



---

**Universidad de Valladolid**

**Facultad de Ciencias Económicas y  
Empresariales**

**Trabajo Fin de Grado**

**Grado en Administración y Dirección de Empresas**

**Análisis de las relaciones entre las  
dimensiones de la pobreza. Una  
aproximación basada en el Índice de  
Pobreza Multidimensional.**

Presentado por:

**Raúl García Hevia**

Tutelado por:

**Mercedes Prieto-Alaiz**

## **Resumen**

El Índice de Pobreza Multidimensional (MPI) es uno de los primeros intentos, en el contexto internacional, de medir la pobreza considerando más dimensiones que la monetaria. En concreto, este índice evalúa las privaciones en diez indicadores distribuidos en tres dimensiones esenciales: *Educación, Salud y Nivel de vida*. Sin embargo, el MPI no es suficientemente sensible a las interacciones entre las privaciones entre dichos indicadores. Este es un aspecto esencial cuando se aborda la pobreza como un fenómeno multidimensional. En este contexto, el presente trabajo se centra en analizar las relaciones entre las tasas de privación de los 10 indicadores del MPI de los 110 países para los que se calcula este índice. Con tal fin, se han generado tres grafos ponderados y no dirigidos utilizando tres medidas de dependencia basadas en rangos. Posteriormente, se ha construido el subgrafo de *Minimum Spanning Tree* para simplificar la representación de las relaciones bivariantes más fuertes entre las tasas de privación en los indicadores.

Los resultados destacan una elevada relación entre las diferentes dimensiones de la pobreza. Se observa también que, en los países con mayores niveles de pobreza, las privaciones en distintos indicadores tienden a estar más interrelacionadas que en aquellos países que presentan menor nivel de privaciones, haciendo la pobreza más persistente y difícil de superar en los casos más severos. En particular, la privación en el acceso a la electricidad se muestra como crítica en los países más pobres. Este estudio tiene implicaciones importantes para el desarrollo de políticas públicas efectivas, resaltando la necesidad de enfoques integrales que consideren la interdependencia de las distintas dimensiones de la pobreza.

## **Palabras clave**

Pobreza, Índice de Pobreza Multidimensional, Grafos, Minimum Spanning Tree, Dependencia

## **Abstract**

The Multidimensional Poverty Index (MPI) is one of the first international attempts to measure poverty by considering more dimensions than the monetary one. Specifically, this index evaluates deprivations in ten indicators distributed across three essential dimensions: *Education, Health, and Living Standards*. However, the MPI is not sufficiently sensitive to the interactions between the deprivations among these indicators, which is an essential aspect when addressing poverty as a multidimensional phenomenon. In this context, this study focuses on analyzing the relationships between the deprivation rates of the 10 MPI indicators in the 110 countries for which this index is calculated. To this end, three weighted and non-directed graphs were generated using three rank-based dependence measures. Subsequently, a Minimum Spanning Tree subgraph was constructed to simplify the representation of the strongest bivariate relationships between the deprivation rates in the indicators.

The results highlight a high level of relationship between the different dimensions of poverty. It is also observed that in countries with higher levels of poverty, deprivations in different indicators tend to be more interrelated than in those countries with lower levels of deprivation, making poverty more persistent and difficult to overcome in the most severe cases. In particular, deprivation in access to electricity is shown to be critical in the poorest countries. This study has important implications for the development of effective public policies, emphasizing the need for integrated approaches that consider the interdependence of the different dimensions of poverty.

## **Keywords**

Poverty, Multidimensional Poverty Index, Graphs, Minimum Spanning Tree, Dependence

## Índice:

1. INTRODUCCIÓN .....	6
2. EL INDICE DE POBREZA MULTIDIMENSIONAL.....	9
3. METODOLOGÍA.....	12
3.1. Grafos .....	12
3.2. Los pesos: medidas de dependencia basadas en rangos.....	16
3.3. Subgrafos.....	20
3.4. Métricas de la red .....	24
4. ANÁLISIS UNIVARIANTE DE LOS DATOS EMPLEADOS .....	27
5. UNA PRIMERA VISIÓN DE LAS RELACIONES ENTRE LAS VARIABLES .....	30
6. RESULTADOS .....	36
6.1. Red generada empleando el coeficiente de dependencia en la cola derecha .....	37
6.2. Red generada empleando el coeficiente de dependencia en la cola izquierda.....	39
6.3. Red generada empleando el coeficiente de correlación de Spearman .....	41
7. CONCLUSIONES .....	44
8. BIBLIOGRAFÍA.....	47
ANEXO 1: Código Python empleado para el tratamiento de datos .....	48
ANEXO 2: Software empleado .....	53
Anexo 2.1. Python .....	53
Anexo 2.2. Spyder .....	53
Anexo 2.3. Bibliotecas de Python.....	53

## Índice de figuras

Figura 3.1: Visualización del problema original de los puentes de la ciudad de Königsberg resuelto por Euler (Hart et al. 2016).....	13
Figura 3.2: Ejemplo de un grafo simple para definir las conexiones aéreas entre cuatro ciudades .....	14
Figura 3.3: Visualización de un bucle (en rojo) dentro de un grafo.....	22
Figura 3.4: Visualización de un grafo y su MST asociado.....	23
Figura 4.1: Diagrama de cajas y bigotes para los datos empleados.....	29
Figura 5.1: Gráficos de dispersión para los datos empleados.....	30
Figura 5.2: Gráficos de dispersión para los datos convertidos en rangos.....	32
Figura 6.1: Red MST para los coeficientes de dependencia de cola derecha.....	37
Figura 6.2: Red MST para los coeficientes de dependencia de cola izquierda.....	39
Figura 6.3: Red MST para los coeficientes de correlación de Spearman.....	42

## 1. INTRODUCCIÓN

A lo largo de la historia, la pobreza ha sido tradicionalmente medida en términos de ingresos, una aproximación que, aunque útil, resulta insuficiente para capturar la verdadera magnitud y diversidad de las experiencias de pobreza, y puede ocultar ciertas manifestaciones del fenómeno. Hoy en día, existe un amplio consenso acerca de la necesidad de tratar la pobreza como un fenómeno multidimensional, véanse las referencias de García-Gómez et al. (2021).

La pobreza no es simplemente la falta de dinero; es también la falta de oportunidades y opciones que permiten a las personas llevar una vida digna y satisfactoria. Por ejemplo, una persona puede tener ingresos suficientes para no ser considerada pobre desde una perspectiva monetaria, pero puede carecer de acceso a servicios de salud adecuados, educación de calidad o condiciones de vivienda dignas.

Uno de las primeras medidas de pobreza desde una perspectiva que va más allá de los ingresos y aplicable a un amplio número de países, es el Índice de Pobreza Multidimensional (MPI, por sus siglas en inglés). Introducido en 2010 por el Programa de las Naciones Unidas para el Desarrollo (PNUD) y la Iniciativa sobre Pobreza y Desarrollo Humano de la Universidad de Oxford, el MPI considera tres dimensiones del bienestar humano: la *Educación*, la *Salud* y el *Nivel de vida*. Estas tres dimensiones se componen a su vez por diez indicadores: *Años de escolaridad* y *Asistencia escolar* asociados a la dimensión *Educación*; *Mortalidad infantil* y *Nutrición* corresponden a la dimensión *Salud*; y *Combustible para cocinar*, *Saneamiento*, *Agua potable*, *Electricidad*, *Vivienda* y *Activos* están asociados a la dimensión *Nivel de vida*.

A pesar de que el MPI es una medida de referencia a nivel internacional, no capta las relaciones entre los diferentes componentes de la pobreza, una cuestión fundamental cuando se aborda la pobreza como un fenómeno multidimensional, véase García-Gómez et al. (2021). Entender cómo los factores que describen la pobreza se interrelacionan y se refuerzan mutuamente resulta crucial para desarrollar políticas públicas efectivas y sostenibles que puedan erradicar la pobreza de manera integral. Precisamente, el objetivo principal de este trabajo es analizar las relaciones entre las privaciones de los indicadores del MPI en 110 países en vías de desarrollo para los que se publica este

indicador. En particular, se busca aproximarnos a la interdependencia entre las variables objeto de estudio, analizando las relaciones bivariantes entre ellas.

Este trabajo tiene por tanto una clara vinculación con los Objetivos de Desarrollo Sostenible (ODS) establecidos por la ONU. Dado que nuestro estudio se centra en la pobreza, está directamente relacionado con el primer objetivo que establece la ONU denominado "Poner fin a la pobreza". Además, la consecución de otros objetivos está íntimamente relacionada con la mejora en ciertos indicadores del MPI. Por ejemplo, el ODS 2 "Hambre cero" está ligado al indicador de *Nutrición*, mientras que el ODS 3 "Salud y bienestar" se relaciona estrechamente con los indicadores de la dimensión *Salud*. El ODS 4 "Educación de calidad" está vinculado a los dos indicadores de la dimensión *Educación*, y el ODS 6 "Agua limpia y saneamiento" está directamente relacionado con los indicadores de *Agua potable y Saneamiento*. Finalmente, el ODS 7 "Energía asequible y no contaminante" se puede asociar con el indicador de *Electricidad*. Por tanto, una mejor comprensión de las dependencias en el nivel de privación de los indicadores del MPI puede ayudar a desarrollar políticas más efectivas de cara a alcanzar los Objetivos de Desarrollo Sostenible.

La metodología empleada es una adaptación de la propuesta en los artículos Wang y Xie (2016), Wen, Yang y Zhou (2019) y Qiao y Han (2023) en el ámbito de las finanzas. Esta metodología se basa en la construcción de un árbol de expansión mínima (*Minimum Spanning Tree*) a partir de grafos que tienen, como nodos, el nivel de privación o pobreza en cada uno de los indicadores del MPI, y, cuyos pesos de las aristas vendrán determinados por tres medidas de dependencia bivariantes basadas en rangos. En concreto, emplearemos el coeficiente de correlación de Spearman y dos medidas de dependencia en las colas de la distribución conjunta: una para la cola derecha y otra para la cola izquierda. De las redes resultantes de emplear como pesos estas dos medidas de dependencia en las colas, será de especial interés el análisis de aquella en la que se emplean los coeficientes de dependencia en la cola derecha. Dado que las variables empleadas son los niveles de privación en los indicadores del MPI, esta red reflejará las relaciones entre dichas variables para los países con un nivel de privación más elevado, es decir, será la red que refleje cómo se dan estas relaciones para los países que se encuentran en una situación de pobreza más severa.

Esta metodología permite visualizar las interconexiones entre las privaciones de los indicadores del MPI de una manera intuitiva y comprensible gracias al uso de los grafos. Generando el *Minimun Spanning Tree*, logramos simplificar el grafo original, poniendo el énfasis en las conexiones más fuertes entre las tasas de privación en los indicadores. Además, este enfoque permite determinar las privaciones clave y las interconexiones que merecen atención prioritaria a la hora de diseñar intervenciones que busquen tener un impacto global sobre la situación de pobreza.

Los resultados revelan que la relación entre el nivel de privación en las variables que componen la pobreza es mucho más intensa en los países más pobres. También se observa que las relaciones entre las tasas de privación en los indicadores del MPI son más fuertes entre los indicadores de una misma dimensión que entre los pertenecientes a diferentes dimensiones. Además, los resultados muestran diferencias importantes en los elementos principales de las distintas redes obtenidas. Esto pone en relieve que, dependiendo del nivel de pobreza del país, unas variables serán más o menos determinantes que otras en cuanto a su impacto sobre el resto de variables. En particular, se ha observado que el nivel de privación en el indicador del MPI *Electricidad* es crítico para los países más pobres, afectando este nivel de privación significativamente al resto de variables, mientras que, en los países menos pobres, las relaciones que presenta esta variable con el resto son mucho más débiles.

La estructura del trabajo es la siguiente. La Sección 2 introduce el Índice de Pobreza Multidimensional. A continuación, se describe la metodología empleada, haciendo especial énfasis en cuatro aspectos: los grafos, las medidas de dependencia basadas en rangos, los subgrafos y las medidas de centralidad en los grafos. En la Sección 4 se realiza un análisis univariante de los datos empleados con el fin de obtener un contexto adecuado sobre los mismos de cara a su posterior análisis. Antes de generar los grafos, en la Sección 5, se comienzan a estudiar las relaciones entre las variables, a través de la representación de diagramas de dispersión y el cálculo de diferentes coeficientes de dependencia basados en rangos. Empleando estos coeficientes se generan los grafos presentados y analizados en la Sección 6, los cuales permiten comprender las relaciones entre las variables que componen el MPI y como varían estas relaciones en función del nivel de pobreza. Por último, la Sección 7 recoge las principales conclusiones del estudio.



## 2. EL ÍNDICE DE POBREZA MULTIDIMENSIONAL

El Índice de Pobreza Multidimensional (MPI) es una medida de pobreza basada en la metodología de Alkire y Foster (2011), adoptada por el Programa de Naciones Unidas para el Desarrollo para realizar un seguimiento de la pobreza en más de 100 países en vías de desarrollo. En el presente trabajo se emplearán datos basados en la revisión Alkire (2018). A diferencia de las medidas tradicionales de pobreza, basadas en los ingresos, el MPI considera que la pobreza está caracterizada por las privaciones que presentan las personas en tres dimensiones: *Salud*, *Educación* y *Nivel de vida*. A su vez, estas dimensiones se describen por una serie de indicadores: dos asociados a la dimensión *Salud*, dos asociados a *Educación* y seis asociados al *Nivel de vida*. La Tabla 2.1 recoge la definición de cada indicador que compone MPI, el umbral de privación de cada uno de ellos y la dimensión a la que está asociado cada indicador. En cada país se selecciona una muestra representativa y se evalúa si las personas de dicha muestra se encuentran en situación de privación según el umbral de cada indicador. Dado que, generalmente, los datos se recogen a nivel de hogar, se supone que todos los miembros del hogar comparten la misma situación de privación. Por ejemplo, si una familia de 5 miembros vive en un hogar que no tiene acceso a electricidad, a los 5 miembros de ese hogar se les considera que sufren privaciones en el indicador asociado.

El cálculo del MPI para cada país implica saber, por un lado, el porcentaje de personas que son pobres multidimensionales (la incidencia de la pobreza) y, por otro, el valor medio de privaciones que tienen los pobres multidimensionales (la intensidad de la pobreza). Para su cálculo, el peso asignado a cada dimensión es el mismo y, dentro de cada dimensión, los indicadores tienen la misma importancia; véase los pesos de cada indicador del MPI en la Tabla 2.1. Asimismo, se considera que una persona es pobre multidimensional si se encuentra privado en una dimensión o si la suma de los pesos de los indicadores del MPI es igual o mayor a  $1/3$ . Por ejemplo, una persona será pobre multidimensional si los niños de su familia no van al colegio ( $1/6$ ) y si su casa no tiene electricidad ( $1/18$ ), no tiene agua ( $1/18$ ) y las condiciones sanitarias son deficientes ( $1/18$ ).

Dimensiones de pobreza	Indicador MPI	Se considera privación si...	Peso
Salud	<b>Nutrición:</b> Evaluación del estado nutricional de los miembros del hogar.	Cualquier persona menor de 70 años para la cual un examen nutricional de resultado de desnutrición.	1/6
	<b>Mortalidad infantil:</b> Registro de muertes de niños menores de cinco años en el hogar.	Cualquier niño ha fallecido en la familia en los cinco años anteriores al estudio.	1/6
Educación	<b>Años de escolaridad:</b> : Número de años de educación formal completados por los miembros del hogar	Ningún miembro del hogar de 10 años o más ha completado cinco años de escolaridad.	1/6
	<b>Asistencia escolar:</b> Inscripción de los niños en edad escolar en instituciones educativa	Cualquier niño en edad escolar no asiste a la escuela hasta la edad en la que debería completar el grado 8.	1/6
Nivel de vida	<b>Combustible para cocinar:</b> Uso de combustibles limpios para cocinar.	El hogar cocina con estiércol, madera o carbón.	1/18
	<b>Saneamiento:</b> Acceso a instalaciones sanitarias adecuadas y su exclusividad para el hogar.	La instalación de saneamiento del hogar no ha mejorado (según las directrices de los ODS) o es compartida con otros hogares.	1/18
	<b>Agua potable:</b> Acceso a una fuente de agua potable segura y cercana.	El hogar no tiene acceso al agua potable mejorada (según las directrices de los ODS) o el agua potable está a más de 30 minutos a pie (ida y vuelta).	1/18
	<b>Electricidad:</b> Disponibilidad de electricidad en el hogar.	El hogar no tiene electricidad.	1/18
	<b>Vivienda:</b> Calidad de los materiales del piso, paredes y techo de la vivienda.	El hogar tiene una vivienda inadecuada: el piso es de materiales naturales o el techo o las paredes son de materiales rudimentarios.	1/18
	<b>Activos:</b> Propiedad de bienes básicos y activos productivos, como medios de transporte, teléfonos, ganado, etc	El hogar no posee más de uno de estos activos: radio, TV, teléfono, computadora, carro de animales, bicicleta, motocicleta o refrigerador, y no posee un coche o camión.	1/18

Tabla 2.1: Indicadores que componen el MPI con sus dimensiones y pesos asociados (Alkire 2018)

Se emplea una muestra de 110 países en desarrollo, la cual cubre aproximadamente 6.1 mil millones de personas o, lo que es lo mismo, un 92% de la población mundial. Al ser un indicador que centra el foco sobre la pobreza aguda, cubre únicamente países en desarrollo, sin incluir por tanto los países ya desarrollados. Otro hecho necesario para que un país sea cubierto por el MPI es la disponibilidad de datos comparables. El MPI utiliza datos de encuestas nacionales comparables a nivel internacional, como las Encuestas Demográficas y de Salud (DHS) y las Encuestas de Indicadores Múltiples por Conglomerados (MICS), por lo que un país no podrá ser cubierto por este indicador si no existen estos datos de encuestas nacionales.

### 3. METODOLOGÍA

A continuación, se describe brevemente la metodología que se va emplear, similar a la empleada en los artículos Wang y Xie (2016), Wen, Yang y Zhou (2019) y Qiao y Han (2023) en el ámbito de las finanzas. En primer lugar, introduciremos las nociones más básicas de los grafos. Posteriormente explicaremos las medidas que emplearemos como pesos de las aristas de dichos grafos. A continuación, se introducirá el concepto de subgrafo, haciendo especial énfasis en el Árbol de Expansión Mínima, ya que será el tipo de subgrafo que emplearemos en este estudio. Por último, expondremos las métricas que emplearemos para analizar las redes generadas a lo largo del estudio.

#### 3.1. Grafos

La teoría de grafos es una rama de las matemáticas que se centra en el estudio de grafos, que son estructuras formadas por **nodos** (o vértices) y las conexiones entre ellos, denominadas **aristas**. Su origen se remonta al año 1736, cuando el famoso matemático suizo Leonhard Euler resolvió el problema de los puentes de Königsberg como recogen Biggs, Lloyd y J. Wilson (1986). Este problema consistía en determinar si era posible recorrer todos los puentes de la ciudad de Königsberg sin cruzar el mismo puente dos veces. Visualmente un grafo, en su forma más simple, se puede concebir como un conjunto de puntos (nodos) conectados por líneas (aristas), se trata por tanto de una representación en forma de red. Empleando este tipo de representación, con resultado similar al expuesto en la parte derecha de la Figura 3.1, Euler demostró que no es posible recorrer todos estos puentes pasando una única vez por cada uno. Asimismo, a partir de la resolución de este problema, Euler estableció los fundamentos de lo que hoy conocemos como teoría de grafos.

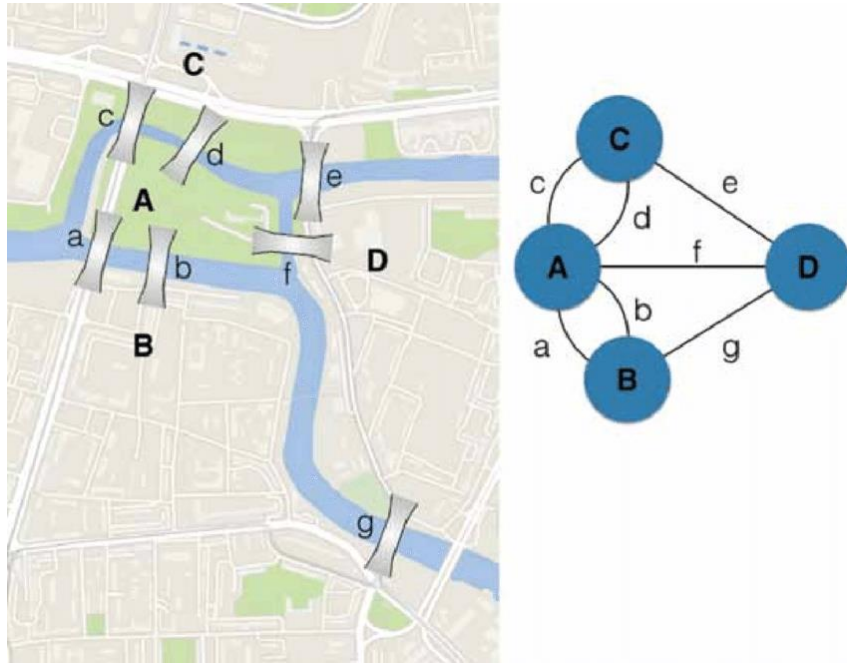


Figura 3.1: Visualización del problema original de los puentes de la ciudad de Königsberg resuelto por Euler, tanto sobre un mapa (izquierda) como en forma de grafo (derecha) (Hart et al. 2016).

Esta primera y sencilla representación de un grafo mostró un gran potencial para simplificar y analizar sistemas complejos de una forma estructurada y clara. La teoría de grafos ha evolucionado enormemente desde la versión propuesta por Euler, aumentando considerablemente su potencial y versatilidad para resolver una amplia variedad de problemas. Por ejemplo, en informática, se usa para diseñar y optimizar redes de comunicación y sistemas de información. En concreto, los algoritmos de búsqueda y optimización mediante grafos son fundamentales para el funcionamiento de motores de búsqueda en internet (i.e. Google, véase Sundkar (2023)), redes de ordenadores y bases de datos. Uno de los principales campos en los que se ha encontrado utilidad a la teoría de grafos es en logística y transporte, donde se emplea para optimizar rutas y minimizar costes al ayudar este tipo de representaciones a encontrar las rutas más eficientes. A parte de todas estas aplicaciones prácticas, los grafos también han sido ampliamente empleados en el ámbito académico para simplificar el análisis de ciertos problemas complejos. Concretamente, se han empleado en el ámbito de las finanzas, especialmente de cara a comprender mejor las relaciones de influencia entre distintos mercados financieros o distintos activos financieros como se puede ver en por ejemplo en Sweeting y Fotiou (2013), Wang y Xie (2016) y Wen, Yang y Zhou (2019).

En su formulación más básica, un grafo es un par de conjuntos  $(V, E)$ , donde  $V$  es un conjunto no vacío cuyos elementos son los nodos (o vértices) y  $E$  es un conjunto de pares de elementos de  $V$  que representan las conexiones (o aristas) entre ellos. Por ejemplo, consideremos las conexiones directas por avión entre Valladolid (Va), Madrid (M), Barcelona (B) y Nueva York (NY). En este caso, tenemos que  $V = \{Va, M, B, NY\}$  y  $E = \{(Va, M), (Va, B), (M, B), (M, NY), (B, NY)\}$ . En la Figura 3.2 se recoge el resultado gráfico de la representación de dicho grafo con las conexiones antes definidas.

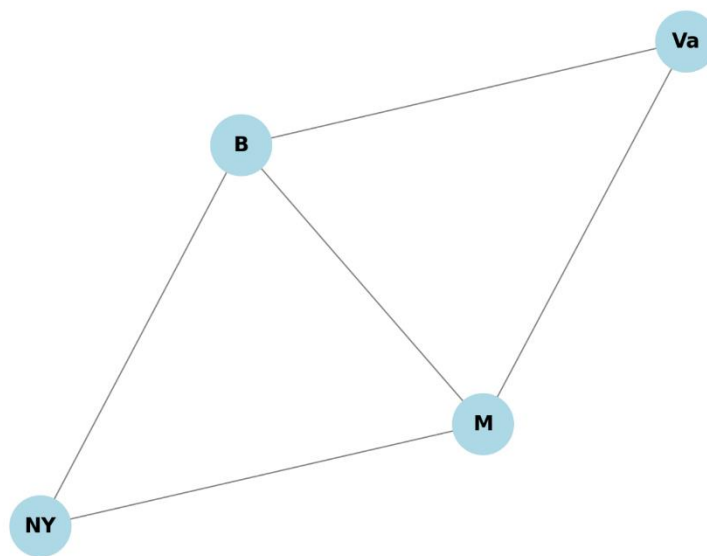


Figura 3.2: Ejemplo de un grafo simple para definir las conexiones aéreas entre 4 ciudades. (Elaboración propia)

En el ejemplo anterior, los nodos pueden estar conectados o no conectados. Si están conectados, las aristas simplemente indican la existencia de una conexión entre dos nodos, sin proporcionar más información sobre esa conexión, en este caso indican la existencia de conexión aérea entre las dos ciudades. Este tipo de grafo se conoce como **no ponderado**. Sin embargo, no todas las conexiones tienen por qué tener la misma importancia. Precisamente, un grafo **ponderado** se caracteriza por el hecho de que cada arista tenga asignado un valor numérico, llamado peso, que permite asignar diferentes niveles de importancia a cada conexión entre los distintos nodos. Siguiendo con el ejemplo anterior, se podrían asignar pesos a cada arista, por ejemplo, en función del coste medio del billete para el vuelo entre las dos ciudades implicadas.

La aparición de los grafos ponderados multiplica las posibles utilidades de los grafos, siguiendo con el ejemplo anterior, el grafo ya no sólo servirá para analizar la existencia de rutas o conexiones entre elementos, sino que podemos comenzar a hablar de optimización de rutas, al poder buscar la ruta que minimice o maximice la suma del indicador asociado a las conexiones (i.e. ruta de mínimo coste, mínimo tiempo, máximo ahorro...).

En el contexto del estudio sobre el nivel de privación de los indicadores del MPI, un grafo no ponderado podría simplemente indicar si existe o no una relación entre los niveles de pobreza de dos indicadores. Por otro lado, un grafo ponderado permitiría asignar un valor cuantitativo a la relación entre ambos niveles de privación, reflejando la intensidad o el grado de dicha relación. Así, un grafo ponderado nos permitirá analizar cuál es la privación que tiene una mayor relación con el resto y, por lo tanto, la privación que tiene una mayor influencia sobre el resto.

Otra clasificación de los grafos es en función de si las aristas tienen asociadas direcciones específicas de conexión o no, distinguiremos entre grafos dirigidos o no dirigidos. En un grafo **no dirigido**, las conexiones entre los nodos no tienen una dirección específica. En este tipo de grafo, cada arista  $(u, v) \in E(u, v)$  indica que existe una conexión bidireccional entre los nodos  $u$  y  $v$ . Esto significa que, si  $u$  está conectado a  $v$ , entonces  $v$  también está conectado a  $u$ . En caso de ser ponderado, esta conexión tendrá asociado el mismo peso en ambas direcciones. Los grafos no dirigidos son útiles para modelar situaciones donde la relación entre los nodos es simétrica. Por otro lado, un **grafo dirigido**, también conocido como dígrafo, incluye aristas con una dirección específica. En un grafo dirigido, cada arista  $(u, v) \in E(u, v)$  indica que hay una conexión dirigida del nodo  $u$  al nodo  $v$ . Esto implica que la relación no necesariamente es bidireccional; puede existir una conexión de  $u$  a  $v$ , pero no necesariamente de  $v$  a  $u$ . En el caso ponderado, pueden existir pares de nodos con conexión en las dos direcciones, pero con diferentes pesos asignados en función de la dirección. Los grafos dirigidos son particularmente útiles para modelar situaciones donde las relaciones entre elementos son asimétricas.

### **3.2. Los pesos: medidas de dependencia basadas en rangos**

Los pesos de las aristas en un grafo ponderado pueden tener diversos significados dependiendo de la aplicación. En una red de transporte, los pesos pueden representar la distancia o el tiempo entre dos puntos, normalmente se utilizarán para optimizar rutas en cuanto a tiempo/distancia recorrida. En análisis de redes eléctricas, los pesos pueden reflejar la capacidad de transmisión o el coste de conexión, ayudando a analizar la eficiencia del suministro. En cada caso, los pesos proporcionan una medida cuantitativa que facilita la optimización y el análisis de las relaciones y estructuras dentro del grafo.

En nuestro caso concreto, estamos interesados en estudiar las relaciones dos a dos entre las tasas de privación de los 10 indicadores del MPI por lo que estos pesos deberán corresponderse con una medida que refleje el grado de relación existente entre las diferentes parejas variables.

Una de las medidas más habituales para analizar las relaciones entre dos variables es el coeficiente de correlación de Pearson. Sin embargo, en nuestro caso, no lo consideramos adecuado por dos motivos. En primer lugar, este coeficiente es adecuado para medir relaciones lineales, es decir, cuando se asume que la distribución conjunta de las dos variables es una distribución gaussiana. Sin embargo, como veremos posteriormente, este supuesto no es adecuado con los datos empleados. En segundo lugar, el coeficiente de correlación de Pearson asume una relación homogénea a lo largo de todo el rango de valores de las variables, ignorando que pueden existir relaciones distintas en diferentes segmentos de dicho rango. En este estudio, dada la heterogeneidad de los países considerados, resulta particularmente interesante emplear medidas que centren su atención en las relaciones que se producen en los países que se encuentran, por un lado, con los mayores niveles de privación y, por otro, aquellos que tengan menores niveles de privación. Por ejemplo, nos interesa saber en qué medida los países que se encuentran entre los peores en acceso a electricidad también están entre los peores en acceso a la educación, o si los países que se encuentran entre los mejores en acceso a electricidad también están entre los mejores en acceso a la educación. Las medidas basadas en rangos permiten salvar ambas deficiencias y algunas de ellas son las que empleemos a lo largo de este estudio.



A continuación, introducimos la notación que nos permitirá definir las medidas basadas en rangos empleadas. Sea  $X_u = (X_{1u}, \dots, X_{iu}, \dots, X_{nu})$  y  $X_v = (X_{1v}, \dots, X_{iv}, \dots, X_{nv})$  los vectores de las observaciones de las variables  $u$  y  $v$ . A partir de los valores de las variables, definimos los vectores de rangos  $R_u = (R_{1u}, \dots, R_{iu}, \dots, R_{nu})$  y  $R_v = (R_{1v}, \dots, R_{iv}, \dots, R_{nv})$  formados por la posición que ocupa cada observación una vez ordenadas las variables de menor a mayor.

La primera medida de dependencia basada en rangos que se va emplear es el **coeficiente de correlación de Spearman**. El coeficiente de correlación de Spearman es el coeficiente de correlación lineal entre los rangos de las variables. Es decir, el coeficiente de correlación de Spearman entre las variables  $u$  y  $v$ ,  $r_{uv}^s$ , se define como:

$$r_{uv}^s = \frac{S_{R_u R_v}}{S_{R_u} S_{R_v}}$$

donde  $S_{R_u R_v}$  es la covarianza entre  $R_u$  y  $R_v$ ; y  $S_{R_u}$  y  $S_{R_v}$  son las desviaciones típicas de  $R_u$  y  $R_v$ , respectivamente. Si no se producen empates, la expresión anterior del coeficiente de correlación de Spearman se puede simplificar de la siguiente manera

$$r_{uv}^s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Siendo  $d_i$  la diferencia en los rangos de dos variables para el  $i$ -ésimo par de datos  $d_i = R_{i_u} - R_{i_v}$ . El coeficiente de correlación de Spearman varía entre -1 y 1. Valores positivos indicarán una relación positiva entre los rangos de las variables, valiendo 1 cuando la concordancia de las dos ordenaciones es perfecta, esto es, cuando el rango de las dos variables es exactamente idéntico. Valores negativos indicarán relación negativa entre los rangos, tomando el valor -1 cuando presenta concordancia o asociación inversa u opuesta, dicho de otra forma, cuando el rango de una variable es el inverso del de la otra Escuder Vallés y Santiago Murgui (2011). Si dos variables son independientes, este coeficiente tomará valores cercanos a 0. Una propiedad importante es que  $r_{uv}^s = r_{vu}^s$ , por lo tanto, asignará el mismo valor a la relación de  $v$  con  $u$  y de  $u$  con  $v$ . Esto es importante a la hora de elegir el tipo de grafo que vamos a emplear, en este caso, al ser los pesos iguales para la conexión  $u v$  que para la conexión  $vu$ , podemos hablar de un grafo no dirigido

El coeficiente de correlación de Spearman captura relaciones que no necesariamente son lineales. Sin embargo, sigue siendo una medida que se define a lo largo de todo el campo de variación de las variables. Esto implica que, al igual que el coeficiente de Pearson, no permite centrarse en partes específicas de ese campo de variación.

Para abordar esta limitación, definiremos medidas que se centran en las relaciones que se producen en los cuantiles inferiores y superiores de las variables. En Sweeting y Fotiou (2013) aparecen con el nombre de coeficientes de dependencia finita en la cola inferior y en la cola superior, ya que son versiones simplificadas de las medidas de dependencia en las colas introducidas por Sibuya (1960). En este trabajo buscamos centrar nuestra atención en los extremos de la distribución, por lo que nos centraremos en las relaciones entre pares de variables cuando ambas variables toman valores inferiores al primer cuartil y cuando ambas variables toman valores superiores al tercer cuartil. Al coeficiente que mide el primer tipo de relaciones, lo denominaremos coeficiente de dependencia en la cola inferior (o cola izquierda), y al coeficiente que mide el segundo tipo, lo llamaremos coeficiente de dependencia en la cola superior (o cola derecha).

Con el fin de definir ambos coeficientes se obtienen los rangos normalizados de las variables, dividiendo los rangos entre el número de observaciones más uno. Sea  $R_u^{nm} = (R_{1u}^{nm}, \dots, R_{iu}^{nm}, \dots, R_{nu}^{nm})$  el vector de rangos normalizados o posiciones relativas de la variable  $u$  tal que  $R_{iu}^{nm} = \frac{R_{iu}}{n+1}$  con  $i = 1, \dots, n$ . De la misma forma se define el vector de rangos normalizados de la variable  $v$ ,  $R_v^{nm} = (R_{1v}^{nm}, \dots, R_{iv}^{nm}, \dots, R_{nv}^{nm})$  donde  $R_{iv}^{nm} = \frac{R_{iv}}{n+1}$  con  $i = 1, \dots, n$ . Los dos coeficientes que vamos a utilizar para medir la dependencia en los cuantiles son una frecuencia relativa condicionada.

En concreto, el **coeficiente de dependencia en la cola inferior** (o cola izquierda) entre las variables  $u$  y  $v$ ,  $q_{uv}^{L,0.25}$ , es la proporción de individuos que presenta valores menores o iguales al primer cuartil de la variable  $v$ , condicionada a que los individuos presenten valores menores o iguales al primer cuartil de la variable  $u$ , es decir,

$$q_{uv}^{L,0.25} = \frac{\sum_i^n I_{\{R_{iu}^{nm} \leq 0.25 \cap R_{iv}^{nm} \leq 0.25\}}}{\sum_i^n I_{\{R_{iu}^{nm} \leq 0.25\}}}$$

donde  $I_{\{\cdot\}}$  es la función indicador del conjunto que aparece entre llaves, es decir, es una función que vale 1 si se cumple la condición definida entre llaves y 0 en caso contrario.

Nótese que  $q_{uv}^{L,0.25} = q_{vu}^{L,0.25}$ , propiedad fundamental para determinar el tipo de grafo que se van emplear. Con las variables utilizadas en nuestro estudio, las tasas de pobreza en los indicadores del MPI,  $q_{uv}^{L,0.25}$  mide la tendencia en los países analizados a que la tasa de pobreza en uno de los indicadores sea baja, inferiores al primer cuartil, si la tasa de privación en el otro indicador es también baja.

El coeficiente  $q_{uv}^{L,0.25}$  varía entre 0 y 1, ya que se trata de una frecuencia relativa. El coeficiente  $q_{uv}^{L,0.25} = 1$ , cuando todos los individuos que presentan valores por debajo del primer cuartil en la variable  $u$  también presentan valores por debajo del primer cuartil de la variable  $v$ . En este caso, diremos que hay una máxima asociación positiva en la cola izquierda. Esto sucede, por ejemplo, cuando las posiciones relativas de las dos variables son iguales en la parte baja de la distribución, pero no necesariamente en la parte alta. Por otro lado, el coeficiente  $q_{uv}^{L,0.25} = 0$ , cuando no hay ningún individuo que presente valores por debajo del primer cuartil en las dos variables simultáneamente. Una situación en la que  $q_{uv}^{L,0.25} = 0$  indicará por tanto que las variables presentan asociación inversa. Si las variables son independientes,  $q_{uv}^{L,0.25} = 0.25$ . Este último valor del coeficiente nos permitirá determinar si las variables muestran una asociación positiva en la cola inferior. Así, diremos que hay una asociación positiva en la cola izquierda si  $q_{uv}^{L,0.25} > 0.25$ , de tal forma que cuanto mayor sea este coeficiente, mayor será el grado de asociación positiva en la cola inferior.

El segundo coeficiente, el **coeficiente de dependencia en la cola superior** (o cola derecha) entre las variables  $u$  y  $v$ ,  $q_{uv}^{U,0.25}$ , es la proporción de individuos que presenta valores mayores al tercer cuartil de la variable  $v$ , condicionada a que los individuos presenten valores mayores al tercer cuartil de la variable  $u$  es decir,

$$q_{uv}^{U,0.25} = \frac{\sum_{i=1}^n I_{\{R_{iu}^{nm} > (1-0.25) \cap R_{iv}^{nm} > (1-0.25)\}}}{\sum_i I_{\{R_{iu}^{nm} > (1-0.25)\}}}$$

También cumple la propiedad  $q_{uv}^{U,0.25} = q_{vu}^{U,0.25}$ . En el caso de nuestras variables, que miden tasas de privación en los indicadores del MPI,  $q_{uv}^{U,0.25}$  mide la tendencia a que la tasa de privación en uno de los indicadores sea alta, superiores al tercer cuartil, cuando la tasa de privación en el otro indicador del MPI sea también alta

De forma similar al coeficiente de dependencia en la cola izquierda,  $q_{uv}^{U,0.25}$  varía entre 0 y 1:  $q_{uv}^{U,0.25}=1$ , cuando todos los individuos que presenta valores por encima del tercer cuartil en la variable  $u$  también presentan valores por encima del tercer cuartil de la variable  $v$  (máxima asociación entre las dos variables en la cola derecha);  $q_{uv}^{U,0.25}=0$ , cuando no hay ningún individuo que presente valores por encima del tercer cuartil en las dos variables simultáneamente; si las variables son independientes,  $q_{uv}^{U,0.25}=0.25$ . De nuevo, 0.25 será el límite que consideraremos para señalar que existe una asociación positiva en la cola superior. Por tanto, diremos que hay una asociación positiva en la cola superior si  $q_{uv}^{U,0.25}>0.25$ , de tal forma que cuanto mayor sea este coeficiente mayor nivel de asociación positiva hay en la cola superior.

En este trabajo, los coeficientes de dependencia en las colas toman como referencia los cuartiles, aunque se podría utilizar como referencia cualquier otro cuartil como se ha hecho en Venter (2001).

Como resumen del marco metodológico del trabajo visto hasta ahora, se emplearán grafos ponderados, no dirigidos y completamente conectados. En estos grafos, los nodos representarán las tasas de privación en cada uno de los 10 indicadores del MPI, y los pesos de las conexiones entre nodos estarán determinados por las tres medidas de dependencia basadas en rangos mencionadas anteriormente, a saber, el coeficiente de correlación de Spearman, el coeficiente de dependencia en la cola inferior y el coeficiente de dependencia en la cola superior.

### **3.3. Subgrafos**

Un subgrafo es una parte de un grafo más grande, formado al seleccionar un subconjunto de sus nodos y de aristas que los conectan. Resultan importantes ya que permiten analizar y entender secciones más pequeñas y manejables de un grafo complejo. Pueden ser muy útiles a la hora de simplificar la resolución de problemas al descomponer un grafo grande en partes más pequeñas que se pueden abordar individualmente.

La forma de generar un subgrafo dependerá del uso que se le quiera dar y la causa por la que se quiera simplificar el grafo original. Por ejemplo, en una red de carreteras

nacional en la que las ciudades son los nodos, un subgrafo podría representar solo las carreteras dentro de una región específica, permitiendo un análisis detallado del tráfico en esa área sin considerar toda la red nacional. Sin embargo, si lo que se quiere es analizar el tráfico a nivel nacional pero sólo en las principales vías, se podrían mantener todas las ciudades en el subgrafo, pero conservar únicamente las autopistas eliminando el resto de vías. Por tanto, resultará esencial definir un criterio para a la hora de generar el subgrafo.

En nuestro caso concreto, emplearemos el subgrafo conocido como **Minimum Spanning Tree (MST)**. Para entender sus características es necesario comprender el concepto de **ciclo** en un grafo. En teoría de grafos, un ciclo, también llamado bucle, es una secuencia de nodos y aristas que comienza y termina en el mismo nodo, sin repetir ninguna arista o nodo (excepto el nodo inicial/final). Formalmente, un ciclo en un grafo  $G = (V, E)$  es una secuencia de nodos  $v_1, v_2, \dots, v_k$  que cumple las siguientes 3 condiciones:

1.  $v_1 = v_k$  (el ciclo comienza y termina en el mismo nodo).
2. Para cada par consecutivo de nodos existe una arista  $(v_i, v_{i+1}) \in E$  con  $1 \leq i < k$  (para cada par consecutivo de nodos existe una arista que los une).
3. Todos los nodos  $v_1, v_2, \dots, v_{k-1}$  son distintos entre sí.

Un ciclo es una característica importante en los grafos, ya que puede indicar la presencia de bucles o redundancias en la estructura de la red. En la Figura 3.3 se representa en color rojo un bucle dentro de un grafo. Por ejemplo, en una red de transporte, un ciclo podría representar una ruta circular que vuelve al punto de partida sin pasar dos veces por la misma carretera. Eliminar estas redundancias en la red suele ser uno de los objetivos principales al generar un subgrafo, ya que para muchas utilidades resulta interesante que únicamente exista un camino entre cada par de nodos de la red.

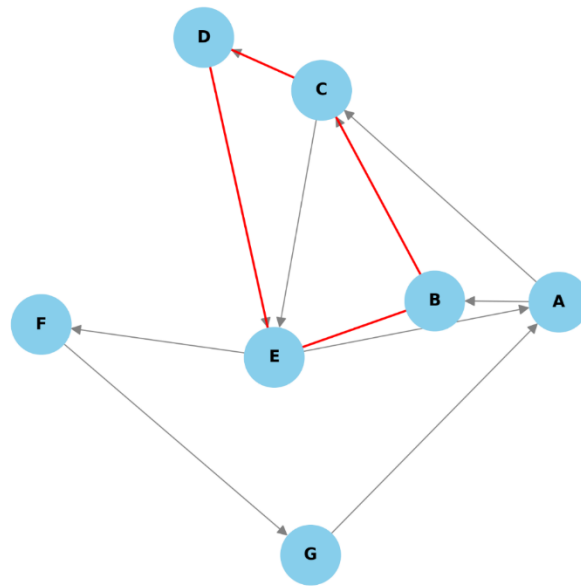


Figura 3.3: Visualización de un bucle (en rojo) dentro de un grafo (Elaboración propia).

El *Minimum Spanning Tree* o Árbol de Expansión Mínima (MST, por sus siglas en inglés) es una estructura fundamental en la teoría de grafos, crucial para resolver problemas de optimización en diversas áreas. En un grafo ponderado, el MST es el subgrafo cuya suma de los pesos asociados a todas sus aristas es mínima de entre todos los que cumplen las siguientes dos condiciones:

1. Conecta todos los nodos del grafo original.
2. No contiene ningún ciclo.

Estas dos condiciones implican que se trate de un subgrafo que contiene el mínimo número de aristas posibles de forma que se mantengan conectados todos los nodos de la red original. Es, por tanto, en término de número de aristas, el subgrafo más simple posible manteniendo todos los nodos originales.

Al minimizar la suma de los pesos de la red se mantienen únicamente las conexiones con pesos más bajos, es decir, se mantienen únicamente las conexiones más fuertes. Este hecho es una propiedad esencial en aplicaciones donde es necesario minimizar costes, como en el diseño de redes de telecomunicaciones, sistemas de distribución eléctrica, y rutas de transporte. Por ejemplo, en una red de distribución de electricidad, el MST asegura que todas las estaciones estén conectadas utilizando la mínima cantidad de cableado, lo cual reduce tanto los costes de instalación como las pérdidas de energía.

Tenemos por tanto la red más simple posible en término de número de aristas y manteniendo las conexiones más fuertes. El subgrafo MST ha de ser visto como una simplificación del grafo original, en el que se evitan las redundancias en los caminos de la red (los bucles) manteniendo las conexiones más fuertes posibles. Es por tanto una simplificación que nos permite destacar las conexiones más fuertes de la red original. En la Figura 3.4 se presenta un grafo y su subgrafo MST asociado.

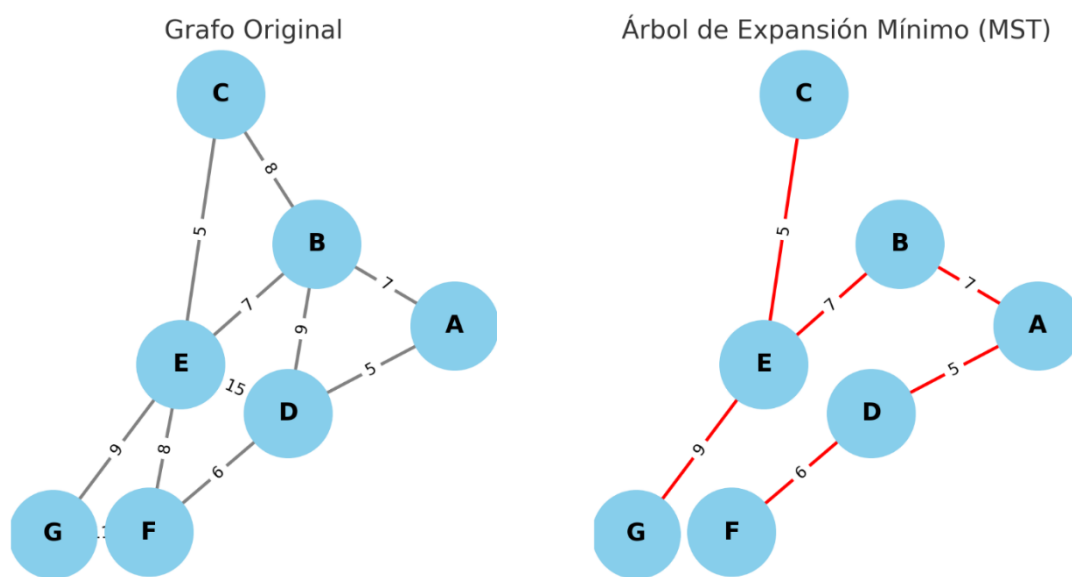


Figura 3.4: Visualización de un grafo y su MST asociado (Elaboración propia).

Los MST sirven para reducir notablemente la complejidad de las redes, al no presentar ningún ciclo y conectar todos los nodos, el número de aristas del subgrafo MST será siempre de  $D-1$  siendo  $D$  el número de nodos del grafo original. Hay que tener en cuenta que, en caso de que originalmente cada nodo presentara conexiones con todos y cada uno del resto de nodos de la red (lo que resulta bastante habitual y es el caso en este estudio) el número original de conexiones sería  $\frac{D \times (D-1)}{2}$ , por lo que, al generar la red MST, esta tendrá  $\frac{D}{2}$  veces menos aristas que la original. Al tener en nuestro caso 10 indicadores de pobreza,  $D = 10$ , se está reduciendo el número de aristas pasando de 45 a 9.

Los principales algoritmos para calcular el MST son el algoritmo de Kruskal y el algoritmo de Prim. El algoritmo de Kruskal construye el MST seleccionando las aristas de menor peso en orden ascendente, añadiéndolas al árbol sin formar ciclos, hasta conectar todos los nodos. Por el contrario, el algoritmo de Prim comienza con un nodo arbitrario y expande el MST seleccionando repetidamente la arista de menor peso que conecta un nodo dentro del árbol con uno fuera de él, hasta incluir todos los nodos, véase Singh (2024). La eficiencia de uno y otro a nivel de consumo de recursos computacionales en el cálculo del MST dependerá del grafo de origen, por lo que cada en función de este será más adecuado uno u otro. La función `nx.draw_networkx` de *Python*, empleada para la realización de este estudio (más información en los Anexos I y II), utilizará en cada caso el algoritmo más eficiente, es por esto que no podemos asegurar que algoritmo ha sido empleado en nuestro caso concreto, si bien el resultado será el mismo independientemente de cuál se emplee.

### 3.4. Métricas de la red

Una vez calculado el MST contamos con distintas métricas que podemos calcular con el fin de analizar más en profundidad las relaciones entre los elementos que las componen. Comenzaremos por presentar una medida que nos permitirá analizar el grado de conexión del grafo en su conjunto, a continuación, presentaremos 3 medidas destinadas a analizar el grado de relevancia de cada nodo dentro del grafo.

La primera que vamos a utilizar es **average shortest path** (camino más corto promedio),  $L$ , la cual es una medida de la eficiencia de la conexión de **un grafo en su conjunto**. Nos ayudará a identificar el nivel de conexión entre todos los elementos del grafo. Concretamente, se refiere a la longitud, entendida como suma de los pesos de las conexiones, media del camino más corto entre todos los pares posibles de nodos en el grafo. En otras palabras, calcula la distancia promedio que se debe recorrer por el camino más corto para ir de un nodo a otro cualesquiera a través de las aristas del grafo. Su fórmula es la siguiente:

$$L = \frac{1}{D \times (D - 1)} \times \sum \sum_{u \neq v} d(u, v)$$



Siendo  $d(u, v)$  la distancia por el camino más corto entre los nodos  $u$  y  $v$  de la red. Esta métrica es importante en varios campos, incluyendo la teoría de redes, la informática, y las ciencias sociales, ya que proporciona una idea de cómo de cercanos están los nodos en promedio dentro de la red (entendiendo la distancia como la suma de los pesos de todas las aristas presentes en el camino más corto entre dos nodos). En nuestro caso, esta métrica nos servirá para medir el grado de proximidad de los diferentes indicadores del MPI, un mayor valor de  $L$  mostrará menos relación entre los diferentes indicadores del MPI y, por tanto, una mayor independencia entre las componentes de la pobreza.

Adicionalmente a esta métrica general de toda la red, emplearemos 3 métricas de la fuerza de influencia **de cada nodo**:

El **grado de centralidad** (GC), mide la influencia de un nodo basándose en la cantidad de conexiones directas que tiene. La idea es analizar el nivel de centralidad de ese nodo en la estructura de la red. Se calcula como el número de conexiones directas de dicho nodo dividido entre el número de nodos menos 1:

$$GC(v) = \frac{\text{Número de conexiones del nodo } v}{D - 1}$$

La **centralidad de cercanía** (CC) mide el grado de cercanía de cada nodo con el resto de nodos de la red. Cuanto mayor sea este indicador, mayor es el grado de cercanía medio con el resto de nodos de la red. Se calcula como el inverso del sumatorio de la distancia más corta entre el nodo en cuestión y el resto de nodos de la red.

$$CC(v) = \frac{n - 1}{\sum_{u \neq v} d(u, v)}$$

La **centralidad de intermediación** (CI) mide la influencia de un nodo basándose en el número de veces que este actúa como puente a lo largo del camino más corto entre otros dos nodos. Se ha de interpretar como el porcentaje de caminos más cortos entre nodos distintos de él mismo en el que hay que pasar por el propio nodo. El objetivo con esta métrica es identificar en qué medida el nodo  $v$  se encuentra en los caminos más cortos entre nodos, es decir, en qué medida está presente en los caminos de unión más intensa entre los distintos nodos. Si por ejemplo tuviéramos un grafo de la red ferroviaria española, Madrid presentaría una elevada centralidad de intermediación, ya que una elevada proporción de los viajes interurbanos en tren requieren del paso por

Madrid si se quiere seguir la ruta más rápida. Sin embargo, siguiendo con este ejemplo, Barcelona presentaría un valor mucho más bajo de esta métrica, ya que, al ser una ciudad mucho más periférica, muy pocos viajes interurbanos en tren requieren pasar por Barcelona. La fórmula para su cálculo es la siguiente:

$$CI(v) = \sum_{u \neq v \neq t} \frac{\alpha_{ut}(v)}{\alpha_{ut}}$$

Siendo  $\alpha_{st}(v)$  el número de caminos más cortos entre los nodos  $s$  y  $t$  y  $\alpha_{st}$  el número total de caminos más cortos entre dichos nodos. En nuestro caso, al trabajar con MST sólo habrá un camino posible entre cada par de nodos, por lo que esta fórmula se simplifica a la siguiente, añadiendo un factor de  $\frac{1}{(D-1)^2}$  con el fin de normalizar y que tome valores entre 0 y 1, ya que  $(D-1)^2$  es el número total de caminos mínimos en una red MST entre nodos diferentes a  $v$ :

$$CI(v) = \frac{1}{(D-1)^2} \times \sum_{u \neq v \neq t} \alpha_{ut}(v)$$

#### 4. ANÁLISIS UNIVARIANTE DE LOS DATOS EMPLEADOS

Esta sección presenta un primer análisis exploratorio del porcentaje de personas que sufren privaciones en cada uno de los 10 indicadores del MPI en los 110 países objeto de estudio. Los datos han sido obtenidos del estudio *2023 Global Multidimensional Poverty Index (MPI)*, a través de su página web <https://ophi.org.uk/global-mpi/2023> OPHI (2023).

A partir de los principales estadísticos descriptivos que aparecen en la Tabla 4.1, se observa una notable heterogeneidad para los niveles de privación presentados tanto dentro de cada indicador como entre diferentes indicadores que componen el MPI.

La variabilidad dentro de cada indicador se puede apreciar a través del coeficiente de variación, el cual fluctúa entre 0.77, para las tasas de pobreza en *Nutrición y Combustible para cocinar*, y 1.2, para la de *Electricidad*. Estos valores, superiores a 0.5, indican una alta variabilidad, lo que sugiere grandes diferencias en el nivel de privación en cada aspecto del MPI entre los diferentes países. Por ejemplo, en Burundi, el 99% de la población sufre privaciones en *Combustible para cocinar*, mientras que en Barbados nadie se encuentra privado en este aspecto.

Si nos centramos en las diferencias entre los distintos indicadores, se observa que, el nivel de privaciones varía de forma importante según el indicador, siendo los indicadores relacionados con el *Nivel de vida* los que presentan mayores niveles de privación. Así, más del 25% de la población de la mitad de los países analizados presenta privaciones en *Combustible para cocinar*, *Saneamiento* y *Vivienda*. En contraste, se encuentra la *Mortalidad infantil*, puesto que en la mitad de los países el porcentaje de personas que vive en un hogar en el que ha fallecido un niño en la familia en los cinco años anteriores al estudio es inferior al 5% de la población.

	Max	País Max.	Min	País Min.	Media	Mediana	Coefficiente de variación	Jarque -Bera p-valor
<b>Nutrición</b>	60%	Niger	1%	Serbia	19%	17%	0,77	0,01
<b>Mortalidad Infantil</b>	19%	Niger	0%	China	4%	3%	0,92	0,01
<b>Años de escolaridad</b>	75%	Niger	0%	Turkmenistán	15%	10%	1,04	0,00
<b>Asistencia escolar</b>	59%	Niger	0%	Ucrania	13%	6%	1,14	0,00
<b>Combustible para cocinar</b>	99%	Burundi	0%	Barbados	47%	48%	0,77	0,00
<b>Saneamiento</b>	90%	Rep. Centroafricana	0%	Seychelles	36%	26%	0,82	0,00
<b>Agua potable</b>	69%	Rep. Centroafricana	0%	Turkmenistán	22%	15%	0,91	0,00
<b>Electricidad</b>	92%	Chad	0%	Jordán y Albania	24%	8%	1,20	0,00
<b>Vivienda</b>	97%	Kenia	0%	Seychelles	35%	27%	0,81	0,01
<b>Activos</b>	71%	Rep. Centroafricana	0%	Turkmenistán	16%	8%	1,14	0,00

Tabla 4.1: Indicadores descriptivos por variable para los datos empleados

La Figura 4.1, que muestra los diagramas de cajas de las variables, ilustra de manera visual las dos características antes mencionadas: las diferencias en el nivel de privación de los indicadores del MPI y la importante variabilidad dentro de cada variable. Este gráfico permite además analizar la forma de la distribución de las variables objeto de estudio, pudiendo observar las posibles asimetrías y la presencia de valores atípicos en cada una de ellas. Como se puede comprobar, hay una clara asimetría, siendo mucho más alargadas las colas superiores que las inferiores (asimetría a la derecha). En esta misma figura se aprecia también una media (representada por un aspa) generalmente superior a la mediana (representada por la línea horizontal), lo que es también una señal de asimetría. Finalmente se observa la presencia valores atípicos en la parte superior. Estas características evidencian una clara muestra de que la distribución de cada variable no es una distribución normal, hecho confirmado por el contraste de Jarque-Bera (véase el p-valor en la última columna de la Tabla 4.1).

La Tabla 4.1 contiene también los países que alcanzan el valor mínimo y máximo. A partir del análisis de estos países, ya se observa una tendencia a que los países con altos niveles de pobreza en un indicador tienden a presentar altos niveles de pobreza en otros indicadores también. Por ejemplo, Níger es el país con el mayor porcentaje de personas con privaciones en cuatro indicadores, mientras que la República Centroafricana muestra el mayor porcentaje en tres indicadores. Esta relación sugiere que los países que tienen más privaciones en un indicador suelen presentar también privaciones elevadas en otros indicadores. Este hecho resalta la importancia del análisis que se realizará en los siguientes apartados.

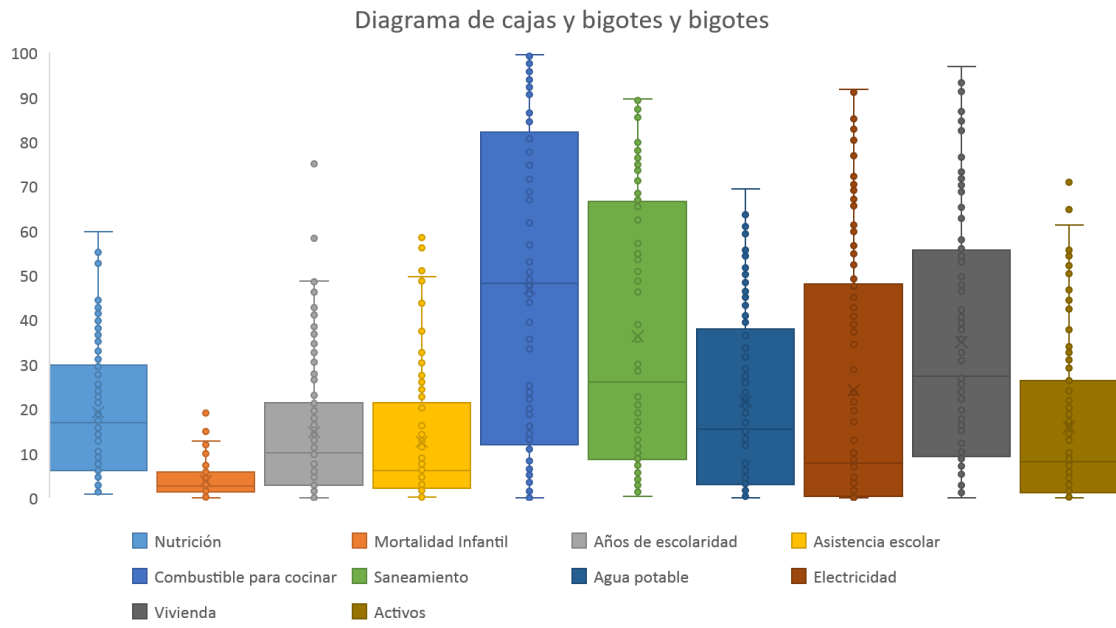
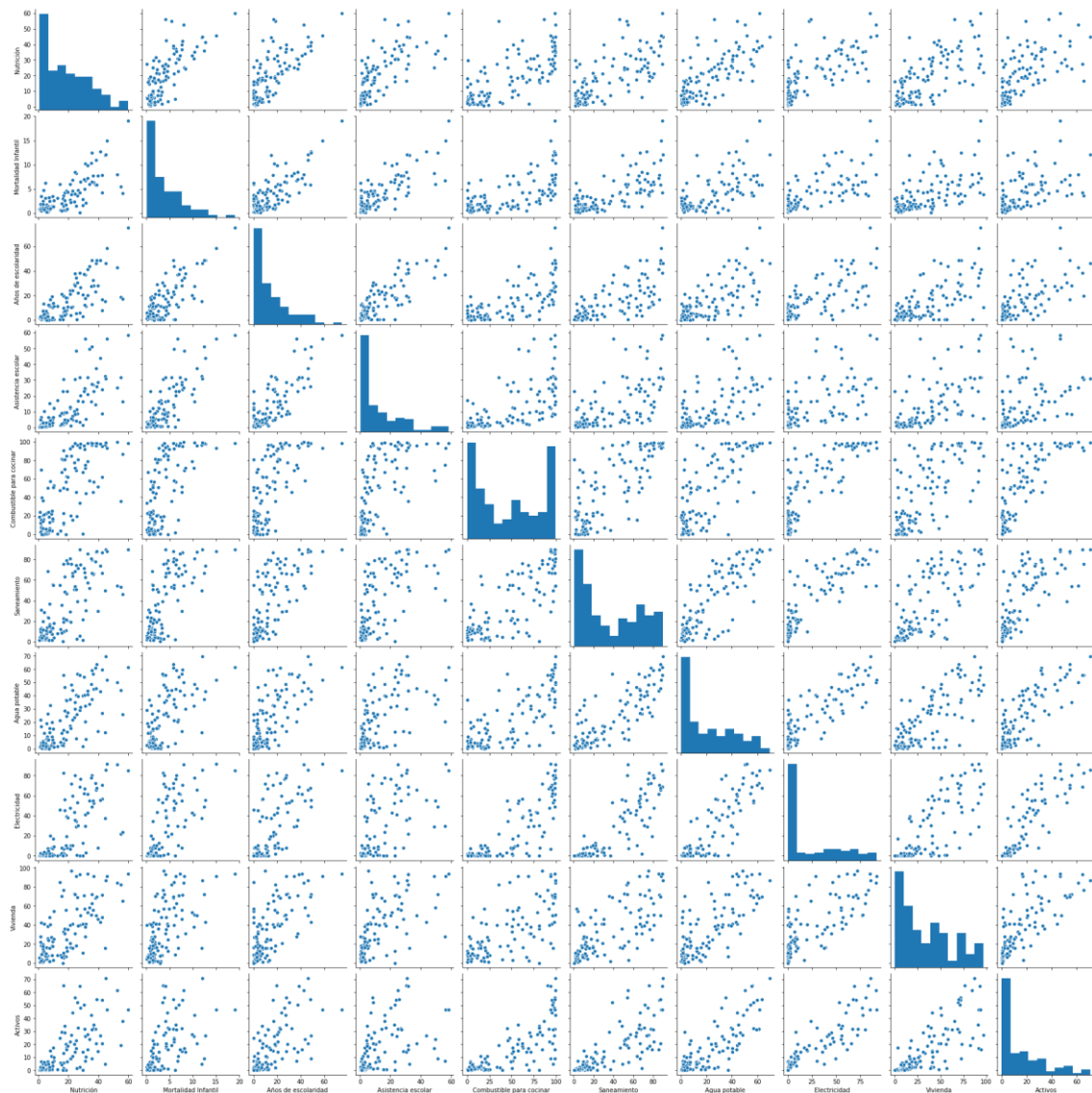


Figura 4.1: Diagrama de cajas y bigotes para los datos empleados

## 5. UNA PRIMERA VISIÓN DE LAS RELACIONES ENTRE LAS VARIABLES

En esta sección se comenzará analizando las relaciones existentes entre las variables a partir de los diagramas de dispersión de las variables originales y de los rangos normalizados. A continuación, presentaremos e interpretaremos los valores de los coeficientes de dependencia basados en rangos presentados en el Apartado 3.2



En la Figura 5.1 vemos los gráficos de dispersión para los porcentajes de privación, mientras que en la 5.2 encontramos el mismo gráfico, pero con las variables ya convertidas en rangos normalizados. En la Figura 5.1, se aprecia claramente que todas las relaciones entre pares de variables son positivas, lo que se ha de interpretar como que un elevado valor de privación en una variable normalmente va unido a un elevado

valor de privación en el resto de variables. Además, se aprecia que las relaciones que se establecen entre cada par de variables, por lo general no son lineales, existiendo diferentes tipos de relaciones. *Combustible para cocinar* y *Mortalidad infantil* guardan una relación que claramente no es lineal, su forma recuerda a la de un logaritmo. La pendiente entre *Activos* y *Electricidad* cambia a lo largo de la representación siendo mayor al inicio y al final, presentando una forma similar a la representación de una función tangente. Otras como *Años de escolaridad* y *Mortalidad infantil* muestran formas que parecen mucho más lineales.

En la Figura 5.2, se comprueba que la relación entre los rangos normalizados también es positiva. Además, en múltiples variables se aprecia que la asociación entre las mismas cambia a lo largo de la distribución. Por ejemplo, entre *Asistencia escolar* y *Años de escolaridad*, vemos que la relación es débil en la parte alta, sin embargo, la dispersión de los puntos disminuye según aumentamos el valor del rango normalizado, como resultado vemos una forma cónica en su respectivo diagrama de dispersión. Otro ejemplo, es *Agua potable* y *Activos* que presenta una relación mucho más elevada en ambos extremos de lo que lo hacen en el centro de la distribución de los rangos normalizados, lo que da como resultado una distribución con forma similar a una elipse. Esto enfatiza la idea original de que, para estudiar las relaciones entre estas variables, es necesario emplear métricas que permitan poner el énfasis en diferentes partes de la distribución ya que aquí se aprecia claramente como la relación entre el nivel de privación de los indicadores del MPI cambia a lo largo de la distribución.

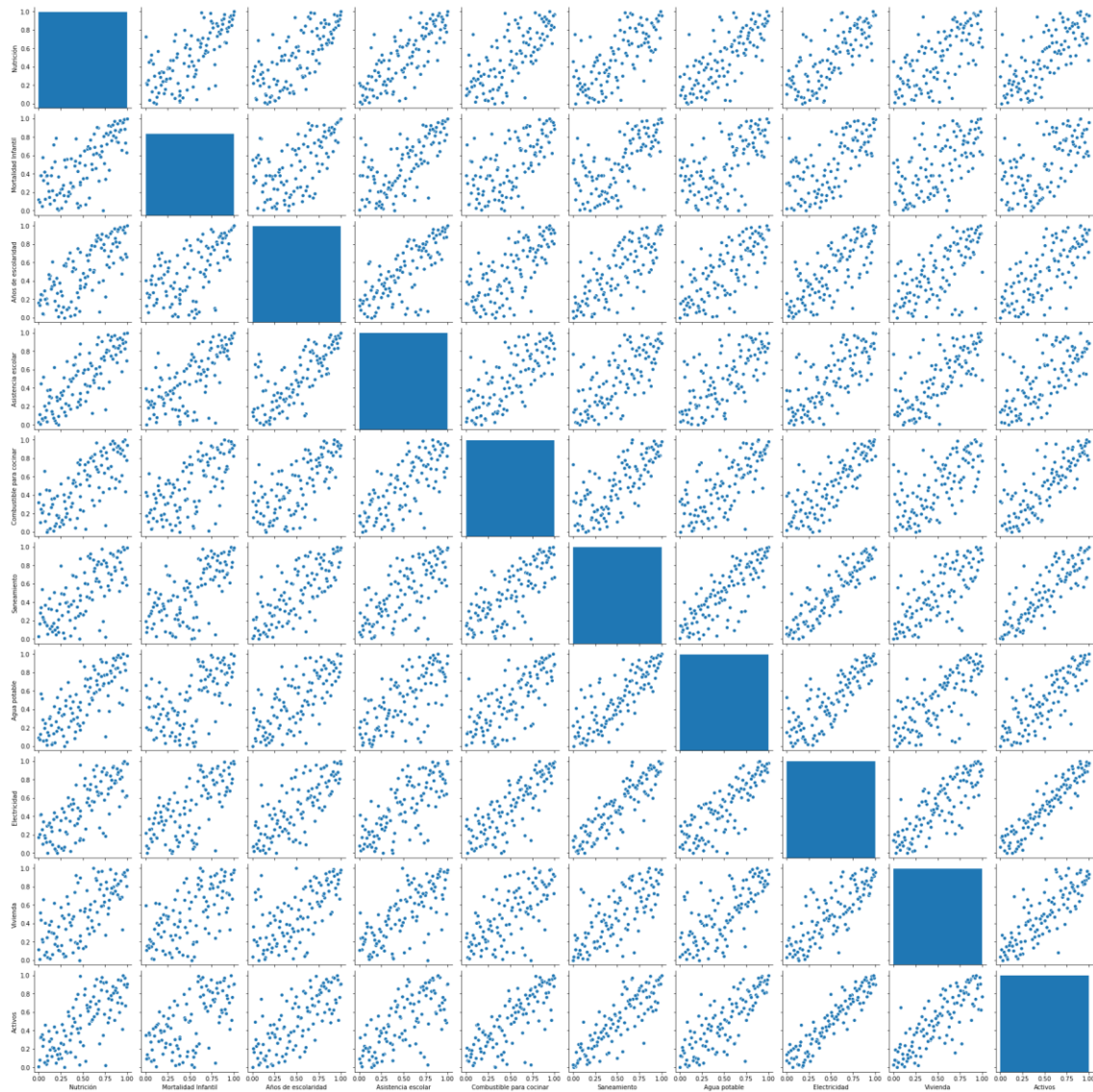


Figura 5.2: Gráficos de dispersión para los datos convertidos en rangos

Con el fin de completar la visión que se muestra en los diagramas de dispersión, las Tablas 5.1, 5.2 y 5.3 recogen, respectivamente, los valores obtenidos para los coeficientes de dependencia de cola derecha, de cola izquierda y de correlación de Spearman. Estas tablas presentan un código de color en el cual un verde más intenso representa un mayor valor del coeficiente en cuestión.



	Nutrición	Mortalidad Infantil	Años de escolaridad	Asistencia escolar	Combustible para cocinar	Saneamiento	Agua potable	Electricidad	Vivienda	Activos
Nutrición	1,00	0,74	0,63	0,78	0,63	0,56	0,67	0,70	0,59	0,59
Mortalidad Infantil	0,74	1,00	0,67	0,74	0,56	0,56	0,56	0,59	0,48	0,48
Años de escolaridad	0,63	0,67	1,00	0,67	0,67	0,63	0,67	0,74	0,59	0,63
Asistencia escolar	0,78	0,74	0,67	1,00	0,59	0,59	0,67	0,67	0,63	0,52
Combustible para cocinar	0,63	0,56	0,67	0,59	1,00	0,78	0,70	0,81	0,67	0,74
Saneamiento	0,56	0,56	0,63	0,59	0,78	1,00	0,70	0,74	0,56	0,63
Agua potable	0,67	0,56	0,67	0,67	0,70	0,70	1,00	0,78	0,67	0,81
Electricidad	0,70	0,59	0,74	0,67	0,81	0,74	0,78	1,00	0,70	0,81
Vivienda	0,59	0,48	0,59	0,63	0,67	0,56	0,67	0,70	1,00	0,67
Activos	0,59	0,48	0,63	0,52	0,74	0,63	0,81	0,81	0,67	1,00

Tabla 5.1: Valores obtenidos para los coeficientes de dependencia en la cola derecha

	Nutrición	Mortalidad Infantil	Años de escolaridad	Asistencia escolar	Combustible para cocinar	Saneamiento	Agua potable	Electricidad	Vivienda	Activos
Nutrición	1,00	0,50	0,50	0,61	0,57	0,61	0,61	0,57	0,54	0,61
Mortalidad Infantil	0,50	1,00	0,32	0,46	0,46	0,32	0,36	0,39	0,46	0,32
Años de escolaridad	0,50	0,32	1,00	0,64	0,54	0,68	0,64	0,61	0,57	0,61
Asistencia escolar	0,61	0,46	0,64	1,00	0,57	0,61	0,64	0,64	0,64	0,64
Combustible para cocinar	0,57	0,46	0,54	0,57	1,00	0,46	0,54	0,61	0,61	0,64
Saneamiento	0,61	0,32	0,68	0,61	0,46	1,00	0,64	0,75	0,68	0,79
Agua potable	0,61	0,36	0,64	0,64	0,54	0,64	1,00	0,64	0,54	0,64
Electricidad	0,57	0,39	0,61	0,64	0,61	0,75	0,64	1,00	0,79	0,89
Vivienda	0,54	0,46	0,57	0,64	0,61	0,68	0,54	0,79	1,00	0,82
Activos	0,61	0,32	0,61	0,64	0,64	0,79	0,64	0,89	0,82	1,00

Tabla 5.2: Valores obtenidos para los coeficientes de dependencia en la cola izquierda

	Nutrición	Mortalidad Infantil	Años de escolaridad	Asistencia escolar	Combustible para cocinar	Saneamiento	Agua potable	Electricidad	Vivienda	Activos
Nutrición	1,00	0,71	0,73	0,79	0,73	0,70	0,76	0,76	0,69	0,73
Mortalidad Infantil	0,71	1,00	0,59	0,70	0,65	0,66	0,61	0,69	0,63	0,61
Años de escolaridad	0,73	0,59	1,00	0,75	0,71	0,76	0,73	0,77	0,63	0,74
Asistencia escolar	0,79	0,70	0,75	1,00	0,74	0,70	0,69	0,73	0,72	0,69
Combustible para cocinar	0,73	0,65	0,71	0,74	1,00	0,75	0,74	0,82	0,69	0,79
Saneamiento	0,70	0,66	0,76	0,70	0,75	1,00	0,82	0,89	0,73	0,85
Agua potable	0,76	0,61	0,73	0,69	0,74	0,82	1,00	0,83	0,72	0,81
Electricidad	0,76	0,69	0,77	0,73	0,82	0,89	0,83	1,00	0,81	0,93
Vivienda	0,69	0,63	0,63	0,72	0,69	0,73	0,72	0,81	1,00	0,82
Activos	0,73	0,61	0,74	0,69	0,79	0,85	0,81	0,93	0,82	1,00

Tabla 5.3: Valores obtenidos para el coeficiente de correlación de Spearman

De un primer análisis, podemos apreciar las siguientes similitudes entre los tres coeficientes basados en rangos:

- Todos indican una asociación positiva entre todos los pares de variables, ya que los coeficientes de dependencia en las colas presentan siempre valores superiores a 0.25 y el coeficiente de correlación de Spearman presenta siempre valores positivos.

- Para los tres coeficientes, el nivel de privación en el indicador *Electricidad* es el que presenta, en términos medios, una mayor relación con el resto de las tasas de privación.
- Para los tres coeficientes, el nivel de privación en el indicador *Mortalidad Infantil* es el que presenta, en términos medios, una menor relación con el resto de las tasas de privación.
- Los tres coeficientes apuntan a una elevada relación entre las seis variables pertenecientes a la dimensión *Nivel de vida* (ubicados en la esquina inferior derecha de las tablas). En general, para los tres coeficientes empleados, se aprecia como los elementos de una misma dimensión suelen mostrar relaciones más fuertes entre ellos mismos que con los indicadores pertenecientes a otras dimensiones.

Aunque hemos observado ciertas similitudes en la dependencia medida por las diferentes medidas empleadas, cabe destacar que los tres coeficientes captan aspectos diferentes en la relación que se da entre las variables. Por ejemplo, las parejas *Electricidad-Asistencia escolar* y *Activos-Nutrición* presentan ambos valores de 0.73 para el coeficiente de correlación de Spearman, sin embargo presentan respectivamente valores de 0.67 y 0.59 para el coeficiente de dependencia en la cola derecha y de 0.64 y 0.61 para el coeficiente de dependencia en la cola izquierda. Esto implica que, si bien la nivel de relación de ambas parejas es similar a lo largo de toda la distribución, si nos centramos en las colas hay cambios importantes, lo que da sentido a emplear indicadores que se centren en estas regiones de la distribución.

Si nos centramos en las diferencias que encontramos en los valores obtenidos para los coeficientes de dependencia en las colas derecha e izquierda, se observa que, en general, los primeros son mayores que los segundos. Por término medio, el valor de los coeficientes de dependencia en la cola izquierda es de 0.58, mientras que para los coeficientes de dependencia en la cola derecha es de 0.66. Se trata de una diferencia considerable la cual indica que la dependencia entre variables es mayor en las colas asociadas a los países más pobres que en aquellas asociadas a los menos pobres.

Este hecho puede interpretarse como que la dependencia entre las variables es más fuerte en situación de pobreza extrema. Es decir, ser pobre en una dimensión aumenta

significativamente la probabilidad de ser pobre en otras dimensiones. Esta observación sugiere que las privaciones están más interrelacionadas en situaciones de mayor pobreza, lo cual hace que la pobreza sea más persistente y difícil de superar, haciendo necesarias enfoques integrales.

Por el contrario, la menor dependencia en la cola izquierda indica que para aquellos países que ya gozan de una mejor situación, la dependencia entre las distintas componentes de la pobreza se vuelve más suave, siendo las variables más independientes entre sí. Esto puede significar que las mejoras en un área no se traducen en la misma medida en que lo hacen en países más pobres multidimensionalmente en mejoras en todas las áreas cuando el nivel ya es elevado, subrayando la necesidad de políticas específicas y multidimensionales para combatir la pobreza en aquellos países que ya gocen de una buena situación en una dimensión concreta.

## 6. RESULTADOS

A continuación, analizaremos las redes generadas empleando las tres medidas de correlación propuestas. En primer lugar, generaremos la red empleando los pesos calculados a partir del indicador de dependencia en la cola derecha. Esta red nos permitirá analizar la dependencia de las variables en los países más pobres. A continuación, emplearemos el coeficiente de dependencia en la cola izquierda, con el fin de analizar la dependencia en los países menos azotados por la pobreza. Por último, emplearemos el coeficiente de correlación de Spearman, lo que nos dará una visión global de la dependencia a lo largo de toda la distribución.

Para la generación de las redes no utilizaremos directamente los coeficientes mostrados en las Tablas 5.1, 5.2 y 5.3 como pesos, sino que emplearemos la siguiente fórmula para calcularlos:

$$w_{uv} = \sqrt{1 - coef_{uv}}$$

siendo  $coef_{uv}$  una de las tres medidas de dependencia basada en rangos entre las variables  $u$  y  $v$ . De esta forma, un peso  $w_{uv}$  bajo se deberá interpretar como una alta relación como una alta relación entre las variables  $u$  y  $v$ . Al buscar minimizar la suma de los pesos de la red al generar la MST lo que estaremos haciendo es quedarnos únicamente con las conexiones más intensas entre variables.

Ciertos autores optan por elevar  $coef_{uv}$  al cuadrado, especialmente cuando se trabaja con indicadores que pueden tomar valores negativos como lo es el coeficiente de correlación de Spearman. En nuestro caso no es necesario, ya que como se menciona anteriormente el coeficiente de correlación de Spearman con nuestros datos siempre toma valores positivos. Tras probar ambas opciones se ha comprobado que sin elevar el coeficiente al cuadrado la interpretación de las redes resultantes resultaba más directa y sencilla.

### 6.1. Red generada empleando el coeficiente de dependencia en la cola derecha

Comenzamos por la red generada a partir de las distancias calculadas empleando el coeficiente de dependencia en la cola derecha entre pares de variables. Recordemos que dicho coeficiente para dos indicadores del MPI mide la proporción de países que se sitúan entre el 25% de los países con mayor nivel de privación en un indicador dado que dichos países se encuentren también entre el 25% de los países con mayor privación en el otro indicador. Por lo tanto, la red que se genera nos servirá para medir el nivel de relación entre las variables para los países con un nivel de privación más elevado. Esta será por tanto la red más importante de las que generemos, al poner el foco en los países más pobres.

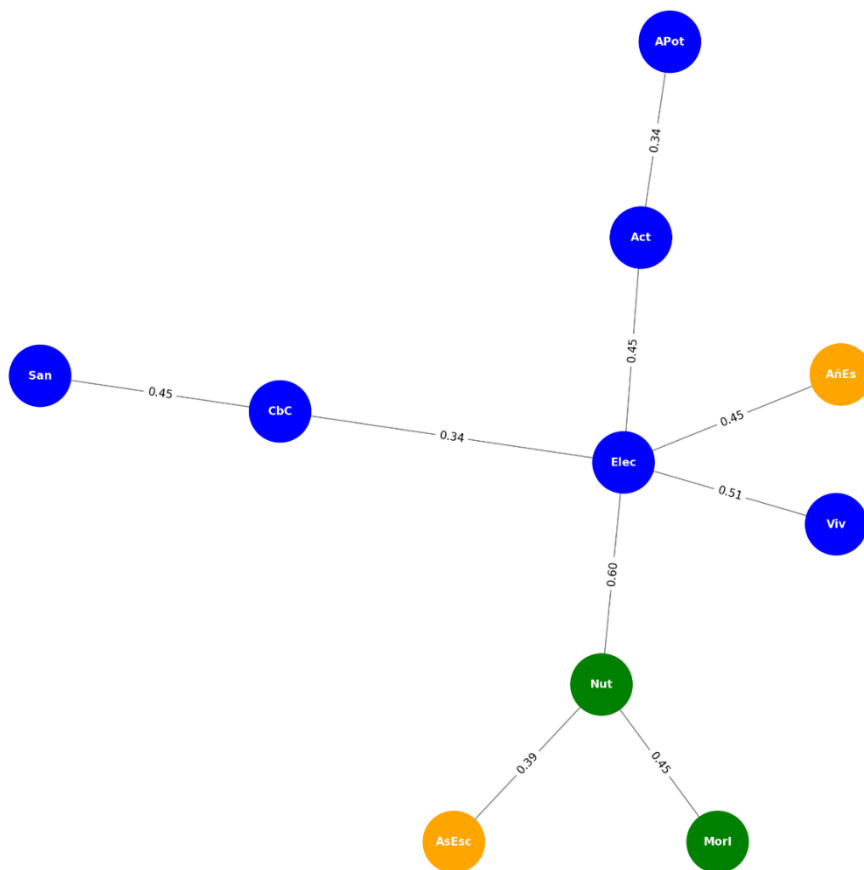


Figura 6.1: Red MST para los coeficientes de dependencia de cola derecha

La Figura 6.1 muestra la red MST generada a partir de los coeficientes de dependencia en la cola derecha. En cuanto a las métricas generales de esta red, la suma de todos los pesos de las aristas es de 3.98, mientras que la media de los pesos de las aristas es de 0.442. El camino más corto promedio es de 1.15. Estas cifras serán de especial utilidad

a efectos comparativos, especialmente con la red generada a partir de los coeficientes de dependencia en la cola izquierda.

Se aprecia como el elemento central en dicha red es el nivel de privación en el indicador *Electricidad*, siendo *Nutrición* el siguiente nodo por importancia en cuanto a número de conexiones. Se aprecia como los dos nodos de la dimensión Salud, que son *Nutrición* y *Mortalidad Infantil*, presentan una conexión directa. Sin embargo, *Años de escolaridad* y *Asistencia escolar*, los cuales componen la dimensión *Educación*, se encuentran distantes en la red. En cuanto a la dimensión del *Nivel de Vida*, observamos que los 6 nodos que la componen están directamente conectados.

En la Tabla 6.1 se recogen los valores de las métricas de centralidad para cada uno de los nodos de la red. Se aprecia, como anticipábamos al visualizar la red, que la variable principal de la red es la tasa de pobreza en *Electricidad*. Esta privación obtiene el valor más elevado en todas las métricas de centralidad, con una gran diferencia respecto al resto. Por tanto, la tasa de privación en *Electricidad* es la que tiene el mayor número de conexiones con otras privaciones (mayor Grado de centralidad), la que está más asociada en promedio otras privaciones (mayor Centralidad de cercanía) y la que se encuentra más veces en los caminos más cortos entre dos privaciones (mayor Centralidad de cercanía). Es por ello que la tasa de privación asociada al indicador *Electricidad* puede ser considerada como la privación más importante en la red generada con los datos de los países con mayores niveles de pobreza en cada indicador MPI.

	Grado de centralidad	Centralidad de cercanía	Centralidad de intermediación
<b>Nutrición</b>	0,33	1,06	0,42
<b>Mortalidad Infantil</b>	0,11	0,72	0,00
<b>Años de escolaridad</b>	0,11	0,87	0,00
<b>Asistencia escolar</b>	0,11	0,74	0,00
<b>Combustible para cocinar</b>	0,22	1,01	0,22
<b>Saneamiento</b>	0,11	0,71	0,00
<b>Agua potable</b>	0,11	0,73	0,00
<b>Electricidad</b>	0,56	1,43	0,86
<b>Vivienda</b>	0,11	0,85	0,00
<b>Activos</b>	0,22	1,01	0,22

Tabla 6.1: Valores para los indicadores de centralidad de la red utilizando el coeficiente de dependencia en la cola derecha

## 6.2. Red generada empleando el coeficiente de dependencia en la cola izquierda

La Figura 6.2 muestra la red generada empleando el indicador de dependencia de cola izquierda, mientras que la Tabla 6.2 recoge los principales indicadores de centralidad asociados a cada variable de dicha red. Es importante tener en cuenta que este coeficiente nos permite estudiar las relaciones de las variables poniendo el foco en los países con menor nivel de pobreza para cada indicador del MPI

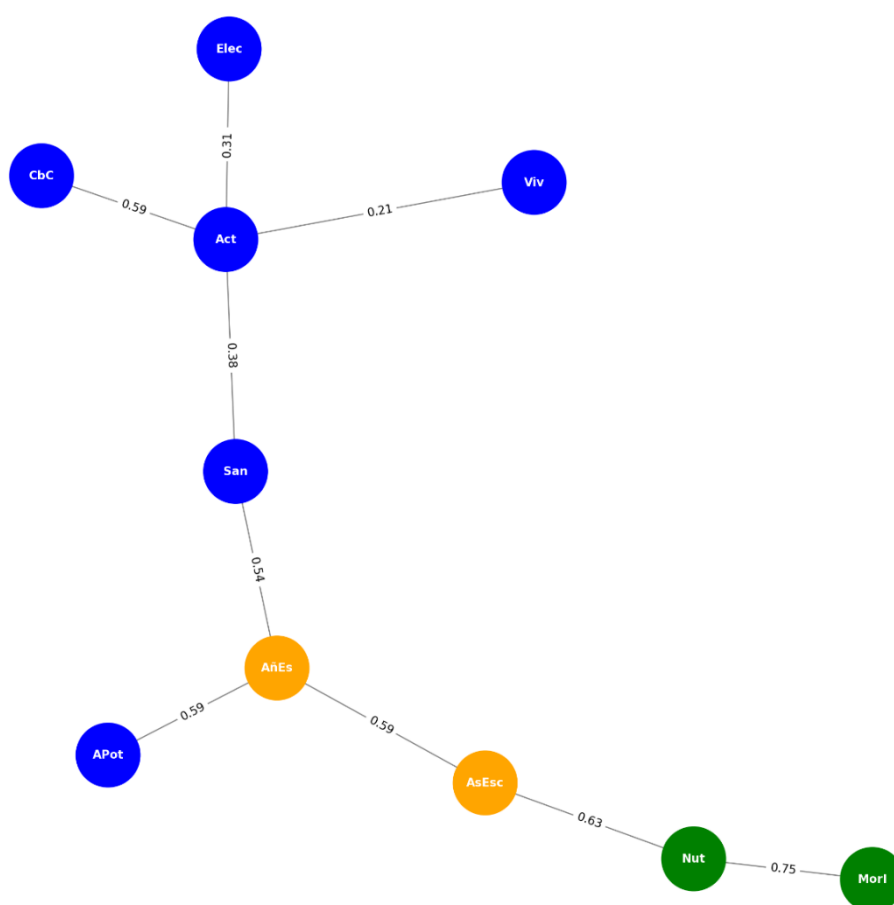


Figura 6.2: Red MST para los coeficientes de dependencia de cola izquierda

La suma de todos los pesos de las conexiones de la red es de 4.59, mientras que la media es de 0.51. Vemos como se trata de valores superiores a los presentados en la red calculada con los indicadores de dependencia en la cola derecha (3.98 y 0.442 respectivamente). Mayores valores implican mayores distancias, lo que se ha de leer como una relación más baja entre las variables. Esto enfatiza la idea presentada en el apartado anterior de que las variables se conectan de forma más intensa entre los países

más pobres. Esta hipótesis se ve soportada por el hecho de que en esta red la distancia mínima promedio es de 1.59 frente a los 1.15 de la red generada con los coeficientes de dependencia en la cola derecha, esto significa que las variables están, en media, más lejanas entre sí y por lo tanto menos relacionadas en la red generada poniendo el énfasis en los países con menor nivel de privaciones.

Respecto al análisis por dimensiones de la red obtenida, vemos en la Figura 6.2 como todos los indicadores están conectados directamente con al menos un indicador de su dimensión a excepción de *Agua potable*, perteneciente a *Nivel de Vida*. Estos resultados se mantienen en la línea de los obtenidos en la red generada a partir de los coeficientes de dependencia en la cola derecha, es decir, se observa una mayor relación entre variables de la misma dimensión.

La variable con más conexiones de la red es la asociada al indicador *Activos*, la cual cuenta con 4 que se traduce a su vez en presentar el mayor indicador de Grado de centralidad. Sin embargo, no es el primero ni en Centralidad de cercanía ni en Centralidad de intermediación. Esto significa que, pese a ser el nodo con más conexiones, no es el nodo por el que pasan más caminos de mínima distancia (lo es *Años de escolaridad* con una Centralidad de intermediación de 0.64 lo que significa que se encuentra en el 64% de estos caminos) ni el nodo que, en media, está más próximo al resto de nodos (lo serán *Saneamiento* y *Años de escolaridad*, al presentar ambos valores de Centralidad de cercanía de 0.87). Por tanto, pese a ser *Activos* el indicador con más conexiones, no podemos decir que el nivel de privación en ningún indicador sea el más importante en cuanto a influencia sobre el nivel de privación en el resto de indicadores ya que no todas las métricas de centralidad coinciden en un mismo elemento de la red.

Pasaremos ahora a comparar la red recogida en la Figura 6.2 y los indicadores de centralidad asociados a esta red recogidos en la Tabla 6.2 con los resultados obtenidos para la red calculada empleando los coeficientes de cola derecha. Llama la atención que el nodo asociado a la tasa de privación en *Electricidad* en el caso anterior era, con una gran diferencia, el nodo más importante y, en este caso, ha pasado a tener una única conexión y valores muy contenidos en todas las métricas de centralidad. Una notable diferencia es que, en este caso, en el que recordemos estamos abordando el estudio de las relaciones entre variables para los países menos pobres, ganan un gran peso las



variables que componen la dimensión *Educación*. Esto se interpreta como que en los países menos pobres la *Educación* tiene un mayor impacto sobre el resto de dimensiones de la pobreza que en los países más pobres.

	Grado de centralidad	Centralidad de cercanía	Centralidad de intermediación
Nutrición	0,22	0,54	0,22
Mortalidad Infantil	0,11	0,41	0,00
Años de escolaridad	0,33	0,87	0,64
Asistencia escolar	0,22	0,70	0,39
Combustible para cocinar	0,11	0,56	0,00
Saneamiento	0,22	0,87	0,56
Agua potable	0,11	0,59	0,00
Electricidad	0,11	0,65	0,00
Vivienda	0,11	0,61	0,00
Activos	0,44	0,80	0,58

Tabla 6.2: Valores para los indicadores de centralidad de la red utilizando el coeficiente de dependencia en la cola izquierda

### 6.3. Red generada empleando el coeficiente de correlación de Spearman

En este apartado se presenta en la Figura 6.3 la red calculada empleando como pesos el coeficiente de correlación de Spearman. La Figura 6.3 recoge las medidas de centralidad asociadas a dicha red.

No tiene sentido hablar en este caso de las métricas generales de la red ya que no resultarían directamente comparables con las obtenidas en los casos anteriores, al estar estas íntimamente relacionadas con los pesos de las aristas y depender estos en este caso directamente del valor del coeficiente de correlación de Spearman, cuyo valor no resulta directamente comparable con los coeficientes de dependencia en las colas que hemos empleado para la generación de las dos redes anteriores.

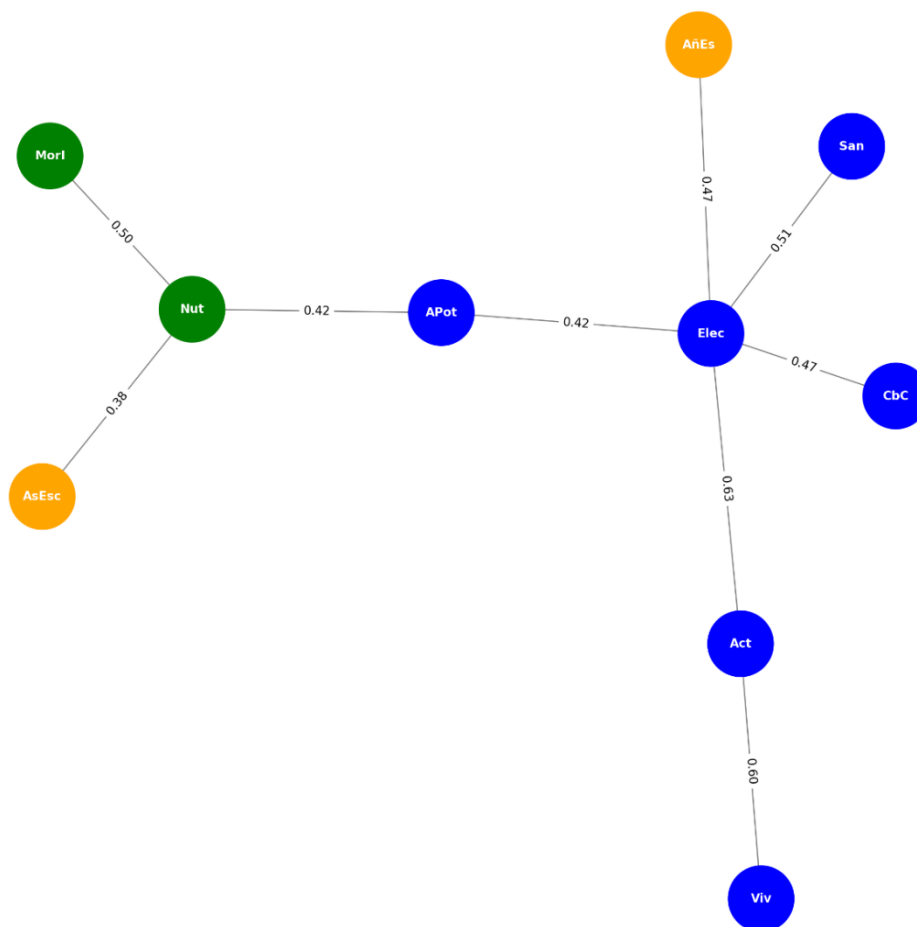


Figura 6.3: Red MST para los coeficientes de correlación de Spearman

Llama la atención como el análisis por dimensiones de la red representada en la Figura 6.3 es idéntico al realizado en el caso de la red calculada a partir de los coeficientes de dependencia en la cola derecha: todos los nodos mantienen al menos una conexión con otro nodo de su dimensión a excepción de los dos asociados a la dimensión *Educación*, que se encuentran alejados en la red. Esto sugiere una vez más una mayor relación entre variables de la misma dimensión, como ya hemos visto en los dos casos anteriores, hecho que ahora se confirma para el conjunto de la distribución.

*Electricidad* es, claramente, el nodo que presenta un valor más elevado para todas las métricas de centralidad. Este hecho es muy similar al observado en la primera red analizada donde *Electricidad* era también la variable principal. Por el contrario, esto es muy diferente al caso de la red calculada con los coeficientes de dependencia izquierda, donde el nodo *Electricidad* tiene un rol secundario en cuanto a centralidad. Esto parece

indicar que *Electricidad* guarda mucha relación con el resto de variables a lo largo de toda la distribución en todos los países menos en aquellos que presentan menores niveles de privación.

	Grado de centralidad	Centralidad de cercanía	Centralidad de intermediación
<b>Nutrición</b>	0,33	0,99	0,42
<b>Mortalidad Infantil</b>	0,11	0,67	0,00
<b>Años de escolaridad</b>	0,11	0,88	0,00
<b>Asistencia escolar</b>	0,11	0,71	0,00
<b>Combustible para cocinar</b>	0,11	0,93	0,00
<b>Saneamiento</b>	0,11	1,00	0,00
<b>Agua potable</b>	0,22	1,26	0,50
<b>Electricidad</b>	0,56	1,43	0,81
<b>Vivienda</b>	0,11	0,80	0,00
<b>Activos</b>	0,22	1,14	0,22

Tabla 6.3: Valores para los indicadores de centralidad de la red utilizando el coeficiente de correlación de Spearman

## 7. CONCLUSIONES

La pobreza es un fenómeno multidimensional que no sólo depende de la falta de ingresos, sino también de la carencia de salud o del no acceso a la educación. Uno de los primeros intentos de medir la pobreza de manera multidimensional ha sido el Índice de Pobreza Multidimensional (MPI). Diseñado dentro del Programa de Naciones Unidas para el Desarrollo (PNUD), este indicador capta las privaciones en 10 indicadores agrupados en 3 dimensiones fundamentales: *Educación, Salud y Nivel de vida*. A pesar de la relevancia del MPI, este índice no es suficientemente sensible a las relaciones entre las variables que lo definen. Esta limitación es crucial, ya que el análisis de la pobreza desde un enfoque multidimensional requiere estudiar cómo interactúan y se refuerzan mutuamente las distintas privaciones. La comprensión de estas interacciones puede proporcionar una visión más completa de la pobreza, permitiendo el desarrollo de políticas más efectivas.

En este contexto, este trabajo está dedicado a analizar cómo interactúan las tasas de privación en los 10 indicadores que componen el Índice de Pobreza Multidimensional (MPI) en 110 países en vías de desarrollo mediante el estudio de las relaciones bivariantes que se establecen entre dichas tasas. Para ello, se han generado tres grafos no dirigidos y ponderados con el objetivo de representar estas relaciones. Los pesos empleados para generar estos grafos se obtuvieron a partir de tres medidas de dependencia basadas en rangos: el coeficiente de correlación de Spearman, una medida de dependencia de la cola inferior y una medida de dependencia de la cola superior. Posteriormente, para cada uno de estos tres grafos, se ha generado el subgrafo de *Minimum Spanning Tree* con el fin de simplificar su representación y hacer énfasis en las conexiones más fuertes del grafo original. Estos subgrafos permiten, por un lado, visualizar las relaciones que se dan entre las diferentes privaciones y, por otro, identificar los indicadores clave que están asociados de forma importante con la presencia o la ausencia de otras privaciones.

Las principales conclusiones extraídas del presente estudio son las siguientes:

1. Independientemente del coeficiente de dependencia empleado, se observa una asociación positiva entre las tasas de privación de todos los indicadores del MPI.

Esto implica que, en los países analizados, hay una tendencia a que estas tasas de privación tienden a ser altas y bajas de forma simultánea.

2. En las redes generadas con los coeficientes de dependencia en las colas, se aprecia una relación más fuerte entre las tasas de privación de los indicadores del MPI en los países con niveles de privación más elevados que en los países con niveles de privación más bajos.
3. Se aprecia, generalmente, una relación más fuerte entre los niveles de privación en indicadores correspondientes a una misma dimensión que entre los niveles de privación en indicadores correspondientes a diferentes dimensiones.
4. Se observan grandes diferencias en cuanto a los elementos más importantes en las redes generadas a partir de los coeficientes de dependencia en las colas.
  - a. En la red generada a partir de los coeficientes de dependencia en la cola derecha (los países con mayores niveles de pobreza en cada indicador), el nivel de privación en Electricidad es la variable más importante de acuerdo con las tres métricas de centralidad empleadas. Sin embargo, para la red generada a partir de los coeficientes de dependencia en la cola izquierda, el nodo *Electricidad* cuenta únicamente con una conexión y presenta valores bajos en los indicadores de centralidad. A la vista de esto, podemos concluir que, el nivel de privación en el indicador del MPI *Electricidad* tendrá una elevada relación con el resto de variables en los países más pobres, sin embargo, su relación con el nivel de privación en el resto de indicadores será más débil en aquellos países menos pobres.
  - b. En la red generada a partir de los coeficientes de dependencia en la cola izquierda (los países con menores niveles de pobreza en cada indicador), existe mayor ambigüedad a la hora de determinar cuál es el elemento principal de la red. No hay un elemento que destaque en todas las métricas de centralidad. En cuanto a nivel de conexiones el elemento principal de dicha red es Activos mientras que las otras dos métricas de centralidad empleadas determinan que el elemento principal de dicha red es Años de escolaridad.
5. En la red generada empleando como pesos el coeficiente de correlación de Spearman, no se advierten muchas diferencias con los resultados obtenidos al

emplear como pesos los coeficientes de dependencia en la cola derecha. Esto indica que, si bien el nivel de relación entre las variables es más fuerte en los países más pobres, el resto de características de dichas relaciones en los países más pobres son similares a las características de las relaciones entre variables en la distribución en su conjunto.

Conviene destacar que, si bien el método empleado busca resaltar las relaciones entre los diferentes factores que componen la pobreza, solo permite analizar las relaciones entre variables dos a dos. Esto implica que puede haber interacciones en las que actúen más de dos variables simultáneamente, cuyo efecto no sea perceptible con el método utilizado. A la vista de estos resultados, se evidencia la necesidad de ahondar en esta línea de estudio, empleando otros indicadores de pobreza, otras métricas de dependencia y otras herramientas analíticas complementarias a los grafos, de forma que se pueda entender mejor la relación entre las diferentes facetas de la pobreza, haciendo énfasis en como varían estas relaciones en función de la situación de mayor o menor pobreza de un país.

## 8. BIBLIOGRAFÍA

- César García Gómez, Ana Pérez, y Prieto-Alaiz, Mercedes. «Copula-based analysis of multivariate dependence patterns between dimensions of poverty in Europe.» *Review of Income and Wealth* 67 (3 2021): 165-195.
- Wang, Gang-Jin, y Chi Xie. «Tail dependence structure of the foreign exchange market: A network view.» *Expert Systems with Applications*, 46, 2016: 164-179.
- Wen, F, X Yang, y W-X Zhou. «Tail dependence networks of global stock markets.» *Int J Fin Econ.*, 2019: 558-567.
- Qiao, Tongshuai, y Liyan Han. «COVID-19 and tail risk contagion across commodity futures markets.» *The Journal of Futures Markets*, 43, 2023: 242-272.
- Sabina Alkire, Usha Kanagaratnam and Nicolai Suppa. «The Global Multidimensional Poverty Index (MPI): 2018 Revision.» Oxford, 2018.
- Sabina Alkire, James Foster. «Counting and multidimensional poverty measurement», *Journal of Public Economics*, 95, 2011, 476-487
- Biggs, Norman L., E. Keith Lloyd, y Robin J. Wilson. *Graph Theory 1736-1936*. Oxford University Press, 1986.
- G Hart, Michale, R.J.F., Ypma, Romero-García, Rafael, Price, Stephen John. «Graph theory analysis of complex brain networks: New concepts in brain mapping applied to neurosurgery», *Journal of Neurosurgery*, 124, 2016.
- Sundkar, R. «How Search Engines use Graph Concepts»  
<https://medium.com/@rahul.sundkar/how-search-engines-use-graph-concepts-321c8f3ffd4e#:~:text=Graphs%20are%20utilized%20to%20represent,%2C%20interests%2C%20and%20social%20relationships.> (2023, May 2) Último acceso: Julio 2024.
- Sweeting, P., & Fotiou, F. (2013). Calculating and communicating tail association and the risk of extreme loss. *British Actuarial Journal*, 18(1), 13–72.  
<https://doi.org/10.1017/S1357321712000347>
- Escuder Vallés, R, y J Santiago Murgui. «Estadística Aplicada». Tirant lo Blanch libros, 2011
- Sibuya, M. «Bivariate extreme statistics.» *Annals of the Institute of Statistical Mathematics*, 195-210. 1960.
- Venter, Gary G. «Tails of Copulas.» 2001.
- Singh, A. (. «Difference Between Prims and Kruskal Algorithm. » (16 de febrero de 2024) Shiksha Online.
- OPHI. Global MPI 2023. 2023. <https://ophi.org.uk/global-mpi/2023>.

## ANEXO 1: Código Python empleado para el tratamiento de datos

```
import numpy as np
import pandas as pd
from scipy.stats import rankdata
from scipy.stats import spearmanr
import matplotlib.pyplot as plt
from networkx import Graph, spring_layout, draw, draw_networkx_edge_labels,
get_edge_attributes, minimum_spanning_tree
import networkx as nx
import seaborn as sns

datos_excel = pd.DataFrame()

archivo_excel = r"C:\Users\raulg\Desktop\Sexto\TFG ADE\table-1-national-results-mpi-
2023_2.xlsx" # Reemplaza "nombre_del_archivo.xlsx" con la ruta de tu archivo Excel
datos_excel = pd.read_excel(archivo_excel, sheet_name= "FinalData")

# Convertir cada valor en 100 menos el mismo
df_transformed = 100 - datos_excel.iloc[:, 1:]

# Función para calcular el rango normalizado de cada columna
def normalize_ranks(df_transformed):
    ranked_df = df_transformed.rank(method='min')
    normalized_ranks = (ranked_df - 1) / (len(df_transformed) - 1)
    return normalized_ranks

# Aplicar la función a cada columna
df_normalized_ranks = normalize_ranks(df_transformed)

sns.pairplot(df_normalized_ranks)
plt.show()
```



```

# Convertir los datos en una matriz
matrizOriginal = datos_excel.iloc[:, 1:].values

matrizOriginal = np.ones_like(matrizOriginal,dtype = float)*100 -matrizOriginal

filas = matrizOriginal.shape[0]
columnas = matrizOriginal.shape[1]

matrizOrdenada = np.zeros_like(matrizOriginal, dtype=float)

percentiles = np.zeros_like(matrizOriginal, dtype=float)
percentilesBinLT = np.zeros_like(matrizOriginal, dtype=float)
percentilesBinUT = np.zeros_like(matrizOriginal, dtype=float)

for col in range(columnas):
    percentil25 = np.percentile(matrizOriginal[:,col],25)
    percentil75 = np.percentile(matrizOriginal[:,col],75)
    matrizOrdenada[:,col] = rankdata(matrizOriginal[:,col])

maximo = np.max(matrizOrdenada[:,1])
matrizOrdenada = matrizOrdenada/(filas)

#percentiles = matrizOrdenada/(filas)

percentilesBinLT = matrizOrdenada <= 0.25
percentilesBinUT = matrizOrdenada >= 0.75

LTCoeff = np.zeros((columnas,columnas))
UTCoeff = np.zeros((columnas,columnas))

```

```
SPCcoeff = np.zeros((columnas,columnas))
```

```
SPPValor = np.zeros((columnas,columnas))
```

```
coeficienteIndiv = spearmanr(matrizOriginal[:,0],matrizOriginal[:,1])
```

```
for i in range(0,columnas):
```

```
    for j in range (i,columnas):
```

```
        auxL = percentilesBinLT[:,i]*percentilesBinLT[:,j]
```

```
        auxU = percentilesBinUT[:,i]*percentilesBinUT[:,j]
```

```
        LTCoeff[j,i] = np.count_nonzero(auxL)/(filas)
```

```
        LTCoeff[i,j] = np.count_nonzero(auxL)/(filas)
```

```
        UTCoeff[j,i] = np.count_nonzero(auxU)/(filas)
```

```
        UTCoeff[i,j] = np.count_nonzero(auxU)/(filas)
```

```
        SPCoeff[i,j],SPPvalor = spearmanr(matrizOriginal[:,i],matrizOriginal[:,j])
```

```
        SPCoeff[j,i],SPPvalor = spearmanr(matrizOriginal[:,i],matrizOriginal[:,j])
```

```
LTCoeff = LTCoeff/LTCoeff[0,0]
```

```
UTCoeff = UTCoeff/UTCoeff[0,0]
```

```
LTCoeff = (np.ones((columnas,columnas))-LTCoeff)**(1/2)
```

```
UTCoeff = (np.ones((columnas,columnas))-UTCoeff)**(1/2)
```

```
SP = SPCoeff
```

```
SPCcoeff = (np.ones((columnas,columnas))-SPCcoeff)**(1/2)
```

```
G = Graph()
```

```
for i in range(len(SPCoeff)):
```

```
    G.add_node(i)
```

```

for i in range(len(SPCoeff)):
    for j in range(i+1, len(SPCoeff)):
        if SPCoeff[i][j] != 0:
            G.add_edge(i, j, weight=SPCoeff[i][j])

mst = minimum_spanning_tree(G)

# Mapear nombres originales de los nodos a nombres deseados
nombres_nodos = {0: "Nutrition", 1: "Child mortality", 2: "Years of school", 3: "School
attendance", 4: "Cooking fuel", 5: "Sanitation", 6: "Drinking water", 7: "Electricity", 8:
"Housing", 9: "Assets"}

# Dibujar el grafo
node_positions = spring_layout(mst) # Asignar posiciones para los nodos

plt.figure(figsize=(100, 100))
pos = nx.spring_layout(mst, scale = 100)
nx.draw_networkx(mst, pos, labels = nombres_nodos, with_labels=True, node_size=500000,
node_color='skyblue', font_size=90, font_weight='bold',width=0, edge_color='gray')
pos = nx.spring_layout(G, scale=100) # Ajustar el parámetro 'scale'
#nx.draw_networkx(G, pos, labels=nombres_nodos, with_labels=True, node_size=500000,
node_color='skyblue', font_size=90, font_weight='bold', width=2, edge_color='gray')

# Añadir etiquetas de peso de los bordes
labels = {(u, v): f"{w:.2f}" for u, v, w in mst.edges.data('weight')}
nx.draw_networkx_edge_labels(mst, pos, edge_labels=labels, font_size = 90)

plt.title("PMFG")
plt.show()
plt.savefig('grafo.jpg', bbox_inches='tight')

```

```
G = mst
```

```
average_shortest_path_length = nx.average_shortest_path_length(G, weight='weight')
```

```
print(f"Average Shortest Path Length: {average_shortest_path_length}")
```

```
# Calcular Degree Centrality
```

```
degree_centrality = nx.degree_centrality(G)
```

```
print("Degree Centrality:", degree_centrality)
```

```
# Calcular Closeness Centrality
```

```
closeness_centrality = nx.closeness_centrality(G, distance='weight')
```

```
print("Closeness Centrality:", closeness_centrality)
```

```
# Calcular Betweenness Centrality
```

```
betweenness_centrality = nx.betweenness_centrality(G, weight='weight', normalized=True)
```

```
print("Betweenness Centrality:", betweenness_centrality)
```

```
# Crear una tabla con los resultados
```

```
influence_metrics = pd.DataFrame({  
    'Degree Centrality': degree_centrality,  
    'Closeness Centrality': closeness_centrality,  
    'Betweenness Centrality': betweenness_centrality  
})
```

```
print(influence_metrics)
```

```
matrix = influence_metrics.values
```

## **ANEXO 2: Software empleado**

Para la realización de este análisis, se utilizó el lenguaje de programación Python, en combinación con el entorno de desarrollo integrado (IDE) Spyder, y varias bibliotecas específicas para el manejo y análisis de datos.

### **Anexo 2.1. Python**

Python es un lenguaje de programación de alto nivel, ampliamente utilizado en diversas disciplinas científicas y de ingeniería debido a su simplicidad y versatilidad. Es conocido por su sintaxis clara y su capacidad para soportar múltiples paradigmas de programación, incluyendo la programación orientada a objetos, la programación funcional y la programación procedural. Python es una herramienta poderosa en el análisis de datos y la ciencia de datos, gracias a la vasta cantidad de bibliotecas disponibles que facilitan tareas complejas.

### **Anexo 2.2. Spyder**

Spyder, cuyo nombre proviene de "Scientific PYthon Development EnviRonment", es un entorno de desarrollo integrado (IDE) diseñado específicamente para científicos e ingenieros. Proporciona una interfaz amigable para escribir, depurar y ejecutar código en Python, y viene con herramientas integradas como un editor de texto avanzado, un depurador interactivo y un explorador de variables. Además, Spyder permite integrar y manejar bibliotecas de Python de manera eficiente, lo que lo convierte en una opción ideal para proyectos de análisis de datos.

### **Anexo 2.3. Bibliotecas de Python**

En programación, una biblioteca es un conjunto de módulos que contienen funciones y métodos que pueden ser reutilizados en diferentes programas. Las bibliotecas permiten a los desarrolladores ahorrar tiempo al proporcionar soluciones ya implementadas para tareas comunes. En este proyecto, se utilizaron las siguientes bibliotecas de Python:

NumPy: Es una biblioteca fundamental para la computación científica en Python. Proporciona soporte para arrays y matrices multidimensionales, junto con una colección de funciones matemáticas de alto nivel para operar con estos arrays de manera

eficiente. En este análisis, NumPy se utilizó para manejar y transformar matrices de datos.

Pandas: Esta biblioteca es esencial para la manipulación y análisis de datos estructurados. Ofrece estructuras de datos flexibles y potentes, como DataFrames, que permiten realizar operaciones complejas de manera sencilla y eficiente. Pandas se empleó para leer el archivo Excel y convertir los datos en matrices manipulables.

SciPy: SciPy es una biblioteca que se basa en NumPy y proporciona herramientas adicionales para la computación científica y técnica. En este trabajo, se utilizó específicamente el módulo `scipy.stats` para calcular estadísticas y correlaciones, como los coeficientes de dependencia y la correlación de Spearman.

Matplotlib: Esta biblioteca se utiliza para la generación de gráficos y visualizaciones de datos. Proporciona una API de trazado flexible y extensa que facilita la creación de gráficos estáticos, animados e interactivos. En el proyecto, Matplotlib se utilizó para dibujar y guardar los gráficos de los árboles de expansión mínima.

NetworkX: NetworkX es una biblioteca para la creación, manipulación y estudio de la estructura, dinámica y funciones de redes complejas. Ofrece herramientas para trabajar con grafos y redes, incluyendo algoritmos para calcular árboles de expansión mínima. Esta biblioteca se utilizó para generar y visualizar los grafos basados en las matrices de distancia calculadas.

Estas herramientas y bibliotecas permitieron realizar un análisis exhaustivo y visualmente comprensible de los datos de los indicadores multidimensionales, facilitando la interpretación de las relaciones y dependencias entre diferentes variables.