

Análisis de la crisis del fentanilo mediante el uso de técnicas de redes y sistemas complejos



DECSAI
Universidad de Granada

Juan Barea Rojo
77391685C
barearojouan@correo.ugr.es

Índice

1. Selección de un dominio, definición de una pregunta de investigación y obtención de un conjunto de datos estructural inicial asociado.....	3
2. Construcción de la red compleja a analizar y visualizar.....	4
2.1 Estructura de las Redes.....	4
2.1.1. Red de Co-ocurrencia Textual.....	4
2.2.1 Red de Hipervínculos.....	5
2.2 Características de las Redes.....	5
2.2.1 Características de Red de Co-ocurrencia Textual.....	5
2.2.2 Características de Red de Hipervínculos.....	6
2.3 Procedimiento de Reducción.....	6
2.3.1. Poda de Aristas en la Red de Co-ocurrencia Textual.....	6
2.3.2. Poda de Nodos en la Red de Co-ocurrencia Textual.....	7
2.3.3. Poda de Enlaces en la Red de Hipervínculos.....	7
2.3.4. Poda de Nodos en la Red de Hipervínculos.....	7
2.4 Resultado de las Redes Reducidas.....	8
3. Métricas y análisis de la propiedades de la red.....	9
3.1. Red de Co-ocurrencia Textual.....	9
3.1.1 Métricas.....	9
3.1.2. Propiedades de la red.....	12
3.2. Red de Hipervínculos.....	13
3.2.1 Métricas.....	13
3.2.2. Propiedades de la red.....	17
4. Descubrimiento de comunidades en la red.....	18
4.1. Comunidades de Co-ocurrencia Textual.....	18
4.2. Comunidades de Hipervínculos.....	23
5. Red de Co-ocurrencia Textual Visualización.....	27
6. Red de Hipervínculos Visualización.....	28
7. Discusión de los resultados obtenidos.....	29
8. Conclusión y consideraciones finales.....	31



1. Selección de un dominio, definición de una pregunta de investigación y obtención de un conjunto de datos estructural inicial asociado.

He seleccionado el dominio de la percepción pública sobre la crisis del fentanilo como foco de mi investigación, dado su impacto significativo en la salud pública y la sociedad en general. La pregunta de investigación que me he planteado es:

¿Cómo se ha reflejado la visión del público sobre la crisis del fentanilo en el último tiempo en wikipedia?

La crisis del fentanilo se ha convertido en una de las emergentes amenazas para la salud pública en las últimas décadas. El fentanilo es un opioide sintético extremadamente potente, utilizado principalmente para tratar el dolor intenso, especialmente en pacientes con cáncer. Sin embargo, su alta potencia y rápida acción lo convierten en una sustancia con un alto potencial de abuso y dependencia. En los últimos años, el uso ilícito de fentanilo ha escalado de manera alarmante, contribuyendo a un aumento significativo en las tasas de sobredosis y muertes relacionadas con opioides. El impacto de la crisis del fentanilo es multifacético, afectando no solo la salud individual de quienes lo consumen, sino también la estructura social y económica de las comunidades. Las consecuencias incluyen la pérdida de vidas humanas, el aumento de la carga sobre los sistemas de salud y justicia, y la desintegración de familias y comunidades enteras. Además, la crisis ha generado un amplio debate público sobre la regulación de los opioides, las estrategias de prevención y tratamiento, y la responsabilidad de diversas entidades en la proliferación de esta sustancia.

He decidido utilizar Wikipedia como fuente principal para mi investigación, considerando que es una plataforma que refleja el conocimiento colectivo y las percepciones de sus usuarios. El artículo base de partida es [Fentanyl] (<https://en.wikipedia.org/wiki/Fentanyl>), a partir del cual planeo analizar cómo se ha desarrollado y actualizado la información sobre esta sustancia en respuesta a los eventos recientes de la crisis. Al examinar las ediciones y discusiones en Wikipedia, espero captar las tendencias y cambios en la comprensión pública del fentanilo, así como identificar los aspectos que más preocupan a la comunidad. Este enfoque me permitirá obtener un conjunto de datos estructurado que refleja las dinámicas de conocimiento y opinión pública sobre el fentanilo. Al comprender cómo se representa la crisis en una fuente tan accesible y ampliamente utilizada como Wikipedia, podré responder de manera más efectiva a mi pregunta de investigación y aportar insights valiosos sobre la percepción pública de este grave problema de salud.



2. Construcción de la red compleja a analizar y visualizar.

Para abordar la pregunta de investigación planteada, procedí a construir dos redes complejas utilizando el conjunto de datos obtenido a partir del scraping del artículo de Wikipedia sobre [Fentanyl] (<https://en.wikipedia.org/wiki/Fentanyl>). A continuación, detallo la estructura de ambas redes, sus características y los métodos empleados para su reducción y optimización.

2.1 Estructura de las Redes

2.1.1. Red de Co-ocurrencia Textual

- **Nodos:**
 - **Representación:** Los nodos de esta red representan **palabras individuales y bigramas** extraídas del texto de los artículos de Wikipedia relacionados con el fentanilo.
 - **Categorías:** Cada nodo está categorizado como *word* (palabra individual) o `bigram` (combinación de dos palabras).
 - **Atributos:** Cada nodo posee un atributo *Attribute* que indica la **frecuencia de aparición** de la palabra o bigrama en el conjunto de datos.
- **Enlaces:**
 - **Representación:** Los enlaces entre los nodos representan la **co-ocurrencia** de palabras y bigramas dentro de una ventana deslizante de tamaño 5 palabras.
 - **Tipo de Red:** La red es **no dirigida**, ya que la co-ocurrencia no implica una dirección específica entre las palabras.
 - **Ponderación:** Los enlaces están **ponderados**; el peso de cada enlace (*Weight*) refleja la **frecuencia de co-ocurrencia** entre los pares de palabras o bigramas.



2.2.1 Red de Hipervínculos

- **Nodos:**
 - **Representación:** Los nodos de esta red representan **artículos de Wikipedia** relacionados con el fentanilo.
 - **Categorías:** Cada nodo está categorizado como *link*.
 - **Atributos:** Cada nodo posee un atributo *Attribute* que actúa como un **placeholder** (valor por defecto *o*), ya que en esta red no se utiliza para frecuencias específicas.
- **Enlaces:**
 - **Representación:** Los enlaces entre los nodos representan **hipervínculos** entre los artículos de Wikipedia.
 - **Tipo de Red:** La red es **dirigida**, ya que los hipervínculos tienen una dirección específica de un artículo a otro.
 - **Ponderación:** Los enlaces están **ponderados**; el peso de cada enlace (*Weight*) se establece en *1*, ya que cada hipervínculo representa una única conexión directa.
 - **Tipos de Enlaces:**
 - **Directed:** Representa la dirección del hipervínculo de un artículo a otro.

2.2 Características de las Redes

2.2.1 Características de Red de Co-ocurrencia Textual

- **No Dirigida:** Las conexiones de co-ocurrencia no tienen una dirección inherente.
- **Ponderada:** Los pesos reflejan la frecuencia de co-ocurrencia, permitiendo identificar las relaciones más fuertes entre términos.



- **No Bipartita:** La red incluye tanto palabras individuales como bigramas, permitiendo conexiones entre cualquier par de nodos según su co-ocurrencia.

2.2.2 Características de Red de Hipervínculos

- **Dirigida:** Las conexiones tienen una dirección clara, indicando la relación de un artículo a otro.
- **Ponderada:** Aunque cada hipervínculo tiene un peso de 1, esto facilita la contabilización de múltiples conexiones entre los mismos nodos si existieran.
- **No Bipartita:** La red conecta artículos entre sí sin restringir las conexiones a dos conjuntos distintos de nodos.

2.3 Procedimiento de Reducción

Dado que ambas redes inicialmente contenían una gran cantidad de nodos y enlaces, era necesario aplicar procedimientos de reducción para hacerlas manejables y enfocadas en las relaciones más significativas.

2.3.1. Poda de Aristas en la Red de Co-ocurrencia Textual

- **Criterios:**
 - **Percentil de Peso:** Eliminé las aristas cuyo peso estaba por debajo del 45º percentil de la distribución de pesos.
 - **Peso Mínimo:** Además, establecí un **peso mínimo de 5** para retener una arista.
 - **Justificación:** Estos criterios aseguran que solo las relaciones de co-ocurrencia más fuertes y frecuentes se mantengan en la red, reduciendo el ruido y facilitando la identificación de patrones relevantes.



2.3.2. Poda de Nodos en la Red de Co-ocurrencia Textual

- Criterios:

- **Percentil de Frecuencia:** Eliminé los nodos cuya frecuencia estaba por debajo del **15º percentil** de la distribución de frecuencias de palabras.
- **Frecuencia Mínima:** Además, establecí una **frecuencia mínima de 5** para retener un nodo.
- **Justificación:** Al aplicar estos filtros, me aseguré de que solo las palabras y bigramas más frecuentes y representativas permanecieran en la red, mejorando la claridad y relevancia del análisis.

2.3.3. Poda de Enlaces en la Red de Hipervínculos

- Criterios:

- **Peso Mínimo:** Establecí un **peso mínimo de 1**, ya que cada hipervínculo representa una conexión única.
- **Justificación:** Dado que cada hipervínculo tiene un peso fijo, la poda se centra en mantener todas las conexiones relevantes sin eliminar ninguna, asegurando una representación completa de las relaciones entre los artículos.

2.3.4. Poda de Nodos en la Red de Hipervínculos

- Criterios:

- **Grado de Entrada y Salida:** Consideré tanto el grado de entrada como el de salida para determinar la relevancia de cada nodo.
- **Filtrado por Relevancia:** Eliminé nodos que no contribuían significativamente a la estructura general de la red, basándome en métricas como el centralidad.
- **Justificación:** Este enfoque garantiza que la red de hipervínculos refleje las conexiones más importantes y significativas entre los

artículos, facilitando un análisis más enfocado de la estructura de conocimiento sobre la crisis del fentanilo.

2.4 Resultado de las Redes Reducidas

Tras aplicar los procedimientos de poda, ambas redes resultantes presentan estructuras más enfocadas y manejables:

- **Red de Co-ocurrencia Textual:**

- **Nodos:** Representan las palabras y bigramas más frecuentes y significativas en el contexto de la crisis del fentanilo.
- **Enlaces:** Reflejan las relaciones de co-ocurrencia más fuertes, facilitando la identificación de términos clave y sus interrelaciones dentro de la narrativa pública sobre el fentanilo.

- **Red de Hipervínculos:**

- **Nodos:** Representan los artículos de Wikipedia más relevantes relacionados con el fentanilo.
- **Enlaces:** Muestran las conexiones dirigidas más significativas entre estos artículos, permitiendo visualizar cómo se estructura el conocimiento sobre el fentanilo en Wikipedia.

Estas redes optimizadas me permitirán analizar de manera más efectiva las tendencias y conexiones principales en la percepción pública sobre la crisis del fentanilo, proporcionando una base sólida para responder a la pregunta de investigación planteada.



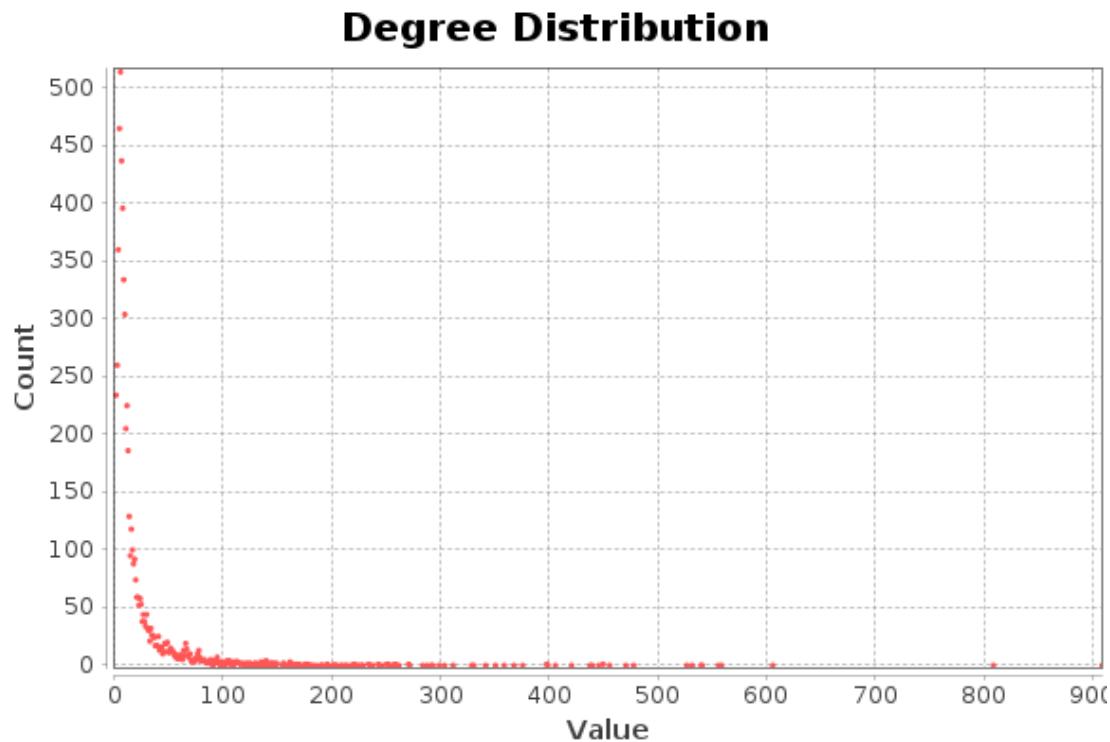
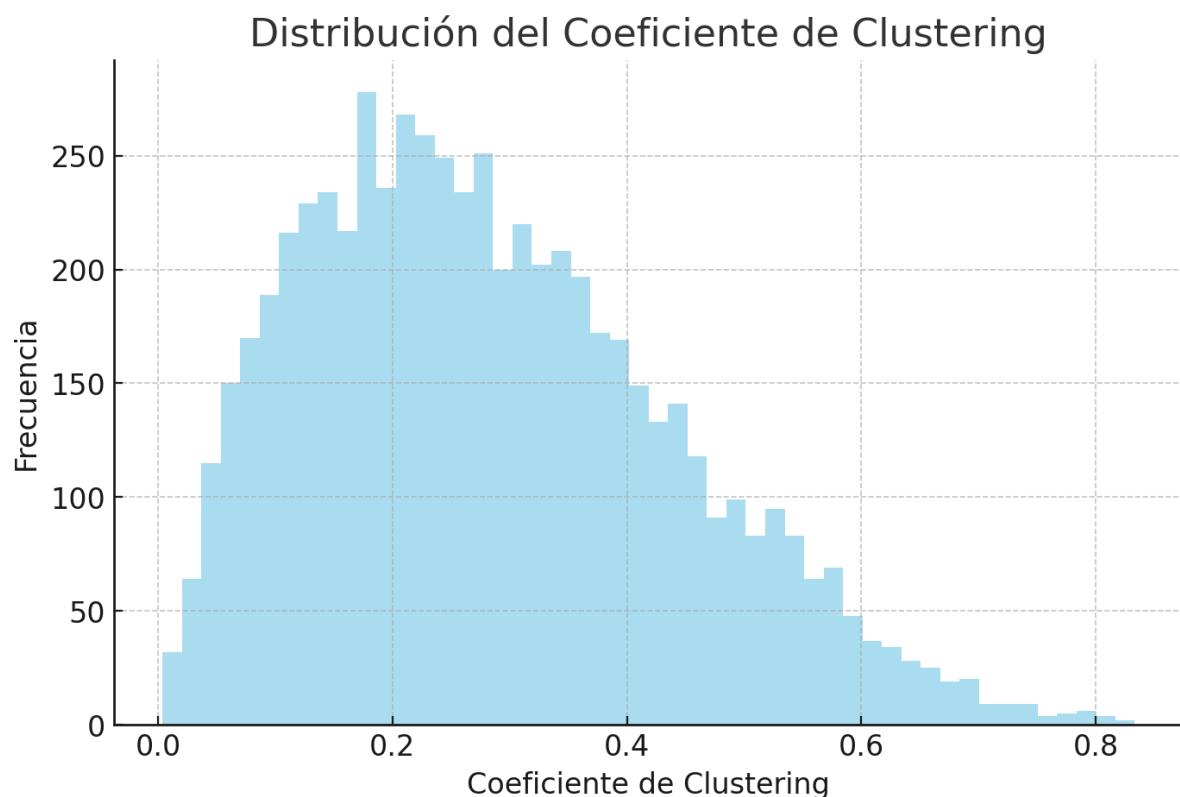
3. Métricas y análisis de la propiedades de la red.

3.1. Red de Co-ocurrencia Textual

3.1.1 Métricas

Medida	Valor
Número de enlaces	63685.0
Número de nodos	6143.0
Densidad	0.003
Grado medio	20,734
Grado medio con pesos	208,779
Diámetro	7.0
Distancia media	2,9921
Distancia media red aleatoria	2.8772
Coeficiente de clustering medio (C)	0,326
Coeficiente de clustering red aleatoria	0.0034
Número de nodos componente conexa	5908.0
Porcentaje de nodos en componente conexa	96.17 %
Número de enlaces componente conexa	63685.0
Porcentaje de aristas en componente conexa	100 %

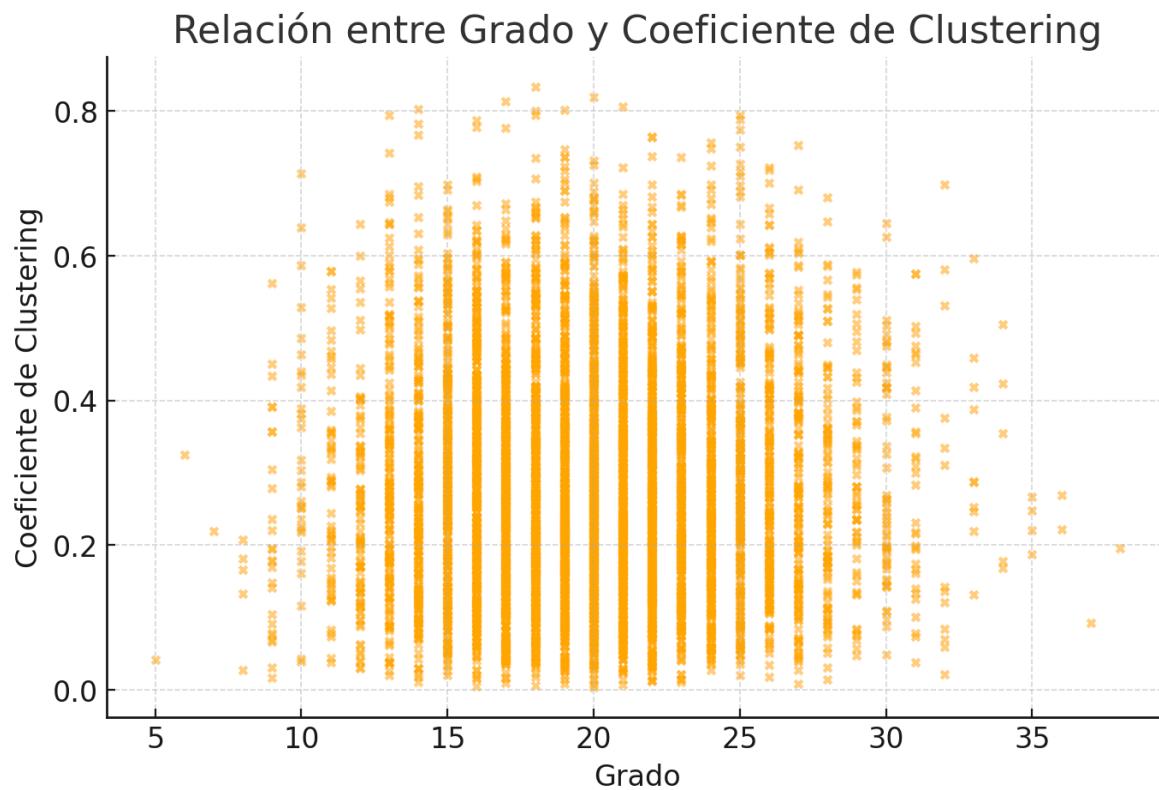


Gráfica del grado medio**Distribución del coeficiente de clustering**

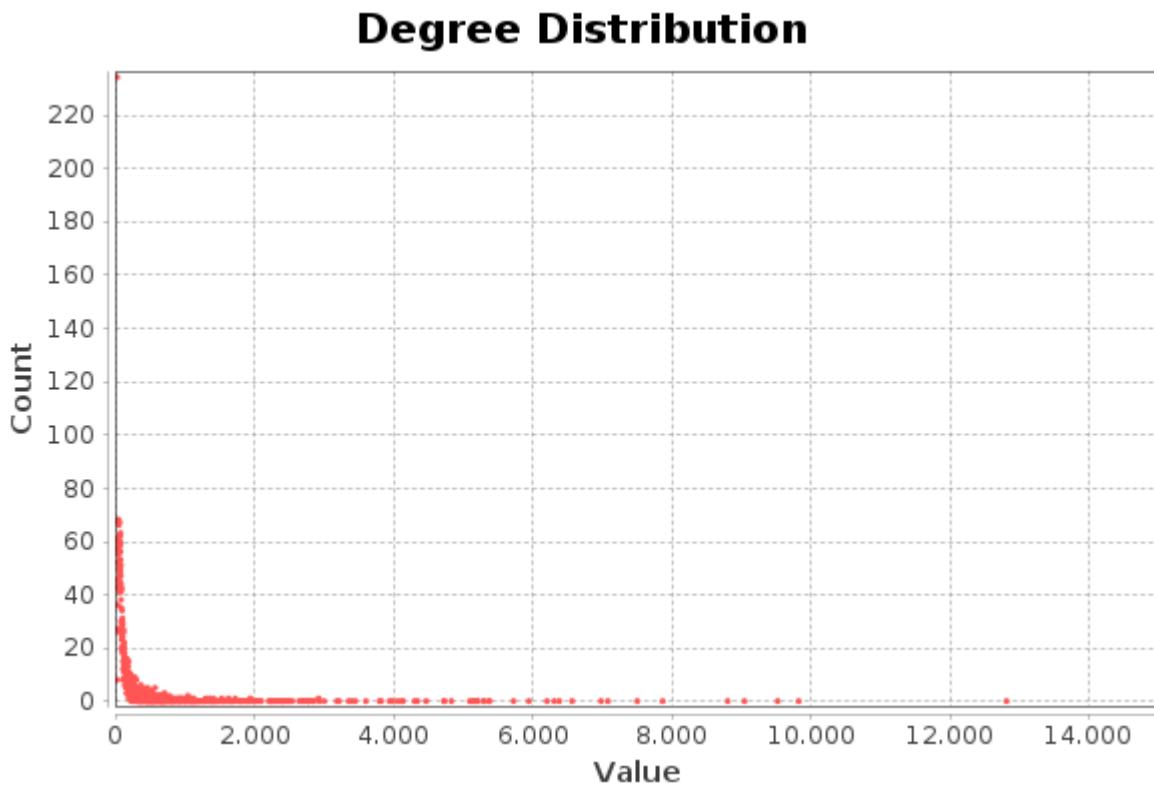
Juan Barea Rojo



Relación entre el grado y Coeficiente de Clustering



Gráfica del grado medio con pesos



3.1.2. Propiedades de la red

La red analizada muestra características que podrían sugerir que es **libre de escala**, ya que su distribución de grados tiene una cola larga, donde la mayoría de los nodos tienen pocos enlaces y unos pocos (hubs) concentran un alto número de conexiones. Este comportamiento es común en redes libres de escala, pero debe destacarse que estamos trabajando con una **red de coherencia contextual**, y no con una red social. Las redes de coherencia contextual están diseñadas para modelar relaciones conceptuales o semánticas dentro de un texto o corpus, lo que significa que los patrones de conectividad responden más a la organización y estructura del contenido que a interacciones humanas. Además, el comportamiento observado puede estar influenciado por un **sesgo implícito** en el diseño de la red, donde se eliminaron aristas y nodos que no alcanzaban un cierto umbral de peso o relevancia. Este sesgo puede haber fortalecido artificialmente los hubs conceptuales, distorsionando la distribución natural de los grados.

En términos de propiedades de **mundo pequeño**, la red cumple con los criterios esperados. La distancia media $\langle d \rangle = 2.9921$ es baja y está muy cerca de la distancia media de una red aleatoria equivalente $\langle d \rangle_{aleatoria} = 2.8772$, indicando que cualquier concepto dentro de la red puede conectarse a otro a través de pocas relaciones intermedias. Asimismo, el coeficiente de clustering medio $\langle C \rangle = 0.326$ es significativamente mayor que el de una red aleatoria $\langle C \rangle_{aleatoria} = 0.0034$, lo que refleja que los nodos tienden a agruparse en comunidades coherentes semánticamente. Sin embargo, es importante resaltar que esta característica también podría haber sido influenciada por las reglas de construcción de la red, como la eliminación de aristas con pesos bajos, que tienden a favorecer la densificación de las comunidades y la formación de triángulos conceptuales.

Por último, la red no se comporta como una red regular ni como una red aleatoria. Una red regular tendría una distribución uniforme de grados y un alto coeficiente de clustering, mientras que una red aleatoria tendría un coeficiente de clustering bajo y una distancia media similar. En cambio, esta red presenta una combinación de alto clustering y una distribución de grados con cola larga, características propias de redes libres de escala y de mundo pequeño. Sin embargo, debido a su naturaleza como red de coherencia contextual, su estructura responde más al diseño del corpus y las relaciones semánticas entre conceptos que a dinámicas sociales o naturales. Los sesgos introducidos, como la priorización de conceptos con mayor peso o relevancia, pueden haber alterado las propiedades de la red, haciendo que los hubs conceptuales reflejen jerarquías implícitas del texto en lugar de un crecimiento espontáneo. Esto refuerza la necesidad de interpretar sus propiedades en el contexto del diseño y objetivo de la red.

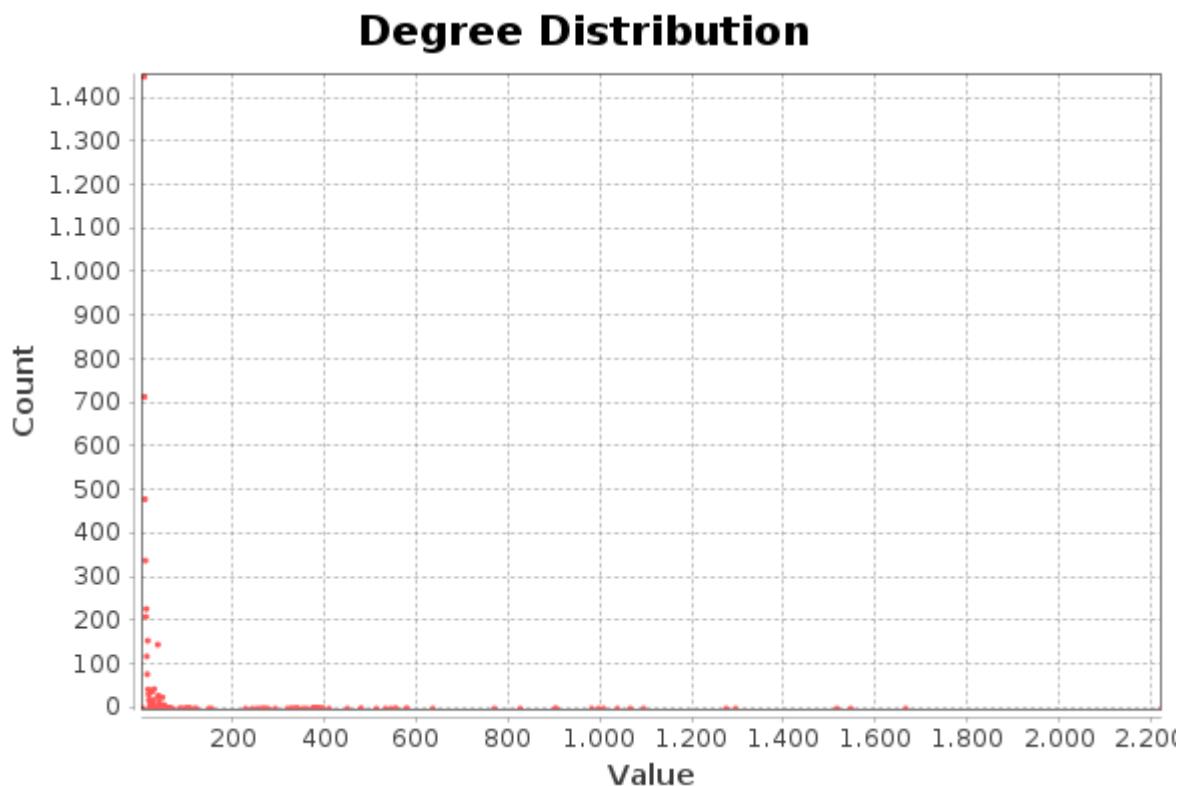
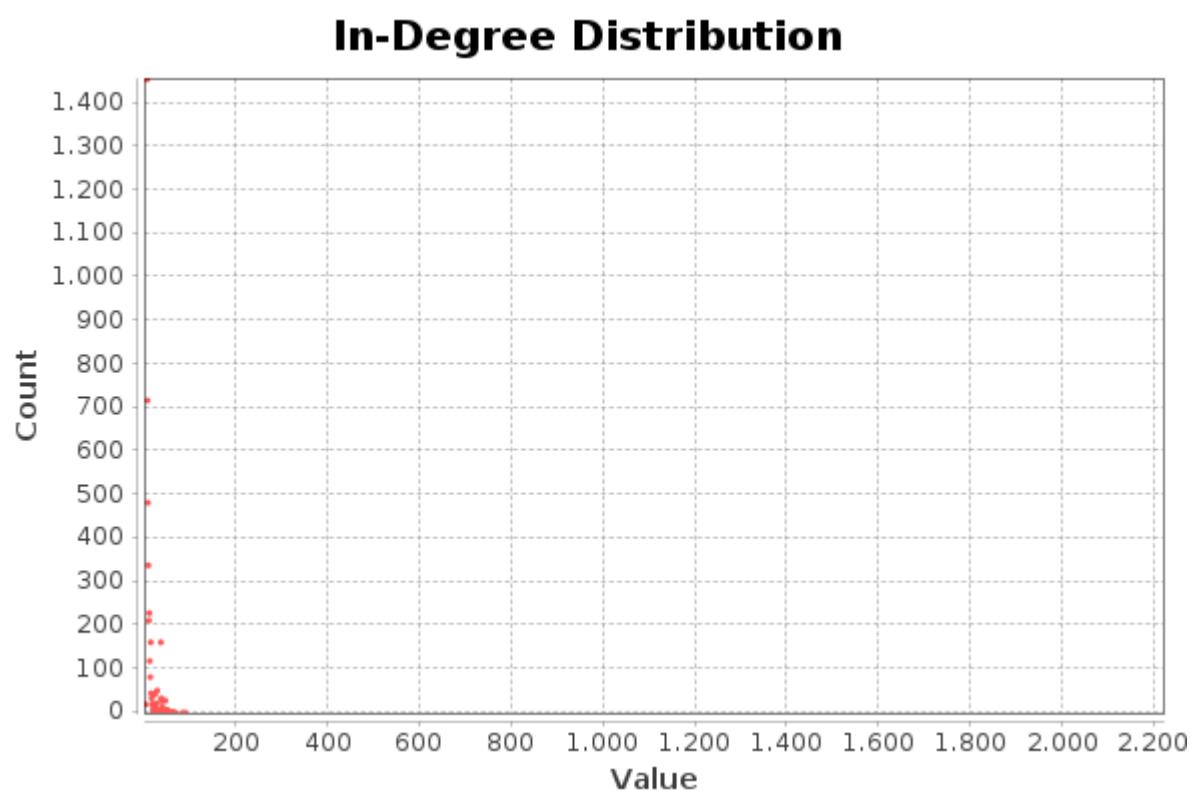


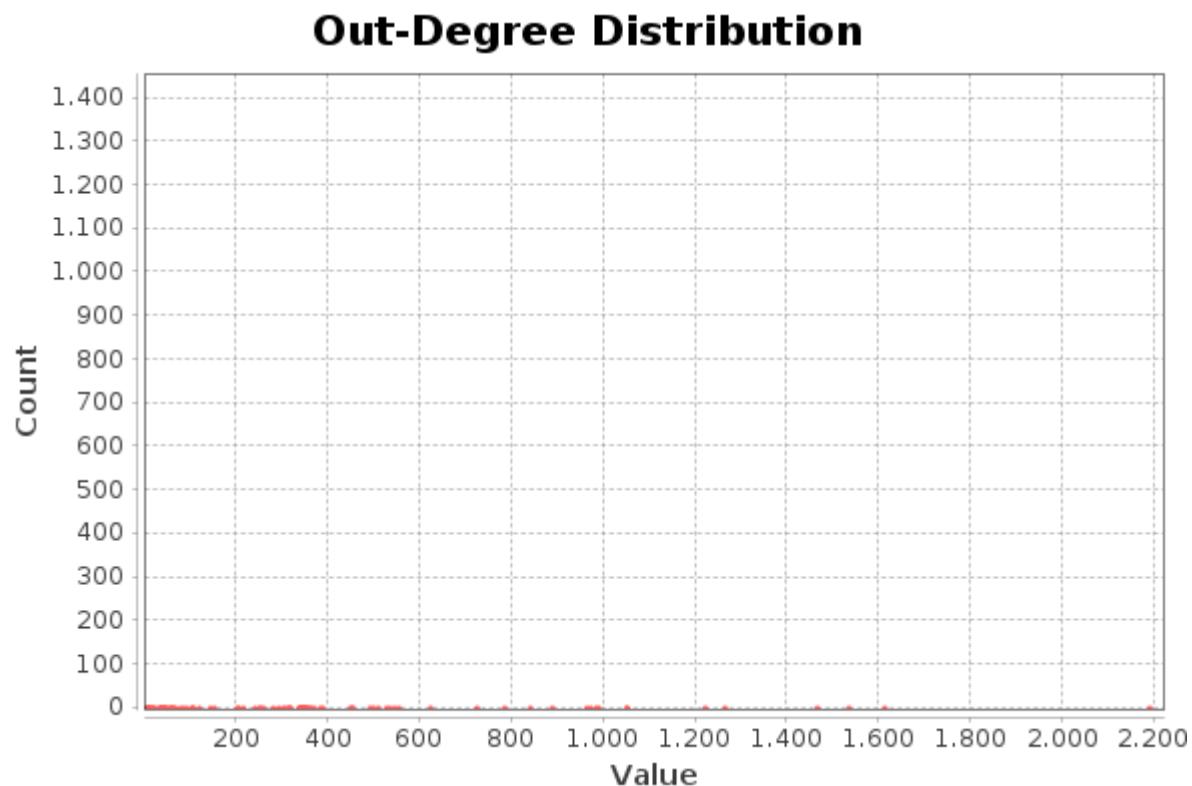
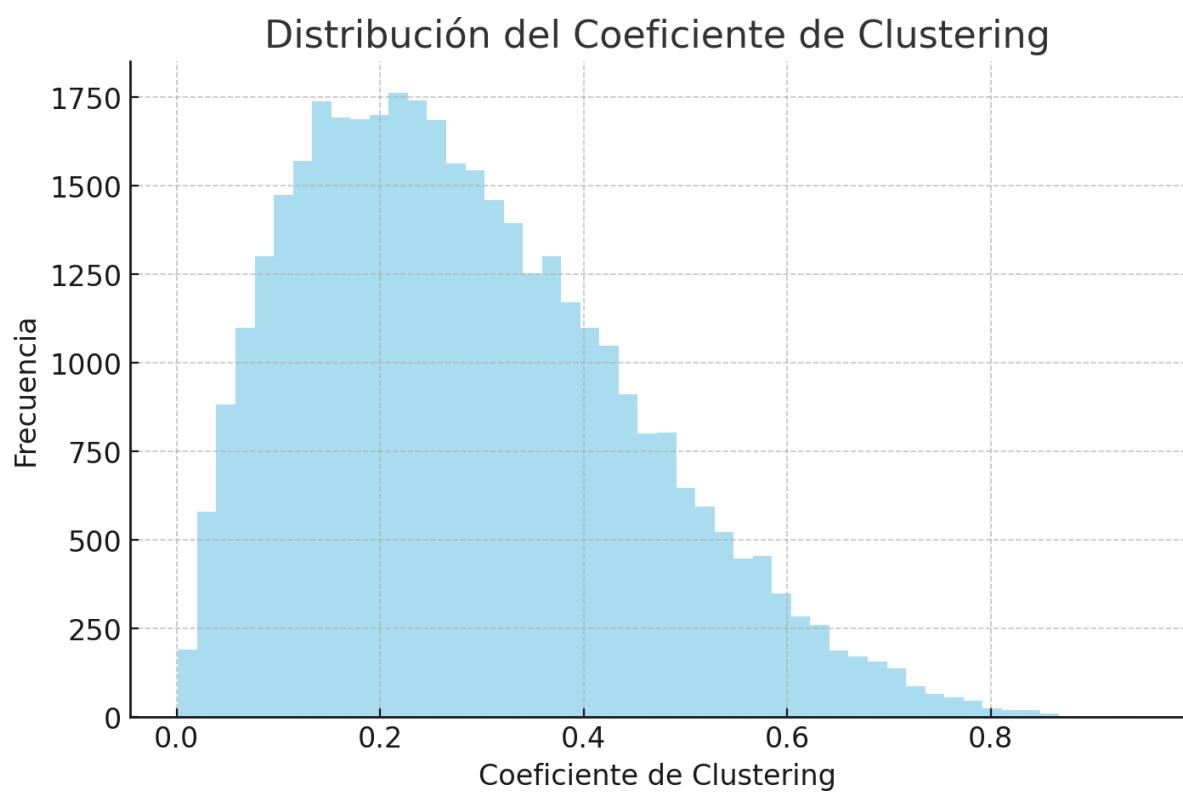
3.2. Red de Hipervínculos

3.2.1 Métricas

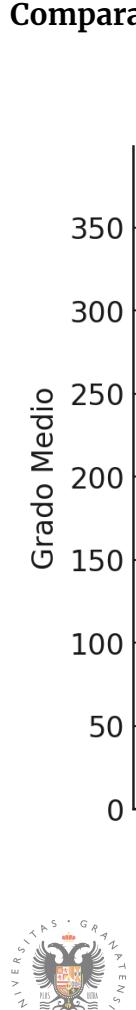
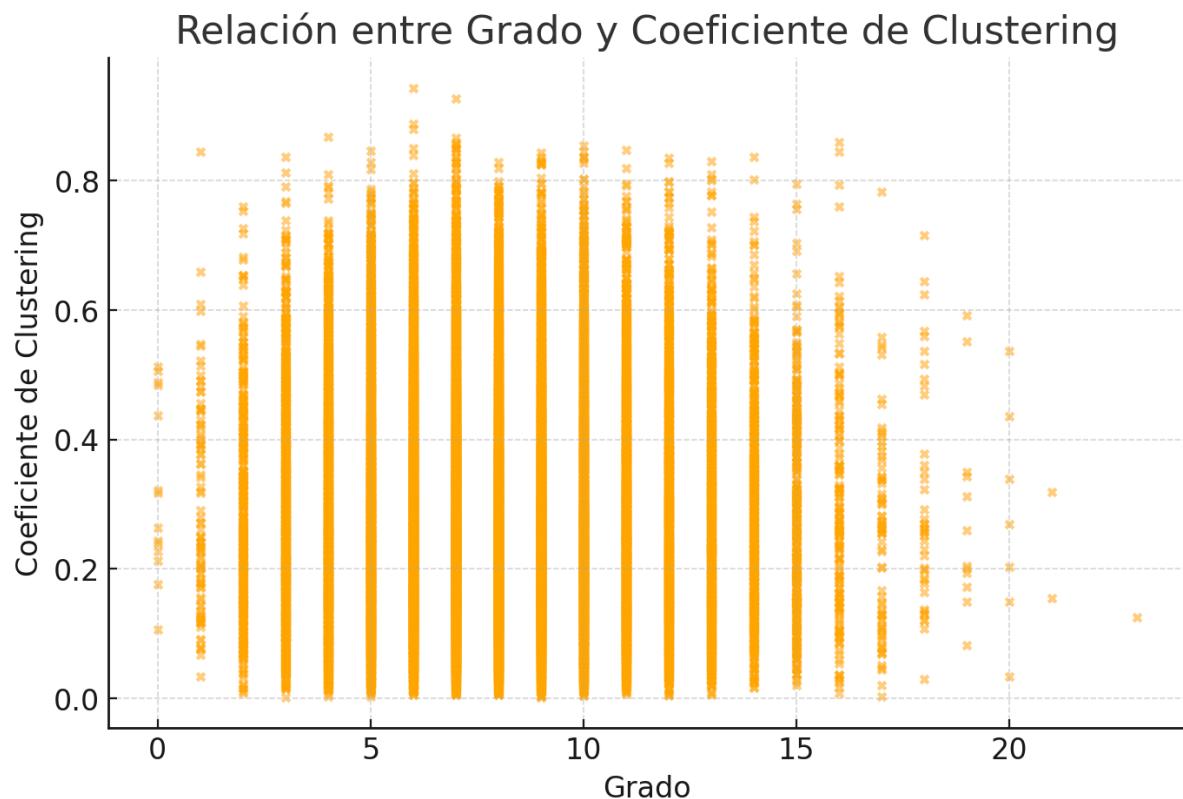
Medida	Valor
Número de enlaces	4505
Número de nodos	37997
Densidad	0,002
Grado medio de entrada	8,434
Grado medio de salida	379.97
Diámetro	5
Distancia media	2.27489
Distancia media red aleatoria	1.7752691
Coeficiente de clustering medio (C)	0.81080094184282
Coeficiente de clustering red aleatoria	0.01
Número de nodos componente conexa	4505
Porcentaje de nodos en componente conexa	100%
Número de enlaces componente conexa	37997
Porcentaje de aristas en componente conexa	100 %



Gráfica del grado medio**Gráfica del grado medio entrada**

Gráfica del grado medio de salida**Distribución del coeficiente de clustering**

Relación entre el grado y Coeficiente de Clustering



3.2.2. Propiedades de la red

La red analizada, basada en hiperenlaces derivados de un nodo base, muestra características que podrían sugerir que es libre de escala, ya que su distribución de grados probablemente presenta una cola larga, donde la mayoría de los nodos tienen pocos enlaces y unos pocos nodos (hubs) concentran un alto número de conexiones. Este comportamiento es común en redes libres de escala y es consistente con redes de hiperenlaces, donde ciertos nodos actúan como concentradores clave al recibir enlaces de manera desproporcionada (como páginas principales o portales relevantes). Sin embargo, cabe destacar que el diseño de esta red está influenciado por la extracción basada en un nodo base, lo que introduce un sesgo implícito. Por ejemplo, los algoritmos utilizados pueden priorizar ciertos nodos o eliminar enlaces que no cumplan con criterios de relevancia, lo que puede distorsionar la distribución natural de los grados al reforzar artificialmente la importancia de ciertos hubs.

En términos de propiedades de mundo pequeño, la red cumple con los criterios característicos. La distancia media $\langle d \rangle = 2.27489$ es baja y cercana a la distancia media de una red aleatoria equivalente $\langle d \rangle_{aleatoria} = 1.77526$. Esto indica que cualquier nodo dentro de la red puede conectarse a otro a través de un número reducido de pasos intermedios. Además, el coeficiente de clustering medio $\langle C \rangle = 0.81080$ es significativamente mayor que el coeficiente de clustering de una red aleatoria equivalente $C_{aleatoria} = 0.01$, lo que refleja que los nodos tienden a formar triángulos o comunidades densas de hiperenlaces. Estas características son típicas de redes donde los hiperenlaces están organizados en torno a grupos temáticos o jerárquicos, como se observa en sistemas de navegación basados en nodos centrales.

Por último, la red no puede clasificarse como regular ni como aleatoria. Una red regular tendría una distribución uniforme de grados y un alto coeficiente de clustering, mientras que una red aleatoria tendría un coeficiente de clustering bajo y una distancia media similar. En contraste, esta red presenta una combinación de alto clustering y baja distancia media, junto con una probable distribución de grados con cola larga. Estas propiedades son características de redes libres de escala y de mundo pequeño. No obstante, debido a su naturaleza como red de hiperenlaces construida a partir de un nodo base, es importante interpretar estas propiedades dentro del contexto del diseño. El sesgo introducido por la selección de enlaces desde un nodo inicial y los filtros aplicados para priorizar relevancia o peso pueden haber alterado la estructura de la red, haciendo que los hubs reflejen patrones de navegación específicos en lugar de una estructura completamente espontánea. Por ello, es fundamental considerar el propósito y las limitaciones del diseño al analizar las propiedades observadas.



4. Descubrimiento de comunidades en la red.

4.1. Comunidades de Co-ocurrencia Textual

La red de co-ocurrencias textuales fue analizada considerando su estructura modular, empleando el método de Lovaina. Este análisis permitió obtener un valor de modularidad de **0,370**, lo que indica una segmentación moderada entre los términos. Aunque las conexiones entre comunidades son relativamente frecuentes, la red presenta zonas de alta densidad interna que facilitan la identificación de agrupaciones temáticas bien definidas. Con este enfoque, se detectaron un total de **245** comunidades, siendo las 10 más relevantes destacadas por su tamaño y relevancia en el contexto textual.

En un análisis complementario, se aplicó el algoritmo de modularidad codicioso (*greedy_modularity_communities*), que mostró una visión diferente de la red. Este método logró identificar una modularidad de **0,3312**, detectando un total de **277** comunidades. A pesar de una menor granularidad en términos de modularidad, el enfoque codicioso permitió observar comunidades con diferentes tamaños, distribuidas de manera desigual en la red.

Entre las comunidades más grandes encontradas con el método codicioso destacan la **Comunidad 1**, que agrupa 2284 nodos, y la **Comunidad 2**, que incluye 2237 nodos. Otras comunidades significativas incluyen la **Comunidad 3**, con 755 nodos, y la **Comunidad 4**, que cuenta con 248 nodos. Además, muchas comunidades más pequeñas están formadas por un número reducido de nodos, algunas incluso de un solo nodo, lo que refleja una menor densidad interna en comparación con las principales agrupaciones.

Sin embargo, es importante señalar que, al tratar con coherencia textual, estos métodos, aunque útiles, aportan menos información contextual que en otras aplicaciones, como el análisis de redes sociales. Esto se debe a que el lenguaje natural no tiende a comportarse de forma idéntica a las interacciones humanas en redes sociales, donde las conexiones suelen reflejar relaciones explícitas, como amistad o afinidad temática. A pesar de ello, existen ciertas similitudes en la forma en que los términos textuales y los usuarios en redes sociales se agrupan en comunidades, lo que permite que estas técnicas sigan siendo relevantes para identificar patrones y temáticas dentro de los textos analizados.



Comunidad 107 (20.92% de la red, 1285 nodos)

Esta es la comunidad más grande, representando más del 20% de la red, con un enfoque en términos geográficos y administrativos. Nodos como **state** (907 conexiones), **chicago** (554 conexiones) y **new** (538 conexiones) actúan como hubs principales, conectando información relacionada con ubicaciones, poblaciones y territorios. Estos términos, por su relevancia, parecen estructurar gran parte de los datos geográficos en el texto.

La presencia de términos como **population** (257 conexiones) y **city** (454 conexiones) refuerza la cohesión de esta comunidad, mostrando un claro vínculo entre descripciones geográficas y datos demográficos. La densidad interna de esta comunidad asegura que la información sobre ubicaciones y entidades geográficas esté interrelacionada de manera coherente y fácilmente accesible dentro del texto.

Comunidad 1 (14.94% de la red, 918 nodos)

Con un enfoque en términos farmacológicos y de sustancias, esta comunidad incluye nodos clave como **drug** (807 conexiones), **effect** (557 conexiones) y **opioid** (525 conexiones). Estos términos conectan temas relacionados con el uso y los efectos de sustancias, posicionándolos como centros de acceso a conceptos importantes en farmacología.

Términos secundarios como **fentanyl** (419 conexiones), **heroin** (340 conexiones) y **morphine** (297 conexiones) muestran una alta relación con tratamientos y sustancias de control, indicando que esta comunidad se especializa en aspectos relacionados con opioides y su impacto médico. La conectividad interna facilita la navegación entre conceptos relacionados, haciendo de esta comunidad un recurso temático altamente enfocado.

Comunidad 4 (12.50% de la red, 768 nodos)

Esta comunidad está centrada en términos asociados con el uso y los efectos del alcohol. Nodos como **alcohol** (604 conexiones), **people** (539 conexiones) y **use** (448 conexiones) son los más relevantes, indicando un enfoque temático en los efectos sociales y de salud del consumo de alcohol.

Otros términos como **risk** (349 conexiones), **pain** (302 conexiones) y **delirium** (290 conexiones) sugieren una relación cercana entre los efectos adversos del alcohol y su impacto en la salud. La estructura de esta comunidad refleja una interrelación



entre el consumo, los riesgos asociados y las posibles consecuencias, lo que refuerza su cohesión temática.

Comunidad 83 (10.48% de la red, 644 nodos)

Esta comunidad aborda temas relacionados con conflictos, gobiernos y derechos humanos. Nodos como **death** (530 conexiones), **taliban** (476 conexiones) y **country** (438 conexiones) destacan por su conectividad, señalando un enfoque en cuestiones políticas y de conflicto.

La aparición de términos como **government** (270 conexiones), **human** (236 conexiones) y **execution** (200 conexiones) sugiere un vínculo temático entre conflictos sociales, gobiernos y derechos humanos. Esta comunidad muestra una estructura coherente, interrelacionando términos que son frecuentemente utilizados en contextos de políticas y conflictos globales.

Comunidad 91 (8.94% de la red, 549 nodos)

Con un enfoque en términos temporales y de entretenimiento, esta comunidad incluye nodos como **time** (469 conexiones), **wrestling** (329 conexiones) y **match** (222 conexiones). Estos términos indican que la comunidad está centrada en actividades deportivas o de entretenimiento, con un enfoque especial en el deporte de lucha libre profesional.

Términos adicionales como **team** (123 conexiones) y **professional** (117 conexiones) refuerzan esta idea, mostrando una relación entre conceptos relacionados con competiciones y organizaciones deportivas. La densidad de enlaces internos refleja la cohesión de esta comunidad, que estructura de manera efectiva la información sobre deportes y entretenimiento.

Comunidad 98 (8.43% de la red, 518 nodos)

Esta comunidad se centra en términos médicos y farmacológicos, con nodos principales como **dose** (310 conexiones), **epidural** (269 conexiones) y **analgesia** (158 conexiones). Estos términos destacan por su relación con la administración de medicamentos y procedimientos médicos específicos.

Otros términos como **administration** (232 conexiones) y **brain** (208 conexiones) muestran que la comunidad aborda tanto aspectos fisiológicos como prácticas



médicas, reflejando un enfoque bien definido en procedimientos terapéuticos y analgésicos.

Comunidad 5 (7.36% de la red, 452 nodos)

Con un enfoque en términos técnicos y científicos, esta comunidad incluye nodos como **system** (256 conexiones), **chemical** (146 conexiones) y **acid** (144 conexiones). Estos términos conectan conceptos básicos en química y sistemas biológicos.

Términos como **different** (119 conexiones) y **example** (138 conexiones) indican que esta comunidad también podría incluir descripciones generales o aplicaciones de conceptos científicos, reforzando su valor como un nodo de conocimiento técnico.

Comunidad 79 (7.11% de la red, 437 nodos)

Esta comunidad está orientada a términos relacionados con salud pública y organizaciones internacionales. Nodos como **health** (448 conexiones), **world** (444 conexiones) y **organization** (160 conexiones) actúan como hubs principales, destacando un enfoque en temas de bienestar y cooperación internacional.

Otros nodos como **public** (199 conexiones) y **care** (154 conexiones) refuerzan la estructura de la comunidad, que conecta conceptos relacionados con políticas de salud y servicios públicos.

Comunidad 103 (4.00% de la red, 246 nodos)

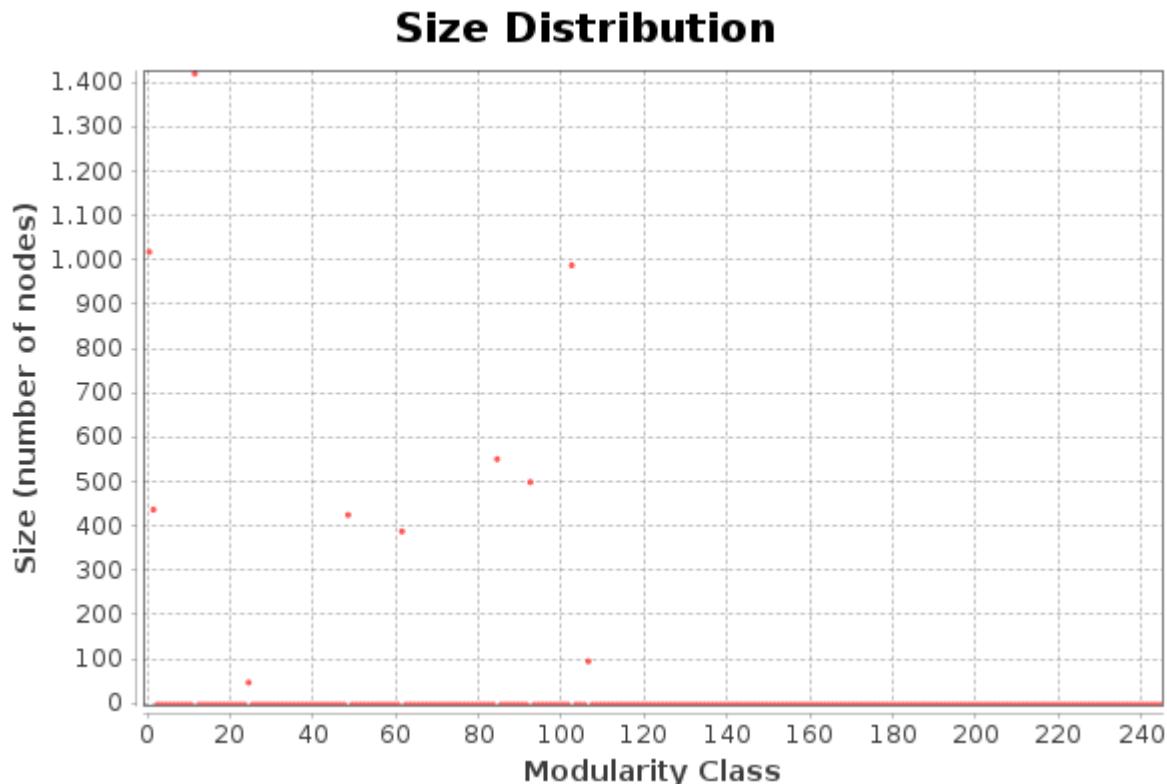
Con términos como **treatment** (374 conexiones), **control** (255 conexiones) y **mifepristone** (226 conexiones), esta comunidad está claramente enfocada en tratamientos médicos y terapias. La presencia de términos como **woman** (216 conexiones) y **combination** (133 conexiones) sugiere que el texto aborda cuestiones de salud reproductiva y tratamientos combinados.

Conclusión

El análisis de la red basada en co-ocurrencias textuales ha permitido identificar 10 comunidades principales con estructuras bien definidas. Nodos altamente



conectados, como **state** en la Comunidad 107 o **drug** en la Comunidad 1, actúan como hubs centrales, facilitando la cohesión interna y la navegación temática. Estas comunidades reflejan la diversidad temática del texto analizado, abarcando desde términos geográficos y administrativos hasta conceptos médicos y científicos. Este diseño permite explorar el texto de manera eficiente, conectando información general y específica a través de una red estructurada y coherente.



4.2. Comunidades de Hipervínculos

La red de hiperenlaces en Wikipedia fue analizada utilizando el método de Lovaina para calcular la modularidad, lo que permitió identificar un valor de **0,485**. Este valor refleja una estructura modular definida, donde las conexiones dentro de las comunidades son significativamente más densas que entre comunidades. Con un total de **7 comunidades**, la red se segmenta de manera temática, agrupando nodos interrelacionados que actúan como hubs o puntos de conexión en temas específicos.

Por otro lado, aplicando el algoritmo de modularidad codiciosa (*greedy_modularity_communities*), se obtuvieron resultados ligeramente diferentes en cuanto a la estructura de la red. Este método logró identificar una mejor modularidad de **0,4284**, con un total de **4 comunidades** detectadas.

En los resultados obtenidos con el algoritmo codicioso, las comunidades se distribuyen de la siguiente manera: la **Comunidad 1** contiene 2216 nodos, la **Comunidad 2** tiene 1273 nodos, la **Comunidad 3** agrupa 726 nodos y la **Comunidad 4** incluye 290 nodos. Estas comunidades también reflejan agrupaciones temáticas, aunque con menor granularidad que las identificadas con el método de Lovaina.

Comunidad 0 (22.44% de la red, 1011 nodos)

La comunidad 0, que constituye más del 22% de la red, está centrada en sustancias químicas y farmacológicas. Nodos como **Arylcyclohexylamine** y **Opioid**, con un alto número de enlaces salientes (1542 y 1091, respectivamente), actúan como hubs dentro de esta comunidad. Estos nodos tienen un papel clave en la distribución de información, conectando páginas relacionadas con drogas, analgésicos y sustancias químicas.

Durante el scraping, se identificaron enlaces hacia nodos como **Desomorphine** y **Methamphetamine**, que, con más de 700 conexiones cada uno, facilitan el acceso a información interrelacionada. Estos nodos también actúan como puentes hacia otras comunidades, ya que conectan conceptos de abuso de sustancias con aplicaciones médicas.

Comunidad 1 (7.44% de la red, 335 nodos)

Esta comunidad, aunque más pequeña, está altamente especializada en opioides sintéticos y sus derivados. Nodos clave como **Fentanyl** y **Dihydroetorphine** poseen



enlaces que los posicionan como centros de acceso a información sobre regulación de opioides y tratamientos analgésicos. **Fentanyl**, en particular, se conecta directamente con nodos sobre legislación y uso clínico, lo que refuerza su rol como un hub temático.

Los nodos secundarios, como **Co-dydramol** y **Paregoric**, tienen un número moderado de enlaces que los convierten en puntos de referencia para explorar opioides menos conocidos o sus aplicaciones específicas. Esta estructura interna permite que la comunidad funcione como una red de recursos especializados dentro del contexto de farmacología analgésica.

Comunidad 2 (8.68% de la red, 391 nodos)

La comunidad 2 se centra en medicamentos antiinflamatorios y analgésicos comunes, con nodos principales como **Hyperforin** e **Ibuprofen/paracetamol**. Estos nodos, con más de 500 enlaces, facilitan la navegación hacia combinaciones farmacológicas ampliamente utilizadas, posicionándose como hubs centrales dentro de la comunidad.

Enlaces hacia nodos como **Acemetacin** y **Nimesulide** reflejan la estructura de la comunidad, donde los temas están interrelacionados a través de medicamentos específicos. Además, estos nodos funcionan como intermediarios que conectan la comunidad con otras relacionadas con tratamientos alternativos y terapias no opioides.

Comunidad 3 (29.97% de la red, 1350 nodos)

La comunidad más grande, que agrupa casi el 30% de la red, integra nodos de gran relevancia temática como **Alcohol (drug)** y **Methadone**, con miles de enlaces cada uno. Estos nodos actúan como hubs principales dentro de la comunidad, conectando conceptos generales sobre sustancias recreativas y medicamentos ampliamente conocidos.

Otros nodos, como **Dextropropoxophene** y **Sedative-hypnotic**, tienen un número considerable de enlaces que los posicionan como puntos secundarios de conexión. Esta estructura permite que la comunidad funcione como un puente entre términos generales y temáticas específicas, facilitando la navegación dentro de la red.

Comunidad 4 (15.34% de la red, 691 nodos)



La comunidad 4 está orientada hacia hormonas y farmacología biológica, con nodos destacados como **Cyproterone acetate** y **Mifepristone**. Estos nodos, con más de 900 enlaces cada uno, son puntos críticos para explorar temas médicos especializados en salud reproductiva y terapias hormonales.

La conectividad de nodos secundarios como **17 α -Hydroxyprogesterone** y **WHO Model List of Essential Medicines** refuerza la cohesión de esta comunidad, donde los enlaces actúan como canales hacia información esencial en farmacología y tratamientos médicos.

Comunidad 5 (5.59% de la red, 252 nodos)

Siendo la comunidad más pequeña, está centrada en conceptos básicos de química. Nodos como **Formaldehyde** y **Ethylene** destacan por su conectividad moderada, lo que los convierte en puntos clave para explorar términos introductorios en química orgánica e inorgánica.

Enlaces hacia nodos como **Chemical formula** demuestran cómo esta comunidad proporciona una base de conocimiento fundamental, actuando como soporte para conceptos más complejos explorados en otras comunidades.

Comunidad 6 (10.54% de la red, 475 nodos)

La comunidad 6 aborda una mezcla temática de conceptos médicos y regionales, con nodos principales como **Virginia (state)** y **Epidural anaesthesia**. Estos nodos tienen enlaces que conectan temas administrativos y clínicos, posicionándose como puntos de convergencia para explorar aspectos legales, médicos y geográficos.

Enlaces hacia nodos como **New Jersey** y **Epidural** refuerzan la diversidad temática de la comunidad, permitiendo que funcione como un puente entre diferentes contextos.

El análisis de la red utilizando el método de Lovaina ha permitido identificar 7 comunidades temáticas bien definidas, donde los enlaces internos actúan como los principales determinantes de la estructura modular. Nodos altamente conectados, como **Alcohol (drug)** en la Comunidad 3 o **Fentanyl** en la Comunidad 1, funcionan como hubs que facilitan la navegación temática dentro de sus respectivas

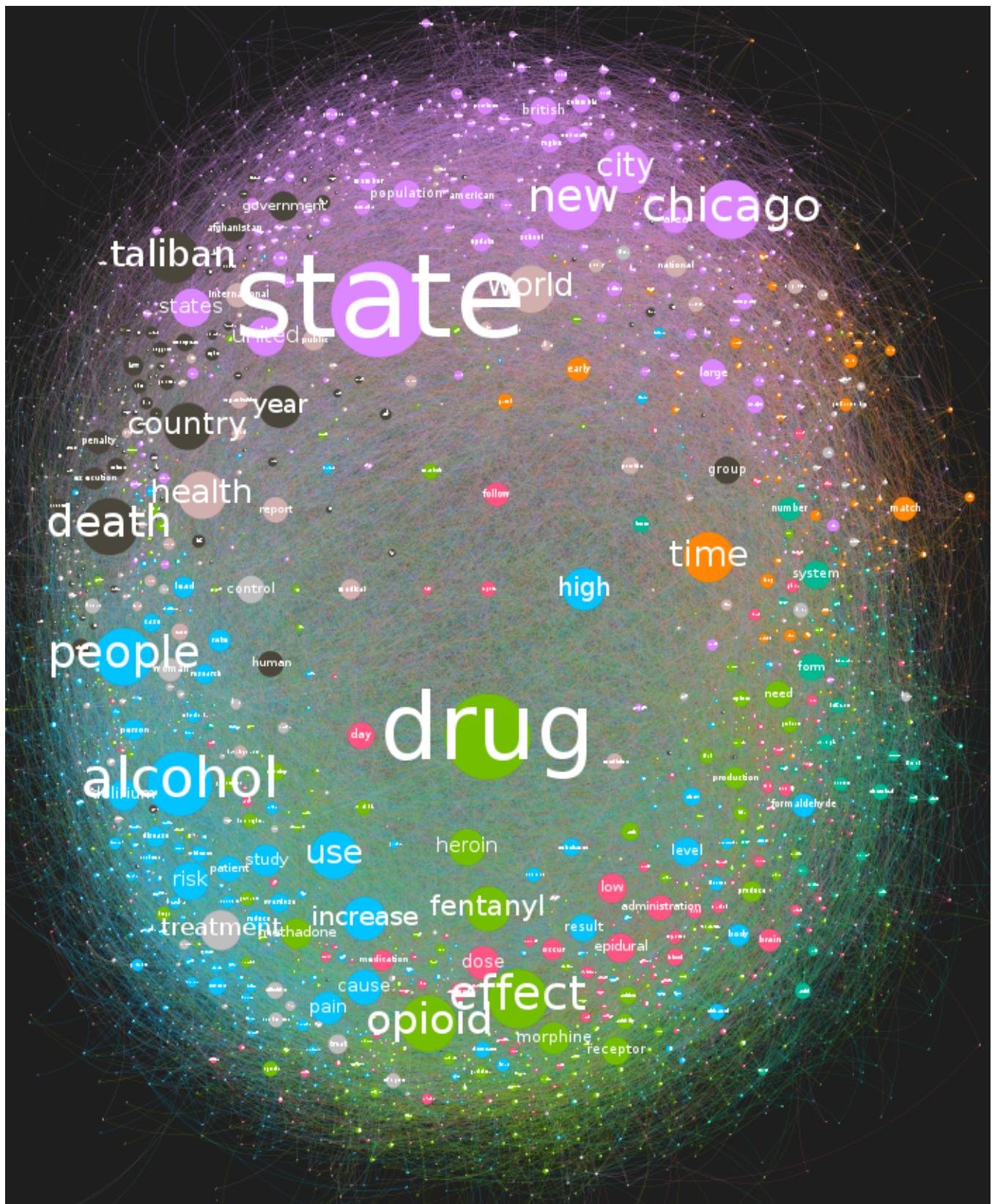


comunidades y hacia otras. Estas comunidades reflejan una organización temática eficaz, que puede servir como base para optimizar la estructura de Wikipedia y mejorar la experiencia de los usuarios al buscar información interrelacionada.

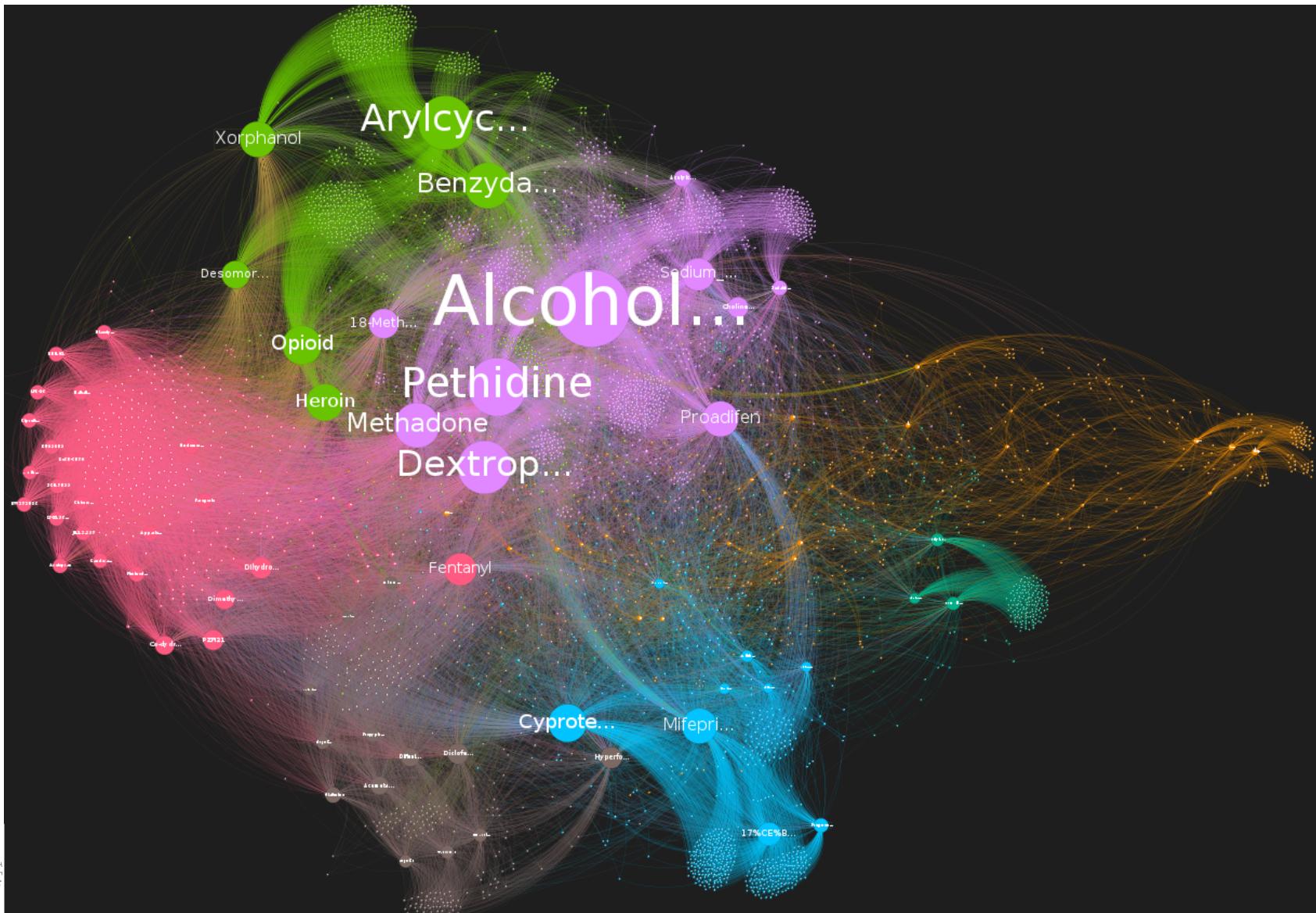
Size Distribution



5. Red de Co-ocurrencia Textual Visualización



6. Red de Hipervínculos Visualización



7. Discusión de los resultados obtenidos.

El análisis de las redes construidas para estudiar la percepción pública de la crisis del fentanilo en Wikipedia, tanto desde la perspectiva de co-ocurrencia textual como desde el análisis de hipervínculos, pone de manifiesto una estructura compleja, altamente interconectada, que refleja cómo los conceptos se relacionan y estructuran en esta plataforma. Estas redes, combinadas, permiten una comprensión más profunda de las conexiones estructurales, la relevancia temática y las implicaciones en el conocimiento público sobre esta crisis.

Desde el análisis de co-ocurrencia textual, la red muestra comunidades temáticas bien definidas, agrupadas en torno a términos clave como "drug," "opioid," y "effect." Estos nodos forman un núcleo farmacológico que conecta aspectos médicos, sociales y regulatorios del fentanilo. Simultáneamente, términos como "state" y "chicago" revelan un enfoque territorial, lo que sugiere que las narrativas sobre el impacto de la crisis están profundamente vinculadas a contextos geográficos específicos. Estas comunidades, aunque bien conectadas internamente, mantienen puentes con otros grupos temáticos, como el de la salud pública y las políticas regulatorias, representados por términos como "health" y "organization."

Por otro lado, el grafo de hiperenlaces en Wikipedia refuerza esta estructura al conectar nodos clave como "Fentanyl," "Alcohol (drug)," y "World Health Organization." Los hiperenlaces permiten transitar rápidamente entre artículos, facilitando la exploración de conceptos relacionados. Visualmente, la red de hiperenlaces presenta zonas densas que representan grupos temáticos con múltiples conexiones internas, mientras que las zonas más dispersas reflejan conceptos periféricos o menos relacionados con el núcleo de la red. Este diseño es inherente a una plataforma que busca facilitar la navegación y el aprendizaje, asegurando que los usuarios puedan explorar tanto temas generales como específicos.

El comportamiento de los enlaces en la red de hipervínculos es notablemente diverso. Nodos altamente conectados, o hubs, consolidan gran parte de los enlaces y funcionan como puntos de acceso clave para la exploración de conceptos. Por ejemplo, "World Health Organization" actúa como un puente temático que conecta políticas internacionales de salud con las discusiones sobre el fentanilo. Simultáneamente, nodos más especializados, como "Anti-fentanyl legislation in the United States," permiten un análisis más profundo de los aspectos regulatorios. Este balance entre enlaces fuertes y débiles asegura una estructura escalable y eficiente, donde los usuarios pueden moverse entre distintos niveles de detalle sin dificultad.



La jerarquía descentralizada de la red de hipervínculos complementa la organización modular de la red de co-ocurrencia textual. Si bien algunos nodos concentran la mayoría de las conexiones, no monopolizan por completo la red, permitiendo que cualquier nodo tenga una oportunidad razonable de ser alcanzado desde otros puntos. Este diseño distribuye la carga de navegación, evita cuellos de botella y asegura que la red sea útil tanto para usuarios que buscan información específica como para aquellos que desean explorar una amplia gama de temas relacionados. Además, fomenta la curiosidad y la interactividad, características fundamentales de un recurso como Wikipedia.

En conjunto, las redes de co-ocurrencia textual y de hipervínculos reflejan cómo los discursos relacionados con el fentanilo están estructurados en torno a múltiples dimensiones temáticas. Por un lado, la red textual resalta las conexiones semánticas entre términos clave, mientras que la red de hipervínculos organiza el conocimiento de manera que facilite su navegación. La representación combinada permite observar cómo se interrelacionan los aspectos farmacológicos, sociales y regulatorios de la crisis, proporcionando una visión multidimensional del problema. El artículo base sobre Fentanyl aporta información relevante y sorprendente que amplía la comprensión de esta crisis. Por ejemplo, destaca que el fentanilo es **30 a 50 veces más potente que la heroína y 100 veces más potente que la morfina**, lo que lo posiciona como un analgésico crucial en contextos médicos, pero también como un peligroso opioide sintético en el mercado ilícito. Desde su síntesis en 1959 por Paul Janssen, el fentanilo se ha utilizado en anestesia, manejo del dolor crónico, cuidados paliativos e incluso en entornos militares, como lollipops de fentanilo para tratar heridos en combate. Sin embargo, su uso ilícito y la adulteración con otras drogas, como cocaína y heroína, han llevado a una epidemia de sobredosis, con más de **71,000 muertes relacionadas en 2021 solo en los Estados Unidos**.

El proceso de scraping realizado fue crucial para capturar la riqueza de esta estructura. A partir del nodo inicial "Fentanyl," que concentró 1030 enlaces y 5629 palabras, se exploraron temas conexos como políticas internacionales ("World Health Organization") y regulaciones locales ("Anti-fentanyl legislation in the United States"). Asimismo, se incluyeron nodos relacionados con tratamientos y alternativas terapéuticas, como "Methadone" y "Cyproterone acetate," ampliando el alcance del análisis hacia aspectos médicos y sociales. En conclusión, la combinación de las redes de co-ocurrencia textual y de hipervínculos proporciona un análisis integral del conocimiento público sobre la crisis del fentanilo. Estas redes no solo reflejan cómo se organiza y distribuye la información en Wikipedia, sino que también destacan la capacidad de esta plataforma para estructurar el conocimiento de manera eficiente y accesible. Al conectar términos con niveles de detalle variados, estas redes ofrecen una herramienta valiosa tanto para la educación como para la concientización pública, subrayando la importancia de comprender esta crisis desde múltiples perspectivas.



8. Conclusión y consideraciones finales

En este estudio, se ha respondido a la pregunta de investigación planteada mediante el análisis de 100 nodos relevantes del artículo de Wikipedia sobre fentanilo y sus discusiones asociadas. Todo el código desarrollado para realizar este análisis, así como los datos estructurados obtenidos, han sido liberados bajo una política de uso libre para fines de investigación, siempre y cuando se mencione al autor. El repositorio público donde se puede acceder a estos recursos es https://github.com/juanbarearojo/wiki_scrapper_fentanilo

El análisis presentado aquí aborda las complejidades inherentes al trabajo con texto, especialmente en un entorno como Wikipedia, donde la naturaleza colaborativa y abierta puede llevar a información inconsistente o incompleta. Una de las principales dificultades fue procesar el texto utilizando bigramas para identificar relaciones semánticas clave, lo cual requirió un balance entre granularidad y relevancia. Además, se implementaron técnicas como el reconocimiento de entidades nombradas (NER) para identificar términos críticos, la eliminación de *stopwords* para reducir el ruido, y el análisis de palabras raíz y derivadas para determinar su significado en contexto. Sin embargo, debido a la limitación a 100 nodos, es posible que parte de la información relevante haya quedado fuera del alcance del análisis.

Por ejemplo, aunque el término **Talibán** apareció destacado en las redes de co-ocurrencia y es ampliamente mencionado en el artículo en relación con el cultivo de opio y el impacto geopolítico de la crisis de opioides, otros temas igualmente importantes, como la **epidemia de fentanilo en Estonia**, no fueron capturados en los nodos seleccionados. Estonia es mencionada en el artículo como un caso significativo, ya que ha sido el hogar de la epidemia de fentanilo documentada más larga del mundo, especialmente tras la prohibición de cultivo de amapolas impuesta por el Talibán en Afganistán. Este detalle subraya la interacción global de los factores sociales y económicos detrás de la crisis, pero quedó fuera del análisis al no haber sido priorizado en la selección de nodos, lo que representa una oportunidad para investigaciones futuras que exploren de manera más amplia las conexiones internacionales.

En conclusión, este trabajo ha permitido identificar tendencias y dinámicas sobre la percepción pública de la crisis del fentanilo, capturando cómo se ha desarrollado el entendimiento colectivo de este grave problema de salud pública en una plataforma tan influyente como Wikipedia. A pesar de las limitaciones mencionadas, los métodos utilizados y los resultados obtenidos proporcionan un marco inicial que puede servir de base para investigaciones adicionales en el ámbito de la percepción pública y el análisis de contenido digital.

