !: 'dems' (estamos hablando de la misma cosa); y también de fusiones: "new york times" : "newyorktimes", "swing voter" : "swingvoter", "citizen journalism" : "citizenjournalism", "swing state" : "swingstate", "democratic party" : "dems", "nov 5th" : "nov5th".
- Eliminamos palabras vacías de contenido, por ejemplo: "the", "to", "a", "they", "just", "that", "for", "how".
- Eliminamos verbos auxiliares o modales, por ejemplo: "have", "should", "can", "could".
- Al final, filtramos todas las listas vacías o con una sola palabras, porque no darán ninguna contribución a la red.

2 Estructura de la red

Ahora que el conjunto de datos está listo, tenemos que construir la red. En nuestra red, como en casi todas las redes semánticas, los nodos serán las palabras y tendremos enlaces entre palabras (nodos) que serán cercanas en los posts. La estructura será ponderada y no dirigida (no vamos a tener una relación "esta palabra punta esta palabra": la red expresará solo una relación de cercanía entre palabras). Tenemos entonces que elegir:

- Cual será el nivel de poda de los nodos;
- Cual será la relación cercanía - enlaces de las palabras;
- Cual será el peso de los enlaces;
- Cual será el nivel de poda de los enlaces.

Para el nivel de poda de los nodos, tenemos más de 2000 palabras diferentes en el conjunto. No son muchísimos nodos, pero no tiene mucho sentido transformar todas las palabras en nodos. Entonces así como se hace en el caso [12], ponemos un umbral: podamos las palabras con una sola ocurrencia. Así el número final de nodos que tendremos en nuestra red será 305.

Para la relación cercanía - enlaces la pregunta será ¿Cuánto cercanas tendrán que ser dos palabras en un post para estar conectadas con un enlace? Estamos hablando de un conjunto de datos creado sobre posts, entonces no necesitamos de conectar muchas palabras. En [12], que será aún aquí nuestra referencia, vamos a poner enlaces entre palabras que serán cercanas de dos en dos. Ejemplo: en la lista "*dem*s lie *wi*n *ele*ction" las solas palabras que no tendrán enlaces entre ellas serán "*dem*s" y "*ele*ction". Una otra posibilidad sería tener enlaces solo entre palabras directamente a continuación: no vamos a usar esta opción, pero es interesante para entender como puede variar la estructura de una red semántica respecto de esta elección. También en nuestra red si tendremos la misma palabra dos veces en los grupos de tres, vamos a poner un auto-enlace. Entonces la red final tendrá 1002 enlaces.

El peso de los enlaces es muy importante en las redes semánticas, tendremos un análisis solo para los enlaces más fuertes de la red. Tenemos muchas opciones: podemos por ejemplo proporcionar el peso de los enlaces a la longitud de un post, donde un post corto es una frase más directa y eficaz la conexión entre palabras tiene que ser más fuerte. Vamos pero a escoger el método más sencillo de todos: ponemos un enlace de peso 1 cada vez que dos palabras serán cercanas y si encontramos una otra vez la misma relación de cercanía, aumentamos de 1 el peso del enlace.

En esta red no necesitamos de podar enlaces: 1002 enlaces no son muchos y la densidad de la red como veremos será baja.

3 Las medidas principales

3.1 Tabla y resultados

En la tabla 1 podemos mirar los valores de la red en las medidas principales. Una nota: nuestra red es ponderada, entonces tenemos que mirar los valores ponderados, ponemos los valores de la red equivalente no ponderada para comparar los resultados. En general, las redes semánticas tienen una estructura de mundo pequeño, caracterizada por una conectividad escasa, una longitud media corta de los caminos entre las palabras y un fuerte clustering local. Además, las distribuciones del número de conexiones siguen leyes de potencia indican un patrón de conectividad scale-free [11].

Vamos a hablar en la próxima sección de las propiedades de redes libres de escala y de mundo pequeño, lo que podemos mirar hasta ahora es que $L \ll L_{max}$ y que la red es escasa: la densidad es muy baja aunque no usemos ningún algoritmo de poda para los enlaces. Entonces el grado medio es bastante bajo también, observamos pero que el grado medio de nuestra red (la ponderada) es más grande de casi el 50% de la red no ponderada: esto indica que la estructura de la red que escogimos hace su efecto y que tenemos muchas palabras que se conectan más de una vez. El diámetro y la distancia media también son muy muy pequeños en respecto del tamaño de la red (con la distancia media ponderada que es 24, 24% más grande de la no ponderada). Sabemos que las distancias en una red aleatoria son varios órdenes de magnitud menores que el tamaño de la red, esta similitud con nuestra red (las redes reales en general) es la razón por la que la distancia media de nuestra red no ponderada

Medida	Valor
Número de nodos N	305
Número de enlaces L	1002
Número máximo de enlaces L_{max}	46360
Densidad D	0.022
Grado medio $\langle k \rangle$	6.57
Grado medio ponderado $\langle k_P \rangle$	9.725
Diámetro d_{max}	6
Diámetro ponderado d_{Pmax}	6
Distancia media $\langle d \rangle$	3.206
Distancia media ponderada $\langle d_P \rangle$	3.983
Distancia media red aleatoria equivalente $\langle d_{aleatoria} \rangle$	3.039
Distancia media red aleatoria ponderada equivalente $\langle d_{Paleatoria} \rangle$	2.515
Coeficiente de clustering medio $\langle C \rangle$	0.309
Coeficiente de clustering medio red aleatoria equivalente $\langle C_{aleatoria} \rangle$	0.0215
Coeficiente de clustering medio red aleatoria ponderada equivalente $\langle C_{Paleatoria} \rangle$	0.0319
Número componentes conexas	3
Número nodos componente gigante N_{cg}	301
Porcentaje nodos componente gigante % N_{cg}	98.69 %
Número aristas componente gigante L_{cg}	997
Porcentaje aristas componente gigante % L_{cg}	99.5 %

Table 1: Las medida principales de la red

y la distancia media de la red aleatoria equivalente son bastante similares. La distancia media en la red aleatoria ponderada equivalente es mucho más baja de nuestra red ponderada. El coeficiente de clustering medio es muy alto, muy más alto de lo que tenemos en la red aleatoria equivalente: el coeficiente de clustering de un nodo de una red aleatoria es pequeño y en nuestra red conectamos las 2 palabras cercanas, entonces es claro que el coeficiente de clustering será alto. Al final, tenemos 3 componentes conexas y una componente gigante que tiene la mayoría de los nodos de la red. Casi todas las palabras, aunque si están en posts diferentes, serán conectadas en la red, esto es muy bueno para el análisis que haremos. Podemos mirar en Gephi los nodos de las 2 componentes conexas aisladas de la componente gigante:

- Nodos : inflation
- Nodos : porsche, 911, cybertruck

Que serán los únicos posts de Elon Musk aislados de la discusión principal.

3.2 Diferencias entre una diferente relación cercanía - enlaces

En esta parte, podemos mirar que pasa si en lugar de escoger una estructura de la red donde las palabras son conectadas dos a dos, usamos una conexión entre palabras que son directamente siguientes en los posts.

- Los nodos serán los mismos.
- Los enlaces serán 729 y entonces la densidad será más baja (0.016).
- La distancia media y el diametro serán casi los mismos (3.62 y 8).
- El grado medio y el grado medio ponderado serán más bajos de consecuencia (4.78 y 5.521), también la diferencia entre el primero y el segundo no será así grande como en nuestra red.
- La diferencia más grande será en el coeficiente de clustering medio, que será de solo 0.094, un valor muy cercano de lo que tenemos en la red aleatoria equivalente. Esto es muy interesante: el coeficiente de clustering de las redes semánticas será muy dependiente de la manera de conectar las palabras y entonces de la estructura escogida para la red.

3.3 Los enlaces más fuertes en la red

Aunque tenemos hasta ahora solo las medidas principales de la red, podemos sin embargo empezar a analizar los enlaces más fuertes en la red, para ver las palabras que se dicen con más frecuencia juntas en los posts de Elon Musk.

Ordenamos los enlaces para el peso en Gephi y vemos los resultados en Figura 1.

Source	Target	Type	Weight
24	1	Undirected	8.0
1	1	Undirected	7.0
45	15	Undirected	6.0
4	12	Undirected	6.0
23	9	Undirected	6.0
2	1	Undirected	5.0
8	113	Undirected	5.0
3	1	Undirected	5.0
29	1	Undirected	5.0

Figure 1: Los 9 enlaces más fuertes en la red

Que listamos aquí en orden descendente:

1. **”early” - ”vote”**, muy interesante. La cosa de la que Musk estarías hablando más antes del 5 de noviembre serían los resultados del voto anticipado
2. **”vote” - ”vote”**, incitar directamente a la gente a votar.
3. **”blue” - ”state”**, hablar de los estados de los democraticos.
4. **”legacymedia” - ”lie”**, muy interesante también. Elon Musk dice que los media tradicionales de comunicación mienten, que son cosas que Trump dijo muchisimo en los ultimos 4 años [6].
5. **”kamala” - ”biden”**, una conexión entre el ex presidente de los Estados Unidos y el actual candidato dems.
6. **”vote” - ”donaldtrump”**, incitar directamente a la gente a votar Donald Trump.
7. **”election” - ”interference”**, puede ser o una referencia a las elecciones actuales, en las que podría haber interferencias ilegales de los dems, o una referencia al asalto a Capitol Hill en enero 2020 [9].
8. **”vote” - ”dems”**, aquí podría estar señalando los efectos negativos de votar a los dems.
9. **”everyone” - ”vote”**, incitar de nuevo la gente a votar.

Haremos un análisis mejor cuando hablaremos de las medidas de centralidad, es claro, pero ahora también que las elecciones son muy centrales en los posts del multimillonario.

4 Las propiedades de la red

4.1 Distribuciones de grados y Coeficientes de clustering

En Figura 2 podemos mirar la distribución (normalizada) de los grados ponderados y no ponderados. Tenemos una distribución que es común en las redes reales: empezamos con valores muy bajos que suben hasta el valor más grande y después bajan de manera casi monótona con una ley de potencia. La distribución normalizada del grado no ponderado tiene menos oscilaciones, tiene la probabilidad más alta por el grado 4 (cerca del 0.2) y tendrá muchos hubs de grado cerca del 40. La distribución normalizada del grado ponderado a diferencia de la primera tiene muchas oscilaciones, tiene la probabilidad más alta por el grado 6 (cerca del 0.14, es también más uniforme en las probabilidades) y tendrá 3 hubs que serán muchos más fuertes que los demás cerca del 75.

En Figura 3 se puede mirar la distribución de los coeficientes de clustering, lo que vemos es coherente con el coeficiente de clustering medio.

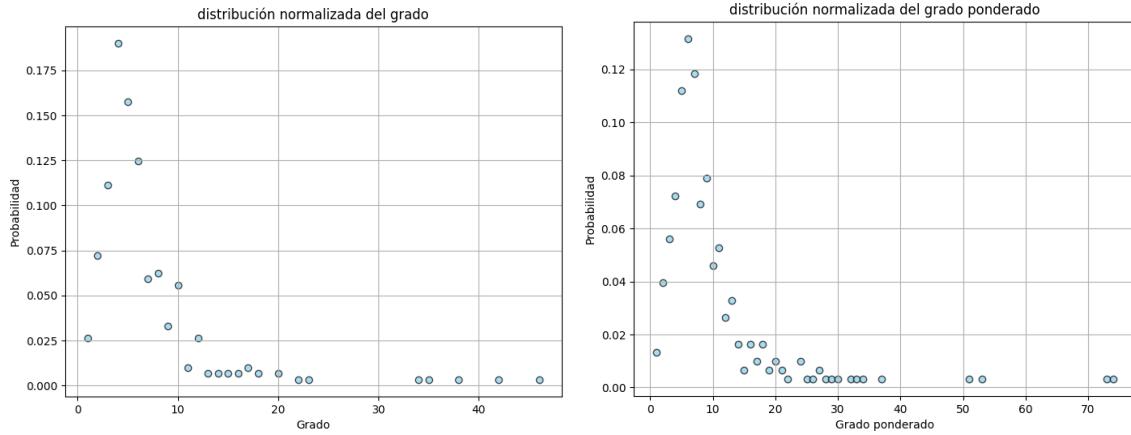


Figure 2: Distribución de grados ponderados y no ponderados

Clustering Coefficient Distribution

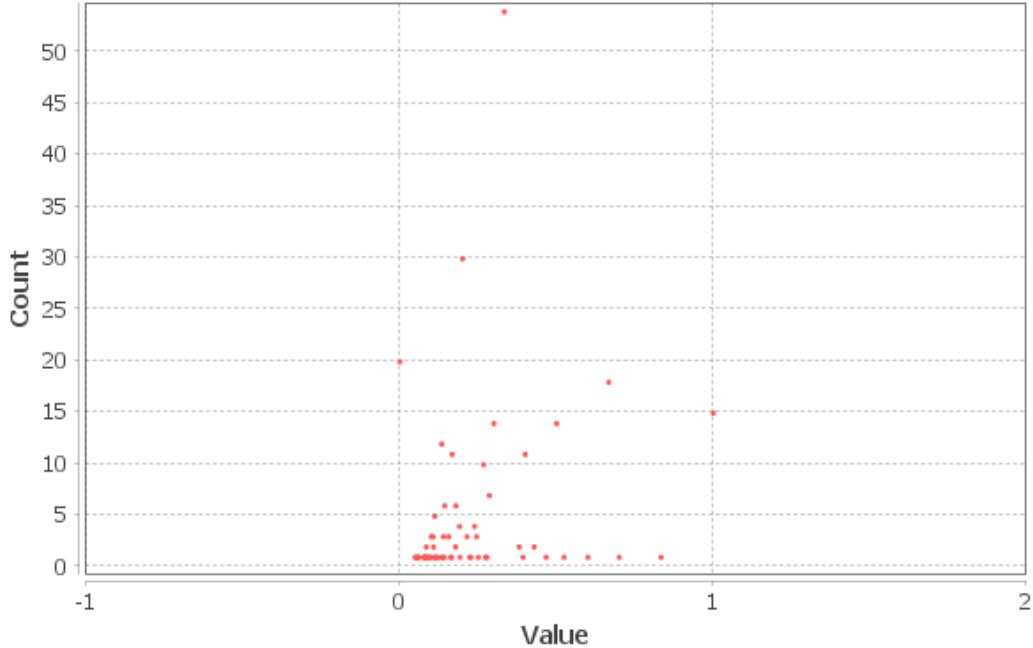


Figure 3: Distribución de los coeficientes de clustering

En Figura 4 podemos mirar la relación entre grado y coeficiente de clustering en el caso de grado ponderado y no ponderado. En el caso del grado ponderado tendremos más oscilaciones, pero el trend es lo mismo en los dos casos: los nodos con un grado más bajo tienen un coeficiente de clustering más alto y viceversa. Esta es una consecuencia de la jerarquía de redes. La razón es que, como será claro en la siguiente sección, nuestra red tiene hubs, palabras con un grado medio mucho más alto que los otros que conecten posts diferentes. Es claro entonces que si Elon Musk habla de Donald Trump tanto en un post donde patrocina un debate con Joe Rogan como en un post donde critica al New York Times, será improbable que Joe Rogan y el New York Times están conectados.

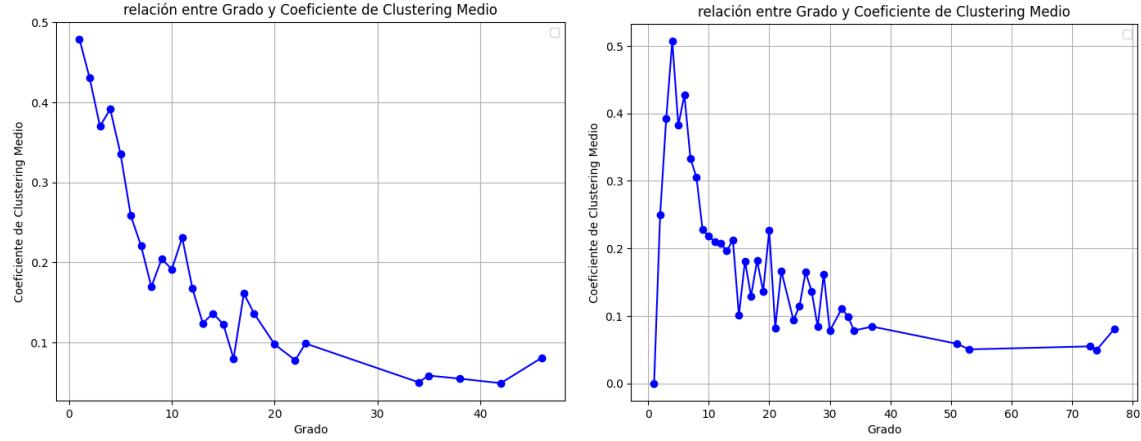


Figure 4: Relacione entre grado (ponderado y no ponderado) y coeficiente de clustering

4.2 ¿Es una red libre de escala?

Casi todas las redes semánticas son redes libres de escala y esta no es una excepción. Las distribuciones de los grados observadas sigue la llamada Ley de la Potencia, también tenemos muchísimos hubs en nuestra red, una cosa que es muy frecuente en las redes libres de escala [20]. Además, con el grado ponderado, $k_{max} = 77$ y $k_{min} = 1$, la diferencia es muy grande, como en las redes libres de escala.

La distribución normalizada de grado tiene una ley $P(k) = k^{-\gamma}$: podemos estimar el coeficiente γ . Lo haremos en Python con la library `scipy` [8] con 3 métodos diferentes: Ordinary Least Squares, Differential Evolution y Powell's Method. El resultado será lo mismo:

$$\gamma \approx 1.82$$

$$\gamma_{ponderado} \approx 1.68$$

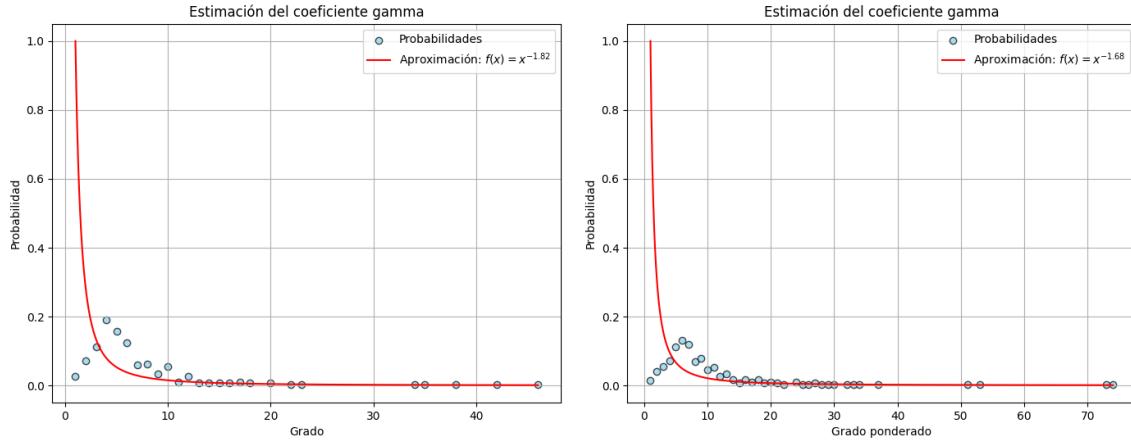


Figure 5: Estimación del coeficiente γ

En Figura 5, podemos mirar el resultado de la estimación. En general, tenemos una red demasiado pequeña para estimar correctamente el coeficiente (si no hubiera considerado los primeros nodos que son "irregulares", tendríamos un coeficiente aún más bajo - 1.54). El valor es más bajo de 2: estamos en el régimen crítico. Entonces la fórmula

$$k_{max} = k_{min} \cdot N^{\frac{1}{\gamma-1}} \approx 1956$$

devuelve un valor sin sentido.

4.3 ¿Es una red de mundo pequeño?

Casi todas las redes semánticas son redes de mundo pequeño y esta no es una excepción. Sabemos que tenemos un mundo pequeño si:

- $\langle d \rangle \approx \langle d_{aleatoria} \rangle$, y la diferencia es solo del 5.59% en el caso de grado no ponderado;
- : $\langle C \rangle > > \langle C_{aleatoria} \rangle$, y el coeficiente de clustering medio de nuestra red es 865.63% más grande que en la red aleatoria equivalente (con pesos).

Entonces las 2 condiciones son válidas: es una red de mundo pequeño.

Después nuestro análisis, podemos decir que el coeficiente de clustering se comporta como una red libre de escala: he dicho antes que tenemos muchos hubs y que hay diferencias muy significativas entre k_{min} y k_{max} . El tamaño de los hubs depende de N: cuanto mayor es el tamaño de la red, mayor es el grado del hub más grande (si tomamos más posts, tendremos más palabras conectadas con los actores principales).

5 Las medidas de Centralidad

Ahora haremos un análisis de redes sociales básico usando medidas de centralidad. Los haremos con las medidas de grado, intermediación, cercanía y vector propio.

5.1 Centralidad de Grado

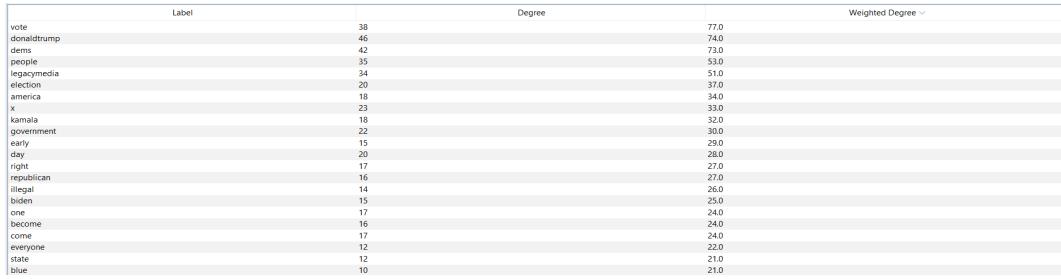


Figure 6: Los actores más centrales para centralidad de grado

¿Las palabras con más conexiones serán las palabras más importantes de la red? En verdad sabemos que la centralidad del grado no tiene en cuenta la estructura global de la red y que no captura las “corredurías”. Sin embargo, tiene sentido hacer un primero análisis con el grado ponderado y no ponderado. Sacamos los primeros 22 nodos con grado ponderado más alto de 20, el resultado es en Figura 6. Tenemos 3 palabras que son las más fuertes de la red: ”vote”, ”donaldtrump” y ”dems”; que tienen un valor de centralidad de vector propio más alto que 70. Despues de dos palabras de grado 53 y 51: ”people”, ”legacymedia”. La siguiente palabra tiene grado 37 y despues de ella el grado desciende regularmente. Podemos hablar de las primeras 3 palabras como los actores principales de la red y la cuarta y la quinta como los actores secundarios. Los otros actores no son así relevantes como los primeros aquí.

Los actores principales, entonces, se refieren todos directamente a las elecciones americanas; por otra parte, los actores secundarios no se refieren directamente a las elecciones, pero puede ser que existe una conexión relevante con el tema también. En las otras 17 palabras, 8 se refieren directamente a las elecciones.

5.2 Intermediación

Los actores con una intermediación alta suelen ocupar una posición que les permite trabajar como interfaces entre subgrupos de actores fuertemente unidos. La intermediación es quizás la medida más importante de esta sección, porque nos dirá que palabras hacen de puente, conectando los distintos posts de Elon Musk. Las palabras con una intermediación más alta son en Figura 7.

Label		Weighted Degree	Betweenness Centrality
donaldtrump	74.0	6693.723768	
dems	73.0	5503.837118	
vote	77.0	4742.161205	
people	53.0	4506.490852	
legacymedia	51.0	3991.978426	
election	37.0	2389.194795	
government	30.0	2267.432663	
x	33.0	2171.138839	
america	34.0	1960.736115	
come	24.0	1880.000001	
become	24.0	1781.125943	
one	24.0	1545.336291	
day	28.0	1544.439917	
kamala	32.0	1279.000000	
time	19.0	1245.660453	
because	20.0	1181.393581	
republican	27.0	1165.669639	
right	27.0	1133.000000	
illegal	26.0	1007.841911	
big	18.0	903.622184	
starlink	16.0	879.094105	
criminal	16.0	850.096852	
biden	25.0	843.056526	
early	29.0	821.1974489	

Figure 7: Los actores más centrales para intermediación

Es claro que Donald Trump es el puente más importante de todos los posts de Elon Musk: su nivel de intermediación es mucho más alto de todos los otros. Vemos también cómo la palabra "dems" es mucho más central ahora que "vote", es el segundo puente más fuerte. "vote", "people" y "legacymedia" son aún puentes fuertes. Vemos también palabras que no tienen un grado muy alto, que pero son importantes para la intermediación: "big", "starlink" y "criminal". "Criminal" puede ser como Musk juzga a los dems o al gobierno actual, "starlink" [21] son los satélites de SpaceX [16], la agencia espacial del miliardario. Observamos también que "X" es muy central tanto aquí como en el grado.

5.3 Cercanía

Label		Weighted Degree	Closeness Centrality
conche	5.0	1.0	
911	7.0	1.0	
cybertruck	2.0	0.453192	
donaldtrump	74.0	0.441176	
vote	77.0	0.429799	
dems	73.0	0.416089	
people	53.0	0.415133	
legacymedia	51.0	0.398936	
x	33.0	0.397878	
one	24.0	0.397875	
election	37.0	0.391134	
america	34.0	0.38961	
day	28.0	0.388066	
government	30.0	0.3861	
because	20.0	0.3861	
time	19.0	0.385109	
right	27.0	0.384076	
kamala	32.0	0.376452	
become	24.0	0.375	
come	24.0	0.373152	
republican	27.0	0.373134	
most	13.0	0.371287	
big	18.0	0.371287	
velo	11.0	0.370377	
everyone	22.0	0.369914	
msg	11.0	0.369914	
future	10.0	0.369458	

Figure 8: Los actores más centrales para cercanía

La cercanía es la importancia a "estar en medio de la red", es una medida normalizada. En Figura 8 podemos mirar las palabras con un nivel de cercanía más alto. Los primeros 3 nodos tienen un valor de cercanía de 1, el máximo. Estas 3 palabras no tienen interés en nuestro análisis, son una componente conexa que no tiene enlaces con la componente gigante. El valor es 1 porque los 3 nodos tienen 3 enlaces entre ellos.

Las 5 palabras con un valor de cercanía más alto en la componente gigante son siempre los mismos: "donaldtrump", "vote", "dems", "people", "legacymedia". Aparte de estas, observamos valores muy altos en "x", "election". "Kamala" es mucho más central aquí que "biden". "msg", que se refiere a Madison Square Garden (donde Trump tenía un mitin electoral el 27 de octubre, que el New York Times llamó "A Closing Carnival of Grievances, Misogyny and Racism" [15]), es una palabra con un grado muy bajo, que pero es bastante central aquí.

5.4 Centralidad de Vector Propio

La centralidad de vector propio se basa en que la centralidad de un nodo concreto depende de cuánto centrales sean sus vecinos. En una evolución de la centralidad de grado que tiene en cuenta los vecinos. En Figura 9 podemos mirar los actores más importantes para esta medida.

Los resultados ahora son mucho diferentes que lo que tuvimos con el grado: los actores principales son siempre 3: "donaldtrump", "vote" y "dems"; con Donald Trump que es entonces el protagonista de los posts de Elon Musk, con un valor de 1. Para los actores secundarios es diferente, porque "legacymedia" y "people" son siempre los segundos más fuertes, pero vemos 3 palabras que destacan antes de que el valor caiga uniformemente con el tiempo: "X", "one" y "day". Con "X" que se confirma como muy central. Tenemos muchas palabras de grado muy bajo que son muy importantes con esta

Label		Weighted Degree	Eigenvector Centrality
donaldtrump		74.0	1.0
vote		77.0	0.952098
dem		73.0	0.911699
legacymedia		51.0	0.662419
people		53.0	0.688817
X		33.0	0.518613
one		24.0	0.493737
day		28.0	0.450664
election		37.0	0.411114
government		30.0	0.391637
right		27.0	0.37588
republican		27.0	0.368017
easy		23.0	0.360205
because		20.0	0.351959
kamala		32.0	0.350852
america		34.0	0.348917
time		19.0	0.332194
everyone		22.0	0.32952
win		11.0	0.312237
propaganda		13.0	0.310163
pennsylvania		16.0	0.306633
hoax		16.0	0.301874
squirrel		13.0	0.299463
future		13.0	0.294936
nov5th		13.0	0.294936

Figure 9: Los actores más centrales para centralidad de vector propio

medida. "future", "nov5th", "squirrel", "propaganda", "win". Aquí 12/25 palabras son directamente conexas con las elecciones y como veremos en la próxima sección la mayoría de las otras serán conexas indirectamente.

Para concluir nuestro análisis podemos decir que los actores más centrales son:

- Donald Trump, el protagonista;
- Vote, un personaje principal;
- Los Democratics, un personaje principal.

Los actores importantes (los personajes secundarios) son:

- Legacy Media;
- People;
- X, que es muy alto tanto en intermediación como en centralidad de vector propio.

6 Descubrimiento de comunidades

Si miramos a la Figura 4 es claro que nuestra red tiene una estructura modular: Los hubs hacen de puente entre distintos grupos; entonces tendremos comunidades. En esta sección, usaremos 4 algoritmos para estudiar las comunidades de nuestra red.

6.1 El método de Lovaina

El algoritmo de Louvain es uno de los algoritmos más populares y eficientes para detectar comunidades en grafos. Se basa en la idea de maximizar el número de enlaces dentro de una comunidad y minimizar el número de enlaces entre comunidades (optimización de modularidad). Su funcionamiento se divide en dos fases: una fase de agrupamiento y una fase de refinamiento [22]. Tiene un tiempo de ejecución de $O(N \cdot \log N)$ y las comunidades resultantes tienden a ser compactas y bien definidas.

En Gephi tenemos que escoger si queremos randomizar o usar pesos (haremos así) y el parámetro más importante es la resolución. Un valor de resolución alto produce comunidades más pequeñas y específicas, mientras que un valor bajo produce comunidades más grandes y generales.

En la Tabla 2 tenemos los resultados de varias ejecuciones del algoritmo sobre nuestra red con diferentes valores de resolución. Para cada valor de resolución hicimos 5 ejecuciones y la modularidad y el número de comunidades resultantes son la media de los valores obtenidos en cada ejecución. Podemos observar los siguientes resultados:

- Con una resolución más alta tendremos menos comunidades;
- El umbral de resolución debajo del cual la modularidad es menor que 0.3 es 0.03;
- El umbral de resolución por encima del cual la modularidad es menor que 0.3 es 2.4;
- El valor de resolución mejor es 0.9;

Resolución	Modularidad	Número de comunidades
0.01	0.236	127
0.03	0.298	105
0.1	0.393	64
0.4	0.469	29
0.7	0.487	19.8
0.9	0.493	15.2
1	0.492	13.8
1.2	0.49	12
1.5	0.481	9.6
2	0.425	6.6
2.4	0.303	4.4
5	0.008	3

Table 2: Los valores de modularidad y el número de comunidades

- El resultado mejor de modularidad es de 0.496, con 15 comunidades.

Por pura curiosidad, si intentamos de ejecutar el algoritmo sin opciones "randomize" y "use weights", los resultados empeorarían. Una otra cosa interesante que podemos analizar es que la modularidad con valores de resolución bajos es baja, mejora con regularidad hasta un punto de máxima y con valores más altos de resolución empeorará.

Vamos ahora a analizar las comunidades encontradas con el método de Lovaina (resolución: 0.9, modularidad: 0.496, 15 comunidades). Desafortunadamente los datos no son muchísimos y la red no es muy grande, entonces no todas 15 comunidades tendrán un sentido real: vamos a analizar solo las que podemos interpretar bien. En las siguientes tablas, los nodos serán ordenados para centralidad de vector propio, entonces los primeros nodos serán los más centrales.

6.1.1 Comunidad 4 - El soporte a MSG



Figure 10: La comunidad 4

Ya hablamos del mitin electoral de Donald Trump a Madison Square Garden como una de las palabras con cercanía más alta. En esta comunidad tenemos un soporte directo al mitin de Trump: "public", "matter", "massive", "see", "goal", "trust" son las palabras de este cluster que se refieren a "MSG". Podemos suponer que Musk en algunos de sus posts está invitando a los votantes a ver el mitin, alabándolo.

6.1.2 Comunidad 5 - El woke mind virus

Elon Musk no utiliza palabras muy dulces para atacar la cultura woke [19]. Aquí podemos mirar a una comunidad que es un ataque contra el "woke mind virus". Tenemos palabras que atacan el virus como: "defeat", "crush"; pero también palabras que advierten del peligro: "save", "life", "extremely", "real".



Figure 11: La comunidad 5

6.1.3 Comunidad 6 - El soporte a X

Esta comunidad es un soporte a la plataforma de Musk X, la interpretación puede aparecer sencilla: "send X link most friend". Es interesante pero mirar que tenemos también palabras como: "Google", "youtube", "fight", "die", "hard". Es como si el multimillonario que ensalza X quisiera desacreditar a las otras plataformas.



Figure 12: La comunidad 6

6.1.4 Comunidad 7 - Kamala y Biden

Tenemos muchas informaciones importantes en esta comunidad. La primera es que Kamala y Biden están en la misma comunidad: en los posts de Musk las palabras que se refieren a ellos son las mismas. Podemos observar palabras que se refieren directamente a ellos: "puppet", "cruel"; o a la campaña electoral / presidencia: "illegal", "administation", "legalize", "immigration", "competent", "program", "sue", "mess". Es claro que Musk quiere desacreditar al primero y al actual candidato demócrata.



Figure 13: La comunidad 7

6.1.5 Comunidad 8 - Vote vote vote

Esta comunidad es una de las más grandes. Este cluster es una otra prueba de la cantidad de posts que Musk hizo sobre las elecciones: habla del voto, de las elecciones del 2020 y del 2024, de los swingstates, del early vote y sus resultados, etc...



Figure 14: La comunidad 8

6.1.6 Comunidad 9 - La ardilla asesinada

Esta comunidad puede ser imposible de interpretar si no se conoce la historia de "Peanut the Squirrel" [17], una pobre ardilla que fue suprimida debido a una prueba positiva para la rabia. Esta historia se ha convertido en una forma de criticar a la administración demócrata en Pennsylvania por parte de los medios e influencers pro-Trump. Aquí tenemos una comunidad dedicada a una ardilla: "kill peanut squirrel", "pet", "house", "corruption".



Figure 15: La comunidad 9

6.1.7 Comunidad 10 - El soporte a Donald Trump

Aquí tenemos una de las comunidades más grandes de la red: es una comunidad de soporte a Donald Trump. Aquí patrocina una entrevista con Joe Rogan: "joe rogan", "tonight", "full", "conversation", "overseas". Se patrocina también un documental sobre el expresidente: "documentary", "watch", "worth". Musk ensalzando su inminente victoria: "win", "believe", "push". Y además critica a los medios de comunicación tradicionales y los críticos: "legacymedia lie", "propaganda", "hoax", "fraud", "crazier", "fraud".



Figure 16: La comunidad 10

6.1.8 Comunidad 12 - Las empresas de Musk

Esta comunidad se compone de las empresas de Musk y de las palabras de elogio para ellas. Las empresas son: "Starlink", "SpaceX", "Neuralink"; las palabras de elogio son: "big", "great", "high number", "humanity", "turn". Se llama también "Mars", "money" y "military", que es bastante raro si pensamos que son empresas de proyectos científicos.



Figure 17: La comunidad 12

6.2 El método de Girvan-Newman

El método de Girvan-Newman es una técnica de clustering jerárquico divisiva. En el algoritmo se calcula el valor de intermediación de todos los enlaces de la red y se elimina el enlace de mayor intermediación, se repite esto hasta que no haya enlaces que eliminar en la red o hasta que el valor máximo de intermediación sea menor que un umbral.

En Gephi podemos escoger 4 parámetros con el algoritmo de Girvan-Newman, que darán lugar a modularidad y tiempos diferentes: probando todas las combinaciones el range de tiempos es [3.802 s, 4.564 s] y el range de modularidad es [0.402, 0.404]. Los resultados son entonces bastante cercanos, lo que es interesante es que la modularidad con el método de Girvan-Newman es mucho más baja de la que tuvimos con Lovaina. La configuración con el mejor resultado de modularidad es esta:

- Respect edge type for shortest path betweeness: YES
- Respect parallel edges for shortest path betweeness: YES
- Respect edge type for modularity computation: YES
- Respect parallel edges for modularity computation: YES
- Processing time: 3.924 s
- Number of communities: 37
- Maximum found modularity: 0.4043

Si tenemos que comparar Girvan-Newman con Louvain, Louvain es mejor para manejar grandes conjuntos de datos. El uso del proceso de maximización de la modularidad de Girvan-Newman presenta inconvenientes, ya que puede dar lugar a que se detecten diferentes comunidades en cada ejecución o incluso a que comunidades enteras queden fuera debido a una división temprana de los enlaces [10].

Lo que pasa con el método de Girvan-Newman en nuestra red es que detecta demasiadas comunidades: la interpretación de las comunidades es imposible. Tenemos una comunidad con más del 23% de los enlaces que tiene los nodos más centrales y las otras comunidades no tienen sentido. Entonces, en nuestro caso, el método de Lovaina es mejor: es más rápido, más eficiente, la modularidad es más alta y la interpretación de las comunidades obtenidas para el máximo nivel de modularidad es excelente.

6.3 El método de Leiden

El método de Leiden es una modificación reciente del de Lovaina que se basa en la corrección de un fallo de Lovaina (donde se pueden generar comunidades erróneas en las que se incluyan nodos que no tienen enlaces entre sí). El algoritmo de Leiden es más rápido y puede dividir grupos en lugar de solo fusionarlos. Se supone que garantiza el “óptimo” si se ejecuta varias veces.

En Gephi, usamos el método de Leiden con los siguientes parámetros y variamos la resolución:

- Quality function: Modularity
- Use edge weights: YES
- Number of iterations: 100
- Number of restarts: 6
- Random seed: RANDOM

En la Tabla 3 tenemos los resultados de varias ejecuciones del algoritmo sobre nuestra red con diferentes valores de resolución. Para cada valor de resolución hicimos 5 ejecuciones y la modularidad y el número de comunidades resultantes son la media de los valores obtenidos en cada ejecución. Podemos observar los siguientes resultados:

- Con una resolución más alta tendremos más comunidades (el contrario de Lovaina);
- La modularidad es inversamente proporcional a la resolución;

- La modularidad converge a 1 si la resolución es cerca de cero;
- El umbral de resolución por encima del cual la modularidad es menor que 0.3 es 6;
- No tenemos un valor de resolución mejor: por los valores más altos tenemos una resolución así baja que la división en comunidades no tiene sentido.

Resolución	Modularidad	Número de comunidades
0.01	0.99	3
0.1	0.9	3
0.5	0.594	5.8
1	0.51	14
2	0.439	25.8
6	0.303	49.8
10	0.221	66.6

Table 3: Los valores de modularidad y el número de comunidades

El comportamiento raro de la modularidad depende de la formulación matemática del algoritmo de Leiden en Gephi. En nuestra red vamos a usar una resolución de 1.1 para ver si Leiden toma comunidades o conexiones que Lovaina no tenía. Las comunidades generadas serán 15 como en el primero caso.

6.3.1 La nueva comunidad de la ardilla

La comunidad es siempre la misma, estamos hablando de nuevo de Peanut the Squirrel. Esta vez, respecto al método de Lovaina, el cluster es más coherente con el topic: con el cluster de Lovaina tuvimos palabras como "Catherine Herridge", que no tenía ninguna conexión con la ardilla.



Figure 18: La comunidad de la ardilla

En este caso, el método de Leiden se revela más eficaz que el método de Lovaine, que incluyó nodos que no tenían enlaces entre sí en el mismo cluster.

6.3.2 La nueva comunidad de Starlink

Con el método de Lovaina "Starlink" estaba en la misma comunidad de "military" y no era claro porque. Con el método de Leiden tenemos un nuevo cluster que incluye con "Starlink" también "russia", "ukraine", "military", "emergency", "provide". Ahora es claro que tenemos una conexión entre Starlink y la guerra en Ucrania, si buscamos en el perfil de X de Elon Musk "Starlink" y "Ukraine" encontramos este post:

"A major factor for why Ukraine was NOT overrun by Russia is the Starlink support I provided, at great risk to SpaceX cyber and physical attack by Russian military forces. Starlink is the BACKBONE of Ukrainian military communications at the front lines, because everything else has been blown up or jammed by Russia. In the words of Zelensky and Vice PM of Ukraine confirms that I am an ally" [13]

El algoritmo de Leiden puso Starlink en un cluster donde se habla de la guerra entre Rusia y Ucrania, resaltando una conexión semántica que no tuvimos antes. En estos dos casos resulta más eficaz que el algoritmo de Lovaina.



Figure 19: La comunidad de la Starlink

6.4 El algoritmo de Ravasz

El algoritmo de Ravasz es un algoritmo de clustering aglomerativo, donde los pasos principales son:

1. Definición de la matriz de similitud;
2. Decisión del mecanismo de agregación de los grupos: unión de grupos individual, completa o promedio;
3. Aplicación del clustering jerárquico aglomerativo;
4. Construcción del dendrograma

Después la ejecución, calculamos el valor de modularidad para cada partición posible y seleccionamos la que maximice el valor de la función.

El algoritmo de Ravasz se puede implementar en Python con las library `scipy` y `community` [7]. Con la library `community` las opciones de agregación de los grupos no son 3 sino 4: existe también la agregación "ward". La agregación ward minimiza la varianza dentro de los conglomerados fusionándolos de forma iterativa en función de su contribución a una objective function que mide la varianza total dentro de los clusters.

En Figura 20 - 23 podemos mirar los 4 dendrogramas generados, el resultado para cada método es el siguiente:

- La modularidad con el método individual es 0.5016;
- La modularidad con el método completo es 0.5008;
- La modularidad con el método promedio es 0.5032 - la mejor;
- La modularidad con el método ward es 0.5017.

Las comunidades identificadas con este método (el promedio) son 14 y la lista se puede consultar en el Apéndice A. La subdivisión es muy similar a la realizada por Leiden, con comunidades más grandes que también esta vez capturan enlaces semánticos que anteriormente se habían escapado (ejemplo: ahora la ardilla está en la misma comunidad que Biden y Harris). A pesar de esto, la subdivisión de conglomerados encontrada para Leiden es superior y más fácil de interpretar en esta tarea.

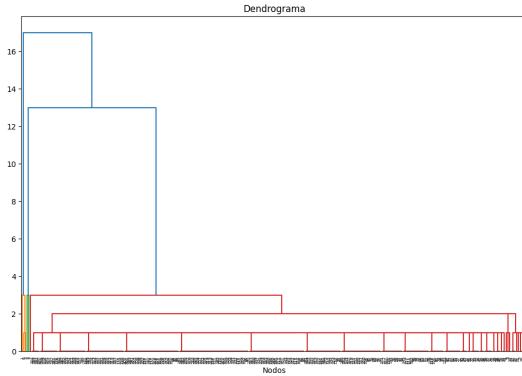


Figure 20: Dendrograma - método individual

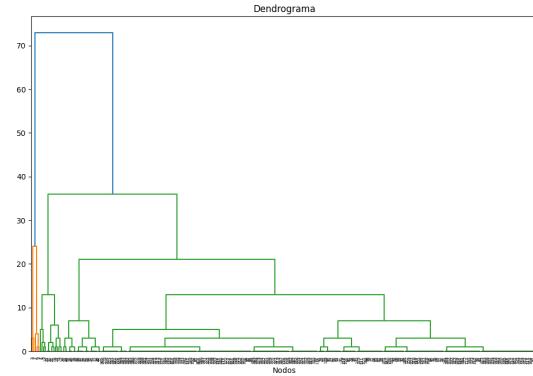


Figure 21: Dendrograma - método completo

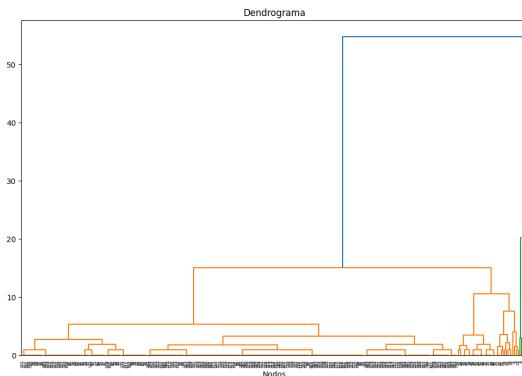


Figure 22: Dendrograma - método promedio

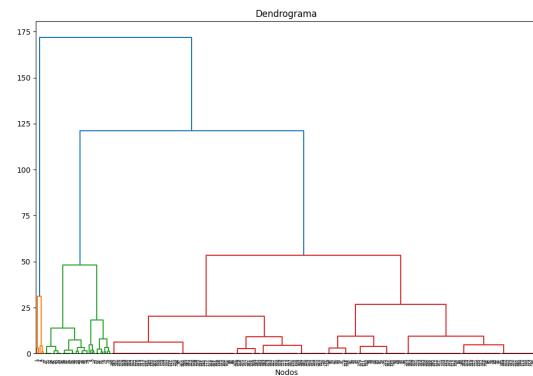


Figure 23: Dendrograma - método ward

7 Visualización de la red

Sabemos que cuando tenemos una densidad cerca de 0.02, no tenemos que usar algoritmos de poda de enlaces para tener una buena visualización. Tenemos entonces 4 visualizaciones diferentes para esta red:

1. En Figura 24, visualizamos los nodos con el color de la comunidad del algoritmo de Lovaina. El tamaño de los nodos es proporcional a su centralidad de vector propio y los enlaces tienen un color más fuerte si el peso es alto. El algoritmo de visualización de Gephi es Force Atlas 2 (que implementa el Kamada Kawai). El objetivo de esta primera visualización es resaltar la estructura modular de la red validando la subdivisión en comunidades que hemos estudiado. El resultado es muy bueno.
2. En Figura 25, visualizamos los nodos con el color de la comunidad del algoritmo de GN. Como he dicho antes, el resultado de este algoritmo no es muy bueno en esta red. Entonces esta vez el objetivo es diferente: resaltar la única comunidad dominante dentro de la subdivisión mostrando sus nodos (que son los más fuertes de la red). Usamos siempre la proporción entre tamaño de los nodos y valor de centralidad de vector propio (esta vez más fuerte) y color de los enlaces y su peso. Usé el algoritmo OpenOrd y después el Noverlap para evitar nodos superpuestos.
3. En Figura 26, visualizamos los nodos con el color de la comunidad del algoritmo de Leiden. La configuración y el objetivo son los mismos que la visualización obtenida con el método Lovaina (el algoritmo es siempre Force Atlas 2). Esta vez, sin embargo, al cambiar la configuración de color intentamos resaltar más los enlaces más fuertes.
4. En Figura 27, ya no visualizamos las comunidades sino que resaltamos los nodos más centrales de la red utilizando el algoritmo de visualización Fruchterman Reingold. MÁS grande es el tamaño de los nodos, más alto será el valor de centralidad de vector propio.

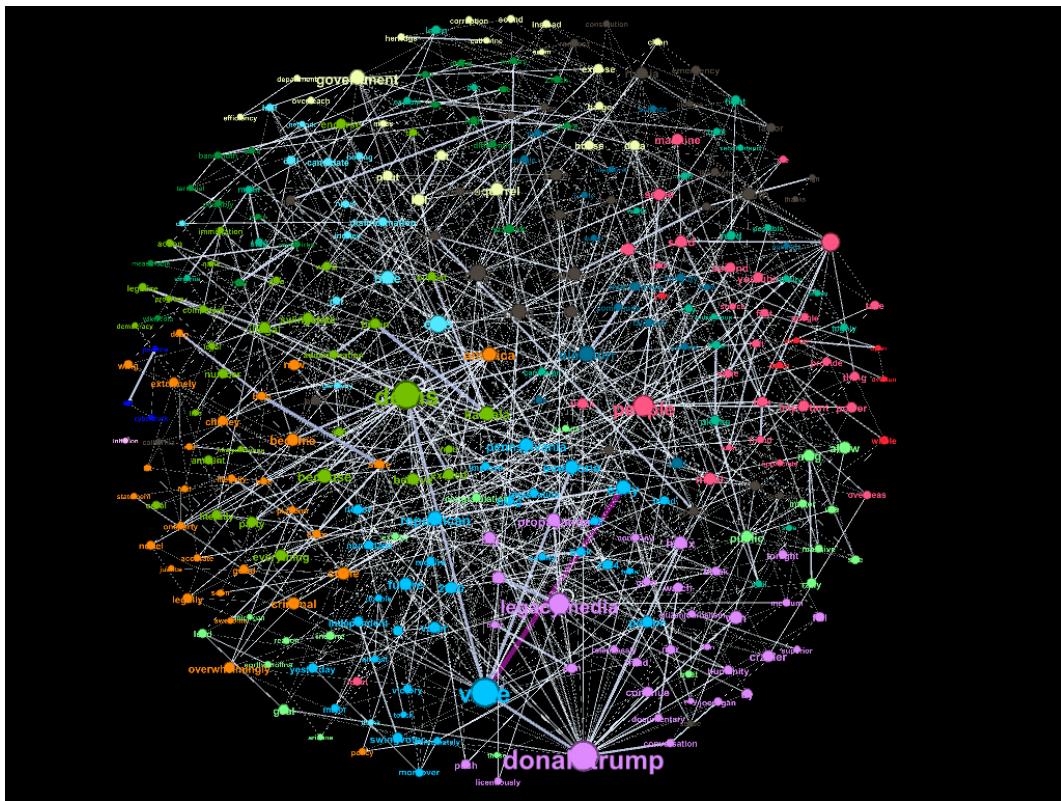


Figure 24: Las comunidades de Lovaina

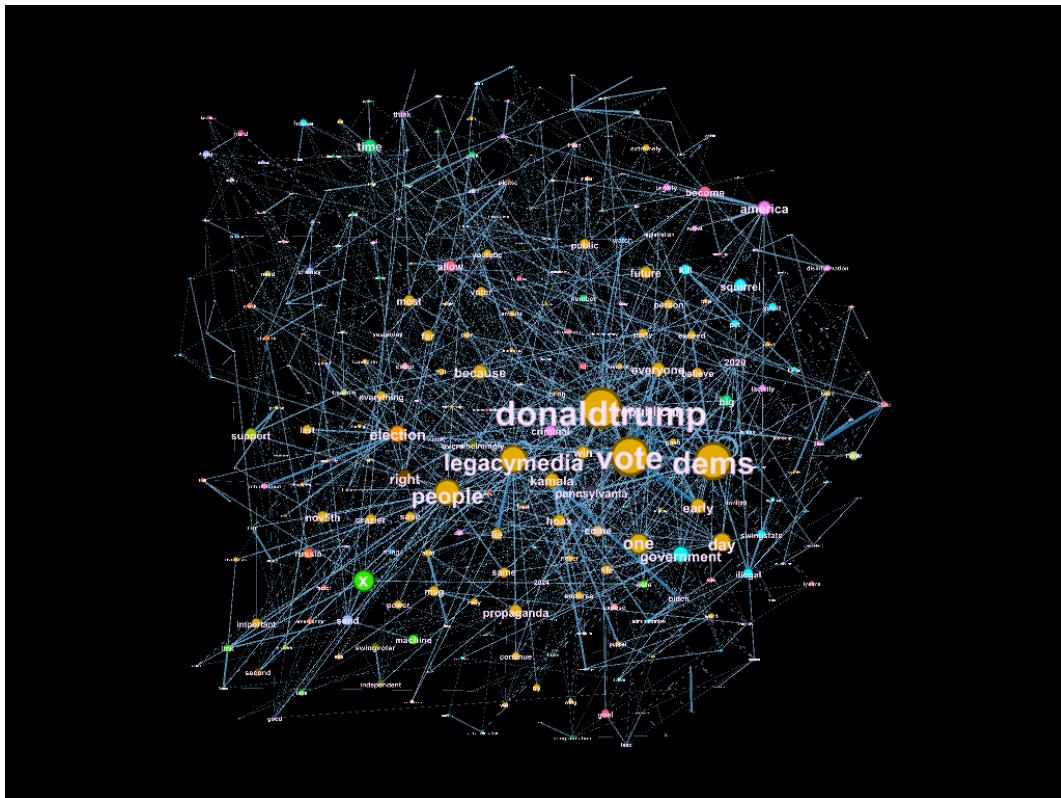


Figure 25: Las comunidades de GN

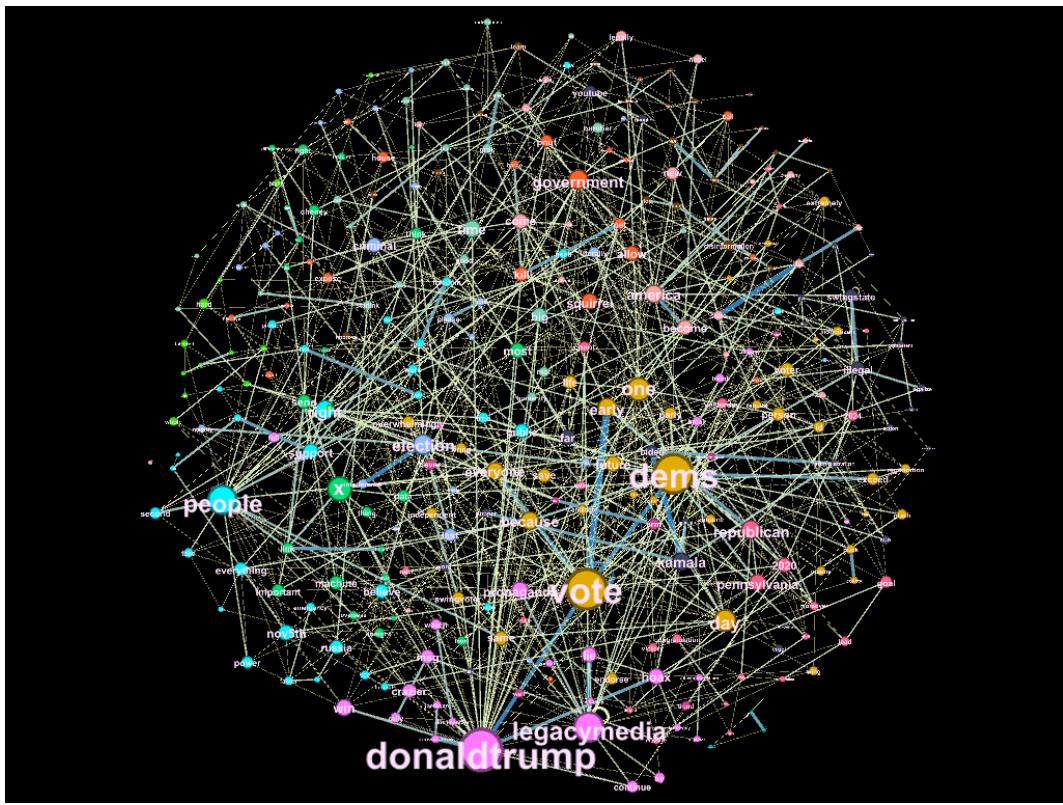


Figure 26: Las comunidades de Leiden

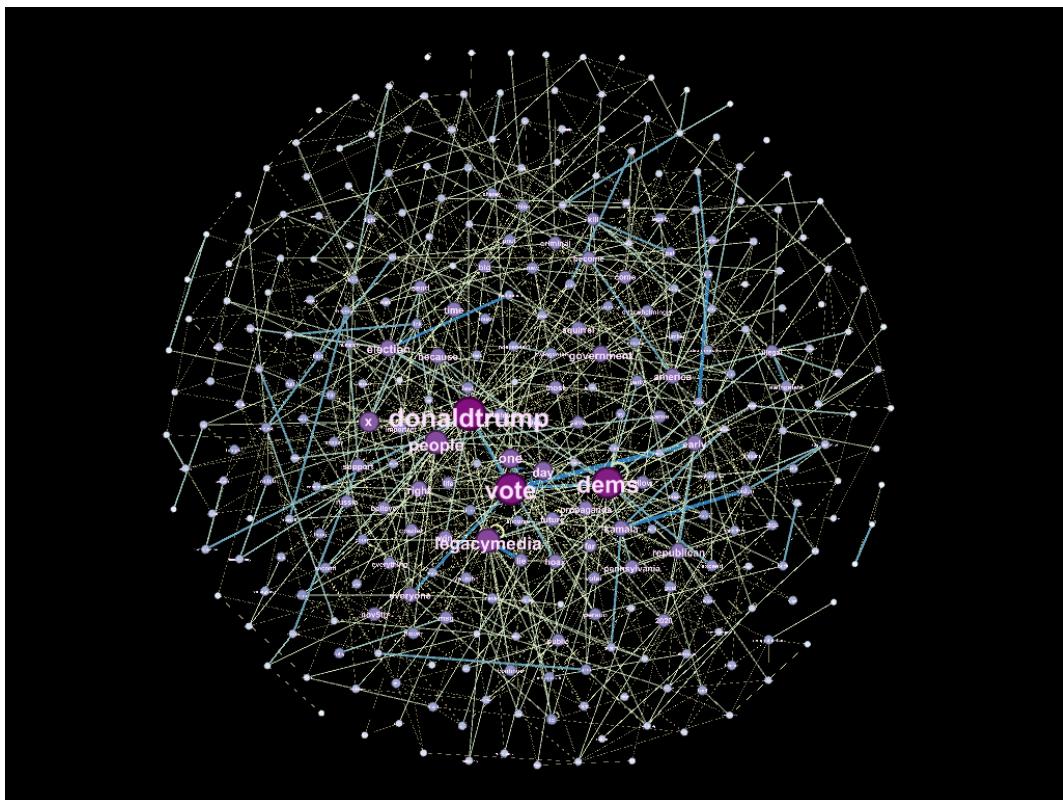


Figure 27: La red con el m odo FR

Con la visualización son aún más obvias las propiedades estructurales más significativas de la red: los enlaces tienen un peso diferente, hay nodos mucho más centrales que otros, la red tiene una estructura muy modular! Esto es claro porque la modularidad es alta en todos los algoritmos de comunidades y cuando se ejecutan algoritmos como Force Atlas 2 o Openord, la red se divide en estas comunidades. Tenemos muchos hubs que conectan nodos de comunidad diferentes como en todas las redes libres de escala y con la visualización de Openord de Figura 25 es claro que la red es una red de mundo pequeño: las distancias son muy bajas.

8 Análisis de resultados final

8.1 Discusión de los resultados obtenidos

En referencia a la pregunta planteada al principio de este análisis

Donald Trump y las elecciones estadounidenses protagonizan las actividades de Elon Musk sobre X antes del 5 de noviembre 2024 (the election day). También se pueden identificar clústeres semánticos centrados en el apoyo al ex-presidente y en el ataque a sus oponentes.

y considerando los resultados obtenidos, especialmente en el análisis de centralidad y comunidades, podemos concluir que

En los últimos 10 días antes del 5 de noviembre 2024 (the election day), el protagonista de las actividades sobre X de Elon Musk fue Donald Trump. Los principales temas tratados son la incitación a la población a votar por el candidato republicano, las críticas a los demócratas y a los "medios tradicionales" y al patrocinio de su propia empresa X

La mayoría de clústeres semánticos son centrados en el ataque a los dems y a sus administraciones, al gobierno actual y a sus ideas políticas. Luego encontramos grupos de apoyo al presidente, patrocinio de sus mítines e incitación al voto.

Los resultados obtenidos son excelentes, no solo pudimos responder a la pregunta de investigación inicial, sino incluso ampliarla. El análisis de centralidad reveló al protagonista (Trump) y a los principales actores de la red semántica, destacando también cómo algunas palabras, aparentemente poco frecuentes, son en cambio puentes de conexión entre diferentes posts. El análisis de las comunidades nos reveló entonces que hay más grupos semánticos centrados en atacar a los oponentes de Trump que en apoyarlo. Está claro hasta qué punto Elon Musk centra su discurso en el ataque a la cultura woke (directamente y de otra manera) y a los medios tradicionales, a los que prefiere el "Citizen Journalism" [3]. Poder encontrar comunidades como la de "Peanut Squirrel" o que conectan "Starlink" directamente con la guerra en Ucrania es realmente destacable.

8.2 Comparación con los resultados existentes

Nuestra red, como muchas redes semánticas, tiene una estructura de mundo pequeño, una conectividad escasa, una distancia media corta y un fuerte clustering local. Es una red libre de escala con el coeficiente $\gamma = 1.82$. ¿Cómo funciona en comparación con otras redes libres de escala?

En [4], donde las redes libres de escalas tienen un $\gamma > 2$, asintóticamente la distribución de grados del modelo Scale Free es independiente de tiempo (y, posteriormente, independiente del tamaño del sistema), lo que indica que a pesar de su continuo crecimiento, la red alcanza un estado estacionario Scale Free. Será lo mismo para nuestra red: independientemente del número de nodos, tendremos siempre hubs (verbos y palabras más frecuentes etc...) y la distribución de grado tendrá siempre la misma forma. En el mismo artículo vemos que el rasgo característico de las redes reales son la coexistencia de clusters y la corta longitud de los caminos. En nuestra red es lo mismo: el coeficiente de clustering es muy alto y los caminos muy cortos. "There is no analytical prediction for the SF model for the Clustering Coefficient": encontramos que el coeficiente de clustering de las redes libres de escala es aproximadamente 5 veces mayor que la de la aleatoria (en nuestro caso mucho más mayor). Sin embargo, el coeficiente de agrupamiento del modelo Scale Free disminuye con el tamaño de la red siguiendo básicamente una ley de potencia. En nuestra red no es lo mismo: el coeficiente de clustering,

	γ	K_{min}	P-VALUE	PERCENT	γ	k_{set}	k_{cut}	P-VALUE
Internet	3.42	72	0.13	0.6%	3.55	8	8500	0.00
WWW (IN)	2.00	1	0.00	100%	1.97	0	660	0.00
WWW (OUT)	2.31	7	0.00	15%	2.82	8	8500	0.00
Power Grid	4.00	5	0.00	12%	8.56	19	14	0.00
Mobile Phone Calls (in)	4.69	9	0.34	2.6%	6.95	15	10	0.00
Mobile Phone Calls (out)	5.01	11	0.77	1.7%	7.23	15	10	0.00
Email-Pre (in)	3.43	88	0.11	0.2%	2.27	0	8500	0.00
Email-Pre (out)	2.03	3	0.00	1.2%	2.55	0	8500	0.00
Science Collaboration	3.35	25	0.0001	5.4%	1.50	17	12	0.00
Actor Network	2.12	54	0.00	33%	-	-	-	0.00
Citation Network (in)	2.79	51	0.00	3.0%	3.03	12	5691	0.69
Citation Network (out)	4.00	19	0.00	14%	-0.16	5	10	0.00

Figure 28: Diferentes redes con su exponente γ

como en los mundos pequeños, será independiente del tamaño de la red (esto sucede estructuralmente si pensamos en cómo configuramos la red).

Cómo se dice en [12], las redes semánticas tienen $\gamma \approx 3$, que es mucho más grande que el nuestro. Si tenemos que caracterizar la red en función del valor del exponente γ , en Figura 28 podemos mirar las redes analizadas en [5]. La única red que tiene un exponente más bajo de la red de Elon Musk es "Science Collaboration" (el segundo): es una red muy diferente, la densidad en redes científicas es mucho más alta y el criterio de cercanía entre dos nodos no es lo mismo. La inconsistencia con los dos resultados mencionados anteriormente se debe probablemente al conjunto de datos excesivamente limitado que tenemos disponible, significativamente pequeño para una red semántica.

References

- [1] Elon Musk’s account on X. <https://x.com/elonmusk>, accessed on 13/12/2024.
- [2] Matthew Ward Agius. Elon musk: Why he supported trump. *DW*, 8/11/2024. <https://www.dw.com/en/elon-musk-why-he-supported-trump/a-70724719>.
- [3] Sonny Albarado. Citizen journalism. *Encyclopedia Britannica*, 22/12/2024. <https://www.britannica.com/topic/citizen-journalism>.
- [4] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, January 2002.
- [5] Albert-Laszlo Barabasi. *Network Science*. Cambridge University Press, 2016.
- [6] Reporters Without Borders. Usa: Trump verbally attacked the media more than 100 times in run-up to election. *Reporters Without Borders*, 25/10/2024. <https://rsf.org/en/usa-trump-verbally-attacked-media-more-100-times-run-election>.
- [7] community API. <https://python-louvain.readthedocs.io/en/latest/api.html>.
- [8] SciPy documentation. <https://docs.scipy.org/doc/scipy/>.
- [9] note =https://en.wikipedia.org/wiki/January_6_United_States_Capitol_attack, accessed on 14/12/2024 January 6 United States Capitol attack.
- [10] London Kasper. Girvan-newman vs. louvain for community detection. *Medium*, Apr 14, 2022. <https://medium.com/smucs/girvan-newman-v-louvain-for-community-detection-33988baab55b>.
- [11] Joshua B. Tenenbaum Mark Steyvers. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science* 29 (2005) 41–78, 2005.
- [12] Eric Pak Levi Oyster Boyd Ching Ashika Mulagada Martin Cenek, Rowan Bulkow. Semantic network analysis pipeline—interactive text mining framework for exploration of semantic flows in large corpus of text. *MDPI*, 2019.
- [13] Elon Musk’s post on the war in Ukraine. <https://x.com/elonmusk/status/1852417130924216744>.
- [14] re—Regular expression operations. <https://docs.python.org/3/library/re.html>.
- [15] Michael Gold Shane Goldmacher, Maggie Haberman. Trump at the garden: A closing carnival of grievances, misogyny and racism. *New York Times*, 27/10/2024.
- [16] SpaceX. <https://it.wikipedia.org/wiki/SpaceX>, accessed on 14/12/2024.
- [17] The Economic Times. Was ‘peanut the squirrel’ killed without reason? test results reveal shocking answers. *The Economic Times*, 13/11/2024. <https://economictimes.indiatimes.com/news/international/global-trends/was-peanut-the-squirrel-killed-without-reason-test-results-reveal-shocking-answers/articleshow/115249210.cms?from=mdr>.
- [18] Natural Language Toolkit. <https://www.nltk.org>.
- [19] Elon Musk tries to describe what a “woke mind virus” is during an interview with HBO. <https://www.instagram.com/msnbc/reel/CryYgDkMAtS/>.
- [20] Yang Wang, Yanqing Hu, Zengru Di, and Ying Fan. The effect of hub nodes on the community structure in scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 390(21):4027–4033, 2011.
- [21] Starlink website. <https://www.starlink.com/>, accessed on 14/12/2024.
- [22] Fernando Carazo y Joaquín Amat. Detección de comunidades en grafos y redes con python. *cienciadedatos.net*, 2023.

A Las comunidades del algoritmo de Ravasz

- Comunidad 1: ['vote', 'republican', 'pennsylvania', 'early', 'congratulation', 'everyone', '2020', 'voter', 'future', 'never', 'almost', 'ballot', 'insane', 'id', 'independent', 'person', 'exactly', 'major', 'victory', '2024', 'count', 'yesterday', 'trend', 'northcarolina', 'well', 'nevada', 'reason', 'swingvoter', 'double', 'those', 'goal', 'touch', 'immediately', 'michigan', 'buck', 'county', 'require', 'moreover', 'lead', 'noncitizen', 'bring', 'improve', 'arizona', 'registration']
- Comunidad 2: ['donaldtrump', 'time', 'big', 'win', 'watch', 'number', 'worth', 'mar', 'joerogan', 'tonight', 'put', 'money', 'half', 'boeing', 'documentary', 'continue', 'humanity', 'full', 'turn', 'conversation']
- Comunidad 3: ['dems', 'day', 'one', 'life', 'hoax', 'because', 'party', 'literally', 'save', 'endorse', 'exceed', 'extremely', 'men', 'push', 'relentlessly', 'candidate', 'request', 'cruel', 'germany', 'thank', 'amount', 'through', 'wing']
- Comunidad 4: ['legacymedia', 'lie', 'propaganda', 'public', 'arm', 'msg', 'use', 'citizenjournalism', 'break', 'matter', 'medium', 'company', 'fraud', 'obviously', 'see', 'rally', 'bear', 'way', 'crazier', 'superior', 'sense', 'trust', 'try', 'licentiously', 'overseas']
- Comunidad 5: ['people', 'right', 'starlink', 'support', 'freedom', 'grok', 'work', 'wave', 'good', 'speech', 'free', 'need', 'nov5th', 'great', 'same', 'high', 'power', 'second', 'everything', 'russia', 'image', 'first', 'take', 'million', 'rule', 'constitution', 'tesla', 'possibly', 'believe', 'factor', 'ukraine', 'provide', 'military', 'california', 'let', 'phone', 'bandwidth', 'terminal', 'sign', 'petition', 'understands', 'thanks', 'emergency', 'wikipedia']
- Comunidad 6: ['x', 'send', 'most', 'friend', 'link', 'think', 'important', 'far', 'level', 'data', 'youtube', 'google', 'machine', 'source', 'cheney', 'die', 'differently', 'front', 'line', 'fight', 'hard', 'appreciate', 'thing', 'write']
- Comunidad 7: ['government', 'kill', 'pet', 'squirrel', 'pnut', 'allow', 'house', 'barge', 'instead', 'sound', 'overreach', 'expose', 'department', 'efficiency', 'corruption', 'catherine', 'herridge', 'clean']
- Comunidad 8: ['mindvirus', 'call', 'news', 'real', 'stop', 'mean', 'disinformation', 'woke', 'massive', 'crush', 'defeat', 'network', 'dnn', 'wake', 'learn', 'explain', 'back', 'measurable', 'shift', 'virginia']
- Comunidad 9: ['kamala', 'illegal', 'biden', 'puppet', 'word', 'swingstate', 'give', 'administration', 'live', 'legal', 'true', 'campaign', 'seem', 'mess', 'legalize', 'action', 'immigration', 'name', 'talk', 'newyorktimes', 'democracy', 'sue', 'competent', 'program']
- Comunidad 10: ['down', 'window', 'year', 'whole', 'overtón', 'open']
- Comunidad 11: ['america', 'state', 'become', 'come', 'blue', 'platform', 'new', 'criminal', 'red', 'fact', 'easy', 'here', 'illegally', 'oneparty', 'deep', 'overwhelmingly', 'policy', 'legally', 'nobel', 'soon', 'awesome', 'justice', 'statement', 'accurate']
- Comunidad 12: ['porsche', '911', 'cybertruck']
- Comunidad 13: ['inflation']
- Comunidad 14: ['election', 'please', 'note', 'last', 'spacex', 'city', 'town', 'hall', 'funny', 'interference', 'capacity', 'urban', 'possible', 'sure', 'mail', 'single', 'figure', 'science', 'neuralink', 'cool', 'highly', 'recommend', 'super', 'talented', 'team', 'integrity', 'community', 'outrageous', 'congrats']