

Technical Report



University of the Basque Country UPV/EHU
Department of Computer Science and Artificial
Intelligence

Analysis of Spanish Text-Thesaurus as a Complex Network.

B. Cases and A. Moujahid

January 2014

San Sebastian, Spain
www.ehu.es/ccia-kzaa
hdl.handle.net/10810/4562

Analysis of Spanish Text-Thesaurus as a Complex Network.

B. Cases and A. Moujahid

Resumen

Basándose en las herramientas de la Teoría de Redes Complejas, este trabajo proporciona un estudio descriptivo del diccionario básico de sinónimos del español Open Thesaurus representado como un grafo. Estudiamos las principales medidas estructurales de la red comparándolas con las de un grafo aleatorio. Los resultados numéricos muestran que Open-Thesaurus es un grafo cuyas propiedades topológicas se aproximan a una red libre de escala (Scale-free network) aunque no presenta la propiedad del pequeño mundo debido a la representación dispersa que realizamos. Encontramos además que las palabras de mayor centralidad de intermediación corresponden a términos que sugieren el vocabulario del psicoanálisis: placer, ayudante (en el sentido de trabajador) y regular (en el sentido de ajustar).

Palabras clave: Semántica, Relación de Sinonimia, Redes Complejas.

Abstract

Based on the theoretical tools of Complex Networks, this work provides a basic descriptive study of a synonyms dictionary, the Spanish Open Thesaurus represented as a graph. We study the main structural measures of the network compared with those of a random graph. Numerical results show that Open-Thesaurus is a graph whose topological properties approximate a scale-free network, but seems not to present the small-world property because of its sparse structure. We also found that the words of highest betweenness centrality are terms that suggest the vocabulary of psychoanalysis: *placer* (pleasure), *ayudante* (in the sense of assistant or worker), and *regular* (to regulate).

Keywords: Thesaurus, synonymy, complex networks.

1. Introducción

La Teoría de Redes Complejas se ocupa de estudiar las propiedades macroscópicas de grafos de gran tamaño a partir del estudio de las interacciones locales, atendiendo tanto a los aspectos estructurales como a los dinámicos. En

(Newman 2003) y (Boccaletti et al. 2006) se pueden encontrar definidos los conceptos básicos.

Un red compleja es un grafo de gran tamaño que posee una topología no trivial desde el punto de vista estadístico: no es una red regular generada por un algoritmo determinista, pero tampoco es un grafo totalmente aleatorio, como los estudiados por Erdős y Rényi a mitad del s. XX. Esta característica es se muestra repetidamente en los datos recogidos de diferentes fuentes.

En las redes sociales los nodos representan individuos y las aristas relaciones sociales, como ser familia, amigo, empleado, vecino, coautor, etc. Las redes biológicas comprenden las redes neurológicas, las llamadas redes ómicas (genómica, proteómica, etc.) y las epidemiológicas. También se aprecia la característica topología de las redes complejas en las redes tecnológicas: redes electricidad, de telefonía, Internet, etc.

Las redes complejas comienzan a ser una herramienta de estudio en las áreas de la Inteligencia Artificial (Bales y Johnson2006), (Mika 2007) y del Procesamiento del Lenguaje Natural, principalmente en la minería de datos sobre Internet (Ortega 2013). Algunos trabajos han abordado el análisis de tesauros desde la perspectiva de las redes complejas (De Jesus Holanda et. al. 2004), (Steyvers y Tenenbaum 2005), pero no tenemos constancia de trabajos que aborden el mismo análisis de tesauros en español.

En este trabajo abordamos el análisis como red compleja del diccionario básico en español Text-Thesaurus que es la base sin modificar del diccionario colaborativo Open Thesaurus (Open-Thesurus) y que contiene grupos de sinónimos de la forma $a; b; c; d; e$ obtenidos del Diccionario de la Real Academia Española. A diferencia de otros trabajos, realizamos una representación dispersa de la relación de sinonimia. Estudiamos si la topología de Text-Thesaurus corresponde a una red libre de escala calculando la curva de regresión exponencial de la distribución de grados y comparándolo con un grafo de igual tamaño generado aleatoriamente por el modelo GN,p de Erdős-Rényi. Analizamos las medidas de centralidad intermediación (Betweenness) y de cercanía (Closseness). También analizamos si Text-Thesaurus presenta la propiedad de pequeño mundo. Al tratarse de una representación dispersa, cabe preguntarse cuál es el índice k de degeneración del grafo y determinar el núcleo de mayor conectividad mediante el algoritmo k -core.

Por último, abrimos el debate acerca de si la representación de redes complejas podría ser la forma idónea de comparar diversas fuentes de datos desde la perspectiva de la Psicología, la Neurología y la Lingüística Computacional. En concreto y a la vista de los resultados descriptivos del diccionario analizado, proponemos el estudio de los conceptos y técnicas del psicoanálisis como una forma de dotar de armazón científico a esta práctica clínica en en las áreas de la psicología, la neurociencia computacional y la Inteligencia Artificial.

2. Representación de Text Thesaurus como un grafo disperso

Se ha seleccionado el diccionario básico de sinónimos Text-Thesaurus bajo licencia GNU. No comprende todo el vocabulario del español: los términos que no disponen de sinónimos en el diccionario de la Real Academia Española (RAE) (por ejemplo lago, para el que el tesoro proporciona una definición), se consideran nodos aislados y no están incluidos en Text-Thesaurus (cada grupo de sinónimos contiene al menos dos palabras). Como el Diccionario Esencial de la RAE (RAE 2013), según se cita en su web, contiene más de $N = 88,000$ entradas léxicas, de los que 18,664 son palabras del diccionario Text Thesaurus (TT) y 69,336 son nodos aislados (no tienen sinónimos y no están en el diccionario). Tomaremos N como el número de nodos.

Formalmente un grafo es un par $G = (V, E)$ compuesto por un conjunto de nodos o vértices V y un conjunto de aristas E , que son pares de nodos (i, j) . Si la relación E es simétrica el grafo es no dirigido, tipo al que nos restringimos en este trabajo. Gráficamente una arista es un segmento que conecta los puntos i y j . El tesoro está organizado por grupos de sinónimos separados por punto y coma y que acaban con un salto de párrafo. Eventualmente un sinónimo puede ir seguido de una etiqueta (fig.) que indica que la palabra se usa en sentido figurado, (noRAE) para indicar que no se encuentra como sinónimo en el diccionario de la RAE y (vulg.) para los vulgarismos. Se han eliminado y por lo tanto no se han registrado como atributos de los arcos estas etiquetas. Cada línea del tesoro $a; b; c; d; e...$ se substituyó usando el editor de textos por $a--b--c--d--e...$, que es la notación del formato .dot de grafos que soporta Graphviz (Bilgin et. al. 1988) y que puede importar la aplicación Gephi (Bastian, Heymann y Jacomy 2009) que usamos para el análisis. Esta representación del grafo de sinonimia es dispersa (sparse): la sinonimia perfecta, entendida como relación de equivalencia de palabras que pueden substituirse entre sí en la misma frase sin cambiar el significado de la misma, requeriría ser reflexiva, simétrica y transitiva. Así, cada grupo de sinónimos debería conformar una clique, es decir, un subgrafo en el que todos los nodos están conectados entre sí. Puesto que el tamaño del fichero de texto alcanza 399 Kb, se optó por representar exclusivamente las secuencias de lectura de cada línea.

3. Propiedades topológicas.

En esta sección vamos a definir las medidas de centralidad generalmente usadas para caracterizar una red. La propiedad topológica más importante que permite describir la estructura de un grafo la podemos encontrar en la distribución estadística de los grados $P(k)$ definida como la probabilidad de que un nodo elegido aleatoriamente de manera uniforme tenga el grado k . En los grafos aleatorios de Erdős-Rényi la distribución de grados es una distribución binomial (de Poisson cuando el número de nodos tiende a infinito) con media $\langle k \rangle = Np$ donde N es el número de nodos y p la probabilidad de que dos nodos diferentes

estén conectados.

Otra propiedad de interés es la tendencia que tienen los nodos de un grafo de formar cliques o grupos. Un k -clique es un k -subgrafo completo de tamaño $k < N$. Ésta característica conocida como *clustering* ha sido observada en muchas redes complejas tanto naturales como tecnológicas. Para un nodo i de grado k_i el coeficiente de clustering es dado por:

$$c_i = \frac{2e_i}{k_i(k_i - 1)} \quad (1)$$

donde e_i es el número de aristas conectando los vecinos del nodo i . El coeficiente medio de clustering caracterizando todo el grafo es simplemente la media de los c_i . En las redes complejas, a diferencia de las redes aleatorias, el coeficiente medio de clustering es relativamente alto indicando mayor transitividad de la red.

Junto con el coeficiente de clustering y la distribución de grados, la longitud media de camino (Average path length) definida como el número medio de pasos a lo largo de los caminos más cortos que conectan dos nodos cualquiera en el grafo, representa otra de las medidas de centralidad más robustas. La longitud media de los caminos mínimos viene dada entonces por:

$$l = \frac{1}{(N(N-1))} \sum_{i,j \in V, i \neq j} d_{ij} \quad (2)$$

La mayoría de las redes complejas se caracterizan por tener una longitud media de camino relativamente corta, está característica es conocida con el nombre del mundo pequeño.

Para cerrar este breve resumen acerca de las medidas de centralidad, introducimos la *betweenness* un concepto muy útil para cuantificar la importancia de los nodos (o aristas) de un grafo. Un camino de longitud k sobre el grafo que conecta los nodos n_0 y n_k es una sucesión de nodos n_0, n_1, n_{k-1}, n_k en la que para todo $i, 0 \leq i \leq k$ el par (n_i, n_{i+1}) es una arista. Un camino mínimo es uno de longitud mínima. Para un nodo i , la *betweenness* se define como:

$$b_i = \sum_{j,k \in V, j \neq k} \frac{n_{jk}(i)}{n_{jk}} \quad (3)$$

$n_{jk}(i)$ es el número total de caminos mínimos que conducen del nodo j al nodo k pasando por el nodo i y n_{jk} es el número total de caminos mínimos conectando los nodos j y k . Por otra parte, la densidad d de un grafo mide la probabilidad de que dos nodos estén conectados y se calcula como la fracción de aristas del grafo, M , entre el número total de aristas posible,

$$d = \frac{2M}{N(N-1)} \quad (4)$$

Nodos totales N	88.000
Nodos aislados	69.336
Nodos no aislados NTT	18.664
Aristas M	17.465
Densidad de TT	1 E -4
Densidad de G	1 E -5
Número de componentes conexas de TT	3.201
Número de componentes conexas de G	72.537
Longitud del camino medio l	13,637
Diámetro	38
Tamaño de la componente gigante	10.592

Cuadro 1: Datos estadísticos básicos del grafo no dirigido

4. Análisis topológico

Una componente conexas es un subgrafo $G' = (V', E')$ de G en el que cada dos nodos i, j de V' están conectados por un camino.

En la tabla 1 se puede ver que el número de componentes conexas del grafo G sin nodos aislados es 3.201, que asciende a 72.537 en el grafo G que incluye los nodos aislados. La longitud del camino mínimo más largo o diámetro es 38, mientras que la longitud media de todos los caminos mínimos es del orden de 13,637. La componente conexas gigante es la de mayor tamaño reuniendo 10.592 nodos. Todas estas medidas formarán parte del análisis del grafo para determinar si presenta las características de una red compleja. Las preguntas que se plantean son las siguientes:

¿Es la sinonimia representada en el grafo G una relación arbitraria generada aleatoriamente o bien muestra una pauta de crecimiento propia de las redes complejas?. ¿Puede ser este diccionario producto de un proceso de crecimiento con probabilidad uniforme?. Para analizarlo, compararemos la distribución de frecuencias de los grados de los nodos con la de un grafo aleatorio de las mismas dimensiones. Para llevar a cabo este análisis, hemos comparado el grafo correspondiente a Text-Thesaurus con un grafo aleatorio de Erdős-Rényi formado por el mismo número de nodos y aristas. El modelo utilizado es el modelo $G_{N,p}$ que consiste en recorrer la lista de las $N(N - 1)/2$ posibles aristas uniendo los nodos con probabilidad p . Tomando como base $N = 88,000$ nodos, la figura 1(A) compara las distribuciones de probabilidad de los grados de los nodos para un grafo aleatorio y para el grafo G con densidad $p = 10^{-5}$.

La distribución de probabilidad de los grados para el diccionario G sigue una ley de potencias en la que hay muchos nodos de grado bajo y muy pocos de grado mayor o igual que 5 hasta 22, que es el máximo. El grado medio es 1,8761 vecinos por nodo si no consideramos los nodos aislados (TT) y baja a 0,398 si los tenemos en cuenta (G). El grado medio pN del grafo aleatorio es de $10^{-5} \times 88,000 = 0,88$. En cuanto a las 10 palabras con mayor grado que se enumeran en la tabla 2, cobrarán mayor sentido en la medida en que avancemos en el análisis. A la vista del análisis estadístico, no parece que el

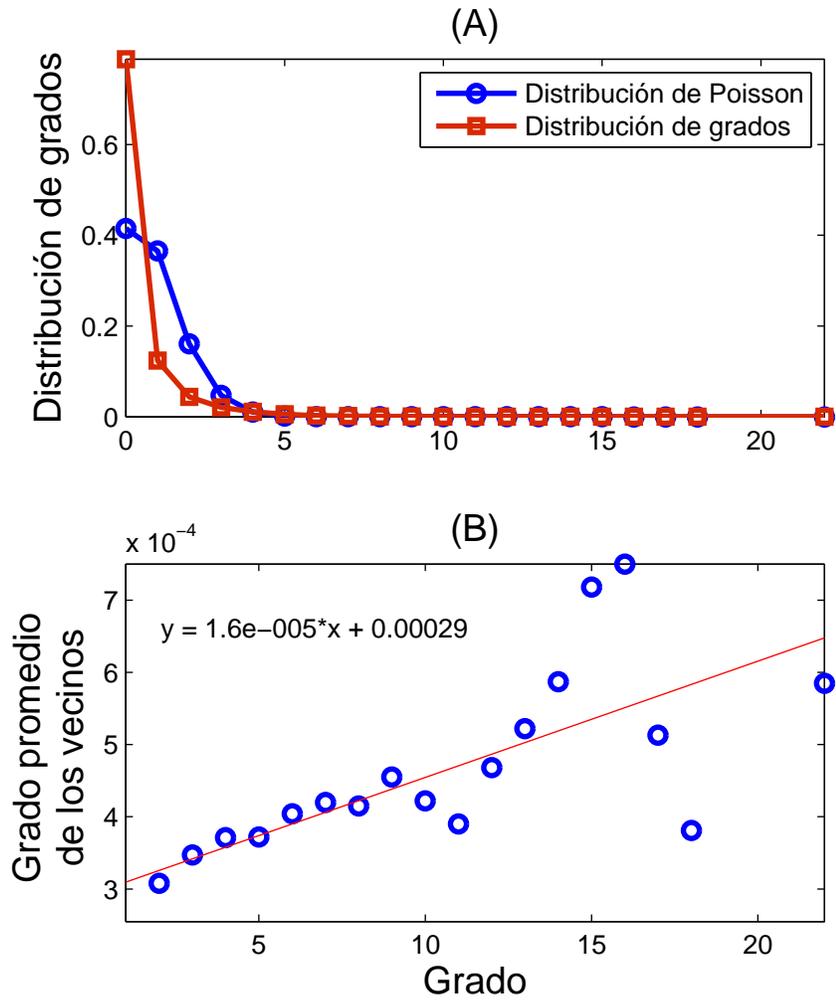


Figura 1: (A) Distribución de grados del grafo G. Los marcadores circulares corresponden a la distribución de Poisson del grafo con $N = 88,000$ nodos y media dada por Np . La distribución de grados es representada por los cuadrados. (B) Resultado del algoritmo KNN.

Grado	Id	Grado	Id
22	gracia	16	confusión
18	engaño	15	necio
17	golpe	15	alboroto
17	lío	14	tonto
17	atractivo	14	molestar

Cuadro 2: Las 10 palabras con mayor grado.

	R2	Ecuación
Logarítmica	0,6555	$y = -0,1374Ln(x) + 0,3381$
Lineal	0,3142	$y = -0,0131x + 0,1858$
Exponencial	0,9368	$y = 0,2904e^{-0,4839x}$
Potencial	0,9361	$y = 3,1354x^{-3,51}$

Cuadro 3: Curvas de regresión.

grafo G corresponda a los parámetros de distribución de grados de un grafo aleatorio de Erdős-Rényi. Nos preguntamos entonces si Text-Thesaurus podría ser una red de enlace preferencial.

Las redes de enlace preferencial (preferential attachment) son grafos cuyo crecimiento está condicionado por los nodos preexistentes en la red cada vez que se añade un nodo nuevo n : la probabilidad de conexión con otro nodo n' es proporcional al grado de n' . Así, los nodos de mayor grado adquieren mayor peso dando lugar a una distribución de probabilidad de grados potencial, es decir, $P(x) = ax^{-\gamma}$. Las redes libres de escala (Barabási, A. L., Réka A. 1999) son las que presentan una distribución potencial de los grados, siendo el rango típico de valores, obtenido analizando redes de diferente procedencia, entre 2.1 y 4. Restringiéndonos al grafo Text-Thesaurus, sin nodos aislados, que normalmente se excluyen del análisis, hemos ajustado varias curvas de regresión a la distribución de probabilidad de los grados. Los resultados se exponen en la tabla 3, el valor R es el coeficiente de correlación que varía entre 0 y 1, indicando cuanto más alto una mayor bondad del ajuste.

Comprobamos así que la distribución exponencial y la potencial, características ambas de las redes complejas (Newman 2005), (Deng, Cai, y Wang . 2011), son las que mejor se ajustan a la distribución de grados de Text-Thesaurus. En dicho diccionario, el exponente gamma alcanza 3,51, situándose en los parámetros de las redes libres de escala según (Newman 2005) en el rango $2, 1 \leq \gamma \leq 4$. Sin embargo, para decidir definitivamente si Text-Thesaurus es una red libre de escala debemos comprobar otros aspectos que tienen que ver con la autosimilaridad o la propiedad de fractalidad: que los subgrafos determinados aleatoriamente de un grafo libre de escala G tengan distribuciones de grado similares a G .

El primer aspecto a estudiar es la correlación entre el grado de un nodo y la media de los grados de sus vecinos a través del algoritmo KNN (K-Nearest Neighbors), que permite distinguir entre redes que presentan fuerte correlación,

las llamadas redes selectivas o assortativas: A mayor (menor) grado del nodo, mayor (menor) grado medio de los vecinos; y redes No selectivas (disassortative): A mayor (menor) grado del nodo, menor (mayor) grado medio de los vecinos.

El algoritmo consiste en asignar a cada nodo un par (k,KNN) donde k es el grado y KNN el grado medio de los vecinos. $KNN(k)$ es la media de todos los valores KNN sobre el mismo grado k . Si $KNN(k)$ es creciente, hablamos de una red selectiva, mientras que valores decreciente de $KNN(k)$ indican una red no selectiva. La figura 1(B) muestra el resultado del algoritmo KNN, utilizando el software Network Workbench (NWB Team. 2006) junto con la recta de tendencia lineal: al ser la pendiente de la recta de regresión prácticamente horizontal podemos afirmar que no existe correlación entre los grados.

El coeficiente medio de clustering $\langle C \rangle = C(i)/N = 0,022$ es muy bajo y su promedio según el grado muestra una ligera tendencia creciente como se aprecia en la Fig. 2(A), aunque la gráfica muestra un tramo claramente creciente, desde el grado 1 al 15, y otro decreciente, del grado 15 al 22. No se puede decir entonces que Text Thesaurus sea un red libre de escala, aunque existe mucha discusión en la literatura y el grafo muestre una ligera tendencia creciente con el grado tanto en la correlación de grados KNN como en el coeficiente de clustering. Podríamos decir que esta representación dispersa apunta, aunque no llega a manifestarse claramente, a ser una red libre de escala.

La segunda propiedad característica de las redes complejas es la de ser un *mundo pequeño* (o small-world) (Watts y Strogatz 1998). Esta idea proviene del estudio de las redes sociales de mediados del s. XX ligada a la idea de los seis grados de separación: entre dos habitantes del mundo existe un camino mínimo de conocidos de longitud 6 que conecta a ambos. Sin embargo, la caracterización científica de la propiedad de mundo pequeño se establece en que la distancia media entre cada dos puntos del grafo (en la misma componente conexa) sea de orden logarítmico con respecto al número de nodos, aproximándose por $l \approx \ln(N)$.

En el grafo G que complementa Text-Thesaurus con los nodos aislados del diccionario esencial de la RAE (Tabla 1), $N=88.000$ y $\ln N=9,834$, claramente inferior a la longitud media $l=13,637$ calculada para el grafo. Por otra parte, el diámetro 38, que es el de la componente conexa gigante, también niega que G sea un pequeño mundo. La estructura es demasiado dispersa para ser un mundo pequeño. A pesar de de que la distribución potencial de nodos expresada en la ecuación 1 tiene un exponente gamma que apunta a una red libre de escala, se demostró (Cohen. y Havlin 2003) que para redes en el rango $1 < \gamma < 3$ la longitud media estimada se aproxima a $\log(\log(N))$. Una vez más, esta representación se queda a las puertas de ser caracterizada como una red libre de escala. Por otra parte, los grafos aleatorios tienen una distancia media estimada de, $l \approx \frac{\ln(N)}{\ln(\langle k \rangle)}$.

Los límites inferiores y superiores de la longitud media de caminos mínimos del grafo GTT están dados respectivamente por $\ln(N)$ y $\frac{\ln(N)}{\ln(\langle k \rangle)}$ (Vease Eq. (5)). Esto indica que la red de diccionario no es un mundo pequeño, pero tampoco es completamente aleatoria.

Nodos totales N_{CG}	10.592 (57,75 %)
Aristas M	12.571 (72,02 %)
Densidad CG	$2 \cdot 10^{-4}$
Longitud del camino medio l	13,639
Diámetro	38
Coefficiente de Clustering	0,025

Cuadro 4: Datos estadísticos básicos

$$\ln(N) = 9,834 < l = 13,637 < \frac{\ln(N)}{\ln(\langle k \rangle)} = 15,630 \quad (5)$$

Restringiéndonos a la componente conexa gigante (CG), los datos estadísticos son muy similares a los del grafo total, mostrados en la tabla 2. La Tabla 2 y la Eq. (6) muestran que incluso es más dispersa que un grafo aleatorio.

$$\ln(N) = 9,268 < \frac{\ln(N)}{\ln(\langle k \rangle)} = 10,34 < l = 13,637 \quad (6)$$

Las redes libres de escala, al ser redes selectivas, tienden a que los nodos de mayor grado se conecten más entre sí. Forman entramados en los que cada uno se rodea de otros de grado similar. Son como el punto central o buje de una rueda que se rodean de nodos de grado decreciente a medida que nos alejamos del centro.

La figura 2(B) muestra los promedios de centralidad de cercanía: tienen una ligera tendencia decreciente a partir de los nodos de grado 5, de forma que los nodos de mayor grado ocupan lugares más centrales. Parece que la representación dispersa apuntara a las propiedades de una red compleja sin llegar a manifestarse claramente. La centralidad de intermediación es mayor cuantos más caminos pasen por el nodo. Si un nodo es tal que al quitarlo aparecen nuevas componentes conexas, todos los caminos mínimos de unas a otras componentes deben pasar por ese nodo intermediario.

En la figura 2(C) se puede apreciar que la betweenness media sí es muy alta para los nodos de mayor grado. Nótese que mientras que el algoritmo KNN y la intermediación de cercanía muestran máximos en el grado $k = 15$, la Betweenness muestra un mínimo. Se diría que los nodos más intermediarios ($k > 15$) son a su vez escasos en triángulos, como puede ocurrir en los árboles.

5. El principio del placer y el principio de realidad en los tesauros

En esta sección, una vez mostradas las medidas estadísticas de la red con el fin de catalogarla o no como red compleja, procederemos al análisis cualitativo de la misma, es decir, analizar el entramado de nodos con mayor centralidad de

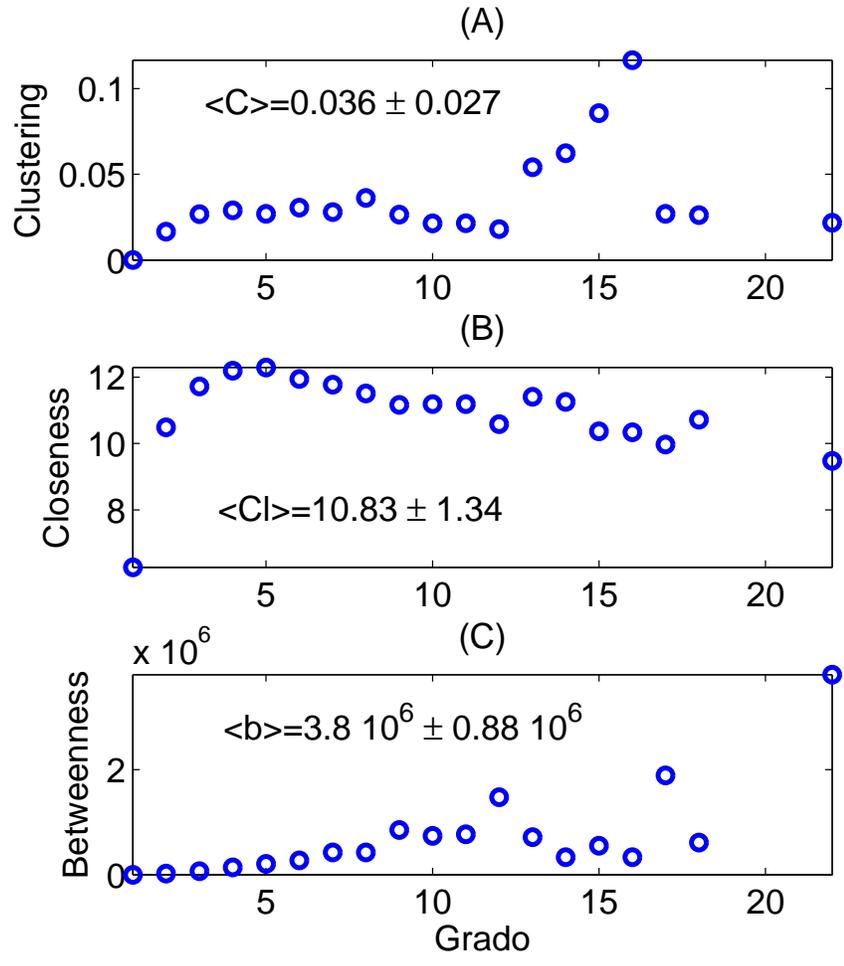


Figura 2: Promedio de los coeficientes de clustering (A), closeness (B) y betweenness (C) según el grado.

intermediación no solo para examinar si están o no relacionados entre sí, sino qué significados agrupan.

El subgrafo de las 20 palabras con mayor Betweenness se muestra en la Fig. 3. Placer es la palabra más central, seguida de deleitar: el gusto, la alegría, seducir, satisfacer, liquidar en la misma componente conexa, sugieren lo que en Psicoanálisis se nombra como Principio del Placer (Freud 1920): nuestra psique persigue satisfacer de forma inmediata y continuada las necesidades básicas.

El proceso de sublimación consiste en la represión del principio del placer a favor del principio de realidad: la no inmediata satisfacción de una necesidad instintiva, como el sexo, para convertirla en una actividad productiva, como trabajo, arte o juego y que a medio plazo contribuirá a la satisfacción del instinto de forma socialmente aceptable. Parece increíble que la tercera y cuarta palabras en valor de la Betweenness sean ayudante y auxiliar, conectadas también con trabajador, guía, donar y defender, que sugieren altruismo (no satisfacción del instinto a corto plazo). La quinta palabra es regular, que aparece con dos significados: ajustar y vulgar. Por último está la componente del juicio, el parecer y la aproximación.

Con respecto al nivel de dispersión que presenta la representación según flujo de lectura de los grupos de sinónimos, estudiamos a continuación el núcleo duro o k -core del grafo (Dorogovtsev, Goltsev y Mendes 2006). El valor k , llamado índice de degeneración de un grafo es el menor valor k tal que cualquier subgrafo tiene al menos un nodo de grado menor o igual que k . Quitando los nodos de grado k incrementalmente desde los de menor grado k hasta el valor máximo que cumpla la condición de degeneración, se obtiene el k -core del grafo y k es el índice de degeneración.

Text-Thesaurus presenta un índice de degeneración $k=4$ y nos muestra el 4-core de la Fig. 4. Se han utilizado los pesos de las aristas (algunas secuencias de sinónimos se repiten en más de un grupo). Todos los nodos tienen cuatro o más vecinos. El follón y sus sinónimos son el corazón del español según Text-Thesaurus, con todas sus connotaciones psicoanalíticas.

6. Conclusiones

A la vista de los resultados cuantitativos, Text-Thesaurus se podría catalogar como una red compleja dispersa. La razón es obvia: la representación secuencial de los grupos de sinónimos, que transitivamente deberían ser cliques, restan triángulos a la estructura, traducándose en la pérdida de la propiedad de pequeño mundo. Cabe estudiar si este problema es común a otros campos en los que se trabaja con secuencias: valgan como ejemplo los resultados del análisis de la red eléctrica en (Watts y Strogatz 1998); la genómica y la proteómica, al tratarse también de secuencias, podrían ser candidatas a presentar el mismo fenómeno de dispersión.

Por otra parte, nos preguntamos si las redes complejas, de las que son muestras las redes neurológicas y las lingüísticas no se perfilan como la mejor herramienta para comparar los efectos que tienen en el cerebro las herramientas

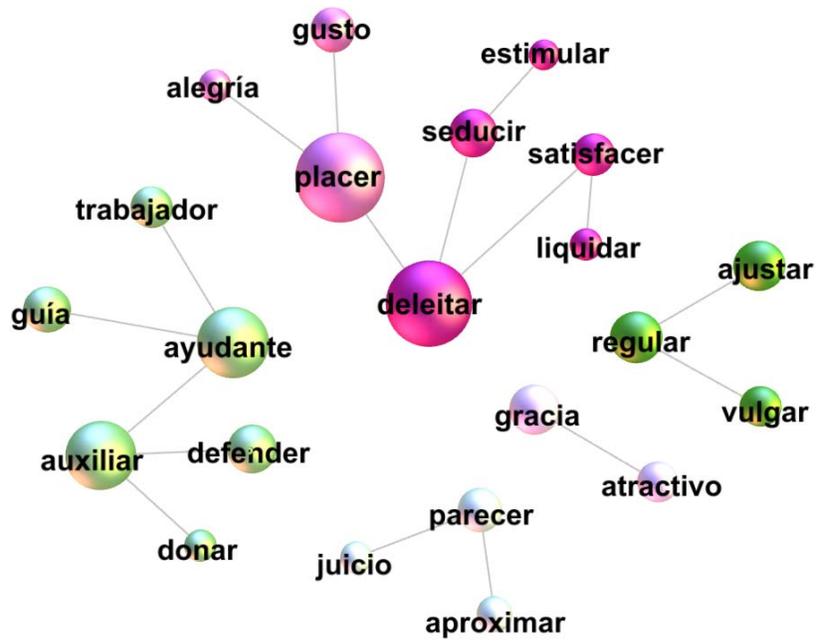


Figura 3: Palabras con mayor valor de Betweenness separadas en diferentes comunidades.

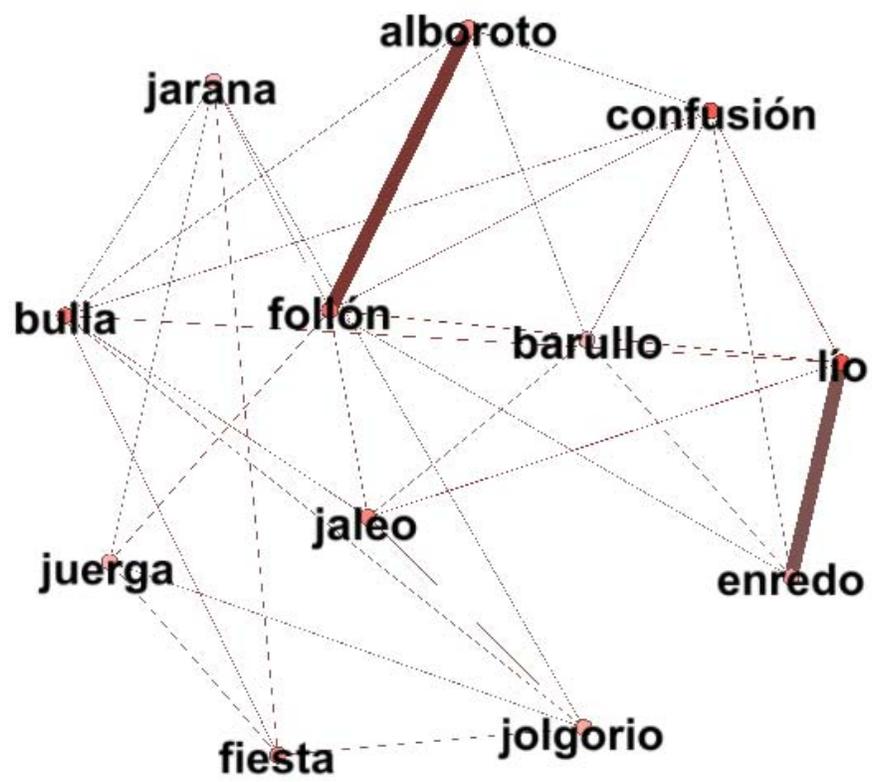


Figura 4: 4-core del grafo Text-Thesaurus.

basadas en el lenguaje que usa la terapia psicoanalítica a través de técnicas como la tomografía funcional y otros sensores fisiológicos. Quizás, desde esta perspectiva, la Neurolingüística podría atacar el duro problema de relacionar el lenguaje y la emoción, el último reto de la Inteligencia Artificial, a través de nuestras respuestas fisiológicas.

Referencias

- [1] Barabási, A. L., Réka A. (1999). "Emergence of scaling in random networks". *Science* 286 (5439): 509-512
- [2] Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media. Aplicación disponible en línea en <https://gephi.org/>
- [3] Bales, M.E., Johnson, S.B. (2006). Graph theoretic modeling of large-scale semantic Networks. *Journal of Biomedical Informatics*, 39 (4), pp. 451-464.
- [4] Bilgin, A. et al. (1988). Graphviz - Graph Visualization Software. Disponible en <http://www.graphviz.org/Home.php>.
- [5] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., y Hwang, D. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175-308, February.
- [6] Cohen, R. y Havlin S. (2003). Scale-Free networks are ultrasmall. *Physical Review Letters*, 90(5):058701, February 2003.
- [7] De Jesus Holanda, A., Pisa, I.T., Kinouchi, O., Martinez, A.S., Ruiz, E.E.S. (2004). Thesaurus as a complex network *Physica A: Statistical Mechanics and its Applications*, 344 (3-4 SPEC. ISS.), pp. 530-536.
- [8] Deng, W., Li, W., Cai, X., Wang, Q.A. (2011). The exponential degree distribution in complex networks: Non-equilibrium network theory, numerical simulation and empirical data. *Physica A: Statistical Mechanics and its Applications*, 390 (8), pp. 1481-1485.
- [9] Dorogovtsev, S. N., Goltsev, A. V. y Mendes, J. F. F. (2008). Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275-1335, October.
- [10] Dorogovtsev, S.N.; Goltsev, J.F.F.; Mendes, JF. (2006). "k-Core Organization of Complex Networks", *Physical Review Letters* 96 (4): 040601
- [11] Freud, S. (1920). Más allá del principio de placer. Obras completas de Sigmund Freud. Volumen XVIII -Amorrortu editores.

- [12] Mika, P. (2007). Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(1):5-15, March 2007
- [13] NWB Team. (2006). Network Workbench Tool. Indiana University, Northeastern University, and University of Michigan, <http://nwb.slis.indiana.edu>
- [14] Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167-256.
- [15] Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323-351.
- [16] Real Academia Española y Asociación de Academias de la Lengua Española, *Diccionario esencial de la lengua española*. Madrid: Espasa, (2006). Disponible en línea en la página web <http://rae.es/recursos/diccionarios/desen>
- [17] Steyvers, M., Tenenbaum, J.B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29 (1), pp. 41-78.
- [18] OpenThesaurus-es - Tesoro en español. Disponible en la página web <http://openthes-es.berlios.de/>.
- [19] Ortega, F. J. (2013). Detection of Dishonest Behaviors in On-Line Networks Using Graph-based Ranking Techniques. *Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*, Monografía 12.
- [20] Watts. D.J., and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393 (6684): 440-442