

The background of the slide is a complex network diagram. It consists of numerous nodes of varying sizes and colors (dark blue, light blue, and grey) connected by thin, light grey lines. Some nodes are highlighted with larger, concentric circles. The overall aesthetic is modern and technical, suggesting themes of data science, networking, or systems analysis.

TOPIC MODELLING

Laura J. Mundaca Retuerto

CONTEXTO

CONOCIMIENTOS
VALIOSOS EN
TIEMPOS CORTOS.

QUÉ HABLA LA
GENTE, SUS
PROBLEMAS Y
NECESIDADES.

GRAN CANTIDAD
DE DATOS
MAYORMENTE NO
ESTRUCTURADOS.

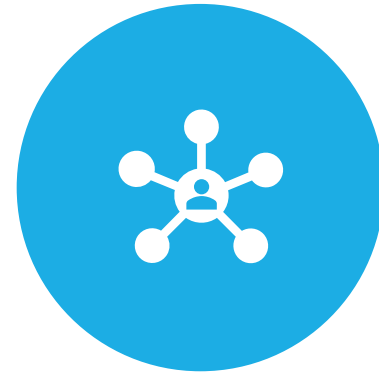
QUÉ ES TOPIC MODELLING



TÉCNICA DE APRENDIZAJE NO
SUPERVISADO

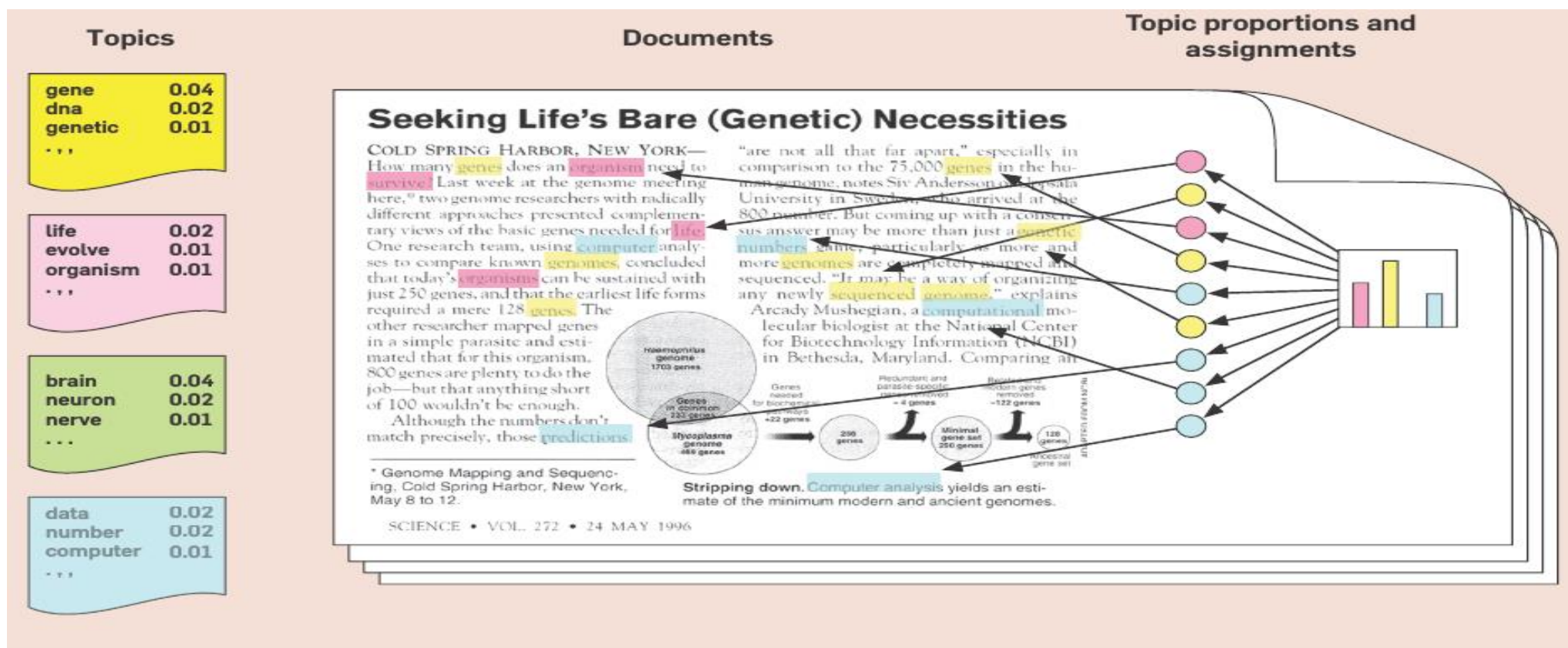


ANALIZAR LOS DATOS



AGRUPA PALABRAS SIMILARES
EN BASE A PATRONES

TOPIC MODELLING - LDA



TOPIC MODELING - LDA

TRATA CADA DOCUMENTO COMO UNA MEZCLA DE TEMAS Y CADA TEMA COMO UNA MEZCLA DE PALABRAS.



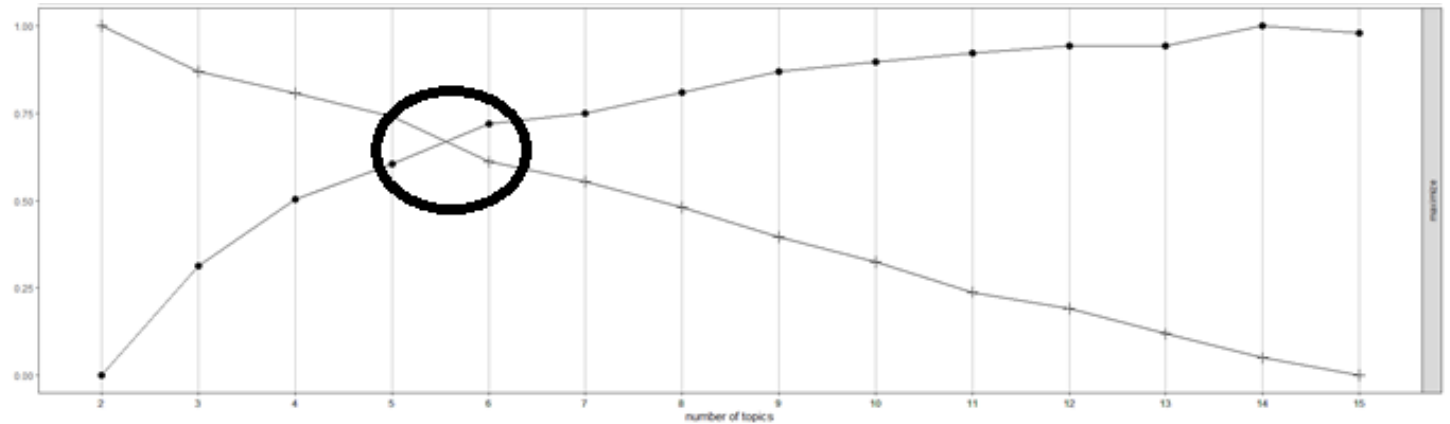
MEZCLA DE TEMAS: DOC1 60% TEMA A 25%
TEMA B 15% TEMA C



MEZCLA DE PALABRAS: LAS PALABRAS SE PUEDEN COMPARTIR EN VARIOS O TODOS LOS TEMAS.

EJEMPLO: ANÁLISIS DE NOTICIAS PERUANAS DE 11 SET AL 13 SET

La gráfica sugiere que pueden ser 5 ó 6 tópicos los encontrados en el conjunto de noticias.



TÓPICOS HALLADOS

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
[1,]	"covid"	"presidente"	"covid"	"nacional"	"economía"	"fiscal"
[2,]	"salud"	"vizcarra"	"millones"	"alcalde"	"ministra"	"nacional"
[3,]	"gobierno"	"congreso"	"salud"	"josé"	"censura"	"elecciones"
[4,]	"nacional"	"martín"	"personas"	"fiscalía"	"alva"	"caso"
[5,]	"pandemia"	"comisión"	"tratamiento"	"lima"	"maría"	"primera"
[6,]	"coronavirus"	"karen"	"casos"	"salud"	"antonietta"	"electoral"
[7,]	"país"	"richard"	"coronavirus"	"informó"	"moción"	"blancos"
[8,]	"casos"	"estado"	"gobierno"	"investigación"	"bancadas"	"cuellos"
[9,]	"ministerio"	"audios"	"país"	"caso"	"podemos"	"ministerio"
[10,]	"vacuna"	"fiscalización"	"mil"	"victoria"	"congresista"	"especial"

• Las 10 palabras concentradas nos indica que la naturaleza de este tópico está basada en las medidas que toma el gobierno para el manejo de la pandemia en una coyuntura económica afectada, donde en esa época, emergían cuestionamientos de la ministra de economía, bajo acusaciones de mal manejo del ministerio.

Tópico 1

• Esta basada en la crisis política del ejecutivo concerniente a los cuestionamientos de corrupción por la contratación del personaje "Richard" (swing) y las fricciones del equipo personal de asesores del presidente.

Tópico 2

• Esta basada en las políticas públicas sanitarias dada por el ejecutivo para el manejo del coronavirus.

Tópico 3

• Las denuncias de corrupción por malos manejos durante la pandemia alertaron a la fiscalía y ministerio público, destacándose en los medios.

Tópico 4

• El cuestionamiento a la ministra de economía de aquel entonces, se ve representado por una censura que grupos políticos promovían desde el congreso.

Tópico 5

• Pese a la coyuntura de pandemia, los medios aún tocaban el tema de corrupción con el caso de cuellos blancos.

Tópico 6

PALABRAS MÁS FRECUENTES EN CADA TÓPICO

