PROJECTE FINAL DE CARRERA

Llicenciatura en Ciències i Tècniques Estadístiques

Títol: Caracterización de tráfico dinámico en redes ópticas elásticas

Autor: Lluís Virgili Correas

Director: Josep Anton Sànchez Espigares

Co-director: Marc Ruiz Ramírez

Agradecimientos

A mi familia por su apoyo durante el transcurso del proyecto.

A mis directores de proyecto por su paciencia y dedicación en la realización de la memoria.

A Luis Velasco y Alberto Castro por su ayuda complementaria en el proceso de obtención de datos y en mi formación en el mundo de las redes ópticas.

Al Grupo de Comunicaciones Ópticas (GCO) de la UPC por cederme un espacio y un equipo para poder hacer el proyecto.

Índice

Lista de figuras ..... **V**

Lista de tablas ..... **VII**

Capítulo 1 Introducción **1**

1.1 Motivación y objetivos 1

1.2 Organización de la memoria 2

Capítulo 2 Conceptos básicos **3**

2.1 Teoría de grafos 3

2.2 Redes ópticas elásticas 6

2.3 Tráfico dinámico 9

Capítulo 3 Diseño experimental **13**

3.1 Obtención de los datos 13

3.1.1 Definición de las variables respuesta 13

3.1.2 Metodología usada 13

3.2 Escenario 18

3.3 Descripción de las topologías 20

3.3.1 Variables explicativas de la topología 20

3.3.2 Análisis descriptivo de las topologías 21

3.4 Resumen 25

Capítulo 4 Modelización **27**

4.1 Modelización de la intensidad 27
4.1.1 Efecto de la capacidad 27
4.1.2 Efecto de la topología 31
4.1.3 Validación conjunta de residuos 36
4.2 Modelización del perfil de tráfico 39
4.3 Validación mediante las redes de prueba 41
4.4 Conclusión 43

Capítulo 5 Casos de aplicación 45
5.1 Dimensionamiento de redes ópticas 45
5.1.1 Definición del problema 45
5.1.2 Caso de estudio 45
5.2 Planificación de redes ópticas 49
5.2.1 Definición del problema 49
5.2.2 Método de resolución 50

Capítulo 6 Conclusiones 53
6.1 Contribuciones 53
6.2 Desarrollo temporal 55
6.3 Trabajo futuro 55

Apéndice A - Topologías usadas 57

Apéndice B - Código implementado 67

Apéndice C - Tabla de resultados 81

Referencias 87

Lista de figuras

Fig. 2-1 Ejemplo para cálculo del coeficiente de clúster 5
Fig. 2-2 Ejemplo para la conectividad algebraica 6
Fig. 2-3 Representación lógica de un enlace de fibra óptica 7
Fig. 2-4 Establecimiento de conexión óptica 8
Fig. 3-1 Ejemplo de selección de la intensidad máxima 18
Fig. 3-2 Histogramas variables explicativas 23
Fig. 3-3 Análisis de componentes principales 24
Fig. 4-1 Intensidad en función del tamaño de Δ y S 28
Fig. 4-2 Ajuste lineal de I a partir de X 30
Fig. 4-3 Análisis de los residuos del modelo 34
Fig. 4-4 Residuos estandarizados para todas las observaciones 36
Fig. 4-5 Normalidad de los residuos 37
Fig. 4-6 Residuos estandarizados en función del tamaño de espectro y Δ 38
Fig. 4-7 Normalidad de los residuos en función del tamaño de espectro y Δ 38
Fig. 4-8 BW medio de los TP's en función del tamaño de Δ 39
Fig. 4-9 Test de Tukey para comparar medias de los TP's en función de Δ 40
Fig. 4-10 Test de Tukey para comparar medias en función de S 41
Fig. 4-11 Diagrama bivariante entre |E| y κ 42
Fig. A-1 Red TP5 57
Fig. A -2 Red TP6 58
Fig. A -3 Red Cost266 58
Fig. A -4 Red Abilene 59
Fig. A -5 Red TP7 59
Fig. A -6 Red Atlanta 60

Fig. A-7 Red France 60
Fig. A -8 Red Geant 61
Fig. A -9 Red EU 61
Fig. A -10 Red Germany 62
Fig. A -11 Red US 62
Fig. A -12 Red Norway 63
Fig. A -13 Red Polska 63
Fig. A -14 Red EONCT 64
Fig. A -15 Red TID 64
Fig. A -16 Red JANOS 65

Lista de tablas

Tabla 2-1: Eficiencia de transporte de 40 Gbps 9
Tabla 3-1: Algoritmo TPmaker 15
Tabla 3-2: Ejemplos de perfil de tráfico máximo 16
Tabla 3-3: Generación de perfiles con salto de un 1% 17
Tabla 3-4: Dimensiones de las redes de experimento 19
Tabla 3-5: Dimensiones de las redes de test 19
Tabla 3-6: Configuraciones de capacidad 20
Tabla 3-7: Límites máximos y mínimos del perfil de tráfico 20
Tabla 3-8: Valores de las variables explicativas de topología 21
Tabla 3-9: Análisis descriptivo de las variables explicativas 22
Tabla 3-10: Ranking de correlaciones entre variables explicativas 23
Tabla 4-1: Parámetros de ajuste exacto para la red Abilene 28
Tabla 4-2: Parámetros de ajuste de X 30
Tabla 4-3: Valor beta de las redes de estudio 31
Tabla 4-4: Bondad de ajuste de ß = b1*variable + b0 32
Tabla 4-5: Resumen de la bondad de ajuste de los modelos obtenidos 35
Tabla 4-6: Parámetros del modelo de intensidad 36
Tabla 4-7: TP's en función de Δ 41
Tabla 4-8: Error de estimación de I para cada red 42
Tabla 4-9: Error de estimación de TP para cada red 43
Tabla 5-1: Previsiones de la demanda de ancho de banda 46
Tabla 5-2: Coste por unidad de ancho de banda de cada servicio 47
Tabla 5-3: Ingresos previstos (u.c.) por el tráfico de negocios previsto (Gbps) 47
Tabla 5-4: Ingresos previstos (u.c.) por el tráfico de Internet previsto (Gbps) 48

Tabla 5-5: Explotación anual para el tráfico de negocios 48
Tabla 5-6: Explotación anual para el tráfico de Internet 49
Tabla 5-7: Algoritmo de minimización de |E| 51
Tabla 6-1: Diagrama de Gantt del presente proyecto..... 55

Capítulo 1.

Introducción

1.1 Motivación y objetivos

En la actualidad, la mayor cantidad de tráfico transportado en las redes de comunicaciones la constituye el tráfico IP, característico de todos aquellos servicios y aplicaciones basadas en Internet. Este volumen de tráfico va en aumento, creciendo en el tiempo con una tendencia exponencial, con lo que las actuales arquitecturas y tecnologías ópticas de transporte tienden a quedarse obsoletas a una velocidad vertiginosa. Además, el tráfico tiende a ser cada vez más fluctuante en el tiempo y más variable en cuanto a su distribución espacial, con lo que la tecnología actual, basada en el establecimiento de conexiones ópticas de ancho de banda fijo, son demasiado rígidas para sacar un máximo aprovechamiento de la capacidad de las redes. La emergente tecnología de redes ópticas elásticas (actualmente en fase de desarrollo) permite tener redes ópticas de transporte de gran capacidad con una mayor adaptabilidad al ancho de banda demandado que las redes ópticas rígidas.

Ante este panorama, las empresas que se encargan de planificar y gestionar las redes ópticas de comunicaciones están muy interesadas en buscar la forma óptima de diseñar y dimensionar las redes elásticas del futuro. Una figura importante para poder tomar decisiones es conocer la cantidad de tráfico que puede admitir una red sin superar un cierto límite que pueda afectar al servicio que se les ofrece a los clientes. Esta cantidad de ancho de banda servido permite, entre otros, conocer cuántos ingresos se pueden esperar de la explotación de la red o si es necesario ampliar el equipamiento para conseguir unos mínimos de ancho de banda deseables.

El objetivo principal de este proyecto es el de obtener modelos estadísticos para la estimación de la cantidad de tráfico que puede servir una red, sin violar ciertos requisitos de servicio, y con el fin de maximizar el ancho de banda total servido. Las redes consideradas en este proyecto (redes *core*) son redes de alcance nacional o

internacional que constituyen el núcleo de la red de transporte de datos, soportando grandes volúmenes de datos que llegan agregados desde otras redes de ámbito metropolitano o local. Además, consideraremos que la tecnología que dota elasticidad a las redes ópticas es la tecnología *flexgrid* (actualmente en proceso de desarrollo), mientras que consideraremos entornos de tráfico dinámico. Este escenario descrito es de crucial interés para las empresas del sector ya que representa el panorama previsto para los próximos años y, por ende, el escenario a considerar para estos problemas de planificación de red. Estos modelos de predicción (actualmente no existentes) estudiarán la influencia de la topología y de la capacidad de las redes sobre ciertas variables de tráfico. Como característica importante, los modelos deberán tener una estructura orientada a su fácil inserción y utilización en problemas de diseño y dimensionamiento de redes ópticas de transporte.

1.2 Organización de la memoria

El resto de la memoria se distribuye de la siguiente manera. En el capítulo 2 se introducirán todos aquellos conceptos previos necesarios para entender el desarrollo del trabajo realizado. Se pondrá especial énfasis en definir aquellas figuras claves relacionadas con teoría de grafos, redes ópticas elásticas, y entornos de tráfico dinámico.

En el capítulo 3 se mostrará el procedimiento usado para la obtención de los valores de la variable respuesta de los modelos, basados en un simulador de redes ópticas de comunicaciones. Además, se hará un análisis descriptivo de las topologías utilizadas para obtener resultados, todas ellas características de redes de transporte nacionales e internacionales reales.

En el capítulo 4 se detallará todo el proceso de obtención de los modelos estadísticos, haciendo uso de diferentes técnicas estadísticas. Se mostrará la validez de los modelos obtenidos a partir de la muestra de redes utilizada para su elaboración y también a partir de un segundo conjunto de redes de prueba no incluidas para el estudio.

En el capítulo 5 se mostraran dos ejemplos de aplicación de los modelos obtenidos para la resolución de problemas de decisión en planificación de redes ópticas elásticas. Concretamente, se presentará un caso de estudio de equipamiento de redes ópticas existentes y se introducirá un método iterativo para el diseño óptimo de la topología de una red óptica nueva.

Finalmente, el capítulo 6 resumirá las principales contribuciones de este proyecto y las futuras extensiones que quedan abiertas tras su realización.

Capítulo 2.

Conceptos básicos

En esta sección se introducen los conceptos necesarios para la comprensión de este trabajo relacionados con las redes ópticas de transporte. En primer lugar, se presentan diversas definiciones relacionadas con grafos del estilo de los que forman las redes ópticas de transporte. En segundo lugar, se introducen los conceptos más representativos sobre la composición y funcionamiento de las redes ópticas elásticas. Finalmente, se definen aquellas figuras más relevantes relacionadas con entornos de tráfico dinámico.

2.1 Teoría de grafos

En este apartado se describen aquellos conceptos básicos referentes a teoría de grafos, que se pueden encontrar fácilmente en la literatura (e.g. [Vm11]).

La topología o morfología de una red puede describirse mediante un grafo $G(N, E)$, donde N representa el conjunto de nodos (o vértices) y E el conjunto de enlaces (o arcos). La dimensión de estos dos conjuntos (es decir, número de nodos y de arcos) se denota como $|N|$ y $|E|$, respectivamente. Los enlaces pueden ser unidireccionales o bi-direccionales (o, alternativamente, dirigidos o no dirigidos). En nuestro trabajo consideraremos que nuestras redes tienen arcos bidireccionales.

Un camino o ruta sobre un grafo es una secuencia de enlaces o nodos contiguos que unen un par de nodos del grafo. Varias de esas rutas que pueden obtenerse en una red permiten caracterizar la morfología del grafo. Una de esas características más evidentes es la conectividad. Por lo general los grafos que representan redes ópticas de transporte nacionales suelen ser, al menos, grafos 2-conexos. Eso quiere decir que, entre cualquier par de nodos, existen al menos dos caminos disjuntos en arcos (es decir, no comparten ningún enlace entre sí).

En cuanto al tamaño de la red, aparte de $|N|$ y $|E|$ como principales medidas absolutas, el valor relativo más usado es el grado nodal medio (δ). Antes de su

definición, cabe puntualizar que el grado nodal de un nodo es el número de arcos incidentes con él. Así, el grado nodal medio es la media de todos esos grados nodales, y puede calcularse exactamente como:

$$\delta = \frac{2 \cdot |E|}{|N|} \tag{2.1}$$

La topología de un grafo puede almacenarse en forma matricial para su uso en cálculos. De todas las matrices características de un grafo, destacamos dos que serán usadas para cálculos posteriores. La primera es la matriz de adyacencias *M*, que es aquella matriz de dimensión |N|x|N|, donde la celda m_{ij} vale 1 si existe un enlace entre los nodos i y j, 0 en caso contrario. A partir de la matriz anterior (que permite almacenar toda la información necesaria para extraer la morfología del grafo) se puede obtener una matriz con información agregada llamada matriz Laplaciana Λ. Esta matriz Λ se trata de una matriz diagonal de tamaño |N|x|N|, donde el elemento *i* de la diagonal contiene el grado nodal del nodo *i*.

Con todas estas definiciones previas, vamos a definir una serie de valores característicos de las topologías, que pueden ser de especial interés para su uso en modelos de estimación de tráfico. En primer lugar, y en relación con los caminos del grafo, uno de los más importantes es el *camino mínimo medio* (*h*). Para su cálculo hace falta definir h_{ij} como el camino más corto en número de enlaces entre el par de nodos *i-j*. Así, el camino mínimo medio se calcula como la media de todos esos caminos mínimos entre cualquier par de nodos. Tomando como referencia estos valores h_{ij}, podemos definir dos nuevas variables, el *radio* (*rad*) y el *diámetro* (*dia*) del grafo de la siguiente manera:

$$rad = \min_{i \in N} \left(\max_{\substack{j \in N \\ i \neq j}} (h_{ij}) \right) \tag{2.2}$$

$$dia = \max_{i \in N} \left(\max_{\substack{j \in N \\ i \neq j}} (h_{ij}) \right) \tag{2.3}$$

Existe una medida que relaciona los caminos mínimos de un grafo y el grado de centralidad de un elemento (nodo o arco) de éste. Esta medida se llama *intermediación* (o *betweeness* en inglés) y para un cierto nodo (o arco) se calcula como el número de caminos mínimos que pasan por ese nodo (o arco). A partir de los valores de intermediación de todos los elementos del grafo, se pueden sacar valores agregados como la media o la desviación estándar. Esta segunda tiene la particularidad de dar información acerca de cómo se distribuyen los caminos mínimos por la red. Así, un valor bajo indicará que todos los nodos (o arcos) son utilizados de forma más o menos uniforme, lo que puede indicar que la carga de una red de transporte con dicha topología esté distribuida, de forma natural, de un modo más balanceado pudiendo ser ello positivo para un mejor uso de la capacidad de la red.

Otras medidas importantes son aquellas que dan información acerca de la conectividad de la red. De entre ellas, destacamos dos importantes: el coeficiente de clúster medio y la conectividad algebraica. El primero de ellos se calcula como la media del coeficiente de clúster de todos los nodos del grafo. La ecuación (1.4) muestra el cálculo del coeficiente de clúster para un nodo n, donde d representa el grado nodal del nodo e y representa el número de arcos entre nodos adyacentes con en el nodo n.

$$c = \frac{2 \cdot y}{d * (d-1)} \tag{2.4}$$

Para ilustrar el cálculo de este coeficiente, procederemos a calcular el coeficiente de clúster del nodo A de la red de la Fig. 2-1. Para calcular el coeficiente de clúster de A, obtenemos que d=4 y que los nodos adyacentes con A son B, C, D, y E. Entre estos cuatro nodos, existe un único arco (C-D), por lo que y=1, dando como resultado c(A)=1/6. Este coeficiente está relacionado con el grado de conectividad de la red, de forma que un valor medio elevado indica una alta conectividad del grafo.

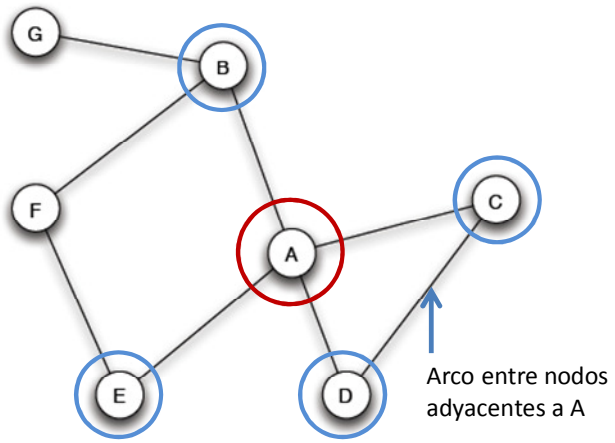


Fig. 2-1 Ejemplo para cálculo del coeficiente de clúster

La conectividad algebraica corresponde al segundo valor propio de la matriz Q formada por la resta de la matriz Laplaciana y de la matriz de adyacencias (es decir, Q= Λ-M). La conectividad algebraica sirve como medida de la dificultad de formar subgrafos a partir de la eliminación de arcos del grafo principal. En otras palabras, informa de lo bien conectados que están los nodos de una red entre sí. La Fig. 2-2 tal muestra el valor de la conectividad algebraica para varias topologías, mostrando los enlaces que hay que quitar para separar la red en al menos dos conjuntos de nodos no conexos (líneas punteadas). Como se puede ver, un valor mayor de conectividad implica un número mayor de enlaces a quitar y, por lo tanto, un mayor grado de conectividad.

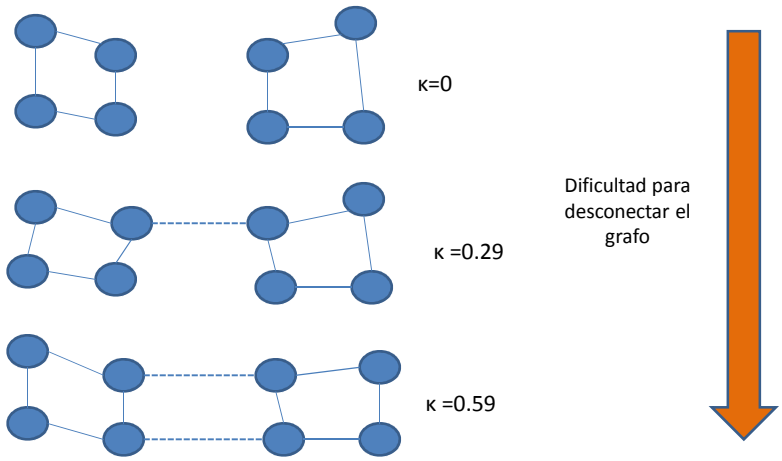


Fig. 2-2 Ejemplo para la conectividad algebraica

2.2 Redes ópticas elásticas

De forma genérica, podemos decir que una red óptica está constituida por una topología como las descritas en la subsección anterior, junto con una serie de equipos y dispositivos ópticos que la caracterizan. Así, un enlace se corresponde con una fibra óptica y un nodo representa todo el equipamiento necesario para poder establecer, enrutar, y eliminar conexiones ópticas a través de las fibras de la red. Este equipamiento de la red añade una nueva dimensión a la topología del grafo: la capacidad. Un enlace de fibra óptica tiene una capacidad útil que denominaremos *espectro* o *ancho espectral* (S), que se mide en GigaHerzios (GHZ) y que viene caracterizado en gran parte por el tipo de equipamiento instalado en los nodos.

Las conexiones ópticas son caminos (o también llamados circuitos) que se establecen en la red y que permiten transmitir datos en forma de onda de luz. En una misma fibra y momento pueden transportarse varias de estas conexiones simultáneamente porque cada una de ellas ocupa una porción del espectro óptico diferente. Es decir, si dividimos el espectro en varios fragmentos (o longitudes de onda), cada uno de ellos solo podrá soportar una conexión en un momento dado. Esta multiplexación del espectro óptico es lo que dota a las redes ópticas de una gran capacidad para el transporte de datos.

Las conexiones ópticas se establecen con el fin de servir las peticiones de conexión de clientes que llegan a la red. Una petición de conexión consta mayoritariamente de la siguiente información: los nodos origen y destino para realizar la conexión, y la cantidad de ancho de banda a transportar, generalmente expresado en Megabits por segundo (Mbps) o Gigabits por segundo (Gbps). El proceso de establecimiento de estas conexiones ópticas se obtiene tras resolver el problema denominado de enrutamiento y asignación de espectro (en inglés, *routing and spectrum assignment*

, RSA). El objetivo es encontrar una ruta que contenga espectro suficiente para albergar el ancho de banda demandado por el cliente, estableciéndose si esa capacidad libre existe o rechazándose en caso contrario.

Una *red óptica elástica* se puede definir como aquella red óptica que tiene la capacidad de adaptar el espectro asignado de las conexiones ópticas a los requerimientos de ancho de banda del tráfico transportado. Estas redes se oponen a las *redes ópticas rígidas*, donde cada conexión óptica ocupa la misma cantidad de espectro. Esto lleva a que en redes ópticas rígidas (como las actualmente utilizadas redes DWDM [ITU-G.694]) sea necesario agrupar mucho el tráfico en conexiones ópticas de gran ancho de banda para no desaprovechar recursos de capacidad. Las redes ópticas elásticas permiten una mayor versatilidad y adaptabilidad con el fin de obtener un mejor aprovechamiento de los recursos ópticos.

De entre las tecnologías que se están estudiando para el desarrollo de redes ópticas elásticas, la que está proporcionando un mejor balance entre rendimiento y complejidad tecnológica es la tecnología *flexgrid* [LI11]. En este tipo de redes, el espectro óptico está dividido en porciones de idéntica anchura, denominadas *segmentos*. La anchura de estos segmentos puede ser diferente según los equipos instalados en la red, así que consideraremos el parámetro Δ como la anchura del segmento en GHz (en general serán valores pequeños). Cada segmento ocupa una posición en el espectro, caracterizada por su frecuencia central (CF). Esto nos permite ordenar los segmentos de forma lógica a partir de sus frecuencias centrales, de menor a mayor. Cada conexión óptica establecida utiliza un número consecutivo de esos segmentos (lo que se llama *contigüidad de espectro*), en función del ancho de banda requerido. Esta porción de espectro formada por segmentos contiguos y destinada al transporte de una conexión se llama *canal*, y se puede caracterizar mediante su CF. Sin entrar en detalles, el ancho de banda que puede transportar cada segmento dependerá del formato de modulación de la onda. Para nuestro estudio consideraremos que el número de Gbps que puede llevar un segmento es de hasta 2*Δ. La Fig. 2-3 muestra un esquema lógico de un enlace de fibra de una red óptica elástica con esta tecnología flexgrid, que es la que vamos a considerar en este trabajo.

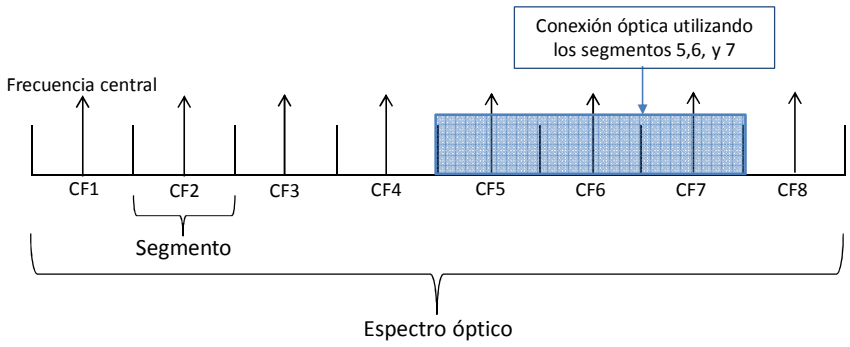


Fig. 2-3 Representación lógica de un enlace de fibra óptica

La Fig. 2-4 muestra un ejemplo de cómo se establecería una petición de conexión entrante en el nodo B con destino al nodo D, demandando una cantidad de gigabits que debería ser servida por 2 segmentos. En la fase de búsqueda de ruta se debe encontrar una que tenga suficiente capacidad utilizable, generalmente la más corta posible (porque es la que utiliza menos recursos). No obstante, una condición que tienen que cumplir las conexiones ópticas es que la porción de espectro usada en cada enlace de la ruta sea la misma (es lo que se denomina *continuidad de espectro*). Como vemos, aunque la ruta formada por los arcos B-A y A-D es la más corta en número de saltos y tiene capacidad libre disponible, no se puede cumplir la condición de continuidad. Es por eso que la ruta finalmente establecida es la que sigue los enlaces B-A, A-C, y C-D, que sí que tiene suficientes segmentos continuos en todos los enlaces.

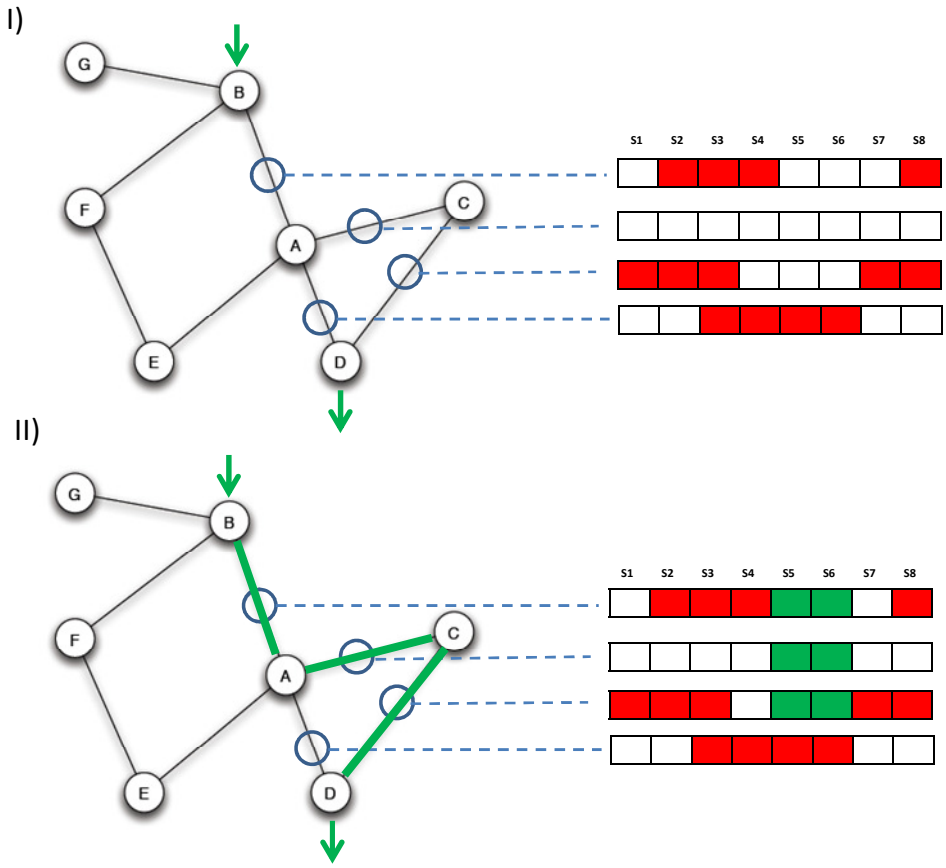


Fig. 2-4 Establecimiento de conexión óptica

Respecto a las distintas clases de clientes que puede transportar una red, para nuestro estudio solo distinguiremos tipos de clientes en función del ancho de banda demandado. Así, consideraremos anchos de banda (o alternativamente velocidades) de 10, 40, 100, y 400 Gbps para identificar nuestras 4 clases o tipos diferentes de clientes. Esta combinación de distintos requerimientos de demanda lleva a que, según sea el tamaño de segmento Δ, el aprovechamiento del ancho espectral S sea

diferente. Por ejemplo, si contemplamos una demanda de 40 Gbps (que con el formato de modulación escogido necesitaría 20 GHz de espectro contiguo), ésta usará una cantidad real de espectro diferente en función de Δ, como muestra la Tabla 2-1. En ella vemos como el uso de segmento de 50 GHz lleva a un uso poco eficiente del espectro ya que se pierde mucho espectro en forma de sobrante (de cada segmento, 30 GHz son malgastados). En cambio, con un tamaño menor se gana en eficiencia, ya que es más fácil adaptarse a las necesidades de ancho de banda de las demandas (es decir, el espectro se divide en una granularidad más fina). No obstante se ha demostrado en estudios de investigación que cuanto mayor es el ancho de banda demandado (se entiende como el promedio de todos los tipos de cliente), mayor es el beneficio que se obtiene de los tamaños grandes de segmento, ya que con tamaños pequeños de éste se produce demasiada fragmentación del espectro que dificulta encontrar grandes porciones de fragmento contiguo y continuo [Ca12]. Es por eso que, cuando nos refiramos a la capacidad de las redes ópticas elásticas, lo haremos especificando no solo el valor de S sino también el de Δ (lo que llamaremos *configuración de capacidad*).

Tabla 2-1: Eficiencia de transporte de 40 Gbps

Δ (GHz)	**# segmentos**	**Eficiencia**
50	1	40%
25	1	80%
12.5	2	80%

La siguiente sección introduce los conceptos más relevantes relacionados con entornos de tráfico dinámico en redes ópticas elásticas.

2.3 Tráfico dinámico

Como hemos dicho anteriormente, las redes ópticas elásticas se despliegan con el objetivo de proporcionar servicio de conectividad a ciertos clientes. Esta conectividad se realiza mediante el establecimiento de caminos, es decir, creando una conexión entre dos extremos de la red, que permanece establecida durante un tiempo hasta que se elimina. Este funcionamiento es similar al de las redes telefónicas, donde se establece una conexión para una llamada entre dos abonados, la conexión permanece establecida durante el tiempo que dura la conversación y se deshace o elimina al final, liberando recursos en la red que podrán ser utilizados por otras conexiones.

Precisamente de las redes telefónicas proviene el término de *intensidad de tráfico* (*I*) para medir la cantidad media de conexiones que satisface una red en un cierto instante de tiempo. De forma genérica, consideraremos que las peticiones de conexión que llegan a un nodo de la red siguen una distribución de Poisson con una

tasa de llegada λ, cuya inversa se define como *inter-arrival time* (*iat*). Por otra parte, el tiempo que esa conexión permanece establecida (si ha podido ser servida) sigue una distribución exponencial con tasa de servicio μ, cuya inversa se denomina *holding time* (*ht*). Entonces, la intensidad se calcula según la ecuación (1.5), donde la unidad de medida es el Erlang.

$$I = \frac{ht}{iat} \tag{2.5}$$

Una característica básica de las redes de telefonía que dieron lugar a toda esta teoría de tráfico dinámico (también denominada *teoría de tele-tráfico* [ITU05]) es que las conexiones tenían un ancho de banda fijo e idéntico. En redes elásticas, como ya hemos visto, según sea la petición de ancho de banda, la conexión ocupará un tamaño de espectro diferente. Eso nos lleva a concluir que la intensidad no es suficiente para caracterizar el tráfico transportado en la red. Por eso, definimos el *perfil de tráfico* (*TP*) como la proporción media de peticiones de cada tipo de cliente que llegan a la red. Este perfil, que tendrá como dimensión el número de clases distintas de cliente, es un vector de distribución de probabilidades cuya suma debe dar 1.

En entornos de tráfico dinámico, el grado de servicio de la red se mide con la probabilidad de que una petición de conexión no pueda ser aceptada en el momento de su llegada. Esta probabilidad se denomina *probabilidad de bloqueo* (*PB*) y es una de las figuras más usadas para evaluar la capacidad efectiva de la red. Así, podemos considerar una probabilidad de bloqueo del 1% como el valor más alto admisible para dar un servicio mínimamente decente a nuestros clientes. Aunque la red tenga capacidad sobrante, no vamos a permitir más intensidad ya que empeoraría el grado de servicio alcanzando un nivel no deseable. Otra vez aquí, nos encontramos con que no es lo mismo rechazar una demanda de 400 Gbps que una de 10 Gbps, ya que la cantidad de ancho de banda rechazado es mucho menor. Así, en redes ópticas elásticas surgen dos medidas de probabilidad de bloqueo: la parcial de cada tipo de cliente y la ponderada, haciendo uso de las proporciones del perfil de tráfico y de sus velocidades. A esta probabilidad de bloqueo ponderada se le puede llamar con mayor precisión como la proporción de ancho de banda rechazado.

Finalmente, para definir la cantidad media de ancho de banda total que transporta una red (BW_{total}) en un determinado momento necesitamos conocer la intensidad global (I) y, para cada tipo de cliente i su proporción en el perfil de tráfico (TP_i), su ancho de banda demandado (bw_i), y su probabilidad de bloqueo parcial (PB_i). Así, el cálculo se puede efectuar como sigue en la ecuación (1.6). Una muy buena aproximación de BW_{total} se puede obtener si no se considera la probabilidad de bloqueo. Esto ocurre cuando restringimos ésta a valores pequeños (e.g. 1%).

$$BW_{total} = I \cdot \sum_{\forall i} bw_i \cdot TP_i \cdot \left(1 - PB_i\right) \tag{2.6}$$

Considerando este entorno de tráfico dinámico y las características de las redes ópticas elásticas definidas anteriormente, el siguiente capítulo explica cómo hemos obtenido los valores de la variable respuesta, que serán fundamentales para la posterior búsqueda de los modelos.

Capítulo 3.

Diseño experimental

Este capítulo consta de tres partes: en primer lugar, se muestra el procedimiento para obtener valores óptimos para las variables de respuesta; en segundo lugar, se definen el escenario utilizado, poniendo especial énfasis en las redes y configuraciones de capacidad utilizadas; por último, se muestra un análisis descriptivo de las variables explicativas de la topología, con el objetivo de encontrar información previa útil para la búsqueda de modelos.

3.1 Obtención de los datos

3.1.1 Definición de las variables respuesta

Como se ha definido en el capítulo introductorio, el objetivo de este trabajo es encontrar modelos de predicción de tráfico para la maximización del ancho de banda total servido. Para ello, necesitamos encontrar dos incógnitas relacionadas entre sí:

I: Intensidad en Erlangs que es capaz de admitir la red.

TP: Vector de perfil de tráfico, con la proporción de conexiones de cada tipo de cliente.

Resulta obvio decir que la obtención de valores de estas variables respuesta dada una red concreta no resulta trivial. Para obtener valores óptimos para I y TP, hemos diseñado un algoritmo basado en la ejecución de un simulador de redes ópticas. Se describe su funcionamiento en la siguiente sección.

3.1.2 Metodología usada

Para obtener los valores de las variables respuesta, hemos diseñado un algoritmo que usa un simulador de tráfico dinámico para redes ópticas elásticas. Este

simulador de tráfico dinámico (*RSAsim*) está implementado en OMNeT [OM] y se trata de un clásico simulador por eventos discretos. Éste recibe como datos de entrada un valor de intensidad de tráfico *I*, el perfil de tráfico *TP*, la topología *G* y la configuración de capacidad de la red (S, Δ). Al finalizar la simulación tras un cierto criterio de parada (e.g. un cierto número de peticiones de conexión generadas), el simulador devuelve la probabilidad de bloqueo obtenida para cada uno de los anchos de banda del perfil de tráfico (vector *PB*). Sin intención de definir exhaustivamente el pseudocódigo del simulador, los eventos más importantes a tener en cuenta para la comprensión de este trabajo son:

1) **Inicialización de las primeras peticiones de conexión**: puesto que consideramos que cada nodo es generador de tráfico con la misma probabilidad (generación uniformemente distribuida), se inicializa el tiempo de simulación de la primera llegada para cada nodo siguiendo la tasa 1/*iat*. Cabe resaltar que hemos fijado un *ht* constante para todos los experimentos, siendo *iat* el único valor variable para cada *I*.

2) **Llegada de una nueva conexión**: cuando se produce este evento para un nodo determinado, se procede de la siguiente forma:

    a. En primer lugar se genera el tiempo de llegada de la próxima petición con origen en ese nodo.

    b. Seguidamente, se escoge el nodo destino de esa petición, escogiendo cualquier otro de los nodos de la red con idéntica probabilidad (trafico uniformemente distribuido). El ancho de banda demandado por esta petición se calcula de forma aleatoria a partir del vector de probabilidades TP.

    c. Luego, se procede a buscar ruta y asignación de espectro factible a partir del algoritmo de RSA presentado en [Ca12]. Sin entrar en detalles, este algoritmo de RSA busca la ruta más corta con canal disponible para albergar el ancho de banda demandado. En el caso de haber más de un canal, el algoritmo elige el canal con menor índice de orden (estrategia *first-fit*).

    d. Si el algoritmo RSA encuentra ruta y asignación para la demanda, se establece esa conexión, reservando los recursos de capacidad necesarios. El tiempo de establecimiento se determina a partir de una distribución exponencial con tasa de servicio 1/*ht*. Se genera el evento de fin de conexión para el tiempo de simulación determinado.

    e. En caso de no encontrar ruta factible, la petición de conexión se rechaza por falta de recursos disponibles y se actualiza la proporción de conexiones bloqueadas del ancho de banda correspondiente.

3) **Fin de conexión establecida:** al producirse este evento, la conexión se elimina, liberando así todos los recursos ocupados por dicha conexión.

Como hemos anunciado anteriormente, hemos diseñado e implementado un algoritmo que ejecuta el simulador *RSAsim* de forma iterativa (cambiando los valores de entrada en cada iteración) con el fin de encontrar los valores de *I* y *TP* que maximizan el ancho de banda total. La Tabla 3-1 muestra el pseudocódigo de este algoritmo (llamado *TPmaker*). Partiendo de un valor bajo de intensidad, el simulador se ejecuta iterativamente para diversos perfiles de tráfico en orden descendiente de ancho de banda medio hasta obtener una ejecución donde todas las probabilidades de bloqueo obtenidas sean inferiores a los valores máximos permitidos. Se incrementa el valor de la intensidad en una cierta medida y se repite el proceso. El algoritmo finaliza cuando se llega a una intensidad para la que no es posible asegurar los valores de probabilidad de bloqueo con ningún perfil de tráfico. La salida del algoritmo está formada por todos los pares intensidad/perfil de tráfico obtenidos. Cabe destacar que, para la generación del perfil de tráfico, se cuenta con unas cotas mínima y máxima para cada tipo de conexión (vectores *TPmin* y *TPmax* respectivamente) y se impone que la proporción de cada cliente debe estar entre esos umbrales mínimo y máximo. Esta limitación tiene sentido porque es lógico que existan unos mínimos que servir de cada una de las clases de tráfico. Además, aunque podamos elegir cuantos recursos podemos destinar a cada clase de tráfico, nunca podremos sacar provecho a recursos que no se utilicen por falta de demanda, lo que justifica el uso de cotas superiores.

Tabla 3-1: Algoritmo TPmaker

```
Entrada: Topología G, configuración capacidad (S, Δ), valores máximos
y mínimos del perfil de tráfico TPmax y TPmin, probabilidades de
bloqueo máximas PB
Salida: Conjunto de pares intensidad/perfil de trafico R

Generar el conjunto Ω con todos los perfiles de tráfico entre TPmax y
TPmin, en orden descendiente de ancho de banda medio.
R=∅
I := 1 erlang, incI= 1 erlang
ite:=1
fin='Falso'
mientras no fin hacer
    para todo tp en Ω hacer
        sol='Falso'
        PBite=simulador(I,tp)
        si PBite(i)<PB(i) para todo i entonces
            sol='Verdadero'
            R(ite)={I, tp}
            I=I+incI
            ite++
            salir bucle para todo
    si sol='Falso' entonces
    fin='Verdadero'
devolver R
```

El procedimiento para obtener los perfiles de tráfico que se testean en el algoritmo es el siguiente: a partir de los límites superior e inferior de cada clase de servicio,

se forma el perfil de tráfico con mayor ancho de banda medio. Este será aquel que cargue el máximo peso posible en las clases con mayor ancho de banda. Vale la pena recordar que, en cualquier caso, la suma del vector TP debe ser 100%. La Tabla 3-2 ilustra el perfil de tráfico con mayor ancho de banda medio para 3 limitaciones distintas de TPmax y TPmin.

Tabla 3-2: Ejemplos de perfil de tráfico máximo

	10 Gbps	**40 Gbps**	**100 Gbps**	**400 Gbps**
Ejemplo 1				
TPmax	100%	100%	100%	100%
TPmin	0%	0%	0%	0%
Perfil máximo	**0%**	**0%**	**0%**	**100%**
Ejemplo 2				
TPmax	80%	80%	80%	80%
TPmin	10%	10%	10%	10%
Perfil máximo	**10%**	**10%**	**10%**	**70%**
Ejemplo 3				
TPmax	50%	40%	30%	15%
TPmin	30%	20%	10%	5%
Perfil máximo	**30%**	**25%**	**30%**	**15%**

Tomando como valor inicial ese perfil máximo y los valores límites TPmin y TPmax, el resto de perfiles de tráfico se generan a partir de desplazar una cierta unidad (por ejemplo, un 1%) desde las demandas con mayor carga hasta las de menor. La Tabla 3-3 ilustra la obtención de todos los perfiles de tráfico que se pueden llegar a ejecutar en el algoritmo a partir de un perfil inicial máximo. Podemos observar como el ancho de banda medio del perfil va disminuyendo progresivamente.

Una vez obtenidos todos los pares de intensidad/perfil de tráfico debemos seleccionar uno sólo de ellos como valores de la variable respuesta a modelar. Tras ejecutar el algoritmo con distintas topologías y configuraciones de capacidad, observamos comportamientos similares a los descritos en la Fig. 3-1. En ella se puede observar el siguiente comportamiento: para los valores más bajos de intensidad, como se puede servir el perfil de tráfico con mayor ancho de banda medio, el ancho de banda total va incrementando linealmente. Una vez se llega a la intensidad a partir de la cual no es posible servir dicho perfil de tráfico (porque se viola alguna probabilidad de bloqueo límite), se van obteniendo otros perfiles de tráfico óptimos que dibujan una curva variable con algunos máximos locales, hasta

Tabla 3-3: Generación de perfiles con salto de un 1%

	10 Gbps	40 Gbps	100 Gbps	400 Gbps	BW medio (Gbps)
Perfil 1 (máximo)	10%	10%	10%	70%	295
Perfil 2	10%	10%	11%	69%	292
Perfil 3	10%	10%	12%	68%	289
			…		
Perfil i	10%	10%	70%	10%	115
Perfil i+1	10%	11%	69%	10%	114.4
Perfil i+2	10%	12%	68%	10%	113.8
			…		
Perfil j	68%	12%	10%	10%	61.6
Perfil j+1	69%	11%	10%	10%	61.3
Perfil j+1 (mínimo)	70%	10%	10%	10%	61

que se llega a la intensidad máxima factible. Con el fin de seleccionar un único par, decidimos escoger uno de los máximos locales de esta curva dentro de un intervalo marcado por el máximo global (e.g. todos los máximos dentro del intervalo del 5%). El máximo escogido es aquel cuyo valor de intensidad es el más alto. Esta estrategia permite elegir, de entre el conjunto de resultados con mayor ancho de banda total, aquel que permite servir a un mayor número de clientes en la red (es decir, el de mayor I). Cabe destacar que, aunque hay valores que no son máximos locales y que dentro del intervalo descrito tienen un valor mayor de I, hemos decidido optar por esta estrategia de selección a causa del comportamiento de la desviación estándar observado en los máximos. Así, hemos observado que, después de lanzar varias ejecuciones y calcular las medias del ancho de banda servido por cada punto de intensidad, la desviación estándar observada en los máximos es en general menor que en el resto de puntos.

Una vez escrito el procedimiento para obtener los valores de las variables respuesta, vamos a proceder a describir el escenario que hemos diseñado para obtener resultados válidos para la obtención de modelos de predicción útiles.

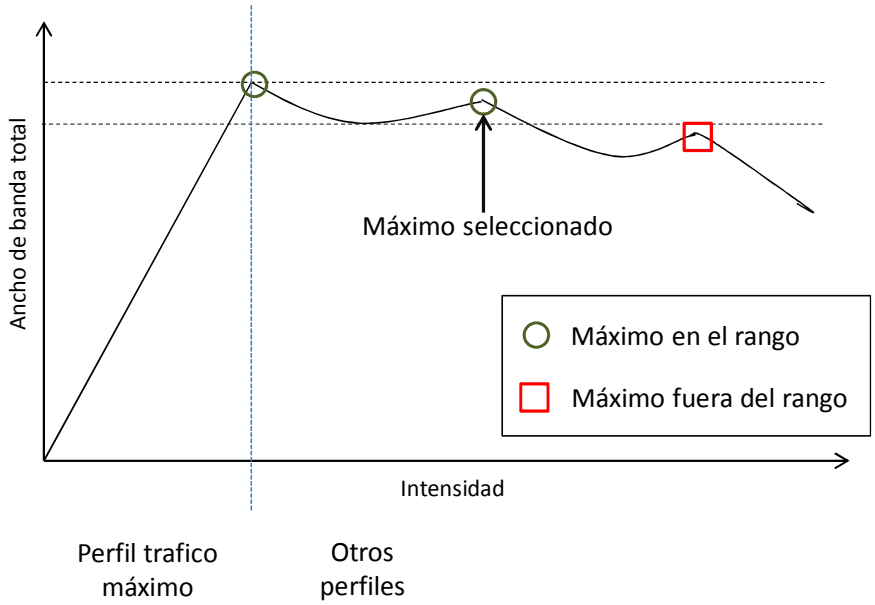


Fig. 3-1 Ejemplo de selección de la intensidad máxima

3.2 Escenario

Para realizar nuestro estudio hemos seleccionado 16 redes ópticas nacionales del conjunto de redes reales disponibles en la literatura (véase, por ejemplo, [Ma03], [SNDlib], [Pa10]). Las topologías de cada una de estas redes pueden consultarse en el apéndice A. Un total de 13, las hemos utilizado para estimar los modelos de intensidad, mientras que las 3 restantes las hemos reservado para validar los modelos una vez construidos. La Tabla 3-4 muestra las dimensiones de las redes de experimento, mientras que la Tabla 3-5 contiene las de las redes de test. Cabe destacar que las redes nacionales suelen tener un tamaño relativamente reducido en comparación con otras redes de transporte (por ejemplo, redes IP, que pueden llegar a contener varios centenares de nodos). Así, podemos observar como el número máximo de nodos, arcos, y de grado nodal medio no supera los 37, 57, y 3.78 respectivamente. No obstante, hemos intentado elegir la mayor diversidad posible entre el conjunto de redes reales y, así, tener un modelo más robusto para predicción y su uso en diseño. Además, podemos comprobar como las redes reservadas para validación están perfectamente en concordancia con las utilizadas para obtener los modelos, cubriendo diferentes tamaños.

Como se introdujo en la definición de los conceptos básicos sobre redes ópticas, cada par de valores (S, Δ) constituye una configuración de capacidad diferente. Así, podemos tener resultados muy diversos sobre la misma red con solo cambiar la configuración de capacidad. Con el fin de cubrir un amplio rango de configuraciones

Tabla 3-4: Dimensiones de las redes de experimento

Red	|N|	|E|	δ
Abilene	12	15	2.50
Atlanta	15	22	2.93
Cost266	37	57	3.08
EU	28	41	2.93
France	25	45	3.60
Germany	17	26	3.06
Norway	27	51	3.78
Polska	12	18	3.00
TP5	10	15	3.00
TP6	10	12	2.40
TP7	21	25	2.38
UK	20	32	3.20
US	14	21	3.00

Tabla 3-5: Dimensiones de las redes de test

Red	|N|	|E|	δ
EON CT	16	23	2.88
TID	21	35	3.33
JANOS	33	55	3.33

de capacidad, hemos elegido 3 tamaños de S y 3 tamaños de Δ, que combinaremos entre sí para obtener 9 configuraciones distintas. La Tabla 3-6 muestra todas esas configuraciones, que se probaran en todas las redes. Para la selección de Δ, se ha tenido en cuenta la propuesta efectuada en el estándar ITU-G.694.1 [ITU-G.694] [G.107] propuesto por la International Telecommunications Union, donde se proponen los tamaños de Δ con granularidad m*6.25GHZ, con m={1,2,4,8}. Los más comunes son 12.5, 25, y 50GHz. Por lo que respecta al tamaño del espectro, hemos utilizado tres dimensiones: S=800, 1600, y 3200 GHz.

Un concepto importante a la hora de encontrar el TP óptimo que maximiza el ancho de banda total servido es la limitación que pueda existir para cada valor del perfil de tráfico (es decir, TPmin y TPmax). Para nuestro estudio, hemos utilizado un rango muy amplio e idéntico para cada clase de cliente. Esto quiere decir que se pueden dar proporciones altas o bajas sea cual sea el ancho de banda demandado.

Tabla 3-6: Configuraciones de capacidad

	S	Δ
C1	800	50
C2	800	25
C3	800	12.5
C4	1600	50
C5	1600	25
C6	1600	12.5
C7	3200	50
C8	3200	25
C9	3200	12.5

La Tabla 3-7 muestra los valores fijados de TPmin y TPmax. Aunque este escenario puede resultar algo alejado de la realidad actual (ya que algunos patrones surgidos bajo este escenario serian difíciles de cumplir), nos sirve para poder estudiar de forma amplia el impacto de todos los factores sobre las variables respuesta.

Tabla 3-7: Límites máximos y mínimos del perfil de tráfico

	10 Gbps	40 Gbps	100 Gbps	400 Gbps
TPmax	80%	80%	80%	80%
TPmin	10%	10%	10%	10%

Con el fin de encontrar relaciones entre las distintas redes usadas y los valores de las variables respuesta, hace falta identificar las topologías a partir de variables explicativas que den información útil acerca de éstas. La siguiente sección se centra en este aspecto.

3.3 Descripción de las topologías

3.3.1 Variables explicativas de la topología

Como hemos detallado anteriormente, simularemos el comportamiento de las variables respuesta en función de unos escenarios que dependerán de la topología y de la configuración de capacidad. De cada una de las redes introducidas en la Tabla 3-8 vamos a usar una serie de características que nos dan información topológica como variables explicativas de nuestros modelos. Estas son:

|N|: Número de nodos (o vértices) de la red.

|E|: Número de arcos (o enlaces) de la red.

δ (o delta): Grado nodal medio de la red.

h: Camino mínimo medio de la red.

rad: Radio de la red.

dia: Diámetro de la red

c: Coeficiente de clúster medio de la red.

sc: Desviación estándar del coeficiente de clúster de la red.

bl: Media del coeficiente de intermediación por enlace.

sbl: Desviación estándar del coeficiente de intermediación por enlace.

bn: Media del coeficiente de intermediación por nodo.

sbn: Desviación estándar del coeficiente de intermediación por nodo.

κ (o kappa): Conectividad algebraica de la red.

A grandes rasgos, tenemos variables que nos hablan claramente de la dimensión de la red (|N| y |E|, principalmente), mientras que el resto están más relacionadas con la forma (h, rad, dia, c, etc). Resultará útil realizar un breve análisis descriptivo en aras de conocer sus medias, correlaciones, etc… para luego disponer de una información previa con tal de conocer las variables susceptibles de entrar en el modelo. La siguiente sección ilustra este análisis.

3.3.2 Análisis descriptivo de las topologías

La Tabla 3-8 muestra los valores de cada una de las variables explicativas definidas anteriormente para cada una de las redes. Se pueden observar valores distintos entre redes dentro de los propios rangos de cada variable, lo que reafirma la diversidad de la muestra elegida.

Tabla 3-8: Valores de las variables explicativas de topología

	networkId	N	E	delta	h	rad	dia	c	c+	sc	bl	sbl	bn	sbn	kappa
1	Abilene	12	15	2.50	2.47	3	5	0.14	1.14	0.30	10.87	4.83	8.08	6.45	0.31
2	Atlanta	15	22	2.93	2.50	3	5	0.37	1.37	0.36	11.95	5.42	10.53	8.91	0.43
3	Cost266	37	57	3.08	3.74	5	8	0.00	1.00	0.00	43.71	27.49	49.33	48.31	0.16
4	EU	28	41	2.93	3.56	4	8	0.00	1.00	0.00	32.94	16.36	34.71	30.07	0.17
5	France	25	45	3.60	2.62	3	5	0.56	1.56	0.37	17.47	12.19	19.44	32.12	0.35
6	Germany	17	26	3.06	2.70	3	6	0.16	1.16	0.24	14.14	7.57	13.62	15.35	0.30
7	Norway	27	51	3.78	3.13	4	7	0.31	1.31	0.23	21.61	12.94	27.81	24.88	0.29
8	Polska	12	18	3.00	2.14	3	4	0.15	1.15	0.17	7.83	2.17	6.25	4.66	0.71
9	TP5	10	15	3.00	1.87	2	3	0.52	1.52	0.40	5.60	3.14	3.90	7.21	0.59
10	TP6	10	12	2.40	2.29	3	5	0.00	1.00	0.00	8.58	2.84	5.80	4.16	0.38
11	TP7	21	25	2.38	3.50	4	8	0.00	1.00	0.00	29.40	10.24	25.00	15.87	0.17
12	UK	20	32	3.20	2.80	4	5	0.21	1.21	0.19	16.62	8.58	17.10	9.91	0.34
13	US	14	21	3.00	2.13	3	4	0.13	1.13	0.16	9.24	3.02	7.36	5.42	0.71

La Tabla 3-9 muestra el análisis de máximos, mínimos, y medias, así como la de los principales cuartiles. En esta tabla se observa que las variables c y sc tienen al menos una observación con valor igual a 0. Con el fin de evitar problemas en el uso de futuras transformaciones durante la modelización (e.g. logaritmos), hemos creado una nueva variable consistente en sumarle 1 a c (llamada c+). Como nota adicional, decir que aunque en esta tabla no acabe de apreciarse, las variables N, E, rad y dia son discretas mientras que el resto son continuas.

Tabla 3-9: Análisis descriptivo de las variables explicativas

	N	E	delta	h	rad	dia	c
Min.	10.00	12.00	2.380	1.870	2.000	3.000	0.0000
1st Qu.	12.00	18.00	2.930	2.290	3.000	5.000	0.0000
Median	17.00	25.00	3.000	2.620	3.000	5.000	0.1500
Mean	19.08	29.23	2.989	2.727	3.385	5.615	0.1962
3rd Qu.	25.00	41.00	3.080	3.130	4.000	7.000	0.3100
Max.	37.00	57.00	3.780	3.740	5.000	8.000	0.5600

	c+	sc	bl	sbl	bn	sbn	kappa
Min.	1.000	0.0000	5.60	2.170	3.90	4.16	0.1600
1st Qu.	1.000	0.0000	9.24	3.140	7.36	6.45	0.2900
Median	1.150	0.1900	14.14	7.570	13.62	9.91	0.3400
Mean	1.196	0.1862	17.69	8.984	17.61	16.41	0.3777
3rd Qu.	1.310	0.3000	21.61	12.190	25.00	24.88	0.4300
Max.	1.560	0.4000	43.71	27.490	49.33	48.31	0.7100

Con el fin de ilustrar la distribución de cada una de las variables explicativas, la Fig. 3-2 muestra los histogramas con las frecuencias de cada nivel, con lo que tenemos una idea general de cómo son las redes ópticas respecto a estas variables. Como nota anecdótica, a la vista de los resultados se puede ver que en general son variables que no siguen una distribución normal.

Para intentar tener una idea previa acerca de la validez de cada una de las variables explicativas como contenedoras de información propia, hemos realizado un análisis de correlaciones a pares, para ver si existen variables muy correlacionadas y, por lo tanto, redundantes a la hora de ser introducidas ambas en un modelo. La Tabla 3-10 muestra un ranking de estas correlaciones, que por razones de espacio, sólo se han puesto aquellas superiores a 0.90. Al leer la definición de las variables, no es de extrañar que variables como, por ejemplo, h y dia o h y rad tengan una correlación alta por lo que poner las dos juntas en un modelo quizás no tendría mucho sentido. Obviamente, c y c+ tienen una correlación perfecta porque se trata de la misma característica. Cabe destacar que el número de nodos y de arcos están muy correlacionados. Esto significa que el tamaño de las redes ópticas aumenta proporcionalmente de una forma constante, sin producirse casos extremos (e.g. pocos nodos y muchos enlaces, o viceversa).

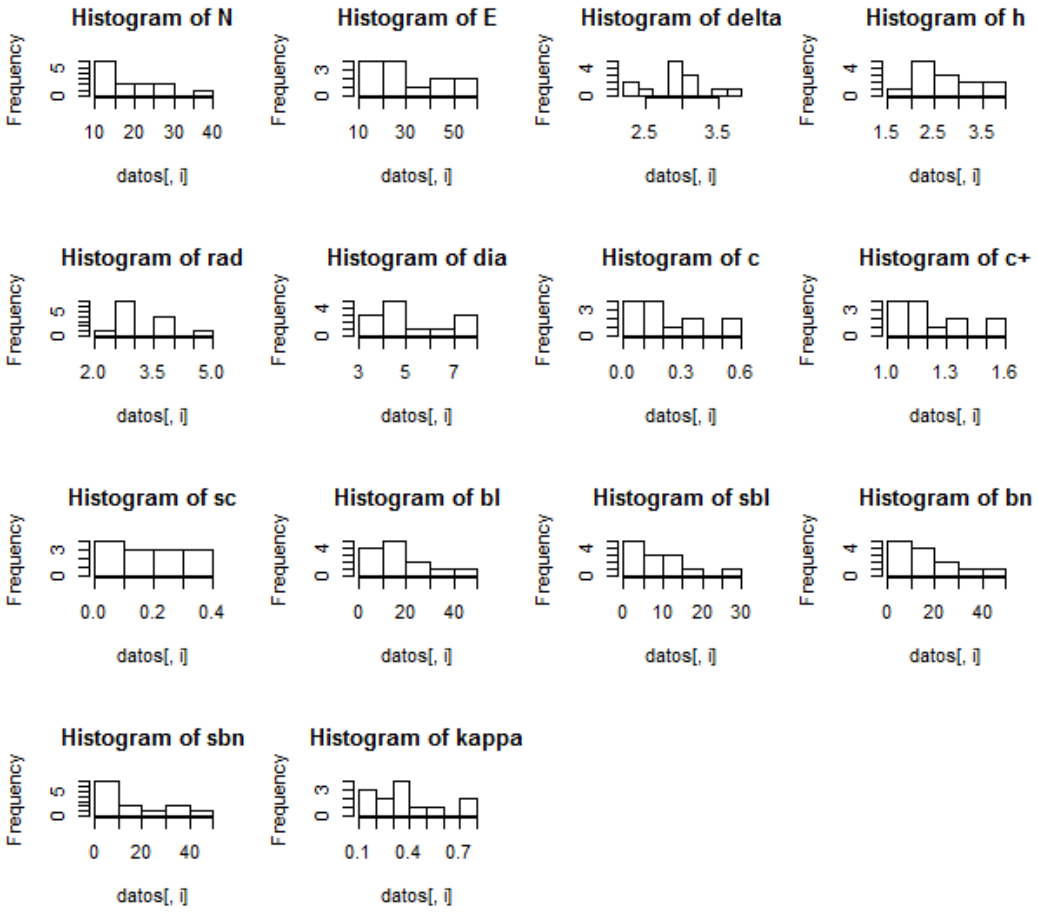


Fig. 3-2 Histogramas variables explicativas

Tabla 3-10: Ranking de correlaciones entre variables explicativas

c	c+	1.00000000
bl	bn	0.98087349
sbl	bn	0.98021000
h	dia	0.97382745
N	bn	0.97188532
N	sbl	0.96989453
sbl	sbn	0.96411993
N	E	0.96342995
h	bl	0.96058485
N	sbn	0.95536218
bl	sbl	0.94689644
h	bn	0.93433711
E	sbn	0.92494062
bn	sbn	0.92347068
N	bl	0.92253077
dia	bl	0.90765926
h	rad	0.90630026
c+	sc	0.90331115
c	sc	0.90331115

Para concluir este apartado, se realizará un análisis de componentes principales con el fin de detectar si hemos elegido un conjunto de redes lo bastante heterogéneo como para utilizar los modelos que se obtengan para un amplio conjunto de casos. El problema es que si se quiere hacer de todas las variables, se tendrán más componentes (variables) que individuos (redes) por lo que se han escogido las variables delta, h, c, bl y bn dado la importancia que suelen tener esas variables en las redes convencionales [Pa10]. La Fig. 3-3 muestra las gráficas más representativas en análisis de componentes principales.

En el gráfico de individuos, se observa que no hay un gran solapamiento de las redes. Eso quiere decir que las redes se diferencian una de otra con lo que nuestro análisis ganará validez al tratar redes con características diferentes. Además, del gráfico de variables, se vuelve a poner de manifiesto la fuerte correlación entre h, bn y bl que se ha visto antes.

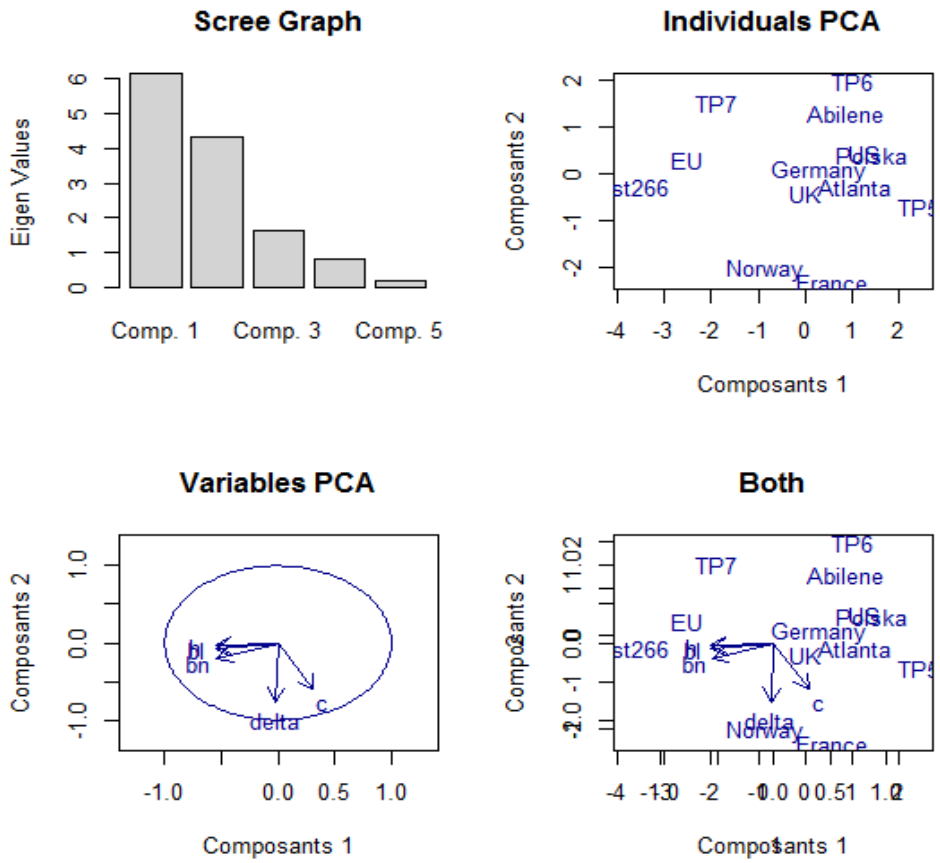


Fig. 3-3 Análisis de componentes principales

3.4 Resumen

En este capítulo hemos visto qué redes, configuraciones de capacidad, y casos hemos considerado para obtener los modelos, así como una clara descripción de la metodología de obtención de los valores de las variables de respuesta. Además, hemos realizado un estudio previo de las topologías y de sus variables explicativas, que nos permiten tener una información previa acerca de las correlaciones existentes entre sí. Esta información nos dará pié a mejorar la búsqueda de modelos de predicción de las variables de respuesta en el capítulo siguiente.

Capítulo 4.

Modelización

En este capítulo, vamos a aplicar una serie de técnicas estadísticas para la obtención de modelos que nos permitan predecir con cierta precisión nuestras variables respuesta. Es obvio decir que, tal y como hemos descrito anteriormente, tanto la topología como la capacidad tendrán efecto sobre dichas variables respuesta. Como validación de los modelos obtenidos, no sólo vamos a analizar la calidad de los residuos sino que vamos a probar la fiabilidad de los modelos en la predicción de redes de prueba que no se han tenido en cuenta para la obtención de los modelos.

4.1 Modelización de la intensidad

4.1.1 Efecto de la capacidad

Dado que la capacidad de la red (i.e. Δ y S) influye de manera muy clara sobre la intensidad que puede soportar ésta, primero estudiaremos dicha relación. Así, nos hacemos la siguiente pregunta: ¿qué relación existe entre las variables I, Δ y S? Con la intención de responder esta pregunta, la

Fig. 4-1 muestra el comportamiento de la intensidad para las redes del experimento respecto a las variables Δ y S. Los valores de esta variable respuesta para todas las configuraciones de capacidad pueden consultarse con más detalle en la tabla del apéndice C.

La primera impresión que se puede observar es que tanto Δ como S influyen en el valor de I. Las curvas que aparecen en las figuras (S=800,1600 y 3200) tienen una dependencia no lineal con respecto a Δ. No obstante, resulta fácil ver cómo se puede formular I mediante la búsqueda de un exponente óptimo (por transformación Box-Cox) y dos parámetros que corresponden al pendiente de la recta (τ) y la constante

(η), todos ellos dependientes de S. La ecuación 4.1 muestra la predicción exacta de I a partir de Δ y S.

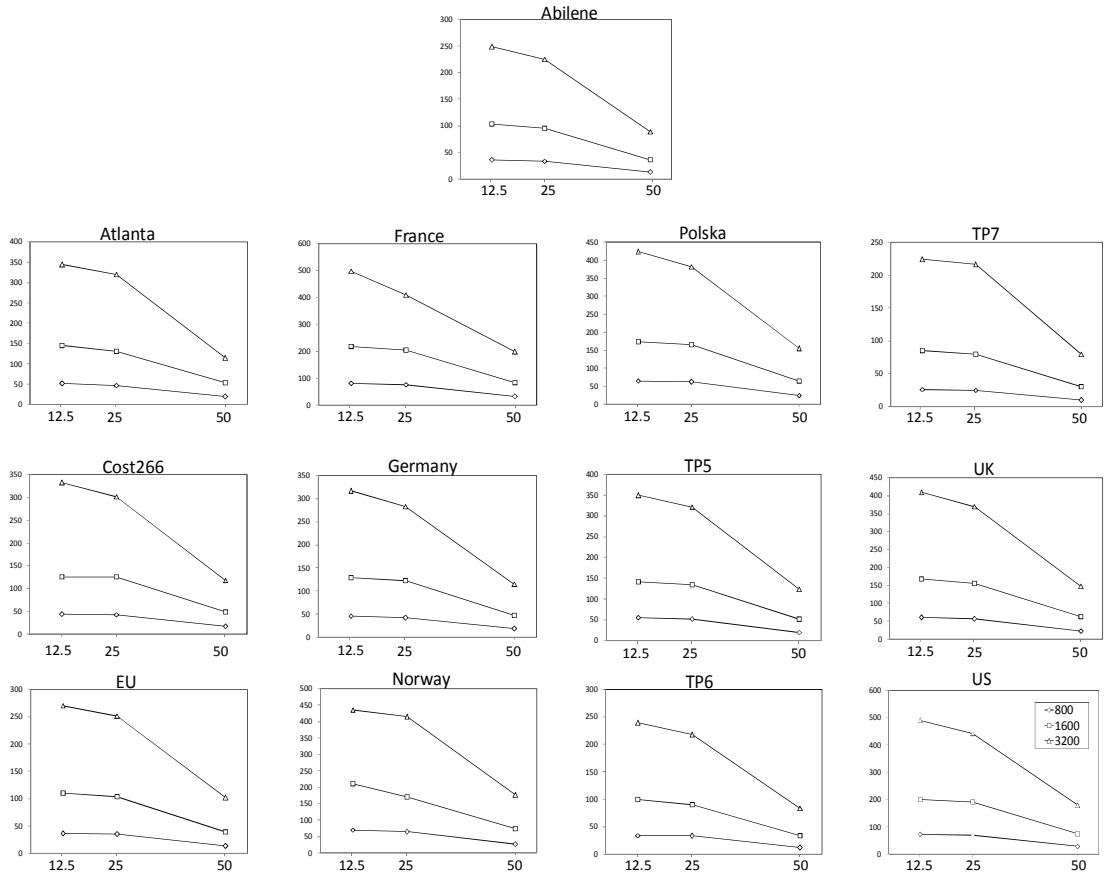


Fig. 4-1 Intensidad en función del tamaño de Δ y S

$$I = \tau(S) \cdot \Delta^{\gamma(S)} + \eta(S) \tag{4.1}$$

Lógicamente, un punto de fricción en esta ecuación es la búsqueda de los parámetros. Tras buscar la transformación óptima y obtener por regresión lineal la pendiente de la recta τ y la constante η, obtenemos el ajuste perfecto para dicha red (ajustamos 9 observaciones con una expresión que depende de 9 parámetros). La Tabla 4-1 muestra, como ejemplo, los valores de dichos parámetros para la red Abilene.

Tabla 4-1: Parámetros de ajuste exacto para la red Abilene

S	γ	τ	η
800	3.39	-4.04e-05	3.62e+01
1600	2.91	-7.87e-04	1.05e+02
3200	2.5	-9.34e-03	2.54e+02

Dado que la magnitud de las intensidades difiere significativamente entre redes distintas (por ejemplo, para S=3200 y Δ=12.5, las redes TP7 y France tienen unos valores de I igual a 224 y 499, respectivamente), planteamos la ecuación (4.2) para el cálculo de I donde todos los parámetros dependen de la topología y la capacidad. Con el propósito de ilustrar esa dependencia, entre paréntesis G indica la topología y S indica el tamaño espectral, como hemos definido anteriormente.

$$I = \tau(G,S) \cdot \Delta^{\gamma(G,S)} + \eta(G,S) \tag{4.2}$$

Este modelo pretende ilustrar la relación entre todas estas variables. Cabe decir que podríamos hacer una predicción exacta de todos los valores de I obtenidos a partir de la estimación por regresión de los parámetros γ, τ, y η de cada red. No obstante no obtendríamos ninguna reducción ni ninguna validez para la estimación de I para otras redes. Con el fin de reducir el número de coeficientes a estimar de nuestro modelo, vamos a tratar de simplificarlo de forma progresiva, eliminando el efecto de la topología en la mayor medida posible sin eliminar una cantidad significante de información. Empezamos por ver si podemos aplicar una transformación única para todas las redes (es decir, un único valor de γ para cada tamaño de espectro).

Para hallar la γ óptima para cada tamaño de espectro, se ha implementado una función (en MATLAB [Mat]) que realiza el siguiente proceso: para un valor de γ dado, se aplica dicha transformación a todas las redes y calcula el valor del coeficiente de determinación de Pearson (R^2). El valor de R^2 más bajo obtenido para una red se toma como bondad de ajuste de ese coeficiente. Se repite este procedimiento para un amplio rango de valores de γ (entre 1 y 9) y se selecciona aquel cuya bondad de ajuste sea la más alta (aquel que cuya red que peor ajusta sea el valor más alto). Tras aplicar este proceso, hemos obtenido que para unos valores de γ de 3.3, 2.4, y 2.2 para 800, 1600, y 3200 GHz respectivamente, el peor ajuste lineal para una red es del 99.46%, lo cual valida el uso de un único parámetro γ para todas las redes, pudiendo reducir el modelo de la ecuación (4.2) de la siguiente forma:

$$I = \tau(G,S) \cdot \Delta^{\gamma(S)} + \eta(G,S) \tag{4.3}$$

El siguiente paso es ver si se puede simplificar τ y η. No obstante, como ya venimos observando desde la obtención de los resultados, podemos concluir que tanto τ como η dependen claramente de la topología. Llegados a este punto, vamos a intentar separar el efecto de la topología del de la capacidad, mediante un modelo multiplicativo formado por dos factores: por una parte, el efecto de la capacidad (factor X) y por otro el de la topología (β). La ecuación (4.4) muestra este modelo:

$$I = \beta(G) \cdot X(S) = \beta(G) \cdot [\tau(S) \cdot \Delta^{\gamma(S)} + \eta(S)] \tag{4.4}$$

Para poder validar este modelo, hace falta ver primero si existen unos valores comunes de τ y η para todas las redes, que permitan estimar I a partir del factor de capacidad X y gracias a una pendiente específica β para cada topología. La manera

de obtener los valores de τ y η comunes ha sido mediante la regresión lineal de todos los valores de I con respecto de la variable Δ transformada (elevada al γ óptimo) Esto nos ilustra claramente el significado de X, que representa la intensidad media de todas las redes como función exclusivamente de la capacidad. La Tabla 4-2 muestra los valores óptimos de τ y η. Con el fin de mostrar la bondad de ajuste de dichos valores, hemos obtenido la recta de regresión de cada una de las redes con respecto de X, obteniendo el valor de la pendiente β y el valor de R^2. La Fig. 4-2 ilustra estas rectas para 3 de las redes utilizadas, que permiten ilustrar el buen ajuste lineal de I a partir de X. Con más detalle, la Tabla 4-3 muestra la correlación lineal entre X e I para cada red, así como la pendiente β que consigue una mejor estimación de I a partir de X. Se puede ver como esta correlación es igual a 99.06% en el peor de los casos, lo cual representa un ajuste muy preciso.

Tabla 4-2: Parámetros de ajuste de X

S	γ	τ	η
800	3.3	-8.08e-05	5.25e+01
1600	2.4	-8.23e-03	1.52e+02
3200	2.2	-4.33e-02	3.67e+02

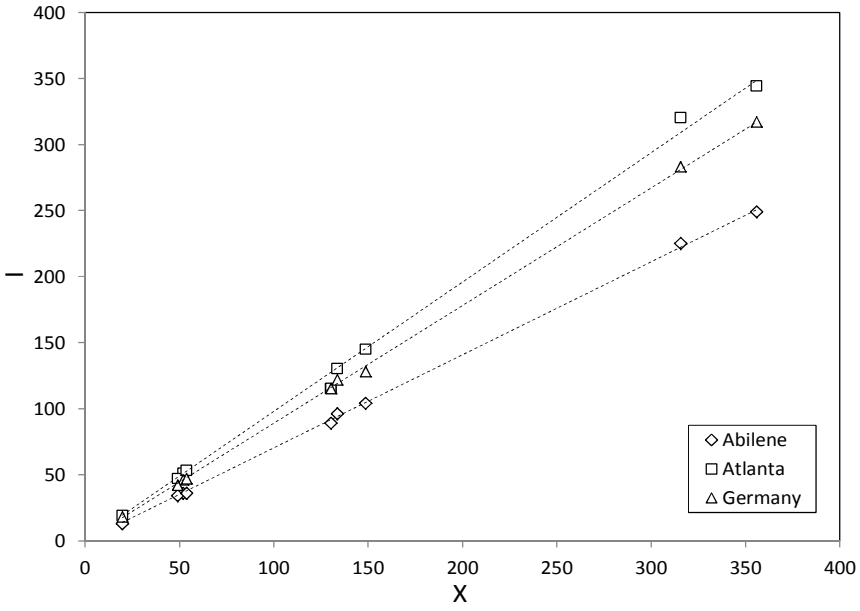


Fig. 4-2 Ajuste lineal de I a partir de X

De la observación de la Fig. 4-2 y de la Tabla 4-3 se ratifica la percepción de que cada topología tiene un efecto diferente sobre el efecto que la capacidad tiene sobre

la intensidad admitida por la red. Por ello, en la siguiente sección se buscará un modelo para estimar el parámetro β a partir de las características de la topología.

Tabla 4-3: Valor beta de las redes de estudio

Red	β	correlación
Abilene	0.7035	0.9995
Atlanta	0.9786	0.9973
Cost266	0.9321	0.9974
EU	0.7728	0.9979
France	1.3903	0.9911
Germany	0.8904	0.9996
Norway	1.2838	0.9944
Polska	1.1988	0.9994
TP5	0.9945	0.9984
TP6	0.6752	0.9990
TP7	0.6371	0.9906
UK	1.1562	0.9997
US	1.3867	0.9996

4.1.2 Efecto de la topología

Antes de iniciar la modelización de β, nos imponemos la restricción de utilizar pocas variables explicativas de la topología para su predicción, ya que solo contamos con 13 observaciones. Por ese motivo, empezamos con modelos lineales simples que incluyan una única variable explicativa [Do02].

La Tabla 4-4 muestra un resumen del ajuste de cada modelo. Como se puede observar, no hay ningún modelo que dé un R^2 aceptable. No obstante, hemos de decir que estos modelos los hemos obtenido para tener una primera idea sobre qué variables influyen más a la hora de estimar β. Los modelos con un R^2 más alto son los que incluyen δ y c. Además, en ambos casos las variables son significativas para un nivel de confianza del 5%. La variable κ parece que también aporta información útil, así que se tendrá en cuenta su peso en modelos con más variables.

*Tabla 4-4: Bondad de ajuste de β = b1*variable + b0*

Variable	b1	p-valor	R^2
N	0.0034	0.7275	0.0115
E	0.0056	0.2959	0.0987
δ	0.5246	0.0007	0.6621
H	-0.1347	0.3148	0.0916
rad	-0.0396	0.7083	0.0132
dia	-0.0632	0.1798	0.1572
c	0.7848	0.0413	0.3266
sc	0.7857	0.1295	0.1963
bl	-0.0053	0.4507	0.0527
sbl	-0.0020	0.8581	0.0030
bn	-0.0019	0.7562	0.0091
sbn	0.0014	0.8144	0.0052
κ	0.7470	0.0625	0.2809

Yendo un paso más allá, probamos modelos lineales con 2 variables. Mediante el procedimiento "step" implementado en R [R], se ha hallado como mejor modelo el que contiene las variables δ y κ. A continuación se muestra la salida de R de dicho modelo:

```
Call:
lm(formula = Betaprima ~ delta + kappa)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.15498 -0.05628 -0.02695  0.07521  0.15583 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.76675    0.20581  -3.726 0.003939 ** 
delta        0.50533    0.06689   7.554 1.94e-05 ***
kappa        0.67834    0.14625   4.638 0.000925 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09492 on 10 degrees of freedom
Multiple R-squared: 0.8928,     Adjusted R-squared: 0.8713 
F-statistic: 41.63 on 2 and 10 DF,  p-value: 1.418e-05
```

A pesar de que todos los coeficientes son claramente significativos y que el valor de R^2 es bastante bueno, seguimos con la exploración de nuevos modelos ya que la precisión de éste no es suficiente para una estimación correcta de ß. Con el fin de no aumentar el número de variables para no incurrir en una sobreestimación de la variable respuesta, aplicamos la transformación logarítmica a todas las variables [Ch97] y volvemos a repetir la obtención de un modelo lineal con 2 variables, ahora con las transformadas. Llegados a este punto, decidimos probar el modelo incluyendo las variables |E| y h, ya que en modelos de estimación similares al nuestro pero para tecnologías ópticas diferentes, aparecen ambas como variables significativas en un modelo con variables logarítmicas [Ve11]. Sin entrar en detalles, los autores en [Ve11] muestran modelos para predicción de la intensidad en redes ópticas rígidas, sin contemplar ni anchos de banda diferente ni configuraciones de capacidad distintas. Por ese motivo, creemos que nuestra componente ß puede tener relación con la intensidad que ellos modelan. Los resultados de este nuevo modelo son los siguientes:

```
Call:
lm(formula = log(Betaprima) ~ log(E) + log(h))

Residuals:
     Min       1Q   Median       3Q      Max
-0.16117 -0.06252 -0.03879  0.06893  0.14012

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.80545    0.18992  -4.241  0.00171 **
log(E)       0.74408    0.08838   8.419 7.51e-06 ***
log(h)      -1.68377    0.20604  -8.172 9.77e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1 on 10 degrees of freedom
Multiple R-squared: 0.8868,     Adjusted R-squared: 0.8641
F-statistic: 39.16 on 2 and 10 DF,  p-value: 1.861e-05
```

Aunque el ajuste no es malo, cabe decir que es peor que el obtenido sin variables logarítmicas. Además, es peor que el ajuste de aquellos modelos que aparecen en [Ve11]. Por ello, debemos descartar el uso de estas variables para nuestro modelo. No obstante, el hecho de usar un modelo con una variable relacionada con el tamaño de la red (|E|) y otra que muestra el grado de conectividad de ésta (h), tiene mucho sentido para tratar de estimar la intensidad que puede admitir una red. Por eso, decidimos probar un modelo que contenga |E| y otra variable relacionada con la conectividad de la red (δ, dia, rad, cluster, bn, bl, y κ) y ver que resultados nos da. De todos ellos, el mejor que hemos obtenido es el que contiene las variables |E| y κ, cuyos resultados se muestran a continuación:

```
Call:
lm(formula = log(Betaprima) ~ log(E) + log(kappa))

Residuals:
      Min        1Q    Median        3Q       Max
-0.056140 -0.014306 -0.003682  0.025508  0.058029

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.17345    0.07502  -15.64 2.34e-08 ***
log(E)       0.55515    0.02685   20.67 1.55e-09 ***
log(kappa)   0.61501    0.02686   22.90 5.69e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03792 on 10 degrees of freedom
Multiple R-squared: 0.9837,     Adjusted R-squared: 0.9805
F-statistic: 302.3 on 2 and 10 DF,  p-value: 1.14e-09
```

A la vista de los resultados, podemos concluir que el ajuste es muy bueno (R^2 mayor del 98%, todos los p-valores por debajo de 1e-08), con lo que decidimos dejar de buscar nuevos modelos y pasar a un análisis exhaustivo de los residuos para β. La Fig. 4-3 muestra el análisis de residuos, cuyo propósito es el de validar las hipótesis de normalidad, linealidad y varianza constante. Además, también sirve para ver si hay observaciones atípicas y proponer posibles mejoras en el modelo.

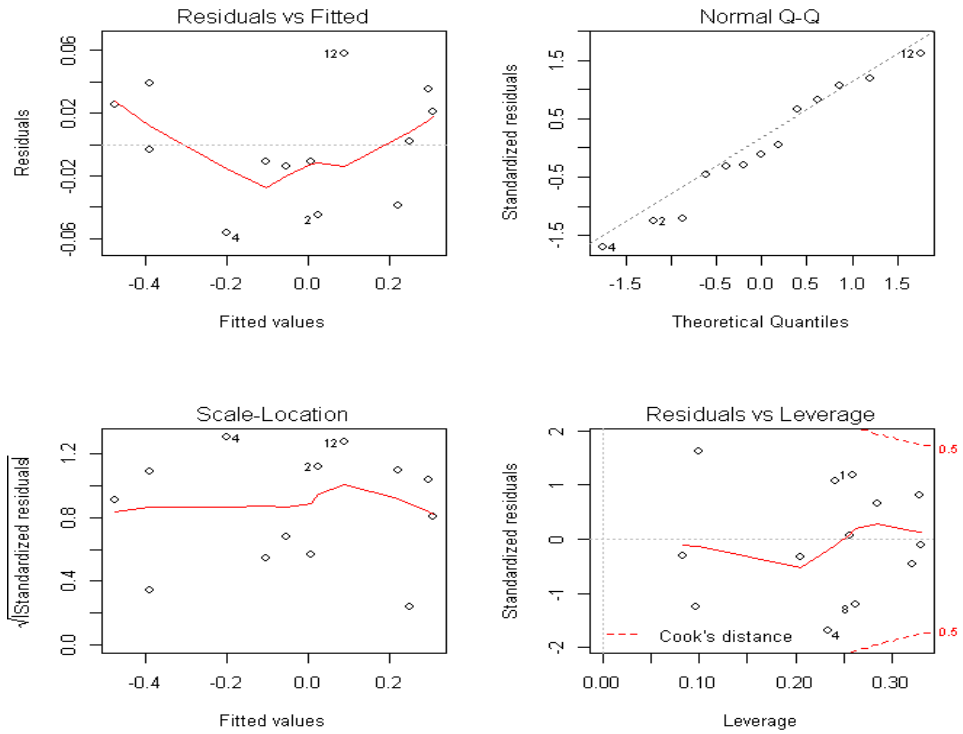


Fig. 4-3 Análisis de los residuos del modelo

Por lo que se ve en, nada nos hace indicar que no se cumpla alguna de las hipótesis puesto que los residuos se ajustan bastante bien a la línea de normalidad, se encuentran dispersos y sin seguir ningún tipo de patrón específico y, finalmente, no se observa ninguna observación atípica.

Antes de concluir con esta sección, probamos buscar un modelo que añada una tercera variable con el fin de evaluar la necesidad de ampliar el modelo. Los resultados muestran una extensión del modelo anterior al que se ha añadido el diámetro de la red.

```
Call:
lm(formula = log(Betaprima) ~ log(E) + log(kappa) + log(dia))

Residuals:
      Min        1Q    Median        3Q       Max
-0.055240 -0.029246  0.002541  0.029249  0.049542

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.10038    0.11148  -9.871 3.99e-06 ***
log(E)       0.56504    0.02930  19.286 1.25e-08 ***
log(kappa)   0.56996    0.05724   9.958 3.71e-06 ***
log(dia)    -0.09163    0.10251  -0.894    0.395
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03831 on 9 degrees of freedom
Multiple R-squared: 0.9851,     Adjusted R-squared: 0.9801
F-statistic: 197.8 on 3 and 9 DF,  p-value: 1.567e-08
```

La conclusión que podemos sacar es clara: no vale la pena añadir otra variable puesto que no se aprecia ninguna mejora significativa. Reforzando esta conclusión, la Tabla 4-5 muestra un resumen de la bondad de ajuste de los modelos, añadiendo el AIC como factor de decisión definitivo. Como se puede observar, el modelo escogido (en negrita), tiene un R^2 ligeramente inferior al modelo consistente en 3 variables pero su AIC es inferior al contar con menos variables.

Tabla 4-5: Resumen de la bondad de ajuste de los modelos obtenidos

Modelo	DF	R^2	AIC
$\beta \sim \delta$	11	0.6621	-6.8205
$\beta \sim \delta + \kappa$	10	0.8928	-19.7419
$\ln(\beta) \sim \ln(\lvert E \rvert) + \ln(h)$	10	0.8868	-18.3762
$\ln(\beta) \sim \ln(\lvert E \rvert) + \ln(\kappa)$	**10**	**0.9837**	**-43.5972**
$\ln(\beta) \sim \ln(\lvert E \rvert) + \ln(\kappa) + \ln(dia)$	9	0.9851	-42.7029

Con estos datos, vemos que el mejor modelo posible es $\ln(\beta) \sim \ln(E) + \ln(\kappa)$. Puesto que con anterioridad se ha validado el modelo con el análisis de residuos, éste es el modelo escogido para predecir β.

4.1.3 Validación conjunta de residuos

En los apartados anteriores hemos construido paso por paso un modelo para la estimación de la intensidad formado por dos componentes relacionadas mediante un producto. La expresión de este modelo se muestra completa en la ecuación (4.5), donde ε constituye el error del modelo. La Tabla 4-6 muestra todos los coeficientes de este modelo:

$$I = \beta(G) \cdot X(S) = \left(e^{b0} \cdot E^{b1} \cdot \kappa^{b2}\right) \cdot \left(\tau(S) \cdot \Delta^{\gamma(S)} + \eta(S)\right) + \varepsilon \tag{4.5}$$

Tabla 4-6: Parámetros del modelo de intensidad

S	γ	τ	η	b0	b1	b2
800	3.3	-8.08e-05	5.25e+01	-1.17345	0.55515	0.61501
1600	2.4	-8.23e-03	1.52e+02			
3200	2.2	-4.33e-02	3.67e+02			

A continuación, se procederá a hacer un análisis de los residuos del modelo completo de la ecuación (4.5). La Fig. 4-4 muestra los residuos estandarizados de cada una de las observaciones. Aunque se observa que hay algunos residuos que toman valores peligrosamente altos cercanos a 3 o incluso a 4 (en valor absoluto), vemos que la gran mayoría (cerca del 95%) están dentro del intervalo de confianza [-2,2]. Además, no se observa que los residuos sigan algún tipo de patrón específico.

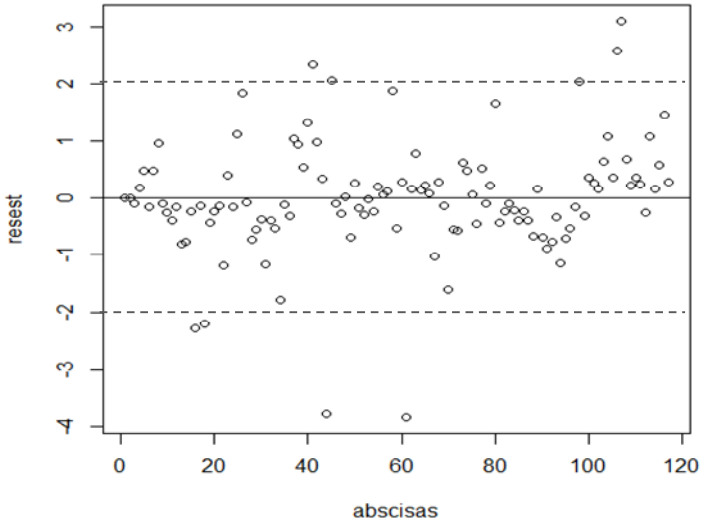


Fig. 4-4 Residuos estandarizados para todas las observaciones

El siguiente paso es comprobar la normalidad de los residuos, como se muestra en la Fig. 4-5. Aquí se observa claramente que los residuos de todas las intensidades no siguen una distribución normal. De hecho, se puede apreciar el fenómeno de las

colas pesadas que penaliza mucho la normalidad de los residuos. No obstante, el hecho de considerar intensidades que varían en un amplio rango de valores puede provocar un efecto como el descrito en la figura. En nuestro caso, la amplitud de ese rango se debe mayoritariamente al uso de configuraciones de capacidad muy distintas. Es por eso que tiene sentido hacer el análisis de residuos separando las predicciones por configuraciones de capacidad.

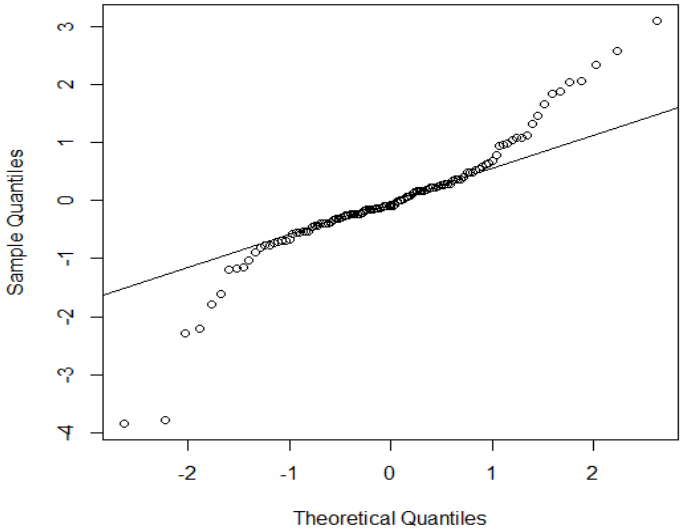


Fig. 4-5 Normalidad de los residuos

Las Fig. 4-6 y Fig. 4-7 muestran el análisis de residuos estandarizados y de normalidad respectivamente para cada una de las configuraciones de capacidad por separado. Aquí se observa que, en la mayoría de los casos, los residuos se sitúan en el intervalo [-2,2] y no se aprecia ninguna forma extraña que invalide la hipótesis de linealidad. Además, aunque hay algunos casos en que la normalidad es más evidente que en otros, se puede aceptar la hipótesis de normalidad sin pérdida de generalidad.

Finalmente, se ha calculado los promedios de todos los errores (en %) en función de Δ, siendo 4'91, 4'49 y 6'05 para los tamaños 12,5 25 y 50, respectivamente. Para los tamaños de espectro son 7'51, 4'69 y 3'27 para 800, 1600 y 3200. No se observan grandes diferencias aunque, en el caso del tamaño de espectro hay un incremento del error para un tamaño de 800. En cualquier caso, el error medio para todos los casos es cercano al 5%, lo que nos permite definitivamente considerar que nuestro modelo efectúa un error bajo en la predicción.

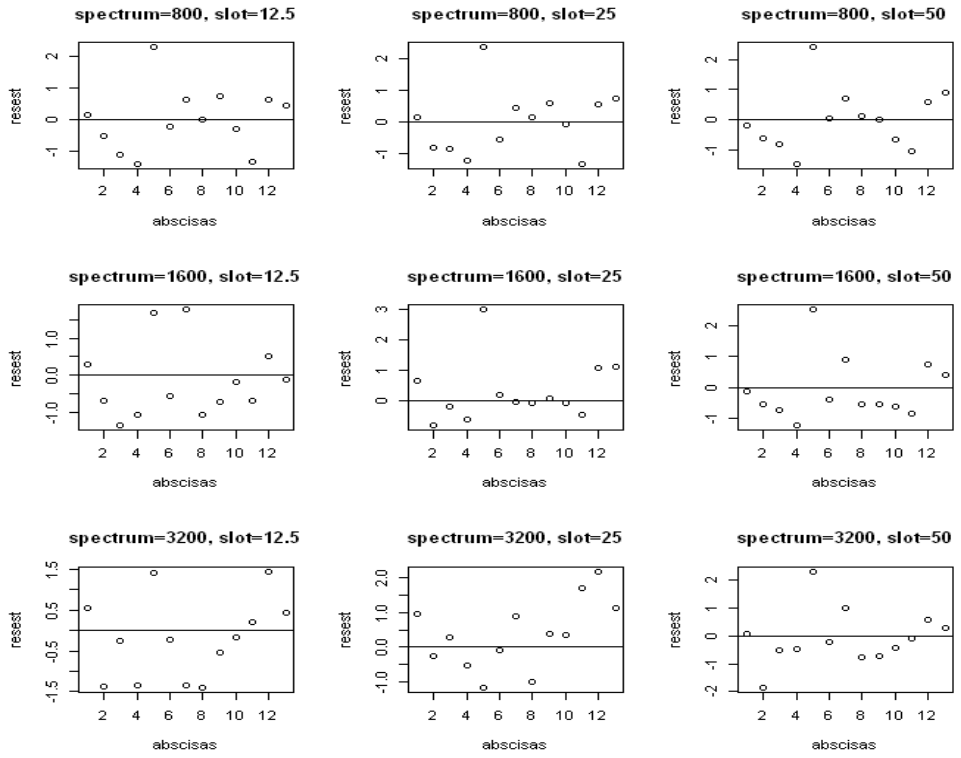


Fig. 4-6 Residuos estandarizados en función del tamaño de espectro y Δ

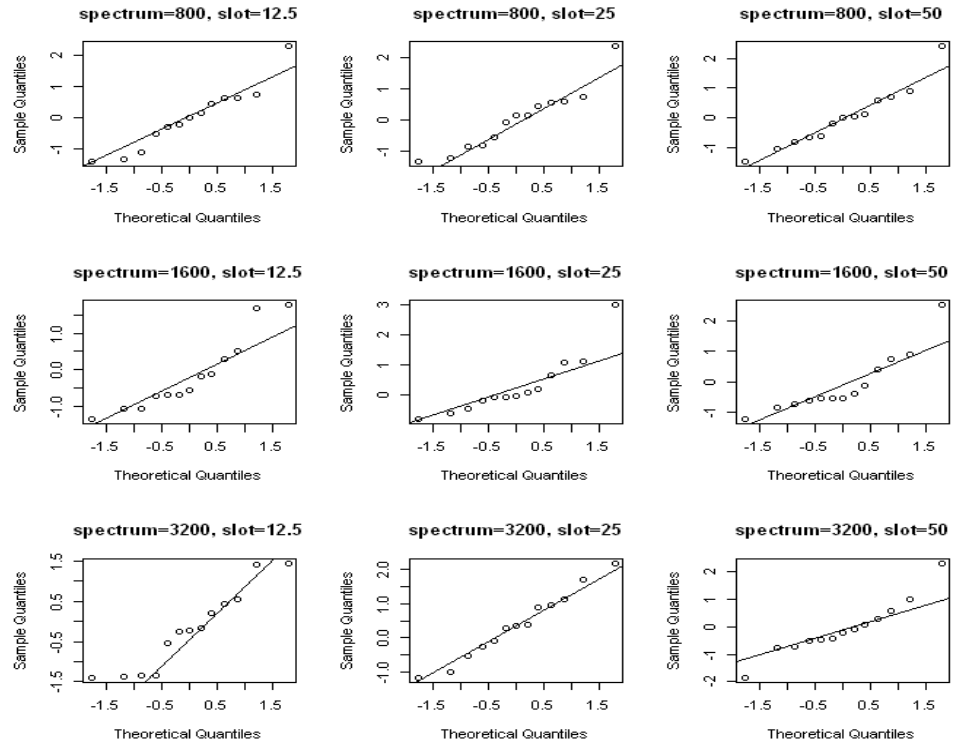


Fig. 4-7 Normalidad de los residuos en función del tamaño de espectro y Δ

4.2 Modelización del perfil de tráfico

Para la predicción del vector de perfil de tráfico (TP), decidimos usar como referencia el valor de ancho de banda medio ya que, como introdujimos en el capítulo 1, cada perfil de tráfico está relacionado unívocamente con un ancho de banda medio. Para ello, será necesario establecer sus dependencias con Δ, S y/o la topología y ver si éstas influyen sobre dicho TP. Para ello, resultará útil hacer un breve análisis exploratorio, como el mostrado en la Fig. 4-8, donde se representan todos los puntos obtenidos respecto a Δ. Se observa que, en principio, Δ tiene una fuerte influencia en el valor del TP. Por el contrario, para una cierta Δ, no se observan grandes diferencias en los valores del TP, hecho que nos hace sospechar que ni el tamaño del espectro ni la topología tendrán influencia en el valor del TP.

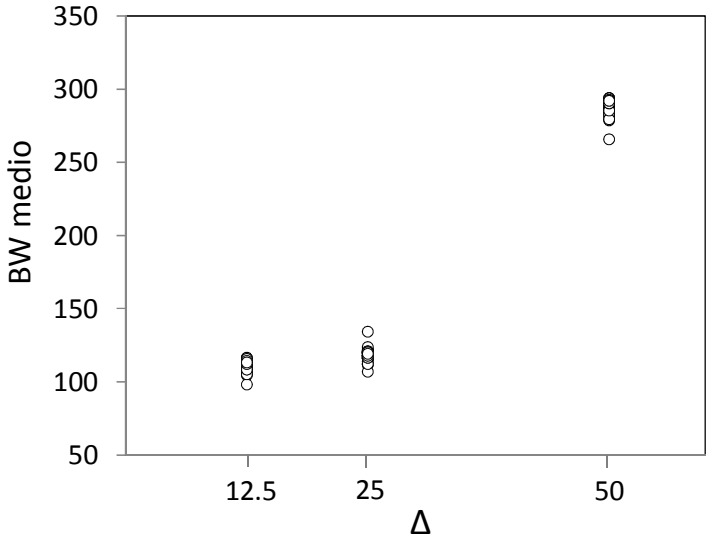


Fig. 4-8 BW medio de los TP's en función del tamaño de Δ

Con tal de demostrar las afirmaciones anteriores, realizamos constrastes de hipótesis para comparar medias y ver así si nuestras intuiciones se cumplen. Los resultados de estos contrastes se ilustran a continuación. En la Fig. 4-9, se muestra el contraste con respecto al tamaño de Δ. Se observa claramente que Δ influye en el valor del perfil de tráfico puesto que, al pasar de un tamaño de Δ a otro, el intervalo (para una α=0.05) que incluye el valor medio de la diferencia de tamaño en ningún caso incluye el 0 así que se determina que la diferencia de medias es significativa en todos los casos.

Conviene comparar este test con la Fig. 4-8 donde se observaba que, a medida que aumentaba Δ, aumentaba el TP. No obstante, si parecía muy claro que la diferencia entre $\Delta_{25\text{-}50}$ era significativa, no lo parecía tanto que la diferencia entre $\Delta_{12.5\text{-}25}$ fuese también significativa. Lógicamente, si la diferencia entre $\Delta_{25\text{-}50}$ es significativa, la diferencia entre $\Delta_{12.5\text{-}50}$ también lo será.

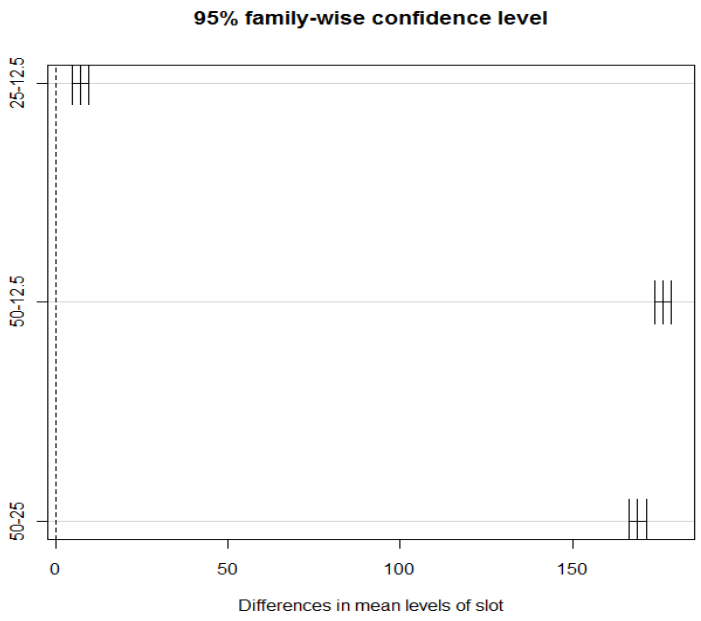


Fig. 4-9 Test de Tukey para comparar medias de los TP's en función de Δ

Una vez hemos demostrado que Δ es significativo, comprobaremos si S y la topología lo son. Para ello, se manipularán los datos de forma que se creará, para cada tamaño de Δ (es decir, se fijará Δ) una matriz donde aparecerá, para cada topología, su tamaño de espectro junto con su valor de TP correspondiente. De esa manera, se podrán hacer comparaciones de medias teniendo en cuenta la topología y el tamaño de espectro. No obstante, un gráfico donde para cada tamaño de Δ se mostrasen todas las posibles comparaciones entre topologías resultaría muy confuso y difícil de interpretar. Para sintetizar, hemos hecho todos esos contrastes y hemos constatado que todos los p-valores de las comparaciones eran superiores a 0.05 por lo que se concluye que no hay diferencias entre redes y, por lo tanto, que la topología no es significativa en el valor del TP.

Repetimos el proceso anterior pero para la variable S. En este caso sí que es más clara la interpretación, por lo que se ilustra estos contrastes en la Fig. 4-10. Tal como se puede observar, el tamaño de espectro no influye en el valor del TP al contener todos los intervalos de confianza el valor 0.

Por lo tanto, se concluye que el vector TP sólo depende de Δ. Así, el modelo consiste únicamente en la media de todos los valores por cada valor de Δ, tal y como se muestra en la Tabla 4-7. En este punto cabe recordar que la obtención del TP a partir del ancho de banda medio es directa ya que existe una relación unívoca entre el vector y su valor promedio.

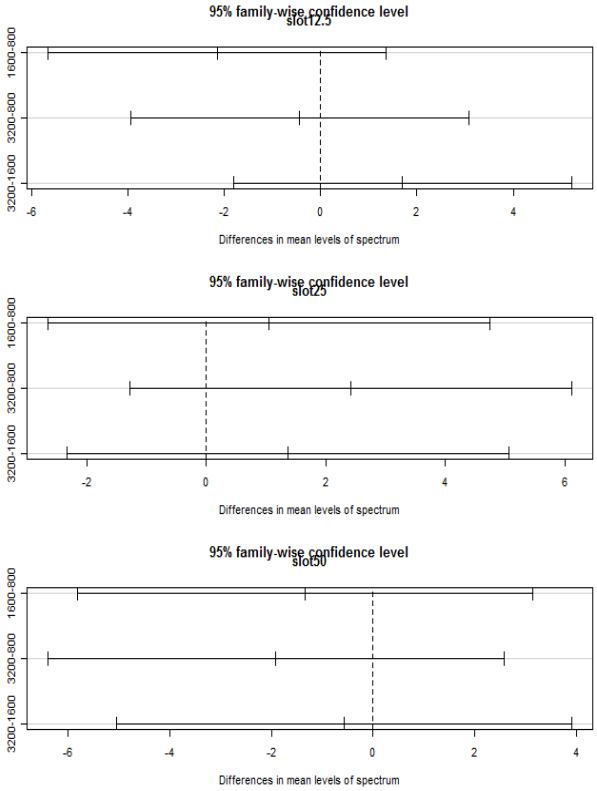


Fig. 4-10 Test de Tukey para comparar medias en función de S

Tabla 4-7: TP's en función de Δ

Δ (GHz)	**10 Gbps**	**40 Gbps**	**100 Gbps**	**400 Gbps**	**BW medio (Gbps)**
12.5	0.1	0.16	0.64	0.1	111.27
25	0.1	0.1	0.69	0.11	118.46
50	0.1	0.1	0.125	0.675	287.32

4.3 Validación mediante las redes de prueba

Como validación final y para ver la utilidad del modelo para la estimación del tráfico en redes diferentes a las utilizadas en el estudio, se han utilizado las topologías EONCT, TID y JANOS (descritas anteriormente) con unos valores de |E| de 23, 35 y 55 y de κ de 0.39, 0.43 y 0.15, respectivamente. En la Fig. 4-11, se muestra un diagrama bivariante entre las variables explicativas del modelo (|E| y κ) de las redes utilizadas en el estudio y, con distintas figuras, aparecen las redes utilizadas en la validación. Este diagrama resultará muy útil a la hora de interpretar los resultados de estas redes.

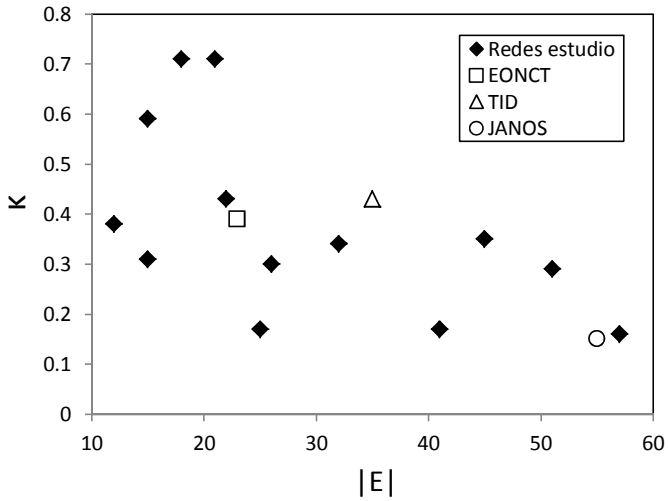


Fig. 4-11 Diagrama bivariante entre |E| y κ

La validación de los modelos se ha realizado comparando estimaciones con resultados de simulación. La manera de proceder en la obtención de los datos de simulación ha sido la misma que la utilizada en las redes del estudio: para cada configuración de capacidad (S, Δ), se han obtenido los valores de I y TP que maximizan el ancho de banda. Así, para cada red, hemos obtenido nueve estimaciones y nueve predicciones (una por cada configuración de capacidad). Siendo I_c^{est} la intensidad estimada para la configuración de capacidad *c* y I_c^{sim} la obtenida por simulación, el error ε_c se define como muestra la ecuación (4.6). Para el TP, hemos comparado el ancho de banda medio, obteniendo los valores estimados (BW_c^{est}) y simulados (BW_c^{sim}), y comparándolos con una expresión similar a la (4.6).

$$\varepsilon_c = \frac{\left| I_c^{est} - I_c^{sim} \right|}{I_c^{sim}} \cdot 100 \tag{4.6}$$

La Tabla 4-8 muestra un resumen del error cometido por cada una de las redes en la estimación de I, resumido con el error medio, error al intervalo del 95%, y error máximo de las distintas configuraciones de capacidad. Del mismo modo, la Tabla 4-9 muestra los mismos errores para la estimación del TP.

Tabla 4-8: Error de estimación de I para cada red

	Medio	**95%**	**Max**
EONCT	1.95%	4.89%	6.25%
TID	4.91%	8.84%	9.33%
JANOS	10.48%	17.11%	19.70%

Tabla 4-9: Error de estimación de TP para cada red

	Medio	95%	Max
EONCT	4.46%	9.99%	11.28%
TID	3.86%	8.43%	8.86%
JANOS	2.57%	6.33%	6.94%

Para la red EONCT, el ajuste de la intensidad resulta muy bueno ya que el error medio cometido es del 1.95%, siendo el error para el 95% de los casos inferior al 5%. La razón de estos buenos resultados viene dada por que esta red, tal como indica la Fig. 4-11 tiene unos valores de |E| y κ similares a los de las redes utilizadas en el estudio. Es decir, para valores similares de |E| a los de la red EONCT se han usado varias redes con un amplio rango de κ.

Para la red TID, el error medio de la intensidad es del 4.91% con un error máximo inferior al 10%. Es remarcable decir que, a pesar de que esta red tiene unos valores de E y κ ligeramente distintos de los de las redes de estudio (conectividad algo mayor que las redes de estudio con mismo número de enlaces), el ajuste continúa siendo muy bueno.

Los resultados de validación de la intensidad para la red JANOS no son tan buenos ya que el error medio es del 10.48% con el límite del intervalo del 95% fijado al valor de 17.11% de error. No obstante, hay que tener en cuenta que la red JANOS tiene una dimensión próxima al límite de nuestro estudio, hecho que la convierten en un valor extremo en comparación con las otras redes de estudio. Eso indica que el modelo es muy sensible a redes que se aproximan, por lo menos, a los tamaños más grandes estudiados. En cualquier caso, el error medio de todas las comparativas juntando todas las redes y configuraciones de capacidad es del 5.78%, lo que podemos considerar como un buen ajuste teniendo en cuenta la gran variedad de casos probados.

Finalmente, respecto a los modelos del TP, vemos como el error parece disminuir con el tamaño de la red. Como resumen, podemos decir que cualquier red tiene un error medio inferior al 5%, con un intervalo del 95% marcado en un valor inferior al 10% de error. Este error supone una bondad de ajuste mayor que la ofrecida por el modelo de intensidad, con lo que podemos concluir que ambos modelos dan una predicción válida para la estimación del ancho de banda total de la red.

4.4 Conclusión

En este capítulo hemos encontrado modelos capaces de estimar con precisión el ancho de banda máximo servido a partir de modelar I y TP. Hemos visto como el modelo de estimación de I se puede separar en dos partes, una dependiente de la

capacidad instalada en la red y otra dependiente de la topología. En esta segunda, hemos obtenido un modelo lineal utilizando variables logarítmicas que utiliza la conectividad algebraica y el número de enlaces como variables significativas. Por lo que respecta al TP, hemos visto que la única variable significativa es Δ, siendo S y la topología no influyentes para su predicción. Hemos reforzado el análisis de la calidad del modelo a partir de estimaciones sobre redes de prueba, obteniendo resultados que se han validado por simulación.

Para demostrar la utilidad de estos modelos, en el siguiente capítulo se muestran dos casos de aplicación para el óptimo diseño y dimensionado de redes ópticas elásticas.

Capítulo 5.

Casos de aplicación

En el capítulo de introducción se hacía referencia al uso de los modelos de estimación de tráfico para el diseño de redes ópticas elásticas. En este capítulo se muestran dos casos de aplicación de los modelos. El primero está orientado al dimensionamiento de redes ópticas, del cual se presenta un detallado caso de estudio con valores obtenidos con nuestros modelos. En segundo lugar, se introduce un problema de diseño de topología, del cual definimos el problema e ilustramos como el uso de los modelos de estimación de tráfico permitiría encontrar la solución óptima.

5.1 Dimensionamiento de redes ópticas

5.1.1 Definición del problema

Este problema se define como un problema de equipamiento de la red. Se trata de encontrar la configuración de capacidad óptima según un cierto criterio de valoración, teniendo en cuenta una serie de restricciones. Bajo esta amplia definición, existen múltiples versiones, aunque todas ellas comparten características comunes, como por ejemplo que la topología de la red viene dada y no se modifica durante la resolución del problema. A continuación se presenta un caso práctico que podría ser de interés para un operador de red.

5.1.2 Caso de estudio

Supongamos que tenemos un operador de red, propietario de la infraestructura de fibra óptica, que decide estudiar la migración a la tecnología característica de las redes ópticas elásticas. Concretamente, se cuestiona como debe equipar la red para obtener el máximo beneficio cumpliendo con sus necesidades y objetivos. Sin entrar

en el análisis de costes, vamos a hacer un análisis de ingresos previstos haciendo uso de los modelos de predicción de intensidad y perfil de tráfico.

La empresa quiere hacer un cálculo de ingresos previstos durante los 5 años posteriores a la implantación de la tecnología, basándose en una previsión de ancho de banda demandado por sus clientes. La Tabla 5-1 muestra unos valores de previsión de ancho de banda de sus clientes, que tomaremos como el máximo tráfico que podrá servir y, por lo tanto, facturar. Destacamos dos clases de tráfico diferentes: una clase destinada a negocios y empresas, caracterizada por conexiones con mayor ancho de banda, y otra dedicada a proveer servicios de internet a usuarios domésticos, con menor ancho de banda por conexión.

Tabla 5-1: Previsiones de la demanda de ancho de banda

Tipo	**Año 1**	**Año 2**	**Año 3**	**Año 4**	**Año 5**
Negocios	3000	4200	5880	8232	11525
Internet	10000	12000	14400	17280	20736

La compañía, propietaria de la red EONCT mostrada en las redes de prueba de los modelos, tiene un ancho espectral total por equipar de 3200 GHz (estos constituyen los *recursos* de la empresa). Como modelo empresarial propio, toma la decisión de equipar 800 GHz para tráfico de negocios (que equipará con Δ=50GHz, por ajustarse más eficientemente a la clase de trafico de negocios) y 1600 GHz para tráfico de Internet (cuya máxima eficiencia se da con Δ=12.5GHz), mientras que se cuestiona qué hacer con los 800 GHz restantes (que llamaremos *recursos variables*, en contraste con los anteriores que son *recursos reservados*). Esta será la principal incógnita a resolver en este estudio, incógnita que solo puede tomar 4 soluciones diferentes:

1) Equipar el espectro con Δ=50GHz y servir tráfico de negocios de los propios clientes.
2) Equipar el espectro con Δ=50GHz y alquilar la infraestructura a otra compañía.
3) Equipar el espectro con Δ=12.5GHz y servir tráfico de Internet de los propios clientes.
4) Equipar el espectro con Δ=12.5GHz y alquilar la infraestructura a otra compañía.

Además de responder a esta pregunta, la empresa quiere tener una estimación de los ingresos previstos para poder comparar con la inversión inicial que supone la migración a la nueva tecnología y, así, saber cuál es el valor neto de la operación. La Tabla 5-2 muestra cuales son los costes que hemos asumido para poder pasar unidades de ancho de banda a un cierto valor económico.

Tabla 5-2: Coste por unidad de ancho de banda de cada servicio

Tipo	Unidades de Coste (u.c.)
Negocios	2
Internet	1
Alquiler	0.5

Como hemos introducido anteriormente, los valores de ancho de banda transportado (necesarios para el cálculo de los ingresos previstos y la toma de decisiones) provienen de los modelos de I y TP estimados. En este caso, al tener una topología fija y variar solo las configuraciones de capacidad, la pendiente β del modelo de cálculo de I permanece constante, mientras que los cambios entre casos se dan en la componente de capacidad X. A continuación se muestran los resultados de estos cálculos.

En primer lugar, calcularemos los ingresos previstos de la parte que ya sabemos que uso va a tener. Así, la Tabla 5-3 muestra el ancho de banda facturable (servido) y los ingresos previstos de cada año para el tráfico de negocios, teniendo en cuenta las previsiones. Además, se muestra el rendimiento de cada año, definido como la cantidad de tráfico servido entre la cantidad máxima de tráfico que se podría servir con el equipamiento instalado (capacidad). Alternativamente, se muestra la cantidad de tráfico que no se puede admitir fruto de tener menos capacidad que demanda. Como se puede observar, hasta el año 3 no se llega a saturar la capacidad reservada para este tipo de tráfico. De hecho, el último año solo podemos servir la mitad de todo el volumen estimado, con lo que podríamos plantear el uso de los recursos variables para este tipo de tráfico. No obstante, esta extensión no tendría sentido durante los primeros años, ya que no llegan a explotarse todos los recursos reservados para tráfico de negocios.

Tabla 5-3: Ingresos previstos (en u.c.) por el tráfico de negocios previsto (en Gbps)

	Año 1	Año 2	Año 3	Año 4	Año 5
Demanda (Gbps)	3000	4200	5880	8232	11525
Capacidad (Gbps)	5669	5669	5669	5669	5669
Servido (Gbps)	3000	4200	5669	5669	5669
Rendimiento (%)	53	74	100	100	100
No admitido (Gbps)	0	0	211	2563	5856
Ingresos (u.c.)	6000	8400	11337	11337	11337

Igual que la Tabla 5-3, la Tabla 5-4 muestra el análisis para el tráfico de Internet. Las conclusiones son similares, aunque aquí la saturación de recursos reservados se da más tarde (a partir del 4 año).

Tabla 5-4: Ingresos previstos (en u.c.) por el tráfico de Internet previsto (en Gbps)

	Año 1	Año 2	Año 3	Año 4	Año 5
Demanda (Gbps)	10000	12000	14400	17280	20736
Capacidad (Gbps)	16300	16300	16300	16300	16300
Servido (Gbps)	10000	12000	14400	16300	16300
Rendimiento (%)	61	74	88	100	100
No admitido (Gbps)	0	0	0	980	4436
Ingresos (u.c.)	10000	12000	14400	16300	16300

Con estos valores, los ingresos estimados para recursos reservados en los 5 años son de 48412 u.c. para tráfico de negocios y de 69001u.c. para tráfico de Internet, dando un total de 117412 u.c.. A continuación, vamos a evaluar como equipar y que uso dar a los recursos variables. Tomando la decisión de equipar los recursos variables para servir tráfico de negocios, la Tabla 5-5 muestra cual sería la explotación anual. Dado que los 2 primeros años, nuestra demanda no supera la cantidad que podemos servir con los recursos reservados (como se constató en la Tabla 5-3), decidimos alquilar la infraestructura a otra compañía. Los ingresos que recibiremos por este alquiler de la infraestructura se computan a partir del ancho de banda máximo que se puede llegar a servir en esos recursos, pero a un coste inferior del que facturaría nuestra empresa si se encargara ella de explotar estos recursos. A partir del 3 año tenemos más demanda de nuestros clientes de la que podemos servir con los recursos reservados, pero no decidimos explotarla nosotros ya que nos sale más a cuenta alquilar los recursos a un tercero. A partir del año 4, explotamos nosotros los recursos variables y podemos así dar más servicio a nuestros clientes. La

Tabla 5-6 muestra el mismo estudio para tráfico de Internet, de la que se puede hacer una interpretación similar.

Tabla 5-5: Explotación anual para el tráfico de negocios

		Año 1	Año 2	Año 3	Año 4	Año 5
Alquiler de infraestructura	**Facturable (Gbps)**	5669	5669	5669	5669	5669
	Precio (u.c.)	0.5	0.5	0.5	0.5	0.5
	Total (u.c.)	2834	2834	2834	2834	2834
Explotación de la empresa	**Facturable (Gbps)**	0	0	211	2563	5856
	Precio (u.c.)	2	2	2	2	2
	Total (u.c.)	0	0	423	5127	11712
Óptimo	**Decisión**	Alquilar	Alquilar	Alquilar	Explotar	Explotar
	Ingresos (u.c.)	2834	2834	2834	5127	11712

Tabla 5-6: Explotación anual para el tráfico de Internet

		Año 1	Año 2	Año 3	Año 4	Año 5
Alquiler de infraestructura	Facturable (Gbps)	5669	5669	5669	5669	5669
	Precio (u.c.)	0.5	0.5	0.5	0.5	0.5
	Total (u.c.)	2834	2834	2834	2834	2834
Explotación de la empresa	Facturable (Gbps)	0	0	0	980	4436
	Precio (u.c.)	2	2	2	2	2
	Total (u.c.)	0	0	0	1959	8871
Óptimo	Decisión	Alquilar	Alquilar	Alquilar	Alquilar	Explotar
	Ingresos (u.c.)	2834	2834	2834	2834	8871

De este modo, hemos encontrado la distribución óptima de cada anualidad, por lo que respecta a la decisión de alquilar o explotar. No obstante, aún no hemos tomado una decisión acerca de qué equipamiento instalar (orientado a negocios o Internet). Para tomar esta decisión final, bastará con sumar los ingresos esperados de cada año con las dos alternativas y elegir el mayor valor. Así, el uso de recursos variables para tráfico de negocios reportaría unos ingresos extras de 25342 u.c., mientras que el uso para tráfico de Internet los daría de 20209 u.c.. Con esta información, solo nos queda concluir que el mejor uso que se le puede dar a los recursos variables es para tráfico de negocios, alquilando la infraestructura a otra compañía durante los primeros tres años y explotando los propios recursos en los dos siguientes. Esta acción reportaría unos ingresos previstos sumando todos los recursos y años de 142754 u.c.

5.2 Planificación de redes ópticas

5.2.1 Definición del problema

En este caso de aplicación, mostraremos como se puede hacer uso de los modelos de estimación de tráfico para diseñar la topología de la red para asegurar un cierto volumen de tráfico objetivo con el mínimo coste posible. Definimos el problema formalmente como:

Dados:

a) Una topología conexa G(N, E), donde N representa el conjunto de nodos y E representa el conjunto de posibles enlaces de fibra óptica uniendo pares de nodos,

b) el tamaño de espectro S y de segmento Δ que se desea instalar,

c) el conjunto de clases de servicio *K*; caracterizados por su ancho de banda (*bw)*, la proporción mínima y máxima en el perfil de tráfico (TPmin y TPmax)*,* y la máxima probabilidad de bloqueo permitida.

d) el volumen de ancho de banda total servido *BWexp* que deberá poder admitir la red una vez diseñada,

Encontrar:

a) una topología G'(*N*, *E**),

b) el valor de *I* y *TP*.

Tal que:

i) $E^* \subseteq E$ (i.e.: el conjunto de arcos sea un subconjunto de la conectividad inicial),

ii) *G'* sea un grafo conexo,

iii) El ancho de banda que puede admitir la red sea igual o superior a *BWexp,*

iv) Todas las clases de servicio tengan una probabilidad de bloqueo inferior o igual al umbral predefinido,

v) La proporción de cada clase de servicio esté entre los valores mínimos y máximos predeterminados.

Objetivo: minimizar el número de enlaces de la red (i.e. min $|E^*|$)

Este tipo de problemas de diseño de topología que parten de una conectividad inicial y devuelven una sub-topología de ésta, son comúnmente utilizados en planificación de redes ópticas (e.g [Ve11, Ru11]). La explicación es la siguiente: la implantación de una nueva tecnología en una red requiere la instalación de equipos en los nodos, que permiten que un cierto enlace tenga la funcionalidad propia de la tecnología. Así, los enlaces de fibra óptica, de por sí, no están ligados a ninguna tecnología concreta. Es por eso que, el diseño de una red nueva se puede realizar a partir de toda la fibra instalada a la que tiene acceso un operador y/o propietario de la red. Además, la instalación de fibra es un proceso muy caro y generalmente evitable (requiere de obra pública de dimensiones considerables), lo que lógicamente nos lleva a considerar la fibra ya desplegada para diseñar nuestra topología de la forma más eficiente.

La obtención de una solución óptima para este problema se puede hacer a partir de la resolución de un modelo de programación matemática. La siguiente sección introduce los aspectos más generales de su formulación y resolución.

5.2.2 Método de resolución

El problema descrito anteriormente se puede formular como un modelo de programación lineal entera con variables binarias, con la excepción de la restricción

que computa el ancho de banda que serviría la red y lo compara con BWexp. La no linealidad radica en que la estimación del ancho de banda se haría con los modelos de estimación de tráfico que hemos desarrollado, y estos incluyen productos y potencias de variables de decisión. No obstante, se puede linealizar con cierta facilidad.

En este problema, la configuración de capacidad viene dada como parámetros, así que diferentes soluciones factibles solo diferirán en la topología (es decir, en el factor ß del modelo de intensidad). Esta ß depende a su vez de dos características: el número de enlaces (nuestra función objetivo a minimizar) y la conectividad algebraica de la red. Dada la naturaleza de nuestro problema, el número de enlaces es un valor entero que varía en un rango relativamente pequeño. Esto nos abre la posibilidad de plantear un modelo iterativo que resuelva un subproblema en cada iteración y que nos permita optimizar el problema principal.

El subproblema buscaría una topología contenida en G y conexa, con un cierto número de enlaces predefinido, con el objetivo de maximizar la conectividad algebraica de la topología resultante. Esta formulación permite encontrar topologías que maximicen la intensidad (y por ende el ancho de banda servido) dado un cierto tamaño. La inclusión de esta formulación en un algoritmo iterativo que modifica el número de enlaces en cada iteración y que comprueba después de cada solución si el ancho de banda resultante es superior que el demandado, daría con la solución óptima del problema. Esto es lo que se ilustra en el algoritmo de la Tabla 5-7. Cabe puntualizar que encontrar una formulación que maximiza la conectividad algebraica no es trivial, pero puede hacerse uso, como función objetivo alternativa, del camino mínimo medio, cuya correlación con la conectividad algebraica ha sido profundamente estudiada [Mo11].

Tabla 5-7: Algoritmo de minimización de |E|

```
Entrada: Topología G(N,E), capacidad (S,Δ), características del
tráfico, BWexp.
Salida: Topología G'(N,E*), BWtotal
nlinks=|N|-1
BWtotal=0;
Mientras BWtotal< BWexp hacer
     nlinks++
     Obtener G' resolviendo el subproblema con nlinks
     Actualizar BWtotal con G' a partir de los modelos de I y TP.
Devolver G', BWtotal
```

Capítulo 6.

Conclusiones

En este apartado, se resumen las contribuciones más importantes surgidas como desarrollo de este proyecto. Además, se muestra cual ha sido el desarrollo temporal de las diferentes fases del proyecto, así como varias líneas de continuación del trabajo realizado.

6.1 Contribuciones y valoración personal

Con el fin de resolver el objetivo principal de este proyecto, en primer lugar nos planteamos la necesidad de encontrar un método que permitiera obtener los datos necesarios para realizar los modelos. Así, diseñamos e implementamos un algoritmo en lenguaje Omnet que, haciendo varias llamadas a un simulador de redes de comunicaciones, conseguía obtener la información deseada.

Una vez obtenida esa información, la segunda y más importante contribución de este proyecto es la obtención de un modelo de predicción de tráfico dinámico para redes ópticas elásticas. Tras la aplicación de diversas técnicas estadísticas, hemos obtenido un modelo para la estimación de la intensidad y otro para el perfil de tráfico. Las conclusiones obtenidas a partir del escrutinio de los modelos son un aporte novedoso de este proyecto. Las más importantes son: a) la independencia entre el efecto de la capacidad y el efecto de la topología para la estimación de la intensidad; b) la relación entre la conectividad algebraica de la topología y la intensidad; y c) la independencia de la topología y del tamaño del espectro en la selección del perfil de tráfico óptimo.

Finalmente, y con el ánimo de ilustrar la aplicabilidad de nuestro trabajo, hemos presentado dos problemas de planificación que pueden resolverse mediante el uso de los modelos. El primero, un problema de dimensionamiento de equipos ópticos, ha sido ilustrado con un ejemplo numérico sobre un supuesto práctico. El segundo,

de diseño de topología, ha sido introducido proponiendo las bases de un método para la obtención de la solución óptima.

Con estas contribuciones podemos concluir que el objetivo principal del proyecto definido en la introducción de la presente memoria ha sido satisfactoriamente resuelto.

Personalmente, para la realización de este proyecto he requerido de una breve formación en el ámbito de las redes de fibra óptica. Entender los conceptos de las variables que entran en el estudio ha resultado fundamental para la realización del trabajo, así como el mecanismo del simulador usado para obtener los datos. Todo este aprendizaje fue transmitido tanto por mis directores de proyecto como por otros integrantes del Grupo de Comunicaciones Ópticas (GCO) de la UPC, que también me cedió un espacio y equipo en sus instalaciones para poder desarrollar el proyecto.

Desde el punto de vista estadístico, como se ha puesto un especial énfasis en los modelos lineales, el hecho de tener unos sólidos conocimientos para obtener buenos modelos y, posteriormente, validarlos mediante un correcto análisis de residuos ha sido primordial. Además, también he experimentado la dificultad del proceso de obtención de datos; en este caso por la gran cantidad de tiempo necesario para tener disponibles el conjunto completo de datos necesario para realizar los modelos.

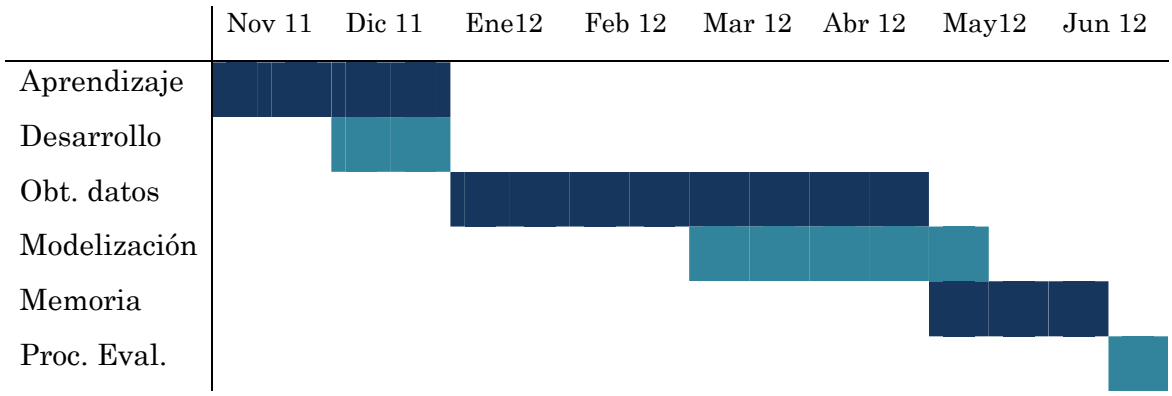
Des del punto de vista computacional, los softwares que he utilizado han sido Excel, R, Matlab, y OMNeT. Debido a que Excel es un software muy común, ya disponía de una base previa que me ha ayudado en el tratamiento de los datos. Toda la parte de obtener correlaciones, analizar modelos, validar modelos y realizar tests estadísticos se ha hecho con el R y, puesto que en la licenciatura se utiliza en muchas asignaturas, el uso de este software no ha sido un impedimento en el desarrollo del proyecto. En cambio, Matlab (utilizado para obtener los valores de las variables explicativas) es el software que me ha dado más trabajo ya que, al no utilizarse demasiado durante la licenciatura, ha sido necesario un buen repaso en el funcionamiento de este software. Finalmente, el uso que he hecho de OMNeT ha sido a un nivel mucho más simple (lanzar ejecuciones, recoger resultados). Una parte importante de este proyecto ha sido el diseño del algoritmo TPmaker, cuya implementación en OMNeT (lenguaje C++) fue desarrollada por un miembro del grupo de investigación en el que he colaborado.

Finalmente, hay que decir que este proyecto también conlleva una buena dedicación en tiempo que dificulta, de manera ostensible, compaginarlo con otras actividades.

6.2 Desarrollo temporal

La Tabla 6-1 muestra el diagrama de Gantt con la distribución temporal de cada una de las fases del proyecto. Estas son: la fase de aprendizaje de conceptos previos; el desarrollo del método de obtención de datos en el simulador; la obtención de datos a partir del diseño de experimentos creado; la obtención de los modelos y su posterior validación; la escritura de la memoria conteniendo los detalles anteriormente descritos más la descripción de los casos prácticos; y la fase de proceso de evaluación del trabajo (depósito, matrícula, y defensa).

Tabla 6-1: Diagrama de Gantt del presente proyecto



De este diagrama, conviene destacar el tiempo dedicado a la obtención de los datos. Los motivos de tan amplia franja temporal son principalmente tres: la alta complejidad de tanto el algoritmo como el simulador; la exhaustividad de las ejecuciones lanzadas (muchas redes y configuraciones de capacidad) que además han sido replicadas varias veces para obtener una mayor precisión en los datos obtenidos; y la limitación en la cantidad de recursos computacionales que se han podido destinar a este proyecto.

6.3 Trabajo futuro

Como continuación inmediata y dado que las conclusiones efectuadas tienen un gran interés en la investigación actual en este tipo de redes, se ha planeado someter los resultados de este trabajo (junto con alguna extensión) a una conferencia de prestigio mundial en el mundo de las redes de comunicaciones.

Las extensiones naturales que se pueden realizar a partir de este proyecto se pueden separar en dos direcciones. El primero sería ampliar los escenarios de aplicabilidad de estos modelos, realizando el mismo tipo de estudios para diferentes configuraciones iniciales. Un ejemplo seria modificar los límites de perfil de tráfico a unos valores más estrictos y/o proporcionales al ancho de banda de cada

demanda. La segunda extensión sería la de extender los estudios de planificación de red usando los modelos. Aquí la continuación más inmediata sería desarrollar el procedimiento de diseño de topología que se introdujo en el capítulo de casos de estudio de la presente memoria. Cabe puntualizar que si no hemos desarrollado ninguna de estas extensiones durante este proyecto ha sido porque quedaban fuera de los objetivos previos, dimensionados para ajustarse a la extensión de un proyecto final de carrera.

Apéndice A.

Topologías usadas

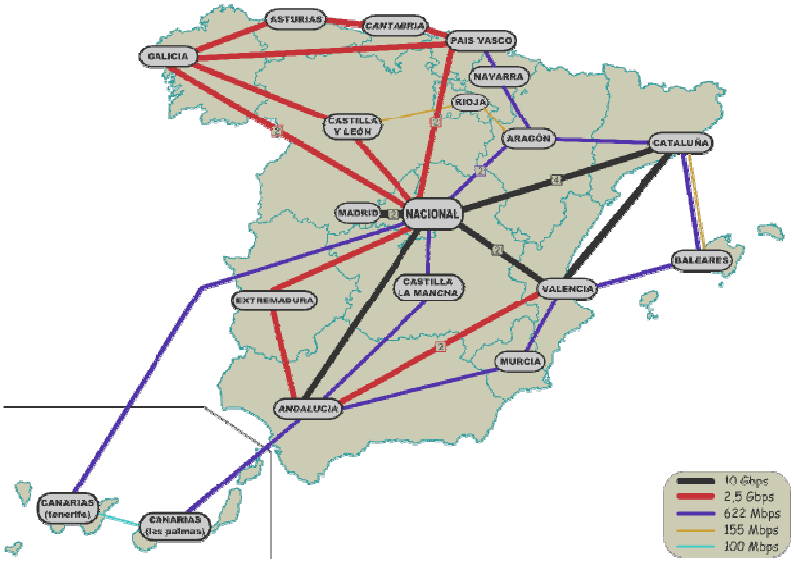


Fig. A-1 Red TP5

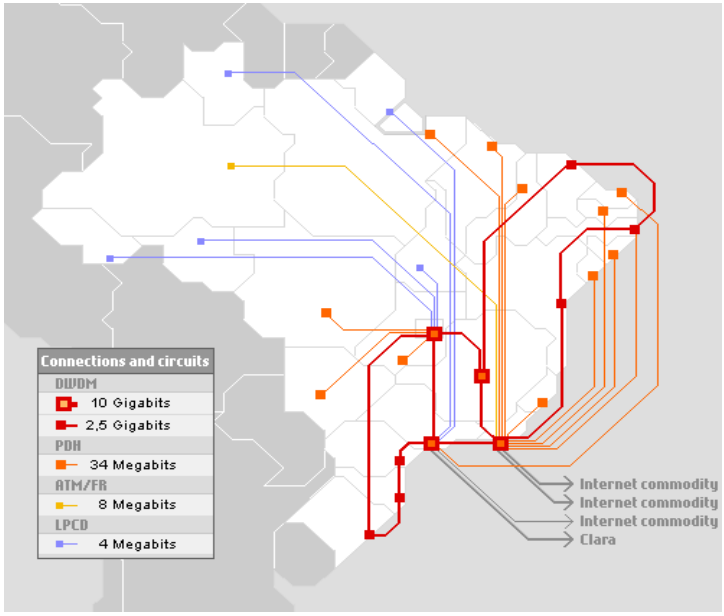


Fig. A -2 Red TP6

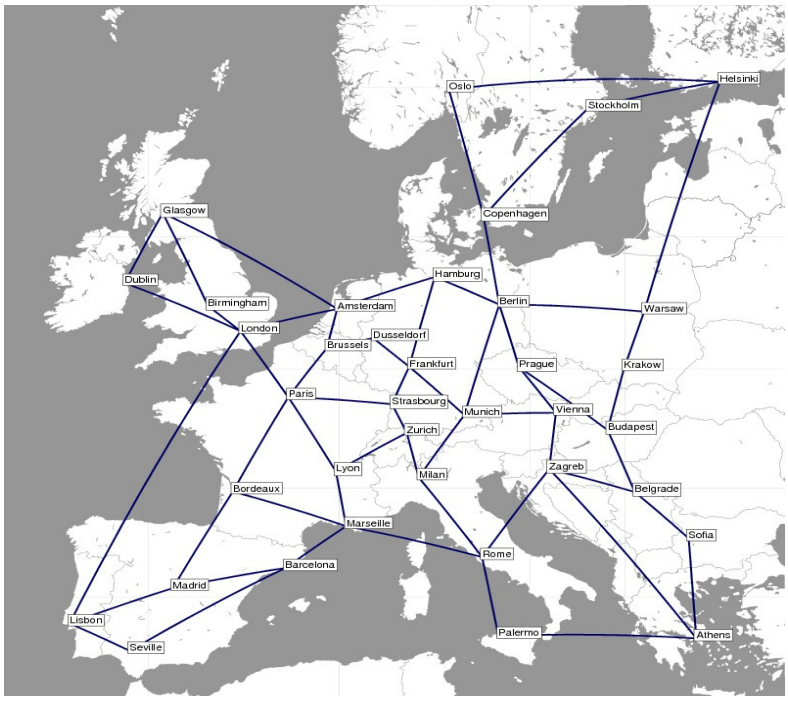


Fig. A -3 Red Cost266

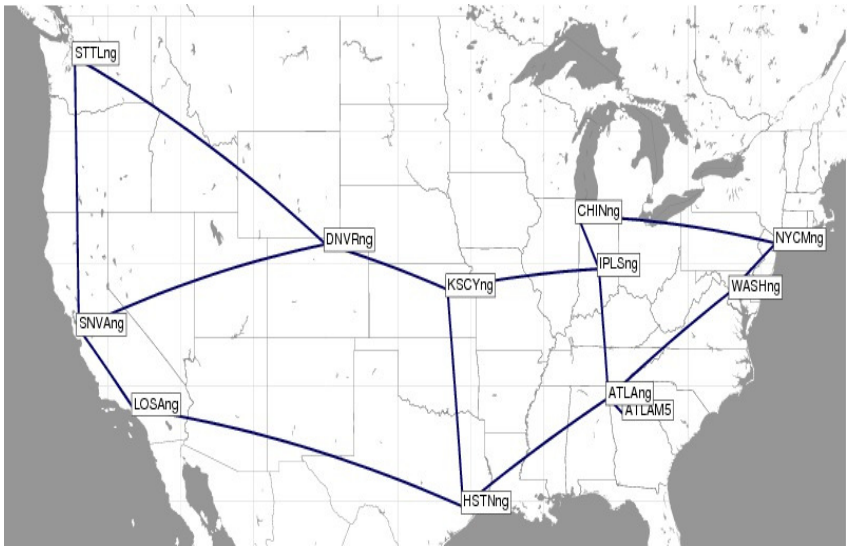


Fig. A -4 Red Abilene

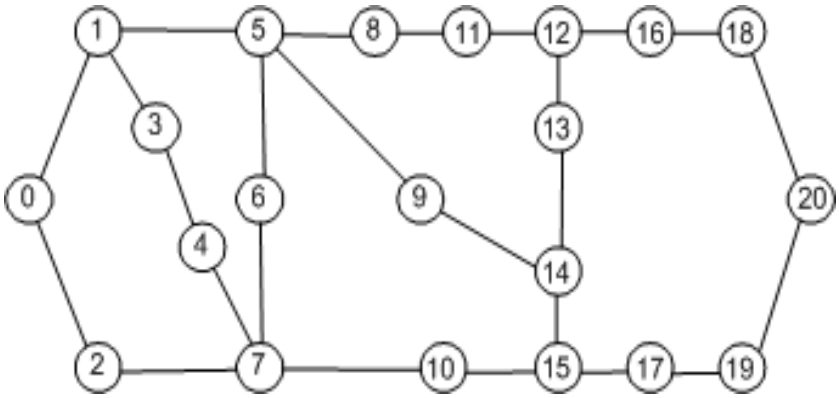


Fig. A -5 Red TP7

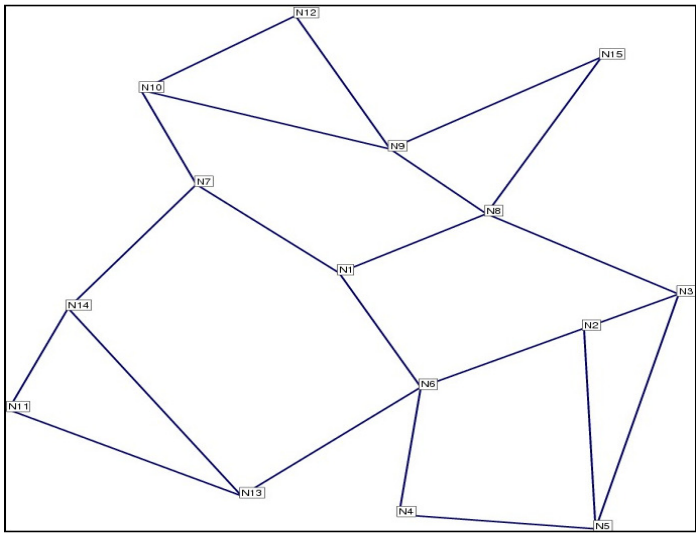


Fig. A -6 Red Atlanta

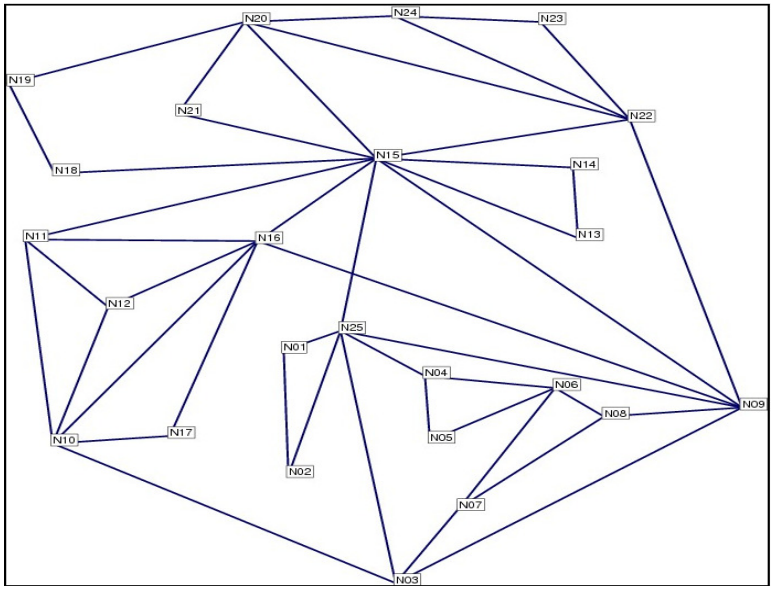


Fig. A-7 Red France

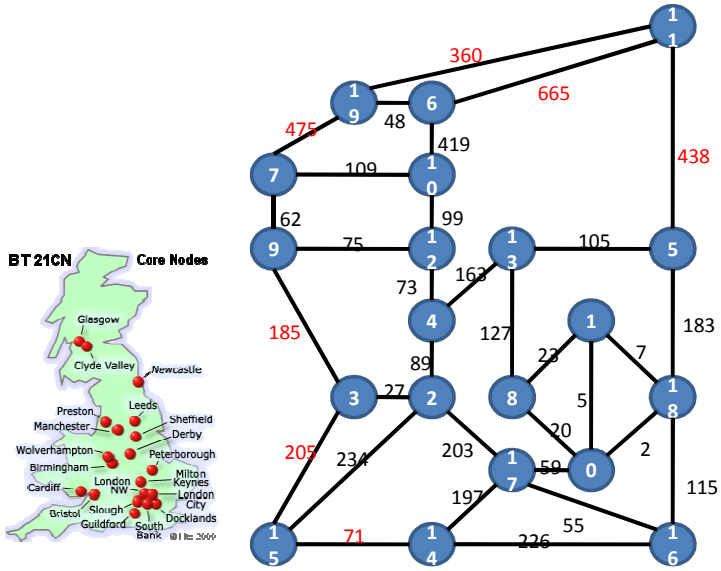


Fig. A -8 Red Geant

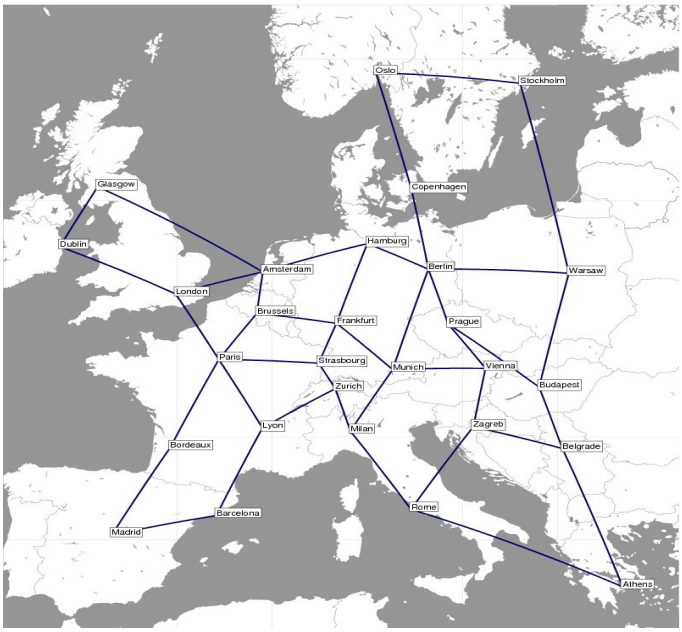


Fig. A -9 Red EU

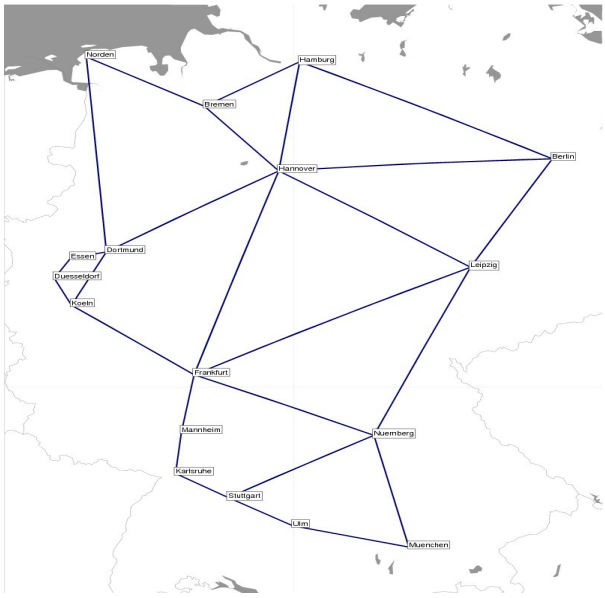


Fig. A -10 Red Germany

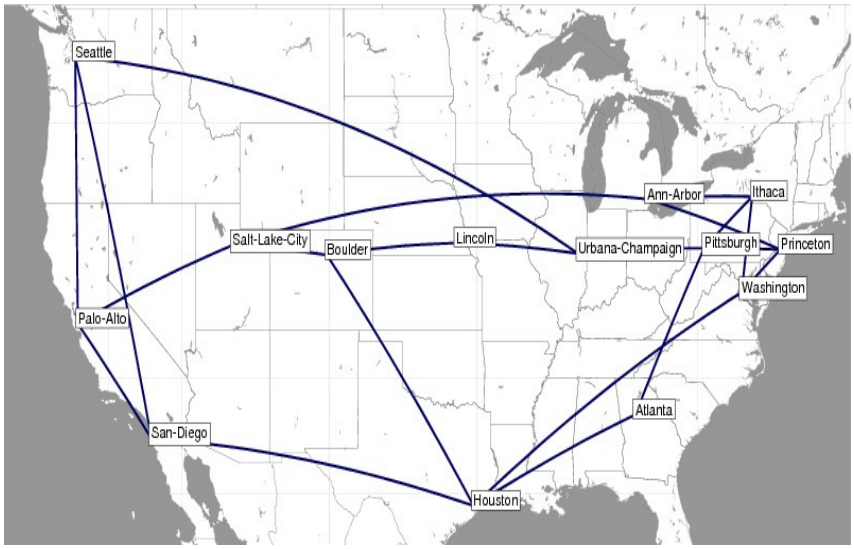


Fig. A -11 Red US

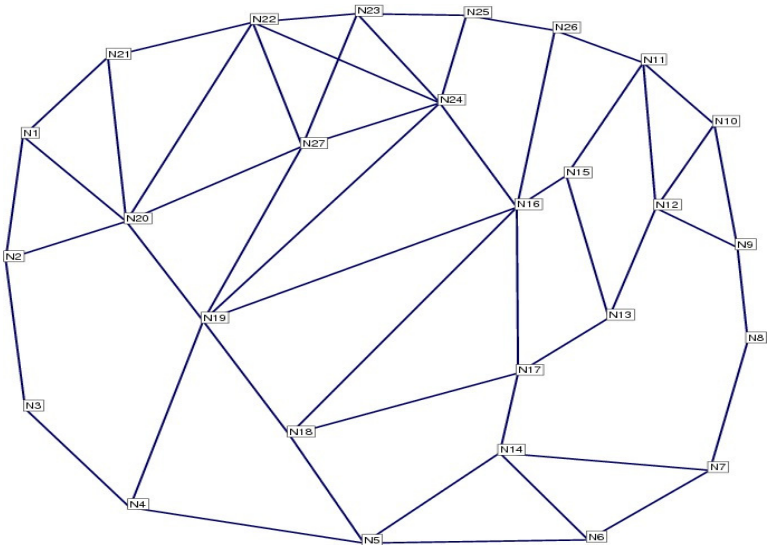


Fig. A -12 Red Norway

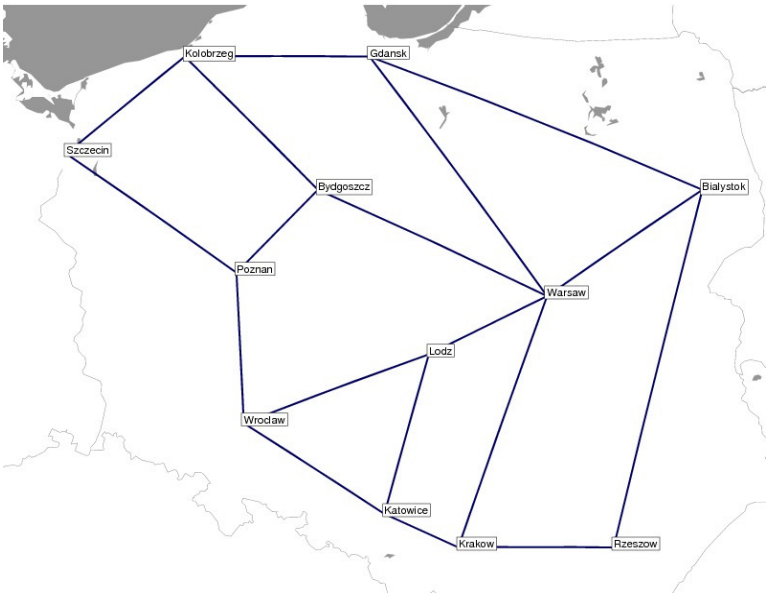


Fig. A -13 Red Polska

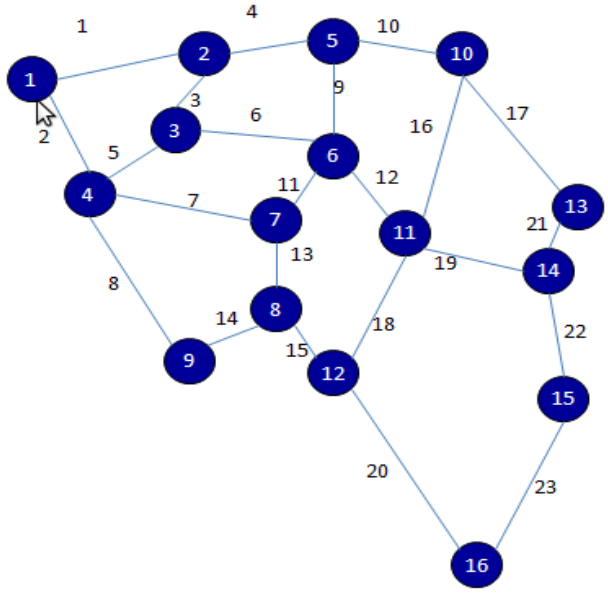


Fig. A -14 Red EONCT

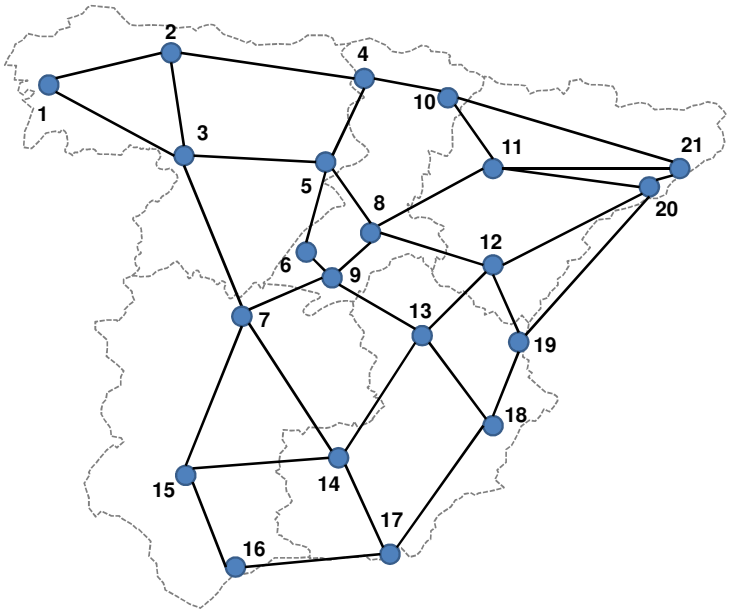


Fig. A -15 Red TID

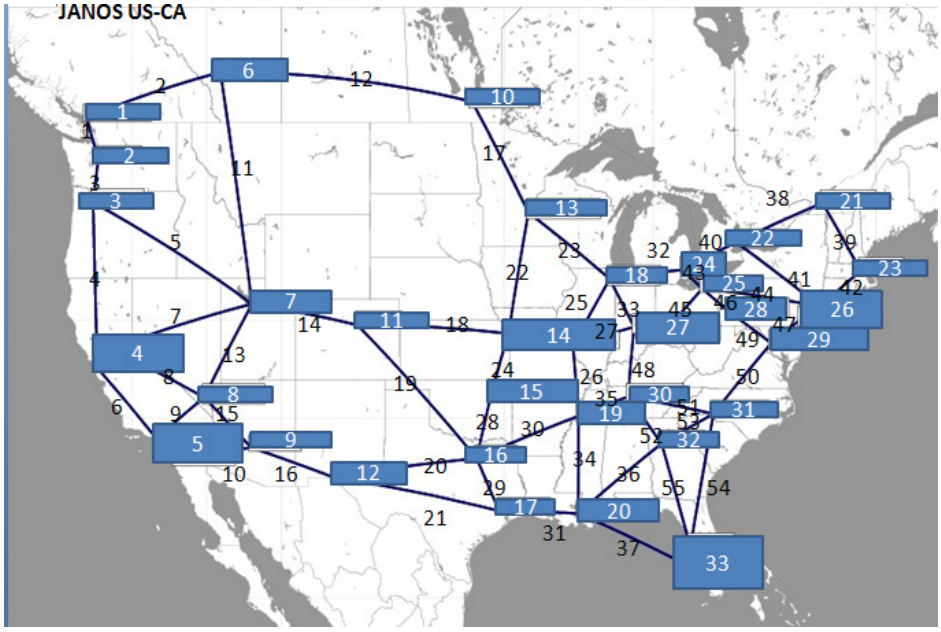


Fig. A -16 Red JANOS

Apéndice B.

Código implementado

Este apéndice ilustra el principal código desarrollado en este proyecto. Así, se muestra el principal código del algoritmo TPmaker implementado en OMNeT ,omitiendo aquellas funciones ya disponibles o que no ofrecen demasiado interés (escritura de datos, pre-procesos,etc). Además, como las instrucciones de R que se han ejecutado para obtener los modelos.

Algoritmo TPmaker

```
int main(int argc, char *argv[])
{
      vector<double>   TPmax,TPmin,TPsimulator,demands, PBdemand;
      double     PBmax = 0,BWsimulator = 0, offered_load = 0,
increment = 0, spectrumWidth = 0, subWavelengthWidth = 0, ini = 0,
end = 0, inc = 0, fragmentation = 0;
      long double runningTime = 0;
      bool  stop = false, TPchange = true, fragFile = false;
      char network[1024];
      int fragIni = 0, fragStep = 0, maxPathRequests = 0,
discardPaths = 0;
        timespec t1, t2;

      strcpy(network,argv[1]);
      spectrumWidth = atof(argv[2]);
      subWavelengthWidth = atof(argv[3]);
      ini = atof(argv[4]);
      end = atof(argv[5]);
      inc = atof(argv[6]);

      loadTestFile(demands,TPmax,TPmin,PBmax,increment,PBdemand,TPs
imulator,maxPathRequests,discardPaths,fragFile,fragIni,fragStep);
```

```
      printResultsHeader(network,spectrumWidth,subWavelengthWidth,d
emands,TPsimulator,PBdemand);

      for(offered_load = ini; offered_load <= end; offered_load +=
inc)
      {
            stop = false,
            TPchange = true;

            writeIniFile(network, offered_load, spectrumWidth,
subWavelengthWidth, maxPathRequests, discardPaths, fragFile,
fragIni, fragStep);

            for(unsigned int i = 0; i < demands.size(); i++)
                  cout << "Demand: " << demands.at(i) << " TPmax: "
<< TPmax.at(i) << " TPmin: " << TPmin.at(i) << "\n";
            cout << "PBmax: " << PBmax << "\nIncrement: " <<
increment << "\nOffered load: " << offered_load << "\n";
            cout << "Max paths requests: " << maxPathRequests << "
Discard paths: " << discardPaths << "\n";
            cout << "Fragmentation file: " << fragFile << "
Fragmentation ini: " << fragIni << " Fragmentation step: " <<
fragStep;
            cout << "\n\n";

            clock_gettime(CLOCK_REALTIME, &t1);
            initializeTPsimulator(TPsimulator,TPmax,TPmin);

            for(int i = TPsimulator.size() - 1; i > 0; i--)
            {
                  while(true)
                  {
                        if(TPchange)
                        {

      runSimulation(demands,TPsimulator,PBdemand,fragmentation);

                              for(unsigned int j = 0; j <
demands.size(); j++)
                                    cout << "Demand: " <<
demands.at(j) << "\tTPsimulator: " << TPsimulator.at(j) <<
"\tPBdemand: " << PBdemand.at(j) << "\n";
                        }

                        if(isFeasible(PBmax,PBdemand))
                        {
                              BWsimulator =
calculateBW(offered_load,demands,TPsimulator,PBdemand);
                              stop = true;
                              break;
                        }
                        else
                        {
```

```
                              if((TPsimulator.at(i) - increment) >=
TPmin.at(i))
                              {

      updateTPsimulator(TPsimulator,TPmax,i,increment);
                                    TPchange = true;
                              }
                              else
                              {
                                    TPchange = false;
                                    break;
                              }
                        }
                  }

                  if(stop)
                        break;
            }
            clock_gettime(CLOCK_REALTIME, &t2);
            runningTime = (t2.tv_sec * pow(10,9) + t2.tv_nsec) -
(t1.tv_sec * pow(10,9) + t1.tv_nsec);

            if(stop)
            {

      printResults(network,offered_load,spectrumWidth,subWavelength
Width,TPsimulator,PBdemand,BWsimulator,fragmentation,runningTime);
                  cout << "BW: " << BWsimulator << "  Time elapsed:
" << runningTime << " nano\n\n";
            }
            else
            {

      printResultsNoSolution(network,offered_load,spectrumWidth,sub
WavelengthWidth);
                  cout << "BW: " << "NO SOLUTION" << "  Time
elapsed: " << runningTime << " nano\n\n";
            }
      }

      return 0;
}
```

Scripts R

Script R para las correlaciones y el ACP

```
##Correlaciones variables
variables<-
read.csv2("C:/Users/Virgh/Desktop/Projecte/codigoVariables/Variables.c
sv",header=FALSE,dec=".")
```

```
rownames(variables)<-
c("TP5","TP6","Cost266","Abilene","TP7","Atlanta","France","EU","UK","
Germany","US","Norway","Polska","Ta1")
colnames(variables)<-
c("nodes","links","avgdegree","avgh","radius","diameter","avgcluster",
"stdcluster","avgbetwlink","stdbetwlink","avgbetwnode","stdbetwnode","
connect")
variables<-variables[,1:13]
cor(variables[,c(1:13)], use="pairwise.complete.obs")
pairs(variables,main="correlacion todas las variables")
variables2<-
cbind(variables[,3],variables[,4],variables[,7],variables[,9],variable
s[,11])
rownames(variables2)<-
c("TP5","TP6","Cost266","Abilene","TP7","Atlanta","France","EU","UK","
Germany","US","Norway","Polska","Ta1")
colnames(variables2)<-
c("avgdegree","avgh","avgcluster","avgbetwlink","avgbetwnode")
pairs(variables2,main="correlacion variables ACP")

##Ranking Correlaciones
ranking<-matrix(c(rep(0,(78*3))),ncol=3,nrow=78)
cont<-1
ranking<-as.data.frame(ranking)
for (i in 1:ncol(variables)){
for (j in 1:ncol(variables)){
if (j>i){
ranking[cont,1]<-colnames(variables[i])
ranking[cont,2]<-colnames(variables[j])
ranking[cont,3]<-abs(cor(variables[,c(i,j)])[1,2])
cont<-cont+1
}}}
library(plyr)
rankord<-arrange(ranking,desc(ranking[,3]))

##Observación de la distribución
library(MASS)
gradoTP5<-c(4,2,2,3,2,7,2,2,3,3)
gradoTP6<-c(2,2,2,3,3,3,3,2,2,2)
gradoCost266<-
c(3,2,3,3,3,4,3,5,5,2,2,3,4,3,2,4,3,3,3,4,2,3,2,3,4,3,3,4,3,2,3,5,3,3,
2,2,3)
gradoAbilene<-c(3,2,2,3,3,3,2,3,3,2,2,2)
gradoTP7<-c(3,2,2,2,2,4,2,4,2,2,2,2,3,2,3,3,2,2,2,2,2)
gradoAtlanta<-c(2,3,3,3,3,3,4,2,2,4,4,3,3,2,3)
gradoFrance<-c(5,4,3,2,2,2,6,6,2,2,4,5,2,10,3,2,3,3,2,5,2,2,4,3,6)
gradoEU<-c(2,2,2,2,2,3,5,3,3,4,3,3,3,3,3,2,2,2,3,4,4,5,3,3,2,3,3,3)
gradoUK<-c(4,3,4,3,3,3,3,3,3,3,3,3,3,3,3,3,3,4,4,3)
gradoGermany<-c(2,2,3,2,4,5,2,2,3,3,6,3,3,4,4,2,2)
gradoUS<-c(3,3,3,3,3,4,2,3,2,3,3,5,2,3)
gradoNorway<-c(3,2,3,4,3,3,2,3,3,4,3,3,4,5,3,3,6,6,4,4,4,3,4,3,6,6,5)
gradoPolska<-c(2,3,3,3,3,3,3,5,3,3,3,2)
gradoTa1<-c(3,3,10,3,6,3,2,11,3,4,2,4,6,3,4,5,5,3,2,6,4,3,4,3)
fitdistr(gradoCost266, "poisson")
ks.test(gradoCost266,"ppois",3.08)
```

```
##ACP
library(amap)
analisis<-acp(variables2)
plotAll(analisis)
```

Script R para los modelos y los contrastes de hipótesis

```
 ##install.packages("RODBC")
library(RODBC)
data=odbcConnectExcel2007(file.choose())
sqlTables(data)
mydat=sqlFetch(data,"Reducida")
datos<-mydat[1:13,1:19]
odbcClose(data)

##MODELOS CON UNA SOLA VARIABLE
attach(datos)
uno<-lm(Betaprima~N)
summary(uno)
dos<-lm(Betaprima~E)
summary(dos)
tres<-lm(Betaprima~delta)
summary(tres)
AIC(tres)
cuatro<-lm(Betaprima~h)
summary(cuatro)
cinco<-lm(Betaprima~rad)
summary(cinco)
seis<-lm(Betaprima~dia)
summary(seis)
siete<-lm(Betaprima~c)
summary(siete)
ocho<-lm(Betaprima~sc)
summary(ocho)
nueve<-lm(Betaprima~bl)
summary(nueve)
diez<-lm(Betaprima~sbl)
summary(diez)
once<-lm(Betaprima~bn)
summary(once)
doce<-lm(Betaprima~sbn)
summary(doce)
trece<-lm(Betaprima~kappa)
summary(trece)
detach(datos)

###MODELO 2 VARIABLES########
attach(datos)
primero<-lm(Betaprima~delta+kappa)
summary(primero)
AIC(primero)
par(mfrow=c(2,2))
plot(primero)
cor(datos[,2:18])

###MODELO LOGARITMICO########
```

```
articulo<-lm(log(Betaprima)~log(E)+log(h))
summary(articulo)
AIC(articulo)
par(mfrow=c(2,2))
plot(articulo)

###MODELO LOGARITMICO(ESCOGIDO)########
logaritmico<-lm(log(Betaprima)~log(E)+log(kappa))
summary(logaritmico)
AIC(logaritmico)
logaritmico2<-lm(log(Betaprima)~log(E)+log(kappa)+(E:kappa))
summary(logaritmico2)
AIC(logaritmico2)
par(mfrow=c(2,2))
plot(logaritmico)

###MODELO 3 VARIABLES###
tri<-lm(log(Betaprima)~log(E)+log(kappa)+log(dia))
summary(tri)
AIC(tri)

###MODELO ALTERNATIVO########
primario<-lm(Betaprima~1)
step(primario,~+nodes+links+avgdegree+avgh+radius+diameter+avgcluster+
stdcluster+avgbetwlink+stdbetwlink+avgbetwnode+stdbetwnode+connect)
prueba<-lm(log(Betaprima)~log(avgdegree)+log(connect)+log(radius))
summary(prueba)

#Añadir radius aumenta R2 pero empeora la normalidad de los residuos
par(mfrow=c(2,2))
plot(logaritmico2)
plot(log(links),residuals(logaritmico2))
plot(log(connect),residuals(logaritmico2))
plot(fitted.values(logaritmico2),residuals(logaritmico2))
#logaritmico<-logaritmico2#(Para no volver a repetir la sintaxis de
los gráficos)

###ANALISIS RESIDUOS MODELO VS BETA########
par(mfrow=c(2,2))
plot(logaritmico)
exp(predict(logaritmico,datos))
detach(datos)
data=odbcConnectExcel2007(file.choose())
sqlTables(data)
mydat=sqlFetch(data,"Todo")
datos<-mydat[1:117,1:34]
odbcClose(data)

###ANALISIS RESIDUOS MODELO VS INTENSIDAD########
attach(datos)
abscisas<-seq(from=1,to=117,by=1)
Error<-Error2
resest<-Error/sd(Error)
plot(abscisas,resest,main="Errores de todos los valores")
abline(h=0)
qqnorm(resest,main="Normalidad residuos de todos los valores")
```

```
qqline(resest)
detach(datos)
espectrum800slot12.5<-subset(datos,(spectrum==800&slot==12.5))
espectrum800slot25<-subset(datos,(spectrum==800&slot==25))
espectrum800slot50<-subset(datos,(spectrum==800&slot==50))
espectrum1600slot12.5<-subset(datos,(spectrum==1600&slot==12.5))
espectrum1600slot25<-subset(datos,(spectrum==1600&slot==25))
espectrum1600slot50<-subset(datos,(spectrum==1600&slot==50))
espectrum3200slot12.5<-subset(datos,(spectrum==3200&slot==12.5))
espectrum3200slot25<-subset(datos,(spectrum==3200&slot==25))
espectrum3200slot50<-subset(datos,(spectrum==3200&slot==50))
abscisas<-seq(from=1,to=13,by=1)

###ANALISIS RESIDUOS MODELO VS INTENSIDAD EN FUNCIÓN DEL ESPECTRO Y
SLOT########
par(mfrow=c(3,3))
attach(espectrum800slot12.5)
resest<-Error/sd(Error)
plot(abscisas,resest,main="spectrum=800, slot=12.5")
abline(h=0)
detach(espectrum800slot12.5)
attach(espectrum800slot25)
resest<-Error/sd(Error)
plot(abscisas,resest,main="spectrum=800, slot=25")
abline(h=0)
detach(espectrum800slot25)
attach(espectrum800slot50)
resest<-Error/sd(Error)
plot(abscisas,resest,main="spectrum=800, slot=50")
abline(h=0)
detach(espectrum800slot50)
attach(espectrum1600slot12.5)
resest<-Error/sd(Error)
plot(abscisas,resest,main="spectrum=1600, slot=12.5")
abline(h=0)
detach(espectrum1600slot12.5)
attach(espectrum1600slot25)
resest<-Error/sd(Error)
plot(abscisas,resest,main="spectrum=1600, slot=25")
abline(h=0)
detach(espectrum1600slot25)
attach(espectrum1600slot50)
resest<-Error/sd(Error)
plot(abscisas,resest,main="spectrum=1600, slot=50")
abline(h=0)
detach(espectrum1600slot50)
attach(espectrum3200slot12.5)
resest<-Error/sd(Error)
plot(abscisas,resest,main="spectrum=3200, slot=12.5")
abline(h=0)
detach(espectrum3200slot12.5)
attach(espectrum3200slot25)
resest<-Error/sd(Error)
plot(abscisas,resest,main="spectrum=3200, slot=25")
abline(h=0)
detach(espectrum3200slot25)
```

```
attach(espectrum3200slot50)
resest<-Error/sd(Error)
plot(abscisas,resest,main="spectrum=3200, slot=50")
abline(h=0)
detach(espectrum3200slot50)
par(mfrow=c(3,3))
attach(espectrum800slot12.5)
resest<-Error/sd(Error)
qqnorm(resest,main="spectrum=800, slot=12.5")
qqline(resest)
detach(espectrum800slot12.5)
attach(espectrum800slot25)
resest<-Error/sd(Error)
qqnorm(resest,main="spectrum=800, slot=25")
qqline(resest)
detach(espectrum800slot25)
attach(espectrum800slot50)
resest<-Error/sd(Error)
qqnorm(resest,main="spectrum=800, slot=50")
qqline(resest)
detach(espectrum800slot50)
attach(espectrum1600slot12.5)
resest<-Error/sd(Error)
qqnorm(resest,main="spectrum=1600, slot=12.5")
qqline(resest)
detach(espectrum1600slot12.5)
attach(espectrum1600slot25)
resest<-Error/sd(Error)
qqnorm(resest,main="spectrum=1600, slot=25")
qqline(resest)
detach(espectrum1600slot25)
attach(espectrum1600slot50)
resest<-Error/sd(Error)
qqnorm(resest,main="spectrum=1600, slot=50")
qqline(resest)
detach(espectrum1600slot50)
attach(espectrum3200slot12.5)
resest<-Error/sd(Error)
qqnorm(resest,main="spectrum=3200, slot=12.5")
qqline(resest)
detach(espectrum3200slot12.5)
attach(espectrum3200slot25)
resest<-Error/sd(Error)
qqnorm(resest,main="spectrum=3200, slot=25")
qqline(resest)
detach(espectrum3200slot25)
attach(espectrum3200slot50)
resest<-Error/sd(Error)
qqnorm(resest,main="spectrum=3200, slot=50")
qqline(resest)
detach(espectrum3200slot50)

###CONTRASTES TP###
data=odbcConnectExcel2007(file.choose())
sqlTables(data)
mydat=sqlFetch(data,"Todo")
```

```
datos<-mydat[1:117,1:38]
odbcClose(data)

Abilene<-matrix(c(rep(0,3)))
Atlanta<-matrix(c(rep(0,3)))
Cost266<-matrix(c(rep(0,3)))
EU<-matrix(c(rep(0,3)))
France<-matrix(c(rep(0,3)))
Germany<-matrix(c(rep(0,3)))
Norway<-matrix(c(rep(0,3)))
Polska<-matrix(c(rep(0,3)))
TP5<-matrix(c(rep(0,3)))
TP6<-matrix(c(rep(0,3)))
TP7<-matrix(c(rep(0,3)))
UK<-matrix(c(rep(0,3)))
US<-matrix(c(rep(0,3)))

##Contrastes segmento
attach(datos)
slot<-as.factor(slot)
a1<-aov(TPbw~slot)
TukeyHSD(a1)
plot(TukeyHSD(a1))
detach(datos)

##Contrastes topologia y espectro
z<-1
Abilene[1]<-datos[z,38]
Abilene[2]<-datos[z+3,38]
Abilene[3]<-datos[z+6,38]
z<-z+9
Atlanta[1]<-datos[z,38]
Atlanta[2]<-datos[z+3,38]
Atlanta[3]<-datos[z+6,38]
z<-z+9
Cost266[1]<-datos[z,38]
Cost266[2]<-datos[z+3,38]
Cost266[3]<-datos[z+6,38]
z<-z+9
EU[1]<-datos[z,38]
EU[2]<-datos[z+3,38]
EU[3]<-datos[z+6,38]
z<-z+9
France[1]<-datos[z,38]
France[2]<-datos[z+3,38]
France[3]<-datos[z+6,38]
z<-z+9
Germany[1]<-datos[z,38]
Germany[2]<-datos[z+3,38]
Germany[3]<-datos[z+6,38]
z<-z+9
Norway[1]<-datos[z,38]
Norway[2]<-datos[z+3,38]
Norway[3]<-datos[z+6,38]
z<-z+9
Polska[1]<-datos[z,38]
```

```
Polska[2]<-datos[z+3,38]
Polska[3]<-datos[z+6,38]
z<-z+9
TP5[1]<-datos[z,38]
TP5[2]<-datos[z+3,38]
TP5[3]<-datos[z+6,38]
z<-z+9
TP6[1]<-datos[z,38]
TP6[2]<-datos[z+3,38]
TP6[3]<-datos[z+6,38]
z<-z+9
TP7[1]<-datos[z,38]
TP7[2]<-datos[z+3,38]
TP7[3]<-datos[z+6,38]
z<-z+9
UK[1]<-datos[z,38]
UK[2]<-datos[z+3,38]
UK[3]<-datos[z+6,38]
z<-z+9
US[1]<-datos[z,38]
US[2]<-datos[z+3,38]
US[3]<-datos[z+6,38]
slotlow<-
rbind(Abilene,Atlanta,Cost266,EU,France,Germany,Norway,Polska,TP5,TP6,
TP7,UK,US)
slotlow<-as.data.frame(slotlow)
a<-
matrix(c(800,1600,3200,800,1600,3200,800,1600,3200,800,1600,3200,800,1
600,3200,800,1600,3200,800,1600,3200,800,1600,3200,800,1600,3200,800,1
600,3200,800,1600,3200,800,1600,3200,800,1600,3200))
a<-as.factor(a)
b<-
factor(c("Abilene","Abilene","Abilene","Atlanta","Atlanta","Atlanta","
Cost266","Cost266","Cost266","EU","EU","EU","France","France","France"
,"Germany","Germany","Germany","Norway","Norway","Norway","Polska","Po
lska","Polska","TP5","TP5","TP5","TP6","TP6","TP6","TP7","TP7","TP7","
UK","UK","UK","US","US","US"))
slotlow<-cbind(b,slotlow,a)
colnames(slotlow)<-c("networkId","TPbw","spectrum")
z<-2
Abilene[1]<-datos[z,38]
Abilene[2]<-datos[z+3,38]
Abilene[3]<-datos[z+6,38]
z<-z+9
Atlanta[1]<-datos[z,38]
Atlanta[2]<-datos[z+3,38]
Atlanta[3]<-datos[z+6,38]
z<-z+9
Cost266[1]<-datos[z,38]
Cost266[2]<-datos[z+3,38]
Cost266[3]<-datos[z+6,38]
z<-z+9
EU[1]<-datos[z,38]
EU[2]<-datos[z+3,38]
EU[3]<-datos[z+6,38]
z<-z+9
```

```
France[1]<-datos[z,38]
France[2]<-datos[z+3,38]
France[3]<-datos[z+6,38]
z<-z+9
Germany[1]<-datos[z,38]
Germany[2]<-datos[z+3,38]
Germany[3]<-datos[z+6,38]
z<-z+9
Norway[1]<-datos[z,38]
Norway[2]<-datos[z+3,38]
Norway[3]<-datos[z+6,38]
z<-z+9
Polska[1]<-datos[z,38]
Polska[2]<-datos[z+3,38]
Polska[3]<-datos[z+6,38]
z<-z+9
TP5[1]<-datos[z,38]
TP5[2]<-datos[z+3,38]
TP5[3]<-datos[z+6,38]
z<-z+9
TP6[1]<-datos[z,38]
TP6[2]<-datos[z+3,38]
TP6[3]<-datos[z+6,38]
z<-z+9
TP7[1]<-datos[z,38]
TP7[2]<-datos[z+3,38]
TP7[3]<-datos[z+6,38]
z<-z+9
UK[1]<-datos[z,38]
UK[2]<-datos[z+3,38]
UK[3]<-datos[z+6,38]
z<-z+9
US[1]<-datos[z,38]
US[2]<-datos[z+3,38]
US[3]<-datos[z+6,38]
slotmedium<-
rbind(Abilene,Atlanta,Cost266,EU,France,Germany,Norway,Polska,TP5,TP6,
TP7,UK,US)
slotmedium<-as.data.frame(slotmedium)
slotmedium<-cbind(b,slotmedium,a)
colnames(slotmedium)<-c("networkId","TPbw","spectrum")
z<-3
Abilene[1]<-datos[z,38]
Abilene[2]<-datos[z+3,38]
Abilene[3]<-datos[z+6,38]
z<-z+9
Atlanta[1]<-datos[z,38]
Atlanta[2]<-datos[z+3,38]
Atlanta[3]<-datos[z+6,38]
z<-z+9
Cost266[1]<-datos[z,38]
Cost266[2]<-datos[z+3,38]
Cost266[3]<-datos[z+6,38]
z<-z+9
EU[1]<-datos[z,38]
EU[2]<-datos[z+3,38]
```

```
EU[3]<-datos[z+6,38]
z<-z+9
France[1]<-datos[z,38]
France[2]<-datos[z+3,38]
France[3]<-datos[z+6,38]
z<-z+9
Germany[1]<-datos[z,38]
Germany[2]<-datos[z+3,38]
Germany[3]<-datos[z+6,38]
z<-z+9
Norway[1]<-datos[z,38]
Norway[2]<-datos[z+3,38]
Norway[3]<-datos[z+6,38]
z<-z+9
Polska[1]<-datos[z,38]
Polska[2]<-datos[z+3,38]
Polska[3]<-datos[z+6,38]
z<-z+9
TP5[1]<-datos[z,38]
TP5[2]<-datos[z+3,38]
TP5[3]<-datos[z+6,38]
z<-z+9
TP6[1]<-datos[z,38]
TP6[2]<-datos[z+3,38]
TP6[3]<-datos[z+6,38]
z<-z+9
TP7[1]<-datos[z,38]
TP7[2]<-datos[z+3,38]
TP7[3]<-datos[z+6,38]
z<-z+9
UK[1]<-datos[z,38]
UK[2]<-datos[z+3,38]
UK[3]<-datos[z+6,38]
z<-z+9
US[1]<-datos[z,38]
US[2]<-datos[z+3,38]
US[3]<-datos[z+6,38]
slothigh<-
rbind(Abilene,Atlanta,Cost266,EU,France,Germany,Norway,Polska,TP5,TP6,
TP7,UK,US)
slothigh<-as.data.frame(slothigh)
slothigh<-cbind(b,slothigh,a)
colnames(slothigh)<-c("networkId","TPbw","spectrum")
attach(slotlow)
a3<-aov(TPbw~networkId+spectrum)
TukeyHSD(a3)
detach(slotlow)
attach(slotmedium)
a4<-aov(TPbw~networkId+spectrum)
TukeyHSD(a4)
detach(slotmedium)
attach(slothigh)
a5<-aov(TPbw~networkId+spectrum)
TukeyHSD(a5)
detach(slothigh)
par(mfrow=c(3,1))
```

```
plot(TukeyHSD(a3,"networkId"),main="slot12.5")
plot(TukeyHSD(a4,"networkId"),main="slot25")
plot(TukeyHSD(a5,"networkId"),main="slot50")
par(mfrow=c(3,1))
plot(TukeyHSD(a3,"spectrum"),main="slot12.5")
plot(TukeyHSD(a4,"spectrum"),main="slot25")
plot(TukeyHSD(a5,"spectrum"),main="slot50")
```

Apéndice C.

Tabla de resultados

A continuación se presenta la tabla con los valores de la variable respuesta para cada una de las redes y configuraciones de capacidad usadas para obtener los modelos:

Red	S	Δ	I	TP(10)	TP(40)	TP(100)	TP(400)	BW medio
Abilene	800	12.5	36	0.10	0.20	0.59	0.11	112.15
Abilene	800	25	34	0.10	0.10	0.69	0.11	118.00
Abilene	800	50	13	0.10	0.10	0.12	0.68	289.75
Abilene	1600	12.5	104	0.10	0.25	0.54	0.11	108.76
Abilene	1600	25	96	0.10	0.10	0.69	0.11	118.00
Abilene	1600	50	36	0.10	0.10	0.13	0.67	287.00
Abilene	3200	12.5	249	0.10	0.16	0.63	0.11	114.40
Abilene	3200	25	225	0.10	0.10	0.68	0.12	120.00
Abilene	3200	50	89	0.10	0.10	0.14	0.66	284.33
Atlanta	800	12.5	51	0.10	0.19	0.60	0.11	112.36
Atlanta	800	25	47	0.10	0.10	0.68	0.12	121.00
Atlanta	800	50	19	0.10	0.10	0.11	0.69	291.40
Atlanta	1600	12.5	145	0.10	0.22	0.57	0.11	110.97
Atlanta	1600	25	130	0.10	0.10	0.68	0.13	122.50
Atlanta	1600	50	53	0.10	0.10	0.14	0.67	284.50
Atlanta	3200	12.5	344	0.10	0.14	0.65	0.11	116.35

Atlanta	3200	25	320	0.10	0.10	0.68	0.12	120.40
Atlanta	3200	50	115	0.10	0.10	0.13	0.67	286.00
Cost266	800	12.5	44	0.10	0.17	0.62	0.11	113.70
Cost266	800	25	43	0.10	0.13	0.66	0.11	116.10
Cost266	800	50	17	0.10	0.10	0.13	0.67	286.00
Cost266	1600	12.5	126	0.10	0.17	0.62	0.11	113.60
Cost266	1600	25	125	0.10	0.16	0.63	0.11	114.16
Cost266	1600	50	48	0.10	0.10	0.13	0.67	286.00
Cost266	3200	12.5	333	0.10	0.32	0.47	0.11	104.65
Cost266	3200	25	302	0.10	0.12	0.68	0.11	117.10
Cost266	3200	50	118	0.10	0.10	0.11	0.69	292.67
EU	800	12.5	36	0.10	0.27	0.52	0.11	107.92
EU	800	25	35	0.10	0.20	0.59	0.11	112.00
EU	800	50	13	0.10	0.10	0.10	0.70	293.80
EU	1600	12.5	110	0.10	0.30	0.50	0.11	106.30
EU	1600	25	104	0.10	0.20	0.59	0.11	112.12
EU	1600	50	39	0.10	0.10	0.13	0.67	287.00
EU	3200	12.5	271	0.10	0.25	0.54	0.11	109.15
EU	3200	25	252	0.10	0.12	0.67	0.11	116.60
EU	3200	50	102	0.10	0.10	0.15	0.65	278.67
France	800	12.5	81	0.10	0.19	0.60	0.11	112.60
France	800	25	76	0.10	0.10	0.69	0.11	119.00
France	800	50	32	0.10	0.10	0.10	0.70	293.80
France	1600	12.5	218	0.10	0.26	0.53	0.11	108.16
France	1600	25	205	0.10	0.10	0.68	0.12	121.00
France	1600	50	83	0.10	0.10	0.13	0.67	286.86
France	3200	12.5	499	0.10	0.18	0.61	0.11	113.20
France	3200	25	410	0.10	0.10	0.64	0.17	134.50
France	3200	50	198	0.10	0.10	0.14	0.66	282.00
Germany	800	12.5	46	0.10	0.19	0.60	0.11	112.48
Germany	800	25	42	0.10	0.10	0.68	0.12	120.88
Germany	800	50	18	0.10	0.10	0.15	0.65	280.00

Germany	1600	12.5	128	0.10	0.19	0.60	0.11	112.84
Germany	1600	25	122	0.10	0.13	0.66	0.11	116.20
Germany	1600	50	47	0.10	0.10	0.13	0.67	286.00
Germany	3200	12.5	317	0.10	0.26	0.54	0.11	108.70
Germany	3200	25	283	0.10	0.11	0.67	0.12	119.65
Germany	3200	50	115	0.10	0.10	0.13	0.67	286.00
Norway	800	12.5	70	0.10	0.23	0.56	0.11	109.96
Norway	800	25	65	0.10	0.13	0.66	0.11	116.05
Norway	800	50	27	0.10	0.10	0.10	0.70	293.80
Norway	1600	12.5	210	0.10	0.43	0.36	0.11	98.20
Norway	1600	25	171	0.10	0.10	0.68	0.12	120.25
Norway	1600	50	73	0.10	0.10	0.13	0.67	287.00
Norway	3200	12.5	436	0.10	0.13	0.66	0.11	116.35
Norway	3200	25	415	0.10	0.11	0.68	0.12	119.80
Norway	3200	50	177	0.10	0.10	0.13	0.67	286.00
Polska	800	12.5	65	0.10	0.20	0.59	0.11	112.12
Polska	800	25	62	0.10	0.10	0.69	0.11	118.00
Polska	800	50	25	0.10	0.10	0.12	0.68	288.40
Polska	1600	12.5	174	0.10	0.18	0.61	0.11	113.00
Polska	1600	25	166	0.10	0.10	0.69	0.11	118.00
Polska	1600	50	65	0.10	0.10	0.13	0.67	287.00
Polska	3200	12.5	423	0.10	0.17	0.62	0.11	113.95
Polska	3200	25	382	0.10	0.10	0.68	0.12	120.00
Polska	3200	50	155	0.10	0.10	0.12	0.68	289.00
TP5	800	12.5	56	0.10	0.25	0.54	0.11	109.10
TP5	800	25	52	0.10	0.28	0.51	0.11	107.20
TP5	800	50	20	0.10	0.10	0.11	0.69	292.00
TP5	1600	12.5	142	0.10	0.15	0.64	0.11	115.12
TP5	1600	25	135	0.10	0.10	0.67	0.13	124.00
TP5	1600	50	52	0.10	0.10	0.11	0.69	292.50
TP5	3200	12.5	350	0.10	0.24	0.55	0.11	109.75
TP5	3200	25	322	0.10	0.10	0.68	0.12	120.00

TP5	3200	50	124	0.10	0.10	0.11	0.69	291.33
TP6	800	12.5	34	0.10	0.14	0.65	0.11	115.40
TP6	800	25	33	0.10	0.10	0.69	0.11	118.00
TP6	800	50	12	0.10	0.10	0.11	0.69	293.00
TP6	1600	12.5	99	0.10	0.24	0.55	0.11	109.66
TP6	1600	25	90	0.10	0.10	0.68	0.12	121.00
TP6	1600	50	34	0.10	0.10	0.12	0.68	289.00
TP6	3200	12.5	239	0.10	0.21	0.58	0.11	111.52
TP6	3200	25	218	0.10	0.10	0.69	0.11	118.00
TP6	3200	50	84	0.10	0.10	0.13	0.67	285.67
TP7	800	12.5	26	0.10	0.24	0.55	0.11	109.60
TP7	800	25	25	0.10	0.12	0.67	0.11	118.00
TP7	800	50	10	0.10	0.10	0.20	0.60	266.00
TP7	1600	12.5	85	0.10	0.31	0.48	0.11	105.31
TP7	1600	25	79	0.10	0.14	0.65	0.11	116.80
TP7	1600	50	30	0.10	0.10	0.14	0.66	283.00
TP7	3200	12.5	224	0.10	0.26	0.53	0.11	108.40
TP7	3200	25	216	0.10	0.20	0.59	0.11	112.15
TP7	3200	50	80	0.10	0.10	0.15	0.65	279.40
UK	800	12.5	60	0.10	0.16	0.63	0.11	114.60
UK	800	25	56	0.10	0.10	0.68	0.12	120.50
UK	800	50	23	0.10	0.10	0.12	0.69	290.50
UK	1600	12.5	168	0.10	0.17	0.62	0.11	113.58
UK	1600	25	155	0.10	0.16	0.63	0.12	116.60
UK	1600	50	62	0.10	0.10	0.13	0.67	285.50
UK	3200	12.5	410	0.10	0.20	0.60	0.11	112.30
UK	3200	25	370	0.10	0.10	0.69	0.11	119.00
UK	3200	50	148	0.10	0.10	0.11	0.69	291.33
US	800	12.5	73	0.10	0.14	0.65	0.11	115.75
US	800	25	70	0.10	0.10	0.68	0.12	120.25
US	800	50	29	0.10	0.10	0.11	0.69	290.80
US	1600	12.5	201	0.10	0.16	0.63	0.11	114.40

US	1600	25	191	0.10	0.10	0.69	0.11	118.00
US	1600	50	75	0.10	0.10	0.12	0.69	290.50
US	3200	12.5	490	0.10	0.18	0.61	0.11	113.35
US	3200	25	442	0.10	0.10	0.69	0.11	119.20
US	3200	50	180	0.10	0.10	0.11	0.69	292.00

Referencias

[Ca12] A. Castro, L. Velasco, M. Ruiz, M. Klinkowski, J. P. Fernández-Palacios, D. Careglio, "Dynamic Routing and Spectrum (Re)Allocation in Future Flexgrid Optical Networks", Computer Networks, online press, DOI: 10.1016, 2012.

[Ch97] R. Christensen, "Log-Linear Models and Logistic Regression", Springer Texts in Statistics, New York, 1997.

[Do02]. A. J. Dobson, "An introduction to generalized linear models", Chapman & Hall/CRC, Florida, 2002.

[ITU05] Telecom Development Bureau, Teletraffic Engineering Handbook, ITU, 2005.

[ITU-G.694] ITU-T G.694.1, "Spectral grids for WDM applications: DWDM frequency grid"

[LI11] Y. Li, F. Zhang, and R. Casellas, "Flexible grid label format in wavelength switched optical network," IETF RFC Draft, Jul. 2011.

[Ma03] S. De Maesschalck et al., "Pan-European Optical Transport Networks: An Availability-based Comparison," Photonic Network Communications, vol. 5, no. 20, pp. 203-225, 2003.

[Mat] MATLAB - http://www.mathworks.es/products/matlab/

[Mo11] B. Mohar, "Eigenvalues, diameter and mean distance in graphs," Graphs and Combinatorics vol. 7, pp. 53-64, 2011.

[OM] OMNeT++. http://www.omnetpp.org/

[Pa10] C. Pavan, R. Morais, J. Ferreira, and A. Pinto, "Generating Realistic Optical Transport Network Topologies,"IEEE/OSA Journal of. Optical Communications and. Networking, vol 2, pp. 80-90, 2010.

[R] R - http://cran.r-project.org/

[Ru11] M. Ruiz, J. Perelló, L. Velasco, S. Spadaro, J. Comellas, and G. Junyent, "GMPLS Control Plane Network Design with Resilience Guarantees", IEEE/OSA Journal of Lightwave Technology (JLT), 29, 37-47, 2011.

[SNDlib] http://sndlib.zib.de/home.action

[Ve11] L. Velasco et al, "Service and Resource Differentiation in Shared-Path Protection Environments to Maximize Network Operator's Revenues", IEEE/OSA Journal of Optical Communications and Networking (JOCN), 3, 117-126, 2011.

[Vm11] P. Van Mieghem, Graph Spectra of Complex Networks, Cambridge University Press, 2011.