



Vicerrectorado de Calidad

Clasificación de las respuestas a preguntas abiertas del cuestionario de satisfacción con el profesorado por medio de redes neuronales y algoritmos de clasificación

DOCENTIA UCM



UNIVERSIDAD
COMPLUTENSE
MADRID

Vicerrectorado de Calidad

Clasificación de las respuestas a preguntas abiertas por medio de redes neuronales

Proceso analítico

Modelo de datos

Cuestionario DOCENTIA-UCM

Topic modeling

Deep learning y clustering

Procesamiento

Adaptación de los datos

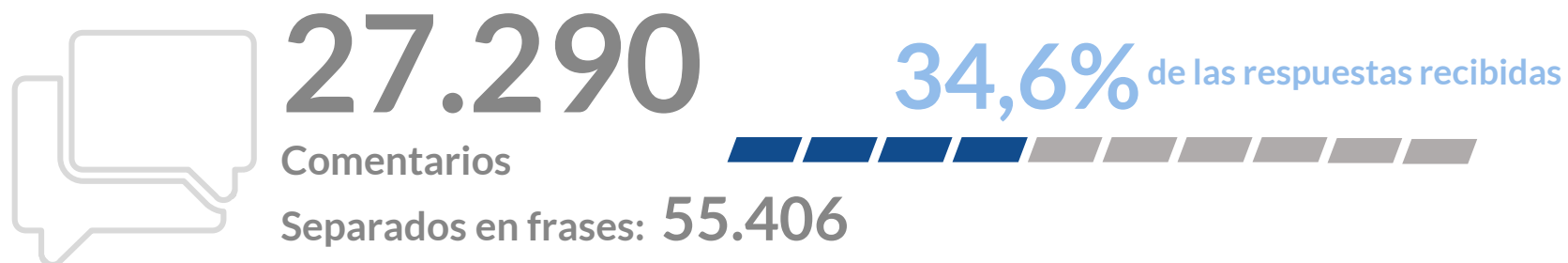
Representación y análisis

Visualización de las redes de relaciones

Cuestionario de satisfacción con el profesorado

Docentia UCM 2019-2020

Respuestas a la pregunta abierta de observaciones del cuestionario DOCENTIA-UCM del curso 2019-2020.



Con este volumen de respuestas abiertas es inviable la clasificación manual de todas ellas mediante su codificación de categorías para obtener al final un análisis de las frecuencias, tal y como se ha realizado este proceso de forma tradicional:

- Clasificar las respuestas elegidas en temas
- Darle un nombre a cada tema.
- Asignar un código a cada patrón general de respuestas

Procesamiento

Depuración por algoritmos



Se realiza un análisis objetivo automático de cada una de las respuestas abiertas por medio de un proceso desarrollado en Python:

- Separar cada uno de los comentarios en **frases**.
- Eliminar **signos de puntuación**.
- Eliminar **acentos**.
- Eliminar ***stopwords*** o palabras vacías: Artículos, pronombres, preposiciones, adverbios.
- **Lematización** de palabras: Proceso que pasa de la forma flexionada de cada palabra (es decir, en plural, en femenino, conjugada, etc), al lema correspondiente, quedándonos con las palabras que por convenio se aceptan como representantes de todas las formas flexionadas de una misma palabra, reduciendo así las variantes morfológicas.

Topic modelling

Red neuronal

El objetivo de este proceso es el de agrupar, de forma automática, los comentarios o frases en conglomerados o temas expuestos por los estudiantes en sus respuestas a la pregunta abierta de observaciones del cuestionario. Como resultado final obtendríamos una clasificación para cada comentario o frase, y el porcentaje de comentarios que hablan de cada tema:

1. Creación de *word embeddings* a través del modelo de red neurona *word2vec* implementado en *Skip-gram*. Agregando a la red las respuestas abiertas de todos los estudiantes a la pregunta de observaciones.

Con el vocabulario generado con las palabras del corpus, el objetivo es entrenar la red neuronal con las sentencias del corpus para que, dada una palabra, nos diga la probabilidad de que cada palabra del vocabulario sea vecina de la primera. De esta manera *Word2Vec* usará vectores de palabras para representar la información semántica de las palabras, aprendiendo el comportamiento de los textos a través de un espacio de incrustación que hace que las palabras semánticamente similares estén muy cerca en este espacio. La incrustación (*embeddings*) son en realidad un mapeo de la palabra desde el espacio original al nuevo espacio multidimensional.

2. Una vez la red neuronal ha creado los vectores de las relaciones entre las palabras y las ha dado un peso, realizamos la agrupación de los comentarios en conglomerados a través de *K-Media* que es un algoritmo de agrupamiento, que tiene como objetivo la partición de un conjunto de observaciones en *k* grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano.
3. Por último creamos los bigramas de frecuencias de co-ocurrencia (utilización conjunta de dos unidades léxicas en una unidad superior) de palabras por parejas y la matriz de relaciones para *Gephi*. Como resultado se obtendrá una matriz de relaciones por cada conglomerado.

Resultados |



UNIVERSIDAD
COMPLUTENSE
MADRID

Vicerrectorado de Calidad

Representación del corpus

En principio parece que hay un eje que son los **profesores y profesoras y las asignaturas**. Un segundo nivel encontramos las **prácticas, las explicaciones, las clases**. Y si seguimos descendiendo las **dudas, los trabajos, el interés**.



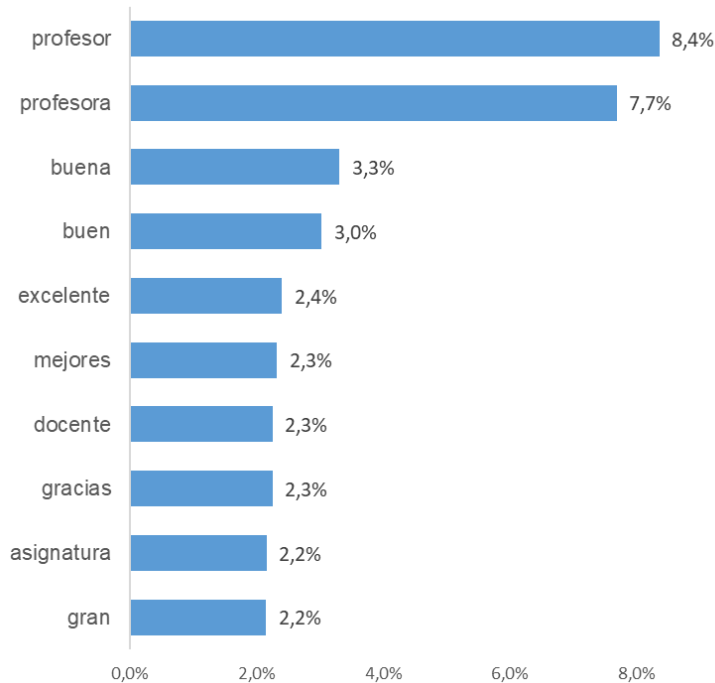
Clustering

Primera aproximación

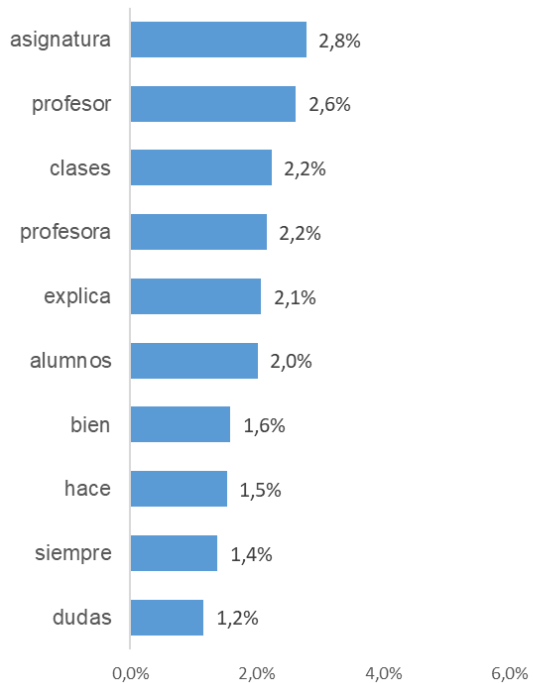
Tras el proceso de aprendizaje de la red neuronal con las sentencias del corpus creamos la matriz de vectores que es analizada con el algoritmo e clasificación K-Medias del que obtenemos tres *clusters* (0, 1 y 2) que acumulan cada uno el 14, 2%, el 26,1 % y el 59,7%, respectivamente, del corpus de sentencias.

La frecuencia de las primeras 10 palabras con mayor aparición en cada uno de los clústeres son:

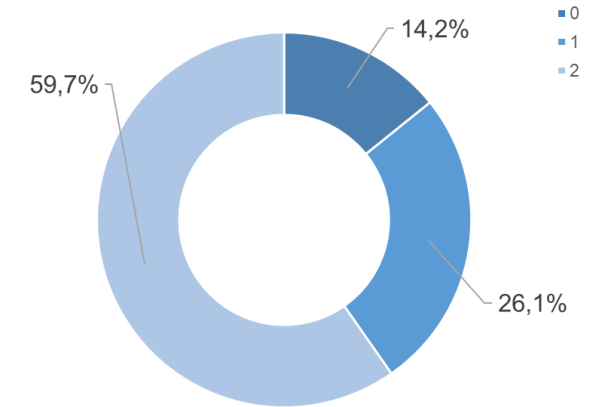
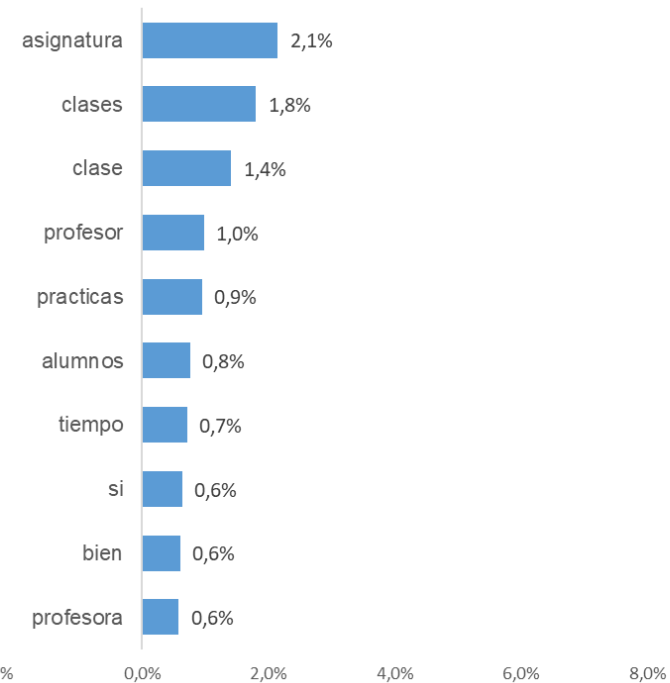
Cluster 0



Cluster 1



Cluster 2

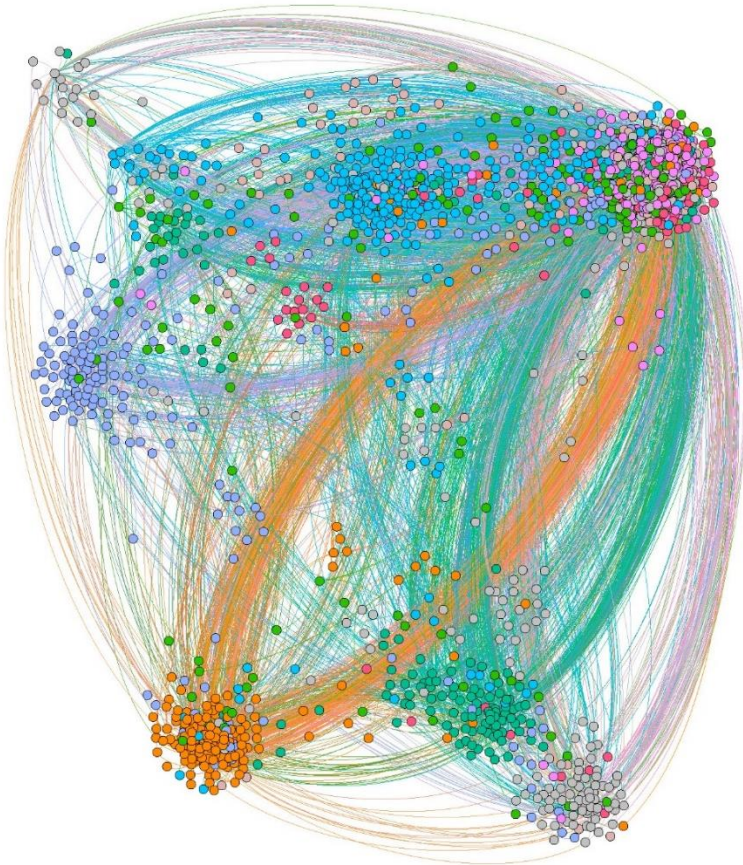


Representación de los *clusters*

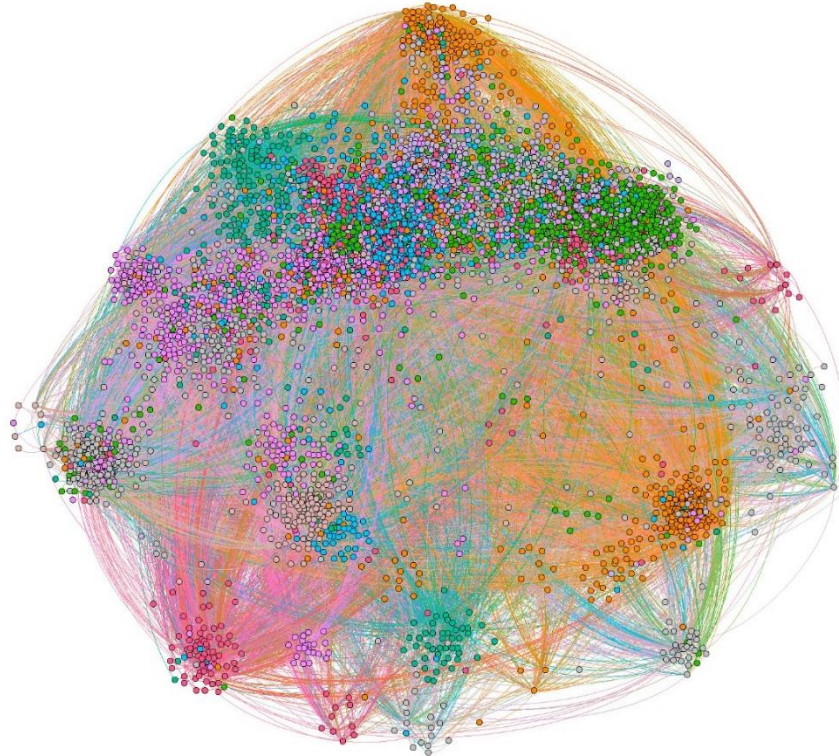


Para poder obtener una visión lo más completa, es imprescindible unificar en una sólo visualización gráfica tanto los procesos de relación de las sentencias del corpus realizado por la red neuronal, como el de clasificación de estas relaciones. Para ello utilizamos Gephi, una herramienta open-source desarrollada para visualizar y analizar grandes gráficos de red, permitiendo localizar patrones y tendencias en una base de datos.

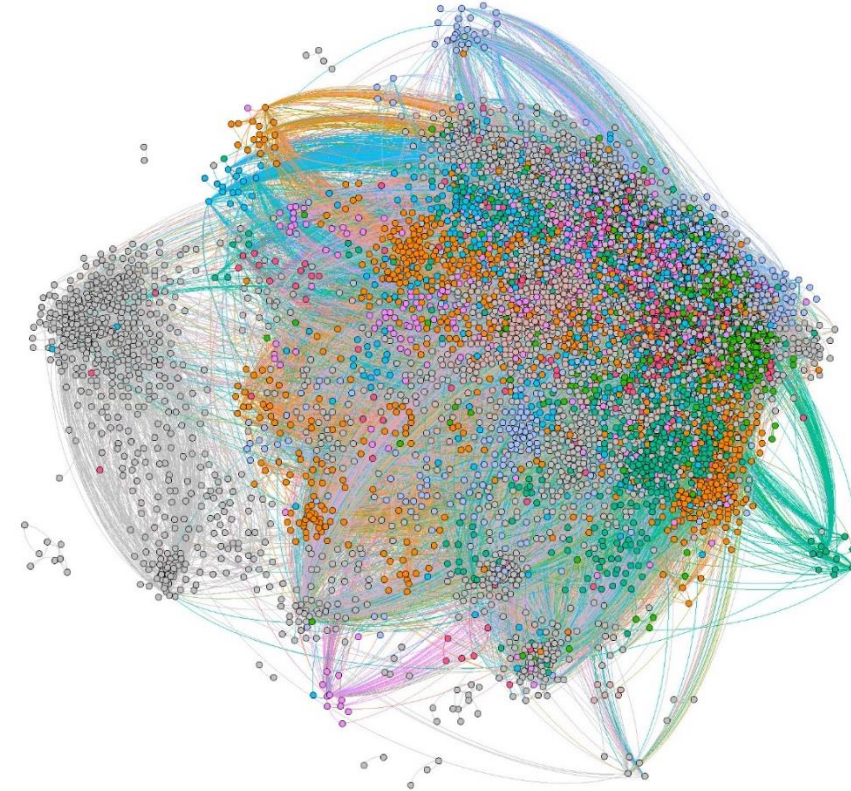
Cluster 0 (todas las palabras)



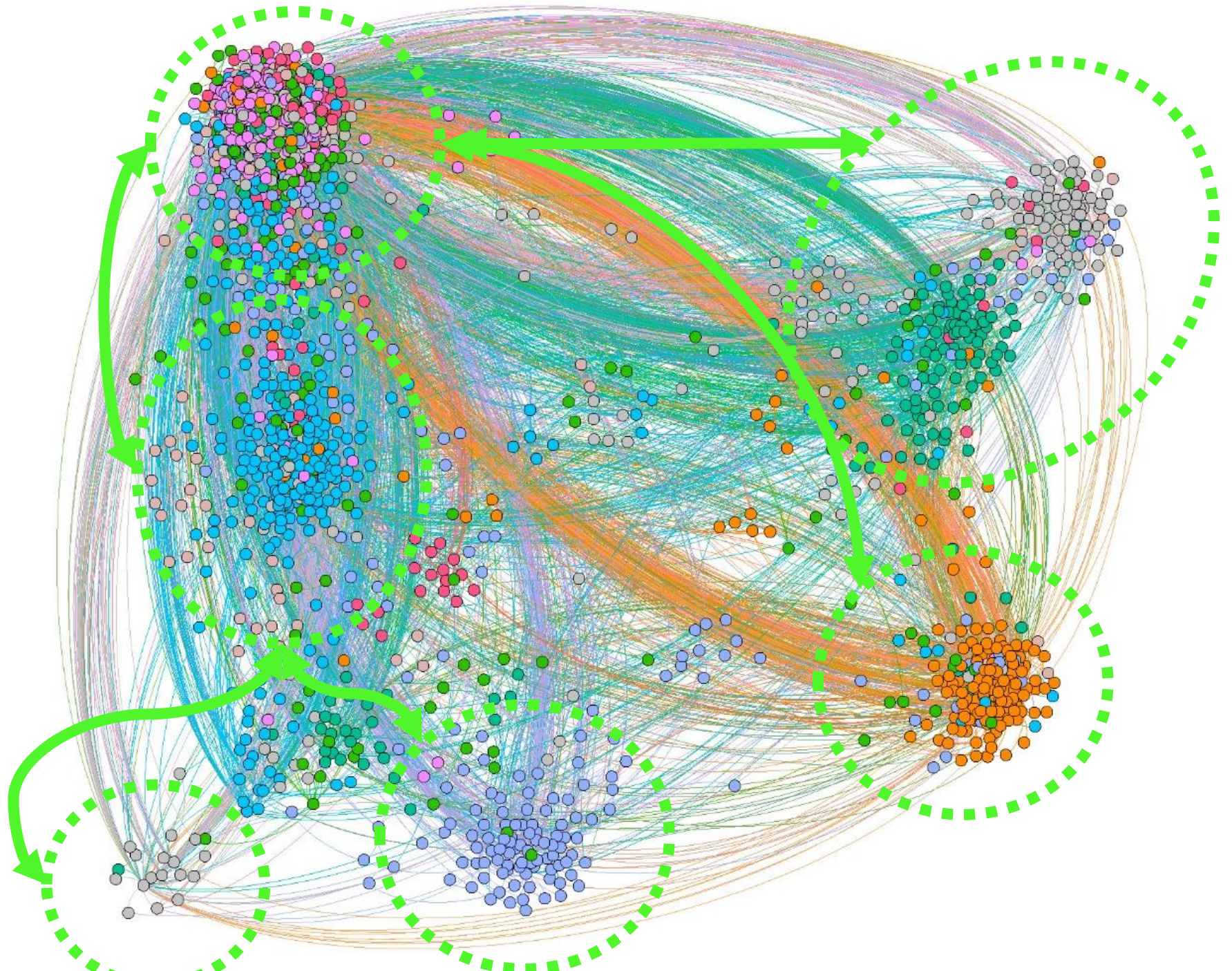
Cluster 1 (todas las palabras)



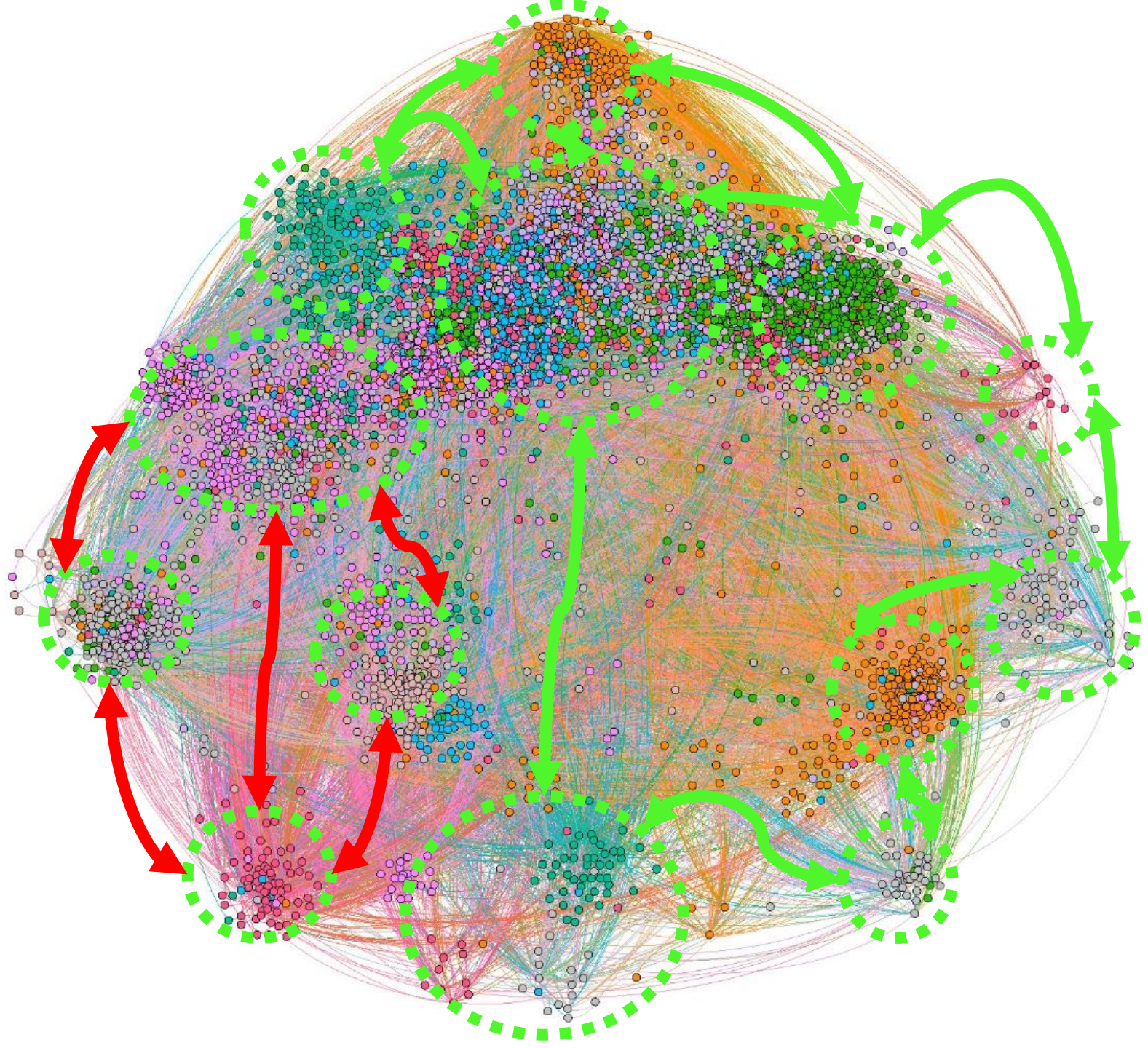
Cluster 2 (todas las palabras)



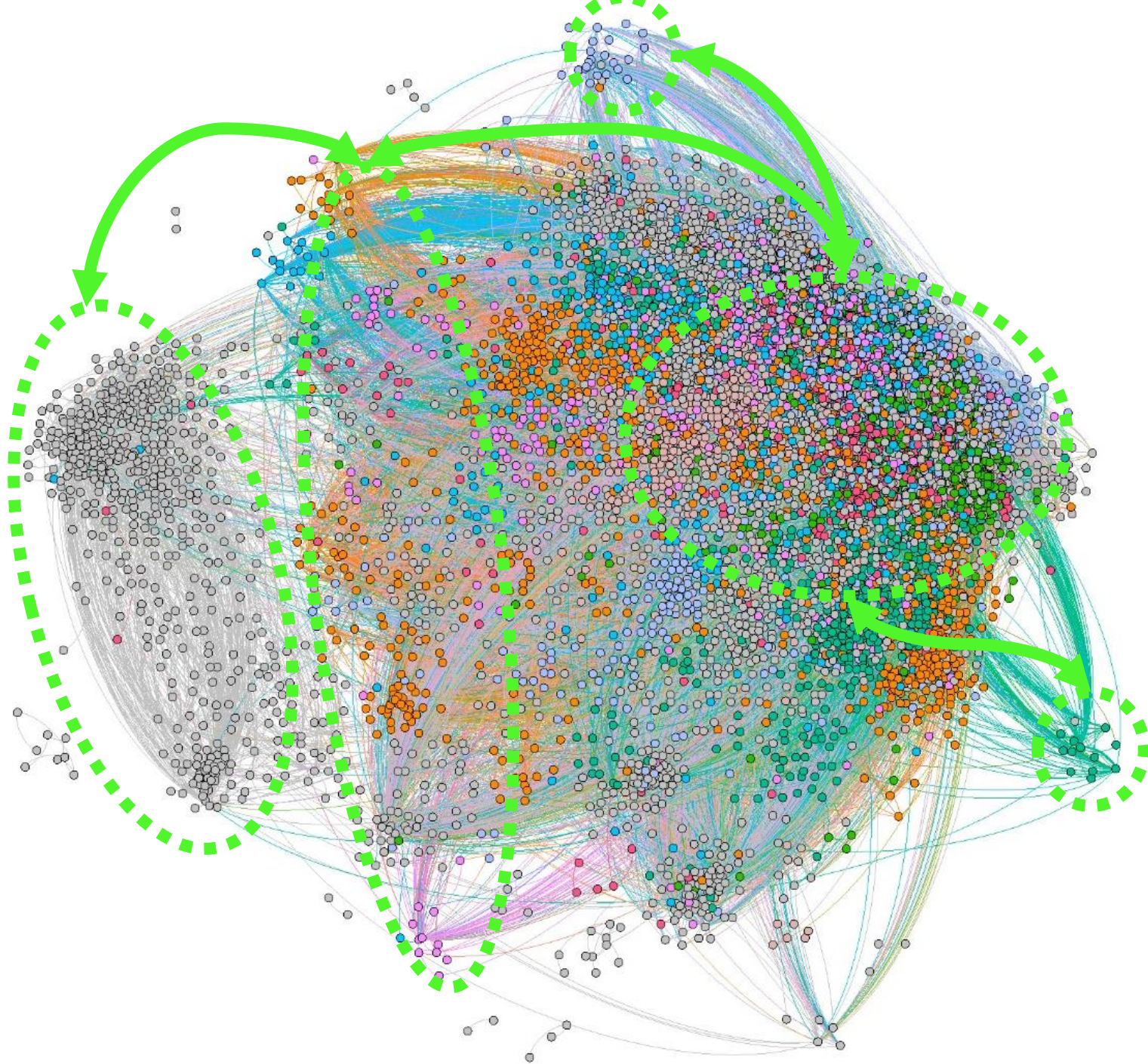
Cluster 0



Cluster 1



Cluster 2



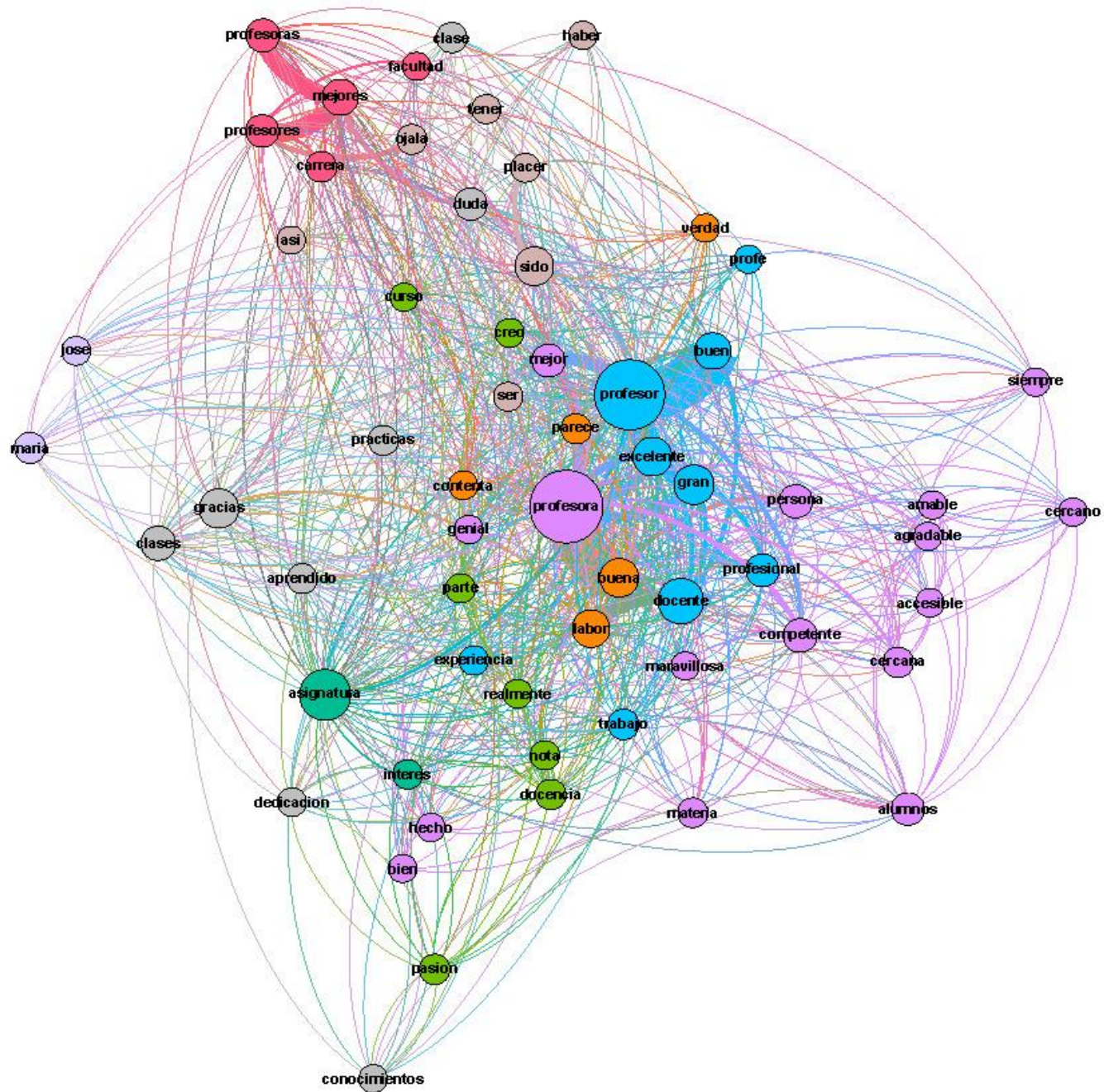
Más cerca

(Principales conceptos)

Cluster 0

14,2%

1. **Profesor**, de manera concreta respecto al profesor los estudiantes opinan que realiza un excelente trabajo, que la experiencia le convierten en un buen docente y gran profesional.
2. **Profesora**, las opiniones de los estudiantes son más positivas en cuanto a la calidad humana de la profesora, expresan que es una persona amable, agradable, genial, y que siempre se muestra accesible y cercana a los alumnos. Respecto a su labor de profesora consideran que es competente y que hace que una asignatura sea mejor.
3. **Profesores y profesoras**, la comparación de las y los profesores en términos generales se realiza dentro de la carrera y dentro de la facultad.
4. **Gratitud**, hay un gran sentimiento de gratitud de los estudiantes hacia sus profesores. Tanto por los conocimientos aprendidos, por la dedicación con que impartieron sus clases y prácticas, así como por la respuesta a las dudas planteadas.
5. **Pasión**, otro aspecto valorado por los estudiantes es la pasión con la que imparte la docencia un profesor, consideran que es algo realmente evidente y que además influye en su calificación.
6. **Temas relacionados**, las opiniones expresadas con menor peso por parte los estudiantes dentro de este clúster son los sentimiento y reacción que genera en el alumno las características de su profesor, así el alumnado opina que una buena labor docente se traduce en un estudiante contento y que es el docente quien genera interés en la asignatura que imparte.



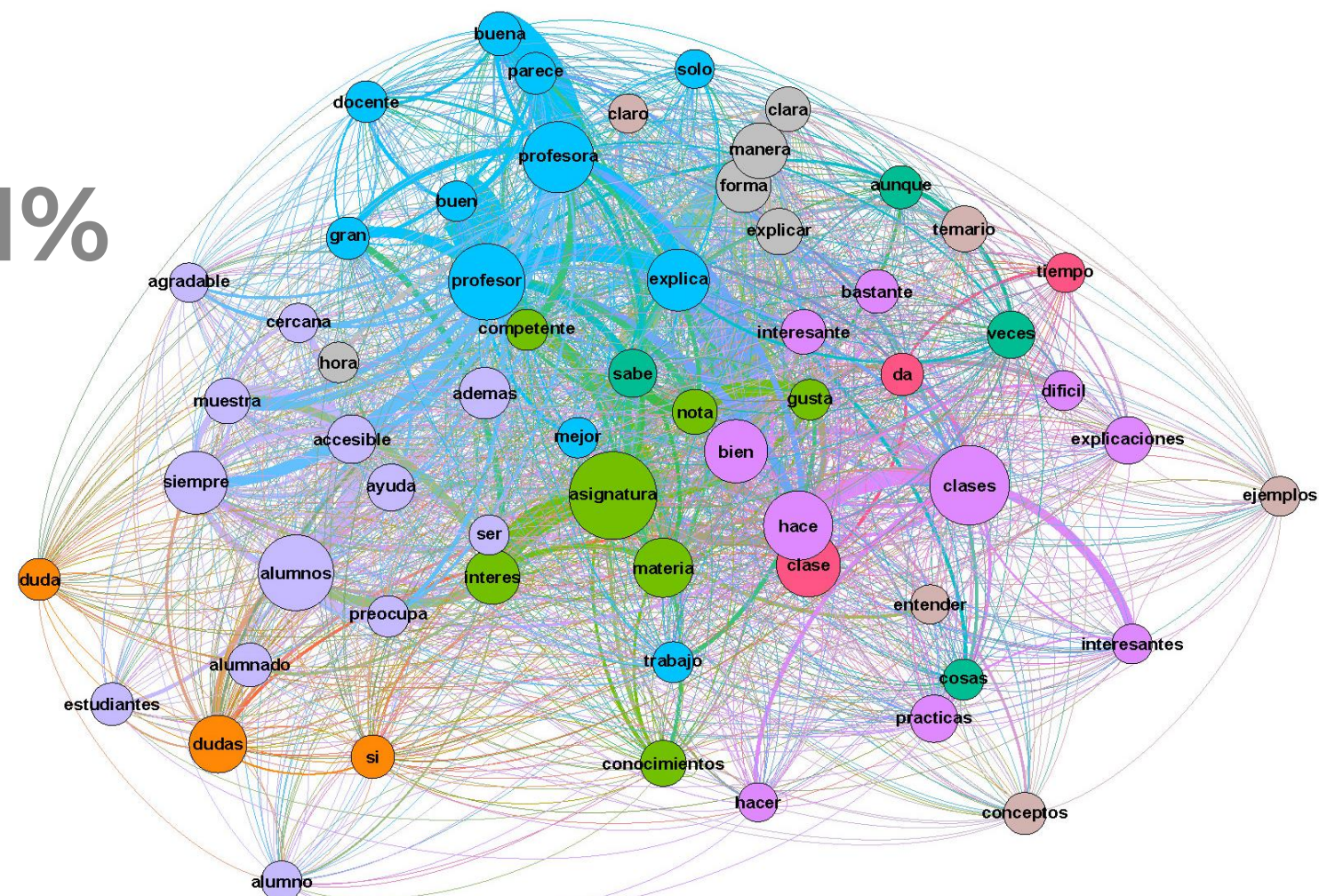
Buena enseñanza

Cluster 1

26,1%

Buena enseñanza: Esta dimensión hace referencia a la percepción que tienen los estudiantes de las prácticas de enseñanza en el curso, grado, facultad, o institución educativa, incluye ítems relacionados con la actividad del docente: proporciona retroalimentación útil y oportuna, da explicaciones claras, motiva a los estudiantes, hace que el curso sea interesante y comprende los problemas de los estudiantes. En el clúster hay palabras como: profesor, profesora, asignatura, clases, forma de explicar, ejemplos, explicaciones, entender, accesible, interés, claro, duda, dudas...

Este clúster se corresponde con la dimensión **Buena enseñanza** del *Course Experience Questionnaire* (CEQ) (Cuestionario sobre Experiencia en el Curso). Esta dimensión se divide en dos dimensiones: buenos materiales y buenas tutorías.



Los 5 temas tratados + 1

1. **Asignatura y materia**, dentro de este tema los estudiantes aprecian si el docente es competente para impartir conocimientos y generar interés y gusto por la asignatura que luego se traduce en la nota, otro aspecto altamente valorado por los estudiantes.
2. **Profesor, profesora y docente**, el estudiante valora el trabajo del docente en relación a sus explicaciones y lo cataloga.
3. **Clases**, otro aspecto evaluado por los alumnos dentro de la buena enseñanza son las clases. El estudiante valora si el docente hace que las clases sean interesante, buenas o si las explicaciones son difíciles, también valora si se hacen prácticas en clase.
4. **Alumnos, estudiantes y alumnado**, la relación del docente con el alumnado es objeto de observación por los estudiantes. Aprecian que el docente siempre se muestre agradable, cercano, accesible y que además se preocupe y ayude.
5. **Temario**, a los estudiantes les interesa que el docente explique claramente el temario y entender los conceptos.
6. **Temas relacionados**, con menor representación dentro de este clúster, pero también apreciados por los estudiantes se encuentran temas como: claridad en la forma o manera de explicar, el tiempo de la clase y la gestión de dudas.

Objetivos y estándares claros de la asignatura

Cluster 2

Objetivos y estándares claros de la asignatura. Esta denominación se corresponde con una de las dimensiones del CEQ que lleva el mismo nombre. Adicionalmente, en el clúster hay palabras como: contenido, evaluación, docentes, exámenes, temario, trabajos, práctica, teoría, online, apuntes, entre otros.

Los 4 temas tratados + 1

1. **Asignatura**, la opinión de los estudiantes respecto a los docentes gira en torno a los contenidos de la materia. Según el alumnado, la mejor manera de aprender los temas es mediante prácticas que permiten captar el contenido de manera interesante.
2. **Clase**, respecto a cada día de clase de determinado curso, el alumno es consciente que cada profesor tiene una forma particular de dar o impartir un tema. Los estudiantes distinguen entre una clase de ejercicios y una de teoría, además, valoran más las horas de ejercicios y piden menos horas de teoría dentro del temario.
3. **Clases**, de manera general los estudiantes consideran que en las clases online los alumnos precisan de más ayuda y es algo que los profesores deben tener en cuenta. Respecto al tiempo, el alumnado considera que muchas veces las horas dedicadas para adquirir conocimientos no son suficientes y por tanto hace difícil adquirir el aprendizaje y es algo que a consideración del estudiante el profesor lo debe tener en cuenta al momento de la evaluación.
4. **Prácticas y trabajos**, según los estudiantes los trabajos, ya sean prácticas en clase o trabajos fuera del aula, deben estar relacionados con el método de evaluación del proyecto formativo o asignatura. Proponen el siguiente orden: primero la realización de prácticas en clase y luego de trabajos en casa, además, que en el contenido evaluado en los exámenes se debe considerar el conocimiento adquirido en las prácticas y trabajos, de manera que influya en la nota final y que no se les evalúe luego de otras cosas.
5. **Temas relacionados**, con menor peso dentro de este clúster se consideran temas como apuntes que toman los estudiantes y la importancia de la asignatura dentro de la carrera.

