

Universidad Nacional de Educación a Distancia (UNED)

E.T.S. de Ingeniería Informática — Máster Universitario en Tecnologías del
Lenguaje

Minería de Textos

Reconocimiento de Entidades Nombradas

Tarea 3 Obligatoria

Autor: Franly Iris Urbina Franco

DNI: 71.997.418-N

Correo: furbina7@alumno.uned.es

Fecha: 13 de febrero de 2026

Índice

Introducción

A continuación, se presenta un análisis basado en `Brown_10-2.graphML`, un fichero que contiene una red léxica de coaparición construida a partir de un corpus textual.

La estructura interna de esta red se organiza en comunidades detectadas mediante modularidad, donde cada nodo corresponde a un término y las palabras tienden a agruparse en campos semánticos diferenciados. Cada arista indica que dos términos aparecen juntos en un mismo contexto y posee un peso asociado que mide la frecuencia de esa relación.

A partir de esta estructura se generan cincuenta tablas, dos por cada comunidad, con los diez nodos más relevantes según dos criterios: el grado, que indica cuántas conexiones distintas tiene un término, y el grado con pesos, que refleja la intensidad total de sus relaciones.

Cada clase modular representa una comunidad de palabras que conforman un vocabulario interno del corpus, lo que permite identificar grupos temáticos, términos que actúan como centros estructurales y palabras que funcionan como enlaces dentro de cada dominio semántico. Ordenar por grado permite observar cuántas palabras diferentes coaparecen con un término al menos una vez, ya que el grado representa el número de nodos distintos con los que está conectado, identificando así palabras presentes en muchos contextos y nodos estructuralmente centrales con alta diversidad de relaciones. Por otro lado, el grado ponderado es la suma de los pesos de todas sus aristas, es decir, cuántas veces en total coaparece ese término con otros, lo que permite detectar qué palabras funcionan como núcleos temáticos más intensos. De este modo se responde a dos cuestiones: cuáles son los nodos con más relaciones distintas y cuáles presentan relaciones más fuertes dentro de cada comunidad.

Condición	Interpretación
Alto grado + alto peso	Conceptos centrales del tema
Alto grado + bajo peso	Términos puente o generales
Bajo grado + alto peso	Términos muy específicos
Bajo grado + bajo peso	Nodos periféricos

Cuadro 1: Interpretación estructural de los nodos según grado y grado ponderado

Los nodos de las tablas representan el significado global de cada comunidad, por lo que el nombre asignado reflejará su campo semántico mediante la elección de un sustantivo representativo de las etiquetas de los nodos, con el fin de caracterizar los núcleos modulares por significado dentro del corpus.

Para obtener las comunidades se siguieron las instrucciones: primero se filtró el componente gigante para eliminar subgrafos irrelevantes; después se calcularon métricas estructurales y se aplicó el algoritmo de modularidad para detectar comunidades y el algoritmo de distribución *ForceAtlas2* para la disposición visual del grafo. Esto permite interpretar las comunidades como agrupaciones semánticas de términos relacionados a partir del análisis de redes léxicas orientado a extraer conocimiento estructural del corpus.

Comunidad 0 - Election

En la Comunidad 0 se observa una estructura jerárquica definida donde los términos *vote* y *election* actúan como centro semántico de la red. Estos nodos mantienen las dos primeras posiciones en ambas métricas, lo que demuestra que no solo poseen una alta conectividad estructural (grado), sino que sus relaciones son las más intensas en términos de coocurrencia recurrente (grado con pesos).

Id	Label	Grado	Grado con pesos
267	<i>vote</i>	51	74.1021
4	<i>election</i>	41	61.9433
135	<i>taxpayer</i>	38	30.4639
100	<i>federal</i>	35	31.6903
64	<i>government</i>	33	39.0482
692	<i>method</i>	32	18.0729
352	<i>tax</i>	30	34.9038
33	<i>voter</i>	29	39.3848
266	<i>force</i>	28	33.8163
22	<i>report</i>	26	30.0436

Top 10 por Grado

Id	Label	Grado	Grado con pesos
267	<i>vote</i>	51	74.1021
4	<i>election</i>	41	61.9433
33	<i>voter</i>	29	39.3848
64	<i>government</i>	33	39.0482
352	<i>tax</i>	30	34.9038
242	<i>candidate</i>	18	34.3173
266	<i>force</i>	28	33.8163
66	<i>propose</i>	26	32.3521
100	<i>federal</i>	35	31.6903
116	<i>state</i>	22	31.1115

Top 10 por Grado con pesos

El núcleo central es estable, ya que ambas tablas comparten el 70 % de palabras en común entre clasificaciones, tanto cuando se mide la diversidad de conexiones como cuando se evalúa su intensidad acumulada.

Sin embargo, el cambio más significativo ocurre con el término *voter*, que en la clasificación de grado con peso se observa en agrupación junto a *government*. Esto indica un posible discurso electoral vinculado a cuestiones institucionales y fiscales por su similitud en pesos con los términos *tax*, *candidate*, *force*, *propose*, *federal* y *state*.

Términos como *taxpayer* o *method* mantienen vínculos fuertes y frecuentes dada su posición en el ranking y correlación con términos de alto grado en peso.

Si bien todas las palabras son léxicamente distintas, se pueden formar grupos de palabras que definen un proceso electoral y, además, connotan un contexto político posiblemente enfocado en propuestas electorales. Es por ello que la palabra que más define el conjunto es *election*, el cual es un término más representativo para esta comunidad y sintetiza el significado común de los nodos centrales.

Comunidad 1 - Prision

En la Comunidad 1 se observa una estructura relativamente compacta donde los términos *term*, *prison*, *permit* y *approve* ocupan posiciones destacadas en ambas métricas, lo que indica que actúan como núcleos estructurales y semánticos dentro de la red. Estos nodos presentan simultáneamente una diversidad notable de conexiones y una intensidad que, en términos de valores, presenta una diferencia numérica mayor. Esto sugiere que todos forman parte del vocabulario central de esta comunidad, otorgándole a *prison* la mayor carga descriptiva.

Id	Label	Grado	Grado con pesos
20	<i>term</i>	21	15.9108
206	<i>prison</i>	17	21.1576
198	<i>permit</i>	14	13.3879
67	<i>believe</i>	12	11.9297
316	<i>approve</i>	12	16.0187
517	<i>violate</i>	12	10.0479
808	<i>crisis</i>	12	4.8851
705	<i>remainder</i>	11	7.1639
265	<i>warn</i>	10	4.7316
1933	<i>split</i>	10	6.8805

Top 10 por Grado

Id	Label	Grado	Grado con pesos
206	<i>prison</i>	17	21.1576
316	<i>approve</i>	12	16.0187
20	<i>term</i>	21	15.9108
315	<i>bond</i>	8	14.1677
198	<i>permit</i>	14	13.3879
67	<i>believe</i>	12	11.9297
517	<i>violate</i>	12	10.0479
789	<i>courtroom</i>	6	8.8003
865	<i>vindication</i>	6	8.8003
920	<i>stagger</i>	6	8.8003

Top 10 por Grado con pesos

Seis de los diez términos coinciden en ambas tablas, lo que refleja coherencia temática entre los nodos más conectados y relacionados. Esta coincidencia indica que los conceptos dominantes mantienen su relevancia independientemente de la existencia de otros términos, existiendo una amplitud estructural con un fuerte vínculo semántico. Destaca *bond* (fianza), que no aparece entre los nodos con mayor grado pero sí alcanza una posición alta en el ranking ponderado, lo que indica que, aunque se relaciona con menos términos distintos, sus conexiones son especialmente intensas dentro del contexto de la comunidad. De forma similar, *courtroom*, *vindication* y *stagger*¹ aparecen solo en el listado ponderado como términos formales, lo que sugiere que refuerza un entorno jurídico y de ámbito penal.

Por el contrario, términos como *crisis*, *remainder*, *warn* o *split* aparecen únicamente en el top por grado, lo que indica que poseen una mayor diversidad de conexiones pero con menor intensidad total, actuando como nodos de enlace o términos más generales que apuntan a procesos normativos, decisiones formales y situaciones de carácter jurídico. Por este motivo, el término más representativo para esta comunidad es *prison*, ya que el ascenso de la palabra *bond* (fianza) refuerza la interpretación legal y punitiva que se le puede atribuir a la temática dominante.

¹que podría venir de *staggering sentence* (sentencia desproporcionada) o *stagger payments* (pagos prorrataeados).

Comunidad 2 - Investigate

En la Comunidad 2 la estructura no presenta un núcleo tan marcado como en otras comunidades. Aunque *sewer*, *brief*, *adjustment* e *investigate* aparecen en ambas métricas, el grado con pesos muestra una distribución más homogénea entre varios términos, sin un liderazgo tan acusado, donde en el ranking ponderado se introducen términos distintos con mayor intensidad.

Id	Label	Grado	Grado	Id	Label	Grado	Grado
			con pesos				con pesos
1498	<i>sewer</i>	19	14.7579	1345	hire	11	16.3325
208	file	17	9.0250	1470	assure	10	15.0955
454	gun	17	4.5346	1498	<i>sewer</i>	19	14.7579
458	sheriff	17	8.1063	383	reject	13	13.4912
233	<i>brief</i>	16	12.0738	378	privilege	11	13.3967
324	<i>adjustment</i>	16	12.0738	21	<i>investigate</i>	15	13.3669
3661	corporate	16	10.2493	204	employe	12	12.8397
21	<i>investigate</i>	15	13.3669	233	<i>brief</i>	16	12.0738
202	pension	15	10.0948	324	<i>adjustment</i>	16	12.0738
1731	object	15	7.5810	1497	provision	12	11.3156

Top 10 por Grado

Top 10 por Grado con pesos

El término *sewer* posee el mayor grado y es el tercer término con mayor grado con pesos. Esto lo convierte en un nodo estructural fuerte que, si aparece junto a palabras como *adjustment*, *corporate*, *pension*, *provision* o *investigate*, podría estar vinculado a gestión pública, contratos municipales o revisiones administrativas relacionadas con servicios urbanos. Es un sustantivo concreto que puede ser candidato a término representativo, pero deja sin representar a varios elementos de la comunidad.

Uno de los términos que más destaca es *gun*, que aparece en grado simple pero no en el ranking ponderado, lo que hace probable que se conecte con muchos términos distintos pero no forme parte del núcleo temático más cohesionado, situándolo como término contextual, no central. Podría estar relacionado con regulación o licencias, si en el corpus existen textos administrativos sobre armas. La comunidad en general combina nodos con alta diversidad de conexiones (*file*, *gun*, *sheriff*, *pension*). Si analizamos la proximidad numérica en el grado con pesos, se sugiere un entorno temático compartido que aborda diferentes partes de un mismo tema. El conjunto apunta a procesos administrativos y procedimentales, donde investigación, contratación, revisión y decisiones formales forman parte del mismo núcleo.

Por otro lado, *hire* y *assure* son menos estructurales aunque lideren el peso ponderado. Por ello, el término más representativo para esta comunidad es *investigate*, ya que aparece entre los nodos con mayor grado y peso y podría englobar el resto de relaciones ya que si el campo semántico dominante es procedimental o administrativo, funciona como verbo articulador.

Comunidad 3 - Police

En la Comunidad 3 se observa una estructura jerárquica, ya que *police* presenta una diferencia numérica amplia respecto al resto de términos en el grado con pesos, consolidándose como núcleo dominante. Otros términos como *construction*, *hearing*, *attorney*, *admit*, *cabinet* y *property* ocupan posiciones destacadas en ambas métricas, constituyendo el núcleo estructural y semántico del subgrafo. Estos nodos combinan una alta diversidad de conexiones con una intensidad elevada de coaparición, situándose como centros semánticos de la red.

Id	Label	Grado	Grado con pesos	Id	Label	Grado	Grado con pesos
772	<i>police</i>	47	59.9574	772	<i>police</i>	47	59.9574
331	<i>construction</i>	34	36.8864	331	<i>construction</i>	34	36.8864
470	<i>hearing</i>	33	30.2936	10	<i>jury</i>	19	31.4183
219	<i>attorney</i>	26	31.0277	219	<i>attorney</i>	26	31.0277
798	<i>admit</i>	25	29.5414	470	<i>hearing</i>	33	30.2936
222	<i>property</i>	24	25.6138	798	<i>admit</i>	25	29.5414
777	<i>arrest</i>	24	22.1554	138	<i>fire</i>	17	29.2208
1000	<i>cabinet</i>	24	26.2667	28	<i>receive</i>	21	28.1654
1	<i>investigation</i>	23	16.9025	1000	<i>cabinet</i>	24	26.2667
137	<i>court</i>	23	14.4315	222	<i>property</i>	24	25.6138

Top 10 por Grado

Top 10 por Grado con pesos

El núcleo principal presenta una coincidencia elevada entre ambas tablas, con siete de los diez términos en común en ambos rankings. Esto indica que los conceptos centrales mantienen relevancia tanto en la amplitud de relaciones como en la intensidad total de sus vínculos.

Entre las diferencias más relevantes se observa que términos como *arrest*, *investigation* y *court* aparecen únicamente en el listado por grado, lo que indica que se relacionan con un número amplio de términos distintos pero con menor intensidad acumulada. Mientras tanto, palabras como *jury*, *fire* y *receive* solo aparecen en el ranking ponderado, lo que indica relaciones más intensas aunque con menos conexiones distintas. El ascenso de *jury* en el ranking ponderado refuerza la dimensión judicial del clúster, ya que mantiene vínculos intensos con los términos centrales.

El campo semántico integra el ámbito penal y judicial con elementos de gestión institucional, lo que sugiere un entorno vinculado a seguridad pública y administración formal. Por este motivo, el término más representativo para esta comunidad es *police*, ya que aparece como el nodo con mayor grado y peso y sintetiza el eje temático dominante.

Comunidad 4 - Diplomatic

En la Comunidad 4 se observa una estructura con un núcleo dominante moderado, donde *consultation* presenta el mayor grado y el mayor peso, situándose como punto central del subgrafo. Junto a *ally*, *diplomatic* y *prevent*, forma el bloque principal que concentra tanto diversidad de conexiones como intensidad de coocurrencia. No obstante, tras *consultation* los valores del grado con pesos se agrupan en un rango próximo, lo que indica una cohesión compartida más que una jerarquía muy marcada.

Id	Label	Grado	Grado con pesos
1149	<i>consultation</i>	49	41.8272
1037	<i>ally</i>	41	34.0270
1229	<i>diplomatic</i>	41	34.0270
1238	<i>intervention</i>	38	28.3684
1228	<i>bloc</i>	35	27.7394
1297	<i>neutralist</i>	35	27.7394
1057	<i>bitter</i>	32	22.5169
1051	<i>tension</i>	31	19.7045
1175	<i>prevent</i>	30	35.9384
1061	<i>allied</i>	29	23.5419

Top 10 por Grado

Id	Label	Grado	Grado con pesos
1149	<i>consultation</i>	49	41.8272
1175	<i>prevent</i>	30	35.9384
972	<i>reaction</i>	26	34.2553
1037	<i>ally</i>	41	34.0270
1229	<i>diplomatic</i>	41	34.0270
169	<i>political</i>	27	33.9719
1022	<i>climate</i>	24	33.1499
1088	<i>global</i>	24	32.5985
1148	<i>timely</i>	24	32.5985
1184	<i>inclination</i>	24	32.5985

Top 10 por Grado con pesos

Entre las variaciones más relevantes se observa que términos como *intervention*, *bloc*, *neutralist*, *bitter*, *tension* y *allied* aparecen únicamente en el ranking por grado. Esto indica que se relacionan con un número amplio de términos distintos, aunque con menor intensidad acumulada, actuando como elementos de contexto general.

Por el contrario, palabras como *reaction*, *political*, *climate*, *global*, *timely* o *inclination* solo aparecen en el listado ponderado. Esto sugiere que mantienen relaciones más intensas dentro de contextos específicos y refuerzan una dimensión más abstracta y estratégica del discurso diplomático. Destaca especialmente *prevent*, cuyo peso es significativamente alto en relación con su grado, lo que indica una función relevante en la cohesión temática.

El conjunto apunta a relaciones internacionales, alianzas y mecanismos de cooperación orientados a evitar conflictos. Aunque *consultation* es el término estructuralmente más fuerte, *diplomatic* sintetiza mejor el ámbito temático general de la comunidad. Por este motivo, el término más representativo para este clúster es *diplomatic*.

Comunidad 5 - Estimate

En la Comunidad 5 se observa una estructura relativamente equilibrada. *estimate* ocupa la primera posición en ambas métricas, pero la diferencia con *proposal* en el grado con pesos es mínima, lo que indica un núcleo compartido más que un liderazgo aislado. Términos como *residential*, *outlay*, *approval*, *victim* e *increase* conforman un bloque fuerte al presentar valores ponderados próximos entre sí.

Id	Label	Grado	Grado con pesos
482	<i>estimate</i>	29	24.4942
559	<i>outlay</i>	21	18.3271
554	<i>residential</i>	20	19.7080
2062	<i>flat</i>	20	13.9847
1192	<i>chain</i>	19	13.6469
2090	<i>builder</i>	19	7.0864
1067	<i>atomic</i>	18	13.7516
1426	<i>shopping</i>	17	8.8281
2076	<i>victim</i>	17	17.3913
459	<i>approval</i>	16	17.4590

Top 10 por Grado

Id	Label	Grado	Grado con pesos
482	<i>estimate</i>	29	24.4942
581	<i>proposal</i>	14	24.3684
554	<i>residential</i>	20	19.7080
559	<i>outlay</i>	21	18.3271
459	<i>approval</i>	16	17.4590
2076	<i>victim</i>	17	17.3913
364	<i>increase</i>	14	17.2862
498	<i>personal</i>	10	14.3840
904	<i>payment</i>	13	14.1531
610	<i>requirement</i>	12	14.0937

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *flat*, *chain*, *builder*, *atomic* y *shopping* aparecen únicamente en el listado por grado, lo que indica conexiones amplias pero menos intensas.

Por el contrario, palabras como *proposal*, *increase*, *personal*, *payment* o *requirement* solo aparecen en el ranking ponderado, lo que señala relaciones más concentradas y recurrentes en contextos específicos.

El término *victim* es relevante porque mantiene grado y peso similares a *approval* e *increase*, lo que sugiere un proceso administrativo donde se estiman daños, se formulan propuestas y se aprueban compensaciones, posiblemente en entornos residenciales.

El campo semántico se vincula a procedimientos económicos y formales, por lo que el término más representativo es *estimate*, ya que lidera en grado y peso y articula el conjunto temático del subgrafo.

Comunidad 6 - Mechanism

En la Comunidad 6 se observa una estructura relativamente cohesionada donde los términos *stem*, *very*, *vary* y *mechanism* aparecen en posiciones destacadas en ambas métricas como núcleo estructural y semántico del subgrafo. La palabra *very* tiende a aparecer de forma extremadamente recurrente, lo que posiblemente sea debido a apariciones junto a términos técnicos de esta comunidad, como *artificial*, *protective* o *mechanism*, para reforzar descripciones por lo que su co-aparición no es aleatoria, sino que está anclada a conceptos técnicos de este clúster específico; su naturaleza adverbial sugiere un uso intensificador más que conceptual.

Id	Label	Grado	Grado con pesos
1332	<i>stem</i>	23	14.7057
1016	<i>very</i>	22	20.5590
3389	<i>vary</i>	20	16.1942
3947	<i>mechanism</i>	20	13.7127
1431	enterprise	18	10.6537
3768	profession	18	10.5577
1126	european	17	9.0358
1039	circumstance	15	7.6644
1111	application	15	10.1116
168	eliminate	14	6.1563

Top 10 por Grado

Id	Label	Grado	Grado con pesos
1016	<i>very</i>	22	20.5590
4968	<i>scheme</i>	11	16.7398
4987	<i>protective</i>	11	16.7398
5014	<i>brand</i>	11	16.7398
5029	<i>paradox</i>	11	16.7398
5045	<i>substitute</i>	11	16.7398
5059	<i>artificial</i>	11	16.7398
3389	<i>vary</i>	20	16.1942
1332	<i>stem</i>	23	14.7057
3947	<i>mechanism</i>	20	13.7127

Top 10 por Grado con pesos

El ranking ponderado muestra un bloque compacto formado por *scheme*, *protective*, *brand*, *paradox*, *substitute* y *artificial*, todos con los mismos valores, lo que indica cohesión temática en torno a conceptos técnicos o abstractos, donde *vary*, *stem* y *mechanism* se integran en este entorno conceptual relacionado con procesos, variación y funcionamiento de sistemas.

El campo semántico sugiere un ámbito conceptual asociado a mecanismos, cambios y condiciones de operación. Por este motivo, el término más representativo es *mechanism*, ya que aparece entre los nodos con mayor grado y peso y sintetiza el eje temático dominante del subgrafo.

Comunidad 7 - Taxi (maniobra de aviación)

En la Comunidad 7 se observa una estructura parcialmente concentrada donde los términos *automatic*, *taxi* y *halt* sugieren una comunidad con conexiones variadas en el grado simple. Esto refleja una amplitud estructural alta pero sin que el núcleo temático quede claramente definido únicamente por esta métrica. Sin embargo, al aplicar la métrica de grado con pesos, la estructura muestra una característica significativa; aparece un bloque sólido de términos relacionados con la aeronáutica: *cockpit*, *bomber*, *runway*, *jet*, *pilot*, *engine* y *landing*, que poseen exactamente el mismo peso de interacción. Esta igualdad numérica indica una coaparición recurrente y concentrada, propia de un subgrupo altamente cohesionado dentro del subgrafo. Estos términos solo aparecen en el ranking ponderado, lo que refuerza que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo característico de términos especializados.

Id	Label	Grado	Grado con pesos
3321	<i>automatic</i>	24	19.0645
3881	<i>farmer</i>	22	14.5975
1721	<i>taxi</i>	21	22.6571
3567	<i>fuel</i>	21	16.9585
1253	<i>disclose</i>	19	12.8941
1326	<i>crime</i>	17	14.2169
1762	<i>mortgage</i>	15	11.7213
817	<i>write</i>	13	6.1059
1449	<i>wage</i>	13	12.6424
3921	<i>halt</i>	12	19.3924

Top 10 por Grado

Id	Label	Grado	Grado con pesos
1721	<i>taxi</i>	21	22.6571
3921	<i>halt</i>	12	19.3924
3928	<i>cockpit</i>	12	19.3924
3932	<i>bomber</i>	12	19.3924
3933	<i>runway</i>	12	19.3924
3936	<i>jet</i>	12	19.3924
3939	<i>pilot</i>	12	19.3924
3950	<i>engine</i>	12	19.3924
3966	<i>landing</i>	12	19.3924
3321	<i>automatic</i>	24	19.0645

Top 10 por Grado con pesos

El término más llamativo es *taxi*, que no debe interpretarse como “taxi urbano”, sino como *taxis*, la maniobra de desplazamiento de una aeronave en pista. La presencia de *bomber* y *jet* refuerza que no se trata de vehículos terrestres, sino de aeronaves específicas. *Taxi* aparece en ambas métricas y lidera el grado con pesos, lo que lo convierte en el nodo de mayor intensidad del bloque aeronáutico.

El término *halt*, que ocupa una posición baja en grado simple, escala hasta la segunda posición ponderada, lo que demuestra que su uso está estrechamente vinculado a este contexto específico de la aviación. Términos como *farmer*, *fuel*, *disclose*, *crime*, *mortgage*, *write* y *wage*, que solo figuran en el grado simple, mantienen amplitud estructural, pero el peso los excluye del núcleo temático cohesionado.

El campo semántico apunta a un ámbito relacionado con la aviación, especialmente asociado a maniobras y operaciones de vuelo. Por este motivo, el término más representativo es *Taxiing*, ya que lidera en intensidad y articula el bloque técnico dominante del subgrafo.

Comunidad 8 - Development

En la Comunidad 8 se observa una estructura cohesionada donde los términos *advance*, *understand*, *development*, *weakness*, *understanding*, *budget*, *process* y *reference* aparecen en posiciones destacadas en ambas métricas, lo que indica que constituyen el núcleo estructural y semántico en una estructura jerárquica con elevada intensidad y coaparición, situándose como el centro conceptual de la red.

Id	Label	Grado	Grado con pesos
1014	<i>understanding</i>	24	15.5003
1060	<i>advance</i>	21	32.8997
1026	<i>understand</i>	20	28.7065
4537	<i>adequate</i>	19	10.4317
670	<i>development</i>	17	25.1008
462	<i>certain</i>	15	9.1241
1567	<i>reference</i>	14	12.7389
862	<i>process</i>	13	14.4028
1697	<i>weakness</i>	13	16.2344
555	<i>budget</i>	12	14.9643

Top 10 por Grado

Id	Label	Grado	Grado con pesos
1060	<i>advance</i>	21	32.8997
1026	<i>understand</i>	20	28.7065
670	<i>development</i>	17	25.1008
1697	<i>weakness</i>	13	16.2344
1014	<i>understanding</i>	24	15.5003
555	<i>budget</i>	12	14.9643
862	<i>process</i>	13	14.4028
1567	<i>reference</i>	14	12.7389
1142	<i>apparent</i>	10	11.4902
1035	<i>difficulty</i>	10	11.3314

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *adequate* y *certain* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total que palabras como *apparent* y *difficulty*, que solo aparecen en el ranking ponderado como parte de la especialización temática de la comunidad.

Esto sugiere que el análisis, la planificación y el desarrollo de procesos constituyen el campo semántico de la comunidad, es decir, un ámbito vinculado a comprensión, evaluación y mejora estructurada dentro de un marco procedimental. Por este motivo, el término más representativo es *development*, ya que identifica los nodos que sostienen la carga semántica del grupo, centrada en el avance y la evolución de procesos, y aparece entre los nodos con mayor grado y peso, sintetizando el eje temático dominante del subgrafo.

Comunidad 9 - Pledge

En la Comunidad 9 se observa una estructura moderadamente dispersa donde los términos *pledge*, *tract*, *animal* y *teamster* aparecen en posiciones destacadas en ambas métricas, lo que indica que constituyen dos bloques estructurales y semánticos basados en los pesos junto con un tercer bloque que destaca por conservar los mismos pesos en palabras como *discredited*, *desperate*, *repay*, *hoodlum* y *conservation* que emergen con un peso idéntico de 10.5921. Este enfoque toma fuerza según la teoría de partición de grafos, donde este fenómeno indica la existencia de un subgrupo con una relación “más fuerte” e intensa entre sí que con el resto de la red dado a que solo aparecen en el ranking ponderado, lo que indica que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo propio de términos más específicos o dependientes de contextos concretos.

Id	Label	Grado	Grado con pesos
1636	<i>pledge</i>	15	11.8739
1661	<i>tract</i>	10	11.7035
2128	<i>explosive</i>	10	3.9676
425	<i>calm</i>	9	4.9719
1490	<i>meaning</i>	9	4.3317
2064	<i>animal</i>	9	7.2759
2108	<i>teamster</i>	9	7.2759
441	<i>violence</i>	8	3.1471
2038	<i>intervene</i>	8	4.8900
4624	<i>wash</i>	8	5.2467

Top 10 por Grado

Id	Label	Grado	Grado con pesos
1636	<i>pledge</i>	15	11.8739
1661	<i>tract</i>	10	11.7035
1568	<i>discredited</i>	6	10.5921
1591	<i>desperate</i>	6	10.5921
1607	<i>repay</i>	6	10.5921
1657	<i>hoodlum</i>	6	10.5921
1665	<i>conservation</i>	6	10.5921
2064	<i>animal</i>	9	7.2759
2108	<i>teamster</i>	9	7.2759
17	<i>manner</i>	3	5.3877

Top 10 por Grado con pesos

Se identifican así dos bloques estructurales definidos por los pesos y un tercer bloque cohesionado por la igualdad numérica de sus valores. Dentro de esta organización, *pledge* actúa como eje central para un grupo de palabras con tintes sociales y procedimentales.

La aparición de términos con alto peso como *repay* y *conservation* en el mismo bloque que *discredited*, *desperate* y *hoodlum* sugiere una temática social conflictiva vinculada a compromisos, descrédito público y posibles tensiones comunitarias. La presencia de términos como *explosive*, *violence*, *intervene*, *calm* o *meaning* refuerza la idea de un entorno narrativo asociado a conflicto y respuesta institucional. Términos como *teamster* (miembro de un sindicato de transporte) y *animal* apuntan a un posible ámbito colectivo o laboral, lo que conecta el bloque de compromiso con dinámicas sociales más amplias.

Dado que *pledge* es el nodo con mayor fuerza dentro del grupo y mantiene presencia destacada en ambas métricas, se considera el término que mejor representa la red y articula los distintos bloques identificados.

Comunidad 10 - Market

En la Comunidad 10 se observa una estructura claramente definida donde los términos *market*, *earning*, *investor*, *production*, *equipment* y *price* aparecen en posiciones destacadas en ambas métricas. Sin embargo, el grado con pesos muestra una jerarquía más precisa, donde *market* marca un valor notablemente superior, jerarquizando la estructura y situándose como eje central temático.

Id	Label	Grado	Grado con pesos
3384	<i>market</i>	35	37.6279
4046	<i>earning</i>	33	20.4574
4032	<i>investor</i>	32	19.7702
4097	<i>marketing</i>	30	16.7314
2173	<i>compare</i>	28	18.6556
4054	<i>corn</i>	27	13.6037
4094	<i>production</i>	27	25.6798
1357	<i>equipment</i>	25	22.4271
4052	<i>competitive</i>	24	16.0284
179	<i>price</i>	23	20.4423

Top 10 por Grado

Id	Label	Grado	Grado con pesos
3384	<i>market</i>	35	37.6279
1448	<i>product</i>	17	29.9264
4094	<i>production</i>	27	25.6798
1357	<i>equipment</i>	25	22.4271
4046	<i>earning</i>	33	20.4574
179	<i>price</i>	23	20.4423
1551	<i>industry</i>	22	20.2054
4032	<i>investor</i>	32	19.7702
2289	<i>yield</i>	12	19.2198
3345	<i>improvement</i>	23	19.0698

Top 10 por Grado con pesos

Mientras que *earning* e *investor* presentan grados altos, su peso es sensiblemente inferior al de *production* y *equipment*, lo que indica que la intensidad temática no se centra únicamente en agentes financieros, sino en la dinámica productiva. El bloque formado por *product*, *production*, *equipment*, *industry*, *price* y *yield*, que aparece reforzado en el ranking ponderado, señala un entorno económico ligado a la actividad industrial y al rendimiento económico.

Términos como *marketing*, *compare*, *corn* y *competitive*, que solo aparecen en el grado simple, amplían el campo semántico pero no sostienen el núcleo de mayor intensidad. Funcionan como extensiones del discurso económico, no como su centro.

Los términos *product*, *production*, *earning*, *price* e *investor* definen un escenario comercial, un ámbito claramente vinculado a la actividad productiva y económica, por lo que el término más representativo para esta comunidad es *market*, ya que, al aparecer como el nodo con mayor grado y peso, destaca como el principal centro de la red, demostrando que en la comunidad analizada es el término que articula la mayoría de las relaciones sobre comercio e industria.

Comunidad 11 - Front

En la Comunidad 11 el grado con pesos permite identificar con claridad la estructura interna del subgrafo. Se distinguen dos bloques definidos por la proximidad numérica de sus valores.

Id	Label	Grado	Grado con pesos
2292	deadlock	17	9.1953
2125	front	15	9.3025
2170	rebel	13	8.5288
1451	catch	12	5.1904
2354	slug	12	7.8427
2442	physician	12	5.6775
1824	tactic	10	4.9889
3169	heat	10	8.6750
1306	err	9	8.4916
2510	treatment	9	8.2077

Top 10 por Grado

Id	Label	Grado	Grado con pesos
2125	front	15	9.3025
5182	chinese	7	9.2556
5188	transport	7	9.2556
5256	ante	7	9.2556
5266	presidency	7	9.2556
2292	deadlock	17	9.1953
3169	heat	10	8.6750
2170	rebel	13	8.5288
1306	err	9	8.4916
2510	treatment	9	8.2077

Top 10 por Grado con pesos

El primero lo forman *front*, *chinese*, *transport*, *ante*, *presidency* y *deadlock*, que concentran la mayor intensidad acumulada. La combinación de estos términos apunta a un contexto político o geopolítico donde aparecen posiciones enfrentadas, bloqueos y escenarios institucionales. *Front* puede interpretarse como frente político o línea de confrontación, mientras que *deadlock* refuerza la idea de estancamiento en un proceso decisional.

El segundo bloque está compuesto por *heat*, *rebel*, *err* y *treatment*, con valores homogéneos aunque algo inferiores. Este grupo introduce un marco de tensión, presión o reacción, donde la presencia de error, respuesta o tratamiento sugiere gestión de conflictos o consecuencias derivadas del escenario principal.

Los términos que solo aparecen con alto grado simple, como *catch*, que podría vincularse a dinámicas de confrontación, y *slug*, que encaja en un discurso de tensión o agresión, amplían el campo léxico pero no forman parte del núcleo de mayor intensidad. Actúan como elementos contextuales dentro del corpus sin definir la temática dominante.

El campo semántico sugiere un escenario de tensión política o diplomática, posiblemente internacional, vinculado a situaciones de bloqueo y confrontación institucional. Por este motivo, el término más representativo es *front*, que destaca por su estabilidad en ambas métricas, ocupando la segunda posición en grado simple y ascendiendo al primer puesto en grado con pesos, lo que lo consolida como el núcleo de mayor intensidad de la comunidad.

Comunidad 12 - Catcher

La Comunidad 12 presenta un núcleo temático muy específico relacionado con el béisbol. La aparición de términos con alto peso como *bunt* (toque de bola), *shortstop* (paracorto) y *teammate* (compañero de equipo), junto con nombres propios que actúan como términos en este corpus, refuerza esta interpretación.

Esto es relevante porque en inglés *mantle*, probablemente en referencia a Mickey Mantle, significa *manto*, cuyo significado no guarda relación con el resto de palabras salvo que se trate del apellido de un jugador. Esto da más fuerza a la definición de un clúster deportivo.

			Grado con pesos
Id	Label	Grado	
2851	drag	18	9.2876
2891	blond	17	15.4827
2894	ultimate	17	12.2663
2901	possess	17	15.4827
2917	mantle	17	15.4827
2360	pound	16	9.7960
2897	peak	16	9.8173
2066	catcher	15	11.3041
2872	recall	15	9.9340
2324	bunt	14	12.7305

Top 10 por Grado

		Grado con pesos
Id	Label	Grado
2891	blond	17
2901	possess	17
2917	mantle	17
2324	bunt	14
2358	shortstop	14
2390	teammate	14
2894	ultimate	17
2066	catcher	15
2919	thin	13
2872	recall	15

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *drag*, *pound* y *peak* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total, actuando como nodos contextuales amplios dentro de la comunidad.

Por el contrario, palabras como *shortstop*, *teammate* y *thin* solo aparecen en el ranking ponderado, lo que indica que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo propio de términos más específicos o dependientes de contextos concretos.

El campo semántico apunta a un ámbito deportivo, en concreto al béisbol, donde aparecen términos asociados a posiciones, jugadas y jugadores. Por este motivo, el término más representativo es *Catcher*, ya que aparece entre los nodos con mayor grado y peso y sintetiza el eje temático dominante del subgrafo. Se descarta *Mantle* porque, aunque podría remitir a un jugador concreto, el bloque temático refleja una semántica más general del deporte.

Comunidad 13 - League

En la Comunidad 13 se observa una estructura altamente cohesionada donde los términos *player*, *pennant*, *rookie*, *pitcher*, *bat*, *ball*, *team*, *league* y *game* aparecen en posiciones destacadas en ambas métricas, lo que indica que constituyen el núcleo estructural y semántico del subgrafo. Estos nodos combinan una elevada diversidad de conexiones con una intensidad de coaparición muy alta, situándose como los principales organizadores dentro de la red.

Id	Label	Grado	Grado con pesos
2357	player	63	141.2842
2286	pennant	54	91.2147
2303	rookie	54	91.2147
2315	pitcher	54	91.2147
2387	bat	54	91.2147
2323	ball	52	118.8597
721	team	49	80.0137
1488	league	49	96.7207
2199	game	49	90.5975
2295	ninth	49	66.9841

Top 10 por Grado

Id	Label	Grado	Grado con pesos
2357	player	63	141.2842
2323	ball	52	118.8597
2307	pitch	47	112.1459
1488	league	49	96.7207
2286	pennant	54	91.2147
2303	rookie	54	91.2147
2315	pitcher	54	91.2147
2387	bat	54	91.2147
2199	game	49	90.5975
721	team	49	80.0137

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que el término *ninth* aparece únicamente en el listado por grado y *pitch* solo en el ranking ponderado. Esto indica que *pitch*, aunque conecta con menos nodos distintos, mantiene vínculos más intensos con términos centrales como *player* o *ball*. La comunidad muestra una estructura semántica cohesionada: el grado simple refleja la amplitud del discurso deportivo, mientras que el grado con pesos concentra el núcleo técnico formado por *pitcher*, *bat* y *pennant*, cuyos enlaces se sitúan entre los de mayor intensidad.

El nodo *player* lidera ambas clasificaciones con una diferencia abrumadora, lo que indica que en la matriz de adyacencia de este término posee el mayor número de enlaces incidentes y que sus relaciones de co-aparición son las más fuertes y frecuentes de todo el corpus analizado.

El campo semántico es claramente homogéneo y está vinculado al ámbito deportivo, concretamente al béisbol, donde aparecen conceptos asociados a jugadores, posiciones, acciones y competición. Por este motivo, el término más representativo es *league* ya que engloba el concepto global y aparece como uno de los nodos en el ranking que sintetiza el eje temático dominante, aunque *player* sea el nodo central estructuralmente.

Comunidad 14 - Academic

En la Comunidad 14 se observa una estructura moderadamente cohesionada donde los términos *speculate*, *grip*, *recruit*, *newsman* y *mad* aparecen en posiciones destacadas en ambas métricas, al observar los términos que ganan peso (*biology*, *enrollment*, *academic*, *scholar*), aparece un núcleo temático muy específico relacionado con la universidad y la investigación con un mismo valor para todos pesos.

Id	Label	Grado	Grado con pesos
5030	<i>recruit</i>	17	13.0521
2591	<i>newsman</i>	16	11.2682
2598	<i>speculate</i>	14	15.1572
2611	<i>grip</i>	14	15.1572
5228	leak	14	8.8128
2318	chore	13	8.0454
3327	discrepancy	13	3.4763
943	qualified	12	5.3733
3182	marvel	12	8.7209
2603	<i>mad</i>	11	11.0337

Top 10 por Grado

Id	Label	Grado	Grado con pesos
2598	<i>speculate</i>	14	15.1572
2611	<i>grip</i>	14	15.1572
5030	<i>recruit</i>	17	13.0521
4969	biology	9	12.2457
5006	enrollment	9	12.2457
5019	academic	9	12.2457
5020	scholar	9	12.2457
2591	<i>newsman</i>	16	11.2682
2603	<i>mad</i>	11	11.0337
923	stay	7	10.6072

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *leak*, *chore*, *discrepancy*, *qualified* y *marvel* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total, actuando como nodos contextuales amplios dentro de la comunidad.

Las nuevas palabras que aparecen en el ranking de grado con peso *biology*, *enrollment*, *academic*, *scholar* y *stay* indican que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo característico de términos más específicos o dependientes de contextos concretos.

Sin embargo, hay palabras claves como *recruit*, *grip*, *newsman* y *speculate* que hacen pensar en reportes y noticias entorno a un sector académico, es por ello que *academic* es el término que mejor define la esencia del clúster si se prioriza un concepto que englobe una temática centrar.

Comunidad 15 - Poet

En la Comunidad 15 se aprecia una estructura dispersa organizada en cuatro grupos. En ambas métricas, los términos *conductor*, *wheel* y *style* ocupan posiciones centrales, por lo que actúan como ejes semánticos de la comunidad.

Un segundo grupo lo forman *poet*, *austere* y *roman*, que apunta a un contexto literario o estético. Un tercer conjunto, compuesto por *viewer*, *sigh* y *meaningless*, introduce una dimensión interpretativa o valorativa. Por último, *dull* y *tag* presentan un grado con pesos muy similar, lo que indica una intensidad de coaparición comparable.

Id	Label	Grado	Grado con pesos
2126	wheel	14	10.2449
2602	conductor	14	12.8794
2669	red	12	6.3482
2060	male	10	5.9002
2665	powerful	10	3.2460
4733	ancient	10	6.8665
2646	lifetime	9	4.2030
4152	dull	8	7.4662
4715	style	8	10.5212
2586	tag	7	7.4639

Top 10 por Grado

Id	Label	Grado	Grado con pesos
2602	conductor	14	12.8794
4715	style	8	10.5212
2126	wheel	14	10.2449
4734	roman	7	8.8502
4765	poet	7	8.8502
4772	austere	7	8.8502
2553	viewer	6	7.7017
2628	sigh	6	7.7017
2635	meaningless	6	7.7017
4152	dull	8	7.4662

Top 10 por Grado con pesos

En la comparación entre ambas métricas se observa que términos como *red*, *male*, *powerful*, *ancient* y *lifetime* aparecen únicamente entre los primeros por grado sin pesos. Esto indica que se conectan con muchos nodos distintos, pero con una intensidad media menor.

La copresencia de *wheel* y *conductor*, junto con *poet*, sugiere una posible referencia a un texto concreto donde estos términos coaparecen, como podría ser el poema *Wheels* de Jim Daniels. Puede afirmarse que existe una coocurrencia sistemática entre estos términos dentro del mismo conjunto documental, lo que sugiere una crítica o análisis de su poemario.

Con base en esto, el término que mejor representa la identidad semántica de la comunidad es *poet*, ya que actúa como núcleo interpretativo del conjunto y da sentido a la descripción del uso conjunto de las palabras.

Comunidad 16 - Golfer

En la Comunidad 16 se observa una estructura temática definida donde términos como *slice*, *golfer*, *par*, *fairway*, *tee* y *bunker* aparecen concentrados en posiciones relevantes, lo que indica la presencia de un subcampo léxico claramente especializado. Estos nodos muestran relaciones intensas entre sí y forman el núcleo semántico dominante del subgrafo.

Id	Label	Grado	Grado con pesos
3082	swim	26	13.1190
2811	nerve	25	16.0043
2976	setting	23	16.5333
2977	italian	23	16.5333
2775	slice	21	19.6605
2822	soft	21	19.6605
1867	chapter	20	16.3190
2731	amateur	20	11.1253
3044	candy	20	8.6244
1563	motor	19	15.1308

Top 10 por Grado

Id	Label	Grado	Grado con pesos
2762	golfer	16	20.6904
2763	par	16	20.6904
2772	fairway	16	20.6904
2818	tee	16	20.6904
2820	bunker	16	20.6904
2775	slice	21	19.6605
2822	soft	21	19.6605
2976	setting	23	16.5333
2977	italian	23	16.5333
1867	chapter	20	16.3190

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *swim*, *nerve*, *amateur*, *candy* y *motor* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total, actuando como nodos contextuales generales dentro de la comunidad.

Por el contrario, los términos *golfer*, *par*, *fairway*, *tee* y *bunker* solo aparecen en el ranking ponderado, lo que indica que mantienen relaciones muy intensas entre sí aunque con menos conexiones distintas, característica propia de vocabulario especializado y cohesionado.

El campo semántico está claramente vinculado al ámbito deportivo del golf, donde aparecen términos propios de reglas, jugadas y elementos del campo. Por este motivo, el término más representativo es *golfer*, ya que describe de forma directa el dominio temático que estructura el subgrafo.

Comunidad 17 - Cocktail

En la Comunidad 17 se observa una estructura temática coherente donde términos como *wear*, *guest*, *chic*, *pink*, *gown*, *cocktail*, *invitation* y *orchestra* aparecen concentrados en posiciones relevantes, lo que indica la presencia de un subcampo léxico claramente asociado a contextos sociales formales. Estos nodos presentan relaciones intensas entre sí y forman el núcleo semántico dominante del subgrafo.

Id	Label	Grado	Grado con pesos
3189	pink	40	36.1221
2967	gown	34	30.8768
3023	entertain	34	30.8768
3054	wear	33	45.5821
3283	chic	31	34.8132
2978	engagement	30	21.7875
4293	wardrobe	30	24.4597
2947	buffet	29	19.6070
2939	publicity	28	29.7137
3097	guest	28	36.5391

Top 10 por Grado

Id	Label	Grado	Grado con pesos
1516	cocktail	26	47.4608
3054	wear	33	45.5821
3280	lady	27	40.4794
2928	marriage	24	38.2980
3097	guest	28	36.5391
3096	orchestra	19	36.4394
3189	pink	40	36.1221
2969	invitation	19	35.9160
3283	chic	31	34.8132
761	color	22	34.4726

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *entertain*, *engagement*, *wardrobe*, *buffet* y *publicity* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total, actuando como nodos contextuales generales dentro de la comunidad.

Por el contrario, palabras como *cocktail*, *lady*, *marriage*, *orchestra*, *invitation* y *color* solo aparecen en el ranking ponderado, lo que indica que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo característico de vocabulario asociado a situaciones concretas y cohesivas.

El campo semántico está claramente vinculado a contextos sociales formales y celebraciones, donde aparecen términos relacionados con vestimenta, asistentes y elementos propios de eventos. Por este motivo, el término más representativo es *cocktail*, ya que describe de forma directa el dominio temático que estructura el subgrafo.

Comunidad 18 - Ballroom/Designer

En la Comunidad 18 se observa una estructura temática cohesionada donde los términos *fun*, *designer*, *yellow*, *ballroom*, *handsome* y *charm* aparecen en posiciones destacadas en ambas métricas, lo que indica que constituyen el núcleo estructural y semántico del subgrafo. Estos nodos presentan simultáneamente una alta diversidad de conexiones y una intensidad elevada de coaparición, situándose como los principales organizadores conceptuales dentro de la red.

Id	Label	Grado	Grado con pesos
3223	<i>fun</i>	32	31.4950
3016	<i>handsome</i>	31	26.6047
4279	<i>designer</i>	30	30.5695
4461	<i>yellow</i>	30	30.5695
3273	<i>ballroom</i>	27	26.7968
4301	<i>charm</i>	26	23.2020
4308	<i>lively</i>	25	20.3991
4340	<i>athlete</i>	25	20.3991
3037	<i>typical</i>	23	12.9775
1531	<i>love</i>	22	18.8029

Top 10 por Grado

Id	Label	Grado	Grado con pesos
3223	<i>fun</i>	32	31.4950
4279	<i>designer</i>	30	30.5695
4461	<i>yellow</i>	30	30.5695
3273	<i>ballroom</i>	27	26.7968
3016	<i>handsome</i>	31	26.6047
4301	<i>charm</i>	26	23.2020
4287	<i>showing</i>	16	21.5013
4342	<i>charming</i>	16	21.5013
4371	<i>bake</i>	16	21.5013
4408	<i>decorate</i>	16	21.5013

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *lively*, *athlete*, *typical* y *love* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total, actuando como nodos contextuales generales dentro de la comunidad.

Por el contrario, palabras como *showing*, *charming*, *bake* y *decorate* solo aparecen en el ranking ponderado, lo que indica que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo característico de términos asociados a situaciones concretas y cohesivas.

El campo semántico está vinculado a contextos sociales y estéticos donde predominan referencias a apariencia, ambiente y actividades recreativas. Por este motivo, los términos más representativos son *ballroom* o *designer* porque en ambas métricas confirma que son los pilares que sostienen la identidad temática de esta partición del grafo.

Comunidad 19 - Legislative

En la Comunidad 19 se observa una estructura temática coherente donde los términos *majority*, *senate*, *legislative*, *lieutenant*, *underlie*, *undermine*, *extension* y *normal* aparecen en posiciones destacadas en ambas métricas, lo que indica que constituyen el núcleo estructural y semántico del subgrafo. Estos nodos combinan una diversidad apreciable de conexiones con una intensidad de coaparición elevada, situándose como los principales organizadores conceptuales dentro de la red.

Id	Label	Grado	Grado con pesos
659	<i>majority</i>	18	12.5393
1917	<i>enjoy</i>	14	2.3023
762	<i>underlie</i>	13	10.2589
764	<i>extension</i>	13	9.2418
3383	<i>consist</i>	12	5.2182
658	<i>legislative</i>	11	10.1177
521	<i>undermine</i>	10	10.2161
587	<i>senate</i>	10	10.2161
644	<i>lieutenant</i>	10	11.7355
701	<i>normal</i>	10	10.2161

Top 10 por Grado

Id	Label	Grado	Grado con pesos
659	<i>majority</i>	18	12.5393
726	<i>associate</i>	7	11.8017
644	<i>lieutenant</i>	10	11.7355
762	<i>underlie</i>	13	10.2589
521	<i>undermine</i>	10	10.2161
587	<i>senate</i>	10	10.2161
701	<i>normal</i>	10	10.2161
658	<i>legislative</i>	11	10.1177
764	<i>extension</i>	13	9.2418
489	<i>fiscal</i>	9	9.0952

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *enjoy* y *consist* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total, actuando como nodos contextuales generales dentro de la comunidad.

Por el contrario, palabras como *associate* y *fiscal* solo aparecen en el ranking ponderado, lo que indica que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo característico de términos asociados a contextos más específicos.

El campo semántico está claramente vinculado al ámbito político e institucional, donde aparecen conceptos relacionados con órganos legislativos, jerarquías y procesos normativos. Por este motivo, el término más representativo es *legislative*, ya que aparece en ambas tablas y resume el eje temático que organiza el subgrafo.

Comunidad 20 - Movement

En la Comunidad 20 se observa una estructura temática heterogénea donde los términos *movement*, *book*, *advanced* y *bag* aparecen en posiciones destacadas en ambas métricas, lo que indica que constituyen el núcleo estructural y semántico del subgrafo. Estos nodos presentan una diversidad apreciable de conexiones junto con una intensidad de coaparición relevante, actuando como puntos organizadores dentro de la red.

Id	Label	Grado	Grado con pesos
1466	<i>govern</i>	20	9.9990
19	<i>conduct</i>	19	9.0764
3476	<i>advanced</i>	19	11.8341
3512	<i>bag</i>	19	11.8341
493	<i>book</i>	16	13.5660
1477	<i>movement</i>	16	14.5587
2206	<i>organized</i>	16	6.6668
4657	<i>fundamental</i>	16	8.0733
1405	<i>type</i>	15	9.6044
1622	<i>division</i>	15	9.5814

Top 10 por Grado

Id	Label	Grado	Grado con pesos
2205	<i>dinner</i>	14	22.3284
799	<i>honor</i>	9	20.4638
2161	<i>father</i>	13	14.6317
1477	<i>movement</i>	16	14.5587
1350	<i>headquarters</i>	12	14.2470
493	<i>book</i>	16	13.5660
3783	<i>guidance</i>	13	13.5374
4637	<i>library</i>	14	13.0460
3476	<i>advanced</i>	19	11.8341
3512	<i>bag</i>	19	11.8341

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *govern*, *conduct*, *organized*, *fundamental*, *type* y *division* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total, actuando como nodos contextuales generales dentro de la comunidad.

Por el contrario, palabras como *dinner*, *honor*, *father*, *headquarters*, *guidance* y *library* solo aparecen en el ranking ponderado, lo que indica que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo característico de términos asociados a contextos concretos y cohesivos.

El campo semántico sugiere un entorno conceptual vinculado a organización, actividad colectiva y contextos institucionales o sociales donde aparecen referencias a dirección, eventos y espacios estructurados. Por este motivo, el término más representativo es *movement*, ya que aparece en ambas tablas y resume el eje temático que articula el subgrafo.

Comunidad 21 - Talent

En la Comunidad 21 se observa una estructura temática definida donde los términos *stick*, *talent* y *vow* aparecen en posiciones destacadas en ambas métricas, lo que indica que constituyen el núcleo estructural del subgrafo. Sin embargo, el conjunto de términos con mayor peso revela un subcampo léxico claramente especializado vinculado al ámbito artístico y escénico.

Id	Label	Grado	Grado con pesos
4579	stick	24	15.5197
2711	personality	21	14.1118
1091	fix	16	6.9702
4655	vow	16	14.3896
1879	hero	15	10.1835
1642	trip	13	9.8713
2900	rare	11	12.6343
816	summer	10	8.6867
1280	huge	10	4.5679
2888	talent	10	14.6705

Top 10 por Grado

Id	Label	Grado	Grado con pesos
4579	stick	24	15.5197
4568	instrument	10	15.3037
4599	skip	10	15.3037
4616	pianist	10	15.3037
4631	ballet	10	15.3037
4632	performer	10	15.3037
4633	symphony	10	15.3037
4651	enthral	10	15.3037
2888	talent	10	14.6705
4655	vow	16	14.3896

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *personality*, *hero*, *trip*, *summer* y *huge* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total, actuando como nodos contextuales generales dentro de la comunidad.

Por el contrario, palabras como *instrument*, *pianist*, *ballet*, *symphony* y *enthral* solo aparecen en el ranking ponderado, lo que indica que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo característico de vocabulario especializado y cohesionado.

El campo semántico está claramente vinculado al ámbito artístico y escénico, donde aparecen conceptos relacionados con interpretación musical y actuación. Por este motivo, el término más representativo es *talent*, ya que aparece dentro de la tabla ponderada y describe de forma directa el dominio temático que organiza el subgrafo.

Comunidad 22 - Battle

En la Comunidad 22 se observa una estructura cohesionada donde los términos *battle*, *communism*, *drastic*, *summon*, *pursue* y *routine* aparecen en ambas métricas, lo que indica que constituyen el núcleo estructural y semántico del subgrafo. Estos nodos combinan diversidad de conexiones con intensidad de coaparición, situándose como los elementos más estables de la comunidad.

Id	Label	Grado	Grado con pesos
3309	<i>drastic</i>	14	8.7519
3337	<i>summon</i>	14	8.7519
3594	<i>pursue</i>	13	7.0110
4112	<i>communism</i>	12	10.9307
4146	<i>routine</i>	12	5.8626
355	<i>battle</i>	11	11.8936
2865	<i>trio</i>	10	9.3312
2909	<i>preach</i>	10	9.3312
2449	hip	9	2.9730
3386	encouraging	9	5.3125

Top 10 por Grado

Id	Label	Grado	Grado con pesos
355	<i>battle</i>	11	11.8936
4112	<i>communism</i>	12	10.9307
2865	<i>trio</i>	10	9.3312
2909	<i>preach</i>	10	9.3312
3309	<i>drastic</i>	14	8.7519
3337	<i>summon</i>	14	8.7519
440	threat	3	8.6965
4148	accompany	5	7.1139
3594	<i>pursue</i>	13	7.0110
4146	<i>routine</i>	12	5.8626

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *hip* y *encouraging* aparecen únicamente en el listado por grado, lo que indica que se conectan con varios términos distintos pero con menor intensidad total, funcionando como elementos contextuales secundarios dentro del subgrafo.

Por el contrario, palabras como *threat* y *accompany* solo aparecen en el ranking ponderado, lo que sugiere relaciones más intensas aunque con menos conexiones distintas, rasgo propio de vocabulario dependiente de contextos específicos.

El campo semántico está asociado a acciones, conflicto y discurso ideológico, donde aparecen términos relacionados con confrontación, convocatoria y posicionamiento político. Por este motivo, el término más representativo es *communism*, ya que sintetiza el eje conceptual dominante y organiza semánticamente el conjunto de relaciones del subgrafo.

Comunidad 23 - Scientific

En la Comunidad 23 se observa una estructura relativamente estable donde los términos *implement*, *significance*, *scientific* y *organ* aparecen en ambas métricas, lo que indica que constituyen el núcleo estructural y semántico del subgrafo. Estos nodos combinan una diversidad notable de conexiones con una intensidad alta de coaparición, por lo que actúan como puntos centrales dentro de la red léxica.

Id	Label	Grado	Grado con pesos
840	<i>significance</i>	19	15.9856
3903	<i>scientific</i>	16	13.2788
4206	<i>implement</i>	16	16.1372
2234	<i>nominate</i>	15	7.6020
2578	<i>organ</i>	15	13.1642
3428	<i>suburban</i>	15	11.4732
1367	<i>enemy</i>	14	7.2114
2954	<i>success</i>	13	9.2594
787	<i>client</i>	12	4.8045
1411	<i>grocery</i>	12	7.1645

Top 10 por Grado

Id	Label	Grado	Grado con pesos
4206	<i>implement</i>	16	16.1372
840	<i>significance</i>	19	15.9856
3903	<i>scientific</i>	16	13.2788
2578	<i>organ</i>	15	13.1642
2653	<i>arrangement</i>	9	12.8308
4904	<i>informal</i>	9	12.6511
4916	<i>illusion</i>	9	12.6511
4923	<i>broad</i>	9	12.6511
4932	<i>fatal</i>	9	12.6511
4955	<i>initiative</i>	9	12.6511

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *nominate*, *suburban*, *enemy*, *client* y *grocery* aparecen únicamente en el listado por grado, lo que indica que se conectan con múltiples palabras distintas pero con menor frecuencia total, funcionando como nodos contextuales amplios dentro del subgrafo.

Por el contrario, palabras como *arrangement*, *informal*, *illusion*, *initiative* o *fatal* solo aparecen en el ranking ponderado, lo que sugiere que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo típico de términos más especializados y dependientes de contextos concretos.

El campo semántico está asociado a procesos conceptuales y terminología analítica vinculada al ámbito científico y metodológico. Por este motivo, el término más representativo es *scientific*, ya que sintetiza el eje temático dominante y organiza semánticamente el conjunto de relaciones del subgrafo.

Comunidad 24 - Murder

En la Comunidad 24 se observa una estructura donde los términos *murder*, *death*, *delegate*, *grade* y *uncle* aparecen en ambas métricas, lo que indica que constituyen el núcleo estructural del subgrafo. Estos nodos presentan simultáneamente diversidad de conexiones e intensidad de coaparición, por lo que actúan como puntos estables dentro de la red léxica.

Id	Label	Grado	Grado con pesos
3361	<i>murder</i>	18	13.8694
2247	<i>mature</i>	17	10.6203
171	<i>elaborate</i>	16	11.1266
444	<i>death</i>	16	13.1834
2210	<i>delegate</i>	16	11.4510
2994	<i>hint</i>	16	6.3101
3640	<i>grade</i>	16	11.2311
2942	<i>uncle</i>	15	11.1851
2943	<i>aunt</i>	15	11.1851
3116	<i>civilian</i>	15	6.1706

Top 10 por Grado

Id	Label	Grado	Grado con pesos
3361	<i>murder</i>	18	13.8694
444	<i>death</i>	16	13.1834
3582	<i>collection</i>	12	12.2605
3520	<i>appeal</i>	10	11.8602
3636	<i>institution</i>	13	11.5116
3662	<i>existence</i>	13	11.5116
2210	<i>delegate</i>	16	11.4510
3583	<i>faculty</i>	12	11.2654
3640	<i>grade</i>	16	11.2311
2942	<i>uncle</i>	15	11.1851

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *mature*, *elaborate*, *hint*, *aunt* y *civilian* aparecen únicamente en el listado por grado, lo que indica que se relacionan con varios términos distintos pero con menor frecuencia total, funcionando como elementos contextuales dentro de la comunidad.

Por el contrario, palabras como *collection*, *appeal*, *institution*, *existence* o *faculty* solo aparecen en el ranking ponderado, lo que indica que mantienen relaciones más intensas aunque con menos conexiones distintas, rasgo propio de términos más específicos o dependientes de contextos concretos.

El campo semántico se organiza en torno a conceptos asociados a mortalidad, consecuencias y contextos sociales relacionados con fallecimiento y responsabilidad. Por este motivo, el término más representativo es *death*, ya que sintetiza el eje temático dominante y estructura el significado global del subgrafo.

Comunidad 25 - Enact

En la Comunidad 25 se observa una estructura donde el término *enact* aparece en ambas métricas junto con varios nodos de fuerte carga jurídica y normativa, lo que indica que el subgrafo se organiza en torno a procesos legales y regulatorios. La coexistencia de términos relacionados con delito, sentencia y legislación muestra que el núcleo semántico está definido por vocabulario propio del ámbito judicial.

Id	Label	Grado	Grado con pesos
529	<i>enact</i>	26	20.7854
894	employer	26	14.4611
778	criminal	25	8.9822
3467	missile	24	10.8210
704	credit	22	13.0036
779	disclosure	21	11.5444
914	excess	21	12.1162
936	freedom	21	14.4427
3491	indictment	20	14.2949
3506	convict	20	14.2949

Top 10 por Grado

Id	Label	Grado	Grado con pesos
2074	sentence	14	22.9521
529	<i>enact</i>	26	20.7854
1362	vehicle	18	17.9736
153	citizen	14	17.9622
2934	sister	15	16.2134
3882	interfere	18	16.0827
3091	die	13	15.7493
837	prosecution	12	15.6911
838	unfair	12	15.6911
855	constitute	12	15.6911

Top 10 por Grado con pesos

Entre las diferencias más relevantes se observa que términos como *employer*, *missile*, *credit*, *disclosure* o *excess* aparecen únicamente en el listado por grado, lo que indica que se relacionan con muchos términos distintos pero con menor intensidad total, funcionando como nodos contextuales amplios dentro del subgrafo.

Por el contrario, palabras como *sentence*, *prosecution*, *constitute* o *unfair* solo aparecen en el ranking ponderado, lo que sugiere relaciones más intensas aunque con menos conexiones distintas, rasgo característico de vocabulario más específico y dependiente de contextos jurídicos concretos.

El campo semántico está vinculado a procedimientos legales, responsabilidad penal y aplicación de normas. Por este motivo, el término más representativo es *criminal*, ya que sintetiza el dominio conceptual predominante y organiza semánticamente el conjunto de relaciones del subgrafo.

Conclusiones

Referencias

- [1] Elavarasan P. *Natural Language processing (Basics to SOTA models) Part-1.* Disponible en <https://digitaldata.science.blog/2022/01/18/natural-language-processing-basics-to-sota-models-part-1/>.
- [2] Wikipedia. *Natural Language Toolkit.* Disponible en https://en.wikipedia.org/wiki/Natural_Language_Toolkit.
- [3] NLTK Team. *Natural Language Toolkit (NLTK) Documentation.* Disponible en <https://www.nltk.org/>. 2026.
- [4] Luis Merino Ulizarna. *NLTK: tus primeros pasos con Procesamiento del Lenguaje Natural.* Disponible en <https://adictosaltrabajo.com/2023/07/27/nltk-python/>. 2026.
- [5] Wikipedia. *spaCy.* Disponible en <https://en.wikipedia.org/wiki/SpaCy>.
- [6] spaCy Developers (Explosion AI). *spaCy: Library Architecture.* Disponible en <https://spacy.io/api>. 2025.
- [7] Explosion AI. *spaCy: Model Architectures.* Disponible en <https://spacy.io/api/architectures>.
- [8] Wikipedia. *BERT (modelo de lenguaje).* Disponible en [https://es.wikipedia.org/wiki/BERT_\(modelo_de_lenguaje\)](https://es.wikipedia.org/wiki/BERT_(modelo_de_lenguaje)).
- [9] Marcello Politi (2022). *Custom Named Entity Recognition with BERT.* Disponible en [https://towardsdatascience.com/custom-named-entity-recognition-with-bert-cf1fd4510804/](https://towardsdatascience.com/custom-named-entity-recognition-with-bert-cf1fd4510804).
- [10] Vikas Yadav and Steven Bethard (2019). *A Survey on Recent Advances in Named Entity Recognition.* Disponible en <https://arxiv.org/abs/1910.11470>.
- [11] DeepAI. (s.f.). *CoNLL-2003 Named Entity Recognition Dataset.* Disponible en <https://data.deepai.org/conll2003.zip>.