

Actualización, análisis y visualización de la producción científica en el campo de las nanociencias y nanotecnologías en México hasta el año 2020

Proyecto CONACYT: Ciencia de Frontera 2019. No. 304320 Una revisión crítica del desarrollo de las nanotecnologías en México acorde a las prioridades socio-económicas nacionales

Eduardo Robles-Belmont
IIMAS, UNAM

Agosto 2021

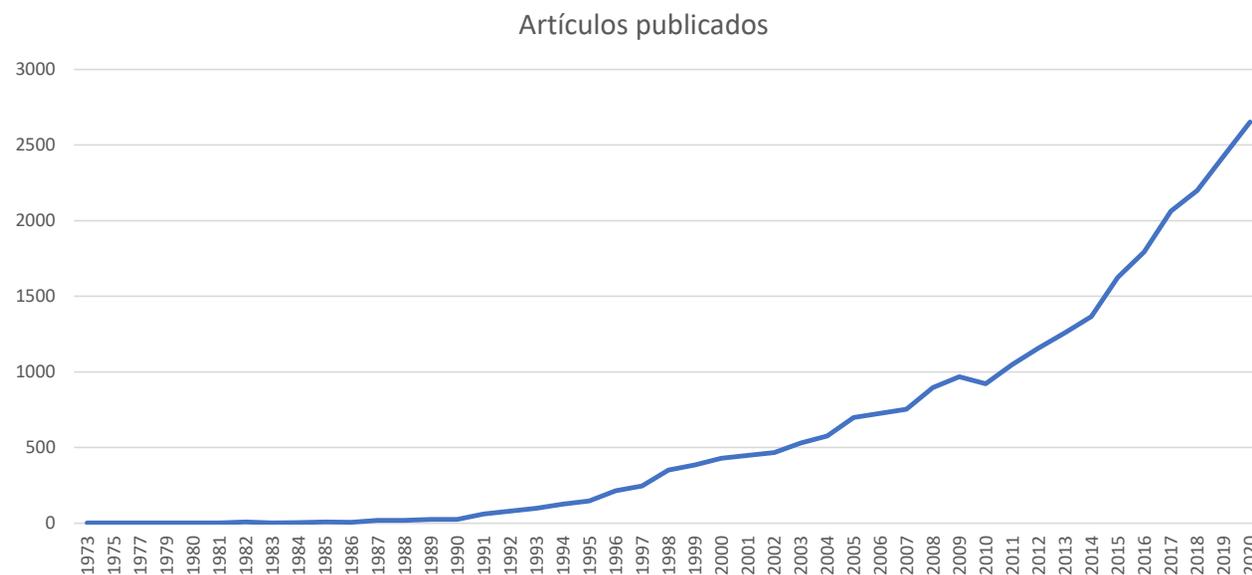
La estrategia de búsqueda es una combinación de diversas propuestas en la literatura en la bibliometría y la cienciometría sobre el caso de las nanociencias y nanotecnologías. Se trata de estrategias de búsqueda de datos en la colección principal de la Web of Science, las cuales han sido revisadas por nanotecnólogos y estrategias evolutivas (método estadístico de representatividad). La combinación se realizó en un estudio previo (Robles-Belmont y Vinck, 2011).

Los resultados reportados en este estudio parten de datos bibliográficos obtenidos de la colección principal de la Web of Science, cuya consulta se realizó el 20 de febrero de 2021 y la actualización de la base de datos WoS fue el 19 de febrero del mismo año. Los resultados retenidos han sido hasta el año 2020, por lo cual es necesario advertir sobre el efecto de la cola de indización en torno a las referencias del último año. Esto se refiere a que a lo largo del año 2021 seguirán siendo indizados documentos científicos publicados en el año 2020, los cuales no tenemos registro en nuestra base de datos actual.

El análisis de los datos se ha hecho con el programa informático Access. Las tablas y gráficos con Excel y RawGraphs, así como otras aplicaciones para visualizar datos que se usarán en función de las necesidades del análisis en el proyecto. Además, el análisis y visualización de los datos se ha realizado con herramientas para el análisis de redes sociales: Pajek, UCINET, VOSViewer, MapEquation, etc. Finalmente el análisis de contenido, semántico y de minería de texto se puede realizar con WordMapper, T-Lab, etc. La elección de las herramientas y metodologías para el análisis de los datos se hará a partir de las necesidades y las preguntas de investigación en el marco del proyecto.

Los resultados presentados en este reporte son parciales y buscan mostrar un panorama muy general de la base de datos actualizada hasta el año 2020. El objetivo central es presentar resultados que animen la reflexión y formulación de preguntas a partir de los datos obtenidos. Un texto esta en proceso de escritura con estos datos y se usarán algunas de las gráficas y visualizaciones, si necesitan para sus textos alguna visualización o gráfica no olviden avisarme para proporcionarles y no usar las mismas imágenes en dos o más textos.

El análisis solo considera los artículos científicos por tratarse de conocimiento certificado (ha pasado por un proceso de discusión, revisión y evaluación).



Tipo de documento	Documentos	% de 30924
Article	26858	86,85%
Proceedings Paper	5391	17,43%
Review	909	2,94%
Book Chapter	328	1,06%
Meeting Abstract	328	1,06%
Editorial Material	107	0,35%
Early Access	96	0,31%
Letter	57	0,18%
Note	39	0,13%
Correction	33	0,11%
Data Paper	17	0,05%
Book Review	2	0,01%
Retracted Publication	1	0,00%
Correction, Addition	1	0,00%
Discussion	1	0,00%
Hardware Review	1	0,00%
News Item	1	0,00%
Retraction	1	0,00%

Perfil disciplinar del campo nano (science map)

Un indicador interesante de mirar concierne a las categorías disciplinarias donde son indizadas las revistas en las cuales se han publicado los artículos científicos. En el caso de la Web of Science tenemos las WoS Categories, con las cuales podemos realizar lo que se conoce como un *science map*, yo prefiero llamarle perfil disciplinar del campo.

Entonces los datos vienen de la WoS, solo para artículos y hasta el año 2020. Estos datos han sido analizados con Pajek y la identificación de clústeres con el Método de Louvain, y finalmente la visualización con VOSviewer. El resultado se expone en la siguiente página, pero antes es pertinente mirar las principales disciplinas donde se han publicado los artículos.

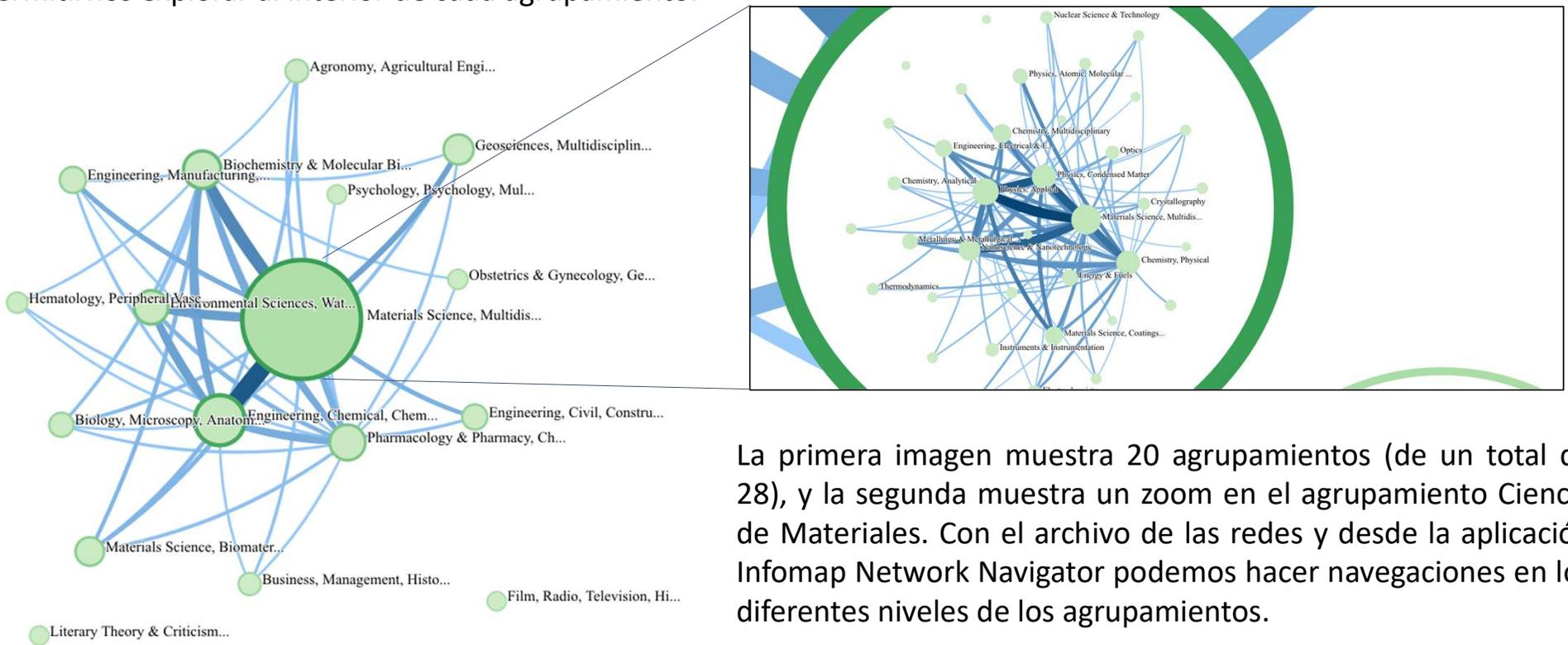
La Tabla siguiente muestra las principales 10 categorías de disciplinas donde se han publicado los artículos en nano en México. Una cuarta parte de la producción científica es en Ciencias de los Materiales, con menor presencia esta Física Aplicada y Físico-química Física de la Materia Condensada.

De las disciplinas mostradas en esta tabla, predomina el aporte de la física y la química, ambas son áreas con una tradición consolidada en México.

Categorías WoS	Artículos	% de 26858
Materials Science, Multidisciplinary	8529	31,76%
Physics, Applied	5029	18,72%
Chemistry, Physical	4394	16,36%
Physics, Condensed Matter	3682	13,71%
Nanoscience & Nanotechnology	2560	9,53%
Chemistry, Multidisciplinary	2072	7,71%
Engineering, Chemical	1524	5,67%
Polymer Science	1339	4,99%
Materials Science, Coatings & Films	1296	4,83%
Electrochemistry	1231	4,58%

Perfil disciplinar del campo nano (science map)

Otra herramienta interesante de usar para el análisis de este tipo de datos es MapEquation, cuyo método es bastante estable y permite hacer análisis multinivel de redes. El análisis de clustering es diferente al anterior, pues además de estar basado en el enfoque de módulos, va más lejos al identificar niveles en la estructura de las redes, así como permitimos explorar al interior de cada agrupamiento.



La primera imagen muestra 20 agrupamientos (de un total de 28), y la segunda muestra un zoom en el agrupamiento Ciencia de Materiales. Con el archivo de las redes y desde la aplicación Infomap Network Navigator podemos hacer navegaciones en los diferentes niveles de los agrupamientos.

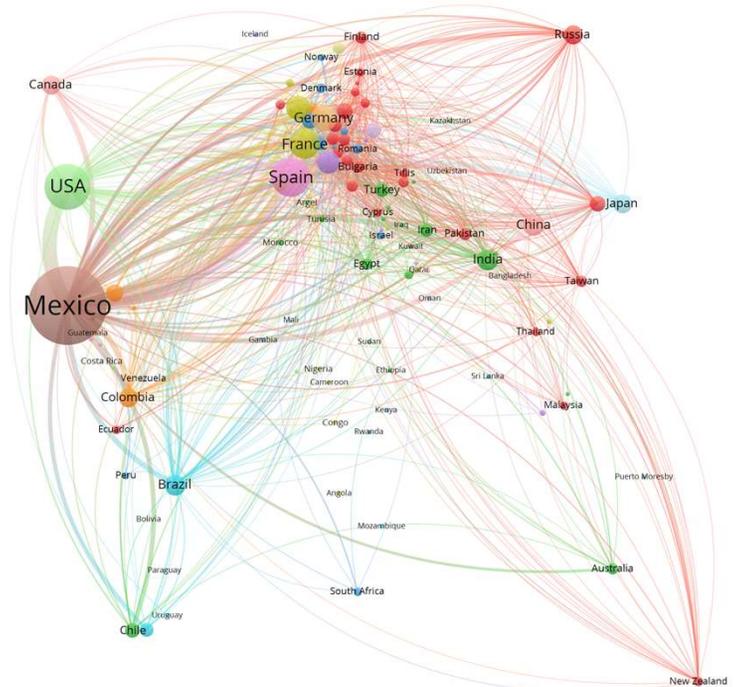
La distribución institucional es otro indicador muy común en los estudios bibliométricos y cuantitativos. Conocer esta distribución nos permite identificar donde se encuentran instaladas las capacidades científicas y tecnológicas, así como la concentración de estas capacidades. Por otro lado, si la estructura de los datos permite hacer cruces con otra información, se pueden pensar indicadores que permitan conocer a partir de la distribución institucional, y espacial, en términos de las disciplinas y temáticas.

Sin embargo, los datos, que las bases de indexación nos permiten descargar, no están normalizados, lo cual dificulta esta tarea porque el trabajo de normalización de las instituciones requiere mucho tiempo. La estrategia que nos queda es hacer las consultas a partir de las instituciones desde el motor de búsqueda de la Web of Science. Entonces, necesitamos formular las preguntas y, a partir de éstas, hacer nuevas consultas de información por institución. Por el momento, lo que tenemos es la siguiente tabla que se limita a exponer la frecuencia de las instituciones de adscripción de los autores en el campo nano en México.

Top	Institución	Artículos	% de 26858
1	UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO	9168	34,14%
2	INSTITUTO POLITECNICO NACIONAL MEXICO	4948	18,42%
3	CINVESTAV CENTRO DE INVESTIGACION Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITECNICO NACIONAL	3355	12,49%
4	UNIVERSIDAD AUTONOMA METROPOLITANA MEXICO	2075	7,73%
5	BENEMERITA UNIVERSIDAD AUTONOMA DE PUEBLA	1515	5,64%
6	UNIVERSIDAD AUTONOMA DE NUEVO LEON	1360	5,06%
7	UNIVERSIDAD AUTONOMA DE SAN LUIS POTOSI	1334	4,97%
8	INSTITUTO MEXICANO DEL PETROLEO	972	3,62%
9	UNIVERSIDAD AUTONOMA DEL ESTADO DE MORELOS	953	3,55%
10	CIMAV CENTRO DE INVESTIGACION EN MATERIALES AVANZADOS S C (901)	901	3,35%

Colaboraciones internacionales

- En cuanto a las colaboraciones internacionales hasta el año 2020, se ha colaborado con un total de 116 países diferentes en el campo de las nanociencias y nanotecnologías.
- Seguramente estas colaboraciones han aumentado, aunque es necesario hacer una comparación entre las colaboraciones nacionales y las internacionales para mejor observar la evolución de ambas (punto pendiente por realizar).
- En general podemos ver en la red de esta lamina que siguen predominando las colaboraciones con USA y Europa.
- A partir de la ultima publicación sobre estas colaboraciones, vemos que en estos 10 últimos años las colaboraciones con USA han disminuido, mientras que con Europa el aumento es más importante, así como con Colombia.
- Los colores de los agrupamientos (análisis de clustering) esta realizado con VOSviewer, pero se pueden hacer otros análisis.
- Podría ser interesante hacer un análisis más detallado sobre las colaboraciones por continente y ver sus trayectorias y dinámicas en el tiempo.



Hasta 2009

Combination	% of 11757 (%)
USA	14.50
Spain	5.88
France	4.49
Germany	3.46
England	3.05
Cuba	2.16
Canada	1.82
Italy	1.81
Russia	1.76
Japan	1.67

Hasta 2020

Top País	Artículos	% de 26858
1 USA	3504	13,05%
2 SPAIN	2317	8,63%
3 FRANCE	1194	4,45%
4 GERMANY	700	2,61%
5 ENGLAND	558	2,08%
6 COLOMBIA	545	2,03%
7 ITALY	514	1,91%
8 CUBA	498	1,85%
9 CANADA	493	1,84%
10 INDIA	480	1,79%



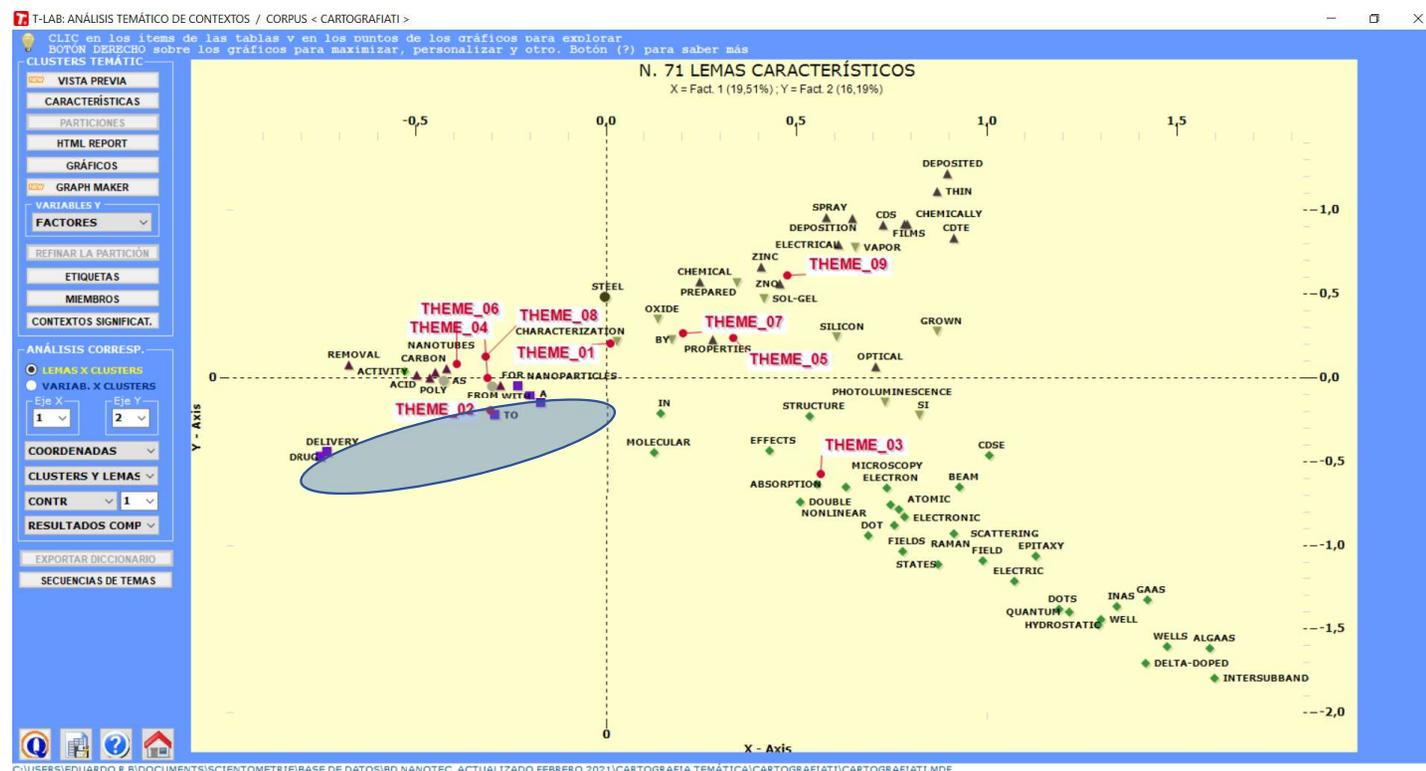
El último indicador que les propongo de explorar es lo que llamamos cartografía de temáticas, las cuales nos muestran un panorama de las temáticas que se han desarrollado en un periodo. Esto va un poco más lejos que algunos estudios bibliométricos ya que se trata de un análisis de contenido aplicado a un corpus (en este caso los títulos de los artículos en nano). Hay diferentes enfoques metodológicos y teóricos: minería de texto, análisis de contenido, y análisis semántico. Las herramientas que usamos en el DMMSS del IIMAS son dos: WordMapper y T-Lab. Hay una tercera que es más potente, pero no tengo acceso por ahora (se trata de Alceste).

Este tipo de visualizaciones me interesa mostrarles ahora pues son de mucha ayuda para la etapa exploratoria de los datos. La idea es tener un primer panorama general de lo que se ha desarrollado en las nanociencias y nanotecnologías en México, y partiendo de esos resultados veamos que preguntas nuevas resultan y hasta dónde estos análisis nos permiten abordarlas.

Dos posibilidades se me ocurren. La primera es dividir el periodo de análisis o hacer el análisis por región, área del conocimiento y otro dato que podamos cruzar con el campo de los títulos de los artículos (el cual es nuestro campo de información de entrada, pero también podemos usar las palabras clave y el resumen, que este último requiere un trabajo importante de preparación del corpus).

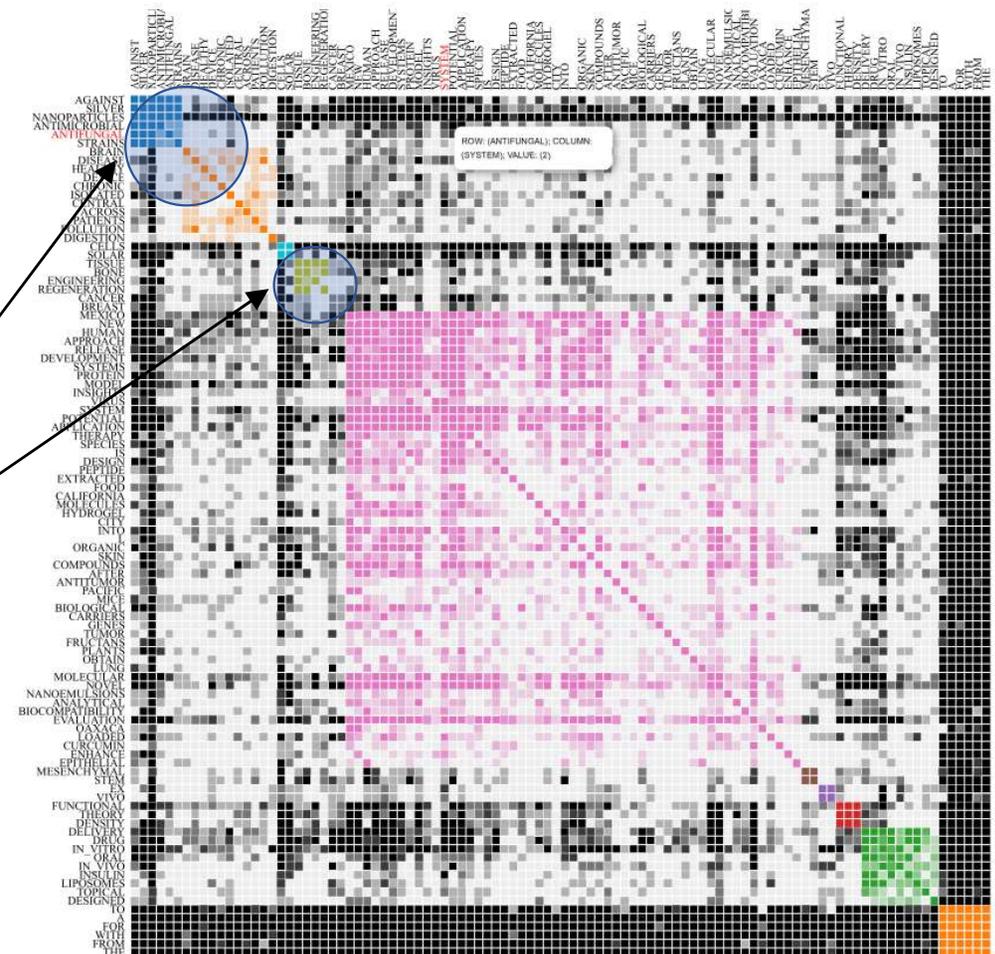
- T-Lab: análisis lingüista, estadístico y gráfico para texto. En este caso aplico análisis temático, pero hay otros análisis complementarios en esta herramienta.
- En la gráfica siguiente se muestra el análisis de los títulos de los 26858 documentos. El resultado arrojó un total de 9 clústeres, los cuales podemos considerar como ejes temáticos.

- En los 9 clústeres encontramos palabras claves que describe cada eje temático. Analizar la cartografía sin el programa es difícil ya que no podemos hacer acercamientos y ver a detalle lo que hay en cada clúster.
- Lo anterior es un límite importante y debemos usar tablas de adyacencia y matrices para poder analizar los resultados.
- Enseguida hacemos un acercamiento (zoom) sobre el clúster 2.

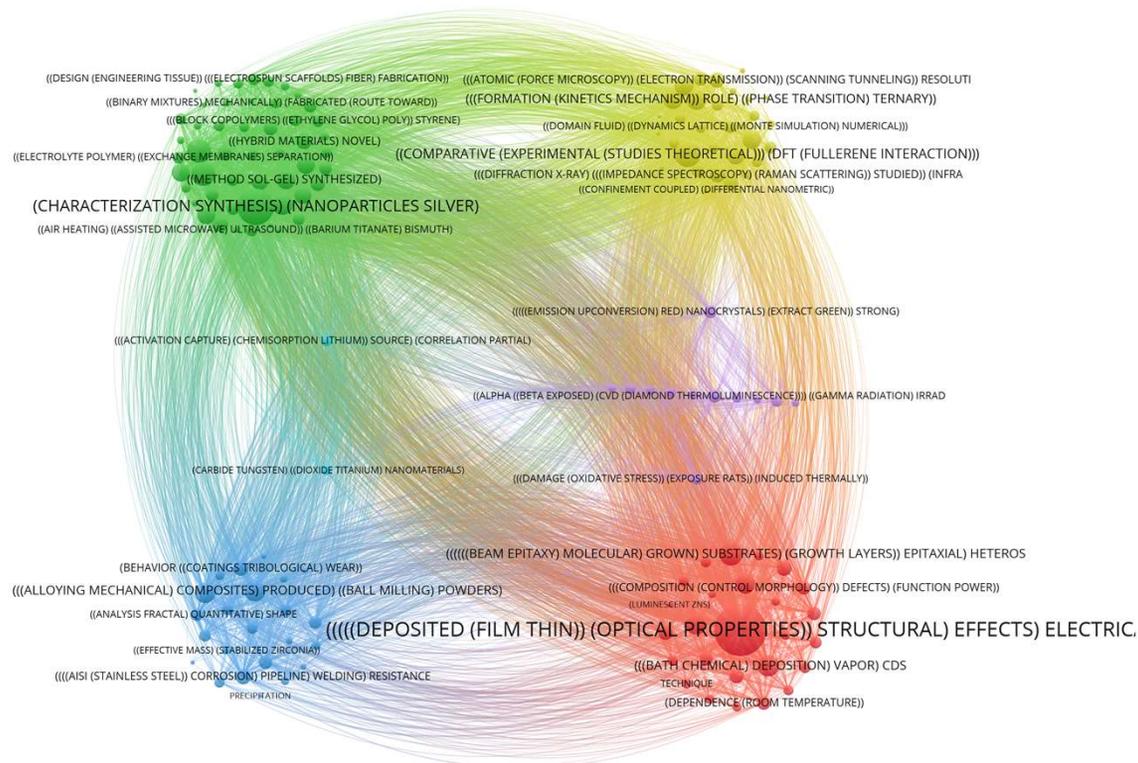


Ejemplo con el clúster 2:

- Dentro de este eje temático encontramos 9 temáticas, las cuales se muestran en la matriz de adyacencia de esta lamina.
- Por ejemplo, la primera (color azul) concierne al desarrollo de aplicaciones para las nanopartículas de plata basadas en sus potencialidades antibacteriales y fungicidas.
- Otro ejemplo es la de color kaki que evoca trabajos sobre regeneración celular para la ingeniería de tejidos.
- Está gráfica contiene bastante información y para leerla es necesario el archivo que les envío en documento adjunto y podrán abrir en cualquier explorador web.
- Para que miren a detalle lo que hay en cada una de los 9 ejes temáticos, les envío en un archivo .zip las matrices de adyacencia.



En la cartografía anterior se visualizan las 146 temáticas identificadas por WordMapper, sin embargo, su lectura no es sencilla por el alto número de temáticas. Para complementar, y facilitar la lectura, a la visualización de WordMapper podemos exportar el archivo de la cartografía de temáticas en formato de archivo para análisis de redes y aplicar otro tipo de análisis a los datos. En este caso lo que hacemos es obtener el archivo .net de los clústeres y analizarlo con Pajek, buscar una distribución que mejore la legibilidad de los datos, para finalmente visualizarlo con VOSviewer.



El resultado de esta nueva visualización nos da 6 ejes temáticos, donde están clasificados los 146 temas identificados con WordMapper y 10455 relaciones entre los temas y un total de 1818400 relaciones.

Esta reinterpretación de los temas muestra ejes temáticos, cuyo análisis requiere ser minucioso con la ayuda de la herramienta VOSviewer (tengo los archivos disponibles por si alguien quiere mirarlos).

- El panorama de este documento es muy general. El objetivo de hacerlo ha sido mostrarles algunas de las cosas que podemos hacer y encuentro interesante y les pueden servir al grupo para sus preguntas de investigación particulares. Para esto, les pido me hagan sus peticiones específicas y vamos trabajando sus solicitudes. Si creen pertinente podemos agendar, para el mes de agosto, un taller corto donde les muestre a detalle los análisis.
- La visualización basada en el Análisis de Redes Sociales y de Clustering muestra las relaciones entre las disciplinas, y a partir de estas relaciones se puede establecer clústeres o agrupamientos de disciplinas relacionadas en la producción de nuevos conocimientos en el campo nano.
- Sobre el perfil disciplinar, he mostrado dos herramientas basadas en el análisis de redes sociales. La primera nos permite tener una perspectiva muy general de las relaciones entre las disciplinas en el desarrollo de las nano en México. La segunda herramienta, método MapEquation, nos permite tener perspectivas más detalladas estas relaciones entre disciplinas. En ambos resultados observamos la centralidad de la Física y la Química, sin embargo, con el método MapEquation tenemos una distribución de los agrupamientos que nos permite observar la búsqueda de aplicaciones de las nano en agronomía, medicina, ingeniarías, etc.