



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

La visibilidad en el análisis de textos:
un enfoque de la teoría de redes

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuaría

PRESENTA:

Ana Victoria Ramos Ruiseco

TUTORA

Dra. Bibiana Obregón Quintana

2017



CIUDAD UNIVERSITARIA, CD.MX.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno.

Ramos
Ruiseco
Ana Victoria
55521738
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
309209955

2. Datos del tutor.

Dra.
Bibiana
Obregón
Quintana

3. Datos del sinodal 1.

Dr.
Pedro Eduardo
Miramontes
Vidal

4. Datos del sinodal 2.

Dra.
Claudia Orquídea
López
Soto

5. Datos del sinodal 3.

Dr.
Roman Anselmo
Mora
Gutiérrez

6. Datos del sinodal 4.

Dr.
Adonis Germinal
Cocho
Gil

7. Datos del trabajo escrito.

La visibilidad en el análisis de textos:
un enfoque de la teoría de redes
78 p.
2017

Agradecimientos y dedicatorias

A Dios

Por permitirme llegar a este momento tan especial en mi vida. Por los triunfos, los momentos difíciles y sobretodo por las personas que ha puesto en mi camino.

A mis padres

Por apoyarme en todo momento, por sus consejos, sus valores, por hacer de mí una persona de bien, pero más que nada, por su amor incondicional. Gracias por ser ejemplos de perseverancia y por siempre creer en mí.

A mi hermano

Por estar siempre a mi lado, ser mi cómplice y por recordarme que siempre la unión hace la fuerza. Te quiero mucho.

A mis abuelos

A mi abuelita Conchita por nunca dejarme sola, por su cariño, paciencia y por ser una segunda mamá para mi hermano y para mí. A los demás abuelitos por cuidarme y guiarme aunque ya no están físicamente.

A mis familiares

Por que cada uno fue parte importante en este proceso. Por estar siempre pendiente de mí y por recordarme a cada paso que siempre podré contar con ustedes.

A la Dra. Bibiana Obregón

Por permitirme trabajar a su lado y por compartir su tiempo y sus conocimientos. No tengo como agradecerle su paciencia y empeño en el desarrollo de este trabajo, pero sobretodo, quiero agradecerle por su dedicación como profesora, México necesita más profesores como usted.

A los sinodales

Dr. Pedro Miramontes, Dr. Germinal Cocho y Dr. Roman Gutiérrez, gracias por su apoyo, comprensión y observaciones, todos sus comentarios enriquecieron mi trabajo y a mí como persona. En especial, agradezco a la Dra. Claudia López por haber sido parte fundamental de mi formación profesional y por su apoyo en este trabajo.

A la UNAM y FCIENCIAS

A la Universidad Nacional Autónoma de México y en especial a la Facultad de Ciencias por permitirme ser parte de una generación de triunfadores y gente productiva para el país.

A mis grandes amigos JG

A esos grandes amigos Fer, Mirza, Alicia, Viri, Roy, Javier, Sam, que con el tiempo se hicieron mis hermanos. Gracias por reír, llorar, cantar, gritar y sobretodo crecer conmigo durante todos estos años. En especial a Alejandro por no dejar que olvide las cosas importantes de la vida.

A Frida,Rebeca y Dámaso

Por su amistad sincera que sin duda, me ha hecho mejor persona.

A mis amigos P6

Por enseñarme que un atardecer en Coyoacán, puede regalarte amigos del alma. En especial, Selene, Sergio, Axel, Olaf, gracias por ser parte de la mejor etapa de mi vida. Adrián y Monse, sin ustedes, nada de esto sería posible, gracias por soportarme en mis peores ratos y por estar conmigo compartiendo los mejores; los amo.

A Lydia, Cynthia y César

Lydia y César gracias al equipo que formamos logramos llegar hasta el final del camino y no podía tener mejor recompensa que su amistad. Los quiero mucho. Cynti gracias por dejarme compartir mi locura y la dupla que formamos, es inigualable.

Al BFUNAM

Por dejarme ser y expresarme a través de lo que más amo, la danza. En especial a Moni y Héctor que sin su apoyo, esto no sería realidad.

Al equipo CONAPO

Martha, Alma, Alfonso, Juan y Ale, no tengo como agradecerles la lección de vida que me dieron: trabajar es lo mejor del mundo cuando tienes un equipo tan maravilloso, talentoso y con personas tan dedicadas como ustedes. Gracias por nunca dudar de mí y por su apoyo incondicional para que este sueño se pudiera realizar.

A mi equipo de trabajo Plan Seguro

Por todo el apoyo de mis jefes César, Toño y Sergio, para poder culminar este trabajo. A Aura, Perla, Mario, Kenny y los Pedros por dejarme aprender cada día de ustedes y por compartir conmigo cada día de locura laboral. En especial a Ale por no dejar que me rindiera en el camino, por cada consejo y vivencia compartida.

A Héctor, mi todo

A ti que has llenado mi vida con tu luz y que me has regalado las mejores aventuras en lo que llevo vivido. Gracias por no soltarme de la mano en los momentos más difíciles y por ser el hombre tan maravilloso que eres. Sobretodo gracias por inspirarme a ser mejor cada día y por hacer que seguir compartiendo cada día contigo, sea mi nuevo sueño. Te amo.

Índice general

Introducción	8
1. La importancia del lenguaje	11
1.1. Lenguas y lenguajes	11
1.2. La Oratoria: el arte de la palabra	12
1.3. Las matemáticas de los lenguajes	13
2. Redes y sus propiedades elementales	17
2.1. Historia	17
2.2. Definiciones y principales modelos de redes	19
2.3. Redes complejas	22
3. Redes de lenguaje	26
3.1. El lenguaje como serie de tiempo	27
3.2. Algoritmos de visibilidad	28
4. Metodología del estudio	31
4.1. Metodología	31
5. Resultados	36
Conclusiones	60
Anexos	63
A. Biografías	64
B. Algoritmo de visibilidad en R	71
C. Textos seleccionados y dónde encontrarlos	74
Bibliografía	74

Índice de cuadros

2.1. Principales modelos de redes y algunas de sus características	20
4.1. Textos considerados por autor y número de palabras de cada uno de ellos	32
4.2. Lenguajes y longitud media de sus palabras	34
5.1. Comparación de los valores de aglomeración por artista	41
5.2. Comparación de los valores de eficiencia por artista	41
5.3. Comparación de los valores de distancia media por artista	42
5.4. Comparación de los valores de aglomeración por orador	43
5.5. Comparación de los valores de eficiencia por orador	44
5.6. Comparación de los valores de distancia media por orador	45

Introducción

Las palabras a lo largo de la historia han levantado corazones en tiempos oscuros, han dado esperanza en tiempos de desesperación, han dado coraje a los débiles y también han honrado a la muerte, pero sobretodo, han cambiado el curso de la historia. Es posible verificar que la historia está repleta de líderes políticos que usaron sus palabras para incitar el odio y empezar guerras, así como de otros cuyas palabras fueron la base para reformar naciones completas.

El poder de las palabras ha sido estudiado no únicamente por lingüistas, sino también por psicólogos, docentes y matemáticos, principalmente. Estos últimos las han estudiado para determinar la autoría de los textos, para medir la frecuencia de las palabras usadas e incluso han intentado cuantificar, con herramientas como la teoría de redes, sus efectos.

Con base en estas ideas, surge la idea de realizar este trabajo cuyo objetivo principal es dar a conocer la fuerte conexión que existe entre las matemáticas y los lingüistas, identificando puntos de acuerdo entre los campos, esto a través del algoritmo de visibilidad (NVG) que permite transformar las series de tiempo en redes. Para conseguirlo, este trabajo se ha estructurado en 5 capítulos que abordan desde qué son los lenguajes y las redes, hasta cómo transformar series de tiempo en dichas estructuras.

Se realizó un análisis de 53 textos escritos, interpretados y/o leídos por algunos de los personajes que se han ido haciendo espacio en la historia de la humanidad y que destacaron por su oratoria o en los casos que corresponda, por sus composiciones musicales. Se cubrió una amplia gama de actividades y vocaciones, de países, épocas históricas y circunstancias personales para lograr generalizar los resultados a diversas áreas de estudio.

Principalmente, se propone que el lenguaje sea visto como una serie de tiempo a la que se le puede aplicar el algoritmo mencionado anteriormente y que como resultado, la estructura de la serie de tiempo se conserva en la red, por lo que se buscaba obtener redes conectadas, no dirigidas e invariantes bajo transformaciones, es decir, redes libres de escala, por lo que, probar dichas afirmaciones representa otro de los objetivos de este trabajo.

El Capítulo 1 de esta tesis, explora el poder de las palabras haciendo énfasis en

la oratoria y los componentes principales en la transmisión de un mensaje, además, se presenta un breve recuento histórico de los principales estudios matemáticos realizados con el lenguaje, pues entre otras cosas, se ha tratado de identificar componentes o tendencias para explicarlo y se ha empezado a apreciar científicamente el poder que tienen las palabras para formar la identidad de la persona y moldear la forma en que percibe el mundo.

Por su parte, el Capítulo 2, presenta un resumen histórico de los hechos más relevantes en la historia de la teoría de redes para que ya en contexto, se puedan estudiar las propiedades elementales de las redes como son el coeficiente de aglomeración o *clustering*, la distancia media, la distribución de grado, entre otras. Lo anterior se hace con la finalidad de poder presentar los cuatro modelos principales de redes que existen: redes regulares, redes aleatorias, redes de mundo pequeño y de libre escala. El capítulo termina al introducir otro escalón importante en el desarrollo de la ciencia de las redes, las redes complejas, que no son más que la caracterización de un sistema complejo.

En el Capítulo 3 se continúa con el tema de redes complejas, pero enfatizado a un sistema complejo conocido por todo el mundo: el lenguaje. En este capítulo se describe el lenguaje como una serie de tiempo y se dan ejemplos de las diferentes redes que podrían formarse a partir de él. Además, se señala el por qué la idea de convertir series de tiempo en redes resulta atractiva y se mencionan algunos métodos de construcción de redes a partir de series de tiempo, entre los que se destaca el algoritmo de visibilidad, desarrollado por Lucas Lacasa y colaboradores en 2008 [?]. En este capítulo se detalla el algoritmo de visibilidad (NVG) que fue usado para realizar el análisis de esta tesis.

En el Capítulo 4 se detalla la metodología seguida para la realización de este trabajo. Se explica con base en qué fueron elegidos los textos y cómo fue implementado el algoritmo en el programa R así como bajo cuáles supuestos. Se brinda también una explicación sobre el idioma en que se consideraron los textos.

Es en el Capítulo 5 donde se presentan los resultados del análisis y para lograr una mayor y más fácil comprensión de todos los lectores, dichos resultados se dividieron en cuatro fases. En la primera se presentan los resultados generales de comparar canciones y discursos, en la segunda, los resultados obtenidos al efectuar comparaciones de textos del mismo tipo, es decir, canciones contra canciones y discursos contra discursos. Se hicieron además algunas comparaciones poco esperadas entre personajes que aparentemente no compartían puntos de vista o incluso momento histórico y estos resultados, se presentan en la tercera fase. Por último, en la fase cuatro, se da una breve explicación de la prueba de *Kolmogorov-Smirnoff* que se usa para intentar demostrar que las redes obtenidas por medio del algoritmo NVG, cumplen con todas las propiedades enunciadas en el quinto párrafo de la presente introducción.

Las conclusiones del análisis pueden consultarse en el último apartado, además de las posibles aplicaciones que podría tener este trabajo en otras áreas de la investigación y algunos comentarios generales sobre la realización de esta tesis. Además, se puede consultar sobre algunas mejoras que podrían hacerse en cuanto a la programación

del algoritmo en R en proyectos posteriores y lineamientos para futuros desarrollos alrededor de este tema.

En el anexo A se puede consultar una breve biografía de cada uno de los personajes de los que se tomaron los textos analizados, esto para comprender por qué fueron o han sido parte fundamental de la historia. Por otra parte, en el anexo B se encuentra el programa detallado y comentado, mientras que el anexo C presenta una tabla con la información de las páginas de internet o libros de los que se obtuvieron los textos analizados, por si se desea consultar alguno en específico.

En el proceso de investigación que ha culminado en la realización de esta tesis, se consultaron diversos libros, artículos de revistas y páginas de internet cuyos títulos, autores y ediciones pueden ser consultados en la sección de Bibliografía.

Capítulo 1

La importancia del lenguaje

*"Ten cuidado con lo que piensas, tus pensamientos se convierten en tus palabras.
Ten cuidado con lo que dices, tus palabras se convierten en tus acciones.
Ten cuidado con lo que haces, tus acciones se convierten en tus hábitos.
Ten cuidado con tus hábitos, tus hábitos se convierten en tu carácter.
Ten cuidado con tu carácter, tu carácter se convierte en tu destino."
-Mahatma Gandhi*

Las palabras que usamos, ya sea para hablar o escribir, influyen en cómo los otros nos perciben. Son un factor en las decisiones que la gente toma sobre ti, pues ellas definen quién eres y moldean tu percepción del mundo, desencadenan emociones, y sobretodo, tienen consecuencias. Tienen el poder de dar vida a través de la honestidad y estímulos, pero también pueden matar mediante mentiras. Para bien o para mal, las palabras definen nuestra realidad, con ellas tenemos garantizados nuestros derechos y obligaciones en diferentes leyes, se conmemoran hechos históricos y muchas veces, han capturado las esperanzas y sueños de una persona.

Las palabras tienen poder, para levantar o aplastar, para inspirar o provocar, para curar o herir, para crear o destruir. Nos incitan a comprar ciertas marcas, a comer ciertas cosas, a odiar a los que piensan diferente, a votar de cierta forma e incluso a matar. La historia está repleta de líderes políticos que usaron sus palabras para incitar el odio y empezar guerras, así como de otros cuyas palabras fueron la base para reformar naciones completas. Quienes escriben blogs o libros entienden este poder, los que trabajan en ventas y publicidad conocen su alcance, aquellos en los medios de comunicación lo comprenden y saben utilizarlo, y los líderes políticos saben casi todo acerca de él. Muchos profesionistas han estudiado el poder de las palabras, entre ellos, los matemáticos que han intentado cuantificar, con herramientas como la teoría de redes, sus efectos.

1.1. Lenguas y lenguajes

Dentro de los fenómenos culturales, que son todo aquello que hace la forma de vida de un pueblo, comunidad o grupo, como por ejemplo las religiones, las tradiciones y

costumbres y las artes, los que tienen la más larga historia continua son las lenguas y lenguajes del mundo. Esta continuidad es la esencia misma del lenguaje, pero como todo acontecer humano, es prácticamente imposible dar una fecha precisa para su nacimiento, aunque evidencia anatómica sugiere que la habilidad de hablar de los humanos inicia entre 1.6 millones y 600 mil años atrás. Sin embargo, la certeza de que se comunicaran ideas complejas ocurre hasta la sofisticación cultural que se asocia a los humanos “modernos” [?].

La historia de las lenguas y lenguajes a través del tiempo puede ser contada de muchas formas diferentes, pero todas ellas comienzan igual, terminando con la idea de que lengua y lenguaje significan lo mismo. A la palabra lengua se le atribuyen distintos significados que dependen del contexto y la disciplina en que se le utilice, según la Real Academia Española, la lengua es un sistema de comunicación verbal y casi siempre escrito, propio de una comunidad humana, mientras que el lenguaje es el estilo y modo de hablar y escribir propio de cada persona. En otras palabras, la lengua es lo que se habla, lo que se emplea para comunicarse, lo que se utiliza para escribir y lo que se aprende al nacer y que a su vez, se comparte con una comunidad entera, mientras que el lenguaje, concierne a las palabras que se emplean dentro de esa lengua, cómo usarlas, cómo escribirlas, los tonos que muestran las intenciones de las personas y además, dicta las normas ortográficas o gramaticales de cada una de las palabras.

Hoy en día, se hablan alrededor de 7,000 lenguas en el mundo y aunque parecen muchas, probablemente sean solamente una fracción de todas las que se han hablado en la historia pues una lengua, como todo lo que tiene vida, está en constante cambio. La lengua es un torrente de innovaciones en la pronunciación, en la construcción de las palabras y de las frases, en el vocabulario, en la entonación, en todo. La mayoría de las innovaciones se dan en vocabulario y, según Antonio Alatorre en [?], son como semillas que caen en piedra, pero algunas caen en buen terreno, germinan y se propagan. Las oraciones se construyen a partir de palabras y, aunado a lo anterior, se concluye que cada palabra tiene su propia historia y ello las ha convertido en las más estudiadas tanto por lingüistas como por psicólogos y matemáticos, principalmente.

1.2. La Oratoria: el arte de la palabra

La Oratoria es una forma de comunicación interpersonal a través de la cual el orador se sirve de una serie de actitudes, así como de técnicas verbales y no verbales, para influir sobre un determinado grupo de personas con la intención de que piensen, sientan y actúen de cierta forma. Proviene de una tradición griega puesta en práctica por eruditos que en su mayoría eran políticos y filósofos, y si llegaban a dominar este arte, les daba prestigio ante la sociedad de aquellos tiempos.

Hablar de la Oratoria es hablar del arte de la palabra, de cómo una persona es capaz de transmitir un mensaje hablando, así que la clave está en la capacidad del orador para hacer llegar aquello de lo que se habla; son importantes los conocimientos que se tengan sobre el mensaje que se quiere transmitir, pero lo más importante es que sea

creíble para el público que lo escucha.

La Oratoria tiene tres componentes: estilo, contenido e impacto. Es decir, deben ser textos agradables tanto para escuchar y leer que se centren en un tema e inspiren los ideales y valores de la audiencia, pues siempre se busca persuadir a ésta sobre algún hecho o idea para cambiar sus mentes y corazones [?].

En vista de que la Oratoria permite utilizar todos los recursos que aporta la lengua, la lingüística y la dicción para transmitir un mensaje, hay que resaltar la importancia de las partes de la comunicación de un mensaje:

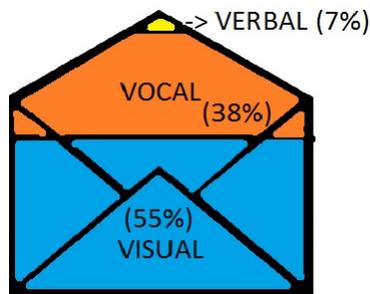


Figura 1.1: Componentes de la comunicación en la transmisión de un mensaje.

Fuente: *Elaboración propia.*

Según [?], el 7% del mensaje se transmite por las palabras que se usan (lo que se dice), el 38% se transmite por la forma en que se dice, es decir, por la voz, pues a través de ella se agrega énfasis y significado. Si se desea transmitir emoción, el habla es rápida y si se desea transmitir sinceridad, el habla es lenta. El 55% se transmite por lo que el cuerpo dice, pues la audiencia cree más en lo que ve que en lo que escucha (ver Figura 1.1).

Es por esto que cada palabra que se usa es importante y se deben usar lo mejor posible para lograr el mayor impacto. Éste puede ser medido de diferentes formas, sin embargo, en el presente trabajo se busca medirlo con herramientas de matemáticas, específicamente de la teoría de redes. Cabe mencionar que, el estudio de los lenguajes y específicamente de las palabras, con herramientas de matemáticas no es nuevo y en la siguiente sección se evidencia este hecho.

1.3. Las matemáticas de los lenguajes

El primer hallazgo matemático del lenguaje es anónimo y se reduce a que es directamente análogo de los números naturales pues no existe “la oración más larga”, es decir, no hay un límite en la longitud de la misma.

Posteriormente, hace alrededor de 150 años, se empezó el estudio de la longitud de las palabras. El primero en hacerlo fue el matemático Augustus De Morgan el 18 de Agosto de 1851 cuando se le ocurrió que la longitud de las palabras podría ser un indicador del estilo de cada escritor, es decir, podría ser un factor que determinara la autoría. Si se generalizan sus ideas, el promedio de las longitudes de las palabras en dos textos escritos por el mismo autor, aún de temas diferentes, es más similar que en dos textos escritos por autores diferentes pero sobre el mismo tema.

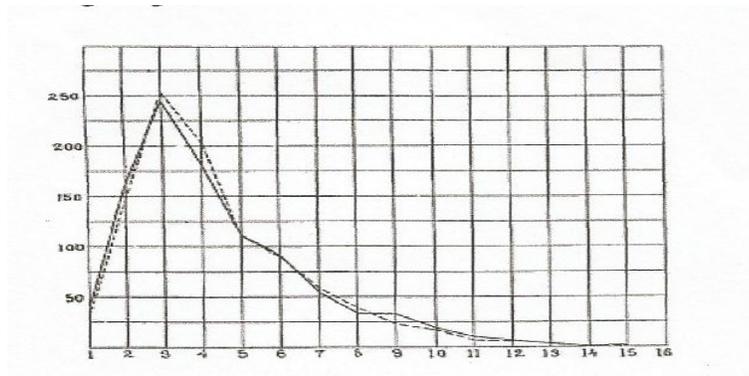


Figura 1.2: Curva obtenida con las frecuencias de la longitud de las palabras en Oliver Twist de Charles Dickens (Mendenhall 1887).

Fuente: [?]

Algunas décadas después, el físico y meteorólogo Thomas Corwin Mendenhall brinda las primeras pruebas empíricas sobre las asunciones de De Morgan, añadiendo que también se debía considerar la frecuencia y no sólo el promedio de las longitudes. Analizó textos de 1,000 palabras de diversos autores como Charles Dickens, y su última aportación fue la *curva normal del escritor* o *curva característica* (en inglés, *word spectrum*), que es una representación gráfica de un arreglo de palabras según su longitud y su frecuencia relativa de ocurrencia (ver Figura 1.2). Sus investigaciones abarcaron idiomas como el francés, alemán, italiano, latín y español, además de realizar los mismos estudios pero con sílabas de las palabras en las oraciones.

Lo que Mendenhall hizo es lo que hoy en día se conoce como análisis de frecuencias o análisis de distribución de frecuencias. En esos tiempos, las matemáticas que conllevaban comparar las distribuciones de frecuencias eran muy escasas y poco entendidas. Es hasta 1976, que Williams señala una falla metodológica en los trabajos de Mendenhall pues éste no consideró que existían diferencias más allá de la autoría en los cálculos, por ejemplo, si uno de los textos estaba en prosa y el otro en verso blanco¹, dejando así claro que la longitud no necesariamente es característica del estilo del autor, más bien, la longitud y la frecuencia de las palabras dependen de otros factores, como por ejemplo, el género [?].

Mendenhall nunca intentó hallar un modelo matemático formal que describiera la distribución de las frecuencias y como consecuencia, ninguna comparación objetiva entre las distribuciones empíricas y teóricas había sido posible, pero se le debe atribuir el haber establecido una base empírica para la investigación de la longitud de las palabras y con ello, haber iniciado toda una línea de investigación que sigue siendo importante en estos tiempos.

¹Poesía escrita en líneas métricas pero sin rima regular y casi siempre en pentámetros.

En el siguiente periodo histórico en los estudios sobre la longitud de las palabras, se propusieron modelos de distribución particulares y se desarrollaron métodos para probar la bondad de los resultados obtenidos, siempre asumiendo que debía existir algún principio organizacional general del que se derivaban ciertos modelos. Los primeros modelos se desarrollaron a finales de los años cuarenta del siglo XX y después la investigación se concentró en dos modelos: la distribución Poisson y la geométrica. Después, de mediados de los años cincuenta en adelante, la distribución Poisson se sometió a varias modificaciones y generalizaciones.

El primer intento de obtener una descripción matemática de la distribución de frecuencias de la longitud de las palabras se le atribuye a Sir William P. Elderton, quien realizó sus estudios en pasajes de algunos escritores ingleses, entre ellos, Gray, Macaulay y Shakespeare. Midió la longitud como el número de sílabas y no el número de letras por palabra. Además de contar las frecuencias y representarlas en forma gráfica, Elderton buscó un modelo estadístico que describiera esas distribuciones que asumía debían seguir la distribución geométrica. Publicó sus resultados en 1949 en el artículo *A Few Statistics on the length of English words*. En pruebas que se han realizado recientemente [?] ha resultado que el ajuste es bueno, sin embargo, solamente es válido para algunos textos en inglés pues la distribución geométrica es adecuada para describir distribuciones monótonas decrecientes y no todos los lenguajes presentan esta característica.

En 1972, Merkyté analizó material elegido al azar de un diccionario lituano asumiendo que hay dos tipos de palabras, raíces y afijos, además, supuso que la distribución de sílabas en las raíces era descrita por una ley geométrica como caso especial de una distribución binomial negativa con $k = 1$, mientras que la de los afijos sería una binomial. El resultado principal al que llega es que las frecuencias de las longitudes no pueden ser explicadas por un sólo proceso, sino que son organizadas por dos procesos que se traslapan, sin embargo, existía un problema en sus investigaciones, pues la combinación de ambos tipos de palabras no necesariamente sigue la composición de las dos distribuciones individuales [?].

Sergei Grigorevich Chebanov se concentró en el proceso del desarrollo de lenguaje e investigó 127 diferentes lenguajes y dialectos de la familia indoeuropea durante veinte años. Asume que la esperanza es una relación entre la longitud media y las frecuencias relativas, así que opta por el modelo Poisson. En los tres textos analizados, obtuvo un ajuste excelente, aunque desafortunadamente el trabajo se ha quedado en el olvido.

Otro modelo teórico fue propuesto por Gustav Herdan y René Moreau, a finales de los cincuenta en el siglo XX y a diferencia de los otros trabajos hechos, consideraron un modelo continuo, la distribución lognormal. La idea fue antes propuesta por Williams. Herdan para probar su hipótesis, únicamente usó técnicas gráficas, específicamente *p-p plots*. La longitud fue medida por letras y fonemas y los resultados fueron convincentes. Moreau propone un tercer parámetro para la distribución pero Herdan no lo considera necesario. Resulta que de ninguna forma es adecuado en la práctica y Michel en 1982, llega a la misma conclusión.

El siguiente gran paso en el estudio de la longitud de las palabras, fueron los análisis teóricos y empíricos del físico alemán Wilhelm Fucks, que basado en distintos supuestos teóricos comparados con los hechos por Chebanov, llegó a conclusiones similares. Este modelo fue el principal entre 1950 y 1970 y la mayor diferencia radica en que el modelo de Fucks era más general y el otro era sólo un caso en particular. Además, desarrolló ideas importantes sobre el funcionamiento, no sólo del lenguaje sino también de otros sistemas de señales.

Por otra parte, hay un estudio muy importante sobre la frecuencia de las palabras. Fue desarrollado alrededor de 1935 por el lingüista de Harvard, George Kingsley Zipf. En ese estudio, afirma que hay un pequeño número de palabras que son utilizadas con mucha frecuencia, mientras que un gran número de palabras son poco empleadas. Dado un texto, se hace una tabla con las palabras y su respectiva frecuencia $f(r)$ y se ordena según la frecuencia de más a menos. Este orden en que aparece cada palabra se le llama rango y se denota por r . Entonces, la ley de Zipf indica que el número de veces que aparece una palabra $f(r)$ es inversamente proporcional a su rango r , de forma que:

$$f(r) = \frac{1}{r^\alpha} \quad (1.1)$$

y demostró en *La psico-biología del lenguaje: Introducción a la filología dinámica* [?] que la frecuencia de palabras exhiben una ley de potencia en términos del rango con $\alpha \approx 1$. Otra forma de calcular la Ley de Zipf, es contar cuantas veces aparece una palabra y dividirla entre el número total de palabras del texto.

A través de este recuento histórico de los principales estudios matemáticos realizados al lenguaje, es evidente que se ha tratado de identificar componentes o tendencias para explicarlo a lo largo del tiempo, y si a eso se le añade que se ha empezado a apreciar científicamente el poder de las palabras para formar la identidad de la persona y la forma en que se percibe el mundo [?, ?], el interés por estudiarlo aumenta.

Capítulo 2

Redes y sus propiedades elementales

Las redes están presentes en el día a día de muchas personas, se esté o no consciente de ellas. Se revisa el correo electrónico, se hacen llamadas telefónicas, se usa el transporte público, etc., y en todos estos casos, se hace uso de las redes y de sus propiedades, es decir, se tiene un conjunto de diferentes elementos (por ejemplo, individuos, computadoras, automóviles) conectados a través de las distintas interacciones que hay entre ellos. Para representar un sistema como una red, es necesario un alto nivel de abstracción y se debe tener en cuenta que lo que se pierde de detalle del sistema, se gana en lo universal de la forma, esto es, en pensar diferentes sistemas como realizaciones de la misma estructura teórica: la red.

2.1. Historia

El concepto de gráfica fue desarrollado por el matemático Leonhard Euler, que en 1736 buscaba responder a la pregunta que los ciudadanos de la antigua ciudad de Königsberg (hoy Kaliningrado, Rusia) se hacían, ¿sería posible pasar únicamente una vez por cada uno de los siete puentes que unían la ciudad con la isla de Kneiphof?. En su afán de conseguirlo, representó a la ciudad de forma en que las porciones de tierra eran círculos y los puentes que las unían eran líneas; generalmente, los especialistas nombran los círculos como nodos o vértices y las líneas como arcos o enlaces. Este sencillo dibujo permitió dar solución al problema: era imposible cruzar únicamente una vez por los puentes ya que, salvo los nodos de partida y llegada, el resto debía tener un número par de arcos, es decir, uno de llegada y uno de salida y en la Figura 2.1, se observa que cada nodo tenía un número impar de arcos. Los puentes de Königsberg constituyen la base de toda una rama de las matemáticas basada en el análisis de gráficas.

A partir de entonces, muchos matemáticos han estudiado las propiedades formales de las gráficas y muchos son los científicos que las han aplicado para solucionar diversos problemas en diferentes áreas, por ejemplo, a circuitos eléctricos (G.R. Kirchhoff

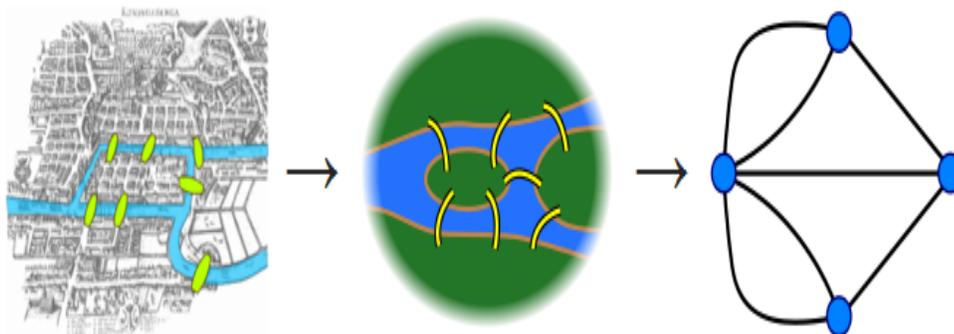


Figura 2.1: Los puentes de Königsberg representados por Leonard Euler como una gráfica.

Fuente: [?]

en 1845), a isómeros de componentes orgánicos (A. Cayley en 1857), entre otros [?]. Sin embargo, el paso crucial en el desarrollo de la ciencia de las redes se da hasta 1932, cuando a Jacob Moreno y Helen Jennings se les ocurre aplicar esta idea a la sociedad. Así, ayudándose de la sociometría, logran explicar porqué 14 niñas huyeron de una escuela secundaria en Nueva York.

Entre 1959 y 1961, Paul Erdős y Alfréd Rényi estudiaron una red en la que consideraron la probabilidad con la que los vértices estaban conectados, este modelo matemático es conocido como red aleatoria y constituye el segundo paso importante en el desarrollo de esta ciencia. En primera instancia esta red fue propuesta en 1951 por Ray Salomonoff y Anatol Rapoport. Este tipo de redes se usa principalmente para comparar aquellas del mundo real con su equivalente aleatorio y poder detectar así la presencia de algún tipo de orden, estructura o grado de aleatoriedad. Su aportación principal consiste en que una vez que la mayoría (entre el 90 y 95 por ciento) de los nodos están conectados entre sí, se forma el componente gigante y basta con su estudio para conocer el comportamiento de toda la red.

Sin embargo, es hasta 1967 que se puede ubicar otro paso crucial en esta ciencia. Stanley Milgram y Jeffrey Travers enviaron decenas de cartas al azar en Kansas y Nebraska pidiendo que mandaran la carta a una persona en Massachusetts, si la conocían, pedían que la carta se mandara directamente pero si no la conocían, debían mandar la carta a alguien que pensaran que podría conocerla. Por cada carta enviada, una debía ser enviada a Milgram para seguir el proceso. Se obtuvo que una tercera parte de las cartas llegó al destino y que ninguna carta fue mandada más de diez veces, además, para que la carta llegara de un lado a otro, en promedio bastaba con mandar 5.5 cartas. De este experimento surge la frase “seis grados de separación” usada incluso en 1990 como título de una película. El resultado principal que obtuvieron es la propiedad de mundo pequeño, que indica que entre cada par de nodos hay una distancia pequeña que

los separa y ésta crece lentamente conforme el número de nodos aumenta.

La propiedad de mundo pequeño fue retomada en 1998, cuando Duncan Watts y Steven Strogatz proponen el modelo de una red de mundo pequeño que satisface la propiedad de mundo pequeño y tiene un alto valor de transitividad [?].

Un año más tarde, Albert-László Barabási y Réka Albert proponen las redes de libre escala en las que son pocos nodos los que tienen una gran cantidad de enlaces, debido a que la naturaleza de estas redes permite que exista la preferencia de enlace. Es decir, si aparece un nuevo nodo, éste tiende a enlazarse con los nodos que ya tienen muchos enlaces y estos con el tiempo, adquirirán más y más enlaces, en pocas palabras, “el rico se hace más rico” [?].

Otro escalón importante en el desarrollo de la ciencia de las redes, ocurre porque se empiezan a utilizar para analizar sistemas complejos. Así, una red compleja no es más que la caracterización de un sistema complejo, donde las partes se representan con nodos y la relación entre ellas, con arcos.

2.2. Definiciones y principales modelos de redes

Un conjunto de nodos unido por arcos es solamente la forma más simple de representar una red. Pueden tener más de un tipo de vértices y/o arcos y además, los arcos pueden tener un peso asociado o tener dirección. Algunas redes pueden también evolucionar a lo largo del tiempo, con vértices o arcos naciendo o desapareciendo y/o sus valores asociados cambiando ¹.

El análisis de redes se divide en dos campos que se complementan [?]: estructura y dinámica. En el primero, se estudian las propiedades estructurales o topológicas de la red, es decir, el objeto es estudiar cómo están conectados los nodos unos con otros y con respecto a la dinámica de la red, se estudian las interacciones entre nodos, sin embargo, en el presente trabajo, se abordan más las propiedades estructurales que las dinámicas. Para poder comenzar el análisis estructural de una red es necesario definir algunos conceptos.

El número de arcos de un nodo se conoce como *grado*, entonces, k_i es el grado del i -ésimo nodo, y en redes con dirección, es posible definirlo para los arcos de salida del nodo (*out*) y para los arcos que llegan al nodo (*in*). A los nodos que tienen una gran cantidad de enlaces se les llama *hubs* y por lo general, en una red real son pocos. Asociado a este concepto, está la *distribución de grado* $P(k)$ que es la probabilidad de que un nodo escogido al azar, tenga k conexiones. En otras palabras, es el porcentaje de vértices en la red que tienen grado k , respecto al total de vértices de la red.

Cuando en una red se tiene que nodos de bajo grado se enlazan con nodos de bajo

¹Existen diferencias en las definiciones de algunos conceptos en la teoría de redes y la teoría de gráficas, como por ejemplo, la forma de llamar a los nodos en teoría de gráficas es vértices. Para mayores referencias consultar [?, ?].

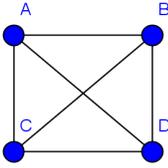
grado, y *hubs* se enlazan con *hubs*, se dice que la red es asortativa. La *asortatividad*, en inglés, *assortative mixing* es una forma de homofilia en la que los nodos tienden a conectarse con otros similares a ellos en el número de conexiones [?]. Se le conoce también como correlación grado-grado. El número resultante está en el intervalo $[-1, 1]$; si los nodos están correlacionados, la red es asortativa y toma el valor 1. Si no hay correlación, la red es disortativa y toma el valor -1. Se dice que la red es neutral si el valor resultante es cero.

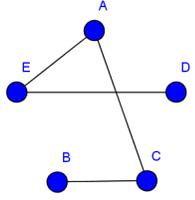
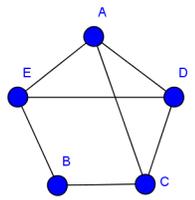
La *ruta más corta* es la menor cantidad de nodos que se deben visitar desde un nodo i hasta llegar a cualquier otro nodo j . La *distancia media* es el promedio de rutas más cortas en una red entre cada par de nodos y su inverso, la *eficiencia*, representa qué tan rápido es posible mover la información dentro de la red. La propiedad de mundo pequeño o seis grados de separación, que indica que hay una distancia pequeña entre cada par de nodos, está relacionada con la distancia media y para una red en la que N es el número de nodos, esta distancia es aproximadamente $\ln(N)$ como lo probaron Watts y Strogatz en [?].

Por otra parte, si un nodo A está enlazado con un nodo B y un nodo B con un nodo C , entonces es muy alta la probabilidad de que A y C estén enlazados formando un triángulo. El *coeficiente de transitividad o aglomeración*, conocido en inglés como *clustering*, mide esta probabilidad por lo que resulta ser un número entre 0 y 1.

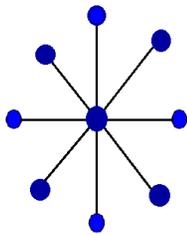
Una vez definidos los conceptos anteriores, es posible clasificar las redes en básicamente cuatro modelos: redes regulares, aleatorias, de mundo pequeño y de libre escala. Las características principales de cada uno se detallan en el Cuadro 2.1.

Cuadro 2.1: Principales modelos de redes y algunas de sus características

Modelo de red	Características	Ejemplos
Regulares 	Todos los nodos tienen el mismo grado, es decir, el mismo número de arcos. Presentan un valor de <i>clustering</i> igual a 1, pero no cumplen la propiedad de mundo pequeño. En muchos casos, es el modelo base para iniciar un estudio.	Representar una familia o un grupo de amigos.

Modelo de red	Características	Ejemplos
<p data-bbox="365 394 479 422">Aleatorias</p> 	<p data-bbox="604 394 989 615">Tienen más o menos el mismo número de arcos, se forma el componente gigante cuando entre el 90 y 95 por ciento de los nodos están enlazados. $P(k)$ tiende a una <i>Poisson</i> en redes grandes. Presentan la propiedad de mundo pequeño.</p>	<p data-bbox="1011 394 1318 489">Se usan para hacer comparaciones con las redes reales del sistema.</p>
<p data-bbox="365 1020 578 1052">De mundo pequeño</p> 	<p data-bbox="604 1020 989 1272">La mayoría de los nodos no están conectados entre ellos pero pueden ser alcanzados desde cualquier otro en un número pequeño de saltos, es decir, cumple la propiedad de mundo pequeño (distancia media aproximadamente $\ln(N)$). Presentan un valor de <i>clustering</i> entre 0.5 y 1.</p>	<p data-bbox="1011 1020 1179 1052">Redes sociales.</p>

Modelo de red	Características	Ejemplos
De libre escala	Unos cuantos nodos dominan la red, es decir, hay presencia de pocos <i>hubs</i> . Su distancia media es muy pequeña $\ln(\ln(N))$ y busca representar la preferencia de enlace, mejor conocida como el fenómeno “del rico se hace más rico” [?]. La distribución de grado sigue una ley de potencia.	Redes biológicas y tecnológicas.



2.3. Redes complejas

Se dice que un sistema es complejo si cumple ciertas características, como por ejemplo, debe cumplir con la auto-organización, interdependencia, independencia y emergencia. La auto-organización indica que las interacciones entre los componentes del sistema se dan de forma espontánea permitiendo que se tenga un orden sin que intervenga algún agente externo dentro o fuera del sistema, en el caso de los lenguajes, es posible pensar en las palabras que cambian con el paso del tiempo o bien, en aquellas que desaparecen y no por ello, el lenguaje deja de usarse o surgen problemas de comunicación. La interdependencia significa que si una parte del sistema es removida, se puede afectar otra parte, por ejemplo, en el lenguaje, si se quitaran los verbos o algo aparentemente menos importante como los artículos, sería imposible expresarse de forma correcta, al igual que expresarse con únicamente alguna de estas partes del lenguaje. En cuanto a la característica de independencia, se refiere a que las partes del sistema tienen la libertad de responder a las demandas propias del sistema, pero solamente si las partes están relacionadas entre sí, funcionarían de forma correcta, es decir, las palabras deben responder las demandas de un texto, por eso es posible formar oraciones nuevas con palabras que son conocidas pero éstas son correctas si, son colocadas de forma que se respeten las reglas gramaticales. En los sistemas complejos, el todo no es la simple suma de las partes que lo componen [?], esta propiedad se conoce como emergencia y en el caso de los lenguajes, un buen ejemplo es que a pesar de que existen verbos, artículos, preposiciones, sustantivos, etc., para que funcione correctamente el sistema, son necesarias las reglas gramaticales y cada una de sus partes.

En vista de que los sistemas complejos pueden ser modelados y por tanto, analiza-

dos mediante redes complejas, resulta necesario hacer énfasis en que las definiciones y los modelos explicados en la sección anterior son válidos en este tipo de redes y por lo tanto, las propiedades más importantes que determinan la estructura de la red son:

- a) Distribución de Grado $P(k)$.- en el caso de los lenguajes, $P(k)$ es libre escala debido a que hay palabras que se usan más que otras, por ejemplo, los artículos y las preposiciones (*hubs*).



Figura 2.2: En la gráfica se observa que son pocos los nodos que tienen muchos arcos, es decir, pocos *hubs*.

Fuente: *Elaboración propia*.

La Distribución de Grado $P(k)$, para una red con N nodos, se define como la fracción de nodos de la red, que tienen grado k y queda expresada como:

$$P(k) = \frac{N_k}{N} \quad (2.1)$$

Donde N_k es el número de nodos de grado k de la red.

- b) Aglomeración o *clustering* C .- es la probabilidad de que dos nodos enlazados directamente a un tercero estén conectados entre sí cuando la red se forma por sonidos [?]. Para los lenguajes, si se construye una red por sonidos similares, puede interpretarse como, si dos palabras suenan parecido, la tercera palabra será similar a cualquiera de las otras dos, un ejemplo sencillo son los trabalenguas, o bien, si el significado de la palabra P es similar al significado de la palabra Q , para cualquier otra palabra R con significado similar a P , tendrá significado similar también a Q . Para el nodo i , con grado k_i , se define esta propiedad como:

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \quad (2.2)$$

Donde L_i es el número de enlaces entre los k_i vecinos del nodo i .

Para obtener el *clustering* de la red, se utiliza el *clustering promedio* $\langle C \rangle$, definido como sigue:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i \quad (2.3)$$

En donde C_i es el *clustering* del nodo i .

- c) Distancia media $\langle d \rangle$.- debido a que es el promedio² de las longitudes de las rutas más cortas en una red entre cada par de nodos, es posible formar frases correctas y rápidas con poco vocabulario, como en el caso de los niños pequeños. En redes no dirigidas, $\langle d \rangle$ queda definida como:

$$\langle d \rangle = \frac{1}{L_{max}} \sum_{i,j>1} d_{ij} \quad (2.4)$$

En donde d_{ij} es la longitud entre el nodo i y el nodo j y L_{max} es el número mayor de enlaces de una red con N nodos y:

$$L_{max} = \frac{N(N-1)}{2} \quad (2.5)$$

- d) Eficiencia E .- en el caso de los lenguajes, mientras más léxico se adquiera, más eficiente se vuelve la comunicación porque las rutas para llegar de una palabra a otra, son más cortas. Se puede definir como:

$$E = \frac{1}{L_{max}} \sum_{i,j>1} \frac{1}{d_{ij}} \quad (2.6)$$

- e) Asortatividad ρ .- para los lenguajes, podemos pensar en palabras que se usan mucho como las conjunciones, que se enlazan con preposiciones (red asortativa) y las preposiciones que se unen con otras palabras menos usadas, para lograr formar frases completas (red disortativa), como se muestra en la Figura 2.3.

²En diferentes artículos, se considera también que la distancia media es la longitud de la ruta más larga, y se define tanto al promedio como a la ruta más larga con el término *diámetro*. Esto se debe a que en redes que tienden al infinito, el promedio y la ruta más larga son muy parecidos [?].

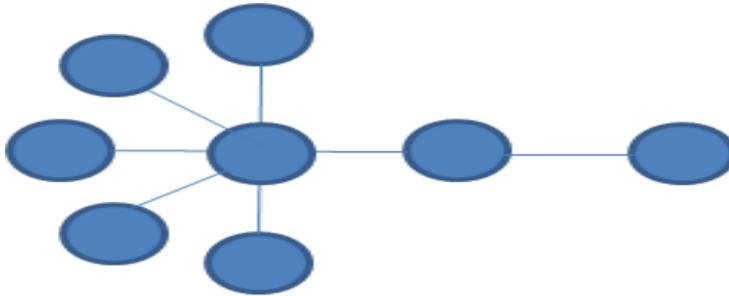


Figura 2.3: Se aprecia como un *hub* se une a un nodo de menor grado, este caso representa una red disortativa.

Fuente: *Elaboración propia.*

Se define como el coeficiente de correlación de Pearson de los grados de aquellos nodos conectado por un enlace [?]. Se expresa como sigue:

$$\rho = \frac{\langle k_i k_j \rangle - \langle k_i \rangle \langle k_j \rangle}{\sqrt{(\langle k_i^2 \rangle - \langle k_i \rangle^2)(\langle k_j^2 \rangle - \langle k_j \rangle^2)}} \quad (2.7)$$

donde k_i y k_j son los grados de los nodos extremos del enlace, y $\langle . \rangle$ significa el promedio de todos los enlaces.

Capítulo 3

Redes de lenguaje

Un lenguaje es un sistema compuesto por reglas gramaticales que se aplican a un vocabulario o léxico, en el que la unidad elemental son las palabras, y la secuencia de éstas permite la transmisión de información o ideas [?]. En la sección anterior, se explicó el por qué un lenguaje es un sistema complejo y, aunque resulte contradictorio, se piensa que la lengua que se habla es más sencilla y que son únicamente los idiomas que no se conocen, los que se consideran de alta complejidad, sin embargo, todos los lenguajes son complejos y se han realizado estudios para comprobarlo, como referencia [?].

Entonces, es posible construir una red con palabras como nodos si se identifica la relación que las une, estas redes son las que se conocen como redes de lenguaje. La relación definida no es única por lo que la red tampoco lo es. Por ejemplo, la relación puede establecerse si las palabras son sinónimos, antónimos o si las palabras suenan parecido aún con significados diferentes [?], como se puede observar en la Figura 3.1. Otra forma podría ser por la posición que ocupan dentro del texto, es decir, existe un enlace con la palabra inmediata previa y posterior gracias a las reglas gramaticales. Al considerar esta última forma de relacionar las palabras, según Guido Caldarelli en [?], es posible entender mejor el significado de una palabra, pues se toman en cuenta sus vecinos en el texto. Aun así e incluso, a pesar de contener mucha información, las redes de lenguaje no son útiles para estudiar el contenido de los textos, ni tampoco las relaciones entre las ideas expresadas en textos diferentes.

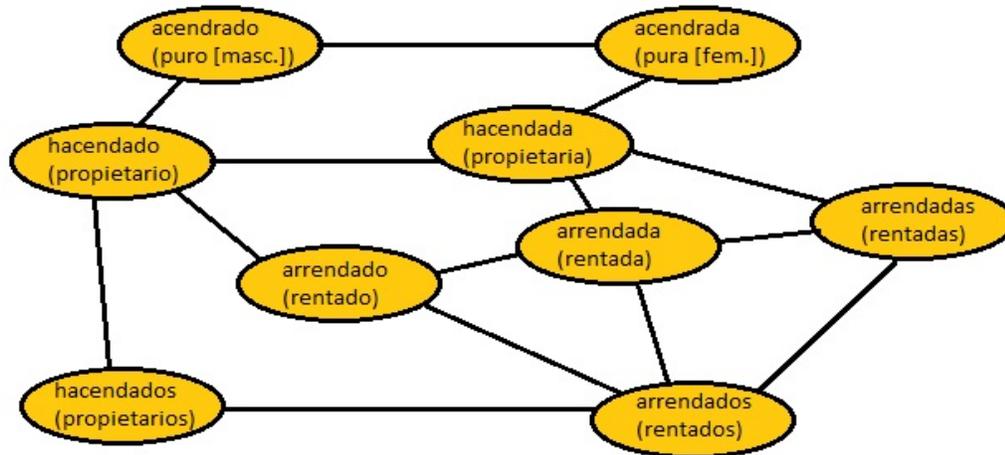


Figura 3.1: Se presentan nueve palabras con el significado de cada raíz debajo. Suenan similar pero no necesariamente tienen significados similares.

Fuente: [?]

3.1. El lenguaje como serie de tiempo

También es posible estudiar un lenguaje como una serie de tiempo. Si una variable es medida secuencialmente en el tiempo o en intervalos predeterminados conocidos como intervalos de muestra, los datos obtenidos forman una serie de tiempo. En otras palabras, el término hace referencia a datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares que en el caso de lenguajes, se interpreta como el registro de una palabra por unidad de tiempo. Las series de tiempo suelen ser representadas por gráficas de líneas o histogramas en los que los intervalos de tiempo se representan en el eje horizontal y en el eje vertical, se representan los valores de la serie.

La idea de convertir series de tiempo en redes resulta atractiva porque representa la unión de dos campos de la ciencia moderna: el análisis no lineal de series de tiempo y la teoría de redes complejas. A lo largo de los años, muchos investigadores se han dedicado a crear formas para construir redes a partir de estas series, por ejemplo, Zhang & Small en 2006, Xu y colaboradores en 2008, y recientemente, Campanharo en 2011 y Donner entre 2010 y 2011 [?]. Entre todos estos métodos de construcción, el *algoritmo de visibilidad*, desarrollado por Lucas Lacasa y colaboradores en 2008 [?], se abordará a detalle en la siguiente sección.

3.2. Algoritmos de visibilidad

Existen varios algoritmos asociados al término visibilidad, aunque en realidad, es posible hablar de uno sólo y definir el resto como sus variantes. El algoritmo principal se conoce como el algoritmo natural de visibilidad o NVG por sus siglas en inglés [?] y se propuso para transformar series de tiempo irregulares en redes con la idea de analizar la complejidad de la serie original, con la ayuda de las metodologías de la teoría de redes. Del NVG se derivan dos modalidades, el algoritmo horizontal y el vertical, que se pueden consultar en [?].

Algoritmo natural de visibilidad (NVG)

Paso 1.- Cada palabra debe ser asociada al número correspondiente al tiempo en que aparece por primera vez en el texto, así, las palabras que aparecen más de una vez (palabras repetidas) solamente tendrán un número asociado.

Paso 2.- Se debe definir la regla de asignación de valores de la serie que puede ser por ejemplo, el número de letras que la forman (longitud) para entonces, realizar una gráfica de barras en la que la serie de palabras queda representada en el eje horizontal según el orden de ocurrencia del texto, y la longitud correspondiente a cada una de estas palabras se expresa en el eje vertical. Así, queda definido un nodo por cada elemento de la serie, es decir, por cada palabra y entonces, la primera palabra del texto será el nodo 1 y quedará caracterizada por la pareja ordenada $(1, l_1)$ donde l_1 es la longitud de esa palabra; la segunda palabra del texto será el nodo 2 y la pareja ordenada correspondiente será $(2, l_2)$ con l_2 representando la longitud de la palabra 2, y así sucesivamente.

Paso 3.- Se selecciona un nodo y éste será enlazado con aquellos que sean representados por barras que puedan ser vistas desde la cima de la barra correspondiente al nodo seleccionado, obteniendo así la red asociada a la serie de la longitud de las palabras, como se puede observar en la Figura 3.2. En esta red cada nodo corresponde, en el mismo orden, a los datos de la serie y dos nodos están conectados si existe visibilidad entre ellos, es decir, si existe una línea recta que conecte ambos datos sin que interseque algún otro dato. Formalmente, el criterio de visibilidad queda establecido como sigue: dos datos arbitrarios (t_a, y_a) y (t_b, y_b) tendrán visibilidad y por ende, estarán conectados en la red asociada si cualquier otro dato (t_c, y_c) que se encuentra entre ellos, es decir, $t_a < t_c < t_b$ satisface:

$$y_c < y_b + (y_a - y_b) \frac{t_b - t_c}{t_b - t_a} \quad (3.1)$$

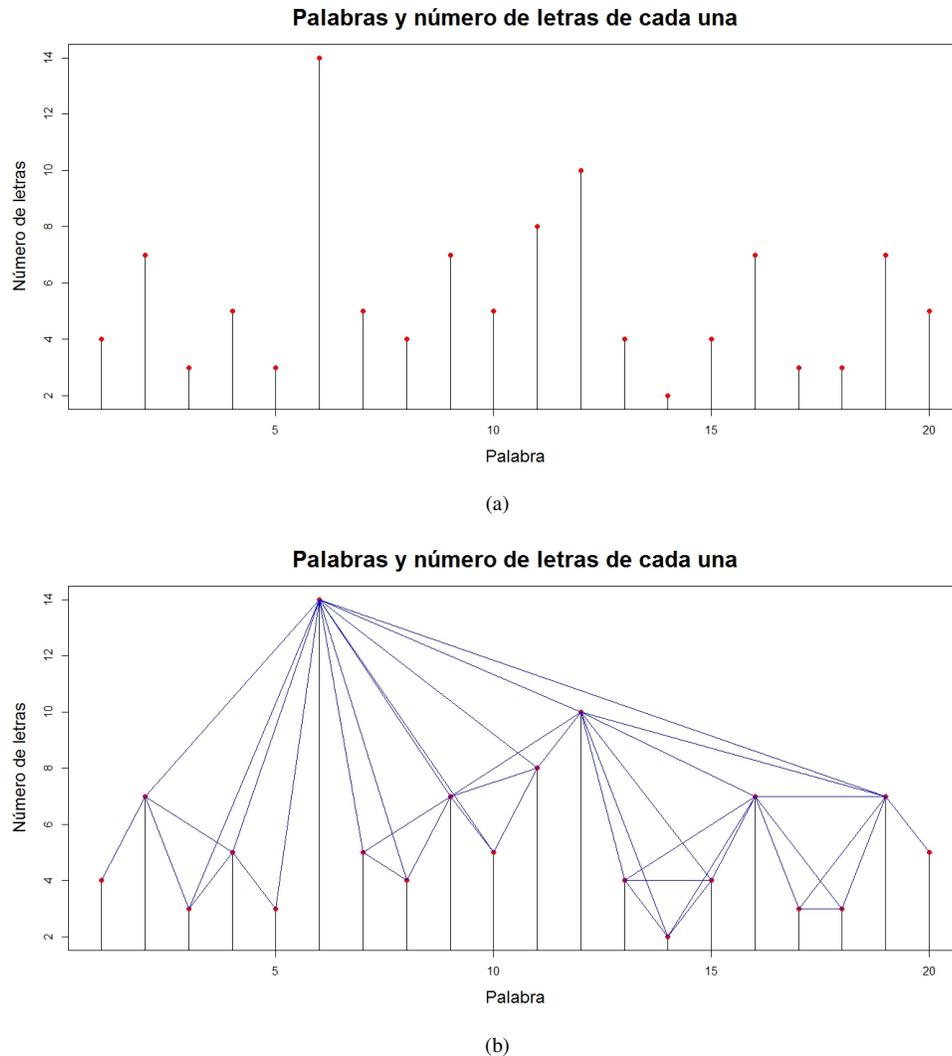


Figura 3.2: (a) Las barras representan la longitud de la secuencia de palabras. (b) Aplicación del algoritmo de visibilidad a la secuencia de longitudes de palabras. Dos palabras están conectadas si existe "visibilidad entre ellas".

Fuente: Elaboración propia.

En [?], se ha probado que la estructura de la serie de tiempo se conserva en la red, esto es, las series periódicas y estocásticas se convierten en redes libre escala; además, se ha comprobado que las redes construidas a partir de una serie de tiempo mediante el NVG son siempre:

a) Conectadas \rightarrow un nodo está enlazado al menos con sus vecinos más cercanos tanto

a la izquierda como a la derecha.

- b) No dirigidas → por construcción del algoritmo.
- c) Invariantes bajo transformaciones de los datos de la serie → el criterio de visibilidad es invariante tanto al cambiar la escala de los ejes horizontal y vertical, como para translaciones en cualquiera de estos ejes (Ver Figura 3.3).

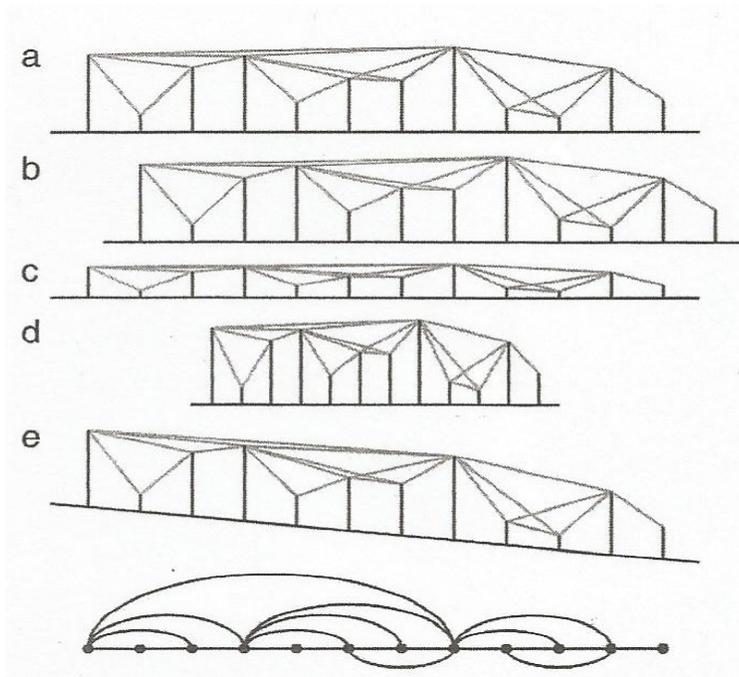


Figura 3.3: (a) Serie de tiempo original con enlaces de visibilidad. (b) Translación de los datos. (c) Reescalamiento vertical. (d) Reescalamiento horizontal. (e) Se añade una línea de tendencia a los datos. Como se puede observar en la red formada y mostrada al final, en todos estos casos, la red de visibilidad permanece invariante.

Fuente: [?]

Capítulo 4

Metodología del estudio

En las siguientes páginas, se describe el diseño de la investigación y cómo fue posible llevarla a cabo en la práctica. Se justifica la elección de métodos y técnicas así como, los supuestos que se asumieron para la realización de este estudio.

En cuanto a la elección de los personajes, se ha cubierto una amplia variedad de actividades y vocaciones, de países y de épocas históricas, aunque la mayoría de ellos, se ubican en el siglo XX. El criterio usado en este trabajo, fue que los personajes debían ser oradores o, en su defecto, compositores que a través de sus discursos o canciones, hallan marcado tendencia y, por tanto, dejado marca en la historia mundial. En el anexo A se puede consultar una breve biografía de los personajes que por sus aportaciones en música, literatura, política, ciencias, entre otras, fueron considerados para este estudio.

4.1. Metodología

Para llevar a cabo la investigación, fue necesario dividir el trabajo en varias etapas que se detallan a continuación:

1.- Elaboración del programa computacional del algoritmo natural de visibilidad (NVG)

En la Sección 3.2 se describió el NVG, cuya aportación principal es la construcción de una red a través de la transformación de series de tiempo irregulares para poder analizarlas posteriormente, con herramientas y metodologías propias de la teoría de redes. Así que, el primer paso de la investigación fue programar este algoritmo en la plataforma *RStudio*, considerando la longitud de la palabra como la cantidad de letras que la forma.

Además del algoritmo, también se hizo uso de las librerías *rgl*, para producir gráficas interactivas, *igraph*, que permite manejar redes de gran tamaño y proporciona funciones y algoritmos para el análisis de redes, y herramientas propias de *R*, para calcular algunas métricas importantes mencionadas en la Sección 2.3. En caso de querer consultar el código del algoritmo NVG, referirse al Anexo B y se encuentra explicado paso a paso para facilitar su comprensión.

2.- Selección de textos para el análisis Los textos en que se basa esta investigación, consisten en discursos, entrevistas o canciones, que fueron escritas por cada personaje mencionado en la sección anterior. En el caso de las canciones, se consideraron algunas que son representativas de dichos cantantes, a pesar de no ser propietarios o copropietarios de ellas.

Se consideró para la selección de dichos textos, es decir, para seleccionar las canciones, discursos y entrevistas, a aquellos escritos en los momentos más trascendentes en la vida pública de cada personaje, así como, a aquellos textos en los que se abordaron temas polémicos en materia política o social. La mayoría de los textos se obtuvieron de las páginas oficiales de cada uno y de las organizaciones que fundaron, por lo que fueron considerados en el idioma original de cada personaje. En caso de querer consultar los textos completos, se puede recurrir al Anexo C en el que se detalla de qué libro o página web fueron obtenidos.

Con base en los trabajos de Mendenhall discutidos en la Sección 1.3, se buscó que los textos utilizados tuvieran al menos una base de mil palabras, esto debido a la asunción estadística en el área de lingüística de que se estaría en presencia de una muestra bien definida, y entonces, sería posible y correcto hacer inferencias sobre el total. Sin embargo, al ser un número de palabras muy grande para una entrevista o una canción, esta condición únicamente se verifica en la mayoría de los discursos analizados.

En el Cuadro 4.1 se presenta el nombre de cada personaje, el nombre e identificador de cada texto considerado, así como el número de palabras de cada uno, con y sin repeticiones.

Cuadro 4.1: Textos considerados por autor y número de palabras de cada uno de ellos

Personaje		Texto	Palabras	Palabras sin repeticiones
Mahatma Gandhi	MG1	Quit India	1,103	420
	MG2	Eve	1,064	434
	MG3	Dandi March	1,003	395
Winston Churchill	WC1	Alliance with Russia	2,020	773
	WC2	Blood, Toil, Tears and Sweat	700	310
	WC3	Be Ye Men of valour	1,359	543
	WC4	The Sinews of peace	4,860	1,383

Personaje		Texto	Palabras	Palabras sin repeticiones
Albert Einstein	AE1	Open Letter to the General Assembly of the United Nation	1,379	511
	AE2	The war is won but peace is not	1,014	454
	AE3	Why do they hate the Jews?	2,784	956
Adolf Hitler	AH1	Before the Reichstag	15,866	2,538
	AH2	Declaration of war against the United States	10,021	2,161
	AH3	Berlin Sports palast	12,394	2,056
	AH4	Last Radio Speech	2,172	770
Nelson Mandela	NM1	Apertura de la defen-sa	10,690	2,116
	NM2	Liberado	1,613	583
	NM3	Presidente	693	310
Margaret Thatcher	MT1	Britain Awake	2,867	871
	MT2	MT's first television interview	108	73
	MT3	The Bruges Speech	3,523	1,079
Elvis Presley	EP1	All shook up	263	101
	EP2	Don't be cruel	221	76
	EP3	That's someone you never forget	76	43
	EP4	Love me tender	106	58
	EP5	I can't stop loving you	76	56
	EP6	Jailhouse rock	275	130
Donald Trump	DT1	Iowa Freedom Speech	4,044	770
	DT2	Announcement Speech	6,703	1,048
	DT3	Arizona Speech	7,101	1,407
Hillary Clinton	HC1	Campaign launch	4,815	1,290
	HC2	Immigration reform	943	453
	HC3	DNC Speech	5,353	1,308
Angela Merkel	AM1	We have no time to lose	3,527	1,048
	AM2	Welcome to Berlin, Mr. Obama	769	333
	AM3	New Year	829	379

Personaje		Texto	Palabras	Palabras sin repeticiones
Madonna	M1	Frozen	230	72
	M2	Hung up	341	77
	M3	Papa do not preach	343	105
	M4	4 minutes	575	123
	M5	Like a virgin	258	91
	M6	Like a prayer	290	80
Michael Jackson	MJ1	Billie Jean	460	121
	MJ2	Smooth Criminal	523	79
	MJ3	We are the world	294	105
	MJ4	Beat it	417	82
	MJ5	Thriller	410	196
	MJ6	Love never felt so good	378	64
Los Beatles	B1	Yesterday	133	61
	B2	I want to hold your hand	190	38
	B3	Yellow submarine	229	70
	B4	Let it be	246	64
	B5	Help	275	78
	B6	Imagine	137	63

3.- Análisis de textos con el programa creado del NVG y de las redes obtenidas En vista de que para el análisis se considera la longitud de las palabras, es importante mencionar que, como se observa en el Cuadro 4.2, la distribución de la longitud de las mismas no es igual en todas las lenguas y que además, el número promedio de palabras que usa un hablante en su vocabulario habitual es distinto según el lenguaje que se considere. Por mencionar un ejemplo, un hablante de español no utiliza más allá de mil palabras, mientras que el hablante promedio de alemán usa aproximadamente cuatro mil palabras en su vocabulario habitual [?].

Cuadro 4.2: Lenguajes y longitud media de sus palabras

Lenguaje	Longitud media
Alemán	5.92
Español	4.96
Inglés	4.50
Francés	4.84

Fuente: [?]

Es por las observaciones anteriores que, se consideraron todos los textos en el idioma inglés, ya que la mayoría se encontraba originalmente en este idioma y

es considerado universal; además, no hace uso de acentos lo que resulta favorable para la investigación, ya que el programa elaborado del NVG, no toma en cuenta las letras acentuadas. Sin embargo, es importante mencionar que, con respecto a los signos de puntuación, al analizar los textos aún en este idioma, los resultados se ven afectados en parte por los apóstrofes que usa la gramática del posesivo y los guiones medios en palabras compuestas propias del inglés, limitante que puede corregirse en una investigación posterior. Para que el programa pudiera cargar el texto sin producir errores, se optó por quitar de ellos los guiones medios y marcarlos como espacios en blanco y en el caso de los apóstrofes, únicamente se quitaron.

Se dividió el análisis en tres fases: la primera consistió en formar dos grupos con los textos, uno con canciones, y otro con discursos y entrevistas, para realizar comparaciones generales entre estos; en la segunda fase se realizaron comparaciones del mismo tipo de textos (canciones con canciones y discursos con discursos) entre el mismo autor y también entre autores (por ejemplo, comparaciones entre todas las canciones de Michael Jackson y después entre las canciones de Michael Jackson y Madonna), para finalizar con la tercera fase del análisis que consistió en comparar autores sin importar si sus textos eran canciones o discursos, es decir, para hacer comparaciones inesperadas que pueden sin duda, dar una prueba más de la utilidad de esta herramienta de análisis, como por ejemplo, Adolf Hitler y The Beatles.

Para la última parte del análisis, y en vista de que se están considerando en algunos casos textos traducidos al inglés, se debe tomar en cuenta que, las traducciones son un modo limitado de describir el pensamiento de un autor y que no todas las palabras tienen un equivalente en todos los idiomas, por lo que los resultados de las comparaciones hechas entre textos originalmente escritos en inglés con los traducidos, no son realmente afines pero sí brindan un panorama general. Para conseguir resultados precisos sería necesario traducir los textos a un idioma neutro, es decir, a un idioma en el que ninguno de los textos se encontrara originalmente, lo cual podría considerarse en un trabajo futuro.

El análisis de los resultados presentados en la siguiente sección, se hizo sobre una base gráfica por lo que va acompañado de algunas imágenes que representan la implementación del NVG y la red generada, así como de cuadros que resumen lo obtenido en las principales métricas por cada personaje, para permitir una interpretación clara y contextualizada.

Capítulo 5

Resultados

Comparar establece semejanzas y diferencias entre dos conceptos, dos elementos o dos realidades. Se trata de acercar dos realidades y descubrir sus semejanzas y diferencias. Es por eso que, la comparación puede ser definida como un recurso del habla o de la escritura que se utiliza para establecer los elementos a partir de los cuales objetos, personas o situaciones son similares o diferentes entre sí.

Las personas estamos constantemente, de manera consciente e inconsciente, comparando, y esto es así porque la comparación es una acción que nos permite comprender una realidad, es decir, es una acción que siempre nos permitirá profundizar en el conocimiento de algo que ya se conoce o que se desconoce aún, por eso es una acción relevante.

A continuación se presentan algunas comparaciones de los resultados obtenidos de esta investigación, que se han encontrado interesantes, intentando con una breve contextualización, dar una explicación de las similitudes y diferencias que dan argumentos para hacer esta comparación válida y de interés general.

Fase 1. Canciones .vs. Discursos

En la Figura 5.1 se muestra el número de palabras con repeticiones y sin repeticiones de todas las canciones consideradas en esta tesis. Es posible notar que la canción con más palabras, contando aquellas que se repiten, es *Four minutes* (M4) con 575 palabras, mientras que hay un empate de canciones con el menor número de palabras, que es 76, y son del mismo autor: *That's someone you never forget* (EP3) y *I can't stop loving you* (EP5). Por otro lado, si no se consideran las palabras repetidas, la canción con más palabras resulta ser *Thriller* (MJ5) y la que menos palabras tiene es *I want to hold your hand* (B2).

En cuanto a los discursos, la Figura 5.2 muestra el número de palabras con repeticiones y sin repeticiones de todos los discursos considerados en la investigación. El que tiene más palabras contando las repeticiones, es *Before the Reichstag* (AH1) con 15 866 palabras, seguido por otro discurso del mismo orador *Berlin Sports Palast*

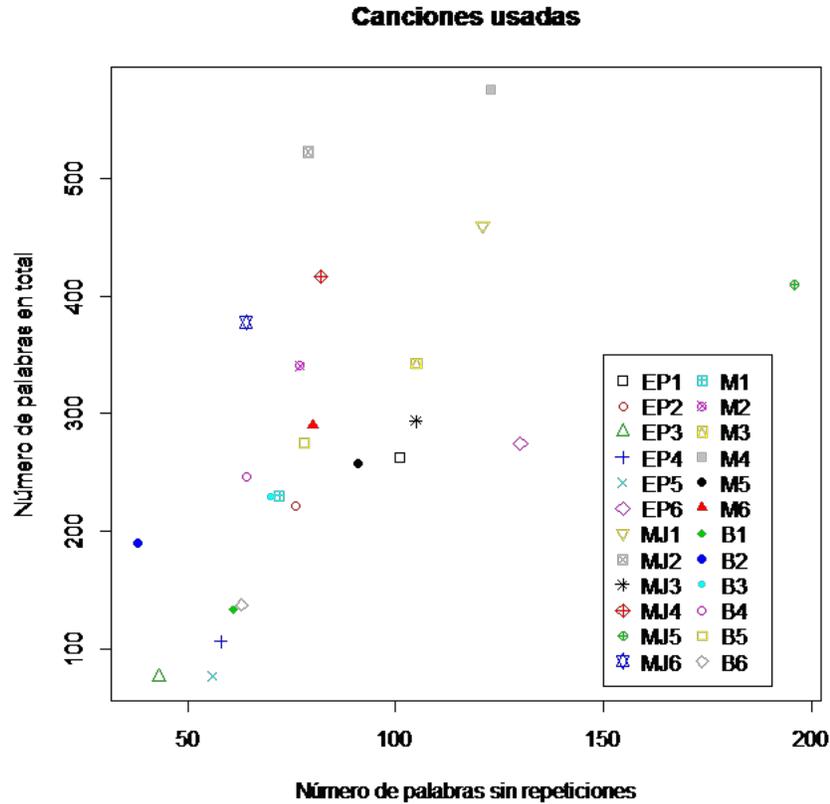


Figura 5.1: Canciones usadas para el análisis.

Fuente: Elaboración propia.

(AH3), mientras que con el menor número de palabras se tiene una entrevista: *MT's first television interview* (MT2). Si se quitan las palabras repetidas, sigue siendo AH1 el discurso que tiene más palabras, seguido de *Declaration of War against the United States* (AH2); en cuanto al discurso con el menor número de palabras al quitar las repetidas, continúa siendo MT2. Cabe resaltar que tanto con repeticiones o sin ellas, son los discursos de *Adolf Hitler* los que encabezan la lista con el mayor número de palabras usadas, en concordancia con que en alemán se usan en promedio más palabras, como se mencionó en el apartado 3 de la Sección 4.1.

Con base en los resultados obtenidos, se encontró que todas las canciones tienen un coeficiente de aglomeración en promedio mayor a 0.43, lo que las coloca por arriba del valor promedio obtenido en los discursos (0.38). Los valores más altos en esta métrica para las canciones, los tienen: *Yellow submarine* (B3), *That's someone you never forget* (EP3) y *Let it be* (B4) con 0.487, 0.474 y 0.470, respectivamente. En cuanto a los

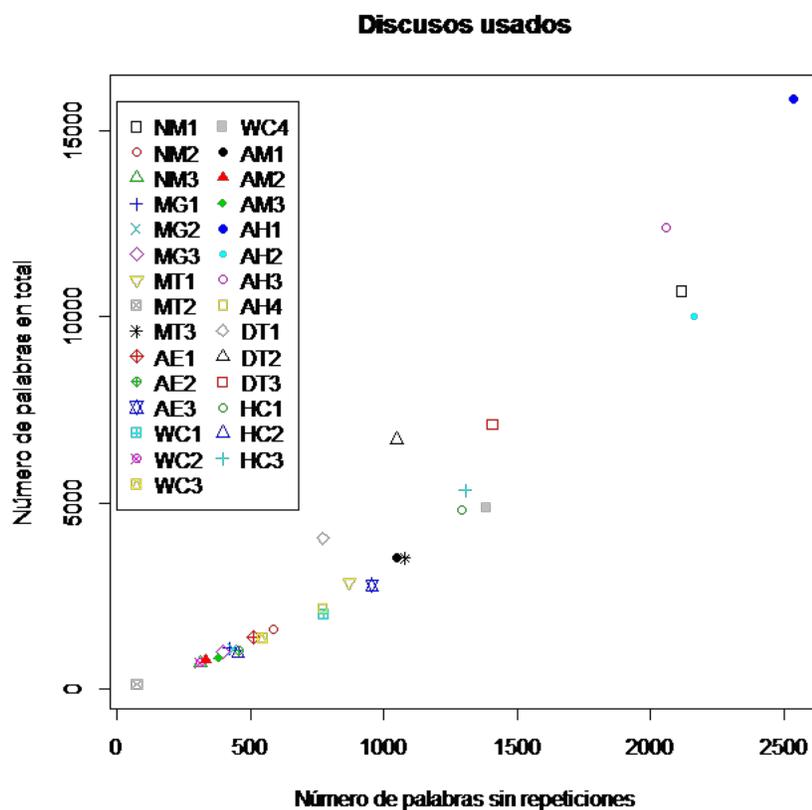
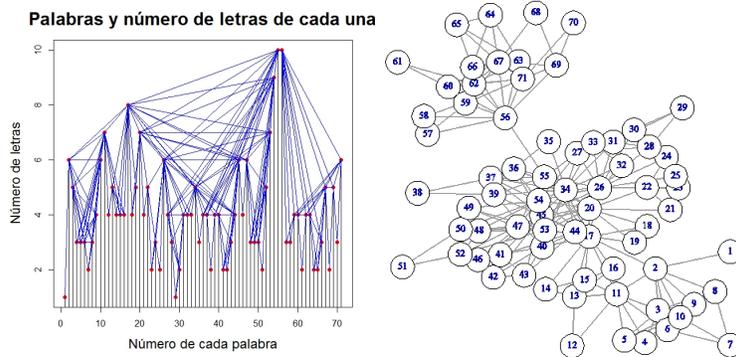


Figura 5.2: Discursos usados para el análisis.

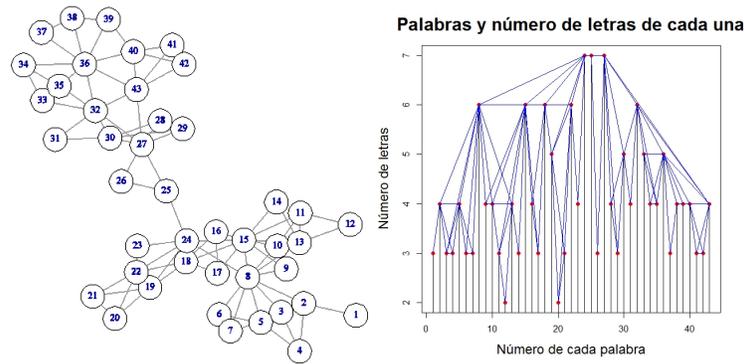
Fuente: Elaboración propia.

discursos, los valores más altos los tienen: *The war is won* (AE2) con 0.408, *MT's first television interview* (MT2) con 0.406 y *Quit India* (MG1) con 0.405. Es posible notar que, todos estos valores se encuentran por debajo de los valores más altos obtenidos en las canciones y se puede explicar debido a la rima que existe en las canciones pues las palabras tienen longitudes similares y esto permite tener una red más homogénea, hecho que es posible comprobar en la Figura 5.3; sobre todo en B3 (ver Figura 5.3(a)) pues su red presenta el grado promedio más alto que es 6.286, sin embargo, EP3 (ver Figura 5.3(b)) tiene el menor grado promedio 4.233, que es incluso menor al promedio general de las canciones (5.392).

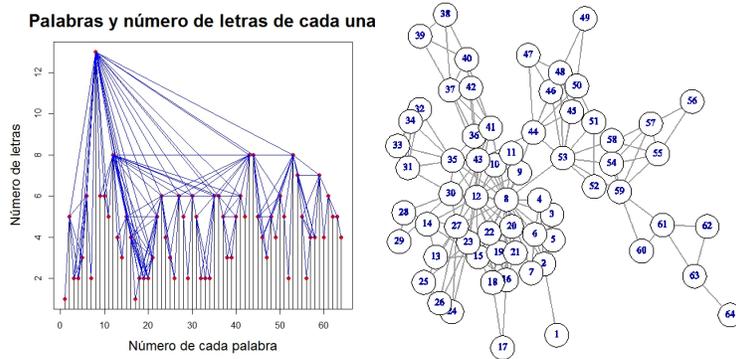
También en cuanto al grado promedio, se tiene que la red de DT2 presenta el valor más alto de los discursos, que es 6.160, mientras que el grado promedio general se encuentra en 5.89. Tanto las canciones como los discursos, tienen un grado promedio mayor a 5, únicamente las canciones EP3 y MJ6 están por debajo con 4.23 y 4.97,



(a) NVG aplicado a la canción B3 y la red obtenida



(b) NVG aplicado a la canción EP3 y la red obtenida



(c) NVG aplicado a la canción B4 y la red obtenida

Figura 5.3: Gráficas de visibilidad y redes asociadas de algunas canciones
Fuente: Elaboración propia.

respectivamente. AM2 tiene el menor grado promedio (5.479) de los discursos, que es incluso menor al obtenido en promedio para los discursos, pero mayor al promedio general de las canciones (5.392).

La eficiencia es otra métrica de la cual cabe resaltar los resultados obtenidos. En promedio, todas las canciones presentan un valor de 0.3, que es mayor al promedio obtenido en los discursos (0.2). Al tener una red construida con base en la longitud de las palabras, las canciones tienen mayor eficiencia por la rima de cada una de ellas. Además, al tratarse del inverso de la ruta más corta entre pares de nodos, podemos agregar que en general, las canciones tienen una distancia media menor que los discursos, que en términos numéricos indica que se necesita alrededor de 3.3 palabras intermedias para llegar de una a cualquier otra en una canción, mientras que en los discursos se necesitan 4.9 palabras para conseguirlo. Por lo tanto, ya que las canciones tienen mayor aglomeración o *clustering* y menor distancia media que los discursos, son más fáciles de recordar pues te llevan de una parte de la canción a otra con mayor facilidad, esto aunado a que tienen menor cantidad de palabras.

En cuanto a la asortatividad, los nodos tienden, por lo general, a conectarse con otros similares a ellos en cuanto al número de conexiones, es decir, tienden a conectarse con nodos de grado similar. Si las palabras usadas en las canciones y los discursos fueran similares en longitud, se esperaría tener una red homogénea, es decir, nodos de alto grado se enlazan con nodos de alto grado, sin embargo, los valores obtenidos en las redes de las canciones analizadas y las propias de los discursos, son muy cercanos a cero y no presentan diferencias significativas, por lo que no es posible establecer un resultado o una tendencia y en ambos casos, se tienen redes neutras.

Al dividir los textos en estos dos grupos, se esperaba obtener diferencias muy marcadas en las métricas, sin embargo, estas diferencias resultaron ser no muy grandes pero evidencian que las canciones son más fáciles de recordar que los discursos aunque de acuerdo a los resultados obtenidos, los discursos no son tan difíciles de aprender.

Fase 2. Canciones .vs. Canciones

En primera instancia, se enlista el número promedio de palabras usadas por cada artista: *Elvis Presley* tiene 77.3, *Michael Jackson* tiene 107.8, *Madonna* cuenta con 91.3 y *The Beatles* con 62.3. La diferencia entre el artista que más palabras en promedio utiliza con el que menos utiliza, es de 45.5 palabras.

En el Cuadro 5.1, se presenta un comparativo con los valores obtenidos del coeficiente de aglomeración por cada artista, es posible observar que los valores más altos los tienen *Los Beatles* y la brecha más grande entre el valor mínimo y máximo la tiene *Madonna*. Esto podría ser factor para corroborar que las canciones del grupo inglés sean más fáciles de aprender que las de la reina del pop pues tienen mayor rima y usan en promedio treinta palabras menos que *Madonna*, 62 contra 92. Vale la pena mencionar que, una palabra puede estar relacionada con la misma palabra cuando esta cambia ligeramente, debido a que tendrían longitudes similares, por ejemplo, *live*, *lives*, *life*, etcétera; así que aunque será más fácil recordar las canciones, también será más proba-

ble confundir con facilidad una palabra por otra; sobre todo en los coros que riman.

Cuadro 5.1: Comparación de los valores de aglomeración por artista

Artista	Mayor <i>clustering</i>	Menor <i>clustering</i>	Promedio por artista
Elvis Presley	<i>That's someone you never forget</i> EP3 0.474	<i>Jailhouse rock</i> EP6 0.406	0.433
Michael Jackson	<i>Smooth Criminal</i> MJ2 0.461	<i>Thriller</i> MJ5 0.386	0.427
Madonna	<i>Like a prayer</i> M6 0.461	<i>Like a virgin</i> M5 0.386	0.425
Los Beatles	<i>Yellow submarine</i> B3 0.477	<i>Help</i> B5 0.417	0.447
General Canciones	<i>Yellow submarine</i> B3 0.477	<i>Like a virgin</i> M5 0.386	0.433

Es posible notar que, el valor promedio general únicamente es rebasado por el promedio de dos artistas, *Elvis Presley* y *Los Beatles*. Además, si se observa el Cuadro 5.2, esto sucede para la eficiencia sólo en el caso de *Los Beatles*.

Cuadro 5.2: Comparación de los valores de eficiencia por artista

Artista	Mayor eficiencia	Menor eficiencia	Promedio por artista
Elvis Presley	<i>I can't stop loving you</i> EP5 0.351	<i>That's someone you never forget</i> EP3 0.256	0.297
Michael Jackson	<i>Love never felt so good</i> MJ6 0.315	<i>Thriller</i> MJ5 0.258	0.291
Madonna	<i>Like a virgin</i> M5 0.326	<i>Four minutes</i> M4 0.276	0.305
Los Beatles	<i>I want to hold your hand</i> B2 0.399	<i>Let it be</i> B4 0.318	0.348

Artista	Mayor eficiencia	Menor eficiencia	Promedio por artista
General Canciones	<i>I want to hold your hand</i> B2 0.399	<i>That's someone you never forget</i> EP3 0.256	0.310

Ya que el grupo inglés presenta el valor más alto de aglomeración y de eficiencia, se puede afirmar que sus canciones son las más fáciles de recordar de los artistas seleccionados y si a estos resultados se les suma que tiene el valor más pequeño de distancia media, se corrobora esta afirmación. La información de esta métrica puede ser consultada en el cuadro 5.3.

Cuadro 5.3: Comparación de los valores de distancia media por artista

Artista	Mayor distancia media	Menor distancia media	Promedio por artista
Elvis Presley	EP3 y EP6 3.9	EP5 2.8	3.4
Michael Jackson	MJ5 3.9	MJ4 y MJ6 3.2	3.5
Madonna	M3 y M4 3.6	M1, M5 y M6 3.1	3.3
Los Beatles	B5 3.2	B2 2.5	2.9
General Canciones	EP3, EP6 y MJ5 3.9	B2 2.5	3.3

En cuanto al grado promedio, también es el grupo británico quien tiene el valor más grande, esto puede traducirse en que las palabras de sus canciones tienen longitudes similares y por eso, el algoritmo de visibilidad permite conectarlas con más palabras.

Si analizamos la asortatividad, todos los valores son muy cercanos a cero, lo que indica que se trata de redes neutrales. Únicamente *Elvis Presley* tiene un valor promedio positivo y el más pequeño lo presentan *Los Beatles* con -0.04.

Además, si comparamos las primeras 5 canciones en algunas métricas, obtenemos que B3 tiene el mayor *clustering* y una de las distancias medias más pequeñas (2.9) seguida por EP3 que tiene una de las distancias medias más grandes (3.9) así que, B3 es más fácil de recordar y, por lo tanto, se puede explicar que la gente sepa mejor esta letra. En cuanto a la eficiencia, también es una canción del grupo la que ocupa el primer lugar, B2 que además tiene la menor distancia media (2.5). Si consideramos los resultados del coeficiente de aglomeración y la eficiencia, resulta que cinco de las seis canciones consideradas del grupo británico aparecen con los valores más altos (a excepción de B5) y las distancias medias más pequeñas también las tienen tres de las seis canciones del grupo.

En resumen y con relación a la conclusión de la primera sección de este Capítulo, es posible afirmar que las canciones de *Los Beatles* son las que cuentan, en promedio, con el menor número de palabras utilizadas, los valores más bajos de distancia media, y los valores más altos de clustering por lo que, podemos afirmar, que son más fáciles de recordar, y por tanto de aprenderse, que las de los demás artistas considerados. Además, los altos valores de eficiencia, indican que hay presencia de más rimas en cada canción.

Fase 3. Discursos .vs. Discursos

En cuanto a los discursos, es posible observar que el valor promedio más alto de *clustering*, lo tiene *Mahatma Gandhi* con 0.4 y si se interpreta que un valor alto implica que las personas podrán recordar con mayor facilidad el discurso, entonces se puede explicar el por qué a través de sus discursos convenció a la gente de tomar acciones para lograr la independencia de la India y por qué es uno de los personajes más citados hoy en día [?].

El Cuadro 5.4 presenta los valores obtenidos en esta métrica por orador y es posible notar que, la brecha más grande entre el valor más grande y el más pequeño lo tiene *Margaret Thatcher*.

Cuadro 5.4: Comparación de los valores de aglomeración por orador

Orador	Mayor <i>clustering</i>	Menor <i>clustering</i>	Promedio por orador
Nelson Mandela	Presidente NM3 0.395	Apertura de la defensa NM1 0.364	0.379
Mahatma Gandhi	<i>Quit India</i> MG1 0.405	<i>Dandi March</i> MG3 0.394	0.400
Margaret Thatcher	<i>MT's first television interview</i> MT2 0.406	<i>Britain Awake</i> MT1 0.361	0.380
Albert Einstein	<i>The war is won but peace is not</i> AE2 0.408	<i>Why do they hate the Jews?</i> AE3 0.366	0.392
Winston Churchill	<i>Be Ye Men of valour</i> WC3 0.383	<i>Alliance with Russia</i> WC1 0.356	0.371
Angela Merkel	<i>New Year</i> AM3 0.403	<i>We have no time to lose</i> AM1 0.377	0.393
Adolf Hitler	<i>Last Radio Speech</i> AH4 0.389	<i>Before the Reichstag</i> AH1 0.361	0.376

Orador	Mayor <i>clustering</i>	Menor <i>clustering</i>	Promedio por orador
Donald Trump	<i>Iowa Freedom Speech</i> DT1 0.384	<i>Arizona Speech</i> DT3 0.358	0.373
Hilary Clinton	<i>Immigration Speech</i> HC2 0.375	<i>DNC Speech</i> HC3 0.365	0.370
General Discursos	<i>The war is won but peace is not</i> AE2 0.408	<i>Alliance with Russia</i> WC1 0.356	0.381

Cabe mencionar que para seis de los nueve oradores (excepto MG, WC y AM), el discurso con mayor coeficiente de aglomeración coincide con aquel que tiene menos palabras. Únicamente MG3 es el discurso con menos palabras y con menor *clustering*. Es posible notar que el valor promedio de aglomeración es rebasado por *Albert Einstein*, *Angela Merkel* y *Mahatma Gandhi*, en el caso de este último, incluso el valor más pequeño supera al promedio general.

El orador que utiliza más palabras en promedio en sus discursos es *Adolf Hitler* (1,881) y es *Mahatma Gandhi* el que utiliza menos palabras en sus discursos (416); la diferencia entre ambos es de 1,465.

En cuanto a la eficiencia, el promedio general es rebasado por los promedios de los primeros cinco oradores que aparecen en el Cuadro 5.5, ordenándolos de mayor a menor son: *Margaret Thatcher* (0.248), *Mahatma Gandhi* (0.224), *Winston Churchill* (0.218), *Albert Einstein* (0.214) y *Nelson Mandela* (0.210).

En vista de que MG presenta el valor más alto de aglomeración y uno de los más altos de eficiencia (el segundo), se puede afirmar que sus discursos son los más fáciles de seguir y, por ende, los más sencillos de recordar de los oradores seleccionados y si a estos resultados se les suma que tiene el segundo valor más pequeño de distancia media (4.5), se corrobora esta afirmación. La información de la distancia media, puede ser consultada en el Cuadro 5.6.

Cuadro 5.5: Comparación de los valores de eficiencia por orador

Orador	Mayor eficiencia	Menor eficiencia	Promedio por orador
Nelson Mandela	Presidente NM3 0.243	Apertura de la defensa NM1 0.180	0.210
Mahatma Gandhi	<i>Quit India</i> MG1 0.235	<i>Eve</i> MG2 0.210	0.224

Orador	Mayor eficiencia	Menor eficiencia	Promedio por orador
Margaret Thatcher	<i>MT's first television interview</i> MT2 0.332	<i>The Bruges Speech</i> MT3 0.202	0.248
Albert Einstein	<i>The war is won but peace is not</i> AE2 0.221	<i>Why do they hate the Jews?</i> AE3 0.207	0.214
Winston Churchill	<i>Blood, Toil, Tears and Sweat</i> WC2 0.248	<i>The Sinews of Peace</i> WC4 0.189	0.218
Angela Merkel	<i>Welcome to Berlin, Mr. Obama</i> AM2 0.208	<i>We have no time to lose</i> AM1 0.186	0.196
Adolf Hitler	<i>Last Radio Speech</i> AH4 0.206	<i>Berlin Sports Palast</i> AH3 0.153	0.177
Donald Trump	<i>Arizona Speech</i> DT3 0.201	<i>Announcement Speech</i> DT2 0.197	0.199
Hilary Clinton	<i>Immigration Speech</i> HC2 0.230	<i>Campaign Launch</i> HC1 0.188	0.205
General Discursos	<i>Blood, Toil, Tears and Sweat</i> WC2 0.248	<i>Berlin Sports Palast</i> AH3 0.153	0.209

Cuadro 5.6: Comparación de los valores de distancia media por orador

Orador	Mayor distancia media	Menor distancia media	Promedio por orador
Nelson Mandela	Apertura de la defensa NM1 5.6	Presidente NM3 4.1	4.8
Mahatma Gandhi	<i>Eve</i> MG2 4.8	<i>Quit India</i> MG1 4.3	4.5
Margaret Thatcher	<i>The Bruges Speech</i> MT3 4.9	<i>MT's first television interview</i> MT2 3.0	4.2
Albert Einstein	<i>Why do they hate the Jews?</i> AE3 4.8	<i>The war is won but peace is not</i> AE2 4.5	4.7

Orador	Mayor distancia media	Menor distancia media	Promedio por orador
Winston Churchill	<i>The Sinews of Peace</i> WC4 5.3	<i>Blood, Toil, Tears and Sweat</i> WC2 4.0	4.6
Angela Merkel	<i>We have no time to lose</i> AM1 5.4	<i>Welcome to Berlin, Mr. Obama</i> AM2 4.8	5.1
Adolf Hitler	<i>Berlin Sports Palast</i> AH3 6.5	<i>Last Radio Speech</i> AH4 4.8	5.7
Donald Trump	<i>Announcement Speech</i> DT2 5.1	<i>Iowa Freedom Speech/Arizona Speech</i> DT1/DT3 5.0	5.0
Hilary Clinton	<i>Campaign Launch</i> HC1 5.3	<i>Immigration Reform</i> HC2 4.3	4.9
General Discursos	<i>Berlin Sports Palast</i> AH3 6.5	<i>MT's first television interview</i> MT2 3.0	4.9

Cabe resaltar también que, MT tiene el valor promedio más pequeño en distancia media, el valor más alto de eficiencia y uno de los mayores valores promedio de *clustering* por lo que sus discursos están también entre los más sencillos de seguir.

En cuanto al grado promedio, es *Donald Trump* quien tiene el valor más grande, 6.0 esto puede traducirse en que las palabras de sus discursos tienen longitudes similares y por eso, el algoritmo de visibilidad permite conectarlas con más palabras. Además, el promedio de los discursos es 5.89, que representa 0.5 más que el de las canciones (5.39). El valor promedio más pequeño lo presenta *Angela Merkel* con 5.72.

Si analizamos la asortatividad, todos los valores son muy cercanos a cero, lo que indica que se trata de redes neutrales. Todos los valores promedio son positivos y el más grande lo presenta *Adolf Hitler* con 0.18, mientras que el valor más pequeño lo presenta la alemana *Angela Merkel* (0.08).

Si comparamos los primeros 5 discursos en algunas métricas, obtenemos que AE2 tiene el mayor *clustering*, pero es MT2 el discurso que tiene no sólo un valor alto en esta métrica (el segundo lugar general) si no también, la distancia media más pequeña (3.0), además de un valor de grado promedio de palabras alto y el mayor valor registrado en eficiencia, así que, unido al hecho de que *Margaret Thatcher* presenta un valor promedio de palabras usadas por discurso de 674.33 y el número de palabras de MT2 es alrededor de un noveno de estas, es más fácil de recordar.

Si consideramos los resultados del coeficiente de aglomeración y la eficiencia, resulta que sólo el discurso MT2 de esa oradora, está presente dentro de los primeros diez discursos. En cambio, todos los discursos considerados de *Mahatma Gandhi* aparecen dentro de los valores más altos. Dentro de los diez valores de distancias medias más pequeñas, también aparecen los discursos de *Mahatma Gandhi*.

En resumen y con relación a la conclusión de las dos primeras secciones, es posible afirmar que los discursos de *Mahatma Gandhi* son los que cuentan, en promedio, con el menor número de palabras utilizadas (416.33), los valores más bajos de distancia media (menor o igual a 4.752), y los valores más altos de clustering (mayor o igual a 0.389) por lo que, podemos afirmar, que dentro de los oradores considerados, son los más fáciles de recordar.

Fase 4. Comparaciones de interés

The Beatles .vs. Mahatma Gandhi

Es bien sabido por todos que Gandhi peleó contra los británicos por medio del pacifismo hasta conseguir la independencia de la India y los *Beatles*, influenciados por estas ideas, con su actitud que los medios definieron como “alegremente desafiante”, retaron al gobierno y al sector conservador británico.

Para la portada del álbum más revolucionario del grupo, que marcó el nacimiento de la música psicodélica y dio origen al rock sinfónico, *Sgt. Pepper's Lonely Hearts Club*, se incluyeron las imágenes de algunos de los personajes más importantes de la historia de la humanidad hasta ese momento y se solicitó incluir las imágenes de *Hitler*, *Gandhi* y *Jesucristo* pero ninguno fue incluido.

En cuanto al análisis, se encontró que son necesarias sólo dos palabras intermedias más para llegar de cierta palabra a cualquier otra en un discurso de *Gandhi* que en una canción de los *Beatles*. Las redes formadas por el grupo resultan estar más conectadas ya que el valor de *clustering* promedio es de 0.447, mientras que el del hindú es en promedio 0.400; cabe resaltar que este valor es el mayor obtenido de todos los discursos.

Madonna .vs. Hilary Clinton

Ambas estadounidenses, desde el inicio de sus respectivas carreras, han luchado para obtener igualdad de género y respeto para las mujeres y para otros grupos vulnerables.

En los meses previos a las elecciones presidenciales de Estados Unidos, la reina del pop ofreció conciertos en varios estados del país e incluso favores sexuales [?], a cambio de votos a favor de *Hilary Clinton*.

Es una de las comparaciones que arroja resultados muy diferentes en casi todas las métricas analizadas, la excepción es el grado promedio, que en *Madonna* es 5.460 y en *Clinton* es 5.906. En primer término, *Madonna* tiene una distancia media más pequeña (3.3) y una mayor eficiencia (0.305) que la candidata presidencial que presenta 4.9 y 0.205, respectivamente; este hecho se traduce en que moverse de una palabra a otra es

mucho más fácil en una canción debido a la rima.

En segundo término, el *clustering* que presenta la reina del pop es mayor al de *Clinton*, por lo que nuevamente se confirma que las canciones forman redes más homogéneas. Además, sus valores de asortatividad indican que se trata de redes neutras por ser tan cercanos a cero, sin embargo el valor promedio de las redes de *Madonna* es negativo.

Michael Jackson .vs. Nelson Mandela

Mandela describió a *Jackson* como un hombre cercano a su familia y le rindió tributo en su muerte. Además, hace unos años tras quedar en libertad después de ser acusado de abuso infantil, *Michael Jackson* se comparó con *Nelson* cuando fue liberado de prisión después de estar ahí injustamente por 27 años.

Ambos hombres son considerados dentro de los tres hombres negros que cambiaron el rumbo del mundo en distintas dimensiones, junto con *Muhammad Ali*.

Mandela presenta un valor menor de *clustering* que *Michael Jackson*, esto se trata principalmente porque sus canciones tienen rima y una longitud similar en las palabras que las componen, por lo que se forman redes homogéneas. Ambos tienen un grado promedio mayor a cinco y similar, es decir, una palabra de cualquiera de ellos tiene al menos 5 con las que, por el algoritmo de visibilidad, podría conectarse.

En cuanto a la distancia media, la que presentan las redes de *Mandela* son una unidad mayor en esta métrica que las de *Jackson*, por lo que sin duda es mucho más sencillo recordar una canción del rey del pop que un fragmento de algún discurso del líder sudafricano. A pesar de estas diferencias en distancia media, sus valores de eficiencia resultan muy similares.

Por último, tanto *Jackson* como *Mandela* tienen un valor de asortatividad promedio muy cercano a cero, la única diferencia radica en que el valor del primero es negativo.

Mahatma Gandhi .vs. Adolf Hitler

Por un lado se tiene al artífice de la independencia de la India que instauró métodos de protesta pacíficos como la huelga de hambre, y por el otro, se tiene al que es considerado la persona más malévola de la historia, profeta del odio y actor intelectual de la masacre más grande de la humanidad.

Ambos personajes son conocidos por sus dotes de liderazgo, sin embargo, la diferencia radica en que *Gandhi* era líder a través de inspiración y *Hitler* lo era a través de intimidación, hecho que se puede verificar si se analiza el contenido de sus discursos.

Entre 1939 y 1940, *Mahatma* escribió una serie de cartas para tratar de evitar que *Hitler* empezara una guerra pero no hay registro que indique si recibió alguna de ellas.

En términos del estudio, *Gandhi* utilizó alrededor de 416 palabras en sus discursos

mientras que *Hitler* usó cuatro veces más palabras (1,881). Lo anterior corrobora que *Adolf* basaba sus discursos en la repetición de ideas por lo que eran muy largos y con una única idea principal [?].

En cuanto a los valores promedio obtenidos de *clustering*, *Gandhi* presenta un valor de 0.4 que es cercano al obtenido por las canciones y difiere por casi tres centésimas con el valor de *Hitler* (0.376), que sorprendentemente resulta uno de los más bajos de los obtenidos por los oradores. Para el grado promedio, los valores de ambos son muy similares sin embargo, el valor de la distancia media indica que para moverse de una palabra a otra en los discursos del alemán, se necesita una palabra más que para hacerlo dentro de los de *Mahatma*.

Donald Trump .vs. Adolf Hitler

Muchos medios de comunicación, tras escuchar los discursos de odio del actual Presidente de Estados Unidos, *Trump*, han sugerido similitudes entre ambos, principalmente por hacer uso del racismo para incitar multitudes.

Hitler promovió la imagen de un líder al que no se le tiene que cuestionar, se presentó como antídoto para todos los problemas sociales y políticos que acontecían en Alemania, y con base en la retórica populista que apela a la nostalgia, a la defensa del país y al resentimiento, propició la unidad a costa de los extranjeros. *Trump* ha seguido estos pasos con la diferencia de que *Hitler* culpaba a los judíos de todos los problemas alemanes y *Trump* culpa a los inmigrantes, en especial, a los mexicanos.

Durante la campaña presidencial, se dijo que *Trump* tenía en su buró una copia de “*Mi lucha*”, la obra autobiográfica de *Hitler*; este hecho fue confirmado por el candidato.

Al igual que los medios de comunicación, también las métricas analizadas sugieren similitudes entre *Donald Trump* y *Adolf Hitler*. Por ejemplo, en *clustering* difieren por tres milésimas, 0.373 contra 0.376, así que las redes formadas por los discursos del alemán, sugieren provenir de textos con palabras de longitud similar. Sin embargo, *Hitler* presenta una mayor distancia media y por tanto, una menor eficiencia que *Donald* pero se le puede atribuir a que en promedio, en alemán se usan un mayor número de palabras que en inglés [?].

Winston Churchill .vs. Adolf Hitler

Existen muchos paradigmas que rodean esta comparación. En primer término, ambos son conocidos por su carisma, se rodearon de personas inteligentes en sus equipos de trabajo, e incluso a ambos les gustaban las artes, específicamente la pintura.

En segundo plano, muchos historiadores han afirmado que ambos llegaron a ser líderes de sus respectivas naciones en gran parte por la época en la que vivieron, la post-guerra y las consecuencias que tuvo esta para sus países.

Existía un contraste en las técnicas que empleaban para elaborar sus discursos,

Hitler tenía un asesor que los preparaba y otro que le enseñó las técnicas de la oratoria, es decir, distraer a la audiencia con movimientos corporales. Por otro lado, *Churchill* pasaba horas e incluso días preparando sus escritos.

En cuanto a su vida familiar, contrasta el hecho de que el padre de *Hitler* era un oficial de bajo rango que lo golpeaba brutalmente, mientras que el padre de *Winston* fue un político conservador importante en Reino Unido. Si bien es cierto que gran parte de sus vidas contrasta, ambos tenían otra cosa en común: no tenían buenas calificaciones.

En ambos casos, sus redes tienen grados promedio cercanos a 6, y las pequeñas diferencias en distancia media y eficiencia, pueden explicarse debido al número de palabras promedio que se utilizó en los discursos.

Los valores obtenidos de asortatividad indican que las redes son neutrales, por lo que no existe una tendencia a unir nodos de alto grado con nodos de alto grado.

Elvis Presley .vs. Albert Einstein

Ambos son considerados genios de la humanidad, el primero por innovar la industria de la música con el *Rock & Roll* y el segundo, por llevar a cabo una revolución en el ámbito de las ciencias físicas.

Según la revista *Forbes*, ambos están en la lista de personas que más ganancias obtienen después de su muerte, son respectivamente el segundo y el octavo lugar de la lista.

Al igual que en las otras comparaciones de cantantes y oradores, los diferentes valores obtenidos en las métricas analizadas, se deben principalmente al hecho de estar comparando canciones con discursos. *Elvis* presenta valores mayores en *clustering* y eficiencia, mientras que *Einstein* tiene un valor más alto en grado promedio, distancia media y asortatividad.

Angela Merkel .vs. Margaret Thatcher

Ambas representan el símbolo de la mujer que se impone en un mundo de hombres. *Merkel* es la primera mujer que logra convertirse en canciller de Alemania, mientras que *Thatcher* fue la primera mujer en convertirse en jefa de gobierno en Reino Unido.

Tanto *Angela* como *Margaret*, tomaron sus cargos pasados los cincuenta años de edad (51 contra 53 años) y desde el momento en que llegaron, fueron conocidas por su carácter firme, incluso se ha llegado a decir en los medios de comunicación que *Merkel* es la nueva “*dama de hierro*” de Europa. Fueron reelegidas por varias ocasiones.

La eficiencia promedio que presentan las redes de *Margaret Thatcher* es 0.248, esta métrica indica qué tan bien se mueve la información dentro de la red, y al ser mayor que la presentada por *Angela Merkel*, es posible concluir que seguir la red formada por los discursos de la británica es más sencillo, a pesar de que el valor de *clustering* que presenta sea menor que el de la alemana (0.380 contra 0.393).

Análisis estadístico

En la Sección 2.2 se abordaron las características más relevantes de los principales modelos de redes, entre ellos se encontraron las redes libre escala cuya cualidad primordial es contar con una distribución de grado que sigue una ley de potencia. Además, el Capítulo 3 fue dedicado a las redes de lenguaje y al algoritmo de visibilidad, por lo que fue posible concluir que en vista de que dicho algoritmo permite transformar series de tiempo, como los lenguajes, en redes, y a que las redes formadas a través de este algoritmo son siempre conectadas y no dirigidas [?], falta demostrar que son invariantes bajo transformaciones y que las redes de lenguaje presentan la estructura de una red libre escala, por lo que sería suficiente demostrar que su distribución de grado sigue una ley de potencia.

Se dice que los datos de K siguen una ley de potencia si $P(K = k) \sim k^{-\alpha}$ donde k es un número positivo y $\alpha > 1$, además, α se conoce como parámetro de escala. La herramienta más utilizada para probar que ciertos datos cumplen con esta ley es un histograma en el que los ejes x y y son logarítmicos; si al hacer el histograma, la distribución se representa con una línea recta entonces se podría afirmar que la distribución sigue una ley de potencia con un valor α dado por la pendiente absoluta de la línea recta que se obtiene a través de una regresión lineal (método de mínimos cuadrados). La función en R, *power.law.fit*, permite realizar este procedimiento en sólo una línea de código.

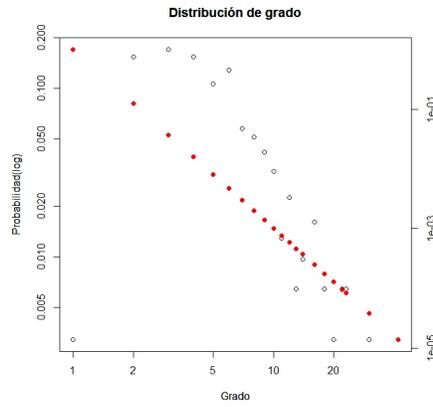
Los datos que se generan a partir de una distribución diferente (una exponencial o una lognormal), siempre pueden ajustarse a un modelo de ley de potencia pero el ajuste puede ser pobre aunado a esto, este método muestra sesgos significativos por lo que los resultados son poco confiables y se necesita una herramienta que indique que tan bueno es el ajuste [?], por eso surgen las pruebas de bondad de ajuste. En la actualidad, existen muchas pruebas y todas comparan los datos observados con los de la distribución hipotética pero una de las más sencillas y la que se aplica en este estudio es la prueba de *Kolmogorov-Smirnoff* (K-S).

Esta prueba permite medir el grado de concordancia que existe entre la distribución real de un grupo de datos y una distribución teórica específica, que en este caso será la ley de potencia. En otras palabras, el objetivo de la prueba K-S es indicar si los datos podrían razonablemente proceder de la distribución teórica especificada. Para poder aplicarla, se toma como hipótesis que los datos provienen de la distribución teórica y para poder aceptar dicha hipótesis, se usa un valor conocido como *p-value*, el cual deberá ser mayor o igual a 0.05 [?]; mientras más cercano a uno sea, es más probable que los datos se extrajeran de una distribución de ley de potencia [?]. Es importante no perder de vista que las pruebas de bondad de ajuste son herramientas que permiten descartar modelos por lo que lo mejor que se puede obtener en este caso, en un sentido estrictamente estadístico, es no descartar el modelo de ley de potencia por los datos observados.

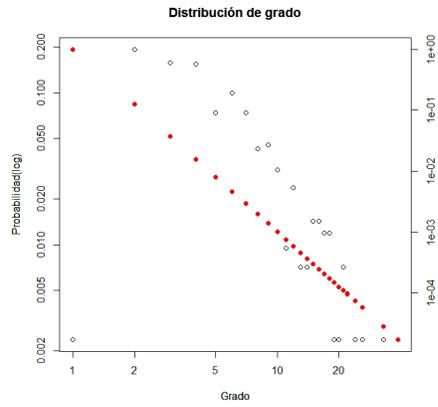
El exponente α varía según el sistema que se esté analizando y muchas propiedades

de la red dependen de este valor; una de las propiedades principales es la distancia media y se ha probado que si α cumple con estar en el intervalo $2 < \alpha < 3$ [?], se satisface que la distancia media se aproxima a $\ln(\ln(N))$, lo cual significa que son redes ultra pequeñas; fuera de estos límites, la red o bien se desconectaría fácilmente ($\alpha > 3$) ya que habría pocos hubs, pues la distribución de grado descendería rápidamente, o sería muy densa ($\alpha < 2$), hasta podría suceder que el número de enlaces de los hubs creciera rápidamente, incluso más rápido que la misma red [?].

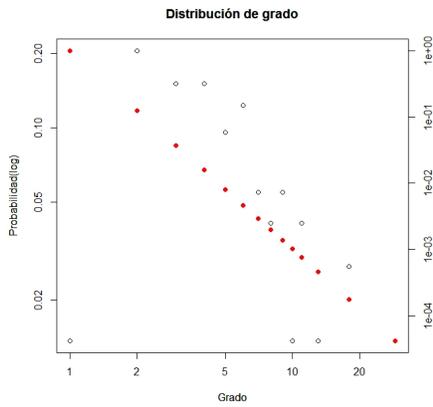
A continuación, presentamos las gráficas con los datos de los discursos con sus valores respectivos de α y p -value, que arrojó la prueba K-S. Debido a que son muchos textos, únicamente se incluyen las gráficas de cada orador que presentan un valor de α y un p -value más cercanos a su respectivo promedio, sin embargo, en todas las distribuciones de grado, se puede observar que las redes de las que se obtuvieron, pueden provenir de una distribución de ley de potencia, por lo que podemos concluir que son redes con tendencia a ser de libre escala.



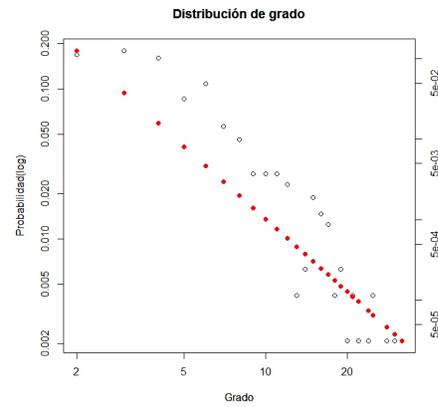
(a) NM3, $\alpha=3.48$



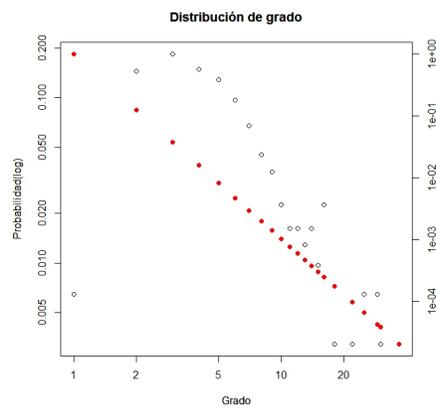
(b) MG1, $\alpha=5.46$



(c) MT2, $\alpha=3.29$



(d) AE3, $\alpha=5.42$



(e) WC2, $\alpha=2.99$

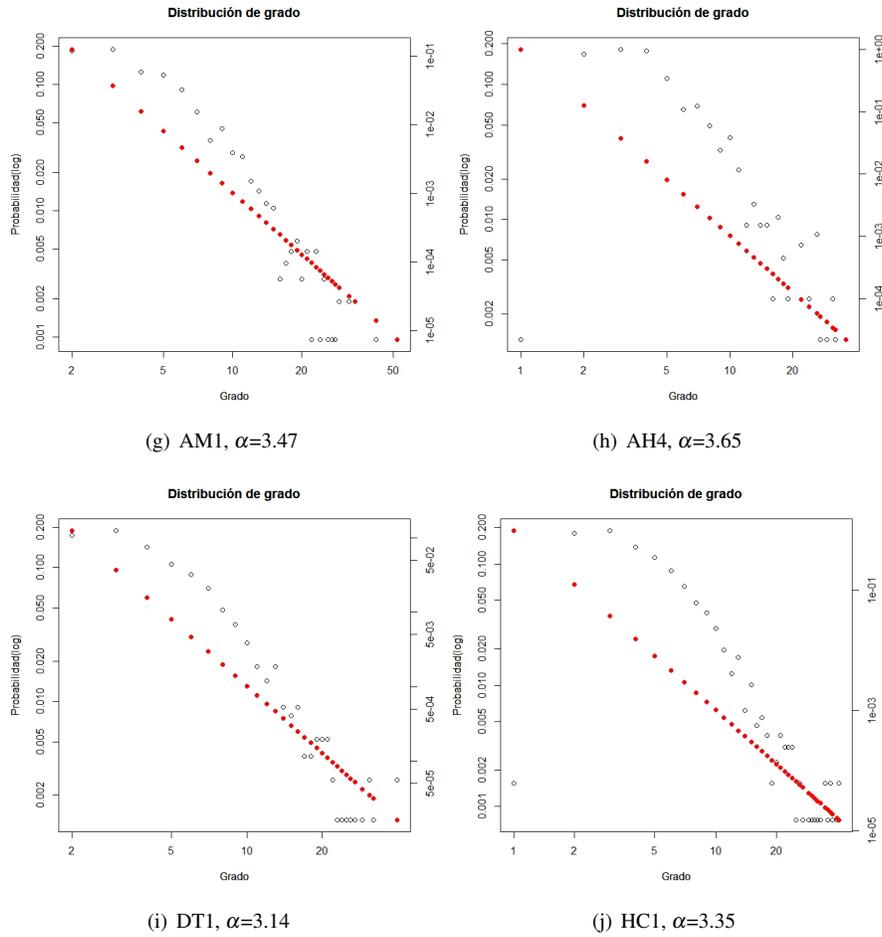


Figura 5.4: Gráficas de la distribución de grado de los discursos utilizados que más se acercan al valor de α promedio de cada orador.

Fuente: *Elaboración propia.*

Un *p-value* cercano a uno, no necesariamente significa que la ley de potencia es la distribución correcta para los datos pues en primer lugar, puede haber otras distribuciones que coincidan con los datos igual de bien o mejor en el rango considerado por lo que se necesita de otras pruebas para descartarlos y decidir qué modelo es el que mejor ajusta a los datos[?] y en segundo lugar, porque el *p-value* es más confiable cuando el tamaño de la muestra es grande (más de mil) pero si la muestra es pequeña, *p-value* altos pueden observarse aún cuando este modelo es incorrecto para los datos. Esto no es una deficiencia del método, simplemente refleja el hecho de que es más difícil descartar la ley de potencia cuando hay menos datos [?].

En cuanto a las canciones, al tratarse de redes pequeñas (menos de mil nodos), se obtuvo un $\alpha > 3$ debido a que hay pocos hubs ya que las palabras riman fácilmente por tener alturas similares. Sin embargo, la distribución que mejor les ajustó (siguiendo el método de descarte expuesto en [?]), fue una distribución exponencial y a continuación se muestran los resultados obtenidos al aplicarles la prueba de K-S. Dicha prueba, se realizó con el comando *KS-test* en R, en el cual, el *p-value* es la probabilidad de tener un valor más grande al que se observaría si dos muestras fueran tomadas de la misma distribución; los valores grandes de un *p-value* indican que estadísticamente no es posible sustentar que provienen de distribuciones diferentes pero los valores pequeños tampoco implican que sí lo hacen ya que estos pueden aparecer en muestras pequeñas (como es el caso), por lo que sería necesario realizar otro tipo de prueba para confirmarlo (por ejemplo, la prueba de χ^2 o métodos no lineales). A continuación se presentan los resultados obtenidos para las canciones:

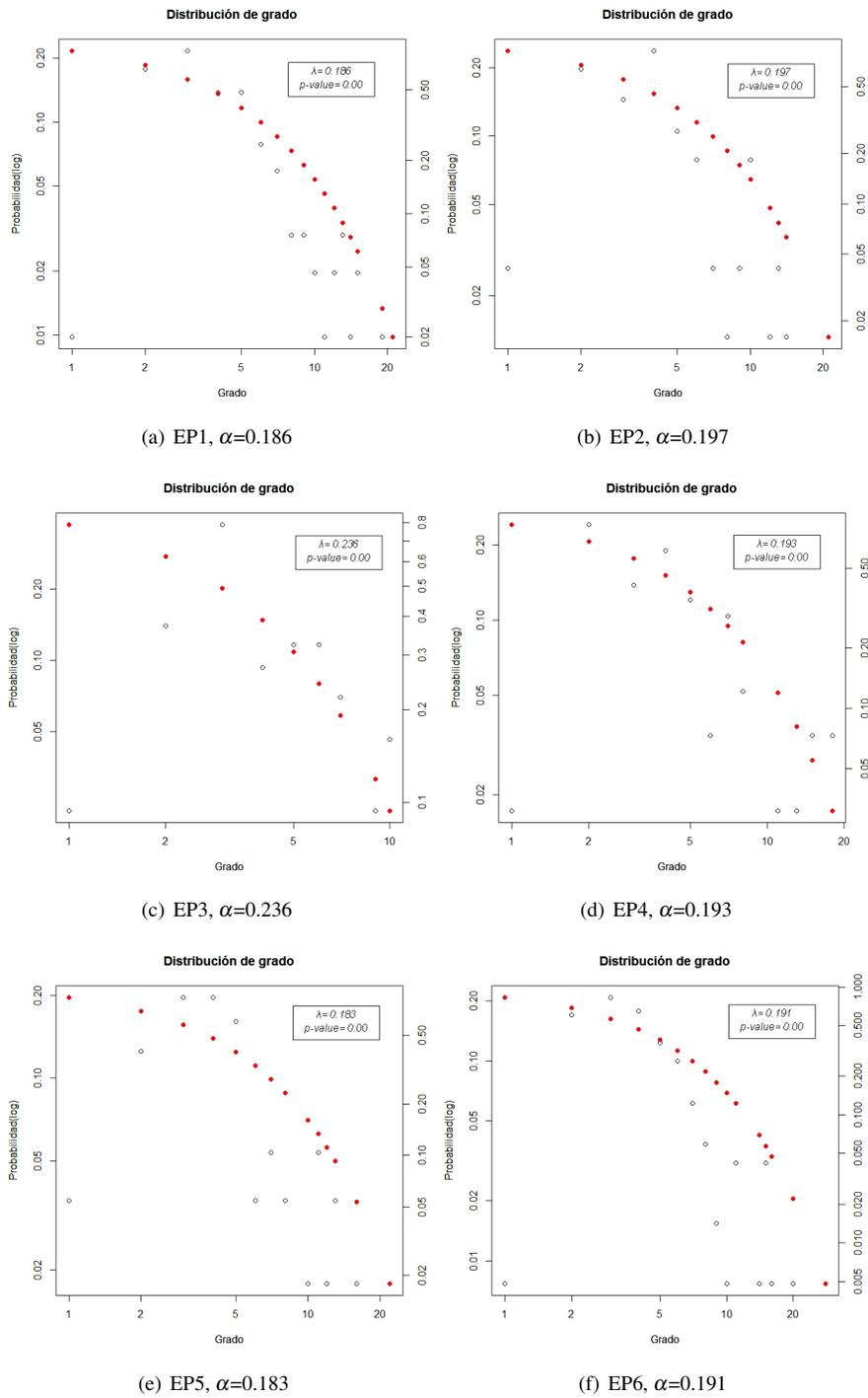


Figura 5.5: Gráficas de la distribución de grado de *Elvis Presley*.
 Fuente: *Elaboración propia*.

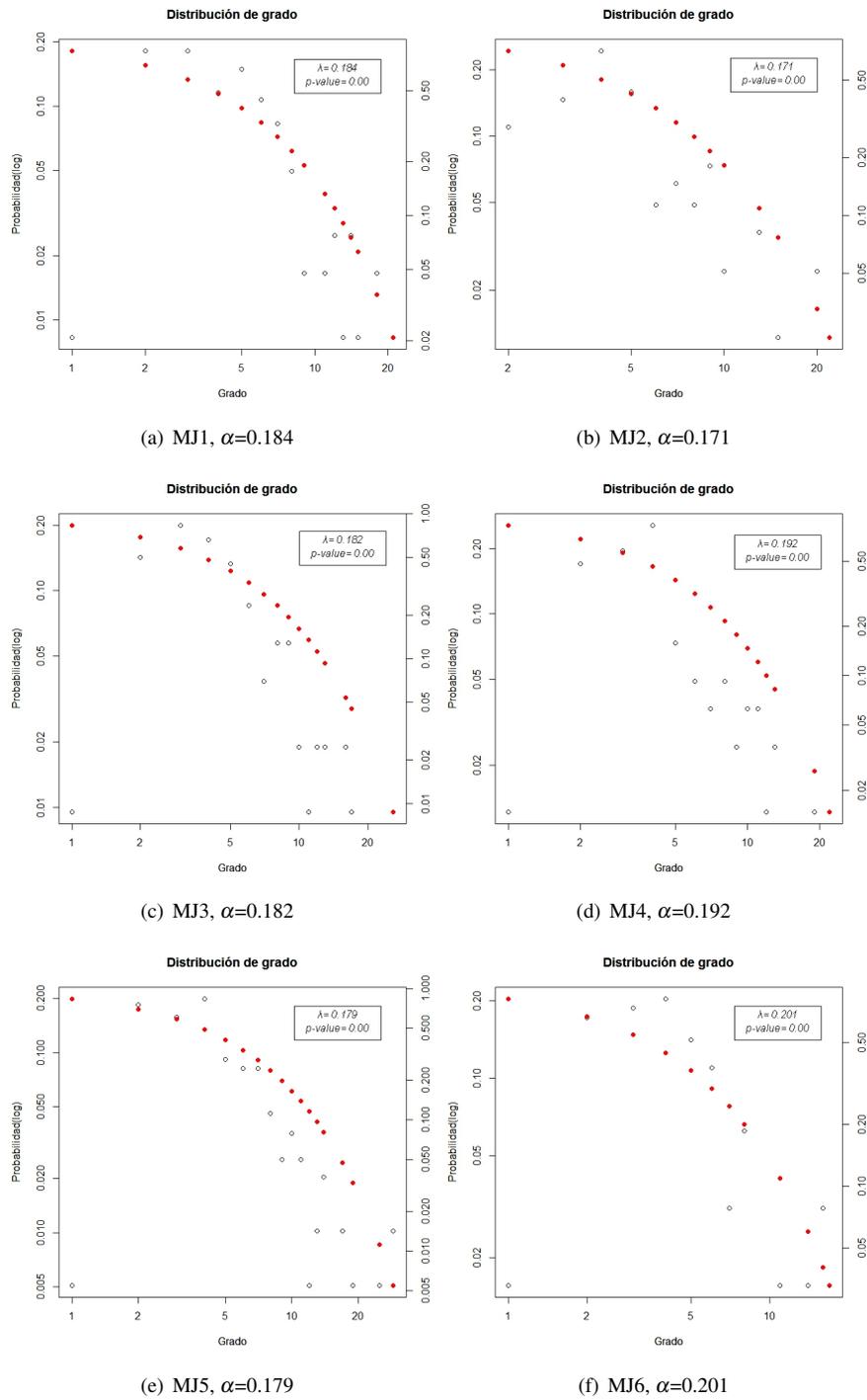


Figura 5.6: Gráficas de la distribución de grado de las redes de *Michael Jackson*.
 Fuente: *Elaboración propia*.

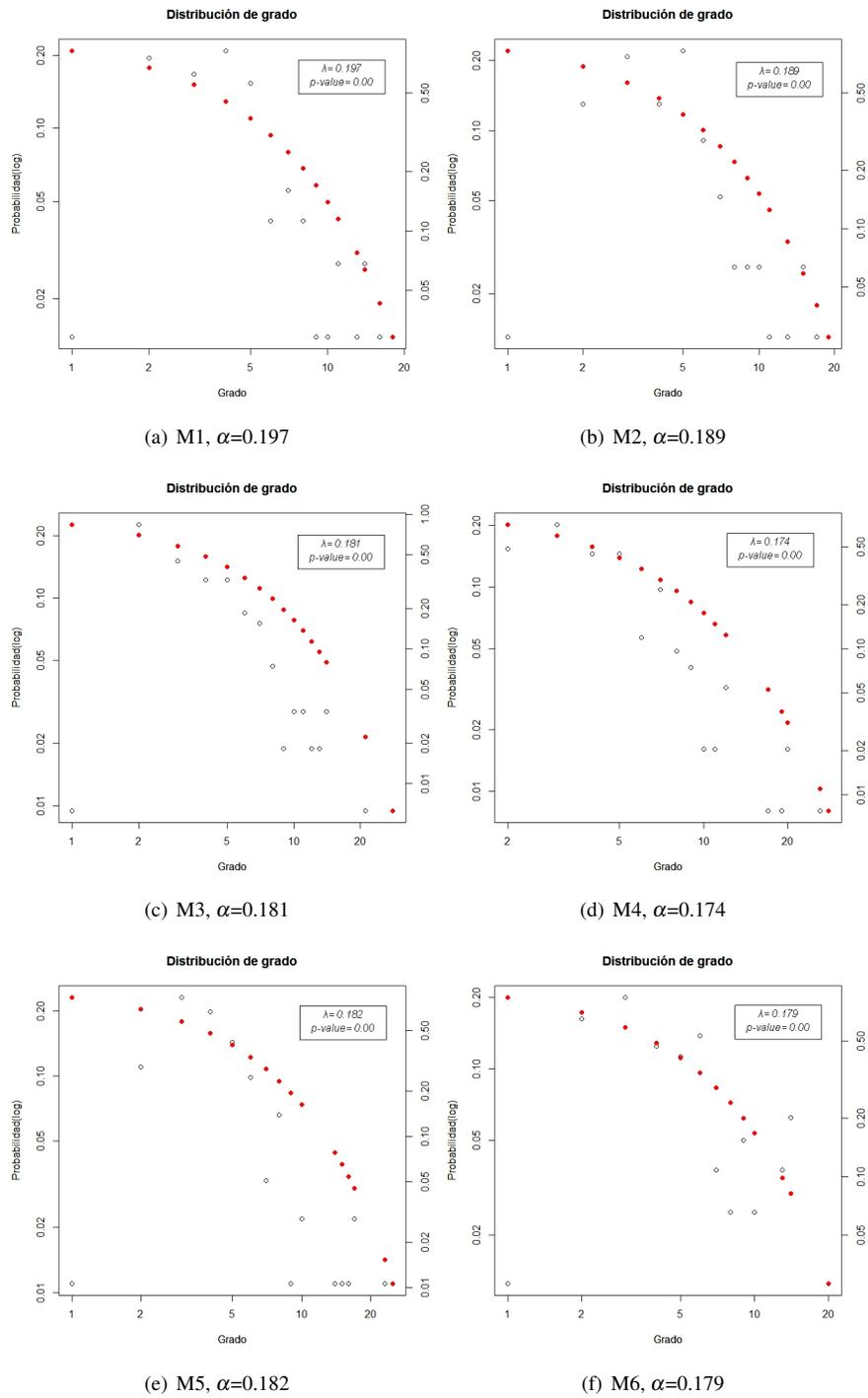


Figura 5.7: Gráficas de la distribución de grado de las redes de *Madonna*.
 Fuente: *Elaboración propia*.

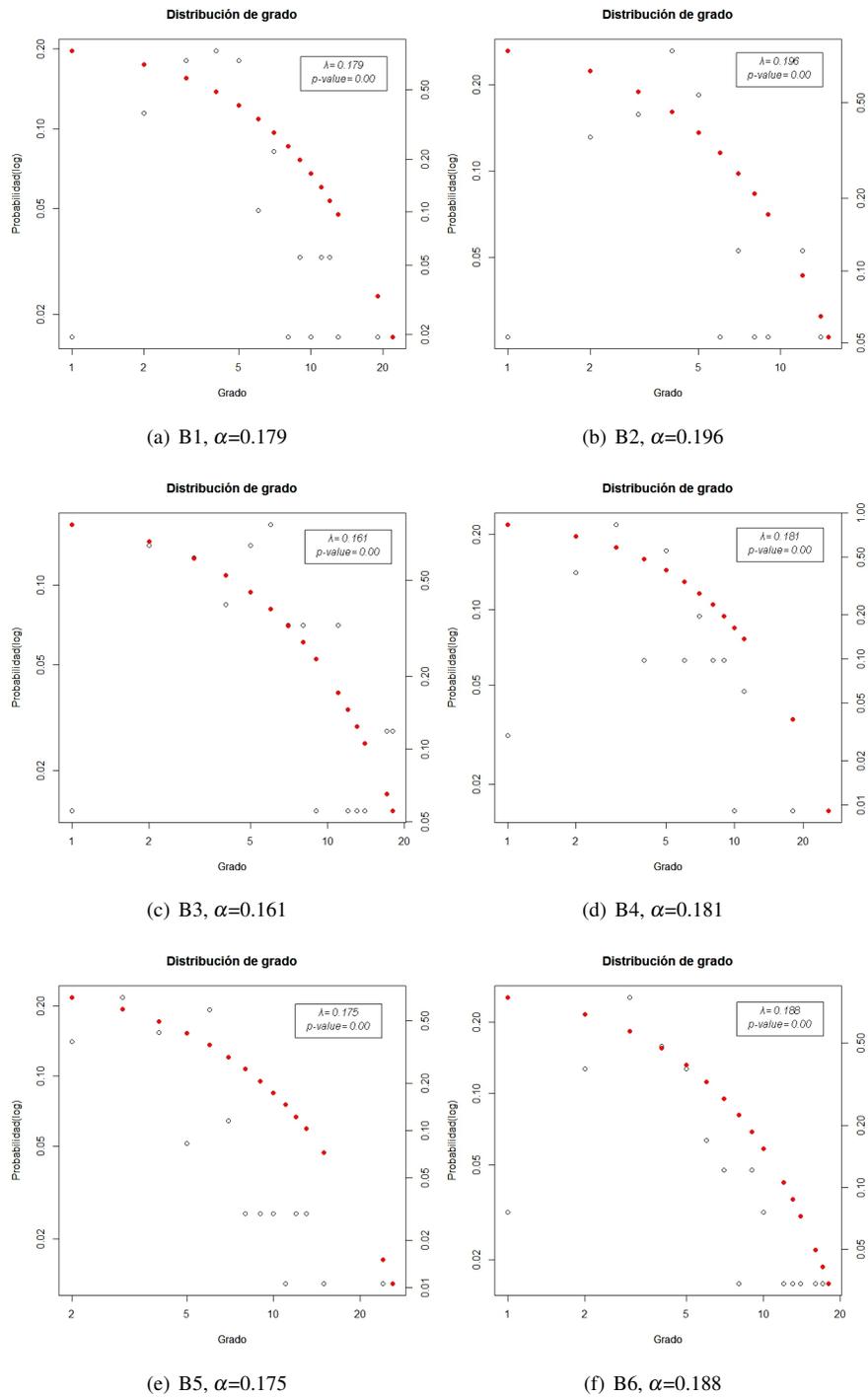


Figura 5.8: Gráficas de la distribución de grado de los *Beatles*.
 Fuente: Elaboración propia.

Conclusiones

Al inicio de este trabajo el objetivo era evidenciar la fuerte conexión que existe entre las matemáticas, específicamente la teoría de redes, y los lingüistas, además de mostrar que lo probado por Lucas Lacasa y colaboradores en 2008, era válido en las redes de lenguaje construidas por medio del algoritmo de visibilidad (NVG), es decir, que cumplían con ser conectadas, no dirigidas e invariantes bajo transformaciones [?].

En vista de que el lenguaje puede entenderse como una serie de tiempo, pues puede registrarse una palabra por unidad de tiempo, es posible transformarlo en redes. Estas redes no son únicas, puesto que es posible construirlas con las palabras como nodos e identificando la relación que las une, por ejemplo, la relación puede establecerse si las palabras son sinónimos, antónimos, o incluso si las palabras suenan parecido. En los casos analizados en este trabajo, la relación en cada uno de ellos quedó definida a través del algoritmo de visibilidad, que establece que dos nodos estarán conectados si existe visibilidad entre ellos, es decir, si existe una línea recta que conecte ambos datos sin que interseque algún otro dato.

En cuanto a la metodología usada en el estudio, es necesario hacer énfasis en la elaboración del programa del algoritmo de visibilidad en la plataforma *RStudio*, ya que fue hecho desde cero y cuenta con algunos puntos que se podrían mejorar en trabajos posteriores, por ejemplo, la inclusión de signos ortográficos para poder analizar los textos traducidos al español o bien, la inclusión de los posesivos del inglés para poder tomar los textos tal cual fueron escritos.

De los resultados obtenidos tras el análisis, destaca que todas las canciones tienen un *clustering* mayor a 0.4 y en los discursos se tienen valores menores a 0.4. Además, la eficiencia en las redes resulta ser mayor en las canciones y, al tratarse del inverso de la distancia entre nodos, podemos agregar que en general, las canciones tienen una distancia media menor que los discursos, es decir, que se necesita alrededor de 3.3 palabras intermedias para llegar de una palabra a cualquier otra en una canción, mientras que en los discursos se necesitan 4.9 palabras para conseguirlo. Estos resultados, mayor *clustering*, mayor eficiencia y menor distancia media, confirman lo que todos suponíamos, que las canciones son más fáciles de recordar que los discursos. En cuanto a

la asortatividad, los nodos tienden a conectarse con otros similares a ellos en cuanto al número de conexiones (grado), si las palabras usadas en las canciones son similares en longitud, se esperaba tener una red homogénea, sin embargo, los valores obtenidos tanto para las redes de las canciones analizadas como para los discursos, son muy cercanos a cero, lo que implica que son redes neutrales.

Bajo el contexto que se aborda en esta investigación, un valor alto de coeficiente de aglomeración, implica que es más fácil recordar ese conjunto de palabras, es decir, la letra de esas canciones o esos discursos. No se puede perder de vista que una palabra puede estar relacionada con la misma palabra cuando esta cambia ligeramente, así que aunque ese texto será más fácil de recordar, también será más probable confundir con facilidad una palabra por otra. Con esto en mente, al comparar canciones con canciones, se obtiene que el grupo *The Beatles* presenta los valores más altos en esta métrica y *Madonna* tiene la diferencia más grande entre su valor máximo y mínimo; para verificar que lo dicho al inicio del párrafo se cumple, bastaría un experimento en el que se pida a la gente memorizar letras de ambos intérpretes y revisar si es verdad que aunque son más fáciles de memorizar las canciones del grupo, también es fácil confundir las palabras. En vista de que el grupo inglés presenta el valor más alto de aglomeración y de eficiencia, se puede afirmar que sus canciones son las más fáciles de recordar de los artistas seleccionados y si a estos resultados se les suma que tiene el valor más pequeño de distancia media, se corrobora esta afirmación.

En cuanto a los discursos, es posible observar que el valor promedio más alto de *clustering*, lo tiene *Mahatma Gandhi* con 0.4, dicho valor iguala al promedio obtenido en las canciones y entonces, se podría explicar el por qué es uno de los personajes más citados hoy en día [?].

A partir del hecho de que las personas estamos constantemente comparando para comprender mejor la realidad, surgió la idea de hacer comparaciones entre los textos de ciertos personajes para mostrar que a pesar de vivir en distintas épocas, tener profesiones que no se relacionan e incluso profesar diferentes religiones, no son tan parecidos o diferentes como era de pensarse, o al menos, según las métricas de las redes obtenidas. Por ejemplo, al comparar a Churchill y Hitler, se obtuvo en todas las métricas analizadas valores similares que indican que se tratan de redes neutrales y con grados promedio cercanos a 6. Las pequeñas diferencias en distancia media y eficiencia, pueden explicarse por el número de palabras promedio que se utilizó en los discursos.

Por otra parte, por la construcción del algoritmo quedo evidenciado que las redes obtenidas a través del algoritmo de visibilidad estaban conectadas y sin dirección, pero para probar que presentan tendencia a ser redes libres escala, fue necesario utilizar la prueba de bondad de ajuste de *Kolmogorov-Smirnoff*. Con base en dicha prueba, y a los valores de α y p - *value* obtenidos, se demostró que todas las redes obtenidas de

los discursos, por medio del algoritmo NVG, presentaron distribuciones de grado con tendencia a ser redes libres de escala, es decir, que si se tomaba parte de la distribución, no cambiaba su forma y eso sucede a consecuencia de que nodos con grados muy diferentes entre sí coexistan en la misma red, es decir, hay presencia de hubs y como se mencionó en capítulos anteriores, esto es una característica básica de este tipo de redes. Además, en las canciones, por el hecho de tratarse de redes con muy pocos nodos y al tener alturas similares, se obtuvo un valor de $\alpha > 3$ y un ajuste no confiable del modelo de ley de potencia así que tras realizar un análisis de descarte como lo plantea Aaron Clauset y colaboradores en [?], se optó por ajustarles una distribución exponencial que arrojó un mejor ajuste tras aplicar la prueba de K-S, sin embargo, por ser redes muy pequeñas, sería preferible aplicar otro tipo de prueba para comprobar este hecho (por ejemplo, la prueba de χ^2 o métodos de regresión no lineales).

Por último, pero no menos importante, cabe recalcar que las aplicaciones que tiene la teoría de redes, no se reducen al campo de las matemáticas ni al de la lingüística, sino que al estar estas estructuras presentes en muchos de los sistemas que usamos día a día, por ejemplo, sistemas de transporte, la web, etc., tiene aplicación en áreas diversas que abarcan desde la biología hasta la economía. De forma específica, el algoritmo de visibilidad ha sido usado para estudiar las características organizacionales de las series de tiempo complejas de diferentes sistemas en economía e incluso de sismología, como referencia se puede consultar [?] o [?], por mencionar algunos.

En cuanto a esta tesis, se espera que sirva como base para futuras investigaciones que implementen el algoritmo NVG, principalmente en el análisis de textos, y que sirva como punto de partida para investigaciones, que tengan como objetivo mostrar que el puente que existe entre las ciencias exactas con las otras áreas de estudio, es más corto de lo que se piensa.

Anexos

Anexo A

Biografías

Las biografías son textos que enseñan cómo detrás de las revoluciones sociopolíticas, estéticas y científicas que han conducido hasta el mundo contemporáneo, están hombres y mujeres que fueron movidos por sus anhelos, a veces nobles y otras ruines.

Están ordenadas cronológicamente para poder recorrer las diferentes etapas de la mano de cada uno de los personajes que definieron el curso de la historia. En estas biografías, se incluyen datos de lo más relevante de sus carreras, para evidenciar el por qué fueron considerados para esta investigación.

Todas las imágenes se obtuvieron del sitio público *Wikimedia Commons*.

Mahatma Gandhi (1869-1948)

*“Si quieres cambiar al mundo,
cámbiate a ti mismo.”*
-Mahatma Gandhi

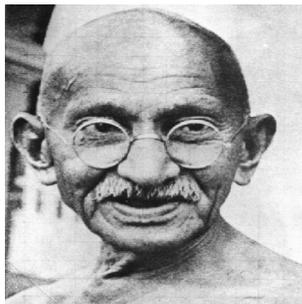


Figura A.1: Mahatma Gandhi. Mayor líder espiritual de la India moderna, al tiempo que es reconocido en el mundo por su política del pacifismo.

Su nombre real era Mohandas Karamchand Gandhi, posteriormente sería llamado Mahatma que significa *gran alma*. Nació el 2 de octubre de 1869 en Porbandar, India y murió asesinado el 30 de enero de 1948.

Fue abogado, político y pensador indio, conocido principalmente por conducir la independencia de la India a través de métodos no violentos como la huelga de hambre. Sus protestas sirvieron como inspiración para movimientos posteriores, por ejemplo, el de Martin Luther King.

Sir. Winston Churchill (1874-1965)

“A menudo me he tenido que comer mis palabras, y he descubierto que son una dieta equilibrada.”

-Sir. Winston Churchill



Figura A.2: Sir. Winston Churchill. Uno de los políticos británicos más notables del siglo XX.

Albert Einstein (1879-1955)

“La mente es como un paracaídas... sólo funciona si la tenemos abierta.”

-Albert Einstein

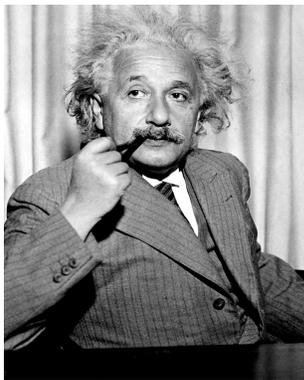


Figura A.3: Albert Einstein. Gran genio que revolucionó las ciencias físicas con la teoría de la relatividad.

Sir. Winston Leonard Spencer Churchill nació el 30 de noviembre de 1874 y falleció el 24 de enero de 1965 en Reino Unido. Es recordado por su papel como primer ministro del Reino Unido durante la Segunda Guerra Mundial aunque destacó también como pintor, historiador, biógrafo y orador.

Fue el gran forjador de la *Gran Alianza* conformada por Estados Unidos, la Unión Soviética y el Imperio británico, esta alianza consiguió la victoria y rendición de Alemania.

Obtuvo el Premio Nobel de Literatura por sus *“Memorias sobre la Segunda Guerra Mundial”*.

Uno de los más grandes genios de la humanidad, elegido por la revista *Time* [?] como el personaje más importante del siglo XX. Nació el 14 de marzo de 1879 en Alemania, y falleció el 18 de abril de 1955.

En el ámbito de las ciencias físicas, llevó a cabo una revolución cuyos efectos aún no han podido medirse en toda su amplitud, sentando las bases del uso de la energía nuclear y cambiando el concepto que se tenía del Universo.

Al concluir la Segunda Guerra Mundial, se dedicó a encarnar la figura del sabio comprometido con la paz y la defensa de los derechos civiles, siendo uno de los líderes del *World Government Movement*.

Adolf Hitler (1889-1945)

*“De las entrañas del oeste de Europa,
un niño nacerá en una familia pobre,
él, con su hablar, seducirá a las masas.”*
-Quatrain 35, Siglo III



Figura A.4: Adolf Hitler. Su sueño autoritario cobró forma en la ideología nazi y en la fundación del Tercer Reich.

Nació el 20 de abril de 1889 en Austria y falleció el 30 de abril de 1945 en Alemania.

Al finalizar la Primera Guerra Mundial, entró al Partido Obrero Alemán (DAP) que años más tarde se convertiría en el partido nazi. Se dio el título de *Führer* o líder máximo, e instituyó el uso de sus dos símbolos míticos célebres: la bandera con la cruz gamada y el saludo *Heil!*, que significa *Salve!*, con el brazo derecho en alto. Intentó derrocar al gobierno y fue a prisión donde escribió *Mein Kampf* (Mi Lucha), una obra autobiográfica donde manifestaba sus ideas sobre la selección racial y el exterminio de las “razas inferiores”.

Nelson Mandela (1918-2013)

*“La mayor gloria de vivir yace no en
nunca caer, sino en levantarse cada vez
que caigas.”*
-Nelson Mandela



Figura A.5: Nelson Mandela. Símbolo de lucha contra la discriminación racial y primer presidente negro de Sudáfrica.

Rolihlahla Mandela, considerado padre de la nación sudafricana, fundador de la democracia y emancipador de la nación, nació el 18 de julio de 1918 y falleció el 5 de diciembre de 2013 en Sudáfrica.

Tras 27 años en prisión por alta traición y acciones contra el gobierno, fue nombrado presidente de Sudáfrica después de las primeras elecciones multirraciales del país en 1994, lo que lo convirtió en el primer presidente negro de esa nación. Recibió más de 250 reconocimientos y títulos.

Se convirtió en símbolo de lucha contra la discriminación racial y evitó que el país estallara en una guerra civil durante la transición del *apartheid*^a a la democracia.

^aSe refiere a un sistema de distritos raciales que arrebató a los negros la condición de ciudadanos y los relegó a vivir en ciertos territorios destinados para ellos.

Margaret Thatcher (1925-2013)

“Si quieren que algo sea hecho, pregúntenle a una mujer.”
-Margaret Thatcher



Figura A.6: Margaret Thatcher. Primera mujer en liderar un partido mayoritario en Reino Unido y la primera jefa de gobierno.

Fue la primera mujer en liderar un partido mayoritario en Reino Unido y la primera jefa de gobierno, por lo que se convirtió en el símbolo de la mujer que se impone en un mundo de hombres; fue reelegida dos veces y es recordada por su política de recortes y su negativa a rectificar sus decisiones, de ahí su apodo de la *dama de hierro*.

Sus ideas serían resumidas en el término *thatcherismo* y aún tienen el poder de provocar filias y fobias tanto en el país como en el mundo.

Margaret Hilda Thatcher nació el 13 de octubre de 1925 en Grantham, Reino Unido y falleció a la edad de 87 años en Londres, el 8 de abril de 2013.

Elvis Presley (1935-1977)

“La verdad es como el sol, puedes ocultarla por un tiempo pero no va a desaparecer.”
-Elvis Presley



Figura A.7: Elvis Presley. Con su voz y movimientos corporales hizo accesible a millones de personas una nueva música denominada *rock and roll*.

Fue un cantante y actor estadounidense que con su voz, movimientos del cuerpo y una gran personalidad, hizo accesible a millones de personas una nueva música que se denominó *rock and roll*, de la cual llegó a ser rey indiscutible. Además, supo encarnar la rebeldía de una generación cansada de guerras [?].

Elvis Aaron Presley nació el 8 de enero de 1935 en Tupelo, Estado Unidos, y falleció el 16 de agosto de 1977 en Memphis.

Aunque su popularidad decayó con la llamada *invasión británica* encabezada por el grupo The Beatles, Elvis es un ícono de la historia de la música y para atestiguarlo, están sus más de cien discos de oro, platino y multiplatino.

Donald Trump (1946-)

“Hagamos a América grande otra vez.”
-Donald Trump



Figura A.8: Donald Trump. Actual Presidente de Estados Unidos.

Donald John Trump nació el 14 de junio de 1946 en Nueva York, Estados Unidos. Es ejecutivo, político, empresario multimillonario y actual presidente de los Estados Unidos, miembro del Partido Republicano. También es presidente de la *Trump Organization* y fundador de *Trump Entertainment Resorts*; así mismo, fue presentador del programa *The Apprentice*.

El 16 de junio de 2015 anunció su candidatura para las elecciones y desde ese día, los medios de comunicación lo confirmaron como el candidato más adinerado, polémico y narcisista que se hubiera presentado en la historia a la contienda por la presidencia de ese país.

Hillary Clinton (1947-)

“Todos los días los americanos necesitan un campeón, y yo quiero ser ese campeón.”
-Hillary Clinton



Figura A.9: Hillary Clinton. Candidata por el partido demócrata a la presidencia de Estados Unidos en 2016.

Hillary Diane Rodham Clinton, más conocida como Hillary Clinton, nació en Chicago el 26 de octubre de 1947. Ha sido Secretaria de Estado, primera dama, senadora, profesora de derecho y activista. Fue la primera mujer electa al senado de Estados Unidos por el estado de Nueva York y se presentó como candidata del Partido Demócrata para contender por la presidencia en 2008, ganando más elecciones primarias que ninguna otra mujer en la historia de Estados Unidos.

Anunció su postulación a la presidencia el 12 de Abril de 2015 y de haber ganado la contienda presidencial, se hubiera convertido, después de 44 hombres, en la primera mujer presidenta de Estados Unidos.

Angela Merkel (1954-)

“Un buen político siempre tiene dudas y, por tanto, revisa permanentemente sus respuestas.”

-Angela Merkel



Figura A.10: Angela Merkel. Primera mujer en asumir la cancillería de Alemania.

Angela Dorothea Kasner, mejor conocida como Angela Merkel, es la primera mujer desde el nacimiento del estado alemán (1870), y la primera persona originaria de la extinta República Democrática Alemana, que asume la cancillería de Alemania. Ha sido reelegida en dos ocasiones (2009 y 2013) y entre sus acciones destacan la recesión nacional durante la crisis económica mundial.

Nació el 17 de julio de 1954 en Hamburgo, Alemania Occidental.

Hasta 2015, son diez años en los que ha sido nombrada por la revista *Forbes* como la mujer más poderosa del mundo.

Madonna (1958-)

“Mucha gente tiene miedo de reconocer lo que quiere...por eso no lo consigue.”

-Madonna

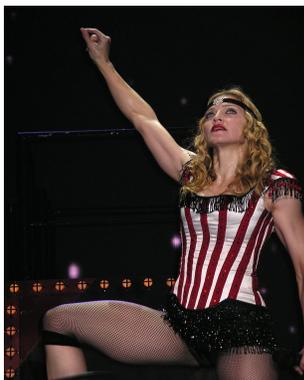


Figura A.11: Madonna. Ícono de la moda y de la música pop que, sigue vigente.

Madonna Louise Veronica Ciccone, nació en Michigan, Estados Unidos el 16 de agosto de 1958. Es cantante, compositora, actriz y productora denominada como la reina del pop. Se ha convertido en un ícono internacional por su ambición, y la innovación y provocación que muestra en sus videos y espectáculos.

Algunos de sus discos están entre los mejores de la historia y ha sido multigañadora de premios *Grammy*; ha vendido alrededor de 200 millones de álbumes y en el 2000, el libro *Guinness* de los récords publicó que es la artista femenina más exitosa de todos los tiempos.

Michael Jackson (1958-2009)

*“Cualquier cosa que canto,
la siento de verdad, no canto, siento.”*
-Michael Jackson



Figura A.12: Michael Jackson. Genio musical y artístico más premiado de la historia por sus innovaciones en coreografía y música.

Cantante y bailarín estadounidense, nacido el 29 de agosto de 1958 en Indiana, que fue la máxima estrella pop de los años 1980 a 1990. Apodado el rey del pop, vendió más de 750 millones de discos en todo el mundo, ganó alrededor de 405 galardones a lo largo de su carrera y popularizó movimientos de baile como el robot o el *moon walk* (caminata lunar) que marcaron tendencia. Se convierte en rey tras su álbum *Thriller* (1982) y el video de su canción *Billie Jean* fue el primero de un artista negro que emitiera el canal MTV.

Es también calificado por el Libro *Guinness* de los récords, como el artista que más ha donado a la caridad en la historia y fue nominado dos veces para recibir el premio Nobel de Paz por sus actos de ayuda comunitaria [?].

The Beatles (1962-1970)

“Somos más populares que Jesús.”
-John Lennon



Figura A.13: The Beatles. Banda británica que transformó por completo el mundo del *rock and roll*.

Grupo británico más admirado y popular de la década de 1960 y uno de los más influyentes en la historia de la música moderna

Los integrantes del grupo fueron John Lennon, Paul McCartney, George Harrison y Ringo Starr, y se distinguían por el corte de cabello y por una actitud desafiante.

Terminaron con la supremacía del *rock and roll* americano y su obra más revolucionaria, el disco *Sgt. Pepper's Lonely Hearts Club Band*, dio origen al rock sinfónico y a la música psicodélica.

En 1970, el grupo se separó y siguieron sus carreras como solistas.

Anexo B

Algoritmo de visibilidad en R

```
#Se instalan las paqueterías necesarias
install.packages("rgl")
install.packages("igraph")

#Se llaman a las librerías de las paqueterías recién instaladas
library(rgl)
library(igraph)

#Se lee el archivo de texto y se guarda en la variable texto
texto<-read.table("C:Tesis/Discursos/AH2.txt",header=F,sep="\r")

#El texto guardado se convierte en letras mayúsculas
texto=toupper(texto$V1)

#Se convierte el texto en una lista separada por espacios
texto_split = strsplit(texto, split=" ")

#Se crea el data.frame a partir de la lista
texto_col = as.character(unlist(texto_split))
texto_col = data.frame(texto_col)
names(texto_col) = c("V1")

#Se eliminan todos los números, signos de puntuación y espacios
texto_col$V1 = sub("[[:digit:]]+", "", texto_col$V1)
texto_col$V1 = sub("[[:punct:]]+", "", texto_col$V1)
texto_col$V1 = sub("[[:space:]]+", "", texto_col$V1)

#Se crea una variable que guarde la longitud de la palabra
texto_col$largo = nchar(texto_col$V1)

#Se añade el texto al programa para poder manejarlo
```

```

attach(texto_col)

#Se crea un data.frame con las palabras y el número de veces que aparecen en el
texto
b <- data.frame(table(V1))

#Se asigna una propiedad más al data frame que es la longitud de cada palabra
b$largo = nchar(as.character(b$V1))

#Media de las longitudes
mean(b$largo)

#Longitud máxima
max(b$largo)

#Longitud mínima
min(b$largo)

#Para obtener las gráficas
#t es un vector de 1 hasta el número de palabras que tenga el texto
t <- c(1:length(b$V1))

#y corresponde a la longitud de las palabras
y <- b$largo

#Se abre una nueva ventana y se grafica el número de palabra con su longitud
#Se añaden puntos en la parte superior de cada barra creada
#Todo esto se guarda en el objeto llamado algorit
x11()
algorit <- plot(t,y,type="h",main="Palabras y número de letras de cada una",cex.main=2,
xlab="Número de cada palabra",ylab="Número de letras",cex.lab=1.5,col=81)
points(t,y,pch=16,col="red")

#Se obtiene la matriz de adyacencia de la gráfica obtenida
#Es decir, con la matriz se sabrá qué palabra se unió con qué palabra
adyac <- matrix(data=0,nrow=length(t),ncol=length(t))

#Criterio de visibilidad
for (a in t)
for (b in t)
if (t[b]-t[a]>0)
if (t[b]-t[a]==1)
segments(t[a],y[a],t[b],y[b],col="blue")
adyac[t[a],t[b]] <- 1 else
contador=0
for (c in t[a+1]:t[b-1])

```

```
if (t[a]<t[c] & t[c]<t[b])
if (y[c]<y[a]+((y[b]-y[a])((t[c]-t[a])/(t[b]-t[a])))
contador<-contador+1
if (contador==length(t[a+1]:t[b-1]))
segments(t[a],y[a],t[b],y[b],col="blue")
if (contador==length(t[a+1]:t[b-1]))
adyac[t[a],t[b]]<-1

#Se crea la red, es decir, g obtiene los arcos de la red
g<-graph.adjacency(adyac,mode="undirected",weighted = NULL)
tkplot(g,vertex.color="white")
```

Anexo C

Textos seleccionados y dónde encontrarlos

Adolf Hitler	[?]
	[?]
	[?]
Albert Einstein	[?]
	[?]
Angela Merkel	[?]
	[?]
Donald Trump	[?]
	[?]
Hilary Clinton	[?]
Mahatma Gandhi	[?]
Margaret Thatcher	[?]
Nelson Mandela	[?]
Winston Churchill	[?]
	[?]

Canciones Las letras de las canciones fueron obtenidas de [?] y se cotejaron con las páginas oficiales de cada cantante.

Bibliografía

- [Hit, 2009] (2009). Hitler's blood oath- the effect nostradamus.
- [Imp, 2015] (2015). Importancia de la oratoria. url *http* : [//www.importancia.org/oratoria.php](http://www.importancia.org/oratoria.php). Accedido 10-01-2016.
- [Alatorre, 2002] Alatorre, A. (2002). *Los 1001 años de la lengua española*. Fondo de cultura económica. Tercera edición.
- [Aldana, 2006] Aldana, M. (2006). Redes complejas.
- [Appleman, 2016] Appleman, E. M. (2016). P2016 race for the white house. url *http* : [//www.p2016.org/](http://www.p2016.org/). Accedido 20-01-2017.
- [AZlyrics, 2016] AZlyrics (2016). Azlyrics. url *https* : [//www.azlyrics.com/](https://www.azlyrics.com/). Accedido 02-01-2017.
- [Barabási, 2014] Barabási, A.-L. (2014). *The Scale-Free Property*.
- [Barabási y Albert, 1999] Barabási, A.-L. y Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286.
- [Bhavan, 2016] Bhavan, M. (2016). Mahatma gandhi information website. url *http* : [//www.gandhi-manibhavan.org/](http://www.gandhi-manibhavan.org/). Accedido 20-11-2016.
- [Bondy y Murty, 1976] Bondy, J. A. y Murty, U. (1976). *Graph Theory with Applications*. MacMillan.
- [Boroditsky, 2011] Boroditsky, L. (2011). How language shapes thought. *Scientific American*.
- [Caldarelli y Catanzaro, 2012] Caldarelli, G. y Catanzaro, M. (2012). *Networks A very short introduction*. Oxford University Press.
- [Clauset *et al.*, 2007] Clauset, A., Shalizi, C. R., y Newmann, M. (2007). Power-law distributions in empirical data. *Carnegie Mellon University Research Showcase, Department of Statistics*, (6).
- [Clinton, 2016] Clinton, H. (2016). Hillary clinton speeches. url *https* : [//hillaryspeeches.com/](https://hillaryspeeches.com/). Accedido 04-02-2017.

- [Cowperwait y Metcalfe, 2009] Cowperwait, P. S. y Metcalfe, A. V. (2009). *Introductory Time Series with R*. Springer.
- [Credimar, 2002] Credimar (2002). *Grandes personajes de la humanidad*, volumen 4.
- [Cusnier, 1999] Cusnier, G. M. (1999). La corte del rey del pop. url www.lacortedelreydelPOP.com/quienes28.htm. Accedido 27-01-2016.
- [Delbyck, 2016] Delbyck, C. (2016). Madonna opens amy schumer's stand-up show by offering blow jobs to hillary voters. url http://www.huffingtonpost.com.mx/entry/madonna-opens-amy-schumers-stand-up-show-by-offering-to-blow-hillary-voters_us_5807aa3de4b0dd54ce36f354. Accedido 20-01-2017.
- [Deutsche Welle, 2015] Deutsche Welle (2015). Angela merkel's new year message. url www.dw.com/en/angela-merkels-new-year-message/av-18953209. Accedido 10-10-2016.
- [Dorogotsev y Mendes, 2003] Dorogotsev, S. N. y Mendes, J. (2003). *Evolution of Networks From Biological Nets to the Internet and WWW*. Oxford University Press.
- [Eidenmuller, 2016] Eidenmuller, M. E. (2016). American rhetoric online speech bank. url <http://www.americanrhetoric.com/speeches/alberteinsteinpostwarworld.htm>. Accedido 22-08-2016.
- [Einstein, 2011] Einstein, A. (2011). *Out of my later years: The scientist, philosopher and man portrayed through his own words*.
- [Foundation, 2016a] Foundation, M. T. (2016a). Margaret thatcher foundation. url <http://www.margaretthatcher.org/document/102939>. Accedido 03-12-2016.
- [Foundation, 2016b] Foundation, N. M. (2016b). Nelson mandela foundation living the legacy. url <http://db.nelsonmandela.org/speeches/>. Accedido 12-11-2016.
- [Fund, 2016] Fund, W. F. (2016). World future fund. url www.worldfuturefund.com/Reports2013/hitlerenablingact.htm. Accedido 05-10-2016.
- [García Bellido y Jornet Meliá, 2010] García Bellido, R. and González Such, J. y Jornet Meliá, J. (2010). Spss: Pruebas no paramétricas kolmogorv-smirnov. *innova-MIDE*.
- [Gilbert, 2012] Gilbert, M. (2012). *Churchill The power of words*. Da Capo Press.
- [Grzybek, 2006] Grzybek, P. (2006). *History and Methodology of Word Length Studies: The State of Art*, volumen 31, pp. 15-90.
- [Gúzman-Vargas et al., 2015] Gúzman-Vargas, L., Obregón Quintana, B., Aguilar-Velázquez, D., Hernández-Pérez, R., y Liebovitch, L. S. (2015). Word-length correlations and memory in large texts: a visibility network analysis. *Entropy*, (17).

- [Hauser y et al., 2002] Hauser, M. D. y et al. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298:1569–1579.
- [Hitler, 2016] Hitler, A. (2016). The speeches of the führer. url [der – fuehrer.org/redenen.htm](http://der-fuehrer.org/redenen.htm). Accedido 05-10-2016.
- [Jaimes, 2015] Jaimes, A. (2015). Introducción a la ciencia de redes. url [https://rpubs.com/Alberto – /IntroduccionCienciaRedes](https://rpubs.com/Alberto-IntroduccionCienciaRedes). Accedido 27-03-2016.
- [John, 2012] John, P. (2012). *Churchill versus Hitler War of Words*. Bennion Kearny.
- [Kalinesti, 2016] Kalinesti, H. (2016). Lista del los 100 filósofos y humanistas mas citados en internet. url [arteysciencias.blogspot.mx/2008/12/lista – del – los – 100 – filsofos – y – humanistas.html](http://arteysciencias.blogspot.mx/2008/12/lista-del-los-100-filsofos-y-humanistas.html). Accedido 08-02-2017.
- [Keller, 2013] Keller, B. (2013). De rebelde a prisionero a presidente. url http://internacional.elpais.com/internacional/2013/12/06/actualidad/13863569367_122339.html. Accedido 17-01-2016.
- [Kracht, 2003] Kracht, M. (2003). *The mathematics of language*, volumen 63. Studies in generative grammar.
- [Lacasa et al., 2008] Lacasa, L., Luque, B., Ballesteros, F., Luque, J., y Nuño, J. C. (2008). From time series to complex networks: The visibility graph. *PNAS*, 105(13).
- [Lande y Snarskii,] Lande, D. y Snarskii, A. Compactified horizontal visibility graph for the language network.
- [Luque et al., 2009] Luque, B., Lacasa, L., Ballesteros, F., y Luque, J. (2009). Horizontal visibility graphs: Exact results for random time series. *Physical Review*, (80).
- [Martínez Arqué, 2011] Martínez Arqué, N. (2011). *Análisis, comparativa y visualización de redes sociales on-line representadas como grafo*. Proyecto Fi de Carrera 2010-2011. Universitat Pompeu Fabra.
- [Merkel, 2016] Merkel, A. (2016). Discursos la historia a través de los discursos de sus líderes. url [http://www.beersandpolitics.com/discursos/angela – merkel/182](http://www.beersandpolitics.com/discursos/angela-merkel/182). Accedido 11-10-2016.
- [Michel y et al., 2011] Michel, J.-B. y et al. (2011). Quantitative analysis of culture using millions of digitalized books. *Science*, 331:176–182.
- [Nuñez et al., 2012] Nuñez, A. M., Lacasa, L., Gomez, P., y Luque, B. (2012). *Visibility algorithms: A short review*, *New Frontiers in Graph Theory*.
- [Ocaña, 2003] Ocaña, J. C. (2003). El sitio web de la historia del siglo xx. url <http://www.historiasiglo20.org/BIO/>. Accedido 27-01-2016.
- [Pagel, 2012] Pagel, M. (2012). *War of words: The language paradox explained*. A deliberate language barrier.

- [Piantadosi *et al.*, 2011] Piantadosi, S. T., Tily, H., y Gibson, E. (2011). Word lengths are optimized for efficient communication. *PNAS*, 108(9).
- [Pratt, 1939] Pratt, F. (1939). *Secret and Urgent: The Story of Codes and Ciphers*. Cryptographic series. Aegean Park Press.
- [Quesada Granja, 2009] Quesada Granja, C. (2009). Distribución por longitud de las palabras de diferentes idiomas. url es.slideshare.net/quesadagranja/distribucion-por-longitud-de-las-palabras-de-diferentes-idomas-presentation. Accedido 14-04-2017.
- [Ruiza, 2016] Ruiza, M. (2004-2016). Biografías y vidas, la enciclopedia biográfica en línea. url <http://www.biografiasyvidas.com>. Accedido 27-01-2016.
- [Samuel *et al.*, 2010] Samuel, A., Strogatz, S. H., y Vitevitch, M. S. (2010). Comparative analysis of networks of phonologically similar words in english and spanish. *Entropy*, (12).
- [Sanz, 2015] Sanz, E. (2015). 15 frases geniales de albert einstein. url <http://www.muyhistoria.es/contemporanea/articulo/quince-frases-geniales-de-albert-einstein>. Accedido 20-01-2016.
- [Solé *et al.*, 2005] Solé, R. V., Corominas Murtra, B., Valverde, S., y Luc, S. (2005). Language networks: their structure, function and evolution. *Santa Fe Institute*.
- [Staff, 2016] Staff, T. (2016). Time. url <http://www.time.com/>. Accedido 20-01-2017.
- [Watts y Strogatz, 1998] Watts, D. y Strogatz, S. (1998). Collective dynamics of small-world networks. *Nature*, 393.
- [Zipf, 1935] Zipf, G. K. (1935). The psycho-biology of language: An introduction to dynamic philology. *MIT Press*.