

Trabajo Fin de Máster
Máster en Organización Industrial y Gestión
Empresarial

Aplicaciones de técnicas de análisis de redes
complejas a redes de dominación en eficiencia

Autora: Laura Calzada Infante

Tutor: Sebastián Lozano Segura

Dep. Organización Industrial y Gestión de Empresas I
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2016



Trabajo Fin de Máster
Máster en Organización Industrial y Gestión de Empresas

Aplicaciones de técnicas de análisis de redes complejas a redes de dominación en eficiencia

Autora:

Laura Calzada Infante

Tutor:

Sebastián Lozano Segura

Dep. Organización Industrial y Gestión de Empresas I

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2016

Trabajo Fin de Máster: Aplicaciones de técnicas de análisis de redes complejas a redes de dominación en eficiencia

Autora: Laura Calzada Infante

Tutor: Sebastián Lozano Segura

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2016

El Secretario del Tribunal

A mi familia

Agradecimientos

En primer lugar me gustaría agradecer a mi tutor de Trabajo Fin de Máster, Sebastián Lozano, por toda su ayuda, comprensión y paciencia. Desde el momento en que le conocí, no ha dejado de sorprenderme y de romperme todos los esquemas, con un despliegue de asertividad, eficiencia y entusiasmo envidiable, ayudándome a ampliar mi visión de la vida y de la profesión.

Quisiera dar también las gracias a Adenso, por su paciencia y por darme la oportunidad de adentrarme en el mundo de la organización y la eficiencia. La vida está llena de sorpresas y nunca sabes dónde te puede llevar.

Quisiera dar las gracias a Santiago, por sus enseñanzas, sus consejos, su atención y su amistad, no hay día que no me acuerde de nuestras charlas y espero se produzcan muchas más.

Quisiera dar las gracias a mi familia, por estar siempre tan cerca. Gracias a vuestro apoyo y consejos. No hay día que no os dedique una sonrisa.

Gracias a Fernando, por su apoyo y comprensión infinita. Sin ti, no hubiera llegado tan lejos. Gracias por cambiarme el prisma.

Y por último y no menos importante gracias a mis amigos por apoyarme, por los buenos y malos momentos que hemos pasado y por los que están por llegar. Gracias por enseñarme tanto.

Laura Calzada Infante

Sevilla, 2016

Resumen

La técnica de Análisis de Envoltura de Datos es muy conocida y permite comparar unidades productivas homogéneas. Dentro de la literatura se han realizado muchas aplicaciones prácticas. Sin embargo, trabajar con los resultados y visualizarlos es una tarea complicada cuando se trata múltiples entradas y múltiples salidas.

En este trabajo se pretende salvar este obstáculo convirtiendo esos resultados en una red dirigida acíclica en la que los nodos sean DMUs y los arcos establecen una relación de dominancia entre cada par de nodos, de forma que en caso de existir dicha relación entre dos nodos, un nodo señale a otro si este último es más eficiente y es un benchmark para el primero.

Además, gracias a la aplicación de las técnicas del Análisis de Redes Complejas sobre esta red de dominancia se puede estructurar la información, determinar caminos cortos hacia la frontera eficiente, realizar un ranking de las unidades productivas,...

Agradecimientos	i
Resumen	i
Índice	i
Índice de Tablas	i
Índice de Figuras	1
1 Objetivo	1
2 Análisis Por Envoltura De Datos	1
2.1 <i>Conceptos fundamentales</i>	1
2.1.1 Productividad y Eficiencia	2
2.1.2 Tecnología	3
2.2 <i>Modelos</i>	5
2.2.1 Modelos de Retornos de Escala Constante	5
2.2.1.1 Modelo Ratio	5
2.2.1.2 Modelo CCR-INPUT	6
2.2.1.3 Modelo CCR-OUTPUT	9
2.2.2 Modelos con Retornos de Escala Variables	11
2.2.2.1 Modelo BCC-INPUT	11
2.2.2.2 Modelo BCC-OUTPUT	13
2.2.3 Modelos FDH (Free Disposal Hull)	14
2.2.3.1 Modelo FDH con orientación de entrada	14
2.2.3.2 Modelo FDH con orientación de salida	15
2.2.3.3 Modelo Aditivo	16
2.2.3.4 Measure of Inefficiency Proportions	16
2.2.3.5 Range-Adjusted Measure	17
2.3 <i>Medidas de eficiencia</i>	18
2.4 <i>Secuencia de targets intermedios</i>	19
2.4.1 Technical Efficiency Improvement Program y Scale Efficiency Improvement Program	19
2.4.2 Selección de targets intermedios	21
2.4.3 Selección de rutas	25
3 Análisis de Redes Complejas	27
3.1 <i>Tipos de redes y definición de la red</i>	27
3.2 <i>Caracterización de las redes</i>	29
3.3 <i>Modelos de redes</i>	34
3.3.1 Redes aleatorias	34
3.3.2 Redes de mundo pequeño	34
3.3.3 Modelos de configuración	35
3.3.4 Modelo libre de escala	35
4 Metodología de análisis de redes de dominancia	37
4.1 <i>Técnicas que emplean el Análisis de Redes Complejas y Análisis de Envoltura de Datos</i>	37
4.1.1 Análisis de redes de colaboración	37
4.1.2 Evaluación de nodos eficientes usando CNA	38
4.2 <i>Redes de dominancia</i>	39

4.2.1	Construcción de la red y definición de conjuntos	39
4.2.1.1	Conjuntos de nodos eficientes	40
4.2.1.2	Conjuntos de nodos dominados	40
4.2.1.3	Estratificación	40
4.2.2	Indices	41
4.2.2.1	Medidas a nivel de nodo.....	41
4.2.2.2	Medidas a nivel de capa.....	44
4.2.2.3	Medidas a nivel de componente	44
4.2.2.4	Medidas a nivel de red.....	45
4.2.3	Filtros.....	46
5	Ilustración de la metodología.....	49
5.1	<i>Juego de datos de CST.....</i>	<i>49</i>
5.2	<i>Juego de datos de Lim.....</i>	<i>53</i>
5.3	<i>Juego de datos de Park</i>	<i>58</i>
6	Conclusiones	63
	Referencias.....	65

ÍNDICE DE TABLAS

Tabla 3.1. Densidad	29
Tabla 3.2. Grado	30
Tabla 3.3. Fuerza	30
Tabla 3.4. Camino medio	30
Tabla 3.5. Diámetro	31
Tabla 3.6. Eficiencia	31
Tabla 3.7. Coeficiente de Clustering red sin pesos	31
Tabla 3.8. Coeficiente de Clustering red con pesos	32
Tabla 3.9. Transitividad	32
Tabla 3.10. Centralidad según el grado	32
Tabla 3.11. Centralidad de intermediación	33
Tabla 5.1. Índices a nivel de red y de componente en la red CST	50
Tabla 5.2. Distribución de los nodos según su capa en el caso CST	51
Tabla 5.3. Índices a nivel de nodo en el caso CST	52
Tabla 5.4. Índices a nivel de nodos eficientes en el caso CST	53
Tabla 5.5. Datos del caso Lim	53
Tabla 5.6. Índices a nivel de red y de componente en el caso Lim	54
Tabla 5.7. Distribución de los nodos según su capa en el caso Lim	56
Tabla 5.8. Índices a nivel de nodo en el caso Lim	57
Tabla 5.9. Índices a nivel de nodos eficientes en el caso Lim	58
Tabla 5.10. Datos del caso Park	59
Tabla 5.11. Índices a nivel de red y de componente en el caso Park	59
Tabla 5.12. Distribución de los nodos según su capa en el caso Park	60
Tabla 5.13. Índices a nivel de nodo en el caso Park	61
Tabla 5.14. Índices a nivel de nodos eficientes en el caso Park	62

ÍNDICE DE FIGURAS

Figura 2.1. Esquema de una DMU	1
Figura 2.2. Tecnología FDH en un modelo con una entrada y una salida	4
Figura 2.3. Tecnología VRS y CRS en un modelo con una entrada y una salida	5
Figura 2.4 Modelo CCR-Input con una entrada y una salida	8
Figura 2.5: Modelo CCR-Input con dos entradas y una salida	8
Figura 2.6 Modelo CCR-Output con una entrada y una salida	11
Figura 2.7 Modelo CCR-Output con una entrada y dos salidas	11
Figura 2.8 Modelo BCC-Input para el caso de una entrada y una salida	12
Figura 2.9 Comparación del modelo CCR-Input y el modelo BCC-Input	13
Figura 2.10: Modelo FDH con orientación de entrada en el caso de un modelo con dos entradas y una salida	15
Figura 2.11 Modelo Aditivo con tecnología VRS, caso de 2 entradas y 1 salida	19
Figura 2.12: Visualización de los pasos intermedios generados por los modelos TEIP y SEIP en un caso con 6 DMUs con una entrada y una salida	21
Figura 2.13 Identificación de los diferentes niveles de frontera eficiente en un conjunto de datos con dos entradas y una salida constante.	23
Figura 2.14: Objetivos reales propuestos para la DMU L que le permiten llegar a la frontera eficiente.	25
Figura 2.15: Caminos posibles que puede tomar L hasta la frontera eficiente, en el juego de datos de los supermercados	25
Figura 3.1: Tipos principales de redes complejas y sus transformaciones	28
Figura 3.2 Generación de redes de mundos pequeños	35
Figura 3.3: Representación de la variación de la media de la longitud geodésica y el clustering en función de la probabilidad p	35
Figura 4.1 Distribución del PageRank de forma simplificada	43
Figura 5.1 Visualización de la red de dominancia del juego de datos de CST	50
Figura 5.2 Subgrafo de esqueleto en la red CST	50
Figura 5.3 Distribución de las distancias máximas a la frontera eficiente del caso CST	51
Figura 5.4 Distribución de los enlaces del caso CST	51
Figura 5.5 Visualización del grado de entrada y de salida en función de las capas en el caso CST	52
Figura 5.6 Visualización de la red de dominancia en el caso Lim	54
Figura 5.7 Distribución de las distancias máximas a la frontera eficiente en el caso Lim	55
Figura 5.8 Subgrafo de esqueleto en el caso Lim	55
Figura 5.9 Distribución de los enlaces en el caso Lim	56

Figura 5.10 Visualización del grado de entrada y de salida en función de las capas en el caso Lim	57
Figura 5.11 Visualización de la red de dominancia del juego de datos de Park	59
Figura 5.12 Subgrafo de esqueleto en el caso Park	59
Figura 5.13 Distribución de las distancias máximas a la frontera eficiente en el caso Park	60
Figura 5.14 Distribución de los enlaces en el caso Park	60
Figura 5.15 Visualización del grado de entrada y de salida en función de las capas en el caso Park	61

1 OBJETIVO

“Una imagen vale más que mil palabras”

Proverbio chino

El Análisis de Redes Complejas ha tenido una gran aplicación en diferentes ciencias. Se basa en la caracterización de un sistema, entendiendo como sistema, a una serie de entes, llamados nodos, que está relacionados entre sí por enlaces y que simbolizan la interacción existente entre dichos nodos. La caracterización, permite comprender como funciona la agrupación, analizar su estructura para determinar cuáles son los elementos más determinantes en una red, cuáles son los principios que permiten a una red crecer hasta convertirse en una red robusta y eficiente e incluso predecir cuál será el futuro de ese sistema.

Por otro lado, la metodología de Análisis por Envoltura de Datos, permite comparar una serie de unidades, con el fin de determinar la eficiencia relativa entre dichas unidades, considerando como eficiencia el cociente entre producción y recursos. Los múltiples modelos de programación lineal que se han desarrollado permiten comparar todas las unidades entre sí, determinar cuáles son las unidades no eficientes y cuáles son sus objetivos a seguir teniendo en cuenta su tamaño.

El Análisis por Envoltura de Datos es una herramienta muy utilizada; sin embargo, es compleja la visualización de sus resultados cuando se analizan múltiples entradas y/o salidas. Por ello, el objetivo de este trabajo es utilizar las herramientas de caracterización y visualización del Análisis de Redes Complejas, para estudiar estos resultados. Para ello se establecerán como nodos las unidades analizadas, y los enlaces partirán de aquellas unidades que no son eficientes y señalarán a las que son de su mismo tamaño y las dominan por ser más eficientes que ellas. Gracias a estas relaciones de dominancia se obtendrá una red dirigida, debido a que los enlaces o arcos tienen una dirección.

En el capítulo 2 se explicarán todos los conceptos necesarios del Análisis por Envoltura de Datos y en el capítulo 3 los del Análisis de Redes Complejas para poder desarrollar la metodología planteada en el capítulo 4. Finalmente en el capítulo 5 se aplicará la metodología a varios juegos de datos, tras la cual se desarrollarán las conclusiones del presente trabajo.

2 ANÁLISIS POR ENVOLTURA DE DATOS

Las personas debemos el progreso a los insatisfechos.

-Aldous Huxley -

En este apartado se pretende explicar los principales conceptos de la metodología conocida como Envoltura de Análisis de Datos, (Data Envelopment Analysis, DEA). Con el fin de desarrollar la base que será necesaria para comprender la técnica que se desarrollará en este trabajo.

El Análisis de Envoltura de datos tiene como objetivo determinar la eficiencia relativa de las unidades que se están estudiando. Este estudio permitiría analizar cuáles son aquellas unidades que realizan una mejor gestión de sus recursos, y que por ello, se consideran modelos a seguir por las unidades de la muestra que poseen un tamaño similar.

El origen de esta herramienta no paramétrica se remonta a 1978, donde Charnes, Cooper y Rhodes publicaron (Charnes et al. 1978), basándose en el concepto de eficiencia desarrollado por Farrell en 1957. El fin de su investigación era analizar la eficiencia del programa de educación “Follow Through” en escuelas públicas de Estados Unidos.

Farrell planteaba en (Farrell 1957) como se podría aumentar la producción de una empresa, haciéndola más eficiente sin tener que aumentar sus recursos. Desarrolló un método que medía la eficiencia técnica de una empresa comparándola con otra hipotética que usaba la misma proporción de recursos y que había sido generada a partir de la media ponderada de otras dos empresas existentes.

2.1 Conceptos fundamentales

La técnica DEA se puede aplicar a cualquier unidad que realice un proceso productivo que consuma unos recursos (inputs o entradas) y obtenga unos resultados (outputs o salidas). Estas unidades homogéneas se denominan Unidades de Decisión (Decision Making Units, DMUs), debido a que cada unidad decide cómo gestionar sus recursos y es responsable de su productividad, al ser capaz de modificar su proceso productivo.



Figura 2.1. Esquema de una DMU

Como consecuencia de la amplia variedad de factores que afectan a un sistema productivo, se deben analizar cuáles son las salidas que se pretenden medir y cuáles son las entradas que serían determinantes en el proceso productivo y que afectan directamente a esas salidas.

A continuación se muestran las principales variables que se tienen en cuenta en un modelo DEA:

- n observaciones o DMUs con $j=1,\dots,n$, con m entradas y s salidas cada una.

- x_{ij} : Entrada i consumida por la unidad j con $i=1,\dots,m$

Siendo X la matriz de entradas de dimensiones $n \times m$

- y_{kj} : Salida k correspondiente a la unidad j con $k=1,\dots,s$

Siendo Y la matriz de salidas de dimensiones $n \times s$

Para poder ejecutar un modelo DEA, la relación que debe existir entre el número de DMUs y el número de entradas y salidas que se analizan es:

$$n > 3(m + s) \quad (2.1)$$

2.1.1 Productividad y Eficiencia

Según (Farrell 1957) se entiende como productividad la relación entre los resultados obtenidos y los recursos consumidos en el proceso productivo. De forma matemática se expresaría como el ratio entre las unidades producidas y las unidades consumidas.

$$Productividad = Eficiencia absoluta = \frac{Resultados}{Recursos} = \frac{Salidas}{Entradas} \quad (2.2)$$

Esta expresión representa la eficiencia absoluta, dado que muestra la proporción de recursos necesarios teniendo en cuenta únicamente los datos del proceso productivo de la unidad que se está analizando.

Sin embargo, una vez escogido los principales factores que representarían los resultados y los recursos de los procesos productivos, para poder agregar los resultados por una parte y los recursos por otro, se les debe asignar un peso a cada uno, que represente su importancia y permita que el ratio sea adimensional.

$$Productividad = \frac{Suma ponderada de salidas}{Suma ponderada de entradas} \quad (2.3)$$

Siendo u_{ij} y v_{kj} los pesos correspondientes a cada entrada y salida respectivamente. La agregación de las entradas y las salidas de la unidad j se expresaría de la siguiente forma.

$$Entradas = \sum_{i=1}^m u_{ij} x_{ij} \quad (2.4)$$

$$Salidas = \sum_{k=1}^s v_{kj} y_{kj} \quad (2.5)$$

Sin embargo, resulta interesante disponer de un índice que permita evaluar la productividad de una unidad respecto de las demás unidades semejantes. Por ello surgió el concepto de eficiencia relativa, donde se toma como referencia una unidad homogénea. En este documento siempre que se hable de eficiencia, se referirá a la eficiencia relativa.

$$Eficiencia\ relativa_j = \frac{Productividad_j}{Productividad_0} \quad (2.6)$$

Dependiendo de qué unidad se tome como referencia se referirá a diferentes tipos de eficiencia relativa:

- Si se toma como referencia la unidad con la máxima eficiencia absoluta, se utiliza el concepto de *eficiencia global*.
- Se habla de *eficiencia técnica* si se considera como referencia la unidad de tamaño similar con la mayor eficiencia absoluta. Se dicen de dos unidades de tamaño similar cuando ambas tienen el mismo orden de magnitud.
- Se habla de *eficiencia de escala* cuando se evalúa el ratio entre la eficiencia global y la eficiencia técnica de una determinada unidad.

Hay que tener en cuenta que la eficiencia relativa depende de la orientación. Se considera *orientación de entrada*, cuando el objetivo de la unidad que se está analizando es disminuir el número de recursos empleados, sin disminuir la producción y se habla de *orientación de salida* si pretende aumentar los resultados del proceso productivo sin variar los recursos que emplea en él.

Por lo tanto cuando se analice la eficiencia técnica, si se obtiene un ratio de 1, se considera que la unidad es eficiente y por ello formará parte de la *frontera eficiente*, que se compone de las unidades más eficientes de la muestra analizada. Las unidades que no tengan un ratio 1, siempre serán inferiores a la unidad y se denominan unidades no eficientes.

En el caso de la eficiencia de escala, si la unidad analizada tiene un ratio 1, entonces coincide la eficiencia global con la eficiencia técnica. Por lo que, dicha unidad tiene el tamaño de escala más productivo (Most Productive Scale Size, MPSS)

Gracias a que la unidad de referencia va a ser siempre una unidad eficiente, el ratio de eficiencia absoluta será igual a la unidad. Por lo que la expresión de eficiencia se expresa finalmente de la siguiente forma:

$$Eficiencia_j = \frac{\sum_{k=1}^s v_{kj} y_{kj}}{\sum_{i=1}^m u_{ij} x_{ij}} \quad (2.7)$$

2.1.2 Tecnología

El concepto de tecnología hace referencia al conjunto de procesos productivos tecnológicamente factibles que se evalúan en el modelo DEA. Existen cuatro hipótesis que permiten definir la tecnología:

1. *Envoltura*: Las observaciones pertenecen al conjunto de posibilidades de producción (T)

$$\{(x_j, y_j) \in T\} \quad (2.8)$$

2. *Free disposability* o *libre disponibilidad*: Libre para desechar o derrochar. Se considera que una unidad puede usar más recursos de los que en verdad necesita y que puede producir menos de lo que puede llegar a producir.

$$\{(x, y) \in T \forall x \geq x, y \leq y\} \quad (2.9)$$

3. *Hipótesis de convexidad*: Considera que es factible cualquier combinación convexa de las unidades existentes.

$$Si (x_1, y_1) \in T \& (x_2, y_2) \in T \rightarrow \lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2) \in T \quad (2.10)$$

4. *Escalabilidad*: Se puede escalar cualquier proceso productivo perteneciente al conjunto de posibilidades de producción.

$$Si (x_1, y_1) \in T \& (x_2, y_2) \in T \rightarrow (\lambda x, \lambda y) \in T \forall \lambda \geq 0 \quad (2.11)$$

Existen tres tipos de tecnología FDH (Free Disposability Hull), VRS (Variable Return to Scale) y CRS (Constante Return to Scale) y cada una cumple una serie de hipótesis.

- La tecnología FDH cumple las hipótesis de envoltura y libre disponibilidad

- La tecnología VRS cumple las hipótesis de convexidad, envoltura y libre disponibilidad
- La tecnología CRS cumple las hipótesis de escalabilidad, convexidad, envoltura y libre disponibilidad.

La tecnología FDH cumple las propiedades de envoltura y de libre disponibilidad, es decir, está compuesta por las unidades observadas más todos aquellos procesos productivos que consumen más recursos que los existentes o consiguen niveles de producción inferiores. Se define de con la siguiente expresión.

$$T_{FDH} = \left\{ (\vec{x}, \vec{y}) : \exists \vec{\lambda} \geq 0, \sum_{j=1}^n \lambda_j = 1; \vec{\lambda}X \leq \vec{x}; \vec{\lambda}Y \geq \vec{y}; \lambda_j \in \{0,1\} \right\} \quad (2.12)$$

Las unidades únicamente se pueden proyectar sobre una DMU existente, debido a que la variable λ es binaria y el sumatorio debe ser 1. En la Figura 2.2 que se muestra a continuación, se aprecia la proyección de una DMU ineficiente sobre una DMU que forma parte de la frontera eficiente. Cabe destacar que los segmentos que unen las DMUs eficientes, no forman parte de la frontera eficiente.

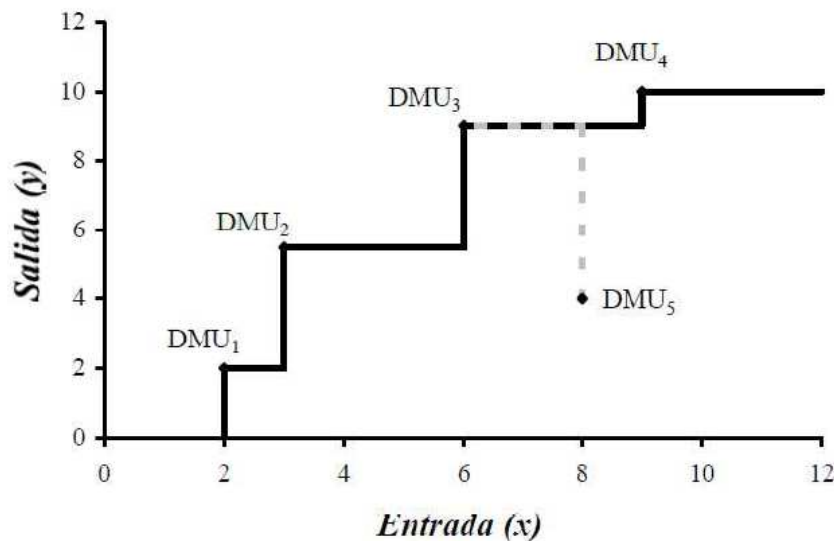


Figura 2.2. Tecnología FDH en un modelo con una entrada y una salida

Fuente: (Fernández 2015)

La tecnología VRS permite la combinación convexa de las unidades existentes, como se puede visualizar en la Figura 2.3. Por lo tanto todas las unidades factibles se encuentran por debajo de la línea continua. La frontera eficiente en esta compuesta por las unidades eficientes y los segmentos que las unen, sin embargo los tramos paralelos a los ejes únicamente forman parte de la frontera de producción admisible.

En esta tecnología se puede apreciar tres tipos de rendimiento de escala diferentes:

- Retornos de escala crecientes (Increasing Return Scale, IRS): se aprecia cuando el incremento porcentual de outputs es mucho mayor que el incremento porcentual de los inputs. En la Figura 2.3 se correspondería con el tramo AB.
- Retornos de escala constantes (Constant Return Scale, CRS): el incremento porcentual de outputs es igual al incremento porcentual de inputs. Las unidades que se encuentran en este tramo de la frontera poseen el tamaño de escala más productivo (MPSS). En la Figura 2.3 se correspondería con los tramos BC y CD.
- Retornos de escala decrecientes (Decreasing Return Scale, DRS): el incremento porcentual de outputs es mucho menor que el incremento porcentual de los inputs.

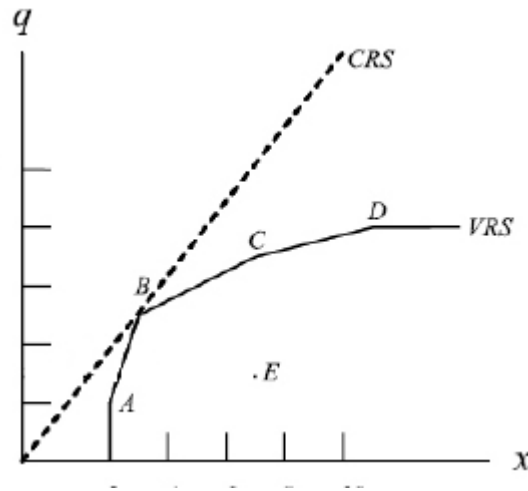


Figura 2.3. Tecnología VRS y CRS en un modelo con una entrada y una salida

Fuente: (Fernández 2015)

La tecnología VRS, se expresa matemáticamente con la siguiente expresión:

$$T_{VRS} = \left\{ (\vec{x}, \vec{y}) : \exists \vec{\lambda} \geq 0, \sum_{j=1}^n \lambda_j = 1; \vec{\lambda}X \leq \vec{x}; \vec{\lambda}Y \geq \vec{y} \right\} \quad (2.13)$$

En la frontera eficiente de la tecnología CRS, se encuentran las unidades con la máxima productividad posible al utilizarse como referencia la eficiencia global máxima de las DMUs existentes. Esta frontera eficiente aparece representada con una línea discontinua en la Figura 2.3.

La expresión matemática que define la tecnología CRS es:

$$T_{CRS} = \{ (\vec{x}, \vec{y}) : \exists \vec{\lambda} \geq 0, \vec{\lambda}X \leq \vec{x}; \vec{\lambda}Y \geq \vec{y} \} \quad (2.14)$$

2.2 Modelos

Existen múltiples modelos DEA dependiendo de qué aspectos se pretenden enfatizar, debido a que depende de ello la eficiencia relativa.

2.2.1 Modelos de Retornos de Escala Constante

A continuación se presentan los modelos básicos en los que se considera como tecnología admisible, la tecnología CRS. Por lo que se consideran factibles como unidades de referencias, la unidad con mayor productividad escalada. Los modelos que se presentan son: Modelo Ratio, Modelo CCR-Input y Modelo CCR-Output, desarrollados por (Charnes et al. 1978)

2.2.1.1 Modelo Ratio

Calcula la eficiencia relativa de cada unidad al compararla con el resto de las unidades que forman parte de la tecnología. Parte de la definición de eficiencia relativa, debido a su objetivo es maximizar la eficiencia absoluta de la unidad que se está analizando (J). Este objetivo equivale a maximizar la eficiencia relativa de la unidad J, porque al tomar como referencia la unidad más eficiente el denominador será constante e igual a la unidad.

$$\max \frac{\sum_{k=1}^s v_{kJ} y_{kJ}}{\sum_{i=1}^m u_{iJ} x_{iJ}} \quad (2.15)$$

s. a.

$$\frac{\sum_{k=1}^s v_{kj} y_{kj}}{\sum_{i=1}^m u_{ij} x_{ij}} \leq 1 \quad j = 1, 2, \dots, n$$

$$v_{kj} \geq \varepsilon \quad k = 1, 2, \dots, p$$

$$u_{ij} \geq \varepsilon \quad i = 1, 2, \dots, m$$

Los pesos que ponderan cada variable son escogidos por la unidad analizada, de forma que esta buscará aquellos pesos que maximicen su eficiencia y se aplicarán al resto de las unidades. Y las restricciones que se emplean impiden que una variable tenga un peso negativo y que la definición de eficiencia no tenga un valor mayor que la unidad

Como resultado se obtendrá la eficiencia de cada unidad, aquellas unidades que tengan como eficiencia la unidad serán las más eficientes de la tecnología, mientras que las que tengan una eficiencia menor que uno serán ineficientes.

Al tratarse de un modelo no lineal se desarrollaron 2 modelos que lo linealizaban que se diferenciaban en la orientación que consideraban, orientación de entrada o de salida.

2.2.1.2 Modelo CCR-INPUT

Este modelo permite linealizar el modelo Radial teniendo en cuenta la orientación de entrada. Para ello maximiza el cociente al maximizar el numerador, dejando constante el denominador. En este modelo aparece una nueva restricción que aparte de mantener el denominador constante, elimina la redundancia de las soluciones.

$$\sum_{i=1}^m u_{ij} x_{ij} = 1 \quad (2.16)$$

A continuación se muestra el modelo en forma multiplicativa, tras linealizar completamente el modelo Radial.

$$\text{Maximizar} \quad \sum_{k=1}^s v_{kj} y_{kj}$$

s. a.

$$\sum_{k=1}^s v_{kj} y_{kj} - \sum_{i=1}^m u_{ij} x_{ij} \leq 1 \quad j = 1, 2, \dots, n \quad (2.17)$$

$$\sum_{i=1}^m u_{ij} x_{ij} = 1$$

$$v_{kj} \geq \varepsilon \quad k = 1, 2, \dots, p$$

$$u_{ij} \geq \varepsilon \quad i = 1, 2, \dots, m$$

Sin embargo, los pesos son complejos de interpretar por ello, normalmente se utiliza las variables del modelo dual, conocido como forma envolvente, que se expresa a continuación:

$$\text{Minimizar} \quad \theta_j - \varepsilon \left[\sum_{k=1}^s h_k^+ + \sum_{i=1}^m h_i^- \right] \quad (2.18)$$

s. a.

$$\begin{aligned} \sum_{j=1}^n \lambda_j x_{ij} &= \theta_J x_{iJ} - h_i^- & i = 1, 2, \dots, m \\ \sum_{j=1}^n \lambda_j y_{kj} &= y_{kJ} + h_k^+ & k = 1, 2, \dots, s \\ \lambda_j &\geq 0 \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k \\ \theta_J &\text{ libre} \end{aligned}$$

Este modelo se resuelve en dos fases.

○ FASE I

El objetivo de esta fase es minimizar los recursos empleados por la unidad J equiproporcionalmente, hasta que se encuentra con la frontera eficiente. Esta fase, se denomina fase radial al avanzar en una dirección que permita que se reduzcan todas las entradas en la misma proporción. La parte de la izquierda de las inecuaciones de las restricciones 1 y 2, determinan la zona admisible de la tecnología

Minimizar θ_J

s. a.

$$\begin{aligned} \sum_{j=1}^n \lambda_j x_{ij} &\leq \theta_J x_{iJ} & i = 1, 2, \dots, m \\ \sum_{j=1}^n \lambda_j y_{kj} &\geq y_{kJ} & k = 1, 2, \dots, s \\ \lambda_j &\geq 0 \quad \forall j \\ \theta_J &\text{ libre} \end{aligned} \tag{2.19}$$

Una vez obtenido el óptimo de la primera fase, se han reducido las entradas equiproporcionalmente lo máximo posible θ_J^* ; sin embargo la unidad J puede no haberse proyectado sobre ninguna unidad de la frontera eficiente, es por ello que entra en funcionamiento la segunda fase.

○ FASE II

Esta fase se llama fase rectangular, debido a que las holguras actúan de forma independiente y se maximizan hasta que la unidad proyectada alcanza la frontera eficiente.

$$\text{Minimizar} \quad -\varepsilon \left[\sum_{k=1}^s h_k^+ + \sum_{i=1}^m h_i^- \right]$$

s. a.

$$\begin{aligned} \sum_{j=1}^n \lambda_j x_{ij} &= \theta_J x_{iJ} - h_i^- & i = 1, 2, \dots, m \\ \sum_{j=1}^n \lambda_j y_{kj} &= y_{kJ} + h_k^+ & k = 1, 2, \dots, s \\ \lambda_j &\geq 0 \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k \end{aligned} \tag{2.20}$$

En la segunda fase se maximizan las holguras hasta que la unidad J obtiene como objetivo (target) un punto de la frontera eficiente. El target sobre el que se proyecta la unidad J es:

$$(x_{ij}^*, y_{kj}^*) \rightarrow (\theta_j x_{ij} - h_i^-, y_{kj} + h_k^-) \quad (2.21)$$

A continuación se muestran dos ejemplos, para visualizar como se produce la proyección de las unidades considerando la orientación de entrada.

En el caso de una entrada y una salida, se puede observar en la Figura 2.4 como las DMUs reducen sus entradas hasta que se encuentran con la frontera eficiente. Las unidades eficientes son las que se proyectan sobre sí mismas y forman parte de la frontera eficiente. En este caso, la única unidad eficiente que existe es la DMU D.

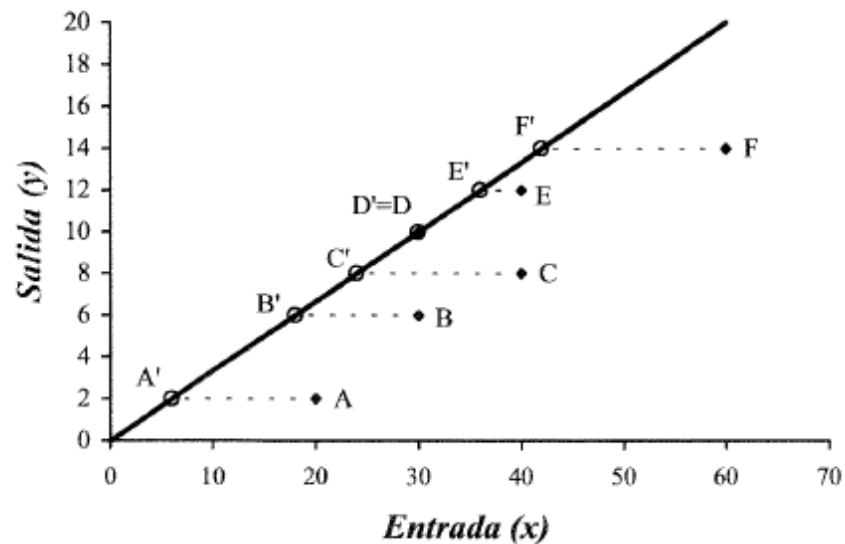


Figura 2.4 Modelo CCR-Input con una entrada y una salida

Fuente: (Villa 2003)

En la siguiente figura, se visualiza un modelo con dos entradas y una salida constante:

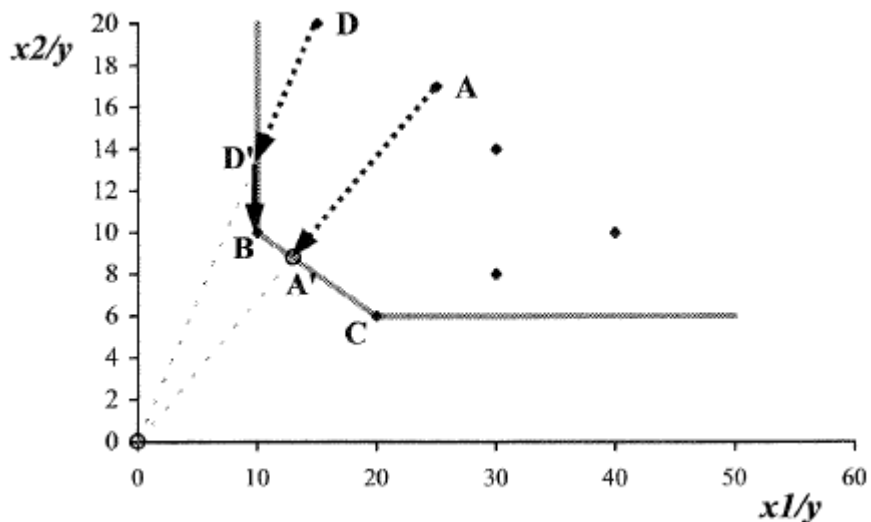


Figura 2.5: Modelo CCR-Input con dos entradas y una salida

Fuente: (Villa 2003)

Las flechas punteadas muestran las reducciones equiproporcionales hasta que se encuentran con la frontera admisible. En el caso de la DMU D, se observa el efecto de la segunda fase en la búsqueda de una unidad virtual que pertenezca a la frontera eficiente, al no pertenecer los hiperplanos paralelos a los ejes en la frontera eficiente. Sin embargo, en la DMU A, no se observa, porque en la fase I, ha alcanzado directamente una unidad de la frontera eficiente.

2.2.1.3 Modelo CCR-OUTPUT

El modelo CCR-Output es el resultado de linealizar el Modelo Ratio, considerando la orientación de salida. Es decir, se mide la eficiencia de las DMUs en función del incremento proporcional de outputs que podrían conseguir sin que se produzca una variación de sus recursos.

El modelo primal CCR-Output, para linealizar la función objetivo modelo radial, mantiene constante el numerador con una restricción que iguala la expresión a la unidad y minimiza el denominador en la función objetivo. De esta forma, al igual que en el modelo CCR-Output, con esta nueva restricción se eliminan duplicidades. La forma multiplicativa del modelo es:

$$\begin{aligned}
 & \text{Minimizar} \quad \sum_{i=1}^m u_{ij}x_{ij} \\
 & \text{s. a.} \\
 & \sum_{k=1}^s v_{kj}y_{kj} - \sum_{i=1}^m u_{ij}x_{ij} \leq 1 \quad j = 1, 2, \dots, n \\
 & \sum_{k=1}^s v_{kj}y_{kj} = 1 \\
 & v_{kj} \geq \varepsilon \quad k = 1, 2, \dots, p \\
 & u_{ij} \geq \varepsilon \quad i = 1, 2, \dots, m
 \end{aligned} \tag{2.22}$$

Mientras que la forma dual se expresa de la siguiente forma:

$$\begin{aligned}
 & \text{Maximizar} \quad \gamma_j + \varepsilon \left[\sum_{k=1}^s h_k^+ + \sum_{i=1}^m h_i^- \right] \\
 & \text{s. a.} \\
 & \sum_{j=1}^n \lambda_j x_{ij} = x_{ij} - h_i^- \quad i = 1, 2, \dots, m \\
 & \sum_{j=1}^n \lambda_j y_{kj} = \gamma_j y_{kj} + h_k^+ \quad k = 1, 2, \dots, s \\
 & \lambda_j \geq 0 \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k \\
 & \gamma_j \text{ libre}
 \end{aligned} \tag{2.23}$$

Al igual que el Modelo CCR-Input, el modelo se ejecuta en dos fases:

- FASE I

En esa fase se incrementan las salidas equiproporcionalmente hasta que la proyección se encuentra con la frontera de admisibilidad.

$$\text{Maximizar} \quad \gamma_j \tag{2.24}$$

$$\begin{aligned}
 & s. a. \\
 & \sum_{j=1}^n \lambda_j x_{ij} \leq x_{iJ} \quad i = 1, 2, \dots, m \\
 & \sum_{j=1}^n \lambda_j y_{kj} \geq \gamma_J y_{kJ} \quad k = 1, 2, \dots, s \\
 & \lambda_j \geq 0 \quad \forall j \\
 & \gamma_J \text{ libre}
 \end{aligned}$$

o FASE II

Esta fase entra en funcionamiento cuando la unidad proyectada que toma como referencia la unidad J, no se encuentra en la frontera eficiente.

$$\text{Maximizar } \varepsilon \left[\sum_{k=1}^s h_k^+ + \sum_{i=1}^m h_i^- \right]$$

$$\begin{aligned}
 & s. a. \\
 & \sum_{j=1}^n \lambda_j x_{ij} = x_{iJ} - h_i^- \quad i = 1, 2, \dots, m \\
 & \sum_{j=1}^n \lambda_j y_{kj} = \gamma_J y_{kJ} + h_k^+ \quad k = 1, 2, \dots, s \\
 & \lambda_j \geq 0 \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k
 \end{aligned} \tag{2.25}$$

A continuación se muestran dos ejemplos que pretenden visualizar como funciona el modelo CCR-Output. El primer ejemplo consta de una entrada y una salida. Como se aprecia en la Figura 2.6 Modelo CCR-Output con una entrada y una salida la frontera eficiente viene definida por las unidades más eficientes, en este caso la DMU D. Estas unidades se aprecian fácilmente en los resultados del modelo porque no tienen holgura y la variable γ_j es igual a la unidad, es decir se proyectan sobre ellas mismas.

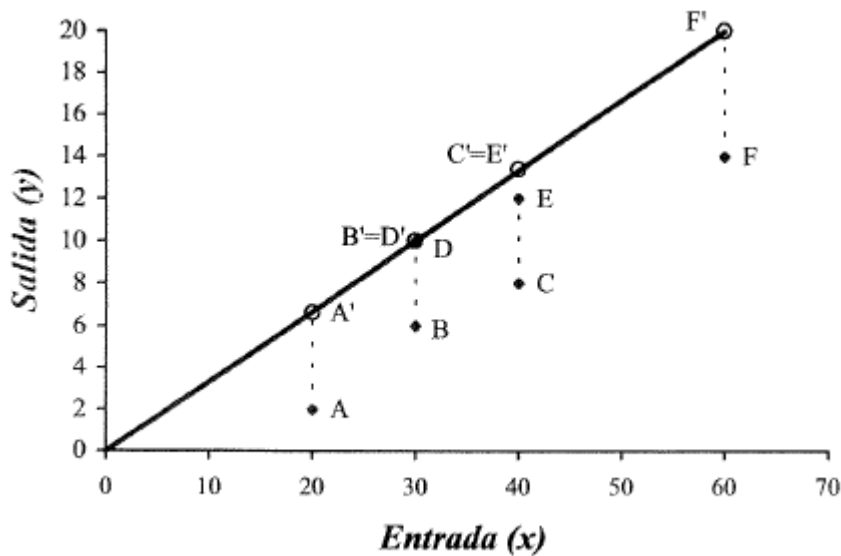


Figura 2.6 Modelo CCR-Output con una entrada y una salida

Fuente: (Villa 2003)

Las unidades ineficientes, aumentan su producción sin variar sus recursos, hasta que se encuentran con la frontera eficiente. En este caso la DMU D es la unidad de referencia para la unidad B. Por otra parte, las unidades E y C tienen la misma unidad de referencia al utilizar la misma cantidad de recursos.

El segundo ejemplo, analiza 6 unidades con una entrada constante y dos salidas. Como se aprecia en la Figura 2.7, al tener las salidas constantes se puede visualizar la frontera eficiente en 2 dimensiones, aunque el modelo tenga 3 dimensiones.

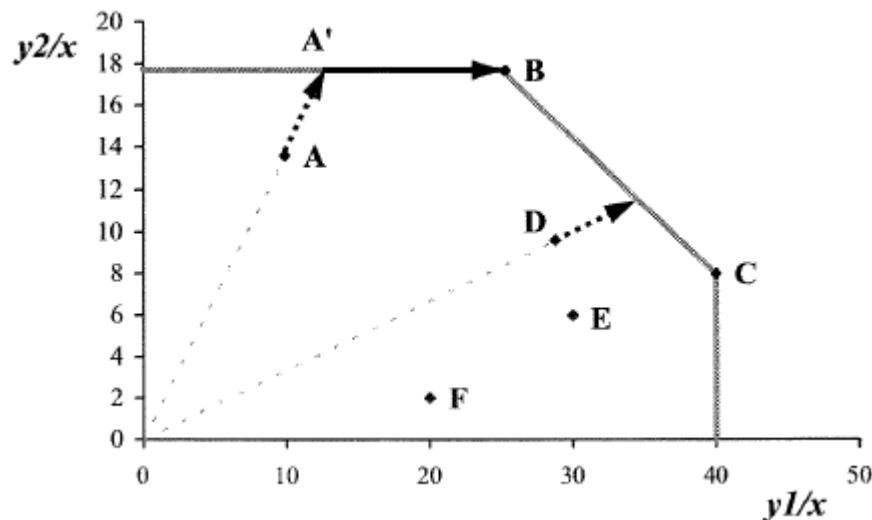


Figura 2.7 Modelo CCR-Output con una entrada y dos salidas

Fuente: (Villa 2003)

En este caso las unidades eficientes son las DMU B y C y la frontera eficiente se encuentra entre ambas unidades. Los tramos paralelos a los ejes que pasan por B y C representan la frontera admisible, por ello cuando una unidad se proyecta sobre esta parte de la frontera actúa la fase 2 para indicar a través de la holgura, la unidad virtual perteneciente a la frontera eficiente que más se aproxima a ella.

2.2.2 Modelos con Retornos de Escala Variables

A continuación se desarrollan los modelos básicos que consideran como tecnología admisible, la tecnología VRS. Los dos modelos que se plantean son BCC-Input y BCC-Output. Ambos partirán del Modelo Ratio; sin embargo deben imponer la condición de que la frontera eficiente sea convexa, dado que en esta tecnología no se considera la propiedad de la escalabilidad.

2.2.2.1 Modelo BCC-INPUT

El modelo fue desarrollado por (Banker, R.D., Charnes, A. Cooper 1984), con el fin de que cada unidad se compare con las unidades eficientes que tengan un tamaño similar y no con la unidad más eficiente de la muestra, como se daba en los modelos de retorno de escala constante. Es decir, se consideran eficientes aquellas unidades que sean combinación convexa de dos unidades eficientes observadas (“peer group”), dando lugar a una región convexa que define la tecnología.

El hecho de que la tecnología debe ser convexa se modela con la siguiente expresión.

$$\sum_{j=1}^n \lambda_j = 1 \quad (2.26)$$

De esta forma cada unidad virtual eficiente se situará entre dos unidades eficientes observadas, donde λ_j indicará el % de similitud que tiene con cada una de ellas.

A continuación se expresa el Modelo BCC-Input en forma envolvente, se aprecia que es similar al Modelo CCR-Input, salvo por el hecho de que la frontera es convexa.

$$\begin{aligned}
 & \text{Minimizar } \theta_j - \varepsilon \left[\sum_{k=1}^s h_k^+ + \sum_{i=1}^m h_i^- \right] \\
 & \text{s. a.} \\
 & \sum_{j=1}^n \lambda_j x_{ij} = \theta_j x_{ij} - h_i^- \quad i = 1, 2, \dots, m \\
 & \sum_{j=1}^n \lambda_j y_{kj} = y_{kj} + h_k^+ \quad k = 1, 2, \dots, s \\
 & \sum_{j=1}^n \lambda_j = 1 \\
 & \lambda_j \geq 0 \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k \\
 & \theta_j \text{ libre}
 \end{aligned} \tag{2.27}$$

A continuación en la Figura 2.8 se muestra la solución de un caso en el que se analizan las unidades productivas con una entrada y una salida.

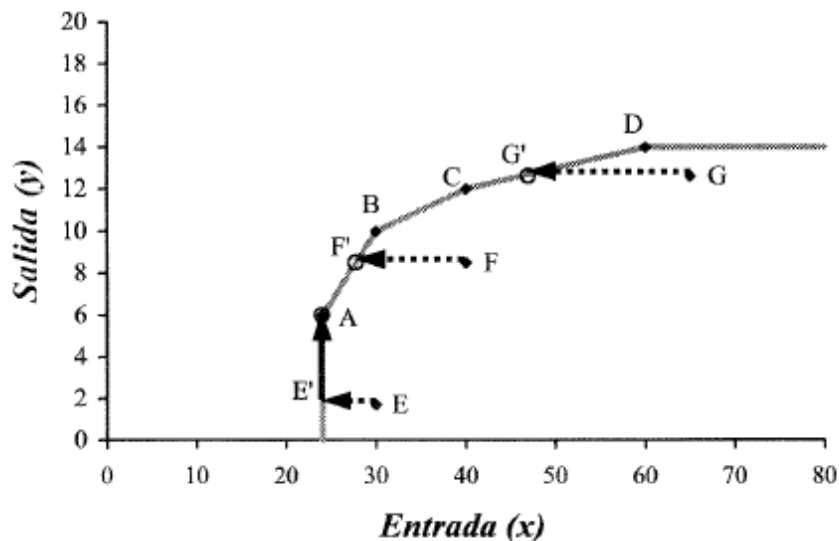


Figura 2.8 Modelo BCC-Input para el caso de una entrada y una salida

Fuente: (Villa 2003)

Como se puede apreciar la frontera eficiente está compuesta por tres tramos entre las unidades A y D. Al igual que en el caso de los modelos de retornos de escala constante las unidades eficientes se proyectan sobre sí mismas, por lo que las únicas proyecciones que se aprecian se corresponden a las de las unidades ineficientes. Las unidades F y G, no tienen holgura porque se proyectan directamente sobre la frontera eficiente, este no es el caso de la unidad E, debido a que se proyecta sobre la frontera admisible.

Si comparamos las diferentes soluciones que se obtienen a la hora de aplicar dos modelos diferentes con la

misma orientación, uno con retorno de escala constante y el otro con retorno de escala variable. Se aprecia como en el CCR-Input tiene menos unidades eficientes que el modelo BCC-Input, como se puede apreciar en la Figura 2.9.

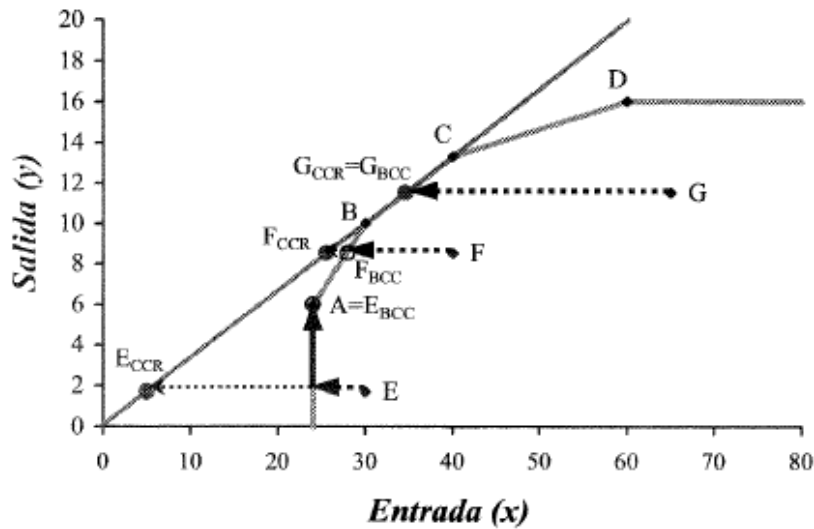


Figura 2.9 Comparación del modelo CCR-Input y el modelo BCC-Input

Fuente: (Villa 2003)

Por otra parte la eficiencia calculada en el modelo BCC-Input será superior a la del modelo CCR-Input, debido a que la frontera eficiente en este último se encuentra más alejado de las unidades ineficientes, al compararse con las unidades de mayor productividad del problema.

Muchas unidades que son eficientes, ya sean virtuales u observadas del modelo BCC-Input serían ineficientes para el modelo CCR-Inputs, siempre que no se encontraran dichas unidades en la parte de la frontera con el tamaño de escala más productivo (MPSS), línea BC.

Las unidades que se encuentran en la zona AB, se encuentran en la zona con retorno de escala creciente (Increasing Return Scale, IRS) son unidades eficientes para las unidades de tamaño similar; sin embargo deberían incrementar sus inputs si quieren alcanzar el tamaño de mayor productividad del problema. Lo mismo ocurre con las unidades de la zona CD, son unidades eficientes cuando se compararan con unidades de tamaño similar pero si quieren alcanzar el tamaño de mayor productividad observada en el problema deberían reducir sus recursos, estas unidades operan con retornos de escala decrecientes (Decreasing Return Scale, DRS)

2.2.2.2 Modelo BCC-OUTPUT

De forma análoga al modelo BCC-Input se construye el modelo de Retorno de Escala Variable con orientación de salida, partiendo del modelo CCR-Output. La expresión del modelo BCC-Output en forma envolvente es:

$$\begin{aligned} & \text{Maximizar } \gamma_j + \varepsilon \left[\sum_{k=1}^s h_k^+ + \sum_{i=1}^m h_i^- \right] \\ & \text{s. a.} \\ & \sum_{j=1}^n \lambda_j x_{ij} = x_{ij} - h_i^- \quad i = 1, 2, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{kj} = \gamma_j y_{kj} + h_k^+ \quad k = 1, 2, \dots, s \end{aligned} \quad (2.28)$$

$$\sum_{j=1}^n \lambda_j = 1$$

$$\lambda_j \geq 0 \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k$$

$$\gamma_j \text{ libre}$$

Las apreciaciones realizadas en el ejemplo del modelo BCC-Input y su comparativa con el modelo CCR-Input, se pueden extrapolar al modelo BCC-Output, con la salvedad de que la orientación, de este último, es de salida.

2.2.3 Modelos FDH (Free Disposal Hull)

Estos modelos utilizan la tecnología FDH, se diferencian en la tecnología VRS en que no permiten la combinación convexa de dos unidades existentes, únicamente se considera la existencia de las unidades observadas y de otras unidades virtuales que consumen más recursos o producen menos que estas. La tecnología FDH, como se había visto el apartado 2.1.2 tiene las propiedades de envoltura y libre disponibilidad.

A continuación se presentan varios modelos con tecnología FDH: Modelo FDH con orientación de entrada, Modelo FDH con orientación de salida, Modelo Aditivo, Measure of Inefficiency Proportions y Range-Adjusted Measure.

2.2.3.1 Modelo FDH con orientación de entrada

A partir de un modelo que considera la tecnología VRS, es sencillo imponer la condición de tecnología FDH, ya que la única condición que habría que imponer es que una unidad tome como referencia una única unidad eficiente existente. Es decir que la variable λ_j sea binaria. La expresión matemática del Modelo FDH con orientación de entrada es:

$$\text{Minimizar } \theta_j$$

$$s. a.$$

$$\sum_{j=1}^n \lambda_j x_{ij} = \theta_j x_{ij} - h_i^- \quad i = 1, 2, \dots, m$$

$$\sum_{j=1}^n \lambda_j y_{kj} = y_{kj} + h_k^+ \quad k = 1, 2, \dots, s \quad (2.29)$$

$$\sum_{j=1}^n \lambda_j = 1$$

$$\lambda_j = \{0, 1\} \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k$$

$$\theta_j \text{ libre}$$

Las holguras no participan en la función objetivo, debido a que con la proyección es radial, una vez que la DMU se proyecta sobre la frontera admisible utiliza las holguras para proyectarse sobre la frontera eficiente. Se puede dar el caso en el que una unidad se proyecte directamente sobre la frontera eficiente, en este caso las holguras serían nulas.

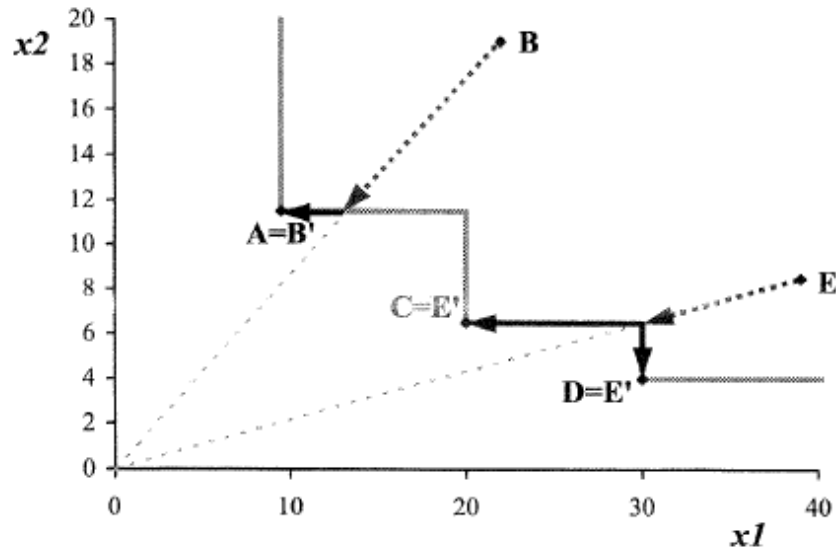


Figura 2.10: Modelo FDH con orientación de entrada en el caso de un modelo con dos entradas y una salida

Fuente: (Villa 2003)

En la Figura 2.10, se muestra la solución de un Modelo FDH con orientación de entrada en un caso con dos entradas y una salida. Es interesante destacar el caso de la unidad ineficiente E, en el que tras chocar con la frontera eficiente puede tomar indistintamente como unidad de referencia la unidad C o D. Todas las variables λ_j tendrán valor cero salvo una que tendrá valor 1. Al tener variables binarias, no se puede resolver por el método simplex; sin embargo se puede utilizar el siguiente algoritmo:

$$\theta_j^* = \min_{j \in D(J)} \left\{ \max_{i=1, \dots, m} \left\{ \frac{x_{kj}}{x_{ki}} \right\} \right\} \quad (2.30)$$

Siendo: $D(J) = \{j. x_{ij} \leq x_{iJ}, \forall i; y_{kj} \leq y_{kJ}, \forall k\}$ Es decir, las unidades que dominan a J.

2.2.3.2 Modelo FDH con orientación de salida

Si se aplican las mismas consideraciones que se han tenido en cuenta en el apartado 2.2.3.1, el modelo que considera la orientación de salida con tecnología FDH tiene la siguiente expresión matemática:

$$\begin{aligned} & \text{Maximizar } \gamma_j \\ & \text{s. a.} \\ & \sum_{j=1}^n \lambda_j x_{ij} = x_{iJ} - h_i^- \quad i = 1, 2, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{kj} = \gamma_j y_{kJ} + h_k^+ \quad k = 1, 2, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \lambda_j = \{0, 1\} \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k \end{aligned} \quad (2.31)$$

γ_j libre

Siendo el algoritmo que lo resuelve:

$$\gamma_j^* = \max_{j \in D(J)} \left\{ \min_{k=1, \dots, p} \left\{ \frac{y_{kj}}{y_{kj}} \right\} \right\} \quad (2.32)$$

2.2.3.3 Modelo Aditivo

El modelo aditivo fue desarrollado inicialmente por (Charnes et al. 1985) y más tarde por (Bardhan et al. 1996) Este modelo se caracteriza porque no tiene orientación, no dispone de fase radial, únicamente tiene fase rectangular como la fase 2 del modelo CCR y BCC. Este modelo se puede utilizar en tecnología CRS y VRS con las restricciones de la región admisible que se emplearon en los apartados 2.2.1 y 2.2.2.

A continuación se expresa el modelo en forma envolvente:

$$\begin{aligned} & \text{Maximizar} \quad \sum_{k=1}^s h_k^+ + \sum_{i=1}^m h_i^- \\ & \text{s. a.} \\ & \sum_{j=1}^n \lambda_j x_{ij} = x_{ij} - h_i^- \quad i = 1, 2, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{kj} = y_{kj} + h_k^+ \quad k = 1, 2, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \lambda_j = \{0, 1\} \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k \end{aligned} \quad (2.33)$$

El hecho de que no tenga orientación, permite que la unidad productiva se compare con más unidades eficientes de la frontera eficiente y tome como referencia aquella que maximice las holguras.

Una característica de este modelo, descubierta por (Ali & Seiford 1990) que también posee el modelo BCC, es que es invariante ante las translaciones, es decir, se puede añadir una constante arbitraria tanto a los recursos como a las salidas y no variarán los valores óptimos del modelo, ni la tecnología, ni la ordenación de las unidades analizadas.

2.2.3.4 Measure of Inefficiency Proportions

En el artículo (Cooper, W. W., Park, K. S. Pastor 1999), se desarrolló el modelo Measure of Inefficiency Proportions (MIP), es un modelo aditivo, por lo que no tiene orientación. Esta medida agrega la proporción de las holguras respecto de la unidad analizada; e intenta encontrar aquellas unidades de referencia que maximice la suma de dichas proporciones. Se rige por la siguiente expresión:

$$\begin{aligned} & \text{Maximizar} \quad MIP = \sum_{i=1}^m \frac{h_i^-}{x_{ij}} + \sum_{k=1}^s \frac{h_k^+}{y_{kj}} \\ & \text{s. a.} \\ & \sum_{j=1}^n \lambda_j x_{ij} = x_{ij} - h_i^- \quad i = 1, 2, \dots, m \end{aligned} \quad (2.34)$$

$$\begin{aligned} \sum_{j=1}^n \lambda_j y_{kj} &= y_{kj} + h_k^+ & k = 1, 2, \dots, s \\ \sum_{j=1}^n \lambda_j &= 1 \\ \lambda_j &= \{0, 1\} \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k \end{aligned}$$

2.2.3.5 Range-Adjusted Measure

Range-Adjusted Measure (RAM) fue desarrollado en (Cooper, W. W., Park, K. S. Pastor 1999), y al igual que el modelo MIP es un modelo aditivo y no tiene orientación. Como consecuencia, permite determinar cuál es la proporción de la holgura respecto de la máxima holgura que se da en la muestra. Para ello define dos parámetros denominados rango que tienen como valor la holgura máxima de cada variable.

$$\begin{aligned} R_i^- &= \max_j \{x_{ij}\} - \min_j \{x_{ij}\} \\ R_k^+ &= \max_k \{y_{kj}\} - \min_k \{y_{kj}\} \end{aligned} \quad (2.35)$$

Si se agregan las proporciones de las holguras respecto de su rango, se obtiene la siguiente expresión:

$$0 \leq \frac{1}{m+s} \left(\sum_{i=1}^m \frac{h_i^-}{R_i^-} + \sum_{k=1}^s \frac{h_k^+}{R_k^+} \right) \leq 1 \quad (2.36)$$

Al tener un valor entre 0 y 1, la medida de eficiencia se calcula restándole a la unidad la ineficiencia, dejando la siguiente expresión

$$0 \leq 1 - \frac{1}{m+s} \left(\sum_{i=1}^m \frac{h_i^-}{R_i^-} + \sum_{k=1}^s \frac{h_k^+}{R_k^+} \right) \leq 1 \quad (2.37)$$

Por lo tanto el modelo se expresa con la siguiente formulación:

$$\begin{aligned} & \text{Maximizar} \quad \frac{1}{m+s} \left(\sum_{i=1}^m \frac{h_i^-}{R_i^-} + \sum_{k=1}^s \frac{h_k^+}{R_k^+} \right) \\ & \text{s. a.} \\ & \sum_{j=1}^n \lambda_j x_{ij} = x_{ij} - h_i^- & i = 1, 2, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{kj} = y_{kj} + h_k^+ & k = 1, 2, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \\ & R_i^- = \max_j \{x_{ij}\} - \min_j \{x_{ij}\} & i = 1, 2, \dots, m \\ & R_k^+ = \max_k \{y_{kj}\} - \min_k \{y_{kj}\} & k = 1, 2, \dots, s \end{aligned} \quad (2.38)$$

$$\lambda_j = \{0,1\} \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k$$

2.3 Medidas de eficiencia

En (Charnes et al. 1978) se define que la unidad eficiente que se usaba como referencia se nombraba con un asterisco (x_i^*, y_k^*) para todos los recursos i y todas las salidas k de esa unidad eficiente, así como todas las variables que se utilizaban para referirse a ella. Siendo una unidad eficiente, aquella que no pueda mejorar sus entradas o salidas sin empeorar otras entradas o salidas.

La medida que analice la eficiencia de una unidad ineficiente (Γ) a partir de los resultados de los modelos anteriores debe tener las siguientes características:

- A. $0 \leq \Gamma \leq 1$
- B. $\Gamma = \begin{cases} 1 & \Leftrightarrow DMU_o \text{ es eficiente} \\ 0 & \Leftrightarrow DMU_o \text{ es ineficiente} \end{cases}$
- C. Γ es invariante independientemente de cual sea el óptimo y de las unidades que definan las variables de entrada y de salida
- D. Γ es monotonica, siendo una función monotónica aquella cuya primera derivada no cambia de signo.

Measure Efficiency Dominance (MED) fue desarrollada por (Bardhan et al. 1996) en el que partía de las restricciones de la tecnología del modelo aditivo para representar la ineficiencia de las unidades proyectadas, apoyándose en la propiedad de la traslación invariante del modelo aditivo. Se genera la proporción ineficiente de la entrada i de la unidad J .

$$\begin{aligned} \sum_{j=1}^n \lambda_j x_{ij} &= x_i^* = x_{iJ} - h_i^- & i = 1, 2, \dots, m \\ 0 &\leq h_i^- = x_{iJ} - x_i^* \leq x_{iJ} \\ 0 &\leq \frac{h_i^-}{x_{iJ}} = \frac{x_{iJ} - x_i^*}{x_{iJ}} \leq 1 \end{aligned} \quad (2.39)$$

Y de la misma forma se genera la proporción ineficiente de la salida k de la unidad J :

$$\begin{aligned} \sum_{j=1}^n \lambda_j y_{kj} &= y_k^* = y_{kJ} + h_k^+ & k = 1, 2, \dots, s \\ 0 &\leq h_k^+ = y_k^* - y_{kJ} \leq y_k^* \\ 0 &\leq \frac{h_k^+}{y_k^*} = \frac{y_k^* - y_{kJ}}{y_k^*} \leq 1 \end{aligned} \quad (2.40)$$

Estas proporciones son adimensionales por lo que se pueden agregar en la siguiente expresión, denominada Measure of Inefficiency Dominance (MID).

$$0 \leq \frac{\sum_{i=1}^m \frac{x_{iJ} - x_i^*}{x_{iJ}} + \sum_{k=1}^s \frac{y_k^* - y_{kJ}}{y_k^*}}{s + m} \leq 1 \quad (2.41)$$

Automáticamente, al tener la medida de ineficiencia valores entre 0 y 1, al restarle a la unidad se obtiene la llamada Measure of Efficiency Dominance (MED):

$$0 \leq 1 - \frac{\sum_{i=1}^m \frac{x_{ij} - x_i^*}{x_{ij}} + \sum_{k=1}^s \frac{y_k^* - y_{kj}}{y_k^*}}{s + m} \leq 1 \quad (2.42)$$

Esta medida será igual a la unidad cuando no existan holguras, es decir cuando la unidad sea eficiente y será cero cuando sea ineficiente.

2.4 Secuencia de targets intermedios

Una DMU ineficiente debe establecer unos targets intermedios antes de llegar a la frontera eficiente bien porque la unidad proyectada es una unidad virtual, ejecutar a la vez múltiples estrategias que mejoren la eficiencia resulta complejo y porque es complicado alcanzar la eficiencia en un solo paso.

A continuación se muestra un ejemplo de la dificultad de proyectarse sobre la frontera eficiente en una tecnología VRS en el caso de dos entradas y 1 salida con valores constantes. La unidad L tiene como objetivo la unidad virtual L', al no existir esta unidad es complejo materializarla como objetivo así como alcanzarlo en un solo paso.

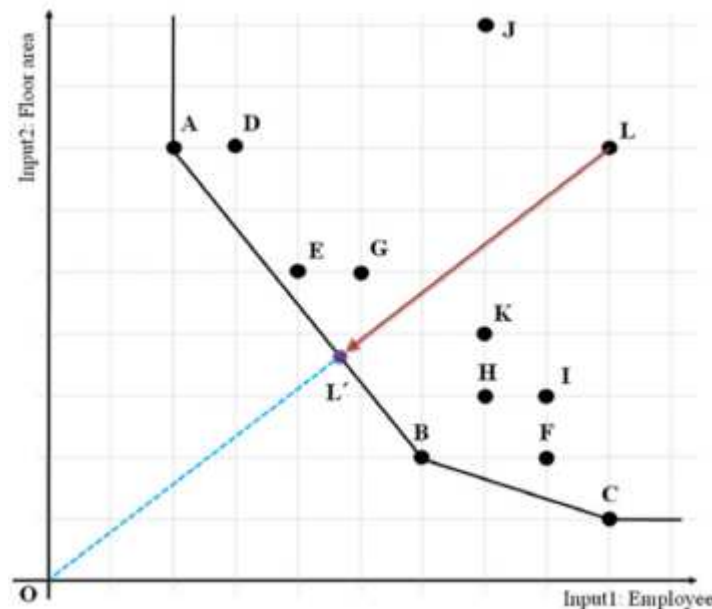


Figura 2.11 Modelo Aditivo con tecnología VRS, caso de 2 entradas y 1 salida

Fuente: (Lim et al. 2011)

Una unidad ineficiente puede estar dominada por otras unidades son ineficientes, por lo tanto no todas las unidades ineficientes son igual de importantes. Si una vez analizada la tecnología, se elimina la frontera eficiente las DMUs que formen la nueva frontera eficiente pertenecerán a Frontera Eficiente de segundo nivel. Si se repite el proceso, se obtendrá la Frontera Eficiente de tercer nivel, así sucesivamente hasta que no queden más DMUs. Este proceso fue expuesto por (Seiford & Zhu 2003) y proporciona una media relativa del atractivo de una DMU y su progreso hacia la frontera eficiente.

2.4.1 Technical Efficiency Improvement Program y Scale Efficiency Improvement Program

Las unidades intermedias se convierten en objetivos específicos que pueden monitorizar y medir el grado de mejora de la unidad analizada. Por otra parte aquellas unidades que son eficientes técnicamente, deben intentar alcanzar la eficiencia global y situarse en la zona de máxima productividad (Most Productive Scale Size). (Lozano & Villa 2010) definieron dos modelos basados en el modelo MIP para tecnología VRS. El primero

determinaba las unidades intermedias que no pertenecían a la frontera eficiente como Technical Efficiency Improvement Program (TEIP) y marcaban el camino a la frontera eficiente, mientras que el segundo, llamado Scale Efficiency Improvement Program (SEIP), tenía como objetivo determinar los pasos intermedios para que una unidad técnicamente eficiente alcanzara la zona MPSS.

El modelo TEIP impone unos límites de mejora que generan unas unidades virtuales que marcan el camino hacia la frontera eficiente. Para ello el decisor determina las siguientes variables:

- α_{i0}^- Máximo reducción relativa del recurso i para la unidad 0
- β_{k0}^+ Máximo aumento relativo de las salidas k para la unidad 0

El modelo que desarrollaron, en el que se analiza el objetivo para la unidad DMU0 en el paso t es el siguiente:

$$\begin{aligned}
 & \text{Maximizar } \delta_{0t} = \sum_{i=1}^m \frac{h_{it}^-}{x_{ij}} + \sum_{k=1}^s \frac{h_{kt}^+}{y_{kj}} \\
 & \text{s. a.} \\
 & \sum_{j=1}^n \lambda_j x_{ij} = x_{ij}^{t-1} - h_{it}^- \quad i = 1, 2, \dots, m \\
 & \sum_{j=1}^n \lambda_j y_{kj} = y_{kj}^{t-1} + h_{kt}^+ \quad k = 1, 2, \dots, s \\
 & \sum_{j=1}^n \lambda_j = 1 \\
 & h_{it}^- \leq \alpha_{i0}^- x_{ij}^{t-1} \quad i = 1, 2, \dots, m \\
 & h_{kt}^+ \leq \beta_{k0}^+ y_{kj}^{t-1} \quad k = 1, 2, \dots, s \\
 & \lambda_{jt} > 0 \quad \forall j; \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k
 \end{aligned} \tag{2.43}$$

Siendo:

- t el índice de los targets intermedios
- δ_{0t} Incremento de la eficiencia técnica en el paso t

La solución óptima que genera el modelo para el paso t para la unidad 0

$$\begin{aligned}
 x_{i0}^t &= \sum_{j=1}^n \lambda_j x_{ij} = x_{ij}^{t-1} - (h_{it}^-)^* \\
 y_{k0}^t &= \sum_{j=1}^n \lambda_j y_{kj} = y_{kj}^{t-1} + (h_{kt}^+)^*
 \end{aligned}$$

El modelo SEIP generado para que las unidades eficientes obtengan la eficiencia global, no debe presentar la restricción de convexidad de la tecnología VRS, para utilizar la tecnología CRS. La siguiente expresión, determina el modelo con paso t de la unidad 0 hasta la unidad con eficiencia global 0'.

$$\begin{aligned}
 & \text{Minimizar } \lambda_{0t} \\
 & \text{s. a.} \\
 & \lambda_{0t} x_{i0} + \lambda_{0't} x_{i0'} = x_{i0}^{t-1} + h_{it}^+ + h_{it}^- \\
 & \lambda_{0t} y_{k0} + \lambda_{0't} y_{k0'} = y_{k0}^{t-1} + h_{kt}^+ + h_{kt}^-
 \end{aligned} \tag{2.44}$$

$$\begin{aligned}\lambda_{0t} + \lambda_{0,t} &= 1 \\ h_{it}^+ &\leq \alpha_{i0}^+ x_{i0}^{t-1} \quad i = 1, 2, \dots, m \\ h_{it}^- &\leq \alpha_{i0}^+ x_{i0}^{t-1} \quad i = 1, 2, \dots, m \\ h_{kt}^+ &\leq \beta_{k0}^+ y_{k0}^{t-1} \quad k = 1, 2, \dots, s \\ h_{kt}^- &\leq \beta_{k0}^+ y_{k0}^{t-1} \quad k = 1, 2, \dots, s \\ \lambda_{0t}, \lambda_{0,t}, h_{it}^+, h_{it}^-, h_{kt}^+, h_{kt}^- &\geq 0 \quad \forall i, k\end{aligned}$$

La unidad que se pretende alcanzar 0^* se ha determinado a través del modelo CRS MIP y permite determinar que unidades pertenecen a la zona MPSS así como qué unidades de dicha zona se convierten en los objetivos de las unidades eficientes que no tienen eficiencia global.

En la Figura 2.12 se muestra el análisis de 6 DMUs con una entrada y una salida, en una tecnología VRS y los pasos que deberían seguir las unidades eficientes E y F hasta la frontera eficiente con el modelo TEIP, así como los pasos que deberían dar las unidades eficientes A, D y la unidad E cuando llega a la frontera eficiente, para alcanzar la eficiencia global.

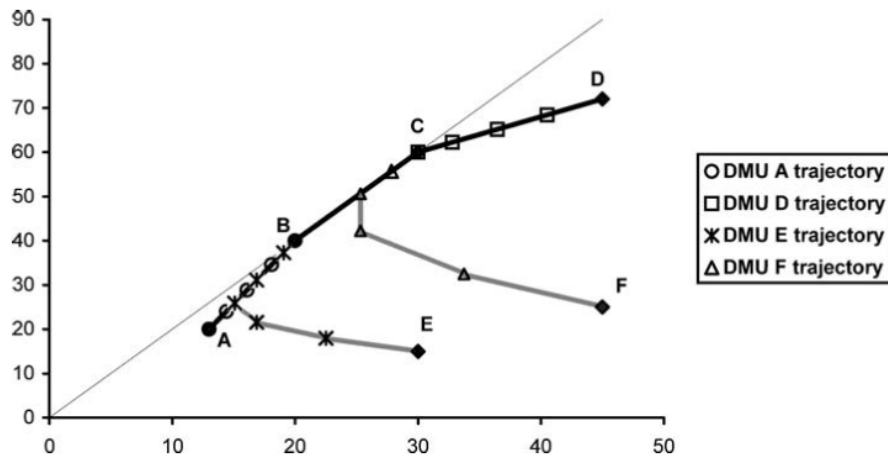


Figura 2.12: Visualización de los pasos intermedios generados por los modelos TEIP y SEIP en un caso con 6 DMUs con una entrada y una salida

Fuente: (Lozano & Villa 2010)

2.4.2 Selección de targets intermedios

En el artículo (Lim et al. 2011) se buscan aquellas unidades intermedias que permitan diseñar un camino hacia la frontera eficiente para todas las unidades que no pertenecen a ella. Serán targets intermedios aquellas unidades que tengan más atractivo para la unidad analizada, que no se encuentren muy alejadas y sean factibles para dicha unidad.

La selección de la unidad objetivo se basa en tres criterios que se ponderan con unos pesos (w_1, w_2, w_3) escogidos por el decisor:

$$Benchmark_k^* = \arg \max \{ S_{kp} = w_1 A_p^* - w_2 P_p^* - w_3 F_{kp} : p \in E^{t-1} \} \quad (2.45)$$

Siendo:

- E^l la frontera eficiente del nivel $l \in [1, L]$
- $j \in F(J^l)$ el conjunto de $DMU_j \in J^l$, donde $J^{l+1} = J^l - E^l$.

J^1 está formado por todas las unidades observadas en el problema

- A_k^* Atractivo
- P_p^* Progreso
- F_{kp} Inviabilidad

En primer lugar hay que agrupar las DMUs en capas de acuerdo a su eficiencia. Para ello se ejecuta el siguiente algoritmo:

- Paso 1: Para $l = 1$, J^l es el conjunto de todas las unidades analizadas
- Paso 2: Evaluar el conjunto de las unidades del conjunto J^l con el siguiente modelo

$$z^*(l, k) = \text{Maximizar } \frac{1}{m + s} \left(\sum_{i=1}^m \frac{h_i^-}{R_i^-} + \sum_{k=1}^s \frac{h_k^+}{R_k^+} \right)$$

s. a.

$$\sum_{j \in F(J^l)} \lambda_j x_{ij} = x_{ij} - h_i^- \quad i = 1, 2, \dots, m$$

$$\sum_{j \in F(J^l)} \lambda_j y_{kj} = y_{kj} + h_k^+ \quad k = 1, 2, \dots, s \quad (2.46)$$

$$\sum_{j \in F(J^l)} \lambda_j = 1$$

$$R_i^- = \max_j \{x_{ij}\} - \min_j \{x_{ij}\} \quad i = 1, 2, \dots, m$$

$$R_k^+ = \max_k \{y_{kj}\} - \min_k \{y_{kj}\} \quad k = 1, 2, \dots, s$$

$$\lambda_j \geq 0 \quad \forall j \in F(J^l); \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k$$

De esta forma se calculan las unidades que pertenecen a la frontera eficiente de nivel l , E^l

- Paso 3: Excluir de las unidades eficientes las obtenidas en el conjunto E^l . Es decir, $J^{l+1} = J^l - E^l$
Si $J^{l+1} = \emptyset$ el algoritmo se para.
- Paso 4: $l = l + 1$ Se vuelve al paso 2

En la siguiente figura, se muestra el resultado de aplicar el algoritmo sobre un conjunto de datos tipo que simulan ser supermercados con dos entradas y una salida constante:

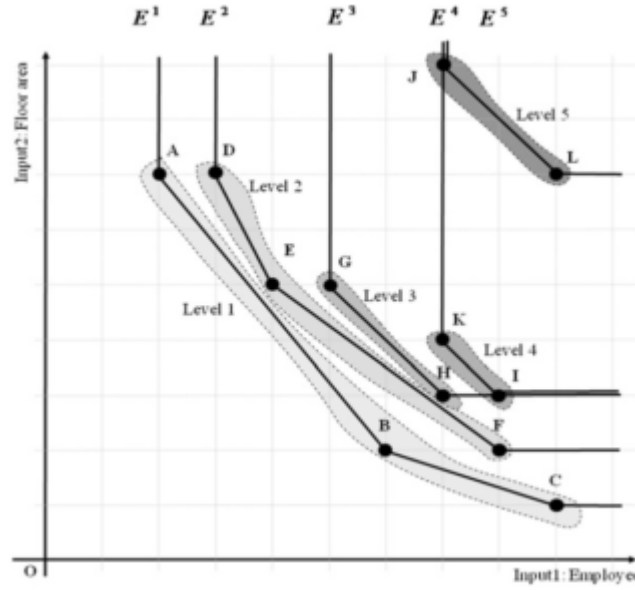


Figura 2.13 Identificación de los diferentes niveles de frontera eficiente en un conjunto de datos con dos entradas y una salida constante.

Fuente: (Lim et al. 2011)

La selección de la unidad objetivo se basa en tres criterios que se ponderan con unos pesos (w_1, w_2, w_3) escogidos por el decisor:

$$Benchmark_k^* = \arg \max \{ S_{kp} = w_1 A_p^* - w_2 P_p^* - w_3 F_{kp} : p \in E^{l-1} \} \quad (2.47)$$

Siendo

- E^l la frontera eficiente del nivel l
- $j \in F(J^l)$ el conjunto de $DMU_j \in J^l$, donde $J^{l+1} = J^l - E^l$.
- J^1 está formado por todas las unidades observadas en el problema

A_k^* Atractivo: En el modelo que se muestra a continuación, se busca calcular el atractivo de la unidad k que pertenece al nivel $l \in (1, L - 1)$, siendo L el número máximo de capas, en función de la distancia de esta unidad a la capa $l+1$. Es decir, cuanto más cerca se encuentre la unidad k de la siguiente capa más ineficiente, más atractivo tendrá.

$$A_k^* = \text{Minimizar} \frac{1}{m+s} \left(\sum_{i=1}^m \frac{h_i^-}{R_i^-} + \sum_{k=1}^s \frac{h_k^+}{R_k^+} \right)$$

s. a.

$$\sum_{j \in F(E^{l+1})} \lambda_j x_{ij} - h_i^- \leq x_{ij} \quad i = 1, 2, \dots, m \quad (2.48)$$

$$\sum_{j \in F(E^{l+1})} \lambda_j y_{kj} + h_k^+ \geq y_{kj} \quad k = 1, 2, \dots, s$$

$$\sum_{j \in F(E^{l+1})} \lambda_j = 1$$

$$\begin{aligned}
R_i^- &= \max_j \{x_{ij}\} - \min_j \{x_{ij}\} \quad i = 1, 2, \dots, m \\
R_k^+ &= \max_k \{y_{kj}\} - \min_k \{y_{kj}\} \quad k = 1, 2, \dots, s \\
\lambda_j &\geq 0 \quad \forall j \in F(E^{l+1}); \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k
\end{aligned}$$

El Progreso P_k^* para una unidad k que pertenece al nivel $l \in (2, L)$, representa la capacidad de avance ante la frontera eficiente l-1. Se determina a través del siguiente modelo:

$$\begin{aligned}
P_k^* &= \text{maximizar} \frac{1}{m+s} \left(\sum_{i=1}^m \frac{h_i^-}{R_i^-} + \sum_{k=1}^s \frac{h_k^+}{R_k^+} \right) \\
\text{s. a.} \\
\sum_{j \in E^{l-1}} \lambda_j x_{ij} + h_i^- &= x_{ij} \quad i = 1, 2, \dots, m \\
\sum_{j \in E^{l-1}} \lambda_j y_{kj} - h_k^+ &= y_{kj} \quad k = 1, 2, \dots, s \\
\sum_{j \in E^{l-1}} \lambda_j &= 1 \\
R_i^- &= \max_j \{x_{ij}\} - \min_j \{x_{ij}\} \quad i = 1, 2, \dots, m \\
R_k^+ &= \max_k \{y_{kj}\} - \min_k \{y_{kj}\} \quad k = 1, 2, \dots, s \\
\lambda_j &\geq 0 \quad \forall j \in F(E^{l-1}); \quad h_i^-, h_k^+ \geq 0 \quad \forall i, k
\end{aligned} \tag{2.49}$$

La inviabilidad F_{kp} deja constancia de las Limitaciones de la unidad $DMU_k = (x_k, y_k)$ para alcanzar la unidad $DMU_p = (x_p, y_p)$ Para ello se representan sus restricciones a través de la siguiente inecuación:

$$A \begin{pmatrix} h^- \\ h^+ \end{pmatrix} \leq b \tag{2.50}$$

Siendo $h^- = x_k - x_p$; $h^+ = y_p - y_k$; A y b una matriz y un vector, respectivamente con las dimensiones adecuadas que representan las restricciones de las unidades. Calculándose la inviabilidad con la siguiente expresión:

$$F_{kp} = \frac{\left\| b - A \begin{pmatrix} h_i^- \\ h_k^+ \end{pmatrix} \right\|}{\|b\|} \tag{2.51}$$

Como resultado se obtiene un camino desde una unidad ineficiente hasta la frontera eficiente, a través de los diferentes niveles de frontera eficiente que existen entre ellos y proyectándose siempre en unidades eficientes existentes. Con el objetivo de mostrar a dicha unidad aquellos objetivos reales que permitan mejorar su efectividad de forma paulatina y siempre dentro de sus capacidades. Es importante remarcar que el camino se diseña buscando los targets intermedios más factibles para la unidad analizada, no en el camino más corto a la frontera eficiente.

En la siguiente figura se muestra el camino diseñado, a través del modelo de (Lim et al. 2011) para la unidad ineficiente L hasta la frontera eficiente. En el mismo conjunto de datos de la Figura 2.13

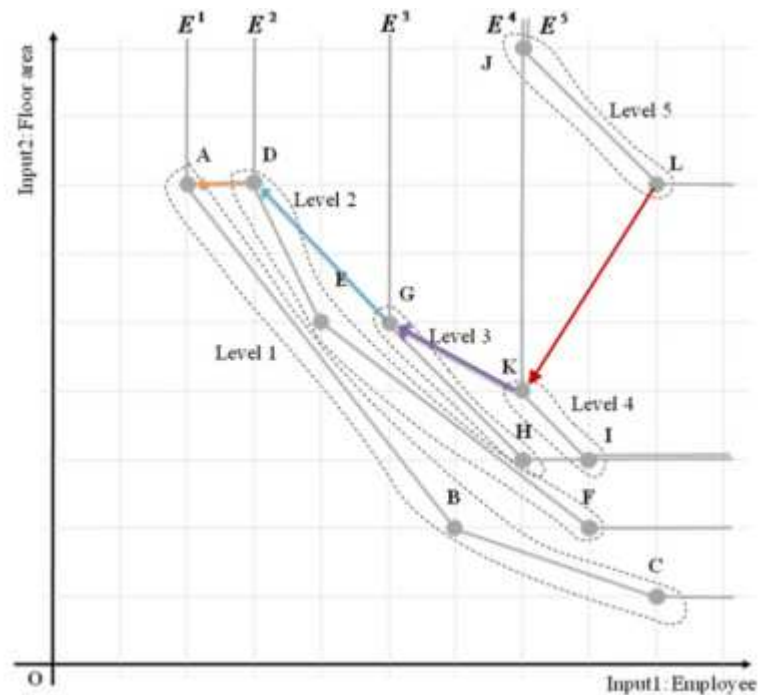


Figura 2.14: Objetivos reales propuestos para la DMU L que le permiten llegar a la frontera eficiente.

Fuente: (Lim et al. 2011)

2.4.3 Selección de rutas

En el artículo (Park et al. 2012), a diferencia de los modelos anteriores, se considera el paso óptimo para determinar los targets intermedios entre las unidades ineficientes y la frontera eficiente. Y por otra parte, en el caso de considerar múltiples objetivos hay que considerar tanto la mejora de los recursos al alcanzar los objetivos como la proximidad a la dirección de mejora. Esto permite obtener un modelo más realista y efectivo.

(Park et al. 2012) aplica su metodología en un caso de la literatura en el que se analiza un conjunto de supermercados, cuyo resultado sobre el nodo ineficiente L, se muestra en la Figura 2.15. En esta figura se observa cómo existen múltiples targets intermedios que se pueden utilizar para llegar a la frontera eficiente.

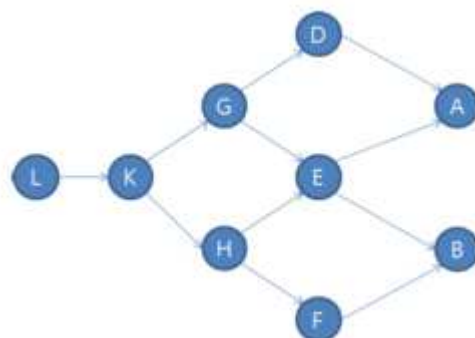


Figura 2.15: Caminos posibles que puede tomar L hasta la frontera eficiente, en el juego de datos de los supermercados

Fuente: (Park et al. 2012)

Se definen dos objetivos que deben cumplir la unidad referencia intermedios

- Gestión de los recursos de forma eficiente.

Cuando se analiza la unidad de referencia de la unidad J, se busca una unidad que consiga obtener como mínimo los mismos niveles de producción con un menor número de recursos. Para ello se realiza una reducción equi-proporcional de todos los recursos con el siguiente modelo:

$$\begin{aligned}
 & \text{Minimizar } \sum_{i=1}^m \alpha_i \\
 & \text{s. a.} \\
 & \lambda_j x_{ij} \leq \alpha_i x_{iJ} \quad i = 1, 2, \dots, m \\
 & \lambda_j y_{kj} \geq y_{kJ} \quad k = 1, 2, \dots, s \\
 & \lambda_j \geq 0
 \end{aligned} \tag{2.52}$$

Definiendo $G_i = x_{ik} - \alpha_i^* x_{ik}$

- Proximidad a la dirección de mejora

Se busca como unidad de referencia (j) aquellas unidades que se encuentren más próximas a la próxima unidad de referencia. Definiendo dirección de mejora de la unidad j (δ_j) a través de la siguiente formulación:

$$\delta_j = 1 - \frac{\cos^{-1} \left(\frac{\sum_{r=1}^m g_r h_r}{\sqrt{\sum_{r=1}^m g_r^2} \sqrt{\sum_{r=1}^m h_r^2}} \right)}{\hat{\delta}} \tag{2.53}$$

Siendo:

- g_r : Diferencia del recurso r entre la unidad evaluada y la última unidad objetivo
- h_r : Diferencia del recurso r entre la unidad evaluada y la unidad j
- $\hat{\delta}$: Umbral definido por el usuario entre 90 y 180.

De forma que δ_j determina la proximidad de la dirección respecto de la dirección de la DMU evaluada y su último objetivo.

Para determinar el camino óptimo desde una unidad ineficiente hasta la frontera eficiente, estableciendo enlaces entre las unidades. De forma que $l_{jj'}$, es un enlace entre la unidad j y la unidad j' que existe en caso de una de ellas sea una unidad de referencia para la otra. En ese caso tiene valor 1 la variable $a_{jj'}$ en otro caso su valor es nulo.

A continuación se desarrolla el modelo que establece las unidades de referencia que componen el camino:

$$\begin{aligned}
 & Sp = \text{Maximizar } \sum_{j=1}^J \sum_{j'=1}^J \delta_{jj'} a_{jj'} \\
 & \text{Minimizar } \sum_{j=1}^J \sum_{j'=1}^J \sum_{i=1}^m G_{ijj'} \\
 & \text{s. a. } \sum_{j'=1}^J a_{jj'} - \sum_{k=1}^J a_{kj} = \begin{cases} 1 & (j = 1) \\ 0 & (j = 2, 3, \dots, J-1) \\ -1 & (j = J) \end{cases} \\
 & a_{jj'} = \{0, 1\} \quad j', j = 1, 2, \dots, J
 \end{aligned} \tag{2.54}$$

3 ANÁLISIS DE REDES COMPLEJAS

“A menudo las personas perfeccionistas y eficaces olvidan la importancia de la eficiencia.”

- Ismael Díaz Lázaro -

Una red compleja es un grafo que está compuesto por nodos que están relacionados entre sí por enlaces. Las redes complejas modelan matemáticamente sistemas de todo tipo que aparecen en la vida real a través de grafos. Esto permite contabilizar que ocurre en la red, cómo se disipa la información, cómo se disipan las enfermedades y cómo se puede evitar, qué provoca el colapso de una red y cómo de estable es según su topología, cómo crece un sistema,.. Estas son algunas de las múltiples preguntas a las que se pretende dar respuesta. En este capítulo se pretende desarrollar los conceptos básicos para poder comprender la técnica que se desarrolla en el siguiente capítulo.

3.1 Tipos de redes y definición de la red

Existen varios tipos de redes dependiendo de la información que aporten los enlaces o arcos, si los arcos tienen una longitud unitaria se dice que es una red sin pesos, en otro caso se dice que es una red con pesos. Por otra parte, si los arcos tienen una dirección, es decir, tienen un nodo origen y un nodo destino simbolizado con una flecha, se dice que la red es dirigida, en otro caso se trata de una red no dirigida.

Por otra parte existen otros tres tipos de redes: las redes multiplexadas en las que existen varios tipos de enlaces, los hipergrafos en los que un enlace puede conectar más de dos vértices a la vez y las redes bipartitas en las que existen dos tipos de nodos y únicamente existen enlaces entre ellos. Estas últimas también se llaman redes de afiliación, porque se usan para determinar qué miembros pertenecen a una organización concreta. Además son fácilmente convertibles a una red normal bien reflejando el número de afiliaciones a las que pertenecen dos personas o bien el número de personas que pertenecen a dos afiliaciones.

Un grafo (V, A) está definido por un conjunto de vértices (V) y un conjunto de arcos (A) , estos arcos pueden ser dirigidos o no dirigidos. Se puede definir un grafo bien a través de una lista de arcos $A = \{(i_1, j_1), (i_2, j_2), \dots, (i_m, j_m)\}$ o bien, a través de una matriz. Existen varios tipos de matrices, a continuación se muestran los tipos de matrices para definir una red sin pesos.

- o Matriz de adyacencia, es una matriz $n \times n$, siendo n el número de nodos, en las que el elemento ij es igual a la unidad si existe un arco que con origen en i y destino en j . Esta matriz será simétrica si se trata de una red no dirigida.
- o Matriz de incidencia para grafos no dirigidos, es una matriz $n \times m$, siendo m el número de arcos, en la que el elemento ij es igual a 1 si el vértice i es uno de los dos extremos del arco j .
- o Matriz de incidencia para grafos dirigidos, es una matriz $n \times m$ en la que el elemento ij es igual a 1 si el vértice i es destino del arco j , o será igual a -1 si el vértice i es el origen del arco j .
- o Matriz de incidencia para grafos bipartitos, es una matriz $n_1 \times n_2$, siendo n_1 el número de nodos de tipo 1 y n_2 el número de nodos de tipo 2, en la que el elemento ij es igual a 1 si existe un enlace entre el nodo i es un nodo de tipo 1 y el nodo j que es un nodo de tipo 2.

Estas matrices también se usan para definir las redes con pesos, la única diferencia es que sus elementos no son iguales a la unidad, sino al peso del arco que se está considerando. Como consecuencia, una red sin pesos es equivalente a una red con pesos iguales a la unidad.

Una vez se dispone de una red dirigida con pesos, se puede convertir en una red no dirigida si se aplica la herramienta de simetría, según (Costa et al. 2007) la matriz que se genera parte de la suma de la matriz de adyacencia original y su traspuesta. También se puede convertir en una red sin pesos considerando que existen todos los arcos de la red original cuyo peso supere un cierto umbral.

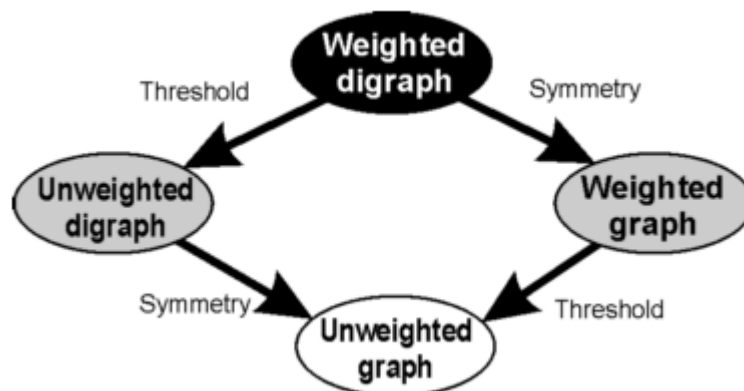


Figura 3.1: Tipos principales de redes complejas y sus transformaciones

Fuente: (Costa et al. 2007)

A continuación se muestran los conceptos de paseo, camino, sendero, ciclo y componentes con el fin de mostrar los diferentes tipos de conjuntos existentes en una red según el criterio de agrupación.

Una sucesión de nodos conectados entre sí constituyen un *paseo*, mientras que en un paseo si ninguno de los nodos se ha recorrido más de una vez se denomina *camino*. La longitud de un camino viene determinada por la suma de los pesos de los arcos que componen el camino, siendo el camino más corto entre un par de nodos el *camino geodésico*.

Un paseo en el que ningún arco se recorre más de una vez se denomina un *sendero*, mientras que un sendero cerrado, es decir que se inicia y finaliza en el mismo nodo, se denomina *ciclo*. Como consecuencia, una red acíclica es una red que no tiene ciclos. Por otra parte, el arco que tiene como origen y destino el mismo nodo se denomina *loop*.

Dentro de una red se define como *componente*, al mayor subgrafo conectado. De forma que todos vértices pertenecen a una única componente, por lo que dos vértices conectados entre sí pertenecen a la misma componente. En el caso de los vértices aislados, cada vértice forma una única componente. La componente con el mayor número de nodos se denomina, *componente gigante*. Matemáticamente se define con la siguiente expresión:

$$(V', A'): \begin{cases} V' \subseteq V \wedge A' \subseteq A \\ i \in V' \wedge j \in V' \Rightarrow ij \text{ están conectados a través de } A' \\ i \in V' \wedge (i, j) \in A \Rightarrow j \in V' \wedge (i, j) \in A' \end{cases} \quad (3.1)$$

En el caso de una red dirigida existen dos tipos de componentes: las componentes débilmente conectadas que no tienen en cuenta el sentido de los arcos y se calculan como si la red fuese no dirigida y las componentes fuertemente conectadas que tienen en cuenta el sentido de los arcos, por lo que dentro de una componente existe un camino entre cada par de nodos del subgrafo.

Dentro de una red dirigida se considera la componente de salida de un nodo, como el conjunto de nodos que pueden ser alcanzados por él, mientras que la componente de entrada de un nodo es el conjunto de nodos que pueden alcanzarlo.

3.2 Caracterización de las redes

Antes de definir los modelos básicos de redes es necesario conocer algunas métricas que permitan caracterizarlas. La nomenclatura que se utilizará en las formulaciones son:

- n : Número de nodos en una red
- m Número de arcos en una red
- A : Matriz de adyacencia que define la red con elementos binarios
- W : Matriz de adyacencia que define los pesos de los arcos de la red
- d_{ij} : Camino geodésico entre el nodo i y el nodo j

A continuación se definen las métricas básicas que caracterizan una red:

- Densidad:

Muestra el ratio entre el número de enlaces existentes en una red y el número posible de enlaces. En una red no dirigida el número de posibles enlaces existentes en una red es la mitad que en una red dirigida para el mismo número de nodos.

Tabla 3.1. Densidad

Red no dirigida	Red dirigida
$\rho = \frac{m}{n(n-1)/2}$	$\rho = \frac{m}{n(n-1)}$

- Grado del nodo i :

Existen varios tipos de grado, en caso de las redes no dirigidas se habla de grado (k_i) al número de nodos que están conectados con el nodo i . En el caso de las redes dirigidas se habla de grado de entrada (k_i^{in}) al número de arcos que tienen como destino el nodo i , mientras que el grado de salida (k_i^{out}) es el número de arcos que

tienen como origen el nodo i . El grado total representa el número de conexiones que tiene el grado i y se calcula como la suma del grado de entrada y el grado de salida.

Tabla 3.2. Grado

Red no dirigida	Red dirigida
$k_i = \sum_j A_{ij}$	$k_i^{in} = \sum_i A_{ij}$
	$k_i^{out} = \sum_j A_{ij}$
	$k_i = k_i^{out} + k_i^{in}$

- o Fuerza del nodo i (s_i).

En el caso de las redes no dirigidas representa la longitud total de los arcos que conectan al nodo i con la red. Si se considera el caso de las redes dirigidas, se hace la misma distinción que se ha realizado en la métrica del grado. El sumatorio de las longitudes de los arcos que tienen como origen el nodo i es el grado de entrada (s_i^{in}), mientras que el grado de salida (s_i^{out}) es el sumatorio de las longitudes de los arcos que llegan al nodo i . Siendo el sumatorio de ambos la fuerza total del nodo i (s_i).

Tabla 3.3. Fuerza

Red no dirigida	Red dirigida
$s_i = \sum_j W_{ij}$	$s_i^{in} = \sum_i W_{ij}$
	$s_i^{out} = \sum_j W_{ij}$
	$s_i = s_i^{out} + s_i^{in}$

- o Camino medio:

Determina la longitud media de los caminos más cortos entre cualquier par de nodos. En caso de que se considere que la matriz entre dos nodos que no están conectados es ∞ , se debe considerar únicamente en la siguiente formulación los caminos entre los nodos que sí están conectados.

Tabla 3.4. Camino medio

Red no dirigida	Red dirigida
$\langle d \rangle = \frac{\sum_{i \neq j} d_{ij}}{n(n-1)}$	$\langle d \rangle = \frac{\sum_{i \neq j} d_{ij}}{n(n-1)/2}$

- o Diámetro

El diámetro representa la máxima distancia geodésica que se observa en toda la red para cada par de nodos. Al igual que en el cálculo del camino medio, si se considera que la distancia entre dos nodos conectados es ∞ , sólo se considerarán los caminos entre nodos que sí están conectados.

Tabla 3.5. Diámetro

Red no dirigida	Red dirigida
$D = \max_{j>i} d_{ij}$	$D = \max_{j \neq i} d_{ij}$

- Eficiencia:

Esta medida determina que la eficiencia con la que manda información el nodo i al nodo j es inversamente proporcional a la distancia que hay entre ellos

Tabla 3.6. Eficiencia

Red no dirigida	Red dirigida
$\eta = \frac{\sum_{i \neq j} d_{ij}}{n(n-1)}$	$\eta = \frac{\sum_{i \neq j} d_{ij}}{n(n-1)/2}$

Esta métrica se corresponde con la inversa de la media armónica.

- Coeficiente de clustering del nodo i :

Es una métrica que permite analizar si los nodos vecinos del nodo i están conectados entre sí, dependiendo del tipo de red que se esté analizando en la literatura existen múltiples variaciones como se puede ver en (Saramäki et al. 2007), en este apartado se van a desarrollar los coeficientes de clustering recogidos por (Fagiolo 2007).

Tabla 3.7. Coeficiente de Clustering red sin pesos

Red no dirigida	Red dirigida
$C_i = \frac{\frac{1}{2} \sum_{j \neq i} \sum_{h \neq (i,j)} a_{ij} a_{ih} a_{jh}}{\frac{1}{2} k_i (k_i - 1)}$ $= \frac{(A^3)_{ii}}{k_i (k_i - 1)}$	$C_i = \frac{\frac{1}{2} \sum_j \sum_h (a_{ij} + a_{ji})(a_{ih} + a_{hi})(a_{jh} + a_{hj})}{k_i^{tot} (k_i^{tot} - 1) - 2d_i^{\leftrightarrow}}$ $= \frac{(A + A^T)_{ii}^3}{2[k_i^{tot} (k_i^{tot} - 1) - 2d_i^{\leftrightarrow}]}$

Siendo:

- $(A^3)_{ii}$ el elemento de la diagonal i en la matriz A^3
- A^T la traspuesta de la matriz A
- $\frac{1}{2} k_i (k_i - 1)$ el número de posibles triángulos que pueden existir como máximo
- d_i^{\leftrightarrow} el número de enlaces que son bidireccionales (hay que tenerlos en cuenta para eliminar los falsos triángulos que se obtendrían en caso de no considerarlos).

El nodo i tiene la posibilidad de formar dos triángulos con cada pareja de vecinos, teniendo como mucho $\frac{k_i^{tot}(k_i^{tot}-1)}{2}$ parejas con las que generar un triángulo. Los falsos triángulos se generan por la pareja de enlaces que generan el enlace bidireccional, por esa razón se eliminan 2 posibles triángulos por enlace bidireccional.

En el caso de que se consideren que los enlaces de la red tengan pesos, con el fin de determinar cuál es el peso de cada vecindad, basándose en el concepto de intensidad de subgrafo, definido como la media geométrica de los pesos de los enlaces del subgrafo.

Tabla 3.8. Coeficiente de Clustering red con pesos

Red no dirigida	Red dirigida
$C_i = \frac{\frac{1}{2} \sum_{j \neq i} \sum_{h \neq (i,j)} \hat{w}_{ij}^{1/3} \hat{w}_{ij}^{1/3} \hat{w}_{jh}^{1/3}}{\frac{1}{2} k_i (k_i - 1)} = \frac{(\hat{W}^{1/3})_{ii}^3}{k_i (k_i - 1)}$	$C_i = \frac{(\hat{W}^{1/3} + (\hat{W}^T)^{1/3})_{ii}^3}{2[k_i^{tot}(k_i^{tot} - 1) - 2d_i^{\leftrightarrow}]}$

Siendo

- \hat{w} la matriz w normalizada, $w_{ij} \in [0,1]$
- $\hat{w}^{1/3}$ la matriz \hat{w} a cuyos elementos se les ha aplicado la cúbica.
- Transitividad:

Una forma de analizar los ciclos de grado 3 de forma global es a través del ratio entre número de triángulos existentes (ciclos compuestos por tres arcos) y el número de tripletas, siendo una tripleta un conjunto formado por dos arcos que conectan al mismo nodo.

Tabla 3.9. Transitividad

Red no dirigida	Red dirigida
$T = \frac{3 \cdot \sum_{k>j>i} A_{ij} A_{ik} A_{jk}}{\sum_{k>j>i} (A_{ij} A_{ik} + A_{jk} A_{ji} + A_{ki} A_{kj})}$	$T = \frac{\sum_{k \neq j \neq i} A_{ij} A_{ik} A_{jk}}{\sum_{k \neq j \neq i} A_{ij} A_{ik}}$

- Centralidad:

La posición de un nodo respecto de los demás nodos de la red, resulta determinante para el control de la información y dependiendo del punto de vista con que se mire, un nodo será más central que otro. A continuación se muestran algunos de los más relevantes:

- ✓ Centralidad según el grado: Se considera que un nodo es central cuantas más conexiones tenga, porque es capaz de recibir y transmitir información fácilmente. Sin embargo no se tiene en cuenta como son los nodos a los que está conectado.

Tabla 3.10. Centralidad según el grado

Red no dirigida	Red dirigida
$DC_i = \frac{k_i}{n - 1}$	$DC_i = \frac{k_i^{in} + k_i^{out}}{n - 1} = \frac{k_i}{n - 1}$

- ✓ Centralidad eigenvector: A diferencia de la centralidad según el grado tiene en cuenta con quién está conectado el nodo, cuanto más importante sean sus conexiones, más importante será. De forma matemática la centralidad del nodo i se expresaría de la siguiente forma: $x_i = \frac{1}{\lambda} \sum_j A_{ij} x_j$, siendo λ el autovalor máximo de A .
- ✓ Centralidad de intermediación: Determina con qué frecuencia se encuentra un nodo en el camino más corto entre cada par de nodos. Siendo $\zeta(j, p)$ el número de caminos más cortos entre j y p y $\zeta(j, p|r)$ el número de caminos más cortos entre j y p que pasan por r .

Tabla 3.11. Centralidad de intermediación

Red no dirigida	Red dirigida
$\beta_r = \frac{1}{(n-1)(n-2)} \sum_{\substack{j>p \\ j \neq r \neq p}} \frac{\varsigma(j,p r)}{\varsigma(j,p)}$	$\beta_r = \frac{2}{(n-1)(n-2)} \sum_{\substack{j>p \\ j \neq r \neq p}} \frac{\varsigma(j,p r)}{\varsigma(j,p)}$

- Vulnerabilidad

Ante la desaparición de un nodo, no todas las redes se comportan de la misma forma, en unas la desaparición de unos pocos nodos puede suponer la desconexión de la red, mientras que otras son más robustas. Por otra parte, algunas redes presentan propiedades jerárquicas, es decir los nodos más cruciales se encuentran en las posiciones de la jerarquía más elevadas. Con el fin de determinar que nodos son más críticos se define en (Gol'dshtein et al. 2004) como la vulnerabilidad de un nodo i al ratio que determina la pérdida de eficiencia de la red cuando desaparece el nodo i y los enlaces que lo conectan.

$$V_i = \frac{E - E_{-i}}{E} \quad (3.2)$$

Siendo la distribución máxima de todos los vértices, la vulnerabilidad de la red. Por otra parte, la distribución de la vulnerabilidad determina si una estructura es jerárquica, si todos los vértices tienen la misma vulnerabilidad estamos ante una estructura no jerárquica. Además Gol'dshtein afirma que existe una relación entre las estructuras simétricas y las estructuras jerárquicas; sin embargo no es directa.

- Asortatividad:

Una red es asortativa si se produce un mayor número de conexiones entre nodos del mismo tipo, en las redes sociales al hecho de que dos personas con características afines (religión, educación...) se le denomina homofilia. Para ello se genera una matriz E en la que cada elemento (e_{st}) representa el número de arcos que conectan los vértices de tipo s con vértices del tipo t .

$$\hat{E} = \frac{E}{\|E\|} \quad (3.3)$$

Siendo $\|E\|$ la suma de los elementos de la matriz E , por lo que \hat{E} es la matriz E normalizada. Por lo que la probabilidad de que un vértice s tenga como vecino un vértice t , se rige por la siguiente expresión.

$$P(t|s) = \frac{\hat{e}_{st}}{\sum_u \hat{e}_{su}} \quad (3.4)$$

$$\sum_t P(t|s) = 1$$

De forma que la asortatividad se define con el siguiente ratio, si se quiere dar el mismo peso a cada grupo, siendo N_t el número de grupos existentes en la red.

$$0 \leq \hat{Q} = \frac{\sum_s P(s|s) - 1}{N_t - 1} \leq 1 \quad (3.5)$$

Siendo $\hat{Q} = 1$ en caso de redes completamente asortativas y $\hat{Q} = 0$ en el caso de redes aleatorias. Sin embargo, si el tamaño de los grupos es significativo, se utiliza la siguiente expresión, que asigna el mismo peso a todos los nodos.

$$0 \leq \mathbb{Q} = \frac{\text{Traza}(e) - \|e^2\|}{1 - \|e^2\|} \leq 1 \quad (3.6)$$

Siendo $\mathbb{Q} = 1$ en el caso de ser una red asortativa y $\mathbb{Q} = 0$ en caso de ser una red aleatoria.

3.3 Modelos de redes

Con el fin de estudiar las propiedades topológicas de las redes reales se han generados múltiples modelos; sin embargo en este apartado se van a presentar los más básicos reflejados en (Newman 2003).

3.3.1 Redes aleatorias

Las redes aleatorias son el modelo más básico de las redes complejas, fue formulada en (Erdős & Rényi 1959), donde definen una red a través n vértices desconectados que se van conectando con m enlaces que se añaden evitando los loops. Otros modelos consideran que se parten de n vértices no conectados entre los que se genera un enlace entre cada par de nodos con una probabilidad p . Este último se conoce como el modelo de Erdős-Rényi (ER) y sus enlaces se generan según distribución binomial $B(\binom{n(n-1)}{2}, p)$, de forma que la distribución del grado de los nodos sigue una binomial $B(n-1, p)$. Sin embargo, cuando el número de nodos tiende a ∞ los enlaces se distribuyen según una Poisson de media $\langle k \rangle = p(n-1)$.

En función de la probabilidad que se emplee la red estará más o menos conectada. Si $p = 1/n$ el grado medio es 1 generándose una gran componente, si la probabilidad es mucho más inferior se obtiene una red con muchos nodos aislados. Si $p = \ln(n)/n$, entonces el grado medio de la red es $\langle k \rangle \approx \ln(n)$ desarrollando una red completamente conectada.

Como características principales de la red, cabe decir que la asortatividad es nula y al igual que el grado medio, el camino medio y el clustering medio dependen de la probabilidad con la que se generan los arcos, rigiéndose por la siguiente expresión.

$$\begin{aligned} \langle d \rangle &\approx \frac{\log(n)}{\log(k)} \quad \text{para } n \rightarrow \infty \\ CC &\cong p = \frac{\langle k \rangle}{n-1}, \text{ siendo nulo para } n \rightarrow \infty \end{aligned} \quad (3.7)$$

3.3.2 Redes de mundo pequeño

Muchas de las redes reales estudiadas presentan una media de caminos geodésicos baja, el hecho de que un nodo pueda alcanzar otro a través de un número pequeño de arcos fue descubierto por Milgram en 1967 que afirmó que los ciudadanos de EEUU estaban conectados de media por 6 conocidos. También se ha observado que estas redes tienen un número elevado de ciclos compuestos por tres arcos, es decir tienen un alto clustering.

En el artículo (Watts & Strogatz 1998), se definieron las redes de mundo pequeño con el fin de generar una red que pudiera representar las propiedades descritas anteriormente, sin embargo hasta 1998 se analizaban redes o completamente aleatorias o completamente regulares, siendo estas últimas aquellas cuyos nodos tienen el mismo grado cuyo valor es mucho más pequeño que el número de nodos de la red. Las redes aleatorias tienen la media de los caminos geodésicos pequeña pero no tienen un elevado clustering, mientras que las redes regulares presentan un alto clustering y una media de caminos geodésicos elevada.

Para generar estas redes, partían de un entramado regular con n vértices, donde cada vértice estaba conectado con sus k vecinos más cercanos en cada dirección, contabilizando un grado total de valor $2k$, siendo $N \gg k \gg \log(N) \gg 1$. Cada enlace de la red era reordenado con una probabilidad p , de forma que si $p = 0$ no se producía ninguna reordenación y si $p = 1$ se generaba una red aleatoria.

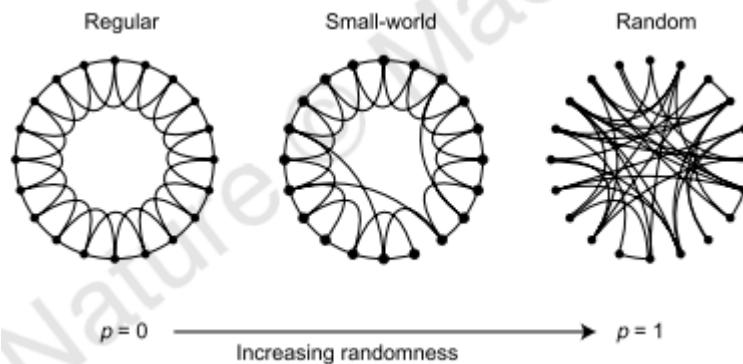
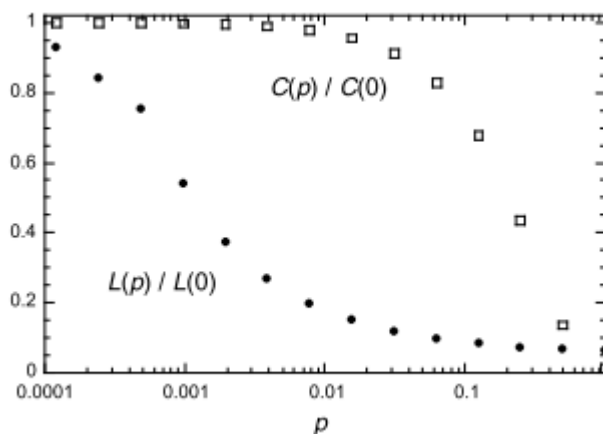


Figura 3.2 Generación de redes de mundos pequeños

Fuente: (Watts & Strogatz 1998)

En la siguiente figura se muestra como se produce la variación de las métricas de la media de la distancia geodésica y el clustering al representar el ratio entre una red que ha sufrido una reordenación de sus enlaces con una probabilidad p y una red que no ha sufrido ninguna reordenación.

Figura 3.3: Representación de la variación de la media de la longitud geodésica y el clustering en función de la probabilidad p

Fuente: (Watts & Strogatz 1998)

3.3.3 Modelos de configuración

Una forma de estudiar las redes reales es comparando sus características con las de redes aleatorias similares, buscando que tengan el mismo grado de distribución. El método más común es generar el número de nodos con el grado deseado para obtener la distribución del grado deseada. Para generar ese grado se utilizan unos tipos de enlaces, llamados stubs que no unen al nodo que están asociados con ningún otro. Aleatoriamente, se seleccionan un par de stubs y se unen para formar un enlace.

Otra forma para generar una red aleatoria dirigida, es partiendo de la red real y seleccionando un par de enlaces que intercambian los nodos que conectan. Sin embargo, cuando se genera una red aleatoria no dirigida se coge un arco y se cambia uno de los extremos con otro.

3.3.4 Modelo libre de escala

Como característica común en muchas de las redes reales con un gran número de nodos, se observa que la distribución del grado de los nodos sigue una distribución de leyes de potencia $P(k) \sim k^{-\gamma}$. Esta distribución muestra que un gran número de nodos tiene pocos enlaces, mientras que un pequeño número de nodos tienen un gran número de enlaces que actúan como hubs. A estas redes se las denomina redes libre de escala.

En el artículo (Barabási & Albert 1999), se afirma que este comportamiento se debe a dos características: el crecimiento continuo de la red con nuevos vértices y que los nuevos vértices tienden a conectarse con nodos que ya están bien conectados. Con estas dos características se generó el modelo de red de Barabasi-Albert. Para obtener esta red se parte de un conjunto de m_0 vértices y a cada paso se genera un nuevo vértice con m enlaces que le conectan con los vértices existentes según una probabilidad proporcional al grado que tengan. Como por ejemplo $P(i \rightarrow j) = \frac{k_j}{\sum_u k_u}$. Siguiendo esta regla conocida como “preferential attachment” se observa el paradigma que aquellos nodos que están más conectados consiguen más conexiones y por ello tienen más probabilidad de conseguir más. A este paradigma se le denomina “rich get richer”.

4 METODOLOGÍA DE ANÁLISIS DE REDES DE DOMINANCIA

“Saber dónde encontrar la información y cómo usarla, éste es el secreto del éxito.”

Albert Einstein

El análisis de redes complejas ofrece una herramienta muy versátil para analizar un sistema. En este trabajo se pretende implementar esta herramienta para analizar los resultados obtenidos con el análisis de envoltura de datos. Se podrán implementar diferentes índices y filtros propios de la técnica de análisis de redes y desarrollar otros nuevos. Para ello se generará una red formada por las DMUs que estarán relacionadas entre sí según la eficiencia relativa existente entre ellas. Esta red permitirá comprender las relaciones entre las unidades

4.1 Técnicas que emplean el Análisis de Redes Complejas y Análisis de Envoltura de Datos

En la literatura se pueden encontrar estudios que utilizan tanto el Análisis de Redes Complejas, como el Análisis de Envoltura de Datos, desarrollando novedosos puntos de vista que permiten ampliar la forma de ver los conceptos dentro de sus propias disciplinas.

A continuación se exponen diferentes técnicas que emplean ambas técnicas.

4.1.1 Análisis de redes de colaboración

En el artículo (Lee et al. 2012), se realiza un estudio acerca de las Instituciones de Investigación Públicas (Public Research Institutions, PRI) en Corea en el ámbito de la ciencia y la ingeniería. Su objetivo es determinar cuál es el impacto de las estructuras de colaboración entre las diferentes instituciones, en la producción de dichas instituciones. Para ello realizan una correlación entre las redes de colaboración y la productividad entre los años 2000 y 2010

Para identificar las cooperaciones entre instituciones se basan en la autoría de los artículos científicos que publicaban y estaban registrados en Scopus. Con estos datos generan la red con el fin de analizar la posición de cada institución respecto a las demás. Con las herramientas de CNA, como la densidad, la eficiencia y el coeficiente de intermediación, determinan como se comporta la red desde el punto de vista estructural. A mayor densidad, la información se transmite mejor; a mayor eficiencia la información se transmite a muchas instituciones con un número limitado de enlaces y a mayor coeficiente de intermediación, mayor control tiene una institución sobre la información que posee. Por otra parte, analizan como son las relaciones a través del eigenvector y la centralidad por cercanía. A mayor eigenvector, mayor capacidad de coordinación entre las instituciones y a mayor la

centralidad por cercanía, mayor será el impacto de la situación actual de la institución, así como sus capacidades potenciales.

Con el fin de analizar la productividad, utilizan el modelo ventana de DEA para analizar la eficiencia de cada institución en el tiempo, considerando que una misma unidad en dos periodos diferentes, son dos unidades diferentes y por ello se comparan entre ellas. Como entradas consideran tanto el número de empleados a tiempo completo, como el coste de la investigación, teniendo en cuenta la deflación del PIB. Como salidas consideran el número de artículos científicos y de patentes así como el dinero generado por las licencias tecnológicas.

Gracias a sus resultados, se determina que una red bien conectada aumenta la productividad; sin embargo una conexión excesiva dificulta la innovación y el desarrollo de nuevas tecnologías.

4.1.2 Evaluación de nodos eficientes usando CNA

En el artículo (Liu et al. 2009), se propone utilizar la métrica centralidad eigenvector para realizar un ranking de las unidades eficientes obtenidas en el Análisis de Envoltura de Datos. Para ello generan una red dirigida con pesos, donde los nodos son las DMUs y los enlaces dirigidos las relaciones entre las unidades generadas por DEA, siendo el peso de estos enlaces proporcional los valores lambda de dichas relaciones.

Con esta metodología determinan que los nodos eficientes más importantes son aquellos que tienen más enlaces, con más peso y con nodos adyacentes bien conectados. De forma que, aquellas unidades que se encuentren mal conectadas, se consideran unidades especializadas.

Con el fin de determinar unos pesos de los arcos representativos de las relaciones existentes, se generan múltiples cálculos de DEA utilizando todas las combinaciones posibles de entradas y de salidas, para luego agregar los valores lambda y calcular que DMUs son más eficientes, independientemente de la combinación de entradas y salidas que se considere. A continuación se muestra la matriz de pesos que se consideraría para generar la red.

$$A = \left[\sum_{t=1}^w \lambda_{jk}^t \right] \quad (4.1)$$

Siendo λ_{jk}^t el valor lambda entre las unidades j y k , siendo k la unidad analizada en el modelo t .

En el artículo (Liu & Lu 2010), se realiza una mejora sobre este modelo que tiende a beneficiar a las DMUs pequeñas, normalizando las lambdas en función del tamaño de la unidad como se muestra en la siguiente expresión

$$\begin{aligned} IW_{ij}^{t,k} &= \frac{\lambda_{jk}^t x_{ij}^t}{\sum_{j \in E} \lambda_{jk}^t x_{ij}^t}, \quad 0 < IW_{ij}^{t,k} < 1 \\ OW_{rj}^{t,k} &= \frac{\lambda_{jk}^t y_{rj}^t}{\sum_{j \in E} \lambda_{jk}^t y_{rj}^t}, \quad 0 < OW_{rj}^{t,k} < 1 \\ IOW_{jk}^t &= \frac{1}{(m+s)} \left(\sum_{i=1}^m IW_{ij}^{t,k} + \sum_{r=1}^s OW_{rj}^{t,k} \right) \end{aligned} \quad (4.2)$$

Y agregándolas, para generar el peso de los arcos dirigidos de la red, según esta otra expresión:

$$A = \left[\sum_{t=1}^w IOW_{jk}^t \right] \quad (4.3)$$

Siendo A_{jk} el arco que va de la DMU k a la DMU j . Esto permite eliminar el efecto que produce el tamaño de la organización y permite que se aplique esta metodología tanto para modelos CRS, como modelos VRS.

Una aplicación práctica de este modelo ha sido desarrollada en (Ho et al. 2014), donde estudian la eficiencia de la transferencia tecnológica de las universidades a través de un modelo DEA de dos etapas en el que primero se

analizaba la innovación científica y finalmente la creación de valor de las universidades. La agrupación de los lambdas tras realizar las distintas combinaciones entre los factores de entrada, de salida e intermediarios de las DMUs se agrupan de forma normalizada. Gracias a la formación de la red se puede realizar un ranking de todas las unidades en función de cada factor de entrada o de salida, determinar que factor influye más a la hora de determinar que unidad es eficiente y cuáles permiten definir los targets intermedios.

4.2 Redes de dominancia

Las redes de dominancia son redes dirigidas con pesos que muestran la relación de eficiencia entre cada par de DMUs, obtenida con las técnicas del Análisis de Envoltura de Datos. La red de dominancia estará compuesta por unos nodos que representan las unidades de decisión y unos enlaces cuya longitud muestra la eficiencia relativa entre los nodos origen y destino. El nodo origen es la unidad dominada, mientras que el nodo destino es la unidad dominante. En este trabajo se utiliza una tecnología FDH para evaluar la eficiencia global de la muestra.

A continuación se desarrolla la metodología que se presenta en este trabajo.

4.2.1 Construcción de la red y definición de conjuntos

El sistema va a estar compuesto por n DMUs, que forman el conjunto D , cada una de ellas se evalúa en función de m entradas y s salidas, que se denotan con los siguientes vectores $\bar{x}_j = (x_{1j}, x_{2j}, \dots, x_{mj})$

$\bar{y}_j = (y_{1j}, y_{2j}, \dots, y_{mj})$, siendo la media de los diferentes inputs y outputs $x_i^{aver} = \frac{1}{n} \sum_{j \in D} x_{ij}$ e

$y_k^{aver} = \frac{1}{n} \sum_{j \in D} y_{kj}$. El conjunto de nodos que dominan al nodo r se define de la siguiente forma

$D(r) = \{j \neq r: x_{ij} \leq x_{ir} \forall i \text{ \& } y_{kj} \geq y_{kr}\} \subset D$ Por lo tanto, para que exista un enlace con origen en r y destino en j , j debe dominar a r , es decir, $j \in D(r)$ De forma que se define como red de dominancia (D, E) donde siendo $E = \{(r, j): r \in D \wedge j \in D(r)\}$ el conjunto de enlaces.

La longitud de los arcos, viene definida por la eficiencia relativa obtenida con DEA, en este trabajo se emplea una métrica aditiva que se muestra en la fórmula (4.4), que adimensionaliza los ratios utilizando la media.

$$e_{rj} = \begin{cases} 0 & \text{si } j \notin D(r) \\ \sum_i \frac{x_{ir} - x_{ij}}{x_i^{aver}} + \sum_k \frac{y_{kj} - k_{kr}}{y_r^{aver}} & \text{si } j \in D(r) \end{cases} \quad (4.4)$$

Sin embargo, se puede aplicar otras métricas, como RAM o MIP, como se muestra respectivamente en las siguientes fórmulas (4.5)

$$e_{rj} = \begin{cases} 0 & \text{si } j \notin D(r) \\ \sum_i \frac{x_{ir} - x_{ij}}{R_i^-} + \sum_k \frac{y_{kj} - k_{kr}}{R_k^+} & \text{si } j \in D(r) \end{cases} \quad (4.5)$$

$$e_{rj} = \begin{cases} 0 & \text{si } j \notin D(r) \\ \sum_i \frac{x_{ir} - x_{ij}}{x_{ij}} + \sum_k \frac{y_{kj} - k_{kr}}{y_{kj}} & \text{si } j \in D(r) \end{cases}$$

Una propiedad que tienen las redes de dominancia debido a la forma en la que se formula la existencia de los enlaces es la propiedad de *transitividad*. Esta propiedad implica que si existe un arco entre (r,p) y entre (p,j) , $p \in D(r) \wedge j \in D(p) \Rightarrow j \in D(r)$, es decir, si p domina a r y j domina a p , entonces j domina a r . Debido a que un nodo sólo puede dominar a otro si consume menos recursos y/o produce más salidas. Como consecuencia se genera una red acíclica.

Una propiedad importante que tiene la métrica propuesta y la métrica basada en RAM al tener el denominador constante, es la propiedad de *aditividad*. Esta propiedad implica que la suma de los enlaces que componen el

camino existente entre los nodos r y j es igual a la longitud del enlace que une los nodos r y j . Es decir, $e_{rp} + e_{pj} = e_{rj}$. Debido a la propiedad de aditividad todos los caminos existentes entre dos nodos son iguales y por ello todos los caminos son geodésicos.

A partir de la red de dominancia dirigida con pesos (D, E) se puede generar la red no dirigida sin pesos. (D, A) Siendo A el conjunto de enlaces representados en la matriz de incidencia según la siguiente expresión.

$$a_{rj} = a_{jr} = \begin{cases} 1 & \text{si } e_{rj} > 0 \Leftrightarrow (r, j) \in E \\ 0 & \text{en otro caso} \end{cases} \quad (4.6)$$

La red (D, A) permite determinar las componentes de la red que está formada por el máximo conjunto de nodos conectados entre sí, de forma que se existe un camino entre cada par de nodos. Cada componente c está formada por los nodos D_c y los enlaces no dirigidos que los conectan $E_c = \{(r, j) \in E : r \in D_c \wedge j \in D_c\} \subseteq E$ siendo la componente c (D_c, E_c) un subgrafo de la red (D, A) . Cada nodo sólo puede pertenecer a una componente, por lo que $D_c \cap D_{c'} = \emptyset \forall c \neq c' \Rightarrow \bigcup_c |D_c| = |D| = n$, siendo el número máximo de componentes NC . Cada nodo aislado se considera una componente.

4.2.1.1 Conjuntos de nodos eficientes

Los nodos eficientes de la red que no están dominados forman el conjunto $D^* = \{r : D(r) = \emptyset\} \subseteq D$. El conjunto de nodos eficientes del nodo r se define de la siguiente forma $D^*(r) = \begin{cases} r & \text{if } r \in D^* \\ D(r) \cap D^* & \text{if } r \notin D^* \end{cases}$. El conjunto de enlaces que proyectan al nodo r sobre la frontera eficiente es $E^*(r) = \{(r, j) : j \in D^*(r)\} \subseteq E(r)$.

De todos los nodos eficientes que dominan a r los nodos más cercanos a r se agrupan en el conjunto $J^*(r) = \arg \min_{j \in D^*(r)} e_{rj}$. Por otra parte una vez que se han definido las componentes de la red, se pueden determinar los nodos eficientes que pertenecen a un conjunto a través de la intersección de conjuntos $D_c^* = \{j : j \in D_c \cap D^*\} = \bigcup_{r \in D_c} D^*(r) = D^*$

4.2.1.2 Conjuntos de nodos dominados

Con el fin de definir los nodos que son dominados por el nodo r se define el siguiente conjunto $D^{-1}(r) = \{j \neq r : x_{ij} \geq x_{ir} \forall i, y_{kj} \leq y_{kr} \forall k\} \subset D$ a este subconjunto está asociado el siguiente $E^{-1}(r) = \{(q, r) : q \in D^{-1}(r)\}$

Por otra parte resulta significativo determinar cuáles son los nodos que únicamente tienen una unidad de referencia dentro de la frontera eficiente, para ello se define el conjunto $D^{-1}(j)$ siendo j una unidad de la frontera eficiente $D^{-1}(j) = \{r : r \in D^{-1}(j) \wedge D(r) \cap D^* = \{j\}\} \subseteq D^{-1}(j)$

4.2.1.3 Estratificación

Gracias a las relaciones de dominancia, los nodos ineficientes encuentran en los nodos que los dominan, un objetivo a seguir. Según la métrica que se ha propuesto todos los caminos entre cada par de nodos tienen la misma distancia; sin embargo el número de nodos intermedios que existen en cada camino varía.

A cada nodo que compone la red, se le va a asignar un atributo, llamado capa, que va a determinar el número máximo de targets intermedios que tiene que alcanzar para encontrarse con la frontera eficiente. De forma matemática se expresaría con la siguiente fórmula.

$$\lambda(r) = \begin{cases} 0 & \text{si } D(r) = \emptyset \\ 1 + \max_{j \in D(r)} \lambda(j) & \text{en otro caso} \end{cases} \quad (4.7)$$

Según esta definición, todos los nodos que pertenecen a la capa 0 componen la frontera eficiente y si esta capa fuera eliminada, la nueva frontera eficiente estaría compuesta por los nodos de la capa 1.

El conjunto de nodos pertenecientes a una capa (q) y a una componente concreta (c) se define como $L_{cq} = \{r \in D_c : \lambda(r) = q\}$

4.2.2 Índices

Con el fin de representar la posición de un nodo respecto de los demás y caracterizar las relaciones entre ellos, se han desarrollado unas métricas, basadas algunas de ellas en la técnica de Análisis de Redes Complejas. Estos índices se agrupan en cuatro categorías dependiendo de a qué nivel se realiza la observación. Los niveles que se consideran son: nivel de nodo, nivel de capa, nivel de componente y nivel de red.

4.2.2.1 Medidas a nivel de nodo

Este tipo de índices se centran en el nodo y sus relaciones más inmediatas y representar su distribución gráficamente son de gran utilidad. A continuación se explican las métricas que se aplican a todos los nodos por igual.

- Grado de entrada del nodo r :

Este índice determina el número de nodos que domina, es decir el número de enlaces que tienen como destino el nodo r . Representa el número de veces que el nodo r es un objetivo para un nodo más ineficiente que él.

$$d_r^{in} = |D^{-1}(r)| \quad (4.8)$$

- Grado de salida del nodo r :

El grado de salida determina el número de veces que un nodo es dominado, es decir, es el número de enlaces que tienen como origen el nodo r .

$$d_r^{out} = |D(r)| \quad (4.9)$$

- Especificidad del nodo r :

Determina el número de enlaces que inician o finalizan en el nodo r , es decir, el número de relaciones existentes con el nodo r

$$\eta_r = |D(r)| + |D^{-1}(r)| = d_r^{in} + d_r^{out} \quad (4.10)$$

- Índice hub del nodo r :

Este índice toma valores mayores cuantas más veces domine y sea dominado el nodo r , siendo nulo cuando sea un nodo extremo de la red.

$$\gamma_r = |D(r)| \cdot |D^{-1}(r)| = d_r^{in} \cdot d_r^{out} \quad (4.11)$$

- Número de nodos eficientes que dominan al nodo r :

Como su propio nombre indica, el número de nodos que dominan al nodo r y pertenecen a la frontera eficiente.

$$|D^*(r)| \quad (4.12)$$

- Máxima distancia del nodo r a la frontera eficiente:

Determina la eficiencia relativa máxima entre el nodo r y los nodos pertenecientes a la frontera eficiente que lo dominan

$$e_r^{max} = \max_{j \in D^*(r)} e_{rj} \quad (4.13)$$

- Mínima distancia del nodo r a la frontera eficiente:

Determina la eficiencia relativa mínima entre el nodo r y los nodos pertenecientes a la frontera eficiente que lo dominan

$$\tau_r^{min} = \min_{j \in D^*(r)} e_{rj} \quad (4.14)$$

- Dentro de esta categoría se definen otro tipo de índices para los nodos que pertenecen a la frontera eficiente
 - Importancia del Benchmark j

Determina el número de nodos que domina el nodo eficiente j

$$d_j^{in} = |D^{-1}(j)| \quad (4.15)$$

- Necesidad de Benchmark j

Este índice muestra el número de nodos que únicamente tiene como target dentro de la frontera eficiente al nodo j

$$|D^{-1}(j)| \quad (4.16)$$

- Potencial del Benchmark j

El número de nodos que domina un nodo eficiente es importante; sin embargo también es importante cuál es la eficiencia relativa entre las unidades. Una forma para analizar esa distancia es el sumatorio de la longitud de los arcos que llegan al punto de referencia j , como una

$$\kappa_j = \sum_{r \in D^{-1}(j)} e_{rj} \quad (4.17)$$

- Radio de ineficiencia del Benchmark j

Este índice determina la eficiencia relativa máxima entre el Benchmark y uno de sus nodos dominados, para poder analizar cómo de lejos se encuentran los nodos ineficientes que se proyectan en esa parte de la frontera eficiente

$$\sigma_j = \max_{r \in D^{-1}(j)} e_{rj} \quad (4.18)$$

- Índice de supereficiencia del Benchmark j

Muestra el número de nodos ineficientes que tienen como objetivo el target referencia j y se encuentran en la capa 1, por lo que estas unidades pasarían a formar parte de la frontera eficiente en caso de que desapareciera la unidad j

$$\varphi_j = |\{r : D(r) = \{j\} \wedge \lambda(r) = q\}| \quad (4.19)$$

- Coeficiente de intermediación del nodo i

Muestra el número de veces el nodo i aparece el camino más corto entre cada par de nodos, representando el número de veces que aparece i como objetivo intermedio en los caminos geodésicos de la componente c

$$\beta_i = \begin{cases} 0 & \text{si } |D(i)| = 0 \vee |D^{-1}(i)| \\ \sum_{\substack{j, r \in D_c \\ j \neq r \neq i}} \frac{\xi(j, r|i)}{\bar{\xi}(j, r)} & \text{en otro caso} \end{cases} \quad (4.20)$$

- Coeficiente de clustering del nodo i

Determina si los nodos que dominan al nodo i y son dominados por i tienen enlaces entre sí.

$$CC_i = \begin{cases} 0 & \text{si } |D(j)| + |D^{-1}(j)| \leq 1 \\ \frac{|D^{-1}(i)| * |D(i)| + \sum_{r \in D^{-1}(j)} |D(r) \cap D^{-1}(i)| + \sum_{r \in D(j)} |D(i) \cap D^{-1}(r)|}{\binom{|D(i)| + |D^{-1}(i)|}{2}} & \text{en otro caso} \end{cases} \quad (4.21)$$

Siendo:

- $\binom{|D(i)| + |D^{-1}(i)|}{2}$ El número de posibles arcos entre los nodos vecinos del nodo i
- $|D^{-1}(i)| * |D(i)|$ Los arcos existentes entre los nodos que dominan al nodo i y los nodos dominados por i, gracias a la propiedad de transitividad.
- $\sum_{r \in D^{-1}(j)} |D(r) \cap D^{-1}(i)|$ Número de arcos existentes entre los nodos dominados por el nodo i
- $\sum_{r \in D(j)} |D(i) \cap D^{-1}(r)|$ Número de arcos existentes entre los nodos que dominan al nodo i
- PageRank

Es una herramienta muy conocida por ser el motor del buscador de Google que permite ordenar las páginas web resaltando las páginas más importantes al principio. Se calcula de forma iterativa y se basa en que cada nodo de la red reparte su puntuación (PR) entre todos los nodos a los que se dirigen sus enlaces de forma proporcional al peso de dichos enlaces. Por lo que la importancia de un nodo dependerá de la importancia de los nodos que le señalen y el número de nodos a los que señalan estos últimos.

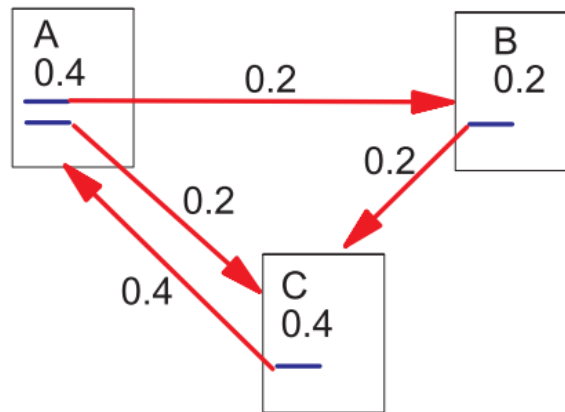


Figura 4.1 Distribución del PageRank de forma simplificada

Fuente: (Page, L., Brin, S., Motwani, R., Winograd 1999)

Este índice se rige por la siguiente expresión:

$$PR_i = \left(\alpha \sum_j PR_j \frac{W_{ji}}{s_j^{out}} + \alpha \frac{1}{n} \sum_j PR_j \delta(s_j^{out} = 0) + \frac{(1 - \alpha)}{n} \right) n \quad (4.22)$$

Los dos primeros sumandos del paréntesis realizan el reparto de los PageRank entrando en funcionamiento el último de ellos cuando el grado de salida del nodo j es 0, siendo en ese caso $R_j \delta(s_j^{out} = 0) = 1$. El coeficiente α es menor que la inversa del máximo autovalor de la matriz W aunque por acuerdo se utiliza normalmente un factor de 0.85 El tercer sumando permite que todas aquellas iteraciones que se hagan dentro de una componente fuertemente conectada sin que haya ningún arco que permita acceder a otras partes de la red, puedan salir de ese bucle reiniciando el cálculo.

Otra de verlo sería que con una probabilidad de 0.85 se realizan los cálculos a través de los nodos que están

conectados entre sí, mientras que con una probabilidad de 0.15 se elige un destino aleatoriamente.

La multiplicación del paréntesis se debe a que la media del PageRank en caso de no multiplicarlo es $1/n$. De forma que según la fórmula propuesta todos los nodos que tengan un $PR > 1$ estarán por encima de la media.

4.2.2.2 Medidas a nivel de capa

- Porcentaje de nodos de la componente c que se encuentra en la capa q :

Se define como el ratio entre los nodos que pertenecen a la capa q dentro de la componente c $|L_{cq}|$ y los nodos que pertenecen a la componente c $|D_c|$

$$v_{cq} = \frac{|L_{cq}|}{|D_c|} \quad (4.23)$$

- Grado medio de entrada de los nodos que pertenecen a la capa q de la componente c :

$$d_{cq}^{averin} = \frac{1}{|L_{cq}|} \sum_{r \in L_{cq}} d_r^{in} \quad (4.24)$$

- Grado medio de salida de los nodos que pertenecen a la capa q de la componente c :

$$d_{cq}^{averout} = \frac{1}{|L_{cq}|} \sum_{r \in L_{cq}} d_r^{out} \quad (4.25)$$

- Distancia mínima y máxima de la capa a la frontera eficiente:

Con el fin de establecer un rango en el cual se encuentra la distancia entre la capa q y la frontera eficiente, se determinan los rangos de distancia máxima y mínima. Formulándose respectivamente

$$\begin{aligned} \theta_{cq}^{max} &\in \left(\min_{r \in L_{cq}} e_r^{max}, \max_{r \in L_{cq}} e_r^{max} \right) \\ \tau_{cq}^{min} &\in \left(\min_{r \in L_{cq}} \tau_r^{max}, \max_{r \in L_{cq}} \tau_r^{min} \right) \end{aligned} \quad (4.26)$$

4.2.2.3 Medidas a nivel de componte

- Porcentaje de nodos que se encuentran en la componente c

Muestra ratio de los nodos de la red que pertenece a la componente c

$$\alpha_c = \frac{|D_c|}{|D|} \quad (4.27)$$

- Porcentaje de enlaces que se encuentra en la componente c

Determina el número de enlaces que pertenecen a la componente c , siendo NE_c el número de enlaces que se encuentran en la componente c y NE el número de enlaces existentes en la red.

$$\hat{\alpha}_c = \frac{|NE_c|}{|NE|} \quad (4.28)$$

Resulta significativo cuando la red está compuesto por una componente gigante y varias componentes compuestas únicamente por un nodo aislado, porque en ese caso el ratio será igual a la unidad en el caso de la componente gigante.

- Porcentaje de nodos eficientes de la componente c

El objetivo de este índice es calcular el ratio que muestre, de los nodos eficientes cuántos pertenecen a la componente c .

$$\pi_c = \frac{|D_c^*|}{|D^*|} \quad (4.29)$$

- Porcentaje de nodos eficientes en la componente c

Este índice busca calcular dentro de una componente, cuántos nodos son eficientes.

$$\widehat{\pi}_c = \frac{|D_c^*|}{|D_c|} \quad (4.30)$$

- La media del grado de los nodos pertenecientes a la componente c

Al analizarse todos los nodos de la componente, a la hora de calcular la media resulta indiferente usar el grado de entrada o el grado de salida.

$$d_c^{aver} = \frac{1}{|D_c|} \sum_{r \in D_c} d_r^{in} = \frac{1}{|D_c|} \sum_{r \in D_c} d_r^{out} \quad (4.31)$$

- Densidad de la componente c

Como se explicó en el apartado 3.2 la densidad se define como el número de enlaces entre el número de posibles enlaces

$$\rho_c = \begin{cases} 1 & \text{si } |D_c| = 1 \\ \frac{m}{(|D_c| - 1) |D_c|} = \frac{d_c^{aver}}{|D_c| - 1} & \text{si } |D_c| > 1 \end{cases} \quad (4.32)$$

- Diámetro de la componente c

Calcula cuál es la longitud máxima de un nodo ineficiente a la frontera eficiente. Gracias a las propiedades de transitividad y aditividad el arco ij tiene como longitud la distancia del camino geodésicos entre i y j .

$$\Delta_c = \max_{r, j \in D_c} e_{rj} \quad (4.33)$$

- Distancia media de la componente c a la frontera eficiente

$$\theta_c^{aver} = \frac{1}{|D_c|} \sum_{r \in D_c} e_r^{max} \quad (4.34)$$

- Mínima eficiencia de mejora total de la componente c

Determina cuál es la distancia mínima que debería recorrer los nodos ineficientes de la componente c en total para poder alcanzar la frontera eficiente.

$$\tau_c = \sum_{r \in D_c \setminus D_c^*} \tau_r^{min} \quad (4.35)$$

4.2.2.4 Medidas a nivel de red

En el caso de que la red esté compuesta por una componente gigante y nodos aislados algunos índices que se detallan a continuación tendrán el mismo valor que sus homólogos en los índices de las componentes.

- Porcentaje de nodos eficientes

$$\hat{\pi} = \frac{|D^*|}{|D|} \quad (4.36)$$

- Diámetro de la red

$$\Delta = \max_{r,j \in D} e_{rj} = \max_{r \in D} e_r^{max} = \max_{j \in D^*} \sigma_j \quad (4.37)$$

- Distancia media a la frontera eficiente

$$\theta^{aver} = \frac{\sum_c |D_c| \theta_c^{aver}}{|D|} = \frac{1}{n} \sum_{r \in D} e_r^{max} \quad (4.38)$$

4.2.3 Filtros

Las redes de dominancia proporcionan un marco en el que se puede visualizar gráficamente quién domina a quién y cuál es la eficiencia relativa existente entre ellos. Una herramienta muy útil sobre todo cuando los datos de las unidades a analizar tienen múltiples dimensiones, que impiden su representación gráfica en 2 o 3 dimensiones. Por otra parte, dentro del Análisis de Redes Complejas, existen múltiples técnicas que permiten generar subgrafos dentro de la red, que poseen características concretas con el fin de segmentar la información proporcionada por la red. A continuación se detallan algunos de los posibles filtros que se pueden aplicar

- Filtro de umbral superior

Genera un subgrafo en el que se eliminan todos aquellos arcos que tienen un valor mayor a un umbral (l) determinado. Siendo el subgrafo resultante (D, E') , donde $E' = \{(r, j) \in E: 0 < e_{rj} \leq l\}$.

- Filtro de umbral inferior

Este filtro al igual que el filtro de umbral superior mantiene en el subgrafo generado todos los nodos de la red, pero elimina aquellos arcos que tengan un valor inferior al marcado por un determinado umbral (l'). Siendo el subgrafo resultante (D, E') , donde $E' = \{(r, j) \in E: l' \leq e_{rj}\}$.

- Filtro de grafo bipartito

Si se toman 2 tipos de nodos dentro de la red: los nodos eficientes y los nodos ineficientes. Y únicamente se muestran los enlaces existentes entre los dos tipos de nodos, sin visualizar los arcos que se encuentran entre los nodos ineficientes, se obtiene un grafo bipartito definido como (D, E') . Siendo $E' = \{(r, j) \in E: r \in D \setminus D^* \wedge j \in D^*(r)\}$

- Filtro de los objetivos eficientes más cercanos

Este filtro es una combinación del filtro de grafo bipartito y el filtro de umbral superior aplicando un umbral dinámico. El objetivo de este filtro es visualizar únicamente aquellos enlaces que marcan la mínima distancia a la frontera eficiente. El subgrafo se define con la siguiente expresión (D, E') siendo $E' = \{(r, j) \in E: j \in D^* \wedge e_{rj} = \tau_r^{min}\} = \{(r, j) \in E: r \in D \setminus D^* \wedge j \in D^*(r)\}$

- Filtro de objetivos eficientes

Este filtro se aplica sobre cada unidad ineficiente, mostrando únicamente los nodos eficientes sobre los que se proyecta y los arcos existentes entre ellos. Si aplicamos el filtro de objetivos eficientes al nodo r , obtendríamos el siguiente subgrafo $(D^*(r), E^*(r))$

- Filtro de nodos dominados

Este se aplica a cualquier nodo eficiente de la red y visualiza aquellos nodos que domina y su relación con ellos. De forma que el subgrafo resultante al aplicar el filtro de nodos dominados sobre el nodo j es $(D^{-1}(j), E^{-1}(j))$

- Filtro de egonetwork

Al aplicar este filtro sobre un nodo p , se visualizan todos los nodos que dominan y son dominados por el nodo p , así como los arcos existentes entre todos los nodos visualizados. Definiéndose con la siguiente expresión

$(D^{ego}(p), E^{ego}(p))$ siendo $D^{ego}(p) = \{p\} \cup D(p) \cup D^{-1}(p)$ y $E^{ego}(p) = E^{-1}(p) \cup E(p)$

○ Filtro de esqueleto

Este filtro se aplica sobre toda la red eliminando los arcos transitivos, definidos como $\{(r, j): \exists p \in D(r) \cap D^{-1}(j)\}$. Gracias a la propiedad de transitividad, se puede reducir el número de enlaces sin perder información en la visualización. El subgrafo resultante tras la aplicación de este filtro es (D, E^S) siendo $E^S = \{(r, j): D(r) \cap D^{-1}(j) = \emptyset\}$

5 ILUSTRACIÓN DE LA METODOLOGÍA

“El genio se compone del dos por ciento de talento y del noventa y ocho por ciento de perseverante aplicación.”

- Ludwig van Beethoven -

En este apartado se va a aplicar la metodología desarrollada a una serie de juegos de datos de la literatura, con el fin de mostrar las ventajas que se adquieren a la hora de visualizar y analizar los resultados obtenidos tras el análisis de la eficiencia en forma de red.

A la hora de realizar el análisis de eficiencia en los juegos de datos, se va a utilizar la tecnología FDH, por lo que la frontera eficiente estará compuesta por unidades existentes que se proyectan sobre sí mismas y serán targets para las unidades ineficientes. La eficiencia relativa entre estas unidades se mide con la siguiente métrica aditiva, por lo que el modelo DEA no tiene orientación de entrada, ni de salida.

$$e_{rj} = \begin{cases} 0 & \text{si } j \notin D(r) \\ \sum_i \frac{x_{ir} - x_{ij}}{x_i^{aver}} + \sum_k \frac{y_{kj} - k_{kr}}{y_r^{aver}} & \text{si } j \in D(r) \end{cases} \quad (4.4)$$

La red se construye con la metodología descrita en el apartado 4.2.1, por ello y gracias a la métrica empleada los enlaces tienen las propiedades de aditividad y de transitividad.

5.1 Juego de datos de CST

El primer juego de datos que se analiza en este capítulo se encuentra en el libro (Cooper, W. W., Seiford, L. M., Zhu 2004), consta de 8 unidades productivas, un factor de entrada y un factor de salida. No tiene dimensiones al ser no representar las DMUs unos procesos reales.

Juego de datos de LIM Tabla 5–2. Ejemplo de Cooper, Seiford y Tone (CST)

	A	B	C	D	E	F	G	H
x1	2	3	3	4	5	5	6	8
y1	1	3	2	3	4	2	3	5

Gracias a los índices a nivel de red y de componente se puede establecer que en este caso, sin necesidad de representar gráficamente las relaciones, que la red está compuesta por 3 componentes. Dos componentes formadas por nodos aislados y la tercera es una componente gigante, cuyas métricas se representan en la Tabla 5.11.

Todos los enlaces de esta red se encuentran dentro de la componente gigante, que posee la mitad de las DMUs eficientes que componen la frontera eficiente y el 75% de los nodos de la red. Estos nodos eficientes representan el 33.33% de los nodos que forman la componente gigante. La componente tiene una densidad muy baja (30%) y está compuesta por 9 arcos, dos de los cuales son transitivos.

Tabla 5.1. Índices a nivel de red y de componente en la red CST

Δ	θ^{aver}	$\hat{\pi}$	$\hat{\alpha}_1$	NC	CC_1^{aver}	τ_1
0.79	0.25	0.50	1	3	0.47	1.84
Δ_1	θ_1^{aver}	$\hat{\pi}_1$	α_1	ρ_1	π_1	d_1^{aver}
0.79	0.34	0.33	0.75	0.30	0.50	1.50

La distancia del nodo más ineficiente a la frontera eficiente es de 0.79 y la distancia mínima que deberían recorrer en total todos los nodos ineficientes de la componente gigante para pertenecer a la frontera eficiente es de 1.84. No obstante, la distancia media de los nodos a la frontera eficiente es más baja que la media, por lo que hay nodos muy próximos a la frontera eficiente. El grado medio dentro de la componente gigantes es de 1.50. El coeficiente clustering es muy bajo, debido a la poca transitividad de la red y al número reducido de capas, ya que no pueden existir arcos entre nodos que pertenezcan a la misma capa.

A continuación se representa la red completa y el subgrafo esqueleto, donde no se encuentran los arcos transitivos. Hay que tener en cuenta que en la Figura 5.1 existe un arco entre G y B que no se visualiza porque están superpuestos los arcos GD y BD. Este arco no existe en la Figura 5.2

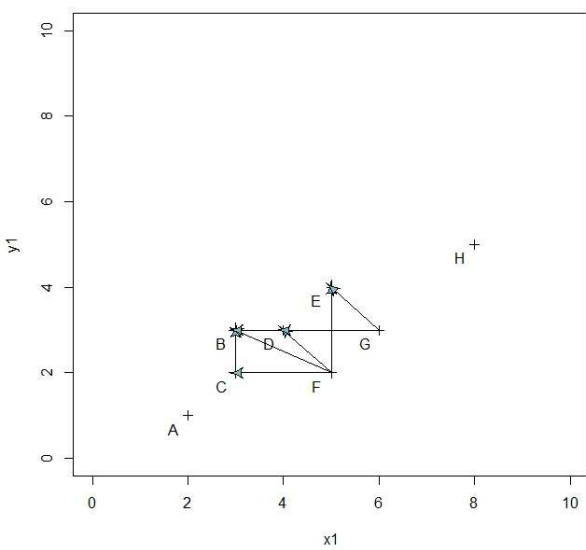


Figura 5.1 Visualización de la red de dominancia del juego de datos de CST

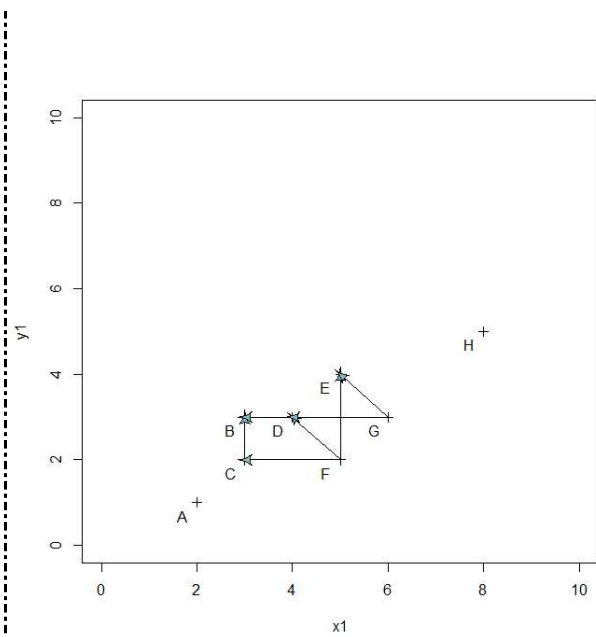


Figura 5.2 Subgrafo de esqueleto en la red CST

Fuente: Propia

Como se puede apreciar que la frontera eficiente está compuesta por 4 DMUs, esta apreciación se encontraba implícita dentro de los índices anteriormente mencionados, porque todos los nodos aislados pertenecen a la frontera eficiente.

Dentro de la Figura 5.3 se aprecia el diámetro de la red de 0.79 y cómo la mitad de los arcos son menores o iguales a 0.35. El hecho de que haya pocos arcos transitivos hace que la distribución de los enlaces del esqueleto

y de la red completa, que se observan en la Figura 5.4, estén muy próximas. El peso máximo de un arco no transitivo es de 0.70 y casi el 60% de estos enlaces son menores o iguales a 0.45.

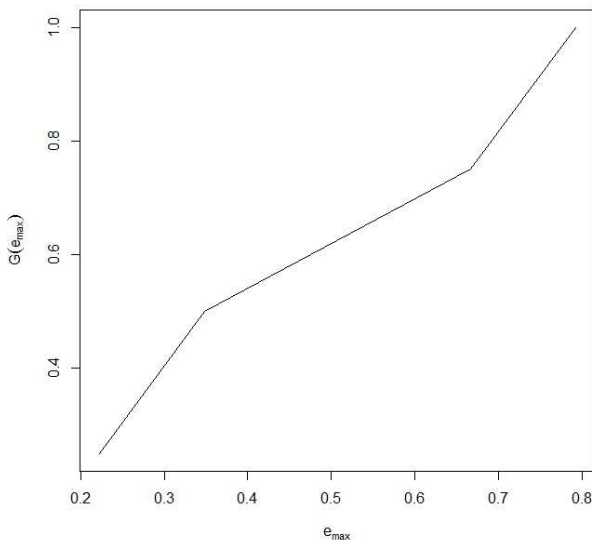


Figura 5.3 Distribución de las distancias máximas a la frontera eficiente del caso CST

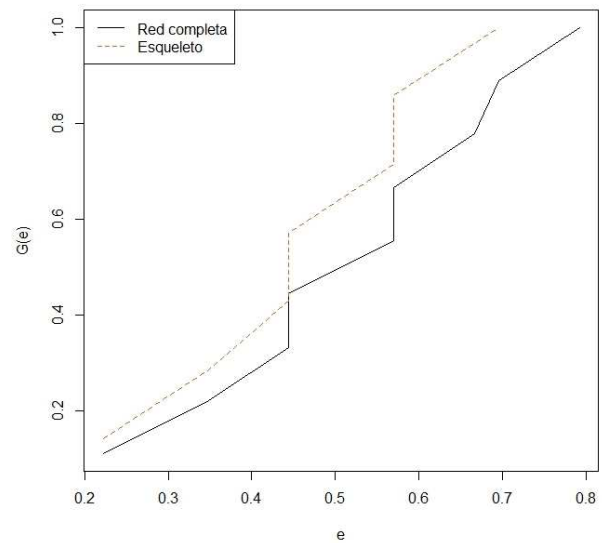


Figura 5.4 Distribución de los enlaces del caso CST

Fuente: Propia

La componente gigante está compuesta por 3 capas, que están formadas cada una por 2 nodos. En este caso el grado medio de entrada es constante en la capa 1 y en la capa 0, pero el grado va a ser siempre mayor cuanto más cerca esté el nodo de la frontera eficiente.

Tabla 5.2. Distribución de los nodos según su capa en el caso CST

Capa	Nº nodos en la capa	% de nodos de la componente c en la capa	θ_{cq}^{max}	τ_{cq}^{min}	$d_{cq}^{aver\ in}$	$d_{cq}^{aver\ out}$
0	2	33.33	[0,0]	[0,0]	1.50	0.00
1	2	33.33	[0.22,0.35]	[0.22,0.35]	1.50	1.00
2	2	33.33	[0.67,0.79]	[0.57,0.70]	0.00	3.50

El grado medio de salida de la última capa es el más elevado de todas, al ser donde se encuentran los nodos ineficientes con más targets intermedios. En esta tabla se puede apreciar como la capa 1 se encuentra muy próxima a la frontera eficiente.

En la Figura 5.5 se aprecia como el grado de entrada es mayor cerca de la frontera eficiente y como a medida que los nodos pertenecen a capas más alejadas de la frontera eficiente va disminuyendo, mientras aumenta el grado de salida debido a que tienen más targets intermedios.

Dentro de la especificidad de un nodo coincide con el grado total del nodo y representa el número de arcos que llegan o salen de la red, mientras que el índice hub no solo muestra la proporción de enlaces, sino también cuáles son los nodos extremos. En este caso los únicos nodos que no son extremos son los nodos C y D porque tienen un índice hub distinto de cero. Por lo que son los únicos nodos que aparecen en el camino intermedio entre los nodos extremos, por ello son los únicos con un coeficiente de intermediación distinto de cero

Tabla 5.3 se observan las características de los nodos, a que componente y capa pertenece cada una así como todos los índices descritos en la metodología. Los nodos pertenecientes a la capa 1, están dominados por un

único nodo, luego la necesidad del benchmark de los nodos eficientes a los que se refiere aumentará. Esto también se aprecia dentro de los nodos eficientes y los nodos aislados, porque todos los nodos que pertenecen a la frontera eficiente se proyectan sobre sí mismos. La distancia máxima y mínima de los nodos ineficientes coincidirán si sólo son dominados por un único nodo, al solo haber una medida a la frontera eficiente.

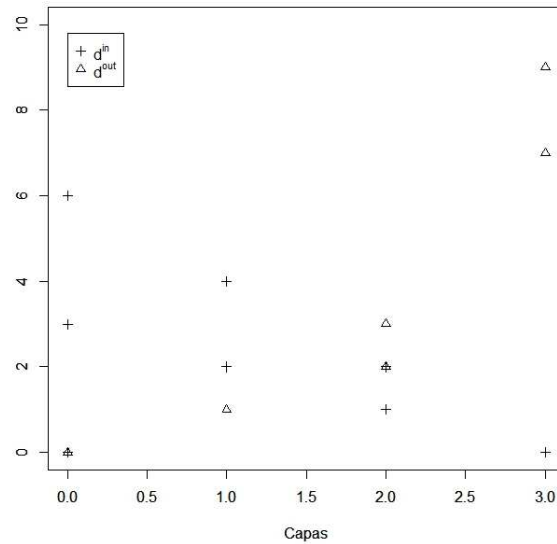


Figura 5.5 Visualización del grado de entrada y de salida en función de las capas en el caso CST

Fuente: Propia

La especificidad de un nodo coincide con el grado total del nodo y representa el número de arcos que llegan o salen de la red, mientras que el índice hub no solo muestra la proporción de enlaces, sino también cuáles son los nodos extremos. En este caso los únicos nodos que no son extremos son los nodos C y D porque tienen un índice hub distinto de cero. Por lo que son los únicos nodos que aparecen en el camino intermedio entre los nodos extremos, por ello son los únicos con un coeficiente de intermediación distinto de cero

Tabla 5.3. Índices a nivel de nodo en el caso CST

DMU_r	Componente	Capa	$ D^*(r) $	e_r^{max}	τ_r^{min}	d_r^{in}	d_r^{out}	η_r	γ_r	β_r	CC_r	PR_r
B		0	1	0.00	0.00	4	0	4	0	0.00	0.50	2.55
E		1	1	0.00	0.00	2	0	2	0	0.00	0.00	1.03
C	1	1	1	0.35	0.35	1	1	2	1	0.33	1.00	0.77
D			1	0.22	0.22	2	1	3	2	0.33	0.67	0.95
F		2	2	0.79	0.70	0	4	4	0	0.00	0.33	0.67
G			2	0.67	0.57	0	3	3	0	0.00	0.33	0.673
A	2	0	1	0.00	0.00	0	0	0	0	0.00	0.00	0.67
H	3	0	1	0.00	0.00	0	0	0	0	0.00	0.00	0.673

Resulta sorprendente encontrar que la DMU C tiene un coeficiente de clustering 1, esto se debe a que domina a un nodo y sólo está dominada por nodo, por lo que el único arco existente entre sus vecinos es un arco transitivo. El nodo D por su parte, se encuentra en la misma capa que el nodo C y su coeficiente clustering es menor debido a que domina a dos nodos de la capa 2 que nunca podrán estar dominados por su pertenencia a la misma capa.

El coeficiente PageRank permite ordenar los nodos de la red según su importancia, el nodo que más enlaces tiene y cuyos vecinos son más importantes es el nodo B, seguido por el nodo E con un PR muy cercano al nodo D esto se debe a que ambos dominan a los mismos nodos. Sin embargo, el nodo E pertenece a la frontera eficiente.

Tabla 5.4. Índices a nivel de nodos eficientes en el caso CST

DMU_r	$ D^{-1}(j) $	$ D^{-1}(j) $	κ_j	σ_j	φ_j
A	0	0	0.00	0.00	0
B	4	2	2.03	0.79	2
E	2	0	1.27	0.70	0
H	0	0	0.00	0.00	0

Si se analizan los nodos eficientes, se observa como B es la DMU más importante al ser un target para la mayor parte de los nodos de la red. Además, es el único nodo eficiente para dos nodos de la red, mientras el resto de la frontera no los tiene. El radio de ineficiencia es similar para los nodos eficientes de la red. El potencial del benchmarking es superior en el nodo B debido a que domina a más nodos y es el único que domina a nodos que se encuentran en la primera capa.

5.2 Juego de datos de Lim

A continuación se usa el juego de datos del supermercado utilizado en (Lim et al. 2011) compuesto por 12 DMUs con 2 entradas y 1 salida. Las dimensiones de las unidades de las entradas son: 10 empleados y 1000 m² de superficie, mientras que las dimensiones de las salidas son 100.000 dólares. A continuación se muestran los datos del problema.

Tabla 5.5. Datos del caso Lim

Tiendas		A	B	C	D	E	F	G	H	I	J	K	L
Empleados	x1	2	6	9	3	4	8	5	7	8	7	7	8
Superficie	x2	7	2	1	7	5	2	5	3	3	9	4	7
Ventas	Y	1	1	1	1	1	1	1	1	1	1	1	1

Todas las unidades, al tener salidas constantes, permite que se pueda representar el conjunto de datos en un gráfico de dos dimensiones y visualizar donde se encuentra la frontera eficiente, así como las relaciones entre las diferentes DMUs. Una vez generada la red de dominancia, se muestra gráficamente en la siguiente figura.

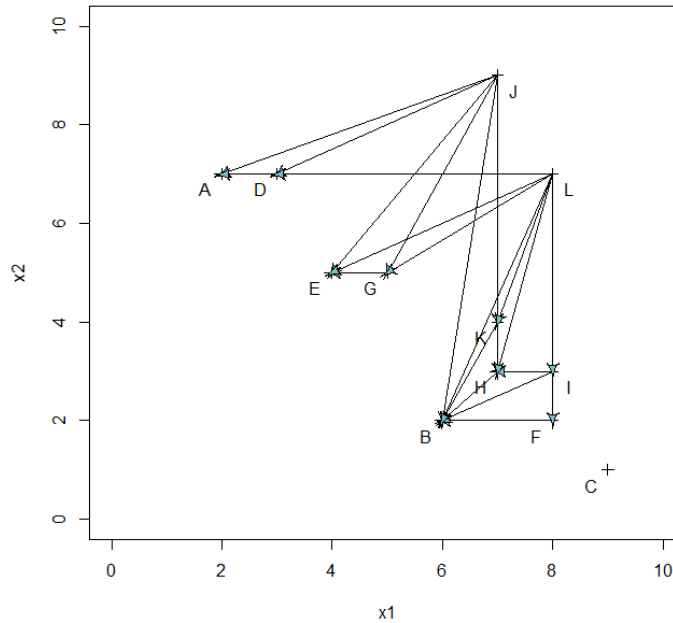


Figura 5.6 Visualización de la red de dominancia en el caso Lim

Fuente: Propia

La Figura 5.6 muestra que dentro de la red existen dos componentes, una componente gigante compuesta por 11 DMUs y una componente formada por un nodo aislado. La frontera eficiente está formado por todas aquellas DMUs que se proyectan sobre sí mismas, estas proyecciones no se representan porque no se ha dispuesto en la definición de la red, por lo que tienen grado de salida nulo. Cabe destacar que las componentes con nodos aislados muestran que la unidad no tiene ninguna relación con ningún otro nodo de la red, pero eso no significa que no se proyecten sobre sí mismas. Como consecuencia, todas las componentes formadas por nodos aislados forman parte de la frontera eficiente.

Tabla 5.6. Índices a nivel de red y de componente en el caso Lim

Δ	θ^{aver}	$\hat{\pi}$	$\hat{\alpha}_1$	NC	CC_1^{aver}	τ_1
1.69	0.44	0.33	1	2	0.65	4.39
Δ_1	θ_1^{aver}	$\hat{\pi}_1$	α_1	ρ_1	π_1	d_1^{aver}
1.69	0.48	0.273	0.92	0.23	0.75	2.27

A nivel de red, se observa que el 33.33% de los nodos son eficiente, siendo la distancia máxima entre dos unidades 1.69. La distancia de este camino geodésico se aprecia entre la unidad J y la unidad B. Sin embargo, la distancia media a la frontera eficiente es de 0.44. Si apreciamos la distribución de las distancias a la frontera eficiente en la Figura 5.7 se observa que el 75% de las distancias tiene un valor igual o inferior a 0.60

La red y la componente gigante están compuestas por el mismo número de arcos, la distribución de la longitud de sus arcos es la misma. Por consiguiente se puede centrar el foco en la componente gigante sin perder información de las relaciones de dominancias que no se encuentren en él. Por lo que los próximos índices que se mencionen se referirán a la componente gigante.

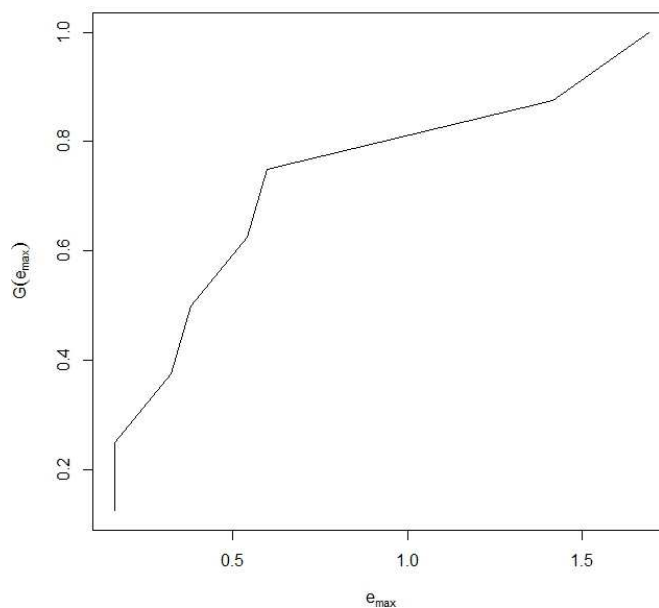


Figura 5.7 Distribución de las distancias máximas a la frontera eficiente en el caso Lim

Fuente: Propia

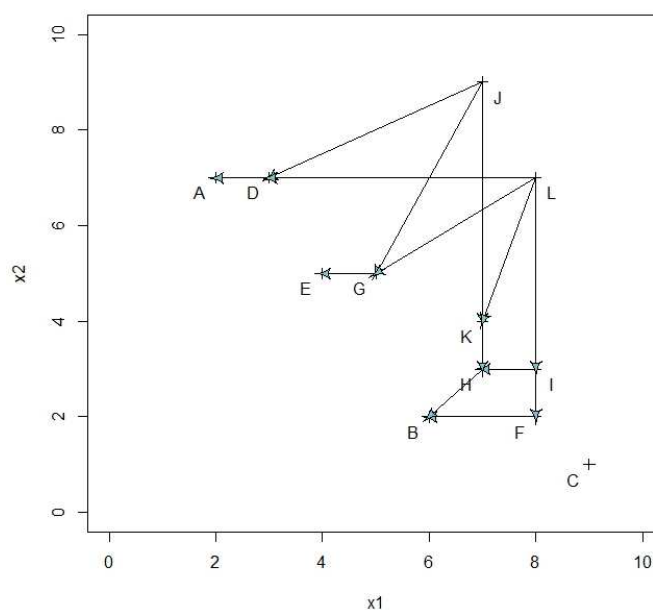


Figura 5.8 Subgrafo de esqueleto en el caso Lim

El número de enlaces de la componente gigante es de 25 arcos, situando el grado medio de los arcos en 2.273 y la densidad de la componente en un 22.7%. En caso de eliminar los arcos transitivos, la subred está constituida por 14 arcos. El porcentaje de nodos eficientes es de 27.3%, constituyendo el 75% de los nodos eficientes de la red.

El diámetro de la componente es el mismo que en la red general y la distancia media a la frontera eficiente es muy similar, tiene un valor de 0.48. No obstante, La distancia que deberían recorrer todos los nodos para convertirse en nodos eficientes es de 4.39.

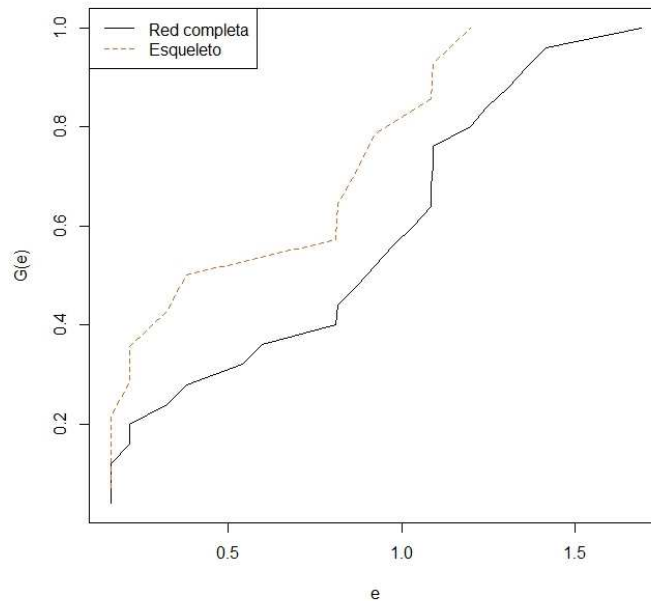


Figura 5.9 Distribución de los enlaces en el caso Lim

Fuente: Propia

La comparación entre la distribución de la longitud de los arcos en la red esqueleto y en toda la red, nos muestra el efecto de los arcos transitivos. La distribución de la red esqueleto, al no mostrar los arcos transitivos, muestran la longitud de los caminos compuestos por un único enlace. Y como es lógico, la curva se desplaza hacia la izquierda al tener una mayor proporción de arcos de menor longitud.

Si se observan los nodos en función de la capa que tienen asignada, se distribuyen tal y como se indica en la siguiente tabla:

Tabla 5.7. Distribución de los nodos según su capa en el caso Lim

Capa	Nº nodos en la capa	% de nodos de la componente c en la capa	θ_{cq}^{max}	τ_{cq}^{min}	d_{cq}^{averin}	$d_{cq}^{averout}$
0	3	27.27	[0,0]	[0,0]	3	0
1	4	36.36	[0.16,0.38]	[0.16,0.38]	2.5	1
2	2	18.18	[0.54,0.60]	[0.54,0.60]	1.5	2.5
3	2	18.18	[1.42,1.69]	[1.250.97]	0	8

El 36.36% de los nodos de la componente se encuentran en la capa 1 a una distancia entre 1 y 2 unidades. La capa 2 está compuesta por casi el 20% de los nodos de la componente y se encuentra a una distancia de 3 unidades de la componente. A medida que las DMUs se alejan de la frontera eficiente se visualiza más distancia entre las distancias máximas y mínimas a dicha frontera, debido a que la zona donde se proyectan es mucho más amplia, como se puede observar en la capa 3. Resulta lógico ver cómo el grado medio de entrada disminuye, mientras el grado medio de salida aumenta a medida que se analizan capas más alejadas de la frontera eficiente. En caso de visualizar el grado medio de entrada y de salida en función de las capas, como se muestra en la Figura 5.10, se observa el mismo patrón.

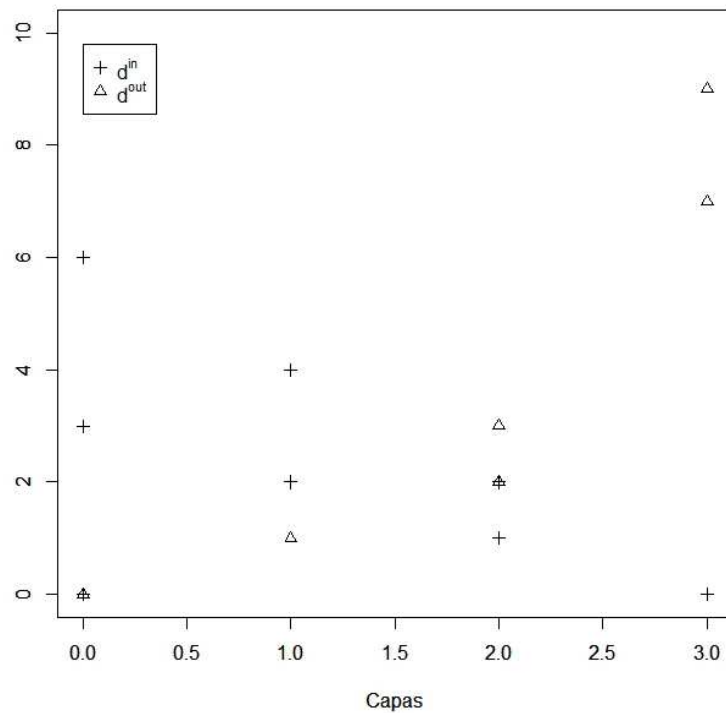


Figura 5.10 Visualización del grado de entrada y de salida en función de las capas en el caso Lim

Fuente: Propia

A continuación se muestran la tabla donde se puede apreciar los valores de los índices expuestos en el apartado 4.2.2.1

Tabla 5.8. Índices a nivel de nodo en el caso Lim

DMU_r	Compo- nente	Capa	$ D^*(r) $	e_r^{max}	τ_r^{min}	d_r^{in}	d_r^{out}	η_r	γ_r	β_r	CC_r	PR_r
A			1	0.00	0.00	3	0	3	0	0.00	0.67	1.32
B		0	1	0.00	0.00	6	0	6	0	0.00	0.60	3.03
E			1	0.00	0.00	3	0	3	0	0.00	0.67	1.34
D			1	0.16	0.16	2	1	3	2	0.50	0.67	0.70
F			1	0.32	0.32	2	1	3	2	0.25	1.00	0.79
G	1	1	1	0.16	0.16	2	1	3	2	0.00	0.67	0.71
H			1	0.38	0.38	4	1	5	4	2.25	0.70	0.98
I			1	0.54	0.54	1	3	4	3	0.58	0.83	0.64
K		2	1	0.60	0.60	2	2	4	4	0.58	0.83	0.70
J			3	1.69	1.25	0	7	7	0	0.00	0.24	0.59
L		3	3	1.42	0.97	0	9	9	0	0.00	0.25	0.59
C	2	0	1	0.00	0.00	0	0	0	0	0.00	0.00	0.59

En esta tabla se muestra igualmente como el grado de entrada aumenta, mientras el grado de salida disminuye en las capas más alejadas de la frontera eficiente, así como los intervalos entre los que se encuentra la distancia de cada capa a la frontera eficiente.

La especificidad del nodo deja patente, el número de relaciones de dominancia que tiene ese nodo, mientras que el índice hub muestra no solo cómo cuantifica esas relaciones sino también indica cuáles son los nodos extremos. Analizando simultáneamente con el PageRank, resulta interesante analizar el caso de los nodos H, I y K. Los nodos I y K tienen la misma especificidad, ambos tienen el mismo clustering pero el nodo K tiene mejor intermediación y PageRank. Esto se debe a que tiene un enlace más que lo conecta con un el nodo B. Por otra parte, el nodo H y el nodo K tienen el mismo índice hub, pero el nodo H tiene mayor importancia dentro de la red no sólo por estar a un menor número de pasos de la frontera eficiente, sino porque su PageRank es superior al dominar más nodos que K.

Uno de los sumandos del coeficiente clustering muestra la existencia de caminos alternativos entre los vecinos de un determinado nodo que no pasen por ese determinado nodo. Sin embargo no se puede establecer una relación directa entre el coeficiente de intermediación y el coeficiente clustering porque en el coeficiente clustering es un ratio, mientras que el coeficiente de intermediación es una suma de ratios.

El coeficiente de intermediación es un buen referente dentro de los nodos ineficientes, donde el PageRank aunque marca la importancia entre ellos esta diferencia es muy sutil. De esta forma queda patente que el nodo H es el nodo que más veces queda en el camino más corto entre cada par de nodos.

El PageRank permite ordenar todos los nodos de la red, pero lo que es más importante los nodos eficientes, dejando patente que el nodo más importante es el nodo B, seguido del nodo A y D. Teniendo en cuenta no solo su grado de entrada sino también la importancia de los nodos que domina.

Si se utilizan los índices a nivel de nodo de las unidades eficientes, se observa que el nodo B es el único benchmark para 4 nodos de la componente, mientras para los nodos A y D solo se da esa situación para un nodo. Los nodos que dominan A y D se encuentran en ambas situaciones de forma global a la misma distancia de la frontera eficiente, mientras que para el nodo B es casi el doble debido a que domina al doble de nodos. El radio de ineficiencia ronda entre los valores 1.25 y 1.69 para los tres nodos eficientes de la componente, debido a que todos dominan a los nodos de la capa más alejada de la frontera eficiente. A continuación se muestran los valores de los índices que se acaban de describir.

Tabla 5.9. Índices a nivel de nodos eficientes en el caso Lim

DMU_r	$ D^{-1}(j) $	$ D^{-^{-1}}(j) $	κ_j	σ_j	φ_j
A	3	1	2.38	1.25	1
B	6	4	4.95	1.69	2
C	0	0	0.00	0.00	0
E	3	1	2.61	1.36	1

5.3 Juego de datos de Park

El juego de datos ha sido extraído de (Park et al. 2012) donde representan un conjunto de 12 supermercados. Se consideran 2 factores de entrada y uno de salida constante, que nos permitirá visualizar las DMUs en un plano horizontal en el espacio dimensional.

A continuación se muestra los datos que componen el problema y en la Figura 5.6 se muestra cuál es la red generada.

Tabla 5.10. Datos del caso Park

	A	B	C	D	E	F	G	H	I	J	K	L
x1	2	4	8	3	4	5	5	6	7	6	6	7
x2	4	2	1	6	3	2	6	3	3	9	4	7
Y	1	1	1	1	1	1	1	1	1	1	1	1

Esta red está compuesta por dos componentes, una componente compuesta por un nodo aislado y una componente gigante que tiene todos los arcos de la red. La frontera eficiente está compuesta por 3 DMUs (A,B,C), la DMU C es una unidad productiva especializada para su tamaño.

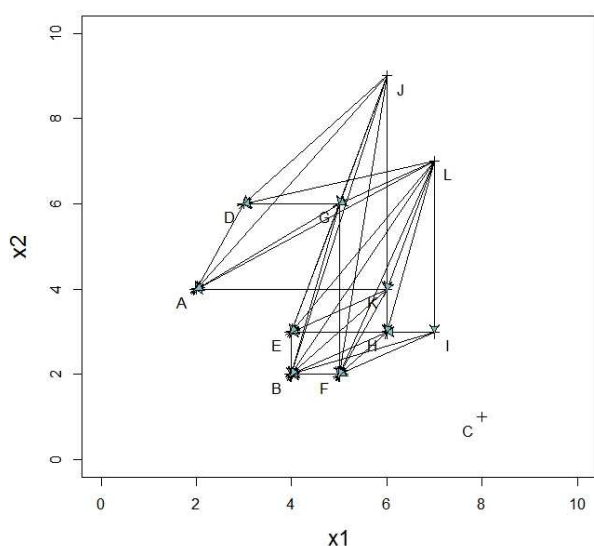


Figura 5.11 Visualización de la red de dominancia del juego de datos de Park

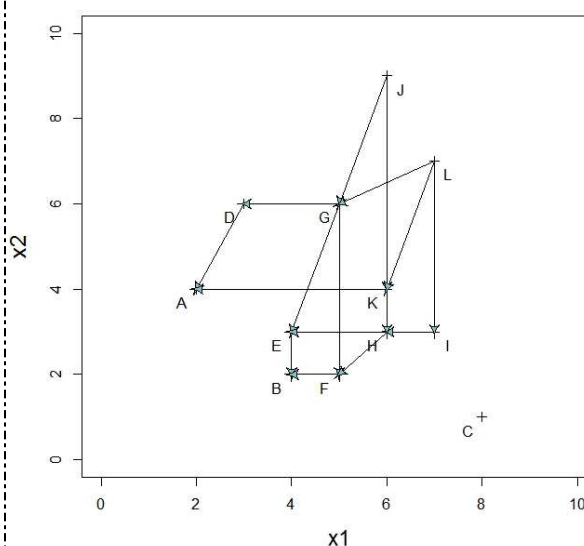


Figura 5.12 Subgrafo de esqueleto en el caso Park

Fuente: Propia

En la Figura 5.6 parece que la red es muy densa; sin embargo su densidad es del 34%, resulta mucho más útil la Figura 5.12 donde los arcos transitivos están implícitos. Al tener la componente gigante pocos targets (sólo el 18% de los nodos son eficientes) y muchos targets intermedios el coeficiente clustering medio es bastante elevado con un valor de 0.72. Como consecuencia, existe un gran número de arcos transitivos, concretamente el 56.76% de los arcos son transitivos, siendo 37 el número de arcos que componen la componente gigante. Por eso se observa un gran desplazamiento hacia la izquierda en Figura 5.7 por parte de los arcos que forman en el esqueleto. El peso máximo de los arcos entre dos nodos sin que exista un target intermedio entre ellos es de 1.20 y un 60% de dichos arcos tiene un valor igual o menor a 1.00.

Tabla 5.11. Índices a nivel de red y de componente en el caso Park

Δ	θ^{aver}	$\hat{\pi}$	$\hat{\alpha}_1$	NC	CC_1^{aver}	τ_1
2.06	0.70	0.25	1.00	2	0.72	7.98

Δ_1	θ_1^{aver}	$\hat{\pi}_1$	α_1	ρ_1	π_1	d_1^{aver}
2.06	0.76	0.18	0.92	0.34	0.67	3.36

Si se observa la Figura 5.14 poco más del 80% de las distancias máximas a la frontera eficiente es menor o igual que 1.5 Siendo la mínima distancia que deberían recorrer las DMUs ineficientes en la componente 1 para pertenecer a la frontera eficiente 7.98. La distancia media de las DMUs a la frontera eficiente tiene un valor de 0.70

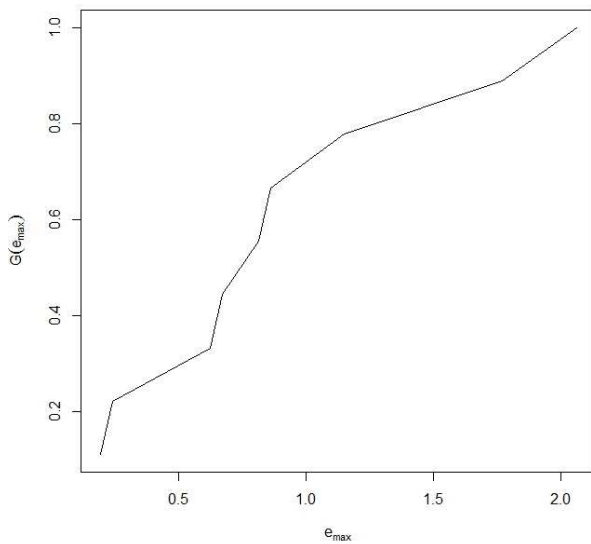


Figura 5.13 Distribución de las distancias máximas a la frontera eficiente en el caso Park

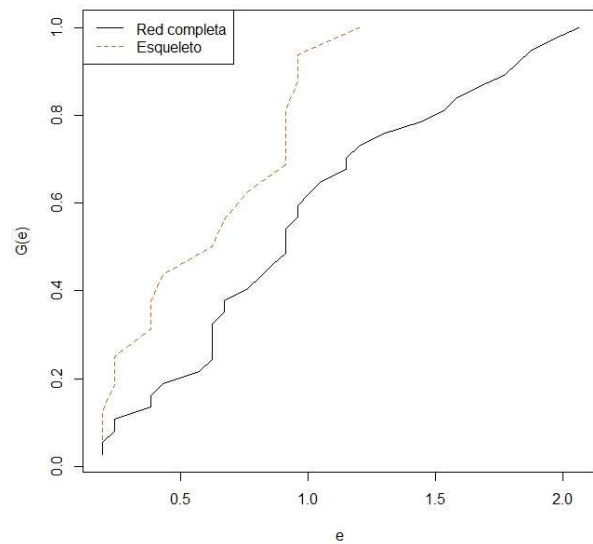


Figura 5.14 Distribución de los enlaces en el caso Park

Fuente: Propia

El grado medio dentro de la componente es 3.36, pero si atendemos al grado de entrada y de salida de cada capa vemos como el grado de entrada es más elevado en la capa 0 y disminuye a medida que la capa se aleja de la frontera eficiente. Todas las capas están compuestas por 2 nodos, salvo la capa 1 que está compuesta por 3. Por lo que los intervalos entre los que se encuentran las distancias máximas y mínimas a la frontera eficiente desde dichas capas, no son en verdad un intervalo en este caso, salvo en la capa1.

Tabla 5.12. Distribución de los nodos según su capa en el caso Park

Capa	Nº nodos en la capa	% de nodos de la componente c en la capa	θ_{cq}^{max}	τ_{cq}^{min}	d_{cq}^{averin}	$d_{cq}^{averout}$
0	2	18.18	[0,0]	[0,0]	4.33	0
1	3	27.27	[0.19,0.67]	[0.19,0.67]	5	1
2	2	18.18	[0.62,1.15]	[0.62,1.05]	3	4
3	2	18.18	[0.81,0.86]	[0.76,0.81]	1.5	4.5
4	2	18.18	[1.77,2.06]	[1.67,1.96]	0	8.5

En la Figura 5.15 se puede observar una correlación negativa entre el grado de entrada y las capas y una correlación positiva entre el grado de salida y las capas. En caso de querer calcularse esta correlación se debería no

tener en cuenta las componentes formadas por nodos aislados.

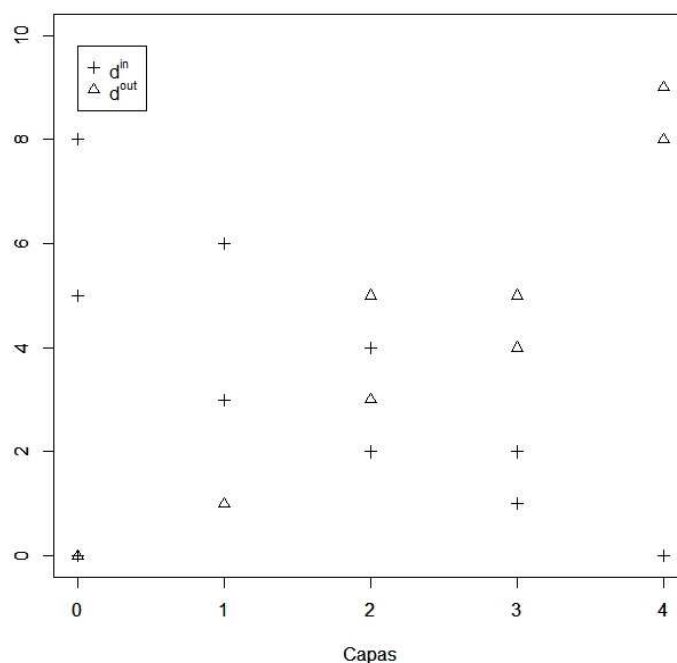


Figura 5.15 Visualización del grado de entrada y de salida en función de las capas en el caso Park

Fuente: Propia

Si observamos la Tabla 5.8 se aprecia que la distancia máxima y la distancia mínima de cada nodo a la frontera eficiente coinciden, salvo en las DMUs G,J,K,L. Esto se debe a que el resto de los nodos sólo tienen un nodo de referencia como muestra $|D^*(r)|$.

Dentro de la componente gigante se aprecian 4 nodos en los extremos, el 66.67% de los nodos tiene una especificidad en el rango entre 7 y 8 y se aprecia mucha más variación dentro del hub, donde el máximo valor lo alcanza el valor H, que alcanza también el mayor coeficiente de intermediación. El coeficiente clustering es muy elevado, debido a la transitividad mencionada previamente, siendo mayor en las capas intermedias 2 y 3.

El nodo B es el más importante de todos los nodos eficientes, debido a su elevado grado de entrada y a la importancia de sus vecinos. Mientras que el nodo B tiene a los dos nodos más importantes de la capa 1 (Los nodos E y F), el nodo A sólo domina al nodo D en esta capa. De hecho el resto de los nodos que domina el nodo A, son 4 de los 6 nodos menos importantes de toda la red.

Tabla 5.13. Índices a nivel de nodo en el caso Park

DMU_r	Componente	Capa	$ D^*(r) $	e_r^{max}	τ_r^{min}	d_r^{in}	d_r^{out}	η_r	γ_r	β_r	CC_r	PR_r
A		0	1	0.00	0.00	5	0	5	0	0.00	0.70	1.48
B			1	0.00	0.00	8	0	8	0	0.00	0.75	3.40
D	1		1	0.67	0.67	3	1	4	3	2.25	0.83	0.66
E		1	1	0.24	0.24	6	1	7	6	1.48	0.71	1.16
F			1	0.19	0.19	6	1	7	6	2.48	0.71	1.20

DMU_r	Componente	Capa	$ D^*(r) $	e_r^{max}	τ_r^{min}	d_r^{in}	d_r^{out}	η_r	γ_r	β_r	CC_r	PR_r
G		2	2	1.15	1.05	2	5	7	10	3.24	0.62	0.59
H		2	1	0.62	0.62	4	3	7	12	4.34	0.81	0.71
I	1	3	1	0.81	0.81	1	4	5	4	1.29	0.90	0.57
K		3	2	0.86	0.76	2	5	7	10	2.09	0.71	0.61
J		4	2	2.06	1.96	0	8	8	0	0.00	0.57	0.53
L		4	2	1.77	1.67	0	9	9	0	0.00	0.56	0.53
C	2	0	1	0.00	0.00	0	0	0	0	0.00	0.00	0.53

Si se atiende a los índices referentes a los nodos eficientes, se observa como ya se ha comentado previamente, que el nodo B tiene más nodos dentro de la capa 1 que el nodo A. La distancia máxima a los nodos que dominan es similar tanto en ambos nodos así como el potencial benchmarking aunque el nodo b domine a 3 nodos más que el nodo A. Por otra parte el nodo B cuenta con una necesidad de Benchmarking mucho mayor que el nodo A al ser el único nodo eficiente para 4 nodos de la red frente a 1 en el caso del nodo A.

Tabla 5.14. Índices a nivel de nodos eficientes en el caso Park

DMU_r	$ D^{-1}(j) $	$ D^{--1}(j) $	κ_j	σ_j	φ_j
A	5	1	6.12	1.96	1
B	8	4	7.71	2.06	2
C	0	0	0.00	0.00	0

6 CONCLUSIONES

“Los momentos finales de una experiencia determinan el recuerdo que conservaremos de la misma.”

- Daniel Kahneman -

Las redes de dominancia permiten tener una mayor comprensión de los resultados obtenidos con el Análisis de Envoltura de Datos. Permite visualizar gráficamente los caminos de las DMUs dominadas hacia la frontera eficiente a través de la red, siendo de gran utilidad en caso de que sea imposible graficar las DMUs debido a la multidimensionalidad de los datos.

En este trabajo, tras realizar un pequeño estado del arte, se ha desarrollado una metodología que permite la creación de una red de dominancia en el caso de que el análisis de DEA emplee una métrica aditiva en una tecnología FDH. Por otra parte se han definido diferentes índices que permiten analizar la posición de cada DMU dentro de la red, su posición respecto de la frontera eficiente, a través sus caminos más cercanos y la estratificación de las unidades dominadas. Estos índices permiten caracterizar la red a diferentes niveles: a nivel de red, de componente, de capa y de nodo. Una vez creada y caracterizada la red se puede visualizar y aplicar una serie de filtros en caso de que resulte complicada la visualización y se quiera únicamente observar los nodos y las relaciones que atiendan a un criterio determinado.

Para explicar mejor el alcance de esta caracterización se ha aplicado la metodología a tres conjuntos de datos. No obstante la aplicación de los filtros definidos no se ha aplicado a los conjuntos de datos expuestos en este trabajo, debido al pequeño tamaño de la red. Sin embargo, resultan muy útiles en el caso de redes grandes.

Dentro de los índices que se han aplicado cabe destacar el PageRank que permite ordenar las DMUs según su importancia que es atribuida en función del número y la importancia de sus vecinos. El coeficiente de intermediación que determina la importancia de un nodo según el número de veces que se encuentre en los caminos más cortos entre cada par de nodos, debido a que su información resultará más eficiente para los nodos que domina cuanto mayor sea el coeficiente.

En el futuro se podría realizar un análisis que permitiera realizar la red de dominancia para modelos de DEA que consideraran una tecnología VRS y/o CRS. Así como el desarrollo de nuevas métricas que permitan caracterizar otros aspectos de la red.

REFERENCIAS

- Ali, A.I. & Seiford, L.M., 1990. Translation invariance in data envelopment analysis. *Operation Research Letters*, 9(6), pp.403–405.
- Banker, R.D., Charnes, A. Cooper, W.W., 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9), pp.1078–1092.
- Barabási, A.-L. & Albert, R., 1999. Emergence of Scaling in Random Networks. *Science*, 286, pp.509–512.
- Bardhan, I. et al., 1996. Models for Evaluating and Measuring Efficiency and Dominance in DEA. *Journal of the Operation Research Society of Japan*, 39(3), pp.322–332.
- Charnes, A. et al., 1985. Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30(1–2), pp.91–107
- Charnes, A., Cooper, W.W. & Rhodes, E., 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), pp.429–444.
- Cooper, W. W., Park, K. S. Pastor, J.T., 1999. RAM : A Range Adjusted Measure of Inefficiency for Use with Additive Models , and Relations to Other Models and Measures in DEA. *Journal of Productivity Analysis*, 11, pp.5–42.
- Cooper, W. W., Seiford, L. M., Zhu, J., 2004. Data envelopment analysis. In *Handbook on data envelopment analysis*,
- Costa, L.D.F. et al., 2007. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1), pp.167–242.
- Erdős, P. & Rényi, A., 1959. On random graphs I. *Publicationes Mathematicae (Debrecen)*, 6, pp.290–297.
- Fagiolo, G., 2007. Clustering in complex directed networks. *Physical Review E*, 76(2), pp.1–8.
- Farrell, M.J., 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), pp.253–290.
- Fernández, S., 2015. Representación del conocimiento sobre el Análisis por Envoltura de Datos (DEA) usando mapas de conceptos. Universidad de Sevilla.
- Gol'dshtein, V., Koganov, G.A. & Surdutovich, G.I., 2004. Vulnerability and Hierarchy of Complex Networks. *Physics*, 16(1), pp.1–4.
- Ho, M.H.C. et al., 2014. A new perspective to explore the technology transfer efficiencies in US universities. *Journal of Technology Transfer*, 39(2), pp.247–275.
- Lee, D.H. et al., 2012. Collaboration network patterns and research performance: The case of Korean public research institutions. *Scientometrics*, 91(3), pp.925–942.
- Lim, S., Bae, H. & Lee, L.H., 2011. A study on the selection of benchmarking paths in DEA. *Expert Systems with Applications*, 38(6), pp.7665–7673.
- Liu, J.S. et al., 2009. A network-based approach for increasing discrimination in data envelopment analysis. *Journal of the Operational Research Society*, 60(11), pp.1502–1510.
- Liu, J.S. & Lu, W.M., 2010. DEA and ranking with the network-based approach: A case of R&D performance. *Omega*, 38(6), pp.453–464.
- Lozano, S. & Villa, G., 2010. Gradual technical and scale efficiency improvement in DEA. *Annals of Operations Research*, 173(1), pp.123–136.
- Newman, M.E.J., 2003. The structure and function of complex networks. *SIAM Review*, 45(2), pp.167–256.

- Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The PageRank Citation Ranking. Bringing Order to the Web,
- Park, J., Bae, H. & Lim, S., 2012. Stepwise Benchmarking Path Selection in DEA. *Smart Innovation, Systems and Technologies*, 16, pp.477–484.
- Saramäki, J. et al., 2007. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 75(2), pp.1–4.
- Seiford, L.M. & Zhu, J., 2003. Context-dependent data envelopment analysis - Measuring attractiveness and progress. *Omega*, 31(5), pp.397–408.
- Villa, G., 2003. *Análisis por Envoltura de Datos (DEA): Nuevos Modelos y Aplicaciones*. Univesidad de Sevilla.
- Watts, D.J. & Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature*, 393(6684), pp.440–442.