

## Paso 2 – Selección de variante de Forman–Ricci y especificación

### 0) Resumen ejecutivo

- **Selección de variante:** Para el MVP se adopta la **versión no ponderada de la curvatura de Forman–Ricci**, definida en grafos sin pesos como  $F(e) = 4 - d_v - d_w$ , donde  $d_v$  y  $d_w$  son los grados de los nodos conectados por la arista <sup>1</sup>. Esta variante es local, combina sólo grados de los nodos y se calcula en tiempo  $O(\deg(v) + \deg(w))$  por arista, permitiendo su aplicación en tiempo casi lineal para grafos grandes <sup>2</sup>.
- **Opción ampliada:** Se considera opcionalmente la **curvatura Forman–Ricci aumentada (AFRC)**, que añade contribuciones de triángulos:  $F^\#(e) = 4 - d_v - d_w + 3\Delta(e)$ , siendo  $\Delta(e)$  el número de triángulos que contienen la arista <sup>3</sup>. Este ajuste contrarresta la penalización por alto grado cuando existe cohesión local y captura mejor los clusters densos. El cálculo de  $\Delta(e)$  requiere intersectar vecinos de  $v$  y  $w$ , con coste  $O(\min(\deg(v), \deg(w)))$  por arista y complejidad total aproximadamente lineal <sup>4</sup>.
- **Descartadas:** Las variantes **ponderada general** y **2-complejo** se documentan, pero se descartan para el MVP. La versión ponderada utiliza pesos de vértice y de arista (ecuación (4.9)) <sup>5</sup>; aunque local, añadir pesos exige normalizar la semántica de pesos y no aporta mejoras significativas en un grafo inicialmente homogéneo. La versión 2-complejo (ecuación (4.11)) incluye triángulos y k-caras de orden >3 <sup>6</sup>; su cálculo exacto no escala bien y requiere contar caras de orden k en las vecindades, lo que incrementa mucho el coste <sup>2</sup>.
- **Riesgos y mitigaciones:** Utilizar la versión no ponderada puede ignorar la influencia de pesos heterogéneos. Se mitigará registrando métricas de enrutado y, si fuera necesario, se probará la versión ponderada en una fase posterior. El uso de AFRC podría sobreajustarse a grafos altamente triangulados; se incluirá un parámetro para activar o desactivar la contribución de triángulos.

### 1) Estado epistemológico

Claim_ID	Claim	Etiqueta	Justificación breve
C7	La curvatura de Forman–Ricci para grafos no ponderados es $F(e) = 4 - d_v - d_w$ .	A	El artículo de Estrada et al. (2025) especifica que, en grafos sin pesos, la curvatura de una arista $e$ se reduce a $4 - k_v - k_w$ <sup>1</sup> .
C8	La curvatura de Forman–Ricci general (ecuación (4.9)) utiliza pesos de vértice y arista, pero reduce a la fórmula no ponderada si $w_v = w_e = 1$ .	A	El artículo de Tian, Lubberts y Weber (2025) presenta la definición general de FRC con pesos $w_e$ , $w_v$ y muestra que para grafos no ponderados se obtiene $RicF(e) = 4 - d_{v_1} - d_{v_2}$ <sup>7</sup> .

Claim_ID	Claim	Etiqueta	Justificación breve
C9	El cálculo exacto de la curvatura 2-complejo requiere identificar todas las caras de orden k vecinas a una arista, lo cual no escala bien; es preferible limitarse a triángulos.	A	El mismo artículo indica que para calcular la FRC 2-complejo debe identificarse cada cara de orden k, lo que limita la escalabilidad y justifica aproximar usando solo triángulos <a href="#">2</a> .
C10	La curvatura Forman–Ricci aumentada incorpora triángulos: $F^{\#}(e) = 4 - d_v - d_w + 3 \Delta(e)$ y se calcula en tiempo casi lineal.	A	La página especializada de EmergentMind explica que AFRC se define añadiendo 3 veces el conteo de triángulos $\Delta(e)$ a la FRC original y que su cálculo para todos los bordes es near-linear usando intersecccionar vecinos <a href="#">3</a> <a href="#">4</a> .
C11	Incorporar contribuciones de cuadrángulos y caras de orden superior no mejora significativamente el rendimiento y aumenta el coste.	B	El análisis en el artículo de Tian et al. observa que incluir caras de orden $k > 3$ no mejora las prestaciones de la FRC en clustering y aumenta el coste computacional <a href="#">2</a> .

## 2) Registro de Evidencia

Claim_ID	Claim	Etiqueta	Fuente(s)	Extracto / Idea clave	Riesgo de error	Cómo verificar
C7	$F(e) = 4 - d_v - d_w$ para grafos no ponderados	A	<a href="#">1</a>	El artículo define la curvatura de Forman–Ricci de una arista en grafos no ponderados como $4 - k_v - k_w$ .	Bajo (fórmula sencilla)	Reproducir en un grafo simple y comprobar signos de curvatura para distintos grados.
C8	Fórmula general FRC con pesos (ecuación 4.9) y reducción	A	<a href="#">7</a>	La definición con pesos y la reducción a $4 - d_v - d_w$ si $\omega_v = \omega_e = 1$ .	Bajo	Implementar la fórmula general y verificar que con todos los pesos iguales a uno se obtiene la fórmula no ponderada.

Claim_ID	Claim	Etiqueta	Fuente(s)	Extracto / Idea clave	Riesgo de error	Cómo verificar
C9	Coste elevado de la FRC 2-complejo e importancia de limitarse a triángulos	A	2	Se indica que calcular la FRC 2-complejo requiere encontrar todas las caras en el vecindario de un borde y que es mejor usar solo triángulos para reducir la carga computacional.	Medio	Medir el tiempo de cómputo de la FRC 2-complejo frente a la FRC con triángulos en grafos de prueba.
C10	AFRC incorpora triángulos y se calcula en tiempo near-linear	A	3 4	La página señala que AFRC suma 3 veces el número de triángulos a la FRC y que el coste es $O(\min(\deg(x), \deg(y)))$ por arista, logrando complejidad casi lineal.	Medio	Implementar la AFRC y comparar el tiempo de cálculo frente a la FRC simple en grafos reales.
C11	Contribuciones de cuadrángulos y caras de orden superior aportan poco y encarecen el cálculo	B	2	El artículo observa que no se gana rendimiento en clustering al incluir k-caras con $k > 3$ y que el coste aumenta, por lo que recomienda ignorarlas.	Alto (basado en un estudio concreto)	Repetir el análisis con otras tareas/escenarios para verificar si las caras de orden superior ofrecen beneficios en otros contextos.

### 3) Diseño/Propuesta: variante seleccionada y especificación

#### 3.1 Variantes consideradas

1. **FRC no ponderada - seleccionada.** Para una arista  $e = (u, v)$  en un grafo no ponderado, la curvatura se define como:

$$F(e) = 4 - d(u) - d(v) \quad [415615785479541 \dagger L362 - L365] .$$

- **Complejidad:**  $O(\deg(u) + \deg(v))$  por arista; global  $O(|E|)$ .
- **Interpretación:** Aristas que unen nodos con altos grados tienen curvatura negativa y se consideran “puentes” susceptibles de estirarse y podarse.
- **Ventajas:** Cálculo local y muy eficiente; sin necesidad de pesos iniciales.

- **Inconvenientes:** Ignora información de pesos heterogéneos o patrones de clustering; penaliza puentes útiles en redes heterofílicas.
- **FRC ponderada (equación (4.9)) – descartada para el MVP.** La definición general usa pesos de vértice  $\omega_v$  y de arista  $\omega_e$ <sup>7</sup>.
- Permite incorporar la semántica de pesos (p. ej. relevancia de nodos).
- Requiere normalizar pesos y fijar la semántica con precisión; añade complejidad matemática sin clara ganancia en grafos con pesos uniformes.
- Será considerada en iteraciones posteriores si el grafo almacena pesos confiables.
- **FRC 2-complejo (ecuación (4.11)) – descartada.** Incluye contribuciones de caras de orden k (triángulos, cuadrángulos, etc.)<sup>6</sup>.
  - Teóricamente más completa (incluye cohesión de 2-caras).
  - Computacionalmente costosa: requiere enumerar todas las k-caras en la vecindad y no escala bien<sup>2</sup>.
  - Los estudios disponibles indican que aportar pentágonos o caras mayores no mejora las métricas de clustering y aumenta la carga<sup>2</sup>.
- **FRC aumentada por triángulos (AFRC) – opción complementaria.** Se define como  $F^{\#}(e) = 4 - d(u) - d(v) + 3 \Delta(e)$ , donde  $\Delta(e)$  es el número de triángulos incidentes en  $e$ <sup>3</sup>.
  - Captura la cohesión local (presencia de triángulos) contrarrestando la penalización por alto grado.
  - **Complejidad:** contar triángulos por arista cuesta  $O(\min(\deg(u), \deg(v)))$  y la complejidad global es near-linear<sup>4</sup>.
  - **Ventaja:** preserva puentes que están insertos en cliques o comunidades, evitando su poda prematura.
  - **Riesgo:** en redes con pocos triángulos o estructura heterofílica, la aportación de  $3\Delta(e)$  puede ser irrelevante o incluso generar sesgos.
  - **Uso propuesto:** introducir un parámetro  $\alpha \in \{0, 1\}$  para activar/desactivar la contribución de triángulos y calibrarlo empíricamente.

### 3.2 Especificación para el MVP

- **Curvatura adoptada:** se implementará inicialmente la **FRC no ponderada** como función `compute_kappa_forman_simple(G, e)` que devuelve  $4 - d(u) - d(v)$ .
- **Parámetro opcional** `alpha_triangles`: valor booleano para incluir la term  $3 \Delta(e)$ . Si `alpha_triangles = True`, la función computará  $F^{\#}(e) = 4 - d(u) - d(v) + 3 \Delta(e)$ , siendo  $\Delta(e)$  el número de vecinos comunes de  $u$  y  $v$ .
- **Interfaz:** la función recibirá el grafo (lista de adyacencia) y la arista; utilizará `degree` y la intersección de las listas de vecinos.
- **Complejidad esperada:**  $O(\deg(u)+\deg(v))$  para la versión simple y  $O(\min(\deg(u), \deg(v)))$  para la versión con triángulos; la ejecución para todas las aristas es  $O(|E| \cdot d_{avg}) \approx \text{linear}$  para grafos espaciados.
- **Verificación:** se implementarán tests unitarios para comprobar que:
  - En grafos sin triángulos (p. ej., árbol), la versión simple y la AFRC coinciden.
  - La fórmula general con `w_v=w_e=1` produce el mismo resultado que la versión simple.

- Los signos de la curvatura se comportan como esperado en grafos toy (línea, estrella, ciclo).

### 3.3 Supuestos y riesgos

- **S3:** Se asume que los grados de los nodos capturan adecuadamente la dispersión geodésica local; si las aristas de alto grado son funcionales (por ejemplo, puentes legítimos), la FRC simple podría penalizarlas indebidamente. **Mitigación:** usar la versión AFRC con `alpha_triangles=True` en escenarios donde se observe que la poda elimina aristas importantes.
- **S4:** Se asume que las contribuciones de cuadrángulos y caras superiores son despreciables para los objetivos de denoising. Si en un futuro se trabajan grafos con muchas cuadrículas (por ejemplo, redes de malla), podría ser necesario reconsiderar la FRC 2-complejo.

## 4) Plan de simulación/experimento

### Hipótesis (medible)

- **H3:** La FRC simple (sin triángulos) mejora la precisión del enrutamiento reduciendo el ruido estructural, en comparación con un grafo sin limpieza.
- **H4:** La AFRC (con triángulos) preserva mejor la conectividad útil cuando existen comunidades densas, evitando podar excesivamente puentes que participan en triángulos.

### Setup

1. Construir grafos sintéticos:
2. Redes de bloques estocásticos (SBM) con parámetros controlados para grado medio y densidad de triángulos.
3. Árboles, ciclos y grafos reticulados (para observar FRC simple vs AFRC).
4. Aplicar el `RicciCleaner` con `compute_kappa_forman_simple` y con `alpha_triangles=True`.
5. Evaluar tareas de enrutado: dada una consulta, medir si el router encuentra el nodo/experto correcto (top-k).
6. Registrar métricas: porcentaje de aristas podadas, precisión del enrutado, conectividad residual (componentes conexas), coste computacional.

### Variables

- Independientes: densidad de triángulos en el grafo, grados medios, `alpha_triangles` (0/1), umbral  $d_{\max}$ .
- Dependientes: `accuracy@k`, número de aristas eliminadas, tamaño de la mayor componente conexa, tiempo de cómputo.

### Criterio de éxito/fallo

- Éxito: la FRC simple logra una mejora significativa (p. ej. +10 %) en precisión de enrutamiento respecto a la base, sin incrementar la latencia más de 20 %.
- Evaluar si la AFRC supera a la FRC simple en grafos con alto número de triángulos y si, en grafos sin triángulos, no degrada el rendimiento.

## 5) Próximos pasos y riesgos principales

1. **Implementar** `compute_kappa_forman_simple` y `compute_kappa_forman_triangle` según la especificación.
2. **Integrar la variable** `alpha_triangles` **en el** `RicciCleaner` para activar la contribución de triángulos.
3. **Diseñar y ejecutar el benchmark sintético** (árboles, SBM, grafos reticulados) para validar H3 y H4.
4. **Analizar resultados:** comparar precisión y coste entre FRC simple y AFRC; ajustar parámetros  $\eta$  y  $d_{\max}$ .
5. **Decidir** si se adopta la versión ponderada en una fase posterior.

**Riesgos principales:** - La FRC simple puede eliminar aristas fundamentales en grafos heterofílicos (e.g. bipartitos). Se mitigará calibrando el umbral y monitorizando la conectividad.  
- La AFRC podría no aportar mejoras en grafos sin triángulos y complicar la implementación; se añadirá como opción y se evaluará empíricamente.  
- El cálculo de triángulos podría ser costoso en grafos densos; se optimizará con técnicas de intersección eficiente.

---

1 formanricci\_communicability\_curvature\_of\_graphs\_and\_networks.pdf

<https://www.cambridge.org/core/services/aop-cambridge-core/content/view/1D39CA14E9513387FB1E353105F57E20/S0956792525000014a.pdf>formanricci\_communicability\_curvature\_of\_graphs\_and\_networks.pdf

2 5 6 7 24-0781.pdf

<https://www.jmlr.org/papers/volume26/24-0781/24-0781.pdf>

3 4 Augmented Forman-Ricci Curvature

<https://www.emergentmind.com/topics/augmented-forman-ricci-curvature-afr>