Dr. Pierre-Nicolas Schwab

pn@intotheminds.com - https://www.linkedin.com/in/pnschwab/

Investigación sobre los factores que afectan a la viralidad de las publicaciones en LinkedIn

Junio 2021

Índice

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu’un hyperlien soit fait vers le site d’IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

Resumen

Resumen

La investigación basada sobre 4.599 millones de publicaciones en LinkedIn en 193 países determinó que sólo 3 factores (de los 6 considerados) tienen un impacto estadísticamente significativo en la viralidad del contenido:

- El tamaño de la red
- El número de palabras
- El número de emojis

Sin embargo, el impacto de estos 3 factores es muy diferente. El tamaño de la red representa el 33,99% de los efectos, el número de palabras el 19,4% y el número de emojis sólo el 2,4%.

El modelo matemático también muestra que los hashtags, el idioma del post o el país de residencia no tienen un efecto significativo.

Resumen de los resultados

- **Sólo 3 factores tienen una influencia estadísticamente significativa** en la visibilidad de las publicaciones de Linkedin: el número de personas en su red, el número de palabras y el número de emojis
- Algunas bvriables como los hashtags, el idioma del post o el país no ejercen una influencia significativa en la visibilidad de los posts de Linkedin
- El **tamaño de su red** representa el 34% de la visibilidad de sus publicaciones en LinkedIn
- El **número de palabras** utilizadas en sus publicaciones de LinkedIn determina el 19,1% de su visibilidad
- **Los emojis** sólo el 2,4%
- **El 51,04% de los contenidos son puestos en línea por usuarios con menos de 1442 conexiones,** sólo el 7% por los que tienen más de 10000 y el 1,3% más de 20000
- **El 80,3% de las publicaciones** publicadas en LinkedIn tienen menos de 92 palabras
- **El 2,2% de las publicaciones** publicadas en LinkedIn tienen más de 200 palabras
- **El 80,12%** de las publicaciones en LinkedIn no contienen emojis

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

Resultados

1. Introducción

Para las empresas, la comprensión de los factores que influyen en la visibilidad en las redes sociales se ha convertido en un reto importante. El rápido aumento de la cantidad de contenidos ha impuesto algoritmos de recomendación que deciden la visibilidad de los contenidos. Sin embargo, el funcionamiento de estos algoritmos es opaco y todavía no se conocen con certeza las reglas que los rigen. Por ello, es necesario un esfuerzo de investigación para observar, desde fuera, la dinámica de publicación de contenidos y deducir los factores determinantes de su éxito.

Esta investigación pretende responder a estas preguntas para una red social especializada en intercambios profesionales: LinkedIn. Con 756 millones de usuarios (en 2021), LinkedIn fue adquirida por Microsoft en 2016. Se estima que 3 millones de usuarios están activos cada semana y, en 2019, se alcanzaron 358.000 millones de actualizaciones del feed de LinkedIn.

La actualización del algoritmo de LinkedIn de 2020 condiciona la recomendación de las publicaciones (y por tanto su visibilidad dentro de la red) al interés que generan. Este interés se mide a través del "dwell time", es decir, el tiempo de interacción con un contenido concreto.

Sin ser recomendado por el algoritmo, el contenido de LinkedIn tiene pocas posibilidades de ser visible. Por lo tanto, es esencial producir contenido que "guste" y se comente para enviar una señal al algoritmo, para que recomiende el contenido a los demás. Sin embargo, sólo el 3,59% del contenido en LinkedIn obtiene más de 100 reacciones (“me gusta” y comentarios).

SOLO EL 3,59% DE LOS CONTENIDOS EN LINKEDIN ALCANZAN LAS 100 REACCIONES

Existen muchas teorías sobre lo que determina el "éxito" de una publicación en LinkedIn. Para objetivar los factores que realmente juegan un papel importante hemos utilizado un conjunto de datos de 4.599 millones de posts de LinkedIn en 193 países. Los datos fueron proporcionados por Linkalyze y las herramientas de ciencia de datos por TIMi.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu’un hyperlien soit fait vers le site d’IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

2. La dinámica de los comentarios y de los “me gusta”

En esta primera parte, encontrará estadísticas sobre el compromiso en LinkedIn. Se propone una nueva lectura, por país y por idioma, que pone de manifiesto las diferencias esenciales.

¿Cuál es el número medio de "Me gusta" y de "Comentarios" por cada publicación en LinkedIn?

El número medio de “me gusta” y comentarios varía en función del idioma. En nuestro conjunto de datos, pudimos encontrar publicaciones de LinkedIn en 15 idiomas. A continuación, encontrará una imagen que muestra los resultados. El checo fue el idioma menos representado, con 6.256 publicaciones, y el inglés el más representado, con 2.556.617 publicaciones.

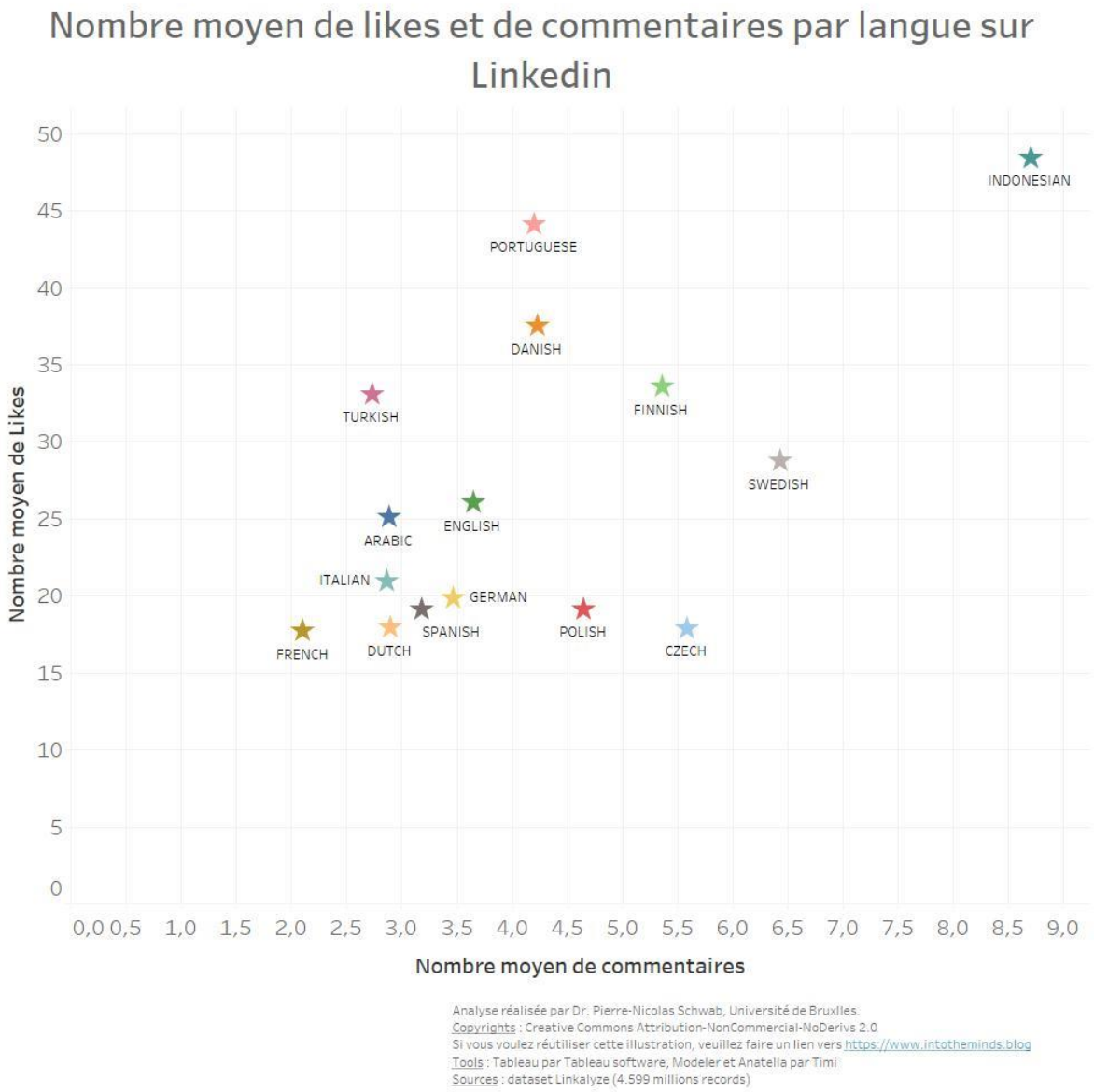


Imagen 1: Número medio de comentarios y de "me gusta" por publicación en LinkedIn, por idioma

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu’un hyperlien soit fait vers le site d’IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

El francés es el idioma en el que las publicaciones de LinkedIn obtienen una menor cantidad de “me gusta”: 2,1 comentarios y 17,8 likes. El indonesio está en el otro extremo del espectro, con 8,7 comentarios y 48,5 me gusta.

Distribución del número de comentarios en LinkedIn según el idioma

El análisis de la distribución de los comentarios por idioma ofrece una lectura diferente del fenómeno. Independientemente del idioma, la mayoría de los posts de LinkedIn no reciben comentarios. Pero esta realidad se contrasta como se muestra en la Imagen 2 (en rojo, la proporción de posts de LinkedIn que no reciben comentarios; en verde, la proporción que recibe más de 20 comentarios).

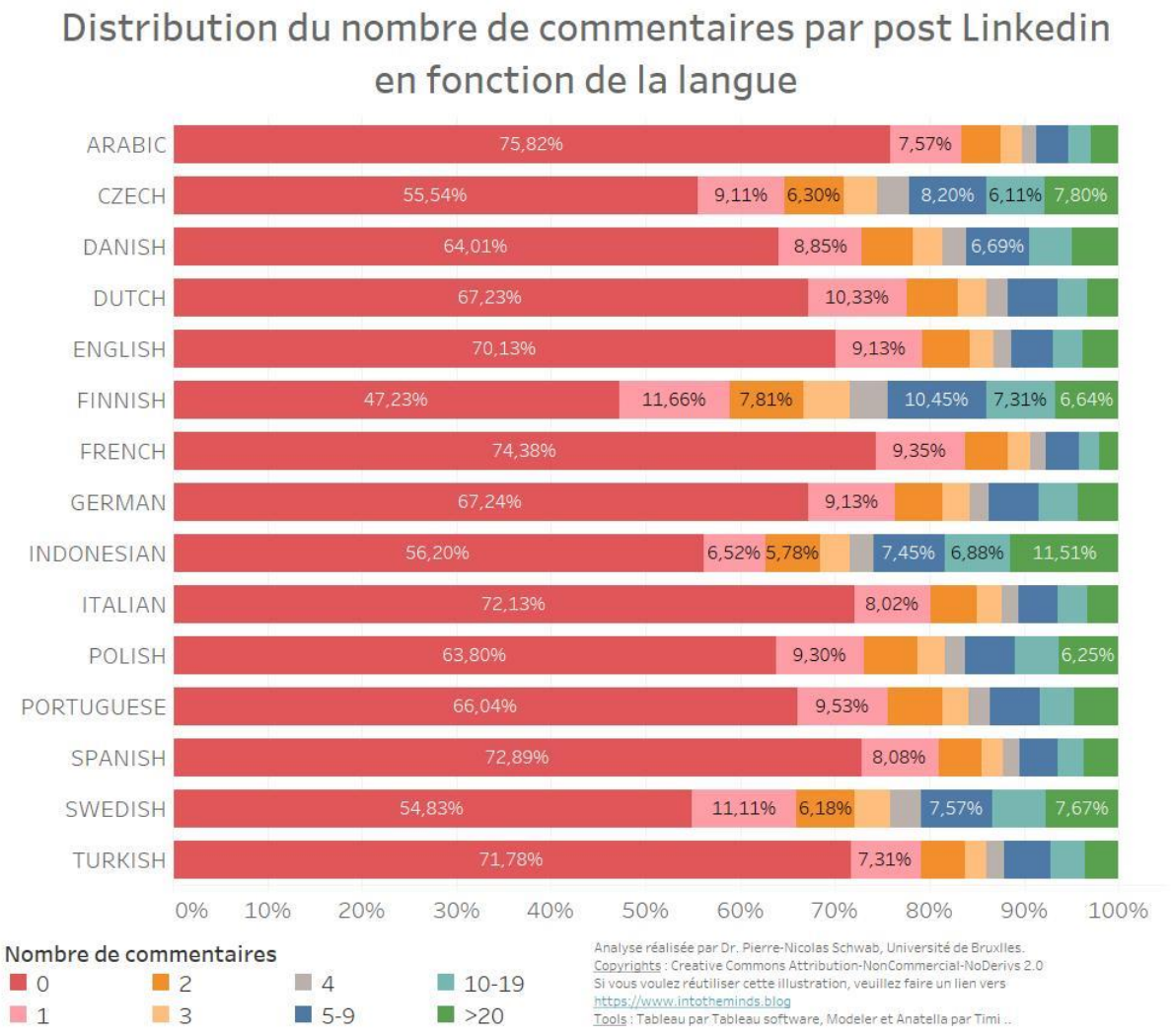


Imagen 2: Distribución de los comentarios en las publicaciones de LinkedIn según el idioma.

Se observan 30 puntos diferentes entre el finlandés, el checo y el sueco, por un lado, y el francés, el inglés o el español, por otro.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu’un hyperlien soit fait vers le site d’IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

Distribución por idiomas del número de “me gusta” en LinkedIn

Las disparidades por idioma también se observan en el caso de los “me gusta”. La diferencia es notable en particular entre el árabe y el finlandés. En inglés, francés y español, casi 1/4 de las publicaciones no obtienen ningún "me gusta", proporción que sólo es superada por las publicaciones en árabe (33,16%).

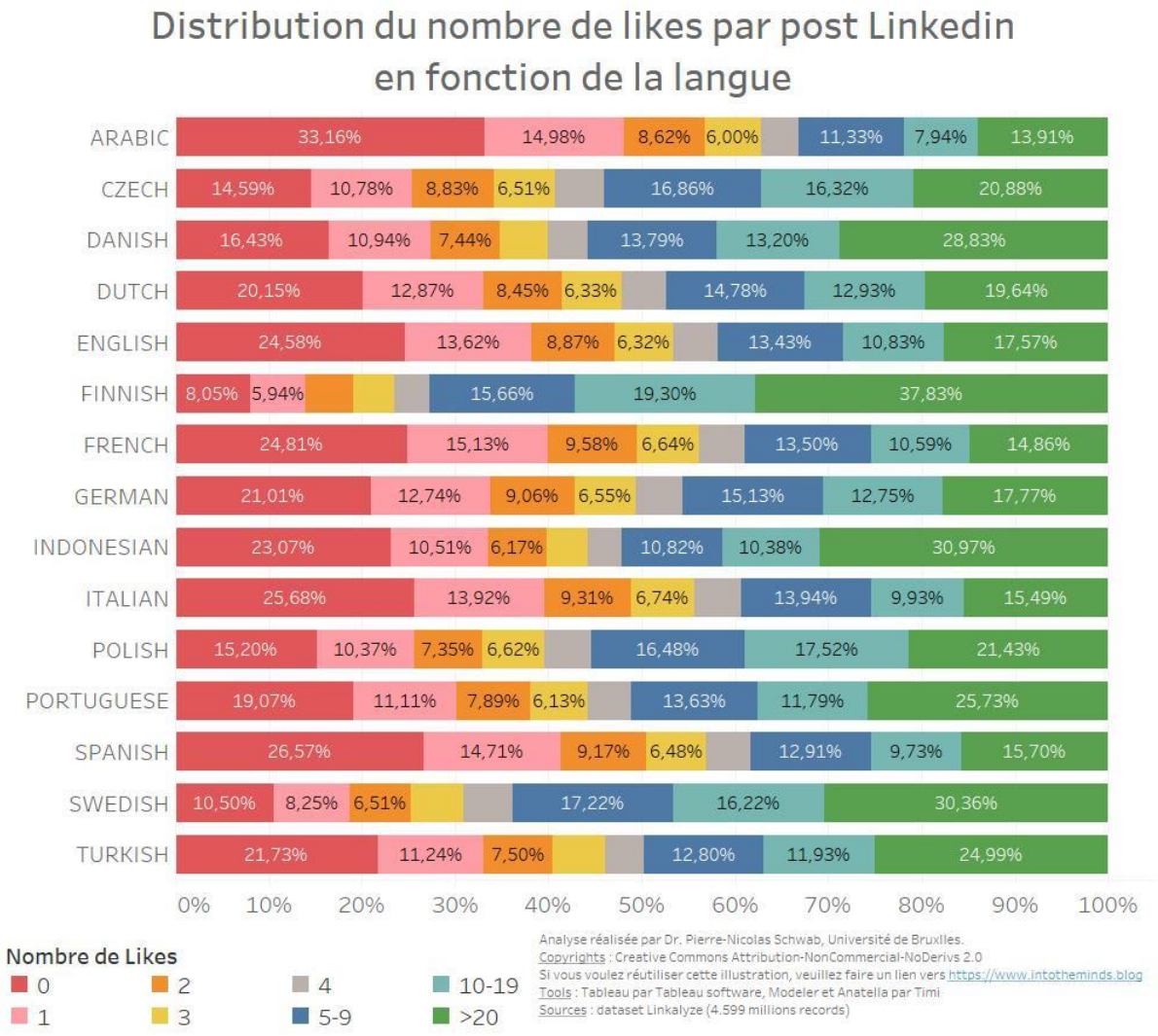


Imagen 3: Distribución del número de “me gusta” según el idioma

¿En qué países se consiguen más “me gusta” en LinkedIn?

No todos los países son iguales en términos de “me gusta” en LinkedIn. Podemos comprobarlo con el análisis de las publicaciones de 193 países.

Para descartar los datos extremos de algunos microterritorios, los datos se representaron como cuartiles. En otras palabras, los datos se normalizaron y luego se asignaron a un segmento en función de su valor.

Esta visualización por cuartiles es útil cuando determinados valores extremos (outliers) "borran" las posibles diferencias.

El mapa que se muestra a continuación facilita la visualización del cuartil al que pertenece cada país y la toma de conciencia de las diferencias entre territorios en términos de “me gusta”.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu’un hyperlien soit fait vers le site d’IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

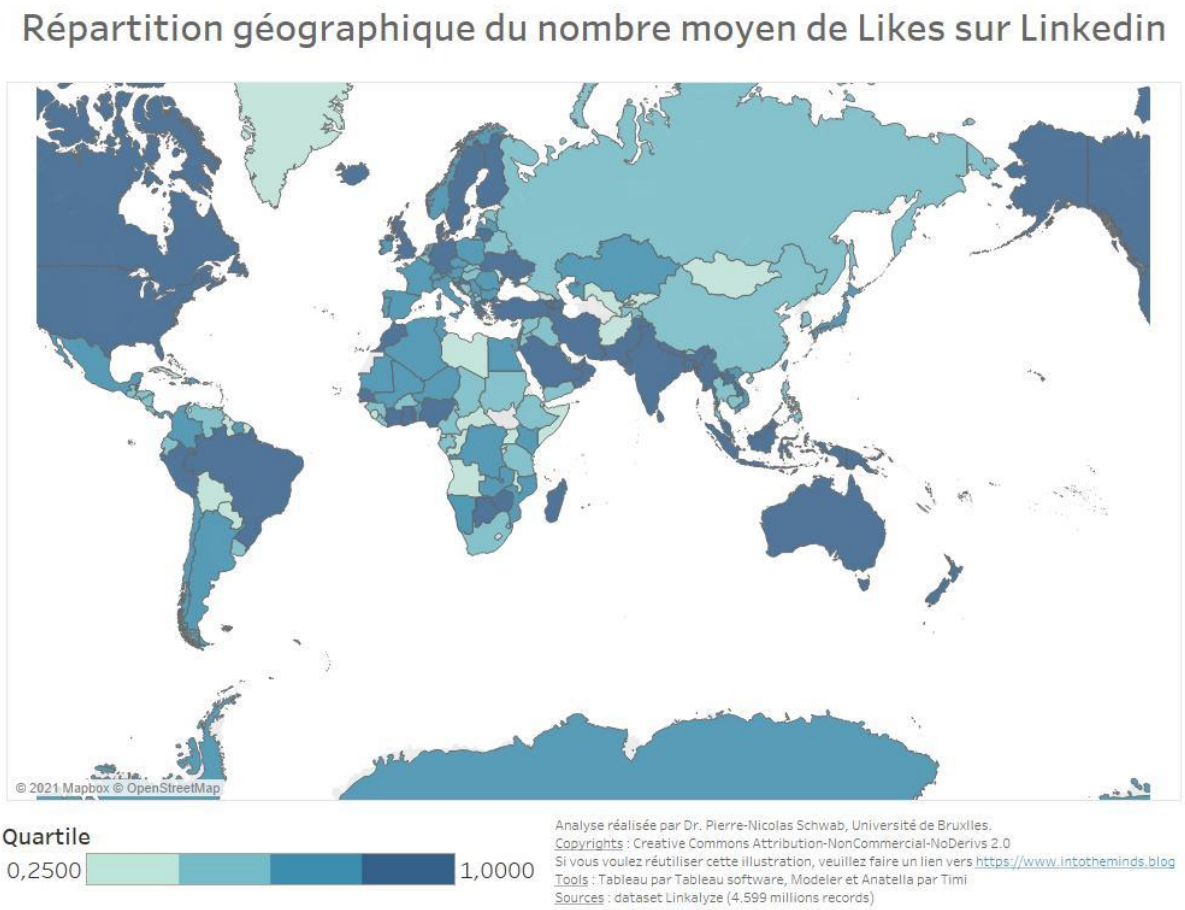


Imagen 4: Distribución del número de likes por cuartil y por país

¿En qué país se comentan más los posts de LinkedIn de media?

Hemos repetido el mismo análisis teniendo en cuenta el número medio de comentarios en LinkedIn. Como se puede ver (Figura 5), los dos mapa presentan algunas diferencias. Con una media de 3,3 comentarios en cada publicación de LinkedIn, Rusia, por ejemplo, se encuentra en el cuartil superior, mientras que está en el segundo cuartil de los "me gusta" (10 "me gusta" de media por publicación).

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

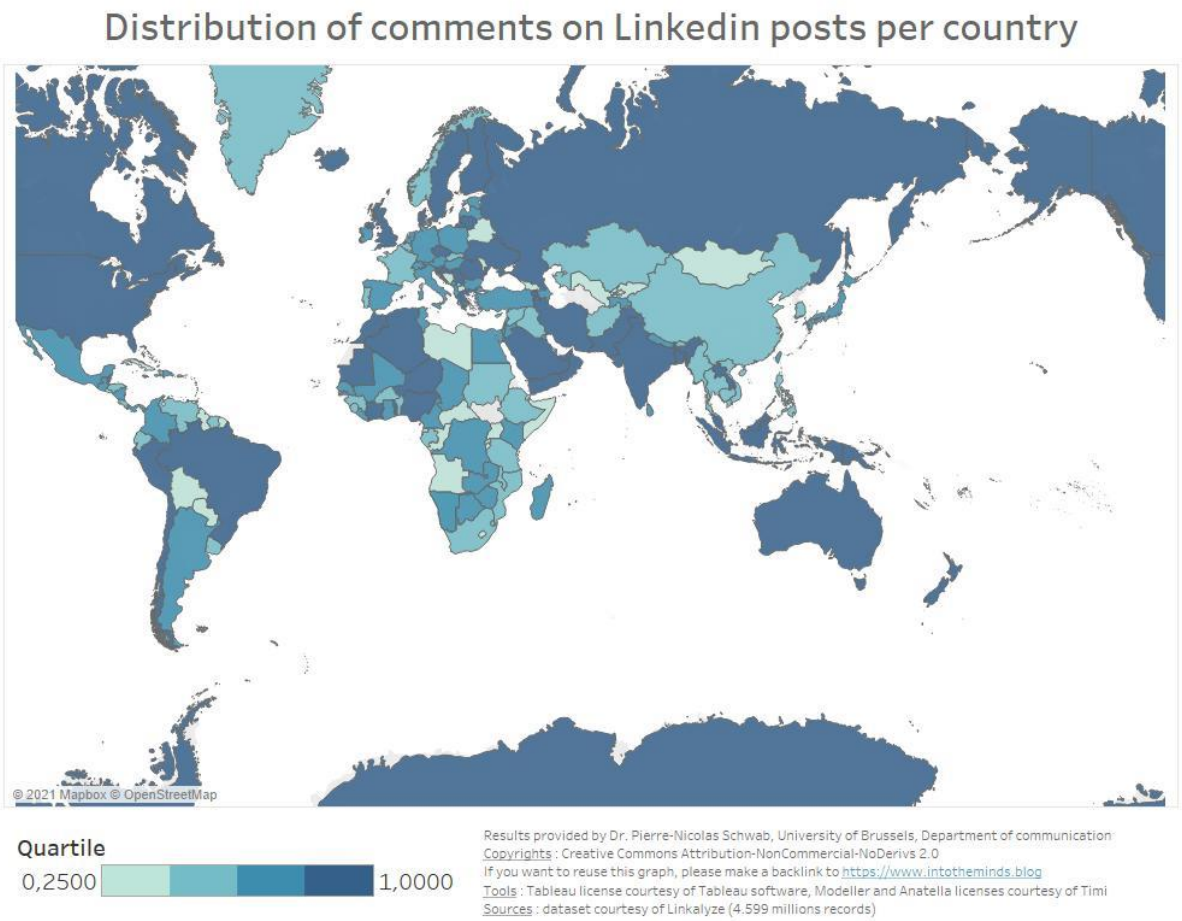


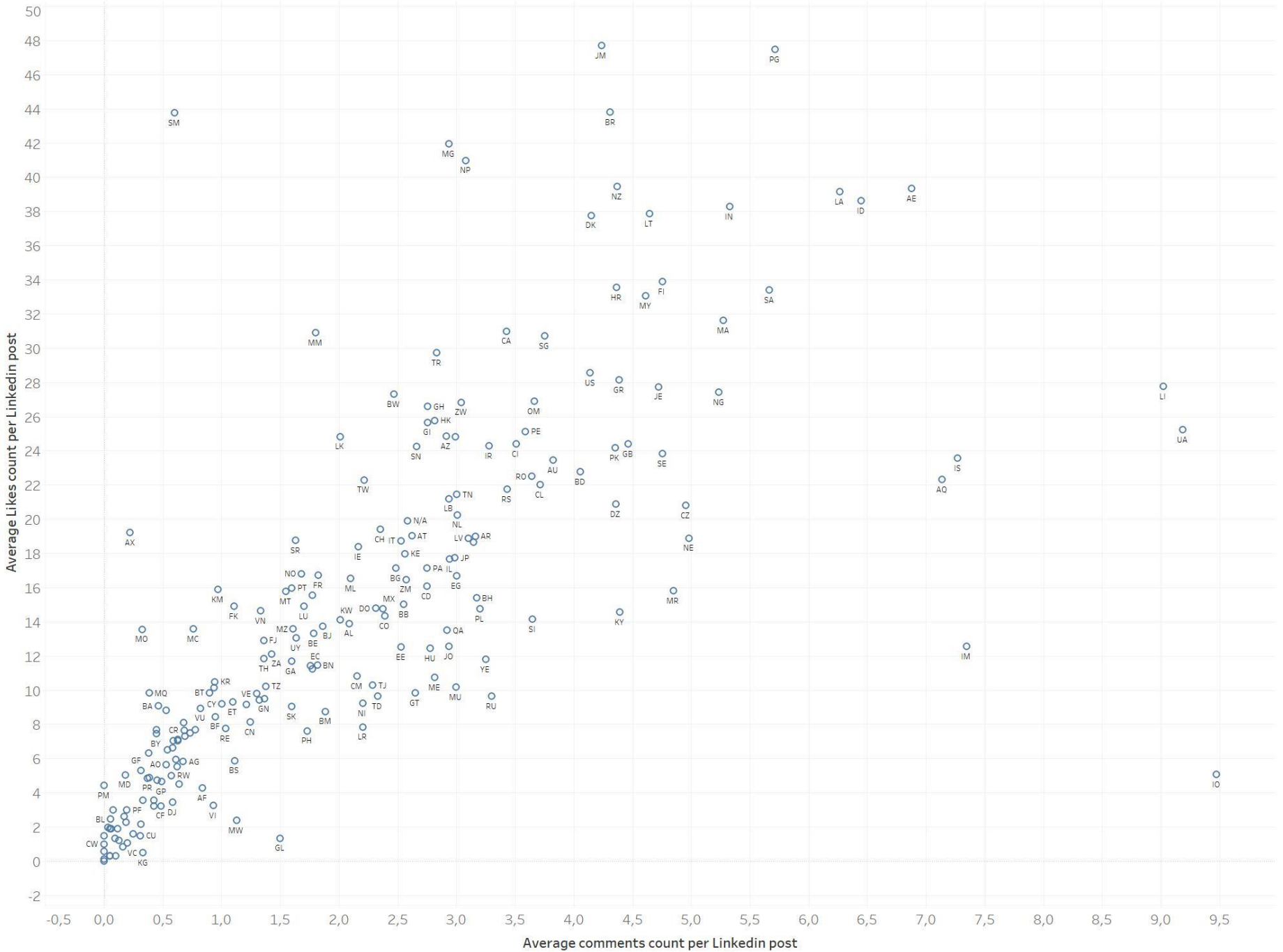
Imagen 5: Distribución del número de comentarios por cuartil y por país

Estadísticas sobre los comentarios y los "me gusta" por país

Por último, propongo visualizar en un "gráfico de dispersión" la posición de cada país respecto al número medio de "me gusta" y comentarios (Figura 6). Se han eliminado algunos valores extremos (Islas Vírgenes, Tonga, Belice) para que el gráfico siga siendo legible.
Para cada país del mundo, es posible acceder a las estadísticas de las pubicaciones de LinkedIn publicadas allí. El código ISO indica los países.

Página siguiente: Imagen 6. Número medio de comentarios y “me gusta” por publicación de LinkedIn por país

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu’un hyperlien soit fait vers le site d’IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.


Average Likes count per Linkedin post
Average comments count per Linkedin post
50
48
46
44
42
40
38
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
-2
-0,5
0,0
0,5
1,0
1,5
2,0
2,5
3,0
3,5
4,0
4,5
5,0
5,5
6,0
6,5
7,0
7,5
8,0
8,5
9,0
9,5
JM
PG
SM
BR
MG
NP
NZ
LA
ID
AE
DK
LT
IN
HR
MY
FI
SA
MA
MM
CA
SG
TR
US
GR
JE
NG
LI
BW
GH
HK
ZW
OM
GI
PE
LK
AZ
IR
CI
SN
PK
GB
SE
UA
AU
IS
RO
BD
TW
RS
CL
AQ
TN
DZ
CZ
LB
N/A
NL
AX
CH
IT
AT
LV
AR
NE
SR
IE
KE
JP
PA
IL
NO
BG
ZM
EG
FR
ML
CD
BH
KM
PT
MT
MX
MR
FK
LU
VN
DO
BB
PL
KW
CO
KY
MO
MC
MZ
AL
QA
SI
FJ
BE
BJ
UY
EC
EE
HU
JO
IM
TH
ZA
GA
BN
YE
KR
CM
TJ
ME
MQ
BT
TZ
GT
MU
RU
BA
CY
VE
GN
NI
TD
VU
ET
SK
BM
BF
CN
LR
RE
PH
CR
BY
GF
AO
AG
BS
RW
PM
MD
PR
GP
AF
PF
DJ
VI
BL
CF
MW
CU
GL
CW
VC
KG
IO

3. Resultados de la modelización predictiva

Hemos estudiamos la influencia de 6 variables en el número de reacciones recibidas. La modelización matemática muestra que sólo 3 variables tienen un impacto estadísticamente significativo en la probabilidad de obtener al menos 100 reacciones (me gusta y comentarios) en una publicación de LinkedIn:

- El tamaño de la red del autor del post
- El número de palabras del post
- El número de emojis que contiene el post

Las otras 3 variables consideradas en el modelo (número de hashtags, idioma, país) no tuvieron un impacto significativo.

a) Efectos relativos de las 3 variables significativas

Las 3 variables (tamaño de la red, número de palabras, número de emojis) no influyen en la obtención de 100 o más reacciones de igual manera. Por ejemplo, los emojis sólo contribuyen en un 2,4%, mientras que el tamaño de la red explica el 33,99% del resultado obtenido. El número de palabras explica el 19,1%.

Estas diferencias se aprecian mejor en la figura 7. En forma de áreas, esta visualización representa el "impacto" relativo de las distintas variables en el resultado.

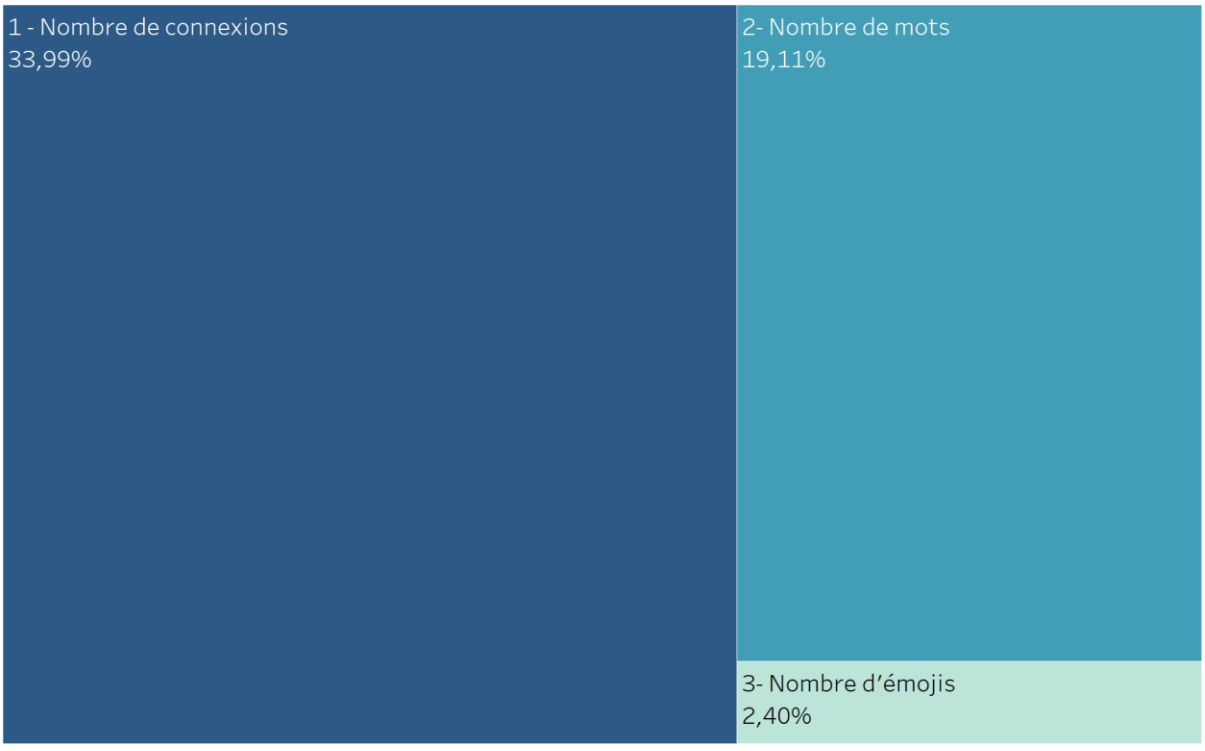


Imagen 7. Visualización de los efectos relativos de las 3 variables significativas sobre la probabilidad de recoger al menos 100 reacciones en un post de LinkedIn.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

b) Efecto del tamaño de la red

El análisis del contenido de la base de datos permite comprender la asimetría existente en la red de LinkedIn en función del tamaño de la red. Así, el 51,14% de los contenidos los suben usuarios con menos de 1442 conexiones. Sólo el 1,3% de los usuarios tiene más de 20.000 conexiones (Figura 8).

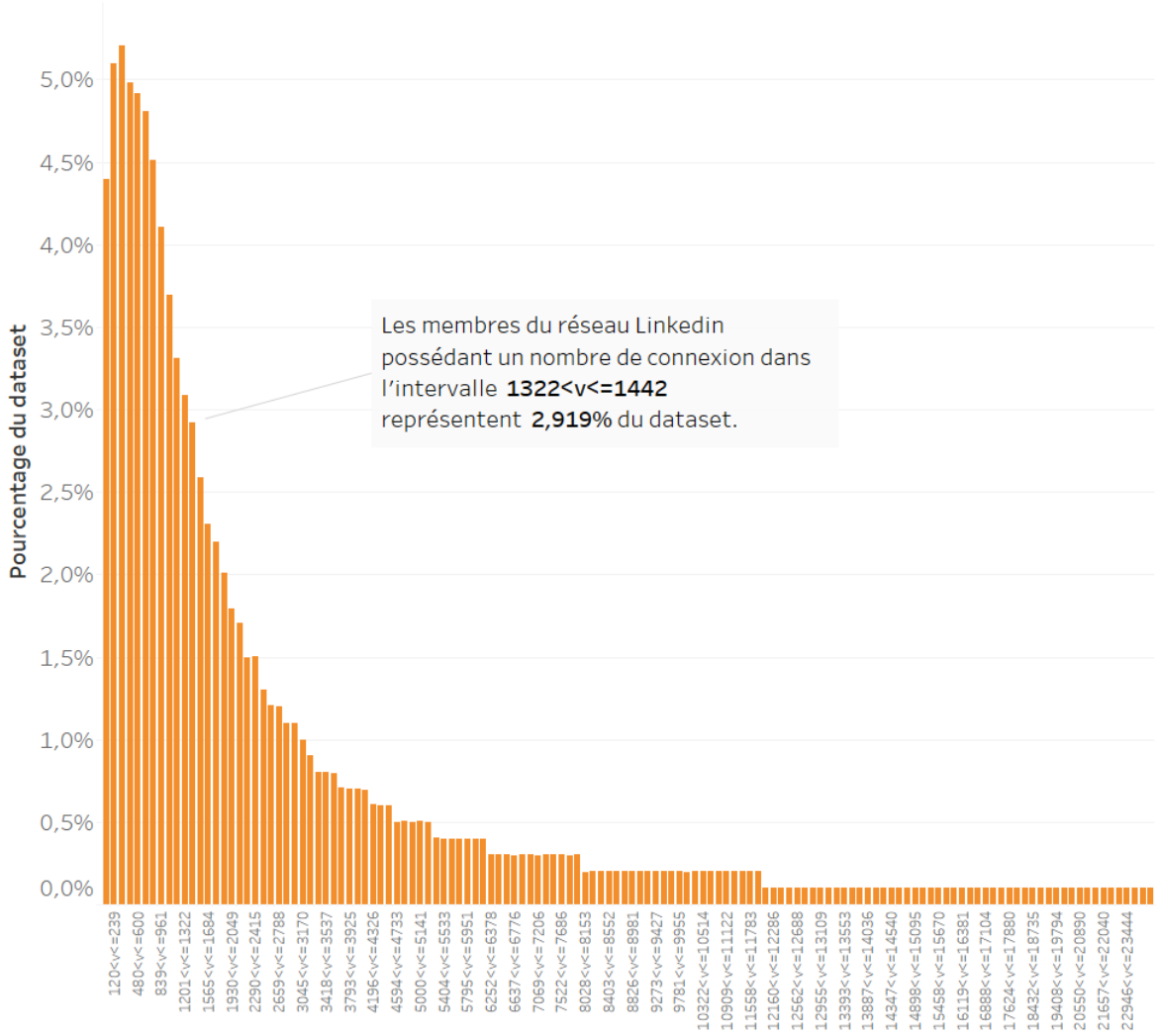


Imagen 8. Distribución de los mensajes de LinkedIn según el tamaño de la red del autor

El análisis de los 4,599 millones de publicaciones de LinkedIn (Figura 8) muestra claramente el efecto del tamaño de la red. Esto es bastante lógico: cuanto más extensa es la red de un miembro de LinkedIn, mayor es la probabilidad de obtener 100 reacciones en una de sus publicaciones.

Sin embargo, es sorprendente ver cuánto influye esta variable: 33.99%. En otras palabras, **alcanzar las 100 reacciones en una publicación de LinkedIn está condicionado en 1/3 por el tamaño de su red.**

Para la mayoría de los usuarios de LinkedIn, la perspectiva de conseguir 100 reacciones sigue siendo muy hipotética. El 50% de los usuarios de LinkedIn tiene menos de 1442 conexiones. Con 1442 conexiones, tenemos una posibilidad entre 50 de conseguir 100 likes/comentarios (2%). Sin embargo, con una red de 24000 contactos, la probabilidad se eleva al 19%. Estamos hablando de casi 10 veces más.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

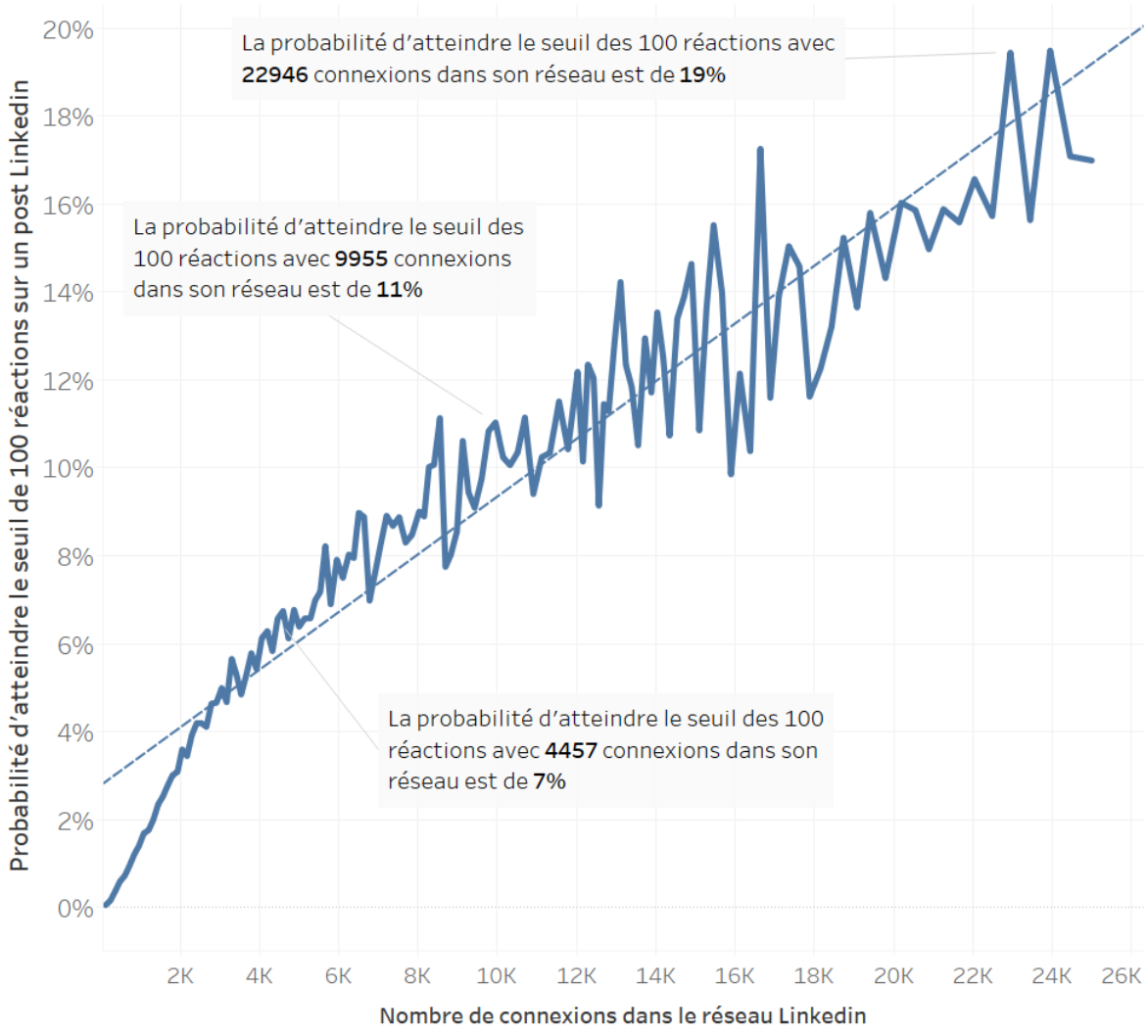


Imagen 9. Efecto del tamaño de la red en la probabilidad de obtener al menos 100 reacciones en una publicación de LinkedIn

Es posible ver los efectos en la Figura 9. Podemos establecer una relación lineal entre el número de conexiones en la red y la probabilidad de alcanzar el resultado (línea de puntos). La probabilidad de obtener 100 reacciones aumenta un 1% por cada 500 conexiones.

c) Efecto del número de palabras

El segundo factor que tiene un efecto significativo es el número de palabras de la publicación en LinkedIn. Este factor influye en casi el 20% de la probabilidad de alcanzar 100 likes o comentarios.

Como podemos ver, la distribución de las publicaciones de LinkedIn según el número de palabras que contienen dista mucho de la media (véase la figura 10). La mitad de las publicaciones de LinkedIn tienen menos de 36 palabras, y el 1,26% (o 50341 publicaciones) sólo contienen 2 palabras. En el otro extremo del espectro, el 0,12% de las publicaciones tienen entre 222 y 225 palabras.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

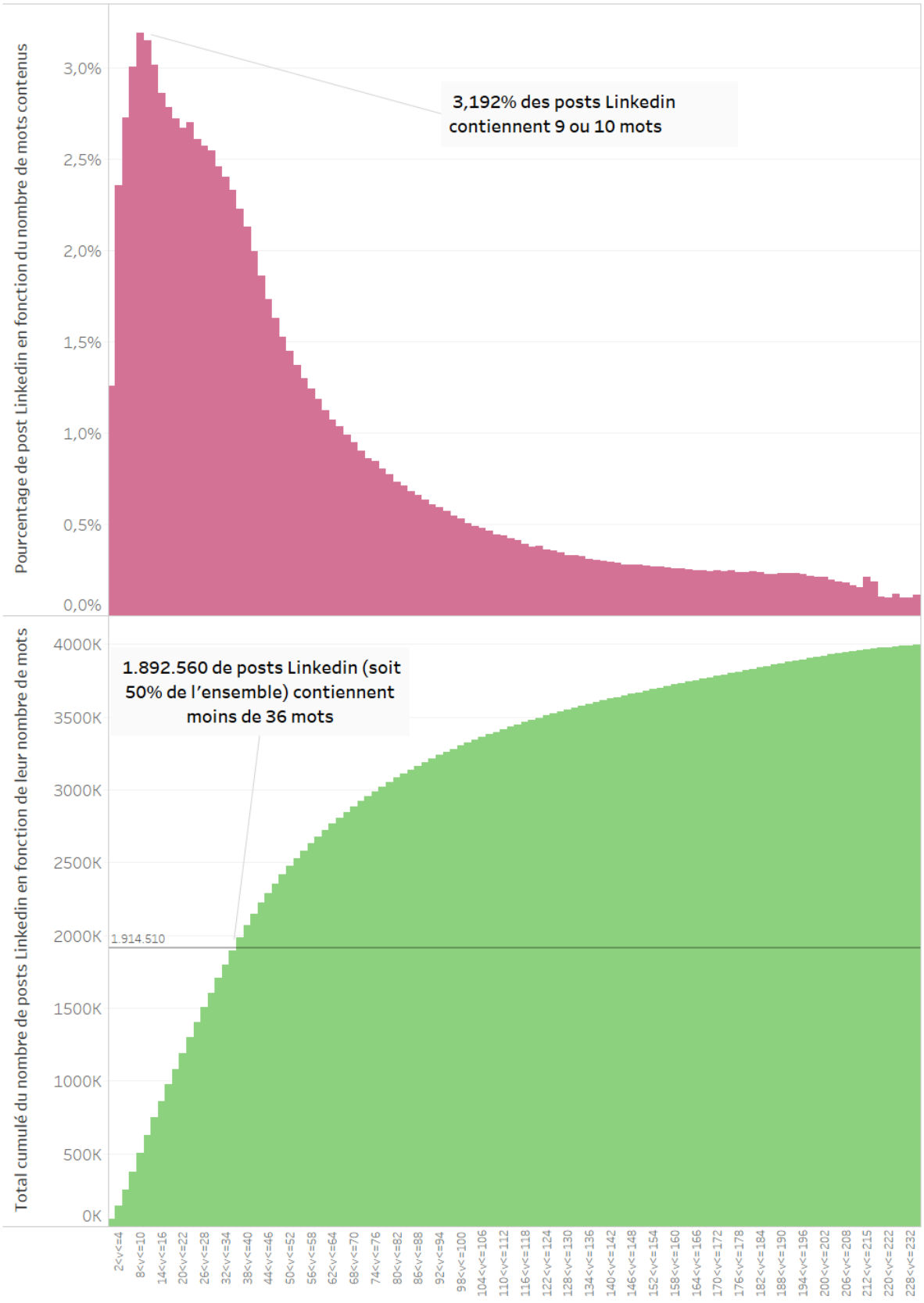


Imagen 10. Distribución estadística de las publicaciones de LinkedIn según el número de palabras

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

La modelización (Figura 11) muestra que una publicación que contiene 232 palabras tiene 6 veces más probabilidades de obtener 100 me gusta/comentarios que lo normal.

Sin embargo, el efecto no es lineal. Se observa un punto de inflexión en torno a las 150-170 palabras. Esto significa que hay un interés indudable en superar este límite a la hora de escribir contenidos.

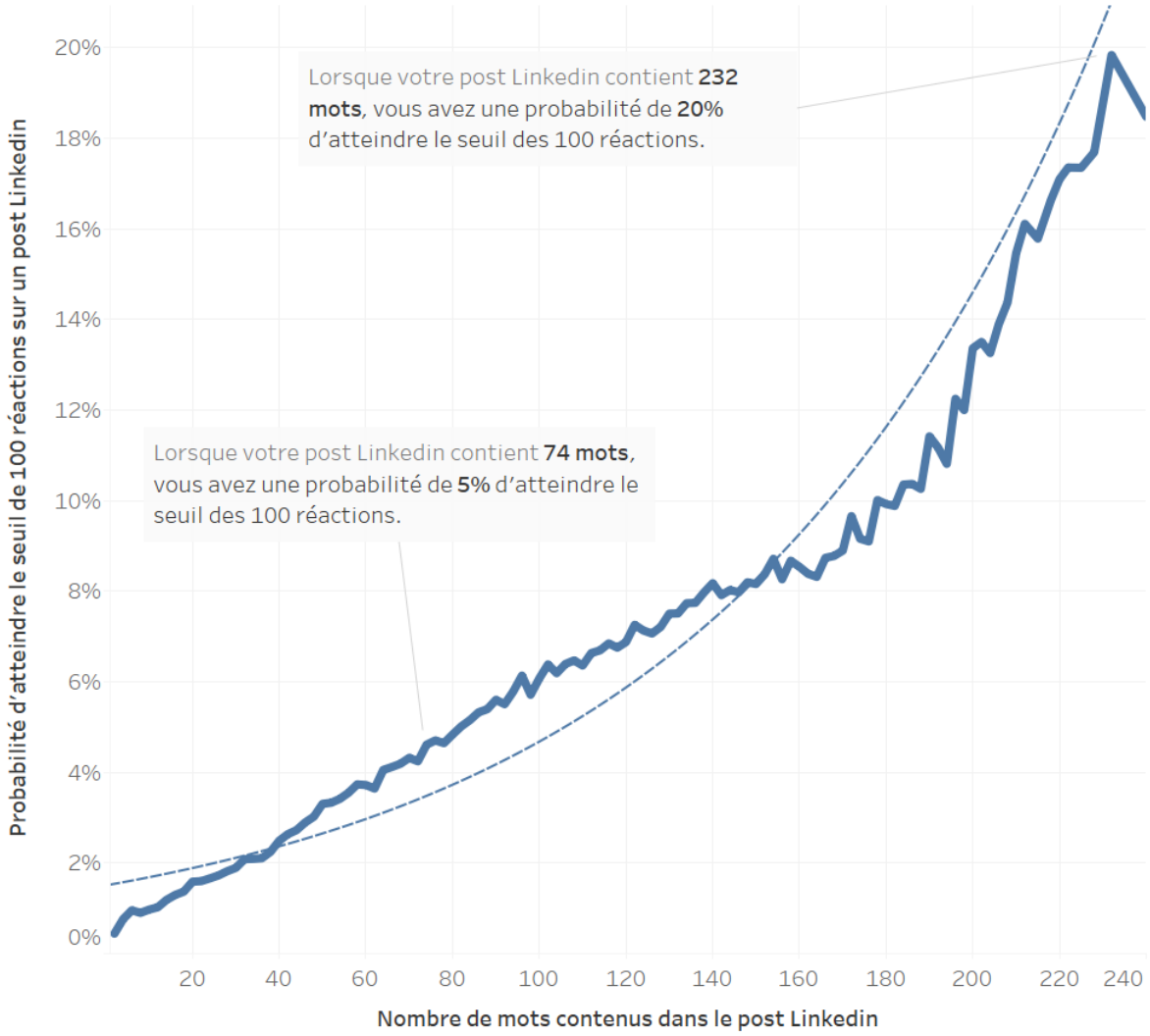


Imagen 11: Influencia del número de palabras en la probabilidad de obtener 100 reacciones en una publicación de LinkedIn

Este efecto se explica por los cambios realizados en el algoritmo de LinkedIn en 2019. El "tiempo de permanencia" se ha convertido en la variable que condiciona la visibilidad de una publicación dentro de la red de LinkedIn y, por tanto, su capacidad de recibir “me gusta” o comentarios. El "tiempo de permanencia" mide el tiempo de interacción con el contenido. Cuanto más tiempo pase el usuario leyendo el contenido, más infiere el algoritmo que ese contenido es interesante y merece ser expuesto a otras personas de la red. Es más

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu’un hyperlien soit fait vers le site d’IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

probable que un contenido de 232 palabras atraiga a un usuario durante más tiempo que un contenido de 50 palabras. Así pues, el efecto era previsible, pero se cuantifica por primera vez.

En cuanto al punto de inflexión visible en torno a las 150-170 palabras, sólo podemos constatar su existencia sin explicar su significado..

d) Efecto del número de emojis

La última variable con un efecto significativo es el número de los emojis.

El análisis estadístico descriptivo muestra de nuevo una distribución que se aleja de la media, ya que el 80,12% de las publicaciones de LinkedIn no contienen ningún emoji (Figura 12). Sólo el 9,58% de las publicaciones de LinkedIn incluye 3 o más emojis. El máximo observado en una sola publicación fue de 1190 emojis.

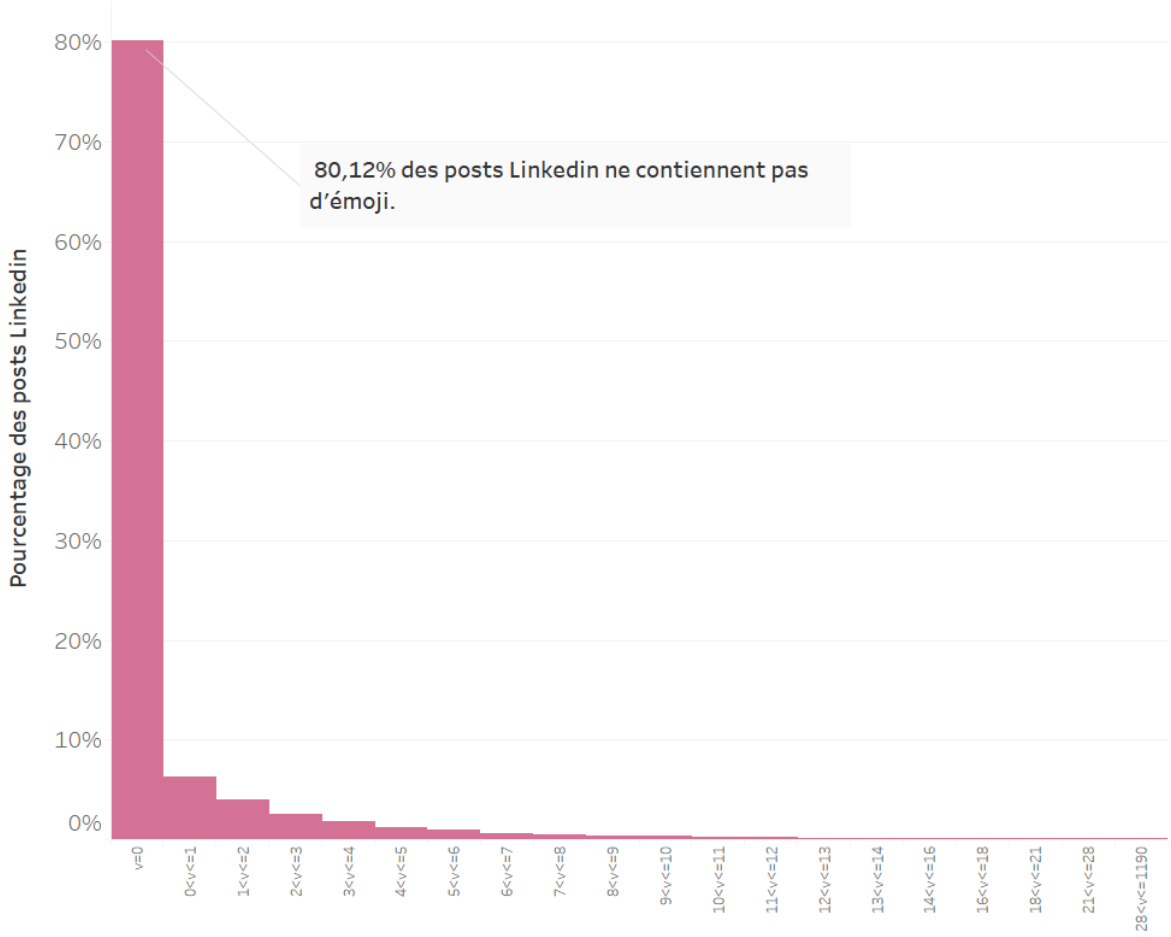


Imagen 12. Distribución estadística del número de posts de LinkedIn según el número de emojis que contienen.

En la Figura 13 se presenta la modelización del efecto de los emojis para alcanzar el umbral de 100 reacciones. Asistimos a un impacto positivo ya desde el primer emoji. En otras palabras, **poner un solo emoji en una publicación de LinkedIn ya aumenta las posibilidades de obtener 100 likes/comentarios**.

Este efecto estadístico se explica por la distribución de los emojis en las publicaciones de LinkedIn. En efecto, **el 80,2% de las publicaciones no contiene emojis**. Por tanto, el efecto se observa en cuanto se supera el umbral de 0.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

Incluir 16 emojis, como se muestra en la Figura 13, es ideal. Con 16 emojis, las posibilidades de superar las 100 reacciones se multiplican por 2,5 en comparación con la media.

El análisis no dice dónde deben incluirse estos emojis ni cuáles hay que elegir.

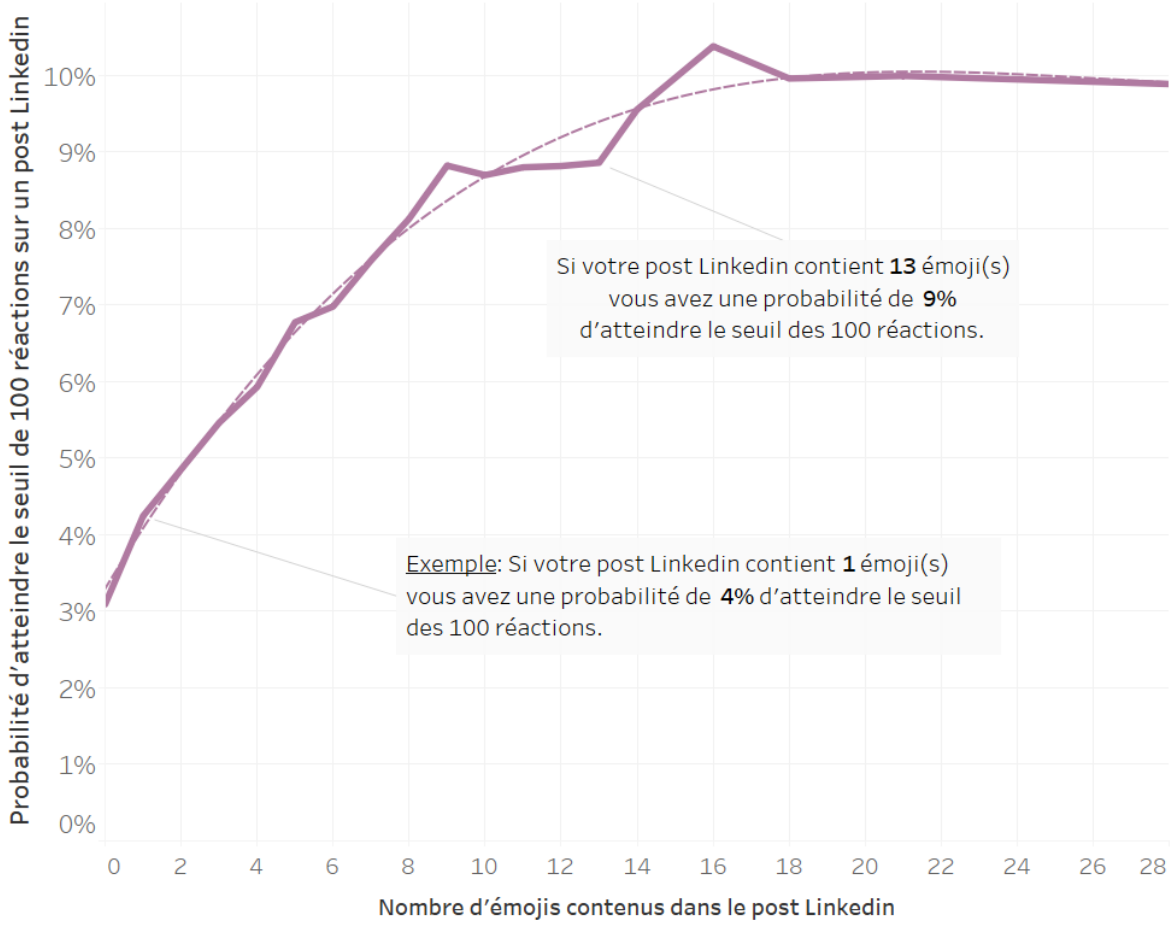


Imagen 13. Influencia del número de emojis en la probabilidad de obtener al menos 100 reacciones en una publicación de LinkedIn

e) Correlaciones entre las variables

A primera vista, se podría pensar que el "número de emojis" es redundante con la variable "número de palabras". ¿Acaso un post de más de 200 palabras no tiene una mayor probabilidad de contener el número "correcto" de emojis?

El análisis muestra que estas 2 variables están efectivamente relacionadas (la correlación lineal entre las 2 variables es del 19%). Por tanto, la cuestión es saber cuál de estas dos variables influye en la viralidad de una publicación de LinkedIn y si ambas son necesarias.

Podemos apreciar el efecto inducido por la eliminación de la variable "número de emojis" directamente en los resultados entregados por "TIMi Modeler". La eliminación de la variable "número de emojis" de un modelo predictivo que contiene la variable "número de palabras" conduce a una pérdida de "poder predictivo" del modelo, materializada en una disminución del 2,4% del AUC (Área Bajo la Curva).

Esto demuestra que la variable "número de emojis" tiene una influencia directa y diferente de la variable "número de palabras" en la probabilidad de alcanzar el umbral de 100 reacciones. Por lo tanto, debemos mantenerla en el modelo.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

En la práctica, al escribir un post en LinkedIn, debemos centrarnos en la longitud del post y en el número correcto de emojis (16).

f) Variables sin efecto significativo

El país y el idioma no influyen en la probabilidad de obtener 100 reacciones en un post de LinkedIn. Es una buena noticia porque significa que **no es necesario vivir en Estados Unidos ni escribir en inglés para lograr este objetivo.**

Este resultado se ha obtenido a partir de un conjunto de datos que abarca todos los países del mundo, pero aún así muestra grandes disparidades en los índices de penetración. El conjunto de datos incluye 604241 usuarios en Estados Unidos, pero sólo 746 en Estonia (ver Figura 14 y Figura 15).

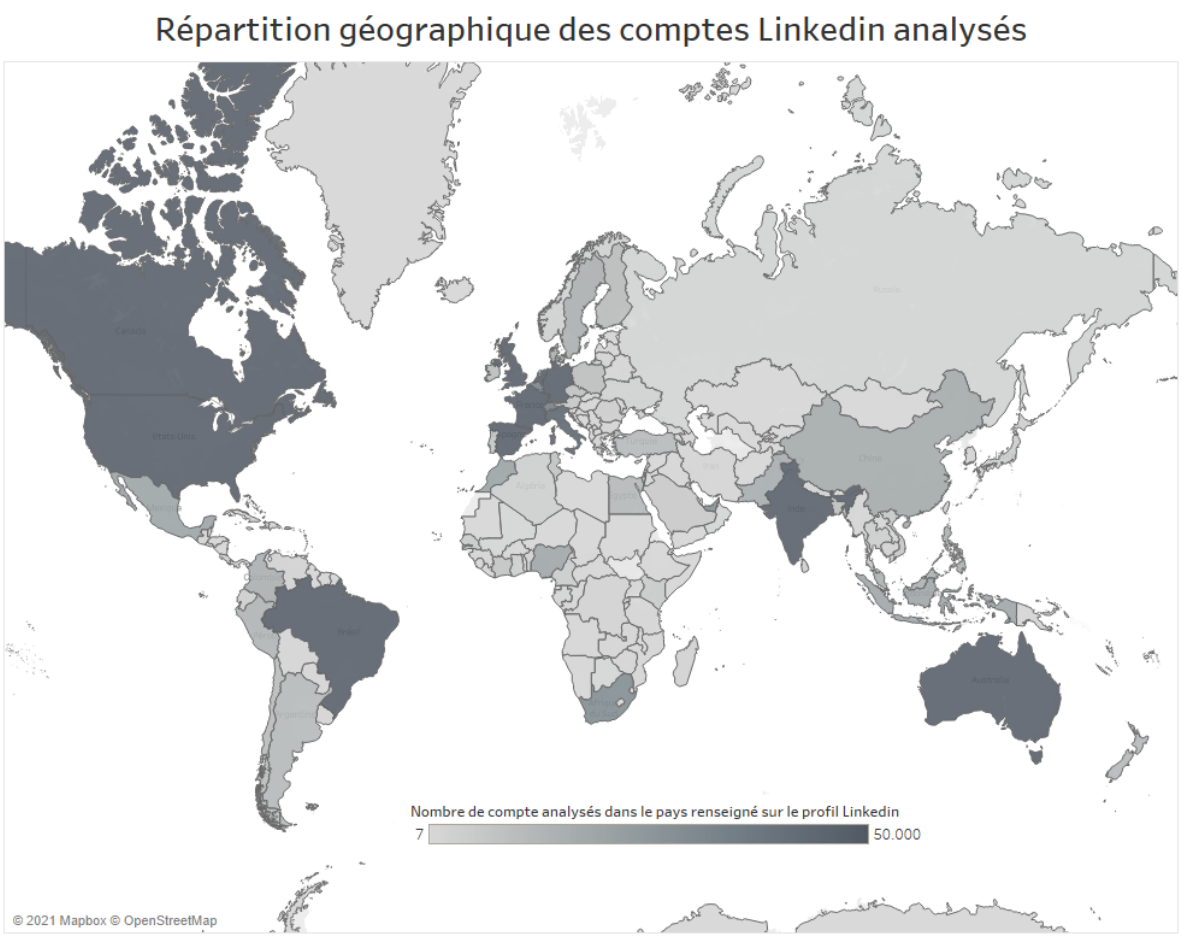


Imagen 14. Distribución geográfica de los usuarios del conjunto de datos (según el país introducido en su perfil de LinkedIn)

Sin embargo, teniendo en cuenta todas las variables en el mismo modelo vemos que los hashtags no tienen ninguna influencia estadística. En otras palabras, **los hashtags no influyen significativamente en la probabilidad de obtener 100 reacciones**. En un primer modelo que no incluía los emojis, los hashtags mostraban un efecto significativo. Pero en cuanto se tuvieron en cuenta los emojis en el modelo, su impacto, aunque modesto (2,4% del total), superó al de los hashtags. Este hallazgo, por lo tanto, socava una suposición común sobre el efecto de los hashtags en la viralidad de las publicaciones de LinkedIn.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

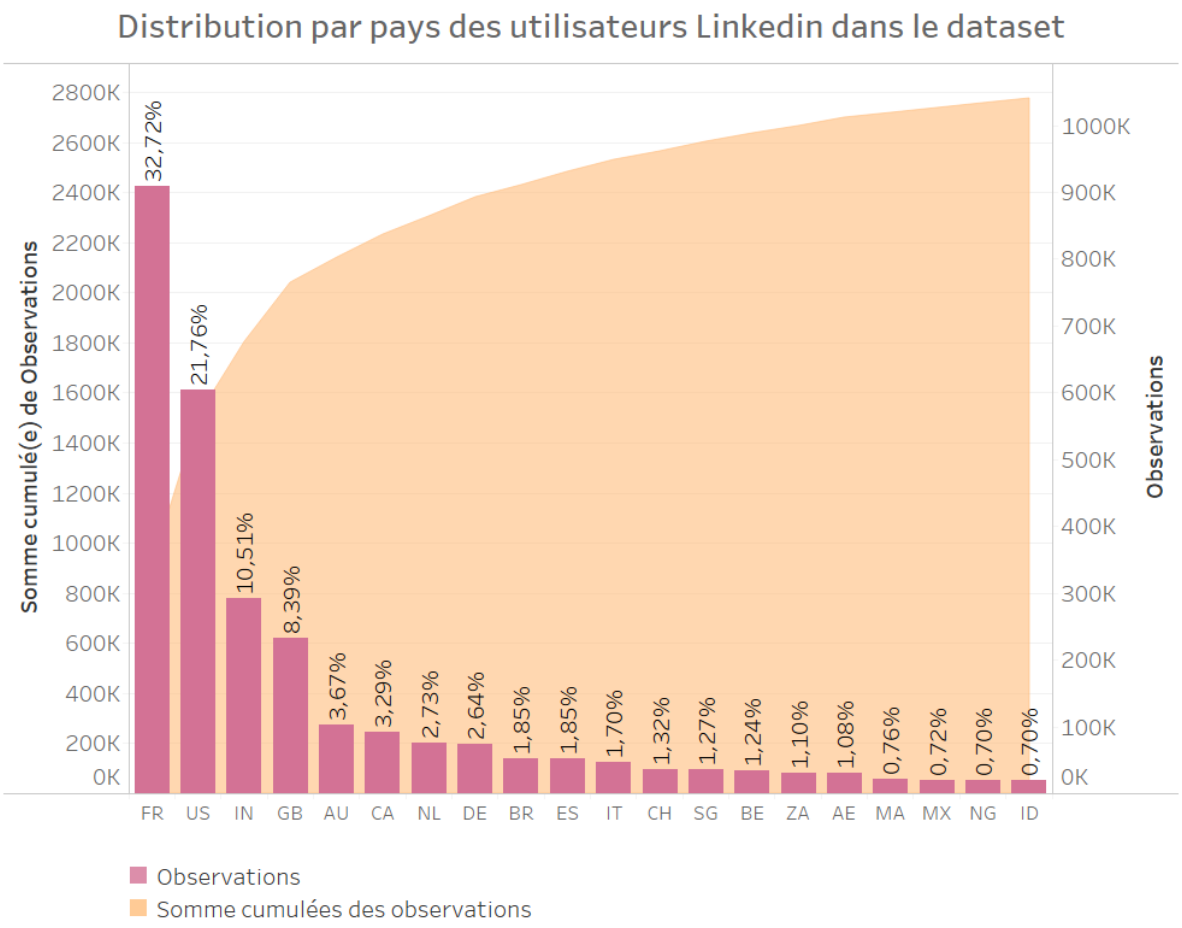


Imagen 15. El 70,3% de los usuarios de LinkedIn en la base de datos se concentra en 22 países

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

Metodología

4. Metodología

a) Preparación de los datos

Los datos se prepararon conciliando 2 bases de datos en formato JSON que contenían

- Los datos del contenido (con 4,599 millones de líneas extraídas y un archivo JSON de 3,1 GB)
- Los datos sobre los autores (con 407.737 líneas extraídas y un archivo JSON de 781 MB)

La conciliación se ha realizado a partir del identificador único del autor mediante el software Anatella proporcionado por la empresa TIMi.

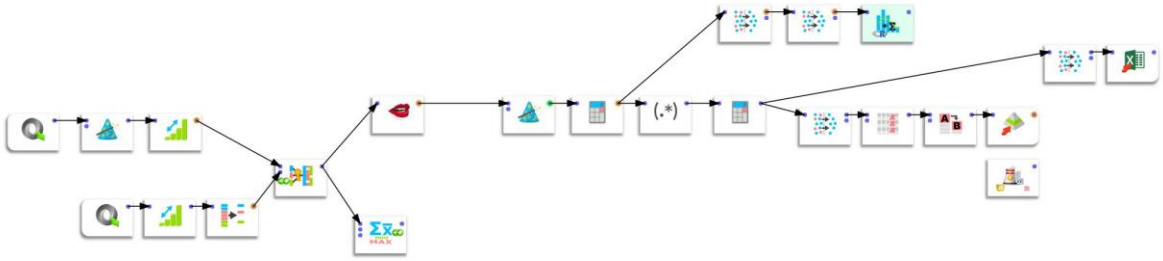


Imagen 16. Esquema de la preparación de los datos con Anatella

Después de la conciliación, se realizó el enriquecimiento de los datos de contenido utilizando el cuadro "DetectLanguage" de Anatella, que utiliza el algoritmo de detección de idiomas denominado "CLD2" (véase Lui y Baldwin, 2014).

Todo el procedimiento de preparación de los datos se ejecuta en 457 segundos (el cuadro "DetectLanguage" consume el 23% del tiempo de cálculo). La memoria RAM consumida es de 2260 MB.

Para evitar los efectos secundarios creados por usuarios conocidos o con muchos suscriptores, se eliminaron del conjunto de datos las personas con más de 25.000 conexiones (dejando 3,999 millones de líneas utilizables).

b) Modelado

El modelado se realizó utilizando Modeler de la empresa TIMi. Modeler permite probar millones de modelos (auto-ML) y encontrar el que ofrece los mejores resultados automáticamente.

La modelización se realiza a partir de los datos brutos exportados en formato .gel (formato propio de TIMi) y designando el "objetivo", es decir, la variable independiente a modelizar. Para esta investigación, elegimos un objetivo binario separando el conjunto de datos en 2 partes: las publicaciones que habían recibido menos de 100 reacciones (me gusta + comentarios) y las que habían recibido 100 o más. El análisis, por tanto, ofrece resultados sobre la probabilidad de alcanzar el umbral de 100 reacciones.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

[1] Lui, M., & Baldwin, T. (2014, abril). Identificación lingüística precisa de los mensajes de Twitter. En Proceedings of the 5th workshop on language analysis for social media (LASM) (pp. 17-25).

Sobre este conjunto de datos de 3,999 millones de líneas, la herramienta "Modeler" construye varios miles de modelos predictivos, selecciona y entrega el mejor modelo, y genera todos los informes de análisis (en formato .docx, .xlsx y .html) en pocos segundos.

La calidad de un modelo de predicción con un objetivo "binario" (es decir, un objetivo con dos estados: o el puesto ha alcanzado las 100 reacciones o el puesto no ha alcanzado las 100 reacciones) suele ilustrarse mediante una curva de elevación. Los detalles de la curva de elevación obtenida con el modelo predictivo construido con el Modelador TIMi se muestran en la curva azul de la Figura 17.

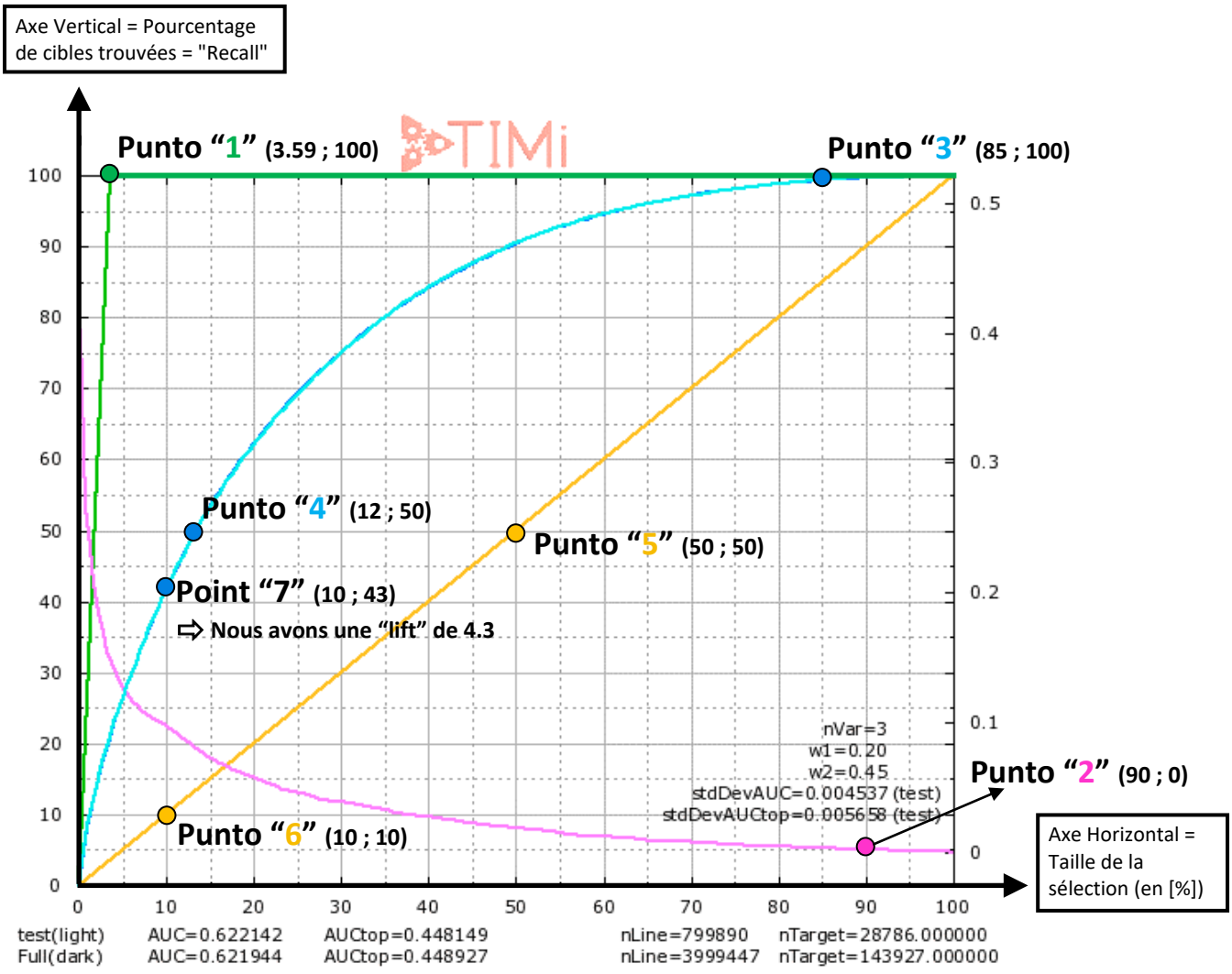


Imagen 17. Ajuste del modelo de predicción utilizado en esta investigación

Para explicar la elevación en el gráfico anterior, tenemos un modelo de predicción "perfecto" que no comete errores. Ahora utilicemos este modelo perfecto para ordenar toda nuestra población (es decir, todas nuestras

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

"publicaciones") en el eje X. Este orden no es aleatorio: se basa en nuestro modelo predictivo, es decir, colocaremos...

- ...a la izquierda, las "publicaciones" que, según nuestro modelo, tienen una probabilidad muy alta de estar en el objetivo (los que tienen una mayor oportunidad de alcanzar las 100 reacciones).
- ...a la derecha, las "publicaciones" que tienen una escasa probabilidad de alcanzar el objetivo.

En la figura 17, el modelo de predicción "perfecto" está ilustrado por la curva de elevación **verde**. Esta curva representa la calidad de nuestra programación. Lo ideal sería que todos los "objetivos" (posts que han alcanzado 100 reacciones) se situaran a la izquierda. En este caso, el tamaño de nuestro objetivo es del 3,59% (es decir, el 3,59% de los posts obtienen 100 reacciones). Esto significa que el modelo predictivo "perfecto" detectará el 100% de nuestros objetivos seleccionando sólo el 3,59% de la población (esta información está representada por el punto "1" en la figura 17).

La curva **verde** (el modelo de predicción perfecto) se compone de dos segmentos lineales. El primer segmento une las coordenadas (0; 0) y (3,59; 100). Representa lo que ocurre cuando utilizamos un modelo predictivo perfecto para seleccionar una población cada vez mayor (desde una selección vacía del 0% hasta un tamaño que representa el 3,59% de todas las publicaciones analizadas). Con un modelo perfecto, cuando finalmente hemos seleccionado el 3,59% de la población (es decir, cuando llegamos al punto "1"), hemos "encontrado" el 100% de los objetivos (las publicaciones que tienen más de 100 reacciones).

En cambio, un modelo de predicción "normal" debe "reclutar" muchas más publicaciones que el 3,59% mínimo de la población para encontrar el 100% del objetivo. Su "ordenación" no es perfecta. La ordenación se basa en los resultados del modelo: los posts que tienen (según nuestro modelo "normal") la mayor probabilidad de estar en el objetivo (100 reacciones o más) se colocan en el eje X de la izquierda.

En el gráfico anterior, la curva **rosa** muestra las diferentes probabilidades calculadas por el modelo de predicción para cada uno de los posts conocidos. Por construcción, la curva **rosa** disminuye constantemente de izquierda a derecha. Por ejemplo, gracias al punto "2" con coordenadas (90; 5), sabemos que un 10% de las publicaciones tienen (según nuestro modelo) una probabilidad nula de estar en el objetivo (son el 10% de las publicaciones situadas a la derecha del punto "2").

En la figura 17, la calidad del modelo predictivo construido aquí está representada por la "curva de elevación" **azul**. El modelo predictivo realizado con "TIMi Modeler" debe:

- ... seleccionar el 85% de la población para encontrar todos los objetivos: es el punto "3" del gráfico con coordenadas (85; 100).
- ... seleccionar el 12% de la población para encontrar la mitad de los objetivos: es el punto "4" del gráfico con coordenadas (12; 50).

En la figura 17, la última curva **amarilla** representa la calidad de un modelo de predicción terrible, ya que este modelo ordena las publicaciones según un orden perfectamente aleatorio. Por ejemplo, por el simple efecto del azar, este modelo totalmente aleatorio conseguirá encontrar el 50% del objetivo seleccionando el 50% del contenido: es el punto "**5**" de la figura 11, con coordenadas (50; 50).

Veamos ahora el punto "**6**" de la figura 17, con coordenadas (10; 10). Situado en la curva **amarilla**, representa el rendimiento de un modelo aleatorio. El modelo obtenido con "TIMi Modeler" (curva **azul**) es mejor que un

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

modelo aleatorio porque no pasa por el punto "6" sino por el punto "7" de coordenadas (42; 10). Como el punto "7" es 4,2 veces mayor que el punto "6", diremos que el modelo TIMi es 4,2 veces mejor que el modelo aleatorio. O que el "lift" del modelo TIMi es 4,2.

Para estimar la calidad de un modelo de predicción, nos interesan dos medidas

- la "elevación" del modelo predictivo (que en este caso es de 4,2)
- el AUC (Área Bajo la Curva) del modelo predictivo.

En TIMi, el AUC se define como el área sombreada de la figura 18.

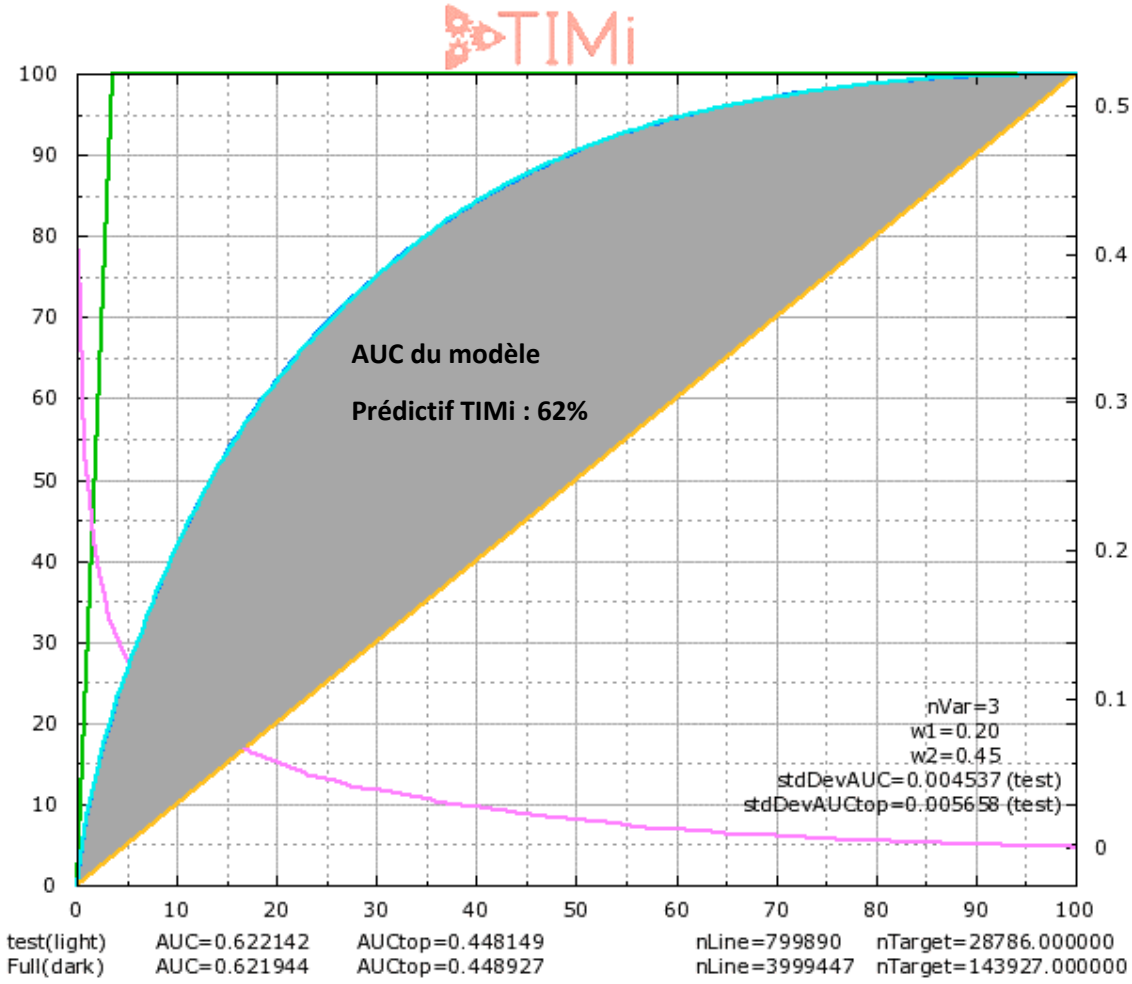


Imagen 18. Definición del AUC de un modelo predictivo TIMi

Por convención

- el AUC del modelo predictivo "perfecto" en **verde** se establece en AUC=100%.

- el AUC del modelo predictivo "aleatorio" en amarillo es AUC=0%.

La siguiente tabla compara las dos medidas de calidad más comunes (elevación y AUC) si se trabaja con modelos predictivos binarios.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

Nombre	Aspectos positivos	Aspectos negativos
Lift	Fácil de entender, nos permite evaluar hasta qué punto el modelo entregado es superior al modelo aleatorio.	No es muy fiable porque solo representa lo que ocurre en un único punto de la curva de elevación (en X=10%, por ejemplo).
AUC	Medición muy fiable porque utiliza toda la curva de elevación desde X=0% hasta X=100%.	Se trata de una medida más abstracta y más difícil de entender, pero aún así existen posibles explicaciones. Por ejemplo, si utilizamos otra convención que fija el AUC del modelo aleatorio en el 50% (en lugar del 0%), podemos decir: *"El AUC es la probabilidad de que, tomando al azar 2 publicaciones situadas en el eje X, estas 2 publicaciones se ordenen en el orden correcto (el que tiene mayor probabilidad de ser "viral" se sitúa a la izquierda)."*

En el modelador TIMi, todos los cálculos tienen como objetivo maximizar el "AUC superior" (una variante mejorada del AUC). Por ejemplo:

- TIMi Modeler crea miles de modelos y selecciona el modelo con el AUC más alto.
- TIMi Modeler simplifica el modelo (durante el procedimiento opcional de poda) eliminando del modelo predictivo todas las variables que no influyen en el AUC.

Así, en este análisis, la variable "país" se eliminó automáticamente del modelo predictivo final entregado por TIMi Modeler. Por otro lado, la variable "tamaño de la red" debe mantenerse porque su ausencia elimina el 33,99% del AUC del modelo.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

TIMi suite

El análisis de datos de LinkedIn que se presenta en este White Paper se ha realizado utilizando dos herramientas de TIMi suite:

Anatella es la primera herramienta utilizada: dedicada a la "preparación de datos".
Las principales ventajas de Anatella son :

- **RAPIDEZ**: Anatella maneja tablas con miles de millones de líneas y miles de columnas en una pequeña infraestructura. Esta potencia permite un análisis más profundo de los datos para obtener resultados mejores y más relevantes. Anatella es tan rápido que es posible analizar tablas con miles de millones de líneas utilizando un simple ordenador portátil, casi en tiempo real. Anatella es un poco como "Excel" pero con esteroides.
- **FÁCIL**: desarrolle transformaciones de datos complejas más rápidamente con una interfaz intuitiva que no requiere código.
- **AUTOMATIZADO**: automatice cualquier proceso empresarial, informático o de otro tipo. Con sólo unos pocos clics, puede poner sus gráficos en producción y ejecutarlos repetidamente.
- **ABIERTO**: con más de 400 cuadros disponibles, todo es posible, casi al instante. Si es necesario, utilice el marco de colaboración y la herramienta de revisión integrada para crear fácilmente nuevas cajas (en R, Python, JavaScript o C).

Modeler es la segunda herramienta utilizada: es una herramienta de "Aprendizaje Automático de Máquinas (auto-ML)".
Las principales ventajas de Modeler son:

- **PRECISIÓN Y RAPIDEZ**: Modeler ofrece los mejores modelos predictivos en pocos segundos. Con estos modelos, obtendrá las mejores tasas de conversión y retención para todas sus campañas de marketing y la mejor estimación del riesgo para proteger su negocio.
- **POTENTE**: no hay límite para el conjunto de datos de entrenamiento: El modelador maneja decenas de miles de variables y millones de líneas.
- **AUTOMATIZADO**: serán suficientes unos pocos clics para beneficiarse de más de 20 años de experiencia en ciencia de datos incorporada directamente en Modeler.
- **ACTUABLE**: Modeler garantiza que sus modelos saldrán del laboratorio de datos y pasarán a producción con un simple arrastrar y soltar.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

Linkalyze

Linkalyze es una investigación de influencias y solución de análisis de LinkedIn. Linkalyze identifica a los influencer de LinkedIn en temas específicos y ayuda a las marcas a definir el valor de una publicación basándose en diferentes criterios.

Contacto: Sylvain Tillon

sylvain@linkalyze.app

+33 609 924 038

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

Conclusiones de la investigación

5. Conclusiones e investigación futura

A través de la investigación sobre los 4.599 millones de publicaciones en LinkedIn podemos extraer lecciones claras y estadísticamente objetivas sobre el papel de los diferentes factores. Demostramos que, entre las variables tenidas en cuenta, sólo 3 tienen un impacto significativo en la probabilidad de alcanzar el umbral de las 100 reacciones: el tamaño de la red (medido en número de conexiones), el número de palabras y el número de emojis.

Sin embargo, no se trata de conclusiones definitivas. En primer lugar, el número de variables que se han tenido en cuenta es limitado, por lo que no representan todos los factores que intervienen. En segundo lugar, es imposible en este momento demostrar una relación causal entre las variables estudiadas y la consecución del umbral de 100 reacciones. ¿El algoritmo de recomendación de LinkedIn tiene en cuenta estas variables para mostrar el post a un número más significativo de personas? ¿O es el propio contenido del post el que, por su número de palabras y los emojis que contiene, atrae más la atención e influye así en el algoritmo de recomendación?

Nuestros siguientes análisis explorarán el conjunto de datos con mayor profundidad. Además de considerar nuevas variables, queremos analizar la semántica de las publicaciones de LinkedIn para averiguar si los temas o el estilo son factores diferenciadores de los posts de LinkedIn que tienen éxito.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

6. Agradecimientos

El autor desea agradecer a la empresa Linkalyze y, en particular, a Sylvain Tillon, Thomas Pons y Laetitia Verrière por su valiosa ayuda y por la disponibilidad de los datos que han permitido la realización de esta investigación.

Gracias también a la empresa TIMi por haber permitido la utilización de Anatella y Modeler, y en particular a su responsable Frank Vanden Berghen por las múltiples interacciones y el apoyo técnico durante la explotación de los resultados.

Les informations contenues dans ce rapport peuvent être réutilisées à la condition expresse qu'un hyperlien soit fait vers le site d'IntoTheMinds https://www.intotheminds.com. Toute utilisation non conforme sera poursuivie.

LA SOLUTION
LA PLUS
RAPIDE DU MARCHÉ

TIMi effectue des transformations complexes sur plusieurs terabytes de données en quelques minutes.

DES MONTAGNES
DE DONNÉES
MANIPULÉES
AISÉMENT

TIMi stocke, traîte et analyse des milliards de lignes et des milliers de colonnes sur un simple PC.

TIMi

CREATIVITY THROUGH EFFICIENCY

CRÉEZ FACILEMENT
DES MODÈLES
PRÉDICTIFS

TIMi libère votre créativité pour inventer des transformations ingénieuses et construire en quelques clicks des modèles fiables et interprétables.

Découvrez La Suite TIMi :

Anatella
Le cœur du framework et un ETL analytique

Modeler
Moteur AUTO-ML en temps réel de TIMi

Stardust
Segmentation et visualisation 3D et VR

Kibella
Tableau de bord en libre service, interactif & illimité

Visite www.timi.eu para un ensayo gratuito en TIMi Suite

IntoTheMinds sprl
Rue Général Capiaumont 11
B-1040 Etterbeek

Tel. (B) : +32 (0)2 347 45 86
Tel. (F) : +33 (0)1 88 32 73 44
info@intotheminds.com

www.intotheminds.com
Facebook/IntoTheMinds
Youtube/IntoTheMinds