

UNIVERSIDAD DE SALAMANCA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADÍSTICA



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

TRABAJO DE FIN DE GRADO

REDES DE COLAS

Tutor: MIGUEL RODRÍGUEZ ROSA

Autor: MANUEL CALVO MEZQUITA

Grado en Estadística

Curso académico 2022-23

UNIVERSIDAD DE SALAMANCA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADÍSTICA



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

TRABAJO DE FIN DE GRADO

REDES DE COLAS

Firmado por:

A handwritten signature in blue ink that reads 'Manuel C.' with a large, sweeping flourish extending to the right.

Autor: Manuel Calvo Mezquita

A handwritten signature in blue ink that reads 'Miguel Rodríguez Rosa' with a large, sweeping flourish extending to the right.

Tutor: Miguel Rodríguez Rosa

Grado en Estadística

Curso académico 2022-23



Certificado del tutor TFG Grado en Estadística

D. Miguel Rodríguez Rosa, profesor del Departamento de Estadística de la Universidad de Salamanca,

HACE CONSTAR:

Que el trabajo titulado “*Redes de colas*”, que se presenta, ha sido realizado por D. Manuel Calvo Mezquita, con DNI 71045168P y constituye la memoria del trabajo realizado para la superación de la asignatura Trabajo de Fin de Grado en Estadística en esta Universidad.

Salamanca, a 5 de septiembre de 2023.

Fdo.: Miguel Rodríguez Rosa

Índice general

Summary	9
Introducción	11
Objetivos	17
1. Redes de colas	19
1.1. ¿Qué son las redes de colas?	19
1.2. El enrutado	22
2. Tipos de redes	25
2.1. Redes abiertas	27
2.2. Redes cerradas	28
3. Redes de jackson	29
3.1. Redes abiertas de Jackson	31
3.1.1. Ecuaciones de tráfico	31
3.1.2. Condiciones de no saturación	34
3.1.3. Teorema de Jackson y sus consecuencias	35
3.1.4. Medidas de rendimiento	35
3.1.5. Ejemplos numéricos	35
3.2. Redes cerradas de Jackson	40
3.2.1. Ecuaciones de tráfico	40
3.2.2. Condiciones de no saturación	43
3.2.3. Medidas de rendimiento	43
3.2.4. Teorema de Jackson y sus consecuencias	44
3.2.5. Ejemplo numérico	44
4. Programa en R	49
5. Conclusiones	51
Bibliografía	53
Anexo	55

Summary

A queueing network is a system where multiple queues exist, and customers flow from one queue to another, potentially without a predetermined flow pattern. This study involves on the concept of queueing networks and the types of queueing networks that exist, including Jackson networks. An attempt will be made to specify a theory appropriately, deducing the conditions of non-saturation, traffic equations, performance measures, and the concept of bottleneck, as well as developing some examples that describe reality.

Introducción

Para comprender bien lo que son las redes de colas primero debemos comprender el significado y las partes de una única cola.

Podemos definir una cola como un sistema en el que los individuos entran, esperan dentro del sistema ser a escogidos por los diferentes métodos de selección, reciben un servicio en el servidor o uno de los servidores que exista en el sistema, para después abandonarlo.

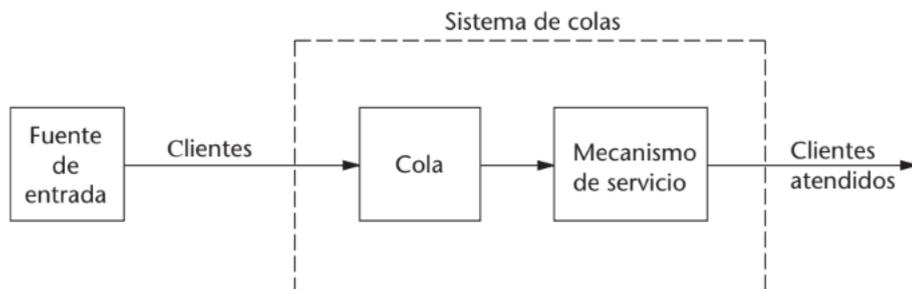


Figura 1: Elementos de una cola con un servidor (Lieberman & Hillier, 2010).

En la figura 1 podemos ver que un sistema de cola cuenta con múltiples partes que definiremos a continuación para comprenderlo a la perfección e identificar en ejemplos prácticos a qué corresponde cada cosa y cuál es su utilidad. Apoyándonos en lo que dice el libro de Lieberman y Hillier (2010):

- Primero definiremos la población de entrada. Esta la conforman los individuos que pueden entrar a nuestra cola desde el exterior. Por lo general, para simplificar los cálculos, se considera infinita cuando así lo es o es muy elevada; pero existen casos en los que esta es finita, debido a limitaciones externas o a que el propio sistema de espera limite la aparición de nuevos individuos para la entrada, por esperas prolongadas, y debe definirse como tal, ya que afectará a los tiempos de llegada del sistema.
- Los tiempos de llegada entre individuos son fundamentales para el cálculo de estimadores de rendimiento y tiempos totales en el sistema. Es la proporción inversa de la tasa de llegadas, definida por la letra λ , que define los individuos que llegan por unidad de tiempo en estado constante. Las tasas de llegada por lo general seguirán distribuciones de Poisson, lo cual facilita el cálculo, pero no todas tienen porqué. Existen infinidad de distribuciones posibles, pero las dividiremos en cuatro grupos principales: markovianas (exponenciales/-

Poisson), Erlang (atribuida al matemático Agner Krarup Erlang, pionero en teoría de colas), deterministas (para distribuciones no probabilísticas) y generales, para todo el resto de distribuciones (Gamma, Beta, Hipereexponencial, Tipo fase...).

Si bien los modelos que no siguen entradas de Poisson no son tan convenientes como se relata en el libro de Lieberman y Hillier (2010), estos existen y algunos de ellos tienen cálculos estructurados, como es el caso de las deterministas y las de Erlang. Para distribuciones generales, estas no imponen restricciones como la exponencial en los tiempos de llegada, mientras que las deterministas asignan una constante fija a estos tiempos y las Erlang manejan tasas de llegada tanto con constante fija como totalmente aleatorias.

- Los tiempos de servicio representan el tiempo que cada individuo tarda en recibir el servicio en su respectivo servidor. Es la inversa de la tasa de servicio (μ). Como en los tiempos de llegada, pueden seguir cualquier distribución, pero aquí sí es más común que sigan distribuciones diferentes a la exponencial, porque el servicio realizado definirá la distribución del tiempo de espera en el servidor. Además, el tipo de distribución modificará las fórmulas de las medidas de rendimiento, que se calcularán gracias a la fórmula de Pollaczek-Khintchine (Lieberman & Hillier, 2010).

La diferencia entre las distribuciones radica en el rango de variación de los tiempos de servicio, que podrá ir desde 0 (casos deterministas) hasta una variación muy grande (casos exponenciales); pero entre el amplio rango de estos dos extremos, por desgracia, se encuentra la mayor parte de casos reales para los que tendremos que usar distribuciones generales. Entre estos casos generales destaca la distribución Erlang, que, básicamente, es la suma de varias variables exponenciales $T_1, T_2, T_3 \dots, T_k$ siendo k el número de variables independientes exponenciales similares entre sí, y lo que indicará el grado de flexibilidad de nuestra distribución. Asimismo, observando su definición podemos afirmar que tanto las distribuciones deterministas como exponenciales no son más que casos específicos de la distribución Erlang, con $k = \infty$ y $k = 1$, respectivamente. Además de por esto, su importancia radica en que, al ser la suma de variables exponenciales, puede tomarse como la suma de k procesos; lo que aplica a servidores que realizan en el mismo servicio k procesos idénticos, como por ejemplo un taller mecánico que desmonte las ruedas de un coche, en el que lo componen cuatro procesos idénticos (desmontar una por una cada rueda del coche).

- Las disciplinas son los métodos de selección de los individuos para que reciban el servicio en el servidor correspondiente. Aunque la más común es la disciplina FIFO (First In First Out, básicamente que el primero que entra al sistema es el primero en recibir el servicio y salir de este) (Ferrer, Accessed: 05-09-2023).

hay muchas otras disciplinas. como la LIFO (Last In First Out, el último en llegar es el primero en recibir el servicio, por ejemplo un montón de papeles que deben ser firmados en los que el de arriba que llego el último es el primero que se firma), SIRO (Service In Random Order, cuando los individuos se seleccionan al azar, como por ejemplo un proceso de empaquetado de fruta, en la que entran grandes volúmenes de fruta y esta se va escogiendo al azar para su empaquetado), Priority (cuando algunos individuos tienen preferencia de selección, por ejemplo el servicio de urgencias en cualquier hospital) y existen tipos que pertenecen mayoritariamente a redes de colas, como el WFQ (Weighted Fair Queued, donde se clasifican en colas según importancia) que describiremos más adelante.

- El número de servidores (c) también influye directamente en la capacidad del sistema, y

dependerá tanto de factores limitantes como la naturaleza de la cola (una cola para recibir un autógrafa de x famoso solo puede tener un servidor, porque solo existe un servidor que pueda realizar el servicio), factores económicos o factores logísticos (por ejemplo, si en una fábrica solo entran dos máquinas que realicen el servicio no se podrá aumentar el número de servidores), como de factores de rendimiento (por ejemplo, si la tasa de utilización de un sistema es muy alta y no permite asegurar un estado estable del sistema, añadiendo más servidores esta disminuirá hasta reducirla por debajo de 1).

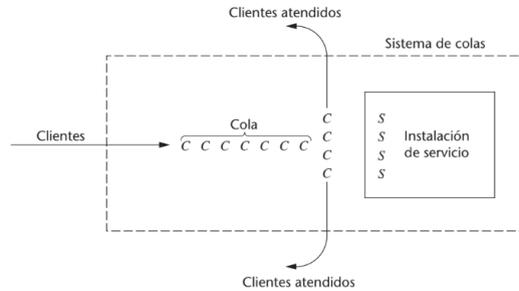


Figura 2: Ejemplo esquemático de una cola con varios servidores

- La capacidad del sistema influye de forma directa tanto en los tiempos de espera de las colas como del sistema. Por lo general, para la comodidad de los cálculos siempre que esta cota sea grande se tomará como infinita, pero en casos donde la saturación sea fácil de alcanzar se establecerá una cota fija llamada K . Para ajustar la capacidad de nuestro sistema es necesario recoger información cuando este esté en estado estable, que sucede, como se explica en Lieberman y Hillier (2010), una vez haya transcurrido un cierto periodo de tiempo y podamos asumir que la tasa de utilización $\left(\rho = \frac{\lambda}{c \cdot \mu}\right)$ es menor que uno, y esta nos será imprescindible para calcular la longitud esperada de la cola (L_q), que estima el número medio de individuos en la cola excluyendo el servicio, y el número de individuos esperados del sistema (L), definidas en el apartado de *Medidas de rendimiento*.
- Las medidas de rendimiento son valores que estiman el funcionamiento de un sistema de cola, calculando tiempos medios de espera tanto de cola como del sistema o número de individuos durante estas esperas. Fundamentalmente son cuatro: el número de individuos presentes en la espera de una cola (L_q) como en todo el sistema (L) mencionados anteriormente y los tiempos de espera medios tanto en la cola (W_q) como en el sistema completo (W). Además, dependiendo de la distribución que sigan los tiempos de servicio, se modificarán las fórmulas de estos. La fórmula de L será:

 - para tiempos Markovianos $L = \sum_{n=0}^K n \cdot P_n$
 - para tiempos Generales $L = \rho + L_q$
 - para tiempos Deterministas $L = \rho + L_q$
 - para tiempos Erlangianos $L = \lambda \cdot W$

Donde el sumatorio recorrerá n desde cero hasta la capacidad del sistema, y la probabilidad P_n es la probabilidad de que haya n individuos en el sistema en un instante.

De forma análoga la fórmula de L_q será:

- para tiempos Markovianos $L_q = \sum_{n=s}^K (n - s) \cdot P_n$

- para tiempos Generales $L_q = \frac{\lambda^2 \cdot \sigma^2 + \rho^2}{2 \cdot (1 - \rho)}$
- para tiempos Deterministas $L_q = \frac{\rho^2}{2 \cdot (1 - \rho)}$
- para tiempos Erlangianos $L_q = \frac{1 + k}{2 \cdot k} \cdot \frac{\lambda^2}{\mu \cdot (\mu - \lambda)}$

Donde el sumatorio recorrerá n desde c hasta la capacidad del sistema, la probabilidad se calculará de la misma forma, indicando n el número de individuos presentes en la cola (incluyendo a los presentes en los servidores).

Ahora, para calcular los tiempos de espera respectivos, aplicaremos la fórmula de Little, que dice:

$$L = \lambda \cdot W$$

por lo que despejando:

- para tiempos Markovianos $W = \frac{L}{\lambda} = W_q + \frac{1}{\mu}$
- para tiempos Generales $W = W_q + \frac{1}{\mu}$
- para tiempos Deterministas $W = W_q + \frac{1}{\mu}$
- para tiempos Erlangianos $W = W_q + \frac{1}{\mu}$

Además, estas fórmulas corroboran que:

$$L_q = \lambda \cdot W_q$$

Por lo que:

- para tiempos Markovianos $W_q = \frac{L_q}{\lambda}$
- para tiempos Generales $W_q = \frac{L_q}{\lambda}$
- para tiempos Deterministas $W_q = \frac{L_q}{\lambda}$
- para tiempos Erlangianos $W_q = \frac{1 + k}{2 \cdot k} \cdot \frac{\lambda}{\mu \cdot (\mu - \lambda)}$

Y con esto obtenemos los tiempos de espera en el sistema y en la cola, respectivamente. Además, también podemos considerar como medida de rendimiento al factor de utilización del sistema (ρ) ya que indica la proporción del tiempo que los servidores están ocupados, y nos ayuda a conocer qué tan alta es la saturación de nuestro sistema. Su fórmula, antes mencionada, es:

$$\rho = \frac{\lambda}{c \cdot \mu}$$

- Para poder dar una descripción rápida de las características de un sistema de cola, se creó la notación de Kendall, que se compone de seis parámetros $A|B|C|K|N|D$: A , para la distribución que sigue la tasa de llegadas (utilizando la letra M para procesos markovianos de Poisson, D para deterministas, E para distribuciones Erlang y G para una distribución general), B para la distribución de la tasa de servicios (que utiliza la misma notación que la tasa de llegadas), C para el número de servidores del sistema, K para la capacidad del sistema, N

para la disciplina utilizada y D para el número de colas en el sistema. Se suele reducir a los tres primeros parámetros, siendo por defecto $K = \infty$, $N = FIFO$ y $D = 1$, un sistema simple con capacidad infinita y utilizando la disciplina FIFO como método de selección. Así, las colas más usuales son $M|M|1$, una cola simple que sigue distribuciones de Poisson con un solo servidor, $M|M|c|K$ que tiene c servidores y capacidad K , por lo que un nuevo usuario es rechazado del sistema cuando ya está a su capacidad máxima, o $M|M|c$ donde la capacidad es ilimitada y varios usuarios pasan a ser servidos por varios servidores.

Las redes de colas surgen por la necesidad de adecuar los sistemas de colas al mundo real que nos ocupa, ya que con el desarrollo presentado hasta ahora no es suficiente, un sistema podría componerse de múltiples procesos (que podemos definir como colas) que se van enlazando unos con otros y que pueden depender o no del anterior.

Aunque las redes de colas parezcan algo futurista que solo aplica a procesos complejos, la realidad es que las redes de colas son una herramienta útil para sistemas que tenemos presentes en el día a día, como un control de seguridad en un aeropuerto (donde varias colas con varios servidores de revisión actúan a la vez, e incluso de la revisión general se puede entrar o no en una revisión de explosivos y/o drogas y además dependiendo del contenido se puede pasar a otra revisión específica del contenido de la maleta por líquidos o materiales punzantes) o una cadena de montaje que requiere de distintos procesos y selecciones.

Ahora una vez definidos los términos principales de este trabajo veremos su estructura:

Tras un primer apartado con los objetivos del trabajo, nuestro primer capítulo tratará de una definición minuciosa del concepto de red de colas, explicando cada una de sus partes y la definición axiomática; en particular, la definición y ejemplos del término “enrutado”.

Nuestro segundo capítulo hablará de los tipos de redes de colas, diferenciando entre redes abiertas y cerradas con sus respectivos subtipos.

Nuestros siguientes capítulos tratarán de un tipo específico y muy importante de redes, que son las redes de Jackson, en los que definiremos para ambos tipos (abiertas y cerradas) sus ecuaciones de tráfico, condiciones de no saturación, etc. Daremos ejemplos numéricos de ambas y para el caso de las redes de Jackson cerradas definiremos el fenómeno de cuello de botella, y aplicaremos un programa en R que resolverá dicho tipo de redes.

Por último, se ofrecerán las conclusiones extraídas del trabajo, la bibliografía utilizada para su realización, donde encontraremos diversos libros, artículos y páginas web, y un anexo con el código del programa en R.

Objetivos

En este trabajo tendremos los siguientes objetivos:

- Dar a conocer ampliamente qué son las redes de colas y por qué surgieron.
- Justificar y convencer por qué las redes de colas son una herramienta útil y fundamental en el panorama actual, por lo que su desarrollo en investigación debería ser mayor.
- Desarrollar en concreto un tipo muy útil de redes, las redes de Jackson, definiéndolas, desarrollando sus partes, teoremas y aportando ejemplos numéricos.
- Programar un código en R que permita resolver redes de Jackson cerradas con el algoritmo MVA (Mean-Value Analysis), para que su uso en grandes sistemas de datos sea fácil y viable.

Capítulo 1

Redes de colas

1.1. ¿Qué son las redes de colas?

Una red de colas, como se muestra en Gross et al. (1998), no es más que un conjunto de estaciones diferenciadas, llamadas nodos, que aplican servicios de distinto tipo, pudiendo tener diferente número de servidores para cada nodo existente. Además, de forma general, en estas redes los individuos entran y salen de cualquier nodo, pudiendo repetir en cualquiera de ellos, omitirlos o seguir cualquier ruta aleatoria o determinada.

Una vez definido el concepto de cola, pasamos a ver sus partes fundamentales:

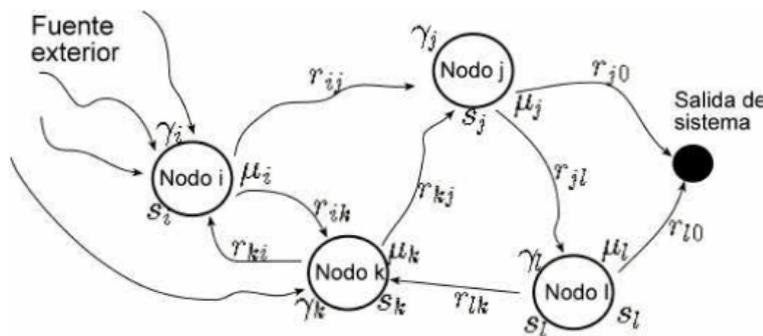


Figura 1: Redes de colas.

Figura 1.1: Elementos de una red de colas simple.

Apoyándonos en las partes descritas en el capítulo de *Introducción* de una única cola, observamos en la figura 1.1 que una red simple consta de:

- La población de entrada. De igual forma que en el sistema de una cola, la conforman los individuos que pueden entrar desde el exterior. Por lo general, para simplificar los cálculos, se considera infinita cuando así lo es o es muy elevada; pero existen casos en los que esta es finita, debido a limitaciones externas o que el propio sistema de red limite la aparición de nuevos individuos. Además, deberemos definir si nuestra población de entrada puede entrar al sistema por cualquiera de las colas presentes en el sistema (como por ejemplo cuando tú eliges en qué cola del supermercado te pones) o debe entrar por una cola determinada (como

cuando al entrar en un parque de atracciones debes entrar por la cola en la que revisan tu entrada y no puedes acceder directamente posicionándote en una de las colas de las atracciones).

- Los tiempos de llegada siguen siendo una pieza clave para los rendimientos en el sistema. Lo ideal será que podamos asumir que los tiempos de llegada al sistema siguen todos la misma distribución, y que, por lo tanto, son todos equivalentes. Esto sucede en los sistemas de colas infinitas que veremos más adelante, lo que simplifica mucho los cálculos. Un ejemplo de estas puede ser nuestro ejemplo del parque de atracciones, que en condiciones ideales a priori todas las atracciones deben tener las mismas tasas de llegada, ya que todas tienen la misma probabilidad de ser escogidas por los individuos, aunque en la realidad siempre habrá atracciones que tengan más afluencia que otras y no sean iguales. Aún así, siempre existirán sistemas en los que no todos los tiempos de llegada a los diferentes nodos sean equivalentes o incluso ni sigan la misma distribución; en estos casos deberemos prestar mucha atención a las diferencias a la hora de calcular medidas de rendimiento o al clasificar estos mismos. Un ejemplo de sistema con tiempos de llegada diferentes puede ser un sistema de renovación de documentos, en el que haya una cola para consultas, con llegada exponencial, y otra para las renovaciones, con cita previa que siga una tasa de llegada determinista.
- Los tiempos de servicio en las redes de colas conforman el mayor problema a la hora de calcular nuestras medidas de rendimiento, puesto que como en muchas ocasiones no seguirán las mismas distribuciones, no se podrá utilizar una fórmula generalizada para estimar los tiempos medios de espera en las colas y sistemas, y además tampoco serán independientes, lo que no nos permitirá utilizar la propiedad de equivalencia ni el teorema de Burke (2.1), que definiremos más adelante, para calcularlas por separado.
- Las disciplinas, como definimos en el capítulo de *Introducción*, son los métodos de selección de los individuos para que reciban el servicio en el servidor correspondiente. Para una red se seguirán usando los mismos tipos: FIFO (First In First Out), LIFO (Last In First Out), SIRO (Service In Random Order), Priority... que se usarán tanto unificados (en redes en las cuales las colas tengan procesos similares, como un súper, que todas las colas serán FIFO) o mezclados (para redes donde unos procesos se distinguen de otros, como en una cadena de limpieza y reciclaje, donde los residuos se seleccionan al azar y luego van pasando por los diferentes procesos mediante una disciplina FIFO). Por último, existen unas disciplinas específicas de redes que mencionamos anteriormente, como la WFQ (Weighted Fair Queued), donde los individuos se clasifican en colas según importancia y el sistema va dando servicio a estas colas según un algoritmo RR (Round Robin) ???. Con él, se va ajustando cada disciplina de la cola al servicio y a la longitud de la cola, y va asignando más capacidad a las colas más largas sin penalizar a las más libres.
- Como ya sabemos, la capacidad influye de forma directa tanto en los tiempos de espera de las colas como del sistema. Las capacidades de nuestro sistema y de nuestras colas se definirán de forma análoga a los métodos para una única cola, pero aquí cobrarán una importancia adicional, ya que dependiendo de la capacidad de nuestras colas, nuestros sistemas podrán simplificarse o no, porque como veremos en el siguiente capítulo de *Tipos de redes*, si nuestras colas son infinitas, del tipo $M|M|c$, el sistema se podrá simplificar como un sistema de colas independientes, y los cálculos se acortarán por la *propiedad de equivalencia* ??.

- Como ya vimos, las medidas de rendimiento determinan la eficiencia del sistema, y sus fórmulas son:

- para tiempos Markovianos:

$$L = \sum_{n=0}^K n \cdot P_n$$

$$L_q = \sum_{n=s}^K (n - s) \cdot P_n$$

$$W = \frac{L}{\lambda}$$

$$W_q = \frac{L_q}{\lambda}$$

- para tiempos Generales:

$$L = \rho + L_q$$

$$L_q = \frac{\lambda^2 \cdot \sigma^2 + \rho^2}{2 \cdot (1 - \rho)}$$

$$W = W_q + \frac{1}{\mu}$$

$$W_q = \frac{L_q}{\lambda}$$

- para tiempos Deterministas:

$$L = \rho + L_q$$

$$L_q = \frac{\rho^2}{2 \cdot (1 - \rho)}$$

$$W = W_q + \frac{1}{\mu}$$

$$W_q = \frac{L_q}{\lambda}$$

- para tiempos Erlangianos:

$$L = \lambda \cdot W$$

$$L_q = \frac{1 + k}{2 \cdot k} \cdot \frac{\lambda^2}{\mu \cdot (\mu - \lambda)}$$

$$W = W_q + \frac{1}{\mu}$$

$$W_q = \frac{1 + k}{2 \cdot k} \cdot \frac{\lambda}{\mu \cdot (\mu - \lambda)}$$

con

$$\rho = \frac{\lambda}{c \cdot \mu}$$

Teniendo estas fórmulas, bastará con definir nuestro tipo de red para adaptar las medidas de rendimiento a ella. Si en nuestra red es obligatorio pasar por todas las colas presentes (como

puede ser en un proceso de montaje), para calcular los tiempos de espera en la cola o en cada nodo bastará con sumar los respectivos tiempos de cada cola o de todos los servicios:

$$L_{red} = \sum_{n=1}^K L_n$$

$$L_{q_{red}} = \sum_{n=1}^K L_{q_n}$$

$$W_{red} = \sum_{n=1}^K W_n$$

$$W_{q_{red}} = \sum_{n=1}^K W_{q_n}$$

siendo K el número de colas existentes en el sistema.

En caso de que no sea necesario pasar por todas las colas, se hará una suma ponderada en función del factor de utilización de cada cola, para determinar cuál es su peso en los tiempos de espera y longitudes de cola.

- El número de servidores influye de la misma manera que en los sistemas de una cola. Aún así, en el caso de los sistemas de redes, nos será útil siempre que se pueda utilizar nodos de un solo servidor para facilitar los cálculos en casos de tiempos de espera con distribuciones diferentes a Poisson. Si los sistemas son Markovianos el número de servidores no será gran problema porque funcionarán como colas independientes. Añadir más de un servidor en los nodos de un sistema de red de colas puede ser muy útil en casos como los sistemas en tándem (que veremos en el próximo capítulo), ya que, como la masificación de una cola puede conllevar atrasos en los posteriores procesos, la adición de otro servidor en un nodo específico puede ayudar a aliviar y regular el tránsito.
- En el caso de las redes de colas, se seguirá utilizando la notación de Kendall como viene descrita en el apartado anterior, que se compone de seis parámetros: $A|B|C|K|N|D$. Lo único es que se especificará por separado cómo son cada una de las colas de nuestro sistema utilizando esta notación, ya que algunos sistemas tendrán todas sus colas iguales (como el caso de un supermercado) o unas serán diferentes a otras (como el caso de una selección de reciclaje, que la primera cola será por ejemplo $M|M|1|\infty|S|IRO|1$ y las demás serán por ejemplo $M|M|1|\infty|FIFO|\infty$).

1.2. El enrutado

El enrutado es el proceso de, elegido un individuo que entra en el sistema, cómo recorre el propio sistema. Es decir, la forma de pasar de un nodo a otro hasta salir del sistema (o no, en el caso de sistemas cerrados). El enrutado es clave en temas de optimización del tiempo de espera y de salida en el sistema, por lo que encontrar la ruta adecuada es imprescindible en el estudio de una red.

Existen infinitud de tipos de enrutamiento, ya sea definido por el individuo (en el que el individuo escoge entrar o no en determinados nodos; como puede ser el ejemplo de un parque de atracciones,

en el que el usuario elige en qué atracciones montarse, el orden, si repetir o no. . .), una ruta única (donde todos los individuos que entran en el sistema siguen una ruta fija preestablecida en la que se pasará de un nodo a otro en orden hasta salir del sistema, o recorrerán el sistema de esa forma infinitamente en caso de sistemas cerrados), por políticas (en el que los individuos serán distribuidos en unas u otras colas y seguirán unas rutas u otras en función de características o aspectos discriminantes (Enrutamiento, 2020); como por ejemplo el enrutamiento en telecomunicaciones, que precisa de políticas concretas de la distribución de la información para que determinados paquetes de información sigan x caminos por cuestiones administrativas), o de forma aleatoria (como por ejemplo un sistema de control de calidad, en el que aleatoriamente unos individuos entran en nodos para que se realice el servicio, que será la inspección de calidad).

En este último caso de un enrutado aleatorio, para la selección de la ruta de los individuos se podrán utilizar métodos probabilísticos comunes, utilizando probabilidades sencillas constantes en el tiempo que seguirán distribuciones establecidas, como la de Bernoulli o la Normal (como es el caso del control del aeropuerto, en el que el individuo entrará en la revisión de explosivos aleatoriamente siguiendo una probabilidad fija con una distribución de Bernoulli o Poisson ya que la probabilidad puede ser muy pequeña), o puede utilizar algoritmos probabilísticos más complejos, en los que las probabilidades varíen y las rutas dependan directamente del tráfico del servidor, como es el ejemplo del algoritmo RED (Random Early Drop), desarrollado por Escalante (2023), que se usa en redes digitales para evitar el congestionamiento de los mecanismos de distribución de la información. Para ello, descarta aleatoriamente paquetes de datos que desean entrar en el sistema en función de la proximidad al congestionamiento que haya en nuestro sistema, con el fin de no descartar conjunta y seguidamente paquetes de manera indiscriminada. Este funciona de la siguiente manera:

Algoritmo RED

- Primero, este algoritmo necesita unos umbrales, mínimo y máximo, entre los que nuestro sistema deberá trabajar. Por debajo del umbral mínimo el algoritmo no funcionará, ni por encima del máximo, ya que el sistema estará colapsado y rechazará paquetes linealmente y no mediante el algoritmo.
- Una vez definidos estos, el algoritmo analizará el tamaño de las colas de nuestros nodos, si están dentro de los umbrales, este calculará una probabilidad concreta de descarte, y variará según fluctúen las colas, utilizando la función de probabilidad del Probabilistic Packet Drop.

Aunque este algoritmo se use mayoritariamente en colas separadas para regular los congestionamientos, también se puede aplicar a redes de colas para calcular enrutamientos en los que, cuando una cola se satura, se envíen estos paquetes a otras colas y el recorrido del sistema se mantenga estable.

Capítulo 2

Tipos de redes

Las redes de colas se pueden clasificar de diversas formas, enfocándose en unos aspectos u otros, como por los tiempos de entrada al sistema (sigan una distribución de Poisson, que es lo habitual, o sigan otros procesos limitados, como los tiempos de entrada establecidos para la renovación de un documento oficial, que precisa de cita previa), por su sistema de entrada y salida del sistema (si es posible entrar y salir del sistema o no) o por la capacidad de sus colas (si puede aceptar un número infinito de individuos o finito).

Este último posee una característica especial, ya que en el caso de capacidades infinitas, por la propiedad de la equivalencia descrita en el libro de Lieberman y Hillier (2010), en modelos de colas iguales con c servidores, tiempos de llegada y de espera λ y W respectivamente, si $c \cdot W > \lambda$ podemos simplificar todas las colas como un modelo $M|M|c$ y, por lo tanto, las probabilidades de capacidad de todo el sistema se obtienen con el producto de las probabilidades respectivas en cada cola, y el tiempo de espera en todo el sistema sumando los distintos tiempos de cada cola.

Por desgracia, en general, las colas suelen ser finitas, en las que los resultados no son tan sencillos y su cálculo es más limitado.

El tipo más abundante en los sistemas de redes son las llamadas redes de Jackson, las cuales son sistemas de redes de colas infinitas que cumplen unas características específicas definidas por su creador, el matemático Jackson (Jackson, 1963). Las características fundamentales, así como el teorema que respalda sus dos tipos (abiertas y cerradas) y sus medidas de rendimiento las desarrollaremos más detalladamente en el siguiente capítulo.

También están las colas tándem, que como se desarrolla en el pdf de de la Fuente-Fernández (2020) son redes en las que el individuo debe visitar seguidamente x nodos antes de salir del sistema. Un ejemplo de ello puede ser un restaurante de autoservicio, en el que vas pasando por diferentes estaciones para que se les asigne su plato y finalmente pagar. Para comprender bien el funcionamiento de estas, desarrollaremos numéricamente el ejemplo anterior utilizando los datos de los apuntes de de la Fuente-Fernández (2020). El enunciado será el siguiente:

“Un autoservicio dispone de tres empleados, un camarero sirve el primer plato, el segundo camarero sirve el segundo plato y el tercero se encarga de la caja. Los dos primeros camareros disponen de suficiente espacio para atender a clientes sin limitación, mientras que el tercer camarero tiene una longitud máxima de cola de dos personas. El autoservicio, modelado como red, muestra que

la tasa media de llegadas a la hora de la comida es de 54 clientes/hora, el primer camarero tiene un tiempo medio de servicio de un minuto y el segundo camarero de treinta segundos. Se solicita:

- Valor máximo del tiempo de servicio del tercer camarero para que su trabajo no interrumpa al de sus compañeros.
- Longitud de las colas que forman el sistema.
- Tiempo medio que un cliente pasa en el autoservicio desde que llega hasta que sale dispuesto para comer.”

Para comenzar, utilizando el Teorema de Burke (Burke, 1956), que dice:

“Para toda cola $M|M|c$, con $c \in [1, \infty)$ y tasa de llegadas λ , la tasa de salidas también seguirá una distribución de Poisson con parámetro λ ”,

(2.1)

como todo son colas $M|M|1$, podemos analizarlas de forma independiente, por lo que los tiempos de espera y probabilidades serán simplemente las sumas y productos de las diferentes colas. Además, como consecuencia del teorema de Burke (2.1), podríamos obtener que cualquier $P(n)$, con n =número de individuos en el sistema, se calculará sumando el producto de las probabilidades específicas de que haya n_i individuos en el nodo i por su respectivo número de individuos, quedando:

$$P(n) = \sum_{i=1}^K n_i \cdot P(n_i).$$

Para nuestro problema, utilizando la definición de L_q para colas Markovianas:

$$L_q = \frac{\lambda^2}{\mu \cdot (\mu - \lambda)} = \frac{\rho^2}{1 - \rho}$$

con esto, para el apartado a), bastará con pasar nuestra tasa de llegadas λ a minutos para tener todas nuestras variables en las mismas unidades, y aplicar las fórmulas de una cola markoviana descritas en el anterior capítulo:

$$\lambda = 54 \text{ clientes/hora} = 0.9 \text{ clientes/minuto.}$$

Primero calculamos el factor de utilización mediante la restricción $L_{q3} = 2$ clientes, que despejando nos queda:

$$\rho_3^2 - (1 - \rho_3) \cdot 2 = 0 \Rightarrow \rho_3 = 0.7321$$

Calcularemos μ_3 despejando de la definición del factor de utilización, que hemos calculado previamente, y nos queda:

$$\mu_3 = \frac{\lambda}{\rho_3} = 1.2294 \text{ clientes/minuto} \Rightarrow 1/\mu_3 = 0.8134 \text{ minutos/cliente} = 48.8034 \text{ segundos/cliente.}$$

Para el apartado b), necesitaremos saber primero las tasas de servicio μ_i en los demás nodos, por lo que, conociendo los tiempos de servicio de los nodos 1 y 2, que es el inverso de μ_i , tenemos que:

$$\mu_1 = \frac{1}{1 \text{ minuto/cliente}} = 1 \text{ cliente/minuto};$$

$$\mu_2 = \frac{1}{0.5 \text{ minutos/cliente}} = 2 \text{ clientes/minuto}.$$

Por lo que usando la fórmula de $L = L_q + \frac{\lambda}{\mu} = \frac{\lambda}{\mu - \lambda}$:

$$L_1 = \frac{0.9}{1 - 0.9} = 9 \text{ clientes},$$

$$L_2 = \frac{0.9}{2 - 0.9} = 0.81 \text{ clientes},$$

$$L_3 = 2 + \frac{0.9}{1.2294} = 2.7321 \text{ clientes}.$$

Para terminar, para el apartado c), bastará con utilizar la fórmula de Little y:

$$W = \frac{L}{\lambda} = \frac{L_1 + L_2 + L_3}{\lambda} = 13.9447 \text{ minutos}.$$

Sabiendo ya la cantidad de formas que existen de clasificar las redes, y habiendo definido brevemente algunas, de cara al siguiente capítulo, en estas secciones nosotros nos centraremos en clasificar las redes de colas según su entrada: abiertas y cerradas.

2.1. Redes abiertas

Las redes de colas abiertas son aquellas en las que los individuos pueden entrar y salir del sistema. Unos ejemplos claros pueden ser un supermercado o un parque de atracciones, en los que se pueden considerar colas infinitas y los individuos van entrando y saliendo con libertad.

Para su resolución, es necesario determinar la forma de llegada de los individuos, con las tasas de llegada siguiendo una distribución de Poisson, y los métodos de selección para la entrada a la i -ésima cola, que seguirán métodos probabilísticos.

Por último se definirá el enrutado de la red, dependiendo de los fines de esta: el tráfico, la cantidad de servidores, un enrutado probabilístico, uno por ruta más corta, uno cíclico, uno por políticas. . .

Dentro de las redes abiertas podemos volver a clasificarlas en cíclicas o acíclicas (de la Fuente-Fernández, 2020):

- **Cíclicas:** Cuando un individuo puede volver a entrar en alguna de las colas por las que ya ha pasado. Esto es útil en procesos con posibilidad de fallo en la realización del servicio, con independencia de la ruta seguida por el individuo.
Ejemplos claros de este tipo son: los procesos de escáner, como el ejemplo del control de seguridad de los aeropuertos, en el que pueden hacerte revisar la maleta cuantas veces haga falta; o el ejemplo del parque de atracciones, donde puedes repetir cola en x atracción cuantas veces desees. El siguiente esquema (2.1) ejemplifica a la perfección este tipo:
- **Acíclicas:** Cuando no se puede volver a entrar en colas por las que ya se ha pasado. Este tipo es necesario en procesos encadenados en los cuales la ruta a seguir es unidireccional, pues

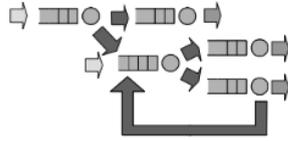


Figura 2.1: Ejemplo de una red abierta cíclica (fuente: apuntes de la profesora Rivas-López (2013)).

no pueden repetirse al azar las estaciones, o donde este proceso sea costoso. Ejemplos de este tipo las son cadenas de montaje, en donde un artículo no puede pasar dos veces por el mismo proceso, ya que conlleva a fallos y pérdidas de tiempo y dinero. De nuevo, apoyándonos en el siguiente esquema (2.2):

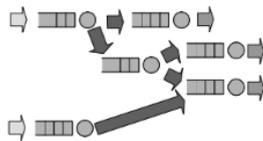


Figura 2.2: Ejemplo de una red abierta acíclica (Rivas-López, 2013).

2.2. Redes cerradas

Las redes cerradas son aquellas en las que no pueden entrar ni salir nuevos individuos. Esto hace que los individuos que pertenecen al sistema lo recorran indefinidamente. Un ejemplo claro de estas son los sistemas de reparación de un conjunto de maquinaria o de forma ideal el agua de una depuradora. Un ejemplo visual lo encontramos en el siguiente esquema (2.3):

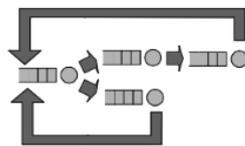


Figura 2.3: Ejemplo de una red cerrada (Rivas-López, 2013).

Capítulo 3

Redes de Jackson

Las redes de Jackson son un tipo de redes de cola infinita en las que los individuos no tienen por qué seguir el orden establecido en las colas infinitas, pueden omitir nodos, y cada individuo puede llegar a cada nodo desde otra instalación o desde afuera en el caso de ser redes abiertas.

Cabe resaltar que James R. Jackson se inspiró en el teorema de Burke (2.1), y aplica los mismos conceptos en su teorema pero especializándose en un tipo concreto de redes, que están presentes en múltiples ámbitos de la vida cotidiana, como el sector en el que Jackson trabaja, fabricación en cola y planeamientos de sistemas (Jackson, 1963).

Dentro de las redes de Jackson, como se muestra en el libro de Gross et al. (1998), durante los años ha habido ampliaciones de casos, como las redes de Jackson con dependencia entre nodos y tiempos de servicio, que, resumidamente, se convertían a una forma parecida de redes de nacimiento y muerte, con sistemas de ecuaciones mucho más complejas que, por desgracia, no se conseguían resolver de una manera eficiente.

Otro tipo de ampliaciones son las redes de Jackson con tiempo de viaje entre servicios, que consideraba los tiempos de trayecto como otra variable con distribuciones de Poisson, o como otro nodo, en el cual se tomaban los tiempos medios para los cálculos, ya que la distribución era indiferente en el estudio, porque los nodos no se veían incluidos en las distribuciones marginales de interés, como relatan Posner y Bernholtz (1968).

Pero, sin duda, el tema más importante y en el que más se ha investigado de las extensiones de las redes de Jackson son las redes de Jackson multiclase. Estas redes contemplan sistemas en los que los individuos tienen diferentes clases, y con ello los tiempos de llegada y de servicio en los diferentes nodos i cambiarán. Estos tipos de redes admiten tres tipos de disciplinas (Processor sharing, amplio servicio y LIFO con servicio de reanudación preventiva) en los que las redes podrán admitir o no a los individuos según su clase, y estos podrán cambiar de clase una vez se les aplica el servicio en un nodo i cualquiera.

Además, gracias al trabajo de muchos matemáticos, como Kelly (1975) en varias de sus publicaciones, conformaron una estructura de notación para redes multiclase de único tiempo de servicio y disciplina FIFO (que podía extenderse a las diferentes disciplinas) especificando los estados de cada nodo. O el posterior trabajo de Baskett et al. (1975) y de Barbour (1976), que modelizaron las conjeturas de Kelly extendiendo su modelo a redes multiclase con tiempos de servicio genera-

les gracias a aproximaciones de la red con los cálculos de una red general descompuesta, lo cual permitió que Gross y Ince (1981) diesen un resultado numérico al típico problema de reparación de máquinas como una red de Jackson cerrada.

Hoy en día el campo de las redes cerradas de Jackson multiclase se ha podido extender y sigue extendiéndose día a día por su importancia en computación, comunicación y logística.

Como se menciona en el libro de Zukerman (2013), para que una red de colas se pueda clasificar como una red de Jackson, esta debe cumplir ciertos requisitos, dependiendo de si hablamos de redes de Jackson abiertas o cerradas, que son:

Condiciones de redes abiertas

- Todos los servicios deben ser iguales. Esto implica que todos sus tiempos de llegada deben estar distribuidos exponencialmente de la misma manera.
- Todos los nodos serán de la siguiente forma: $M|M|c_i|\infty|FIFO|1$.
Por tanto, podemos ver claramente que cumplirán el Teorema de Burke (2.1) y los diferentes nodos se podrán considerar independientes entre sí, así como afirmar que los tiempos de llegada cuando el individuo proviene de otra cola sigan una distribución de Poisson.
- Todos los nodos deben estar en un estado estable; es decir, su factor de utilización debe ser menor que 1, $\rho_i < 1$, siendo $i = 1, \dots, K$ los diferentes nodos del sistema.
- Al menos un nodo de los K nodos presentes en nuestro sistema debe ser abierto, es decir, que permita la entrada de individuos desde el exterior al sistema, teniendo ese nodo i una tasa de llegadas desde el exterior $\gamma_i > 0$.
Pueden darse sistemas donde algunos o la mayoría de los nodos tengan su probabilidad de entrada nula, pero siempre deberá existir al menos uno en el cual esta sea no nula.
- Un individuo se moverá de una cola i a otra cola j con probabilidad r_{ij} , o saldrá del sistema con probabilidad $r_{i0} = 1 - \sum_{j=1}^K r_{ij}$, donde el sumatorio de las probabilidades recorre todos los estados restantes, que se compondrá de los diferentes cambios a colas distintas y la re-entrada en el mismo nodo del que sale el individuo.
De forma análoga al punto anterior, esta probabilidad r_{i0} puede ser nula en algunos o la mayoría de nodos del sistema, pero debe existir al menos uno en el que sea no nula para que el sistema se pueda definir como abierto.

En el caso de las redes cerradas, muchas condiciones se exigirán de la misma forma, pero por el diseño de la propia red algunos variarán, quedando de la siguiente forma:

Condiciones de redes cerradas

- Todos los servicios deben ser iguales. Esto implica que todos sus tiempos de llegada deben estar distribuidos exponencialmente de la misma manera.
- Todos los nodos serán de la siguiente forma: $M|M|c_i|\infty|FIFO|1$.
Por tanto, podemos ver claramente que cumplirán el Teorema de Burke (2.1) y los diferentes nodos se podrán considerar independientes entre sí, así como afirmar que los tiempos de llegada cuando el individuo proviene de otra cola sigan una distribución de Poisson.

- Todos los nodos deben estar en un estado estable; es decir, su factor de utilización debe ser menor que 1, $\rho_i < 1$, siendo $i = 1, \dots, K$ los diferentes nodos del sistema.
- El paso de un nodo i a otro nodo j tendrá una probabilidad r_{ij} , para las cuales $\sum_{j=1}^K r_{ij} = 1 \quad \forall i = 1 \dots, K$. Existe la posibilidad de que $i = j$, que definirá el concepto de bucle en el sistema de red de colas (como por ejemplo cuando una máquina en un sistema de reparaciones no cumpla los estándares y se le vuelva a aplicar el trabajo inmediatamente), o de que $r_{ij} = 0$, que definirá en algunos casos rutas inexistentes para los individuos (por ejemplo, un sistema en el que los bucles no estén contemplados).

Podemos observar que los tres primeros puntos se cumplen de la misma manera, y que solo cambian los dos últimos. Por lo tanto, tomaremos estos tres primeros como características generales de las redes de Jackson, y los otros dos como características específicas de los tipos, ya que solo afectan a su naturaleza de entrada.

3.1. Redes abiertas de Jackson

Las redes abiertas de Jackson tienen una gran importancia, como las cerradas, en sistemas de comunicaciones y ordenadores.

En este apartado veremos las ecuaciones de tráfico de la red, sus condiciones de no saturación, el teorema de Jackson que respalda todo este tipo de colas único, y sus consecuencias, así como las medidas de rendimiento del sistema, y ejemplos numéricos que nos ayudarán a visualizar el funcionamiento de estas redes.

3.1.1. Ecuaciones de tráfico

Las ecuaciones de tráfico hacen referencia al cálculo de las tasas de llegada a cada nodo i . Se componen de un sistema de ecuaciones que consideran tanto las llegadas del exterior como de otros nodos de dentro del sistema, quedando de la siguiente manera:

$$\lambda_i = \gamma_i + \sum_{j=1}^K \lambda_j \cdot r_{ji} \quad (3.1)$$

Las cuales tienen un sistema de estados:

$$P_{n_1, \dots, n_k} = (1 - \rho_1) \cdot \rho_1^{n_1} \cdot \dots \cdot (1 - \rho_K) \cdot \rho_K^{n_K} \quad (3.2)$$

Que se puede ver como un producto de marginales de colas $M|M|1$ independientes con tasa de llegadas λ_i y tasa de servicio μ_i . Aún así, no se debe confundir como que el sistema se puede descomponer como K procesos de Poisson de una cola $M|M|1$ ya que aunque en la práctica funciona así, en la teoría estos procesos internos no siguen distribuciones de Poisson, como es el caso de los bucles.

Además, como se nombra en el libro de Gross et al. (1998), estas ecuaciones de tráfico y sistema de estados no son más que una versión restrictiva de la ecuación de balance global en la que, par-

tiendo de los siguientes estados:

$$n_1, \dots, n_i, n_j, \dots, n_K \equiv \bar{n}$$

$$n_1, \dots, n_i - 1, n_j, \dots, n_K \equiv \bar{n}; i^-$$

$$n_1, \dots, n_i + 1, n_j, \dots, n_K \equiv \bar{n}; i^+$$

$$n_1, \dots, n_1 + 1, n_j - 1, \dots, n_K \equiv \bar{n}; i^+ j^-$$

tenemos que para un sistema con un servidor y uno o más individuos en cada nodo, la ecuación de paso del estado n a cualquier otro se compone por:

$$\sum_{i=1}^K \gamma_i \cdot P_{\bar{n}; i^-} + \sum_{j=1}^K \sum_{\substack{i=1 \\ i \neq j}}^K \mu_i \cdot r_{ij} \cdot P_{\bar{n}; i^+ j^-} + \sum_{i=1}^K \mu_i \cdot r_{i0} \cdot P_{\bar{n}; i^+} = \sum_{i=1}^K \mu_i \cdot (1 - r_{ii}) \cdot P_{\bar{n}} + \sum_{i=1}^K \gamma_i \cdot P_{\bar{n}} \quad (3.3)$$

Del que se obtiene una ecuación de estados global de la siguiente manera:

$$P_{\bar{n}} = C \cdot \rho_1^{n_1} \cdot \rho_2^{n_2} \cdots \rho_K^{n_K} \quad (3.4)$$

Como se describe en Gross et al. (1998), la utilidad de la ecuación 3.3 es que da lugar a un conjunto de ecuaciones de la forma

$$\lambda_i \cdot P_{\bar{n}; i^-} = \mu_i \cdot P_{\bar{n}},$$

que lo que nos indican es que es equivalente el paso de un estado al otro y viceversa; lo que nos induce a métodos lineales diferenciales que conducen a la forma de distribución conjunta que vemos en la ecuación 3.2.

Para comprobar que esto es cierto, basta con demostrar la relación de la ecuación 3.2 con 3.4.

Para ello, primero demostraremos que la ecuación 3.4 cumple la ecuación 3.3 y a partir de ahí bastará con demostrar que para el caso de la ecuación 3.2 nuestra constante C de la otra ecuación vale $C = \prod_{i=1}^K (1 - \rho_i)$

El primer paso para demostrar la primera equivalencia será nombrar a $R^{\bar{n}} = \rho_1^{n_1} \cdot \rho_2^{n_2} \cdots \rho_K^{n_K}$, despejamos C de la ecuación 3.4 y agrupamos en 3.3, $C \cdot R^{\bar{n}}$:

$$C \cdot R^{\bar{n}} \cdot \sum_{i=1}^K \frac{\gamma_i}{\rho_i} + C \cdot R^{\bar{n}} \cdot \sum_{j=1}^K \sum_{\substack{i=1 \\ i \neq j}}^K \mu_i \cdot r_{ij} \cdot \frac{\rho_i}{\rho_j} + C \cdot R^{\bar{n}} \cdot \sum_{i=1}^K \mu_i \cdot r_{i0} \cdot \rho_i \stackrel{?}{=} C \cdot R^{\bar{n}} \cdot \sum_{i=1}^K \mu_i \cdot (1 - r_{ii}) + C \cdot R^{\bar{n}} \cdot \sum_{i=1}^K \gamma_i \quad (3.5)$$

Ahora, eliminando $C \cdot R^{\bar{n}}$ y expresando $\rho_i = \frac{\lambda_i}{\mu_i}$ llegamos al siguiente supuesto:

$$\sum_{i=1}^K \frac{\gamma_i \cdot \mu_i}{\lambda_i} + \sum_{j=1}^K \sum_{\substack{i=1 \\ i \neq j}}^K \mu_i \cdot r_{ij} \cdot \frac{\lambda_i \cdot \mu_j}{\lambda_j \cdot \mu_i} + \sum_{i=1}^K \mu_i \cdot r_{i0} \cdot \frac{\lambda_i}{\mu_i} \stackrel{?}{=} \sum_{i=1}^K (\mu_i - \mu_i \cdot r_{ii} + \gamma_i) \quad (3.6)$$

Una vez obtenido el supuesto, para resolverlo deberemos reescribir utilizando la ecuación 3.6, aplicando al nodo j en vez del nodo i por cuestiones de notación, la igualdad:

$$\sum_{\substack{i=1 \\ i \neq j}}^K \lambda_i \cdot r_{ij} = -\lambda_j \cdot r_{jj} + \lambda_j - \gamma_j \quad (3.7)$$

y nos quedaría:

$$\sum_{i=1}^K \frac{\gamma_i \cdot \mu_i}{\lambda_i} + \sum_{j=1}^K \frac{\mu_j}{\lambda_j} \cdot (-\lambda_j \cdot r_{jj} + \lambda_j - \gamma_j) + \sum_{i=1}^K \mu_i \cdot r_{i0} \cdot \frac{\lambda_i}{\mu_i} \stackrel{?}{=} \sum_{i=1}^K (\mu_i - \mu_i \cdot r_{ii} + \gamma_i) \quad (3.8)$$

Cambiando j por i en el segundo sumatorio y reagrupando obtenemos:

$$\sum_{i=1}^K \left(\frac{\gamma_i \cdot \mu_i}{\lambda_i} + \frac{\mu_i}{\lambda_i} \cdot (-\lambda_i \cdot r_{ii} + \lambda_i - \gamma_i) + r_{i0} \cdot \lambda_i \right) \stackrel{?}{=} \sum_{i=1}^K (\mu_i - \mu_i \cdot r_{ii} + \gamma_i) \quad (3.9)$$

lo que, despejando y anulando términos queda:

$$\sum_{i=1}^K r_{i0} \cdot \lambda_i = \sum_{i=1}^K \gamma_i \quad (3.10)$$

Y este supuesto se puede corroborar ya que la parte de la izquierda hace referencia al flujo total

de salida de la red y la parte derecha al flujo total de entradas, que en este tipo de redes deben ser iguales.

Con esto demostrado, ahora sólo queda comprobar que $C = \prod_{i=1}^K (1 - \rho_i)$ que es fácilmente obtenible utilizando

$$\sum_{n_1=0}^{\infty} \cdots \sum_{n_K=0}^{\infty} C \cdot \rho_1^{n_1} \cdot \rho_K^{n_K} = 1$$

que se deduce pensando en que la probabilidad de la combinación de todos los posibles estados debe ser la unidad. Por lo que, despejando C :

$$C = \left(\sum_{n_1=0}^{\infty} \cdots \sum_{n_K=0}^{\infty} \rho_1^{n_1} \cdot \rho_K^{n_K} \right)^{-1} \quad (3.11)$$

Esto, aplicando la igualdad

$$\sum_{n_i=0}^{\infty} \rho_i^{n_i} = \frac{1}{1 - \rho_i}$$

podemos simplificar:

$$C = \left(\frac{1}{(1 - \rho_1) \cdots (1 - \rho_K)} \right)^{-1} \quad (3.12)$$

que, haciendo el inverso nos queda:

$$C = (1 - \rho_1) \cdots (1 - \rho_K) = \prod_{i=1}^K (1 - \rho_i) \quad (3.13)$$

3.1.2. Condiciones de no saturación

Para comprobar la no saturación de ningún nodo el sistema ($\rho_i < 1$) primero deberemos calcular todas las tasas de llegada con las ecuaciones de tráfico 3.1 descritas en el apartado anterior. Con esto y las tasas de servicio μ_i , utilizaremos las fórmulas de las tasas de utilización ya descritas anteriormente, quedando:

$$\rho_i = \frac{\lambda_i}{c_i \cdot \mu_i} \quad \forall i \in \{1, \dots, K\}, \quad (3.14)$$

siendo c_i el número de servidores del nodo i .

3.1.3. Teorema de Jackson y sus consecuencias

El teorema de Jackson es una adaptación del teorema de Burke a nuestro tipo específico de redes, que se enuncia como vemos en Rivas-López (2013):

“Dado un sistema de red de Jackson que cumpla las condiciones de no saturación, la fórmula para calcular la probabilidad concreta de clientes en cada nodo simultáneamente es la que sigue:

$$P(N_1 = n_1, \dots, N_K = n_K) = p(n_1, \dots, n_K) = \prod_{i=1}^K p_i(n_i) \quad (3.15)$$

donde $p_i(n_i) = P(N_i = n_i)$ es la probabilidad de que haya n_i clientes en el nodo i según las ecuaciones de una cola $M|M|c_i$ ”

Las principales consecuencias de este teorema se ven reflejadas en las medidas de rendimiento de la red, que al poder expresarse como un sistema de colas independientes $M|M|c$, las longitudes de las colas de cada nodo L_{q_i} , el número de clientes en cada nodo del sistema L_i , y el número de clientes en toda la red L_{red} , así como los tiempos de espera, tanto en cada cola W_{q_i} , como en cada nodo W_i , o en la red W_{red} , se calcularán de forma análoga a si fuese una cola separada $M|M|c_i$ en el caso de las medidas para cada nodo, y las correspondientes a la red no serán más que la suma de las de todos los nodos de la red.

3.1.4. Medidas de rendimiento

Como hemos descrito en el apartado anterior, las medidas de rendimiento de los nodos seguirán el mismo patrón que una cola $M|M|c_i$ por lo que serán:

$$L_i = \frac{\rho_i}{(1 - \rho_i)} ; L_{q_i} = \frac{\rho_i^2}{1 - \rho_i} \quad (3.16)$$

$$W_i = \frac{L_i}{\lambda_i} ; W_{q_i} = \frac{L_{q_i}}{\lambda_i} \quad (3.17)$$

$$L_{red} = \sum_{i=1}^K L_i ; \lambda_{red} = \sum_{i=1}^K \gamma_i ; W_{red} = \frac{L_{red}}{\lambda_{red}} \quad (3.18)$$

3.1.5. Ejemplos numéricos

Por último, para comprender mejor el funcionamiento y uso de todas las fórmulas descritas anteriormente, desarrollaremos dos ejemplos numéricos, para los dos tipos de redes abiertas (cíclicas y acíclicas):

Ejemplo 1: Red acíclica

De la figura podemos extraer rápidamente su forma y sus entradas. Como se indica, es una red de Jackson abierta con dos nodos de entrada (nodo 1 y nodo 6) acíclica.

Con esto pasaremos a plantear las ecuaciones de tráfico de los distintos nodos con las fórmulas 3.1:

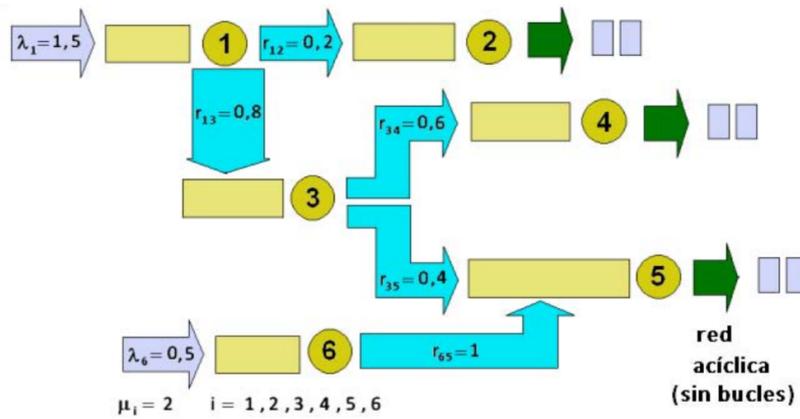


Figura 3.1: Red acíclica (de la Fuente-Fernández, 2020).

Ecuaciones de tráfico

- $\lambda_1 = \gamma_1 = 1.5$ clientes/unidad de tiempo,
- $\lambda_2 = \lambda_1 \cdot r_{12} = 0.3$ clientes/u.t.,
- $\lambda_3 = \lambda_1 \cdot r_{13} = 1.2$ clientes/u.t.,
- $\lambda_4 = \lambda_3 \cdot r_{34} = 0.72$ clientes/u.t.,
- $\lambda_5 = \lambda_3 \cdot r_{35} + \lambda_6 \cdot r_{65} = 0.98$ clientes/u.t.,
- $\lambda_6 = \gamma_6 = 0.5$ clientes/u.t.

A continuación, para poder seguir con los cálculos, debemos comprobar que el sistema está en un estado estable y para ello deberemos comprobar las condiciones de no saturación, es decir, que todos los nodos cumplen

$$\rho_i < 1 \quad \forall i \in \{1, \dots, 6\}$$

Estas tasas de ocupación se calcularán con la fórmula 3.14, y, sabiendo que todos los nodos tienen $c_i = 1$ servidor y $\mu_i = 2$ clientes/u.t.:

Condiciones de no saturación

- $\rho_1 = \frac{\lambda_1}{c_1 \cdot \mu_1} = 0.75 < 1$
- $\rho_2 = \frac{\lambda_2}{c_2 \cdot \mu_2} = 0.15 < 1$
- $\rho_3 = \frac{\lambda_3}{c_3 \cdot \mu_3} = 0.6 < 1$
- $\rho_4 = \frac{\lambda_4}{c_4 \cdot \mu_4} = 0.36 < 1$
- $\rho_5 = \frac{\lambda_5}{c_5 \cdot \mu_5} = 0.49 < 1$
- $\rho_6 = \frac{\lambda_6}{c_6 \cdot \mu_6} = 0.25 < 1,$

podemos afirmar que existe un estado estable en la red.

Por último, calcularemos las distintas medidas de rendimiento desarrolladas en el apartado anterior.

Empezando por el número de clientes en cada cola y cada nodo, con sus respectivas fórmulas presentes en 3.16:

Número de clientes por nodo y cola

- Nodo 1: $L_1 = \frac{\rho_1}{1 - \rho_1} = 3$ clientes ; $L_{q1} = \frac{\rho_1^2}{1 - \rho_1} = 2.25$ clientes,
- Nodo 2: $L_2 = \frac{\rho_2}{1 - \rho_2} = 0.1765$ clientes ; $L_{q2} = \frac{\rho_2^2}{1 - \rho_2} = 0.0265$ clientes,
- Nodo 3: $L_3 = \frac{\rho_3}{1 - \rho_3} = 1.5$ clientes ; $L_{q3} = \frac{\rho_3^2}{1 - \rho_3} = 0.9$ clientes,
- Nodo 4: $L_4 = \frac{\rho_4}{1 - \rho_4} = 0.5625$ clientes ; $L_{q4} = \frac{\rho_4^2}{1 - \rho_4} = 0.2025$ clientes,
- Nodo 5: $L_5 = \frac{\rho_5}{1 - \rho_5} = 0.9608$ clientes ; $L_{q5} = \frac{\rho_5^2}{1 - \rho_5} = 0.4708$ clientes,
- Nodo 6: $L_6 = \frac{\rho_6}{1 - \rho_6} = 0.3$ clientes ; $L_{q6} = \frac{\rho_6^2}{1 - \rho_6} = 0.083$ clientes.

Seguimos por los tiempos de espera en cada cola y cada nodo, con sus fórmulas (3.17), obtenidas a partir de la fórmula de Little:

Tiempos de espera por nodo y cola

- Nodo 1: $W_1 = \frac{L_1}{\lambda_1} = 2$ u.t. ; $W_{q1} = \frac{L_{q1}}{\lambda_1} = 1.5$ u.t.,
- Nodo 2: $W_2 = \frac{L_2}{\lambda_2} = 0.5882$ u.t. ; $W_{q2} = \frac{L_{q2}}{\lambda_2} = 0.0882$ u.t.,
- Nodo 3: $W_3 = \frac{L_3}{\lambda_3} = 1.25$ u.t. ; $W_{q3} = \frac{L_{q3}}{\lambda_3} = 0.75$ u.t.,
- Nodo 4: $W_4 = \frac{L_4}{\lambda_4} = 0.7813$ u.t. ; $W_{q4} = \frac{L_{q4}}{\lambda_4} = 0.2813$ u.t.,
- Nodo 5: $W_5 = \frac{L_5}{\lambda_5} = 0.9804$ u.t. ; $W_{q5} = \frac{L_{q5}}{\lambda_5} = 0.4804$ u.t.,
- Nodo 6: $W_6 = \frac{L_6}{\lambda_6} = 0.6$ u.t. ; $W_{q6} = \frac{L_{q6}}{\lambda_6} = 0.16$ u.t.

Por último, calculamos la tasa de llegadas a la red, así como el número de clientes y el tiempo de espera conjunto con las fórmulas presentes en 3.18:

Tasa de llegadas, número de clientes, y tiempo de espera de la red

- $\lambda_{red} = \sum_{i=1}^6 \gamma_i = 2$ clientes/u.t.,
- $L_{red} = \sum_{i=1}^6 L_i = 6.5331$ clientes,
- $W_{red} = \frac{L_{red}}{\lambda_{red}} = 3.2666$ u.t.

Ejemplo 2: Red cíclica

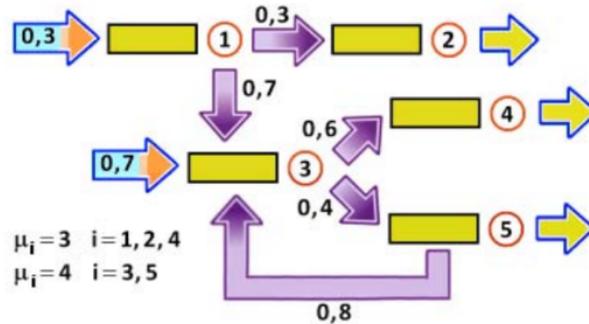
red cíclica (un cliente puede volver a la misma cola)

Figura 3.2: Red cíclica (de la Fuente-Fernández, 2020).

En la figura 3.2 vemos que es una red de Jackson abierta con dos nodos de entrada (nodo 1 y nodo 3) cíclica.

Con esto pasaremos a plantear las ecuaciones de tráfico de los distintos nodos con las fórmulas 3.1. Como en este caso nuestra red es cíclica, en aquellos nodos que inicien un bucle deberemos considerar el mismo en la tasa de llegadas, quedando:

Ecuaciones de tráfico

- $\lambda_1 = \gamma_1 = 0.3$ clientes/u.t.,
- $\lambda_2 = \lambda_1 \cdot r_{12} = 0.09$ clientes/u.t.,
- $\lambda_3 = \gamma_3 + \lambda_1 \cdot r_{13} + \lambda_5 \cdot r_{53} = \gamma_3 + \lambda_1 \cdot r_{13} + \lambda_3 \cdot r_{35} \cdot r_{53} \Rightarrow \lambda_3 = 1.3382$ clientes/u.t.,
- $\lambda_4 = \lambda_3 \cdot r_{34} = 0.8029$ clientes/u.t.,
- $\lambda_5 = \lambda_3 \cdot r_{35} = 0.5353$ clientes/u.t.

En este tipo de ejercicios se debe prestar mucha atención al modelamiento de las entradas en los nodos que forman parte de un bucle y resolver bien la ecuación que se obtiene en ellos.

De nuevo como en el anterior ejemplo, debemos comprobar que el sistema está en un estado estable, viendo si se cumple

$$\rho_i < 1 ; \forall i \in \{1, \dots, 5\}$$

Estas tasas de ocupación se calcularán con la fórmula 3.14, y sabiendo que $c_i = 1$ servidor y $\mu_1 = \mu_2 = \mu_4 = 3$ clientes/u.t.; $\mu_3 = \mu_5 = 4$ clientes/u.t.:

Condiciones de no saturación

- $\rho_1 = \frac{\lambda_1}{c_1 \cdot \mu_1} = 0.1 < 1$
- $\rho_2 = \frac{\lambda_2}{c_2 \cdot \mu_2} = 0.03 < 1$
- $\rho_3 = \frac{\lambda_3}{c_3 \cdot \mu_3} = 0.3346 < 1$

- $\rho_4 = \frac{\lambda_4}{c_4 \cdot \mu_4} = 0.2676 < 1$
- $\rho_5 = \frac{\lambda_5}{c_5 \cdot \mu_5} = 0.1338 < 1$

A la vista de los resultados podemos afirmar que existe un estado estable en la red.

Por último, volveremos a calcular las distintas medidas de rendimiento.

Empezando por el número de clientes en cada cola y cada nodo, con sus respectivas fórmulas presentes en 3.16:

Número de clientes por nodo y cola

- Nodo 1: $L_1 = \frac{\rho_1}{1 - \rho_1} = 0.1$ clientes ; $L_{q1} = \frac{\rho_1^2}{1 - \rho_1} = 0.01$ clientes,
- Nodo 2: $L_2 = \frac{\rho_2}{1 - \rho_2} = 0.0309$ clientes ; $L_{q2} = \frac{\rho_2^2}{1 - \rho_2} = 0.0009$ clientes,
- Nodo 3: $L_3 = \frac{\rho_3}{1 - \rho_3} = 0.5028$ clientes ; $L_{q3} = \frac{\rho_3^2}{1 - \rho_3} = 0.1682$ clientes,
- Nodo 4: $L_4 = \frac{\rho_4}{1 - \rho_4} = 0.3655$ clientes ; $L_{q4} = \frac{\rho_4^2}{1 - \rho_4} = 0.0978$ clientes,
- Nodo 5: $L_5 = \frac{\rho_5}{1 - \rho_5} = 0.1545$ clientes ; $L_{q5} = \frac{\rho_5^2}{1 - \rho_5} = 0.0207$ clientes.

Ahora calculamos los tiempos de espera en cada cola y cada nodo, con sus fórmulas (3.17), obtenidas a partir de la fórmula de Little:

Tiempos de espera por nodo y cola

- Nodo 1: $W_1 = \frac{L_1}{\lambda_1} = 0.370$ u.t. ; $W_{q1} = \frac{L_{q1}}{\lambda_1} = 0.037$ u.t.,
- Nodo 2: $W_2 = \frac{L_2}{\lambda_2} = 0.3433$ u.t. ; $W_{q2} = \frac{L_{q2}}{\lambda_2} = 0.01$ u.t.,
- Nodo 3: $W_3 = \frac{L_3}{\lambda_3} = 0.3757$ u.t. ; $W_{q3} = \frac{L_{q3}}{\lambda_3} = 0.1257$ u.t.,
- Nodo 4: $W_4 = \frac{L_4}{\lambda_4} = 0.4552$ u.t. ; $W_{q4} = \frac{L_{q4}}{\lambda_4} = 0.1218$ u.t.,
- Nodo 5: $W_5 = \frac{L_5}{\lambda_5} = 0.2886$ u.t. ; $W_{q5} = \frac{L_{q5}}{\lambda_5} = 0.0387$ u.t.

Por último, calculamos la tasa de llegadas a la red, así como el número de clientes y el tiempo de espera conjunto con las fórmulas presentes en 3.18:

Tasa de llegadas, número de clientes, y tiempo de espera en la red

- $\lambda_{red} = \sum_{i=1}^5 \gamma_i = 1$ cliente/u.t.,
- $L_{red} = \sum_{i=1}^5 L_i = 1.1648$ clientes,
- $W_{red} = \frac{L_{red}}{\lambda_{red}} = 1.1648$ u.t.

3.2. Redes cerradas de Jackson

Como ya vimos en el apartado de redes de Jackson abiertas, las redes cerradas tienen un amplio uso en telecomunicaciones e ordenadores, definiendo los sistemas de transmisión de datos y rutas de CPU.

Aquí, de forma análoga a las redes abiertas, veremos sus ecuaciones de tráfico, condiciones de no saturación, el teorema de Jackson y sus consecuencias, así como las medidas de rendimiento y ejemplos numéricos, los cuales tendrán ligeras variaciones debido a la ausencia de entradas del exterior.

3.2.1. Ecuaciones de tráfico

Siguiendo la misma ecuación global 3.3 del apartado anterior; ahora, con la ausencia de γ_i en el sistema, esta quedará:

$$\sum_{j=1}^k \sum_{\substack{i=1 \\ i \neq j}}^K \mu_i \cdot r_{ij} \cdot P_{\bar{n};i+j} = \sum_{i=1}^K \mu_i \cdot (1 - r_{ii}) \cdot P_{\bar{n}} \quad (3.19)$$

De aquí, con la misma función global de estados 3.4 y desarrollando volvemos a demostrar 3.19 obteniendo:

$$\sum_{j=1}^k \lambda_j^* \cdot r_{ji} = \sum_{j=1}^k \mu_j \cdot r_{ji} \cdot u_j \quad (3.20)$$

ya que se puede ver que

$$\lambda_j^* = \mu_j \cdot u_j$$

de la fórmula 3.14.

Con esto, podemos volver a utilizar 3.4 para los estados; pero ahora no podrá separarse sin más y deberá evaluarse de la siguiente manera, como se describe en el libro Gross et al. (1998):

$$\sum_{n_1+n_2+\dots+n_K=N} C u_1^{n_1} \cdot u_2^{n_2} \cdot \dots \cdot u_K^{n_K} = 1 \quad (3.21)$$

que despejando C obtenemos:

$$C = \left(\sum_{n_1+n_2+\dots+n_K=N} u_1^{n_1} \cdot u_2^{n_2} \cdot \dots \cdot u_K^{n_K} \right)^{-1} \quad (3.22)$$

Lo que comúnmente se denota como $C^{-1} = G(N)$, que aplicado al caso general con c_i servidores nos queda en una función de los estados, llamada Teorema de Jackson:

$$P_{n_1, n_2, \dots, n_k} = \frac{1}{G(N)} \cdot \prod_{i=1}^k \frac{u_i^{n_i}}{a_i(n_i)} \quad (3.23)$$

siendo a_i una funcion a trozos que toma valores de la siguiente forma:

$$a_i(n) = \begin{cases} n! & \text{si } n \leq c_i \\ c_i! \cdot c_i^{n-c_i} & \text{si } n \geq c_i \end{cases} \quad (3.24)$$

Aunque funciona, cuando nuestro tamaño del sistema N y nuestro número de colas K sean muy altos, los cálculos de 3.23 serán muy difíciles, por lo que el matemático J. P. Buzen desarrolló un algoritmo iterativo que, como se muestra en el libro de de la Fuente-Fernández (2020), funciona de la siguiente manera:

renombramos $G(N) = g_m(n)$ con $m = K$ y $n = N$, definimos:

$$f_i(n) = \frac{u_i^n}{a_i(n)} \forall i \in \{1, \dots, k\}; \forall n \in \{0, \dots, N\}$$

para obtener la recurrencia de $g_m(n)$ de la siguiente manera:

$$g_m(n) = \sum_{n_1 + \dots + n_m = n} \prod_{i=1}^m f_i(n_i) = \sum_{i=0}^n f_m(i) \cdot g_{m-1}(n-i) \quad (3.25)$$

con la que aplicando la fórmula podemos obtener rápidamente:

$$g_m(n) \rightarrow \begin{cases} g_1(n) = f_1(n) \\ g_m(0) = f_m(0) = 1 \quad \text{si } m \in \{1, \dots, K\} \end{cases}$$

Así, la probabilidad de que haya n_K individuos en el nodo K es:

$$p_m(n_K) = p_{\bullet \dots \bullet n_K} = \frac{f_m(n_K) \cdot g_{m-1}(N - n_K)}{g_m(N)} \forall m \in \{1, \dots, K\} \quad (3.26)$$

de lo que obtenemos las siguientes medidas de rendimiento:

$$L_m = \sum_{i=1}^N i \cdot p_m(i) \forall m \in \{1, \dots, K\} \quad (3.27)$$

$$W_m = \frac{L_m}{\lambda_m} \quad (3.28)$$

Para calcular los tiempos de cada nodo necesitaremos calcular las tasas de llegadas reales λ_m . Estas llegadas reales se obtienen de la fórmula

$$\lambda_m = c \cdot \lambda_m^* \quad (3.29)$$

en la que las llegadas relativas λ_m^* se calculan según la fórmula:

$$\lambda_m^* = \sum_{j=1}^K \lambda_j^* \cdot r_{ji} \quad (3.30)$$

estas tasas de llegadas relativas no son más que valores proporcionales de las tasas de llegadas reales, por lo que para poder obtener todas las tasas de llegadas relativas y reales a través de la constante c deberemos asumir que uno de los nodos tiene las tasas de llegada iguales a las tasas de salida, por lo que para ese nodo k :

$$\lambda_m = \mu_m \cdot \sum_{i=1}^n p_m(i) \quad (3.31)$$

Con esto y la fórmula 3.30 podemos despejar c de la igualdad 3.29:

$$c = \frac{\lambda_m}{\lambda_m^*}$$

para obtener el valor de c y así poder calcular las restantes λ_{m-1} .

Aún así, este método sigue siendo costoso y enrevesado de calcular, por lo que el método más utilizado para el cálculo de redes de Jackson cerradas es el algoritmo MVA (Mean Value Analysis) que, como se detalla en de la Fuente-Fernández (2020), se resume en la propiedad del observador aleatorio, que dice:

“Cuando en un sistema cerrado llega un cliente al nodo k , el resto del sistema se puede considerar como un sistema estable de $M - 1$ individuos”

Con este algoritmo, los tiempos de llegada relativos se calculan:

$$\lambda_i^* = \sum_{j=1}^K \lambda_j^* \cdot r_{ji} \quad (3.32)$$

con esto, se calcularán las medidas de rendimiento, que serán algoritmos recursivos en los que las medidas para cada m se calcularán con las medidas para $m - 1$, partiendo de $L_i(0) = 0$.

Existe una aproximación del modelo MVA, conocido como AMVA, que, como se describe en el libro de de la Fuente-Fernández (2020), que utiliza la misma fórmula para calcular los tiempos de espera, la cual es:

$$W_i(m) = \frac{1}{\mu_i} + \frac{L_i(m-1)}{c_i \cdot \mu_i} \quad (3.33)$$

Pero aquí, las llegadas reales y las longitudes de cola son aproximaciones del modelo real. Estas son:

$$\lambda_m = \frac{m}{\sum_{i=1}^K W_i(m) \cdot \lambda_i^*} \quad (3.34)$$

$$L_i = \lambda_i^* \cdot \lambda_m \cdot W_i(m) \quad (3.35)$$

que, con la aproximación de Bard-Schweitzer se puede considerar L_i como:

$$L_i(m) \approx \frac{m}{m-1} \cdot L_i(m-1) \quad (3.36)$$

lo cual suele ser más rápido y computable que nuestro algoritmo original MVA , pero al perder información los cálculos son menos exactos que en el anterior.

3.2.2. Condiciones de no saturación

Como vimos en las redes abiertas, para que se cumpla las condiciones de no saturación ($\rho_i < 1$) primero deberemos calcular todas las tasas de llegada y las tasas de servicio μ_i . Utilizaremos las fórmulas de la tasa de utilización ya descritas en apartados anteriores, quedando:

$$\rho_i = \frac{\lambda_i}{c_i \cdot \mu_i} \forall i \in \{1, \dots, K\} \quad (3.37)$$

siendo c_i el número de servidores del nodo i .

La única diferencia es que, en oposición al tipo de redes de Jackson abiertas, nuestras tasas de ocupación se calcularán una y otra vez para cada uno de los casos específicos de número de individuos m , así como nuestros respectivos $\lambda_i(m)$; lo que, en porcentaje, representará una nueva medida de rendimiento llamada U_i que se utilizará en caso de que un nodo tenga el llamado cuello de botella, que se explicará más detenidamente en el apartado de medidas de rendimiento, así como U_i .

3.2.3. Medidas de rendimiento

Como hemos visto en el apartado de *Ecuaciones de tráfico*, las medidas de rendimiento en redes cerradas se calculan con un algoritmo iterativo que calcula los tiempos de espera y longitudes de cola para cada nodo de los respectivos números de clientes m en base a los cálculos del caso anterior $m - 1$.

Las fórmulas utilizadas de L_i y W_i son:

$$W_i(m) = \frac{1}{\mu_i} + \frac{L_i(m-1)}{c_i \cdot \mu_i} \quad (3.38)$$

$$L_i(m) = m \cdot \frac{\lambda_i^* \cdot W_i(m)}{\sum_{j=1}^K \lambda_j^* \cdot W_j(m)} \quad (3.39)$$

con estas medidas, obtenemos las tasas de llegadas reales con la igualdad:

$$\lambda(m) = \frac{L(m)}{W(m)}$$

Por lo general, estas tres medidas de rendimiento nos darán una estimación de cómo está nuestro sistema, pero hay casos en los que en algún nodo concreto los tiempos de espera y longitud de cola no se estabilizan y tienden a infinito; esto indica que dicho nodo esta sobrecargado de llegadas y absorbe los individuos, distorsionando los resultados de los demás.

A este fenómeno se le conoce como cuello de botella, en cuyo caso haremos uso de una nueva medida de rendimiento: U_i .

U_i es el porcentaje de ocupación de cada nodo, definido por la expresión:

$$U_i = 100 \cdot \rho_i = 100 \cdot \frac{\lambda_i}{c_i \cdot \mu_i}$$

Esta nueva medida de rendimiento nos ayudará a ver gráficamente la ocupación de un nodo y detectar sobrecargas de flujo o posibles sobrecargas futuras en función del tamaño de m .

Para evaluar tamaños del sistema m pequeños su cómputo es rápido y se puede hacer manualmente, pero en situaciones reales el tamaño de m deberá ser elevado; por lo que en este trabajo se diseñará un programa con R que computará todos los valores de L_i , W_i , λ_i y los U_i así como sus gráficas y el grafo del sistema que permitirá calcular de forma rápida y sencilla sistemas de redes amplios y con tamaños elevados. Esto lo desarrollaremos mejor, así como el funcionamiento del programa, en el siguiente capítulo.

3.2.4. Teorema de Jackson y sus consecuencias

Para el caso de redes cerradas también existe un teorema de estados que define la probabilidad de cada uno de los estados del sistema.

El teorema de Jackson para redes cerradas es similar al teorema para redes abiertas 3.15, pero en él se incluye la constante $G(N)$ descrita en el apartado de *Ecuaciones de tráfico* (3.23), y en este caso, en vez de las tasas de ocupación ρ_i se utilizarán los respectivos u_i .

3.2.5. Ejemplo numérico

Por último, calcularemos un ejemplo de redes cerradas para comprender mejor el funcionamiento de este tipo de redes, así como el del algoritmo MVA para las medidas de rendimiento.

Ejemplo red cerrada

“Tenemos un sistema que cumple las siguientes características:

- cinco nodos
- con tasas de salidas 1, 4, 5, 8 y 10 clientes/u.t.
- con enrutados:
 - desde el nodo 1: 25 % a cada uno de los otros cuatro nodos.
 - desde el nodo 2: 75 % de volver al nodo 1 y 25 % de quedarse en el nodo 2.
 - desde el nodo 3: 80 % de volver al nodo 1 y 20 % de quedarse en el nodo 3.
 - desde el nodo 4: 87.5 % de volver al nodo 1 y 12.5 % de quedarse en el nodo 4.
 - desde el nodo 5: 90 % de volver al nodo 1 y 10 % de quedarse en el nodo 5.
- con $m = 200$ iteraciones (desde 1 a 200 clientes en el sistema).

y se pide calcular:

- a) Las ecuaciones de tráfico.
- b) Las respectivas medidas de rendimiento que se consideren oportunas.
- c) Graficar tanto las medidas como el sistema.”

Para el apartado a) calcularemos el producto matricial de $\vec{\lambda}_i = \vec{\lambda}_j \cdot r$ para hallar las respectivas ecuaciones, obteniendo:

Ecuaciones de tráfico

$$\left\{ \begin{array}{l} \lambda_1^* = \lambda_2^* \cdot r_{21} + \lambda_3^* \cdot r_{31} + \lambda_4^* \cdot r_{41} + \lambda_5^* \cdot r_{51} \\ \lambda_2^* = \lambda_2^* \cdot r_{22} + \lambda_1^* \cdot r_{12} \\ \lambda_3^* = \lambda_1^* \cdot r_{13} + \lambda_3^* \cdot r_{33} \\ \lambda_4^* = \lambda_1^* \cdot r_{14} + \lambda_4^* \cdot r_{44} \\ \lambda_5^* = \lambda_1^* \cdot r_{15} + \lambda_5^* \cdot r_{55} \end{array} \right. \quad (3.40)$$

Como obtenemos un sistema indeterminado, para hallar una de las infinitas soluciones suponemos que $\lambda_1^* = 1$ y despejamos las demás, obteniendo:

$$\lambda_1^* = 1; \lambda_2^* = 0.3; \lambda_3^* = 0.3125; \lambda_4^* = 0.2857; \lambda_5^* = 0.27$$

cuyas unidades son número de clientes por unidad de tiempo.

Para el apartado b) haremos uso del algoritmo MVA para calcular los respectivos $L_i(m)$ y $W_i(m)$ con sus fórmulas 3.39 Y 3.38 obteniendo:

$$m = 1 \Rightarrow \left\{ \begin{array}{l} W_1(1) = 1.0000; W_2(1) = 0.2500; W_3(1) = 0.2000; W_4(1) = 0.1250; W_5(1) = 0.1000 \\ L_1(1) = 0.8269; L_2(1) = 0.0689; L_3(1) = 0.0516; L_4(1) = 0.0295; L_5(1) = 0.0229 \end{array} \right. \quad (3.41)$$

$$m = 2 \Rightarrow \left\{ \begin{array}{l} W_1(2) = 1.2067; W_2(2) = 0.2672; W_3(2) = 0.2103; W_4(2) = 0.1286; W_5(2) = 0.1022 \\ L_1(2) = 1.6916; L_2(2) = 0.1248; L_3(2) = 0.0921; L_4(2) = 0.0515; L_5(2) = 0.0398 \end{array} \right. \quad (3.42)$$

$$m = 3 \Rightarrow \left\{ \begin{array}{l} W_1(3) = 1.4229; W_2(3) = 0.2812; W_3(3) = 0.2184; W_4(3) = 0.1314; W_5(3) = 0.1039 \\ L_1(3) = 2.5849; L_2(3) = 0.1702; L_3(3) = 0.1240; L_4(3) = 0.0682; L_5(3) = 0.0524 \end{array} \right. \quad (3.43)$$

•
•
•

$$m = 198 \Rightarrow \left\{ \begin{array}{l} W_1(198) = 49.9744; W_2(198) = 0.372; W_3(198) = 0.265; W_4(198) = 0.145; W_5(198) = 0.112 \\ L_1(198) = 196.8978; L_2(198) = 0.4888; L_3(198) = 0.3267; L_4(198) = 0.1637; L_5(198) = 0.1229 \end{array} \right. \quad (3.44)$$

$$m = 199 \Rightarrow \begin{cases} W_1(199) = 50.2244; W_2(199) = 0.372; W_3(199) = 0.265; W_4(199) = 0.145; W_5(199) = 0.112 \\ L_1(199) = 197.8977; L_2(199) = 0.4888; L_3(199) = 0.3267; L_4(199) = 0.1637; L_5(199) = 0.1229 \end{cases} \quad (3.45)$$

$$m = 200 \Rightarrow \begin{cases} W_1(200) = 50.4744; W_2(200) = 0.372; W_3(200) = 0.265; W_4(200) = 0.145; W_5(200) = 0.112 \\ L_1(200) = 198.8976; L_2(200) = 0.4888; L_3(200) = 0.3267; L_4(200) = 0.1637; L_5(200) = 0.1229 \end{cases} \quad (3.46)$$

Como se puede apreciar, $L_1(m)$ no se estabiliza y tiende a ∞ por lo que se produce un cuello de botella en el nodo 1 y se deberá hacer uso de U_i por lo que calcularemos tanto las tasas de llegada reales como los respectivos $U_i(m)$ obteniendo:

Tasas de llegada reales (λ_i)

$$\lambda_1(1) = 0.8269; \lambda_2(1) = 0.2756; \lambda_3(1) = 0.2584; \lambda_4(1) = 0.2362; \lambda_5(1) = 0.2296 \quad (3.47)$$

$$\lambda_1(2) = 1.4018; \lambda_2(2) = 0.4672; \lambda_3(2) = 0.4380; \lambda_4(2) = 0.4005; \lambda_5(2) = 0.3893 \quad (3.48)$$

$$\lambda_1(3) = 1.8167; \lambda_2(3) = 0.6055; \lambda_3(3) = 0.5677; \lambda_4(3) = 0.5190; \lambda_5(3) = 0.5046 \quad (3.49)$$

•
•
•

$$\lambda_1(198) = 3.9399; \lambda_2(198) = 1.3133; \lambda_3(198) = 1.2312; \lambda_4(198) = 1.1257; \lambda_5(198) = 1.0944 \quad (3.50)$$

$$\lambda_1(199) = 3.9402; \lambda_2(199) = 1.3133; \lambda_3(199) = 1.2312; \lambda_4(199) = 1.1257; \lambda_5(199) = 1.0944 \quad (3.51)$$

$$\lambda_1(200) = 3.9405; \lambda_2(200) = 1.3133; \lambda_3(200) = 1.2312; \lambda_4(200) = 1.1257; \lambda_5(200) = 1.0944 \quad (3.52)$$

Tasa de ocupación (U_i) en %

$$U_1(1) = 20.6726; U_2(1) = 6.8908; U_3(1) = 5.1681; U_4(1) = 2.9532; U_5(1) = 2.2969 \quad (3.53)$$

$$U_1(2) = 35.0454; U_2(2) = 11.6818; U_3(2) = 8.7613; U_4(2) = 5.0064; U_5(2) = 3.8939 \quad (3.54)$$

$$U_1(3) = 45.4176; U_2(3) = 15.1392; U_3(3) = 11.3544; U_4(3) = 6.4882; U_5(3) = 5.0464 \quad (3.55)$$

•
•
•

$$U_1(198) = 98.4991; U_2(198) = 32.8330; U_3(198) = 24.6247; U_4(198) = 14.0713; U_5(198) = 10.9443 \quad (3.56)$$

$$U_1(199) = 98.5066; U_2(199) = 32.8330; U_3(199) = 24.6247; U_4(199) = 14.0713; U_5(199) = 10.9443 \quad (3.57)$$

$$U_1(200) = 98.5140; U_2(200) = 32.8330; U_3(200) = 24.6247; U_4(200) = 14.0713; U_5(200) = 10.9443 \quad (3.58)$$

Como se sospechaba, en el nodo 1 la tasa de saturación se acerca mucho al 100 %, lo que hace que se genere un cuello de botella en él y los demás nodos no trabajen correctamente.

En los siguientes gráficos se ilustra el crecimiento de las distintas medidas en función de m , así como el grafo de la red con sus enrutados:

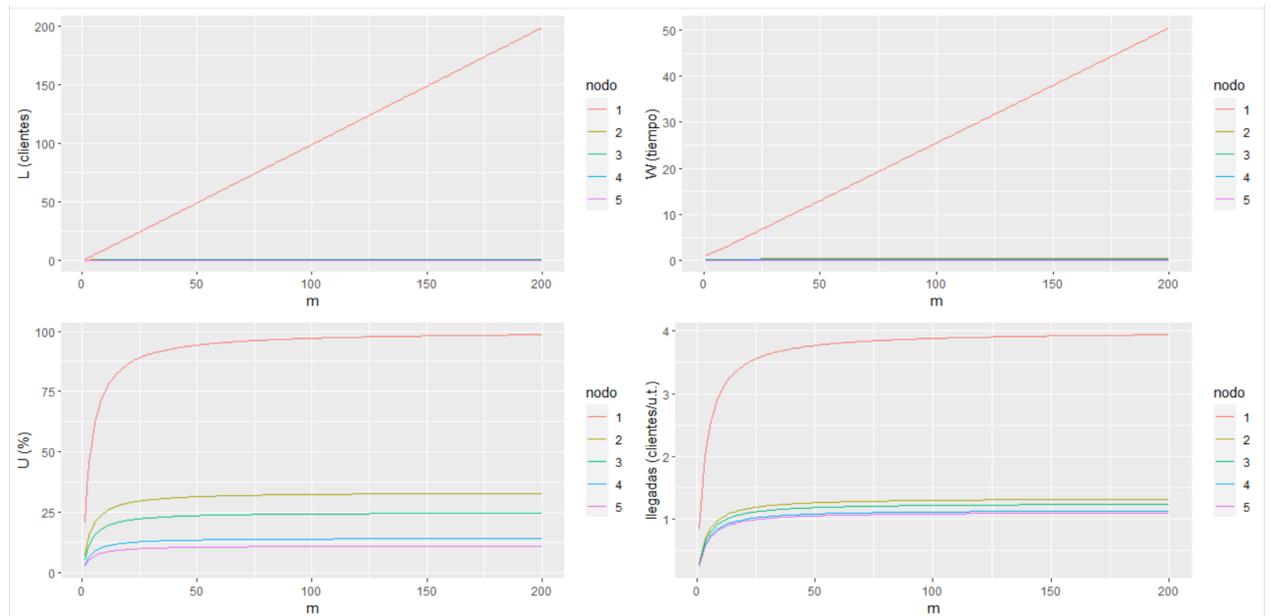


Figura 3.3: Gráficas de L_i, W_i, U_i, λ_i

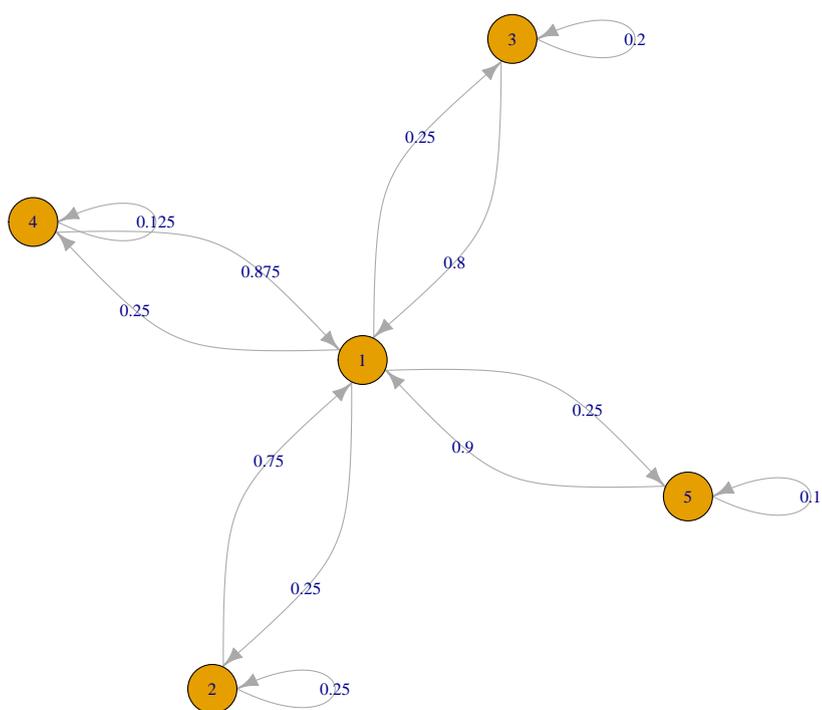


Figura 3.4: Grafo ponderado.

Capítulo 4

Programa en R

En este capítulo describiremos el funcionamiento del programa mencionado en el capítulo anterior.

Este programa resolverá problemas de redes cerradas, calculando las ecuaciones de tráfico, todas las medidas de rendimiento y sus gráficos; así como un grafo ponderado de nuestra red.

Para ello, la función tendrá cinco valores de entrada:

- *nodos*, que hará referencia al número de nodos que tiene nuestro sistema.
- *salidas*, un vector de tamaño $K = n^\circ$ de nodos que hará referencia a la tasa de salidas de cada nodo.
- *enrutados*, una matriz de tamaño $K \cdot K$ con las probabilidades de salida de cada nodo a los restantes o a sí mismo.
- *servidores*, un vector de tamaño K que hará referencia al número de servidores de cada nodo.
- *m*, que hará referencia al número de clientes sobre los que se evaluará el sistema.

Es importante que todos los datos introducidos en nuestra función tengan el tamaño correcto, ya que si no los cálculos no serán correctos o podrá haber imposibilidad de cálculo.

Con todos estos datos introducidos, nuestra función nos devolverá:

- Un vector de tamaño K que hará referencia a las llegadas relativas de cada nodo.
- Cuatro matrices de datos: $L_i(m)$, $W_i(m)$, $U_i(m)$, $\lambda_i(m)$. Estas matrices indicarán todos los cálculos de las respectivas medidas de rendimiento, distribuidas por nodos en las columnas y por tamaño de m en las filas, desde 1 hasta m .
- Un grafo ponderado que representará nuestro sistema para una mejor visualización.
- Los gráficos de las cuatro medidas de rendimiento, para visualizar mejor su crecimiento a lo largo de m .

el código interno de la función se ve así:

```

rendimiento<-function(nodos,salidas,servidores,enrutados,m){

# resolver el sistema de las ecuaciones de trafico para calcular las tasas de llegadas relativas
# usando la función solve
llegadas_relativas<-c(1,solve(diag(1,nodos)-t(enrutados[-1,-1]), enrutados[1,-1]))
print(llegadas_relativas)
# crear las matrices con las sucesiones de L y W (L empieza en 0 y w en 1)
L<-matrix(0,nrow=m+1,ncol=nodos)
W<-matrix(0,nrow=m,ncol=nodos)

# iterar
for(i in 1:m) {

# calcular W_i a partir de L_{i-1}
W[i,]<-1/salidas+L[i,]/(servidores*salidas)

# calcular L_i a partir de W_i
L[i+1,]<-i*llegadas_relativas*W[i,]/(sum(llegadas_relativas*W[i,]))
}

# quitar la fila 0 de L
L<-L[-1,]

# calcular las tasas de llegadas reales
llegadas_reales<-L/W

# calcular los porcentajes de utilizacion (U)
U<-matrix(0,nrow=m,ncol=nodos)

```

Figura 4.1

```

U[i,]<-100*llegadas_reales[i,]/(servidores*salidas)
}
# poner etiquetas a las filas y columnas de las tablas de las cuatro medidas de rendimiento (L, W, U y las tasas de llegada
colnames(L)<-unlist(lapply(1:nodos,function(i){paste("L_",i,sep=""))}))
colnames(W)<-unlist(lapply(1:nodos,function(i){paste("W_",i,sep=""))}))
colnames(U)<-unlist(lapply(1:nodos,function(i){paste("U_",i,sep=""))}))
colnames(llegadas_reales)<-unlist(lapply(1:nodos,function(i){paste("lambda_",i,sep=""))}))
rownames(L)<-rownames(W)<-rownames(U)<-rownames(llegadas_reales)<-unlist(lapply(1:m,function(i){paste("m=",i,sep=""))}))

# sacar las tablas de las cuatro medidas de rendimiento en la consola
print(L)
print(W)
print(U)
print(llegadas_reales)

# sacar el grafo en la consola
# usando la función graph.adjacency
grafo<-graph.adjacency(enrutados,mode="directed", weighted = T)
window(x)
plot.igraph(grafo,edge.curve=0.5, edge.label = E(grafo)$weight)
# crear los gráficos de las cuatro medidas de rendimiento
l<-lapply(list(L,W,U,llegadas_reales),function(i){return(data.frame("m"=rep(1:m,times=nodos),"i"=as.vector(i),
"nodo"=factor(rep(1:nodos,each=m)))))})
labs<-c("L (clientes)", "W (tiempo)", "U (%)", "llegadas (clientes/u.t.)")
plot_l<-lapply(1:4,function(j){return(ggplot(l[[j]], aes(x=m,y=i,group=nodo,color=nodo))+geom_line()+labs(y=labs[j]))})

# sacar los cuatro gráficos en la consola
plot_grid(plot_l[[1]],plot_l[[2]],plot_l[[3]],plot_l[[4]],ncol = 2, byrow = T)

```

Figura 4.2

Capítulo 5

Conclusiones

Para concluir este estudio, resumiremos los puntos más importantes extraídos de la realización del trabajo.

Para comenzar, cabe resaltar la similaridad entre los sistemas de una cola y de redes de colas. Como hemos visto en sus respectivas definiciones, casi todos los parámetros son similares, las nomenclaturas son iguales y tienen las mismas partes resaltables. En la mayor parte de las fórmulas estas son iguales, y cuando no, la mayoría simplemente son agrupaciones de las mismas fórmulas, como hemos visto en los casos de redes de colas infinitas que funcionan como colas aisladas $M|M|c$.

Es por ello que en la realidad siempre será útil trabajar con sistemas de redes y no intentar pensar en colas aisladas, ya que el desarrollo actual da muy poca presencialidad a este tipo de sistemas.

Además hemos visto los amplios tipos de redes, y como cada uno de ellos está presente en un ámbito u otro de la vida. Por ejemplo, las colas tándem que representan la mayor parte de procesos de montaje y fabricación o nuestro tipo más resaltante, las redes de Jackson, que son tan amplias que representan desde complejos sistemas de transmisión de datos en telecomunicaciones hasta un simple modelado de un parque de atracciones.

Así, debemos concienciarnos de la presencia de la teoría de redes en nuestro día a día, y dejar de considerarlo como una simple teoría matemática compleja que solo pueda aplicar a sectores informáticos o clases universitarias; sino a las situaciones más cotidianas de la vida que nos pueden hacer comprender mejor el mundo que nos rodea.

También, ya profundizando más en el conocimiento de estas debemos resaltar la presencia y utilidad de las redes cerradas. En contra de lo que a priori se puede imaginar de una red cerrada, que podría ser un caso más restrictivo y menos útil que una abierta, las redes cerradas (en especial las de Jackson) son un tipo ampliamente presente en sectores de telecomunicaciones y ordenadores, y confeccionan los sistemas de transmisión CPU, ya que estos sistemas teóricamente abiertos pueden considerarse “semicerrados” y los cálculos se realizarán como tal, abriendo un nuevo mundo para dichos sectores, en los que aún queda mucho por estudiar y desarrollar.

Otros temas de interés son conceptos clave que en el día a día entendemos pero que merecen pro-

fundizarse más en ellos para comprender bien el desarrollo de nuestra teoría, como puede ser los conceptos de enrutado, los distintos algoritmos de selección, el concepto de “cuello de botella” y su uso en redes. . . , y otros conceptos que no han sido descritos a fondo en este trabajo pero que se podrán encontrar más detallados en la bibliografía utilizada como son los procesos de nacimiento y muerte, distribuciones gamma y su uso en colas, procesos multiclase e infinidad de ramas que ampliarían nuestro conocimiento de teoría de redes.

Por último, debe mencionarse que aunque en nuestro tipo más importante (las redes de Jackson) los cálculos parecen tediosos y poco aplicables a casos reales, mediante nuestro programa en R hemos demostrado que no es difícil crear códigos en cualquier lenguaje de programación que faciliten el cálculo de estos y nos permitan graficar y obtener resultados para grandes tamaños de iteración. El programa descrito en el apartado anterior se adjuntará al anexo de este trabajo, para el libre uso de cualquier lector y así ayudar a que los cálculos de este tipo de redes se faciliten.

Bibliografía

- Barbour, A. D. (1976). Networks of queues and the methods of stages. *Advances in Applied Probability*, 8, 584-591.
- Baskett, F., Chandy, K. M., Muntz, R. R., & Palacios, F. G. (1975). Open, closed and mixed networks of queues with different classes of costumers. *Journal of the Association for Computing Machinery*, 22, 248-260.
- Burke, P. J. (1956). The output of a queueing system. *Operations research*, 4(6), 699-704.
- de la Fuente-Fernández, S. (2020). *Instrumentos estadísticos avanzados; series y redes de colas*. Universidad Autónoma de Madrid.
- Enrutamiento. (2020). *Enrutamiento basado en políticas* [Accessed: 05-09-2023]. https://es.wikipedia.org/wiki/Enrutamiento_basado_en_pol%C3%ADticas
- Escalante, M. (2023). Algoritmos de Colas: RED (Random Early Drop). *ABC Xperts*.
- Ferrer, V. (Accessed: 05-09-2023). *Modelando sistemas con teoría de colas*. <https://vicentferrer.com/teoria-de-colas/#:~:text=Las%20podemos%20clasificar%20en%20cola,por%20la%20tasa%20de%20abandono>
- Gross, D., & Ince, J. (1981). The machine repair problem with heterogeneous populations. *Operations Research*, 29, 532-549.
- Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (1998). *Fundamentals of queueing theory*. John wiley & sons, Inc.
- Jackson, J. R. (1963). Jobshop-like Queueing Systems. *Management Science*, 10(1), 131-142.
- Kelly, F. P. (1975). Networks of queues with customers of different types. *Journal of Applied Probability*, 12, 542-555.
- Lieberman, G. J., & Hillier, F. S. (2010). *Introducción a la investigación de operaciones*. MC Graw Hill.
- Posner, M., & Bernholtz, B. (1968). Closed finite queueing networks with time lags. *Operations research*, 16(5), 962-976.
- Rivas-López, M. J. (2013). *Apuntes Redes de colas, Asignatura Procesos Estocásticos*. Universidad de Salamanca.
- Zukerman, M. (2013). *Introduction to Queueing Theory and Stochastic Teletraffic Models*. University of Hong Kong.

