



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

POPULARIDAD LOCAL EN GRAFOS DEL TIPO BARABÁSI-ALBERT:  
IMPLICANCIAS EN EL MODELO MULTIPARTITO NUMBER-IN-HAND

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

JAVIERA FRANCISCA URRUTIA ESPÍNDOLA

PROFESOR GUÍA:  
IVÁN RAPAPORT ZIMERMANN

PROFESOR CO-GUÍA:  
DANIEL REMENIK ZISIS

MEMBROS DE LA COMISIÓN:  
JOSÉ SOTO SAN MARTÍN

Agradece financiamiento de Fondecyt 1090156 y Fondecyt 1130061.

SANTIAGO DE CHILE  
AGOSTO 2013

RESUMEN DE LA MEMORIA  
PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL MATEMÁTICO  
POR: JAVIERA URRUTIA ESPÍNDOLA  
FECHA: 19/08/2013  
PROF. GUÍA: IVÁN RAPAPORT Z.

POPULARIDAD LOCAL EN GRAFOS DEL TIPO BARABÁSI-ALBERT:  
IMPLICANCIAS EN EL MODELO MULTIPARTITO NUMBER-IN-HAND

La presente memoria tiene como objetivo el estudio del problema de reconstrucción de redes en sistemas distribuidos que utilizan mensajes pequeños. Esto resulta de gran interés en la actualidad debido a la existencia de redes de gran tamaño, tales como la *World Wide Web* y las redes sociales, que no necesariamente se encuentran almacenadas en su totalidad en un solo lugar. Debido a esto surge el interés de estudiar aspectos particulares de la topología de las redes de este tipo, coloquialmente denominadas *Small World Networks*.

En primer lugar se elabora un protocolo donde todos los nodos o procesadores, utilizando su información local, colaboran para reconstruir el grafo completo intercambiando información en rounds. Este protocolo es capaz de reconstruir cualquier grafo en 2 rounds. En particular, funcionará utilizando una cantidad pequeña de información local para clases de grafos que exhiben una topología especial, que se denominará *Popularidad Local* (se dice que un grafo es *k-Localmente-Popular* si para todo nodo en la red, éste tiene a lo más  $k$  vecinos con grado mayor o igual a él).

Posteriormente, se expone el modelo desarrollado por Barabási y Albert, quienes motivados por simular de mejor forma la topología de redes que surgen en la actualidad, proponen un modelo que evoluciona en etapas discretas. Partiendo con un grafo pequeño, en cada etapa se agrega un nuevo nodo que se conecta con  $m$  nodos existentes en el grafo utilizando la regla *Preferential Attachment*, en la cual la probabilidad de conectarse con un nodo depende de su grado.

Luego se introduce la noción de *Popularidad-Local-Débil* para grafos aleatorios. Se analiza la topología de los árboles del tipo *Small World* utilizando el modelo de Barabási y Albert con parámetro  $m = 1$  y se demuestra que esta clase de grafos es  $\frac{9}{4}$ -*Débil-Localmente-Popular*. Por lo tanto, estos grafos se pueden reconstruir en 2 rounds utilizando información local de tamaño logarítmico.

Finalmente se estudia experimentalmente la topología de los grafos generales de Barabási y Albert y se conjetura que estos grafos son *k-Débil-Localmente-Populares* donde  $k$  sólo depende de  $m$  y no depende del tamaño del grafo. Ésto podría constituir un gran avance dado que permitiría reconstruir eficientemente las grandes redes de la actualidad.

*A mis padres,  
Sadie y Tito.*

## AGRADECIMIENTOS

En primer lugar quiero agradecer a mis padres, por su amor incondicional, dedicación y respeto. A mi padre por inculcarme su razonamiento, dedicarme siempre todo su tiempo y atención y por motivarme a entrar en el mundo de las matemáticas. Gracias a él elegí este camino y estaré siempre agradecida por ello. A mi madre por estar siempre disponible para conversar conmigo, ser mi psicóloga personal y hacerme ver muchas cosas que han sido fundamentales en mi crecimiento como persona.

Agradecer a mis profesores. A mi profesor guía Iván Rapaport, por su apoyo y ayuda durante todo este trabajo y por su paciencia y buena onda. A Daniel Remenik por su gran ayuda en el área probabilista de esta memoria, y también por su paciencia y muy buena disposición siempre. A José Soto por sus acotaciones y comentarios que ayudaron a mejorar este trabajo.

Agradecer a mi tía Sole, y a mis primas Beatriz y Maricruz por darme tantas veces asilo en su casa, por los buenos momentos en el living del depa, y por apoyarme siempre.

Agradecer a mis amigos que estuvieron presentes a lo largo de todo este proceso. A Vale por su apoyo incondicional, sus consejos, las conversaciones, los buenos momentos y carretes y sobre todo por estar siempre en los malos momentos y fracasos. A Mauro, por los buenos momentos, las sesiones de estudio y su apoyo ante fracasos académicos (como esa vez que fuimos al cine después de mi primer fracaso en Análisis). A Tomás y Fran, que a pesar de estar lejos siempre los siento cerca, se que me apoyan y que están presentes. A Vale Núñez por las entretenidas juntas en su departamento, por su comprensión y buena onda siempre. A Bruno por las conversaciones y consejos y por incentivar a seguir esta carrera.

Agradecer a mis compañeros y amigos de generación Andrés, Benjamín Palacios, Francisco y Contador con los que compartí interminables sesiones de estudio, por su apoyo y generosidad con sus conocimientos, por los buenos carretes y la buena onda siempre. A Pedro por poner siempre la cuota de humor, por sus locuras y buena onda. A Riffo por apoyarme siempre en éxitos y sobre todo en fracasos. A Benja Obando por ser una gran persona y enseñarme, quizás sin darse cuenta, valores importantes. A Amitai por su apoyo y generosidad y por su ayuda probabilista en este trabajo.

Agradecer a mis compañeros y amigos de la 435, Tish, Víctor, Muns, Alf, Kiko y Toño, por darme asilo en su oficina, los buenos carretes y la buena onda.

Agradecer también a los funcionarios del DIM, Oscar Mori, Gladys Cavallone y Eterin por su ayuda y buena disposición siempre.

Finalmente agradecer a Fondecyt por financiar este trabajo.

# Tabla de Contenido

Tabla de Contenido	v
<b>1. Introducción</b>	<b>1</b>
1.1. Organización	2
<b>2. El Modelo Multipartito Number-In-Hand</b>	<b>3</b>
2.1. Modelos Preliminares	3
2.2. Modelo Multipartito Number-In-Hand	4
<b>3. Problema de Reconstrucción</b>	<b>6</b>
3.1. Problema	6
3.1.1. Imposibilidad de Reconstrucción en 1 Round	7
3.2. Protocolo Reconstrucción	7
3.2.1. Protocolo	7
3.2.2. Correctitud	7
3.3. Popularidad Local	8
<b>4. El Modelo Barabási-Albert</b>	<b>9</b>
4.1. Modelos Preliminares	9
4.2. Modelo de Barabási-Albert	10
<b>5. Preferential Attachment y Popularidad Local: Resultados Analíticos para Árboles</b>	<b>12</b>
5.1. Árboles Preferential Attachment	12
5.1.1. Proceso	12
5.2. Análisis para $v_0$	13
5.2.1. Análisis	13
5.2.2. Urna de Pólya-Eggenberger	15
5.2.3. Cálculo	16
5.3. Análisis General	24
<b>6. Preferential Attachment y Popularidad Local: Resultados Experimentales para el Caso General</b>	<b>28</b>
6.1. Grafos de Barabási-Albert	28
6.1.1. Proceso	28
6.2. Simulaciones	29
6.2.1. Definiciones Previas	29

---

6.2.2. Resultados . . . . .	29
6.3. Análisis . . . . .	37
<b>7. Conclusión y Trabajo Futuro</b>	<b>38</b>
<b>Bibliografía</b>	<b>39</b>

# Capítulo 1

## Introducción

En la presente memoria se estudian los grafos del tipo *Small World* y la posibilidad de ser reconstruidos en un sistema distribuido. El estudio de las redes *Small World* está creciendo rápidamente dada la gran disponibilidad de información distribuida en las redes. Por ejemplo una red compleja con más de 3 millones de vértices es la *World Wide Web*, donde un vértice es un documento y un arco es un link que los conecta [1]. Otra red compleja de gran tamaño es la de citas en publicaciones científicas, donde los vértices representan papers y los arcos representan links a artículos citados por el paper [2].

En este trabajo se estudian grafos del tipo *Small World* generados por el modelo de Barabási y Albert [3]. Es de gran interés el estudio de este tipo de redes porque surgen en variados sistemas, como por ejemplo las redes sociales. Por lo tanto estudiar su topología y en particular la posibilidad de reconstruirlas usando una cantidad pequeña de información local, resulta interesante y útil.

En redes de gran tamaño, como las redes sociales, la información completa del grafo no necesariamente se encuentra almacenada en un solo lugar (se requeriría de mucho espacio). Sin embargo, ésta sí se encuentra repartida entre los nodos o personas. Resulta interesante saber si es o no posible reconstruir la red completa utilizando una pequeña cantidad de información aportada por cada uno de los usuarios.

Inspirados en ello, el problema en particular que se trata es el siguiente. Dada una red de nodos o procesadores que se interconectan de algún modo, donde los arcos no necesariamente representan links de traspaso de información entre ellos, se quiere reconstruir la red utilizando una pequeña cantidad de información local proporcionada por cada uno de los nodos o procesadores. En particular en el caso de redes del tipo *Small World*, cada persona sólo tiene conocimiento de su vecindad. En base a esto debe proporcionar información que ayudará a reconstruir toda la red.

Para resolver este problema se utiliza el modelo Number-In-Hand. Éste consiste en agentes donde cada uno conoce su propio input, y todos deben comunicarse y colaborar para responder una pregunta o realizar algún cálculo. Existen variadas formas en las cuales los jugadores se pueden comunicar. En particular, en este trabajo los agentes co-responderán a procesadores o nodos en una red o grafo, y se comunicarán utilizando el modelo de pizarra compartida.

Haciendo uso de este modelo, se expone un protocolo en 2 rounds que reconstruye cualquier grafo simple, no dirigido. En particular, se demuestra que este algoritmo funciona bien para cierta clase de grafos, los cuales inspiran la introducción de un nuevo concepto: *Popularidad Local* (se dice que un grafo es *k-Localmente-Popular* si para todo nodo en la

red, este tiene a lo más  $k$  vecinos con grado mayor o igual a él). El algoritmo funcionará con mensajes pequeños para las clases de grafos  $k$ -Localmente-Populares con  $k$  constante.

En la presente memoria se estudia la propiedad de *Popularidad-Local-Débil* (un grafo aleatorio es  $k$ -Débil-Localmente-Popular si para todo nodo en la red, el *valor esperado* de vecinos con grado mayor o igual a él es menor o igual a  $k$ ), en clases de grafos que se construyen utilizando la regla de *Preferential Attachment*, modelo expuesto por Barabási y Albert [3]. Se demuestra que los árboles de *Barabási-Albert* son  $\frac{9}{4}$ -Débil-localmente-populares, independiente del tamaño del grafo, y se conjetura que los grafos generales de *Barabási-Albert* también son  $k$ -Débil-Localmente-Populares, con una constante  $k$  que no depende del tamaño del grafo.

## 1.1. Organización

En el Capítulo 2 se exponen ciertos modelos distribuidos que han suscitado interés en los últimos años. En particular, se presenta el modelo Number-In-Hand.

Posteriormente, en el Capítulo 3 se presenta el Protocolo Reconstrucción en 2 Rounds utilizando una pizarra compartida. Se demuestra su correctitud y se introduce el concepto de *Popularidad Local*.

Luego en el Capítulo 4 se exponen ciertos modelos previamente utilizados para construir redes del tipo *Small World*. Se presenta el modelo de *Barabási-Albert*.

Luego en el Capítulo 5, se define el proceso mediante el cual se forman los árboles de *Barabási-Albert* y se demuestra que estos son 3-Débil-Localmente-Populares.

En el Capítulo 6 se exponen resultados experimentales para grafos de *Barabási-Albert* generales que permiten conjeturar que éstos también son  $k$ -Débil-Localmente-Populares, con  $k$  constante que no depende del tamaño del grafo.

Finalmente, en el Capítulo 7, se concluye y se dan las pautas para un futuro trabajo.



# Capítulo 2

## El Modelo Multipartito Number-In-Hand

Un sistema distribuido es una red interconectada de procesadores que generalmente se representa por un grafo donde los arcos corresponden a alguna relación entre los nodos y cada nodo representa un procesador. Como los nodos no poseen conocimiento global, surgen nuevas nociones de complejidad y variados modelos.

### 2.1. Modelos Preliminares

Bajo el paradigma de que la comunicación es mucho más lenta y costosa que los procesos locales, el análisis de complejidad se centra principalmente en el intercambio de mensajes entre los procesadores. En el modelo clásico *CONGEST* [4], donde la red es representada por un grafo cuyos nodos representan los procesadores y los arcos representan links entre procesadores, los  $n$  procesadores pueden mandar un mensaje en cada round de tamaño  $O(\log n)$ , por cada uno de los arcos adyacentes a él. De este modo, este modelo restringe sólo la cantidad de información que se envía a través de los arcos, centrando el análisis en los costos de comunicación.

Se han propuesto muchas variaciones al modelo *CONGEST* con el objetivo de enfocarse en otros aspectos de un sistema distribuido. Por ejemplo, Linial introdujo el modelo *LOCAL* [5, 4], enfocándose en el hecho de que la información que se posee es local. En este modelo la restricción en el tamaño de los mensajes se elimina de modo que cada nodo pueda mandar mensajes de tamaño ilimitado en cada round. Este modelo se centra en la pregunta "What cannot be computed locally?" [6]. Problemas difíciles como encontrar un recubrimiento mínimo no se pueden aproximar usando el modelo *LOCAL* [6].

En este trabajo, los arcos no necesariamente representarán canales de comunicación entre los procesadores. Así, la comunicación de cada procesador no está restringida únicamente a sus vecinos. Se trabajará en un modelo en el cual existe un grafo subyacente y la información que cada nodo posee se limita solamente a su vecindad. En este escenario, todos los nodos deben responder una pregunta sobre alguna propiedad del grafo, de manera colectiva. En [7] se introduce un nuevo modelo en el cual se le permite a cada nodo enviar una cantidad de información acotada por  $O(\log n)$ , donde  $n$  es el tamaño de la red. En este modelo cada nodo transmite un mensaje a una entidad central que debe ser capaz de entregar el output correcto, recopilando toda la información enviada por los nodos.

En [7] se analiza si una cantidad logarítmica de información local entregada por cada nodo es o no suficiente para responder preguntas estructurales básicas de la red, llegando a la conclusión de que en general la respuesta es negativa. Por ejemplo, se demuestra que preguntas simples como "¿Contiene  $G$  un cuadrado?", "¿Es el diámetro de  $G$  a lo más 3?", no se pueden responder utilizando este modelo. Sin embargo, ciertas clases de grafos sí se pueden reconstruir. Entre ellos los grafos planares, los grafos de treewidth acotado, y en general, los grafos  $k$  degenerados.

El modelo usado en [7] puede ser visto también del siguiente modo: dada una pregunta sobre el grafo, cada nodo escribe de manera simultánea un mensaje, de tamaño  $O(\log n)$ , en una zona global de memoria compartida, una pizarra. Luego, cada nodo debe ser capaz de responder la pregunta usando sólo el contenido de la pizarra. Esto inspira nuevos modelos, en los cuales se pueden relajar ciertas condiciones. En efecto, en [8] la condición de simultaneidad se relaja, y se permite que los nodos escriban en la pizarra de manera secuencial: esto les permite tomar en cuenta la información que ya existe en la pizarra al momento de crear su propio mensaje.

## 2.2. Modelo Multipartito Number-In-Hand

En general la Computación Distribuida es capturada de manera más precisa por modelos de comunicación con múltiples agentes o multipartitos. Por ejemplo, por el modelo Number-In-Hand, donde cada agente o procesador conoce su propio input (en contraste con el modelo Number-In-Forehead donde cada jugador conoce el input de todos los otros agentes excepto el suyo). El modelo Number-In-Hand no suscitó mucho interés inicial en el mundo de la complejidad comunicacional. Sin embargo, recientemente han surgido nuevas técnicas que han llevado avances en este modelo [9, 10]. En particular, en [11] se presentan cotas para el problema Distributed Task Allocation.

En la presente memoria se considerará el modelo de comunicación Multipartito Number-In-Hand, en el cual participan  $n$  agentes. La red interconectada se modelará como un grafo  $G = (V, E)$  simple, no dirigido y conexo, de tamaño  $n$ . Cada nodo  $v \in V$  tiene una etiqueta única  $ID(v) \in \{1, \dots, n\}$ . Cada nodo corresponde a un agente o procesador que conoce su  $ID$ , el número total de nodos en la red  $n$  y los  $ID$ 's de sus vecinos  $\{ID(u) | u \in N(v)\}$ , donde  $N(v)$  denota el conjunto de vecinos de  $v$ . Los agentes colaboran para computar una respuesta al problema en cuestión, asumiendo que todos los agentes o nodos tienen capacidad de proceso ilimitada.

Para poder resolver la tarea o pregunta en conjunto, los nodos o procesadores se deben comunicar. En este trabajo se asume comunicación síncrona en rounds. Se utiliza el modelo clásico de comunicación mediante una pizarra compartida en la cual los nodos escriben un mensaje que es visible para todos los demás. De este modo, en un round todos los nodos escriben de manera simultánea un mensaje en la pizarra utilizando su información local; y posteriormente todos los nodos leen la pizarra y realizan el proceso correspondiente. En este trabajo interesa medir:

- *Complejidad en Rounds*: Número de rounds de comunicación que se requieren, donde en cada round todos los nodos escriben un mensaje en la pizarra, de manera simultánea.

- *Tamaño de los Mensajes*: Número de bits que escriben los nodos en la pizarra en cada round.

Otro modelo de comunicación ampliamente difundido es aquel en el que cada nodo envía un mensaje a cada uno de sus vecinos a través de sus arcos, modelo en el cual se asume que los nodos inicialmente no saben nada del grafo excepto su tamaño. Este modelo resulta interesante cuando los arcos en el grafo subyacente representan interconexiones entre los procesadores. Sin embargo, en este trabajo los arcos no necesariamente representan conexiones entre los procesadores (por ejemplo los arcos pueden representar llamadas telefónicas).

# Capítulo 3

## Problema de Reconstrucción

En el presente capítulo se resolverá, de manera distribuida, el problema de Reconstrucción de Grafos utilizando el Modelo Multipartito Number-In-Hand.

En [7] se demostró que la clase de grafos  $k$ -degenerados se puede reconstruir utilizando el Modelo Number-In-Hand en 1 round con mensajes de tamaño  $O(\log n)$ . Ahora bien, para poder usar el protocolo que se expone en [7], es necesario conocer a priori la degenerancia del input, lo cual no es un requisito trivial de satisfacer. Con esto en mente, en este capítulo se busca un protocolo que reconstruya *cualquier* grafo simple, no dirigido, pero que para una clase particular de grafos  $k$ -degenerados, el protocolo funcione con mensajes de  $O(\log n)$ .

A continuación se presenta el problema que se resolverá en este capítulo. Y luego su imposibilidad de ser resuelto en 1 round, lo que motiva la búsqueda de una solución en 2 rounds.

### 3.1. Problema

El problema a tratar será el de la reconstrucción de la clase de grafos simples, no dirigidos de tamaño  $n$ . Se quiere generar un protocolo Number-In-Hand que, dado  $G$  como input, sea capaz de reconstruir  $G$  (es decir, que entregue como output su matriz de adyacencia). Y que además, si  $G$  es un árbol, funcione con mensajes de tamaño logarítmico.

Se utilizará el Modelo Multipartito Number-In-Hand, en el cual en cada round todos los nodos, de manera simultánea, podrán escribir un mensaje usando su información local (que consiste en su  $ID$  y los  $ID$ 's de sus vecinos). Al final de cada round todos los nodos podrán leer la pizarra, con capacidad de proceso y memoria ilimitada.

En lo que sigue se demostrará que este problema no puede ser resuelto usando el Modelo Number-In-Hand en 1 round y con mensajes de tamaño  $O(\log n)$  en el caso de los árboles.

Posteriormente, se presentará el protocolo en 2 rounds, se demostrará que este reconstruye cualquier grafo. En particular, se verá que para cierta clase de grafos este protocolo funciona con mensajes de tamaño logarítmico.

### 3.1.1. Imposibilidad de Reconstrucción en 1 Round

Primero se verá que no es posible encontrar un protocolo en 1 round que reconstruya cualquier grafo simple, no dirigido, y que además cuando el input sea un árbol este use mensajes de tamaño  $O(\log n)$ .

**Lema 3.1.** *Sea  $\mathcal{P}$  protocolo que resuelve el problema expuesto en 3.1, en 1 round. Se tiene que,  $\forall n \in \mathbb{N}$  existe un árbol  $T_n$ , tal que al menos uno de sus nodos envía un mensaje  $\Omega(\log n)$ .*

*Demostración.* Sea  $\mathcal{P}$  protocolo que resuelve el problema expuesto en 3.1, en 1 round y sea  $n \in \mathbb{N}$ . Supongamos, por contradicción, que todos los grafos de tamaño  $n$  envían sólo mensajes de tamaño  $O(\log n)$ . Así, cada grafo envía en total  $O(n \log n)$  bits. Esto no es suficiente para distinguir los  $2^{\frac{n(n-1)}{2}}$  grafos de tamaño  $n$ . Por lo tanto, existe al menos un grafo  $G_n$  con al menos un nodo  $v$  que envía un mensaje  $\Omega(\log n)$ . Recordemos que el mensaje que cada nodo manda sólo depende de su vecindad y del número total de nodos  $n$ . Por lo tanto, basta construir un árbol  $T_n$ , con un nodo con la misma vecindad que  $v$  en  $G_n$  y tamaño  $n$ .  $\square$

## 3.2. Protocolo Reconstrucción

En la presente sección se presenta un protocolo que reconstruye los grafos simples y no dirigidos, en 2 rounds, utilizando el Modelo Multipartito Number-In-Hand.

**Definición 3.2.** Sea  $G = (V, E)$  grafo simple, no dirigido y sea  $v \in V$  nodo cualquiera. Denotaremos  $N(v)$  al conjunto de vecinos de  $v$  y  $d_G(v)$  al grado de  $v$  en  $G$ .

### 3.2.1. Protocolo

Sea  $G = (V, E)$ ,  $V = \{v_0, v_1, v_2, \dots, v_n\}$ .  $G$  grafo simple, no dirigido.

- Round 1. Cada nodo  $v_i$  escribe un mensaje en la pizarra que se compone de su propio  $ID = i$ , y su grado  $d_G(v_i)$ . Luego de esto, cada nodo  $v_i$  lee la pizarra y recuerda la información escrita por sus vecinos  $N(v_i)$ .
- Round 2. Cada nodo  $v_i$  escribe un mensaje que se compone de los  $ID$ 's de sus vecinos que tengan grado mayor o igual a  $v_i$ .

### 3.2.2. Correctitud

**Proposición 3.3.** *El Protocolo 3.2.1 reconstruye cualquier grafo  $G = (V, E)$  simple y no dirigido.*

*Demostración.* Sea  $G = (V, E)$  simple y no dirigido. Corremos el algoritmo en  $G$ . Al terminar el round 1 cada nodo sabrá la cantidad de nodos  $n$  en la red. Es decir, se conocerá  $V$ . Sea  $uv \in E$ . Claramente, o bien  $d_G(u) \geq d_G(v)$ , o bien  $d_G(v) \geq d_G(u)$ . Por lo tanto,  $uv$  aparecerá, al menos una vez, en la pizarra en el round 2. Esto se tiene para todo  $e \in E$ , con lo cual todos los nodos son capaces de generar la matriz de adyacencia de  $G$ .  $\square$

### 3.3. Popularidad Local

El Protocolo 3.2.1 inspira una nueva noción, a la que llamaremos *Popularidad Local*, que se define a continuación.

**Definición 3.4.** Sea  $G = (V, E)$  grafo simple, no dirigido,  $V = \{v_0, v_1, \dots, v_n\}$ . Sea  $v_i \in V$ . Denotaremos  $m_i$  al número de vecinos de  $v_i$  que tienen grado mayor o igual que el grado de  $v_i$ .

**Definición 3.5.** Sea  $G = (V, E)$  grafo simple, no dirigido. Diremos que  $G$  es  $k$ -*Localmente-Popular* si  $\forall v \in V, m_i \leq k$ .

Se demostró que el Protocolo 3.2.1 reconstruye cualquier grafo  $G$  simple, no dirigido. Sin embargo, puede ocurrir que los mensajes en el round 2 sean largos para cierto tipo de grafos. Ahora bien, en el caso de los grafos  $k$ -*Localmente-Populares*, los mensajes en el round 2 resultan de tamaño  $O(\log n)$ , donde  $n$  es el tamaño del grafo.

Esto inspira el estudio de los grafos  $k$ -*Localmente-Populares*, noción que lleva a pensar inmediatamente en grafos del tipo *Small World*. Es por eso que en el Capítulo 5 se analiza una versión *débil* de esta propiedad en los árboles de Barabási-Albert.

*Observación 3.6.* No todos los árboles son  $k$ -*Localmente-Populares*. Por ejemplo, un árbol construido por capas, en el cual los nodos de la capa  $n$  tienen  $n + 1$  vecinos, los de la capa  $n + 1$  tienen  $n + 2$  vecinos, etc. No es  $k$ -*Localmente-Popular*, para ningún  $k$ . Como se muestra en la Figura 3.1.

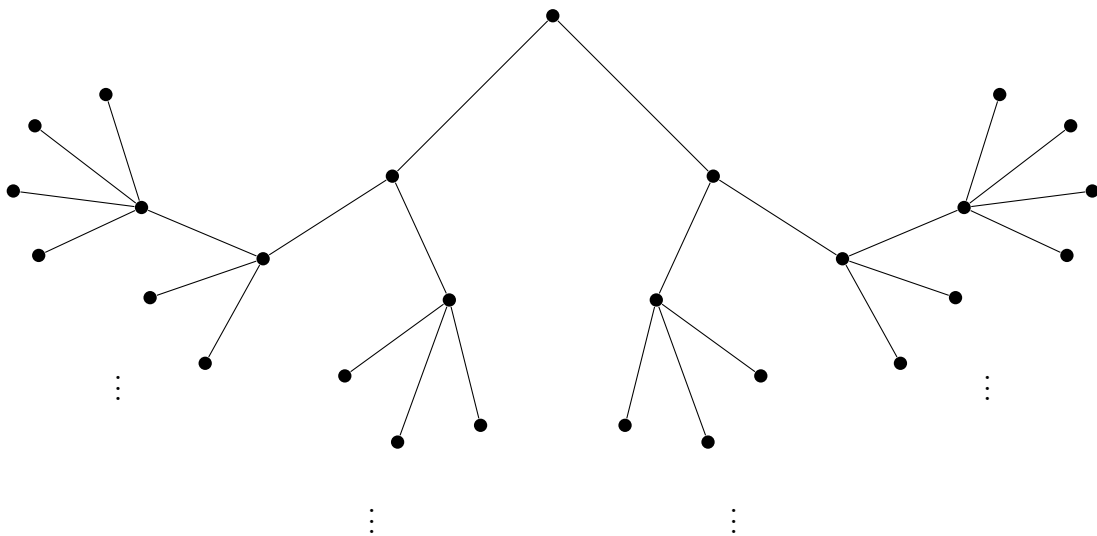


Figura 3.1

# Capítulo 4

## El Modelo Barabási-Albert

Se cree que casi cualquier par de personas en el mundo se pueden conectar a través de una cadena corta de conocidos intermedios, típicamente de largo 6. Este fenómeno, coloquialmente llamado *6 grados de separación* [12], ha sido sujeto de muchísimo estudio reciente en la comunidad física y matemática. A los grafos que exhiben esta característica se les llama *Small World Networks* [13].

Redes aleatorias de este tipo, con topologías complejas, son comunes en la naturaleza, describiendo sistemas tan diversos como la red de páginas web [1], redes de publicaciones científicas [2] o redes de neuronas [14]. Recientemente se ha demostrado que muchas de las grandes redes para las cuales su información topológica es conocida, exhiben una distribución de grados que corresponde a una potencia.

### 4.1. Modelos Preliminares

Probablemente el modelo más antiguo e investigado de redes aleatorias fue introducido por Erdős y Rényi [15]. En este modelo se tienen  $N$  nodos en la red y cada par de nodos se conecta con un arco con probabilidad  $p_{er}$ , generando una red aleatoria.

Otro modelo que ha sido objeto de atención últimamente es el modelo propuesto por Watts y Strogatz [13]. En este modelo la red comienza con un anillo de  $N$  nodos, donde cada nodo está conectado a sus vecinos más cercanos. Luego cada arco es modificado con probabilidad  $p_{ws}$ , donde la modificación consiste en cambiar uno de los nodos por otro elegido al azar, con la restricción de que no puede haber más de un arco entre cada par de nodos.

Una característica común a estos modelos es que ambos predicen que la distribución de la cantidad de nodos de grado  $k$ ,  $P(k)$ , corresponde a una potencia con cutoff exponencial. Es decir tiene la forma  $P(k) = k^\gamma e^{-\alpha k}$ . Sin embargo, como se mencionó anteriormente, muchos sistemas en la naturaleza poseen una distribución de grados que corresponde a una potencia, es decir  $P(k) = k^\gamma$ . Para entender el origen de esta discrepancia, Barabási y Albert proponen un nuevo modelo que se acerca más al comportamiento de estas redes en la naturaleza.

## 4.2. Modelo de Barabási-Albert

Los dos modelos anteriores asumen que la red comienza con un número fijo de nodos  $N$  que se conectan o reconectan de manera aleatoria, sin modificar  $N$  en el tiempo. Sin embargo, muchas redes en la naturaleza son abiertas. Es decir, continuamente se están agregando nuevos nodos y  $N$  crece con el tiempo (como por ejemplo la red de las páginas en internet). Consecuentemente, Barabási y Albert proponen un modelo que crece con el tiempo.

Por otro lado, los modelos anteriores asumen que la probabilidad de que dos nodos estén conectados es uniforme, en contraste con muchas redes de la naturaleza, en las cuales un nodo que se agrega a la red tiene más posibilidades de conectarse con un nodo que ya tiene varios vecinos. Por ejemplo una nueva publicación científica tiene más posibilidades de citar a una publicación que ya es reconocida y citada que a una desconocida. A este comportamiento se le denominó *Preferential Attachment* [16].

Teniendo esto en mente Barabási y Albert proponen un modelo que incorpora estos dos ingredientes y demuestran analíticamente que la distribución de los grados en un grafo generado con este modelo tiene en efecto una distribución de potencia, es decir tiene la forma  $P(k) = k^{-\gamma}$ , donde  $P(k)$  denota la cantidad de nodos de grado  $k$  [3].

### Crecimiento

Se parte con un grafo  $G^0$  simple, no dirigido, con un número pequeño de nodos  $m_0$ . En cada paso se agrega un nuevo nodo con  $m \leq m_0$  arcos que se conectarán con nodos que ya pertenecen al grafo.

### Preferential Attachment

**Definición 4.1.** Sea  $G = (V, E)$  un grafo simple, no-dirigido. Sea  $v \in V$ , se denotará  $d_G(v)$  al grado de  $v$  en  $G$ .

Cuando se eligen los nodos con los cuales el nuevo nodo se conecta, se asume que la probabilidad de que un nuevo nodo se conecte con el nodo  $v_i$  depende del grado de  $v_i$ .

El proceso de crecimiento del grafo evoluciona en tiempos discretos. Su estado en cada tiempo  $n \geq 0$  es un grafo conexo, simple, no-dirigido  $G^n = (V^n, E^n)$ , cuyo conjunto de vértices es  $V^n$  y de arcos es  $E^n$ , y en cada etapa se agrega un nodo con  $m$  arcos distintos.

A la llegada del  $n$ -ésimo nodo, se eligen sus  $m$  vecinos del siguiente modo:

1. Elegimos un  $w_n^1 \in V^{n-1}$  de acuerdo a la siguiente distribución de probabilidad,

$$\forall v \in V^{n-1}, \quad \mathbb{P}(w_n^1 = v | G^{n-1}) = \frac{d_{G^{n-1}}(v)}{\sum_{u \in V^{n-1}} d_{G^{n-1}}(u)}.$$

2. Elegimos un  $w_n^2 \in V^{n-1} \setminus \{w_n^1\}$  de acuerdo a la siguiente distribución de probabilidad,

$$\forall v \in V^{n-1} \setminus \{w_n^1\}, \quad \mathbb{P}(w_n^2 = v | G^{n-1}) = \frac{d_{G^{n-1}}(v)}{\sum_{u \in V^{n-1} \setminus \{w_n^1\}} d_{G^{n-1}}(u)}.$$



3. Elegimos un  $w_n^3 \in V^{n-1} \setminus \{w_n^1, w_n^2\}$  de acuerdo a la siguiente distribución de probabilidad,

$$\forall v \in V^{n-1} \setminus \{w_n^1, w_n^2\}, \quad \mathbb{P}(w_n^3 = v | G^{n-1}) = \frac{d_{G^{n-1}}(v)}{\sum_{u \in V^{n-1} \setminus \{w_n^1, w_n^2\}} d_{G^{n-1}}(u)}.$$

⋮

- $m$ . Elegimos un  $w_n^m \in V^{n-1} \setminus \{w_n^1, w_n^2, \dots, w_n^{m-1}\}$  de acuerdo a la siguiente distribución de probabilidad,

$$\forall v \in V^{n-1} \setminus \{w_n^1, w_n^2, \dots, w_n^{m-1}\}, \quad \mathbb{P}(w_n^m = v | G^{n-1}) = \frac{d_{G^{n-1}}(v)}{\sum_{u \in V^{n-1} \setminus \{w_n^1, w_n^2, \dots, w_n^{m-1}\}} d_{G^{n-1}}(u)}.$$

Finalmente se agrega el nuevo nodo  $v_n$  y los nuevos arcos:

$$V^n = V^{n-1} \cup \{v_n\} \text{ y } E^n = E^{n-1} \cup \{w_n^1 v_n, w_n^2 v_n, \dots, w_n^m v_n\}.$$

# Capítulo 5

## Preferential Attachment y Popularidad Local: Resultados Analíticos para Árboles

En el presente capítulo se presentarán resultados analíticos para la clase de árboles construídos en base a la regla de *Preferential Attachment*. En particular se introducirá la noción de *Popularidad-Local-Débil* para grafos aleatorios y se demostrará que los árboles construídos utilizando el modelo de *Barabási-Albert* son  $\frac{9}{4}$ -*Débil-Localmente-Populares*.

### 5.1. Árboles Preferential Attachment

Los árboles en esta clase se construyen mediante un proceso discreto en el cual en cada paso se agrega un nodo al árbol, usando la regla de *Preferential Attachment*.

A continuación se define la notación y conceptos que se utilizarán y luego el proceso mediante el cual los árboles se construyen.

**Definición 5.1.** Sea  $T = (V, E)$  un árbol simple, no-dirigido y sea  $v \in V$ . Se denotará  $N(v)$  al conjunto de vecinos de  $v$  y  $d_T(v)$  al grado de  $v$  en  $T$ .

#### 5.1.1. Proceso

El proceso de crecimiento del árbol evoluciona en tiempos discretos. Su estado en cada tiempo  $n \geq 0$  es un árbol simple, conexo y no-dirigido  $T^n = (V^n, E^n)$  con raíz  $r$ , cuyo conjunto de vértices es  $V^n$  y de arcos es  $E^n$ . Inicialmente  $T^0$  contiene a la raíz  $r$  y a un nodo  $v_0$ . Ambos se unen a través de un arco  $rv_0$ , i.e.,  $V_0 = \{r, v_0\}$  y  $E_0 = \{rv_0\}$ . En el proceso de crecimiento del árbol no participará la raíz  $r$  (lo que se requiere es que  $v_0$  comience el proceso con grado no nulo), como se verá a continuación.

Sea  $n \geq 1$  y  $T^{n-1} = (V^{n-1}, E^{n-1})$ . Con  $V^0 = \{r, v_0\}$ ,  $E^0 = \{rv_0\}$ .

1. Elegimos un único  $w_n \in V^{n-1} \setminus \{r\}$  de acuerdo a la siguiente distribución de probabilidad,

$$\forall v \in V^{n-1} \setminus \{r\}, \quad \mathbb{P}(w_n = v | T^{n-1}) = \frac{d_{T^{n-1}}(v)}{\sum_{u \in V^{n-1} \setminus \{r\}} d_{T^{n-1}}(u)}$$

$$2. V^n = V^{n-1} \cup \{v_n\} \text{ y } E^n = E^{n-1} \cup \{w_n v_n\}.$$

Claramente, a lo largo de este proceso,  $T^n$  es un árbol  $\forall n \in \mathbb{N}$  (dado que en cada paso  $n$ , un nuevo nodo  $v_n$  se agrega al árbol  $T^{n-1}$  uniéndose a un solo nodo  $w_n$  que ya pertenecía al árbol).

## 5.2. Análisis para $v_0$

En esta sección se analizará lo que sucede con el nodo  $v_0$ . Para luego, en la Sección 5.3, extender los resultados de  $v_0$  a todo el árbol.

El proceso de crecimiento del árbol dinámico descrito en 5.1.1 tiene inherentemente asociado un proceso estocástico, que se define a continuación.

**Definición 5.2.** Sea  $T^N = (V^N, E^N)$  el árbol construido de acuerdo a lo definido en 5.1.1. Sea  $n, i \in \mathbb{N}, n, i \leq N$ . Se denotará  $X_{i,n}$  al número de vecinos del nodo  $v_i$  en el paso  $n$  del proceso de construcción del árbol. Esto define un arreglo de variables aleatorias  $\{X_{i,n}\}_{0 \leq i, n \leq N}$ , para las cuales imponemos que si  $n < i$ ,  $X_{i,n} = 0$ . Es decir, en tiempos anteriores al tiempo de llegada del nodo  $v_i$ , se dice que  $v_i$  tiene 0 vecinos.

**Definición 5.3.** Sea  $T^N = (V^N, E^N)$  el árbol construido de acuerdo a lo definido en 5.1.1. Sea  $n, i \in \mathbb{N}, n, i \leq N$ . Se denotará  $M_{i,n}$  al número de vecinos con grado mayor o igual a  $v_i$  en el paso  $n$  del proceso de construcción del árbol. Esto define un arreglo de variables aleatorias  $\{M_{i,n}\}_{0 \leq i, n \leq N}$ , para las cuales imponemos que si  $n < i$ ,  $M_{i,n} = 0$ . Es decir en tiempos anteriores al tiempo de llegada del nodo  $v_i$ , se dice que  $v_i$  tiene 0 vecinos con grado mayor o igual a él.

**Definición 5.4.** Sea  $T^N = (V^N, E^N)$  el árbol construido de acuerdo a lo definido en 5.1.1. Diremos que  $T$  es *k-Débil-Localmente-Popular* si  $\forall v \in V^N, \mathbb{E}(M_{i,n}) \leq k, \forall i, n \leq N$ .

En esta sección se quiere analizar qué sucede con  $\mathbb{E}(M_{0,n})$  definida en 5.3. Es decir, se desea calcular el valor esperado de la cantidad de vecinos que tienen grado mayor o igual a  $v_0$ . Para luego, en la Sección 5.3 extender este resultado a todos los nodos en el árbol.

Desde el punto de vista de las redes sociales, se quiere calcular, para cualquier individuo en la red, el valor esperado de la cantidad de amigos que son tan o más populares que él.

Se verá que, en esperanza,  $v_0$  es popular dentro de su grupo de amistades. En particular, se demostrará que  $\mathbb{E}(M_{0,n}) \leq 2, \forall n \in \mathbb{N}$ . Es decir, se demostrará que, el número esperado de vecinos de  $v_0$  con grado mayor o igual a él es  $\leq 2$ , a través de todo el proceso.

### 5.2.1. Análisis

**Definición 5.5.** Sea  $T^N = (V^N, E^N), N \in \mathbb{N}$  el árbol dinámico construido de acuerdo a 5.1.1. Se definen las siguientes variables aleatorias,

- $\{N_k\}_{k \in \mathbb{N}}$ , donde  $N_k$  es el instante en que el  $k$ -ésimo vecino de  $v_0$  llega al árbol. Si  $k \geq N$ ,  $N_k = \infty$ .
- $\{d_n^0\}_{n \in \mathbb{N}}$ , donde  $d_n^0$  es el grado del nodo  $v_0$  en el paso  $n$  del proceso 5.1.1. En  $n = 0$ ,  $d_0^0 = 1$ .

- $\{d_n^k\}_{n \in \mathbb{N}}$ ,  $k \in \mathbb{N}$ , donde  $d_n^k$  es el grado del  $k$ -ésimo vecino de  $v_0$  en el paso  $n$  del proceso 5.1.1. Si en el tiempo  $n$  el  $k$ -ésimo vecino de  $v_0$  aún no ha llegado al árbol, imponemos  $d_n^k = 0$ .
- $\{S_n^k\}_{n \in \mathbb{N}}$ ,  $k \in \mathbb{N}$ , donde  $S_n^k = \mathbb{I}_{\{d_n^k \geq d_n^0\}}$ .

Para  $\mathbb{E}(M_{0,n})$ ,  $n \in \mathbb{N}$ , definida en 5.3, se tiene lo siguiente:

$$\begin{aligned} \mathbb{E}(M_{0,n}) &= \mathbb{E} \left( \sum_{k=1}^{d_n^0} S_n^k \right) \\ &= \mathbb{E} \left( \sum_{k=1}^n S_n^k \right) \\ &= \sum_{k=1}^n \mathbb{E}(S_n^k) \end{aligned}$$

Por lo tanto, el problema de acotar  $\mathbb{E}(M_{0,n})$  se reduce a acotar  $\mathbb{E}(S_n^k)$ . Para esto se define lo siguiente.

**Definición 5.6.** Sea  $T^N = (V^N, E^N)$ ,  $N \in \mathbb{N}$  el árbol dinámico construido de acuerdo a 5.1.1 y sean  $u, v \in V^N$ . Se denotará  $T_{uv}^n$  al árbol inducido por los nodos  $u, v$  y sus vecinos hasta el tiempo  $n$ . Si  $u$  o  $v$  aún no han llegado al árbol en el tiempo  $n$ ,  $T_{uv}^n = \phi$ .

**Definición 5.7.** Sea  $T^N = (V^N, E^N)$ ,  $N \in \mathbb{N}$  el árbol dinámico construido de acuerdo a 5.1.1. Sea  $k \in \mathbb{N}$ , se denotará  $u_k$  al  $k$ -ésimo vecino de  $v_0$ . Se definen los siguientes tiempos de parada,

- $\sigma_0^k = N_k$ .
- $\sigma_{m+1}^k = \inf_{n \geq \sigma_m^k} \{T_{v_0 u_k}^n \subsetneq T_{v_0 u_k}^{\sigma_m^k}\}$   $m \in \mathbb{N}, m \leq N$ .

Los tiempos de parada definidos en 5.7 corresponden a los tiempos de llegada de vecinos a  $v_0$  o a  $u_k$  luego de la llegada de  $u_k$ . Estos son los tiempos de interés para calcular el valor de  $S_n^k$ .

**Proposición 5.8.** Sea  $T^N = (V^N, E^N)$ ,  $N \in \mathbb{N}$  el árbol dinámico construido de acuerdo a 5.1.1. Sea  $k \in \mathbb{N}$ ,  $M \in \mathbb{N}$ .

$$\mathbb{E}(S_{\sigma_n^k}^k) \leq M, \quad \forall n \in \mathbb{N} \implies \mathbb{E}(S_n^k) \leq M, \quad \forall n \in \mathbb{N}$$

*Demostración.* Si  $\sigma_n^k \leq m < \sigma_{n+1}^k$ , entonces  $S_m^k = S_{\sigma_n^k}^k$ . Es decir la variable aleatoria  $\{S_n^k\}_{n \in \mathbb{N}}$  sólo puede cambiar de valor en los tiempos  $\sigma_n^k$ , por lo tanto basta demostrar una cota para  $\mathbb{E}(S_{\sigma_n^k}^k)$ ,  $n \in \mathbb{N}$ .  $\square$

Con el objetivo de encontrar una cota para  $\mathbb{E}(S_{\sigma_n^k}^k)$  se introducirán a continuación las Urnas de Pólya-Eggenberger.

### 5.2.2. Urna de Pólya-Eggenberger

Una urna de *Pólya-Eggenberger* inicialmente contiene  $r$  bolas rojas y  $b$  bolas blancas. En cada instante se extrae una bola, luego se repone y se agregan  $c$  bolas adicionales del mismo color.

**Definición 5.9.** Se denotará  $\{J_n\}_{n \in \mathbb{N}}$  a la fracción de bolas blancas luego de la  $n$ -ésima extracción.

**Proposición 5.10.**  $\{J_n\}_{n \in \mathbb{N}}$  es una martingala.

*Demostración.* En efecto, sea  $\mathcal{F}_n = \sigma(J_0, J_1, \dots, J_n)$  y supongamos que en el tiempo  $n$  hay  $i$  bolas rojas y  $j$  bolas blancas en la urna. Entonces,

$$J_{n+1} = \begin{cases} \frac{j+c}{i+j+c} & \text{con probabilidad } \frac{j}{i+j} \\ \frac{j}{i+j+c} & \text{con probabilidad } \frac{i}{i+j} \end{cases}$$

y con esto se tiene lo siguiente,

$$\begin{aligned} \mathbb{E}(J_{n+1} | \mathcal{F}_n) &= \frac{j+c}{i+j+c} \cdot \frac{j}{i+j} + \frac{j}{i+j+c} \cdot \frac{i}{i+j} \\ &= \frac{(j+c+i)j}{(j+c+i)(i+j)} \\ &= \frac{j}{i+j} \\ &= J_n. \end{aligned}$$

□

Como  $J_n \geq 0$ , por el teorema de convergencia de martingalas [17], se tiene que existe  $J_\infty$ , variable aleatoria límite de  $J_n$ . Además, gracias al resultado de Eggenberger y Pólya [18], se sabe que,

$$J_\infty \sim \beta\left(\frac{b}{c}, \frac{r}{c}\right) \quad (5.1)$$

donde  $\beta\left(\frac{b}{c}, \frac{r}{c}\right)$  denota la distribución de la Beta con parámetros  $\frac{b}{c}$  y  $\frac{r}{c}$ , cuya densidad es,

$$f_{\beta\left(\frac{b}{c}, \frac{r}{c}\right)} = \frac{\Gamma\left(\frac{b+r}{c}\right)}{\Gamma\left(\frac{b}{c}\right)\Gamma\left(\frac{r}{c}\right)} x^{\frac{b}{c}-1} (1-x)^{\frac{r}{c}-1}.$$

**Definición 5.11.** Se denotará  $\{R_n\}_{n \in \mathbb{N}}$  al número de bolas rojas en la urna luego del paso  $n$ . Del mismo modo, se denotará  $\{B_n\}_{n \in \mathbb{N}}$  al número de bolas blancas luego del paso  $n$ .

**Definición 5.12.** Se denotará  $\{\tilde{B}_n\}_{n \in \mathbb{N}}$  al número de extracciones de color blanco, hasta el paso  $n$ . Notar que  $\tilde{B}_n = B_n - 1$ .

**Proposición 5.13.** *Sea  $n \in \mathbb{N}$  fijo. En una urna de Pólya-Eggenberger que comienza con  $b$  bolas blancas y  $r$  bolas rojas, s.p.g.  $b \leq r$ . Se tiene que,*

$$\mathbb{P}(B_n \geq R_n) = \sum_{i=(r-b)+\lceil \frac{n-(r-b)}{2} \rceil}^n \frac{b(b+1) \cdots (b+i-1)r(r+1) \cdots (r+n-i-1)}{(b+r)(b+r+1) \cdots (b+r+n-1)} \cdot \binom{n}{i}$$

*Demostración.* La probabilidad de que las primeras  $i$  extracciones correspondan a bolas blancas y las siguientes  $n-i$  a bolas rojas es la siguiente,

$$\frac{b(b+1) \cdots (b+i-1)r(r+1) \cdots (r+n-i-1)}{(b+r)(b+r+1) \cdots (b+r+n-1)}$$

notemos que cualquier otro orden de extracción de  $i$  bolas blancas y  $n-i$  bolas rojas es sólo una permutación de los numeradores, con el mismo denominador y hay  $\binom{n}{i}$  órdenes posibles. Con esto se tiene que,

$$\mathbb{P}(\tilde{B}_n = i) = \frac{b(b+1) \cdots (b+i-1)r(r+1) \cdots (r+n-i-1)}{(b+r)(b+r+1) \cdots (b+r+n-1)} \binom{n}{i}.$$

Así, basta con sumar sobre los  $i$  que resultan en más bolas blancas que rojas en el paso  $n$  para obtener lo deseado.  $\square$

### 5.2.3. Cálculo

Se quiere analizar la interacción entre el nodo  $v_0$  y su vecino  $u_k$ , a partir del momento de llegada de su  $k$ -ésimo vecino,  $u_k$ .

Así la interacción entre el nodo  $v_0$  y su vecino  $u_k$  en los tiempos  $\sigma_n^k$  se modela como una urna de Pólya-Eggenberger, en la cual inicialmente hay  $r = k + 1$  bolas rojas, que representan los vecinos que tiene  $v_0$  al comienzo de la interacción y  $b = 1$  bolas blancas, que representa el hecho de que  $u_k$  tiene a  $v_0$  como único vecino.

En cada paso, se extrae una bola de la urna, se repone y se agrega una bola del mismo color. Es decir, el proceso corresponde a una urna de Pólya-Eggenberger con  $c = 1$ .

**Definición 5.14.** Sea  $k \in \mathbb{N}$  fijo. Se denotará  $\{R_n^k\}_{n \in \mathbb{N}}$  al número de bolas rojas en la urna luego del paso  $n$  y  $\{B_n^k\}_{n \in \mathbb{N}}$  al número de bolas blancas luego del paso  $n$ , donde  $r = k + 1$  es el número inicial de bolas rojas, y  $b = 1$  es el número inicial de bolas blancas.

**Definición 5.15.** Sea  $k \in \mathbb{N}$  fijo. Se denotará  $\{J_n^k\}_{n \in \mathbb{N}}$  a la fracción de bolas blancas en la urna luego del paso  $n$  del proceso descrito anteriormente. Adicionalmente, se define  $J_\infty^k$ , como la variable aleatoria dada por el siguiente límite,

$$J_n^k \xrightarrow[\text{c.s.}]{n \rightarrow \infty} J_\infty^k.$$

Se sabe que  $J_\infty^k$  existe por 5.2.2 ( $r = k + 1$  número inicial de bolas rojas,  $b = 1$  número inicial de bolas blancas).

**Lema 5.16.** Sean  $n, k, j \in \mathbb{N}$ ,  $i + 1 \leq n$ . Se tiene que,

$$\mathbb{P}(\tilde{B}_n^k = i) = \mathbb{P}(\tilde{B}_n^k = i + 1) \cdot \frac{k + n - i}{n - i}.$$

*Demostración.* De la Proposición 5.13 sabemos que,

$$\begin{aligned} \mathbb{P}(\tilde{B}_n^k = i) &= \frac{1 \cdot 2 \cdots i \cdot (k+1)(k+2) \cdots (k+n-i)}{(k+2)(k+3) \cdots (k+n+1)} \binom{n}{i} \\ &= \frac{1 \cdot 2 \cdots i \cdot (k+1)(k+2) \cdots (k+n-i)}{(k+2)(k+3) \cdots (k+n+1)} \cdot \frac{n!}{(n-i)! \cdot i!} \\ &= \frac{(k+1)(k+2) \cdots (k+n-i)}{(k+2)(k+3) \cdots (k+n+1)} \cdot \frac{n!}{(n-i)!} \\ &= \frac{(k+1) \cdot (k+n-i)!}{(k+n+1)!} \cdot \frac{n!}{(n-i)!}. \end{aligned}$$

Y con esto,

$$\begin{aligned} \mathbb{P}(\tilde{B}_n^k = i+1) &= \frac{(k+1) \cdot (k+n-i-1)!}{(k+n+1)!} \cdot \frac{n!}{(n-i-1)!} \\ &= \mathbb{P}(\tilde{B}_n^k = i) \cdot \frac{n-i}{k+n-i}. \end{aligned}$$

□

**Lema 5.17.** Sea  $k \in \mathbb{N}$  fijo,  $k \geq 2$ . Entonces,

$$\forall n \in \mathbb{N}, \quad \mathbb{P}\left(J_n^k \geq \frac{1}{2}\right) \leq \mathbb{P}\left(J_\infty^k \geq \frac{1}{2}\right)$$

*Demostración.* Notemos primero que como se tiene lo siguiente,

$$J_n^k \xrightarrow[\text{c.s.}]{n \rightarrow \infty} J_\infty^k$$

entonces también se tiene,

$$\mathbb{P}\left(J_n^k \geq \frac{1}{2}\right) \xrightarrow{n \rightarrow \infty} \mathbb{P}\left(J_\infty^k \geq \frac{1}{2}\right)$$

dado que la convergencia c.s. implica la convergencia en distribución.

Sea  $k \in \mathbb{N}$  fijo, se analizará el caso  $k$  par y  $k$  impar por separado. Notemos primero que,

$$\mathbb{P}\left(J_n^k \geq \frac{1}{2}\right) = \mathbb{P}(B_n^k \geq R_n^k).$$

- **$k$  par:** Sea  $n \in \mathbb{N}$ , por lo tanto  $2n+1$  es impar y  $2n+2$  su sucesor par. Notemos que en el paso  $n$  en la urna habrá en total  $(k+n+2)$  bolas. Por lo tanto en el momento  $2n+1$  habrá  $(k+2n+3)$ , una cantidad impar de bolas en la urna.

$$\begin{aligned} &\mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k) \\ &= \sum_{i \in \mathbb{N}} \mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k | B_{2n+1}^k = i) \cdot \mathbb{P}(B_{2n+1}^k = i) \\ &= \mathbb{P}\left(B_{2n+2}^k \geq R_{2n+2}^k | B_{2n+1}^k = \frac{k+2n+2}{2}\right) \cdot \mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+2}{2}\right) \\ &\quad + 1 \cdot \mathbb{P}\left(B_{2n+1}^k \geq \frac{k+2n+2}{2} + 1\right) \\ &\quad + 0 \cdot \mathbb{P}\left(B_{2n+1}^k \leq \frac{k+2n+2}{2} - 1\right), \end{aligned}$$

dado que si en la extracción  $2n+1$  las bolas blancas ya tienen una ventaja de 1 bola, en el paso siguiente con probabilidad 1 se tendrá que  $\mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k)$ . Igualmente si tienen una desventaja de 3 bolas, lo anterior se tendrá con probabilidad 0.

Además el primer término al ser producto de probabilidades, es  $\geq 0$ . Usando esa cota, se obtiene:

$$\begin{aligned} \mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k) &\geq 0 + \mathbb{P}\left(B_{2n+1}^k \geq \frac{k+2n+2}{2} + 1\right) \\ &= \mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k). \end{aligned}$$

Ahora bien, como se sabe que la sucesión  $(\mathbb{P}(B_n^k \geq R_n^k))_{n \in \mathbb{N}}$  converge y además demostramos que,

$$\forall n \in \mathbb{N}, \quad \mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k) \geq \mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k),$$

basta con demostrar que la sucesión de los  $n$  pares,  $(\mathbb{P}(B_{2n}^k \geq R_{2n}^k))_{n \in \mathbb{N}}$  es monótona creciente para tener lo deseado.

Sea  $n \in \mathbb{N}$  cualquiera. Se demostrará que,

$$\mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k) \geq \mathbb{P}(B_{2n}^k \geq R_{2n}^k).$$

En efecto, primero notemos que en el paso  $2n+2$  habrá  $(k+2n+4)$  bolas en la urna, es decir una cantidad par. Lo mismo ocurre en el paso  $2n$  cuando habrán  $(k+2n+2)$  bolas. Se procederá condicionando con respecto a la cantidad de bolas blancas en el paso  $2n$ .

$$\begin{aligned} \mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k) &= \sum_{i \in \mathbb{N}} \mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k | B_{2n}^k = i) \cdot \mathbb{P}(B_{2n}^k = i) \\ &= \mathbb{P}\left(B_{2n+2}^k \geq R_{2n+2}^k | B_{2n}^k = \frac{k+2n+2}{2}\right) \cdot \mathbb{P}\left(B_{2n}^k = \frac{k+2n+2}{2}\right) \\ &\quad + \mathbb{P}\left(B_{2n+2}^k \geq R_{2n+2}^k | B_{2n}^k = \frac{k+2n+2}{2} - 1\right) \cdot \mathbb{P}\left(B_{2n}^k = \frac{k+2n+2}{2} - 1\right) \\ &\quad + 1 \cdot \mathbb{P}\left(B_{2n}^k \geq \frac{k+2n+2}{2} + 1\right) \\ &\quad + 0 \cdot \mathbb{P}\left(B_{2n}^k \leq \frac{k+2n+2}{2} - 2\right). \end{aligned}$$

Tenemos que,

$$\begin{aligned} \mathbb{P}\left(B_{2n+2}^k \geq R_{2n+2}^k | B_{2n}^k = \frac{k+2n+2}{2} - 1\right) &= \frac{\frac{k+2n+2}{2} - 1}{k+2n+2} \cdot \frac{\frac{k+2n+2}{2}}{k+2n+3} \\ &= \frac{1}{4} \cdot \frac{k+2n}{k+2n+3}, \end{aligned}$$



dado que el único caso favorable corresponde a extraer una bola blanca en las 2 extracciones siguientes,  $2n + 1$  y  $2n + 2$ .

Además,

$$\begin{aligned} \mathbb{P}\left(B_{2n+2}^k \geq R_{2n+2}^k \mid B_{2n}^k = \frac{k+2n+2}{2}\right) &= 1 - \frac{\frac{k+2n+2}{2}}{k+2n+2} \cdot \frac{\frac{k+2n+2}{2} + 1}{k+2n+3} \\ &= 1 - \frac{1}{4} \cdot \frac{k+2n+4}{k+2n+3}, \end{aligned}$$

dado que el único caso desfavorable corresponde a extraer una bola roja en las 2 extracciones siguientes,  $2n + 1$  y  $2n + 2$ .

Gracias al Lema 5.16 (tomando  $i = \frac{k+2n+2}{2} - 2$ ), se tiene que,

$$\begin{aligned} \mathbb{P}\left(B_{2n}^k = \frac{k+2n+2}{2} - 1\right) &= \mathbb{P}\left(\tilde{B}_{2n}^k = \frac{k+2n+2}{2} - 2\right) \\ &= \mathbb{P}\left(\tilde{B}_{2n}^k = \frac{k+2n+2}{2} - 1\right) \cdot \left(\frac{k+2n - \frac{k+2n+2}{2} + 2}{2n - \frac{k+2n+2}{2} + 2}\right) \\ &= \mathbb{P}\left(B_{2n}^k = \frac{k+2n+2}{2}\right) \cdot \left(\frac{k+2n+2}{2n-k+2}\right). \end{aligned}$$

Juntando todo esto, se tiene finalmente lo siguiente,

$$\begin{aligned} \mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k) &= \mathbb{P}\left(B_{2n}^k \geq \frac{k+2n+2}{2} + 1\right) \\ &+ \left(1 - \frac{1}{4} \cdot \frac{k+2n+4}{k+2n+3}\right) \cdot \mathbb{P}\left(B_{2n}^k = \frac{k+2n+2}{2}\right) \\ &+ \left(\frac{1}{4} \cdot \frac{k+2n}{k+2n+3}\right) \cdot \left(\frac{k+2n+2}{2n-k+2}\right) \cdot \mathbb{P}\left(B_{2n}^k = \frac{k+2n+2}{2}\right). \end{aligned}$$

Pero,

$$\begin{aligned} 1 - \frac{1}{4} \cdot \frac{k+2n+4}{k+2n+3} + \frac{1}{4} \cdot \frac{k+2n}{k+2n+3} \cdot \frac{k+2n+2}{2n-k+2} &\geq 1 \\ \iff \frac{1}{4} \cdot \frac{k+2n}{k+2n+3} \cdot \frac{k+2n+2}{2n-k+2} &\geq \frac{1}{4} \cdot \frac{k+2n+4}{k+2n+3} \\ \iff \frac{k+2n}{1} \cdot \frac{k+2n+2}{2n-k+2} &\geq \frac{k+2n+4}{1} \\ \iff \frac{k+2n+2}{2n-k+2} &\geq \frac{k+2n+4}{2n+k}, \end{aligned}$$

lo que es cierto cuando  $k$  es par,  $2n \geq k$  y  $k \geq 2$ .

Por lo tanto,

$$\begin{aligned}\mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k) &\geq \mathbb{P}\left(B_{2n}^k \geq \frac{k+2n+2}{2} + 1\right) + \mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+2}{2}\right) \\ &= \mathbb{P}\left(B_{2n+1}^k \geq \frac{k+2n+2}{2}\right) \\ &= \mathbb{P}(B_{2n}^k \geq R_{2n}^k).\end{aligned}$$

Lo cual concluye la demostración para el caso  $k$  par.

- **$k$  impar:** Sea  $n \in \mathbb{N}$ , por lo tanto  $2n$  es par y  $2n+1$  su sucesor impar. En el momento  $2n$  habrá  $(k+2n+2)$  bolas en la urna, que al igual que en caso anterior, es impar.

$$\begin{aligned}\mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k) &= \sum_{i \in \mathbb{N}} \mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k | B_{2n}^k = i) \cdot \mathbb{P}(B_{2n}^k = i) \\ &= \mathbb{P}\left(B_{2n+1}^k \geq R_{2n+1}^k | B_{2n}^k = \frac{k+2n+1}{2}\right) \cdot \mathbb{P}\left(B_{2n}^k = \frac{k+2n+1}{2}\right) \\ &\quad + 1 \cdot \mathbb{P}\left(B_{2n}^k \geq \frac{k+2n+1}{2} + 1\right) \\ &\quad + 0 \cdot \mathbb{P}\left(B_{2n}^k \leq \frac{k+2n+1}{2} - 1\right),\end{aligned}$$

dado que si en la extracción  $2n$  las bolas blancas ya tienen una ventaja de 1 bola, en el paso siguiente con probabilidad 1 se tendrá que  $\mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k)$ . Igualmente si tienen una desventaja de 3 bolas, lo anterior se tendrá con probabilidad 0.

Además el primer término al ser producto de probabilidades, es  $\geq 0$ . Usando esa cota, se obtiene:

$$\begin{aligned}\mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k) &\geq 0 + \mathbb{P}\left(B_{2n}^k \geq \frac{k+2n+1}{2} + 1\right) \\ &= \mathbb{P}(B_{2n}^k \geq R_{2n}^k).\end{aligned}$$

Así, análogo al caso anterior, basta con demostrar que la sucesión de los  $n$  impares,  $(\mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k))_{n \in \mathbb{N}}$  es monótona creciente para tener lo deseado.

Sea  $n \in \mathbb{N}$  cualquiera. Se demostrará que,

$$\mathbb{P}(B_{2n+3}^k \geq R_{2n+3}^k) \geq \mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k).$$

En efecto, en el paso  $2n+3$  habrá  $(k+2n+5)$  bolas en la urna, es decir una cantidad par. Lo mismo ocurre en el paso  $2n+1$  cuando habrán  $(k+2n+3)$  bolas. Condicionando con respecto a la cantidad de bolas blancas en el paso  $2n+1$ , se obtiene lo siguiente.

$$\begin{aligned}
& \mathbb{P}(B_{2n+3}^k \geq R_{2n+3}^k) \\
&= \sum_{i \in \mathbb{N}} \mathbb{P}(B_{2n+3}^k \geq R_{2n+3}^k | B_{2n+1}^k = i) \cdot \mathbb{P}(B_{2n+1}^k = i) \\
&= \mathbb{P}\left(B_{2n+3}^k \geq R_{2n+3}^k | B_{2n+1}^k = \frac{k+2n+3}{2} - 1\right) \cdot \mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+3}{2} - 1\right) \\
&\quad + \mathbb{P}\left(B_{2n+3}^k \geq R_{2n+3}^k | B_{2n+1}^k = \frac{k+2n+3}{2}\right) \cdot \mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+3}{2}\right) \\
&\quad + 1 \cdot \mathbb{P}\left(B_{2n+1}^k \geq \frac{k+2n+3}{2} + 1\right) \\
&\quad + 0 \cdot \mathbb{P}\left(B_{2n}^k \leq \frac{k+2n+2}{2} - 2\right).
\end{aligned}$$

Tenemos que,

$$\begin{aligned}
\mathbb{P}\left(B_{2n+3}^k \geq R_{2n+3}^k | B_{2n+1}^k = \frac{k+2n+3}{2} - 1\right) &= \frac{\frac{k+2n+3}{2} - 1}{k+2n+3} \cdot \frac{\frac{k+2n+3}{2}}{k+2n+4} \\
&= \frac{1}{4} \cdot \frac{k+2n+1}{k+2n+4}
\end{aligned}$$

dado que el único caso favorable corresponde a extraer una bola blanca en las 2 extracciones siguientes,  $2n+2$  y  $2n+3$ .

Además,

$$\begin{aligned}
\mathbb{P}\left(B_{2n+3}^k \geq R_{2n+3}^k | B_{2n+1}^k = \frac{k+2n+3}{2}\right) &= 1 - \frac{\frac{k+2n+3}{2}}{k+2n+3} \cdot \frac{\frac{k+2n+3}{2} + 1}{k+2n+4} \\
&= 1 - \frac{1}{4} \cdot \frac{k+2n+5}{k+2n+4}
\end{aligned}$$

dado que el único caso desfavorable corresponde a extraer una bola roja en las 2 extracciones siguientes,  $2n+2$  y  $2n+3$ .

Análogo al caso anterior (tomando  $i = \frac{k+2n+3}{2} - 2$ ), se tiene que,

$$\begin{aligned}
\mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+3}{2} - 1\right) &= \mathbb{P}\left(\tilde{B}_{2n+1}^k = \frac{k+2n+3}{2} - 2\right) \\
&= \mathbb{P}\left(\tilde{B}_{2n+1}^k = \frac{k+2n+3}{2} - 1\right) \cdot \frac{k+2n - \frac{k+2n+3}{2} + 3}{2n - \frac{k+2n+3}{2} + 3} \\
&= \mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+3}{2}\right) \cdot \frac{k+2n+3}{2n-k+3}.
\end{aligned}$$

Juntando todo esto, se tiene finalmente lo siguiente,

$$\begin{aligned} & \mathbb{P}(B_{2n+3}^k \geq R_{2n+3}^k) \\ &= \mathbb{P}\left(B_{2n+1}^k \geq \frac{k+2n+3}{2} + 1\right) \\ &+ \left(1 - \frac{1}{4} \cdot \frac{k+2n+5}{k+2n+4}\right) \cdot \mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+3}{2}\right) \\ &+ \left(\frac{1}{4} \cdot \frac{k+2n+1}{k+2n+4}\right) \cdot \left(\frac{k+2n+3}{2n-k+3}\right) \cdot \mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+3}{2}\right). \end{aligned}$$

Pero,

$$\begin{aligned} & 1 - \frac{1}{4} \cdot \frac{k+2n+5}{k+2n+4} + \frac{1}{4} \cdot \frac{k+2n+1}{k+2n+4} \cdot \frac{k+2n+3}{2n-k+3} \geq 1 \\ \iff & \frac{1}{4} \cdot \frac{k+2n+1}{k+2n+4} \cdot \frac{k+2n+3}{2n-k+3} \geq \frac{1}{4} \cdot \frac{k+2n+5}{k+2n+4} \\ \iff & \frac{k+2n+1}{1} \cdot \frac{k+2n+3}{2n-k+3} \geq \frac{k+2n+5}{1} \\ \iff & \frac{k+2n+3}{2n-k+3} \geq \frac{k+2n+5}{2n+k+1} \end{aligned}$$

lo que es cierto cuando  $k$  es impar,  $2n+1 \geq k$ ,  $k \geq 3$ .

Por lo tanto,

$$\begin{aligned} \mathbb{P}(B_{2n+3}^k \geq R_{2n+3}^k) &\geq \mathbb{P}\left(B_{2n+1}^k \geq \frac{k+2n+3}{2} + 1\right) + \mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+3}{2}\right) \\ &= \mathbb{P}\left(B_{2n+1}^k \geq \frac{k+2n+3}{2}\right) \\ &= \mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k). \end{aligned}$$

Lo cual concluye la demostración. □

**Lema 5.18.** Sea  $k \in \mathbb{N}$ ,  $k \geq 2$ .

$$\mathbb{E}(S_n^k) \leq \mathbb{P}\left(J_\infty^k \geq \frac{1}{2}\right), \quad \forall n \in \mathbb{N}.$$

*Demostración.* Gracias a la Proposición 5.8, basta demostrar que,

$$\mathbb{E}(S_{\sigma_n}^k) \leq \mathbb{P}\left(J_\infty^k \geq \frac{1}{2}\right).$$

Así,

$$\begin{aligned} \mathbb{E}(S_{\sigma_n}^k) &= \mathbb{E}\left(\mathbb{I}_{\{d_{\sigma_n}^k \geq d_{\sigma_n}^0\}}\right) \\ &= \mathbb{P}(d_{\sigma_n}^k \geq d_{\sigma_n}^0) \\ &= \mathbb{P}\left(J_{\sigma_n}^k \geq \frac{1}{2}\right) \\ &\leq \mathbb{P}\left(J_\infty^k \geq \frac{1}{2}\right). \end{aligned}$$

La última desigualdad se tiene gracias al Lema 5.17. Lo cual concluye la demostración.  $\square$

**Teorema 5.19.** *Sea  $T^N = (V^N, E^N)$ ,  $N \in \mathbb{N}$  el árbol dinámico construido de acuerdo a 5.1.1. Se tiene que,*

$$\forall n \in \mathbb{N}, \quad \mathbb{E}(M_{0,n}) \leq \frac{5}{4}.$$

*Es decir el número esperado de vecinos de  $v_0$ , con grado mayor o igual a él, está acotado por 2 a lo largo de todo el proceso.*

*Demostración.* Para demostrar lo requerido, primero calcularemos  $\mathbb{P}(J_\infty^k \geq \frac{1}{2})$ . Por (5.1) sabemos que,

$$J_\infty^k \sim \beta(1, k+1)$$

con lo cual la función de densidad de  $\mathbb{P}(J_\infty^k \geq \frac{1}{2})$  es la siguiente,

$$f_{\beta(1,k+1)} = \frac{\Gamma(k+2)}{\Gamma(k+1)\Gamma(1)}(1-x)^k.$$

Así tenemos que,

$$\begin{aligned} \mathbb{P}\left(J_\infty^k \geq \frac{1}{2}\right) &= \int_{\frac{1}{2}}^1 f_{\beta(1,k+1)} dx \\ &= \frac{\Gamma(k+2)}{\Gamma(k+1)\Gamma(1)} \int_{\frac{1}{2}}^1 (1-x)^k dx \\ &= \frac{(k+1)!}{k!} \int_{\frac{1}{2}}^1 (1-x)^k dx \\ &= (k+1) \left[ -\frac{(1-x)^{k+1}}{k+1} \right]_{x=\frac{1}{2}}^{x=1} \\ &= \left(1 - \frac{1}{2}\right)^{k+1} \\ &= \frac{1}{2^{k+1}}. \end{aligned}$$

Sea  $n \in \mathbb{N}$  fijo,

$$\begin{aligned}
\mathbb{E}(M_{0,n}) &= \mathbb{E}\left(\sum_{k=1}^{d_n^0} S_n^k\right) \\
&= \mathbb{E}\left(\sum_{k=1}^n S_n^k\right) \\
&= \sum_{k=1}^n \mathbb{E}(S_n^k) \\
&\leq \mathbb{E}(S_1^k) + \sum_{k=2}^n \mathbb{P}(J_\infty^k \geq \frac{1}{2}) \\
&\leq \mathbb{P}(d_n^k \geq d_n^0) + \sum_{k=2}^{\infty} \mathbb{P}(J_\infty^k \geq \frac{1}{2}) \\
&\leq 1 + \sum_{k=2}^{\infty} \frac{1}{2^{k+1}} \\
&\leq \frac{5}{4}.
\end{aligned}$$

Notemos que se utilizó el Lema 5.18 para acotar los términos desde  $k = 2$  en adelante. Para el término  $k = 1$ , acotamos por 1.  $\square$

### 5.3. Análisis General

En esta sección se analizará el crecimiento y distribución del número de vecinos de cada nodo en el árbol. En particular, se demostrará que el número de vecinos de cualquier nodo en el árbol se distribuye del mismo modo que el número de vecinos de  $v_0$ , salvo un ajuste en el tiempo que se precisará en lo que sigue.

**Definición 5.20.** Sea  $T = (V, E)$  árbol con raíz  $r$ , simple no dirigido y sea  $v \in V$ . Se llamará *predecesor* de  $v$  al nodo  $p_v \in V$  que está en el camino único de  $v$  a  $r$  y que se encuentra a distancia 1 de  $v$ . Se denotará  $T_v = (V_v, E_v)$  al subárbol que tiene como raíz al predecesor  $p_v$  de  $v$ , como conjunto de nodos a  $p_v$ ,  $v$  y a todos los nodos  $w$  tal que en su único camino a  $r$ , tienen que pasar por  $v$ . El conjunto de arcos  $E_v$  es el conjunto de arcos inducidos por el conjunto  $V_v$ . Los nodos en  $V_v \setminus \{p_v, v\}$  se llamarán *descendientes* de  $v$ .

**Definición 5.21.** Sea  $T^N = (V^N, E^N)$  el árbol construido de acuerdo a lo definido en 5.1.1. Sea  $n, i \in \mathbb{N}, n, i \leq N$ . Se denotará  $Y_{i,n}$  a la suma de los grados de los nodos que pertenecen al subárbol  $T_{v_i}^n \setminus \{p_{v_i}\}$  (definido en 5.20) sobre la suma total de grados, en el paso  $n$  de construcción del árbol dinámico.

Es decir,

$$Y_{i,n} = \frac{\sum_{v \in T_{v_i}^n \setminus \{p_{v_i}\}} d_{T^n}(v)}{\sum_{v \in T^n} d_{T^n}(v)} = \frac{\sum_{v \in T_{v_i}^n \setminus \{p_{v_i}\}} d_{T_{v_i}^n}(v)}{\sum_{v \in T^n} d_{T^n}(v)}, \quad n, i \in \mathbb{N}$$

En particular, si  $n = i$ ,  $Y_{i,i} = \frac{1}{2i+1}$ , dado que en el paso  $i$  la suma total de grados en el árbol será  $2i + 1$  y sólo  $v_i$  será parte del subárbol  $T_{v_i}$ , cuyo grado en el tiempo  $n = i$  será 1.

**Definición 5.22.** Se define la siguiente familia de  $\sigma$ -álgebras,

$$\mathcal{F}_n = \sigma(X_{i,m}, 0 \leq i \leq m, m \leq n), \quad n \in \mathbb{N}.$$

Donde,  $X_{i,m}$  corresponde al número de vecinos del nodo  $v_i$  en el paso  $m$  (definida en 5.2).

**Lema 5.23.**  $\forall i \in \mathbb{N}$ , se tiene lo siguiente

$$\mathbb{E}(Y_{i,n}) = \frac{1}{2i+1}, \quad \forall n \in \mathbb{N}, n \geq i.$$

*Demostración.* Sea  $i \in \mathbb{N}$  fijo. Se demostrará que  $(Y_{i,n})_{n \geq i}$  es una Martingala. Con esto basta dado que se sabe que  $\mathbb{E}(Y_{i,i}) = \frac{1}{2i+1}$ .

En efecto, supongamos que en el paso  $n$  hay  $j$  nodos en el subárbol  $T_{v_i} \setminus \{p_{v_i}\}$ . Se tiene que  $Y_{i,n} = \frac{2(j-1)+1}{2n+1} = \frac{2j-1}{2n+1}$ ,

$$\begin{aligned} \mathbb{E}(Y_{i,n+1} | \mathcal{F}_n) &= \frac{2(j+1)-1}{2(n+1)+1} \cdot \frac{2j-1}{2n+1} + \frac{2j-1}{2(n+1)+1} \cdot \left(1 - \frac{2j-1}{2n+1}\right) \\ &= \frac{2j+1}{2n+3} \cdot \frac{2j-1}{2n+1} + \frac{2j-1}{2n+3} \cdot \left(1 - \frac{2j-1}{2n+1}\right) \\ &= \frac{(2j+1)(2j-1)}{(2n+3)(2n+1)} + \frac{2j-1}{2n+3} - \frac{(2j-1)^2}{(2n+3)(2n+1)} \\ &= \frac{4j^2 - 1 + (2j-1)(2n+1) - (4j^2 - 4j + 1)}{(2n+3)(2n+1)} \\ &= \frac{-1 + 4nj + 2j - 2n - 1 + 4j - 1}{(2n+3)(2n+1)} \\ &= \frac{4nj + 6j - 2n - 3}{(2n+3)(2n+1)} \\ &= \frac{(2j-1)(2n+3)}{(2n+3)(2n+1)} \\ &= \frac{2j-1}{2n+1} = Y_{i,n}. \end{aligned}$$

Con lo cual se demostró que para  $i$  fijo  $(Y_{i,n})_{n \in \mathbb{N}}$  es martingala. Por lo tanto

$$\mathbb{E}(Y_{i,n}) = \mathbb{E}(Y_{i,i}) = \frac{1}{2i+1}, \quad \forall n \in \mathbb{N}, n \geq i.$$

□

**Corolario 5.24.** La esperanza de la fracción de nodos en el árbol  $T_{v_i} \setminus \{p_{v_i}\}$  en el paso  $n$  también se mantiene constante e igual a  $\frac{1}{i+1}$ .

Se quiere demostrar que lo que ocurre dentro de cada subárbol  $T_{v_i}$  es una "réplica", en un sentido que se precisará, del árbol completo  $T$  si nos fijamos sólo en los eventos que ocurren en  $T_{v_i}$  e ignoramos todo lo que ocurre fuera de él. Para esto, se partirá por demostrar el siguiente lema que dice que la probabilidad de elegir un nodo  $u$  dentro de un subárbol que contiene a  $u$  condicionado a que el nodo se elige dentro de ese subárbol, es igual que elegir un nodo dentro del subárbol como si el subárbol fuera el árbol dinámico completo.

**Lema 5.25.** *Sea  $T^n = (V^n, E^n)$ ,  $n \in \mathbb{N}$  el árbol dinámico construido de acuerdo a 5.1.1. Además, sea  $i \in \mathbb{N}$ ,  $i \leq n$ . Asumimos s.p.g. que en el instante  $n$ , la cantidad de nodos en  $T_{v_i}^n \setminus \{p_{v_i}\}$  es  $j$ .*

*Adicionalmente, se define un árbol  $\tilde{T}$  de tamaño  $j$  (excluyendo la raíz  $r$  que no participa del proceso). El cual corresponde a un árbol construido de acuerdo a 5.1.1 en tiempo  $j-1$ , es decir  $\tilde{T} = (\tilde{V}, \tilde{E})$ , donde  $\tilde{V} = \{r, v_0, v_1, \dots, v_{j-1}\}$ .*

*Con esto, teniendo en mente 5.1.1 se tiene lo siguiente,*

$$\mathbb{P}(w_{n+1} = u | w_{n+1} \in T_{v_i}^n \setminus \{p_{v_i}\}, T_{v_i}^n \cong \tilde{T}) = \mathbb{P}(w_j = \pi(u) | T^{j-1} = \tilde{T}), \quad \forall u \in T_{v_i}^n \setminus \{p_{v_i}\}.$$

Donde,  $\pi : V(T_{v_i}^n) \rightarrow V(\tilde{T})$  corresponde al isomorfismo entre  $T_{v_i}^n$  y  $\tilde{T}$ .

*Demostración.* Desarrollando el lado izquierdo de la igualdad,

$$\begin{aligned} \mathbb{P}(w_{n+1} = u | w_{n+1} \in T_{v_i}^n \setminus \{p_{v_i}\}, T_{v_i}^n \cong \tilde{T}) &= \frac{\mathbb{P}(w_{n+1} = u \cap w_{n+1} \in T_{v_i}^n \setminus \{p_{v_i}\} \cap T_{v_i}^n \cong \tilde{T})}{\mathbb{P}(w_{n+1} \in T_{v_i}^n \setminus \{p_{v_i}\} \cap T_{v_i}^n \cong \tilde{T})} \\ &= \frac{d_{\tilde{T}}(u)}{\sum_{v \in T^n} d_{T^n}(v)} \cdot \frac{\sum_{v \in T^n} d_{T^n}(v)}{\sum_{v \in \tilde{T} \setminus \{r\}} d_{\tilde{T}}(v)} \\ &= \frac{d_{\tilde{T}}(u)}{\sum_{v \in \tilde{T} \setminus \{r\}} d_{\tilde{T}}(v)} \\ &= \frac{d_{\tilde{T}}(u)}{2j + 1} \end{aligned}$$

y luego el lado derecho,

$$\begin{aligned} \mathbb{P}(w_j = \pi(u) | T^{j-1} = \tilde{T}) &= \frac{d_{\tilde{T}}(u)}{\sum_{v \in \tilde{T} \setminus \{r\}} d_{\tilde{T}}(v)} \\ &= \frac{d_{\tilde{T}}(u)}{2j + 1} \end{aligned}$$

Lo cual concluye la demostración. □

**Definición 5.26.** Sea  $T^N = (V^N, E^N)$ ,  $N \in \mathbb{N}$ , árbol dinámico construido de acuerdo a 5.1.1. Sea  $i \in \mathbb{N}$ ,  $i \leq N$ , se definen los siguientes tiempos de parada:



- $\tau_0^i = i$ .
- $\tau_{m+1}^i = \inf_{n \geq \tau_m^i} \{V_{v_i}^n \subsetneq V_{v_i}^{\tau_m^i}\} \quad m \in \mathbb{N}, m \leq N$ .

**Teorema 5.27.** *Sea  $T^N = (V^N, E^N)$ ,  $N \in \mathbb{N}$ , árbol dinámico construido de acuerdo a 5.1.1. Sean además  $i, n, m \in \mathbb{N}$ ,  $i \leq n \leq N$  y  $m \leq n - i + 2$ . Entonces,*

$$T_{v_i}^{\tau_m^i} \sim T^m$$

*Demostración.* Sea  $i \in \mathbb{N}$  fijo. Se demostrará lo enunciado mediante inducción,

- **c.b.:**  $T_{v_i}^{\tau_0^i} \sim T^0$ .
- **h.i.:**  $T_{v_i}^{\tau_m^i} \sim T^m$ .
- **p.i.:** Sabemos por hipótesis inductiva que  $T_{v_i}^{\tau_m^i} \sim T^m$  y además sabemos, gracias al Lema 5.25, que en el tiempo  $\tau_{m+1}^i$  el nodo que llega a  $T_{v_i}^{\tau_{m+1}^i}$  se conecta con cualquiera de los nodos en el subárbol con la misma probabilidad con la que lo haría si  $T_{v_i}^{\tau_{m+1}^i}$  fuera el árbol, así tenemos que  $T_{v_i}^{\tau_{m+1}^i} \sim T^{m+1}$ .

Y con esto se concluye lo enunciado. □

En el Teorema 5.27 se demostró que lo que ocurre al interior de un subárbol, condicionado a los eventos que ocurren dentro de él, se distribuye igual que el árbol completo.

**Corolario 5.28.** *Sea  $T^N = (V^N, E^N)$ ,  $N \in \mathbb{N}$  el árbol dinámico construido de acuerdo a 5.1.1. Además, sean  $i, n \in \mathbb{N}$ ,  $i \leq n \leq N$ . Entonces,*

$$\mathbb{E}(X_{i,n} \mid |T_{v_i}^n| = m + 2) = \mathbb{E}(X_{0,m}). \quad (5.2)$$

$$\mathbb{E}(M_{i,n} \mid |T_{v_i}^n| = m + 2) \leq \mathbb{E}(M_{0,m}) + 1. \quad (5.3)$$

Donde  $M_{i,n}$  denota el número de vecinos de  $v_i$  con grados mayor o igual a él (definido en 5.3).

*Demostración.* Notemos que (5.2) se deduce directo del Teorema 5.27.

En el caso de (5.3), se deduce Teorema 5.27 notando que,  $v_0$  tiene como predecesor (definido en 5.20) a  $r$  cuyo grado es 1 (y éste no aumenta), lo cual es siempre menor estricto que el grado de  $v_0$ . En general, el predecesor  $p_v$  de un nodo  $v$  cualquiera en la red, puede tener grado mayor o igual a él. □

**Teorema 5.29.** *Los árboles de Barabási-Albert son  $\frac{9}{4}$ -Débil-Localmente-Populares, independiente del tamaño del árbol.*

*Demostración.* Gracias al Corolario 5.28, se tiene que  $\mathbb{E}(M_{i,n}) \leq 1 + \frac{5}{4}$ ,  $\forall i, n \in \mathbb{N}$ . Con lo cual se concluye. □

# Capítulo 6

## Preferential Attachment y Popularidad Local: Resultados Experimentales para el Caso General

En el presente capítulo se presentarán resultados experimentales de *Popularidad Local Débil* para grafos del tipo *Barabási-Albert*.

### 6.1. Grafos de Barabási-Albert

Los grafos del tipo *Barabási-Albert* se construyen mediante un proceso discreto en el cual en cada paso se agrega un nodo al grafo, que se une a  $m$  nodos distintos que ya pertenecen al grafo, usando la regla de *Preferential Attachment*.

A continuación se define la notación y conceptos que se utilizarán y luego el proceso mediante el cual estos grafos se generan.

**Definición 6.1.** Sea  $G = (V, E)$  grafo simple, no-dirigido. Sea  $v \in V$ , se denotará  $N(v)$  al conjunto de vecinos de  $v$  y  $d_G(v)$  al grado de  $v$  en  $G$ .

#### 6.1.1. Proceso

Sea  $m \in \mathbb{N}$  fijo. El proceso de crecimiento del grafo evoluciona en tiempos discretos  $n = 0, 1, 2, 3, \dots$ . Su estado en cada tiempo  $n \geq 0$  es un grafo conexo, simple, no-dirigido  $G^n = (V^n, E^n)$ , cuyo conjunto de vértices es  $V^n$  y de arcos es  $E^n$ , y en cada etapa se agrega un nodo con  $m$  arcos distintos. Proceso descrito en 4.2.

Inicialmente  $G^0 = K^{m+1}$ . Es decir, se inicia con un clique de tamaño  $m + 1$ ,  $V^0 = \{v_0, v_1, \dots, v_m\}$ . En cada tiempo  $n \geq 1$  se agrega un nuevo nodo a la red, de acuerdo al proceso descrito en 4.2.

## 6.2. Simulaciones

### 6.2.1. Definiciones Previas

**Definición 6.2.** Sea  $G^N = (V^N, E^N)$ ,  $N \in \mathbb{N}$  el grafo construido de acuerdo a lo definido en 6.1.1. Sea  $n, i \in \mathbb{N}$ ,  $n, i \leq N$ . Se denotará  $M_{i,n}$  al número de vecinos con grado mayor o igual que el nodo  $v_i$  en el paso  $n$  del proceso de construcción del grafo. Esto define un arreglo de variables aleatorias  $\{M_{i,n}\}_{0 \leq i, n \leq N}$ . Notemos que en el tiempo  $i$  llega el nodo  $v_{m+i}$  a la red. Por lo tanto para  $i \geq m$ , el tiempo de llegada del nodo  $v_i$  es  $i - m$ . Imponemos que para tiempos anteriores al tiempo de llegada del nodo  $i$  ( $n < i - m$ ),  $v_i$  tiene 0 vecinos con grado mayor o igual a él,  $M_{i,n} = 0$ . Notar además que  $M_{i,0} = m$ ,  $\forall 0 \leq i \leq m$ .

**Definición 6.3.** Sea  $m \in \mathbb{N}$  y  $G^N = (V^N, E^N)$  el grafo construido de acuerdo a lo definido en 6.1.1, donde cada nodo nuevo se conecta con  $m$  nodos distintos ya existentes en el grafo. Sea  $n \in \mathbb{N}$ ,  $n \leq N$ . Se define la variable aleatoria  $Q_n^m$  como sigue:

$$Q_n^m = \frac{\sum_{i=0}^{m+n} M_{i,n}}{m+n+1}.$$

Es decir,  $Q_n^m$  corresponde al promedio de  $M_{i,n}$  sobre todos los nodos, hasta el paso  $n$ .

**Definición 6.4.** Se denotará  $\bar{Q}_n^m$  al *promedio empírico* de  $Q_n^m$ .

### 6.2.2. Resultados

A continuación se exponen los resultados de la simulación del crecimiento de grafos como se describió en 6.1.1.

En lo que sigue se grafica la promedio empírico de la variable aleatoria  $Q_n^m$  en función del tamaño del grafo  $n$ . En todos los casos la promedio empírico fue calculada con 500 experimentos. Se generaron distintos grafos para los distintos valores de  $n$ , partiendo con grafos de tamaño  $n = 100$ , hasta grafos de tamaño  $n = 1500$ . Se presentan los gráficos para diferentes valores de  $m \in \{1, 2, 3, 4, 5, 6, 7\}$  con la tabla de datos correspondiente.

- $m=1$

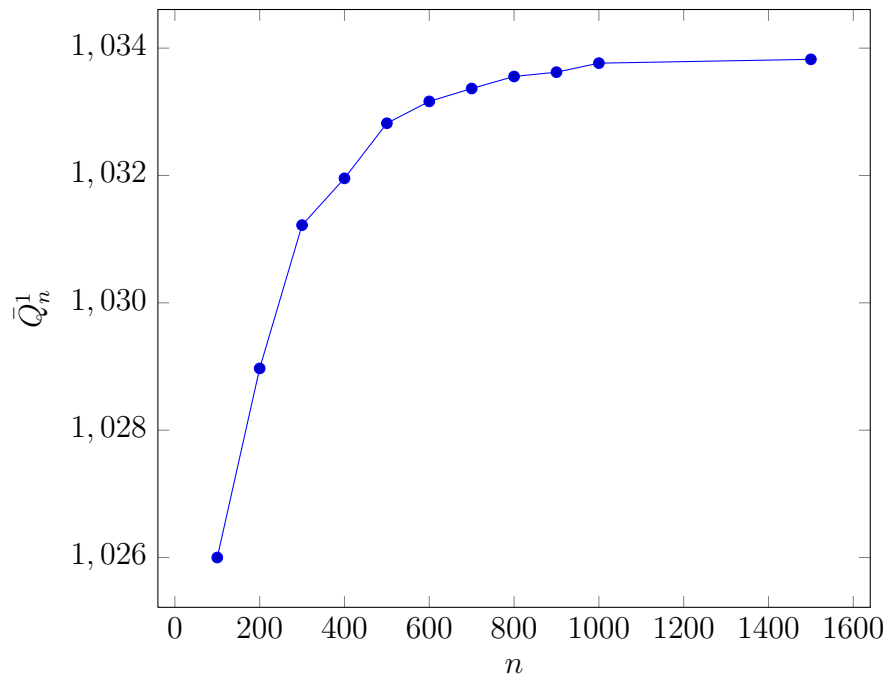


Figura 6.1

Cuadro 6.1

<i>Número de nodos n</i>	<i>Prom. Empírico <math>\bar{Q}_n^1</math></i>
100	1,0260
200	1,0289
300	1,0312
400	1,0319
500	1,0328
600	1,0331
700	1,0333
800	1,0335
900	1,0336
1000	1,0337
1500	1,0338

- $m=2$

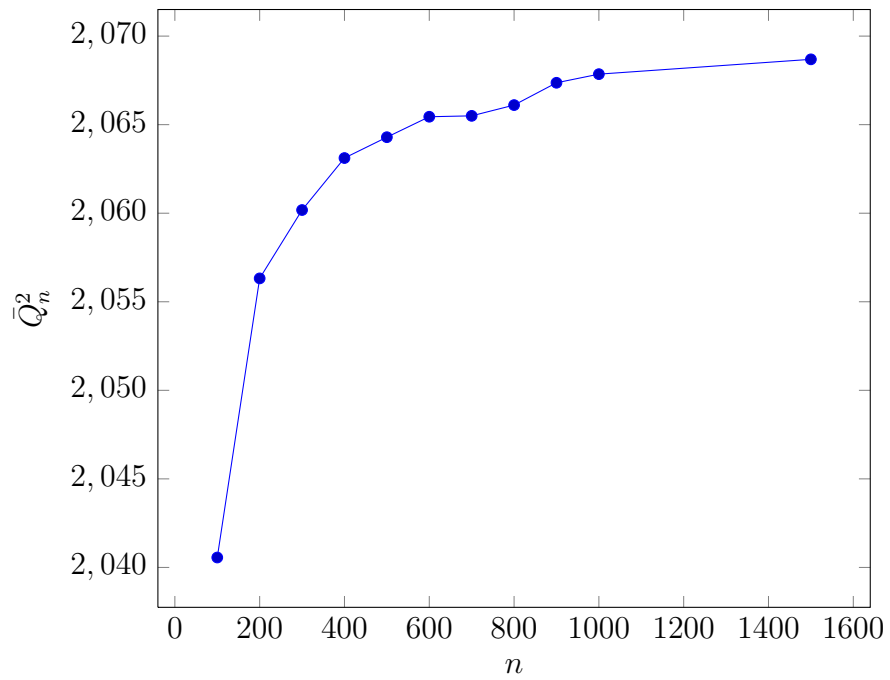


Figura 6.2

Cuadro 6.2

<i>Número de nodos</i> $n$	<i>Prom. Empírico</i> $Q_n^2$
100	2,0406
200	2,0563
300	2,0602
400	2,0631
500	2,0643
600	2,0654
700	2,0655
800	2,0661
900	2,0674
1000	2,0679
1500	2,0687

- $m=3$

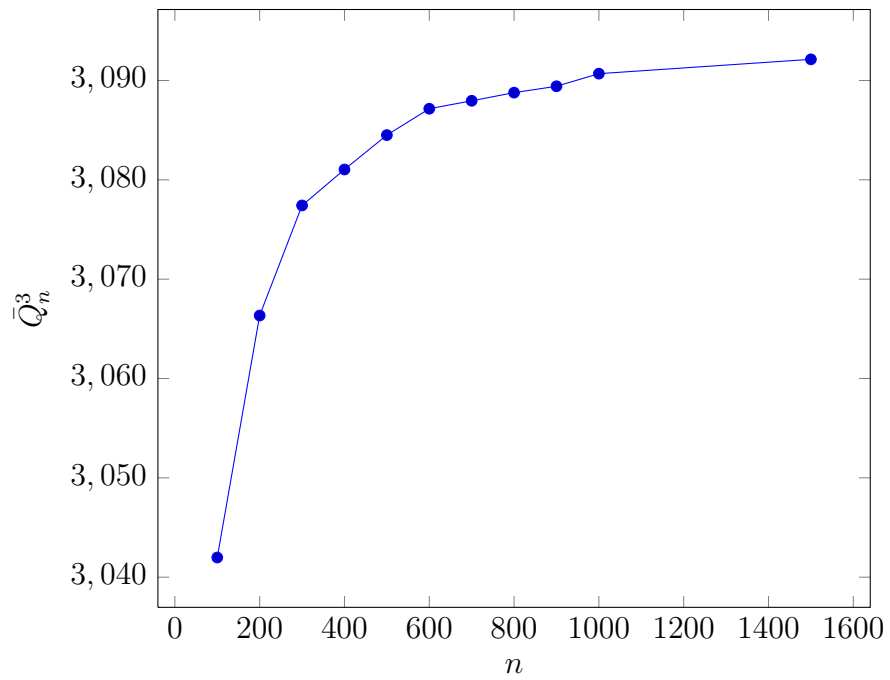


Figura 6.3

Cuadro 6.3

<i>Número de nodos n</i>	<i>Prom. Empírico <math>\bar{Q}_n^3</math></i>
100	3,0419
200	3,0663
300	3,0774
400	3,0810
500	3,0845
600	3,0872
700	3,0879
800	3,0888
900	3,0894
1000	3,0907
1500	3,0922

- $m=4$

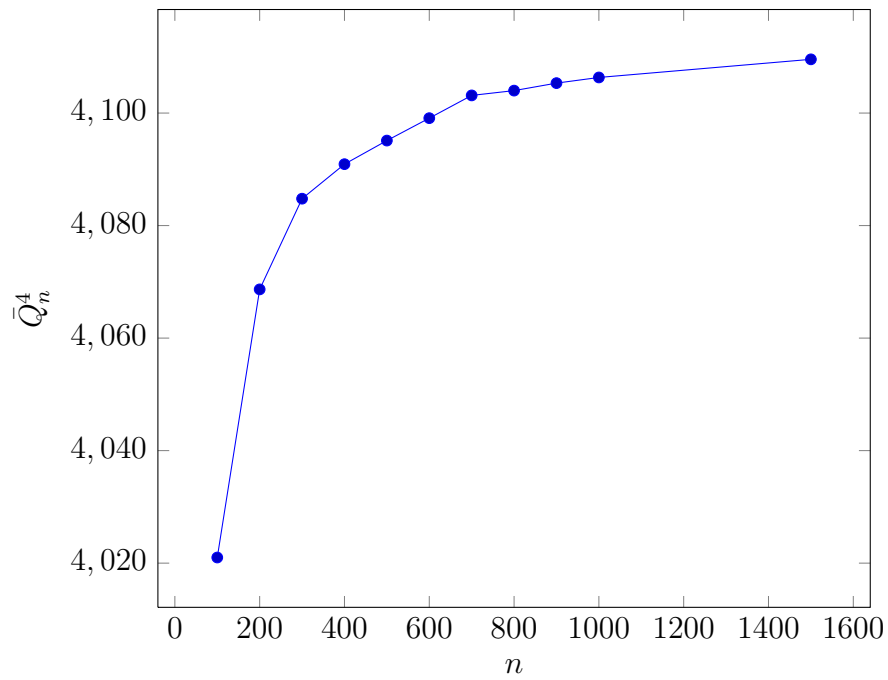


Figura 6.4

Cuadro 6.4

<i>Número de nodos <math>n</math></i>	<i>Prom. Empírico <math>Q_n^4</math></i>
100	4,0209
200	4,0687
300	4,0848
400	4,0909
500	4,0951
600	4,0991
700	4,1031
800	4,1039
900	4,1053
1000	4,1063
1500	4,1095

- $m=5$

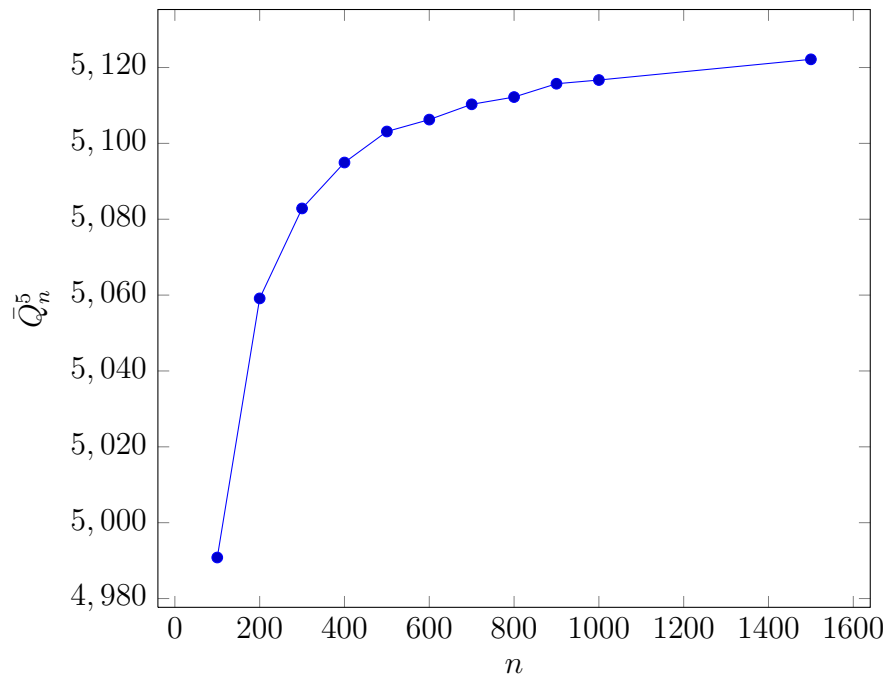


Figura 6.5

Cuadro 6.5

<i>Número de nodos n</i>	<i>Prom. Empírico <math>\bar{Q}_n^5</math></i>
100	4,9908
200	5,0591
300	5,0828
400	5,0949
500	5,1031
600	5,1063
700	5,1103
800	5,1122
900	5,1157
1000	5,1167
1500	5,1222



- $m=6$

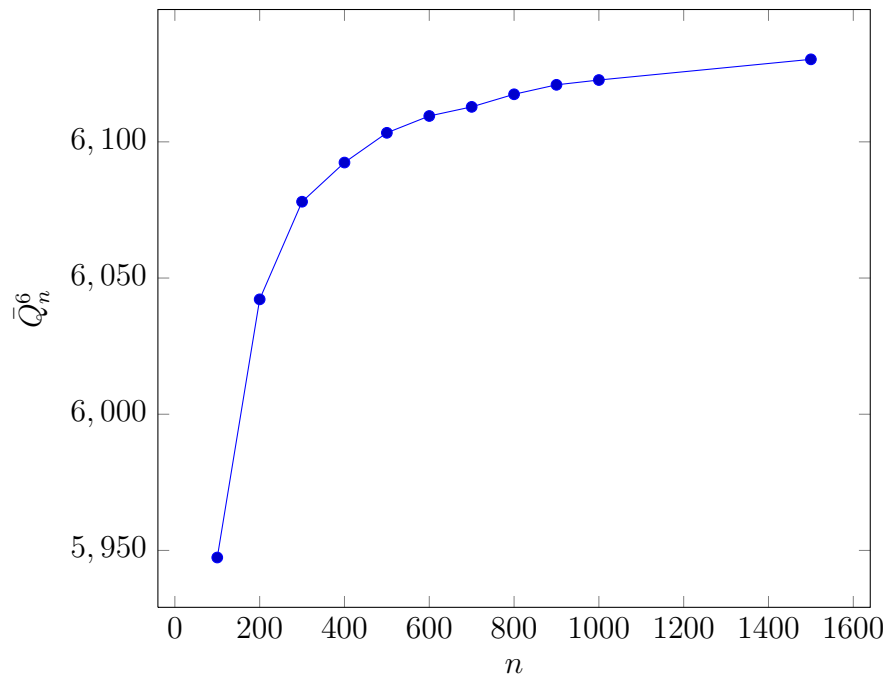


Figura 6.6

Cuadro 6.6

<i>Número de nodos <math>n</math></i>	<i>Prom. Empírico <math>\bar{Q}_n^6</math></i>
100	5,9474
200	6,0422
300	6,0780
400	6,0924
500	6,1033
600	6,1095
700	6,1128
800	6,1175
900	6,1209
1000	6,1227
1500	6,1303

- $m=7$

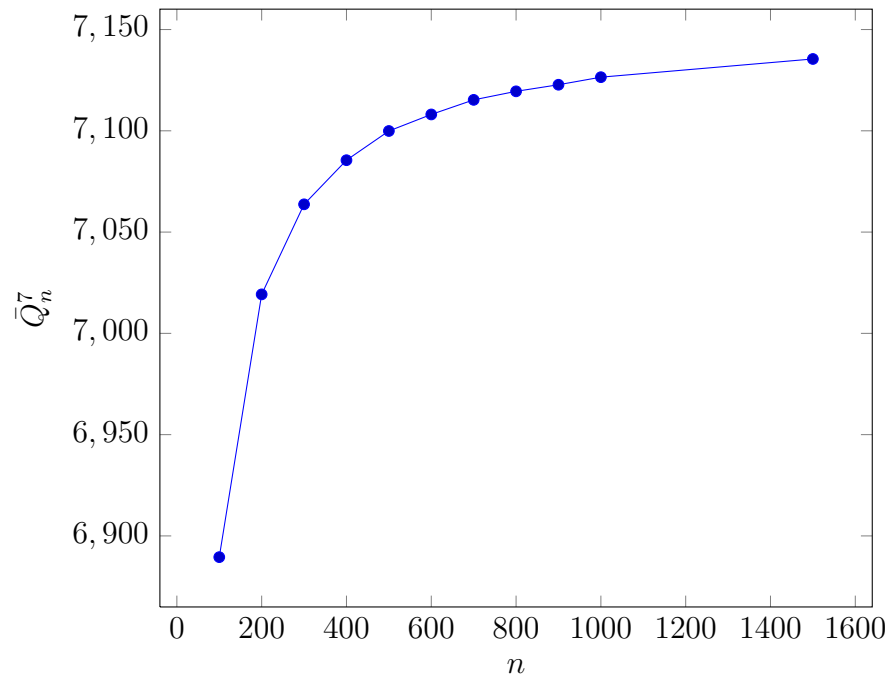


Figura 6.7

Cuadro 6.7

<i>Número de nodos <math>n</math></i>	<i>Prom. Empírico <math>\bar{Q}_n^7</math></i>
100	6,8895
200	7,0192
300	7,0637
400	7,0855
500	7,0999
600	7,1081
700	7,1153
800	7,1195
900	7,1227
1000	7,1265
1500	7,1355

### 6.3. Análisis

A partir de los gráficos de la Sección 6.2.2, es importante destacar lo siguiente:

- *Rango:* A pesar de haber graficado desde grafos de tamaño  $n = 100$  hasta grafos de tamaño  $n = 1500$ , el promedio empírico de  $Q_n^m$  no experimentó una variación sustancial en función de  $n$ ,  $\forall m \in \{1, 2, 3, 4, 5, 6, 7\}$ .
- *Comportamiento:* También es importante destacar que  $\bar{Q}_n^m$  exhibe un comportamiento típico de funciones que se acercan a una asíntota cuando  $n$  crece (dado que su crecimiento se estanca considerablemente para grafos de mayor tamaño).

Las observaciones anteriores son válidas para todos los experimentos realizados, con  $m \in \{1, 2, 3, 4, 5, 6, 7\}$ . Esto nos lleva a creer que la  $\mathbb{E}(Q_n^m)$  está acotada  $\forall n \in \mathbb{N}$ .

**Conjetura 6.5.** *Los grafos de Barabási Albert son  $k$ -Débil-Localmente-Populares, donde  $k$  sólo depende de  $m$ . Más aún,  $k \approx m$ .*

# Capítulo 7

## Conclusión y Trabajo Futuro

La presente memoria se ha enfocado en el estudio de la topología de grafos de gran tamaño que han suscitado interés en la actualidad, tales como la *World Wide Web* o las redes sociales. Se estudió la posibilidad de reconstruir estos grafos utilizando una cantidad de pequeña de información local.

A estas redes, como la *World Wide Web* o las redes sociales, que poseen una topología particular se las agrupa en un sólo concepto, redes *Small World*. Y se modelan utilizando un modelo dinámico introducido por Barabási y Albert [3] que evoluciona en etapas discretas y utiliza la regla de *Preferential Attachment*. Partiendo con un grafo pequeño, en cada etapa se agrega un nuevo nodo que se conecta con  $m$  nodos existentes en el grafo, utilizando la regla de *Preferential Attachment*, en la cual la probabilidad de conectarse con un nodo depende de su grado.

Se presentó un protocolo en 2 rounds en el cual los nodos en la red colaboran para reconstruir el grafo, utilizando un modelo distribuido en el cual cada nodo escribe un mensaje, utilizando sólo su información local, en una pizarra visible para todos.

Se demostró que los árboles de *Barabási-Albert* exhiben una propiedad que se denominó *Popularidad Local Débil*. Se dice que un grafo es *k-Débil-Localmente-Popular* si para todo nodo en la red, el valor esperado de vecinos que tienen grado mayor o igual que él es a lo más  $K$ . Esta propiedad hace posible reconstruir los árboles de *Barabási-Albert* utilizando el protocolo anterior en 2 rounds y con mensajes de tamaño logarítmico.

Posteriormente se conjetura que los grafos generales de *Barabási-Albert* también exhiben la propiedad de *Popularidad Local Débil*. Lo cual, al igual que en el caso de los árboles, haría posible su reconstrucción usando información local de tamaño logarítmico. Esto es potencialmente de gran utilidad debido a que redes de gran tamaño como la *World Wide Web* no se encuentran almacenadas en su totalidad en un sólo lugar o procesador, es por ello que la posibilidad de reconstruirla y así averiguar cualquier propiedad de la red utilizando sólo información local de tamaño logarítmico es un gran avance.

Ahora bien, este trabajo deja abierto el problema de demostrar que en efecto los grafos *Barabási-Albert*, son *k-Débil-Localmente-Populares*, independiente del tamaño del grafo. Esto haría posible, en definitiva, la reconstrucción de cualquier red del tipo *Small World* utilizando información local de tamaño logarítmico.

# Bibliografía

- [1] Steve Lawrence and C. Lee Giles. Searching the World Wide Web. *Science*, 280(5360):98–100, 1998.
- [2] S. Redner. How popular is your paper? an empirical study of the citation distribution. *European Physical Journal B*, 4(2):131–134, August 1998.
- [3] A.-L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272(1-2):173–187, 1999.
- [4] David Peleg. *Distributed computing: a locality-sensitive approach*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.
- [5] Nathan Linial. Locality in distributed graph algorithms. *SIAM J. Comput.*, 21(1):193–201, February 1992.
- [6] Fabian Kuhn, Thomas Moscibroda, and Roger Wattenhofer. What cannot be computed locally! In *In Proceedings of the 23 rd ACM Symposium on the Principles of Distributed Computing (PODC)*, pages 300–309, 2004.
- [7] Florent Becker, Martín Matamala, Nicolas Nisse, Ivan Rapaport, Karol Suchan, and Ioan Todinca. Adding a referee to an interconnection network: What can(not) be computed in one round. *CoRR*, abs/1009.4447, 2010.
- [8] Florent Becker, Adrian Kosowski, Nicolas Nisse, Ivan Rapaport, and Karol Suchan. Allowing each node to communicate only once in a distributed system: shared whiteboard models. In *Proceedings of the 24th ACM symposium on Parallelism in algorithms and architectures*, SPAA '12, pages 11–17, New York, NY, USA, 2012. ACM.
- [9] André Gronemeier. Asymptotically optimal lower bounds on the  $\text{and}$ -multi-party information complexity of the  $\text{and}$ -function and disjointness. In *In STACS 2009*, pages 505–516, 2009.
- [10] Jeff M. Phillips, Elad Verbin, and Qin Zhang. Lower bounds for number-in-hand multiparty communication complexity, made easy. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 486–501. SIAM, 2012.
- [11] Andrew Drucker, Fabian Kuhn, and Rotem Oshman. The communication complexity of distributed task allocation. In *Proceedings of the 2012 ACM symposium on*

- 
- Principles of distributed computing*, PODC '12, pages 67–76, New York, NY, USA, 2012. ACM.
- [12] Ergin Elmacioglu and Dongwon Lee. On six degrees of separation in dblp-db and more. *SIGMOD Rec.*, 34(2):33–40, June 2005.
- [13] Duncan J Watts. *Small worlds: The dynamics of networks between order and randomness*. Princeton University Press, Princeton, NJ, 1999.
- [14] Christof Koch and Gilles Laurent. Complexity and the nervous system. *Science*, 284(5411):96–98, 1999.
- [15] Paul Erdos and Alfred Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hungary. Acad. Sci.*, 5:17–61, 1960.
- [16] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64:025102, Jul 2001.
- [17] R.G. *Discrete Stochastic Processes*. Kluwer, 1996.
- [18] Hosam Mahmoud. *Polya Urn Models*. Chapman & Hall/CRC, 1 edition, 2008.