Agregación robusta en Descubrimiento Causal Federado bajo heterogeneidad de clientes

Nuria Rodríguez-Barroso[1] and M. Victoria Luzón[2]

[1] Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Granada, España. Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional (DaSCI)
rbnuria@ugr.es

[2] Departamento de Lenguajes y Sistemas Informáticos, Universidad de Granada, Granada, España. Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional (DaSCI)
luzon@ugr.es

Resumen El descubrimiento causal federado permite inferir relaciones estructurales entre variables a partir de datos distribuidos, evitando la centralización de información potencialmente sensible. Sin embargo, la heterogeneidad entre clientes plantea retos importantes, ya que diferencias en tamaño muestral, ruido o distribución pueden afectar tanto al aprendizaje local como a la posterior agregación de los grafos causales. En este trabajo se estudia el comportamiento de distintas estrategias simples de agregación de estructuras causales locales en escenarios federados heterogéneos. En particular, se comparan enfoques basados en voto mayoritario, ponderación por tamaño muestral y ponderación por acuerdo estructural entre clientes, utilizando grafos sintéticos y el algoritmo PC como método de descubrimiento local. Los resultados muestran que la agregación federada puede ofrecer una alternativa robusta frente al aprendizaje centralizado o estadísticos globales cuando los datos proceden de distribuciones heterogéneas. Asimismo, se observa que el tamaño muestral no siempre constituye un indicador fiable de calidad, especialmente cuando clientes dominantes presentan datos ruidosos. En conjunto, el estudio sugiere que reglas de agregación simples, interpretables y de bajo coste comunicacional pueden ser competitivas para el descubrimiento causal federado en entornos sensibles o distribuidos.

Keywords: Descubrimiento causal · Aprendizaje federado · Agregación robusta de grafos · Heterogeneidad de clientes

1. Introducción

El descubrimiento causal (DC) busca inferir relaciones estructurales entre variables a partir de datos, proporcionando una representación más interpretable que los modelos puramente predictivos [17,20,5]. Esta capacidad resulta especialmente relevante en dominios como la industria, la salud o la energía, donde comprender las dependencias entre procesos, sensores o variables operativas puede apoyar la toma de decisiones, la optimización y el diseño de intervenciones.

Sin embargo, en muchos escenarios reales los datos no se encuentran centralizados. Pueden estar distribuidos entre organizaciones, plantas industriales, dispositivos o infraestructuras distintas, y su intercambio puede estar limitado por razones de privacidad, confidencialidad, regulación [1,2] o competencia. En este contexto, el aprendizaje federado (AF) permite aprender de forma colaborativa sin compartir los datos originales de los clientes [15,9]. La combinación de ambos paradigmas da lugar al descubrimiento causal federado (DCF) [3], cuyo objetivo es estimar estructuras causales globales a partir de datos distribuidos.

El DCF introduce retos específicos. Los clientes pueden diferir en tamaño muestral, nivel de ruido, distribución de los datos o calidad de las observaciones. Esta heterogeneidad ha sido identificada como uno de sus retos centrales [11,7]. Además, centralizar todos los datos no siempre constituye una referencia óptima: cuando las muestras proceden de distribuciones heterogéneas o de distintos entornos, su unión puede alterar las dependencias estadísticas utilizadas por los algoritmos de DC [18].

En este trabajo se estudia una aproximación simple e interpretable al DCF basada en la agregación de grafos causales locales, comparada con FedPC [21], método del estado del arte que comparte estadísticos de independencia condicional. La hipótesis central es que, bajo heterogeneidad de calidad, la binarización local del DC actúa como filtro de ruido, haciendo que la agregación de grafos binarios sea más robusta que la agregación de estadísticos continuos, a pesar de compartir menos información.

Las principales contribuciones del trabajo son: (i) un estudio empírico de reglas simples de agregación para DCF bajo distintos tipos de heterogeneidad, incluyendo la comparación con FedPC [21]; (ii) la evidencia empírica de que la agregación de grafos binarios supera en robustez a la agregación de estadísticos continuos bajo heterogeneidad de calidad, y la interpretación de este resultado; y (iii) la evaluación de una estrategia basada en acuerdo estructural que no requiere datos adicionales ni incrementa el coste comunicacional respecto al voto uniforme.

El resto del artículo se organiza como sigue. La Sección 2 introduce los fundamentos. La Sección 3 describe las reglas de agregación estudiadas. La Sección 4 presenta el diseño experimental. La Sección 5 discute los resultados obtenidos. Finalmente, la Sección 6 recoge las conclusiones y líneas futuras.

2. Fundamentos y trabajo relacionado

Esta sección resume los conceptos necesarios para contextualizar el trabajo: DC basado en grafos, AF bajo heterogeneidad y DCF mediante agregación de estructuras locales.

2.1. Descubrimiento causal

El DC busca inferir relaciones causales entre variables a partir de datos. Una representación habitual son los grafos dirigidos acíclicos (DAGs), donde los nodos

representan variables aleatorias y las aristas codifican relaciones causales directas [17,20,19]. Bajo supuestos como Markov causal, fidelidad y suficiencia causal, las independencias condicionales observadas permiten restringir el conjunto de estructuras compatibles con los datos [5]. En general, no se identifica un único DAG, sino una clase de equivalencia representada mediante un Grafo Acíclico Dirigido Parcialmente Completado (CPDAG) [20]. Por ello, suele distinguirse entre la recuperación del esqueleto causal y la orientación de las aristas.

Los métodos de DC se clasifican en enfoques basados en restricciones, en puntuación y en optimización continua [5,23]. En este trabajo se utiliza PC, método basado en restricciones que aprende la estructura mediante tests de independencia condicional [20,10]. Para datos continuos gaussianos, PC emplea el test Fisher-Z, basado en correlaciones parciales.

2.2. Aprendizaje federado

El AF es un paradigma de aprendizaje distribuido que permite entrenar modelos de forma colaborativa sin compartir los datos de entrenamiento [9,22]. En su formulación clásica, un servidor coordina el proceso: distribuye el modelo global a los clientes, recibe las actualizaciones calculadas localmente y las agrega para obtener una nueva versión del modelo [15]. De este modo, la información compartida se limita a parámetros, gradientes u otras actualizaciones derivadas del entrenamiento, mientras que los datos permanecen en cada nodo.

Uno de los principales retos del AF es la presencia de distribuciones de datos no IID. Esto es, situaciones en las que los clientes presentan diferencias entre sus distribuciones locales, tamaños muestrales, niveles de ruido o calidad de los datos [12,14]. Esta heterogeneidad entre los datos puede afectar tanto al entrenamiento local como la agregación global, dado que las contribuciones de cada cliente pueden representar tanto información útil como sesgos propios de cada partición [13].

2.3. Descubrimiento causal federado

El DCF [3] combina DC y AF para inferir estructuras causales globales desde datos distribuidos sin centralizar las observaciones, lo que resulta relevante en entornos regulados como la sanidad o la industria. La literatura ha abordado el DCF mediante distintas estrategias: adaptar los tests de independencia condicional al entorno federado [8,21], formular el DC como optimización continua distribuida [16,4], o proponer estrategias de agregación ponderada de estimaciones locales [6].

Una alternativa más simple que consiste en que cada cliente aprenda localmente un grafo causal y el servidor agregue las estructuras binarias mediante una regla de votación o ponderación. Esta estrategia ofrece bajo coste comunicacional e interpretabilidad, y puede ser más robusta bajo heterogeneidad de calidad: la binarización local del test de independencia actúa como filtro, impidiendo que el ruido de medición propague sesgos al estadístico global. Su principal reto es decidir cómo combinar grafos de distinta calidad, tamaño o distribución.

3. Agregación federada de estructuras causales

Se considera un esquema de DCF en dos fases: (1) cada cliente aprende localmente una estructura causal a partir de sus propios datos; (2) el servidor combina las estructuras locales para obtener un grafo causal global.

Sea un escenario de AF con K clientes, donde el cliente k dispone de un conjunto de datos local D_k, con n_k muestras sobre un conjunto común de p variables. Tras aplicar un algoritmo de DC local, cada cliente obtiene un grafo G_k, que se representa mediante una matriz de adyacencia $A^{(k)} \in \{0,1\}^{p\times p}$, donde $A^{(k)}_{ij} = 1$ indica que el cliente k ha identificado una relación estructural entre las variables X_i y X_j. El servidor recibe las matrices locales y construye una matriz agregada A^*. De forma general, cada posible arista recibe una puntuación:

$$s_{ij} = \sum_{k=1}^{K} w_k A^{(k)}_{ij},$$

donde w_k es el peso asignado al cliente k. Una arista se incorpora al grafo global cuando $s_{ij} \geq \tau$; en los experimentos se fija $\tau = 0.5$.

Voto uniforme Asigna el mismo peso a todos los clientes:

$$w_k^{\text{uni}} = \frac{1}{K}.$$

La puntuación s_{ij} equivale a la proporción de clientes que han incluido la arista (i, j) en su grafo local. Su principal ventaja es la simplicidad, ya que no requiere información adicional sobre los clientes. Sin embargo, ignora diferencias de tamaño muestral o calidad de los datos.

Ponderación por tamaño muestral Pondera a cada cliente de acuerdo con su número de muestras:

$$w_k^{\text{size}} = \frac{n_k}{\sum_{\ell=1}^{K} n_\ell}.$$

Refleja la hipótesis de que más datos implican estructuras locales más fiables. Puede fallar cuando un cliente dominante obtiene datos ruidosos.

Ponderación por acuerdo estructural Asigna pesos según la consistencia de cada grafo local con el consenso de la federación. Primero se construye A^{ref} mediante voto uniforme con $\tau = 0.5$. Después, se calcula el F_1 de esqueleto entre el grafo del cliente k y la referencia:

$$a_k = \frac{2\,\text{TP}_k}{2\,\text{TP}_k + \text{FP}_k + \text{FN}_k}, \tag{1}$$

donde TP_k, FP_k y FN_k son las aristas del esqueleto correctamente recuperadas, incorrectamente incluidas y omitidas en $A^{(k)}$ respecto a A^{ref}. Los pesos se

obtienen normalizando: $w_k^{\text{agr}} = a_k / \sum_\ell a_\ell$. Frente a alternativas basadas en distancia de Hamming normalizada, el F_1 resulta más apropiado en grafos causales dispersos, pues normaliza por las aristas efectivamente en juego y mantiene una señal de calidad discriminativa independientemente de la densidad.

FedPC (estado del arte). Como referencia comparativa se incluye FedPC [21], método federado basado en restricciones en el que el servidor coordina la búsqueda del esqueleto PC mediante tests de independencia condicional distribuidos. En lugar de intercambiar grafos binarios, cada cliente k responde a las consultas del servidor con la transformación de Fisher-Z de su correlación parcial estimada, $z_k = \frac{1}{2} \ln \frac{1+r_k}{1-r_k}$. El servidor combina los estadísticos mediante media ponderada por grados de libertad:

$$Z_{\text{fed}} = \frac{\sum_k (n_k - |S| - 3)\, z_k}{\sqrt{\sum_k (n_k - |S| - 3)}}. \tag{2}$$

Bajo datos gaussianos IID, $Z_{\text{fed}} \sim \mathcal{N}(0,1)$ y es equivalente a PC centralizado sin transferencia de datos. El coste comunicacional es $O(n_{\text{tests}} \cdot K)$ valores reales, superior al de los métodos de agregación de grafos. La ponderación por $n_k - |S| - 3$ implica que los clientes con más datos dominan el estadístico combinado, con la misma vulnerabilidad estructural que la ponderación por tamaño.

4. Diseño experimental

El objetivo de la evaluación es analizar cómo distintas reglas de agregación afectan a la calidad del grafo causal global en escenarios de DCF con clientes heterogéneos. Para ello se generan datos sintéticos a partir de DAGs conocidos, lo que permite comparar las estructuras aprendidas con la estructura verdadera.

Generación de datos. Se consideran grafos causales sintéticos con $p \in \{5, 10, 20\}$ variables y $K \in \{3, 5, 10\}$ clientes. Para cada configuración se genera un DAG aleatorio (orden topológico aleatorio, $q = 0.3$, pesos en U(0.5; 1.0) con signo equiprobable; ≈3, ≈13 y ≈57 aristas esperadas para $p = 5$, 10 y 20) y se simulan datos con un SEM lineal gaussiano, $\mathbf{X} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{E}$, $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. En el caso equilibrado, cada cliente recibe $n_{\text{base}} = \text{máx}(300,\, 60p)$ muestras (300, 600 y 1 200 para $p = 5$, 10 y 20). En los escenarios desbalanceados, el total $N = K \cdot n_{\text{base}}$ se reparte con el 50 % al cliente dominante y el resto a partes iguales.

Para los clientes designados como ruidosos, se añade ruido de medición independiente $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0},\, \sigma_m^2\, \mathbf{I})$ sobre las observaciones generadas por el SEM, con $\sigma_m = 2.0$. A diferencia de escalar el ruido estructural del SEM (que preserva las correlaciones parciales entre variables), el ruido de medición rompe dichas correlaciones, degradando de forma efectiva la calidad de la inferencia causal local sin alterar la estructura causal verdadera. En el escenario **Sesgado**, el último tercio de los clientes recibe un desplazamiento de media de 3.0 unidades sobre el primer tercio de las variables, introduciendo una distribución marginal distinta sin modificar la estructura causal.

Se estudian seis escenarios de heterogeneidad: **IID** (distribuciones y tamaños similares), **Desbalanceado** (un cliente con el 50 % de los datos), **Ruidoso (1/3)** (un tercio de los clientes con $\sigma_m = 2.0$), **Ruidoso (1/2)** (la mitad con $\sigma_m = 2.0$), **Grande+ruidoso** (el cliente dominante en datos también es el ruidoso) y **Sesgado** (algunos clientes con desplazamiento de media en sus variables).

Estos escenarios permiten estudiar cuándo el tamaño muestral es un buen indicador de fiabilidad y cuándo puede resultar perjudicial para la agregación.

Métodos comparados. Se comparan seis enfoques: **Centralizado** (PC sobre todos los datos agrupados), **Local** (media del rendimiento de los grafos locales), **Voto uniforme**, **Ponderación por tamaño**, **Ponderación por acuerdo**, y **FedPC** [21] (ver Sección 3) .

En todos los clientes se emplea PC con $\alpha = 0.05$, test Fisher-Z y tamaño máximo del conjunto de condicionamiento igual a 3. Cada configuración se repite con 10 semillas aleatorias y se reportan las medias.

Métricas. La evaluación se centra en el esqueleto causal: SHD, precisión, recall y F_1 de esqueleto. Se incluye el F_1 de orientación como análisis complementario. El coste comunicacional se estima como el número de valores enviados por los clientes al servidor.

5. Análisis de resultados

En esta sección se analizan los resultados obtenidos para las distintas estrategias de agregación. Se estudia, en primer lugar, el comportamiento global de los métodos. A continuación, se analiza el efecto del tipo de heterogeneidad, del tamaño del grafo, del número de clientes y del coste comunicacional.

Análisis global. La Tabla 1 muestra los resultados medios sobre todos los escenarios, tamaños de grafo y números de clientes. Los métodos federados basados en agregación estructural superan al enfoque centralizado en la recuperación del esqueleto causal. En particular, el voto uniforme y la ponderación por acuerdo obtienen el mejor F_1 de esqueleto, con un valor de 0.639, frente al 0.564 del centralizado. La ponderación por acuerdo obtiene también el mejor SHD medio y el mejor F_1 de orientación. Además, FedPC alcanza 0.616, mejor que centralizado y local, pero significativamente inferior a los métodos de agregación de grafos.

Este resultado sugiere que, bajo heterogeneidad entre clientes, agrupar todos los datos o compartir estadísticos no siempre es la mejor opción para DC. La mezcla de distribuciones locales distintas puede alterar las independencias condicionales estimadas por PC y FedPC, mientras que el aprendizaje local seguido de agregación estructural puede actuar como un mecanismo de filtrado frente a errores locales.

Tabla 1. Resultados globales medios.

Método	SHD↓	skel-F_1↑	orient-F_1↑	Comm.↓	T(s)↓
Centralizado	23.33	0.564	0.371	62997	1.72
Local	22.06	0.597	0.401	**42**	**0.11**
Voto uniforme	21.39	**0.639**	0.439	**42**	0.71
Pond. tamaño	21.71	0.632	0.433	48	0.71
Pond. acuerdo	**21.37**	**0.639**	**0.440**	**42**	0.71
FedPC	22.01	0.616	0.432	14431	1.00

Análisis por escenario. La Tabla 2 resume el comportamiento por escenario. En el caso **IID**, FedPC ($F_1 = 0.650$) iguala al centralizado ($F_1 = 0.651$), confirmando la equivalencia teórica entre ambos enfoques. Los métodos de agregación de grafos obtienen $F_1 \approx 0.649$, siendo todos los enfoques comparables en ausencia de heterogeneidad.

Tabla 2. Resultados por escenario.

Escenario	Método	SHD↓	skel-F_1↑	orient-F_1↑
IID	Centralizado	21.29	**0.651**	0.460
	Local	21.41	0.632	0.440
	Voto uniforme	21.13	0.649	0.451
	Pond. tamaño	21.13	0.649	0.451
	Pond. acuerdo	**21.10**	0.649	0.453
	FedPC	21.27	0.650	**0.474**
Desbalanceado	Centralizado	20.87	0.661	0.473
	Local	21.66	0.616	0.409
	Voto uniforme	21.32	0.640	0.442
	Pond. tamaño	**20.77**	**0.666**	**0.476**
	Pond. acuerdo	21.36	0.639	0.442
	FedPC	20.90	0.661	0.472
Ruidoso (1/3)	Centralizado	24.79	0.489	0.276
	Local	22.26	0.588	0.387
	Voto uniforme	**21.40**	**0.639**	**0.437**
	Pond. tamaño	**21.40**	**0.639**	**0.437**
	Pond. acuerdo	21.43	**0.639**	0.436
	FedPC	21.89	0.629	0.436
Ruidoso (1/2)	Centralizado	25.03	0.512	0.289
	Local	22.71	0.561	0.366
	Voto uniforme	21.60	0.634	0.436
	Pond. tamaño	21.60	0.634	0.436
	Pond. acuerdo	**21.48**	**0.636**	**0.440**
	FedPC	22.40	0.609	0.422
Grande+ruidoso	Centralizado	24.87	0.510	0.335
	Local	22.35	0.582	0.391
	Voto uniforme	**21.49**	**0.632**	**0.430**
	Pond. tamaño	23.99	0.564	0.361
	Pond. acuerdo	21.51	**0.632**	0.429
	FedPC	22.99	0.568	0.380
Sesgado	Centralizado	23.14	0.563	0.390
	Local	21.99	0.604	0.414
	Voto uniforme	21.39	**0.641**	0.437
	Pond. tamaño	21.39	**0.641**	0.437
	Pond. acuerdo	**21.32**	**0.641**	**0.442**
	FedPC	22.61	0.578	0.406

Los escenarios **Grande+ruidoso** y **Sesgado** representan dos fuentes distintas de heterogeneidad que penalizan igualmente a FedPC y al enfoque centra-

lizado. En **Grande+ruidoso**, la ponderación por tamaño cae a $F_1 = 0.564$ y FedPC a $F_1 = 0.568$, mientras que el voto uniforme y la ponderación por acuerdo se mantienen en 0.632. El cliente dominante en datos es también el ruidoso, y tanto la ponderación por n_k (grafos) como la ponderación por $n_k - |S| - 3$ (FedPC) amplifican su influencia: el ruido de medición atenúa la correlación parcial hacia cero, sesgando z_k hacia la hipótesis nula y eliminando aristas verdaderas en el estadístico global. En **Sesgado**, FedPC cae de $F_1 = 0.650$ (IID) a 0.578, y el centralizado a 0.563, mientras que los métodos de grafos permanecen en ≈ 0.641. Aquí la fuente es la heterogeneidad marginal: los z_k de los clientes sesgados reflejan una distribución distinta a la de los demás; al combinarlos bajo el supuesto de homogeneidad, Z_{fed} agrega estadísticos incomparables, sesgando el test global. En ambos casos los métodos de agregación de grafos son robustos por el mismo motivo: cada cliente aplica PC sobre sus propios datos internamente consistentes y emite un voto binario válido, inmune a la heterogeneidad entre clientes.

En los escenarios **Ruidoso (1/3)** y **Ruidoso (1/2)**, FedPC ($F_1 = 0.629$ y 0.609) también queda por debajo de los métodos de agregación (≈ 0.639 y 0.634–0.636), mientras que el centralizado se deteriora aún más (0.489 y 0.512). Esto confirma que la binarización local actúa como filtro eficaz: si la correlación local supera el test de Fisher-Z individual, el cliente vota 1; el ruido moderado no altera esta decisión binaria, pero sí sesga el estadístico continuo.

En **Desbalanceado**, FedPC (0.661) alcanza prácticamente el nivel del centralizado (0.661) y la ponderación por tamaño (0.666), confirmando que cuando el cliente dominante tiene buena calidad, compartir estadísticos es tan informativo como centralizar.

Tabla 3. Resultados por número de variables.

p	**Método**	**SHD↓**	**skel-F_1↑**	**orient-F_1↑**
5	Centralizado	1.53	0.870	0.532
	Local	1.13	0.902	0.571
	Voto uniforme	**0.78**	**0.989**	0.645
	Pond. tamaño	0.90	0.967	0.627
	Pond. acuerdo	**0.78**	**0.989**	**0.646**
	FedPC	1.04	0.940	0.612
10	Centralizado	11.53	0.621	0.444
	Local	9.82	0.635	0.462
	Voto uniforme	8.81	0.681	0.503
	Pond. tamaño	9.29	0.678	0.499
	Pond. acuerdo	**8.78**	0.682	0.506
	FedPC	9.47	**0.685**	**0.525**
20	Centralizado	56.93	0.202	0.136
	Local	55.24	**0.255**	0.171
	Voto uniforme	54.57	0.247	0.169
	Pond. tamaño	54.95	0.252	**0.172**
	Pond. acuerdo	**54.54**	0.247	0.170
	FedPC	55.52	0.222	0.159

Efecto del tamaño del grafo. Para $p = 10$, FedPC obtiene el mayor skel-F_1 (0.685) y orient-F_1 (0.525), ligeramente por encima de la ponderación por acuer-

do (0.682, 0.506). En ausencia de clientes muy ruidosos, compartir estadísticos aporta señal adicional respecto a los grafos binarios. Para $p = 20$, todos los métodos se degradan notablemente: con ≈57 aristas esperadas y solo 1 200 muestras por cliente, PC con conjunto de condicionamiento máximo de tamaño 3 no puede recuperar de forma fiable las independencias condicionales de orden superior, lo que hace que el error local sea la limitación dominante independientemente de la estrategia de agregación. Para $p = 5$, los métodos de agregación de grafos obtienen un F_1 de esqueleto cercano a la perfección (0.989), mientras que FedPC se queda en 0.940.

Efecto del número de clientes. La Figura 1 muestra la evolución del SHD con el número de clientes. Los métodos federados de agregación reducen ligeramente el SHD al aumentar K, coherentemente con la idea de consenso. El centralizado mantiene SHD superior en la mayoría de configuraciones, especialmente bajo heterogeneidad. FedPC sigue una tendencia intermedia entre el centralizado y los métodos de agregación.

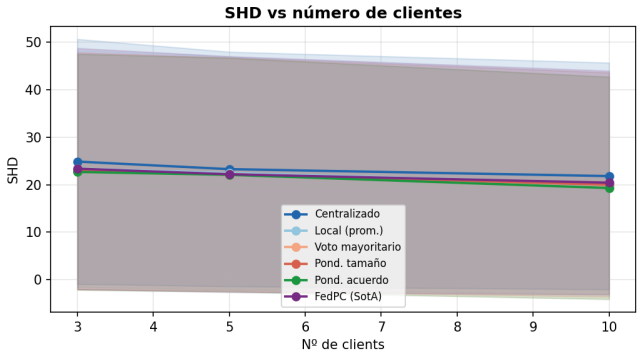


Figura 1. SHD según el número de clientes.

Coste comunicacional. La Figura 2 resume el coste comunicacional entre clientes y servidor. El centralizado requiere transferir todos los datos crudos (≈63000 valores por instancia). Los métodos de agregación de grafos comunican solo las aristas locales (≈42 valores). FedPC ocupa una posición intermedia (≈14400 valores), pues cada test de CI requiere un flotante por cliente; el número de tests crece con la complejidad del grafo. La ponderación por acuerdo mantiene el mismo coste que el voto uniforme: los pesos se calculan en el servidor a partir de los grafos ya recibidos, sin comunicación adicional.

Discusión general. Los resultados apuntan a tres conclusiones:

1. La agregación federada de grafos binarios es más robusta que el aprendizaje centralizado bajo heterogeneidad. La mezcla de datos locales de distin-

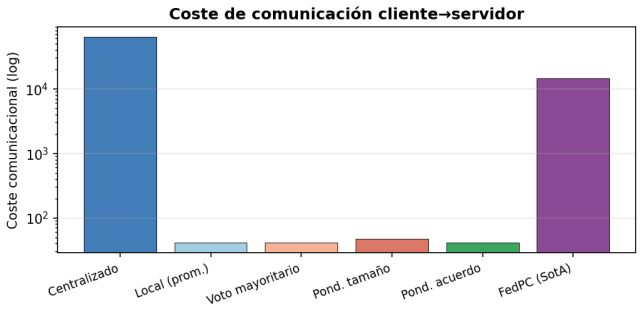


Figura 2. Coste comunicacional cliente–servidor en escala logarítmica.

ta calidad corrompe las independencias condicionales estimadas por PC; el aprendizaje local seguido de agregación actúa como mecanismo de filtrado.

2. La ponderación por tamaño muestral solo es adecuada cuando tamaño y calidad están correlados; en caso contrario amplifica la influencia de clientes poco fiables, y FedPC sufre la misma vulnerabilidad al ponderar los estadísticos por n_k.
3. La agregación de grafos binarios es más robusta que la agregación de estadísticos continuos (FedPC) bajo heterogeneidad de calidad. El mecanismo es la binarización local como filtro de ruido: cada cliente decide de forma binaria si existe dependencia, decisión que es robusta a ruido moderado siempre que la señal local supere el umbral del test. El ruido de medición no altera esa decisión binaria, pero sí sesga el estadístico continuo que FedPC agrega. La ponderación por acuerdo estructural ofrece el mejor compromiso: se comporta como el voto uniforme en escenarios estables y evita la degradación de la ponderación por tamaño y de FedPC en presencia de clientes dominantes ruidosos.

6. Conclusiones

En este trabajo se ha analizado la agregación federada de estructuras causales locales como aproximación simple e interpretable al DCF, comparando tres reglas de agregación de grafos con FedPC, método del estado del arte que comparte estadísticos de independencia condicional.

Los resultados muestran que la agregación de grafos binarios supera en robustez a FedPC bajo heterogeneidad de calidad entre clientes. El motivo está en que la binarización local del algoritmo PC actúa como filtro de ruido que impide que el ruido de medición propague sesgos al estadístico global. FedPC, en cambio, suma estadísticos ponderados por tamaño muestral, con la misma vulnerabilidad estructural que la ponderación por tamaño sobre grafos. En escenarios con clientes homogéneos o simplemente desbalanceados, FedPC es competitivo y puede superar a los métodos de agregación. La ponderación por acuerdo estructural

mantiene el mismo coste comunicacional que el voto uniforme e incorpora una noción básica de fiabilidad estructural sin acceder a los datos locales, ofreciendo el mejor equilibrio general.

Como trabajo futuro, se plantea extender la evaluación a datos reales, modelos no lineales y escenarios con variables parcialmente compartidas, así como estudiar mecanismos de agregación que combinen acuerdo estructural e incertidumbre local con garantías de privacidad más estrictas.

Acknowledgments. Esta investigación es parte del Proyecto TSI-100927-2023-1, financiado por Plan de Recuperación, Transformación y Resiliencia Europeo, de la Unión Europea-Next Generation EU a través del Ministerio para la Transformación Digital y de la Función Pública. Este trabajo también ha recibido apoyo del proyecto PIN-FLEX del Centro Mixto UGR-MADOC y de los Proyectos de Generación de Conocimiento PID2023-149128NB-I00 y PID2023-150070NB-I00, financiados por MICIU/AEI/10.13039/501100011033 y por FEDER, UE.

Disclosure of Interests. Los autores declaran no tener conflictos de intereses que sean relevantes para el contenido de este artículo.

Referencias

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (May 2016)
2. of European Union, C.: REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS (2021)
3. Gao, E., Ng, I., Chen, W., Zhang, K., Gong, M., Shen, Y., Song, Q., Cui, S.: Federated causal discovery. In: NeurIPS 2022 Workshop on Causality for Real-World Impact (2022)
4. Gao, E., Chen, J., Shen, L., Liu, T., Gong, M., Bondell, H.D.: FedDAG: Federated DAG structure learning. Transactions on Machine Learning Research (2023)
5. Glymour, C., Zhang, K., Spirtes, P.: Review of causal discovery methods based on graphical models. Frontiers in Genetics **10**, 524 (2019)
6. Guo, X., Yu, K., Liu, L., Li, J.: FedCSL: A scalable and accurate approach to federated causal structure learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 12235–12243 (2024)
7. Guo, X., Yu, K., Wang, H., Cui, L., Yu, H., Li, X.: Sample quality heterogeneity-aware federated causal discovery through adaptive variable space selection. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. pp. 4071–4078 (2024)
8. Huang, J., Guo, X., Yu, K., Cao, F., Liang, J.: Towards privacy-aware causal structure learning in federated setting. IEEE Transactions on Big Data **9**(6), 1525–1535 (2023)
9. Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al.: Advances and open

problems in federated learning. Foundations and Trends in Machine Learning **14**(1–2), 1–210 (2021)
10. Kalisch, M., Bühlmann, P.: Estimating high-dimensional directed acyclic graphs with the PC-algorithm. Journal of Machine Learning Research **8**, 613–636 (2007)
11. Li, L., Ng, I., Luo, G., Huang, B., Chen, G., Liu, T., Gu, B., Zhang, K.: Federated causal discovery from heterogeneous data. In: Proceedings of the International Conference on Learning Representations (2024)
12. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine **37**(3), 50–60 (2020)
13. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated Optimization in Heterogeneous Networks. In: Proceedings of Machine Learning and Systems. vol. 2, pp. 429–450 (2020)
14. Lu, Z., Pan, H., Dai, Y., Si, X., Zhang, Y.: Federated learning with non-iid data: A survey. IEEE Internet of Things Journal **11**(11), 19188–19209 (2024)
15. McMahan, H.B., Moore, E., Ramage, D., Hampson, S., Aguera y Arcas, B.: Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 54, pp. 1273–1282 (2017)
16. Ng, I., Zhang, K.: Towards federated bayesian network structure learning with continuous optimization. In: Proceedings of the 25th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 151, pp. 8095–8111 (2022)
17. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press, 2 edn. (2009)
18. Perry, R., von Kügelgen, J., Schölkopf, B.: Causal discovery in heterogeneous environments under the sparse mechanism shift hypothesis. In: Advances in Neural Information Processing Systems. vol. 35, pp. 10904–10917 (2022)
19. Peters, J., Janzing, D., Schölkopf, B.: Elements of Causal Inference: Foundations and Learning Algorithms. MIT Press (2017)
20. Spirtes, P., Glymour, C., Scheines, R.: Causation, Prediction, and Search. MIT Press, Cambridge, MA, 2 edn. (2000)
21. Wang, Z., Ma, P., Wang, S.: Towards practical federated causal structure learning. In: Machine Learning and Knowledge Discovery in Databases: Research Track. Lecture Notes in Computer Science, vol. 14170, pp. 359–375. Springer (2023)
22. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology **10**(2), 1–19 (2019)
23. Zheng, X., Aragam, B., Ravikumar, P.K., Xing, E.P.: DAGs with NO TEARS: Continuous optimization for structure learning. In: Advances in Neural Information Processing Systems. vol. 31 (2018)