

TAXONOMÍA DE LOS MODELOS DE TOPOLOGÍA DE INTERNET

José Ignacio Alvarez-Hamelin¹

¹CONICET y Facultad de Ingeniería, U.B.A., Av. Paseo Colón 850, C1063ACV - Buenos Aires - Argentina, ignacio.alvarez-hamelin@cnet.fi.uba.ar
<http://cnet.fi.uba.ar/ignacio.alvarez-hamelin/>

Palabras clave: Modelos, topología de Internet, distribuciones leyes de potencia, k -núcleos

Resumen. Este trabajo tiene por objeto presentar los modelos de topología de Internet y pone en relieve las problemáticas propias de dicho campo. Es un área de trabajo interdisciplinaria que reúne la informática, la ingeniería en redes y la mecánica estadística, entre otras especialidades. Se presentan, siguiendo el orden cronológico, algunos de los modelos más significativos de la topología de Internet. Además, se exponen algunos de los parámetros usados para la caracterización de dichas redes. Se expone en particular la descomposición en k -núcleos, que mediante su aplicación en la visualización de redes permite una rápida validación de los modelos. Complementariamente, los k -núcleos se relacionan con la posibilidad de ruteo de una red así como con su robustez.

1. INTRODUCCIÓN

El modelado de la topología de Internet surge debido a la necesidad de disponer de redes para la simulación de protocolos de comunicación. Se utiliza un grafo no dirigido para representar la red. El modelo más simple es el grafo aleatorio de Erdős y Rényi (1959), en el que simplemente se define una probabilidad fija de conexión entre dos vértices cualesquiera. A partir del artículo de Waxman (1988) se comienza con el modelado de topologías aplicado a redes de datos como Internet. Este modelo intenta incluir componentes geográficas al modelo de grafo aleatorio, posicionando a los vértices sobre un plano de dimensiones finitas.

En este punto es conveniente distinguir dos niveles de la topología de Internet. El primer nivel es el constituido por los ruteadores y enlaces de datos, mientras que el segundo está formado por los Sistemas Autónomos (AS) (Hawkinson y Bates, 1996). Cada AS está compuesto por un conjunto de ruteadores y enlaces de datos, siendo además administrados en forma centralizada. Claramente el nivel AS es macroscópico y en él se pueden distinguir básicamente las funciones de “cliente” (por ejemplo una universidad), o bien “tránsito” (aquellos que sirven para interconectar a los clientes). Posteriormente los AS dedicados a la interconexión fueron estableciendo otras jerarquías, llamadas *tiers*, donde el nivel 1 corresponde al núcleo de la red (*backbone*), los *tiers* 2 interconectan los *tiers* 1 y 3, y los *tiers* 3 brindan servicio a los AS clientes (o *stubs*).

Los siguientes pasos en el modelado fueron la reproducción de la jerarquía de niveles. Sin embargo, el artículo de Faloutsos et al. (1999) conmueve a la comunidad porque muestra que la distribución de grados de los vértices, en grafos representativos de la topología de Internet, sigue una ley de potencias que difiere sustancialmente de aquella que poseen los grafos aleatorios. Esto desata una intensa actividad en el desarrollo de modelos, no sólo en el área informática sino también en la mecánica estadística, y posteriormente en la ingeniería de redes. El artículo (Faloutsos et al., 1999) está basado en los primeros mapas obtenidos de Internet, tanto a nivel de los AS como de los ruteadores. Actualmente existen un sinnúmero de modelos con distribución de grados siguiendo una ley de potencia, pero en este trabajo nos concentraremos en los más significativos.

En las siguientes secciones veremos los modelos agrupados por categorías. Posteriormente presentaremos algunos de los parámetros utilizados para analizarlos. Finalmente mostraremos cómo los k -núcleos pueden ser utilizados en para validar modelos.

2. MODELOS BASADOS EN GRAFOS ALEATORIOS

Los modelos presentados en esta sección tienen su origen en los grafos aleatorios definidos por Erdős y Rényi (1959): $\mathcal{G}_{n,p}$, donde n indica la cantidad de vértices y p indica la probabilidad de que exista una arista entre dos vértices cualesquiera. Fijados estos parámetros se obtienen grafos de grado medio $\bar{g} = n \cdot p$.

Una primera dificultad de este modelo es que no asegura que el grafo resultante sea conexo, propiedad necesaria para modelar redes. Este problema se presenta en los casos de probabilidad p baja, y la solución es generar varios grafos distintos hasta encontrar alguno conexo. Asimismo es importante destacar que los vértices de $\mathcal{G}_{n,p}$ poseen un grado similar a \bar{g} , es decir que la varianza de los grados es del mismo orden o menor que el grado medio \bar{g} .

2.1. Waxman

Waxman (1988) introduce el concepto de ubicación geográfica en los modelos de Internet con el objeto de simular sus protocolos de comunicaciones. En este trabajo se propone reem-

plazar la probabilidad fija de los grafos aleatorios por la siguiente expresión:

$$p = \alpha e^{-d/\beta L} \quad (1)$$

donde d es la distancia euclidiana entre dos vértices del grafo, L es la distancia máxima, α y β son los dos parámetros del modelo. El aumento de α incrementa el número de aristas, mientras que el aumento de β favorece las aristas de largo alcance. En particular se pueden obtener grafos de características similares para varios pares de valores de α y β , lo que sugiere que en realidad conforman un solo parámetro.

Si bien este modelo no es un grafo aleatorio, la distribución de grados posee un coeficiente de varianza < 1 (Medina et al., 2000), tal como ocurre en los grafos aleatorios.

2.2. Modelos jerárquicos

Citaremos en primer lugar los modelos jerárquicos propuestos por Zegura et al. (1997): el *M-level* y el *Transit-Stub*. El primero es una superposición de grafos aleatorios, donde se genera un grafo aleatorio (el nivel superior), luego algunos vértices son reemplazados a su vez por grafos aleatorios de forma recursiva hasta llegar al nivel inferior, generando así los m niveles.

El segundo modelo propuesto por Zegura et al. (1997) tiene en cuenta la división de los Sistemas Autónomos (AS) (Hawkinson y Bates, 1996) entre “clientes” (*stub*) y de “interconexión” (*transit*). El modelo comienza por producir un grafo aleatorio G con n vértices que corresponde a los AS de interconexión. Luego cada vértice i de dicho grafo se reemplaza por un grafo aleatorio G'_i de n' vértices. Finalmente, por cada vértice de G'_i se generan k grafos aleatorios $G''_{i,k}$ de n'' vértices, correspondientes a los AS clientes o *stub*.

Los dos modelos presentados en esta sección carecen de consideraciones geográficas. Doar (1996) propone un modelo geográfico donde se distinguen tres niveles de jerarquía: WAN (*Wide Area Network*), MAN (*Metropolitan Area Network*) y LAN (*Local Area Network*). Comenzando por el nivel más elevado (WAN), se ponen vértices al azar en el plano, descartando aquellos que son próximos. Luego se los conecta usando un árbol que minimiza las distancias euclidianas. A continuación se agregan algunas aristas más para incrementar la redundancia de caminos. Luego por cada vértice se crean varias MAN, utilizando el mismo método que el usado para las WAN pero con una grilla de mayor resolución. Finalmente se crean las LAN como vértices que se conectan en topología de estrella a un vértice correspondiente a la MAN más cercana.

Todos estos modelos son similares a una superposición de grafos aleatorios y responden a la característica básica de tener una distribución de grados con bajo coeficiente de varianza.

3. MODELOS CON DISTRIBUCIÓN LEY DE POTENCIA

Posteriormente, Faloutsos et al. (1999) muestran que la distribución de grados de varias exploraciones de la topología de Internet (tanto a nivel de AS como de ruteadores) no corresponden a un grafo aleatorio, sino que dichas distribuciones siguen una ley de potencias, es decir una distribución donde la varianza es mucho mayor que el valor medio, o incluso infinita. Esto desencadenó una gran actividad en el área del modelado, pero también en el desarrollo y estudio de los métodos de exploración y análisis de parámetros. Si bien la distribución de grados es un parámetro importante, rápidamente aparecieron otros que permitieron distinguir entre un mapa de AS y uno de ruteadores (Pastor-Satorras y Vespignani, 2004), hablaremos de estos parámetros en la sección 4.1.

3.1. Bárabasi-Albert

El primer modelo que cumple con esta característica es el propuesto por [Barabási y Albert \(1999\)](#), que propone el “acoplamiento preferente” (*preferential attachment*) como mecanismo originario de las leyes de potencia en la distribución de grados. En los grafos aleatorios Erdős-Rényi cada vértice se conecta con probabilidad fija sobre el conjunto de vértices. En este modelo se reemplaza esta probabilidad fija por el acoplamiento preferente:

$$p(i) = \frac{g(i)}{\sum_{\forall j \in G} g(j)} \quad (2)$$

donde $g(i)$ es el grado del vértice i , y G es grafo actual. El modelo es dinámico y permite obtener, mediante iteraciones, un grafo que incrementa su tamaño con tiempo, emulando el crecimiento. Esta característica fue adoptada por los modelos sucesivos, ya que esto permite el pronóstico. El algoritmo básico es el siguiente:

Modelo Barabási y Albert (1999)	
1	inicio: grafo completo de 4 vértices
2	repetir: (bucle principal)
3	crear un nuevo vértice v
4	repetir m veces:
5	conectar v a u , según probabilidad $p(u)$ (ver ecuación 2)

Es decir que cada nuevo vértice se conecta a m vértices utilizando el acoplamiento preferente, que es el originario de la ley de potencias en la distribución de grados. Si bien el modelo original es bastante limitado, sólo produce leyes de potencia con exponente 3. Posteriormente varias generalizaciones fueron propuestas permitiendo otros exponentes ([Yamasaki et al., 2005](#)).

Pese a que el mecanismo del acoplamiento preferente permite construir grafos, cuya distribución de grados sigue leyes de potencia, plantea algunos inconvenientes. El primero de ellos es la escala absoluta de “popularidad”, es decir que los vértices más populares (o con grado elevado) atraen más que los otros. El segundo es cómo los vértices pueden conocer la popularidad de todos los vértices. Finalmente el modelo no contempla los aspectos geográficos.

3.2. Optimización multicriterio

Siguiendo otra línea de trabajo [Fabrikant et al. \(2002\)](#) proponen un modelo extremadamente simple, donde el mecanismo que lleva a la aparición de la ley de potencias es la optimización de dos parámetros: la distancia euclidiana y la distancia a un centro (predeterminado) del grafo. La competencia de estos dos parámetros, en ciertos rangos, permite la obtención de grafos con una distribución de potencias. Específicamente se minimiza una función de costo, para un nuevo vértice i que se quiere conectar a la red mediante otro vértice j , de la siguiente forma:

$$w(i, j) = d(i, j) + \xi \cdot h(j) \quad (3)$$

donde $d(i, j)$ es la distancia euclidiana entre los vértices i y j , $h(j)$ es la distancia de j a la raíz r en el grafo, y ξ es el parámetro del modelo. La raíz r es un vértice especial que indica el “centro” del grafo. Este es también un modelo dinámico, donde el grafo aumenta su tamaño según el siguiente algoritmo:

Modelo Fabrikant et al. (2002)

-
- 1 **inicio:** crear el vértice r con posición aleatoria
 - 2 **repetir:** (bucle principal)
 - 3 crear un nuevo vértice v con posición aleatoria
 - 4 conectar v al vértice u que minimiza $w(v, u)$ (ver ecuación 3)
-

Este modelo incluye la posición geográfica de los vértices. Variando el parámetro ξ se obtiene una estrella, para $\xi \rightarrow \infty$, o bien un árbol donde cada nuevo vértice se conecta al más cercano, para $\xi = 0$. Sólo se obtiene un grafo con una ley de potencias para los valores $\frac{1}{\sqrt{n}} < \xi \leq \frac{1}{4}$, donde n es el número de vértices.

Aunque la ley de potencias aparece de una forma más elegante que en el modelo anterior, este modelo tiene varios problemas. El primero es que el grafo obtenido es un árbol, mientras que las redes en general poseen bucles. En segundo lugar, el centro de la red r no debería ser estático. Por último, el carácter *global* no es realista ya que los nuevos *usuarios* no tienen forma de conocer Internet en forma global.

Extensión del modelo Fabrikant et al. (2002). En el artículo de Alvarez-Hamelin y Schabanel (2004) se presenta una extensión que soluciona los dos primeros problemas, obteniendo redes más cercanas a las reales. La modificación más significativa es que una vez conectado el nuevo vértice, se agrega una arista e más que minimiza la siguiente función:

$$W(e) = d(e) + \frac{\gamma}{n} \cdot \sum_{i=1}^n (h'(i) - h(i)) \quad (4)$$

donde $d(e)$ es la longitud euclidiana del arista e , $h(i)$ es la distancia en el grafo de i a la raíz r sin la nueva arista e , mientras que $h'(i)$ es con la nueva arista e . El algoritmo completo es:

Modelo Alvarez-Hamelin y Schabanel (2004)

-
- 1 **inicio:** crear un vértice x con posición aleatoria, tal que la raíz $r = x$
 - 2 **repetir:** (bucle principal)
 - 3 crear un nuevo vértice v con posición aleatoria
 - 4 conectar v a los m vértices u_i que minimicen $w(v, u_i)$ (ver ecuación 3)
 - 5 crear q nuevas aristas e_j que minimicen $W(e_j)$ (ver ecuación 4)
 - 6 elegir $r = i$ en forma aleatoria proporcionalmente al grado $g(i)$ del vértice i
-

Este modelo generaliza y extiende el modelo anterior, ya que la raíz no queda fija, el nuevo vértice se conecta con m aristas, y además se incluyen q nuevas aristas que minimizan una función de costo, inspirada en la propuesta de Fabrikant et al. (2002).

Aunque este modelo genera redes realistas (distribución de grados y otros parámetros), presenta dos problemas. El primero es que sigue siendo un cálculo global, mientras que el segundo problema es la complejidad en tiempo: $\mathcal{O}(n^4)$ (los autores están trabajando en un nuevo algoritmo con $\mathcal{O}(n^2 \log n)$).

Si bien en el artículo original el modelo estaba perfilado para representar los AS, es apropiado para la representación de la topología de ruteadores (ver sección 4.2).

3.3. Competencia y adaptación

Serrano et al. (2005) han presentado un modelo basado en la “competencia” por los usuarios y la “adaptación” de los AS a las condiciones de tráfico. Si bien el modelo es extremadamente simple, ya que los usuarios generan el mismo tipo de tráfico, éste intenta capturar el aspecto económico del problema. El punto de partida son las observaciones sobre las tasas de crecimiento del número de usuarios, el número de AS y el número de enlaces entre los AS. Estas tasas son exponenciales y se pueden modelar como: $W(t) \sim W_0 e^{\alpha t}$ para el número de usuarios, $N \sim N_0 e^{\delta k}$ representa el número de AS, y $E \sim E_0 e^{\beta k}$ el número de conexiones o enlaces de datos entre los AS, todos ellos expresados en número por mes. Se verifica además que $0,036 \simeq \alpha \lesssim \delta \lesssim \beta \simeq 0,029$. La pregunta que surge es si estas desigualdades son reales o son producto de fluctuaciones estadísticas. Los autores argumentan que existen dos mecanismos para compensar el incremento de usuarios: la creación de nuevos AS o bien la incorporación de nuevos enlaces entre AS ya existentes. Cuando estos mecanismos operan al mismo tiempo, la tasa de creación de AS δ debe ser menor a la tasa de creación de usuarios α , ya que parte del tráfico está absorbido por los enlaces entre los AS existentes, es decir β .

Cada usuario nuevo será incorporado a un AS i con probabilidad

$$\Pi_i = \frac{\omega_i}{W(k)} \quad , \quad (5)$$

donde ω_i es el número de usuarios del AS i , y $W(k) = \omega_0 N_0 e^{\alpha k}$. Esta ecuación es el acoplamiento preferente aplicado a los usuarios, donde el AS más popular es aquel que tiene más usuarios.

Luego, los usuarios son demandantes de ancho de banda, ya sea para recibir o transmitir información. Por lo tanto los sistemas autónomos se adaptan estableciendo conexiones o incrementando el ancho de banda con otros AS. En este trabajo el ancho de banda se considera discreto. Entonces, el ancho de banda $b_i(k)$ del sistema autónomo i en el tiempo k , sigue la siguiente ecuación¹:

$$b_i(k) = 1 + \frac{2B(k) - N(k)}{W(k) - \omega_0 N(k)} (\omega_i(k) - \omega_0) \quad , \quad (6)$$

donde $B(k)$ es el ancho de banda total en la red, y el término $\frac{2B(k) - N(k)}{W(k) - \omega_0 N(k)} \sim \frac{2B(k)}{W(k)}$.

Finalmente, una vez determinado el ancho de banda necesario, se seleccionan pares de AS (i, j) , en función de la variación de ancho de banda, es decir el nuevo ancho de banda requerido, y se conectan proporcionalmente según la función de costo

$$D(d_{i,j}, \omega_i, \omega_j) = e^{-d_{i,j}/d_e(\omega_i, \omega_j)} \quad , \quad (7)$$

donde $d_{i,j}$ es la distancia euclidiana entre el AS i y el j , y $d_e(\omega_i, \omega_j) = \omega_i, \omega_j / \kappa W(k)$, con κ el número de usuarios por unidad de distancia.

El algoritmo completo resulta entonces:

¹El detalle de las deducciones se puede encontrar en el artículo original (Serrano et al., 2005)

Modelo Serrano et al. (2005)

-
- 1 **repetir:** (bucle principal)
 - 2 $\Delta W(k) = \omega_0 N_0 (e^{\alpha k} - e^{\alpha(k-1)})$ nuevos usuarios eligen un AS según la probabilidad dada por la ecuación (5)
 - 3 $\Delta N(k) = N_0 (e^{\beta k} - e^{\beta(k-1)})$ nuevos AS son creados con ω_0 usuarios, tomados de los ya existentes en los antiguos AS. Cada nuevo AS se posiciona en el plano en forma aleatoria, siguiendo la dimensión fractal 1,5 (Yook et al., 2002)
 - 4 Cada AS incrementa su ancho de banda en $\Delta b_i(k) = b_i(k) - b_i(k-1)$ de acuerdo a la ecuación (6)
 - 5 Todo par de vértices (i, j) , elegido con probabilidad proporcional a $\Delta b_i(k)$ y $\Delta b_j(k)$, se conectan con probabilidad $D(d_{i,j}, \omega_i, \omega_j)$ (ver ecuación (7))
-

Este modelo presenta un muy buen comportamiento para varios de los parámetros que se encuentran en las exploraciones de la topología de los AS. Adicionalmente, es el primer modelo que intenta incluir a los usuarios como el origen de su dinámica, que introduce el factor económico (representado por la ecuación (7) y la competencia de los usuarios), y que utiliza un posicionamiento físico de los AS. Este último punto es discutible porque los AS son objetos de dos dimensiones, ya que se despliegan a lo largo de la superficie de la tierra. Sin embargo, la necesidad de introducir un costo obliga la representación (aunque abstracta) de la distancia. Como veremos en la sección 4.2, es un muy buen modelo para representar la topología de los AS.

3.4. Optimización del flujo máximo

Li et al. (2005) han presentado un modelo de red donde se parte de una configuración que sigue la distribución de grados y que posteriormente se optimiza de forma tal de maximizar el flujo entre todos los pares de vértices. Dicha optimización se realiza reconectando varios vértices al mismo tiempo de manera de no alterar el grado de ninguno de ellos. Por el momento los autores han producido sólo redes de tamaño reducido para mostrar la diferencia de propiedades, básicamente a nivel del flujo máximo obtenido, comparando esta red con las otras producidas con otros modelos.

Un posible algoritmo que realice esto se puede describir como:

Modelo Li et al. (2005)

-
- 1 Generar un grafo G con n vértices, que cumpla con la distribución de grados deseada
 - 2 **repetir:** (hasta que se alcance la optimización deseada)
 - 3 elegir 2 pares de vértices vecinos (a, b) y (c, d)
 - 4 calcular el flujo total del grafo usando las siguientes aristas:
 - 5 $S_1 = (a, b)$ y (c, d) (grafo sin modificar)
 - 6 $S_2 = (a, d)$ y (c, b)
 - 7 $S_3 = (a, c)$ y (b, d)
 - 8 elegir la solución S_i que maximiza el flujo total
-

Si bien en este algoritmo no está definida la optimización deseada ni la forma en la que se eligen las aristas, proporciona los lineamientos básicos propuestos en el artículo.

En particular, este modelo presenta el punto de vista de la ingeniería en redes de datos, donde el problema se trata desde la aplicación: el diseño de una red eficiente, y no desde el marco teórico abstracto. Además, este punto de vista tiende a ser más local que global, es decir

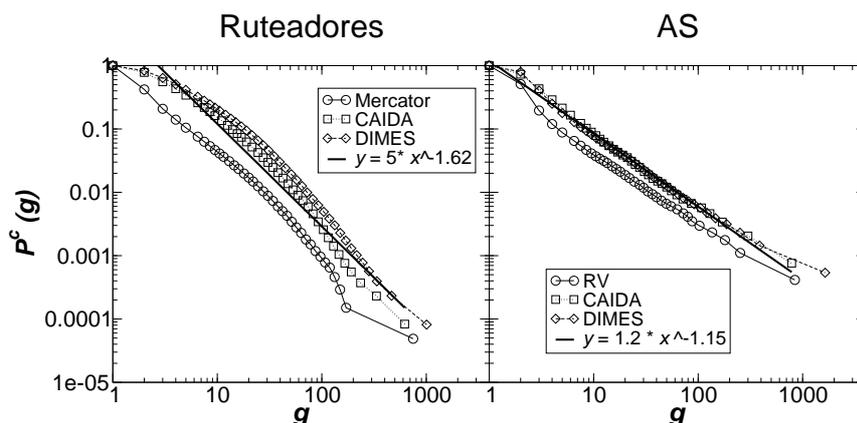


Figura 1: Distribución acumulativa de grados para algunos mapas de ruteadores y AS. Se muestra la ley de potencia que interpola los mapas de CAIDA, si se considera la distribución no acumulativa se debe sumar 1 al exponente.

se diseña una red *completa* donde se optimiza todo el tráfico. En otras palabras, dado el tipo de optimización este método no permite crear redes demasiado grandes, debido a su complejidad de cálculo. En este sentido es una propuesta *local*, ya que se ocupa de redes administradas de forma centralizada, lo que es opuesto a Internet (un conjunto de redes administradas en forma independiente). De todas maneras es importante encontrar un punto de contacto entre las visiones local y global, es decir un equilibrio entre lo teórico y lo práctico.

4. PARÁMETROS DE ANÁLISIS

En esta sección presentaremos algunos de los parámetros más significativos utilizados para el análisis y validación de los modelos. Nuevamente, la lista no es completa, pero pone énfasis en aquellos que han mostrado características estructurales importantes.

4.1. Grafos

Presentaremos aquí algunos de los parámetros tomados de la teoría de grafos.

Distribución de grados. Existen varias formas de expresarla, una de ellas es:

$$P(g) = \frac{1}{n} \sum_{\forall i/g(i)=g} 1 \quad (8)$$

donde $P(g)$ representa la probabilidad que un vértice tenga grado g , $g(i)$ es el grado del vértice i , y n es el número total de vértices. El artículo de Faloutsos et al. (1999) se basó en esta representación para obtener un exponente entre 2,1 y 2,4. Otra posibilidad es la distribución acumulativa:

$$P^c(g) = 1 - \sum_{0 \leq x \leq g-1} P(x) \quad (9)$$

La ventaja de esta representación es que posee menos fluctuaciones, permitiendo una visión más clara de la tendencia de los datos. En este caso, si consideramos que los datos siguen una ley exponencial, el exponente de los datos no acumulados es la pendiente estimada a partir de $P^c(g)$ sumándole 1. En la figura 1 se muestran los gráficos correspondientes a diferentes exploraciones

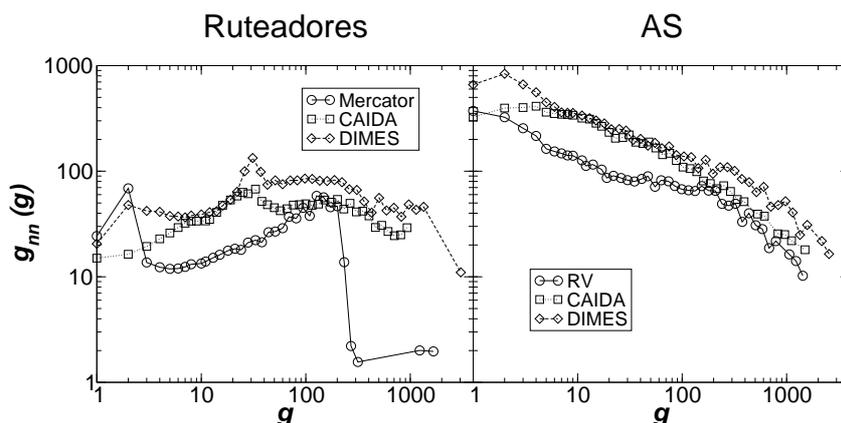


Figura 2: Distribución del grado medio de los vecinos de los vértices de grado g .

de Internet, tanto de ruteadores como de AS. Govindan y Tangmunarunkit (2000) exponen uno de los primeros mapas de ruteadores obtenidos llamado Mercator. Luego tenemos los mapas obtenidos por CAIDA (2003), donde presentamos un mapa de los ruteadores del 2003, y uno de AS del 2005. A continuación presentamos los mapas obtenidos por DIMES (2005), ambos del 2005. Finalmente mostramos el mapa de AS obtenido a partir de Router-Views (2005), correspondiente al año 2005.

Distribución del grado medio de los vecinos Otro parámetro interesante es observar cómo los vértices están conectados. El grado medio de los vecinos a un vértice de grado g se define como:

$$g_{nn}(g) = \frac{1}{n_g} \sum_{j/g_j=g} \frac{1}{|V(j)|} \sum_{i \in V(j)} g_i, \quad (10)$$

donde $V(j)$ es el conjunto de los vecinos del vértice j , $|V(j)|$ su cardinalidad, g_i es el grado del vértice i , y n_g es el número de vértices de grado g . En el caso que los vértices de grado elevado posean un g_{nn} pequeño, entonces ellos actuarán de concentradores, implicando que los vértices de grado bajo tengan como vecinos otros de grado elevado. Este comportamiento está definido como *disassortative* (Newman, 2002) y se observa en la red de AS. En cambio, cuando los vértices de grado elevado tengan como vecinos a otros de grado también elevado, se denomina *assortative* (Newman, 2002). Este último comportamiento es observado en algunas redes sociales. En la figura 2 podemos ver que los $g_{nn}(g)$ obtenidos para los mapas de AS son *disassortative*, mientras que los mapas de ruteadores tienen muy poca variación (menos de una década). Las definiciones de *assortative* o *disassortative* tienen sentido cuando las pendientes son marcadas, si no las pequeñas variaciones pueden ser originadas por los sesgos en la obtención de los mapas.

Distribución del coeficiente de clustering El coeficiente de *clustering* indica qué probabilidad tienen los vecinos de un vértice de estar interconectados entre sí:

$$cc_i = \frac{2 n_{\text{link}}}{g_i(g_i - 1)}, \quad (11)$$

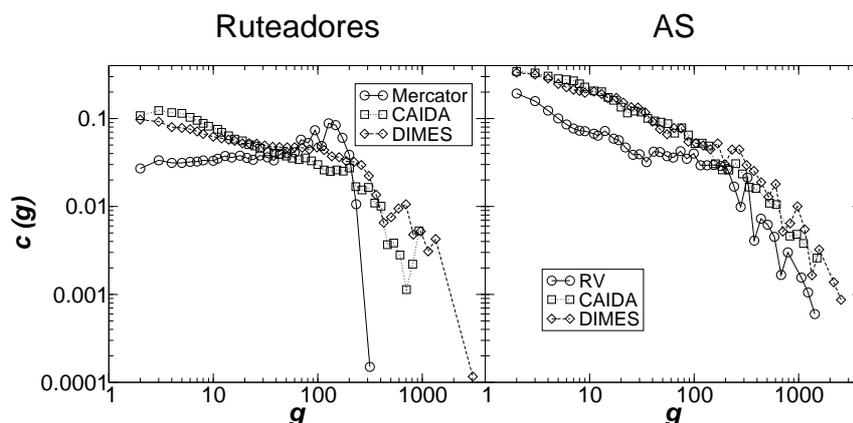


Figura 3: Distribución del coeficiente de *clustering* medio en función del grado g .

donde n_{link} representa el número de conexiones o aristas entre los vecinos de i , y g_i es el grado del vértice i . Su distribución en función del grado es:

$$c(g) = \frac{1}{n_g} \sum_{j/g_j=g} cc_j . \quad (12)$$

Estas curvas pueden verse en la figura 3. Es importante notar que existe una relación entre $c(g)$ y $g_{nn}(g)$, ya que para tener un elevado coeficiente de *clustering* es necesario que el grado de los vecinos sea elevado. Otra observación útil es que estos dos parámetros permiten la diferenciación entre el grafo que representa a los AS y el de los ruteadores. En el primer caso el comportamiento es *disassortative*, mostrando ambas curvas una recta con pendiente negativa en la representación logarítmica. En cambio, para el grafo de los ruteadores, ambas curvas tienen un comportamiento casi estático (ver figura 2).

4.2. k -núcleos

Comencemos por definir k -núcleo: *Un subgrafo $H = (C, E|C)$ inducido por el conjunto $C \subseteq V$ es un k -núcleo o un núcleo de orden k sii $\forall v \in C : \text{grado}_H(v) \geq k$, y H es el máximo subgrafo con esta propiedad.*

Una forma de obtener la descomposición en k -núcleos es eliminando recursivamente todos los vértices de grado menor que k , hasta que todos los vértices restantes tengan grado mayor o igual a k . Se demuestra que la complejidad es $\mathcal{O}(n)$ para un grafo conexo con n aristas (Batagelj y Zaversnik, 2002).

Podemos además definir *capa* de la siguiente manera: *Un vértice i tiene número de capa² c , si dicho vértice pertenece al c -núcleo pero no al $(c + 1)$ -núcleo. Tomamos como notación c_i el número de capa del vértice i .*

La importancia de este parámetro, el número de capa, es su relación con la conectividad. Si bien un k -núcleo no es k -conexo en general, en ciertos casos prácticos se puede dar esta conectividad (Bollobàs, 1983; Carmi et al., 2006). Si un grafo es k -conexo, entonces existen k caminos independientes entre todo par de vértices. En otras palabras, dado dos vértices u y v , existen al menos $\min(c_u, c_v)$ caminos independientes entre ambos. En lenguaje de redes de

²Denominado *shell index* en inglés.

datos esto significa que si alguno de esos caminos está obstruido, ya sea por congestión o por una falla momentánea, entonces se puede elegir otro camino para llegar al destino. Incluso, si necesitamos una red que cumpla con cierta calidad de servicio, es decir que respete ciertas cotas mínimas para algún parámetro (demora, ancho de banda, etc.), cuanto más caminos existan entre sus vértices, más posibilidades tendrá de encontrar aquel que verifique la calidad deseada.

Otro aspecto significativo es que la descomposición en k -núcleos (k -core decomposition) se ha utilizado recientemente para la diferenciación de tipos de redes mediante su visualización (Alvarez-Hamelin et al., 2006), y para el análisis de redes Alvarez-Hamelin et al. (2005).

Presentamos aquí algunos de los resultados de la visualización utilizando LaNet-vi (2005). La idea principal es ubicar en el centro la capa con el $k_{\text{máx}}$ -núcleo en forma de círculo (su diámetro es proporcional al número de elementos), y luego las otras capas en circunferencias concéntricas que se alejan del $k_{\text{máx}}$. En la derecha figura una escala de colores con el número de capa c_i , mientras que en la izquierda se encuentra la escala del grado de los vértices representado por el tamaño de los mismos. Esta última escala es logarítmica y figura el grado máximo. Si bien los vértices están ubicados inicialmente en una circunferencia, éstos pueden desplazarse hacia el centro según el valor promedio de c_v de sus vecinos v . Por último señalamos que no se grafican todas las aristas, sino un porcentaje de las mismas elegidas con probabilidad uniforme. Los detalles de la visualización están descriptos extensivamente en Alvarez-Hamelin et al. (2006).

En la figura 4 presentamos el resultado de un grafo de Erdős y Rényi (1959), con $\bar{g} = 10$. Vemos que la mayor parte de los vértices están en el núcleo máximo, donde $k_{\text{máx}} < \bar{g}$. Luego hay algunos vértices en otras capas, pero son muy pocos.

El modelo Barabási y Albert (1999) está presentado en la figura 5, con $m = 2$. En este caso hay solamente 3 capas, ya que la capa 3 está formada por los cuatro primeros vértices del grafo inicial, mientras que cada vértice que se agrega con sólo dos aristas, son los que pueblan la capa 2, y la capa 1 es conjunto vacío.

La figura 6 muestra la visualización de un mapa de AS obtenido por CAIDA (2003). Aquí vemos principalmente que la mayoría de las aristas van hacia el $k_{\text{máx}}$ -núcleo, y que además el tamaño (es decir el grado) de los vértices está correlacionado con el número de capa. Estas dos observaciones muestran que el modelo es altamente jerárquico.

Para comparar con la figura anterior, se muestra el modelo Serrano et al. (2005) en la figura 7. Aquí podemos ver que posee gran similitud con los AS, ya sea por el gran número de capas, la correlación del grado-capas, y que la mayor parte de las aristas van hacia el $k_{\text{máx}}$ -núcleo. De ahí que podemos concluir que es un muy buen modelo. En el artículo original se muestra que también sigue otras de las propiedades observadas en los mapas de AS.

A continuación tenemos un mapa de ruteadores obtenido por CAIDA (2003) (figura 8). Comparando con la figura 6 vemos que son verdaderamente diferentes. En particular el número de capas es mayor, todas las capas están densamente pobladas, las aristas van principalmente entre capas y no a la capa central, y no existe correlación grado-capas como ocurre en el mapa de AS. Esto último se puede ver ya que en las capas exteriores tenemos vértices de gran tamaño o grado elevado.

Finalmente presentamos en la figura 9 el modelo Alvarez-Hamelin y Schabanel (2004). Si bien este grafo es relativamente pequeño (10^3 nodos), se ve que las aristas conectan principalmente capas, y que no hay correlación grado-capas. En este sentido es un modelo interesante para el grafo de los ruteadores. Sin embargo, sería deseable analizar un grafo de la magnitud que posee el de los ruteadores (10^5 aproximadamente), pero tal como se dijo en la sección 3.2, el tamaño actual está limitado por la complejidad del algoritmo (tema en el que están trabajando los autores para mejorarla).

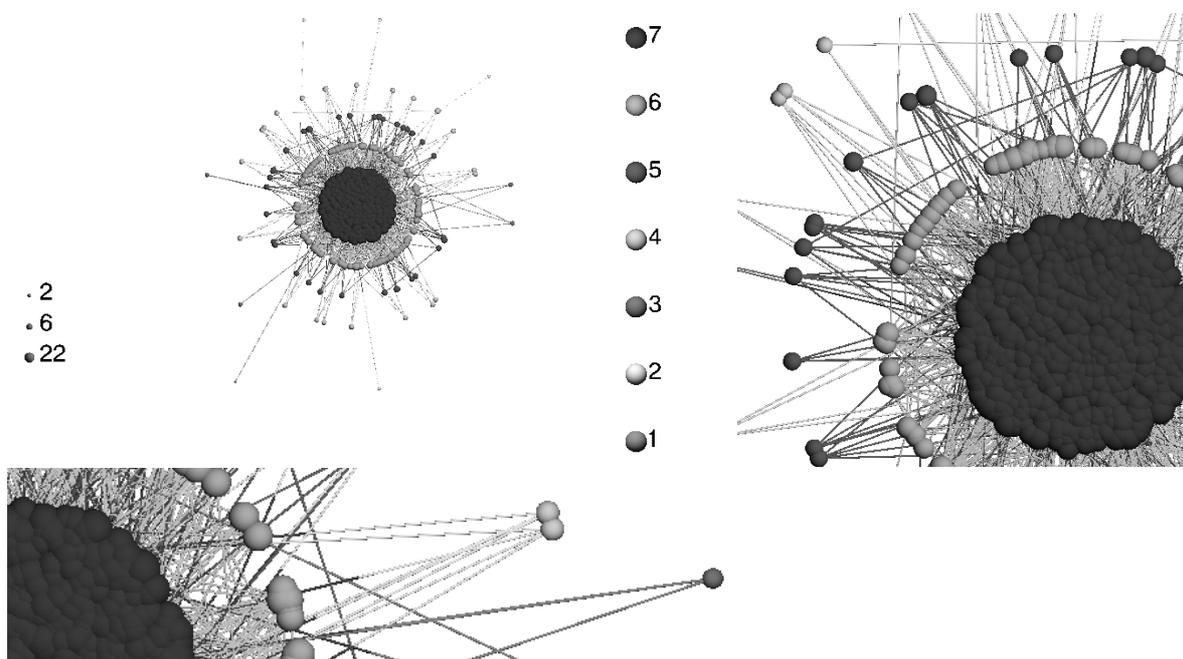


Figura 4: Visualización, utilizando la descomposición en k -núcleos, de un grafo de Erdős y Rényi (1959), con 1000 vértices y grado medio $\bar{g} = 10$.

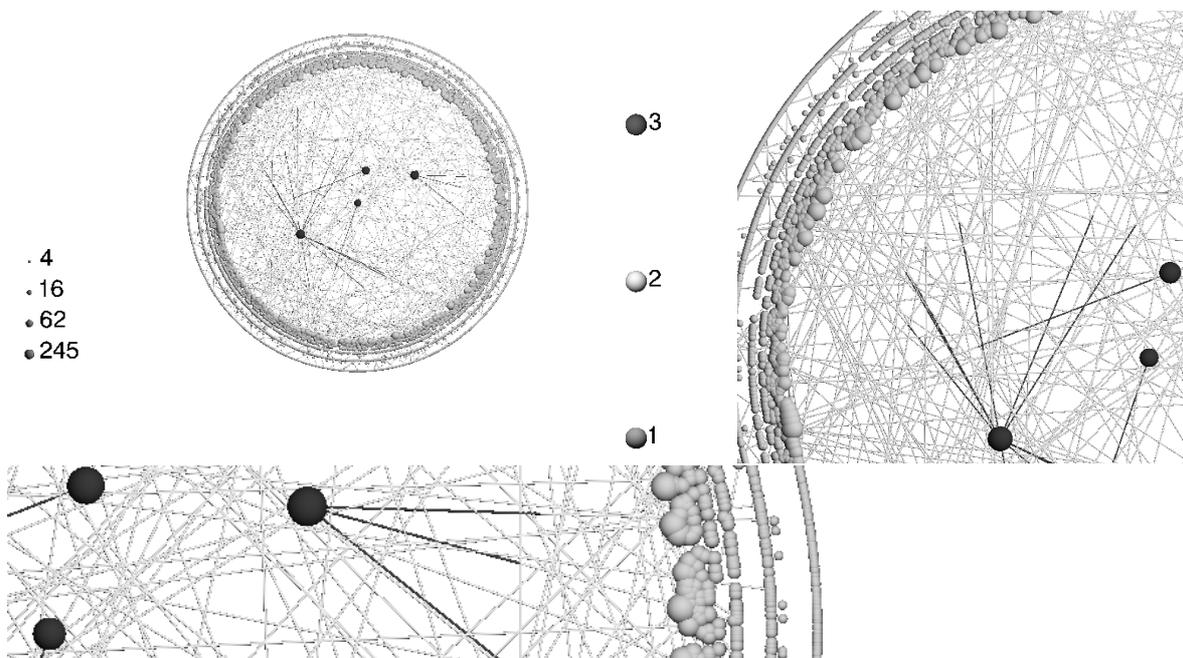


Figura 5: Visualización, utilizando la descomposición en k -núcleos, de un grafo de Barabási y Albert (1999), con 10000 vértices y $m = 2$.

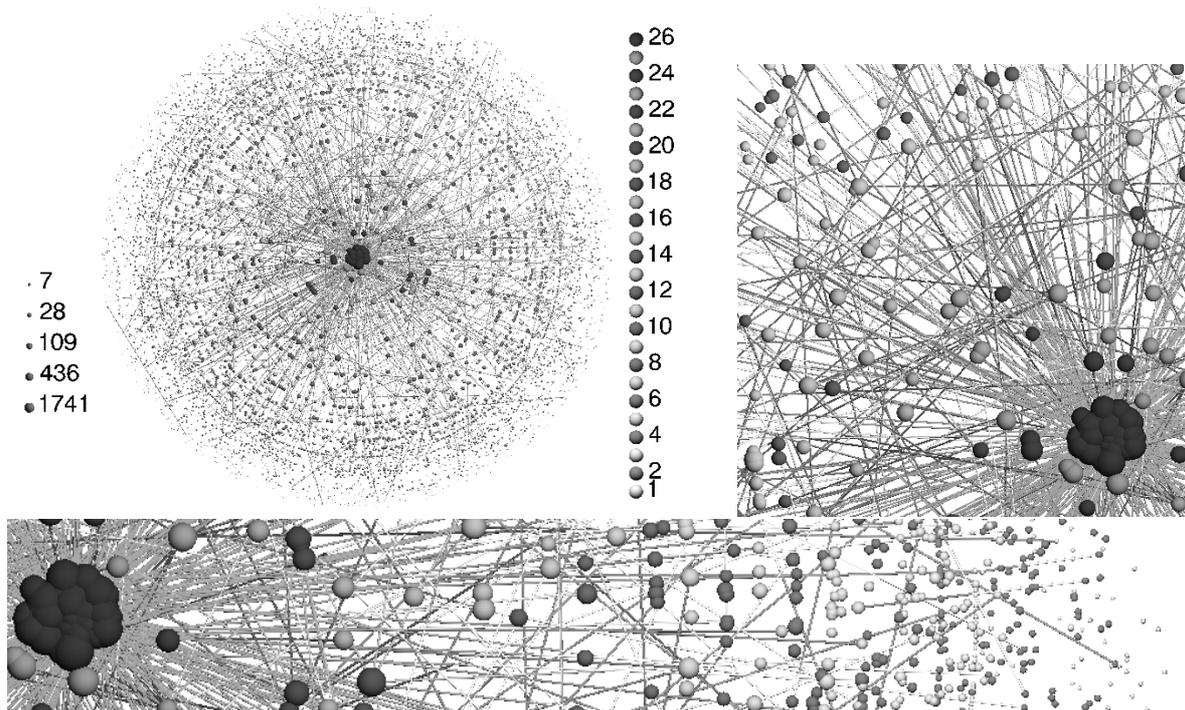


Figura 6: Visualización, utilizando la descomposición en k -núcleos, del mapa de AS obtenido por CAIDA (2003), en abril del 2005.

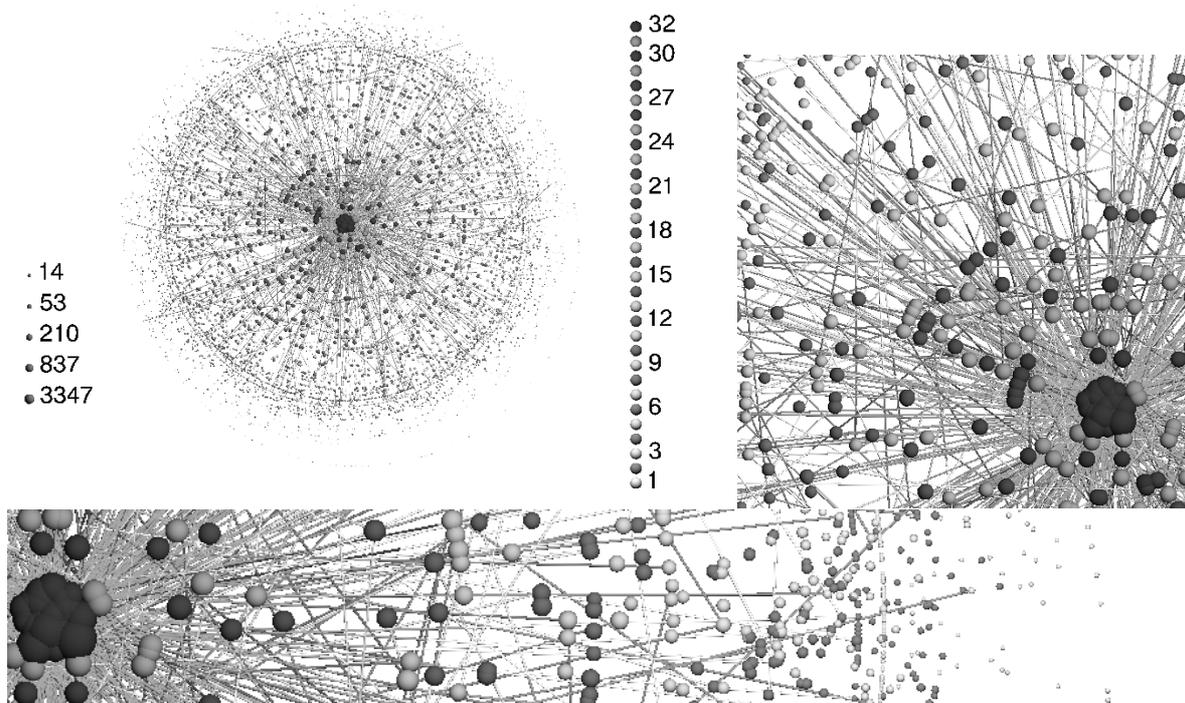


Figura 7: Visualización, utilizando la descomposición en k -núcleos, de un grafo obtenido según el modelo Serrano et al. (2005).

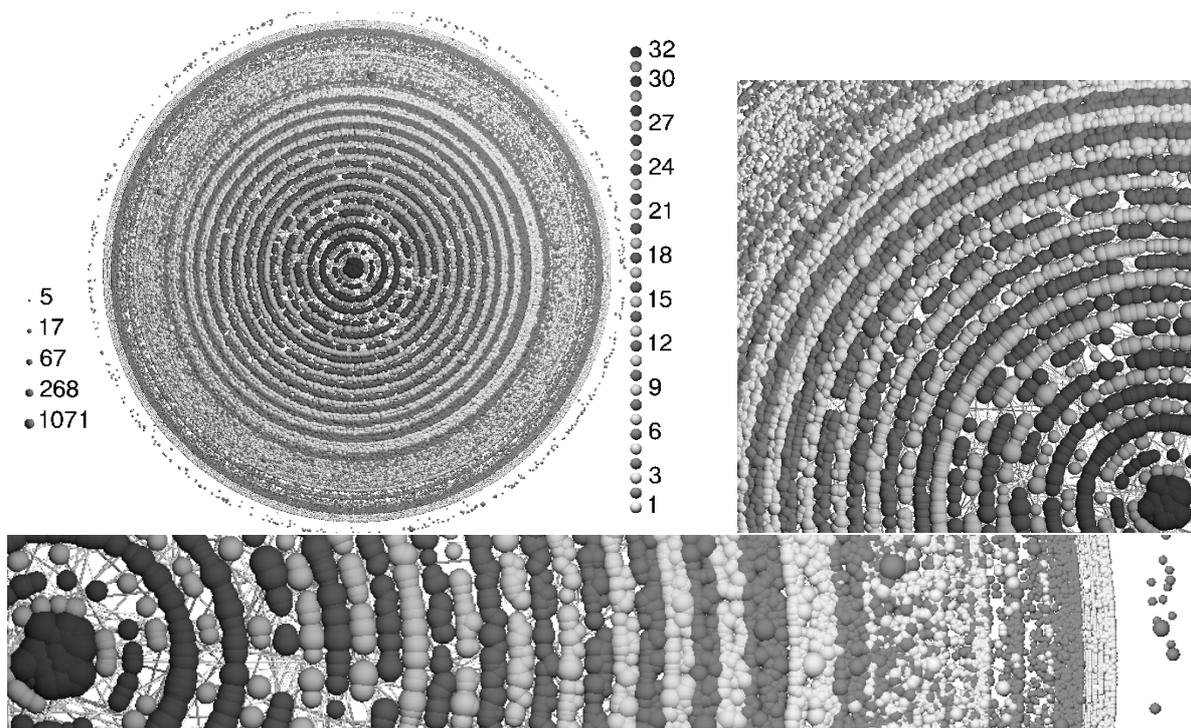


Figura 8: Visualización, utilizando la descomposición en k -núcleos, del mapa de ruteadores obtenido por CAIDA (2003), en el 2003.

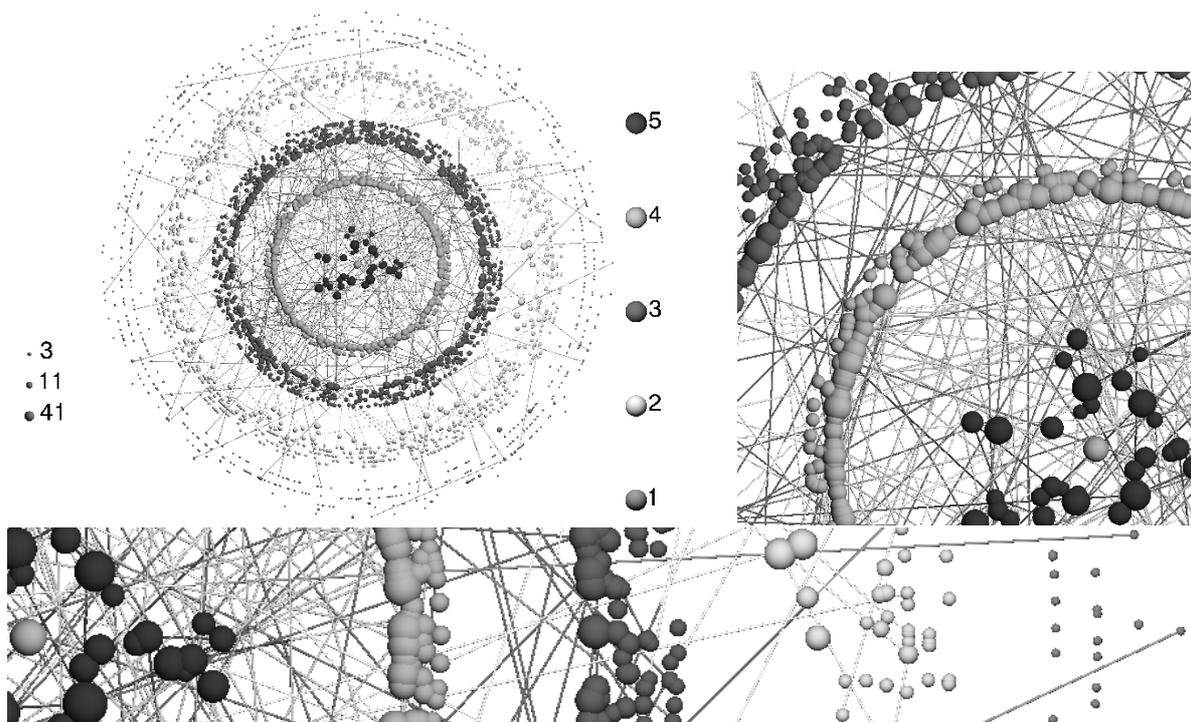


Figura 9: Visualización, utilizando la descomposición en k -núcleos, de un grafo obtenido según el modelo Alvarez-Hamelin y Schabanel (2004), con 1000 vértices.

5. CONCLUSIONES

La actividad en el modelado de la topología de Internet continúa siendo importante por varias razones. Podemos citar en primer lugar la obtención de mapas, ya sea a nivel de AS o de ruteadores. Estos mapas son sesgados e incompletos, por lo que hay que determinar en qué grado lo son, y de esta manera saber cuánto afectan a los parámetros de caracterización (Petermann y De Los Rios, 2004; Clauset y Moore, 2005; Dall'Asta et al., 2006).

En segundo lugar, si bien el modelado de los AS obtiene buenos resultados, todavía no está bien resuelto el nivel de ruteadores. Un desafío de mayor nivel es realizar un modelo conjunto, ya que los AS se obtienen a partir de los ruteadores, organizados según entidades administrativas.

Por último, podemos decir que muchos de los parámetros están relacionados con la teoría de grafos, y poco tienen que ver con el objetivo de estas redes. Recientemente se comenzó a utilizar la descomposición en k -núcleos, parámetro que está vinculado con el ruteo.

Es importante tomar en cuenta que esta actividad es asimismo cíclica: se realizan nuevas exploraciones, se analizan los resultados con los parámetros clásicos y nuevos parámetros, se realizan nuevos modelos y finalmente se validan los modelos utilizando la comparación con los mapas obtenidos por las exploraciones.

Agradecimientos. A Mariángels Serrano y Marián Boguñá por facilitar un grafo obtenido con su modelo, y por sus valiosas sugerencias. También a Ana Krichmar por sus apreciadas recomendaciones.

REFERENCIAS

- J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, y A. Vespignani. k -core decomposition: a tool for the analysis of large scale Internet graphs. *CoRR*, cs.NI/0511007, 2005.
- J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, y A. Vespignani. Large scale networks fingerprinting and visualization using the k -core decomposition. In Y. Weiss, B. Schölkopf, y J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 41–50, Cambridge, MA, 2006. MIT Press.
- J. I. Alvarez-Hamelin y N. Schabanel. An Internet Graph Model Based on Trade-Off Optimization. *Eur. Phys. J. B, special issue on "Applications of networks"*, 38(2):231–237, march II 2004.
- A.-L. Barabási y R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- V. Batagelj y M. Zaversnik. Generalized Cores. *CoRR*, cs.DS/0202039, 2002.
- B. Bollobás. The evolution of sparse graphs. *Graph Theory and Combinatorics*, 1984:35–57, 1983.
- CAIDA. Cooperative Association for Internet Data Analysis, 2003. URL <http://www.caida.org/tools/measurement/skitter/>.
- S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, y E. Shir. MEDUSA - New Model of Internet Topology Using k -shell Decomposition. *cond-mat/0601240*, 2006.
- A. Clauset y C. Moore. Accuracy and scaling phenomena in the Internet. *Phys. Rev. Lett.*, 94: 018701, 2005.
- L. Dall'Asta, J.I. Alvarez-Hamelin, A. Barrat, A. Vazquez, y A. Vespignani. Exploring networks with traceroute-like probes: theory and simulations. *Theor. Comput. Sci.*, 351(1):6–24, 2006.

- DIMES. Distributed Internet MEasurements & Simulations, 2005. URL <http://www.netdimes.org/science.html>.
- M. Doar. A better model for generating test networks. In *IEEE Globecom*, November 1996.
- P. Erdős y A. Rényi. On random graphs I. *Publ. Math. (Debrecen)*, 6:290–297, 1959.
- A. Fabrikant, E. Koutsoupias, y C. H. Papadimitriou. Heuristically Optimized Trade-Offs: A New Paradigm for Power Laws in the Internet. *LNCS*, 2380:110–, Jun 2002.
- M. Faloutsos, P. Faloutsos, y C. Faloutsos. On power-law relationship of the Internet topology. *Comput. Commun. Rev.*, 29:251–263, 1999.
- R. Govindan y H. Tangmunarunkit. Heuristics for Internet Map Discovery. In *IEEE INFOCOM 2000*, pages 1371–1380, Tel Aviv, Israel, March 2000. IEEE.
- J. Hawkinson y T. Bates. RFC-1930: Guidelines for creation, selection, and registration of an Autonomous System (AS). Technical report, IETF, March 1996.
- LaNet-vi. LARge NETwork VISualization tool, 2005. URL <http://xavier.informatics.indiana.edu/lanet-vi/>.
- L. Li, D. Alderson, R. Tanaka, J. C. Doyle, y W. Willinger. Towards a theory of scale-free graphs: Definition, properties, and implications (extended version). *Internet Math.*, 2(4): 431–523, 2005.
- A. Medina, I. Matta, y J. Byers. On the origin of power laws in internet topologies. *Computer Communications Review*, 30(2):18–28, apr 2000.
- M. E. J. Newman. Assortative Mixing in Networks. *Phys. Rev. Lett.*, 89:208701–, 2002.
- R. Pastor-Satorras y A. Vespignani. *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, 2004.
- T. Petermann y P. De Los Rios. Exploration of scale free networks. *Eur. Phys.J. B*, 38:201, 2004.
- Router-Views. University of Oregon, Router Views Project, 2005. URL <http://www.routeviews.org/>.
- M. A. Serrano, M. Boguñá, y A. Diaz-Guilera. Competition and adaptation in an Internet evolution model. *Phys. Rev. Lett.*, 94:038701, 2005.
- B. M. Waxman. Routing of Multipoint Connections. *IEEE Journal on Selected Areas in Communications*, 6(9):1617–1622, December 1988.
- K. Yamasaki, K. Matia, D. Fu, S. V. Buldyrev, F. Pammolli, M. Riccaboni, y H. E. Stanley. A Generalized Preferential Attachment Model for Complex Systems. *CoRR*, physics/0502082, 2005.
- S. H. Yook, H. Jeong, y A.-L. Barabási. Modeling the internet's large-scale topology. *Proc Natl Acad Sci USA*, 99(21):13382–6, Oct 2002.
- E. W. Zegura, K. L. Calvert, y M. J. Donahoo. A quantitative comparison of graph-based models for Internet topology. *IEEE/ACM Transactions on Networking*, 5(6):770–783, 1997.