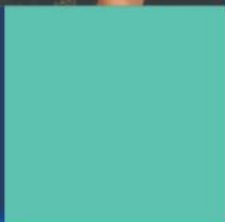


Aviso: Alertas para la Vigilancia del Sistema Educativo.

Machine Learning y Minería de Grafos para el monitoreo costo-efectivo de redes sociales.



Aviso: Alertas para la Vigilancia del Sistema Educativo

Machine Learning y Minería de Grafos para el monitoreo costo-efectivo de redes sociales

Jefa de la Oficina de Seguimiento y Evaluación Estratégica

Lourdes Patricia Vargas Vilchez

Elaboración de Contenidos:

Erik Carl Candela Rojas

Severo Alfredo Aquino Baldeon



Esta obra está bajo una Licencia [Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/).
Vea una copia de esta licencia en <https://creativecommons.org/licenses/by-nc-nd/4.0/>

MINISTERIO DE EDUCACIÓN

Oficina de Seguimiento y Evaluación Estratégico

Diciembre - 2023

Sede Central: Calle Del Comercio N° 193 Lima - Lima - San Borja - 15021 Perú

Teléfono: (01) 615-5800

<https://www.gob.pe/minedu>

Avise: Alertas para la Vigilancia del Sistema Educativo

Machine Learning y Minería de Grafos para el monitoreo costo-efectivo de redes sociales.

Presentación:

En la actual era digital, las redes sociales han emergido como una importante fuente de información que proporciona al sector público una valiosa ventana hacia las opiniones, inquietudes y demandas de la ciudadanía.

El Ministerio de Educación del Perú (Minedu) reconoce esta importancia y, a través de la Oficina de Seguimiento y Evaluación Estratégica (OSEE), se emprendieron esfuerzos para aplicar diversas técnicas de analítica avanzada en la información generada por las redes sociales con la finalidad de obtener una mayor comprensión y visibilidad de las opiniones, tendencias y necesidades presentes en el sector educativo en el Perú.

Producto de este esfuerzo, la OSEE desarrolló la innovación «Alertas para la vigilancia del sistema educativo» (Avise). Esta innovación hace uso de técnicas de *Machine Learning* y Minería de Grafos para analizar la información generada en las redes sociales con el objetivo de brindar información oportuna a las autoridades educativas para que puedan responder, de forma más efectiva, las problemáticas y necesidades de la ciudadanía.

La información oportuna que brinda Avise facilita la identificación de los temas más significativos dentro del ámbito educativo, los cuales están siendo debatidos en las redes sociales. Además, permite la identificación de usuarios que desempeñan un papel clave en la promoción e influencia sobre otros usuarios en temas relacionados con la educación en el Perú. En última instancia, Avise posibilita la evaluación del sentimiento expresado en los comentarios compartidos por los usuarios que expresan sus opiniones sobre temas educativos.

El presente documento busca proporcionar una breve descripción de las técnicas analíticas que fueron empleadas en la creación de Avise, así como el proceso de descubrimiento de información valiosa y la generación de recomendaciones a partir de un caso de uso práctico.

Palabras clave: *Avise, Social Network Analysis, Social Media Analytics, Social Science Graphs, Social Media Monitoring, Influential Users, Sentiment Analysis, Education Trends, Early Warning, Artificial Intelligence, Natural Language Processing, Machine Learning.*

Abstract:

In the current digital era, social networks have emerged as an important source of information providing the public sector with a valuable window into the opinions, concerns, and demands of the citizenry. The Ministry of Education of Peru (Minedu) acknowledges this importance, and through the Office of Monitoring and Strategic Evaluation (OSEE), efforts were undertaken to apply various advanced analytics techniques to the information generated by social networks in order to gain a better understanding and visibility of opinions, trends, and needs in the educational sector in Peru.

As a result of these efforts, the OSEE developed the innovation «Alerts for the Surveillance of the Educational System» (Avisé). This innovation employs Machine Learning and Graph Mining techniques to analyze information generated on social networks with the aim of providing timely information to educational authorities so they can more effectively respond to the issues and needs of the citizenry.

The timely information provided by Avisé facilitates the identification of the most significant topics within the educational scope that are being debated on social networks. Additionally, it allows the identification of users who play a key role in promoting and influencing others on education-related topics in Peru. Ultimately, Avisé enables the evaluation of the sentiment expressed in comments shared by users expressing their opinions on educational issues.

This document aims to provide a brief description of the analytical techniques employed in the creation of Avisé, as well as the process of discovering valuable information and generating recommendations based on a practical use case.

Keywords: *Avisé, Social Network Analysis, Social Media Analytics, Social Science Graphs, Social Media Monitoring, Influential Users, Sentiment Analysis, Education Trends, Early Warning, Artificial Intelligence, Natural Language Processing, Machine Learning.*

1. LA INNOVACIÓN

1.1 Motivación

Las redes sociales representan una fuente valiosa de información que alberga una amplia variedad de opiniones en forma de comentarios. Muchos de estos comentarios expresan preocupaciones y demandas vinculadas al ámbito educativo peruano. Obtener *insights*¹ de estos comentarios resulta todo un desafío debido a la complejidad para analizar este tipo de información no estructurada, con el fin de convertirla en conocimientos concretos que puedan ser de utilidad para las autoridades educativas.

Tener acceso a este conocimiento nos ayudaría a visibilizar los siguientes escenarios:

1. En primer lugar, se podría visibilizar muchos aspectos relevantes, como las inquietudes de los usuarios, patrones de demanda de necesidades públicas del sector, el interés que tiene la población por ciertos temas e incluso la existencia de noticias falsas sobre las políticas educativas del Perú, con la oportunidad de ser aclaradas por el Minedu.
2. En segundo lugar, es crucial tener en cuenta que muchas de estas opiniones son expresadas por usuarios influyentes en comunidades digitales². Si no los identificamos con antelación, estaríamos desaprovechando la oportunidad de generar un mayor *engagement*³ en los mensajes del Minedu, ya que no estarían contextualizados con los términos o expresiones empleados por estas comunidades.
3. Por último, muchas de estas opiniones expresan emociones que, al ser ignoradas, ocasionaría una desconexión emocional y una visión poco veraz de la percepción pública sobre temas del sector de educación en el Perú.

En resumen, contar con este conocimiento podría llevar al Ministerio de Educación a tener información valiosa y oportunidades de mejora para tener una mejor conexión con la comunidad educativa y la sociedad.

De acuerdo con el Reglamento de Organización y Funciones (ROF) del Minedu, aprobado por Decreto Supremo N° 001-2015-MINEDU, la Oficina de Seguimiento y Evaluación Estratégica (OSEE) de la Secretaría de Planificación Estratégica (SPE) es la responsable de coordinar el proceso de producción, integración y análisis de información estadística, de desempeño y de impacto de la política educativa.

En el marco de estas competencias, la OSEE propuso una solución que atienda la necesidad de procesar y transformar información no estructurada de redes sociales en conocimientos accionables para la gestión. La solución planteada corresponde a la innovación denominada «Alertas para la Vigilancia del Sistema Educativo» (Avisé), el cual emplea técnicas de «*Machine Learning* y Minería de Grafos»⁴ para generar un reporte que brinde información clave para el análisis de las temáticas, los usuarios influyentes y el sentimiento de los comentarios que se originan en redes sociales.

De esta forma, la innovación busca ser una herramienta costo-efectiva⁵ que ayude en las estrategias comunicacionales de la institución al ofrecer una visión complementaria de las preocupaciones y necesidades de la población expresada en redes sociales, permitiendo así que las políticas estén mejor sintonizadas con las demandas de la sociedad.

¹ Un *insight* es una revelación o comprensión profunda que surge de la interpretación y análisis detallado de datos o información, proporcionando una visión más rica y perspicaz sobre un tema o situación, lo que a menudo conduce a una toma de decisiones más informada y estratégica.

² Dentro del marco de este documento, el término «comunidad digital» se refiere a un conjunto de usuarios en una red social que comparten uno o varios intereses comunes (ver Anexo 2).

³ El *engagement* es el grado de participación activa e interés genuino que una audiencia o comunidad muestra hacia una actividad o tema de interés, lo que puede manifestarse en una conexión emocional.

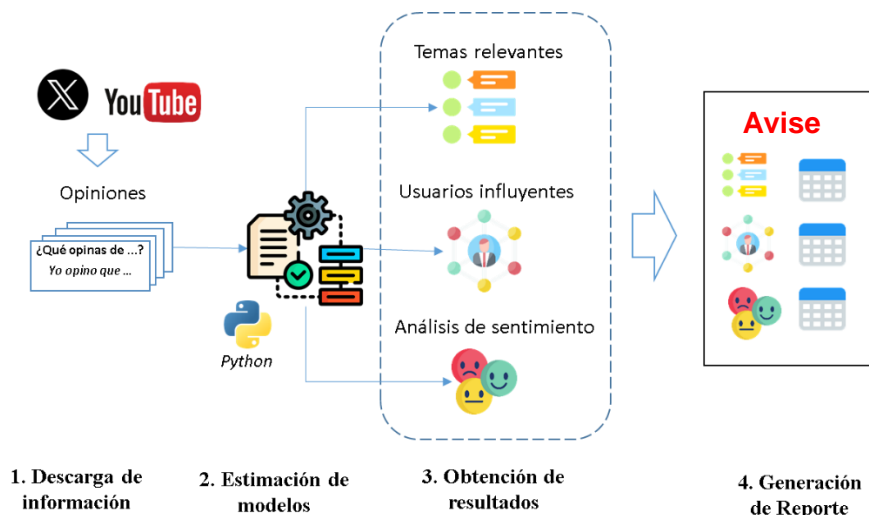
⁴ Ver anexo 2 para obtener información más detallada sobre los términos mencionados.

⁵ Avisé ha sido desarrollado haciendo uso únicamente de aplicaciones y algoritmos de código abierto, los cuales están disponibles de manera gratuita. Además, la información utilizada proviene de datos públicos de las redes sociales.

1.2 Metodología de la Innovación

Para elaborar el reporte de Avise a partir de datos recopilados en redes sociales, es esencial aplicar una serie de procesos de transformación de información y modelos analíticos. Además, resulta imperativo mantener un flujo de información que se integre con los modelos analíticos a fin de facilitar la generación de *insights*. La Figura 1 proporciona una visión general de los pasos de la metodología utilizada en la construcción de Avise:

Figura 1. Pasos para la generación del reporte Avise



1. Se procede a la descarga automatizada de los comentarios provenientes de las redes sociales.
2. Los comentarios son sometidos a procesos analíticos para estimar modelos que posibiliten la identificación de temáticas, usuarios influyentes, comunidades digitales y el sentimiento de los comentarios.
3. Se generan los resultados obtenidos por cada modelo analítico.
4. Estos resultados se integran de manera coherente en el reporte Avise.

A continuación, se profundizará en los modelos analíticos utilizados, los cuales desempeñaron un papel crucial en la identificación de temáticas relevantes, usuarios influyentes, comunidades digitales y el análisis del sentimiento de los comentarios.

1.2.1 Identificación de Temas Relevantes

La identificación de temas de interés ha representado una actividad relevante para entidades gubernamentales y líderes políticos, ya que les permite discernir los temas predominantes en las conversaciones sostenidas en las redes sociales y así ajustar sus estrategias de comunicación. Por ejemplo, Iyer (2017) realizó un estudio que empleó técnicas de aprendizaje supervisado y análisis de grafos para identificar temas de interés presentes en los tweets publicados por los representantes gubernamentales de los EE.UU para conocer la agenda política vigente (Iyer y otros, 2017).

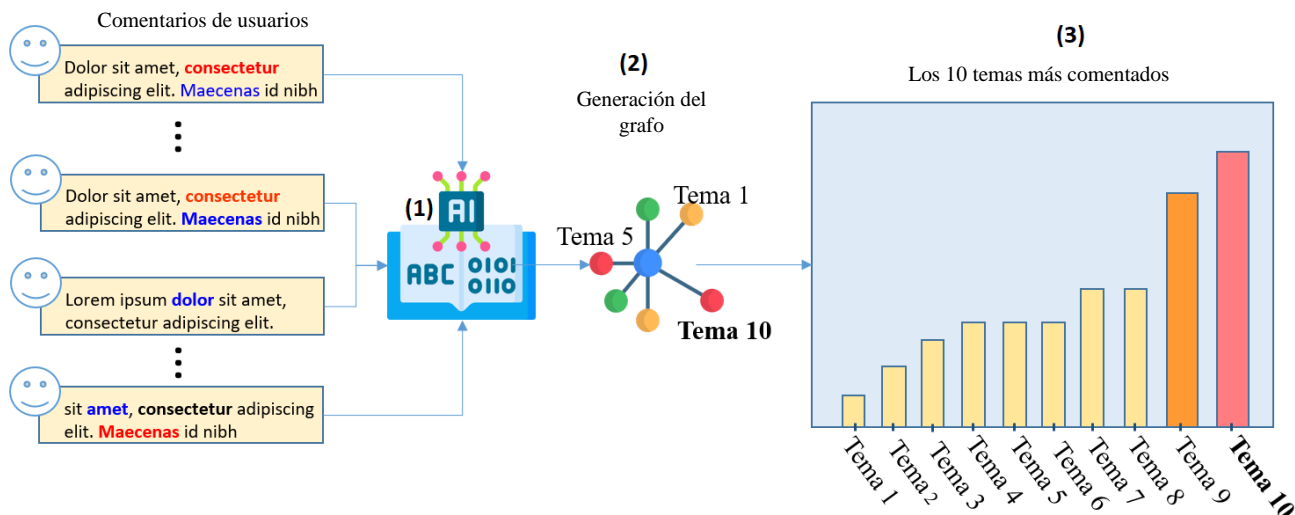
En el Minedu, se ha llevado a cabo una experiencia similar, en la que se utilizó el análisis de grafos para identificar los temas que existen en las respuestas de preguntas abiertas provenientes de instrumentos de recojo de información (Candela Rojas & Cañari Huerta, 2022). Avise aprovecha esta reciente experiencia al implementarla en el ámbito de las redes sociales, construyendo un grafo que facilita la inferencia de los temas más relevantes.

Para la concepción del grafo se consideró que cada nodo representa el texto de un comentario o tweet. La representación de cada arista que une dos nodos diferentes se apoyó en la correlación o similitud entre los textos asociados a cada nodo. El peso asignado a cada arista se estableció en función del valor de la correlación.

Para automatizar la construcción del grafo, se emplearon técnicas de «procesamiento del lenguaje natural»⁶ (NLP, por sus siglas en inglés) y algoritmos de minería de grafos. Se utilizó el algoritmo Louvain⁷ para agrupar los nodos con textos similares, donde cada grupo representa una temática.

La Figura 2 ilustra el proceso descrito, donde las temáticas identificadas se representan en un gráfico de barras organizadas en función de la cantidad de comentarios asociados a cada una.

Figura 2. Identificación de Temas Relevantes



1. Se descarga la información de los comentarios y se le aplican técnicas de NLP.
2. Se construye el grafo agrupando los nodos según la similitud en sus textos.
3. Se presenta visualmente el listado de los 10 temas más comentados.

Beneficios:

Al aplicar este proceso en el sector educativo del Perú, se podría obtener los siguientes beneficios:

- **Abordar rápidamente inquietudes y preguntas de los usuarios:** Facilita una respuesta ágil y precisa a las preocupaciones y preguntas planteadas por los usuarios.
- **Desmentir noticias falsas relacionadas con las políticas del sector:** Ayuda a detectar y refutar de manera efectiva noticias falsas o desinformación relacionadas con las políticas educativas o acciones que realiza el Minedu.
- **Identificar patrones en la demanda de necesidades públicas del sector:** Permite reconocer patrones consistentes en las demandas y necesidades públicas dentro del sector educativo, lo que permite una respuesta más precisa a las demandas de la comunidad digital.
- **Revisar y ajustar las prioridades y estrategias de comunicación:** Brinda información valiosa para la reevaluación de prioridades y objetivos en las estrategias de comunicación del Minedu, garantizando que estén alineadas con las preocupaciones y expectativas actuales de la sociedad.
- **Mejorar la focalización de difusión:** Mejora la efectividad de las campañas comunicacionales al entender los intereses de la ciudadanía, lo que posibilita adaptar los mensajes y llegar a un público objetivo más receptivo.

⁶ El procesamiento del lenguaje natural (NLP) es un área de la informática que emplea métodos para analizar, modelar y comprender el lenguaje humano (Vajjala et al., 2020).

⁷ Algoritmo empleado para identificar comunidades o agrupaciones de nodos en un grafo (ver Anexo 2).

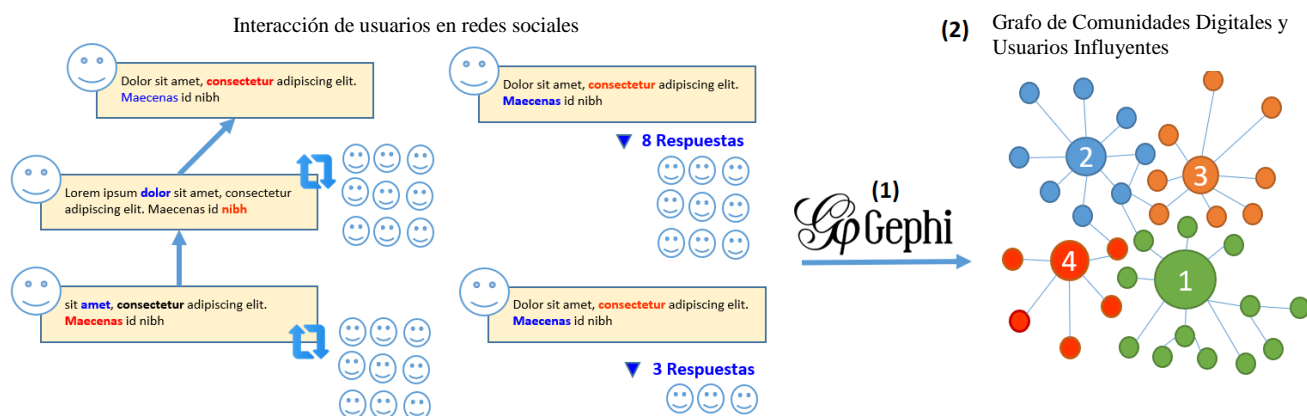
1.2.2 Identificación de usuarios influyentes y comunidades digitales

El análisis de la influencia de usuarios es un tema que se ha investigado en diversas áreas del sector público y privado empleando técnicas desde análisis cualitativos, cuantitativos y hasta métodos más sofisticados de análisis de influencia social. Un ejemplo de ello es el estudio que buscó identificar usuarios influyentes sobre calentamiento global empleando datos de la red social de Twitter (Leon Payano, 2019).

Este tipo de análisis también se ha realizado en un contexto peruano. Reyes Silva (2019) identificó a los usuarios influyentes sobre temas relacionados a la defensa de derechos de las mujeres por el día internacional de la mujer, aplicando técnicas de procesamiento del lenguaje natural y análisis de grafos en la red social de Twitter (Reyes Silva, 2019).

En Avisa, la representación de usuarios influyentes y comunidades digitales se plasma en un grafo. Cada nodo en este grafo representa a un usuario, y las aristas simbolizan las interacciones entre ellos. En el ámbito de Twitter, estas interacciones se manifiestan mediante los *retweets*, mientras que en el caso de YouTube, se evidencian cuando un usuario comenta en respuesta a un comentario existente o a un video. La Figura 3 brinda un resumen gráfico de este procedimiento.

Figura 3. Detección de comunidades digitales y usuarios influyentes.



1. Empleamos Gephi⁸ para identificar grupos de usuarios a partir de sus interacciones mediante *retweets* o comentarios en publicaciones.
2. Representamos visualmente el grafo que incluye todos los nodos y aristas, destacando las comunidades digitales diferenciadas por colores y los usuarios influyentes con los nodos más grandes.

Una ventaja de representar las interacciones en un grafo es que permite organizar⁹ a los usuarios en grupos. En el ejemplo referencial de la Figura 3, existen cuatro grupos¹⁰ de usuarios diferenciados por colores (celeste, naranja, rojo y verde), donde cada uno representa una comunidad digital.

Además, esta herramienta también posibilita la identificación de usuarios influyentes, quienes desempeñan un importante papel incentivando las interacciones del grafo, y cuya influencia se puede cuantificar mediante una variedad de indicadores de centralidad¹¹. En el ejemplo de la Figura 3, los nodos más influyentes (nodos 1, 2, 3 y 4) se muestran con un tamaño significativamente mayor.

⁸ Herramienta especializada para la exploración y análisis de grafos.

⁹ El criterio empleado para ordenar los nodos en el grafo se denomina *Fruchterman-Reingold* (ver Anexo 2).

¹⁰ Un grupo o módulo es un conjunto de nodos que están interconectados entre sí y pocos conectados con otros grupos. Avisa emplea el algoritmo *Louvain* para realizar esta agrupación (ver Anexo 2).

¹¹ La centralidad es una medida que se utiliza para evaluar el grado de influencia de un nodo dentro del grafo. Avisa emplea el Grado y *Page Rank* como indicadores de centralidad (ver Anexo 2).

Beneficios:

A continuación, se listan algunos de los beneficios:

- Mejorar la focalización de estrategia comunicacional: Facilitará la visibilidad de las comunidades de usuarios influyentes con el objetivo de mejorar la focalización de la estrategia comunicacional a las comunidades más relevantes y de interés para el sector educativo.
- Ampliar el alcance de la audiencia: Al comprender los intereses específicos de las comunidades digitales, se podría generar publicaciones más pertinentes y de mayor interés.
- Mejorar el *engagement* de las publicaciones: Al crear contenido relevante para las comunidades digitales, se estimula una mayor participación por parte de los usuarios, incluyendo la posibilidad de que los usuarios influyentes de las comunidades digitales lo compartan con un público más amplio.

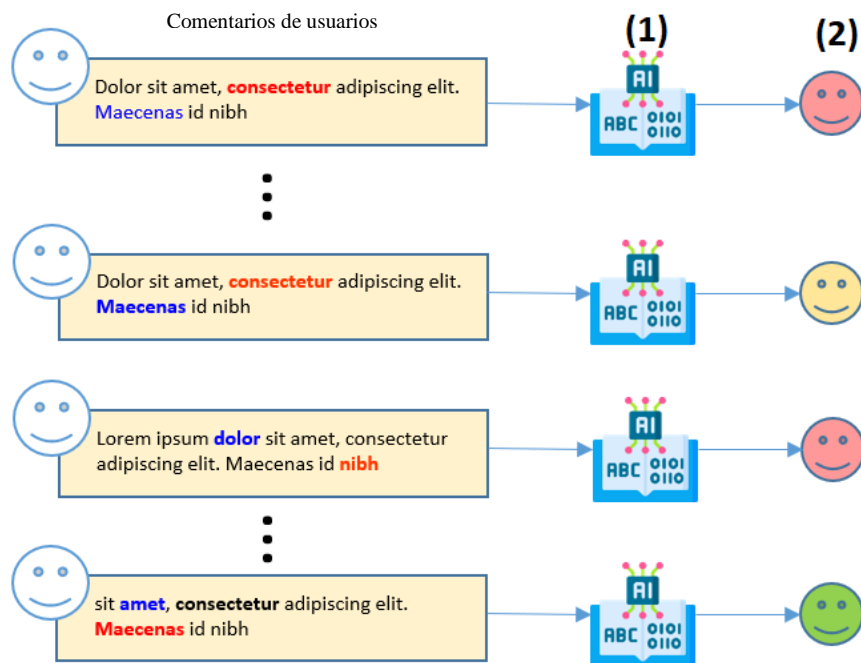
1.2.3 Análisis de sentimiento

El análisis de sentimiento permite determinar la actitud, emoción o polaridad expresada en un fragmento de texto. Este tipo de análisis resulta útil para comprender la percepción del público en términos de sentimiento expresado en los comentarios que realizan los usuarios en las redes sociales.

Un ejemplo de lo anterior es el trabajo realizado por el Banco Mundial y Global Pulse quienes colaboraron en un proyecto de investigación que analizaba el contenido y el sentimiento de los tweets para comprender mejor la opinión pública en torno a las reformas de políticas en un subsidio nacional al gas propano en el Salvador (UN Global Pulse, 2015).

En Avise, se analiza cada comentario de los usuarios y los clasifica en Positivo, Negativo o Neutro como se muestra de forma referencial en la Figura 4.

Figura 4. Clasificación de comentarios en Positivo, Negativo o Neutro.



1. Se someten los comentarios de los usuarios a un análisis de sentimiento.
2. Este proceso resulta en la clasificación de cada comentario como Positivo (marcado en verde), Negativo (señalado en rojo) o Neutral (identificado en amarillo).

Beneficios:

Este tipo de análisis brinda los siguientes beneficios:

- Percepción pública: Al ser un termómetro de los sentimientos de las redes sociales, permite una mejor comprensión de la percepción pública hacia la educación en cierto periodo, lo que puede ayudar a identificar áreas de mejora y fortalezas.
- Evaluación por recepción de políticas: Facilita la evaluación de la recepción de políticas educativas y programas mediante el análisis de sentimiento en los comentarios, proporcionando orientación para implementar ajustes y mejoras.
- Prevención de crisis: Ayuda con la identificación de señales de alerta temprana de sentimientos negativos que puede significar la aparición de crisis o problemas mayores en el sistema educativo.

1.3 Estrategia de análisis

En el punto 1.2 se exploró el panorama integral de la innovación, describiendo sus distintos procesos de transformación de información y los modelos analíticos empleados. En este nuevo punto, se propone la elaboración de una estrategia de análisis que utilice la información generada de los modelos analíticos. El objetivo es discernir *insights* significativos que, posteriormente, puedan convertirse en conocimiento práctico y aplicable para la gestión. A continuación, se detallarán los dos tipos de análisis que se llevarán a cabo en el caso de uso que se explora en el punto 2 del documento.

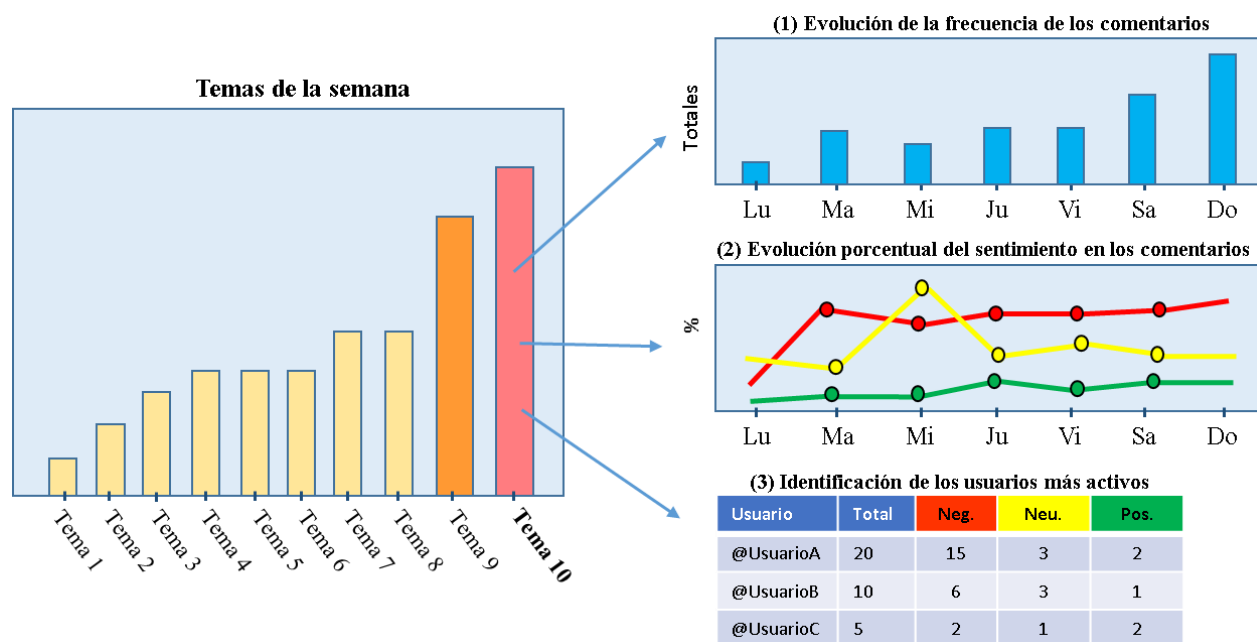
1.3.1 Análisis de la temática más relevante,

Un análisis más exhaustivo comienza con la evaluación del tema más debatido para comprender su evolución semanal, la proporción de los sentimientos negativos en cada día y la identificación de los usuarios más activos¹² en los comentarios. En la Figura 5 se ilustra cómo a partir de la identificación de un tema ampliamente discutido (Eje. Tema 10), se lleva a cabo un análisis más profundo que se detalla a continuación:

1. Evolución de la frecuencia de los comentarios: Este primer análisis se centra en la evaluación de cómo varía la cantidad de comentarios relacionados con el «Tema 10» a lo largo de una semana. Este enfoque proporciona una valiosa visión de la tendencia y participación ciudadana asociada a este tema.
2. Evolución porcentual del sentimiento en los comentarios: El segundo análisis se enfoca en analizar las fluctuaciones en los sentimientos expresados en los comentarios del «Tema 10». A medida que avanza la semana, se calcula el porcentaje de comentarios que son negativos, neutrales y positivos. Este enfoque proporciona una comprensión más profunda de cómo evolucionan las actitudes y emociones de los usuarios durante la semana de análisis.
3. Identificación de los usuarios más activos: En este tercer aspecto, se destaca quiénes son los usuarios más activos en términos de la cantidad de comentarios que han realizado sobre el «Tema 10». Además, se incluye un desglose detallado del número total de comentarios positivos, negativos y neutrales que estos usuarios han generado en relación con el «Tema 10».

¹² Un usuario activo se refiere a aquel que comenta regularmente sin depender de respuestas o *retweets*, distinguiéndose así del usuario influyente.

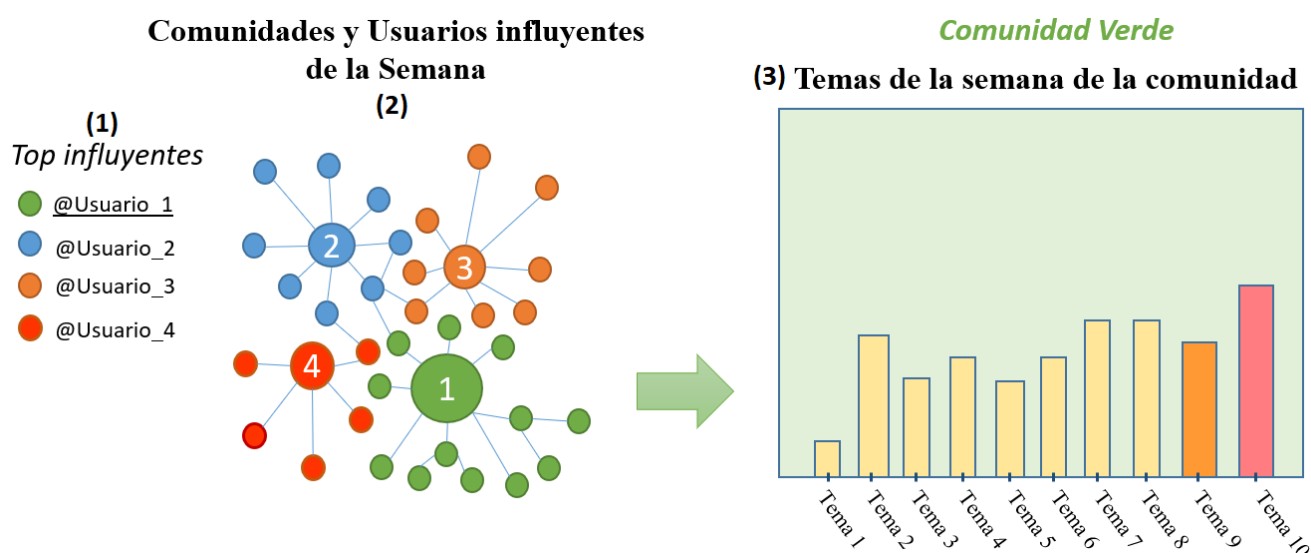
Figura 5. Análisis detallado de la temática más resaltante



1.3.2 Análisis de la comunidad digital más influyente.

Un aspecto igualmente significativo se vincula con los intereses de las comunidades digitales en las redes sociales y los usuarios influyentes que impulsan las conversaciones en estos entornos. Como se puede apreciar en la Figura 6, a partir de los miembros de la comunidad de color verde, que es la más predominante en este ejemplo, es posible identificar los temas más discutidos y quién es el usuario más influyente que dinamiza todas las temáticas de interés en dicho espacio.

Figura 6. Temas de interés de la comunidad más influyente.



1. Los usuarios más influyentes por recibir mayor cantidad de respuestas o *retweets*.
2. Comunidades digitales formadas por las interacciones de los usuarios.
3. Los 10 temas principales de lo que están hablando la comunidad digital más grande.

2. CASO DE USO

En esta sección, se explorará un caso práctico que demuestra cómo Avise puede generar conocimientos accionables que contribuyan a mejorar las estrategias de comunicación del Minedu utilizando datos de las redes sociales de Twitter y YouTube.

2.1 Datos

Para el presente caso, se descargaron publicaciones o comentarios que realizaron los usuarios que van desde el 03/07/2023 hasta el 09/07/2023 (semana de análisis en adelante) de las redes sociales de Twitter y YouTube. Para la descarga de esta información se emplearon los criterios descritos en la Tabla 1.

Tabla 1. Criterios para la descarga de información.

N°	Red Social	Criterios
1.	Twitter	Alcance: <i>Tweets</i> publicados dentro del territorio peruano. Filtro: <i>Tweets</i> que contienen al menos un <i>keyword</i> del Anexo 1.
2.	YouTube	Alcance: Comentarios y títulos de videos de canales enfocados en el sector educativo del Perú. Filtro: Comentarios y títulos que contienen al menos un <i>keyword</i> del Anexo 1.

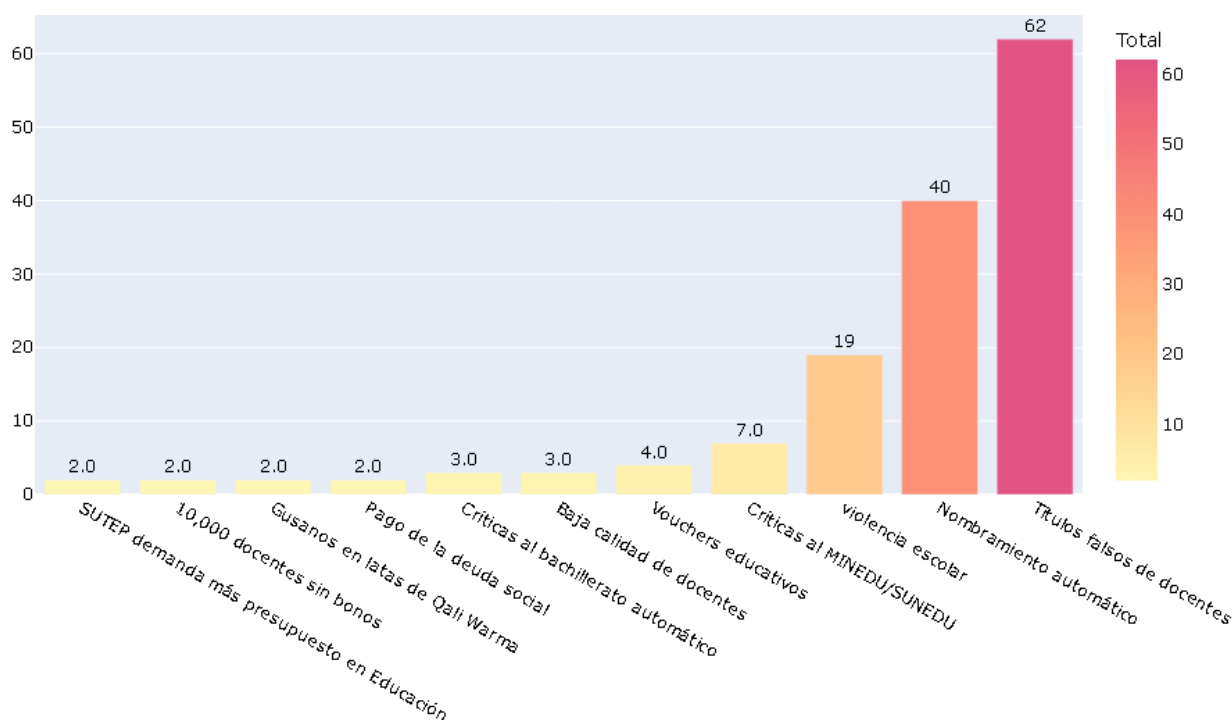
2.2 Análisis de la Red Social de Twitter

En este punto se describe el «Análisis de la temática más relevante» y «Análisis de la comunidad digital más influyente», tal como se detallaron en los puntos 1.3.1 y 1.3.2 previamente.

2.2.1 Análisis de la temática más relevante en Twitter

Como se muestra en la figura 7, la temática más discutida en la semana de análisis corresponde a «Títulos falsos de docentes» y es la temática más representativa, seguido del «Nombramiento automático». Es importante destacar que no se han tenido en cuenta los *retweets*, así como los *tweets* que no incluyan algún *keyword* especificado en el anexo 1.

Figura 7. Temáticas relevantes de la semana en Twitter.



Si profundizamos en el tema de los «Títulos falsos de docentes», podemos descubrir nuevos hallazgos. Por ejemplo, en la figura 8 se presenta la evolución de esta temática a lo largo de la semana de estudio, donde se observa una disminución en el interés a medida que avanza la semana.

Además, en la figura 9 se refleja el sentimiento de los comentarios, los cuales han sido mayoritariamente neutros y negativos.

Figura 8. Evolución de la frecuencia semanal del tema «Títulos falsos de docentes».

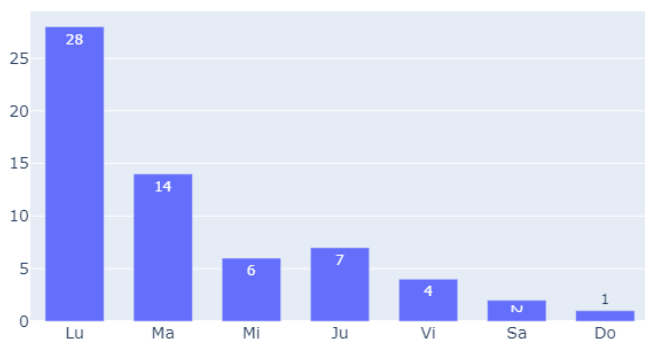


Figura 9. Evolución % del sentimiento semanal del tema «Títulos falsos de docentes».

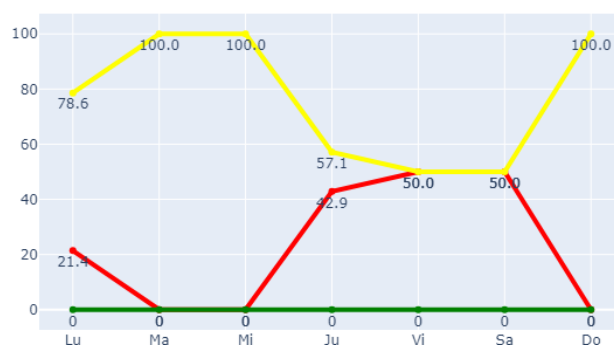


Tabla 2. Usuarios con más comentarios realizados en Twitter sobre el tema «Títulos falsos de docentes».

Usuario	Total	Neg.	Neu.	Pos.
prensachorri	5	0	5	0
elcomercio_peru	4	1	3	0
diariocorreo	3	0	3	0
Latina_pe	3	1	2	0
Latina_Noticias	3	1	2	0

Por otro lado, la tabla 2 revela que los mensajes relacionados con los «Títulos falsos de docentes» no están centralizados en un único usuario, ya que el primer usuario identificado como @prensachorri apenas realizó 5 comentarios. Le siguen @elcomercio_peru y @diariocorreo con 4 y 3 comentarios respectivamente, indicando una distribución más amplia de la discusión en múltiples cuentas.

2.2.2 Análisis de la comunidad digital más influyente en Twitter

La figura 10 muestra la interacción de los usuarios a partir de los *retweets* que recibieron sus comentarios, representados en un grafo. A diferencia del análisis de temáticas detallado en el punto 2.2.1, en el análisis de la comunidad digital se tienen en cuenta los *retweets* realizados por usuarios. Este adquiere relevancia al construir el grafo, donde los *retweets* determinan la orientación de las aristas que conectan los nodos, siendo cada nodo una representación de un usuario. Los nodos están agrupados y diferenciados por colores, donde cada color representa una comunidad digital.

Se observa que la comunidad digital de color morado tiene una mayor representatividad. Además, la tabla 3 presenta el ranking de usuarios influyentes según Grado¹³ y Page Rank¹⁴.

¹³ El grado indica el número de *retweets* que recibió el usuario influyente por sus *tweets*.

¹⁴ Page Rank indica el número de *retweets* que recibió el usuario influyente por sus *tweets* por parte de otros usuarios con cierto nivel de influencia.

Figura 10. Grafo de comunidades digitales de Twitter.

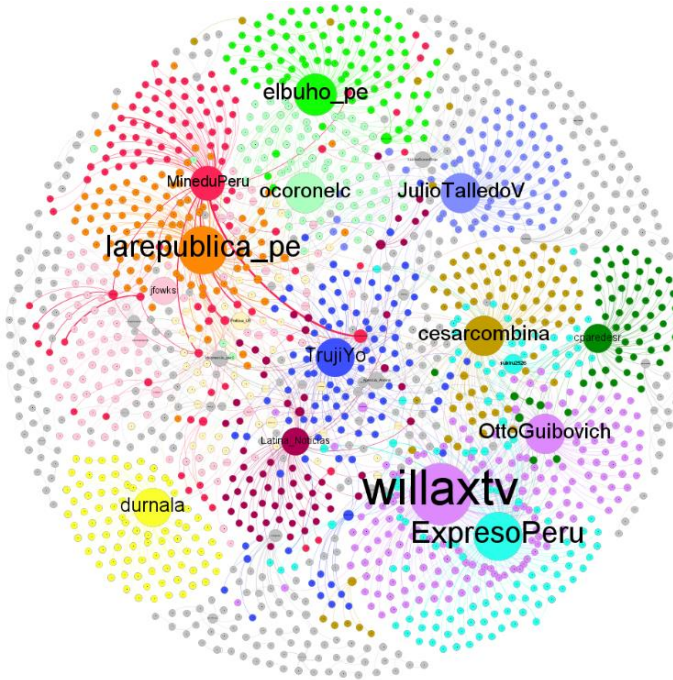


Tabla 3. Ranking de usuarios influyentes de Twitter.

N	Usuario influyente por Grado (G)	G	Usuario influyente por Page Rank (P.R)	P.R
1	willaxtv	155	elbuho_pe	0.128
2	ExpresoPeru	118	larepublica_pe	0.045
3	larepublica_pe	117	Politica_LR	0.034
4	elbuho_pe	100	willaxtv	0.028
5	cesarcombina	95	jfowks	0.026
6	ocoronele	95	durnala	0.021
7	JulioTalledoV	94	JulioTalledoV	0.019
8	OttoGuibovich	94	ExpresoPeru	0.019
9	TrujiYo	91	MineduPeru	0.019
10	Durnala	90	ocoronele	0.018

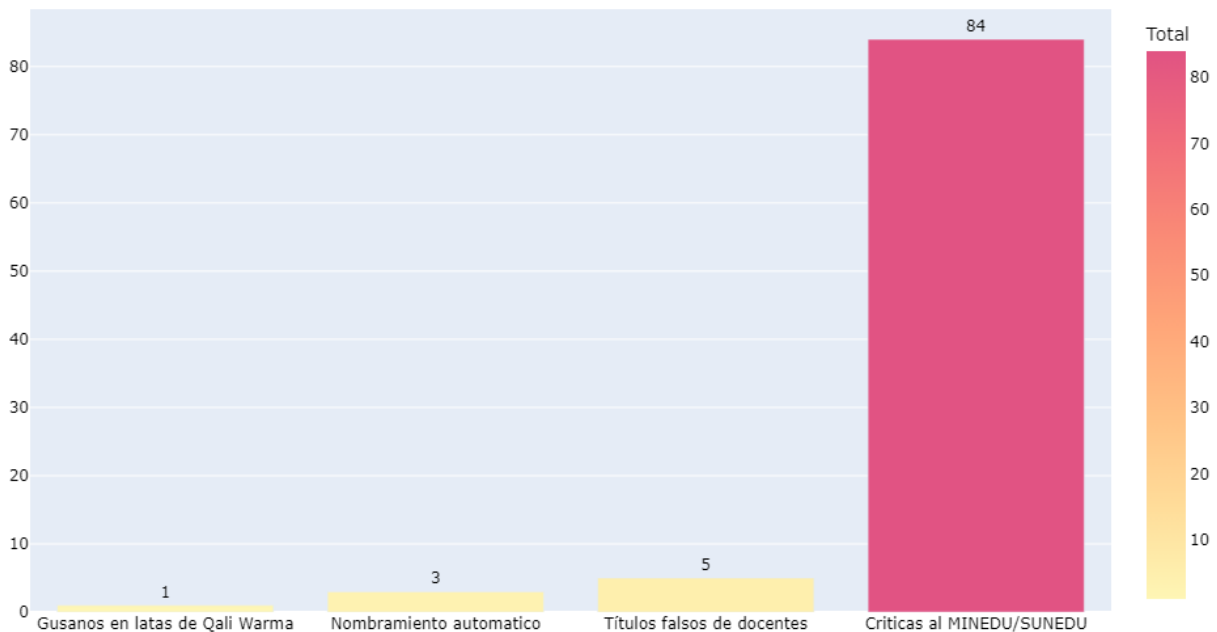
G: Grado

P.R: Page Rank

En la Tabla 3, se destaca la notable influencia del usuario @willaxtv en el ámbito educativo durante la semana de análisis, al acumular la mayor cantidad de *retweets* en respuesta a sus publicaciones. Le siguen de cerca @ExpresoPeru y @larepublica_pe. Por otro lado, el usuario @elbuho_pe resalta por recibir *retweets* de otros usuarios con un nivel destacado de influencia.

Un análisis adicional se enfoca en los intereses específicos de la comunidad identificada por el color Morado, la cual está representada principalmente por @willaxtv y @OttoGuibovich. De acuerdo con la Figura 11, la temática que ha recibido la mayor cantidad de *retweets* en esta comunidad es «Críticas al MINEDU/SUNEDU», seguida por «Títulos falsos de docentes» con una menor incidencia.

Figura 11. Temas de interés de la comunidad de color morado en Twitter.



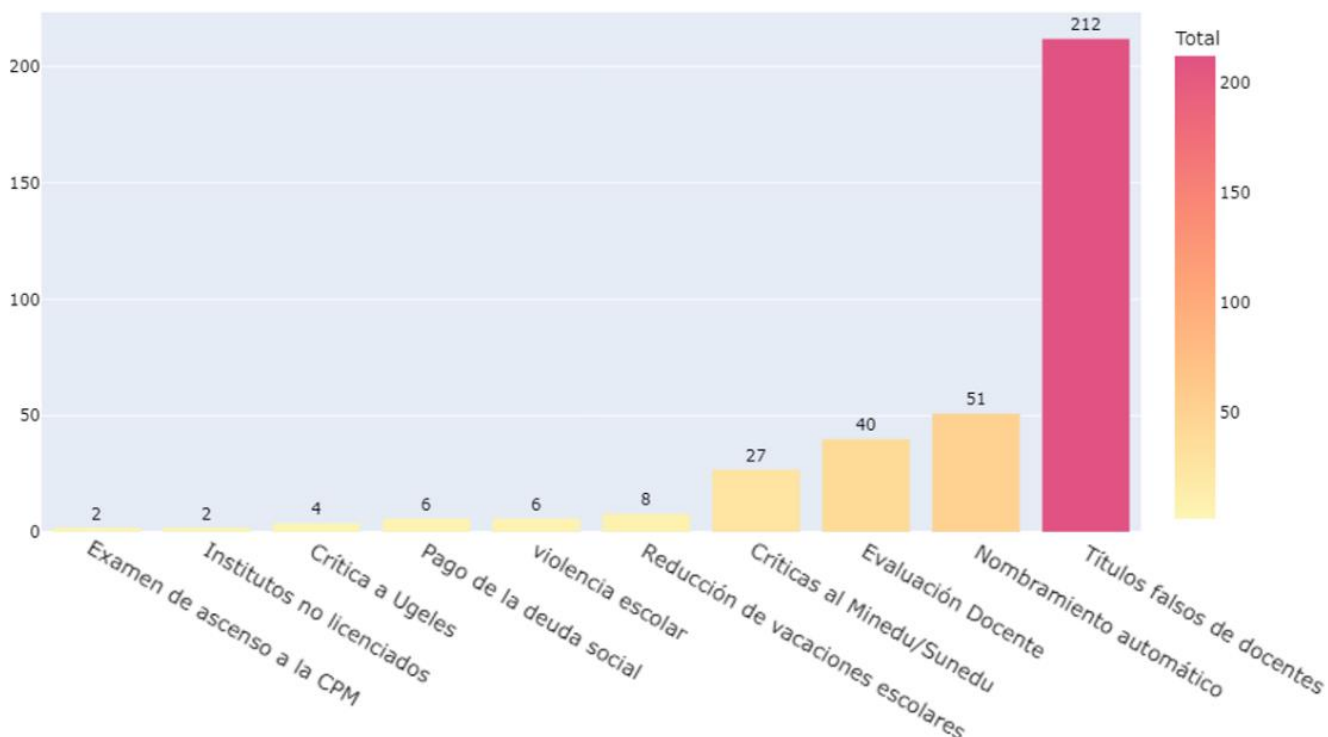
2.3 Análisis de la Red Social de YouTube

En esta sección, se abordan dos aspectos clave: el «Análisis de la temática más relevante» y el «Análisis de la comunidad digital más influyente», los cuales fueron previamente explorados en los puntos 1.3.1 y 1.3.2. Estos análisis han sido aplicados a los comentarios y títulos de los videos descargados a lo largo de la semana de análisis en la plataforma de YouTube.

2.3.1 Análisis de la temática más relevante en YouTube

Como se muestra en la figura 12, la temática más discutida corresponde a «Títulos falsos de docentes» y es la temática más representativa durante la semana de análisis, seguido del «Nombramiento automático».

Figura 12. Temas educativos en YouTube



Si profundizamos en el tema de «Títulos falsos de docentes», podemos descubrir nuevos hallazgos. Por ejemplo, en la figura 13 se presenta la evolución de esta temática a lo largo de la semana de estudio, donde se observa una disminución en el interés a medida que avanza la semana. Además, en la figura 14 se refleja el sentimiento de los comentarios, los cuales han sido mayoritariamente negativos.

Figura 13. Evolución de la frecuencia semanal del tema «Títulos falsos de docentes».

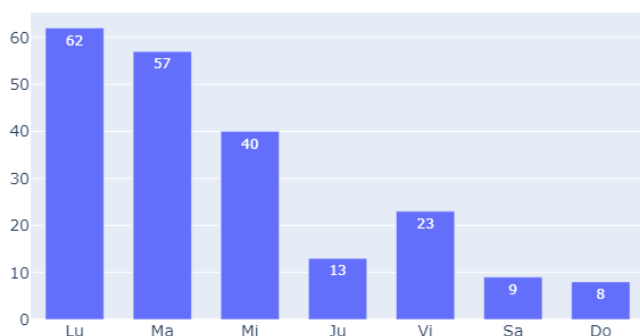


Figura 14. Evolución % del sentimiento semanal del tema «Títulos falsos de docentes».

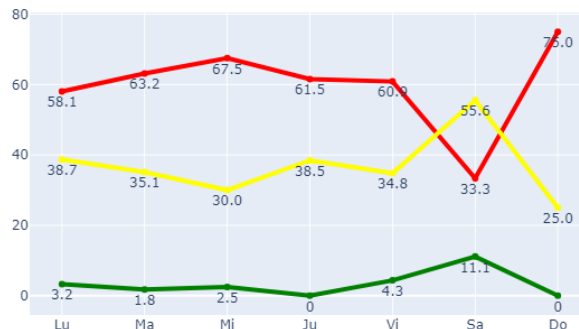


Tabla 4. Usuarios con más comentarios realizados en Twitter sobre el tema «Títulos falsos de docentes».

Usuario	Total	Neg.	Neu.	Pos.
@willyfreyre9309	8	8	0	0
@rociocac5285	4	4	0	0
@ernaldinalliuyagarcia1627	4	1	3	0
@constantinoyalverde9299	3	3	0	0
@review3666	3	3	0	0

Por otro lado, la tabla 4 revela que los comentarios relacionados con los «Títulos falsos de docentes» no están centralizados en un único usuario, ya que el primer usuario @willyfreyre9309 apenas realizó 8 comentarios. Le siguen @rociocac5285 y @ernaldinalliuyagarcia1627 con 4 comentarios cada uno. El resto de los participantes presenta un número de comentarios igual o inferior a 3, destacando así una dispersión de la conversación en varias cuentas.

2.3.2 Análisis de la comunidad digital más influyente en YouTube

La figura 15 muestra un grafo construido a partir de los datos de los usuarios de YouTube, donde las respuestas o comentarios determinan la orientación de las aristas que conectan los nodos, siendo cada nodo una representación de un usuario. Los nodos están agrupados y diferenciados por colores, donde cada color representa una comunidad digital. Se observa que la comunidad digital de color morado tiene una mayor representatividad. Por otro lado, la tabla 5 presenta el ranking de usuarios influyentes según Grado¹⁵ y Page Rank¹⁶.

Figura 15. Grafo de comunidades digitales de YouTube.

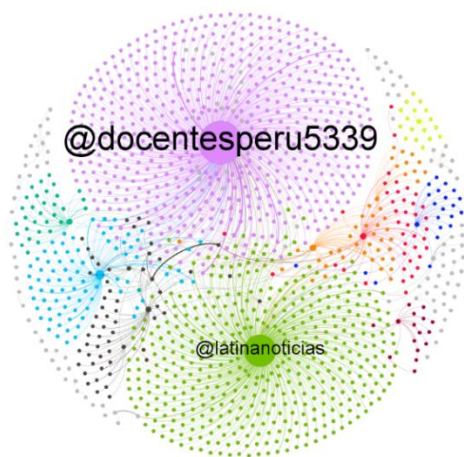


Tabla 5. Ranking de usuarios influyentes de YouTube.

N	Usuario influyente por Grado (G)	G	Usuario influyente por Page Rank (P.R)	P.R
1	@docentesperu5339	655	@bardot392	0.003
2	@latinanoticias	471	@i.e.646pequenosconstructor9	0.003
3	@educacionenaccion	89	@emperadorputin8749	0.002
4	@ElProfeFer	44	@laurafernandezloayza6906	0.002
5	@rppnoticias	43	@noetarazonaramos7732	0.002
6	@exitosape	36	@nordico04	0.002
7	@MineduPeruOficial	35	@maritzanietovarela263	0.002
8	@micanalperu2740	21	@darlingfrankcastrosudario4234	0.002
9	@PeruanoInformado	17	@enfoqueair4625	0.002
10	@ultimahoraperu	16	@lizkf	0.002

G: Grado

P.R: Page Rank

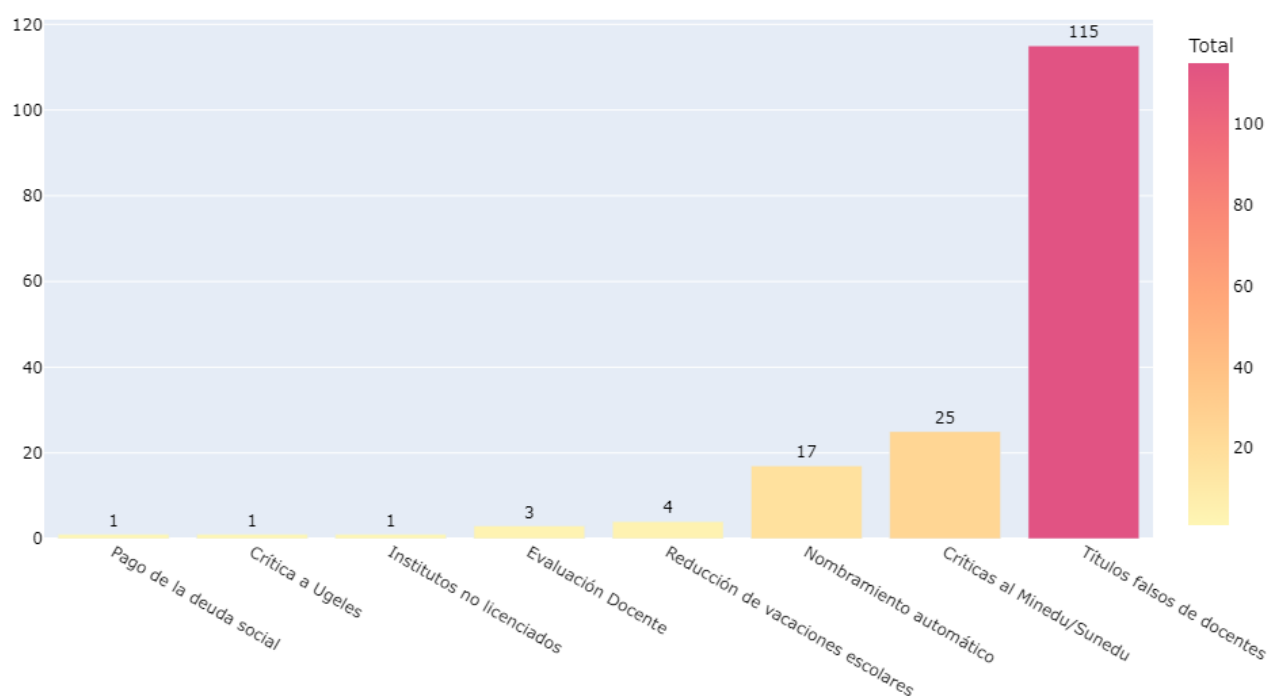
Como se detalla en la Tabla 5, el usuario @docentesperu5339 se destaca como el más influyente en el sector educativo, acumulando la mayor cantidad de respuestas o comentarios a sus publicaciones, seguido por @latinanoticias y @educacionenaccion. Asimismo, no se evidencia usuarios que influncian a otros usuarios influyentes en YouTube por los valores muy bajos de Page Rank obtenidos.

Un análisis adicional se centra en los intereses de la comunidad de color Morado, representada por @docentesperu5339. La Figura 16 revela que la temática más destacada de la comunidad de color Morado, medida por la cantidad de comentarios descargados, corresponde a los «Títulos falsos de docentes».

¹⁵ El grado indica el número de comentarios que han recibido el usuario influyente.

¹⁶ El Page Rank indica el número de comentarios que han recibido el usuario influyente por otros usuarios influyentes.

Figura 16. Temas de interés de la comunidad de color morado en YouTube.



3. CONOCIMIENTOS ACCIONABLES

A partir del caso mencionado, se listan recomendaciones prácticas con el objetivo de proporcionar pautas claras sobre cómo Avise puede contribuir al enriquecimiento de las estrategias de comunicación.

1. **Focalización de mensajes:** Se recomienda tomar en cuenta las temáticas más debatidas y de interés del sector educativo que se debate en las redes sociales. Esto contribuirá a la creación de mensajes más oportunos y útiles.

En el caso que se presenta, se identificó que el tema más comentado fue «Títulos falsos de docentes», generando un total de 62 tweets en Twitter y 212 comentarios en YouTube en una semana. Aunque se observa una tendencia a la baja, es relevante señalar un alto porcentaje de comentarios negativos. Contar con este conocimiento puede ser fundamental para tomar decisiones informadas sobre cómo ajustar la estrategia de comunicación en relación con el tema de «Títulos falsos de docentes».

2. **Evaluación de mensajes:** Es esencial seguir la evolución de la temática a lo largo del período de análisis, ya que brinda la oportunidad de comprender el impacto y la percepción de los comentarios de los usuarios frente a comunicados de la institución.

En el caso de uso analizado, se observó una disminución en la tendencia de la temática «Títulos falsos de docentes», posiblemente vinculada al comunicado 021-2023-MINEDU, donde se informó de manera oportuna sobre las acciones emprendidas por el Minedu para abordar dicha problemática.

3. **Alertas de manipulación de información:** Una alta concentración de comentarios en uno o pocos usuarios en torno a una temática específica pueda indicar ciertos sesgos o distorsiones sobre esta temática. Esto no necesariamente significa que la temática en sí sea falsa, pero podría sugerir que el debate sobre ella está siendo influida de manera desproporcionada por un individuo o grupo particular.

Para la temática «Títulos falsos de docentes» se pudo evidenciar que no hubo un grupo de usuarios que monopolizaban el debate, por lo que siguiera que la temática identificada es representativa.

4. Focalización de mensajes por comunidad digital: Durante la semana de análisis en la plataforma de YouTube, se identificó la comunidad digital de color morado como la más representativa en el ámbito educativo. Esta comunidad estaba predominantemente involucrada en el debate de la temática «Títulos falsos de docentes» y su usuario influyente era @docentesperu5339. Esta información no solo aporta un valor intrínseco como recurso informativo para entender el impacto de la temática en los usuarios, sino que también se erige como un insumo valioso que posibilita la creación de estrategias más personalizadas.

Por ejemplo, en situaciones donde, a pesar de haber difundido un comunicado institucional, la tendencia de la temática no disminuye como se anticiparía, se vuelve esencial realizar un análisis detallado en las comunidades más influyentes. Este enfoque tiene como objetivo descubrir perspectivas novedosas que aporten al desarrollo de comunicados más efectivos, ajustados a la dinámica y las necesidades emergentes de dichas comunidades.

4. CONCLUSIONES Y FUTURAS ACCIONES

- El presente informe evidencia cómo Avise facilita la generación de *insights* a partir de la exploración de datos no estructurados, como los comentarios de usuarios en las redes sociales de Twitter y YouTube. Esta capacidad simplifica de manera significativa el proceso de análisis, proporcionando una visión más clara y profunda de las opiniones de los usuarios.
- Se validó la capacidad de los modelos analíticos para identificar temas de interés, usuarios influyentes, comunidades digitales y el sentimiento de los comentarios de los usuarios en relación con temas vinculados al sector educativo en Perú.
- Se comprobó que la realización de un análisis complementario, al integrar los resultados de los modelos analíticos, posibilita la generación de nuevos conocimientos e ideas que enriquecen la comprensión de aspectos específicos. Estos hallazgos no habrían sido evidentes si se hubieran examinado de manera independiente los resultados de los modelos analíticos.
- En este documento, se ha presentado un caso de uso de como Avise busca contribuir con las estrategias de comunicación del Minedu. Sin embargo, Avise va más allá de ser simplemente una herramienta para el diseño de estrategias de comunicación. Es una solución costo-efectiva que podría desempeñar un papel crucial en el monitoreo de políticas educativas al proporcionar una perspectiva adicional sobre las preocupaciones y demandas recurrentes que los usuarios expresan en las redes sociales.
- El reporte Avise está inmerso en un proceso constante de mejora, donde se están implementando nuevas estrategias de visualización de datos y se están actualizando las metodologías analíticas utilizadas. El objetivo primordial de estas mejoras es disminuir las tasas de error en los modelos analíticos y perfeccionar el proceso de generación de *insights*.

BIBLIOGRAFÍA

- Candela Rojas, E. C., & Cañari Huerta, E. G. (2022). *Machine Learning para la categorización de respuestas de preguntas abiertas*. MINEDU. <https://repositorio.minedu.gob.pe/handle/20.500.12799/8677>
- Dugu, N., & Perez, A. (2015). Directed Louvain : maximizing modularity in directed networks.
- ElPeruano. (2023 de Enero de 2023). Usuarios peruanos incrementan consumo de redes sociales y juegos online. *Ciencia y Tecnología*.
- Fruchterman, T. M., & Reingold, E. M. (1991). Graph Drawing by Force – Directed Placement. *Software: Practice and experience*, 1129-1164.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow*. O'REILLY.
- Iyer, R., Wong, J., Tavanapong, W., & Peterson, D. (2017). *Identifying Policy Agenda Sub-Topics in Political Tweets based on Community Detection*. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017.
- Kido, G. S., Igawa, R. A., & Barbon Jr, S. (2016). *Topic Modeling based on Louvain method in Online Social Networks*. XII Brazilian Symposium on Information Systems.
- Leon Payano, M. A. (2019). *Identificación de líderes de opinión mediante el modelo PROV-DM y técnicas de minería de grafos*.
- Needham, M., & Hodler, A. (2019). *Graph Algorithms*. O'Reilly Media, Inc.
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The pagerank citation ranking: Bringing order to the web. *stanford University*.
- Reyes Silva, P. D. (2019). *Redes en Twitter y la defensa de la mujer peruana (ante el acoso y el feminicidio)*. Universitat Oberta de Catalunya (UOC).
- Samatova, N. F., Hendrix, W., Jenkins, J., Padmanabhan, K., & Chakraborty, A. (2013). *Practical graph mining with R*. CRC Press.
- Trudeau, R. J. (1993). *Introduction to Graph Theory*. Courier Corporation.
- UN Global Pulse. (2015). Using Twitter Data to Analyse Public Sentiment on Fuel Subsidy Policy Reform in El Salvador. *Global Pulse Projec*, Series, no.13.
- Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2020). *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*. O'Reilly Media.

ANEXO 1

MINEDU2023

- from:presidenciaperu educación
- from:presidenciaperu colegio
- Boluarte educacion
- dinna educacion
- dinna educativo
- año escolar
- sistema educativo
- reforma educación
- reforma educativa
- from:MineduPeru
- Ministerio Educación
- Magnet Márquez
- suspender clases
- suspension clases
- minedu
- VamosAlCole
- Fondep
- Juegos Escolares Deportivos y Paradeportivos
- Miriam Ponce
- Fátima Altabás
- Educación Intercultural
- gestión educativa
- EducaciónDeCalidad
- Gestión Pedagógica
- Miriam Ponce
- EducaciónInicialPerú
- minedupe
- ReasignaciónDocente2023
- capacitación docente
- from:dreIm_lima
- escuela privada
- escuela pública
- colegio público
- violencia escolar
- ViolenciaEscolar
- reasignación docente
- from:proniedperu
- institución educativa
- EscuelasBicentenario
- certificado estudios
- simon minedu
- escale minedu
- AñoEscolar2023
- PRONIED
- SUNEDU
- Ministra Educación
- desastre escuela
- recuperacion aprendizaje
- deserción escolar
- suspensión clases
- educación técnico productiva
- Educación Inicial
- EducaciónInicial
- SIDI
- #CADEeducación
- #Maestros
- #EducacionSinDoctrina
- #profesores
- servicio educativo
- formación docente
- matricula colegio
- estudiante discapacidad
- EducaciónInclusiva
- PNESTP
- MineduTV
- Educación Artística
- Educación Física
- salud mental escuela
- SíseVe
- Siagie
- Escuela Bicentenario
- acoso escolar
- Pronabec
- IPD
- #aprendoencasa
- educación especial
- #MinisteriodeEducación

DENGUE

- colegio dengue
- educación dengue
- escolar dengue
- minedu dengue
- alumno dengue
- alumna dengue
- escuela dengue

SUTEP

- SUTEP
- from:SUTEP_Peru
- Sindicato Único Trabajadores Educación

VOUCHER EDUCATIVO

- vaucher educativo
- voucher educativo
- váuchers educativos
- Proyecto Legislativo 2951
- váucher educativo
- cupon educativo
- cupon Milton Friedman
- Ley 2951
- P.L 2951

CNCL 2023

- CNCL 2023
- Concurso Comprensión Lectora
- #ElPerúLee!
- elperulee
- #Leerestádemoda
- 077-2023-Minedu
- from:MineduPeru lectura
- from:MineduPeru lectora

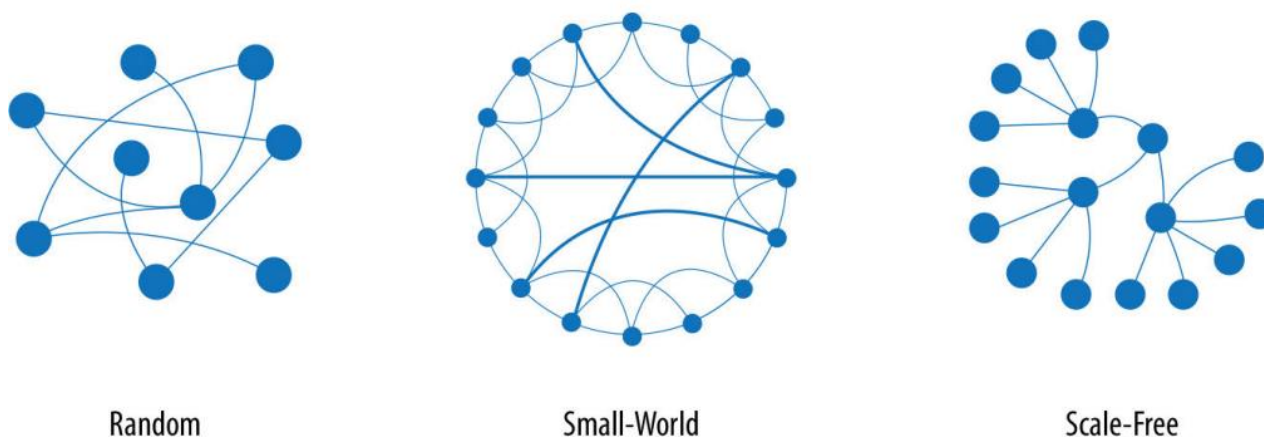
NOMBRAMIENTO AUTOMÁTICO

- Nombramiento automatico
- #NombramientoAutomatico
- meritocracia docente
- meritocracia maestro
- Reforma magisterial automático
- nombramiento maestros contratados
- reforma magisterial meritocrática
- Carrera Publica Magisterial
- Bloque magisterial nombramiento
- nombramiento docente contratados
- nombramiento profesores contratados
- nombramiento docente sin evaluación
- nombramiento docente automática
- docentes ley tres más contrato
- ley 1587
- Pasión Dávila nombramiento

ANEXO 2

- **Machine Learning (ML) o aprendizaje automático:** Es la ciencia (y arte) de programar las computadoras para que puedan aprender de los datos (Géron, 2019).
- **Grafo:** Es un conjunto de puntos y líneas. Los puntos se llaman «nodos» y las líneas se llaman «aristas». Los nodos forman un grupo finito y no vacío, mientras que las aristas, si hay alguna, son simplemente líneas que conectan dos de estos nodos. En resumen, un grafo representa conexiones entre nodos a través de aristas (Trudeau, 1993).

Figura 17. Tipos de grafos.



Nota. Adaptado de *Graph Algorithms* (p. 17), por Needham & Hodler, 2019, O'Reilly Media, Inc.

- Random network o grafo aleatorio: No cuenta con jerarquías ni patrones discernibles. Todos los nodos tienen igual probabilidad de conectarse entre sí.
 - Small-World network o grafo de mundo pequeño: Es como una red social donde, aunque tengas unos pocos amigos cercanos, es probable que estés conectado a otras personas a través de amigos en común, creando una especie de red compacta.
 - Scale-Free network o grafo libre de escala: la distribución de conexiones sigue una ley de potencia, lo que significa que hay unos pocos nodos altamente conectados y muchos nodos con pocas conexiones.
- **Minería de grafos:** Es un campo en crecimiento que tiene como objetivo descubrir conocimientos novedosos e *insight* a partir de datos que están representados en forma de grafo (Samatova et al., 2013).
 - **Centralidad:** La centralidad constituye una métrica esencial para comprender la importancia de un nodo dentro del grafo. Existen diferentes tipos de algoritmos de centralidad creados para medir la importancia (Needham & Hodler, 2019). En el desarrollo de Avise, se abordaron dos tipos de centralidad que se detallan a continuación:

Tabla 6. Tipos de centralidad.

Nº	Centralidad	Descripción	Uso
1.	Grado	Mide la cantidad de conexiones directas que un nodo tiene en un grafo (Samatova et al., 2013).	<p>Ventaja:</p> <ul style="list-style-type: none"> - Fácil de calcular y entender. - Identifica usuarios populares con muchos seguidores. <p>Desventaja:</p> <ul style="list-style-type: none"> - No considera la calidad de las conexiones. - No evalúa la influencia real de un usuario en la red
2.	Page Rank	Se fundamenta en el estudio de Page, Lawrence, et al (1998). En el contexto de un grafo, <i>Page Rank</i> evalúa la centralidad en un grafo asignando valores a los nodos en función de cuántos y qué tan importantes son los enlaces que reciben de otros nodos en el grafo.	<p>Ventaja:</p> <ul style="list-style-type: none"> - Se enfoca en la calidad de los enlaces entrantes en lugar de simplemente contar la cantidad de enlaces. <p>Desventaja:</p> <ul style="list-style-type: none"> - Computacionalmente costoso - Requiere mayor tiempo para su ejecución.

- **Detección de Comunidades:** Comprende la identificación y análisis de conjuntos cohesivos de nodos dentro de un grafo. Esto permite analizar diferentes tipos de comunidades dentro de un grafo para comprender las tendencias de los grupos para atraer o repeler a otros nodos (Needham & Hodler, 2019).
- **Comunidad digital:** En el contexto de Avise, cada conjunto cohesivo de nodos identificado en la detección de comunidades representa una comunidad digital.
- **Modularidad:** La modularidad es un enfoque utilizado para identificar comunidades al segmentar un grafo en módulos (o *clusters*) más amplios y luego evaluar la intensidad de estas agrupaciones. En contraste con la simple observación de la concentración de conexiones dentro de un *cluster*, este método compara las densidades de relaciones en *clusters* particulares con las densidades entre *clusters*. La métrica que evalúa la calidad de estas agrupaciones se conoce como modularidad (Needham & Hodler, 2019).
- **Louvain:** Algoritmo empleado para identificar comunidades o agrupaciones de nodos en un grafo. Dentro de una comunidad, los nodos presentan una densidad de conexiones significativamente superior en comparación con los nodos externos a esa comunidad (Dugu & Perez, 2015). Se cuenta con evidencia que respalda el uso de Louvain para la extracción de temas en plataformas sociales en línea, tales como Twitter y YouTube (Kido y otros, 2016).
- **Fruchterman-Reingold:** Algoritmo de disposición o posicionamiento de nodos en un grafo. Se basa en la idea de que los nodos son partículas cargadas eléctricamente que se repelan entre sí, mientras que sus aristas son como resortes que mantienen unido a los nodos. Bajo esta idea, esta técnica permitirá distribuir los nodos en un plano de tal forma que aquellos nodos que tienen más conexiones estarán más agrupados, mientras que los nodos con menos conexiones estarán aislados (Fruchterman & Reingold, 1991).