FastAnalytics S.A.S

El futuro habla en datos

¡Lo anticipamos con decisiones!

Presentado en

2do Workshop Internacional de estadística

5 y 6 noviembre de 2026

FCEN Facultad de Ciencias Exactas y Naturales

Universidad Nacional de Colombia

Reconstrucción de redes criminales espaciotemporales para su disrupción en la Amazonía

De comunicaciones parciales a vértices y aristas no observados: construcción matemática, evidencia estadística y decisión aplicada

Ponente Andrés Pérez-Coronado

Coautores: Gustavo Canavire-Bacarreza · Santiago Ortíz · Henry Laniado

FastAnalytics S.A.S · Presentación científica — versión para audiencia matemática

2do Workshop Internacional de Estadística 2026

Universidad Nacional de Colombia, Sede Manizales · Métodos, Modelamiento y Aplicaciones

FastAnalytics S.A.S

El futuro habla en datos

¡Lo anticipamos con decisiones!

PATROCINADOR

2do Workshop Internacional de estadística

5 y 6 noviembre de 2026

UNIVERSIDAD NACIONAL DE COLOMBIA

El futuro habla en datos.

¡Lo anticipamos con decisiones!

FastAnalytics es una empresa de investigación, ciencia y tecnología que combina el modelado matemático, la analítica de datos y la inteligencia artificial para generar predicción estratégica.

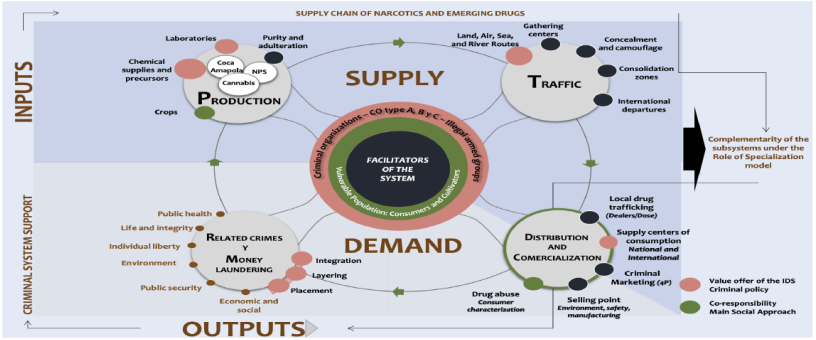
MODELADO MATEMÁTICO

ANALÍTICA DE DATOS · INTELIGENCIA ARTIFICIAL

fastanalytics.co



El sistema de drogas ilícitas (IDS): dónde vive la red



Illicit Drug System (IDS), built and translated from [Rodríguez, 2023].

Síntesis. El IDS muestra que el tráfico es un eslabón de un sistema; nuestra red es su capa logística.

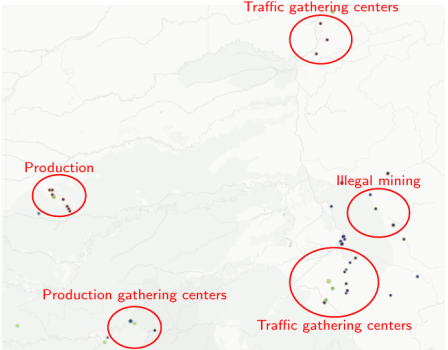
El área de estudio: la Amazonía compartida por Colombia, Venezuela y Brasil





Síntesis. ≈ 706 mil km² de selva: 40 % mayor que España, 12 % menor que Turquía.

El territorio habla: zonas de alto agrupamiento



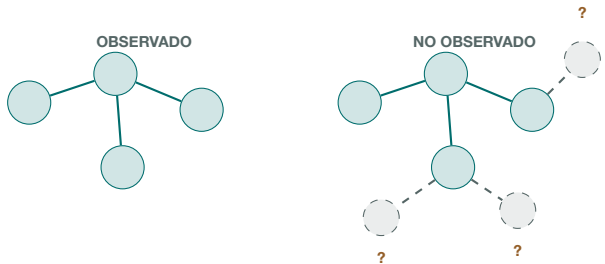
Síntesis. La coincidencia entre alto clustering e infraestructura es evidencia espacial independiente de la predicción.

La pregunta se formula sobre vértices y aristas no observados

¿Podemos reconstruir una red criminal espaciotemporal prediciendo sus vértices y aristas no observados?

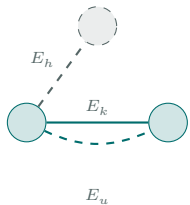
$G = (V, E)$ con $E \subseteq V \times V$ (dirigida: el orden importa)

Vértices: celdas con comunicaciones · Aristas: pares ordenados (u, v) de actividad consecutiva



La tipología separa lo observado de lo que se va a inferir

Elemento	Notación	Definición
Vértice conocido	V_k	Ubicación observada en inteligencia
Arista conocida	E_k	Enlace temporal entre vértices conocidos
Arista desconocida	E_u	Predicha entre vértices conocidos
Arista oculta	E_h	Involucra al menos un vértice oculto
Vértice oculto	V_h	Ubicación inferida mediante aristas ocultas
Vértice desconocido	V_u	Fuera del alcance geográfico del estudio



sólida: conocida
dashed teal: desconocida
gris punteada: oculta

Síntesis. Una arista oculta involucra al menos un vértice oculto; no necesariamente dos.

586 comunicaciones definen la ventana observable de 2018

586	6	304	2018
registros HF	grupos A–F	días activos	2 ene – 31 dic

LIMITACIONES DE OBSERVACIÓN

ubicación aproximada · múltiples hablantes · transmisores móviles · territorios superpuestos · afiliaciones duales · áreas sin monitoreo.

Síntesis. Cada registro contiene tiempo y ubicación; su ausencia no demuestra ausencia de actividad.

El dato básico es un evento ordenado en espacio y tiempo

$$e_a = (l_a, t_a), \qquad l_a = (x_a, y_a)$$
$$t_1 \leq t_2 \leq \cdots \leq t_{N_E}$$

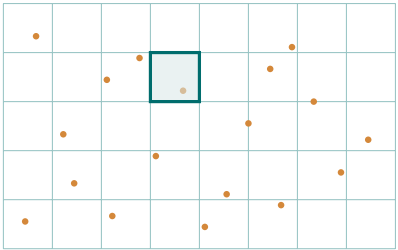


el **tiempo** ordena los eventos antes de conectarlos

Cada comunicación aporta una **ubicación aproximada** y una **marca temporal**. El orden cronológico precede a cualquier conexión.

Síntesis. Primero se ordenan los eventos; después se agregan en celdas y ventanas.

La resolución espacial controla la agregación de eventos



la **celda** (0,03°) define el vértice

≈ 3,33 km · área ≈ 9,6 km^2

Celda	Mín.	Q1	Med.	Media	Q3	Máx.
0,01°	1	1	1	1,24	1	12
0,02°	1	1	1	1,48	1	22
0,03°	1	1	1	1,76	2	25
0,04°	1	1	1	2,00	2	25
0,05°	1	1	1	2,20	2	39
0,06°	1	1	1	2,47	2	43

Elegida: 0,03° ≈ 3,33 km · área ≈ 9,6 km^2

Síntesis. La celda define el vértice: una resolución mayor agrega más eventos, pero reduce precisión espacial.

Las conexiones ocurren entre días activos consecutivos

$$I_k = [(k-1)\Delta t,\ k\Delta t)$$

$K = 304$ días activos

$\Delta t = 1$ día



Síntesis. Un salto une días activos consecutivos, no necesariamente días calendario consecutivos.

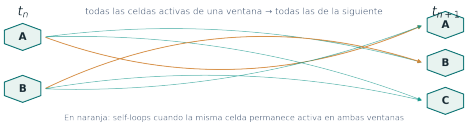
Conexión completa entre celdas activas de t_n y t_{n+1}

PESO DE CADA PAR ORDENADO · ECUACIÓN 1

$$w_{ij} = \sum_{k=1}^{K-1} n_i^{(k)}\, n_j^{(k+1)}$$

$n_i^{(k)}$: eventos de la celda i en el día activo k.
$K = 304$: días activos de 2018.
El producto agrega *coactividad*; la suma acumula transiciones.

IMAGEN · REGLA COMPLETA ENTRE VENTANAS



cada origen activo se conecta con **todos** los destinos activos; la misma celda puede formar un self-loop (naranja)

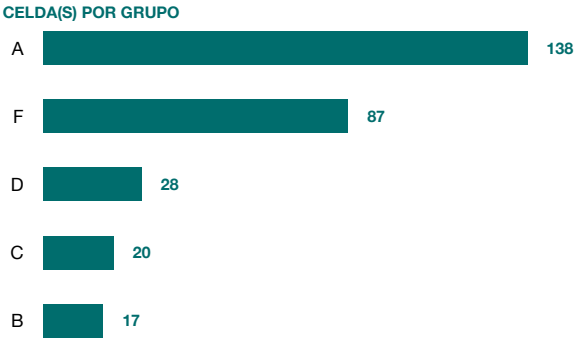
Síntesis. Cada origen activo se conecta con todos los destinos activos; la misma celda puede formar un self-loop.

La red observada contiene 333 vértices y 1.105 aristas

333
celdas ocupadas (vértices)

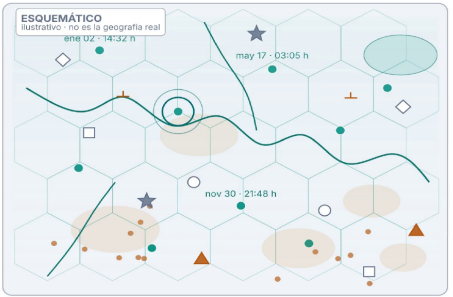
1.105
aristas dirigidas

Las frecuencias describen presencia en celdas; no constituyen por sí mismas una medida de centralidad.



Síntesis. La red observada es la referencia para evaluar reconstrucción y descubrir elementos faltantes.

Cada vértice hereda su contexto ambiental



«cada vértice hereda su contexto ambiental completo»

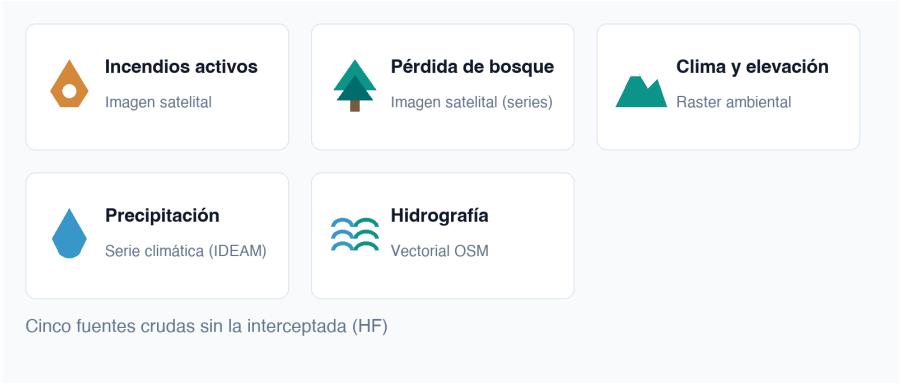
Capas de contexto (se acumulan)

- Comunicaciones HF interceptadas
  puntos con sello de hora
- Incendios NASA FIRMS/VIIRS
  año 2018
- Pérdida de bosque Hansen
  años 2000–2021
- Clima WorldClim + elevación
  tiempo de viaje a ciudad/puerto
- Hidrografía OSM
  ríos y cuerpos de agua
- Infraestructura OSM (validación)
  aeródromos · combustible · policía
  cultivos · canteras · muelles

Fuentes (artículo):
HF: Policía Nacional · NASA FIRMS/VIIRS
Hansen/GEE · WorldClim · OpenStreetMap

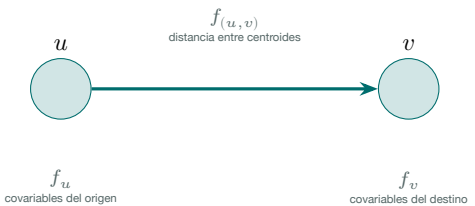
Síntesis. El hueso son las comunicaciones HF; la carne, las capas remotas y climáticas; el contraste, la infraestructura OSM para validar cercanías.

Datos satelitales y ambientales



Síntesis. Cinco fuentes crudas (sin la interceptada HF): lo que se vio desde el espacio y el ambiente físico del vértice.

Cada par dirigido se convierte en una observación de clasificación



$$P_{u,v} = \Psi(f_u, f_v, f_{(u,v)})$$

VARIABLE OBJETIVO

$$G_{uv} \in \{0, 1\}$$

1 = arista observada

0 = par sin arista observada

se excluyen self-loops en este dataset

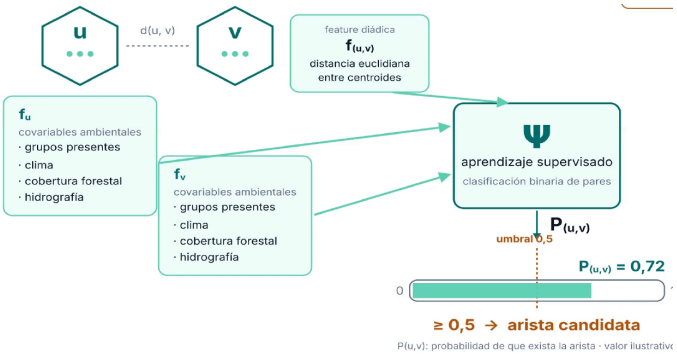
DATOS DESBALANCEADOS

333 vértices $\Rightarrow 333 \times 332 \approx 111$ mil pares dirigidos

1.105 aristas observadas $\Rightarrow \approx 1\,\%$ positivos

Síntesis. La reconstrucción se formula como clasificación binaria sobre pares de celdas.

Predicción de aristas = clasificación binaria (esquema)



Síntesis. Una fila por par (u, v): covariables de ambos vértices más la distancia entre centroides alimentan a Ψ; el umbral 0,5 decide la candidatura.

PCA y SMOTE: preprocesar bajo desbalance

Estandarizar covariables → **PCA** 80 % varianza → **Dividir** 80 / 20 → **SMOTE** minoría ≈ 1 %

PCA · COMPONENTES PRINCIPALES

SMOTE · MINORÍA SINTÉTICA



$$x_{\text{new}} = x_i + \delta\,(x_{nn} - x_i), \quad \delta \in [0, 1]$$

Interpola ejemplos sintéticos de la clase minoritaria (aristas) entre vecinos cercanos hasta equilibrar el entrenamiento.

Síntesis. Del vértice crudo al ejemplo balanceado: PCA compacta la información ambiental; SMOTE equilibra la clase minoritaria antes de entrenar.

Especificidad y regresión logística

ESPECIFICIDAD · CARGA DE VALIDACIÓN

$$\text{Specificity} = \frac{TN}{TN + FP}$$

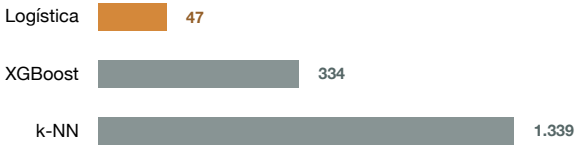
FP = candidato potencial y trabajo de verificación; no arista confirmada

REGRESIÓN LOGÍSTICA

$$\mathbf{p} = \frac{1}{1 + e^{-\mathbf{X}\beta}}$$

$\mathbf{X}$: componentes principales (80 % varianza) · β: pesos ajustados por máxima verosimilitud

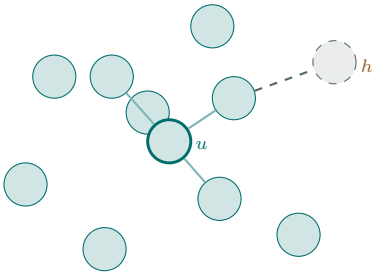
FALSOS POSITIVOS · PRUEBA



barras en escala de raíz cuadrada

Síntesis. Menos FP → lista más manejable para analistas; no mayor certeza sobre cada candidato.

La homofilia reduce el espacio de búsqueda de vértices ocultos



$k = 3$ vecinos por vértice

los ocultos comparten ambiente (homofilia)

333 conocidos

PCA ambiental

↓

$k = 3$ **vecinos**

por vértice conocido

↓

Grupo candidato

búsqueda reducida

↓

500 semillas

estabilidad de la selección

Supuesto: los vértices ocultos comparten características ambientales con los observados. Es una *hipótesis de homofilia*, no una propiedad demostrada.

Síntesis. KNN restringe dónde buscar; las 500 iteraciones miden persistencia de selección.

La frecuencia de predicción mide estabilidad entre semillas

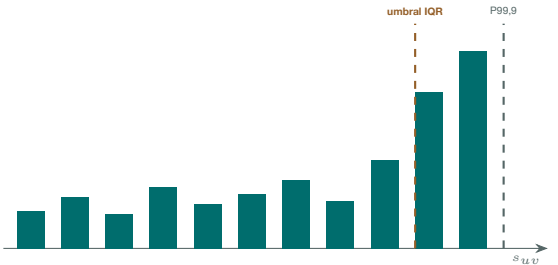
$$s_{uv} = \sum_{r=1}^{500} \mathbb{I}\left(p_{uv}^{(r)} > 0.5\right)$$

s_{uv}: clasificaciones positivas del par (u, v) entre 500 semillas (notación auxiliar de presentación)

DOS CRITERIOS

IQR
aristas desconocidas

Percentil 99,9
aristas ocultas



frecuencia de selección en 500 corridas

Síntesis. Una frecuencia alta indica selección persistente; no equivale a probabilidad de existencia.

La reconstrucción conserva la estructura principal del Chronnet

Métrica	Observado	Reconstruido
Vértices	333	331
Aristas	1.105	1.073
Densidad	0,00999	0,00982
Grado medio	6,64	6,48
Clustering	0,107	0,107
Diámetro	16	16

97 %
aristas recuperadas

99,4 %
vértices retenidos

La comparación evalúa preservación estructural; los exponentes son 2,65 y 2,70.

Síntesis. La reconstrucción conserva la estructura principal del Chronnet observado.

El umbral IQR selecciona 107 aristas entre vértices conocidos

107

aristas desconocidas candidatas

$$s_{uv} > Q_3 + 1.5\,IQR = 46$$

la regla selecciona valores atípicamente altos de frecuencia entre las 500 ejecuciones

Ambos extremos son **vértices conocidos**.

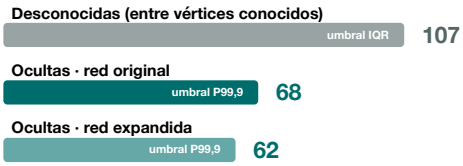


frecuencia s_{uv} · fuera de la cerca IQR → candidata

Síntesis. El resultado es una lista priorizada para verificación, no 107 conexiones confirmadas.

Dos procedimientos convergen en 84 vértices ocultos únicos

ARISTAS PREDICHAS · UMBRAL APLICADO



Las 68 y 62 son **disjuntas** (cero en común); los vértices que revelan **sí se traslapan**: 19 comunes.

COMPOSICIÓN DE LOS 84 VÉRTICES OCULTOS



35 únicos (red original) + 30 únicos (red expandida) + 19 comunes = 84 vértices ocultos a verificar.

Síntesis. Las 68 y 62 aristas no se traslapan; los vértices que revelan sí: 35 + 30 + 19 comunes = 84.

Infraestructura crítica y cadena de procesamiento

Conjunto de datos	Tipo	Fuente / resolución	Uso en el modelo
Infraestructura crítica	Vectorial OSM	OpenStreetMap / Overpass Turbo · aeródromos, combustible, policía, cultivos, canteras, muelles	Validación espacial por proximidad (buffers 5–30 km)

De la descarga a las covariables del modelo



Síntesis. Siete fuentes en total se funden en un solo vector por vértice; la infraestructura OSM además valida por proximidad.

De una prueba de rangos a una regla de decisión territorial

HIPÓTESIS SOBRE MEDIANAS

$$H_0 : \text{Mediana}(C_{\text{hidden}}) \geq \text{Mediana}(C_{\text{random}})$$
$$H_1 : \text{Mediana}(C_{\text{hidden}}) < \text{Mediana}(C_{\text{random}})$$

Wilcoxon unilateral de rangos · 500 repeticiones · buffers de 5, 10, 20 y 30 km.

$$R = \frac{1}{500} \sum_{r=1}^{500} \mathbb{I}(p_r \leq 0.05)$$

R: proporción de pruebas significativas (notación auxiliar)

REGLA OPERACIONAL DEL ARTÍCULO



Aplicación: canteras → atracción · muelles → trade-off · siempre por escala y grupo de vértices.

Síntesis. Patrón mixto o inconcluso no implica irrelevancia; exige lectura por infraestructura y escala.

Validación institucional · Policía Nacional

DIJIN · DIRECCIÓN DE INVESTIGACIÓN CRIMINAL E INTERPOL

El **Centro Nacional de Análisis Criminal (CENAC)** confirmó la **relevancia operativa** de los vértices ocultos y aristas candidatas predichos: coinciden con zonas y patrones conocidos por la inteligencia policial.

https://www.policia.gov.co/jefatura-nacional-del-servicio-de-policia/dijin

DIRAN · DIRECCIÓN DE ANTINARCÓTICOS

La Dirección de Antinarcóticos validó el resultado en contexto de Amazonía: la capa logística de la red es coherente con los sistemas de tráfico que la Policía confronta en el territorio.

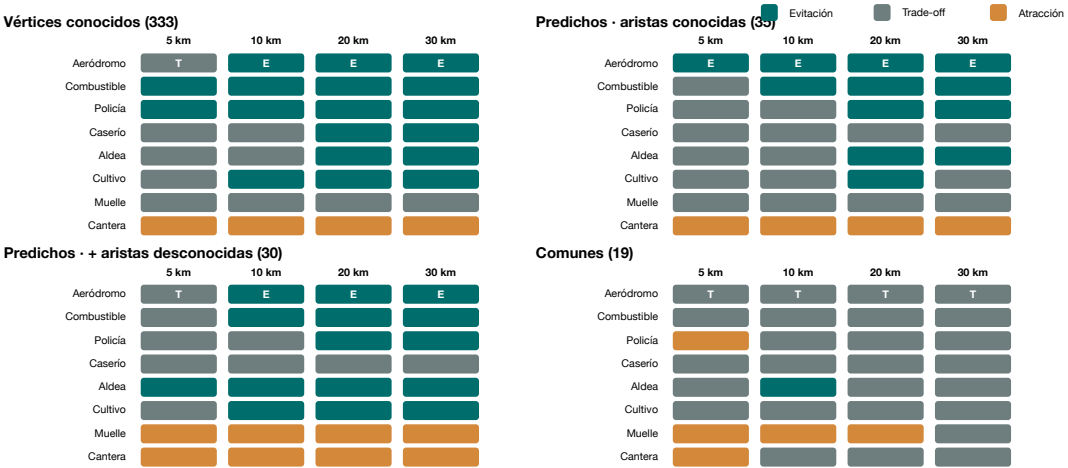
https://www.policia.gov.co/jefatura-nacional-del-servicio-de-policia/diran

Qué valida y qué no: la coincidencia respalda relevancia y *ordena la verificación de campo*; no confirma punto por punto la presencia física del delito.

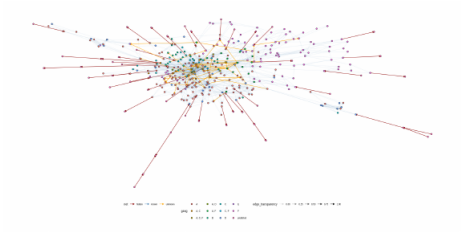
Síntesis. La validación operativa convierte las predicciones en hipótesis priorizadas para la inteligencia policial.

Tres lógicas espaciales — síntesis

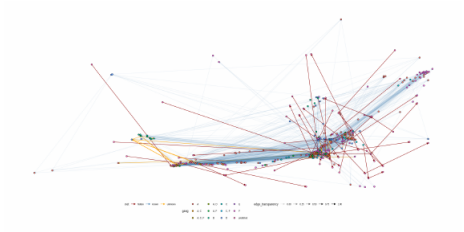
Evitación: significativo ($p \leq 0,05$) en ≥80 % de 500 corridas · Atracción: ≤20 % y mediana $p \geq 0,8$ · Trade-off: el resto · Controles: 500 semillas · umbrales distintos



El Chronnet reconstruido, en abstracto y en el mapa



The complete reconstructed network, visualized by connection strength.



The same complete network, visualized by its real-world latitude and longitude.

Izquierda: la red completa por fuerza de conexión · Derecha: la misma red en sus coordenadas reales (lat-lon).

Gracias



FastAnalytics S.A.S

El futuro habla en datos
¡Lo anticipamos con decisiones!

Andrés Pérez-Coronado · **Ponente**

andres.perezc@fastanalytics.co

2do Workshop Internacional de Estadística 2026 · Universidad Nacional de Colombia, Sede Manizales

fcen.unal.edu.co/2wie2026

Taller y materiales (ES · EN): fastanalytics.co/encuentro-estadistica