

Medición de la Interdisciplinariedad: Un Panel de Indicadores Multicomponente para la Evaluación de la Investigación

A. Rivero y A.I. Scaffold

2026

Abstract

Revisamos el panorama de indicadores bibliométricos para medir la investigación interdisciplinaria y proponemos un panel estructurado de tres componentes que combina la diversidad de Rao-Stirling, la coherencia de red (fuerza media de enlace) y un proxy del efecto transdisciplinar. Basándonos en trabajos recientes que exponen la baja consistencia y limitada validez de constructo de los enfoques existentes basados en un único indicador, organizamos la literatura en una taxonomía de cuatro dimensiones conceptuales — diversidad, coherencia, difusión y novedad — y cuatro familias metodológicas. Utilizando un conjunto de datos universitarios de juguete, demostramos que ningún indicador escalar puede discriminar la genuina integración interdisciplinaria de la amplitud polimática o la especialización estrecha. El panel completo, por el contrario, caracteriza de forma única cada tipo de investigador. Demostramos analíticamente que esta discriminación es robusta ante perturbaciones de la matriz de similaridad intercategorías. Un estudio de caso a nivel departamental con siete investigadores confirma el poder discriminatorio del panel a escala realista, distinguiendo correctamente integradores, polímatas y especialistas puente donde los indicadores únicos fallan. Estos resultados sugieren que la evaluación de la investigación interdisciplinaria — ya sea a nivel institucional o de agencias nacionales — requiere un enfoque multidimensional en lugar de puntuaciones compuestas.

Introducción

La investigación interdisciplinaria (IDR) es ampliamente considerada esencial para abordar desafíos científicos y sociales complejos. Las iniciativas de política científica en Estados Unidos, Europa y otros lugares han buscado fomentar la interdisciplinariedad, frecuentemente bajo el supuesto de que cruzar fronteras disciplinarias conduce a resultados de investigación más impactantes (National Academies, 2005). Sin embargo, la medición de la interdisciplinariedad sigue

siendo problemática. A pesar de décadas de trabajo, la literatura bibliométrica no ha convergido en indicadores apropiados.

Wang y Schneider (2020), al probar 23 medidas de interdisciplinariedad en cuatro familias metodológicas, encontraron correlaciones sorprendentemente bajas incluso entre medidas diseñadas para capturar la misma dimensión, concluyendo que “ningún indicador único puede identificar inequívocamente” la investigación interdisciplinar. Leydesdorff, Wagner y Bornmann (2019) mostraron que el índice de diversidad de Rao-Stirling ampliamente utilizado está dominado por su componente de disparidad, produciendo clasificaciones anómalas y bajo poder discriminatorio. Cantone (2024) argumentó que la interdisciplinariedad es un “constructo polisémico” cuyas múltiples dimensiones semánticas no pueden capturarse mediante ningún valor numérico único. Estos hallazgos apuntan a un desajuste fundamental entre la naturaleza multidimensional de la IDR y los indicadores escalares utilizados para medirla.

Este artículo realiza dos contribuciones. Primero, relevamos el panorama de indicadores y lo organizamos en una taxonomía de cuatro dimensiones conceptuales (diversidad, coherencia, difusión, novedad) cruzadas con cuatro familias metodológicas (basadas en referencias, en citas, en texto, en redes). Segundo, proponemos un panel específico de tres componentes — diversidad, coherencia y efecto transdisciplinar — y demostramos su poder de discriminación en un conjunto de datos de juguete con garantías analíticas de robustez.

El resto del artículo está estructurado como sigue. La Sección 2 revisa los fundamentos conceptuales de la medición de la interdisciplinariedad. La Sección 3 presenta la taxonomía de indicadores. La Sección 4 define nuestro panel, demuestra su poder de discriminación y prueba resultados de robustez. La Sección 5 discute implicaciones para la evaluación a nivel institucional y nacional. La Sección 6 identifica problemas abiertos. La Sección 7 aplica el panel a un estudio de caso a nivel departamental. La Sección 8 concluye.

Fundamentos Conceptuales

El Marco de Diversidad de Stirling

La dificultad de medir la interdisciplinariedad es fundamentalmente un problema de medición de diversidad. Stirling (2007) demostró que cualquier caracterización significativa de la diversidad requiere atención a tres propiedades — *variedad* (cuántas categorías están representadas), *balance* (cuán uniformemente distribuidas están) y *disparidad* (cuán diferentes son las categorías entre sí) — y que índices estándar como la entropía de Shannon o el índice de concentración de Herfindahl capturan solo las dos primeras. Este marco tripartito se ha convertido en el lente conceptual estándar a través del cual se analizan los indicadores de interdisciplinariedad en la literatura bibliométrica (Porter y Rafols, 2009; Leydesdorff y Rafols, 2011; Wang y Schneider, 2020). La influencia del marco se extiende más allá de la bibliometría: Stirling lo desarrolló originalmente en

contextos de diversidad ecológica y tecnológica, y su adopción por la comunidad cientométrica refleja un reconocimiento de que la interdisciplinariedad, como la biodiversidad, no puede reducirse a un recuento de categorías sin atender a las distancias entre ellas. El índice de Rao-Stirling $\Delta = \sum_{i \neq j} d_{ij} p_i p_j$, que operacionaliza las tres propiedades en una única expresión, se ha convertido consecuentemente en el punto de partida más ampliamente utilizado para el diseño de indicadores (Rafols y Meyer, 2009; Zhang, Rousseau y Glanzel, 2016).

Evolución Histórica

El estudio moderno de la interdisciplinariedad tiene raíces en debates de política científica de los años 1970, pero los esfuerzos de medición cuantitativa se aceleraron solo después de que la OCDE codificara una tipología tripartita que sigue siendo influyente hoy (OECD, 1998; Morillo, Bordons y Gomez, 2003). Bajo esa tipología, la investigación *multidisciplinaria* se nutre de diferentes perspectivas disciplinares sin integrarlas; la investigación *interdisciplinaria* alcanza una síntesis teórica, conceptual o metodológica coherente; y la investigación *transdisciplinaria* implica una integración mutua de epistemologías disciplinares que puede trascender las fronteras existentes por completo. Aunque las fronteras entre estas categorías siguen siendo objeto de debate, la distinción pone en primer plano una pregunta crítica para el diseño de indicadores: ¿debería la medición apuntar a la amplitud de los insumos disciplinares (una propiedad multidisciplinaria) o a la profundidad de su integración (una propiedad interdisciplinaria o transdisciplinaria)?

Wagner et al. (2011) elaboran estas definiciones en una revisión integral de la medición de la IDR. La investigación multidisciplinaria “yuxtapone perspectivas disciplinares, añadiendo amplitud y conocimiento disponible — el producto no es más ni menos que la simple suma de sus partes”. La investigación interdisciplinaria “integra datos, métodos, herramientas, conceptos y teorías disciplinares separados para crear una visión holística” — el producto es “diferente de, y mayor que, la suma de sus partes”. Los enfoques transdisciplinarios “son marcos comprehensivos que trascienden el alcance estrecho de las cosmovisiones disciplinares”. Señalan que el uso común “rara vez distingue entre las direcciones de entrada y salida de la IDR” (Wagner et al., 2011, fn. 10), sin embargo la distinción importa para la medición: un investigador puede basarse en los métodos de múltiples disciplinas al diseñar un estudio (interdisciplinariedad del lado de entrada) mientras publica exclusivamente en un campo (concentración del lado de salida), o inversamente puede publicar en muchos campos sin integración metodológica. Un cuarto modo, la investigación *transdisciplinar*, implica referenciar literatura de otro campo sin ningún intento de integración (Aksnes, Karlstrøm y Piro, 2026; Hammarfelt, 2020).

Aksnes, Karlstrøm y Piro (2026), relevando 1.498 publicaciones con calificaciones de IDR autoinformadas, encontraron que el 42 por ciento de los artículos fueron calificados como *tanto* multidisciplinares como interdisciplinares, y un 23 por ciento adicional como parcialmente ambos. Esta superposición empírica confirma

que la distinción multi/inter no es una partición limpia sino un espectro, y que los indicadores escalares únicos de diversidad (que agregan amplitud sin considerar la integración) no pueden distinguir entre estos modos. Como lo expresan Choi y Pak (2006) vía Abramo, D'Angelo y Zhang (2018), multi-, inter- y transdisciplinariedad forman “un continuo de niveles crecientes de involucramiento de múltiples disciplinas”. La direccionalidad del flujo de conocimiento también importa. Cuando la salida de una disciplina sirve como entrada para otra sin síntesis, la literatura lo denomina *multidisciplinariedad secuencial* (Stokols et al., 2003); cuando el insumo prestado transforma la disciplina receptora, es *interdisciplinariedad instrumental* (Klein, 2008). El intercambio bidireccional constituye *interdisciplinariedad recíproca*. Zhou, Guns y Engels (2023) formalizan la perspectiva de flujo mediante su marco de flujo de conocimiento interdisciplinar (IKF), que caracteriza los intercambios a lo largo de tres dimensiones: amplitud, intensidad y homogeneidad. Estas distinciones tienen implicaciones directas para el panel introducido en la Sección 4: la diversidad (Δ) captura amplitud independientemente de la dirección, la coherencia (S) distingue integración de yuxtaposición, y el efecto transdisciplinar (E) mide difusión más allá de la disciplina de origen.

La evidencia empírica sobre tendencias de largo plazo agudiza esta pregunta. Porter y Rafols (2009) analizaron registros de publicación en diez áreas temáticas desde 1975 hasta 2005 y documentaron crecimiento generalizado en marcadores superficiales de interdisciplinariedad: el número promedio de autores por artículo aumentó aproximadamente un 75 por ciento (de 1,3 a 2,0 en matemáticas, de 3,0 a 6,1 en investigación y educación médica), el número promedio de referencias por artículo creció aproximadamente un 50 por ciento, y la diversidad de disciplinas citadas se expandió comparablemente. Las tasas de publicación de autor único declinaron marcadamente en todos los campos (de 71 a 37 por ciento en matemáticas; de 40 a 20 por ciento en física y biología; de 12 a 4 por ciento en química). Sin embargo, las puntuaciones de integración basadas en el índice de Rao-Stirling mostraron solo un aumento promedio modesto de aproximadamente 5 por ciento durante el mismo período (excluyendo matemáticas, donde el aumento alcanzó el 39 por ciento desde una base muy baja). La conclusión fue impactante: la ciencia estaba “volviéndose más interdisciplinaria, pero en pequeños pasos”, con citas principalmente alcanzando campos vecinos y solo un crecimiento modesto en conexiones cognitivas distantes.

Morillo, Bordons y Gomez (2001, 2003) proporcionaron evidencia complementaria a nivel de revista y categoría. Entre 1981 y 1996, el sistema de clasificación de revistas ISI añadió 38 nuevas categorías temáticas, de las cuales 21 aparecieron solo en ingeniería — un incremento del 154 por ciento en el recuento de revistas de ese campo. Las nuevas categorías exhibieron sistemáticamente mayor interdisciplinariedad: el 69 por ciento fueron multi-asignadas a más de una categoría (comparado con el 55 por ciento de categorías más antiguas), y mostraron un 28 por ciento más de fuerza de enlace intercategorías y un 30 por ciento mayor diversidad disciplinar. Un completo 80 por ciento de las nuevas categorías cayeron en clústeres caracterizados por alta interdisciplinariedad. Este patrón sugiere

que el crecimiento de la ciencia procede mediante fragmentación e hibridación simultáneas — nuevas especialidades emergen en las fronteras disciplinares y heredan un carácter interdisciplinar desde sus orígenes.

El Problema de la Polisemia

La interdisciplinariedad misma no es un concepto unitario. Cantone (2024) enfatiza que la IDR es un “constructo polisémico” — un término que porta múltiples significados parcialmente superpuestos a través de comunidades académicas y contextos de política. Un físico colaborando con biólogos, un científico de datos aplicando métodos a través de dominios, y un científico social sintetizando teorías desde múltiples disciplinas son todos llamados “interdisciplinares”, sin embargo la naturaleza y profundidad de su cruce de fronteras difieren cualitativamente. Esta polisemia significa que cualquier sistema de medición debe o bien seleccionar una definición operacional específica de interdisciplinariedad o bien acomodar explícitamente múltiples dimensiones. La taxonomía presentada en la Sección 3 sigue la segunda estrategia.

El problema de polisemia se ve agravado por una distinción cognitivo-social que atraviesa todas las definiciones (Cantone, 2024). Los enfoques cognitivos miden la interdisciplinariedad mediante la diversidad de insumos de conocimiento — referencias, métodos, marcos teóricos — y por tanto rastrean amplitud epistemológica. Los enfoques de agrupación la miden mediante estructuras sociales — coautorías, afiliaciones institucionales, composiciones de paneles. Los enfoques semánticos analizan contenido textual — palabras clave, resúmenes, texto completo — para detectar cruces de fronteras temáticas. Cada uno captura una faceta diferente del fenómeno, y puntuaciones altas en una dimensión no necesitan correlacionar con puntuaciones altas en otra. La consecuencia práctica es que investigadores que se autoidentifican como interdisciplinares pueden no registrarse como tales en indicadores bibliométricos, y viceversa (Zwanenburg, 2022). Esta discordancia entre interdisciplinariedad autoinformada y medida no es meramente un problema de calibración; refleja interpretaciones genuinamente diferentes de lo que el término significa.

Morillo, Bordons y Gomez (2003) introdujeron una distinción adicional entre “gran interdisciplinariedad” — conexiones entre áreas disciplinares distantes — y “pequeña interdisciplinariedad” — conexiones entre categorías vecinas dentro del mismo campo amplio. Química Aplicada, por ejemplo, exhibió un 83 por ciento de revistas multi-asignadas y un 55,6 por ciento de enlaces externos (entre áreas), mientras que Ciencia de Polímeros mostró solo un 39 por ciento de multi-asignación y un 33,3 por ciento de enlaces externos. La distinción grande/pequeño tiene implicaciones directas para el diseño de indicadores: un índice sensible solo a la variedad confundirá estos dos patrones cualitativamente diferentes, mientras que uno que incorpore disparidad los distinguirá.

Desafíos Epistemológicos

Más allá de la polisemia, la medición de la interdisciplinariedad confronta varias dificultades epistemológicas que restringen lo que los indicadores pueden legítimamente pretender capturar. La más fundamental es la inestabilidad de los sistemas de clasificación disciplinar de los que todos los indicadores bibliométricos dependen. Las categorías temáticas ISI Web of Science — la clasificación más ampliamente utilizada — alcanzan solo aproximadamente 50 por ciento de alineación con soluciones de clústeres basadas en citas (Boyack, 2005), y la correspondencia con clasificaciones derivadas de redes es similarmente imperfecta (Leydesdorff, 2006). Según recuentos recientes, el sistema comprende 254 categorías temáticas, con un 39 por ciento de revistas asignadas a más de una categoría. Aunque Rafols y Leydesdorff mostraron que estas desalineaciones tienen efectos limitados en mapas de ciencia agregados, pueden afectar sustancialmente valores de indicadores a nivel individual.

Un problema relacionado concierne a disciplinas “horizontales” — categorías amplias como Biología, Física, Química, o la categoría cajón de sastre de Ciencias Multidisciplinares — que exhiben multi-asignación artificialmente baja precisamente porque las políticas de clasificación de revistas limitan la dispersión excesiva entre categorías (Morillo, Bordons y Gomez, 2003). Revistas como Nature, Science y PNAS aparecen en Ciencias Multidisciplinares pero producen puntuaciones bajas de interdisciplinariedad bajo indicadores estándar a pesar de publicar trabajo que abarca el espectro disciplinar completo. Esto señala una preocupación epistemológica más profunda: las categorías que usamos para definir fronteras disciplinarias son en sí mismas artefactos de convención administrativa e histórica, no características estables del paisaje de conocimiento (Cantone, 2024). Cualquier indicador construido sobre tales clasificaciones hereda su contingencia.

Validez Conceptual vs. Empírica

Una complicación adicional es la brecha entre definiciones conceptuales y operacionalización empírica. Un indicador puede tener una motivación teórica clara (ej., “diversidad de insumos de conocimiento”) pero fallar en discriminar significativamente cuando se aplica a datos reales. Wang y Schneider (2020) documentaron este problema sistemáticamente: medidas que deberían ser teóricamente equivalentes producen clasificaciones inconsistentes y a veces contradictorias cuando se aplican al mismo conjunto de datos. Leydesdorff et al. (2019) mostraron que los valores de diversidad de Rao-Stirling frecuentemente difieren solo en el tercer lugar decimal entre investigadores, limitando el poder discriminatorio práctico. Estos hallazgos subrayan la necesidad de validación empírica de cualquier indicador propuesto, no meramente justificación teórica.

La crisis de validez tiene múltiples dimensiones (Zwanenburg, 2022). La *validez de contenido* está amenazada por la disparidad entre la riqueza conceptual de la interdisciplinariedad — que abarca variedad, balance y disparidad — y la tendencia de los indicadores individuales a capturar solo una o dos de estas

facetas. La *validez de dominio* está socavada por la multiplicidad de sistemas de clasificación y la ambigüedad de asignaciones de revistas a categorías. La *validez de coherencia* requiere consistencia entre operacionalizaciones alternativas del mismo constructo, sin embargo estudios empíricos encuentran repetidamente que medidas nociionalmente equivalentes producen resultados divergentes. Rafols (2019) argumentó que la mayoría de los indicadores existentes carecen de validez analítica (no miden lo que afirman medir) o robustez social (no son percibidos como significativos por las comunidades que evalúan), y que las métricas responsables requieren ambas. Los principios de robustez, humildad, transparencia y sensibilidad a la diversidad epistémica que Rafols articuló proporcionan un marco normativo que cualquier sistema de medición propuesto debería satisfacer.

La confusión de interdisciplinariedad con calidad de investigación es un problema distinto pero relacionado. Los indicadores no registran una propiedad preexistente de “calidad” o “excelencia”; más bien, como argumentó Rafols (2019), *promulgan* estas categorías — fuera de prácticas de evaluación, tales propiedades no tienen existencia independiente. Cuando los indicadores de interdisciplinariedad se usan junto o en combinación con proxies de calidad basados en citas, el riesgo de razonamiento circular es agudo: el trabajo interdisciplinar puede recibir más citas precisamente porque alcanza audiencias más amplias, no porque sea intrínsecamente superior. El panel propuesto en la Sección 4 trata la calidad como ortogonal a la caracterización de interdisciplinariedad, siguiendo el principio de que el panel debería describir el *tipo* de cruce de fronteras sin adjudicar su mérito.

Pluralismo Metodológico

La diversidad de perspectivas conceptuales revisadas arriba ha dado lugar a un conjunto de herramientas de medición correspondientemente diverso. En el extremo más establecido, el índice de diversidad de Rao-Stirling captura variedad, balance y disparidad en una única expresión, distinguiéndolo de la entropía de Shannon y el índice de Herfindahl, que no incorporan ninguna medida de distancia intercategorías (Porter y Rafols, 2009). A nivel de revista y categoría, el índice de coseno de Salton normaliza recuentos de revistas compartidas entre dos categorías por la media geométrica de sus tamaños, proporcionando una medida simétrica de fuerza de enlace intercategorías que varía de cero a uno (Morillo, Bordons y Gomez, 2003). Los patrones de multi-asignación — si las categorías secundarias de una revista caen dentro de la misma área amplia (interna, o “pequeña” interdisciplinariedad) o entre áreas (externa, o “gran” interdisciplinariedad) — ofrecen una perspectiva estructural complementaria.

Diferentes indicadores son apropiados en diferentes niveles de agregación. La multi-asignación de revistas en el sistema ISI proporciona una visión de nivel macro que es fácil de aplicar pero gruesa en resolución. Los patrones de citas y referencias JCR operan a nivel de categoría y son más sensibles a dinámicas disciplinares. Los análisis detallados a nivel de sección (ej., usando secciones de Chemical Abstracts) ofrecen precisión a nivel de revista pero requieren infraestruc-

tura específica de dominio (Morillo, Bordons y Gomez, 2001). Propuestas más recientes abogan por métodos semicualitativos, contextuales — incluyendo mapas de superposición sobre un mapa de ciencia base que visualizan variedad, balance y disparidad simultáneamente — como alternativas o complementos a índices escalares (Rafols, 2019). El principio rector es el de “indicadores en plural”: ninguna métrica única es suficiente, y la evaluación responsable requiere triangulación entre métodos y niveles de análisis. El panel multicomponente desarrollado en la Sección 4 está diseñado en este espíritu.

Una Taxonomía de Indicadores de Interdisciplinariedad

Organizamos la literatura de indicadores a lo largo de dos ejes: la *dimensión conceptual* de interdisciplinariedad que se mide, y la *familia metodológica* del indicador. Esto produce un mapa estructurado del campo que clarifica qué captura realmente cada indicador y dónde permanecen brechas.

Indicadores de Diversidad

Los indicadores de diversidad miden la heterogeneidad de los insumos de conocimiento de un investigador — típicamente la dispersión disciplinar de referencias citadas. Son la familia más extensamente estudiada de medidas de interdisciplinariedad, y cualquier evaluación creíble de IDR a nivel de artículo, autor o institución debe lidiar con las elecciones conceptuales y computacionales incrustadas en su diseño. Esta subsección proporciona un tratamiento sistemático de las principales familias de índices, las matrices de similaridad que requieren, la evidencia empírica sobre su validez, y las dificultades prácticas que surgen cuando se aplican a datos bibliométricos reales.

El marco de tres propiedades de Stirling

El fundamento teórico de la mayoría de los indicadores modernos de diversidad es la descomposición de Stirling (2007) de la diversidad en tres propiedades: *variedad* (el número de categorías distintas a las que se asignan elementos), *balance* (la uniformidad de la distribución de elementos entre esas categorías), y *disparidad* (el grado de diferencia entre las categorías mismas). Una medida de diversidad plenamente satisfactoria debería ser sensible a las tres propiedades simultáneamente. En la práctica, sin embargo, muchos índices ampliamente utilizados capturan solo una o dos de las tres, lo que es una fuente principal de desacuerdo entre estudios empíricos.

Medidas que capturan solo variedad y balance

Varios índices clásicos de ecología y economía han sido adaptados para medición de IDR. Estos índices registran variedad y balance pero son ciegos a la disparidad

— tratan un cambio de Química Orgánica a Química Analítica idénticamente a un cambio de Química Orgánica a Sociología.

Entropía de Shannon. Para una publicación cuyas referencias caen en categorías $i = 1, \dots, n$ con proporciones p_i , la entropía de Shannon es $H = -\sum_i p_i \ln p_i$. Alcanza su máximo $\ln n$ cuando las referencias están distribuidas uniformemente entre n categorías y es igual a cero cuando todas las referencias caen en una única categoría.

Diversidad de Simpson. El índice de Simpson $D_{\text{Sim}} = 1 - \sum_i p_i^2$ da la probabilidad de que dos referencias tomadas al azar pertenezcan a diferentes categorías. Está acotado entre 0 y $1 - 1/n$ y es más sensible a categorías dominantes que la entropía de Shannon.

Índice de Brillouin. Un análogo de muestra finita de la entropía de Shannon, definido como $HB = [\log(\sum_i c_i)! - \sum_i \log c_i!]/\sum_i c_i$ donde c_i es el recuento de referencias en la categoría i . Wang y Schneider (2020) encontraron que Shannon y Brillouin están casi perfectamente correlacionados ($r = 1.00$ en su Tabla 5), haciéndolos empíricamente redundantes.

Coeficiente de Gini invertido. El coeficiente de Gini mide concentración; $1 - G$ lo convierte en un indicador de balance. Como Shannon, es insensible a la identidad de las categorías entre las que se distribuyen las referencias.

Una regularidad empírica importante es que estos índices que no incorporan disparidad están moderada a fuertemente correlacionados entre sí (Wang y Schneider, 2020, Tabla 5, con correlaciones de Pearson por pares entre Simpson, Shannon, Brillouin y $1 - G$ variando de 0,60 a 1,00), pero solo débilmente correlacionados con índices que incorporan disparidad. Las dos familias capturan así aspectos genuinamente diferentes de la diversidad.

El índice de diversidad de Rao-Stirling

El índice de Rao-Stirling es el indicador más ampliamente utilizado que incorpora las tres propiedades de Stirling (Porter y Rafols, 2009; Rafols y Meyer, 2009). En su forma estándar se define como

$$\Delta = \sum_{\substack{i,j \\ i \neq j}} d_{ij} p_i p_j$$

donde p_i es la proporción de referencias en la categoría i y $d_{ij} = 1 - s_{ij}$ es la disimilitud entre las categorías i y j , derivada de una matriz de similaridad $\mathbf{S} = [s_{ij}]$. Cuando todas las categorías son máximamente disimilares ($d_{ij} = 1$ para todo $i \neq j$), el índice de Rao-Stirling se reduce al índice de Simpson. Cuando la matriz de similaridad porta estructura real, el índice penaliza la diversidad entre categorías cercanas y recompensa la diversidad entre distantes.

Generalización alfa-beta. Stirling (2007) propuso una familia más general $D_{\alpha,\beta} = \sum_{i \neq j} d_{ij}^\alpha (p_i p_j)^\beta$, donde los exponentes α y β gobiernan el peso relativo dado a disparidad versus balance. La elección convencional de Rao-Stirling establece $\alpha = \beta = 1$. Aumentar α amplifica la contribución de pares de categorías altamente dispares; aumentar β amplifica la contribución de distribuciones bien balanceadas. En la mayoría del trabajo cientométrico empírico se adopta $\alpha = \beta = 1$ sin discusión, pero la sensibilidad de resultados a estas elecciones de parámetros ha recibido investigación limitada. Los investigadores deberían estar conscientes de que cambiar estos exponentes puede desplazar clasificaciones de artículos o campos, incluso cuando los datos subyacentes son idénticos.

El estudio de ocho variantes. Wang y Schneider (2020) probaron ocho variantes del índice de Rao-Stirling cruzando dos elecciones de nivel de clasificación (promedio de publicación individual RS_P versus agregado RS_G) con cuatro especificaciones de matriz de disimilaridad (usando o bien el coseno vectorial de Salton SC o el coseno escalar de Ochiai SO , cada uno convertido a disimilaridad por o bien $1 - s$ o $1/s$). Los resultados son aleccionadores. Las correlaciones de Pearson entre variantes usando la misma fórmula de coseno pero diferentes transformaciones de disimilaridad pueden ser tan bajas como $r = 0,30$ (entre RS_G[1 - SC] y RS_G[1/ SC]); las correlaciones entre variantes usando diferentes fórmulas de coseno con la misma transformación pueden ser tan bajas como $r = 0,18$ (entre RS_P[1 - SC] y RS_P[1 - SO]). En un análisis en profundidad de cinco categorías temáticas seleccionadas de Web of Science (Nanociencia, Bioquímica, Bibliotecología y Ciencias de la Información, Derecho y Matemáticas), las clasificaciones producidas por diferentes variantes frecuentemente se contradecían entre sí. Por ejemplo, Matemáticas fue clasificada 221^a de 224 categorías por RS_P[1 - SC] pero 79^a por RS_G[1 - SC], a pesar de la fuerte correlación de Spearman general ($\rho = 0,91$) entre estas dos variantes. Wang y Schneider concluyeron que “las mediciones actuales de interdisciplinariedad deberían interpretarse con mucha precaución”.

Medidas de diversidad verdadera tipo Hill

Zhang, Rousseau y Glanzel (2016) propusieron medidas tipo Hill adaptadas de la literatura de diversidad ecológica (Hill, 1973; Jost, 2006; Leinster y Cobbold, 2012). La forma general es

$${}^q D^S = \left(\sum_{i=1}^N p_i \left(\sum_{j=1}^N s_{ij} p_j \right)^{q-1} \right)^{1/(1-q)}$$

donde $\mathbf{S} = [s_{ij}]$ es una matriz de similaridad con $s_{ii} = 1$ y $0 \leq s_{ij} = s_{ji} \leq 1$, y q es un parámetro de sensibilidad. El caso especial $q = 2$ produce

$${}^2D^S = \frac{1}{\sum_{i,j} s_{ij} p_i p_j}$$

que está relacionado con, pero es distinto de, el índice de Rao-Stirling: sustituyendo $d_{ij} = 1 - s_{ij}$, se obtiene ${}^2D^S = 1/(1 - \Delta)$. Mientras que el índice de Rao-Stirling está acotado entre 0 y 1, ${}^2D^S$ varía de 1 (una única categoría, o categorías perfectamente similares) a N (todas las categorías igualmente abundantes y máximamente disimilares), lo que le confiere la interpretación de un “número efectivo de disciplinas”.

La ventaja matemática de las medidas tipo Hill es que satisfacen seis propiedades deseables que los índices basados en entropía violan (Jost, 2006, 2009): simetría, independencia de salida cero, el principio de transferencia, invarianza de escala, el principio de replicación y normalización. El principio de replicación es especialmente importante para discusiones de política: si m portafolios de investigación igualmente diversos y no-superpuestos se agrupan, una medida de diversidad verdadera debería dar al portafolio agrupado una diversidad de $m \cdot D_0$. La entropía de Shannon y el índice de Simpson fallan esta prueba. Solo cuando se trabaja con diversidades verdaderas tiene sentido discutir cambios porcentuales en diversidad — una propiedad que hace las medidas tipo Hill particularmente atractivas para estudios longitudinales y comparativos. En una demostración empírica, Zhang, Rousseau y Glanzel (2016) mostraron que ${}^2D^S$ discrimina más efectivamente que el índice de Rao-Stirling entre revistas que abarcan un rango desde matemáticas especializadas a ciencia multidisciplinaria.

Multi-asignación y el enfoque de Morillo-Bordons-Gomez

Una tradición completamente diferente construye indicadores de interdisciplinariedad desde la multi-asignación de revistas a categorías de clasificación en lugar de desde análisis de listas de referencias. Morillo, Bordons y Gomez (2001) introdujeron un conjunto de indicadores para el sistema de categorías temáticas ISI (ahora Clarivate): el porcentaje de revistas multi-asignadas en una categoría, el porcentaje de revistas asignadas a categorías fuera del área de investigación, y la concentración de referencias entre categorías (medida vía el índice de Pratt). Estos indicadores fueron validados mediante un estudio de caso en Química, donde Química Aplicada — una disciplina “horizontal” con 83% de revistas multi-asignadas — puntuó consistentemente más alto que Ciencia de Polímeros (39% multi-asignado) en todos los indicadores.

En un estudio de seguimiento, Morillo, Bordons y Gomez (2003) extendieron el análisis a todas las categorías ISI y establecieron una tipología de cuatro clústeres de disciplinas basada en porcentaje de multi-asignación, porcentaje de enlaces externos, diversidad de enlaces, y fuerza de enlaces (el último medido por el coseno de Salton de conjuntos de revistas compartidas). Dos de los cuatro clústeres fueron caracterizados como reflejando, respectivamente, “gran interdisciplinariedad” — en la que enlaces conectan categorías de diferentes áreas

de investigación (ej., Biotecnología vinculando Ciencias de la Vida e Ingeniería) — y “pequeña interdisciplinariedad” — en la que enlaces permanecen dentro de la misma área (ej., Trasplante vinculando Cirugía e Inmunología). Esta distinción entre integración de categorías distantes y cercanas anticipó el énfasis posterior en disparidad en el marco de Stirling.

Estos indicadores de multi-asignación tienen ventajas distintas: son directos de computar, no requieren una matriz de similaridad, y pueden aplicarse a nivel macro (áreas de investigación) así como nivel meso (categorías y revistas). Su desventaja es la sensibilidad al esquema de clasificación ISI mismo — particularmente para categorías “horizontales” como Química o Física, donde ISI limita artificialmente la multi-asignación. Tampoco operan a nivel de artículo individual, lo que limita su aplicabilidad en evaluación a nivel de investigador.

Estimación de matriz de similaridad

Todas las medidas sensibles a disparidad dependen de una matriz \mathbf{S} (o su complemento $\mathbf{D} = \mathbf{1} - \mathbf{S}$) codificando relaciones por pares entre categorías de clasificación. El enfoque estándar construye \mathbf{S} desde flujos de citas intercategorías usando una similaridad de coseno. Históricamente, esta construcción extiende métodos de mapeo de citas revista-revista desarrollados para trabajo de clasificación a escala JCR (Leydesdorff, 2006). Wang y Schneider (2020) distinguieron dos variantes: el coseno vectorial de Salton $SC(i, j) = \sum_k c_{ik} c_{jk} / \sqrt{\sum_k c_{ik}^2 \cdot \sum_k c_{jk}^2}$, que compara los perfiles de citación de dos categorías, y el coseno escalar de Ochiai $SO(i, j) = (c_{ij} + c_{ji}) / \sqrt{(\sum_k c_{ik} + \sum_k c_{ki})(\sum_k c_{jk} + \sum_k c_{kj})}$, que usa intercambio de citas bilateral directo. Estas dos formulaciones de coseno pueden producir paisajes de similaridad sustancialmente diferentes: las matrices basadas en SO están extremadamente sesgadas a la izquierda (la mayoría de pares tienen disimilaridad cercana a 1), lo que significa que las variantes de Rao-Stirling basadas en SO se aproximan al índice de Simpson en la práctica.

Un grado de libertad adicional es la transformación de similaridad a disimilaridad. La elección estándar $d_{ij} = 1 - s_{ij}$ es intuitiva pero no la única opción; $d_{ij} = 1/s_{ij}$ también ha sido utilizada (Jensen y Lutkouskaya, 2014). Wang y Schneider (2020) mostraron que la combinación de variante de coseno y transformación de disimilaridad explica más varianza en resultados de Rao-Stirling que cualquier otra decisión metodológica única.

El requerimiento de una gran base de datos de citas es en sí mismo una barrera de entrada. Cantone, Zheng, Tomaselli y Nightingale (2025) recientemente propusieron una alternativa: estimar matrices de similaridad directamente desde modelos de lenguaje grandes (LLMs). En su protocolo experimental, ChatGPT, Gemini y Claude recibieron cada uno prompts para producir estimaciones numéricas de similaridad para pares de disciplinas bajo dos taxonomías. A través de 228 matrices muestreadas (16.200 estimaciones individuales), evaluaron cinco propiedades: precisión (varianza inversa entre prompts idénticos repetidos), acuerdo (correlación entre modelos), resiliencia (sensibilidad a reformulación

semánticamente trivial de nombres de disciplinas), robustez (sensibilidad a reordenamiento), y explicabilidad. Gemini logró estimaciones más cercanas a matrices tradicionales basadas en citas; Claude mostró un perfil balanceado; y ChatGPT exhibió resiliencia superior a variación de prompts. Aunque ninguno de los modelos alcanzó acuerdo perfecto con líneas base basadas en citas, los autores concluyeron que la estimación basada en LLM está “suficientemente bien” adaptada para la tarea y ofrece una alternativa de bajo costo sin base de datos que podría democratizar el acceso a medición de IDR sensible a disparidad.

Normalización, calibración y granularidad de clasificación

Una dificultad generalizada con todos los indicadores de diversidad es su dependencia del sistema de clasificación. Una publicación puede parecer más interdisciplinaria bajo el esquema de 250 categorías de Web of Science que bajo el esquema de 40 categorías de la OCDE simplemente porque mayor granularidad crea más fronteras de categorías a cruzar. Zhang, Rousseau y Glanzel (2016) demostraron esto directamente: las clasificaciones de revistas por ${}^2D^S$ cambiaron sustancialmente cuando se midieron contra los 68 subcampos ECOOM versus los 16 campos principales ECOOM, con correlaciones de Spearman entre las dos clasificaciones de solo 0,79. Esta sensibilidad de clasificación se amortigua, aunque no se elimina, cuando la medida incorpora disparidad — dividir un campo en subcampos produce subcampos que son similares, y su contribución a índices sensibles a disparidad se atenúa correspondientemente (Zwanenburg, Nakhoda y Whigham, 2022).

La normalización de campo plantea complicaciones adicionales. Un artículo en Matemáticas citando tres categorías puede representar mayor amplitud de conocimiento, relativo a normas disciplinarias, que un artículo en Biomedicina citando diez. Sin corrección de línea base, las puntuaciones brutas de diversidad penalizan campos que son naturalmente especializados y recompensan campos que son inherentemente difusos. Leydesdorff, Wagner y Bornmann (2019) propusieron una medida de diversidad descompuesta (DIV) como producto de componentes de variedad, balance ($1 - \text{Gini}$) y disparidad normalizados, permitiendo que cada componente sea inspeccionado por separado. Si estos componentes deberían combinarse multiplicativamente (Leydesdorff, Wagner y Bornmann, 2019) o aditivamente (Mutz, 2022) sigue siendo una pregunta abierta que la definición de IDR sola no resuelve (Zwanenburg, Nakhoda y Whigham, 2022).

Evaluación de validez: la evaluación de Zwanenburg

La evaluación de validez más sistemática hasta la fecha es la de Zwanenburg, Nakhoda y Whigham (2022), quienes evaluaron 21 medidas de IDR contra ocho criterios derivados de una síntesis de 25 definiciones de interdisciplinariedad. Los ocho criterios están organizados bajo cuatro encabezados:

1. *Aplicabilidad:* (1a) Multi-objeto — la medida debería ser aplicable a

artículos, autores, instituciones, revistas y disciplinas; (1b) Independencia de tamaño — las puntuaciones no deberían variar meramente porque el objeto de estudio representa más o menos publicaciones.

2. *Evidencia de integración*: (2) La medida debería basarse en evidencia de que el conocimiento es efectivamente integrado, no meramente que múltiples disciplinas están representadas. El análisis de referencias a nivel de artículo proporciona evidencia adecuada; recuentos agregados de citas a nivel de revista no.
3. *Identificación de disciplina*: (3a) Asignación válida de referencias a disciplinas; (3b) Identificación de todas las disciplinas fuente (completitud); (3c) Bajo sesgo de clasificación — la medida no debería producir puntuaciones muy diferentes cuando se aplica a clasificaciones de diferente granularidad.
4. *Captura de diversidad*: (4a) Sensibilidad a las tres dimensiones de diversidad (variedad, balance, disparidad); (4b) Descomponibilidad en puntuaciones separadas para cada dimensión.

De las 21 medidas evaluadas, solo seis cumplieron el criterio de evidencia de integración de conocimiento (criterio 2): el índice de Rao-Stirling, la medida tipo Hill, la medida de Coherencia, el indicador DIV, el indicador de diversidad general (d_{ive}), y el Índice de Simpson Inverso aplicado a nivel de artículo. Las 15 restantes o bien operaban en niveles agregados que excluyen evidencia de integración, o se basaban en superposición de clasificación en lugar de evidencia basada en citas. Dentro de las seis que cumplieron el criterio 2, las medidas de Rao-Stirling y tipo Hill también satisfacían el criterio 4a (todos los aspectos de diversidad capturados) y el criterio 3c (bajo sesgo de clasificación), pero ninguna era descomponible en puntuaciones separadas de variedad, balance y disparidad (criterio 4b). Solo la medida DIV y el indicador de diversidad general d_{ive} cumplieron todos los cuatro criterios 1a, 1b, 2, 3c, 4a y 4b.

Zwanenburg, Nakhoda y Whigham advirtieron que ninguna medida única satisfacía los ocho criterios, y que el criterio de asignación de disciplina válida (3a) permanecía sin resolver para cada medida que dependía de mapeos revista-categoría — aproximadamente el 30% de referencias en su base de datos institucional fueron asignadas a múltiples categorías WoS, creando ambigüedades de asignación que se propagan en puntuaciones de diversidad infladas.

Diversidad de autores versus referencias

La elección de *qué* diversificar introduce una dimensión adicional. La mayoría de indicadores miden diversidad sobre la lista de referencias, pero una tradición alternativa mide diversidad sobre las afiliaciones disciplinarias de coautores. Abramo, D'Angelo y Zhang (2018) compararon los dos enfoques usando 43.667 publicaciones universitarias italianas, particionadas en artículos de autor único (por construcción no interdisciplinares bajo el método de autores), artículos multi-autor de campo único, y artículos de múltiples campos. Encontraron convergencia general: la diversidad de lista de referencias aumentaba con el número de sectores disciplinarios distintos (SDSs) reflejados en la autoría, y la

disparidad era mayor cuando los SDSs de los autores abarcaban diferentes áreas disciplinares universitarias (UDAs) en lugar de la misma UDA. Sin embargo, surgieron excepciones individuales impactantes. Las tres publicaciones con la puntuación más alta de diversidad integrada en todo el conjunto de datos fueron artículos de autor único, precisamente los “integradores intrapersonales” cuya amplitud de conocimiento no puede detectarse mediante el método de autores. Este hallazgo resalta una limitación estructural de enfoques basados en autores y subraya el valor complementario del análisis basado en referencias para identificar integración de conocimiento individual.

Resumen

El panorama de indicadores de diversidad es rico pero fragmentado. Medidas que incorporan disimilaridad por pares (Rao-Stirling, tipo Hill, DIV) forman un clúster empírico; medidas que no (Shannon, Simpson, recuentos de multi-asignación) forman otro, y las correlaciones entre grupos son débiles (Wang y Schneider, 2020). La elección de especificación de matriz de similaridad explica tanta varianza en resultados como la elección de fórmula de índice. Ninguna medida única satisface todos los criterios de validez (Zwanenburg, Nakhoda y Whigham, 2022), y la estimación de similaridad basada en LLM abre una alternativa prometedora pero todavía inmadura a matrices derivadas de citas (Cantone, Zheng, Tomaselli y Nightingale, 2025). Para evaluación aplicada, la implicación es clara: los indicadores de diversidad deberían reportarse junto con sus especificaciones computacionales (esquema de clasificación, método de similaridad, transformación de disimilaridad, nivel de agregación), y las conclusiones que dependen de una única variante de indicador deberían tratarse con precaución.

Indicadores de Coherencia

Los indicadores de diversidad, por ricamente especificados que estén, responden solo una pregunta: *¿cuán heterogéneos son los insumos de conocimiento?* Son silenciosos sobre si esos insumos heterogéneos han sido tejidos en un tejido intelectual unificado o meramente colocados lado a lado. Rafols y Meyer (2009) introdujeron el concepto de *coherencia* para llenar este vacío, definiéndola como “la medida en que tópicos, conceptos, herramientas, datos, etc. específicos usados en un proceso de investigación están relacionados” (p. 175). Mientras que la diversidad captura la amplitud categórica de referencias, la coherencia captura la estructura relacional entre los ítems dentro de esas categorías — la intensidad de su integración mutua.

La distinción importa para la evaluación. Alta diversidad sola no garantiza que fuentes de conocimiento dispares hayan sido sintetizadas; puede reflejar mera yuxtaposición o amplitud polimática a través de literaturas no relacionadas. La coherencia proporciona la señal faltante. Además, la interpretación funcional de la coherencia depende de la unidad de análisis. Alta coherencia en la lista de referencias de un artículo indica que el artículo se construye sobre una

especialidad establecida, internamente conectada. Alta coherencia a través de las publicaciones de un centro de investigación indica que el centro está logrando su misión integradora. Baja coherencia, inversamente, señala que cuerpos de conocimiento previamente no relacionados están siendo puestos en contacto — un estado de integración interdisciplinaria *potencial* que puede madurar con el tiempo.

Operacionalización mediante acoplamiento bibliográfico

Rafols y Meyer (2009) operacionalizaron la coherencia mediante acoplamiento bibliográfico: dos publicaciones están enlazadas en la medida en que comparten referencias, y la densidad de la red resultante sirve como indicador de coherencia para el conjunto. La similaridad entre dos publicaciones a y b se computa usando el coseno de Salton,

$$s_{ab} = \frac{\mathbf{r}_a \cdot \mathbf{r}_b}{\|\mathbf{r}_a\| \|\mathbf{r}_b\|}$$

donde \mathbf{r}_a es el vector de referencias binario de la publicación a . La normalización de coseno controla por el número total de referencias en cada publicación, una propiedad deseable que protege contra artefactos impulsados por tamaño.

De la matriz de similaridad resultante se derivan dos indicadores de coherencia a nivel de red:

Fuerza media de enlace (S). La media de las entradas fuera de la diagonal de la matriz de acoplamiento bibliográfico normalizada,

$$S = \frac{2}{N(N-1)} \sum_{a < b} s_{ab}$$

donde N es el número de publicaciones. En una red binaria S se reduce a densidad de red ordinaria; en una red valuada captura tanto la proporción de enlaces realizados como su intensidad promedio. S está acotado entre 0 y 1 y se encontró que era invariante de escala a través de tamaños de red de 10 a 1.275 nodos en la muestra de referencia de kinesina de Rafols y Meyer.

Longitud media de camino (L). La distancia promedio de camino más corto entre todos los pares de nodos en la red de similaridad binarizada. La binarización requiere un umbral τ bajo el cual las similaridades por pares son tratadas como cero; Rafols y Meyer adoptaron $\tau = 0.05$ (equivalente a requerir al menos una referencia compartida en una bibliografía de 20 referencias) para suprimir enlaces espurios surgiendo de referencias generales altamente citadas. Valores más bajos de L indican un cuerpo de trabajo más compacto, internamente conectado. En su muestra de motores moleculares, S y L estaban altamente correlacionados ($r \approx 0.95$), sugiriendo que los dos indicadores capturan esencialmente la misma propiedad estructural y que S sola puede ser suficiente en muchas aplicaciones.

La elección de acoplamiento bibliográfico — en lugar de co-citación — como la relación subyacente es deliberada. El acoplamiento bibliográfico mira hacia adelante: refleja las fuentes de conocimiento que los autores eligieron utilizar al momento de escribir, en lugar de los patrones de audiencia que emergen después de la publicación. Esto lo hace aplicable a publicaciones recientes para las cuales una ventana de citación aún no se ha acumulado.

Evidencia empírica: ortogonalidad respecto de diversidad

Rafols y Meyer (2009) probaron sus indicadores de coherencia en 12 artículos extraídos de la literatura de motores moleculares. Diversidad y coherencia se encontraron no correlacionados — las dos dimensiones ofrecieron “perspectivas ortogonales” sobre interdisciplinariedad. Los valores de coherencia abarcaron un rango amplio (S desde 0,024 hasta 0,113). En un extremo, Noji (1997) exhibió $S = 0,024$: su red de referencias mostró una división clara entre las literaturas de bioenergética y motores lineales, conectadas solo mediante un único artículo de revisión. Esta baja coherencia señaló un acto seminal de integración en el que dos hebras de investigación previamente separadas estaban siendo puestas en contacto por primera vez. En el otro extremo, Tomishige (2002) exhibió $S = 0,113$: se basaba en lo que para entonces se había convertido en una especialidad interdisciplinar establecida, y sus referencias formaban un clúster denso, internamente conectado.

Wang y Schneider (2020) confirmaron el hallazgo de ortogonalidad a una escala mucho mayor, computando 16 medidas de interdisciplinariedad para 224 Categorías Temáticas de Web of Science. Su indicador de coherencia — adaptado de Wang (2016), quien lo operacionalizó como el número de enlaces de citas entre referencias citadas pertenecientes a diferentes categorías ponderadas por su disimilitud — mostró solo correlaciones débiles a moderadas con la familia de diversidad: $r = 0,23$ con proporción de multi-asignación, $r = 0,44$ con diversidad de Simpson, $r = 0,46$ con entropía de Shannon, y $r = 0,50$ con el coeficiente de Gini invertido. La correlación con centralidad de intermediación fue insignificante ($r = -0,03$), y la correlación con el coeficiente de clúster fue negativa ($r = -0,36$). Estos resultados colocan la coherencia en un clúster empírico distinto de medidas de diversidad categórica, reforzando la afirmación de que captura una dimensión genuinamente independiente de interdisciplinariedad.

El marco diversidad–coherencia

La observación conjunta de diversidad y coherencia da lugar a una matriz interpretativa útil, propuesta como marco bidimensional por Rafols y Meyer (2009). Baja diversidad combinada con alta coherencia caracteriza investigación disciplinar especializada — trabajo estrechamente integrado dentro de un único paradigma. Baja diversidad con baja coherencia indica que especialidades distantes dentro de la misma disciplina están siendo conectadas, sin todavía alcanzar integración completa. Alta diversidad con baja coherencia representa la forma más naciente de integración interdisciplinar: cuerpos de conocimiento

hasta ahora no relacionados están siendo yuxtapuestos por primera vez, como en el caso Noji (1997). Finalmente, alta diversidad con alta coherencia marca el estado maduro de investigación interdisciplinaria especializada, donde fuentes de conocimiento antes distantes han sido tejidas en un tejido intelectual estable.

El marco implica una trayectoria de integración de conocimiento que se mueve de baja a alta coherencia con el tiempo: la integración pionera gradualmente se consolida en especialidades interdisciplinares establecidas. Para propósitos de evaluación, esta trayectoria habilita una distinción entre integración en etapa temprana (alto potencial, baja consolidación) y campos interdisciplinares maduros (alto potencial realizado), un matiz que las medidas escalares de diversidad solas no pueden capturar.

Operacionalizaciones alternativas

Centralidad de intermediación. Leydesdorff y Rafols (2011) exploraron la centralidad de intermediación como un indicador alternativo relacionado con coherencia a nivel de revista. La centralidad de intermediación de Freeman, definida como

$$g_i = \sum_{\substack{j,k \\ j \neq k \neq i}} \frac{g_{ijk}}{g_{jk}}$$

donde g_{jk} es el número total de geodésicas entre nodos j y k y g_{ijk} el número de esas geodésicas que pasan por i , mide la medida en que una revista ocupa una posición intermedia en la red de citas. Sin embargo, la intermediación bruta está confundida por tamaño: revistas multidisciplinares grandes como *Nature* y *Science* puntúan alto simplemente por su centralidad de grado. Leydesdorff y Rafols abordaron esto computando intermediación en redes normalizadas por coseno, después de lo cual revistas de ciencias sociales emergieron como los puentes interdisciplinares más prominentes entre 8.207 revistas JCR. En un análisis factorial rotado (Bollen et al., 2009), la intermediación cargó cerca del origen — casi ortogonal tanto a indicadores basados en citas como vectoriales — sugiriendo que captura una dimensión posicional distinta.

Sensibilidad a medida de distancia. Leydesdorff y Rafols (2011) también documentaron una sensibilidad impactante de la diversidad de Rao-Stirling a la elección de matriz de distancia. Cuando computaron el indicador usando $(1 - \cos)$ versus distancias euclidianas relativas entre las mismas 8.207 revistas, la correlación de rango-orden de Spearman entre las dos clasificaciones de interdisciplinariedad resultantes fue $\rho = -0,012$ en la dirección citada y $\rho = -0,015$ en la dirección citante — efectivamente cero y, en el último caso, nominalmente negativa. En un análisis factorial rotado la variante basada en eucliana cargó en un componente diferente de todos los otros indicadores, confirmando que las dos formulaciones de distancia capturan características estructurales fundamentalmente diferentes de la red de citas. Este hallazgo dramático subraya

que la elección de medida de distancia no es un detalle técnico menor sino un determinante de primer orden de interdisciplinariedad medida.

Homogeneidad IKF. Zhou, Guns y Engels (2023) propusieron el marco de Flujo de Conocimiento Interdisciplinario (IKF), que descompone relaciones de citas intercampos en tres aspectos: *amplitud* (la fracción de publicaciones que citan una disciplina externa dada), *intensidad* (la participación de citas salientes dirigidas a esa disciplina), y *homogeneidad* (la fracción de referencias de una disciplina que son co-citadas por la disciplina objetivo). La dimensión de homogeneidad está más directamente relacionada con el concepto de coherencia de Rafols y Meyer: mide similaridad cognitiva vía la superposición de bases de conocimiento. Empíricamente, la homogeneidad correlaciona moderadamente con amplitud ($R^2 = 0,47$) pero solo débilmente con intensidad ($R^2 = 0,25$), indicando que estos aspectos capturan facetas diferentes de la relación interdisciplinaria. Un patrón revelador es que baja homogeneidad combinada con alta amplitud caracteriza disciplinas metodológicas (ej., Matemáticas Aplicadas citada por Ecología o Genética): sus herramientas se difunden ampliamente a pesar de una gran distancia cognitiva del campo receptor.

Consideraciones computacionales

Todos los indicadores de coherencia requieren un cómputo de similaridad o acoplamiento por pares cuyo costo crece como $O(N^2)$ en el número de publicaciones o revistas. Para centralidad de intermediación el costo es mayor, en $O(N^3)$ en la implementación ingenua, porque se requiere enumeración de caminos más cortos en la red completa. En la práctica, el coseno de Salton es preferido para construir la matriz de acoplamiento porque es no-paramétrico, maneja vectores de referencias dispersos naturalmente, y normaliza por longitud de publicación. La matriz de co-ocurrencia misma puede obtenerse eficientemente vía multiplicación de matrices ($\mathbf{A}\mathbf{A}^\top$ para acoplamiento bibliográfico, $\mathbf{A}^\top\mathbf{A}$ para co-citación). La selección de umbral al binarizar redes valuadas sigue siendo un compromiso práctico entre filtrado de ruido y pérdida de información; la elección común de $\tau = 0,05$ es adecuada para listas de referencias de tamaño moderado pero puede necesitar ajuste para campos con prácticas de citación sustancialmente diferentes.

Resumen

Los indicadores de coherencia complementan la diversidad midiendo la profundidad de integración de conocimiento en lugar de la amplitud de dispersión categórica. Múltiples operacionalizaciones están disponibles — densidad de acoplamiento bibliográfico, centralidad de intermediación, y homogeneidad de co-citación — y la evidencia empírica muestra consistentemente que están solo débilmente correlacionados con medidas de diversidad (típicamente $r < 0,5$), confirmando que la coherencia constituye una dimensión de medición independiente. Combinada con diversidad, la coherencia habilita una caracterización más rica que distingue integración potencial de realizada, una distinción que

la siguiente subsección desarrolla al examinar indicadores que explícitamente combinan ambas dimensiones.

Indicadores de Difusión

Fundamentos conceptuales

Los indicadores examinados en las subsecciones precedentes — diversidad, coherencia y sus compuestos — todos miran *hacia atrás* desde una publicación al conocimiento que utiliza. Los indicadores de difusión invierten la dirección causal: miran *hacia adelante* desde una publicación a los nuevos cuerpos de investigación que la citan, midiendo el alcance interdisciplinar de salidas de conocimiento en lugar de la heterogeneidad de insumos de conocimiento (Cantone, 2024; Leydesdorff, Wagner y Bornmann, 2019). Formalmente, el vector de referencias $\mathbf{p}(x)$ que subyace a medidas de integración es reemplazado por un vector de citas $\mathbf{q}(x)$, donde $q_i(x)$ denota la proporción de publicaciones *citantes* que pertenecen a la categoría disciplinar i .

Cantone (2024) sitúa la difusión al final de una cadena causal temporal: la cognición precede la producción, y la difusión sigue a ambas. Mientras que la integración captura la amplitud disciplinar de insumos que dieron forma a una pieza de investigación, la difusión captura la amplitud disciplinar de las comunidades que subsecuentemente la absorbieron. Esta asimetría es consecuencial para la evaluación. Un estudio que integra conocimiento desde muchos campos puede sin embargo permanecer confinado a su disciplina de origen en términos de audiencia; inversamente, un estudio de base estrecha puede difundirse ampliamente si sus métodos o hallazgos resultan transferibles. La difusión constituye por tanto una dimensión de medición independiente, una que complementa la integración en lugar de duplicarla.

Definiciones operacionales

La operacionalización más simple de difusión es la *fracción de citas recibidas desde fuera del campo primario*, análoga a la proporción de referencias externas (p_{fuera}) usada para integración. Medidas más informativas aplican la misma maquinaria de diversidad introducida en la Sección 3.1 pero al vector de citas $\mathbf{q}(x)$ en lugar del vector de referencias $\mathbf{p}(x)$.

Leydesdorff, Wagner y Bornmann (2019) proponen el indicador *DIV*, que descompone la diversidad en sus tres componentes de Stirling y los aplica a la dirección *citada*. Para una revista c ,

$$\text{DIV}_c = \frac{n_c}{N} (1 - G_c) \frac{\sum_{i \neq j} d_{ij}}{n_c(n_c - 1)}$$

donde n_c es el número de categorías de Web of Science con participaciones de

citas no-cero (variedad), N el número total de categorías disponibles, $d_{ij} = 1 - \cos(\mathbf{v}_i, \mathbf{v}_j)$ la disparidad entre categorías i y j , y G_c el coeficiente de Gini midiendo la desigualdad de la distribución de citas entre categorías. El componente de balance entra como $(1 - G_c)$: uniformidad perfecta produce $G_c = 0$ y balance máximo, mientras que concentración en una única categoría produce $G_c \rightarrow 1$ y balance que desaparece. Gini se computa vía la fórmula de orden ascendente,

$$G = \frac{\sum_{i=1}^n (2i - n - 1) x_i}{n \sum_{i=1}^n x_i}$$

donde las x_i son las participaciones de citas ordenadas en orden no-decreciente. La factorización triple hace DIV *monótonamente* creciente en cada componente, una propiedad que el índice de Rao-Stirling carece porque RS combina variedad y balance *ex ante* mediante el índice de concentración de Simpson (Leydesdorff, Wagner y Bornmann, 2019).

Una medida complementaria es la *coherencia* de la distribución citante,

$$C = \sum_{i \neq j} p_{ij} d_{ij}$$

que captura la distancia cognitiva promedio entre categorías de citas co-ocurrentes. Cuando se aplica a la dirección citada, alta coherencia indica que las comunidades citantes mismas abarcan partes distantes del paisaje disciplinar — una señal fuerte de amplitud de difusión.

Dinámicas temporales y desafíos de medición

A diferencia de indicadores basados en referencias, que son fijos al momento de publicación, las medidas de difusión son inherentemente dinámicas. Las citas se acumulan durante meses y años, así que una puntuación de difusión computada un año después de publicación puede diferir sustancialmente de una computada cinco años después. Esto crea una dependencia de ventana temporal que las medidas de integración no enfrentan. Además, las citas no están bajo control del autor: la misma publicación puede atraer citas desde campos inesperados dependiendo de cambios en moda de investigación, relevancia de política, o adopción metodológica (Cantone, 2024).

Surgen tres complicaciones adicionales. Primero, la difusión está confundida con *impacto científico*: publicaciones altamente citadas reciben citas de más categorías simplemente en virtud de su volumen de citas, incluso si cada cita individual proviene de dentro de la disciplina de origen. Cualquier medida de difusión debe por tanto interpretarse junto con recuentos totales de citas.

Segundo, una fracción no trivial de publicaciones recibe citas cero o casi cero dentro de ventanas de evaluación típicas, haciendo puntuaciones de difusión indefinidas o degeneradas — un problema de censura sin análogo limpio en el lado de referencias. Tercero, las prácticas de citación varían entre disciplinas tanto en volumen como en latencia, significando que puntuaciones brutas de difusión no son directamente comparables entre campos sin normalización.

Estos desafíos hacen los indicadores de difusión más apropiados para análisis longitudinales o de series temporales, donde la evolución temporal del alcance interdisciplinario puede rastrearse explícitamente en lugar de congelarse en un corte arbitrario.

Patrones empíricos desde estudios a gran escala

Leydesdorff, Wagner y Bornmann (2019) computaron DIV en ambas direcciones citante y citada para 11.487 revistas indexadas en el *Journal Citation Reports* (JCR 2016). En la dirección citada (difusión), *PLOS ONE* alcanzó la puntuación DIV más alta (0,142), seguida por *Science* (0,125) y *Nature* (0,124) — revistas cuyo alcance editorial es suficientemente amplio para atraer comunidades citantes desde todo el mapa disciplinario. Notablemente, el índice de diversidad de Rao-Stirling aplicado a los mismos datos produjo una clasificación sustancialmente diferente: *Daedalus-US* (0,939), *Qualitative Inquiry* (0,936) y *Critical Inquiry* (0,927) lideraron, ilustrando cómo la elección de medida puede alterar conclusiones empíricas incluso cuando los datos subyacentes son idénticos.

La divergencia es instructiva. RS está dominado por su término de disparidad, que favorece revistas citadas por unas pocas categorías muy distantes; DIV da peso proporcional a variedad y balance, recompensando revistas que atraen citas de muchas categorías en participaciones aproximadamente uniformes. En un análisis factorial sobre el conjunto completo de revistas, DIV en la dirección citada cargó en el mismo factor que centralidad de intermediación ($\rho = 0,66$) y factor de impacto, mientras que RS cargó separadamente ($\rho = 0,41$ con intermediación). Este patrón sugiere que DIV captura un rol estructural de *intermediación* — revistas que sirven como conductos entre comunidades disciplinarias — mientras que RS captura un aspecto diferente de alcance intercampo que está menos alineado con posición de red.

Distinción de interdisciplinariedad basada en tópicos

La difusión a veces se confunde con un concepto relacionado pero distinto: *interdisciplinariedad de tópico*, definida como el grado en que el contenido de una publicación aborda temas de múltiples disciplinas. Xiang, Romero y Teplitskiy (2025) desenredan dos dimensiones que frecuentemente se confunden en trabajo empírico. *Interdisciplinariedad de tópico* se mide mediante la clasificación disciplinar del título y resumen de una publicación (ej., etiquetas de conceptos de OpenAlex), y refleja lo que el trabajo *aborda*. *Interdisciplinariedad de base de conocimiento* se mide mediante la composición disciplinar de la lista de

referencias, y refleja en lo que el trabajo *se basa*. Las dos correlacionan solo moderadamente ($r = 0,56$), confirmando que no son intercambiables.

Críticamente, las dos dimensiones portan asociaciones opuestas con resultados de revisión por pares. En un análisis de 128.950 manuscritos STEM enviados entre 2018 y 2022, Xiang, Romero y Teplitskiy (2025) encuentran que un aumento de una desviación estándar en interdisciplinariedad de base de conocimiento eleva la probabilidad de aceptación en 0,9 puntos porcentuales, mientras que el mismo aumento en interdisciplinariedad de tópico la *baja* en 1,2 puntos porcentuales. El término de interacción es positivo y significativo ($\beta = 0,042$): referencias amplias mitigan la penalidad incurrida por encuadre interdisciplinar. Estos hallazgos subrayan que la *dirección* del cruce de fronteras disciplinares importa: la integración (referencias) es recompensada; el abarcamiento temático (contenido) es penalizado a menos que esté respaldado por insumos de conocimiento demostrablemente amplios.

Relación con medidas de integración

Porque difusión e integración usan la misma maquinaria matemática — diversidad de Stirling, balance de Gini, matrices de disparidad — aplicada a vectores de entrada diferentes, una pregunta natural es si están empíricamente relacionados. Leydesdorff, Wagner y Bornmann (2019) reportan que la correlación entre DIV en la dirección citante (integración) y DIV en la dirección citada (difusión) es positiva pero lejos de la unidad, confirmando que los dos capturan fenómenos diferentes. Alta integración no garantiza alta difusión, y viceversa.

Cantone (2024) especula que el trabajo integrativo puede servir como un *factor macro antecedente* para difusión: investigación que sintetiza conocimiento desde múltiples campos puede volverse cognitivamente accesible a una audiencia más amplia, creando un efecto “super-propagador”. Esta hipótesis permanece no probada en la literatura y requeriría diseños longitudinales vinculando puntuaciones de integración al momento de publicación con trayectorias de difusión en años subsecuentes.

Preocupaciones de validez y estatus conceptual

El estatus conceptual de la difusión dentro de un marco de medición de interdisciplinariedad es objeto de debate. Cantone (2024) argumenta que la difusión “desafía la definición completa de IDR” porque mide un *efecto futuro* de la investigación en lugar de una propiedad del *proceso* de investigación. Bajo esta visión, la difusión es un *resultado* de la actividad científica — similar al impacto — en lugar de un *atributo* de la actividad misma. Si una publicación es citada por disciplinas distantes depende de factores mayormente externos al esfuerzo integrativo de los autores: políticas editoriales de revistas citantes, la disponibilidad del trabajo en bases de datos relevantes, y tendencias más amplias en política de investigación.

Esta tensión conceptual admite dos resoluciones. Se puede tratar la difusión

como una dimensión legítima de interdisciplinariedad, con el argumento de que el alcance interdisciplinar es en sí mismo una forma de cruce de fronteras que los marcos de evaluación deberían capturar. Alternativamente, se puede tratar la difusión como una *consecuencia* de interdisciplinariedad que es informativa para política pero no debería confundirse con la medición de producción interdisciplinar. La distinción no es meramente semántica: determina si los indicadores de difusión pertenecen a un panel diseñado para caracterizar investigación *tal como se produce* o en una evaluación separada de investigación *tal como se recibe*. Wang y Schneider (2020) notan que diferentes indicadores pueden capturar “diferentes entendimientos de un concepto tan multifacético como la interdisciplinariedad”, una advertencia que se aplica con particular fuerza al límite integración–difusión.

Orientación práctica y brechas de investigación

Dados los desafíos delineados arriba, los indicadores de difusión son más adecuados para evaluación retrospectiva de impacto y evaluación longitudinal en lugar de selección de proyectos *ex ante* o decisiones de financiamiento. Son más informativos cuando se aplican con ventanas temporales explícitas (ej., ventanas de citación de cinco o diez años) y cuando se acompañan de recuentos totales de citas que permiten a analistas distinguir alcance interdisciplinar genuino del efecto mecánico de alto volumen de citas.

Varias brechas de investigación permanecen. Los coeficientes de Gini del lado de citas y puntuaciones DIV no se computan rutinariamente en kits de herramientas bibliométricas estándar, limitando su adopción práctica. La estructura temporal de la difusión — si los patrones de citación interdisciplinar se estabilizan o continúan evolucionando décadas después de publicación — ha recibido poca atención sistemática. Finalmente, el bucle de retroalimentación hipotetizado entre integración y difusión — mediante el cual trabajo ampliamente integrativo atrae audiencias citantes más amplias, lo que a su vez estimula integración adicional — permanece como una pregunta empírica abierta que diseños de panel longitudinales están bien posicionados para abordar.

Indicadores de Novedad

Fundamento conceptual

Los indicadores de novedad abordan una pregunta distinta de la diversidad: si la investigación combina elementos disciplinares de maneras que no son meramente heterogéneas sino genuinamente *no conformes* relativo a la práctica establecida. Una publicación puede exhibir alta diversidad — basándose en muchos campos distantes — pero baja novedad si su combinación particular de campos se ha vuelto rutinaria, constituyendo efectivamente una nueva “disciplina interdisciplinar” (Cantone, 2024). La novedad captura *innovación* disciplinar, no *amplitud* disciplinar.

El núcleo conceptual es divergencia respecto de un punto de referencia. Dado un vector de proporción disciplinar $p(x)$ para una unidad de investigación x y una

distribución de referencia $p(E)$ representando composición disciplinar esperada, la novedad se operacionaliza como función de la discrepancia entre los dos:

$$\nabla = f(|p(x) - p(E)|).$$

La elección de referencia, la forma funcional f , y el tratamiento de similaridad disciplinar $z(i, j)$ distinguen los enfoques competitores revisados abajo. Un caso límite revelador conecta novedad con diversidad: cuando $p(x)$ y $p(E)$ tienen soportes no superpuestos, la divergencia cuadrática se descompone como $\sum_i [p_i(x) - p_i(E)]^2 = \sum_i p_i(x)^2 + \sum_i p_i(E)^2$, donde cada sumando es un término de tasa-de-repetición (Herfindahl) — una medida de diversidad (Cantone, 2024). La divergencia generaliza así la diversidad incorporando expectativas previas.

Enfoques de divergencia estadística

Varias familias de medidas de divergencia han sido propuestas para cuantificación de novedad. La más simple es la suma de diferencias cuadráticas entre $p(x)$ y $p(E)$, pero no ha visto adopción empírica. La divergencia chi-cuadrado y medidas teórico-informacionales relacionadas (Kullback–Leibler, información mutua) sufren de un defecto común: son indefinidas cuando $p_i(E) = 0$ para cualquier categoría i donde $p_i(x) > 0$, forzando a analistas a ignorar precisamente las contribuciones disciplinares más novedosas (Cantone, 2024).

Dos alternativas evitan esta singularidad. El índice de *Jaccard probabilístico* compara distribuciones mediante su superposición:

$$\nabla_{\text{PJ}}(x) = 1 - \frac{\sum_i \min(p_i(x), p_i(E))}{\sum_i \max(p_i(x), p_i(E))}.$$

Esta es una variante normalizada de entropía mutua generalizada con interpretación puramente frecuentista (Moulton y Jiang, 2018). Es fácil de computar pero difícil de interpretar en términos sustantivos (Cantone, 2024). La *distancia de Hellinger* adopta un enfoque geométrico:

$$\nabla_{\text{Hel}}(x) = \frac{1}{\sqrt{2}} \sqrt{\sum_i \left(\sqrt{p_i(x)} - \sqrt{p_i(E)} \right)^2}.$$

La distancia de Hellinger ofrece escalamientos estables y propiedades métricas bien entendidas pero depende de una interpretación euclíadiana del espacio de probabilidad que es difícil de comunicar a audiencias no especialistas (Cantone, 2024).

Ni el Jaccard probabilístico ni la distancia de Hellinger dan cuenta de similaridad intercategorías $z(i, j)$. Una extensión ponderada por similaridad reemplaza cada proporción p_i con una versión suavizada

$\psi_{i,z}(x) = [\sum_j p_j(x) z(i, j)] / [\sum_i \sum_j p_j(x) z(i, j)]$, que puede entonces sustituirse en cualquier fórmula de divergencia (Cantone, 2024). Esta extensión acerca la medición de novedad a índices de diversidad conscientes de disparidad, pero al costo de requerir la misma infraestructura de matriz de similaridad.

Método de permutación

La medida de novedad más influyente en la literatura bibliométrica es el enfoque de combinaciones atípicas de Uzzi et al. (2013). En lugar de comparar distribuciones agregadas, examina todas las co-ocurrencias por pares de categorías disciplinarias dentro de un cuerpo de investigación. Para cada par observado (i, j) , se computa una puntuación z contra un modelo nulo aleatorizado:

$$\nabla_{\text{Uzzi}}(p_{i,j}, x) = \frac{p_{i,j}(x) - e_{i,j}(x)}{\sigma[e_{i,j}(x)]}, \quad p_{i,j}(x) > 0,$$

donde $e_{i,j}(x)$ y $\sigma[e_{i,j}(x)]$ son la media y desviación estándar de la frecuencia de pares bajo redes de citas permutadas que preservan recuentos de referencias. El método produce una *distribución* de puntuaciones de atipicidad dentro de x , de la cual Uzzi et al. derivan dos estadísticos sumarios no paramétricos capturando conformidad y novedad.

Las afirmaciones empíricas que emergieron de este marco fueron altamente influyentes para política de investigación. Uzzi et al. (2013) reportaron que la ciencia de más alto impacto está “principalmente fundamentada en combinaciones excepcionalmente convencionales” pero simultáneamente presenta “intrusión de combinaciones inusuales”, y que los equipos son más propensos que autores solitarios a insertar emparejamientos novedosos en dominios de conocimiento familiares.

Trabajo de validación subsecuente, sin embargo, ha arrojado serias dudas sobre si la medida de Uzzi captura novedad como constructo distinto de diversidad. Bornmann (2019), usando evaluaciones expertas de F1000Prime como verdad de base, encontró que la medida de combinaciones atípicas *no* correlacionaba con novedad evaluada cualitativamente. Fontana et al. (2020) confirmaron este hallazgo y mostraron que la medida de Uzzi en cambio correlaciona con diversidad de Rao-Stirling — precisamente el constructo que fue diseñada para complementar, no replicar. Esta confusión de novedad y diversidad dentro de un único indicador socava el marco interpretativo original y pone en duda conclusiones de política extraídas de él. Cantone (2024) identifica una razón estructural para la confusión: el método de permutación confunde las dimensiones de novedad y diversidad en una única medida en lugar de separarlas sistemáticamente. Además, el método impone requerimientos de datos muy altos, necesitando listas completas de referencias y densidad de red suficiente para estimación confiable de modelo nulo.

Novedad temporizada

Wang et al. (2017) proponen un enfoque temporal que yace conceptualmente entre no conformidad estadística y ser pionero: para cada par (i, j) de categorías disciplinares con $p_{i,j} > 0$, registran la marca temporal $t_0(i, j)$ de su primera ocurrencia observada en un corpus de referencias. La puntuación de novedad temporizada es entonces:

$$\text{Novelty}(x) = \sum_{i,j} t_0(i, j) \cdot [1 - z(i, j)],$$

ponderando recencia de primera combinación por disparidad intercategorías. La medida es intuitiva — recompensa investigación que instancia emparejamientos disciplinares no observados previamente, especialmente entre campos cognitivamente distantes.

Sin embargo, la novedad temporizada enfrenta múltiples limitaciones serias. Requiere datos históricos de citas suficientes para identificar primeras ocurrencias confiablemente; en muestras conteniendo solo publicaciones recientes, todos los pares aparecen novedosos por construcción, produciendo puntuaciones engañosas (Cantone, 2024). La medida también es altamente sensible a granularidad de taxonomía, haciéndola inapropiada para sistemas de clasificación de grano grueso. Más críticamente, la medida carece de un mecanismo para calibrar innovación contra difusión: una combinación “pionera” que ningún trabajo subsecuente cita plantea la pregunta de si constituye innovación genuina o meramente una anomalía improductiva. Bornmann (2019) y Fontana et al. (2020) no encontraron concordancia entre puntuaciones de novedad temporizada y novedad evaluada por expertos, cuestionando la validez epistémica de la medida.

Especificación de punto de referencia

Todas las medidas de novedad basadas en divergencia dependen de la elección de punto de referencia $p(E)$, y esta elección está lejos de ser neutral. Tres enfoques han sido propuestos. Goyanes et al. (2020) adoptan un *previo uniforme* (peso igual entre todas las categorías observadas), que Cantone (2024) critica como poco realista: “virtualmente ninguna aplicación real espera una distribución perfectamente balanceada”. Uzzi et al. (2013) usan *permutación aleatorizada*, intercambiando citas mientras preservan recuentos de referencias para generar un modelo nulo impulsado por datos — el enfoque preferido para análisis citacionales. Cantone y Nightingale (2024) proponen un *punto de referencia jerárquico* en el que la distribución disciplinar de una unidad contenedora (ej., una revista) sirve como $p(E)$ para sus artículos constituyentes, explotando el anidamiento natural de unidades de publicación. Cada enfoque porta supuestos sobre qué constituye composición disciplinar “esperada”, y no ha emergido consenso sobre mejor práctica.

Exclusión del panel

A pesar de su importancia conceptual, los indicadores de novedad son excluidos del panel de medición propuesto en este artículo. Las razones son tanto operacionales como epistémicas. En el lado operacional, todos los enfoques viables requieren infraestructura de datos — redes de citas completas, líneas base temporales históricas, o aparato de permutación de listas de referencias — que excede sustancialmente lo que kits de herramientas bibliométricas estándar proporcionan. Los métodos de permutación imponen “requerimientos muy altos” (Cantone, 2024), y métricas basadas en redes rara vez son insesgadas en tamaños de muestra típicos. En el lado epistémico, los dos enfoques más prominentes — el método de permutación de Uzzi y la novedad temporizada de Wang — han fallado ambas validación externa, no mostrando concordancia con novedad evaluada por expertos (Bornmann, 2019; Fontana et al., 2020) y correlacionando en cambio con constructos de diversidad que las medidas fueron diseñadas para distinguir. La ausencia de medidas de novedad de la revisión de 23 indicadores de Wang y Schneider (2020) atestigua además su integración limitada en práctica estándar. La novedad sigue siendo una dimensión conceptual importante de interdisciplinariedad, y trabajo futuro en operacionalizaciones validadas puede justificar su inclusión; por ahora, el panel se enfoca en dimensiones — diversidad, coherencia y difusión — para las cuales las herramientas de medición tienen fundamento empírico más fuerte.

Cobertura Matemática y Mapa de Calificación

Para hacer explícito el alcance matemático de los indicadores revisados, la tabla abajo mapea las principales referencias portadoras de fórmulas usadas en esta revisión a sus condiciones de calificación. Este es un dispositivo de higiene de afirmaciones: cada familia de fórmulas está atada a al menos una condición de frontera que restringe interpretación.

Familia de fórmula	Referencias primarias	Condición de calificación usada en esta revisión
Rao-Stirling / integración ($\Delta = \sum d_{ij} p_i p_j$)	Porter y Rafols (2009); Rafols y Meyer (2009); Leydesdorff y Rafols (2011)	Los valores no son invariantes a granularidad de taxonomía, construcción de matriz de similaridad, o elección de métrica de distancia.
Diversidad verdadera basada en similaridad (qD^S , $2D^S$)	Zhang, Rousseau y Glanzel (2016); Hill (1973); Jost (2006, 2009); Leinster y Cobbold (2012)	Las cantidades tipo entropía deben interpretarse en una escala de número efectivo; variantes basadas en similaridad y disparidad no son intercambiables numéricamente.

Familia de fórmula	Referencias primarias	Condición de calificación usada en esta revisión
Índices de variedad-balance (Shannon, Simpson, Herfindahl, Gini)	Porter y Rafols (2009); Leydesdorff, Wagner y Bornmann (2019); Mutz (2022)	Estos índices no pueden, por sí mismos, identificar distancia cognitiva; la elección de operador de agregación (aditivo vs multiplicativo) cambia clasificaciones.
Coherencia (S) vía acoplamiento bibliográfico	Rafols y Meyer (2009); Jensen y Lutkouskaya (2014)	Las estimaciones de coherencia dependen de umbralización/binarización de acoplamiento y convenciones de construcción de red.
Alternativas basadas en centralidad	Leydesdorff y Rafols (2011); Bollen et al. (2009)	La intermediación e indicadores de grafo relacionados mezclan interdisciplinariedad con efectos de tamaño/posición a menos que se normalicen cuidadosamente.
Difusión / efecto transdisciplinar (E)	Leydesdorff, Wagner y Bornmann (2019); Xiang, Romero y Teplitskiy (2025); Lariviere y Gingras (2010)	La adopción transdisciplinar debe normalizarse por campo; la difusión debería tratarse como distinta de diversidad de entrada.
Descomposición de flujo de conocimiento (B, I, H)	Zhou, Guns y Engels (2023)	Los vectores de flujo distribucional responden preguntas de intercambio direccional pero no son reemplazos directos para componentes de panel escalares.
Novedad vía combinaciones atípicas / emergencia temporizada	Uzzi et al. (2013); Wang et al. (2017); Bornmann (2019); Fontana et al. (2020)	Los indicadores de novedad capturan atipicidad bajo supuestos explícitos de modelo nulo; la validación externa permanece limitada para uso en política.
Estabilización de superposición casi-cero	Moulton y Jiang (2018)	Las variantes de Jaccard probabilístico mejoran comportamiento de superposición cero pero aún requieren modelado de disparidad complementario para afirmaciones de interdisciplinariedad.

Familia de fórmula	Referencias primarias	Condición de calificación usada en esta revisión
Cuantificación de incertidumbre y diagnósticos de validez	Zwanenburg, Nakhoda y Whigham (2022); Nakhoda, Whigham y Zwanenburg (2023)	Las estimaciones a nivel individual requieren reporte de intervalos; la umbralización de estimación puntual sola es frágil para decisiones.
Corpus de referencia de coherencia empírica	Noji et al. (1997); Tomishige et al. (2002)	Estos son casos de prueba de referencia para comportamiento de indicadores, no fórmulas normativas para calidad de interdisciplinariedad.
Encuadre de calidad transdisciplinaria	Stokols et al. (2003); Klein (2008); Borlaug y Svartefoss (2025)	Los juicios de calidad requieren criterios evaluativos explícitos más allá de valores de panel bibliométrico.

Este mapa no afirma que cada artículo citado contribuya una fórmula novedosa. Más bien, hace explícito cómo las matemáticas que *se* usan en la revisión son calificadas antes de traducirse en orientación de evaluación.

Más Allá de Escalares: Enfoques Basados en Distribución

Trabajo reciente ha desafiado el supuesto de que la interdisciplinariedad debería medirse mediante indicadores escalares en absoluto. Zhou, Guns y Engels (2023) proponen un marco de Flujo de Conocimiento Interdisciplinario (IKF) que caracteriza la relación entre dos disciplinas cualesquiera a lo largo de tres aspectos: *amplitud* (qué fracción de publicaciones cita una disciplina externa dada), *intensidad* (cuán profundamente comprometidas están esas publicaciones citantes), y *homogeneidad* (similaridad cognitiva vía superposición de co-citación). Formalmente, dada una matriz de citas M ($n \times n$) y entidades X (citante) e Y (citada), la amplitud es $B(X, Y) = |X'|/|X|$, donde X' es el subconjunto de publicaciones en X que citan al menos una publicación en Y . La intensidad restringe el denominador solo a citas salientes desde X' : $I(X, Y) = \sum_{i \in X, j \in Y} M_{ij} / \sum_{i \in X, j=1}^n (M_{ij} \delta_i)$, donde $\delta_i = 1$ si y solo si $i \in X'$. La homogeneidad mide superposición de base de conocimiento: $H(X, Y) = \sum_{i \in X, \gamma=1}^n M_{i\gamma} \varphi_{\gamma, Y} / \sum_{i \in X, j=1}^n M_{ij}$, donde $\varphi_{\gamma, Y} = 1$ si la publicación γ también es citada por Y . Cada aspecto es así una fracción bien definida, y la triple (B, I, H) caracteriza conjuntamente la *forma* del intercambio de conocimiento interdisciplinario — produciendo un vector de distribución en lugar de un único número y respondiendo “qué está interdisciplinado” en lugar de meramente “cuán interdisciplinario”.

Cantone (2024) adopta un enfoque sistémico complementario, descomponiendo el problema de medición en una tubería de elecciones analíticas: selección

de la unidad de análisis (artículo, autor, institución), elección de taxonomía disciplinar, método de clasificación, definición operacional (dimensión y fórmula), y estrategia de agregación. Este encuadre hace explícito que los valores de indicadores dependen de una cadena de decisiones metodológicas, cada una de las cuales introduce potencial inconsistencia.

Ambos enfoques sugieren que el campo se está alejando de resúmenes de único número hacia caracterizaciones más ricas, multidimensionales. Nuestro panel ocupa un terreno medio: es multidimensional (tres componentes) pero produce un perfil compacto, interpretable en lugar de una distribución de alta dimensión.

Resumen de Taxonomía

La siguiente tabla organiza indicadores principales por dimensión conceptual y familia metodológica:

Dimensión	Basada en referencias	Basada en citas	Basada en texto	Basada en redes
Diversidad	Rao-Stirling, Simpson, Shannon, Gini, Hill	—	Similaridad semántica	—
Coherencia	Densidad de acoplamiento bibliográfico	—	—	Centralidad de intermedio, agrupamiento
Difusión	—	Citas transdisciplinarias, diversidad citante	—	—
Novedad	—	—	—	Métricas de recombinación

Emergen cuatro observaciones. Primero, la celda basada en referencias/diversidad está densamente poblada mientras otras celdas permanecen dispersas — la literatura se ha enfocado desproporcionadamente en medir diversidad de insumos. Segundo, coherencia y difusión son mayormente ortogonales a diversidad (confirmado empíricamente por Wang y Schneider, 2020), pero reciben mucha menos atención. Tercero, métodos basados en texto y redes están subrepresentados relativo a su potencial. Cuarto, ningún indicador existente abarca múltiples dimensiones, motivando el enfoque multicomponente que desarrollamos a continuación.

Un Panel Multicomponente

Contra el telón de fondo del panorama de indicadores relevado arriba, presentamos ahora un panel específico de tres componentes diseñado para abarcar tres de las cuatro dimensiones identificadas: diversidad, coherencia y difusión.

Definición del Panel

Definimos tres indicadores que juntos caracterizan la interdisciplinariedad del portafolio de producción de un investigador. Cada uno captura una dimensión distinta de integración de conocimiento.

Diversidad de Rao-Stirling

Sea p_i la proporción de referencias en la categoría i entre las publicaciones de un investigador, y sea s_{ij} la similaridad de coseno entre categorías i y j (computada desde patrones de citas agregados). Defina la distancia $d_{ij} = 1 - s_{ij}$. El índice de diversidad de Rao-Stirling es

$$\Delta = \sum_{i,j} d_{ij} p_i p_j = 1 - \sum_{i,j} s_{ij} p_i p_j$$

Esta es la variante con $\alpha = \beta = 1$ de la heurística de diversidad generalizada de Stirling (2007). Se reduce al índice de diversidad de Simpson cuando todas las categorías son máximamente dispares ($s_{ij} = 0$ para $i \neq j$), y es igual a cero cuando todas las referencias caen en una única categoría. El índice captura variedad, balance y disparidad simultáneamente (Porter y Rafols, 2009).

Coherencia de Red: Fuerza Media de Enlace

Sea que un investigador tenga n publicaciones, y sea \mathbf{r}_k el vector de referencias de la publicación k sobre el conjunto de categorías. Definimos la fuerza media de enlace como

$$S = \frac{1}{\binom{n}{2}} \sum_{k < l} \cos(\mathbf{r}_k, \mathbf{r}_l)$$

donde $\cos(\cdot, \cdot)$ denota similaridad de coseno. Esta es una medida de acoplamiento bibliográfico: publicaciones que comparten muchas referencias en categorías similares tienen alta similaridad por pares. Un valor alto de S indica que las publicaciones del investigador forman un cuerpo de trabajo coherente; un valor bajo indica contribuciones desconectadas entre tópicos no relacionados.

El indicador de coherencia fue introducido conceptualmente por Rafols y Meyer (2009), quienes lo operacionalizaron como la densidad media de redes de acoplamiento bibliográfico. Proporciona una perspectiva de abajo hacia arriba que complementa la medida de diversidad de arriba hacia abajo.

Proxy de Efecto Transdisciplinar

Para cada una de las publicaciones de un investigador, sea k^* su categoría primaria (la categoría con la mayor participación de referencias). El proxy de efecto transdisciplinar es

$$E = \frac{\text{citas de artículos cuya categoría primaria } \neq k^*}{\text{total de citas recibidas}}$$

donde la suma se agrupa entre todas las publicaciones de un investigador. Un valor alto de E indica que el trabajo del investigador es usado entre fronteras disciplinarias — produce impacto transdisciplinar, no meramente insumos transdisciplinarios. Esto distingue integración genuina (alta Δ , S moderada, alta E) de amplitud polimática (alta Δ , baja S , baja E).

Discriminación en Datos de Juguete

Ilustramos el panel con un conjunto de datos universitarios de juguete consistiendo de cinco categorías temáticas y tres arquetipos de investigador.

Configuración

Las cinco categorías son: física de materia condensada (C_1), ciencia de materiales (C_2), química física (C_3), biología molecular (C_4), y matemáticas aplicadas (C_5). Sus similaridades por pares están dadas por la matriz

	C_1	C_2	C_3	C_4	C_5
C_1	1.00	0.60	0.40	0.10	0.30
C_2	0.60	1.00	0.50	0.15	0.20
C_3	0.40	0.50	1.00	0.35	0.10
C_4	0.10	0.15	0.35	1.00	0.05
C_5	0.30	0.20	0.10	0.05	1.00

Cada investigador tiene cinco publicaciones con vectores de referencias sobre estas categorías. Las proporciones de categorías agregadas p_i se derivan de estos vectores (no especificadas independientemente), asegurando consistencia interna entre los tres indicadores.

- **Investigador A** (integrador interdisciplinario): Cada publicación referencia múltiples categorías distantes (ej., materia condensada y biología molecular en un único artículo). Agregado: $p_A = (0.300, 0.075, 0.175, 0.350, 0.100)$.
- **Investigador B** (polímata): Cinco publicaciones de campo único, una por categoría. Agregado: $p_B = (0.200, 0.200, 0.200, 0.200, 0.200)$.
- **Investigador C** (especialista): Todas las publicaciones concentradas en física de materia condensada y ciencia de materiales. Agregado: $p_C = (0.636, 0.273, 0.061, 0.000, 0.030)$.

Ejemplo Computacional Trabajado

Para hacer el cómputo del panel completamente transparente y reproducible, trazamos el cálculo de los tres indicadores para cada investigador desde los vectores de referencias brutos a nivel de publicación. Los datos completos se presentan abajo; las proporciones p_i reportadas arriba se derivan de estos vectores, no se especifican independientemente.

Vectores de referencias de publicación. Cada entrada $r_{k,i}$ da el número de referencias desde la publicación k a la categoría C_i :

Publicación	C_1	C_2	C_3	C_4	C_5	Total
A1	3	0	2	3	0	8
A2	4	1	0	3	0	8
A3	2	2	0	3	1	8
A4	0	0	3	3	2	8
A5	3	0	2	2	1	8
A total	12	3	7	14	4	40
B1	5	0	0	0	0	5
B2	0	5	0	0	0	5
B3	0	0	5	0	0	5
B4	0	0	0	5	0	5
B5	0	0	0	0	5	5
B total	5	5	5	5	5	25
C1p	4	3	0	0	0	7
C2p	5	1	0	0	0	6
C3p	3	3	1	0	0	7
C4p	4	2	1	0	0	7
C5p	5	0	0	0	1	6
C total	21	9	2	0	1	33

Paso 1: Diversidad (Δ). El cómputo procede en tres etapas. Primero, derivar proporciones agregadas: para Investigador A, $p_1 = 12/40 = 0.300$, $p_2 = 3/40 = 0.075$, $p_3 = 7/40 = 0.175$, $p_4 = 14/40 = 0.350$, $p_5 = 4/40 = 0.100$. Segundo, computar el índice de concentración de Herfindahl $H = \sum_i p_i^2$:

$$H_A = 0.300^2 + 0.075^2 + 0.175^2 + 0.350^2 + 0.100^2 = 0.258750$$

Tercero, computar la suma ponderada por similaridad completa $\sum_{i,j} s_{ij} p_i p_j = H + 2 \sum_{i < j} s_{ij} p_i p_j$. Los términos cruzados dominantes para Investigador A son:

Par (i, j)	s_{ij}	p_i	p_j	$s_{ij} p_i p_j$
(C_1, C_4)	0.10	0.300	0.350	0.010500
(C_1, C_3)	0.40	0.300	0.175	0.021000

Par (i, j)	s_{ij}	p_i	p_j	$s_{ij} p_i p_j$
(C_3, C_4)	0.35	0.175	0.350	0.021438
(C_1, C_2)	0.60	0.300	0.075	0.013500
(C_1, C_5)	0.30	0.300	0.100	0.009000

Sumando todos los diez pares fuera de diagonal y aplicando el factor de simetría produce $\sum_{i,j} s_{ij} p_i p_j = 0.440625$, de donde $\Delta_A = 1 - 0.440625 = 0.559375$.

Para Investigador B, la distribución uniforme $p_i = 0.200$ da $H_B = 0.200$ y $\sum_{i,j} s_{ij} p_i p_j = 0.420$, así que $\Delta_B = 0.580$. Para Investigador C, la distribución concentrada da $H_C = 0.4839$ y $\Delta_C = 0.245$.

Paso 2: Coherencia (S). Computamos todas las $\binom{5}{2} = 10$ similaridades de coseno por pares entre vectores de referencias de publicación. Para Investigador A, valores representativos incluyen $\cos(\mathbf{r}_{A1}, \mathbf{r}_{A2}) = 0.878$ (A1 y A2 ambos referencian C_1 y C_4 , compartiendo un puente física–biología) y $\cos(\mathbf{r}_{A2}, \mathbf{r}_{A4}) = 0.376$ (el par más débil, ya que A4 evita C_1 enteramente). La media sobre los diez pares es $S_A = 7.334/10 = 0.733$.

Para Investigador B, cada vector de publicación es ortogonal a cada otro (cada uno referencia exactamente una categoría, y ningún par comparte una categoría), así que $\cos(\mathbf{r}_{Bk}, \mathbf{r}_{Bl}) = 0$ para todo $k \neq l$, dando $S_B = 0.000$. Esta coherencia cero es diagnóstica de polimatía: amplitud máxima sin integración entre publicaciones.

Para Investigador C, todas las publicaciones se agrupan en el vecindario C_1-C_2 , produciendo cosenos por pares uniformemente altos (rango 0.675 a 0.964) y $S_C = 0.881$.

Paso 3: Efecto transdisciplinario (E). Para cada publicación, asignamos la categoría primaria como $k^* = \arg \max_i r_{k,i}$ (empates rotos por índice más bajo). El desglose de citas es:

Publicación	Primaria	Citas totales	De primaria	De otras
A1	C_1	6	2	4
A2	C_1	5	2	3
A3	C_4	4	2	2
A4	C_3	5	2	3
A5	C_1	5	2	3
A total		25	10	15

Así $E_A = 15/25 = 0.600$: sesenta por ciento de las citas del Investigador A se originan fuera de la categoría primaria del artículo citante, confirmando impacto transdisciplinario genuino. Para Investigador B, casi todas las citas provienen de dentro del propio campo de cada publicación ($E_B = 1/16 = 0.063$); para Investigador C, una pequeña fracción de citas llega desde campos vecinos ($E_C = 4/19 = 0.211$).

Resultados

Los valores completos del panel son:

Investigador	Δ	S	E	Tipo
A (integrador)	0.559	0.733	0.600	Interdisciplinar
B (polímata)	0.580	0.000	0.063	Amplitud polimática
C (especialista)	0.245	0.881	0.211	Disciplinar

La observación crítica es que $\Delta_A \approx \Delta_B$ (0.559 versus 0.580): la diversidad sola no puede distinguir el integrador del polímata. Ambos investigadores se basan en un rango amplio de categorías, y el índice de Rao-Stirling correctamente reporta alta diversidad para ambos. La distinción yace en cómo esa diversidad está estructurada.

El indicador de coherencia S revela la diferencia: las publicaciones del Investigador A comparten referencias entre fronteras de categorías ($S = 0.733$), mientras que las publicaciones del Investigador B no tienen superposición alguna ($S = 0$). El efecto transdisciplinar E confirma esto al nivel de impacto: el trabajo del Investigador A es citado entre disciplinas ($E = 0.600$), mientras que las contribuciones de campo único del Investigador B son citadas casi exclusivamente dentro de sus propios campos ($E = 0.063$).

Ningún componente único del panel logra discriminación completa. La diversidad sola falla en A versus B. La coherencia sola falla en distinguir A (moderada-alta) de C (muy alta) sin el contexto de diversidad. El efecto transdisciplinar separa A de tanto B como C, pero no puede por sí mismo distinguir integradores de especialistas cuando la diversidad es desconocida. Solo la triple completa caracteriza únicamente cada tipo.

Análisis de Sensibilidad y Robustez

Una preocupación natural es si la discriminación del panel depende de los valores precisos de la matriz de similaridad intercategorías s_{ij} . Porque Δ es el único componente del panel que usa s_{ij} , el análisis de robustez se centra en dos preguntas: (i) cómo responde Δ a perturbaciones de la matriz de similaridad, y (ii) ¿están afectados los otros componentes en absoluto? Abordamos ambas analíticamente y mediante experimentos numéricos.

Resultado analítico: fórmula de perturbación uniforme

Proposición. Bajo una perturbación aditiva uniforme $s_{ij} \rightarrow s_{ij} + \varepsilon$ para todo $i \neq j$ (con entradas diagonales sin cambio):

$$\Delta_{\text{nuevo}} = \Delta_{\text{viejo}} - \varepsilon (1 - H)$$

donde $H = \sum_i p_i^2$ es el índice de concentración de Herfindahl.

Prueba. Escriba $\Delta = 1 - \sum_{i,j} s_{ij} p_i p_j$. Bajo la perturbación, la suma cambia por $\varepsilon \sum_{i \neq j} p_i p_j$. Dado que $\sum_{i,j} p_i p_j = (\sum_i p_i)^2 = 1$ y $\sum_i p_i^2 = H$, la suma fuera de diagonal es igual a $1 - H$, dando el resultado. \square

La fórmula hace la dependencia en concentración de investigador explícita: un portafolio más concentrado (H más alto) experimenta un cambio absoluto menor en Δ para la misma magnitud de perturbación, porque menos pares de categorías distintas contribuyen a la suma fuera de diagonal.

Corolario (evolución de brecha). La brecha con signo entre dos investigadores cualesquiera evoluciona linealmente:

$$\Delta_{A,\text{nuevo}} - \Delta_{B,\text{nuevo}} = (\Delta_A - \Delta_B) + \varepsilon (H_A - H_B)$$

Esta brecha se desvanece en la perturbación crítica

$$\varepsilon^* = -\frac{\Delta_A - \Delta_B}{H_A - H_B}$$

en cuyo punto la clasificación de diversidad de los dos investigadores se invierte.

Punto de inversión crítica

Para los datos de juguete, los índices de Herfindahl exactos son $H_A = 0.258750$ y $H_B = 0.200000$, dando $H_A - H_B = 0.058750$. La brecha de diversidad exacta es $\Delta_A - \Delta_B = 0.559375 - 0.580000 = -0.020625$. El punto de inversión crítica es por tanto

$$\varepsilon^* = \frac{0.020625}{0.058750} = 0.35106$$

Este valor requiere que *cada* similaridad fuera de diagonal en la matriz sea desplazada por más de 0.35 — una perturbación excediendo 35% de la escala de similaridad — antes de que el orden de diversidad de los Investigadores A y B se invierta. Dado que la incertidumbre realista en estimaciones de similaridad basadas en citas es mucho menor (Wang y Schneider, 2020, reportan correlaciones inter-variante de 0.30–0.91, correspondiendo a desplazamientos absolutos mucho menores en entradas individuales s_{ij}), la casi-igualdad $\Delta_A \approx \Delta_B$ es estructuralmente robusta en lugar de un artefacto de la matriz particular elegida.

Una nota técnica sobre cómputo: el valor exacto $\varepsilon^* = 0.35106$ debe derivarse de cantidades intermedias no redondeadas. Usar la brecha redondeada 0.021 y diferencia de Herfindahl redondeada 0.059 produce la aproximación 0.356, una discrepancia de 1.4% que, aunque pequeña, ilustra cómo el redondeo en etapas intermedias puede acumularse en cantidades derivadas.

Análisis de evolución de brecha

El corolario arriba implica que la brecha $\Delta_A - \Delta_B$ evoluciona como función lineal de ε con pendiente $H_A - H_B = 0.058750 > 0$. Tres regímenes son distinguibles:

- Para $\varepsilon < 0$ (las categorías se vuelven menos similares): la brecha se amplía a favor de B, pero la magnitud absoluta permanece pequeña.
- Para $0 \leq \varepsilon < \varepsilon^*$: $\Delta_A < \Delta_B$, con la brecha reduciéndose desde su valor original de 0.021 hacia cero.
- Para $\varepsilon > \varepsilon^*$: $\Delta_A > \Delta_B$, pero la brecha crece lentamente (pendiente 0.059 por unidad de ε).

A lo largo de este rango, la separación entre el par de alta diversidad $\{A, B\}$ y el especialista C evoluciona como

$$\overline{\Delta}_{AB} - \Delta_C = \overline{\Delta}_{AB,0} - \Delta_{C,0} - \varepsilon (\overline{1 - H}_{AB} - (1 - H_C))$$

donde $\overline{\Delta}_{AB,0} = 0.5697$ y $\overline{1 - H}_{AB} = 0.7706$, $(1 - H_C) = 0.5161$. El coeficiente sobre ε es -0.255 , significando que la separación A/B-C decrece lentamente a medida que las similaridades aumentan pero permanece por encima de 0.24 incluso en $\varepsilon = 0.35$. La capacidad del panel para separar especialistas de investigadores amplios se preserva a través de todas las magnitudes de perturbación realistas.

Experimentos de perturbación no uniforme

La perturbación uniforme es un peor caso en un sentido preciso: desplaza todas las similaridades en la misma dirección, maximizando el efecto acumulativo sobre Δ . La incertidumbre del mundo real en matrices de similaridad es más heterogénea. Por tanto probamos dos escenarios adicionales que modelan patrones realistas de incertidumbre de matriz.

Escenario 1: Vecinos más cercanos. Categorías adyacentes (aquellas con distancia de índice $|i - j| = 1$) se vuelven más similares por 0.10, mientras que categorías distantes ($|i - j| \geq 2$) se vuelven menos similares por 0.05. Esto modela una situación donde fronteras disciplinarias de grano fino se vuelven borrosas mientras la macro-estructura de conocimiento se preserva.

Escenario 2: Desplazamiento uniforme +0.10. Todas las similaridades fuera de diagonal aumentan por 0.10, correspondiendo a una base de datos de citas en la cual campos se han vuelto más interconectados (ej., mediante el auge de métodos impulsados por datos aplicados entre disciplinas).

Los resultados se resumen abajo:

Escenario	Δ_A	Δ_B	Δ_C	$ \Delta_A - \Delta_B $	$ \overline{\Delta}_{AB} - \Delta_C $
Original	0.559	0.580	0.245	0.021	0.32

Escenario	Δ_A	Δ_B	Δ_C	$ \Delta_A - \Delta_B $	$ \bar{\Delta}_{AB} - \Delta_C $
Vecinos más cercanos	0.557	0.572	0.214	0.015	0.35
Uniforme	0.485 +0.10	0.500	0.194	0.015	0.30

En ambos escenarios, tres propiedades se preservan: (i) la brecha A–B permanece pequeña (0.015, más estrecha que el original 0.021), confirmando que la diversidad sola no puede separar integradores de polímatas independientemente de especificación de matriz; (ii) la separación de C permanece grande (0.30 o más), asegurando identificación clara de especialistas; y (iii) el orden relativo $\Delta_C < \Delta_A < \Delta_B$ se mantiene. La perturbación no uniforme (Escenario 1) efectivamente *aumenta* la separación A/B–C porque portafolios de especialistas, concentrados en categorías vecinas, son más afectados por cambios de similaridad de vecinos que portafolios diversos.

Invarianza de coherencia y efecto transdisciplinar

Una ventaja distintiva del enfoque multicomponente es que solo uno de los tres indicadores del panel depende de la matriz de similaridad. El indicador de coherencia S se computa desde similaridades de coseno por pares entre vectores de referencias de publicación \mathbf{r}_k :

$$S = \frac{1}{\binom{n}{2}} \sum_{k < l} \frac{\mathbf{r}_k \cdot \mathbf{r}_l}{\|\mathbf{r}_k\| \|\mathbf{r}_l\|}$$

Esta cantidad depende exclusivamente de los vectores de referencias mismos, no de ninguna estructura de similaridad intercategorías. El efecto transdisciplinar E depende de flujos de citas y asignaciones de categoría primaria (determinadas por $\arg \max_i r_{k,i}$), que son igualmente independientes de s_{ij} . Tanto S como E son por tanto *exactamente invariantes* bajo cualquier perturbación de la matriz de similaridad, sea uniforme o no uniforme.

Esta invarianza tiene una consecuencia práctica: la discriminación entre Investigador A ($S = 0.733$, $E = 0.600$) e Investigador B ($S = 0.000$, $E = 0.063$) es completamente no afectada por la elección de matriz de similaridad. El panel multicomponente es así sustancialmente más robusto que cualquier enfoque de único indicador basado solo en diversidad, porque los canales de coherencia y efecto transdisciplinar no portan incertidumbre de matriz de similaridad alguna.

Resumen de robustez

El análisis de sensibilidad produce tres conclusiones. Primero, la fórmula analítica de perturbación $\Delta_{\text{nuevo}} = \Delta_{\text{viejo}} - \varepsilon(1 - H)$ hace desplazamientos de diversidad completamente predecibles: no hay efectos de umbral ni sorpresas no lineales bajo el punto de inversión ε^* . Segundo, la perturbación crítica requerida para invertir incluso la brecha de diversidad más pequeña en nuestros datos ($\varepsilon^* = 0.351$) excede por mucho la incertidumbre realista en estimación de similaridad. Tercero, y más importante, los indicadores de coherencia y efecto transdisciplinar son completamente inmunes a perturbación de matriz de similaridad, asegurando que el poder discriminatorio del panel se preserve incluso cuando el componente de diversidad está sujeto a incertidumbre de especificación.

Marco de Interpretación

La triple del panel (Δ, S, E) está diseñada no meramente como instrumento de medición sino como herramienta práctica de clasificación para contextos de evaluación. Esta subsección proporciona un marco sistemático para traducir perfiles de panel en decisiones de evaluación.

Taxonomía de patrones

Cuatro patrones recurrentes emergen de la observación conjunta de los tres componentes del panel:

Patrón 1: Alta Δ , alta S , alta E — integrador genuino. El investigador se basa en un rango amplio de categorías disciplinares (alta diversidad), las teje en un cuerpo de trabajo coherente con acoplamiento bibliográfico sustancial entre publicaciones (alta coherencia), y produce investigación que es citada entre fronteras disciplinares (alto efecto transdisciplinar). Este es el perfil canónico de integración interdisciplinar. En los datos de juguete, Investigador A ejemplifica este patrón con $(\Delta, S, E) = (0.559, 0.733, 0.600)$.

Patrón 2: Alta Δ , baja S , baja E — polímata. El investigador publica a través de muchos campos, produciendo alta diversidad categórica, pero las publicaciones son mutuamente incoherentes (acoplamiento bibliográfico bajo o cero) y cada una es citada primariamente dentro de su propio campo (bajo efecto transdisciplinar). Este perfil indica amplitud sin integración — una colección de contribuciones disciplinares independientes en lugar de un programa de investigación sintetizado. Investigador B ejemplifica este patrón con $(\Delta, S, E) = (0.580, 0.000, 0.063)$.

Patrón 3: Baja Δ , alta S , baja E — especialista. El investigador trabaja dentro de un clúster disciplinar estrecho, produciendo baja diversidad pero alta coherencia interna. El impacto transdisciplinar es limitado porque el trabajo se dirige a una audiencia especializada. Un investigador exhibiendo este perfil quien ha sido clasificado como “interdisciplinario” por una agencia de evaluación es probablemente mal clasificado y debería redirigirse a evaluación

disciplinar estándar. Investigador C ejemplifica este patrón con $(\Delta, S, E) = (0.245, 0.881, 0.211)$.

Patrón 4: Baja Δ , baja S , cualquier E — emergente o datos insuficientes. Cuando tanto diversidad como coherencia son bajas, el panel señala o bien un investigador de carrera temprana cuyo registro de publicación es demasiado disperso para estimación estable, o un investigador en un campo emergente donde las categorías disciplinarias no se han estabilizado aún. En cualquier caso, el perfil cuantitativo debería interpretarse con precaución, y evaluación experta cualitativa puede ser más apropiada.

Umbrales de clasificación

Para despliegue operacional, proponemos umbrales de decisión ilustrativos:

Clasificación	Δ	S	E
Integrador genuino	≥ 0.40	≥ 0.30	≥ 0.30
Polímata (no integrativo)	≥ 0.40	< 0.15	< 0.15
Especialista (reclasificar)	< 0.35	cualquiera	cualquiera
Ambiguo (revisión experta)	resto	resto	resto

Estos umbrales se derivan del análisis de datos de juguete y deberían entenderse como puntos de partida en lugar de límites universales. Las agencias de evaluación deben calibrar umbrales a su contexto específico, tomando en cuenta la granularidad del sistema de clasificación disciplinaria, las normas de citación de los campos bajo revisión, y los objetivos de política del ejercicio de evaluación. La validación en casos conocidos — investigadores cuyo estatus interdisciplinario ha sido establecido mediante revisión por pares o consenso experto — es esencial antes del despliegue operacional.

Casos borde y modos de fallo

Seis modos de fallo han sido identificados que pueden comprometer clasificación basada en panel, junto con sus mitigaciones recomendadas:

1. *Recompensa de amplitud-sin-profundidad:* Alta Δ recompensada independientemente de evidencia de integración. Mitigación: requerir que S y E excedan umbrales mínimos antes de clasificar como “integrador”. Esta es precisamente la discriminación que el panel está diseñado para proporcionar.
2. *Penalidad de publicación no estándar:* Las revistas interdisciplinarias frecuentemente tienen factores de impacto más bajos. Mitigación: usar indicadores de citación normalizados por campo; no comparar factores de impacto entre fronteras disciplinarias.

3. *Normas de citación incommensurables*: Las tasas de citación difieren por un orden de magnitud entre campos (ej., matemáticas versus biología molecular). Mitigación: normalizar E por líneas base de citación específicas de campo antes de comparación intercampo.
4. *Persistencia de mala clasificación*: Un especialista entra en una pista de evaluación “Interdisciplinaria” y permanece ahí indefinidamente. Mitigación: disparador automático de reclasificación cuando $\Delta < 0.35$.
5. *Dispersión de datos de carrera temprana*: Investigadores con menos de aproximadamente 15 publicaciones producen estimaciones de panel inestables. Mitigación: imponer un umbral mínimo de publicaciones, o reportar intervalos de confianza siguiendo la metodología de bootstrapping de Nakhoda, Whigham y Zwanenburg (2023).
6. *Gaming vía coautoría estratégica*: Añadir coautores desde campos distantes puede inflar Δ sin integración genuina. Mitigación: restringir cómputo de Δ a publicaciones de autor correspondiente; verificar contra S , que permanecerá baja si las publicaciones coautoradas son incoherentes.

El enfoque de panel no reclama aplicabilidad universal. Evaluaciones de publicación única, campos altamente colaborativos donde asignación de categoría primaria es ambigua, y campos emergentes cuyas fronteras disciplinarias no se reflejan aún en matrices de similaridad existentes todos representan situaciones donde el panel cuantitativo debería suplementarse o reemplazarse por juicio experto. Rafols (2019) ha argumentado que los indicadores deberían contextualizarse y sujetos a validación de partes interesadas; el marco presentado aquí está diseñado en ese espíritu, proporcionando entrada cuantitativa estructurada a — no un sustituto de — evaluación informada.

Medición en la Práctica

La taxonomía y panel presentados arriba tienen implicaciones para cómo se evalúa la interdisciplinariedad en la práctica. Esta sección proporciona orientación operacional para implementar el panel de indicadores, basándose en la tubería sistemática de cinco pasos propuesta por Cantone (2024): (1) selección de unidad de análisis, (2) elección de taxonomía, (3) método de clasificación, (4) definición operacional, y (5) estrategia de agregación. Organizamos la discusión en torno a cinco etapas procedurales que un practicante debe navegar: extracción de datos, mapeo de categorías, construcción de matriz de similaridad, control de calidad, e implementación de software.

Autoevaluación Institucional

Las universidades con acceso completo a datos internos — incluyendo registros de proyectos, bases de datos de publicaciones, composición de personal y presupuestos de investigación — están bien posicionadas para computar indicadores

de diversidad y coherencia directamente. El índice de diversidad de Rao-Stirling requiere solo una clasificación disciplinar de referencias citadas y una matriz de similaridad; el indicador de coherencia requiere datos de acoplamiento bibliográfico a nivel de publicación. Ambos son computables desde datos estándar de repositorio institucional. El efecto transdisciplinar, sin embargo, requiere datos de citas que las instituciones deben típicamente obtener de bases de datos externas (Web of Science, Scopus, o el de acceso abierto OpenAlex). Esta brecha de datos es el principal obstáculo práctico al despliegue de panel completamente interno. Donde datos de citas no están disponibles, un perfil de dos componentes (Δ, S) aún proporciona discriminación útil entre integradores y polímatas. El despliegue interno de tal panel podría apoyar autoevaluación estratégica sin depender de sistemas de clasificación externos, en el espíritu de métricas responsables defendido por Rafols (2019).

Agencias Nacionales de Evaluación

Un desafío distinto surge cuando una agencia nacional de evaluación debe evaluar un investigador cuya clasificación oficial es “Interdisciplinar” — es decir, un investigador que no encaja en ningún panel disciplinar único. Los procedimientos de evaluación estándar asignan revisores de una única disciplina, creando un desajuste estructural. El panel de indicadores puede apoyar evaluación más justa proporcionando evidencia objetiva del tipo y grado de cruce de fronteras: un perfil de alta Δ , alta S , alta E justifica revisores de múltiples campos, mientras que un perfil de alta Δ , baja S puede indicar un polímatha quien puede ser evaluado campo por campo. La composición del comité de evaluación debería reflejar la estructura revelada por el panel.

Un ejemplo concreto es el sistema español de evaluación de investigación nacional. En 2023, la Comisión Nacional Evaluadora de la Actividad Investigadora (CNEAI) creó *Campo 0: Interdisciplinar y Multidisciplinar*, la primera pista de evaluación dedicada para investigadores interdisciplinares dentro de la evaluación de productividad de seis años (*sexenio*). Mandatado por el Artículo 11.7 de Ley Orgánica 2/2023, que requiere valoración positiva de “los resultados de la investigación multidisciplinar e interdisciplinar” entre todos los campos, Campo 0 operacionaliza precisamente la distinción multi/inter discutida arriba. Sus criterios definen contribuciones *interdisciplinares* como aquellas “diseñadas o estructuradas mediante la aplicación de perspectivas, teorías o métodos asociados a diferentes disciplinas” — una definición orientada a entrada midiendo la integración de métodos diversos en diseño de investigación (mapeable a alta diversidad de referencias Δ combinada con alta coherencia S). Separadamente, trayectorias *multidisciplinares* son reconocidas cuando respaldadas por “al menos dos contribuciones en campos disciplinares diferentes” — una definición orientada a salida midiendo amplitud de publicación entre campos (mapeable a alta variedad $N \geq 2$). Notablemente, investigación *transdisciplinaria* está ausente de los criterios de Campo 0 en las tres ediciones publicadas hasta la fecha (2023–2025), consistente con la brecha de medición bibliométrica: transdisciplinariedad,

que involucra socios no académicos y trasciende epistemologías disciplinares, carece de indicadores bibliométricos estándar. El panel (Δ , S , E) podría operacionalizar la medición del lado de entrada que la pista interdisciplinar de ANECA requiere; la diversidad de campo de publicación la complementaría para el criterio multidisciplinar del lado de salida.

Procedimientos de Extracción de Datos

El primer paso práctico es extraer registros bibliográficos estructurados desde una base de datos de citas. Tres fuentes principales están en uso actual. Web of Science (WoS) ha sido la elección estándar para estudios de interdisciplinariedad, ofreciendo Categorías Temáticas de Revistas (aproximadamente 254 categorías en ediciones recientes) organizadas en el Science Citation Index y Social Sciences Citation Index. Wang y Schneider (2020) usaron el conjunto de datos combinado JCR 2016, que cubre 11.487 revistas. Scopus proporciona una alternativa con diferente estructura de categorías y cobertura más amplia en algunos campos, mientras que la plataforma de acceso abierto OpenAlex ofrece una taxonomía mantenida por la comunidad sin barrera de suscripción.

La elección de base de datos determina tanto los metadatos disponibles como la estructura taxonómica. Los campos de metadatos clave incluyen tipo de documento (artículos son preferidos; revisiones y editoriales son típicamente excluidos a menos que sean específicamente relevantes), año de publicación (instantáneas de año único o ventanas temporales definidas), y — más críticamente — listas de referencias. Las referencias representan la “base de conocimiento” sobre la cual se construye un trabajo y son preferidas sobre recuentos de citas para medir integración de conocimiento, porque reflejan elecciones intelectuales deliberadas por los autores en lugar de la recepción post-publicación del trabajo. Los enlaces de citas, por contraste, son más apropiados para medidas de difusión e impacto pero son dinámicos y menos estables con el tiempo.

Para enfoques de clasificación basados en semántica o colaboración, campos adicionales se vuelven relevantes: título, resumen y palabras clave apoyan modelado de tópicos y clasificación basada en IA (Cantone, 2024), mientras que afiliaciones de autores y datos de coautoría habilitan enfoques organizacionales. El volumen de datos puede ser sustancial — el conjunto de datos JCR 2016 solo contiene más de 3 millones de enlaces inter-revista y 50 millones de citas totales (Leydesdorff, Wagner y Bornmann, 2019), requiriendo atención a eficiencia computacional desde el inicio.

Protocolos de Mapeo de Categorías

Una vez que registros bibliográficos son extraídos, cada publicación debe asignarse a una o más categorías disciplinarias. Existen cuatro enfoques principales, cada uno con compensaciones distintas.

Asignación basada en revista es el método más común. Cada artículo hereda las

categorías temáticas de su revista de publicación. Para revistas multi-asignadas, los recuentos se dividen proporcionalmente o cada categoría recibe un recuento completo. La proporción p_i de categoría i en el perfil de referencias de un investigador es entonces $p_i(x) = c_i(x) / \sum_j c_j(x)$, donde $c_i(x)$ cuenta referencias en categoría i . Este enfoque es estable, explicable y computacionalmente barato, pero confunde disciplinariedad de revista con disciplinariedad de artículo — un artículo interdisciplinario publicado en una revista disciplinar hereda la clasificación estrecha de esa revista.

Clasificación cognitiva (basada en referencias) mapea cada referencia citada a las categorías temáticas de su revista, construyendo un perfil disciplinar desde la lista de referencias en lugar de desde la revista de publicación. Este es el enfoque más común para medir integración de conocimiento (Cantone, 2024). Una variante de segundo orden usa referencias de referencias (Rafols y Meyer, 2009), pero esto no resuelve la asimetría fundamental mediante la cual el artículo focal es tratado como potencialmente interdisciplinario mientras sus referencias son tratadas como monodisciplinarias en virtud de sus asignaciones de revista.

Clasificación semántica usa título, resumen y palabras clave como entrada a algoritmos de aprendizaje supervisado, modelos de tópicos, o modelos de lenguaje grandes. Cantone (2025) evaluó tres LLMs para clasificación disciplinar y encontró que Gemini 1.5 Pro más cercanamente aproximaba asignaciones tradicionales basadas en citas, ChatGPT 4o fue más resiliente a variaciones de denominación, y Claude 3.5 Sonnet ofreció un perfil balanceado. Las ventajas de clasificación semántica — intuitividad y aplicabilidad a artículos sin asignaciones de revista claras — se compensan por explicabilidad limitada y sensibilidad a diseño de prompts.

Clasificación basada en colaboración deriva disciplinariedad de artículo desde las identidades disciplinares de sus autores, usando antecedentes de grado, afiliación departamental, o trayectoria de carrera. Este enfoque enfrenta desafíos recursivos severos (clasificar un autor requiere clasificar su trabajo previo), baja señal con recuentos pequeños de autores, y preocupaciones éticas sobre reducir individuos a etiquetas disciplinarias. Es más útil en análisis institucionales donde metadatos a nivel de autor están disponibles y bien curados.

La elección entre estos enfoques no es neutral. Cantone (2024) observa que las medidas se agrupan por método de clasificación en lugar de por dimensión conceptual: medidas basadas en revista y referencias correlacionan entre sí más fuertemente que cualquiera correlaciona con medidas basadas en semántica de la misma dimensión. Los practicantes deberían seleccionar el enfoque que mejor corresponde a su pregunta de investigación y documentar la elección explícitamente.

Construcción de Matriz de Similaridad

Las medidas de diversidad que incorporan disparidad — incluyendo diversidad de Rao-Stirling y el indicador DIV — requieren una matriz de valores de

disimilaridad por pares d_{ij} entre categorías disciplinares. No todos los enlaces intercategorías representan grados iguales de cruce de fronteras: una referencia de física a matemáticas representa menos disparidad que una referencia de física a historia del arte. La construcción de esta matriz es por tanto una elección metodológica crítica.

El enfoque estándar computa similaridad de coseno desde vectores de citas intercategorías. Wang y Schneider (2020) distinguieron dos variantes. El coseno vectorial de Salton $SC(i, j) = \sum_k c_{ik} c_{jk} / \sqrt{\sum_k c_{ik}^2 \cdot \sum_k c_{jk}^2}$ usa el perfil de citas completo de cada categoría, donde c_{ik} representa citas desde categoría i a categoría k . El coseno binario de Ochiai $SO(i, j)$ usa una versión simetrizada de recuentos de citas cruzadas directos. Un hallazgo crítico es que la correlación entre las medidas de disimilaridad resultantes $1 - SC$ y $1 - SO$ es solo 0.54, y cae a 0.30 cuando se usa la transformación inversa $1/SC$ versus $1/SO$. La elección de fórmula de similaridad por tanto afecta dramáticamente estimaciones de diversidad.

La similaridad de coseno tiene varias propiedades deseables para esta aplicación: es no paramétrica, acotada en $[0, 1]$, invariante a escala absoluta (combinaciones lineales preservan valores de coseno), y naturalmente ignora entradas cero en vectores dispersos (Leydesdorff, Wagner y Bornmann, 2019). El procedimiento estándar es construir una matriz de citas $I \times I$ entre categorías y convertir similaridades a disimilaridades vía $d_{ij} = 1 - \cos(i, j)$.

Cuando datos de citas no están disponibles, existen enfoques alternativos. Cantone (2024) describe una normalización de matriz de confusión adecuada para esquemas de clasificación donde probabilidades de mala clasificación son conocidas. Más recientemente, Cantone (2025) ha explorado estimación de modelos de lenguaje grandes, en la que se solicita a un LLM proporcionar puntuaciones de similaridad para todos los pares de categorías. Este enfoque elimina la necesidad de acceso a base de datos de citas pero introduce nuevas preocupaciones: precisión (varianza entre consultas idénticas repetidas), acuerdo entre modelos (aún limitado), y robustez a variaciones triviales de denominación en etiquetas de categorías.

Una preocupación distribucional adicional afecta la interpretación de resultados. Wang y Schneider (2020) encontraron que valores de disimilaridad basados en Ochiai ($1 - SO$) están extremadamente sesgados a la izquierda, con la mayoría de valores concentrados entre 0.95 y 1.0. Bajo esta distribución, la diversidad de Rao-Stirling efectivamente se reduce al índice de Simpson porque todos los pares intercategorías reciben pesos de disparidad casi idénticos. La disimilaridad basada en Salton ($1 - SC$) produce una distribución más uniforme y así preserva el rol pretendido del componente de disparidad. Los practicantes deberían examinar la distribución de su medida de disimilaridad elegida y evitar formulaciones que produzcan ponderación degenerada.

Procedimientos de Control de Calidad

La multiplicidad de elecciones metodológicas válidas crea un problema de “grados de libertad del investigador” (Wang y Schneider, 2020): muchas combinaciones defendibles de base de datos, taxonomía, método de clasificación, fórmula de similaridad y nivel de agregación pueden producir resultados sustancialmente diferentes. Un protocolo riguroso de control de calidad debería abordar al menos cuatro niveles.

Integridad de datos. Antes de computar cualquier indicador, el analista debería validar asignaciones revista-categoría (verificando lógica de multi-asignación), evaluar completitud de referencias (referencias faltantes sesgan diversidad hacia abajo), definir la ventana de citas (retrospectiva de todo tiempo versus una ventana fija como cinco años), y documentar el tratamiento de autocitas.

Consistencia metodológica. Wang y Schneider (2020) demostraron que medidas que pretenden capturar la misma dimensión frecuentemente exhiben correlaciones sorprendentemente bajas. Entre sus 23 variantes de indicadores, las medidas se agruparon por enfoque metodológico (basado en superposición versus basado en disimilaridad) en lugar de por dimensión conceptual. Indicadores que “deberían” correlacionar altamente (por ejemplo, diferentes operacionalizaciones de diversidad) a veces mostraron correlaciones bajo 0.3. Como paso mínimo de validación, cualquier nuevo indicador debería compararse contra alternativas establecidas en el mismo conjunto de datos, con correlaciones esperadas y reales reportadas.

Nivel de agregación. Una distinción crítica separa medición elemental (a nivel de artículo) y colectiva (a nivel de portafolio). En el enfoque elemental, cada artículo recibe su propia puntuación de indicador y la puntuación del investigador es la media o mediana. En el enfoque colectivo, todas las referencias de todos los artículos se agrupan en una única distribución antes de computar el indicador. Wang y Schneider (2020) encontraron que valores de diversidad de Rao-Stirling elementales y colectivos correlacionan a 0.91 cuando se usa la misma medida de disimilaridad, pero la correlación cae a 0.18 cuando se aplican diferentes fórmulas de disimilaridad al mismo nivel de agregación — confirmando que la matriz de similaridad, no el nivel de agregación, es la fuente dominante de variación. La elección de nivel debería guiarse por la pregunta de investigación: la medición elemental captura el artículo típico, mientras que la medición colectiva caracteriza la base de conocimiento general.

Sensibilidad a granularidad. Taxonomías más finas (como 254 categorías temáticas WoS) producen interdisciplinariedad medida más alta que taxonomías más gruesas (como 40 categorías OCDE), porque más fronteras de categorías están disponibles para cruzar (Cantone, 2024). No hay granularidad universalmente “correcta”; el nivel apropiado depende del propósito del análisis. Como verificación de robustez, los practicantes deberían repetir el análisis en múltiples niveles de granularidad y reportar si las conclusiones son estables.

Una forma adicional de validación compara medidas de diversidad contra indi-

cadores estructurales independientes. Leydesdorff, Wagner y Bornmann (2019) proponen comparar diversidad de Rao-Stirling con centralidad de intermediación en la red de citas de revistas. Alta correlación proporcionaría evidencia convergente de que ambas capturan aspectos de posicionamiento interdisciplinar; divergencia indicaría que las medidas capturan fenómenos diferentes. Comparar indicadores del lado citante (integración de conocimiento) y lado citado (difusión de conocimiento) por separado puede clarificar además qué dimensión se está midiendo.

Cuantificación de incertidumbre. Incluso cuando las elecciones metodológicas se mantienen fijas, la naturaleza estocástica de listas de referencias introduce incertidumbre de medición. Nakhoda, Whigham y Zwanenburg (2023) propusieron un enfoque de bootstrap no paramétrico para cuantificar esta incertidumbre para el índice de Rao-Stirling. Su procedimiento toma las N asignaciones de categorías temáticas reconocidas de una publicación, las remuestrea con reemplazo para producir $B = 500$ réplicas bootstrap de tamaño N , computa el índice de Rao-Stirling para cada remuestra, y construye un intervalo de confianza al 95% corregido por sesgo desde la distribución resultante. A través de 42.660 publicaciones, el ancho mediano de intervalo de confianza fue aproximadamente 0.15, pero valores variaron desde cero (cuando todas las referencias cayeron en una única categoría) a más de 0.6. Artículos con menos de diez referencias categorizadas exhibieron intervalos particularmente anchos, indicando que estimaciones puntuales de interdisciplinariedad son poco confiables para listas de referencias cortas. Los autores mostraron además que combinar el intervalo de confianza bootstrap con el número de referencias produce un filtro de confiabilidad más efectivo que cualquier criterio solo, habilitando a practicantes marcar publicaciones cuyas puntuaciones de interdisciplinariedad no deberían interpretarse al valor nominal.

Implementación de Software

Varios recursos de software apoyan el cómputo de indicadores de interdisciplinariedad. Leydesdorff, Wagner y Bornmann (2019) proporcionan una rutina públicamente disponible que acepta matrices de citas en formato Pajek y computa diversidad de Rao-Stirling, el indicador DIV, coeficiente de Gini, índice de Simpson, entropía de Shannon, y componentes separados de disparidad y variedad. Wang y Schneider (2020) combinaron consultas SQL en una base de datos WoS interna con el paquete R *sna* para centralidad de intermediación y scripts R personalizados para otras medidas. En la actualidad, ningún paquete integrado único cubre el panel completo de indicadores propuesto en esta revisión; ensamblar uno desde componentes existentes es un siguiente paso natural.

La escalabilidad computacional requiere atención cuando se trabaja con conjuntos de datos grandes. La matriz de similaridad involucra $O(n^2)$ comparaciones por pares para n categorías — manejable para 254 categorías WoS (aproximadamente 32.000 pares) pero potencialmente costoso si se usan taxonomías de grano más fino. Representaciones de matrices dispersas son apropiadas porque la mayoría de recuentos de citas intercategorías son cero. La matriz de similaridad

debería precomputarse una vez y reutilizarse entre todos los artículos. Los cálculos de indicadores a nivel de artículo son independientes y así naturalmente paralelizables.

La reproducibilidad demanda que cada elección metodológica se documente: versión de base de datos (ej., JCR 2016 o JCR 2023), taxonomía y su granularidad, método de clasificación, fórmula de similaridad y transformación de disimilaridad, nivel de agregación, y cualesquiera criterios de filtrado aplicados. Análisis de sensibilidad — repitiendo el cómputo con elecciones alternativas en cada etapa — deberían acompañar los resultados principales (Cantone, 2024). Una estrategia de implementación práctica es comenzar con un año único, una taxonomía única, y clasificación basada en referencias, validar contra resultados publicados en datos comparables, y solo entonces introducir complejidad adicional. Diseño de tubería modular, con etapas separadas de extracción, clasificación y medición, facilita tanto validación incremental como la eventual sustitución de componentes a medida que los métodos mejoran.

Problemas Abiertos y Direcciones Futuras

Varias preguntas importantes permanecen sin resolver y merecen investigación adicional.

Primero, la relación entre interdisciplinariedad autoinformada y bibliométrica es poco entendida. Aksnes, Karlstrøm y Piro (2026), relevando más de 3.000 publicaciones entre todos los campos, encontraron que medidas de interdisciplinariedad autoinformada y bibliométrica “rara vez corresponden”. Probando entropía de Shannon, la medida de diversidad verdadera (2D_S), y la descomposición DIV* contra autoevaluaciones de investigadores, obtuvieron correlaciones variando de 0.13 a 0.18, explicando solo 2–3% de varianza. Los investigadores evalúan interdisciplinariedad basándose en dinámicas de colaboración e integración metodológica, no patrones de referencias. Esto plantea preguntas fundamentales sobre validez de constructo: si incluso una batería de indicadores bibliométricos complementarios falla en capturar lo que investigadores mismos quieren decir por interdisciplinariedad, la brecha debe reconocerse en cualquier marco de evaluación.

Segundo, la estimación de matrices de similaridad disciplinar — una entrada crítica a diversidad de Rao-Stirling y medidas relacionadas — ha dependido tradicionalmente de datos de acoplamiento de citas. Cantone (2025) ha explorado recientemente el uso de modelos de lenguaje grandes (ChatGPT 4o, Claude 3.5 Sonnet, Gemini 1.5 Pro) para estimar matrices de similaridad directamente desde etiquetas disciplinares, encontrando acuerdo parcial con estimaciones basadas en citas. Si se valida, este enfoque podría reducir los requerimientos de datos para computar indicadores de diversidad, aunque robustez a variaciones triviales de denominación permanece como preocupación.

Tercero, cuantificación de incertidumbre para medidas de interdisciplinariedad

está mayormente ausente de la literatura. Estimaciones puntuales de diversidad o coherencia se reportan sin intervalos de confianza, haciendo difícil evaluar si diferencias observadas entre investigadores o instituciones son estadísticamente significativas. Nakhoda, Whigham y Zwanenburg (2023) identificaron tres fuentes de incertidumbre en medidas basadas en citas — comportamiento arbitrario de referenciación, referencias no categorizadas, y herencia de categoría revista-artículo inválida — y propusieron un método de bootstrapping para estimar intervalos de confianza para el índice de Rao-Stirling. Su hallazgo de que intervalos de confianza pueden abarcar hasta 0.6 puntos subraya el riesgo de sobreinterpretar pequeñas diferencias en puntuaciones de diversidad.

Cuarto, la relación entre interdisciplinariedad y calidad de investigación permanece controvertida. Las medidas de calidad basadas en citas parecen penalizar trabajo interdisciplinar en el corto plazo: portafolios disciplinares balanceados están asociados con recuentos de citas más bajos sobre ventanas de evaluación típicas (Cantone, 2024, citando Larivière y Gingras, 2010). Sin embargo, evidencia reciente sugiere que IDR alcanza mayor y más duradero impacto sobre horizontes temporales más largos — la “penalidad” se entiende mejor como un rezago de difusión entre comunidades disciplinares. El tipo de interdisciplinariedad también importa: Xiang, Romero y Teplitskiy (2025), analizando 128.950 manuscritos entre 62 revistas, encontraron que interdisciplinariedad de base de conocimiento (referencias diversas) está asociada con tasas de aceptación más altas, mientras que interdisciplinariedad de tópico (cruzar fronteras temáticas disciplinares) está asociada con tasas de aceptación más bajas — las dos dimensiones tienen efectos opuestos sobre resultados de revisión por pares. Un panel que caracteriza el *tipo* de cruce de fronteras — como hace el nuestro — proporciona el contexto estructural necesario para interpretar indicadores de calidad correctamente. Importantemente, Xiang et al. también encontraron que revistas designadas como “interdisciplinares” por su editor no mostraron penalidad contra ninguna forma de interdisciplinariedad, sugiriendo que los sesgos observados son específicos a espacios disciplinares en lugar de inherentes al trabajo interdisciplinar mismo. Este hallazgo refuerza el caso para contextos de evaluación interdisciplinar dedicados. Notablemente, nuestro efecto transdisciplinar *E* se define como una fracción (no un recuento absoluto de citas), evitando la confusión de volumen de citas con interdisciplinariedad que afecta algunas medidas de difusión.

Quinto, los enfoques basados en distribución de Zhou et al. (2023) ofrecen una dirección prometedora para enriquecer paneles escalares. Su marco IKF descompone lo que Rao-Stirling agrega, revelando *qué* disciplinas contribuyen a diversidad medida y *cuán profundamente* cada una contribuye. Integrar tal información distribucional en marcos de evaluación prácticos es un desafío abierto.

Sexto, enfoques participativos e informados por diseño ofrecen una alternativa metodológica al despliegue de indicadores puramente cuantitativo. Marres y de Rijcke (2020) proponen “indicadores comprometidos” que reconocen el rol

dual de indicadores: no solo representar patrones de investigación sino también organizar comunidades de interpretación. Su metodología combina análisis cientométrico con talleres de partes interesadas y mapeo interactivo, enfatizando que los indicadores son entidades diseñadas cuyas formas materiales e interactivas incluyen o excluyen actores en procesos de evaluación. Este enfoque aborda la preocupación de Rafols (2019) de que los indicadores, originalmente desarrollados como herramientas para informar toma de decisiones, corren el riesgo de volverse “dispositivos productores de ignorancia” cuando se despliegan mecánicamente sin interpretación contextual. El desafío yace en escalar métodos participativos — que son intensivos en labor y específicos de contexto — a marcos de evaluación institucionales y nacionales mientras se preserva su capacidad de hacer aflorar significados contestados de interdisciplinariedad.

Séptimo, la medición de investigación *transdisciplinaria* permanece como una frontera abierta. La tipología tripartita de la OCDE y su elaboración por Wagner et al. (2011) distinguen transdisciplinariedad de multi- e interdisciplinariedad por su integración de epistemologías disciplinares y, crecientemente, por su compromiso con socios no académicos (Borlaug y Svarcefoss, 2025). Sin embargo ningún indicador bibliométrico estándar captura esta dimensión. Los indicadores de diversidad miden amplitud de insumos de conocimiento; los indicadores de coherencia miden integración de la base de conocimiento; pero ninguno detecta si la investigación trasciende fronteras académicas para comprometer conocimiento de practicantes, contextos de política, o partes interesadas comunitarias. Esta brecha de medición tiene consecuencias prácticas: la ANECA española, al diseñar Campo 0 para evaluación interdisciplinar, explícitamente cubre investigación “interdisciplinar y multidisciplinar” pero omite criterios transdisciplinares enteramente (BOE-A-2023-25537 hasta BOE-A-2025-26118) — un reconocimiento pragmático de que lo que no puede medirse no debería requerirse. Trabajo futuro podría explorar enfoques híbridos combinando paneles bibliométricos con evidencia cualitativa de compromiso de partes interesadas, a lo largo de las líneas de los indicadores participativos de Marres y de Rijcke (2020), para cerrar esta brecha.

Estudio de Caso: Evaluación a Nivel Departamental

La demostración con datos de juguete en la Sección 4 estableció que el panel de tres componentes puede distinguir arquetipos de investigador en principio. Ahora aplicamos el panel a un escenario más realista: evaluar la interdisciplinariedad de siete investigadores en un departamento universitario de física y ciencia de materiales. Este estudio de caso ilustra la operación práctica del panel a una escala que refleja ejercicios reales de evaluación institucional, usando un conjunto más rico de categorías disciplinares, volúmenes de publicación realistas, y datos institucionales suplementarios.

Contexto Departamental y Datos

Considere un departamento hipotético de Física y Ciencia de Materiales en una universidad de investigación de tamaño medio. El departamento alberga siete investigadores en varias etapas de carrera: tres profesores senior (doce a quince años post-doctorado), dos profesores de carrera media (ocho a nueve años), y dos investigadores de carrera temprana (cuatro a cinco años). La pregunta de evaluación es si cada investigador califica para una pista de financiamiento de investigación interdisciplinaria, una clasificación que conlleva consecuencias para asignación de revisores, composición de panel, y expectativas de reporte.

El paisaje disciplinar es capturado por seis categorías temáticas de Web of Science relevantes al portafolio de investigación del departamento:

Categoría	Etiqueta	Relación con departamento
C_1	Física, materia condensada	Núcleo
C_2	Ciencia de materiales	Núcleo
C_3	Química, física	Adyacente
C_4	Óptica	Adyacente
C_5	Ingeniería, eléctrica	Adyacente
C_6	Nanociencia y nanotecnología	Adyacente

Las similaridades por pares entre estas categorías, derivadas de similaridad de coseno en la red de citas de revistas WoS, son:

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	1.00	0.60	0.40	0.35	0.30	0.55
C_2	0.60	1.00	0.50	0.25	0.40	0.65
C_3	0.40	0.50	1.00	0.30	0.20	0.45
C_4	0.35	0.25	0.30	1.00	0.45	0.35
C_5	0.30	0.40	0.20	0.45	1.00	0.50
C_6	0.55	0.65	0.45	0.35	0.50	1.00

Esta matriz exhibe una estructura más rica que el ejemplo de juguete de la Sección 4: las categorías núcleo (C_1, C_2) tienen alta similaridad mutua (0.60) y conexiones moderadas a campos adyacentes; nanociencia (C_6) es intrínsecamente interdisciplinaria, con similaridad por encima del promedio a cinco de las seis categorías; y el par más distante es química–ingeniería ($s_{35} = 0.20$), reflejando genuina distancia epistemológica.

El portafolio de publicaciones de cada investigador se resume mediante un vector de distribución de referencias \mathbf{p} sobre las seis categorías, derivado de las listas de referencias agregadas de todas las publicaciones en el período de evaluación. La Tabla 1 presenta estos perfiles junto con recuentos de publicaciones, totales

de citas, y la fracción de citas recibidas desde fuera de la categoría primaria de cada investigador.

Tabla 1. Perfiles de investigadores. Abreviaturas de etapa de carrera: S = senior, M = carrera media, E = carrera temprana. El vector de referencias $\mathbf{p} = (p_1, \dots, p_6)$ da la proporción de referencias a cada categoría.

Investigador	Etapa	Pubs	\mathbf{p}	Citas	Citas transdisciplinarias
Chen	S	42	(0.35, 0.30, 0.20, 0.05, 0.05, 0.05)	520	180 (35%)
Al-Rahman	M	35	(0.20, 0.20, 0.20, 0.20, 0.15, 0.05)	280	30 (11%)
Kowalski	E	18	(0.65, 0.25, 0.06, 0.02, 0.01, 0.01)	85	10 (12%)
Nguyen	S	55	(0.25, 0.25, 0.15, 0.10, 0.15, 0.10)	890	420 (47%)
Romero	M	28	(0.10, 0.15, 0.05, 0.05, 0.10, 0.55)	195	140 (72%)
Karlsson	E	12	(0.50, 0.30, 0.10, 0.05, 0.03, 0.02)	45	8 (18%)
Osei	S	48	(0.70, 0.20, 0.05, 0.03, 0.01, 0.01)	680	75 (11%)

Varias características de este conjunto de datos merecen comentario. La distribución de referencias de Al-Rahman es casi uniforme entre las primeras cuatro categorías, asemejándose al arquetipo polimático de la Sección 4. La distribución de Romero está dominada por una única categoría adyacente (C_6 , nanociencia), sin embargo su fracción de citas transdisciplinarias es la más alta en el departamento. Estos perfiles contrastantes presagian las discriminaciones que el panel revelará.

Cómputo del Panel

Trazamos el cómputo de los tres componentes del panel para el conjunto completo de investigadores.

Diversidad (Δ). Aplicando la fórmula de Rao-Stirling $\Delta = \sum_{i,j} (1 - s_{ij}) p_i p_j$ a cada vector de referencias produce los valores en la Tabla 2. Ilustramos el cómputo para dos investigadores cuyos perfiles son de particular interés evaluativo.

Para Nguyen ($\mathbf{p}_N = (0.25, 0.25, 0.15, 0.10, 0.15, 0.10)$), la suma fuera de diagonal involucra quince pares distintos. Las contribuciones dominantes vienen de los pares de alto peso y alta distancia: el término (C_1, C_5) contribuye $2 \times 0.25 \times 0.15 \times 0.70 = 0.053$; el término (C_1, C_4) contribuye $2 \times 0.25 \times 0.10 \times 0.65 = 0.033$; y el término (C_3, C_5) contribuye $2 \times 0.15 \times 0.15 \times 0.80 = 0.036$. Sumando todos los quince pares da $\Delta_N = 0.464$.

Para Al-Rahman ($\mathbf{p}_F = (0.20, 0.20, 0.20, 0.20, 0.15, 0.05)$), la distribución casi uniforme genera muchos términos de magnitud comparable. El total evalúa a $\Delta_F = 0.610$, el más alto en el departamento — una consecuencia de peso disperso entre categorías con distancias mutuas sustanciales.

Coherencia (S). La fuerza media de acoplamiento bibliográfico se computa desde similaridades de coseno por pares entre vectores de referencias de publi-

cación. Este indicador captura si las publicaciones diversas de un investigador forman un todo integrado o representan contribuciones desconectadas.

Efecto transdisciplinario (E). La fracción de citas recibidas desde fuera de la categoría primaria de cada publicación se agrupa entre el portafolio completo del investigador.

Tabla 2. Valores de panel para los siete investigadores.

Investigador	Δ	S	E	Patrón
Chen	0.390	0.42	0.35	Integrador moderado
Al-Rahman	0.610	0.04	0.11	Polímata
Kowalski	0.190	0.53	0.12	Especialista
Nguyen	0.464	0.50	0.47	Integrador fuerte
Romero	0.374	0.50	0.72	Especialista puente de nicho
Karlsson	0.244	0.58	0.18	Especialista carrera temprana
Osei	0.176	0.58	0.11	Especialista

El rango de valores de diversidad (0.176 a 0.610) es más amplio que en el ejemplo de juguete, reflejando tanto variación disciplinaria genuina como el efecto de un conjunto de categorías más grande con distancias por pares heterogéneas. Los valores de coherencia se separan limpiamente en dos grupos: la coherencia casi cero de Al-Rahman ($S = 0.04$) está en marcado contraste con la coherencia moderada a alta de todos los otros investigadores ($S \geq 0.42$), indicando que sus publicaciones diversas esencialmente no comparten referencias.

Interpretación y Clasificación

Los valores de panel en la Tabla 2 apoyan una clasificación estructurada de cada investigador contra la pregunta de evaluación. Adoptamos umbrales de decisión indicativos: $\Delta \geq 0.40$ para diversidad sustancial, $S \geq 0.30$ para coherencia significativa, y $E \geq 0.30$ para impacto transdisciplinario significativo. Estos umbrales son ilustrativos; en la práctica, serían calibrados al contexto disciplinario local (ver Sección 5).

Nguyen ($\Delta = 0.464$, $S = 0.50$, $E = 0.47$): **integrador genuino**. Nguyen excede los tres umbrales cómodamente. Sus publicaciones abarcan cuatro categorías principales con peso sustancial, sin embargo comparten una base de referencias común ($S = 0.50$) que indica integración sistemática de conocimiento en lugar de incursiones desconectadas. Casi la mitad de sus citas (47%) se originan fuera de la categoría primaria, confirmando que su trabajo alcanza genuino impacto transdisciplinario. Datos institucionales refuerzan esta evaluación: 22 de 55 publicaciones involucran coautores de otros departamentos (química, ingeniería, escuela de medicina), y todos sus cuatro subsidios fueron otorgados mediante paneles de financiamiento interdisciplinarios. El panel recomienda

clasificar a Nguyen para la pista interdisciplinar, con un comité de revisión abarcando al menos tres de sus categorías activas.

Chen ($\Delta = 0.390$, $S = 0.42$, $E = 0.35$): integradora límite. Chen cae justo bajo el umbral de diversidad ($\Delta = 0.390$ versus el corte de 0.40) pero cumple los criterios de coherencia y efecto transdisciplinar. Su patrón de publicación — fuerte en materia condensada y ciencia de materiales, con compromiso sistemático en química física — sugiere una integradora emergente cuyo alcance interdisciplinar está concentrado entre campos cercanamente relacionados. La disparidad moderada entre sus categorías activas (la mayoría de similaridades por pares exceden 0.35) mantiene su valor de Rao-Stirling bajo el umbral, aunque su práctica de investigación es sustantivamente interdisciplinar. El panel recomienda clasificación como integradora emergente, con reevaluación periódica.

Romero ($\Delta = 0.374$, $S = 0.50$, $E = 0.72$): especialista puente de nicho. Romero presenta el caso más instructivo para interpretación de panel. Su diversidad está bajo el umbral, reflejando una posición concentrada en nanociencia ($p_6 = 0.55$). Sin embargo su efecto transdisciplinar es excepcional: 72% de sus citas provienen desde fuera de nanociencia, indicando que su trabajo especializado sirve como puente conectando nanociencia a materia condensada, ciencia de materiales e ingeniería. Su coherencia ($S = 0.50$) confirma que este rol de puente se sostiene mediante un programa de investigación integrado, no publicaciones transdisciplinares ocasionales. Datos de coautoría corroboran la interpretación: 18 de 28 publicaciones involucran colaboradores de otros departamentos. El panel correctamente identifica a Romero como un caso requiriendo revisión experta — no encaja en el perfil estándar de integrador, pero su rol estructural en la red de investigación departamental puede ser igualmente valioso para propósitos de financiamiento interdisciplinar.

Al-Rahman ($\Delta = 0.610$, $S = 0.04$, $E = 0.11$): polímata. Al-Rahman tiene la diversidad más alta en el departamento, con peso casi uniforme entre cuatro categorías. Sin embargo su coherencia es esencialmente cero ($S = 0.04$), indicando que sus publicaciones en diferentes campos no comparten base de referencias común — cada una constituye una contribución independiente a una conversación disciplinar separada. Su efecto transdisciplinar es correspondientemente bajo (11%): a pesar de publicar ampliamente, su trabajo es citado casi exclusivamente dentro de la categoría donde cada artículo fue publicado. Este es el patrón clásico de amplitud-sin-integración identificado por nuestro ejemplo de juguete. Notablemente, Al-Rahman tiene cero coautorías interdepartamentales, confirmando que su amplitud disciplinar no se traduce en integración colaborativa. El panel recomienda evaluación disciplinar estándar, no la pista interdisciplinar.

Kowalski, Karlsson y Osei: especialistas. Los tres investigadores restantes tienen valores de diversidad bajo 0.35, colocándolos claramente en la categoría de especialista. Osei ($\Delta = 0.176$) es el más concentrado, con 70% de referencias en física de materia condensada; su alta coherencia ($S = 0.58$) y bajo efecto transdisciplinar ($E = 0.11$) describen un investigador disciplinar enfocado y productivo. Kowalski y Karlsson son investigadores de carrera temprana cuya

baja diversidad refleja volumen de publicación limitado en lugar de un perfil disciplinar establecido. Para Karlsson (12 publicaciones), intervalos de confianza bootstrap producen $\Delta \in [0.18, 0.31]$ al nivel del 95%, sugiriendo que la evaluación de panel cuantitativa debería diferirse hasta que su portafolio alcance aproximadamente 20 publicaciones. Los tres son apropiadamente dirigidos a evaluación disciplinar estándar.

Un resultado notable son los datos institucionales sobre diversidad de coautoría. Para los siete investigadores, la fracción de publicaciones interdepartamentales se alinea con la clasificación del panel: Nguyen y Chen tienen colaboración interdepartamental sustancial (40% y 55% respectivamente), Romero tiene la tasa más alta (64%), mientras que Al-Rahman, Osei, Kowalski y Karlsson tienen cero artículos interdepartamentales. Esta convergencia entre indicadores bibliométricos y datos de proceso institucional incrementa confianza en las clasificaciones del panel.

Comparación con Enfoques de Único Indicador

El estudio de caso proporciona una demostración concreta de por qué enfoques de único indicador son inadecuados. Comparamos la clasificación del panel contra tres clasificaciones de único indicador.

Entropía de Shannon ($H = -\sum p_i \log_2 p_i$) aplicada a las distribuciones de referencias produce la siguiente clasificación: Al-Rahman ($H = 2.43$), Nguyen (2.32), Chen (2.05), Romero (1.85), Karlsson (1.45), Kowalski (1.22), Osei (1.08). Bajo esta medida, Al-Rahman — el polímata con cero integración — clasifica como el investigador más interdisciplinario en el departamento, delante de Nguyen, el integrador genuino.

Diversidad de Rao-Stirling sola produce una clasificación cualitativamente similar: Al-Rahman ($\Delta = 0.610$), Nguyen (0.464), Chen (0.390), Romero (0.374), Karlsson (0.244), Kowalski (0.190), Osei (0.176). De nuevo, Al-Rahman lidera. Ambos enfoques de solo-diversidad recompensan amplitud independientemente de si esa amplitud se acompaña de integración de conocimiento.

Razón de citas transdisciplinares sola (E) reordena la clasificación sustancialmente: Romero ($E = 0.72$), Nguyen (0.47), Chen (0.35), Karlsson (0.18), Kowalski (0.12), Al-Rahman (0.11), Osei (0.11). Esta medida correctamente degrada a Al-Rahman pero eleva a Romero — una especialista en nanociencia — a la posición superior, confundiendo impacto de especialista-puente con investigación genuinamente integrativa.

Cada indicador único produce un “investigador más interdisciplinario” diferente, y cada uno mal clasifica al menos un perfil. El panel completo evita estos errores porque opera en tres dimensiones ortogonales simultáneamente. El perfil polimático de Al-Rahman (Δ alta, S cerca de cero, E baja) se detecta inequívocamente; el rol de especialista-puente de Romero (Δ moderada, S moderada, E muy alta) se marca para revisión experta en lugar de clasificación

automática; y el estatus de integrador de Nguyen (Δ , S y E todas sobre umbral) se confirma con alta confianza. Esta discriminación triple es la ventaja práctica primaria del panel.

Limitaciones

Varias limitaciones del estudio de caso deberían reconocerse, ya que ilustran desafíos más amplios para despliegue de panel.

Calibración de umbrales. Los umbrales de decisión usados aquí ($\Delta \geq 0.40$, $S \geq 0.30$, $E \geq 0.30$) son ilustrativos, no empíricamente validados. Su calibración requiere benchmarking contra casos con estatus interdisciplinario conocido — por ejemplo, investigadores financiados mediante mecanismos interdisciplinares cuyo trabajo ha sido independientemente evaluado por paneles expertos. Hasta que tal benchmarking se realice, los umbrales deberían tratarse como parámetros ajustables que instituciones establecen según normas disciplinares locales.

Temporalidad de matriz de similaridad. La matriz de similaridad se deriva de la red de citas de revistas WoS 2016, mientras la evaluación cubre publicaciones hasta 2025. Las fronteras disciplinarias cambian con el tiempo: nanociencia (C_6), por ejemplo, puede haber sido más distinta de física de materia condensada (C_1) hace una década de lo que es hoy. Usar una matriz de similaridad estática introduce un sesgo sistemático que afecta particularmente a investigadores trabajando en las fronteras de campos en rápida convergencia. Una matriz actualizada, computada desde datos de citas contemporáneos con el período de evaluación, mitigaría esta preocupación.

Inestabilidad de carrera temprana. Para Karlsson (12 publicaciones) y Kowalski (18 publicaciones), los valores de panel se computan desde datos relativamente dispersos. Remuestreo bootstrap para Karlsson produce intervalos de confianza al 95% de $\Delta \in [0.18, 0.31]$, un rango que abarca la frontera entre especialista y diversidad moderada. Más generalmente, el indicador de coherencia S es sensible al tamaño de portafolio porque el número de comparaciones por pares crece cuadráticamente con el número de publicaciones. Para portafolios pequeños, una única publicación atípica puede alterar sustancialmente S . Una recomendación práctica es suplementar puntuaciones de panel con intervalos de confianza y diferir decisiones de clasificación para portafolios bajo aproximadamente 20 publicaciones.

Agregación temporal. El estudio de caso agrega la producción completa de carrera de cada investigador, enmascarando trayectorias potencialmente importantes. Las publicaciones tempranas de Kowalski están concentradas en materia condensada, pero su trabajo más reciente muestra expansión hacia ciencia de materiales y química física — una trayectoria que agregación a nivel de carrera oscurece. Una variante con ventana del panel (ej., computada sobre una ventana móvil de tres años) capturaría tales dinámicas, al costo de estabilidad estadística reducida para investigadores con tasas anuales de publicación más bajas.

Limitaciones de acoplamiento bibliográfico. El indicador de coherencia S mide integración mediante referencias compartidas. La coherencia casi cero de Al-Rahman ($S = 0.04$) puede subestimar conexiones metodológicas latentes entre sus publicaciones si esas conexiones operan mediante técnicas o conceptos compartidos en lugar de literatura compartida. Medidas basadas en texto — como análisis de co-palabras de resúmenes o similaridad de modelo de tópicos — podrían complementar acoplamiento bibliográfico en casos donde se sospecha integración metodológica pero no se refleja en superposición de referencias.

Conclusiones

La medición bibliométrica de investigación interdisciplinaria permanece como un problema no resuelto. Nuestra revisión del panorama de indicadores revela un campo fuertemente concentrado en medidas de diversidad — particularmente Rao-Stirling y sus variantes — mientras que dimensiones de coherencia, difusión y novedad reciben comparativamente poca atención. La evidencia empírica, especialmente el hallazgo de Wang y Schneider (2020) de baja consistencia entre 23 medidas y la demostración de Leydesdorff et al. (2019) de poder discriminatorio limitado, fuertemente sugiere que ningún indicador único es adecuado.

Este encuadre se alinea con las cuatro preguntas motivadoras planteadas para este proyecto. El panel está diseñado para separar evidencia de impacto/calidad de afirmaciones simples de amplitud (OQ1), permanecer computable con datos institucionales más dependencias externas claramente declaradas (OQ2), distinguir integración interdisciplinaria de acumulación polimática (OQ3), y apoyar protocolos de evaluación auditables a nivel de agencia mediante evidencia multidimensional explícita en lugar de clasificación de puntuación única (OQ4).

El panel de tres componentes que proponemos — diversidad (Δ), coherencia (S), y efecto transdisciplinario (E) — aborda esta inadecuación abarcando tres dimensiones ortogonales. Nuestra demostración con datos de juguete muestra que el panel caracteriza únicamente integradores, polímatas y especialistas donde cualquier componente único falla. El estudio de caso a nivel departamental de la Sección 7 confirma este poder discriminatorio a escala realista: el panel correctamente identifica a Al-Rahman como polímatas a pesar de tener la puntuación de diversidad más alta, marca el rol de especialista-puente de Romero para revisión experta, y confirma el estatus de integrador de Nguyen entre las tres dimensiones — clasificaciones que ningún indicador único logra. El resultado analítico de robustez — que la discriminación basada en diversidad se preserva bajo perturbaciones de hasta 35% de la matriz de similaridad — proporciona confianza de que el enfoque no es un artefacto de ajuste de parámetros.

Rafols (2019) ha argumentado persuasivamente que los indicadores de ciencia y tecnología deberían contextualizarse, ser multidimensionales, y sujetos a validación de partes interesadas. Nuestro panel está diseñado en este espíritu: presenta tres dimensiones separadas en lugar de colapsarlas en una puntuación

túnica, y su interpretación depende del contexto de evaluación. El despliegue práctico de tales paneles — ya sea para autoevaluación institucional o revisión de agencia nacional — requiere atención a las elecciones metodológicas relevadas en la Sección 5, los problemas abiertos identificados en la Sección 6, y las lecciones prácticas ilustradas en la Sección 7. La validación empírica en datos universitarios reales, construyendo sobre el estudio de caso ilustrativo presentado aquí, es el siguiente paso natural.

Referencias

- Abramo, G., D'Angelo, C. A., and Zhang, L. (2018). A comparison of two approaches for measuring interdisciplinary research output: The disciplinary diversity of authors vs the disciplinary diversity of the reference list. *Journal of Informetrics*, 12(4):1182–1193.
- Aksnes, D. W., Karlstrøm, H., and Piro, F. N. (2026). Self-reported and bibliometric interdisciplinarity measures rarely correspond: a survey-based comparative analysis of indicators and researcher perceptions. *Scientometrics*, 131:189–208.
- Bollen, J., Van de Sompel, H., Hagberg, A., and Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE*, 4(6):e6022.
- Borlaug, S. B. and Svarbefoss, S. M. (2025). Evaluating transdisciplinary research quality. In Sivertsen, G. and Langfeldt, L. (eds.), *Challenges in Research Policy*, pp. 13–20. Springer, Cham.
- Bornmann, L., Tekles, A., Zhang, H. H., and Ye, F. Y. (2019). Do we measure novelty when we analyze unusual combinations of cited references? A validation study of bibliometric novelty indicators based on F1000Prime data. *Journal of Informetrics*, 13(4):100979.
- Boyack, K. W., Klavans, R., and Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3):351–374.
- Cantone, G. G. (2024). How to measure interdisciplinary research? A systemic design for the model of measurement. *Scientometrics*, 129:4937–4982.
- Cantone, G. G. (2025). Estimation of disciplinary similarity with large language models. *Scientometrics*, 130(10):5345–5373.
- Choi, B. C. K. and Pak, A. W. P. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clinical and Investigative Medicine*, 29(6):351–364.
- Fontana, M., Iori, M., Montobbio, F., and Sinatra, R. (2020). New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research Policy*, 49(7):104063.
- Goyanes, M., Demeter, M., Grané, A., Albarrán-Lozano, I., and Gil de Zúñiga, H. (2020). A mathematical approach to assess research diversity: Operationalization and applicability in communication sciences, political

- science, and beyond. *Scientometrics*, 125(3):2299–2322.
- Hammarfelt, B. (2020). Discipline. *Knowledge Organization*, 47(3):244–256.
 - Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2):427–432.
 - Jensen, P. and Lutkouskaya, K. (2014). The many dimensions of laboratories' interdisciplinarity. *Scientometrics*, 98(1):619–631.
 - Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2):363–375.
 - Jost, L. (2009). Mismeasuring biological diversity: Response to Hoffmann and Hoffmann (2008). *Ecological Economics*, 68(4):925–928.
 - Klein, J. T. (2008). Evaluation of interdisciplinary and transdisciplinary research: A literature review. *American Journal of Preventive Medicine*, 35(2S):S116–S123.
 - Larivière, V. and Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology*, 61(1):126–131.
 - Leinster, T. and Cobbold, C. A. (2012). Measuring diversity: the importance of species similarity. *Ecology*, 93(3):477–489.
 - Leydesdorff, L. (2006). Can scientific journals be classified in terms of aggregated journal–journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science and Technology*, 57(5):601–613.
 - Leydesdorff, L. and Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics*, 5(1):87–100.
 - Leydesdorff, L., Wagner, C. S., and Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics*, 13(1):255–269.
 - Marres, N. and de Rijcke, S. (2020). From indicators to indicating interdisciplinarity: A participatory mapping methodology for research communities in-the-making. *Quantitative Science Studies*, 1(3):1041–1055.
 - Morillo, F., Bordons, M., and Gómez, I. (2001). An approach to interdisciplinarity through bibliometric indicators. *Scientometrics*, 51(1):203–222.
 - Morillo, F., Bordons, M., and Gómez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13):1237–1249.
 - Moulton, R. and Jiang, Y. (2018). Maximally consistent sampling and the Jaccard index of probability distributions. In *Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM)*, pages 347–356. IEEE.
 - Mutz, R. (2022). Diversity and interdisciplinarity: Should variety, balance and disparity be combined as a product or better as a sum? An information-theoretical and statistical estimation approach. *Scientometrics*, 127(12):7397–7414.

- Nakhoda, M., Whigham, P., and Zwanenburg, S. (2023). Quantifying and addressing uncertainty in the measurement of interdisciplinarity. *Scientometrics*, 128:6107–6127.
- National Academies. (2005). *Facilitating Interdisciplinary Research*. Washington, DC: National Academies Press.
- Noji, H., Yasuda, R., Yoshida, M., and Kinosita, K. (1997). Direct observation of the rotation of F1-ATPase. *Nature*, 386(6622):299–302.
- OECD (1998). *Interdisciplinarity in Science and Technology*. Paris: OECD Directorate for Science, Technology and Industry.
- Porter, A. L. and Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3):719–745.
- Rafols, I. (2019). S&T indicators in the wild: contextualization and participation for responsible metrics. *Research Evaluation*, 28(1):7–22.
- Rafols, I. and Meyer, M. (2009). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2):263–287.
- Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15):707–719.
- Stokols, D., Fuqua, J., Gress, J., Harvey, R., Phillips, K., Baezconde-Garbanati, L., Unger, J., Palmer, P., Clark, M. A., Colby, S. M., Morgan, G., and Trochim, W. (2003). Evaluating transdisciplinary science. *Nicotine & Tobacco Research*, 5(Suppl 1):S21–S39.
- Tomishige, M., Klopfenstein, D. R., and Vale, R. D. (2002). Conversion of Unc104/KIF1A kinesin into a processive motor after dimerization. *Science*, 297(5590):2263–2267.
- Uzzi, B., Mukherjee, S., Stringer, M., and Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157):468–472.
- Wang, J., Veugelers, R., and Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8):1416–1436.
- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I., and Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1):14–26.
- Wang, Q. and Schneider, J. W. (2020). Consistency and validity of interdisciplinarity measures. *Quantitative Science Studies*, 1(1):239–263.
- Xiang, S., Romero, D. M., and Teplitskiy, M. (2025). Evaluating interdisciplinary research: Disparate outcomes for topic and knowledge base. *Proceedings of the National Academy of Sciences*, 122(16):e2409752122.
- Zhang, L., Rousseau, R., and Glänzel, W. (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology*, 67(5):1257–1265.
- Zhou, Q., Guns, R., and Engels, T. C. E. (2023). Towards indicating inter-

disciplinarity: Characterizing interdisciplinary knowledge flow. *Journal of the Association for Information Science and Technology*, 74(11):1325–1340.